Going Viral in a Pandemic: Social Media and Allyship in the Black Lives Matter Movement

Vladimir Avetian * Annalí Casanueva Artís †

Sulin Sardoschau ‡ Kritika Saxena §

January 2022
Preliminary Draft

Abstract

How do modern social movements broaden coalitions? Triggered by the viral video footage of George Floyd's murder, the Black Lives Matter (BLM) movement gained unprecedented scope during the pandemic. Using Super Spreader Events in the early stages of the pandemic as a source of exogenous variation at the county level, we find that exposure to COVID-19 mobilized "new allies" to join the movement for the first time. We present evidence consistent with the hypothesis that the pandemic increased the use of social media and particularly *twitter* among those not directly affected by the movement's grievances (i.e. more affluent, whiter and more rural counties), thereby exposing a broader section of the population to BLM-related content and the viral protest trigger. Social media can serve as an effective mobilization tool outside of traditional coalitions.

Keywords: BLM, COVID-19, protest, social media
JEL classification: P16, D7

*Sciences Po: vladimir.avetian@sciencespo.fr
†Paris School of Economics: annali.casanueva@gmail.com
‡Humboldt University : sulin.sardoschau@hu-berlin.de
§Graduate Institute of International & Development Studies: kritika.saxena@graduateinstitute.ch The authors would like to thank Davide Cantoni, Ruben Durante, Sergei Guriev, Alexey Makarin, Lore Vandewalle, Maiting Zhuang, and Ekaterina Zhuravskaya for their useful comments and suggestions. We would also like to thank Marlon Seror for his invaluable inputs. We are thankful to all the participants of the 25th Annual conference of the Society for Institutional and Organizational Economics (SIOE), Households in Conflict (HiCN) conference, the AFSE conference, brown bag seminar at The Graduate Institute of International and Development Studies, the Science Po internal seminar, the Empirical Economics Seminar at Ludwigs Maximilian University Munich, the Development Economics Network Berlin and the Applied Young Economists Webinar for their useful feedback.

1 Introduction

There is a far more representative cross-section of America out on the streets [...]
That didn't exist back in the 1960s. That broad coalition.
- Barack Obama, June 3rd 2020

Social movements are integral to democratic politics and can bring about social, economic and institutional change (Ostrom, 1990; Madestam et al., 2013; Della Porta and Diani, 2015). Protesters take to the streets in order to put pressure on politicians and appeal to the broader public in the hopes of influencing policies that address their grievances. The effectiveness of social movements depends on their ability to mobilize allies and build coalitions, thereby inspiring reform through collective action (Olson, 1989; Della Porta and Diani, 2020).

Traditionally, such protest mobilization was organized at the local level. For instance, the Civil Rights Movement in the 1960s depended heavily on local chapters as decision making, mobilization, coordination and persuasion tools (Morris, 1986). Today, social movements depend less on face-to-face interactions and have shifted their activism into the virtual space.[1] The Black Lives Matter (BLM) movement - successor to the Civil Rights Movement - was born on Twitter in 2013 and relies heavily on social media to communicate and mobilize (Mundt et al., 2018; McKersie, 2021). The #BlackLivesMatter hashtag has become one of the most frequently used hashtags on twitter, peaking at 8.8 million tweets per day in May 2020.[2] Videos on twitter about the murder of George Floyd at the hands of the police officer Derek Chauvin were watched over 1.4 billion times within two weeks.[3] The ensuing protest in May of 2020 were labeled the "largest" and the "broadest" social movement in the history of the United States.[4]

What led to the broadening of the movement's coalition during the pandemic? The determinants of modern social movements - in particular, their ability to mobilize new allies - remain poorly understood. In this paper, we investigate the role of social media in mobilizing new protesters in the context of the BLM movement during the pandemic. We focus on this context for various reasons. The BLM movement *i*) experienced a substantial increase in support, particularly in places that have not traditionally protested for the BLM cause before, *ii*) BLM draws heavily on social media as a persuasion tool and *iii*) was subject to an unexpected protest trigger during the pandemic.

We approach this question in two parts. First, we establish a causal link between exposure to COVID-19 and protest participation at the county level, using Super Spreader Events as a source of exogenous variation. We show that exposure to COVID-19 is associated with an increase in protest behavior but only among those counties that have never protested for a BLM-related cause before (we call these "new allies").[5] Second, we investigate the underlying mechanisms.

[1] As McKersie (2021) notes: "Even though an organization like BLM does not have a constituent base like the CCCO, through which affiliated congregations and neighborhood organizations issued calls for participants, current BLM organizations more than compensate by utilizing the power of social media to mobilize participants for protests."

[2] See PEW Research Center (2020)

[3] See Listing of Twitter Videos with GF and BLM hashtag

[4] See New York Times and Washington Post

[5] Note that for our purposes we define new allies as counties, rather than individuals since we are not able to identify whether individual protesters are joining the movement for the first time and only observe new BLM protests in specific locations.

Specifically, we show that the uptake of twitter and online activities more generally (proxied by new twitter accounts, Google searches, and residential stay) *before* the murder of Floyd was higher among these counties. Moreover, we provide evidence that (instrumented) baseline twitter penetration and new twitter accounts created during the pandemic can - at least in part - account for the broadening of the BLM coalition and increase protest participation. Lastly, we verify our proposed mechanism with survey data and examine various alternative mechanisms.

Based on these results, we hypothesize that the binding constraint of modern social movements no longer lies in the presence of local chapters or social ties on the ground but rather depends on increased exposure to their existing social media content and messaging. Previous work has shown that social media can solve the collective action and coordination problem for individuals already sympathetic to a political cause in the early stages of social media expansion (Enikolopov et al., 2018; Manacorda and Tesei, 2020). In contrast, we focus on the role of social media as a protest mobilization tool that can *broaden* alliances and target regions of the country, and potentially fractions of society, that are less impacted by and aware of the movement's grievances.

Our identification is based on a small "window of opportunity" between the end of March and mid April of 2020 during which COVID-19 was prevalent enough but lock-down stringency lax enough to allow for so called Super Spreader Events (SSE) to occur. These events are characterized by the presence of one highly infectious individual (a super-spreader) and took place mainly at birthday parties, nursing homes or prisons. We exploit cross-sectional variation in the number of SSEs within a 50 kilometer radius from the county border but not within the county 6 weeks prior to the murder of George Floyd to construct our instrument for exposure to COVID-19 at the county level. We include state fixed effects and a vast set of county level controls, most notably, the number of past BLM events between 2014 and 2019, as well as socio-demographic variables and proxies for political leaning and social capital.

A causal interpretation is based on the assumption that - conditional on controls and state fixed effects - SSEs in the neighboring counties 6 weeks before GF's murder only impacted BLM protest in the county through it's spreading of COVID-19. We provide various pieces of evidence that support this assumption. Most importantly, we *i*) show in a placebo test that SSEs do not predict past BLM events, and using LASSO *ii*) we weigh SSEs by their inverse probability of occurrence and *iii*) include a control variable that captures the pre-pandemic protest propensity.[6]

In addition, we provide three alternative identification strategies. First, we construct a different instrument, using large scale mobile phone mobility data by *SafeGraph* to measure touristic flows to one of the largest SSEs in the US - Florida spring break in March 2020. We can identify the home counties and vacation locations of these mobile phone users and construct an index of exposure to spring break returnees to instrument COVID-19 at the county level. Second, we employ a difference in differences approach, for which we scrape information on all similar BLM protest triggers since 2014 to estimate the differential response to a protest trigger before and after the pandemic.[7] Third, we use a LASSO based matching approach, comparing

[6] We describe the LASSO selected model in detail in Appendix B.3

[7] We define protest triggers as deadly force used by police against a Black person that received national media coverage

counties with similar pre-pandemic protest probabilities.

We find robust evidence that exposure to COVID-19 increased BLM protest. This holds for various iterations of our SSE instrument (varying distance, time lag, and cases associated with SSEs) and alternative identification strategies. Specifically, we estimate that a one standard deviation increase in the number of COVID-19 related deaths in the county at the time of GF's murder (approximately 25 deaths per 100K inhabitants), increases the likelihood of a BLM event occurring in the three weeks following the murder by 5%.

We find evidence in line with the hypothesis that the pandemic has mobilized new allies to join the movement. Our baseline result is entirely driven by counties with no prior BLM protest history. In fact, the effect of COVID-19 on the likelihood of observing a BLM protest doubles in size and is more precisely estimated. We also find that it is mostly non-Black, affluent and suburban counties that protest in response to COVID-19, confirming that these new allies come from parts of the country that have typically been less affected by the grievances of the movement.

Next, we move to the dynamics and diffusion of BLM protest, focusing on the sub-sample of counties with no prior BLM events. We look at each week separately and we do not find that our results are driven by any specific week. In addition, we do not find evidence that counties that protest first, inspire subsequent protest in neighboring countries or that counties closer to Minneapolis (the location of GF's murder) are more likely to protest in response to COVID-19. Combined, we take this as evidence that the protests did not ripple through the US in space and time but created a simultaneous uproar throughout the country with counties being more exposed to the pandemic reacting more to the protest trigger.

A battery of robustness checks probes the validity of our results. Our results hold when we exclude coastal counties and states, account for spatial correlation across counties, and use different time frames for the outcome and definitions of the independent variable. In another exercise, we verify that the effect is not driven by a substitution away from some locations to others. It is possible that the pandemic did not broaden but only scattered geographically. We tackle this issue in two ways. First, we consider the structure and scope of protests. A scattering of protests would indicate that BLM protest might increase at the extensive margin (likelihood of a BLM protest) but not at the intensive margin (number of participants and number of events). We don't find a significant decrease in the number of participants, and conversely, even detect an increase in the number of BLM protests. Second, we control for having a traditional protester as a neighbor and having a neighbor that is currently protesting (as well as their respective interactions with instrumented COVID-19). A scattering of protests would imply a higher likelihood of observing a protest for counties with neighbors that are currently or typically protesting. Reassuringly, none of these exercises confirm a pandemic-induced scattering of BLM protests.

In the second part of the paper, we analyse the mechanisms behind the broadening of the BLM coalition. We start by repeating the above analysis, this time using an index of online

on CNN, the Washington Post or has a dedicated Wikipedia page activity as our main outcome variable. Online activity is measured *before* the protest trigger but after the outbreak of the pandemic in the United States (i.e. the first detected case on January 20, 2020 to Floyd's murder on May 25th). We construct this index, using the first principle component of three variables: *i*) the log cumulative number of new twitter accounts, which we obtain by scraping and geo-coding information on the creation date of new twitter accounts at the county level from approximately 45 million tweets, *ii*) Google searches for the term "twitter", hypothesizing that new users will Google the term first to then create an account and *iii*) Google mobility data at the county level, assuming that increased residential stay (time spent at home) as well as lower social, work and leisure mobility is associated with more time spent online. We find that the pandemic increased online activities and that this effect is stronger for the sub-sample of counties that have never protested before.[8]

In a next step, we investigate the direct effect of twitter penetration on protest behavior. Specifically, we test two hypotheses consistent with the idea that an increase in the use of social media has sparked the broadening of the BLM coalition during the pandemic. First, we show that new twitter accounts created during the pandemic but before the murder of Floyd increase the likelihood of observing a BLM protest and more so in places with higher COVID-19 exposure. Second, we use pre-pandemic twitter penetration as a measure for social media presence at the extensive margin and interact it with exposure to COVID-19 as a proxy for social media use at the intensive margin.[9] We find that the pandemic had a substantially larger impact on BLM protest in places with a high baseline twitter penetration. We address the concern that our results could capture underlying factors that drive both twitter penetration and protest participation, replicating the SXSW instrument for baseline twitter penetration used by Müller and Schwarz (2020). Reassuringly, our results hold.

In order to probe the social media mechanism further, we use individual-level survey data. Interpreting these results with caution, we find that individuals living in a county with higher COVID-19 deaths, are more likely to receive news about George Floyd through social media than through other channels.[10] We also find, that COVID-19 exposure is associated with larger sympathy for the movement and higher salience of racial injustice among respondents (controlling for race, gender, education, income, and political leaning) without changing attitudes towards other progressive issues, such as "illegal" immigration.

Lastly, we consider alternative explanations for why exposure to COVID-19 could be associated with higher levels of protest beyond the increase in the use of social media. First, the pandemic may have increased overall salience of racial inequality *before* the murder of Floyd. We test this by interacting COVID-19 with a proxy for disproportional death burden on Blacks and the number of BLM related search terms on Google before the protest trigger. Second, we

[8]We use a normalized index of search activity for term 'twitter' provided by Google Trends. Search activity indices are provided as integers from zero to 100 with unreported privacy threshold. Each observation is a number of the searches of the given term divided by the total searches of the geography and time range, which is then normalized between regions such that the region with the largest measure is set to 100.The Google Trends data is defined on a designated market area (DMA) level.

[9]We measure by sampling and geo-locating all tweets containing the word "the" during one week in December 2019, and interact it with exposure to COVID-19.

[10]The data set does not contain information on the location of the respondent but only whether they live in a low, medium or high COVID-19 county. Therefore, we cannot employ our instrument for exposure to COVID-19.

investigate whether the pandemic has decreased the opportunity cost of protesting. We interact COVID-19 with the unemployment rate at the county level and stringency at the state level. If individuals choose to protest in lieu of going to work or engage in social activities, we should see a larger effect in counties with higher unemployment rates or stricter stringency measures. Third, we look at the effect of COVID-19 on other protests. If the pandemic increased overall agitation and propensity to protest, then we would expect this to also hold for other causes beyond BLM. We show that these channels are unlikely to drive our results.

This paper participates in the nascent literature on the effect of the internet on political outcomes (Falck et al., 2014; Lelkes et al., 2017; Boxell et al., 2017; Campante et al., 2018; Guriev et al., 2019) and the effect of social media on xenophobia, polarization, political preferences, social capital and protests more specifically (Acemoglu et al., 2018; Enikolopov et al., 2018; Bursztyn et al., 2019; Enikolopov et al., 2020; Manacorda and Tesei, 2020; Müller and Schwarz, 2020; Zhuravskaya et al., 2020; Müller and Schwarz, 2021; Fujiwara et al., 2021; Campante et al., 2021). To the best of our knowledge, we are the first to investigate the role of social media in *broadening* political coalitions through persuasion, rather than mobilizing individuals that are already sympathetic to the movements grievances.

Typically, these papers consider (the lack of) protest mobilization as a collective action problem, where access to information reduces coordination costs and therefore increases participation. For instance, Cantoni et al. (2019) and Bursztyn et al. (2021) show in an experimental setting in Hong-Kong that information about other people's turnout encourages individual protest participation and that this has longer-run effects on the propensity to protest if a sufficiently large fraction of the network is mobilized. They conclude that one-time mobilization shocks can have persistent effects on the dynamics of social movements. Most similar to our study, Enikolopov et al. (2020) show that social media helps to solve the collective action problem in a one-shot setting, where the expansion of a social media platform coincides with a contested election in Russia. Similarly, Manacorda and Tesei (2020) exploit the expansion of mobile phone reception in Africa to show that access to information and communication technologies will only increase protest if economic grievances are high and opportunity costs are low (e.g., during economic downturns). In contrast to these papers, we are able to identify for which groups exposure to social media is particularly effective and how it can persuade individuals at the margin.

Our analysis also contributes to a large literature that analyzes the determinants of social movements and protests, ranging from macro level drivers, such as local institutions or socio-economic conditions (Lipsky, 1968; Eisinger, 1973; McCarthy and Zald, 1977; Besley and Persson, 2011; Dube and Vargas, 2013; Berman et al., 2017), to micro level drivers, including individual decision making processes (Ellis and Fender, 2011; Guriev and Treisman, 2015; Sangnier and Zylberberg, 2017) and different aspects of individual and social psychology as well as protest as a collective action problem (Guriev and Treisman, 2015; Sangnier and Zylberberg, 2017; Passarelli and Tabellini, 2017; Cantoni et al., 2019; Enikolopov et al., 2020; Manacorda and Tesei, 2020; González and Prem, 2020; Hager et al., 2020; Bursztyn et al., 2021).

The remainder of the paper is organized as follows. In section 2, we provide some background on the BLM movement and present some motivating evidence. Section 3 describes our main data sources. We present our empirical strategy in section 4 before moving to our main results in section 5. Section 6 will shed light on the underlying mechanisms, focusing on social media. Section 7 concludes.

2 Background and Motivating Evidence

The Black Lives Matter (BLM) movement was born on social media after the acquittal of George Zimmerman in the deadly shooting of a Black teenager, named Trayvon Martin. The movement was founded by three Black activists, Alicia Garza, Patrisse Cullors, and Opal Tometi in July of 2013 with the aim to end systemic racism, abolish white supremacy and state-sanctioned violence (Black Lives Matter, 2020), and more generally, to "fundamentally shape whites attitudes toward blacks" (Mazumder, 2019).

Over the following months, an ever increasing but small number of activists gathered under the hashtag #BlackLivesMatter on Twitter and Facebook. In August of 2014, after a court decision to not indite the responsible police officer in the fatal shooting of Michael Brown in Ferguson, #BLM became one of the most widely used hashtags on twitter (the hashtag was use 1.7 million times in the three weeks following the court decision, as compared to 5000 tweets in all of 2013, see Freelon et al. (2016); Anderson and Hitlin (2016)), manifesting its role as a mainstream social media phenomenon. The shooting of Michael Brown was accompanied by a large wave of protest in the city of Ferguson. The consequences of this shooting rippled through all of American society, generating counter-movements under the hashtag #AllLivesMatter and #BlueLivesMatter and mobilizing protesters (for and against the cause) far beyond the city's borders.

BLM played a crucial role in transforming localized activism into a coordinated movement across various locations within and outside of the United States. The founders state that "[...] when it was time for us to leave, inspired by our friends in Ferguson, organizers from 18 different cities went back home and developed Black Lives Matter chapters in their communities and towns — broadening the political will and movement building reach catalyzed by the #BlackLivesMatter project" (Black Lives Matter, 2020). The *Black Lives Matter Global Network infrastructure* was designed to provide decentralized actors with resources and guidelines to organize protests, receive information about the movement, and coordinate through social media.

In the subsequent years, the BLM movement expanded geographically and demographically, attracting an unprecedented number of participants after the murder of George Floyd in Minneapolis on May 25th 2020. Protesters took to the streets as a video of the murder of George Floyd (GF) went viral on social media, showing how GF suffocated under the choke-hold of police officer Derek Chauvin. The video spurred unrest in Minneapolis but the protests quickly expanded to other parts of the United States, including places that had never engaged in BLM portests before. The number of BLM protest quadrupled in May and June of 2020, compared to previous peaks in 2016 (see Figure 9).

The surge in BLM protests in the spring of 2020 is all the more remarkable as the COVID-19 pandemic was well on its way. At the time of George Floyd's murder almost 100,000 COVID-19 related deaths had been recorded in the United States and the country was just recovering from the first wave of the pandemic (see Figure 2). Tough lockdown and social distancing measures were imposed in many counties to prevent the spread of the pandemic. Average lockdown stringency peaked in the month of May (Hale et al., 2020) and the Center for Disease Control and Prevention urged the public to "remain out of congregate settings, avoid mass gatherings, and maintain distance from others when possible" (CDC, 2020).

A key motivating observation for our study is the exceptionally high level of participation in BLM protests after the murder of George Floyd (see Figure 9). While the outbreak of the pandemic and the peak in BLM protest coincided, the surge in protests may still be driven by counties that were less exposed to the pandemic. If we split the sample into above and below median COVID-19 related deaths at the county level and plot the BLM protest in 2020 in the left panel of Figure 3, we also find a geographical link between exposure to COVID-19 and BLM protest. In the right panel of Figure 3, we plot the evolution of tweets that mention the hashtags *#BLM* or *#BlackLivesMatter*. Using an algorithm that assign tweets to geographic locations, we are able to assign these tweets to counties that experience above and below median COVID-19 related deaths. We find that locations that were more affected by COVID-19 increase their online protest activity. These descriptive plots suggest that - despite the fear of contagion and the stringency of social distancing measures - there is both a temporal and a geographical relationship between COVID-19 intensity and occurrence of BLM protests.

Lastly, we find that - in line with public perception - the BLM movement has broadened in scope. We divide the counties based on counties that always protest for BLM and those that protested for the first time after GF's murder.[11] Figure 4 plots counties that had at least one BLM protest pre-pandemic and also protested after GF's death in black. Counties that recorded their first BLM protest only after GF's murder are shown in green ("new allies"). Our data reveals that the geographic spread of first time protesters does not follow the typical coastal geographic clusters and are spreading across all of the United States. Interestingly, new allies make up half of the counties protesting in the weeks following Floyd's murder.

In sum, this motivating evidence delivers three takeaways. First, the BLM movement has gained unprecedented scope during the pandemic. Second, there is a geographic link between COVID-19 exposure and online and offline BLM protest. Third, a meaningful proportion of protesters in 2020 come from counties that have never protested for a BLM related cause before. We use these observations to guide our empirical analysis.

3 Data

3.1 COVID-19 pandemic

COVID-19 Data on COVID-19 related deaths and cases in the USA at the county level comes from the New York Times. This data set provides the cumulative count of cases and deaths

[11] We use data from *Elephrame* on BLM events between 2014 and 2020 and describe this data set in more detail in the next section and Appendix D.

every day for each county in the USA, starting from January 21, 2020 when the first COVID-19 case was reported in the country. A key limitation of COVID-19 cases data is that it depends on the testing facility and availability of the test kits in the region. We therefore mainly rely on COVID-19 related deaths as a measure of exposure to the pandemic. We also obtain data on daily COVID-19 hospitalizations and deaths by race and ethnicity at the state-level from the Center for Disease Control and Prevention.

Super spreader events We collect data on COVID-19 super spreading events from a project started by independent investigators and researchers from London School of Hygiene and Tropical Medicine (Swinkels, 2020). Data are put together based on news reports of super spreader events and so one key limiting factor is that if the event was not identified as a super spreader event in the media, it is not included in the data set. We overcome this limitation by focusing on one popular super spreading event, which is the Florida spring break (described in appendix B) for our IV. We assign each event to a county. For the whole period, we identify a total of 1074 super spreader events in the USA. Most commonly, events occur in nursing homes, prisons, factories, and retribution or medical centers. Figure C1 shows the distribution of these events by their type and Table C1 provides descriptive statistics about each type of event. This mainly shows that variation for our identification is not limited to one type of event.

Lockdown stringency We use data from the Oxford COVID-19 Government Response Tracker (Hale et al., 2020) to measure the restrictiveness of policy under the current pandemic. Use of this data is inspired by recent work which shows that stringent policies lead to lower mortality, mobility and consequently spread of infection during pandemic (Jinjarak et al., 2020; Askitas et al., 2020). This data provides four key indices (i) an overall government response index, (ii) a containment health index, (iii) an economic support index, and (iv) an original stringency index which captures the strictness of lockdown style policies. Each of this index reports values between 1 to 100 and varies across states and weeks.

3.2 BLM movement and other protests

Black Lives Matter protest This data comes from the crowd-sourced platform Elephrame. It provides information on the place and date of each BLM protest and estimated number of participants, as well as a link to a news article covering the protest. We extracted all protests' records from June 2014 to September 2020 and geo-coded their location.

Other protests We add information on non BLM-related protests from the US Crisis Monitor, a joint project between ACLED and the Bridging Divides Initiative (BDI) at Princeton University, that collects real-time data on different types of political violence and protests in the US from Spring 2020 up to date.

Notable deaths We collect data on all notable black deaths that have occurred in the country since 2014. Not every black death at the hands of the police gets media coverage, something which is crucial for generating public discourse and action. We put together details of deaths of all black people in the hands of the local police authorities that got media coverage. Notable deaths are defined as deaths that got covered in a major national daily like the Washington Post or CNN and/or has a dedicated Wikipedia page.

Use of deadly force by police We obtain this from the collaborative platform Fatal Encounters. They start in 2000 and contain the name, gender, race, and age of each victim and the specific address where the death occurred, among other variables.

3.3 Big Data

Twitter We collect the universe of tweets with BLM related hashtags during the three weeks following the murder of George Floyd. We also collect data to estimate the Twitter usage pre-pandemic per county. To do that, we sample the universe of tweets containing the word "the" in December 2019. Finally, to reproduce the instrument for Twitter usage used by Müller and Schwarz (2020) we collect the list of followers of the account of the SXSW festival, which provided an initial boost to Twitter usage. Tweets and users are geo-located at the county level using the location currently indicated in the users profile.

Google mobility We use data on mobility provided by Google to understand the mechanism of observing protests during pandemics. This data collects information on the time a person spent on certain key mobility tasks like the time spent in parks, being at home, doing groceries, in the transit stations and finally at their workplace (as identified by Google). This information is then aggregated at the county level to measure the aggregate daily mobility.

Safe Graph mobility We rely on two datasets provided by the SafeGraph. Both of them are based on anonymized mobile data. The SafeGraph aggregates data from around 45 million smartphones on the level of US Census Block Groups. With the help of the first dataset, Monthly Patterns (MP), we can answer such questions as: who visited each «point of interest», where they came from and where they go. The set of «points of interests» consists of millions of places such as hotels, restaurants, public parks, malls and other establishments. The MP data allows us to observe home locations on the level of the US Census Block Group, which we can use to construct our variable of tourism flows that happened during March, 2020.

3.4 County level data

Control variables We include *unemployment* data available on a monthly basis at the county level from the Local Area Unemployment Statistics of the US bureau of Labor Statistics and the total population, population by ethnicity, income statistics (such as Black poverty rate and median household income (all in 2018), as well as past Republican vote share (in 2012 and 2016) from the American Community Survey. Data on *community resilience* comes from the United States Census Bureau. These estimates measure the capacity of individuals and households to absorb, endure, and recover from the health, social, and economic impacts of a disaster such as a hurricane or a pandemic.[12] We use a dummy for *rural* counties which is constructed from the

[12] `https://www.census.gov/data/experimental-data-products/community-resilience-estimates.html` For each county the population living under each of 11 risk factors is estimated and these factors are then aggregated into 3 composite risk factors- (i) population with 0 risk factors; (ii) population with 1-2 risk factors and; (iii) population with 3 or more risk factors. These risk factors are based on household and individual's socio-economic and health conditions. For

Office of Management and Budget's February 2013 delineation of metropolitan and micropolitan statistical areas.[13] The measure of *social capital* that we use aggregates the information on the number of various local organizations.[14]

Survey Data We use data from the American Trends Panel survey conducted by Pew Research center to estimate the link between COVID-19 death rates and change in use of social media and public opinion on racial disparities and BLM movement. We analyse data from wave 68 that took place between June 4th and June 10th, 2020. This data set does not include information on the county of the respondent but only the exposure to COVID-19 (categorized in low, medium and high) in their county of residence at the time of the interview.

3.5 Descriptive statistics

Table 1 provides an overview of our main variables. The average likelihood of observing a BLM-related protest at the county level between May 25th and June 14th lies at about 10%. There are 0.26 number of events on average in the three weeks following Floyd's murder and average number of participants is approximately 300 with a maximum of over 320K participants.[15]. The average number of cumulative COVID-19 related deaths is at 34 (or 0.114 per 100K) by May 25th 2020.

In addition, we report detailed summary statistics for the different sub-samples in Table **??**. We show four sub-samples. Those that have never protested for a BLM related cause before the pandemic and do not protest after (the vast majority of 2.636 counties, which is approximately 85% of all counties); those that protest for the first time after the pandemic (N=132) and those that are "traditional" protesters and stop protesting (N=163) and those that continue to protest (N=177). Overall, the first time protesters make up nearly 50 percent of all counties that protested during the pandemic.

Overall, we find that counties that protest after the murder of Floyd experienced a higher exposure to COVID-19. New allies (no event before, has event after) have a significantly lower Black population share than all other sub-samples. Traditional protesters (has events before, has events after) are expectedly more Democratic leaning and have a higher median household income.

---

our analysis we look at populations within each county that are classified as living under 1-2 risk factors and 3 or more risk factors.

[13] 2013 NCHS Urban-Rural Classification Scheme for Counties, Vintage 2012 postcensal estimates of the resident U.S. population. NCHS Urbanization levels are designed to be convenient for studying the difference in health across urban and rural ares. This classification has 6 categories: large "center" metropolitan area (*inner cities*), large "fring" metropolitan area (*suburbs*), median metropolitan area, small metropolitan area, micropolitan area and non-core (nonmetropolitan counties that are not in a micropolitan area).

[14] This includes: (a) civic organizations; (b) bowling centers; (c) golf clubs; (d) fitness centers; (e) sports organizations; (f) religious organizations; (g) political organizations; (h) labor organizations; (i) business organizations; and (j) professional organizations.

[15] The average sets the number of participants in places with no BLM protests as zero

4 Empirical Strategy

4.1 Baseline Estimating Equation

To study the effect of exposure to COVID-19 on BLM, we estimate

$$BLM_c = \beta_0 + \beta_1 Covid_{cs} + \mathbf{X}_c \beta_{\mathbf{X}} + \delta_s + \epsilon_{cs} \tag{1}$$

where Y is a dummy variable for the presence of a BLM protest in county c during the three weeks following the murder of George Floyd.[16] We are interested in the coefficient β_1, which captures the effect of one additional COVID-19 related case per 1000 inhabitants in county c of state s at the time of George Floyd's murder on May 25th 2020. In addition to state fixed effects δ_s, the vector $\mathbf{X}_c$ includes an array of county level controls (we describe all these variables in detail in Table 1). Specifically, we include variables that are associated with the participation in the BLM movement, such as a dummy for urban counties and Black population share as well as poverty rate among Blacks. Most importantly, we also include two major determinants of BLM protest after the murder of George Floyd, namely the number of BLM events before the murder (starting 2014) and the use of deadly force by police (i.e. number of Black people that died during an encounter with the police, excluding suicides, for two time periods: from summer 2014 to 2019 and in 2020 until May 25th). We also control for underlying political and attitudinal factors and socio-economic drivers of protest and social media use, such as the vote share for Republicans in the 2012 and 2016 presidential elections, median hh income, unemployment rate, community resilience, as well as two proxies for social capital (number of civil organizations and number of religious organizations). We cluster standard erros at the state level.

4.2 IV Estimation: Super Spreader Events

A key empirical challenge in ascertaining the causal impact of exposure to COVID-19 on BLM protests is that both occurrences could be driven by third (unobserved) factors. For instance, tight-knit and socially active communities may both increase the spread of the virus and protest more for a BLM related cause. Alternatively, counties that are in favor of lax social distancing rules (and thus more aligned with the president's views at the time) are less likely to engage in BLM protests. Additionally, we may be concerned that BLM protests themselves could spur the onset of COVID-19 infections. While we can assuage the latter concern by measuring COVID-19 exposure at baseline (e.g. before the murder of George Floyd and the onset of BLM protests), we address the former concern with an instrumental variable approach.

We exploit plausibly exogenous variation in the occurrence of Super Spreader Events (SSEs) to causally identify the effect COVID-19 on BLM protest at the county level. Specifically, we construct the IV as the sum of all SSEs that occur within 50 km of the county border but not

[16] We restrict the sample for our main outcome of interest to the three weeks after the death of George Floyd, that is the period from May 25th to June 14th for several reasons: we can capture a large share of the protest behavior (66 percent of BLM protests following GF's murder can be observed in this three week window) while limiting potential confounding factors to arise. Our results hold when we extend this window to six or eight weeks, or reduce it to two weeks (see Table A.4)

within the county until 6 weeks before the murder of George Floyd. We show the geographic spread of our instrument in Figure 5. The first stage is written as:

$$Covid_c = \zeta_0 + \zeta_1 Z_{cs} + \mathbf{X}_{cs}\zeta_{\mathbf{X}} + \gamma_c + \eta_{cs}, \tag{2}$$

$$Z_c = \sum_{m=1}^{t-6} SSE_{csm}^{neighbor} \tag{3}$$

4.2.1 Identifying assumption and instrument validity

The key identifying assumption of this instrument is that - given the set of controls and state fixed effects - SSEs only affect BLM protest through an increase in exposure to COVID-19. We exploit three features of our IV to argue for the validity of the exclusion restriction: i) epidemiological features of super spreader events, ii) the temporal feature, e.g. the short window of opportunity for SSEs to arise , and iii) exposure to SSEs *outside* of the county. We then provide some evidence that supports the plausibility of the exclusion restriction.

Super Spreader Events are defined as the presence of a highly infectious person (a super spreader) in a context where they can infect a large number of people. Super-spreaders are individuals who are an order of magnitude more contagious than others. This phenomenon, well-known in epidemiology, is instrumental in infectious disease spread (e.g. Galvani and May (2005)) and of particular importance for COVID-19, where 70–80% of transmissions can be traced back to just 10–20% of cases (Adam et al., 2020; Endo et al., 2020; Miller et al., 2020). It is important to note that these events do not have to be large gatherings or mass events. The majority of the approximately 1000 SSEs in our data[17] take place in prisons, nursing homes, and at birthday parties. SSE are qualified by the presence of a highly infectious individual. The size of the event is only relevant in as much as it increases the likelihood of a super-spreading individual being present. Therefore, not all mass gatherings are SSEs and not all SSEs are mass gatherings. This is relevant for the exclusion restriction as far as it alleviates concerns about SSEs being a proxy for a county's propensity to organize large public events, including BLM events. In fact, the overwhelming majority of SSEs is recorded – as expected – in the medical care sector (see Figure C1).

Next, we illustrate in Figure 6 that the overwhelming majority of SSEs (solid blue line) occurred between the second week of March and the last week of April. This time-period presented a window of opportunity for SSEs to arise for two main reasons. First, the infectious environment was prevalent enough to bring forth a significant amount super-spreader individuals. Second, lock-down measures were not yet stringent enough (in addition to the lack of public awareness) to restrict group gatherings and encourage mask-wearing. The red dotted line of Figure 6 shows that the increase in the number of new COVID-19 cases coincided with the increase in SSEs. The green dashed line illustrates that state-issued stringency measures (as measured by the stringency index from the Oxford COVID-19 Government Response Tracker)

[17] Data recorded by scientists from the London School of Hygiene and Tropical Medicine

peaked around the time that SSEs leveled off. We argue that during this time window, the occurrence of SSEs was mainly driven by the presence of a highly infectious person, rather than heterogeneity in risk preferences or other underlying factors that could drive both SSEs and BLM protest. We only include SSEs until April 13th 2020 - 6 weeks prior to Geroge Floyd's murder, to account for the fact that SSEs further into the pandemic may be more endogenous.

Lastly, we improve on the plausibility of the exclusion restriction by exploiting SSEs *outside* of the county but not within the county. Specifically, we use the number of SSEs within a 50km (or approximately 30 mile) radius from the county border in which we measure exposure to COVID-19 and BLM. We illustrate the construction of our instrument in Figure 7 using the example of Arizona. To create this instrument, we rely on the geo-location information of the super spreader events and county borders. We indicate as red dots the relevant SSEs used for our IV in this illustrating case. We first draw a circle from the location of each super spreader event and then use the SSEs whose circle intersects with the county boundary to instrument COVID-19 deaths. We argue that SSEs in geographic proximity but not in the county itself are even less likely to affect BLM events in the county other than through COVID-19 exposure.

In Figures 8 and 5 we show the geographical distribution of our instrument across US counties. In Figure 8, we show the number of SSEs 6 weeks prior at the county level. In Figure 5, we show the identifying variation of our instrument, e.g. the number of SSEs in 50 km proximity to the county border until April 13th.

We provide various checks to probe validity of the identifying assumption in Table A.1. Specifically, we investigate whether - despite the features of our instrument described above - SSEs capture some underlying factors that co-determine BLM protest. We always present results for the full sample and the sub-sample of counties that never experienced a BLM protest before. Firstly and importantly, we show that SSEs in neighboring counties do not predict the likelihood of past BLM between 2014 and 2019. If our instrument was related to some unobserved heterogeneity that drives BLM events, we should observe a direct effect of SSEs on past BLM events. Reassuringly, this is not the case.

In addition, we consider the following possibility: the likelihood of being treated by our instrument is not the same across all counties. For instance, counties neighboring large cities may have a higher probability of having an SSE in close proximity. This heterogeneity in the probability of being treated could be related to certain county characteristics that relate to their intrinsic probability to participate in BLM protest. We address this issue by weighting each observation (i.e. each county) by their inverse probability of being treated, using LASSO.[18] In doing so, we give more weight to counties that had a low a-priori likelihood of being treated by the instrument. As shown in Appendix Table B2, this weighting procedure does not change our results, further alleviating concerns about a violation of the exclusion restriction.

Lastly, we expand on the idea of controlling for overall BLM protest probability, beyond the important but simple (discrete) measure of past BLM protests. Using LASSO, we select the subset of relevant county-level variables that determine past BLM events and create a propensity

[18]We describe this approach in more detail in Appendix B.3

score of protesting, based on the selection of these variables.[19] This gives us a continuous measure of protest probability that also covers counties that did not end up protesting for a BLM related cause in the past despite having all the features typically associated with protesters. We include this variable as an additional control in column 4 of Table A.1 and confirm that our results remain robust to the inclusion of this variable. Finally, we group counties in sets of 10, 100 and 1000 with similar propensity to protest and add a group fixed effect (Column 5 to 7 of Table A.1)

Overall, the features of our instrument (epidemiological feature, small window of opportunity, geographic distance) and the empirical exercises examining the plausibility of the exclusion restriction, lend confidence to a causal interpretation of our IV estimation. We will discuss further robustness checks concerning the construction of our instrument in the next section.

4.2.2 First stage results and instrument robustness

We probe the robustness of our instrument in Appendix Table A.2 and A.3 (Appendix A provides a more detailed description of these exercises). We report the first stage coefficient of our preferred specification were the instrument is the cumulative number of Super Spreader Events (SSE) in neighbouring counties within a 50km radius up until 6 weeks prior to the murder of George Floyd. We include the full set of fixed effects and controls as specified in our baseline estimation. In the top panel, we show results for the full sample; in the bottom panel we only focus on the sub-sample of counties with no prior BLM protests. We show both the coefficient for SSE on COVID-19 ("first stage coefficient") and the second stage results (IV: COVID). In this section, we focus on the first stage robustness but preview that our second stage is largely robust to these changes.

In column 1 of Table A.2, we show that one additional SSE increases the number of COVID-19 deaths by 0.93 per 100 000 population for the full sample. The first stage F statistics lie well above the conventional threshold (Kleibergen-Paap F of 36) and find a slightly smaller coefficient and a weaker first stage (Kleibergen-Paap F of 27) for the sub-sample of counties that have never protested before. In columns 2 to 4, we consider the baseline time lag of 6 weeks, i.e. SSEs until April 13th 2020, but vary the distance to the border between 25km and 200km. Our results hold but as expected, the coefficient decreases and the first stage becomes weaker if we move too far from the county border. Next, we use the number of cases associated to SSEs and our results largely hold. Then, we keep the 50km distance but vary the time lag of SSEs until the protest trigger, reducing it to five weeks and expanding it to seven and eight weeks in column 6 to 8 and our results hold as well.

In Appendix Table A.3, we continue our set of robustness checks. Again, we report in column 1 our baseline. In column 2, we exclude SSEs that took place in prison as they may differently impact the public perception of exposure to the pandemic and may also be related to factors that drive BLM protest. Next, in column 3, we also include the number of SSEs in county to account for correlation between neighboring and own SSEs. Then we consider the specific distance to

[19]We describe this approach in more detail in Appendix B.3

the geo-located SSE. We include both the simple linear distance and squared distance to the SSE in columns 4 and 5. Then, we also consider the extent of the overlap of the 50km radius and the counties territory in column 6. Our results remain robust to changes in the definition of the instrument.

5 Main Results

In this section, we present the results for exposure to COVID-19 and BLM protest and show whether this mobilization is in fact due to new allies joining the movement. Additionally, we provide an array of robustness checks and two alternative identification strategies.

5.1 COVID-19 and BLM

We present our main results in Panel A of Table 2, showing 2SLS and OLS results for all counties.[20] We successively introduce control variables, starting with our basic controls, e.g. unemployment (just before the murder of GF) and use of deadly force by police, as well as state fixed effects. We then introduce a large set of controls, most importantly past BLM events, which capture the overall propensity of BLM protest to occur, including all its underlying determinants. Interestingly, the coefficient halves when we include two indicators that capture the county's vulnerability to the pandemic: high risk factor (an indicator that measures how well prepared counties are to dampen the consequences of a health crisis, including health infrastructure, health coverage and pre-conditions) and median household income, which is an indicator for the non-institutional, individual-level economic resource counties have to deal with the pandemic.

Our preferred specification is presented in column 7 and includes the whole set of controls. We find that one additional death per 10 000 population increases the likelihood of at least one BLM event occurring in the three weeks following the death of George Floyd by between 2 and 6 percentage points (p.p.) depending on the specification. An increase of one standard deviation in the number of deaths per thousand increase the likelihood of at least one BLM event occurring by between 5 and 14 p.p.

Throughout all of our estimations (including the robustness checks following in the next section) the IV estimates exhibit larger coefficient compared to the OLS. In the absence of exogenous variation in changes to the COVID-19 infectious environment, the OLS underestimates the role of COVID-19 as a trigger for BLM protests. The bias in the OLS could stem from unobserved within state county-level determinants that drive both BLM protest and lower levels of COVID-19 exposure.[21] This could be due - for instance - to underlying attitudes that disapprove of the Trump administration (beyond those that are captured in the past Republican vote shares and the inclusion of state fixed effects). For instance, more progressive counties, such as Travis county (capital Austin Texas) could be more favorable towards the BLM movement and at the same time more cautious vis a vis the pandemic outbreak and adhere to stricter social distancing

[20] We present in Appendix Table A.6 the reduced form regression with the presence of BLM events as outcome.

[21] Since the treatment (exposure to COVID) is measured before the protest trigger, reverse causality is not the driver behind the difference in magnitude.

rules than Montgomery, Texas. Using mobile phone mobility data, we find that counties that protested for BLM after the murder of George Floyd also decrease their workplace and leisure mobility, while increasing residential stay. This is in line with Dave et al. (2020) that show that BLM protesters adhere more to social distancing measures.

5.2 Allyship: Sub-sample Analysis and Heterogeneity

As shown in Figure 4, we observe that more than half the counties that take to the streets in response to Floyd's murder have never protested for a BLM related cause before. We turn to the sub-samples of counties with and without protest history in Panels B and C of Table 2. Focusing on column 7 of Panel B, we find that the effect doubles in size and is more precisely estimated as compared to the full sample. Specifically, we find that a one standard deviation increase in the number of deaths (25 per 100 000), increases the probability of protesting by 10%. In Panel C, we show that traditional protesters are not responding to the exposure to COVID-19, revealing that our baseline result is entirely driven by these "new allies".

In a second step, we consider socio-demographic and political heterogeneity of counties. In Table 3, we interact exposure to COVID-19 with various baseline characteristics for the full sample of counties and report the coefficient of the interacting variable in the bottom row.[22] Again, we show the baseline effect in column 1 for reference. In columns 2 and 3, we consider heterogeneity by race as recorded in the American Community Survey in the year 2018.[23] The coefficient of the interacting variable indicates that - as expected - counties with a higher non-Black and non-white population share are less likely to protest overall. This is in line with our prior that those who are most affected by the movements grievances are typically protesting. However, counties with a higher non-Black population share (including whites, Hispanics, Asians and "others") are more likely to respond to exposure to COVID-19, confirming the idea of a broadening BLM coalition. Interestingly, if we look at the effect of counties with higher non-white population shares (this includes other minorities beyond Blacks), we do not see the same response, indicating that whites are driving the results in column 2.

In column 4, we move to the economic prosperity of the county, as proxied by the median household income. Richer counties are more likely to protest overall and that these counties are protesting even more in response to the pandemic. This is in line with two, mutually non-exclusive interpretations. First, the literature on protest and conflict highlights that individuals need basic resources to be able to engage in protest in the first place (Bates et al., 2002; Bazzi and Blattman, 2014; Besley and Persson, 2011). Only more affluent households may be able to protest when the resources other households are depleted due to the pandemic. Second, it is possible that - similar to the non-Black counties in the previous columns - richer counties

[22] For this exercise, we prefer to analyze heterogeneity over the full sample as we want to identify the features of new allies versus traditional protesters and therefore. Focusing on the sub-sample of counties with no prior BLM protest would tell us something about differences in socio-demographic and political variables for those who continue to not protest and those that start protesting. These are important for the mechanisms and we repeat this analysis only focusing on counties that never had a BLM protest before in Table C3. However, for now, we want to establish the features of protesting counties more generally.

[23] Self reported racial identification with the categories: white, Black, Asian, Hispanic and "other"

become aware of the racial inequalities through the murder of George Floyd and start to protest in response.

As expected, counties with higher vote shares for Donald Trump in the 2016 elections (vote share Republican reported in column 5) are less likely to participate in BLM protest overall. However, the coefficient of the interaction term is negative, insignificant and very noisy, indicating that the political leaning is less relevant for the likelihood of a BLM event occurring in response to higher exposure to COVID-19. Conditional on state fixed effects this may not be surprising as they capture a large share of the variation in political leaning.

In columns 6 to 9, we consider different classifications for a county's degree of urbanization as defined by the 2013 NCHS Urban-Rural Classification Scheme for Counties. Typically, BLM protest occur in large metropolitan areas, like New York or Los Angeles and less frequently in smaller cities, suburban or rural areas. In column 6, we look at the effect of the pandemic on counties that are not part of a large city. This can reach from fairly big sub-urban areas like Bergen County, New Jersey (adjacent to Bronx County in New York) to small rural areas like Mariposa County, California. Similarly, we also consider suburban counties in column 7, which includes counties like Bergen County, New Jersey (adjacent to Bronx County in New York). Both of these county types experience an increase in BLM protest in response to the pandemic. Unsurprisingly, small towns and rural areas are less responsive to COVID-19 exposure.

Overall, these results confirm our prior that the pandemic mobilized new allies to join the movement for the first time during the pandemic. These allies are comprised of counties with no prior BLM protest history and are characterized by a higher share of non-Black and affluent population in the suburbs and medium sized cities.

5.3 Dynamics and Diffusion of Protests

We turn to the dynamics and diffusion of BLM protest, focusing on the sub-sample of counties that has never experienced a BLM protest before. Our aim is to investigate whether there is a "ripple effect" of protest through space and time. One could imagine that counties with no history of organizing and coordinating BLM protests need more time to set up ad-hoc protests and will only start protesting after some time. It is also possible that these counties are more likely to be inspired by counties in close proximity, imitating and learning from from other BLM events. Another possibility is that they respond more when they are in closer proximity of Floyd's murder (Minneapolis), potentially because they learn more quickly about the event, feel more affected or are inspired by the early protests there.

We test all of these hypotheses in Table 4. In columns 2 to 4, we split our initial outcome into three separate and distinct periods, considering as an outcome the presence of BLM events in week 1, week 2 and week 3. We do not find any evidence that new protesters are starting to protest later on in the observation period.[24] In fact, the effect size decreases substantially when we move to protests that occurred three weeks after the murder of George Floyd.[25]

[24] In the robustness checks section we limit and expand the time frame of the outcome.

[25] Since we measure COVID-19 just before Floyd's murder it is possible that COVID-19 death trajectories have diverged substantially across counties (although deaths are more "sticky" than cases and are not too likely to change substantially

In order to jointly investigate learning effects of new allies more specifically, we exploit information on the exact timing of BLM protests. We construct a dummy variable equal to one if the county has a neighbor that protested first in the three weeks following the murder of George Floyd. If counties learn about the movements' cause and how to organize protest from their neighbors, then we should see a positive effect of having a "first mover" county as a neighbor. We test this in column 5, and do not find evidence of a learning or imitation effect.

Lastly, we analyze the geographic diffusion of protest. The viral video footage of Floyd's murder at the hands of police officer Derek Chauvin in Minneapolis inspired large scale protest in the city, already on May 26th 2020. President Trump infamously tweeted that "when the looting starts the shooting starts", referring to the escalation of protests in Minneapolis on May 27th. Minneapolis quickly became one of the main focal points in the Black Lives Matter movement. In columns 6 and 7 of Table 4, we investigate whether proximity to the earliest and largest protest hub affected the protest behavior of new allies. We use the distance and squared distance to Minneapolis and find no significant impact of proximity to Minneapolis. If anything, counties further away may respond slightly more to COVID-19 exposure, with the caveat that the first stage of the interaction term becomes weak in column 7.

We take these results as evidence that learning or imitation through time and space was not a major determinant of BLM protest diffusion.[26]. In Section 6, we will provide a rationale for why we see no such effect. Specifically, we will argue that an increase in the use of social media before the protest trigger led to protest mobilization, which in turn, is less dependent on learning through time and geographic proximity.

5.4 Broadening vs Scattering of Protests

We consider the possibility that our results are driven by a scattering, rather than a broadening of the BLM protest. Specifically, we may observe new counties protesting for various reasons unrelated to the idea of "new allies". We describe these tests in more detail in Appendix A.3 and briefly reiterate here.

First, the pandemic may have changed the scope and structure of BLM protests. We may expect that counties observe an increase in BLM protest at the extensive margin (the likelihood of observing a protest) but an overall decrease at the intensive margin (number of participants and number of events). In Table 5, we show the total number of participants (column 2) and the number of participants per protest (column 3) does not decrease substantially. Conversely, we even find a significant increase in the number of events in column 4. Combined, we take this as evidence that protests did not become smaller or less frequent in response to the pandemic, indicating that our results are not masking changes in the structure of BLM protests.

Similarly, the pandemic and its restrictions on mobility may have led to a geographic spread of the protest, substituting large protests in cities with smaller protests in suburbs. We may detect "new allies" in our analysis simply be traditional protesters are now protesting in a decentralized

within a 3 week time window). Therefore, COVID-19 deaths at baseline may simply become less predictive of protest behavior through time.

[26] In the Section A.3, we will also investigate the possibility of a substitution effect of protest

way - in suburbs and closer to their homes. While we addressed part of this concern by looking at the structure of BLM protest, the previous exercise cannot capture potential substitution effects between neighboring counties.

We therefore run two additional exercises. First, we create a dummy equal to one if the county has a neighbor that is a traditional protester (i.e. has protested for a BLM related cause between 2014 and 2019). Controlling for this variable in column 5 of Table 5 makes sure that we hold constant the protest propensity of neighbors. In column 6, we also include the interaction between COVID-19 and this dummy variable. If traditional protesters are scattering to suburbs during the pandemic, we would expect the interaction to be driving our result - which is not the case. Similarly, we construct a dummy variable that takes the value one if a neighbor is currently protesting (i.e. at any point during the three weeks following Floyd's murder).Again, we include this variable as a control in column 7 and as an interaction in column 8. Overall, we do not find evidence that the pandemic led to a geographic substitution of protest but rather to a true mobilization of new allies.

5.5 Summary of Robustness Checks

In the previous section, we have provided an array of checks on the plausibility of the exclusion restriction and robustness of our instrument to changes in definition (in the first stage and reduced form). We describe these in more detail in Appendix A. In the top row of each panel of Appendix Tables A.2 and A.3, we show the second stage results and - reassuringly - find consistent results throughout. The coefficient of COVID-19 on the likelihood of BLM protest among counties with no prior BLM history remains positive, significant and similar in magnitude.

We now move on to robustness of our results to changes in sample composition, spatial correlation, and definition of the treatment and outcome variable. First, in column 3 and 4 of Table A.4, we exclude counties and whole states on the coasts and our results hold. We do this for two reasons: first, counties and states next to the ocean will mechanically have fewer neighboring counties with SSEs. Second, when thinking about a "broadening" of the BLM coalition, we want to verify that this does not only apply to states with already progressive leanings. In columns 5 to 7, we shorten the time horizon to 2 weeks and to 6 and 8 weeks after the murder of George Floyd. In the last column we use COVID-19 related cases, instead of deaths. All of these checks yield consistent results. We provide further robustness checks in Table A.5. In column 2, we run an IV Probit regression instead of a 2SLS. In columns 7 and 8, we replace the state clustering by spatial clustering, allowing correlation in a 50 km radius for column 7, and between neighbors for column 8. Columns 9 omits clustering altogether. Reassuringly, our results are not sensitive to these changes.

5.6 Alternative Identification Strategies

We complement our preferred estimation strategy in two ways: i) we design an alternative instrument ii) we exploit the panel dimension of our data set to estimate an instrumented Differences in Differences model and iii) we perform a LASSO matching approach comparing counties with similar pre-pandemic protest probability. We give a brief summary of the approaches here and describe the respective strategies in more detail in Appendix B. All of these approaches confirm the baseline results.

Alternative Instrument: Florida Spring Break

Instead of collecting information on multiple independent SSEs as in the previous section, we now focus on one single, large-scale event that is known to have contributed substantially to the spread of COVID-19, namely the Florida Spring Break in March of 2020 (Mangrum and Niekamp, 2020). We use *SafeGraph* mobile phone data with over 45 million data entries to identify spring break tourists and their home counties and calculate the share of devices that were present at one of the main spring break beaches in March of 2020 relative to all devices of the origin county. As expected, the first stage for this instrument (reported in Table B1 is below the conventional threshold, when we include the full set of controls the F-Stats become weak but the results qualitatively hold.

Difference in Differences: Notable Deaths Sample

We expand our data set and include BLM events at the county-week level starting in 2014. We scrape information on all police related deaths of Blacks since July 2014 that were covered in a major national newspaper like the Washington Post, that received TV coverage by CNN and/or have a dedicated Wikipedia page. We include county and state-week fixed effects to account for all time-invariant county level heterogeneity and common time varying characteristics at the state level. We interact these "Notable Deaths" (time variation) with the instrumented exposure to COVID-19 (county variation). In this instrumented Difference in Differences Approach, we exploit differences in protest behavior following a "notable" death in the presence and absence of COVID-19. We show the results in Table B3 and we find a sufficiently strong first stage and a strongly significant effect consistent with our baseline results.

LASSO Matching: Propensity to Protest

We additionally exploit the previously constructed dataset of notable deaths and BLM events to construct a measure of the propensity of a county to protest after a notable death. The controls used in the model are selected using LASSO logit regression. We use this propensity measure to construct a matching of counties with and without COVID-19 deaths and with a similar propensity to protest. The results (presented in Table B2) are highly significant and consistent with our baseline results.

6 Mechanisms

In this section, we investigate the drivers behind the COVID-19 induced increase in BLM protest. We hypothesize that the pandemic increased the use of social media which in turn lead to the mobilization of a broader set of protesters in response to the highly viral protest trigger: the video footage of George Floyd's murder.

First, we show that the pandemic is indeed associated with a higher online presence (proxied by higher residential stay, new twitter accounts and more Google searches for twitter), particularly among new allies. Then, we show that the increase in protest is driven by those counties with a higher twitter penetration - both at baseline and during the pandemic. Next, we confirm the proposed mechanism by using survey data on social media news consumption about George Floyd and attitudes towards BLM. Lastly, we consider alternative (non-exclusive) mechanisms on pandemic-induced increase in BLM protest, considering *i*) pandemic-induced salience of racial inequality *ii*) lower opportunity costs of protesting and *iii*) increased overall agitation and propensity to protest.

6.1 COVID-19 and the Use of Social Media

So far, we have shown that exposure to the COVID-19 pandemic increased BLM protest. A key hypothesis that we test in this section is that this increased activity is due to increased use of to social media during the pandemic. We show some descriptive figures in Appendix Figure C2 and C3 to motivate this hypothesis. We see that in the period prior to the protest trigger mean stringency (as proxied by the Oxford Government Response Tracker) increased substantially. Stringency measures mostly included recommendations to socially distance (and interestingly, mask wearing recommendations - a sub-category in this index - only started many weeks later). In C3, we use Google mobility data and show that residential stay increased, whereas other types of mobility (particularly, work, transit, and retail) decreased substantially. We believe that the period between March and May coincided with a decrease in social activities and increase in online activities, which we measure more explicitly in the following.

We create an index of "online activity" that comprises the first principle component of three variables: *i*) the log cumulative number of new twitter accounts, which we obtain by scraping and geo-coding information on the creation date of new twitter accounts at the county level from approximately 45 million tweets. *ii*) The normalized index of search activity for term 'twitter' provided by Google Trends, hypothesizing that new users will Google the term first to then create an account. The Google Trends data is defined on a designated market area (DMA) level. *iii*) Google mobility data at the county level, assuming that increased residential stay (time spend at home) as well as lower social, work and leisure mobility is associated with more time spent online. All of these variables are measured between January 2020 and May 24th 2020, i.e. after the outbreak of the pandemic but before the murder of George Floyd. We limit the observation period, such that the BLM events themselves do not impact online activity but we are still able observe the pandemic-induced increase in online activity.

In Table 7, we show the results for the full sample (Panel A), new allies (Panel B) and traditional protesters (Panel C). Again, we use the instrumented exposure to cumulative COVID-19 deaths until May 24th as a main explanatory variable. In column 1, we confirm that the pandemic has led to an increase in online activity as measured by our online activity index for all of the three sub samples. Notably, the effect is largest for the subset of counties with no prior BLM protest history and the magnitude of the effect is twice as large as the effect for traditional protesters.

We then zoom into the specific sub-components of the index and find in column 2 that increased exposure to the pandemic had no effect on the raw number of new twitter accounts created until May 24 (just before George Floyd's murder) for the full sample, or the sample of traditional protesters but is large and significantly positive for the sub-sample of new allies. When we consider the log of new twitter accounts in column 3, we find an even stronger effect for the sub-sample of new allies. Focusing on twitter search terms on Google as an additional proxy for the use of twitter in column 4, we find that - again - search terms only significantly increased among new allies. Lastly, we show residential stay, using Google mobility data at the county level in the month leading up to George Floyd's murder and find that for all samples there has been an increase in residential stay - and more so among new allies. We assume that higher residential stay is likely associated with higher online activity.

Consistent with our prior, we find that the pandemic has increased online activity and particularly the use of twitter- but only among those counties that never protested for a BLM related cause before. We investigate the effect of twitter further in the next section.

6.2 Twitter and BLM protest

The literature on the effect of social media on protest and other political outcomes typically exploits the geographic expansion of access to social media in contexts where they are not yet widely available (Enikolopov et al., 2020; Manacorda and Tesei, 2020; Müller and Schwarz, 2020). In our setting, we consider COVID-19 as a shock to the use of social media at the intensive margin, in locations where these platforms have been available for over a decade.

In the previous section, we have established that the pandemic is associated with higher online activities. However, it is possible that some counties among those that never protested before access social media more frequently during the pandemic but that those are not the same counties that also start protesting. In this section, we establish a more direct link between online activity, particularly twitter usage, on protest behavior.

Specifically, we take twitter penetration at pre-pandemic baseline in December of 2019 (we detail the construction of this variable in Appendix D) and twitter penetration during the pandemic and interact those variables with (instrumented) exposure to COVID-19 as a measure of online activities at the intensive margin. We caveat now that baseline twitter penetration may be related to unobserved factors that co-determine BLM protest. Additionally, new twitter accounts are a bad control as they are co-determined by exposure to COVID-19. We will address this point in the subsequent analysis but focus, for now, on the following heterogeneity. We estimate a second stage regression of the form:

$$
\begin{aligned}
BLM_{cs} = & \beta_0 + \beta_1 \widehat{\text{Covid}_c} + \beta_2 \, Twitter_c \\
& + \beta_3 \widehat{\text{Covid}_c \times Twitter_c} \\
& + \mathbf{X}_c \beta_{\mathbf{X}} + \delta_s + \epsilon_{cs}
\end{aligned}
\tag{4}
$$

where $Twitter_c$ is either (i) the number of users posting about BLM registered in 2020 before May 24 in county c of state s, or (ii) the number of users from the county observed in a sample of tweets collected on December 2019. The logarithm of this number (plus one, to avoid missing values) is interacted with COVID 19 deaths.[27] We instrument COVID-19 deaths and their interaction with users by SSEs and their interaction with $Twitter_c$.

The results, presented in columns 1 and 2 of Table 7 for the sample of counties that didn't have a BLM protest before Floyd's murder, show a positive and significant coefficient for both interactions, meaning that higher Twitter penetration in the county at pre-pandemic baseline and higher number of new users during the pandemic are both associated with more BLM protests in response to COVID-19.

As previewed above, these results cannot be interpreted causally: while we have an instrument for COVID-19, the number of pre-existing and new Twitter users is endogenous and potentially correlated with the error term. Even with the fixed effects and various controls, twitter usage at baseline could be driving BLM protest differentially for counties with higher COVID-19 exposure. We focus only on pre-pandemic

In order to address this concern, we instrument pre-pandemic twitter penetration in December of 2019.[28] Sepcifically, we reproduce the SXSW instrument for Twitter usage described by Müller and Schwarz (2019). SXSW (South by Southwest) is a yearly festival taking place in Austin, Texas. During the March 2007 edition, Twitter was heavily promoted, leading to a rapid increase in the social network's popularity. In order to reproduce this instrument, we collect the location of all followers of the @SXSW account of the South by Southwest festival and the date they joined Twitter.

The dataset we end up with is not entirely identical: some users created on or before March 2007 might have started or stopped following SXSW later. They might also have changed their location between the time Müller and Schwarz collected their dataset and when we collected ours (2019 versus November 2021). Finally, our geolocation method might be different.[29]

Following Müller and Schwarz (2020), we compute for each county the number of followers whose account was created in March 2007 and the number of users whose account was created before this date. With our data collection and user localization strategy, this leads to users being located in 172 counties, only 67 of which did not have BLM events before (Müller and Schwarz

[27] We use the logarithm instead of the actual number of Tweets to overcome potential problems with outliers

[28] We cannot use the same instrument for new Twitter accounts created during the pandemic as pre-existing users are very likely to have a direct effect on BLM events without involving the creation of new accounts.

[29] We automatically geocode the location given by the user using Nominatim, as described in the Data section. Müller and Schwarz (2019) do not detail their geolocation method. Fujiwara et al. (2021) indicates that 58% of users that joined between 2006 and 2008 are geocoded; we attribute 52% of users to US counties (excluding imprecise locations and locations outside the US)

find 155 affected counties). To increase the number of treated counties, and thus the power of our identification, we also consider users in neighboring counties created during this period: assuming that Twitter presence diffuses in part along geography (again following the Müller and Schwarz approach), these counties should also have a higher number of Twitter users. We find 817 such counties, 618 of which did not have a BLM protest before.

We estimate the number of observed Twitter users in December 2019 using the number of users that joined during SXSW controlled by the number of SXSW followers that joined before,[30] with the following regression:

$$\begin{aligned}\log(1 + \text{Users}_{sc}) =& \xi_0 + \xi_1 \log(1 + \text{SXSWUsers}_{sc}) + \xi_2 \log(1 + \text{PreSXSWUsers}_{sc}) \\ &+ \mathbf{X}_{cs}\xi_{\mathbf{X}} + \gamma_s + \eta_{cs} \end{aligned} \tag{5}$$

where SXSWUsers_{sc} is the number of SXSW followers who created their account in March 2007 in the county and neighboring counties, and PreSXSWUsers_{sc} is the number of SXSW followers in the county and neighboring counties that created their account before March 2007.

The results of this first stage regression are reported in Appendix Table C5. The coefficient of SXSW users is positive and highly significant, and the first stage is strong (F = 13.02). We re-run the above specification, this time instrumenting pre-existing Twitter users by the SXSW instrument. The results for the second stage are presented in column 3 of Table 7. The coefficient of the interaction is positive and significant. We report per-coefficient F statistic of weak identification following Sanderson and Windmeijer (2016): while the coefficient of COVID-19 is only weakly identified in this case, the effect of its interaction with pre-existing users is strongly identified.We also report the reduced form regression Appendix Table **??**. The coefficient of SSEs interacted with SXSW users is also significant and positive. Thus, Twitter penetration in the county have a positive effect on the reaction to COVID-19 deaths, confirming that social media plays a role in triggering protests.

6.3 Survey on News Consumption and Attitudes towards BLM

In this section we probe the social media mechanisms further by exploiting individual-level survey data. We ask whether exposure to COVID-19 at the individual level caused a shift in news consumption away from traditional media towards social media. We then investigate whether this shift is accompanied by a change in attitudes towards Blacks and the Black Lives Matter movement more generally.

It is important to note, that a causal interpretation of these results not possible as we do not have information on the precise location of the respondent; we only have information on the severity of exposure to COVID-19 at their county of residence, at the time of the interview in June 2020. However, the rich set of individual-level controls and placebo checks assuage concerns about omitted variable bias.

We use survey data from the Pew research center to conduct individual-level multivariate

[30]This variable controls for the interest in SXSW festival and also acts as a proxy control for the general interest in Twitter in the county.

regressions on different outcomes, controlling for various characteristics of the respondent: race, whether or not they live in a metropolitan area, gender, age, education, income and whether or not they lean towards the democratic party. Table 8 shows the results. Column 1 - 3 investigate the intensity and form of news consumption in the context of George Floyd's murder. Higher levels of COVID-19 are positively and significantly associated with more news consumption about George Floyd and more social media news consumption about George Floyd. In column 3, we show that individuals in counties with higher COVID-19 exposure also consume relative more news about Geroge Floyd on social media, confirming a change in the information set - or at least their source.

Then, we analyze whether this change in mode of news consumption is accompanied by a change in attitudes. In column 4, we find that individuals are more likely to report that higher hospitalization rates of Blacks during the pandemic is caused by circumstances beyond their control, rather than personal choices or lifestyle. Respondents are also more likely to agree with the statement that the BLM protest arises because of structural racism and not as an excuse for criminal behavior. In order to rule out that exposure to COVID-19 in the earlier stages of the pandemic is just a proxy for more progressive leaning counties, we use an additional question that deals with an unrelated progressive issue: legal status for undocumented immigrants. Individuals living in counties with higher exposure to COVID-19 are not more likely to grant more rights to undocumented immigrants, alleviating some of concern on unobserved heterogeneity.

6.4 Alternative mechanisms

We have established that social media use instigated by the COVID-19 pandemic broadened the coalition for BLM movement by bringing in new allies. These new allies were more likely to be non-Black, rural and affluent. In Table 9, we now look at whether there are other mechanisms through which the pandemic could have influenced BLM protests, focusing on the sub-sample of new allies. We show equivalent results for the full sample in Appendix Table C4.

The first alternative mechanism we test is a rise in the salience of racial inequality due to the pandemic itself and not through exposure to BLM related content online. For instance, an a-priori indiscriminate virus should affect Whites and Blacks equally but if racial disparities in death rates arise, then people may be more inclined to believe in systemic disadvantages for the Black community. We therefore hypothesize that counties facing higher proportion of Black deaths due to COVID-19 are more likely to protest after the trigger of George Floyd's death. Column 1 of Table 9 shows that counties in the presence of COVID-19 are not more likely to protest due to increased death burden of Blacks. Additionally, we check whether new allies showed more interest in BLM related issues *before* the murder of George Floyd. We test this in column 2, using BLM search terms on Google in the month leading up to George Floyd's murder. We do not find that interest in racial injustice increased before the protest trigger.

The next channel that we test is the opportunity cost channel. It is possible that new allies joined the movement, particularly more affluent and white counties, because they had a lower opportunity cost of protesting during the pandemic. We proxy lower opportunity costs in two ways: first, economic opportunity costs using the unemployment rate before the protest trigger and second, social opportunity costs, e.g. stringency of social distancing measures at the state level. Columns 3 and 4 of Table 9 show the result for this channel. We find that counties experiencing higher unemployment or have stricter lock-down measures are not more likely to protest.

Lastly, we investigate whether COVID-19 has generally increased agitation in the public space. It is possible that these new allies just protest more in general and not because they have been exposed to new content and messaging online. We therefore look at the effect on other protests, using ACLED protest data. We exclude BLM-related protests from this data set and expand the observation period to 3 months post George Floyd to make sure, we do not capture a substitution effect between BLM protest and other protests right after the BLM protest trigger. We report the results in column 5 and do not find an effect of COVID-19 on other protests. Interestingly, when looking at a sub-set of protest, namely COVID-19 related protests (which are largely comprised of anti-mask protests), we do not find any evidence for a pandemic induced increase.

Taken together, we find convincing evidence that the pandemic induced increase in social media usage which in turn led to the mobilization of new counties. While we cannot fully rule out that other mechanisms are at play simultaneously, we believe that they are unlikely the driver of our results.

7 Conclusion

Protests are an important tool to bring about social change and hold politicians and institutions accountable. In this paper, we show that a higher exposure to the COVID-19 pandemic lead to a higher level of protest in reaction to the murder of George Floyd in regions that had not protested for Black Lives Matter before, and explore possible factors that help explain this expansion of the BLM coalition during the pandemic.

We show that a key mobilizing factor was an increase in the use of social media during the pandemic. Counties that had never protested before the pandemic experienced an increase in Twitter penetration and overall online activity during the pandemic. We support these results with survey evidence and rule out competing mechanisms.

Our research highlights the importance of social movements' online presence. Changes in access to social media may increase political mobilization for those at the margin. However, our research also ties into the potential drivers of an increasing political polarization in the United States. If this effect is symmetric across the ideological spectrum, we may expect similar forms of political mobilization in response to other protest triggers, as the attack on the Capitol on January 6, 2021 illustrates.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Abouk, R. and B. Heydari (2020). The immediate effect of covid-19 policies on social-distancing behavior in the united states. Public Health Reports, 0033354920976575.

Acemoglu, D., T. A. Hassan, and A. Tahoun (2018). The power of the street: Evidence from egypt's arab spring. The Review of Financial Studies 31(1), 1–42.

Adam, D., P. Wu, J. Wong, E. Lau, T. Tsang, S. Cauchemez, G. Leung, and B. Cowling (2020). Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infections in hong kong.

Anderson, M. and P. Hitlin (2016). The hashtag# blacklivesmatter emerges: Social activism on twitter. Pew Research Center 15.

Askitas, N., K. Tatsiramos, and B. Verheyden (2020). Lockdown strategies, mobility patterns and covid-19. arXiv preprint arXiv:2006.00531.

Bates, R., A. Greif, and S. Singh (2002). Organizing violence. Journal of Conflict Resolution 46(5), 599–628.

Bazzi, S. and C. Blattman (2014). Economic shocks and conflict: Evidence from commodity prices. American Economic Journal: Macroeconomics 6(4), 1–38.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017, June). This Mine Is Mine! How Minerals Fuel Conflicts in Africa. American Economic Review 107(6), 1564–1610.

Besley, T. and T. Persson (2011). The logic of political violence. The quarterly journal of economics 126(3), 1411–1445.

Black Lives Matter (2020). Home - black lives matter. https://blacklivesmatter.com/. Accessed: 2020-10-31.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. Proceedings of the National Academy of Sciences 114(40), 10612–10617.

Bursztyn, L., D. Cantoni, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2021). Persistent political engagement: Social interactions and the dynamics of protest movements. American Economic Review: Insights 3(2), 233–50.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social media and xenophobia: evidence from russia. Technical report, National Bureau of Economic Research.

Campante, F., R. Durante, and F. Sobbrio (2018). Politics 2.0: The multifaceted effect of broadband internet on political participation. Journal of the European Economic Association 16(4), 1094–1136.

Campante, F. R., R. Durante, and A. Tesei (2021, September). Media and social capital. Working Paper 29230, National Bureau of Economic Research.

Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2019, May). Protests as Strategic Games: Experimental Evidence from Hong Kong's Anti-Authoritarian Movement. Quarterly Journal of Economics 134(2).

CDC (2020, March). Risk Assessment and Management. Technical report, Centers for Disease Control and Prevention.

Colella, F., R. Lalive, S. O. Sakalli, and M. Thoenig (2019, August). Inference with Arbitrary Clustering. IZA Discussion Papers 12584, Institute of Labor Economics (IZA).

Dave, D. M., A. I. Friedson, K. Matsuzawa, J. J. Sabia, and S. Safford (2020). Black Lives Matter Protests, Social Distancing, and COVID-19. Technical report, National Bureau of Economic Research.

Dave, D. M., D. McNichols, and J. J. Sabia (2021). Political violence, risk aversion, and non-localized disease spread: Evidence from the us capitol riot. Technical report, National Bureau of Economic Research.

Della Porta, D. and M. Diani (2015). The Oxford handbook of social movements. Oxford University Press.

Della Porta, D. and M. Diani (2020). Social movements: An introduction. John Wiley & Sons.

Dube, O. and J. F. Vargas (2013). Commodity price shocks and civil conflict: Evidence from colombia. The review of economic studies 80(4), 1384–1421.

Eisinger, P. K. (1973). The conditions of protest behavior in american cities. The American Political Science Review 67(1), 11–28.

Ellis, C. and J. Fender (2011). Information cascades and revolutionary regime transitions. Economic Journal 121(553), 763–792.

Endo, A., S. Abbott, A. J. Kucharski, S. Funk, et al. (2020). Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. Wellcome Open Research 5(67), 67.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social media and protest participation: Evidence from russia. Econometrica 88(4), 1479–1514.

Enikolopov, R., M. Petrova, and K. Sonin (2018). Social media and corruption. American Economic Journal: Applied Economics 10(1), 150–74.

Falck, O., R. Gold, and S. Heblich (2014). E-lections: Voting behavior and the internet. American Economic Review 104(7), 2238–65.

Freelon, D., C. McIlwain, and M. Clark (2016). Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice. Technical report, Center for Media & Social Impact.

Friedson, A. I., D. McNichols, J. J. Sabia, and D. Dave (2020). Did california's shelter-in-place order work? early coronavirus-related public health effects. Technical report, National Bureau of Economic Research.

Fujiwara, T., K. Müller, and C. Schwarz (2021). The effect of social media on elections: Evidence from the united states. Technical report, National Bureau of Economic Research.

Galvani, A. P. and R. M. May (2005). Dimensions of superspreading. Nature 438(7066), 293–295.

González, F. and M. Prem (2020). Police repression and protest behavior: Evidence from student protests in chile. Available at SSRN 3705486.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2019). 3g internet and confidence in government. Available at SSRN 3456747.

Guriev, S. and D. Treisman (2015). How modern dictators survive: an informational theory of the new authoritarianism. Technical report, National Bureau of Economic Research.

Hager, A., L. Hensel, J. Hermle, and C. Roth (2020). Political activists as free-riders: Evidence from a natural field experiment.

Hale, T., S. Webster, A. Petherick, T. Phillips, and B. Kira (2020). Oxford covid-19 government response tracker. Blavatnik School of Government 25.

Heckman, J. J. (1987). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. University of Chicago Center for Mathematical studies in Business and Economics.

Ince, J., F. Rojas, and C. A. Davis (2017). The social media response to black lives matter: how twitter users interact with black lives matter through hashtag use. Ethnic and racial studies 40(11), 1814–1830.

Jinjarak, Y., R. Ahmed, S. Nair-Desai, W. Xin, and J. Aizenman (2020). Accounting for global covid-19 diffusion patterns, january-april 2020. Technical report, National Bureau of Economic Research.

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of econometrics 95(2), 391–413.

Lasry, A., D. Kidder, M. Hast, J. Poovey, G. Sunshine, K. Winglee, N. Zviedrite, F. Ahmed, K. A. Ethier, C. P. H. L. Program, et al. (2020). Timing of community mitigation and changes in reported covid-19 and community mobility—four us metropolitan areas, february 26–april 1, 2020. Morbidity and Mortality Weekly Report 69(15), 451–457.

Lelkes, Y., G. Sood, and S. Iyengar (2017). The hostile audience: The effect of access to broadband internet on partisan affect. American Journal of Political Science 61(1), 5–20.

Lipsky, M. (1968). Protest as a political resource. The American Political Science Review 62(4), 1144–1158.

Madestam, A., D. Shoag, S. Veuger, and D. Yanagizawa-Drott (2013). Do political protests matter? evidence from the tea party movement. The Quarterly Journal of Economics 128(4), 1633–1685.

Manacorda, M. and A. Tesei (2020). Liberation technology: Mobile phones and political mobilization in africa. Econometrica 88(2), 533–567.

Mangrum, D. and P. Niekamp (2020). Jue insight: College student travel contributed to local covid-19 spread. Journal of Urban Economics, 103311.

Mazumder, S. (2019, May). Black Lives Matter for Whites' Racial Prejudice:Assessing the Role of Social Movements in Shaping Racial Attitudes in the United States. Technical report, Harvard University.

McCarthy, J. D. and M. N. Zald (1977). Resource mobilization and social movements: A partial theory. American journal of sociology 82(6), 1212–1241.

McKersie, R. B. (2021). The 1960s civil rights movement and black lives matter: Social protest from a negotiation perspective.

Miller, D., M. A. Martin, N. Harel, T. Kustin, O. Tirosh, M. Meir, N. Sorek, S. Gefen-Halevi, S. Amit, O. Vorontsov, et al. (2020). Full genome viral sequences inform patterns of sars-cov-2 spread into and within israel. medRxiv.

Morris, A. D. (1986). The origins of the civil rights movement. Simon and Schuster.

Müller, K. and C. Schwarz (2019). From hashtag to hate crime: Twitter and anti-minority sentiment. Available at SSRN 3149103.

Müller, K. and C. Schwarz (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. Available at SSRN 3149103.

Müller, K. and C. Schwarz (2021). Fanning the flames of hate: Social media and hate crime. Journal of the European Economic Association 19(4), 2131–2167.

Mundt, M., K. Ross, and C. M. Burnett (2018). Scaling social movements through social media: The case of black lives matter. Social Media+ Society 4(4), 2056305118807911.

Olson, M. (1989). Collective action. In The invisible hand, pp. 61–69. Springer.

Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge university press.

Passarelli, F. and G. Tabellini (2017). Emotions and political unrest. Journal of Political Economy 125(3), 903–946.

Rosenbaum, P. R. and D. B. Rubin (1983, 04). The central role of the propensity score in observational studies for causal effects. Biometrika 70(1), 41–55.

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. Journal of Econometrics 190(2), 212–221. Endogeneity Problems in Econometrics.

Sangnier, M. and Y. Zylberberg (2017, August). Protests and trust in the state: Evidence from African countries. Journal of Public Economics 152, 55–67.

Swinkels, K. (2020). Covid-19 superspreading events database.

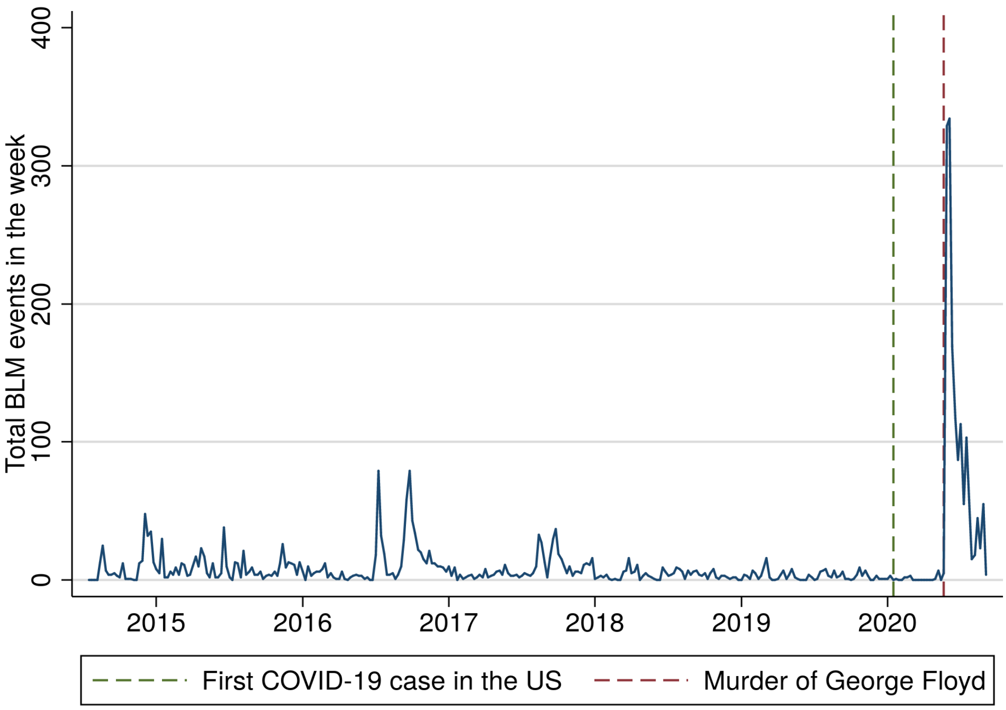
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58(1), 267–288.

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. Annual Review of Economics 12, 415–438.

Figures and Tables

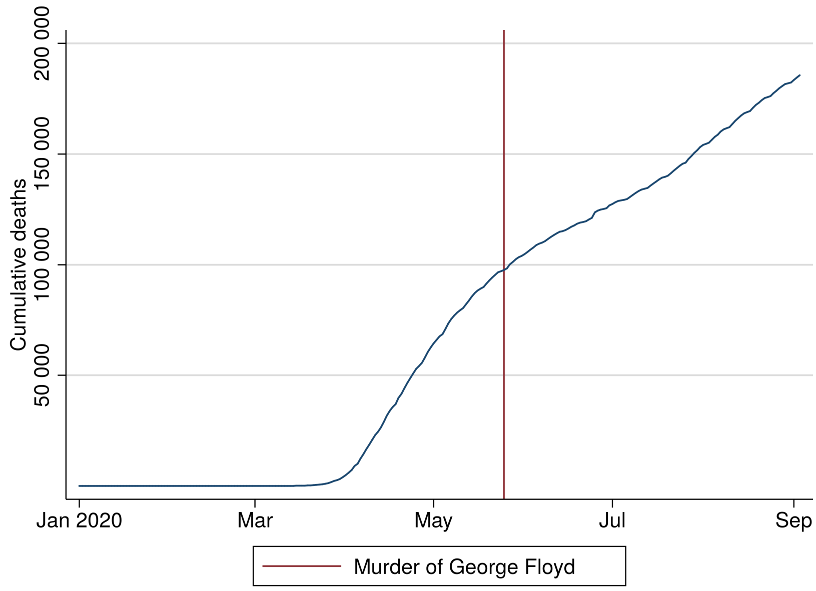
Figure 1: BLM events over time



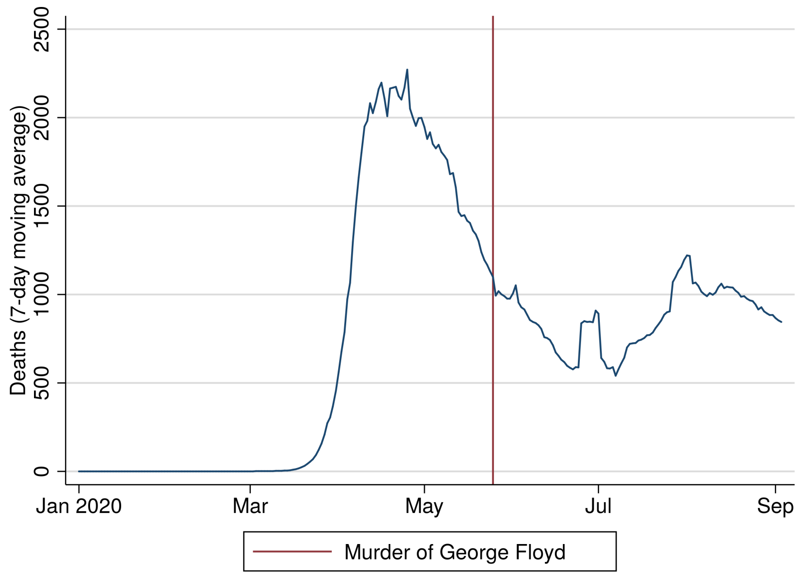
Note: Number of BLM events per week in the US from June 2014 to September 2020. The green line denotes the week of the first confirmed COVID-19 case in the US (January 21, 2020), and the red line denotes the week of the murder of George Floyd (May 25, 2020).

Figure 2: COVID-19 deaths and timing of GF's murder

(a) Cumulative deaths



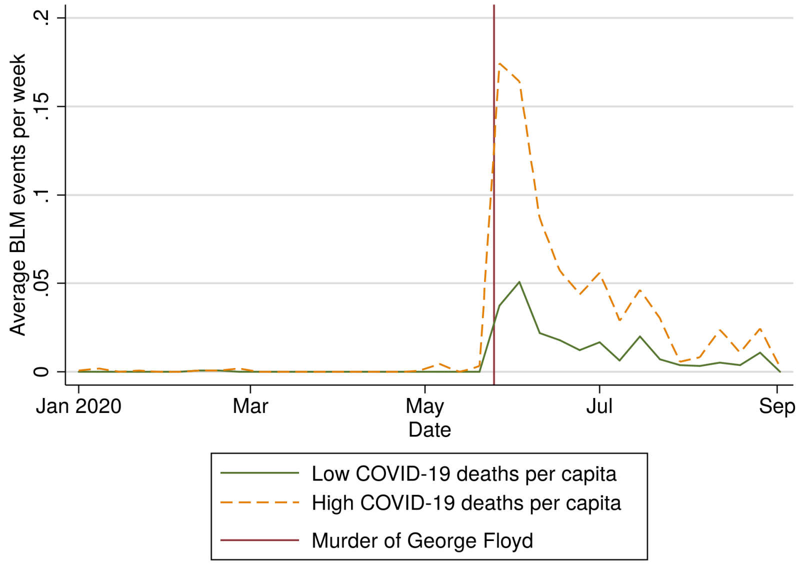
(b) New deaths



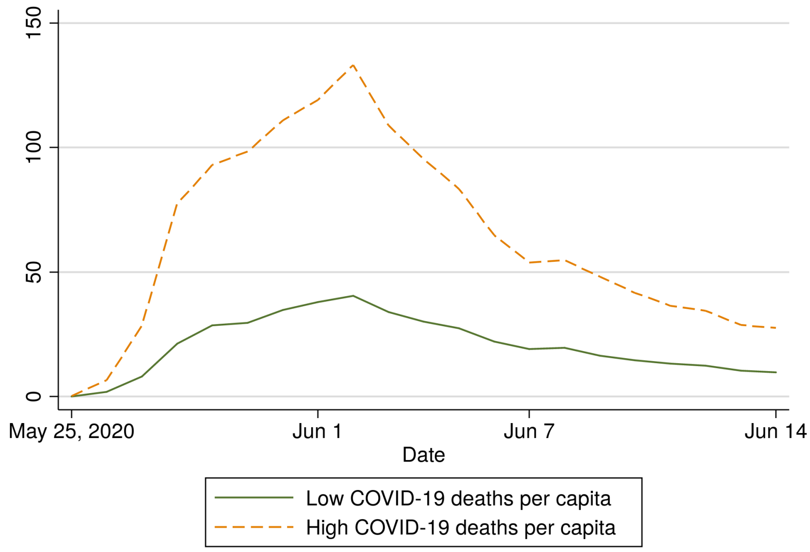
Note: Number of cumulative COVID-19 deaths and daily new COVID-19 deaths in the US between January and September 2020. New COVID-19 deaths are presented as a 7-day moving average. The red line denotes the day of the murder of George Floyd (May 25, 2020).

Figure 3: BLM events and tweets in counties with above and below median COVID-19 deaths per-capita

(a) Average BLM protests per week

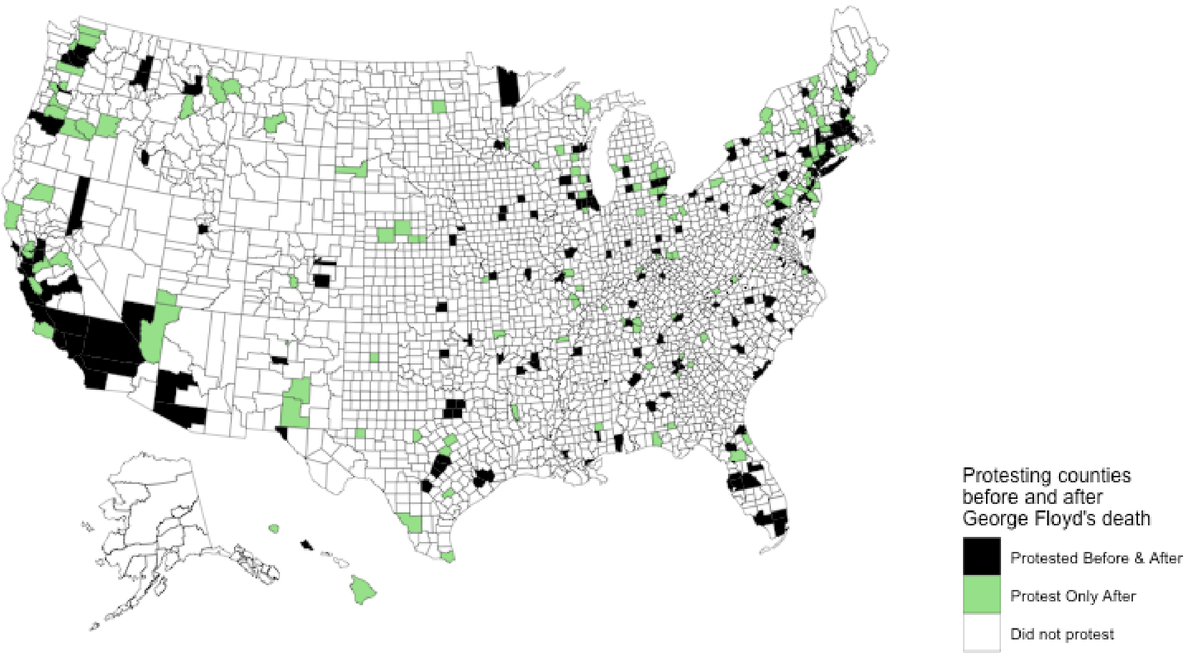


(b) Average tweets mentioning BLM per day



Note: Evolution of two variables over time in counties with below and above median COVID-19 deaths per capita. Subgraph (a) presents the average number of BLM protests per week between January and September 2020. The red line represents May 25, 2020, the day of the murder of George Floyd. Subgraph (b) presents the average number of daily tweets mentioning "BLM" or "Black Lives Matter" from May 25, 2020 (date of the murder of George Floyd) to June 14.

Figure 4: Spatial distribution of US counties based on their BLM protest activities before and after George Floyd's murder



Note: Own visualization based on data from *Elephrame*. This map represents whether US counties that protested in the three weeks following the murder of George Floyd (May 25 to June 14, 2020) already held a BLM protest before the murder of George Floyd. Counties in black protested both before and after the murder of George Floyd. Counties in green are new allies, whose first BLM protest was after George Floyd's murder. Counties in white did not protest after the murder.

Figure 5: Variation across counties of our instrument capturing SSE within 50 kms radius of the county but excluding SSE within the county

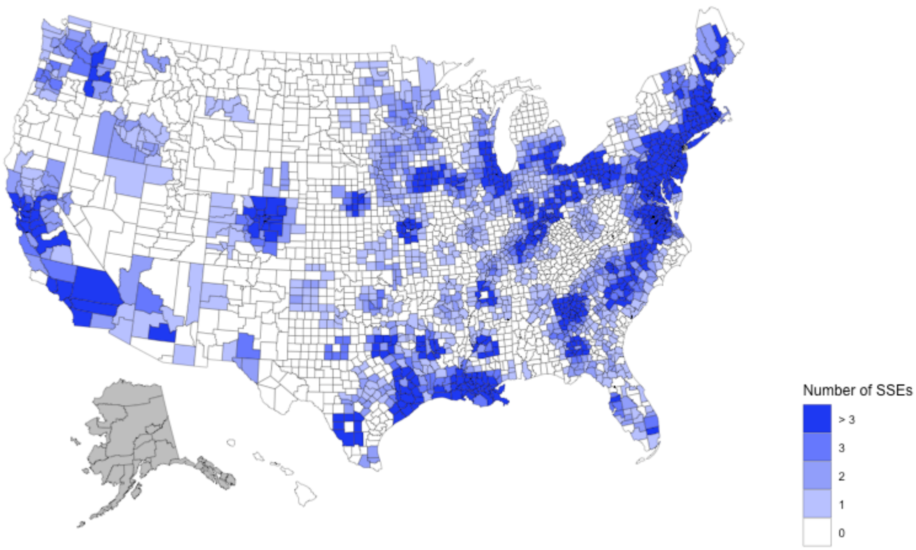


Figure 6: Evolution of Super-Spreader Events, average state-level stringency and number of new COVID-19 cases (daily)

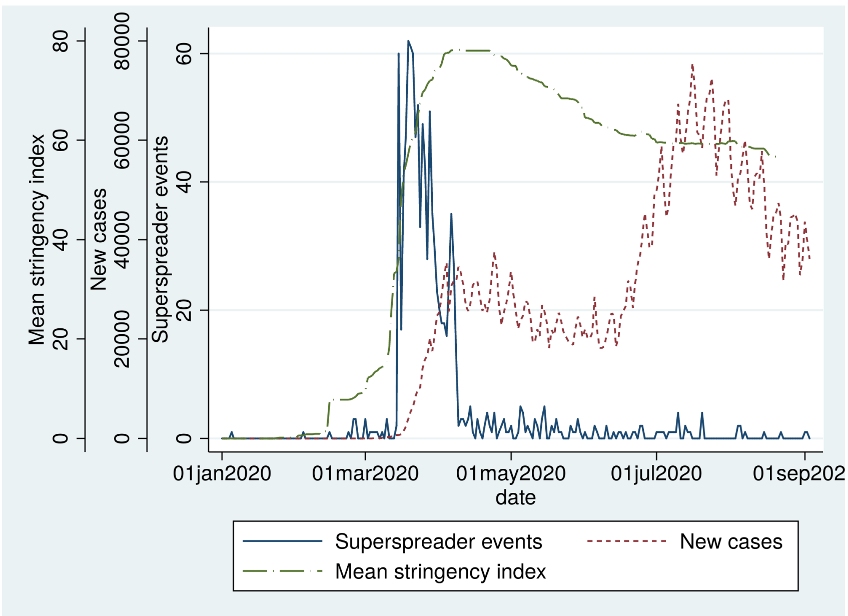


Figure 7: exemplary case for the construction of the super-spreading events instrument

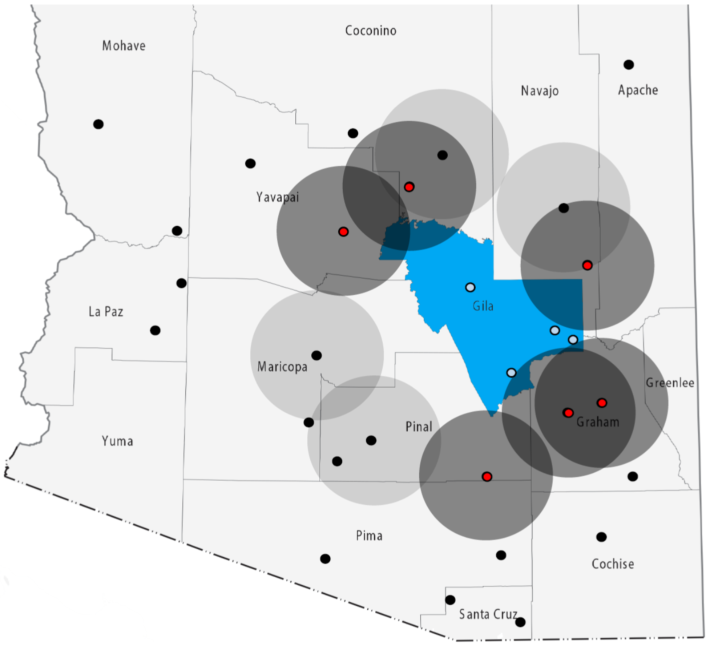


Figure 8: Spread of actual SSE by counties 6 weeks prior to George Floyd's murder

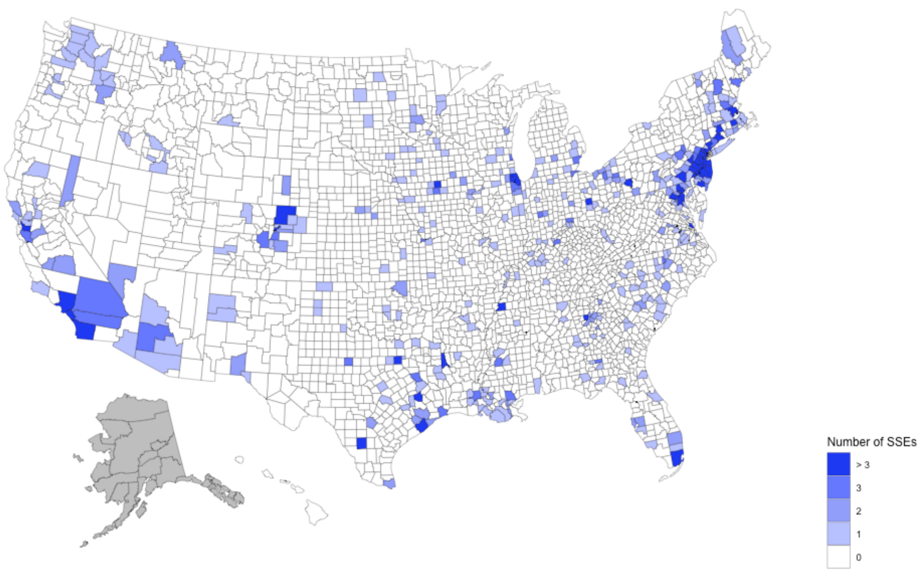
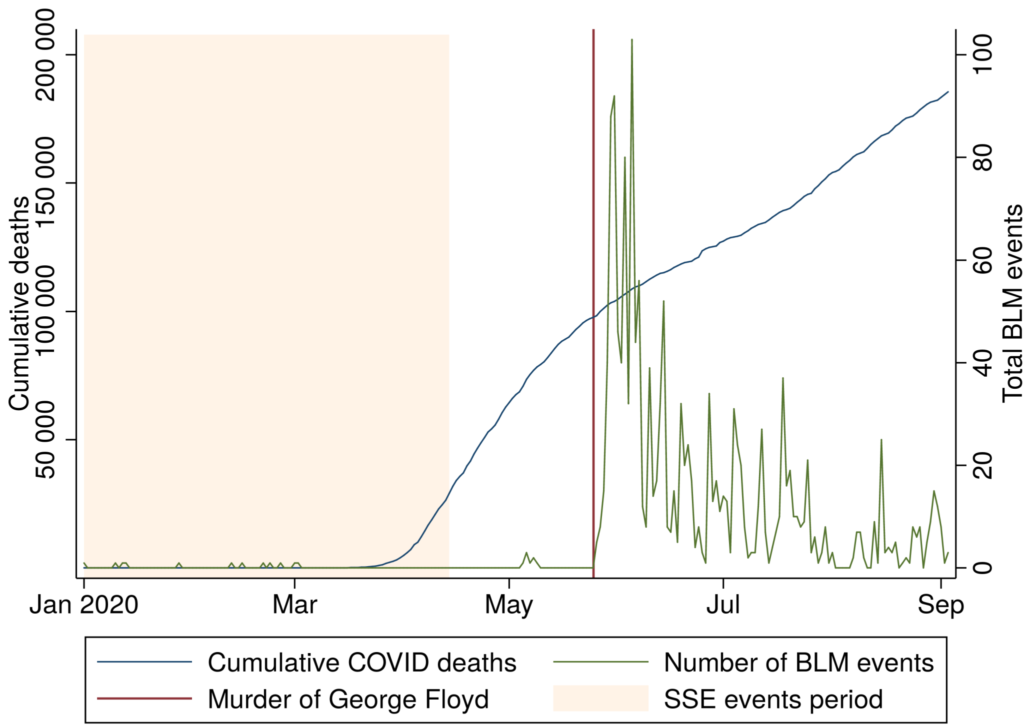


Figure 9: COVID deaths and BLM events over time



Note: Cumulative COVID-19 deaths and BLM events per day from January to September 2020. The red line denotes the week of the murder of George Floyd (May 25, 2020), and the orange shaded area is the period we consider for superspreader events.

Table 1: Summary statistics

From 25th of May to 14th of June 2020:	N	Mean	SD	Min	Max
Presence of BLM events	3108	0.099	0.299	0.000	1.000
Number of BLM events	3108	0.265	1.474	0.000	36.000
Participants in BLM events	3108	299	6082	0	323687
On the 25th of May 2021:					
COVID deaths (total)	3108	34.09	408.17	0.000	21132
COVID cases (total)	3108	588.7	4606.42	0.000	209195
COVID deaths (per 1000)	3108	0.114	0.252	0.000	2.935
COVID cases (per 1000)	3108	2.801	5.678	0.000	145.513
Superspreader events, 6+ weeks ago, neighboring counties	3108	3.119	10.035	0.000	143.000
County characteristics:					
Black police-related deaths (2014-2019)	3108	0.696	3.295	0.000	84.000
Black police-related deaths (2020)	3108	0.048	0.305	0.000	6.000
Unemployment rate (year average)	3107	4.691	1.550	0.708	19.650
Black population share	3108	0.100	0.147	0.000	0.875
Urban counties	3108	0.021	0.142	0.000	1.000
BLM events (2014-2019)	3108	0.686	5.264	0.000	174.000
Black poverty rate	3108	0.281	0.225	0.000	1.000
Population share with 3+ risk factors	3108	25.904	5.022	10.685	48.448
Vote share for republicans (2016)	3108	0.633	0.156	0.041	0.960
Vote share for republicans (2012)	3108	0.596	0.148	0.060	0.959
Median household income (2016)	3108	48810	13288	20170	129150
Social capital	3108	456	1358	0	37547
Notable Deaths	3108	0.0105	0.123	0	3

Note: Summary of main variables used in our analysis. The sample consists of 3,108 US counties. We report the number of observations, the mean, the standard deviation as well as the minimum and maximum value of each of the variables.

Table 2: Covid-19 exposure and BLM protest

	Presence of BLM events						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: All counties							
IV: COVID	0.556***	0.559***	0.573***	0.578***	0.258**	0.222*	0.215*
(deaths/1000)	(0.140)	(0.140)	(0.147)	(0.142)	(0.115)	(0.119)	(0.121)
OLS: COVID	0.0948**	0.0904**	0.0725*	0.0662	0.0366	0.0346	0.0323
(deaths/1000)	(0.0434)	(0.0436)	(0.0416)	(0.0406)	(0.0264)	(0.0268)	(0.0264)
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,106
F first stage	40.63	40.56	36.09	35.01	38.10	37.44	36.05
Mean dep. var.	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988
Panel B: *new allies* (counties with no BLM protest before)							
IV: COVID	0.549***	0.549***	0.595***	0.665***	0.425**	0.405**	0.404**
(deaths/1000)	(0.164)	(0.164)	(0.179)	(0.178)	(0.178)	(0.184)	(0.187)
OLS: COVID	0.0469*	0.0469*	0.0484*	0.0527*	0.0423*	0.0398*	0.0385*
(deaths/1000)	(0.0276)	(0.0276)	(0.0266)	(0.0274)	(0.0233)	(0.0225)	(0.0221)
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,767
F first stage	43.48	43.48	40.31	28.01	26.83	27.35	27.04
Mean dep. var.	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477
Panel C: *traditional protesters* (counties with BLM protest before)							
IV: COVID	0.416*	0.431*	0.378	0.423*	0.123	0.0495	0.0104
(deaths/1000)	(0.221)	(0.220)	(0.248)	(0.220)	(0.293)	(0.268)	(0.266)
OLS: COVID	0.314***	0.305***	0.251**	0.229**	0.0705	0.0743	0.0682
(deaths/1000)	(0.0915)	(0.0938)	(0.110)	(0.0977)	(0.106)	(0.100)	(0.102)
Observations	333	333	333	333	333	333	333
F first stage	36.32	37.34	37.37	37.23	28.87	28.55	28.09
Mean dep. var.	0.514	0.514	0.514	0.514	0.514	0.514	0.514
Past BLM events		Y	Y	Y	Y	Y	Y
Black population			Y	Y	Y	Y	Y
Black poverty			Y	Y	Y	Y	Y
Urban				Y	Y	Y	Y
3+ risk factors					Y	Y	Y
Median hh income					Y	Y	Y
Past Republican vote						Y	Y
Social capital							Y
Use of deadly force	Y	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths per 1000 population on the presence of at least one Black Lives Matter event during the three weeks following the murder of George Floyd. Panel A presents 2SLS estimation, using number of super-spreader events in neighbouring counties (50km radius) six weeks prior as an instrument and OLS results for all US counties. Panel B presents these results for the sub-sample of counties with no BLM protest before the murder of George Floyd. Panel C presents these results for the sub-sample of counties with at least one BLM protest before the murder of George Floyd. All specifications include state fixed effects and control for the unemployment rate of the county and the number of Black people that died during a police encounter. Each column include sequentially different sets of additional controls. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 3: COVID deaths interacted with county characteristics - All counties

	Presence of BLM events								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All counties									
COVID deaths/1000	0.215*	-0.908*	0.276	-0.176	0.256	-0.314	-0.0301	0.295**	0.243**
	(0.121)	(0.487)	(0.204)	(0.301)	(0.175)	(0.188)	(0.147)	(0.112)	(0.120)
… × Non-black population share		1.301**							
		(0.548)							
… × Non-white population share			-0.163						
			(0.411)						
… × Median household income				4.17e-06*					
				(2.43e-06)					
… × Vote Republican 2016					-0.102				
					(0.393)				
… × Not large cities						0.608***			
						(0.155)			
… × Suburban areas							0.321***		
							(0.112)		
… × Smaller towns								0.0391	
								(0.137)	
… × Rural areas									-0.155
									(0.159)
Interacting variable		-0.191	-0.111**	2.24e-06**	-1.056***	-0.566***	-0.0572**	0.0703***	-0.0652***
		(0.185)	(0.0537)	(9.26e-07)	(0.191)	(0.157)	(0.0251)	(0.0220)	(0.0223)
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,106	3,106	3,106
All controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths per 1000 inhabitants on first-time BLM protest, interacted with county characteristics. All specifications include state fixed effects and all standard controls. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Protest dynamics and diffusion

	Presence of BLM events	Only week 1	Only week 2	Only week 3	Presence of BLM events		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Counties without BLM protests before							
IV: COVID (deaths/1000)	0.404**	0.242	0.237	0.0517	0.559**	-0.153	3.307
	(0.187)	(0.159)	(0.246)	(0.0726)	(0.225)	(0.806)	(2.464)
× Neighbor protested first					-0.127		
					(0.165)		
× Distance to Minneapolis						0.000371	-0.00770
						(0.000532)	(0.00473)
× Distance to Minneapolis (squared)							3.73e-06*
							(2.11e-06)
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,767
F first stage	27.04	27.04	27.04	27.04	12.89	29.44	16.61
F interaction					28.33	21.22	11.51
F interaction sq							7.753
Mean of dependent variable	0.0477	0.0130	0.0249	0.0159	0.0477	0.0477	0.0477
All controls	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths on BLM events in a county in the 3 weeks following the murder of George Floyd. Column 1 presents our main regression. Columns 2, 3 and 4 restrict the outcome to the first, second, and third week after the Floyd's murder. Column 4 presents the interaction with whether a neighboring county protested before the county of interest following May 25th. Column 5 and 6 interact with the distance to Minneapolis, and the distance and its square. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Analysis of spillovers

	Presence of BLM events	Total participants	Participants per event	Number of events	Presence of BLM events			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Counties without BLM protests before								
IV: COVID (deaths/1000)	0.404**	-81.00	-723.4	0.621**	0.410**	0.306	0.412**	0.175
	(0.187)	(261.4)	(589.2)	(0.247)	(0.189)	(0.343)	(0.191)	(0.257)
× Neighbor protested historically						0.116		
						(0.345)		
× Neighbor protested currently								0.236
								(0.240)
Neighbor protested historically					-0.0114	-0.0223		
					(0.0120)	(0.0317)		
Neighbor protested currently							-0.0103	-0.0289
							(0.0143)	(0.0253)
Observations	2,767	2,767	120	2,767	2,767	2,767	2,767	2,767
F first stage	27.04	27.04	56.59	27.04	27.08	13.75	26.60	13.47
F interaction						16.57		32.07
Mean of dependent variable	0.0477	21.03	312.4	0.0636	0.0477	0.0477	0.0477	0.0477
All controls	Y	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths on BLM events in a county in the 3 weeks following the murder of George Floyd. Column 1 presents our main regression. Columns 2 present the total number of participants in all events in the county, column 3 the average number of participants per event, column 4 the number of events taking place in the county. Columns 5 to 8 use as outcome the presence of BLM events. Columns 7 and 6 control by whether a neighbor had a BLM protest before Floyd's murder. Column 6 additionally interacts with this variable. Columns 7 and 8 control by whether a neighbor had a BLM protest in the 3 weeks following Floyd's murder. Column 8 additionally interacts with this variable. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: COVID-19 exposure and social media use

	PC 1	New Twitter accounts	New Twitter accounts (log)	Google searches for Twitter	Residential stay
	(1)	(2)	(3)	(4)	(5)
Panel A: all counties					
IV: COVID (deaths/1000)	1.042**	-0.709	0.690*	12.94*	3.155***
	(0.420)	(20.17)	(0.376)	(6.453)	(0.592)
Observations	1,332	3,106	3,106	3,056	1,351
F first stage	27.65	36.05	36.05	35.71	27.49
Mean of dependent variable	0	4.586	0.586	60.64	12.08
Panel B: *new allies* (counties with no BLM protest before)					
IV: COVID (deaths/1000)	1.717***	17.88**	1.317***	18.28**	3.885***
	(0.516)	(7.871)	(0.339)	(8.838)	(0.931)
Observations	1,014	2,767	2,767	2,733	1,025
F first stage	20.21	27.04	27.04	26.05	20.20
Mean of dependent variable	-0.201	1.808	0.420	59.98	11.45
Panel C: *traditional protesters* (counties with BLM protest before)					
IV: COVID (deaths/1000)	0.202	-37.13	-0.374	5.724	2.437***
	(0.443)	(62.07)	(0.395)	(6.164)	(0.886)
Observations	312	333	333	320	320
F first stage	25.27	28.09	28.09	26.54	26.13
Mean of dependent variable	0.652	27.47	1.931	66.43	14.12
All controls	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths per 1000 population on use of social media. Column 1 shows the first principal component of the three outcomes of interest: new Twitter accounts, Google searches for Twitter, and residential stay. Table C7 details the construction of the principalc component. Column 2 shows estimates for new twitter accounts created between April 13 to May 24. Column 3 shows results for Google searches for twitter during the same period and column 4 for residential stay. Panel A presents 2SLS estimation, using number of super-spreader events in neighbouring counties (50km radius) six weeks prior as an instrument and OLS results for all US counties. Panel B presents these results for the sub-sample of counties with no BLM protest before the murder of George Floyd. Panel C presents these results for the sub-sample of counties with at least one BLM protest before the murder of George Floyd. All specifications include state fixed effects and standard controls. Each column include sequentially different sets of additional controls. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Effect of Twitter presence on protest

	(1)	(2)	(3)
	Uninstrumented users		Instrumented users
VARIABLES	Presence of BLM events		Presence of BLM events
COVID (deaths/1000)	-0.599	-0.0444	-0.578
	(0.409)	(0.277)	(0.568)
× Log(Preexisting users)	0.245***		0.232*
	(0.0880)		(0.118)
× Log(New users)		0.205**	
		(0.0834)	
Log(Preexisting users)	0.0128		0.0406
	(0.00854)		(0.0453)
Log(New users)		0.0193*	
		(0.0102)	
Mean of dep. var	0.0477	0.0477	0.0477
F COVID	11.35	15.28	8.530
F users			19.31
F interaction	47.35	60.91	18.87
Observations	2,767	2,767	2,767
Instruments	SSE		SSE and SXSW
All controls	Y	Y	Y
Pre-SXSW users			Y
State fixed effects	Y	Y	Y

Note: Column 1 and 2 show the effect of uninstrumented pre-existing or new users interacted with COVID deaths (instrumented by SSE) on the presence of BLM events in a county. Column 3 shows an IV estimate of the model of column 1, with pre-existing users instrumented by SXSW users. The first stage regression is reported on Table C5. We present results for the sub-sample of counties with no BLM protest before the murder of George Floyd. All specifications include state fixed effects and all standard controls. First stage F statistic for weak identification per second-stage coefficient (F COVID, F users, F interaction) following Sanderson and Windmeijer (2016). Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Survey data: COVID-19, news consumption and attitudes towards BLM, Blacks and COVID-19

	News consumption			**Attitudes towards Blacks, BLM & COVID-19**			**Placebo**
	Follow news about GF (1)	Receive news about GF on social media (2)	Ratio social media to overall GF news (3)	Higher Black COVID hospitaliz. not their fault (4)	Protest because structural racism (5)	Protest because criminal behaviour (6)	Illegal immigration (7)
COVID-19 deaths per capita (category)	0.0480*** (0.00964)	0.0343** (0.0152)	0.0225* (0.0134)	0.0115* (0.00645)	0.0259*** (0.00907)	-0.0254** (0.0109)	-0.00641 (0.00540)
Black	Y	Y	Y	Y	Y	Y	Y
Metropolitan area	Y	Y	Y	Y	Y	Y	Y
Female	Y	Y	Y	Y	Y	Y	Y
Age	Y	Y	Y	Y	Y	Y	Y
Education	Y	Y	Y	Y	Y	Y	Y
Income	Y	Y	Y	Y	Y	Y	Y
Democrat	Y	Y	Y	Y	Y	Y	Y
Observations	9,201	9,121	9,111	9,212	9,190	9,183	9,212

Note: Relation between living in a county with different levels of COVID-19 deaths per capita on different outcomes related to news consumption and attitudes towards Blacks, BLM and COVID-19. Columns 1 to 3 present the estimates for outcomes related to news consumption. In particular, column 1, 2 and 3 show respectively: the interest in George Floyd related news, the amount of GF related news received through social media and the ratio of the variable of column 2 over the variable of column 1. Columns 4 to 6 show the results for the outcomes related to attitudes towards BLM and racism awareness. Column 4 corresponds to the likelihood of answering that the higher COVID-19 mortality rate faced by Blacks is due to their disadvantaged circumstances instead of to their personal life style choices. Columns 5 and 6 correspond to the likelihood of answering that the protest following George Floyd's death is related with structural racism or to criminal behaviour respectively. Finally, column 7 shows a placebo result. The exact framing of the questions is as follows: column 1: *"How closely have you been following news about the demonstrations around the country to protest the death of George Floyd, a black man who died while in police custody?"*; column 2: *How much, if any, news and information about the demonstrations to protest the death of George Floyd have you been getting on social media (such as Facebook, Twitter, or Instagram)?*; column 4:*Do you think the reasons why black people in our country have been hospitalized with COVID-19 at higher rates than other racial or ethnic groups have more to do with... Circumstances beyond people's control*; column 5: *How much, if at all, do you think each of the following has contributed to the demonstrations to protest the death of George Floyd? Longstanding concerns about the treatment of black people in the country*; column 6: *Some people taking advantage of the situation to engage in criminal behavior*; column 7: *Which comes closer to your view about how to handle undocumented immigrants who are now living in the U.S.? There should be a way for them to stay in the country legally, if certain requirements are met* All columns include controls for various characteristics of the respondent: race, whether or not they live in a metropolitan area, gender, age, education, income and whether or not they lean towards the democratic party. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Alternative Mechanisms

	Presence of BLM				Other Protests	COVID-19 Protests
	(1)	(2)	(3)	(4)	(5)	(6)
***New allies*: counties with no BLM protest before**						
COVID (deaths/1000)	0.444**	0.585	0.507	0.0423	0.279	0.087
	(0.1994)	(0.342)	(0.536)	(0.980)	(0.224)	(0.104)
. . . ×Black death burden	1.391					
	(1.476)					
. . . ×Google BLM search		-0.004				
		(0.022)				
. . . ×Unemployment			-0.003			
			(0.054)			
. . . ×Stringency				0.006		
				(0.264)		
Interacting variable	-0.257	0.0006	0.0047			
	(0.176)	(0.001)	(0.008)			
Observations	2,767	2,647	2,767	2,768	2,767	2,767
F stat COVID	31.95	19.09	24.93	85.33	41.12	41.12
F stat Interaction	3.89	25.33	13.61	94.27		
Mean of dependent variable	0.0477	0.0477	0.0477	0.0477	0.0321	0.010
All controls	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths per 1000 population on presence of BLM protest. Column 1 shows estimates for instrumented COVID deaths. Columns 2 to 4 show heterogeneous effects for Black death burden weeks prior to GF's murder, Google searched for BLM 3 weeks prior to GF's murder, unemployment and stringency 3 weeks after GF's murder. Column 5 presents results for other protests. We present these results for the sub-sample of counties with no BLM protest before the murder of George Floyd. All specifications include state fixed effects and standard controls. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Online Appendix - Preliminary

Appendix A: Robustness Checks

In this section, we describe in more detail the robustness checks performed for our analysis. We focus on three dimensions: *i*) robustness to changes in the definition and construction of our instrumental variable *ii*) robustness of our main results to sample composition, spatial correlation and other confounding factors and *iii*) the possibility that our results are driven by a re-location of protesters across time and space rather than a "broadening" of the BLM coalition. We present our results in Tables A.2 to A.5.

A.1 Instrument Robustness

Group size.

Changing the radius around SSEs. In the baseline specification, we choose the 50km threshold as a distance of the SSE to the county border, as it is approximately two times the average radius of a county in the US.[31] To make sure that this choice is not driving our results, we change the radius of influence to 25 km, 100 km and 200 km (columns 2, 3 and 4 of Table A.2 respectively). Results for the sample of counties with no BLM protest before are robust to all these changes in radius (panel B). Estimates for all counties shown in panel A though remain similar in magnitude to the baseline magnitude but loose significance for change in radius to 100 km.

Cases related to SSEs. In our baseline specification, we consider the number of SSE events. In column 5 of Table A.2, we use the number of COVID-19 cases associated with these SSE instead of the number of events. Results hold only for all counties but not for the sub-sample of counties that did not protest before. We interpret this as the burden of COVID was felt mostly through deaths rather than cases. Further, as explained before, using COVID cases is prone to measurement and reporting errors.

Changing the time window of SSEs. In our baseline specification, we take into account the SSEs that occurred in a specific time window that we call "window of opportunity" where there were enough cases to observe SSEs and the social distancing measures were not applied strictly or widely enough. In columns 6 to 8 of Table A.2 we change the time window to check that the results are not driven by the specific window chosen. In particular, instead of stopping counting SSEs on April 13th (6 weeks before the murder of George Floyd); we stop counting on April 20th, 5 weeks before the murder of George Floyd (column 6), on April 6th, 7 weeks before (column 7) and on March 30th, 8 weeks before (column 8). Results are robust to change in the specific time window chosen.

Excluding SSEs in prisons. A non-negligible part of SSEs occurred inside prisons. It is likely that by the nature of prisons, the geographical spread of cases stemming from an SSE in a prison could be quite different from SSEs in other locations. In column 2 of Table A.3 we exclude SSEs that occurred in prisons. Results hold for both samples.

[31] For reference, the average radius of a county is 28 km and the average radius of a state is 220 km.

Controlling for SSEs in the county. Our first stage compares the effect of having an SSE outside the county within 50 km of the county border and excluding the effect of SSE that take place within its border. Therefore, in our analysis, a county is "not affected" by SSE if its border is either further than 50 km from the SSE, or the SSE happened within its boundaries. We expect the effect of SSEs to be different between these groups: presumably, counties far away will have no COVID-19 cases from this SSE, while the county where the SSE took place will have a lot of cases and deaths caused by the event. To provide a more accurate model for the first stage, we add as a control the number of SSEs that occurred within the county itself. Estimates are presented in column 3 of Table A.3 and show that the results of the baseline specification are robust to the addition of this control for the counties with no BLM before and become imprecisely estimated for the sample of all counties.

Weighting SSEs by distance. In our baseline specification, we count any SSE that occurred in a 50 km radius outside the border of a county as an additional SSE affecting the county. However, an SSE 1 km away from the border is likely to have a different level of influence from a SSE 49 km away. To ensure that this simplification is not driving the results, we refine the level of influence in three different ways. First we weight the SSEs by a linear function decreasing with distance (column 4 of Table A.3). Second, we repeat the analysis but with a quadratic function (column 5 of Table A.3). Finally, we weight by the percentage of the county that overlaps with the 50 km radius circle around the SSE (column 6 of Table A.3). The results are robust to all three different specifications, except the overlap specification in the full sample (panel A, column 6): in that case, the magnitude decreases considerably and the effect becomes insignificant, confirming that our main results are driven by the subsample of counties with no past BLM protest.

Weighting SSEs by the inverse probability of occurrence. The probability of being near a county that has a SSE is not constant over all counties. For instance, counties neighboring cities have likely a higher probability of being treated by our instrument as its neighbor is more likely to host an SSE. To overcome this possible violation of the exclusion restriction[32], we weight each observation (i.e. each county) by the inverse probability of being treated. Using LASSO (a regularized regression procedure that performs variable selection and avoids overfitting, Tibshirani 1996), we select relevant variables predicting (by a logit model) the probability of having a neighbor with an SSE among a set of county characteristic including a large set of socio-demographic and economic characteristics extracted from the American Community Survey (such as population, population density, race distribution, age groups, poverty rates, among others), indicators for different levels of urbanization, geographical indications (latitude, longitude, and state dummies), as well as the minimum and maximum of these variables for neighboring counties. We use the LASSO selected model to predict the probability of a county having a neighbor with an SSE, then weight the observations by the inverse of this probability. Doing this means that counties with a higher probability of having a neighbor with an SSE that actually had a neighbor with an SSE are weighted less that counties with a lower probability of being treated that are actually treated. Estimates are presented in column 7 of Table A.3 and show that our results are robust to this weighting.

[32] This could be a violation of the exclusion restriction because the probability of being treated by our instrument at a certain level is not uniform and this heterogeneity could be related to certain county characteristics that could be, at their turn, related to the intrinsic probability of protesting.

A.2 Robustness of Main Results

Placebo estimate. If our instrument were to pick up any underlying factors correlated with the overall likelihood to protest for a BLM related cause, then this would challenge a causal interpretation of our estimates. In order to probe the plausibility of the exclusion restriction, we estimate the effect of instrumented COVID-19 on the likelihood of observing past BLM protests. If indeed, our insturment was correlated with the county unobservables that also predict the likelihood of observing BLM protests then we expect to see a statistically significant relationship between our instrumented COVID-19 and likelihood of observing a BLM protest in the past. In column 2 of Table A.4, we show that exposure to COVID-19 does not predict the presence of BLM events between 2014 and 2019. We take this as an additional piece of evidence for the plausibility of our identifying assumption.

Excluding coastal counties and states. Coastal states and counties might behave differently, either with regard to our instrument or to the process of COVID-19 contagion. Coastal regions are generally denser, which increases the chance of having an SSE (Figure 5 shows the density of SSEs). On the other hand, our instrument behaves differently as half of the potential area where SSEs affecting the county could happen is actually in the ocean. Coastal regions are also more internationally connected, and were the first affected by COVID-19 in the US (the first reported case was in the state of Washington, and the first reported death in California). We show that our main result for the counties with no BLM protest before are robust to excluding coastal counties (column 3 of Table A.4), as well as coastal states (column 4). Estimates for panel A remain with similar magnitude but become imprecisely estimated.

Changing the time window of protests. In our baseline specification, we choose the three week window following Floyd's murder since it captures the vast majority of BLM related protest occurring (see Figure 3), while being close enough to the exposure to COVID-19 on May 24th, right before the protest trigger. We show that our main results (panel B) are robust to reducing this time window to 2 weeks and expanding this time window to 6 and 8 weeks (columns 5 to 7 of Table A.4 respectively). For panel A the estimate for column 5 decreases in magnitude and becomes insignificant.

COVID-19 cases instead of deaths. In our baseline specification we use the number of COVID-19 deaths per thousand in the county as an explanatory variable for protest. In column 8 of Table A.4, we show that the results hold when using the number of COVID-19 related cases instead of the number of deaths.

Changing the estimation method from a 2SLS to a Probit. In our baseline specification the effect of COVID-19 is additive. It might be the case that the effect would be multiplicative of some characteristics of the counties. Using a Probit model accounts for this possibility. Non-linear models with many covariates (typically when using fixed effects) suffer from the incidental parameter problem resulting in bias of the estimates (Heckman, 1987; Lancaster, 2000; Wooldridge, 2015). To reduce the extent of this problem, we omit the state fixed effects which significantly reduce the number of covariates. Results, keeping an OLS first stage, but using a Probit second stage, are shown in column 2 of Table A.5. Results are positive and significant for the subsample of interest (panel B) and positive but imprecisely estimated for the subsample of all counties.

Controlling for propensity to protest. Our main specification already controls for the number of BLM events that took place in the county in the previous year. While this gives some indication of the county's propensity to protest, this is essentially an imprecise measure of this fact, since counties having a non-zero probability to protest might simply not have protested before by random chance. We re-use the propensity to protest that we constructed for our matching-based alternative identification (the construction of this propensity measure is detailed in Appendix B.3) as a control in our regression. We first use it directly as a control (column 3 of Table A.5). Results hold. We also include fixed effects for different levels of the propensity to protest. In particular, we group observations by groups of 1000, 100 and 10 units with similar propensity to protest and add fixed effects for each group. Results are shown in columns 4 to 6 of Table A.5. This is essentially a matching-like strategy, where the fixed effects ensure that observations with similar propensity are compared. Results are robust to the inclusion of fixed effects for the panel of interest (panel B) and become insignificant for some specifications of the whole sample.

Accounting for spatial correlation. Observations are likely to be spatially correlated for several reasons. For instance, there could be spatially-correlated unobserved factors influencing the decision to protest (such as weather conditions or available TV and radio stations). Spatially correlated observations lead to incorrect standard errors. Clustering by state does not entirely remove these errors because correlation across state borders remain (Colella et al., 2019). To overcome this problem, we use Conley standard errors that allow for spatial correlation within a certain distance. Column 7 of Table A.5 shows the estimates when allowing spatial correlation between observations in a 50 km radius. Column 8 of Table A.5 shows the estimates when allowing spatial correlation with all neighboring counties. Results remain robust.

Estimation without clustering The inclusion of clustering when adding fixed effects at the same level is discussed in the literature (Abadie et al., 2017). Our preferred specification clusters at the state level and includes state fixed effects. Column 9 of Table A.5 shows that our results also hold when we do not cluster the standard errors.

A.3 Broadening versus Scattering of Protest

In this section we discuss the possibility that spatial spillovers from BLM protest (say, from the cities to the suburbs) are driving our results. Specifically, we investigate whether the observed broadening of the coalition is in fact just a substitution of protesters in time and space. In fact, it is possible that we observe new counties protesting for various reasons unrelated to the idea of "new allies". First, the pandemic may have changed the scope and structure of BLM protests (smaller but more numerous). Second, neighboring counties may inspire subsequent protest in close proximity.[33] Third, the pandemic and its restrictions on mobility may have led to a geographic spread of the protest, substituting large protests in cities with smaller protests in suburbs. In the following we address the concern that the pandemic may have simply led to a substitution of protest locations and frequencies, rather than a true broadening of the coalition.

Number of participants and protests. If the observed increase in the number of "new allies" is simply driven by a substitution of protest across space (e.g. re-location of protesters themselves or creation of multiple smaller protest events), we should observe that the number of protests increases while the number of participants should decrease. We show in columns 2 to 4

[33]If SSEs and BLM protests themselves have spill-over effects, we may falsely attribute an increase in protest to the pandemic.

of Table 5 that neither is the case. We take this as first indicative evidence that the pandemic does not change the structure of these protests.

Moreover, we consider the possibility that individuals that protest might, in response to the pandemic, decide to protest closer to home and not protest in the city center of the neighboring county. For instance, protesters could be affected by restrictions and closures of public transport, preventing them from going to a demonstration further away. They might also consider that a smaller, more local demonstration is safer as they would come into contact with less people, limiting the risk of spreading coronavirus between communities.

Traditional protesters as neighbors. While we should pick up some of this in the number of participants and protests in the previous analysis, we test this more systematically by constructing a dummy variable equal to one if one of the county's neighbors is a "traditional protester" (e.g. had a BLM related protest before May 25th 2020), including it both separately and as an interaction term. In column 5 and 6 of Table 5, we show that having a traditional protester as a neighbor does not increase the probability of protesting overall within the sample of counties that had never protested before. More importantly, the interaction term between exposure to COVID-19 and having a traditional protester as a neighbor in column 6 is not significant and if anything reduces the likelihood of protesting in response to the pandemic. This seems to indicate that the displacement effect is not the driver behind our results.

Recent protesters as neighbors. Lastly, it is possible that protests in one county could inspire protests in neighboring counties over time. While this would not go against the idea of a broadening BLM coalition, it indicates that protests during the pandemic inspire subsequent protests in neighboring counties. We therefore construct an indicator similar to the "traditional protester as neighbor" but apply this to the period after Floyd's murder. More specifically, we construct a dummy variable that indicates whether the county has a neighbor that protested *before* they start to protest. This allows us - even in our cross-sectional setup - to account for spillovers in time. However, this approach suffers from an important caveat: protests in neighboring counties during the pandemic could be endogenous and therefore a bad control. Nevertheless, we look at these effects in columns 7 and 8 of Table 5 with these caveats in mind.

If spillovers exist, we would expect that having a neighbor that recently protested increases the likelihood of observing a protest yourself. We include this variable as a control in columns 7 and 8 of Table 5 and find no change in our results when we only include current neighbours that start protesting before as a control and no effect when this is included as an interaction term. tHis provides suggestive evidence that these temporal spillovers across neighboring counties are not driving our main results.

Overall, these exercises alleviate concerns that the observed broadening of the BLM coalition is driven by the substitution of existing protesters across space and time.

Table A.1: Instrument validity

	Presence of BLM events	Past BLM events	Presence of BLM events				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: all counties							
IV: COVID (deaths/1000)	0.215*	-0.0523	0.248*	0.249**	0.242**	0.116	0.0994
	(0.121)	(0.215)	(0.139)	(0.114)	(0.116)	(0.129)	(0.125)
First stage coefficient:	0.00930***	0.00928***	0.00965***	0.00936***	0.00940***	0.00945***	0.00931***
	(0.00155)	(0.00155)	(0.00158)	(0.00153)	(0.00154)	(0.00153)	(0.00133)
Reduced form: SSEs	0.00200	-0.000485	0.00239	0.00233*	0.00227*	0.00109	0.000919
	(0.00128)	(0.00194)	(0.00152)	(0.00122)	(0.00127)	(0.00132)	(0.00124)
Observations	3,106	3,106	3,105	3,002	3,106	3,106	3,106
F first stage	36.05	35.73	37.35	37.47	37.15	38.02	48.84
Mean of dep. var.	0.0988	0.108	0.0952	0.102	0.0988	0.0988	0.0988
Panel B: counties with no BLM protest before							
IV: COVID (deaths/1000)	0.404**		0.363*	0.423**	0.405**	0.348*	0.341*
	(0.187)		(0.195)	(0.185)	(0.184)	(0.186)	(0.183)
First stage coefficient:	0.00751***		0.00781***	0.00761***	0.00758***	0.00768***	0.00738***
	(0.00144)		(0.00164)	(0.00142)	(0.00143)	(0.00149)	(0.00142)
Reduced form: SSEs	0.00303*		0.00284*	0.00322*	0.00307*	0.00268	0.00248
	(0.00163)		(0.00163)	(0.00164)	(0.00162)	(0.00164)	(0.00160)
Observations	2,767		2,766	2,663	2,767	2,767	2,767
F statistic	27.04		22.55	28.56	28.20	26.72	27.06
Mean of dep. var.	0.0990		0.0883	0.102	0.0990	0.0990	0.0990
SSE probability weighting			Y				
Propensity to protest				Y			
Propensity to protest group: size					1000	100	10
All controls	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y

Note: Variations of the baseline specification of the effect of the number of SSE in neighboring counties on the presence of at least one Black Lives Matter event during the weeks following the murder of George Floyd. Column 1 corresponds to our baseline specification. Columns 2 presents the effect on past BLM events as a placebo. In columns 3 observations are weighted by the inverse probability of observing a SSE affecting the county if a SSE is observed, no SSE if no SSE is observed. Column 4 adds a control for the propensity to protest. Columns 5 to 7 add fixed effects for groups of propensity to control of size 1000, 100 and 10 respectively. All specifications include the whole set of controls and state fixed effects. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Robustness checks - I

	Presence of BLM events during 3 weeks after May 25th							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: all counties				
IV: COVID (deaths/1000)	0.215*	0.278**	0.275	0.373*	0.213*	0.209*	0.225*	0.240*
	(0.121)	(0.130)	(0.175)	(0.207)	(0.121)	(0.121)	(0.119)	(0.120)
First stage coefficient:	0.00930***	0.0141***	0.00388***	0.00132***	6.34e-05***	0.00919***	0.00962***	0.0112***
	(0.00155)	(0.00215)	(0.000703)	(0.000379)	(1.24e-05)	(0.00154)	(0.00164)	(0.00208)
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,106	3,106
F first stage	36.05	42.84	30.38	12.08	26.23	35.66	34.63	28.91
Mean of dep. var.	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988
				Panel B: counties with no BLM protest before				
IV: COVID (deaths/1000)	0.404**	0.503*	0.499**	0.503**	0.219	0.383**	0.440**	0.410**
	(0.187)	(0.266)	(0.191)	(0.225)	(0.215)	(0.188)	(0.193)	(0.187)
First stage coefficient:	0.00751***	0.0126***	0.00304***	0.000901***	4.86e-05***	0.00738***	0.00770***	0.00926***
	(0.00144)	(0.00331)	(0.000309)	(0.000272)	(1.00e-05)	(0.00139)	(0.00154)	(0.00170)
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,767	2,767
F first stage	27.04	14.40	97.13	10.95	23.40	28.12	24.87	29.78
Mean of dep. var.	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477
Distance	50 km	25 km	100 km	200 km	50 km	50 km	50 km	50 km
Lag	6 weeks	6 weeks	6 weeks	6 weeks	6 weeks	5 weeks	7 weeks	8 weeks
SSE measure	count	count	count	count	cases	count	count	count
All controls	Y	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

Note: Variations of the baseline specification of the effect of the number of SSE in neighboring counties on the presence of at least one Black Lives Matter event during the weeks following the murder of George Floyd. Column 1 correspond to our baseline specification. Columns 2 to 4 vary the distance at which SSE are counted from 25 to 200km. Column 5 uses the number of cases attributed to an SSE instead of the number of SSEs. Columns 6 to 8 vary the time at which SSEs are counted (usually 6 weeks), showing values for 5, 7 and 8 weeks. All specifications include the whole set of controls and state fixed effects. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Robustness checks - II

	Presence of BLM events during 3 weeks after May 25th						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: all counties							
IV: COVID (deaths/1000)	0.215*	0.255**	0.235	0.275**	0.307***	0.0228	0.248*
	(0.121)	(0.109)	(0.149)	(0.108)	(0.109)	(0.136)	(0.139)
First stage coefficient:	0.00930***	0.0100***	0.00842***	0.0207***	0.0274***	0.0154***	0.00965***
	(0.00155)	(0.00177)	(0.00176)	(0.00341)	(0.00452)	(0.00209)	(0.00158)
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,105
F first stage	36.05	31.92	22.92	36.72	36.83	54.03	37.35
Mean of dep. var.	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988	0.0952
Panel B: counties with no BLM protest before							
IV: COVID (deaths/1000)	0.404**	0.482**	0.401*	0.515**	0.581**	0.343*	0.363*
	(0.187)	(0.180)	(0.232)	(0.217)	(0.266)	(0.203)	(0.195)
First stage coefficient:	0.00751***	0.00798***	0.00653***	0.0178***	0.0239***	0.0146***	0.00781***
	(0.00144)	(0.00159)	(0.00130)	(0.00448)	(0.00637)	(0.00348)	(0.00164)
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,766
F first stage	27.04	25.03	25.08	15.73	14.09	17.69	22.55
Mean of dep. var.	0.0477	0.0477	0.0477	0.0477	0.0477	0.0477	0.0494
Excluding SSEs in prisons		Y					
Control SSE in county			Y				
Measure				linear	square	overlap	
SSE probability weighting							Y
All controls	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y

Note: Variations of the baseline specification of the effect of the number of SSE in neighboring counties on the presence of at least one Black Lives Matter event during the weeks following the murder of George Floyd. Column 1 corresponds to our baseline specification. Column 2 excludes SSEs that took place in prisons. In column 3, a control is added for superspreader events in the county 6 weeks before the murder of George Floyd. Columns 4 to 6 vary the effect of SSEs depending on the distance, either decreasing linearly (column 5), quadratically (column 6) or based on the overlap between the 50 km circle around the SSE and the county. In columns 7, observations are weighted by the inverse probability of observing a SSE affecting the county if a SSE is observed, no SSE if no SSE is observed. All specifications include the whole set of controls and state fixed effects. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Robustness checks - III

	Presence of BLM events during X weeks after May 25th							
	3 weeks	Past events	3 weeks	3 weeks	2 weeks	6 weeks	8 weeks	3 weeks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: all counties								
IV: COVID (deaths/1000)	0.215*	-0.0523	0.253	0.324	0.0694	0.289**	0.244**	
	(0.121)	(0.215)	(0.185)	(0.279)	(0.122)	(0.108)	(0.119)	
IV: COVID (cases/1000)								0.0135*
								(0.00694)
First stage coefficient:	0.00930***	0.00928***	0.00819***	0.00811***	0.00930***	0.00930***	0.00930***	0.149***
	(0.00155)	(0.00155)	(0.00110)	(0.00109)	(0.00155)	(0.00155)	(0.00155)	(0.0539)
Observations	3,106	3,106	2,882	1,839	3,106	3,106	3,106	3,106
F first stage	36.05	35.73	55.75	55.13	36.05	36.05	36.05	7.633
Mean of dep. var.	0.0988	0.108	0.0833	0.0712	0.0821	0.123	0.134	0.0988
Panel B: counties with no BLM protest before								
IV: COVID (deaths/1000)	0.404**		0.531**	0.344*	0.440**	0.575***	0.470***	
	(0.187)		(0.219)	(0.173)	(0.194)	(0.171)	(0.166)	
IV: COVID (cases/1000)								0.0312***
								(0.0116)
First stage coefficient:	0.00751***		0.00808***	0.00877***	0.00751***	0.00751***	0.00751***	0.0974**
	(0.00144)		(0.00119)	(0.000699)	(0.00144)	(0.00144)	(0.00144)	(0.0363)
Observations	2,767		2,616	1,697	2,767	2,767	2,767	2,767
F first stage	27.04		45.85	157.2	27.04	27.04	27.04	7.181
Mean of dep. var.	0.0477		0.0428	0.0371	0.0354	0.0665	0.0763	0.0477
Excluding coastal			counties	states				
All controls (except past BLM)	Y	Y	Y	Y	Y	Y	Y	Y
Past BLM events	Y		Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

Note: Variations of the baseline specification of the effect of the number of SSE in neighboring counties on the presence of at least one Black Lives Matter event during the weeks following the murder of George Floyd. Column 1 correspond to our baseline specification. Column 2 predicts past BLM events as a placebo. Columns 3 and 4 exclude coastal counties and states. In columns 5, 6 and 7, the presence of BLM events is measured in the 2, 6 and 8 weeks following May 25. Column 8 looks at the effect of COVID cases instead of deaths. All specifications include the whole set of controls and state fixed effects, escept column 2 where past BLM events are removed as a control. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Robustness checks - IV

	Presence of BLM events								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: all counties									
IV: COVID (deaths/1000)	0.215*		0.249**	0.242**	0.116	0.1000	0.215**	0.215**	0.215**
	(0.121)		(0.114)	(0.116)	(0.129)	(0.125)	(0.101)	(0.103)	(0.0883)
IV Probit: COVID (deaths/1000)		0.344							
		(0.215)							
First stage coefficient:	0.00930***	0.0105***	0.00936***	0.00940***	0.00945***	0.00929***	0.00930***	0.00930***	0.00930***
	(0.00155)	(0.000634)	(0.00153)	(0.00154)	(0.00153)	(0.00133)	(0.00141)	(0.00144)	(0.000647)
Observations	3,106	3,106	3,002	3,106	3,106	3,106	3,106	3,106	3,106
F first stage	36.05	276.3	37.47	37.15	37.98	48.63			206.6
Mean of dep. var.	0.0988	0.113	0.102	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988
Panel B: counties with no BLM protest before									
IV: COVID (deaths/1000)	0.404**		0.423**	0.405**	0.341*	0.338*	0.404*	0.404**	0.404***
	(0.187)		(0.185)	(0.184)	(0.186)	(0.182)	(0.234)	(0.205)	(0.128)
IV Probit: COVID (deaths/1000)		0.878***							
		(0.230)							
First stage coefficient:	0.00751***	0.00861***	0.00761***	0.00758***	0.00770***	0.00739***	0.00751***	0.00751***	0.00751***
	(0.00144)	(0.000762)	(0.00142)	(0.00143)	(0.00149)	(0.00142)	(0.00110)	(0.00116)	(0.000884)
Observations	2,767	2,767	2,663	2,767	2,767	2,767	2,767	2,767	2,767
F statistic	27.04	127.7	28.56	28.20	26.64	27.08			72.09
Mean of dep. var.	0.0990	0.0992	0.102	0.0990	0.0990	0.0990	0.0990	0.0990	0.0990
All controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Propensity to protest			Y						
Propensity to protest group: size				1000	100	10			
State clustering	Y	Y	Y	Y	Y	Y			
Spatial clustering							50 km	neighbors	
State fixed effects	Y		Y	Y	Y	Y	Y	Y	Y

Note: Variations of the baseline specification of the effect of the number of SSE in neighboring counties on the presence of at least one Black Lives Matter event during the weeks following the murder of George Floyd. Column 1 correspond to our baseline specification. Column 2 estimates an IV Probit model (with an OLS first stage) and omits state fixed-effects. Column 3 adds a control for the propensity to protest. Columns 4 to 6 add fixed effects for groups of propensity to control of size 1000, 100 and 10 respectively. Column 7 and 8 replace the state clustering by spatial clustering, allowing correlation in a 50 km radius for column 7, and between neighbors for column 8. Columns 9 omits clustering altogether. All specifications include the whole set of controls, except column 2 where state fixed effects are removed. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level except for columns 7 to 9. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Reduced form: superspreader events on the presence of BLM events.

	Presence of BLM events						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cumulative SSE 6 weeks ago, not in county, less than 50km away	0.00577***	0.00581***	0.00560***	0.00566***	0.00242**	0.00209	0.00200
	(0.00132)	(0.00133)	(0.00129)	(0.00136)	(0.00117)	(0.00128)	(0.00128)
Past BLM events		Y	Y	Y	Y	Y	Y
Black population			Y	Y	Y	Y	Y
Black poverty			Y	Y	Y	Y	Y
Urban				Y	Y	Y	Y
3+ risk factors					Y	Y	Y
Median hh income					Y	Y	Y
Past Republican vote						Y	Y
Social capital							Y
Use of deadly force	Y	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y
Mean of dependent variable	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988	0.0988
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,106
R^2	0.167	0.175	0.178	0.203	0.228	0.254	0.254

Note: Estimation of the effect of the number of SSE in neighbouring counties (50km radius) six weeks prior to the death of George Floyd on the presence of at least one Black Lives Matter event during the three weeks following the murder of George Floyd. All specifications include state fixed effects and control for the unemployment rate of the county and the number of Black people that died during a police encounter. Each column include sequentially different sets of additional controls. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix B: Alternative Estimation Strategies

B.1 Alternative Instrument: Florida Spring Break

In our preferred empirical strategy, we chose smaller and decentralized SSEs to argue for a causal relationship between COVID-19 and BLM. Here, we add another cross-sectional instrumental variable: the spatial distribution of touristic flows originating in major Florida Spring Break destination during March of 2020. Instead of collecting information on multiple independent SSEs as in the previous section, we now focus on one single, large-scale event that is known to have contributed substantially to the spread of COVID-19 (Mangrum and Niekamp, 2020).

Despite the fact that COVID-19 infections had surged in Florida's main spring break destinations and despite the fact that the Center for Disease Control had issued multiple warnings, Florida Governor DeSantis failed to implement social distancing orders until April 1st 2020[34]. We exploit this unique, large scale event to track the diffusion of COVID-19 infections that originated in Florida during spring break and then spread across the United States. In order to track these movements, we benefit from exceptionally rich data on cell phone mobility provided by SafeGraph. We can identify spring breakers' home counties – locations where they most likely have returned after vacationing in highly infectious spring break locations.

Specifically, we pick three Florida vacation destinations: Miami Beach, Panama Beach and Fort Lauderdale. These three destinations caught the attention of the media in early March which reported congestion of tourists who did not respect social distancing measures (BBC, CNN). We are using anonymised mobile data for the period from March, 1, 2020 to April 1, 2020, covering the majority of spring break periods across the country. With the help of the Monthly Patterns data (MP), we measure unique devices that visited specific «points of interest» in one of three popular spring break destinations mentioned above.

The SafeGraph data provides us with a rich set of points of interests, which include more than 3000 places such as restaurants, bars, hotels, gyms, public parks, malls and other establishments. Using this data, we measure the Number of devices that «pinged» in each of point of interests during March, 2020. The MP data also allows us to observe home locations on the level of the US Census Block Groups (CBG). An individual "home" is defined as a place where user's devices pinged most often in the night time between 6 PM and 7 AM during the baseline 6-week period determined by the SafeGraph.

Using this information, we calculate the number of unique visitors to points of interests in three cities in Florida and group this number by device home counties. Given that cell phone data is anonymized, each device is counted as many times as it has visited different places (such as restaurants and shops) in a given touristic destination. Therefore, this measure captures both intensity of tourism flow from the county and mobility of these tourists during their spring break. Since higher mobility is associated with higher chances of disease contraction, our variable captures both extensive and intensive margins of COVID-19 spread. We see this variable as an improvement over ones used in literature examining stay at home behaviour (Abouk and Heydari (2020); Lasry et al. (2020); Friedson et al. (2020); Dave et al. (2020); Dave et al. (2021)). The exposure to COVID-19 is therefore instrumented by the number of spring-break tourists.

$$Z_c = \frac{\sum_{POIs} pings_{POI,c}}{devices_c} \tag{6}$$

We normalise this variable calculating a ratio of the total number of devices detected in spring breakers' home counties at March , 1, 2020 to account for differences in population size and differences in resident device coverage between counties in the SafeGraph data. In

[34] Local officials had started to close some of the beaches for public access in mid March

Figure B1 the map of (log) number of devices by counties is presented. Figure B2 shows our resulting measure of "spring breakers" inflow split into five categories: high flow, moderate-high flow, moderate-low flow, low flow, no flow (missing).

We use the same set of controls and connotations as in our baseline cross-sectional estimation. Our estimating equation writes as:

$$BLM_c = \beta_0 + \beta_1 \widehat{Covid}_{cs} + \mathbf{X}_c \beta_{\mathbf{X}} + \delta_s + \epsilon_{cs}$$

We present our 2SLS results in Table B1. We use the same set of controls as in the previous cross-sectional estimations, successively introducing socio-economic, demographic and political control variables. The inclusion of the Black population rates and Black poverty index in column 3 substantially decreases the F-Statistic (see First Stage results in Table B1). When including the full set of controls, the instrument remains at 7.3, well below the conventional threshold. However, for all specifications we find a positive coefficient for COVID-19 on the presence of a BLM event and where the first stage is sufficiently strong, we find a positive and statistically significant sign.

B.2 Difference in Differences: Notable Deaths Sample

With this empirical approach, we use data on BLM at the county-week level starting in 2014 and exploit differences in protest behavior following what we call a "notable" death. Deaths of Blacks at the hands of the police have been - not only in the case of George Floyd - a trigger for BLM protests across the country. Roughly, more than 300 Blacks die each year in the US either due to police brutality or under police custody. However, not all of these deaths result in media coverage, which is crucial for generating public discourse or action. Many of these events only received public traction since they were - mostly by chance - recorded through a phone camera. We construct a data set of all police related Black deaths since July 2014 covered in a major national daily newspapers like the Washington Post, received TV coverage by CNN and/or has a dedicated Wikipedia page.

We now exploit the full potential of our panel data by interacting out main COVID-19 variable with a dummy variable for a notable death occurring in a certain week. Following the sample selection of our baseline estimation, we use information on BLM orotest ub counties in the 3 weeks after the recorded notable death (we can reduce this to 2 weeks and expand it to 4 weeks without significantly changing the first and second stage results). This data set structure allows us to observe counties' protest behavior after a protest trigger. Following a difference in differences logic, we then look at whether the reaction following this trigger differs in counties that were more exposed to the COVID-19 pandemic. Again, we use the SSE IV to account for the fact that COVID-19 exposure may be endogenous to past and present protest behavior.

$$Covid_{ct} = \zeta_0 + \zeta_1 Notable\_deaths + \zeta_2 Z_{cst} + \zeta_3 Notable\_deaths \times Z_{cst} + \mathbf{X}_{cs} \zeta_{\mathbf{X}} + \gamma_c + \theta_{st} + \eta_{cst}, \quad (7)$$

$$Z_{cst} = \sum SSE_{cst}^{neighbor} \quad (8)$$

The second stage is written as:

$$BLM_{cst} = \beta_0 + \beta_1 Notable\_deaths_t + \beta_2 \widehat{Covid}_{cst} + \beta_3 Notable\_deaths_t \times \widehat{Covid}_{cst} + \mathbf{X}_{cs} \zeta_{\mathbf{X}} + \mu_c + \delta_{st} + \epsilon_{cst}$$

where, $Notable\_deaths_{cst}$ is a dummy variable that takes the value of one in the three weeks following a nationally covered deaths and zero otherwise. We include county and state-week fixed effects, as well as all Black police-related deaths at the county level. This is a crucial control as it allows us to exploit the "extra" trigger that nationally covered deaths create, above and beyond the local level of deadly force used by local police. The key coefficient of interest is β_3.

Table B3 shows the results of this estimation. Columns 1 and 3 report the effect of notable deaths up to 4 weeks since it occurred and columns 2 and 4 report for up to 3 weeks. In both cases we find that the effect of notable deaths in predicting the likelihood of observing a BLM protest is significantly higher in the presence of COVID death burden. The results control for county specific time trends as shown in columns 3 and 4.

B.3 LASSO Matching: Propensity to Protest

We again exploit data on past protests, this time to predict the propensity of a county to protest in response to a notable death using a wide variety of observable county characteristics.

More precisely, we start by estimating the following logit model:

$$\log \frac{\Pr(BLM_{ci} = 1)}{1 - \Pr(BLM_{ci} = 1)} = \beta_0 + \beta_1 X_c + \varepsilon_{ci}$$

where c runs over counties and i over notable deaths before 2020. BLM_{ci} is a dummy variable equal to one if there was a BLM protest in county c in the three weeks following notable death i, and X_c is a vector of controls. The controls include the usual controls included in our main specification, as well as an array of county characteristics taken from the American Community Survey (such as population and umemployment by race and age groups, education levels, poverty rates), as well as past voting record in 2012, 2016 and 2020, geographic information (such as the county's geographic coordinates), urbanization indicators, and state fixed effects. We select the most relevant subset of variables with LASSO regression (Tibshirani, 1996). This avoids overfitting and gives confidence in using the model to predict the propensity to react to another notable death. This model is estimated on the subset composed on all counties, and we compute the estimated propensity to protest for each county.

We then perform a propensity score matching-like estimation: we consider the binary treatment where counties are considered treated if they had at least one COVID-19 related death on or before May 24th. We match counties with similar historical propensities to protest, and consider as outcome where these counties held a BLM protest in the 3 weeks following the murder of George Floyd. The results are presented in Table B2 for the whole sample, and the subsamples of counties that did and did not protest before. For each of these samples, the propensity-to-protest model is estimated on the whole sample. The results in each case are positive and significative; their magnitude is not comparable with our main specification as the treatment is different. Unlike our main specification, with this estimation strategy, the effect on counties that had BLM events is significant and much higher in magnitude than the effect on counties that did not have BLM events before. This might be consistent with a multiplicative effect of protest: the relative increase (relative to the probability of having a BLM event after the death of George Floyd) is roughly similar.

Note that this is not a proper propensity score matching (Rosenbaum and Rubin, 1983): we are matching not on the propensity to have a COVID death but on the (past) probability to hold a protest. With an usual propensity score matching, we would need to be concerned about unobservable characteristics of the county that affect both the treatment probability and the outcome. In this case, we can also get bias from observable characteristics of the counties that may influence the probability of treatment and protests, but did not influence the past propensity to protest as much. One such example would be the quality of the health system: it raises both the probability of deaths from COVID, and people are likely more concerned about the quality of the health care system than they were for past protests. In the robustness checks section, we use this propensity as a control in our main specification instead.

Figure B1: Number of devices (log) by US counties pinged during March 1st, 2020

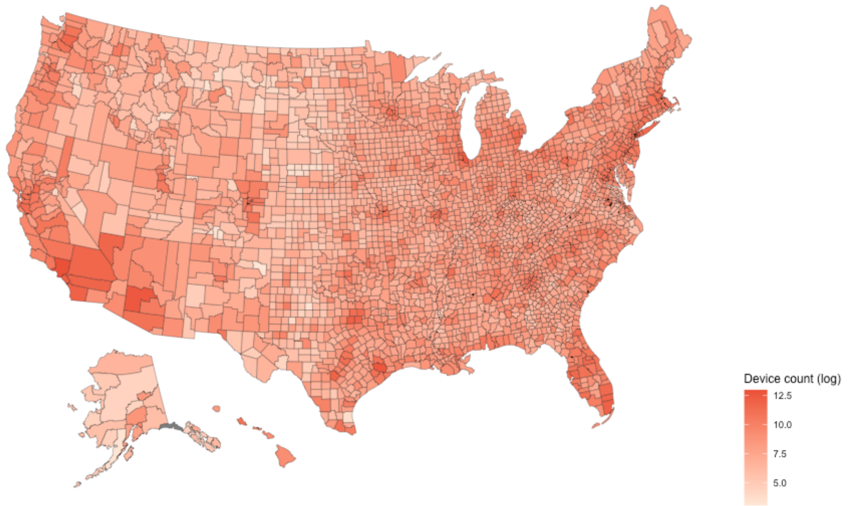


Figure B2: Spring Breakers by US counties. Own visualization based on SafeGraph data.

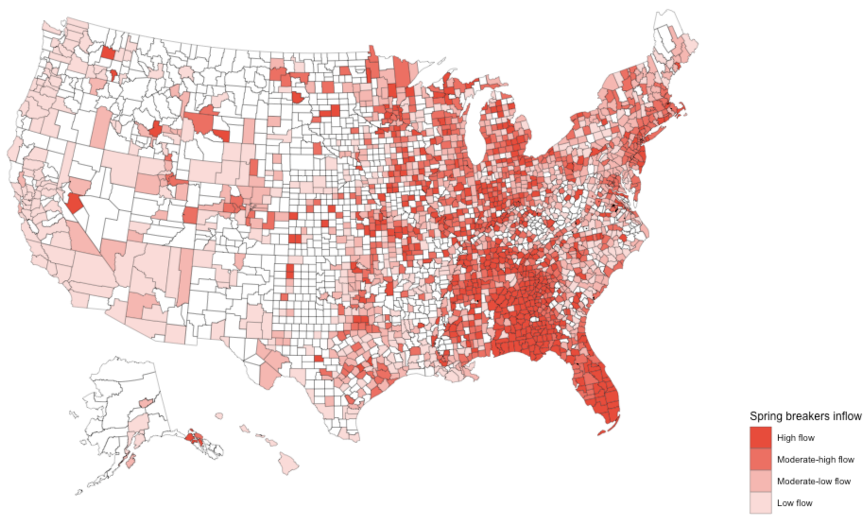


Table B1: Spring breakers IV: Covid-19 deaths on the presence of BLM events, 2SLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Presence of BLM events			
Panel A: IV							
Covid deaths per thousands	1.513***	1.472***	1.714**	1.717**	1.392**	0.828	0.832
	(0.550)	(0.550)	(0.733)	(0.731)	(0.645)	(0.514)	(0.533)
Panel B: OLS							
Covid deaths per thousands	0.0972***	0.0931***	0.0736***	0.0669***	0.0375*	0.0356	0.0333
	(0.0222)	(0.0221)	(0.0227)	(0.0224)	(0.0223)	(0.0219)	(0.0219)
Panel C: First stage							
Visits per device	0.558***	0.548***	0.448***	0.449***	0.446***	0.445***	0.430***
	(0.164)	(0.164)	(0.160)	(0.160)	(0.158)	(0.159)	(0.159)
Past BLM events		Y	Y	Y	Y	Y	Y
Black population			Y	Y	Y	Y	Y
Black poverty			Y	Y	Y	Y	Y
Urban				Y	Y	Y	Y
3+ risk factors					Y	Y	Y
Median hh income					Y	Y	Y
Past Republican vote						Y	Y
Social capital							Y
Deadly forces	Y	Y	Y	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y
Observations	3,039	3,039	3,039	3,039	3,039	3,039	3,038
F first stage	11.53	11.14	7.898	7.916	7.985	7.791	7.305

Cross-sectional 2SLS estimation of the effect of the cumulative number of COVID-19 related deaths per thousand population the day before the death of George Floyd on the likelihood of having at least one BLM event during the first three weeks after George Floyd's death. All specifications include state fixed effects, the cumulative number of black police-related deaths since 2014 and the mean unemployment rate for to period May 2019- May 2020. Columns (1) is the baseline specification. Column (2), (3), (4), (5), (6), (7), (8) include one by one additional set of controls and column (9) include all controls together. Cross-sectional data at the county level. We report Kleibergen-Paap rkWald F statistic. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B2: Matching on past propensity to protest

	(1)	(2)	(3)
VARIABLES	Presence of BLM events		
	All sample	Never protested before	Protested before
Average Treatment Effect	0.117***	0.0439***	0.324***
	(0.0110)	(0.00866)	(0.0537)
Observations	3,108	2,768	340
Mean of dep. var.	0.0994	0.0477	0.521
Propensity to protest	Y	Y	Y

Note: Estimation of the effect of having at least one COVID-19 death on presence of BLM protests. The average treatment effect is evaluated by matching on the past propensity to protest after a notable death. Column 1 presents the results for the whole sample, column 2 for counties that never protested before and column 3 for counties that did protest before. Propensity-to-protest model estimated on the full sample using logit LASSO regression using all available controls. Standard errors (in parentheses) are not clustered. *** p<0.01, ** p<0.05, * p<0.1

Table B3: Notable Deaths Regression

	(1)	(2)	(3)	(4)
	Presence of BLM			
Covid deaths per thousand	0.0595***	0.0597***	0.0450***	0.0451***
	(0.0166)	(0.0166)	(0.0116)	(0.0116)
Notable deaths × Covid deaths	1.4926***	2.0714***	1.4935***	2.0707***
	(0.1053)	(0.1095)	(0.1057)	(0.1102)
Notable deaths	-0.0389***	-0.0391***	-0.0410***	-0.0412***
	(0.0125)	(0.0128)	(0.0127)	(0.0130)
Black police-related deaths	Y	Y	Y	Y
Unemployment	Y	Y	Y	Y
Weeks post Notable Death	4	3	4	3
County FE	Y	Y	Y	Y
State-Week FE	Y	Y		
County Week Trend			Y	Y
Observations	96286	96286	96329	96329
F First Stage (COVID)	18.03	17.92	32.23	32.09
F First Stage (Interaction)	13.05	13.87	14.59	14.97

Note: Estimation of the effect of Notable deaths and COVID-19 deaths on different Black Lives Matter measures. This table presents 2SLS results, using the cumulative number of all super-spreader events in neighbouring counties (50km radius) as an instrument. Columns (1) and (3) presents the effect of instrumented cumulative number of COVID-19 deaths and notable deaths on the likelihood of having a BLM event in the county within 4 weeks of the notable death. Column (2) and (4) presents the effect of instrumented cumulative number of COVID-19 deaths and notable deaths on the likelihood of having a BLM event in the county within 3 weeks of the notable death. All specifications include county fixed effects and two time varying controls (the number of black police-related deaths and the unemployment rate both at a county level) along with either state-week fixed effects or county week time trend to increase precision. Weekly data by county from year 2014 until the 14th June 2020. Standard errors clustered at the county level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix C: Additional Figures and Tables

Figure C1: Distribution of Super Spreader Events in the US by their type

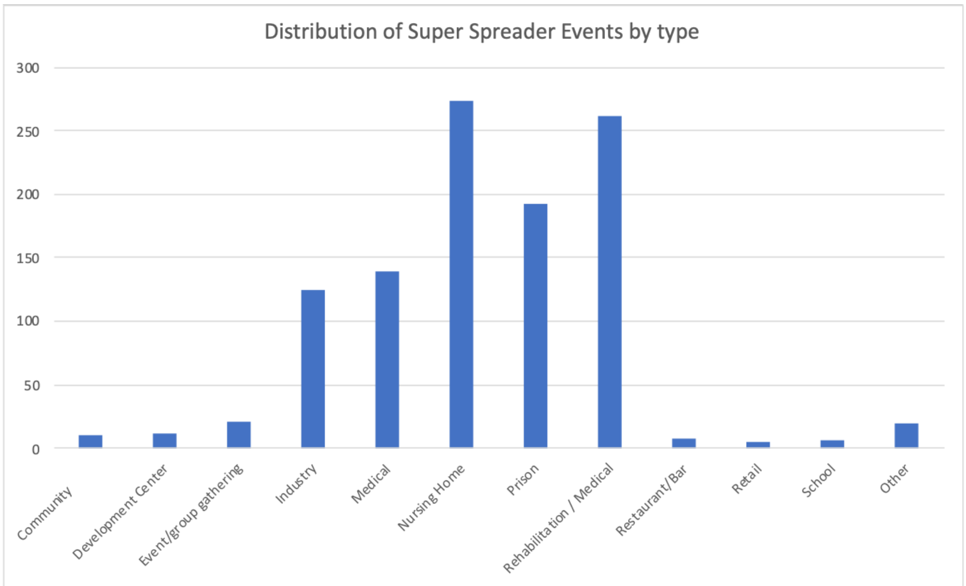
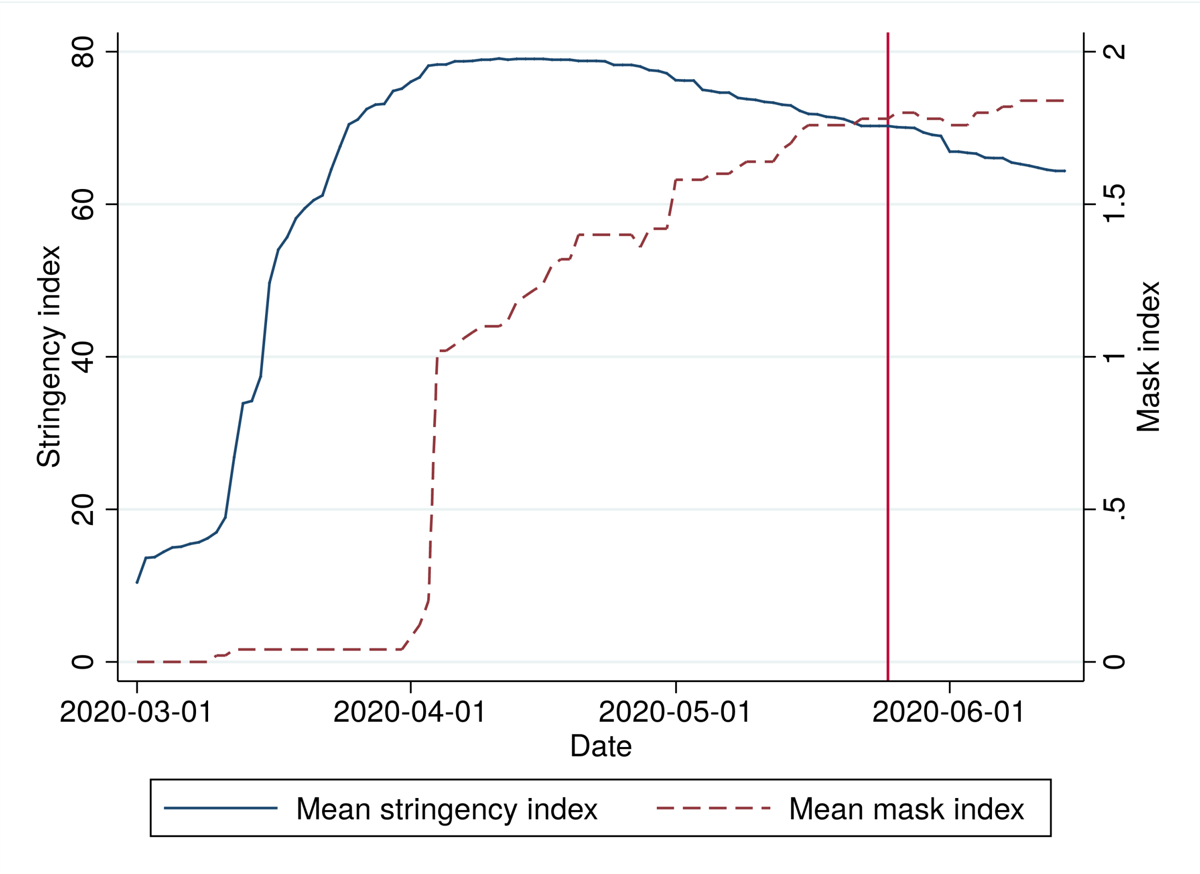
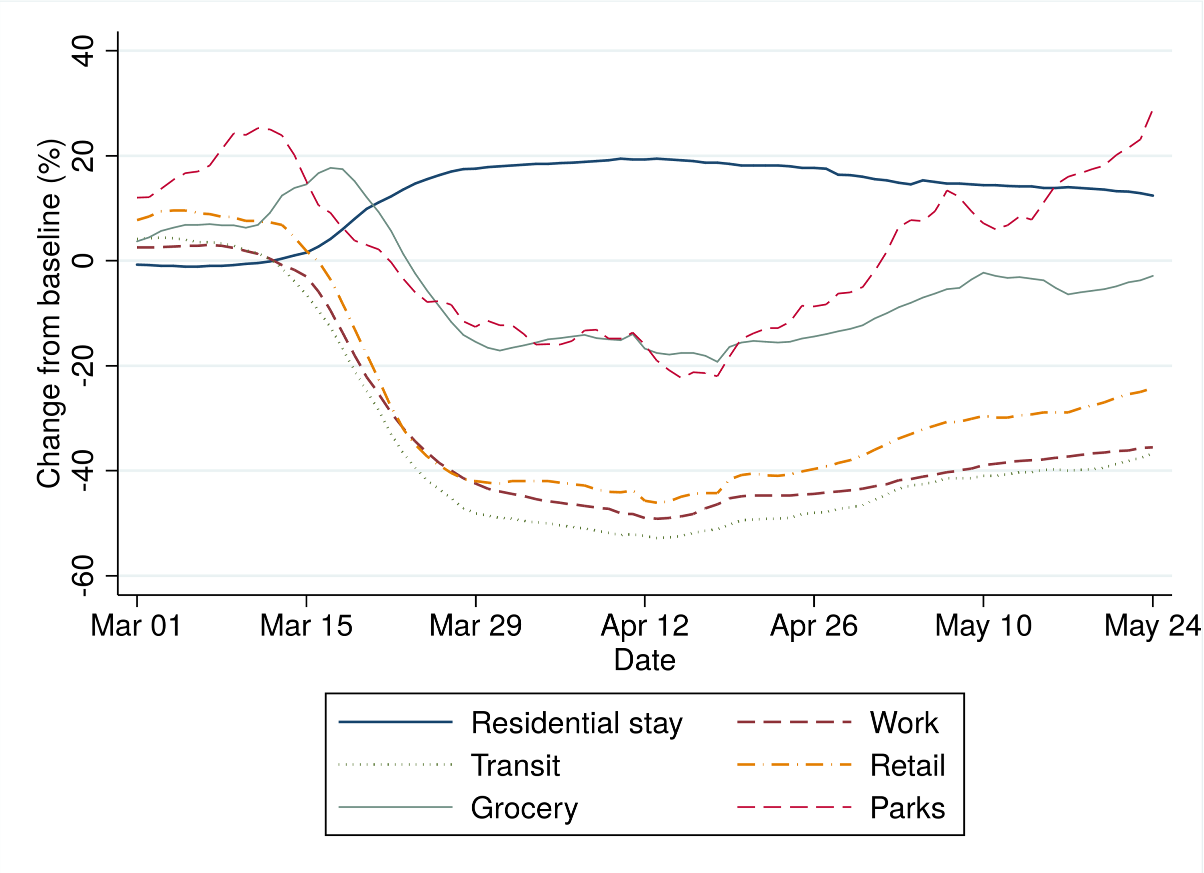


Figure C2: Evolution of lockdown stringency index, and masks recommendations



Note: This graph represents two indicators of average health and lockdown measures in the US over the period from March 1st to June 14th 2020. The blue continuous lines represents the mean lockdown stringency index. The red dashed lined isolates only the indicator for mask recommendations and mandates. The vertical line corresponds to the murder of George Floyd.

Figure C3: Evolution of mobility index



Note: This graph represents the components of the Google Community Mobility index: residential stay, and mobility to different types of places, between March 1st and May 24th, 2020. The index is relative to the average mobility to these places in the same day of the week between January 3 and February 6, 2020. The displayed value is an average of the 7 previous days.

Table C1: Summary statistics for super spreading events by their type

Type of SSE event	Total events	Total Events 6 weeks before GF's murder	Mean	Standard Deviation	Total Cases
Community	11	9	1.364	0.505	504
Development Center	12	12	3.833	1.404	1612
Event/group gathering	21	13	3	1.549	1083
Industry	125	87	15.656	8.642	17825
Medical	140	134	36.586	17.037	13731
Nursing Home	273	261	80.597	37.073	26684
Prison	193	187	45.487	19.674	49747
Rehabilitation / Medical	262	251	89.618	41.009	26979
Restaurant/Bar	8	4	1.5	0.535	1306
Retail	5	0	1	0	68
School	7	2	1.286	0.488	218
Other	20	15	2.5	1.051	1592

All super spreading (SSE) in the USA by their type. Total events are total number of SSE event of each type occurring till 29 August. Total Events 6 weeks before GF's murder is sum of all SSE events by their type that occurred 6 weeks before GF's death. Total cases is sum of all reported COVID-19 positive cases attributed to each type of SSE event.

Table C2: Summary statistics, depending on whether counties had protests before the murder of George Floyd

	All counties					No BLM event before					Has BLM event before				
From 25th of May to 14th of June 2020:	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Presence of BLM events	3107	0.099	0.299	0.000	1.000	2768	0.048	0.213	0.000	1.000	339	0.519	0.500	0.000	1.000
Number of BLM events	3107	0.256	1.381	0.000	36.000	2768	0.064	0.322	0.000	5.000	339	1.823	3.730	0.000	36.000
Participants in BLM events	3107	279.8	5988.7	0	323688	2768	21.0	172.1	0	5500	339	2392.5	18008.4	0	323688
On the 25th of May 2021:															
COVID deaths (total)	3107	24.6	141.3	0	3304	2768	8.4	46.4	0	1025	339	157.1	382.2	0	3304
COVID cases (total)	3107	462.1	2441.7	0	72010	2768	164.5	663.3	0	15169	339	2892.5	6673.3	0	72010
COVID deaths (per 1000)	3107	0.113	0.248	0.000	2.935	2768	0.099	0.230	0.000	2.935	339	0.225	0.345	0.000	2.010
COVID cases (per 1000)	3107	2.794	5.666	0.000	145.513	2768	2.596	5.662	0.000	145.513	339	4.413	5.437	0.000	40.048
Superspreader events, 6+ weeks ago, neighboring	3107	3.081	9.807	0	143	2768	2.327	7.564	0	143	339	9.236	19.310	0	140
County characteristics:															
Black police-related deaths (2014-2019)	3107	0.684	3.227	0	84	2768	0.207	0.724	0	15	339	4.575	8.623	0	84
Black police-related deaths (2020)	3107	0.047	0.301	0	6	2768	0.014	0.131	0	3	339	0.313	0.782	0	6
Unemployment rate (year average)	3107	4.691	1.550	0.708	19.650	2768	4.713	1.575	0.708	17.442	339	4.510	1.323	2.492	19.650
Black population share	3107	0.100	0.147	0.000	0.875	2768	0.093	0.146	0.000	0.875	339	0.158	0.143	0.009	0.727
Urban counties	3107	0.020	0.141	0	1	2768	0.001	0.027	0	1	339	0.180	0.385	0	1
BLM events (2014-2019)	3107	0.631	4.248	0	117	2768	0.000	0.000	0	0	339	5.779	11.661	0	117
Black poverty rate	3107	0.281	0.225	0.000	1.000	2768	0.283	0.236	0.000	1.000	339	0.263	0.099	0.000	0.600
Population share with 3+ risk factors	3107	25.90	5.02	10.68	48.45	2768	25.96	5.07	10.68	48.45	339	25.45	4.60	11.76	39.45
Vote share for republicans (2016)	3107	0.633	0.156	0.041	0.960	2768	0.656	0.141	0.083	0.960	339	0.445	0.144	0.041	0.818
Vote share for republicans (2012)	3107	0.596	0.148	0.060	0.959	2768	0.614	0.140	0.060	0.959	339	0.455	0.132	0.073	0.823
Median household income (2016)	3107	48807	13289	20171	129150	2768	47522	12362	20171	129150	339	59298	15738	28626	120937
Social capital	3107	1	1	0	7	2768	1	1	0	7	339	1	0	0	3
Notable Deaths	3107	0.010	0.116	0	3	2768	0.001	0.033	0	1	339	0.080	0.330	0	3

Note: Summary of main variables used in our analysis. We report the number of observations, the mean, the standard deviation as well as the minimum and maximum value of each of the variables.

Table C3: COVID deaths interacted with county characteristics - Counties without BLM events before

	Presence of BLM events								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Counties without BLM events before									
COVID deaths/1000	0.404**	-0.776	0.421	-0.214	1.083*	-0.202***	0.134	0.457**	0.433**
	(0.187)	(0.920)	(0.294)	(0.257)	(0.618)	(0.0485)	(0.221)	(0.201)	(0.195)
... × Non-black population share		1.305							
		(1.048)							
... × Non-white population share			-0.0523						
			(1.003)						
... × Median household income				7.03e-06**					
				(2.94e-06)					
... × Vote Republican 2016					-1.386				
					(1.236)				
... × Not large cities						0.242***			
						(0.0555)			
... × Suburban areas							0.333*		
							(0.176)		
... × Smaller towns								-0.320	
								(0.237)	
... × Rural areas									-0.179
									(0.187)
Interacting variable		-0.143	-0.102	1.31e-06	-0.328*	0.230***	-0.0416	-0.0186	
		(0.306)	(0.0822)	(8.56e-07)	(0.183)	(0.0703)	(0.0284)	(0.0228)	
Interacting variable		-0.143	0.102	0.230***	1.31e-06	0.0701**	-0.328*		
		(0.306)	(0.0822)	(0.0703)	(8.56e-07)	(0.0303)	(0.183)		
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,767	2,767	2,767
All controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note: Estimation of the effect of COVID-19 deaths per 1000 inhabitants on first-time BLM protest, interacted with county characteristics. We present results for the sub-sample of counties with no BLM protest before the murder of George Floyd. All specifications include state fixed effects and all standard controls. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table C4: Alternative Mechanisms

	Presence of BLM				Other Protests
	(1)	(2)	(3)	(4)	(5)
Panel A: All counties					
COVID (deaths/1000)	0.279**	0.570*	0.252	0.890	0.170
	(0.119)	(0.289)	(0.424)	(1.066)	(0.138)
. . . ×Black death burden	1.017				
	(0.888)				
. . . ×Google BLM search		-0.015			
		(0.010)			
. . . ×Unemployment			0.006		
			(0.030)		
. . . ×Stringency				-0.007	
				(0.0146)	
Interacting variable	-0.195	0.001	0.008*	0.001	
	(0.176)	(0.001)	(0.005)	(0.0013)	
Observations	3,106	3,056	1,351	3107	3,106
F stat COVID	25.59	22.14	27.49	96.71	31.4
F stat Interaction	12.46	58.19	27.49	96.04	
Mean of dependent variable	0.099	0.099	0.099	0.099	0.081
All controls	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y		Y

Note: Estimation of the effect of COVID-19 deaths per 1000 population on presence of BLM protest. Column 1 shows estimates for instrumented COVID deaths. Columns 2 to 4 show heterogeneous effects for Black death burden weeks prior to GF's murder, Google searched for BLM 3 weeks prior to GF's murder, unemployment and stringency 3 weeks after GF's murder. Column 5 presents results for other protests. Panel A presents 2SLS estimation for all counties. Panel B presents these results for the sub-sample of counties with no BLM protest before the murder of George Floyd. All specifications include state fixed effects and standard controls. We report Kleibergen-Paap rkWald F statistic. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table C5: Effect of SXSW users on Twitter presence

VARIABLES	(1) Log(Preexisting users)	(2) Log(New users)	(3) Presence of BLM events
Log(SXSW users)	0.373***	0.193***	0.0151
	(0.103)	(0.0505)	(0.0175)
SSE			-0.00117
			(0.00257)
$\times$ SXSW users			0.00439**
			(0.00172)
Mean of dep. var	1.738	0.420	0.0477
F first stage	13.02		
Observations	2,767	2,767	2,767
Instruments			
All controls	Y	Y	Y
Pre-SXSW users	Y	Y	Y
State fixed effects	Y	Y	Y

Note: Column 1 shows the first stage regression for predicting existing Twitter users at the end of 2019 in the county using SXSW followers that joined Twitter during the festival in the county and its neighboring counties. Column 2 shows the same effect on the users created during COVID-19. Column 3 shows the reduced-form effect of SXSW followers interacted with superspreader event on the presence of protest. We present results for the sub-sample of counties with no BLM protest before the murder of George Floyd. All specifications include state fixed effects and all standard controls. Standard errors (in parentheses) are clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1

Table C6: Summary statistics, depending on whether tweets have been located in a county

	All counties					Counties with tweets					Counties without tweets				
From 25th of May to 14th of June 2020:	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Presence of BLM events	3107	0.099	0.299	0.000	1.000	142	0.070	0.257	0.000	1.000	2965	0.101	0.301	0.000	1.000
Number of BLM events	3107	0.256	1.381	0.000	36.000	142	0.218	1.483	0.000	17.000	2965	0.257	1.376	0.000	36.000
Participants in BLM events	3107	279.8	5988.7	0	323688	142	222.3	2376.3	0	28290	2965	282.5	6108.5	0	323688
On the 25th of May 2021:															
COVID deaths (total)	3107	24.6	141.3	0	3304	142	9.3	48.7	0	432	2965	25.3	144.2	0	3304
COVID cases (total)	3107	462.1	2441.7	0	72010	142	193.2	946.5	0	8110	2965	475.0	2490.2	0	72010
COVID deaths (per 1000)	3107	0.113	0.248	0.000	2.935	142	0.094	0.205	0.000	1.507	2965	0.114	0.250	0.000	2.935
COVID cases (per 1000)	3107	2.794	5.666	0.000	145.513	142	2.533	4.730	0.000	30.975	2965	2.807	5.707	0.000	145.513
Superspreader events, 6+ weeks ago, neighboring	3107	3.081	9.807	0	143	142	2.415	6.013	0	40	2965	3.113	9.952	0	143
County characteristics:															
Black police-related deaths (2014-2019)	3107	0.684	3.227	0	84	142	0.479	3.050	0	29	2965	0.694	3.235	0	84
Black police-related deaths (2020)	3107	0.047	0.301	0	6	142	0.014	0.118	0	1	2965	0.049	0.307	0	6
Unemployment rate (year average)	3107	4.691	1.550	0.708	19.650	142	4.062	1.491	0.708	11.533	2965	4.721	1.547	1.642	19.650
Black population share	3107	0.100	0.147	0.000	0.875	142	0.100	0.156	0.000	0.784	2965	0.100	0.147	0.000	0.875
Urban counties	3107	0.020	0.141	0	1	142	0.021	0.144	0	1	2965	0.020	0.141	0	1
BLM events (2014-2019)	3107	0.631	4.248	0	117	142	0.852	6.408	0	64	2965	0.620	4.117	0	117
Black poverty rate	3107	0.281	0.225	0.000	1.000	142	0.223	0.266	0.000	1.000	2965	0.284	0.223	0.000	1.000
Population share with 3+ risk factors	3107	25.90	5.02	10.68	48.45	142	26.14	5.16	16.95	41.57	2965	25.89	5.01	10.68	48.45
Vote share for republicans (2016)	3107	0.633	0.156	0.041	0.960	142	0.661	0.193	0.041	0.960	2965	0.632	0.154	0.083	0.946
Vote share for republicans (2012)	3107	0.596	0.148	0.060	0.959	142	0.632	0.178	0.073	0.910	2965	0.595	0.146	0.060	0.959
Median household income (2016)	3107	48807	13289	20171	129150	142	49124	15420	22120	119153	2965	48791	13181	20171	129150
Social capital	3107	1	1	0	7	142	2	1	0	6	2965	1	1	0	7
Notable Deaths	3107	0.010	0.116	0	3	142	0.000	0.000	0	0	2965	0.010	0.119	0	3

Note: Summary of main variables used in our analysis. We report the number of observations, the mean, the standard deviation as well as the minimum and maximum value of each of the variables.

Table C7: Principal component analysis of online presence

(a) Correlation between measures

	New Twitter accounts	New Twitter accounts (log)	Google searches for Twitter	Residential stay
New Twitter accounts	1			
New Twitter accounts (log)	0.379***	1		
Google searches for Twitter	0.0558*	0.234***	1	
Residential stay	0.0770**	0.355***	0.520***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Principal components

	Eigenvalue	Difference	Proportion	Cumulative
PC1	1.845167	.7217762	0.4613	0.4613
PC2	1.123391	.5386709	0.2808	0.7421
PC3	.5847198	.137997	0.1462	0.8883
PC4	.4467228	.	0.1117	1.0000

(c) Factor loadings

	PC1	PC2	PC3	PC4
New Twitter accounts	.3265664	.7302487	.5601388	.2152575
New Twitter accounts (log)	.5339498	.3756387	-.6544048	-.3815068
Google searches for Twitter	.5247868	-.4478813	.484595	-.5377442
Residential stay	.5769323	-.3536025	-.1522054	.7203804

Note: The first table reports the correlation between the online presence measures. The second table reports the eigenvalues of the four principal components. The third table reports the loading of the different components.

Appendix D: Data Appendix

D.1 Twitter Data

Twitter usage during the protests Twitter data is an important source of information when studying social events and protests. Previous work on BLM events has made use of this data to further understand this movement (Ince et al., 2017). We collected tweets using the Twitter Academic Research API. In particular, we collected all tweets that contain the keywords "BLM", "Black Lives Matter", "Black Life Matters" or "George Floyd",[35] including retweets, between May 25 and June 14. For each tweet, we extract the time and text of the tweet, the user, the user's stated location, and account creation date. We based the assignation of tweets to a geographical location on the location stated by the user in their profile. Not all users state a location and among those who do, not all state a valid location (e.g., "in the heart of Justin Bieber") so we restrict the sample to the users that state a valid location that can be matched to a USA county (in particular, we exclude users whose location only mentions a state). The location is an arbitrary text field which is not meant to be machine-readable. We use the Nominatim geocoding engine (based on the Open Street Map database) to find the coordinates of the most likely match for the location. We then filter out all locations outside the US and all locations that are too vague (i.e. that map the whole country or a whole state). Finally, we map these coordinates to counties using the US Census Bureau cartographic boundary files. We end up with 2.76 million tweets.

Pre-existing Twitter usage and instrument For the study of mechanisms, we use a proxy of pre-existing Twitter usage measured in December 2019. This is measured by sampling all tweets containing the word "the" during random intervals in one week of December 2019. One million tweets were collected from 765 000 users. Users were attributed to counties using the location in their profile. To study causally the effect of pre-existing Twitter usage on the reaction to COVID-19, we collected data to reproduce the SXSW instrument used by Müller and Schwarz (2019): we collected in November 2021 the locations of all 639 915 followers of the @SXSW Twitter account as well as the date they joined the network.

Google Searches We also use the Google Trends data to analyze patterns of search activity before and after the death of George Floyd. Each variable is a normalized index of search activity for given search term. The indices are specified on a Nielsen's Designated Market Area (DMA) level. A DMA is a region of the United States that consists of counties and ZIP-codes. There are 210 DMA regions covering the US. Search activity is averaged across the period of interest: each observation is a number of the searches of the given term divided by the total searches of the geography and time range, which is then normalized between regions such that the region with the largest measure is set to 100. The important limitation of the Google Trends data is that an index of search activity is an integer from zero to one hundred with an unreported privacy threshold. The search terms that were used in the analysis are presented in Table **??**.

Safe Graph In our alternative identification strategy we employ an instrumental variable based on data provided by the data company Safegraph. The Safegraph data is GPS location data that reveal the spatial mobility of population between the points of interests. For the region of interest (three vacation destinations in Florida: Miami Beach, Panama Beach and Fort Lauderdale) the SafeGraph data provide rich set of points of interests, which include more than

[35] These keywords are considered both in when appearing separated with space, or without spaces as a hashtag (e.g. #BlackLivesMatter)

3000 places such as restaurants, bars, hotels, gyms, public parks, malls and other establishments. Using this data, we measure the Number of devices that "pinged" in each of point of interests during March, 2020. Using these data we can also observe home locations on the level of the US Census Block Groups (CBG). An individual "home" is defined as a place where user's devices pinged most often in the night time between 6 PM and 7 AM during the baseline 6-week period determined by the SafeGraph.

Elephrame Elephrame is a crowd-sourced platform that collects data on Black Lives Matter and other protests. It provides information on the place and date of each BLM protest and estimated number of participants, as well as a link to a news article covering the protest. We extracted all protests' records from June 2014 to September 2020 and geo-coded their location. The observation period starts with the first BLM demonstration for Eric Garner on 7/19/2014 and consist of any public demonstration or public art installation focused on "communicating the value of a Black individual or Black people as a whole". Each observation is manually collected by the creator of Elephrame Alisa Robinson from sources that include press, protests' organizers, participants and observers.

Notable Deaths To exploit the panel aspect for our estimation, we create a county week varying panel data set of notable deaths. Notable deaths are defined as deaths of African American at the hands of a police officer and which make it to national media and/or have a dedicated Wikipedia page. This data set includes personal information of the victim like their name, age, sex and race. It also has details about the event like the county and zip code of the place where shooting took place, cause of death, whether the victim was armed, if a video of the incidence was shot by onlookers and if the police officer wore a body camera. We also collect information on date of the shooting, date of the official verdict from this incidence and whether the police officer was convicted. From 2014 till 2020, we have 34 notable deaths from all over the country. Average age of victim is 34 years, 31 out of 34 are men. All victims in our data are Black.