University of Essex
Department of Mathematical Sciences

MA831: Capstone Project Dissertation

To What Extent Do Demographic Factors and Mobility Restrictions Affect COVID-19 Case Rates in Germany and England?

Bernhard Finke

Supervisor: **Dr Andrew Harrison**

7 April, 2021
Colchester

Contents

List of Figures

CHAPTER 1

Introduction

Since the observation of the first case of COVID-19 in Wuhan, China in December 2019, the highly contagious disease has spread rapidly to all regions of the world. Due to the recency of the discovery of the novel coronavirus, many important questions about the virus's nature remain without definite answers. One of the main areas of uncertainty remains the question of which factors affect transmission of COVID-19, and to what extent. Though it is generally accepted that social distancing and reduced contact with others helps to depress case rates [11], it is harder to say how best to achieve these goals, or what underlying demographics lead to a greater risk of catching COVID-19. For governments, understanding which situations lead to a higher risk of COVID-19 spread can help to guide restriction guidelines. On the one hand, governments can use this information to effectively target high-risk activities, while also avoiding Draconian measures which may not even help to keep the coronavirus at bay.

This report aims to answer two related questions numerically: which demographic factors lead to higher case rates in an area, and how specific government restrictions on movement and associated changes in behaviour lead to changes in COVID-19 case rates? To tighten the scope of the analysis, we will be looking only at two nations: England and Germany. On the one hand both are rich Western European liberal democracies, so the countries are similar enough to warrant comparison. On the other hand government responses have differed greatly at times, so we can contrast the two approaches, with Germany's pandemic response widely considered 'a model for tackling the disease' [7], at least until the beginning of the second wave, while the British response has been characterized as 'a fatal failure' [40].

Specifically, we will begin by comparing overall national case statistics for the two countries and the impact of lockdowns. Additionally, we will be breaking down both countries into first-level geographical divisions, in Germany the 16 federal states and in England the 9 regions, and comparing total cases by region to potential explanatory factors, such as poverty rate. For Germany, where the data are readily available, we will also look at differences in case distributions across gender and age groups to determine which groups are at highest risk of infection. Then, we will move on to investigating the impact of social distancing by looking at mobility data for both countries. First, we will look at the influence of particular events, such as the imposition and loosening of restrictions on overall mobility, before moving on to quantifying the relationship of changes in mobility with changes in case rates. The analysis uses mobility and case data for both countries up to the end of 2020 [43][12][20].

CHAPTER 2

Literature Review

As the onset of the pandemic occurred so recently, much of the academic research on COVID-19 is still in the process of publication. Nevertheless, a handful of articles have been released which answer questions about demographic factors and COVID-related mobility. These studies vary significantly in the timeframe of data used, statistical methodologies, as well as the regions investigated.

In their research article, Roy and Ghosh [32] attempt to identify which input factors have a significant impact on COVID-19 cases. Their analysis contrasts case data from the 50 states of the United States of America over the period January to July 2020. A similar study was carried out by Velasco, Tseng and Chang [44], which instead focused on 141 different countries, using data up to 10 December, 2020. Roy and Ghosh [32] carried out their analysis using time series while Velasco et. al [44] used total cases per million as of 10 December as a metric for the overall severity of COVID-19 spread in a country. Though Roy and Ghosh's article [32] used supervised learning methods and Velasco et. al [44] relied on ordinary least squares regression analysis, both papers came to the conclusion that testing numbers were a significant factor in COVID-19 case prevalence. While Roy and Ghosh [32] found population density to be a significant factor generally, Velasco et. al [44] only found this to be true for countries with a population less than 10 million. Factors that were determined to have no impact on incidence include ethnicity, gender, healthcare index, amount of homeless, and GDP. Roy and Ghosh [32] also found that post-lockdown case rates were strongly affected by pre-lockdown figures, as well as that a significant drop in test-positivity rates occurred three weeks into a lockdown.

A study by Pijls et. al [28] identified differences in infection rates for sex and age groups. The paper is a meta-analysis of 59 studies, 50 of which were on Chinese patients. Pijls et. al found that men had a higher risk of COVID-19 infection than women. Further, they also concluded that those aged 70 years or older had a 65% higher risk of infection than those aged under 70.

Tupper, Boury, Yerlanov, and Colijn [41] find that for events without inter-group mixing, distancing is a more effective intervention than other reduction methods including face masks, hand sanitizer, and strict bubbling. Here, distancing can refer to spacing individuals farther apart, stopping close-range droplet transmission and reducing the amount of attendees. Mendolia, Statrunova, and Yerokhin [24] attempted to differentiate between changes in mobility due to voluntary behavioural changes and changes due to government-mandated measures in 133 countries. Overall, the study suggests that, in most countries, multiple types of mobility were lowered by over 50% within 20 days of first infection, compared to the baseline. This reduction was slightly lower for grocery and workplace mobility, while residential mobility increased. The analysis found decreases in mobility associated with prior increases in cases and deaths, but only for the first 200 cases and two deaths. Nevertheless, the study concludes that around 50% of permanent reductions in mobility are explained by government restrictions, with only 14% associated with information about COVID-19 cases in the country. This suggests that reducing mobility and increasing social distancing is most effectively done through government restrictions. A similar study by Pullano, Valdano, Scarpa, Rubrichi, and Colizza [29] found a 65% decrease in total trips per day in France between the period before announcement and implementation of lockdown and the period after lockdown took effect. Additionally, weekly changes in mobility by region were strongly associated with COVID-19 cases and deaths registered in the week prior.

Finally, we turn to the effect of change in mobility on change in infection rates. In an analysis of cases by New York City borough, Roy and Ghosh [32] identified that inter-borough mobility is a better indicator of spread than inter-borough distance. A study on the effects of mobility restrictions in Shenzen, China [48] found that a reduction in mobility between 20 and 60% led to a significant effect on COVID-19 spread. A greater reduction in mobility led to a greater reduction in cases, with a 60 percent reduction in mobility associated with a 91% reduction in cases.

The overview of available literature above lends credibility to our major hypotheses, namely: that demographic factors, such as population density, have an effect on the level of COVID-19 infections, that there are significant differences between sex and age groups, that government restrictions have a strong effect on mobility, and that changes in mobility are associated with changes in case incidence. As much of the currently available analysis focuses on specific countries, it is worthwhile to investigate whether trends in countries which have already been studied are reproducible in England and Germany.

CHAPTER 3

Timeline of the Pandemic in Germany and England

This chapter gives an overview of significant events in the timeline of the COVID-19 pandemic for Germany and England, particularly those related to lockdowns restrictions.

TABLE 3.1 Timeline - Germany, 2020 [22]

27th January	First confirmed cases in Germany
1st March	100 confirmed cases in Germany
20th March	Bavaria and Saarland impose curfews
23rd March	Contact ban issued (first lockdown begins)
20th April	Shops with a retail space of up to 800 $m.^2$ allowed to reopen
6th May	Merkel declares first wave 'over', meeting of two households allowed, all shops allowed to open [33]
2nd November	Second 'partial' lockdown begins
1st December	Lockdown restrictions tightened, meetings of more than 5 people no longer allowed
16th December	Restrictions further tightened, retail businesses and schools closed

TABLE 3.2 Timeline - England, 2020 [2]

29th January	First confirmed case in England
5th March	100 confirmed cases in England and first COVID-19 related death
20th March	Pubs, restaurants, gyms and other social venues ordered to close, places of work closed
23rd March	First lockdown begins as daily cases number more than 2000
10th May	First easing of lockdown, outdoor exercise allowed
22nd May	Newspapers report that a top government advisor, Dominic Cummings, broke lockdown regulations in March [46]
28th May	England launches Test & Trace System, allowing general access to COVID-19 tests
1st June	Stay-at-home order revoked, as U.K. records lowest increase in daily cases and deaths since end of March
4th July	Pubs, restaurants and hotels reopen
17th July	Use of public transport for non-emergencies allowed again [18]
3rd August	Eat Out to Help Out scheme begins, aimed to help restaurants stay in business
31st August	Eat Out to Help Out scheme ends
20 September	Kent variant first discovered, variant is considered much more infectious [30]
5th November	Second national lockdown begins, with non-essential retail closed
2nd December	National lockdown ends, replaced by regional tier system

CHAPTER 4

Methodology

4.1 Datasets

4.1.1 England COVID-19 Statistics

This project chooses to analyse cases in only England, rather than the whole of the U.K., as the coronavirus response of the four U.K. nations differed significantly. Thus it would be difficult to separate innate regional differences and differences in coronavirus policy. The COVID-19 case statistics for England were found on the section for open data of the U.K. government website [43]. The main dataset used in analysis of English cases consists of overall new cases by day and English region. The specific metric used to tally cases is 'new cases by specimen date', the date on which a test was received from a patient. While 'new cases by publish date' is the more widely used method to count daily cases, measuring by specimen date should give a more accurate impression of the current situation, as there is a shorter lag between a new infection and its appearance in the official figures.

Most of the analysis of this data will compare changes in daily cases over time, though in section 5.3, total cases will be aggregated by region, as a metric for the severity of spread in each geographical area. Additional data containing demographic statistics for each region were attained from various government reports and databases [17][26][15]. Also in section 5.3, we will use a different version of the dataset, which collects daily cases not by region but by upper tier local authority to investigate the effect of population density on spread [42].

4.1.2 Germany COVID-19 Statistics

The COVID-19 case statistics for Germany were sourced from the German COVID-19 Data Hub, operated by the Environmental Systems Research Institute, and backed by the Robert Koch Institute, the main research institute of the German Federal Ministry of Health [12]. The dataset splits daily cases up by federal state, as well as by district, a rough equivalent of upper tier local authority in England. Unlike the English dataset, the German data use 'new cases by publish date' to quantify cases by date.

Additionally, the dataset is also broken up by age group and sex. This additional information will be used in section 5.4 to investigate whether significant differences exist in the distribution of cases across these groups. As with the English data, most of the analysis will revolve around looking at daily cases over time, with the analysis in section 5.3 comparing cumulative cases by federal state to additional demographic factors [27][47][38]. As the data already differentiate cases by district, no additional dataset will be required in the analysis of population density and spread.

4.1.3 COVID-19 Community Mobility Reports

Once we have investigated trends in case statistics, we will move on to trends in mobility. To do this we will use the Google Community Mobility Reports for both England and Germany, developed by Google specifically for use in aiding public health officials [20]. The dataset uses mobile phone location data to chart changes in mobility, compared to a baseline, the median mobility value over the 5-week period from January 3rd to February 6th 2020.

The dataset contains the percentage change of mobility, given by a combination of number of visits and length of stay, from the baseline period. The dataset breaks up changes both by country overall, as well as by geographical sub-division. Change in mobility is broken up into six different types: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential.

In chapter 6 we will explore the mobility datasets for both countries, looking specifically at the impact of certain events on mobility. Finally, we will combine the case data and mobility data to determine the effect of mobility changes on changes in spread.

4.2 Statistical Methods Used

All data analysis was completed using R version 3.6.1. Firstly the data was prepared for analysis using the Tidyverse packages: dplyr and tidyr. This usually consisted of renaming variables, summarising data to remove unnecessary variables, and reshaping data frames. The package ggplot2 was used to produce plots. Statistical testing was then carried out using the following methods.

4.2.1 Pearson Correlation Coefficient

The Pearson correlation coefficient is a measure of linear dependence [35]. It can take values from -1 to 1, corresponding to negative and positive correlation respectively. The further the correlation coefficient is from 0, the more closely related the two variables are. In R, the correlation coefficient is implemented in the function *cor.test(method = "pearson")*.

4.2.2 Simple & Multiple Linear Regression

Linear regression is used to predict a response variable, usually y, based on a number of explanatory variables, usually x [21]. In simple linear regression, we use only one explanatory variable. The associated F-statistic of a multiple linear regression tells us whether at least one of the explanatory variables has a significant relation to the output variable. Looking at the t-statistic associated with each x variable can tell us whether this variable has a significant impact on the output. We can assess the accuracy of a regression model using it's R-squared. R-squared takes a value from 0 to 1, with 1 meaning all output variance is explained by variance in explanatory variables. In R, linear regression is implemented in the function *lm()*.

4.2.3 Unpaired Two-Sample t-Test

This kind of t-test is used to compare the means of two different independent groups [37]. The null hypothesis is that the means are equal, while the alternative hypothesis is that they are not. In R, the unpaired two-samples t-test is implemented in the function *t.test(alternative="two.sided", var.equal=FALSE)*.

4.2.4 One-Way Analysis of Variance Test

The one-way analyis of variance, or ANOVA, is a method to compare means between more than two groups [36]. It compares the variance within samples, also known as common variance, to the variance between sample means. The lower the ratio below 1, the less significant the differences are between means. In R, the ANOVA test is implemented in the function *aov()*. A Tukey Honest Significant Differences Test (HSD) can then be used to calculate pairwise comparisons between groups means. A Tukey HSD Test is implemented in the function *TukeyHSD()*.

4.2.5 Cross-Correlation Function

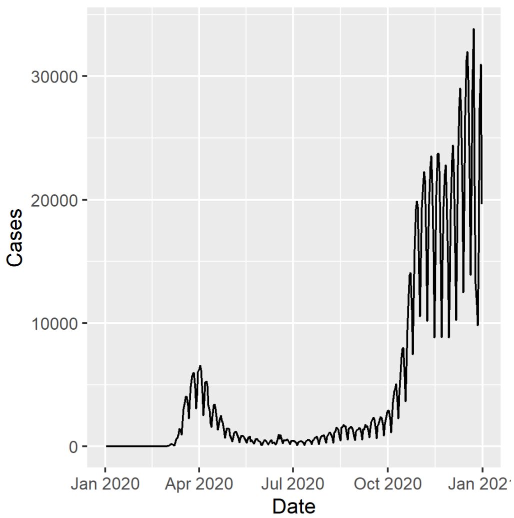
The cross-correlation function (CCF) can be used to determine lags of a predictor variable on a response variable [9]. CCF calculates sample correlations for a number of different lags between the x and y variables. In R, CCF is implemented in the function *ccf()*.

CHAPTER 5

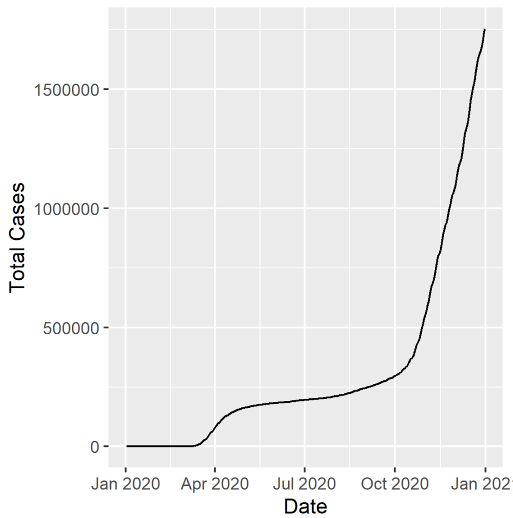
Investigating COVID-19 Case Statistics

5.1 Overview of National Case Statistics

We will begin our analysis by looking at some charts illustrating COVID-19 cases over time for England and Germany.



(a) Daily New Cases.

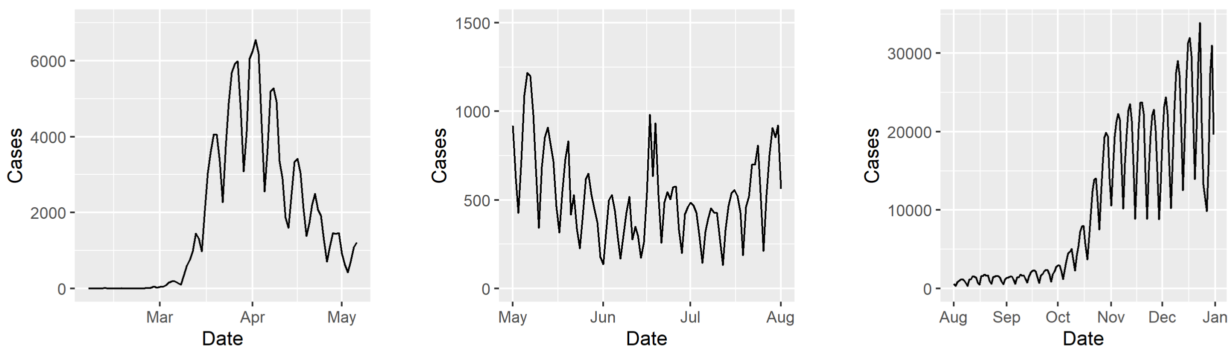


(b) Cumulative Cases.

Figure 5.1: Case Overview Germany, Until 31st December 2020.

Figure 5.1 shows the development of COVID-19 cases in Germany. Up to the end of 2020, Germany had experienced 1,755,091 cases or 2,109.43 cases per 100,000 individuals. We use cases per 100,000 as a per capita estimate as this is the metric used by both England and

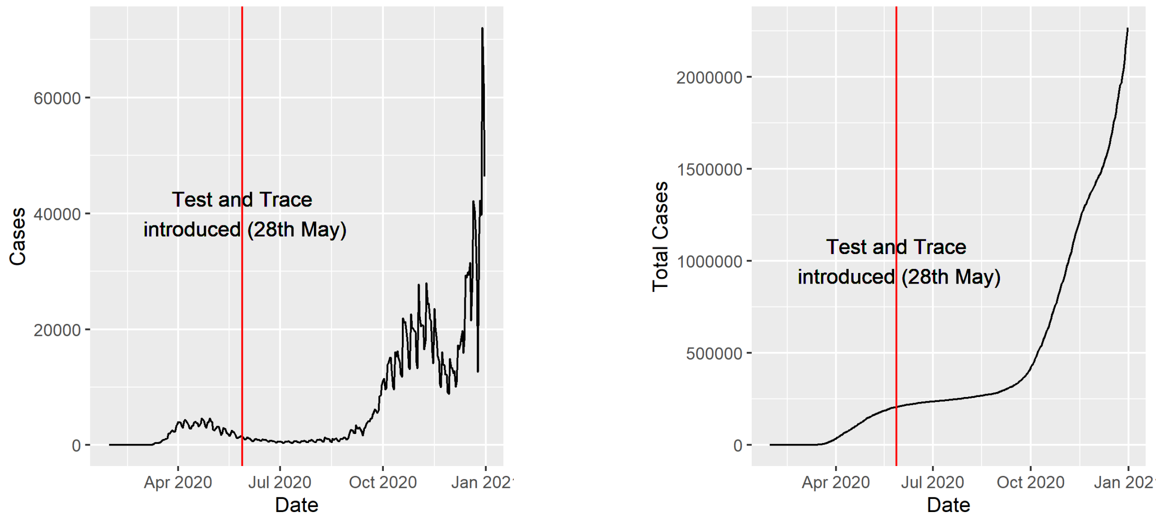
Germany to determine population-adjusted cases. The highest case figure for a single day in 2020 was 33,870, on the 23rd of December. From the plot we can see that the pandemic in 2020 is roughly divisible into three sections: the first wave, which takes place roughly until May and sees the first peak of cases, the summer, the period ranging from May to August beginning around the time Merkel declared the first wave to be over and where cases are much lower and remain low for a few months, and the second wave, beginning in August with a slow uptick in cases until October when cases begin to shoot up to record highs week after week. Panel 5.1b, especially in the second half of the year, highlights the exponential growth of cases often referred to by journalists. We can see that, if case numbers continued to grow as they did in December 2020, the situation would spiral out of control.



(a) First Wave. (b) Summer. (c) Second Wave.

Figure 5.2: Cases per day in Germany, 2020.

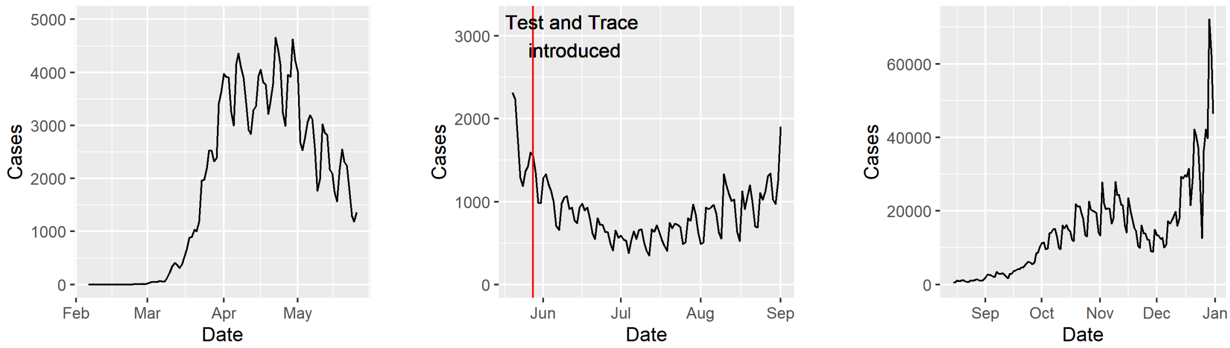
To investigate daily cases in more detail, we break up the plot into the three sections listed above, noting the different scales used in the resulting panels. We can now determine a weekly pattern, with cases highest on Thursdays. This does not necessarily reflect actual weekly fluctuations in infections, instead it is more likely due to the way individual testing stations report cases, with many stations closed on weekends. From panel 5.2a we can now determine that the peak of the first wave occurred on the 2nd of April and saw 6551 new infections. Additionally, we can see that this period is roughly symmetrical, with the highest case numbers occurring two weeks before or after the 2nd of April. In panel 5.2b, we also notice that the plot is roughly symmetrical, with cases generally decreasing week on week until early July, when case numbers begin rising again. This period also sees the lowest case count since March 3rd: July 12th, which had 131 cases. Finally, panel 5.2c shows us the slow increase followed by an explosion in cases around October to November, 2020.



(a) Daily New Cases. (b) Cumulative Cases.

Figure 5.3: Case Overview England, Until 31st December 2020.

We turn our attention to similar plots for England. The red lines show the date on which Test and Trace was instituted, beginning mass testing. Estimates of COVID-19 infections before this date may be inaccurate as only patients with severe symptoms were tested. In fact, models developed using data on deaths, or other more accurate indications of COVID-19 spread, estimate the peak number of cases per day in the first wave as between 99,000 and 178,000 [34], at least 20 times greater than the estimate given by the data. Despite the inaccurate numbers in the first wave, the data can still be used for analysis, provided that raw case numbers before the 28th of May are not compared to those after Test and Trace was introduced. In 2020, England saw 2,267,756 confirmed cases or 4,029.27 cases per 100.000. The highest number of cases occurred on the 29th of December, with 72,081 cases. Again, we can observe the same three sections: the first wave, roughly until the end of May and seeing the first peak of cases, the summer, beginning around the end of the first lockdown, where cases remain relatively low until August, and the second wave, where cases first increase slowly followed by rapidly increasing case figures in October. In panel 5.3b, we also see the exponential shape of cumulative cases.



(a) First Wave. (b) Summer. (c) Second Wave.

Figure 5.4: Cases per Day in England.

Again, splitting up the case chart shows a regular weekly fluctuation, though for England Wednesday is the day of the week with the most cases. Panel 5.4a does not show a symmetrical pattern of cases in the first wave, unlike the German first wave. In fact, we can see that cases remained at a similar very high rate for five weeks, with the absolute peak of 4,661 cases reached on the 22nd of April, before beginning to drop off at a much slower pace than the preceding rise in cases throughout March. The period of relatively low cases between the first and second wave seems to occur about a month later than the equivalent period for Germany, reflecting the fact that the first lockdown was lifted roughly a month later in England. England also saw its lowest case count since early March in this period. In fact, this occurred on the same date as Germany's lowest count, the 12th of July, though with 349 cases on this day, England saw more than twice as many cases as Germany. Unlike in the German second wave, the plot for England shows a temporary decrease in cases around the beginning of December, though cases shoot back up within two or three weeks.

To compare case rates between countries we turn to the plot in figure 5.5, which graphs daily cases for both countries using our population adjusted metric. From the graph we can begin to see why England has almost double the number of confirmed cases per capita as Germany. The only extended period where confirmed case rates were lower for England than for Germany was around April, though due to England not having implemented mass testing yet, it is unlikely that there were actually fewer infections occurring in England during that time. We can also see clearly that the period of stable low case rates during summer is shorter in England. Cases begin to rise again quickly in late August in England, while in Germany this increase does not take place until October. England surpasses the threshold

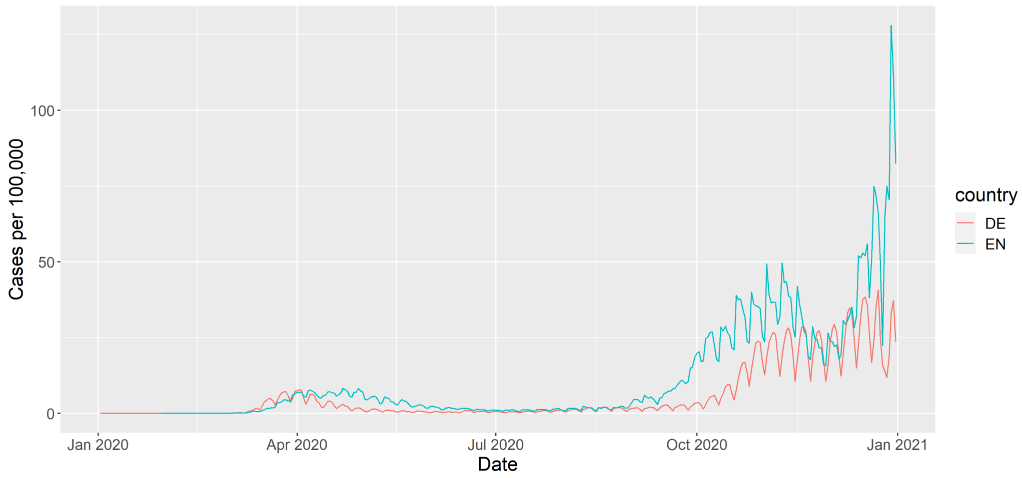


Figure 5.5: Daily Cases per 100,000, Germany and England.

value of 3 cases per 100.000 on the 1st of September, while it takes until the 30th of September for the same to happen in Germany. For both countries we can see a reduction in the growth rate of cases in late November, not long after the imposition of a second lockdown in both countries, though this drop is much more pronounced for England. In the next section we will investigate in more detail the impact of lockdowns on case rates.

5.2 Further Metrics and Impact of Lockdowns

To help quantify changes in case rates more accurately we will introduce two other metrics for measuring COVID-19 case spread, both of which eliminate weekly fluctuations in case reporting. The first alternative metric is 7-day rolling cases per 100,000, that is the sum of cases over the last 7 days adjusted for population. Both the English and German government use 7-day rolling case per 100,000 to estimate the current level of spread. As of the time of writing, Germany uses the threshold of 35 cases per 100,000 to determine when it is safe to remove lockdown restrictions [45]. The second metric we introduce is percentage difference in cases compared to the same day of the prior week. This metric is particularly useful as it allows us to estimate the rate of change of cases. The percentage difference in cases gives us an intuitive way of looking at cases, as we can tell whether the spread of cases is increasing exponentially or slowing down based on whether the percentage is positive.

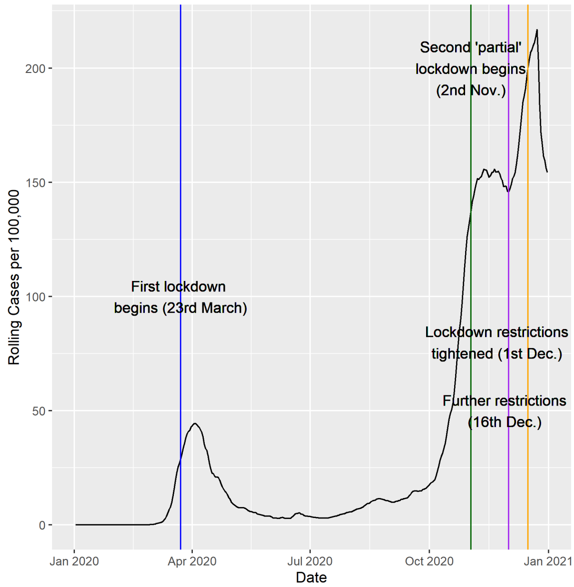
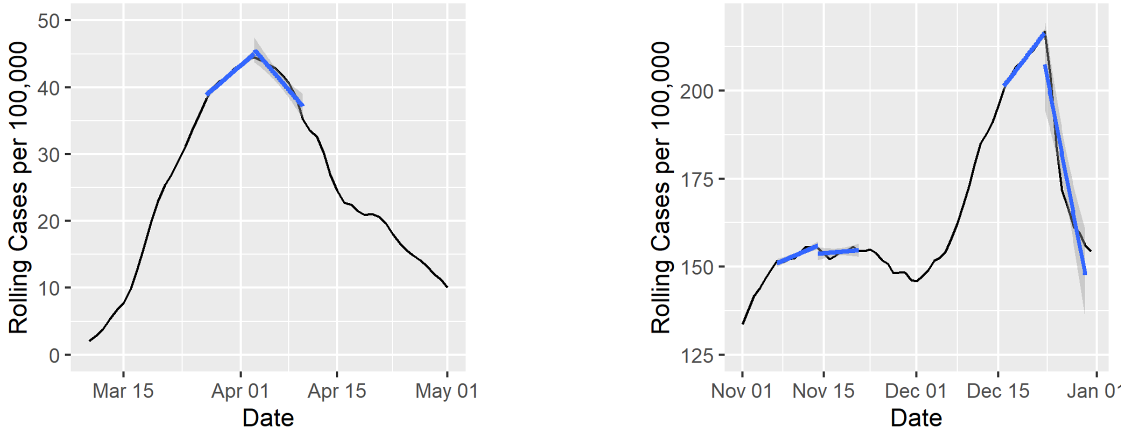


Figure 5.6: Seven Day Population-Adjusted Incidence Rate in Germany.

From the plot in figure 5.6, we can see that using rolling aggregate case rates effectively removes weekday fluctuations. The graph includes dates of the impositions of lockdown restrictions for Germany. After the beginning of the first lockdown, cases peak within 11 days, on the 3rd of April. We can determine that the lockdown was generally effective with cases lowering continually until June, well beyond the end of the restrictions. The same cannot be said of the second, 'partial', lockdown. While cases initially peak on the 14th of November, within 12 days, similarly to the first lockdown, case rates remain around the 150 rolling cases per 100,000 mark for around a month. Noticing that the current restrictions were not effectively reducing cases, the German government decided to tighten the lockdown on the 1st of December, reducing allowed group sizes from 10 down to 5, with a maximum of two households. As the measure was not effective at reducing cases, as shown by the continual rise from the 1st of December, further restrictions were imposed, additionally requiring the closing of retail businesses and schools. Cases peaked again on the 23rd of December, 7 days after the second tightening of lockdown. The peak was followed by a significant drop in cases for the last 8 days of 2020, indicating that the further restrictions were more effective at reducing cases than the partial lockdown on its own, or the first additional restrictions.

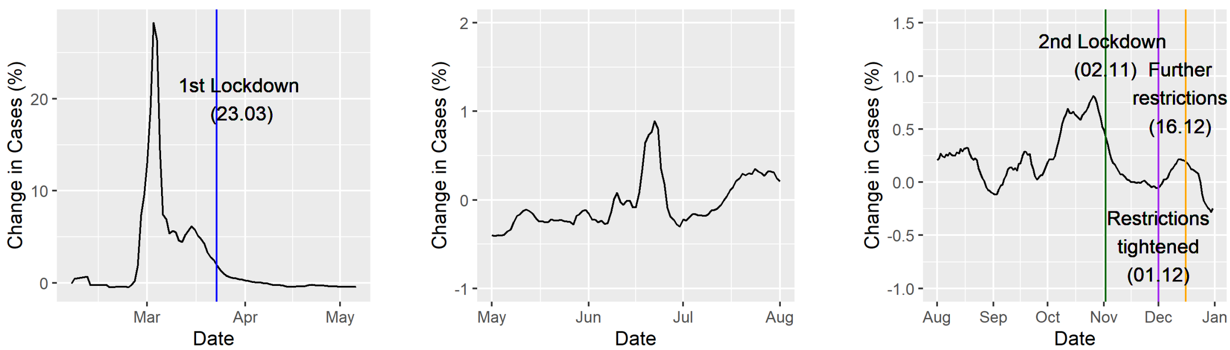
The plots in figure 5.7 attempt to quantify how effective lockdowns were at reducing cases by fitting regression lines around the lockdown-measure related stationary points in case graphs. The linear model is fit once for the week before the peak and once for the week



(a) First Lockdown. (b) Second Lockdown.

Figure 5.7: Plot of Stationary Points Related to Lockdown Measures (DE).

after the peak. In the week before the impact of the first lockdown is seen, rolling case rates per 100,000 increase on average by 0.88 a day. In the week after, incidence rates change on average by -1.20 a day, a change of -236.4%. In contrast, the local peak which emerged after the imposition of the second lockdown sees a far smaller average reduction from 0.71 to 0.14 additional rolling cases a day, or -78.9%. The peak on the 23rd of December associated with the second tightening of restrictions sees the largest change, down from 2.15 more rolling cases a day to -8.57, or -498.6%. The speed at which case figures decrease after the 23rd of December implies that the closing of retail businesses and schools has a more important effect on cases than restrictions on maximum group size, though it is not clear if the entirety of the decrease in case rates comes from government restrictions or if other factors, such as the Christmas period, also helped to drive down cases.



(a) First Wave. (b) Summer. (c) Second Wave.

Figure 5.8: Percentage Difference in Cases Compared to Prior Week (DE).

In figure 5.8, we turn to our second new metric: percentage difference in cases week on week. As the plots show the rate of change of cases, zeroes correspond to stationary points in the case plot. In panel 5.8a we see the acceleration in the rate of change of cases in early March, as cases began spreading exponentially, followed by a decline which begins well before the imposition of lockdown. The rate of change of cases remains between 0 and -0.5% from the 3rd of April until the 17th of June. From September to the 23rd of December, the percentage difference in cases is generally above 0, reflecting the period of rapidly increasing cases. We will return to this metric in section 6.2 to compare mobility with case rates.

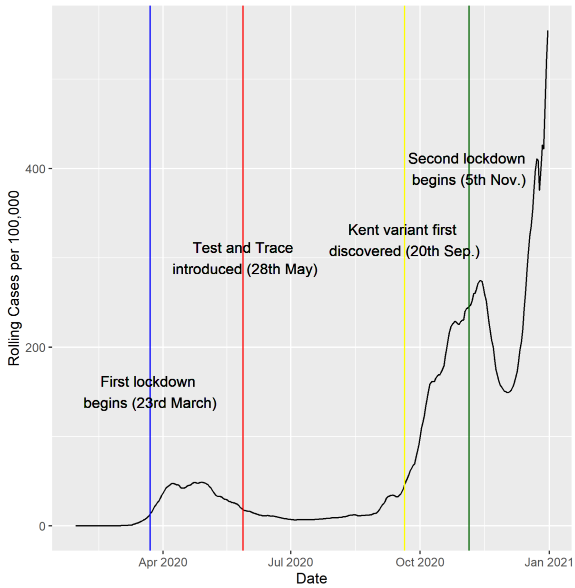
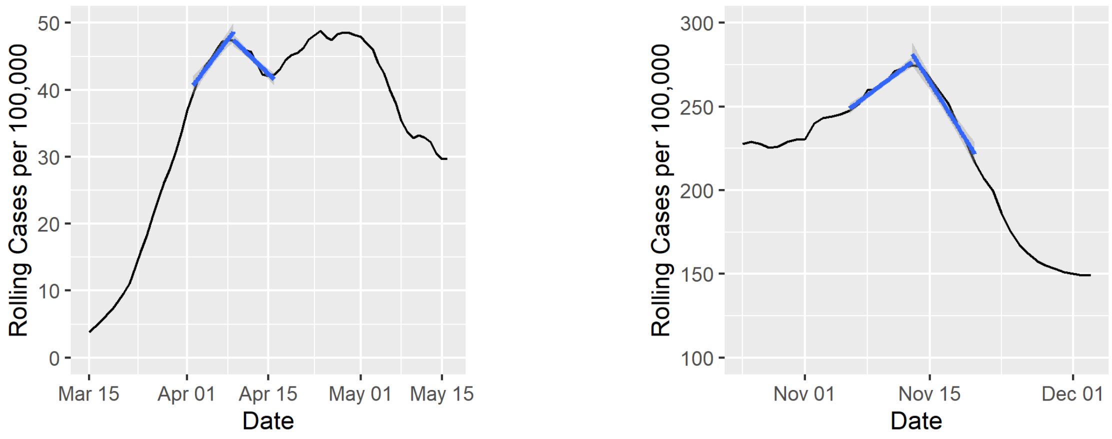


Figure 5.9: Seven Day Population-Adjusted Incidence Rate in England.

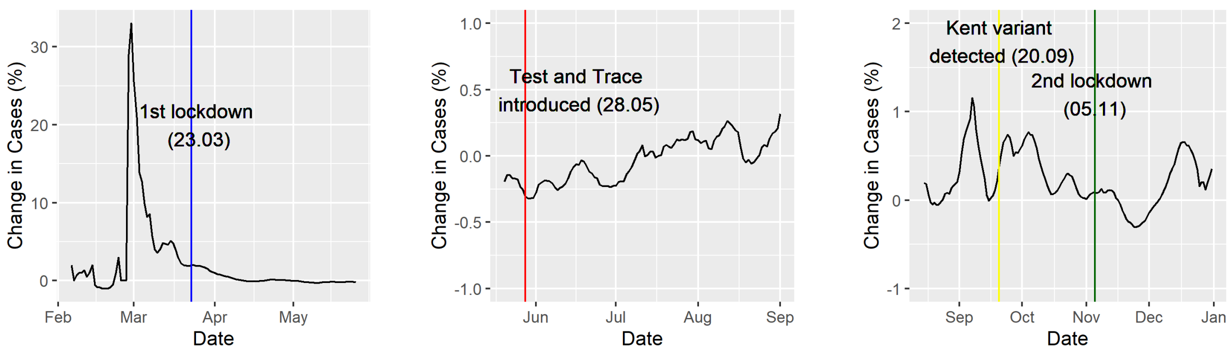
The plot in figure 5.9 of rolling cases rates per 100,000 for England again effectively removes weekly fluctuations. Besides lockdown dates, the plot also includes the date the Kent mutation of COVID-19 was first discovered. It is considered much more infectious than the non-mutated form of the virus and began spreading heavily throughout South East England, contributing to two-thirds of cases in England by December [19]. This may help to explain why COVID-19 cases shot up so quickly in December. The first lockdown led to cases initially peaking on the 9th of April, within 17 days. Unlike in Germany, cases did not decrease continually, but increased again for a few days until the secondary peak on the 29th of April, another 20 days later. In contrast, the impacts of the second national lockdown appear to have taken effect much more quickly, with cases peaking on the 13th of November, 8 days after the lockdown began. Additionally, case incidence sinks quickly until early December, when the Kent variant began spreading rapidly through Southern England and the national lockdown was replaced by the tier system, which set restrictions locally rather than nationally.



(a) First Lockdown. (b) Second Lockdown.

Figure 5.10: Plot of Stationary Points Related to Lockdown Measures (EN).

In the week before the first peak associated with the imposition of the first lockdown, rolling cases per 100,000 were on average increasing by 1.14 daily. The week after saw average case incidence change by -0.84 daily, a change of -173.7%. In contrast, the average case rates around the peak on the 13th of November, associated with the second lockdown, see a larger decrease from 3.99 additional cases to -8.54, a change of -314.0%.



(a) First Wave. (b) Summer. (c) Second Wave.

Figure 5.11: Percentage Difference in Cases Compared to Prior Week (EN).

The plot in figure 5.11 shows the same general trends in percentage difference in cases as the equivalent graph for Germany. Again we see the same increase in cases around the start of March, followed by a normalisation. Here we can see more clearly that the rate of change of cases remains the same for a few days until the effect of lockdown begins to drive cases down. In late April we see the percentage change in cases increase above 0 again, until the second peak on the 29th of April. From this point on, the rate of change of cases remains between 0 and -0.5% until the 9th of July. During summer 2020, there is a positive trend in the rate of change of cases lasting until early September. We do not see a jump in the rate of change of cases associated with the introduction of Test and Trace, due to the fact that increases in testing happened incrementally, rather than all at once. From September to mid November, 8 days after the imposition of the second lockdown, percentage change in cases remains both high and strictly positive, this being the period where England saw the first large rise in cases after the summer. We can see that the imposition of lockdown did reduce cases effectively, however soon after the end of lockdown cases began to spread exponentially again.

Though criticism of the English government's response in the first wave often focuses on the idea that the country was locked down too late [13], both England and Germany began their first national lockdown on the same date: the 23rd of March. In fact, according to the official case figures available, England could be considered to have locked down much more promptly than Germany, with England's 7-day rolling cases per 100,000 value at 11.0 on the 23rd of March, less than half Germany's value of 29.0. Praise for Germany's response often seems to ignore the fact that German authorities were equally slow to impose lockdowns. Instead, articles usually focus on the quick uptake of mass testing and effective communication of lockdown rules [14]. The lack of effective communication may help to explain why cases in England took so long to reduce in April. As late as the 17th of March, the prime minister's official approach to combating the pandemic remained herd immunity, by allowing the virus to spread through the healthy section of the population, rather than keeping cases low [5]. Confusion around the measures may have led to a reduced or slowed-down uptake of strict lockdown restrictions. This hypothesis will be re-examined in Chapter 6. For the second lockdown in November, we have seen that England's approach of a second total lockdown was more effective than Germany's partial lockdown. Where the measures in England contributed to a significant drop in cases for the period of its duration, Germany had to rely on imposing additional restrictions to begin reducing their rates. Overall, lockdowns do seem to be an effective way of reducing the spread of COVID-19.

5.3 Breaking Up Cases by Regional Subdivisions

In this section we would like to determine if there are any noticeable differences in the regions of both countries. We take total population-adjusted cases until the 31st of December 2020 as a metric for the severity of COVID-19 spread in each region, similarly to the approach used in the paper by Velasco et. al [44]. We then compare total cases to a number of possible explanatory factors. These include poverty rate, population density, total regional GDP, and GDP per capita. For Germany we would also like to investigate whether a state being a part of the former German Democratic Republic (GDR) has an effect on total case incidence.

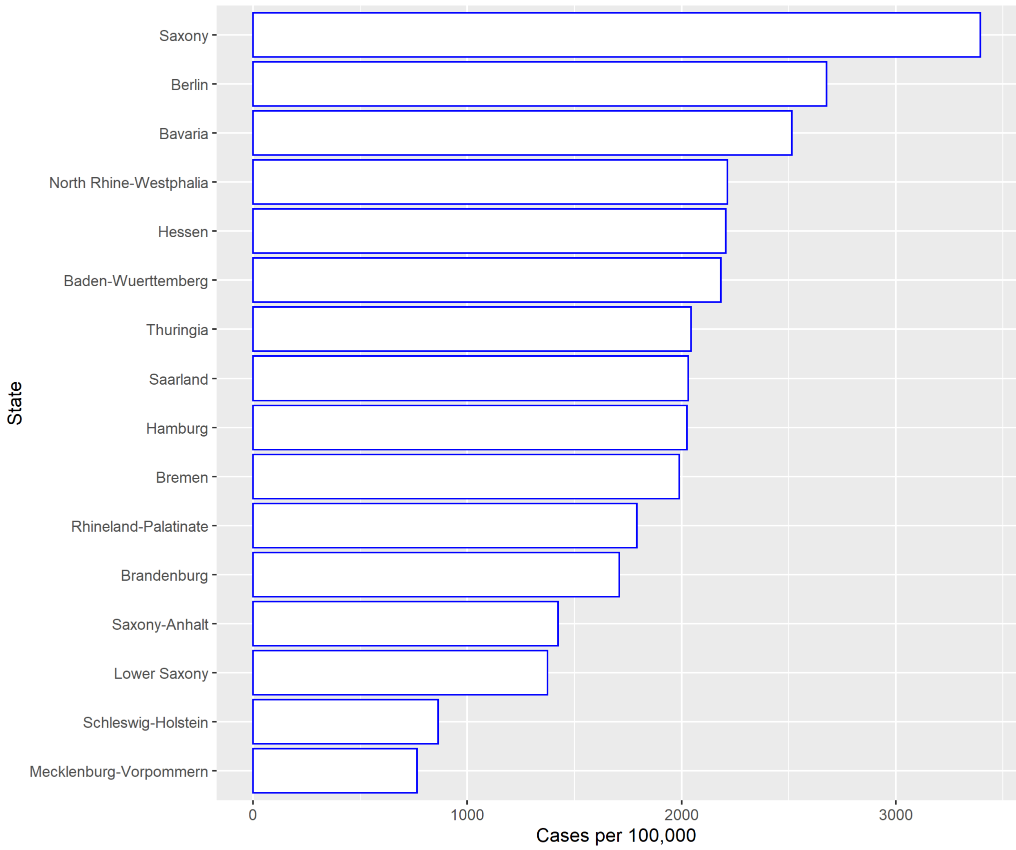


Figure 5.12: Total Cases in Germany by State.

The plot in figure 5.12 gives us an overview of the differences in total case count for each of the 16 federal states. We can see there are large variations in the spread of COVID-19 between states. The worst affected state, Saxony, has more than four times as many cases per capita as Mecklenburg-Vorpommern, 3394 and 764 respectively. To determine whether these differences can be explained by demographic factors, we carry out Pearson Correlation tests. A full table of values used can be found in Appendix A.

Factor	Correlation Coefficient	p-Value
Poverty Rate	-0.0830	0.7599
Population Density	0.4327	0.0941
Regional GDP	0.2981	0.2622
GDP per capita	0.2795	0.2944
Former GDR	-0.0885	0.7444

Table 5.1: Correlations Between Regional COVID-19 Spread and Explanatory Factors (DE).

Table 5.1 shows us that there were no significant correlations between explanatory factors and total cases. Despite this, we can gain some insights by interpreting the signs of the correlation coefficients. As regional GDP and GDP per capita had a positive correlation, this implies that richer states have seen slightly higher COVID-19 rates. A few outliers with very high population density skew the bulk of the data, so the logarithm of population density was used for the correlation calculation. Population density had the greatest correlation, as suggested by the research completed by Roy and Ghosh [32] and by Velasco et. al [44]. This is likely due to the fact that in areas of greater population density, there will be more instances of close contact between people. States in the former GDR have both lower economic output and greater poverty, as well as lower population density, so it makes sense that these states have a lower COVID-19 incidence. Three of these five states: Mecklenburg-Vorpommern, Saxony-Anhalt, and Brandenburg, have among the lowest case rates. Saxony however is an outlier, being the worst affected state, which reduces the correlation coefficient significantly. Some scientists and politicians argue the East's lower incidence may be due to lower disposable incomes in the area, which leads to fewer citizens traveling abroad [4]. As many of the first cases in Germany were imported from holiday destinations in Southern Europe, this may help to explain why case rates for East Germany were much lower in the first few months. Once the lockdown began, it will have been easier for states with lower rates to keep these low. It is also worthwhile mentioning that Schleswig-Holstein, the least economically productive state in West Germany, also has the lowest incidence of West German states.

The plot in 5.13 demonstrates clear regional differences in COVID-19 spread in England. The regions are clustered into four groups: the north, including the North West, North East, and Yorkshire and the Humber, London, the midlands, including the West and East Midlands,

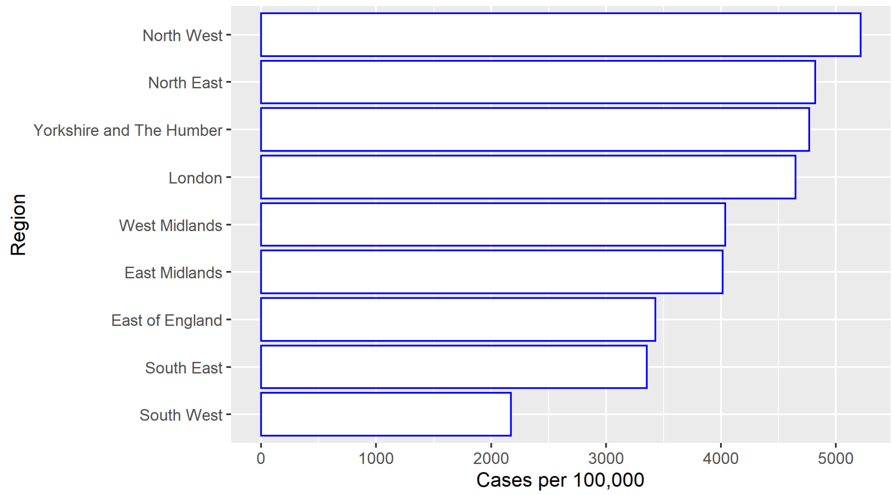


Figure 5.13: Total Cases in England by Region.

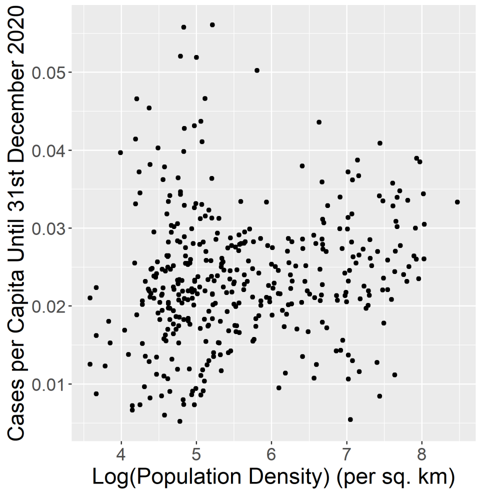
and the south, including East of England, South East and South West. This implies that differences in case incidence by region in England are strongly associated with how far north a region is. Scientists have hypothesised that this is due to greater deprivation in the north, as low-income workers are less likely to be able to work from home and poor high density housing conditions increase the risk of transmission [8].

Factor	Correlation Coefficient	p-Value
Poverty Rate	0.6914	0.0391
Population Density	0.3616	0.339
Regional GDP	0.0478	0.9027
GDP per capita	0.0780	0.842

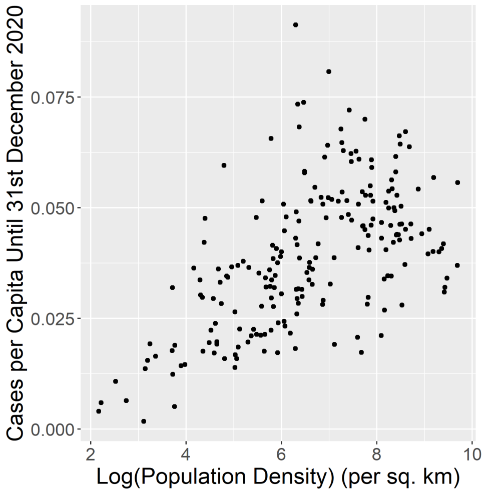
Table 5.2: Correlations Between Regional COVID-19 Spread and Explanatory Factors (EN).

Pearson's correlation test shows that both Regional GDP and GDP per capita have very low correlation with case rates. Higher population density again correlates with higher case rates, though this correlation is not statistically significant. For England we do however find a significant correlation with poverty rate, at a 95% significance level. The correlation coefficient of 0.6914 tells us that there is a moderate association between a higher poverty rate and higher case rates, as discussed above.

It is surprising that while poverty had a significant effect on case rates in England, it had no measurable effect in Germany. This highlights how regional and cultural differences make it difficult to compare COVID-19 rates between countries. Since population density was a key factor highlighted by past research, we will investigate its impact in more detail. We do this by running similar correlation tests for both countries, this time breaking each country up into smaller areas, for England the upper tier authorities and for Germany the administrative districts.



(a) Districts (Germany) [39].



(b) Upper Tier Local Authorities (England) [25].

Figure 5.14: Log of Population Density vs. Total Cases per Capita.

For Germany, we cannot see a clear trend in population density versus case incidence. The correlation coefficient of 0.1505 is only slightly above 0, which shows that higher population density is not strongly associated with greater COVID-19 spread. In England on the other hand, we can see a linear trend emerging between log of population density and cases per capita. The correlation coefficient of 0.5786 shows that population density is a much more important factor in determining how hard-hit by COVID-19 an area is in England than in Germany. Both coefficients are significant at a 95% significance level.

In this section, we were able to determine differences in case rates by area for both countries. The largest differences were geographic, with Eastern Germany less affected by COVID-19 than the West, and Southern England less affected than the North. For England, there is evidence that population density is positively correlated with case incidence. Partially due to the differences in both case incidence and poverty between south and north England, poverty rate was correlated with COVID-19 spread. The method of quantifying COVID-19 spread by total cases per capita is not guaranteed to provide representative results however, as the choice of a different end date to sum by could completely change the correlation values, for example due to differences in spread between the first and second wave.

5.4 Breaking Up Cases by Demographics (Germany)

In this section, we break up case statistics over time by sex and age group, to determine whether different demographic groups face a higher risk of catching COVID-19.

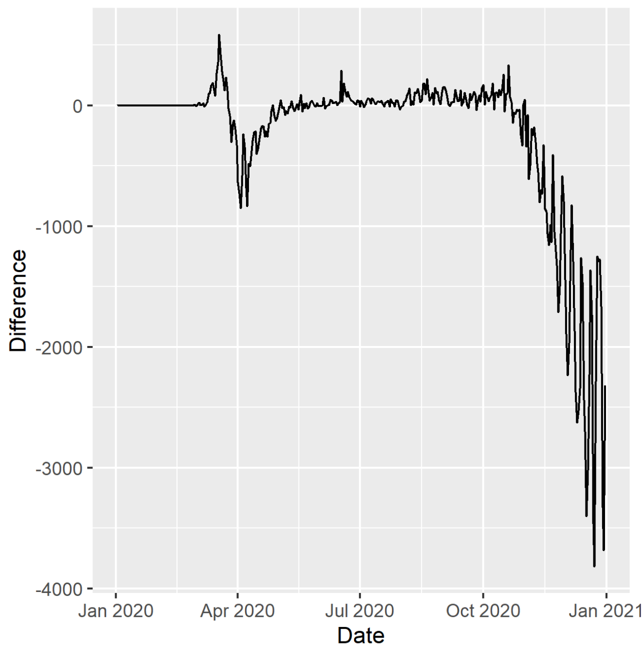


Figure 5.15: Difference in Cases between Sexes (Male - Female) (DE).

The plot above shows the difference in daily case incidence between genders. It is calculated as male cases on a day minus female cases, so a positive value means more men were infected on this date. The metastudy by Pijls et. al [28] found that men were more likely to catch COVID-19 than women. This paper was mainly focused on data from China, so would have highlighted trends which appeared early on in the course of COVID-19 spread in a country. The plot shows that in Germany too, men were more likely to catch COVID-19 in the first weeks. From the 25th of March until the 1st of May however, more women caught the virus than men. During the summer months, until the 23rd of October, men see higher rates of case incidence again. Finally, as the second wave took off, women again faced a much higher risk of infection, with the difference in cases by sex increasing continually as cases rose. Overall, mean daily cases was 2637.7 for women and 2382.5 for men, so women faced a 10.7% higher risk of COVID-19 infection. An unpaired two-sample t-test did not determine that this was a significant difference, $p = 0.39$. The periods where men were more likely to catch COVID-19 were generally periods with lower case incidence and fewer restrictions. A study by Bwire [6] reasoned that men faced higher infection rates due to having a less responsible attitude towards the pandemic. In contrast, during the periods of lockdown women have much higher case rates. This may be due to the fact that women are overrepresented in the healthcare and service sector [10]. Both sectors were deemed essential and continued working during lockdown, meaning workers in these sectors were at higher risk of coming into contact with an infected individual.

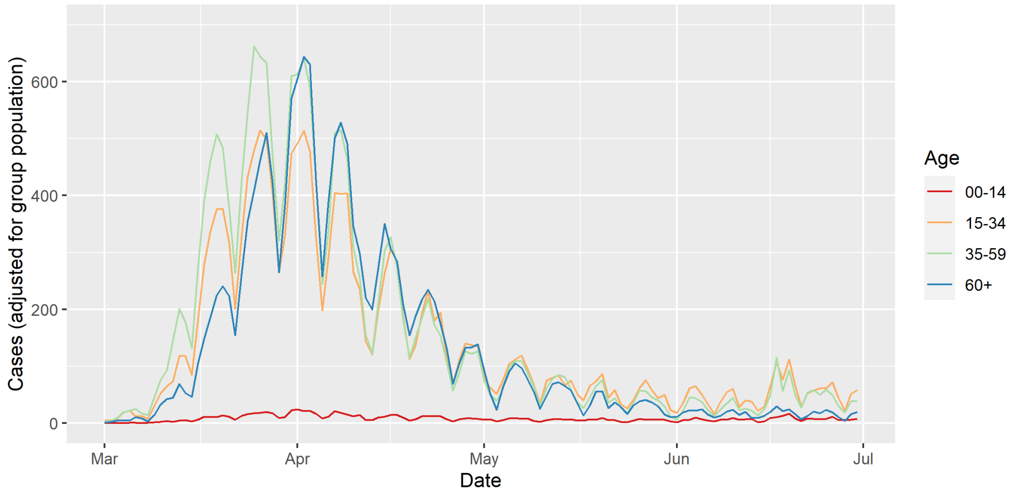


Figure 5.16: Daily Cases in Germany by Age Group, March-June.

The plots in figures 5.16 and 5.17 show COVID-19 cases by age group, adjusted for each group's percentage of the total population [1]. A one-way ANOVA test finds that there are significant differences between age groups at a 95% significance level. One fact that is apparent from both plots is that children, aged 14 and under, had much lower COVID-19 incidence rates than all other age groups, confirmed by a Tukey's Test. This may provide evidence that it is much safer to reopen schools than to remove other restrictions which target older people. The Tukey's Test also informs us that there is no significant difference between incidence rates for 15-34- and 35-59-year-olds. The plot do however show time-lagged differences between the middle-aged groups, 15-34 and 35-59, and the population aged 60 or above, again confirmed by the Tukey's Test. We notice that for most of March, case rates were higher for the middle-aged groups, but by the peak of the first wave rates equalized. From May we see the 60+ age group incidence rates falling faster than rates for middle-aged groups.

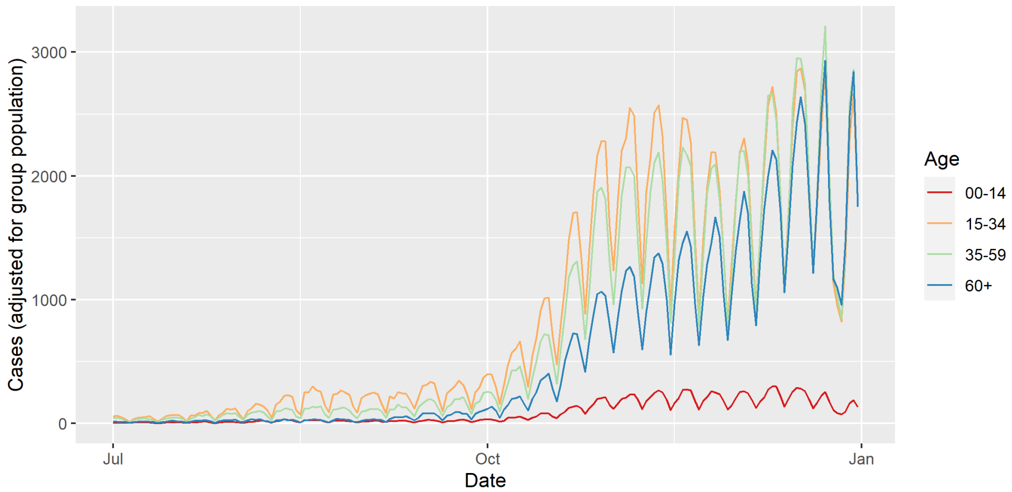


Figure 5.17: Daily Cases in Germany by Age Group, July-December.

For the second half of the year, we can determine a similar trend, with rates first rising for those aged 15-59, before the 60+ age group catches up. As those in the 15-59 age bracket are at lower risk of severe infection or death, this group might be less cautious than the 60+ year old population, many of whom were told to shield for their own protection. This could help to explain the higher rates in younger people during periods where there is no lockdown. In fact, the beginning of the second wave in England was often blamed on young people breaking COVID-19 restrictions, indicating that similar trends may be apparent in both countries [31]. As spread within the younger population grows, they may begin infecting their older relatives, increasing infections in the 60+ group with a lag of a few weeks. As the population of the five Eastern German states skews older than the rest of Germany, the later increase in infections in the population older than 60, helps to explain why these states had lower COVID-19 incidence going in to the first lockdown.

We were able to find differences in case incidence for different demographics in Germany. Though this difference was not significant, women have tested positive around 10% more frequently than men. While children are less likely to catch coronavirus, 15-34- and 35-59-year-olds have similar rates. Though 15-59-year-olds are at lower risk of dying of COVID-19, high infection rates in these ages usually precede increased infections among the elderly.

CHAPTER 6

Investigating Mobility Data

6.1 Analysing the Impact of Restrictions on Mobility

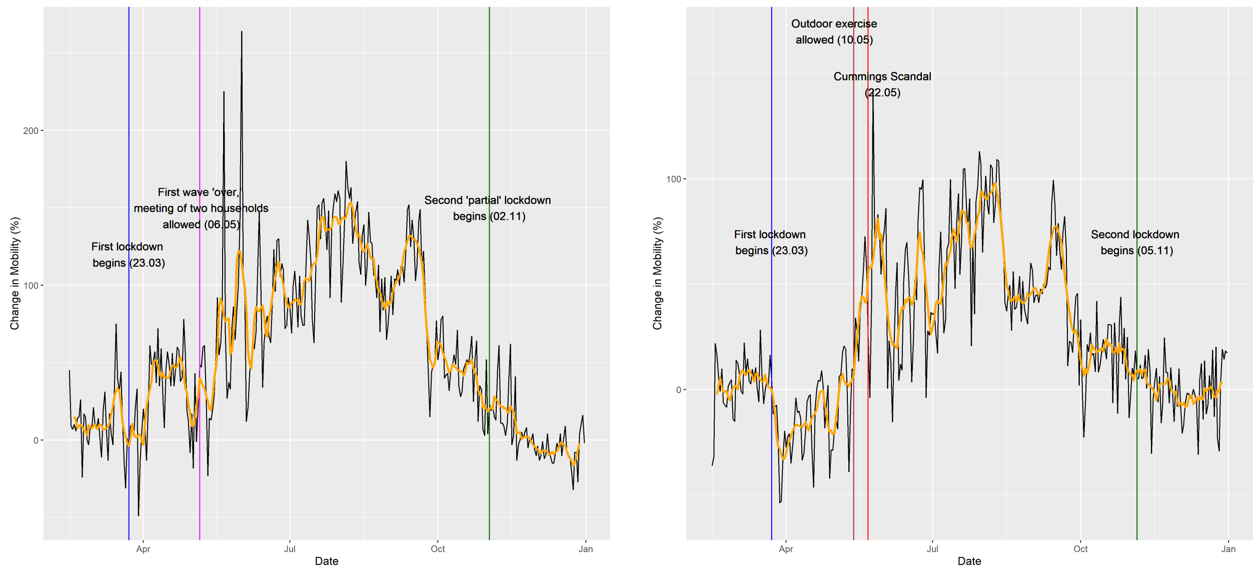
Having investigated case-related data for England and Germany, we will now begin our analysis of mobility data for both countries. As previously mentioned, the Google Community Mobility Reports break changes in mobility up into the 6 subdivisions shown in figure 6.1.

Figure 6.1: Percentage Change in Mobility from Baseline, Germany and England.

The plots show that overall mobility for grocery and pharmacy, retail and recreation, transit stations, and workplaces reduced well below pre-lockdown figures in March. Mobility slowly increases until October/November, but mostly stays under baseline values, before again reducing. We can also see that grocery and pharmacy mobility reduced less drastically, and fluctuated less than other types of mobility. Since groceries and pharmacies are essential businesses, they remained open for the whole year and were similarly frequented throughout 2020. For parks we saw an initial decrease before mobility increases beyond the baseline. As the baseline figures were calculated during winter months, people would have been visiting parks much less than during the summer months. It is possible that if park mobility were compared to values for the past year, mobility changes would be under 0. Residential mobility increases as people spent more time at home. We see that in periods of higher mobility for other groups, residential mobility is lowered and vice versa. The large drop in mobility at the end of December for some types of mobility is associated with the Christmas period, when people generally stayed at home with their family.

Assuming that baseline figures reflect pre COVID-19 mobility accurately, we can determine that, despite indications from higher case rates and a less effective first lockdown in England, people in England actually reduced their mobility more than people in Germany. Multiple t-tests to compare the means of each type of mobility by country, found significant differences in mean mobility change at a 95% significance level. For groceries and pharmacies, English mobility changed on average by -13.1% compared to only -2.1% for Germany. For parks the mean average increase in mobility of 55.8% for Germany was much higher than the 24.3% increase in England. England also saw people staying at home more often, with a 13.4% increase compared to Germany's 6.7%. For retail and recreation mobility England again saw a greater change, -39.6%, versus Germany's -21.6%. Transit stations were also less frequented by the English, with a 41.5% decrease, compared to -26.0% for Germans. Finally, workplaces also saw a greater decrease in mobility in England, -37.5% to Germany's -21.3%.

To look at the specific impact of restrictions and other related events on mobility, we look at some relevant types of mobility again, this time overlaying important dates from our timeline.



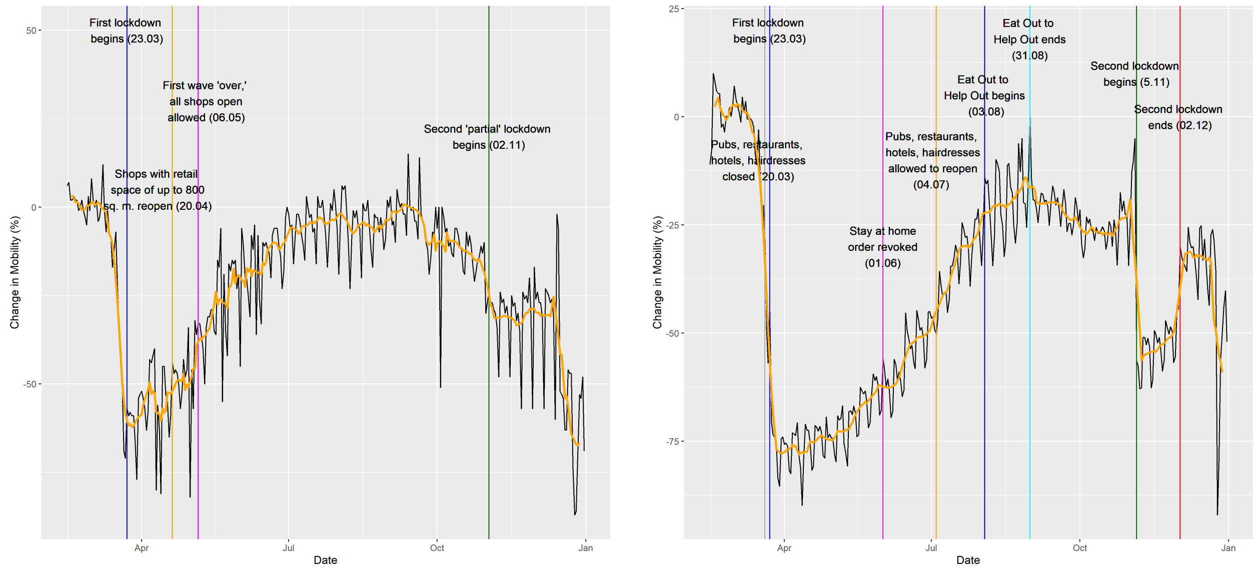
(a) Germany.

(b) England.

Figure 6.2: Change in Parks Mobility From Baseline.

We begin by looking at changes in park mobility in the plots in figure 6.2. The graph shows mobility changes in black, with the orange line representing the rolling average for the past 7 days. For both countries, parks mobility reduces a little around the time of first lockdown. As the temperature increases during the spring and summer, we see parks mobility increasing well above the baseline level. Mobility first peaks around the end of July, with a secondary smaller peak around the end of September. From September, people begin visiting parks much less commonly. Overall, the shapes of both mobility graphs imply that parks mobility is much more associated with climate than government restrictions, with virtually no difference in parks mobility after the imposition of second lockdown in the two countries. In Germany, we can see a small spike in mobility on the 6th of May associated with the first rollback of restrictions. Mobility seems to increase more quickly after this date, implying that people were generally visiting parks less during the first lockdown. The same is true for England, where parks mobility increased rapidly after the 10th of May. For England, the highest daily value of parks mobility occurred on the 25th of May, with 142% greater mobility than the baseline, just three days after newspapers first reported the Cummings scandal. On the 22nd of May it was revealed that Dominic Cummings, a top UK government advisor, had broken lockdown restrictions by travelling from London up to County Durham in March. The revelation was followed by more people reporting they were not following government guidelines [23], implying this peak may be associated with a reaction to the revelations.

However, with Germany's daily peak in mobility of 264% on the 1st of June, only a few days later, the large increase may again be more associated with the weather than with restrictions.

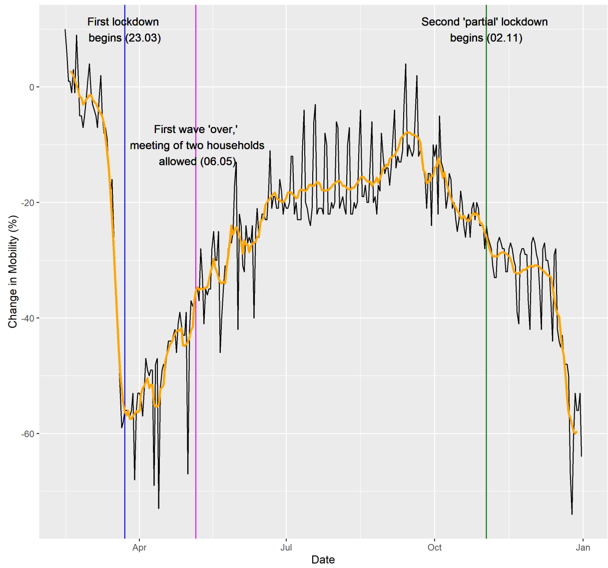


(a) Germany.

(b) England.

Figure 6.3: Change in Retail and Recreation Mobility From Baseline.

For the plot in figure 6.3 we do see reductions in retail and recreation mobility leading up to and continuing for a few days into the beginning of first lockdown. Additionally, there is a very large day to day decrease associated with the introduction of the second lockdown. This reduction is smaller for the partial lockdown in Germany. Again, we see a slow decrease in mobility from the end of September. In Germany, the two reductions in government restrictions on the 20th of April and the 6th of May did precede increases in mobility, though the upward trend began a few days before measures were reduced. Similarly, in England the abolition of the stay-at-home order and reopening of businesses on the 1st of June and the 4th of July respectively did precede increases in mobility, though these were part of a general upwards trend seen since April. On the other hand, the end of the second lockdown did see a large immediate increase in mobility. England instituted the Eat out to Help Out policy from the 3rd of August to the 31st, aiming to get people to eat out at restaurants in an effort to alleviate COVID-19 related economic pressures on the hospitality industry. As we would expect, during this period retail and recreation mobility increased further, with the downward trend in mobility beginning at the end of August. Research has suggested the policy may have inadvertently driven up COVID-19 infections by between 8 and 17% [16]. Changes in retail and recreation mobility seem to be quite strongly associated with changes in government restrictions.



(a) Germany.

(b) England.

Figure 6.4: Change in Transit Station Mobility From Baseline.

We would also like to look at transit station mobility, as we would expect time spent at transit stations to reflect the overall level of movement in a country. The plot in figure 6.4 highlights the same trends seen in retail and recreation mobility, with strong drops around the impositions of lockdowns, an upward trend in the summer months, and a gradual reduction from the end of September. Again, we see that England's second lockdown was more effective in reducing mobility than Germany's partial lockdown, though mobility returned to similarly high levels again as the national lockdown was replaced with the tier system. While transit station mobility continued increasing after the first removal of restrictions in Germany on the 6th of May, there was actually a short-term drop in mobility after similar restrictions were abolished on the 1st of June in England. Additionally, allowing the use of public transport for non-emergencies seems to have had no effect on the frequency with which transit stations were visited in England.

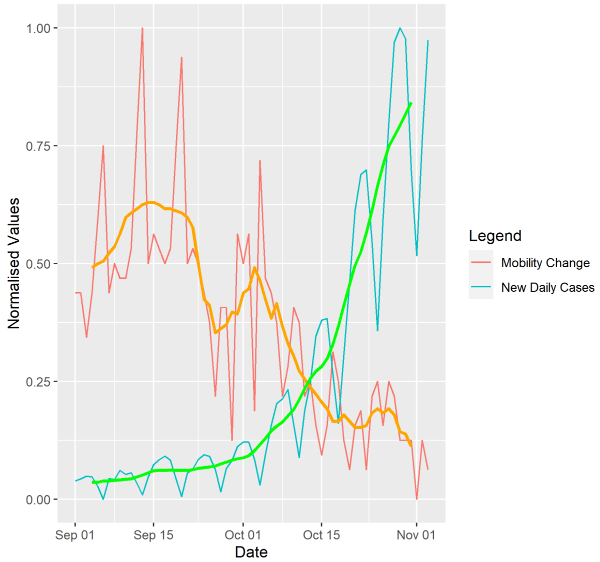
Having looked at the effects of certain restrictions and events on mobility, we can say that lockdowns are the best way to reduce mobility. This is especially true for retail and recreation mobility and transit station mobility compared to park mobility, which seems to be more associated with the climate. Though we were able to explain most of the trends in mobility, there were no additional restrictions associated with the decrease in mobility in September. One possible hypothesis we will look at in the next section, is that this reduction in mobility was a voluntary response to climbing case figures around the time.

6.2 Analysing Interactions Between Mobility and Cases

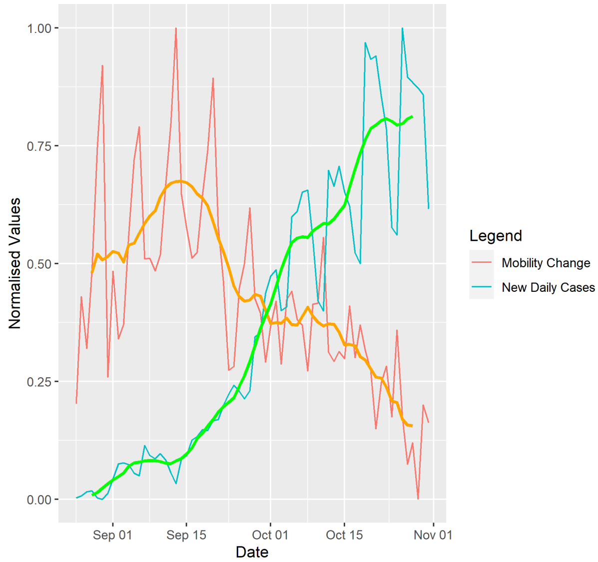
6.2.1 Rising Case Numbers and Voluntary Changes in Mobility

We would like to determine if people's mobility behaviour changes as a result of increasing case numbers. Unfortunately, as the mobility data only contains values from the 15th of February, we can not analyse whether first cases in each country affected mobility. We can however look at the pre-lockdown mobility decrease as in the study by Mendolia, Statrunova, and Yerokhin. We look at the average mobility change over the last three days before first lockdown as a compromise between wanting to eliminate daily fluctuations and representing the large decrease in mobility seen in the last few days before lockdown. The four types of mobility we would expect to decrease are grocery and pharmacy, retail and recreation, transit stations, and workplaces. Grocery and pharmacy mobility changed the least from the baseline in the days before lockdown, with mobility changing by -17% in Germany and increase by 9% in England. The increase in mobility in England may be due to panic buying before the lockdown. Transit station, and retail and recreation mobility reduced greatly for both countries. In Germany transit station mobility decreased by -56.3% and retail and recreation mobility dropped by -63.3%, well below the -50% mark highlighted by the research completed by Mendolia, Statrunova, and Yerokhin [24]. For England the reduction was slightly lower, -40.7% for transit stations and -44.3% for retail and recreation. This difference in pre-lockdown mobility may be a key reason why cases took longer to peak in England than Germany. The decrease in workplace mobility is comparable for both countries: -30.3% for Germany and -28% for England. Over the three-day period residential mobility increased by 13% in Germany and 10.6% in England, again compared to the baseline value. Again we see a larger change for Germany.

To analyse specifically the impact of rising case numbers on mobility, we will look at the period in which mobility drops before the second lockdown in both countries. As no changes in national COVID-19 regulations occurred during this time, changes in mobility were based entirely on voluntary changes in behaviour. We will look at transit station mobility specifically.



(a) Germany.



(b) England.

Figure 6.5: Change in Transit Station Mobility Versus Change in Daily Cases.

To show both changes in mobility and changes in cases in one plot, the values were normalised between 0 and 1. Date ranges were chosen to incorporate the beginning of the increase in cases and drop in mobility, but not the beginning of second lockdown. For both countries, mobility drops as cases begin to increase again after the low figures throughout summer. We use Pearson's Correlation Tests to determine the impact of increasing cases for each country. We will compare mobility to lagged case values, performing one test with a lag of one day, and one with a lag of one week to see how quickly people responded to increases in cases. For Germany, correlation was -0.65 with a one-day lag and -0.64 with a seven-day lag. This implies that past cases did have a moderately negative relationship with changes in mobility. For England, the correlation coefficients were -0.72 and -0.76 for a one- and seven-day lag respectively. Mobility in England was therefore even more strongly associated with increases in cases, especially so for changes in cases from the prior week. The p-values of all four tests were significant at a 95% significance level. Though we cannot be sure that changes in mobility occurred specifically as a result of rising case figures, the fact that no new national restrictions were imposed during this time does lend some credibility to this hypothesis.

6.2.2 Impact of Changes in Mobility on Case Rates

In this section we are interested in how changes in mobility precede changes in case rates. We have already seen that large drops in mobility, associated with the implementation of a lockdown, have a negative effect on case rates. To build a model, we first use cross-correlation functions to determine the optimal lag between each type of mobility and cases. We use only the data from April 1st so we do not include the large pre-lockdown fluctuations in mobility. To remove weekly fluctuations, we again use percentage difference in cases week on week as a metric for changes in spread. To remove weekly fluctuations in mobility data we use the difference in mobility compared to 7 days prior. All cross-correlation function plots can be found in Appendix B.

Comparing optimal cross-correlation values, we find that differences in cases lag behind changes in transit station mobility by 8 days in Germany and 9 days in England. For retail mobility, the lag is 6 days in Germany and 11 days in England. Continuing on to residential mobility we see the lag is 9 days for both countries. For groceries and pharmacies, the lag was 9 days for Germany and 11 days in England. The lag between workplace mobility changes and changes in cases was 7 days for both countries. For parks mobility, spikes in cross-correlation were less clear. This is not surprising as parks mobility seems to follow a mostly seasonal trend and is less affected by restrictions on mobility. For England the best lag for parks mobility was over 20 days, while the best estimate for Germany was 9 days. Overall, the estimates of lag correspond relatively well to similar estimates made in a research paper by Badr et. al [3]. The paper found that changes in mobility are only perceptible after 9-12 days, slightly longer than some of the lower lag values we found.

Having determined the optimal lag values to use, we fit a linear model to understand how well we can predict case figures from lagged mobility values. The linear model uses a combination of all mobility types but parks, lagged by their optimal values, to predict changes in spread of cases.

Looking at Tables 6.1 and 6.2, we can see that the linear model is not very effective at predicting changes in cases. For the model fit on German data, the intercept is the only significant value, while the model for the English data finds that retail and recreation and grocery and pharmacy mobility are significant factors. However, the model predicts that retail and recreation mobility has a positive relationship with cases, which we would not

Coefficient	Estimate	p-Value
Intercept	0.0776	0.000020
Retail and Recreation	-0.0002	0.8759
Residential	0.0152	0.1931
Workplaces	-0.0001	0.9218
Transit Stations	-0.0056	0.0597
Grocery and Pharmacy	-0.0006	0.7084

Table 6.1: Table of Linear Model Coefficients (Germany).

Coefficient	Estimate	p-Value
Intercept	0.1026	0.00000004
Retail and Recreation	0.0099	0.000621
Residential	0.0186	0.435892
Workplaces	-0.0031	0.318084
Transit Stations	-0.0013	0.868753
Grocery and Pharmacy	-0.0099	0.040778

Table 6.2: Table of Linear Model Coefficients (England).

expect to be the case. In fact, the multiple R-squared for both models was very low, 0.044 for Germany and 0.054 for England, meaning only 4-5% percent of the variation in cases is explained by the models. In contrast, a model simply based on the previous daily percentage difference in cases in England was able to explain over 96% of variation.

Though we were able to determine likely estimates for the lag between changes in mobility and changes in cases, fitting a linear model to predict changes in cases proved less successful. It may be the case that small changes in overall mobility have little measurable effect on case rates, especially as the mobility data does not reflect other important factors in the success of social distancing policies, such as proximity to others, or percentage of people wearing face masks.

CHAPTER 7

Conclusion

Through our analysis of COVID-19 case data and the Google Mobility Reports dataset, we have been able to come to a few conclusions about COVID-19 Cases in England and Germany. Most importantly, we saw that lockdowns have the greatest effect on reducing both case rates and mobility. On the other hand, during periods with fewer restrictions, mobility increased. Both countries also saw geographic differences in the severity of the spread of COVID-19, for England a North-South divide and for Germany a West-East divide. Areas with higher population density were generally more hard-hit in both countries. In Germany, men and women faced different incidence rates at different periods of the pandemic. While older people were less often infected at the beginning of the pandemic, differences between age groups evened out later on. Children aged 14 and under were the only group who saw continually lower COVID-19 rates. Mobility was responsive to both government restrictions and increases in case rates.

Though we focused on case and mobility data, the analysis suggests that it would be worthwhile extending the research to other quantifiable factors. Testing and death data for example could be used to attempt to estimate how many people were actually infected in the English first wave. Both cases and mobility showed seasonal trends, so it would also be interesting to look at climate data and compare it to case spread.

Though the end of the pandemic in Europe finally seems within reach, the results of this investigation remain highly relevant to policy makers. As England and Germany continue with their vaccination programs, it will be important to keep case rates low, to avoid additional mutations. Additionally, as of March 2021, Germany is experiencing the beginning of a third wave, fuelled by case rises in the young population. It remains to be seen whether the government will avoid making the same mistakes it made in the second wave, which would lead to further deaths among the elderly.

Appendix A

Tables of Regional Explanatory Variables for COVID-19 Cases

Appendix A contains full lists of the explanatory factors used in the correlation calculations in section 5.4, as well as the total case rate, for the geographic subdivisions in each country.

Table for English Regions

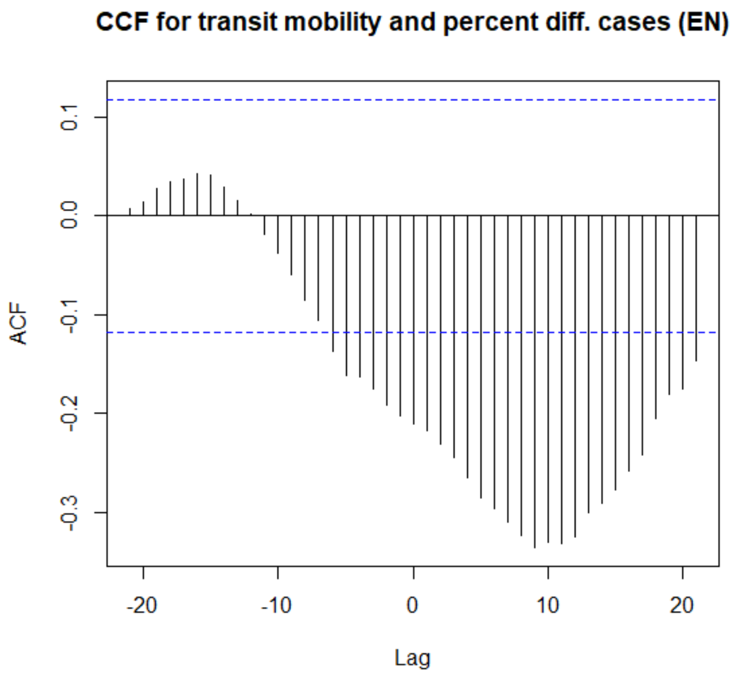
Region	Total Cases (per 1,000)	Poverty Rate [17]	Population Density (per km^2) [26]	Gross Domestic Product (Billion Pounds) [15]	GDP per capita (Pounds)
East Midlands	40.13	0.21	310	124.6	25770
East of England	34.29	0.20	326	186.5	29907
London	46.5	0.28	5701	487.1	54358
North East	48.2	0.24	311	62.6	23454
North West	52.1	0.23	520	207.5	28266
South East	33.5	0.19	481	311.3	33911
South West	21.7	0.19	236	158.1	28112
West Midlands	40.4	0.24	457	159.8	26930
Yorkshire and The Humber	47.7	0.22	357	141.7	25754

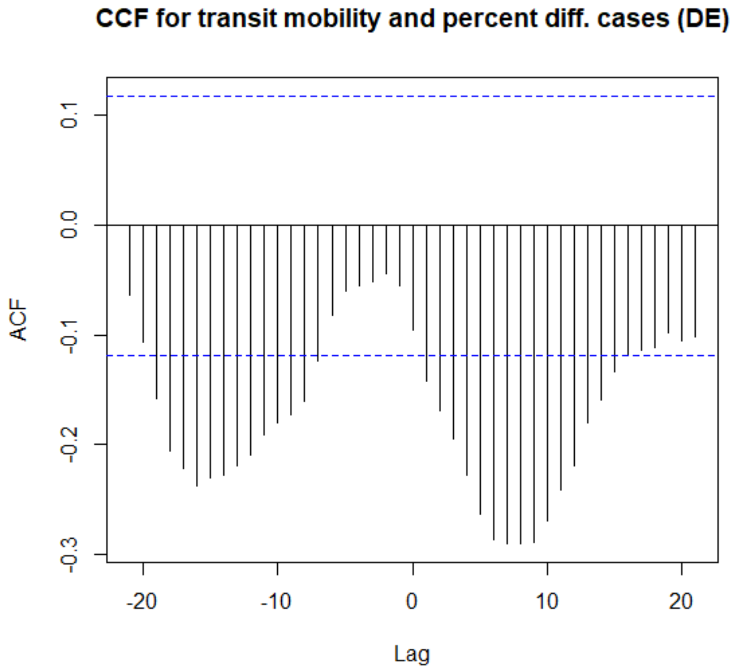
Table for German States						
State	Total Cases (per 1,000)	Poverty Rate [27]	Population Density (per km^2) [47]	Gross Domestic Product (Billion Euro) [38]	GDP per capita (Euro)	Ex-East Germany
Baden-Wuerttemberg	21.82	0.123	293.9	524.3	47234	0
Bavaria	25.15	0.119	175.7	632.9	48221	0
Berlin	26.76	0.193	3692.9	152.3	41755	0
Brandenburg	17.10	0.152	83.3	74.3	29114	1
Bremen	19.90	0.249	1552.5	33.6	48908	0
Hamburg	20.25	0.150	2260.1	123.3	66757	0
Hessen	22.07	0.161	282.8	294.4	46819	0
Mecklenburg-Vorpommern	7.65	0.194	69.4	46.6	28980	1
Lower Saxony	13.73	0.171	163.3	307.0	38404	0
North Rhine-Westphalia	22.14	0.185	514.5	711.4	39639	0
Rhineland-Palatinate	17.92	0.156	201.0	145.0	35418	0
Saarland	20.32	0.170	389.2	36.2	36677	0
Saxony	33.94	0.172	220.2	128.1	31459	1
Saxony-Anhalt	14.23	0.195	111.8	63.5	28929	1
Schleswig-Holstein	8.65	0.145	177.2	97.8	33678	0
Thuringia	20.44	0.170	135.3	63.9	29958	1

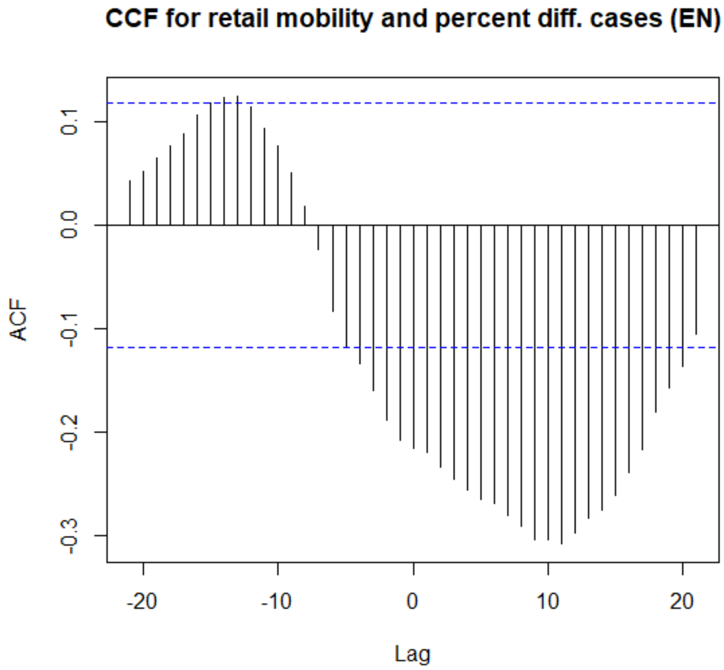
Appendix B

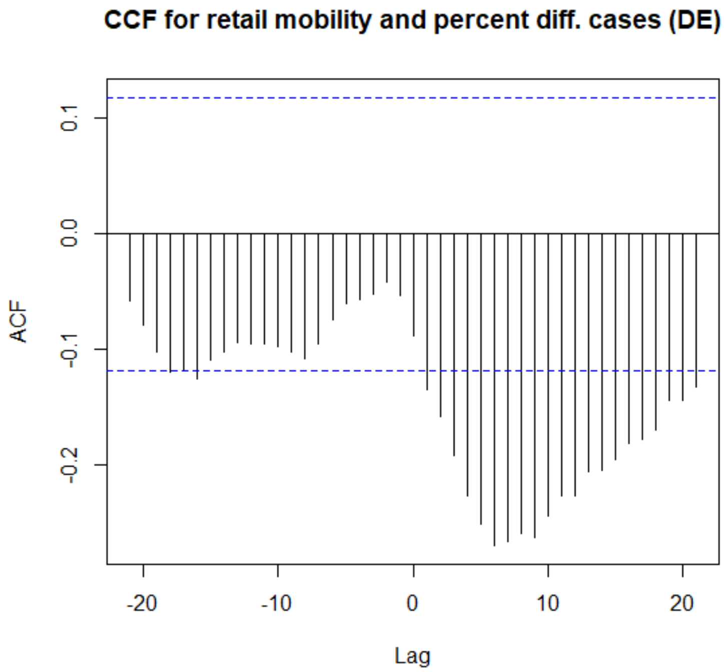
CCF Plots Mobility and Cases

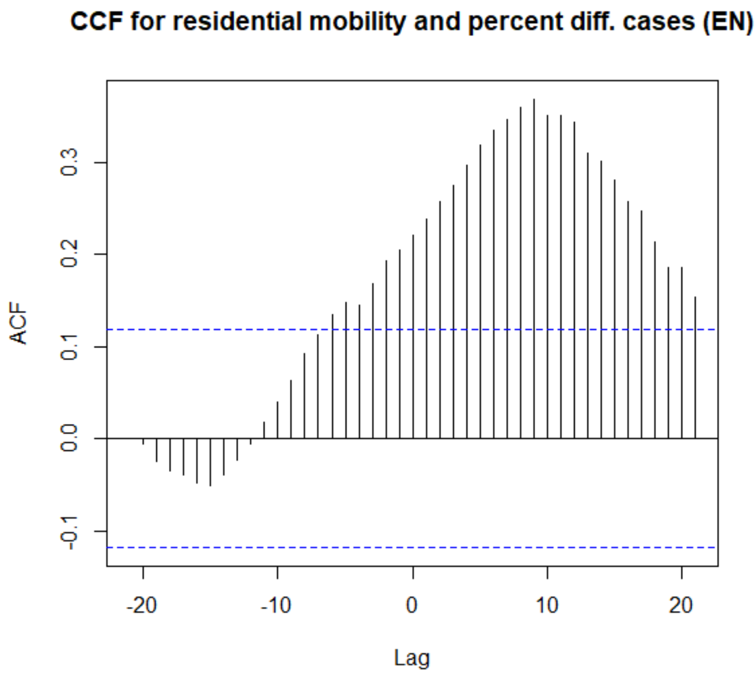
Appendix B consists of cross-correlation function plots for each type of mobility compared to the change in case rates. We can determine the lag with the auto-correlation by looking at the highest spike in each graph. As we are looking at how changes in mobility precede changes in cases we are only interested in values for positive lags. We can see that the ACF is usually negative, except for residential mobility. For parks mobility in England, the spike occurs with a lag of more than 20 days, well out of the range we would expect. For Germany, we see much greater spikes in ACF values for negative values.

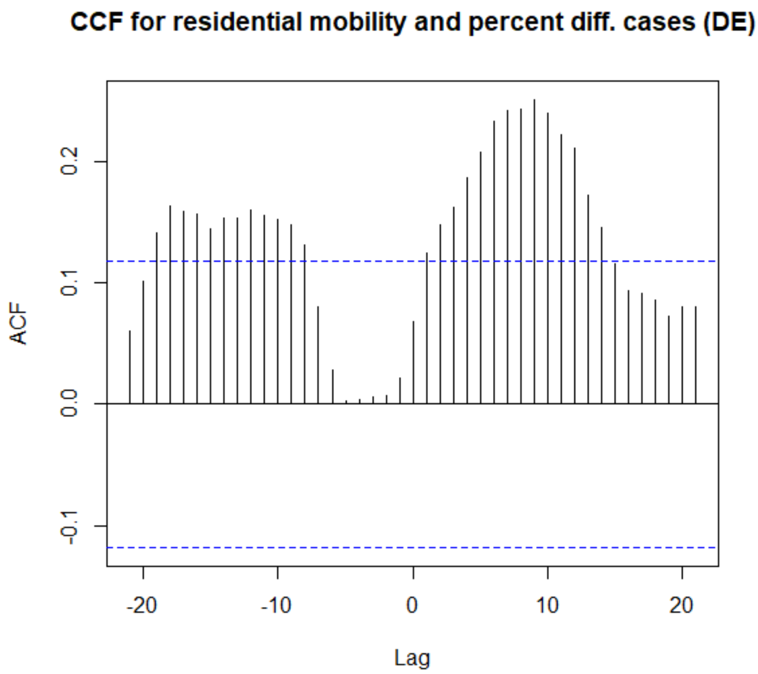


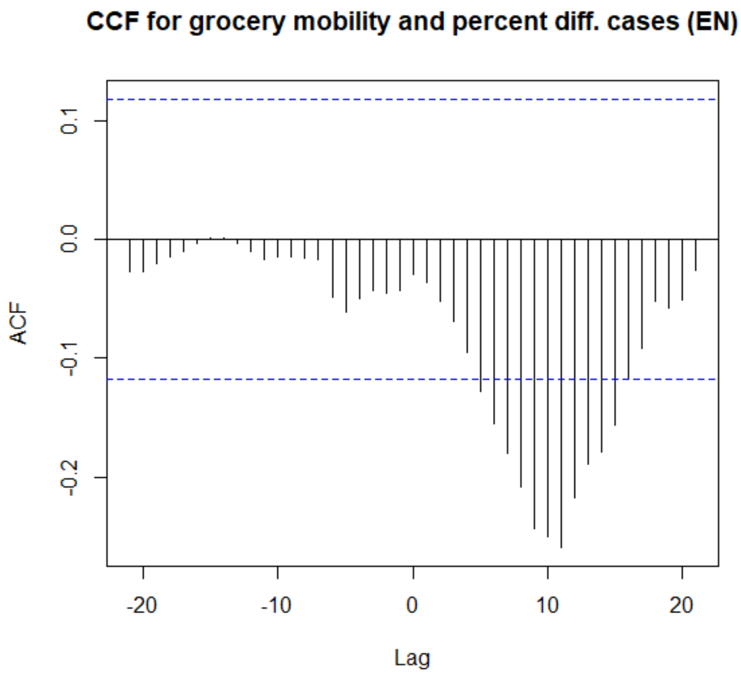

CCF for transit mobility and percent diff. cases (DE)
ACF
0.1
0.0
-0.1
-0.2
-0.3
-20
-10
0
10
20
Lag

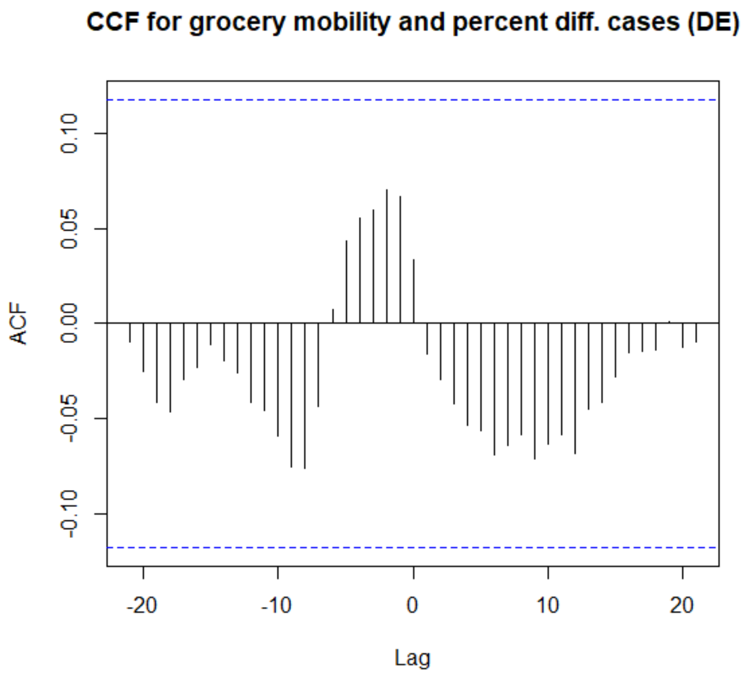

CCF for retail mobility and percent diff. cases (EN)
ACF
Lag
0.1
0.0
-0.1
-0.2
-0.3
-20
-10
0
10
20

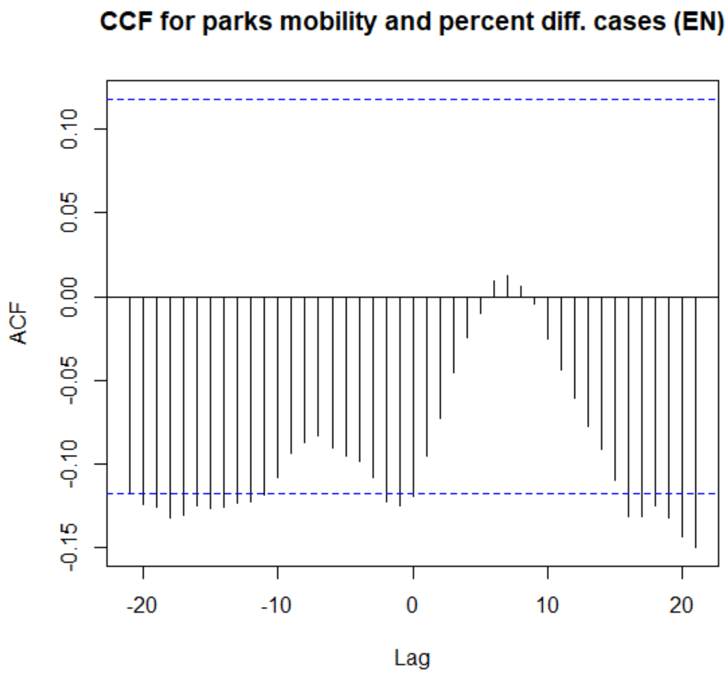

CCF for retail mobility and percent diff. cases (DE)
ACF
Lag
0.1
0.0
-0.1
-0.2
-20
-10
0
10
20

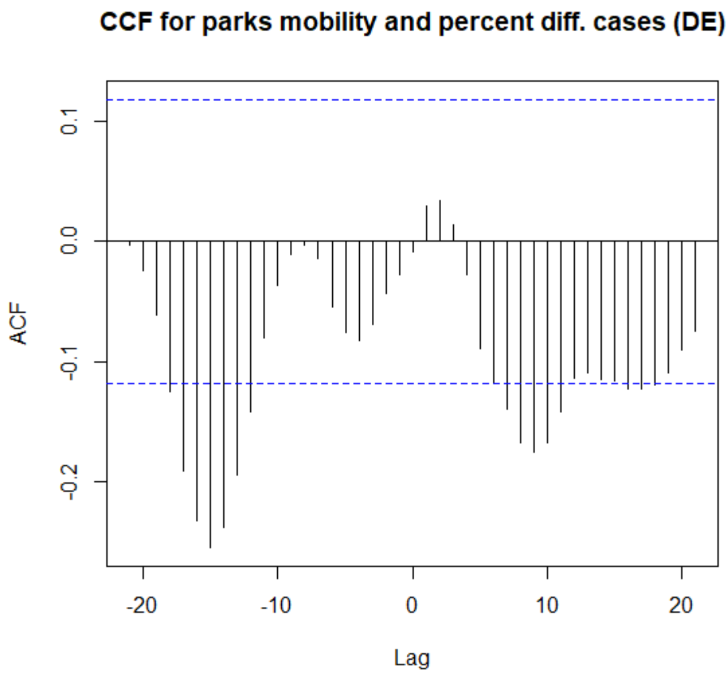

CCF for residential mobility and percent diff. cases (EN)
ACF
0.3
0.2
0.1
0.0
-0.1
-20
-10
0
10
20
Lag

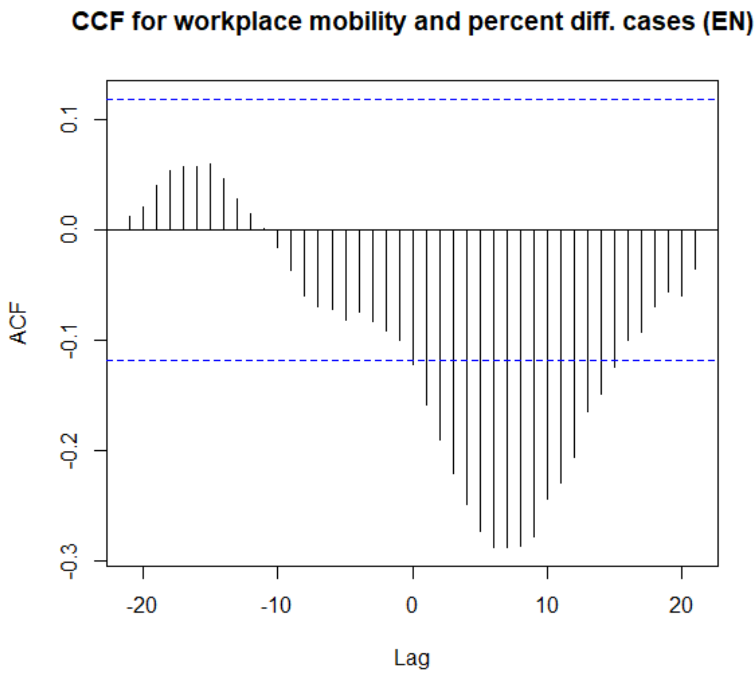

CCF for residential mobility and percent diff. cases (DE)
ACF
0.2
0.1
0.0
-0.1
-20
-10
0
10
20
Lag

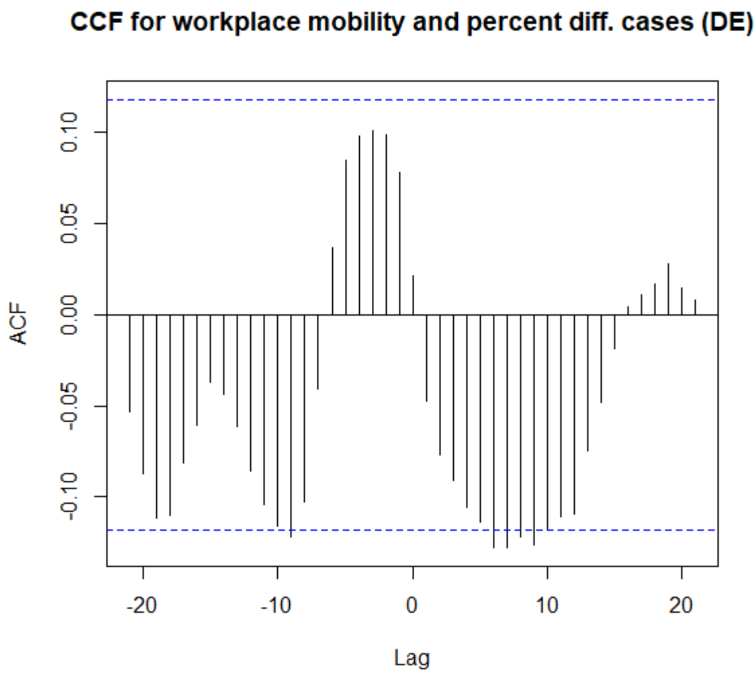

CCF for grocery mobility and percent diff. cases (EN)
ACF
0.1
0.0
-0.1
-0.2
-20
-10
0
10
20
Lag


CCF for grocery mobility and percent diff. cases (DE)
ACF
0.10
0.05
0.00
-0.05
-0.10
-20
-10
0
10
20
Lag


CCF for parks mobility and percent diff. cases (EN)
ACF
0.10
0.05
0.00
-0.05
-0.10
-0.15
-20
-10
0
10
20
Lag


CCF for parks mobility and percent diff. cases (DE)
ACF
0.1
0.0
-0.1
-0.2
-20
-10
0
10
20
Lag


CCF for workplace mobility and percent diff. cases (EN)
ACF
0.1
0.0
-0.1
-0.2
-0.3
-20
-10
0
10
20
Lag


CCF for workplace mobility and percent diff. cases (DE)
ACF
0.10
0.05
0.00
-0.05
-0.10
-20
-10
0
10
20
Lag

APPENDIX C

R Code

```
library(dplyr)
library(tidyr)
library(lubridate)
library(ggplot2)
library(zoo)
library(RColorBrewer)
library(readxl)


setwd("WD")

# Section 5.1, 5.2: Case Summaries

# Work with cases data set for Germany:

# additional data by state (Germany)
population_thousands_de <- c(11100, 13125, 3669, 2552, 687, 1847, 6288, 1608, 7994, 17947, 4094, 987, 4072, 2195, 2904, 2133)
poverty_rate_de <- c(0.123, 0.119, 0.193, 0.152, 0.249, 0.150, 0.161, 0.194, 0.171, 0.185, 0.156, 0.170, 0.172, 0.195, 0.145, 0.170)
ex_gdr_de <- c(0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1)
pop_density_de <- c(293.9, 175.7, 3692.9, 83.3, 1552.5, 2260.1, 282.8, 69.4, 163.3, 514.5, 201, 389.2, 220.2, 111.8, 177.2, 135.3)
city_de <- c(0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
gdp_billions_de <- c(524.3, 632.9, 153.2, 74.3, 33.6, 123.3, 294.4, 46.6, 307.0, 711.4, 145.0, 36.2, 128.1, 63.5, 97.8, 63.9)
population_de <- sum(population_thousands_de)


# select relevant columns, rename to English, convert to date datatype and arrange in order of date and state
rki <- as_tibble(read.csv("RKI_COVID19.csv")) %>%
  select(Bundesland, Altersgruppe, Geschlecht, AnzahlFall, AnzahlTodesfall, Meldedatum) %>%
  rename('state'='Bundesland', 'age_group'='Altersgruppe', 'sex'='Geschlecht', 'cases'='AnzahlFall', 'deaths'='AnzahlTodesfall', 'date
      '='Meldedatum') %>%
  mutate(date = as.Date(date)) %>%
  filter(date <= as.Date("2020-12-31")) %>%
  arrange(date, state) %>%
  mutate(cases_per_thousand = cases / population_de)
rki$state<- recode(rki$state, 'Baden-WÃ¼rttemberg' = 'Baden-Wuerttemberg', 'Bayern' = 'Bavaria', 'Niedersachsen' = 'Lower Saxony', '
    Nordrhein-Westfalen' = 'North Rhine-Westphalia',
                  'Rheinland-Pfalz' = 'Rhineland-Palatinate', 'Sachsen' = 'Saxony', 'Sachsen-Anhalt' = 'Saxony-Anhalt', 'ThÃ¼ringen'
                      = 'Thuringia')
```

```r
# summary statistics for country
cases_de <- sum(rki$cases)
deaths_de <- sum(rki$deaths)
cases_de_per_thousand <- cases_de / population_de

# daily cases for country
daily_cases_de <- group_by(rki, date) %>%
  summarise(sum_cases = sum(cases)) %>%
  mutate(sum_cases_pop = sum_cases / population_de * 100) %>%
  mutate(rolling_cases_week = rollapplyr(sum_cases, width = 7, FUN = sum, partial = TRUE)) %>%
  mutate(rolling_cases_week_pop = rollapplyr(sum_cases_pop, width = 7, FUN = sum, partial = TRUE)) %>%
  mutate(cumulative_cases = cumsum(sum_cases)) %>%
  mutate(percentage_difference_cases = (rolling_cases_week - lag(rolling_cases_week, 7)) / lag(rolling_cases_week, 7)) %>%
  mutate(country = "DE")

max(daily_cases_de$sum_cases)

# plots

# sum_cases

fig1DE <- ggplot(daily_cases_de, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases")

ggsave("fig1DE.png", scale=0.7)

# cumulative_cases

fig2DE <- ggplot(daily_cases_de, aes(x=date, y=cumulative_cases)) + geom_line() +
  xlab("Date") + ylab("Total Cases")

ggsave("fig2DE.png", scale=0.7)

# sum cases breakdown

fig3DE <- ggplot(daily_cases_de, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-02-06", "2020-05-06"))) + ylim(0, 7000)

ggsave("fig3DE.png", scale = 0.5)

fig4DE <- ggplot(daily_cases_de, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-05-01", "2020-08-01"))) + ylim(0, 1500)

ggsave("fig4DE.png", scale = 0.5)

fig5DE <- ggplot(daily_cases_de, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-08-01", "2020-12-31")))

ggsave("fig5DE.png", scale = 0.5)


# rolling_cases_week_pop

fig6DE <- ggplot(daily_cases_de, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-12-01")), color="purple") + # Restrictions tightened
  geom_vline(aes(xintercept=as.Date("2020-12-16")), color="orange") + # Further restrictions
  geom_text(aes(label="First lockdown \nbegins (23rd March)", x=as.Date("2020-03-23"), y=100)) +
  geom_text(aes(label="Second 'partial'\nlockdown begins\n(2nd Nov.)", x=as.Date("2020-11-02"), y=200)) +
  geom_text(aes(label="Lockdown restrictions\ntightened (1st Dec.)", x=as.Date("2020-11-22"), y=80)) +
  geom_text(aes(label="Further restrictions\n(16th Dec.)"), x=as.Date("2020-11-28"), y=50)
```

```
ggsave("fig6DE.png", scale=1.1)

fig6aDE <- ggplot(daily_cases_de, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  scale_x_date(limits = as.Date(c("2020-03-10", "2020-05-01"))) + ylim(0, 50) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-03-27"), to=as.Date("2020-04-03"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-04-03"), to=as.Date("2020-04-10"), by=1)), method="lm"
      )

ggsave("fig6aDE.png", scale=0.5)

summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-03-27"), to=as.Date("2020-04-03"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-04-03"), to=as.Date("2020-04-10"), by=1)), rolling_cases_
    week_pop ~ date))

fig6bDE <- ggplot(daily_cases_de, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  scale_x_date(limits = as.Date(c("2020-11-01", "2020-12-31"))) + ylim(125, 220) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-11-07"), to=as.Date("2020-11-14"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-11-14"), to=as.Date("2020-11-21"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-12-16"), to=as.Date("2020-12-23"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-12-23"), to=as.Date("2020-12-30"), by=1)), method="lm"
      )

ggsave("fig6bDE.png", scale=0.5)

summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-11-07"), to=as.Date("2020-11-14"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-11-14"), to=as.Date("2020-11-21"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-12-16"), to=as.Date("2020-12-23"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_de %>% filter(date %in% seq(from=as.Date("2020-12-23"), to=as.Date("2020-12-30"), by=1)), rolling_cases_
    week_pop ~ date))

# percentage_difference_cases

# zoom in first few months, massive spike, rest

fig7DE <- ggplot(daily_cases_de, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-02-06", "2020-05-06"))) +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_text(aes(label="1st Lockdown\n (23.03)", x=as.Date("2020-03-30"), y=20))


ggsave("fig7DE.png", scale = 0.5)

fig8DE <- ggplot(daily_cases_de, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-05-01", "2020-08-01"))) + ylim(-1, 2)

ggsave("fig8DE.png", scale = 0.5)

fig9DE <- ggplot(daily_cases_de, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-08-01", "2020-12-31"))) + ylim(-1, 1.5) +
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-12-01")), color="purple") + # Restrictions tightened
  geom_vline(aes(xintercept=as.Date("2020-12-16")), color="orange") + # Further restrictions
```

```
  geom_text(aes(label="2nd Lockdown \n(02.11)", x=as.Date("2020-11-02"), y=1.2)) +
  geom_text(aes(label="Restrictions\ntightened\n(01.12)", x=as.Date("2020-12-01"), y=-0.6)) +
  geom_text(aes(label="Further\nrestrictions\n(16.12)", x=as.Date("2020-12-13"), y=0.8))

ggsave("fig9DE.png", scale = 0.5)




# UK data:

# additional data by region
pop_thousands_en <- c(4835, 6236, 8961, 2669, 7341, 9180, 5624, 5934, 5502)
poverty_en <- c(0.21, 0.20, 0.28, 0.24, 0.23, 0.19, 0.19, 0.24, 0.22)
pop_density_en <- c(310, 326, 5701, 311, 520, 481, 236, 457, 357)
gdp_billions_en <- c(124.6, 186.5, 487.1, 62.6, 207.5, 311.3, 158.1, 159.8, 141.7)
pop_en <- sum(pop_thousands_en)

# read uk data, arrange by date, remove NA from data
en <- as_tibble(read.csv("region_2021-01-18.csv")) %>%
  mutate(date = as.Date(date)) %>%
  filter(date <= as.Date("2020-12-31")) %>%
  arrange(date) %>%
  arrange(areaName) %>%
  rename('cases' = 'newCasesBySpecimenDate') %>%
  mutate(cases_per_thousand = cases / pop_en) %>%
  replace_na(list(cases = 0, deaths = 0))


# summary statistics for country
cases_en <- sum(en$cases)
cases_en_per_thousand <- cases_en / pop_en


# daily cases for country
daily_cases_en <- group_by(en, date) %>%
  summarise(sum_cases = sum(cases)) %>%
  mutate(sum_cases_pop = sum_cases / pop_en * 100) %>%
  mutate(rolling_cases_week = rollapplyr(sum_cases, width = 7, FUN = sum, partial = TRUE)) %>%
  mutate(rolling_cases_week_pop = rollapplyr(sum_cases_pop, width = 7, FUN = sum, partial = TRUE)) %>%
  mutate(cumulative_cases = cumsum(sum_cases)) %>%
  mutate(percentage_difference_cases = (rolling_cases_week - lag(rolling_cases_week, 7)) / lag(rolling_cases_week, 7)) %>%
  replace(is.na(.), 0) %>%
  mutate(country = "EN")

daily_cases_en$percentage_difference_cases <- recode(daily_cases_en$percentage_difference_cases, "Inf" = 0)

max(daily_cases_en$sum_cases)


# plots for england

# sum_cases

fig1EN <- ggplot(daily_cases_en, aes(x=date, y=sum_cases)) + geom_line() + xlab("Date") + ylab("Cases") +
  geom_vline(aes(xintercept=as.Date("2020-05-28")), color="red") +
  geom_text(aes(label="Test and Trace\n introduced (28th May)", x=as.Date("2020-05-28"), y=40000))

ggsave("fig1EN.png", scale=0.7)

# cumulative_cases

fig2EN <- ggplot(daily_cases_en, aes(x=date, y=cumulative_cases)) + geom_line() +
  xlab("Date") + ylab("Total Cases") + geom_vline(aes(xintercept=as.Date("2020-05-28")), color="red") +
  geom_text(aes(label="Test and Trace\n introduced (28th May)", x=as.Date("2020-05-28"), y=1000000))
```

```
ggsave("fig2EN.png", scale=0.7)

# sum_cases zoom in

fig3EN <- ggplot(daily_cases_en, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-02-06", "2020-05-25"))) + ylim(0, 5000)

ggsave("fig3EN.png", scale = 0.5)

fig4EN <- ggplot(daily_cases_en, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-05-20", "2020-09-01"))) + ylim(0, 3200) +
  geom_vline(aes(xintercept=as.Date("2020-05-28")), color="red") +
  geom_text(aes(label="Test and Trace\n introduced", x=as.Date("2020-06-12"), y=3000))

ggsave("fig4EN.png", scale = 0.5)

fig5EN <- ggplot(daily_cases_en, aes(x=date, y=sum_cases)) + geom_line() +
  xlab("Date") + ylab("Cases") +
  scale_x_date(limits = as.Date(c("2020-08-15", "2020-12-31")))

ggsave("fig5EN.png", scale = 0.5)


# rolling_cases_week_pop

fig6EN <- ggplot(daily_cases_en, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  geom_vline(aes(xintercept=as.Date("2020-05-28")), color="red") +
  geom_text(aes(label="Test and Trace\n introduced (28th May)", x=as.Date("2020-05-28"), y=300)) +
  geom_vline(aes(xintercept=as.Date("2020-09-20")), color="yellow") + # Kent variant
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  geom_text(aes(label="First lockdown \nbegins (23rd March)", x=as.Date("2020-03-23"), y=150)) +
  geom_text(aes(label="Second lockdown \nbegins (5th Nov.)", x=as.Date("2020-11-05"), y=400)) +
  geom_text(aes(label="Kent variant first \ndiscovered (20th Sep.)", x=as.Date("2020-09-20"), y=320))

ggsave("fig6EN.png", scale=1.1)

fig6aEN <- ggplot(daily_cases_en, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  scale_x_date(limits = as.Date(c("2020-03-15", "2020-05-16"))) + ylim(0, 50) +
  geom_smooth(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-04-02"), to=as.Date("2020-04-09"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-04-09"), to=as.Date("2020-04-16"), by=1)), method="lm"
      )

ggsave("fig6aEN.png", scale=0.5)

summary(lm(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-04-02"), to=as.Date("2020-04-09"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-04-09"), to=as.Date("2020-04-16"), by=1)), rolling_cases_
    week_pop ~ date))

fig6bEN <- ggplot(daily_cases_en, aes(x=date, y=rolling_cases_week_pop)) + geom_line() +
  xlab("Date") + ylab("Rolling Cases per 100,000") +
  scale_x_date(limits = as.Date(c("2020-10-25", "2020-12-03"))) + ylim(100, 300) +
  geom_smooth(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-11-06"), to=as.Date("2020-11-13"), by=1)), method="lm"
      ) +
  geom_smooth(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-11-13"), to=as.Date("2020-11-20"), by=1)), method="lm"
      )

ggsave("fig6bEN.png", scale=0.5)
```

```
summary(lm(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-11-06"), to=as.Date("2020-11-13"), by=1)), rolling_cases_
    week_pop ~ date))
summary(lm(data = daily_cases_en %>% filter(date %in% seq(from=as.Date("2020-11-13"), to=as.Date("2020-11-20"), by=1)), rolling_cases_
    week_pop ~ date))


# percentage_difference_cases

# zoom in first few months, massive spike, rest

fig7EN <- ggplot(daily_cases_en, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-02-06", "2020-05-25"))) +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_text(aes(label="1st lockdown \n(23.03)", x=as.Date("2020-03-26"), y=20))

ggsave("fig7EN.png", scale = 0.5)

fig8EN <- ggplot(daily_cases_en, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-05-20", "2020-09-01"))) + ylim(-1, 1) + geom_vline(aes(xintercept=as.Date("2020-05-28")),
      color="red") +
  geom_text(aes(label="Test and Trace\n introduced (28.05)", x=as.Date("2020-06-16"), y=0.5))

ggsave("fig8EN.png", scale = 0.5)

fig9EN <- ggplot(daily_cases_en, aes(x=date, y=percentage_difference_cases)) + geom_line() +
  xlab("Date") + ylab("Change in Cases (%)") +
  scale_x_date(limits = as.Date(c("2020-08-15", "2020-12-31"))) + ylim(-1, 2) +
  geom_vline(aes(xintercept=as.Date("2020-09-20")), color="yellow") + # Kent variant
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  geom_text(aes(label="2nd lockdown \n(05.11)", x=as.Date("2020-11-05"), y=1.2)) +
  geom_text(aes(label="Kent variant\n detected (20.09)", x=as.Date("2020-09-20"), y=1.8))

ggsave("fig9EN.png", scale = 0.5)



# Compare sum_pop_cases by country
daily_cases_both <- full_join(daily_cases_de, daily_cases_en)
fig10 <- ggplot(daily_cases_both, aes(x=date, y=sum_cases_pop, group=country, color=country)) + geom_line() +
  xlab("Date") + ylab("Cases per 100,000") + theme(text = element_text(size = 16))

ggsave("fig10.png", width = 5.98*2)




# Section 5.3: Regional differences

# tibble containing all demographic data for germany

state_rki <- group_by(rki, state)
states_de <- cbind(summarise(state_rki, cases=sum(cases), deaths=sum(deaths)), population_thousands_de, poverty_rate_de, ex_gdr_de,
    pop_density_de, city_de, gdp_billions_de) %>%
  mutate(cases_per_capita = cases / population_thousands_de / 1000) %>%
  mutate(cases_per_thousand = cases / population_thousands_de) %>%
  mutate(gdp_per_capita = gdp_billions_de / population_thousands_de*1000000)

# correlation: between each factor and cases_per_capita, lm graphs for better fits
cor.test(states_de$cases_per_capita, states_de$poverty_rate_de)
cor.test(states_de$cases_per_capita, log(states_de$pop_density_de))
cor.test(states_de$cases_per_capita, states_de$gdp_billions_de)
cor.test(states_de$cases_per_capita, states_de$gdp_per_capita)
cor.test(states_de$cases_per_capita, states_de$ex_gdr_de)
```

```
t.test(states_de$cases_per_capita ~ states_de$ex_gdr_de)
t.test(states_de$cases_per_capita ~ states_de$city_de)

fig11DE <- ggplot(states_de, aes(x=reorder(state, cases_per_thousand), y=cases_per_thousand*100)) + geom_bar(stat="identity", fill="
      white", color="blue") + coord_flip() +
  ylab("Cases per 100,000") + xlab("State")

ggsave("fig11DE.png", scale=0.4, width=21)


# tibble containing all demographic data for england

state_en <- group_by(en, areaName)
regions_en <- cbind(summarise(state_en, cases=sum(cases)), pop_thousands_en, poverty_en, pop_density_en, gdp_billions_en) %>%
  mutate(cases_per_thousand = cases / pop_thousands_en) %>%
  mutate(cases_per_capita = cases / pop_thousands_en / 1000) %>%
  mutate(gdp_per_capita = gdp_billions_en / pop_thousands_en*1000000)

# correlation: between each factor and cases_per_capita, lm graphs for better fits
cor.test(regions_en$cases_per_capita, regions_en$poverty_en)
cor.test(regions_en$cases_per_capita, log(regions_en$pop_density_en))
cor.test(regions_en$cases_per_capita, regions_en$gdp_billions_en)
cor.test(regions_en$cases_per_capita, regions_en$gdp_per_capita, method="spearman")

fig11EN <- ggplot(regions_en, aes(x=reorder(areaName, cases_per_thousand), y=cases_per_thousand*100)) +
  geom_bar(stat="identity", fill="white", color="blue") + coord_flip() + xlab("Region") + ylab("Cases per 100,000")

ggsave("fig11EN.png", scale=0.7, width=10)



# Lower-level population density

# population density by district

rki2 <- read.csv("RKI_COVID19.csv", encoding="UTF-8") %>%
  group_by(Landkreis) %>%
  summarize(cases=sum(AnzahlFall))

kreise_de <- read_excel("04-kreise.xlsx", sheet="Kreisfreie Städte u. Landkreise") %>%
  tail(-7) %>%
  select("...3", "...5", "...6", "...9") %>%
  rename('Landkreis' = '...3', 'Area (sq km)' = '...5', 'Population' = '...6', 'Pop_Density' = '...9') %>%
  na.omit() %>%
  filter('Landkreis' != 'Insgesamt')

pop_dens_de <- left_join(rki2, kreise_de, by='Landkreis') %>%
  na.omit()
pop_dens_de$`Area (sq km)` <- as.numeric(pop_dens_de$`Area (sq km)`)
pop_dens_de$Population <- as.numeric(pop_dens_de$Population)
pop_dens_de$`Pop_Density` <- as.numeric(pop_dens_de$`Pop_Density`)

pop_dens_de <- mutate(pop_dens_de, cases_pop = cases/Population)

fig12DE <- ggplot(pop_dens_de, aes(x=log(Pop_Density), y=cases_pop)) + geom_point() +
  xlab("Log(Population Density) (per sq. km)") + ylab("Cases per Capita Until 31st December 2020") +
  theme(text = element_text(size = 18))


ggsave("fig12DE.png")

cor.test(log(pop_dens_de$Pop_Density), pop_dens_de$cases_pop)


# Population density by local area
```

```
uk_utla <- read.csv("utla_2021-02-13.csv") %>%
  mutate(date = as.Date(date)) %>%
  filter(date < as.Date("2021-01-01")) %>%
  group_by(areaName) %>%
  summarize(cases=sum(newCasesBySpecimenDate)) %>%
  rename("Name" = "areaName")

uk_pop_dens_utla <- read_excel('ukmidyearestimates20192020ladcodes.xls', sheet="MYE 5") %>%
  tail(-3)
colnames(uk_pop_dens_utla) <- uk_pop_dens_utla[1,]

uk_pop_dens_utla <- uk_pop_dens_utla[-1, ] %>%
  select("Name", "Area (sq km)", "Estimated Population mid-2019") %>%
  na.omit()

uk_pop_dens_utla[427, ] <- c("Hackney and City of London", 19+3, 281120+9721)
uk_pop_dens_utla[428, ] <- c("Cornwall and Isles of Scilly", 3545+16, 569578+2224)

uk_pop_dens_utla$"Estimated Population mid-2019" <- as.numeric(uk_pop_dens_utla$"Estimated Population mid-2019")
uk_pop_dens_utla$"Area (sq km)" <- as.numeric(uk_pop_dens_utla$"Area (sq km)")

uk_pop_dens_utla$populationDensity <- uk_pop_dens_utla$`Estimated Population mid-2019`/uk_pop_dens_utla$`Area (sq km)`

uk_utla_full <- left_join(uk_utla, uk_pop_dens_utla, by="Name")

uk_utla_full$casesPop <- uk_utla_full$cases / uk_utla_full$`Estimated Population mid-2019`

fig12EN <- ggplot(uk_utla_full, aes(x=log(populationDensity), y=casesPop)) + geom_point() +
  xlab("Log(Population Density) (per sq. km)") + ylab("Cases per Capita Until 31st December 2020") +
  theme(text = element_text(size = 18))

ggsave("fig12EN.png")

cor.test(uk_utla_full$casesPop, log(uk_utla_full$populationDensity))



# Section 5.4: Cases by demographics

# daily cases by sex
daily_sex_cases_de <- group_by(rki, date, sex) %>%
  summarise(sum_cases = sum(cases)) %>%
  spread(sex, sum_cases) %>%
  rename('F' = 'W', 'unknown' = 'unbekannt') %>%
  replace(is.na(.), 0) %>%
  mutate(m_greater = M - F) %>%
  gather('M', 'F', 'unknown', 'm_greater', key='sex', value='sum_cases')

unknown <- daily_sex_cases_de %>%
  filter(sex=="unknown")
sum(unknown$sum_cases) / sum(daily_sex_cases_de$sum_cases) # what percentage of cases have unknown sex

t_test_sex <- daily_sex_cases_de %>%
  filter(sex=="F" | sex=="M")
t.test(t_test_sex$sum_cases~t_test_sex$sex, alternative = "two.sided", var.equal=FALSE)

(2637.683-2382.513)/2382.513


age_regrouped <- rki
age_regrouped$age_group <- recode(rki$age_group, "A00-A04" = "Child", "A05-A14" = "Child", "A15-A34" = "Young", "A35-A59" = "Middle",
     "A60-A79" = "Old", "A80+" = "Old")

daily_age_cases_de <- group_by(age_regrouped, date, age_group) %>%
  summarise(sum_cases = sum(cases)) %>%
```

```
  spread(age_group, sum_cases) %>%
  mutate(Child = Child/ 100 * 12.8) %>%
  mutate(Young = Young/ 100 * 34.2) %>%
  mutate(Middle = Middle/ 100 * 24.3) %>%
  mutate(Old = Old/100 * 28.7) %>%
  select(-unbekannt) %>%
  gather("Child", "Young", "Middle", "Old", key="Age", value="sum_cases") %>%
  replace(is.na(.), 0)

daily_age_cases_de$Age <- recode(daily_age_cases_de$Age, "Child" = "00-14", "Young" = "15-34", "Middle" = "35-59", "Old" = "60+")

aov_age <- aov(sum_cases ~ Age, data=daily_age_cases_de)
summary(aov_age)
TukeyHSD(aov_age)

# daily_sex_cases_de, m_greater

fig13 <- daily_sex_cases_de %>%
  filter(sex=="m_greater") %>%
  ggplot(aes(x=date, y=sum_cases)) + geom_line() + xlab("Date") + ylab("Difference")

ggsave("fig13.png", scale=0.8)

# daily_age_cases_de


fig14a <- ggplot(daily_age_cases_de, aes(x=date, y=sum_cases, group=Age, color=Age)) + geom_line() +
  xlab("Date") + ylab("Cases␣(adjusted␣for␣group␣population)") +
  scale_x_date(limits = as.Date(c("2020-03-01", "2020-06-30"))) + ylim(0, 700) +
  scale_color_brewer(palette="Spectral")

ggsave("fig14a.png", width=11.4, scale=0.7)

fig14b <- ggplot(daily_age_cases_de, aes(x=date, y=sum_cases, group=Age, color=Age)) + geom_line() +
  xlab("Date") + ylab("Cases␣(adjusted␣for␣group␣population)") +
  scale_x_date(limits = as.Date(c("2020-07-01", "2020-12-31"))) +
  scale_color_brewer(palette="Spectral")

ggsave("fig14b.png", width=11.4, scale=0.7)






# Chapter 6: Mobility data


# UK mobility

mobility_uk <- read.csv('2020_GB_Region_Mobility_Report.csv') %>%
  mutate(date = as.Date(date)) %>%
  filter(sub_region_2 == "") %>%
  rename('retail_and_recreation' = 'retail_and_recreation_percent_change_from_baseline', 'grocery_and_pharmacy' = 'grocery_and_
       pharmacy_percent_change_from_baseline',
         'parks' = 'parks_percent_change_from_baseline', 'transit_stations' = 'transit_stations_percent_change_from_baseline',
         'workplaces' = 'workplaces_percent_change_from_baseline', 'residential' = 'residential_percent_change_from_baseline') %>%
  gather('retail_and_recreation', 'grocery_and_pharmacy', 'parks', 'transit_stations', 'workplaces', 'residential',
         key='type_mobility', value='percent_change_from_baseline') %>%
  filter(date < as.Date("2021-01-01"))

mobility_uk$percent_change_from_baseline <- na.locf(mobility_uk$percent_change_from_baseline)


# Constituent counties and populations of regions
```

```
summarise_regions <- function(region){
  ret <- mobility_uk %>%
    filter(sub_region_1 %in% row.names(region) & sub_region_2 == "") %>%
    mutate(percent_change_from_baseline = percent_change_from_baseline * (region[as.character(sub_region_1),] / sum(region))) %>%
    group_by(date, type_mobility) %>%
    summarise(percent_change_from_baseline = sum(percent_change_from_baseline))
  return(ret)
}

east_midlands_counties <- c('Derby', 'Leicester', 'Nottingham', 'Lincolnshire', 'Northamptonshire', 'Derbyshire', 'Nottinghamshire', '
    Leicestershire')
# excluded Rutland as mostly NAs, also only 30,000 population
east_midlands_pop <- c(270000, 508000, 768000, 1087000, 747000, 1053000-270000, 1154000-768000, 1053000-508000)
east_midlands <- as.data.frame(east_midlands_pop, row.names=east_midlands_counties)
colnames(east_midlands) <- c("Population")

em_mob <- summarise_regions(east_midlands) %>% mutate(Region = "east_midlands")


east_of_england_counties <- c('Bedford', 'Central Bedfordshire', 'Cambridgeshire', 'Peterborough', 'Essex', 'Hertfordshire', 'Norfolk'
    , 'Suffolk', 'Luton', 'Southend-on-Sea', 'Thurrock')
east_of_england_pop <- c(173000, 288000, 852000, 202000, 1832000-183000-174000, 1184000, 903000, 758000, 213000, 183000, 174000)
east_of_england <- as.data.frame(east_of_england_pop, row.names=east_of_england_counties)
colnames(east_of_england) <- c("Population")

eoe_mob <- summarise_regions(east_of_england) %>% mutate(Region = "east_of_england")


london_counties <- c('Greater London')
london_pop <- c(8899000)
london <- as.data.frame(london_pop, row.names=london_counties)
colnames(london) <- c("Population")

lo_mob <- summarise_regions(london) %>% mutate(Region = "london")


north_west_counties <- c('Blackburn with Darwen', 'Blackpool', 'Cheshire East', 'Cheshire West and Chester', 'Cumbria', 'Greater 
    Manchester', 'Borough of Halton', 'Lancashire', 'Merseyside', 'Warrington')
north_west_pop <- c(149000, 139000, 384000, 343000, 498000, 2812000, 129000, 1498000-149000-139000, 1423000, 210000)
north_west <- as.data.frame(north_west_pop, row.names=north_west_counties)
colnames(north_west) <- c("Population")

nw_mob <- summarise_regions(north_west) %>% mutate(Region = "north_west")


north_east_counties <- c('County Durham', 'Darlington', 'Hartlepool', 'Middlesbrough', 'Northumberland', 'Redcar and Cleveland', '
    Stockton-on-Tees', 'Tyne and Wear')
north_east_pop <- c(866000-106000-93000-197000, 106000, 93000, 174000, 320000, 137000, 197000, 1136000)
north_east <- as.data.frame(north_east_pop, row.names=north_east_counties)
colnames(north_east) <- c("Population")

ne_mob <- summarise_regions(north_east) %>% mutate(Region = "north_east")


south_east_counties <- c('Bracknell Forest', 'Buckinghamshire', 'Brighton and Hove', 'Portsmouth', 'Southampton', 'East Sussex', '
    Hampshire', 'Isle of Wight', 'Kent', 'Medway', 'Milton Keynes',
                'Oxfordshire', 'Reading', 'Slough', 'Surrey', 'West Berkshire', 'West Sussex', 'Windsor and Maidenhead', 'Wokingham')
south_east_pop <- c(122000, 806000-248000, 290000, 238000, 269000, 844000-290000, 1844000-238000-269000, 141000, 1846000-278000,
    278000, 248000,
                    687000, 161000, 164000, 1189000, 158000, 858000, 151000, 171000)
south_east <- as.data.frame(south_east_pop, row.names=south_east_counties)
colnames(south_east) <- c("Population")

se_mob <- summarise_regions(south_east) %>% mutate(Region = "south_east")
```

```
south_west_counties <- c('Bath and North East Somerset', 'Bristol City', 'Plymouth', 'Cornwall', 'Devon', 'Dorset', 'Gloucestershire',
        'North Somerset', 'Somerset',
                  'South Gloucestershire', 'Swindon', 'Torbay' , 'Wiltshire')
south_west_pop <- c(193000, 463000, 262000, 568000, 1194000-262000-130000, 772000, 916000-285000, 215000, 965000-193000-215000,
                    285000, 222000, 130000, 720000-222000)
south_west <- as.data.frame(south_west_pop, row.names=south_west_counties)
colnames(south_west) <- c("Population")

sw_mob <- summarise_regions(south_west) %>% mutate(Region = "south_west")


west_midlands_counties <- c('Stoke-on-Trent', 'Herefordshire', 'Shropshire', 'Staffordshire', 'Warwickshire', 'West Midlands', '
      Worcestershire')
west_midlands_pop <- c(256000, 192000, 498000, 1131000-256000, 571000, 2916000, 592000)
west_midlands <- as.data.frame(west_midlands_pop, row.names=west_midlands_counties)
colnames(west_midlands) <- c("Population")

wm_mob <- summarise_regions(west_midlands) %>% mutate(Region = "west_midlands")


yorkshire_and_the_humber_counties <- c('Kingston upon Hull', 'East Riding of Yorkshire', 'North East Lincolnshire', 'North
      Lincolnshire', 'North Yorkshire', 'South Yorkshire', 'West Yorkshire', 'York')
yorkshire_pop <- c(259000, 600000-259000, 159000, 172000, 1158000-210000-197000-106000-137000, 1402000, 2320000, 210000)
yorkshire_and_the_humber <- as.data.frame(yorkshire_pop, row.names=yorkshire_and_the_humber_counties)
colnames(yorkshire_and_the_humber) <- c("Population")

yh_mob <- summarise_regions(yorkshire_and_the_humber) %>% mutate(Region = "yorkshire_and_the_humber")


england_regions <- c("east_midlands", "east_of_england", "london", "north_west", "north_east", "south_east", "south_west", "west_
      midlands", "yorkshire_and_the_humber")
england_pop <- c(sum(east_midlands), sum(east_of_england), sum(london), sum(north_west), sum(north_east), sum(south_east), sum(south_
      west), sum(west_midlands), sum(yorkshire_and_the_humber))
england <- as.data.frame(england_pop, row.names=england_regions)
colnames(england) <- "Population"



mobility_en <- rbind(em_mob, eoe_mob, lo_mob, ne_mob, nw_mob, se_mob, sw_mob, wm_mob, yh_mob) %>%
  mutate(percent_change_from_baseline = percent_change_from_baseline * (england[as.character(Region),] / sum(england))) %>%
  group_by(date, type_mobility) %>%
  summarise(percent_change_from_baseline = sum(percent_change_from_baseline)) %>%
  mutate(country_region_code = as.factor("EN"))



ggplot(mobility_en, aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  facet_wrap(~type_mobility) +
  ggtitle("Change in Different Types of Mobility From Baseline, England") + xlab("Date") + ylab("Change in Mobility (%)") +
  theme(plot.title = element_text(face="bold", hjust="0.5", size = 11))



# English Mobility Graphs per Type

fig16EN <- mobility_en %>%
  filter(type_mobility == "parks") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  ylim(-60, 170) +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-05-13")), color="brown") + # Allowed outside for exercise
  geom_vline(aes(xintercept=as.Date("2020-05-22")), color="red") + # Cummings
```

```
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=70)) +
  geom_text(aes(label="Outdoor exercise\nallowed (10.05)", x=as.Date("2020-05-01"), y=170)) +
  geom_text(aes(label="Second lockdown \nbegins (05.11)", x=as.Date("2020-11-05"), y=70)) +
  geom_text(aes(label="Cummings Scandal\n(22.05)", x=as.Date("2020-05-31"), y=145)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig16EN.png", scale=1.4)


fig17EN <- mobility_en %>%
  filter(type_mobility == "retail_and_recreation") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  geom_vline(aes(xintercept=as.Date("2020-03-20")), colour="darkgrey") + # First closings
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-06-01")), color="magenta") + # Stay-at-home order revoked
  geom_vline(aes(xintercept=as.Date("2020-07-04")), color="orange") + # Pubs, restaurants, hotels, hairdressers allowed to reopen
  geom_vline(aes(xintercept=as.Date("2020-08-03")), color="darkblue") + # Eat Out to Help Out start
  geom_vline(aes(xintercept=as.Date("2020-08-31")), color="cyan") + # Eat Out to Help Out ends
  geom_vline(aes(xintercept=as.Date("2020-12-02")), color="red") + # Lockdown ends
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=20)) +
  geom_text(aes(label="Stay at home\n order revoked \n(01.06)", x=as.Date("2020-06-01"), y=-30)) +
  geom_text(aes(label="Pubs, restaurants,\n hotels, hairdresses\n closed (20.03)", x=as.Date("2020-03-15"), y=-10)) +
  geom_text(aes(label="Pubs, restaurants,\n hotels, hairdresses\n allowed to reopen\n(04.07)", x=as.Date("2020-07-01"), y=-10)) +
  geom_text(aes(label="Eat Out to \nHelp Out begins\n (03.08)", x=as.Date("2020-08-03"), y=5)) +
  geom_text(aes(label="Eat Out to \nHelp Out ends\n(31.08)", x=as.Date("2020-08-31"), y=18)) +
  geom_text(aes(label="Second lockdown \nbegins (5.11)", x=as.Date("2020-11-05"), y=10)) +
  geom_text(aes(label="Second lockdown \nends (02.12)", x=as.Date("2020-12-02"), y=0)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig17EN.png", scale=1.4)

fig18EN <- mobility_en %>%
  filter(type_mobility == "transit_stations") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-05")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-06-01")), color="magenta") + # Stay-at-home order revoked
  geom_vline(aes(xintercept=as.Date("2020-07-17")), color="violet") + # Public transport unrestricted
  geom_vline(aes(xintercept=as.Date("2020-12-02")), color="red") + # Lockdown ends
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=10)) +
  geom_text(aes(label="Stay at home\n order revoked \n(01.06)", x=as.Date("2020-06-01"), y=-30)) +
  geom_text(aes(label="Use of Public \nTransport for non-emergencies\n allowed (17.07)", x=as.Date("2020-07-17"), y=-10)) +
  geom_text(aes(label="Second lockdown \nbegins (05.11)", x=as.Date("2020-11-05"), y=0)) +
  geom_text(aes(label="Second lockdown \nends (02.12)", x=as.Date("2020-12-02"), y=-20)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig18EN.png", scale=1.4)



# Germany Mobility


mobility_de <- read.csv('2020_DE_Region_Mobility_Report.csv') %>%
  mutate(date = as.Date(date)) %>%
  rename('retail_and_recreation' = 'retail_and_recreation_percent_change_from_baseline', 'grocery_and_pharmacy' = 'grocery_and_
        pharmacy_percent_change_from_baseline',
          'parks' = 'parks_percent_change_from_baseline', 'transit_stations' = 'transit_stations_percent_change_from_baseline',
          'workplaces' = 'workplaces_percent_change_from_baseline', 'residential' = 'residential_percent_change_from_baseline') %>%
  gather('retail_and_recreation', 'grocery_and_pharmacy', 'parks', 'transit_stations', 'workplaces', 'residential',
          key='type_mobility', value='percent_change_from_baseline') %>%
  filter(date < as.Date("2021-01-01")) %>%
  filter(is.na(sub_region_2))
```

```
mobility_de$sub_region_1 <- recode(mobility_de$sub_region_1, 'Baden-WÃ¼rttemberg' = 'Baden-Wuerttemberg')
mobility_de$percent_change_from_baseline <- na.locf(mobility_de$percent_change_from_baseline)

mobility_de_overall <- mobility_de %>%
  filter(sub_region_1 == "") %>%
  arrange(date, type_mobility) %>%
  select(date, type_mobility, percent_change_from_baseline, country_region_code)


# Mobility by Type Germany

ggplot(mobility_de_overall, aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second "Partial" Lockdown
  facet_wrap(~type_mobility) +
  ggtitle("Change in Different Types of Mobility From Baseline, Germany") + xlab("Date") + ylab("Change in Mobility (%)") +
  theme(plot.title = element_text(face="bold", hjust="0.5", size = 11))

fig16DE <- mobility_de_overall %>%
  filter(type_mobility == "parks") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-05-06")), color="magenta") + # First wave over, max of two households
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=120)) +
  geom_text(aes(label="First wave 'over,'\n meeting of two households\n allowed (06.05)", x=as.Date("2020-05-06"), y=150)) +
  geom_text(aes(label="Second 'partial' lockdown \nbegins (02.11)", x=as.Date("2020-11-02"), y=150)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig16DE.png", scale=1.4)

fig17DE <- mobility_de_overall %>%
  filter(type_mobility == "retail_and_recreation") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue", label="hi") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-05-06")), color="magenta") + # First wave over, max of two households
  geom_vline(aes(xintercept=as.Date("2020-04-20")), color="orange") + # Large shops allowed to reopen
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=50)) +
  geom_text(aes(label="First wave 'over,'\n all shops open\n allowed (06.05)", x=as.Date("2020-05-10"), y=30)) +
  geom_text(aes(label="Second 'partial' lockdown \nbegins (02.11)", x=as.Date("2020-11-02"), y=20)) +
  geom_text(aes(label="Shops with retail\n space of up to 800\n sq. m. reopen (20.04)", x=as.Date("2020-04-10"), y=5)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig17DE.png", scale=1.4)

fig18DE <- mobility_de_overall %>%
  filter(type_mobility == "transit_stations") %>%
  ggplot(aes(x=date, y=percent_change_from_baseline)) + geom_line() +
  xlab("Date") + ylab("Change in Mobility (%)") +
  geom_vline(aes(xintercept=as.Date("2020-03-23")), colour="blue") + # First lockdown
  geom_vline(aes(xintercept=as.Date("2020-11-02")), colour="darkgreen") + # Second Lockdown
  geom_vline(aes(xintercept=as.Date("2020-05-06")), color="magenta") + # First wave over, max of two households
  geom_text(aes(label="First lockdown \nbegins (23.03)", x=as.Date("2020-03-23"), y=10)) +
  geom_text(aes(label="First wave 'over,'\n meeting of two households\n allowed (06.05)", x=as.Date("2020-05-06"), y=-10)) +
  geom_text(aes(label="Second 'partial' lockdown \nbegins (02.11)", x=as.Date("2020-11-02"), y=10)) +
  geom_line(aes(y=rollmean(percent_change_from_baseline, 7, na.pad=TRUE)), color="orange", size=1.05)

ggsave("fig18DE.png", scale=1.4)



# overall mobility changes both countries
```

```r
mobility_both <- full_join(mobility_de_overall, mobility_en)
fig15 <- ggplot(mobility_both, aes(x=date, y=percent_change_from_baseline, group=country_region_code, color=country_region_code)) +
      geom_line() +
   facet_wrap(~type_mobility) + xlab("Date") + ylab("Change␣in␣Mobility␣(%)") +
   theme(text = element_text(size = 16))

ggsave("fig15.png", scale = 1.5, width = 7)

parks_both <- mobility_both %>% filter(type_mobility == "parks")
t.test(parks_both$percent_change_from_baseline ~ parks_both$country_region_code)

rar_both <- mobility_both %>% filter(type_mobility == "retail_and_recreation")
t.test(rar_both$percent_change_from_baseline ~ rar_both$country_region_code)

resid_both <- mobility_both %>% filter(type_mobility == "residential")
t.test(resid_both$percent_change_from_baseline ~ resid_both$country_region_code)

gam_both <- mobility_both %>% filter(type_mobility == "grocery_and_pharmacy")
t.test(gam_both$percent_change_from_baseline ~ gam_both$country_region_code)

work_both <- mobility_both %>% filter(type_mobility == "workplaces")
t.test(work_both$percent_change_from_baseline ~ work_both$country_region_code)

tran_both <- mobility_both %>% filter(type_mobility == "transit_stations")
t.test(tran_both$percent_change_from_baseline ~ tran_both$country_region_code)


# Section 6.2: # Did first cases have an effect on mobility

transit_en_sept <- mobility_en %>%
   filter(type_mobility == "transit_stations") %>%
   full_join(daily_cases_en, by="date") %>%
   filter(date >= as.Date("2020-08-25")) %>%
   filter(date <= as.Date("2020-10-31")) %>%
   mutate(mob_capped = (percent_change_from_baseline+45.63)/(-17.84+45.63)) %>%
   mutate(cases_capped = (sum_cases-970)/(22578-970))


transit_de_sept <- mobility_de_overall %>%
   filter(type_mobility == "transit_stations") %>%
   full_join(daily_cases_de, by="date") %>%
   filter(date >= as.Date("2020-09-01")) %>%
   filter(date <= as.Date("2020-11-03")) %>%
   mutate(mob_capped = (percent_change_from_baseline+28)/(4+28)) %>%
   mutate(cases_capped = (sum_cases-580)/(19875-580))


fig19EN <- ggplot(transit_en_sept, aes(x=date, y=mob_capped, color="Mobility␣Change")) + geom_line() +
   geom_line(aes(y=cases_capped, color="New␣Daily␣Cases")) +
   geom_line(aes(y=rollmean(mob_capped, 7, na.pad=TRUE)), color="orange", size=1.05) + ylim(0, 1) +
   geom_line(aes(y=rollmean(cases_capped, 7, na.pad=TRUE)), color="green", size=1.05) + labs(x = "Date", y = "Normalised␣Values", color
         = "Legend")

ggsave("fig19EN.png")

cor.test(transit_en_sept$mob_capped, lag(transit_en_sept$cases_capped))
cor.test(transit_en_sept$mob_capped, lag(transit_en_sept$cases_capped, 7))

fig19DE <- ggplot(transit_de_sept, aes(x=date, y=mob_capped, color="Mobility␣Change")) + geom_line() +
   geom_line(aes(y=cases_capped, color="New␣Daily␣Cases")) +
   geom_line(aes(y=rollmean(mob_capped, 7, na.pad=TRUE)), color="orange", size=1.05) + ylim(0, 1) +
   geom_line(aes(y=rollmean(cases_capped, 7, na.pad=TRUE)), color="green", size=1.05) + labs(x = "Date", y = "Normalised␣Values", color
         = "Legend")

ggsave("fig19DE.png")
```

```
cor.test(transit_de_sept$mob_capped, lag(transit_de_sept$cases_capped))
cor.test(transit_de_sept$mob_capped, lag(transit_de_sept$cases_capped, 7))


# COMPARE BOTH COUNTRIES TO sum_cases, ccf to determine forecasting ability and best lag, lm

mobility_en$seven_day_diff <- (mobility_en$percent_change_from_baseline - lag(mobility_en$percent_change_from_baseline, 42))
mobility_de_overall$seven_day_diff <- mobility_de_overall$percent_change_from_baseline - lag(mobility_de_overall$percent_change_from_
    baseline, 42)

transit_en <- mobility_en %>%
  filter(type_mobility == "transit_stations") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))

retail_en <- mobility_en %>%
  filter(type_mobility == "retail_and_recreation") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))

resident_en <- mobility_en %>%
  filter(type_mobility == "residential") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))

grocery_en <- mobility_en %>%
  filter(type_mobility == "grocery_and_pharmacy") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))

parks_en <- mobility_en %>%
  filter(type_mobility == "parks") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))

work_en <- mobility_en %>%
  filter(type_mobility == "workplaces") %>%
  full_join(daily_cases_en, by="date") %>%
  filter(date > as.Date("2020-03-31"))


ccf(transit_en$seven_day_diff, transit_en$percentage_difference_cases, main = "CCF for transit mobility and percent diff. cases (EN)")
ccf(retail_en$seven_day_diff, retail_en$percentage_difference_cases, main = "CCF for retail mobility and percent diff. cases (EN)")
ccf(resident_en$seven_day_diff, resident_en$percentage_difference_cases, main = "CCF for residential mobility and percent diff. cases 
    (EN)")
ccf(grocery_en$seven_day_diff, grocery_en$percentage_difference_cases, main = "CCF for grocery mobility and percent diff. cases (EN)")
ccf(parks_en$seven_day_diff, parks_en$percentage_difference_cases, main = "CCF for parks mobility and percent diff. cases (EN)")
ccf(work_en$seven_day_diff, work_en$percentage_difference_cases, main = "CCF for workplace mobility and percent diff. cases (EN)")

transit_de <- mobility_de_overall %>%
  filter(type_mobility == "transit_stations") %>%
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))

retail_de <- mobility_de_overall %>%
  filter(type_mobility == "retail_and_recreation") %>%
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))

resident_de <- mobility_de_overall %>%
  filter(type_mobility == "residential") %>%
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))

grocery_de <- mobility_de_overall %>%
  filter(type_mobility == "grocery_and_pharmacy") %>%
```

```
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))

parks_de <- mobility_de_overall %>%
  filter(type_mobility == "parks") %>%
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))

work_de <- mobility_de_overall %>%
  filter(type_mobility == "workplaces") %>%
  full_join(daily_cases_de, by="date") %>%
  filter(date > as.Date("2020-03-31"))


ccf(transit_de$seven_day_diff, transit_de$percentage_difference_cases, main = "CCF for transit mobility and percent diff. cases (DE)")
ccf(retail_de$seven_day_diff, retail_de$percentage_difference_cases, main = "CCF for retail mobility and percent diff. cases (DE)")
ccf(resident_de$seven_day_diff, resident_de$percentage_difference_cases, main = "CCF for residential mobility and percent diff. cases
    (DE)")
ccf(grocery_de$seven_day_diff, grocery_de$percentage_difference_cases, main = "CCF for grocery mobility and percent diff. cases (DE)")
ccf(parks_de$seven_day_diff, parks_de$percentage_difference_cases, main = "CCF for parks mobility and percent diff. cases (DE)")
ccf(work_de$seven_day_diff, work_de$percentage_difference_cases, main = "CCF for workplace mobility and percent diff. cases (DE)")


summary(lm(retail_en$percentage_difference_cases ~ lag(retail_en$seven_day_diff, 11) + lag(resident_en$seven_day_diff, 9) +
           lag(work_en$seven_day_diff, 7) + lag(transit_en$seven_day_diff, 9) + lag(grocery_en$seven_day_diff, 11)))

summary(lm(retail_en$percentage_difference_cases ~ lag(retail_en$percentage_difference_cases)))

summary(lm(retail_de$percentage_difference_cases ~ lag(retail_de$seven_day_diff, 6) +
           lag(resident_de$seven_day_diff, 9) + lag(work_de$seven_day_diff, 7) + lag(transit_de$seven_day_diff, 8) +
           lag(grocery_de$seven_day_diff, 9)))
```

Bibliography

[1] Anon. (2017). *Germany age breakdown (2017).* Encyclopdia Britannica. Retrieved from https://britannica.com/place/Germany/Population-structure

[2] Aspinall, E. (2021, March 8). *COVID-19 Timeline.* British Foreign Policy Group. Retrieved from https://bfpg.co.uk/2020/04/covid-19-timeline

[3] Badr, H, S., Du, H., Marshall, M., Dong, E., Squire, M, M., Gardner, L, M. (2020, July 1). *Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study.* The Lancet. Retrieved from https://thelancet.com/journals/laninf/article/PIIS1473-3099(20)30553-3/fulltext

[4] Bender, R. (2020, October 3). *Covid-19 Sparks New East-West Divide in Germany-30 Years After Reunification.* Wall Street Journal. Retrieved from https://wsj.com/articles/covid-19-sparks-new-east-west-divide-in-germany30-years-after-reunification-11601717400

[5] Burdeau, C. (2020, March 17). *Boris Johnson's Talk of 'Herd Immunity' Raises Alarms.* Courthouse News Service. Retrieved from https://courthousenews.com/boris-johnsons-talk-of-herd-immunity-raises-alarms

[6] Bwire, G, M. (2020, June 4). *Coronavirus: Why Men are More Vulnerable to Covid-19 Than Women?* US National Library of Medicine National Institutes of Health. Retrieved from https://ncbi.nlm.nih.gov/pmc/articles/PMC7271824/

[7] Chazan, G. (2020, June 4). *How Germany got coronavirus right.* Financial Times. Retrieved from https://ft.com/content/cc1f650a-91c0-4e1f-b990-ee8ceb5339ea

[8] Culbertson, A. (2020, October 14). *Coronavirus: Why has the North been hit harder now?* Sky News. Retrieved from https://news.sky.com/story/coronavirus-why-has-the-north-been-hit-harder-now-12103152

[9] Department of Statistics. (n.d.). *8.2 Cross Correlation Functions and Lagged Regressions.* PennState, Eberly College of Science. Retrieved from https://online.stat.psu.edu/stat510/lesson/8/8.2

[10] Drees, J. (2020, March 13). *Gender roles and the coronavirus: Why women may have higher risk of catching COVID-19.* Becker's Hospital Review. Retrieved from https://beckershospitalreview.com/care-coordination/gender-roles-and-the-coronavirus-why-women-may-have-higher-risk-of-catching-covid-19.html

[11] Dutch National Institute for Public Health and the Environment. (2021, February 23). *The spread of COVID-19.* Retrieved from https://rivm.nl/en/novel-coronavirus-covid-19/spread

[12] Environmental Systems Research Institute. (2021, January 18). *RKI Covid19.* [Data file]. Retrieved from https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets

[13] Farr, C. (2020, July 16). *The UK did some things right to fight the coronavirus, but it locked down too late.* CNBC. Retrieved from https://cnbc.com/2020/07/16/how-the-uk-fought-the-coronavirus.html

[14] Farr C. (2020, July 21). *Germany's coronavirus response is a master class in science communication.* CNBC. Retrieved from https://cnbc.com/2020/07/21/germanys-coronavirus-response-masterful-science-communication.html

[15] Fenton, T. (2019, December 19). *Regional economic activity by gross domestic product, UK: 1998 to 2018.* Office for National Statistics. Retrieved from https://ons.gov.uk/economy/grossdomesticproductgdp/bulletins/ regionaleconomicactivitybygrossdomesticproductuk/1998to2018

[16] Fetzer, T. (2020, October). *Subsidizing the spread of COVID19: Evidence from the UK's Eat-Out-to-Help-Out scheme.* CAGE. Retrieved from https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/wp.517.2020.pdf

[17] France-Devine, B. (2020, June 18). *Poverty in the UK: statistics.* House of Commons Library. Retrieved from https://commonslibrary.parliament.uk/research-briefings/sn07096/

[18] Gallagher, J. (2020, July 17). *Coronavirus: Boris Johnson sets out plan for 'significant normality' by Christmas.* BBC. Retrieved from https://www.bbc.co.uk/news/uk-53441912

[19] Gallagher, J. (2020, December 20). *New coronavirus variant: What do we know?* BBC. Retrieved from https://www.bbc.co.uk/news/health-55388846

[20] Google. (2021, January 24). *COVID-19 Community Mobility Reports.* [Data file]. Retrieved from https://google.com/covid19/mobility

[21] kassambara. (2018, March 10). *Multiple Linear Regression in R.* STHDA. Retrieved from http://sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r

[22] Marcus, I. (2020, December 20). *Chronology 2020: Germany and the Corona Pandemic.* The Berlin Spectator. Retrieved from https://berlinspectator.com/2020/12/20/chronology-germany-and-the-coronavirus-8/

[23] Martin, D. (2020, June 1). *More people abandoning lockdown rules after Dominic Cummings controversy say DMU researchers.* Leicester Mercury. Retrieved from https://leicestermercury.co.uk/news/leicester-news/more-people-abandoning-lockdown-rules-4180023

[24] Mendolia, S., Stavrunova, O., & Yerokhin, O. (2020, October). *Determinants of the Community Mobility during the COVID-19 Epidemic: The Role of Government Regulations and Information.* Institute of Labor Economics. Retrieved from https://iza.org/de/publications/dp/13778

[25] Office for National Statistics. (2020, June 24). *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland.* [Data file]. Retrieved from https://ons.gov.uk/peoplepopulationandcommunity/populationandmigration/ populationestimates/datasets/ populationestimatesforukenglandandwalesscotlandandnorthernireland

[26] Park, N. (2020, May 6). *Population estimates for the UK, England and Wales, Scotland and Northern Ireland, provisional: mid-2019.* Office for National Statistics. Retrieved from https://ons.gov.uk/peoplepopulationandcommunity/populationandmigration/ populationestimates/bulletins/annualmidyearpopulationestimates/mid2019

[27] Pieper, J., Schneider, U., Schroeder, W. (2020, November). *Der Paritaetische Armutsbericht 2020.* Deutscher Paritaetischer Wohlfahrtsverband. Retrieved from http://derparitaetische.de/schwerpunkt/armutsbericht/

[28] Pijls, B, G., Jolani, S., Atherley, A., Derckx, R, T., Dijksttra, J, I, R., Frannsen, G, H, L., ... Zeegers, M, P. (2020, December 14). *Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies.* BMJ Open. Retrieved from https://bmjopen.bmj.com/content/11/1/e044640

[29] Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., Colizza, V. (2020, October 28). *Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study.* Lancet Digital Health, Retrieved from https://thelancet.com/journals/landig/article/PIIS2589-7500(20)30243-0/fulltext

[30] Rose, J. (2021, January 7). *Scientists think they know when, where and how the new coronavirus strain started in Kent.* KentLive. Retrieved from https://kentlive.news/news/kent-news/scientists-think-know-when-how-4866103

[31] Rosney, D. (2020, September, 7). *Coronavirus: Young people breaking rules risk 'second wave'*. BBC. Retrieved from https://bbc.co.uk/news/newsbeat-54056771

[32] Roy, S., & Ghosh, P. (2020, October 23). *Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking.* PLoS ONE. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241165

[33] Schumacher, E.. (2020, May 6). *Merkel cautiously optimistic as she announces lockdown rollback.* Deutsche Welle. Retrieved from https://www.dw.com/en/merkel-cautiously-optimistic-as-she-announces-lockdown-rollback/a-53346427

[34] Scobie, S. (2020, December 9). *How many cases of Covid-19 are there?* nuffieldtrust. Retrieved from https://nuffieldtrust.org.uk/news-item/how-many-cases-of-covid-19-are-there-a-q-a

[35] Statistical tools for high-throughput data analysis. (n.d.). *Correlation Test Between Two Variables in R.* STHDA. Retrieved from http://sthda.com/english/wiki/correlation-test-between-two-variables-in-r

[36] Statistical tools for high-throughput data analysis. (n.d.). *One-Way ANOVA Test in R.* STHDA. Retrieved from http://sthda.com/english/wiki/one-way-anova-test-in-r

[37] Statistical tools for high-throughput data analysis. (n.d.). *Unpaired Two-Samples T-test in R.* STHDA. Retrieved from http://sthda.com/english/wiki/correlation-test-between-two-variables-in-r

[38] Statistik Portal. (2020, September 24). *Bruttoinlandsprodukt, Bruttowertschoepfung.* Statistische Aemter des Bundes und der Laender. Retrieved from https://statistikportal.de/de/vgrdl/ergebnisse-laenderebene/bruttoinlandsprodukt-bruttowertschoepfung

[39] Statistische Aemter des Bundes und der Laender. (2020, September 2). *Kreisfreie Staedte und Landkreise nach Flaeche, Bevoelkerung und Bevoelkerungsdichte am 31.12.2019.* [Data file]. Retrieved from https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/04-kreise.html

[40] Stone, J. (2020, June 26). *'Fatal failure': How the world's media covered the UK's botched response to coronavirus.* Independent. Retrieved from https://independent.co.uk/news/uk/politics/coronavirus-uk-cases-government-boris-johnson-covid-19-death-toll-a9588111.html

[41] Tupper, P., Boury, H., Yerlanov, M., & Colijn, C. (2020, November 2). *Event-specific interventions to minimize COVID-19 transmission.* Proceedings of the National Academy of Sciences of the United States of America. Retrieved from https://pnas.org/content/117/50/32038

[42] UK Government. (2021, February 13). *Daily COVID Cases and Deaths by UTLA, England.* [Data file]. Retrieved from https://coronavirus.data.gov.uk/details/download

[43] UK Government. (2021, March 11). *Daily COVID Cases and Deaths by Region, England.* [Data file]. Retrieved from https://coronavirus.data.gov.uk/details/download

[44] Velasco, J, M., Tseng, W., & Chang, C. (2021, January 14). *Factors Affecting the Cases and Deaths of COVID-19 Victims.* International Journal of Environmental Research and Public Health. Retrieved from https://www.mdpi.com/1660-4601/18/2/674

[45] Waschinski, G. (2021, February 10). *35 sind die neuen 50: Das steckt hinter dem Grenzwert, der jetzt ueber Oeffnungen entscheidet*. Handelsblatt. Retrieved from https://handelsblatt.com/politik/deutschland/26901656.html

[46] Weaver, M. (2020, May 22). *Pressure on Dominic Cummings to quit over lockdown breach.* The Guardian. Retrieved from https://theguardian.com/politics/2020/may/22/dominic-cummings-durham-trip-coronavirus-lockdown

[47] zensus2011. (2011). *Zensus 2011.* Statistische Aemter des Bundes und der Laender. Retrieved from https://ergebnisse2011.zensus2022.de/datenbank/online/

[48] Zhou, Y., Xu, R., Hu, D., Yue, Y., Li, Q., Xia, J. (2020, August). *Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data.* Lancet Digital Health, Retrieved from https://thelancet.com/journals/landig/article/PIIS2589-7500(20)30165-5/fulltext