Social mobility and gonorrhoea in Germany during 2020

Mihály Sulyok,1,2 Mark Walker3

ABSTRACT

Objectives The incidence of STIs is likely to be related to levels of social activity and mobility. Novel datasets detailing levels of social activity were made widely available during the COVID-19 pandemic. These allow the relationship between activity and STI incidence to be examined.

Methods The correlation between social activities and the reported number of gonorrhoea cases between March and December 2020 in Germany was studied. Regression through Autoregressive Integrated Moving Average (ARIMA) time series modelling identified those activities associated with case numbers.

Results ARIMA regression identified a significant association with ‘transit’ activity within the Apple data and ‘parks’ within Google.

Conclusions This study illustrates the potential newly available measures of social activity provided for STI research. Reductions in STI incidence are likely to have occurred due to COVID-19 social restrictions. Although other studies report reductions in infectious diseases during this period, few examine the potential social factors mediating this. The results illustrate the continual need for sexual health services throughout the pandemic.

INTRODUCTION

That a relationship exists between social activity and the incidence of STIs is intuitive, increased levels of mobility and social interaction provide increased opportunities for potential transmission of these conditions. Although a relationship is well established for some activities, notably international travel,1,2 it has not been demonstrated for other forms of social activity. This is understandable; there are considerable practical problems in ascertaining daily levels of social activity and mobility, within communities.3 Traditionally, mobility was quantified with censuses or surveys of small populations.3 New mobile phone technology means collection of data for large populations is now feasible. The potential of such datasets has already been recognised.4

The impact of COVID-19 on society has been large. Various restrictions on everyday life were imposed throughout 2020 in order to hinder viral transmission. Numerous studies have examined the effect of these travel and social restrictions on COVID-19 incidence.5,6 These restrictions have had an unintended effect on the incidence of other infectious diseases. For example, in Germany, reductions in reported numbers of a range of notifiable infectious diseases have been described.5 Although such studies report reductions in the incidence of infectious diseases, how social activity or mobility mediated this remains unexplored.

In response to the COVID-19 pandemic, datasets providing mobility information were made widely available. Google’s ‘Community Mobility Reports’ (CMR) details activity at different locations, by those accessing Google applications.9 Apple’s ‘Mobility Trends’ (MT) shows the volume of direction requests on Apple maps.10 The Oxford ‘Coronavirus Governmental Response Tracker’ collates governmental restrictions into a single indexed value of restriction severity.11 The availability of the Apple and Google datasets allows the influence of social activities on disease incidence to be examined.

Gonorrhoea, caused by Neisseria gonorrhoeae, is an ideal candidate for examination. There have been consistent annual increases in cases. Gonorrhoea is now the second most commonly diagnosed STI across Europe11 and the third in Germany.12 Gonorrhoea infection is distinct and noticeable, meaning diagnosis is sought promptly.13 The time from infection to symptom onset is typically 1–5 days.14 This means trends in reporting are likely to mirror incidence. Changes in population activity and mobility are likely to quickly become apparent in reported case data.

Here, the relationship between reported cases of gonorrhoea since the advent of COVID-19 and social activities was examined using the newly available mobility datasets. Since the 1970s, Autoregressive Integrated Moving Average (ARIMA) has gained widespread popularity as a method for time series forecasting.15 Despite its simplicity and ease of application, it has been proven more accurate than other more complicated methods. Additionally, ARIMA can perform dynamic regression; identifying the effect of external regressors on a response variable.16 Although conceived for time series, ARIMA has been far less frequently used for regression than forecasting, despite being superior to other techniques.17 Here, ARIMA regression was used to identify those activities of potential influence on gonorrhoea.

Investigating STIs throughout the COVID-19 pandemic is particularly important. Evidence suggests sexual activity continued, even when social activities were legally restricted.18 However, accessibility to sexual health services was restricted throughout 2020.19 This poses the risk of future rises in STI incidence in the future. Study is of broader interest; knowledge of the relationship between social activity and STIs increases epidemiological understanding of these conditions.
METHODS

Data

The number of gonorrhoea cases reported to the Robert Koch Institute (RKI) and obtainable from the Survstat Database was downloaded on 19 February 2021.\(^{20}\) RKI reports gonorrhoea cases where reduced antibiotic susceptibility against azithromycin, cefixime or ceftriaxone has been observed and which are thus notifiable to RKI (Klaus Jansen, RKI. Personal Communication).

Activity data were obtained from Google’s CMR,\(^{9}\) Apple’s MT\(^{10}\) and the Oxford Tracker.\(^{7}\) CMR shows daily volume of activity at six location categories: ‘retail and recreation’, ‘grocery and pharmacy’, ‘parks’, ‘workplace’, ‘transit’ and ‘residential’ indexed against median baseline values from January and February 2020. MT shows daily volume of direction requests in categories: ‘driving’, ‘transit’ and ‘walking’, indexed against 13 January 2020. The Oxford Tracker provides a single indexed overall ‘Stringency Index’ (SI).

For MT and CMR, mean weekly values for each activity category were calculated. SI was recorded at weekly intervals. Data were examined visually. Differences in case numbers between years were examined using Analysis of Variance (ANOVA) testing and Tukey Honest Significant Difference (HSD) testing.

Correlations

Relationships were examined using Spearman rank cross-correlations. Normality was examined using Shapiro-Wilk. Data were used from 14 February until 20 December 2020. Data were not used after this due to fluctuations in case numbers apparently related to Christmas and New Year holiday periods. Access to sexual health services during this time is likely to be limited. Delays to testing and subsequent reporting may occur due to staff absences.

First, contemporaneous correlations were ascertained, then cross-correlations were performed to ascertain strongest correlations, with activity data correlated across a time frame of −7 to +7 weeks. This encompasses symptom onset in most cases, which in most cases is 2–5 days.\(^{14}\) Expected was that strongest correlations would be observed when there was a positive lead to activity data; symptom onset and diagnosis is expected to occur several days following changes in activity. Results for contemporaneous Spearman rank ρ, and the maximum correlation following cross-correlation with activity data, are reported.

Modelling

ARIMA modelling was used for dynamic regression, allowing the potential influence of external regressors on case numbers to be ascertained. The method outlined by Hyndman and Athanasopoulos was followed.\(^{16}\)

a. Model assumptions: data were graphed and visually examined. ARIMA modelling assumes that the time series under examination is stationary. This was checked with Kwiatkowski-Phillips-Schmidt-Shin (KPSS) testing and examination of autocorrelation.

b. Model establishment: an ARIMA model was established with social mobility categories being used as explanatory variables accounting for cases.

c. Model diagnostics: model residuals were plotted and examined. The associated autocorrelation plot and histogram were examined. A Ljung-Box test was performed to test randomness of autocorrelations.

RESULTS

Trends in reported numbers

Figure 1 shows the number of gonorrhoea cases during 2020. No apparent trend in the number of cases reported throughout the year is observable. A notable peak in reported cases occurred in the summer and autumn, possibly corresponding to easing of COVID-19 social restrictions.\(^{5}\) A significant decline in reported cases in 2020 is apparent compared with previous years (ANOVA; df=4, F=4.247, p<0.002; online supplemental files 2; 3).

Mobility

Figures showing mobility for each index examined are provided in the online supplemental files 4–6. Table 1 shows the strength of contemporaneous correlations and cross-correlation between reported cases and activities. As expected, there are positive

Table 1 Correlations between indices of social activity and the number of gonorrhoea cases reported to RKI

<table>
<thead>
<tr>
<th>Index</th>
<th>Contemporaneous correlation ρ</th>
<th>Cross-correlation, values are maximum ρ observed (time lag/lead in weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple MT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driving</td>
<td>0.29</td>
<td>0.34 (0)</td>
</tr>
<tr>
<td>Transit</td>
<td>0.34</td>
<td>0.27 (+2)</td>
</tr>
<tr>
<td>Walking</td>
<td>0.26</td>
<td>0.30 (+2)</td>
</tr>
<tr>
<td>Google CMR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail and recreation</td>
<td>0.05</td>
<td>0.51 (+1)</td>
</tr>
<tr>
<td>Grocery and pharmacy</td>
<td>0.34</td>
<td>0.50 (+1)</td>
</tr>
<tr>
<td>Parks</td>
<td>0.20</td>
<td>0.43 (+1)</td>
</tr>
<tr>
<td>Workplace</td>
<td>0.34</td>
<td>0.28 (+2)</td>
</tr>
<tr>
<td>Transit</td>
<td>0.06</td>
<td>0.45 (+1)</td>
</tr>
<tr>
<td>Residential</td>
<td>0.08</td>
<td>−0.45 (+3)</td>
</tr>
<tr>
<td>Oxford Restriction Tracker</td>
<td>0.229</td>
<td>−0.39 (+1)</td>
</tr>
</tbody>
</table>

Value in brackets shows number of lead/lag days for activity data which resulted in maximum correlation. CMR, Community Mobility Reports; MT, Mobility Trends; RKI, Robert Koch Institute.
relationships between reported case numbers and those categories indicative of social activity and mobility when examining contemporaneous correlations. Of note, the strongest contemporaneous correlations were observed with ‘transit’ (0.34) in MT, and ‘grocery and pharmacy’ and ‘workplace’ (both 0.34) in CMR. Activity within these categories is likely to reflect general levels of activity among society. However, there is a negative correlation for ‘residential’ activity and ‘SI’, both indicative of sedentary activity.

With cross-correlation analyses, what is notable are not the strength of correlation, but the lead periods resulting in highest figures as a delay between social activity changes, and a subsequent effect on case numbers is expected. Along with delays in subsequent testing and reporting, strongest correlations with a lead time of +1 to +4 weeks could well be realistic. For most categories, this was indeed the case, with the strongest relationships occurring with lead times +1 to +3 weeks.

Modelling

Exploratory analyses

Shapiro-Wilk testing indicated that case numbers for 2020 and activity data were non-normally distributed. Results of KPSS testing on case numbers indicated that they were stationary (KPSS level=0.404, lag parameter=3, p=0.07). This was corroborated through the associated Autocorrelation function (ACF) plot (online supplemental file 7). Differentiation was testing on case numbers indicated that they were stationary 

Model establishment

A model comprising a single autoregressive component was identified as fitting data best, ARIMA (1,0,0). The log likelihood was −125.19. The model Akaike information criterion (AIC) was 276.39. Model coefficients are provided (table 2). Examination of variable significance indicates that within MT, ‘transit’ was significant (estimate: −0.284; SE: 0.0799; p<0.001) and within CMR, ‘parks’ (estimate: −0.0936; SE: 0.0936, p<0.001).

Model diagnostics

Residuals were distributed evenly, suggesting model fitting was good online supplemental file 8. The residual histogram indicates normal distribution. Ljung-Box testing showed that residuals are randomly distributed (Q=0.409, p=0.523).

DISCUSSION

The principal aim of this investigation was to demonstrate how novel datasets detailing levels of social activity could be used in the study of STI epidemiology. Previous studies have typically relied on survey data, often using small sample sizes. The advantage of Google and Apple data is that data for large numbers of individuals are aggregated, meaning activity patterns are likely to represent community activity. Additionally, information for discrete location types is provided. These datasets provide ideal tools to examine the relationship between activity and STIs.

Case numbers

No apparent trend was observable in reported case numbers throughout 2020. A general rise in case numbers might be expected as restrictions eased following spring implementation. Compared with previous years, the number of reported cases was lower in 2020 than previous years. However, disentangling the underlying cause of this is difficult. The lower numbers for 2020 may be reflective of a decline in incidence because of the influence of COVID-19 which limited levels of social interaction. They could however also be a result of reduced access to sexual health services or an unwillingness to visit healthcare facilities due to fears of COVID-19 infection. Access to sexual health services was much reduced during 2020. Changes to the definition of reported cases by RKI could also have played a part in reporting, making comparison with previous years difficult.

Relationship with activities

Expected were positive correlations between reported case numbers and social activity; and conversely negative correlations...

### Table 2 ARIMA model coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>SE</th>
<th>T statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>−0.476</td>
<td>0.137</td>
<td>−3.48</td>
<td>0.0013</td>
</tr>
<tr>
<td>MT: driving</td>
<td>0.224</td>
<td>0.162</td>
<td>1.38</td>
<td>0.174</td>
</tr>
<tr>
<td>MT: transit</td>
<td>−0.284</td>
<td>0.0799</td>
<td>−3.55</td>
<td>0.00091</td>
</tr>
<tr>
<td>MT: walking</td>
<td>0.0712</td>
<td>0.151</td>
<td>0.473</td>
<td>0.639</td>
</tr>
<tr>
<td>Oxford: SI</td>
<td>−0.0401</td>
<td>0.118</td>
<td>−0.341</td>
<td>0.735</td>
</tr>
<tr>
<td>CMR: retail</td>
<td>0.224</td>
<td>0.171</td>
<td>1.31</td>
<td>0.197</td>
</tr>
<tr>
<td>CMR: grocery</td>
<td>−0.224</td>
<td>0.175</td>
<td>−1.28</td>
<td>0.208</td>
</tr>
<tr>
<td>CMR: parks</td>
<td>0.405</td>
<td>0.343</td>
<td>1.18</td>
<td>0.244</td>
</tr>
<tr>
<td>CMR: workplace</td>
<td>−0.0936</td>
<td>0.0396</td>
<td>−2.36</td>
<td>0.0226</td>
</tr>
<tr>
<td>CMR: residential</td>
<td>0.613</td>
<td>1.38</td>
<td>0.44</td>
<td>0.66</td>
</tr>
<tr>
<td>Intercept</td>
<td>20</td>
<td>4.86</td>
<td>4.12</td>
<td>0.00016</td>
</tr>
</tbody>
</table>

ARIMA, Autoregressive Integrated Moving Average; CMR, Community Mobility Reports; MT, Mobility Trends; SI, Stringency Index.

Sulyok M, Walker M. Sex Transm Infect 2021;0:1–5. doi:10.1136/sextrans-2021-055159
between those reflecting increased staying at home. This was seen in cross-correlations, most notably for the SI index, which provides a composite measure of social activity.

For most categories, the strongest cross-correlations were obtained when activity data were time shifted +1to +3 weeks against case data. This is consistent with the expected delay between infection, symptom onset and diagnosis. Although correlations were not strong for any category, this may reflect gonorrhoea epidemiology. In a large proportion of cases, gonorrhoea is asymptomatic.13 14 Thus, the actual strength of association is probably greater than that actually observed. Many factors influence incidence, not only social activity and mobility.21 However, the direction of correlations and patterns in cross-correlation followed that, which was expected.

Category importance
A key feature of this research was the attempt to identify which activities were associated with gonorrhoea infections. This is not to imply that increases in activity in one specific location type are directly responsible for corresponding changes in STIs. For example, examination of CMR data indicated that ‘parks’ activity was associated with case numbers. This does not necessarily mean there is a connection. Both parks activity and STI incidence could be influenced by the same underlying factors, or both reflect general patterns in social behaviour. Warmer weather and longer days could be possible factors affecting social activity in general and having a knock-on effect on STI incidence. Social activity in general may well be seasonal in nature. Further research could examine the causative reasons underlying the association. It must be remembered that 2020 was a most unusual year; outdoor activities could have assumed far greater importance in this year when other forms of recreational activity were restricted.

Factors affecting incidence
Many factors have been shown to influence gonorrhoea incidence. Particularly, social aspects, such as social deprivation, high population density and low socioeconomic status, were associated with incidence.21–23 Higher levels of incidence are observed within high-risk populations, including the young and non-white.24 Geospatial aspects are also likely to affect gonorrhoea incidence. Studies examining census data have shown how location and disease incidence are related.22 23 Mobility has been postulated as a potential factor in modelling studies.24 Cases are concentrated among particular ethnic groups.25 Studies examining chlamydia incidence have shown a relationship with mobility at the local residential scale.26

COVID-19 and sexual health
A number of studies have reported reductions in STIs during 2020.24 25 However, disentangling whether this was due to a decline in incidence or because of COVID-19-related factors is difficult. Reported case numbers for gonorrhoea remained low throughout 2020, even subsequent to easing of restrictions and improved access to sexual services, which is suggestive that there was a general decline in incidence. Postponement to the seeking of medical attention due either to COVID-19 fears or reduced medical access could have led to a subsequent surge in STI cases.27 28 Although there was a peak in reported cases in autumn, this was not sustained in nature suggesting that this did not occur.

Strengths and limitations
Few studies have examined the relationship between STIs and social activity at this scale. Thus, a key strength of this research is that it adds to research on this topic. Gonorrhoea was considered an ideal candidate for examination being one of the most common sexual conditions in Germany.13 14 In comparison, other sexual conditions occur much less frequently. Additionally, diagnosis of other conditions such as HIV and syphilis depends on a much greater extent of physician referral, meaning that these conditions were much more greatly influenced by restrictions to sexual health services during 2020.19 Testing for gonorrhoea is routinely offered by sexual health clinics, and can even be done postally. Additionally, symptoms for other STIs only become apparent some time following infection, meaning influence of mobility changes on case numbers would not be as apparent as for those where symptom onset is rapid such as gonorrhoea.

Another strength is that the restriction measures imposed to combat COVID-19 created an ideal experimental situation, artificially manipulating levels of societal activity. This created a unique situation which could not be replicated normally. The magnitude of activity changes that occurred was greater than would ever be observed normally. It is generally acknowledged that knowledge of gonorrhoea epidemiology in Germany is poor.29 The data used for regression were obtained from the RKI Survstat Database and relate to only notified cases of gonorrhoea.12 This is likely to reflect the general incidence of gonorrhoea.29 Trends in such data are likely to reflect corresponding changes in incidence generally. However, further study of this would be of interest. Another limitation is that mobility data may not reflect activity within those populations where gonorrhoea transmission is centred. However, mobility phone coverage in all population groups is now high in Germany.30

Avenues for further study
An obvious next step would be to examine whether mobility and activity differ among those specific groups most affected by this condition. Additionally, mobile phone data offer the future potential for spatial analyses of user movement, and how that relates to disease epidemiology. For example, whether the mobility patterns and spatial usage of those groups in which gonorrhoea cases are known to be concentrated differ from the general population would be of interest. This has been shown for other conditions.21 This would offer an interesting new avenue for investigation.

Public health implications
Case of gonorrhoea continued to be notified to RKI despite the considerable social restrictions during 2020; if anything this emphasises the fact that cases continued to occur throughout the pandemic despite the presence of such restrictions. This underlines the importance of continuous access to sexual health services; these should be maintained in the face of further restrictions. The study also illustrates the social nature of STIs. The methodology introduced, with ARIMA modelling being used to perform dynamic regression, suggests how further studies could use this underused technique and the data sources identified here to identify those activities potentially associated with the transmission of a range of infectious conditions.
CONCLUSION
In response to the COVID-19 pandemic, various novel datasets detailing volumes of social activity at different categories of location and mobility by different modes of transport were made widely available. This provided a unique opportunity to investigate which activities were associated with STI transmission. Gonorrhoea makes an ideal candidate for study being one of the most common sexually transmitted conditions, and because of the short period from infection to symptom onset. Strongest correlations were observed with a time lead of +1 to +3 weeks to cases, indicative of the time from infection to symptom onset and subsequent diagnosis. Dynamic regression performed as part of ARIMA modelling suggested an association between parks and transit activity. Results illustrate the social nature of STI and offer the potential to study further in future.

Key messages
► The interaction between social activity and STIs is complex, but of importance in understanding the epidemiology of these conditions.
► In response to COVID-19, Apple and Google released novel datasets detailing levels of social activity and mobility. The relationship with gonorrhoea case numbers was studied for Germany.
► Social activities were cross-correlated with case numbers. Dynamic regression as part of Autoregressive Integrated Moving Average modelling was conducted to identify those activities most strongly associated with reported gonorrhoea cases.
► The strongest relationship occurred with a lead-on activity data +1 to +3 weeks. Regression found an association between ‘transit’ activity within Apple data and ‘parks’ within Google.

REFERENCES
8 Ullrich A, Schranz M, Rexroth U. The impact of the COVID-19 pandemic and associated public health measures on other notifiable infectious diseases under national surveillance in Germany, week 1-2016–Week 32-2020. GZHW.
spearmanf <- function(x) {
  cor.test(gon_and_activity$goncases, x)$estimate
}

library(readr)
gon_and_activity <- read_csv("~/Downloads/gon_master.xlsx - Sheet2.csv", col_types = cols(date = col_date(format = "%d/%m/%Y")))

summary(gon_and_activity)
##       date               goncases     drivingweekmean
##  Min.   :2020-02-16   Min.   : 4.00   Min.   : 46.14   Min.   :31.18
##  1st Qu.:2020-05-03   1st Qu.: 8.00   1st Qu.: 84.34   1st Qu.:77.82
##  Median :2020-07-19   Median :10.00   Median :111.54   Median :106.48
##  Mean   :2020-07-19   Mean   :11.02   Mean   :109.02   Mean   : 98.72
##  3rd Qu.:2020-10-04   3rd Qu.:13.00   3rd Qu.:135.64   3rd Qu.:130.24
##  Max.   :2020-12-20   Max.   :28.00   Max.   :157.24   Max.   :152.76
##      transitweekmean    stringency        retail          grocery
##  Min.   :-16.71   Min.   :-58.00   Min.   :-65.14   Min.   :-30.0000
##  1st Qu.:  8.00   1st Qu.: 39.71   1st Qu.: -41.86   1st Qu.:  -6.0000
##  Median :111.76   Median : 59.72   Median : -12.00   Median :  0.8571
##  Mean   :107.86   Mean   : 57.82   Mean   : -22.82   Mean   :  2.4921
##  3rd Qu.:137.49   3rd Qu.: 64.81   3rd Qu.:  -5.00   3rd Qu.:  3.1429
##  Max.   :155.94   Max.   : 64.81   Max.   :  5.50   Max.   :  27.0000
##      parks           transit         workplace       residential
##  Min.   :-16.71   Min.   :-58.00   Min.   :-65.14   Min.   : 0.000
##  1st Qu.:  8.00   1st Qu.: -39.71   1st Qu.: -25.57   1st Qu.:  3.143
##  Median :43.57   Median : -21.43   Median : -17.43   Median :  5.143
##  Mean   :51.32   Mean   : -26.49   Mean   : -21.64   Mean   :  6.902
##  3rd Qu.:90.43   3rd Qu.: -16.29   3rd Qu.: -14.57   3rd Qu.:  11.000
##  Max.   :149.43   Max.   :  2.00   Max.   :  0.00   Max.   : 18.857

lapply(gon_and_activity[2:12], shapiro.test)
## $goncases
##
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.91635, p-value = 0.003187

## $drivingweekmean
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.94724, p-value = 0.03998

## $transitweekmean
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.93305, p-value = 0.01204

## $walkingweekmean
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.94737, p-value = 0.04044

## $stringency
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.86318, p-value = 8.142e-05

## $retail
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.8735, p-value = 0.0001564

## $grocery
## Shapiro-Wilk normality test
## $parks
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.87986, p-value = 0.000237
##
## ## $transit
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.91429, p-value = 0.002722
##
## ## $workplace
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.93845, p-value = 0.01887
##
## ## $residential
## Shapiro-Wilk normality test
## data: X[[i]]
## W = 0.91802, p-value = 0.003626

spearmanf<-function(x) {
  cor.test(gon_and_activity$goncases, x)$p.value
}

p<-as.data.frame(lapply(gon_and_activity[3:12], spearmanf))
p<-as.vector(p)
padj<-p.adjust(p,"holm",10)
padj

## drivingweekmean transitweekmean walkingweekmean stringency
## retail
## 0.22996525 0.34164990 0.26041078 0.22996525
## 0.05745214
##
## grocery parks transit workplace
## residential
```r
## 0.34164990 0.20367280 0.06505678 0.34164990 0.08303123
est<- as.data.frame(lapply(gon_and_activity[3:12], spearmanf))
est
##
## drivingweekmean transitweekmean walkingweekmean stringency
## 1 0.038622       0.1693287       0.0651027 0.03832754
## 2 0.08303123
##
## retail
grocery      parks transit workplace residential
## 1 0.1138833 0.02909611 0.007228531 0.1732345   0.0103789
## 2 0.005745214
contemp<- rbind(est, padj)

library(Hmisc)
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
##
## format.pval, units
ccfspear<- function(y){ ccfspearmanx <- sapply( -3:3, function(l)
    cor.test(y, Hmisc::Lag(gon_and_activity$goncases,l), method =
    "spearman", use = "complete.obs", exact=FALSE)$estimate)
}
est<- as.data.frame(lapply(gon_and_activity[3:12], ccfspear))

ccfspear<- function(y){ ccfspearmanx <- sapply( -3:3, function(l)
    cor.test(y, Hmisc::Lag(gon_and_activity$goncases,l), method =
    "spearman", use = "complete.obs", exact=FALSE)$p.value)
}
p<- as.data.frame(lapply(gon_and_activity[3:12], ccfspear))
p<- as.matrix(p)
padjunct<- matrix(p.adjust(p, "holm", 70), nrow=7, ncol=10)
padjunct
```

Sulyok M, Walker M. Sex Transm Infect 2021;0:1–5. doi: 10.1136/sextrans-2021-055159
[1,]  1.000000 1 1 1.0000000 1.000000 1.000000 0.07516131 1.000000  
[2,]  1.000000 1 1 1.0000000 1.000000 1.000000 0.28852541 1.000000  
[3,]  1.000000 1 1 1.0000000 0.2789515 1.0000000 1.000000 0.27863624  
[4,]  0.929459 1 1 0.3776859 0.1061867 0.40303813 0.26705075 0.09989371  
[5,]  1.000000 1 1 1.0000000 0.0239737 0.02832612 0.18313202 0.18855380  
[6,]  1.000000 1 1 0.4588861 0.1746424 0.51717661 0.40303813 0.28739363  
[7,]  1.000000 1 1 1.0000000 0.3203195 0.22256502 0.47933840 0.26705075  
[8]    est  

## drivingweekmean transitweekmean walkingweekmean stringency retail  
## 1  0.09999678 -0.09641674 0.03457985 -0.04364172 0.2761435  
## 2  0.16091375 -0.06874648 0.08057057 -0.04581780 0.3194382  
## 3  0.24237221  0.08331324 0.19972545 -0.23447763 0.4135688  
## 4  0.34166365  0.23987181 0.29962494 -0.39219850 0.4568015  
## 5  0.32926902  0.24582025 0.29843541 -0.31520661 0.5154748  
## 6  0.33186199  0.27035591 0.30566585 -0.3880019 0.4458953  
## 7  0.24862941  0.20196058 0.21927845 -0.23784260 0.4140939  

## grocery parks transit workplace residential  
## 1  0.1953446 0.4859078 0.2673252 0.1191726 -0.2686112  
## 2  0.2986783 0.4149801 0.3079655 0.1519113 -0.3164139  
## 3  0.3210346 0.3373550 0.4144321 0.2603637 -0.4157579  
## 4  0.3822199 0.4141077 0.4596441 0.1960678 -0.4338489  
## 5  0.5096071 0.4377524 0.4353386 0.2241331 -0.4380836  

Sulyok M, Walker M. Sex Transm Infect 2021;0:1–5. doi: 10.1136/sextrans-2021-055159
```r
library(ggplot2)
library(tidyverse)

gon_lon <- gather(gon_and_activity, category, value, goncases:residential)

ggplot(gon_lon, aes(date, value, group=category, col=category)) + geom_line() + scale_y_log10()
```

```r
library(ggfortify)
autoplot(ts(gon_and_activity[2:12], start=c(2020,7), frequency=52.18), ncol=3, geom="ribbon")
```
library(tsibble)

##
## Attaching package: 'tsibble'
##
## The following objects are masked from 'package:base':
##
## intersect, setdiff, union

library(fable)

## Loading required package: fabletools

library(feasts)

gon <- as_tsibble(gon_and_activity)

## Using `date` as index variable.

gg_tsdisplay(gon) # looks stationary

## Plot variable not specified, automatically selected `y = goncases`
gon %>%
  features(goncases, unitroot_kpss) # and is stationary

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1     0.405      0.0751

fit <- gon %>%
  model(arima = ARIMA(goncases ~ drivingweekmean + transitweekmean +
                    walkingweekmean + stringency + retail + grocery + transit + parks +
                    workplace + residential))

fit

## # A mable: 1 x 1
## arima
## # 1 <LM w/ ARIMA(1,0,0) errors>

glance(fit)

## # A tibble: 1 x 8
## .model sigma2 log_lik     AIC     AICc     BIC ar_roots ma_roots
## <chr>  <dbl>   <dbl>    <dbl>    <dbl>    <dbl>    <list>    <list>
## 1 arima 20.7  -125.  276.  288.  300.  <cpl [1]> <cpl [0]>
### Series: goncases
### Model: LM w/ ARIMA(1,0,0) errors

### Coefficients:
<table>
<thead>
<tr>
<th></th>
<th>ar1</th>
<th>drivingweekmean</th>
<th>transitweekmean</th>
<th>walkingweekmean</th>
</tr>
</thead>
<tbody>
<tr>
<td>stringency</td>
<td>-0.0401</td>
<td>0.2241</td>
<td>-0.2839</td>
<td>0.0712</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.1177</td>
<td>0.1623</td>
<td>0.0799</td>
<td>0.1506</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>retail</th>
<th>grocery</th>
<th>transit</th>
<th>parks</th>
<th>workplace</th>
<th>residential</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.2237</td>
<td>-0.2240</td>
<td>0.4052</td>
<td>-0.0936</td>
<td>-0.2369</td>
<td>0.6133</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.1709</td>
<td>0.1754</td>
<td>0.3429</td>
<td>0.0396</td>
<td>0.1959</td>
<td>1.3828</td>
</tr>
</tbody>
</table>

### sigma^2 estimated as 20.71: log likelihood=-125.19
### AIC=276.39   AICc=288.13   BIC=299.88

#### coefficients(fit)

# A tibble: 12 x 6
## .model term      estimate std.error statistic  p.value
## <chr>  <chr>             <dbl>     <dbl>     <dbl>    <dbl>
## 1 arima ar1        -0.476     0.137     -3.48  0.00113
## 2 arima drivingweekmean 0.224     0.162      1.38  0.174
## 3 arima transitweekmean -0.284    0.0799    -3.55  0.000911
## 4 arima walkingweekmean 0.0712    0.151      0.473 0.639
## 5 arima stringency  -0.0401    0.118      -0.341 0.735
## 6 arima retail      0.224     0.171      1.31   0.197
## 7 arima grocery     -0.224    0.175      -1.28  0.208
## 8 arima transit     0.405     0.343      1.18   0.244
## 9 arima parks       -0.0936   0.0396     -2.36  0.0226
## 10 arima workplace  -0.237    0.196      -1.21  0.233
## 11 arima residential 0.613     1.38       0.443 0.660
## 12 arima intercept  20.0      4.86       4.12   0.000160

#### accuracy(fit)

# A tibble: 1 x 10
## .model .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE ACF1
## <chr> <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 arima Training -0.0393 3.90 3.06 -12.1 31.7 0.536 0.519 -0.0922
## Series: goncases
## Model: LM w/ ARIMA(1,0,0) errors
##
## Coefficients:
##           ar1  drivingweekmean transitweekmean walkingweekmean
## stringency    -0.4765           0.2241          -0.2839           0.0712
##             -0.0401
## s.e.   0.1369           0.1623           0.0799           0.1506
## 0.1177
## retail   grocery   transit    parks   workplace   residential
## intercept 0.2237  -0.2240   0.4052  -0.0936    -0.2369       0.6133
## 20.0278
## s.e.  0.1709   0.1754   0.3429   0.0396     0.1959       1.3828
## 4.8608
##
## sigma^2 estimated as 20.71:  log likelihood=-125.19
## AIC=276.39   AICc=288.13   BIC=299.88

```r
gg_tsresiduals(fit)  #looks fine
```

```r
augment(fit) %>%
  features(.innov, ljung_box)
```

## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
## 1    fit  20.0   none
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>arima</td>
<td>0.409</td>
</tr>
</tbody>
</table>
95% family-wise confidence level

Differences in mean levels of as.factor(Year)
Driving

Transit

Walking