Technical Appendix

Team 8

Contents

Preface

The aim of this document is to provide the decisions, code and visualisations substantiating the appended report and its conclusions.

Packages

The below installs (if uncommented) and imports the libraries required for later analysis.

```r
# "Uncomment" the relevant lines below

# install.packages("finalfit")
# install.packages("knitr")
# install.packages("rio")
# install.packages("dplyr")
# install.packages("readxl")
# install.packages("stringr")
# install.packages("ggplot2")
# install.packages("gridExtra")
# install.packages("zoo")
# install.packages("ggcorrplot")
# install.packages("lubridate")
# install.packages("runner")
# install.packages("flexsurv")
# install.packages("survival")
# install.packages("survminer")
# install.packages("cmprsk")
# install.packages("remotes")
# remotes::install_github("nset-ornl/wbstats")
# remotes::install_github("joachim-gassen/tidycovid19")
# install.packages("ggrepel")
# install.packages("grid")
# install.packages("RColorBrewer")
# install.packages("tidyverse")
# install.packages("maps")

library(finalfit)
library(knitr)
library(rio)
library(dplyr)
library(readxl)
library(stringr)
library(ggplot2)
library(gridExtra)
library(zoo)
library(ggcorrplot)
library(lubridate)
library(runner)
library(flexsurv)
library(survival)
library(survminer)
library(cmprsk)
library(remotes)
library(tidycovid19)
library(ggrepel)
library(grid)
library(RColorBrewer)
library(tidyverse)
```

Data

Data were joined offline from the original sources of OxCGRT and tidycovid19. The joined data was subset by countries in Europe, North America, and South America using the "continent" variable in OxCGRT. The cleaned data is imported below

```
EAData <- read.csv("./data/processed/europe_americas.csv")
```

We then change the confirmed variable from cumulative cases to the daily cases, and removing missing values.

```
EAData$DaysSinceJan1 <- as.numeric(as.Date(EAData$date) - as.Date("01-01-2020","%d-%m-%Y"))

firstdiff <- function(x) {
  shifted <- c(0,x[1:(length(x)-1)])
  result = x-shifted
  which_negative = which(result<0)
  result[which_negative] = NA
  return(result)
}

EAData <- EAData %>%
  mutate(daily_confirmed = firstdiff(confirmed)) %>%
filter(!is.na(continent)) %>%
  filter(!is.na(C6_Stay.at.home.requirements)) %>%
  filter(!is.na(H4_Emergency.investment.in.healthcare)) %>%
  mutate(daily_confirmed = if_else(is.na(daily_confirmed), 0, daily_confirmed))
```

Here we are performing some simple data manipulation to get the data set into a style we can work with. We filter out any NA values and change the confirmed variable to the daily confirmed cases.

Before starting any further analysis, in order to make our models more accurate we have mutated the retail, workplaces and grocery data to become a rolling mean of the previous 7 days. This is because data for the weekends was very different to the data on week days, leading to spikes every week.

Also adding an estimate of current active cases once confirmed in with 14 days until removed and a lag of 2 days symptoms before test result.

```
EAData$week <- week(EAData$date)
EAData <- EAData %>% group_by(country) %>% mutate(m_retail = rollmean(gcmr_retail_recreation, k = 7, fil


EAData <- EAData %>%
  group_by(country) %>%
  mutate(
    date = ymd(date),
    active = sum_run(
      x = daily_confirmed,
      k = 14,
      lag = 2
    )) %>%
      mutate(
        active_per_pop = active/population
      ) %>%
  ungroup()
```

```r
euroData <- EAData %>% filter(continent == "Europe")

northData <- EAData %>% filter(continent == "North America")

southData <- EAData %>% filter(continent == "South America")
```

Next we create some simple visualisations to get a feel of the data.

```r
read_csv("./data/processed/acaps_npi.csv", col_types = cols()) %>%
  mutate(npi_date = ymd(date_implemented)) %>%
  rename (npi_type = category) %>%
  select(iso3c, npi_date, npi_type) -> npi

EAData %>%
  group_by(iso3c) %>%
  filter(deaths >= 10) %>%
  summarise(edate = min(date)) -> ctry_edate

EAData %>%
  select(iso3c, country) %>%
  unique() -> ctry_names


npi %>%
  left_join(ctry_edate, by = "iso3c") %>%
  filter(!is.na(edate)) %>%
  mutate(npi_edate = as.numeric(npi_date - edate)) %>%
  left_join(ctry_names, by = "iso3c") %>%
  select(iso3c, country, npi_date, npi_edate, npi_type) -> npi_edates
```

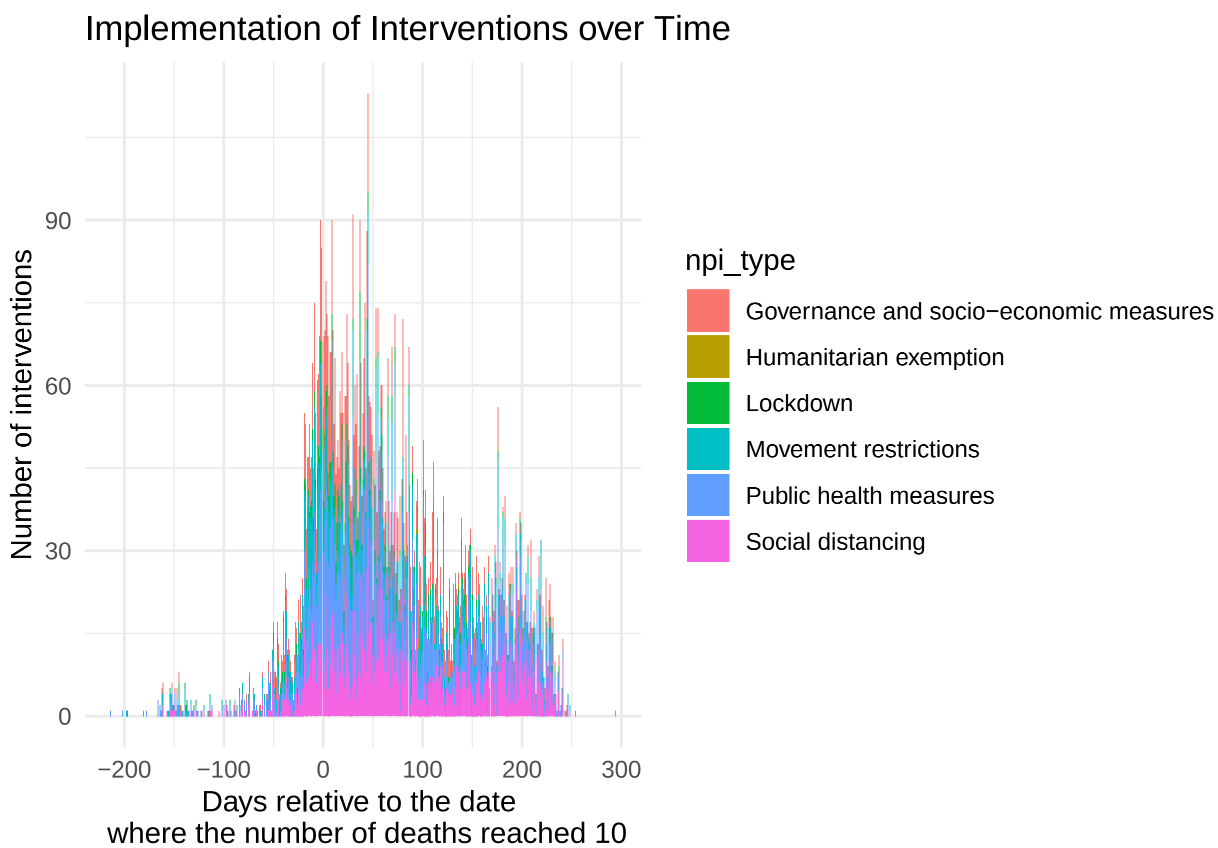
1 Exploratory Data Analysis

1.1 Broad Analysis

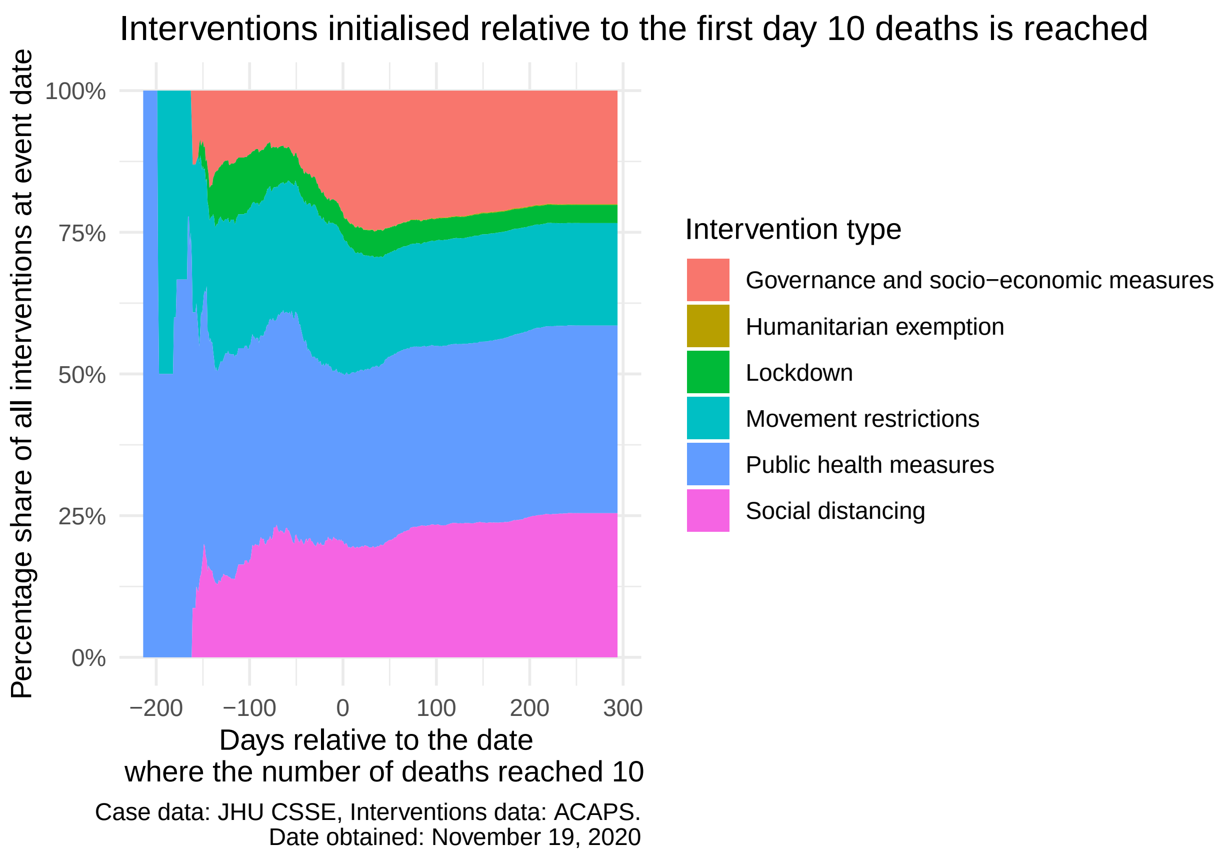
1.1.1 Interventions

Aim to visualise the types of interventions, and when they came into place. The visualisations should be informative to compare countries.

```r
lab_x <- "Days relative to the date \n where the number of deaths reached 10"


ggplot(npi_edates, aes(x = npi_edate, fill = npi_type)) +
  geom_bar(position = "stack") + theme_minimal() +
  labs(title = "Implementation of Interventions over Time",
       x = lab_x,
       y = "Number of interventions")
```




```
npi_edates %>%
  group_by(npi_edate, npi_type) %>%
  summarise(
    npi_count = n()
  ) %>%
  ungroup() %>%
  arrange(npi_type, npi_edate) %>%
  group_by(npi_type) %>%
  mutate(npi_count =  cumsum(npi_count)) %>%
  complete(npi_edate = min(npi_edates$npi_edate):max(npi_edates$npi_edate)) %>%
  fill(npi_count) %>%
  replace_na(list(npi_count = 0)) %>%
  ggplot(aes(x = npi_edate, fill = npi_type, y = npi_count)) +
  theme_minimal() + labs(
    x = lab_x,
    y = "Percentage share of all interventions at event date",
    title = "Interventions initialised relative to the first day 10 deaths is reached",
    fill = "Intervention type", caption="Case data: JHU CSSE, Interventions data: ACAPS.\n Date obtaine
  geom_area(position = "fill") +
  scale_y_continuous(labels = scales::percent)
```



Next classifying based on Covid "prevalence" per capita, where prevalence is defined as the number of cases per test conducted. Further social interaction measures are also defined and explained below.

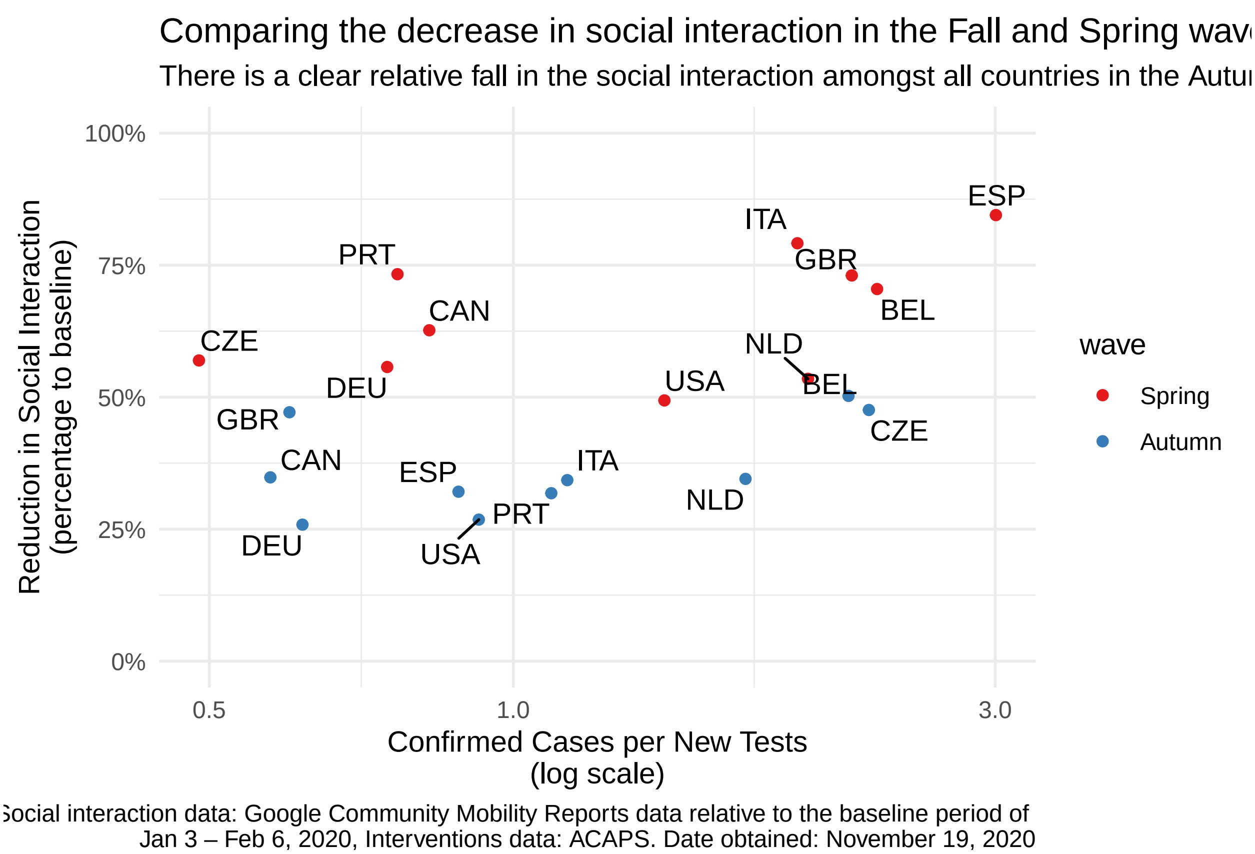
```
df <- read.csv("./data/processed/europe_americas.csv")

ctries <- df %>%
  group_by(iso3c) %>%
  filter(!is.na(deaths), population >= 10e6) %>%
  filter(date == max(date)) %>%
  summarise(deaths_per_mio_pop = deaths * 1e6/population) %>%
  filter(deaths_per_mio_pop > 100) %>%
  pull(iso3c)

ave_measures <- df %>%
  arrange(iso3c, date) %>%
  group_by(iso3c) %>%
  filter(iso3c %in% ctries) %>%
  mutate(
    new_cases = confirmed - lag(confirmed),
    total_tests = na.approx(total_tests, na.rm = FALSE),
    new_tests = total_tests - lag(total_tests),
    ave_pos_test_rate = rollsum(
      (confirmed - lag(confirmed))/new_tests,
      7, na.pad=TRUE, align="right"
    ),
    ave_new_cases_wk_per_100e5 = rollsum(
```

```r
      new_cases*1e5/population, 7, na.pad=TRUE, align="right"
    ),
    ave_soc_dist_google = rollmean(
      (gcmr_retail_recreation +  gcmr_transit_stations +
        gcmr_workplaces)/3, 7, na.pad=TRUE, align="right"
    ),
    ave_soc_dist_apple = rollmean(
      (apple_mtr_driving + apple_mtr_walking + apple_mtr_transit)/3,
      7, na.pad=TRUE, align="right"
    )
  ) %>%
  filter(
    max(
      (date < lubridate::ymd("2020-04-01")) * ave_new_cases_wk_per_100e5,
      na.rm = TRUE
    ) > 10
  ) %>%
  select(
    iso3c, country, date, population, ave_new_cases_wk_per_100e5,
    ave_pos_test_rate, ave_soc_dist_apple, ave_soc_dist_google, new_tests
  )

smp_countries <- unique(ave_measures$country)
```

```r
my_palette <- c(brewer.pal(8, "Set1"), "lightblue")

ave_measures %>%
  filter(date < lubridate::ymd("2020-06-01")) %>%
  summarise(
    cases = max(ave_pos_test_rate, na.rm = TRUE),
    tests = mean(new_tests*1e5/population, na.rm = TRUE),
    soc_dist = -min(ave_soc_dist_google, na.rm = TRUE)/100,
    .groups = "drop"
  ) %>% mutate(wave = "Spring") %>%
  select(iso3c, wave, cases, soc_dist, tests) -> spring_wave

ave_measures %>%
  filter(date > lubridate::ymd("2020-09-01")) %>%
  summarise(
    cases = max(ave_pos_test_rate, na.rm = TRUE),
    tests = mean(new_tests*1e5/population, na.rm = TRUE),
    soc_dist = -min(ave_soc_dist_google, na.rm = TRUE)/100,
    .groups = "drop"
  ) %>% mutate(wave = "Autumn") %>%
  select(iso3c, wave, cases, soc_dist, tests) -> fall_wave

soc_dist_by_wave <- rbind(spring_wave, fall_wave)
soc_dist_by_wave$wave <- factor(soc_dist_by_wave$wave, c("Spring", "Autumn"))
```

The analysis here is constricted to countries which have a population of more than 10 million that have had at least 100 Covid19 related deaths per million residents This ensures that the analysis is limited to reasonably large countries that have been significantly affected by Covid-19. Furthermore, to focus on countries that experienced both, the Spring and the Autumn waves, we constrict to the countries that had a peak daily new infections higher than 30 per 100,000 residents before April 2020.

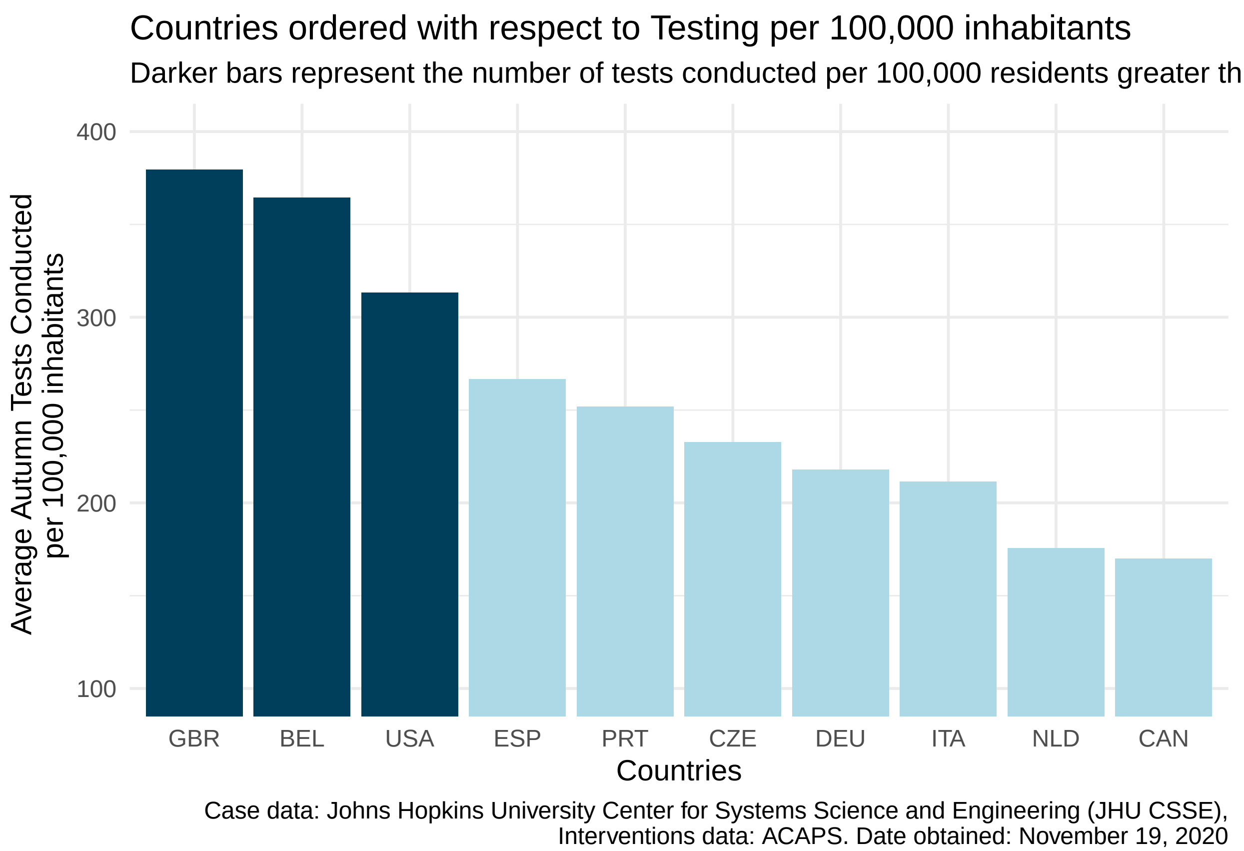
```
ggplot(soc_dist_by_wave, aes(
  x = cases, y = soc_dist, color = wave, label = iso3c)) +
  geom_point() +
  geom_text_repel(color = "black") +
  scale_x_continuous(trans = "log10") +
  scale_y_continuous(labels = scales::percent, limits = c(0, 1)) +
  theme_minimal() +
  scale_color_manual(values = my_palette) +
  labs(
    x = "Confirmed Cases per New Tests\n(log scale)",
    y = "Reduction in Social Interaction\n(percentage to baseline)",
    title = "Comparing the decrease in social interaction in the Fall and Spring waves",
    subtitle = "There is a clear relative fall in the social interaction amongst all countries in the Au
    caption="Social interaction data: Google Community Mobility Reports data relative to the baseline p
```



The plot above displays the reduction in social interaction with a baseline period set to be Jan 3 – Feb 6, 2020 against the log of confirmed cases per new tests. Social interaction represents the minimum rolling mean of average recreation, transit and workplaces Google mobility data across the countries relative to the baseline for the Spring and Autumn waves The x-axis here is the max rolling mean of confirmed cases per new tests for the 2 waves. Per new tests accounts for the increase in testing over the months.

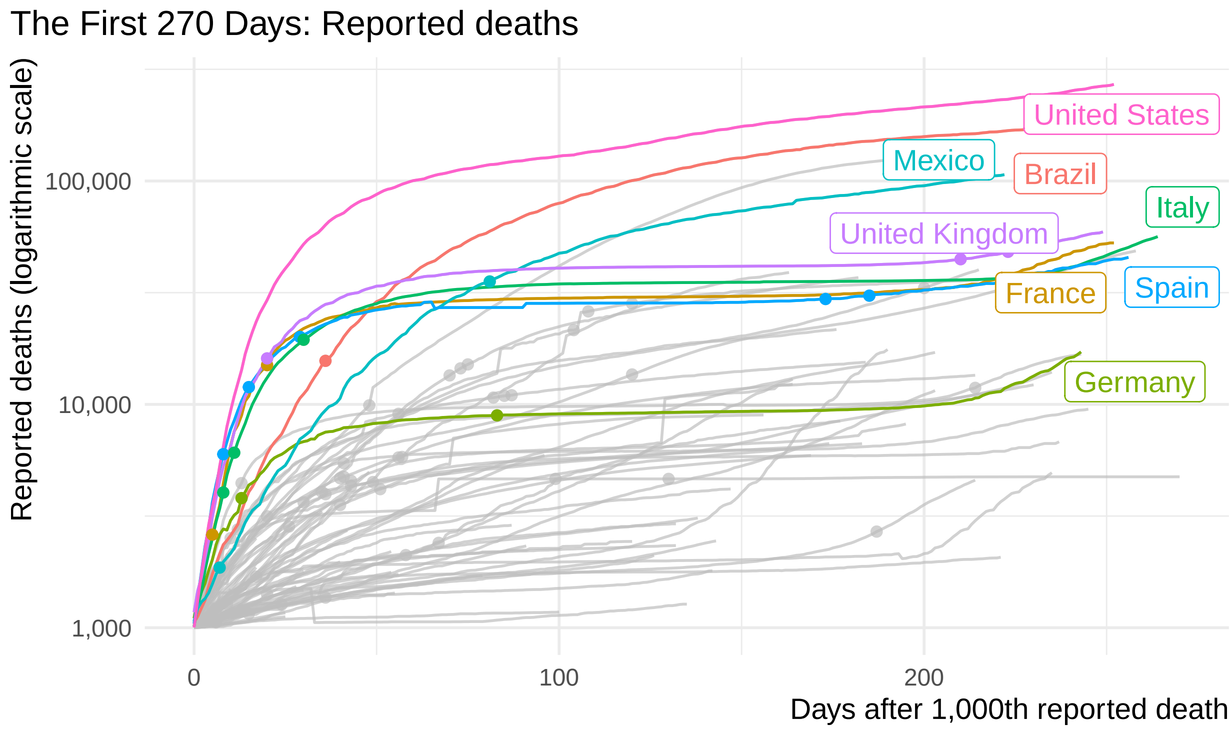
The graph shows a general fall in the social interactions in the Autumn waves relative to the Spring.

```r
fall_testing = fall_wave[-c(6,11), ]

ggplot(fall_testing, aes(x=reorder(iso3c, -tests), y=tests, label = iso3c)) +
  geom_bar(stat = "identity", fill = ifelse(fall_testing$tests>300, "#003F5C", "lightblue")) + coord_ca
```



The plot above displays the average daily new tests conducted per 100,000 inhabitants for countries from the 1st of September. The average daily new tests conducted per 100,000 inhabitants can be seen as metric of how many COVID-19 tests are available to the public for a population. In a business sense, this could be seen as employees having a greater easiness in getting a test and hence faster test results implying employees can act on that faster; be it self-isolation or getting back to work. Hence, recommend the business to consider focusing operations in these Great Britain, Belgium and the USA. Possible flaw here there will be more tests readily available because of the fact that country has more cases.

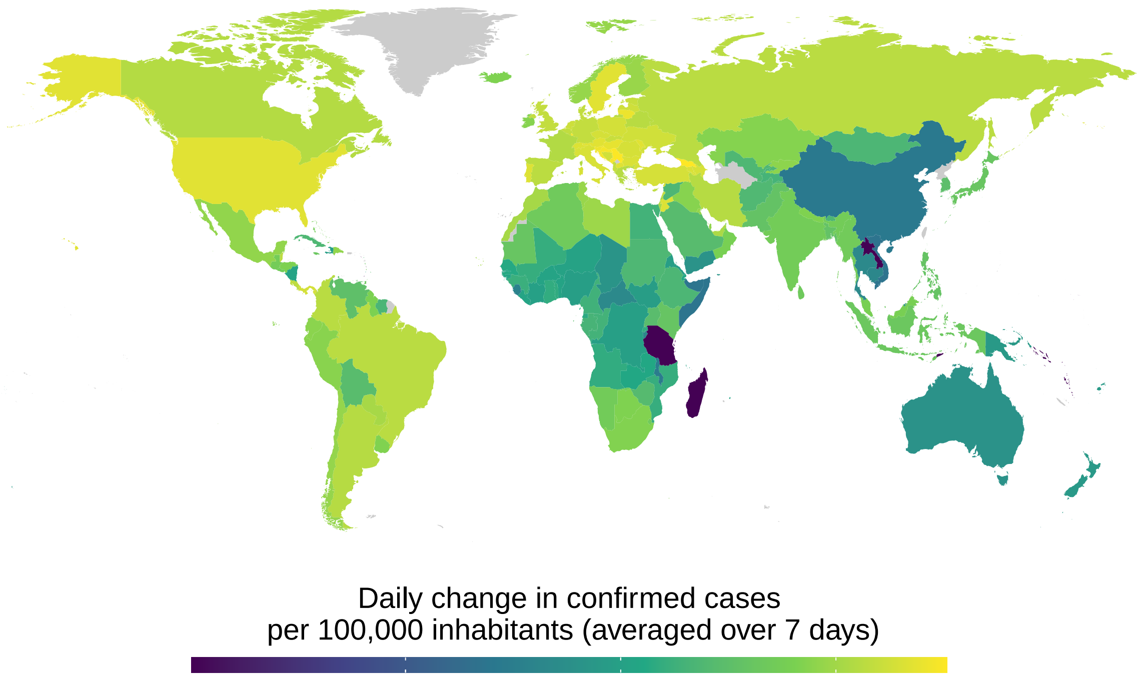
```r
library(tidycovid19)
merged <- download_merged_data(cached = TRUE, silent = TRUE)
plot_covid19_spread(merged, highlight = c("ITA", "ESP", "GBR", "FRA",
                                           "DEU", "USA", "BRA", "MEX"),
intervention = "lockdown", edate_cutoff = 270)
```



Case data: Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Interventions data: Data obtained on December 03, 2020. The sample is limited to countries with at least 7 days of data. Dots indicate governmental interventions of type 'lockdown'. Code: https://github.com/joachim−gassen/tidycovid19.

```
map_covid19(merged, type = "confirmed", per_capita = TRUE)
```

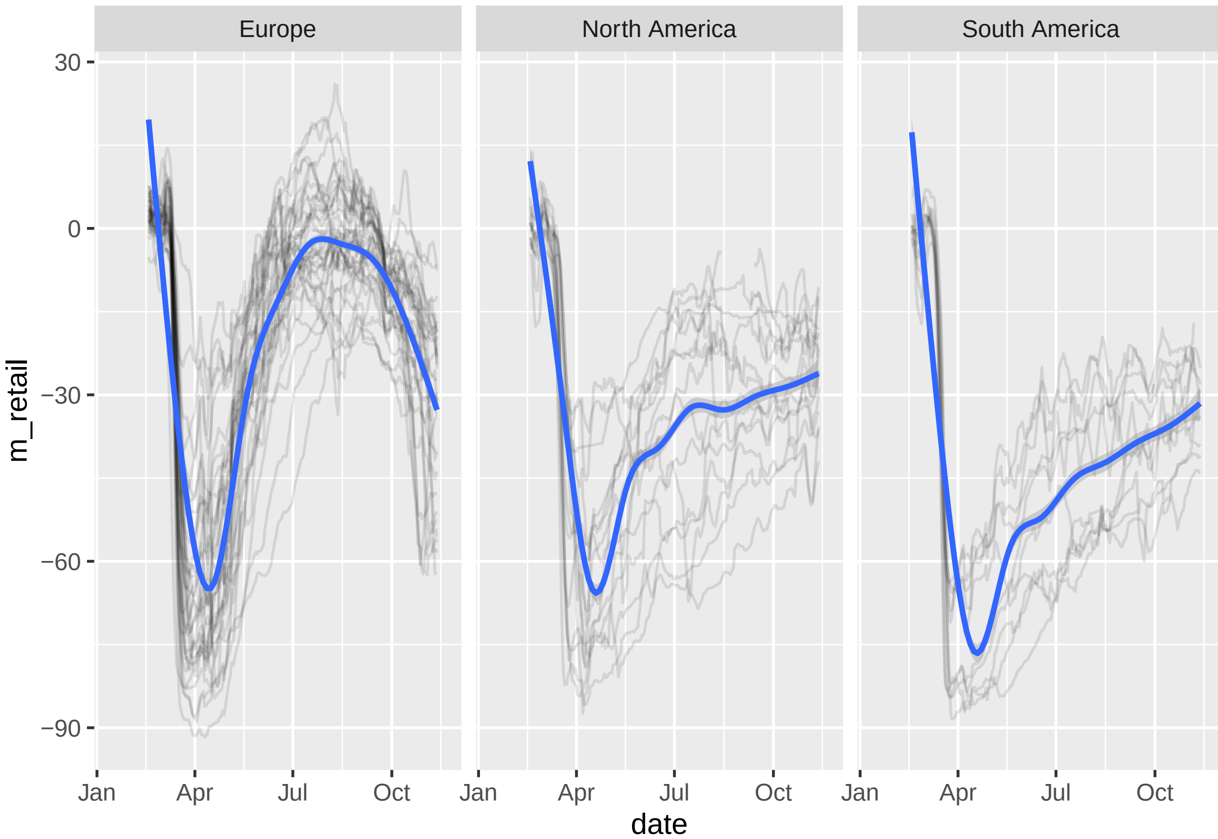
Covid19: Confirmed cases (new cases per day) as of December 01, 2020



Case data: Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Population data: Worldbank. Data obtained on December 03, 2020. Code: https://github.com/joachim−gassen/tidycovid19.

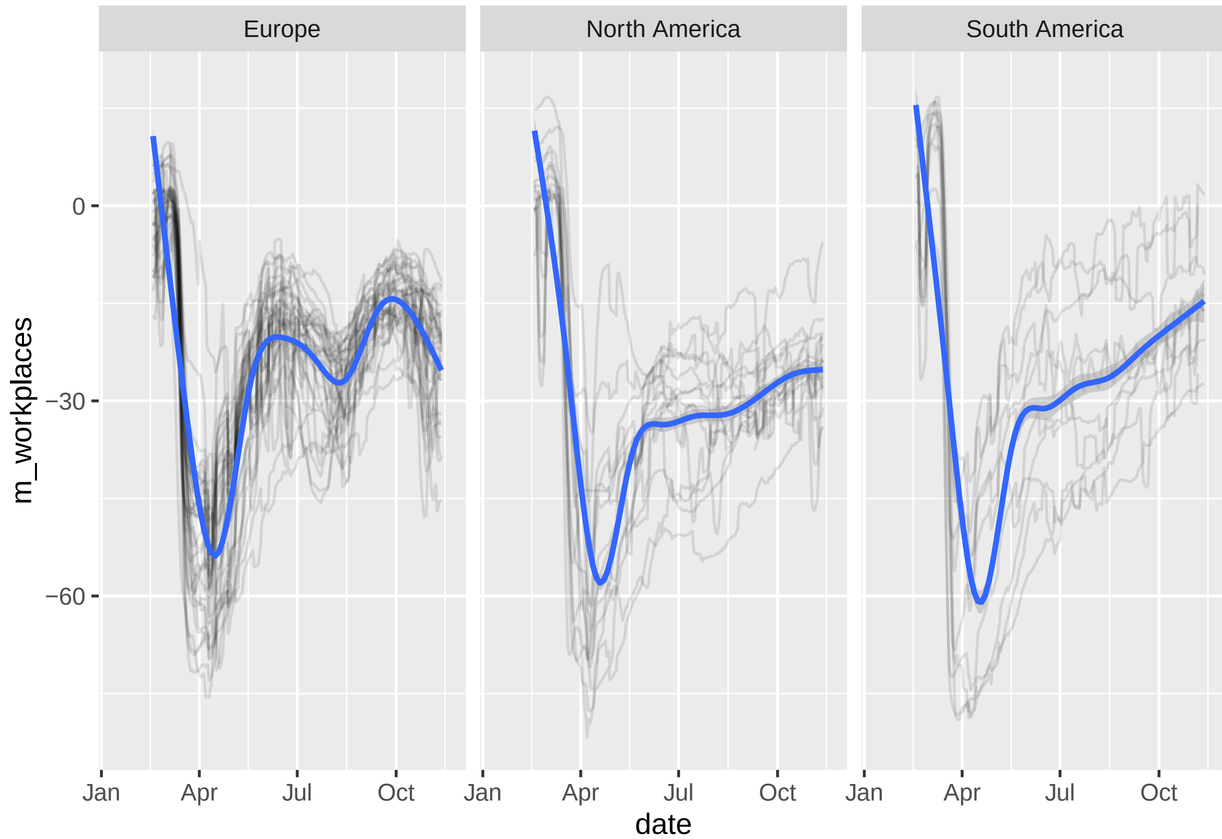
1.1.2 Retail Recreation

```
ggplot(aes(x=date, y=m_retail), data = EAData) + geom_line(aes(group = country),
                                                           alpha = 0.1) +
  facet_wrap(~ continent) + geom_point(alpha = 0) + geom_smooth(method = "gam")
```



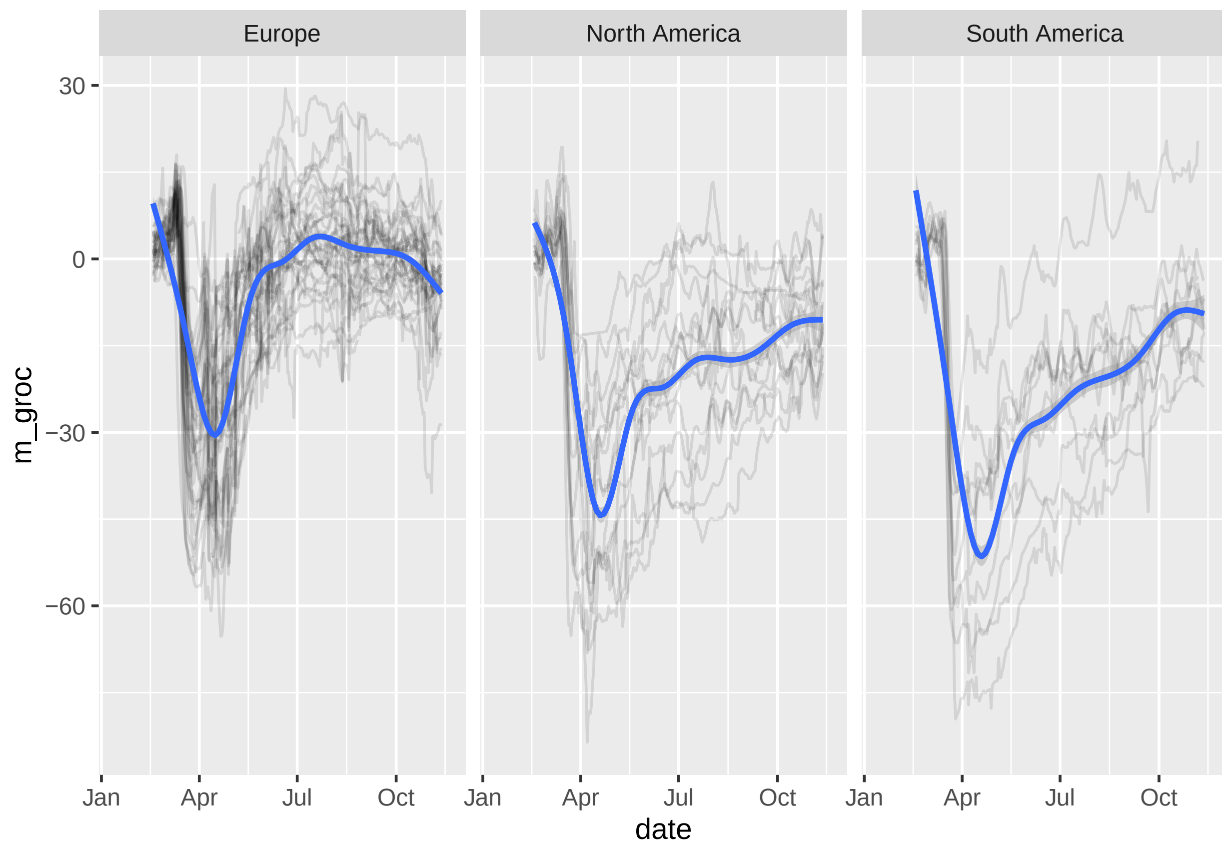
1.1.3 Workplaces

```
ggplot(aes(x=date, y=m_workplaces), data = EAData) +
  geom_line(aes(group = country), alpha = 0.1) +
  facet_wrap(~ continent) + geom_point(alpha = 0) + geom_smooth(method = "gam")
```



1.1.4 Grocery Pharmacy

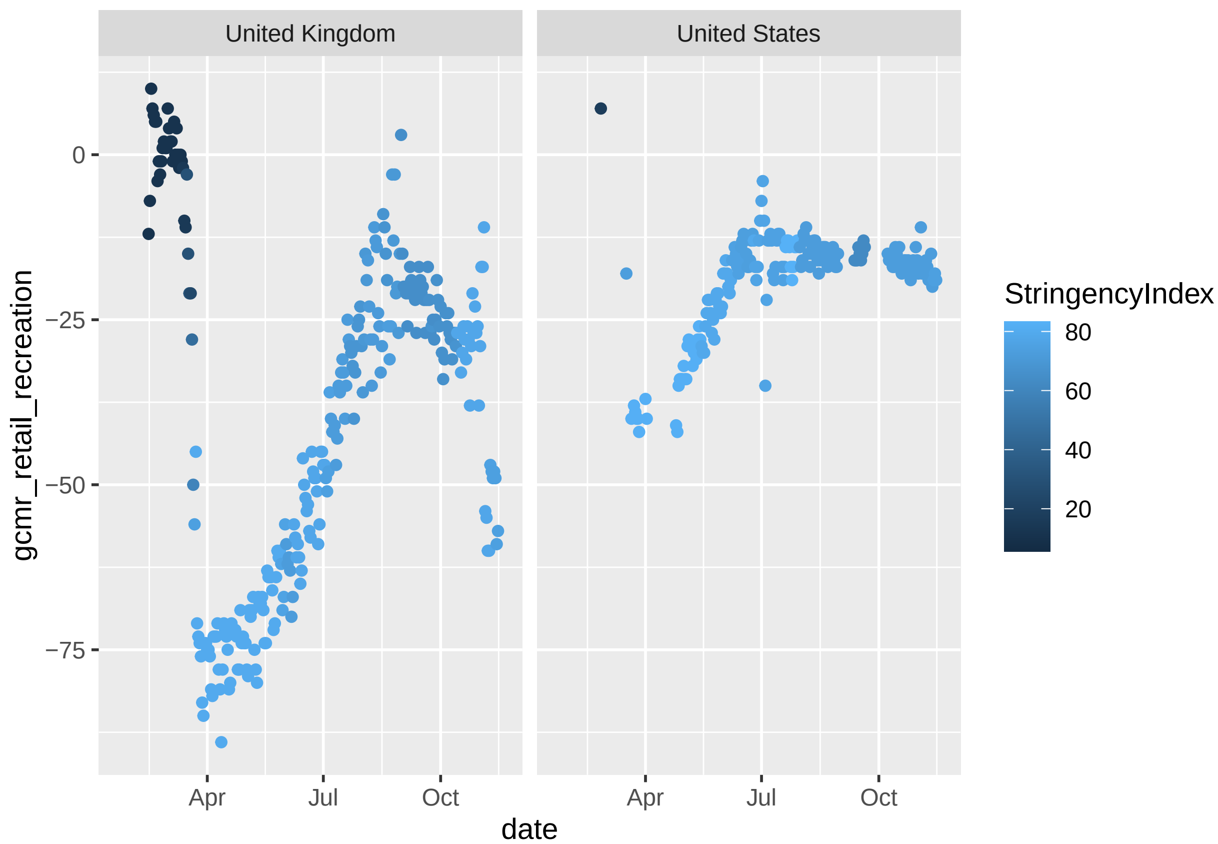
```
ggplot(aes(x=date, y=m_groc), data = EAData) + geom_line(aes(group = country),
                                                     alpha = 0.1) +
  facet_wrap(~ continent) + geom_point(alpha = 0) + geom_smooth(method = "gam")
```



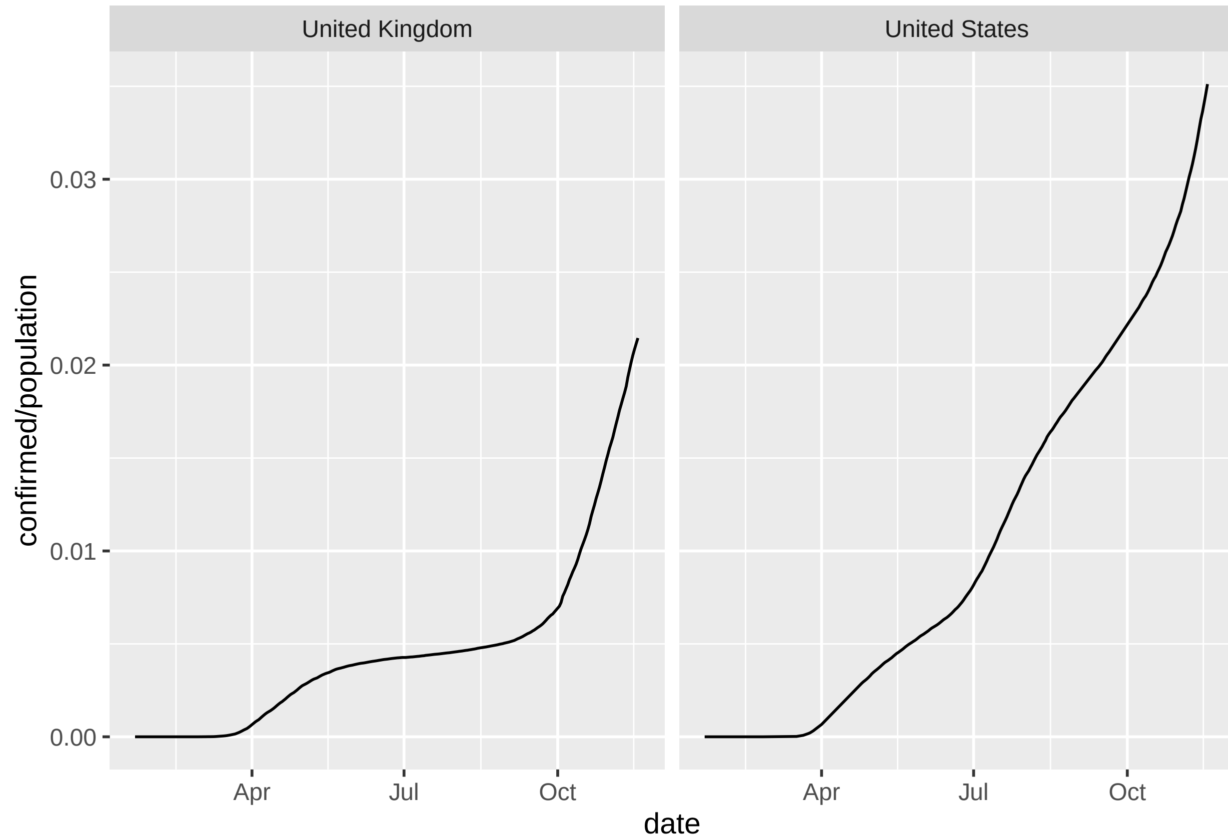
1.1.5 UK / USA Plotting

Now we are curious as to how the GCMR Retail & Recreation data changes over time depending on the stringency index in the oxford data set. So we plot two graphs for the UK and the US.

```
EDADATA <- EAData %>%
  filter(country == "United Kingdom" | country == "United States") %>%
  select(date, confirmed, soc_dist, mov_rest,
         gcmr_retail_recreation, country,
         StringencyIndex, ContainmentHealthIndex,
         population)

ggplot(EDADATA, aes(x = date, y = gcmr_retail_recreation,
                    color = StringencyIndex)) +
  geom_point() +
  facet_wrap(~ EDADATA$country)
```




```
ggplot(aes(x= date, y= confirmed/population ), data = EDADATA) +
  geom_line() + facet_wrap(~ country)
```



We also plot the graphs of their daily cases. If cases decreased, it may imply more people would attend retail and recreational areas. However, it appears that the daily cases are continuing to rise so this unlikely to be case.

Now we can see patterns in the data we want to explore the graphs between Countries with varying levels in the Stringency Index. So we take the mean stringency index for each country and plot the quartiles to see if there is any variation.

```
max_ <- function(...) max(..., na.rm = T)
mean_ <- function(...) mean(..., na.rm = T)

MeanStringencyData <- EAData %>% group_by(country) %>%
  summarise(mean_stringency = mean(StringencyIndex),
            total_confirmed = max_(confirmed))

EAData <- left_join(x = EAData, y = MeanStringencyData)

quantile(EAData$mean_stringency, na.rm = TRUE)
```

```
##        0%      25%      50%      75%     100%
## 10.40232 45.26671 54.61179 63.29863 77.17713
```

```
HStringency <- EAData %>% filter(mean_stringency > 64.41)

MStringency <- EAData %>% filter(46.27 < mean_stringency,
                                 mean_stringency < 64.41)
```

```
LStringency <- EAData %>% filter(mean_stringency < 46.27)
```

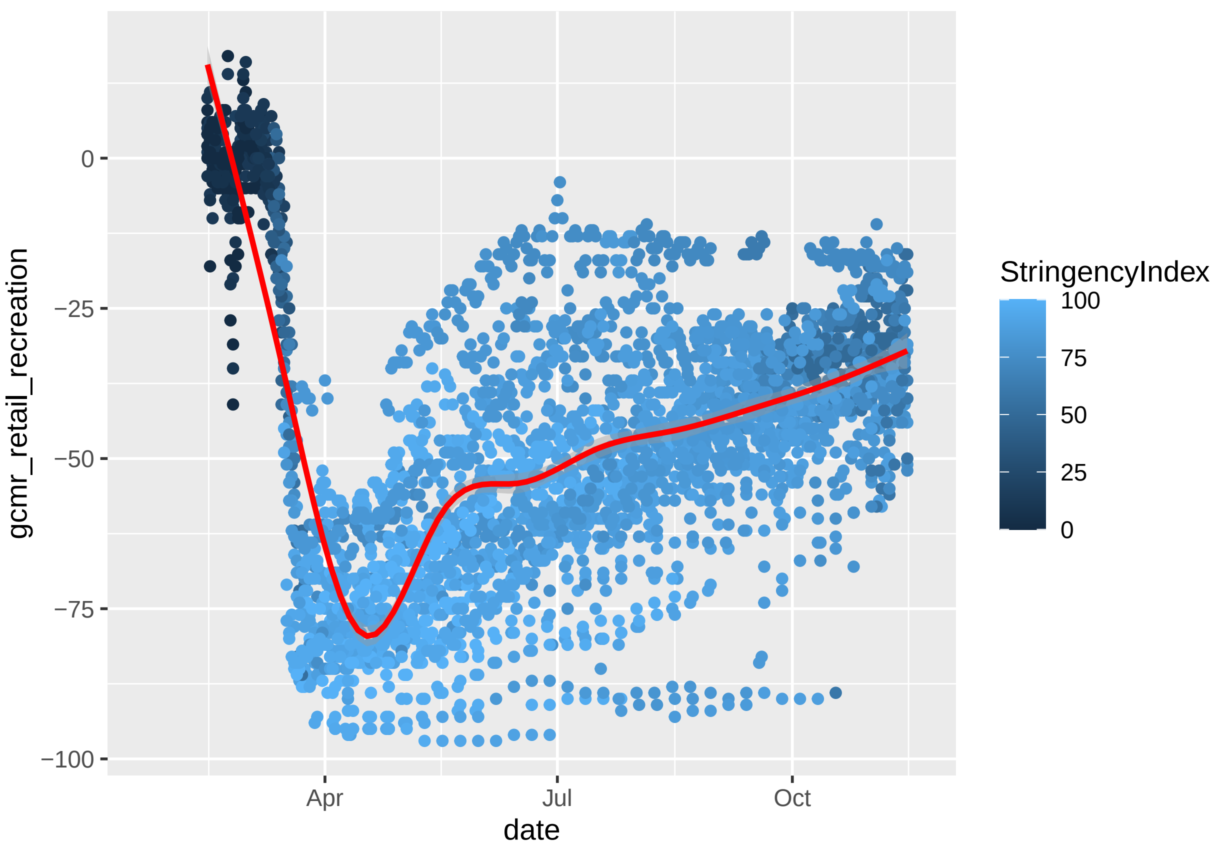
1.2 Focused analysis

1.2.1 Retail and recreation

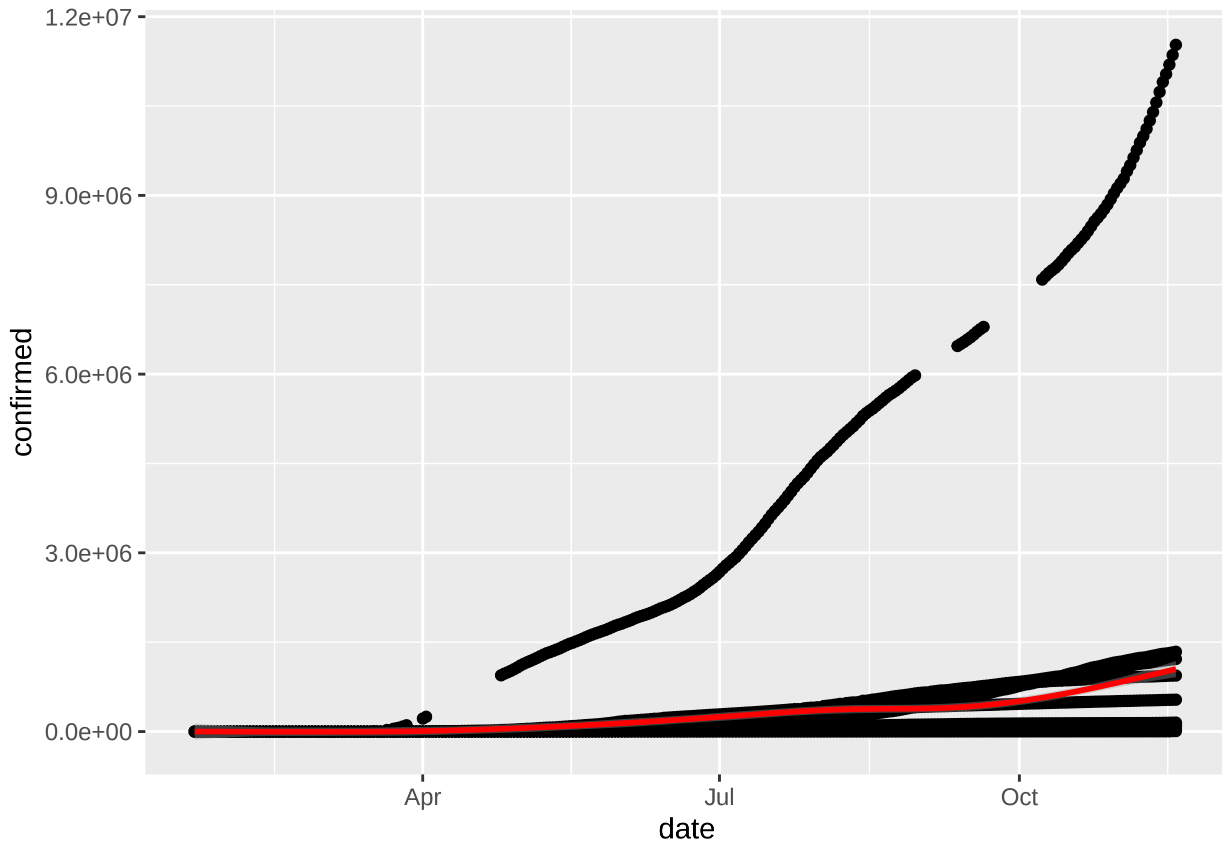
1.2.1.1 Global stringency quartiles

Upper:

```
ggplot(aes(x = date, y = gcmr_retail_recreation, color = StringencyIndex),
       data = HStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

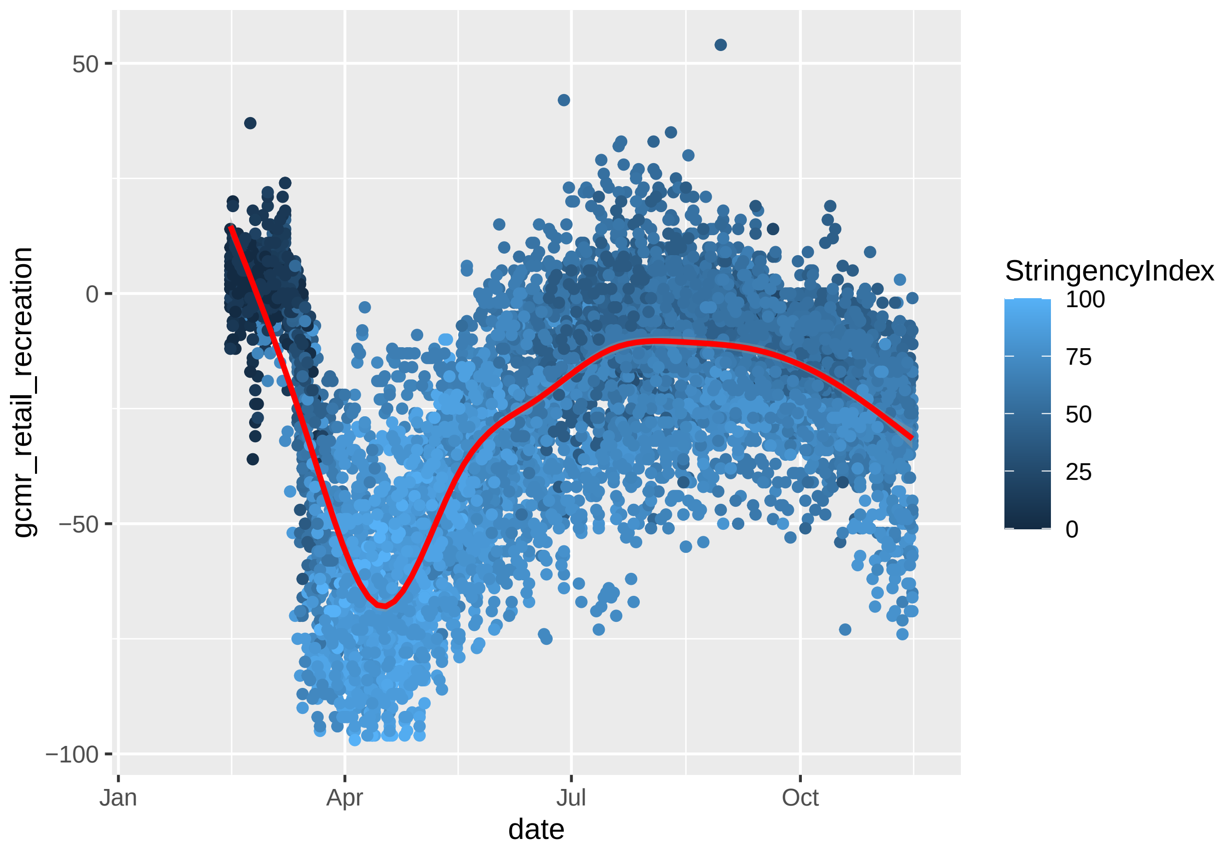



```
ggplot(aes(x= date, y= confirmed), data = HStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

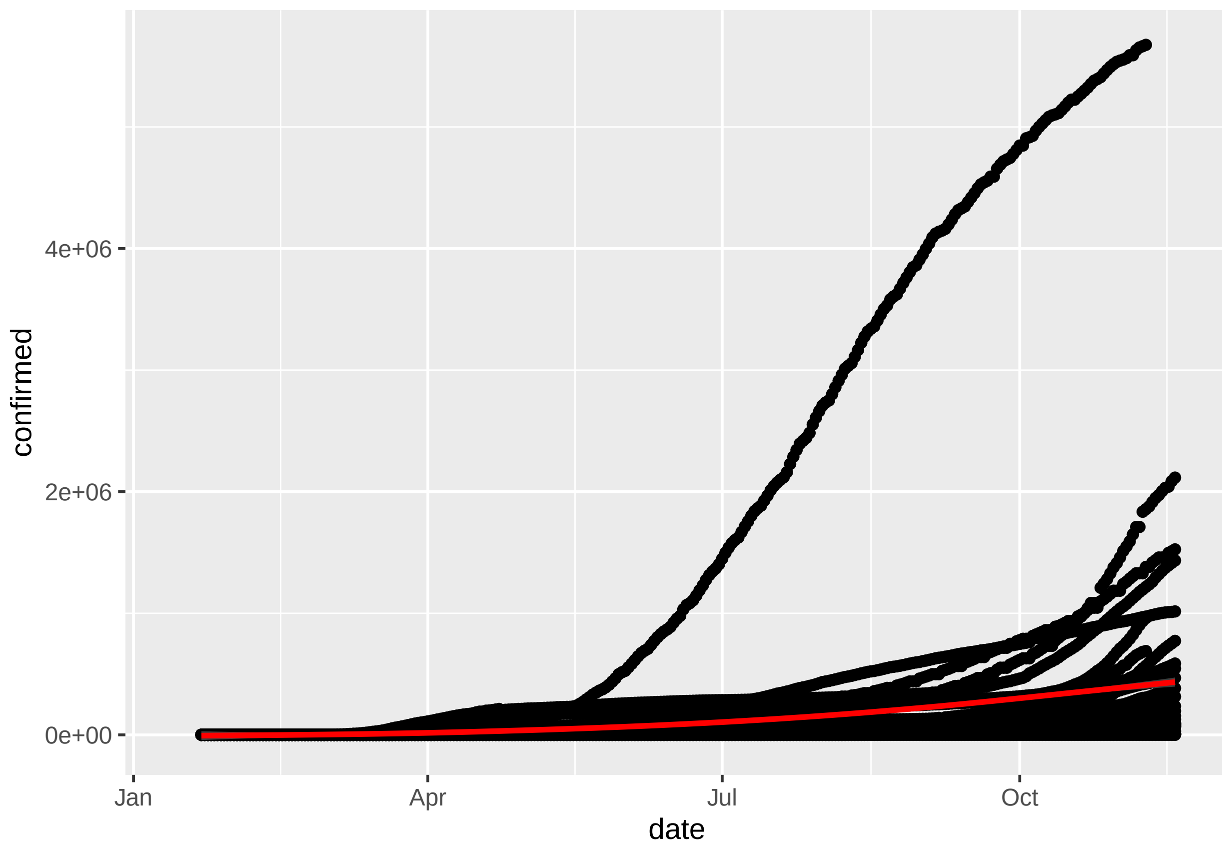


Middle:

```
ggplot(aes(x = date, y = gcmr_retail_recreation, color = StringencyIndex),
       data = MStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

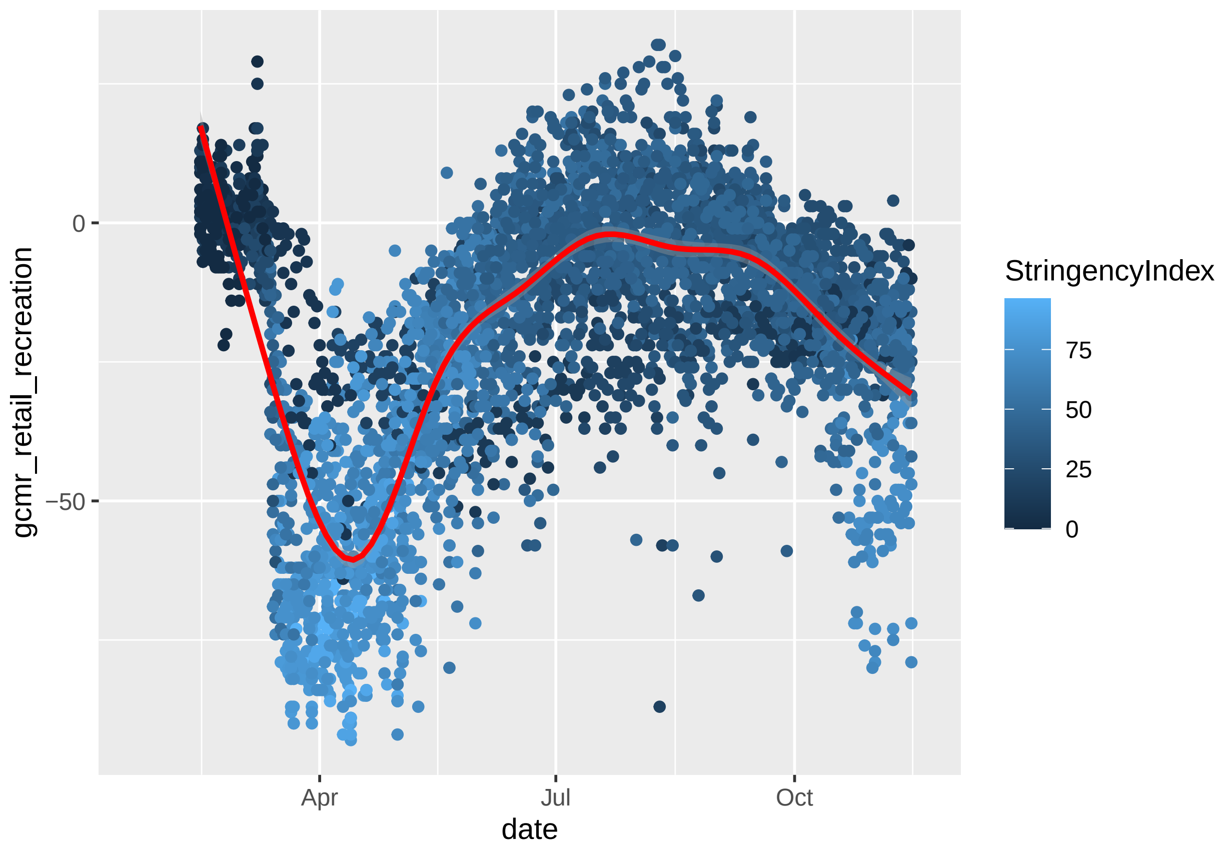



```
ggplot(aes(x= date, y= confirmed), data = MStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

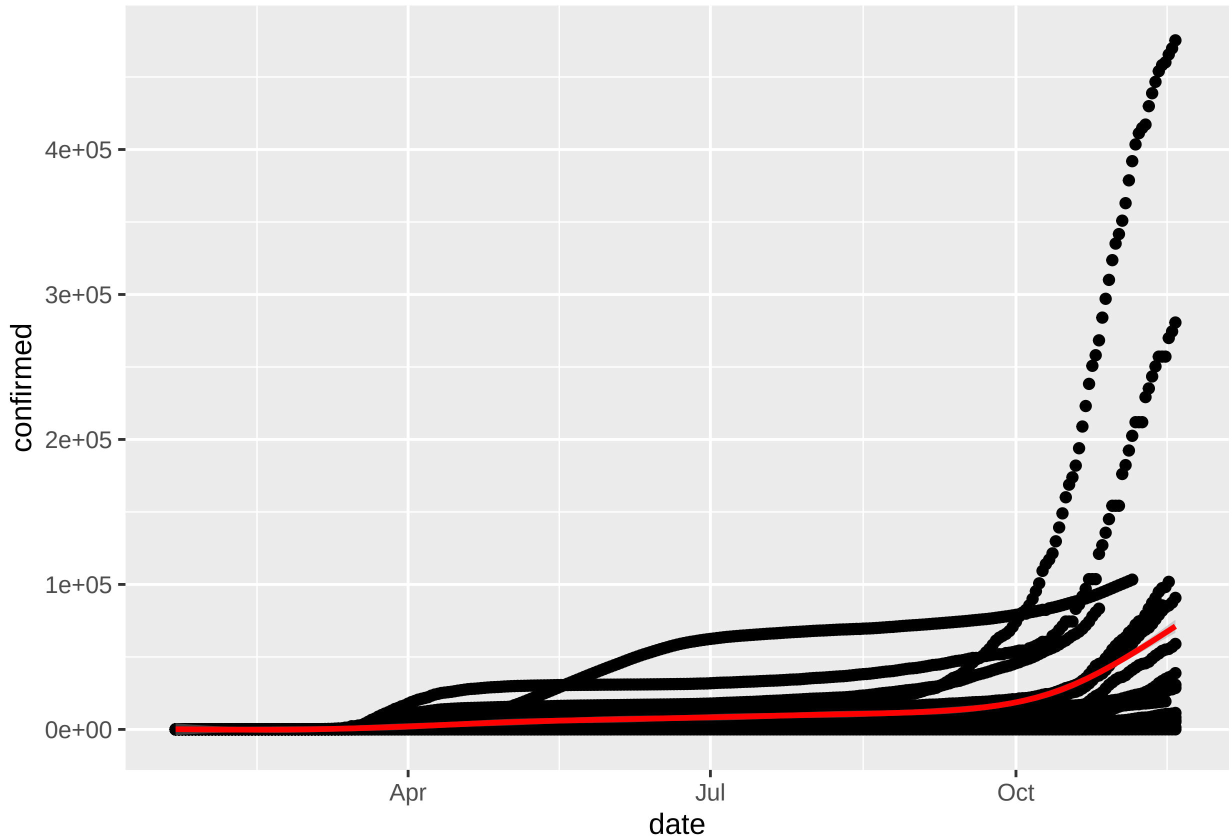


Lower:

```
ggplot(aes(x = date, y = gcmr_retail_recreation, color = StringencyIndex),
       data = LStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

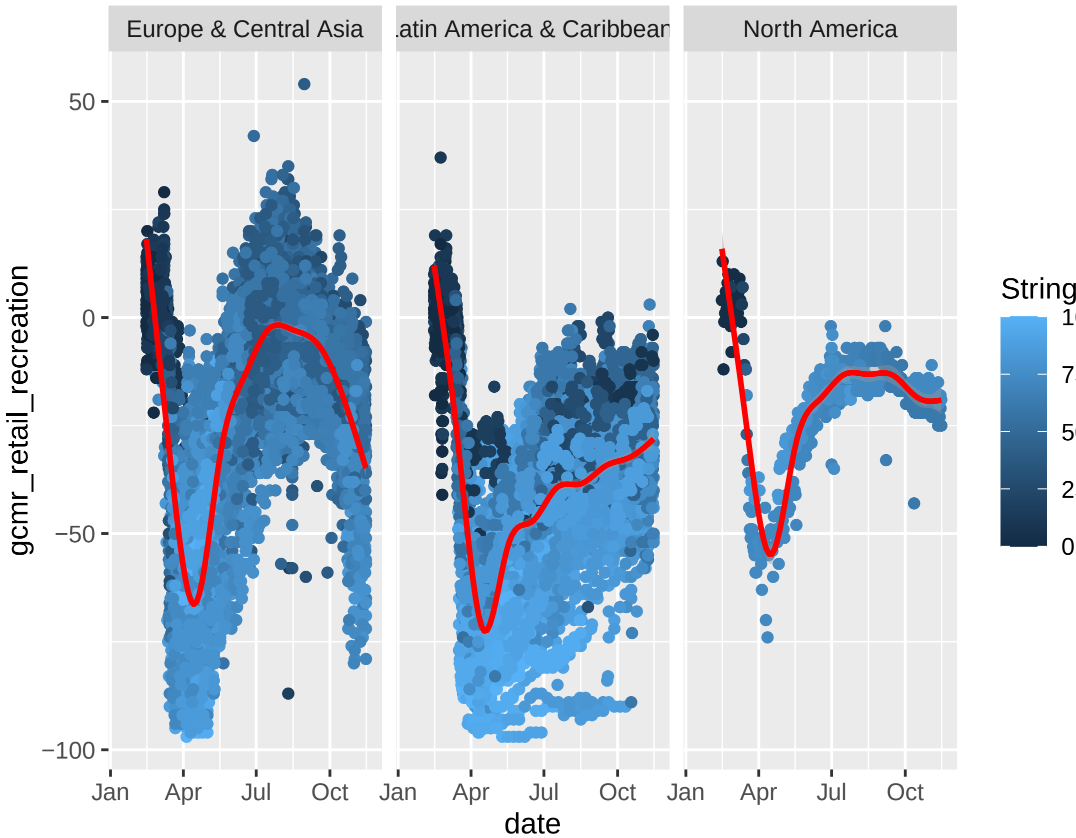



```
ggplot(aes(x= date, y= confirmed), data = LStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```



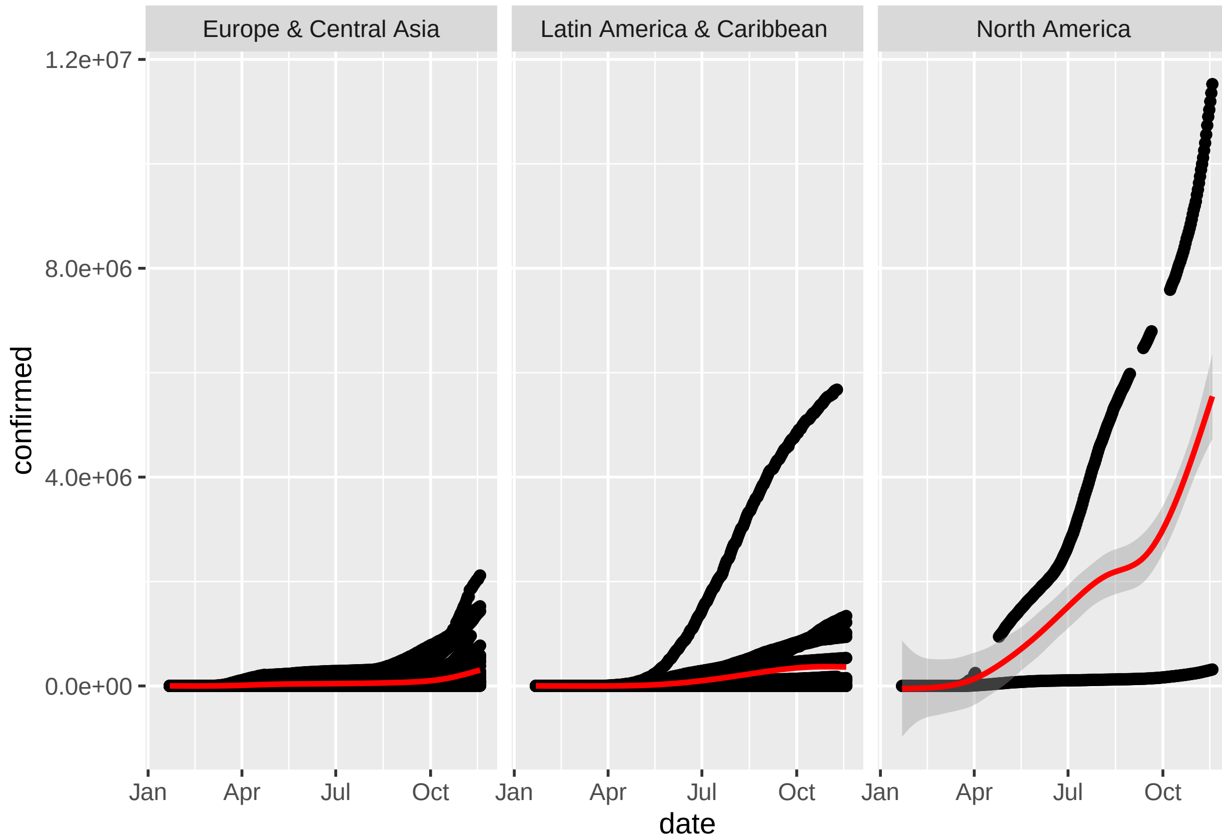
As we are looking to compare how the effects may differ between countries we next group countries by continent visualise trends.

```
ggplot(aes(x = date, y = gcmr_retail_recreation, color = StringencyIndex),
       data = EAData) +
  geom_point() +
  facet_wrap(~ region) +
  geom_smooth(colour = "red", method = "gam", group = 1)
```



1.2.1.2 Regional groups

```
ggplot(aes(x= date, y= confirmed), data = EAData) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1) + facet_wrap(~ region)
```



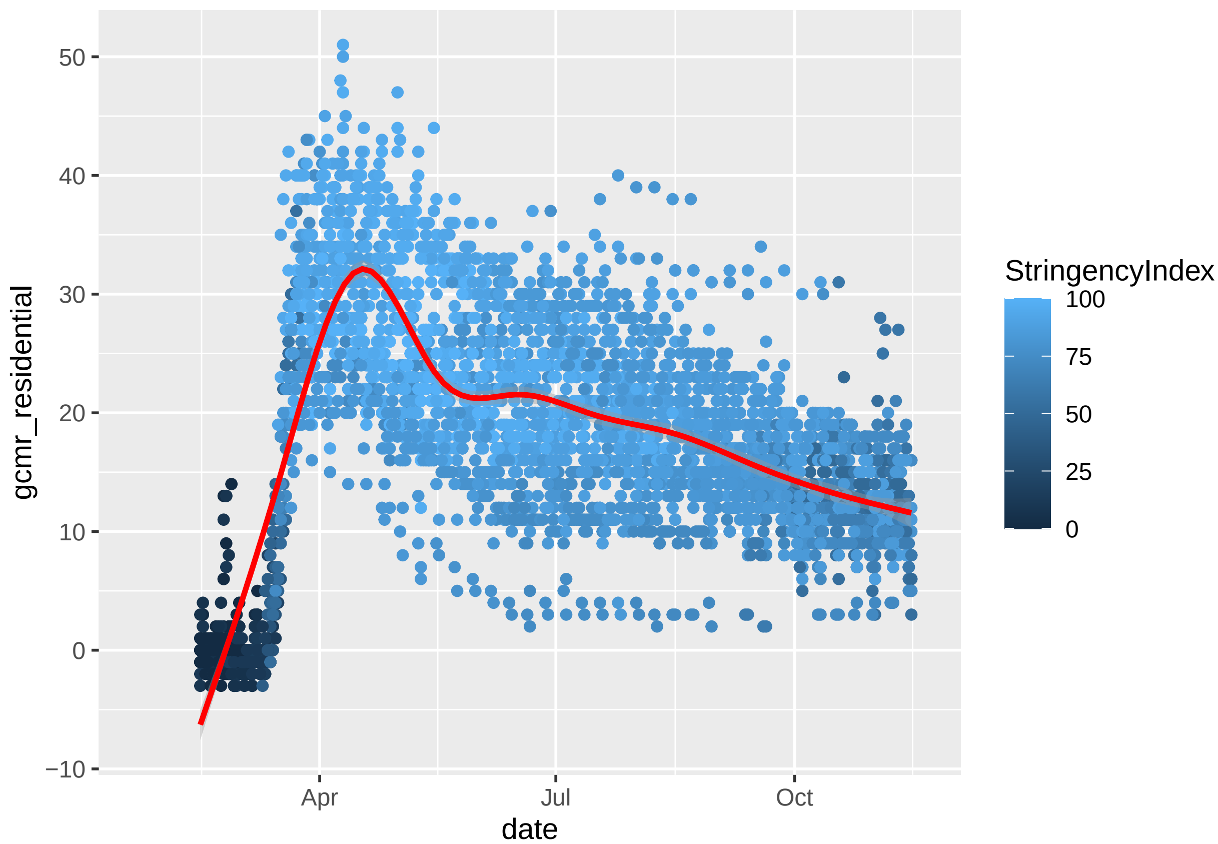
We now repeat our analysis for the other GCMR variables in the tidycovid19 dataset

1.2.2 Residential

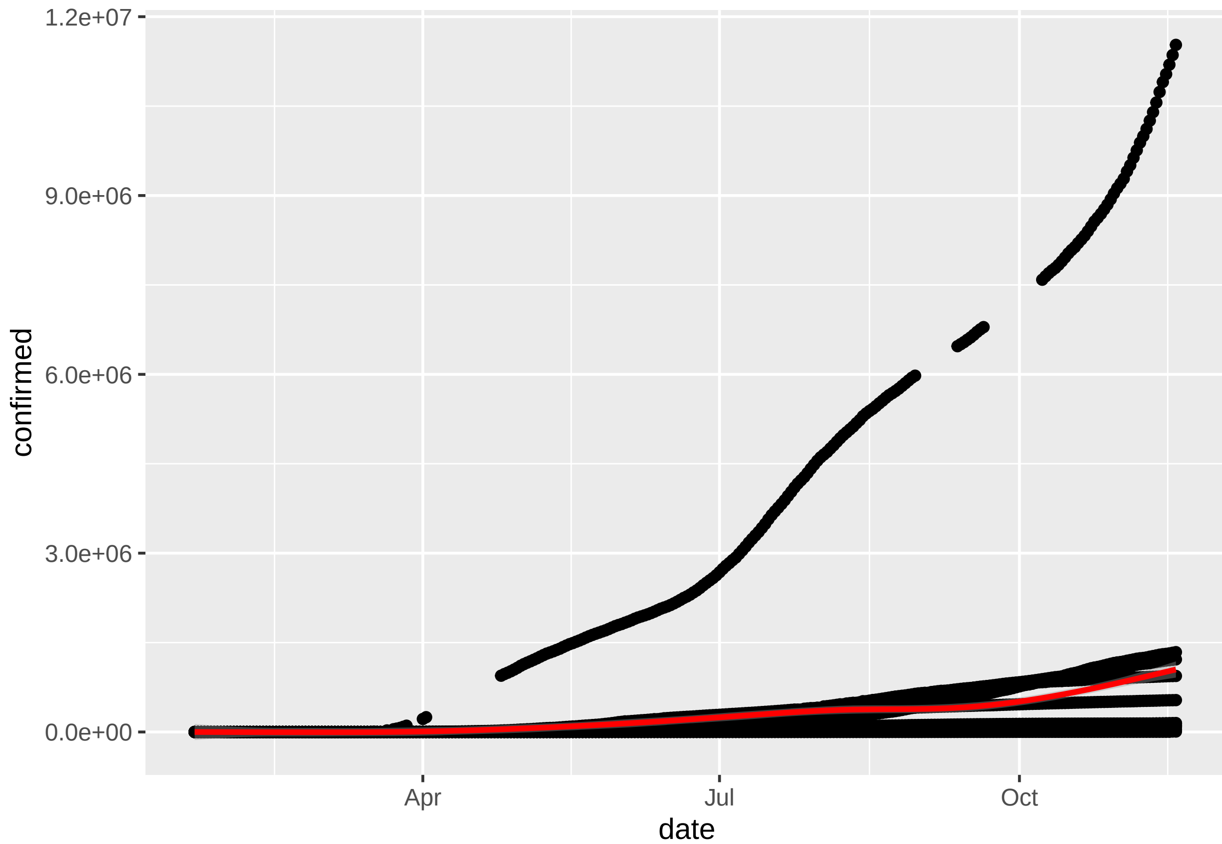
1.2.2.1 Global stringency quartiles

Upper:

```
ggplot(aes(x = date, y = gcmr_residential, color = StringencyIndex ),
       data = HStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

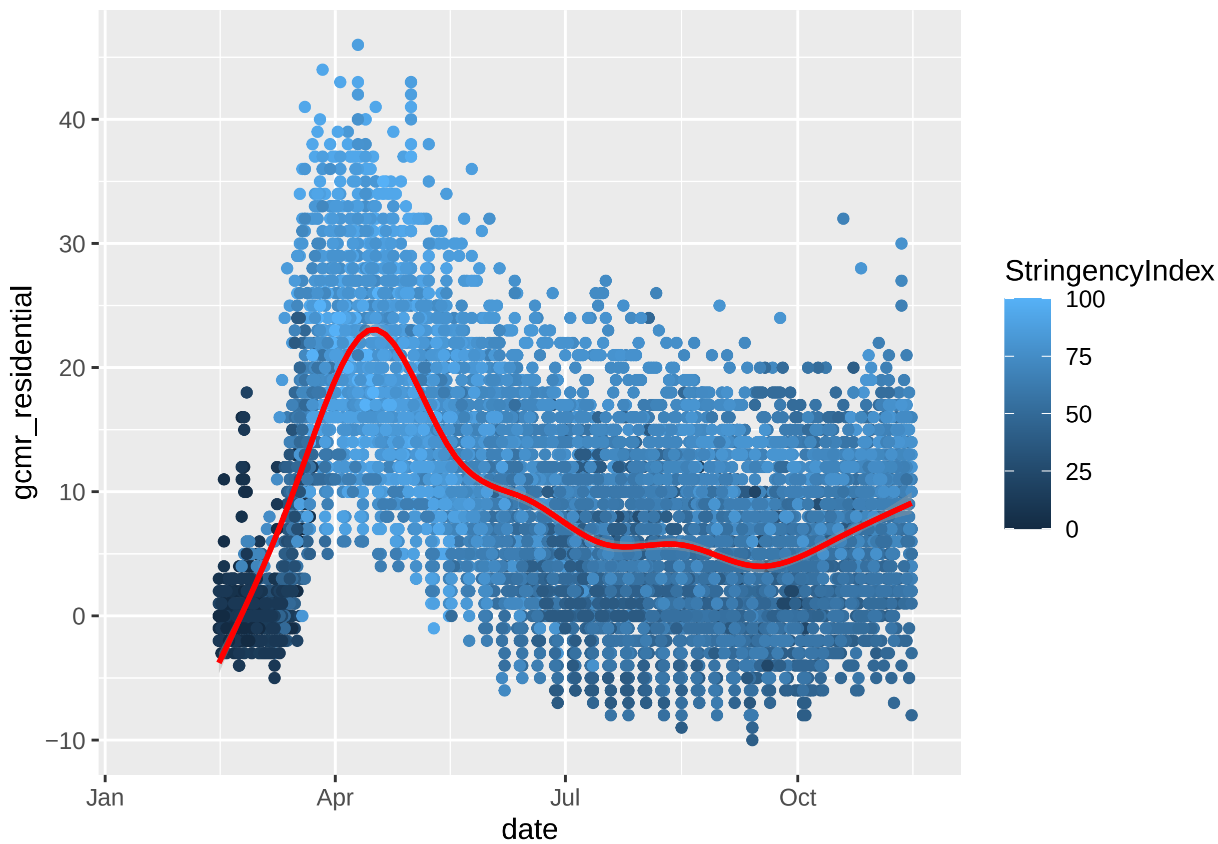



```r
ggplot(aes(x= date, y= confirmed), data = HStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

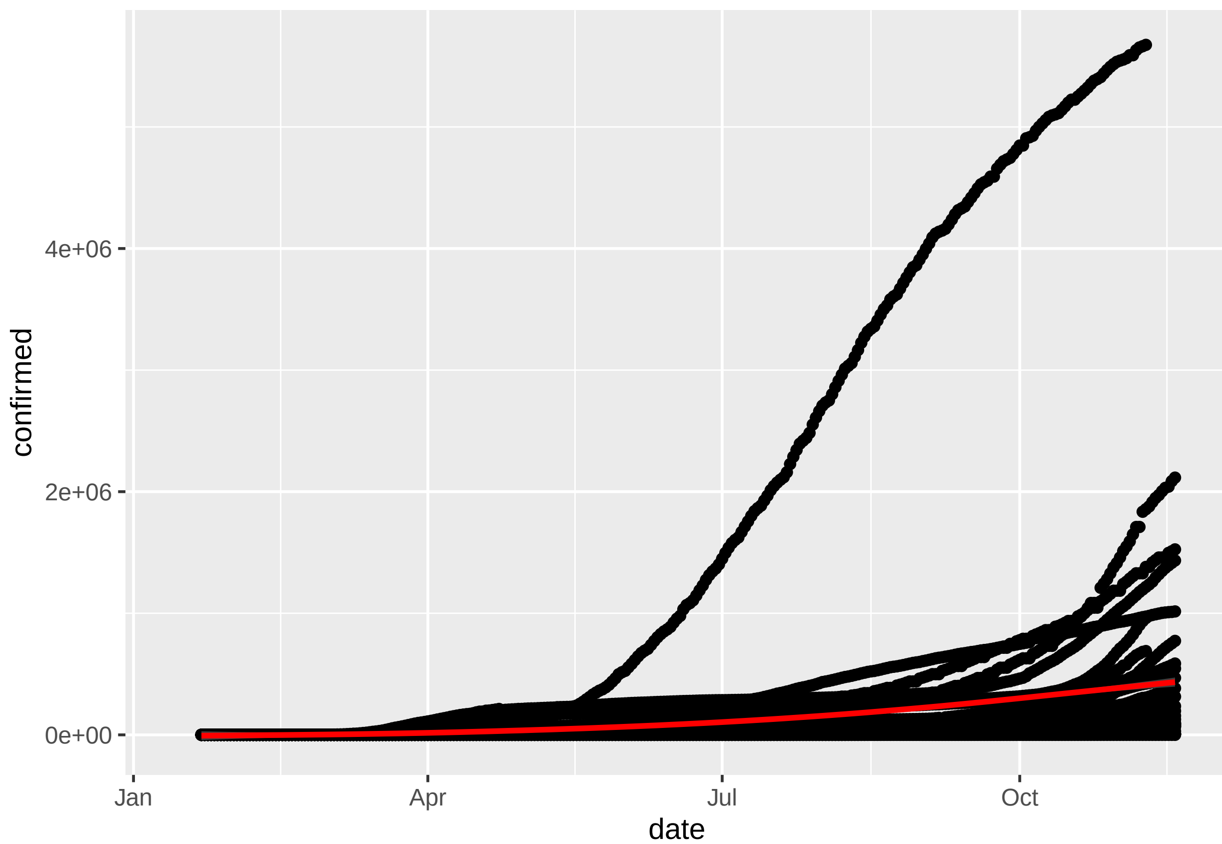


Middle:

```
ggplot(aes(x = date, y = gcmr_residential, color = StringencyIndex ),
       data = MStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

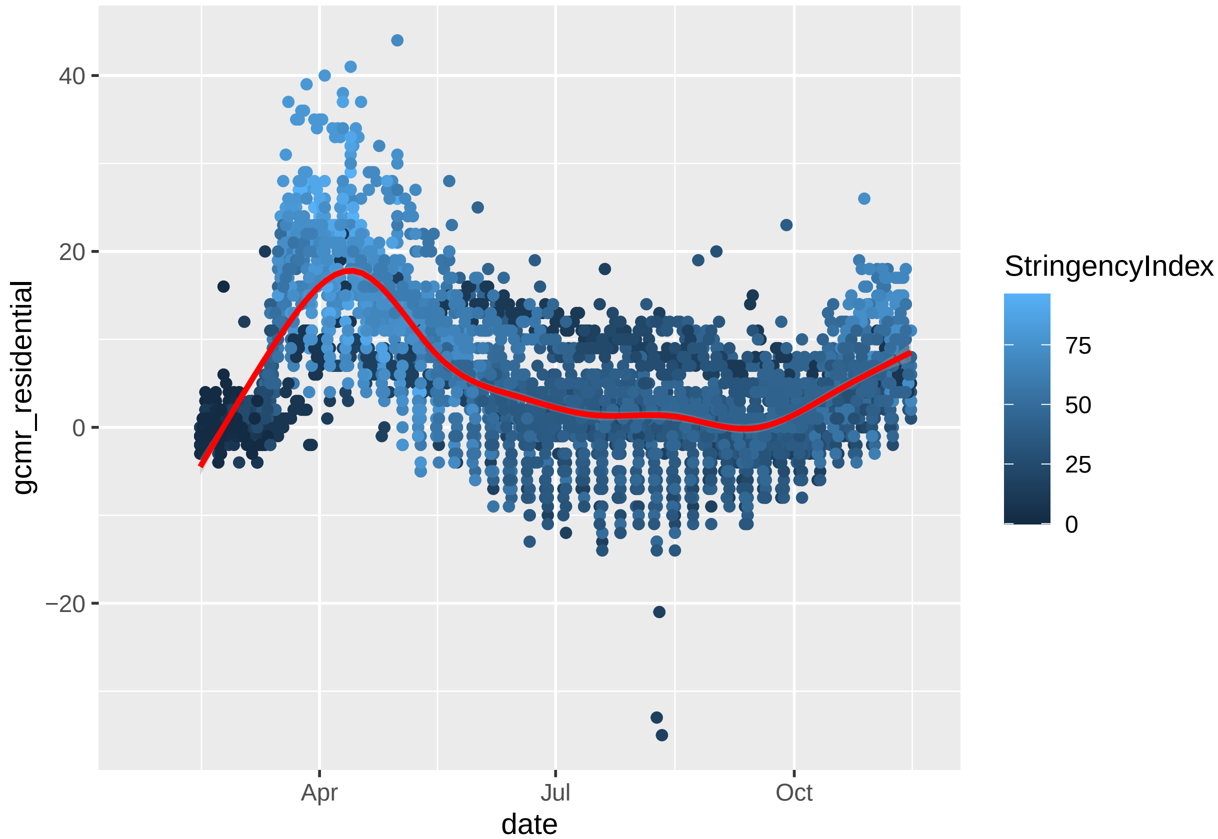



```r
ggplot(aes(x= date, y= confirmed), data = MStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

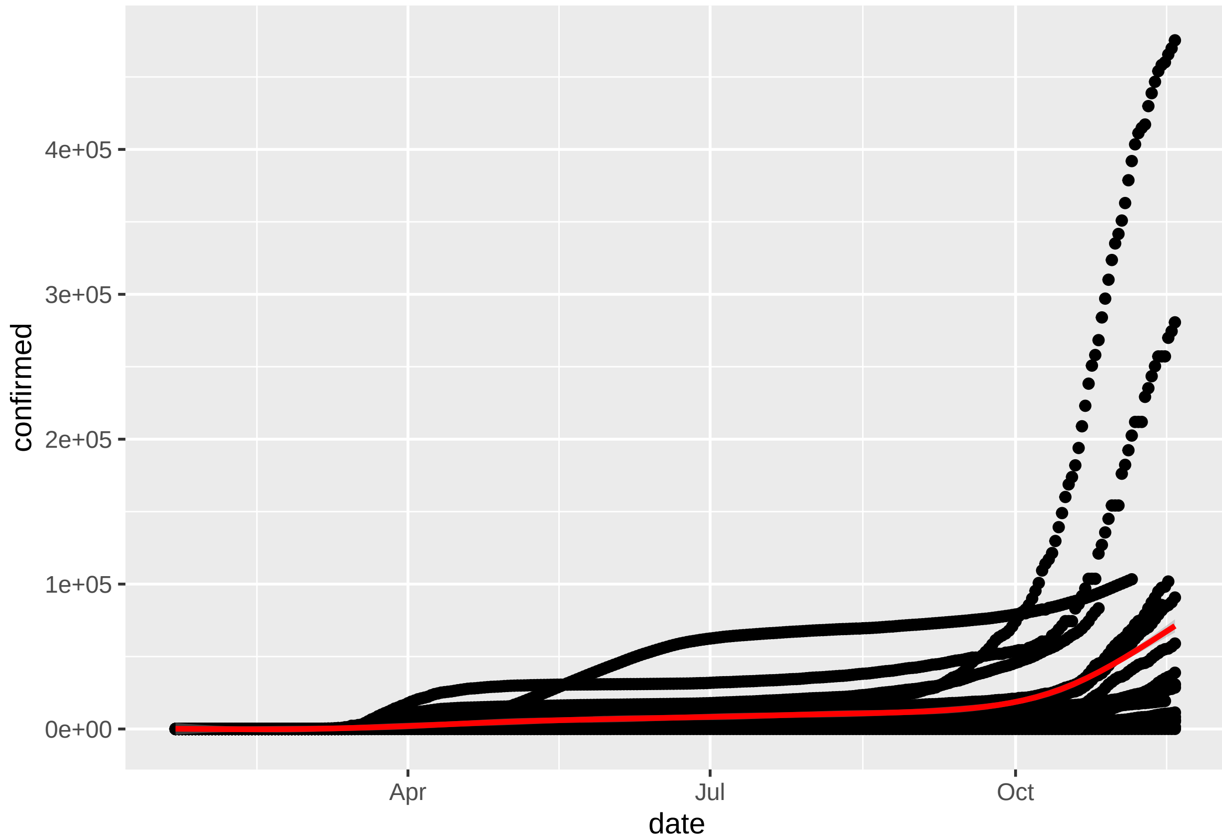


Lower:

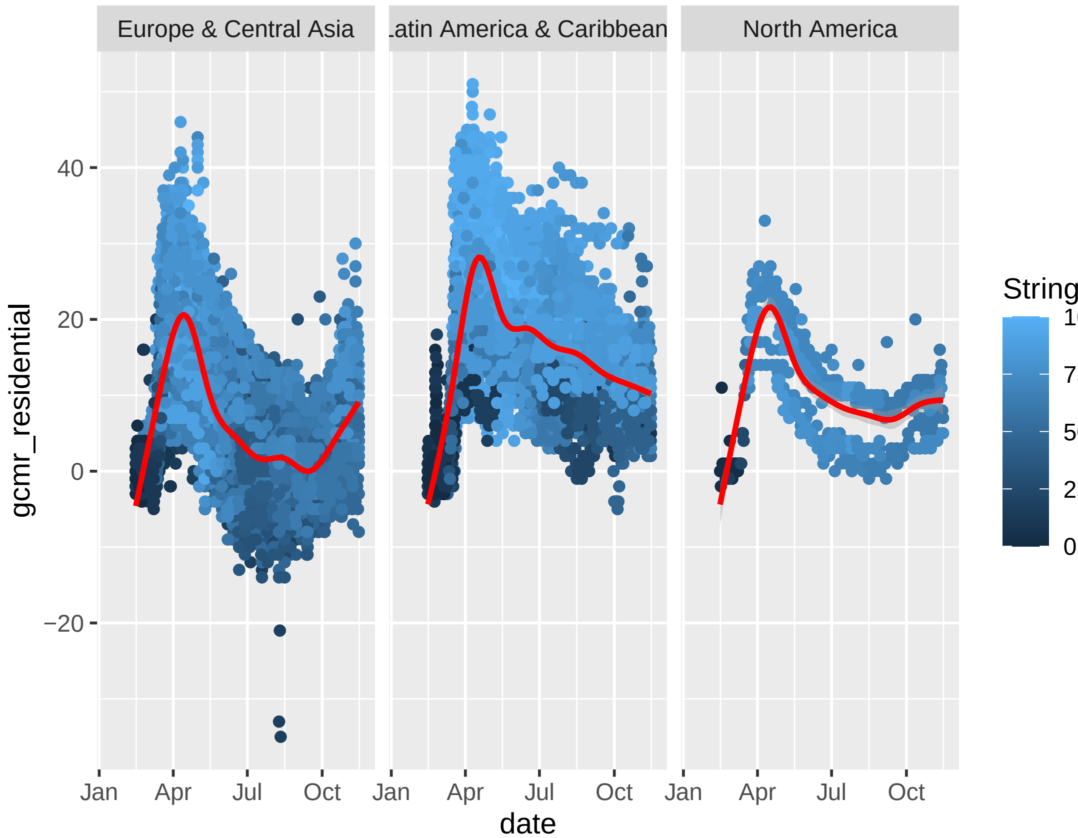
```r
ggplot(aes(x = date, y = gcmr_residential, color = StringencyIndex ),
       data = LStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```




```r
ggplot(aes(x= date, y= confirmed), data = LStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

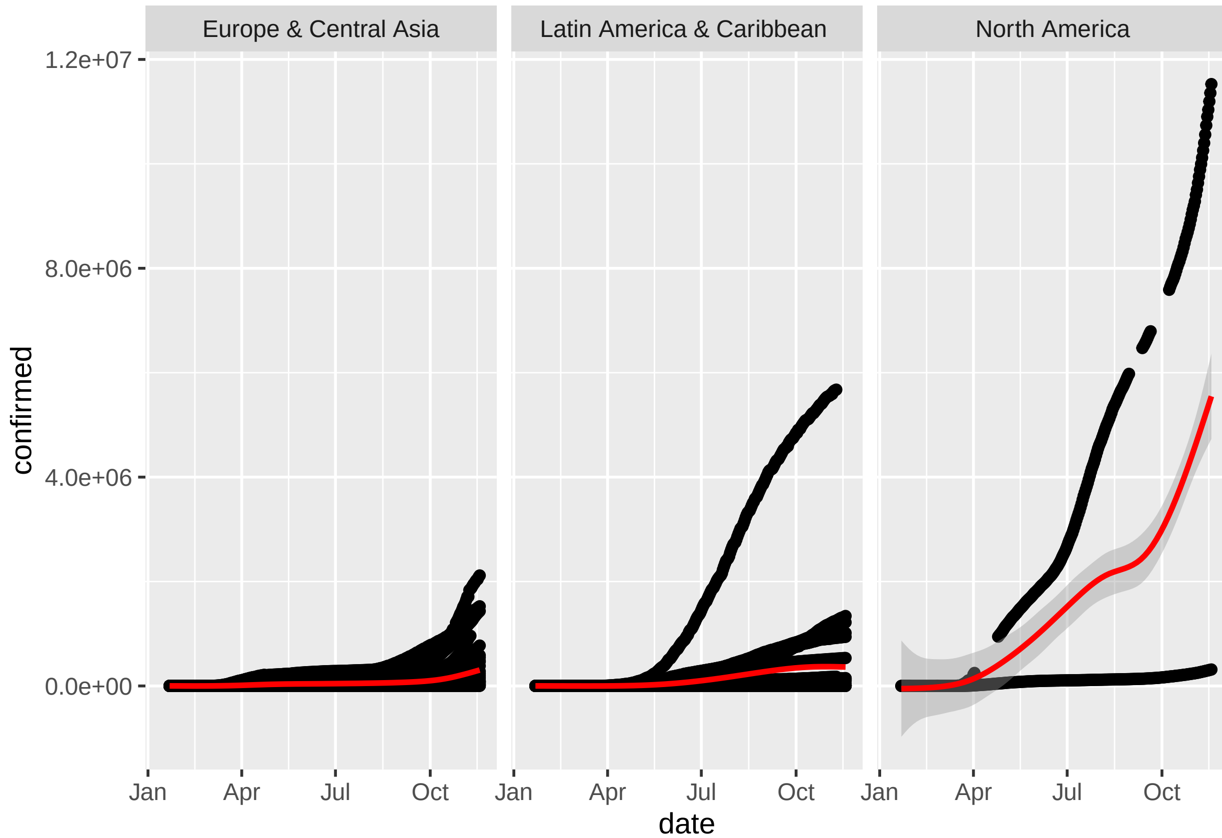



```
ggplot(aes(x = date, y = gcmr_residential, color = StringencyIndex ),
       data = EAData) +
  geom_point() + facet_wrap(~ region) +
  geom_smooth(colour = "red", method = "gam", group = 1)
```



1.2.2.2 Regional groups

```
ggplot(aes(x= date, y= confirmed), data = EAData) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1) + facet_wrap(~ region)
```

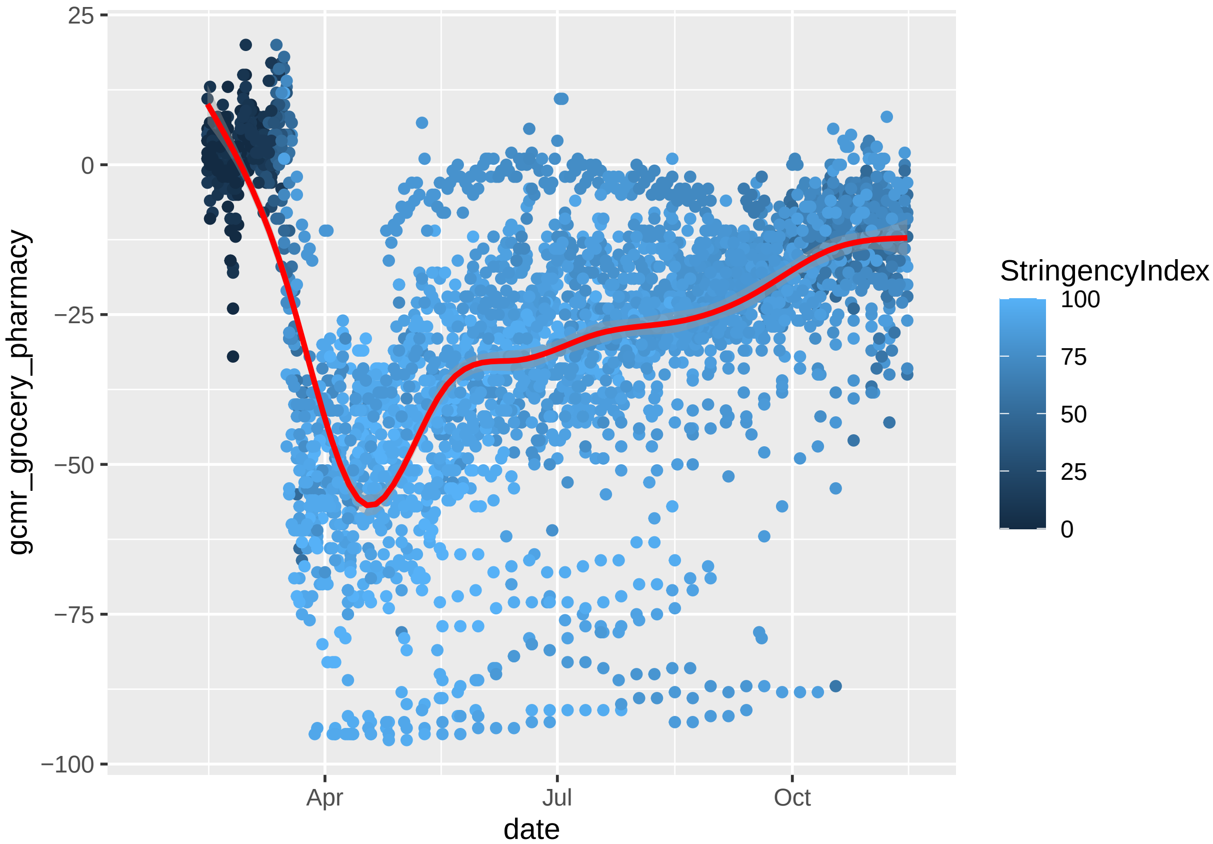


1.2.3 Grocery and pharmacy

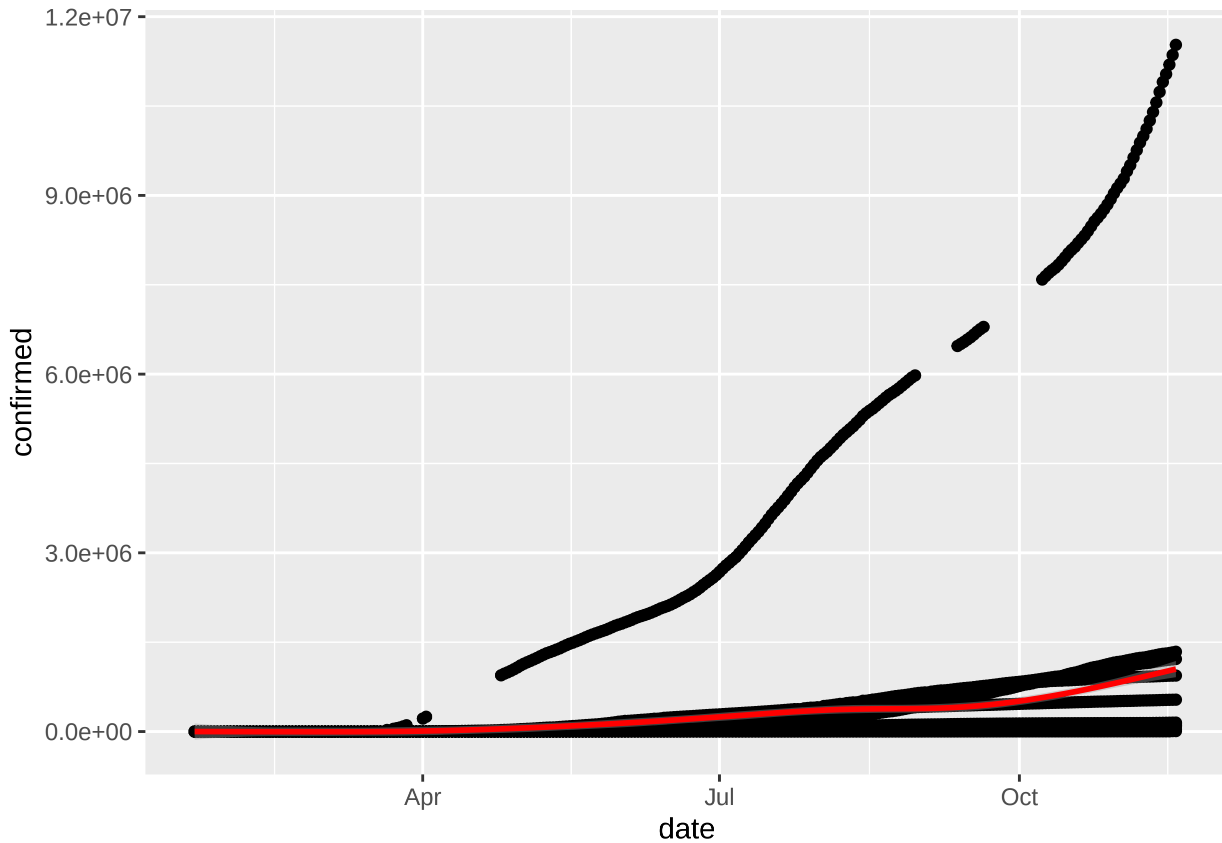
1.2.3.1 Global stringency quartiles

Upper:

```
ggplot(aes(x = date, y = gcmr_grocery_pharmacy, color = StringencyIndex ),
       data = HStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

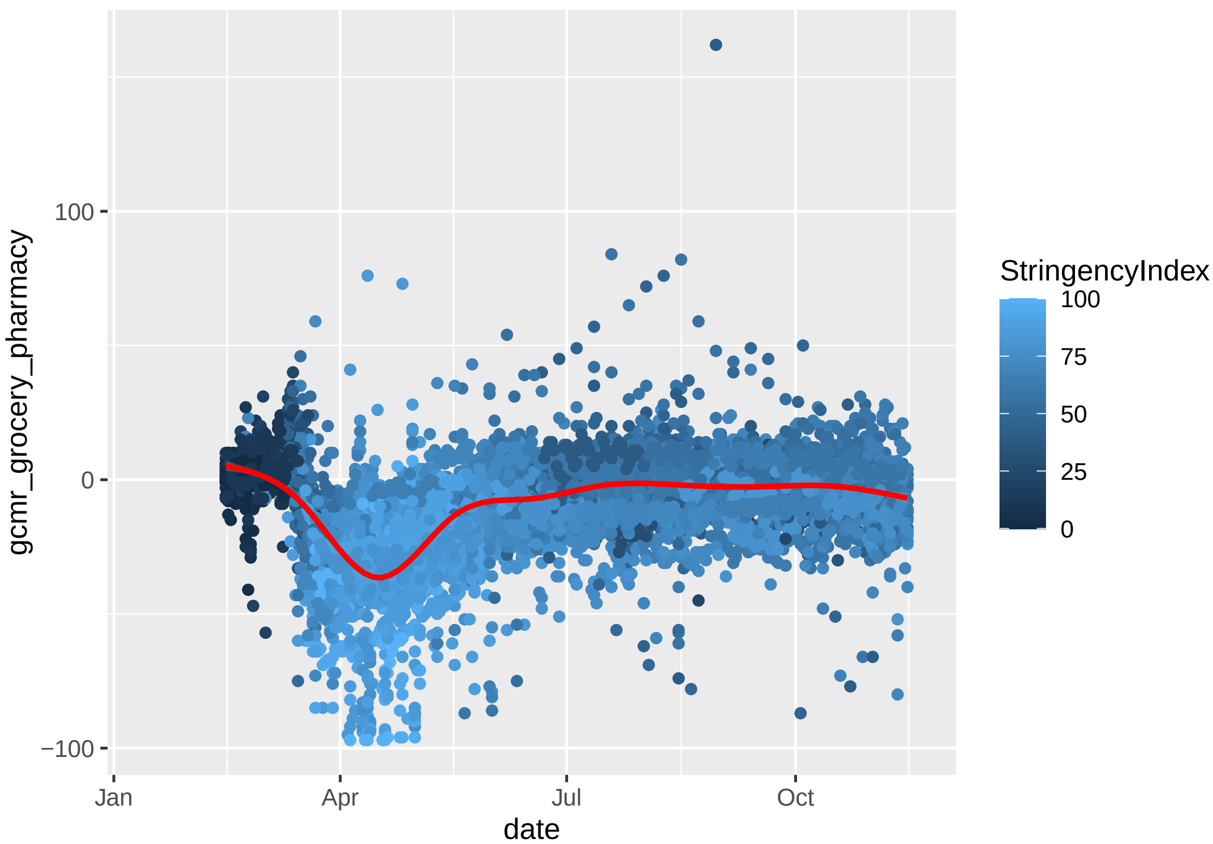



```r
ggplot(aes(x= date, y= confirmed), data = HStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

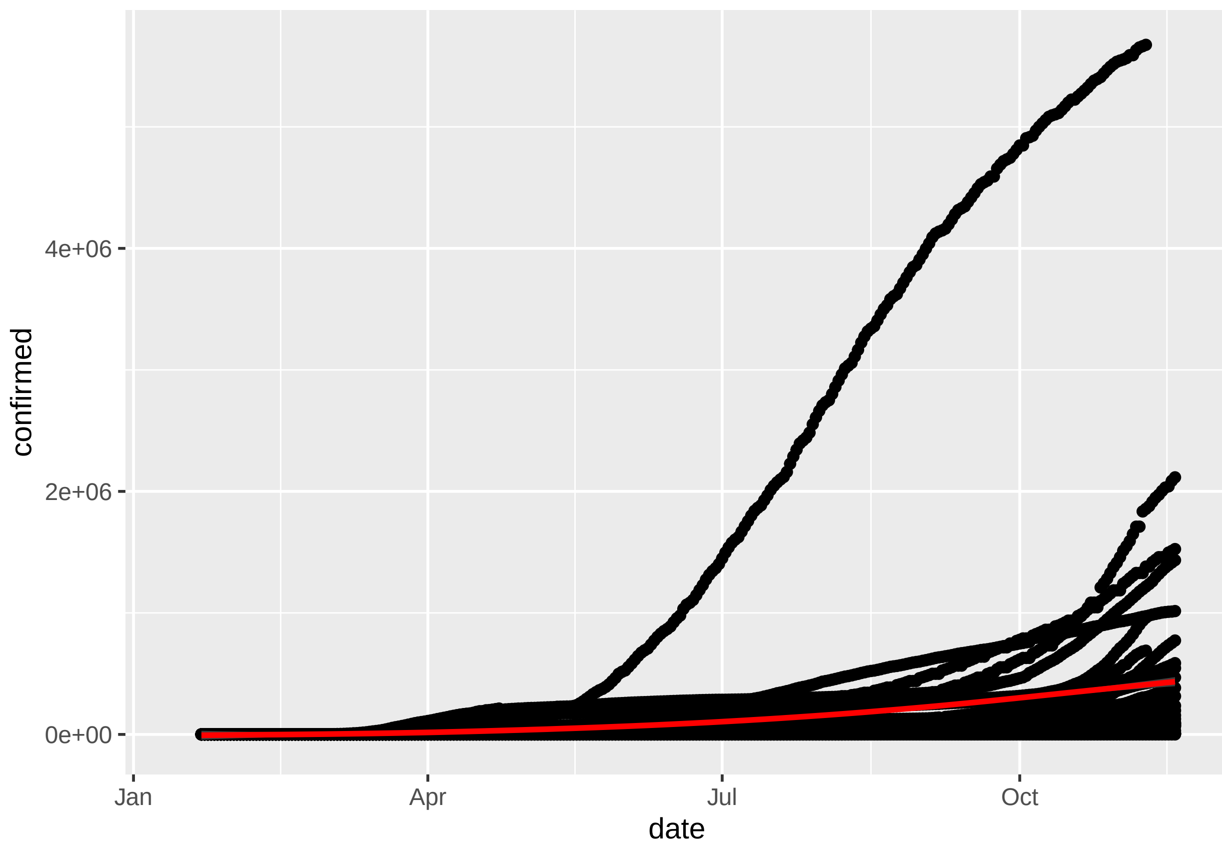


Middle:

```
ggplot(aes(x = date, y = gcmr_grocery_pharmacy, color = StringencyIndex ),
       data = MStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

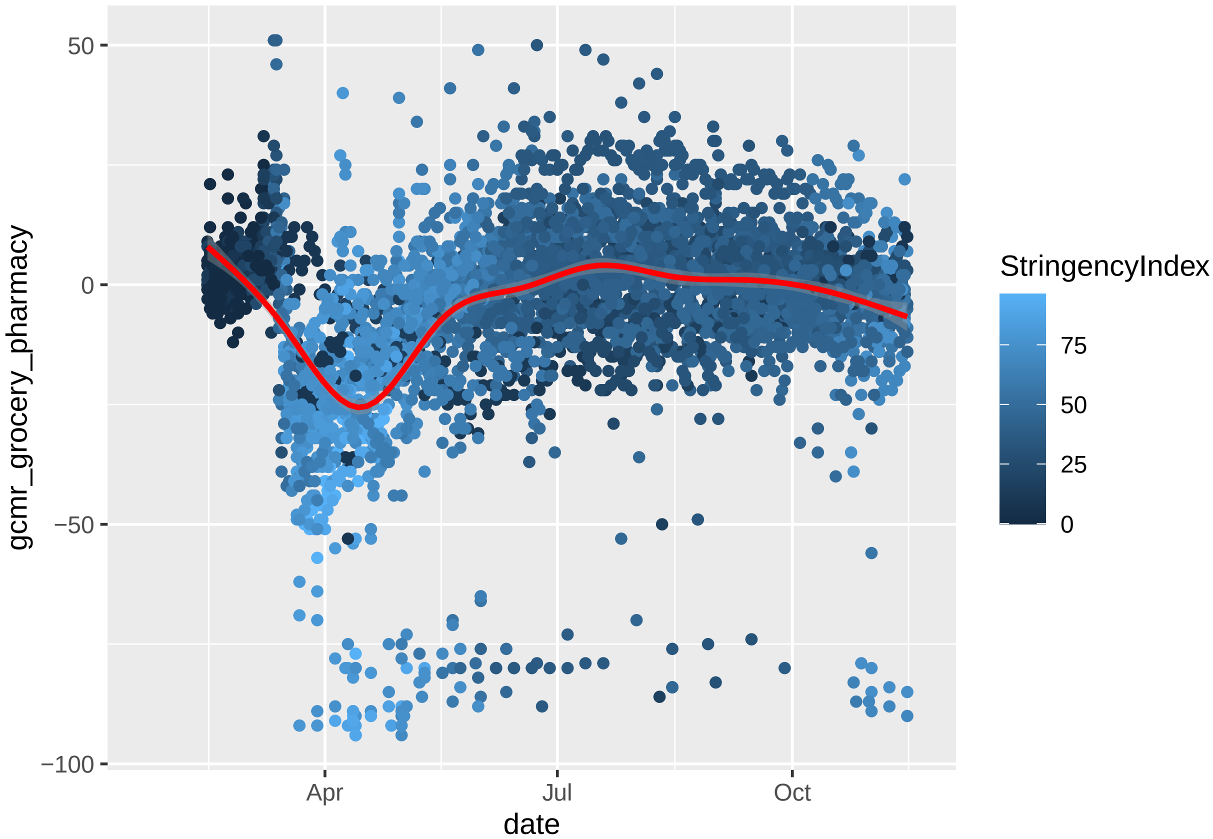



```r
ggplot(aes(x= date, y= confirmed), data = MStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

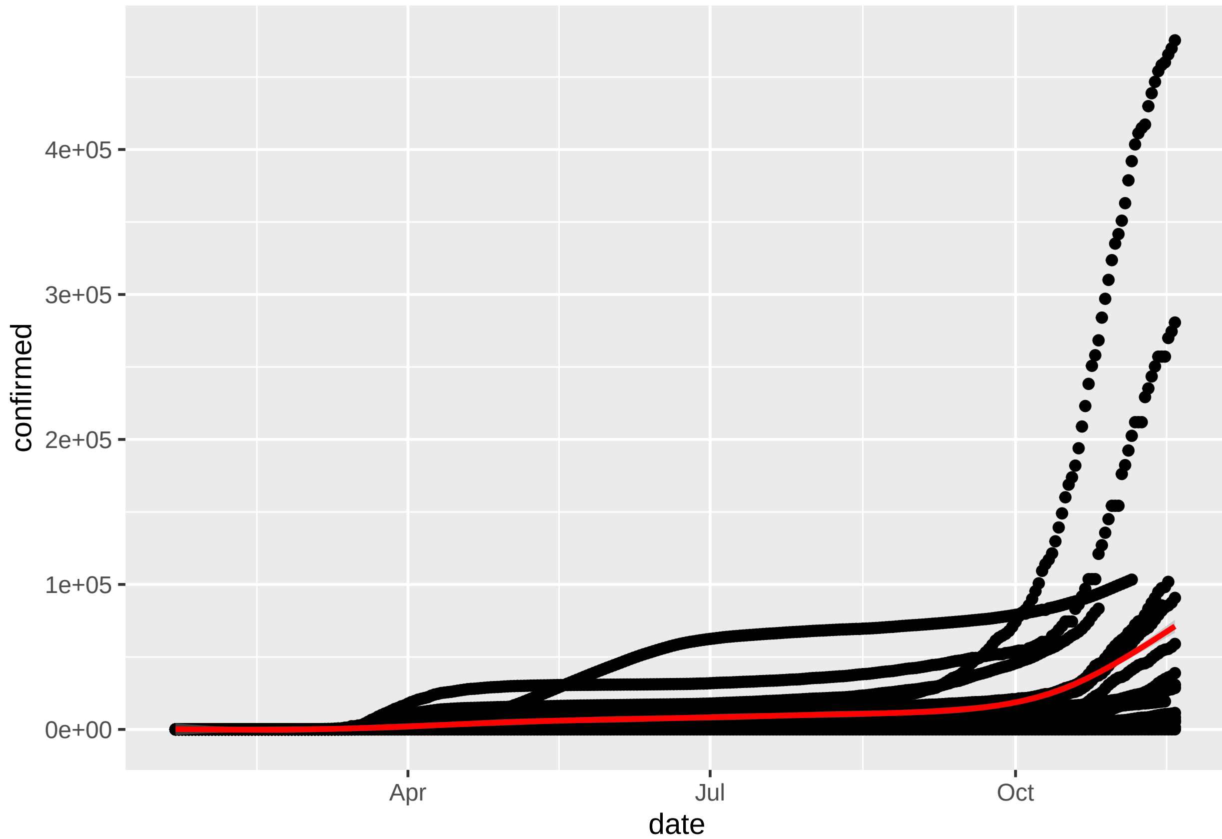


Lower:

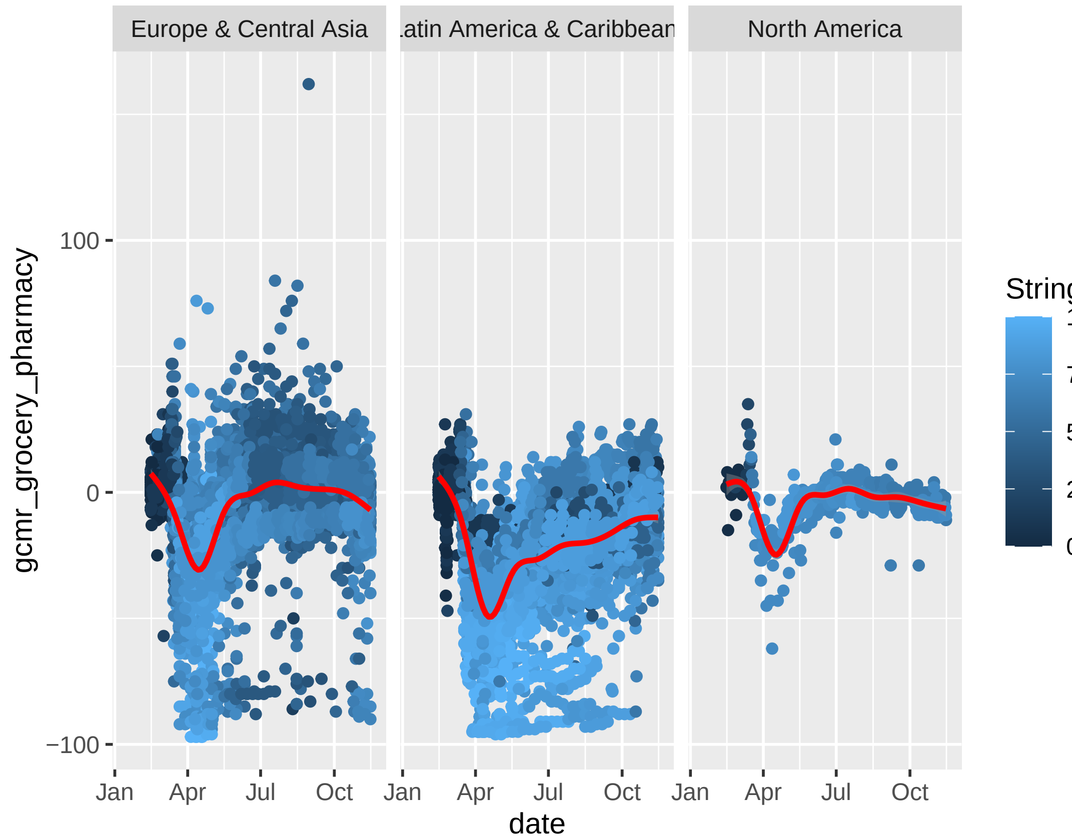
```
ggplot(aes(x = date, y = gcmr_grocery_pharmacy, color = StringencyIndex ),
       data = LStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```




```
ggplot(aes(x= date, y= confirmed), data = LStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

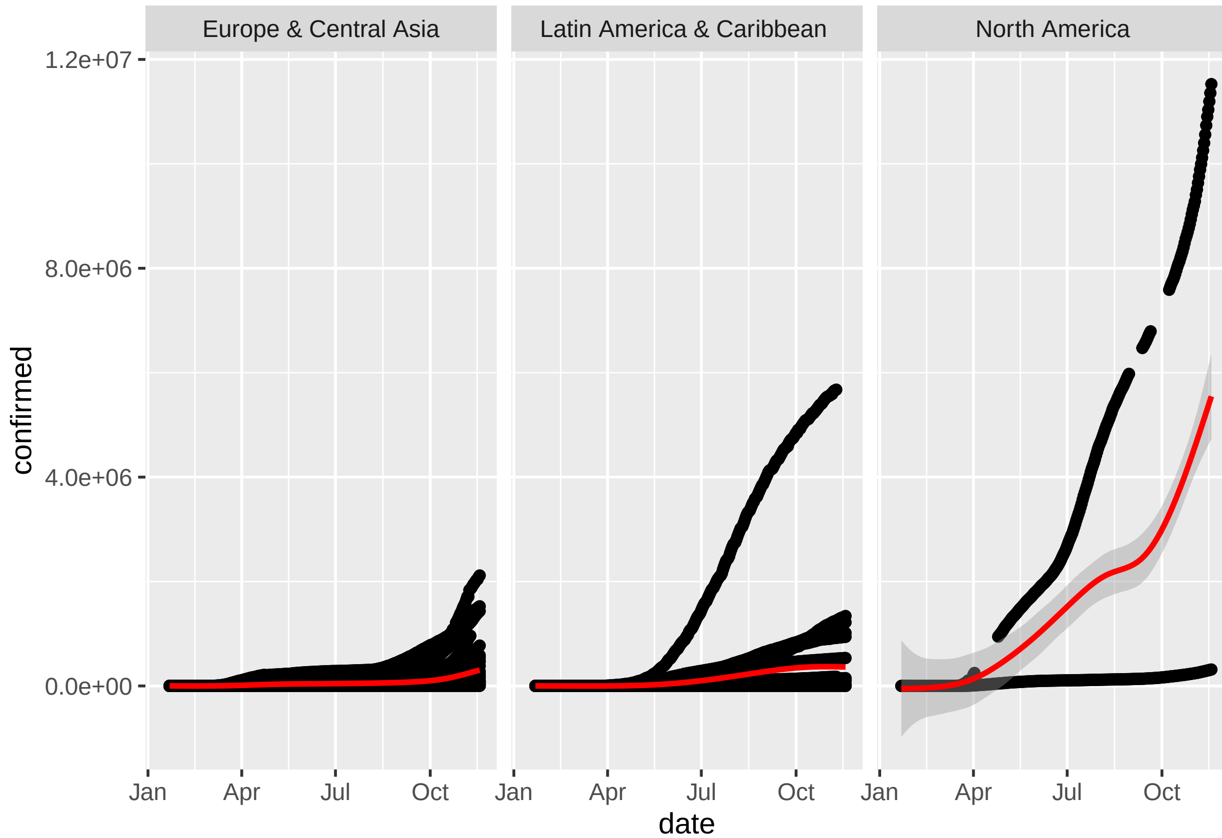



```
ggplot(aes(x = date, y = gcmr_grocery_pharmacy, color = StringencyIndex ),
       data = EAData) +
  geom_point()+ facet_wrap(~ region) +
  geom_smooth(colour = "red", method = "gam", group = 1)
```



1.2.3.2 Regional Groups

```r
ggplot(aes(x= date, y= confirmed), data = EAData) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1) + facet_wrap(~ region)
```

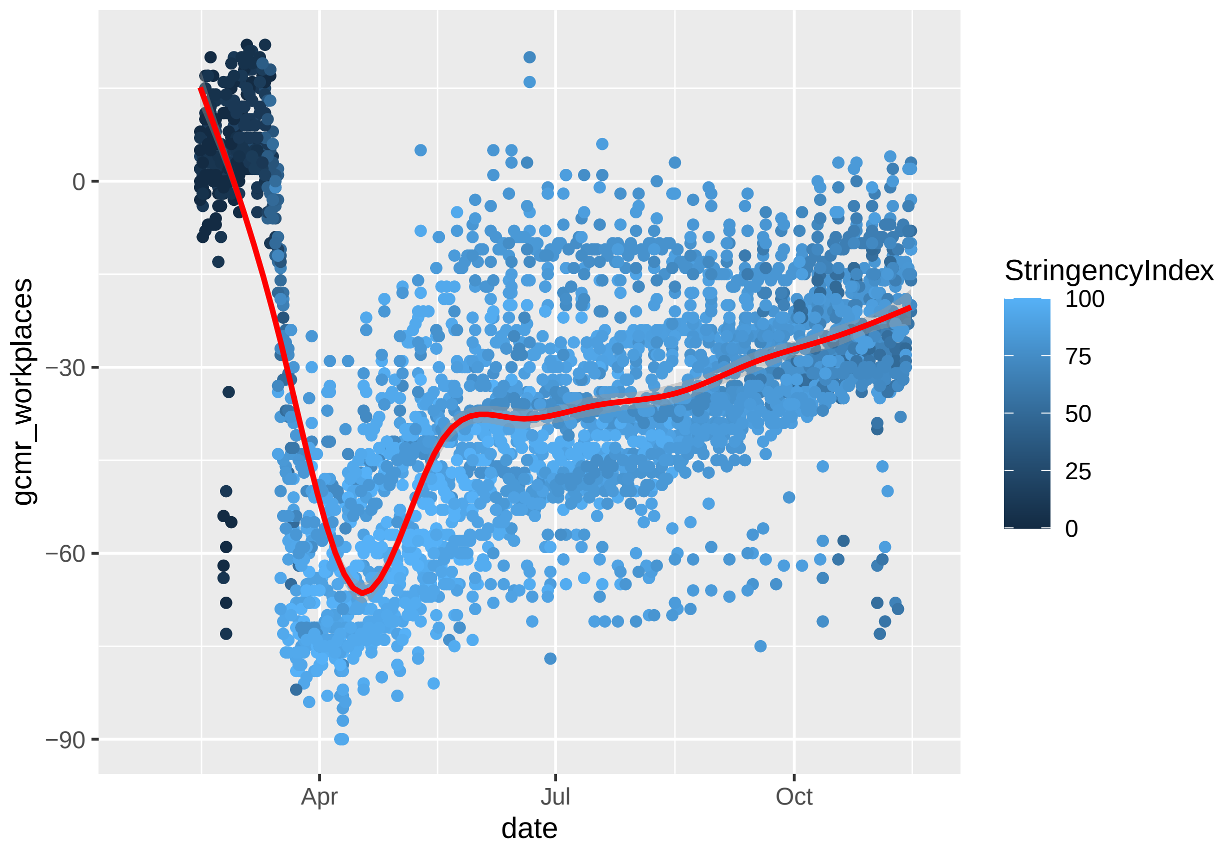


1.2.4 Workplaces

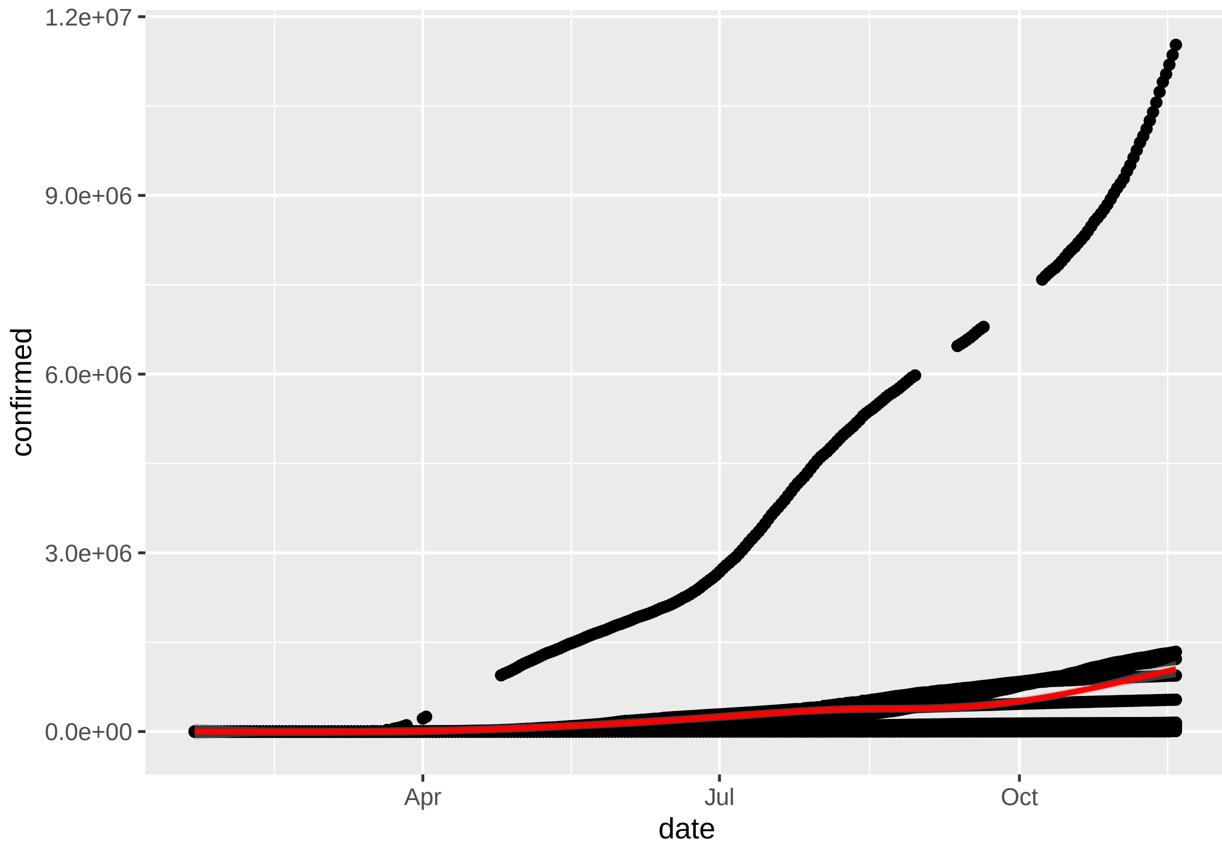
1.2.4.1 Global stringency quartiles

Upper:

```
ggplot(aes(x = date, y = gcmr_workplaces, color = StringencyIndex ),
       data = HStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

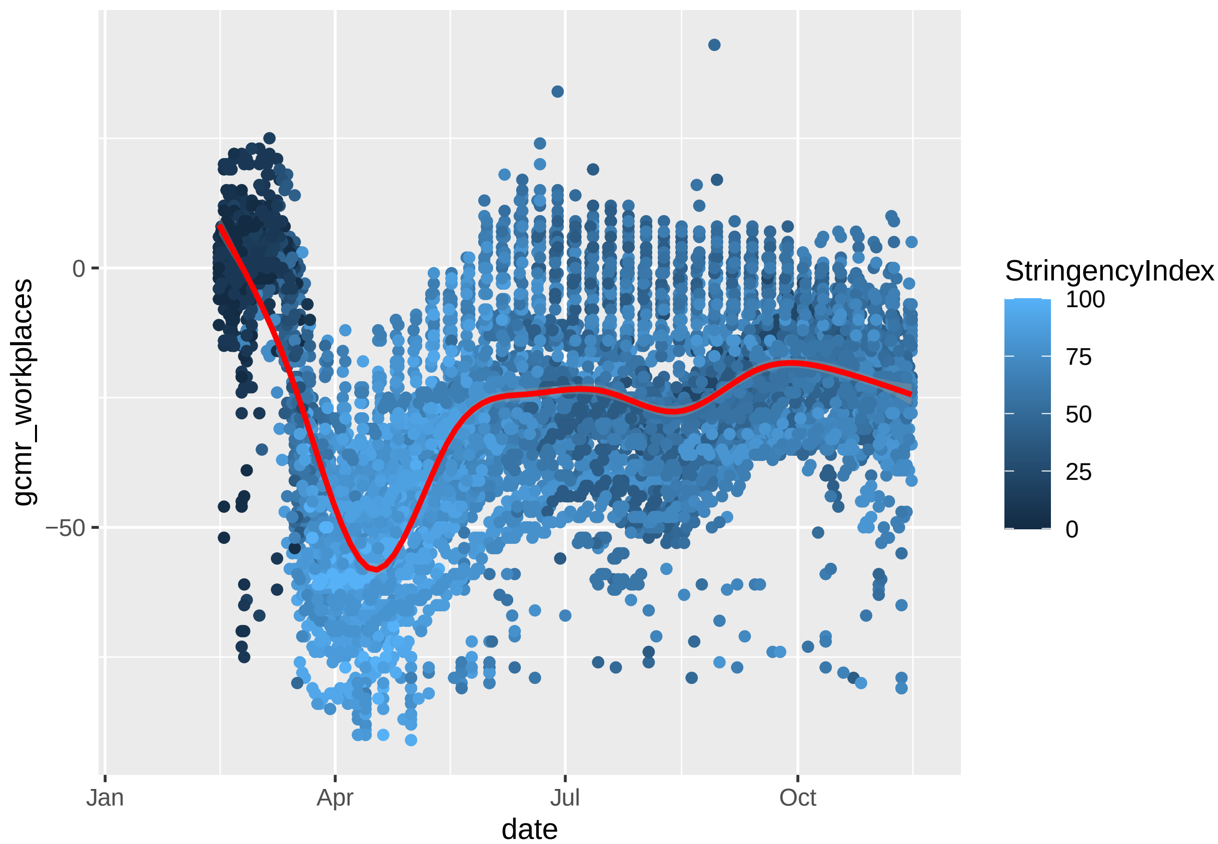



```r
ggplot(aes(x= date, y= confirmed), data = HStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

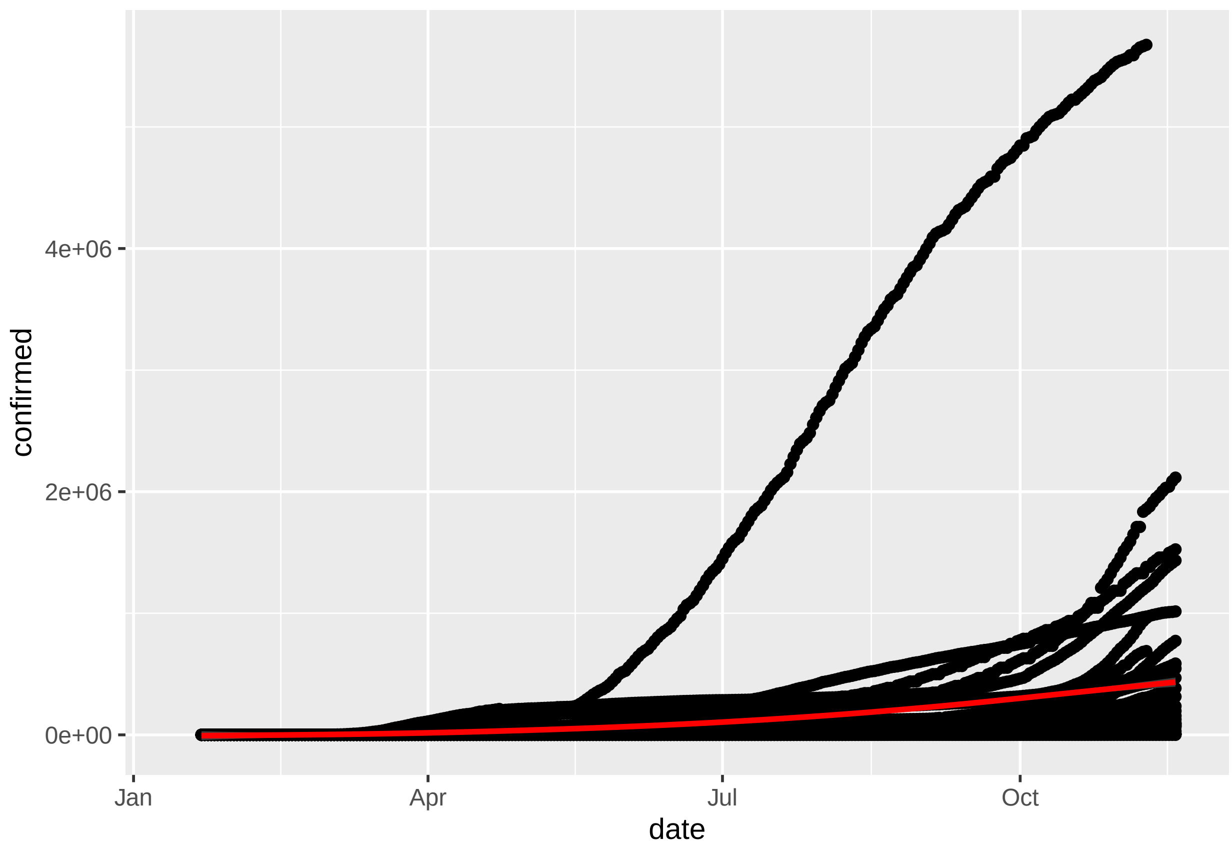


Middle:

```
ggplot(aes(x = date, y = gcmr_workplaces, color = StringencyIndex ),
       data = MStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

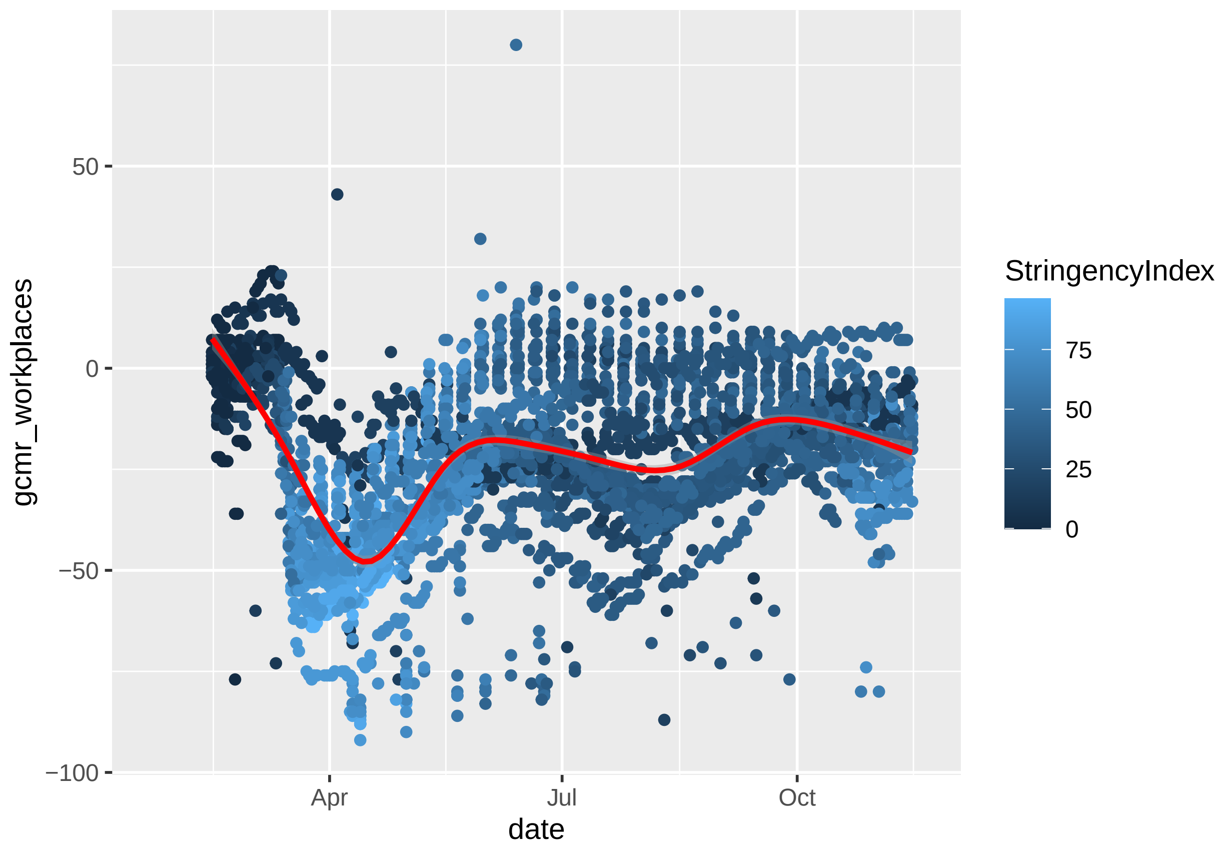



```r
ggplot(aes(x= date, y= confirmed), data = MStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

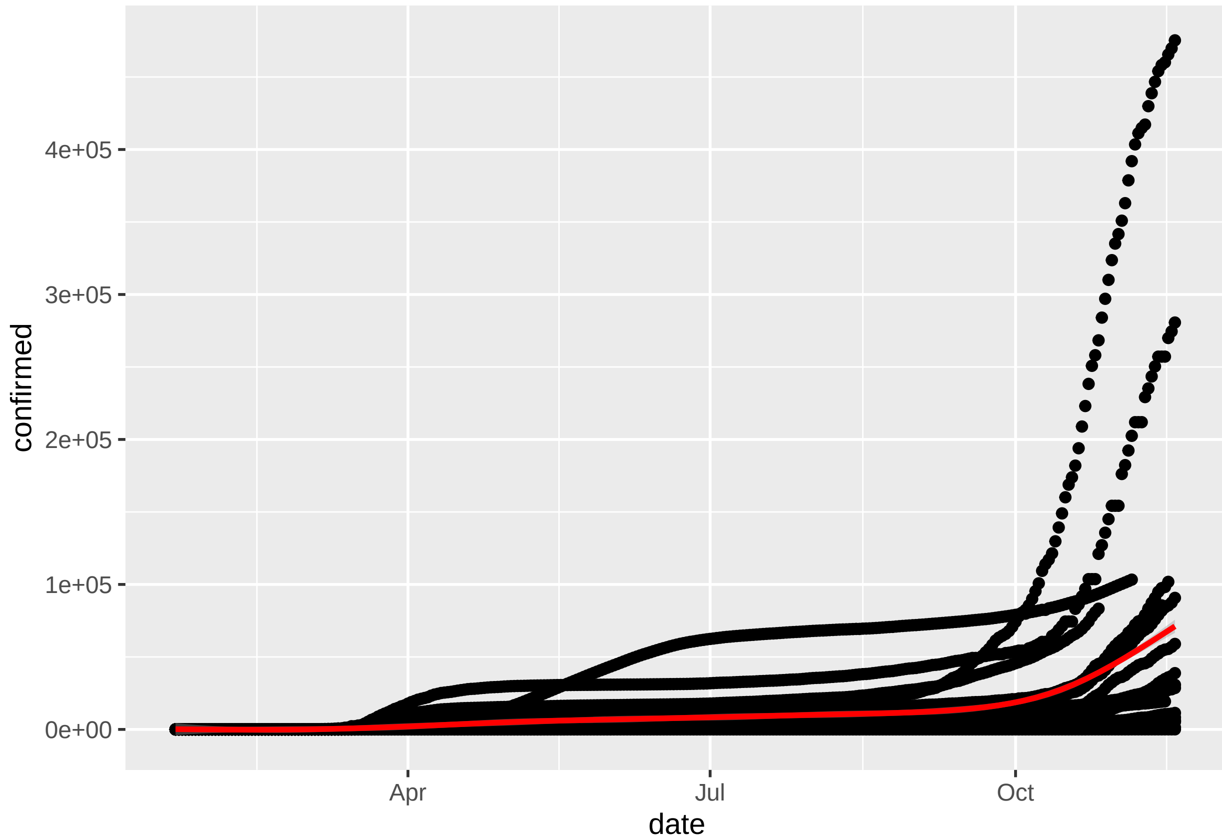


Lower:

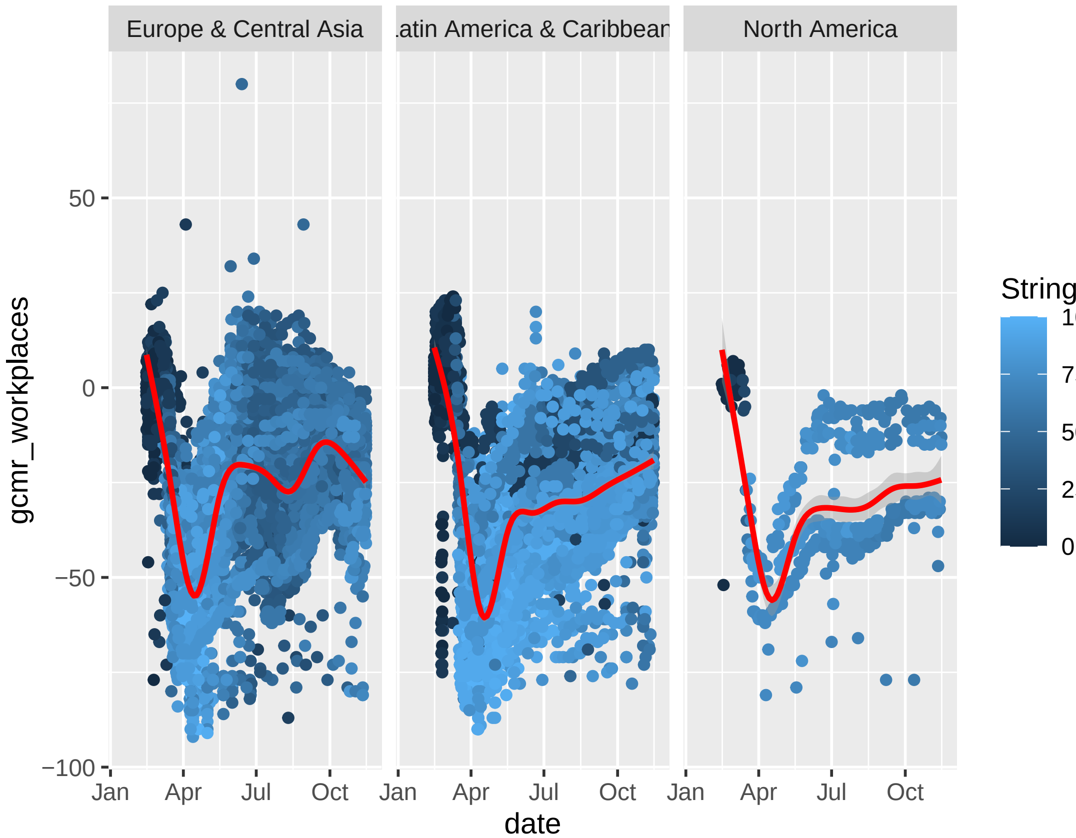
```
ggplot(aes(x = date, y = gcmr_workplaces, color = StringencyIndex ),
       data = LStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```




```
ggplot(aes(x= date, y= confirmed), data = LStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

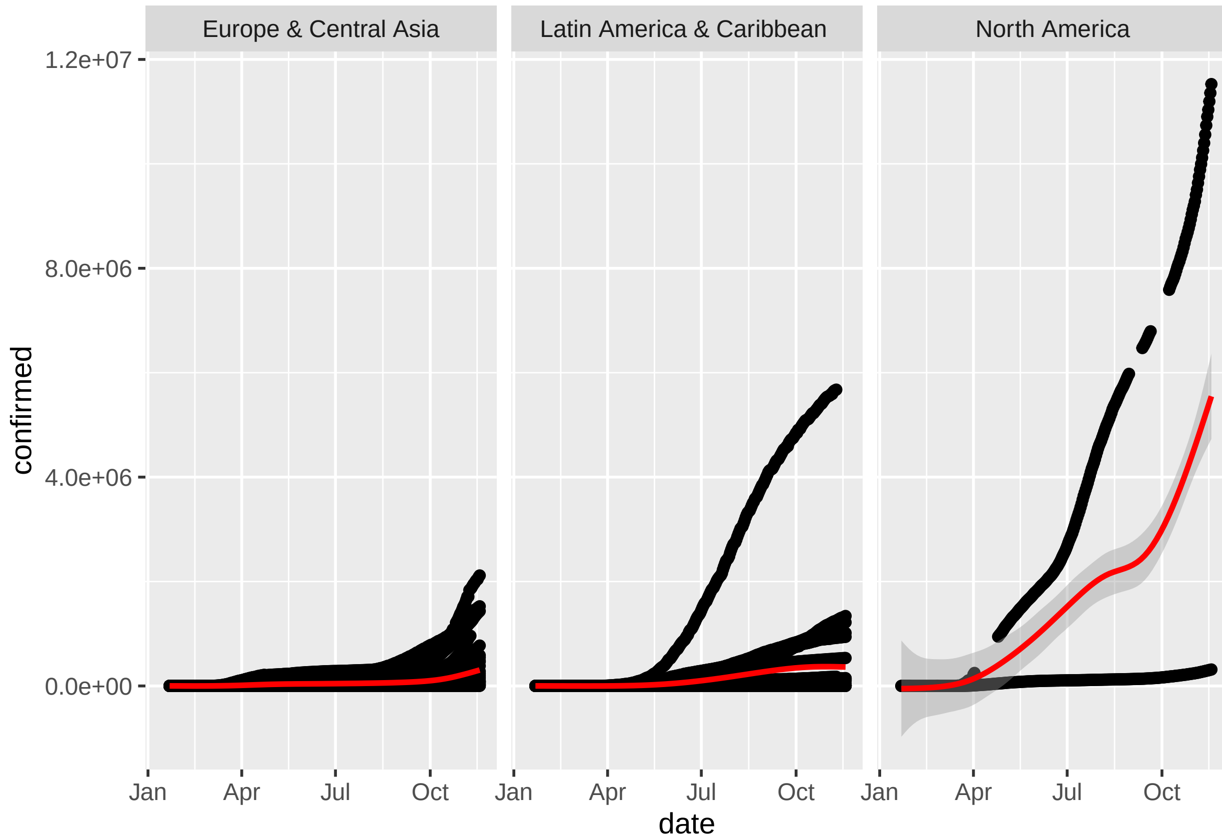



```r
ggplot(aes(x = date, y = gcmr_workplaces, color = StringencyIndex ),
       data = EAData) +
    geom_point()+ facet_wrap(~ region) +
  geom_smooth(colour = "red", method = "gam", group = 1)
```



1.2.4.2 Regional groups

```
ggplot(aes(x= date, y= confirmed), data = EAData) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1) + facet_wrap(~ region)
```

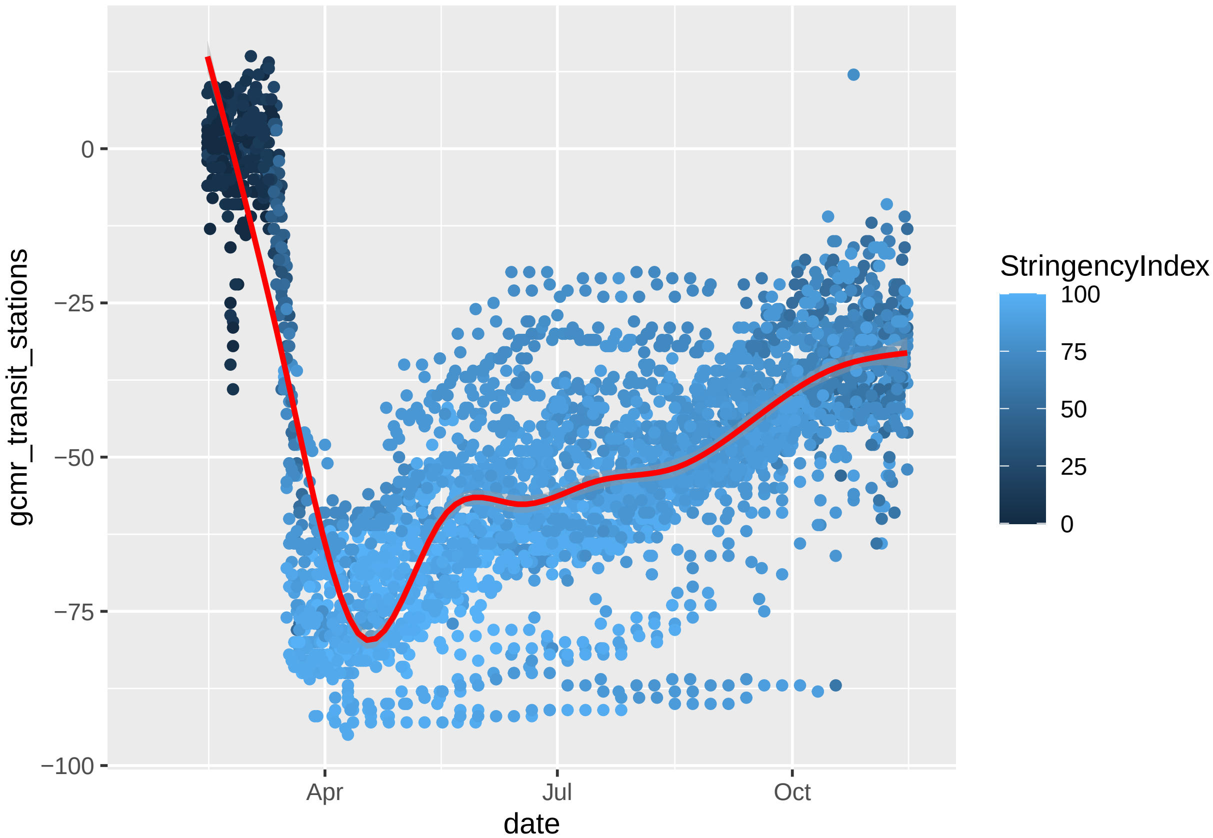


1.2.5 Transit stations

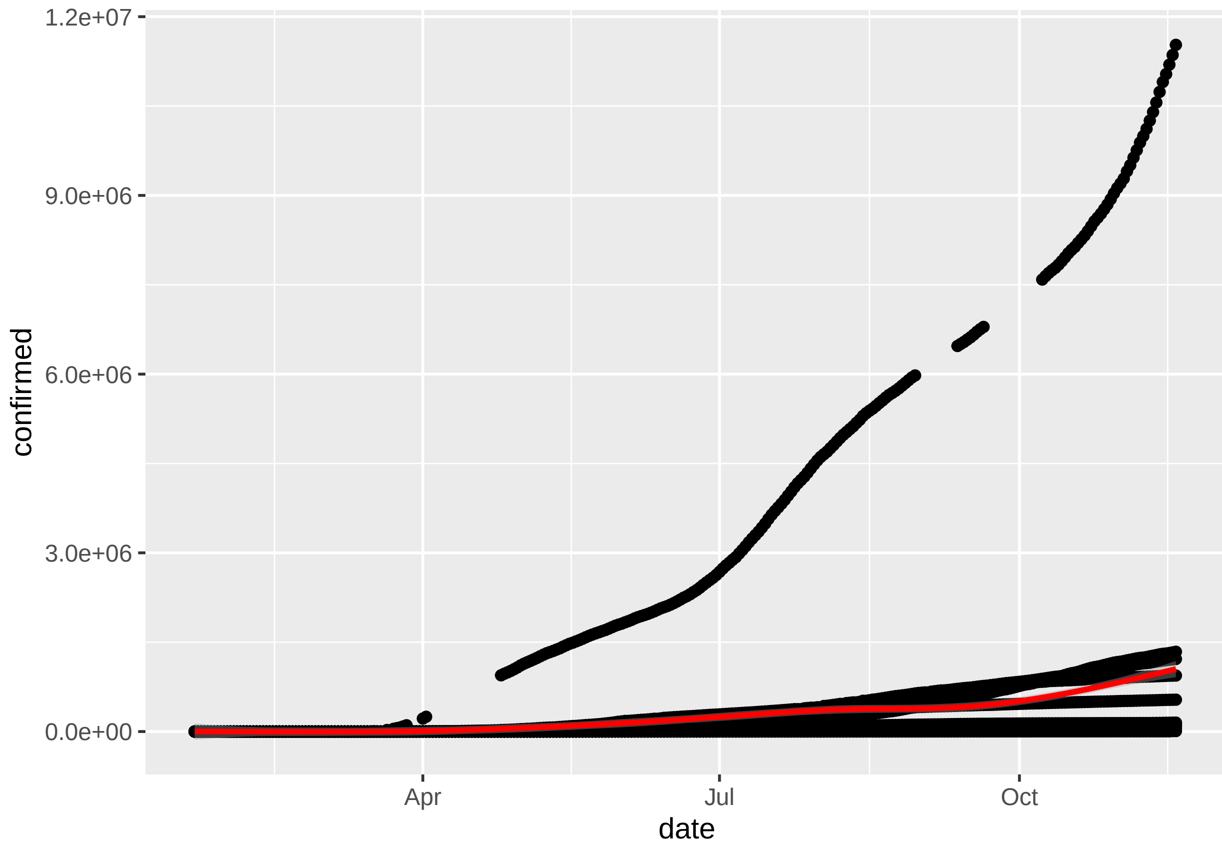
1.2.5.1 Global stringency quartiles

Upper:

```
ggplot(aes(x = date, y = gcmr_transit_stations, color = StringencyIndex ),
       data = HStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

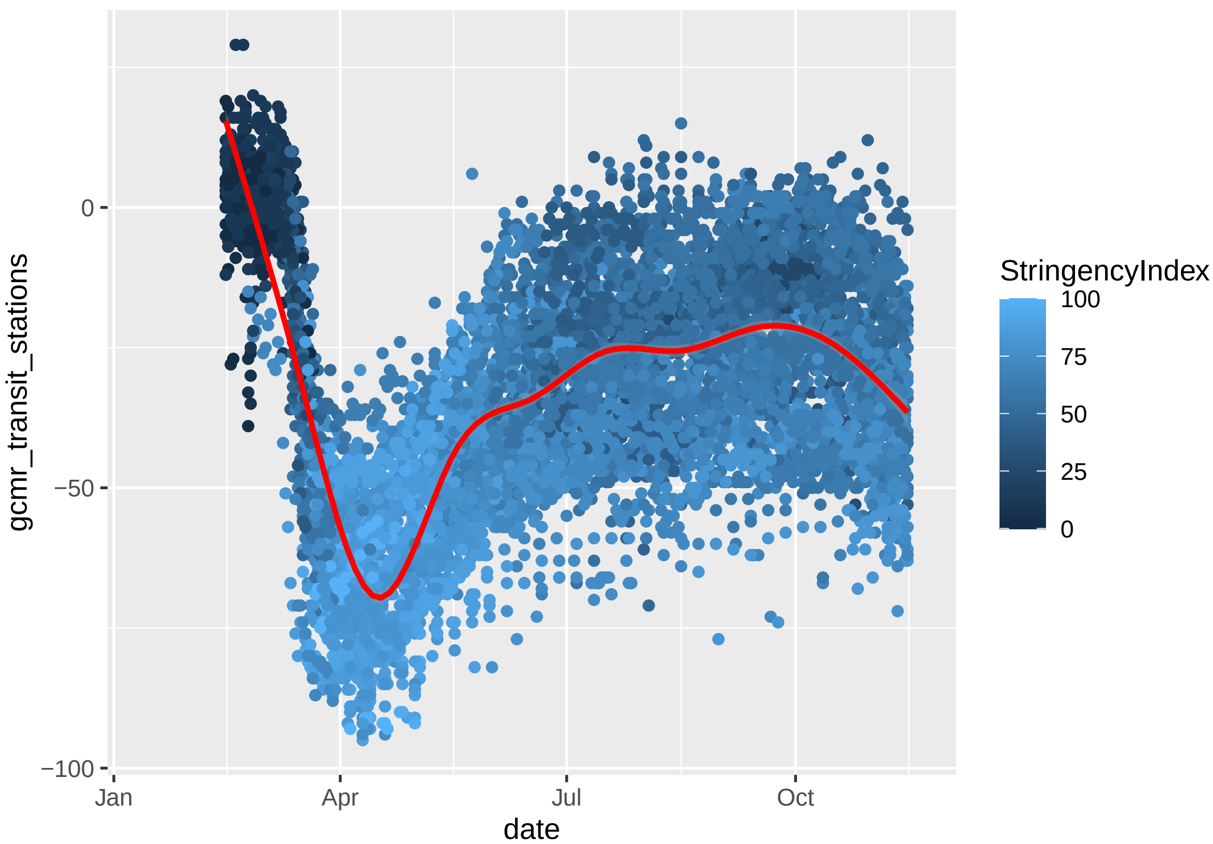



```
ggplot(aes(x= date, y= confirmed), data = HStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

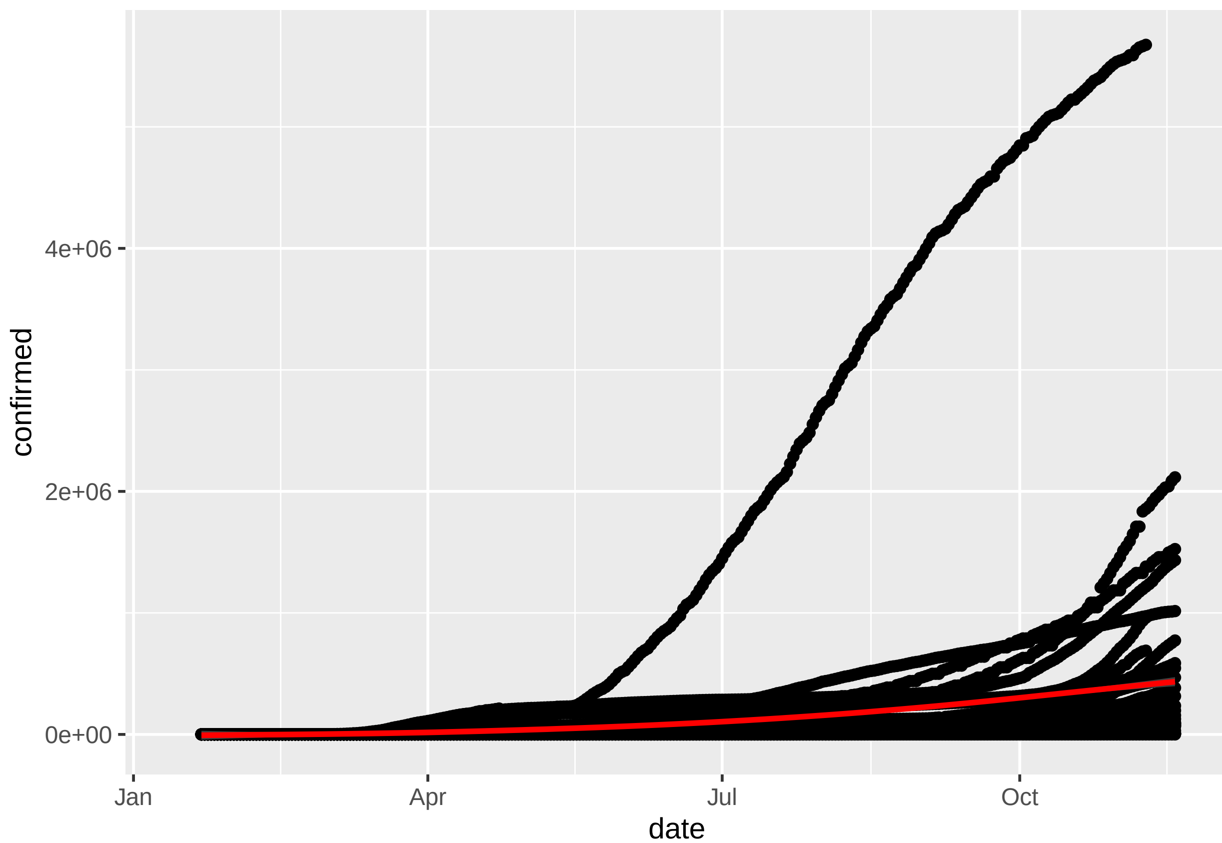


Middle:

```
ggplot(aes(x = date, y = gcmr_transit_stations, color = StringencyIndex ),
       data = MStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```

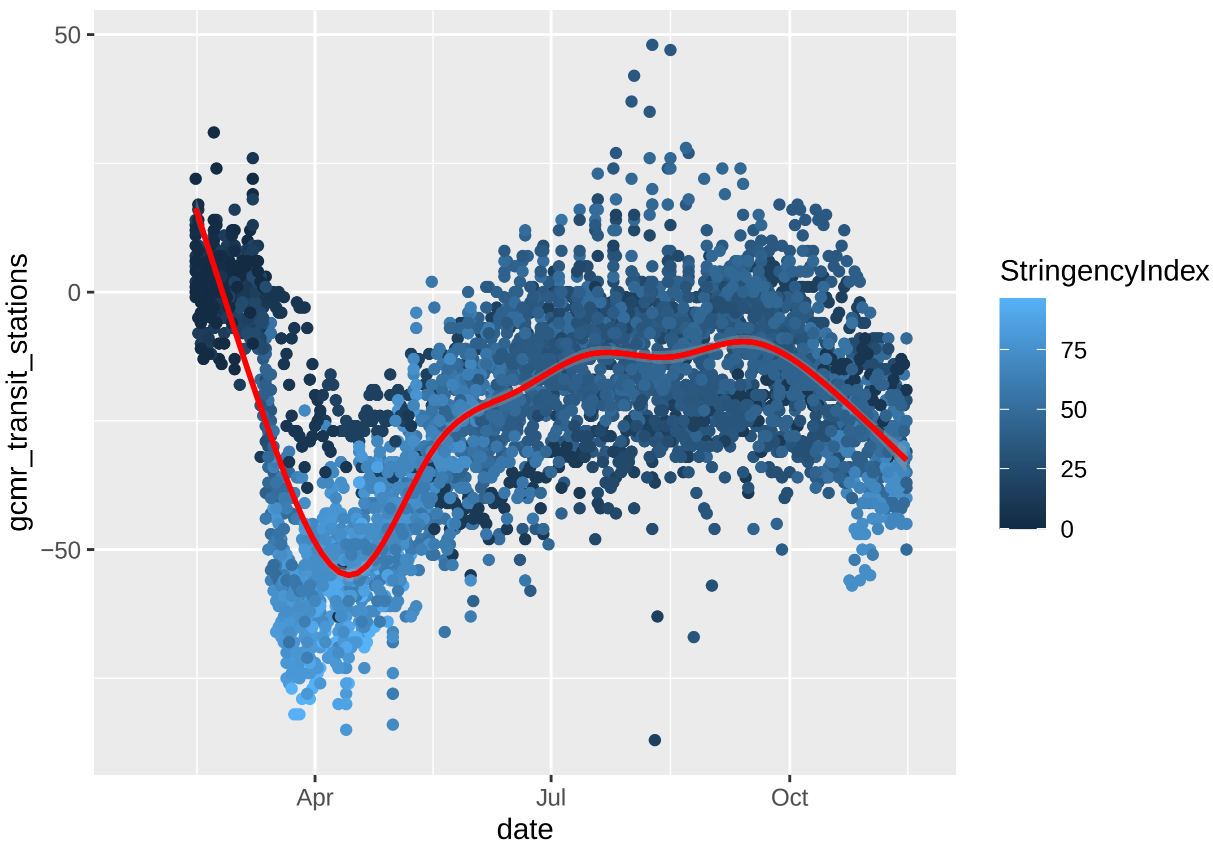



```
ggplot(aes(x= date, y= confirmed), data = MStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

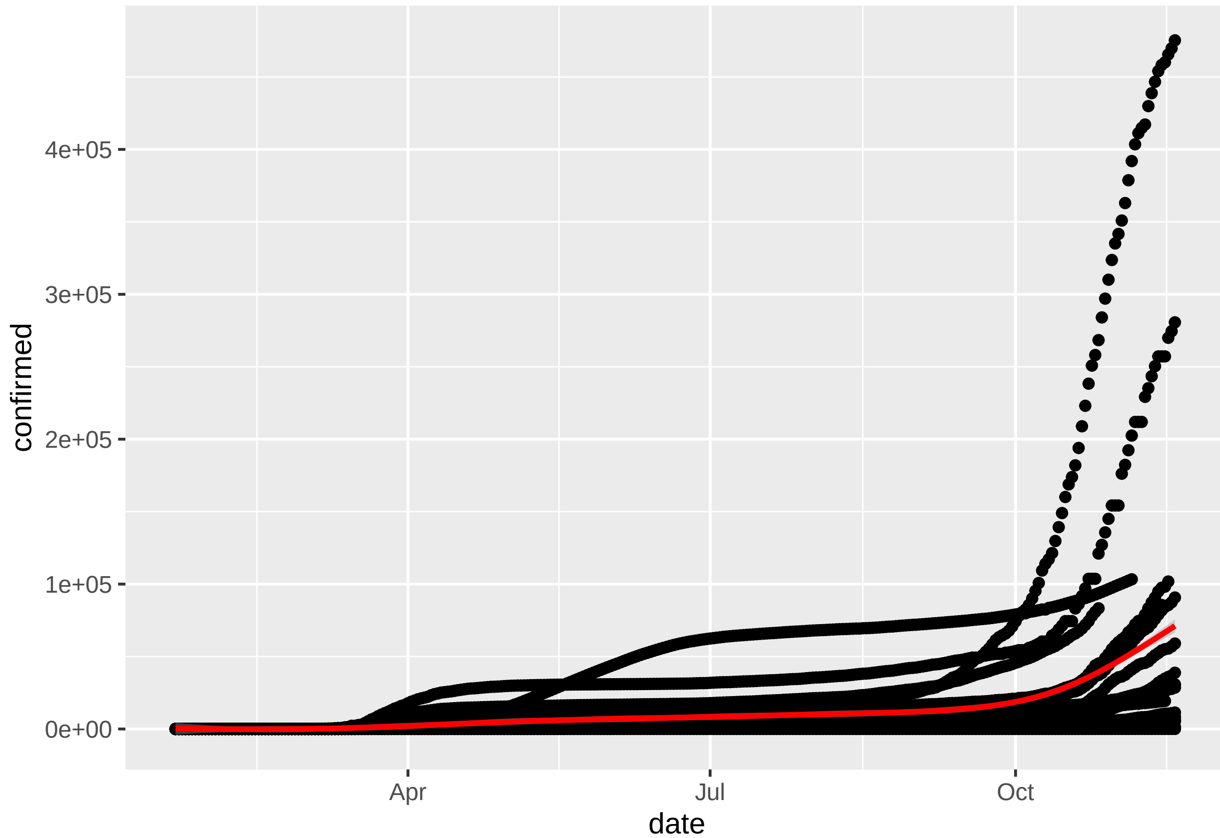


Lower:

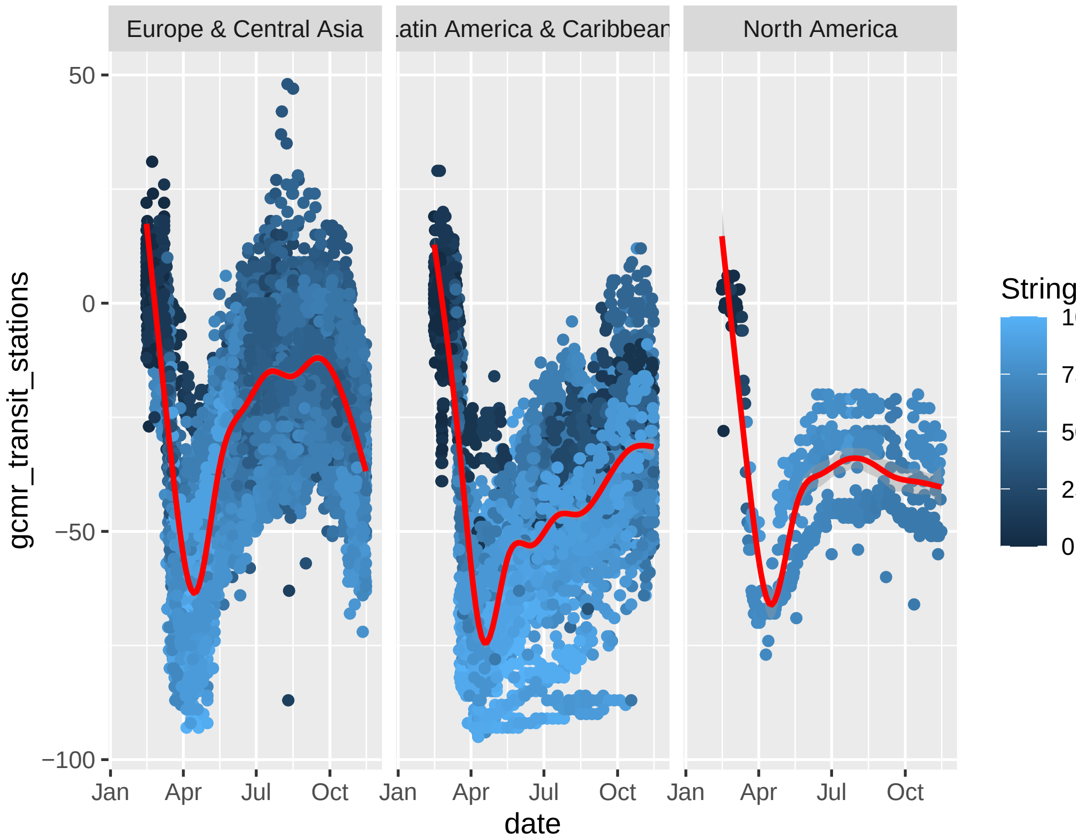
```
ggplot(aes(x = date, y = gcmr_transit_stations, color = StringencyIndex ),
       data = LStringency) +
    geom_point() + geom_smooth(colour = "red", method = "gam", group = 1)
```




```r
ggplot(aes(x= date, y= confirmed), data = LStringency) + geom_point() +
  geom_smooth(colour = "red", method = "gam", group = 1)
```

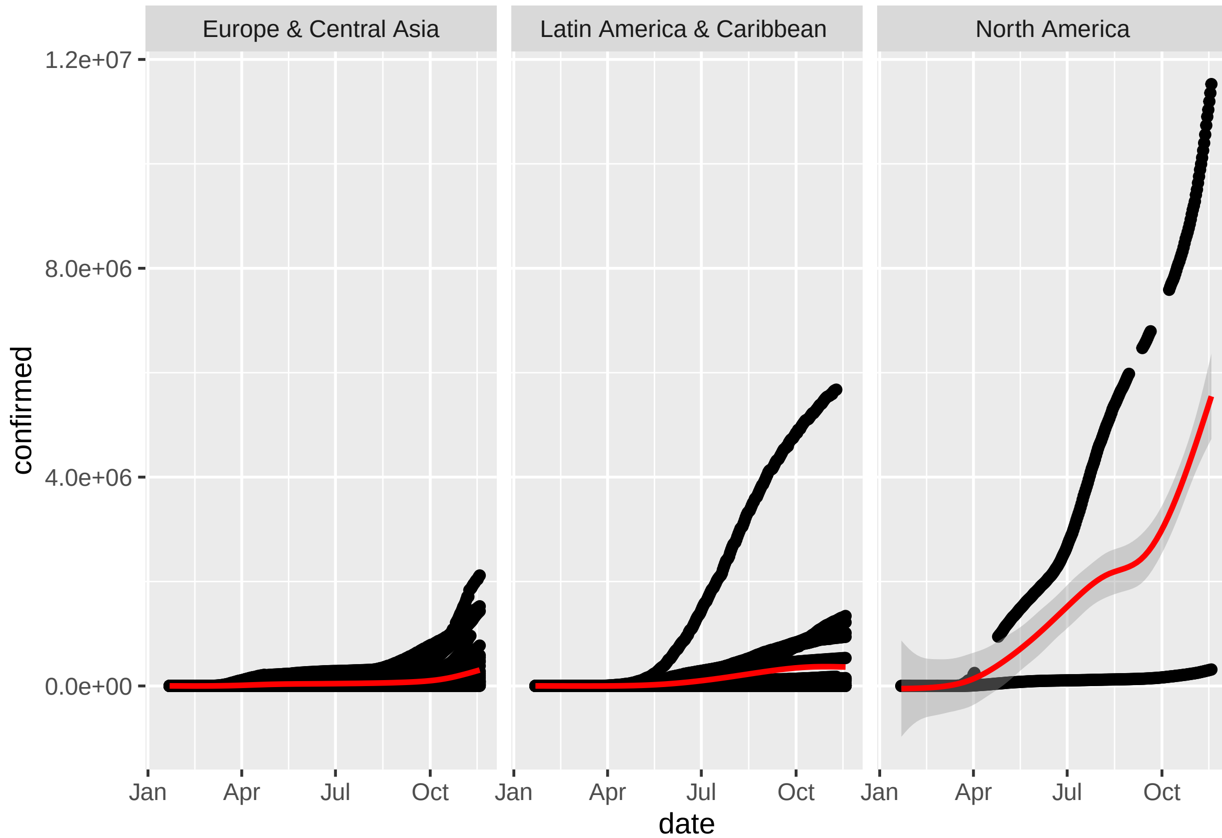



```
ggplot(aes(x = date, y = gcmr_transit_stations, color = StringencyIndex ),data = EAData) +
    geom_point()+ facet_wrap(~ region)  + geom_smooth(colour = "red", method = "gam", group = 1)
```



1.2.5.2 Regional groups

```
ggplot(aes(x= date, y= confirmed), data = EAData) + geom_point() + geom_smooth(colour = "red", method =
```

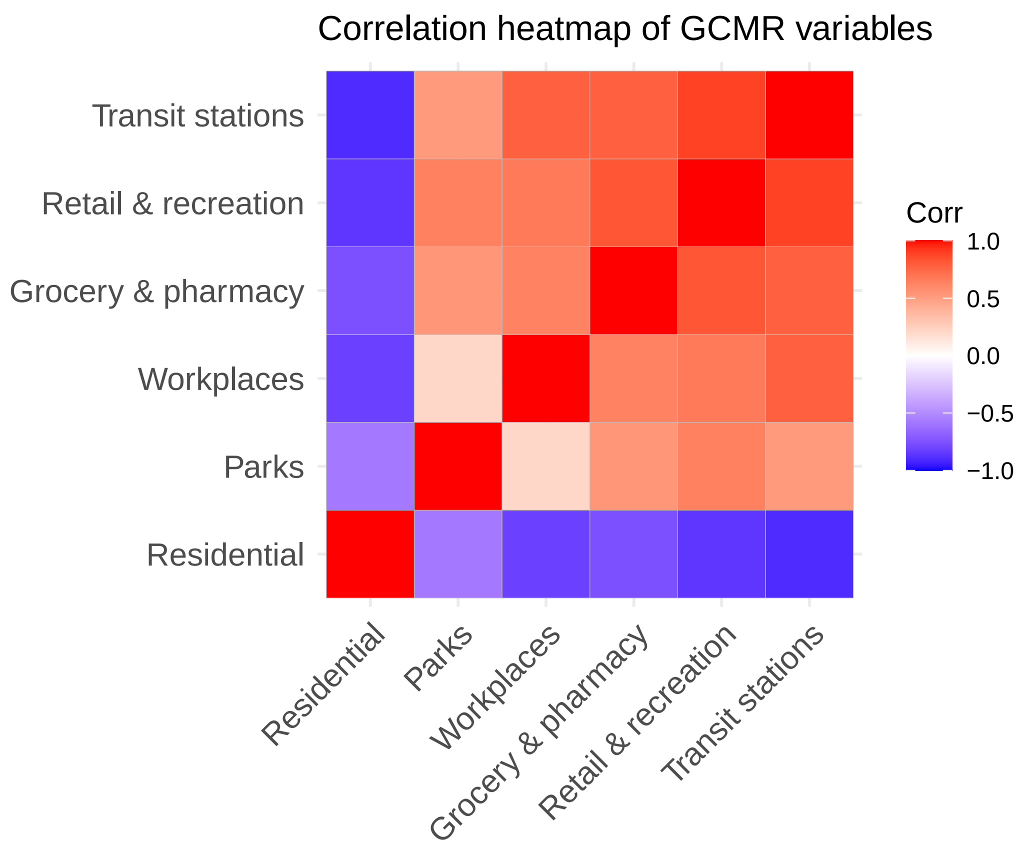


2 Correlation analysis & testing

2.1 Correlation of GCMR variables

Pearson correlation between the variables in Google Community Mobility Reports is visualised below.

```
gcmr_only <- EAData %>% select(gcmr_retail_recreation, gcmr_grocery_pharmacy,
                               gcmr_parks, gcmr_transit_stations,
                               gcmr_workplaces, gcmr_residential)
names(gcmr_only) <- c("Retail & recreation", "Grocery & pharmacy", "Parks",
                      "Transit stations", "Workplaces", "Residential")
gcmr_cor <- cor(gcmr_only, use = "pairwise.complete.obs", method = "pearson")
ggcorrplot(gcmr_cor, hc.order = TRUE) +
  ggtitle("Correlation heatmap of GCMR variables")
```



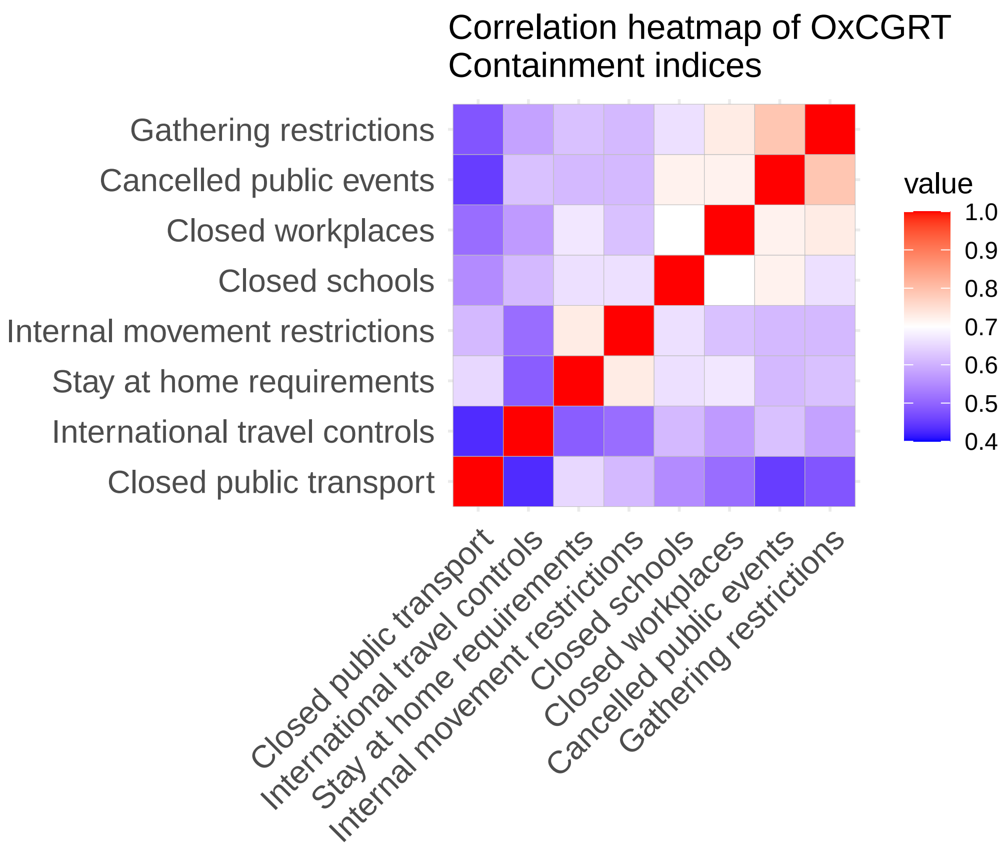
Notably, residential activity is significantly negatively correlated with all other variables. This is unsurprising, as it can be safely assumed that when a person is their residence, they are not in one of the other places denoted by the variable, and vice versa. There is a strong correlation between transit station activity and retail and recreation activity. This perhaps can be attributed to people using transit stations to commute to their retail and recreation activities.

2.2 Correlation of Oxford containment indices

The below heatmap visualises the Pearson correlation between containment variables provided by the OxCGRT data.

```
containment_only <- EAData %>% select(C1_School.closing, C2_Workplace.closing,
                                      C3_Cancel.public.events,
                                      C4_Restrictions.on.gatherings,
                                      C5_Close.public.transport,
                                      C6_Stay.at.home.requirements,
                                      C7_Restrictions.on.internal.movement,
                                      C8_International.travel.controls)
names(containment_only) <- c("Closed schools", "Closed workplaces",
                             "Cancelled public events",
                             "Gathering restrictions",
                             "Closed public transport",
                             "Stay at home requirements",
                             "Internal movement restrictions",
                             "International travel controls")
```

```r
containment_cor <- cor(containment_only, use = "pairwise.complete.obs", method = "pearson")
ggcorrplot(containment_cor, hc.order = TRUE) +
  ggtitle("Correlation heatmap of OxCGRT \nContainment indices") +
  scale_fill_gradient2(limit = c(0.4,1), low = "blue", high =  "red", mid = "white", midpoint = 0.7)
```



It should be noted that the scale has changed, in order to better visualise the correlation. All indices are positively correlated. The least positively correlated indices are between international travel controls, and closed public transport. Worldwide, international travel restrictions were very strict, and remained in place for months; however, public transport was not fully closed in many countries, as it was a necessary part of transit for essential health workers.

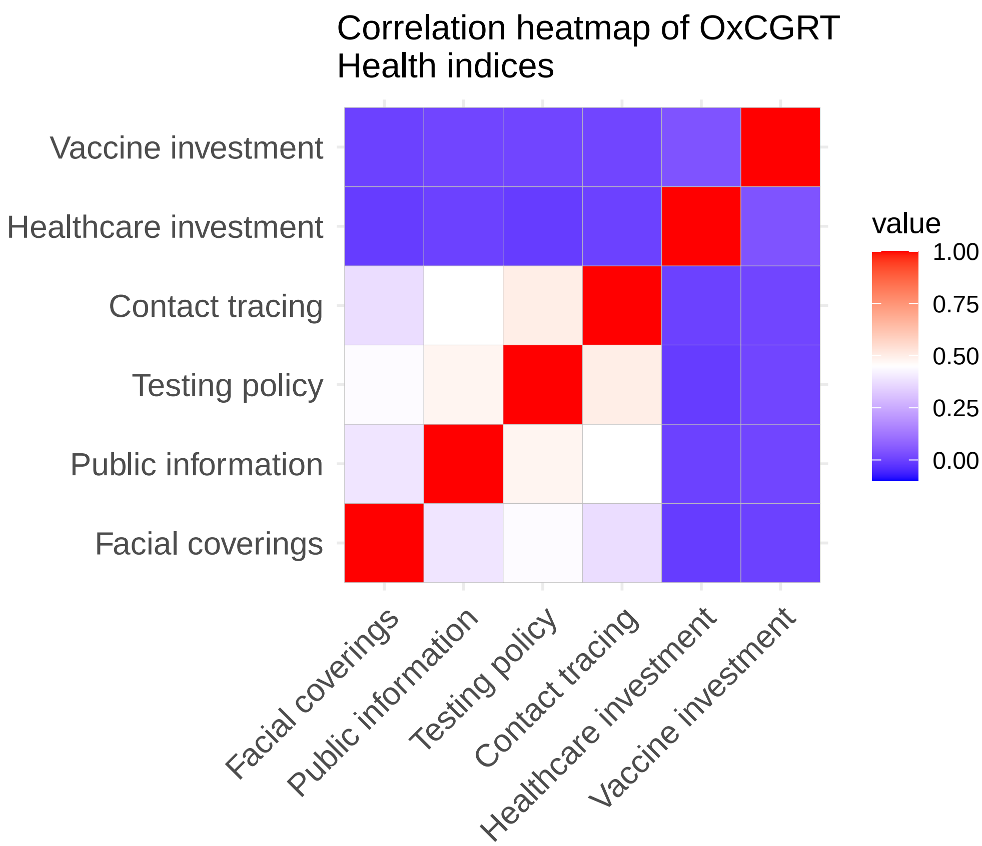
Cancelling big events and gathering restrictions are very closely correlated, as they often represent the same type of government intervention - preventing large numbers of people from congregating.

2.3 Correlation of Oxford health indices

The below heatmap visualises the Pearson correlation between health variables provided by the OxCGRT data.

```r
health_only <- EAData %>% select(H1_Public.information.campaigns,
                                 H2_Testing.policy,
                                 H3_Contact.tracing,
                                 H4_Emergency.investment.in.healthcare,
                                 H5_Investment.in.vaccines,
                                 H6_Facial.Coverings)
```

```r
names(health_only) <- c("Public information", "Testing policy",
                         "Contact tracing",
                         "Healthcare investment",
                         "Vaccine investment",
                         "Facial coverings")
health_cor <- cor(health_only, use = "pairwise.complete.obs", method = "pearson")
ggcorrplot(health_cor, hc.order = TRUE) +
  ggtitle("Correlation heatmap of OxCGRT \nHealth indices") +
  scale_fill_gradient2(limit = c(-0.1,1), low = "blue", high =  "red", mid = "white", midpoint = 0.45)
```



As with the containment indices, the scale has been adjusted to represent the true distribution of correlations. There are no notable correlations with these health indices. Facial coverings, public information campaigns, testing policies, and contact tracing are weakly positively correlated. These variables represent basic healthcare management steps in the face of a pandemic, and are likely to coexist.

2.4 Combined correlation

After having plotted some graphs to get a feel for how the data looks and behaves we want to know if any variables are correlated so perhaps we can explore them in more detail. Here we explore if there is any correlation between the individual, C - containment and closure policies and H - health system policies with the Retail, Workplace and Groceries variables.

```r
correlation_df <- data.frame(
  row.names = c("C1","C2","C3","C4","C5",
                "C6","C7","C8","H1","H2",
```

```
               "H3","H4","H5","H6"),
  Retail_Corr = c(C1 = cor(EAData$gcmr_retail_recreation,
                           EAData$C1_School.closing,
                           use = "complete.obs"),
                C2 = cor(EAData$gcmr_retail_recreation,
                           EAData$C2_Workplace.closing,
                           use = "complete.obs"),
                C3 = cor(EAData$gcmr_retail_recreation,
                           EAData$C3_Cancel.public.events,
                           use = "complete.obs"),
                C4 = cor(EAData$gcmr_retail_recreation,
                           EAData$C4_Restrictions.on.gatherings,
                           use = "complete.obs"),
                C5 = cor(EAData$gcmr_retail_recreation,
                           EAData$C5_Close.public.transport,
                           use = "complete.obs"),
                C6 = cor(EAData$gcmr_retail_recreation,
                           EAData$C6_Stay.at.home.requirements,
                           use = "complete.obs"),
                C7 = cor(EAData$gcmr_retail_recreation,
                           EAData$C7_Restrictions.on.internal.movement,
                           use = "complete.obs"),
                C8 = cor(EAData$gcmr_retail_recreation,
                           EAData$C8_International.travel.controls,
                           use = "complete.obs"),
                H1 = cor(EAData$gcmr_retail_recreation,
                           EAData$H1_Public.information.campaigns,
                           use = "complete.obs"),
                H2 = cor(EAData$gcmr_retail_recreation,
                           EAData$H2_Testing.policy,
                           use = "complete.obs"),
                H3 = cor(EAData$gcmr_retail_recreation,
                           EAData$H3_Contact.tracing,
                           use = "complete.obs"),
                H4 = cor(EAData$gcmr_retail_recreation,
                           EAData$H4_Emergency.investment.in.healthcare,
                           use = "complete.obs"),
                H5 = cor(EAData$gcmr_retail_recreation,
                           EAData$H5_Investment.in.vaccines,
                           use = "complete.obs"),
                H6 = cor(EAData$gcmr_retail_recreation,
                           EAData$H6_Facial.Coverings,
                           use = "complete.obs")
  ),
  Workplace_Corr = c(C1 = cor(EAData$gcmr_workplaces,
                           EAData$C1_School.closing,
                           use = "complete.obs"),
                C2 = cor(EAData$gcmr_workplaces,
                           EAData$C2_Workplace.closing,
                           use = "complete.obs"),
                C3 = cor(EAData$gcmr_workplaces,
                           EAData$C3_Cancel.public.events,
                           use = "complete.obs"),
```

```r
                   C4 = cor(EAData$gcmr_workplaces,
                            EAData$C4_Restrictions.on.gatherings,
                            use = "complete.obs"),
                   C5 = cor(EAData$gcmr_workplaces,
                            EAData$C5_Close.public.transport,
                            use = "complete.obs"),
                   C6 = cor(EAData$gcmr_workplaces,
                            EAData$C6_Stay.at.home.requirements,
                            use = "complete.obs"),
                   C7 = cor(EAData$gcmr_workplaces,
                            EAData$C7_Restrictions.on.internal.movement,
                            use = "complete.obs"),
                   C8 = cor(EAData$gcmr_workplaces,
                            EAData$C8_International.travel.controls,
                            use = "complete.obs"),
                   H1 = cor(EAData$gcmr_workplaces,
                            EAData$H1_Public.information.campaigns,
                            use = "complete.obs"),
                   H2 = cor(EAData$gcmr_workplaces,
                            EAData$H2_Testing.policy,
                            use = "complete.obs"),
                   H3 = cor(EAData$gcmr_workplaces,
                            EAData$H3_Contact.tracing,
                            use = "complete.obs"),
                   H4 = cor(EAData$gcmr_workplaces,
                            EAData$H4_Emergency.investment.in.healthcare,
                            use = "complete.obs"),
                   H5 = cor(EAData$gcmr_workplaces,
                            EAData$H5_Investment.in.vaccines,
                            use = "complete.obs"),
                   H6 = cor(EAData$gcmr_workplaces,
                            EAData$H6_Facial.Coverings,
                            use = "complete.obs")
  ),
  Grocery_Corr = c(C1 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C1_School.closing,
                            use = "complete.obs"),
                   C2 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C2_Workplace.closing,
                            use = "complete.obs"),
                   C3 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C3_Cancel.public.events,
                            use = "complete.obs"),
                   C4 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C4_Restrictions.on.gatherings,
                            use = "complete.obs"),
                   C5 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C5_Close.public.transport,
                            use = "complete.obs"),
                   C6 = cor(EAData$gcmr_grocery_pharmacy,
                            EAData$C6_Stay.at.home.requirements,
                            use = "complete.obs"),
                   C7 = cor(EAData$gcmr_grocery_pharmacy,
```

```r
                                    EAData$C7_Restrictions.on.internal.movement,
                                    use = "complete.obs"),
                        C8 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$C8_International.travel.controls,
                                    use = "complete.obs"),
                        H1 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H1_Public.information.campaigns,
                                    use = "complete.obs"),
                        H2 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H2_Testing.policy,
                                    use = "complete.obs"),
                        H3 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H3_Contact.tracing,
                                    use = "complete.obs"),
                        H4 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H4_Emergency.investment.in.healthcare,
                                    use = "complete.obs"),
                        H5 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H5_Investment.in.vaccines,
                                    use = "complete.obs"),
                        H6 = cor(EAData$gcmr_grocery_pharmacy,
                                    EAData$H6_Facial.Coverings,
                                    use = "complete.obs")
  ))

kable(correlation_df)
```

	Retail_Corr	Workplace_Corr	Grocery_Corr
C1	-0.6221818	-0.5189212	-0.4504805
C2	-0.6291966	-0.5538510	-0.4649939
C3	-0.5600977	-0.4524492	-0.4058998
C4	-0.5080346	-0.4703083	-0.3478436
C5	-0.5360284	-0.4043785	-0.4488337
C6	-0.6692673	-0.5128903	-0.5535797
C7	-0.6204898	-0.4815517	-0.4770798
C8	-0.4232938	-0.3947373	-0.3532599
H1	-0.2350113	-0.2775067	-0.1433123
H2	0.0769589	-0.0698679	0.0822914
H3	0.0725148	-0.0419604	0.1152478
H4	-0.0128634	-0.0078554	-0.0022266
H5	-0.0103667	-0.0047282	-0.0075104
H6	-0.0734493	-0.0543538	-0.0508921

2.5 Shapiro normality testing

For modelling purposes we thought it may be a good idea to test if either GCMR Retail & Recreation or Workplaces were normally distributed. So we tested them with the Shapiro-Wilk normality test. However as the data set was to big it had to be split by continents, and as the Europe set was still to large it was evaluated visually by QQ-plots.

```r
Test.A <- EAData %>% filter(region == "North America")

shapiro.test(Test.A$gcmr_retail_recreation)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Test.A$gcmr_retail_recreation
## W = 0.91143, p-value = 8.99e-16
```

```r
shapiro.test(Test.A$gcmr_workplaces)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Test.A$gcmr_workplaces
## W = 0.95305, p-value = 6.203e-11
```

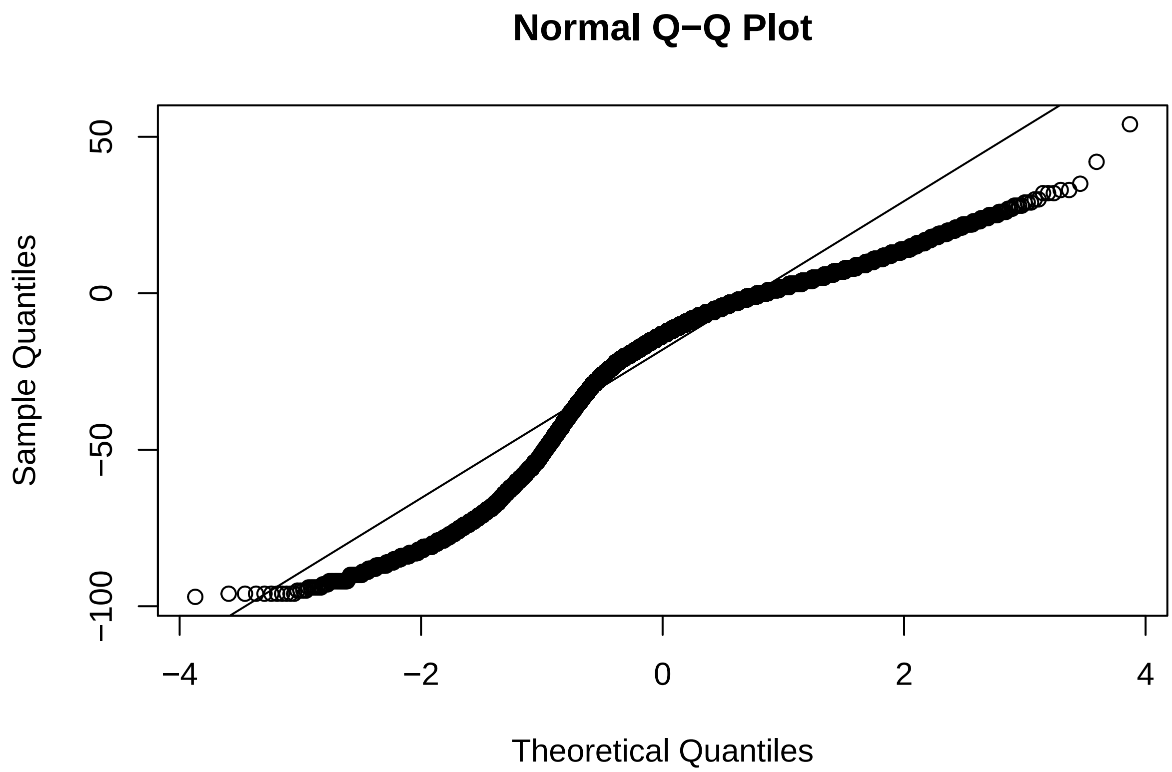
```r
Test.B <- EAData %>% filter(continent == "South America")

shapiro.test(Test.B$gcmr_retail_recreation)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Test.B$gcmr_retail_recreation
## W = 0.98607, p-value = 9.606e-16
```

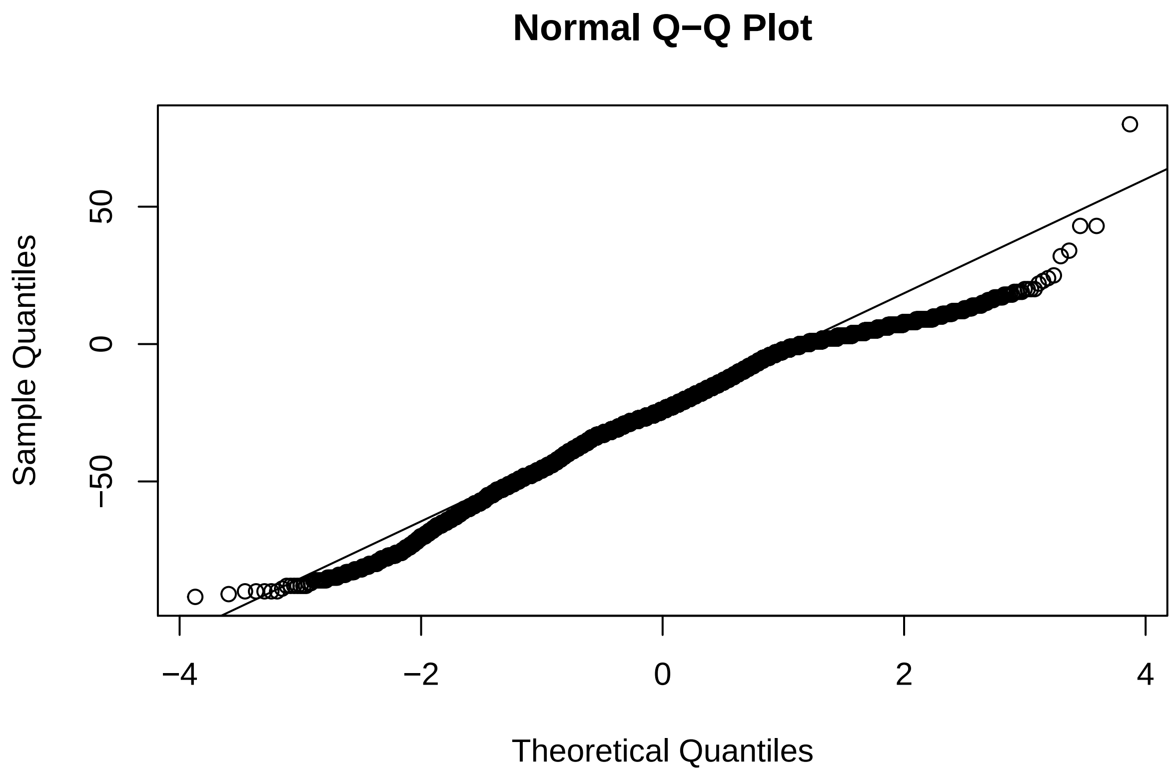
```r
shapiro.test(Test.B$gcmr_workplaces)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Test.B$gcmr_workplaces
## W = 0.98843, p-value = 4.057e-14
```

```r
Test.C <- EAData %>% filter(continent == "Europe")

qqnorm(Test.C$gcmr_retail_recreation)
qqline(Test.C$gcmr_retail_recreation)
```




```r
qqnorm(Test.C$gcmr_workplaces)
qqline(Test.C$gcmr_workplaces)
```



The tests show that the relationship between changes in retail & recreation / grocery and pharmacy and new confirmed cases of COVID-19 in all regions were significantly different from the normal distribution.

3 Modelling

3.1 Retail

In this section we will be creating various models of the retail variable.
We split these models by continent as we've seen there are different patterns for each continent.

3.1.1 Europe simple

Starting with Europe we create a simplified minimum and maximum linear model and use backwards stepwise regression to find the most appropriate model for retail

```
min_modelE1 <- lm(m_retail ~ 1, data = euroData)
max_modelE1 <- lm(m_retail ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + (active_per_pop), data = euroData)

euro_rr_backward1 <- step(max_modelE1, direction = 'backward',
scope = list('lower' = min_modelE1))
```

```
## Start:  AIC=50333.37
## m_retail ~ 1 + C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events +
##     C4_Restrictions.on.gatherings + C5_Close.public.transport +
##     C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
##     C8_International.travel.controls + (active_per_pop)
##
##                                       Df Sum of Sq     RSS   AIC
## <none>                                             2414216 50333
## - C4_Restrictions.on.gatherings         1      2445 2416660 50340
## - C8_International.travel.controls      1      7742 2421958 50360
## - C5_Close.public.transport             1      8656 2422872 50364
## - active_per_pop                        1     24085 2438301 50421
## - C2_Workplace.closing                  1     30999 2445215 50446
## - C3_Cancel.public.events               1     73170 2487386 50600
## - C1_School.closing                     1     87685 2501901 50652
## - C7_Restrictions.on.internal.movement  1     88999 2503215 50657
## - C6_Stay.at.home.requirements          1    157550 2571766 50900
```

And here is our most appropriate model

```r
summary(euro_rr_backward1)
```

```
##
## Call:
## lm(formula = m_retail ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     (active_per_pop), data = euroData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.237  -9.579  -0.851  11.337  45.201
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             8.5863     0.4866  17.647  < 2e-16 ***
## C1_School.closing                      -4.1044     0.2272 -18.067  < 2e-16 ***
## C2_Workplace.closing                   -3.1350     0.2918 -10.742  < 2e-16 ***
## C3_Cancel.public.events                -5.1192     0.3102 -16.504  < 2e-16 ***
## C4_Restrictions.on.gatherings          -0.5641     0.1870  -3.017  0.00256 **
## C5_Close.public.transport              -1.9510     0.3437  -5.676 1.42e-08 ***
## C6_Stay.at.home.requirements           -7.8431     0.3239 -24.217  < 2e-16 ***
## C7_Restrictions.on.internal.movement   -5.2104     0.2863 -18.202  < 2e-16 ***
## C8_International.travel.controls        0.9396     0.1750   5.368 8.14e-08 ***
## active_per_pop                      -1023.7397   108.1169  -9.469  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.39 on 8987 degrees of freedom
##   (3131 observations deleted due to missingness)
## Multiple R-squared:  0.5613, Adjusted R-squared:  0.5609
## F-statistic:  1278 on 9 and 8987 DF,  p-value: < 2.2e-16
```

in which we can see that all explanatory variables have been kept.

3.1.2 Europe complex

We now create a more complex model which involves more explanatory variables and their interactions.

```
min_modelE2 <- lm(m_retail ~ 1, data = euroData)
max_modelE2 <- lm(data = euroData, m_retail ~ 1 + (active_per_pop) +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.trac

euro_rr_backward2 <- step(max_modelE2, direction = 'backward',
scope = list('lower' = min_modelE2))
```

After stepwise regression analysis we our result of the most appropriate model is as follows

```
summary(euro_rr_backward2)
```

```
##
## Call:
## lm(formula = m_retail ~ active_per_pop + C1_School.closing +
##     C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.tracing +
##     H4_Emergency.investment.in.healthcare + gdp_capita + pop_density +
##     C1_School.closing:H1_Public.information.campaigns + C1_School.closing:H2_Testing.policy +
##     C1_School.closing:H3_Contact.tracing + C1_School.closing:H4_Emergency.investment.in.healthcare +
##     C2_Workplace.closing:H1_Public.information.campaigns + C2_Workplace.closing:H2_Testing.policy +
##     C2_Workplace.closing:H3_Contact.tracing + C2_Workplace.closing:H4_Emergency.investment.in.health
##     C3_Cancel.public.events:H1_Public.information.campaigns +
##     C3_Cancel.public.events:H2_Testing.policy + C3_Cancel.public.events:H3_Contact.tracing +
##     C4_Restrictions.on.gatherings:H1_Public.information.campaigns +
##     C4_Restrictions.on.gatherings:H2_Testing.policy + C4_Restrictions.on.gatherings:H3_Contact.tracin
##     C5_Close.public.transport:H1_Public.information.campaigns +
##     C5_Close.public.transport:H2_Testing.policy + C5_Close.public.transport:H3_Contact.tracing +
##     C6_Stay.at.home.requirements:H1_Public.information.campaigns +
##     C6_Stay.at.home.requirements:H2_Testing.policy + C6_Stay.at.home.requirements:H3_Contact.tracing
##     C7_Restrictions.on.internal.movement:H1_Public.information.campaigns +
##     C7_Restrictions.on.internal.movement:H3_Contact.tracing +
##     C8_International.travel.controls:H1_Public.information.campaigns +
##     C8_International.travel.controls:H2_Testing.policy + C8_International.travel.controls:H4_Emergenc
##     data = euroData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.971  -8.304   0.448   8.518  68.283
##
## Coefficients:
##                                                                               Estimate
## (Intercept)                                                                  4.821e+00
```

```
## active_per_pop                                                           -1.704e+03
## C1_School.closing                                                         4.758e+00
## C2_Workplace.closing                                                     -4.108e+01
## C3_Cancel.public.events                                                  -2.790e+01
## C4_Restrictions.on.gatherings                                             2.543e+01
## C5_Close.public.transport                                                -2.597e+01
## C6_Stay.at.home.requirements                                             -4.237e+01
## C7_Restrictions.on.internal.movement                                      5.747e+01
## C8_International.travel.controls                                         -7.199e+00
## H1_Public.information.campaigns                                           1.975e+00
## H2_Testing.policy                                                         8.471e-01
## H3_Contact.tracing                                                        3.825e+00
## H4_Emergency.investment.in.healthcare                                    -2.933e-09
## gdp_capita                                                               -1.346e-04
## pop_density                                                               5.833e-03
## C1_School.closing:H1_Public.information.campaigns                        -6.355e+00
## C1_School.closing:H2_Testing.policy                                       1.945e+00
## C1_School.closing:H3_Contact.tracing                                     -5.383e-01
## C1_School.closing:H4_Emergency.investment.in.healthcare                  -2.870e-09
## C2_Workplace.closing:H1_Public.information.campaigns                      1.340e+01
## C2_Workplace.closing:H2_Testing.policy                                    2.383e+00
## C2_Workplace.closing:H3_Contact.tracing                                   3.705e+00
## C2_Workplace.closing:H4_Emergency.investment.in.healthcare               -3.053e-09
## C3_Cancel.public.events:H1_Public.information.campaigns                   1.615e+01
## C3_Cancel.public.events:H2_Testing.policy                                -1.464e+00
## C3_Cancel.public.events:H3_Contact.tracing                               -2.764e+00
## C4_Restrictions.on.gatherings:H1_Public.information.campaigns            -1.079e+01
## C4_Restrictions.on.gatherings:H2_Testing.policy                          -1.828e+00
## C4_Restrictions.on.gatherings:H3_Contact.tracing                         -1.521e+00
## C5_Close.public.transport:H1_Public.information.campaigns                 1.190e+01
## C5_Close.public.transport:H2_Testing.policy                               1.392e+00
## C5_Close.public.transport:H3_Contact.tracing                             -2.545e+00
## C6_Stay.at.home.requirements:H1_Public.information.campaigns              1.400e+01
## C6_Stay.at.home.requirements:H2_Testing.policy                            2.684e+00
## C6_Stay.at.home.requirements:H3_Contact.tracing                           1.796e+00
## C7_Restrictions.on.internal.movement:H1_Public.information.campaigns     -3.309e+01
## C7_Restrictions.on.internal.movement:H3_Contact.tracing                   3.260e+00
## C8_International.travel.controls:H1_Public.information.campaigns          2.796e+00
## C8_International.travel.controls:H2_Testing.policy                        1.399e+00
## C8_International.travel.controls:H4_Emergency.investment.in.healthcare   4.396e-09
##                                                                          Std. Error
## (Intercept)                                                               1.020e+00
## active_per_pop                                                            9.357e+01
## C1_School.closing                                                         2.124e+00
## C2_Workplace.closing                                                      1.254e+01
## C3_Cancel.public.events                                                   2.845e+00
## C4_Restrictions.on.gatherings                                             7.512e+00
## C5_Close.public.transport                                                 1.448e+01
## C6_Stay.at.home.requirements                                              4.448e+00
## C7_Restrictions.on.internal.movement                                      1.405e+01
## C8_International.travel.controls                                          8.627e-01
## H1_Public.information.campaigns                                           6.286e-01
## H2_Testing.policy                                                         5.977e-01
## H3_Contact.tracing                                                        5.294e-01
```

```
## H4_Emergency.investment.in.healthcare                                    1.967e-09
## gdp_capita                                                               6.535e-06
## pop_density                                                              1.640e-03
## C1_School.closing:H1_Public.information.campaigns                        1.054e+00
## C1_School.closing:H2_Testing.policy                                      3.002e-01
## C1_School.closing:H3_Contact.tracing                                     3.176e-01
## C1_School.closing:H4_Emergency.investment.in.healthcare                  1.362e-09
## C2_Workplace.closing:H1_Public.information.campaigns                     6.291e+00
## C2_Workplace.closing:H2_Testing.policy                                   3.722e-01
## C2_Workplace.closing:H3_Contact.tracing                                  3.986e-01
## C2_Workplace.closing:H4_Emergency.investment.in.healthcare               1.412e-09
## C3_Cancel.public.events:H1_Public.information.campaigns                  1.372e+00
## C3_Cancel.public.events:H2_Testing.policy                                4.369e-01
## C3_Cancel.public.events:H3_Contact.tracing                               4.556e-01
## C4_Restrictions.on.gatherings:H1_Public.information.campaigns            3.749e+00
## C4_Restrictions.on.gatherings:H2_Testing.policy                          2.279e-01
## C4_Restrictions.on.gatherings:H3_Contact.tracing                         2.644e-01
## C5_Close.public.transport:H1_Public.information.campaigns                7.252e+00
## C5_Close.public.transport:H2_Testing.policy                              4.644e-01
## C5_Close.public.transport:H3_Contact.tracing                             5.009e-01
## C6_Stay.at.home.requirements:H1_Public.information.campaigns             2.196e+00
## C6_Stay.at.home.requirements:H2_Testing.policy                           4.049e-01
## C6_Stay.at.home.requirements:H3_Contact.tracing                          4.082e-01
## C7_Restrictions.on.internal.movement:H1_Public.information.campaigns     7.026e+00
## C7_Restrictions.on.internal.movement:H3_Contact.tracing                  3.879e-01
## C8_International.travel.controls:H1_Public.information.campaigns         4.498e-01
## C8_International.travel.controls:H2_Testing.policy                       2.012e-01
## C8_International.travel.controls:H4_Emergency.investment.in.healthcare  1.133e-09
##                                                                         t value
## (Intercept)                                                               4.729
## active_per_pop                                                          -18.212
## C1_School.closing                                                         2.241
## C2_Workplace.closing                                                     -3.277
## C3_Cancel.public.events                                                  -9.806
## C4_Restrictions.on.gatherings                                             3.385
## C5_Close.public.transport                                                -1.794
## C6_Stay.at.home.requirements                                             -9.527
## C7_Restrictions.on.internal.movement                                      4.090
## C8_International.travel.controls                                         -8.345
## H1_Public.information.campaigns                                          3.141
## H2_Testing.policy                                                         1.417
## H3_Contact.tracing                                                        7.225
## H4_Emergency.investment.in.healthcare                                    -1.491
## gdp_capita                                                              -20.600
## pop_density                                                               3.558
## C1_School.closing:H1_Public.information.campaigns                        -6.032
## C1_School.closing:H2_Testing.policy                                       6.478
## C1_School.closing:H3_Contact.tracing                                     -1.695
## C1_School.closing:H4_Emergency.investment.in.healthcare                  -2.107
## C2_Workplace.closing:H1_Public.information.campaigns                      2.130
## C2_Workplace.closing:H2_Testing.policy                                    6.403
## C2_Workplace.closing:H3_Contact.tracing                                   9.293
## C2_Workplace.closing:H4_Emergency.investment.in.healthcare               -2.162
## C3_Cancel.public.events:H1_Public.information.campaigns                  11.768
```

```
## C3_Cancel.public.events:H2_Testing.policy                            -3.350
## C3_Cancel.public.events:H3_Contact.tracing                           -6.068
## C4_Restrictions.on.gatherings:H1_Public.information.campaigns        -2.879
## C4_Restrictions.on.gatherings:H2_Testing.policy                      -8.022
## C4_Restrictions.on.gatherings:H3_Contact.tracing                     -5.752
## C5_Close.public.transport:H1_Public.information.campaigns             1.641
## C5_Close.public.transport:H2_Testing.policy                           2.996
## C5_Close.public.transport:H3_Contact.tracing                         -5.082
## C6_Stay.at.home.requirements:H1_Public.information.campaigns          6.374
## C6_Stay.at.home.requirements:H2_Testing.policy                        6.629
## C6_Stay.at.home.requirements:H3_Contact.tracing                       4.401
## C7_Restrictions.on.internal.movement:H1_Public.information.campaigns -4.709
## C7_Restrictions.on.internal.movement:H3_Contact.tracing               8.406
## C8_International.travel.controls:H1_Public.information.campaigns      6.216
## C8_International.travel.controls:H2_Testing.policy                    6.954
## C8_International.travel.controls:H4_Emergency.investment.in.healthcare 3.879
##                                                                     Pr(>|t|)
## (Intercept)                                                         2.30e-06
## active_per_pop                                                       < 2e-16
## C1_School.closing                                                   0.025074
## C2_Workplace.closing                                                0.001055
## C3_Cancel.public.events                                              < 2e-16
## C4_Restrictions.on.gatherings                                       0.000715
## C5_Close.public.transport                                           0.072841
## C6_Stay.at.home.requirements                                         < 2e-16
## C7_Restrictions.on.internal.movement                                4.36e-05
## C8_International.travel.controls                                     < 2e-16
## H1_Public.information.campaigns                                     0.001688
## H2_Testing.policy                                                   0.156457
## H3_Contact.tracing                                                  5.41e-13
## H4_Emergency.investment.in.healthcare                               0.136047
## gdp_capita                                                           < 2e-16
## pop_density                                                         0.000376
## C1_School.closing:H1_Public.information.campaigns                   1.68e-09
## C1_School.closing:H2_Testing.policy                                 9.76e-11
## C1_School.closing:H3_Contact.tracing                                0.090065
## C1_School.closing:H4_Emergency.investment.in.healthcare             0.035115
## C2_Workplace.closing:H1_Public.information.campaigns                0.033191
## C2_Workplace.closing:H2_Testing.policy                              1.60e-10
## C2_Workplace.closing:H3_Contact.tracing                              < 2e-16
## C2_Workplace.closing:H4_Emergency.investment.in.healthcare          0.030628
## C3_Cancel.public.events:H1_Public.information.campaigns              < 2e-16
## C3_Cancel.public.events:H2_Testing.policy                           0.000810
## C3_Cancel.public.events:H3_Contact.tracing                          1.35e-09
## C4_Restrictions.on.gatherings:H1_Public.information.campaigns       0.004003
## C4_Restrictions.on.gatherings:H2_Testing.policy                     1.17e-15
## C4_Restrictions.on.gatherings:H3_Contact.tracing                    9.13e-09
## C5_Close.public.transport:H1_Public.information.campaigns           0.100838
## C5_Close.public.transport:H2_Testing.policy                         0.002739
## C5_Close.public.transport:H3_Contact.tracing                        3.81e-07
## C6_Stay.at.home.requirements:H1_Public.information.campaigns        1.93e-10
## C6_Stay.at.home.requirements:H2_Testing.policy                      3.57e-11
## C6_Stay.at.home.requirements:H3_Contact.tracing                     1.09e-05
## C7_Restrictions.on.internal.movement:H1_Public.information.campaigns 2.53e-06
```

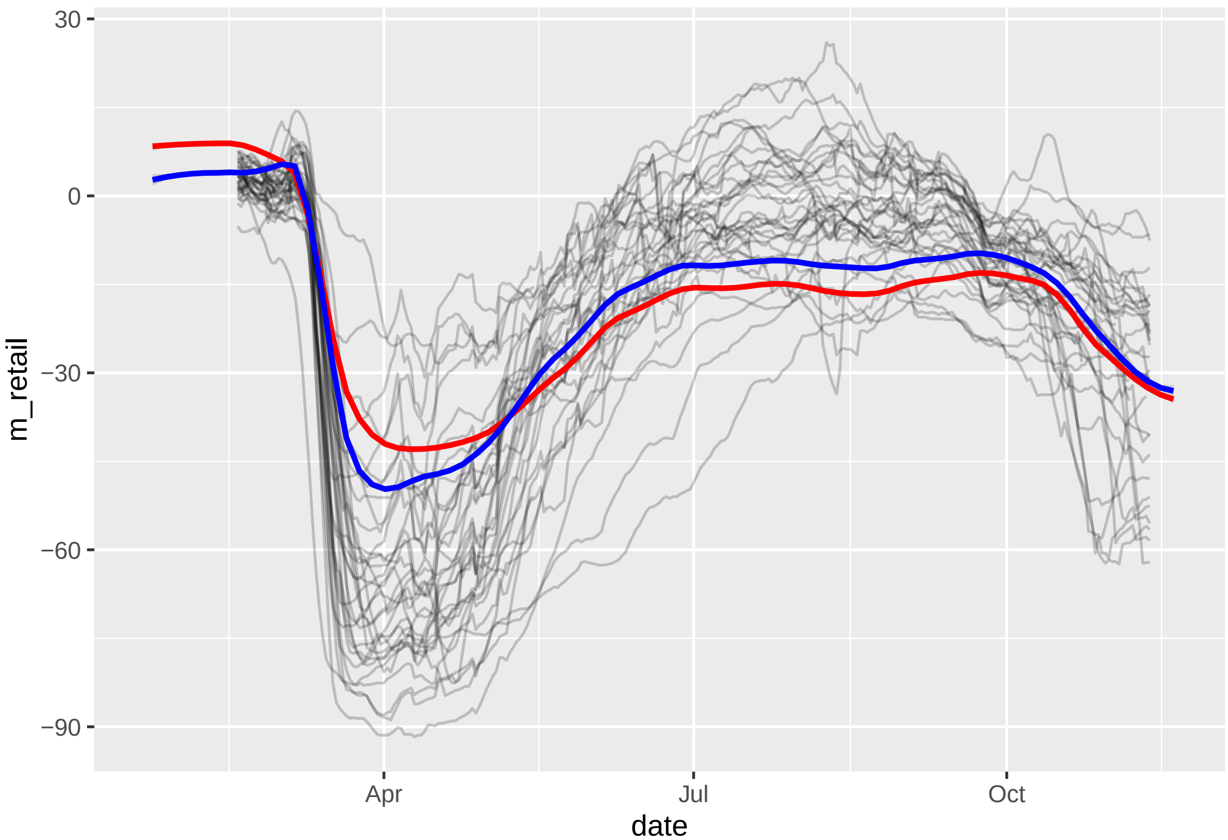
```
## C7_Restrictions.on.internal.movement:H3_Contact.tracing                  < 2e-16
## C8_International.travel.controls:H1_Public.information.campaigns       5.33e-10
## C8_International.travel.controls:H2_Testing.policy                     3.80e-12
## C8_International.travel.controls:H4_Emergency.investment.in.healthcare 0.000106
##
## (Intercept)                                                            ***
## active_per_pop                                                         ***
## C1_School.closing                                                      *
## C2_Workplace.closing                                                   **
## C3_Cancel.public.events                                                ***
## C4_Restrictions.on.gatherings                                          ***
## C5_Close.public.transport                                              .
## C6_Stay.at.home.requirements                                           ***
## C7_Restrictions.on.internal.movement                                   ***
## C8_International.travel.controls                                       ***
## H1_Public.information.campaigns                                        **
## H2_Testing.policy
## H3_Contact.tracing                                                     ***
## H4_Emergency.investment.in.healthcare
## gdp_capita                                                             ***
## pop_density                                                            ***
## C1_School.closing:H1_Public.information.campaigns                      ***
## C1_School.closing:H2_Testing.policy                                    ***
## C1_School.closing:H3_Contact.tracing                                   .
## C1_School.closing:H4_Emergency.investment.in.healthcare                *
## C2_Workplace.closing:H1_Public.information.campaigns                   *
## C2_Workplace.closing:H2_Testing.policy                                 ***
## C2_Workplace.closing:H3_Contact.tracing                                ***
## C2_Workplace.closing:H4_Emergency.investment.in.healthcare             *
## C3_Cancel.public.events:H1_Public.information.campaigns                ***
## C3_Cancel.public.events:H2_Testing.policy                              ***
## C3_Cancel.public.events:H3_Contact.tracing                             ***
## C4_Restrictions.on.gatherings:H1_Public.information.campaigns          **
## C4_Restrictions.on.gatherings:H2_Testing.policy                        ***
## C4_Restrictions.on.gatherings:H3_Contact.tracing                       ***
## C5_Close.public.transport:H1_Public.information.campaigns
## C5_Close.public.transport:H2_Testing.policy                            **
## C5_Close.public.transport:H3_Contact.tracing                           ***
## C6_Stay.at.home.requirements:H1_Public.information.campaigns           ***
## C6_Stay.at.home.requirements:H2_Testing.policy                         ***
## C6_Stay.at.home.requirements:H3_Contact.tracing                        ***
## C7_Restrictions.on.internal.movement:H1_Public.information.campaigns   ***
## C7_Restrictions.on.internal.movement:H3_Contact.tracing                ***
## C8_International.travel.controls:H1_Public.information.campaigns       ***
## C8_International.travel.controls:H2_Testing.policy                     ***
## C8_International.travel.controls:H4_Emergency.investment.in.healthcare ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.57 on 8956 degrees of freedom
##   (3131 observations deleted due to missingness)
## Multiple R-squared:  0.7002, Adjusted R-squared:  0.6988
## F-statistic: 522.9 on 40 and 8956 DF,  p-value: < 2.2e-16
```

Using our models we now try to create predictions as to outcome of the retail variable

```
Europe_daily_means2 <- euroData %>%
  group_by(date) %>%
  summarise_all(mean)

Europe_daily_means2$country <- "AVERAGE"

Europe_daily_means2$predicted_retail_recreation1 <- unname(predict(euro_rr_backward1, newdata = Europe_
Europe_daily_means2$predicted_retail_recreation2 <- unname(predict(euro_rr_backward2, newdata = Europe_
```

To visualise our predictive models we overlay them on the plot of the mean retail vs date

```
ggplot() + geom_line(aes(y=m_retail, x=date, group = country), data = euroData, alpha = 0.2) + geom_smo
```



We now repeat the same process for North America and South American.

3.1.3 North America simple

```
min_modelN1 <- lm(m_retail ~ 1, data = northData)
max_modelN1 <- lm(m_retail ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + (active_per_pop), data = northData)
```

```r
north_rr_backward1 <- step(max_modelN1, direction = 'backward',
scope = list('lower' = min_modelN1))
```

```
## Start:  AIC=19952.75
## m_retail ~ 1 + C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events +
##     C4_Restrictions.on.gatherings + C5_Close.public.transport +
##     C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
##     C8_International.travel.controls + (active_per_pop)
##
##                                       Df Sum of Sq    RSS   AIC
## - active_per_pop                       1        92 675011 19951
## <none>                                             674919 19953
## - C3_Cancel.public.events              1       439 675358 19953
## - C1_School.closing                    1      3277 678196 19969
## - C7_Restrictions.on.internal.movement  1      3536 678455 19971
## - C8_International.travel.controls     1     16699 691618 20045
## - C4_Restrictions.on.gatherings        1     30388 705307 20121
## - C5_Close.public.transport            1     41857 716776 20183
## - C2_Workplace.closing                 1    105759 780678 20513
## - C6_Stay.at.home.requirements         1    107519 782438 20521
##
## Step:  AIC=19951.28
## m_retail ~ C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events +
##     C4_Restrictions.on.gatherings + C5_Close.public.transport +
##     C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
##     C8_International.travel.controls
##
##                                       Df Sum of Sq    RSS   AIC
## <none>                                             675011 19951
## - C3_Cancel.public.events              1       469 675480 19952
## - C1_School.closing                    1      3200 678211 19968
## - C7_Restrictions.on.internal.movement  1      3503 678514 19969
## - C8_International.travel.controls     1     17353 692364 20047
## - C4_Restrictions.on.gatherings        1     30525 705536 20120
## - C5_Close.public.transport            1     41852 716863 20182
## - C2_Workplace.closing                 1    105669 780680 20511
## - C6_Stay.at.home.requirements         1    107645 782656 20520
```

```r
summary(north_rr_backward1)
```

```
##
## Call:
## lm(formula = m_retail ~ C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls,
##     data = northData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.260  -8.318   0.965  10.055  32.935
##
```

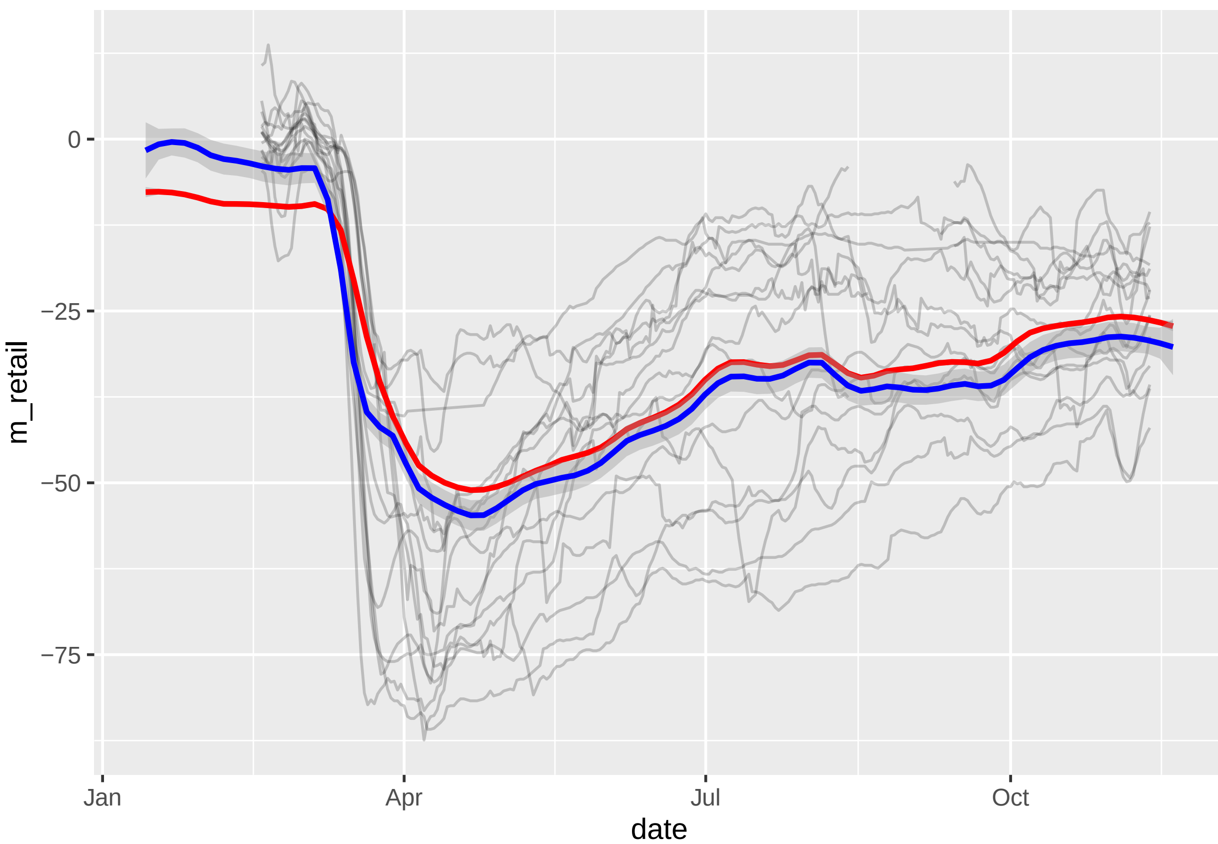
```
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          -7.6669     0.5599 -13.693  < 2e-16 ***
## C1_School.closing                    -1.8327     0.4289  -4.273 1.97e-05 ***
## C2_Workplace.closing                 -8.8860     0.3619 -24.553  < 2e-16 ***
## C3_Cancel.public.events               1.1805     0.7217   1.636    0.102
## C4_Restrictions.on.gatherings         3.8999     0.2955  13.196  < 2e-16 ***
## C5_Close.public.transport            -5.9075     0.3823 -15.452  < 2e-16 ***
## C6_Stay.at.home.requirements         -9.0983     0.3671 -24.782  < 2e-16 ***
## C7_Restrictions.on.internal.movement  1.6839     0.3767   4.471 8.02e-06 ***
## C8_International.travel.controls     -2.1220     0.2133  -9.950  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.24 on 3851 degrees of freedom
##   (1186 observations deleted due to missingness)
## Multiple R-squared:  0.607,  Adjusted R-squared:  0.6062
## F-statistic: 743.5 on 8 and 3851 DF,  p-value: < 2.2e-16
```

3.1.4 North America Complex

```
min_modelN2 <- lm(m_retail ~ 1, data = northData)
max_modelN2 <- lm(data = northData, m_retail ~ 1 + active_per_pop +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.tra

north_rr_backward2 <- step(max_modelN2, direction = 'backward',
scope = list('lower' = min_modelN2))
```

```
north_daily_means2 <- northData %>%
  group_by(date) %>%
  summarise_all(mean)

north_daily_means2$country <- "AVERAGE"

north_daily_means2$predicted_retail_recreation1 <- unname(predict(north_rr_backward1, newdata = north_da
north_daily_means2$predicted_retail_recreation2 <- unname(predict(north_rr_backward2, newdata = north_da
```

```
ggplot() + geom_line(aes(y=m_retail, x=date, group = country), data = northData, alpha = 0.2) + geom_smo
```



3.1.5 South America simple

```
min_modelS1 <- lm(m_retail ~ 1, data = southData)
max_modelS1 <- lm(m_retail ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + active_per_pop, data = southData)

south_rr_backward1 <- step(max_modelS1, direction = 'backward',
scope = list('lower' = min_modelS1))
```

```
## Start:  AIC=13694.29
## m_retail ~ 1 + C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events +
##     C4_Restrictions.on.gatherings + C5_Close.public.transport +
##     C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
##     C8_International.travel.controls + active_per_pop
##
##                                       Df Sum of Sq    RSS   AIC
## <none>                                             443322 13694
## - C7_Restrictions.on.internal.movement  1      1244 444566 13700
## - C3_Cancel.public.events               1      2234 445557 13706
## - C6_Stay.at.home.requirements          1      2892 446215 13710
## - C4_Restrictions.on.gatherings         1      3763 447086 13715
## - C5_Close.public.transport             1     12644 455966 13768
```

```
## - active_per_pop                          1     12935 456257 13769
## - C1_School.closing                       1     17521 460843 13796
## - C2_Workplace.closing                    1     21218 464540 13817
## - C8_International.travel.controls        1     98347 541670 14228
```

```r
summary(south_rr_backward1)
```

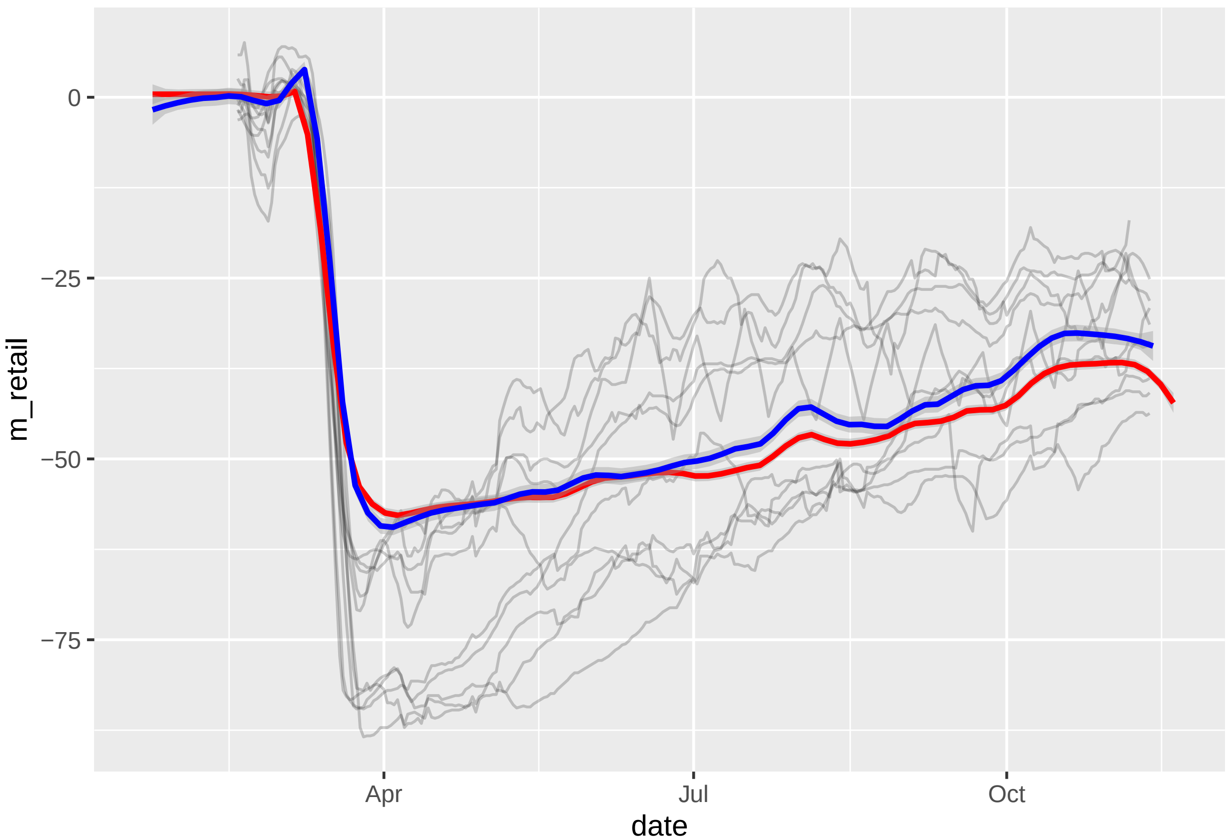
```
##
## Call:
## lm(formula = m_retail ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     active_per_pop, data = southData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.388  -7.206   0.422   9.508  32.851
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             0.9659     0.7856   1.230 0.218986
## C1_School.closing                      -5.7446     0.5596 -10.265  < 2e-16 ***
## C2_Workplace.closing                   -5.6252     0.4980 -11.296  < 2e-16 ***
## C3_Cancel.public.events                 3.0864     0.8420   3.665 0.000252 ***
## C4_Restrictions.on.gatherings           1.6957     0.3565   4.757 2.07e-06 ***
## C5_Close.public.transport              -3.5892     0.4116  -8.720  < 2e-16 ***
## C6_Stay.at.home.requirements           -2.4623     0.5904  -4.170 3.14e-05 ***
## C7_Restrictions.on.internal.movement   -1.9620     0.7173  -2.735 0.006277 **
## C8_International.travel.controls       -6.5681     0.2701 -24.319  < 2e-16 ***
## active_per_pop                       2394.7057   271.5207   8.820  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.9 on 2666 degrees of freedom
##   (929 observations deleted due to missingness)
## Multiple R-squared:  0.6545, Adjusted R-squared:  0.6533
## F-statistic: 561.1 on 9 and 2666 DF,  p-value: < 2.2e-16
```

3.1.6 South America complex

```r
min_modelS2 <- lm(m_retail ~ 1, data = southData)
max_modelS2 <- lm(data = southData, m_retail ~ 1 + active_per_pop +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.tra

south_rr_backward2 <- step(max_modelS2, direction = 'backward',
scope = list('lower' = min_modelS2))
```

```r
south_daily_means2 <- southData %>%
  group_by(date) %>%
  summarise_all(mean)

south_daily_means2$country <- "AVERAGE"

south_daily_means2$predicted_retail_recreation1 <- unname(predict(south_rr_backward1, newdata = south_da
south_daily_means2$predicted_retail_recreation2 <- unname(predict(south_rr_backward2, newdata = south_da
```

```r
ggplot() + geom_line(aes(y=m_retail, x=date, group = country), data = southData, alpha = 0.2) + geom_sm
```



3.1.7 Model evaluation

In order to evaluate the fit of each model we use the adjusted R^2 to see how well each of them fit

```r
PRESS <- function(linear.model) {
    pr <- residuals(linear.model)/(1 - lm.influence(linear.model)$hat)
    PRESS <- sum(pr^2)
    return(PRESS)
}
pred_r_squared <- function(linear.model) {
    lm.anova <- anova(linear.model)
    tss <- sum(lm.anova$"Sum Sq")
    # predictive R^2
```

```
    pred.r.squared <- 1 - PRESS(linear.model)/(tss)
    return(pred.r.squared)
}
```

```
eval_df <- data.frame(model = c("Simple Euro", "Complex Euro", "Simple North America", "Complex North Am
                      predictedRsq = c(pred_r_squared(euro_rr_backward1), pred_r_squared(euro_rr_backwar
                      adjustedRsq = c(summary(euro_rr_backward1)$adj.r.squared, summary(euro_rr_backward

kable(eval_df)
```

model	predictedRsq	adjustedRsq
Simple Euro	0.5603093	0.5608729
Complex Euro	0.5700079	0.6988455
Simple North America	0.6051114	0.6061850
Complex North America	-1823.6071976	0.7176047
Simple South America	0.6521390	0.6533074
Complex South America	0.7446616	0.7537349

3.1.8 Comparing simple models

We want to see how much these models differ and so we create a table with the coefficients of each explanatory variable. Here we can now see where the major similarities and differences are between each regions linear model.

```
summary_df <- data.frame(Europe = euro_rr_backward1$coefficients,
                         Europe_p_values = format(summary(euro_rr_backward1)$coefficients[,4], scientifi

                         North_America = c(north_rr_backward1$coefficients[1:3], "Excluded", north_rr_ba
                         North_p_values = c(format(summary(north_rr_backward1)$coefficients[1:3,4], scie

                         South_America = south_rr_backward1$coefficients,
                         South_p_values = format(summary(south_rr_backward1)$coefficients[,4], scientifi

row.names(summary_df) = c("Intercept", "C1", "C2", "C3", "C4", "C5", "C6", "C7", "C8", "Active per pop")
kable(summary_df)
```

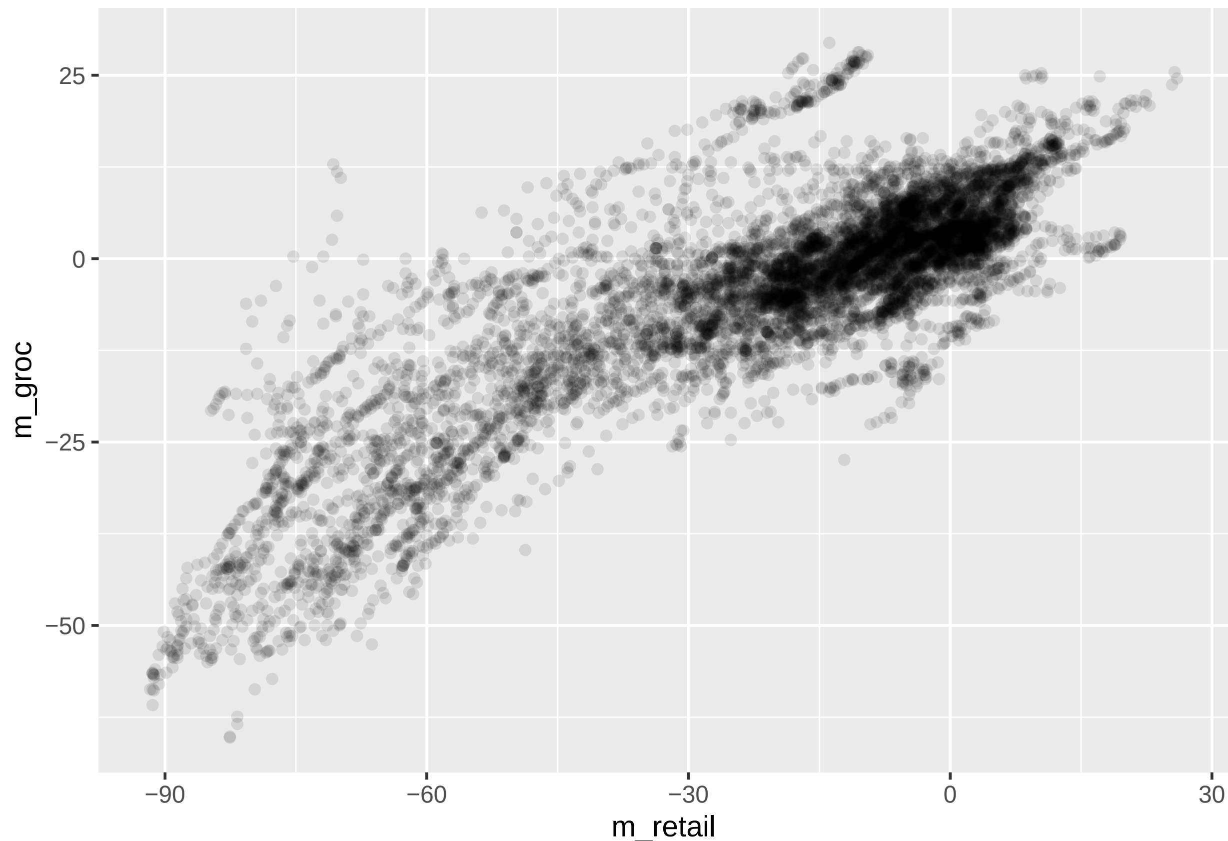
	Europe	Europe_p_values	North_America	North_p_values	South_America	South_p_values
Intercept	8.5862720	1.5e-68	-7.66693629520528	1.0e-41	0.9659284	2.2e-01
C1	-4.1043882	1.1e-71	-1.83271191256698	2.0e-05	-5.7445908	2.9e-24
C2	-3.1350025	9.4e-27	-8.88603940734178	8.4e-124	-5.6252373	6.2e-29
C3	-5.1191967	2.6e-60	Excluded	NA	3.0864380	2.5e-04
C4	-0.5640850	2.6e-03	1.18047398274546	1.0e-01	1.6956773	2.1e-06
C5	-1.9510243	1.4e-08	3.89985424936626	6.4e-39	-3.5892150	4.8e-18
C6	-7.8431404	1.4e-125	-5.90751725144384	2.6e-52	-2.4623315	3.1e-05
C7	-5.2103584	1.0e-72	-9.0983130353492	6.4e-126	-1.9619515	6.3e-03

	Europe	Europe_p_values	North_America	North_p_values	South_America	South_p_values
C8	0.9396019	8.1e-08	1.68392718408995	8.0e-06	-6.5680598	3.7e-118
Active per pop	-1023.7397458	3.6e-21	-2.12197190235636	4.8e-23	2394.7056823	2.0e-18

3.1.9 Grocery and pharmacy vs retail and recreation

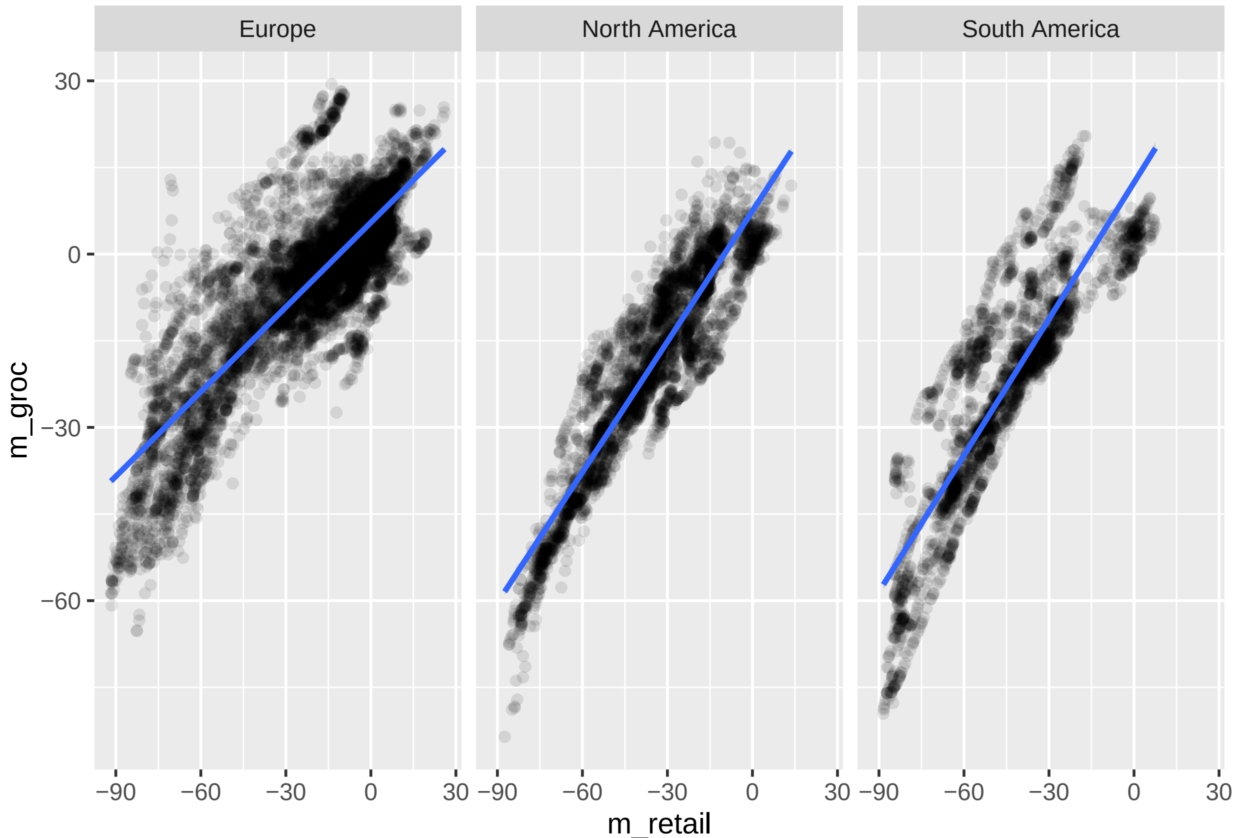
Exploring the data further we find that there seems to be some correlation between Retail and Grocery

```
ggplot(aes(x=m_retail, y=m_groc), data = euroData) + geom_point(alpha = 0.1)
```



So we plot Mean retail vs Mean grocery for each continent.

```
ggplot(aes(x=m_retail, y=m_groc), data = EAData) + geom_point(alpha = 0.1) + facet_wrap(~ continent) + g
```



We can see this is confirmed with a simple test.

```
cor(y = euroData$m_retail, x = euroData$m_groc, method = "pearson", use = "complete.obs")
```

```
## [1] 0.8341088
```

3.1.10 Workplace by predicted retail

So using the predictions from our complex linear models for Retail we create another linear model for each continent where mean of workplaces is the observed variable.

```
euroData$predicted_retail_recreation2 <- unname(predict(euro_rr_backward2, newdata = euroData))
northData$predicted_retail_recreation2 <- unname(predict(north_rr_backward2, newdata = northData))
southData$predicted_retail_recreation2 <- unname(predict(south_rr_backward2, newdata = southData))

euro_retail_workplaces_lm <- lm(m_groc ~ predicted_retail_recreation2, data = euroData)
summary(euro_retail_workplaces_lm)
```

```
##
## Call:
## lm(formula = m_groc ~ predicted_retail_recreation2, data = euroData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -47.238  -5.458   0.103   7.050  32.453
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 5.384019   0.156729   34.35   <2e-16 ***
## predicted_retail_recreation2 0.484625   0.005338   90.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.46 on 8946 degrees of freedom
##   (3180 observations deleted due to missingness)
## Multiple R-squared:  0.4795, Adjusted R-squared:  0.4794
## F-statistic:  8242 on 1 and 8946 DF,  p-value: < 2.2e-16
```

```r
north_retail_workplaces_lm <- lm(m_groc ~ predicted_retail_recreation2, data = northData)
summary(north_retail_workplaces_lm)
```

```
##
## Call:
## lm(formula = m_groc ~ predicted_retail_recreation2, data = northData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.471  -6.793  -0.158   7.620  35.413
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  6.67902    0.39568   16.88   <2e-16 ***
## predicted_retail_recreation2  0.73236    0.01002   73.07   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.18 on 3875 degrees of freedom
##   (1169 observations deleted due to missingness)
## Multiple R-squared:  0.5794, Adjusted R-squared:  0.5793
## F-statistic:  5339 on 1 and 3875 DF,  p-value: < 2.2e-16
```

```r
south_retail_workplaces_lm <- lm(m_groc ~ predicted_retail_recreation2, data = southData)
summary(south_retail_workplaces_lm)
```

```
##
## Call:
## lm(formula = m_groc ~ predicted_retail_recreation2, data = southData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.236  -8.688   0.326   7.867  37.535
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.49888    0.68026   15.43   <2e-16 ***
## predicted_retail_recreation2  0.74528    0.01393   53.51   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.73 on 2674 degrees of freedom
##   (929 observations deleted due to missingness)
## Multiple R-squared:  0.5171, Adjusted R-squared:  0.5169
## F-statistic:  2863 on 1 and 2674 DF,  p-value: < 2.2e-16
```

3.2 Workplaces

Here we follow the same modeling and analysis process as for Retail but instead for Workplaces

3.2.1 Europe simple

```
Wmin_modelE1 <- lm(m_workplaces ~ 1, data = euroData)
Wmax_modelE1 <- lm(m_workplaces ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + (active_per_pop), data = euroData)

euro_w_backward1 <- step(Wmax_modelE1, direction = 'backward',
scope = list('lower' = Wmin_modelE1))
```

```
## Start:  AIC=42140.3
## m_workplaces ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     (active_per_pop)
##
##                                         Df Sum of Sq     RSS   AIC
## <none>                                               979558 42140
## - C3_Cancel.public.events                1       334  979891 42141
## - C8_International.travel.controls       1      3037  982595 42166
## - C5_Close.public.transport              1      6148  985705 42194
## - active_per_pop                         1      6264  985821 42196
## - C7_Restrictions.on.internal.movement   1     10305  989863 42232
## - C4_Restrictions.on.gatherings          1     13126  992684 42258
## - C1_School.closing                      1     28926 1008484 42400
## - C2_Workplace.closing                   1     49024 1028581 42577
## - C6_Stay.at.home.requirements           1     54108 1033666 42621
```

```
summary(euro_w_backward1)
```

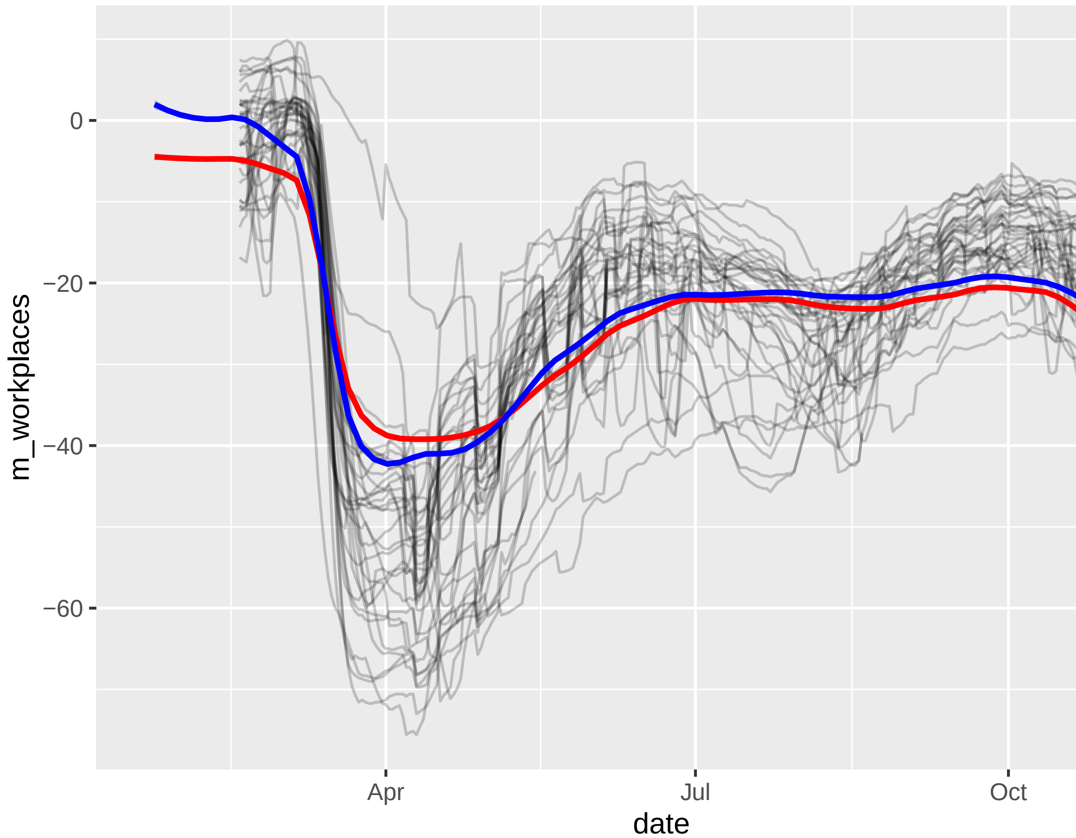
```
##
## Call:
## lm(formula = m_workplaces ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     (active_per_pop), data = euroData)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.437  -6.642   1.358   7.172  38.213
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             -4.3582     0.3103 -14.043  < 2e-16 ***
## C1_School.closing                       -2.3638     0.1453 -16.272  < 2e-16 ***
## C2_Workplace.closing                    -3.9495     0.1864 -21.183  < 2e-16 ***
## C3_Cancel.public.events                 -0.3459     0.1979  -1.747   0.0806 .
## C4_Restrictions.on.gatherings           -1.3143     0.1199 -10.961  < 2e-16 ***
## C5_Close.public.transport               -1.6450     0.2193  -7.501 6.92e-14 ***
## C6_Stay.at.home.requirements            -4.6106     0.2072 -22.254  < 2e-16 ***
## C7_Restrictions.on.internal.movement  -1.7750     0.1828  -9.712  < 2e-16 ***
## C8_International.travel.controls        -0.5889     0.1117  -5.273 1.38e-07 ***
## active_per_pop                         522.3048    68.9809   7.572 4.05e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.45 on 8966 degrees of freedom
##   (3152 observations deleted due to missingness)
## Multiple R-squared:  0.5576, Adjusted R-squared:  0.5572
## F-statistic:  1256 on 9 and 8966 DF,  p-value: < 2.2e-16
```

```
Wmin_modelE2 <- lm(m_workplaces ~ 1, data = euroData)
Wmax_modelE2 <- lm(data = euroData, m_workplaces ~ 1 + (active_per_pop) +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.trac

euro_w_backward2 <- step(Wmax_modelE2, direction = 'backward',
scope = list('lower' = Wmin_modelE2))
```

```
Europe_daily_means2 <- euroData %>%
  group_by(date) %>%
  summarise_all(mean)

Europe_daily_means2$country <- "AVERAGE"

Europe_daily_means2$predicted_wp1 <- unname(predict(euro_w_backward1, newdata = Europe_daily_means2))
Europe_daily_means2$predicted_wp2 <- unname(predict(euro_w_backward2, newdata = Europe_daily_means2))
```

```
ggplot() + geom_line(aes(y=m_workplaces, x=date, group = country), data = euroData, alpha = 0.2) + geom_
```



3.2.1.1 Europe complex

3.2.2 North America simple

```
Wmin_modelN1 <- lm(m_workplaces ~ 1, data = northData)
Wmax_modelN1 <- lm(m_workplaces ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + active_per_pop, data = northData)

north_w_backward1 <- step(Wmax_modelN1, direction = 'backward',
scope = list('lower' = Wmin_modelN1))
```

```
## Start:  AIC=18323.51
## m_workplaces ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     active_per_pop
##
##                                        Df Sum of Sq    RSS   AIC
## - C7_Restrictions.on.internal.movement  1         1 417166 18322
## <none>                                              417165 18324
## - active_per_pop                        1       238 417403 18324
## - C4_Restrictions.on.gatherings         1       659 417824 18328
## - C3_Cancel.public.events               1      1721 418887 18338
```

```
## - C5_Close.public.transport            1      5401 422566 18372
## - C1_School.closing                    1      7552 424718 18392
## - C8_International.travel.controls     1     15439 432605 18464
## - C6_Stay.at.home.requirements         1     38062 455227 18664
## - C2_Workplace.closing                 1     54383 471548 18802
##
## Step:  AIC=18321.51
## m_workplaces ~ C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events +
##     C4_Restrictions.on.gatherings + C5_Close.public.transport +
##     C6_Stay.at.home.requirements + C8_International.travel.controls +
##     active_per_pop
##
##                                    Df Sum of Sq    RSS   AIC
## <none>                                          417166 18322
## - active_per_pop                    1       237 417403 18322
## - C4_Restrictions.on.gatherings     1       659 417825 18326
## - C3_Cancel.public.events           1      1790 418956 18336
## - C5_Close.public.transport         1      5492 422658 18371
## - C1_School.closing                 1      7552 424718 18390
## - C8_International.travel.controls  1     15803 432969 18465
## - C6_Stay.at.home.requirements      1     40971 458137 18687
## - C2_Workplace.closing              1     55049 472215 18806
```

```
summary(north_w_backward1)
```

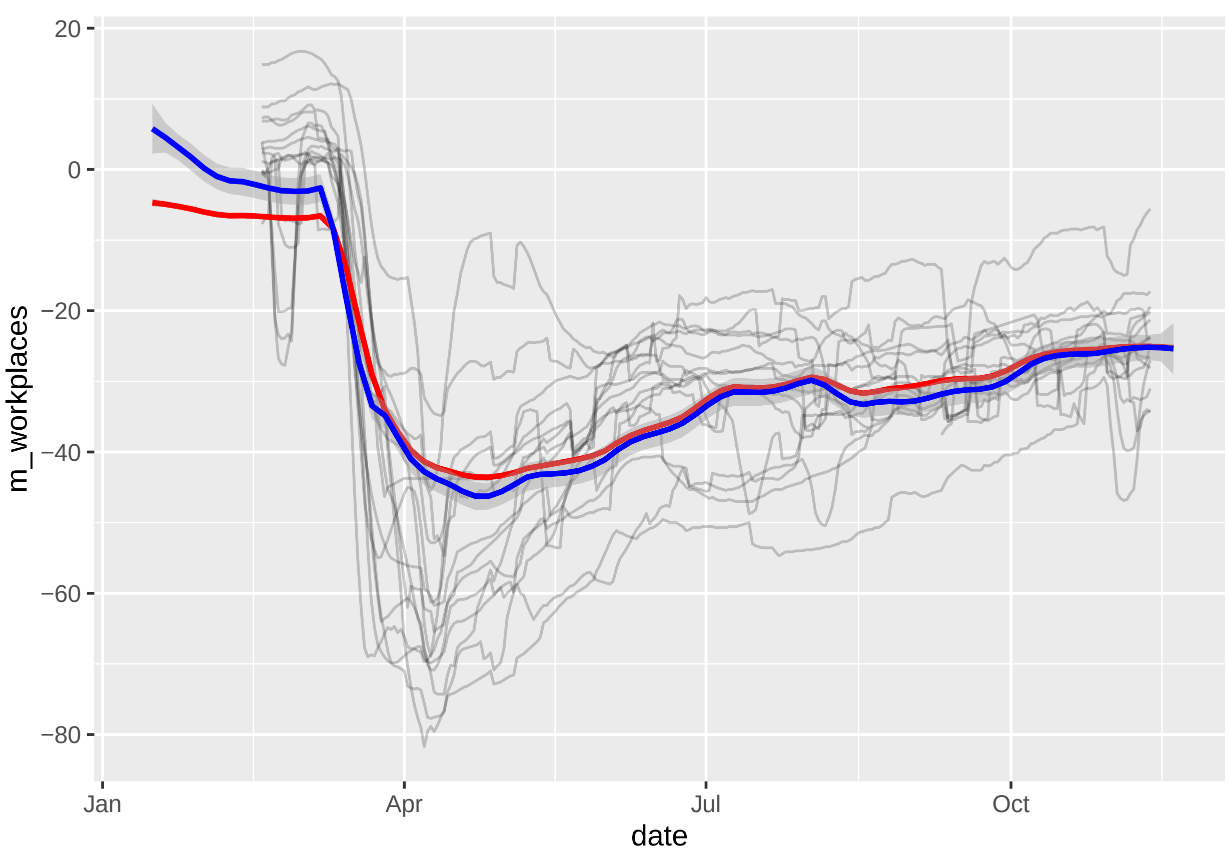
```
##
## Call:
## lm(formula = m_workplaces ~ C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C8_International.travel.controls + active_per_pop, data = northData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.112  -5.701   1.014   6.990  36.978
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -4.8534     0.4291 -11.310  < 2e-16 ***
## C1_School.closing                 -2.7841     0.3308  -8.416  < 2e-16 ***
## C2_Workplace.closing              -6.3621     0.2800 -22.724  < 2e-16 ***
## C3_Cancel.public.events            2.2430     0.5474   4.097 4.26e-05 ***
## C4_Restrictions.on.gatherings      0.5639     0.2268   2.486    0.013 *
## C5_Close.public.transport         -2.0923     0.2915  -7.177 8.46e-13 ***
## C6_Stay.at.home.requirements      -5.3309     0.2719 -19.604  < 2e-16 ***
## C8_International.travel.controls  -2.0172     0.1657 -12.175  < 2e-16 ***
## active_per_pop                   336.1483   225.3155   1.492    0.136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.33 on 3913 degrees of freedom
##   (1124 observations deleted due to missingness)
## Multiple R-squared:  0.6364, Adjusted R-squared:  0.6357
## F-statistic: 856.1 on 8 and 3913 DF,  p-value: < 2.2e-16
```

3.2.3 North America complex

```
Wmin_modelN2 <- lm(m_workplaces ~ 1, data = northData)
Wmax_modelN2 <- lm(data = northData, m_workplaces ~ 1 + active_per_pop +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.trac

north_w_backward2 <- step(Wmax_modelN2, direction = 'backward',
scope = list('lower' = Wmin_modelN2))
```

```
north_daily_means2 <- northData %>%
  group_by(date) %>%
  summarise_all(mean)

north_daily_means2$country <- "AVERAGE"

north_daily_means2$predicted_wp1 <- unname(predict(north_w_backward1, newdata = north_daily_means2))
north_daily_means2$predicted_wp2 <- unname(predict(north_w_backward2, newdata = north_daily_means2))
```

```
ggplot() + geom_line(aes(y=m_workplaces, x=date, group = country), data = northData, alpha = 0.2) + geom
```



3.2.4 South America simple

```
Wmin_modelS1 <- lm(m_workplaces ~ 1, data = southData)
Wmax_modelS1 <- lm(m_workplaces ~ 1 +
C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls + active_per_pop, data = southData)

south_w_backward1 <- step(Wmax_modelS1, direction = 'backward',
scope = list('lower' = Wmin_modelS1))
```

```
## Start:  AIC=14366.79
## m_workplaces ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     active_per_pop
##
##                                        Df Sum of Sq    RSS   AIC
## <none>                                              569982 14367
## - C4_Restrictions.on.gatherings         1       959 570940 14369
## - C6_Stay.at.home.requirements          1      1968 571949 14374
## - C7_Restrictions.on.internal.movement  1      3270 573251 14380
## - C5_Close.public.transport             1      3346 573327 14380
## - C3_Cancel.public.events               1     11543 581524 14418
## - C2_Workplace.closing                  1     22662 592644 14469
## - C1_School.closing                     1     28747 598728 14496
## - active_per_pop                        1     30741 600723 14505
## - C8_International.travel.controls      1     38018 608000 14538
```

```
summary(south_w_backward1)
```

```
##
## Call:
## lm(formula = m_workplaces ~ 1 + C1_School.closing + C2_Workplace.closing +
##     C3_Cancel.public.events + C4_Restrictions.on.gatherings +
##     C5_Close.public.transport + C6_Stay.at.home.requirements +
##     C7_Restrictions.on.internal.movement + C8_International.travel.controls +
##     active_per_pop, data = southData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -53.83 -10.11   0.90  10.85  35.24
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             7.7801     0.8908   8.734  < 2e-16 ***
## C1_School.closing                      -7.3583     0.6346 -11.596  < 2e-16 ***
## C2_Workplace.closing                   -5.8136     0.5647 -10.296  < 2e-16 ***
## C3_Cancel.public.events                 7.0156     0.9548   7.348 2.66e-13 ***
## C4_Restrictions.on.gatherings           0.8558     0.4042   2.117  0.03431 *
## C5_Close.public.transport              -1.8463     0.4667  -3.956 7.83e-05 ***
```

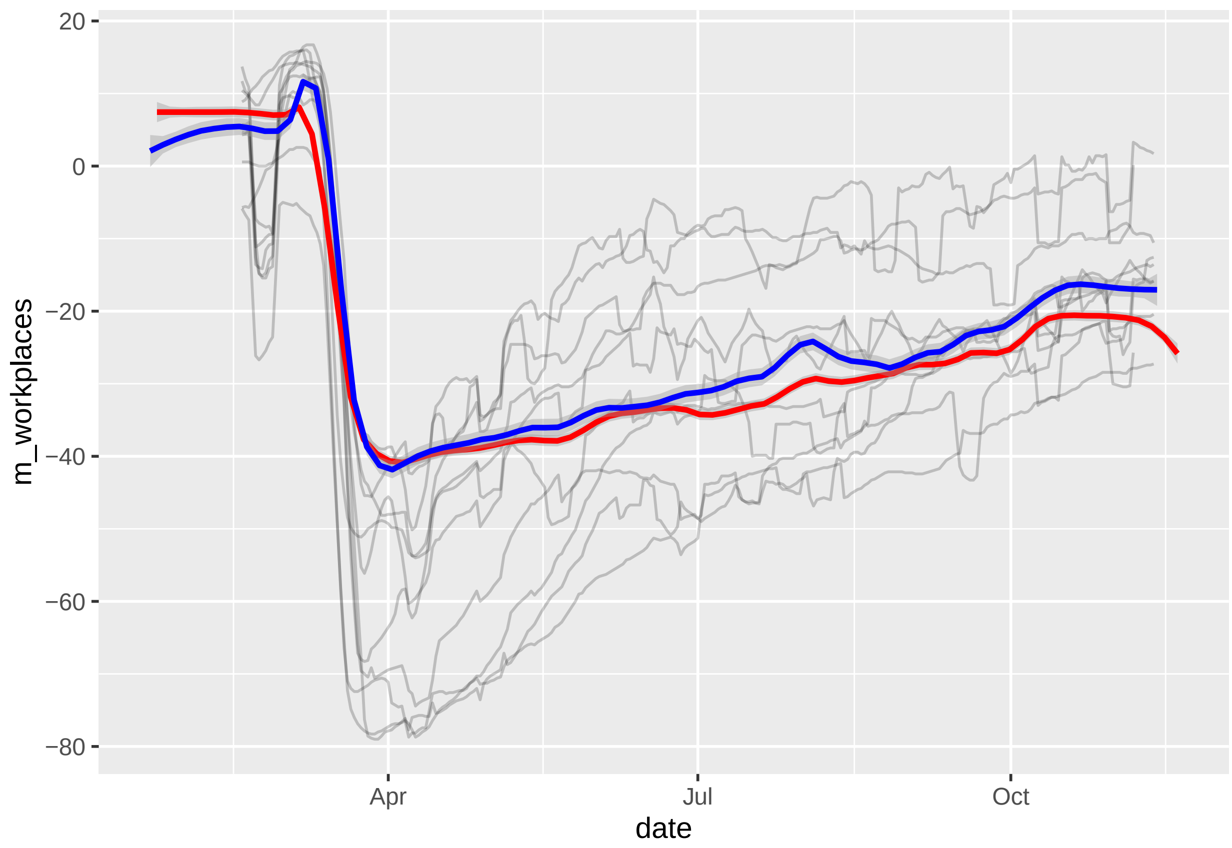
```
## C6_Stay.at.home.requirements           -2.0311     0.6695  -3.034  0.00244 **
## C7_Restrictions.on.internal.movement   -3.1808     0.8134  -3.911 9.43e-05 ***
## C8_International.travel.controls       -4.0837     0.3062 -13.335  < 2e-16 ***
## active_per_pop                       3691.7623   307.8744  11.991  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.62 on 2666 degrees of freedom
##   (929 observations deleted due to missingness)
## Multiple R-squared:  0.5109, Adjusted R-squared:  0.5093
## F-statistic: 309.4 on 9 and 2666 DF,  p-value: < 2.2e-16
```

3.2.5 South America Complex

```
Wmin_modelS2 <- lm(m_workplaces ~ 1, data = southData)
Wmax_modelS2 <- lm(data = southData, m_workplaces ~ 1 + active_per_pop +
(C1_School.closing + C2_Workplace.closing + C3_Cancel.public.events + C4_Restrictions.on.gatherings +
C5_Close.public.transport + C6_Stay.at.home.requirements + C7_Restrictions.on.internal.movement +
C8_International.travel.controls)*(H1_Public.information.campaigns + H2_Testing.policy + H3_Contact.tra

south_w_backward2 <- step(Wmax_modelS2, direction = 'backward',
scope = list('lower' = Wmin_modelS2))
```

```
south_daily_means2 <- southData %>%
  group_by(date) %>%
  summarise_all(mean)

south_daily_means2$country <- "AVERAGE"

south_daily_means2$predicted_wp1 <- unname(predict(south_w_backward1, newdata = south_daily_means2))
south_daily_means2$predicted_wp2 <- unname(predict(south_w_backward2, newdata = south_daily_means2))
```

```
ggplot() + geom_line(aes(y=m_workplaces, x=date, group = country), data = southData, alpha = 0.2) + geom
```



3.2.6 Comparing simple models

```
summary_df1 <- data.frame(Europe = c(euro_w_backward1$coefficients),
                          Europe_p_values = c(format(summary(euro_w_backward1)$coefficients[,4],
                                                     scientific = T, digits = 2)),

                          North_America = c(north_w_backward1$coefficients[1:7],
                                            "Excluded",
                                            north_w_backward1$coefficients[8:9]),
                          North_p_values = c(format(summary(north_w_backward1)$coefficients[1:7,4], scie

                          South_America = south_w_backward1$coefficients,
                          South_p_values = format(summary(south_w_backward1)$coefficients[,4], scientifi
row.names(summary_df1) = c("Intercept", "C1", "C2", "C3", "C4", "C5", "C6", "C7", "C8", "Active per pop
kable(summary_df1)
```

	Europe	Europe_p_values	North_America	North_p_values	South_America	South_p_values
Intercept	-4.3582474	2.5e-44	-4.85343679745938	3.3e-29	7.7801488	4.3e-18
C1	-2.3637889	1.1e-58	-2.78410247228975	5.4e-17	-7.3582576	2.3e-30
C2	-3.9495067	3.2e-97	-6.36208267703902	1.8e-107	-5.8135797	2.1e-24

	Europe	Europe_p_values	North_America	North_p_values	South_America	South_p_values
C3	-0.3458809	8.1e-02	2.24304062620445	4.3e-05	7.0155776	2.7e-13
C4	-1.3142903	8.8e-28	0.56386487036465	1.3e-02	0.8558271	3.4e-02
C5	-1.6450285	6.9e-14	-2.0922580493741	8.5e-13	-1.8462609	7.8e-05
C6	-4.6105983	7.7e-107	-5.33090743097943	1.1e-81	-2.0310857	2.4e-03
C7	-1.7749503	3.4e-22	Excluded	NA	-3.1808498	9.4e-05
C8	-0.5889418	1.4e-07	-2.01717468101061	1.7e-33	-4.0836893	2.6e-39
Active per pop	522.3047568	4.0e-14	336.148305055131	1.4e-01	3691.7622573	2.6e-32

We look to see if there is any correlation between Retail and Workplaces, as there was for Retail and Grocery

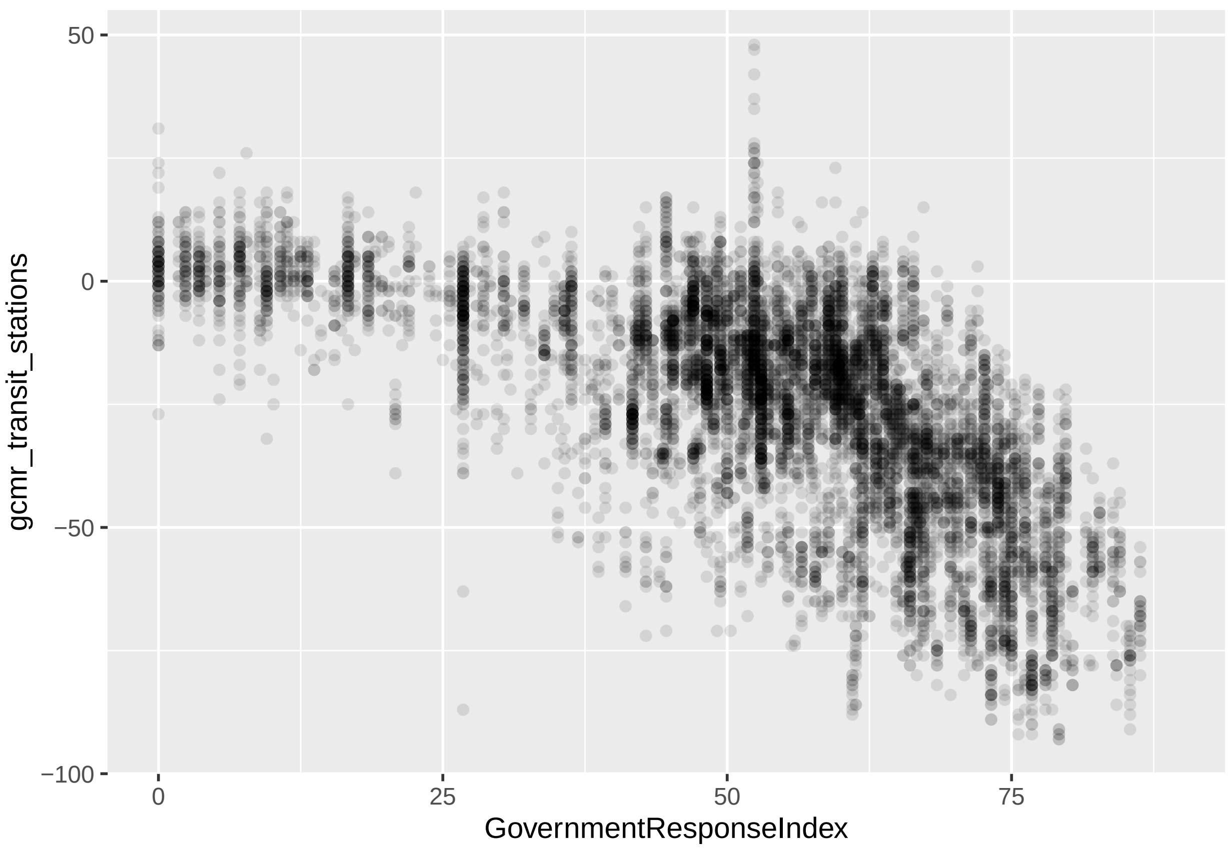
```
cor(y = euroData$m_retail, x = euroData$m_workplaces, method = "pearson", use = "complete.obs")
```

```
## [1] 0.8161911
```

3.3 Transit

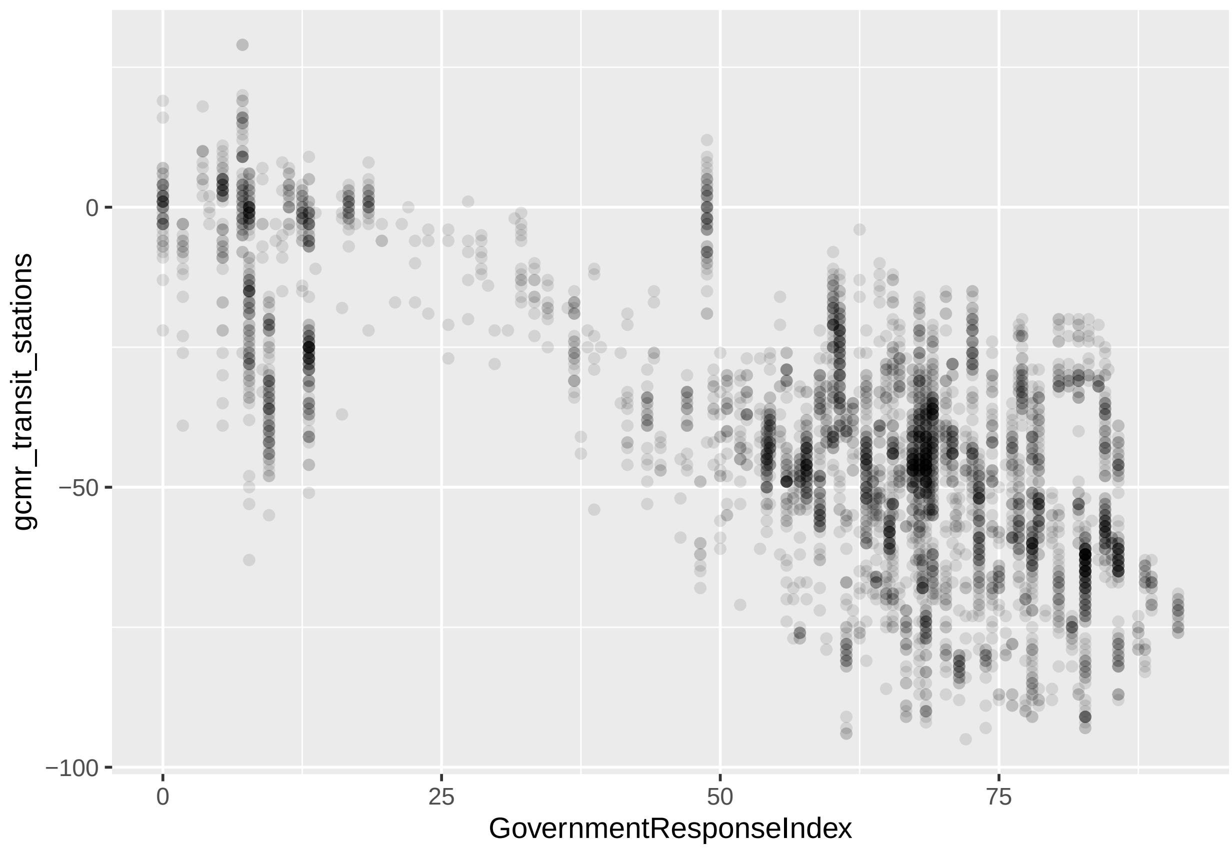
3.3.1 Europe

```
ggplot(data = euroData, aes(x = GovernmentResponseIndex, y = gcmr_transit_stations)) + geom_point(alpha
```



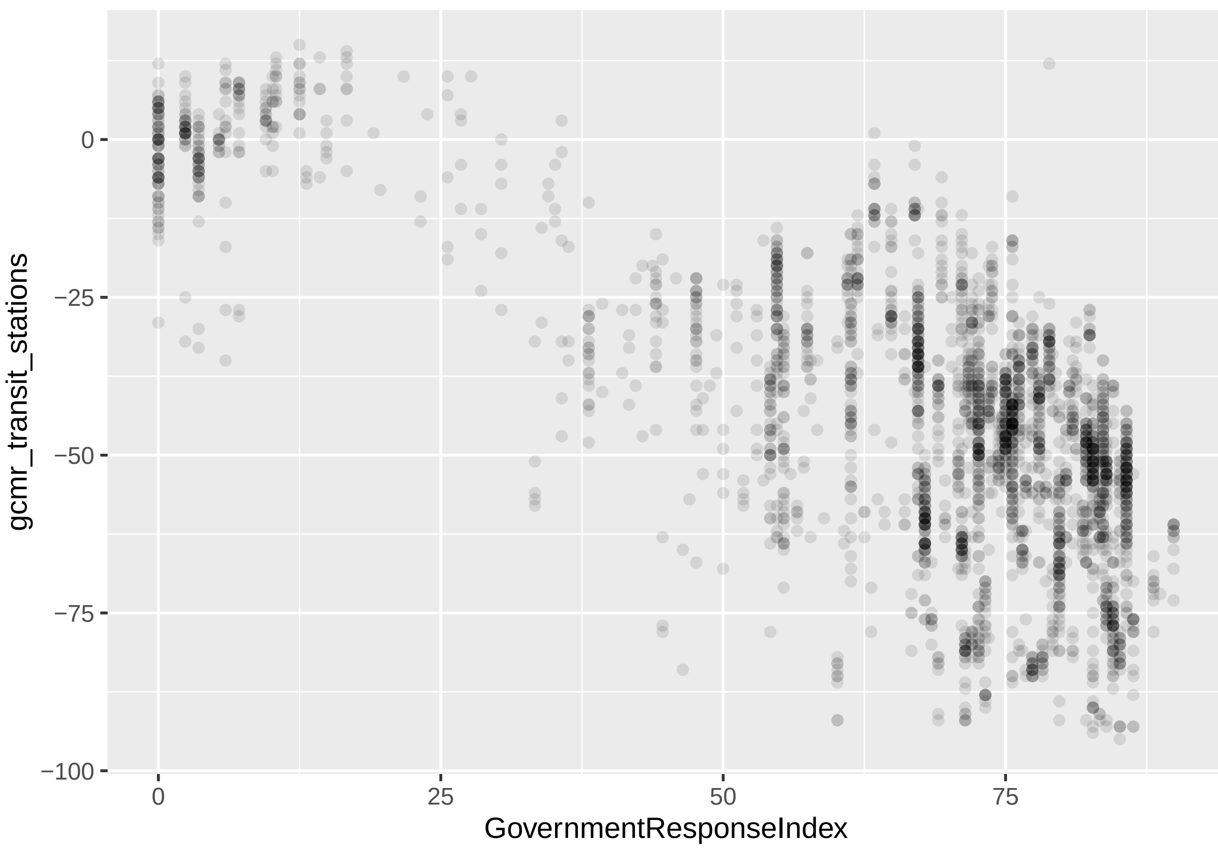
3.3.2 North America

```
ggplot(data = northData, aes(x = GovernmentResponseIndex, y = gcmr_transit_stations)) + geom_point(alpha
```



3.3.3 South America

```
ggplot(data = southData, aes(x = GovernmentResponseIndex, y = gcmr_transit_stations)) + geom_point(alpha
```



4 Survival Analysis

Survival analysis typically aims to identify the likelihood of entering a small number of states. The below example examines the transition from lowest retail activity in a country to its baseline, following a lockdown-induced crash.

4.1 Retail

4.1.1 Survival Objects

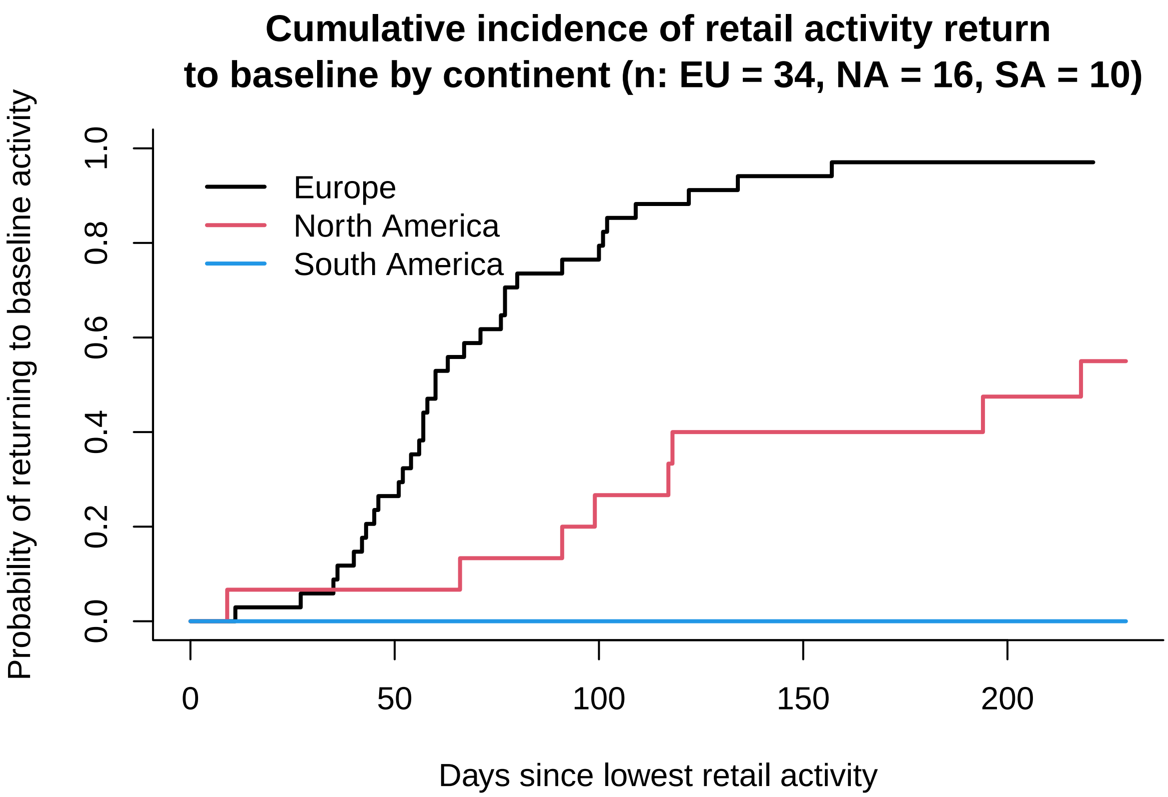
```
minretail <- EAData %>% group_by(country) %>% slice_min(gcmr_retail_recreation) %>% distinct(date)
names(minretail) <- c("country", "mindate")
euna <- left_join(EAData, minretail)

normalretail <- euna %>%
  group_by(country) %>%
  filter(date > mindate, gcmr_retail_recreation > -8) %>% # the next date where retail greater than -8%
  slice_min(gcmr_retail_recreation) %>%
  slice_min(date) %>%
  distinct(date)
names(normalretail) <- c("country", "maxdate")
```

```r
minmax<-left_join(minretail,normalretail)
minmax$mindate <- ymd(minmax$mindate)
minmax$maxdate <- ymd(minmax$maxdate)
minmax <- minmax %>% mutate(
  etime = ifelse(!is.na(maxdate),
                 as.duration(mindate %--% maxdate) / ddays(1),
                 as.duration(mindate %--% ymd("2020-11-19") / ddays(1))),
  event = ifelse(!is.na(maxdate),
                 1,
                 0)
  )

minmax <- left_join(minmax, select(euna, country, continent), by = "country")
surv <- minmax %>% distinct(etime, event, continent) %>% slice_min(etime)

surv$continent <- case_when(
  surv$continent == "Europe" ~ "EU",
  surv$continent == "North America" ~ "NA",
  surv$continent == "South America" ~ "SA"
)
```

4.1.2 Cumulative incidence of hazard of attaining baseline of retail

With the survival object ready, the cumulative incidence of re-attaining baseline for countries can be visualised.

```r
cumincfit <- cuminc(
  ftime = surv$etime,
  fstatus = surv$event,
  group = surv$continent
)

plot(cumincfit,
     col = c(1,2,4),
     lty = c(1,1,1),
     lwd = 2,
     wh = c(0,3),
     xlab = "Days since lowest retail activity",
     ylab = "Probability of returning to baseline activity",
     main = "Cumulative incidence of retail activity return\n to baseline by continent (n: EU = 34, NA =
legend(-3, 1, c("Europe", "North America", "South America"),
       col = c(1,2,4),
       lwd = 2,
       box.lty = 0,
       bty = 'y')
```



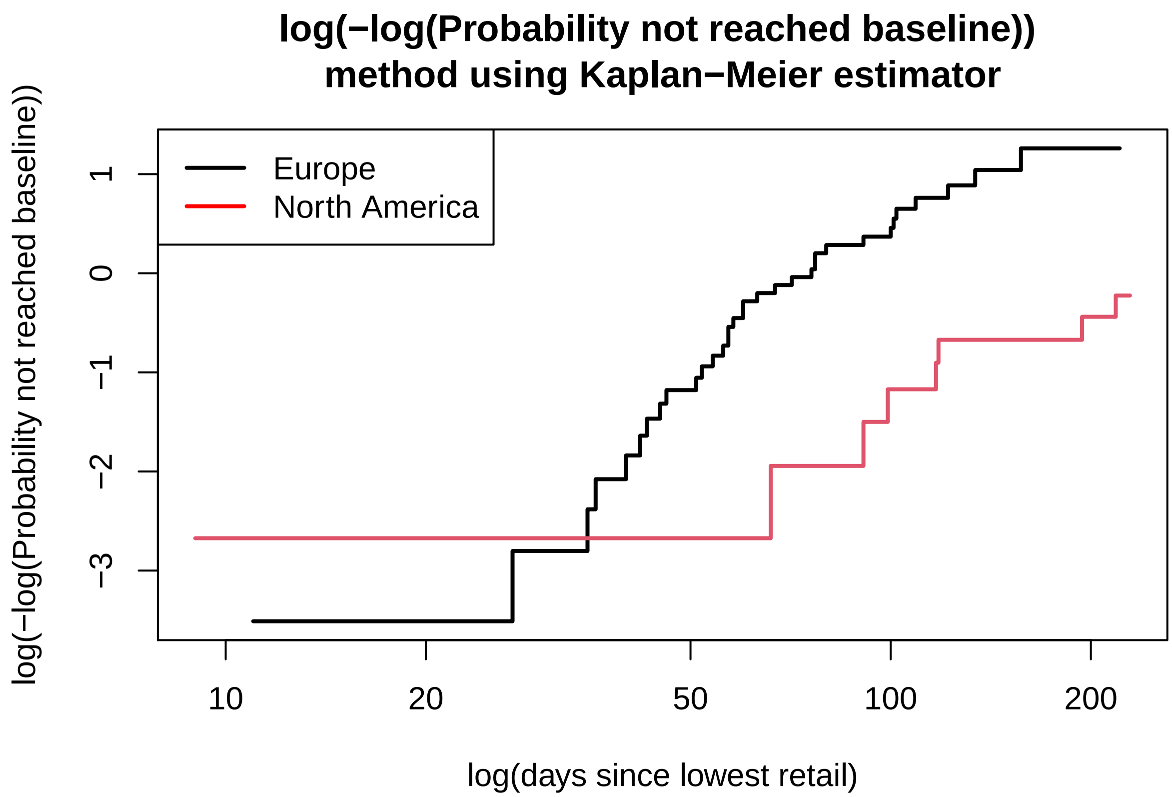
The above cumulative incidence plot of the return to baseline retail activity highlights that European countries were quicker to return to baseline activity than North America. No countries in South America have yet returned to baseline retail activity. South American businesses may be seeing long-term depression in activity, while European business activity returned to baseline in 80% of countries within 100 days.

Cannot be assumed hazard of EU vs NA represents a proportional hazard to the risk of returning to baseline activity.

4.1.3 Checking proportional hazards assumption in retail/recreation

Proportional hazards on a log(-log) plot should be parallel.

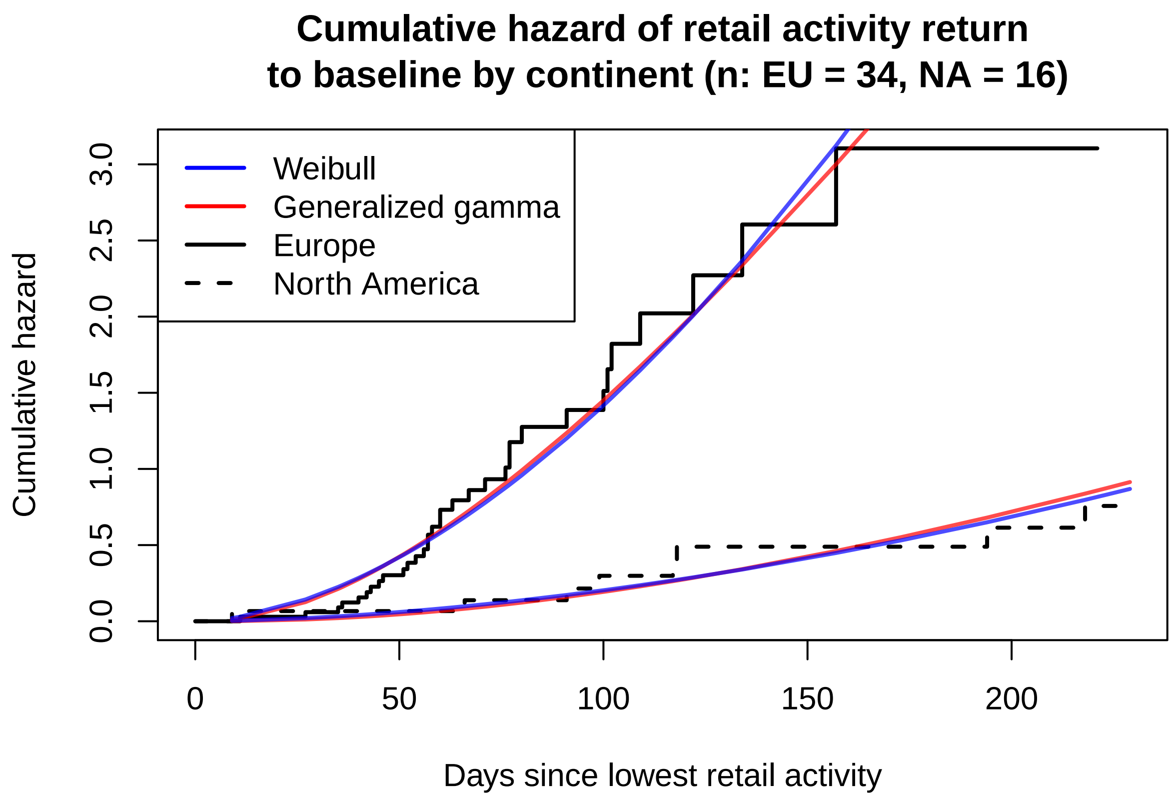
```
surv_euna <- surv %>% filter(continent != "SA")
retailfit_euna <- survfit(Surv(etime, event) ~ continent, surv_euna)
plot(retailfit_euna,
     col = c(1,2),
     lwd = 2,
     fun = "cloglog",
     xlab = "log(days since lowest retail)",
     ylab = "log(-log(Probability not reached baseline))",
     main = "log(-log(Probability not reached baseline)) \nmethod using Kaplan-Meier estimator")
legend("topleft", lty = c(1,1), lwd = c(2,2), col = c("black", "red"), c("Europe", "North America"))
```



4.1.4 Modelling non-proportional hazards

Proportional hazards assumption not met. Therefore must consider Accelerated Failure Time (AFT) models such as generalised gamma and Weibull.

```
gg <- flexsurvreg(Surv(surv_euna$etime, surv_euna$event) ~ continent, data = surv_euna, dist="gengamma"
wb <- flexsurvreg(Surv(surv_euna$etime, surv_euna$event) ~ continent, data = surv_euna, dist="weibull")

plot(retailfit_euna,
     cumhaz = TRUE,
     lty = c(1,2),
     lwd = 2,
     xlab = "Days since lowest retail activity",
     ylab = "Cumulative hazard",
     main = "Cumulative hazard of retail activity return\n to baseline by continent (n: EU = 34, NA = 1
lines(gg,
      col = rgb(1,0,0,0.7), ci = FALSE,
      type = "cumhaz")
lines(wb,
      col=rgb(0,0,1,0.7),
      type = "cumhaz",
      ci=FALSE)
legend("topleft", lty=c(1,1,1,2), lwd=c(2,2), col = c("blue","red","black","black"), c("Weibull","Genera
```

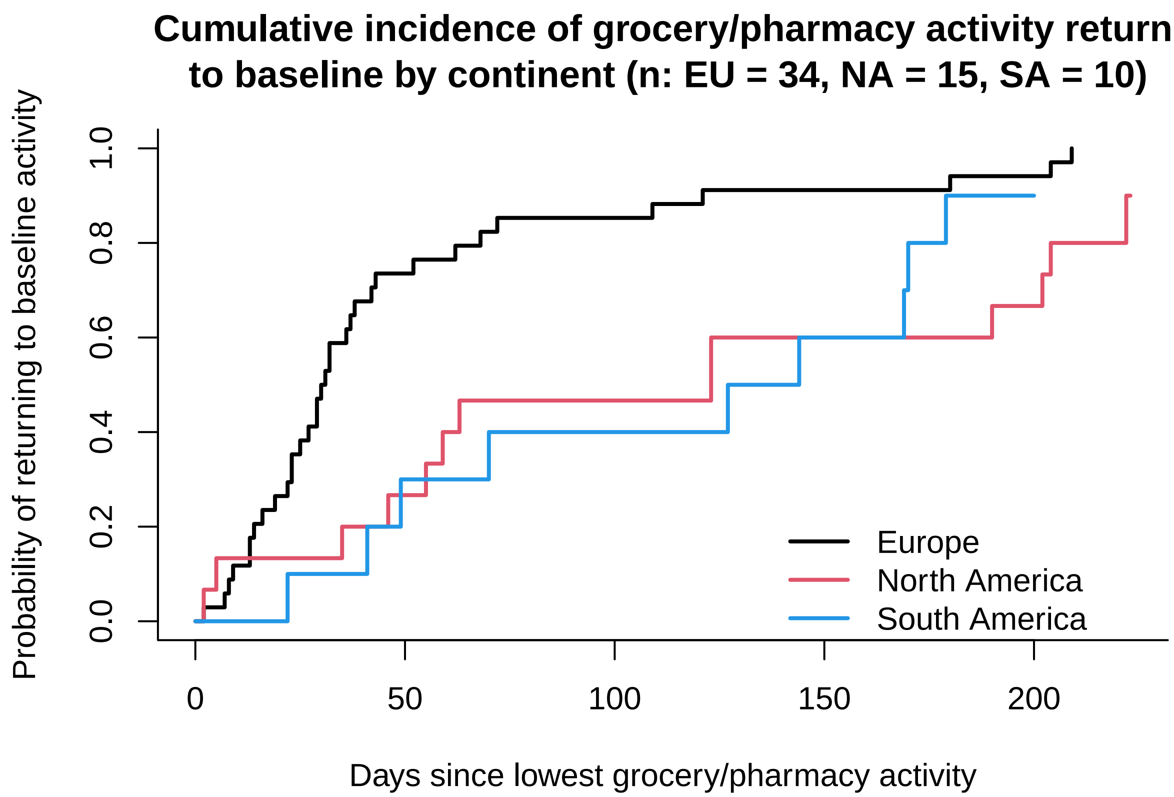


By inspection these fit well on the cumulative hazard plot, for the hazard of returning to baseline retail activity.

4.2 Grocery and pharmacy

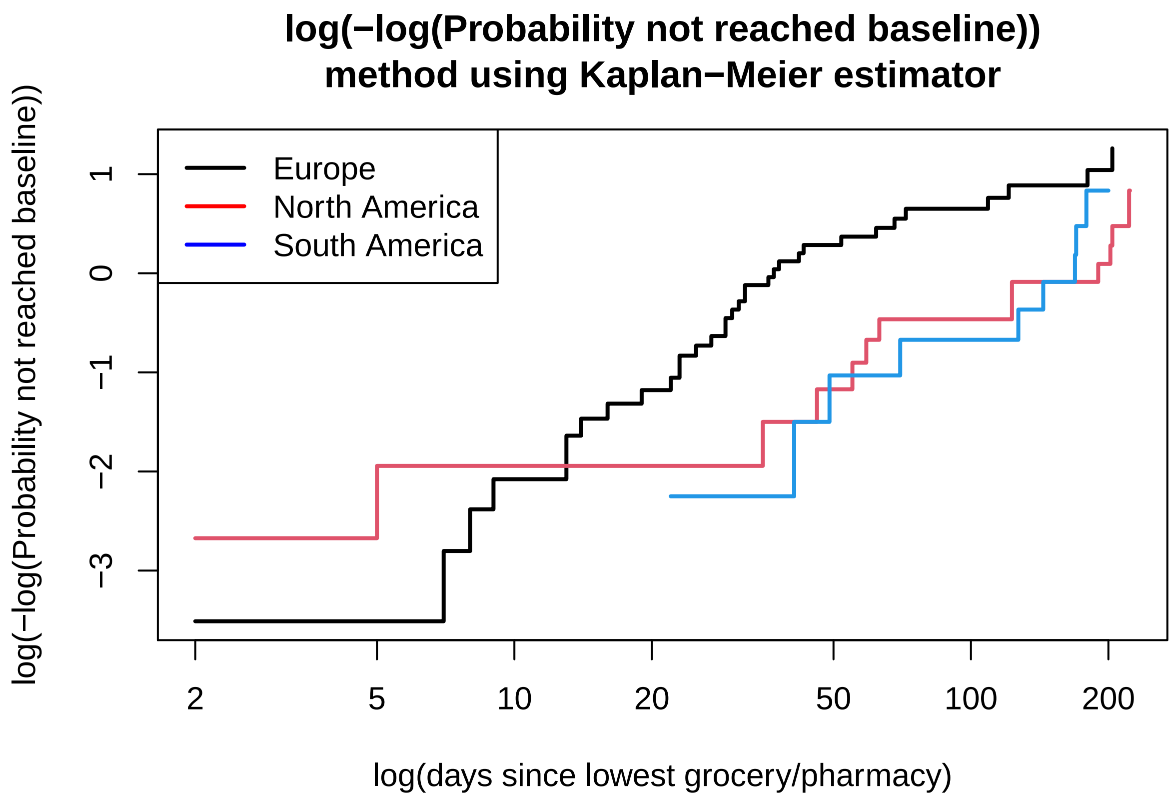
```
mingroc <- euna %>% group_by(country) %>% slice_min(gcmr_grocery_pharmacy) %>% distinct(date)
names(mingroc) <- c("country", "mindate")
euna_g <- left_join(euna, mingroc)

normalgroc <- euna_g %>%
  group_by(country) %>%
  filter(date > mindate, gcmr_grocery_pharmacy > -8) %>%
  slice_min(gcmr_grocery_pharmacy) %>%
  slice_min(date) %>%
  distinct(date)
names(normalgroc) <- c("country", "maxdate")

minmax_g<-left_join(mingroc,normalgroc)
minmax_g$mindate <- ymd(minmax_g$mindate)
minmax_g$maxdate <- ymd(minmax_g$maxdate)
minmax_g <- minmax_g %>% mutate(
  etime = ifelse(!is.na(maxdate),
                 as.duration(mindate %--% maxdate) / ddays(1),
                 as.duration(mindate %--% ymd("2020-11-19") / ddays(1))),
```

```r
  event = ifelse(!is.na(maxdate),
                 1,
                 0)
  )

minmax_g <- left_join(minmax_g, select(euna, country, continent), by = "country")
surv_g <- minmax_g %>% distinct(etime, event, continent) %>% slice_min(etime)

surv_g$continent <- case_when(
  surv_g$continent == "Europe" ~ "EU",
  surv_g$continent == "North America" ~ "NA",
  surv_g$continent == "South America" ~ "SA"
)

cuminc_gfit <- cuminc(
  ftime = surv_g$etime,
  fstatus = surv_g$event,
  group = surv_g$continent
)

plot(cuminc_gfit,
     col = c(1,2,4),
     lty = c(1,1,1),
     lwd = 2,
     wh = c(0,3),
     xlab = "Days since lowest grocery/pharmacy activity",
     ylab = "Probability of returning to baseline activity",
     main = "Cumulative incidence of grocery/pharmacy activity return\n to baseline by continent (n: EU
legend(135,0.25, c("Europe", "North America", "South America"),
       col = c(1,2,4),
       lwd = 2,
       box.lty = 0,
       bty = 'y')
```



4.2.1 Testing proportional hazards assumption

```
surv_gfit <- survfit(Surv(etime, event) ~ continent, surv_g)
plot(surv_gfit,
     col = c(1,2,4),
     lwd = 2,
     fun = "cloglog",
     xlab = "log(days since lowest grocery/pharmacy)",
     ylab = "log(-log(Probability not reached baseline))",
     main = "log(-log(Probability not reached baseline))\nmethod using Kaplan-Meier estimator")
legend("topleft", lty = c(1,1,1), lwd = c(2,2,2), col = c("black", "red","blue"), c("Europe", "North Ame
```



Can be assumed proportional hazards. If treating the baseline continent as South America, then the below Cox Proportional Hazards table is below:

```
surv_g$continent <- relevel(as.factor(surv_g$continent), "SA")
coxph(Surv(surv_g$etime, surv_g$event) ~ surv_g$continent)
```

```
## Call:
## coxph(formula = Surv(surv_g$etime, surv_g$event) ~ surv_g$continent)
##
##                         coef exp(coef) se(coef)      z      p
## surv_g$continentEU  0.8425    2.3222   0.3845  2.191 0.0284
## surv_g$continentNA -0.2569    0.7734   0.4499 -0.571 0.5679
##
## Likelihood ratio test=12.83  on 2 df, p=0.001635
## n= 59, number of events= 56
```

Yielding the hazard ratio of 2.32 ($p = 0.02$) for Europe attaining baseline grocery/pharmacy activity. In "businessy" terms, the probability of countries returning to baseline grocery/pharmacy activity at any time is 2.32 times higher in Europe than that of South America.