Ananlysis of COVID Data

Ryan Greenup

May 19, 2020

Contents

Preliminary

Load Packages and Data

```
 if (require("pacman")) {
    library(pacman)
  }else{
    install.packages("pacman")
    library(pacman)
  }
  pacman::p_load(xts, sp, gstat, ggplot2, rmarkdown, reshape2, ggmap,
                 parallel, dplyr, plotly, tidyverse, reticulate,
                  ↪  UsingR, Rmpfr,
                 swirl, corrplot, gridExtra, mise, latex2exp,
                  ↪  tidyverse, xts, maptools, plyr, ggplot2, maps,
                  ↪  viridis)

mise()
```

Load the Data

```r
covid <- read.csv("/home/ryan/Notes/DataSci/Visual_Analytics/Assessment 2/owid-covid-data.csv")
```

Set Working Directory

Introduction

- in December 2012 first cases of *COVID-19* were reported, the disease has since attributed to the *SARS-CoV2* virus.
- The disease became endemic throughout China before spreading throughout Europe in an epidemic fashion and finally reaching the rest of the globe as a pandemic outbreak.

Chloropleth Map

A Chloropleth map of the number of deaths can offer an insight into the impact that the disease has had with respect to individual countries.

The Total deaths should be scaled relative to the population of the country, that way countries with a smaller and sparser population will still be represented by the visualisation (this is quite important given that many countries such as Italy have a small population compared to the US and much of Asia [**2020n**]).

A worldwide Chloropleth map visualising the total number of deaths attributed to *COVID-19* is shown in figure 1 and a Europe-centric visualisation is shown in 2.

Discussion

Worldwide

The first plot appears to show a very limited amount of difference in deaths attributable to *COVID-19* across regions other than the North America and Europe. While first-world countries such as New Zealand and Australia are somewhat insulated from the disease by virtue of geography and population density, it's striking that much of Asia and Russia have such low levels of outbreak.

This could be attributed to the fact that a more power-cetric regime such as in China, Russia, North Korea, etc. may have more capacity to:

1. Diminish the spread of the diseasy by implementing policy decisions,
   (a) whereas countries such as the US and Europe have a much higher expectation of civil liberties.
2. Control the spread of information for want of international reputation.

A similar disease, *MERS*, emerged in 2012 in Middle-Eastern Regions [**woodley2020**] and a Korean outbreak of the *MERS* disease occured in 2015 [**serrano2015**], these outbreaks likely prepared Korea, the Middle East and other Asian Regions regions for an outbreak which helps explain the dichotomous nature of the deaths attributable to *COVID-19* for those Countries.

Europe

A closer look at Europe shows that Belgium and Italy have been the most affected by this disease, it isn't very clear why those regions have been impacted so significantly but this could be indicative of policy decisions and warrants further research.

Technique

Woldwide Map

First the data must be aggregated in order to retrieve the total number of deaths, this can be acheived by taking the maximum of the total deaths across countries (the total number of death rates will be a strictly positive and monotone trend, otherwise the outbreak would be an entirely different type of pandemic!), this can be performed by using the `aggregate` function as demonstrated in figure 1.

```r
fatalprop <- aggregate(total_deaths_per_million ~ location, covid, max)
## Order the Values in Descending Order
fatalprop <- fatalprop[order(-fatalprop$total_deaths_per_million),]
## Rename USA
covid$location[covid$location=="United States"] <- "USA"
```

Listing 1: Use Aggregate to aggregate total number of deaths

It is next necessary to rename `location` to `region` so map data will be consistent with the provided data set, this is shown in listing 2.

```r
## Rename to facilitate joining with map
names(fatalprop) <- c("region", "total_deaths_per_million")
```

Listing 2: Rename Features for consistency

For a broad overview of the data, small regions such as San Marino and Belgium will not be visible and will skew the colour pallete, so instead they should be removed and instead a seperate plot of Europe will be creted as shown in figure 2, this removal is performed in listing 3.

Next it is necessary to retrieve map data, this can be done using the `map_data` function, this data may then be combined by region with the provided data set using the `left_join` function, this is shown in listing 4.

Finally this data frame can be plotted by using `ggplot2` and the `geom_map` layer, modifying the `theme` layer will allow to provide a natural background, this is demonstrated in listing 5 and the output is provided in figure 1.

```r
## San Marino will be shown by italy and this skews the results
## Belgium and San Marino are very hard to visualise from above
## They skew the rsults and so will be removed.
fatalprops <- fatalprop %>% filter(region!="San Marino")
fatalprops <- fatalprop %>% filter(region!="Belgium")
```

Listing 3: Filter out small dense regions to prevent scale issues

```r
## Retrieve the map data
some.eu.maps <- map_data("world", region = fatalprops$location)

## Join the Data Frames Together
fatalmap <- left_join(fatalprops, some.eu.maps, by = "region")
```

Listing 4: Combine Map Data with Provided Data

```r
 ggplot(fatalmap, aes(map_id = region)) +
  geom_map(map = fatalmap,  color = "grey", aes(fill =
   ↪  total_deaths_per_million), lwd = 0.1, alpha = 0.6)+
  expand_limits(x = fatalmap$long, y = fatalmap$lat)+
  scale_fill_gradient(high = "darkred", low = "white") +
  guides(fill = guide_legend("Total Deats \n per Million")) +
   # Change the colors of background
   # and the color of grid lines to white
   theme(
     panel.background = element_rect(fill = "lightblue",
                                     colour = "lightblue",
                                     size = 0.5, linetype = "solid"),
     legend.position = c(0.6, 0.1),
     legend.direction = "horizontal",
     legend.background = element_rect(fill = "white", size = 0.1,
      ↪  colour = "darkblue", linetype = "solid")) +
   labs(x = "Longitude", y = "Latitude", title = TeX("Total Deaths
    ↪  Attributed to \\textit{COVID-19}"))
#   geom_text(data = region_lab_df, aes(y = lat, x = long, label =
 ↪  region), size = 1)
```

Listing 5: use ggplot2 to create a chloropleth map from data, output in figure 1

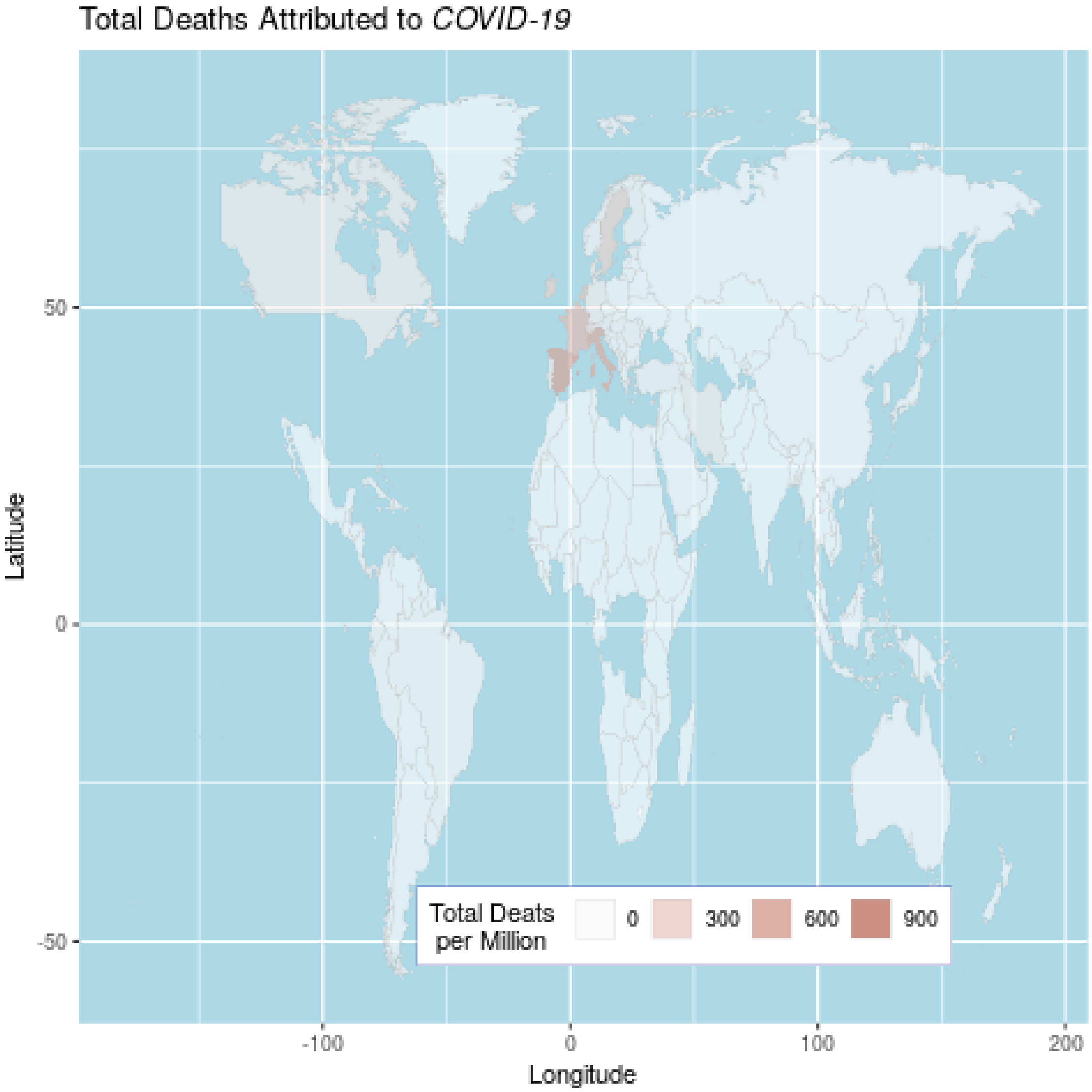


Figure 1: Chloropleth map of total deaths attributed to *COVID-19* (per Million people)

Europe Centric

The chloropleth map clearly shows that the disease has caused more fatalities per capita in Europe and so the plot will be adjusted central to Europe.

As before it is necessary to rename the features of the dataset, however in this instance small European countries such as Belgium should be retained (San marino is a very small italian provice that isn't detectable in the visualisation and skews the pallete, for this reason it will be removed), this is demonstrated in figure 6

```r
## Rename to facilitate joining with map
names(fatalprop) <- c("region", "total_deaths_per_million")

## San Marino will be shown by italy
 fatalprop <- fatalprop %>% filter(region!="San Marino")
```

Listing 6: Rename the features of the data and remove San Marino

In this map it will be desirable to have labels for the European countries (whereas this would have made the worldwide map too busy), so this will be implemented by using `dyplyr` to generate a second data set as shown in figure 7 which can then be used to generate a plot as shown in figure 8, this produces the output shown in figure 2.

```r
fatalmap <- left_join(fatalprop, some.eu.maps, by = "region")

## Filter out only Europe
fatalmap <-  fatalmap %>%
  filter(30 <  lat & lat < 65) %>%
  filter(-30 <  long & long < 35)

## Create Label Data Frame
region_lab_df <- fatalmap %>%
  dplyr::group_by(region) %>%
  dplyr::summarise(long = mean(long), lat = mean(lat))
```

Listing 7: use `dplyr` to reduce the plot size and create a data frame of country labels

Time Series

Bar Chart

Pie Chart

```r
library(ggrepel)
library(ggrepel)
ggplot(fatalmap, aes(map_id = region, label = region)) +
  geom_map(map = fatalmap,
           aes(fill = total_deaths_per_million),
           color = "white") +
  expand_limits(x = fatalmap$long, y = fatalmap$lat) +
  scale_fill_viridis_c(option = "C") +
  scale_fill_gradient(high = "darkred", low = "white") +
  guides(fill = guide_legend("Total Deats \n per Million")) +
  # Change the colors of plot panel background to lightblue
  # and the color of grid lines to white
  theme(
    panel.background = element_rect(
      fill = "lightblue",
      colour = "lightblue",
      size = 0.5,
      linetype = "solid"
    ),
    legend.position = c(0.1, 0.6),
    legend.direction = "vertical",
    legend.background = element_rect(
      fill = "white",
      size =
        1.1,
      colour = "darkblue",
      linetype = "solid"
    )
  ) +
  labs(
    x = "Longitude",
    y = "Latitude",
    title = TeX("Total Deaths Attributed to \\textit{COVID-19}")
  ) +
  geom_text_repel(
    data = region_lab_df,
    aes(y = lat, x = long, label = region),
    size = 2,
    col = "black",
    nudge_y = 0.7,
    nudge_x = -0.5,
    min.segment.length = 0.3,
    force = 2
  )
```

Listing 8: Generate a Chloropleth map centred on Europe using ggplot2

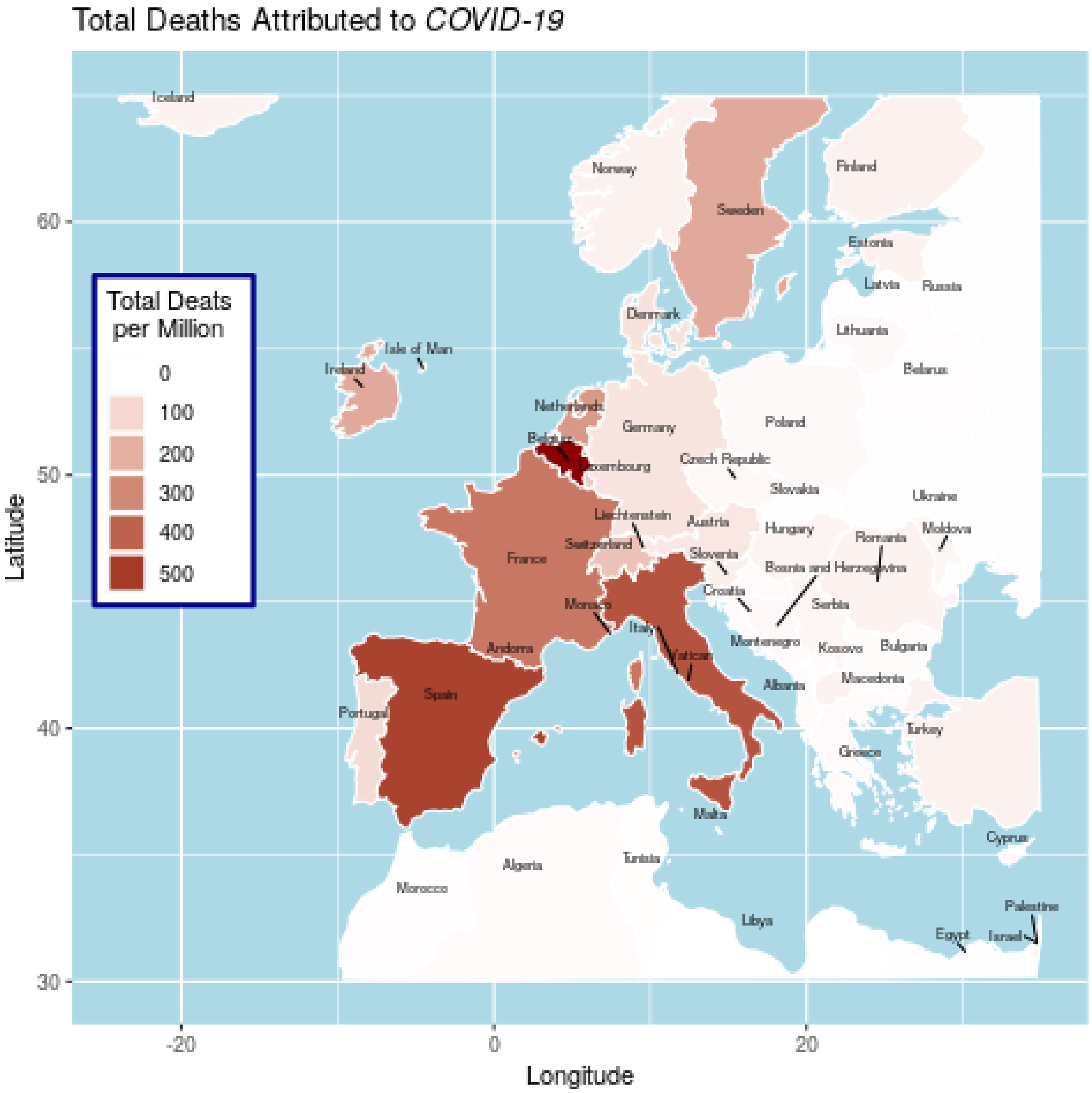


Figure 2: Europe Centred Chloropleth of Deaths Attributed to *COVID-19*

Spider Chart / Star Plot

Multiple Line Charts

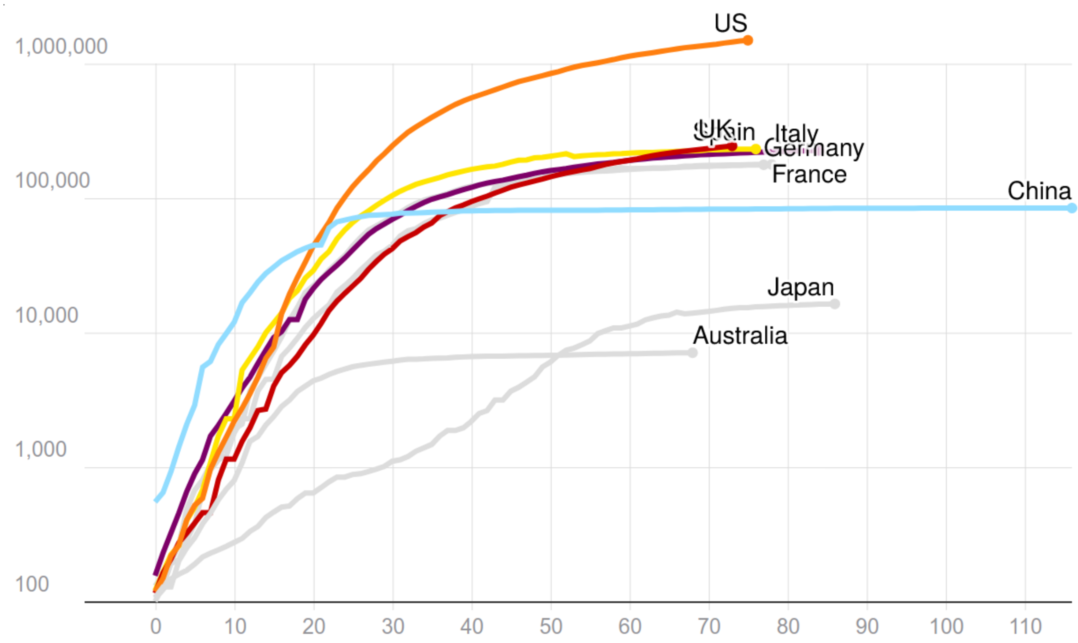
Parallell Co-ordinates

each line is a country each column is a feature like testing, death and cases.

This Stack Post shows how to make them curvy

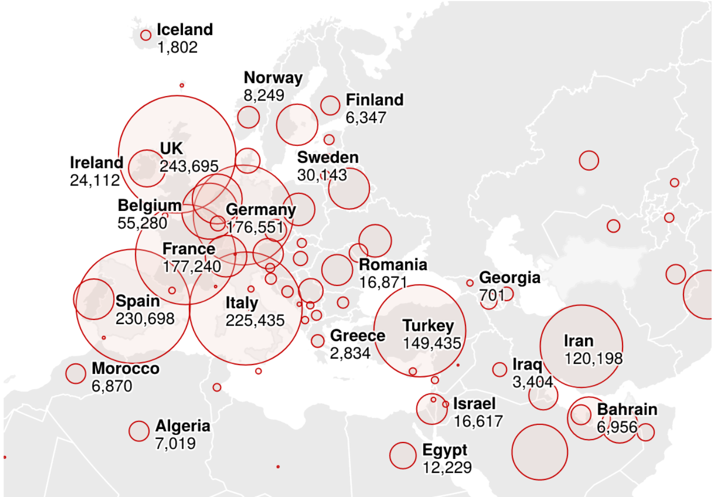
3D Scatter Plot

Log Scaled from 100th case ATTACH



Bubble Plot ATTACH

Guardian



Animation of 3d Chloropleth heatmap

visualisation

The total number of deaths per country can be analysed using

Technical Details

Advantages compared to other methods

Disasadvantages

Discussion on analysis results

Discussion on other Aspects

Literature review of related work

For Each Visualisation

Technical Details

Advantages compared to other methods

Disasadvantages

Discussion on analysis results

Discussion on other Aspects

Literature review of related work

References

../../../../Studies/Papers/references