World Covid Analysis

STAT614-002
Chaytanya Kumar, Dennis Poludnev, Youjia (Yuka) Chen

Introduction

- We decided to do a regression analysis on COVID related data.
- Initially we were looking in to the local Montgomery COVID data on the government website, which was in the large part already analysed through tableau.
- Then we decided to find more data related to the bigger scope.
- We really wanted to focus on how multiple countries are fairing with COVID and what factors might contribute to better fight this and possible future pandemics.

Data

- We found a good dataset which is constantly updated through github.

  Link: https://github.com/owid/covid-19-data/tree/master/public/data

  - The main variables such as total death and vaccination are updated weekly from a credible sources: John Hopkins and CSSC, ECDC.
- The data is very vast and has missing values in form of NA.
- The NA values are due to the nature of using multiple global health care organizations to compile one data frame.
- Therefore we decided to clean the data first then take a random sample for data analysis and hypothesis testing.

Metrics to measure

Confirmed cases

Confirmed deaths

Tests and Positivity

Vaccination

Hospital and ICU

Policy response

Other variables of Interest

Tidying the Data.

- We used R studio to clear our as much NA values as possible.
- Then we decided to randomly select 30 samples.
- The samples were selected based on the country column.
- We did that in order to search for causation variables and make analysis based on total death as the response variable and to find out why some countries had harder time with COVID pandemic than the others.
- The R studio was an excellent tool to use for a quick data cleaning, multiple linear model analysis and visualization.

Hypothesis

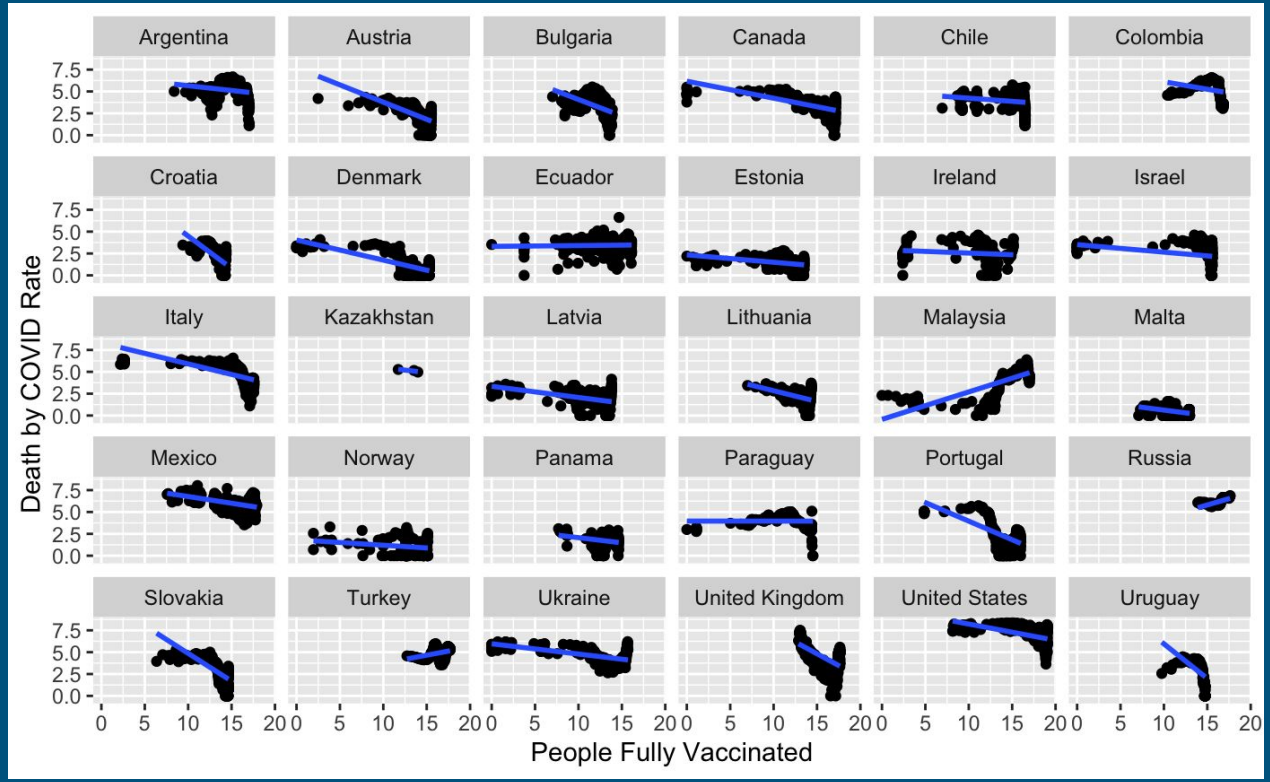
- Developing countries had harder time controlling the COVID pandemic than the developed countries.

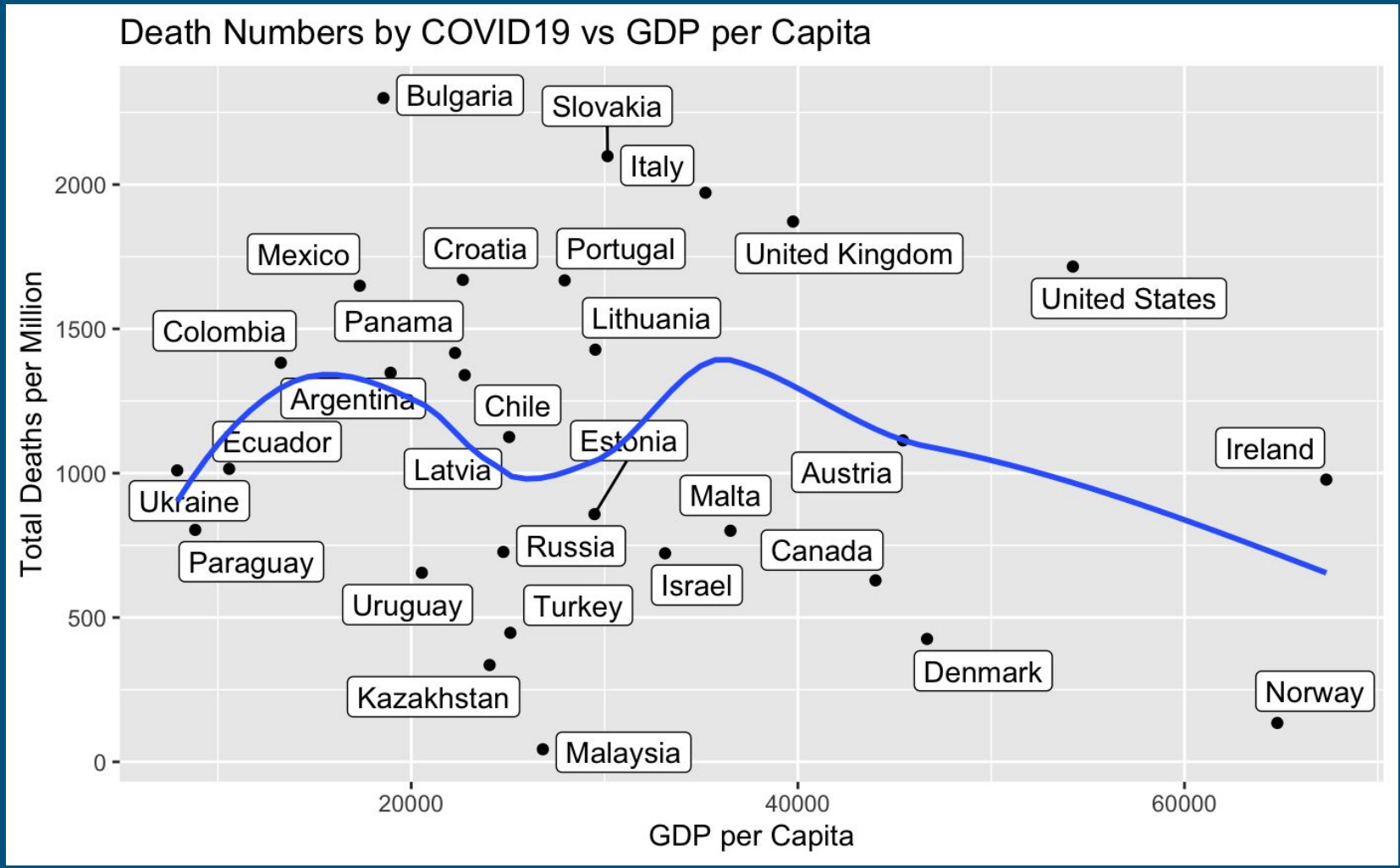
- This hypothesis focused on GDP as the main independent variable and total deaths due to covid as the response variable.

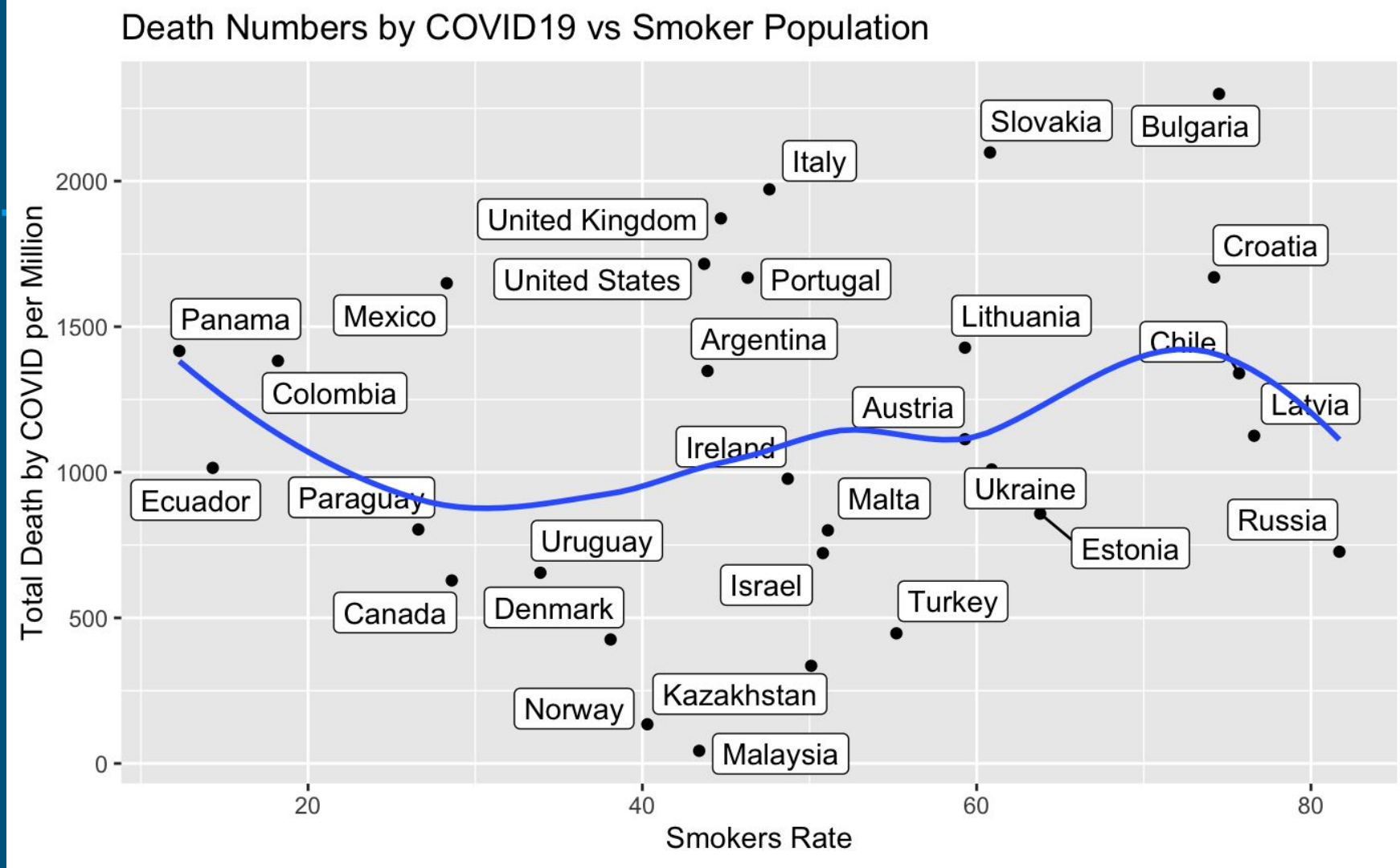
Analysis Results.

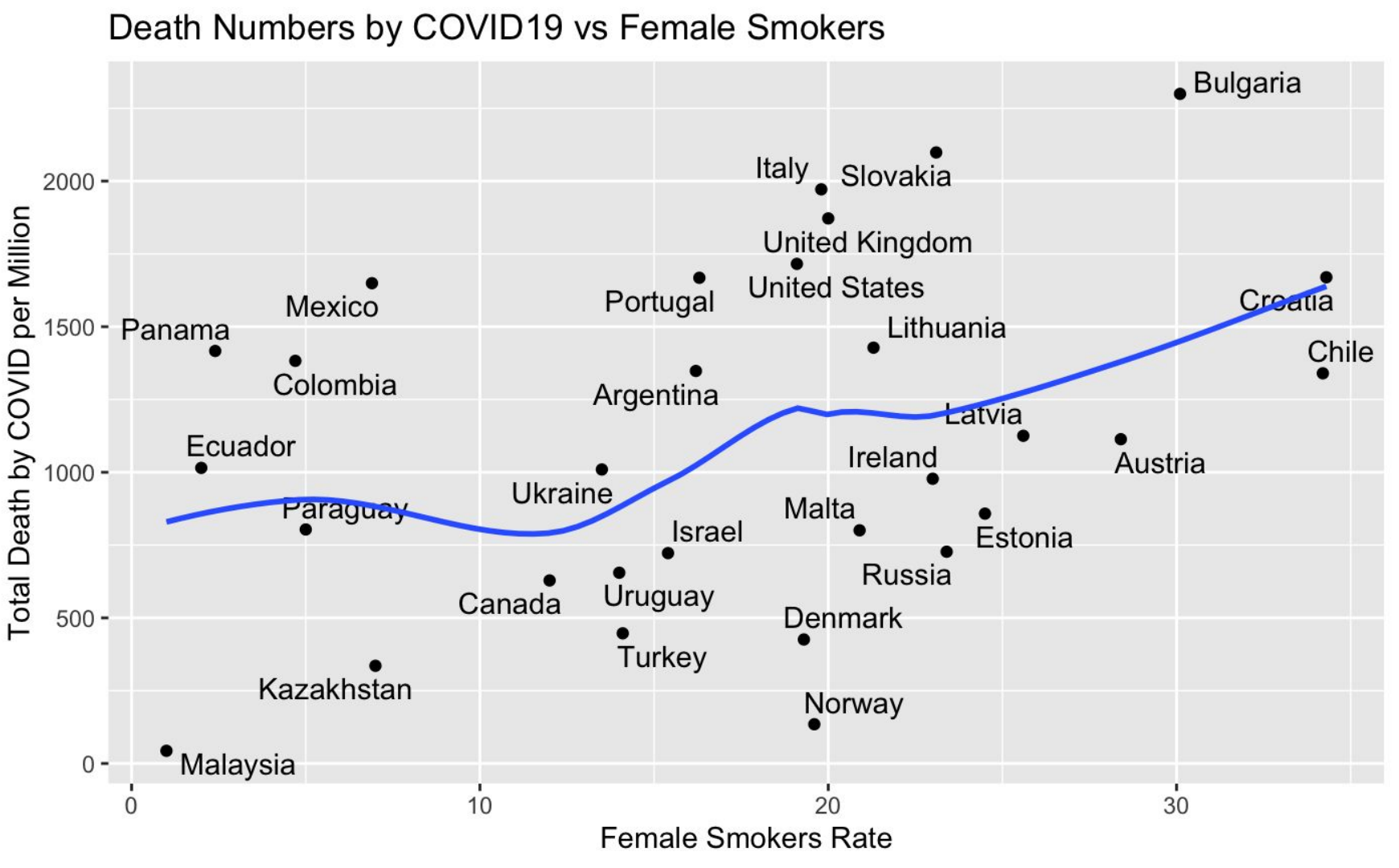
```
Residuals:
    Min      1Q  Median      3Q     Max
-641.38 -342.33  -89.45  284.22  951.69

Coefficients:
                                Estimate Std. Error
(Intercept)                    6.357e+02  6.828e+02
X30samples$total_vaccinations  4.336e-06  2.418e-06
X30samples$gdp_per_capita     -4.809e-03  8.232e-03
X30samples$extreme_poverty     3.018e+02  1.197e+02
X30samples$smoker_total        1.285e+01  6.382e+00
X30samples$population_density  1.811e-01  3.607e-01
X30samples$diabetes_prevalence -4.986e+01  3.663e+01
                               t value Pr(>|t|)
(Intercept)                      0.931   0.3615
X30samples$total_vaccinations    1.793   0.0861 .
X30samples$gdp_per_capita       -0.584   0.5647
X30samples$extreme_poverty       2.520   0.0191 *
X30samples$smoker_total          2.014   0.0559 .
X30samples$population_density    0.502   0.6204
X30samples$diabetes_prevalence  -1.361   0.1867
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 500.3 on 23 degrees of freedom
Multiple R-squared:  0.4239,    Adjusted R-squared:  0.2736
F-statistic:  2.82 on 6 and 23 DF,  p-value: 0.03311
```

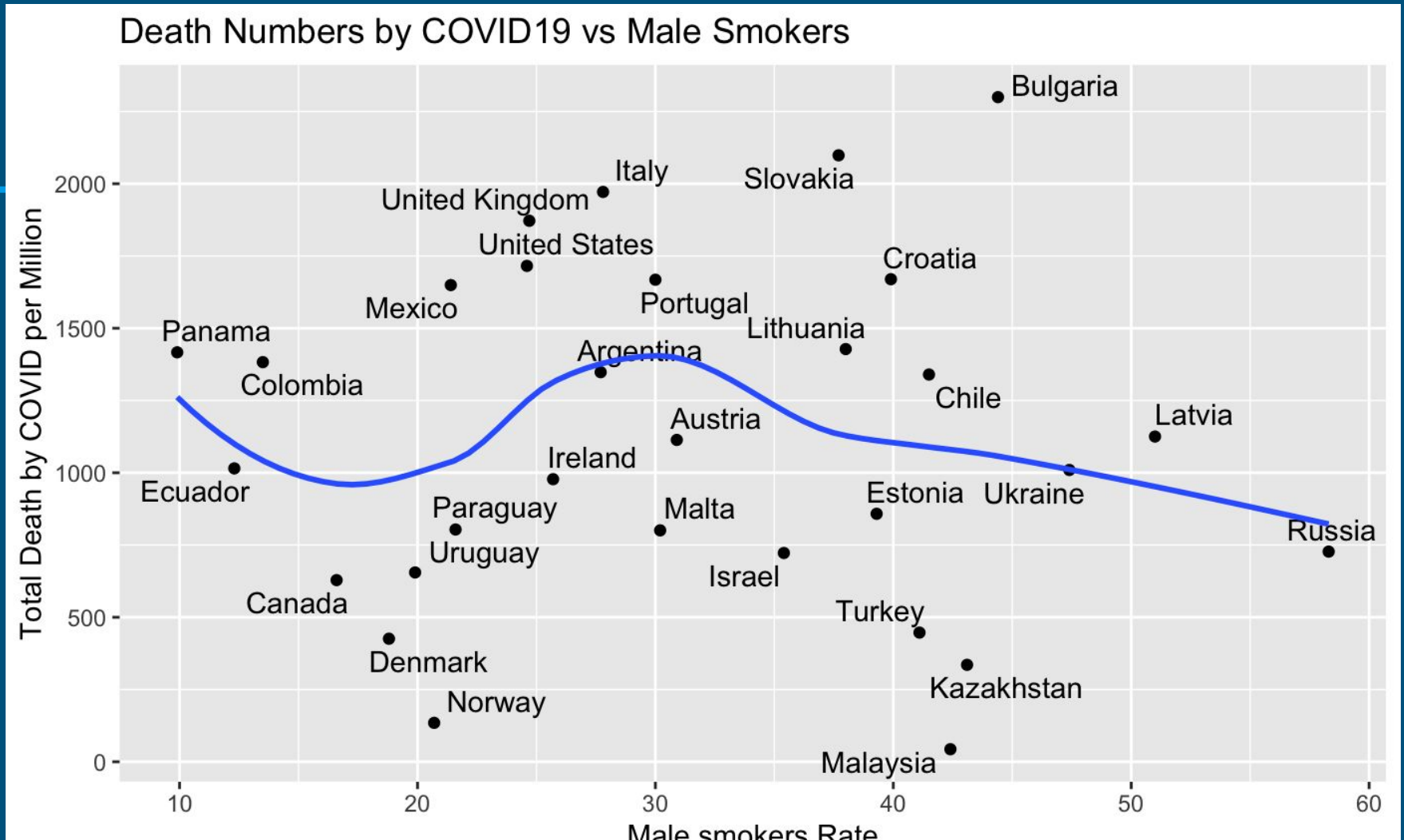
Data Visualization.

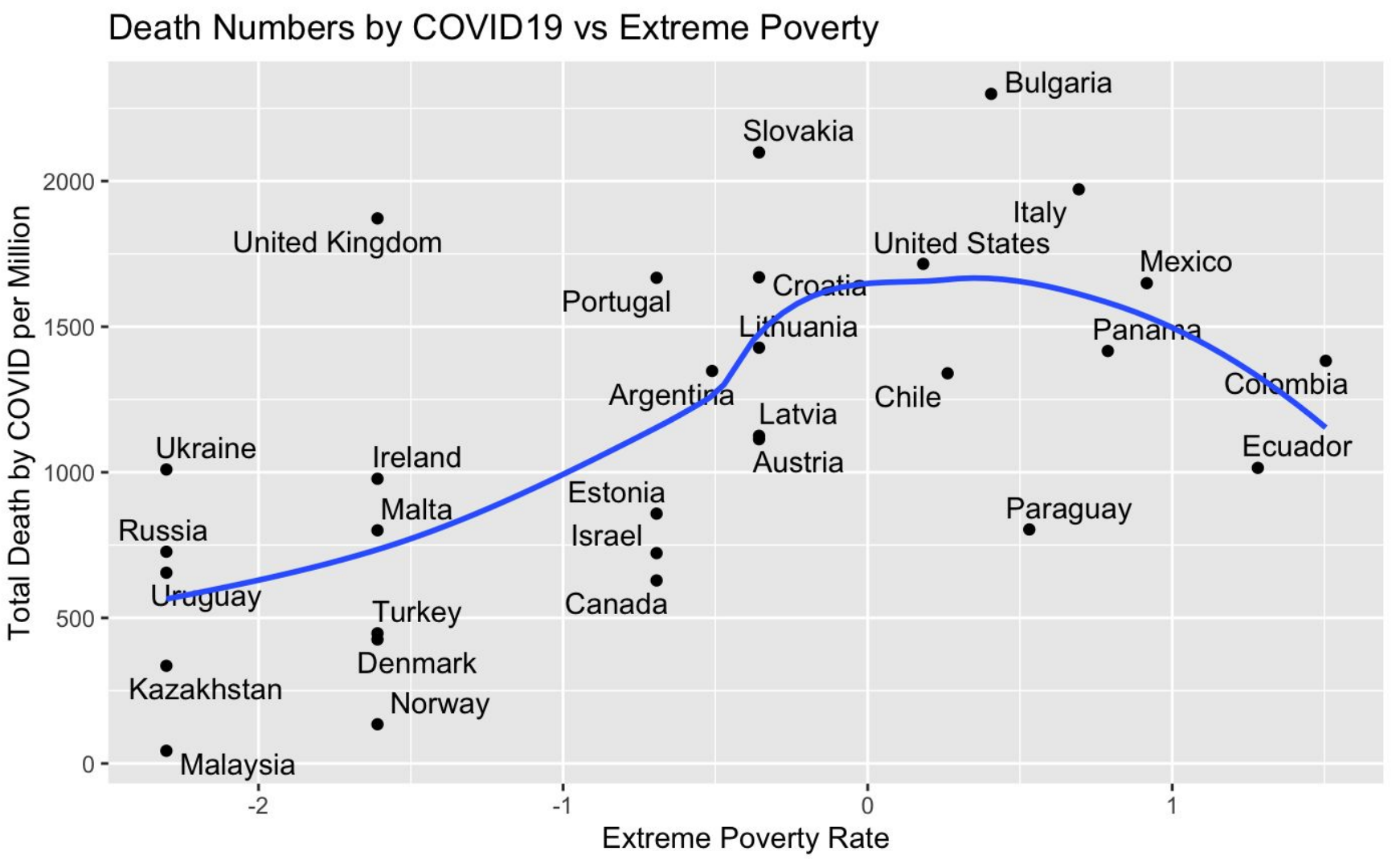


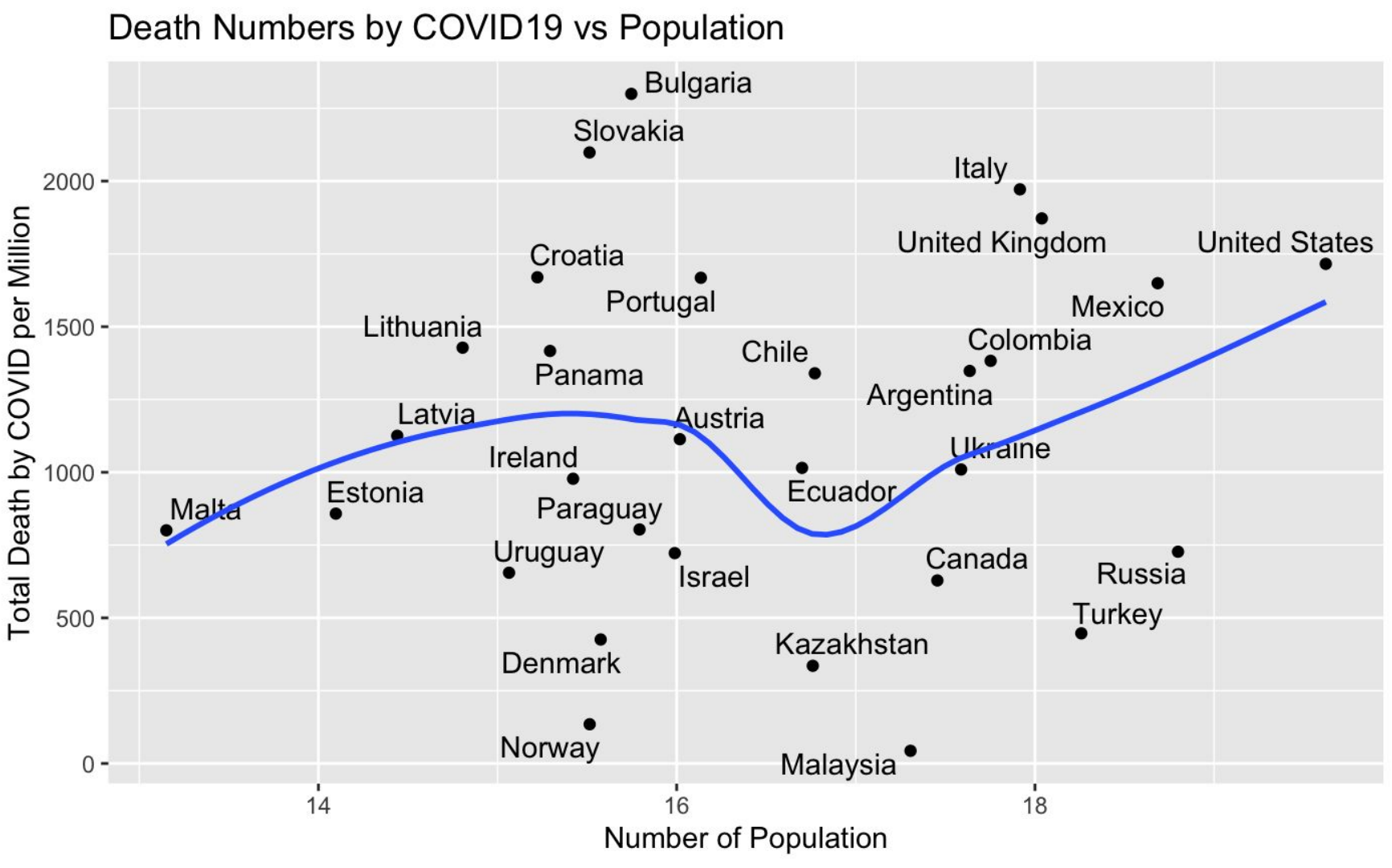

Death Numbers by COVID19 vs GDP per Capita
Total Deaths per Million
GDP per Capita
0
500
1000
1500
2000
20000
40000
60000
Bulgaria
Slovakia
Italy
Mexico
Croatia
Portugal
United Kingdom
United States
Colombia
Panama
Lithuania
Argentina
Chile
Ecuador
Estonia
Ireland
Latvia
Austria
Ukraine
Malta
Paraguay
Russia
Canada
Israel
Uruguay
Turkey
Denmark
Kazakhstan
Norway
Malaysia


Death Numbers by COVID19 vs Smoker Population
Total Death by COVID per Million
Smokers Rate
0
500
1000
1500
2000
20
40
60
80
Slovakia
Bulgaria
Italy
United Kingdom
Croatia
United States
Portugal
Panama
Mexico
Lithuania
Argentina
Chile
Colombia
Austria
Latvia
Ireland
Ecuador
Paraguay
Malta
Ukraine
Russia
Uruguay
Estonia
Israel
Turkey
Canada
Denmark
Kazakhstan
Norway
Malaysia


Death Numbers by COVID19 vs Female Smokers
Total Death by COVID per Million
Female Smokers Rate
0
500
1000
1500
2000
0
10
20
30
Bulgaria
Italy
Slovakia
United Kingdom
United States
Portugal
Mexico
Panama
Colombia
Lithuania
Croatia
Chile
Argentina
Latvia
Ecuador
Ireland
Austria
Ukraine
Paraguay
Malta
Estonia
Israel
Russia
Canada
Uruguay
Denmark
Turkey
Kazakhstan
Norway
Malaysia


Death Numbers by COVID19 vs Male Smokers
Total Death by COVID per Million
Male smokers Rate
0
500
1000
1500
2000
10
20
30
40
50
60
Bulgaria
Slovakia
Italy
United Kingdom
United States
Croatia
Mexico
Portugal
Lithuania
Panama
Argentina
Colombia
Chile
Austria
Latvia
Ireland
Ecuador
Estonia
Ukraine
Paraguay
Malta
Russia
Uruguay
Israel
Canada
Turkey
Denmark
Kazakhstan
Norway
Malaysia


Death Numbers by COVID19 vs Extreme Poverty
Total Death by COVID per Million
2000
1500
1000
500
0
-2
-1
0
1
Extreme Poverty Rate
Bulgaria
Slovakia
Italy
United Kingdom
United States
Mexico
Portugal
Croatia
Lithuania
Panama
Argentina
Chile
Colombia
Latvia
Ukraine
Ireland
Austria
Ecuador
Estonia
Malta
Paraguay
Russia
Israel
Uruguay
Canada
Turkey
Denmark
Kazakhstan
Norway
Malaysia


Death Numbers by COVID19 vs Population
Total Death by COVID per Million
Number of Population
0
500
1000
1500
2000
14
16
18
Bulgaria
Slovakia
Italy
United Kingdom
United States
Croatia
Portugal
Mexico
Lithuania
Colombia
Chile
Panama
Argentina
Latvia
Austria
Ukraine
Ireland
Estonia
Ecuador
Malta
Paraguay
Uruguay
Canada
Israel
Russia
Turkey
Kazakhstan
Denmark
Norway
Malaysia

Conclusion