IBM Data Science – Capstone Project

COVID-19 / Medical Facility Analysis

Alex Newman May 2020

# Executive Summary

There is no correlation (0.14) between death rates and the number of medical research labs within a 4km radius of the associated capital city center.

The categorisation of European countries by death rates disproves the perception that countries such as Italy (medium), Spain (medium) & the UK (medium) are the worst affected in Europe is not true. In comparison Belgium (high) and San Marino (very high) have a far higher death rate.

The quality of the data used in this investigation is questionable due to the lack of documentation regarding the categorisation employed by FourSquare (or most probably a lack of due diligence by the wannabe data scientist) and restrictions experienced in obtaining data regarding medical research facilities (licensing limitations). An example of unusual data is the lack of medical research labs in Moscow. Based on similar cities in population in Europe you would expect many more facilities.

Table of Contents

[Executive Summary 2](#_Toc40607631)

[Introduction / Business Problem 4](#_Toc40607632)

[Data 5](#_Toc40607633)

[FourSquare Medical Facility information 5](#_Toc40607634)

[Wikipedia World Population Data 5](#_Toc40607635)

[Wikipedia COVID-19 Statistics 5](#_Toc40607636)

[European Capital Cities 5](#_Toc40607637)

[Methodology 6](#_Toc40607638)

[Results 7](#_Toc40607639)

[Correlation 7](#_Toc40607640)

[Medical Research Facility Results 8](#_Toc40607641)

[Medical Research Facility in Brussels 9](#_Toc40607642)

[Categorisation 10](#_Toc40607643)

[Category Results 11](#_Toc40607644)

[Map of European Countries by Death Rate Percentage 12](#_Toc40607645)

[Discussion 13](#_Toc40607646)

[Recommendations 13](#_Toc40607647)

[References 14](#_Toc40607648)

# Introduction / Business Problem

After a recent surge in public attacks against medical laboratories due COVID-19, the World Health Organisation (WHO) have engaged the services of a wannabe data scientist.

The WHO require an investigation into the conspiracy theory involving the prevalence of medical laboratories and the death rate of the associated country.

Specifically, they require:

1. European countries to be assigned to a category based on COVID-19 death rate
2. Establish if there is a correlation between medical laboratories and death rate

# Data

## FourSquare Medical Facility information

A search of the Medical Research Facilities was conducted for each of the 41 European Country Capitals. Due to licensing restrictions we were only able to pull down a maximum of 50 venues per city. The radius of our search was reduced to 4km to prevent that threshold being reached in a majority of cases.

No documentation could be found regarding the criteria for classifying a venue as a medical research facility. Interestingly a search of Moscow returned zero results.

URL: <https://foursquare.com>

Example provided in the Appendices.

## Wikipedia World Population Data

World Population Data was scraped from the following Wikipedia page. The data was updated regularly from a variety of sources.

URL: <https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population>

Example Provided in the Appendices.

## Wikipedia COVID-19 Statistics

Covid-19 stats are published on Wikipedia accessed from the URL below. The information contains country reference along with total number of cases, deaths and recovered patients (where available). The information is published data from the relevant source associated with the country.

URL: <https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data#covid19-container>

Example Provided in the Appendices.

## European Capital Cities

A list of 44 European Capital Cities along with the associated country. Only 41 of these cities were identifiable against world population and COVID-19 stats – naming conventions were the primary reason. This information could be cleaned up in future iterations – it was decided for simplicity that they countries would be excluded from the first iteration of analysis

URL: <https://www.nationsonline.org/oneworld/capitals_europe.htm>

# Methodology

1. Pull and Clean datasets
2. Create COVID-19 Death Category

* join covid-19 data with country population data and calculate death rate (deaths / population)
* create a three tiered category based on death rate (high medium low)
* produce European map of death rate categories

1. 4square 'medical' data

* geocode all European capital cities
* pull all medical labs for each ~44 European capital cities within a 2km radius
* count all medical labs for each ~44 European capital cities within a 2km radius
* produce a European map medical lab counts

1. Investigate a correlation between Death rate of country and the number of medical facilities

* plot death rate against number of facilities
* investigate results

# Results

## Correlation

There was no correlation (0.14) between death rate of a country and the number of medical research facilities within a 4km radius of the country’s capital city.

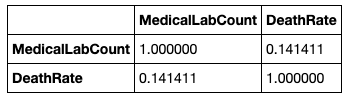


Figure - Death Rate and Medical Research Facility Correlation table

The Histogram clearly demonstrates the scattering of points.

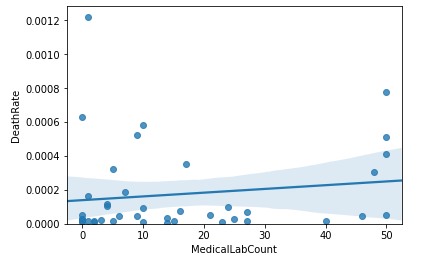


Figure - Histogram European Countries Medical Research Facility and Death Rate

## Medical Research Facility Results

No Medical Research Facilities were return through a 4km radius search of FourSquare location data of Moscow. As we can see from the table below this is very unusual for a European city of this size and raises serious question of the accuracy of this information for this particular European City.

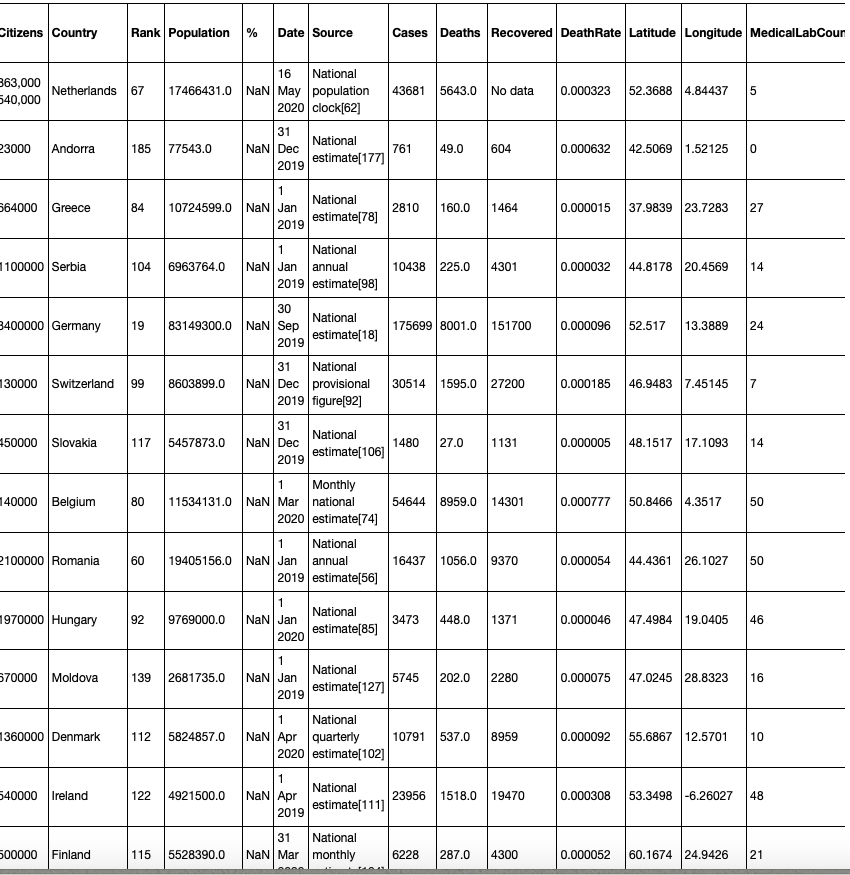


Figure Medical Research Facility Data from FourSquare

## Medical Research Facility in Brussels

50 medical research facilities were returned from a FourSquare search of Brussels (Belgium).

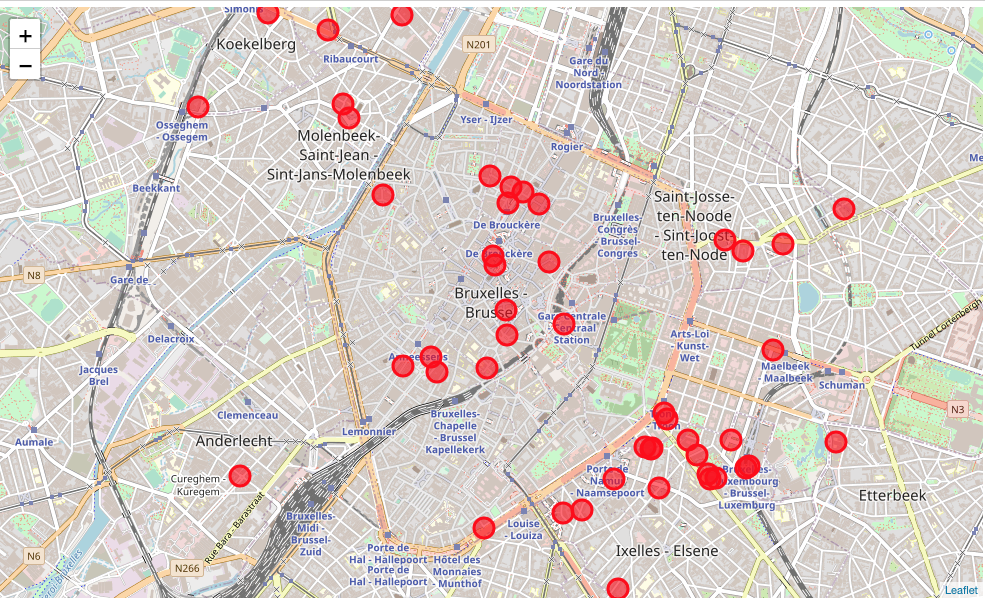


Figure Medical Research Facilities around Brussels

Medical Research Facilities in London

50 medical research facilities were returned from a FourSquare search of London (UK).

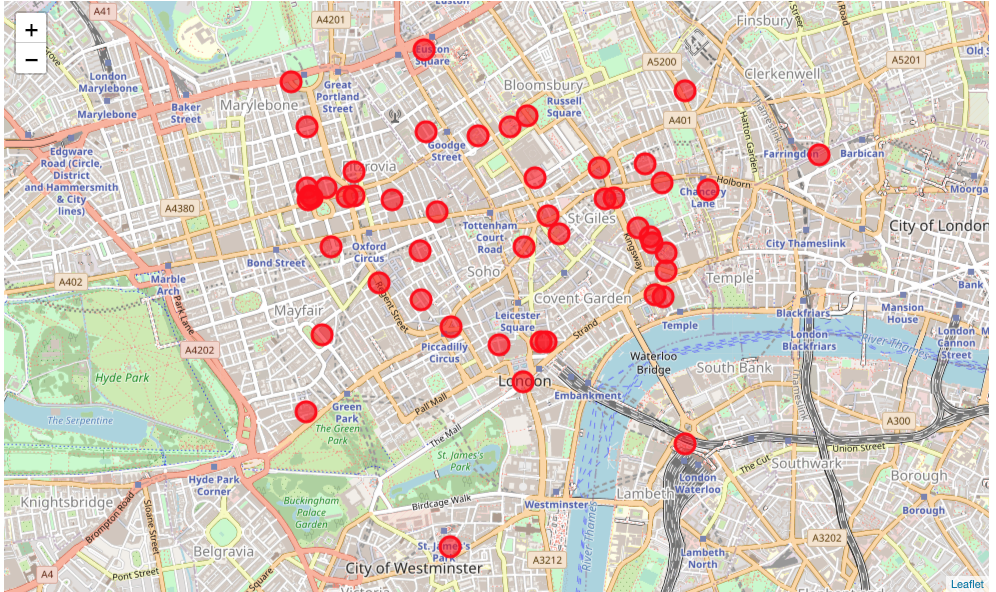


Figure Medical Research Facilities around London

## Categorisation

Death rates across European Countries vary widely. A box plot was produced to examine the range of rates so that an appropriate categorisation could be applied.

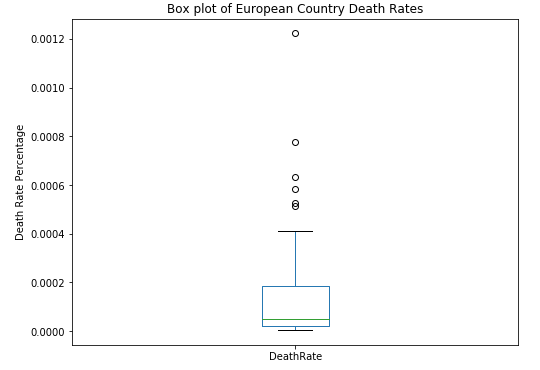


Figure - Box Plot of the Death Rate of European Countries

As can be seen from the Box Plot above there are a number of outliers. Interestingly Italy, Spain and the UK were did not have the highest death rates in Europe despite the media attention that was focused on these three countries. San Marino and Belgium have recorded much higher death rates at the time of producing this report.

A total of 41 counties were measured – resulting in the following table describing the Death Rate Category:

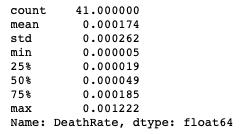


Figure - Table describing the Death Rate Attribute for European Countries

Using the figures from the above table a 5 categories were applied to the Death Rate figure classifying countries into Very high, high, medium, low and very low.

## Category Results

The following table was produce appending the 5 categories to the Death Rate and Medical Research facility Count data:

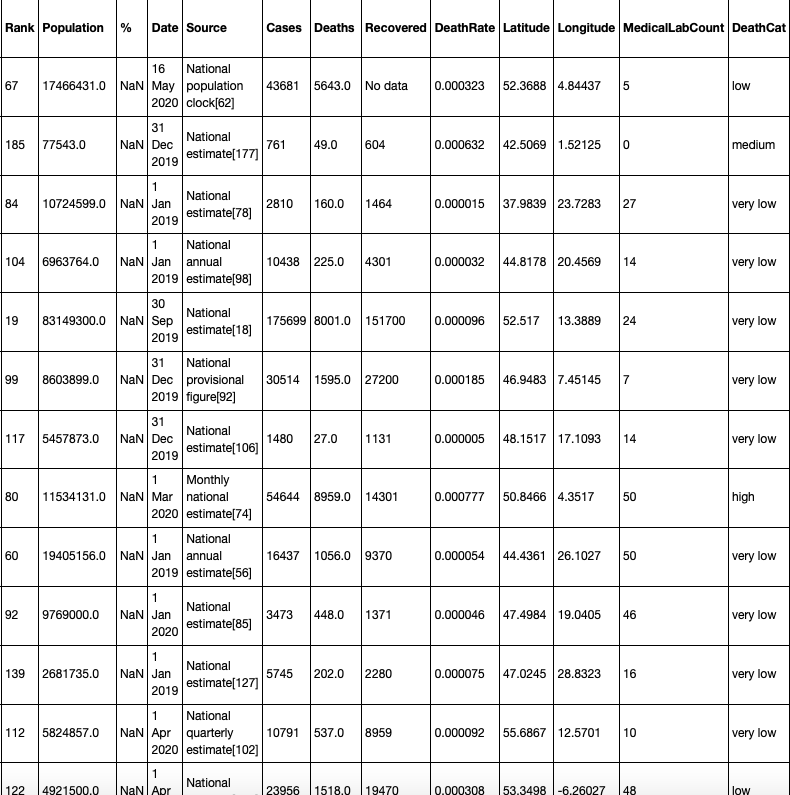


Figure Death Risk Categories for European Countries

## Map of European Countries by Death Rate Percentage

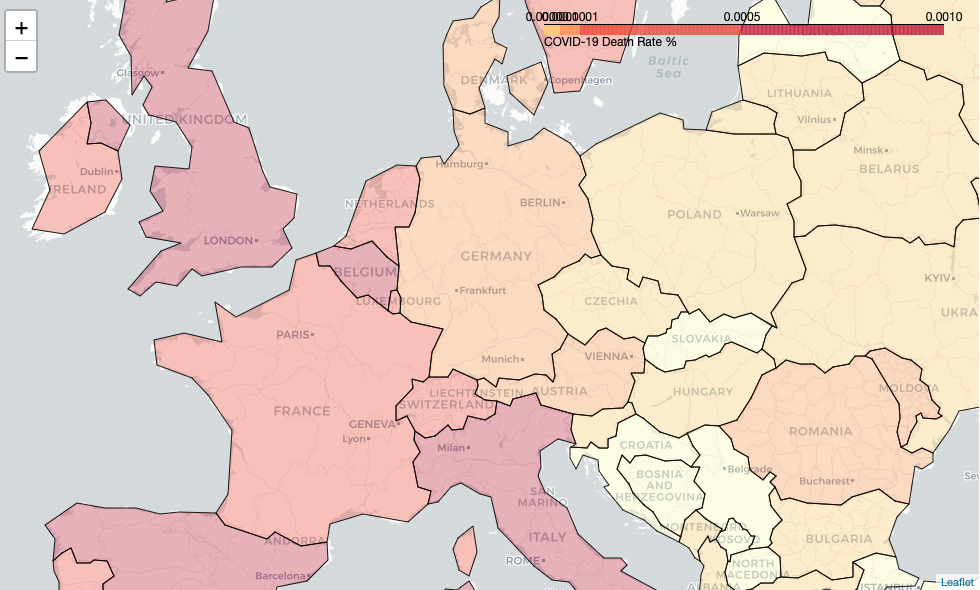


Figure European Death Rate Choropleth Map

## Discussion

Interestingly Italy, Spain and the UK were rated as having medium death rates despite the media attention of these countries due to their perceived high death rates.

The quality of FourSquare data is very variable depending on the country concerned. The query on Moscow medical reasearch facilities netted zero returns. The lack of data out of Russia could have seriously affected the correlation result.

Recommendations

Further research is required to determine the accuracy of the FourSquare data and to investigate if alternative data sources would yield different results.

Increasing the Foursquare search criteria to 100km potentially would yield significantly different results. This would require a modification to the current license employed by our wannabe data scientist.

As the COVID-19 crisis is still happening it is a further recommendation that the analysis be continued to see if trends start materializing over time.

# References

<https://labs.cognitiveclass.ai/tools/jupyterlab/>

<https://dataplatform.cloud.ibm.com/projects?context=wdp>

<https://www.nationsonline.org/oneworld/capitals_europe.htm>

<https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data#covid19-container>

<https://foursquare.com>

<https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population>

## Appendices

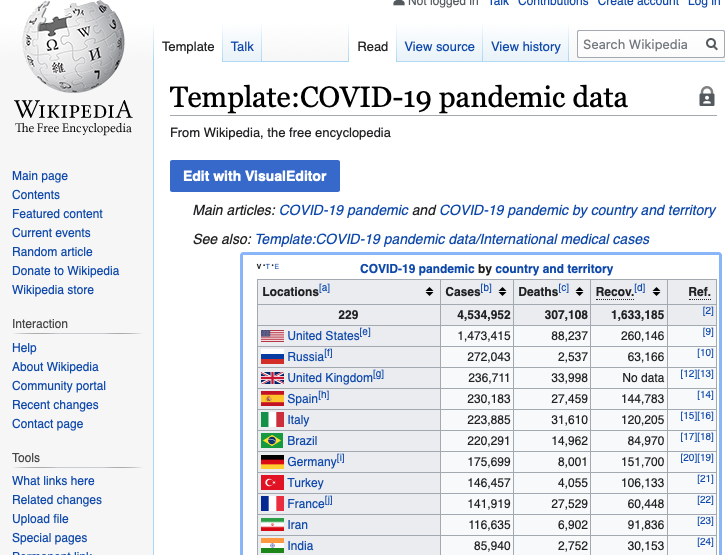


Figure Wikipedia COVID-19 stats



Figure Wikipedia Country Population stats



Figure Countries online European Capital City information