**Introduction to Data Scientist 11372 (UG)**

**Final Assessment Part D – Documentation and Reporting**

Tuan Anh (Vincent) Nguyen – u3196825

Table of Contents

[Task 2 Exploratory Data Analysis 2](#_Toc71819702)

[Q1: 2](#_Toc71819703)

[Q2: 3](#_Toc71819704)

[Q3: 4](#_Toc71819705)

[Q4: 5](#_Toc71819706)

[Q5: 5](#_Toc71819707)

[Q6: 6](#_Toc71819708)

[Q7: 7](#_Toc71819709)

[Q8: 9](#_Toc71819710)

[Q9: 10](#_Toc71819711)

[Q10: 11](#_Toc71819712)

[Task 3: Data-Driven Modelling: 12](#_Toc71819713)

[Q1: 12](#_Toc71819714)

[Q2: 12](#_Toc71819715)

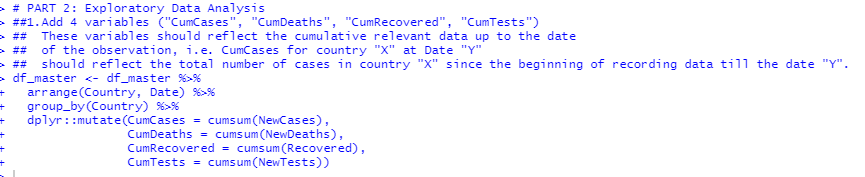
[Q3: 13](#_Toc71819716)

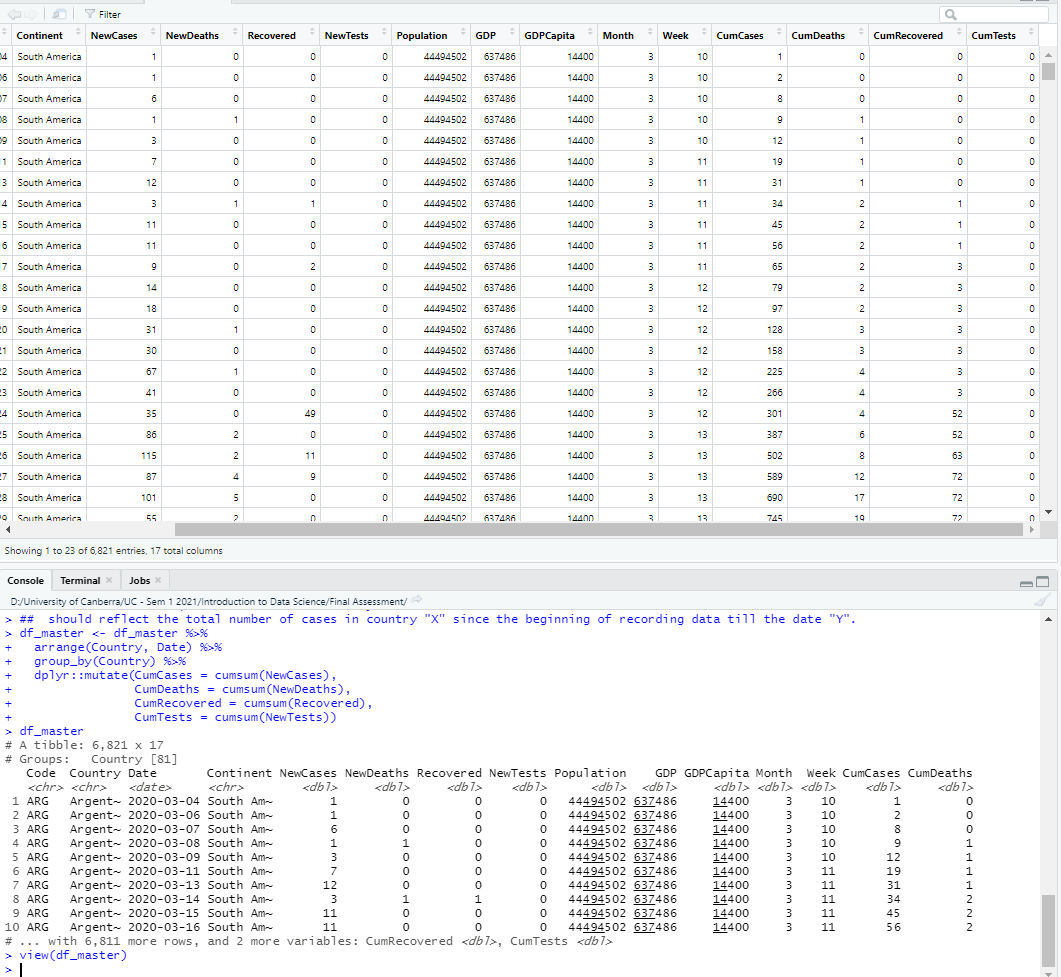
[Q4: 14](#_Toc71819717)

[Q5: 16](#_Toc71819718)

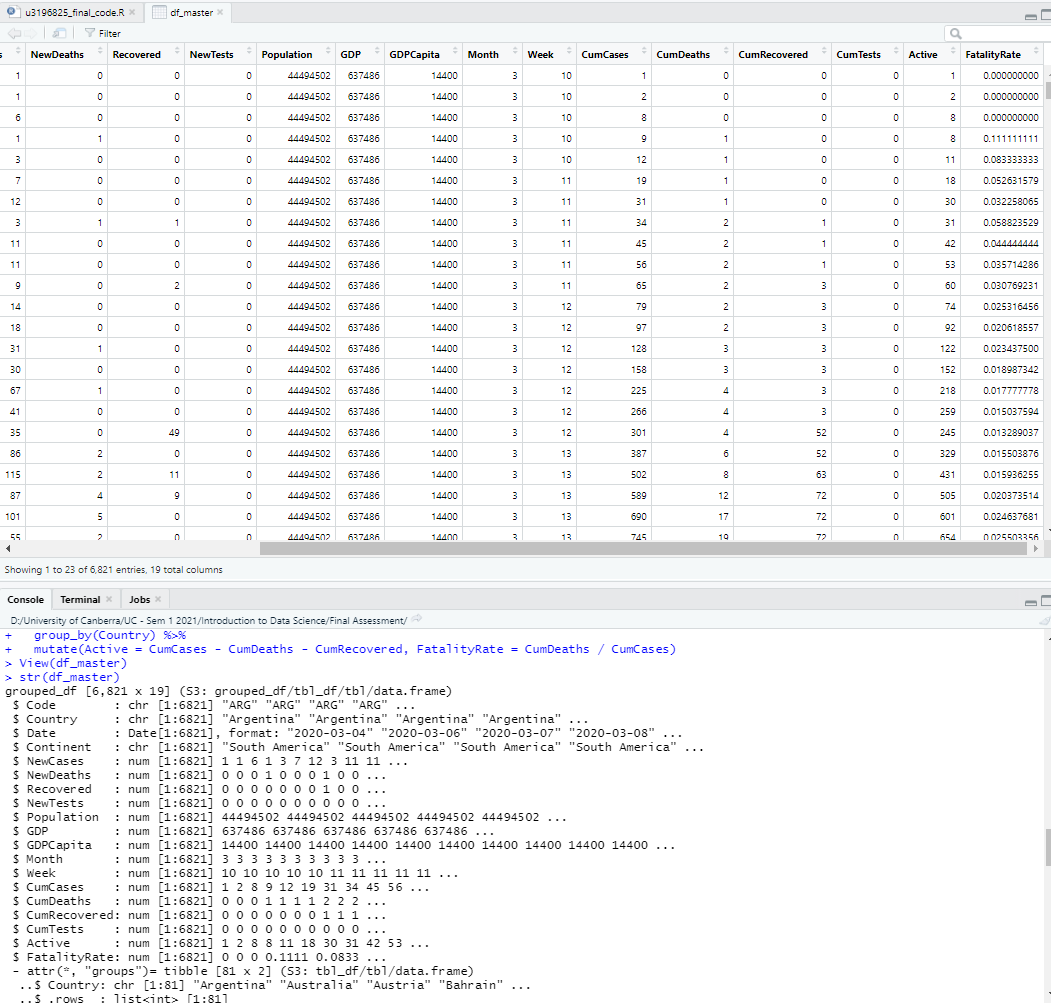
# Task 2 Exploratory Data Analysis

## Q1:

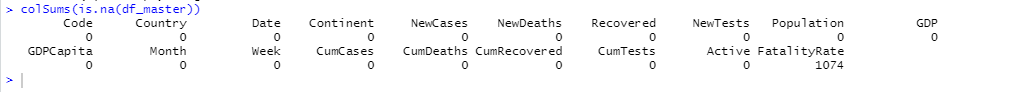


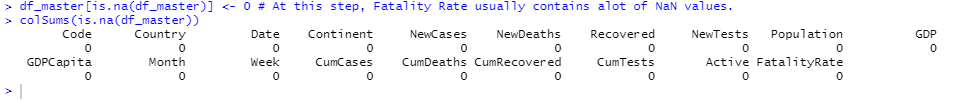


## Q2:

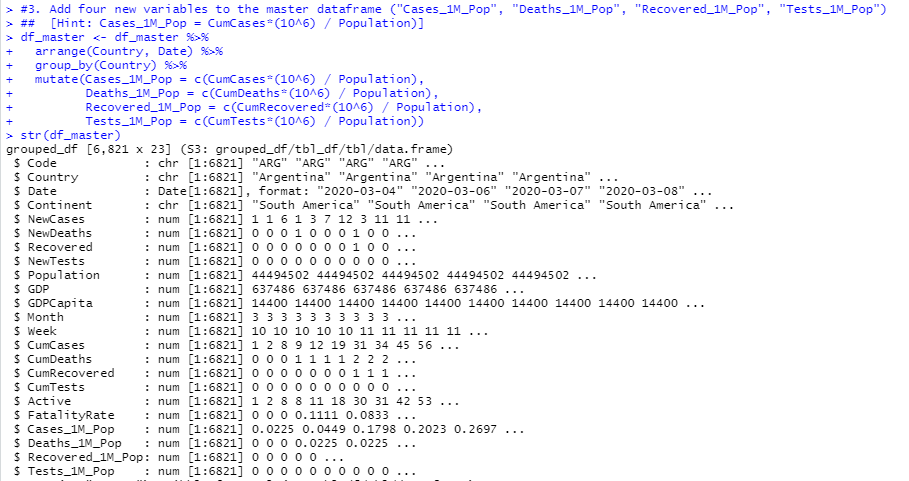


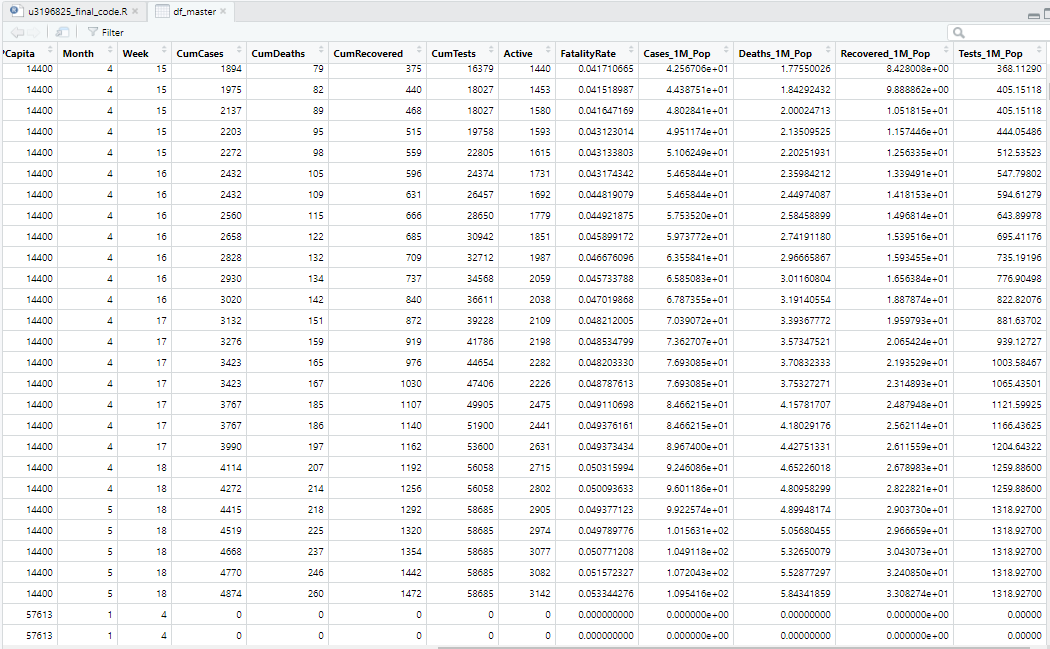
NOTE: however, at this state, I discovered the there are number of NaN(s) in Fatality Rate (due to divide by 0 in some rows). So, I change them into 0 for easier calculation.



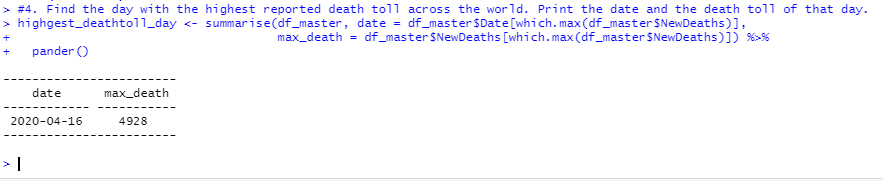


## Q3:

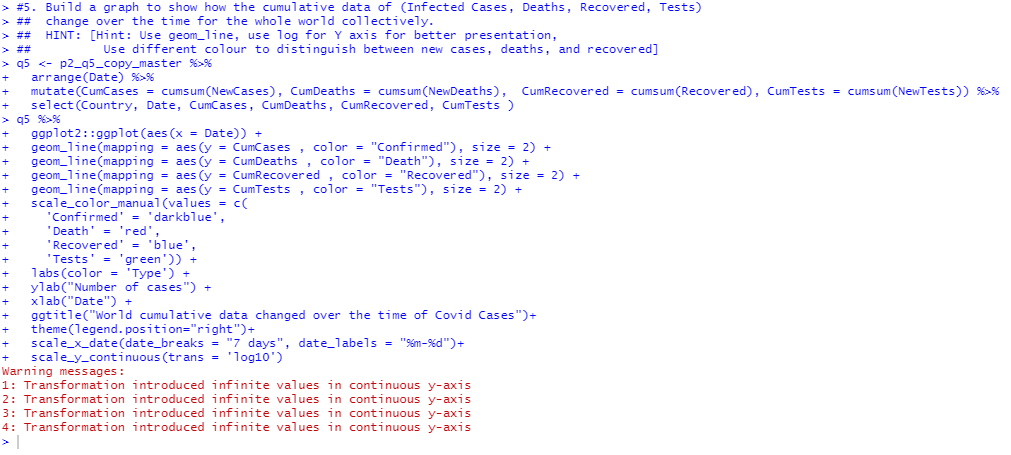


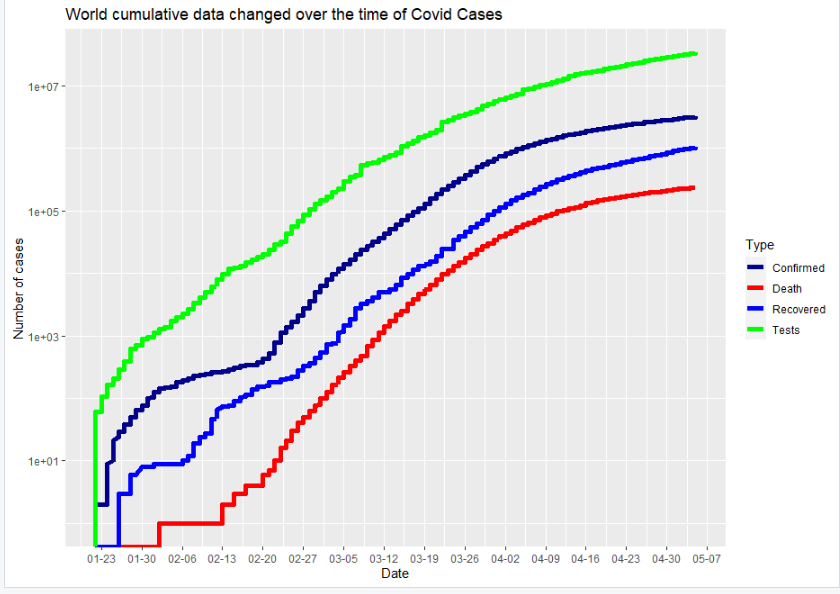


## Q4:

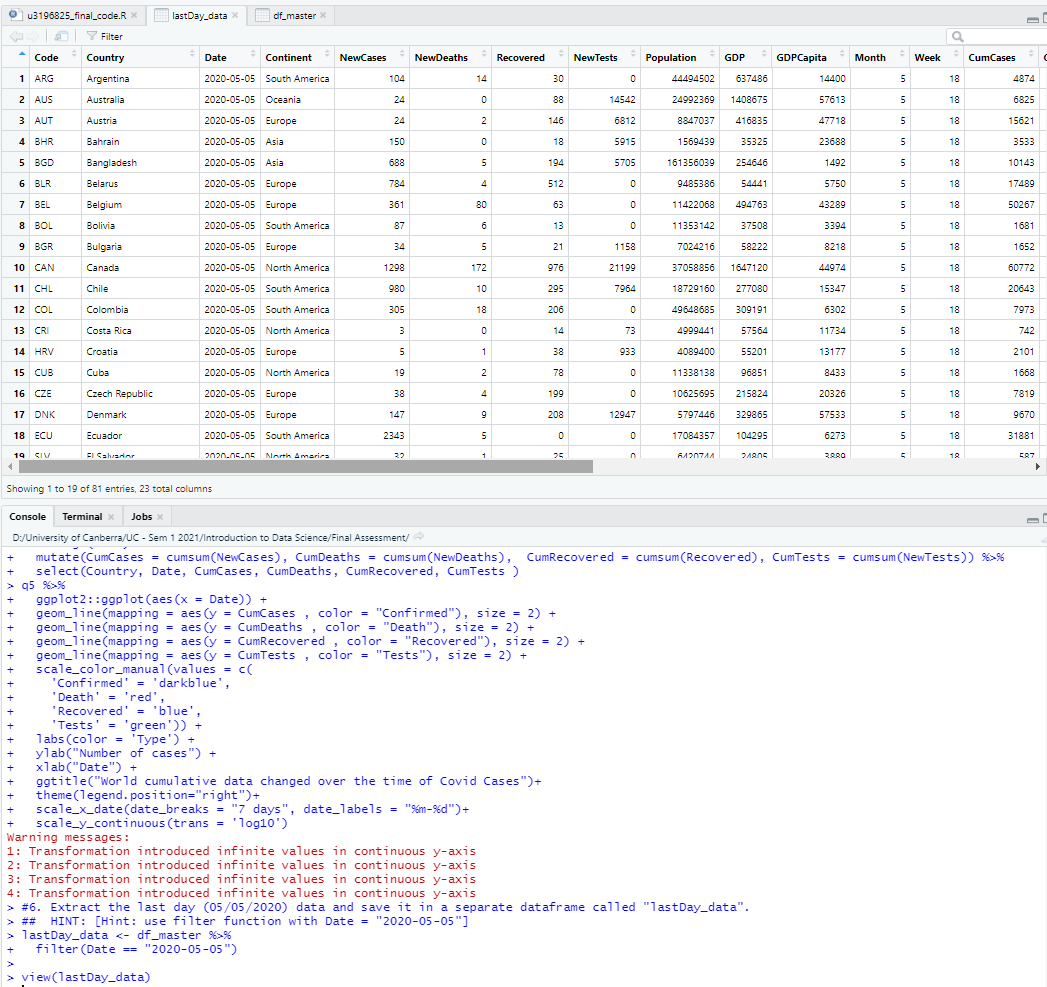


## Q5:

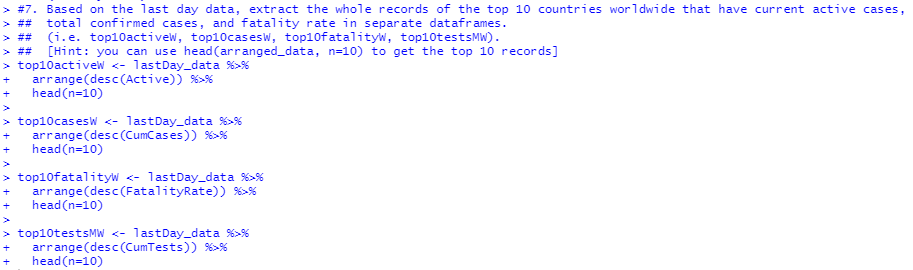




## Q6:

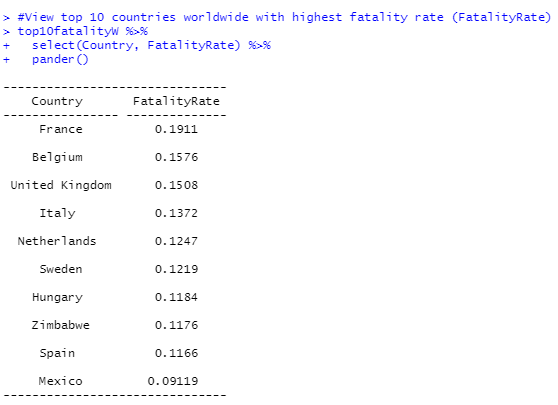


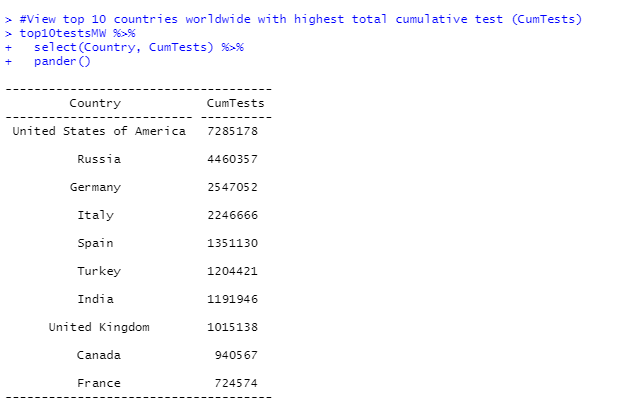
## Q7:



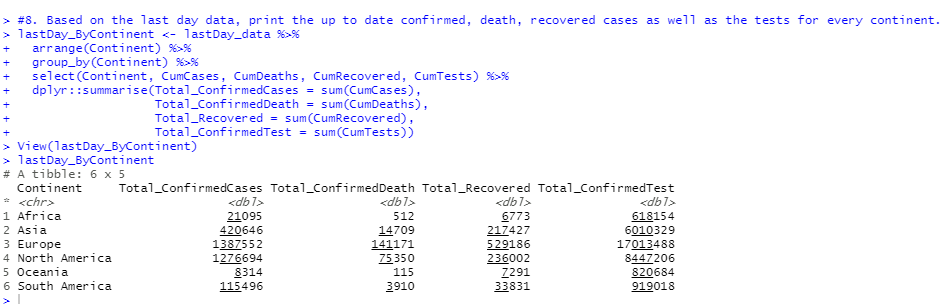




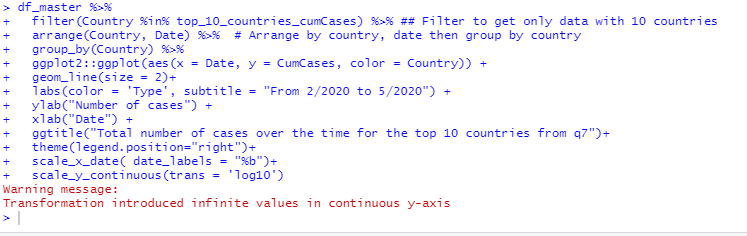


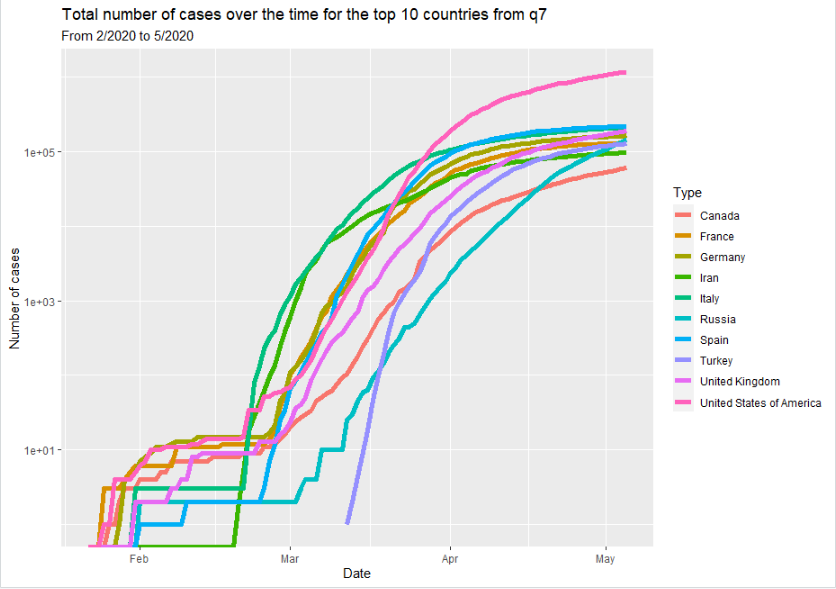


## Q8:

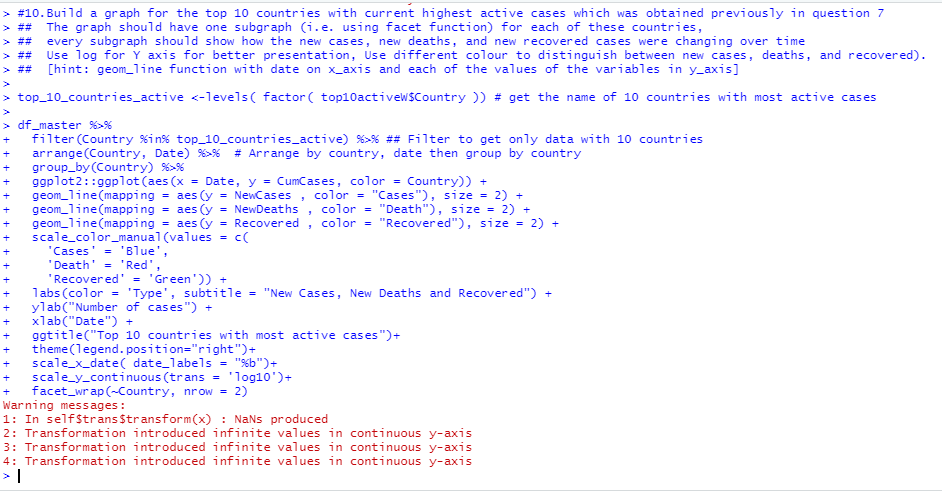


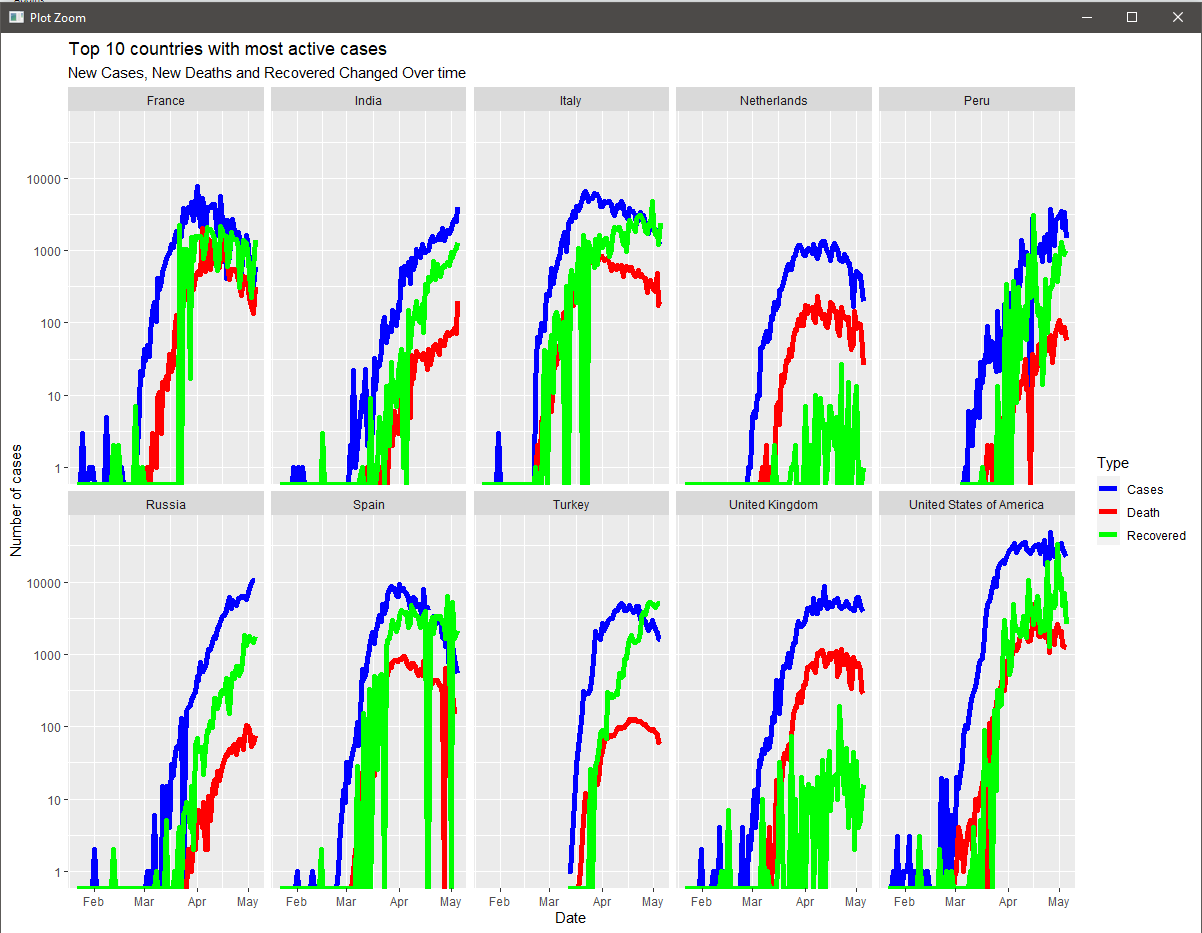
## Q9:





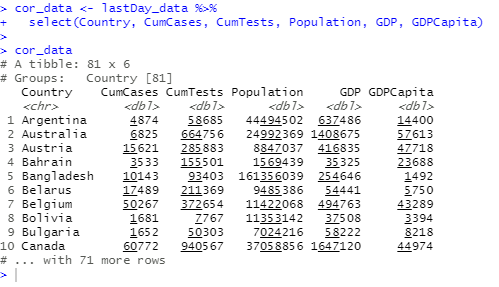
## Q10:



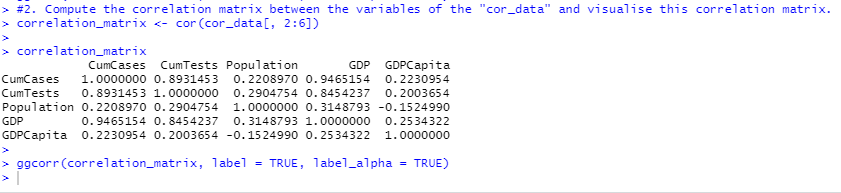


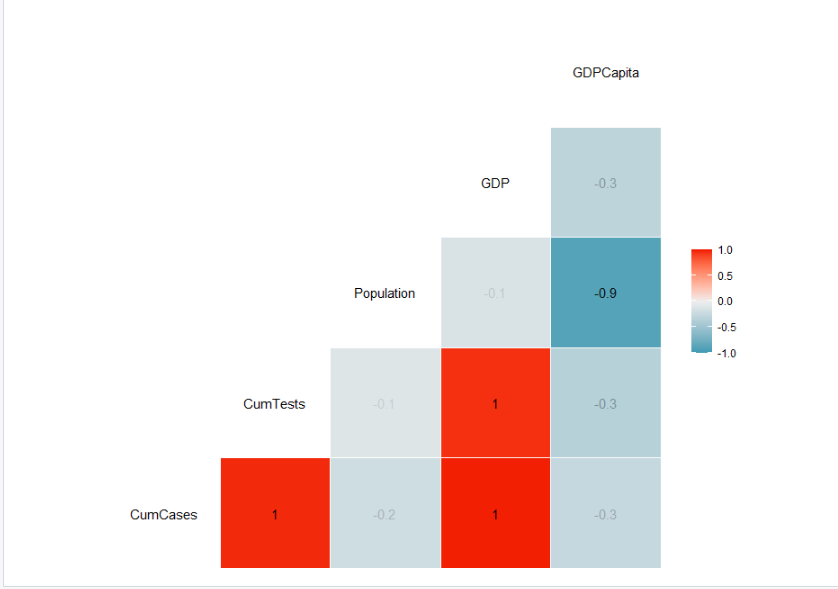
# Task 3: Data-Driven Modelling:

## Q1:

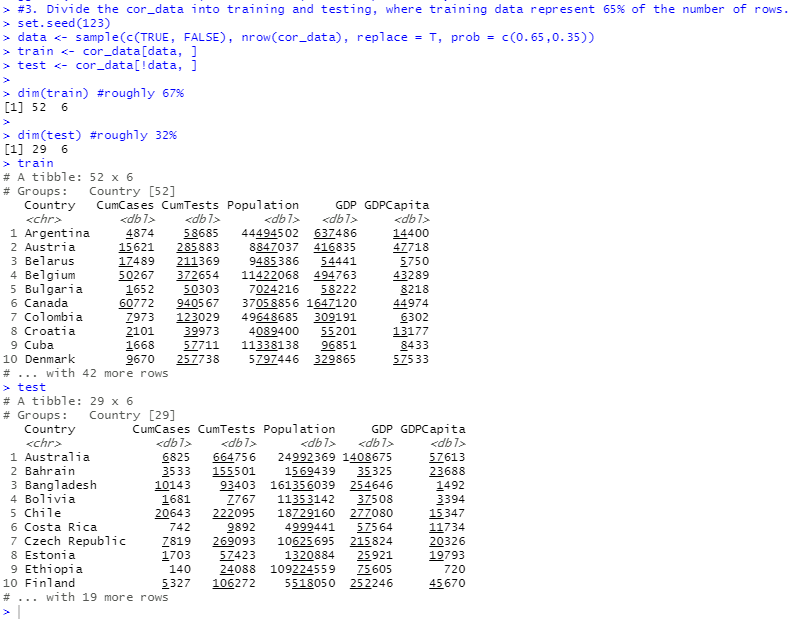


## Q2:

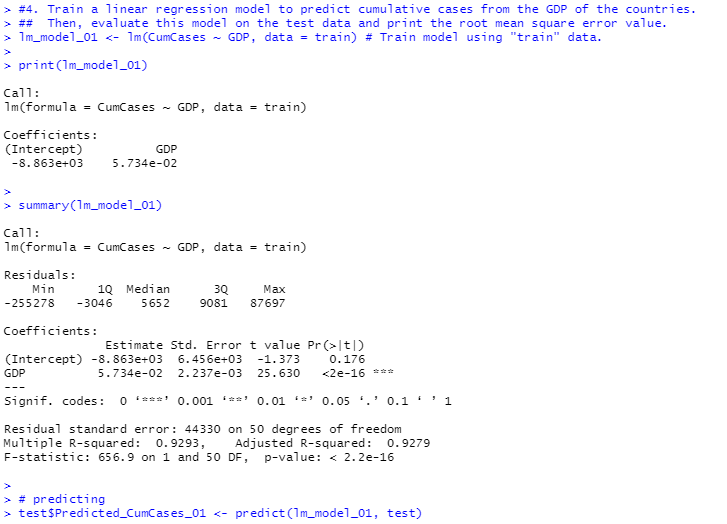




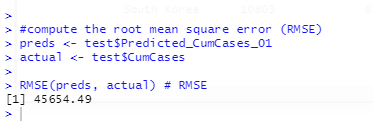
## Q3:



## Q4:







## Q5:

