# 

# 

Joint-Human Machine Learning

Assignment 1

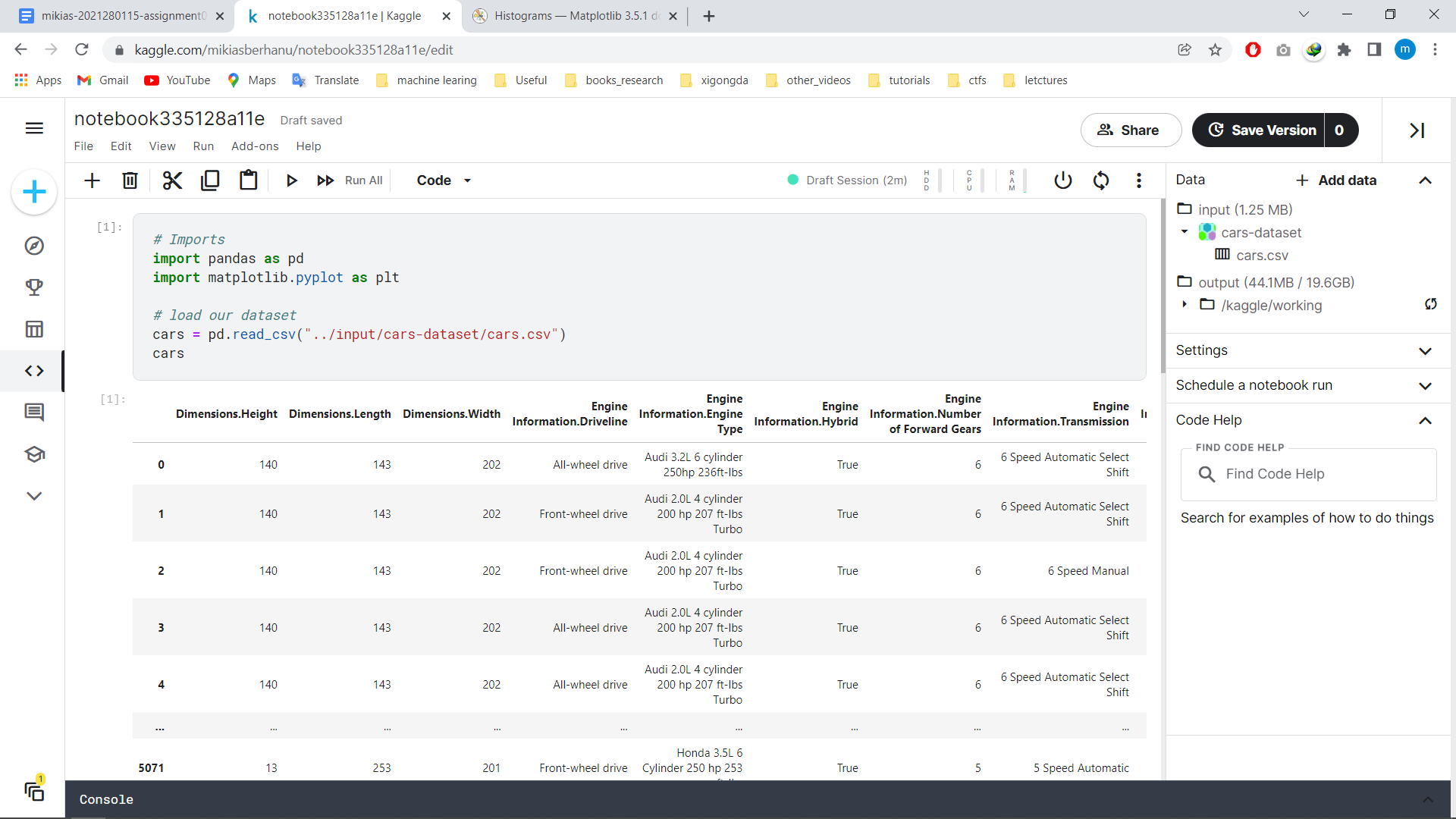
Mikias Berhanu, 2021280115

# 

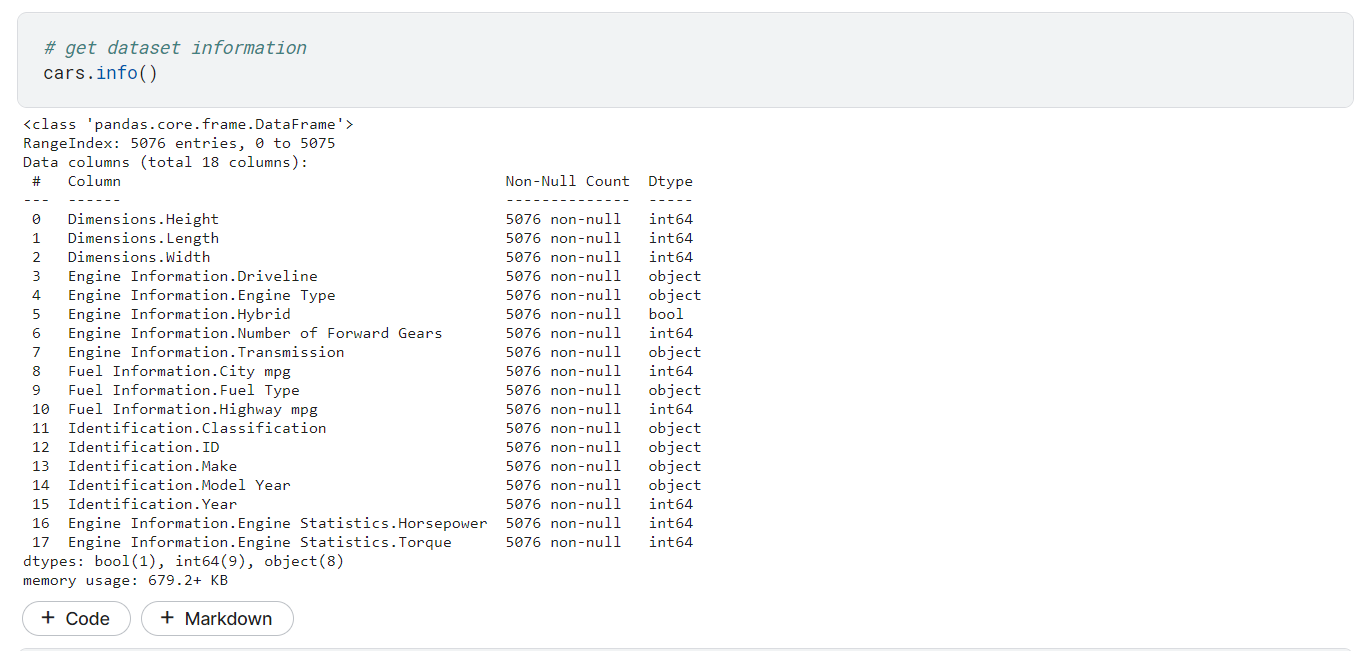
# Data Visualization using Charts and Different Plots

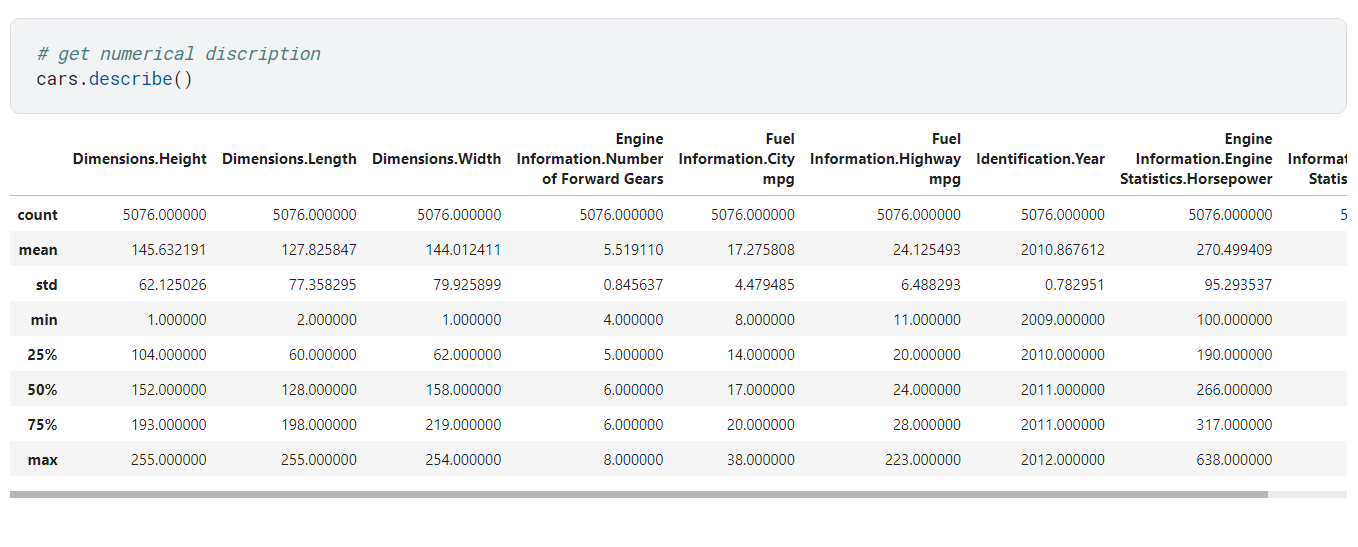
Data Visualization is a way of representing data in a graphic form which is an efficient way to describe data especially when it’s in the form of time series. Basically it's a process by which a large amount of data and metrics is translated into charts and graphs. Data visualization is important when it comes to understanding data and getting insights from it. Data visualization makes it easier for the human brain to understand patterns, trends, outliers etc… For this assignment I will use python programming language and a popular python library called Matplotlib which is used for plotting different graphs, charts and more and also pandas library which is used for reading and manipulating datasets with different formats. I’m going to use the kaggle notebook which is an online platform used for running data science and machine learning projects.

## Lab Assignment 1

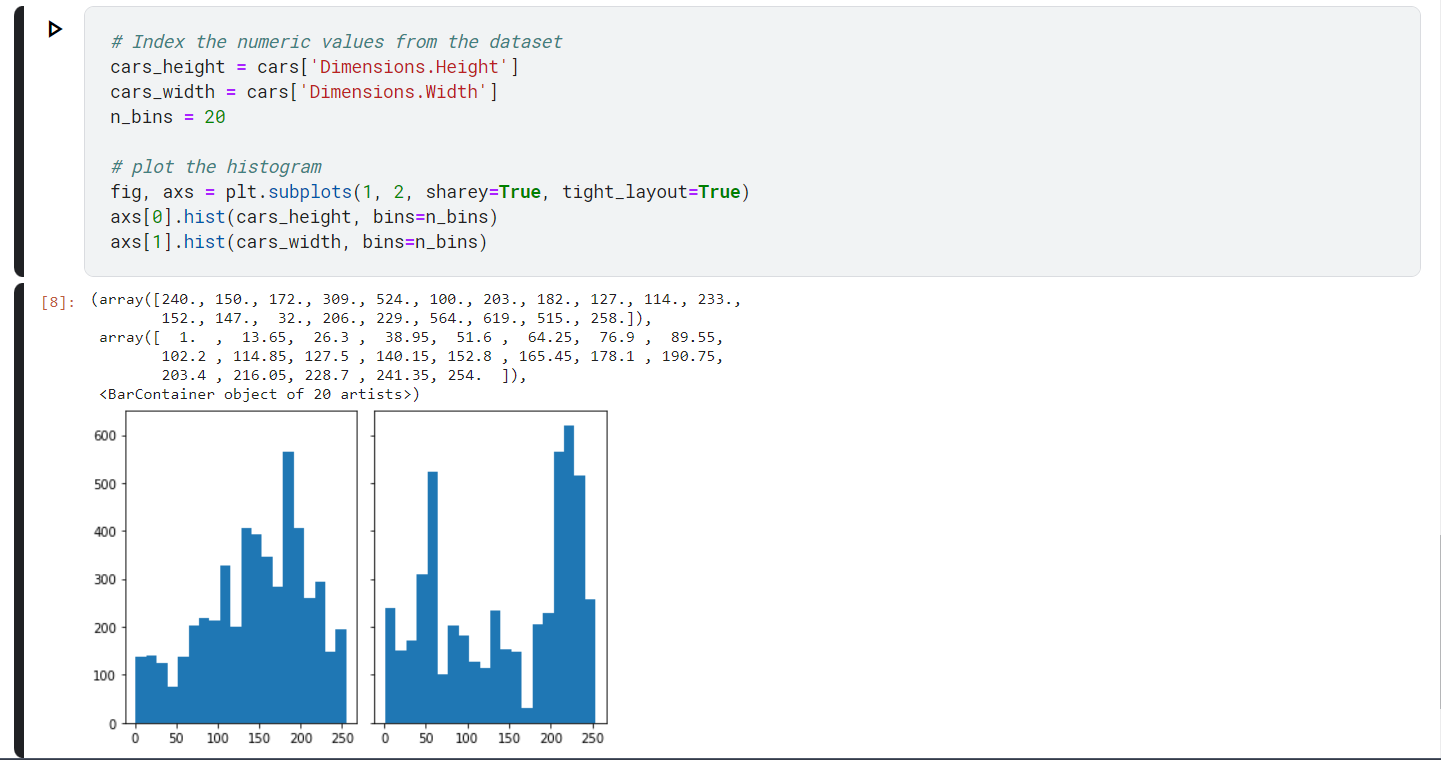
The first thing to do is to load our dataset to our workspace using the pandas library since the dataset is in the form of csv we can use the read csv method. 

Next we can get information and numerical description of our dataset using panda’s functions which makes everything easy to use. The info method tells us basic information about the dataset whereas the describe method tells the numeric structure of the dataset.





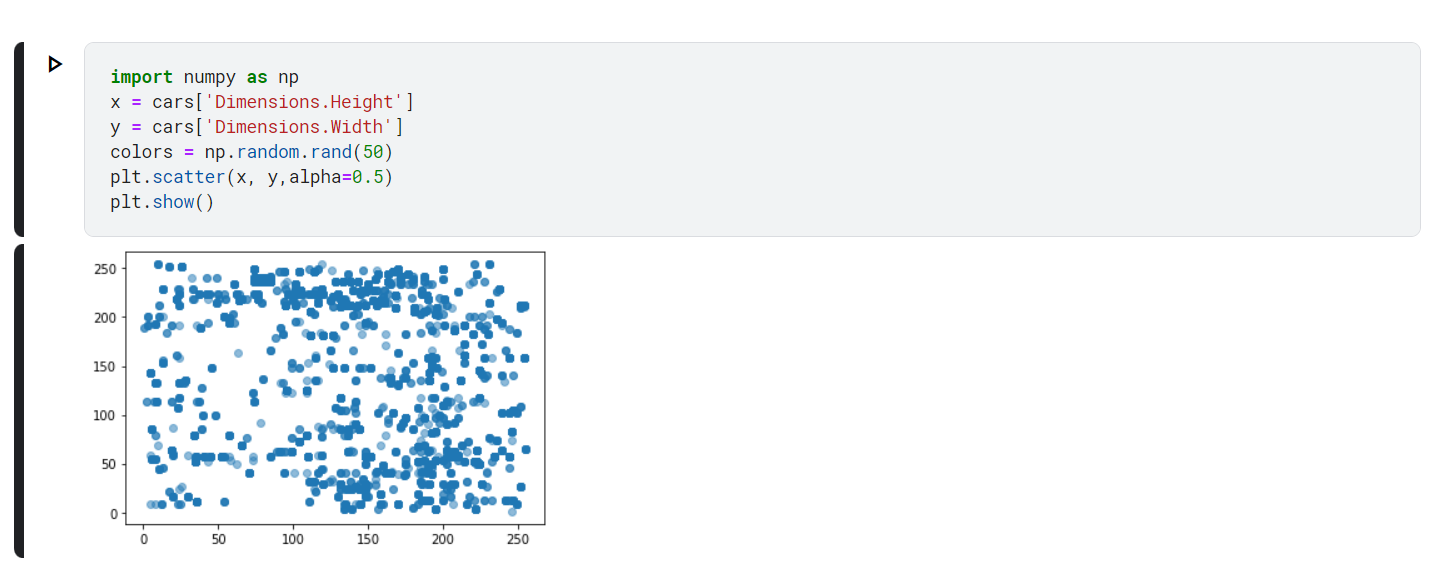
Now we can use the Matplotlib library to plot the histogram for our dataset. As an example if we want to get the histogram for the car’s width and height we can use the hist function from matplotlib. From the figure below we can see that the car's magnitude seems to be higher in some instances compared to the car’s height. The bin values are used to group the numerical data with equal width.



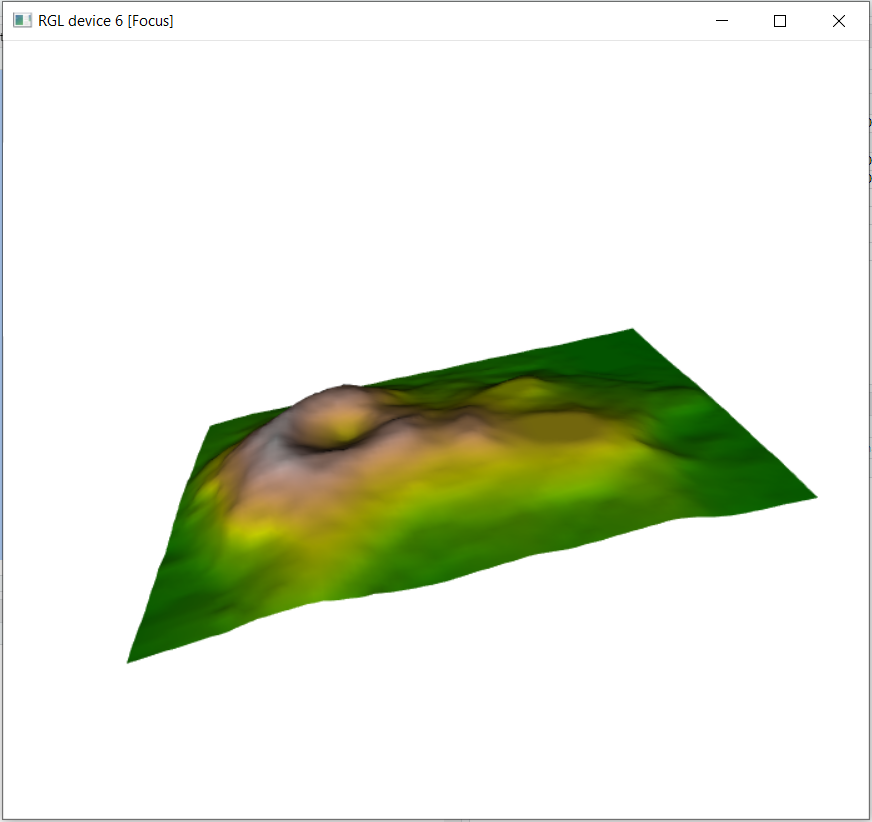
We can also plot a pie chart using the same technique for this case we can plot the car’s Fuel Information and HorsePower. For this I used the mean value of each column as a value for the pie chart, of course the values might change depending on the situation and the purpose of the pie chart.



Another graph we can plot is scatter plot which is a mathematical graph where data points are plotted on the cartesian coordinate plane. Usually this is done between two data points; this helps us to observe the relationship between the two variables. Let's plot the scatter plot to see the relationship between a car's width and car’s height.

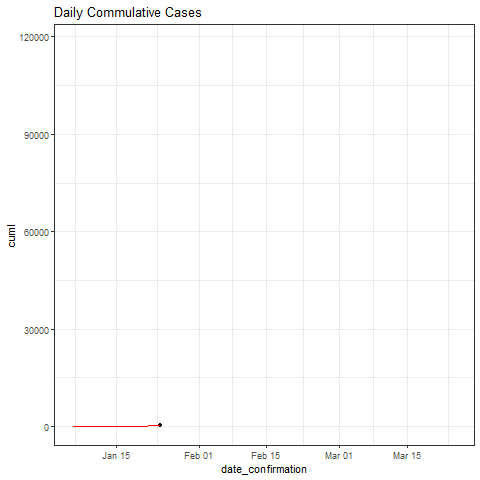


R programming language and R studio comes with a number of datasets and one of them is the volcan0 dataset which contains topographic data of the 50 volcano fields in Auckland volcanic field. We can plot 3D plots to understand the geography of the volcan and the dataset we have. The 3D plot clearly shows the geography of the volcano, the surrounding environment and also the volcan0 it self.



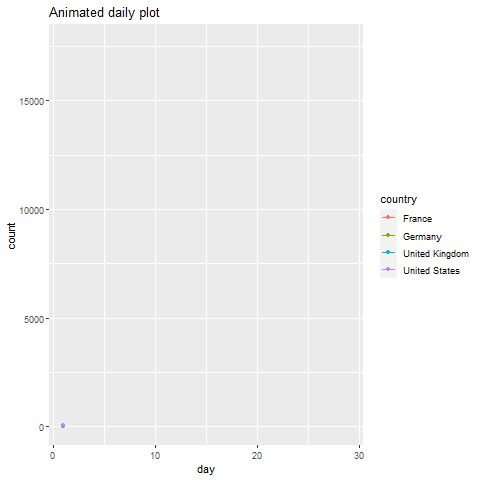
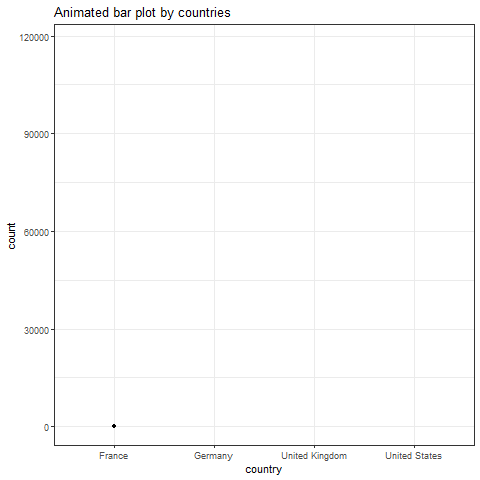
## Lab Assignment 2

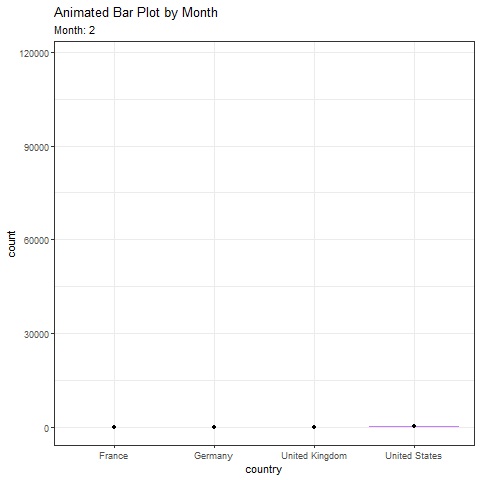
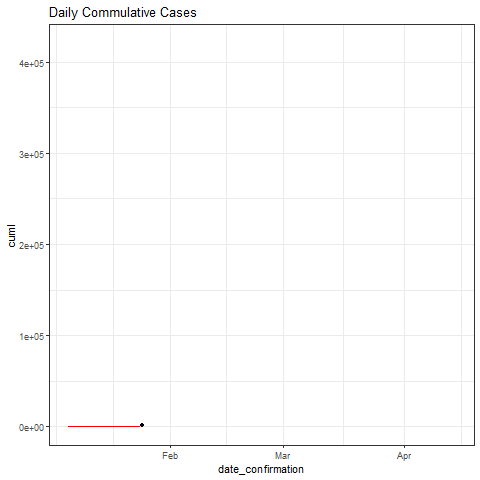
The second assignment is plotting animated time series using the covid dataset. This dataset has a number of information like travel history, source, symptoms, admission to hospital etc… We will plot a graph to see how to covid case spiked using the date of confirmation confirmation data, we will format the confirmation date so that it can be easily interpreted on the graph. This is done using the R programming language inside R studio. The graph shows that from january 15 the cases slowly start to increase and after march 15 the cases are at their highest possible number.



## Lab Assignment 3

The third assignment is plotting animated line plots and bar charts on the latest COVID dataset. The line plot is plotted for the United States, France, United Kingdom and Germany on a daily basis. This shows that the covid cases increase day by day almost exponentially from February to April. The bar plots are plotted for the same countries on a monthly basis and clearly shows that the United States reported the highest number of COVID cases compared to the other three countries and Germany reported the lowest COVID cases compared to the other three nations.





*The source code and images of this assignment are linked* [*here*](https://github.com/mikias21/JHML/tree/main/assignment-01)