**Name : Vishvesh Vikas Kadam**

**Student ID : 18210026**

**Data Visulization Assignment (CA682)**

**1 .Introduction**

My topic for Data Visualization Assignment is ‘What’s wrong with American health care’. For this I have collected data from different sources so that I can extract meaningful insights about US health care spending and life expectancy. Once I have clean data and everything is in place, I am showing visualization on three different charts like Multivariate Line chart, bar chart and World Map.

I have used Python libraries for creating my visualizations.

Library used :

1. Matplotlib
2. Plotly
3. Imageio
4. Pandas

**2. Dataset :**

So for the Line chart animation, I have fetched the following data :

Buzzfeed – Life Expectancy and Health care spendings. I downloaded the data from github site : <https://github.com/BuzzFeedNews/2017-05-us-health-care/tree/master/data>

Under OECD, we have the 2 dataset.csv files.

For Bar chart:

Buzzfeed - From same github link, under IFHP, we have Health prices.csv data.

For WorldMap, I have taken fetched data from 2 different sources-

Kaggle - ‘Life Expectancy data’: <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

I am considering only Country and Infant death column from above dataset

Data.worldbank: <https://data.worldbank.org/indicator/SP.DYN.AMRT.MA?view=map&year=2017>

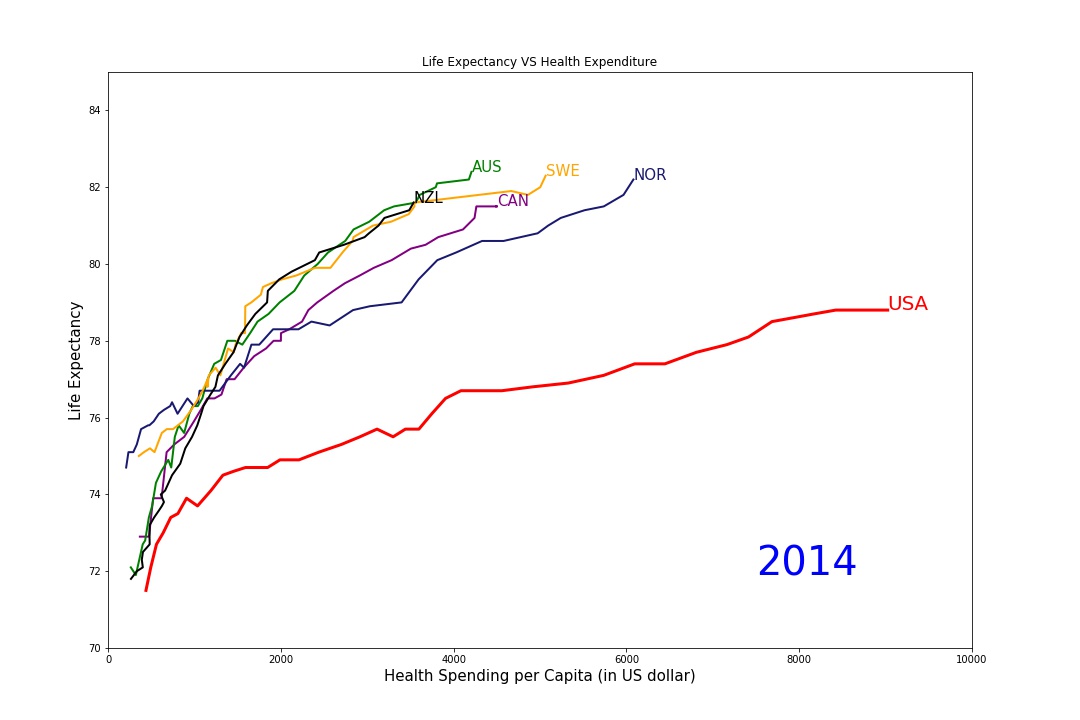
I used the above data just to get country codes.

**3. Process:**

Following was the approach that I followed while doing the visualization approach.

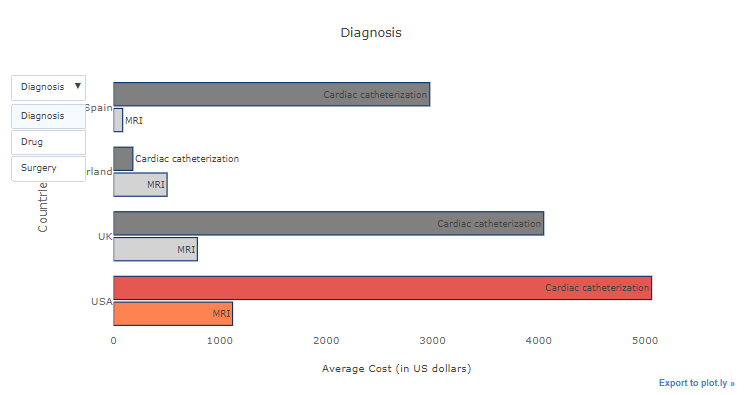
For Line chart animation:

1. I merged the two dataset ([health\_spending\_per\_cap.csv](https://github.com/BuzzFeedNews/2017-05-us-health-care/blob/master/data/OECD/health_spending_per_cap.csv) and [life\_expect\_birth.csv](https://github.com/BuzzFeedNews/2017-05-us-health-care/blob/master/data/OECD/life_expect_birth.csv)) from buzzfeed site.
2. The issue I faced was for Canada, there was missing years in Life Expectancy data. So for that I had to clean the data using my own algorithm and forward fill function in python.
3. Then, I merged the two dataset and used it to plot line graph of top developed countries iteratively from year 1973 to 2014. I used matplotlib for this chart.
4. Then, I used ‘imageio’ library create a gif animation from my output.



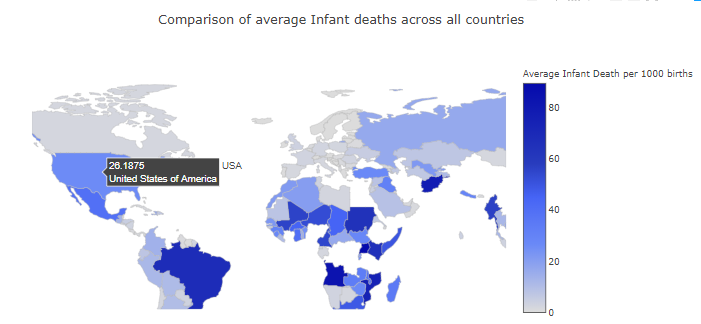
For my Bar chart interaction:

1. I imported prices.csv data.
2. There was spelling mistake for ‘Diagnosis’ at one place, which I corrected in python.
3. Then, I created multiple traces for bar chart and added update menu buttons for changing the medical type like Diagnosis, Drugs and Surgery.



For my last chart, i.e. World Map:

1. I imported 2 datasets : Life Expectancy data from Kaggle and mortality data from data.worldbank.org just to get country codes.
2. I did vlookup to get final data including country codes.
3. I obtained average infant death values for all years.
4. I saved it to new dataset called who.csv
5. Then, using plotly, I mapped infant death on map for countries having rate less than 100.



4. Report :

Following are that factors I considered while doing the visualization:

* I could only use world map to plot one variable, and so did my primary comparison on Line chart.
* Also the colour coding for map is done using hue so that we can get the comparison of other countries
* For Line chart, I have mostly used colors based on the countries national colors.
* Issues I faced was cleaning and filling Canada data as it had many null values.
* For bar chart, I choose to grey out other country spending so as to show the contract between US and other countries.

Following factors I could have included were:

* Including 2 level drop down. For example – after selecting drug: choose specific drug to further filter my bar chart.
* Also, I could have used interactive on line charts by using sliders for year. But as I liked the idea of animation better, I opted for that.