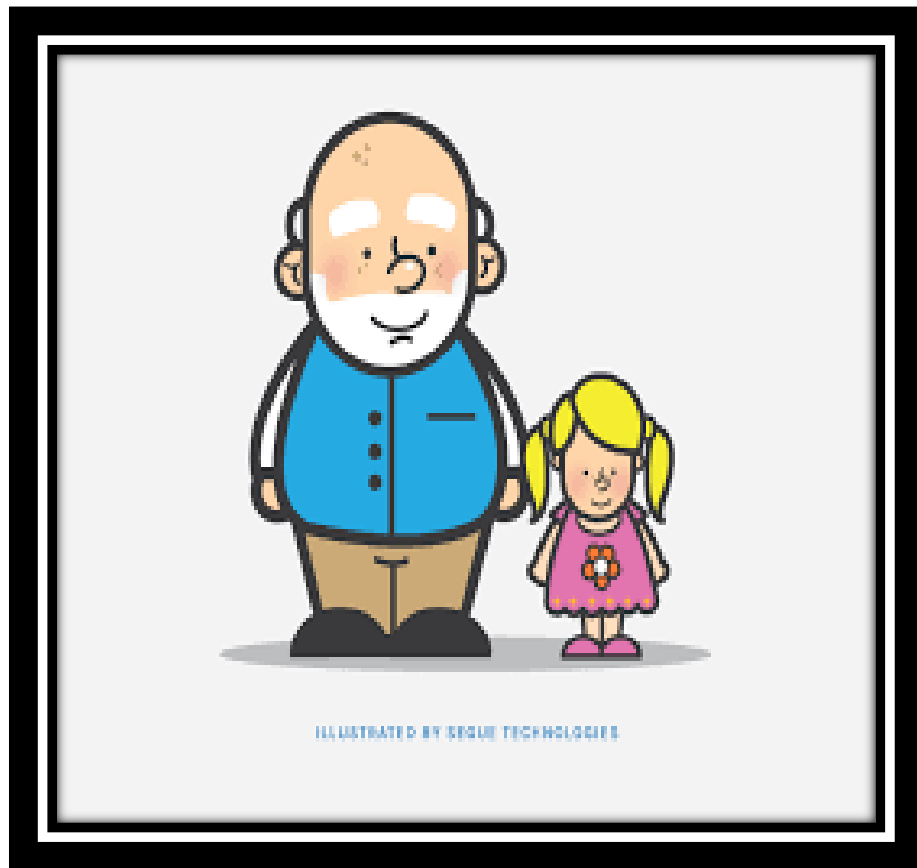


FIT5147-VISUALISATION PROJECT

LIFE STYLE ANALYSIS-(REPORT)



By: VIPUL JAD

Student ID.: 29520762

CONTENTS

INTRODUCTION	3
FOCUS	3
ABOUT DATASET	3
THE FIVE SHEET DESIGN	4
Sheet 1	4
SHEET 2	5
SHEET 3	6
SHEET 4	6
SHEET 5 Final design realization	7
IMPLEMENTATION	8
THE OVERVIEW	8
THE WORLD	8
IN THE MOTION	9
THE PREDICTION	9
THE LIBRARIES USED	9
USER GUIDE	10
CONCLUSION	13
REFLECTION	14
REFERENCES	14
APPENDIX	15

INTRODUCTION

In the present era we have realized that the life expectancy of humans has increased exponentially. Humans are living longer than they used to. But are they living a healthy lifestyle?

In this project we try to challenge the hypothesis that humans living a healthy lifestyle therefore life expectancy is increased with the fact that the drastic increase in the life expectancy is nowhere dependent on the healthy lifestyle of humans.

We will also try to prove the fact that humans are not living a healthy lifestyle using one of the unhealthiest and readily available thing "SUGAR".

It may be because of the unawareness in the society or negligence of the people, but sugar consumption has been drastically increasing around the world and in this visualization report we will try to address this problem too.

FOCUS

This Visualization report is mainly focused on people who are trying to live healthy or lose weight. Mainly people who are more into fitness and healthy lifestyle are the focus of this project. Additionally, Health department and common man can also take advantage of these visualizations as it delivers some useful facts about healthy living and life expectancy.

ABOUT DATASET

The datasets were collected from www.gapminder.org. Total of 6 different datasets were fetched and Wrangled using **Python-3**.

After Wrangling 4 different datasets were made as follows.

1. The Main Dataset which contains all the data for all the years (data.csv)
2. The Dataset having the Average Life Expectancy of each Continent calculated using Python (Continents.csv)
3. For Radar Chart we needed to Normalize/Standardize each column within the range of (0 to 1) for a better Visualization. This is done using the following equation

$$Xi = \frac{Xi - Xmin}{Xmin - Xmax}$$

4. The Outliers may obstruct the information to be delivered by our Visualization therefore another dataset is made by removing the outliers in GDP and Population.

Below are the attributes used for our narrative:

- **Year:** The exact year in which the data was collected.
- **Country:** The Name of the Country.
- **Male_BMI:** The Average BMI of all the male population within a Country.
- **Female_BMI:** The Average BMI of all the female population within a Country.
- **Average_BMI or Av_BMI:** The Average of BMI's of both male and female population per country calculating by using the following formula .

$$AV_BMI = \frac{Male\ BMI + Female\ BMI}{2}$$

- **GDP:** Represents the Gross Domestic Product or Purchasing Power Parity in Billions.
- **Population:** Represents the population of each Country in exact figures.

- **Sugar:** The average amount of sugar consumed by each country per day per person in grams.
- **Life Expectancy:** The Average Life Expectancy of a human in a particular Country.
- **Continent:** The Continent in which the Country is Located.

The normalized dataset has values from 0 to 1 for all the numeric columns. For example. China is having the highest population therefore its value in the normalized data will be exact 1. This creates a very non-meaningful visualization if we compare it with countries with much lower population such as Australia which is having the value of around zero in the graph.

To Overcome the above scenario, I removed the outlying countries in Population and GDP and allowed the user to choose an option to accept or reject the outliers.

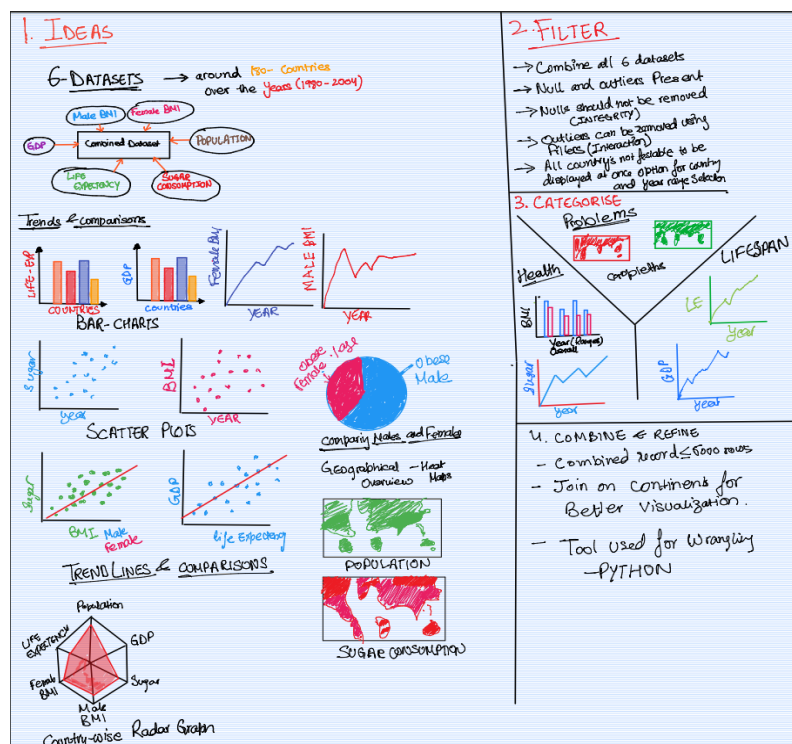
Unfortunately, the datasets I merged are from the year 1980 to 2004. The reasons for it is that

- For some of the datasets like Sugar and life expectancy the not a large number of datasets were present and therefore it would be impossible to make a combined dataset of all the information till the recent years.
- I am using R shiny for Visualization and it is slower in scaling a large dataset. It would take an impeccable amount of time for loading the visualization (graph) onto the web browser if the dataset was a bit bigger.

THE FIVE SHEET DESIGN

The design was concluded using a 5-sheet design methodology. From brainstorming the potential options to final realization of the dashboard it helped me to make a visualization dashboard presenting the story I wanted to convey. My final realizations are provided in the following 5 sheets.


SHEET 1



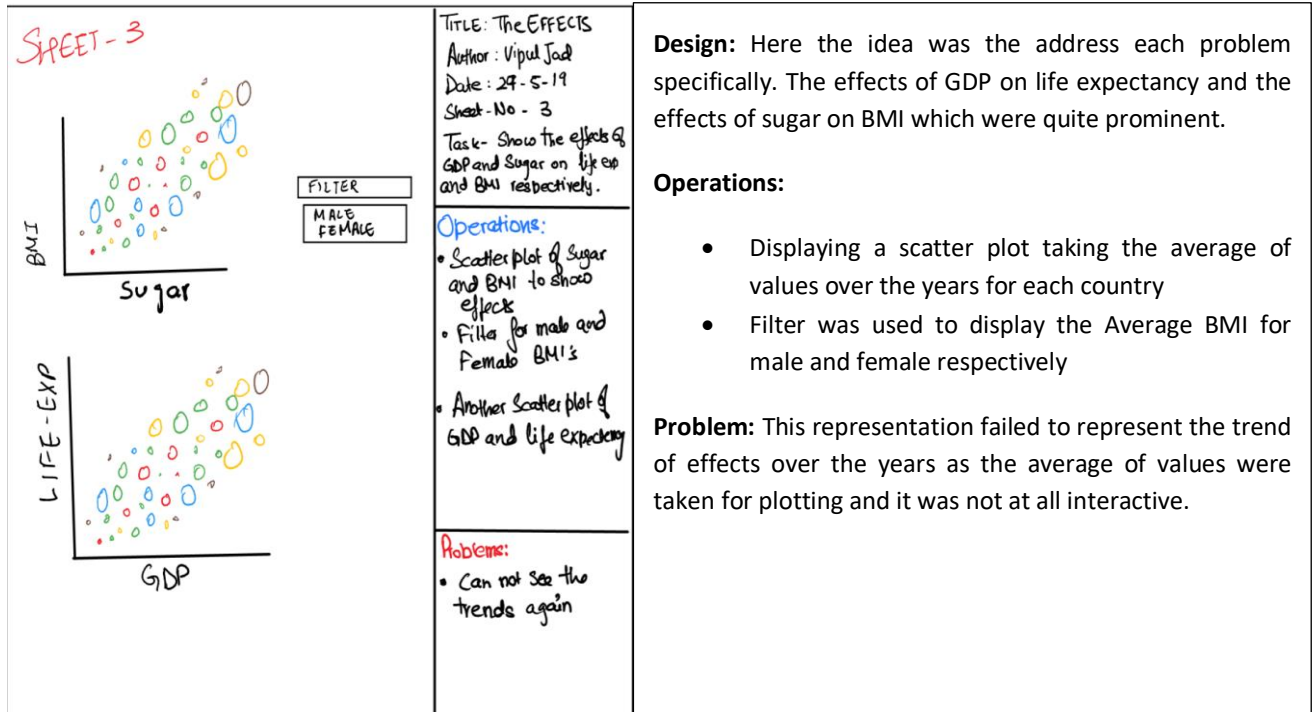
This is the first sheet consisting of all the potential Ideas and brainstorming which are explained in detail as follows.

- **The Ideas:** All the possible/ potential Ideas I had in mind to make the visualization dashboard convey my story. I was having many ideas like scatterplots, radar charts, line graphs choropleth Maps, pie charts and bar graphs.
- **The Filter:** Filtering of Ideas was done in this step. Outliers needed to be handled and the Nulls were needed to be rejected. Colour coding for countries needed to be handled in scatter plots as it was not meaningful enough to color-code all the countries uniquely.
- **Categorise:** Grouping of Ideas was done here. I needed to state an overview for my problem, then convey each problem specifically. Initially I made three categories for my visualization. The Overview/Context, Sugar - BMI effects and GDP - life expectancy effects.
- **Combine and refine:** The Ideas were combined, and the next sheets were made

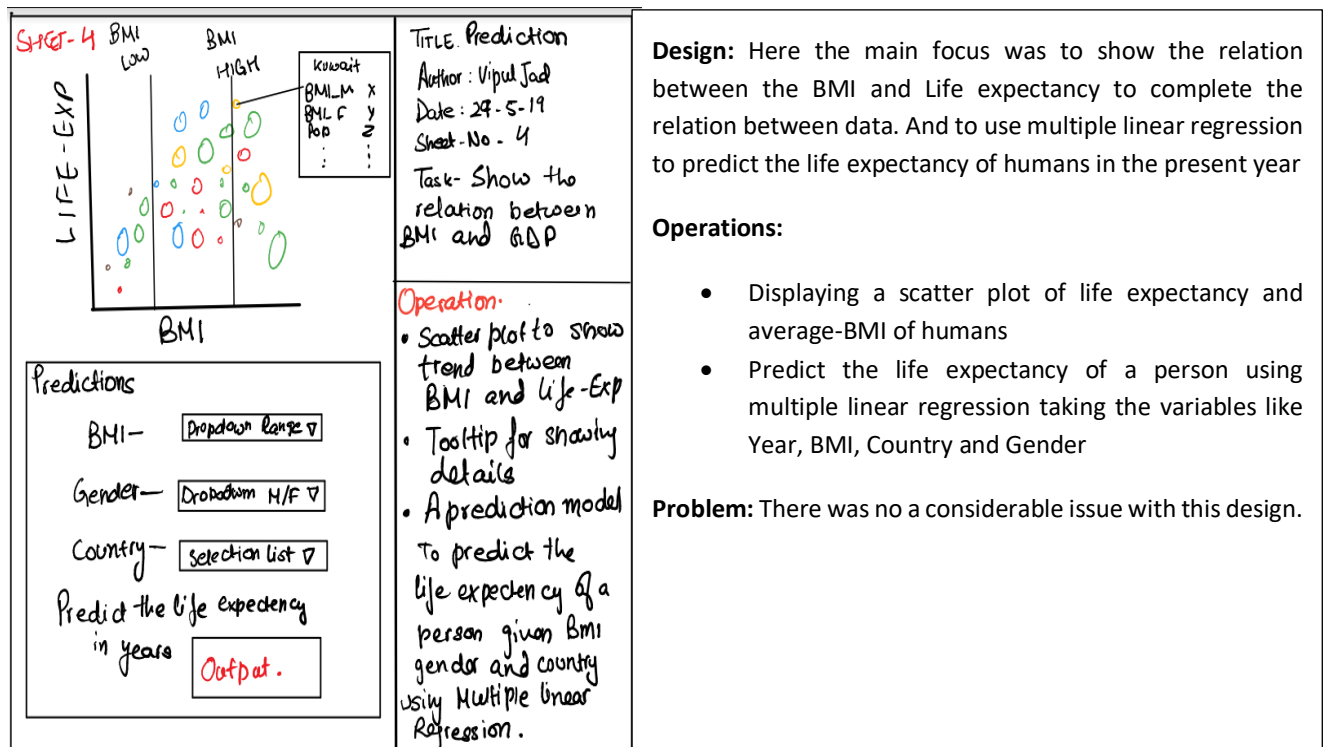
SHEET 2

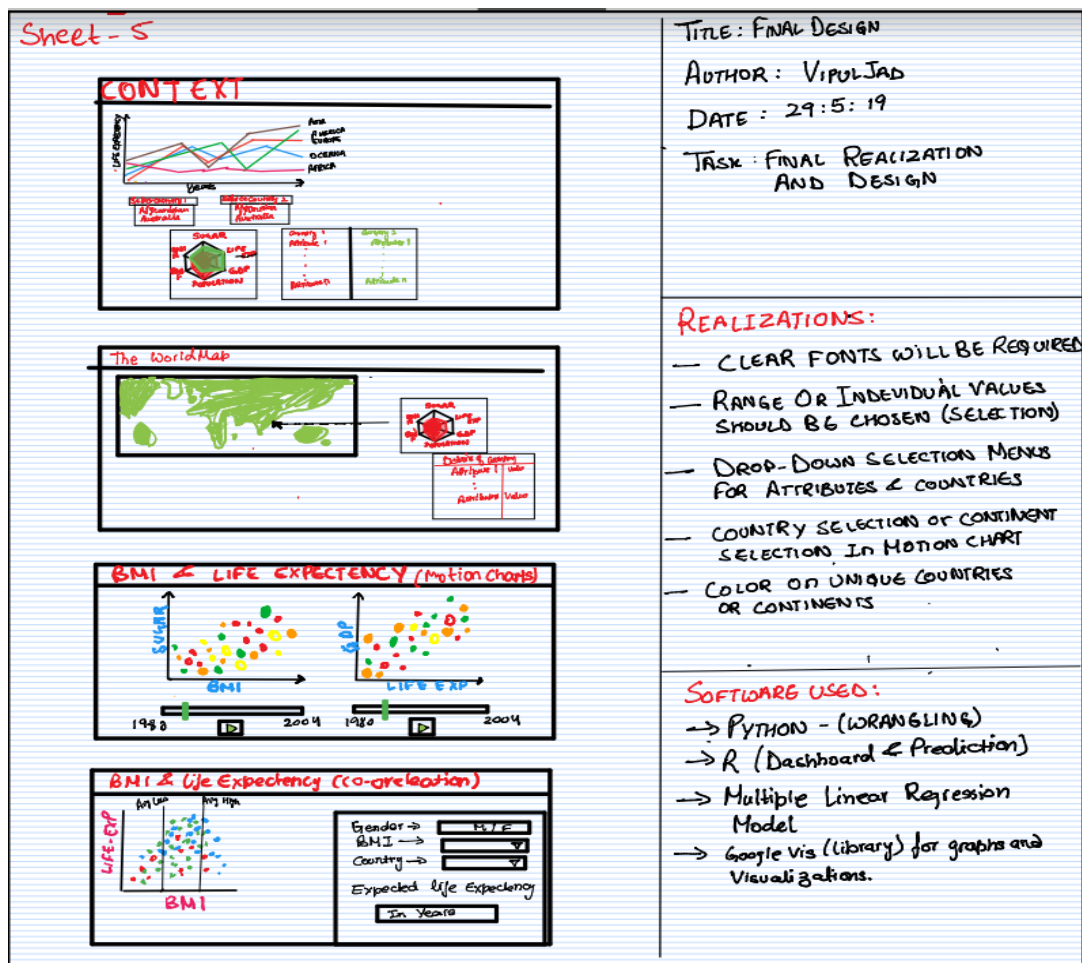
<p>SHEET-2</p> 	<p>TITLE: Overview Author: Vipul Jod Date: 29-5-19 Sheet No.: - 2 Task: deliver an overview of the problem</p> <p>Operations: → Give an overview Using a Heatmap of BMI and life expectancy → Filter - BMI - life expectancy → Radar graph delivers all the attributes of each country on selection</p> <p>Problems: - Trends can not be displayed</p>	<p>Design: The Idea to present an overview for the situation was to display a choropleth representation of average BMI or GDP on the map and displaying all the attributes of the country on clicking the country on the map.</p> <p>Operations:</p> <ul style="list-style-type: none"> • It had to give an Overview of the situation. The best option I had was a choropleth representation. • Filter was used to display the Average BMI or GDP on the Choropleth map. • On clicking on a country, a Radar Chart should be displayed and present the details about the country. <p>Problem: This representation was good for gathering the information but could not set an overview of the situation and trend could not be displayed.</p>
--	---	---

SHEET 3



SHEET 4





- **Design:** The final design has 4 tabs in it.
 - **The Overview/Context:** Delivering an overview of the situation or the problem. It has a motion Line Chart displaying the trend of life expectancy over the years. Followed by a comparison menu for comparing any two countries using a Radar chart and data table for more precise comparison
 - **The World:** It displays a worldwide overview of the situation using a choropleth map and on clicking on a country on the map a Radar chart Pops up displaying the attributes of the particular country followed by a data table to display its attributes and values.
 - **The Trends:** Displaying the two motion charts for delivering the effects of sugar on BMI and GDP on life expectancy over the years. An option for removing the outliers will also be provided. As motion charts are better at representing the trend of effects of a variable over the years it is highly suitable for our representation
 - **The Co-relation/ Prediction:** In this we first must display the trend of relation between the BMI and life expectancy using a cluster coloured scatter plot and afterwards taking the input from the user about Country, Gender and BMI we can predict their life expectancy using multiple linear regression

- Realizations: Range of values should be chosen like year or country therefore selection boxes will be provided and colour coding in motion charts should be done according to the continent rather than the country.
- Some changes were made while iterating through the 2nd, 3rd, 4th sheet like adding a linear motion chart and comparison menu in the first tab

IMPLEMENTATION

The dashboard was implemented using **R-studio** with **shiny library**. More details of the implementation are provided below.

The Dashboard contains 4-Tabs each of them are explained below respectively.

THE OVERVIEW

- This is the opening tab of the dashboard which contains a Motion Line chart and a Radar Chart followed by a table for comparing countries.
- The motion chart is made using **GOOGLEVIS** library and displays the trend of growth of Average life expectancy of different continents around the globe over the years.
- Each country in the motion chart is labelled with a unique colour for differentiation.
- Selection options are also provided with the motion chart to select and follow the trend of a continent.
- Following the motion chart is the comparison menu which takes a year and any two countries as input and compares them using Radar graph and a table to display the values.
- Radar graph displays all the attributes of the countries selected.
- Option for removal of outlying countries are also provided for a better visualization.
- Countries like China and India have significantly higher population than other countries therefore they must be removed.
- And the countries like United States and Kuwait have a much higher GDP than other countries and they must also be removed for better visualization.
- The radar chart was made using **PLOTLY** library and table using the **DT** library.
- During the normalization of data for the radar chart some countries have a very higher value for GDP and population.
 - All the numerical values were normalized between 0 and 1.
 - Following equation was used for normalization.

$$Xi = \frac{Xi - Xmin}{Xmin - Xmax}$$

THE WORLD

- This is the second tab of my dashboard starting with the choropleth representation of World Map with Average BMI as the colouring factor
- Slider is provided for selecting the data for a year on top to view the choropleth Map of each year if needed and respective values will be displayed in the radar chart and table below.
- On clicking on any country on the map a Radar chart will be displayed conveying the attributes of the country and a data table is displayed for viewing the attributes of the selected year and the country.
- The Map and radar chart are made using **PLOTLY** library and table is made using **DT** library.

IN THE MOTION

- This is the third tab delivering the trend of the effects
- It has two sub tabs. The first tab “*SUGAR EFFECTS*” has a motion chart delivering the effects of sugar on the BMI
- And the second one “*THE LIFESPAN*” is also a motion chart which delivers the effects of GDP on life expectancy an option for removing outliers is also provided for better visualizations as in the GDP data there are many outlying countries.
- Both visualizations were made using **GOOGLEVIS** library
- Both Motion Charts were used to display the trends of effects of sugar on BMI and GDP on Life Expectancy as in my case the motion chart is the best suitable way to represent a trend over the years
- The user can select a country either by clicking on the bubble or selecting the country from a selection checkbox provided on the bottom right corner of the chart
- Various option of selecting the size and colour of the bubble is provided
- Many other representations such as bar and line graph options are also provided if the user wishes to change the layout
- These motion charts clearly display the trends of the effects of both of our cases which is necessary for conveying my story.

THE PREDICTION

1. This is the last tab for the dashboard which starts with a scatterplot for displaying the final and most needed relation between the life expectancy and BMI.
2. This visualization lays a base for our prediction as we needed to show a relationship between life expectancy and BMI.
3. The Prediction is concluded using a Multiple Linear Regression Model using the formula
4. The dataset is splitted for male and female population and model is created respectively for each gender and Country
5. Input is taken from the user of the BMI, GENDER and COUNTRY and the prediction is made
6. As the GDP and country are related for each year therefore we can use the Country variable instead of GDP resulting in the following formula

$$\text{Life Expectancy}[i] = a + b_1 * \text{Year}[i] + b_2 * \text{Country}[i] + b_3 * \text{BMI}[i]$$

THE LIBRARIES USED

- **googleVis** – For presenting the motion charts (Library by google).
- **ggplot2** – For visualizations like scatterplots.
- **shiny** – For making interactive dashboards.
- **shinythemes** – For importing useful themes made by MIT.
- **plotly** – For plotting the Choropleths and Radar Charts.
- **DT** – For displaying the output table.
- **Countrycode** – For fetching the country code of each country – used in choropleths.
- **Shinydashboard**- for interactive functions like fluid row and box.

USER GUIDE

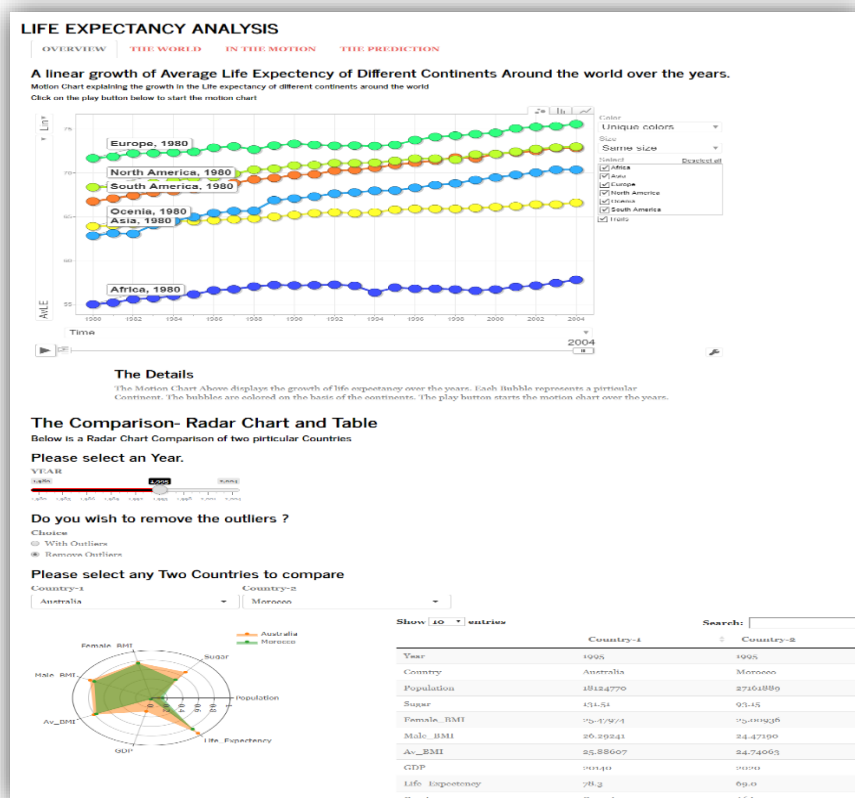
Please follow the following guide to go through the project.

- Before Starting
 1. Open the R script file in the submitted zip folder.
 2. Please install all the above mentioned libraries.
 3. After successfully installing the libraries run the shiny app through the script
 4. Please open the file in a Chromium based browser like Firefox or Chrome
 5. Please allow the flash player for the motion charts to load through the web page settings
 6. Please note if the flash player is not initialized in the browser the motion charts will not be displayed.
 7. The motion charts will take a while to load as the dataset is very large and complex.

- Going through the project .

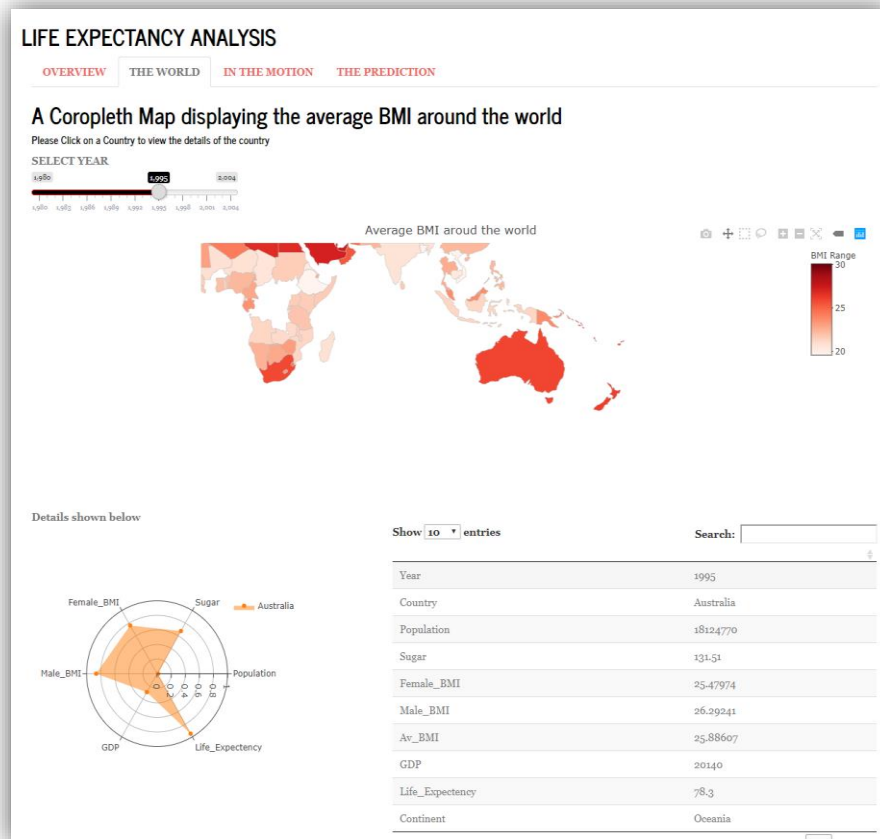
The project has 4 indevedual tabs each of them representing pirticulae story. This guide will provide a step by step process of going through the dashboard

1. On initially loading the dashboard into a browser please wait for the page to load completely as the motion charts take a while to load into the browser and make sure the flash player is allowed.
2. On loading the initial page will be displayed as follows.

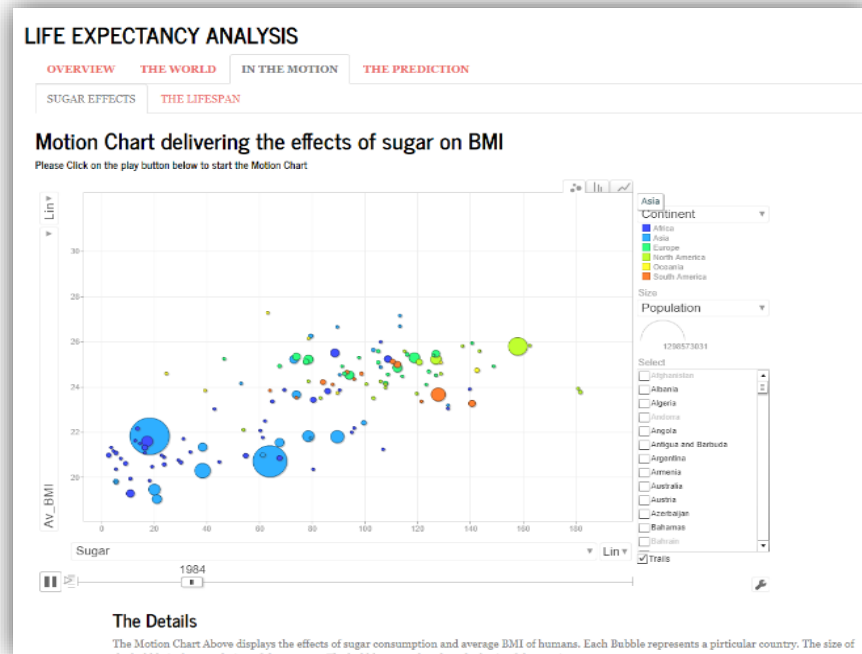


3. In the overview tab initially the Motion chart will be displayed please uncheck and check again the Trails option on the right hand side of the motion chart. It is set by default but not working initially with shiny library because of application integration issues of shiny and googleVis.
4. Click play button on the bottom left of the motionchart and the motionchart will start playing.

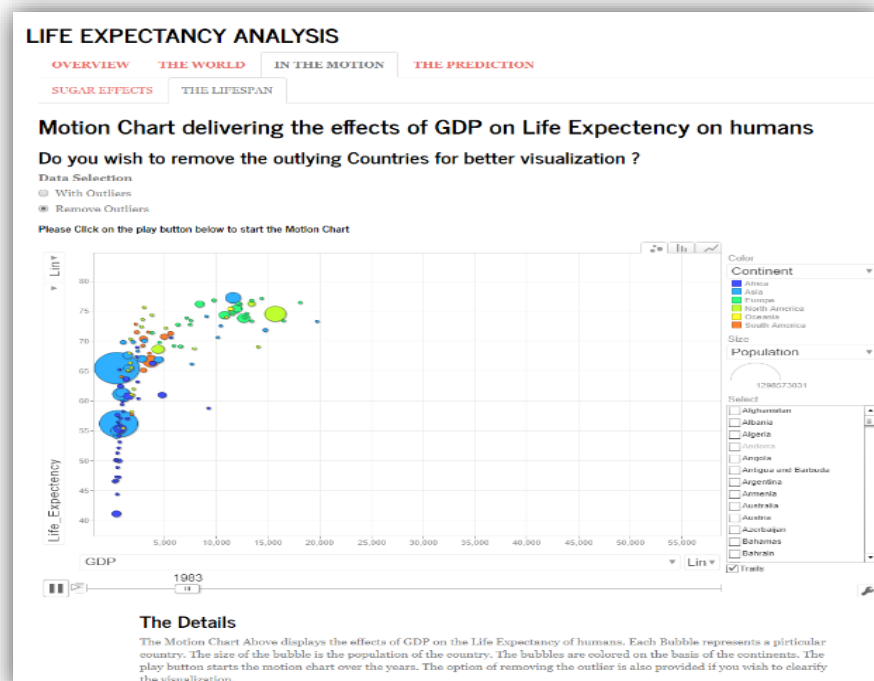
- After the motionchart is finished scroll down and in the comparison manu select an Year through the slider provided.
- An option to remove the outliers is provided if you wish to remove the outliers in the data
- Please select any two countries you wish to compare.
- After selecting both the countries a Radar chart will be displayed followed by a data table.
- The radar chart provides a overview of both countries with different colors and the datatable provides the exact value of both countries.
- Scroll back up and select the next tab -THE WORLD. Wait for the choropleth map to load.



- You will see a slider to select a particular year to view please select a preferred year.
- The choropleth map displays the density of average BMI around the world.
- Click on any country in the map. A radar chart will be displayed visualizing all the attributes of the country followed by a data table which shows the exact value of the country.
- Again scroll back up and select the third tab –“IN THE MOTION”.
- This Tab has two sub tabs the first one is “SUGAR EFFECTS” shown below.

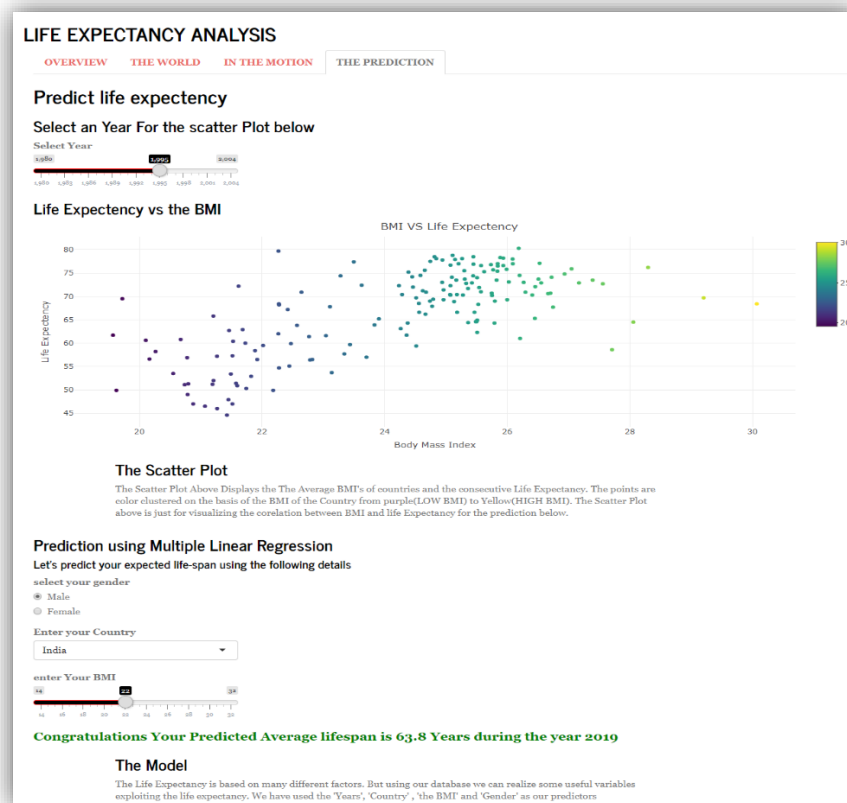


16. The motion chart will be displayed as shown above. The options to choose the color option and size are provided (by default the colour is set to the continents and size is the population).
17. If you wish to track the trend of one or more countries please click on the bubble or select the country from the selection box on the right hand side.
18. Please select the second subtab "THE LIFESPAN".



19. Here is the second motion chart displaying the effects of GDP on life expectancy.
20. You have the option to remove the outliers if you wish to.
21. The interactions are same as the previous motion chart.
22. Play button starts the motion chart.

23. Finally select the last tab “THE PREDICTION”.



24. Initially you will see a slider input for the year select any year you wish to see to visualise the particular year data on the scatter plot below.
25. The scatter plot displays the relation between life expectancy and BMI.
26. Finally the predict the average life expectancy of the user.
27. Start with selecting your gender followed by country and finally enter your BMI.
28. The predicted average life expectancy will be displayed for the year 2019.

CONCLUSION

To conclude, This is a interactive dashboard created using shiny and many other useful libraries for visualization. As the audience I target are mostly common people and fitness enthusiasts the dashboard has to be more simplistic and minimal. The visualizations should be clear enough to easily understand and provide useful information.

My visualizations are mostly motion charts which are very useful when visualizing a trend over time as the user can clearly understand the effects that are being displayed over the years and it is very interactive.

Following them by Choropleths which provide a overview of the situation that is needed to be presented. But just the Choropleths are not enough for the user to understand the message or change therefore backing them by Radar graphs and making it interactive was necessary. Through which user can understand the situation of the country along with the geographic locations.

Finally the predictions were made using Multiple linear regression which may not be the most prominent message but gives the user an idea of how the life expectancy is being affected.

REFLECTION

Through this project I have learned a lot about making dashboards, motion charts and choropleths and most importantly making user interactive dashboards using R shiny. How to take the mouse interaction events and get the data from it.

Data manipulation was also one of the most important part of the visualization for me as there were many outliers and the data was very much complex for visualization. The diversity of data made it impossible for me to find a meaningful visualization to show the trends of the events happening over the years.

Many improvements could be made with the dashboard but due to the time constraint and limited resources this dashboard is the best I could achieve.

Level of difficulty:

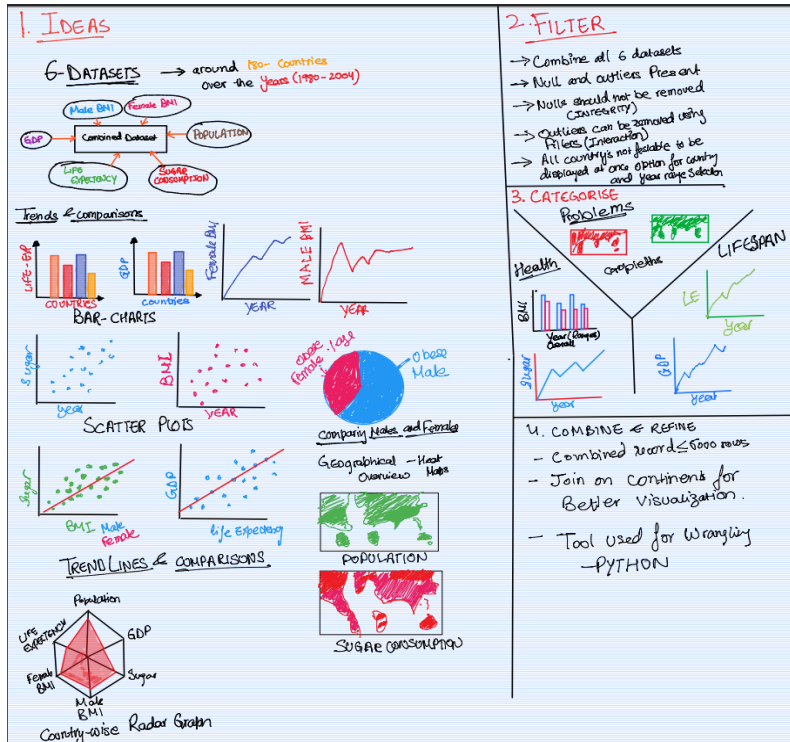
1. Large data (more than 4000 rows of data).
2. Normalizing the data for radar plots.
3. Finding and removing the outliers..
4. Multiple Linear regression.
5. Dynamically behaving on the user input.
6. Dynamically manipulating and subsetting the data on the basis of user input.
7. Taking mouse input by the user for the Choropleths and plotting the respective radar graph and table.
8. More simplistic design.

REFERENCES

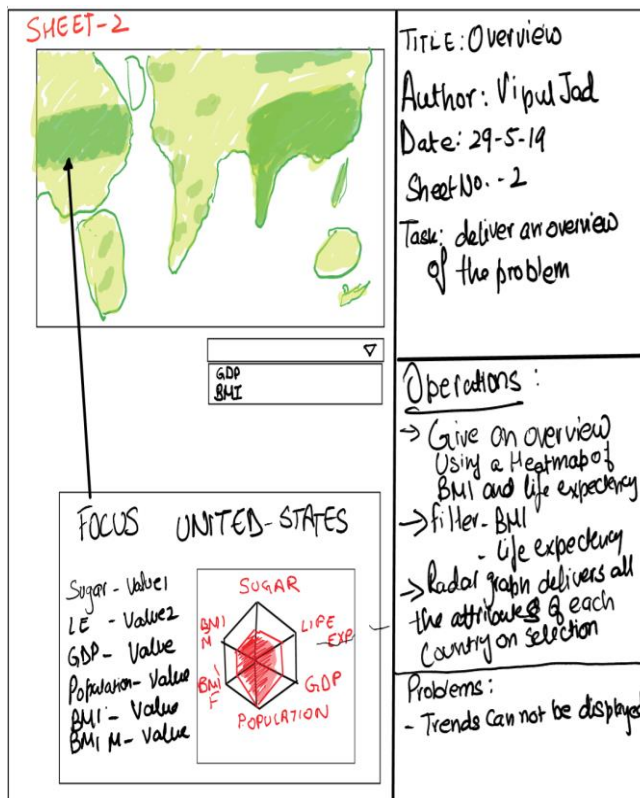
- Gapminder.org. (2019). *How Does Income Relate to Life Expectancy?*. [online] Available at: <https://www.gapminder.org/answers/how-does-income-relate-to-life-expectancy/> [Accessed 10 Jun. 2019].
- Roser, M. (2019). *Life Expectancy*. [online] Our World in Data. Available at: <https://ourworldindata.org/life-expectancy> [Accessed 10 Jun. 2019].
- Diabetes.co.uk. (2019). *BMI Calculator - Body Mass Chart, BMI Formula and History*. [online] Available at: <https://www.diabetes.co.uk/bmi.html> [Accessed 10 Jun. 2019].
- Wittwer, J. (2019). *Printable BMI Chart*. [online] Vertex42.com. Available at: <https://www.vertex42.com/ExcelTemplates/bmi-chart.html> [Accessed 10 Jun. 2019].
- Stack Overflow. (2019). *Stack Overflow - Where Developers Learn, Share, & Build Careers*. [online] Available at: <https://stackoverflow.com/> [Accessed 10 Jun. 2019].
- Shiny.rstudio.com. (2019). *Shiny*. [online] Available at: <https://shiny.rstudio.com/> [Accessed 10 Jun. 2019].

APPENDIX

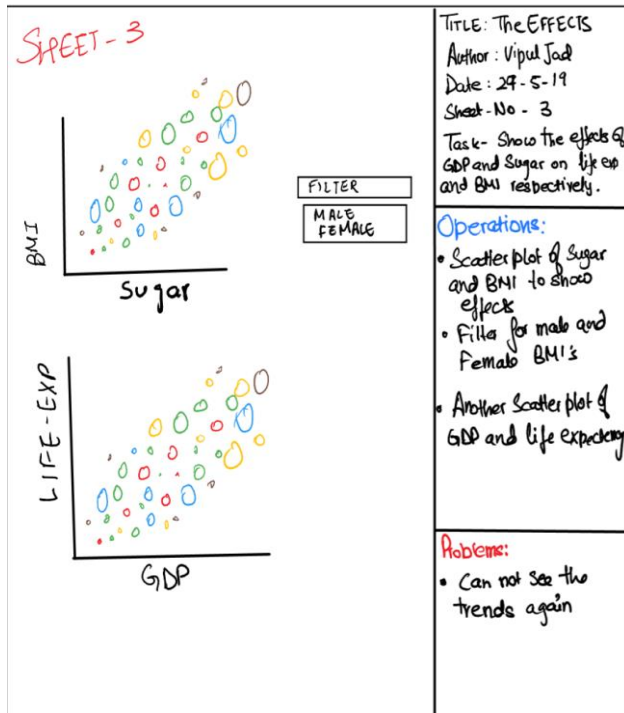
- Sheet 1



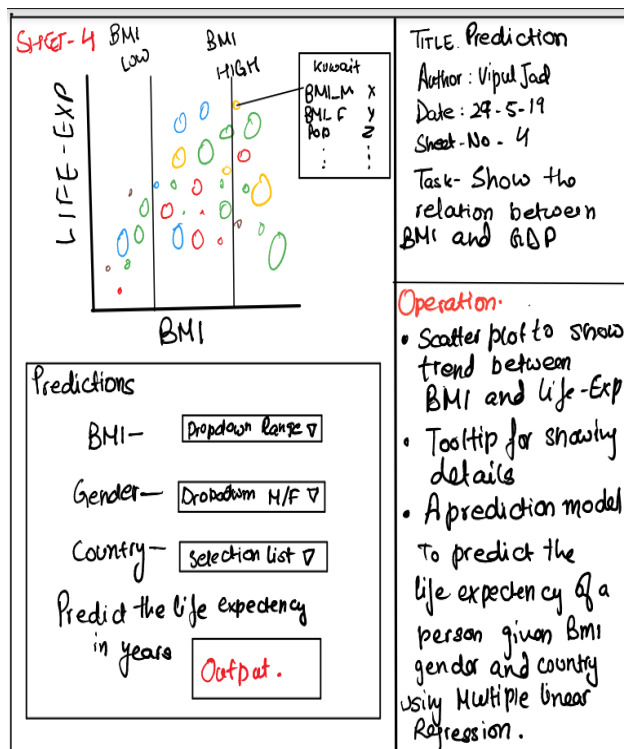
- Sheet 2



- Sheet 3

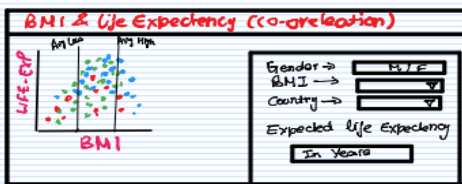
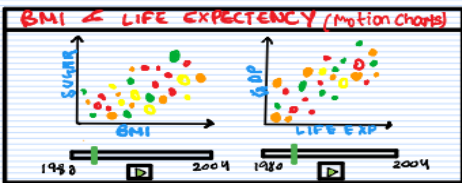
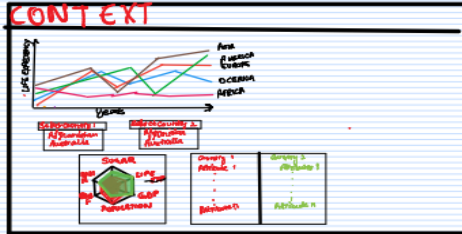


- Sheet 4



- Sheet 5

Sheet - 5



TITLE: FINAL DESIGN

AUTHOR: VIPUL JAD

DATE: 29:5:19

TASK: FINAL REALIZATION
AND DESIGN

REALIZATIONS:

- CLEAR FONTS WILL BE REQUIRED
- RANGE OR INDIVIDUAL VALUES SHOULD BE CHOSEN (SELECTION)
- DROP-DOWN SELECTION MENUS FOR ATTRIBUTES & COUNTRIES
- COUNTRY SELECTION OR CONTINENT SELECTION IN MOTION CHART
- COLOR ON UNIQUE COUNTRIES OR CONTINENTS

SOFTWARE USED:

- PYTHON - (WRANGLING)
- R (Dashboard & Prediction)
- Multiple Linear Regression Model
- Google Vis (Library) for graphs and Visualizations.