OBESITY EPIDEMIC IN THE USA

FIT5147 Narrative Visualisation
Project Report

MD SAADMAN HOSSAIN
STUDENT ID: 31043313
LAB:13
TUTOR: MOHAMMAD HAQQANI

INTRODUCTION

Malnutrition and food shortages have been plaguing humanity from the start of our civilisation. Only recently (after 18th century) due to technological advancement food supply has gradually increased. Globalisation has created the opportunity for people to access a tremendous variety of foods all over the world. This with many other reasons has created a fast-growing problem throughout the world. After world war 2, an overabundance of quick and easy food sources combined with lowered levels of physical activity (technological advancements in work environments and other areas) has led to an ever-growing problem of obesity.

For the purposes of this narrative visualisation project, I have chosen united states of America's obesity epidemic situation to highlight why we need to reduce the growth of obesity and what factors are related to it. The population of USA is multicultural and data sources from USA are also generally comprehensive. The aim for this project is to show the general public:

- How different states in USA are doing in terms of tackling obesity.
- How we can slow down the epidemic.
- Explore various socio-economic factors (race, education, gender, income, age) which can give us insights towards reduction of obesity levels.

I would like to address the following questions:

- Which states in the USA have the most and least obese/overweight population?
- Do activity levels if individuals affect the level of obesity in the US population?
- Does age or race have an effect on the obesity levels?
- Are there any differences in levels of obesity by gender? Why is this the case?
- Which factors are responsible for an unhealthy diet? Is there any relation between unhealthy food habits and obesity?

- Do social and economic conditions affect obesity levels?

Source of the data set-

Nutrition, physical activity and obesity – Behaviour Risk Factors Surveillance System. The data is collected by Centers for disease Control and Prevention and can be found in the USA government federal database (public use).

DESIGN

Before going ahead with the design of the narrative visualisation, I was considering all the available data sources and a good way to portray the data through interactive visualisations. I found some obesity data from CDC which contains obesity and physical activity data for USA. It also contains the behavioural risk factors associated with obesity.

Using the Five Design Sheet Methodology (FDS):

The first step of the FDS is to brainstorm to find all the possible ideas for visualisations. the diagram below contains the brainstorming portion of my FDS.

In this part of the FDS I was interested to explore all the ideas I could possible generate for the visualisations from the data source. the data source contains obesity, physical activity, unhealthy food habits data along with the socio-economics factors data for each US state and for years 2011 to 2018. The data also contains spatial data relating to the geographical locations for the US states along with categorical data for education, race, income, gender and age of the population. The time series data ranging from 2011 to 2018 is also useful for the visualisations.

Firstly, I thought about implementing a choropleth map which will show the obesity levels and physical activity levels for each US state. And it will have an interactive time slider which will enable the user to scroll through time and see the levels change over time. The concept of a choropleth map can also be represented using bubble

chart, dot density plot and descending bar chart (some of my alternate ideas).

The socio-economic factors in the data set are related to 3 variables (obesity, physical activity, and unhealthy food habits). To represent the socio-economic factors, I was interested to use pie charts, stacked bar charts and even boxplots (although I later discarded this idea as It would not be easily understood by the general audience).

I was also interested to show the relation of obesity and an unhealthy lifestyle. To portray this relation, I was considering using a scatterplot which allows the user to select the variables and display a scatterplot. I also contemplated using a bubble chart, in which the bubbles are the states, x-axis is unhealthy lifestyle and y axis is obesity. For the time series data, I considered using a line chart showcasing the levels of obesity, physical activity and unhealthy lifestyle over the years for the whole country (aggregated).

Design 1

In the first design, I considered implementing a choropleth map showcasing the obesity and physical activity levels for all the US states. Ans clicking on each state would generate a bar graph containing the aggregate percentages of the selected socio-economic factors.

Here, the layout includes a choropleth map which displays the intensity or aggregate percentage of either obesity or physical activity for the US states. There is a dropdown which allows selection of the desired variables. An interactive time slider will allow the user to switch from one year to the next. Hover/click on the states to display individual states information. We will also have a trend line appearing which shows the states' percentages over time. After clicking a state on the choropleth map the information generated will also include a bar graph which has a selection tab for the socio-economic factors.

The advantage of the choropleth map for this project's purposes include the ability to show case the varying levels of either obesity or physical activity levels across all the us states on a map. This allows for easy understandability just by looking at the map. I will also be using shades of red for obesity levels and blue for activity levels. The bar graph is also very useful when displaying categories. i.e. If I am able to showcase the obesity levels of difference races in the USA through the use of a bar graph (useful insights can be generated from it, which is then portrayed to the audience).

One of the disadvantages for this design is that it is a bit difficult to implement. The design is able to portray most of the messages I wanted to convey to the audience, but it contains a lot of unnecessary functionalities which are not very important. Also, it is not able to highlight the relation of obesity and an unhealthy lifestyle very efficiently.

Design 2

In the second design I opted for a descending bar chart which shows the states in USA in a bar chart. It includes an interactive selection of either obesity or physical activity to graph. The users will also be able to click on the bar or hover on it to display more information relating to the particular states. the click will generate a dot plot showing the percentages over time for the state and also generate a pie chart. The pie chart will include a drop-down selection for socio economic factors. i.e. if obesity and race is chosen, it will display a pie chart of obesity by race.

Some of the advantages for this design include easy implementation, clear view of the top states in terms of obesity or physical activity, the pie chart shows the contrasting categories pretty well. This design is also unable to show the correlation of obesity and an unhealthy lifestyle properly which is a major concern for my next design. It is

also unable to highlight properly why there is a discrepancy of obesity levels amongst the genders.

Design 3

The third design uses a bubble chart to display all the US states. here the bubble size is the percentage and the bubble colour correspond to the states. the operation for the bubble chart begins with a selection of obesity, physical activity or unhealthy lifestyle percentages. We will also have another drop down which enables the selecting of any particular US state. The default option produces a scatter plot with obesity against unhealthy lifestyle as seen above. The user will be able to select the variables for which to sketch the scatterplot. There will also be another drop down which will enable the user to select the socio-economic factors. It produces boxplots showing the distributions for the categories of socio-economic factors. for example, if the user chooses race, then the a boxplot of different races in the particular state and their obesity levels will be generated.

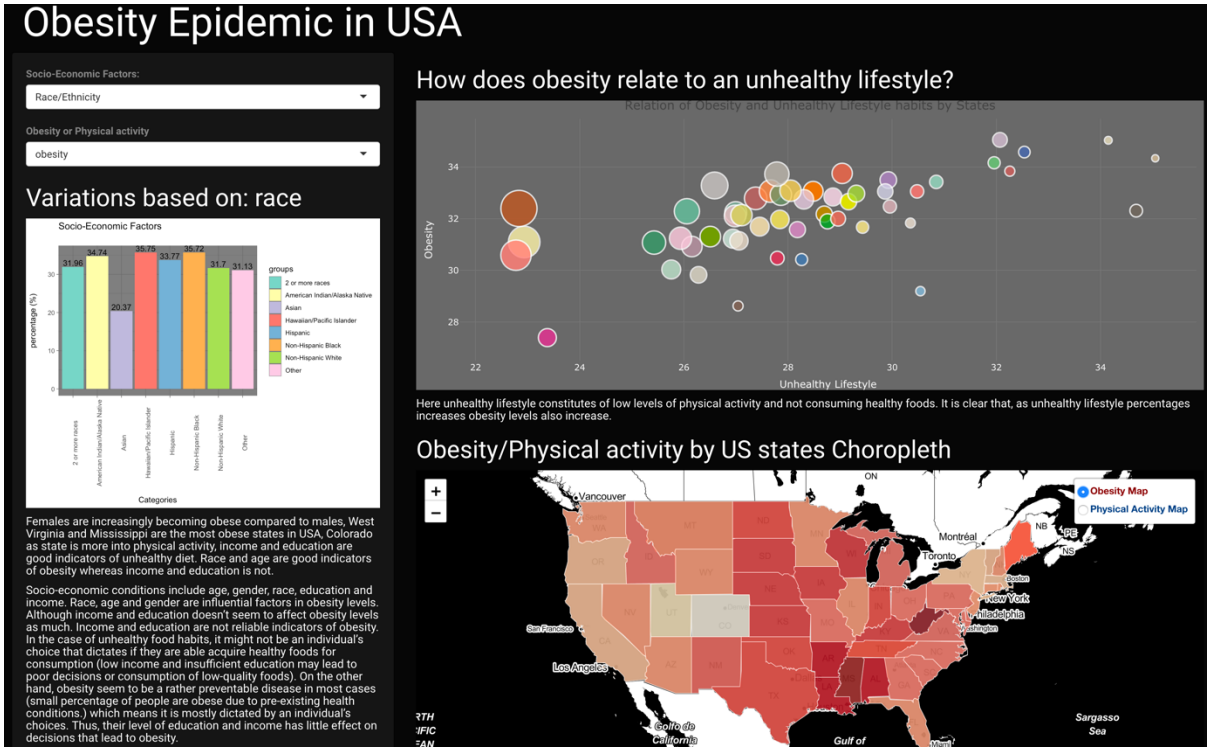
Advantages for this design include ease of implementation, ability to answers all the desired question for successful portrayal of messages. One of the disadvantages for this design is that it might be too complicated for an average audience to comprehend. i.e. the information shown in the boxplot and scatterplot maybe difficult to grasp for the general public. Also, this design allows too much interaction and too many options for the user. it will be difficult for me to compose a narrative story with so many options.

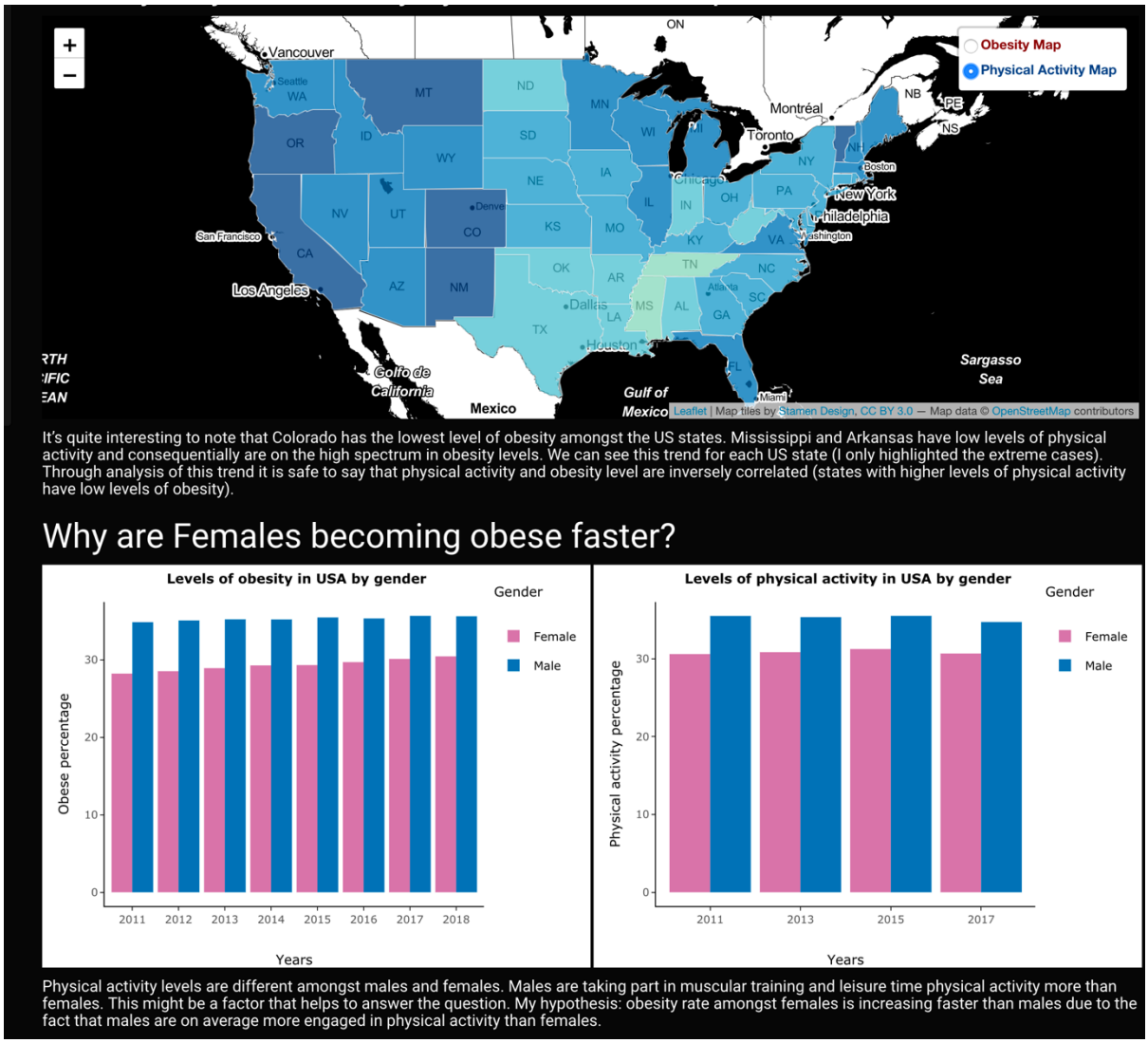
Realisation

After evaluation of the alternate designs (design 1,2 and 3), it was clear to me that the initial designs had pros and well as cons. To rectify the cons of those designs, I combined them into the realisation design taking the best aspects of the initial designs and trying to compensate for the cons. For the spatial data, I found that the choropleth map is the best suited. To show the correlation of obesity

and unhealthy lifestyle, I used a bubble chart which displays the states as bubbles. x-axis is unhealthy food habits and y-axis is obesity. For the socio-economic factors I used a bar chart which has interactive drop downs to select the factors and variable (obesity, physical activity). To show the disparity on gender in terms of obesity, I used a grouped bar chart. The realisation design is able successfully convey all the messages to the intended audience (general public). The realisation design offers a lot of interactive controls and also helps in portraying the current obesity situation in USA.

Final Design:





It's quite interesting to note that Colorado has the lowest level of obesity amongst the US states. Mississippi and Arkansas have low levels of physical activity and consequentially are on the high spectrum in obesity levels. We can see this trend for each US state (I only highlighted the extreme cases). Through analysis of this trend it is safe to say that physical activity and obesity level are inversely correlated (states with higher levels of physical activity have low levels of obesity).

Why are Females becoming obese faster?

Physical activity levels are different amongst males and females. Males are taking part in muscular training and leisure time physical activity more than females. This might be a factor that helps to answer the question. My hypothesis: obesity rate amongst females is increasing faster than males due to the fact that males are on average more engaged in physical activity than females.

As seen from the above pictures, the final design is able to communicate my findings to the audience. There are various user interactions in my final design which will be discussed in further details in the user guide section.

Implementation:

R was used as the platform to create this project. In particular Shiny (R package) was used to create an interactive web page for my narrative

visualisation. Previously in the data exploration project, I was able to create a csv file (orderdata.csv) which is a wrangled version of the raw data file that I sourced from CDC website. For the implementation of the shiny webpage, further wrangling needed to take place. I used various **libraries** from R such as:

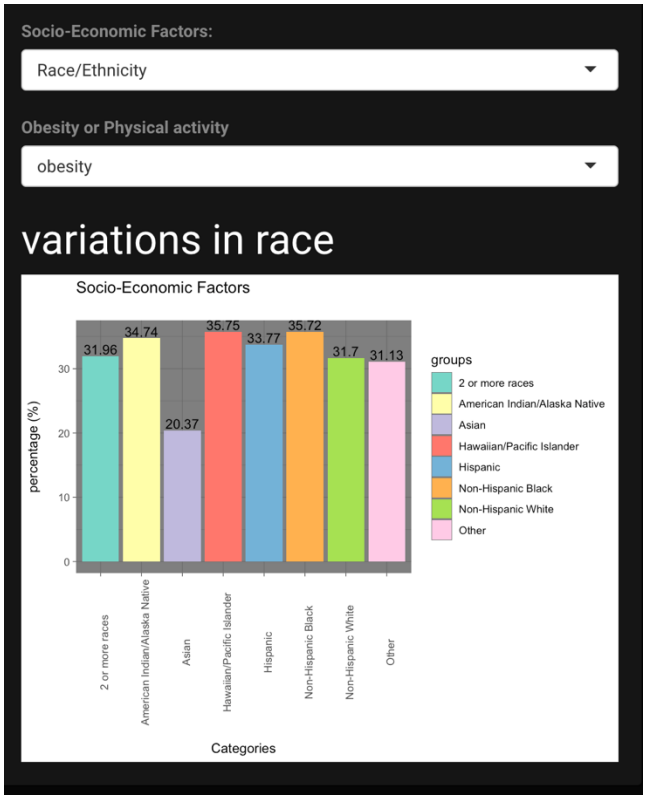
- Shiny: to create the interactive shiny app.
- Leaflet: required to implement the choropleth map and make it interactive.
- Ggplot2: ggplot2 is used for the creation of my bar charts.
- Rgdal: this library is used to use the readOGR function to read US shape files.
- Dplyr: dplyr library is very useful in the wrangling process.
- Plotly: this library in conjunction with ggplot2 is very useful to make interactive charts.
- Shinythemes: allows for customizability for the shiny app.
- Rcolorbrewer – for various color palette genrations.

Implementation steps:

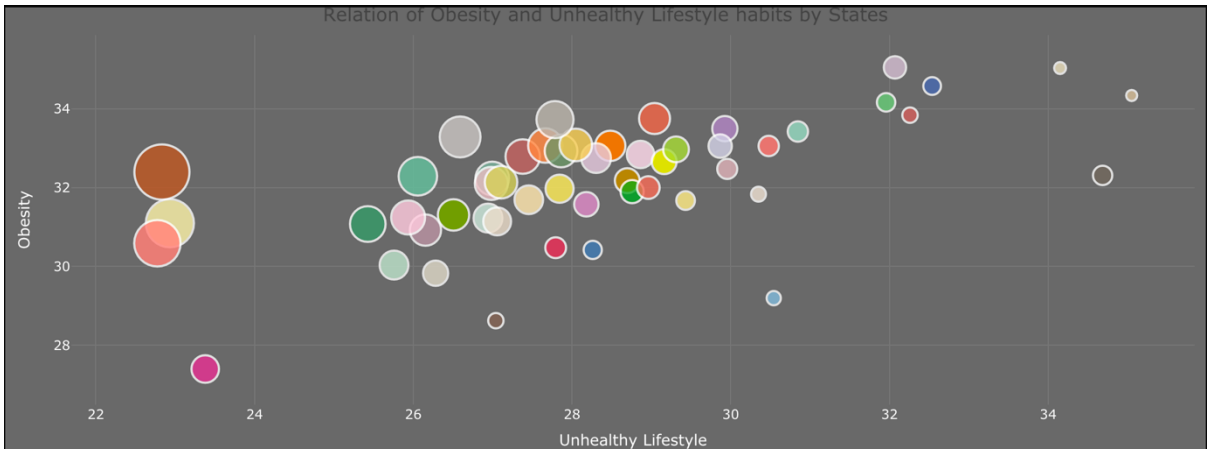
- Reading the wrangled dataset as well as US shape files into R.
- Align the dataset with the shapefile for plotting purposes.
- Create colour palettes for the visualisations.
- Various data frames for use in the shiny app extracted from the main data set.
- The shiny app includes a sidebar panel which has two inputs and one of the plots.
- The main panel contains the choropleth map, bubble chart and the bar charts for gender.
- Ggplot command enclosed in plotly used to make the ggplot graphs interactive inside the shiny app.
- For the leaflet map, I used separate polygons for obesity and physical activity to display them interactively.
- Finally, the bubble chart was created using plotly. On hover it displays the state names and percentage disparity between obesity and physical activity for the state.

User Guide:

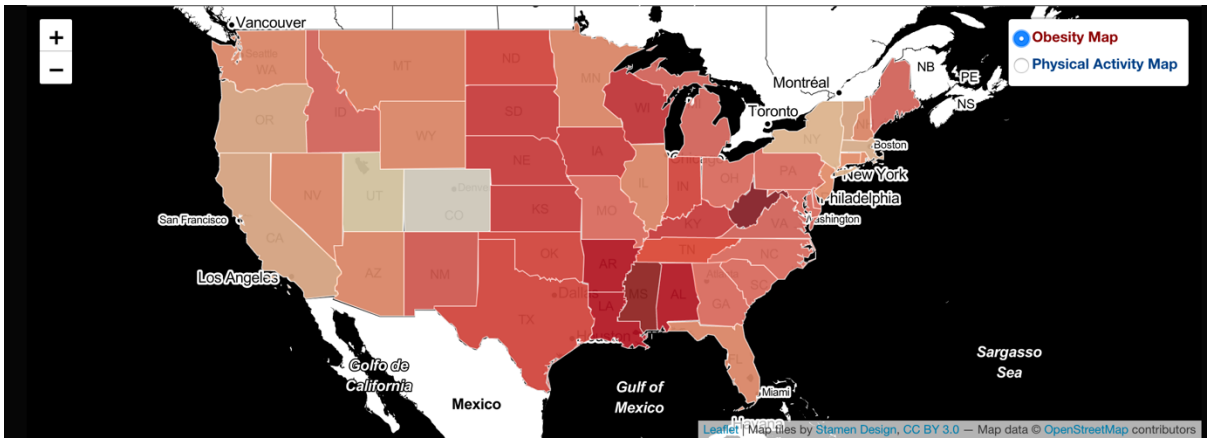
One of my goals for this project was to make the visualisation intuitive and interactive so that the general public is able to understand and navigate through the web app with ease.

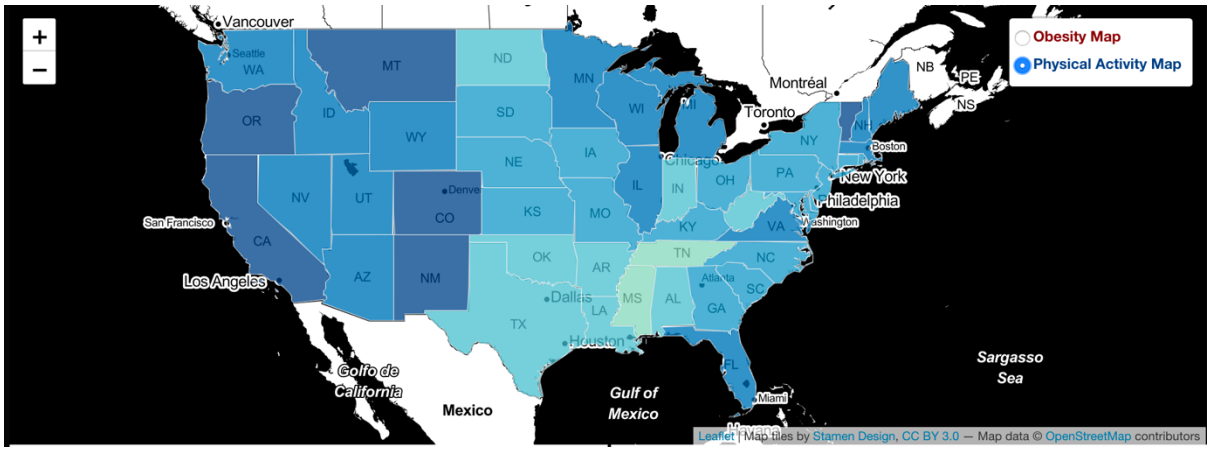


In this section, the user is able to select the socio-economic factor (race,age,gender,income,education) and either obesity or physical activity in the second drop down. The selection will alter the bar chart on display accordingly.

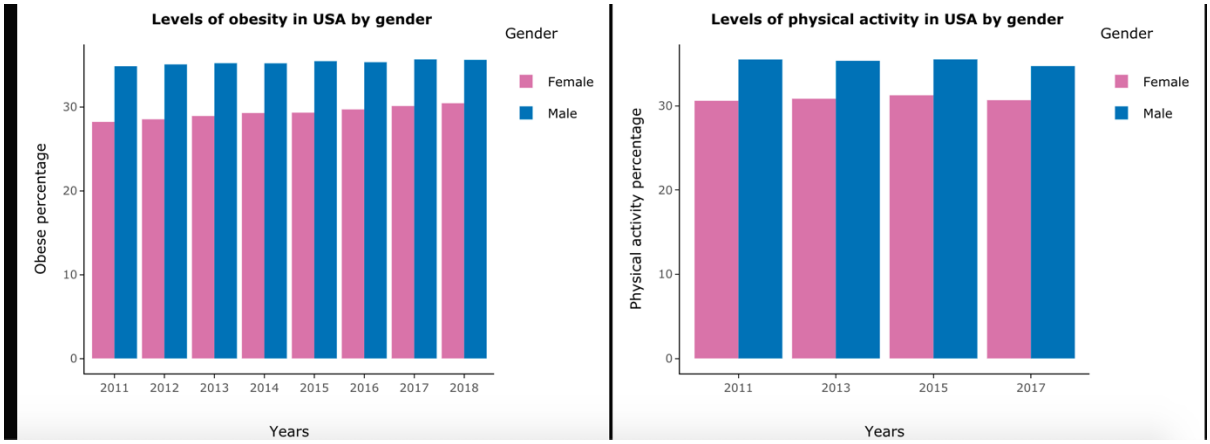


Next up, we have a bubble chart on the right side of the web page. Here users can see the relationship between an unhealthy lifestyle and obesity. Here unhealthy lifestyle constitutes of low levels of physical activity and not consuming healthy foods. Users can clearly see that, and unhealthy lifestyle percentages increases obesity levels also increase. The bubbles represent US states and bubble size is the disparity between percentages of obesity and unhealthy habits. Users can hover on top of the bubbles to view the state names as well as the variance in percentages for each state. It's also noticeable that as percentages increase the bubble sizes decrease which means obesity and unhealthy lifestyle has a stronger correlation at higher levels of each variable.





Then we have the choropleth maps for the obesity and physical activity levels. Through these maps I am trying to communicate the relation of obesity and physical activity to the audience. We can see from the maps that, the states with higher physical activity levels have lower obesity levels and vice versa. The users are able to toggle between the maps by selecting either options on the top right of the map. The maps also have interactivity built in; users are able to hover on the map to highlight the states and click on the states to get more information regarding the states (state name and percentage). Leaflet also provides zoom and drag, which the users can utilise to see some other states and territories in the remote areas (Alaska, Virgin Islands).



Last but not least we have the grouped bar charts for the gender. Here users can see clearly that male populations throughout the years have higher levels of obesity compared to females. But female obesity levels are rising which is an interesting find. The right bar chart highlights the reason behind this phenomenon. Females are less involved in physical activity then males, which might be a good indicator of their growth in obesity levels. users are able to hover on top of the bars to see the year and percentage information as well.

Conclusion:

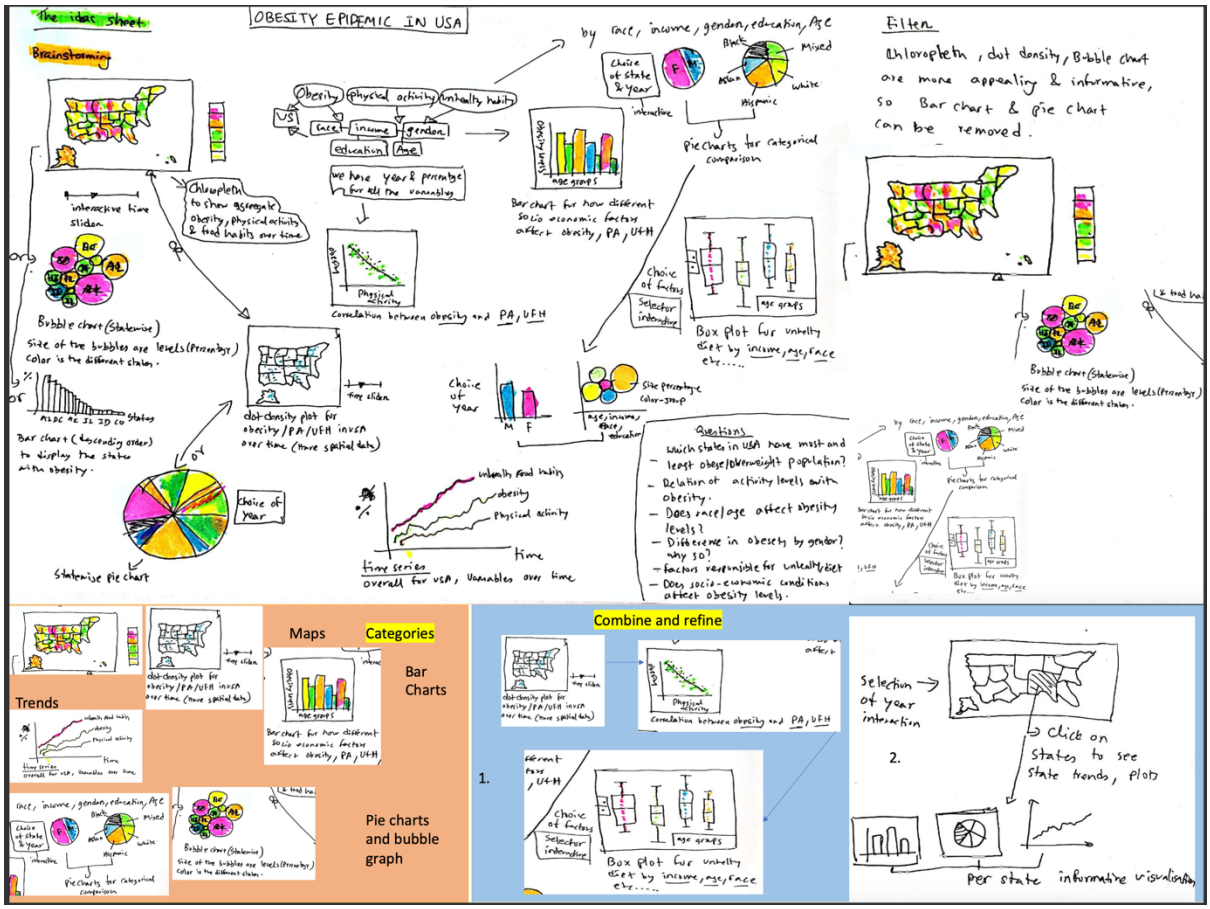
By analysing the obesity data for USA, I was able to find trends and insights from the data such as females are increasingly becoming obese compared to males, West Virginia and Mississippi are the most obese states in USA, Colorado as state is more into physical activity, income and education are good indicators of unhealthy diet. I also learned that obesity and unhealthy food habits are correlated, race and age are good indicators of obesity whereas income and education is not.

This project has been a great learning process for me. I have learned how to wrangle data and analyse it to create visualisations. Through the use of shiny in R, I am able to provide interactive visualisations for the audience and portray the current obesity situations in USA. In hindsight if I had mastery in the use of D3 to create the visualisations, I think I would be able to create even more vivid, intuitive and interactive visualisations for this project.

Bibliography:

Dataset used:
https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system

Appendix:


The idea sheet
OBESITY EPIDEMIC IN USA
Brainstorming
Obesity
physical activity
unhealthy habit
US
race
income
gender
education
Age
by race, income, gender, education, Age
we have year & percntge for all the variables
Chloropleth to show aggregate obesity, physical activity & food habits over time
interactive time slider
Bubble chart (Statewise)
Size of the bubbles are levels (Percentage)
Color is the different states.
Bar chart (descending order) to display the states with obesity.
dot-density plot for obesity/PA/UFH over time (more spatial data)
Statewise Pie chart
Choice of Year
Time series
overall for USA, Variables over time
Bar chart for how different socio economic factors affect obesity, PA, UFH
Pie charts for categorical comparison
Box plot for unhelthy diet by income, age, race etc.....
Correlation between obesity and PA, UFH
Questions
Which states in USA have most and least obese/Overweight population?
Relation of activity levels with obesity.
Does race/age affect obesity levels?
Difference in obesety by gender? why so?
factors responsible for unhealthy diet
Does socio-economic conditions affect obesity levels.
Filter
Chloropleth, dot density, Bubble chart are more appealing & informative, so Bar chart & pie chart can be removed.
Trends
Maps
Categories
Bar Charts
Pie charts and bubble graph
Combine and refine
1.
2.
Selection of year interaction
Click on States to see State trends, plots
per state informative visualisation

First Design

Information

Title : FDS sheet 2 (1st Design)
Author : MD. Saadman Hossain
Data : CDC obesity data
Date : 03:06:2020

Operations:

→ Layout is a chloropleth map displaying intensity/avg percentage.
→ Drop down selection of obesity, physical activity & Unhealthy food habits percentages.
→ Interactive time slide to show how percentages in each state changes over time.
→ click on states to display individual state information. (Focus)
→ We have a trend line appearing which shows particular States percentage levels over time
→ selection for ~~factor~~ socio-economic factor which shows Bar graphs per state. (age, race, gender, education, income)

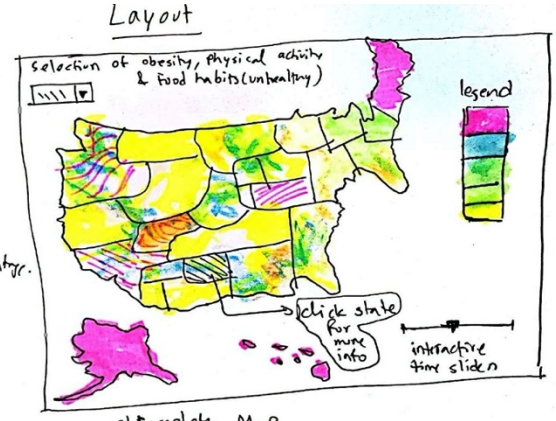
Pros

~~- relatively easy implementation~~
+ can see intensity of obesity levels (darker states have more obesity).
+ can see how activity levels & unhealthy food habits affect obesity

Cons

- dont answer all questions
- Hardon implementation
- dont show correllations of different factors with obesity, PA & UFH.

Layout



chloropleth Map

Focus

→ drop down to select the variable for which to see the bard graph

Second design

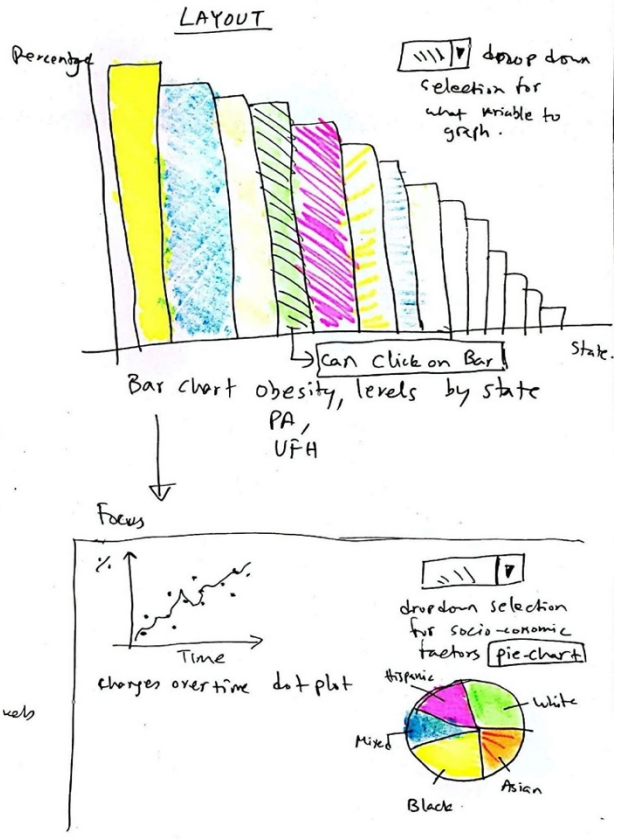
Information

Title: FDS sheet 3 (2nd Design)
Author: MD. Saadman Hossain
Data: CDC Obesity Data
Date: 07.06.2020

Operations:

→ Bar chart for obesity; PA, UFH from most to least. Interactive seletion of what variable to graph.
→ can click on Bar to display more info in the focus.
→ default shows a dot plot, percentages over time.
→ drop down selection for socio economic factors. i.e if obesity is chosen & race is chosen, it will show pie-chart of obesity by race.

Pros	Cons
– easier implementation	– doesn't show correlation
– can see clearly the top states.	– doesn't answer why levels vary by gender.
– pie-chart breakdown of a socio economic factor gives good idea.	



Third design

Information:

Title: FDS sheet 4 (3rd Design)
Author: MD. Sandman Hossain
Data: CDC obesity data
Date: 07.06.2020

operation

→ Select obesity, Unhealthy food habits, Physical activity to display Bubble chart

→ Drop down to select State, default produces scatter plot (drop down to select X and Y Variables) which shows correlation of var X against var Y. So ~~for~~ ie. unhealthy food habits vs obesity.

→ Drop down to choose socio-economic factor, produces breakdown for each state. i.e. distributions of different age groups for a particular state.

Pros

+ implementation may be easier
+ Boxplots give us a good idea how each factor is broken down

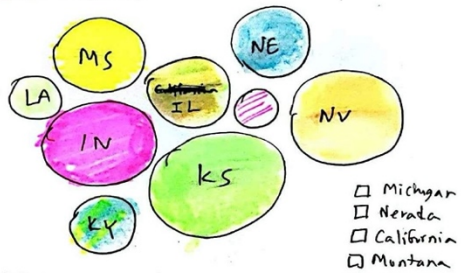
Cons

- doesn't answer all questions.

Layout

[▼] drop down to select variable.

Bubble chart

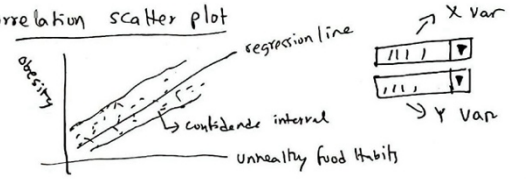


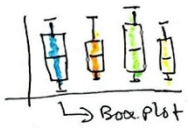
→ Bubble size is percentage
→ Bubble color is state

Focus

→ ~~click~~ [▼] drop down to select state.

Correlation scatter plot





= [▼] → Choose socio economic factor Breakdown each state.

↳ Box plot

Realisation

Information:

Title: PDS sheet 5 (final design)
Author: MD. Saadman Hossain
Data: CDC obesity data
Date: 07.06.2020

Description:

→ chloropleth map with an interactive time slider to display change over the years. selection of variable.
→ clickable states to toggle further visualisations.
→ drop down to select socio economic factor to create pie-chart.
→ trend line will show change of the variable over time.
→ correlation scatter plot with interactive variable selection. ie. Food Habit vs obesity.

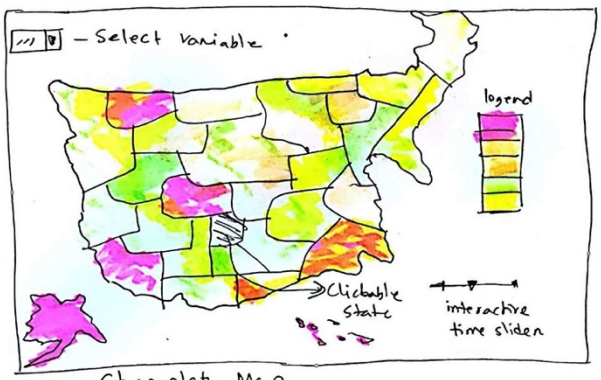
Software Requirements:

R, R packages, shiny for interactive viz, maybe D3.

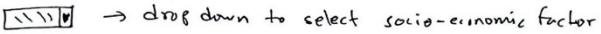
Time requirement: 1 week

Pros	Cons
- Answer almost all questions - Finished product should be decent.	- bit harder to implement

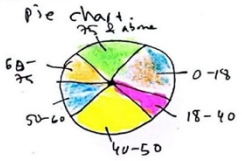
Layout

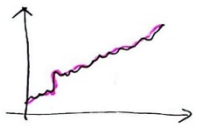


Chonopleth Map



pie chart





Trend line for variable over time for a state

Correlation scatterplot