Narrative Visualisation Project

Changes in smoking and drinking rates in counties in each U.S. state and the number of related cancer cases from 2002 to 2012.

Student Name: PEIYU LIU

Student ID: 31153291

Tutor: Angel Das and Mohit Gupta

Lab: Tutorial 3, Thursdays at 4 pm6 pm

Table of contents

1.Introduction

Alcohol is one of the most popular drinks in the world nowadays. Alcohol is everywhere in people's daily activities or social activities. Whether teenagers or adults, they all drink alcohol. Even some people are addicted to alcohol, especially the population of males who are addicted to alcohol are much higher than females. The population of females who are drinking alcohol is increasing dramatically in most states. Significantly, the female drinking rate in Kentucky has increased by 73% from 2002 to 2012.

Although in most states, the drinking rate of males is represented as a slight decrease, males are still the leading group who drink alcohol. Each year the alcohol consumption of the male group is much higher than the female group. Each year, the male will consume more than 4000-millilitre alcohol, but women use no more than 1450 millilitres.

Many people regard smoking as one way to release pressure. However, Cigarettes weaken the body's immune system, and it is not healthy for our bodies. Smoking rates in most states in the USA are decreasing slightly. For Both females and males, the population of smoking is decreasing.

The numbers of breast, oral, lung, liver, and thyroid cancer cases are increasing obviously. Significantly thyroid and lung cancer cases are increasing dramatically. Female breast cancer also increases. One of the reasons is that more women like drinking alcohol.

Cigarettes and alcohol are carcinogens. The higher the intake of carcinogens, the higher the risk of cancer.

This narrative project is about the changes in smoking and drinking rate in each US state and the impacts of relevant cancer cases. Additionally, this project will display five cancer case data: breast, oral, lung, liver, and thyroid cancer.

The target audience of this narrative project is the public who want to know the data changes. However, the dataset is complex and has various location variables, years, and gender and cancer type categories.

Therefore, the visualization should be direct and easy to understand to convey all information to the general public. The visualization should not require viewers to have relevant academic backgrounds. Add interaction and animation to improve user experience. Moreover, add an introduction and guide label to help the user know how to act with visualization.

2.Design

Considering the target audience is the general public, my diagram should be straightforward and interesting. Before writing code to realize my visualization project, I used five design-sheet methods to collect my ideas and choose methods to design a narrative project.

Sheet 1

At the beginning of the five design-sheets, in sheet 1 design, it is a brainstorming stage, so I put all my ideas in sheet 1. It contains all the diagrams that I plan to use. I filter heat map, treemap, box whiskey map and stack bar map because they are not suitable for my data. I plan to use the choropleth map as media to display all states in the USA. It can help viewers understand this is a project about the USA location. I plan to use two-dimensional diagrams like line charts or bar charts. These two-dimensional charts are helpful to display the whole alcohol consumption trend. Moreover, I want to use a scatter plot and pie chart to display how cancer cases are distributed each year. In sheet one, I combined and refined the bar chart, line chart, scatter chat, and pie chart to display five different cancer case changes.

Sheet 2

In sheet 1, I plan to use a choropleth map as the diagram to draw the smoking rate in each state in the USA. So in sheet 2, I use the dataset of smoking rate to draw a USA choropleth map and add a tooltip method to display the delicate value. When viewers interact with this diagram, mouse events are used to realize interaction. The choropleth map in sheet 2 will focus on each part of the map should be a state, and the value of this state can be popup to tooltip to show the result to viewers. The value will be aggregated in the system to show changes in 10 years(from 2002 to 2012)to the viewer. The advantage of using a choropleth map is that the map can help the audience know this is a visualization about the USA directly. Furthermore, using a tooltip to show results can increase user experience. The disadvantage is if the audiences are not familiar with the USA map when they have to check each place when they want to find one particular place.

Sheet 3

In sheet 3, bar charts combine line charts to show how male group and female groups drink alcohol in the USA from 2002 to 2012. add one selection like two buttons. One button shows a bar chart about the male group, and another one shows the female group. It can avoid drawing two duplicate charts. This bar chart should focus on the years from 2002 to 2012 and the population rate. Bar charts can also display changing trends. For each year, the rate is increasing or decreasing. Line charts are used to show the changing trend in alcohol consumption. A line chart can display the whole trend directly. Moreover, add a tooltip to display the actual value when the audience mouse on this line.

Sheet 4

In sheet 4, it combines the scatter diagram, bar diagram, line diagram and pie diagram that I mentioned in sheet 1. Use combination diagrams to represent how female breast cancer cases, oral cancer cases, lung cancer cases, liver cancer cases and thyroid cancer cases changed from 2002 to 2012 in the USA. All diagrams should focus on year distribution and the total population of cancer cases. Add popup interaction to help audiences see detailed information like real population value when using a mouse to interact with a diagram. I do not plan to add value on bars' top of dots. Too many numbers will make the screen not clear.

Sheet 5

In sheet 5, I consider all ideas I design in sheet 2,3,4. Sheet 5 is a similar layout to my actual narrative project. In the beginning, audiences can see a USA map with a popup function to display the smoking value of each state. Then use the same map method to show drinking rate changes in each state. After two main maps, use a bar chart with selection to show any level of drinking rate in the USA each year from 2002 to 2012. selection is divided data by different genders. Then use a line chart to show the binge level of drinking rates in the USA each year from 2002 to 2012. two lines mean different gender groups. One is the male group, and the other is the female group. And then use two line charts to show alcohol consumption in the USA each year from 2002 to 2012. When audiences put a mouse on this line, it can show the exact value of consumption and a vertical line to help audiences confirm which year. Finally, draw five two-dimensional charts to represent five different cancer cases changes.

3.Implementation

The main technology to reformat and data wrangling is python. I used libraries like Pandas, NumPy and Spark Dataframe to process my dataset. Then use D3 version 4 to produce a narrative visualization project. I import some code templates like choropleth map and bar chart with selection. The main diagrams are choropleth maps with tooltips to show smoking and drinking rate in each US state. Bar charts and line chart combined to show the whole changing trend of alcohol consumption. Scatter diagram, bar diagram, line diagram and pie diagram to show cancer cases changes.

Map

Reference(use to draw map):
map: https://www.d3-graph-gallery.com/graph/choropleth_hover_effect.html
tooltips: https://stackoverflow.com/questions/20070919/adding-tooltip-in-d3-js-map

Choropleth maps are for spatial data, and audiences can directly know this is a narrative diagram about smoking and drinking data in the USA. Add interaction methods like mouse events and popup text descriptions when audiences use a mouse to explore this visualization.

In the beginning, convert data to Geojson format and load data to D3 function.

Label: to guide audiences to use the mouse to explore states on the map.

Title: inform audiences this map diagram is about smoking or drinking rate changes in the USA states(exclude Alaska and Hawaii).

Colour: set colours on maps using state id. The smoking map is green like a cigarette, and the drinking map is blue like water.

Area: load Geojson data and then use coordinates of each state to draw their shape on the map. Set mouse events styles, when audiences put the mouse on each state, this part of the map will highlight. Then the rest of the states' parts are lighter in colour.

Tooltip: the system will return a description of the analysis result to audiences; instead of displaying all information on the map directly, it can increase the interest in maps. It can avoid too much text on maps.

Animation: add different styles to different mouse events, like highlighting and lighter colours. Help audiences to identify their actions like exploring each state.

Show description of data after audiences put the mouse on maps.

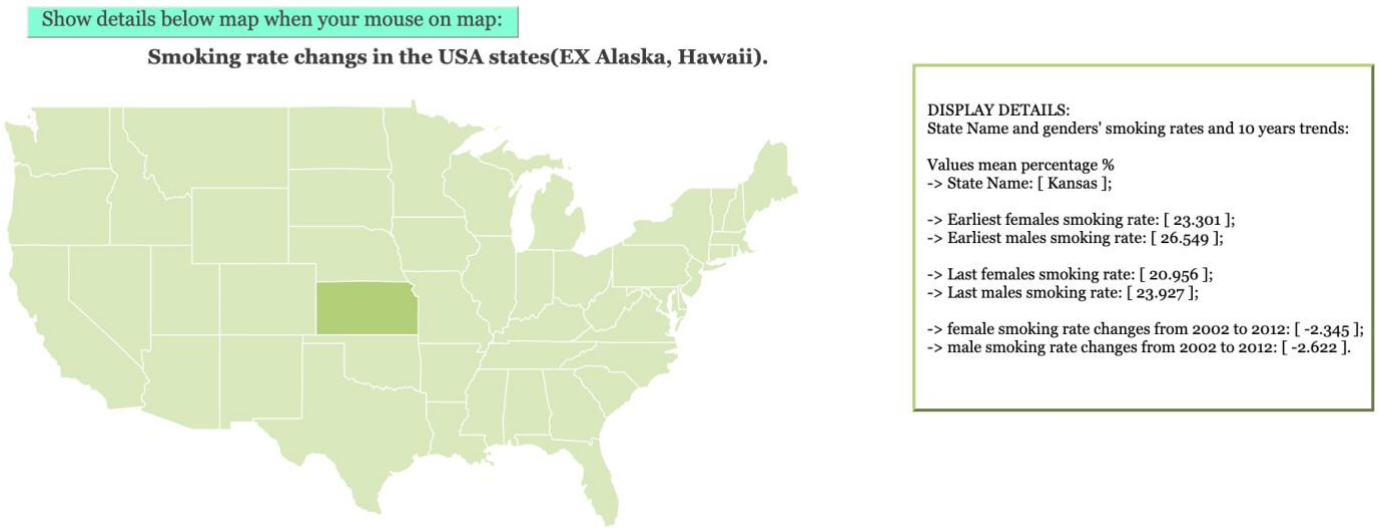


Figure 1: Screenshot of map

Line chart

Reference(use to draw line): https://bl.ocks.org/Qizly/8f6ba236b79d9bb03a80

line charts help me to display time-series data and represent how value changes from 2002 to 2012. a line chart can show a direct whole trend to audiences. Charts are suitable for the data of alcohol consumption each year and cancer cases in each year.

Line: show the whole trend over a ten years period.

Colour: for multiple line charts, pink is about the female group, and blue is about the male group.

Mouse events: when audiences put the mouse on particular lines, the system will popup the exact value to help audiences know the real value in my data.

Title: a guide in which data is displayed here.

Label: guide audience to interact with bar diagram.

Axis: x is time series, y is population data.

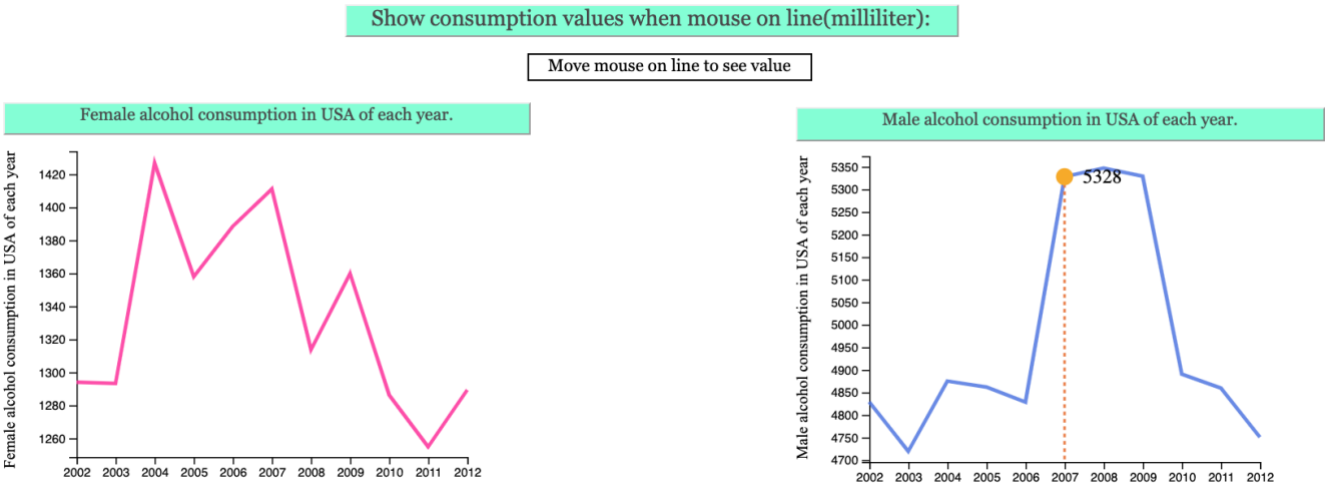


Figure 2: Screenshot of line chart

Scatter chart

Reference(use to draw scatter plot):https://www.d3-graph-gallery.com/graph/scatter_tooltip.html

scatter charts can represent data distribution, and they can combine with a line chart to show data trends.

I use a scatter chart combined with a line chart to draw cancer cases changes in a 10 year period. It can show the changing trend.

Point: match each year and the population of each year.

Colour: set colour makes the diagram more interesting.

Mouse events: the system will popup data descriptions when audiences interact with the diagram.

Amination: the system will change the selected scatter colour that is selected by audiences and show a tooltip to describe this scatter.

Title: a guide in which data is displayed here.

Label: guide audience to interact with bar diagram.

Axis: x is time series, y is population data.

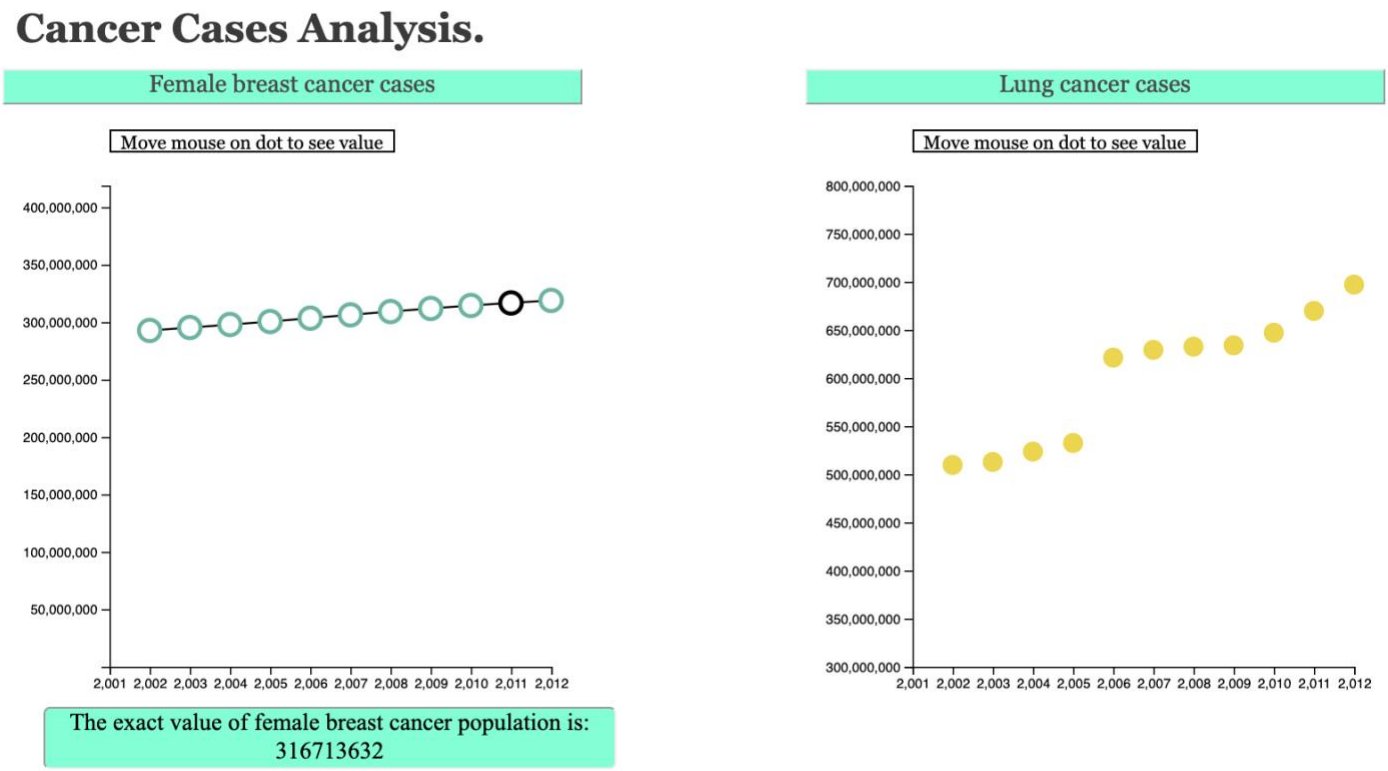


Figure 3 : Screenshot of scatter plot and combine with line

Pie chart

Reference(use to draw pie):http://bl.ocks.org/dbuezas/9306799

The pie chart can represent the data distribution of the total data.

Use a pie chart to draw liver cancer cases data.

Pie: the size of each part on the pie chart can display different population values. The value is big, and the pie part is big. It can also determine whether the data is growing by the length of each part of the pie chart.

Title: guide audience this diagram is about which data.

Label: display each year and exact value.

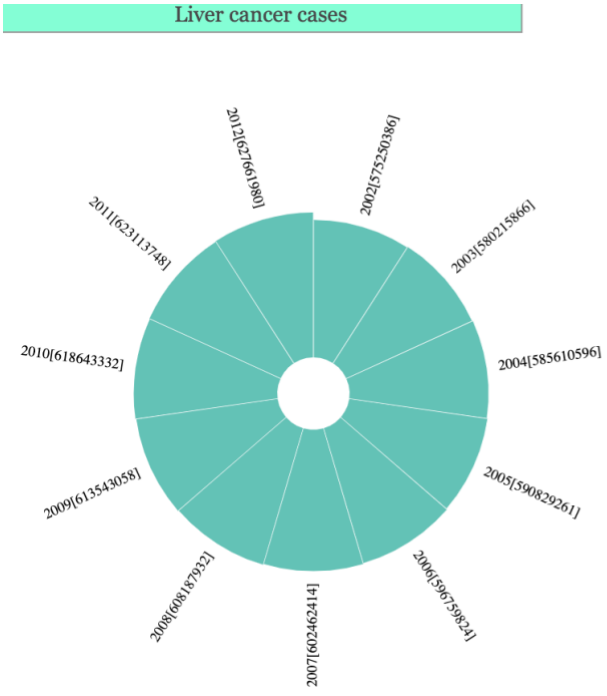


Figure 4: Screenshot of pie chart

Bar chart

Reference(use to draw bar chart): https://bl.ocks.org/alandunning/274bf248fd0f362d64674920e85c1eb7

Bar charts can represent each year's values, and audiences can also see changes from each bar on this bar chart. Add a filter button to show different data on the bar chart.

Bar: the height of bars means different values in each group.

Colour: set colour to the diagram makes the diagram more attractive. Some bars also set colours depending on the value range(cancer cases range).

Mouse events: system will represent one tooltip that contains a basic description and exact value of this bar on this diagram.

Button: change variable in my data and change bar diagram to represent the different groups(female and male groups).

Amination: when audiences click the button, then the diagram will change with a duration animation.

Title: a guide in which data is displayed here.

Label: guide audience to interact with bar diagram.

Axis: x is time series, y is population data.

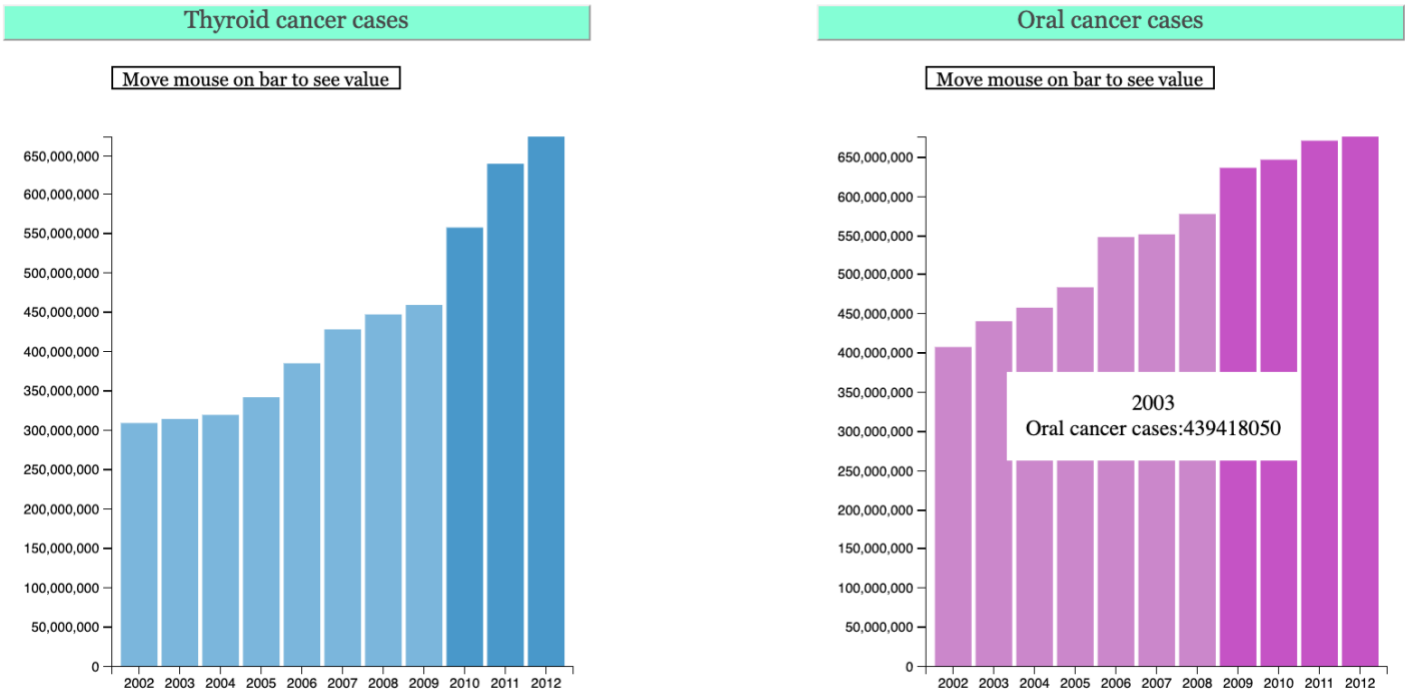


Figure 5: Screenshot of bar chart

Reflective thinking about challenging parts

Data wrangling: At the beginning, I selected many datasets and multiple types like XLSX, CSV, pdf, JSON. I use python to process these data and filter unneeded data, then output data that I need to realize data visualization using D3. Dealing with this raw data has taken many days. Especially when I convert data to Geojson, I use much time to understand the structure of this type of data format.

D3 CSS: When I use D3 to draw graphics, I set the CSS style of position for each graphic as relative. But I found that when I zoomed out of my page, the position of the graphics might change, and I did not know how to solve the problem. So finally, I set the position to absolute, fixed the position of each graphic and text. The disadvantage is that when the computer size of the audience is larger than the screen size of my computer, there will be a blank part on the right.

D3: When I use the d3 template-choropleth map to draw my smoking and drinking maps. I want to set different colours for each state, and the colour is according to the value in this place. So I change the result value by female smoking/drinking rate in this place. But the colour has not changed. I printed the console log to check return data, everything is going well, but the colour still has not changed.

4.User guide

Audiences who open this narrative project can see the project title and short conclusion about background and questions.

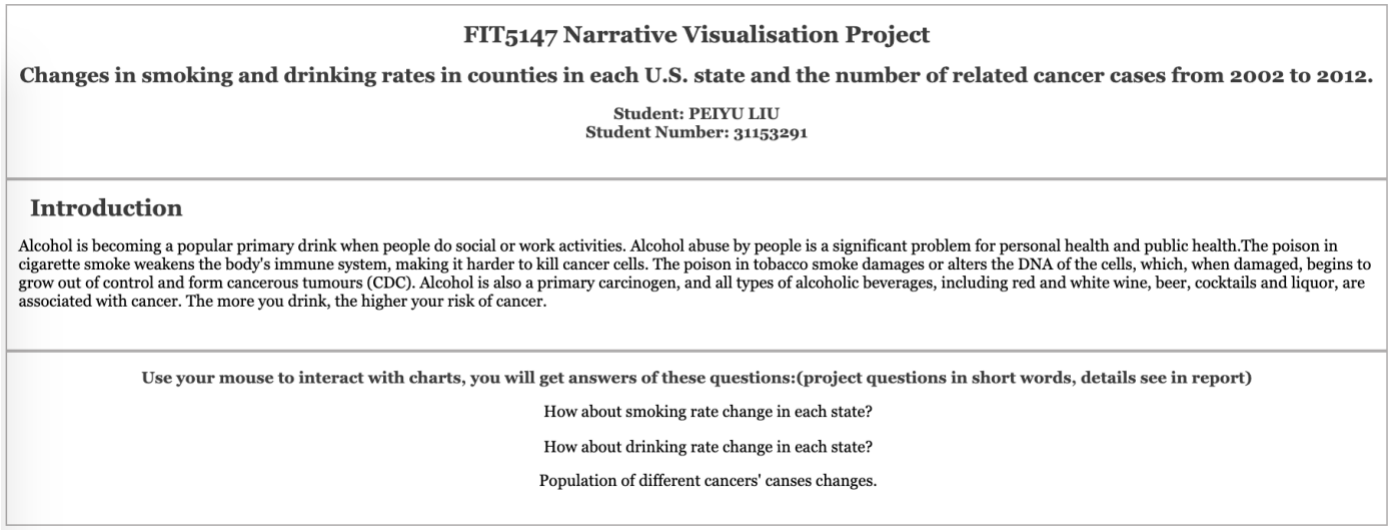


Figure 6: welcome banner

The first diagram is a choropleth map about the smoking rate in the US states. Add label and title to guide audiences, put the mouse on map and details will show in an aside box. The output information is about female and male smoking rate changes from 2002 to 2012. Behind each diagram will be a description box to represent the analysis result and target.

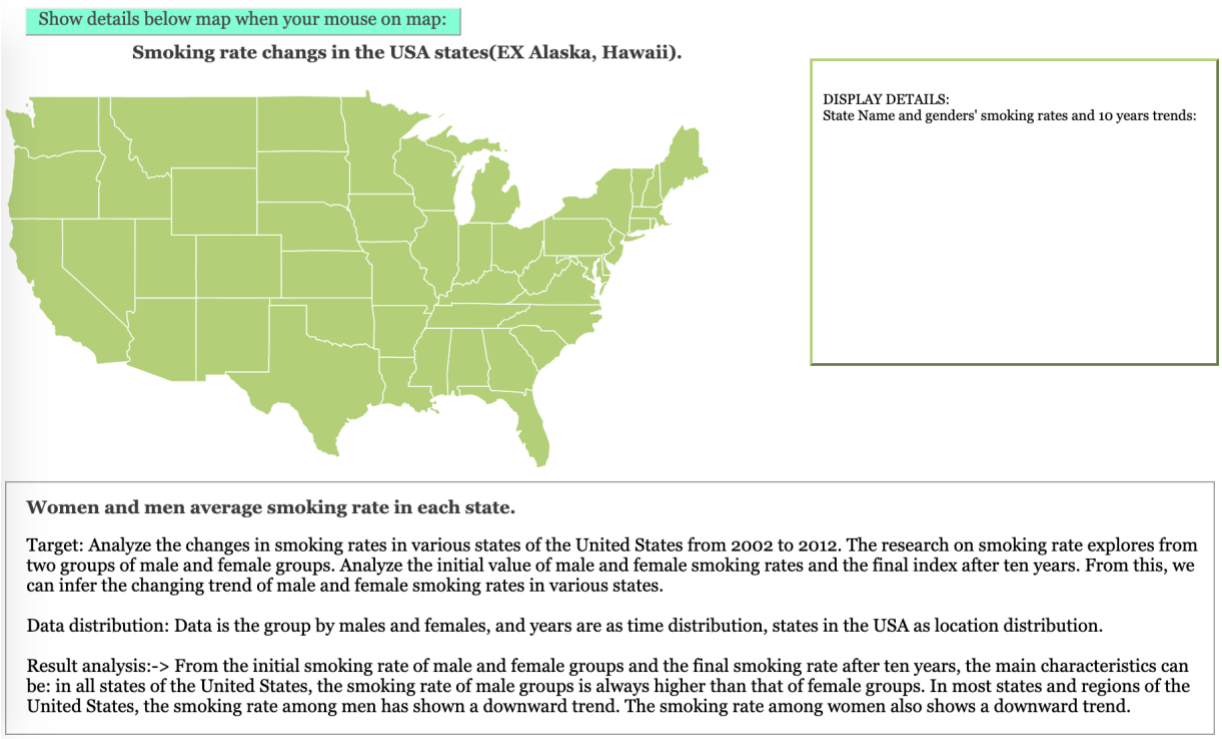


Figure 7: smoking rate in the USA

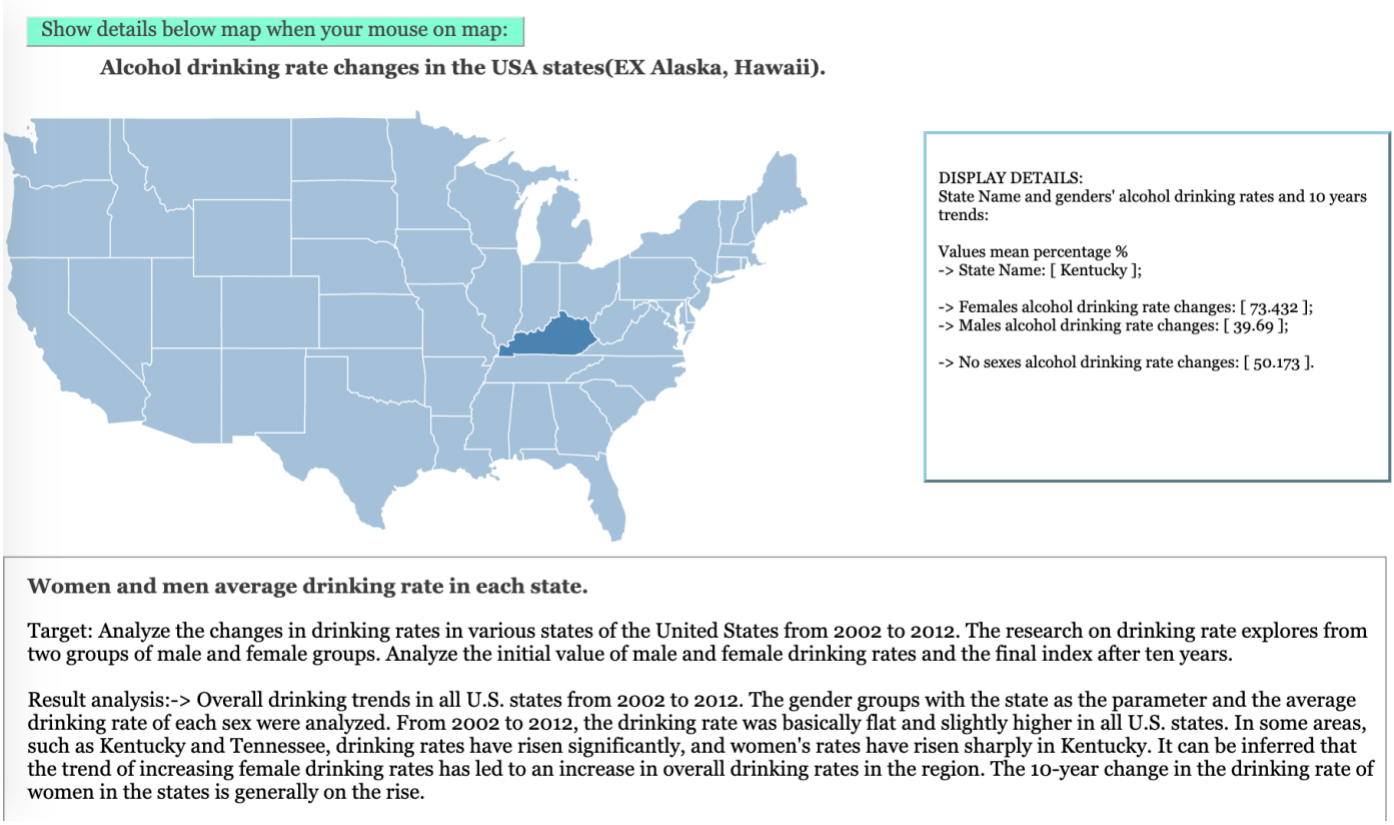


Figure 8: mouse event on map

The second diagram is a choropleth map about the drinking rate. It has the same structure and function as the smoking map. So all actions are the same. Behind is the result description of drinking rate analysis.

The next part is one bar chart representing the drinking rate (any level: people drink more than one cup) of the female and male groups. Follow the guide in a blue banner, and click two buttons to change the related group's bar chart.

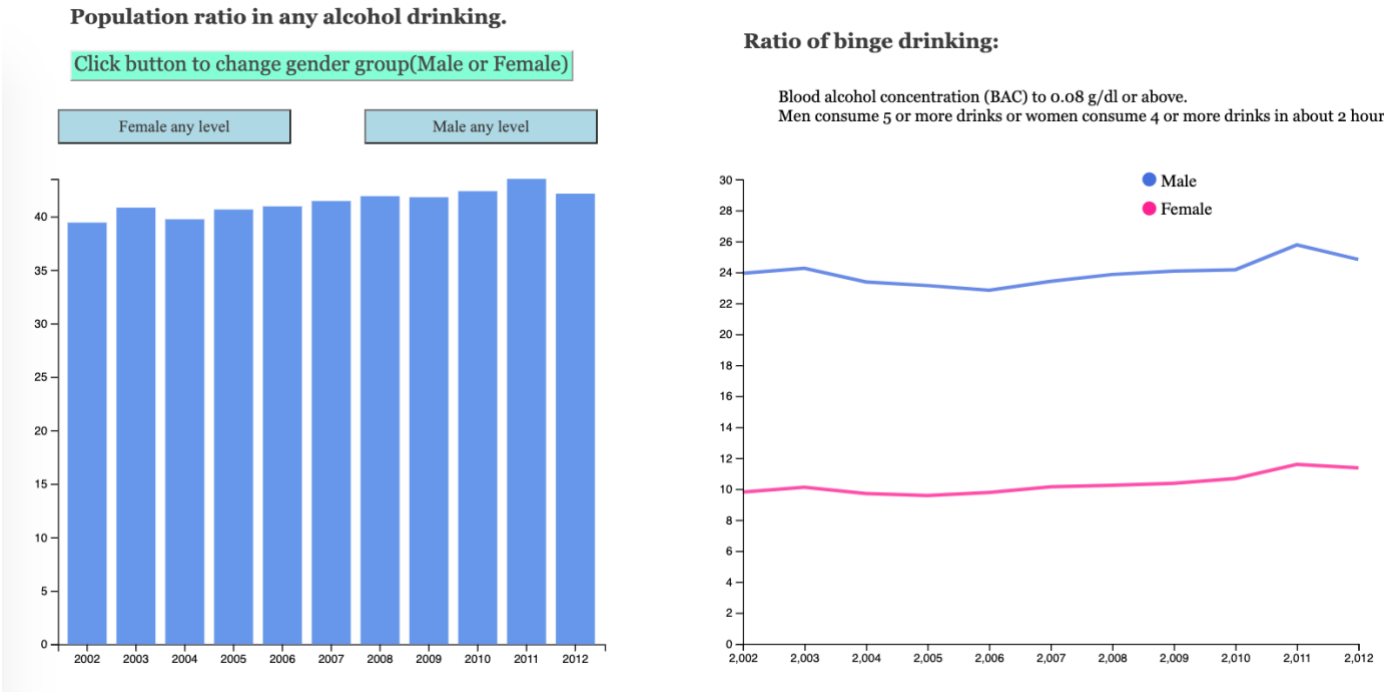


Figure 9: no popup function(only show trend)

The line chart on the ride size compares two genders who drink alcohol at binge level(men consume 5 cups or more, women consume four or more in 2 hours). These two charts are with no popup function, showing each group's trend. from the height of the bar and line, audiences can know the male is the leading group drinking alcohol.

The next part is a scatter chart and a cartoon to make the website more colourful and interesting. Audiences can click the button to show the average alcohol consumption of females and males in the USA. From the value in the Y-axis, audiences can know males consume much more alcohol each year than females.

The behind two line charts support average alcohol consumption data. Audiences can put a mouse on each line, and the system will popup exact values to help audiences know the actual value. Then in the text box, it shows the result about alcohol consumption.

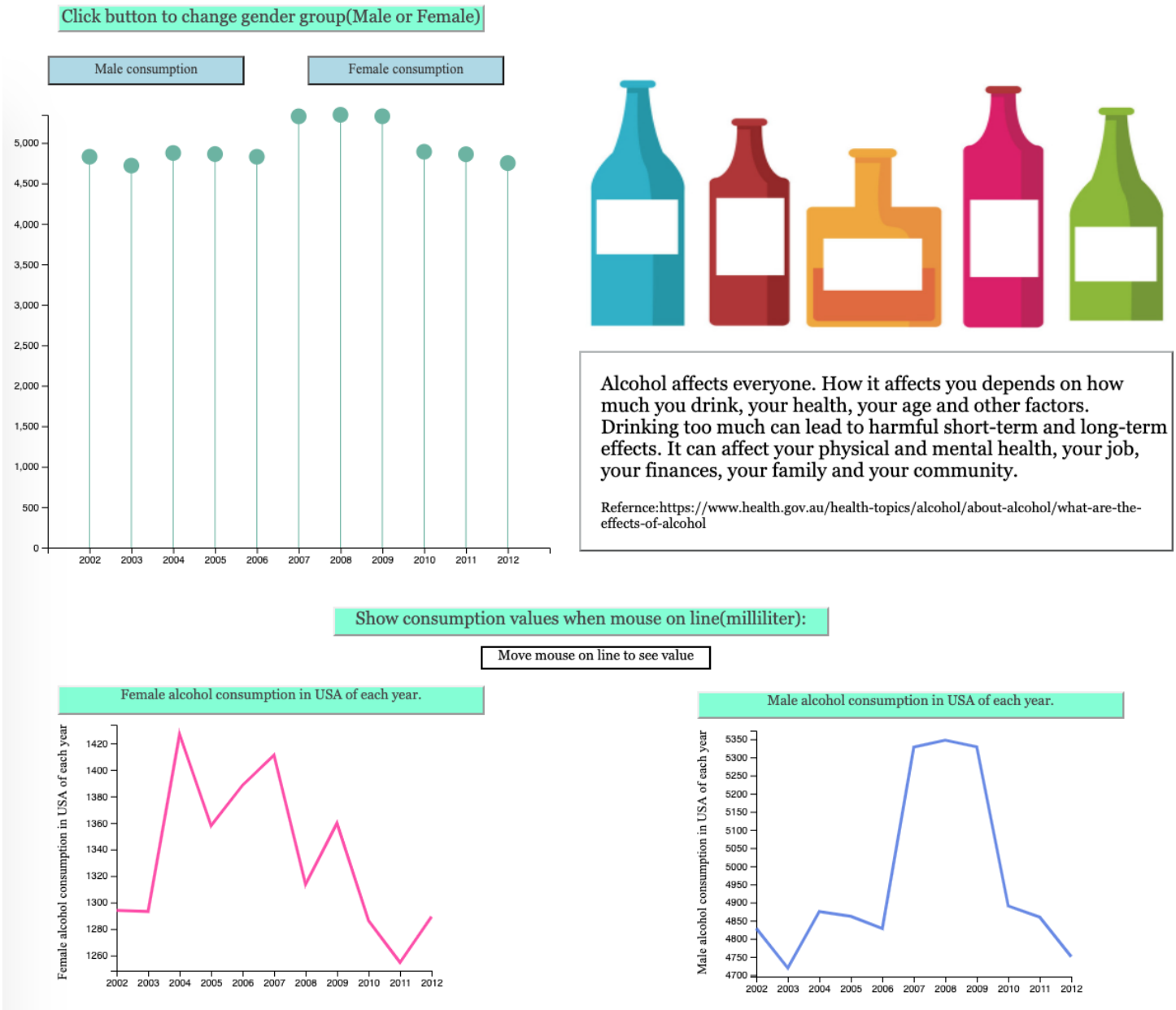


Figure 10: interaction on line chart

The final part of the project is an analysis of changes in five cancer cases. It has a title to introduce each cancer and a label to guide audiences to interact with diagrams to know more information. At the end of the right side is the analysis result of the changing trend of cancer cases.

Figure 11: audiences can interact with diagrams

5.Conclusion

Finding results

Smoking rates declined in most parts of the United States, with a slight decline in both male and female groups, essentially flat. Most states have higher smoking rates than women, and men are the leading group of smokers.

Women's drinking rates rose in most U.S. states, particularly Tennessee, up 46.04 per cent and Kentucky, up 73 per cent. The rate of male drinking remained virtually unchanged in most regions, with an increase and decrease rate of no more than 10 per cent. In terms of the overall population, most U.S. states have an upward trend in alcohol consumption rates, mainly due to a sharp increase in female drinking rates. Despite the significant increase in female drinking rates, men have always been the leading drinking group, with more than 50 per cent of men drinking alcohol compared to about 40 percent of women. The proportion of men who drink alcohol at binge level is much higher than that of women, at about 25 per cent.

In comparison, the proportion of women remains at 10 per cent. The leading group of alcohol consumers in the United States each year is men, who average more than 4719 millilitres per year. Women's annual average alcohol is no more than 1500 millilitres.

The number of breast, throat, oral, lung and liver cancers in women is rising each year. In particular, lung cancer, oral cancer and throat cancer showed a significant upward trend. This is significantly related to the increase in alcohol consumption and the absence of decreased smoking rates in the states, where alcohol and tobacco are primary carcinogens.

Reflective improvement

Narrative visualization brings complex data to life. Whether or not audiences can analyze data, they can see trends in data changes from the graphs of data visualization. Data visualization is beneficial in the era of big data. It can be difficult and complex data through a simple graphical display. People will be attracted by graphics and are more willing to understand the data. In this project, I learned if I looked for the correct raw data and then cleaned and formatted the data by processing. I understand that graphical visualization is not suitable for all data. I need to choose the correct visualization based on the data. By completing this project, I learned how

to get CSS, Html, javascript together to do what I want to realize. Also, have a specific understanding of the layout of the previous page.

There are some areas for improvement in the project. Perhaps with the specific value of each bar at the top of the bar charts. Graphic presentations can add more interesting dynamic animations. Instead of using absolute's mode to set up the page, the page can be set with relative length and width.

6.Bibliograph(APA7)

Alan, D.(2019, Oct 19). *Bar Chart with Tooltip with Tooltip D3 V4.*Observable. https://bl.ocks.org/alandunning/274bf248fd0f362d64674920e85c1eb7

David, B.(2021, June 4) *Pie charts labels.* Observable. http://bl.ocks.org/dbuezas/9306799

Yong, L.(2019, Nov 4). *Line Chart with Tooltip 1*. Observable. https://bl.ocks.org/Qizly/8f6ba236b79d9bb03a80

The D3 Graph Gallery. *Choropleth map with hover effect in d3.js.* https://www.d3-graph-gallery.com/graph/choropleth_hover_effect.html

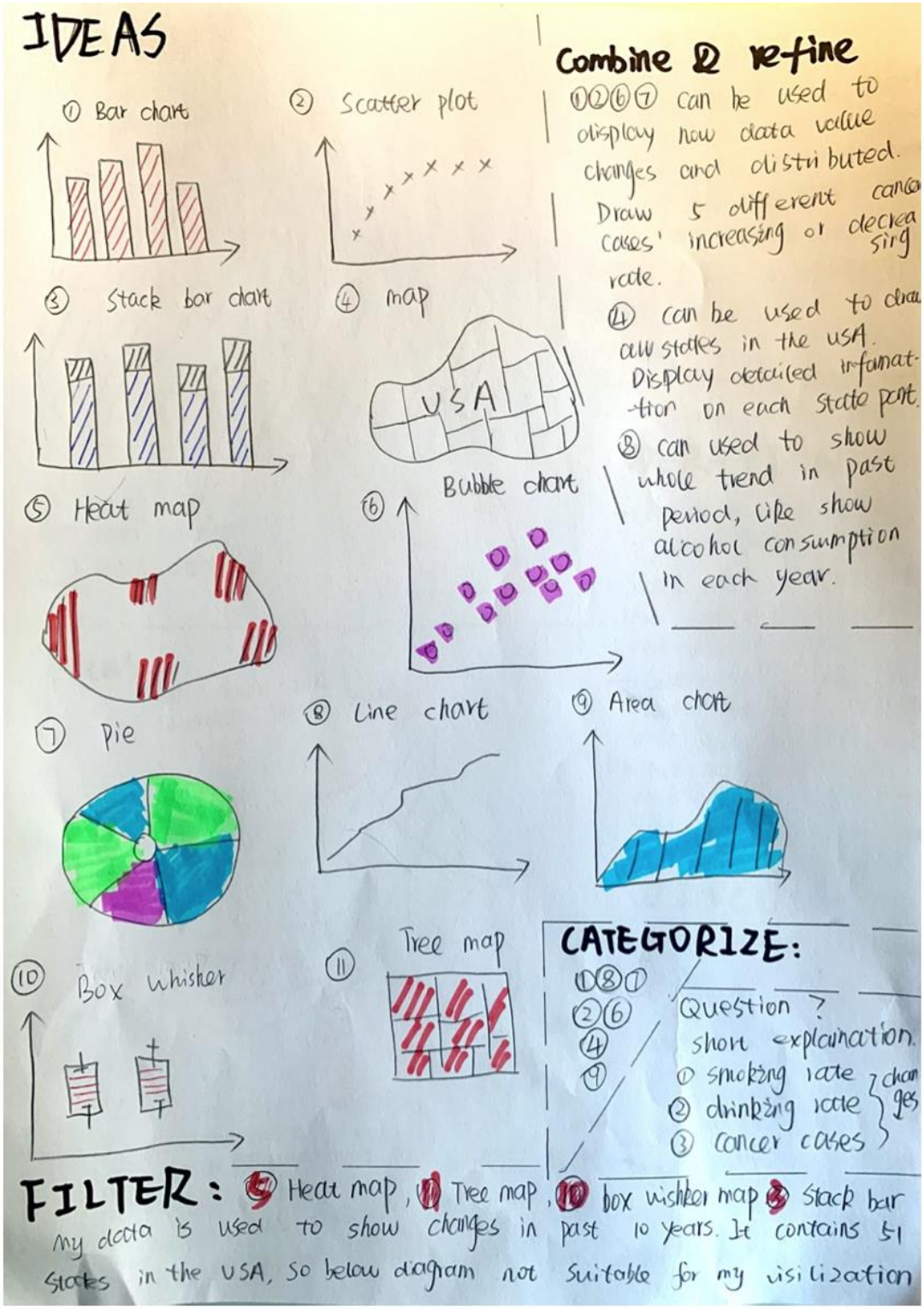
The D3 Graph Gallery.*Button to change input data in barplot.*https://www.d3-graph-gallery.com/graph/barplot_button_data_simple.html

The D3 Graph Gallery. *Lollipop chart*.https://www.d3-graph-gallery.com/lollipop.html

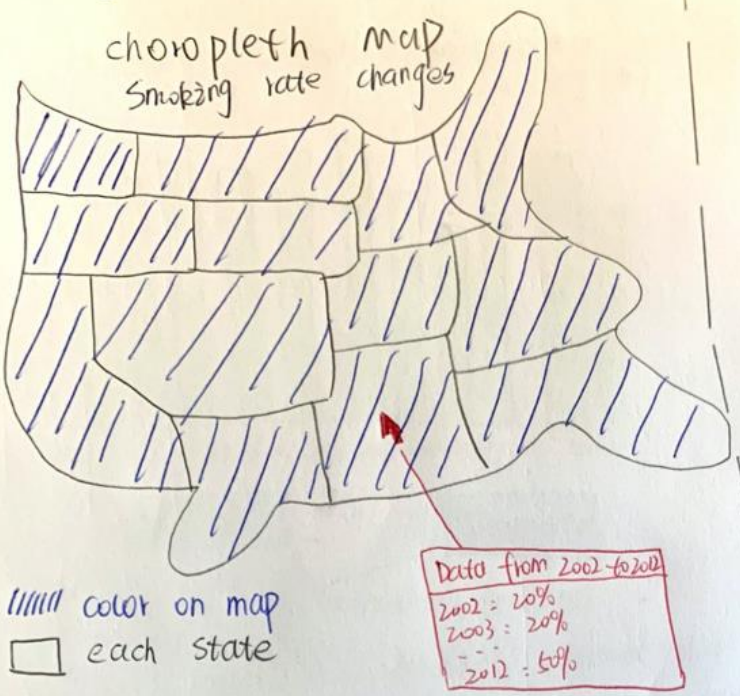
The D3 Graph Gallery. *Scatterplot with tooltip in d3.js.* https://www.d3-graph-gallery.com/graph/scatter_tooltip.html

git: https://github.com/SukiGit2021/FIT5147_dataset.git

7.Appendix


IDEAS
① Bar chart
② Scatter plot
③ Stack bar chart
④ map
USA
⑤ Heat map
Bubble chart
⑥
⑦ Pie
⑧ Line chart
⑨ Area chart
⑩ Box whisker
⑪ Tree map
Combine & refine
①②⑥⑦ can be used to display how data value changes and distributed.
Draw 5 different cases' increasing or decrea sing rate.
④ can be used to draw all states in the USA. Display detailed infomat-tion on each state port.
⑧ can used to show whole trend in past period, like show alcohol consumption in each year.
CATEGORIZE:
①⑧⑨
②⑥
④
⑤
Question ?
short explaination.
① smoking rate
② drinking rate
③ cancer cases
changes
FILTER: ⑤ Heat map, ⑪ Tree map, ⑩ box whisker map ③ Stack bar
My data is used to show changes in past 10 years. It contains 51 states in the USA, so below diagram not suitable for my visilization

LAYOUT



INFO

Title: changes in smoking and drinking rates in each U.S state and the number of related cancer cases from 2002 to 2012.

Author: PEIYU LIU 31153291

Date: 24/10/2021

sheet: 2

Task: represent data and draw a USA map.

OPERATION

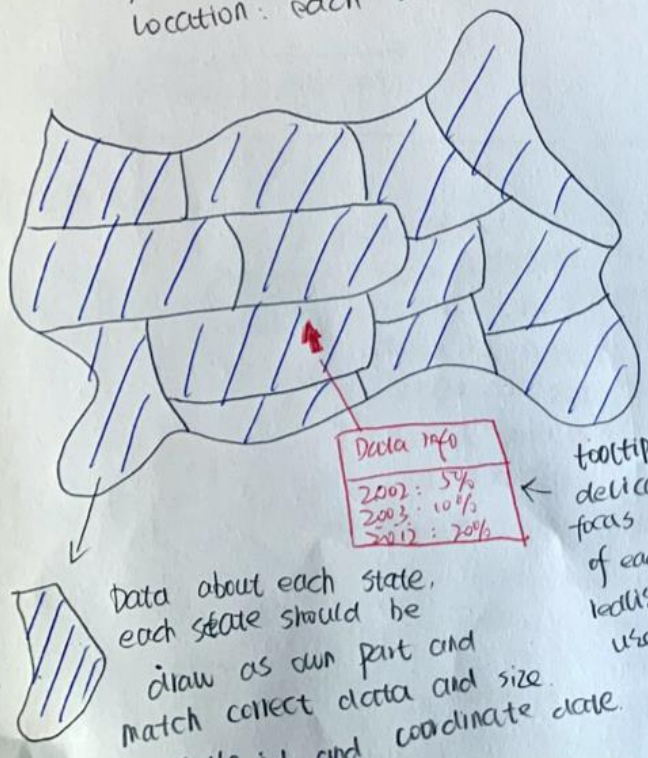
→ Convert raw data to geojson format, each State has own coordinate use coordinate to draw each state individually.

→ Set color on map acc-cording to female smoking rate. show different color.

→ add mouse events on each part of map. when user mouse on map, sha on tooltip to show value.

Focus

year: 2002 - 2012
location: each State



tooltip show delicate values. focus on real value of each year. realise function use mouse events

Data about each state, each state should be draw as own part and match correct data and size. Use state id and coordinate data.

DISCUSS

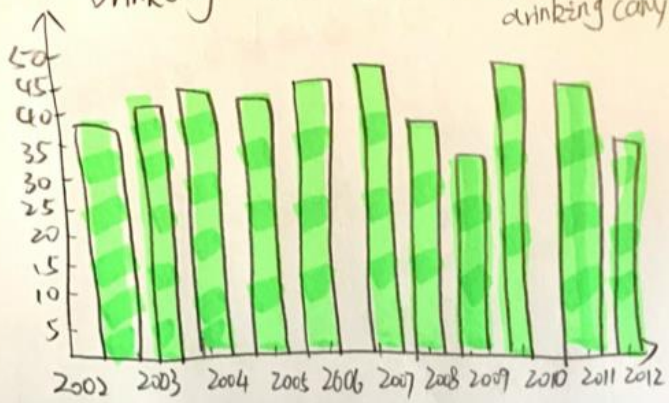
+ve: User can see space spatial. like this is a USA map. and information is about each state. use mouse event to interact with user instead of just show all information on the scre

-ve. If user wants to find information about one state, user need to move mouse to find, can not just search name

LAYOUT

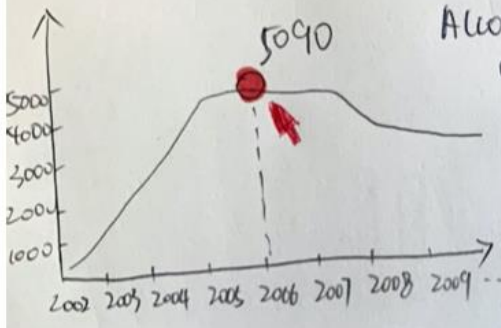
Drinking rate (%)

x: year: 2002 - 2012
y: population rate of drinking cany (ever)



Male Female

selection (genders)



5090

Alcohol consumption (each year in USA)

Year: 2002 ↓ 2012

INFO

Title: changes in smoking and drinking rates in each US. State and the number of related cancer cases from 2002 to 2012.
Author: PEIYU LIU 31153291
sheet : 3
Date: 24/10/2021

OPERATION

-> Use alcohol consumption data to draw line chart to show whole trend about this data. increasing or decreasing ?

-> add tootip related mouse event, when mouse on. show dedicated value.

-> Use alcohol using rate data to draw bar chart to show whole trend of alcohol using rate, add button to change genders. show females and males

Focus

Male Female

selection: show different bar chart. like filtering data by gender. People can see females' trend and male trend individually.

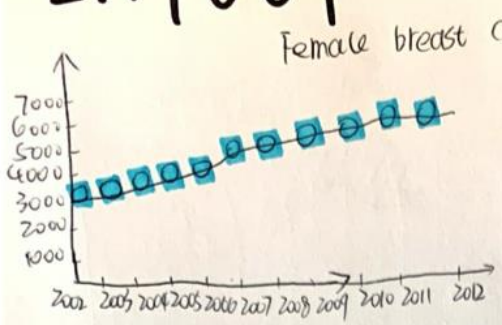
tooltip shows real value, when mouse move.
x axis is time,
y axis is value,
mouse event show interaction like display data.

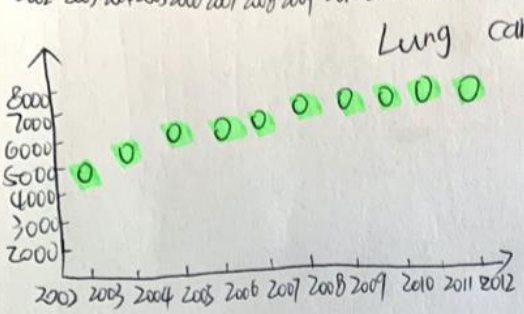
DISCUSS

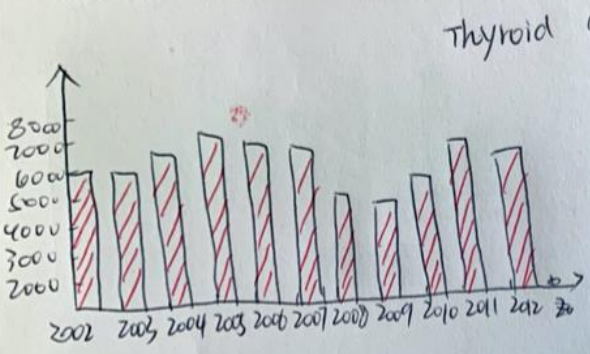
+ve: Bar chart can show data contrast clearly. can directly see trend and value. Button to change data can improve user experience. Line chart can also show trend clearly.

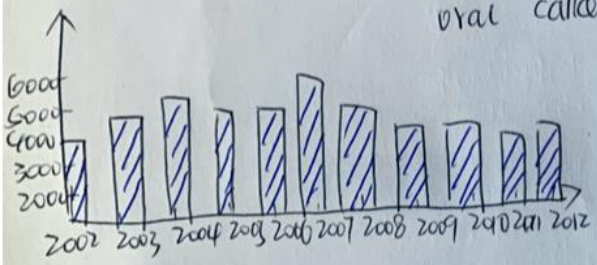
-ve: It can not show each State individually. If show all states. will have 51 bars and lines. Combine each year of all states.

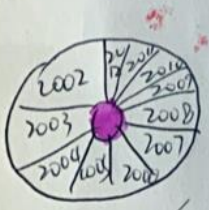
LAYOUT











Focus: mouse event to show real value

tooltip: value = 12345

value cases: 189657

Focus on each year, and related case population.

INFO

Title: changes in smoking and climbing rates in each US state and the number of related cancer cases from 2002 to 2012.

Author: PEIYU LIU 31153291

Sheet: 4

Date: 24/10/2021

Task: draw line chart to show trend of changes.

OPERATION

→ process 5 different types of cancer.

→ time from 2002 to 2012.

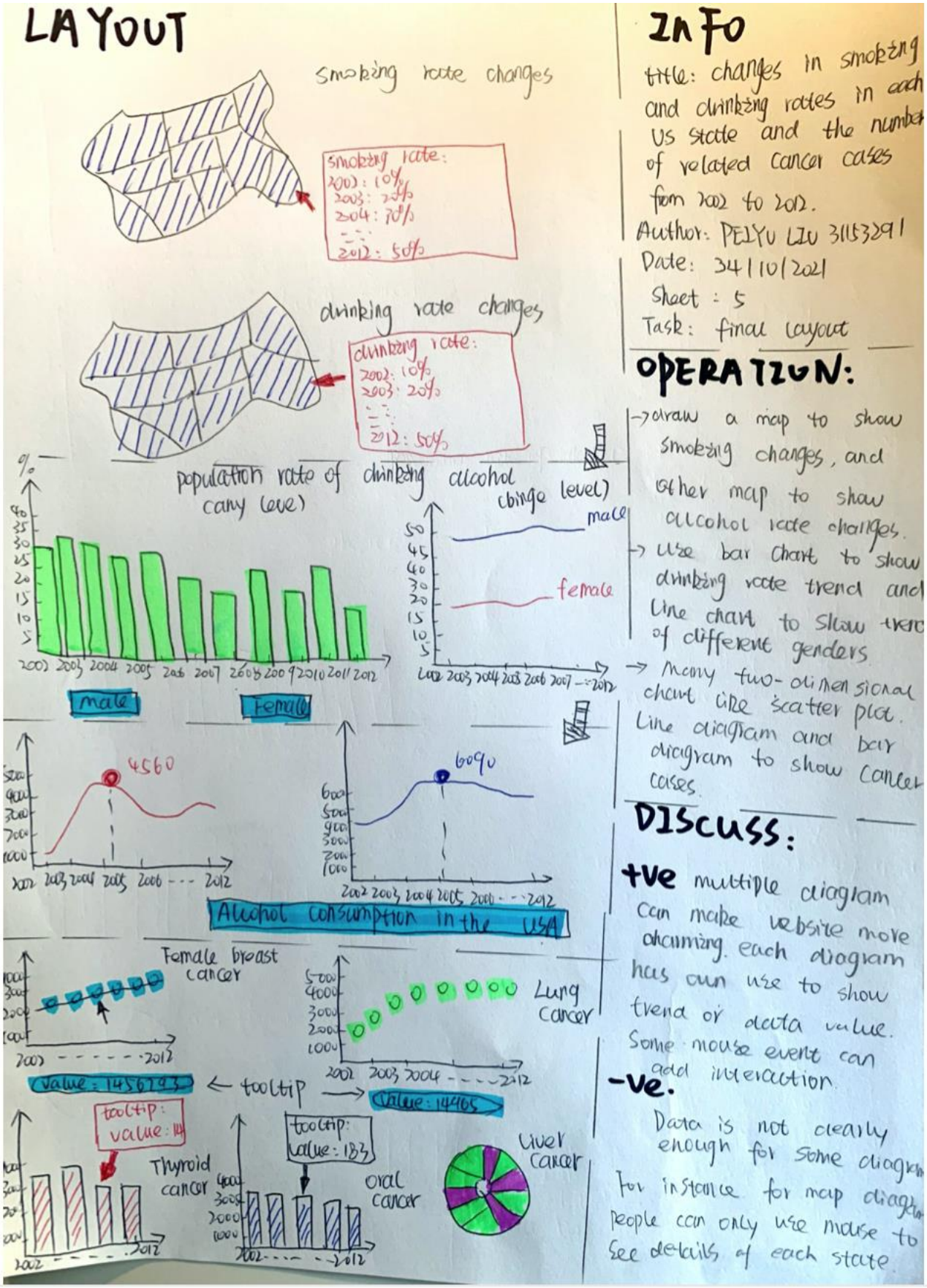
→ each year has own case population value.

→ two-dimensional diagram like line chart, dot chart and bar chart to show trend. tooltip to show value

DISCUSS

+ve: It's clearly to show increasing or decreasing trend add mouse event to improve user experience.

-ve: population value is represent all states together in this year. Can not show each State in this year.


LAYOUT
Smoking rate changes
Smoking rate:
2001: 10%
2003: 20%
2004: 30%
2012: 50%
drinking rate changes
drinking rate:
2002: 10%
2003: 20%
2012: 50%
population rate of drinking alcohol
(any level)
(binge level)
male
female
4560
6090
Alcohol consumption in the USA
Female breast cancer
Lung cancer
value: 1456793
tooltip
value: 14465
tooltip:
value: 114
tooltip:
value: 183
Thyroid cancer
oral cancer
Liver cancer
INFO
title: changes in smoking and drinking rates in each US state and the number of related cancer cases from 2002 to 2012.
Author: PEIYU LIU 3115329 1
Date: 34/10/2021
Sheet: 5
Task: final layout
OPERATION:
→ draw a map to show smoking changes, and other map to show alcohol rate changes.
→ use bar chart to show drinking rate trend and line chart to show trend of different genders
→ many two-dimensional chart like scatter plot. Line diagram and bar diagram to show cancer cases
DISCUSS:
+ve multiple diagram can make website more charming. each diagram has own use to show trend or data value. Some mouse event can add interaction
-ve.
Data is not clearly enough for some diagram
For instance for map diagram people can only use mouse to see details of each state