FIT3179
Assignment 2

Sayan Datta (31955436)

Link to Visualisation:
https://github.com/sdat0005/FIT3179-A2/blob/main/index.html

Repository:
https://github.com/sdat0005/FIT3179-A2.git

Domain:

The domain of the visualisation is examining the quality of life across various counties in the United States in the years 2010 to 2015, taking into account several factors such as unemployment rate, monthly rent, cost of living, grocery prices, etc.

Why:

The visualisations aims to help readers asses the different living conditions in the different counties in America. It compares the different factors such as cost of living, monthly rent, etc spanning across the years 2010 to 2015 enabling users to explore regional differences and trends. This allows the users to make decisions about the areas where they might want to live as they get to know about the areas with better affordability and job prospects.

Who:

The visualisation is aimed at people who are considering to relocate to United States and are curious about the differences in the living conditions within the different regions in America. It is also useful for people who want to compare affordability, job prospects and rental costs across the counties in USA, helping their decision about where they might want to live.

Data:

Sources:

URL:
https://www.kaggle.com/datasets/jayrav13/unemployment-by-county-us
Author: Jay Ravalia and US Department of Labor's Bureau of Labor Statistics
Relevance: It provides useful information about job availability across the different regions in America for a period of time.

URL:
https://www.numbeo.com/cost-of-living/region_rankings.jsp?title=2015&region=021
Author: Numbeo
Relevance: This dataset gives information about the affordability in different regions allowing comparison of expenses across different counties and years.

URL: https://www.kaggle.com/datasets/zillow/rent-index
Author: Zillow, Abigail Larion
Relevance: This dataset is useful for tracing rental costs across time and region allowing readers to asses the living expensses in America.

Preparation:

The datasets required a lot of data wrangling in order to prepare for the visualisations. Firstly, the data from numbeo isn't directly downloadable and has a separate table for each year. Hence, the data had to be manually parsed for each table and combined into a single table for all years. The unemployment rate data is also in the form of year and month in each row, but only year data is needed for creating the map visualisation. Hence, the data had to be grouped by year and the aggregate average of the unemployment rate had to be calculated for each year in each county.

The map had divisions based on county and the livability indexes

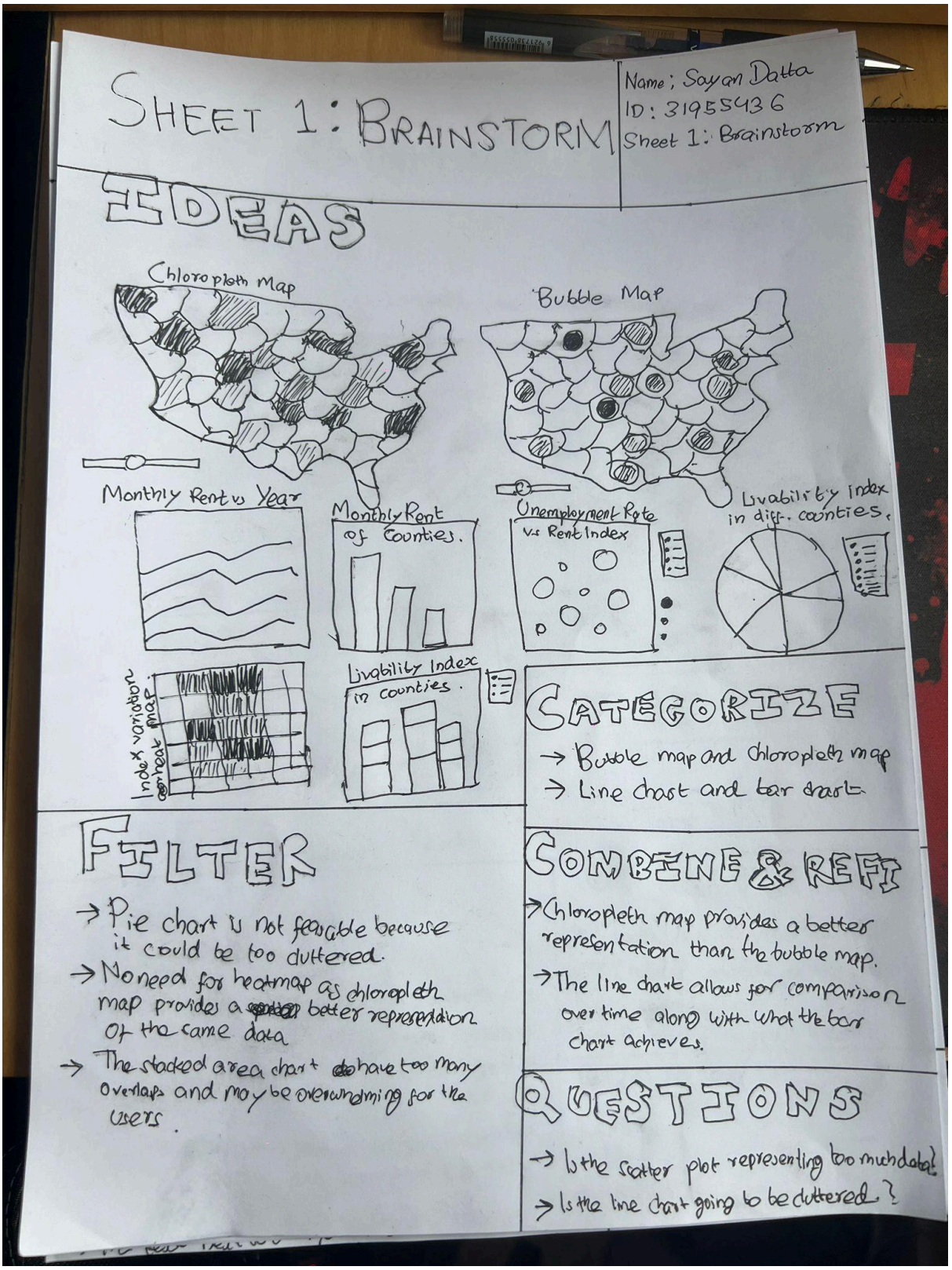
contained data based on each city and therefore wasn't being mapped on the map correctly. Hence, the respective counties for each city in the database had to added as a new county column which was done using a python script. For the line chart, the zillow rent index dataset had a column for each month-year, hence the data had to be unpivoted. The bubble chart required the unemployment dataset and the livability index database to be linked together but they didn't have any key so a county-year pair key needed to be created in order to joint those two tables to create bubble chart database.

Rationale:

The map idiom was chosen because it is easier to spot the variations of the different factors like unemployment rate and cost of living in different regions across the United States. Bubble charts are also effective for comparing different metrics, for example, by emphasizing areas where high rents might coincide with high unemployment rates, meaning limited job prospespects. The line chart is ideal for show changing trend over time, in this case, helps to observe whether the rental prices are increasing or decreasing in different counties.

The map and bubble chart visualisations also include year sliders and dropdown selectors to add interactivity, allowing users to customise their view to specific years and observing how trends changer over the years. The dropdown selectors in the map allow the users to focus on a specific aspect of the data such as unemployment rate or rent index or grocery index, etc. Adding interactivity like these increases engagement among the users and provides them with a personalised experience.

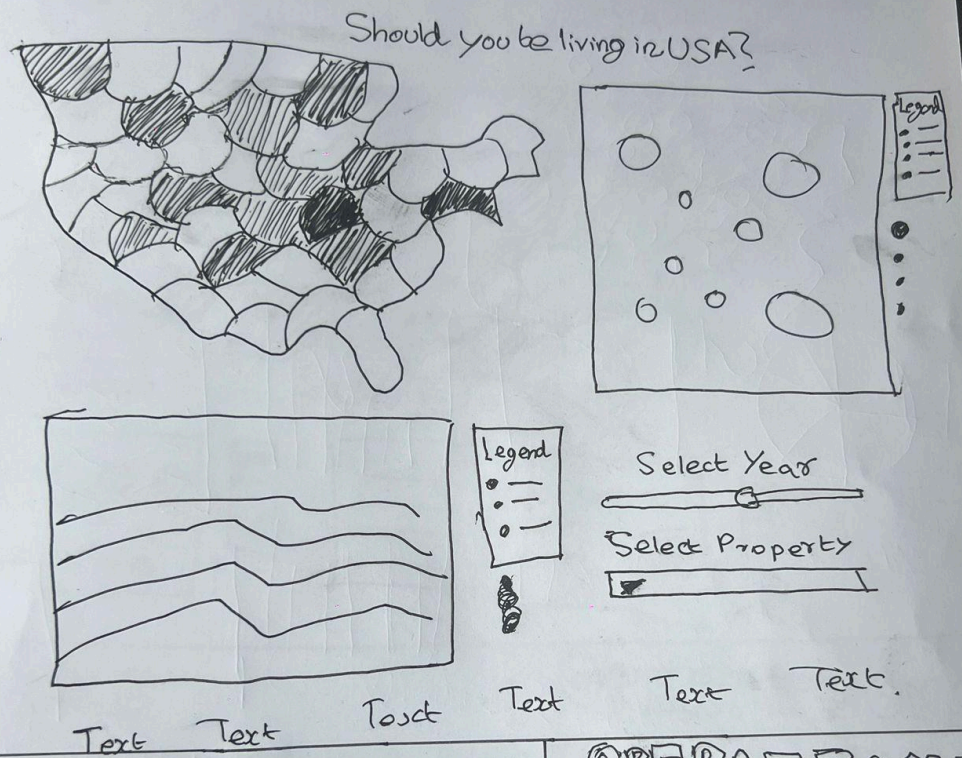
Five Design Sheets:



SHEET 2: INITIAL IDEAS

Name: Sayan Datta
ID: 31955436
Sheet 2: Initial Ideas

LAYOUT



FOCUS

→ The choropleth map provides the distribution of the different indexes in the different regions of USA allowing users to identify the regions of choice.

→ The bubble chart is illustrating the relationship between rent index and unemployment rate.

→ Line chart shows the trend of rent over time.

OPERATIONS

→ Tooltip on all visualisations.
→ Clickable legends on line chart
→ Slider for year.
→ Dropdown to select index.

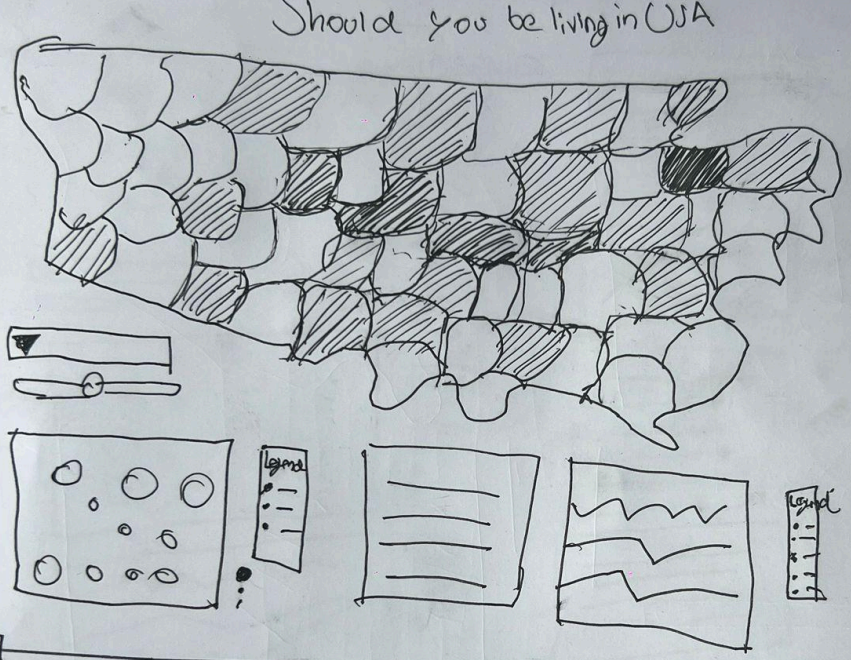
PRO AND CONS

Cons.
→ No space for texts
→ No visual hierarchy

SHEET 3 : INITIAL IDEAS

Name : Sayan Datta
ID : 31955436
Sheet 3 : Initial Ideas.

LAYOUT



FOCUS

→ This layout mainly focuses on the map visualization and uses the other idioms just as a supporting visualization.

→ The text box in the middle grabs the user's attention.

→ The bigger map at the top and the smaller visualizations and text at the bottom creates a visual heirarchy for the user.

OPERATIONS

→ Common year slider affecting both map and the bubble chart.

→ Drop down selector for map to select property.

→ Select and highlight counties on map.

PROS AND CONS

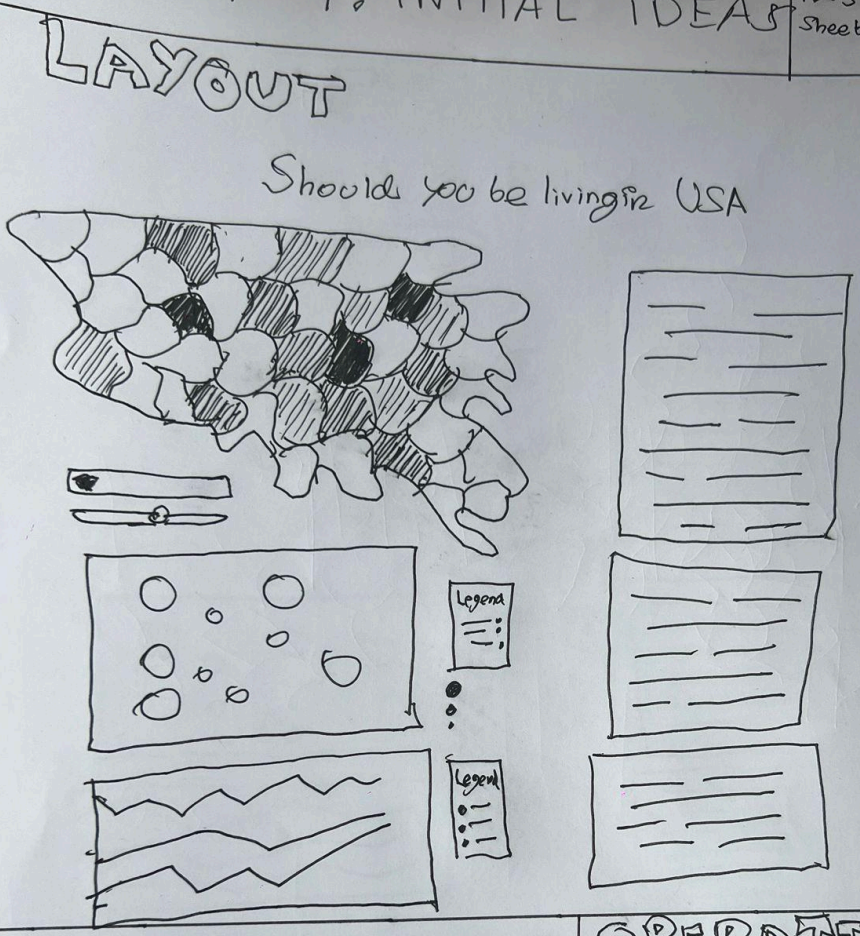
→ Good visual heirarchy

→ The supporting maps may be too small

SHEET 4: INITIAL IDEAS

Name: Sayan Datta
ID: 31955436
Sheet 4: Initial ideas

LAYOUT



FOCUS

→ All the idioms are same size so have equal importance in the visualization.

→ There is a clear visual hierarchy

→ The texts right next to the graph allows ~~the eyes to~~ easy comprehension of the data

~~The eyes~~

OPERATIONS

→ The year slides allows filtering by year for both map and bubble charts.

→ Map zoom in feature and selecting counties.

→ Click on line graph legend to select the lines on legend.

PROS AND CONS

→ Good user readability.

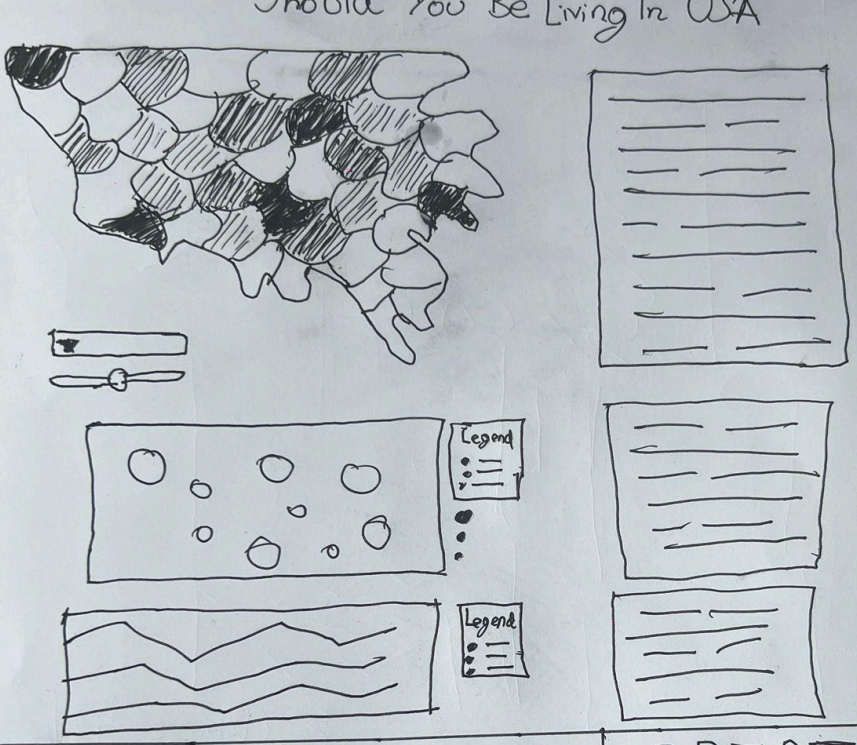
→ Contains interactivity.

→ Contains visual hierarchy

SHEET 5: REALIZATION

Name: Sayan Datta
ID: 31955436.
Sheet 5; Realization.

LAYOUT



OPERATIONS

→ Equal sized charts give visual importance to all the idioms equally.

→ The dropdown selector on map allows user to view regional variances for different livability indexes.

→ The year slider adds filter to both map

→ The ~~xxx~~ items on the line graph are selectable to highlight a specific line in the line graph.

→ The map is zoomable to view the regions clearly.

→ The text next to graphs allow user readability.

DETAILS

→ Dependencies: Python for data cleaning, data wrangling and preparation. Vega lite online for creating the visualizations. Liveweave for coding HTML and CSS with live preview and github for hosting.

→ Time and effort: A day or two for data prep, a week to create visualizations and a day or two to make the website