Alpha Release Design Document

Design Process

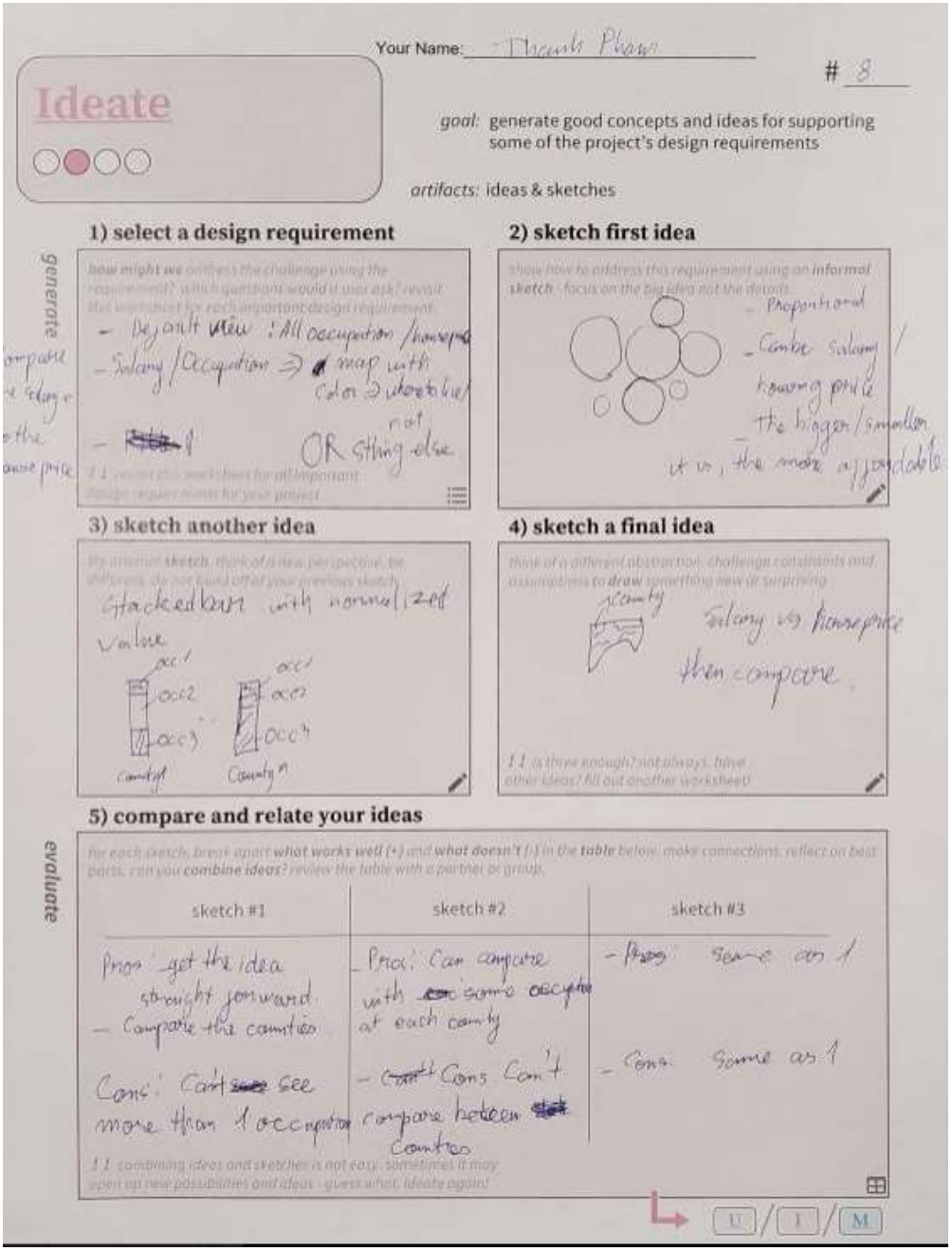

Fig 1: The design sheet showing the ideation process that was used to come up with some potential visualizations and the design requirements.

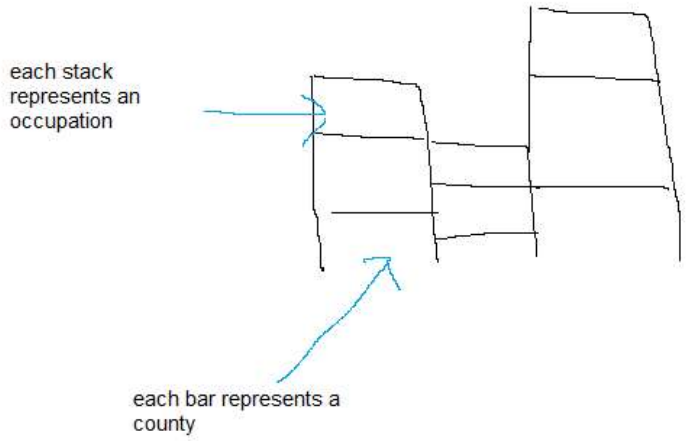

Fig 1.1: A clearer pic of bar chart.

Design Assumptions:

1) Since the wage data for various occupations is at state level, we are assuming that the average wages remain the same for all the counties within a state.

Design Requirements

Based on the design process and activity (Fig 1), the following requirements are identified as part of the design requirements.

Requirement ID	Requirement Description	Priority
DR_001	Users must be able to filter for multiple counties	High
DR_002	Users must be able to filter for multiple occupations	High
DR_003	Users must be able to view visualization showing the affordability per the selections. Affordability is defined as: Affordability = (Average Salary of the occupation / House price in the county)*100	High
DR_004	The visualization must be able to change dynamically per the filters applied by user	High
DR_005	Users must NOT be asked for any personally identifiable	High

	information like name, address, and Date of Birth.	
DR_006	Users must be able to filter the visualization for multiple states	High

Data Pre-processing

- Dealing with missing data: Elimination of following states from salary data which do not have corresponding housing data
    - ‘GU': Guam
    - 'PR': Puerto Rico
    - 'VI: US Virgin Islands
- Dealing with incomprehensible data: Conversion of following symbols with quantitative data in the salary data:
    - * = indicates that a wage estimate is not available (NaN)
    - ** = indicates that an employment estimate is not available (NaN)
    - # = indicates a wage equal to or greater than $100.00 per hour or $208,000 per year

Visualizations

Based on the design requirements listed above, we came up with a few visualization ideas including a stacked bar chart, heat map, a geographical map (choropleth), a line chart and a bubble chart. As part of the alpha release, we were able to code a stacked bar chart and a heat map showing affordability metric. Both of these visualizations meet all of the design requirements listed above except DR_006 as we are planning to expand our data to other states after alpha release.

- Stacked bar chart
  Shows affordability (a new metric derived from housing and wages data) for chosen occupations in chosen counties for different states (Maryland for alpha release).
  We are transforming the data to define a new metric ‘Affordability’, which is defined as:

  Affordability = (Average Salary of the occupation / House price in the county)*100

  The greater this ratio, the more affordable are the houses in a county for an occupation. The stacked bar chart shows the affordability metric. The user can select multiple counties and occupations (Fig 2) to further filter it and it does not require any Personally Identifiable Information (PII) from the user. It changes per the filters selected by the user (Fig 3). This is how the user will be able to interact with the visualization. The usage of different colors here reflect the affordability of the houses for various occupations in a particular county.

Fig 2: County and Occupation Filters available to the user. The user can select multiple counties and occupations to filter the data visualized in the stacked bar chart

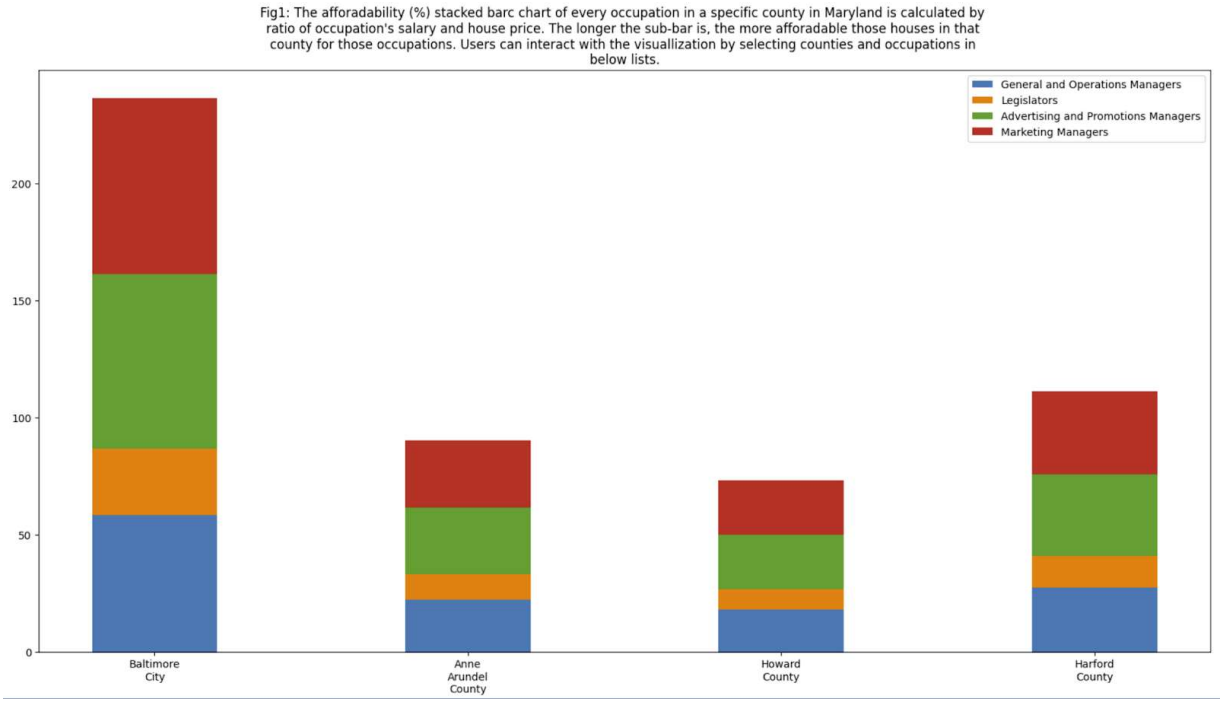

Fig 3: The affordability (%) stacked bar chart of every occupation in a specific county in Maryland is calculated by ratio of occupation's salary and house price. The longer the sub-bar is, the more affordable those houses in that county are for those occupations. Users can interact with the visualization by selecting counties and occupations in below lists.

.

- Heatmap

  The heatmap also visualizes the affordability on a map visualization using the same affordability metric. The greater the affordability, the darker the saturation of the color in the county.

  This heat map shows the affordability metric. The user can select multiple counties and occupations to further filter it (Fig 4) and it does not require any Personally Identifiable Information (PII) from the user. It changes per the filters selected by the user (Fig 5). This is how the user will be able to interact with the visualization.

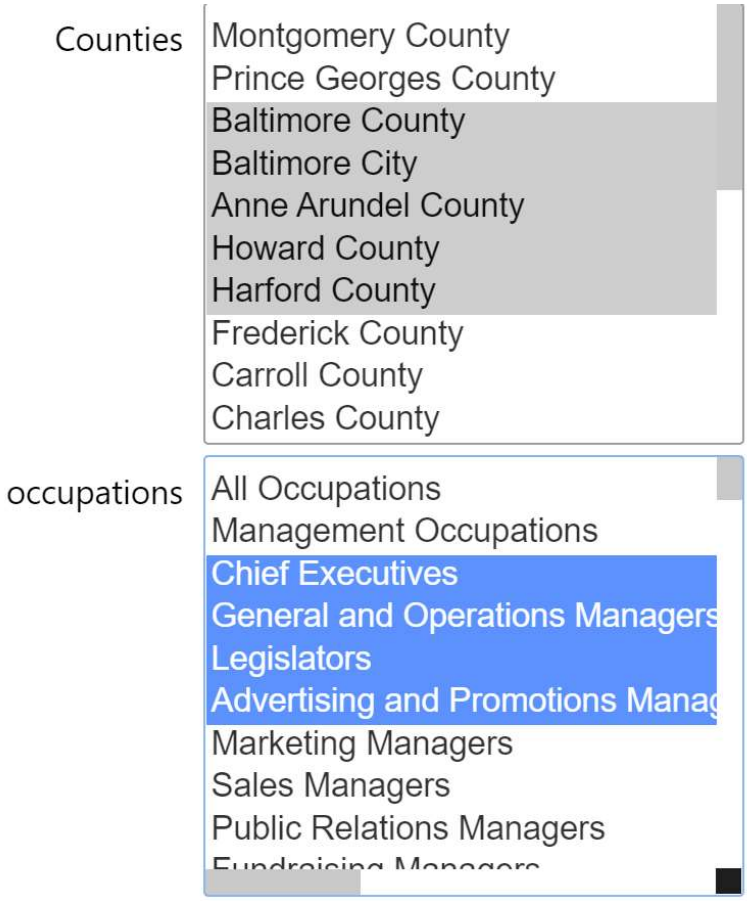

Fig 4: County and Occupation Filters available to the user. The user can select multiple counties and occupations to filter the data visualized in the heat map

Fig2: The afforadability (%) heatmap of every occupation in a specific county in Maryland is calculated by ratio of occupation's salary and house price. The darker the cell is, the more afforadable those houses in that county for those occupations. Users can interact with the visuallization by selecting counties and occupations in below lists.

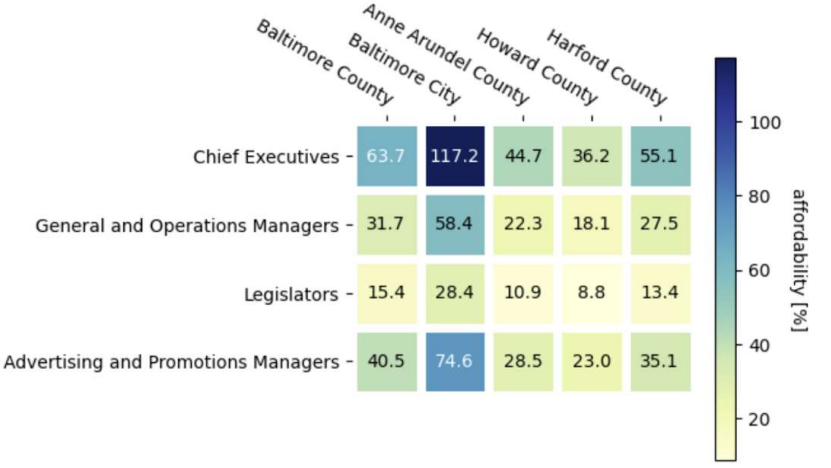

Fig 5: The affordability (%) heatmap of every occupation in a specific county in Maryland is calculated by ratio of occupation's salary and house price. The darker the cell is, the more affordable those houses in that county for those occupations. Users can interact with the visualization by selecting counties and occupations in below lists.

- Geographic Map
  The geographic map also visualizes the affordability on a map visualization using the same affordability metric. The greater the affordability, the darker the saturation of the color in the county.

- Bubble Chart
  This visualization was part of the design process to visualize the different salaries within a state.

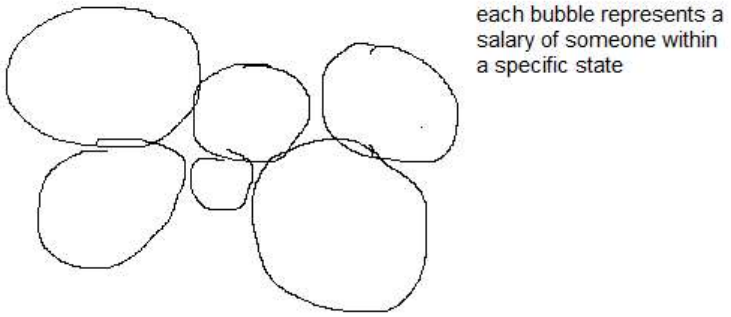

Fig 6: A sketch of the bubble chart representation created during design process

- Line chart
  This visualization was part of the design process to visualize the trend of housing prices within a county in a defined time period

Comparison between Stacked Bar Chart and Heat Map

	Stacked Bar Chart	Heat Map
Trade off	Easy to implement and understand but not scalable	Easy to implement and more scalable but require some background knowledge
Pros	Easy stand out for outliers, Simple and easy to comprehend	Provides rich insights in terms of usability and are easy to understand

Cons	As the counties and occupation selected by the user increases, it becomes very difficult for user to understand the data	Can be interpreted in different ways, hence, require background information
Scalability Limits	It is good for 3-4 layers. As the layers exceed this threshold, it becomes confusing for the user to understand.	It is more scalable than Stacked Bar chart
Cognitive Load	It has less cognitive load.	It has more cognitive load

Future Plan

- Creation of Choropleth map for the affordability metric for state of Maryland
- Expanding the visualizations to other states
- Introducing a filter for state selection by the user

References:

https://matplotlib.org/stable/gallery/images_contours_and_fields/image_annotated_heatmap.html

https://matplotlib.org/stable/gallery/lines_bars_and_markers/bar_stacked.html

https://ipywidgets.readthedocs.io/en/stable/examples/Widget%20List.html#