Alpha release: Rakshith Venkatachalapathy

I came across Zillow research data and Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type for homes across the US.
My goal was to take this large data set present its findings using visualization techniques that we learnt in this class. Since I moved to the US, I've always wondered what it costs to live in the US and how much people invest in real estate. By visualizing this dataset, I am hoping that the above questions would be answered through simple visualization.

The idea is to create an interactive dashboard with 4 different panes. All these panes are synchronized and allow synchronized interactions.

These four panes have different visualizations as described in the revised document.

Features completed:

1. Data cleaning: Initially the dataset contained a lot of rows and columns.
   - 2.8k rows and 300+ columns
   - Each row is a US county
   - Data has been sourced from: https://www.zillow.com/research/data/
   - This data had to be scrapped for columns like RegionName, StateCodeFIPS, MunicipalCodeFIPS

2. Generate geo location details for the counties. This is done in tableau which provides feature to concatenate columns. By concatenating the StateCodeFIPS and MunicipalCodeFIPS, I was able to generate a county ID which was helpful in plotting the choropleth map.

3. Once the data cleaning was done and with the availability of the county id, I was able to plot the choropleth map for visualizing the median house selling price details for one particular month of 2020. This was done with the help of D3.

   Color and scale: For the choropleth map, I used a sequential color scale. I used sequential colors of blue to represent the counties in the choropleth map. I used five distinguishable shades of blue to represent the different counties. A legend is also provided to help distinguish the five colors.

   Data ink for choropleth map: The map will have some background as there are a few counites missing and the background will be seen.

4. With the availability of the county ID, I was able to plot the bar graph for 3 neighboring counties (with D3). This constitutes P3 of the project objectives.

Color and scale: To represent the bar graph, I used blue to fill the bars to match the color scale in the choropleth map. The bars are thick, and this width of the bar is going to distinguish it from the other bar graph form P2(P2 yet to be implemented).
Data ink for the bar graph: The data ink on the bar graph is very low. A little data ink should be added to help the viewers distinguish the bars. This is can be done with the help of adding a few lines to support the bar chart visualization.

5. Design:
   The overall design is very simple. The usage of simple visualization techniques like bar graphs, choropleth map and line charts are simple to understand and easy to interpret the data being shown to the end user.

   Lie factor:
   For the visualizations created for alpha release:
   - With the choropleth map, there are 5 distinguishable shades of blue and it is difficult to tell the difference between the first two colors on the legend as there are a lot of counties of the same color. The data that is visualized is the median selling price of a house and it can be misunderstood for population density because of the color changes on the east and the west coast.
   - With the bar chart, it is difficult to tell the exact value of the bar even though the axis has the value. This is because of the interval size of the axis.

   Visual Encoding:
   - In case of the choropleth map, sequential color scale is used with the color blue. This scale has 5 different shades of blue to help distinguish the median sale price.
   - In case of the bar chart, the length of the bar varies according to the median sale price which helps the user to distinguish between the bars.

Upcoming milestone:

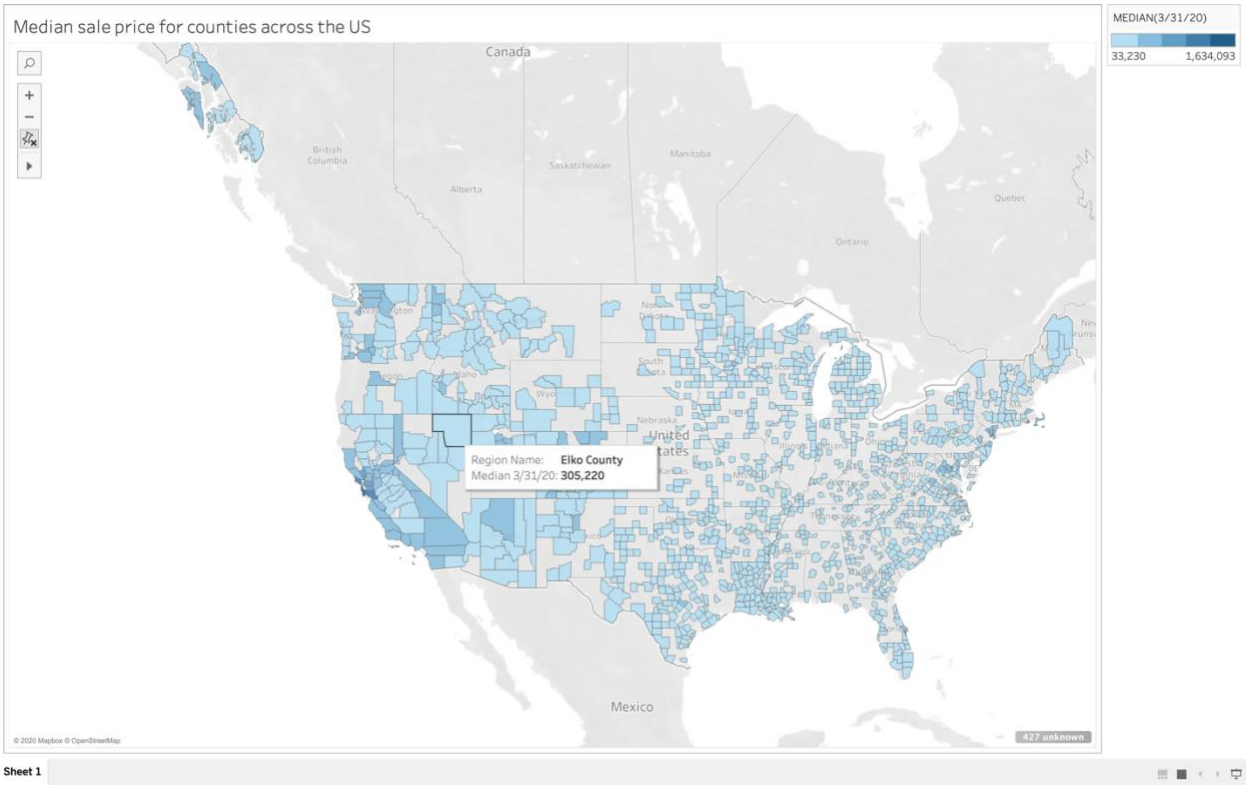
1. The dataset does not contain the median values for a few counties. These counties are the ones which are sparsely populated and do not contain any data. I need to come up with a way to represent the same so that it looks uniform on the choropleth map and reduce the data ink.
   The data needs to be put into a particular format to represent these null values.
2. Work on the dashboard. The dashboard needs to have 4 panes together and these panes have all the primary objectives from P1-P4 as described in the revised document.

Roadblocks:

1. The dataset does not contain the median values for a few counties. These counties are the ones which are sparsely populated and do not contain any data. I need to come up with a way to represent the same so that it looks uniform on the choropleth map.

2. This is very evident in the visualization that I have created using in tableau as shown below.
3. This involves coming up with a color scale and represent these null values in a proper format to represent the same.

P1 – Choropleth map



P3 – bar chart for the neighboring counties