A3: White Hat/Black Hat Visualization

Amindu Abeydeera, Divjot Muchhal, Nikhil Paruchuri

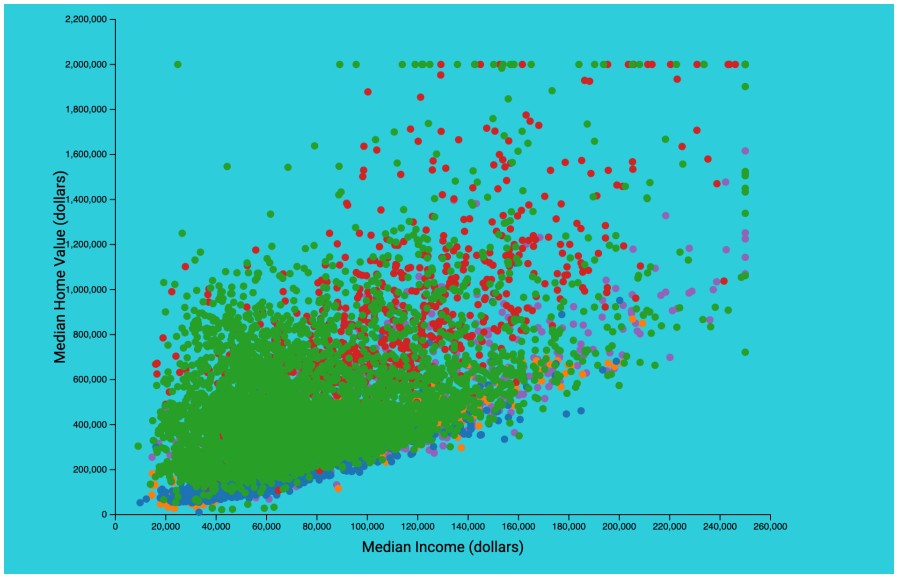
Development Process:

Regarding our development process, the task was divided fairly evenly amongst us. Having 3 group members made it easy to split up the work fairly. As opposed to the last project, our schedule for the week was a bit more relaxed, so we didn't have to utilize additional slack days to complete the project and had enough time to put in careful work toward each visualization and the writing process. We decided to follow the same development process as our last project, as it gave us the most success. This process involved dividing our workflow into 4 parts: the creation of our visualization on Github, the write-up, the demo video, and the actual implementation of our visualizations. Nikhil decided to work on the white hat visualization, finding the relationship between Median income and Median home value, and its demo video, as well as setting up the GitHub repository for the project and deploying the webpage. Amindu worked on the black hat visualization, finding the relationship between the top 10 high-population areas and population count, as well as its demo video. Divjot was responsible for the write-up about both of the visualizations. We decided to plan what we wanted to accomplish, starting with which dataset to implement. We collectively agreed on the Gentrification Dataset. In terms of time spent on this assignment, each member spent roughly 2.5 hours planning and implementing their visualizations. Most of the time spent on the visualizations was allocated toward debugging various code issues and visual issues which will be mentioned in more detail

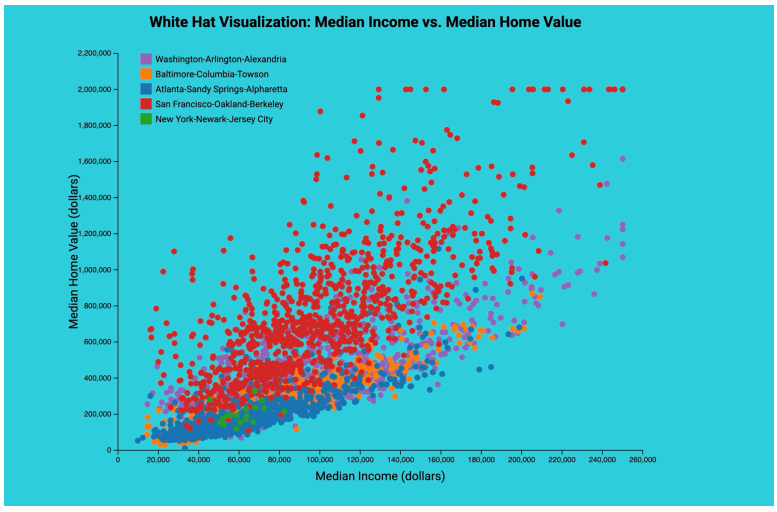
when covering the visualizations separately. Additionally, a good amount of time was also spent analyzing our visualizations at the end of development to find additional improvements or deceptiveness that could be added depending on the visualization being worked on. Other than the difficult debugging and visualization fixes, however, the rest of the assignment went fairly smoothly. Overall, our planning, workflow, and actual development were smooth and collaborative allowing us to create a meaningful White hat visualization and a deceptive black hat visualization.

White Hat:

For this project, the dataset"Gentrification and Demographic Analysis" is the basis for the visualizations. Buzzfeed created this dataset to observe how the character and demographics of neighborhoods change as more wealthy individuals and businesses move in and potentially displace existing residents. It contains tract-level census information on five major U.S. metropolitan areas—New York, Washington D.C., San Francisco, Seattle, and Los Angeles—covering demographic and socioeconomic indicators for 8,281 individual census tracts. Among the key demographic measures are the total population and population aged 25, along with data on bachelor's degrees or higher for educational trends. Economic indicators include median home value and median household income, both of which are significant in measuring affordability, displacement risk, and neighborhood change. The dataset also includes detailed racial and ethnic composition, with counts for categories such as White alone, Black alone, Asian alone, Native American, Pacific Islander, two or more races, and Hispanic or Latino. Each tract is linked with a city and metro area, allowing for both intra-city and cross-city comparisons. The white hat visualization aimed to present the relationship between median income(x-axis) and median home value(y-axis) for various census tracts or “metropolitan areas” while prioritizing transparency, clarity, and accuracy. For clear communication, the visualization is built with simplicity in mind, with an intuitive scatter plot to show this relationship. The color scale is applied to differentiate the metro areas, helping the viewers distinguish the points and identify trends within the region. Tooltips were also used to provide detailed data including census tract ID, income, and home value making the chart both informative and interactive. For transparent data transformation and aggregation, the data is filtered by metro area allowing viewers to see data from all regions. However, one issue we had to resolve was the New York metro area data points ended up covering a significant portion of our data, leading to less clarity in our visualization, as shown in this image:



Fixing this issue took a lot of discussion and time, but our final solution after various discussions was to filter out a lot of the New York Metro (Green) area by census number instead of location, which ensured randomness. This allowed for a clearer visualization providing clarity for the user. There were still values from the “green points” because it represented more than one metro area so we were not losing too much data. This was the result afterward:

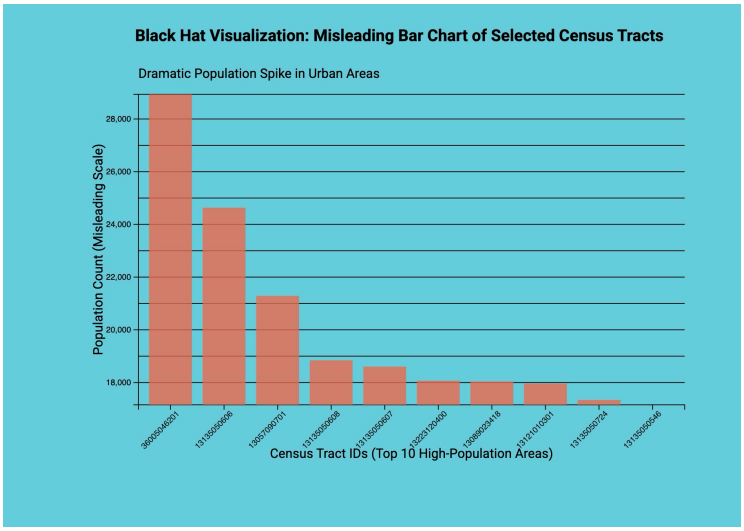


Additionally, for marks and encodings, the plot uses circles to represent the individual census tractors with a uniform size. The positions of the circles are mapped to the median income and home value, providing an accurate representation of the data. The color coding and the color

legend clarify any visual doubt to the viewers about which metro area is which. Lastly, the chart includes clear titles and axis labels, helping contextualize the data present. Overall, each design choice for the white hat visualization makes it engaging, informative, and transparent for viewers.

Black hat:

This black hat visualization aims to intentionally manipulate design elements to create a misleading narrative based on 2 key aspects: deceptiveness and subtlety. With deceptiveness, this black hat visualization has the y-axis not starting at 0. For a non-zero starting label, the data differences are exaggerated making small disparities appear bigger than they may seem. For example, the distance between the starting tick to the first tick is 18000, whereas for the rest of them, it's 2000. This over-exaggeration can be seen when we look at the bar to the very left and the bars to the right in the image below:



The difference between the highest population increase and the lowest is around 12-14k, which is not extreme for urban areas even though the graph makes it seem like it is. Furthermore, the selective filtering of the data set(data transformations), also known as “cherry-picking” is another form of deceptiveness. Only the top 10 census tracts by population are shown, which disregards a macroscopic view of the dataset. Showing only a small biased subset of the data makes viewers believe the areas shown represent all census tracts, which in reality they do not.

This distorts the viewer's perception, leading to inaccurate representation of the data. The last form of deceptiveness comes from the misleading chart title and axis labels. The chart title has language(“dramatic” and “spike”) that evokes an emotional aspect of the viewers, suggesting the chart shows a very important trend, but the data may not even support this claim. Additionally, the labels both push toward misleading narratives with the x-axis label making viewers assume it represents the most significant areas of the dataset when it is just a subset of the data. Subtlety is shown through the rotated x-axis labels introducing a sense of complexity without any real purpose. It may look like the rotation is a design choice to better fit the data but it only adds visual clutter and makes it harder for viewers to read and understand. This subtle clutter can distract the viewer from the overall point of the data bringing potential confusion. Another form of subtlety in this visual is the transparent bars and grid lines, creating a false sense of depth, complexity, and accuracy. The opacity of the bars gives this illusion that there may be some hidden data and that the data has more meaning and value than what it gives off. Additionally, the use of the grid lines can make the viewers think there is some accuracy factor or some sort of precision when it comes to the data when in reality it is not. It contributes to the exaggeration of the visual differences with the data points. Overall, the whole visualization design with its marks, encodings, data transformations, and titles/labels contributed to the key components of black hat visualization of deceptiveness and subtlety. These design choices were meant to push a false narrative towards the viewers making viewer perception distorted and unable to use the data for any real analysis.