

Whitehouse Logs Write Up

The audience for this presentation will be the White House staff that logs the visitors into the White House. They may be familiar with the data. However, they might not be familiar with the importance behind why they're logging the information for the visitor. For the presentation, I tailored it to someone that may not have a good understanding of the data and made it easy to comprehend.

The purpose of the presentation is to bring awareness that the logs are not being completed properly. Thus, this is the call to action; the White House visitor logs need to be properly completed to prevent security risks to the white house and its guests. I want the audience to see that there is a significant number of visitors to the white house throughout the year and many of the logs are not properly filled out for who or where they're visiting. If we're unable to track this information, it makes it difficult to identify the locations to increase security to prevent any visitor from doing any actions that threaten the safety of the white house and its guests. I'm hoping that the audience walks away having a better understanding of why we need the logs filled out properly. This would fall under a change of direction from the current habits for whomever is entering the log information.

The medium of choice for this presentation is a power point. This was chosen as this type of presentation is ideal for staff training. Ideally there will be a classroom in which I would be able to present this information to the staff. The session can then be recorded and a link of it can be emailed to the staff that were unable to attend the classroom training.

While designing the visualizations I was working towards yearly trends of visitors to the white house using the 2023 data. I applied the figure-ground to the bar graph visualization. However, I wanted the audience to see the overall trend of how the data was being entered, so this was not applied to the stacked bar graph. Proximity was used for the scatter plots. The data points were separate enough that the audience wouldn't consider them the same, as well as the difference in color. Continuity was used in the later scatter plots with the line of best fit. This made the data appear more related to each other. However, I did use line colors to match the data point color. This creates the separation of known and unknown visitors. This would also help to avoid common fate. Symmetry was easier to see in the pie chart. However, this concept was also applied to the remaining graphs for the axis and overall design.

The colors chosen for the visuals were primarily blue. However, when I was making visuals for data that contained two items, a light blue and orange were used to prevent from

using colors that would be misconceived for color-blind individuals. The text I chose was Arial. This tends to be clean and easy to read. I did increase the font size for the title and abbreviate the months for the x-axis to make the visual easier to read. The title was moved to the upper left of the visual and any legends were moved to the upper right to keep the flow of the z-shape for English readers. Spacing for the bar charts were left as the standard from python as this appeared to be the right amount of white space between the columns.

The data set was built by combining 12 csv files and was cleaned by using mode to fill in missing values for categorical and mean for the continuous values. The dates were filled with forward fill. I then changed the dates from object to a datetime type. I reduced the data frame to contain only the columns that I found interesting for my subject. I converted the names of individuals to known or unknown and created a new column that contained the month the data was released. There do not appear to be any legal or regulatory guidelines for the data. The data source only mentions that the president will release this information according to the disclosure policy. I was unable to locate information pertaining to how the data may be used.

The transformations I used could have caused data leakage from a previous month into the next from using the forward fill method. The mode and median could have been affected by one population of data that may make a specific category over-represented in the data. With the changes I made I did anonymize the data by removing visitor names. While cleaning the data I assumed that the data was already organized by month, hence I used forward fill to fill in any missing dates. I also, assumed that the release date would contain the visits for that month. I assumed that the visitor log would have been similar month to month. I should have uploaded them individually to prevent causing any chance of creating inflated numbers for monthly visitor numbers. I did filter the data by name. However, the visualization should be representative of the change showing that so many visitors had a known visitor and others were unknown. The data was sourced and acquired in an ethical manner as this is public access information that was released by the president as part of the White House disclosure policy.