Scott Elmore

3/11/20

DSC 465

Final Project Report

**QB Heat Map:**

In doing the heat map, we wanted to show how each NFL team’s quarterback (QB) performed in a metric that is shown to be a good descriptor and predictor of success, EPA. This deviates from typical analyses that focus on passing yards and touchdowns as metrics that relate to QB success, because EPA tries to account for the fact that not all yards gained should be given equal weight, and not all touchdowns are because of a QB’s proficiency. The heat map lists each QB in order of their sum of EPA over the course of a season, so the heat map shows a definite trend of color changing vertically from top to bottom. The dimensions of the chart are created by ‘binning’ the yard line on the field at which the play happened from. We decided to use 10-yard bins because that’s how a football field looks like in real life, and it helps the audience understand that the visual is supposed to relate to an actual field when viewing. In the first draft, we kept the yard line as a continuous variable, but with 99 possible yard lines that a play could start from, the chart was too sparse by showing some encoding at each individual yard line as opposed to the 10 yard bins, so it was best to create a categorical dimension via the bins. Next to the QB’s name is some text that shows if the QB played in the Superbowl, because keeping true to the theme of our project, we want to show where the Superbowl stood out among the pack. We also wanted to emphasize this further by adding a Tableau feature that ‘highlights’ the Superbowl QBs whenever the mouse is over one of them. This is an example of a pre-attentive attribute, because the highlighted QBs have their entire row shown with more brightness, and the non-highlighted rows are dimmed, which can be seen immediately by the audience. We used a diverging color scale to show average EPA that changes hue when going from a positive value to a negative value, and changes saturation when the value approaches the center. The center was decided to be the average EPA of all QBs. In addition to color, the other encoding is length of the bar which depicts the count of plays at the given yard line bin. This encoding is important, because the count of plays can vary from one yard-line bin to the next, and it is important to keep in mind when telling the story of success via EPA. The length of the bar can be reliably compared vertically amongst QBs because they share an axis as described in Weber’s Law. The comparison horizontally amongst different axes is a little tougher, but still reliable form of comparison for the audience.

The second chart that can be seen on the page is an interactive plot that displays the pass distribution for each QB. This chart takes advantage of the interactive filtering capabilities of Tableau, because each QB is filtered down depending on which row is hovered over and can be filtered even further depending on which 10-yard bin is also being hovered over. If the chart on the left is an overview of the success of a QB, the chart on the right is a good look into how they made this success. The ‘Pass Locations’ chart has to non-aggregate measures because we want to show how each data point looks in comparison to the others. The chart is essentially a scatter plot with the x-axis representing the location on the field the throw happened (left, middle, or right) and the y-axis representing the yards down the field the throw happened. The y-axis is held constant to again better provide a fixed baseline for QB comparison when hovering over different QBs. The encodings are colored circles to represent another EPA measure, but this time a slightly different version of EPA is used. We thought it would make most sense to use ‘Air EPA’ because this shows the value of a QB’s throw without any effect of what the receiver did after they caught it. Air EPA measures the EPA of a play of only the yards gained by the throw, not the catch and run after. The color is once again diverging, but the color is fixed at the min and max, to better provide baseline values for comparison. It is important to also use a less opaque color in order to show density at some of the yard positions, because there can be overcrowding at some spots. This chart is shown vertically as opposed to horizontally like the other chart, because it shouldn’t display a fixed football field, but rather from the view of the QB, and where their throws ended up.

**49ers Defense Boxplots:**

This plot was done using ggplot2, so it required some coding to make the visual. We decided to use python to wrangle and modify the data to make it better suited for ggplot2 functionality. The point of the chart is to show exactly how dominant the 49ers defense was, and why it led to them being in the Superbowl. We pulled 9 statistics from the data that all try to measure defensive success, and while some measures are more important than others, the entirety of all nine displays a comprehensive summary. The first thing necessary to do for this visualization was create a dataset that only had the defensive statistics we wanted to show. We did this by pulling those dimensions and rows where the fields are not null into an excel table using Tableau’s export feature. Then we pull the table into a python panda’s data frame using .read\_excel(). But in order to get a team by team breakdown, the data had to be aggregated up to the team level, so we used .groupby() to get the team level data. Next, to make the column names categorical variables themselves, we employed the .melt() to get a set of data that has as rows ‘team name’, ‘variable’, and ‘value’. Here is where we started making plots to see what the distributions looked like and which plot type would look best. We liked the way the violin + sina plots looked, but they lack the quartile breakdown like the boxplot, which goes a long way to show the dominance of the 49ers defense. After using color to differentiate between the teams, it was clear that all the colors messed up the message of the plots. So, we went back to make another variable that was only True if the 49ers were the team in the row of the data frame. This allowed for only two colors, which is red (49ers colors) if True, gray if False. The boxplot needed the jitter plot to be attached also in order to see the individual team breakdown as points, without the overlapping of dots along the same axis. The plots look much less cluttered with the new color scheme and jitter, and the red stands out in a pre-attentive way to attract the audience to seeing where the 49ers land in the distribution of all NFL teams. The plots do a great job of showing how the defense performed above the median in every category and in most cases, above the upper quartile. It should be noted that in some plots, being closer to 0 is better, like for 3rd down conversion and yards gained. This chart started as an exploratory visual because it was just a way to view the distributions of the defensive statistics, but it became a explanatory visual because of the way the dots looked inside the boxplot, with the 49ers always being closer to the good side of the distribution than the bad. It is typical for sports journalists to cherry pick data to prove a narrative, but here the narrative is clear when looking at all 9 statistics at once, the 49ers defense was very, very good.