*Visualization Descriptions*

Olivia Ericsson, Yuna Lee, Henry Barrera

[Link to website](https://oliviaericsson1.github.io/)

**Title: What are the main causes/factors that contribute to a stroke?**

***1: Heat Map (Static Visualization #1)***

The heatmap displays a covariance matrix using shades of red to represent the intensity of covariance between variables; darker shades indicate higher covariance. We chose a default red color gradient because it intuitively signifies stronger values and highlights the higher covariance. This gradient smoothly transitions across continuous covariance values. This color setting improves readability and comprehension, helping users quickly spot areas of strong variable relationships. The figure size was increased to (10,10) in order to get more visibility into each feature correlation as it is harder to see with a smaller figure size. From our heatmap plot it can be concluded that hypertension, heart disease, and age are variables that are most correlated with stroke.

***2: D3 Scatterplot (Static Visualization #2)***

This D3 scatterplot displays the relationship between two imperative variables, which are glucose levels and age. Using this design helps us find trends that can be seen, such as individuals who have had a stroke generally have higher glucose levels, with a noticeable concentration of stroke cases above the 150 mg/dL glucose level. Additionally, it helps us see outliers - there were only 2 individuals below the age of 20 who suffered from a stroke, and they both had below 80 mg/dL glucose levels. Moreover, color was an important factor in this plot because there is a differentiator between stroke (dark blue) and no stroke (light pink), making it easy for the user to read. Lastly, by locating the data points of stroke patients front, we allowed users to see the relationship of two variables for stroke patients more easily.

***3: Altair plot consists of histograms and bar plots (Interactive Visualization #1)***

Mark is represented by a bar, and the length of the line is the channel. We use a histogram for numerical/continuous variables and a bar chart for categorical variables. Since the data values are not evenly distributed across the variable's unique values, we used the percentage of stroke occurrences instead of the count. Also, because we are using percentages, which total 100, we included only the percentage of stroke patients to mitigate confusion. For interactivity, we added a tooltip that displays the exact percentage value when you hover the mouse over a bar.

***4: Altair Scatterplot (Interactive Visualization #2)***

To effectively display the relationship between numerical and categorical variables with stroke occurrence, we utilized a dual visualization approach. On the left, we used a scatterplot to illustrate the relationship between age and BMI. The data points that represent patients who had strokes are highlighted in a warm-tone orange, making them easily distinguishable. In contrast, non-stroke patients' data points are rendered in a cool-tone blue and made more transparent to ensure that the stroke patients' data stands out.

Additionally, we enhanced interactivity: selecting a region on the scatterplot triggers the bar chart on the right to update. This chart displays the count of patients by different smoking statuses, where the color of each bar indicates the average glucose level of the data represented. This interactive feature allows users to explore how different factors correlate with stroke occurrence in a dynamic and informative way.

***5: Altair Boxplot (Interactive Visualization #3)***

The boxplot design features three plots side-by-side, allowing for an effective comparison of different variables at a glance. This enables users to simultaneously assess the distribution of multiple variables (age, average glucose level, and BMI) for non-stroke and stroke patients. Each boxplot is designed to show the side-by-side boxplot, in which the bars represent the non-stroke and stroke patients. The interactive feature which shows the detailed statistics such as the median, Q1, Q3, minimum, and maximum for each bar makes it easier for users to identify key statistical trends and outliers directly. The color is set by default.