DS4200 Final Project

Sarah Hudson, Gavin Kornitsky, Massimo Prag, Katrina Shonka

**Link to GitHub:** [**https://github.com/katrinashonka/DS4200FinalProject**](https://github.com/katrinashonka/DS4200FinalProject)

**Link to website:** [**https://katrinashonka.github.io/DS4200FinalProject/**](https://katrinashonka.github.io/DS4200FinalProject/)

**Design Idea Paragraphs**

**Visualization 1: Age Histogram with Gender Filter (made with D3)**

This visualization was designed to answer the high-level task: who is most affected by panic attacks? Creating a histogram allows viewers to explore the demographics (age and gender) of the patients included in our dataset to see broad trends in the prevalence of panic attacks. As with most histograms, the marks used were area (rectangular bars) and lines (for the density line), and the channels used were x position (the order of bins on the x axis) and y position (the height of the top of each bar and line). The bars were colored light blue to match the theme of our website, and the density line was colored dark blue to highlight the differences in the trends in the distribution. Our findings in this histogram help viewers understand which demographics the rest of the insights from our dataset (such as symptoms and risk factors) are most applicable to.

**Visualization 2: Tree Map of Symptoms and Lifestyle Factors by Panic Attack Severity**

This design gives us our highest level view of the counts of panic severity for those that had certain reported lifestyle choices and their reported symptoms. We chose a treemap for this visualization since we wanted to demonstrate the containment of the proportions of high, medium, and low severity panic attacks within each symptom or lifestyle factor. The mark of this plot was area (the boxes), and the channels were size (representing the proportion of data fitting a given category) and color (representing low, medium, high in increasingly dark colors).

**Visualization 3: Histogram of Panic Duration Distribution by Trigger**

This interactive histogram displays the panic duration distribution and severity count based on a trigger. We used a drop-down to choose which trigger or all triggers, and the time is in minutes from about 4 minutes to an hour. This plot answers the high-level question of whether there is a difference in the distribution of the severity of panic attacks when they are caused by different triggers, such as PTSD, social anxiety, stress, and more. The mark of this visualization is area (bars), and the channels are position (with the x position according to each bin of duration minutes and the y position indicating the count for each group) and color (indicating low, medium, or high severity).

**Visualization 4: Feature Importance (Random Forest Bar Chart)**

This bar chart displays the top five most important features from the model trained on standardized numeric data. After removing non-binary participants and encoding categorical variables, the model was trained on a sample of the dataset to preserve the distribution of panic severity levels. The bar chart visualizes the computer feature importance from the trained model, with horizontal bars sorted by descending importance values. The x-axis represents the relative importance of each variable to predicting panic severity. Light blue bars align with our site's theme, and the horizontal layout improves readability. By leveraging machine learning, this visualization emphasizes which variables were most important in real world panic severity classification.

**Visualization 5: Stacked Bar Chart: Panic Severity by Heart Rate**

This plot is a static stacked bar chart, which helps the viewers answer the question of whether panic attack severity is worse for different heart rate ranges. The mark of this visualization was area (bars), and the channels used were position, size, color, and text. The x position is determined by the relevant heart rate range, which is listed from smallest to largest. The y position is determined by the height of each box and also by whether the category is low, medium, or high. The colors of the boxes were also selected to range from light to dark blue, representing low, medium, and high severity panic attacks. Finally, since stacked bar plots can be more difficult to interpret than side-by-side bar plots, the numerical value of each bar was typed in the center.

**Visualization 6: Interactive Scatterplot: Duration vs Heart Rate & Panic Attack Severity**

These visualizations were added to incorporate the feedback we received from our in-class presentation; because we initially had many bar plots, it was recommended that we also include a scatterplot with a zoom-in interactive option that generates a second plot. The first visualization aimed to investigate whether there was a relationship between heart rate and panic attack duration, and the second visualization aimed to show the relationship between duration and panic severity score. The mark used was points, and the channels were position (x and y coordinates based on the values of the each continuous variable being plotted).

**Visualization 7: Grouped Bar Chart: Therapy/Medication Impact**

This visualization was designed to answer the mid-level task: do patients using medical interventions such as therapy and medication have more or less severe panic attacks? By filtering the data to identify subgroups of patients (those with no interventions, those with therapy but not medication, those with medication but not therapy, and those with both interventions), the counts of panic attacks in different severity classifications were assessed. This grouped bar chart was sorted by patients with different therapy/medication combinations with bars for counts of low, medium, and high panic severity score categories in each group. The mark of the visualization was area (rectangular bars), and the channels were x position and y position. The groups were ordered from left to right on the x-axis by least to most interventions, and within each group, the bars were ordered from low to high severity. The y positions of the top of each bar indicated the counts within the intervention group. The bars were colored light blue to match the theme of our website.