



This visualization depicts the yield difference of barley across different sites and varieties in the years of 1931 and 1932. The x-axis in the graph captures the varieties of barley and the y-axis captures the sites in Minnesota where they were grown. The shade of the boxes in the matrix communicate the differences across the two years. As the legend shows, a shade of red implies a positive change, and a shade of blue implies a negative change. Overall, the goal of this visualization is to show the downward trend of yield amounts across the two years, as evident by the blueness of the graph; and draw attention to the exception of Morris where the yield amounts go up across all varieties, as evident by the redness of the row.

An important factor I considered in creating this visualization was that all varieties and sites in this data were paired with each other (there was a fully-connected relationship map). Thus, I concluded that a matrix form was the best fit for communicating these site-variety relationships in my visualization. I set the x and y positions in my visualization to these two nominal data types. Then the next step was figuring out how to show the downward trend in the numbers from the year 1931 to the year 1932 in this matrix. Instead of using one color, which would have not brought enough attention to Morris, I decided to use a color map that included two colors and their shades to construct my matrix. This allowed me to emphasize the exception, that is the site at Morris much more clearly, and it also allowed me to center the data around white (no shade of any color), which would imply a net difference of zero. In choosing the colors, I went with the hot-cold color map (red-blue), a familiar coloring scheme that I thought tapped into the learned association viewers have between hot-cold and positive-negative numbers. Thus, my visualization was now able to effectively communicate the differences with little extra effort (the legend).

An aspect that is obscured away in my visualization is the actual yield amounts in these two years. As much as my visualization does a good job at capturing the overall trend in yield across all varieties and sites, it fails to communicate a sense of how important these differences are in relation to the yield amounts. For example, if the yield amounts were well above 1000, then these differences might have minimal impact. On the flip side, if the yield amounts were much lower, than these differences would be colossal. In constructing this visualization, this was an aspect I chose to not communicate, for the sake of bringing attention to the overall trend.

In constructing this visualization, I used matplotlib, a python library that I often used in AI classes, useful for creating powerful static visualizations.