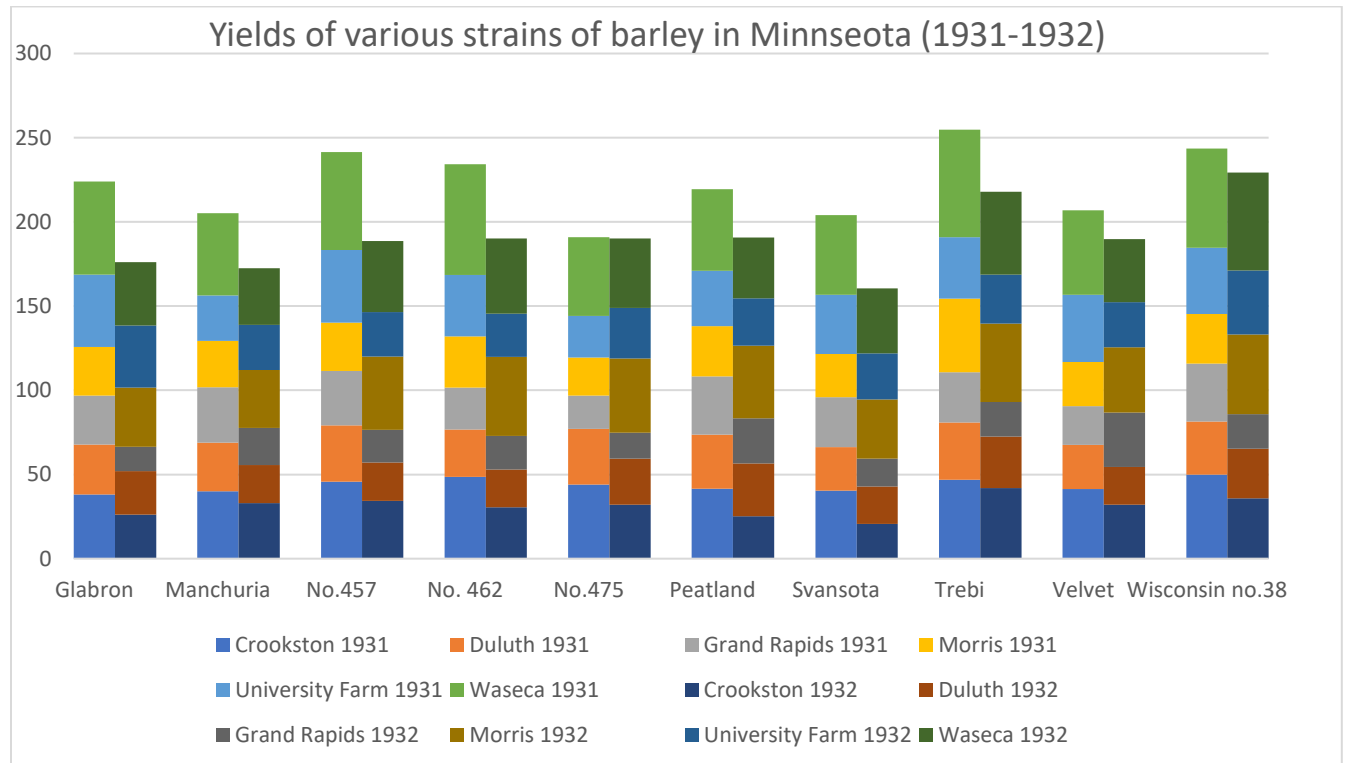


CS 448B assignment 1

Written by Sergio Sardar



My design process was driven by my intended target audience for the data set we were provided. Considering the vast variety of people who might be interested in trends that arise from the data, I decided to limit the target audience to people that are trying to assess the 'best' variety of Barley to grow in the wide, or more specific, Minnesota area. Such an audience would definitely be most interested in the relative yields of each strain of barley. Thus, I attempted to emphasize the relative yields of each strain of barley the most effectively. A stacked bar chart seemed to be best suited to visualize the data. In order to do so, I made the y-axis represent the numeric yield of barley in a given year. The x-axis is labelled with the various types of barley that were grown in the experimental period. Each strain of barley has a pair of bars associated with it, with the left bar in each pair representing the total yield in 1931 for that particular strain of barley. The right bar represents the total yield in 1932. Furthermore, each bar is divided by sections of color, with each section of color representing the location the

barley was grown in. Intuitively, the length of each section of color represents the yield of barley cultivated at the corresponding location. I defend my choices below.

Justification

As mentioned above, I designed this visualization for people that are trying to assess the most appropriate variety of Barley to grow in the Minnesota area. The relative yields of each strain of barley are obviously crucial in choosing the appropriate strain of barley. This is the aspect of the data I tried the hardest to effectively communicate. According to Bertin's "levels of organization", length is the best visual variable to depict this quantitative data. The relative lengths of the bars allow the user to quickly estimate the relative yield of each strain of barley. Length intuitively corresponds with the "size" of data, so naturally readers figure that the length of each bar represents the yield of each strain. In light of this, a bar chart seemed to be the best choice to visualize the data. The strains of barley are nominal, independent variables so it made sense to use labels for them on the x-axis. I decided to separate the yields for 1931 and 1932 by depicting two bars for each strain of barley. The ability to compare yields for a specific year is valuable if the reader wants to account for variables unique to each year that was not expressed in the data. For instance, if the reader knew that 1932 had significantly less rainfall than 1931, and Minnesota is expected to experience severe drought in the next few years, it would be more apt to consider the bars for 1932 and disregard those for 1931. This potentially crucial information would've been lost had I decided to add the yields for both years per strain together.

I decided to represent the fourth dimension of the provided data set, the locations the barley were grown at, using colors. I chose color as the visual variable to represent this since it works well for nominal data such as location names. It was worthwhile to depict this dimension since readers can account for conditions unique to each location while choosing an appropriate strain of barley to grow in the future. The amount of colors used might make it more difficult for users to analyze yields at individual locations. However, since this is a rather fine grain aspect of the data to analyze (most people would probably be interested in the total yields per year of each strain), the difficulty with deciphering the color coding scheme is worth the struggle for the information gained in return. I decided to shade the bars for 1931 slightly darker than the bars for 1932 to make the two clearly stand out from each other, and to make it clear that each bar in a pair represented a different year's yield.

I used Microsoft excel in order to make this graph.