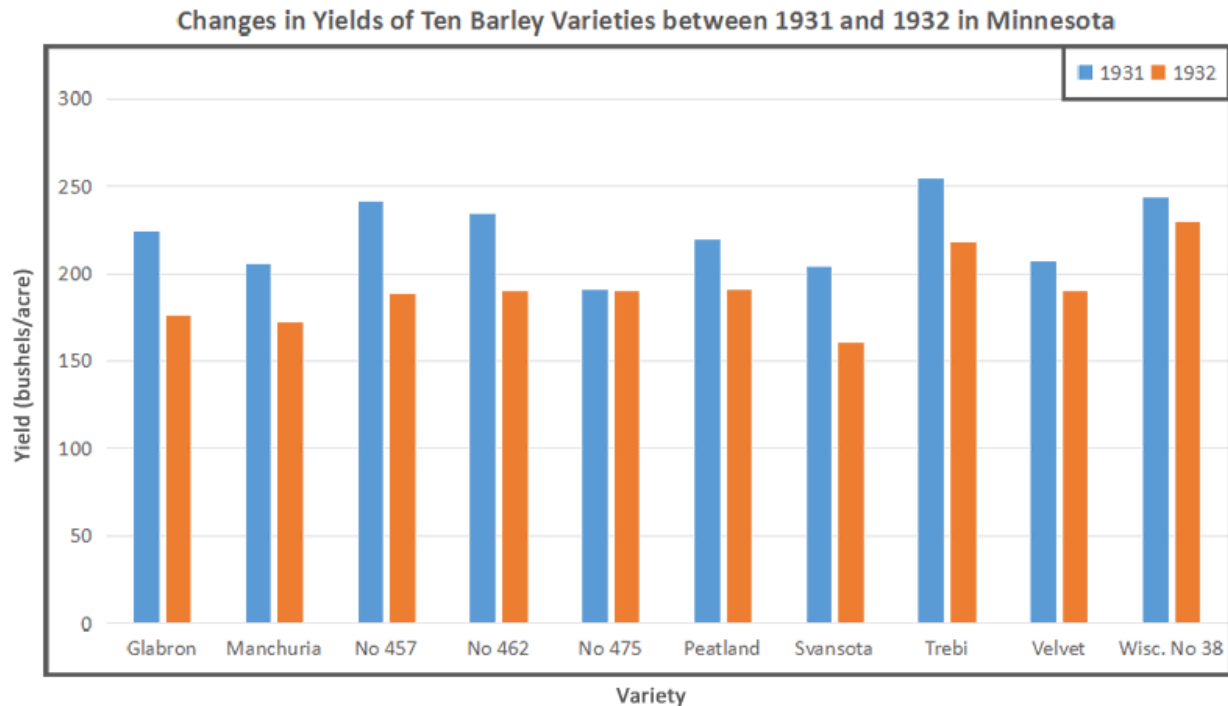


CS 448B Assignment 1: Visualization Design

Shan Jinglin

1 Graph



2 Description

On reading the data set, I envisioned my graph to communicate a story mainly using data in *Yield* variable. I personally found variety to be more interesting than experimental stations to become a feature that relates single records, and thus, I downplayed *sites* variable by summing up the barley yield of each variety in 1931 and 1932 respectively across all sites (using Excel). I decided to show the change in yields of different barley varieties in Minnesota in one year (1931-1932) to our readers with this graph. Given that our graph mainly compares yield change of each variety, it is the best to group data of one variety together, and in each variety group show yield number in 1931 and 1932.

Therefore, the graph has to visually encode such “grouping” relationship. Noticing that *Variety* is nominal, when we look up the Mackinlay’s ranking, we can find that the top two visual variables for encoding a nominal variable are *position* and *hue*. Regarding the purpose of this graph, which is to present “change in yields” while “variety” is one criteria to relate data, I found it unnecessary to give each variety a distinct hue. Instead I employed position to set apart varieties. As to *Yield*, which is quantitative, I used *length* rather than *position* because bars are easier to give readers a visual impact (larger area) than small dots. Colored area looks fulfilling, and it is better at giving readers an impression of “amount” or “yield”. To

faithfully represent data, all bars share the same width, while their heights are proportion to their own quantities.

Indeed to show “change”, a line connecting data points is an option. However, since there is a discontinuity between varieties and in each variety there are only two data points, two visually distinct bars are enough for readers to notice a difference in height. To create a dramatic comparison between two bars in each variety, I used blue and orange to color bars representing data in 1931 and 1932 respectively. These two colors form strong contrast with each other so that it is easier for readers to distinguish two bars, and notice the a decrease in length across all varieties. If our readers take a closer look at the graph, they may also get an idea of how significant each variety’s decrease is.

In the end, I added thick black borders using Photoshop to highlight the legend (blue stands for 1931, orange stands for 1932) and the variable names on the axis. The thick black border excludes elements that are not data of interest from main graphing area that we want our readers to focus on. In addition, the thick black border around the legend makes the legend more noticeable so that it is quicker for readers to learn the meaning of colors in this graph.