The Morris Mystery: Change in Barley Yield between 1931 and 1932

Sophia Pink
October 2nd, 2017
Assignment #1
CS 448B
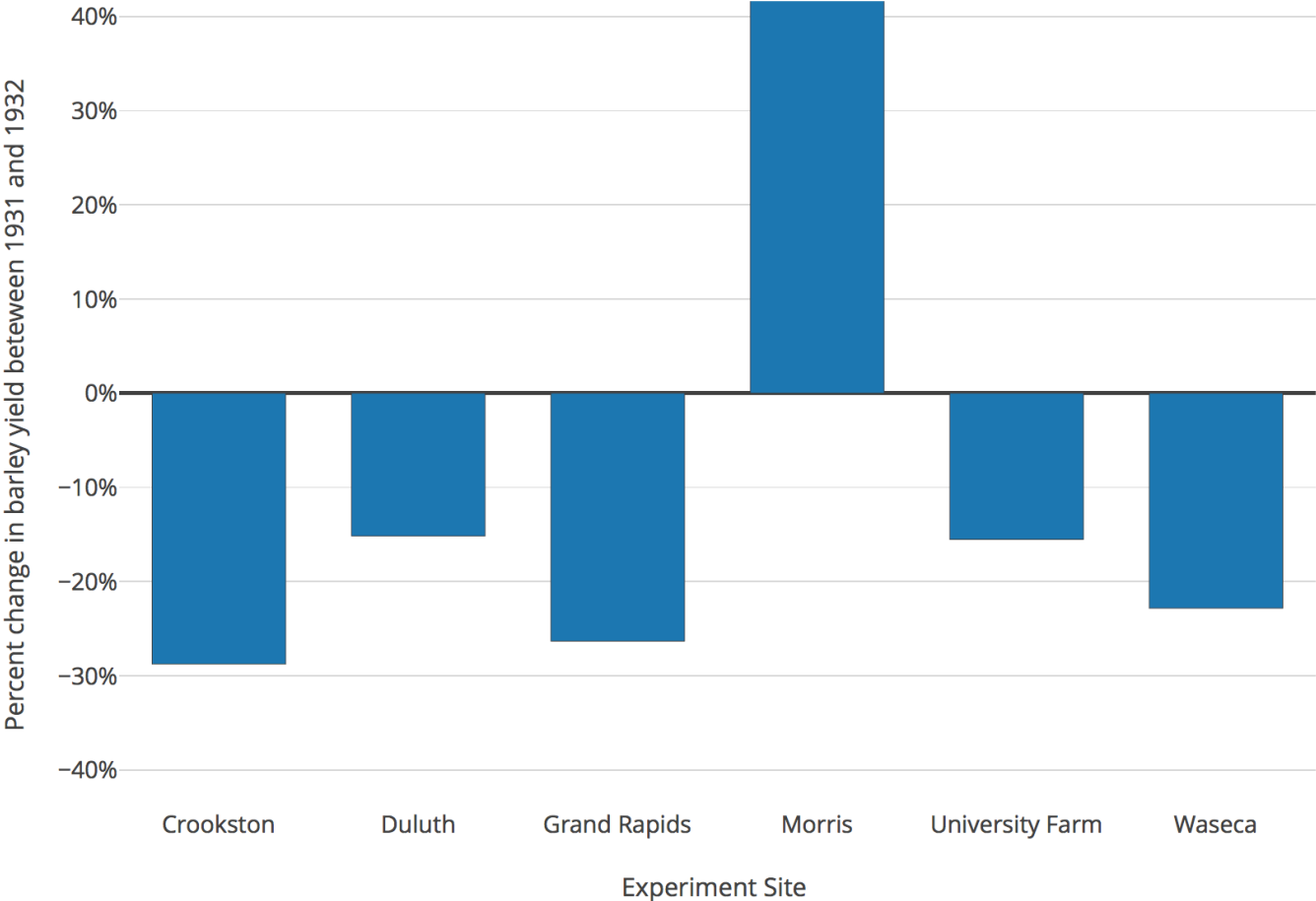
The Morris Mystery: Design Rationale

I designed this graphic for a popular audience who is reading an article about barley. I did this to force myself to communicate a specific point to people who might not be looking at the graphic for a long time. This makes is different from a graphic for barley experts, who may want more details, or for an academic audience, who may want indicators of variance or statistical significance.

I wanted to communicate clearly and quickly that something odd happened in Morris between 1991 and 1932. Although the yield at five sites decreased, the yield at Morris increased. I did not represent the varieties of barley in the graphic because there weren't any large differences, and I thought that adding that extra variable would confuse the viewer rather then clarify. However, this means that the viewer does not know that there were multiple varieties of barley from looking at the graphic.

I represented two variables in the image: site and percent change in yield between 1931 and 1932. I found the second variable by calculating the average percent change in yield at each test site across all the varieties. I used percent change instead of absolute change because viewers likely do not have a sense of the magnitude of one bushel/acre. For example, most people don't know what a decrease of 12 bushels/acre means, but anyone can understand a 30% decrease in yield. One issue with percent change is that is does not show the original relative differences between the yields – for example, it could be that Morris had an extremely low yield in 1931, and is still low despite the increase. However, all the sites had yields within the same order of magnitude, so I did not think that this transformation was deceiving the viewer.

I decided to use a bar chart for three main reasons. First, a bar chart uses position to encode sites, a nominal variable, and uses position and length to encode percent change, a quantitative variable. According to Mackinlay, these are among the most effective encodings for this type of data. The second reason I used a bar chart is that it is easy to compare the differences between the sizes of the bars. There are only six sites, so there is space for a viewer to look at all the bars and compare the differences. Third, bar charts have a clear anchoring at 0%, which emphasizes the difference between an increase in yield and a decrease in yield. I made the horizontal axis thicker than the other horizontal lines to draw the viewer's eyes toward that line. I also set up the y-axis to go from -40% to 40%, so that the 0% line was directly in the middle of the graph, highlighting the turning point even more.

I kept the color of all the bars constant to not confuse the viewer with an additional encoding. I also listed the sites in alphabetical order so as not impose any other order on the nominal variable. Although this chart does not show all facets of the data, I think that these design decisions make it easy for a reader who knows nothing about barley to understand that something unusual happened to barley growth in Morris between 1931 and 1932.

*Tools:*

I manipulated the data in Microsoft Excel, and I used Excel to create quick drafts of different types of charts. Once I decided on what I wanted the visualization to look like, I used Plotly (the web app version) to create the final version.