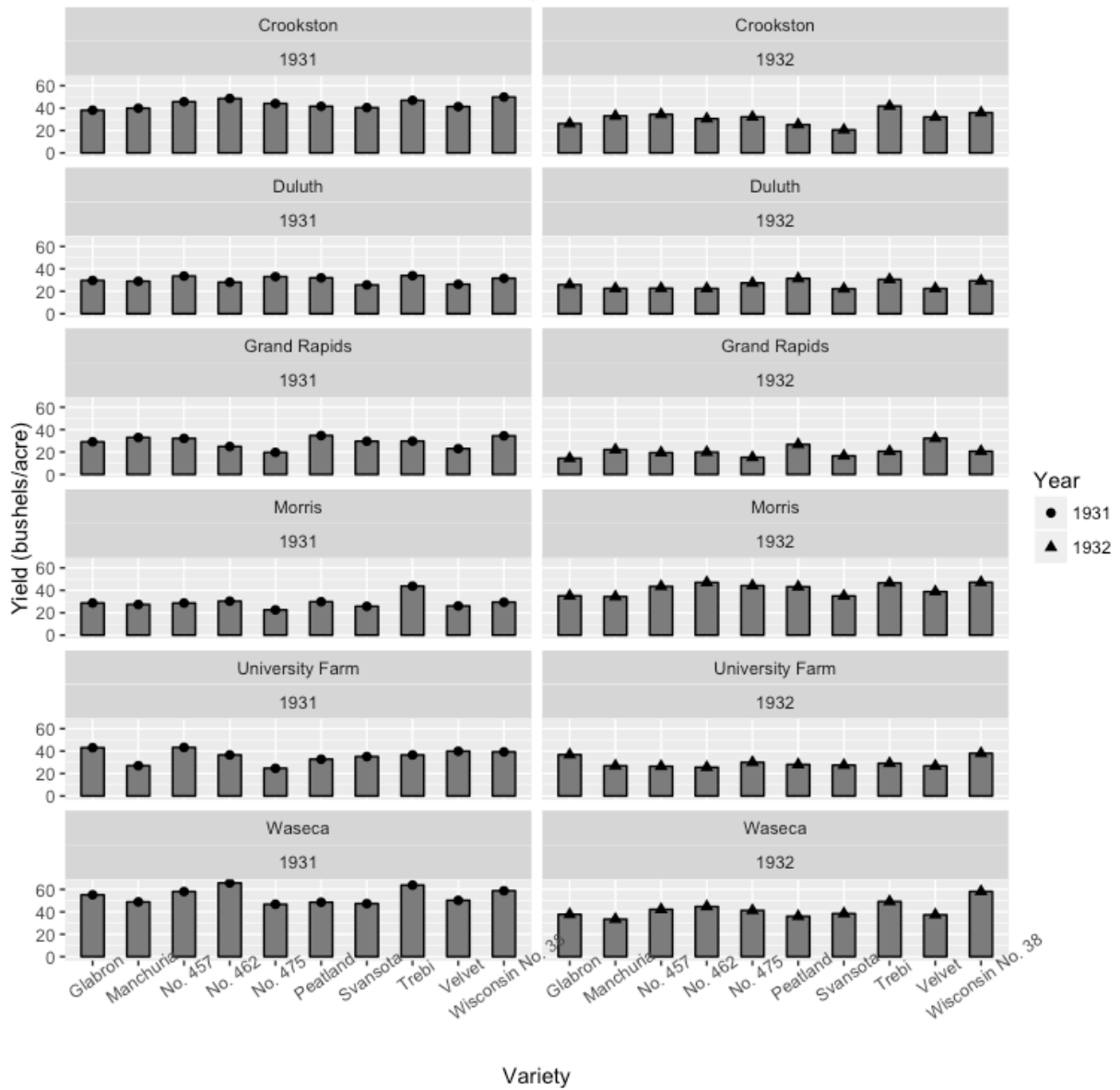


Kali Cornn  
 CS 448B – Assignment 1

Barley Data



### **Overview/Goals**

In exploring the dataset, my goal was to show the differences in barley yield between 1931 and 1932. To do so, I created bar graphs, grouped by site, with one bar for each variety of barley. I also had two columns in my visualization, one (left column) for 1931 data and one (right column) for 1932 data, such that each row in the visualization corresponded to a different site (Crookston, Duluth, etc.) In this way, the reader can look at one row at a time to see differences in yield between the two years. In each bar graph, the x-axis represented the variety of barley, and the y-axis represented barley yield (in bushels per acre).

### **Design Decisions**

I did not transform the data because I wanted to represent the barley yield as is. There was no need to remove any of the individual data records because there were no missing records or record values. I also chose to keep all variables in my visualization, since there were not many variables in the first place. Since the data are not unordered, I used one color to fill in the bars – rather than different-colored hues.

In creating the bar charts, I made sure to maintain the same y-scale and the same breaks on the y-axis breaks, in order to ensure that data was represented the same way across the six different barley sites.

### **Obscured Data**

From the visualization alone, it is difficult to determine the exact percent change in barley yield for each site and variety between 1931 and 1932. However, determining relative positive or negative change can be easily determined by looking at each row of the visualization.

### **Technology Used**

I used R to create my plots; specifically, I used the ggplot2 library to develop the visualization. I used such ggplot2 functions as `facet_wrap()` to wrap the data into two columns and `theme()` to angle the x-axis labels of the barley variety.