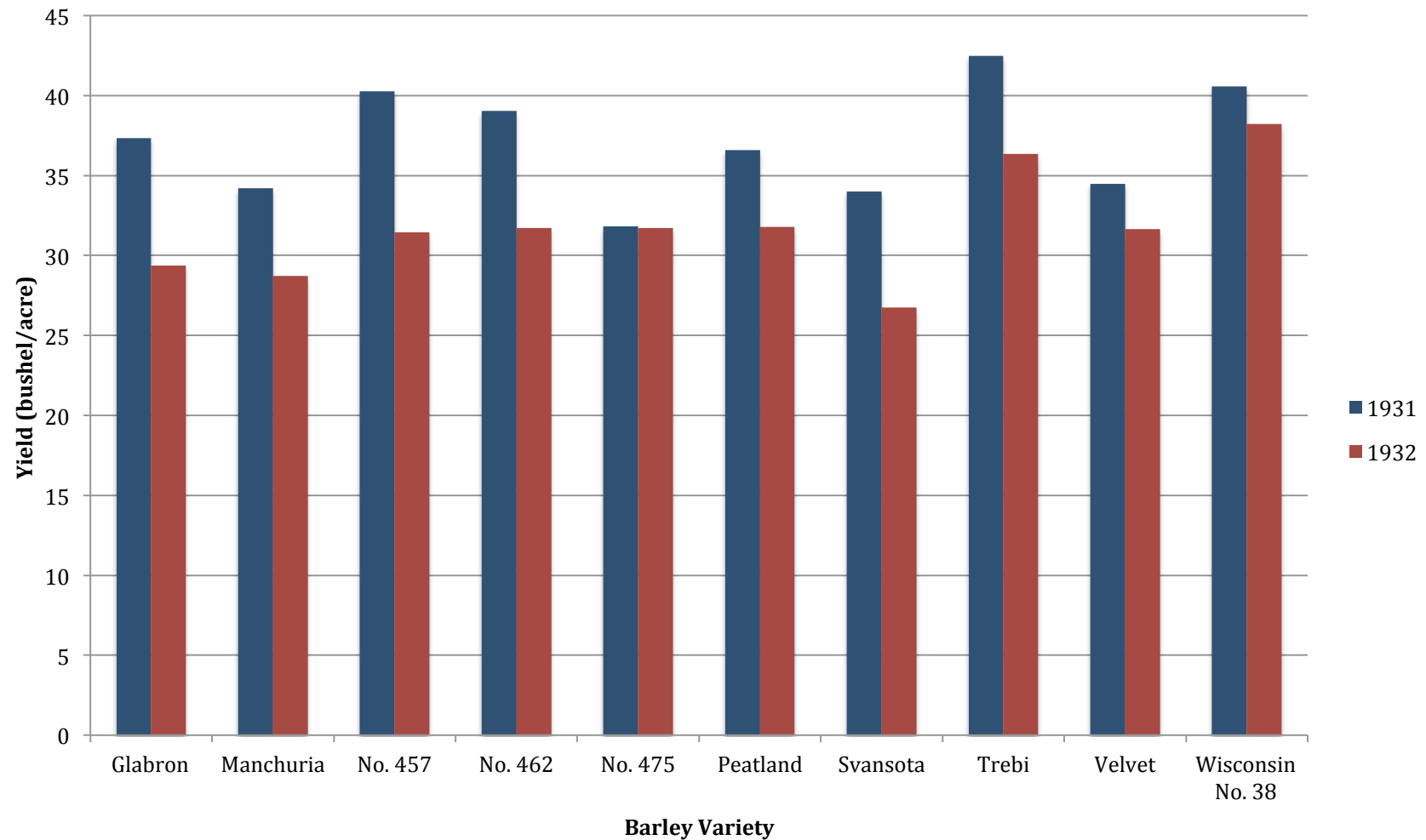


## Change in Barley Production from 1931 - 1932



After several iterations, I have decided that the graph above presents the data set in the most intuitive and digestible manner. In order to create the most meaningful visualization, we first have to decide which variables are the most pertinent in formulating a relationship from the data. In my eyes, the data set is first and foremost, relating the variety of barley to the change in yield per year. Therefore, I filtered and coalesced the data set to make this relationship as obvious as possible and the main focus of the visualization.

First, I decided to filter out the “sites” variable by simply plotting the median yield for each strain, rather than the yield for each strain at every site. I did this because I believe the sites variable exists only to gather the most randomized and unbiased data set. It exists to gather the most realistic data on barley strain growth rather than as a standalone variable. I decided to not obfuscate the graph more by presenting the “sites” data and rather chose to focus on presenting its affect on the main relationship: barley variety vs. yield.

I also decided to do a clustered column graph for each year because all strains declined in growth from 1931 to 1932. There is also a reason that both 1931 and 1932 growth rates were included in this data set. The most fertile strain changed from 1931 to 1932, which would not have been obvious had I left out one of the years or amalgamated the data into a single column graph.

For these reasons, I designed the data visualization shown above. Overall, I wanted to make the relationship between variety of barley and change in yield exceedingly obvious and the product of this data set.