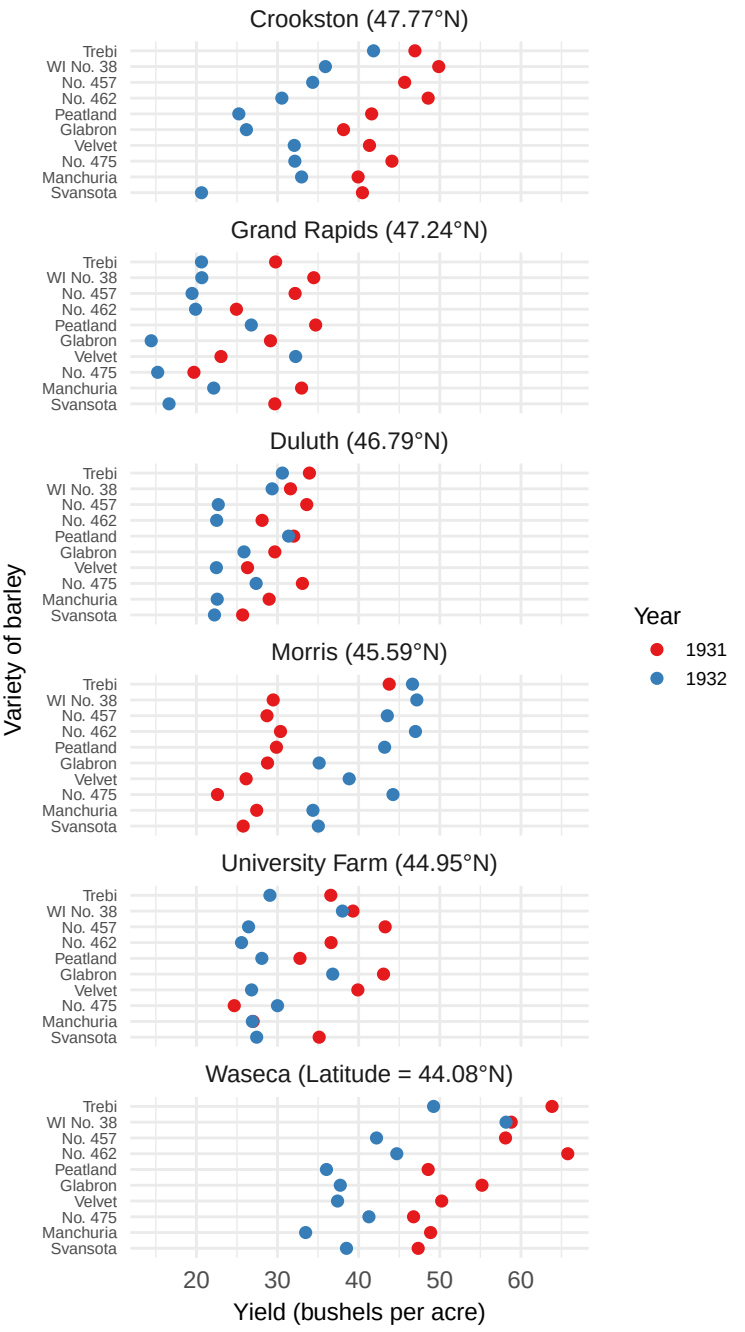
Barley yield by variety and location



Latitude data from latlong.net

Tools used

I used the ggplot2 package in R to create this visualization.

Write-up

The following describes the visual encodings used in my visualization:

- Yield is represented on a lateral scale on the x-axis. I chose to put it on the x-axis so that variety could go on the y-axis, which would make the labels for the varieties easier to read. I also put it on the x-axis so that the viewer can make horizontal comparisons between the different locations.
- Variety of barley is represented on the y-axis. Again, this makes the labels easier to read. It also allows the viewer to compare across different barley types.
- Variety is also sorted by total yield (i.e., the variety with highest yield across all towns and years is at the top). I did this so that it is easier to see which dots correspond to which variety. Without sorting, they were arranged randomly, and it was more difficult to tell the dots apart. This also allows the viewer to more easily see that there are some barley types (e.g., Trebi) that tend to do pretty well no matter the location (with some exceptions).
- The facets encode location. This is so that the viewer can look at the trends in yield for the different locations easily. The facets are arranged vertically so that it is easier to compare yields, as mentioned previously.
- The facets are also arranged by decreasing latitude. This is so the viewer can see the effect of latitude on yield. I put the highest latitude at the top because we are used to thinking of north as up and south as down.
- Color encodes year. Year is a nominal variable here (I would argue that order doesn't matter for the purposes of this plot), so color works well. There are only two years, so the different colors do not become overwhelming.