



The data set we were given consisted of two nominal variables (category and variety), a quantitative variable (yield) and an ordinal variable (year). I chose to base my visualization off a stacked and clustered bar chart because they make it easy to directly compare different categories of nominal data against a quantitative variable. As the dataset was trivariate, I also needed the corresponding visualization to have three dimensions. In this case, the x-axis corresponds to site groups (further distinguished by variety), the y-axis corresponds to yield, and the different bars correspond to the year of that yield. Displaying the data in this manner helps to visualize a general pattern: across the board, most sites saw a dip in yield in 1932 when compared to 1931.

I made a few particular design choices when considering the choice of visualization: first, I noted that every combination of category and variety had a separate entry in the dataset for 1931 and 1932, and so I decided to combine the yield and year fields into the Yield (in 1931) and Yield (in 1932) fields. This allowed me to stack the bars on top of each other and allow for direct comparison of yield against year at a glance.

I chose to eschew visual encodings like color, thickness, size, etc. as a means of distinguishing between different data, as I thought that having that many different sites/varieties of barley would require a lot of different visual variations, which would be confusing to follow. Instead, I split the data into groups of sites, where each group is comprised of all the varieties of barley grown at that site. This created a hierarchy within the data, with site being the most important variable, as I felt that changes in site would have a greater difference than changes in variety, and that there were likely patterns within the yield to be found at varieties of barley grown at the same site. It also allowed the dataset to fit within two axes, as I combined site and variety into a single axis.