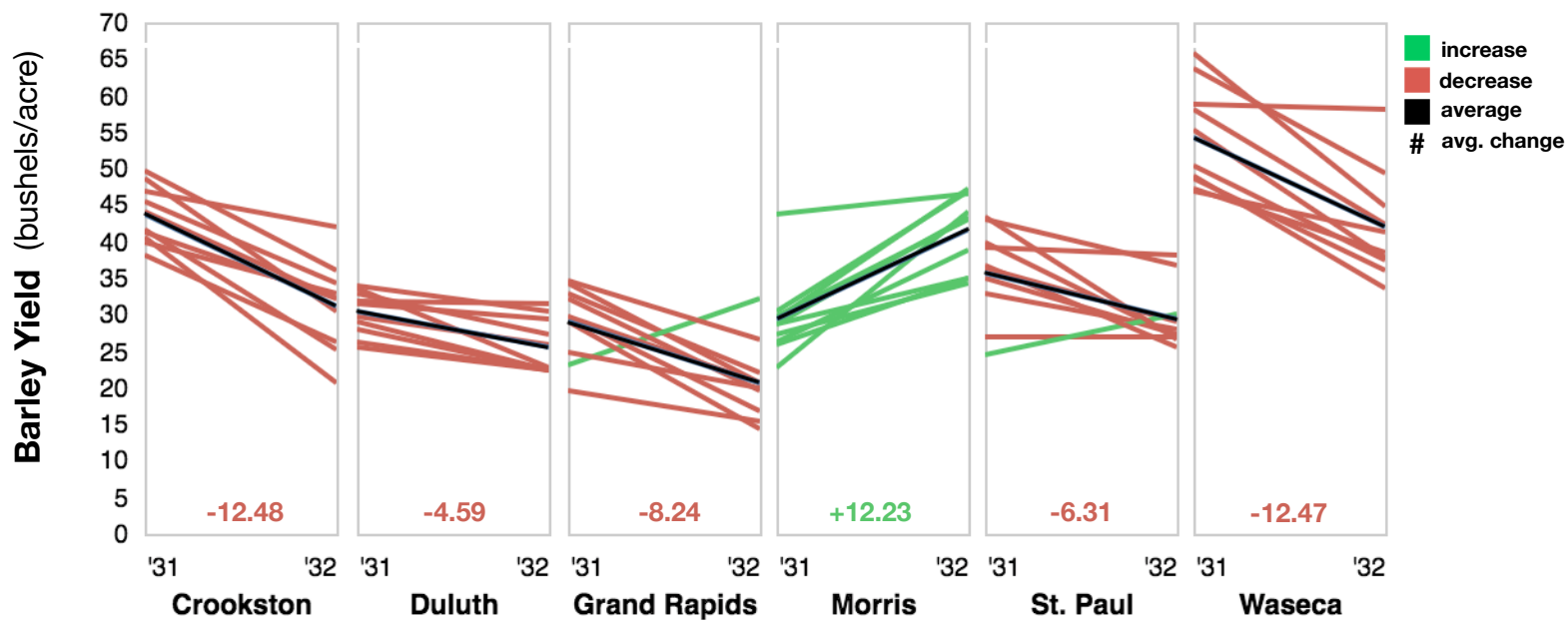


Crop Yield of Ten Varieties of Barley in Six Minnesota Towns, 1931-1932



For my visualization, I created a slope graph of the change in yield between 1931 and 1932 of each variety of barley. The graph is broken down into six plots by location, and colored by trend (either increase in yield or decrease). I included a summary statistic in two forms: a line on each plot showing the average change across all varieties, and the corresponding numerical value of average change. According to Tufte, one of the main functions of visualization is to encourage the eye to compare data, a principle that guided many of the decisions I made when creating my graph. My goal in creating this visualization was to illustrate an interesting feature of the data—that every town had lower barley yields in 1932 than in 1931 except for one exception, the town of Morris, which had a clear increase in barley yield.

I chose a slope graph because it clearly shows change in yield over time, as position is the clearest indicator of quantitative data (Bertin 1967). I chose to put location on the x-axis because I thought location provided the most interesting variation when compared to yield. Because I encoded this data in a simple way, I could use other types of encoding to illustrate trends. In order to direct the viewer's attention to the case of Morris, I colored each line based on its slope (either positive or negative). This way, instead of seeing dozens of lines all with small variations in slope, the viewer sees most lines in red but a clustering of green lines for Morris, and notices that Morris has a trend different from all other towns. I could have chosen a gradient to more clearly indicate the slope of each line, but displaying the exact slope of each line is not the purpose of the plots. Instead, red versus green is a dichotomy that viewers know intuitively, and it signals the difference in Morris much more clearly.

I also included some other statistics on the graph. The presence of these numbers adds concreteness to the trend I displayed—instead of learning that yield in some towns decreased because most of the lines are red and slope downwards at different degrees, the viewer has more concrete knowledge of the trend because it is summarized (averaged) in the black line. This line helps bring focus to the chart, and helps support the notion that the important message in this data is how the data trends differently in Morris than the other towns, and that the exact change in yield of each variety is not particularly important. Despite this, I thought it was still important to include a line for each variety, instead of just displaying the average slopes on a single plot. This is in order to provide context and credibility for the Morris trend, to show that the increase in average yield in Morris is not the result of one abnormally high value, but the result of every single variety increasing in yield. I also added another summary statistic—the numbers at the bottom of each plot representing the average increase in yield. It is somewhat difficult for the viewer to visually compare the increases and decreases across towns because the x-axes are side by side, so having the numbers at the bottom of each plot can help to relate the plots to each other.

Additionally, there were a few parts of my visualization that may have been poor graphical practice, but I felt were necessary. Firstly, the broken x-axis may be confusing because it requires viewers to jump back and forth in time as they move across a visualization. It may even lead viewers to think there is an ordinal trend to the towns. However, I decided this was necessary for a few reasons. I could have kept both axes consistent and plotted all of the data on one large plot, but that would have caused too much clutter for the viewer to discern the clear meaning of the data, and thus seemed a worse option. I also could have kept the x-axis consistent and not the y-axis, but since the x-axis is only two values, it is much simpler for the viewer to recreate the simple two-value axis than the one with a quantitative range. I

also did not label each variety of barley—this is because the differences among individual varieties are not the focus of the visualization; rather, the sum of all of the data illustrates the trend.

For this visualization, I used vega lite for plotting and Google Sheets to perform the simple data analysis calculations (such as averages), both of which worked well on this small dataset.