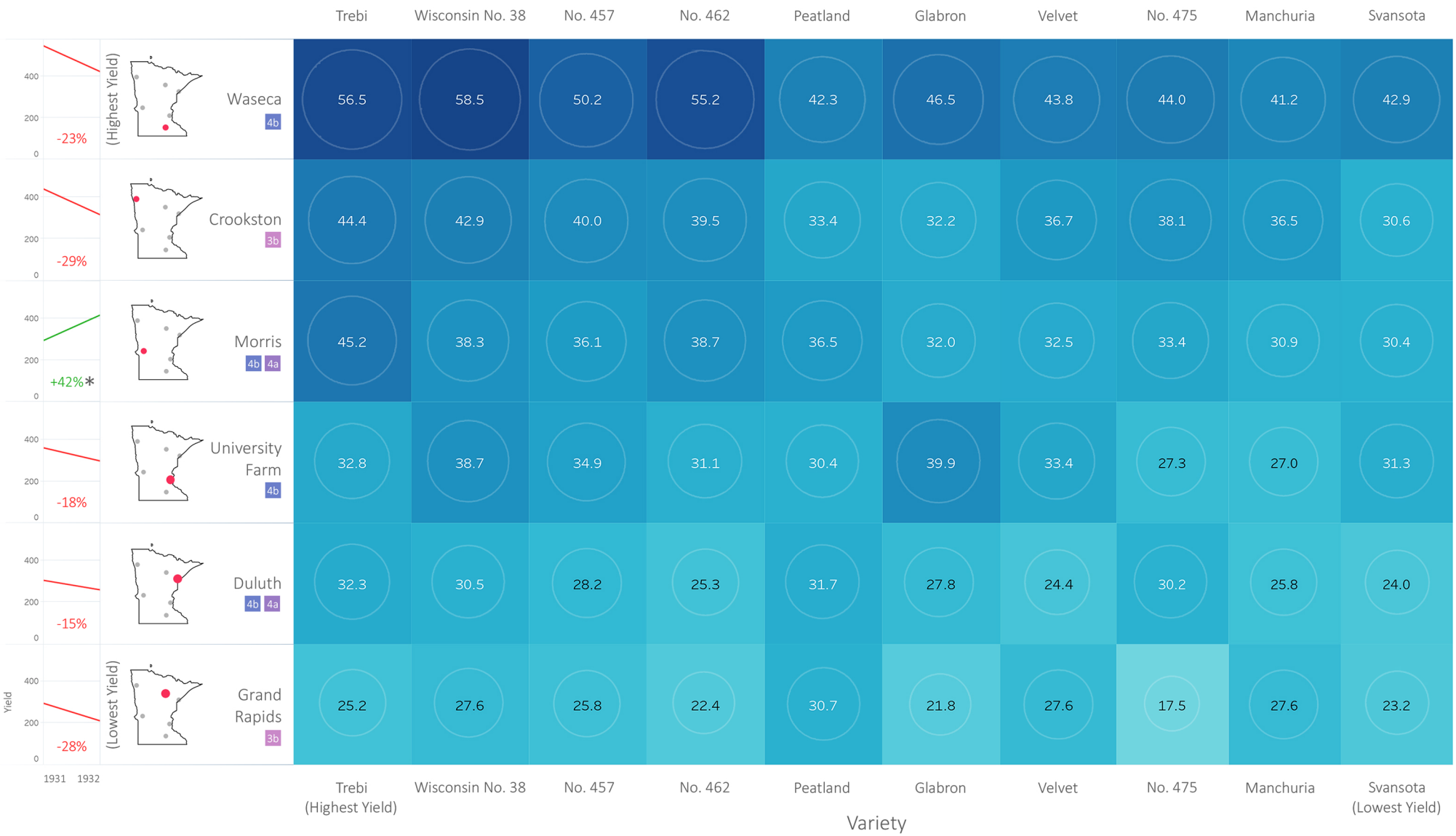
Average Barley Bushel Yield per Acre for 1931-1932

Broken down by Variety vs. Site Location. Color shows average yield per acre. The colored cells are labeled by average yield per acre.



*(The site in Morris appears to have enjoyed a 42% increase in yield from 1931 to 1932. It's possible that the yields recorded at the Morris site in 1931 and 1932 were inverted.)

Jack Reidy - CS 448B Assignment 1

Jack Reidy
CS 448B - Assignment 1

This visualization is designed for practical application of the data set. The visualization facilitates direct reference of site specific variety performance. To that end, the visualization presents average yield for the time period. The highlight table is a functional model for quickly identifying the highest performing varieties of barley for a given location. Additionally, the size, value, and color encoding of the highlight table allows the viewer to read the visualization at more than one perceptual resolution. A farmer considering which barley variety to plant is most interested in data relevant to the location of their farm. With this table, the farmer is able to trace horizontally from the nearest site location to the highest performing strain of barley. The farmer can also quickly determine the relative performance of any site-variety pair against another variety or against another location.

Average yield for each site-variety pair is represented by a cell in the highlight table. The table is designed for reference of quantitative data at multiple levels- from quick glance to close inspection. The visualization enables multi-level reference by visualizing average yield in table cells encoded with value (raw average number), color (blue-white gradient), and size (area ring). Value allows the viewer to calculate the exact difference between cells. Color allows the viewer to interpret the data more generally. Size reinforces the comparative difference in quantities.

Cells are colored according to a blue gradient scale with white corresponding to zero and dark blue corresponding to the cell with highest average yield. The visualization does not offer a color legend because a gradient does not effectively communicate precise yield value. Bertin suggests that color is not an appropriate encoding model for quantitative data, but the highlight table demonstrates that color is functional in a subset of visualizations.

The highlight table is sorted vertically and horizontally. In the vertical, rows are sorted in descending order by the cumulative average yield for each site location. In the horizontal, columns are sorted in left-to-right order by the cumulative average yield for each barley variety. This arrangement is the most informative ordering. Variety labels are displayed above and below the table for convenience. The viewer can more easily identify a cell when labels are in close proximity. Variety labels are centered with each corresponding column for readability. Site location labels are aligned middle-right for readability.

The visualization is augmented by two secondary representations. First, site locations are mapped and labeled with USDA plant hardiness zones because this information would be relevant in practical use. Coordinates for each location were added to the data set. Second, year-to-year yield line graphs are included because the highlight table cannot communicate time. As noted in the visualization, it appears that yearly results recorded for the Morris site were probably inverted. Presented differently, the bad data from the Morris site would be more obvious.

This visualization, which presents average site yield, cannot capture every narrative embedded in the data. Because the highlight table does not plot results for each site-variety-year triple, the visualization cannot communicate the raw data directly. The highlight table is not the best model for precise

comparison in part because data points cannot be encoded by position. The visualization contains a challenging amount of information. Admittedly, it is not the most readable representation. Given that the objective of this visualization is to communicate practical information from the data, the complexity is (hopefully) justified.

The visualization was drafted in Tableau and formatted in Photoshop.