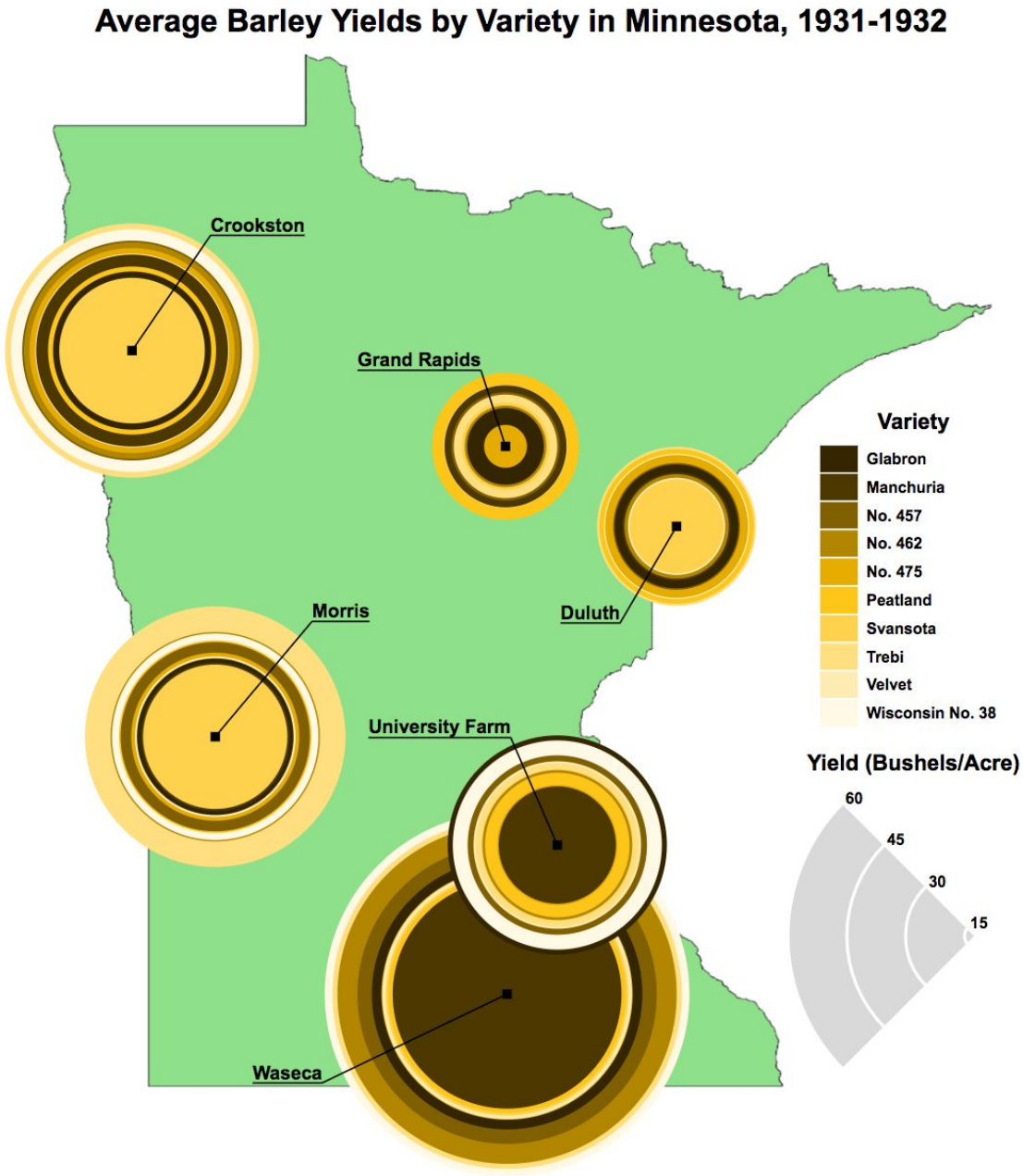
Greg Ramel (gramel)

CS448B Assignment 1 Writeup

Visualization



Rationale

The provided dataset captured four variables: site, variety, yield, and year. My first decision was to determine which of those variables most needed to be visualized, and which could be massaged. As there were only two years, my first reaction was to not do any sort of time series. I did have a brief thought that it could be showing downturn during dust bowl droughts, but the main droughts didn't start until three years later (1934). I also wrote a script to look at changes in variety yield and site yield year-to-year (controlling for the other of the two variables) and didn't find any significant or interesting decline. So, I decided to collapse the dataset into an average yield .json file over two years with a Python script, and focus on the more variable site and variety yields. The visualization itself is done entirely in the Raphael JS JavaScript illustration library, rendering SVG elements over a background image of Minnesota.

I was left with three variables over which to represent the two-year average: site location, variety, and yield. Location was the most important to visualize, as that seemed to be the most important to the original data. Given the geographic nature, I picked MacKinlay's most precise representation of a point to mark the cities. Since it was all in one geographic region (Minnesota), I centered each city on the appropriate location on a map of the state (I used Photoshop to download, erase the background, and find city locations by overlaying it with a Google Maps screenshot). I wanted to represent the yields of each variety at the specific city location, so I chose concentric circles whose radius represented yield. This fit with MacKinlay's assessment of area as a moderately accurate middle-ground representation. I did scale the data slightly to make the circles more visually distinct - the minimum yield was around 17 bushels per acre, so I applied a linear transformation of (x - 12) * 2.5 to scale accordingly without distorting the relative weight of the values.

Variety was then represented with the less precise color. The overall effect was to show how certain cities had notably larger yields (southern and river sites) as well as that certain varieties like Wisconsin no. 38 yielded much more consistently than something like Manchuria (though cities had different orderings in general). For color, I wanted to evoke agricultural themes - the variety colors were shades of grain-like amber. Yields are sorted to allow each to be rendered, largest to smallest, in front of the last. I colored Minnesota a pale green to continue that theme, and allow the circles to pop adequately. The reference for variety shades could fit in a nice chromatic legend, and I used quarter-circles to show the scale of yield without having the reference figure be too large. The shape of the state allowed for a good place for the legend on the right. I wanted also to keep the precise city location marked distinctly, so that became a small square, with the names outside the yields and connected to the icon. Aside from the state outline, I shot for a flat, clean design overall.