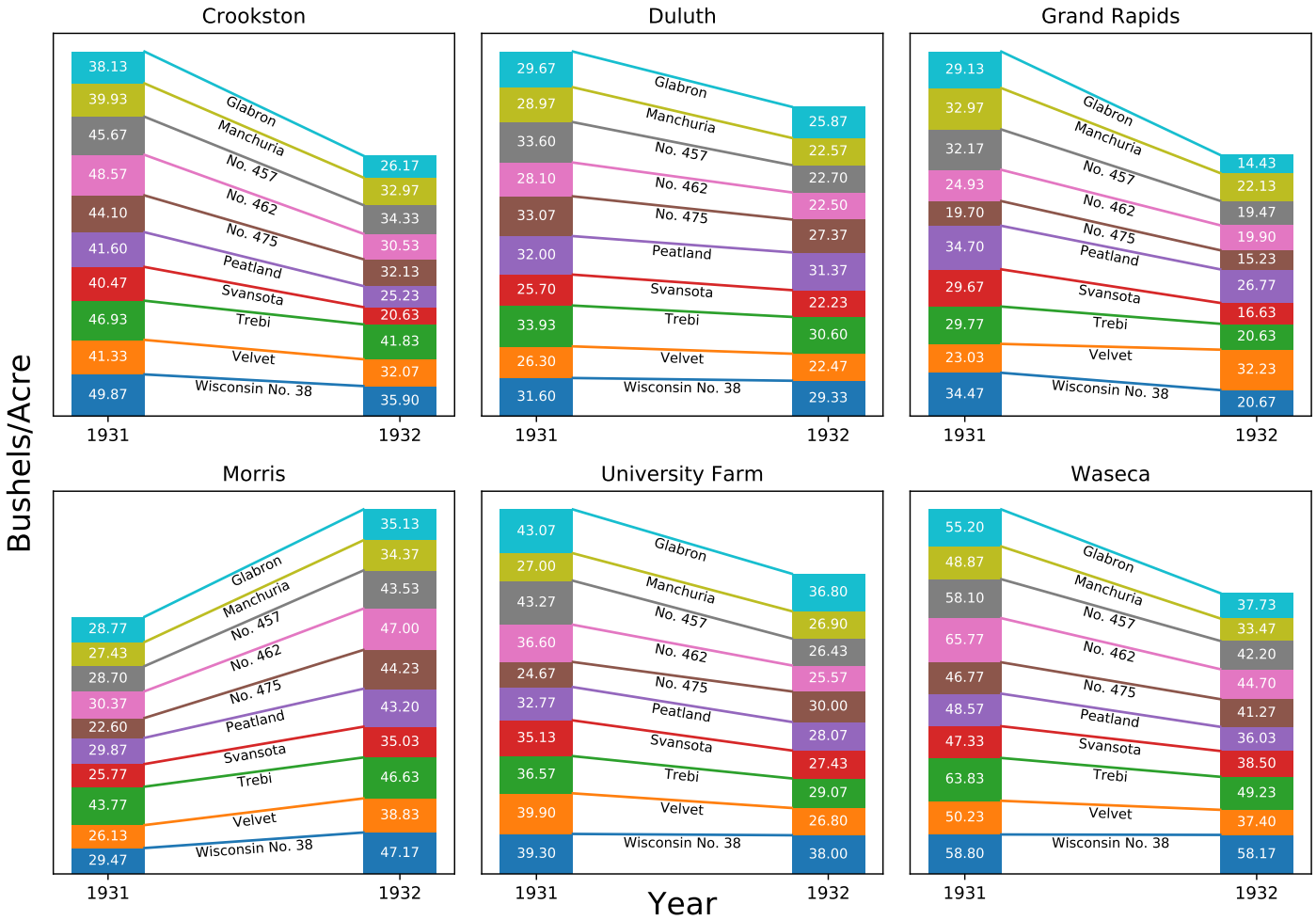
Crookston
38.13
39.93
45.67
48.57
44.10
41.60
40.47
46.93
41.33
49.87
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
26.17
32.97
34.33
30.53
32.13
25.23
20.63
41.83
32.07
35.90
1931
1932
Duluth
29.67
28.97
33.60
28.10
33.07
32.00
25.70
33.93
26.30
31.60
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
25.87
22.57
22.70
22.50
27.37
31.37
22.23
30.60
22.47
29.33
1931
1932
Grand Rapids
29.13
32.97
32.17
24.93
19.70
34.70
29.67
29.77
23.03
34.47
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
14.43
22.13
19.47
19.90
15.23
26.77
16.63
20.63
32.23
20.67
1931
1932
Morris
28.77
27.43
28.70
30.37
22.60
29.87
25.77
43.77
26.13
29.47
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
35.13
34.37
43.53
47.00
44.23
43.20
35.03
46.63
38.83
47.17
1931
1932
University Farm
43.07
27.00
43.27
36.60
24.67
32.77
35.13
36.57
39.90
39.30
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
36.80
26.90
26.43
25.57
30.00
28.07
27.43
29.07
26.80
38.00
1931
1932
Waseca
55.20
48.87
58.10
65.77
46.77
48.57
47.33
63.83
50.23
58.80
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
37.73
33.47
42.20
44.70
41.27
36.03
38.50
49.23
37.40
58.17
1931
1932
Bushels/Acre
Year

The barley data of dot-plotting fame was described by Cleveland as possibly having an error in the Morris site records due to how strange the data looked relative to other sites in Minnesota. Later, Wright brought in more data and statistical analyses to claim that there was nothing wrong with the data and that year-to-year variations explained the apparent discrepancy [1]. In my visualization, I seek to tell Cleveland's original story, but without dot plots.

While the dot plot elucidated a pattern of higher yields in the Morris site, I believe that the story could have been more clearly told by expressing in an obvious way the total yields (numeric) of each site positionally. I do so (using Python's Pandas and Matplotlib) by stacking the various barley yields by year (numeric, but treated nominally per the data-screw-up narrative), in bar charts by site (nominal) such that individual variant (nominal) yields (numeric) are encoded by length and reinforced by explicit labels (visually downplayed relative to the dot plots; labels are white for better contrast). Individual variants are consistently color coded (appropriate for their nominal nature), but labeled along connecting lines between years to avoid forcing viewers to shift attention between the plot and some distant legend. Individual variations among variants are not particularly important for my narrative and are thus downplayed, but lines connected between the bases and the peaks of individual variant yields within each stack across years encode year-to-year variation of individual variants as a delta between angles (with the intent that a viewer may discern increase/decrease, though not necessarily by how much without following the lines to the variant yield numbers within each stacked bar) [2]. Nominal variables are sorted lexicographically (i.e. the order of the subplots for each site and the stacking order for variants) for consistency.

Since my subplot axes are of consistent units, I save space by moving labels to the outer borders of the collection of subplots. Y-axis tick marks are omitted since they are effectively redundant (for the story I'm trying to tell) with the individual variant yield numeric labels in each bar stack, and because they would clutter the graphic. X-axis tick marks and labels are replicated for all subplots to avoid forcing viewers to jump far away from each individual plot to get numeric context. I'm less worried about this for units, because I believe critical viewers will tend to be able to keep in mind units better than specific numbers. Along a similar vein, I omit a full-graphic title since the most probable viewer is a grader or peer in CS448B who already has the context in mind.

While it is questionable whether or not bushels of different variants are comparable and thus may be added togther in the way I have done so, I don't believe this is a problem. Since I do not attempt to quantify or emphasize the numeric difference between sums of variant yields (only the qualitative difference in an obvious way), I don't believe this should be considered 'lying'. Furthermore my graphic makes no attempt to make variant-by-variant comparisons obvious across sites, or even within sites.

[1] http://www.tandfonline.com/doi/full/10.1080/00031305.2013.801783

[2] I'm aware of possible angle illusions, but I don't think the individual variant increases/decreases are important enough for me to care too much