Assignment 1 Xin Jiang (xinj6@)

Research Questions (RQs)

This visualization was created to understand the following questions (ordered by importance):

1. Whether there is a difference in yield between the two years for each variety?
2. Which variety(s) of barley performs best(or worst) **in each year** measured by yield in bushels per acre in each year?
3. Which variety(s) of barley performs best(or worst) **on average** measured by yield in bushels per acre averaged over the two years?

Given the limited real estate of the visualization, the following questions will not be addressed:

1. Differences in yield between experiment stations
2. Overall yield of the 10 varieties between the two years
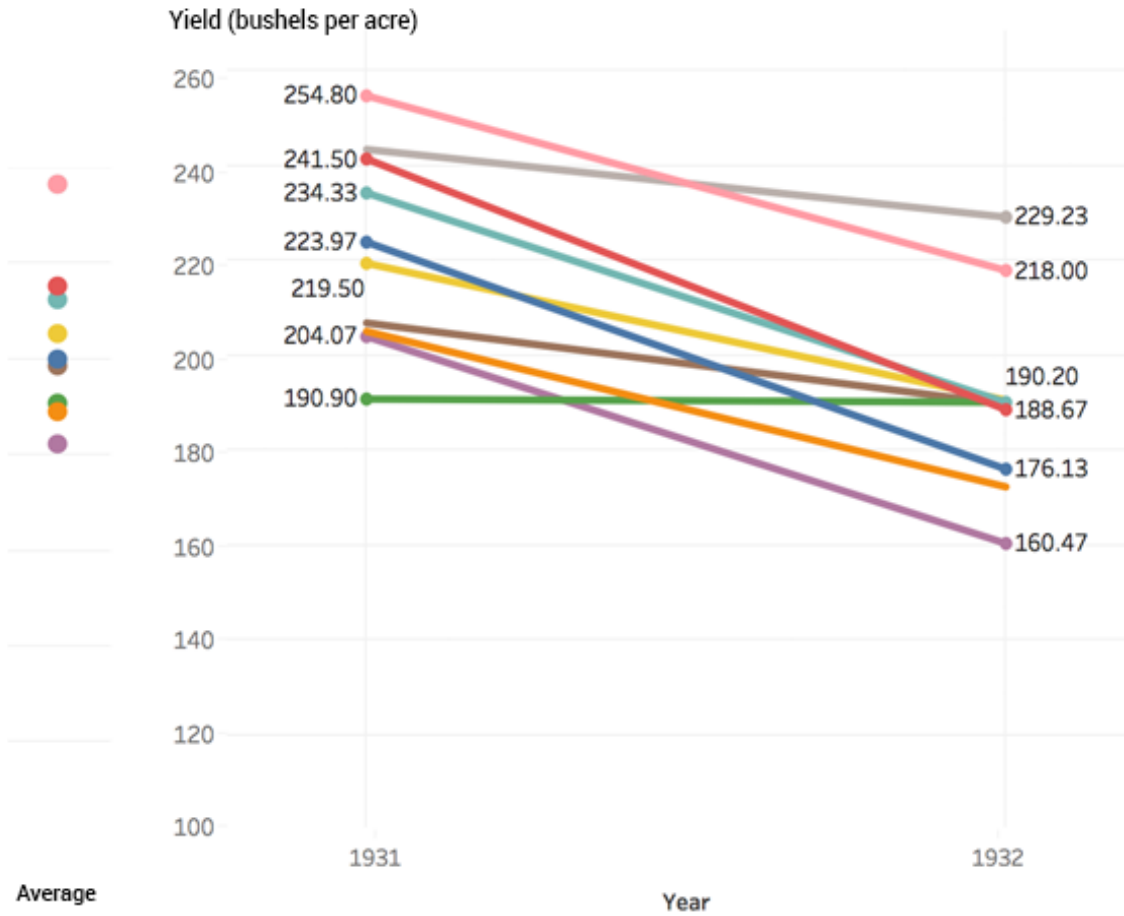
Visualization Design

Some key aspects being considered while designing the visualization are:

- Identified the types of variables of interest:
    - Dimensions (ordered by importance)
    1. Year (Quantitative)
    2. Variety (Nominal)
    3. Experiment sites (Nominal) not visualized
    - Measure: yield (Quantitative)
- Data transformations:
    - Get the overall yield across the six experiment sites for each variety in each year.
    - Get the averaged yield across the two years by averaging the yields of the two years for each variety.
- Identified the importance of the relationship between variables based on the importance of RQs:

    Year and Yield > Variety and Yield > Experiment site and Yield (not visualized)
- Identified visual types for variables:

    Variables and relationships have been ranked by importance, based on Mackinlay's ranking principle, for quantitative variables, Position > Length > Angle > ……, for nominal variables: Position > Hue > Textures > ……, thus the following visual encoding are selected:
    1. Year - Quantitative - Important - x-position
    2. Yield - Quantitative - Important - y-position
    3. Variety - Nominal - somewhat important - hue
    4. Experiment sites - Nominal - not important - not visualized
    5. *Using lines to reveal the trend between the two years
- Label scale: Since all the yield data are in the range of 100 - 260, given the limited real estate, starting y-axis from 100 instead of 0 makes the trend in the data easier to be discovered.
- Two-section design: since average yield is different from yield per year by it's nature, the visualization should include two parts: one part (left) shows average yield and the other part (right) shows yield in each year.

Data Visualization

Yield (bushels per acre)

	1931	1932
Trebi	254.80	218.00
No. 457	241.50	188.67
No. 462	234.33	190.20
Glabron	223.97	176.13
Peatland	219.50	
Velvet	204.07	
No. 475	190.90	
Wisconsin No. 38		229.23
Svansota		160.47

Average

Year

Variery

- Glabron
- Manchuria
- No. 457
- No. 462
- No. 475
- Peatland
- Svansota
- Trebi
- Velvet
- Wisconsin No. 38

Conclusions based on visualization

The visualization helped us answer the research questions effectively:

1) Except for No.475, there is a drop in yield in the year 1932 compared to year 1931.

2) In year 1931, the ranking of the yield of varieties are:

   Trebi > Wisconsin No. 38 > No.457 > No. 462 > Glabron > Peatland > Velvet > Manchuria > Svansota > No. 475

   In year 1932, the ranking of the yield of varieties are:

   Wisconsin No. 38 > Trebi > No. 462 > No. 475, Peatland, Velvet > No. 457 > Glabron > Manchuria > Svansota

   Based on the rankings, we could conclude that Trebi has the best performance in year 1931, while Wisconsin No. 38 has the best performance in year 1932. No. 475 has the worst performance in year 1931, while Svansota has the worst performance in year 1932.

3) The ranking of the yield of varieties on average are:

   Trebi > Wisconsin No. 38 > No. 457 > No. 462 > Peatland > Glabron > Velvet > No. 475 > Manchuria > Svansota

   Based on the rankings, we could conclude that Trebi has the best performance on average over the two years, and Svansota has the worst performance on average over the two years.

Appendix

Description of the tools

I used **Tableau** to create the line chart that reveals the relationship between varieties and yields in the two years, and used **Photoshop** to add the visualization that reveals the relationship between varieties and yields on average of the two years.