

EDA - Assignment 6

Jivitesh Poojary

March 09, 2017

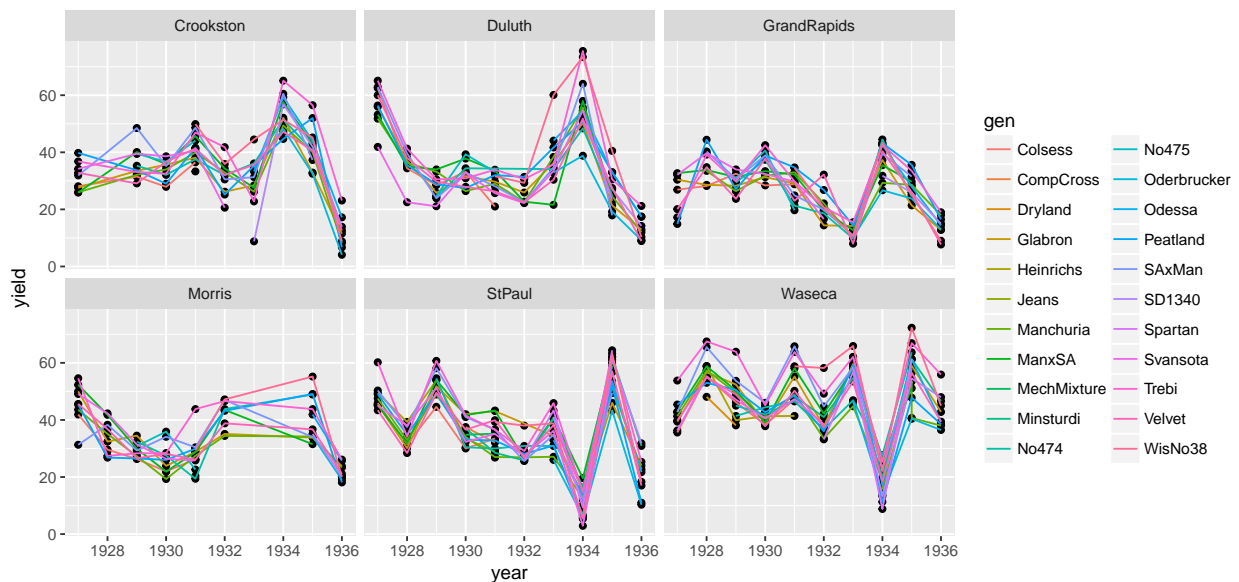
Answer 1

Q1. Draw an appropriate faceted graph showing how barley yield varied by year at each location, using color as necessary. (Note: Lines might be better than dots here.) When looking at in successive years, was it more common for the yields to move in the same direction at all locations, or was it more common for the yields to increase at some locations and decrease at others?

ANSWER: - From the plots we can see that the barley yield varies with location

- Some values appear to be missing in the data. The below inferences are drawn from analysing the available data.
- For Crookston we see that the yield remains the same with slight aberrations, there is a peak at 1934 followed by a drop at 1935 and 1936
- Duluth begins with a high yield and falls till 1928, the yield remains almost the same till 1933, where it increases and reaches a maximum value at 1934 followed by a drop at 1935 and 1936
- For GrandRapids we see that on an average the yield is around 30 there is a drop at 1933 followed by a peak at 1934, followed again by a drop at 1935 and 1936
- Morris begins with a higher yield and drops till 1930 where it starts increasing from 1931, the data seems to be missing for years 1933 and 1934. The yield drops from 1935 to 1936.
- For StPaul we see that the yield changes a lot, we observe a peak at 1929 and 1935. We also see that there is a drop in production at 1934 and 1936.
- Waseca has a higher yield on average compared to other locations. The average yield here is around 50. The most notable feature about the curve here is that the yield drops in 1934 and then continues with the pattern before.

Year to Yield – faceted plot on site



Answer 2

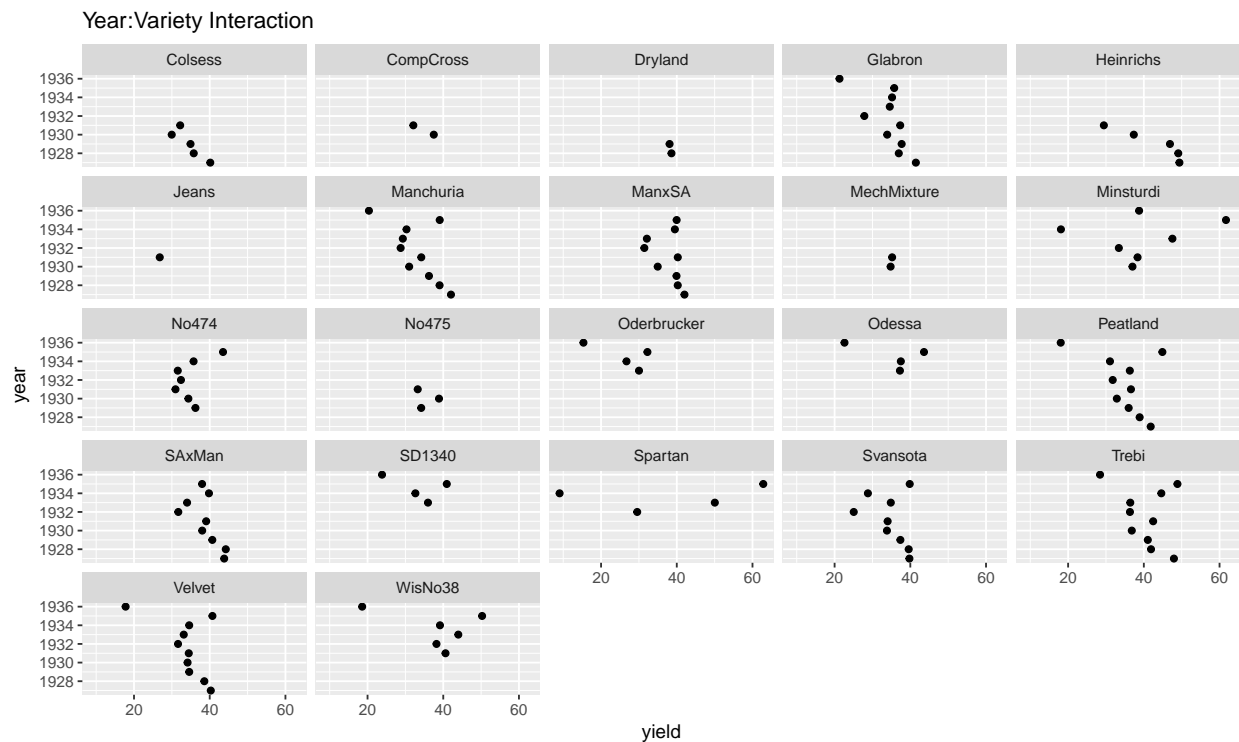
Q2. Fit a model with yield as the response and gen (variety), year, and site as explanatory variables, with the goal of determining whether Morris 1931-1932 is an anomaly. Justify why you chose this model and not some other one. Because of outliers, you should use a robust fitting method.

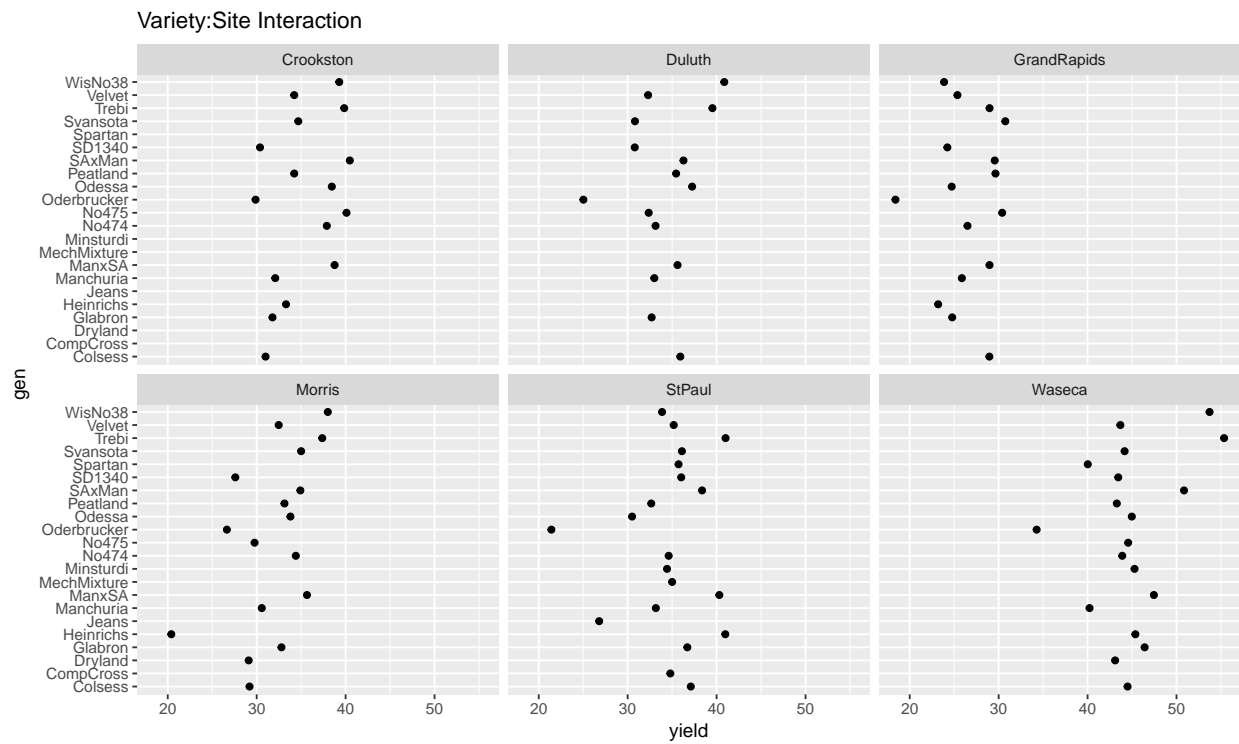
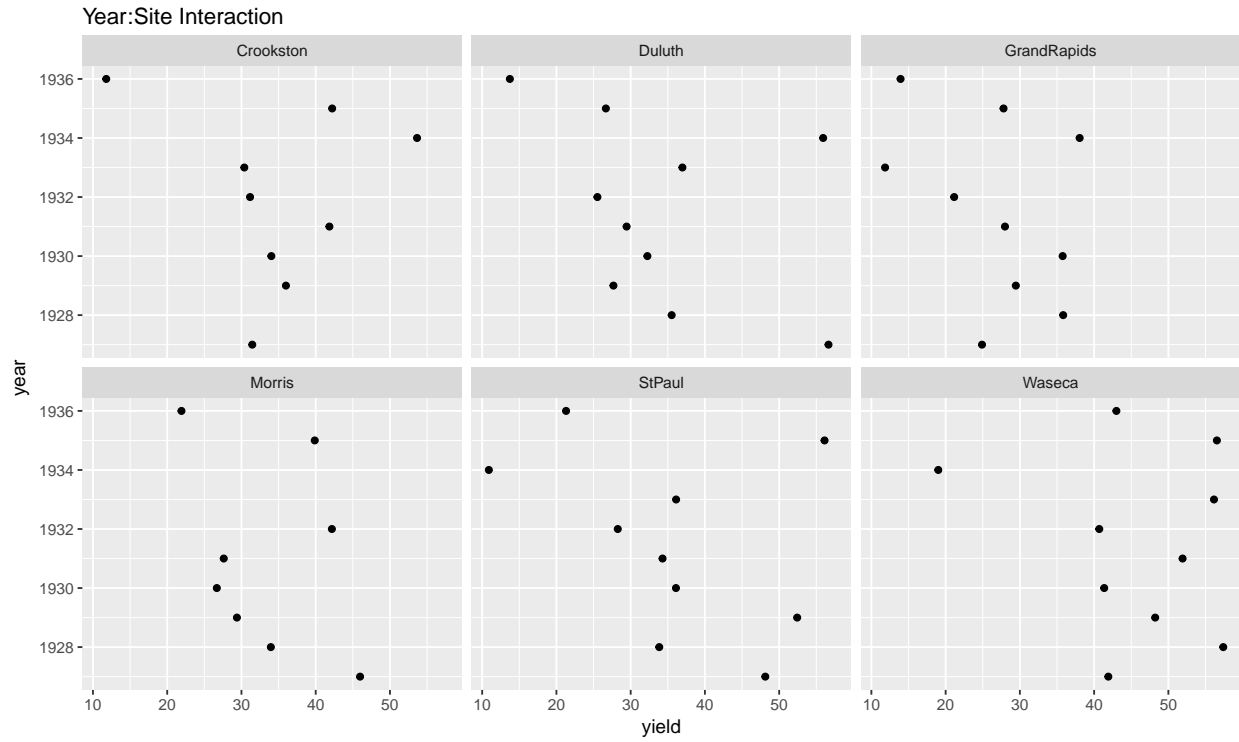
ANSWER: - As there are many outliers in the data we need to apply a robust model on our data

- From the plots below we see that there seems to be some interaction between year and site. For the year:variety plot we see that there could be some interaction but it seems to be minor. The mean value appear to be distributed around 40. Similarly, we also see that the variety:site plot does not have a great deal of interaction.
- We are thus making use of on year:site interaction in our model.
- Both the additive model and the model with interaction is mentioned below.

```
barley.rlm = rlm(yield ~ gen + year + site, psi = psi.bisquare, data = barley)
```

```
barley.rlm2 = rlm(yield ~ gen + year * site, psi = psi.bisquare, data = barley)
```



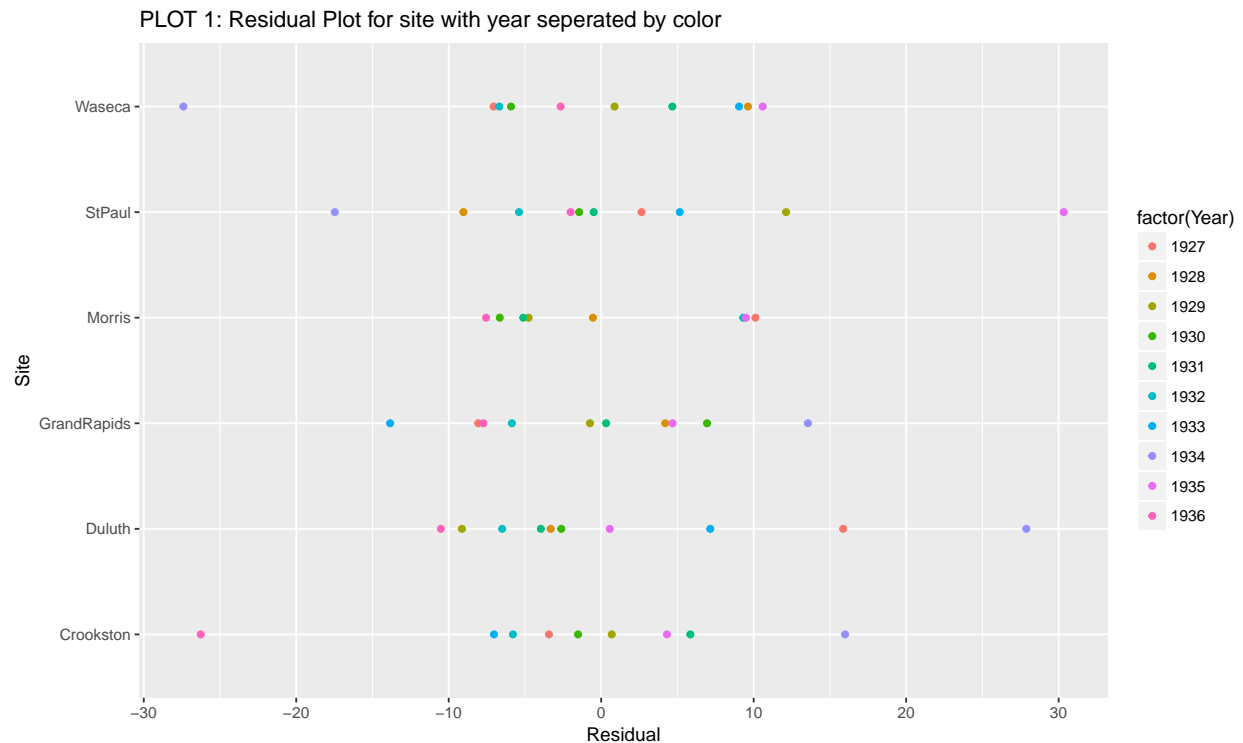


Answer 3

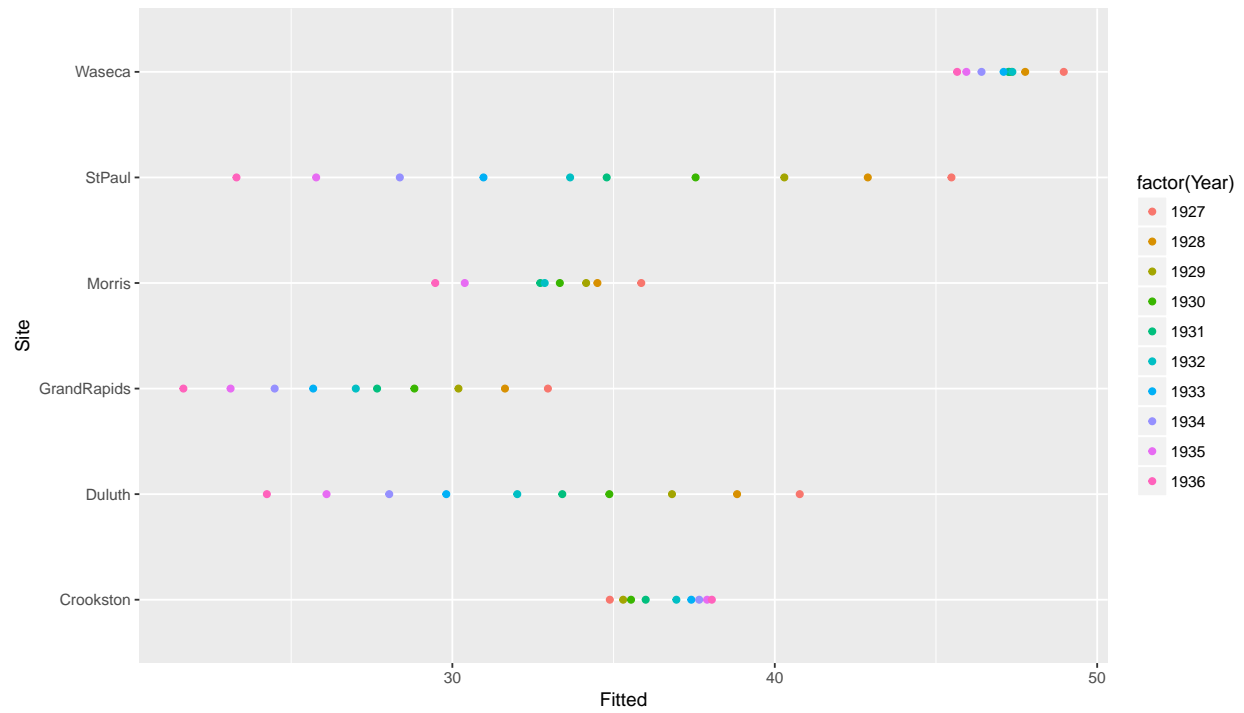
Q3. Draw a set of faceted plots of the residuals with the goal of determining whether Morris 1931-1932 is a mistake, or whether it can be explained as natural variation. As best as you can tell, was there a mistake?

ANSWER: - I have drawn 4 plots below for our analysis.

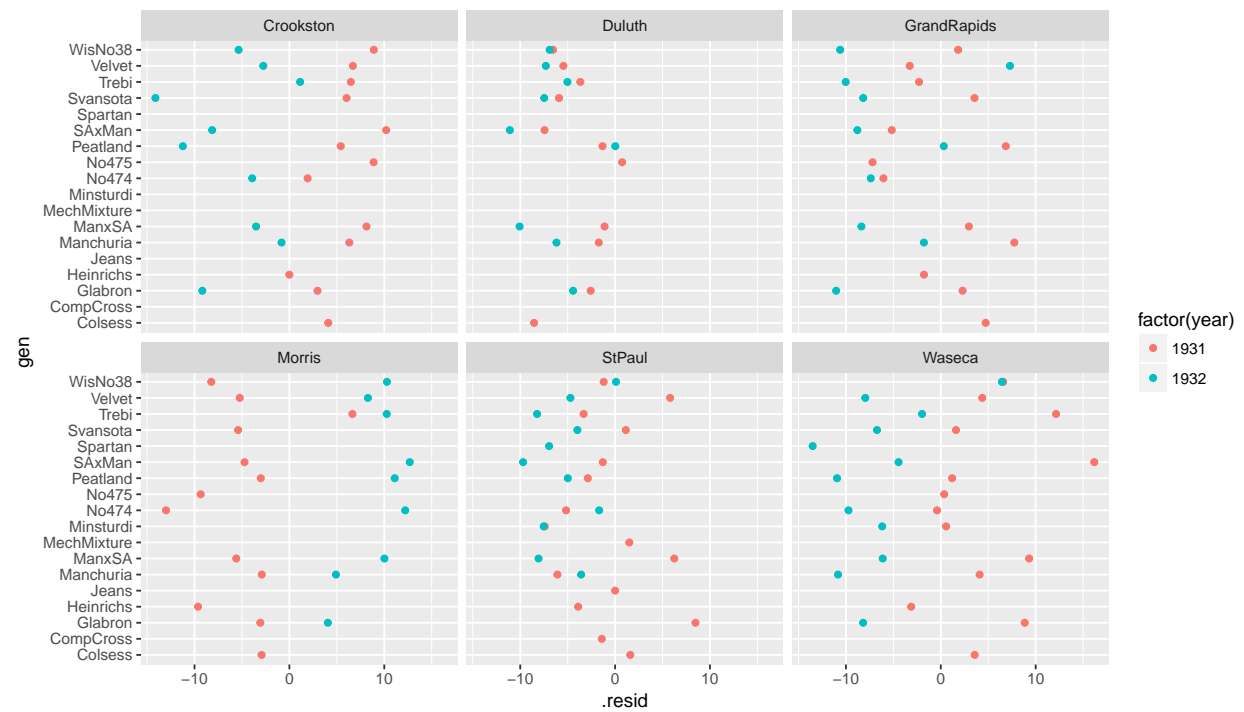
- PLOT 1: I could not apply the plot with dummy coef on this data. From this plot we observe that the residual vary with location for the years 1931 and 1932. This difference cannot be attributed to anomaly.
- PLOT 2: It is a plot of fitted values vs site seperated by color for year. It just tells us the anomaly for Crookston. Morris appears to be fitting properly in this case.
- PLOT 3: This plot again confirm our conclusion that Morris does not show an aberrant behaviour.
- PLOT 4: The residuals for Morris appear to be on the other side with respect to years 1931 and 1932. However it cannot be considered as an anomaly as we have seen this kind of variation in the past years for some other locations as well. We can attribute this aberration to natural phenomenon.



PLOT 2: Fitted Plot for site with year separated by color



PLOT 3: Residual Plot for year 1931–32 faceted on Site



PLOT 4: Fitted Plot for year 1931–32 faceted on Site

