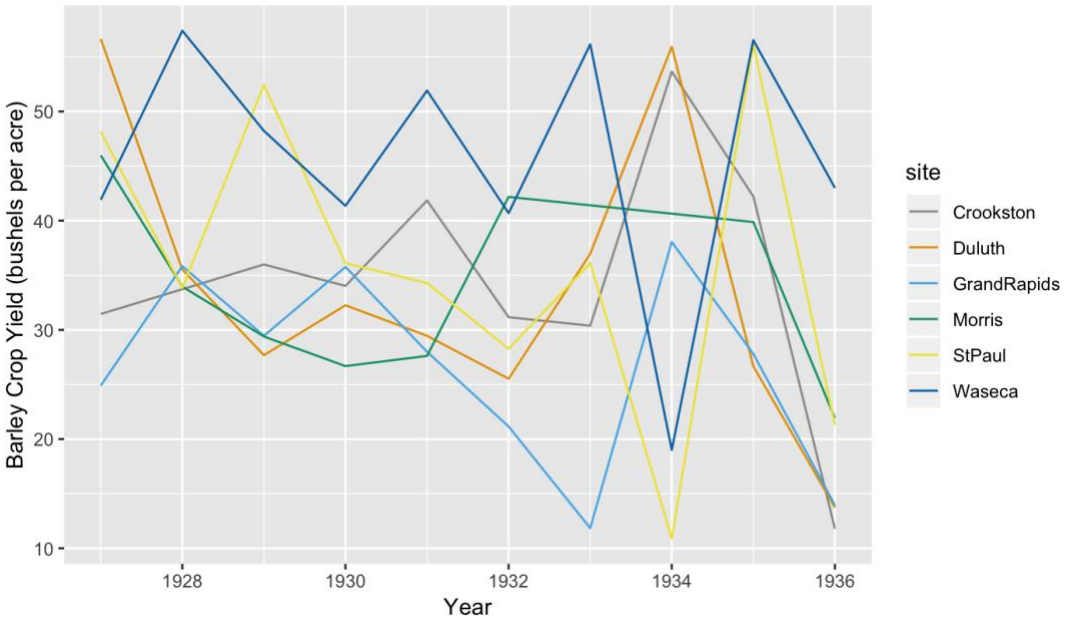
Problem set 6: Minnesota barley revisited

S670 Spring 2019

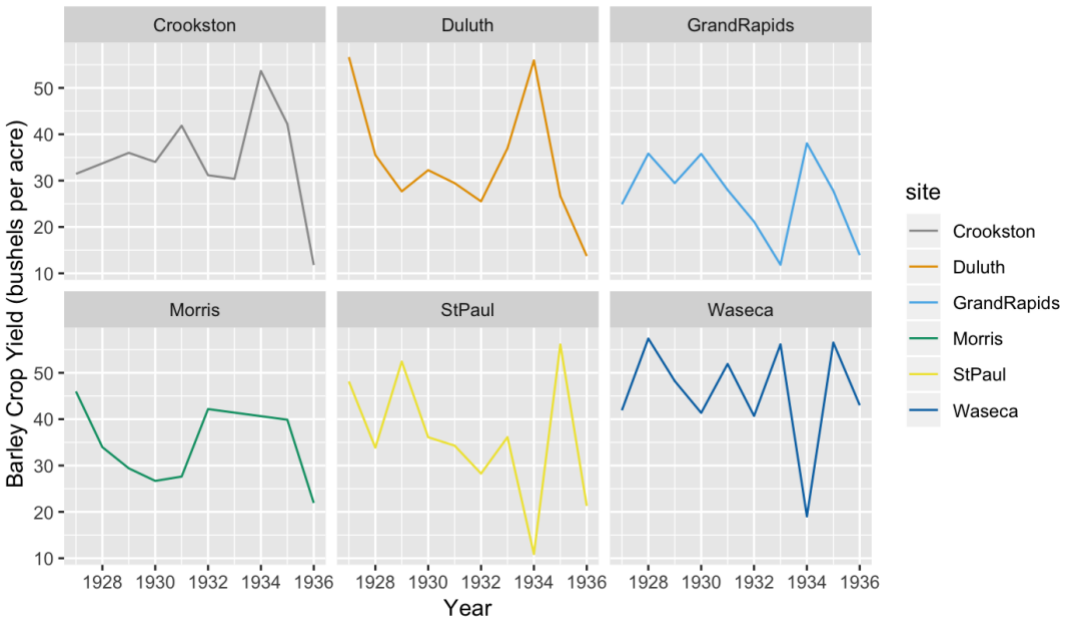
Team Virginia: Pei-Yi Cheng, Sumeet Mishra, Ling Tan, Kimball Wu.

Q1.

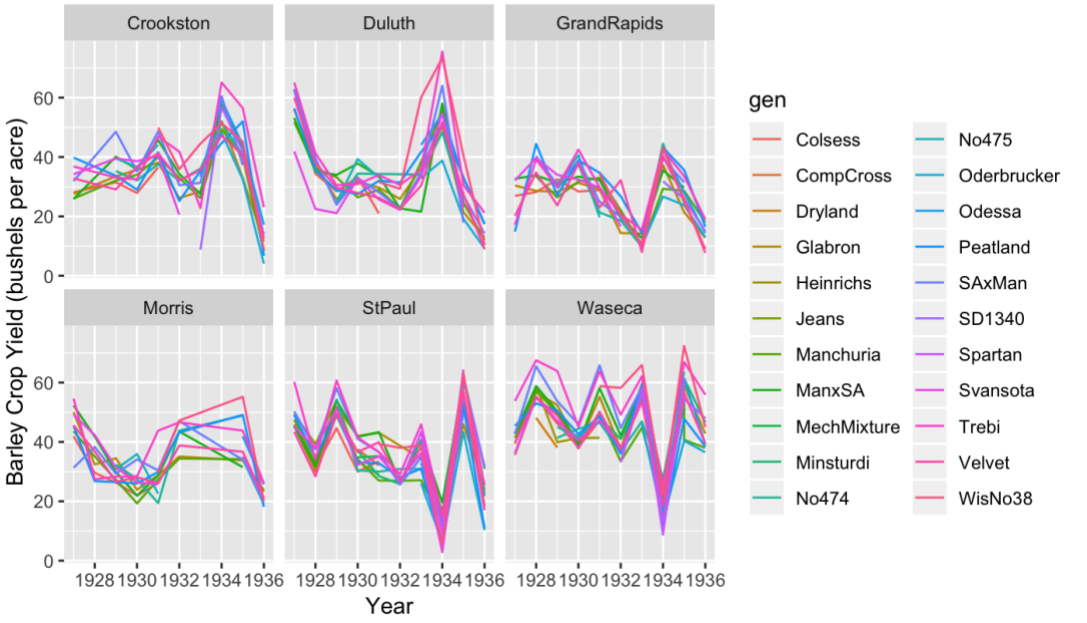
Graph 1:Plot of the Relationship between Average Yield and Year faceted by Site



Graph 2:Plot of the Relationship between Average Yieldand Year faceted by Site



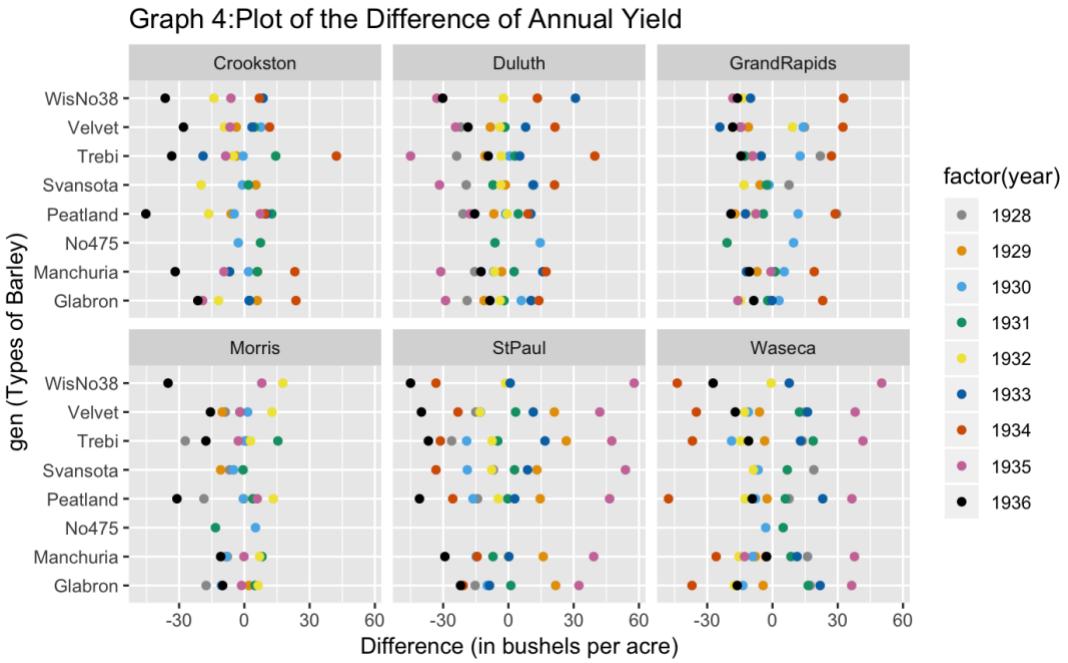
Graph 3:Plot of the Relationship between Yield of Each Type of Barleyand Year face



As we can see from the faceted graphs above, there is substantial barley yield variability across sites:

1. StPaul and Waseca have a similar trend of yield in the year 1929 to 1936. We can observe that yield in these two locations has dropped a considerable amount in the year 1934, and also increased a great amount in the year 1935.
2. From the year 1931-1936, yield in Crookston, Duluth and GrandRapids has similar increasing and decreasing trend; however, in the year 1931, yield in Crookston is increased from the previous year, but yield in GrandPapids and Duluth has dropped.
3. In the year 1934, yield in StPaul and Waseca has dropped exponentially, but yield in Crookston, Duluth and GrandRapids has increased exponentially; moreover, in the year 1933 and 1934, no yield data has been included for Morris.
4. Moreover, on Morris, other than the missing data in 1932-1933, yields over previous years do not share a similar pattern with other sites. Yields decreased from the year 1827-1930, increased a little in the year 1931.
5. For the year 1936, yield in all locations has dropped exponentially.
6. Moreover trend of the yield of each type of barley facet by site, we can observe each type of barley in each site has a similar trend of yield over years. Therefore, the yield in each location does not have large variability across type.

Q2.

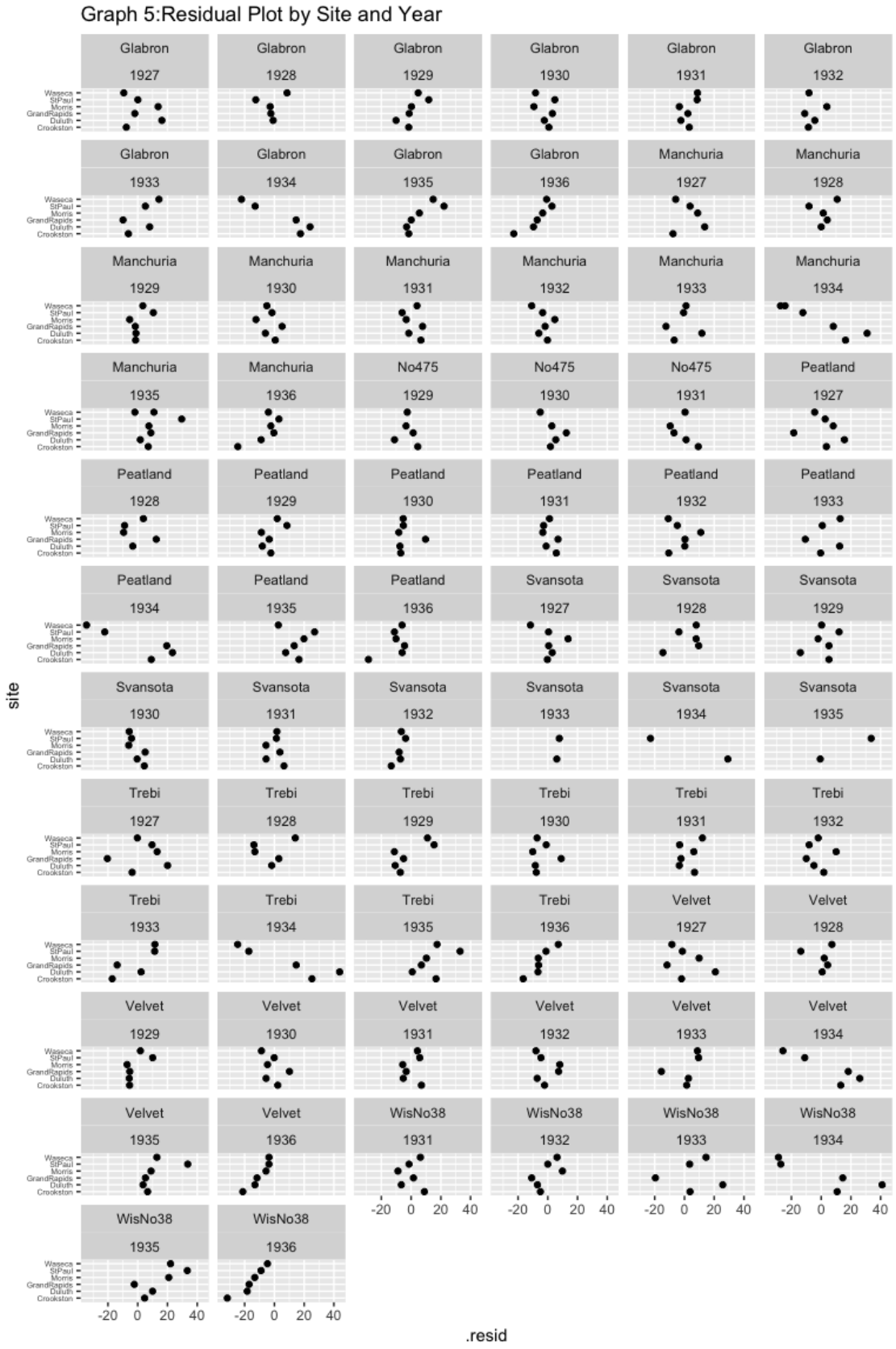


Some of the barley varieties not being planted in all locations, also, some of the barley varieties was not being planted in some years. Therefore, we have filtered out barley types that were not been planted in some year. After filtering 8 type of barley varieties left in the dataset.

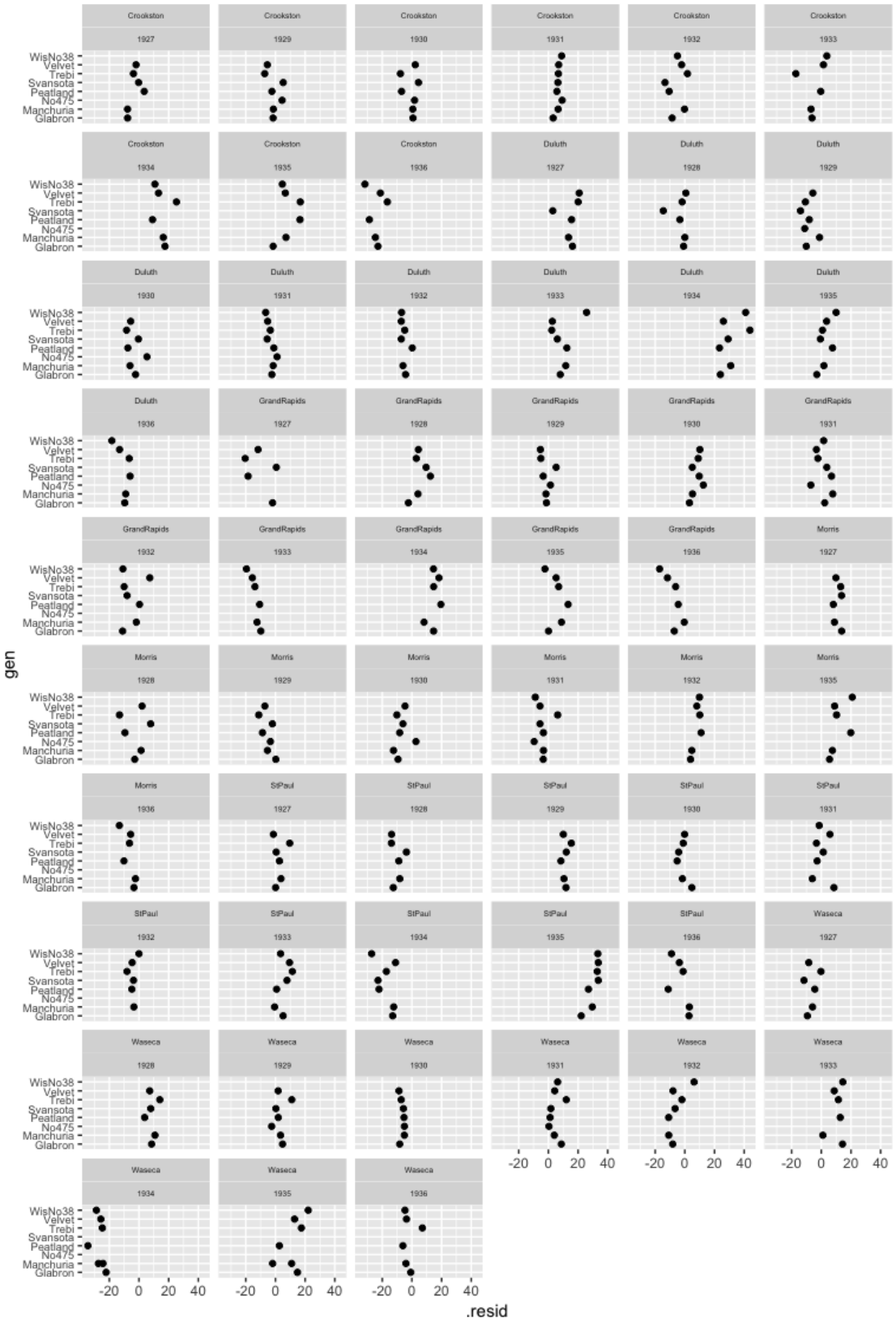
First we can observe from the Graph 4 difference plot, there is very modest variation in gen-by-gen, hence we do not need a gen interaction.
However, when we exam the differences of yield of each year at different sites, we have detected great amount of variation in some years. Although some year variation is modest, but there is some obvious variation across different sites in different years, especially in year 1934-1936. Moreover, Morris has large amount of variation compares to other sites over years. Therefore, we need a year:site interaction to our model. We examine residual plots in Q3 to determine whether to include other interactions, and we find that other interaction effects are small or negligible.

Model :
*rlm(yield ~ gen + year * site, psi = psi.bisquare)*

Q3.


Graph 5:Residual Plot by Site and Year
Glabron 1927
Glabron 1928
Glabron 1929
Glabron 1930
Glabron 1931
Glabron 1932
Glabron 1933
Glabron 1934
Glabron 1935
Glabron 1936
Manchuria 1927
Manchuria 1928
Manchuria 1929
Manchuria 1930
Manchuria 1931
Manchuria 1932
Manchuria 1933
Manchuria 1934
Manchuria 1935
Manchuria 1936
No475 1929
No475 1930
No475 1931
Peatland 1927
Peatland 1928
Peatland 1929
Peatland 1930
Peatland 1931
Peatland 1932
Peatland 1933
Peatland 1934
Peatland 1935
Peatland 1936
Svansota 1927
Svansota 1928
Svansota 1929
Svansota 1930
Svansota 1931
Svansota 1932
Svansota 1933
Svansota 1934
Svansota 1935
Trebi 1927
Trebi 1928
Trebi 1929
Trebi 1930
Trebi 1931
Trebi 1932
Trebi 1933
Trebi 1934
Trebi 1935
Trebi 1936
Velvet 1927
Velvet 1928
Velvet 1929
Velvet 1930
Velvet 1931
Velvet 1932
Velvet 1933
Velvet 1934
Velvet 1935
Velvet 1936
WisNo38 1931
WisNo38 1932
WisNo38 1933
WisNo38 1934
WisNo38 1935
WisNo38 1936
Waseca
StPaul
Morris
GrandRapids
Duluth
Crookston
-20 0 20 40
site
.resid

Graph 6:Residual Plot by gen(Types of Barley) and Year



From the residual plot by site and year, the residual does not have an obvious pattern in various site over years. Furthermore, the residual at Morris is not significant it does not stand out from other sites. The yield residual in some other site has greater variation than Morris, for instance, in Duluth and Waseca. Therefore, there are not enough evidence to conclude data at Morris is a mistake. The variation can probably explain as natural variation, since the barely at Morris failed to yield in the next two consecutive years.