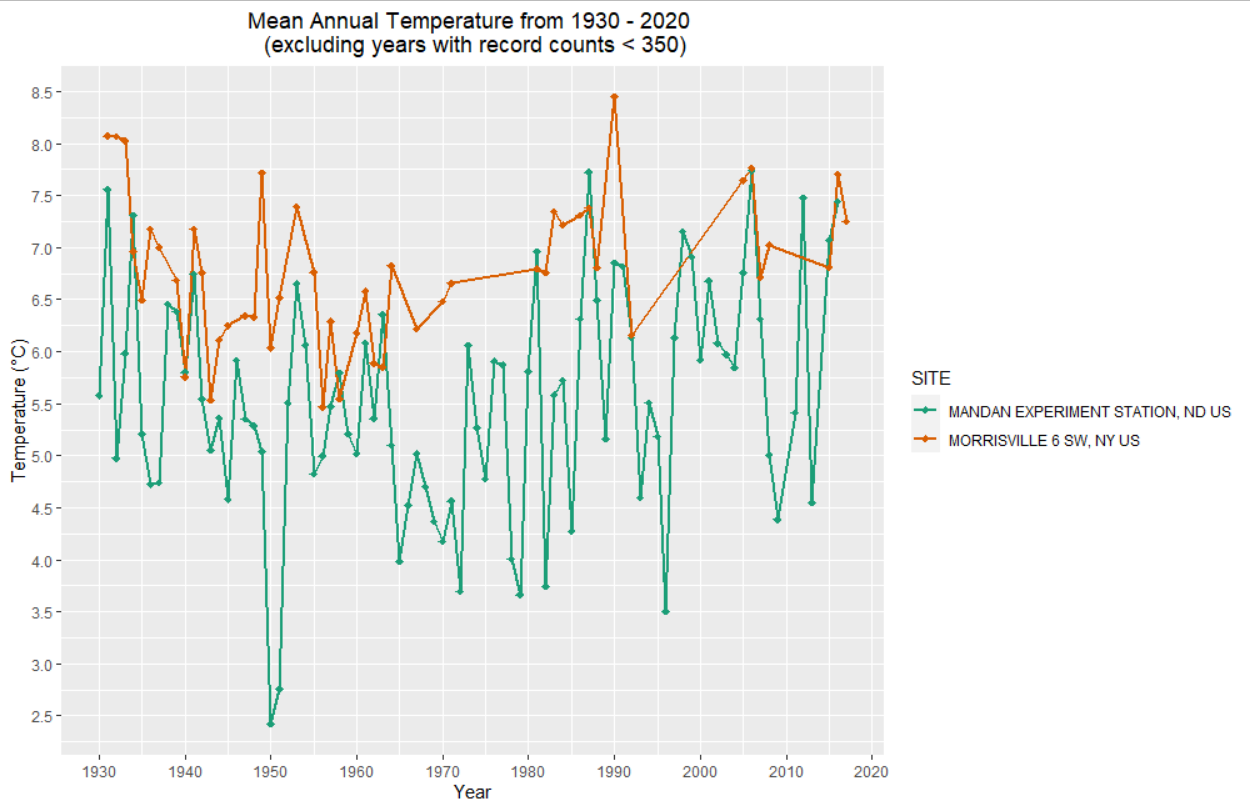
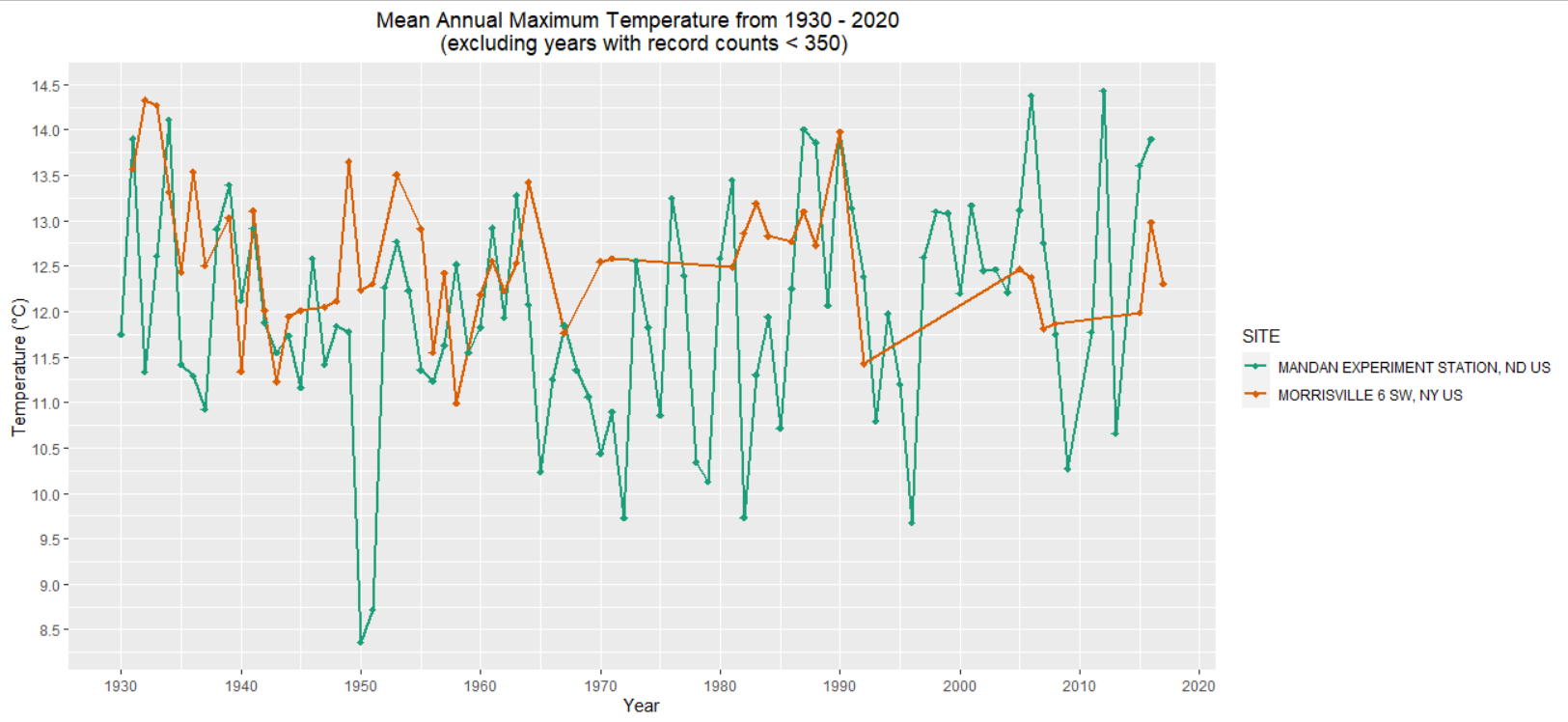
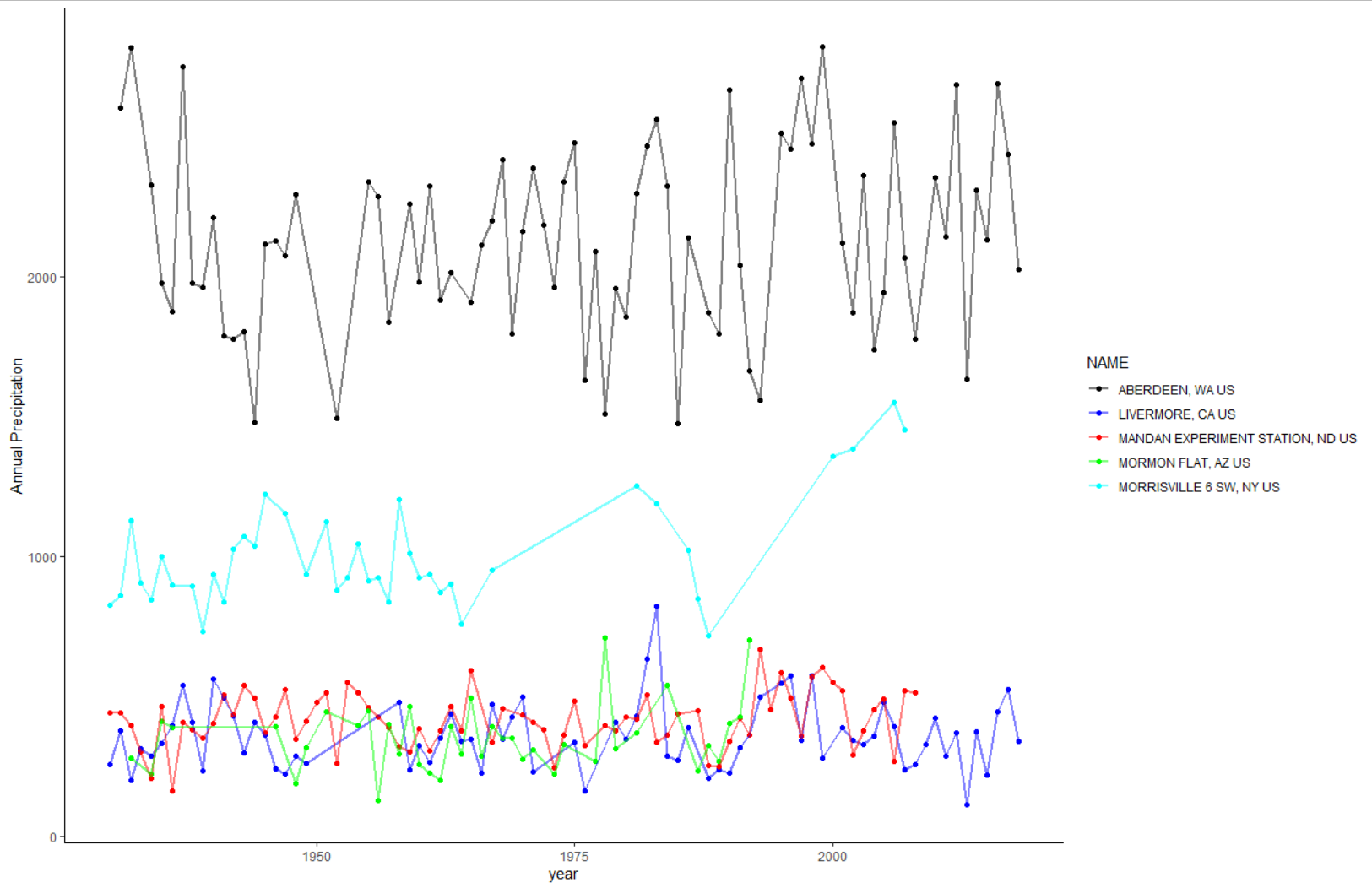
1. When plotting the total precipitations in each year, it is hard for the graph to show how many records each year has. The graph cannot represent the actual change in yearly pattern if they have dissimilar basis – namely, if one year has 2 data point and another one has 365, they would not be comparable, and the graphs would be misleading if only showing the total.

2. Before 1980s the total precipitation varies but are mostly stable in a range, but at the beginning of 1980s both sites experienced a sharp increase in total precipitations, followed by a sharp drop shortly after. By 1990s the trend from NY and CA began to differ, with the precipitation in CA going back to normal variations around the old range, while the number in NY goes up compared to before. Another aspect worth mentioning is that between 1950 – 1960 CA has some missing data points, while there are two great gaps before and after 1980 for NY, so I would not make any conclusive assumptions purely based on this graph.

3. 

4. 

From the graph shown above, Mandan, ND has a higher variability in the mean annual maximum temperature comparing to the data at Morrisville, NY. Overall, the NY data are gradually decreasing by a really small margin, but because of the missing data in the middle not much of the pattern can be collected. For ND data, on the other hand, looks healthier with more data point, and it is easy to observe one sharp decrease in the year of 1950 and 1951, then the average max temperature quickly recovers to the previous state; each local minima seems to space out ten years apart, while the local maxima has been increasing for the past 100 years.

5. 

I used blue, red, and green to exaggerate the contrast for close points, also increased the point alpha to make them stand out more.

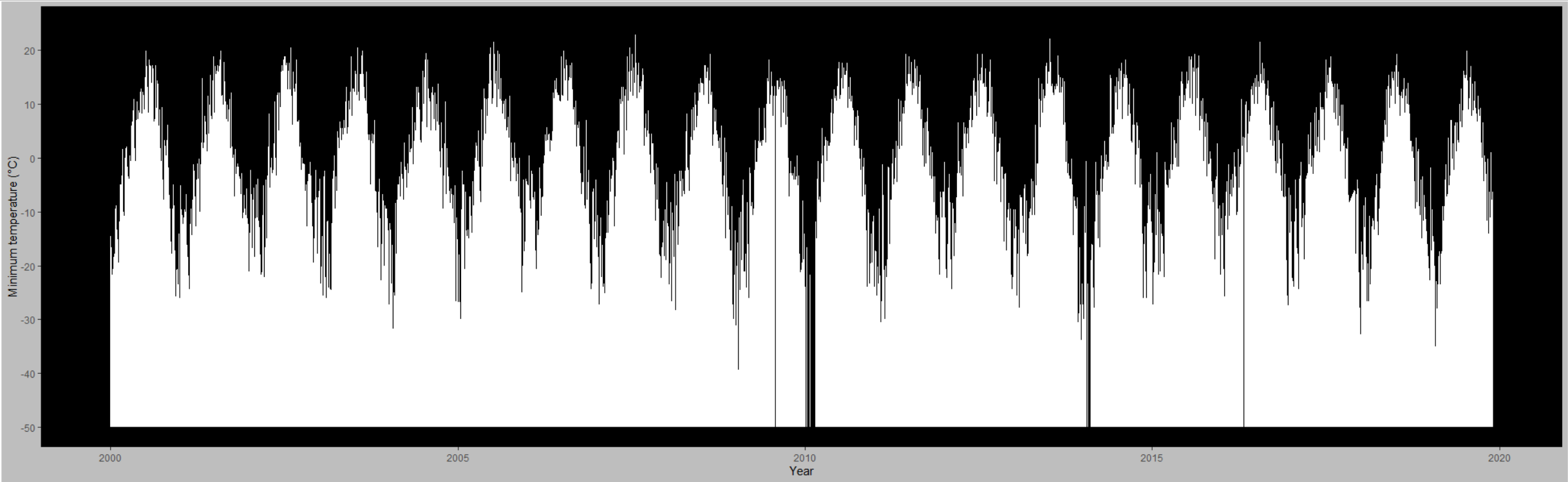
6. a. sub$DATE <- as.Date(sub$DATE,"%b-%d-%y")

b. sub$DATE <- as.Date(sub$DATE,"%m/%d/%Y")

c. sub$DATE <- as.Date(sub$DATE,"%d-%B-%y")

7. The temperature is relatively symetrical and overall looks unimodel from the macro level. The January is the lowest point, as the year goes on, though the temperature fluctuates, the overall pattern increases to a maximum around July, then as the year continues, the temperature drops back to the lowest point again around the next January. Two sharp drops that looks outside of the ordinary happened once during June and once during November. Precipitations are less structural, with moderate amount of rains in January and March, and a bit precipitation between these two month, then followed by a dry period til July, when the precipitation are bacl to a similar volumes to March, gradually dropping as time goes on to November, then a sharp increase to the global maximum in November with very concentrated recurrences. December is also relatively dry, and when it got close to the next year there were a bit precipitation again.

8. For Livermore in CA, the temperature still fluctuates from a day to the following, and the trend generally increases from the lowest point in January, with a rather flat slope of increase till it reaches the maximum around August, then decreases back down. Differences are the first, the graph is not as symetrical, indecating that in 1975 the year may starts warmer than the previous year; moreover, the overall temperature has a wider spread (variance), and the hotest day between the two location goes to AZ, while the coldest day probably does to CA. As for precipitation, the max for CA happenes in March, vs the max for AZ happenes in November; there’s sort of a unimodel distribution within Jan – May, followed by a 5 months of dry period, which is longer than in AZ. A good amount of precipitation comes back after October, but not nearly as much as earlier in the year.

9. 

The above graph is available to for zooming or looking at sectionally for details, but the general idea is that the overall distribution in NY for past 20 years has ben relatively similar. Admittedly the max in each year changes, but one can observe the yearly unimodel distribution that takes pplace in the middle of the year, while the minima are at either the beginning or the end of the year.

10. https://github.com/guozhaosengzs/ENVDS/blob/master/activity5/activity5\_script.R