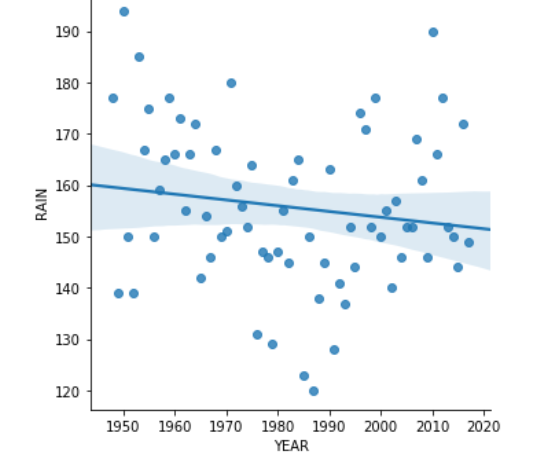
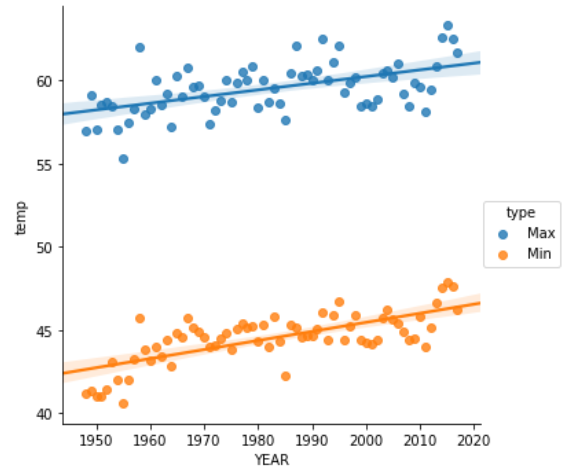
Statistical Question and Conclusion

The dataset that I worked with was the rain in Seattle data set. The data set was a collection of information about rain since the 1940's, which included the date, if it rained or not, the amount if so, and the max and min temperatures. Perhaps prematurely, I came up with the hypothesis that rain has been increasing in Seattle steadily. What my analysis ultimately showed was that it is more or less stayed similar through the time period. There are a few outliers on both sides, but the ultimate conclusion is that there was not a statistical significance of rainy days increasing over time.



Outcome of EDA

Interestingly enough, there was an additional outcome to my exploratory data analysis besides just that my hypothesis was incorrect. Since the dataset was quite large, I first tried plotting the min and max temperatures. This ended up as a bit of a confusing mess due to the volume of the data points. To transform the data, I grouped the data by year, and took the mean temperature of that year and plotted that. You can see from the plot that there is a clear increase over time of the average temperatures per year. In terms of variables that could have helped the analysis, based on what data I had, I had the exact information that I needed. There was not any challenges that I did not face, but I also approached the problem knowing in advance what kind of visuals and analysis that I wanted to perform. Perhaps as a learning point, I might take home to not set a hypothesis too early. I felt that by setting it early, it helped target what I looked into any why, though it was not necessarily based on a strong understanding of data at the time.