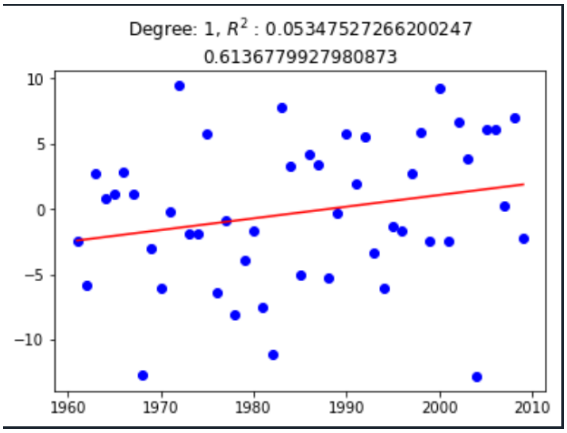
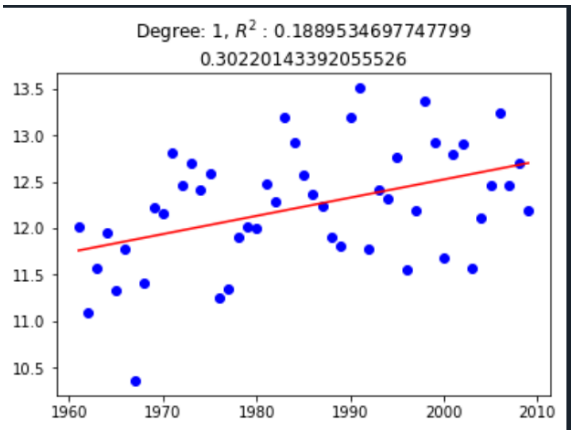
A4.I



A4.II



What difference does choosing a specific day to plot the data for versus calculating the yearly average have on our graphs (i.e., in terms of the R 2 values and the fit of the resulting curves)? Interpret the results.

Nothing, given the time intervals, the data points appear to not differ. I believe due to the date range of decades the difference of choosing one day over another would be minimized. Hence why the data looks similar.

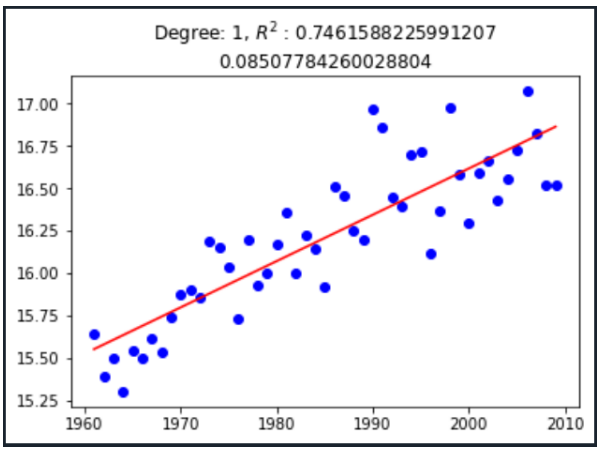
Why do you think these graphs are so noisy? Which one is more noisy?

The graphs are noisy due to the weather of each year. Climate is determined by the average of weather data points. The graphs appear to be the same and one is not noisier than the other.

How do these graphs support or contradict the claim that global warming is leading to an increase in temperature? The slope and the standard error-to-slope ratio could be helpful in thinking about this.

The information provided from the graph only displays temperature for one city over just a few decades. This localized information is not enough to determine the validity of climate change.

B



How does this graph compare to the graphs from part A (i.e., in terms of the R^2 values, the fit of the resulting curves, and whether the graph supports/contradicts our claim about global warming)? Interpret the results.

The graph shows a stronger positive trend compared to the previous graphs. Which would more agree with the assertion of global warming.

Why do you think this is the case?

There are more cities in different areas that contributed to the data which allowed more data points to feed the model for a more accurate picture.

How would we expect the results to differ if we used 3 different cities?

Using only 3 cities I would expect the results to be similar to using only one.

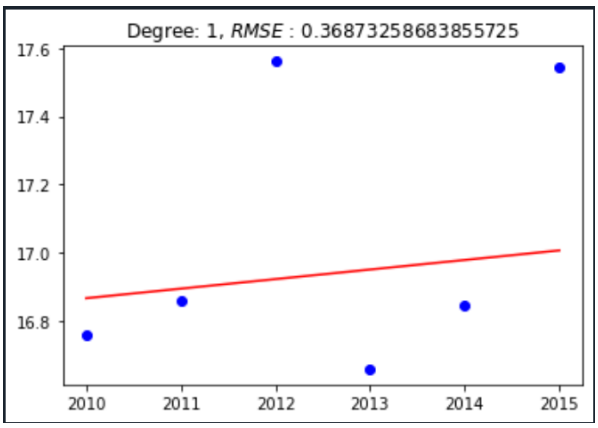
What about 100 different cities?

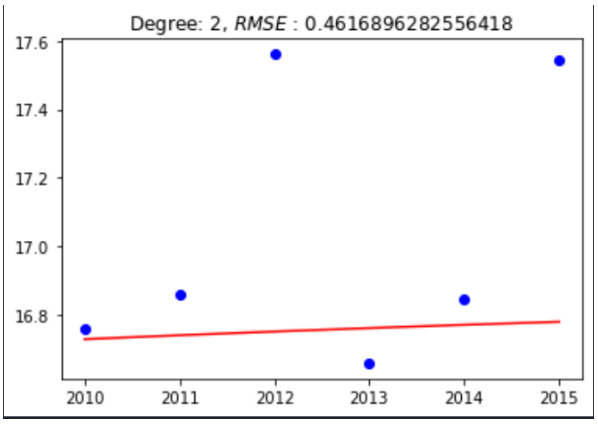
With more data points the data should more accurately reflect the temps from a larger geographic area.

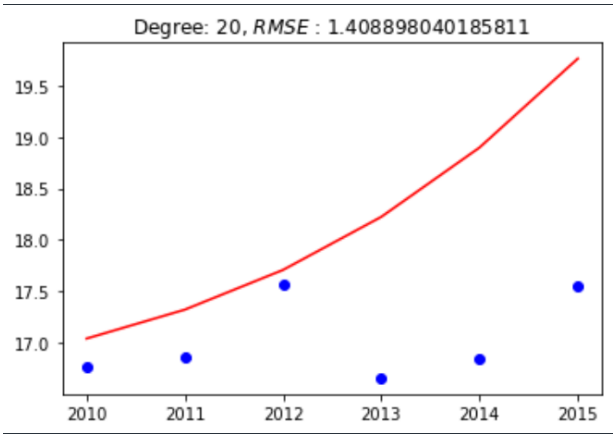
How would the results have changed if all 21 cities were in the same region of the United States (for ex., New England)?

The results would reflect the average temperature of the area, so the trend line would not be to steep or out of the norm. however, this would result in a more regionalize model that could not be used for other areas and would not inform the basis of global warming.

D.II







How did the different models perform?

The models appear to differ but with increasing accuracy.

How did their RMSEs compare?

Which model performed the best?

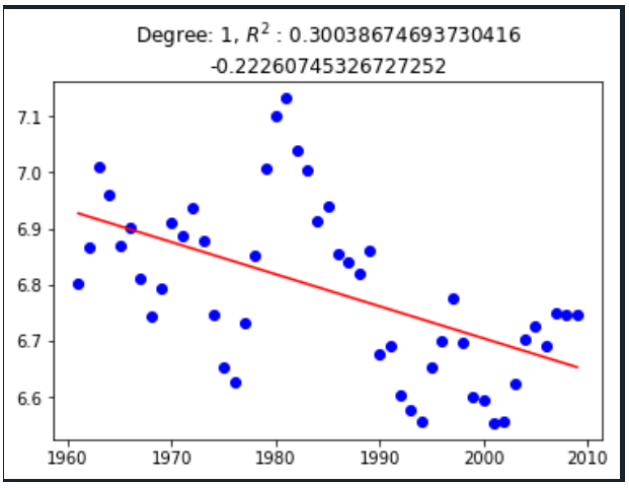
The best model was the 20 as it appeared to have more data points and provided a better picture. Which model performed the worst? Graph 1 did the worst as the trend line appeared to be too linear.

Are they the same as those in part D.2.I? Why?

If we had generated the models using the A.4.II data (i.e. average annual temperature of New York City) instead of the 5-year moving average over 22 cities, how would the prediction results 2010- 2015 have changed?

The results would not have changed much as the temps would have been more consistent in just one area.

E.



Does the result match our claim (i.e., temperature variation is getting larger over these years)?

No, temperature variation appears to be getting lower over the years.

Can you think of ways to improve our analysis?

The date ranges would need to be increased and more geographically differing data points would need to be included.