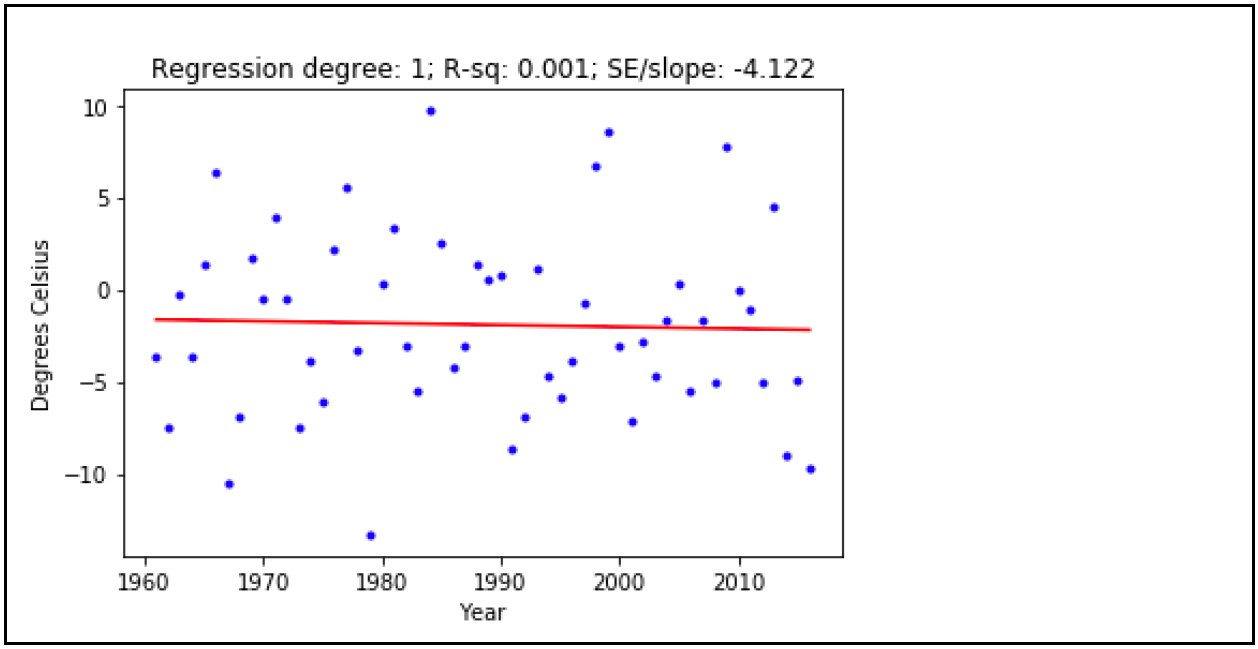
Name: Lydia Yu
Kerberos: lydiayu

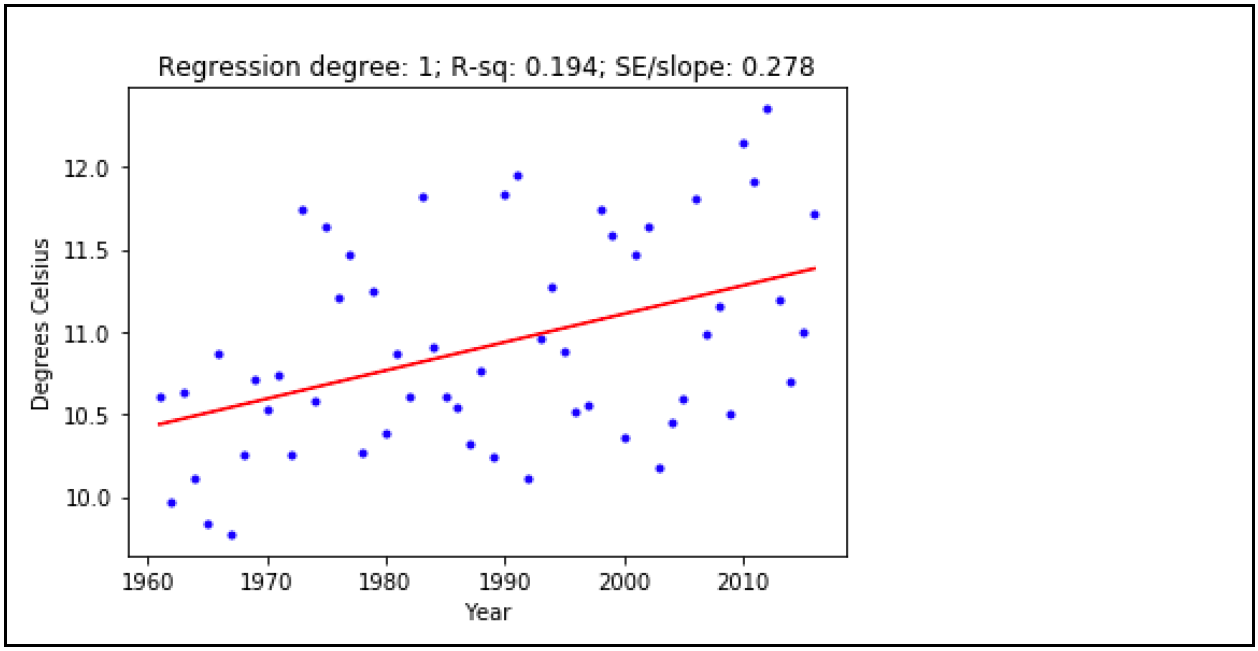
Problem Set 5: Modeling Temperature Change

Problem 4

Plot 4A: *Average Daily Temp for Boston on 2/12 (1961-2016)*



Plot 4B: *Average Yearly Temp for Boston (1961-2016)*



4.1 What difference does choosing a specific day to plot the data versus calculating the yearly average have on the goodness of fit of the model? Interpret the results.

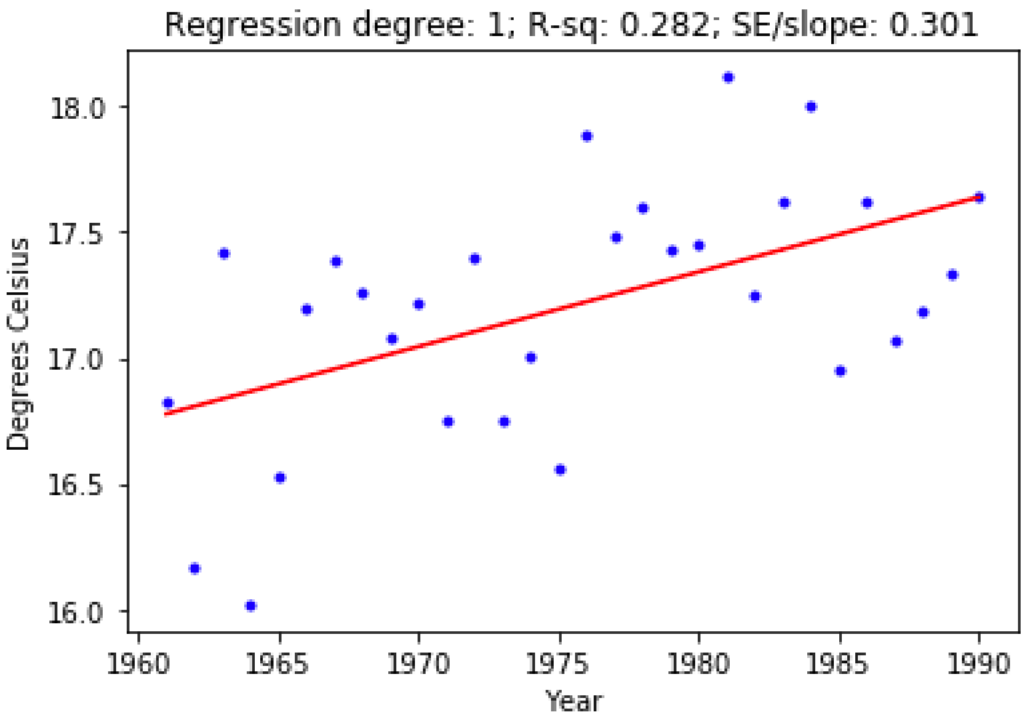
Calculating the yearly average increases the R-sq value, meaning that model has a better fit to the data.

4.2 Why do you think these graphs are so noisy?

For both graphs, you are only looking at one very specific set of data, so there will not be a very strong trend because in the same city, its temperature on a specific date will always be around the same value (+/- a few degrees) and its yearly average temperature will also always be around the same value (+/- a few degrees), but there is no trend to the differences in degrees each year.

Problem 5

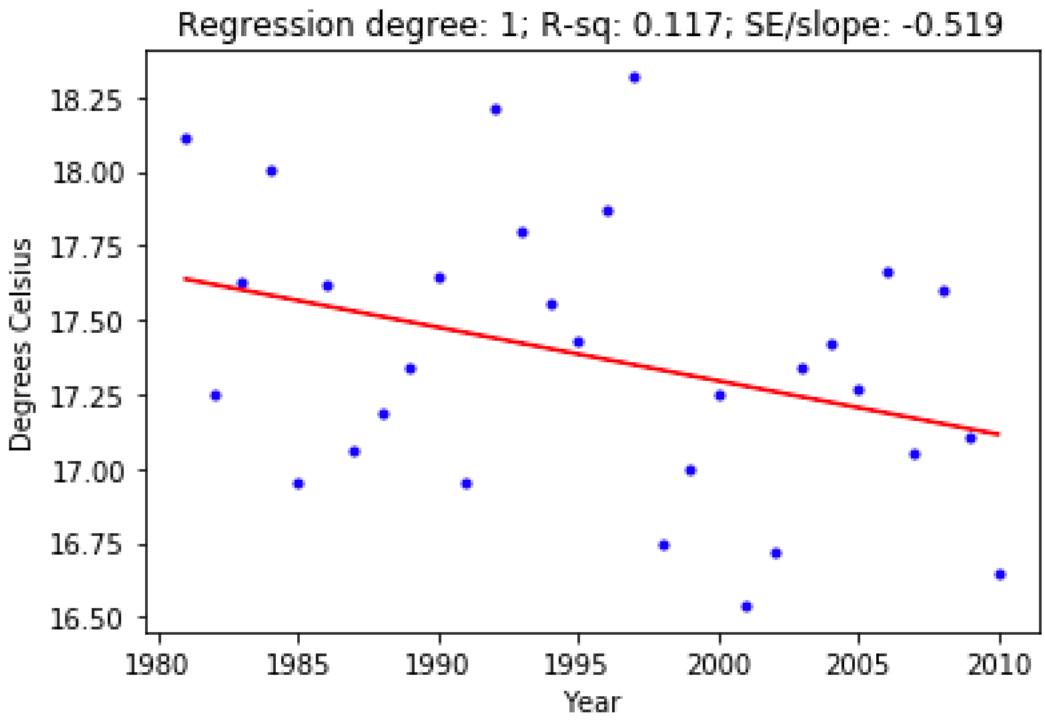
Plot 5.A *Increasing Interval (Los Angeles, length=30)*



5.1 What was the start and end year for your window? What was the slope?

Start year was 1961, end year was 1990. Slope was 0.0297 degrees Celsius per year.

Plot 5.B *Decreasing Interval (Los Angeles, length=30)*



5.2 What was the start and end year for your window? What was the slope?

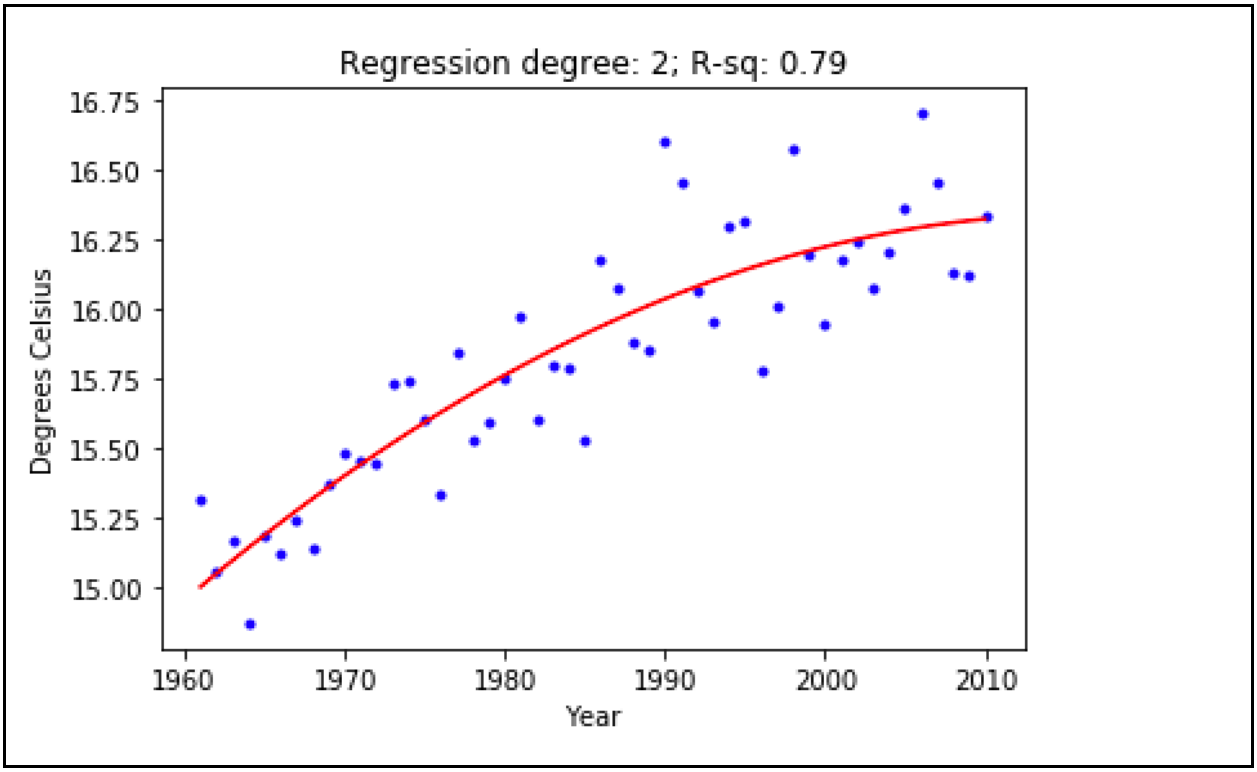
The start year was 1981, end year was 2010. The slope was -0.018 degrees Celsius per year.

5.3 Considering *both* plots, what conclusions might you make with respect to how temperature is changing over time?

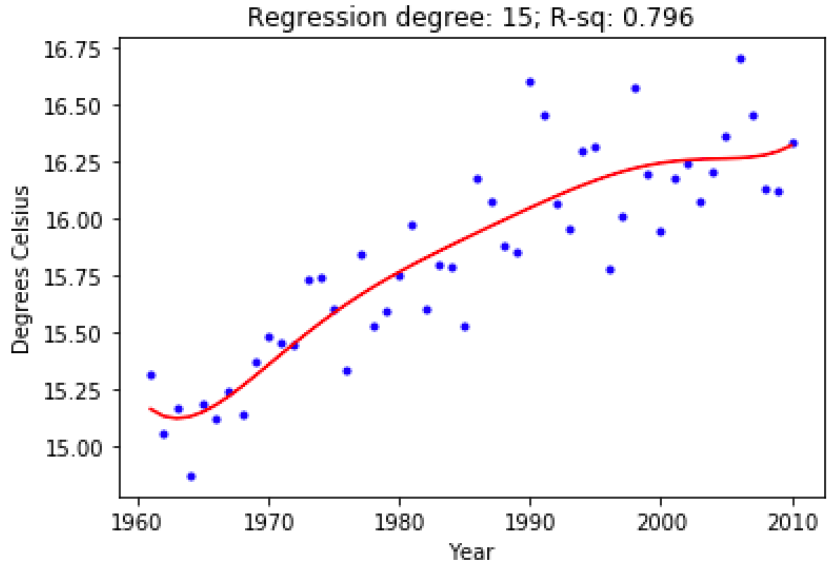
In earlier years, temperature in LA seems to be increasing, but in more recent years, temperature seems to be decreasing.

Problem 6

Plot 6.A *Training Data, Degree 2*



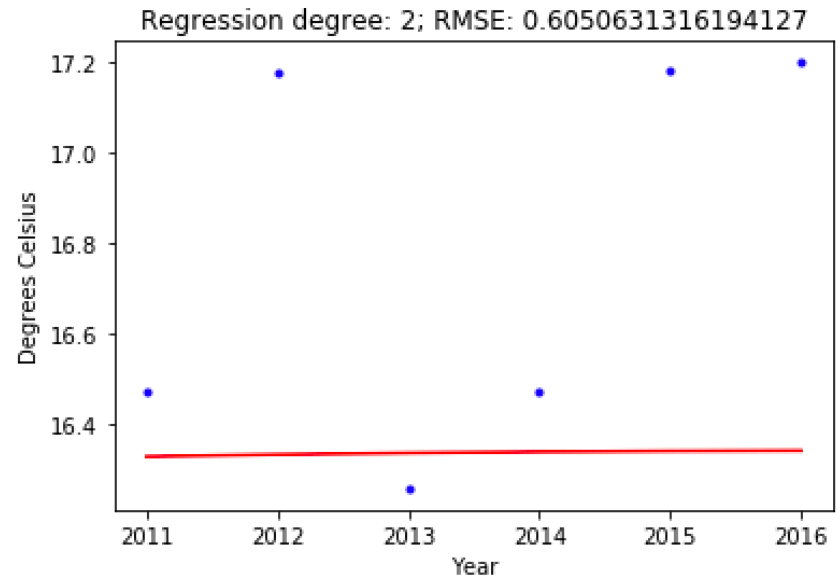
Plot 6.B *Training Data, Degree 15*



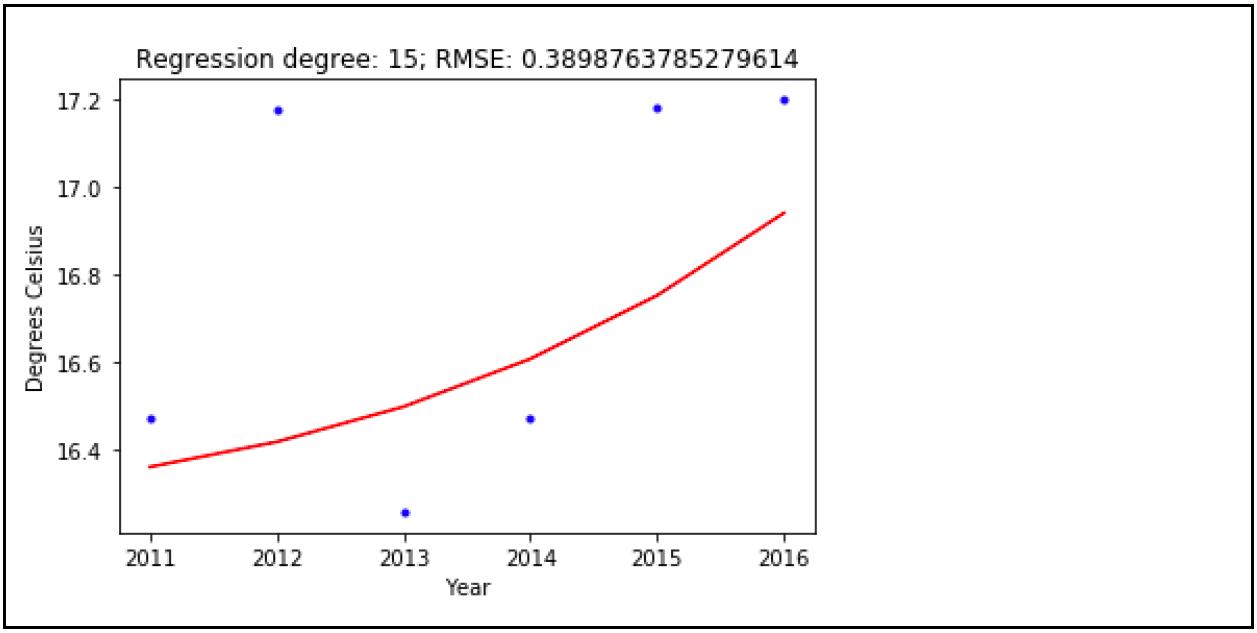
6.1 How do these models compare to each other in terms of R^2 and fitting the data?

The R^2 for degree 15 is about the same as that for degree 2, only a little bit higher, so the fit for the degree 15 model is only slighter better than that of degree 2.

Plot 6.C *Test Data, Degree 2*



Plot 6.D *Test Data, Degree 15*



6.4 Which model performed the best? Which model performed the worst? Is this different from the training performance in the previous section? Why?

The degree 15 model performed the best while the degree 2 model performed the worst because the degree 15 model had an RMSE of 0.39, which is lower than the RMSE for the degree 2 model, which means that the predicted values for the degree 15 model were closer to the actual values.

6.5 If we had generated the models using the data from Problem 4B (i.e. the average annual temperature of Boston) instead of the national annual average over the 22 cities, how would the prediction results on the national data have changed?

Since Boston has colder temperatures than the national average in general, using just Boston as the training data would predict lower temperatures for the national temperatures.