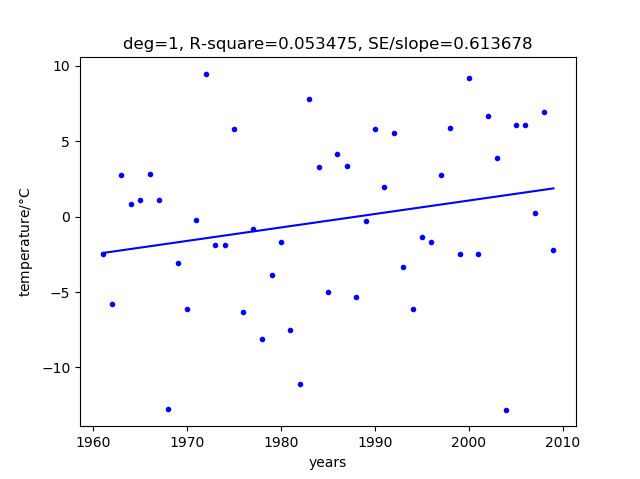
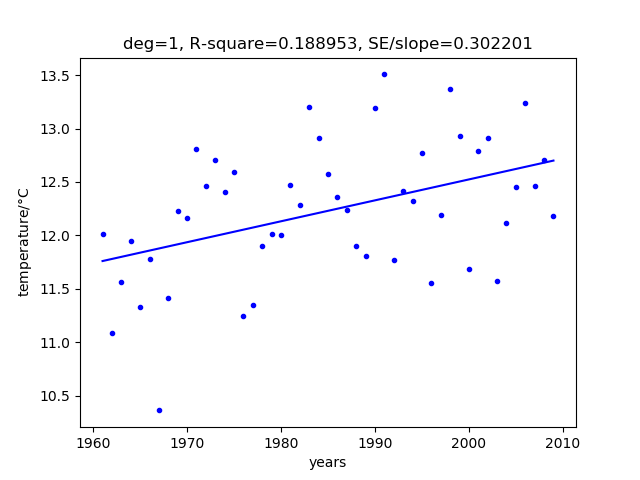
PS5\_answer

Prob A.4.I Temperature on January 10th for New York from 1961 to 2009



Prob A.4.II Averaged yearly temperature for New York from 1961 to 2009

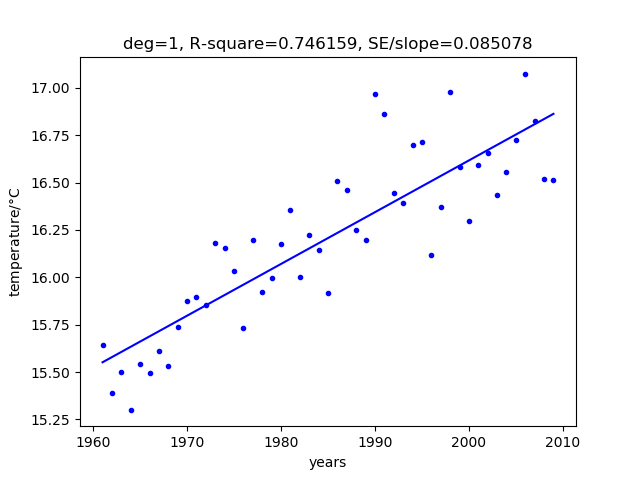


A(a). The fitting result for yearly average temperature vs years is better than temperature on January 10th vs years since it has larger R2 and smaller SE/slope.

A(b). These graphs are noisy because their R2 is so small. The first one is more noise.

A(c). We can see that in the second figure, the fitting has 0.3 standard-error-to-slope ratio, which means the standard error is not large compared to slope. So the slope could be taken as effective and the yearly average temperature is increasing slightly with years.

Prob B

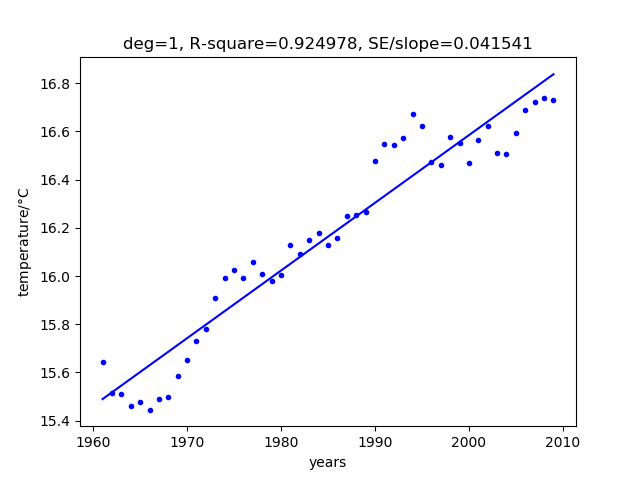


B(a), (b) The R2 is much larger and the SE/slope is much smaller than the two cases in Prob A. Therefore the fitting result here is more acceptable. We can see an obvious trend of temperature increasing as time goes from 1961 to 2010.

B(c) If it only includes 3 cities, the R2 would be smaller than the result of 21 cities. If it incorporates 100 different cities, the R2 would be larger than the result of 21 cities.

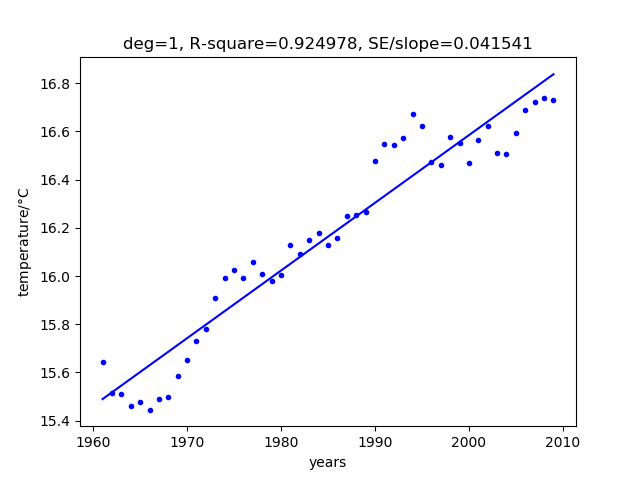
B(d) If all the cities are in the same region, for example, in New York state, then the situation of yearly average temperature for New York city would be similar to yearly average temperature for New York state since the temperature difference is not big inside New York state.

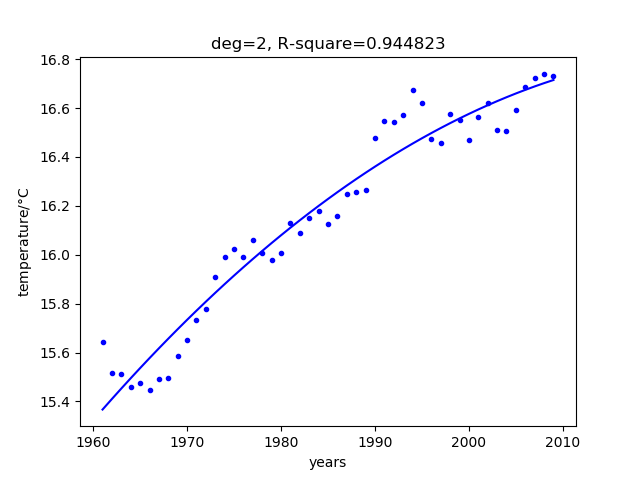
Prob C

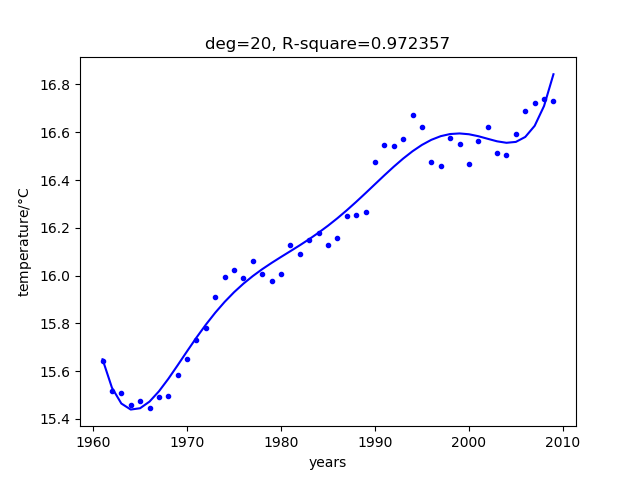


C(a), (b). The R2 is 0.92 for 5-year moving average temperature, which is much better than result in case A and B. The SE/slope is smaller than that in case A and B. The trend for temperature increasing is more clearly shown in this graph.

Prob D.2.I





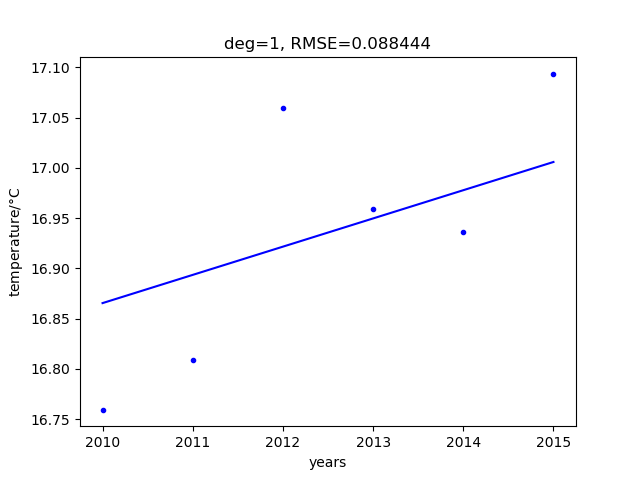


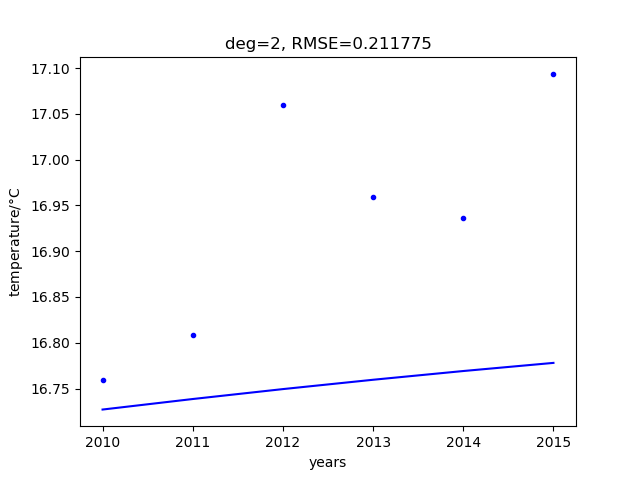
D.2.I (a) They all have good fitting result and their R2s is higher than 0.92.

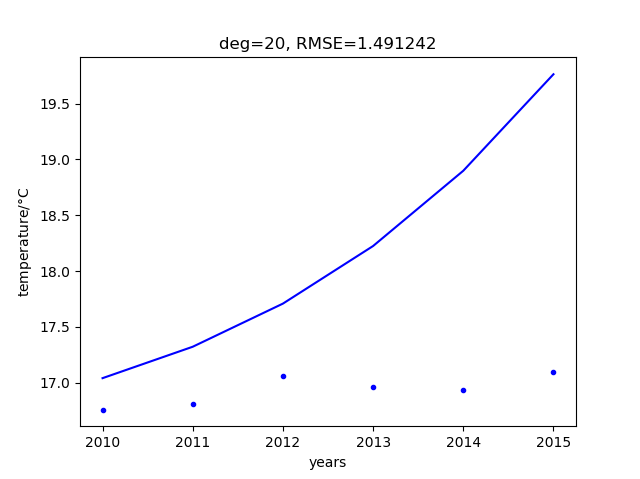
D.2.I (b) The deg 20 has best R2, because it could modify more parameters in the polynomial formula to fit the data.

D.2.I (c) The deg 20 model has the best R2, so it fits the data best. However, the deg 20 model may be overfitting, and it needs to be further tested.

Prob D.2.II





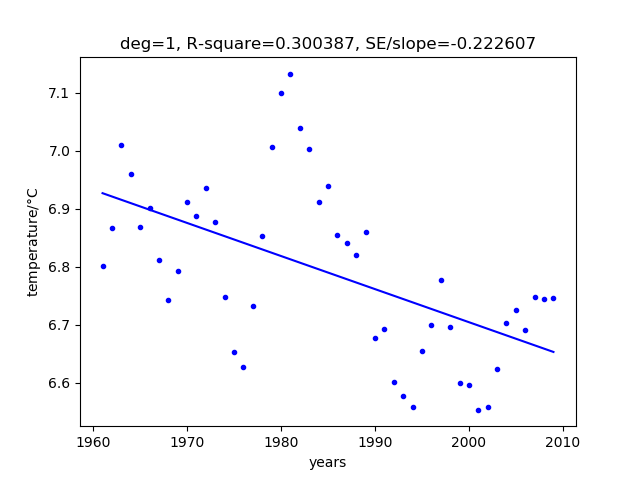


D.2.I (a) These models predict the data differently. The deg 1 model has the smallest RMSE, the deg 20 model has the largest RMSE.

D.2.I (b) The deg 1 model perform the best prediction and the deg 20 model predict the worst. This is reversed situation compared with D.2.I. Because in the model fitting process, deg 20 model is overfitting, which makes it terrible for predicting data.

D.4.I (c) Since the temperature data of New York city is very fluctuate, so all these three models with degree 1, 2, and 20 are not good to fit. And therefore, all of their predicting abilities are also bad.

Prob E



E(a) The temperature variation drops a little as time goes.

E(b) (1). We can do statistics by dividing cities into two categories, hot region and cold region to see if both regions are becoming warmer.

(2). Enlarge the domain of investigation to the global region. Take more data from cities of other countries, not only in the USA.