Homework 1

For this and all homeworks, you must explain your approach and how you got to the answer.

1. Wilks exercise 2.2.
2. Wilks exercise 2.5.
3. Perform an exploratory data analysis of the T_x and T_n time series, as well as their difference (the diurnal temperature range). Write a summary paragraph describing what you've found both statistically and scientifically. There is not a "right" way to do this data exploration, but you should at least create/examine:
   (a) histograms and kernel density estimates of the daily data as a function of month or season (i.e. create four or twelve subplots showing the empirical distributions of the daily data; do not temporally average the data). Bonus question: try using a bin size of 0.5°C. You may notice strange behavior in your histograms. Any idea why?
   (b) summary statistics for winter T_x including the mean, median, standard deviation, mean absolute deviation, representative quantiles, skewness, and kurtosis. Calculate each of these quantities without using the built in functions in your language of choice, and print out the results. You can drop any missing values before doing the calculations.
4. Read Wilks section 3.4. Apply the Box-Cox transformation to winter (DJF) T_x such that it is closer to a normal distribution. Provide justification for your choice of parameters in the transform, and discuss the result (i.e. how successful was your transform).
5. For the winter (DJF) T_x data,
   (a) Fit a gamma distribution using Thom's approximation to the maximum likelihood estimates (see Wilks section 4.4.3).
   (b) Fit a gamma distribution using the built-in functions of the software package of your choice. How similar/different are they? How did the software package calculate the parameters? Be careful of different parameterizations of the gamma.
   (c) Often temperature data is provided in tenths of degrees Celsius, so that it can be stored as integer values. Without converting individual data values, determine the values of the two parameters that would have resulted if the data had been expressed in tenths of degrees Celsius, rather than degrees Celsius.

(d) Construct a histogram of the temperature data, and superimpose the fitted gamma function.

(e) Comment on the goodness of fit of fit of the distribution, including creating a quantile-quantile plot (see section 4.5 in Wilks)

6. On August 13, 2013, the maximum temperature recorded in Los Angeles was 33.3°C. Use the entire Los Angeles T_x data set to provide context for such a warm temperature.

   (a) Find the fraction of JJA days that were at least this warm.

   (b) Fit a normal distribution to JJA T_x. Convert the 33.3°C value into a standard anomaly. Use the table in the back of Wilks to find the probability of observing a day at least this hot, based on the normal fit. Interpret the result, describing and discussing the appropriateness of any assumptions that you have made (form of the distribution, independence, stationarity).

   (c) Read the Wilks section about the generalized extreme value distribution.

   (d) Find the maximum value of T_x for each year in the data set. Plot as a histogram and demarcate the value from August 13, 2013 in some way. Find the fraction of years that feature a maximum T_x hotter than 33.3°C. Interpret this result, and in particular discuss how this approach avoids the issue of very warm days tending to be tightly clustered in time that may mar the interpretation in (a) and (b).

7. Wilks exercise 5.1.