

Statistical Exploration of Climate Data

Tamra Carpenter
Jon Kettenring
Robert Vanderbei

Instructor Module

Version 1
January 5, 2013

Module Summary:

The students will learn some basic concepts in statistical thinking about data, with emphasis on exploratory data analysis. The module will analyze daily temperature data collected over 55 years at a single location – McGuire Air Force Base (AFB) in southern New Jersey. The analysis explores the question, “Is there any observable temperature trend over this time period at McGuire AFB?” The challenge is to see a potentially small change within a data set that has both seasonal variability and high daily variability. We will do basic plots to help the students view data in different ways, introduce methods for removing seasonality, and use averaging to reduce day-to-day variability.

This module might be viewed as a “case study” in data analysis. It will give students a taste of what it’s like to do “real world” data analysis. Students will work with a large noisy data set and look at it in different ways to try to answer a specific question. The module does not, however, provide an answer to the question on temperature change that it addresses – it is about the *process* of data analysis. Each individual analysis (corresponding to a figure in the module) leads us to a new set of questions, which in turn leads to further analyses. This is often the way data analysis proceeds in practice. As the adage goes, “It’s not the destination, it’s the journey.”

This module is created in association with the Mathematics of Planet Earth project.

Target Audience: Introductory undergraduate statistics students; students in a first course on exploratory data analysis.

Prerequisites: Graphing, basic statistical ideas like averages, medians, and variance. It would be helpful for the students to be familiar with basic mathematical notation, such as summations and subscripting to denote terms in a series. **This is especially true for the discussion on moving averages in Inset 3.**

Mathematical Fields: Statistics, specifically exploratory data analysis, graphical data analysis, and very basic ideas in viewing and working with time series data.

Application Areas: Climate Analysis

Notes to the Instructor on Use of Module:

There is both an instructor version and a student version of the module. The student version is a subset of the instructor version, created by deleting all text that appears in red or blue in the instructor version.

If the class has already learned most or all of the basic techniques prior to doing the module, you may want to have them read it prior to class, so that you can jump into discussions. If they have not previously seen this material, you may want to begin it in class before giving the module to them.

The module is driven by the data analysis steps represented in the figures within the module. There are questions following each figure that you may want to use to stimulate class discussion. The blue text in the Instructor Module gives some sample answers, but most of the discussion can be very open-ended. These are just sample thoughts.

The text that appears in red in the Instructor Module provides some notes that are not directly related to the discussion of the figures but may be helpful to the instructor.

This module was inspired by a study on the same data by Robert Vanderbei [7]. You may want to look at this paper for some other ideas related to analyzing this set of data. In particular, the cover picture is taken from that paper and is based on looking at weather data from NOAA collection sites around the world. Red dots are places with positive temperature trend; blue had negative trend; and white had no perceptible trend.

Approximate Length: While these materials can be adapted for a variety of contexts, we envision two main classroom uses. The first is as a classroom discussion driven by the different plots, without individual hands on exploration of data. Used in this way, the module should take roughly one 70-80 minute class period. This may be the best approach for use with less advanced students. With less advanced students, you may also prefer to skip any discussion of the “Insets.” If used in a course on data analysis, you may want to spend more time on the material and allow students to explore a different data set on their own or in small groups.

Technology Software Needs: This module is presented in a way that allows anything from no use of technology to use of a statistical or mathematical package that allows an instructor to duplicate all of the examples of the module on his/her own and replicate them on other similar data sets. Without technology, the instructor will simply use the plots generated in the text and discuss their features with the class.

For instructors wanting students to interact with data, we provide the MATLAB code to generate each of the individual plots discussed in the text. Thus, the instructor can easily reproduce all studies in the module using MATLAB or reprogram the analyses for some other package. There is a MATLAB file associated with each of the figures in the module that contain data plots. Each such figure has a corresponding MATLAB file called

FigureXX.m, where XX is replaced with the figure number. The MATLAB code is provided in the Appendix to this module.

The MATLAB code is “no frills”, so it does not do things that more polished codes would, such as check for missing data. The name of the McGuireAFB data set is hard-coded in each file, so to use them with other data sets, you will need to change the data file name in the first two lines of each MATLAB file.

The file for Figure 7 cannot be used with any other data sets (without further modification) because the intercept and slope values for the trend lines are hard-coded and specific to the McGuire AFB data.

Data Sets: The temperature data from McGuire Air Force Base that is used throughout the module is provided in the accompanying file called McGuireAFB.dat. Other data sets to allow students to rerun the analysis for different locations also accompany the module. The initial release of the module includes additional data sets for temperatures recorded in Fairbanks Alaska, New Orleans Louisiana, and Raleigh/Durham North Carolina. The associated files are called respectively, Fairbanks.dat, NewOrleans.dat, and Raleigh.dat.

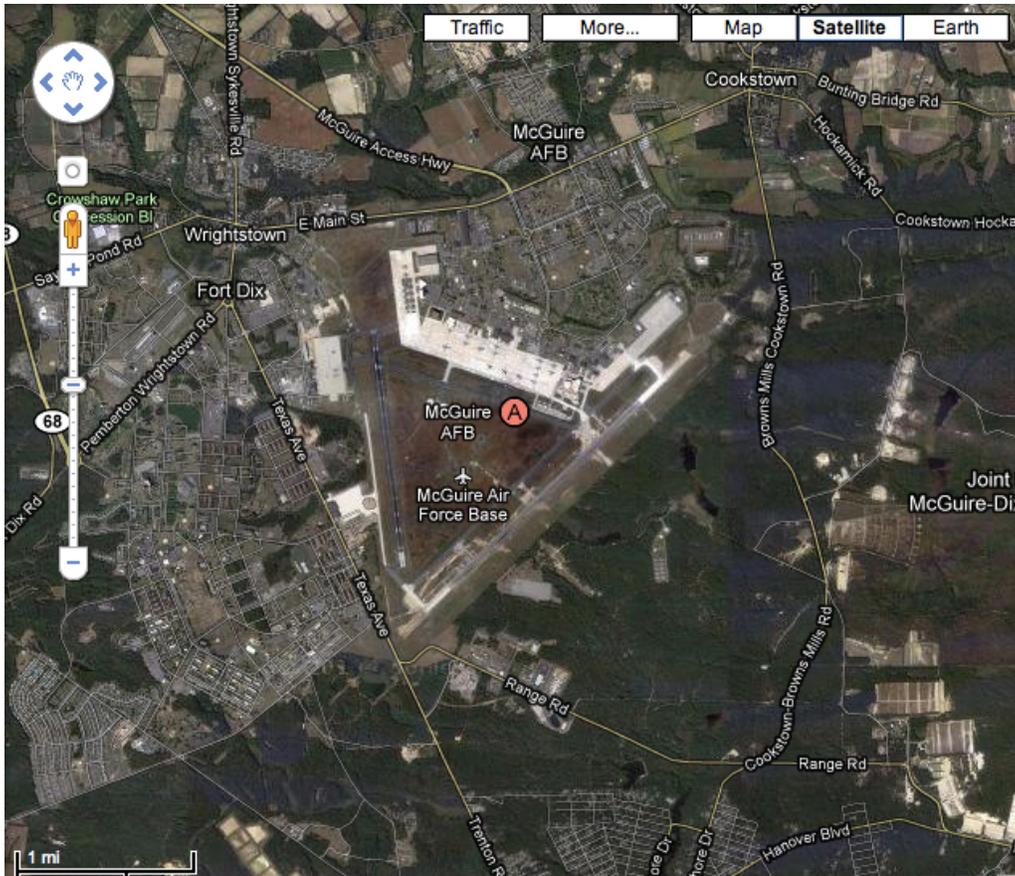
The data files are in two-column plain text format. The first column contains a date and the second column contains the average temperature for that date in degrees Fahrenheit. The files start on January 1, 1955 and proceed chronologically to the most recent dates. The dates are expressed as year, month, day, so that January 1, 1955 is 19550101.

Adventurous users can explore the National Oceanic and Atmospheric Administration (NOAA) website [5,6] for additional (or updated) data sets, but should be careful to note that many files contain missing data, and many locations have data split between files (which is the case for McGuire AFB). Included with the module is the Unix shell script we used to grab the annual data files for Fairbanks, New Orleans, and Raleigh/Durham and then assemble the relevant pieces of data (date and temperature) into a single file for each location.

List of Module Files: This module includes an instructor version (weather-module-final-instructor-V1.pdf), a student version (weather-module-final-student-V1.pdf), the McGuire AFB temperature data set (McGuireAFB.dat), and a presentation containing each of the module figures (Module-Figures-for-Presentation.pdf). Additional files (Fairbanks.dat, NewOrleans.dat, Raleigh.dat) containing temperature data at other locations are also provided but are optional. We expect to add more files at other US locations in the future. Finally, the shell script we used to extract the data files for Fairbanks, New Orleans and Raleigh-Durham from the NOAA website is also included and called getDIMACSdata.sh.

One of the many quotes (or possibly misquotes) attributed to Yogi Berra says, “You can see a lot just by observing.” This module applies that principle to data analysis.

It examines daily average temperature data collected from January 1, 1955 to August 13, 2010 at a weather station located at McGuire AFB in southern New Jersey. The data set contains average temperature readings for a total of 20,309 days. This is a relatively large amount of data that presents a variety of real data analysis challenges. The module will walk through an approach for deciding how to view a relatively large amount of data containing seasonality and high day-to-day variability.



McGuire AFB

Throughout the module we use the McGuire AFB temperature data set to try to answer the question: *Has it gotten warmer at McGuire AFB or not?*

Note that we are not trying to identify the causes for any change. We are simply asking whether we can see a change.

Once you finish this module, you will see that Yogi Berra was correct – you really can see a lot just by observing. You can also use the same techniques to explore the same question at other locations. There are many weather sites around the world that collect similar information, so this same analysis can be repeated to answer the same question

for locations all over the world. (You can learn more about the available data from the National Oceanographic and Atmospheric Administration website [5].)

A good place to start is just plotting the data. A plot of the average temperature (in degrees Fahrenheit) at McGuire AFB by day is given in Figure 1.

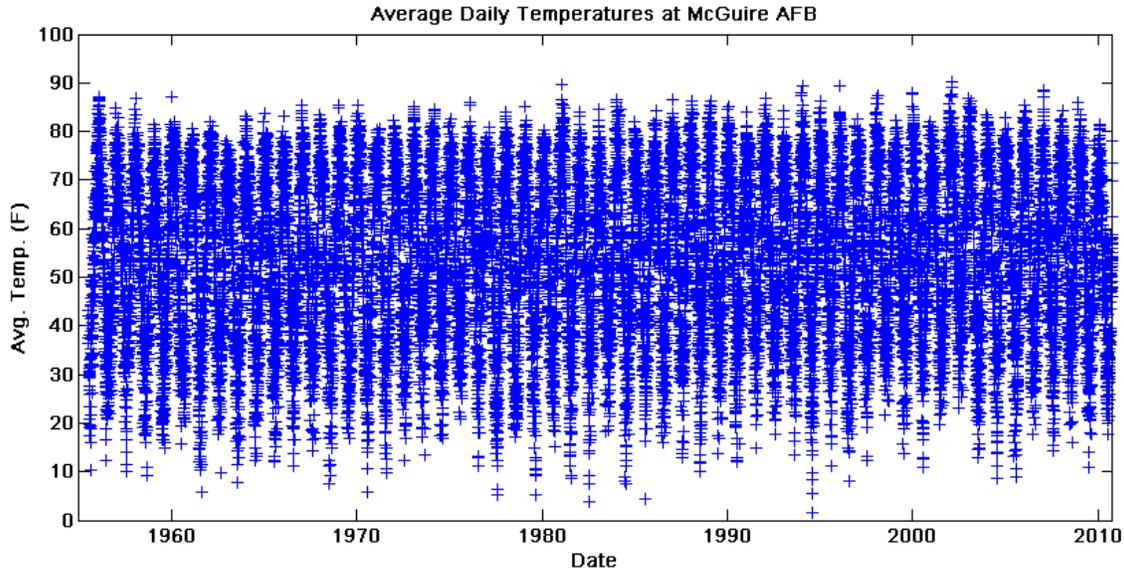


Figure 1: Plot of average daily temperature at McGuire AFB over time

Discussion related to Figure 1:

a) What can you learn from this plot?

There is a lot of data! (Too much to see what's really going on.)

There is a large amount of variability and a strong seasonal effect – summers are warm and winters are cold and there are 55 of each. (It may even be helpful to ask the students to count them. Doing so will give a sense for the year-to-year temperature variation.)

You can get a sense of the range of the average daily temperatures – largely between 0 and 90 degrees Fahrenheit. (Keep in mind that these are average daily temperatures, not highs or lows. The daily average temperature is defined as the average of the high and low reading for each day.)

Looking from left to right there does not appear to be an obvious trend in the data either increasing or decreasing over time.

b) Is this enough to conclude that there has not been any change in temperature?

Hopefully, everyone will agree that this is not enough analysis to draw any conclusion!

There is a large amount of data plotted over a relatively small area, which makes it hard to see what if anything is happening on average.

c) Can you suggest other ways to look at the data that might help see more clearly?

There might be a variety of suggestions to reduce the crowding of data. These could include plotting a subset of the data or plotting averages over longer periods of time, such as weekly, monthly, or yearly averages. We'll explore both of these ideas. Use of color might also be suggested.

One simple strategy might be to plot the daily temperatures for an early year in one color on top of the daily temperatures for a recent year in another color and see whether they appear offset or different in some way. Figure 2 does this for the years 1955 (in blue) and 2000 (in red).

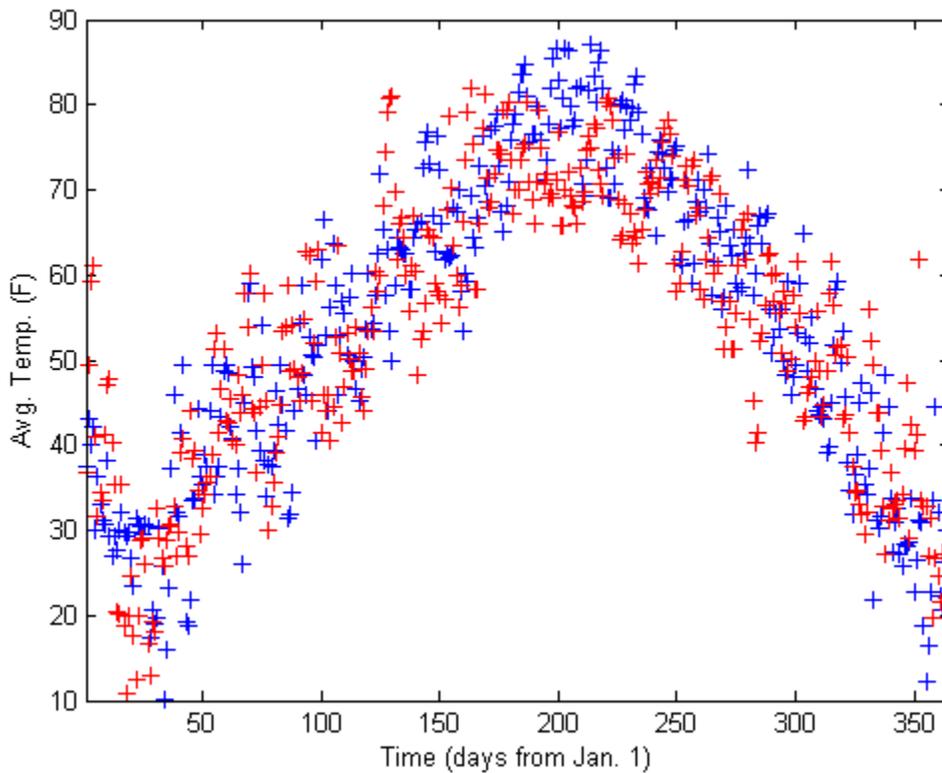


Figure 2: Daily temperature at McGuire AFB, year 1955 (in blue) and 2000 (in red)

Figure 2 discussion: What are some of the things that you see in this plot?

One year is not obviously warmer or cooler than the other. There is a large seasonal effect. The points are fairly tightly grouped in a band that rises and then falls with the seasons. There are a few points that lie substantially off this band, and more of these seem to be from 2000 than 1955. It appears that 2000 may have more variability than 1955. The hottest temperatures occurred in 1955, but the largest upward deviations from the central band occurred in 2000.

We could repeat this type of analysis including additional years in different colors or for different pairs of years, but these methods are only looking at a small amount of the data. Also, the large seasonal effects in Figures 1 and 2 would overshadow any (much smaller) trend in temperature that may have occurred. The methods that we'll look at next will give us less cluttered plots that are not dominated by seasonal variations, and they will use data over the entire 55-year period.

Boxplots are one way to summarize data to get a sense of the overall distribution. They display the median, upper and lower quartiles, and maximum and minimum values of the data. (Inset 1 provides a quick review of distributions and their quantiles.) The basic structure of a boxplot is shown in Figure 3. The “box” is delimited by the upper and lower quartiles, and emanating from the box are “whiskers” to the extreme values in the data. The placement of the median within the box and the relative length of the whiskers give a sense of the spread and skewness (which describes asymmetry) in the data.

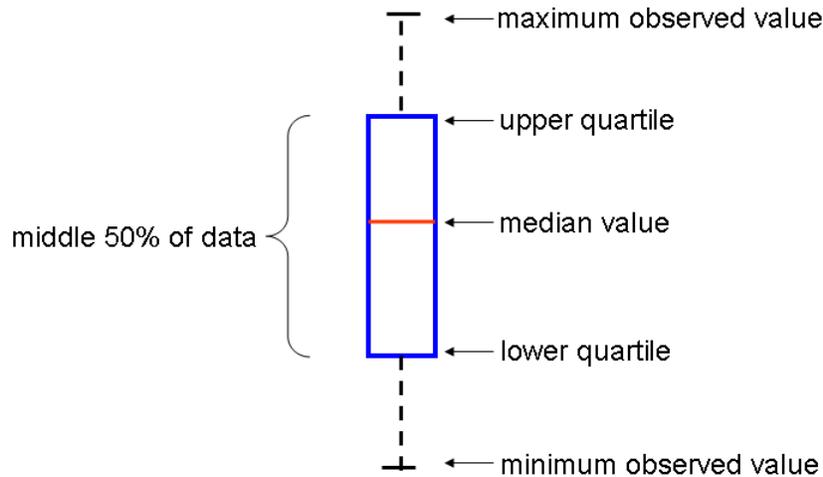
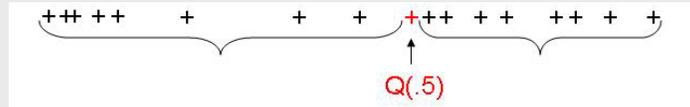


Figure 3: Anatomy of a boxplot

Inset 1: A Quickie on Quantiles. The quartiles and the median are special “quantiles” of a data distribution. The simplest way to think of a *quantile* is in terms of percentiles. When someone says their SAT score is at the 80th percentile, it means that they scored better than 80 percent of the people who took the test. The .8 quantile of a data set is a value such that 80% of the data values are below it and 20% are above it. If we let $Q(p)$ for $p \in (0,1)$ denote the p^{th} quantile in the data then $p*100\%$ of the data values fall below $Q(p)$ and $(1-p)*100\%$ are above it.



The *median* is the .5 quantile of the data set. It splits a data set so that there are an equal number of data values above and below it. The upper and lower *quartiles* are respectively the .75 and .25 quantiles in the data set.

The *interquartile range* is defined to be the difference $Q(.75) - Q(.25)$. Relating this to Figure 3, the interquartile range is depicted by the length of the box in the boxplot. We have based this discussion of quantiles on the material in [1], which provides a much greater context for using quantiles to explore data sets. See also [3] for related information. **In the future, we hope to expand this module to include an optional additional section on computing quantiles, order statistics, and using quantile plots. In particular, that section would provide a more precise treatment of how to compute the quantiles of a data set. Our goal here is simply to introduce the concept.**

Figure 4 shows 55 boxplots of the McGuire AFB temperature data – one for each year. The plots are arranged sequentially, proceeding from 1955 (year 1 on the left) to 2009 (year 55 on the right).

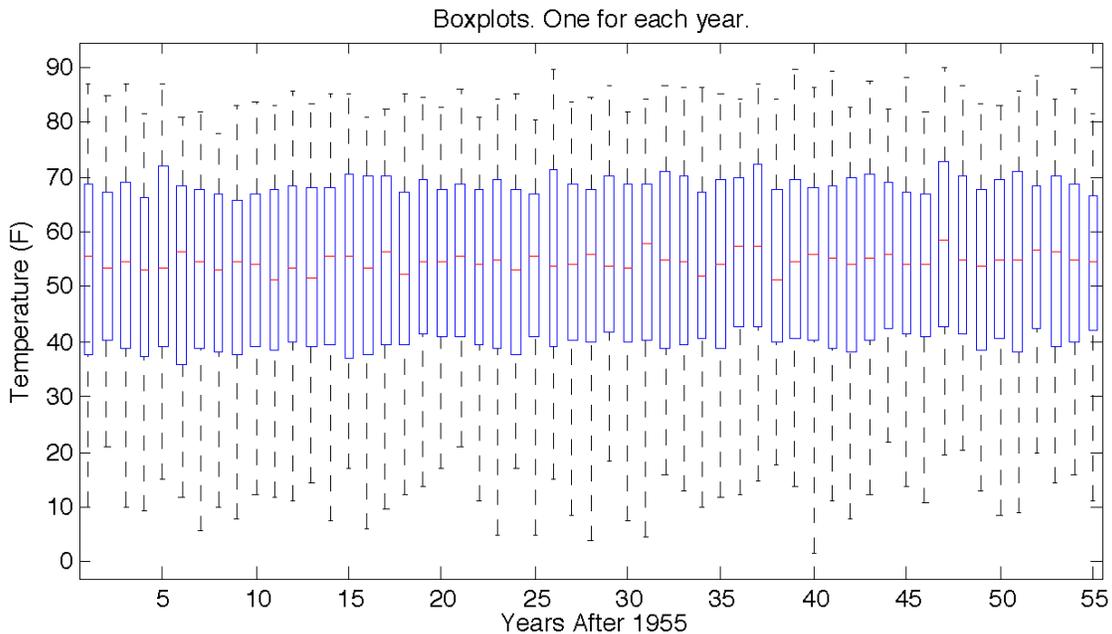


Figure 4: Boxplots of McGuire AFB temperature data by year

The boxplots reduce the original data from 365 data points for each year to just 5: the median (shown in red), the quartiles delimiting the box, and the extreme values. This greatly reduces visual clutter.

The sequential arrangement of the boxplots can help you get a rough sense of whether the distribution is changing over time. In particular, you can follow the red median “ticks” across the page to see whether you think the median is changing systematically over time. Related to our question of temperature change, there does not appear to be an obvious trend. The “boxes” show data variability, which also does not seem to be perceptibly changing over time. The length of the “whiskers” corresponding with the extremes values appear longer on the low side than on the high side, suggesting that the distribution has a heavier lower tail relative to the upper tail.

Teacher note: Continuing with the theme of graphical methods for data analysis, you could discuss skewness measures or explore the symmetry of the distribution using quantile plots, but since this is not directly relevant to the question of temperature trend, we have not included it here. This is something we hope to include in the optional extension that looks at quantile plots.

Discussion related to Figure 4:

a) Is Figure 4 helpful?

It has certainly reduced visual clutter and masked the seasonal variability, so in that sense it is helpful. It is easier to assess the individual yearly distributions and to compare them across time.

b) Does Figure 4 show anything that helps answer our question on temperature change?

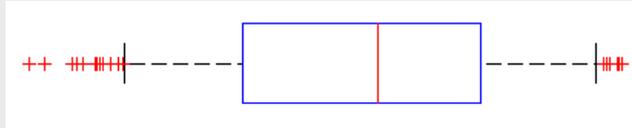
There does not appear to be a discernable temperature change, but you might want to ask the students to check for trend by “fitting” the red median ticks with a straightedge. In doing that, some students may notice a slight upward trend.

Some students may also notice that the scale on the temperature axis goes from 0 to 90 degrees. Since a temperature trend is likely to be quite small, a small trend will be difficult to see on this scale. This is an important observation and something that we will look at later in the module.

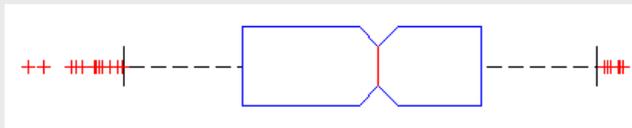
Boxplots can be “embellished” to provide more detailed information as discussed in Inset 2 below and in [1].

Inset 2: The Basics on Boxplots. There are several variations on boxplots that can provide additional information about a dataset. Some boxplots give more detailed information about the distribution’s tails. In such cases, the black lines at the ends of the whiskers may not necessarily extend all the way to the most extreme values.

These lines are called “fences”. Values beyond the fences are explicitly indicated in the boxplot and are called “outside values” (which may or may not be “outliers”). The fences are often defined to be the last data value within a window that extends above and below the interquartile range by a length that is a multiplicative factor of the interquartile range. In picture below, we used .5 as the multiplicative factor. More typically, that factor is 1.5.



A “notched” boxplot shows confidence intervals around the median. Thus, in comparing two datasets, if the notches about their medians do not overlap, then the medians are considered statistically significantly different.



When groups of boxplots are viewed together, still other variations can give a sense for the size of the respective datasets by adjusting the width of the respective boxes. Examples are shown in [1] and [4].

Another way to remove seasonality from data in a series through time is to compare points that should be the same with respect to seasonal effects. For instance, there should be no seasonal effect if you plot only the temperature readings taken on January 1 of each year or only those taken on August 17. This is another plot that you can try, but note that you would have 365 separate data sets. Let’s look for a more holistic approach.

The temperature readings are a sequence of observations in time, or a **time series**. A characteristic property of a time series is that the observations are not independent over time. Seasonality is one example of this lack of independence – you would expect the temperature on August 17, 2010 to be more like the temperature on August 17, 1955, than to the temperature on February 15, 2010. A simple model of data in a time series is to view each observation as being the realization of a random variable made up of a trend through time, (one or more) seasonal effects, and remaining effects that are not a function of time.

The temperature data has a seasonal component with a period of 365 days. Letting T_t denote the temperature reading at time t , the following differences remove the seasonal component:

$$D_t = T_t - T_{t-365}, \text{ (for } t = 366, \dots, 20,309\text{)}.$$

These differences are plotted in Figure 5. The red line through the plot just shows the zero value. If there were no trend in temperature, you would expect the differences to be randomly distributed about zero. If there were a linear trend in temperature, then the differences should be randomly distributed about the average yearly change.

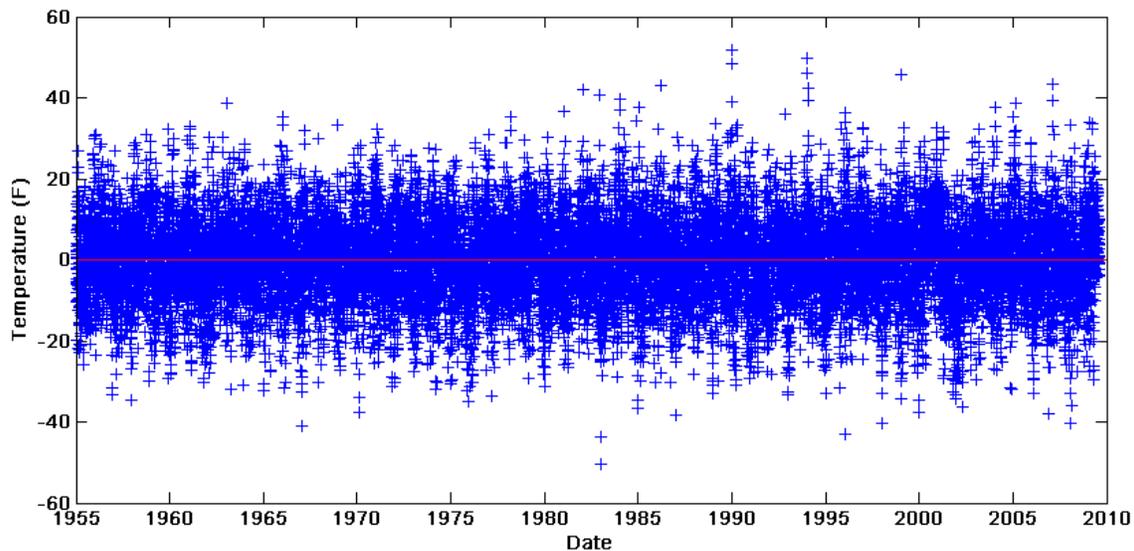


Figure 5: Plot of one year differences in McGuire AFB temperature data

Discussion related to Figure 5:

a) Do you see evidence of a trend in Figure 5?

Looking at the plot, there is no obvious trend up or down.

b) Why might you not be able to see a trend in Figure 5, even if one exists?

The range of the differences is large and the rate of annual temperature change is likely to be small if one exists. This would be very hard to see.

Perhaps the most striking feature in Figure 5 is that the range of the differences is large – there is roughly a ± 50 degree range. This is not due to seasonality, but simply reflects the large day to day variations in temperature. **Some students may be surprised that the range is still so large when seasonality is removed. We note that the variance of a difference of uncorrelated random variables is the sum of the variances; thus, differencing removes the seasonality but increases variance.**

If there were no trend in the temperature over time, you would expect the average of the differences to be zero. The average of the differences is actually 0.0289 °F. This seems

small, but note that it is the average *annual* change. Viewed over the 55 years of observations, it translates into a 1.59 °F increase in average daily temperature at McGuire AFB. Alternatively, it translates into an increase of 2.89°F per century. (This, by the way, appears to be consistent with EPA analysis from 1901-2005:

http://www.epa.gov/climatechange/science/recenttc_tempanom.html) Viewed another way, the 1.59 °F change at McGuire AFB is about the same as the difference in average annual temperature between New York City and Philadelphia [6].

This may or may not be statistically significant, but it does suggest that something may be happening to the temperature at McGuire AFB; moreover, it illustrates how difficult it is to see small trends in highly variable data.

Teacher note: As an aside, the standard deviation of the differences is 10.460 °F, while the standard deviation of the mean of the differences is 0.074 °F. If it fits with other material in the class, instructors may want to discuss significance and the fact that small differences in a large data set may still be significant. However, in this case the standard deviation relative to the mean is high.

Some students may ask whether and how we handled leap years in this differencing. These differences are based on a 365-day year, not the 365.25 day orbit of the Earth around the Sun. This means that each difference covers 0.25 days less than a whole year and that particular quarter of a day shifts relative to Earth's orbit by 0.25 days per year for three years and then gets reset to where it started. Hence, the effect should be very small.

We are searching for a small signal (in this case a temperature change), if any, within data that are very noisy (due to day to day variations). One way to smooth out some of this variability is to use averaging. Figure 6 plots yearly average temperature over time. (The red +'s are the annual average values.)

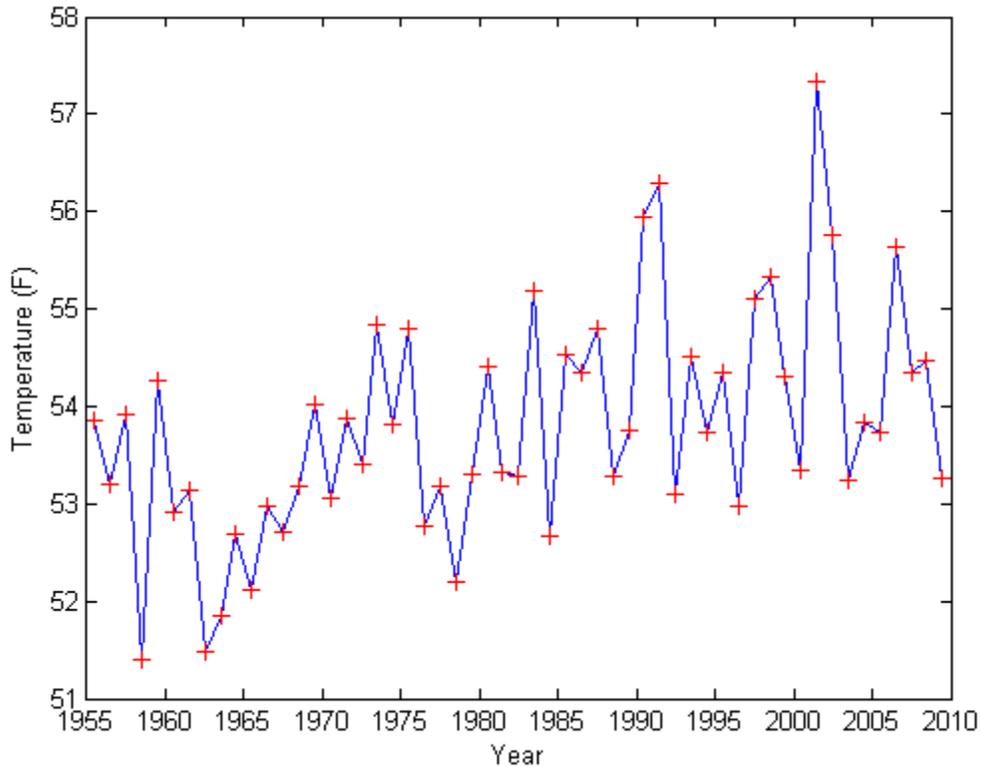


Figure 6: Plot of yearly average temperature at McGuire AFB

Discussion related to Figure 6:

a) Do you see a trend in this plot?

In this case, it does appear that there may be a slight upward trend, but it is still dominated by lots of variability. Again, you might want to ask the students to use a straight edge to try to fit a line through the red ticks. This time most will probably see an upward trend.

b) Why might you possibly be able to see a trend in this plot when you could not see one previously?

Variability is reduced considerably – the range for the plotted values is now only about 6 °F, so a small change will be easier to detect. By averaging over a one-year interval, each point represents a full year of data, so seasonality is also removed.

To help us see whether there is a trend in Figure 6, we can overlay “trend lines” as shown in Figure 7. The solid red line in Figure 7 is the one that minimizes the sum of the absolute deviations between the line and the data values, and dashed line minimizes the sum of squared deviations. In both cases, the slope of the trend line indicates an increase of over 3 degrees per century. More specifically, the line that minimizes the absolute deviations has a trend of 3.23 degrees per century and the one that minimizes squared deviations has a trend of 3.68 degrees per century. The method for fitting these lines is beyond the scope of this module. We use them here only to help decide whether we see a trend.

Teacher note: If you use the MATLAB code with other data set, Figure 7 is not directly applicable because the slopes are hardcoded for McGuire AFB.

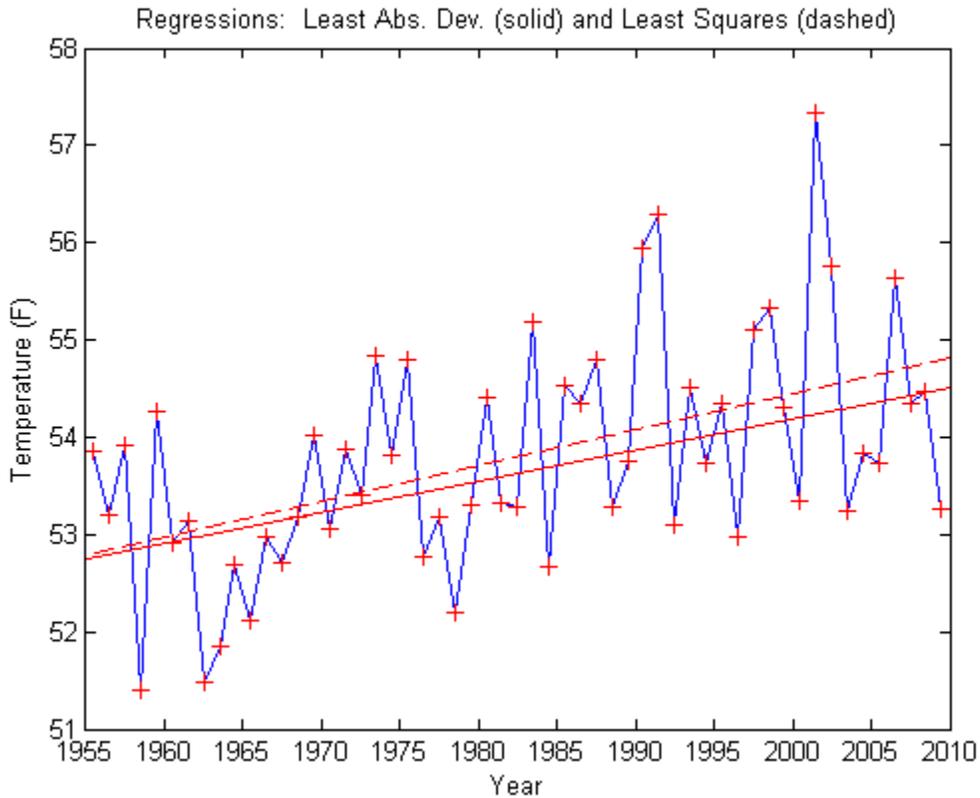


Figure 7: Plot of yearly average temperature at McGuire AFB with regression lines overlaid

Question related to Figure 7: Can you explain why minimizing squared differences would result in a line with greater trend?

Minimizing the sum of squared deviations would be more influenced by the largest deviations, and in this case, the largest deviations are from high values particularly the one in the year 2002 and those in 1991 and 1992.

This could be a logical place to stop, but if time permits, you may want to include the following discussion of moving averages.

Both the boxplots in Figure 4 and the yearly averages in Figures 6 and 7 removed the annual seasonal effects by considering the data in one-year chunks. In “chunking” based on calendar year, they also greatly reduce the number of data points – the yearly averages replace 365 data points with a single point, while the boxplots represent a year with five observations as described above.

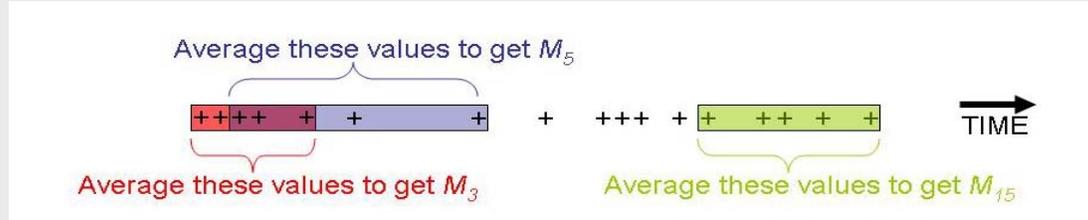
There is no reason that we can only average a calendar year at a time. Taking moving averages is a common way to smooth out short-term fluctuations in a time series while still preserving the slowly varying trend. A **moving average** of order z , computes

averages over “sequential chunks” of z observations moving through time. A moving average is itself a time series computed from the original series.

Inset 3: A Meander on Moving Averages. Moving averages are often used in looking at noisy sequential data, so let’s take a closer look at how they are computed. Let T denote the original time series and let T_t be the reading at time t . (In our case T is the series of temperature readings at McGuire AFB, so T_t is the average temperature on day t .) To compute the moving average series (call it M) of order $2s + 1$, we’ll compute each value M_t as the average of the $2s+1$ values of T centered at time t . In other words, to get M_t we compute the average of T_t together with the s readings before it and the s readings after it.

$$\text{Thus, } M_t = \frac{1}{2s + 1} \sum_{u=-s}^s T_{t+u} \text{ for } t = s+1, \dots, n-s \text{ (where } n \text{ is the length of } T\text{).}$$

The figure below illustrates calculating selected elements of a moving average of order 5 (which is $2s+1$ when $s=2$) in a data set with 17 observations.



To obtain the next point in the moving average, move the “time window” ahead to include the $2s + 1$ data values centered at T_{t+1} . You could use the formula above to compute M_{t+1} , but note that as you slide the time window ahead to be centered on $t+1$ the “oldest” element leaves the window as a new one is added. Thus, the time windows for M_t and M_{t+1} contain $2s$ common elements. This means that we can efficiently calculate M_{t+1} from M_t as follows:

$$M_{t+1} = M_t + \frac{1}{2s + 1} (T_{t+s+1} - T_{t-s})$$

When you are done, the series M_t will contain $2s$ fewer values than the original data set, but the short term fluctuations will be smoothed out.

The series of one-year moving averages for the McGuire AFB temperature data is plotted in Figure 8.

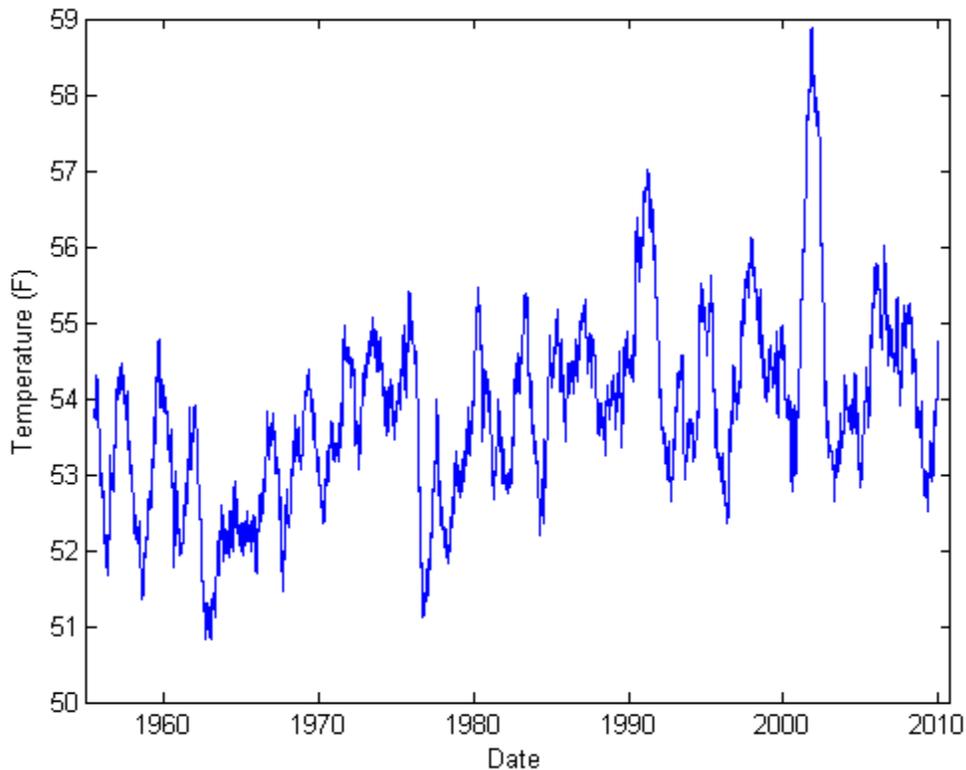


Figure 8: One year moving averages on the McGuire AFB temperature data

Discussion related to Figure 8:

a) What do you notice in this plot?

Again, we can notice that averaging reduces the range of observations – the range here is about 8°F as compared to about 90 °F in the original plot in Figure 1. It appears that there is a slight trend moving right.

b) Can you explain why this plot, in fact, contains all of the data points from the plot in Figure 6?

The points plotted in Figure 6 are just averages over the one year “chunks” that exactly correspond to a calendar year. Figure 8 plots averages over all one-year chunks, including those that correspond with calendar years.

Figure 9 shows the data points from Figure 6 on top of the one in Figure 8. You can observe that the moving average gives you a better sense of the temperature evolution between the red data points, and you can also see that the yearly averages don't quite catch the temperature peaks such as the low average temperatures in 1977 and the highs in 2002.

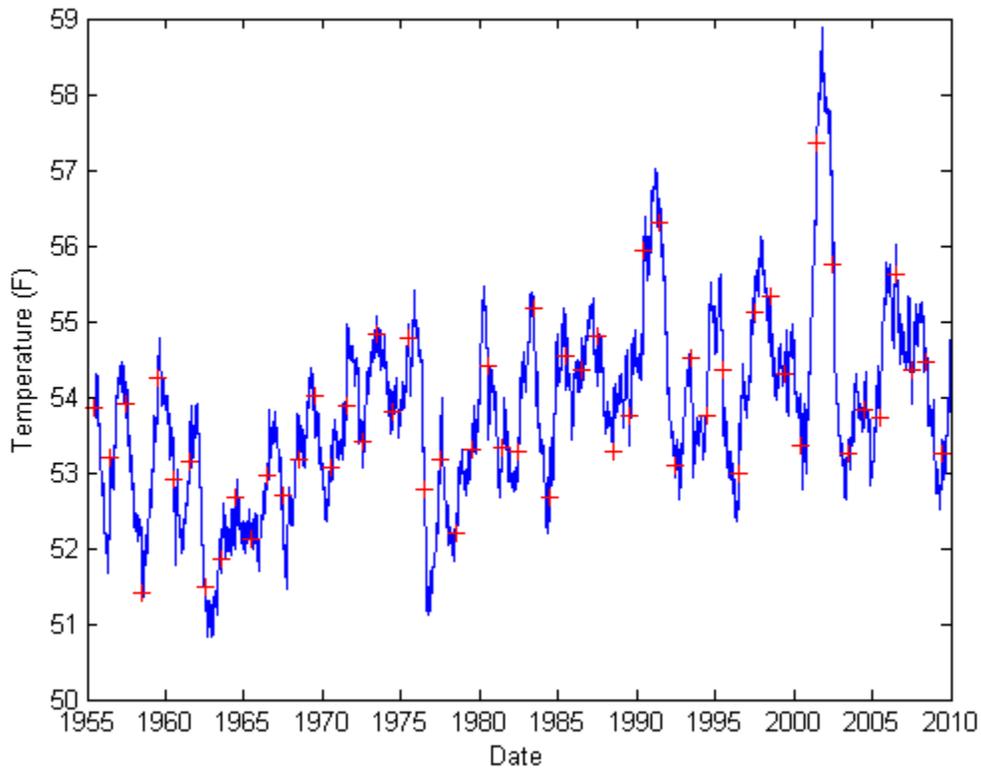


Figure 9: Annual average temperature (red) overlaying one-year moving averages (blue)

We have come pretty far in exploring this data.

Final discussion questions:

- a) Do you think you have a better sense for the data now than you did at the beginning? Explain.
- b) What questions about the data might you want to ask next?

Final Projects:

- a) The module has not answered the question we began with: “Is there any observable temperature trend over this time period at McGuire AFB?” What do you think? Support your position with evidence from the graphs.
- b) Use data from another location to conduct a study similar to what was done for McGuire AFB. Do you think there is any observable temperature trend at this location?

References

- [1] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. Tukey, “Graphical Methods for Data Analysis,” Duxbury Press, Boston, MA, 1983.
- [2] M. Frigge, D. C. Hoaglin, B. Iglewicz, “Some Implementations of the Boxplot,” *The American Statistician*, **43**(1), pp. 50-54, 1989.
- [3] R.J. Hyndman and Y. Fan, “Sample Quantiles in Statistical Packages,” *The American Statistician*, **50**(4), pp. 361-365, 1996.
- [4] R. McGill, J. W. Tukey and W. A. Larsen, “Variations of Boxplots,” *The American Statistician*, **32**(1), pp. 12-16, 1978.
- [5] NOAA, Climate data format and download instruction, 2011.
<ftp://ftp.ncdc.noaa.gov/pub/data/g sod/readme.txt>.
- [6] NOAA, Data Tables, Normal Daily Mean Temperature, Degrees F,
<http://www1.ncdc.noaa.gov/pub/data/ccd-data/nrmavg.txt>.
- [7] R. J. Vanderbei, “Local Warming”, *SIAM Review*, 54(3), pp. 597-606, 2012.
Available on-line,
<http://www.princeton.edu/~rvdb/tex/LocalWarming/LocalWarmingSIREVrev.pdf>.

Appendix: MATLAB Files for Producing Figures in Text

Figure 1

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);

figure(1);
time = 1955+1/2 + (1:size(T))/365.25;
plot(time(1:end),T(1:end),'b+');
xlabel('Date');
ylabel('Avg. Temp. (F)');
title('Average Daily Temperatures');
xlim([1955 2011]);
% You may want to stretch out the figure horizontally to see the
seasonal patterns
```

Figure 2

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);

figure(2);
time = (1:size(T));
j = 0;
plot(time(1:365),T(j+1:j+365),'b+');
hold on;
j = 19719;
j = 16435;
plot(time(1:365),T(j+1:j+365),'r+');
hold off;
xlabel('Time (days from Jan. 1)');
ylabel('Avg. Temp. (F)');
xlim([1 365]);
```

Figure 4

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);

figure(4);
TbyYear = reshape(T(1:55*365),365,55);
boxplot(TbyYear);
xlabel('Years After 1955');
ylabel('Temperature (F)');
```

```
%The figure may have overwritten x-axis labeling.
%If that's the case, you can just stretch the figure to make it wider.
```

Figure 5

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);
```

```
figure(5);
time = 1956 + (1:size(T))/365;
plot(time(1:end-365), T(366:end) - T(1:end-365) , 'b+');
hold on;
plot([1955 2011], [0 0], 'r');
hold off;
xlabel('Date');
ylabel('Temperature (F)');
xlim([1955 2011]);
```

Figure 6

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);
```

```
figure(6);
window = ones(365,1);
Tw = conv(T,window, 'valid');
time = 1955+1/2 + (1:size(Tw))/365.25;
yearlytime = time(1:365:end)';
yearlytemp = Tw(1:365:end)/365;
yearlytimetemp = [yearlytime'; yearlytemp]';
plot(yearlytime,yearlytemp, 'b');
hold on;
plot(yearlytime,yearlytemp, 'r+');
hold off;
xlabel('Year');
ylabel('Temperature (F)');
xlim([1955 2010]);
```

Figure 7

```
% Note that this figure uses data from another source.
% Slopes and intercepts are specific to McGuireAFB.
% Cannot be used with other data sets.
```

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);
```

```
window = ones(365,1);
Tw = conv(T,window, 'valid');
time = 1955+1/2 + (1:size(Tw))/365.25;
yearlytime = time(1:365:end)';
yearlytemp = Tw(1:365:end)/365;
yearlytimetemp = [yearlytime'; yearlytemp]';
save -ascii 'yearlyTemps.dat' yearlytimetemp;
save -ascii 'yearlyDates.dat' yearlytime;
```

```
figure(7);
plot(yearlytime,yearlytemp, 'b');
hold on;
plot(yearlytime,yearlytemp, 'r+', ...
      [1955 2010], [52.7412 52.7412+0.0322755*(2010-1955)], 'r',...
      [1955 2010], [52.7908 52.7908+0.0368361*(2010-1955)], 'r--');
```

```
hold off;
xlabel('Year');
ylabel('Temperature (F)');
title('Regressions: Least Abs. Dev. (solid) and Least Squares
(dashed)');
xlim([1955 2010]);
```

Figure 8

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);

figure(8);
window = ones(365,1);
Tw = conv(T>window,'valid');
time = 1955+1/2 + (1:size(Tw))/365.25;
plot(time(1:end),Tw(1:end)/365);

xlabel('Date');
ylabel('Temperature (F)');
xlim([1955 2010.8]);
```

Figure 9

```
load -ascii data/McGuireAFB.dat;
T = McGuireAFB(:,2);

figure(9);
window = ones(365,1);
Tw = conv(T>window,'valid');
time = 1955+1/2 + (1:size(Tw))/365.25;
plot(time(1:end),Tw(1:end)/365,'b');

hold on;
window = ones(365,1);
Tw = conv(T>window,'valid');
time = 1955+1/2 + (1:size(Tw))/365.25;
plot(time(1:365:end),Tw(1:365:end)/365,'r+');
hold off;
xlabel('Date');
ylabel('Temperature (F)');
xlim([1955 2010]);
```