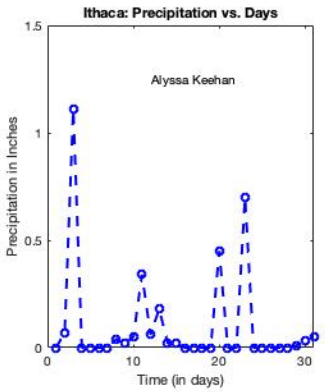
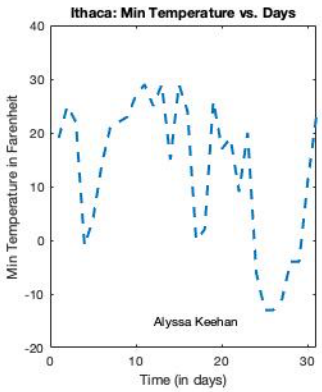
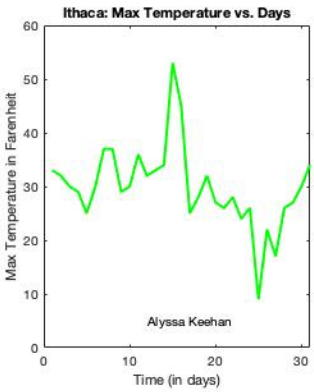
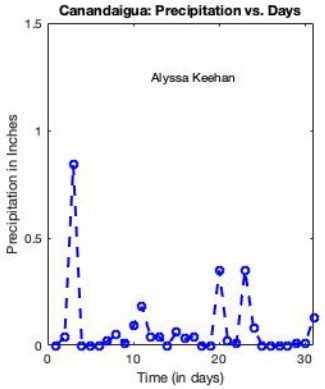
Question 1 and 2 plots

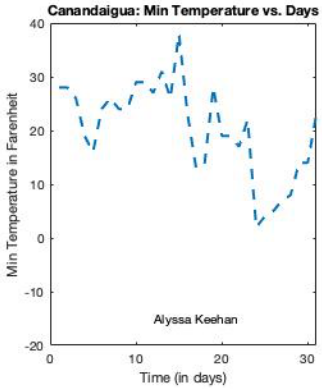
```
% first Im going to initialize each array from both of the datasets
idays = table2array(ithaca(:,1));
iprecip = table2array(ithaca(:,2));
itmin = table2array(ithaca(:,3));
itmax = table2array(ithaca(:,4));
cdays = table2array(canandaigua(:,1));
cprecip = table2array(canandaigua(:,2));
ctmin = table2array(canandaigua(:,3));
ctmax = table2array(canandaigua(:,4));
% plotting
subplot(2,3,1);
plot(idays,iprecip,'b--o', LineWidth = 2);
title('Ithaca: Precipitation vs. Days');
xlabel('Time (in days)');
ylabel('Precipitation in Inches');
ylim([0 1.5]);
text(12,1.25,'Alyssa Keehan');
subplot(2,3,2);
plot(idays,itmin,'--', LineWidth = 2);
title('Ithaca: Min Temperature vs. Days');
xlabel('Time (in days)');
ylabel('Min Temperature in Farenheit');
ylim([-20 40]);
text(12,-15,'Alyssa Keehan');
subplot(2,3,3);
plot(idays,itmax,'g', LineWidth = 2);
title('Ithaca: Max Temperature vs. Days');
xlabel('Time (in days)');
ylabel('Max Temperature in Farenheit');
ylim([0 60]);
text(12,5,'Alyssa Keehan');
subplot(2,3,4);
plot(cdays,cprecip,'b--o', LineWidth = 2);
title('Canandaigua: Precipitation vs. Days');
xlabel('Time (in days)');
ylabel('Precipitation in Inches');
ylim([0 1.5]);
text(12,1.25,'Alyssa Keehan');
subplot(2,3,5);
plot(cdays,ctmin,'--', LineWidth = 2);
title('Canandaigua: Min Temperature vs. Days');
xlabel('Time (in days)');
ylabel('Min Temperature in Farenheit');
ylim([-20 40]);
text(12,-15,'Alyssa Keehan');
subplot(2,3,6);
plot(cdays,ctmax,'g', LineWidth = 2);
title('Canandaigua: Max Temperature vs. Days');
xlabel('Time (in days)');
ylabel('Max Temperature in Farenheit');
ylim([0 60]);
text(12,5,'Alyssa Keehan');
```

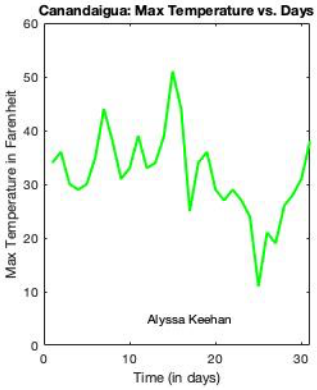

Ithaca: Precipitation vs. Days
Alyssa Keehan
Precipitation in Inches
Time (in days)


Ithaca: Min Temperature vs. Days
Alyssa Keehan
Min Temperature in Farenheit
Time (in days)


Ithaca: Max Temperature vs. Days
Alyssa Keehan
Max Temperature in Farenheit
Time (in days)

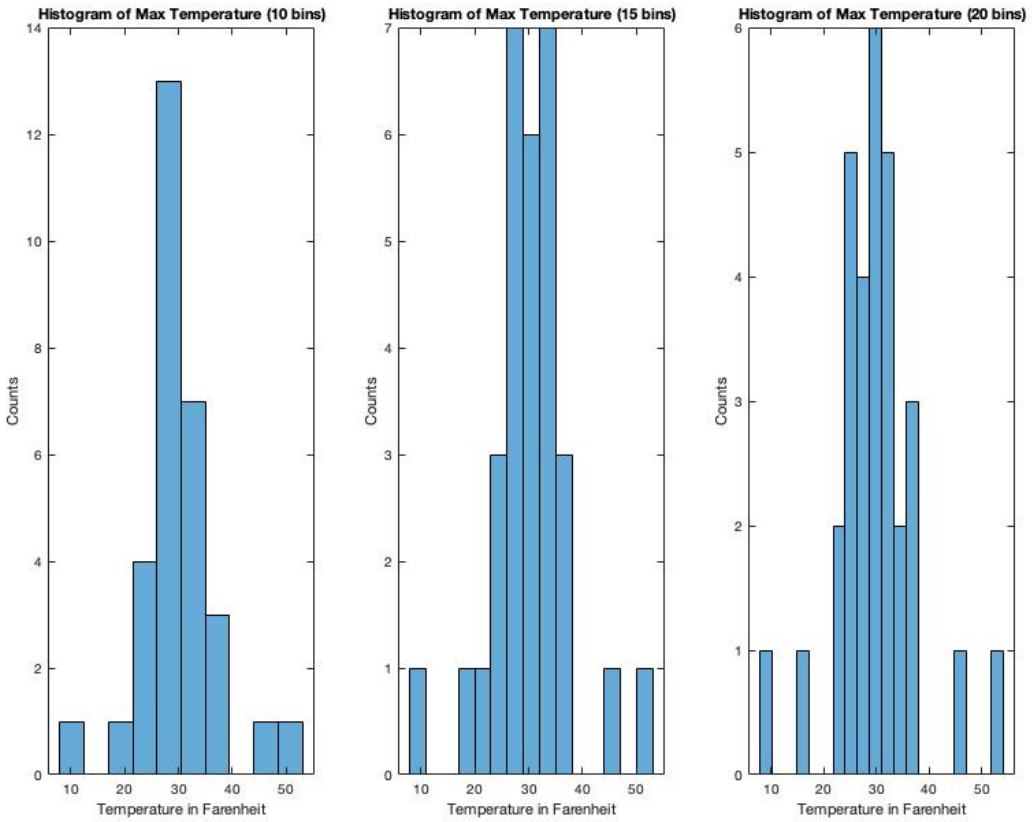

Canandaigua: Precipitation vs. Days
Alyssa Keehan
Precipitation in Inches
Time (in days)


Canandaigua: Min Temperature vs. Days
Alyssa Keehan
Min Temperature in Farenheit
Time (in days)


Canandaigua: Max Temperature vs. Days
Alyssa Keehan
Max Temperature in Farenheit
Time (in days)

Question 3 Plot 1

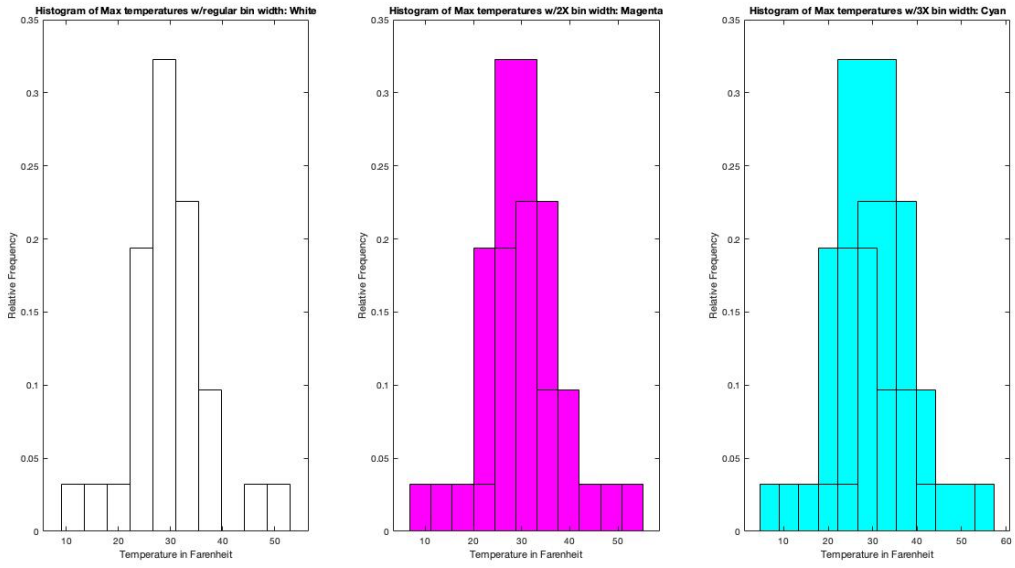
```
nDays = 31;
subplot(1,3,1);
histogram(itmax,10);
title('Histogram of Max Temperature (10 bins)');
xlabel('Temperature in Farenheit');
ylabel('Counts');
subplot(1,3,2);
histogram(itmax,15);
title('Histogram of Max Temperature (15 bins)');
xlabel('Temperature in Farenheit');
ylabel('Counts');
subplot(1,3,3);
histogram(itmax,20);
title('Histogram of Max Temperature (20 bins)');
xlabel('Temperature in Farenheit');
ylabel('Counts');
```



All three of the histograms have the same general shape of being normally distributed. As we increase the binwidth, the height of the histogram's bars decrease.

Question 3 plot 2

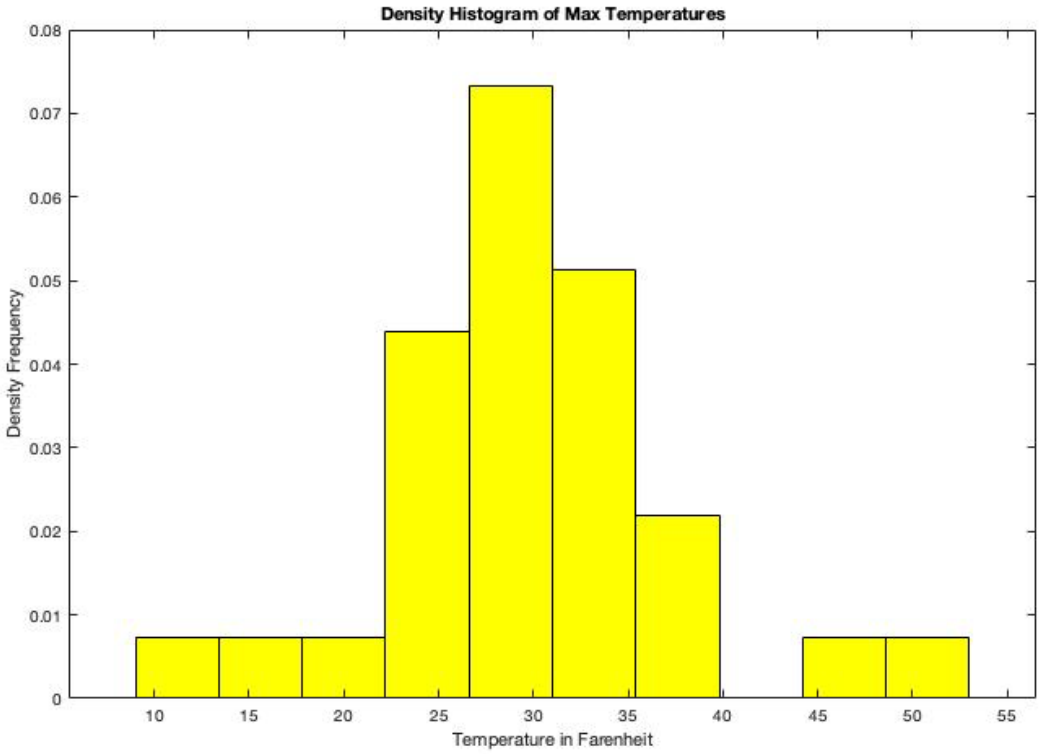
```
% output the # of data in each bin and the bin centers
[nDataInBin,binCentr] = histogram(itmax,10);
nDataInBin; % 1,1,1,6,10,7,3,0,1,1
binCentr; %11.2,15.6,20.0,24.4,28.8,33.2,37.6,42.0,46.4,50.8
% convert the counts to relative frequencies
relFreq = nDataInBin/nDays;
subplot(1,3,1);
bar(binCentr,relFreq,1,'w');
title('Histogram of Max temperatures w/regular bin width: White');
xlabel('Temperature in Farenheit');
ylabel('Relative Frequency');
subplot(1,3,2);
bar(binCentr,relFreq,2,'m');
title('Histogram of Max temperatures w/2X bin width: Magenta');
xlabel('Temperature in Farenheit');
ylabel('Relative Frequency');
subplot(1,3,3);
bar(binCentr,relFreq,3,'c');
title('Histogram of Max temperatures w/3X bin width: Cyan');
xlabel('Temperature in Farenheit');
ylabel('Relative Frequency');
```



binCentr represents the horizontal axis which is used to build each bar in the bar chart. relFreq represents the vertical axis and determines the height of each bar. The third argument represents the width of the bars. If you look at the width when that parameter is set to 2 or three, the bin widths are much larger, and you can see how they are essentially twice the original size. The last argument changes the color of the bins. This argument usually just takes the first letter of the unique color as the argument. Another argument you can input into the bar function are style, which determines how you want to output the bar chart.

Question 4

```
binWidth = binCentr(2)-binCentr(1);
densFreq = relFreq/binWidth;
bar(binCentr,densFreq,1,'y');
title('Density Histogram of Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Density Frequency');
```

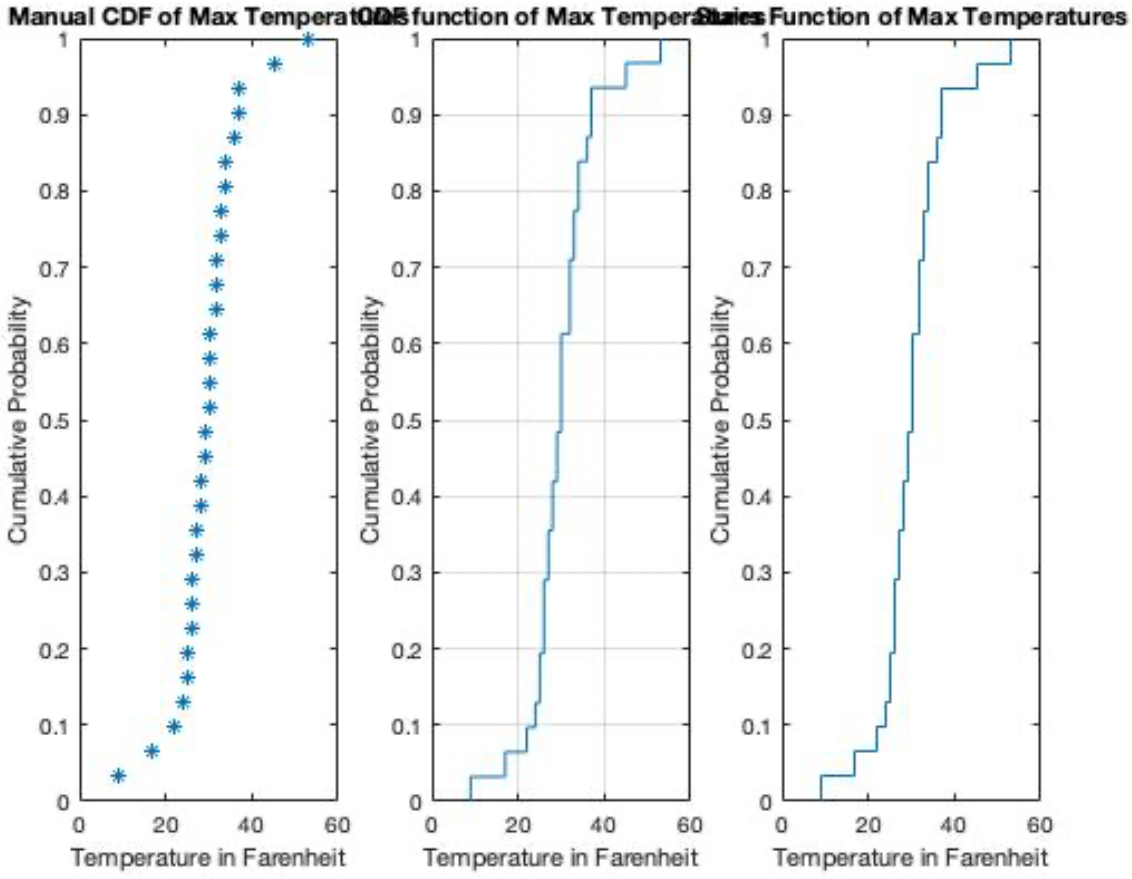


To verify that the total area of the density histogram integrates to 1, I will multiply the bin widths with the density frequencies and add them together.

```
sum(densFreq*binWidth); % the result is in fact 1
```

Question 5

```
tmaxS = sort(itmax);
p=(1:nDays)/nDays;
subplot(1,3,1);
plot(tmaxS,p,'*');
title('Manual CDF of Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Cumulative Probability');
subplot(1,3,2);
cdfplot(itmax);
title('CDF function of Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Cumulative Probability');
[Fx,xS] = ecdf(itmax);
subplot(1,3,3);
stairs(xS,Fx);
title('Stairs Function of Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Cumulative Probability');
```



Both the auto matlab functions are lines and looks more like staircases while the plot function is more rounded off.

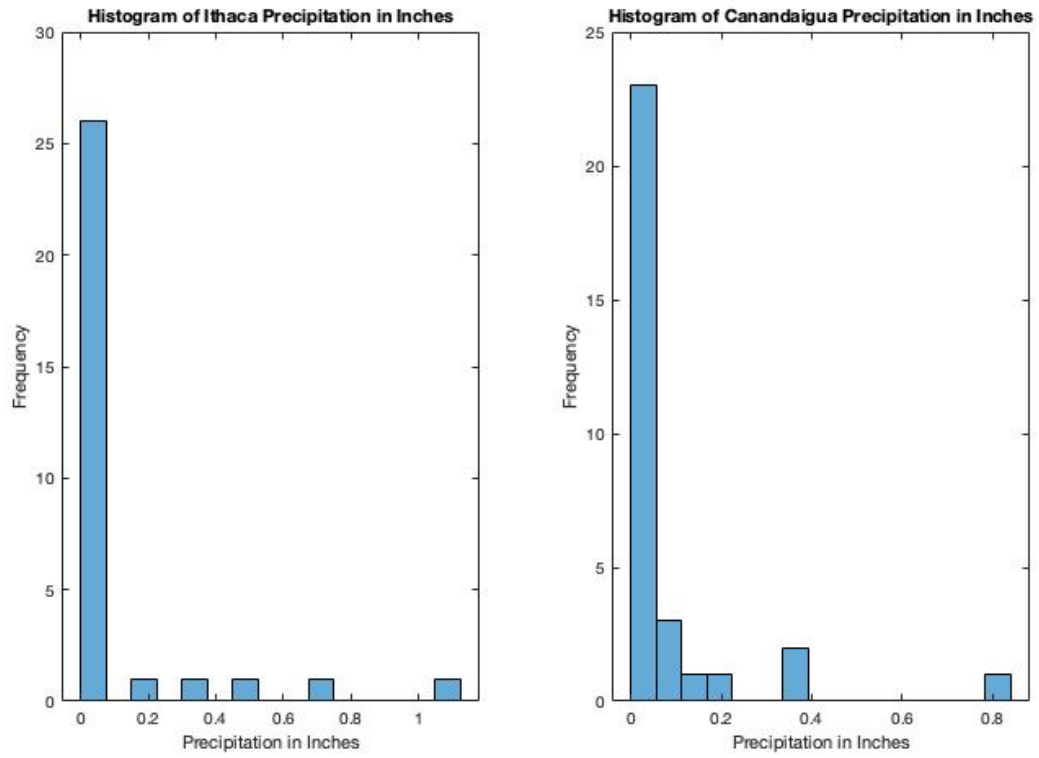
Question 6

```
xS(1) = xS(1)-0.001;
y20 = interp1(xS,Fx,20); % 0.0839
y40 = interp1(xS,Fx,40); % 0.9476
```

The cumulative probability value at a max temperature of 20 degrees is 0.0839 while the cumulative probability value at a max temperature of 40 degree is 0.9476.

Question 7

```
subplot(1,2,1);
histogram(iprecip,15);
title('Histogram of Ithaca Precipitation in Inches');
xlabel('Precipitation in Inches');
ylabel('Frequency');
subplot(1,2,2);
histogram(cprecip,15);
title('Histogram of Canandaigua Precipitation in Inches');
xlabel('Precipitation in Inches');
ylabel('Frequency');
```



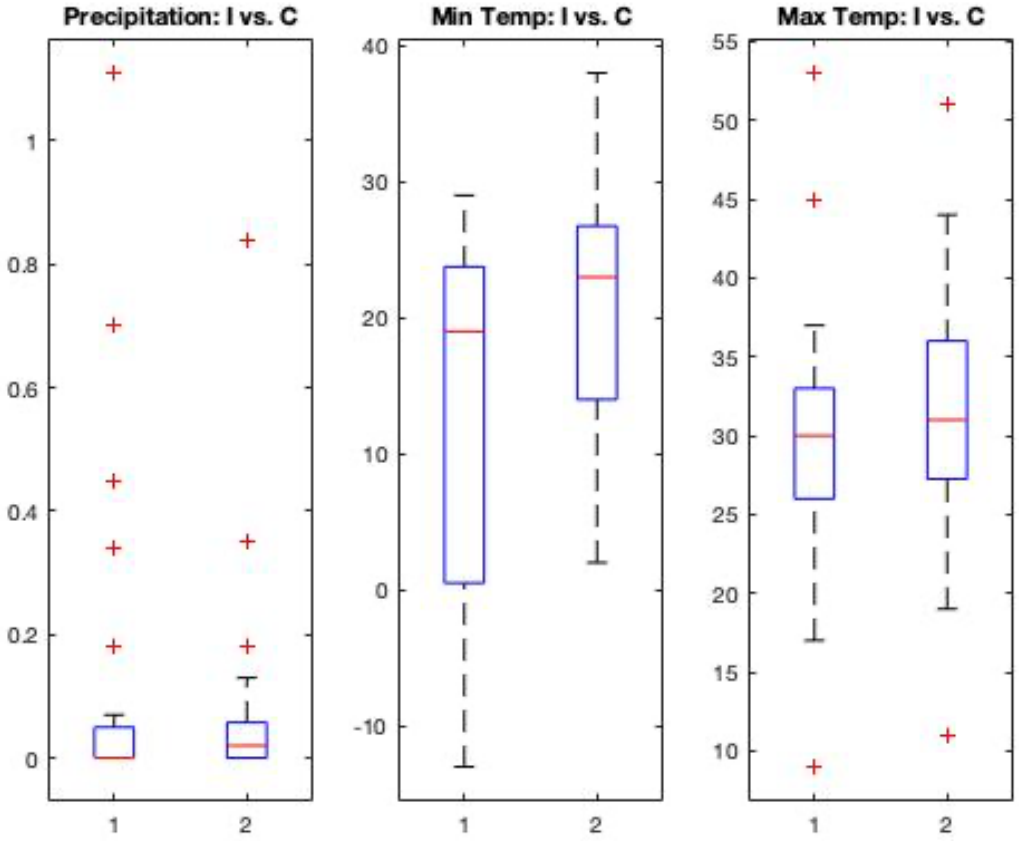
While both histograms are positively skewed, the Ithaca Precipitation measurements have a higher mode at the lowest bin and have a larger range overall. Meanwhile the Canandaigua measurements gradually level off more than Ithaca's sudden drop at the second bin.

```
mean(iprecip); % The mean precipitation for Ithaca is 0.1016 inches.
std(iprecip); % The standard deviation for Ithaca precipitation is 0.2429 inches.
mean(cprecip); % The mean precipitation for Canandaigua is 0.0774 inches.
std(cprecip); % The standard deviation for Canandaigua is 0.1676 inches.
```

It's an interesting observation to see that the mean and standard deviation for the precipitation in Ithaca is higher than for Canandaigua. I also think it is interesting that the standard deviation for both datasets are higher than the means. This is something good to take note of because it means that less than 1 standard deviation lower is gonna be negative inches which should never happen.

Question 8

```
subplot(1,3,1);
boxplot([iprecip cprecip]);
title('Precipitation: I vs. C');
subplot(1,3,2);
boxplot([itmin ctmin]);
title('Min Temp: I vs. C');
subplot(1,3,3);
boxplot([itmax ctmax]);
title('Max Temp: I vs. C');
```



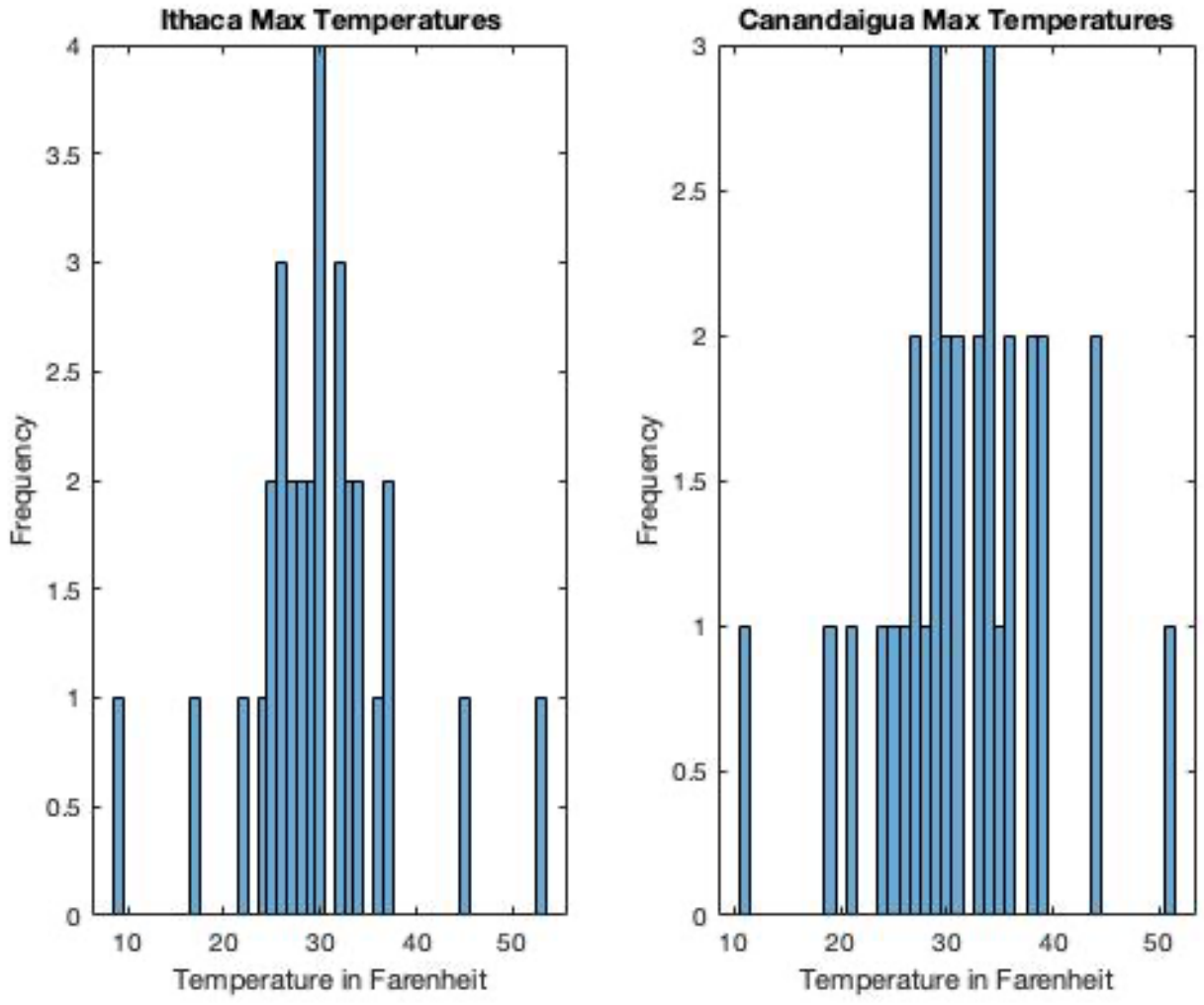
Boxplots are produced in such a way that the actual box is representative of the middle 50% of the data (25th percentile to the 75th percentile) and the line inside is the median. The bars on the end represent the range of the data that is within probable range and the plus signs are considered outliers.

For the precipitation variables, I see that the median and bar for Ithaca is smaller than the median for Canandaigua, but the overall range including the outliers extend farther for Ithaca's values than Canandaigua's. For the minimum temperature, the range and median for Ithaca spans and is lower than Canandaigua's but the interquartile range for Ithaca is larger than

Canandaigua's. Finally, for the max temperatures, the medians are similar, but the Ithaca range has lower
values and is slightly shorter than Canandaigua's. Also, Ithaca has three outliers (one on the lower end and two on the upper end) for max temperature while Canandaigua's max temperature have just one outlier on both sides.

Question 9

```
kurtosis_ithaca = sum((itmax - mean(itmax)).^4)/((std(itmax)^4)*nDays); % 5.1923
kurtosis_canandaigua = sum((ctmax - mean(ctmax)).^4)/((std(ctmax)^4)*nDays); % 3.5863
median(itmax) % 30
mean(itmax) % 29.8710
median(ctmax) % 31
mean(ctmax) % 31.7742
subplot(1,2,1);
histogram(itmax);
title('Ithaca Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Frequency');
subplot(1,2,2);
histogram(ctmax);
title('Canandaigua Max Temperatures');
xlabel('Temperature in Farenheit');
ylabel('Frequency');
```



Based on the histograms alone, I can tell that they look pretty Gaussian/Normally distributed. When I calculated the skewness for the two stations, I noticed that the median for Ithaca's max temperature was slightly higher than the mean but the median for the Canandaigua max temperatures were very slightly lower than its mean. This means that although practically unnoticeable, the Ithaca max temperature distribution is slightly negatively skewed while the Canandaigua max temperature is slightly positively skewed. As for the kurtosis values, both have values above 3 which indicate high peakedness, but the kurtosis for Canandaigua seems to be flatter than that of Ithaca's max temperature distribution.