# Box and Whisker Plots for Local Climate Datasets:  Interpretation and Creation using Excel 2007/2010


PETER C. BANACOS
*NOAA/NWS Burlington, Vermont*


## ABSTRACT


*This paper describes the creation and use of box and whisker plots in the statistical analysis of local climate and other hydrometeorological datasets. Box and whisker plots offer a pictorial summary of important dataset characteristics including the central tendency, dispersion, asymmetry, and extremes, arrived at through percentile rank analysis and the plotting of maximum and minimum dataset values.  Since box and whisker plots display measures of central tendency and spread free from the assumption of a normal distribution, they provide an effective way of identifying asymmetrical attributes in meteorological datasets. Additionally, the underlying statistics are more resistant toward individual outliers than other methods, such as mean and standard deviation. Common measures of variability, such as standard deviation, may be interpreted based upon an assumption of an underlying standard normal distribution for climate and weather analysis purposes, and might also prove too abstract for non-technical users of climate data. Lastly, the graphically compact nature of box and whisker plots facilitates side-by-side comparison of multiple datasets, which can otherwise be difficult to interpret using more complete representations, such as the histogram.  A box and whisker plotting convention geared for meteorological applications is described herein, with examples shown using climate data from the WFO Burlington, Vermont forecast area. The appendix includes instructions for creating the box and whisker plot format advocated in this paper, and a sample template is also available for download.*

—————————————————————

—————————

Corresponding author address: Peter C. Banacos, NOAA/NWS/Weather Forecast Office, 1200 Airport Drive, South Burlington, VT  05403. E-Mail: peter.banacos@noaa.gov

## 1.     Introduction

Conveying the normal variability of weather conditions or specific weather events can be of critical importance as a decision and planning tool for engineers, agriculturists, recreational enthusiasts, and others with weather sensitive interests. However, the large variability of weather in mid-latitudes is not necessarily easy to summarize in statistical terms. Traditional 30-year climate means and extremes may not be effective in demonstrating natural fluctuations in weather about the mean that are typical on daily, seasonal, or annual time scales. Common measures of variability, such as standard deviation, are based upon the assumption of a standard normal distribution, and might also prove too abstract for non-technical users of climate data. What is often needed is a simple graphical summary that portrays the statistical dispersion in a manner that is easy to interpret for a wide range of users.

Focusing climate statistics in terms of the variability of conditions rather than the central tendency also helps place observed or anticipated weather events into a historical context. This provides operational forecasters with a reference point to identify the occurrence of unusual weather conditions, the value of which has been established in other studies (e.g., Grumm and Hart 2001); specifically, it contributes to improved situational awareness. Putting into perspective weather events as they occur is also of strong interest to the media and the general public. In the graphical era of the Internet, the ability to quantify and view a current weather event (e.g., heat wave, snow amount, etc.) against a range of past events of the same type is desirable. Interpretation of such information can form the basis of

further discussion, as might routinely take place between the National Weather Service (NWS) and external users during or following significant weather events.

One way of graphically focusing on statistical variability is by way of the box and whisker plot (Tukey 1977), first proposed by statistician John Tukey in 1970. Plotting conventions have varied since, based on application and user preferences. The goal of this paper is to advocate a form of the box and whisker plot for climate and other hydrometeorological datasets. Box and whisker plots describe data in a manner that is (1) pictorially compact and makes easy comparison with like datasets, (2) retains the ability to interpret asymmetric aspects of the data and data extremes, and (3) is useful to both operational forecasters and external users of climate related datasets. The remainder of the paper is organized as follows. The basic structure of the box and whisker plot is explained in Section 2. In Section 3, interpretation of box and whisker plots is discussed. Some example applications of box and whisker plots are shown and described in Section 4, followed by conclusions in Section 5. Lastly, an appendix is included to show the steps necessary to create box and whisker plots in Excel 2007/2010, which is available at most NWS offices.

## 2.     The box and whisker plot

The form of the box and whisker plot advocated here is a graphical 7-number summary of a given dataset, which includes: the median, the interquartile range (shown by the box), the outer range (shown by the whiskers), and the climatological extremes (Fig. 1). The definition and computation of each of these values is described below.

## a. The median

The median is the middle observation in a ranked dataset (or mean of the two middle observations for an even numbered dataset) and is a measure of the central tendency of the data. The median is equivalent to the $50^{th}$ percentile in a percentile rank analysis with the same number of observations below as above the median. An advantage of the median is its resistance against outlying values for $n \geq 3$, where $n$ is the number of observations. Whereas the mean can be skewed by an extreme outlying observation, especially for relatively small datasets, the median is unaffected and therefore remains robust. In Figure 1, the median is displayed as a solid bar within the box and the median value would be plotted alongside.

## b. The interquartile range

The box represents the middle 50% of the ranked data and is drawn from the lower quartile value to the upper quartile value (i.e., the $25^{th}$ to $75^{th}$ percentile). The lower (upper) quartile is computed by taking the median of the lower (upper) half of the ranked data. The difference between the upper and lower quartile values is referred to as the interquartile range (IQR), and the height of the box is proportional to the statistical disparity or spread of the inner 50% of the ranked data. The box portion of the plot visually stands out, which is a desirable aspect drawing the users attention to the central half of the data (Frigge et al., 1989). The box is standardized as representing the IQR in published applications of box and whisker plots (Schultz 2009). For large, reliable samples, there is a 50% chance that future observations will be within the box portion of the graph (i.e., relative frequency can be interpreted as a probability of occurrence).

The quartile values can be plotted adjacent to the top and bottom of the box to quantify these data for the reader.

## c. The outer range

The whiskers represent an outer range and are drawn as vertical lines extending outward from the ends of the box. Unlike the box, plotting conventions for the whiskers vary (Schultz 2009). For example, Massart et al. (2005) draw the end of the top whisker to the upper quartile + (1.5 x IQR), and the end of the bottom whisker to the lower quartile – (1.5 x IQR). While this choice is arbitrary, the goal of this methodology is to flag outliers as those observations which lie beyond the whiskers; in some applications these individual outliers are plotted as dots. Other conventions include extending the whiskers to the minimum and maximum values of the whole dataset (McGill et al. 1978), or using the $10^{th}$ and $90^{th}$ percentiles to define the ends of the whiskers (Cleveland 1985).

The author adopts the $10^{th}$ and $90^{th}$ percentile for the ends of the whiskers, with data values plotted alongside. In interpreting large and reliable datasets with this convention, there is a 10% probability of future occurrence beyond the values at the ends of the whiskers; an example of this is shown in the freeze climatology in Section 4. Meteorological applications of box and whisker plots have effectively employed the $10^{th}$ and $90^{th}$ percentile for the whisker ends (e.g., Brooks 2004, Thompson et al. 2007). However, whiskers extending to the dataset maximum and minimum values have also been used (e.g., Dupilka and Reuter 2006).

## d. Maximum and minimum values

Meteorologists, climatologists, and the general public are often interested in

climatological extremes, so it is useful to add the maximum and minimum values of the dataset to the traditional box and whisker plot. These values are denoted by the "x" in Figure 1. Plotting the data value and date of occurrence of the extremes can also serve as a handy reference.

## 3. Interpreting box and whisker plots

### a. *Data patterns for individual box and whisker plots*

There are several common patterns associated with box and whisker plots for meteorological applications, as idealized in Figure 2. The length of the IQR (as shown by the box) is a measure of the relative dispersion of the middle 50% of a dataset, just as the length of each whisker is a measure of the relative dispersion of the dataset outer range ($10^{th}$ to $25^{th}$ percentile and $75^{th}$ to $90^{th}$ percentile). This dispersion can be comparatively small (Fig. 2a) or large (Fig. 2b). Likewise, the maximum and minimum values may lie close to the whisker ends (Fig. 2a) or far away (Fig. 2b), which is a measure of how markedly different the extremes are from the remainder of the sample.

When the length of the IQR is small compared to the whiskers, this suggests a middle clustering of data about the median with long tails representing a large dispersion of the relative outliers (Fig. 2c). On the other hand, a large IQR compared to the whiskers can be indicative of a clustering of observations near the $25^{th}$ and $75^{th}$ percentile, or a bimodal distribution (Fig. 2d). In order to confirm a bimodal distribution it is useful to investigate the distribution more thoroughly, such as via a histogram.

A key advantage of the box and whisker plot is the ability to visualize dataset skewness. In Figures 2a-d, there is

zero skewness in the idealized data; the data are perfectly symmetric about the median. An example of upward or positive skewness is shown in Figure 2e; in this case the median is shifted toward the lower portion of the box with a wider range of observations in the upper quartile as compared to the lower quartile. The opposite is true in Figure 2f, which is an example of downward or negative skewness. The whiskers in Figure 2e and Figure 2f also exhibit the same skewness character as the IQR.

Knowledge of skewness tells the user whether deviations from the median are more likely to be positive or negative. Assuming a representative sample, the distribution shown in Figure 2e and Figure 2f would suggest meteorological data limits are approached closer to the median in the negative (positive) direction, and that far outliers are less probable in that direction. Understanding data asymmetries can be useful, and are otherwise lost in classical statistics based on the normal distribution (Massart et al. 2005).

### b. *Comparing box and whisker plots and quantifying data differences*

Another advantage of the box and whisker plot is the ability to compare multiple datasets side-by-side, as idealized in Figure 3. Important characteristics of each dataset (central tendency, skewness, dispersion, and extremes) are easy to interpret and visualize. Qualitatively, the relative overlap between each box and whisker plot indicates the degree to which each dataset is similar in its dispersion, with an emphasis on the IQR owing to the graphing methodology (e.g., the box stands out relative to the rest of the data).

While qualitative features stand out, care is needed in quantifying dataset differences particularly for small *n*. In

Figure 3, there is visually the least overlap between datasets A-C, followed by datasets B-C, and finally A-B. Whether or not these differences are statistically significant is partly a function of the sample size, and ultimately requires significance testing.

In Figure 3, sample size is included at the bottom of the graph beneath each plot; this can be valuable especially if *n* varies between datasets, as this information might not otherwise be known to the reader. Instructions for including the sample size are included in the appendix, and significance test methods are built into most spreadsheet programs.

## 4. Box and whisker application examples

In this section, we show two applications of the box and whisker plot methodology applied to meteorological and hydrological data in the WFO Burlington, VT forecast area.

### a. *Dates of first and last freeze*

The date of first and last freeze (taken as the first and last seasonal occurrences of a minimum temperature at or below $32^{o}F$) has important agricultural implications, and varies from year-to-year and location-to-location. Figure 4 shows box and whisker plots of the last freeze dates leading out of the cold season (Fig. 4a) and the first freeze dates leading out of the warm season (Fig. 4b) for select locations in the WFO Burlington, Vermont forecast area. These plots reveal the dispersion in these dates both at individual sites and between locations over a period of 30 years (sites nearer to Lake Champlain tend to have longer growing seasons). These graphs immediately convey the variability to the reader in a way the mean date cannot, and

more practically than the standard deviation would.

The IQR is relatively large for most sites, spanning roughly two-weeks for the middle 50% of freeze dates both for the first and last freezes. It can be interpreted that half of all years fall within the IQR. Likewise, it could be interpreted that freeze dates beyond the ends of the whiskers are a once-per-decade event. Note also that the 30-year extremes and dates of occurrence are shown for each location. There are asymmetries in these data as well. While a complete meteorological discussion is beyond the scope of this paper, note that South Hero, Vermont shows a downward skew toward later dates. Note that the range between the $75^{th}$ percentile (top of the box) and the median is only 5 days, whereas the median to the $25^{th}$ percentile covers 8 days. Likewise, the top whisker end lies 8 days from the median versus 11 days for the bottom whisker end.

### b. *Lake Champlain lake level*

Box and whisker plots can be applied to hydrological applications as well, such as the daily Lake Champlain lake level binned by month (Fig. 5). These data show the relatively high lake level associated with the "spring melt" in April and May, and large dispersion presumably associated with the variable timing of the seasonal melt, amount of snowpack in the basin, and precipitation. Thereafter, a logarithmic decrease in lake level occurs throughout the summer to a minimum in early fall, with a smaller IQR noted showing smaller variation in the lake level. A secondary maximum in lake level is evident in late fall into early winter associated with more frequent synoptic-scale precipitation and lesser evaporation rates associated with cooler air temperatures. The Lake Champlain flood stage of 100 feet has

been added to the plot to show the data as compared to important operational thresholds; flood stage has only been exceeded in March, April, May, June, and December. The highest level of 101.88 feet occurred on a day in April 1993 and the lowest level of 93.02 feet occurred on a day in November 1953; extreme data is included for Lake Champlain back through 1950.

## 5. Conclusions

In this paper, we have described the definition, use, and advantages of box and whisker plots in the display and interpretation of climate and other hydrometeorological datasets. The box and whisker plot pictorially summarizes the central tendency, dispersion, skewness, and extremes of a dataset in an easy to read format, and one that makes comparisons between multiple datasets easy, as shown in the included examples. The box and whisker methodology does not rely on the statistical assumption of a normal distribution. Instructions for creating the box and whisker format used here is included in the appendix for Microsoft Excel 2007/2010.

There are many potential applications of box and whisker plots in the analysis of local climate and hydrologic data at NWS offices. Likewise, the creation and interpretation of such plots make great student projects[1]. Box and whisker plots can be created from local climate databases and made available on office websites for public viewing and decision support or planning purposes. It is the author's view that the box and whisker format shown here provides a

significant amount of information without unduly taxing the reader. Thus, their application should be agreeable for operational purposes within the NWS and for external users.

## References

Brooks, H. E., 2004: On the relationship of tornado path length and width to intensity. *Wea. Forecasting*, **19**, 310-319.

Cleveland, W. S., 1985**:** *The Elements of Graphing Data*, Hobart Press, 297 pp.

Dupilka, M. L., and G. W. Reuter, 2006: Forecasting tornadic thunderstorm potential in Alberta using environmental sounding data. Part I: wind shear and buoyancy. *Wea. Forecasting*, **21**, 325-335.

Frigge, M., D. C. Hoaglin, and B. Iglewicz, 1989: Some implementations of the boxplot. *The American Statistician*, **43**, 50-54.

Grumm, R. H., and R. Hart, 2001: Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Wea. Forecasting*, **16**, 736-754.

---

[1] See the following URL for one such example: http://www.erh.noaa.gov/btv/climo/snowclimo/snowfall.shtml

Massart, D. L., J. Smeyers-Verbeke, X. Capron, and K. Schlesier, 2005: Visual presentation of data by means of box plots. *LC-GC Europe*, **18**(4), 2-5.

McGill, R., J. W. Tukey, W. A. Larsen, 1978: Variations of box plots. *The American Statistician*, **32**, 12-16.

Schultz, D. M., 2009: *Eloquent Science: A Practical Guide to Becoming a Better Writer, Speaker, and Atmospheric Scientist.* Amer. Meteor. Soc., 440 pp.

Thompson, R. L., C. M. Mead, and R. Edwards, 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Wea. Forecasting,* **22**, 102-115.

Tukey, J. W., 1977: *Exploratory Data Analysis*. Addison-Wesley, 688 pp.

**Figures**
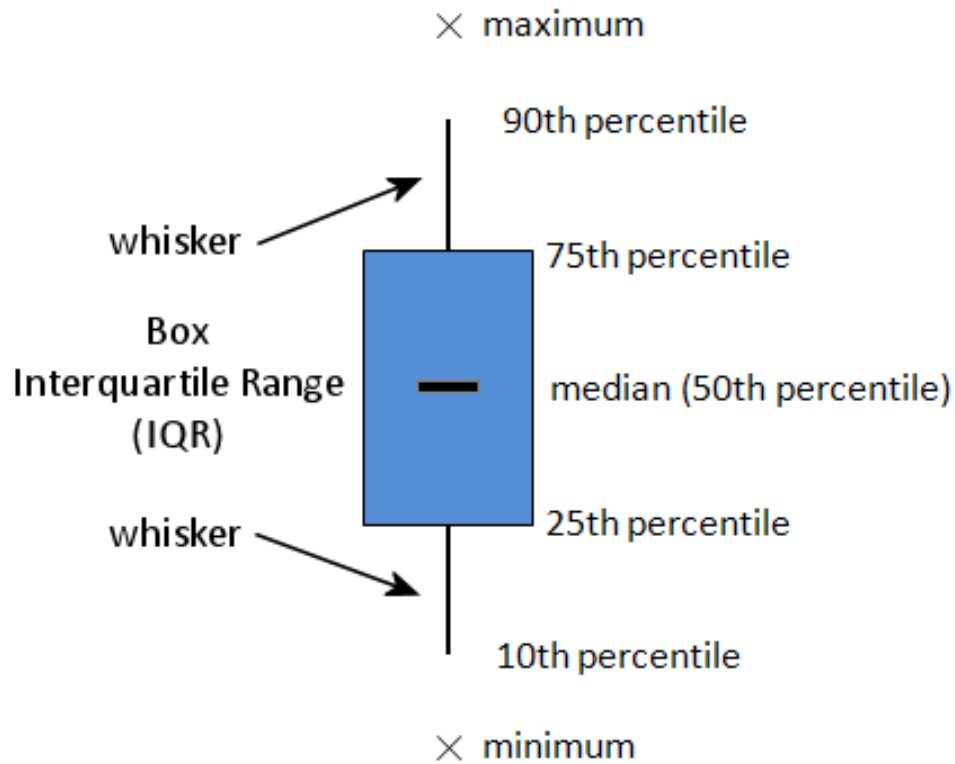


Figure 1. A generic box and whisker plot with extreme data points added. The height of the box portion is given by the interquartile range of the dataset, and extends from the $25^{th}$ to $75^{th}$ percentile. The horizontal bar within the box denotes the median value. The ends of the whiskers are drawn to the $10^{th}$ and $90^{th}$ percentile values. The extreme values are labeled with an "x" at the maximum and minimum points. The descriptors to the right of the box and whisker plot are replaced with data point values in the format shown in Figs. 4-6.
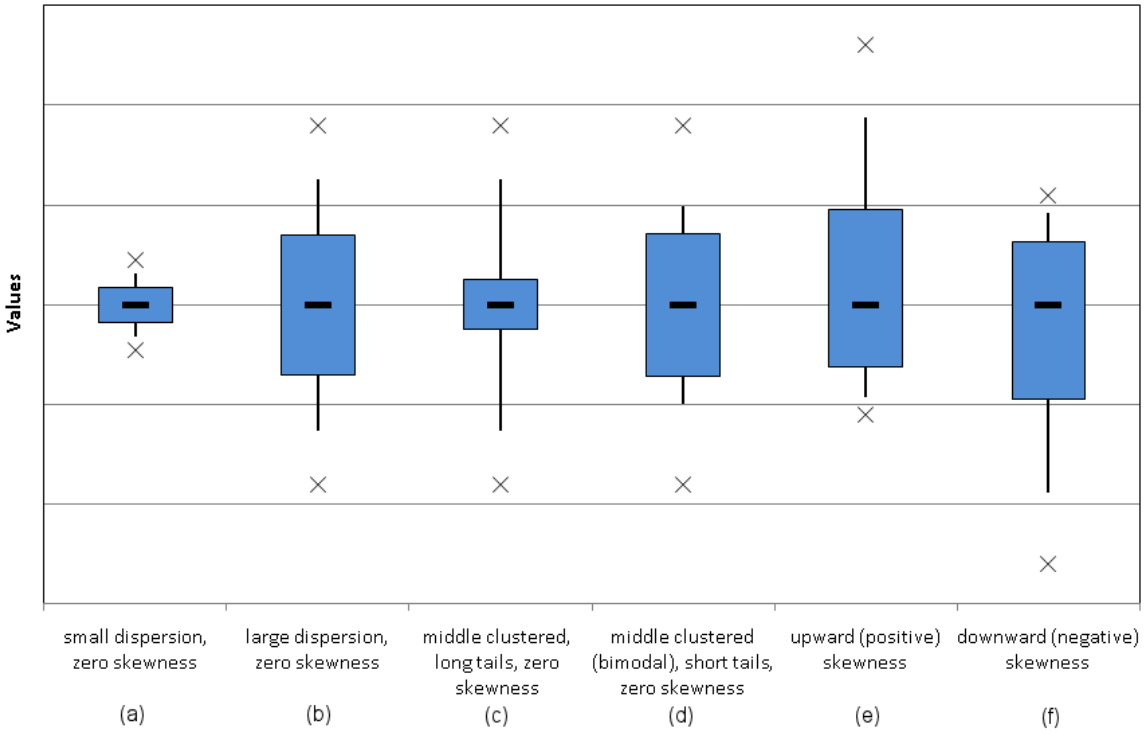
Figure 2. Idealized box and whisker plots for six data distributions. The datasets exhibit (a) small dispersion, (b) large dispersion, (c) middle clustering (with long tails), (d) middle clustering based on a bimodal distribution (with short tails), (e) upward (positive skewness), and (f) downward (negative) skewness. Datasets (a)-(d) are symmetric about the median and have zero skewness. The value scale along the ordinate is arbitrary, but linear.
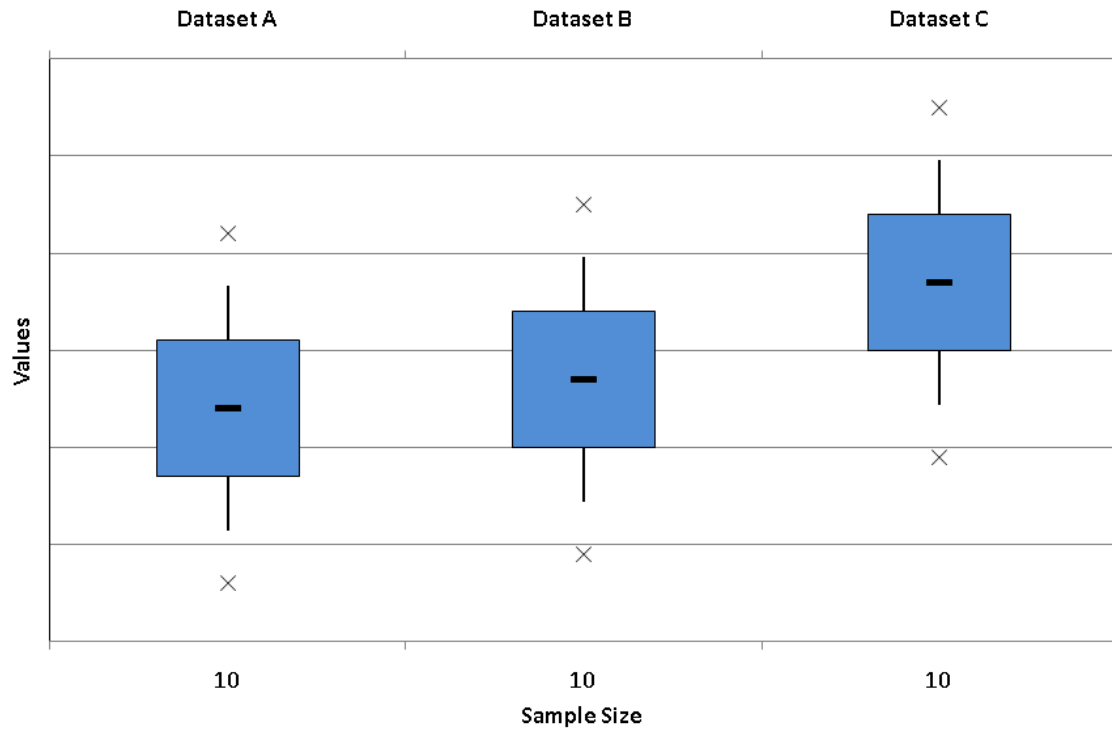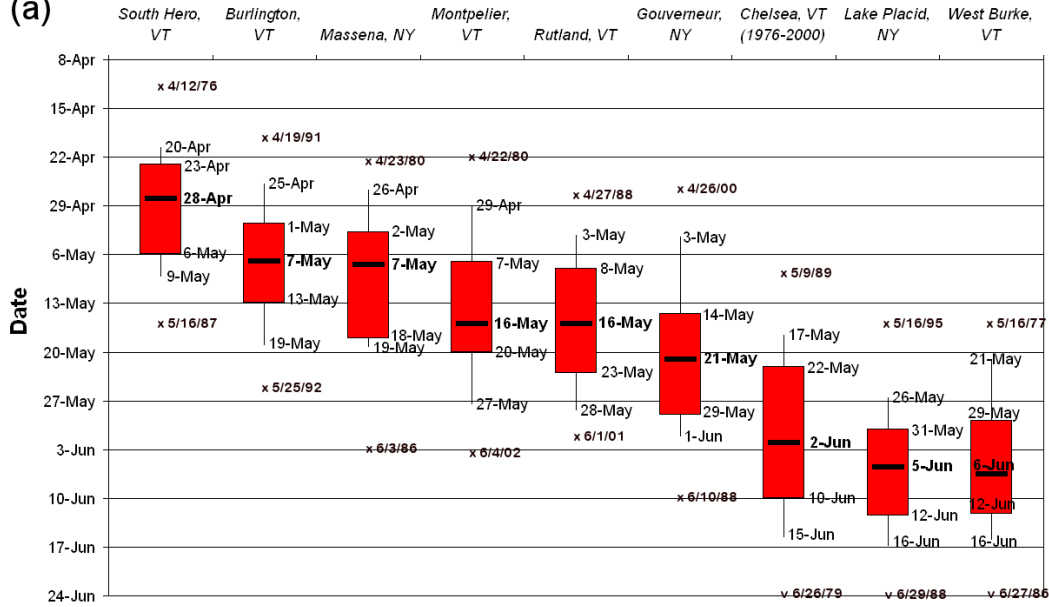
Figure 3. Box and whisker plots for three idealized data distributions (datasets "A", "B", and "C"). Number below each plot represents a dataset of sample size $n$. The value scale along the ordinate is arbitrary, but linear.
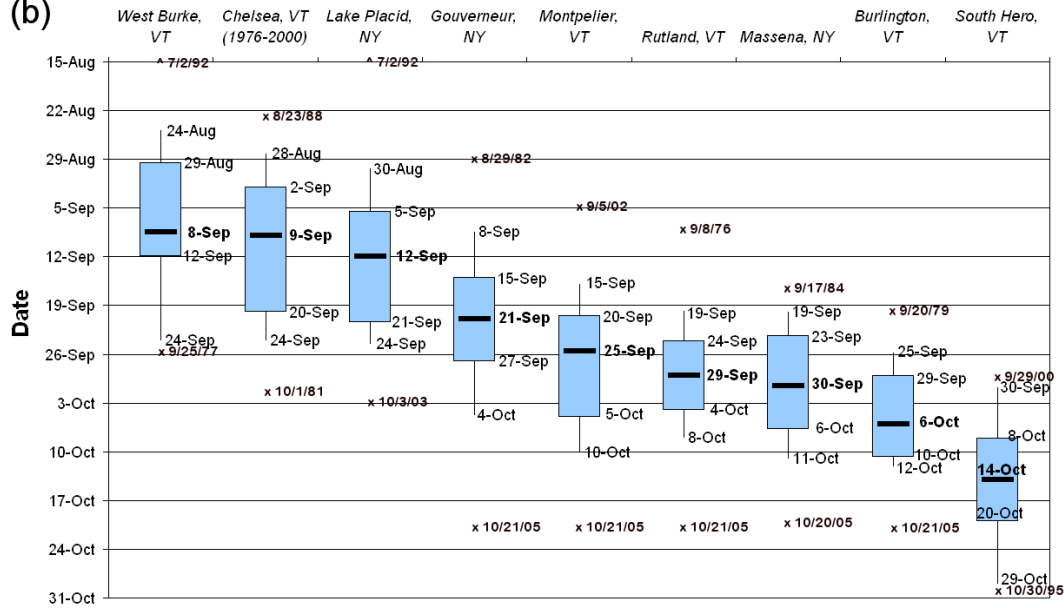
Figure 4. Box and whisker plots of (a) last spring freeze data and (b) first fall freeze data for locations in the WFO BTV forecast area, based on 1976-2005 data.
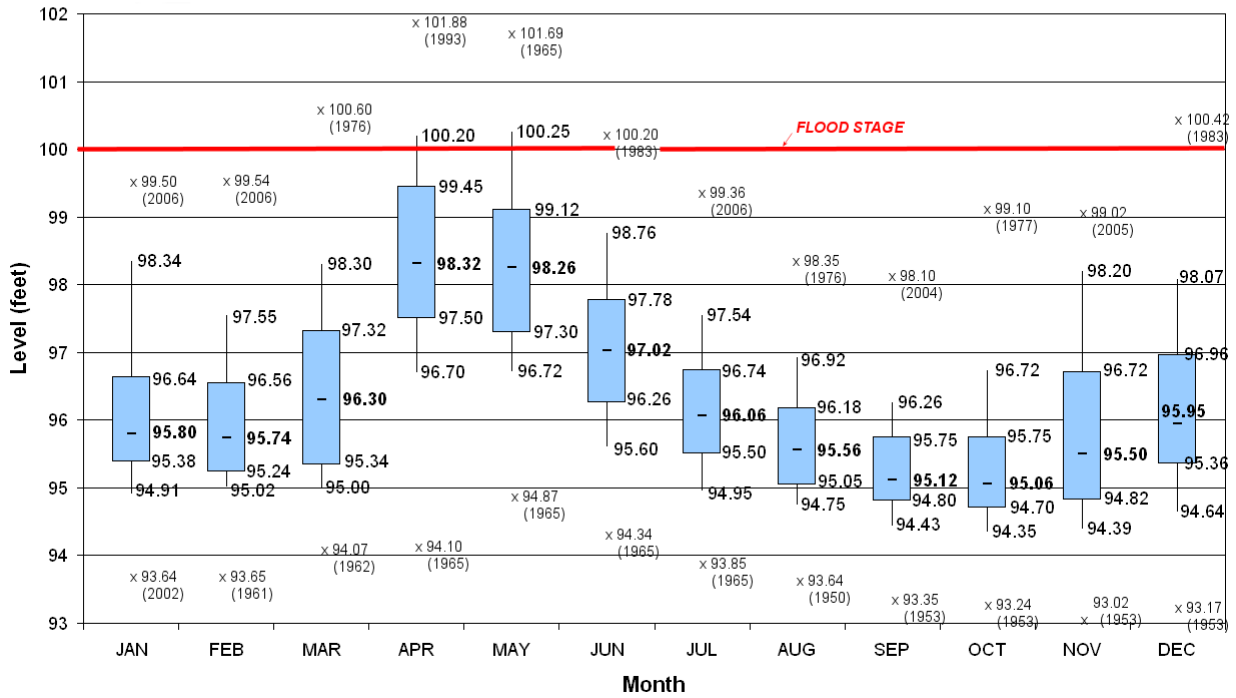
Figure 5. Box and whisker plot of monthly Lake Champlain Lake Level, based on daily readings at the King Street Ferry Dock in Burlington, Vermont.

# APPENDIX

## Creating Box and Whisker Plots in Microsoft Excel 2007 / 2010

The creation of box and whisker plots in Microsoft Excel 2007/2010 is described here since (1) it is the standard spreadsheet software at NWS offices as of this writing, and (2) multiple steps are required to produce these plots. One can download the finished template; visit ftp://ftp.werh.noaa.gov/share/BTV/BoxPlots/Sample.xlsx. Be sure to save a copy of this spreadsheet to your local computer/network. The template can be used to get started quickly by entering in your own data. Otherwise, follow the steps below to generate your own box and whisker plots from scratch (note: ideally, climate analysis includes at least 30-years of data. For simplicity, this example includes only 10 years of data for three months for ease of explanation).

**Step 1.** Type or import your climate data into Excel. If necessary, ASCII text files can be opened within Excel; the "import text wizard" will convert fixed width or delimitated text data into a spreadsheet format with data divided into individual cells.

**Step 2.** Compute percentile rank statistics in the exact order shown below.

|  | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 |  |  |  |  |  |
| 2 |  | 2000-09 JJA Rainfall Totals at Burlington, VT (BTV) |  |  |  |
| 3 |  |  | June | July | August |
| 4 |  | 2000 | 3.55 | 3.16 | 3.67 |
| 5 |  | 2001 | 2.32 | 0.77 | 4.32 |
| 6 |  | 2002 | 6.73 | 3.35 | 1.16 |
| 7 |  | 2003 | 2.98 | 3.48 | 2.24 |
| 8 |  | 2004 | 3.61 | 7.65 | 7.87 |
| 9 |  | 2005 | 3.78 | 5.22 | 4.17 |
| 10 |  | 2006 | 6.77 | 2.95 | 4.36 |
| 11 |  | 2007 | 2.26 | 6.29 | 1.46 |
| 12 |  | 2008 | 5.21 | 7.07 | 3.68 |
| 13 |  | 2009 | 5.25 | 4.62 | 2.32 |
| 14 |  |  |  |  |  |
| 15 |  | Calculations | June | July | August |
| 16 |  | 25th Percentile | 3.12 | 3.21 | 2.26 |
| 17 |  | 10th Percentile | 2.31 | 2.73 | 1.43 |
| 18 |  | Median | 3.70 | 4.05 | 3.68 |
| 19 |  | 90th Percentile | 6.73 | 7.13 | 4.71 |
| 20 |  | 75th Percentile | 5.24 | 6.02 | 4.28 |
| 21 |  | Maximum | 6.77 | 7.65 | 7.87 |
| 22 |  | Minimum | 2.26 | 0.77 | 1.16 |
| 23 |  | N | 10 | 10 | 10 |

| 15 |  | Calculations | June | July | August |  |
|---|---|---|---|---|---|---|
| 16 |  | 25th Percentile | =PERCENTILE(C4:C13,0.25) | =PERCENTILE(D4:D13,0.25) | =PERCENTILE(E4:E13,0.25) |  |
| 17 |  | 10th Percentile | =PERCENTILE(C4:C13,0.1) | =PERCENTILE(D4:D13,0.1) | =PERCENTILE(E4:E13,0.1) |  |
| 18 |  | Median | =MEDIAN(C4:C13) | =MEDIAN(D4:D13) | =MEDIAN(E4:E13) |  |
| 19 |  | 90th Percentile | =PERCENTILE(C4:C13,0.9) | =PERCENTILE(D4:D13,0.9) | =PERCENTILE(E4:E13,0.9) |  |
| 20 |  | 75th Percentile | =PERCENTILE(C4:C13,0.75) | =PERCENTILE(D4:D13,0.75) | =PERCENTILE(E4:E13,0.75) |  |
| 21 |  | Maximum | =MAX(C4:C13) | =MAX(D4:D13) | =MAX(E4:E13) |  |
| 22 |  | Minimum | =MIN(C4:C13) | =MIN(D4:D13) | =MIN(E4:E13) |  |
| 23 |  | N | =COUNT(C4:C13) | =COUNT(D4:D13) | =COUNT(E4:E13) |  |

13

**Step 3.** Highlight your percentile rank analysis labels and data (as shown below), then from the Excel top menu select Insert → Line → Line with Markers:
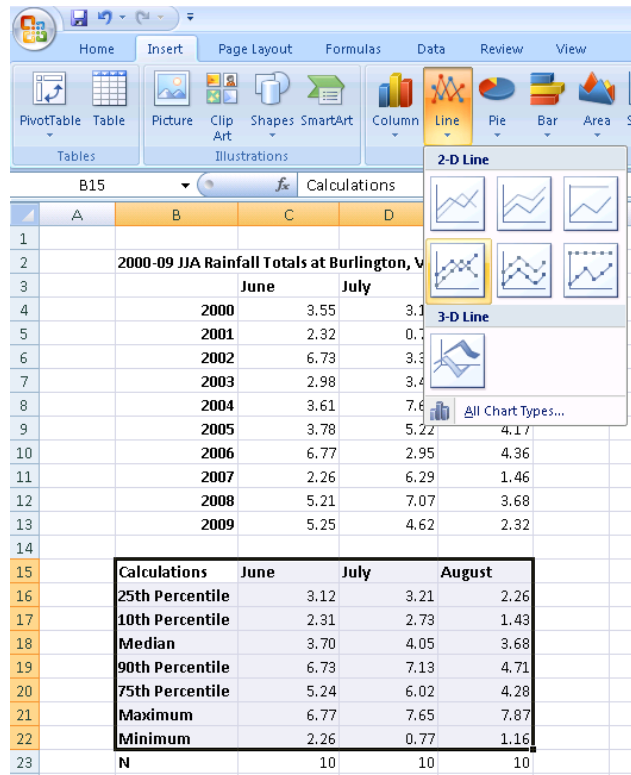


**Step 4.** If the number of columns (3) is less than the number of rows (7, as it is here), once the graph appears, you'll need to click on the graph, select Design then "Switch Row/Column"

 to get the following:
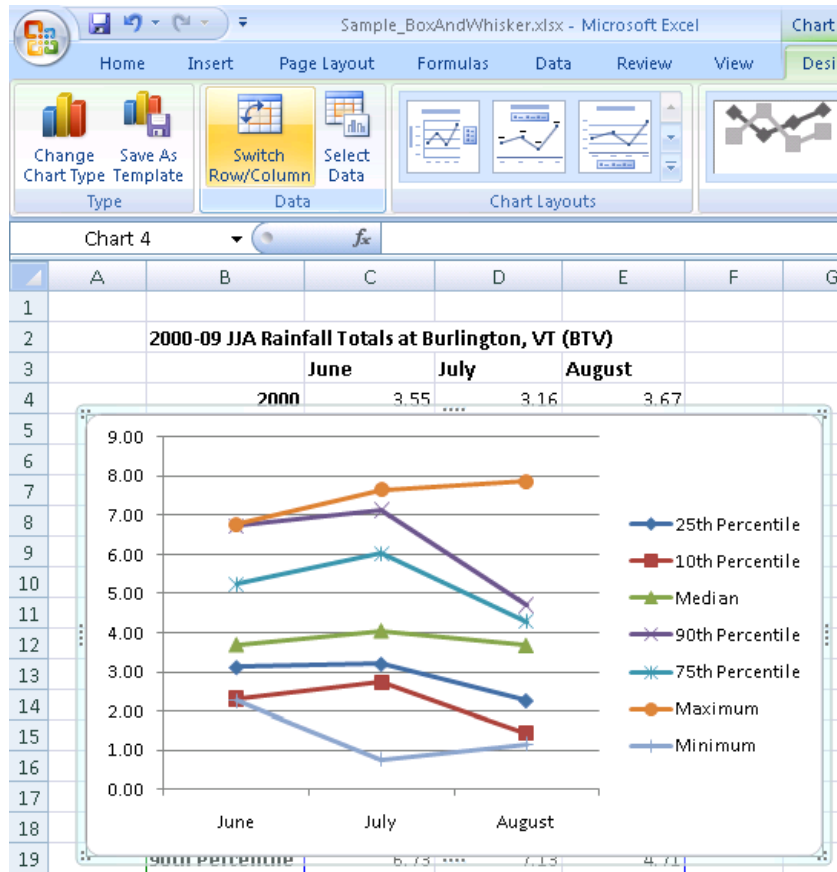
If the number of columns is greater than the number of rows, continue on to the next step.

**Step 5.** Right click on the chart and select "Move Chart…" and "New sheet" to move the graph to a separate sheet. This will make the graph easier to work with.

**Step 6.** Now we need to modify a series of chart attributes to turn the data points and lines into a box and whisker plot format. Right-click on the line corresponding to the maximum value, and select "Format Data Series…". Select "Plot Series On" → Secondary Axis and "Line Color" → "No Line".

**Step 7.** Repeat Step 6 for the minimum value line.

**Step 8.** At this point, you will need to modify the secondary axis (on the right side of the chart) to match the values on the left side. Do this by right clicking on the secondary axis and selecting "Format axis". Under "Axis Options" select the "Fixed" radio buttons for "Minimum", "Maximum", "Major Unit", and "Minor Unit" and set to the same values as along the primary axis. You may want to do the same for the primary vertical axis (on the left side of the chart) to ensure these values remain fixed as you work with your chart. At this point, your chart should look like this:
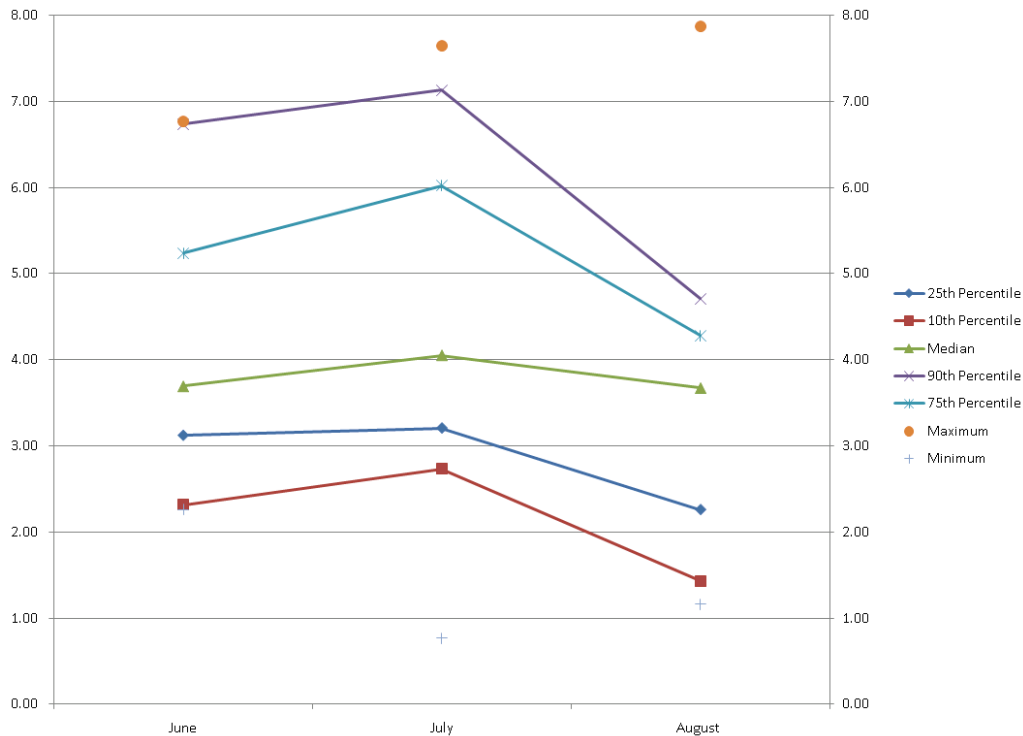
**Step 9.** Now we convert data to the box and whisker format.

**<u>Median</u>**: Right click on the median (middle) line, then select "Format Data Series". Select "Line Color" → "No line". Next, change the marker style by selecting "Marker Options". Select "Built-in" and change the marker type to the big horizontal line, and increase the size to "15". As you make these changes, you should see the modifications take place simultaneously in your chart area. Next select "Marker Fill", "Solid fill", and change the color to black with a transparency of 0%. Then select "Marker Line Color", "Solid line", and change it to black also. In order for the median line to show up in the foreground of the box later, select "Series Options", plot on "Secondary Axis". We are now done formatting the median. Click "Close".

**<u>Box</u>**: Right click on the 75th percentile, then select "Format Data Series". Select "Line Color" → "No line". Repeat for the 25th percentile. Right click again on the 75th percentile. On the top menu, select layout "Up/Down bars" → "Up/Down Bars". The bar should appear between the 25th and 75th percentile. If it does not, ensure you have the 25th, 75th percentiles plotted against the *Primary Axis* with the maximum, minimum, and median values plotted against the *Secondary Axis*. To add color fill to the box, right click on the box and select "format up bars". Then select "fill" → "Solid fill", and modify the color as desired. The median line should still display within the box. If it does not, ensure the median is plotted against the secondary axis. Lastly, we remove the individual marker points. Right click on the 75th percentile. Select "Format Data Series" and under marker options select "None". Repeat for the 25th percentile.

**<u>Whiskers</u>**: Now we need to add the whiskers. Do this by right-clicking on the 90th percentile, and then select "Format Data Series". Select "Line Color" → "No line". Repeat for the 10th percentile. Right click again on the 90th percentile. Select the chart, and from the chart tools menu at the top, select "layout" → "Lines" → "High/Low Lines". This will add the whiskers

16

from the 10th to 90th percentile points. If the whiskers do not connect from the 10th to 90th percentile points, ensure they are plotted on the *Primary Axis*. Lastly, we remove the individual marker points. Right click on the 90th percentile. Select "Format Data Series" and under marker options select "None". Repeat for the 10th percentile.

**Maximum and Minimum Values:** Meteorologists and climatologists are often interested in extreme (record) values, so we include them here along with the traditional box and whisker. The markers you select are personal preference. Here will we replace the default markers with "x" to denote the values.

Right click one of the maximum values to highlight all the maximum data points. Select "Format Data Series" → "Marker Options" → "Built-In" → Type: "x". The default size of 7 is used in this example. Under "Marker Line Color" we select "solid line", and switch the color to black with a transparency of 0%. Click close. Now we will add data labels. Right click on the maximum data points and select "Add Data Labels". The precipitation values corresponding to these data labels should appear to the right of each "x". You can also manually add the year the record occurred by left clicking within each text box until a cursor appears. In this case, we will parenthetically add the year to each label. Repeat the steps in the paragraph above for the minimum data points.

**Legend**: Since we have been removing the markers for the individual data points (except the maximum and minimum values), the legend is not particularly helpful. You can opt to simply right click the legend and select "delete". We will add additional interpretative details to the graph title a bit later on.

At this point, your graph should look something like this:
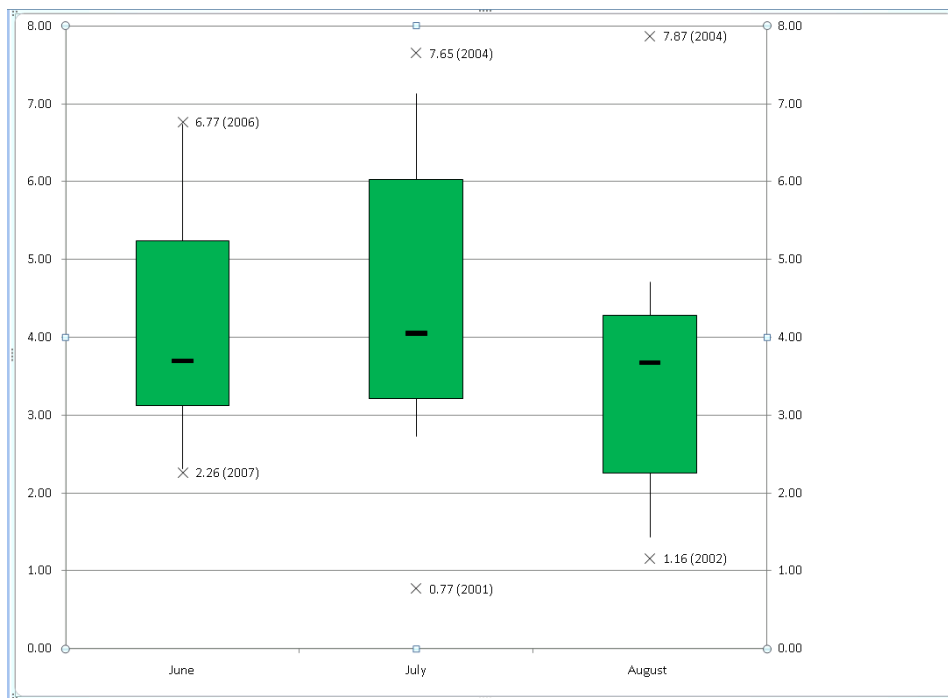
**Step 10.** The remaining steps are mainly for aesthetics/personal preference.

It is often advantageous to add data labels showing the values at the 10th, 25th, median, 75th, and 90th percentile points. Similar to the maximum and minimum data labels, if you right mouse click over the ends of the whiskers, box, and on the median, you can select the option for "Add Data Labels". The individual precipitation values will appear. You may need to manually move the labels slightly to avoid overlap with other data labels. In the case example, we will also make the median data labels bold face and slightly larger font to stand out.

We will also make the axes labels (months, precipitation) bold face for aesthetics.

When dealing with variable sample sizes between data categories, it is often desirable to add the sample size for each box and whisker to the secondary horizontal axis for the convenience of the viewer. To do this, select the chart and from the Excel top menu select "Layout" → "Axes" → "Secondary Horizontal Axis" → "Show left to right axis". The default adds the same labels as along the Primary Horizontal Axis. To change this, right mouse click on the horizontal axis you wish to modify, select "Select Data…". The Data Source Text Box will appear as follows:
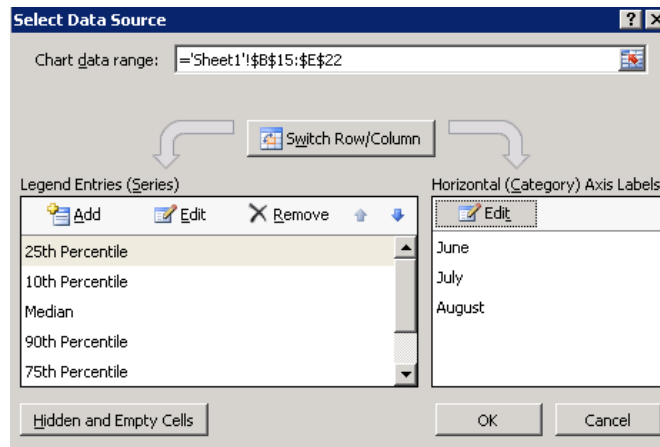


Select "Edit" under "Horizontal (Category) Axis Labels". A window will appear to highlight the Axis label range. Highlight the row of cells corresponding to the sample size, as follows:

| Calculations | June | July | August |
| --- | --- | --- | --- |
| 25th Percentile | 3.12 | 3.21 | 2.26 |
| 10th Percentile | 2.31 | 2.73 | 1.43 |
| Median | 3.70 | 4.05 | 3.68 |
| 90th Percentile | 6.73 | 7.13 | 4.71 |
| 75th Percentile | 5.24 | 6.02 | 4.28 |
| Maximum | 6.77 | 7.65 | 7.87 |
| Minimum | 2.26 | 0.77 | 1.16 |
| N | 10 | 10 | 10 |
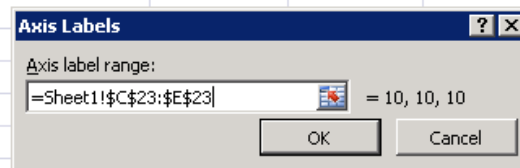
Click "OK". The axis should now show the sample size along one of the horizontal axes (below the graph in this example). That leaves us with something that looks like this:
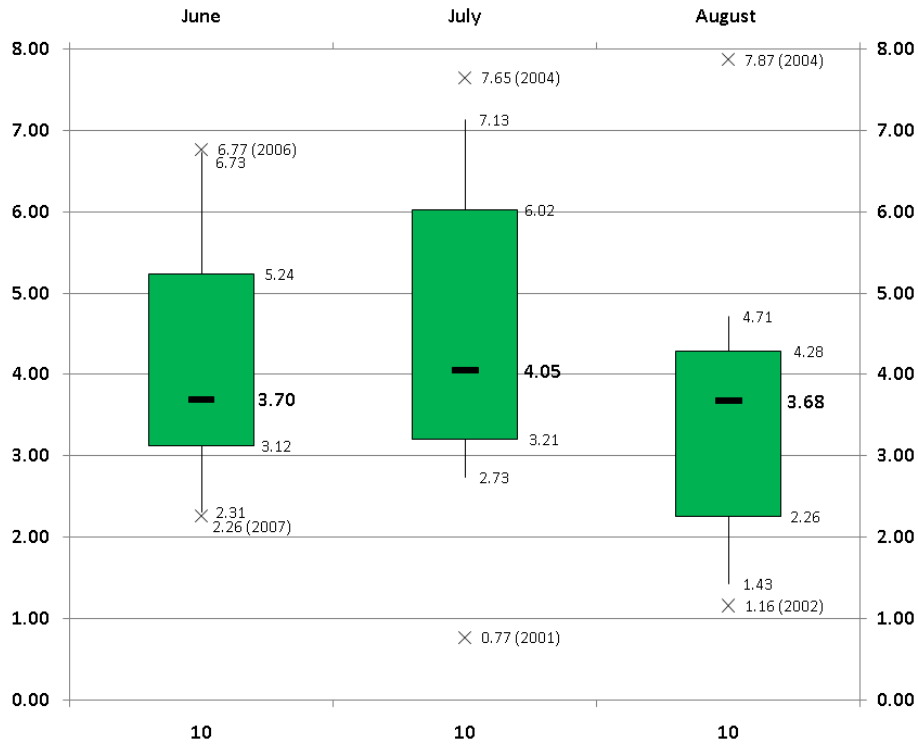


**Step 11.** The final step is to add a title, axis labels, and logos (as desired).

<u>**Title**</u>: Every good graph should include a description of the data displayed (location, time frame, data type, etc.). Since box and whisker plots can vary, it is also useful to include some information about the graphing convention.

To make the title, select the chart and from the Excel top menu, select "Layout"→ "Centered Overlay Title". Place the title within the default text box that appears.

<u>**Axes:**</u> Each axis should be labeled. Select the chart and from the Excel top menu, select "Layout" → "Axis Titles". Create axis titles for each of the four axes in this case.

<u>**Logos**</u>: Can be added as desired, or as necessary based on the dataset being used. Since these are official NWS data, we will add NWS/NOAA logos to the chart.

Our final box and whisker plot looks like this:

# Monthly Precipitation at Burlington, VT (JJA 2000-2009)

Solid Bar - Median Monthly Value; Boxes 25th-75th Percentile; Whiskers 10th/90th Percentile
(x - extreme monthly values 2000-2009)

**Month**

|  June  |  July  |  August  |
|--------|--------|----------|

- June: × 6.77 (2006); 6.73; 5.24; 3.70; 3.12; × 2.31; 2.26 (2007)
- July: × 7.65 (2004); 7.13; 6.02; 4.05; 3.21; 2.73; × 0.77 (2001)
- August: × 7.87 (2004); 4.71; 4.28; 3.68; 2.26; 1.43; × 1.16 (2002)

Precipitation (in)

Sample Size (# of Years): 10, 10, 10