# Excel Review Module 2: Charting

**Version Summer 2024**

This document is part of a series of labs which are intended to develop some skills in Excel which are oriented to students of Web-GIS information technology. Some students took Module 1 at the start of Intro to GIS. This is \*NOT\* a requirement of Advanced GIS but is offered to students who would like to acquire a few new Excel tools and sharpen up their table-logic, as optional learning.

While not specifically a mapping exercise, students will find the skills useful for supporting some mapping workflows and information products.

## Purpose of Exercise

This lab starts with a modest data management workflow, typical of many GIS project needs.

Then the lab proceeds to show how Excel can be used for the creation of a few charts for visualization of the data present in tables. The exercise ends with a light example of inferential statistics.

### Background Article

Read about the General Linear Model of inferential statistics: <https://conjointly.com/kb/general-linear-model/>

## Scenario

How can simple pie and line charts, and scatter plots help understand whether the presence of raptors is related to the temperature and humidity variations at the Pinnacles National Monument?

A map of the location of the Pinnacles, as well as a public layer of landcover types, is here:

A map of a mountain range

Description automatically generated

<https://arcg.is/1LnKqe0>

This exercise is based on observations of three species of raptors at the Pinnacles National Monument in central California.

The dataset includes collection and compilation of average monthly climate data, provided by the National Weather Service on a website.

Data analysis tools will demonstrate the use of linear regression for inferring relationships through scatterplots and the general linear model.

# Data Management

All exercises start with some data management. In this case, the source data is an Excel file containing counts of raptors observed monthly at the Pinnacles National Monument near Soledad, California.

**Summary** (steps are shown below in more detail)

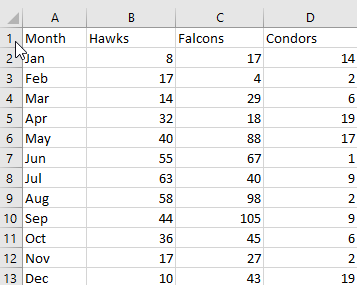
* Obtain the data file from Canvas. This is an XLSX file showing counts of three raptors, by month.
* Compile climate variables by copying them from a website and pasting these into your spreadsheet. The climate data shows average low and high temperature by month and average precipitation by month.
* But … The orientation of the climate data from the website is different than the raptor count table. You will use the *Transpose* tool in Excel to revise the orientation of the climate data. (This is a very common problem and a ***useful Excel trick***.)
* Use the paste-values command to put the revised climate data into the table containing the raptor counts.
* Evaluate the regression between climate variables and raptor presence/absence

Details for these steps are shown below …

## Get the source table of raptor counts:

This table contains counts by month for one year showing number of hawks observed, Get Source table of raptor counts, for hawks, falcons, and California Condors.

The Pinnacles Monument is famous for the condors that reside there during parts of the year. The source table looks like the following screenshot:

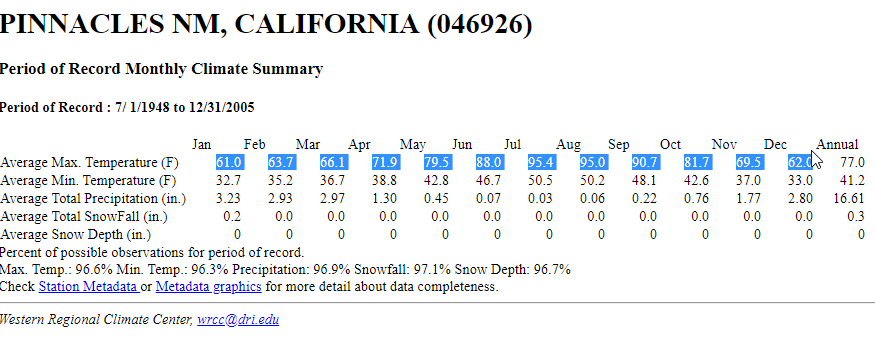


## Get the Climate Data

Go to website to get climate data:

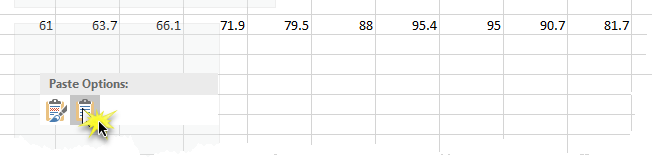
<https://wrcc.dri.edu/cgi-bin/cliMAIN.pl?capinn+sca>

The data is organized by columns:



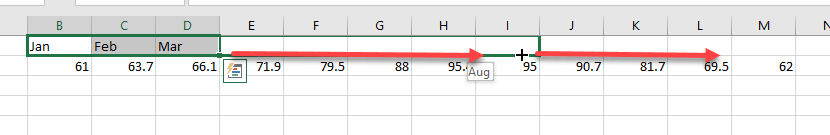
Select the highlighted cells (Ave Max Temp F) and copy these to clipboard (Ctrl C)

Paste these into a worksheet. Select option to “match destination formatting”



Enter on the row above “Jan Feb Mar”

Drag handle to right to complete to December.



Now you have copied data into spreadsheet – but notice that the data table is horizontally organized.

## Save your worksheet in a workbook .xlsx file

Click Save and provide a name of your Excel sheet, in a folder of your own choosing.

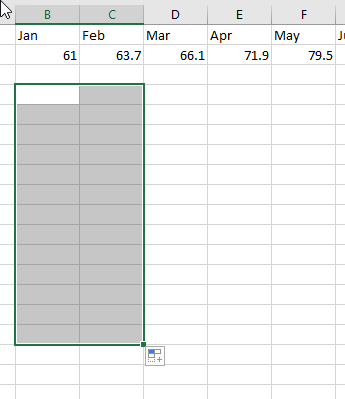
Save this as an Excel Workbook (.xlsx) format.

## Transpose the Climate Data

Sometimes you get data from a website which is organized in rows. But often you need to re-organize it into columns. Or vice versa.

The Transpose function can manage the data to change the orientation.

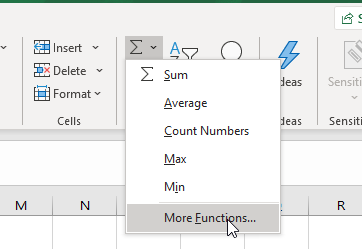
Use your mouse to select a group of *empty* cells which are at least 2 wide and 12 deep. In the screenshot below, that is the area selected in the green box, shown as near the data table.



On the formula line type – note the starting = sign, and the open parenthesis after the word TRANSPOSE

=TRANSPOSE(

When you start on the Excel entry line with an “=” sign, this signals Excel to generate a computation or a function. There are other ways to initiate the function, such as using the function button and searching “More Functions” to find “transpose”.



There are scores of functions available in Excel, many worth trying out to use in your own research and other projects.

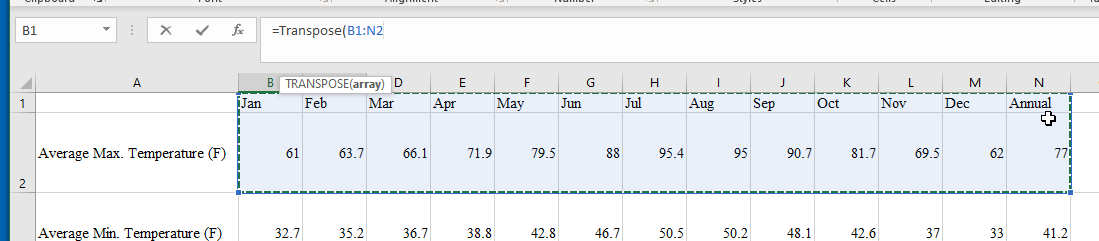
This expression =TRANSPOSE(”) initiates the transposition function of Excel.

With the one parenthesis opened, the function line is waiting for the rest of the parameters to place into the function.

The parameters will point to the array of raw data values in your sheet.

Click on the **Maximum temperature** array to select all of the source values (including the months).

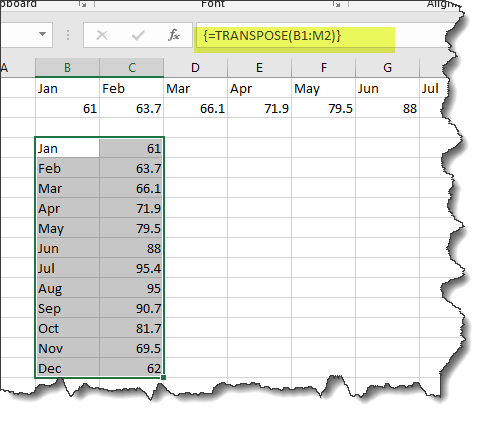
There are other rows in the source table for Minimum temp and humidity we will cover in a moment. But to see how Transpose works in the simplest case … use the following selection:



**Do not close the parenthesis**. Instead, click together Control-Shift-Enter.

The formula changes to show the open/closed brackets { …. } which designate an *array function*.

The result in the XLS is:



As you can verify for yourself, the values that were in rows are now transposed to columns.

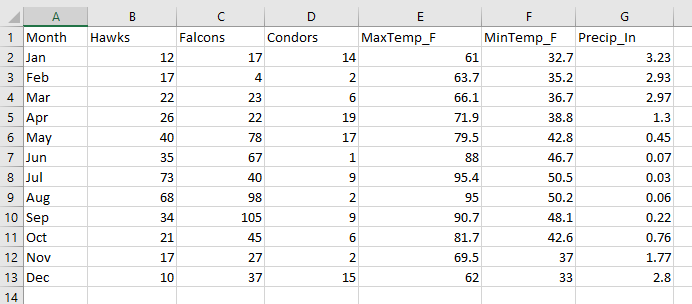
Copy the C column values (Control-C). Use the “Special” Paste-Values function to paste these into your compilation sheet, aligned by month to the existing raptor counts.

Repeat these steps to create a table that includes a column for average maximum monthly temp, average minimum monthly temp, and average total precipitation.

Paste-Values is a very important function for data management. If you merely use Paste, the numbers in the pasted columns will retain a reference to the cells they derive from. Sometimes you want to retain the reference, so that if the calculations change, your pasted values change accordingly. But sometimes you just want numbers, unencumbered by their source computation logic. Paste-Values puts the numbers (or other data) into your target cells after eliminating any logic, computations, or references used to compute the cells. Read more: <https://support.microsoft.com/en-us/office/paste-options-8ea795b0-87cd-46af-9b59-ed4d8b1669ad>

Well worth being aware of this Excel trick!

Your compilation sheet should look similar to:



Now You have a table of numbers, where both the bird observation data and the climate data are organized in a “flat file” format, by month.

The term “flat” is shorthand for a simple data matrix format for this table. Each record is a row, and all rows have the same columns. Each column is a single data type (in this table either integer or decimal number types). There is no text in each numerical column (which would violate the data type for that column). There are no merged cells, no sums or subtotals. There are no colors or highlights or in-cell notes. This simple (or “flat”) data matrix enables Excel to function more as a database, rather than as a “nice” document.

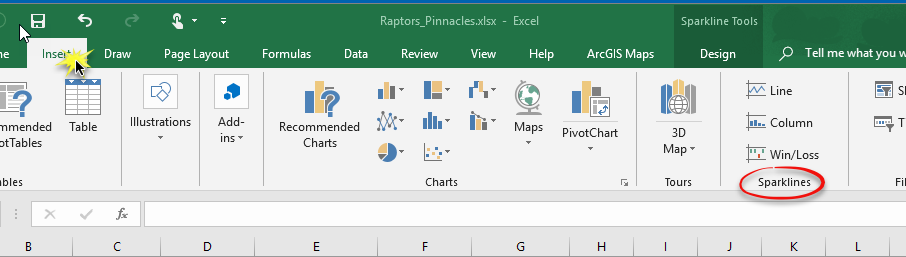
## Review Data Distribution Using Sparklines

It is often helpful to have a quick view of the data distribution. You can use sparklines to get a rapid visualization of a dataset, before making the steps to create more formal charts.

Sparklines are the function in Excel that creates a tiny histogram within a cell.

Click on the bottom of the column of counts for Hawks, in the cell just below December. In my sheet, this is cell B14.

On the Insert tab, find the Sparklines section.



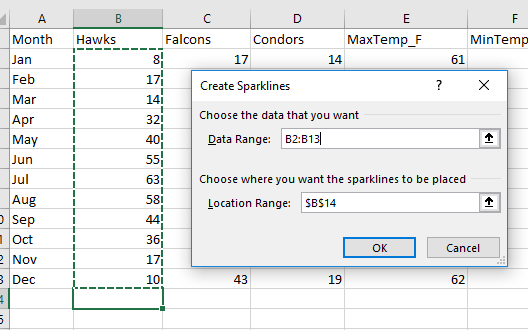
Select the option for “Line”

A screenshot of a computer screen

Description automatically generated

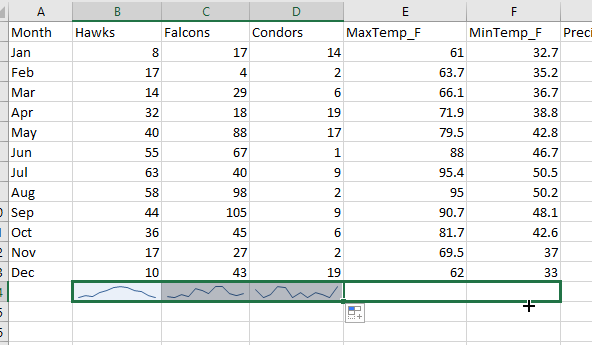
|  |  |  |
| --- | --- | --- |
| Your specific Location Range cell could be different than in the example above. |  |  |

This selects the Data Range. You already selected the Location Range.



Click OK and you should see a plot of the data range appear in B14.

In the next screenshot you can see how to grab that column B sparkline and extend it across the other columns, to generate the same type of data summary graphic.



It’s a simple chart, but surprisingly informative, considering the brevity of its presentation.

# Charting

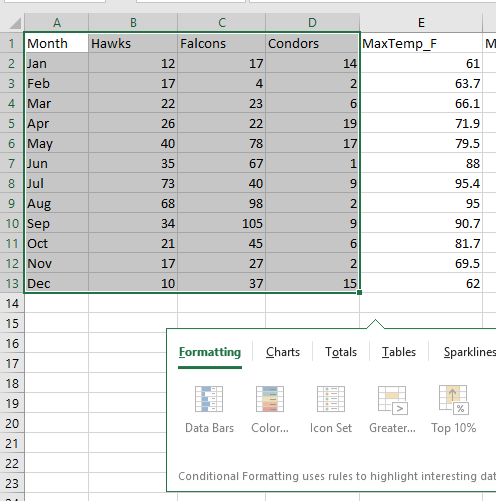
Excel has much more powerful charting functions which are useful if you want to understand some of the relationships and statistics in a data table. It’s usually easier to use Excel for data ‘exploration’ through charting. As compared to making charts in the GIS, which, though possible, is often more cumbersome.

Some students may already know how to use R for charting. Students tell me that R has the best and broadest options for making complex and informative graphics. Excel is easier (if you don’t know R), and may accomplish much of what you need to visualize with simple out-of-the-box buttons, as in this exercise.

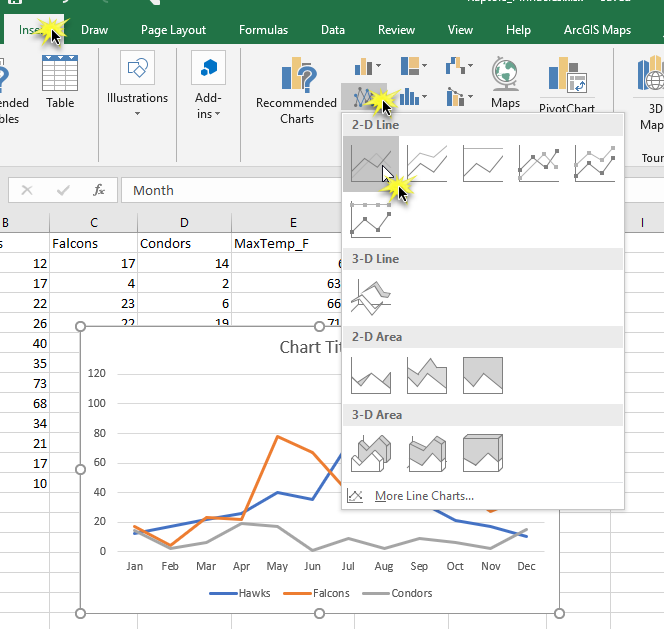
## Line Chart

We can use charting to explore the relative values of raptor counts by month at Pinnacles.

Select the cells with values to chart. Include the months, as well as the counts of the three types of raptors.

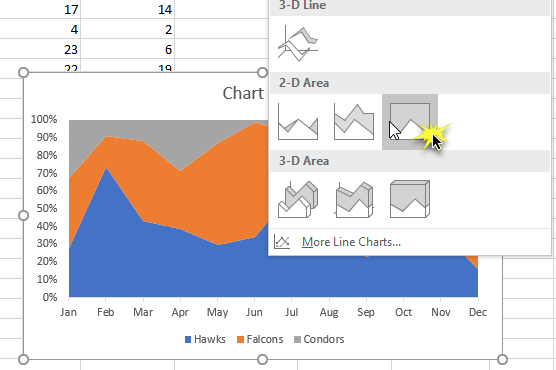


Under the Insert menu, find Charts and the most simple “2D Line” template.



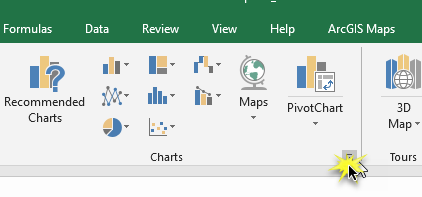
You can hover-over all the templates present to see which style of chart provides you the most valuable presentation, depending on your need.

Do you want to see *total counts* by raptor type? Or *percentage of total* by each type? There are many options to start setting up your chart.



You are not limited by the chart type or choice you made initially.

Click the pop-out for charts …



This will open a broad selection of potential charts. Excel has made it quite easy to view the options and help you identify the type of visualization you want, for your personal research purpose.

You can use their “recommended” charts, or click to “all charts” to see the wide variety. Here showing a specialized “surface” chart, for examples.

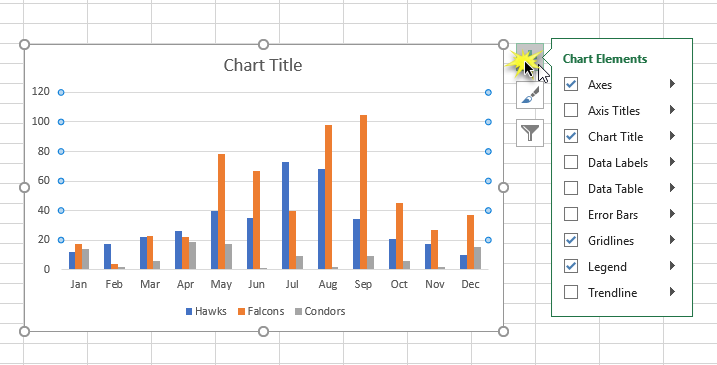
|  |  |
| --- | --- |
| Recommended Charts (example) | All Charts (example) |
|  |  |

Once you like a chart “style” you can click once to insert the chart into your spreadsheet.

You may prefer to cut and paste the chart into a separate tab in the worksheet document. It is a bit easier to keep track of the chart that way and you can change the dimensions to make it into the format you can later paste into your report or other documents.

Once the chart is inserted to the worksheet, you can make modifications to it, such as colors, titles, legends.

Click once on the chart. This will show a highlight with handles you can use to make the chart larger or smaller.



## Chart Layout Options

Once activated, there are controls next to the chart useful for changing the specifics of the chart presentation.

Click on the box that says “Chart Title”. You can overwrite that default with a better title.

The green cross opens an options menu to enable chart elements.

Enable the “Axis Titles” to be visible. Click on each of those to insert a word or short phrase

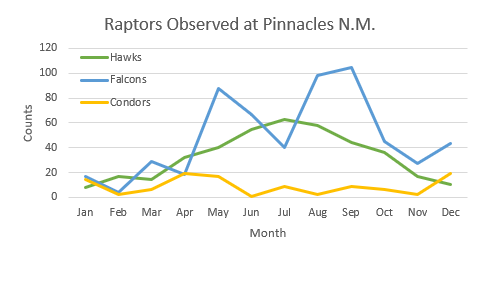
Move the legend to be inside the plot area but not overlapping the data values. (Unclick “Show the legend without overlapping the chart, that should let you position the legend anywhere you like within the chart.).

The paintbrush icon opens a dialogue for changing the chart style, and a color menu to change colors for the data elements.

Experiment with different colors. Notice that your color “ramp” selection carries through to the legend colors (of course!).

There are methods to change the range of colors available, as well.

As an example, my chart looks like the following (but you likely made your own design decisions, differently.



## Proportional Charts

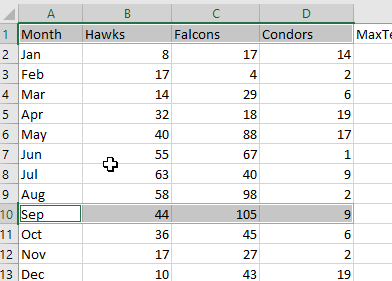
Sometimes you may only need to show comparative proportions of observations, rather than counts.

A “Pie Chart” (or “Rosette”) is commonly used for this purpose.

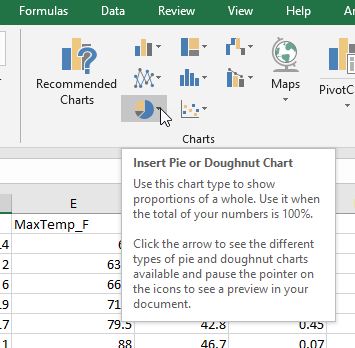
Select a month for all three raptor types. In the example I am going to compare the proportions of the three raptor types observed during September.

Select the header cells (these will provide titles for your values in the chart).

Then using the Ctrl key, select a month to create in your chart.



Under the Insert menu, select the “pie” family.

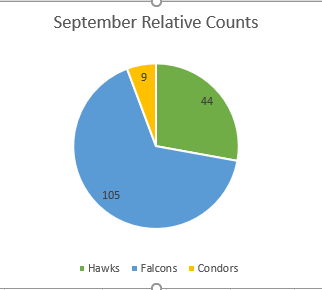


Using the Chart Elements tool, enable the data values to be shown.

Enter a title.

Pick colors to agree with your selection for the line chart, which you created above.

Example:

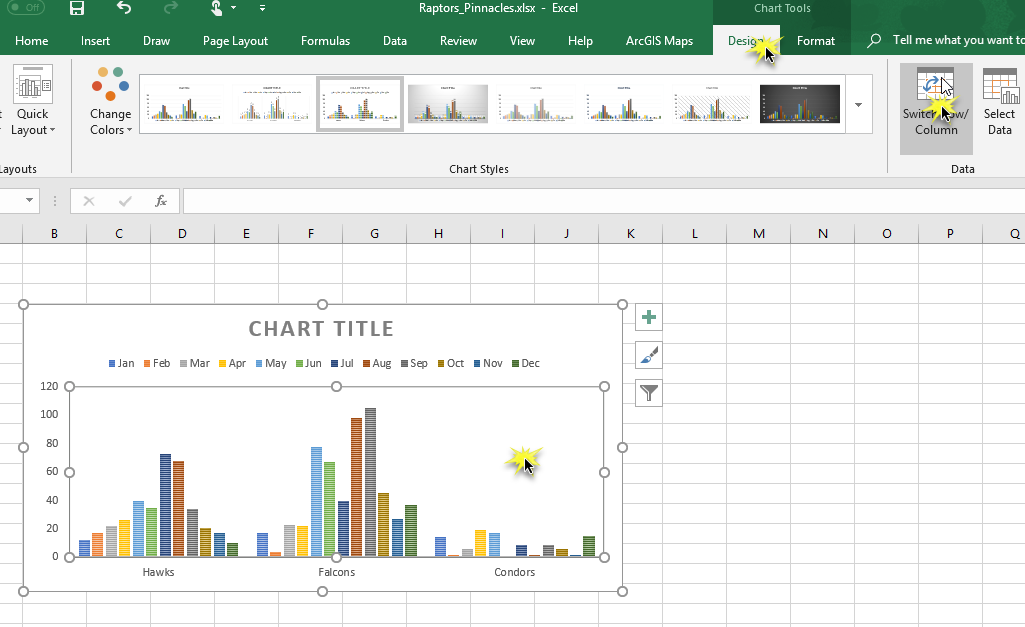


## Reverse the row and column priority

The default layout places the time (months) onto the independent (X) axis, and the raptor counts by species on the dependent (Y) axis. You can reverse that order.

Click once on the interior data part of the chart.

Click the Design tab and find “Switch Row/Column”. This will show the types of the raptors on the X axis and the counts by month on the Y axis.



# Scatter plots (Simple regression analysis)

Is there a significant relationship between the counts of raptors, and any of the climate variables you have added into the data table of raptor observations?

Scatterplots are an entry point to the more deep topic of statistical prediction.

Sometimes a quick view of relationships can be interpreted from a scatterplot. Scatter plots also allow a quick assessment of linear regression.

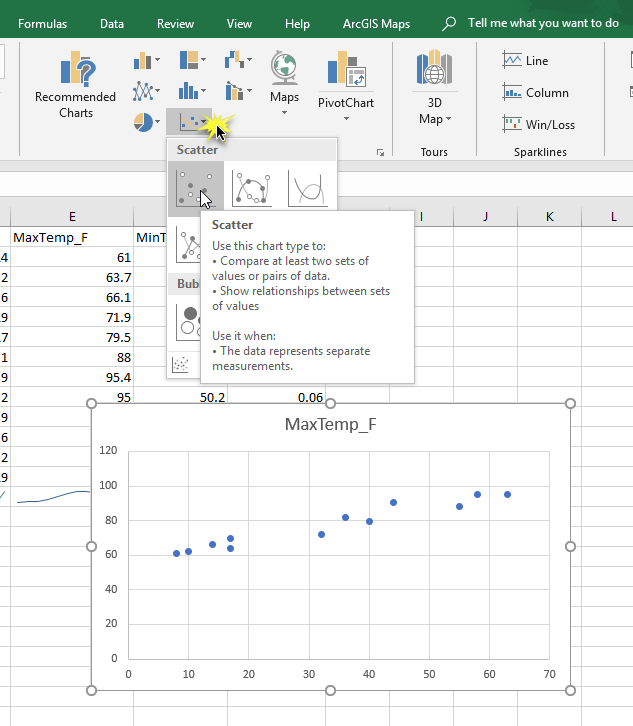
## Select data to put into scatter plot

Select two columns. For the purpose of the exercise, we will analyze hawks in relation to climate variables.

The next section will require a bit of trial and error to use the shift key and control key to make table selections work together. The first time it might be a bit awkward.

|  |  |  |
| --- | --- | --- |
| Select the range of values for the Hawks column. | Hold down the Ctrl key and select the Maxtemp column. |  |

Choose the chart option for “Scatter”



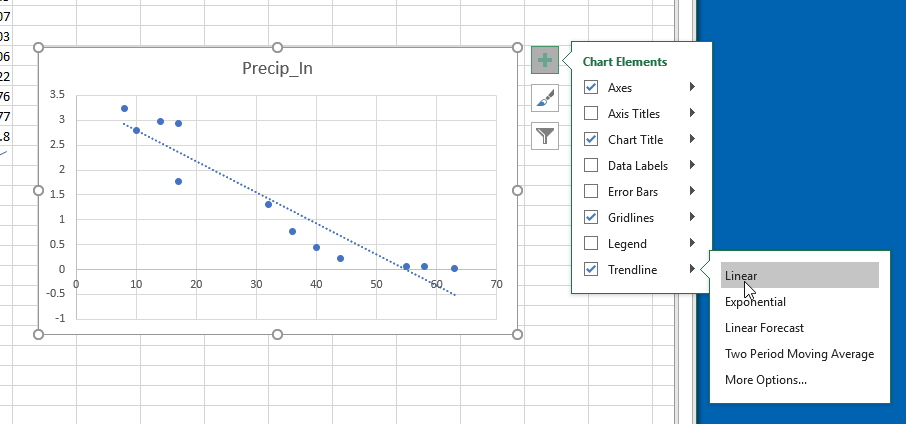
This seems to imply a pretty strong linear relationship between temperature and counts of hawks. That is, the data appears to show that when temperature is higher, more raptors are observed.

Run this for Falcons and Condors by selecting different columns. You can see that the condor relationship is less significant.

## Evaluate regression values related to average monthly precipitation

Use a scatter plot to compare Hawks with inches of average precipitation. This also has a pretty significant trendline.

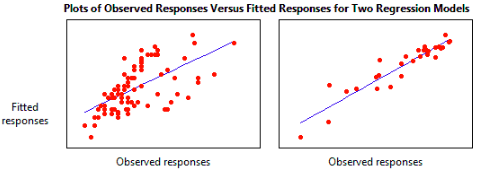
Use a more detailed scatter chart layout to evaluate the regression properties of the hawks relative to precipitation. Using the Chart Elements menu, show the trendline for a linear relationship.



Under “More Options” show the R-squared value. This represents the correlation coefficient of the data relationship.

The range of R-squared is from 0 to 100. A value of 0 indicates a poor fit between data and the fitted regression line. A value of 100 is a “perfect” fit between data and the fitted regression line. So the higher your R-squared, then the better your data fits the computed regression line. That makes your model more reliable and results more likely to be predictive. If values of R-squared are low, then you probably should not try to predict or model any variables from this regression.

For example, the two following graphics show (on the left) a low R-squared value. On the right is a better fit with a higher R-squared value. (These are graphic exhibits unrelated to the raptor data and shown just for reference.)



The R-squared value on the left example is 38. The R-squared in the example on the right example is 87.

(Read more about the value of R-square values, from this public blog:

<http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>)

Display the linear regression equation.

If you want more practice, repeat the analysis for Condors and Falcons. The significance of the observations and relationship to weather is different for each species, as you might expect.

# For upload to Canvas:

Create a screenshot of the scatter plot. Review the scatterplot and other graphics.

What are some questions about the data source that might affect the reliability of the statistical analysis?

Based on your understanding of R-squared and your plots, what can you infer about the relative significance (or “strength”) of the relationship between hawks and condors and average monthly precipitation values?