

## Building a Chart Using Trick or Treat Data – a step by step guide

By Jeffrey A. Shaffer

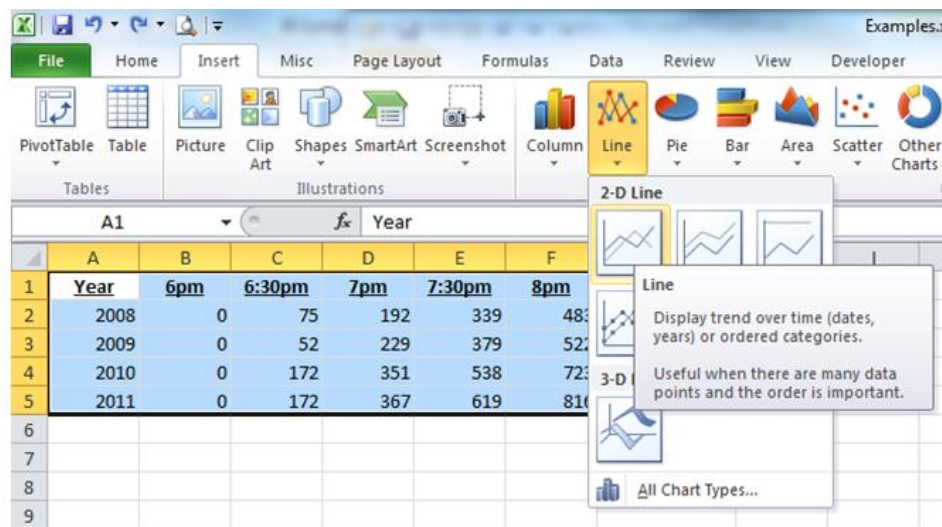
Each year my home is bombarded on Halloween with an incredible amount of Trick or Treaters. So what else would an analytics and data visualization guy do during this holiday, but to gather and display data over the years. Every business has some sort of data that looks like this, for example customer count, sales count or maybe inventory count. I've captured this data with the help of a friend and I use this as an example in how to build a data visualization in this step by step guide.

As a starting point this is what the information looks like in Microsoft Excel (and yes, we did actually have 869 Trick or Treaters this past Halloween).

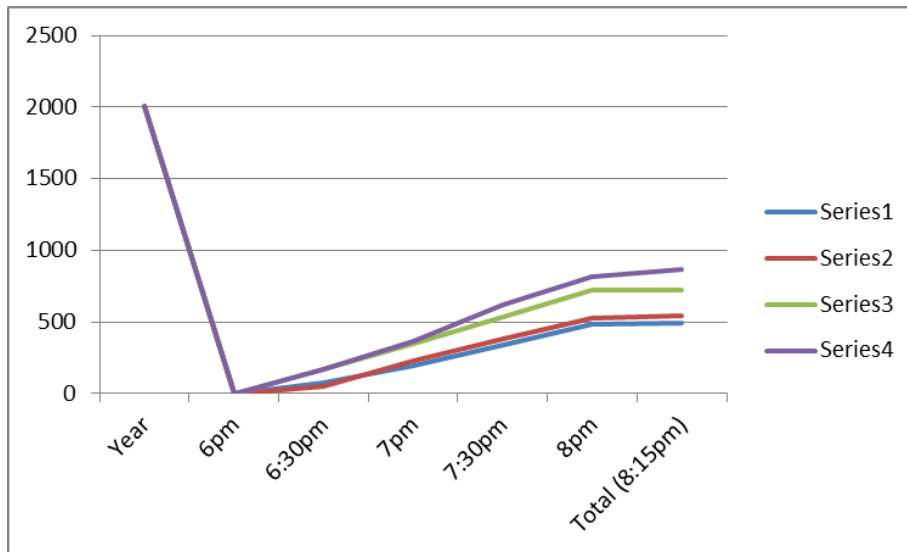
	A	B	C	D	E	F	G
1	<u>Year</u>	<u>6pm</u>	<u>6:30pm</u>	<u>7pm</u>	<u>7:30pm</u>	<u>8pm</u>	<u>Total (8:15pm)</u>
2	2008	0	75	192	339	483	492
3	2009	0	52	229	379	522	542
4	2010	0	172	351	538	723	726
5	2011	0	172	367	619	816	869

Since we are dealing with time series data, in this case 30 minute increments mostly, then a good choice is always a line chart. Line charts are the best way to visualize most time series data whether it be yearly, quarterly, monthly, daily or hourly information. There are exceptions to this, for example when comparing categorical information over time a bar chart could also work well. In this case, I simply have a running total over time so a simple line chart is best.

Highlight the cells and click the Insert Tab, Line and then select the 2D Line Chart:



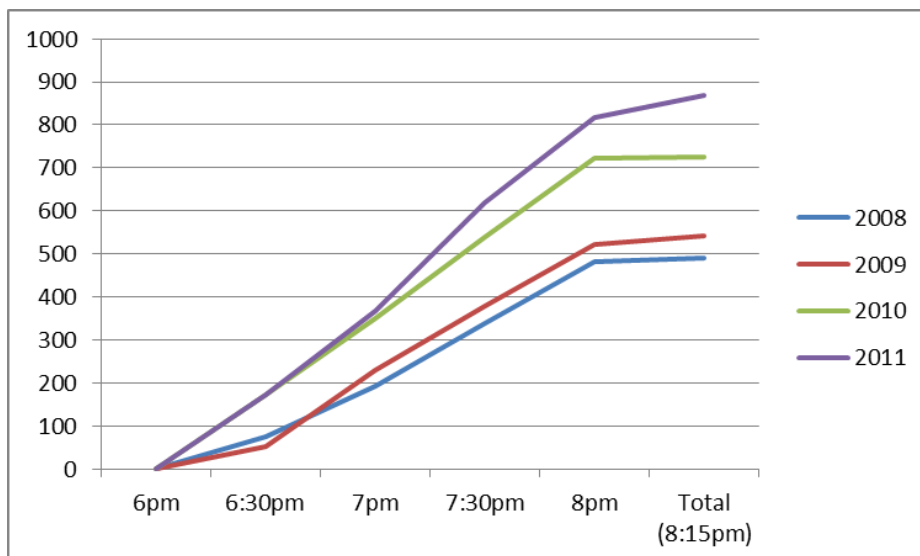
By default you will get a line chart that will look like this:



Excel assumes that the “Year” is another category of data instead of the series. There are a number of ways to fix this, but the easiest is to simply delete the cell with the label “Year” in it so that it looks like this.

	A	B	C	D	E	F	G
1		<u>6pm</u>	<u>6:30pm</u>	<u>7pm</u>	<u>7:30pm</u>	<u>8pm</u>	<u>Total (8:15pm)</u>
2	2008	0	75	192	339	483	492
3	2009	0	52	229	379	522	542
4	2010	0	172	351	538	723	726
5	2011	0	172	367	619	816	869

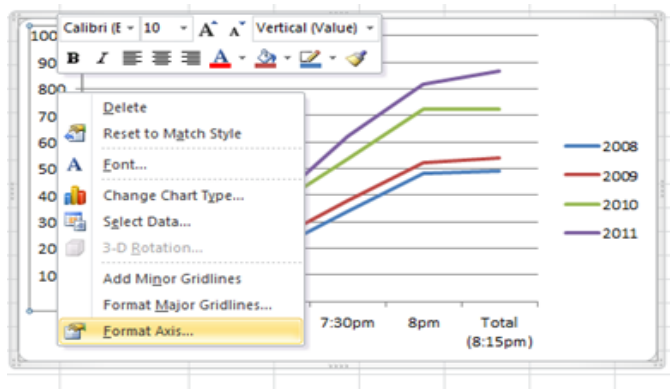
Follow the same instructions as above and you should now have a chart that will look like this.



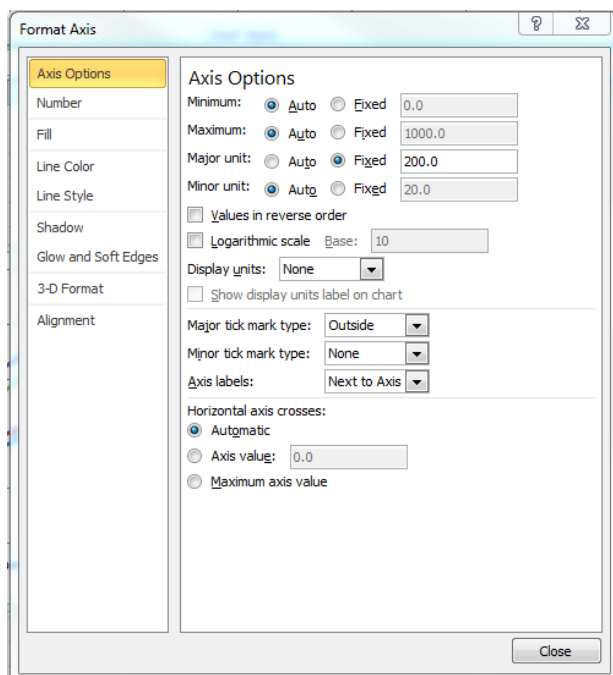
Notice in the new chart that the year in now the series and each color representing a different year. Also notice that Excel creates a very poor data visualization by default. There are a number of things that need to be cleaned up. There are 3 things to clean up right away.

- Axis labels at intervals of 100, creating 10 gridlines. This is unnecessary and will often times create a moiré effect.
- The gridlines are black by default so they are front and center and can often dominate the visual focus.
- By default Excel will automatically insert a border around any chart that is created.

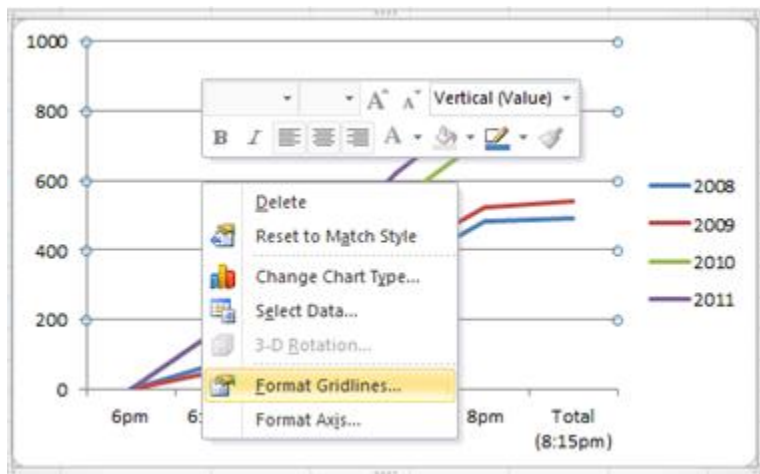
These are all very easy to fix. Right-click on the axis and select Format Axis.



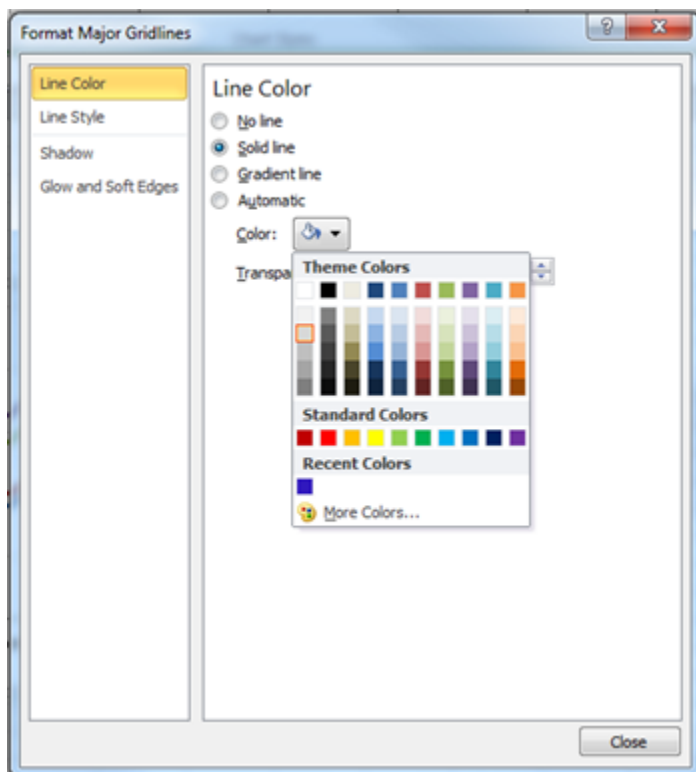
Set the Major Unit to Fixed and enter a value of 200. This will cut the number of gridlines in half. Keep in mind that this fixed value will depend on the data. Use careful judgment in determining the right cut-off, but try to avoid a graph with too many gridlines.



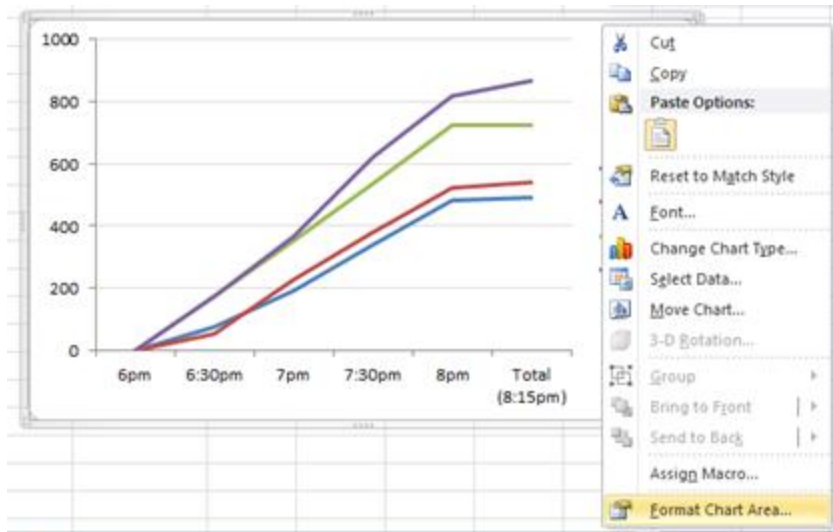
Then click on the gridlines, right click and select Format Gridlines.



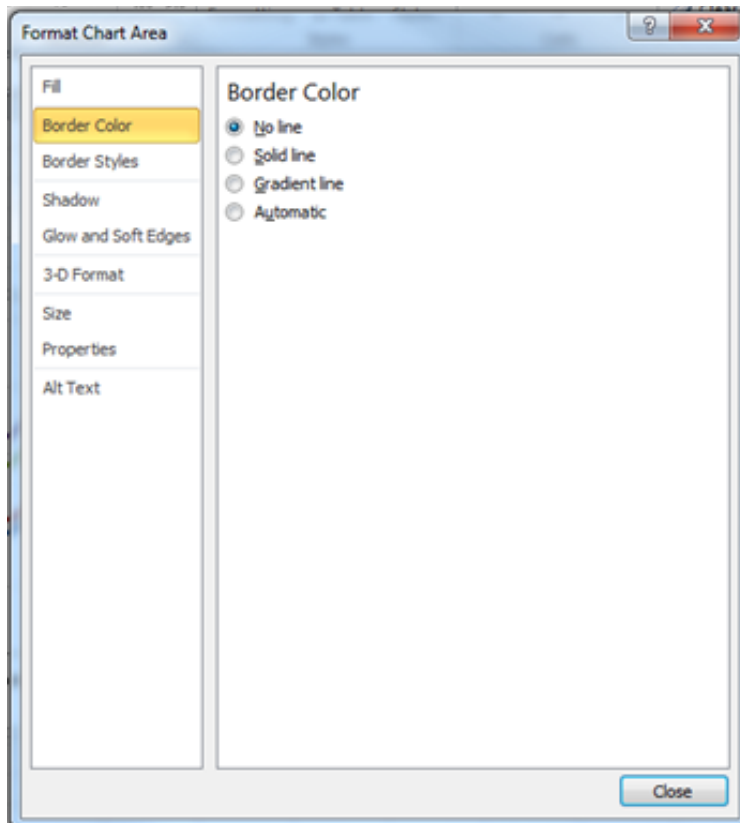
Then select Solid Line and click the drop down box for Color. I typically recommend a light gray color and often use the second one down on the left hand side of the color palette. This will mute the gridlines so that they don't dominate the graph.



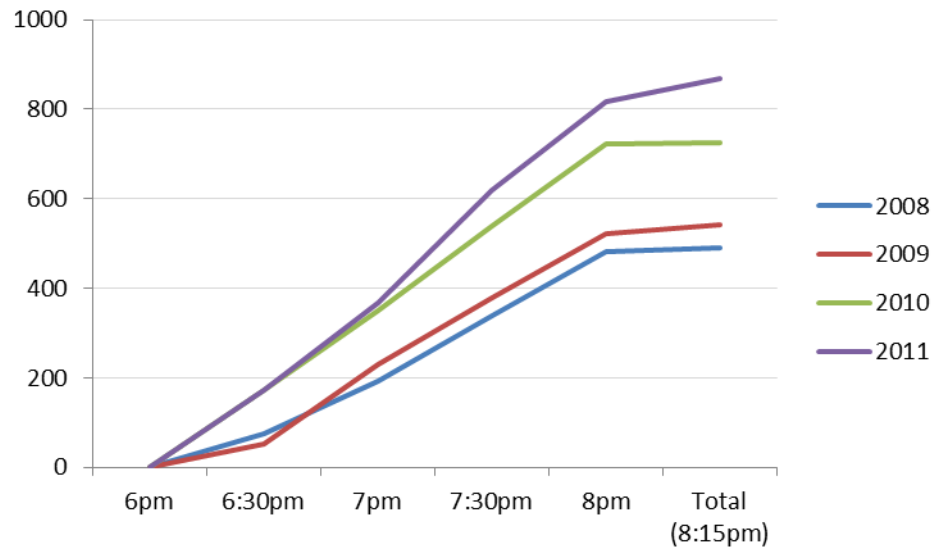
As a last step, right click on the entire chart area and select Format Chart Area.



Then click Border Color and select No Line. This will eliminate the border around the entire chart area.



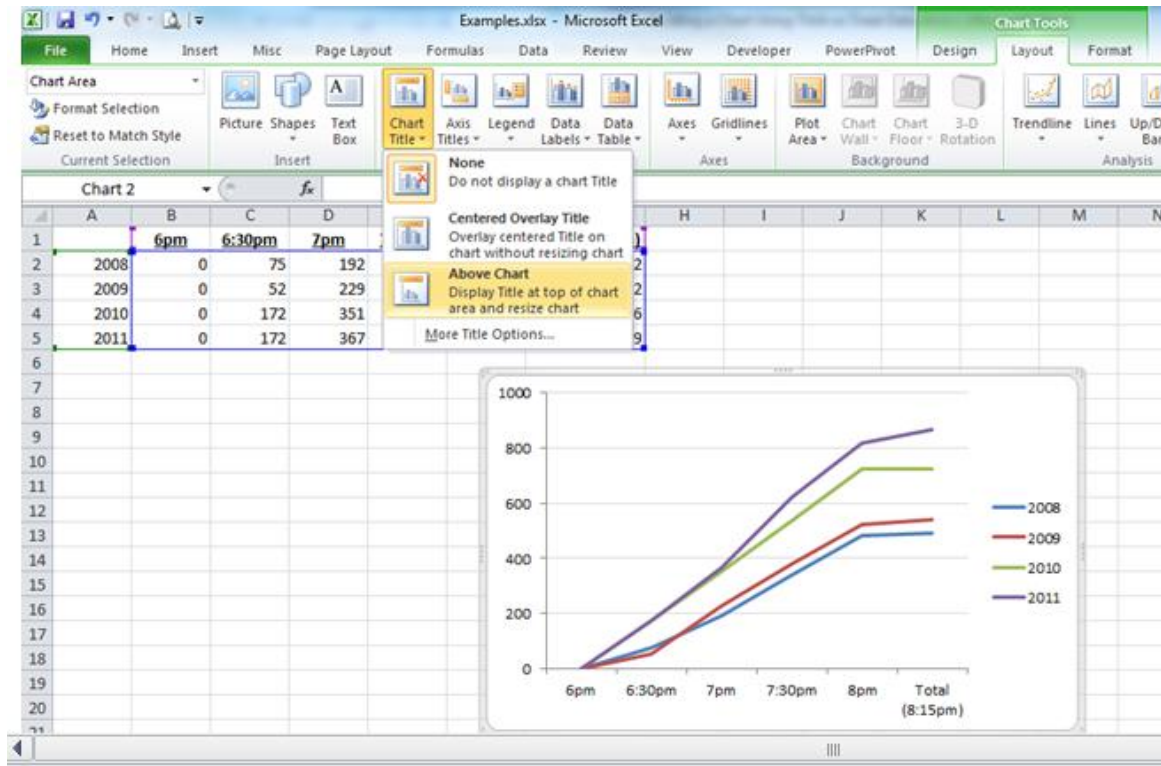
The chart should now look like this.



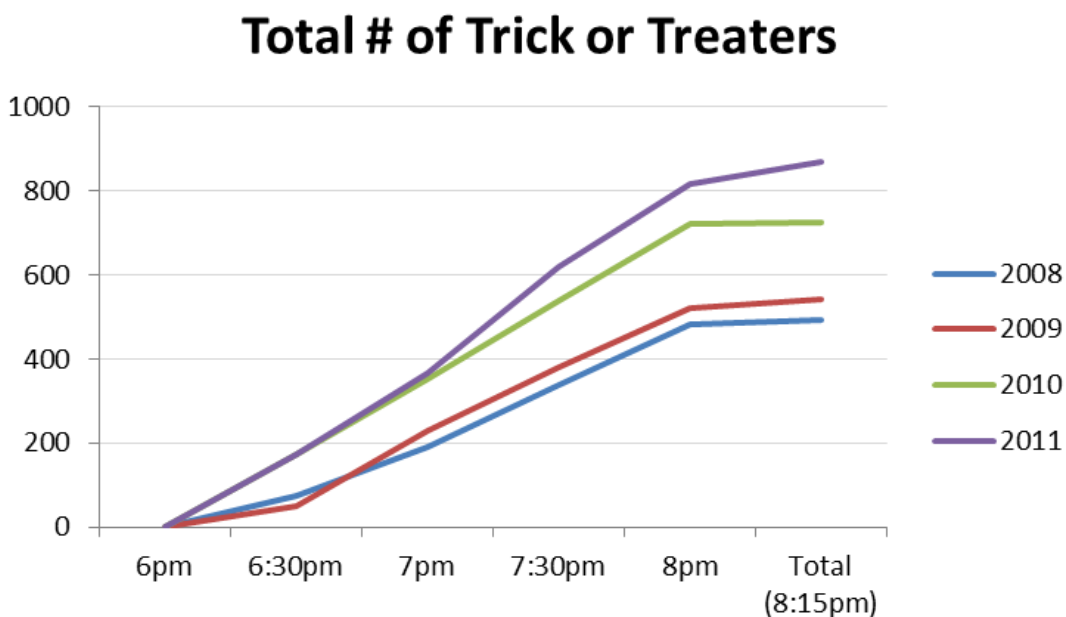
These 3 steps can be used with almost any chart in Excel and will go a long way in making a chart cleaner and easier to read. However, there are a number of things we can do to make this even better.

- There is no title on this chart. What does this chart represent?
- There is no data available on this chart. There are only estimates based on the y-axis that allow the reader to approximate the values of each series over time.
- The default color choices in Excel will look like every other default chart in the world. It will also make it very hard for a color blind person to distinguish between the line colors.
- The series is ordered in the order of the original data set. This is not optimal for our legend as it is backwards when compared to the ends of the line chart. Also, when we add a data table in the next few steps it will be even more confusing.

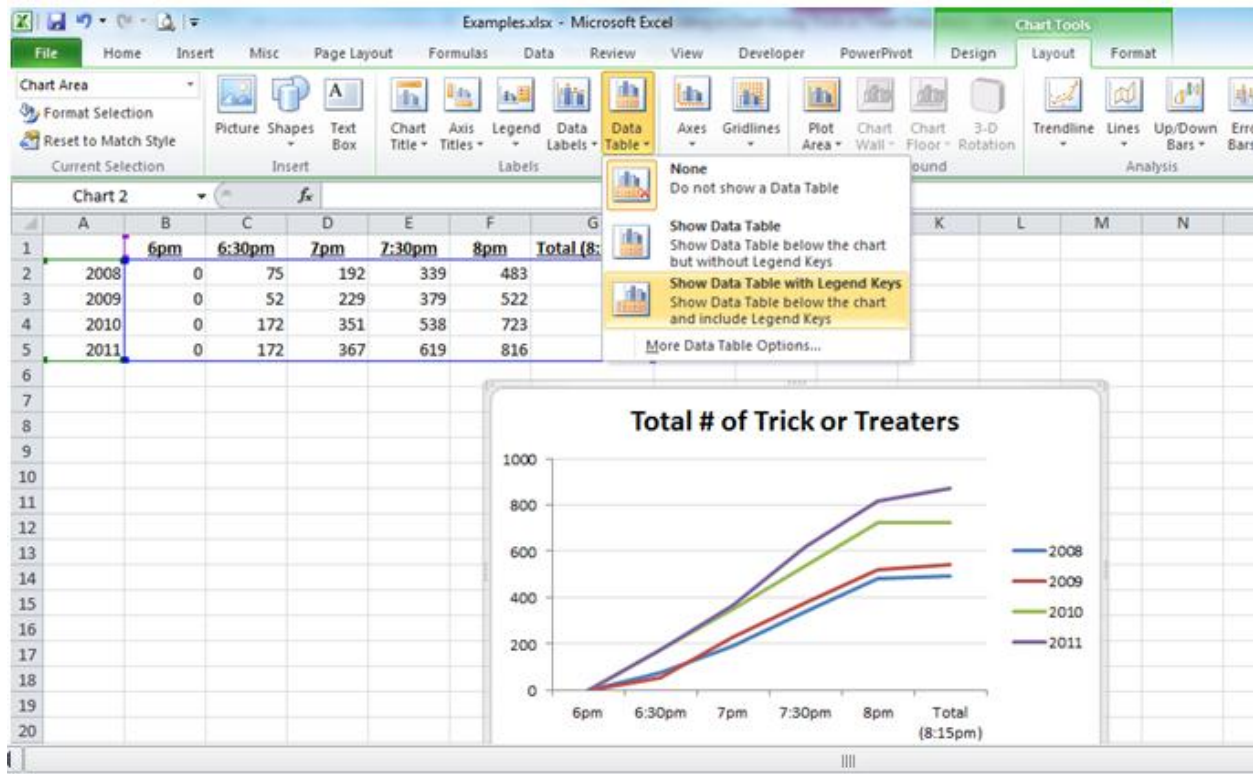
Again, there are all simple changes. First, click on the chart area and then select the Layout tab, click Chart Title and select Above Chart.



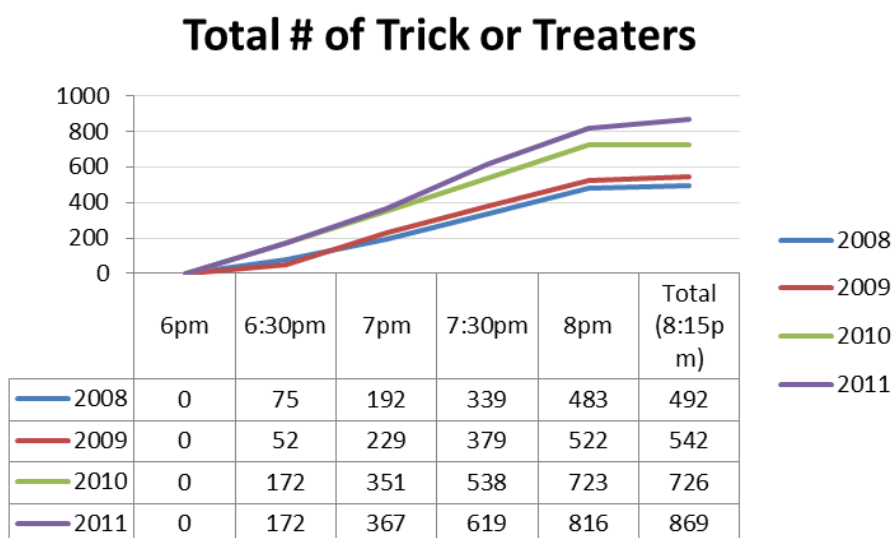
Enter an appropriate title. Use this opportunity to include a good description of the chart. You may also want to include a date or information about who created the chart. For now, enter the text "Total # of Trick or Treaters". The font size, color and bold can also be adjusted, but let's leave that for now.



Next, click on the chart area again and select the Layout tab again, but this time click Data Table and select Show Data Table with Legend Keys.



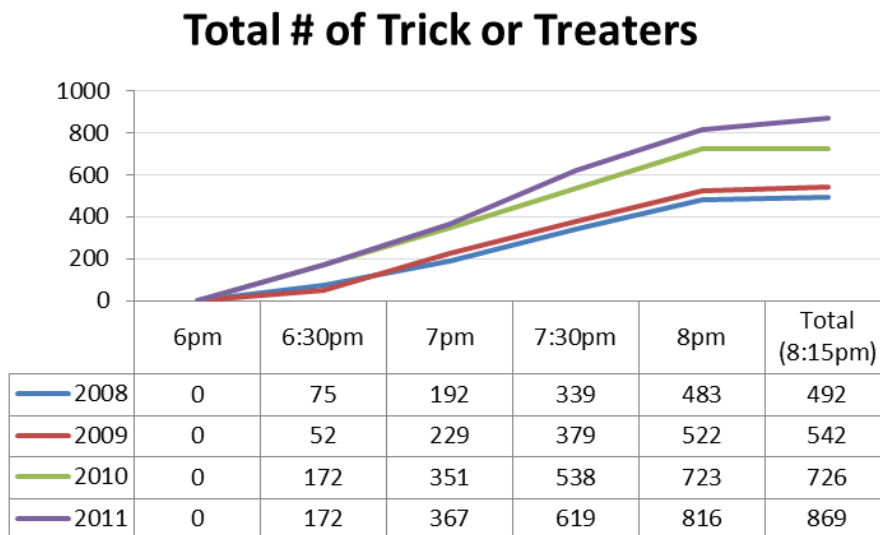
The chart now looks like this:



There is now some data in our chart, providing the user with useful information for comparison. However, there are also 2 legends and the Total column is too narrow for the label.



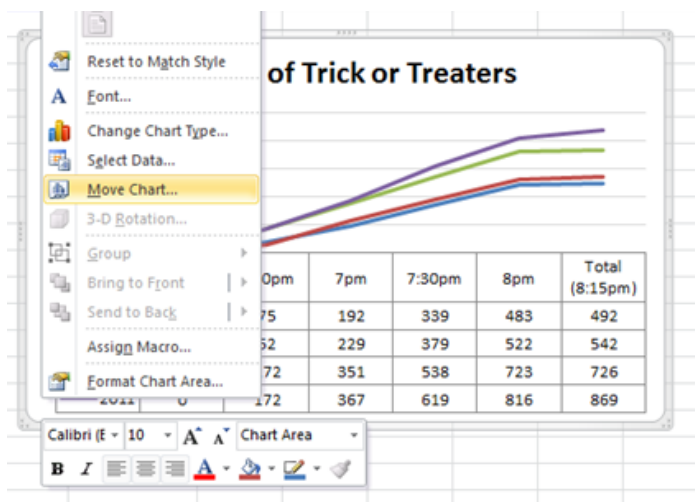
Click on the legend on the right and hit the delete key on the keyboard. This should solve both issues and the chart should look like this.



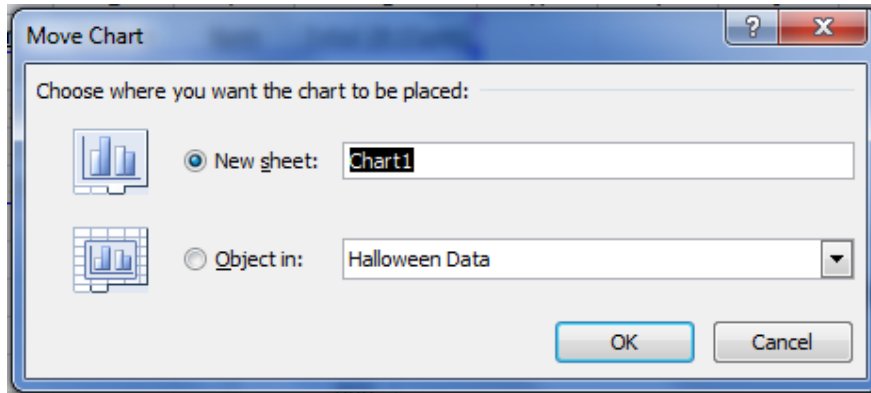
Adding the data table has now lowered the height of the chart forcing the gridlines closer together and making it look too busy again. This is an important lesson. There are two important factors that cause issues with gridlines (other than the default black color that is now muted to a light gray).

- 1.) The number of gridlines, specifically too many.
- 2.) The distance of the lines from one another. In this case we have the exact same number of lines as before (5), but because the distance between the lines is now closer the gridlines appear to be too many for the space.

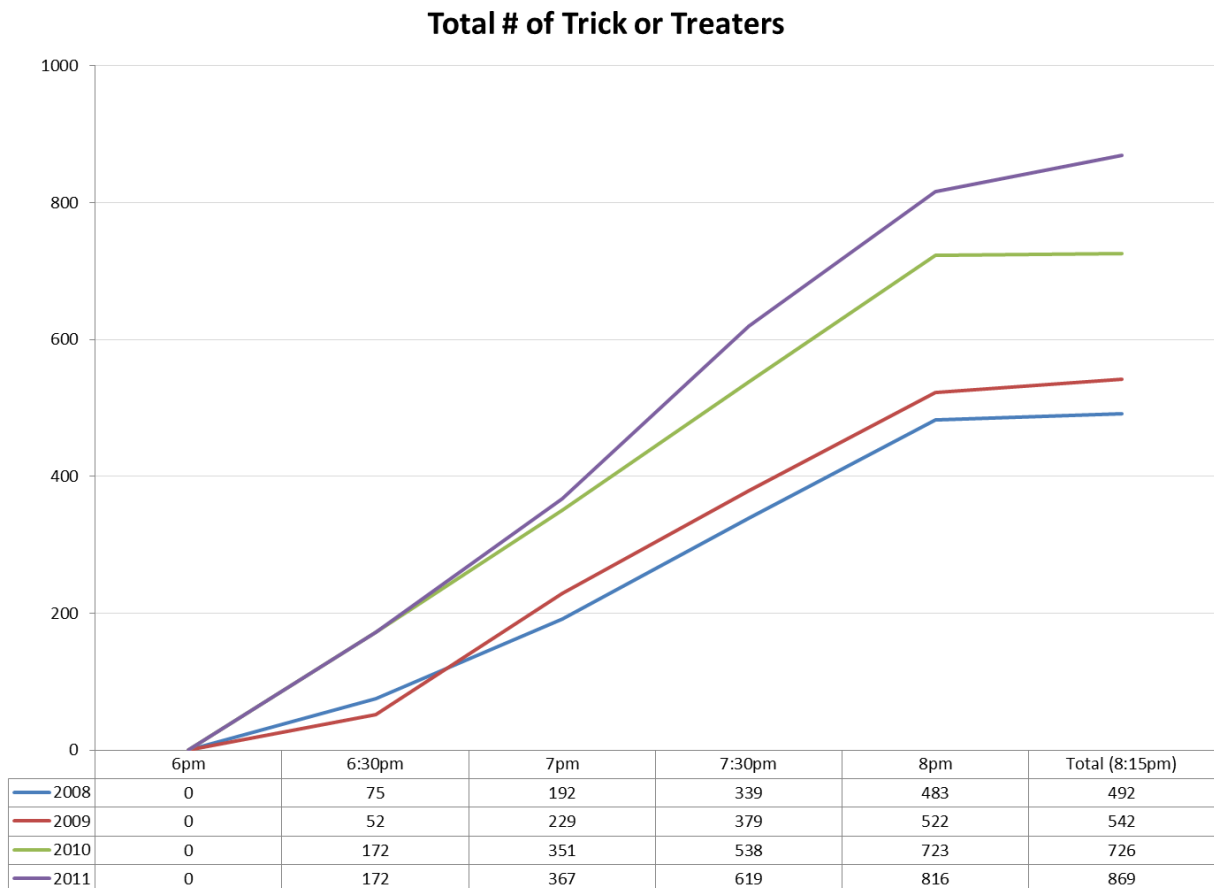
This is easily fixed by simply resizing the chart window. Instead of resizing though, right click on the chart area and select Move Chart.



Then select New Sheet and click OK.



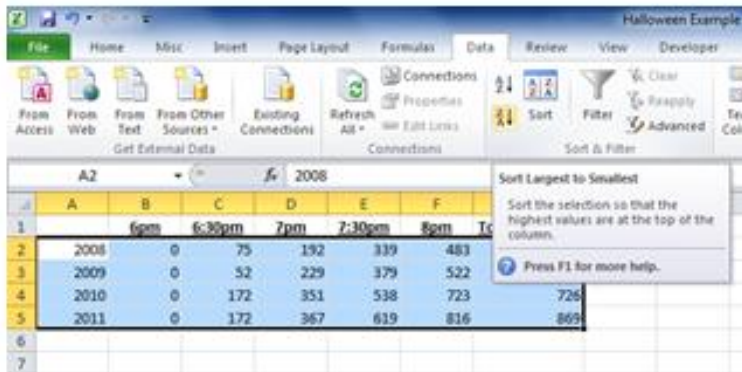
The chart now looks like this.



Before we continue with further formatting, let's take this opportunity to fix the series order in the data table. Typically chronological data in a database table has the oldest date at the top of the table because new data is appended to the bottom of the tables as new records. However, visually it would

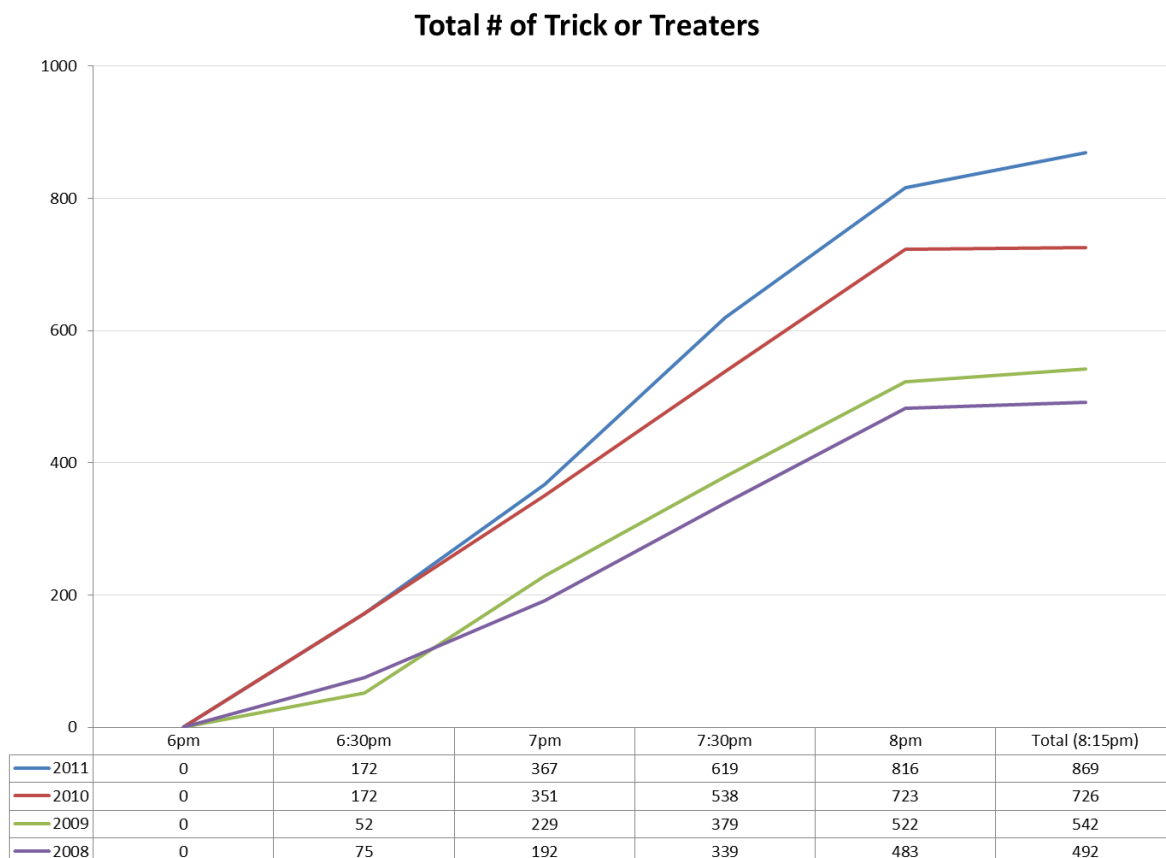
be better to show the most current data at the top. This is also important in this case because the lines happen to trend up in that same order so if the values were flipped it would be much easier to read and more user-friendly. This could also have been done in an earlier step.

Simply highlight the original data set taking everything underneath the headers. Then click the Data tab and sort Z to A.

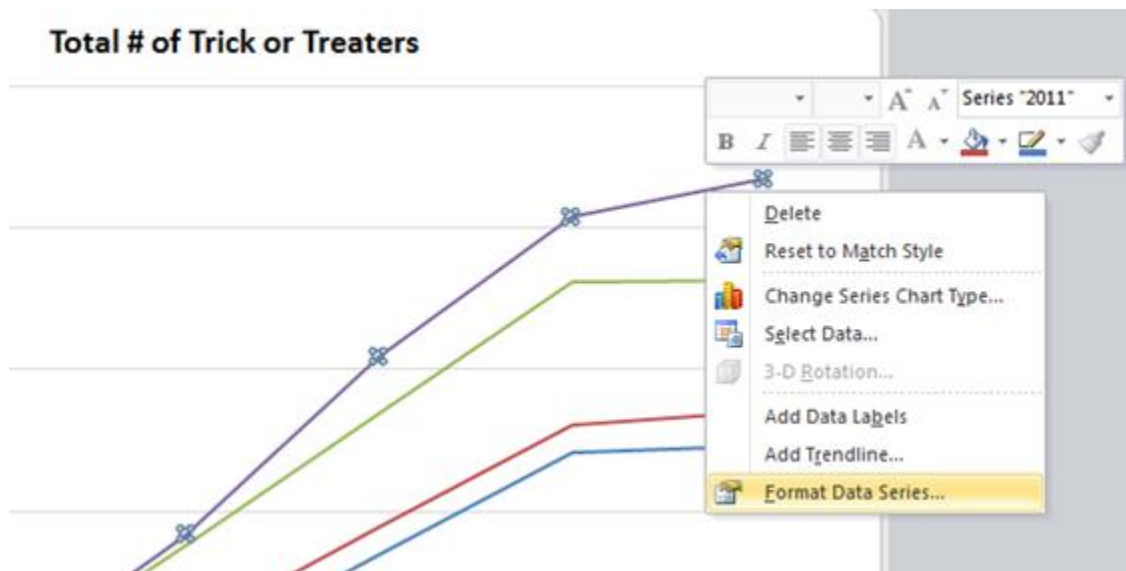


	A	B	C	D	E	F
		6pm	6:30pm	7pm	7:30pm	8pm
2008	0	75	192	339	483	726
2009	0	52	229	379	522	726
2010	0	172	351	538	723	726
2011	0	172	367	619	816	869

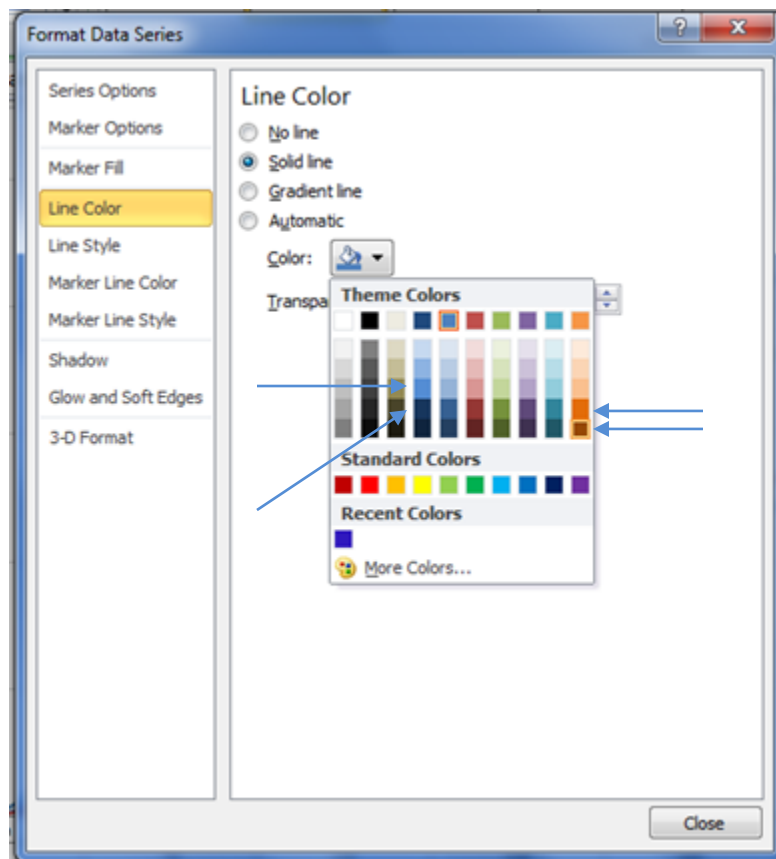
The chart will automatically adjust and look like this.



The next step is to adjust the color of the lines. Click on the top line, right click and select Format Data Series.



Then select Line Color, Solid Line and select the Color dropdown box. Select Dark Orange.



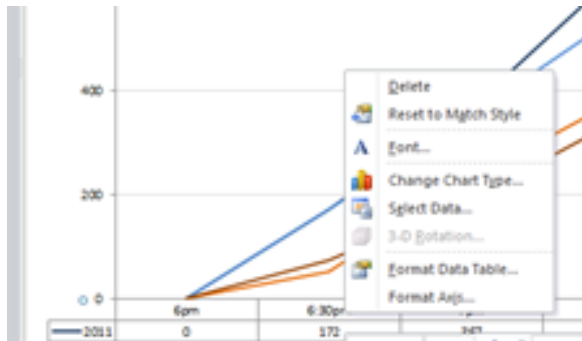
Repeat these two steps for the second line selecting a lighter orange, the third line with a light blue and the lowest line with a darker blue. There are many other options here, for example one color with just light to dark or other alternatives. Try to avoid a red vs. green comparison though.

Note – in Excel 2010 you do not have to close the dialog box to select the next line. You can simply keep the dialog box open and click on the next line in the series to change the format.

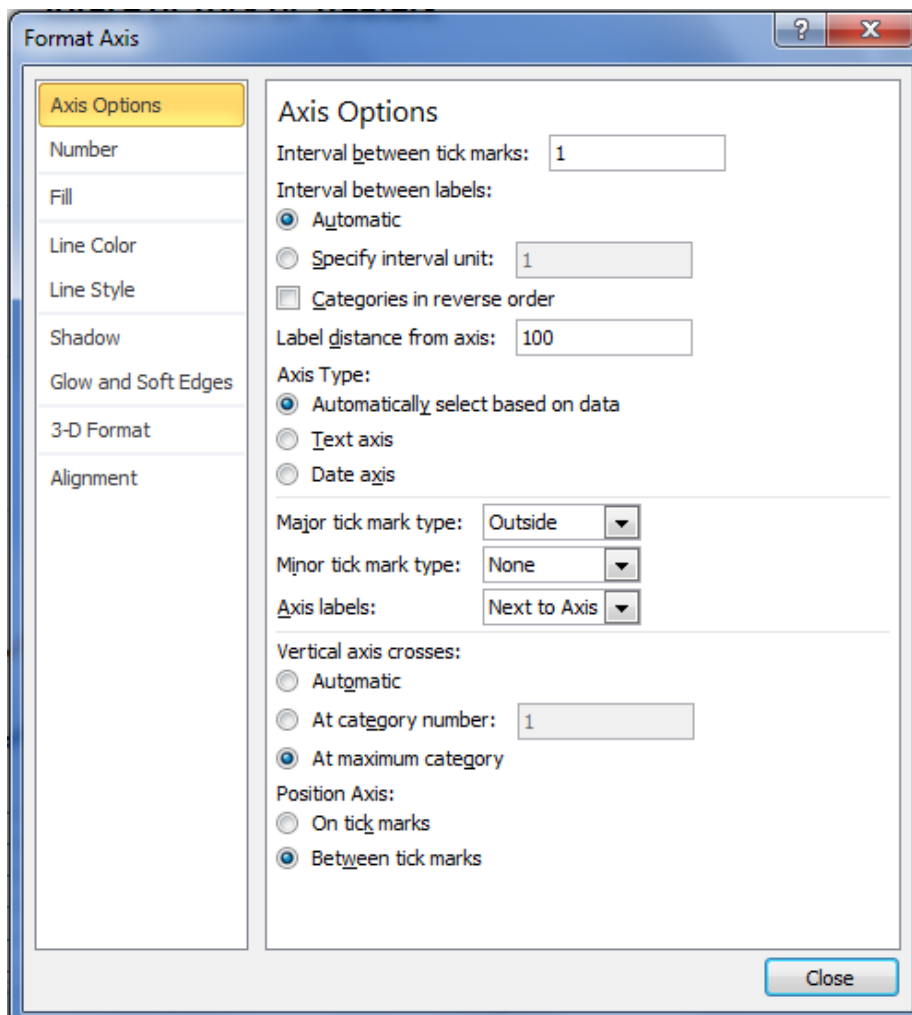


This chart is fairly clean now and would be perfectly acceptable in this form, but there are some additional things that can be added to this chart. The Economist magazine has pioneered the technique of placing the y-axis labels on the right side for time series data. This allows the axis labels to line up with the visual of following the lines up from left to right with the most important points at the end.

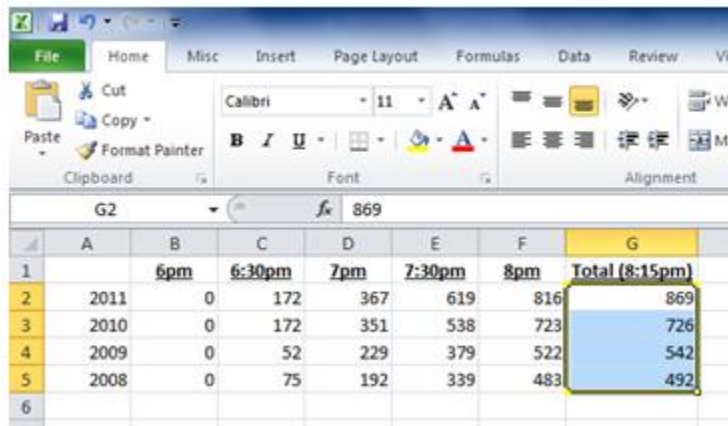
To do this in Excel, right click on the **x-axis** (i.e. the time) and select Format Axis. The easiest thing to do is right-click over top of 6pm or 6:30pm on the x-axis.



Then under “Vertical Axis Crosses:”, select At Maximum Category and click Close.

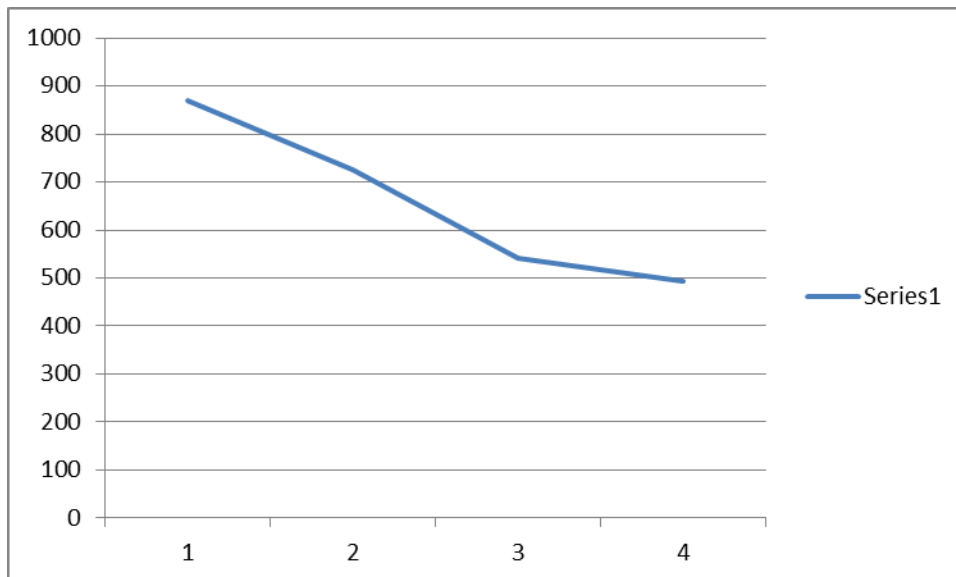


One thing that might add additional context to this chart would be a chart that would show the total number by year. This can be done by inserting a smaller line chart inside of the larger line chart. Back on the data sheet in Excel, select the 4 values in the totals column and click Line and insert another line chart.



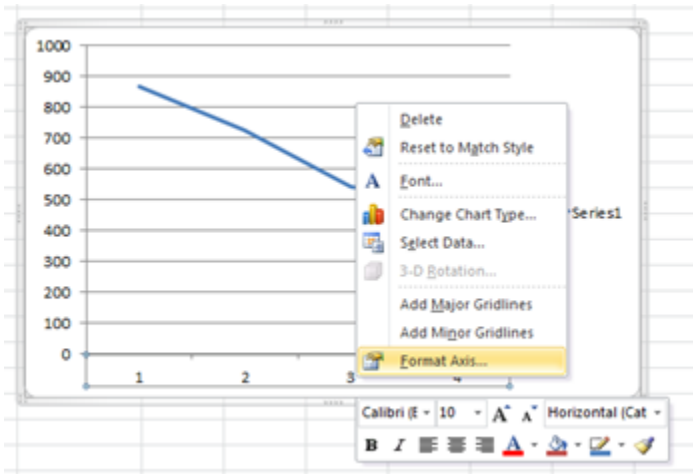
	A	B	C	D	E	F	G
1		6pm	6:30pm	7pm	7:30pm	8pm	Total (8:15pm)
2	2011	0	172	367	619	816	869
3	2010	0	172	351	538	723	726
4	2009	0	52	229	379	522	542
5	2008	0	75	192	339	483	492
6							

The line chart will look like this.



Remember, we re-ordered the data so now it's backwards. We need to reverse the order.

Right click on the x-axis and click format axis.



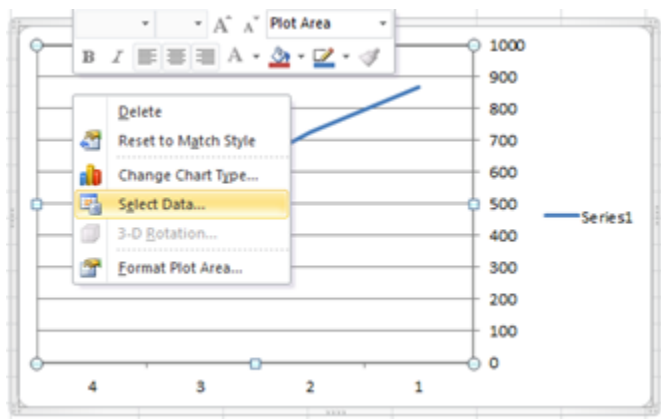
Check Categories in reverse order and click Close.

The 'Format Axis' dialog box is shown. The 'Axis Options' tab is selected. The 'Categories in reverse order' checkbox is checked. The 'Interval between tick marks' is set to 1. The 'Label distance from axis' is set to 100. The 'Axis Type' is set to 'Automatically select based on data'. The 'Major tick mark type' is set to 'Outside'. The 'Minor tick mark type' is set to 'None'. The 'Axis labels' are set to 'Next to Axis'. The 'Vertical axis crosses' is set to 'Automatic'. The 'Position Axis' is set to 'Between tick marks'. A 'Close' button is at the bottom right.

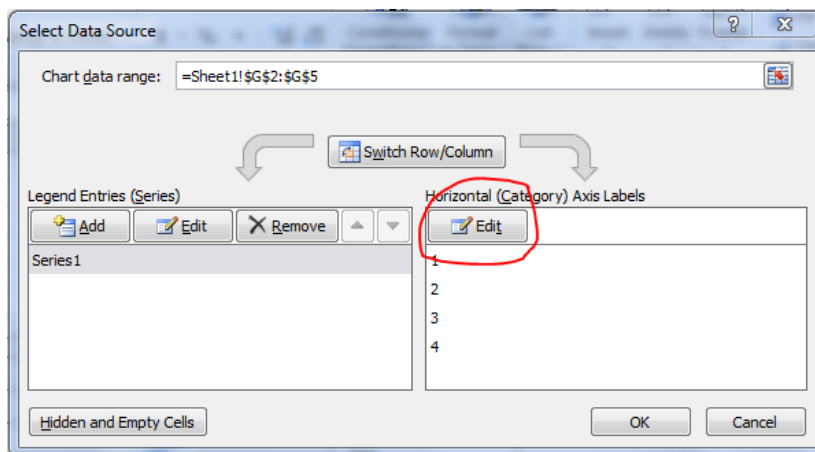
Category	Value
1	850
2	700
3	550



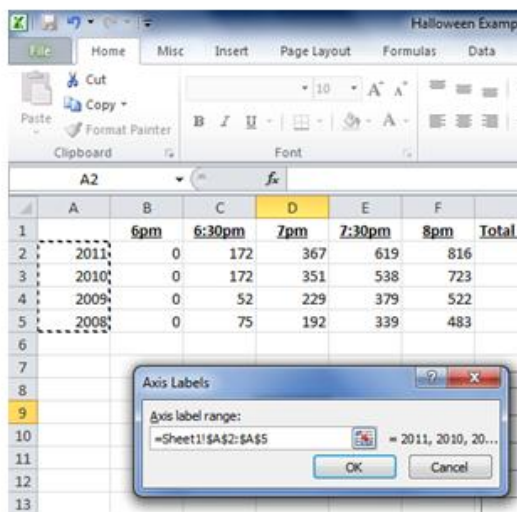
Right-click the chart area and click Select Data.



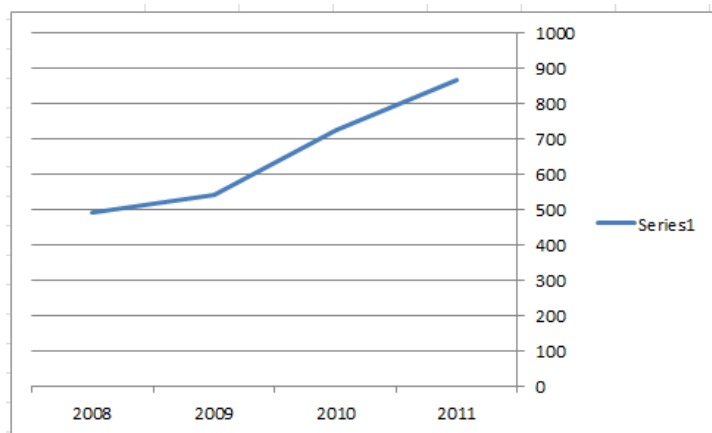
Click Edit on the right hand side window.



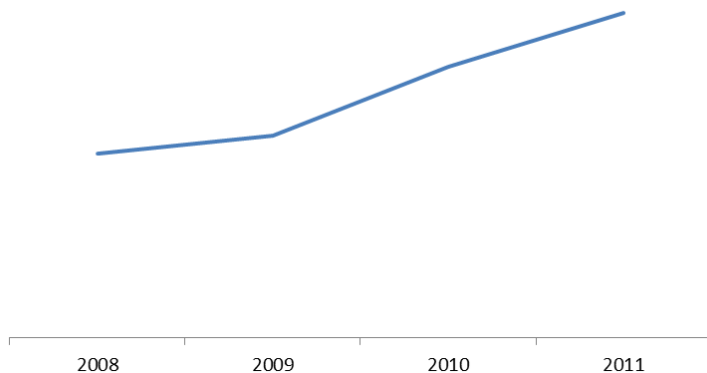
Highlight the years in column A and click OK.



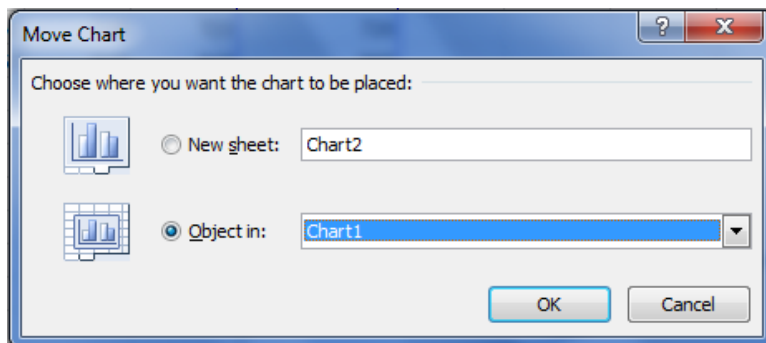
The new chart should look like this now.



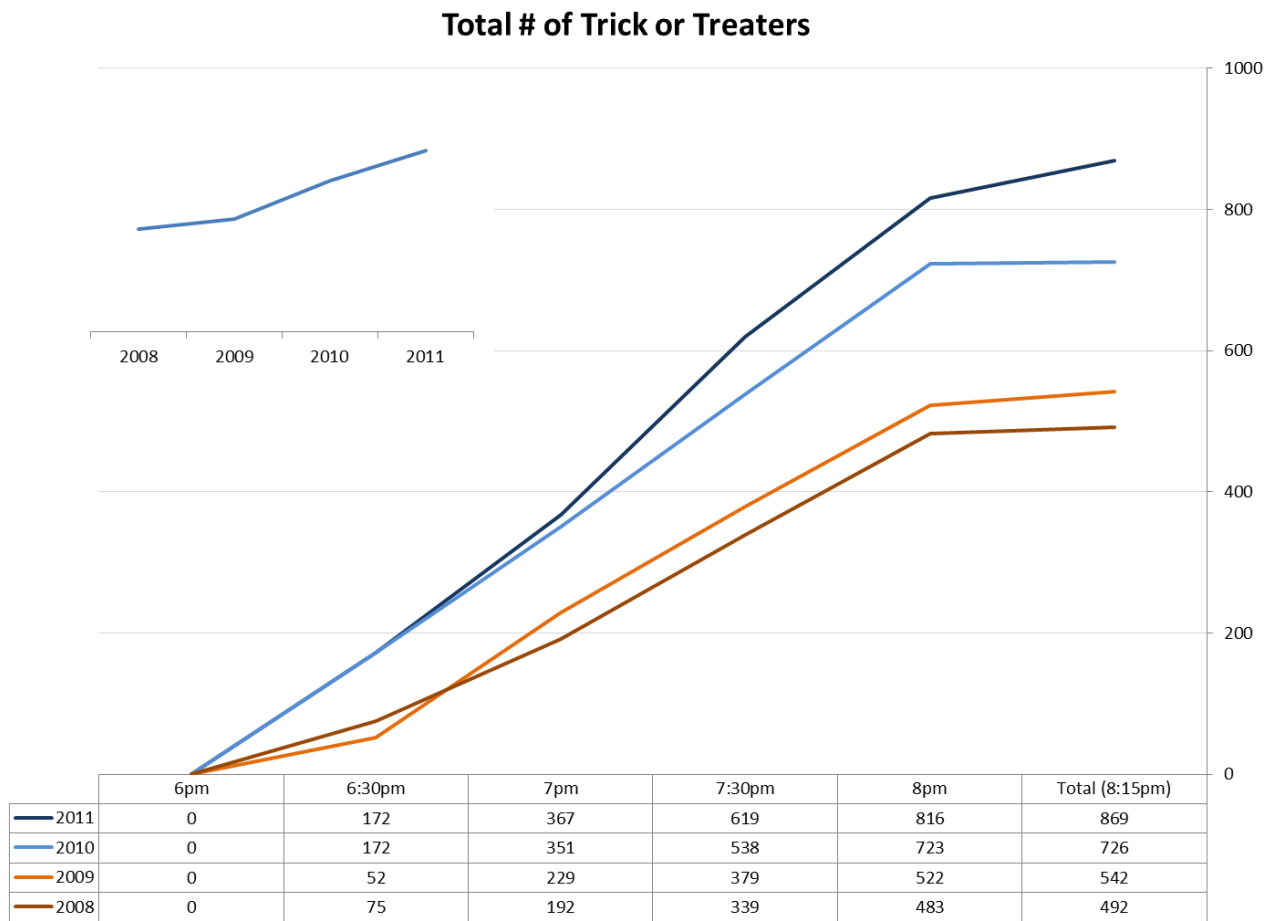
Click on the Legend where Series 1 is located and press the delete key. Then click directly on the gridlines and press the delete key. Click on the y-axis and press the delete key. Finally remove the border around the chart as previously outlined. The chat will now look like this.



Now right click on the chart area and select Move Chart. Select Object in Chart 1 tab (which should be the name of the previous chart built).

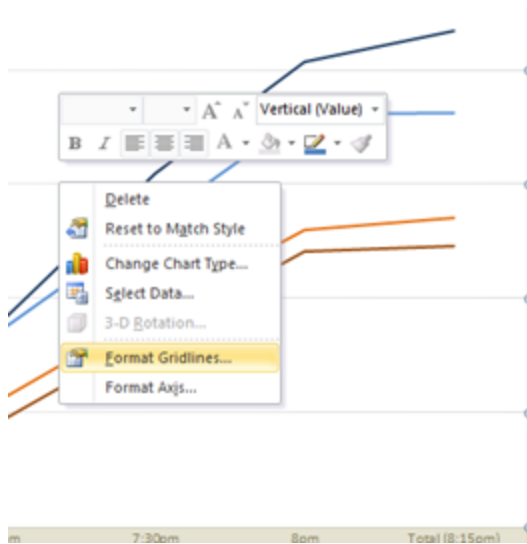


The new chart should now be a smaller chart within the original line chart. Move the smaller chart to the upper left corner of the larger chart and resize the small chart to be a little smaller.

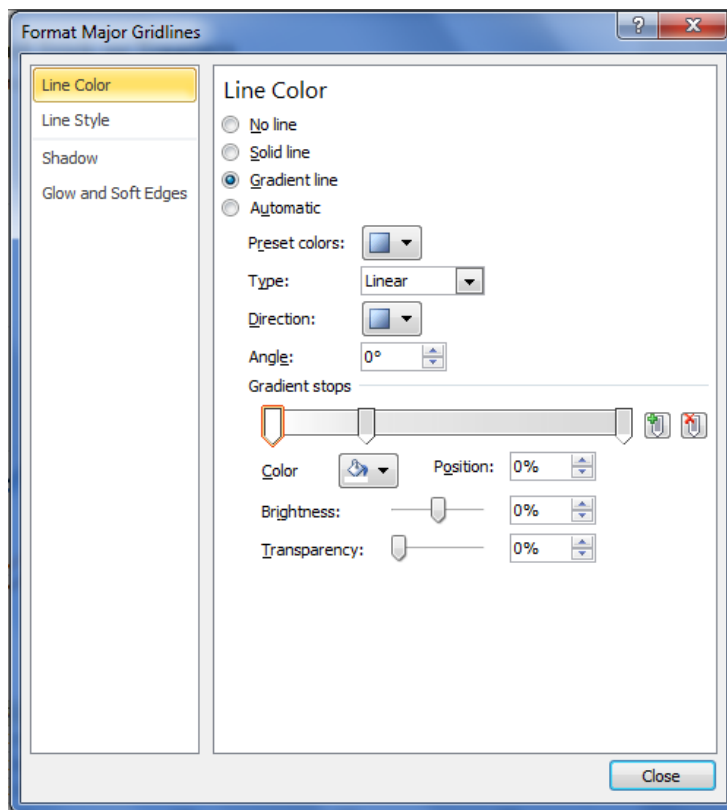


Here are some optional items that will add a few final touches to the chart.

- Right Click on the y-axis and select Format Axis. Change the maximum value on the y-axis to be fixed at 990 so that they top line and change the minimum value on the y-axis to be 0. This will eliminate the top gridline and the label for 1000.
- Right click on the small line chart and select Format Chart Area. Select solid fill and choose the gray that is second from the top. This will put a light border around the small chart.
- Select gridlines, right click and select format gridlines, select Gradient Line.

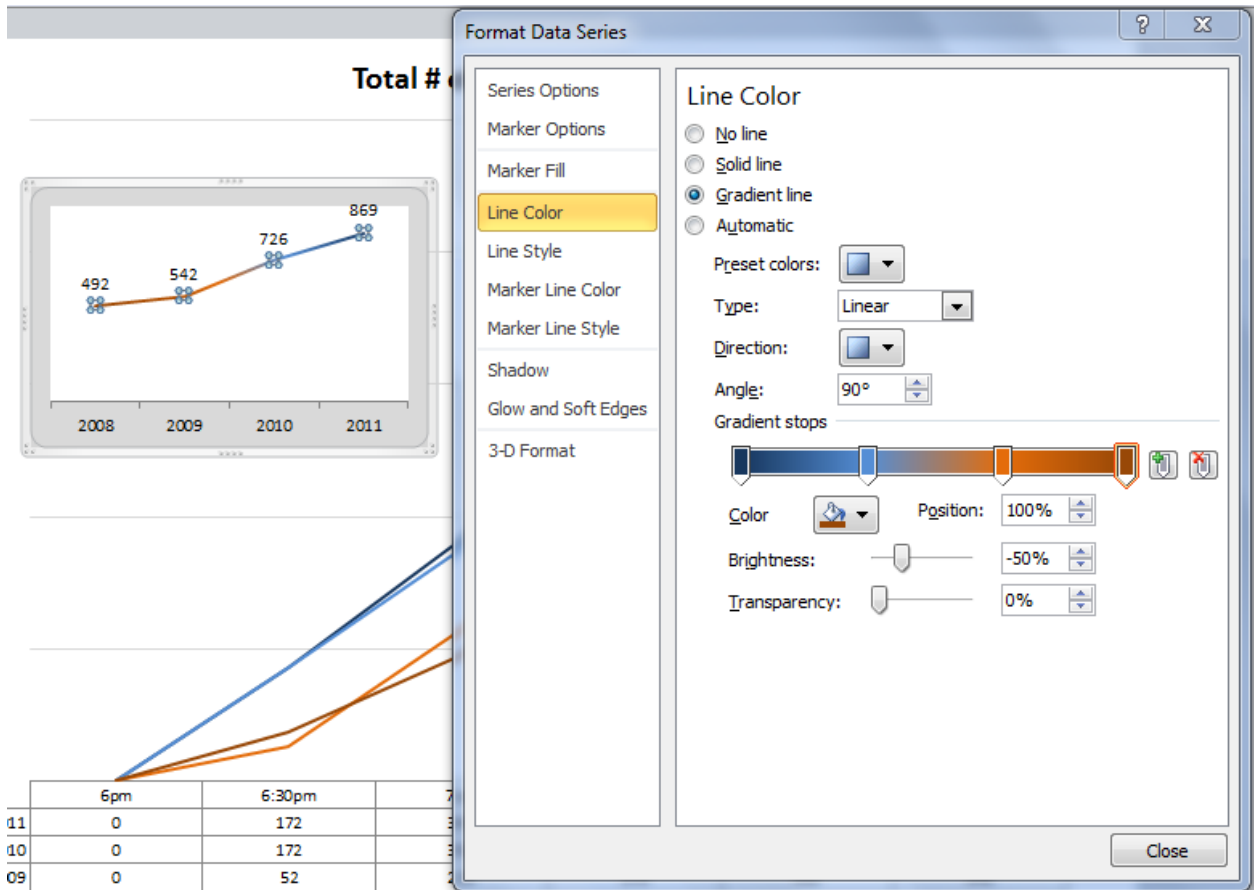


- Change direction to go light on the left to dark on the right. Change the color on the left to white and change the color in the middle to the lightest gray and change the color on the right to be the second gray. Now the gridlines will be blank on the left side of the graph where they aren't needed and show lightly on the right side of the graph where they are useful.



- Right click on the line in the small chart and click Add Data Labels. Right click on the data labels and select Format Data Labels. Change label position to Above.

- Select y-Axis and data table and change the line color to a lighter gray to match the gridlines.
- Select the line in the small graph, right click and select Format Data Series. Select line color and then select Gradient line. Select direction from left to right. Select left color as dark blue. Move stop over and set to light blue, add a third stop for light orange and set the last stop to dark orange (remember it's in reverse order).



Here is the final version. I added a transparent title bar in light gray over top of the times to outline the headings from the data table. This has to be done manually, under the Insert tab and select Insert Shapes. I created a rectangle and made it light gray to match the smaller line chart. If you use this method to add final touches make sure you do this at the end, because this is manual and if you make other changes to the chart then this box might need to be updated as well.

We now have a useful line chart, a mini chart with the totals built-in and a data table with our data in a user-friendly manner. A few small things give the chart a little extra finish.

