Week 5: Data Visualization

Table of Contents

[Opening Story 1](#_Toc464154733)

[Reading 2](#_Toc464154734)

[Example: Alzheimer's Research Study 2](#_Toc464154735)

[Common Graphical Tools 4](#_Toc464154736)

[Bar Charts 4](#_Toc464154737)

[Pie Charts 8](#_Toc464154738)

[Time Series Plots 12](#_Toc464154739)

[Scatter Plots 15](#_Toc464154740)

## Opening Story

Embed Video Here

A college student is sitting in a small (5 to 10 people) business meeting as the new intern for that company. After some discussion of the meeting topic the boss gives the intern the assignment of collecting all the data that was discussed at the meeting and putting it together into a report. The report is to be shared with the group at the next meeting in a day or two.

We watch the student back at their office cubicle struggle with how to summarize all of the stacks of data (perhaps on lots of different papers). It is clearly demonstrated that the student is spending a lot of time and energy and getting no where. They feel completely overwhelmed with the task with no idea of what to do. Eventually they go to a nearby co-worker to ask for help. The co-worker is a recently graduated student that just got a full-time job with the company. (Somehow this person looks like a person that everyone would want to be when they are graduated and have a job.) The intern explains their task and inability to do what was asked. The co-worker says something that demonstrates that they know how to help the intern solve their problem using Excel and that they learned how to solve the problem in their college math class.

The video shouldn't demonstrate how to solve the problem, just that the problem is now going to be solved. In other words, that the intern will know what to do after finishing the conversation with the co-worker.

Students in our Math 108 course will be responsible to create the report requested of the intern in the video by the end of this week's lesson.

## Reading

Humans have an amazing ability to identify patterns in images. A classic example is demonstrated in the following picture. Despite the camouflaging stripes of the tigers, the human eye quickly locates the two tigers crouching in the grass.



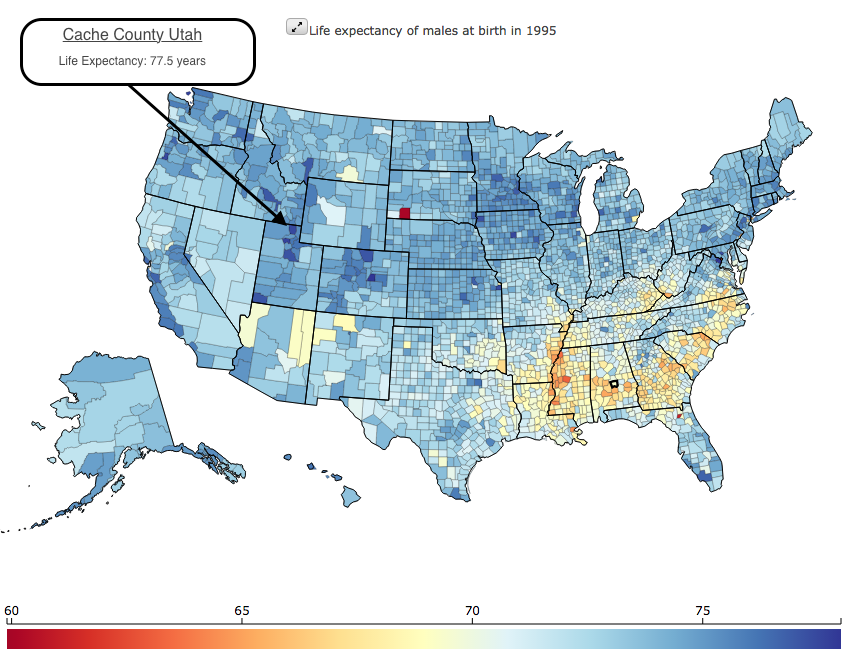
*Data visualization* is the science of communicating important patterns from data through the use of graphics. Figuratively speaking, the goal of a good graphic is to help people quickly identify the "tiger in the grass." In other words, the goal of a good graphic is to reveal important patterns in data.

The insight a graphic provides should enable a more powerful decision on the part of the user. To do this, graphics must communicate the truth data has to offer on a situation of interest.

### Example: Alzheimer's Research Study

Alzheimer's is a serious disease resulting in memory loss that most typically occurs in adults aged 65 and older. Individuals affected with the disease progressively lose the ability to function with normal daily tasks until they eventually die from the disease. There is currently no known cure for Alzheimer's, but much research is being done to try to better understand the disease and how to prevent it.

One of the largest studies of Alzheimer's was undertaken in 1995 in the USA. It is known as the study on [memory health and aging](http://www.usu.edu/epicenter/htm/studies/memorystudy) is still going on today. Before the study officially started, one of the first questions the researchers were faced with was where to conduct the study. They needed a population that was old enough and healthy enough to observe possible factors that contribute to the development of Alzheimer's disease. In making this decision they would have used a graphic similar to the one shown here. Notice how the graphic readily communicates which counties have the longest life expectancy by showing them in dark blue.

[](http://vizhub.healthdata.org/us-health-map/)

**Figure 1**: The counties shaded in dark blue show those counties in the USA with the longest life expectancies for males. There are several counties showing dark blue. A detailed examination of each of the counties shows Cache County Utah having one of the longest life expectancies in the USA at 77.5 years (for males).

The researchers eventually selected Cache County Utah because, as they stated, "The elderly of Cache County have a longer life expectancy, higher educational attainment, and lower incidence of chronic disease (which can complicate the diagnosis of dementias) than other similar populations."

## Common Graphical Tools

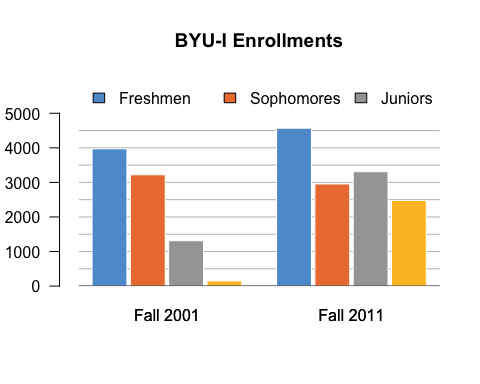
While only the imagination limits what a good graphic can be, there are several types of graphics that are commonly used to communicate information. Learning how and when to use these common graphical tools will benefit you both as a student and as an employee.

### Bar Charts

When data has several categories, the number of observations in each category can be visually summarized with a bar chart.

#### Example: Enrollment of BYU-I Students

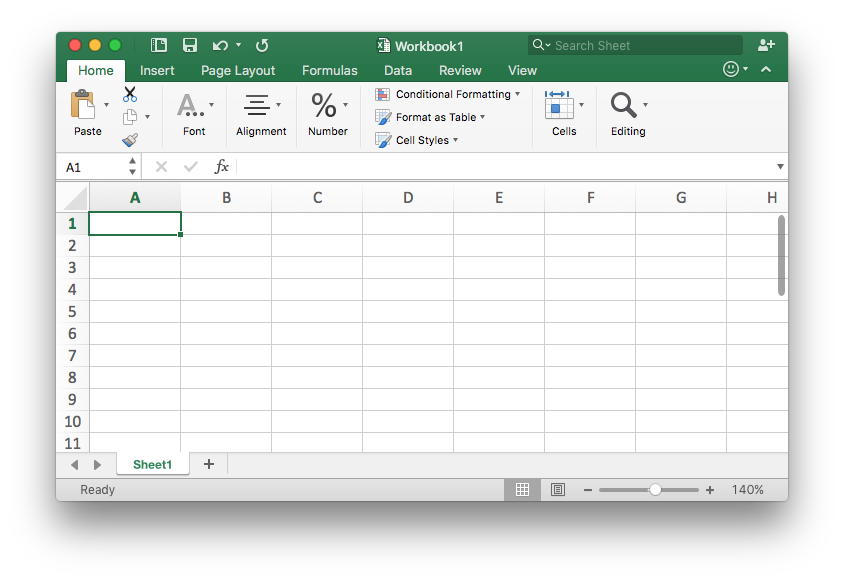
In the Fall of 2001, Ricks College offically transitioned to become BYU-Idaho. This transition changed the two-year junior college into a four-year university. Understandably there were relatively few Juniors and almost no Seniors during the first semester of the transition. Ten years later however, the the number of Juniors and Seniors had stabalized to be roughly equal with the Sophomores. Interestingly, the number of incoming Freshman was still much greater than any of the Sophomore, Junior, or Senior classes. (The data for this chart are publicy available data from the [BYU-Idaho Enrollment Statistics](http://www2.byui.edu/IR/stats/index.htm) website)



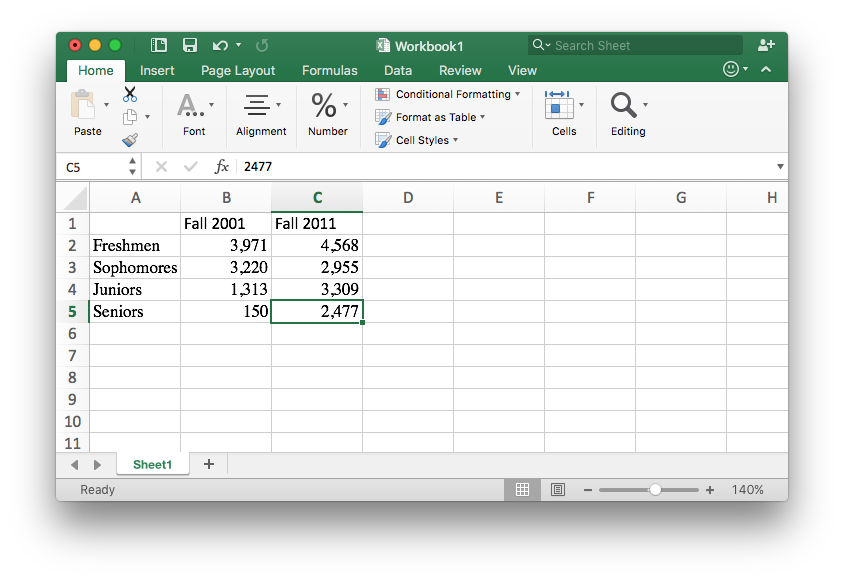
#### Excel Instructions

To make the above bar chart:

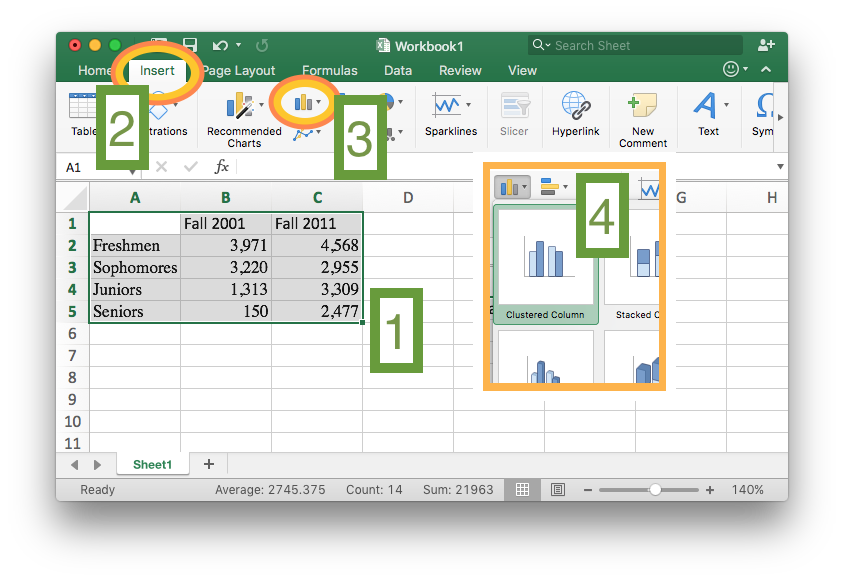
1. Open a "New Workbook" in Excel.



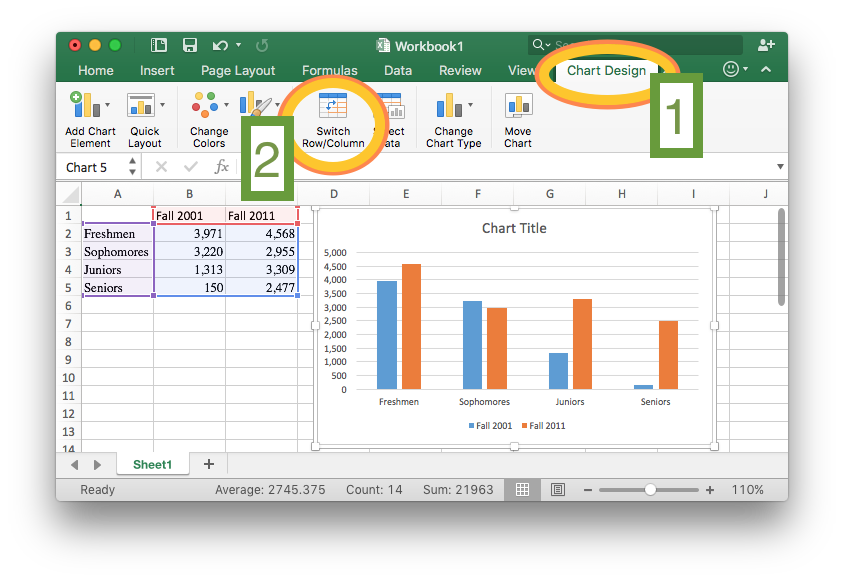
1. Enter the data as found on the [BYU-Idaho Enrollment Statistics](http://www2.byui.edu/IR/stats/index.htm) website.



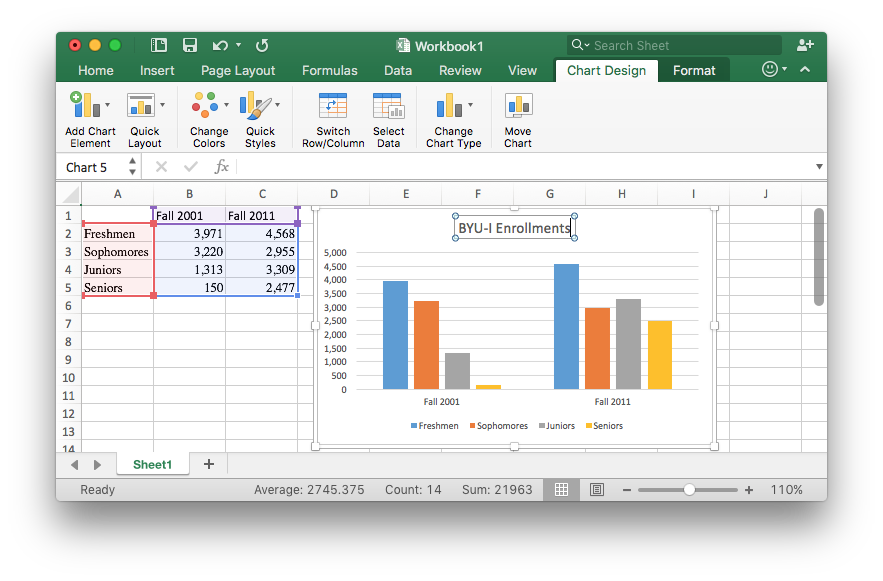
1. [1] Highlight the data table you created.
   * [2] Select the "Insert" tab in the top menu.
   * [3] Select the "Barplot" icon.
   * [4] Select the "Clustered Column" icon.



1. Select the "Chart Design" tab in the top menu.
   * Select the "Switch Row/Column" icon.



1. Give the chart a title.

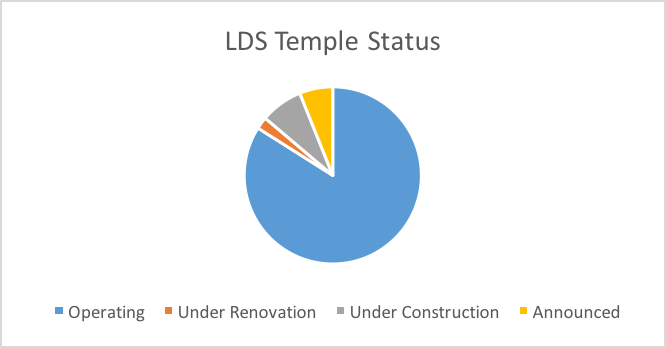


### Pie Charts

The pie chart is a sometimes useful alternative to the bar chart. Like the bar chart is is for displaying data that has several categories. It displays the percentage of the whole "pie", or dataset, that belongs to each "slice", or category, in the data.

#### Example: Current State of LDS Temples

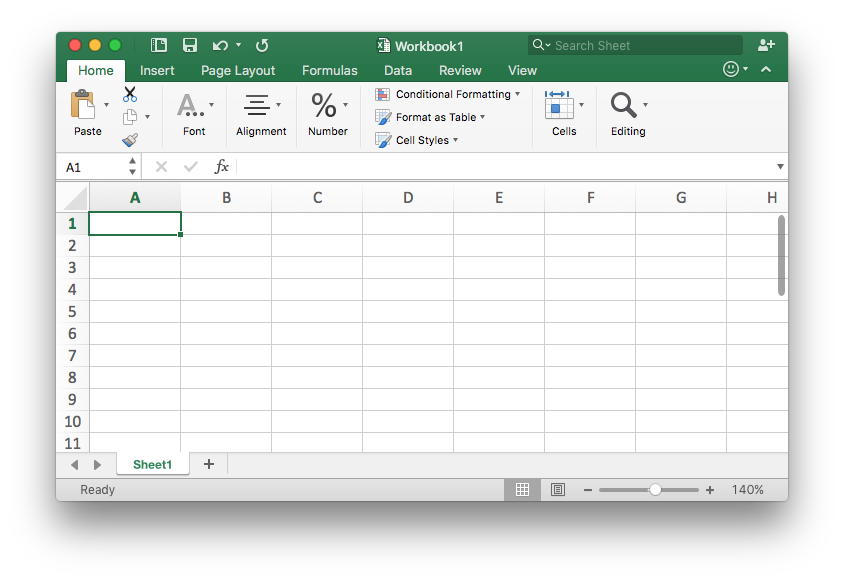
The Church of Jesus Christ of Latter-day Saints places great emphasis on the building of Temples which each bear the engraved words, "Holiness to the Lord: The House of the Lord." According to a recent [official news release](https://www.lds.org/church/temples/find-a-temple?lang=eng&sort=name&sort-direction=up) of the LDS Church, there are currently 181 LDS Temples. The article states that there are "152 operating, 4 under renovation, 14 under construction, 11 announced." The following pie chart depicts this information.



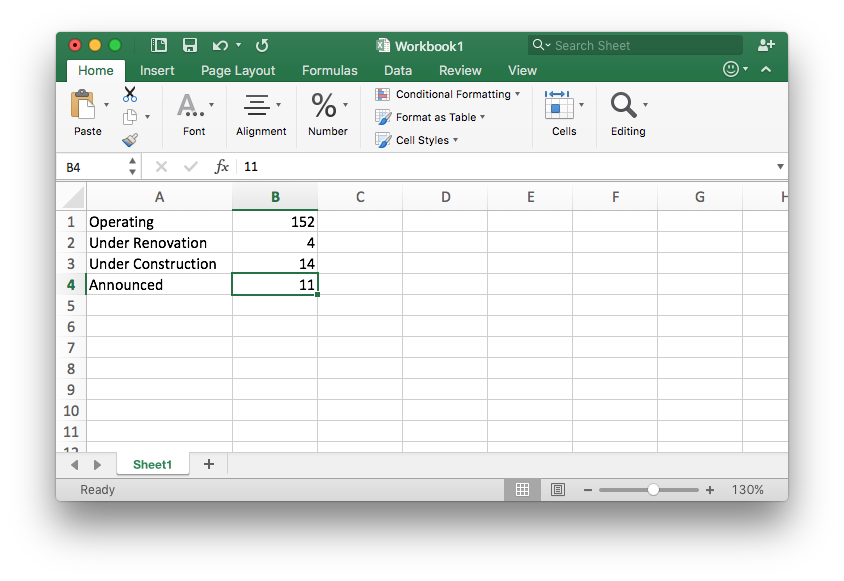
#### Excel Instructions

To make the above pie chart:

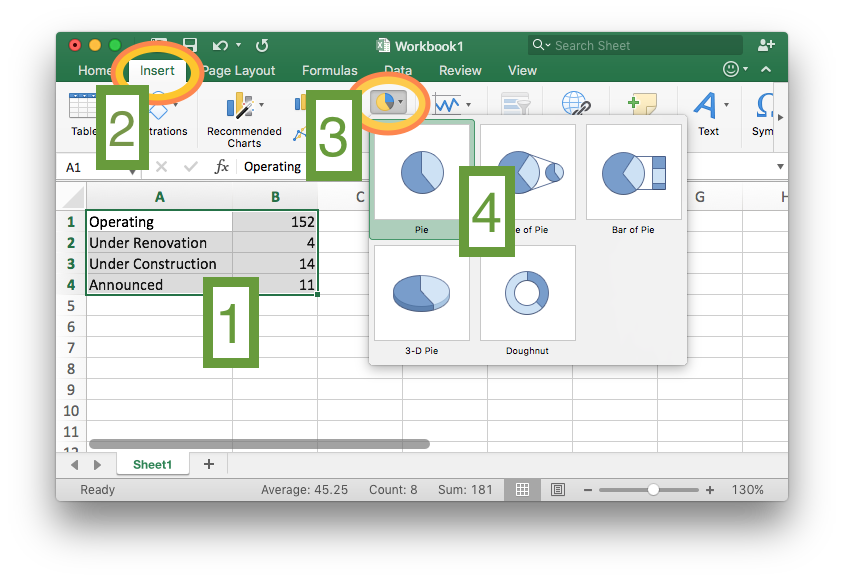
1. Open a "New Workbook" in Excel.



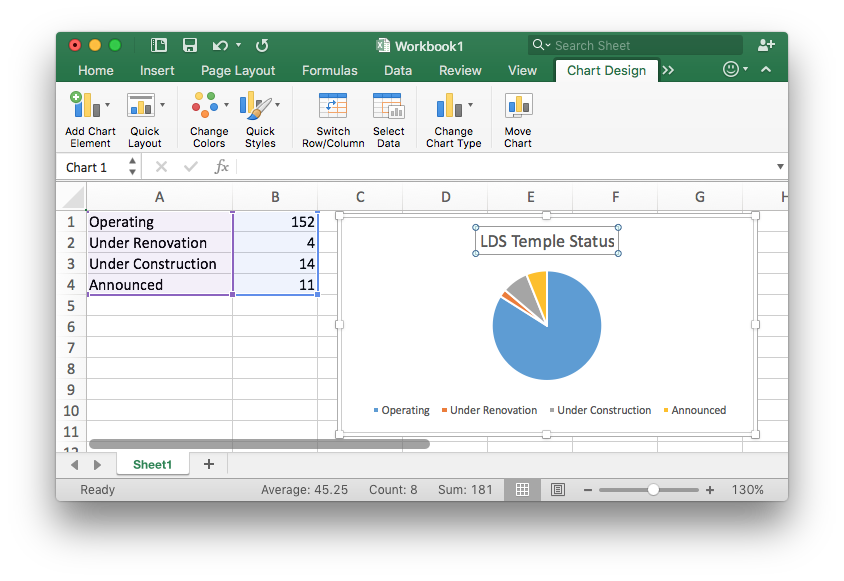
1. Enter the data as found on the [LDS Church news release](https://www.lds.org/church/temples/find-a-temple?lang=eng&sort=name&sort-direction=up).



1. [1] Highlight the data table you created.
   * [2] Select the "Insert" tab in the top menu.
   * [3] Select the "Piechart" icon.
   * [4] Select the "Pie" icon.



1. Give the chart a title.

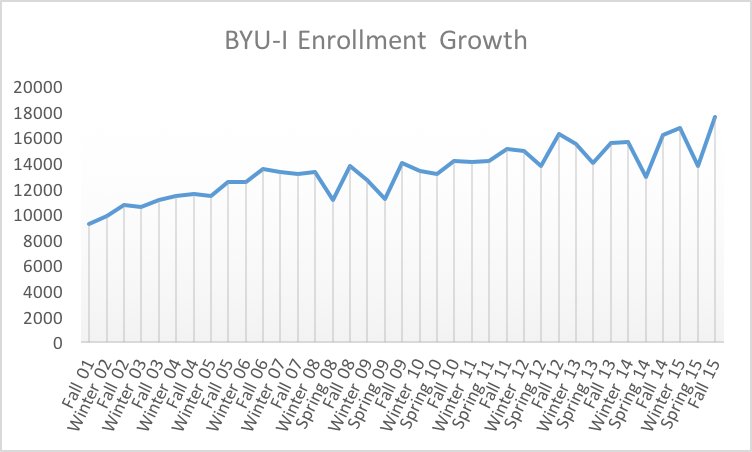


### Time Series Plots

We are often interested in looking at patterns and trends in data over time. This is usually done with time series plots. These plots reveal important information about growth, decay, or repeating (cyclical) patterns.

#### Example: Growth of BYU-Idaho Enrollments

The following graphic depicts the first 15 years of growth of the BYU-Idaho enrollments since BYU-Idaho officially began in the Fall semester of 2001. The count represents the total of part-time and full-time students that were living on campus during that semester. The data was taken from the [BYU-Idaho Housing](http://www.byui.edu/housing/owners) website, under the "Historical" link.

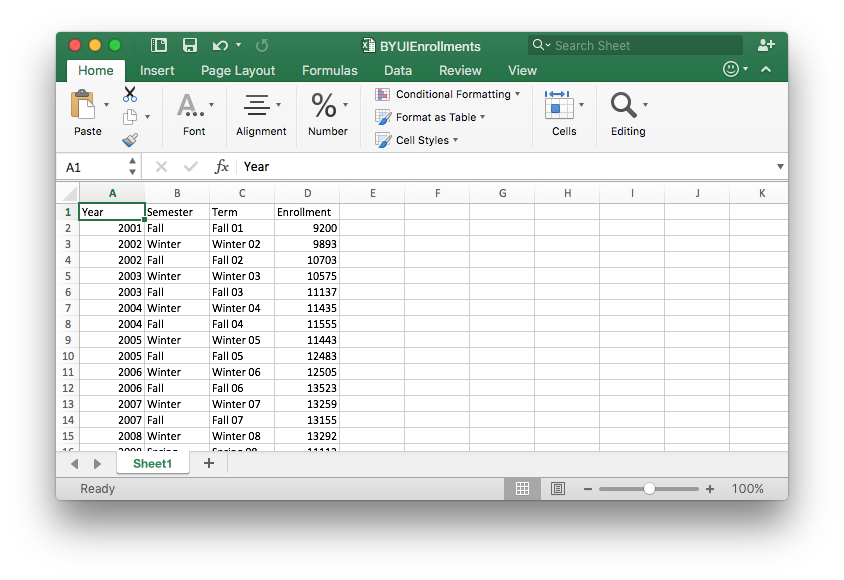


Notice in the graphic that the Spring semester is not officially part of the data until 2008. Also, note the cyclical pattern within each year that typically the Spring semester has the lowest enrollments of a given year. The overall growth of the university is also visible in the graph showing "a steady, upward course" [[1]](http://www2.byui.edu/Presentations/Transcripts/Devotionals/2001_09_18_Eyring.htm).

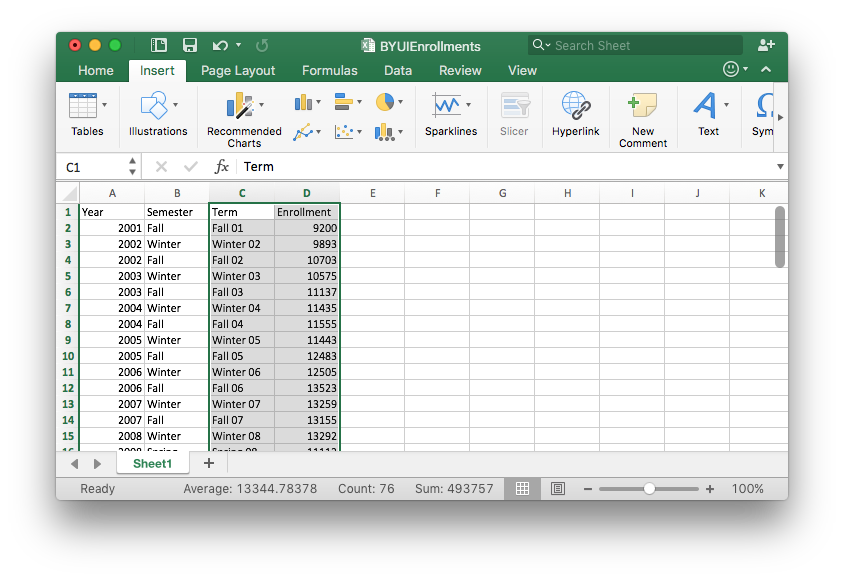
#### Excel Instructions

To make the above time series plot:

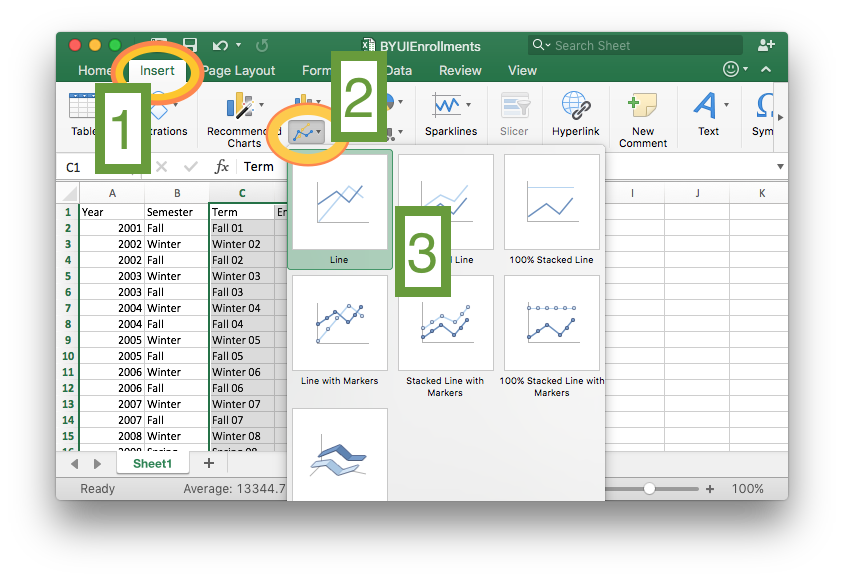
1. Download the [BYUIEnrollments.xlsx](https://content.byui.edu/items/8b0b2956-4527-45e5-bf42-c6b428c8ef67/1/) dataset and open it in Excel.



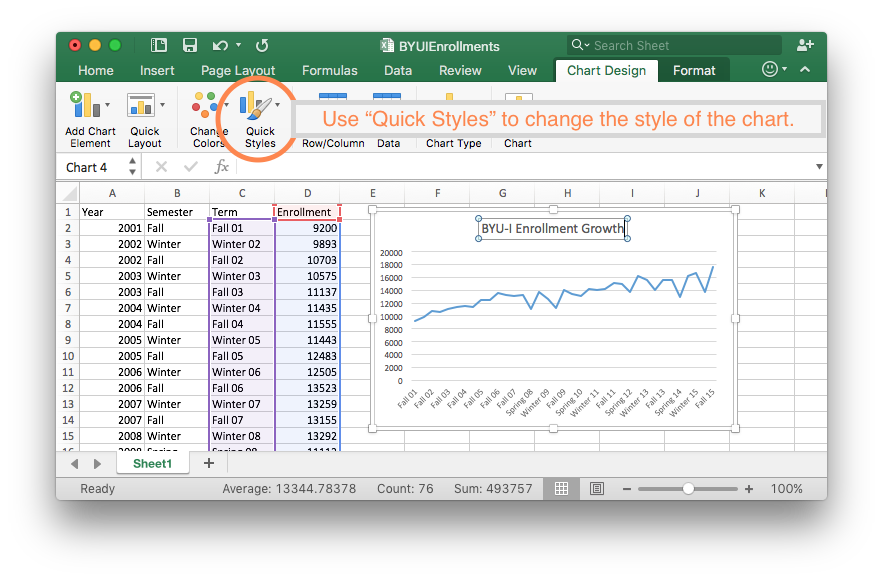
1. Highlight the Term and Enrollment columns of the dataset. (Be sure to highlight all the way to the bottom of the data that goes to row 38.)



1. Then,
   * [1] Select the "Insert" tab in the top menu.
   * [2] Select the "Line Chart" icon.
   * [3] Select the "Line" icon.



1. Edit the chart title and change the styling of the chart if you want.



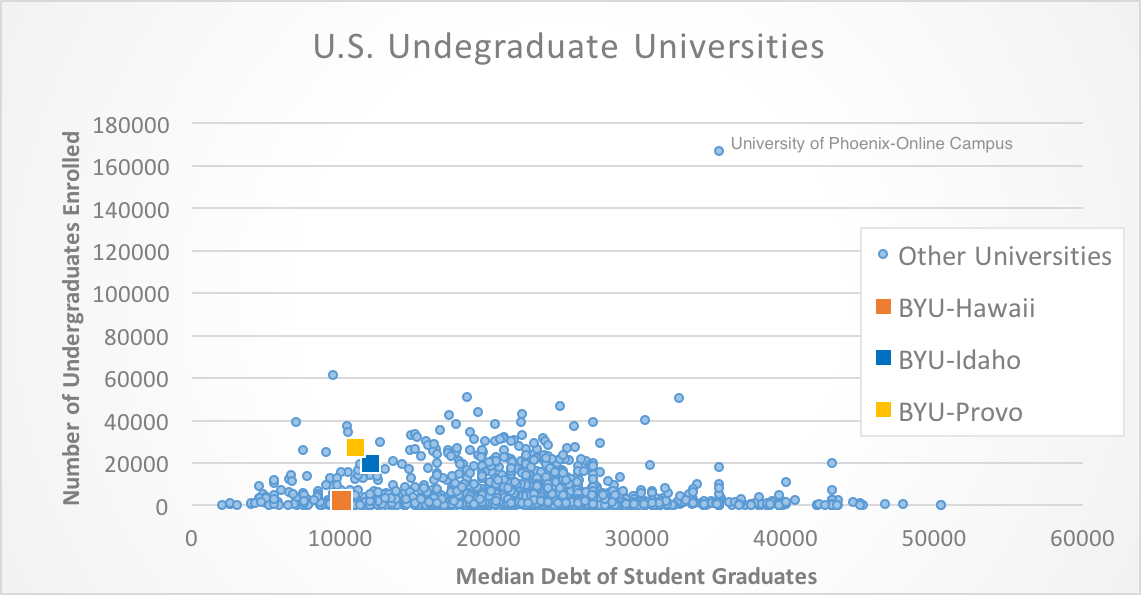
### Scatter Plots

The scatter plot is a more general version of the time series plot. The scatter plot does not require time to be part of the plot. It provides a visual understanding of how two different measurements are related.

#### Example: Undergraduate Debt

The US Department of Education has put together a useful resource called the [College Scorecard](https://collegescorecard.ed.gov/) for college bound students to make important decisions about where to attend college. An important consideration for college students is the amount of student loan debt they may take on during college. Do students that attend bigger universities end up with more debt? The answer is no.

The graphic below uses data from the College Scorecard website to show that there is not any correlation between the size of the university and the median debt load of its graduates. It is interesting to note that all three Brigham Young University schools are on the lower end of the median student loan debt spectrum. However, BYU-Idaho, has the highest median student loan debt of the three while being in the middle for the number of undergraduates enrolled.

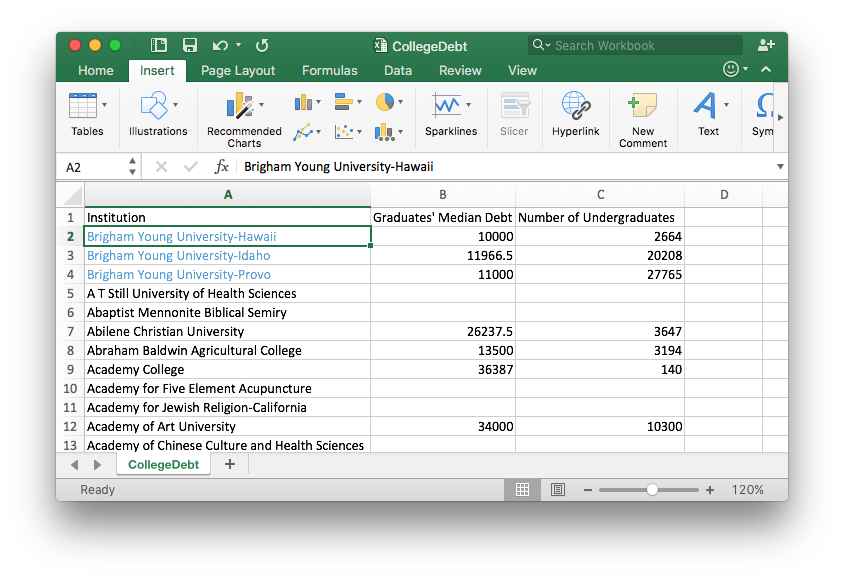


Clearly there are other factors that are not considered here that would provide better explanations of what brings on student debt.

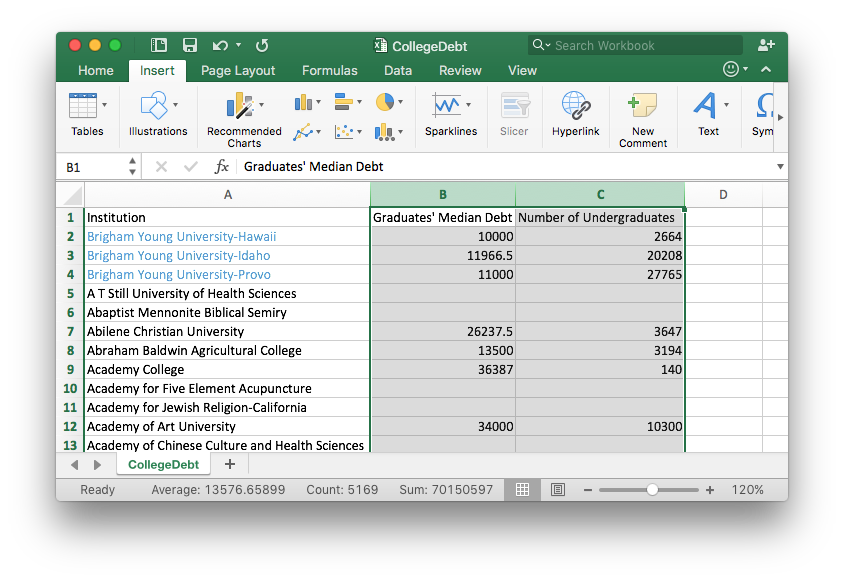
#### Excel Intructions

To make the above scatter plot:

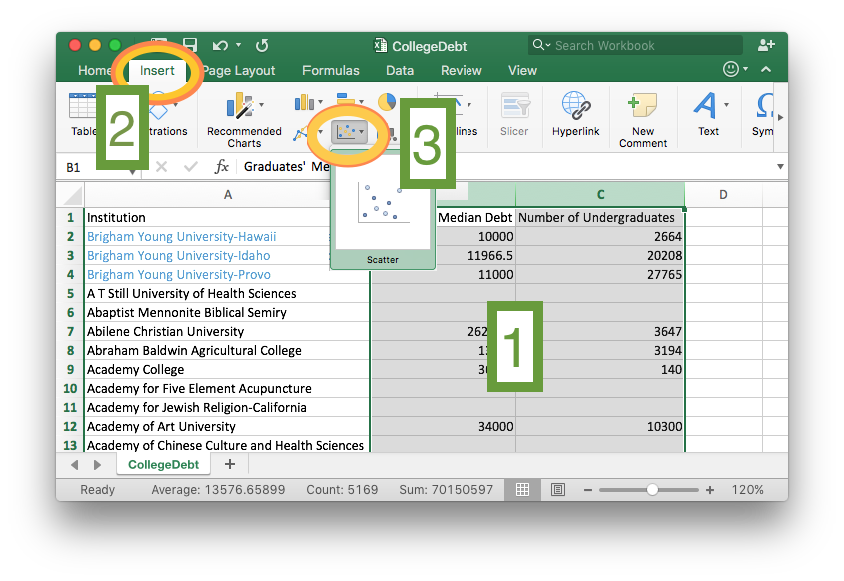
1. Download the **Error! Hyperlink reference not valid.** dataset and open it in Excel.



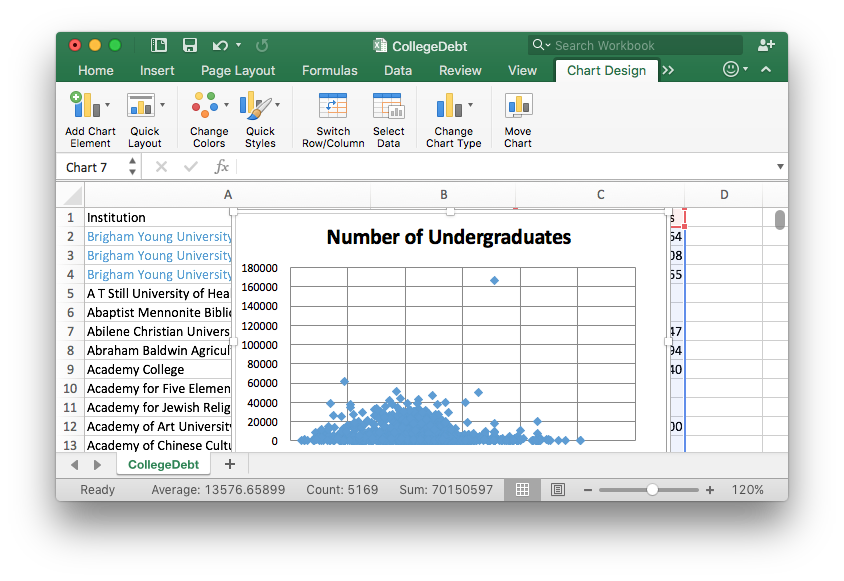
1. Since there are 2,953 rows in this dataset, highlighting the data is done as follows.
   * First, click on the B column label, which is above the Graduates' Median Debt entry. This will highlight the entire B column.
   * Second, hold down Shift and click on the C column label, which is above the Number of Undergraduates entry. This will expand the selection to include the C column.



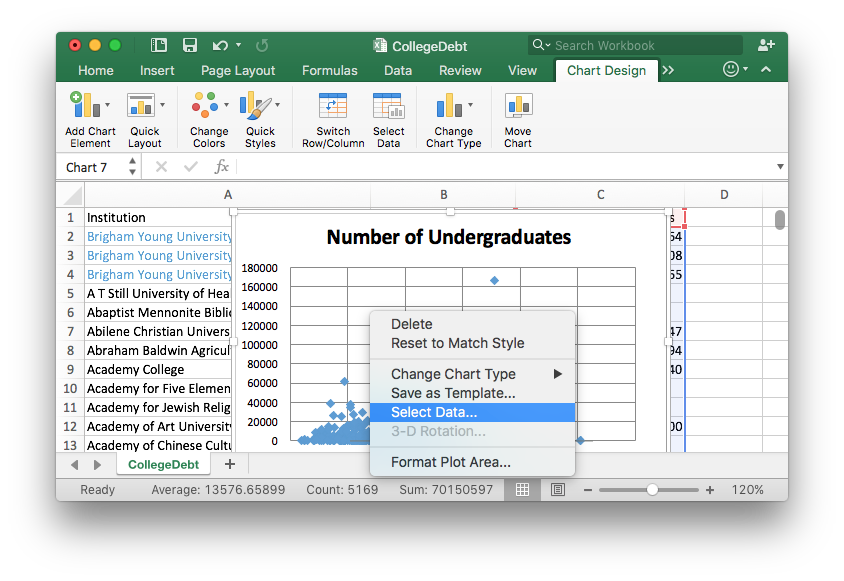
1. Once the data is selected [1], then
   * [2] Select "Insert".
   * [3] Select the "Scatterplot" icon and the "Scatter" option.



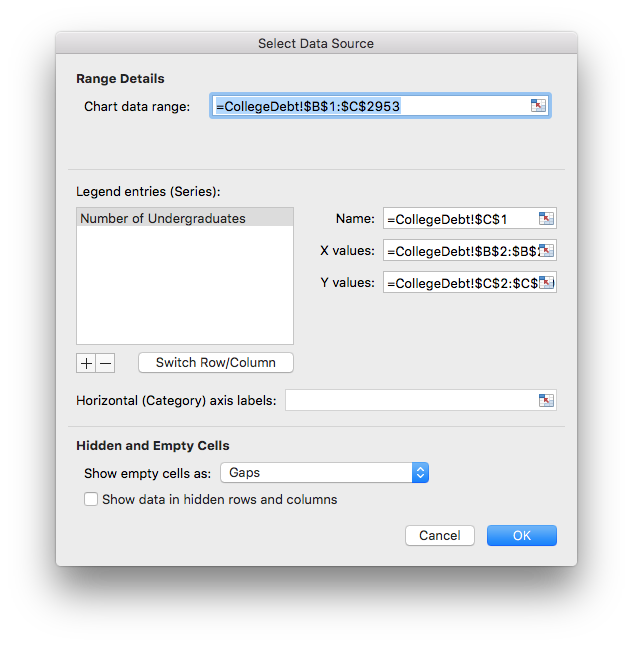
1. This is a scatterplot and it would be sufficient to stop here.



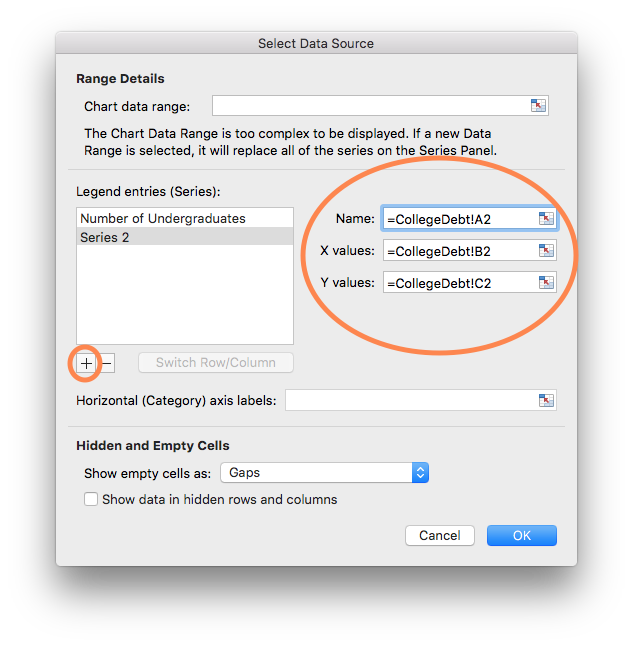
5A. **[Bonus Step]** To add the BYU colleges to the plot, right click on the plot and choose "Select Data..." from the menu.



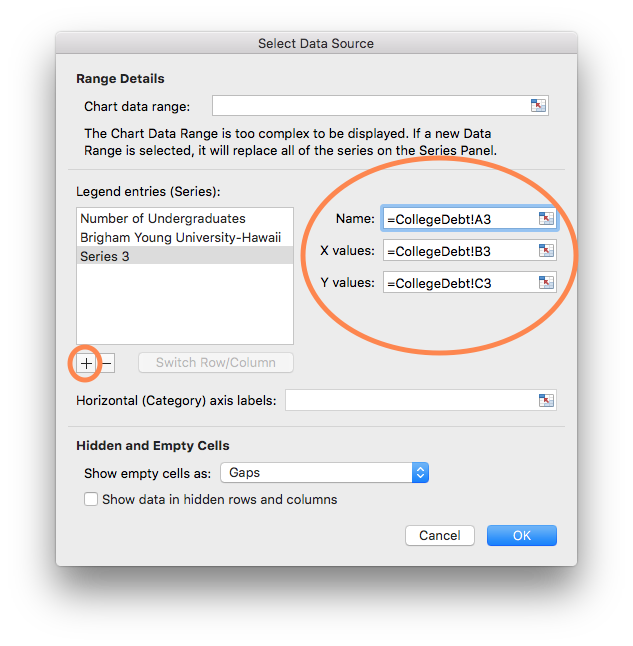
5B. **[Bonuse Step]** A new window will appear entitled "Select Data Source."



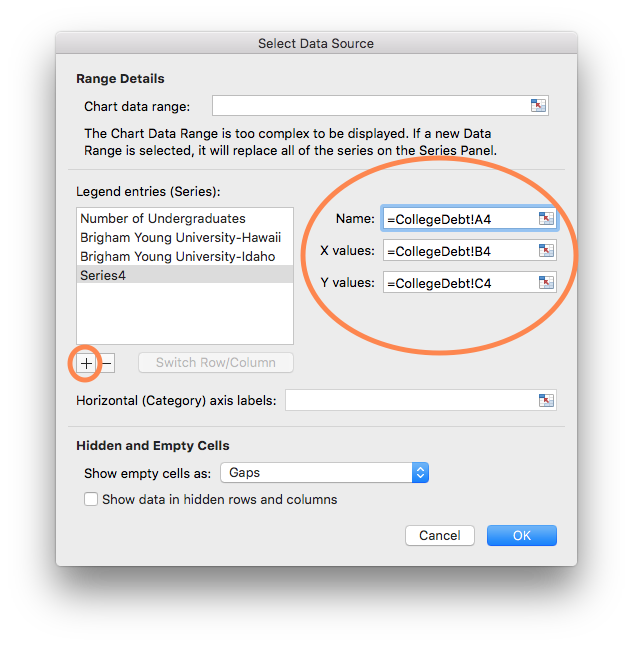
5C. **[Bonus Step]** Click the + sign and enter the following information exactly as shown.



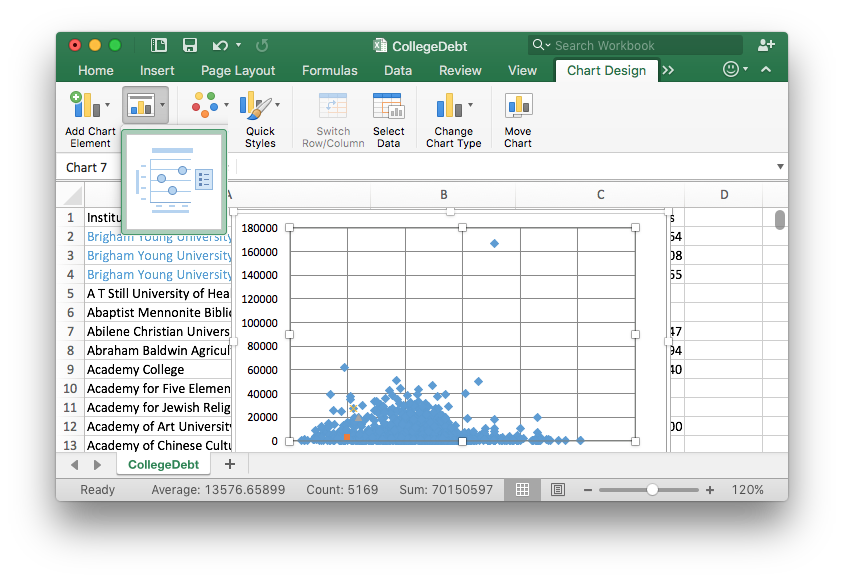
5D. **[Bonus Step]** Note that the information for Brigham Young University-Hawaii has appeared. Now, click the + sign *again* and enter the following information exactly as shown.

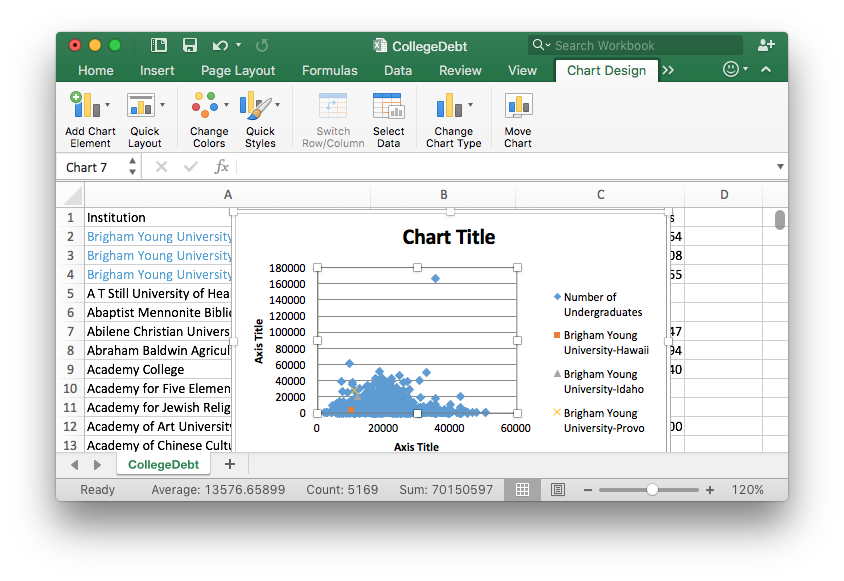


5E. **[Bonus Step]** The information for Brigham Young University-Idaho has appeared. Again, click the + sign and enter the following information exactly as shown, then press OK.



5F. **[Bonus Step]** Finally, select the "Quick Layout" tab in the top left menu and select the first option.



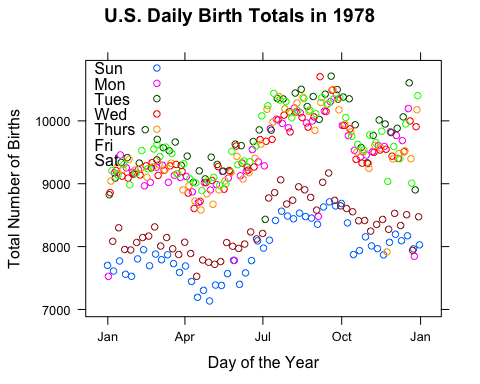


5G. To complete the process use the "Quick Styles" menu to find a style you like, then double click on the different elements of the chart to complete the finishing touches.

##### Interactive Content

What patterns do you notice in the following graphic?

xyplot( births ~ date, data=Births78, groups=wday, type=c("p"), auto.key=list(corner=c(0,1)), xlab="Day of the Year", ylab="Total Number of Births", main="U.S. Daily Birth Totals in 1978")



##### Answers include:

* There are typically fewer births on Saturdays and Sundays than the other days of the week.
* July, August, September, and October show higher numbers of births than the other months of the year.

barplot(apply(HairEyeColor,1,sum), col=c("black","tan4","red4","wheat"), ylab="Frequency", main="Hair Color of College Students")

