© M. Lin. Do not share or post.

4803 Assignment 2: Data Visualization in Excel

This week we review some of the useful visualizations in Excel, especially some that may actually be harder to do in Tableau. We also discussed the first set of visualization principles, especially in terms of color. **I strongly recommend that you read the first chapter of the textbook, "Big Book of Dashboards", which is freely available to you through GT library. You can find the link to the book on the syllabus.**

To complete your assignment as smoothly as possible, I strongly recommend that you either use Excel on the VLAB virtual machine (mycloud.gatech.edu), or use a windows-based PC with either Office 365 or Excel version 2016 or later. When you use the virtual machine, please know that the connection speed depends on not just your network speed, but also the number of simultaneous users (your classmates) and the overall GT network traffic volume. Therefore, using VLAB during the day is usually more difficult than using it in the evenings.

The following are some of the main visualization types that we did in this week's demo:

- Map chart
- Sparklines
- Heatmap
- Box plot and histogram
- Donut chart and waterfall chart

For this assignment, please complete the following. Please keep in mind that although we will be starting from a dataset, in some cases you might need to transform your data or prepare it in other ways before you can create the visualization you are expecting. This is true for future assignments in other tools such as Tableau as well.

What you need to submit:

Please only submit one Excel file titled "**Excel_viz_LastName_FirstNameInitial.xlsx**". Part 1 and Part 2 below should be separate worksheets of this same Excel file. Name your worksheets based on the serial number of questions below – e.g. the worksheet containing your answers for Question 1 should be named "**Q1**", etc.

Part 1: Continue with the car data that you worked on for the previous assignment – this time however I've provided you the **full dataset** (more data than before). Please use the Excel file that came with this assignment. You might need to create some Pivot Tables before creating visualizations. Please also note (as shown in the demo) that some visualizations cannot be created based on the Pivot Table results directly. Also, be sure to format your pivot table output correctly, per the videos (e.g., tabular format, no subtotals or grand totals).

1. (10%) Create a boxplot without outliers of the MSRPs for all cars in the data for the year of 2015.
2. (10% Create a histogram (with a bin to capture the outliers if needed) of the City MPGs of all cars in the data for the year 2010

© M. Lin. Do not share or post.

3. (10%) Suppose you are asked to PRESENT in a meeting the following table: The rows should represent "Vehicle Style", the columns should represent "Number of Doors", and the cells should represent "Average City MPG", based on this dataset. As mentioned in the videos, in general we should not present a big table with a bunch of numbers. Use heatmap or conditional formatting to show this table but with added visual cues.
4. (10%) A line chart shows the change over time. Create such a chart to show the number of "Makes" (manufacturers) captured in the dataset over time.
5. (10%) Use spark lines to repeat #4 above, but break it down by "Vehicle Style", i.e., each row should be a line chart (in the form of a spark line) for each vehicle style.
6. (15%) Create a donut chart to show, for the year 2010, the proportion of cars in the data that has the word "Luxury" in the "Market Category" variable, as compared to those that do not.

   Tip 1: Searching for a string within another string requires the "search" function, the "isnumber" function, and potentially the "if" function, e.g., if it contains that word, 1, otherwise 0.

   Tip 2: This is an important skill (using the IF function or the like) to help you highlight one category against another. You can use this to convert continuous variables into categorical or binary variables, or categorical variables into binary variables.

Part 2: Explore and Explain using the sample-superstore Excel file (the one you use for the demos)

7. (20%)Use the sample-superstore data to create a map chart (filled map; choropleth) of your choice (EXCEPT what you saw in the demo video).

   Note that in addition to states, you have quite a few other geographical variables. Play with the data and see if Excel shows what you have in mind. If it shows the entire globe, it means that the geo locations are not recognized (which points to the limitations of Excel map charts).

8. (15%) Suppose you are working as an intern for this store, and you have been given this dataset to look into. Based on what you have known about it so far, what is one thing that you thought is most interesting that you observed from the data (through any of the visualizations that you have done on it), or something that you want to explore further? Write it down as a sentence in less than 100 words (type it in at the top of the worksheet for this question). Then decide on the best type of Excel visualization to get that point across and add it to the worksheet. If it is something too complicated that you don't yet know how to create (e.g., the calculation is complicated), you can describe the viz verbally too. The key takeaway for you to think about is the match between the finding and the viz.