**Lab 1**

**Exploring Data Using Excel**

In this first lab, you will practice how to carry out descriptive stats and visualization using Excel. First, you will need to get familiar with Excel commands to set up the datasheet and dataset. Then, you will learn Excel functions for performing descriptive stats for continuous and categorical values as listed in Appendix A in your text. Lastly, you will learn ways to create various charts for seeing the big picture.

**Setting Up the Stage**

* Open the Lab 1 workbook and create a datasheet named Stats. Copy the dataset to your new datasheet.
* Create named ranges for data in the Position and Expenses columns. (Formulas|Define Names)

**Formulas for Descriptive Stats**

* Use functions as suggested to calculate the following stats for Expenses.
* Format the values with a conventional currency format.

|  |  |  |
| --- | --- | --- |
| Max | 1800.11 | max |
| Min | 56.75 | min |
| Median | 759.365 | median |
| Mean | 812.351 | average |
| Q1 | 424.5225 | quartile |
| Q3 | 1164.755 | percentile |
| IQR | 740.2325 |  |

* Use the Advanced filtering command (Data|Sorting & Filtering|Advanced) to generate a list of distinct position values, as shown in the first column in the table below. (Instructions available at <http://superuser.com/questions/49614/how-do-i-get-the-distinct-unique-values-in-a-column-in-excel>).
* Add **total** to the end of the list. Use functions as suggested to calculate frequencies, as shown in the second column.
* Format the proportion values in the third column as necessary.

|  |  |  |  |
| --- | --- | --- | --- |
| center | 5 | 25% | countif |
| guard | 5 | 25% | countif |
| forward | 5 | 25% | countif |
| center | 5 | 25% | countif |
| Total | 20 |  | count |

**Data Visualization**

Use the Insert tab

* Bar chart (for Position data)
* Box chart (for Expenses data)

Use the Data Analysis command on the Data tab (You may need to load [the Analysis ToolPak](https://support.office.com/en-us/article/load-the-Analysis-ToolPak-b6814e9e-5860-4113-ba51-e3a1b9ee1bbe) first:   
File | Options | Add-ins 🡪 Analysis ToolPak 🡪 Go…)

* Histogram (for Expenses data)

**Apply to a Big Dataset**

We’ve seen how it work on a make-up data set. Let’s try it out with data collected from a real world application: PM2.5 values measured hourly at US Embassy in Beijing during Jan to Nov 2016. (A copy of the data is available in the Misc folder.)

Use the Value column for continuous feature, and QC Name column as the categorical feature, and repeat what you did for the smaller dataset.

**Due by Wed, 1/18, before class.**