

Data Assignment 1

The Assignment

By **Tuesday, September 22**, you should complete the following assignment and submit a single Word document to me **by e-mail** with the subject "Data 1: <YOUR LAST NAME>." Your final document will consist of (A) a statement of rationale and methodology, (B) frequency table and bar chart for categorical data, and (C) a five-number summary and histogram for the quantitative data. There is a sample assignment on the MLP site, with data taken from the first day survey from another class.

Note that it is perfectly ok if you work with a friend, but (a) you should tell me who you worked with, and (b) you and your friend should not choose the exact same variables. Also note that this assignment uses *Excel* (instead of *StatCrunch*) so that you build some facility with this ubiquitous software package. Because of the statistical limitations of *Excel*, this is the only assignment that will use it.

Getting Started

Get a version of **BigSurveyResults.xls** from the MLP site and save it to your T: drive folder. (The data will need to be cleaned up, so you will save yourself time on future assignments by keeping your clean version handy.) Choose one categorical variable and one quantitative variable. Fix any inconsistently entered data or obvious errors before proceeding.

It is a good idea to delete all other columns on your version of the spreadsheet. (Select a column by clicking on its letter heading, then right click on the heading and select "Delete.") Save this smaller version with a different name so your good copy of the complete data set is intact.

Rationale and Methodology

Begin with a sentence that explains which questions you chose (write out the question) and why you chose them. Then briefly describe any work you had to do to "fix" the data set before you began your analysis. Also, this is where you should acknowledge any students who helped you, if anyone. Remember that you should not use the exact same variables as your collaborators.

Frequency Table Instructions

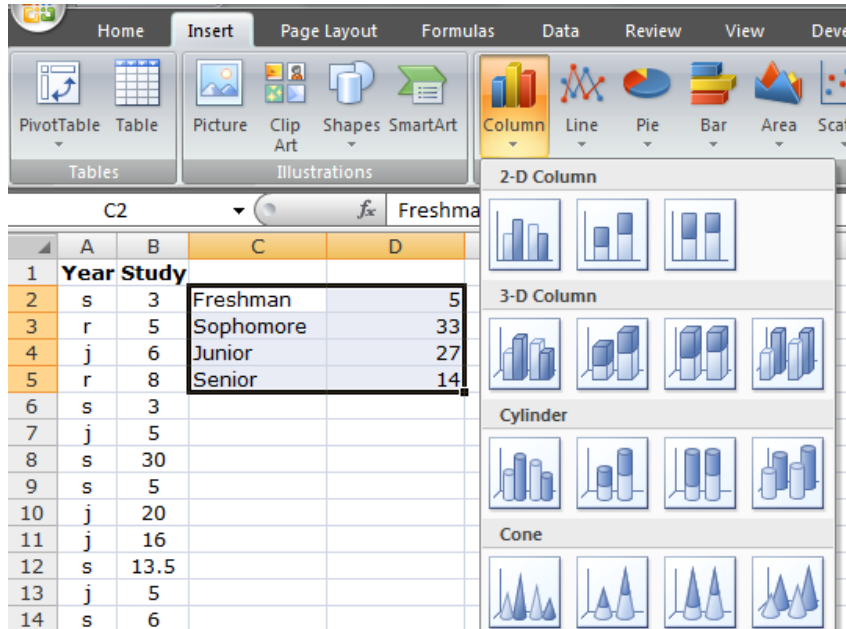
You will use the Excel function COUNTIF to count the number of matches to each possible categorical answer. In the instructions below, my categorical data is in cells A2 through A80, and I changed the existing entries to simply be f, s, j, r for consistency.

1. Start by typing descriptive names (i.e., what you want to appear on the chart) for the categories. Since my example uses "class year," I typed Freshman, Sophomore, Junior, Senior in cells C3 through C6.
2. In cell D2, I use the formula "=COUNTIF(A2:A80, "f")" which will count the number of people who answered "f".
3. In cell D3, I use the formula "=COUNTIF(A2:A80, "s")" which will count the number of people who answered "s".
4. In cell D4, I use the formula "=COUNTIF(A2:A80, "j")" which will count the number of people who answered "j".
5. In cell D5, I use the formula "=COUNTIF(A2:A80, "r")" which will count the number of people who answered "r".

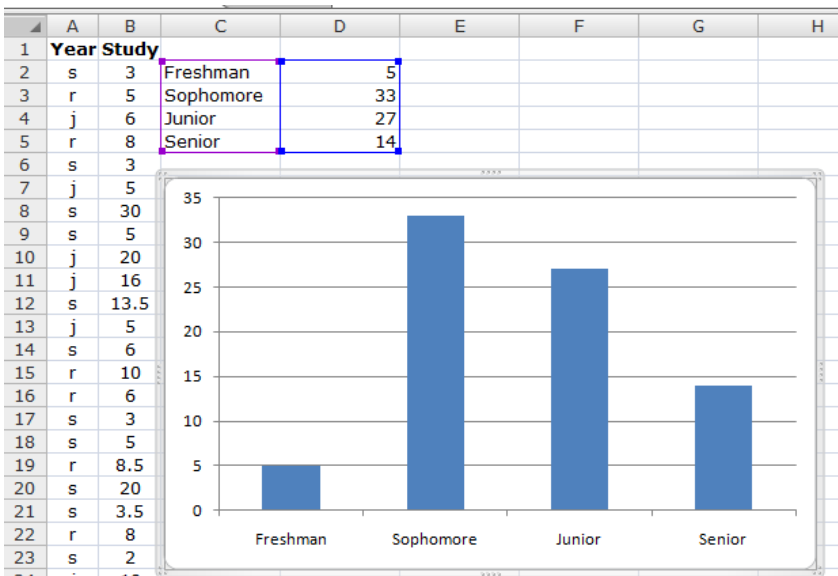
A	B	C	D	E
Year Study				
s	3	Freshman	=COUNTIF(A2:A80,"f")	
r	5	Sophomore	COUNTIF(range, criteria)	
j	6	Junior	27	
r	8	Senior	14	

Bar Chart Instructions

Select the cells that contain the descriptive labels and the counts, as shown below. Then choose the Insert tab and select the 2-D column chart that looks like a regular bar chart.



The result of this is the chart shown below. Google “create bar chart in excel 2013” (or whatever Excel version you have) and watch a video. Note that in Excel they call a vertical bar chart a “column chart” but it is ok if you want to make it horizontal by choosing “bar chart.” For style points, make the chart look fancy and cool. Once you have a chart you like, right click it to copy, and then paste it to your word document. Be clear (either in commentary or axis labeling) what the vertical axis measures (count, proportion, or percentage).



Five Number Summary Instructions

1. This explanation assumes the data is in cells B2 through B80.
2. In cells C82 through C86, type MIN, Q1, MED, Q3, MAX to serve as labels.
3. In cell B82 enter the formula “=QUARTILE(B2:B80,0)”
4. In cell B83 enter the formula “=QUARTILE(B2:B80,1)”
5. In cell B84 enter the formula “=QUARTILE(B2:B80,2)”
6. In cell B85 enter the formula “=QUARTILE(B2:B80,3)”
7. In cell B86 enter the formula “=QUARTILE(B2:B80,4)”

79	f	4	
80	f	3.5	
81			
82		1	MIN
83		4	Q1
84		6	MED
85	=QUARTILE(B2:B80,3)		
86	=QUARTILE(B2:B80,4)		
87			

Histogram Instructions

Define bins as shown in the picture below, and then use Google to find a video for “creating a histogram in Excel 2013” or whatever version you have. Once you have the histogram dialog box open, see the settings in the picture below (like “selecting” Chart Output) to get the histogram as shown.

	A	B	C	D	E	F	G
1	Year Study					BINS	
2	s	3	Freshman	5		5	
3	r	5	Sophomore	33		10	
4	j	6	Junior	27		15	
5	r	8	Senior	14		20	
6	s	3				25	
7	j	5				30	
8	s	30				35	
9	s	5				40	
10	j	20					
11	j	16					
12	s	13.5					
13	j	5					
14	s	6					
15	r	10					
16	r	6					
17	s	3					
18	s	5					
19	r	8.5					
20	s	20					
21	s	3.5					
22	r	8					
23	s	2					
24	j	10					
25	j	3.5					
26	s	5					
27	s	5					
28	s	8					
29	r	10					
30	r	1					
31	j	9.5					

Histogram

Input

Input Range: \$B\$2:\$B\$80

Bin Range: \$F\$2:\$F\$9

Labels

Output options

Output Range:

New Worksheet Ply:

New Workbook

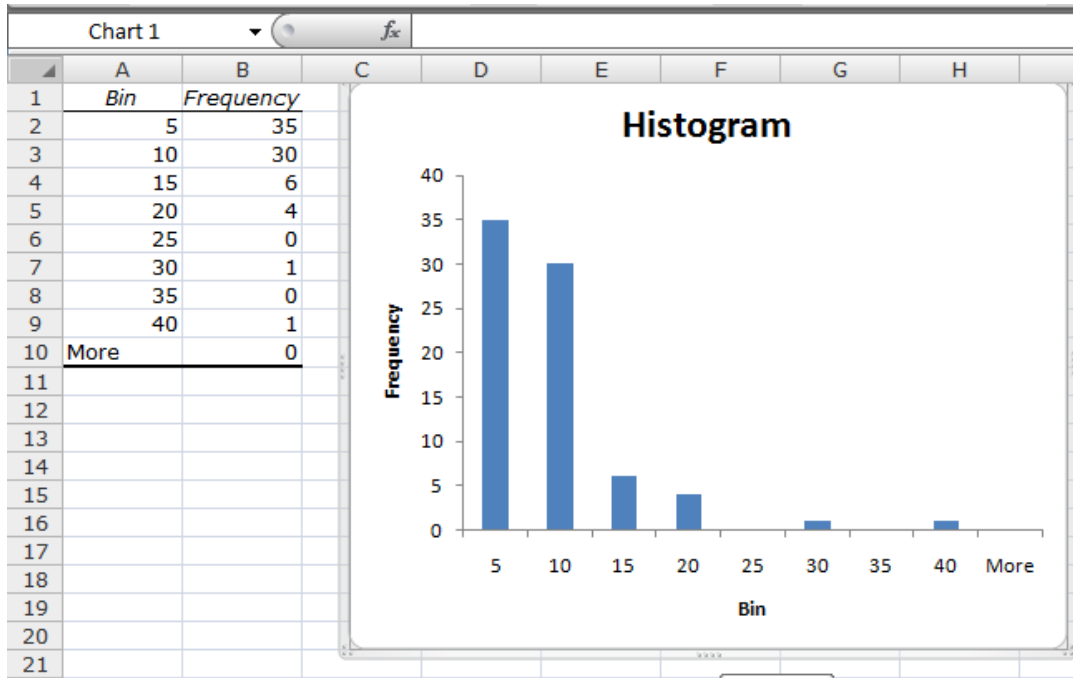
Pareto (sorted histogram)

Cumulative Percentage

Chart Output

OK Cancel Help

Once the histogram has been created (as in the picture below) you can right click to copy it, and then paste it into your *Word* document report.



Assignment Rubric (20 points total)

- ___/2 Rationale and methodology is complete
- ___/4 Well-formatted table correctly displays counts and proportions/percentages
- ___/4 Clearly labeled bar chart that matches the frequency table
- ___/4 Correct five number summary
- ___/4 Clearly labeled histogram
- ___/3 Intangibles such as extra effort in use of other Excel features or Word formatting to enhance the appearance of your submission, or extra commentary on the results.