

# Assignment 1

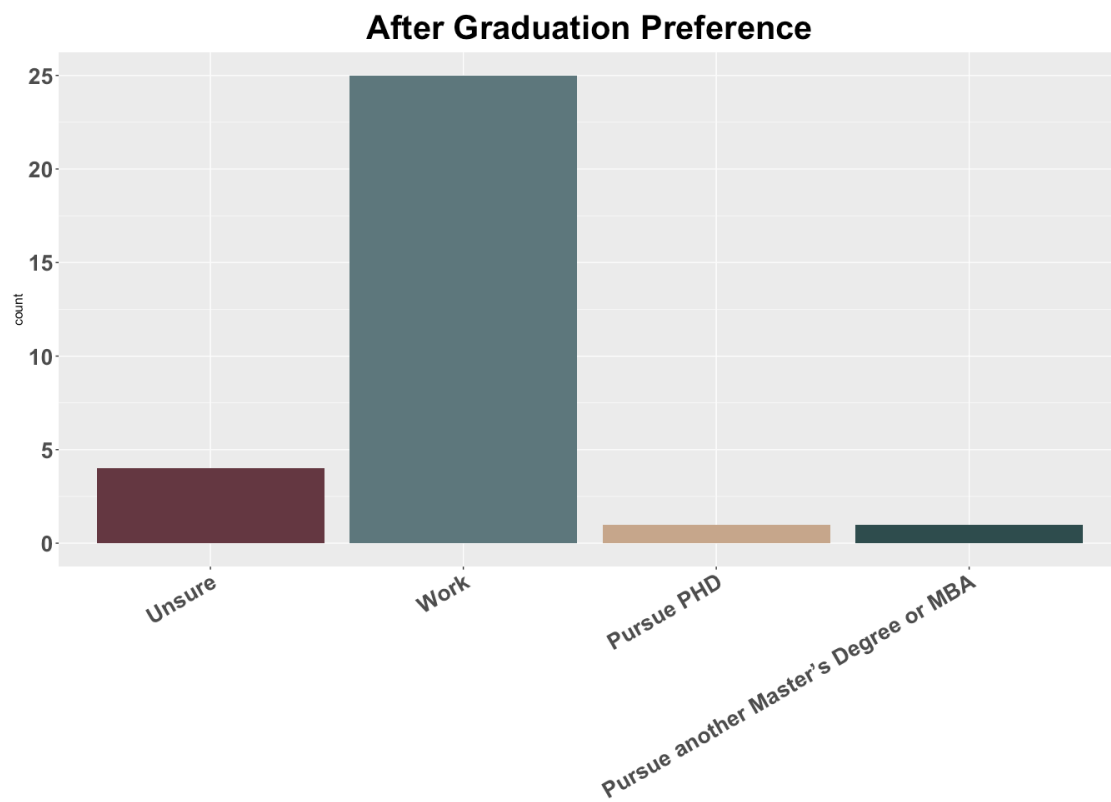
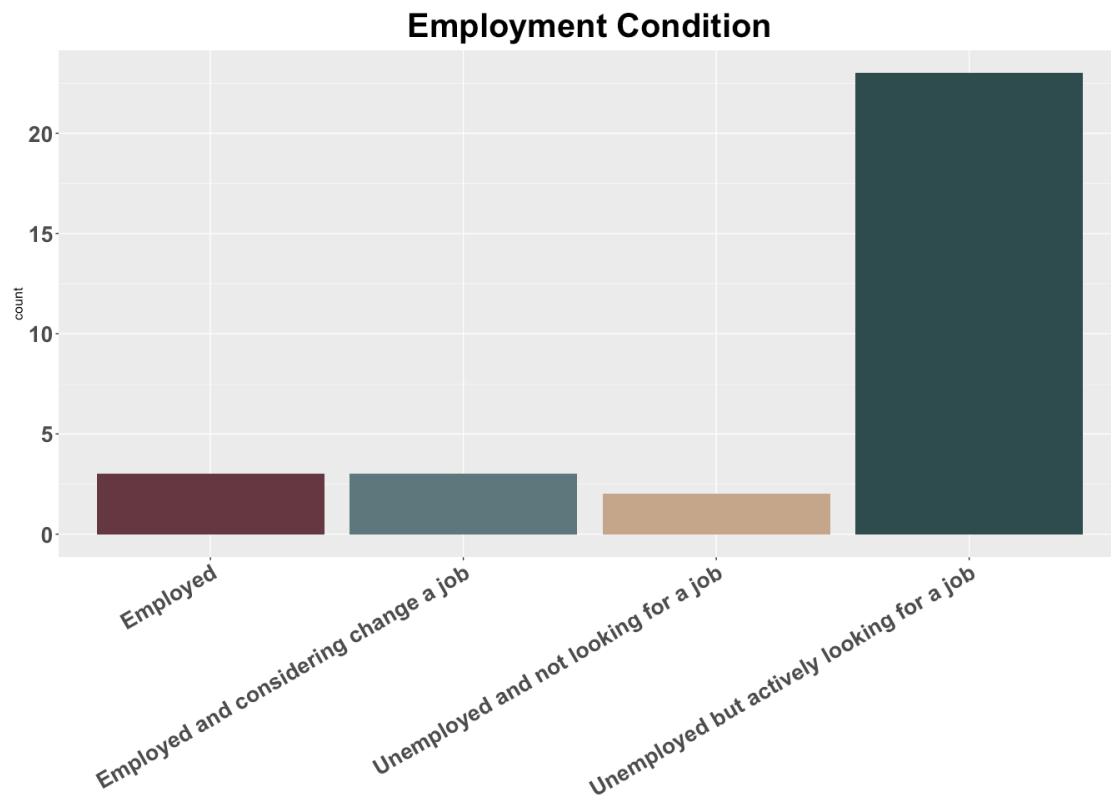
Yuwen Wu(336)

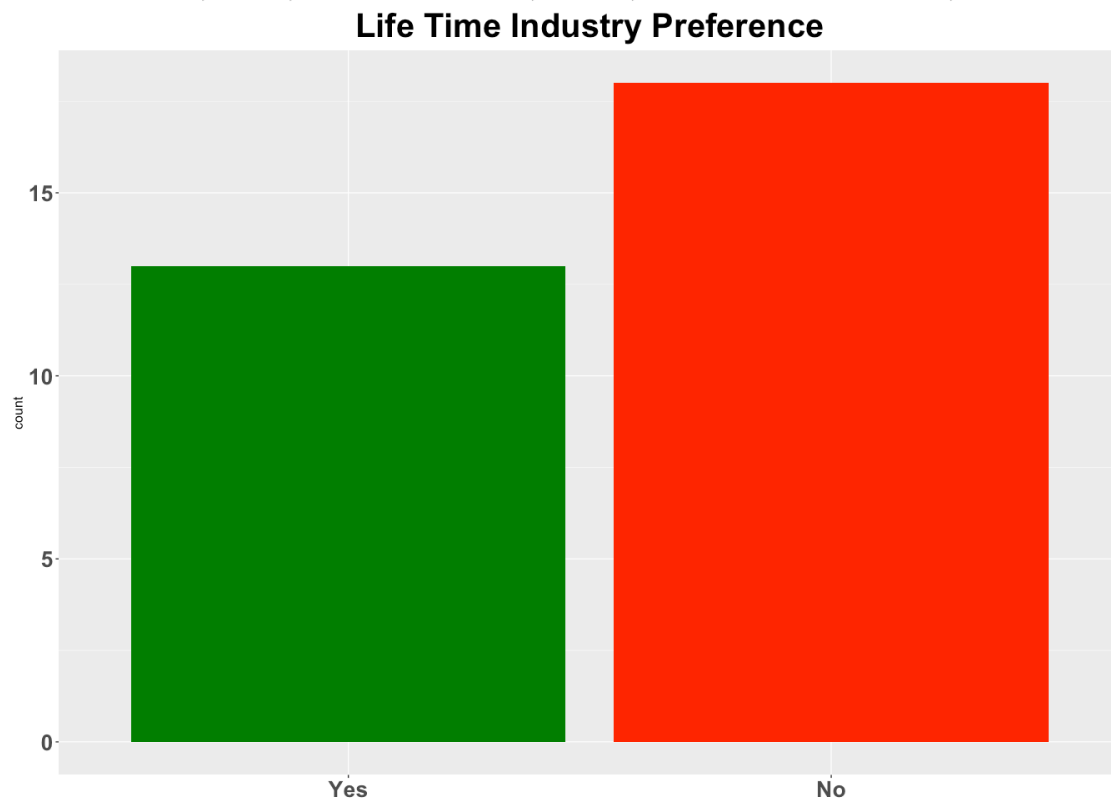
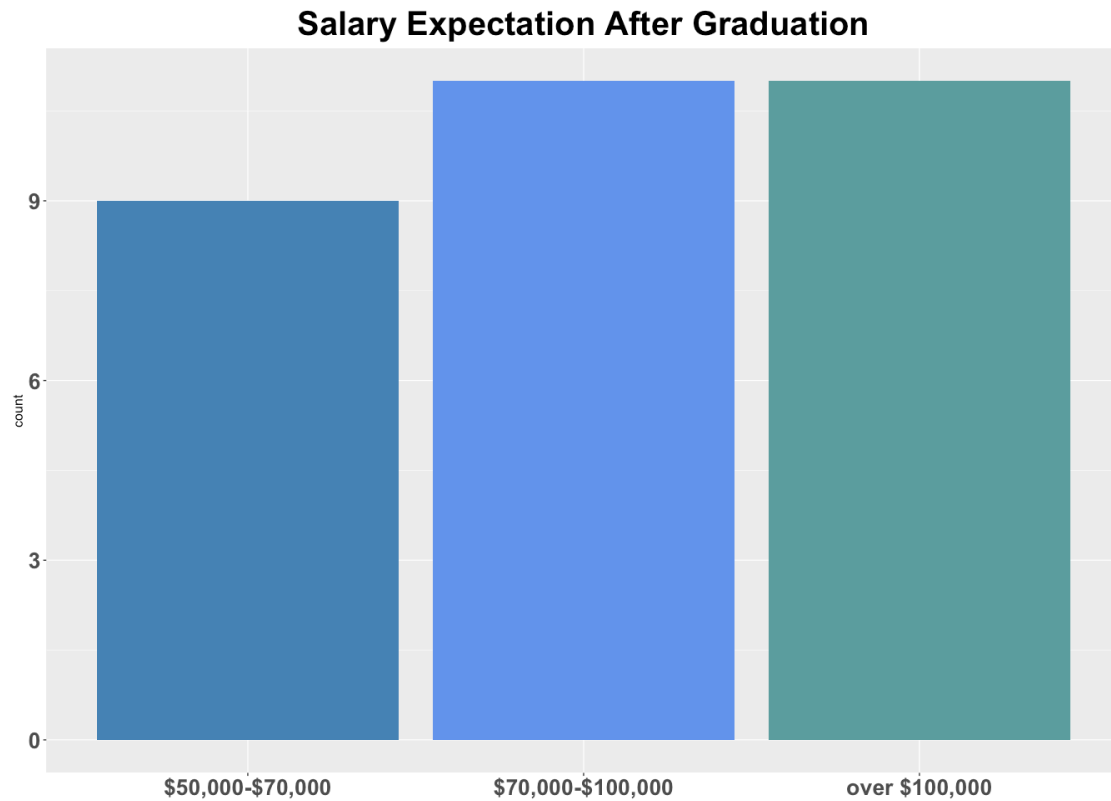
9/23/2019

I am allocated at Group D. in our group question, we have several easy answer questions for everyone such as the laptop brand, gender and course related questions. For the first question we asked: "Who is your laptop manufacturer" was a question people interpreted as we intended and we provided multiple selection for choosing, making people easy to understand the question. For most of our questions such as the question "Which operation system do you use?" and "Did you take any previous courses of Professor Nagaraja", We provided all possible answer to be chosen. In this case, we could collect data easily and without much missing or misinterpretation. For questions provided by other groups, in my mind, the open questions are really difficult to quantified but the answer provided by survey taker are meaningful. So, for our group, we could contribute more open questions to our survey.

For the people who doesnot answer the questions, in my mind, people may too busy to check their mailbox and then may miss the survey. Additionally, people may not willing to tell any information about themselves. Therefore, the missing data here should be Missing at Random, which is the probability that an observation is missing commonly depends on information for that subject that is present, i.e., reason for missing is based on people forgetting or unwilling to telling anything about themselves."

Note: the data here is from our class survey, without any solid analysis and full backgroud checking survey. Therefore, the data only use for data measurement and visualization practice not for any publication or reference.





The data source of these charts are based on students at Data Measurement and Data Visualization course of Fordham Gabelli Graduate School of Business. There are 5 missing

data which could be regarded as random missing cause students may miss email or too busy to do this survey at beginning of semester, but they may also be not willing to provide any private info at this survey. Data quality may not be reliable since people could misunderstand the question or just give wrong information and there were some questions they may feel not quite sure how to answer and questions are based on personal condition and then done it personally.

From above bar charts, most students in Measurement and Data Visualization course are unemployed and actively seeking for a job right now. Only two of students are unemployed and not looking for a job at the last semester which is now. But, from the survey, they all choose to work after graduation. Therefore, they may decide to job seeking after graduation rather than at last semester. Also, there is one student decide to pursue Ph.D degree after graduation and one student decide to pursue MBA or other master degree after graduation.

From salary expectation, students have expected average higher than 70K\$/yr per year after graduation. I can guess, this is why most of student actively seeking jobs right now! Since most students major in business analytics and applied statistics in this course, student happy to work in different industry through whole life. They are really open to various industry.

##	KEY U.S.	LAST	CHG	% CHG
## 1	DJIA	28645.26	23.87	0.08%
## 2	Nasdaq Composite	9006.62	-15.77	-0.17%
## 3	S&P 500	3240.02	0.11	0.00%
## 4	Russell 2000	1669.03	-8.64	-0.51%
## 5	OTHER U.S.	LAST	CHG	% CHG
## 6	DJ Transportation Average	10936.70	-13.90	-0.13%

This matrix is found from *Wall Street Journal SPX index* page. This page displays daily key U.S. stock index last day close price and change point (basic point) and percentage change of the stock price of that day. The S&P 500, is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. The S&P 500 is a capitalization-weighted index and the performance of the 10 largest companies in the index account for 21.8% of the performance of the index.

From the second chapter of *The Tyranny of Metrics*. There are several "Recurring Flaws" listed:

### **Measuring the most easily measurable**

Since the capitalization-weighted index would overweight large capital company performance and underweight small capital company performance, it would not present the true performance of the equity market. However, the capitalization-weighted is easy to measure, therefore, this Metrics has the flaw of Measuring the most easily measurable.

### **Measuring inputs rather than outcomes**

In this Metrics, it is measured the inputs of all stock composite in capital weighted. However, these movements do not show any possible outcomes of economy or other social problem such as unemployment rate.

### **Degarding information quality through standardization**

As mentioned above, S&P 500 index is based on average method of mathematics average method. During this process, the data quality maybe dagarded. This indexe may ignore so called outliers and smooth the data during average process.

### **Gaming through creaming**

This flaw not appear in above matrix since these matrix are built for customer needs and work for institutions.

### **Improving numbers by lowering standards**

For above metrix, this flaw should be not existed since financial data is really high quality and without any mistake from data institutions.

### **Improving numbers through omission or distortion of data**

The index is weighted by a float-adjusted market cap. It only measures the shares available to the public. It does not count those held by control groups, other companies, or government agencies. (S&P Dow Jones Indices)

A committee selects each of the index's 500 corporations based on their liquidity, size, and industry. (S&P Dow Jones Indices)

It rebalances the index quarterly, in March, June, September, and December. To qualify for the index, a company must be in the United States and have a market cap of at least \$6.1 billion. At least 50% of the corporation's stock must be available to the public. Its stock price must be at least \$1 per share. It must file a 10-K annual report. At least 50% of its fixed assets and revenues must be in the United States.(the balance)

From above, this data indeed has some omissions.

### **Cheating**

This flaw, in my mind, does not show in above metrix since this is qualified financial data.

### **Reference**

<https://www.thebalance.com/what-is-the-sandp-500-3305888>. The S&P 500 and How It Works

S&P Dow Jones Indices. "Float Adjustment Methodology," Page 2. Accessed Dec. 10, 2019.