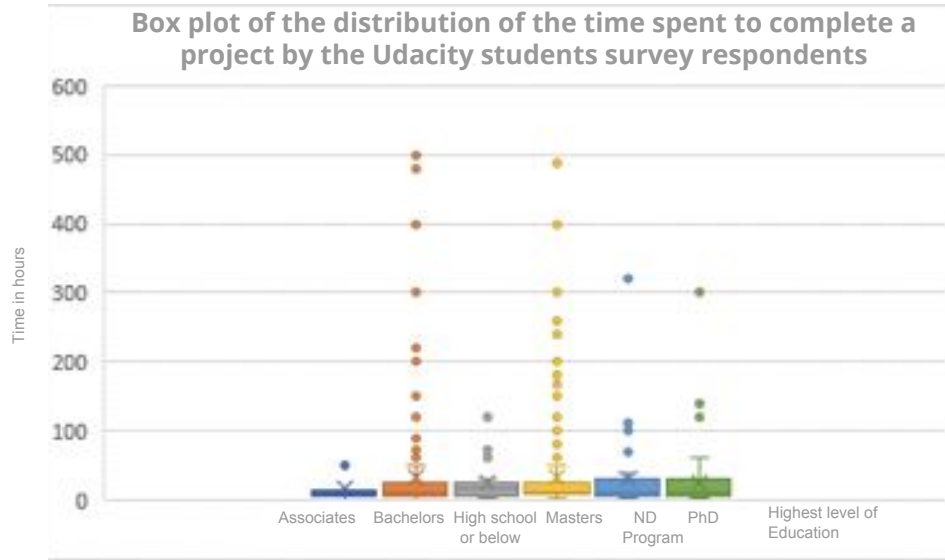# Data Foundations Nanodegree by Udacity

Project: Analyze Survey Results

The data used for this project was collected from Udacity students who responded to a survey. The data was cleaned, and data that didn't make sense was removed according to the criteria stated in the Info sheet of the attached Excel Workbook. I didn't use any external resource when creating this submission.

This presentation explores the four following questions, always bearing in mind that the survey respondents are not the entire population of Udacity students.

- Is there a relation between the time spent to complete a project and the level of Education?
- Do unemployed students recommend Udacity more than employed students?
- Is there a significant study time difference based on the country of origin of the students?
- Do students who take the most common Nanodegree Program study less than other Nanodegree students?

Is there a relation between the level of Education and the time spent to complete a project?

Box plot of the distribution of the time spent to complete a project by the Udacity students survey respondents

Time in hours

Associates Bachelors High school or below Masters ND Program PhD Highest level of Education

Here are the box plots for the survey respondents. I removed two extreme data points from the data set as they seemed unrealistic (800 and 1000). All distributions appeared to be right-skewed. Therefore the mean for each is higher than the median.

The mean for the student respondents Associates seems to be the lowest at about 15.5 hours, whereas for the other students it is comprised between about 24 and 32 hours. Such a difference might be due to the much smaller sample size of the Associates (12 students).

The standard deviation for the Associates and High School or below is respectively 18 and 30 hrs, while for the rest of the students it is above 43 hrs, and even 70 hrs for the Masters. This might be due to outliers who took from 200 to 500 hrs to complete a project. It could also be related to the difficulty of the project of the different Nanodegrees they were enrolled.

If we exclude the Associates, the first quartile is about 7 hours and the third quartile is about 27.5 hours for the rest of the students. The time to complete a project seems to be rather consistent despite the respondents holding different levels of education.

Do unemployed students recommend Udacity more than employed students?
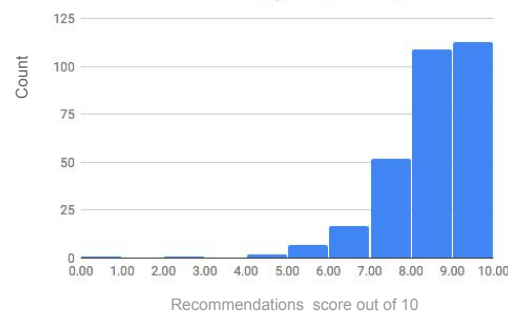
Bar chart of the average recommendations score by Udacity student respondents
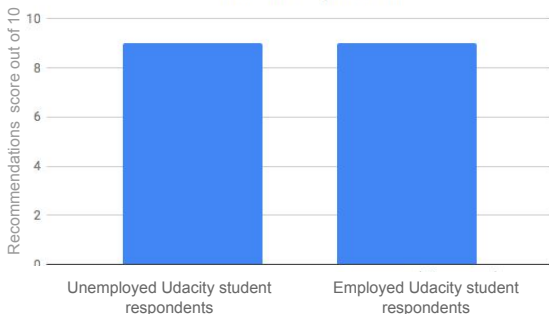
Histogram of the distribution of recommendations by unemployed Udacity student respondents
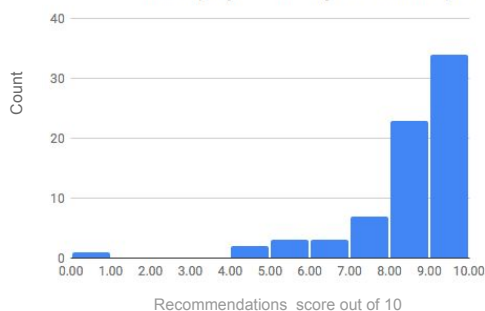
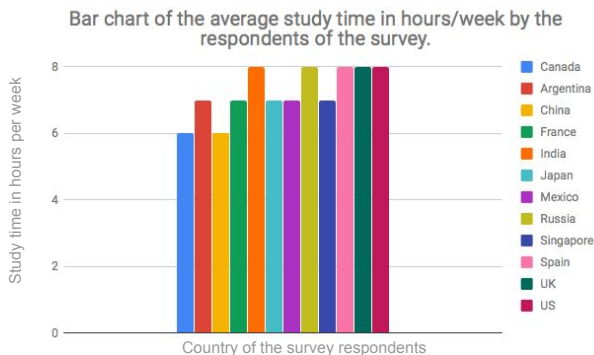**Histogram of the distribution of recommendations by employed Udacity student respondents**

First it is important to mention the difference in sample size. There are 133 unemployed respondents and 620 employed respondents, which means that there are about 5 times more employed respondents.

The bar chart of the average recommendations score shows that unemployed Udacity students seem to recommend it as much as employed students, with both categories displaying an average score of 9 out of 10.

Both histograms of the distribution of recommendations reveal a left-skewdness. While both have a mode of 10 and a mean of 9, it is worth noticing that the unemployed respondents data set has a standard deviation of about 2, whereas the employed one is of about 1. This standard deviation difference could imply that employed students have given higher recommendation scores than unemployed students. But we have to keep in mind that this variability might be due to the difference in sample size, or to other factors about the data set that are unknown to us at the time of the report. .

Is there a significant difference in study time based on the country of origins of the students?

Bar chart of the average study time in hours/week by the respondents of the survey.

Legend: Canada, Argentina, China, France, India, Japan, Mexico, Russia, Singapore, Spain, UK, US

Box plot of the distribution of the study time by country of residence of the Udacity students survey respondents
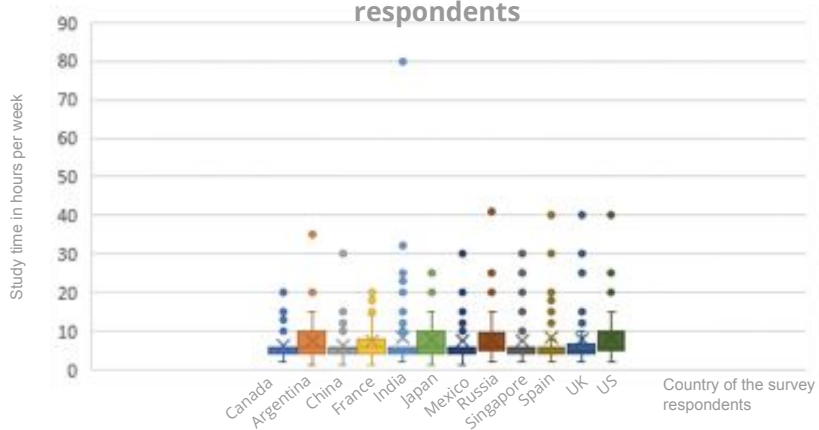
Here is the bar chart of the average study time in hrs/week by the respondents, and the box plots of the distribution of their study time.

Across the 12 countries the average study time is about 7 hours per week which could imply that respondents study for about the same time per week based on the mean.

The mode is 6 hrs/week across all countries, which seems to indicate a consistent length of study habit across all respondents.

All distributions appear to be right-skewed. The mean for each is higher than the median.

All countries have a Median of 6, except for China and Singapore with 5.5.

India is the country with the highest standard deviation of about 12 hrs per week. The variability in commute times might be due to outliers with long study times such as 80 hours per week.

Argentina and Japan have the largest interquartile ranges with 6 and 5.5 respectively. The middle 50% of the dataset is the most spread out for the respondent students from these countries, which might indicate more variability in study time habits.

Lastly we should bear in mind that this dataset shows the country of origin of the respondents to the survey, therefore it does not necessarily reflect a trend of the whole Udacity population.

Do students who take the most common Nanodegree study less than other Nanodegree students?
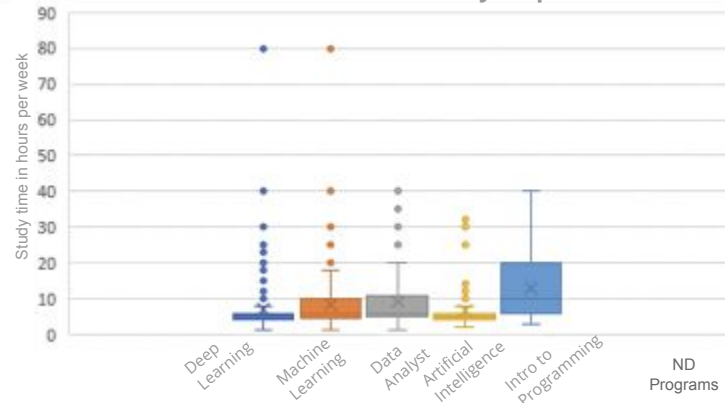

The five most common Nanodegree Programs taken by the students from the survey sample
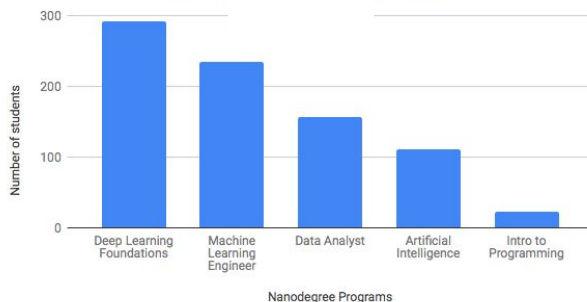
Average study time of the respondents from the five most common Nanodegree Programs
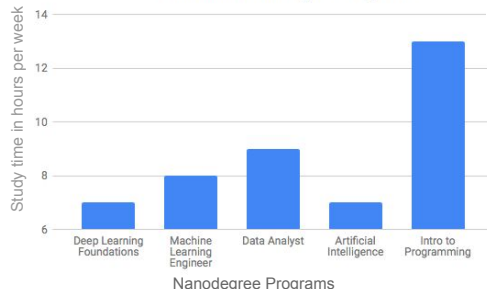
Box plot of the distribution of study time for the five most common ND of the survey respondents

First, as can be seen in the bar chart of the five most common ND programs taken by the respondents, most students study Deep Learning Foundations.

However, Intro to Programming has the highest average study time with 13 hrs/week. The respondent sample might not provide us with reliable insights because of its small size of 23, and we should be cautious when comparing data. This Intro to Programming sample contrasts with the sample of the Deep Learning Foundations students who dedicate 7 hrs/week to studying.

A look at the box plots reveal that the data is right-skewed. Accounting for the outliers, both Deep Learning Foundations and Machines Learning Engineers students reach a very high Maximum of 80 hrs/week with a Minimum of 1, which are the two largest ranges.

The Deep Learning Foundation students study time dataset had a median of 6, similarly to Machine Learning Engineer and Data Analyst students. It has the second lowest Standard Deviation of about 6.65 and an IQR of 2 which shows that most of the data is near the mean. It also has a mode of 6 just like the other Nanodegree students study time dataset.