

Survey Data Analysis

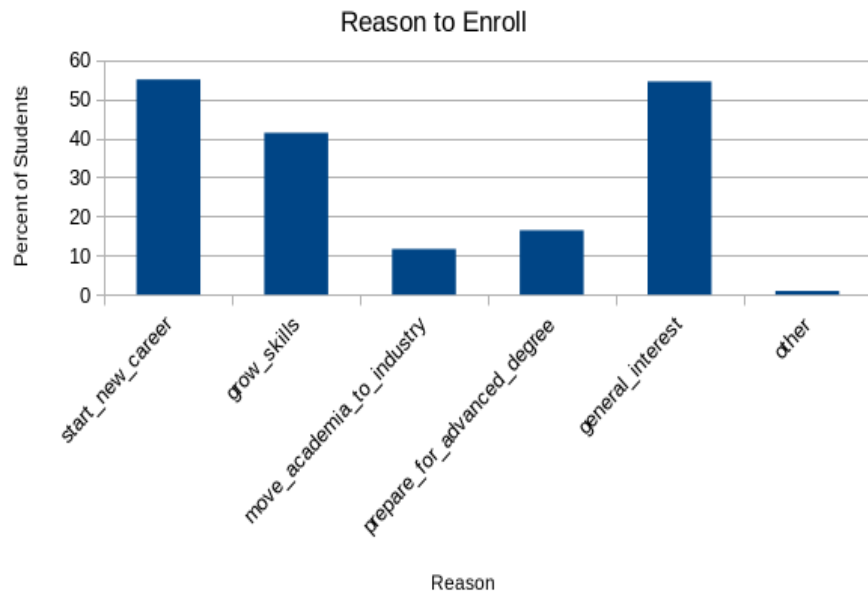
by

Silvana Avramaska-Lukarska

In this project we will analyze a real dataset about current Udacity students across a number of different programs. We will use spreadsheet for the analysis.

Please note that the data is from Survey Respondents and is not from the entire Udacity Student population.

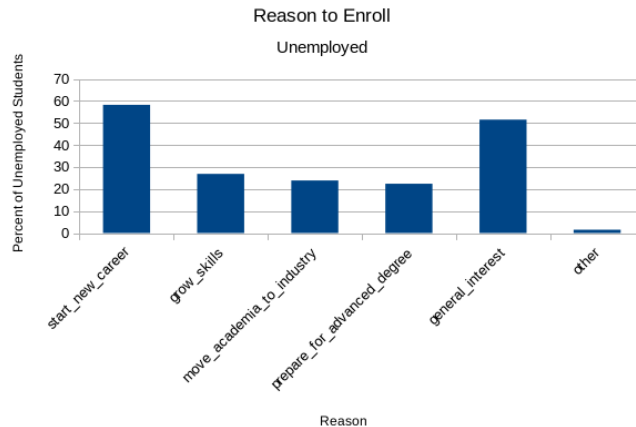
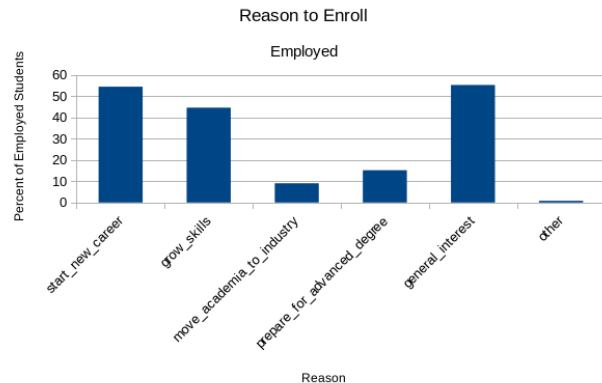
What are the most popular reasons to enroll at Udacity?



Here we look at the different reasons for students in this study to enroll at Udacity. For that we use a bar chart.

We notice that the most popular reason to enroll is to “start a new career” (55.04%), closely followed by “general interest” (54.51%). The least popular one is “other” (0.93%) and “move from academia to industry” (11.67%). We note as well that the percentages do not add up to 100% meaning that some students have indicated more than one reason to enroll.

Is the most popular reason to enroll at Udacity different for employed and unemployed students?



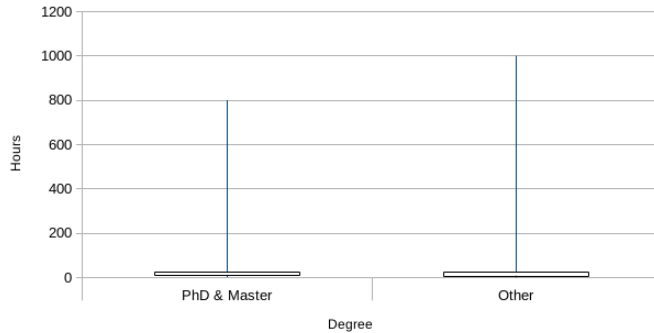
Here we look at the reasons to enroll for employed vs. unemployed students.

We notice that for the employed students the bar chart is pretty similar to the bar chart for all students with the slight difference that here the most important reason to enroll is “general interest” (55.16%), followed by “start new career” (54.35%) and “grow skills (44.52%”, whereas for unemployed students “start new career” (58.21%) and “general interest” (51.49%) seem to be far more popular than “grow skills” (26.87%), “move from academia to industry” (23.88%) and “prepare for advanced degree” (22.39%).

What is the average time to complete a project for students with different level of education?

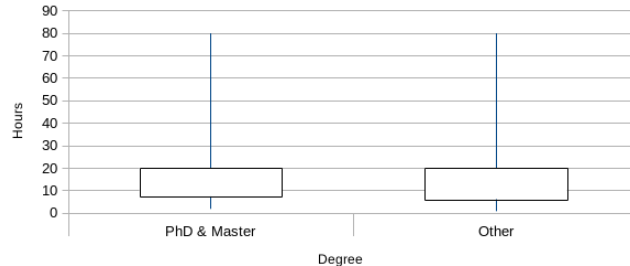
Average Time to Complete a Project

(with Outliers)



Average Time to Complete a Project

(without Outliers)

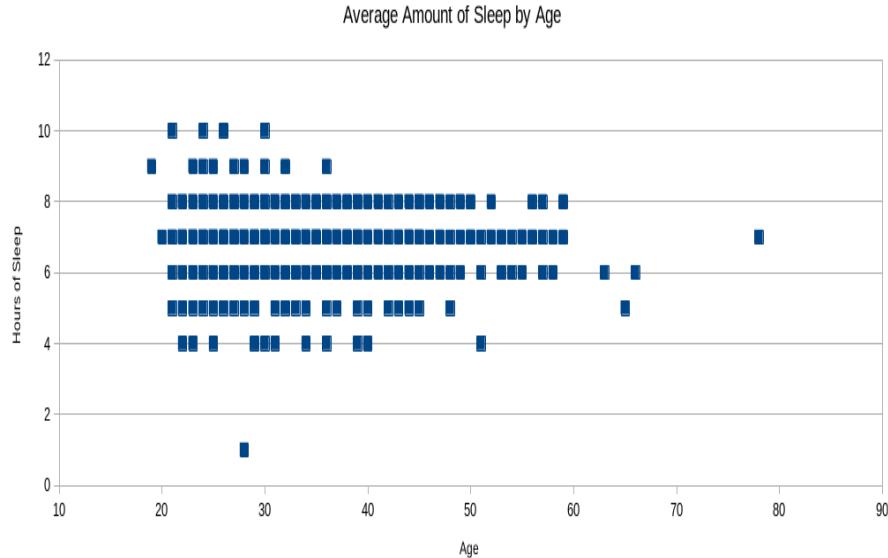


Here we look at the average time needed to complete a project for graduate students (with PhD or Master degree) vs. others. For that we use box plots.

There are 2 plots: the first one is before removing the outliers and the second one is after the outliers have been removed. Here outlier is defined as an observation lying outside ($Q1 - 3 \cdot IQR$, $Q3 + 3 \cdot IQR$).

There does not seem to be much difference between the average time to complete a project for graduate students vs others. For both groups the mean and standard deviation before removing the outliers are at approx. 33 and 77 hours, resp. However, after removing the outliers the mean for both groups is at approx. 17 but the variability of the time that a graduate student needs to complete a project (15.67) is slightly smaller than the variability of the time of the other group of students (17.56). For both groups in both situations (before and after removing the outliers) the biggest portion of the students (mode) reported to have spent an average of 10 hours on a project. We note that the mode is smaller than the mean which is to expect since all the distributions are right skewed.

Is there a relationship between the age and the average amount of sleep that students get?



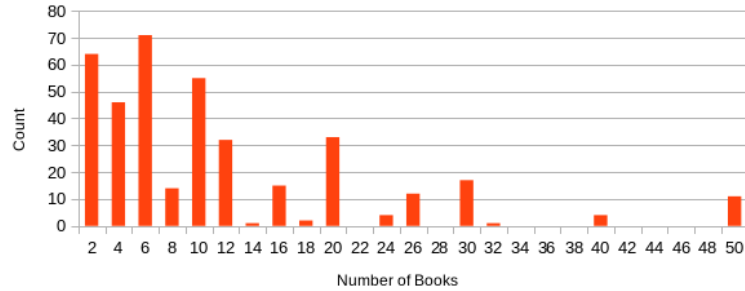
Here we look at the relationship between age and the average amount of sleep per night that the students in this study get. For this we use a scatter plot.

Surprisingly, it does not seem that age plays a big role in the amount of sleep that the students get. As expected, most of the people who sleep longer (more than 8 hours) are among the younger students. However, the people who get less sleep on average (less than 5 hours) are also the ones that are less than 55 years old). At the same time, the students that are more than 60 years old have reported to sleep between 5 and 7 hours which is within 2 standard deviations (2×0.99) from the mean (6.92 hours).

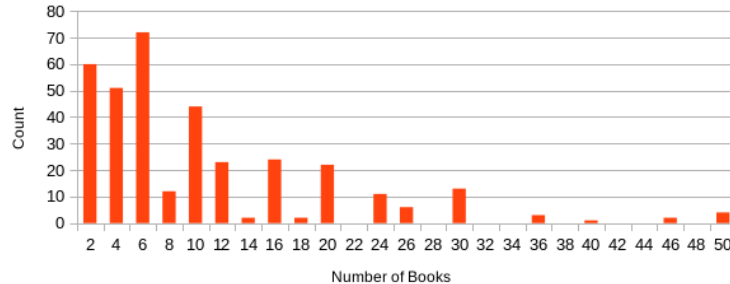
Note that there is a very big difference (9 hours) in the number of hours of sleep that students have reported. While the maximum (10 hours) seems plausible, the minimum (1 hour) might be an input error.

Do graduate students read more books per year on average compared to other students?

Number of Books
Graduate Students
(without Outliers)



Number of Books
Other Students
(without Outliers)



Here we look at the average number of books the students in this study read per year. We compare the data for graduate (PhD and Master) vs. other students. For this we use histograms. On the left are shown the histograms after removing the outliers (observations lying outside of $(Q1 - 3 \cdot IQR, Q3 + 3 \cdot IQR)$).

As expected, both distributions are right skewed. In addition, it does not seem to be any difference in the distribution of the average number of books graduate students read compared to the number of books other students read. However, on average (we have used the median since the distributions are skewed), the graduate students read 8 books per year, whereas the other students read 6 books per year.