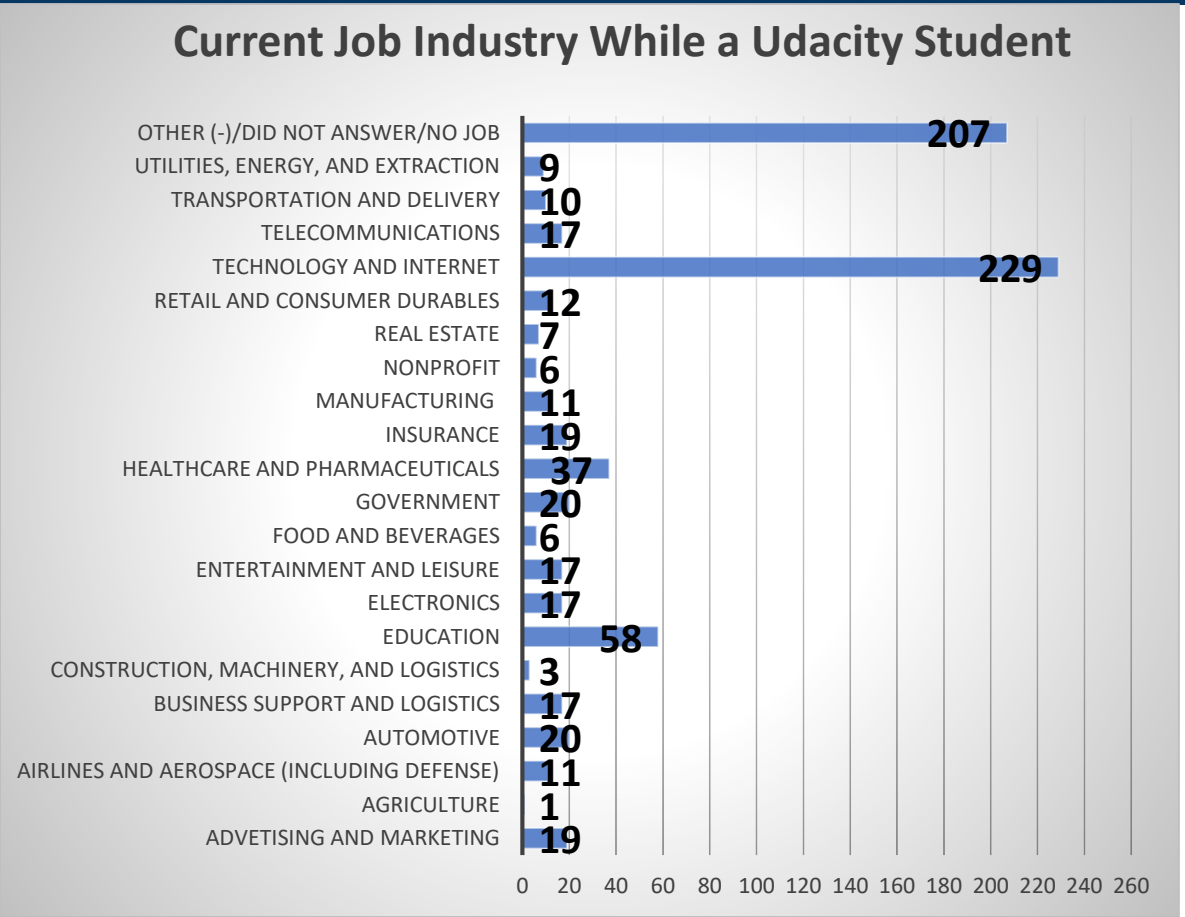
Project: Analyzing Survey Data

By: Brian Berns

What are the most common job industries Udacity students currently work in?



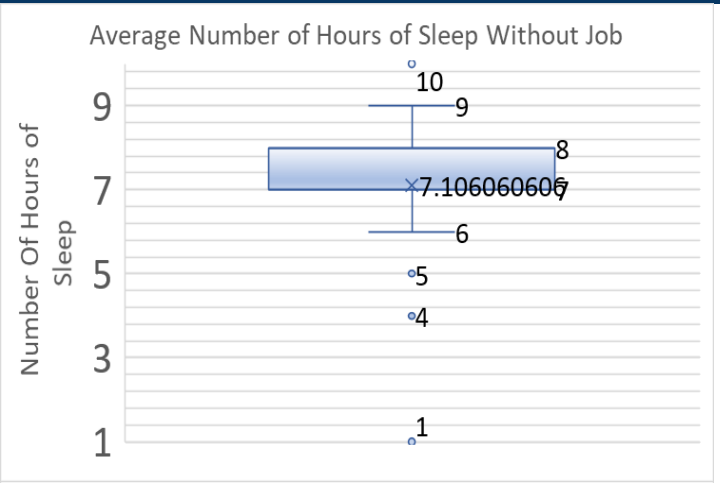
Shown to the left is a clustered bar chart tallying the number of Udacity students in a specific job industry they currently work in.

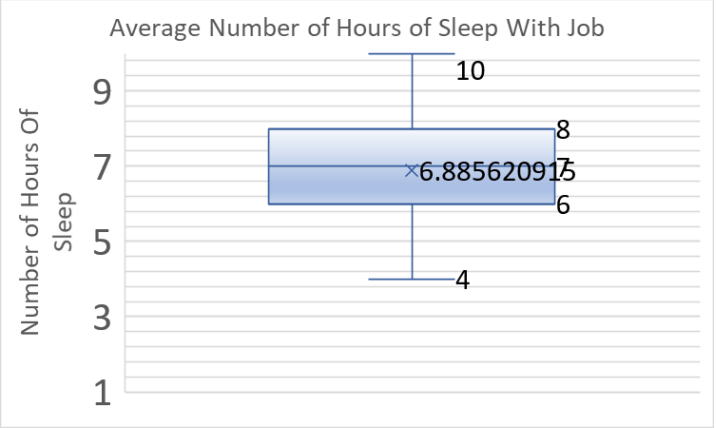
The industries that employ the most amount of Udacity students are the “Technology and Internet” and “Education” sectors.

The industries that employ the least amount of Udacity students are the “Agriculture” and “Construction/Machinery/Logistics” sectors.

There were 207 people that left this question blank on the survey. There could be a variety of reasons this may have happened, including industries not being listed, current student, jobless, or not wanting to share info, etc.

Employed/Not Employed Vs. Average Number of Hours of Sleep





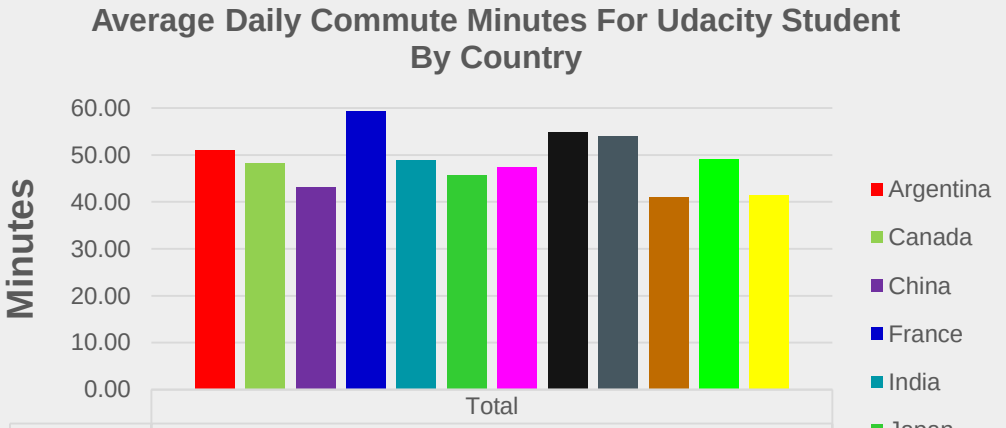
	With	Without
Minimum	4	1
Q1	6	7
Median	7	7
Q3	8	8
Max	10	10
Mean	6.882736	7.082707
S.D	0.921651	1.23267

These two box and whisker plots were used to see if there was a difference between the average number of hours of sleep Udacity students had, if they had a job versus did not have a job. The only data omitted in these charts were answers of incomprehensible amounts of sleep, such as 45/65/85/9141984 hours. I also omitted the data if they did not list the number of hours of sleep.

From looking at the two plots, and the data points in the table next to them, I can see that students with a job slept less hours on average versus students without a job. You can also tell from the plots and the Standard Deviation (in the table) that students without a job had a bigger range of data. Since the standard deviation is higher for hours of sleep without a job, those hours are farther from the average, which means that it has a bigger spread of sleep hours. You can also tell that students without a job had a bigger range of values because its IQR (which is 2) is bigger than the IQR of students with a job (which is 1).

One other statistic I can pull from the data is the fact that in both sets, the mean and the median are almost the same, which means these are very close to being normally distributed.

What is the average number of hours a Udacity student has to commute to work? (by Country)



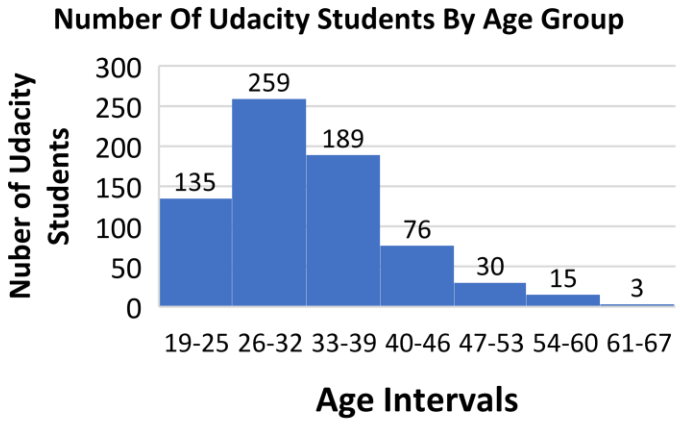
	Total
Argentina	50.99
Canada	48.34
China	43.13
France	59.29
India	48.93
Japan	45.71
Mexico	47.40
Russia	54.99
Singapore	54.05
Spain	41.02
UK	49.07
US	41.56

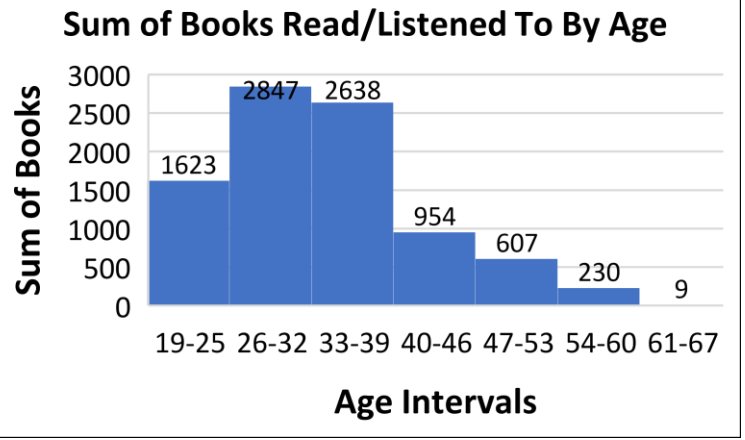
This Bar Graph was used to find the average number of commute minutes a Udacity student travels per day. I organized them by country.

I found that students in France had the highest average minutes to commute to/from work. I also found that Spain and the United States had the lowest average minutes to commute to/from work.

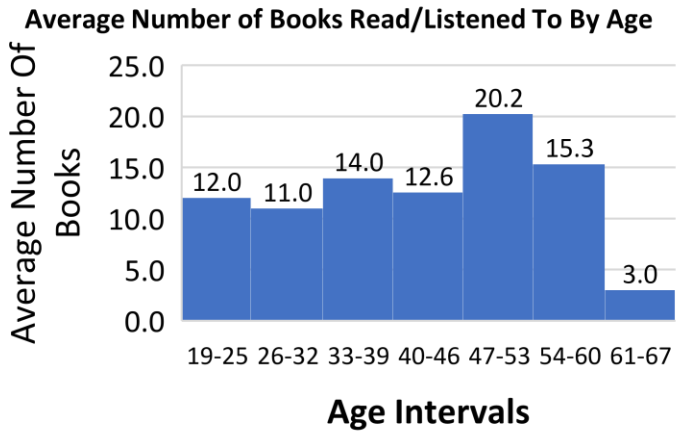
The only data I omitted was if the number of commute minutes per day was astronomical (such as 600, 400, and 360 minutes respectively). I realized that a good amount of students (143 students to be exact), who took this survey listed 0 minutes for a commute. This may be due to the fact they may work at home, they may still be a student, or they do not have a job at the moment. Since the data had so many “0 minute commutes”, this may be why some of the averages between countries are so close to one another.

Comparing Age to Number of Books Listened To By Udacity Students Per Year





The data I omitted was any age not in the range for a Udacity student (18-65). I also omitted any data in the number of books column that was left blank or had an unreasonable amounts of books read (such as 300 and 600 books).



For this set of data, I compared the ages of Udacity students to the number of books they read per year. For this particular question, I wanted to compare the data a couple different ways. To do this I looked at the numbers themselves, and the distribution. What I realized first is that the “Number of Udacity Students by Age Group”, and “Sum of Books Read/Listened to By Age” follows the same distribution. Both graphs are skewed to the right, which means that the mean is less than the median. I saw that, from just our numbers from the survey, a majority of the students were younger. This is why sum books is also the greatest in the younger age intervals.

But if you notice in the “Average Number of Books Read/Listened to By Age”, the data is actually skewed a bit to the right. This means that the mean is greater than the median. Which means even though cumulatively younger people read MORE books in total as a group, the older students read more books per individual.