

Project 2

Analyze Survey Data



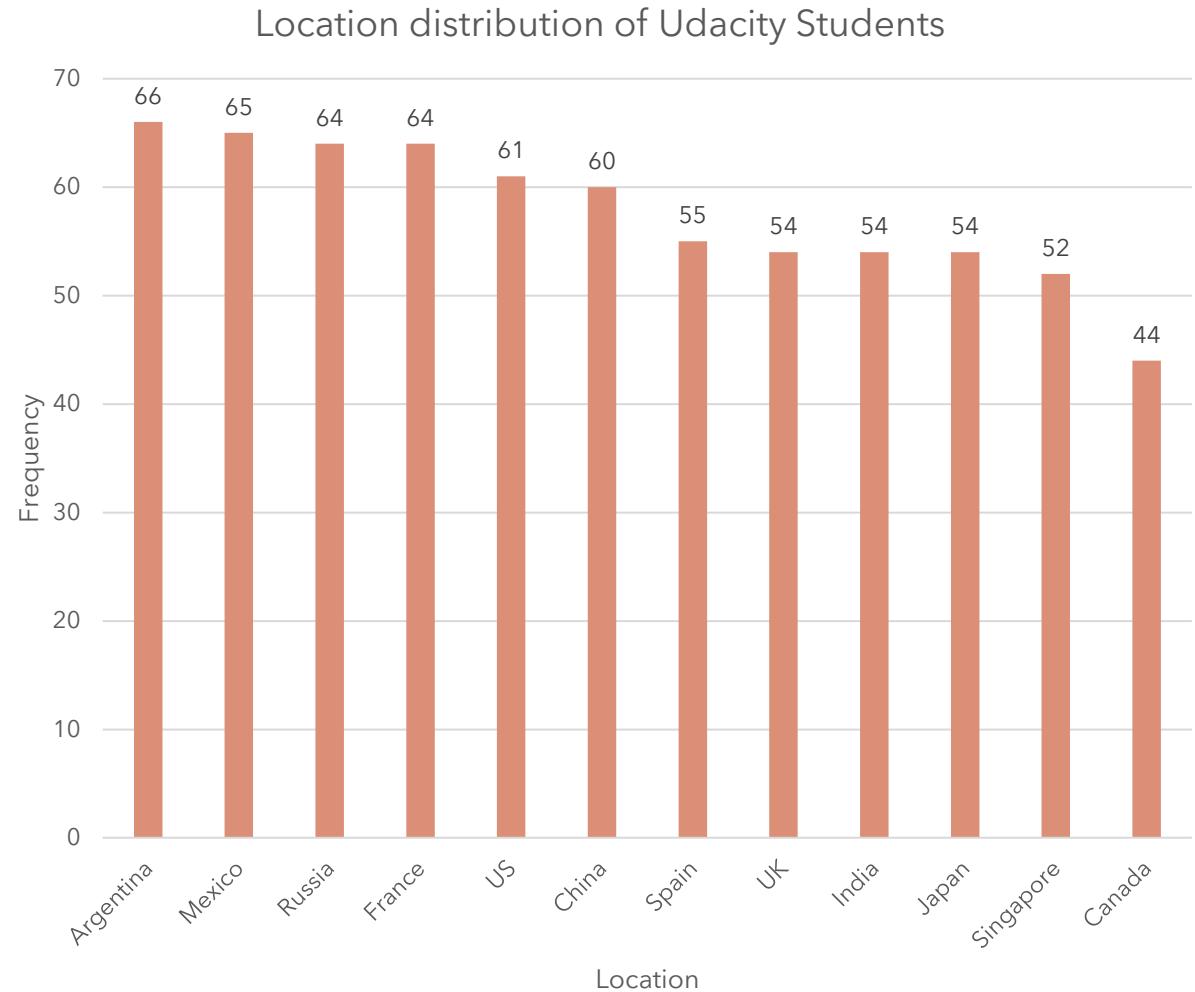
Research questions

- What ages are most representative of Udacity students?
- How was Udacity discovered by students?
- Which countries have the most students?
- Are there differences between countries?
- What are the Education Levels of students?
- Are there differences between Education Levels and studying hours?

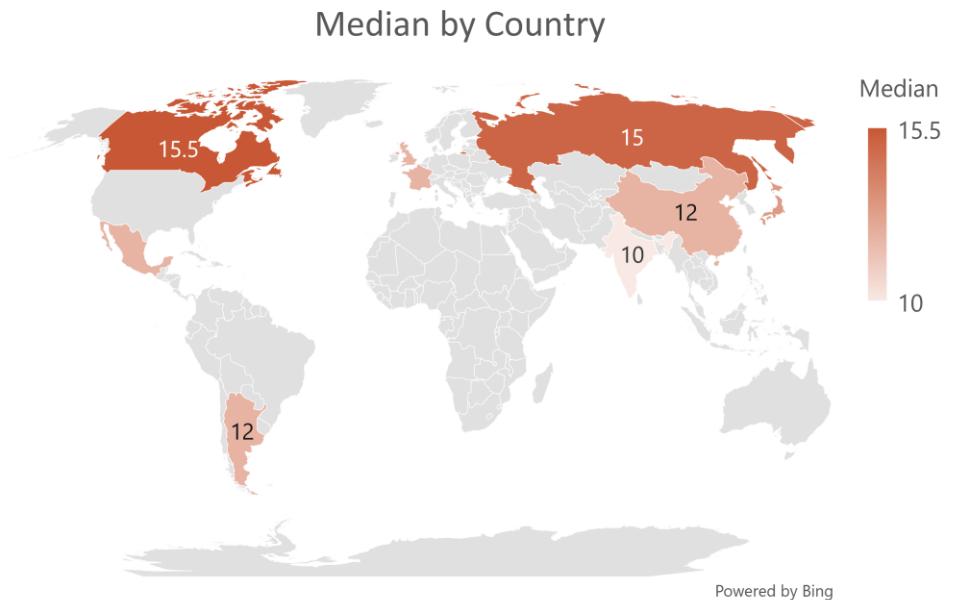
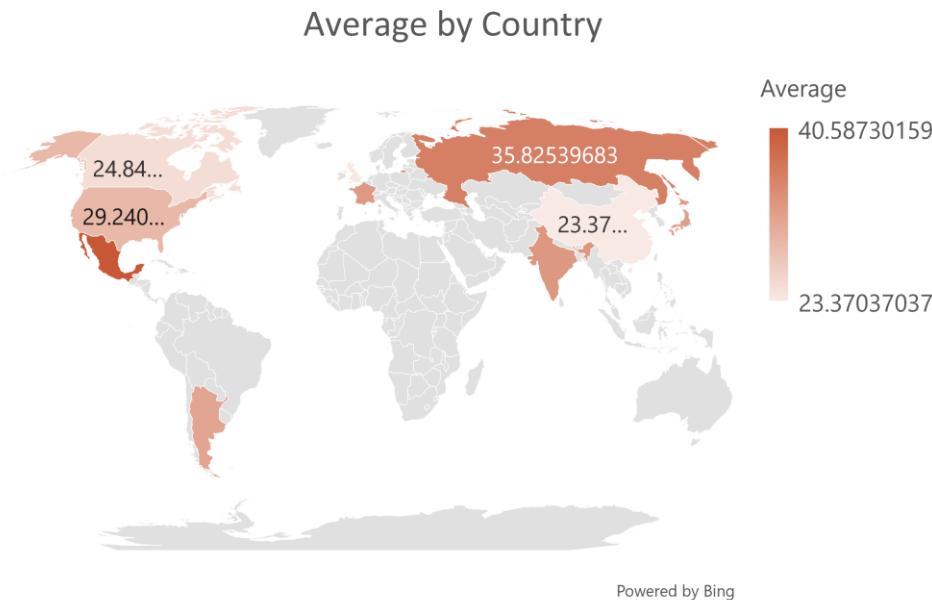
Pre-Processing

- Removed incorrect ages, such as 2, 3 and 120 year old students.
- Removed rows with invalid hours of sleep per night, such as 1, 4, 65 and 85. In this case, above 5 hours and below 14 were considered as valid answers.
- Respondents who work at Udacity removed.

- Argentina and Mexico are the locations with the highest number of students, while Canada is the one with the smallest.

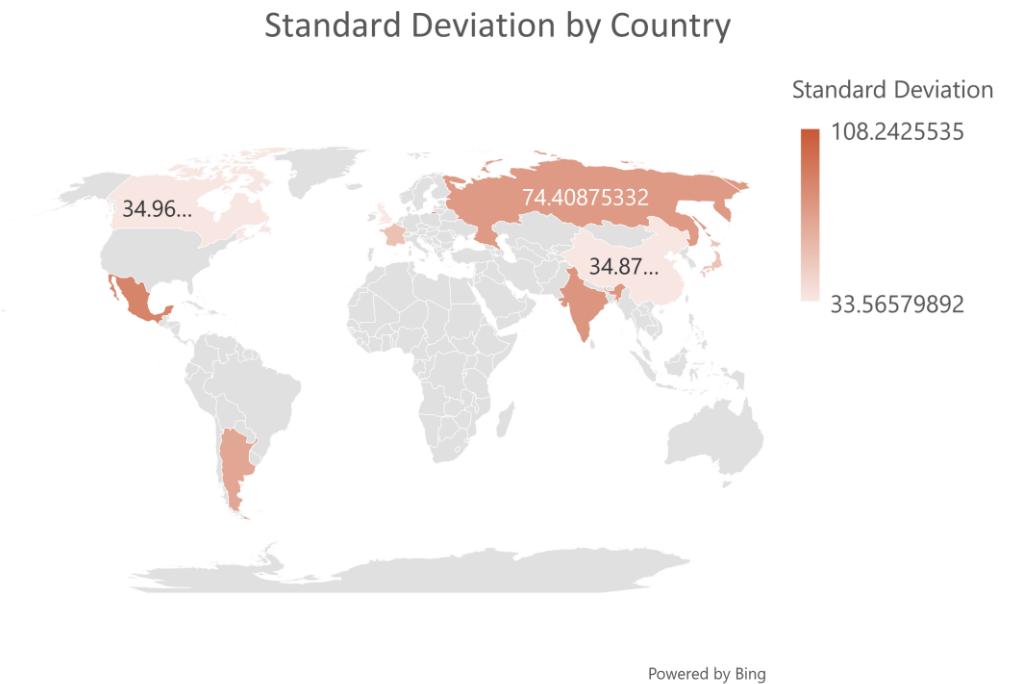
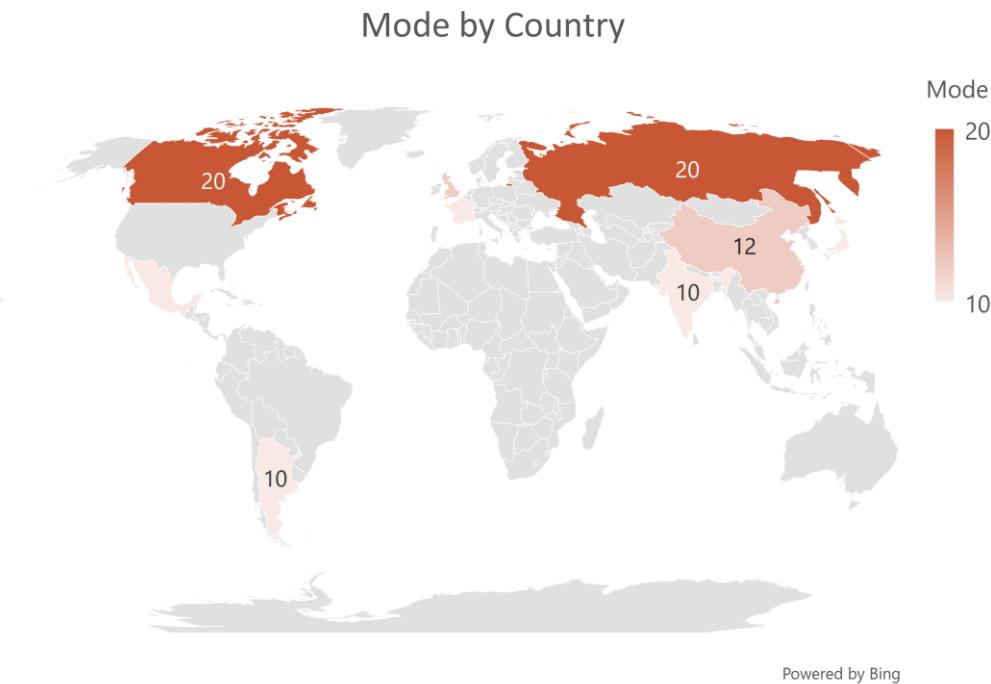


Hours to complete a project by Country



We can see that the Country which takes more hours to complete is Mexico (40.58), followed by Russia (35.82). Despite that, when placing the average and median side by side for these countries, we can clearly see that their medians are way smaller, 12 and 15 for Mexico and Russia respectively, suggesting that some extreme values are pulling the mean upwards.

Hours to complete a project by Country



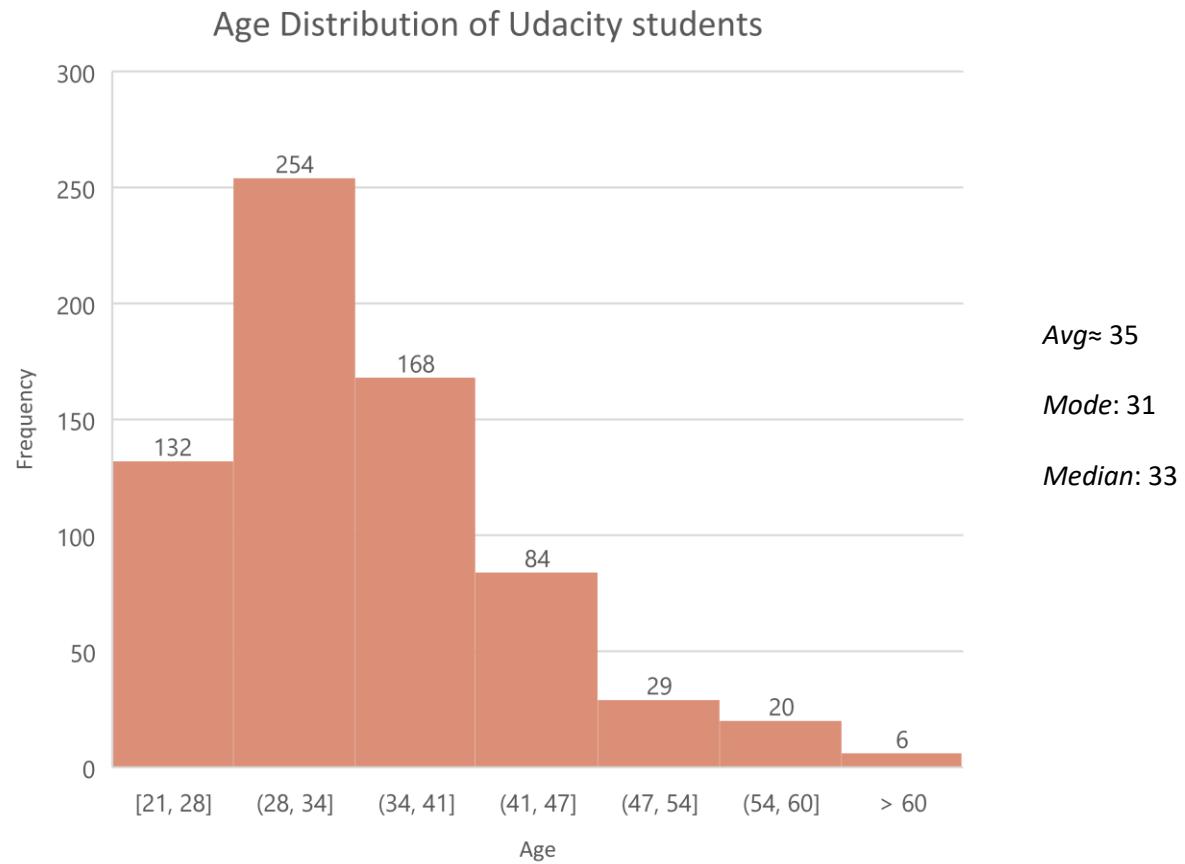
To complement the two graphs in the previous slide, we can take a look at our Standard Deviation, which confirmed our hypothesis that there are extreme values in our dataset, which are really big for Mexico and Russia (85.74 and 74.4).

Two countries that also has a very big standard deviation is Indian and Argentina.

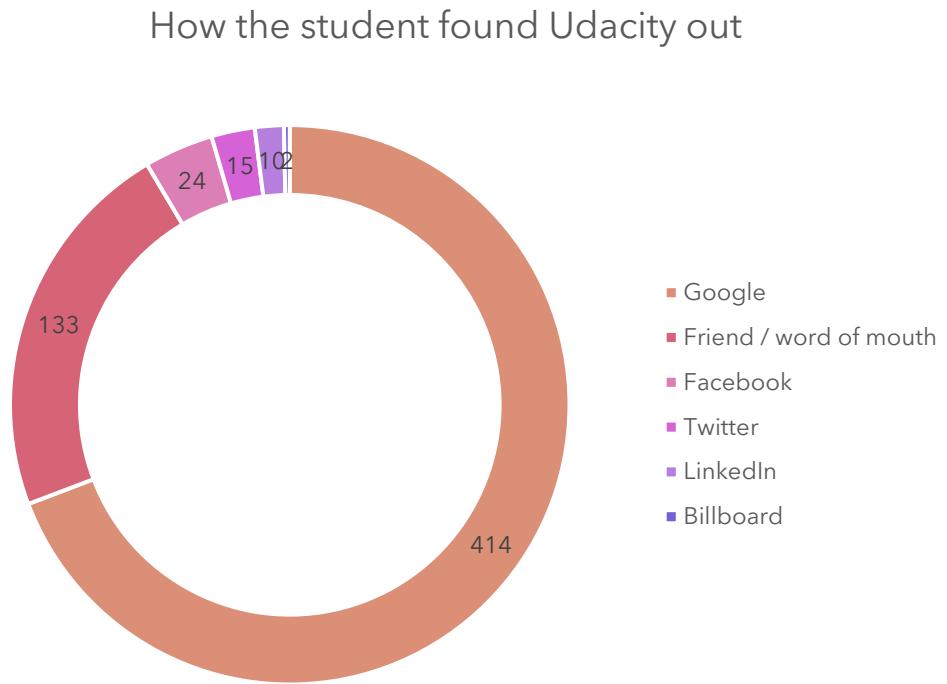
The two countries with the most repeated value were Canada and Russia (20).

Age Distribution of Udacity Students

- The average, mode and median are 35, 31 and 33, these indicate that the distribution is relatively young.
- The chart on the right shows a right skewed distribution, which, in this case, confirms that students tend to be on the younger side.
- The vast majority of students were younger than 47 years old, while most were between 27 and 34.

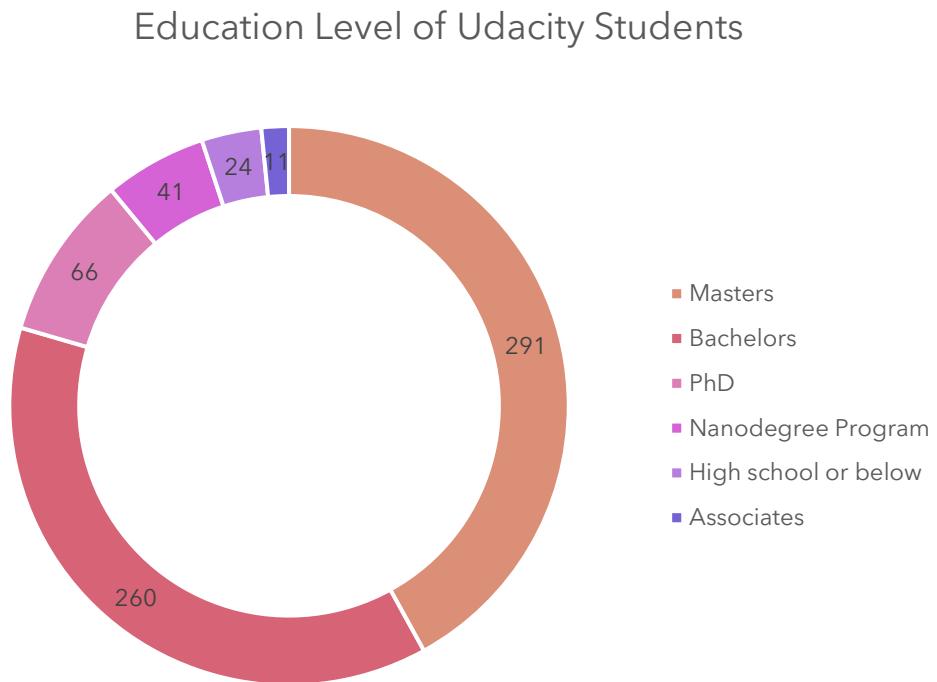


How was Udacity discovered?



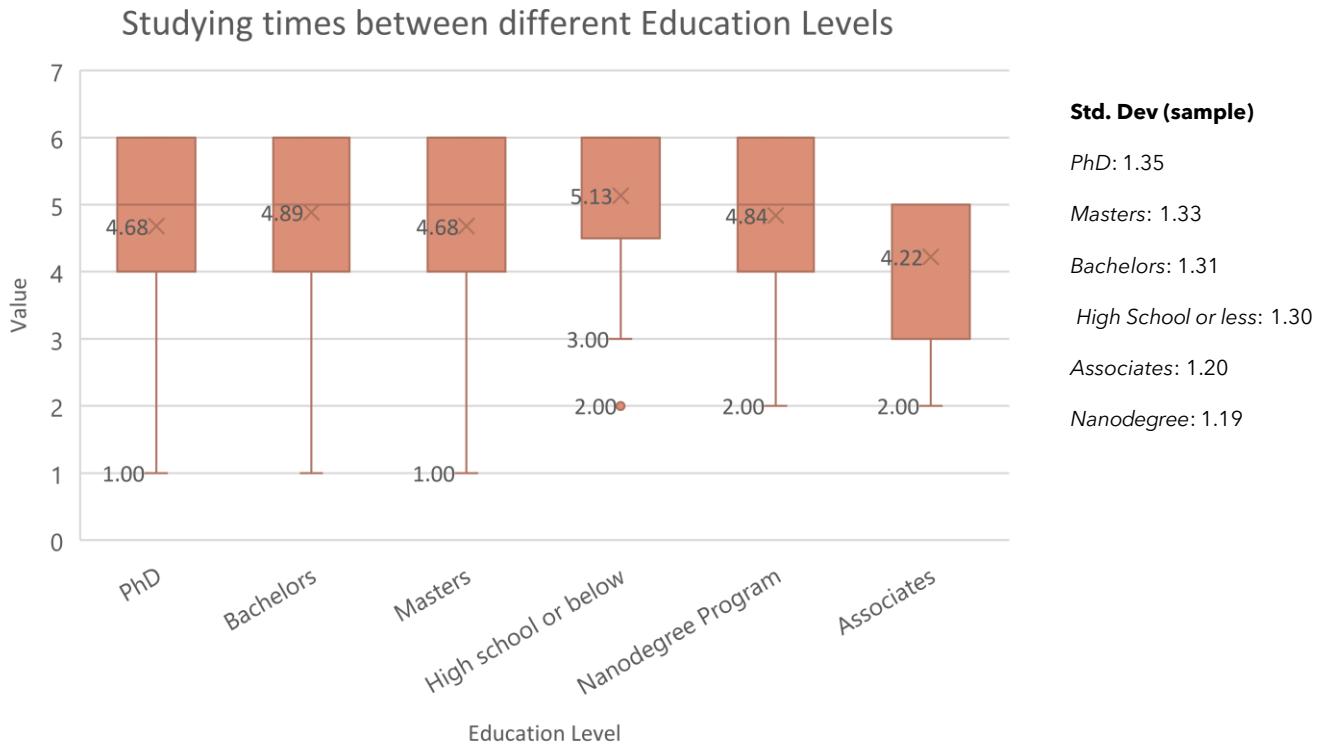
- Google was the bigger contributing factor to students finding out Udacity (414), followed by word of mouth (133).
- Other sources weren't as representative.

Education Level of Udacity Students



- Most Udacity students in the sample hold Master degrees (291), closely followed by Bachelors (260).
- The least education level is associates.

Differences in studying times between Education Level



- From these box and whisker plots, we can have a very complete picture to compare studying times among the different education levels.
 - From it, we can see that there seems to be little difference between PhD to Master holders, although students with High school degrees or below, seem to be studying, on average more hours (5.13).
 - Furthermore, High school or lower students, are reliably studying more hours, as can be seen by the smalled interquartile range, with a minimum value of 3, excluding an outlier of 2.
 - Associate degrees are studying less hours, with a maximum value of 5 and a minimum of 2.
 - Standard deviations range from 1.19 to 1.35. PhD's studying times hold the bigger standard deviation of 1.35, compared to nanodegrees of 1.19, which means that studying times for PhDs are more disperse than nanodegrees.
- Note: Blank studying times removed for this analysis



Limitations

The sample appears to be only of students who have completed a course or at least a project on Udacity, this doesn't allow us to have the bigger picture, as it is biased in that regard, since there could definitely be differences between course finishing students and non-finishing ones.

Sources

- Standard Deviation if:

<https://www.youtube.com/watch?v=d7HJqNbPBvg>