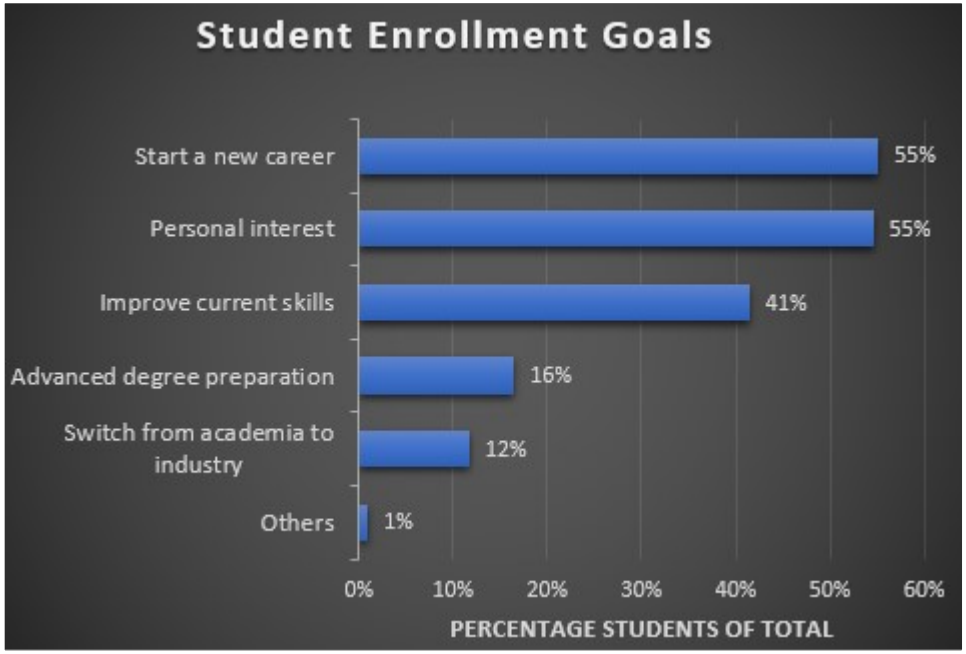
What are the top reasons for enrolling in Nanodegree Programs ?

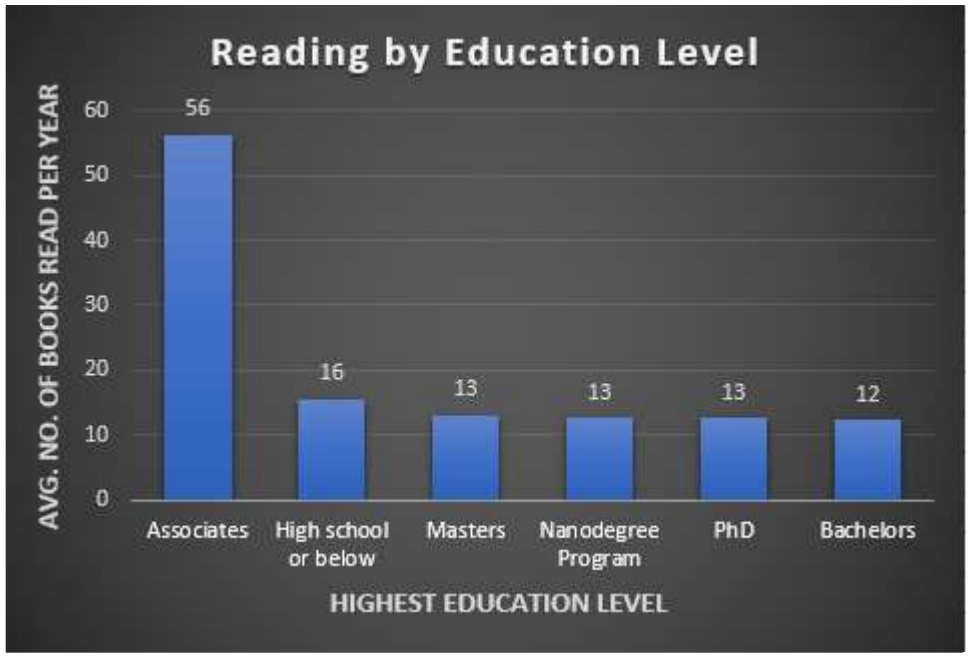


Here is the breakdown of reasons for joining the Udacity Nanodegree programs given by the surveyed 753 students.

Of the surveyed, 415 students, which is 55% of the total stated 'Starting a new career' as their main motive for joining. This is followed by 411 students having a 'Personal interest in the topic'. Then we have 41% i.e. 312 students wanting to 'Improve their skills in current role'.

We also got a sizable proportion of students joining the nanodegree programs to prepare themselves for advanced degrees (124 students – 16%) or to switch jobs from academia to industry (88 students – 12%).

Does the Avg. number of Books read vary based on Education level ?



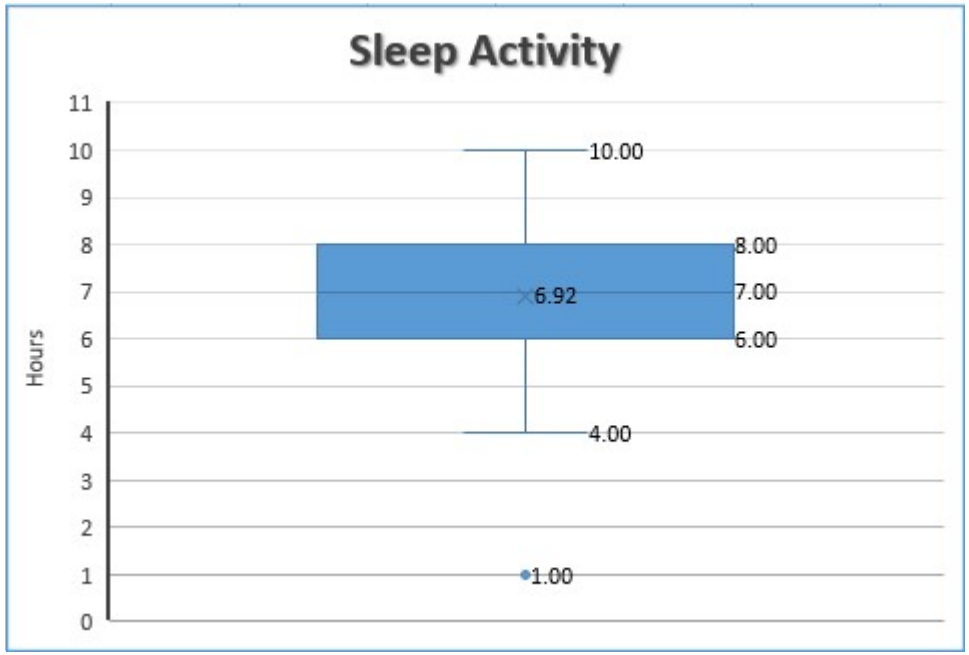
Given is the distribution of the avg. no of books reads by students across different education levels.

The overall mean of the distribution is about 13 books with a S.D of about 29 (we are rounding decimal places here). The variability is mainly because of some outlier values like 600 among the associates which clearly manifests in very high mean and deviation for the group. There are also some higher values like 300, 200 books read noted among the masters which explains for a higher deviation when compared to the rest.

If we get rid of the outliers, we then see the distribution across the groups to be closer to the overall mean of 12 books with the associates being on the lower side and the high schoolers on the upper side.

	Associates	Bachelors	High school or below	Masters	Nanodegree	PhD	Overall
StdDev	171	18	16	23	13	13	29
StdDev w/o outliers	10	18	16	13	13	13	15
Mean w/o outliers	7	12	16	12	13	13	12

Sleep activity of surveyed student population



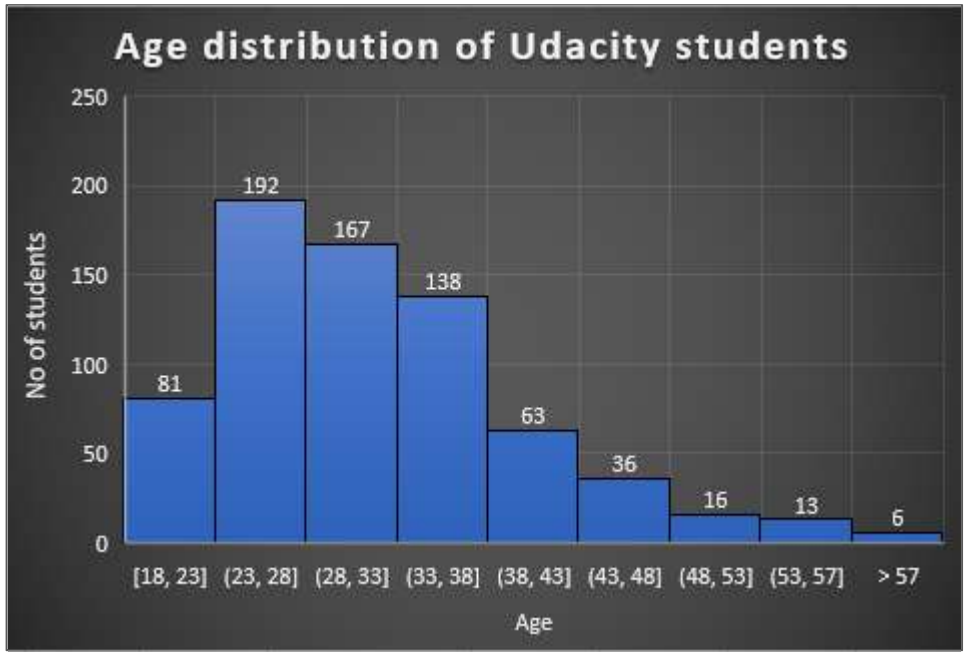
Here we try to understand the sleep activity of the students surveyed using a line histogram plot.

In the given distribution we have the median and mode centered at 7 hrs. The mean is at 6.92 hrs. and the standard deviation equal to 0.99 hr. The distribution is dense around the mean with over 90% of surveyed population sleep within the interquartile range of 6-8 hours

About 67% of distribution (502 students) are within one standard deviation of the mean i.e. in the range 6.92 +/- 0.99 hrs. and about 95% (715 students) of the surveyed student population sleep between 4.94 – 8.9 hrs., two standard deviation of the mean. We have an outlier of 1 hr. and a very few missing ones but they only account for 1% of the total population and hence have very minimal effect on the overall distribution.

Cleaning steps documented in excel

Distribution of surveyed student population by Age



Here is the histogram for age distribution of surveyed student population.

The distribution is slightly skewed to the right with mode < median < mean.

Over 50% of the surveyed student population fall within the interquartile range of 27-37 years, whereas over 90% fall within the range of 19-54 years.

There are more students enrolled under the age of 31 years, however the ages are more spread apart to the right from 31 to 77, which explains the variability (standard deviation of 8.37) we have.

Cleaning steps documented in excel