James Kulp
10/21/24

DS 3001 Visualization Assignment written answers

Question 1

2) The Data Dimensions are 63 columns by 3798 observations per the shape function. Some of the variables include The city and state where the college is located, public versus private status, etc, although many of the variables are cut off by ellipses when displaying the head function.

3) The majority of colleges in these data are four year institutions approximately 60% (2259 total). Among that group the most common control designation is private not-for-profit at 31% of all colleges on the list, slightly over half of all four year colleges (1180 total). Of two year institutions almost 2/3rds are public institutions, making approximately 41% of all the colleges in these data public institutions. Although in totality most public institutions is the largest category, they only constitute approximately 27% of 4 year institutions which is lower than I would've expected.

5) Both the kernel plots and the statistical descriptions show that the average 4 year graduation rate ~33.2% exceeding the 2 year average of ~21.%. The kernel plot further shows the distribution of graduation rates from these 2 year institutions is skewed further to the left (towards 0%), the statistical description shows the quartiles and 50% of the data is below 13% graduation rate, much lower than the more evenly distributed 4 year college graduation rate. A similar trend exists between the public and private designations, the public graduation rate is skewed to the left and much more centered around mean value of only 17% whereas private, either for profit or not, both have much more even distributions visible in the kernel density plots and the quartile statistics.

6) Via the graph the three kinds of institutions with higher mean graduation rates are 4 year private not-for-profit institutions, 2 year private for profit and not for profit institutions. The other three institution types have lower means, and 2 year public institutions have an especially skewed and mean centered distribution around 10%. Based on the distribution and its mean it appears that 4 year, private, not for profit institutions have the best graduation rates out of all the control/level groups. The 2 year private, for profit, institutions by graduation rate average appear to do well, but the distribution is much more spread out than that of the 4 year, private, not for profit group.

7) The peak for all institutions is approximately 8000 in aid, the average of means across all levels and control categories. 4 year institutions grant over twice as much as 2 year institutions which makes sense given the duration of the programs. Of these four year programs there seems to be two means one below 10,000 and one slightly above it. Private, not for profit

institutions give the most aid out of any control group the other two grant around half the amount.

8) There appears to be a correlation between aid and graduation rates especially demonstrated by four years, not for profit institutions where the two values increase together. This would make sense as a leading cause for students not finishing their degree would likely be cost concerns, therefore administering more aid to alleviate some of those issues would allow more students to stay in school as opposed to seeking employment. This is a causal claim though which isn't establishable simply off the graph's results, many other factors play into a student's decision to complete their degree or not. For example all of the two year programs seem to administer roughly the same amount of aid regardless of control, and yet their graduation rates are distributed very evenly in terms of the vertical axis of graduation rate.