

Lab 1

David Gerard

2019-05-01

Instructions

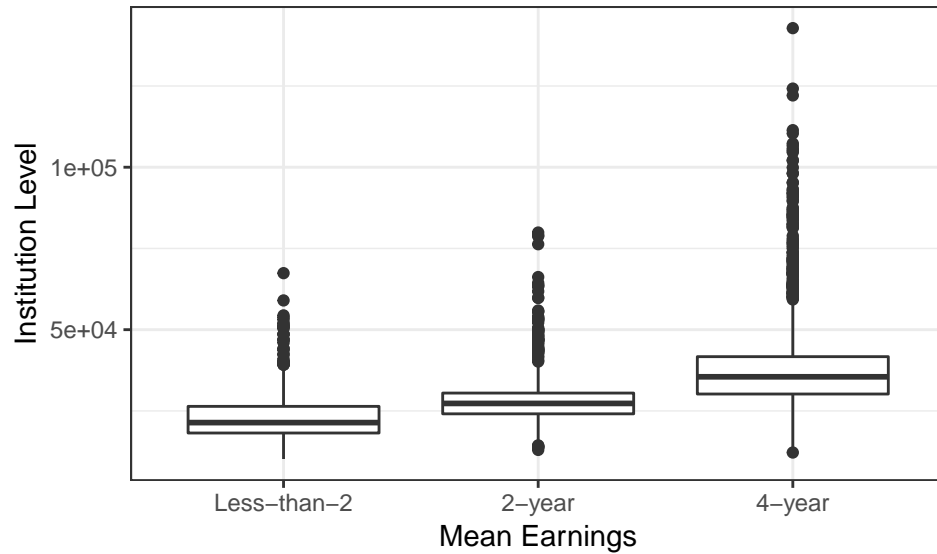
- Make sure to commit after you answer each question.
- Push to GitHub when you are done.
- Learning Objectives
 - Review material from [STAT 412/612](#).
 - Practice committing code to a git repo and pushing to GitHub.

College Scorecard

The data in “college_score.csv” contains a subset of the variables found in the 2016 to 2017 [College Scorecard](#) database. These data contain information on colleges in the United States. The variables included are:

- UNITID and OPEID: Identifiers for the colleges.
- INSTNM: Institution name
- ADM_RATE: The Admission Rate.
- SAT_AVE: Average SAT equivalent score of students admitted.
- UGDS: Enrollment of undergraduate certificate/degree-seeking students
- COSTT4_A: Average cost of attendance (academic year institutions)
- AVGFACSAL: Average faculty salary
- GRAD_DEBT_MDN: The median debt for students who have completed
- AGE_ENTRY: Average age of entry
- ICLEVEL: Level of institution (1 = 4-year, 2 = 2-year, 3 = less than 2-year).
- MN_EARN_WNE_P6: Mean earnings of students working and not enrolled 6 years after entry (so students who graduated in the 2009 to 2010 academic year).

1. Use relative paths to load in the data frame from college_score.csv.
2. If you used the default settings for reading in the data, then some of your variables are probably characters when they should be numerics. Which ones? Why were they read in as characters? Fix this.
3. Does average faculty salary seem to be associated the mean earnings of students six years after graduating? Justify with an appropriate plot.
4. Does the level of the institution seems to be associated with the mean earnings of students six years after graduating? Reproduce this plot in R that explores this relationship (Hint: use factors):



5. Is the length of the name of an institution (in number of **characters**) associated with the mean earnings six years after graduation? Justify with an appropriate plot.
6. Is the length of the name of an institution (in number of **words**) associated with the mean earnings six years after graduation? Justify with an appropriate plot. Make sure that you appropriately account for overplotting if this is an issue.
7. What are the shortest- and longest-named institutions (in terms of number of characters)? Try doing this in one `filter()` call.
8. Devise your own ranking of universities. Which ones are the best under your ranking?