

Assignment 5

EDH7916 | Spring 2020

Benjamin Skinner

NOTE This assignment needs to be completed by the start of the next class. That means everything pushed to your remote GitHub repo before class starts.

Remember, I encourage you to save your work, commit smaller changes, and push to your remote GitHub repo often rather than wait until the last minute.

Use the the IPEDS data sets we used in class, `hd2007.csv`, joined with a new IPEDS data set, `ic2007mission.csv`, to answer the questions. You will need to join them. You may also need to look up the data dictionaries for each file. Click the “continue” button on this page to see the data and accompanying dictionary files.

You **do not** need to save the final output as a data file: just having the final result print to the console is fine. For each question, I would like you to try to pipe all the commands together. Throughout, you **should** account for missing values to the best of your ability by dropping them.

For each question, show your data work and then answer the question in a short (1-2 sentence(s)) comment.

Questions

NB You will need to join the two IPEDS data sets to answer these questions using the common `unitid` key. Note that column names in `hd2007.csv` are uppercase (`UNITID`) while those in `ic2007mission.csv` are lowercase (`unitid`). There are a few ways to join when the keys don’t exactly match.

One is to set all column names to the same case. If you want to use `left_join()` starting with `hd2007.csv`, you can first use the the `{dplyr}` verb `rename_all(tolower)` in your chain to lower all column names.

See the help file for `left_join()` for other ways to join by different variable names.

1. How many chief administrator names start with “Dr.”?
NB Many chief administrators are listed on more than one line due to branch campuses. Make sure to take this into account by keeping only distinct names.
2. **BONUS** How many chief administrator names end with the title “PH.D.” or some variant?
3. Among those schools that give their mission statement:
 1. How many repeat their institutional name in their mission statement?
 2. How many use the word *civic*?
 3. Which top 3 states have the most schools with mission statements that use the word *future*?
 4. Which type of schools (public, private-non-profit, private-for-profit) are most likely to use the word *skill* in their mission statement?
4. Among the schools that closed in 2007 or 2008 and give a date with at least a month and year:
 1. Which has been closed for the longest time? How many months has it been from its close date to the beginning of this current month (1 February 2020)?
 2. How many days were there between the first school to close and the last?