

STAT 140: Homework 1

YOUR NAME HERE

Due Date: Wednesday, 9/11/2019

PROJECT

Identify a research question that you are interested in investigating, and a corresponding data set that can be used to investigate your question. You will be exploring the question and data throughout the semester, and you will use them for your final paper. If you are having trouble finding data, possible resources include, but are not limited to:

- <http://www.data.gov>
- <https://github.com/fivethirtyeight/data>
- <https://www.cdc.gov/nchs>
- <http://jse.amstat.org>
- <http://data.un.org/>

If you find data as part of a research article, please formulate your own question to investigate that is different from that used in the article.

If you work in a research lab and collect data yourself, you may use that data for your project, as long as you have permission from your professor (for that lab).

Come see me if you are having problems identifying a research question/data set.

Questions

- 1) What is your research question?
- 2) What is the population of interest?
- 3) Where did you find your data?
- 4) Describe how your data arose.
- 5) How were the data collected?
- 6) Who are the subjects in this study?
- 7) How many individuals (subjects) are included (use R for this part)? Hint: consider the functions we used to explore data in our second class meeting.
- 8) What variables are present in your data set?
- 9) Is there any missing data? If so, do you know why it is missing? Hint: missing data will often be read in as NA, but sometimes people use other ways to denote that an observation is missing (e.g. any dots (.) where you don't expect them?)