

STAT 231: Problem Set 7A

Sean Wei

due by 5 PM on Monday, October 26

In order to most effectively digest the textbook chapter readings – and the new R commands each presents – series A homework assignments are designed to encourage you to read the textbook chapters actively and in line with the textbook’s Prop Tip of page 33:

“Pro Tip: If you want to learn how to use a particular command, we highly recommend running the example code on your own”

A more thorough reading and light practice of the textbook chapter prior to class allows us to dive quicker and deeper into the topics and commands during class. Furthermore, learning a programming language is like learning any other language – practice, practice, practice is the key to fluency. By having two assignments each week, I hope to encourage practice throughout the week. A little coding each day will take you a long way!

Series A assignments are intended to be completed individually. While most of our work in this class will be collaborative, it is important each individual completes the active readings. The problems should be straightforward based on the textbook readings, but if you have any questions, feel free to ask me!

Steps to proceed:

1. In RStudio, go to File > Open Project, navigate to the folder with the course-content repo, select the course-content project (course-content.Rproj), and click "Open"
2. Pull the course-content repo (e.g. using the blue-ish down arrow in the Git tab in upper right window)
3. Copy ps7A.Rmd from the course repo to your repo (see page 6 of the GitHub Classroom Guide for Stat231 if needed)
4. Close the course-content repo project in RStudio
5. Open YOUR repo project in RStudio
6. In the ps7A.Rmd file in YOUR repo, replace "YOUR NAME HERE" with your name
7. Add in your responses, committing and pushing to YOUR repo in appropriate places along the way
8. Run "Knit PDF"
9. Upload the pdf to Gradescope. Don’t forget to select which of your pages are associated with each problem. *You will not get credit for work on unassigned pages (e.g., if you only selected the first page but your solution spans two pages, you would lose points for any part on the second page that the grader can’t see).*

1. “Tell the truth. Don’t steal. Don’t harm innocent people.”

In the textbook, the authors state, “Common sense is a good starting point for evaluating the ethics of a situation. Tell the truth. Don’t steal. Don’t harm innocent people. But, professional ethics also require a neutral, unemotional, and informed assessment.”

(1a) Assuming the numbers reported in Figure 6.1 are correct (truthful), do you think Figure 6.1 is an *unethical* representation of the data presented? Why or why not?

ANSWER: Yes, this graph seems like an unethical representation of the data presented. This is because on first glance, it seems like the graph conveys the fact that murders committed using firearms after the enactment of the “Stand Your Ground” law in 2005 has decreased. However, on closer inspection the y-axis is flipped - in this unconventional case, up corresponds to decreasing values. The real story behind this visualization is that the number of deaths has actually increased since 2005. Thus, this graph is misleading by trying to convey an idea that is not clear on first glance.

(1b) Pulling from the examples in the textbook, provide one example of a more nuanced ethical situation (one that you perhaps found surprising or hadn’t considered before).

ANSWER: I was pretty surprised by the ethical situation regarding Vioxx in Section 6.3.5. Although there were studies that were positive for Vioxx, the abstract reports show a clear risk after its approval. From my understanding, I thought all possible scenarios would be explored to provide insights about any negative information about the drug before it was approved. The shocking thing for me was that even after this report came out, the drug wasn’t withdrawn for 3 more years. I was extremely shocked by the fact that around 20 million Americans were able to take this drug before it got withdrawn, especially since the poor statistic seemed pretty clear in my opinion.

2. Does publishing a flawed analysis raise ethical questions?

In the course so far, we've touched upon some of the ethical considerations discussed in this chapter, including ethical acquisition of data (e.g., abiding by the scraping rules of a given website) and reproducibility. At the end of Section 6.3.4 (the "Reproducible spreadsheet analysis" example), the authors ask: Does publishing a flawed analysis raise ethical questions?

After reading Section 6.4.1 ("Applying the precepts") for the "Reproducible spreadsheet analysis" example, re-consider that question: Does publishing a flawed analysis raise ethical questions? And, a follow-up question for consideration: Does it depend on who published the flawed analysis (e.g., a trained data scientist? an economist who conducts data science work? a psychologist who works with data? a clinician who dabbles in data science?)

In 4-6 sentences, respond to those questions and explain your response.

ANSWER: Although it is not an ethical obligation to reach 100% correct research results, I still believe that publishing flawed analyses raise ethical questions. The textbook states "The obligation is to do everything feasible to ensure that the conclusions faithfully reflect the data and the theoretical framework in which the data are analyzed" (page 138). The flawed analysis in the "Reproducible spreadsheet analysis" example violates this claim because selectively including data and oddly weighting summary statistics does not "faithfully reflect the data." Thus, I believe that tampering with the data to reach a conclusion is not ethical, as "researchers should be truthful in their reporting of research" (page 138). Even though people may come from different backgrounds, this should not excuse flawed analyses because of their obligation to be truthful with regards to reporting data. No matter one's background in data science, there is a universal professional standard of data science that should be met when reporting findings.