PSYC3830: Data tutorial assignment

This assignment is worth 25 points.

In order to apply what you've learned about the internet as a rich source of psychology data, **your midterm assignment will be to construct a simple, step-by-step tutorial to present to our class on how to download data from one data source, and some of the things you can do with that data once you have it in hand.** What makes this a tutorial is that your document will include both code and explanations of what the code does. These tutorials will also live on the syllabus website and can be accessed after the class ends so you can conduct psychology research online on your own. You can find examples in the *Resources* section of this document.

Note about coding: Students often feel nervous about coding because it can feel radically different from the skills required in other courses. However, psychology research is increasingly only possible with some coding knowledge, which is why it is included in this course. *If you don't know how to code, or don't feel confident in your coding skills, that is okay.* You will not be graded on your ability to code in this course, but rather, your ability to construct a tutorial. I am available to help with all coding questions/struggles. We will also devote some time in class to basic coding principles and troubleshooting strategies.

Data tutorial assignment points breakdown

- Rough draft with pseudocode: 10 points, due 10/3
- Data tutorial: 10 points, due 10/17
- Presentation: 5 points, in class on 10/17

Getting started

- First, decide whether you want to work alone or with a partner. You may choose your partner or have me pair you with another student who is also interested in working with a partner. If you complete it with a partner, you will do the entire assignment together, as a unit (i.e., joint submissions, joint presentation).
- Then, decide on your data source. You can find a list of pre-approved data sources in the Resources section, along with resources for learning how to use code to interact with that type of data. You may use a different data source than the ones listed; just check that it's okay with me first.
- Finally, decide on a coding language for your tutorial. Do you want to use Python or R? Usually, there are packages available for both languages, although there tends to be more available in Python.

Rough draft with pseudocode (10 points, due by end of day on October 3rd)
As a first step in this assignment, you will hand in an outline/rough draft of your tutorial, with pseudocode. Pseudocode is a term used to describe something in between code and a

regular description of what you're doing. It's a good first step when coding. Your rough draft should adhere to the following:

- It must be in the correct type of document (R Markdown for R, Jupyter Notebook for Python)
- It should describe, step by step, what you're planning to demonstrate with your tutorial
- Your pseudocode will include examples of functions you may use
- You will include coding questions for me or things you are struggling to figure out, if any

I will provide you with feedback on your rough draft to ensure you are headed in the correct direction. I will also answer any questions that you include.

Data tutorial (10 points, due by class time on October 17th)

This is the crux of the assignment. You should have a complete tutorial that takes 5-10 minutes to complete. The tutorial should demonstrate how to access one type of data that can be found on the internet, show what types of information can be accessed, and describe some examples of ways the data can be used. The tutorial should include a mix of code and plain text, and it should be self-contained and self-explanatory. Below, you can find some examples of submissions that would receive different grades:

- A 10-point draft is in the correct type of document (R Markdown for R, Jupyter Notebook for Python) and has the steps of the tutorial clearly laid out. The tutorial has a clear flow/arc that mixes plain-text instructions/descriptions with code that demonstrates the tutorial concepts. The tutorial builds on itself, and each chunk of code is clearly explained. The tutorial can be understood on its own, simply by reading it and running the code.
- A 7-point draft is in the correct type of document and is the correct length. The narrative of the tutorial is not clear at times. It relies too heavily on either plain-text or code, and does not mix them together smoothly and evenly. Some code chunks aren't clearly explained, or some plain-text descriptions are not demonstrated with code. The tutorial occasionally doesn't make sense without external explanation.
- A 4-point draft isn't easily accessible and/or doesn't take 5-10 minutes to complete. The tutorial does not build on itself and is difficult to follow. It would not necessarily be useful to someone hoping to learn how to access this type of data. Code is not clearly explained and/or does not run correctly.

Presentation (5 points, in class on October 17th)

In class, you will walk your peers through the tutorial you've created, explaining it as you go. The purpose of this part of the assignment is to give you experience teaching others about coding, while also providing your peers the ability to work with a wide variety of data sources online. Remember that this is a presentation; don't merely read what you've written, but explain it and clarify it if people are confused. Go slowly, and check in with the rest of the class (who are temporarily your students) as you guide them through the tutorial.

Resources

Examples

- [Rough draft example]
- [R Tutorial example], [Python tutorial example]

General coding resources

- [Our class tutorials]

Pre-approved data sources

For each data source, I've included links to some Python/R libraries for working with that type of data. Many of the libraries come with their own detailed documentation and tutorials. You'll likely find these useful when building your own tutorial.

- HTML-based websites:
  - Python: Beautiful Soup
  - R: rvest
- Facebook
  - Python: Python-facebook, Facebook SDK
- Google Books
  - Python: Google-books-api-wrapper
  - R: ngramr
- Reddit
  - Python: PRAW
  - R: RedditExtractoR
- Tik Tok
  - Python: TikTok-API
  - R: traktok
- Wikipedia
  - Python: Wikipedia, Wikipedia-API
  - R: wikipediatrend documentation, getwiki documentation
- X/Twitter
  - Python: Tweepy, snscrape (works for other social media too)
- News
  - Python: GoogleNews, newsapi-python, GNews
  - R: tidyrss, rvest