Welcome to the Putting Things on Your Computer part of data analysis!

I'm Alyssa Smith, a PhD student at the Network Science Institute, and I'll be teaching you a bit about downloading & processing social media data. In order to help the workshop run effectively, we'll try to get the preparation work done ahead of time. I've put together these instructions in the hopes that everything will be mostly set up when you come to class, and we'll have to do minimal debugging the day of. The guides I've linked here will be a lot more helpful at explaining how to get things set up than I will be on my own, but they still might not cover every scenario.

If you run into trouble with any of these steps, I encourage you to (1) Google the error messages / problem you're having and (2) if you are still stuck after 10-15 minutes of Googling, email me at smith.alyss@northeastern.edu. I tend to be very responsive with email. If you find issues with any part of this document, contact me so I can fix it! Feel free to add any helpful tips or tricks you find to this document as well.

Without further ado, here's what you should aim to accomplish before class on Tuesday 2/8/2022.

1. **Check if you have Python on your computer; if you don't, install Python.** Python is a programming language that is great for collecting & working with data. Sometimes Python is already on your computer (more likely if you're using a Mac); other times you'll have to install it. This guide explains what's going on and goes through your options better than I likely can. Here's another guide that is a little bit less bare-bones and perhaps more user-friendly.
2. **Install pip, Python's package manager.** Packages are like plug-ins that help you do things that Python can't do on its own. You can think of them kind of like browser extensions, I think. Here's a guide for MacOS, and here's one for Windows.
3. **Install the following packages using pip (**in order to install **PACKAGE_NAME** using pip, you'll type **pip install PACKAGE_NAME** into your terminal, then press enter.)**:**
    a. Matplotlib ("pip install matplotlib")
    b. Pandas ("pip install pandas")
    c. Tweepy ("pip install tweepy")
    d. Jupyter ("pip install jupyter")
    e. Numpy ("pip install numpy")
4. **Get developer credentials from Twitter (optional if you only care about Reddit data).** Twitter has a great guide to doing this here. Developer credentials let you access Twitter through computer programs so that Twitter knows who you are and roughly why you're accessing Twitter with a computer program. You're going to need a **bearer token**, and you're going to need to keep it somewhere safe so you can use it in class.
5. **Make sure jupyter works.** Type "jupyter notebook" in your terminal, then press ENTER. You should see a whole bunch of stuff scrolling, then your browser should launch. If this doesn't happen, let me know. Jupyter is a way to run small chunks of code at a time in a

notebook-like configuration. That way you can display things and make mistakes without too much trouble.

6. **Download the jupyter notebook (class.ipynb) attached to this email.** Make sure you can navigate to it when the jupyter notebook server is running (as in step 5). It will likely be in your Downloads folder. Test it out if you'd like! To run a cell of code, press shift-enter.

Again, if you're having any issues and Google can't solve them, it is totally okay to ask for help! A lot of this is not intuitive and can be difficult depending on the idiosyncrasies of how your computer is set up, so feel free to reach out with questions/concerns/etc.

Also, if you're thinking about the ethical ramifications of any research you're going to be doing, a colleague recommended this book as a great resource. We'll discuss the ethical issues around social media research a bit in class, but it's always good to have a reference as well!