Econ 5253 - Spring 2023 **Problem Set 6** **Due: Mar. 21 beginning of class**

This problem set will give you practice with cleaning and visualizing data.

As with the previous problem sets, you will submit this problem set by pushing the document to *your* (private) fork of the class repository. You will put this and all other problem sets in the path `/DScourseS23/ProblemSets/PS6/` and name the file `PS6_LastName.*`. Your OSCER home directory and GitHub repository should be perfectly in sync, such that I should be able to find these materials by looking in either place. Your directory should contain at least three files:

- `PS6_LastName.R` (you can also do this in Python or Julia if you prefer)
- `PS6_LastName.tex`
- `PS6_LastName.pdf`
- `PS6a_LastName.png`
- `PS6b_LastName.png`
- `PS6c_LastName.png`

1. Type `git pull origin master` from your OSCER `DScourseS23` folder to make sure your OSCER folder is synchronized with your GitHub repository.
2. Synchronize your fork with the class repository by doing a `git fetch upstream` and then merging the resulting branch.
3. Find some data that interests you and clean it. (This could be data you scraped from PS5, or it could be a different data set that you intend to use for your final project; see also: Kaggle datasets). Note also that cleaning the data is not a requirement; I just do not expect you to come across data that requires no cleaning whatsoever. In your .tex file, tell me about the steps you took to clean and transform the data.

   Write code in R, Python, or Julia to execute the steps described in your .tex file. Your submitted script should be completely reproducible by me.
4. Using R, Python, or Julia, create three visualizations of the data you used in the previous question. Some things to keep in mind as you create these visualizations:
   - The visualization should inform the viewer
   - The code used to generate the visualization should be as readable as possible
   - The visualization should look as sleek as possible

Submit your visualizations as .png files (see problem set instructions above) and also insert them as figures into your .tex file. Consult the help resources on `overleaf` for pointers on how to insert an image into a .tex file.

5. Include in your .tex file an explanation of what the images are communicating. How are they helpful for understanding your data set?

6. Compile your .tex file, download the PDF and .tex file, and transfer it to your cloned repository on OSCER using your SFTP client of choice (or via `scp` from your laptop terminal). You may also copy and paste your .tex file from your browser directly into your terminal via `nano` if you prefer, but you will need to use SFTP or `scp` to transer the PDF.[1]

7. You should turn in the following files: .tex, .pdf, .png (3 of them), and any additional scripts required to reproduce your work. Make sure that these files each have the correct naming convention (see top of this problem set for directions) and are located in the correct directory (i.e. `~/DScourseS23/ProblemSets/PS6`).

8. Synchronize your local git repository (in your OSCER home directory) with your GitHub fork by using the commands in Problem Set 2 (i.e. `git add`, `git commit -m "message"`, and `git push origin master`). More simply, you may also just go to your fork on GitHub and click the button that says "Fetch upstream." Then make sure to pull any changes to your local copy of the fork. Once you have done this, issue a `git pull` from the location of your other local git repository (e.g. on your personal computer). Verify that the PS6 files appear in the appropriate place in your other local repository.

[1] If you want to try out something different, you can compile your .tex file on OSCER by typing `pdflatex myfile.tex` at the command prompt of the appropriate directory. This will create the PDF directly on OSCER, removing the requirement to use SFTP or `scp` to move the file over.