Friday, 09/08 Activities to be completed before coming to class

Task 1: Python and notebooks

- 1. Make sure that you have a functioning version of Canopy on your laptop, run it to see that it's working properly.
- 2. Learn how to use Package Installer from Canopy to either install or update the following packages: jupyter, pandas, matplotlib, plotly. Alternatively, you can use pip or easy_install.
- 3. Refresh your memory about how to run Jupyter notebooks. In CS111 we have shown how they can be opened from within Canopy, but it is possible to start them from the command line as well. Here is <u>a short video</u> that has very useful information about Jupyter notebooks (only 7 minutes, and if you run it with 1.5 speed even shorter).
- 4. Create a simple Notebook to solve the Class Activity from Tuesday (using Python to calculate the descriptive statistics of a simple list). Try both the pandas code (once you have installed pandas), as well as pure Python code, as we started writing in class.

Collaboration Policy: Feel free to create a study group and do these things together with your peers. However, everyone should work on their computer and only discuss issues in English, without sharing code. The best way to learn is to do the tasks yourself.

Help: I have office hours on Friday evening, 5:30-7:00, feel free to stop by (add your name in <u>my</u> <u>calendar's appointment slots</u>). Alternatively, there will be help room on Sunday and Monday evening 7-9pm in E160A.

Expectation of completion: Give it your best shot to complete all / most of these tasks before class. However, if you get stuck, it's okay to continue working on them even after Friday.

Task 2: Reading about Data Science (circa 15 minutes) [before class]

Read pages 1-5 (pg. 5 inclusive) from the paper <u>"Tidy Data"</u> by Hadley Wickham (who has created many important R tools). Though we will not use R in this class, the concepts of Tidy Data are reflected in the Python package "pandas" that we'll be using.

Task 3: Thinking / Writing [before class]

After completing the reading about Tidy Data, take a piece of paper and start planning the variables that you would like to use to describe the observations of your data collection. Think about the missing data as well: the times when you were supposed to take a photo, but you didn't, they also need to be part of the dataset. Also think about how to organize the data, do you need a single table, more than one table? Once you have made these decisions, take another piece of paper and start drawing the table(s), and put some concrete values in them for 2-3 rows, from the data you have collected so far. This task will benefit by being done independently, in order to avoid "group think".

Bring your pieces of paper in class and we'll have a conversation together to decide a common structure for our table(s), so that we can start filling them out (next week).