SI 506 Section 6

All files you need are available on our Canvas site, Files > Section Resources > Section 006.

EXERCISE 1 (~ 40 minutes)

Step 1: Create a folder and a text file

- Create a new folder called `506section7` on your computer.
- Save the provided file `alice.txt` inside that folder -- this file contains text of the first 3 chapters from Alice in Wonderland

Step 2: Plan & Write code!

- In the same folder (`506section7`), create a new file called `section7.py` where you will write a Python program as follows.
- Write a program that opens `alice.txt` and creates a dictionary containing the number of occurrences of each character in `alice.txt`
  - Note: convert all characters to lowercase for simplicity
- Print out the number of occurrences of the ten most frequently-occurring characters.

- Your output should resemble this (the most frequent character is the empty space):

```
  4412
e 2058
t 1582
o 1323
a 1290
h 1139
n 1094
s 1067
i 1025
r 844
```

Hint: Recall file opening, iteration over lines/characters, dictionary accumulation, sorting.

Step 3: (Challenge/Optional) Visualize!

- Choose the top 10 frequently occurring characters and print a visualization representing the relative frequency of their occurrence.
- The most frequent character should always have 20 pound signs (#).
- Your output should resemble this:

```
  : ####################
e : #########
t : #######
```

```
o : #####
a : #####
h : #####
n : ####
s : ####
i : ####
r : ###
```

Hints:

- To compute the number of pound sign characters for a character `ch`, divide its frequency of occurrence by the frequency of the most-frequent character, and multiply the result by 20 (the maximum number of pound signs we want to display).
- You can invoke the int() function to convert floating point values to the number of pound sign characters (#) you need to print.

EXERCISE 2 (~ 20 minutes)

Step 1: Download and open

- We will work once again with a dictionary of states and populations, based on data available on Wikipedia.
- Download the Python file with the states dictionary from canvas:

SI 506 Canvas > Files > Section Resources > Section 006 > `states_dictionary.py`

Step 2: Write a program

- Write a program in that file that sorts the states and territories by population
    - **Good idea:** Write comments to plan it, like you've seen in lecture, and then translate them to code bit by bit, trying out your code repeatedly to see if it does what you expect!
- The program should then print out the 10 states with the highest population, each on one line.

Step 3: Extra challenge! (If time)

- Using the same logic as in step 3 of the previous exercise, create a visualization representing the relative frequency of the population of the 10 most highly populated states.
- In other words, estimate the relative frequency of each state's population. Remember, the most populated state should have 20 pound signs (#).
- This is the expected output:

```
California : ####################
```

```
Texas : ##############
Florida : ##########
New York : ##########
Pennsylvania : ######
Illinois : ######
Ohio : #####
Georgia : #####
North Carolina : #####
Michigan : #####
```