CST 383: Intro to Data Science
Based on Dr. Glenn Bruns

Reading: Data Science Handbook, chapter 2

1. Assume that we have a set of data, all of which are the same type, such as integers or strings. How would you store it for better efficiency: a) Use a Python list for flexibility b) Use a NumPy array?

2. The code below is based on population data from 1939 from Wikipedia. Which country is stored in "x" ? (capitalization not important)

```
country = np.array(["British Empire", "Japan", "Soviet Union",
                    "United States", "Germany", "Italy",
                    "Brazil", "Yugoslavia"])
pop = np.array([550_398_825, 304_119_000, 168_500_000,
                147_050_800, 86_755_281, 56_532_624,
                40_289_000, 15_490_000])
#trick: use _ to read the values easier
x = country[pop < 20_000_000]
```

Hint: https://jakevdp.github.io/PythonDataScienceHandbook/02.06-boolean-arrays-and-masks.html

3. Suppose x is a numeric array of length 10. If you evaluate expression `x < 0`, what is the length of the result?

Hint: https://jakevdp.github.io/PythonDataScienceHandbook/02.06-boolean-arrays-and-masks.html

4. According to our book, which of the following is the default sort algorithm for NumPy? a) mergesort b) Quicksort?

5. If I want to combine the results of two mask arrays (logic values), would you use the operators a) `&` and `|` or b) `and` and `or`?

Hint: https://jakevdp.github.io/PythonDataScienceHandbook/02.06-boolean-arrays-and-masks.html#Aside:-Using-the-Keywords-and/or-Versus-the-Operators-&/|