

Homework 2

Present all Python code and all visuals in a word (or pdf) file.

1. Dejunkify the bar plot created by the code

```
import matplotlib as mpl
mpl.get_backend()
import matplotlib.pyplot as plt
import numpy as np

plt.figure()

major = ['Business', 'Health professions', 'Social sciences', 'Engineering',
'Biological', 'Psychology']
graduates = [386201, 244909, 159967, 121956, 118663, 116432]

plt.bar(major, graduates)

plt.title('Top Six Bachelor Degrees awarded in 2017-2018', alpha=0.8)
plt.show()
```

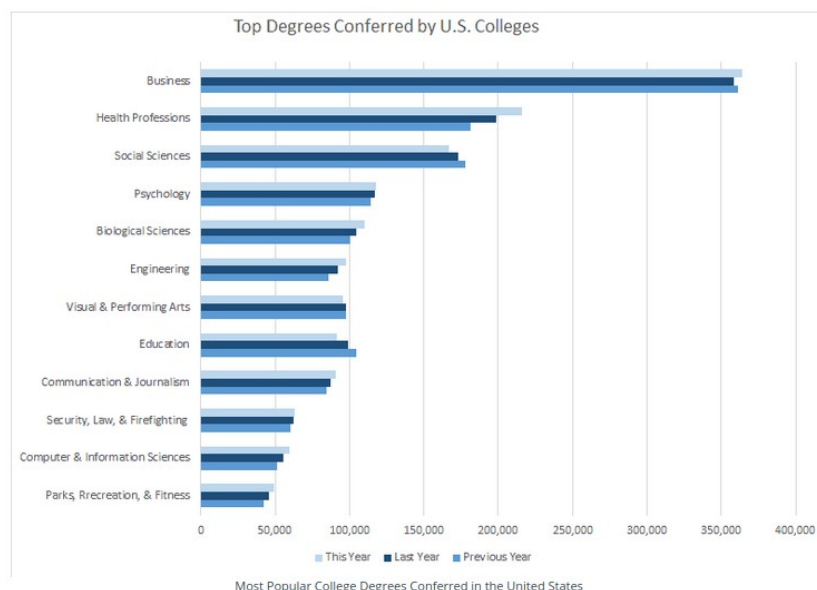
source of data: https://nces.ed.gov/programs/digest/d19/tables/dt19_322.10.asp

How to dejunkify

- Remove the Box
- Fix the labels
- Remove the tick marks on the x-axis and the y-axis
- Remove the y-axis labels
- Add the number of graduates to the bars (as in the Darkhorse Analytics dejunkifying example),
- have the bars lightly shaded and emphasize one major your choice.

Make sure to turn in code and your new plot. Also include a description of the functions and methods you used to create the plot. I would like you to be able to teach your technique to the class well enough so that the rest of the class can use your technique.

2. Create a bar plot in Python for the last three years (2015-16, 2016-17, 2017-18) for the most popular eight majors using the same data that looks like (which is for different years)



Data source: https://nces.ed.gov/programs/digest/d19/tables/dt19_322.10.asp

Barplot example source: <https://www.matchcollege.com/top-majors>

3. Look at the data in the file `daylight.csv`. The data in the file looks like this

```
5,1,5,9:22:20,9.3722
10,1,10,9:27:50,9.4639
```

and you can read it as

The fifth day of the year, Jan 5, there is 9.3722 hours of daylight

The tenth day of the year, Jan 5, there is 9.4639 hours of daylight

The file `alldaylight.csv` has every day, but you don't need that.

Graph the data very “nicely.” I would like to see a clear presentation of the data

Extra credit: Can you find a function to approximate this? A linear regression could work, but you are not using a line. It is a sine function with minimum on Dec 21.

Data for daylight is from the website <https://www.timeanddate.com/sun/usa/new-york?month=5&year=2020> accessed on Jun 17th.