

DS2500

1/17-Fri

Admin

- no lab/Off on Monday
- Hw 1 out now, due 1/24 at 9pm
 - autograded vs. viz/code
 - late policy (78 hours)

Agenda

1. Summary/Catch-up
2. CSV Files + objects
3. Python

1. Summary/Catch-up

any library (ex. matplotlib)

- once per project - **install**
"Python Packages", search for lib

- once per .py file - **import**

ex. `import matplotlib.pyplot as plt`

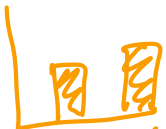
error - "hey I've never heard of that module :)"

- installation
- typo?

data viz library

→ what do I want to create?

Bar Plot



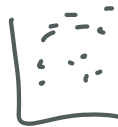
- compare things/cats
(not too many bars)

Line chart



- change over time

Scatterplot

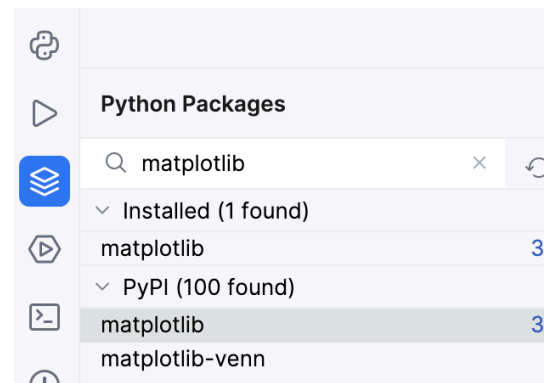


- understand the
rel b/w 2
variables

Histogram



- distribution
of data



F-strings (ex) printing

no f-string
print("arg is", x)
 └── literal ─┘ └── variable ─┘

f-string
print(f"arg is {x}")

literal "un ~"
15
20.5

variable x = ~
name = ~

f" literal {var} money \${var}"

(ex) plot title

no f-string
plt.title("Year" + str(yr))

f string
plt.title(f"Year {yr}")

optional parameters

→ normally...

def spawn(x):

spawn(4)

optional param:

- may or may not be provided when calling
- default values if not provided

(ex)

```
plt.bar(x, y)
plt.bar(x, y, color="pink")
plt.bar(x, y, color="pink", w=.5)
plt.bar(x, y, w=.5)
```

def bar(x, y, color='blue', w=1):

~~~~~

→ caller can provide color/width or else, defaults are used

## 2. CSV Files + objects

→ today's file PwHL Data (sample)

| city | name | wins | losses |
|------|------|------|--------|
| —    | —    | —    | —      |
| —    | —    | —    | —      |
| ...  |      |      |        |

pwhl.csv

comma-sep values (flat file)

goal for today: create objects out of the file

Source of data / Ethical concerns:

- source: wikipedia
- correctness? not sure!
- any personal info?  
no!
- data collected w/ consent?  
yes!
- who is not repped?  
PwHL is new (rep of pro athletes?)  
non-binary athletes?  
players of color?

CSV File → objects

### (1) Read the CSV File

- open the file
- read each line of file
- reader object (variable contents of file)
- variable (list) to store data in file
- csv library
- what do we do w/ header?

data is a 2D list of strings  
csvr is a file object

```
{ ["boston", "8", "9"],  
  ["ny", "5", "12"] }
```

```
import csv → string  
  
def read_csv(filename):  
    data = []  
    with open(filename, 'r') as inf:  
        csvr = csv.reader(inf)  
        for row in csvr:  
            data.append(row)  
  
    return data
```

### 3. Python

→ goal: read in csv file  
each row into an object

```
obj = HockeyTeam("booster", 8, 9)
```

Read in read csv function (separate file)

- optional param: # lines to skip  
default = 1

→ data in a 2D list of strings (list)

for row in list: — iterate by value

→ row = ['boos', 'teet', '8', '9']

one row → one object

row → object, overall a list of objects

When to convert w, l from string to int?

— after read, before call constructor

— in constructor

— on the fly, before we do math/plotting