DS2500
2/11-Tues.

Admin

- exam Fri 2/14
- mini pressos [slides due 2/23 9pm]
- semester project [proposal due 2/28 9pm]

Agenda

1. Math rand-up
2. Data structures + mutability
3. Giving a Talk
4. Exam prep

1. Math Rand-up

Distance/Similarity

	?	Quant/Qual
• Euclidean	• two points on a plane • default distance	Quant
• Hamming	• order matters • where differ?	Qual
• Jaccard	• sets	Qual
• Haversine	• actual distance on earth	Quant (lat/long)
• Manhattan	• grid distance • high dimensionality	Quant

	?	range?
correlation	• relationship b/w 2 vars • r-value	-1 to +1
normalization	• features on different scales	0 to 1 min max

• all features can't the same

[?]

variance — σ^2 — [comparison]

std deviation — $\sigma = \sqrt{variance}$ — avg difference from mean [one dataset alone]

2. Data structures + mutability

- tuples ( )
  - ↳ like a list, but immutable

when do we use tuples?
- return multiple things from a function — return x,y
- protect our data
- dictionary items
- easily swap values — a,b = b,a

- sets { }
  - ↳ same as in math
    - order doesn't matter
    - no dupes

s = {1,2,4} ✓
s.add(6) ✓
s.remove(1) ✓
s[0] = -8 !!
s = {{1,2}, {1}} !!
s = {(1,2), (1)} ✓
s = {1,2,2,2,4,2} ✓
↳ {1,2,4}
s = {[1,2], [2]} !!

Mutability
↳ immutable: cannot be modified once created

Data structures we know:

strings — immutable!
lists — mutable!
dictionaries ↳ mutable, but keys must be immutable
tuples — immutable!
sets — mutable, but items must be immutable

```
def clean_currency(lst):
    ''' given a list that includes currency and
        numbers in the form $x,xxx, clean them up and convert
        to float
    '''
    for i in range(len(lst)):
        lst[i] = lst[i].replace("$", "")
        lst[i] = lst[i].replace(",", "")
        lst[i] = float(lst[i])
```

from last week!
$xx,xxx.xx
into
xxxxx.xx

↳ didn't return any thing!
modifies the given list in the function

lst = ['~', '~', '~'] (strings)
clean_currency(lst)
↳ lst = [~, ~, ~] (floats)

same as...
lst = ['~', '~']
lst2 = lst
any change to lst2 affects lst and vice-versa

in memory:
lst → ['~', '~']
lst2 ↗

4. Exam Prep

Topics: up through 2/7

- everything from class/lab/hw
- Python gen'l → multiple choice, fill in the blank, what does this code do?, write some code
  - classes + objects
  - for loops (by pos, by val), range
  - functions (optional params)
  - mathematical operations
  - reading csv
  - data types
  - lists (10, 20)
  - dictionaries

- Errors + Unit testing
  - identify error types
  - example input, expected output
  - pytest / assert
  - can't == w/ floats

3.000000001 == 3.0

- Distance / similarity
  - compute distance between 2 things
  - what type makes sense?
  - types of data

- Data Science math
  - normalization
  - correlation
  - variance
  - line of best fit (→ prediction model)
  - MSE
  - std dev
  - mean, median, mode, range

$$\frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Logistics

- in class on Fri
- on paper
- 8.5x11 inch cheat (one side)
- no other material / device >
- entire lecture period
- scan, on gradescope
- DAS accommodation?