DS2000

9/20 - Fri :)

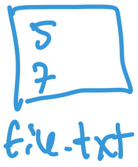
Admin

- HW1 due 9pm
- HW2 due 9/27 9pm

Agenda

1. Data File by time
2. Bigger data file
3. Python

0. Review

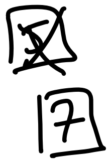


```
with open (____, "r") as infile:
    num = ____ .readline()
```

→

```
with open("file.txt", "r") as infile:
    num = int(infile.readline())
    num = int(infile.readline())
```

num ⇝



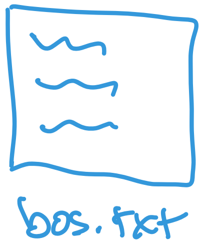
5 is gone :(

1. Data File by time

Structure of any DS program

1. gather data
2. computations
3. communication

last time



one salary
all same year

today

2015
2016
2017
asst.txt

avg salary for that year

What keep?
What else?

Data →

min max range :)

mean ?

median ?

std dev

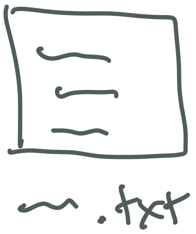
under/over threshold :)
↳ poverty
livable wage

by year

increase per year
%, $

2. Bigger Data Files

last time: 3 lines in file



```
with/open/as:
    sal_one = ~
    sal_two = ~
    sal_three = ~
```

↳ good for when...

- few lines in file
- know how many lines

today: 15 lines in file

- create 15 variables?!
- write 15 nearly-identical lines of code?!

↳ to the rescue: for loop!

- lots of lines
- don't know how many

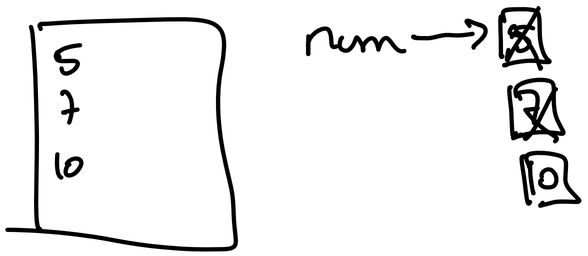
"for" ~ repeat some code
(in execution, not writing)

```
with open(filename, "r") as inf:
    for line in inf:
        num = float(line)
```

shortcut for:

```
with open(filename, "r") as inf:
    num = float(inf.readline()) ✓
    num = float(inf.readline()) ✓
    num = float(inf.readline()) ✓
    ⋮
```

con:
↳ num gets overwritten
(no sep variables)


5
7
10
num
5
7
10