

DS2000
1/17 (Tue)

Admin

- HW1 due Fri 4pm
- Quiz 1 4:30 today - 9:50 AM Fri

Agenda

1. Data Sources
2. File Processing
3. Python

Recap ~ Result?

7 // 2 3

7 / 2 3.5

7 % 2 1

6 % 2 0

8 % 2 0

print(2 / 3) 0.66666666

print(round(2 / 3, 2))

0.67

//, % usually ints
ok with floats

7.0 // 2.5 2

1. Data Sources

1. gather data
2. computation
3. communication

from user (input)
from a file
↳ `...txt`

DS Integrity

- How was the data collected?
- Do we have confidence in accuracy?
- What can I share?

Today's data:

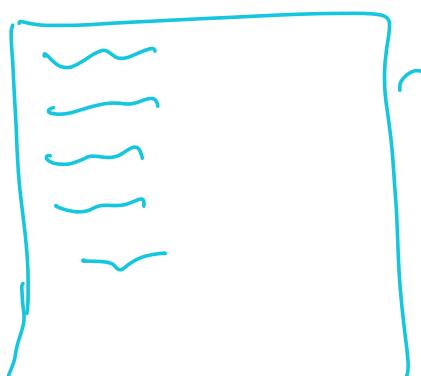
- Salaries of Boston employees
- source: `data.boston.gov`

Salary, overtime, title, name

↳ don't share

2. File Processing

Our file



`boston-salaries.txt`

→ one number per line
Salary from 2021

Q: what do we want
to know?

- differential in salaries (highest vs. 2nd highest) ...
- avg
 - min
 - max
 - upper/lower quartile
- median
- range
- std dev
- range avg/vars
- mode

In Python: no input!

- name of file (constant)
- open the file

variable!

with open ("~", "r") as infile:

↳ Python vocab ↳ read (string)
 ↳ name of file (string)

- read a line (one at a time)

x = infile.readline()

↳ x is variable (string) ↳ read one line

x = int(infile.readline(1))

↳ x is an int

- Save .txt file in same folder as .py file