Nearest Neighbor Classification

with almost no background

Nate Derbinsky

+

Laney Strange

Northeastern University
Khoury College of Computer and Information Sciences

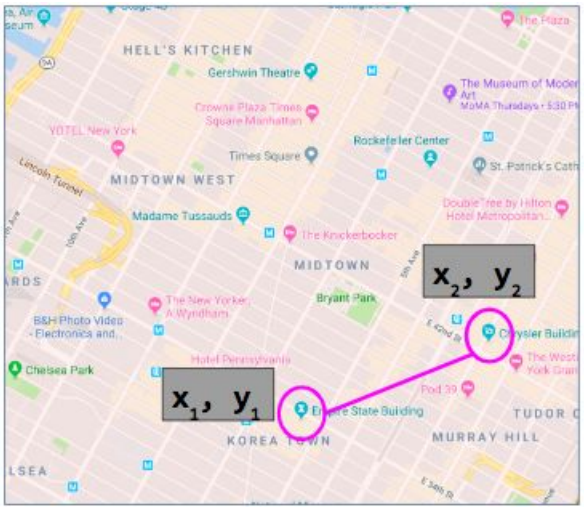
The Assignment

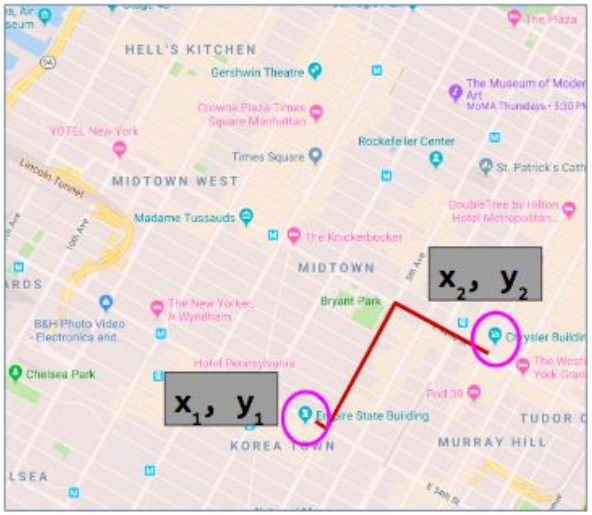
`Nearest Neighbors`

The Assignment

Why did Netflix tell me to watch Bodyguard?

The Assignment






```
Write distance functions
```

The Assignment

Write distance functions

Euclidean, Manhattan, Hamming

The Assignment

```
[ [42.35, -71.18], [42.33, -73.09], [42.57, -71.88],
  [43.33, -71.29], [44.34, -72.07], [41.88, -70.01],
  [41.41, -71.11], [40.72, -73.99], [40.72, -73.98] ]
```

Given a community of data points...

The Assignment

```
[ [42.35, -71.18], [42.33, -73.09], [42.57, -71.88],
  [43.33, -71.29], [44.34, -72.07], [41.88, -70.01],
  [41.41, -71.11], [40.72, -73.99], [40.72, -73.98] ]
```

[42.03, -71.99]?

...find the Nearest Neighbor of a new point

Materials

- Homework specification
- Unit tests
- Sample solution
- Introductory slides

Learners

- CS1 students
- Adults at a workshop

Learners

- CS1 students
- Adults at a workshop


Women's
community of
code

Goals of the Workshop

1. Exposure to Machine Learning

Goals of the Workshop

1. Exposure to Machine Learning
2. Write a program!

Goals of the Workshop

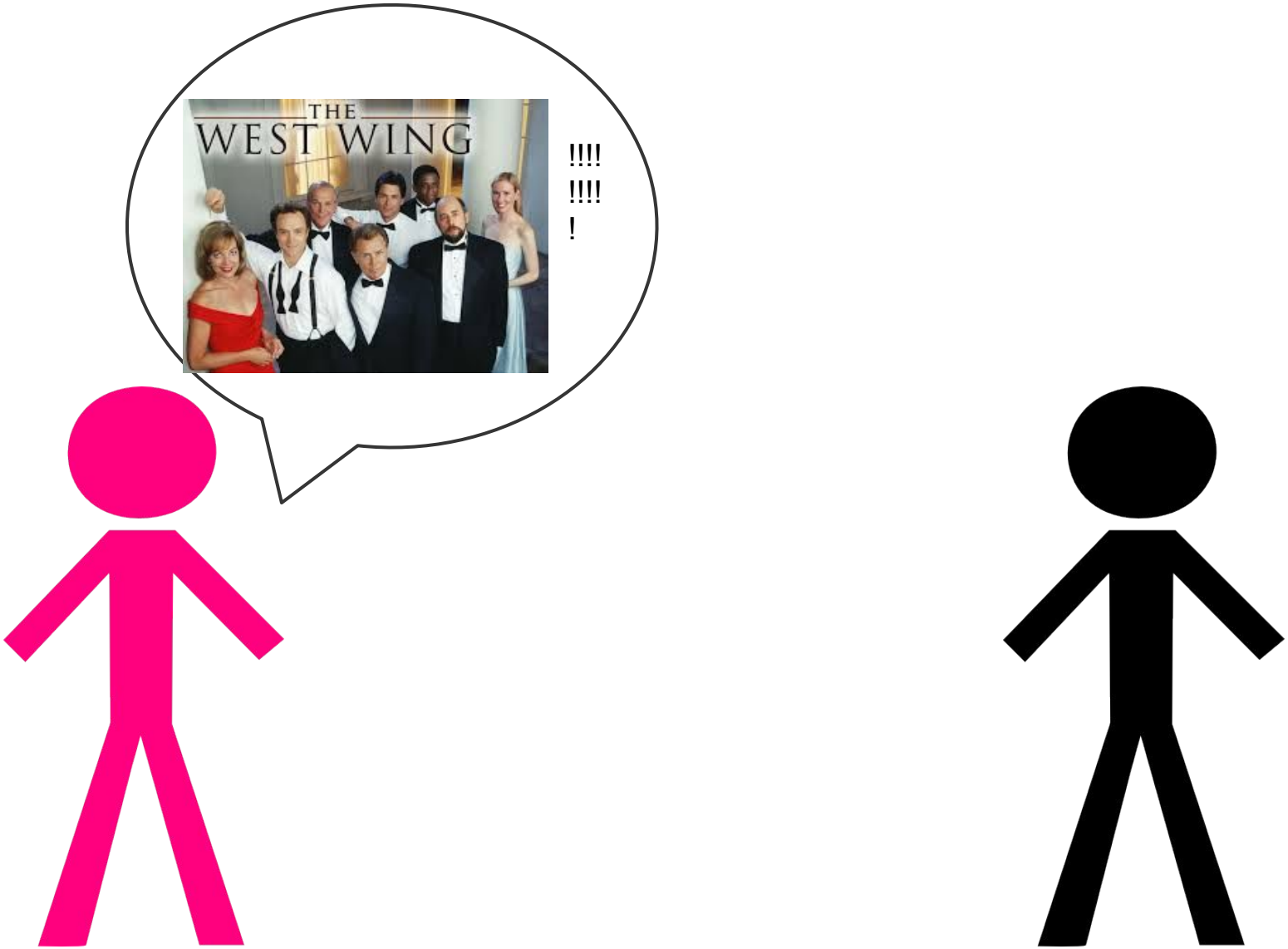
1. Exposure to Machine Learning
2. Write a program!
3. Recognize NN from “real life”

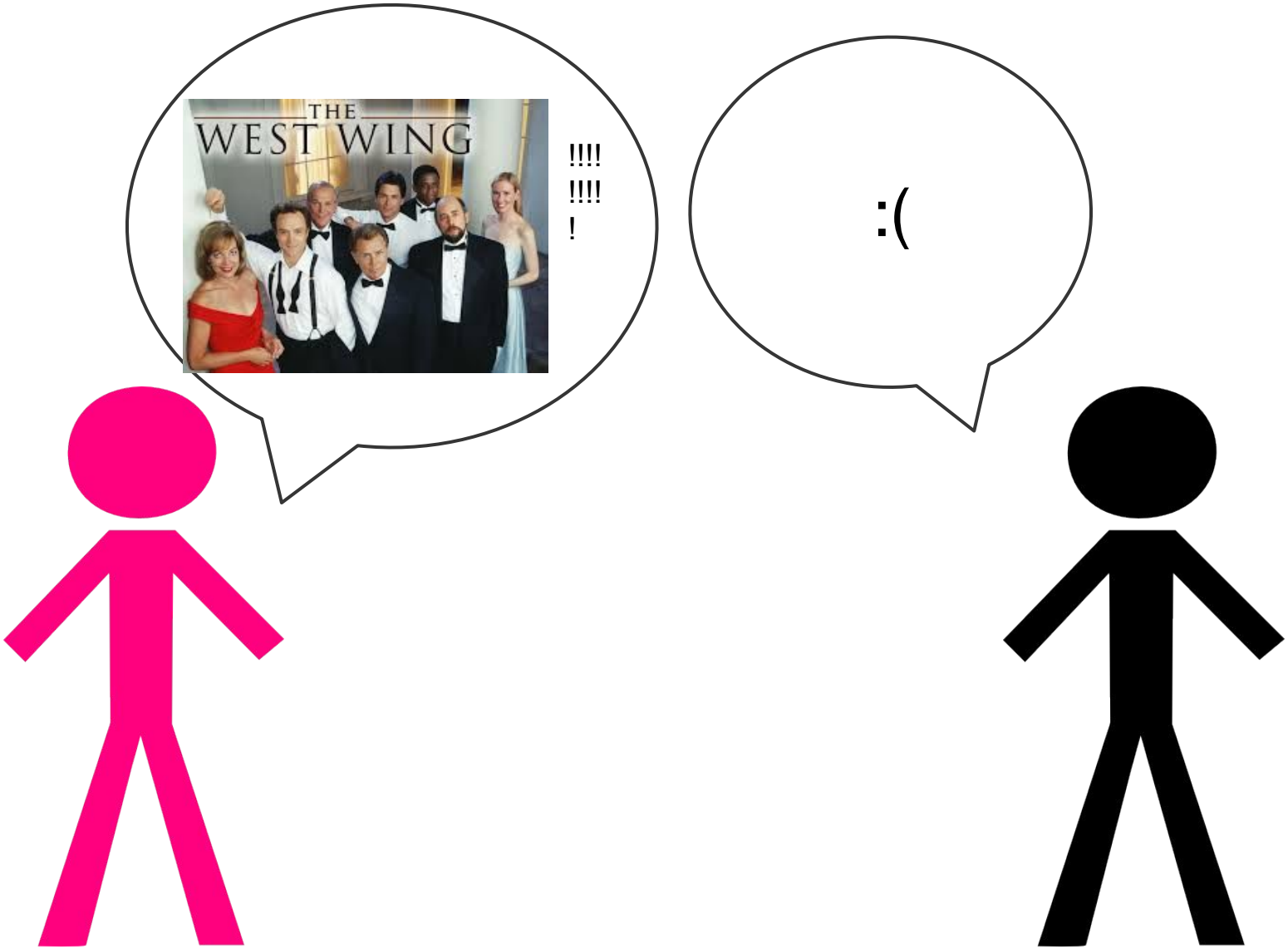
Goals of the Workshop

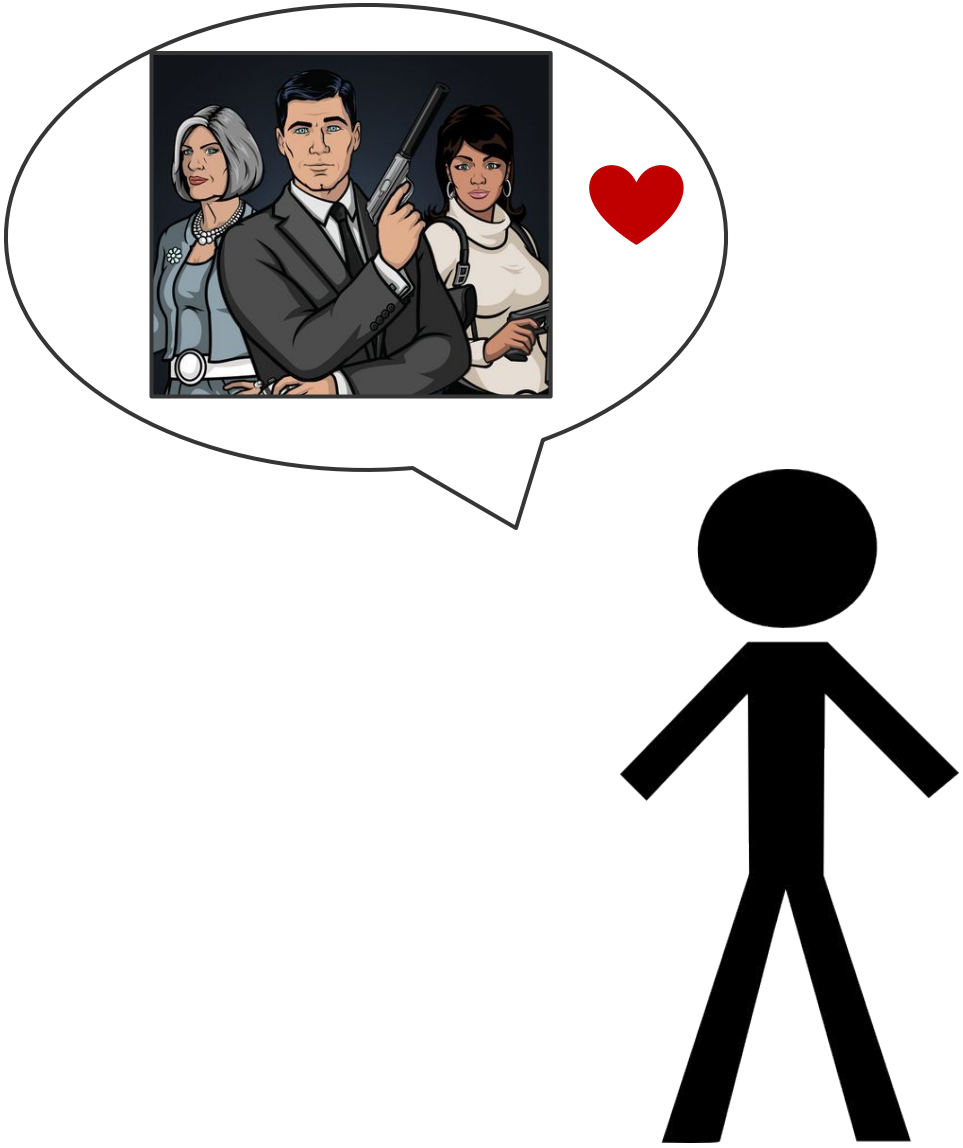
1. Exposure to Machine Learning
2. Write a program!
3. Recognize NN from “real life”

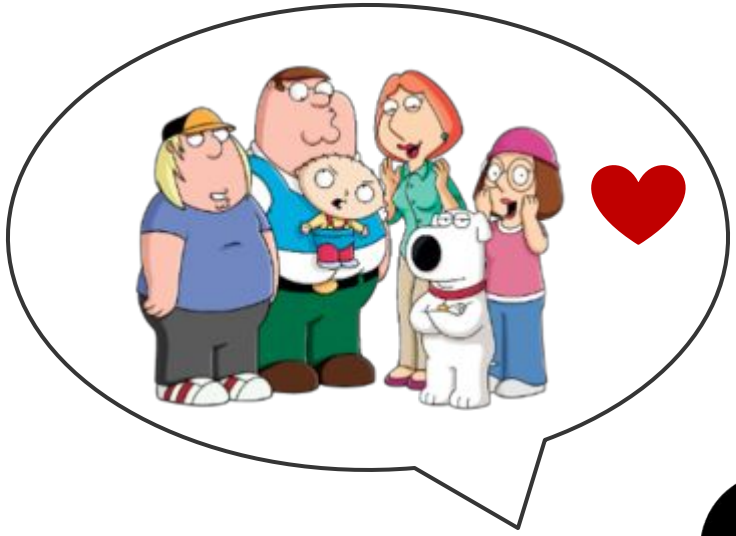
“Oh. That wasn’t so tough!”

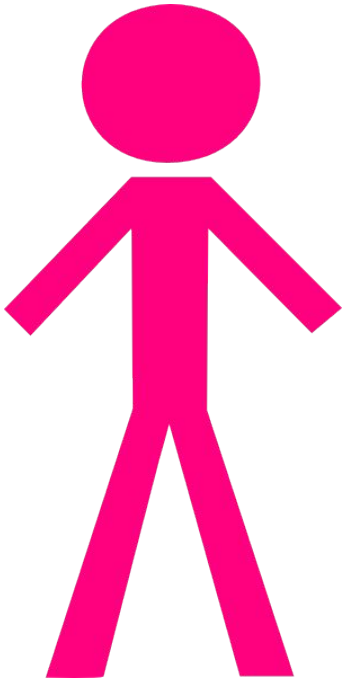
1. Exposure to Machine Learning

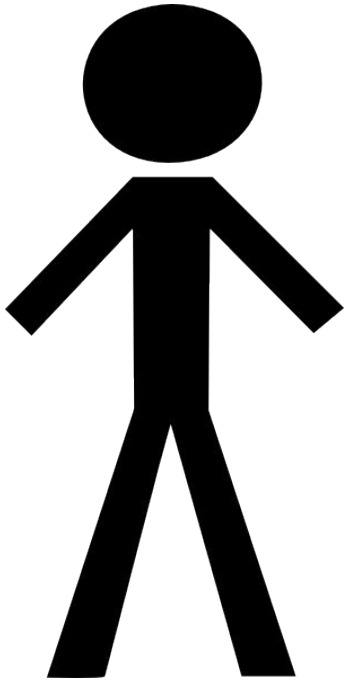

THE
WEST WING
!!!!
!!!!
!

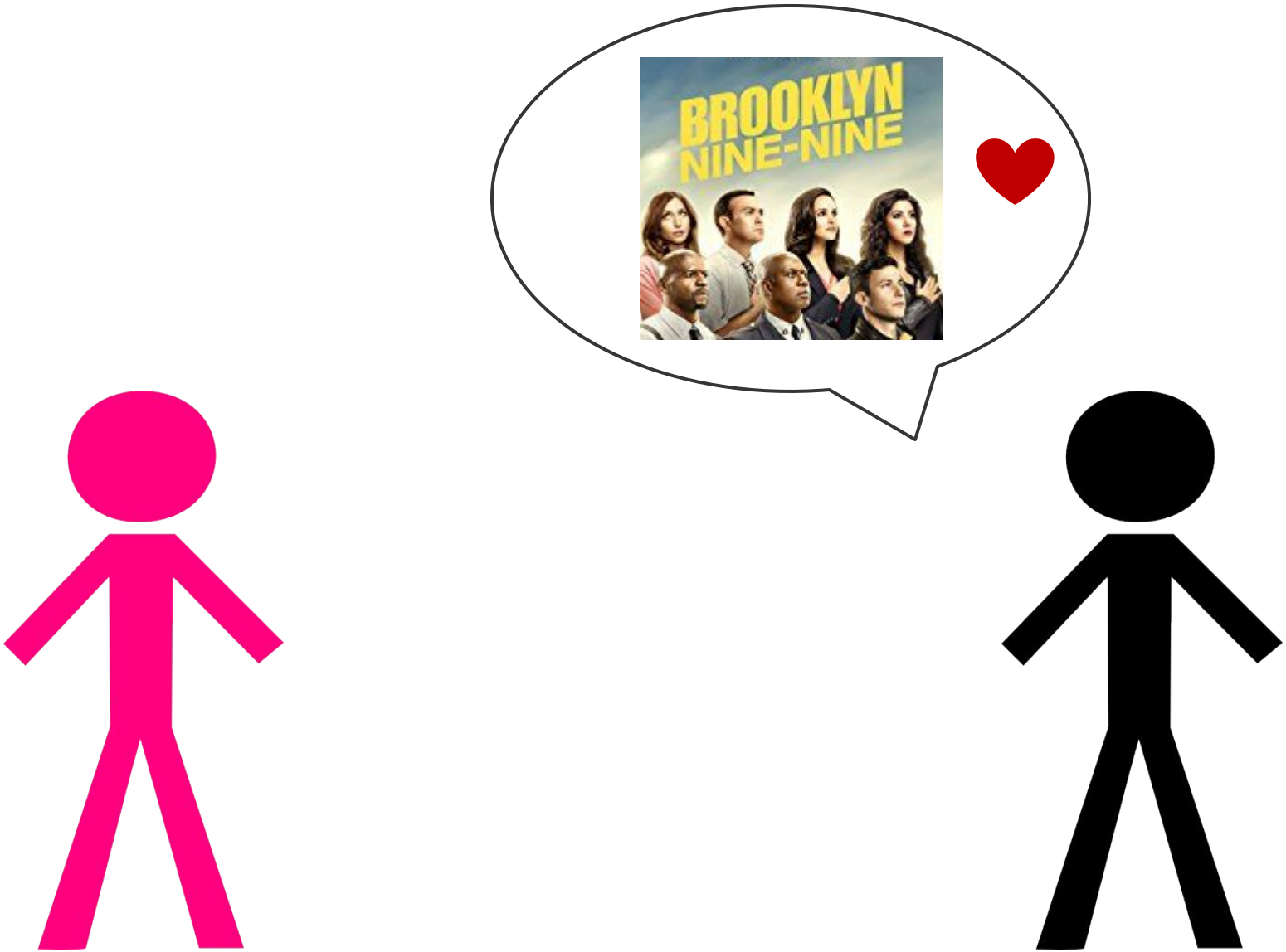

THE
WEST WING
!!!!
!!!!
!
:(

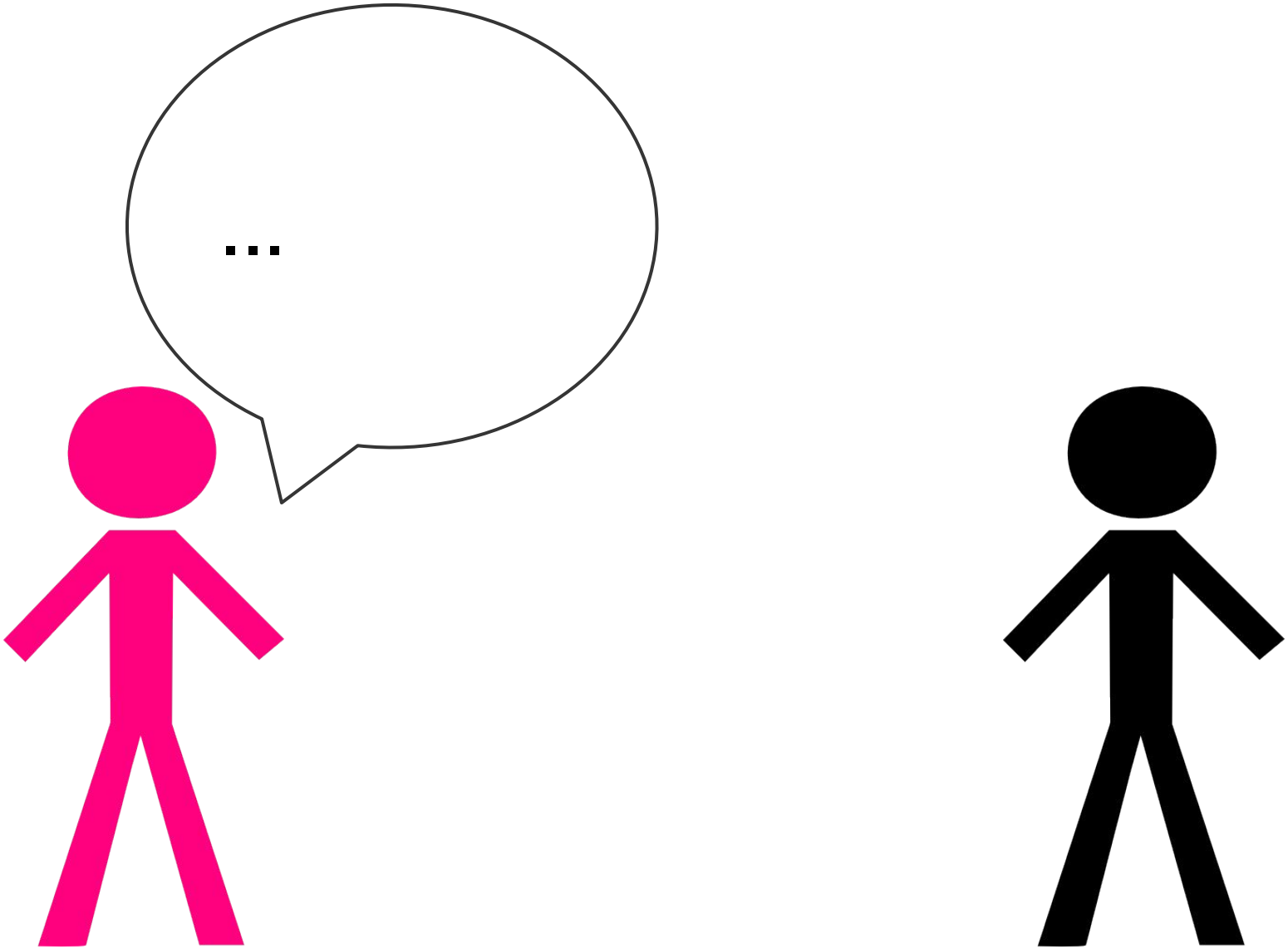


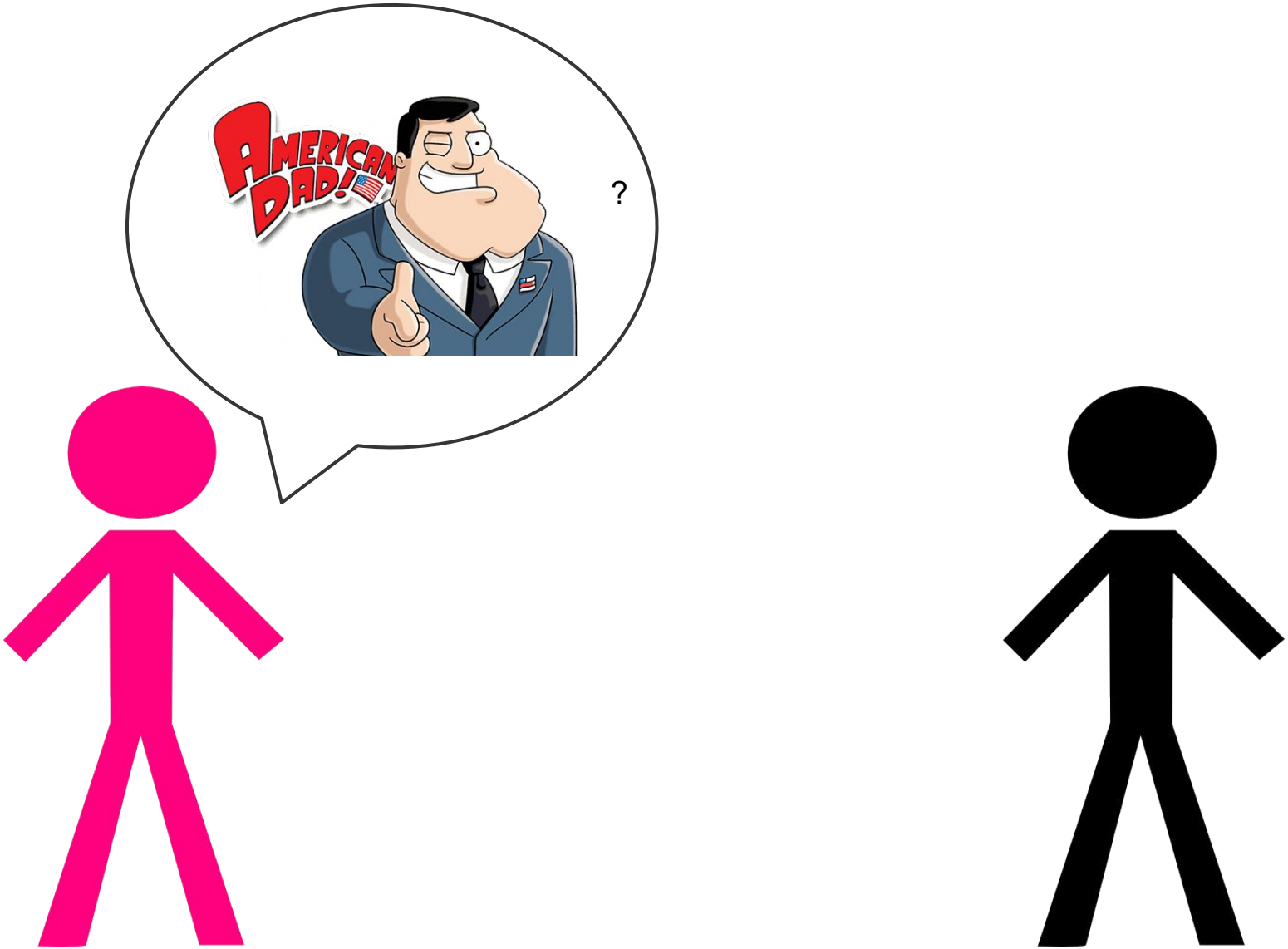


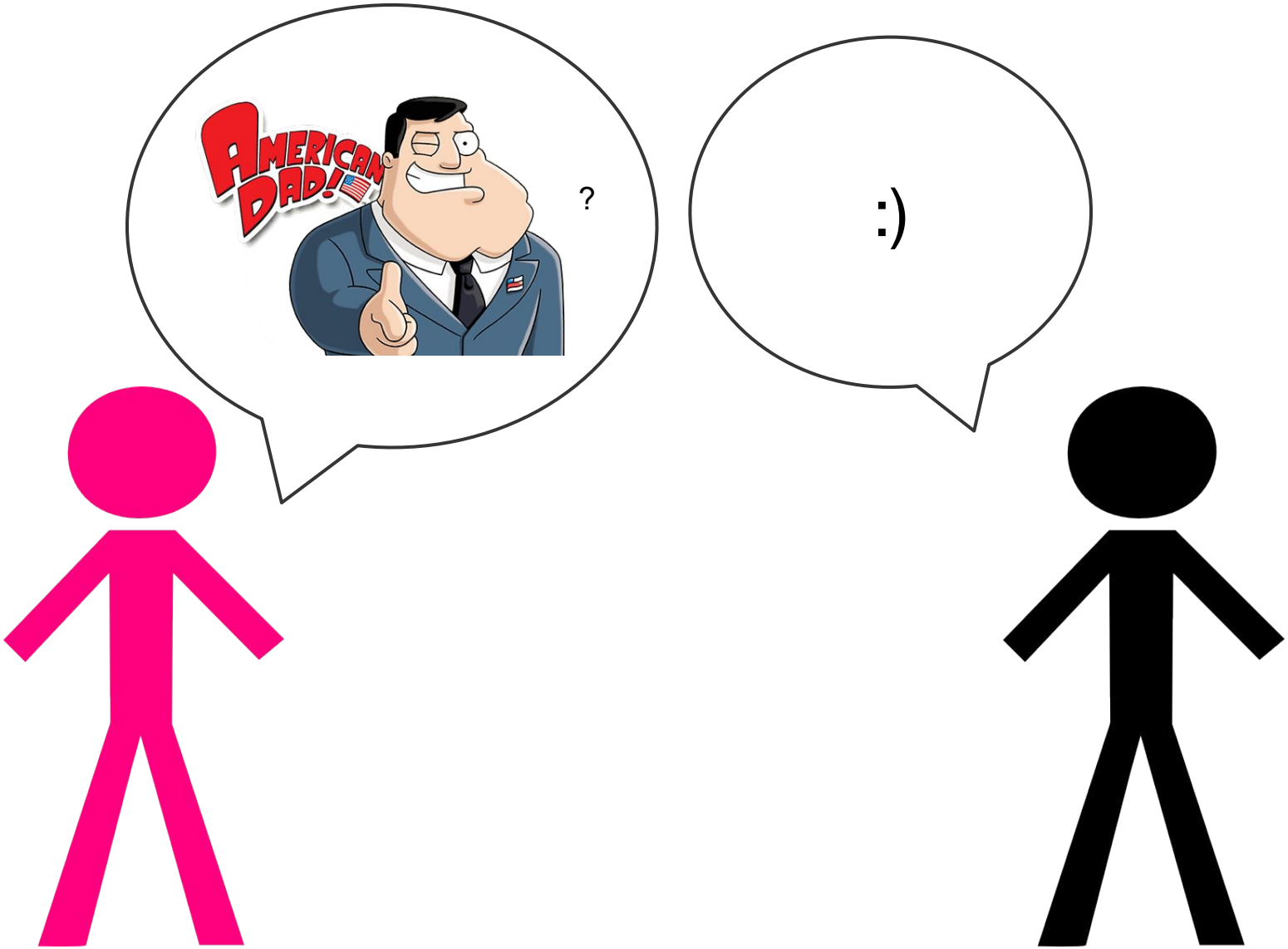





BROOKLYN
NINE-NINE




AMERICAN DAD!
?

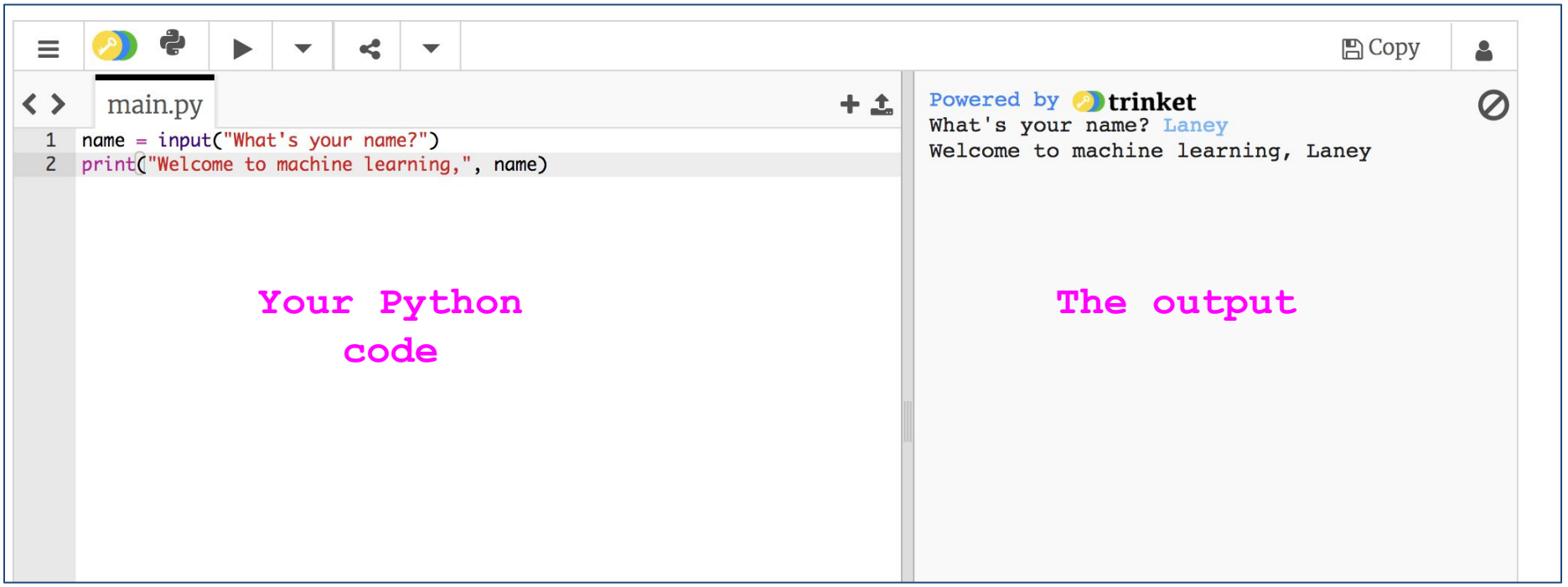

AMERICAN DAD!
?
:)

2. Write a program!

2. Write a program!

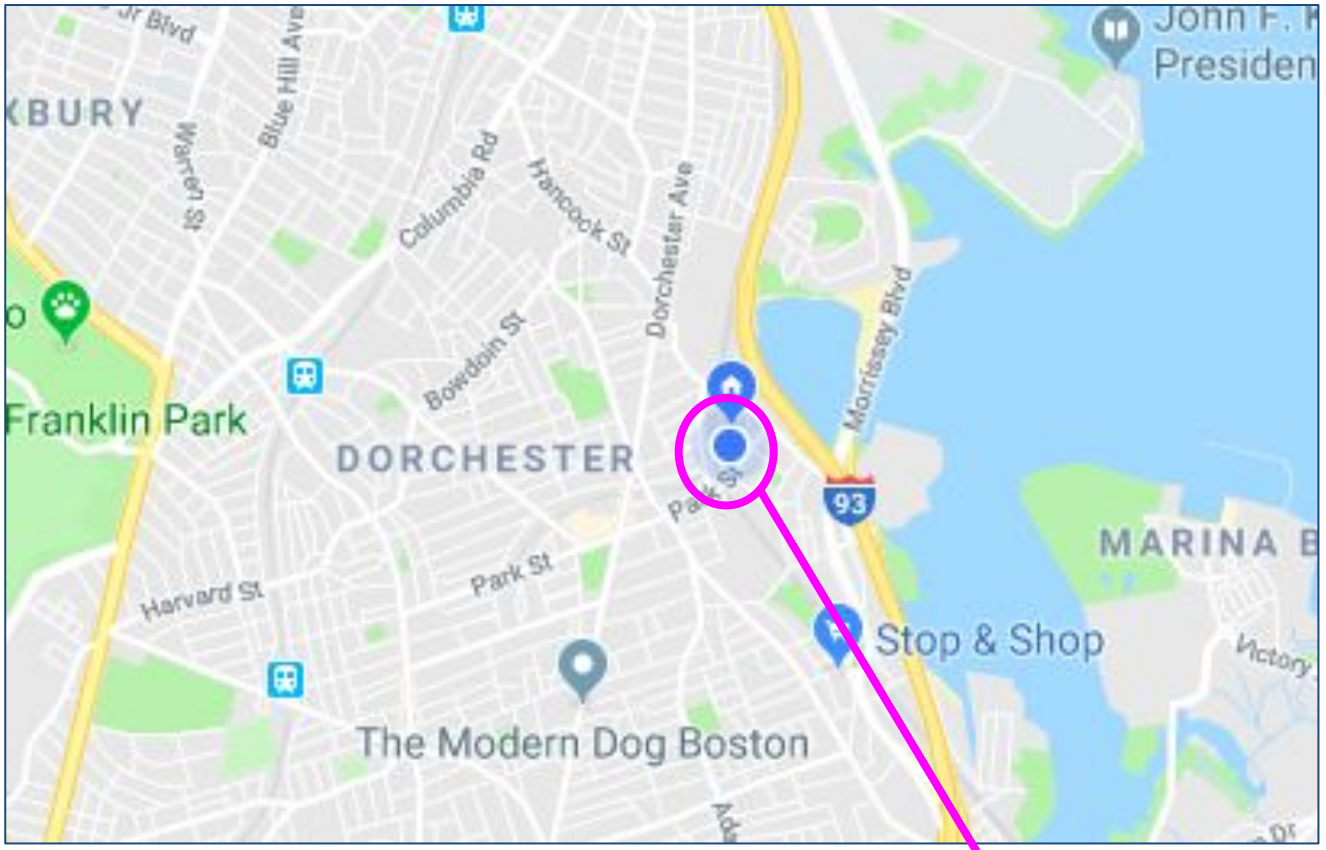
Often for the first time. It's exciting!

https://trinket.io/python3

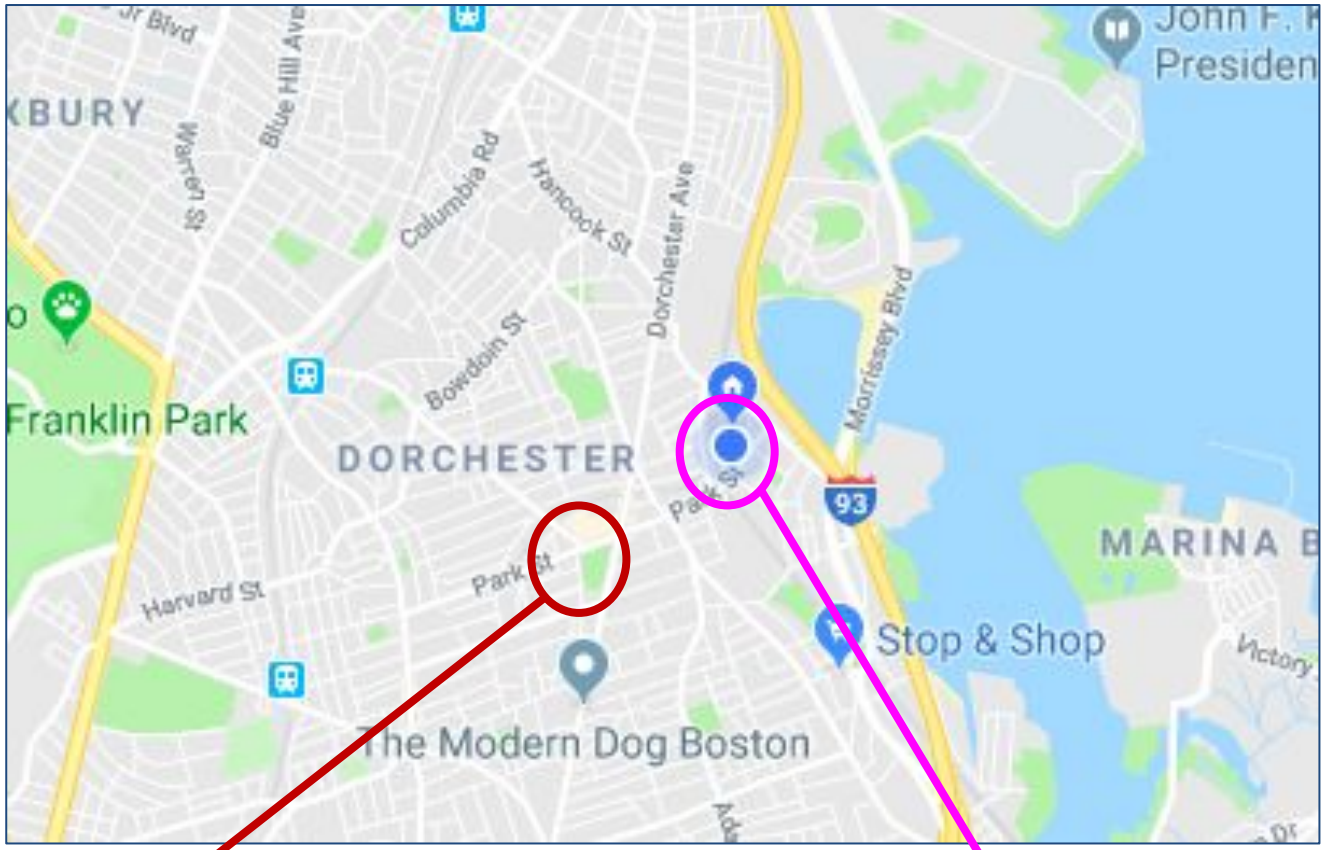

Copy
main.py
1 name = input("What's your name?")
2 print("Welcome to machine learning,", name)
Your Python code
Powered by trinket
What's your name? Laney
Welcome to machine learning, Laney
The output

3. Recognize Nearest Neighbors from “real life”

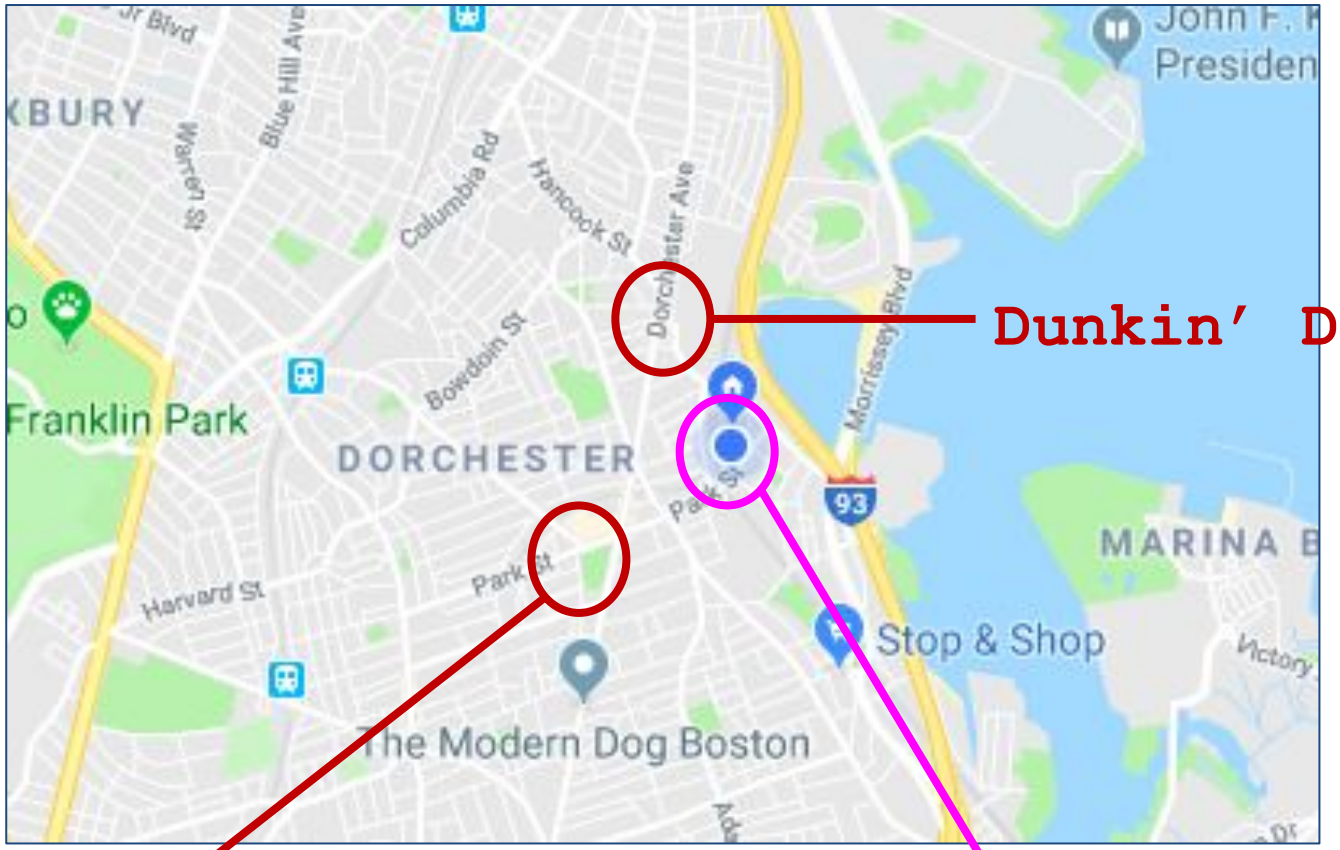
Euclidean Distance



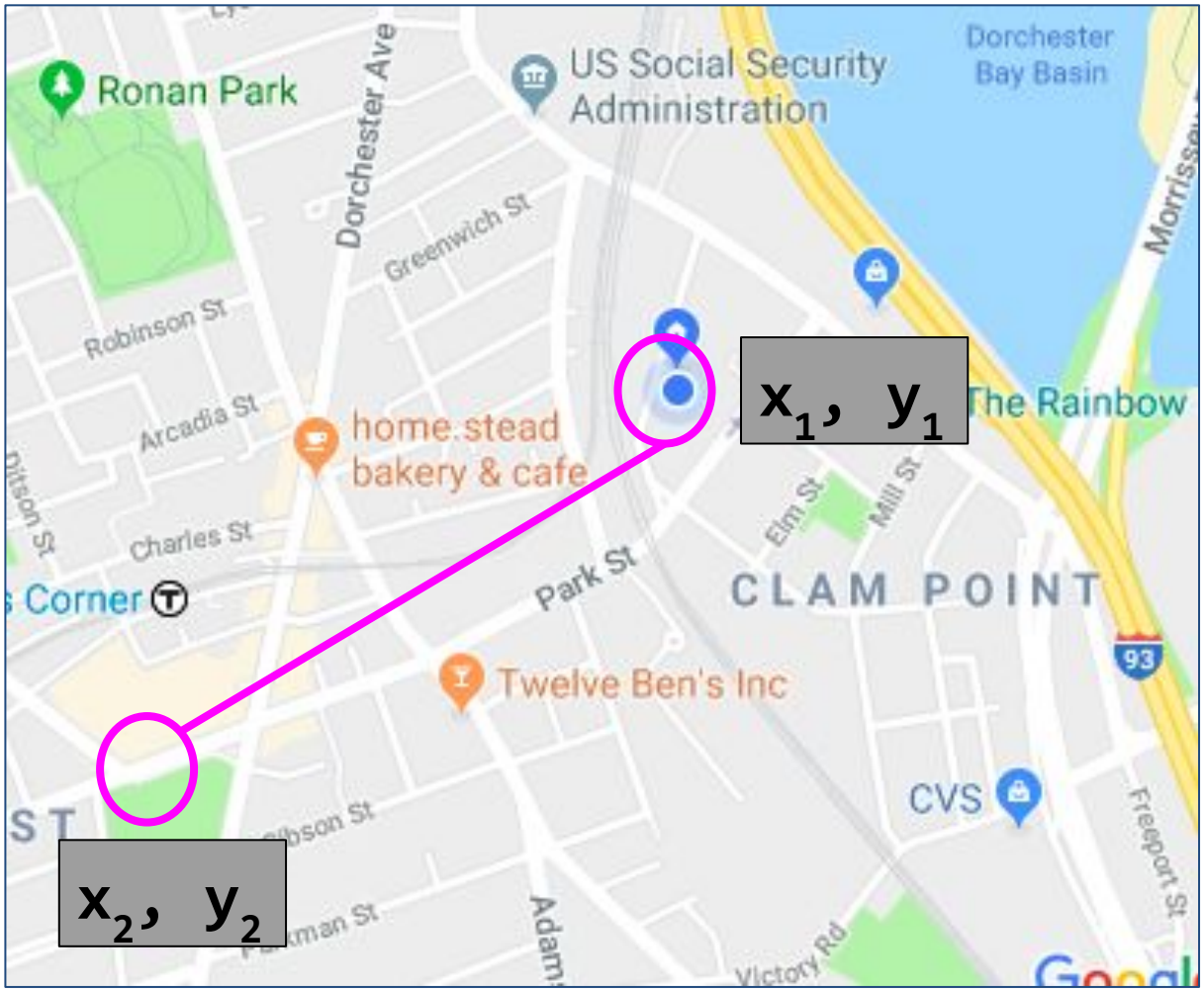
Euclidean Distance



Euclidean Distance



Euclidean Distance



Euclidean Distance

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Euclidean Distance

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Define the function in Python

Euclidean Distance

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Define the function in Python

Step by step

Euclidean Distance

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Define the function in Python

Step by step

Did it work?

Now the NN Algorithm

Now the NN Algorithm

```
[ [42.35, -71.18], [42.33, -73.09], [42.57, -71.88],
  [43.33, -71.29], [44.34, -72.07], [41.88, -70.01],
  [41.41, -71.11], [40.72, -73.99], [40.72, -73.98] ]
```

Given a community of data points...

Now the NN Algorithm

```
[ [42.35, -71.18], [42.33, -73.09], [42.57, -71.88],
  [43.33, -71.29], [44.34, -72.07], [41.88, -70.01],
  [41.41, -71.11], [40.72, -73.99], [40.72, -73.98] ]
```

[42.03, -71.99]?

...find the Nearest Neighbor of a new point

“Keep Going!”

- How else are two cafes similar?
  - Price range
  - Do they have food
  - Sit-down or takeaway

Outcomes

Outcomes

- 15 women participated

Outcomes

- 15 women participated (and one guy!)

Outcomes

- 15 women participated (and one guy!)
- Survey results (all qualitative)
  - Learned a lot
  - This was fun!
  - When's the next one?

Outcomes

- 15 women participated (and one guy!)
- Survey results (all qualitative)
  - Learned a lot
  - This was fun!
  - When's the next one?
- 2 are auditing my Data Science class right now

Outcomes

- 15 women participated (and one guy!)
- Survey results (all qualitative)
  - Learned a lot
  - This was fun!
  - When's the next one?
- 2 are auditing my Data Science class right now
- Volunteers LOVED IT

Using the Assignment

- Host a workshop!

Using the Assignment

- Host a workshop!
- Assign in CS1
  - Functions
  - Arithmetic Operators
  - (Conditionals)
  - (Lists)
  - (Loops)

```
Scales up/down
nicely! (In our
experience. :)
```

The End :)

Questions?

The CS1 Version

- Have assigned this problem four times

The CS1 Version

- Have assigned this problem four times
- (We don’t give them the code)

The CS1 Version

- Have assigned this problem four times
- (We don't give them the code)
- They write:
  - Euclidean distance
  - Manhattan distance
  - Hamming distance
  - Absolute value

The CS1 Version

- Have assigned this problem four times
- (We don't give them the code)
- They write:
  - Euclidean distance
  - Manhattan distance
  - Hamming distance
  - Absolute value
- Then they find:
  - Nearest neighbor of a new data point

The CS1 Version

Concepts:

The CS1 Version

Concepts:

- Functions
  - First time writing them

The CS1 Version

Concepts:

- Functions
  - First time writing them
- Arithmetic Operators
  - Don't use `pow` or `sqrt` or `abs`

The CS1 Version

Concepts:

- Functions
  - First time writing them
- Arithmetic Operators
  - Don't use `pow` or `sqrt` or `abs`

Could ALSO include:

The CS1 Version

Concepts:

- Functions
  - First time writing them
- Arithmetic Operators
  - Don't use `pow` or `sqrt` or `abs`

Could ALSO include:

- Lists
  - Data points with > 2 coordinates

The CS1 Version

Concepts:

- Functions
  - First time writing them
- Arithmetic Operators
  - Don't use `pow` or `sqrt` or `abs`

Could ALSO include:

- Lists
  - Data points with > 2 coordinates
- Conditionals
  - Don't use `min` to find nearest neighbor