

Twitter Trends

John DeNero & Aditi Muralidharan
University of California, Berkeley

A Hook Into Data Science

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

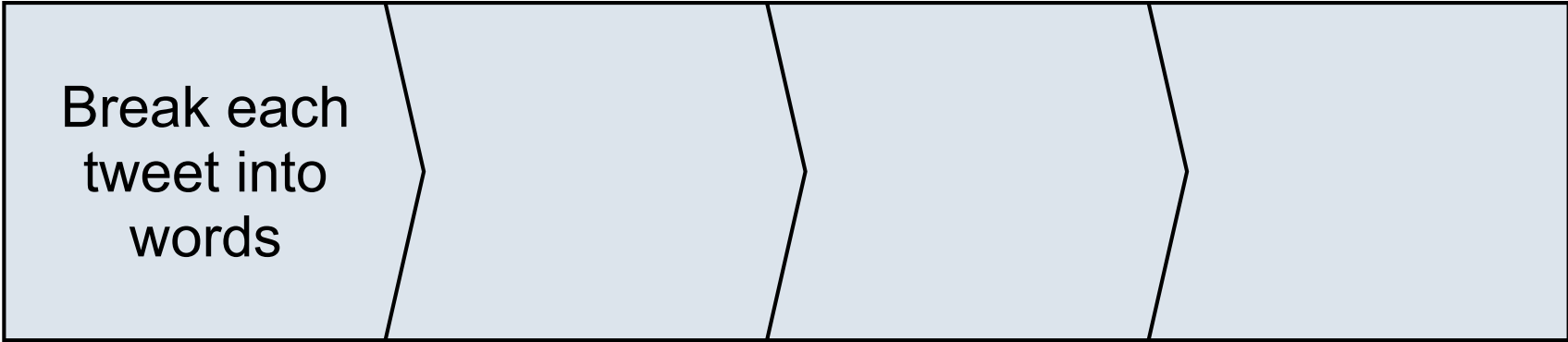
What do people tweet?
Draw their feelings on a map
to discover *trends*.

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

What do people tweet?
Draw their feelings on a map
to discover *trends*.

Break each
tweet into
words



A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

What do people tweet?
Draw their feelings on a map
to discover *trends*.

Break each
tweet into
words

Find all
tweets that
contain a
query word

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

What do people tweet?
Draw their feelings on a map
to discover *trends*.

Break each
tweet into
words

Find all
tweets that
contain a
query word

Group those
tweets by
US state

A Hook Into Data Science

- Second project (of four) in our CS 1 course, based on *The Structure and Interpretation of Computer Programs*.
- Uses Python built-in data types for sequences and maps.
- Students should: process lots of **real data**, create a useful and attractive **visualization**, & understand **data abstraction**.

What do people tweet?
Draw their feelings on a map
to discover *trends*.

Break each
tweet into
words

Find all
tweets that
contain a
query word

Group those
tweets by
US state

Compute
the average
sentiment of
those tweets

What Does America Think of Texas?

What Does America Think of Texas?

I **love** the Texas summer but a high of 111 is **crazy**

What Does America Think of Texas?

+0.625

I **love** the Texas summer but a high of 111 is **crazy**

What Does America Think of Texas?

+0.625

-0.5

I **love** the Texas summer but a high of 111 is **crazy**

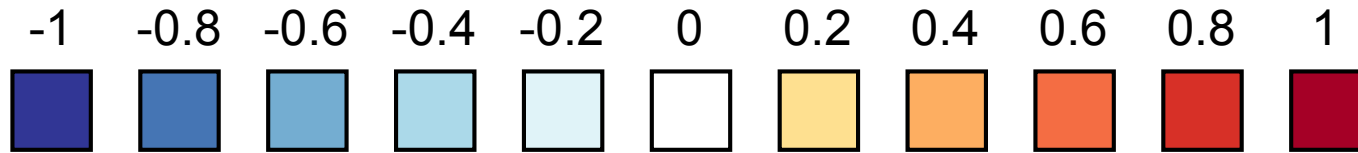
What Does America Think of Texas?

+0.625

-0.5

I **love** the Texas summer but a high of 111 is **crazy**

What Does America Think of Texas?

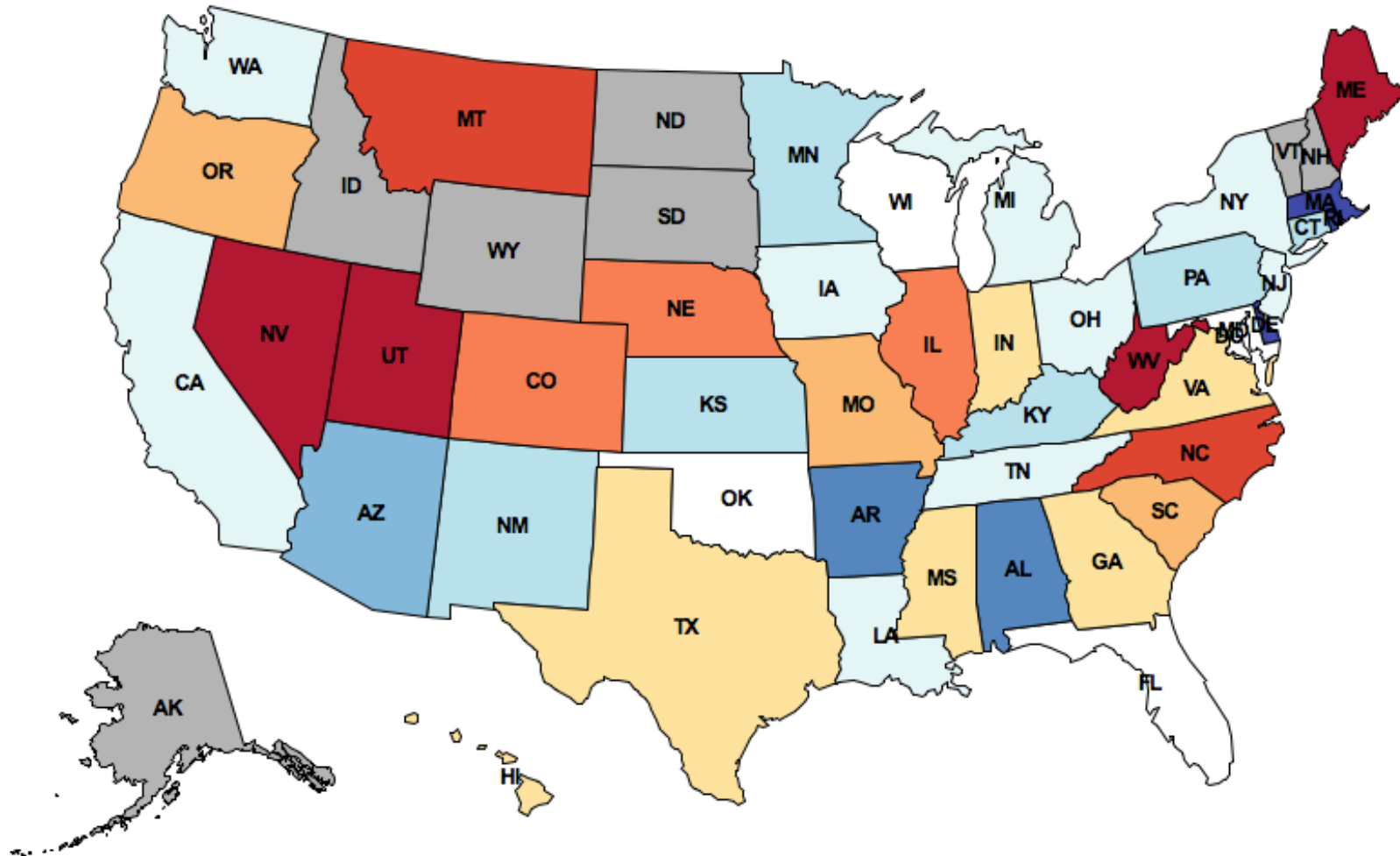
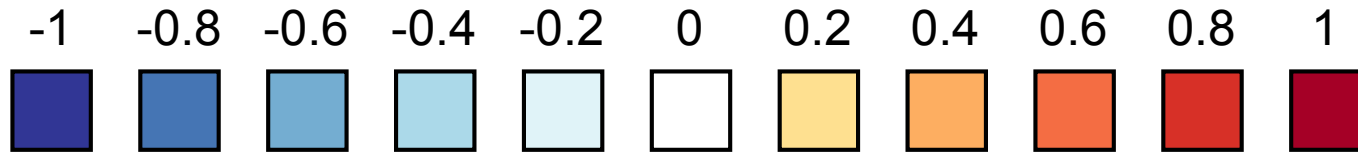


+0.625

-0.5

I **love** the Texas summer but a high of 111 is **crazy**

What Does America Think of Texas?



Finding the Centroid of a State



Finding the Centroid of a State



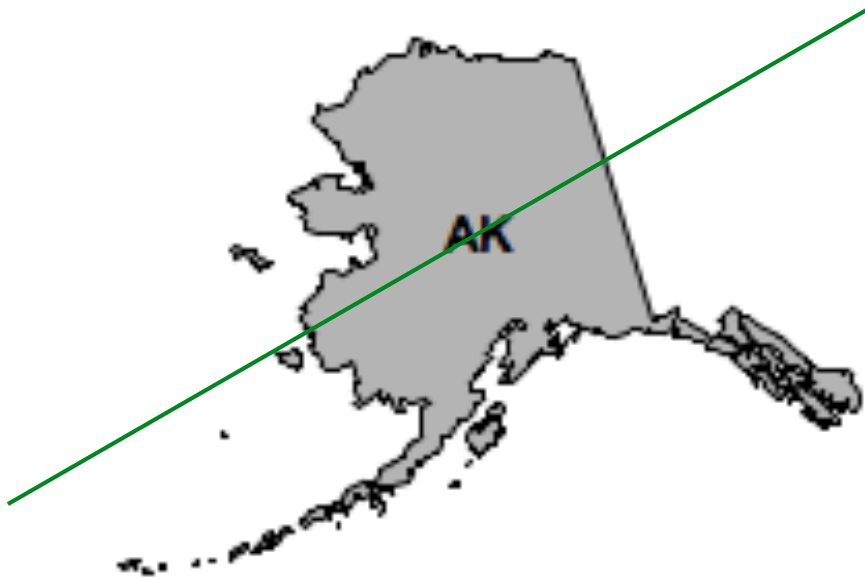
Finding the Centroid of a State



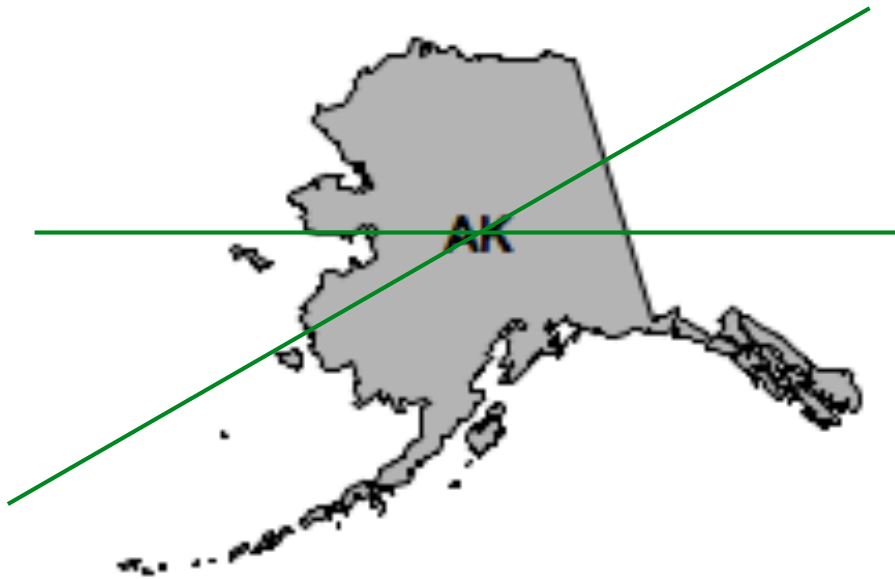
Finding the Centroid of a State



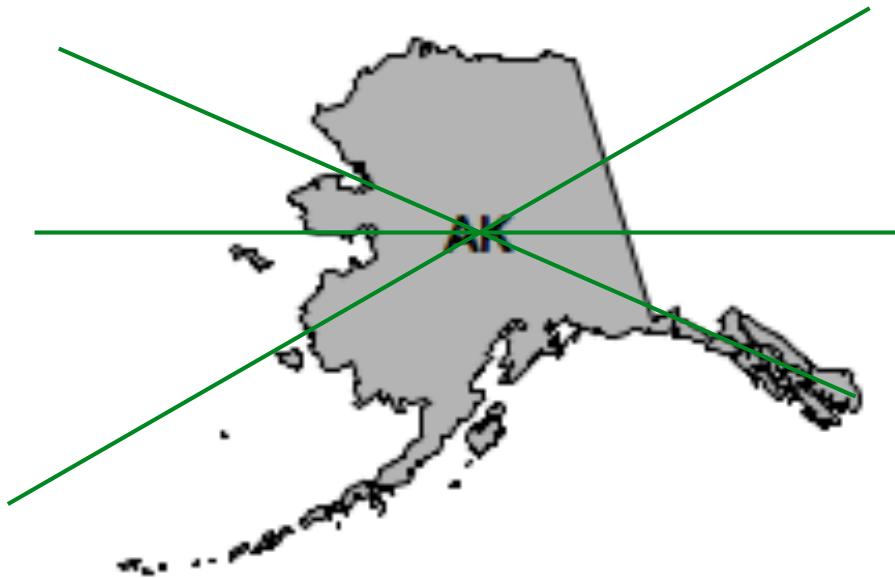
Finding the Centroid of a State



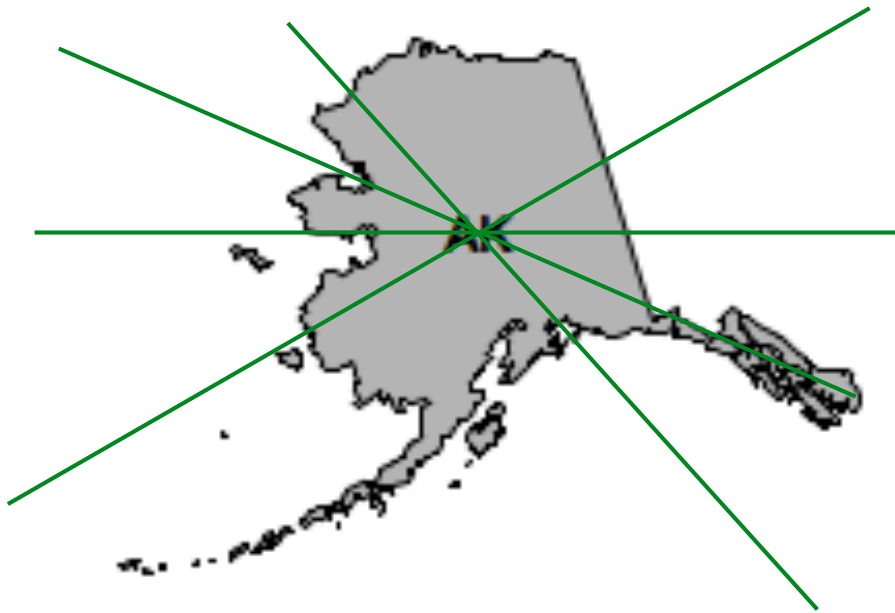
Finding the Centroid of a State



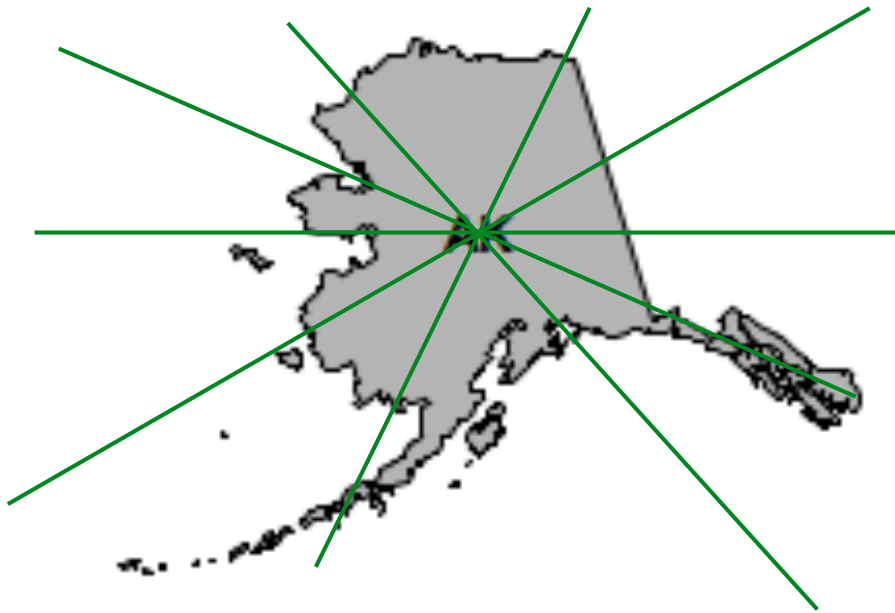
Finding the Centroid of a State



Finding the Centroid of a State

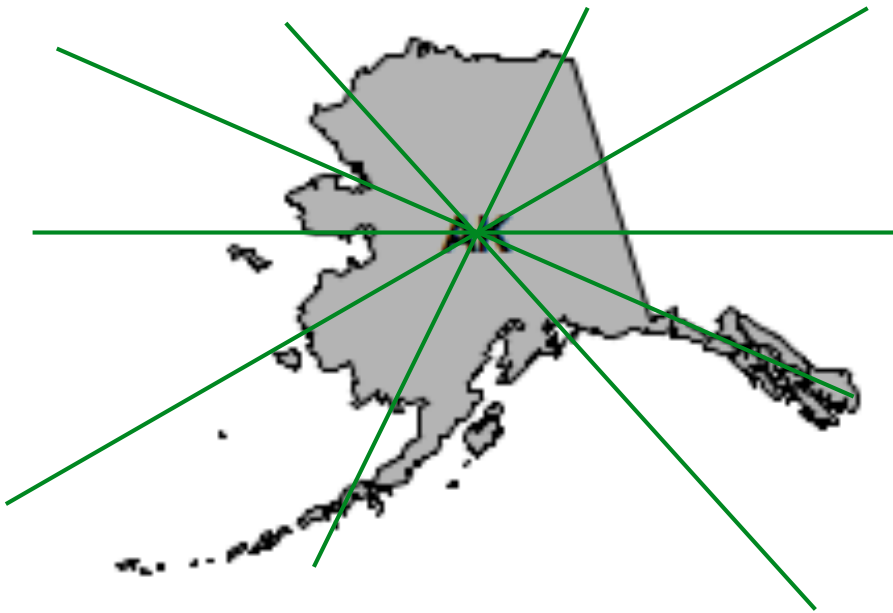


Finding the Centroid of a State



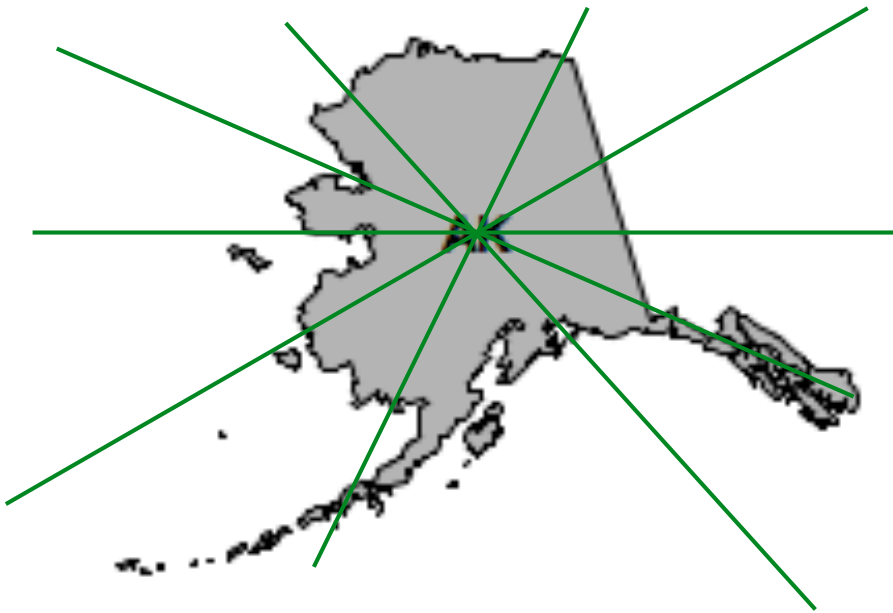
Finding the Centroid of a State

- Each state is represented by a sequence of polygons.



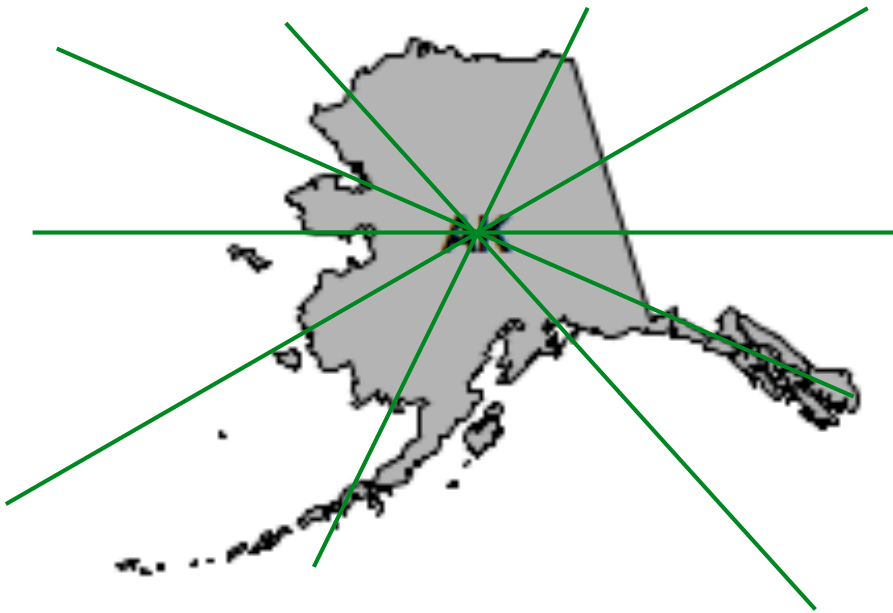
Finding the Centroid of a State

- Each state is represented by a sequence of polygons.
- Each polygon is represented by a sequence of positions.



Finding the Centroid of a State

- Each state is represented by a sequence of polygons.
- Each polygon is represented by a sequence of positions.



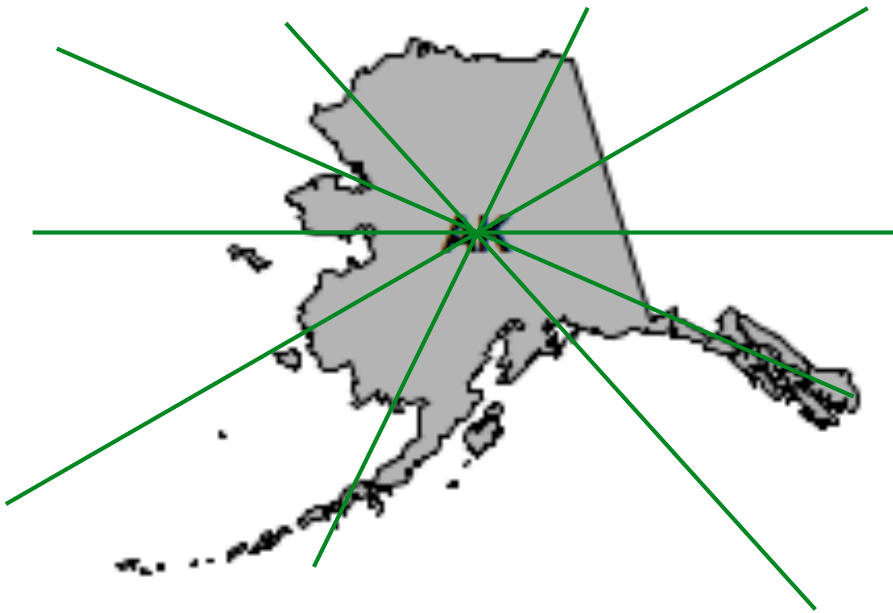
$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

Finding the Centroid of a State

- Each state is represented by a sequence of polygons.
- Each polygon is represented by a sequence of positions.



$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

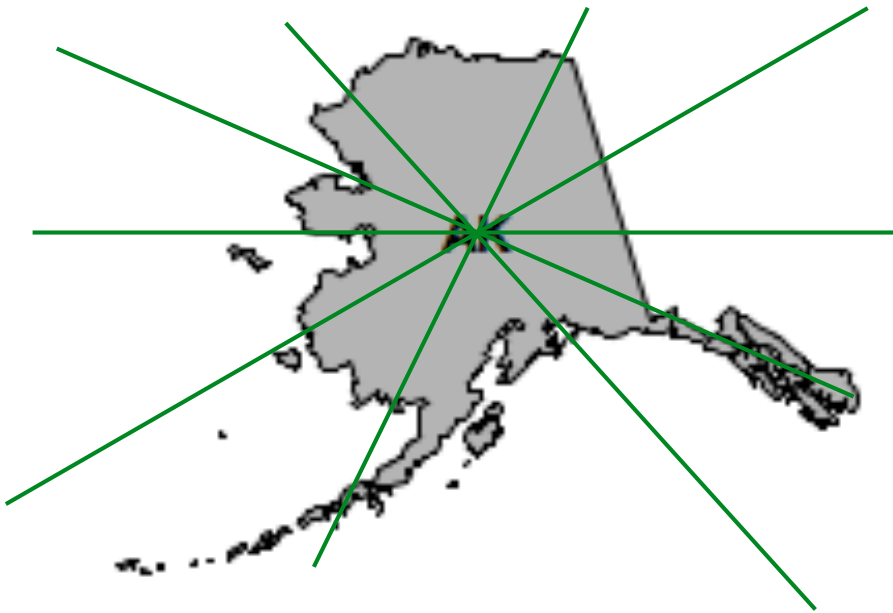
$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

- Students need simple unit tests to solve this problem.

Finding the Centroid of a State

- Each state is represented by a sequence of polygons.
- Each polygon is represented by a sequence of positions.



$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

- Students need simple unit tests to solve this problem.
- (!) Some students encounter floating point approximations.

Checking for Data Abstraction

An *abstract data type* is defined by its behavior,
and its use should be independent of its representation.

Checking for Data Abstraction

An *abstract data type* is defined by its behavior, and its use should be independent of its representation.

```
def make_position(lat, lon):  
    """Return a position..."""  
    return (lat, lon)  
  
def latitude(position):  
    """Return the latitude..."""  
    return position[0]  
  
def longitude(position):  
    """Return the longitude..."""  
    return position[1]
```

Checking for Data Abstraction

An *abstract data type* is defined by its behavior, and its use should be independent of its representation.

```
def make_position(lat, lon):  
    """Return a position..."""  
    return (lat, lon) lambda x: lat if x else lon  
  
def latitude(position):  
    """Return the latitude..."""  
    return position[0] position(true)  
  
def longitude(position):  
    """Return the longitude..."""  
    return position[1] position(false)
```

Survey Results

Survey Results

- Compared to three other projects (2 games, 1 interpreter)

Survey Results

- Compared to three other projects (2 games, 1 interpreter)
- Which project did you enjoy the most? (21.4% overall)
 - Female (23.9%) versus male (20.8%)
 - Started programming after 19th birthday (24.2%)
 - Taking first computer science course (19.0%)
 - Final grade of an A (14.5%), B (25.7%), or C (16.7%)

Survey Results

- Compared to three other projects (2 games, 1 interpreter)
- Which project did you enjoy the most? (21.4% overall)
 - Female (23.9%) versus male (20.8%)
 - Started programming after 19th birthday (24.2%)
 - Taking first computer science course (19.0%)
 - Final grade of an A (14.5%), B (25.7%), or C (16.7%)
- Which project taught you the most? (7.8% overall)
 - Female (8.2%) versus male (7.8%)
 - Final grade of an A (3.2%), B (8.8%), or C (13.9%)