DSFA
Spring 2020

Lecture 7

Functions

Announcements

- HW 3 out, due Friday at 5:59PM
- Prelim 1 coming up, 2/27, 7:30-9PM, more info soon.
- Change in late policy, to be announced on Piazza; "slip days" for assignments.

Can we predict the height of a child given the heights of the child's parents?

The Density Scale

Histogram Axes

By default, `hist` uses a scale (`normed=True`) that ensures the area of the chart sums to 100%

- The horizontal axis is a number line (e.g., years)
- The vertical axis is a rate (e.g., percent per year)
- The area of a bar is a percentage of the whole

How to Calculate Height

The [40, 60) bin contains 45 out of 200 movies

- "45 out of 200" is 22.5%
- The bin is 60 - 40 = 20 years wide

$$\text{Height of bar} = \frac{22.5 \text{ percent}}{20 \text{ years}}$$

$$= 1.125 \text{ percent per year}$$

Height Measures Density

$$\text{Height} = \frac{\text{\% in bin}}{\text{width of bin}}$$

- The height measures the percent of data in the bin *relative to the width of the bin*.

- The height is said to be the **density** of the bin; the density (height) times the width gives the % in the bin.

Area Measures Percent

Area   =   % in bin   =   Height x width of bin

- "How many individuals in the bin?" Use area.
- "How crowded/dense is the bin?" Use height.

Discussion Question

What's the height of each bar in these two histograms?

```
actress.hist(1, bins=[0,15,25,85])
```

```
actress.hist(1, bins=[0,15,35,85])
```

What are the vertical axis units?

Name	2016 Income (millions)
Jennifer Lawrence	61.7
Scarlett Johansson	57.5
Angelina Jolie	40
Jennifer Aniston	24.75
Anne Hathaway	24
Melissa McCarthy	24
Bingbing Fan	20
Sandra Bullock	20
Cara Delevingne	15
Reese Witherspoon	15
Amy Adams	15
Kristen Stewart	12
Amanda Seyfried	10.5
Tina Fey	10.5
Julia Roberts	10
Emma Stone	10
Natalie Portman	8.5
Margot Robbie	8
Meryl Streep	6
Mila Kunis	4.5

Clicker question

What are the vertical axis units?

A. Counts
B. %
C. % per millions $
D. % per $

Comparing Histograms

Overlaid Graphs
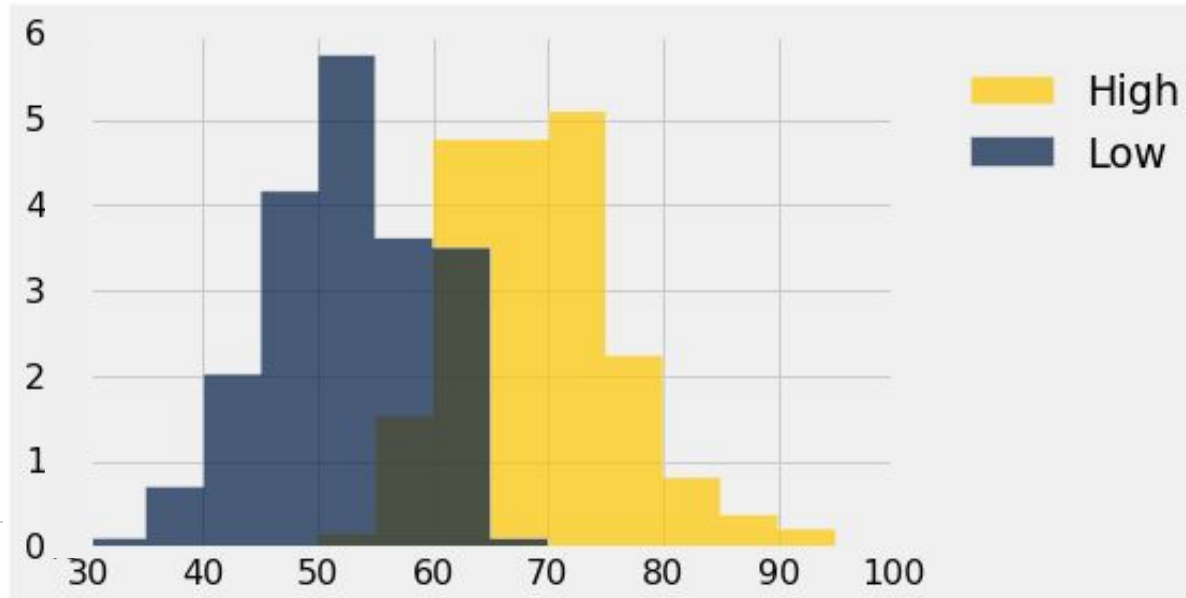
For visually comparing two populations

(Demo)

Discussion Question

This histogram describes a **year** of daily temperatures
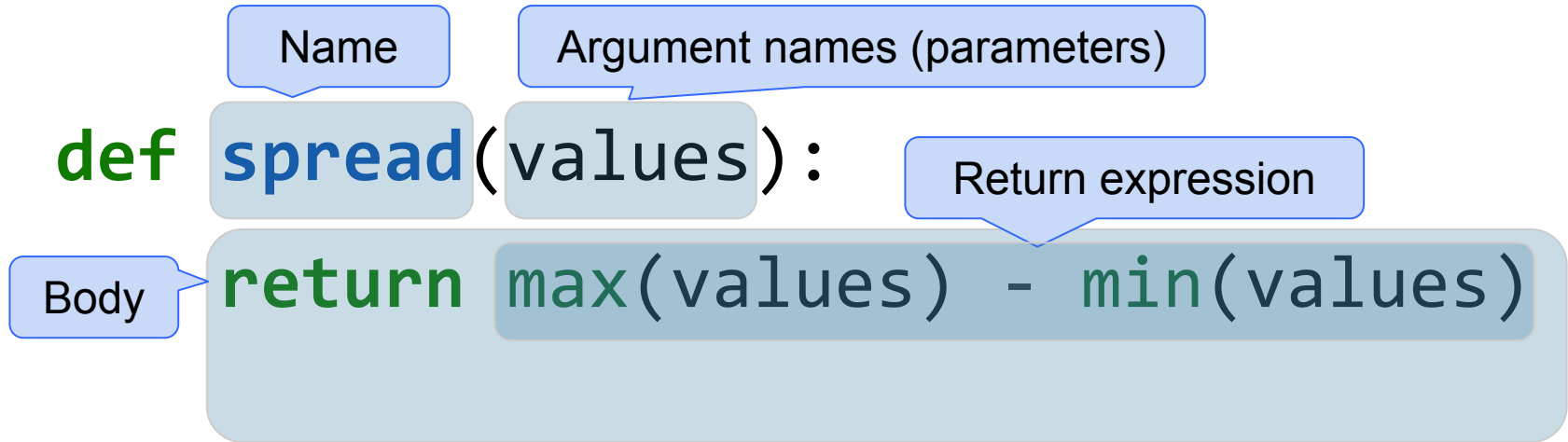
Try to answer these questions:

- What proportion of days had a high temp in the range 60-69?

- What proportion had a low of 45 or more?

- How many days had a difference of more than 20 degrees between their high & low temperatures?

Defining Functions

Def Statements

User-defined functions give names to blocks of code

Name

Argument names (parameters)

Return expression

```python
def spread(values):
    return max(values) - min(values)
```

Body

(Demo)

Discussion Question

What does this function do? What kind of input does it take? What output will it give? What's a reasonable name?

```python
def f(s):
    return np.round(s / sum(s) * 100, 2)
```

(Demo)

Apply

Apply

The `apply` method creates an array by calling a function on every element in input column(s)

- First argument: Function to apply
- Other arguments: The input column(s)

```
table_name.apply(function_name, 'column_label')
```

(Demo)

Example: Prediction

Sir Francis Galton

- 1822 - 1911 (knighted in 1909)

- A pioneer in making predictions

- Particular interest in heredity

- Charles Darwin's half-cousin

(Demo)