YData: An Introduction to Data Science

Lecture 09: Functions

Jessi Cisewski-Kehe
Statistics & Data Science, Yale University
Spring 2020

Credit: data8.org

Announcements

- HW03 due Thursday by 11:59 PM
- Friday, 2/7, lecture rescheduled for Wednesday, 2/5, 4 - 4:50 PM in OML 202

  It will be recorded and posted on Canvas
- Project 1 is posted (early) on our calendar in case you want to get started

  Checkpoint Fri 2/14; Due Fri 2/21

Histogram Review

Histogram Axes

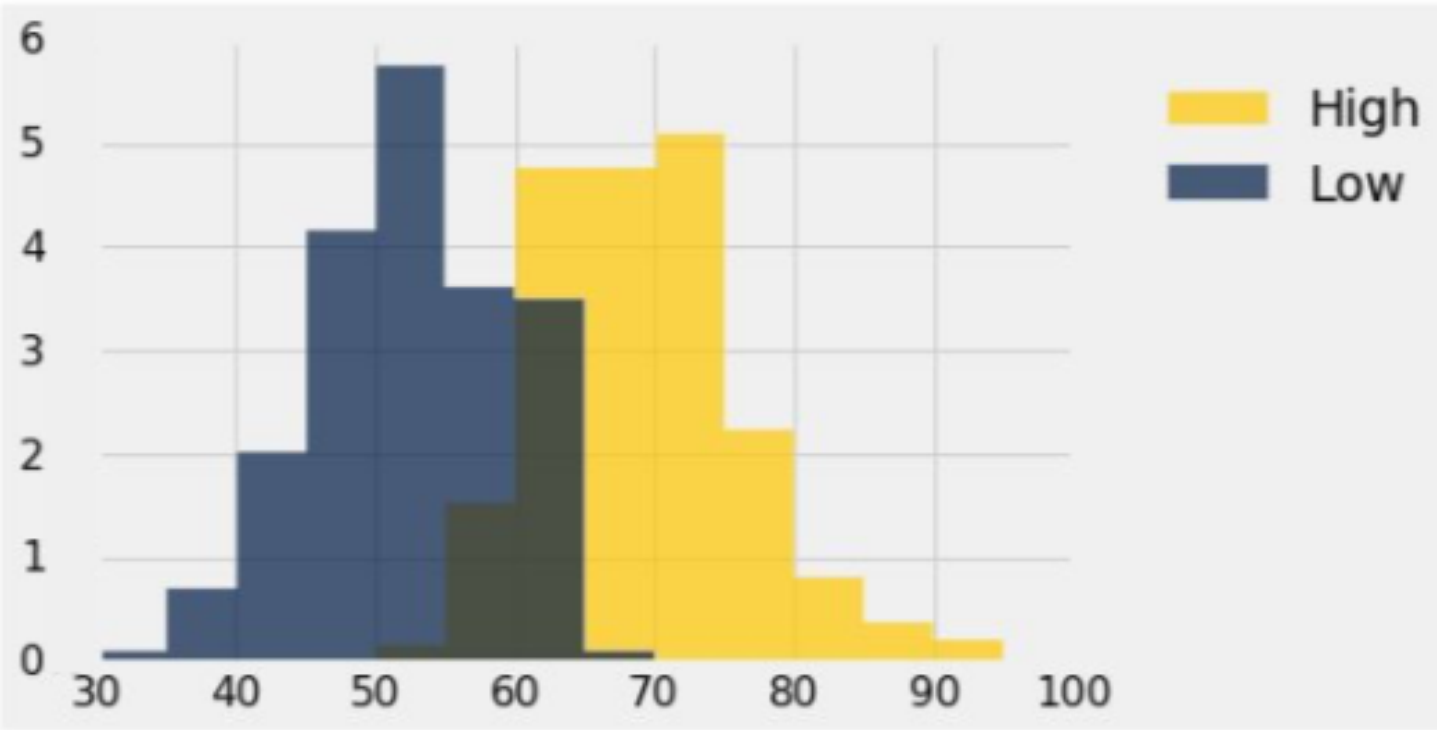
- By default, `hist` uses a scale (`normed=True`) that ensures the area of the chart sums to 100%
- The area of each bar is a percentage of the whole
- The horizontal axis is a number line (e.g., years), and the bins sizes don't have to be equal to each other
- The vertical axis is a density (e.g., percent per year)

(DEMO)

Discussion Question

This histogram describes a **year** of daily temperatures
Answer these questions, if possible:

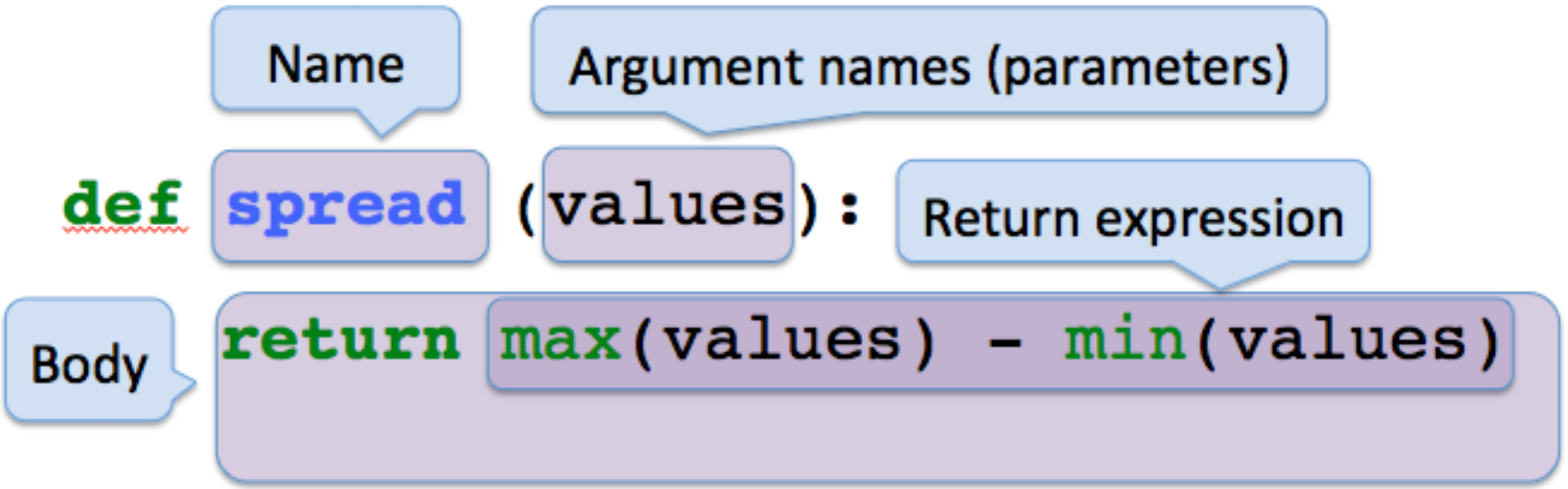
- What proportion of days had a high temp in the range 60-70?
- What proportion had a low of 45 or more?
- How many days had a difference of more than 20 degrees between their high & low temperatures?



Defining Functions

Def Statements

User-defined functions give names to blocks of code



(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?

```
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 (and troublesome) interest in heredity
- Charles Darwin's half-cousin

(DEMO)