



DATA 1202
Spring 2024

Lecture 8

Functions

Announcements

- **HW 4** due Wednesday at **5pm**.
- **No Lab this week**
 - No class Thursday - Friday for Winter Carnival

Histograms

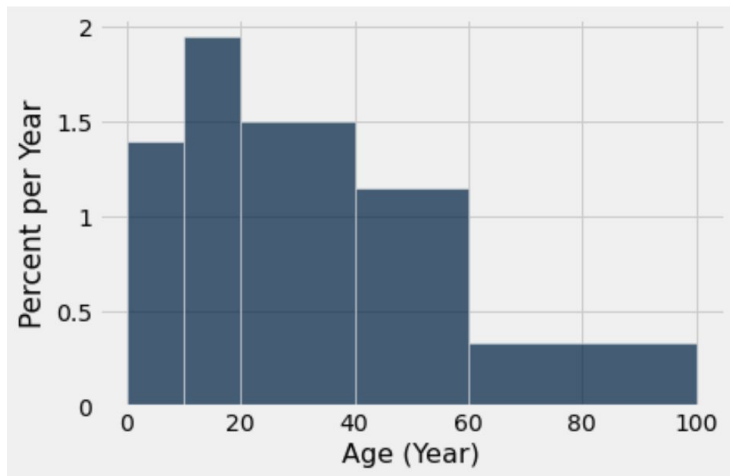
Area and Height

Area of bar = % in bin = Height x width of bin

- “How many individuals in the bin?” Use **area**.
- “How crowded is the bin?” Use **height**.

(Demo)

Discussion Questions



Compare the bins $[10, 20)$ and $[20, 40)$.

- Which one has more movies?
Answer: $[20, 40)$, bigger area
- Which one is more crowded?
Answer: $[10, 20)$, taller

Bar Chart or Histogram?

To display a distribution:

Bar Chart

- Distribution of categorical variable
- Bars have arbitrary (but equal) widths and spacings; in any order
- **height (or length)** and **area** of bars proportional to the percent of individuals

Histogram

- Distribution of numerical variable
 - Horizontal axis is numerical: drawn to scale, no gaps, bins can be unequal
 - **Area** of bars proportional to the percent of individuals; **height** measures density
-

Charts Wrap Up

Summary

- **Line graph**: sequential data (over time, etc.)
 - **Scatter plot**: relation between two numerical variables
 - **Bar chart**: distribution of one categorical variable *or* relation between a categorical and a numerical variable
 - **Histogram**: distribution of one numerical variable
-

Discussion Question

You have data about daily temperatures as shown. Which type of chart would show the answer to each question?

- Are there more cloudy than sunny days?
- What percentage of days have a high at least 72°?
- Do days with hotter highs tend to have hotter lows?

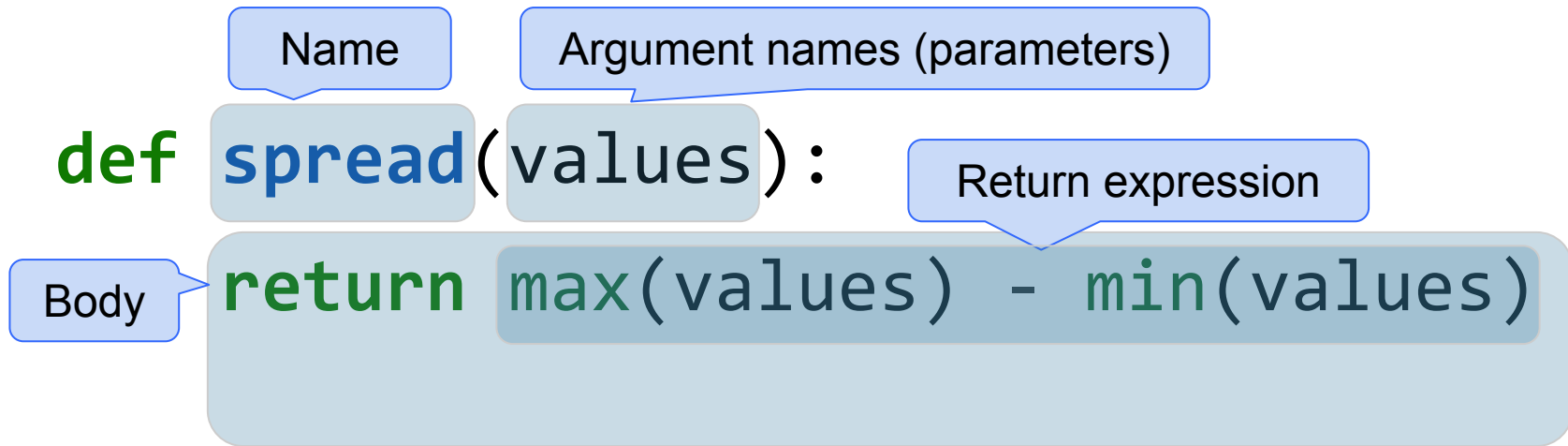
Day	High	Low	Sky condition
1	55.1	43.7	Cloudy
2	57.2	46	Sunny
3	56.8	45.9	Cloudy

... (362 rows omitted)

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

apply

1. Calls a function on every element in the input column(s)
2. Produces an array containing the output of the function on each input column element.
 - First argument: Function to apply
 - Other arguments: Specified input column(s)

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

(Demo)
