Histograms, Data Density, Functions

Prof. Abby Stylianou

Logistics

- Homework 2 out today, due before class next Tuesday

More on distributions

Let’s finish up our demo from last class…

How to Calculate Bin Height

The [40, 65) bin contained 51 out of 200 movies

- “51 out of 200” = 25.5%
- The bin is 65 – 40 = 25 years wide
- Height of bar = (25.5%) / 25 years
  = 1.02 percent per year

Height Measures Density

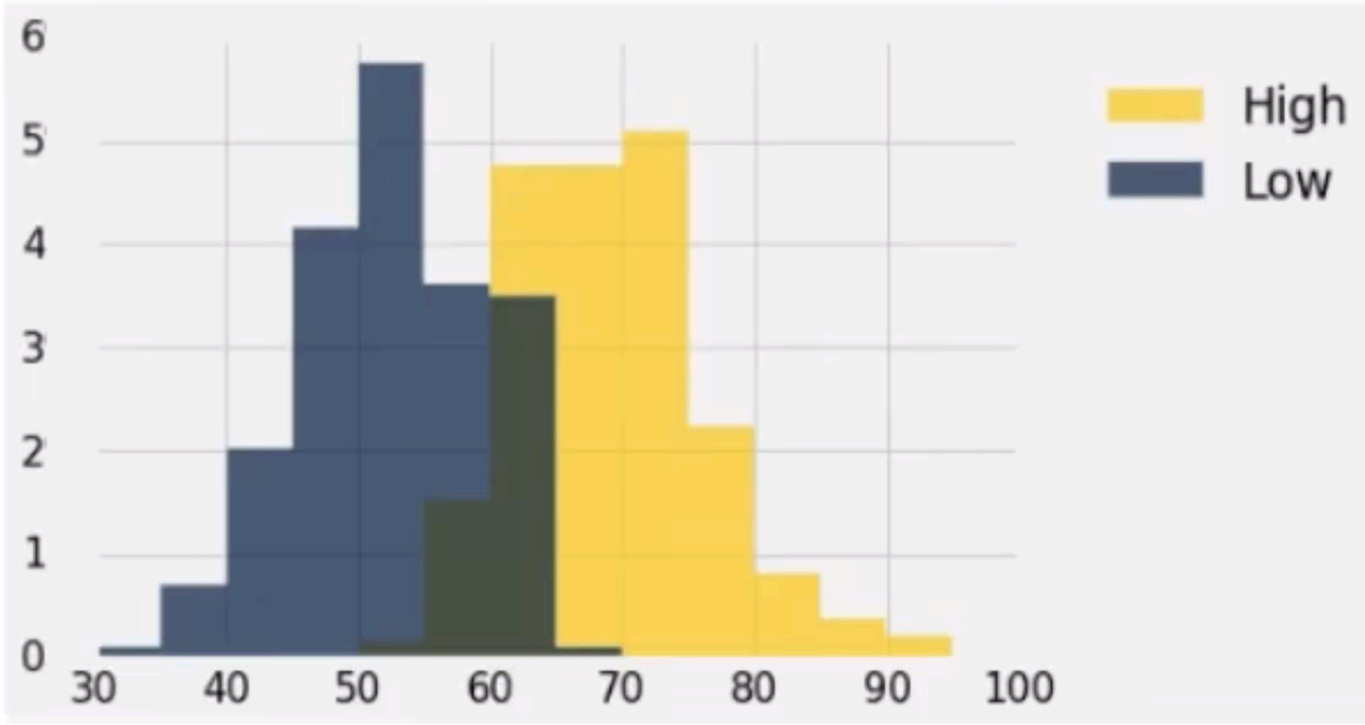
- Height $= \dfrac{\%\text{ in bin}}{\text{width of bin}}$

- Height measures the percent of data in the bin *relative to the amount of space in that bin.*
- Height measures crowdedness, or **density**
- Units: percent per unit on the horizontal axis

Discussion Questions

This histogram describes a year of daily temps

1) What proportion of days had a high temp in the range 60-69?
2) What proportion had a low of 45 or more?
3) How many days had a difference of more than 20 degrees between their high and low temperatures?



Bar Chart vs. Histogram?

Bar Chart

- Distribution of categorical variable
- Bars have arbitrary (but equal) widths and spacings
- Height (or length) of bars proportional to the percent of individuals

Histogram

- Distribution of numerical variable
- Horizontal axis is numerical: to scale, n gaps, bins can be unequal
- Area of bars is proportional to % of individuals; height measure density

Functions

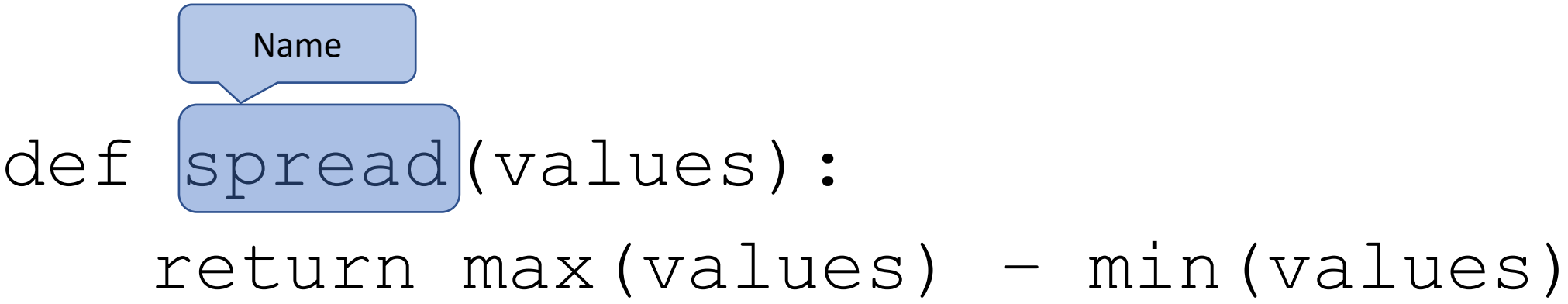
def Statements

User-defined functions give names to blocks of code

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

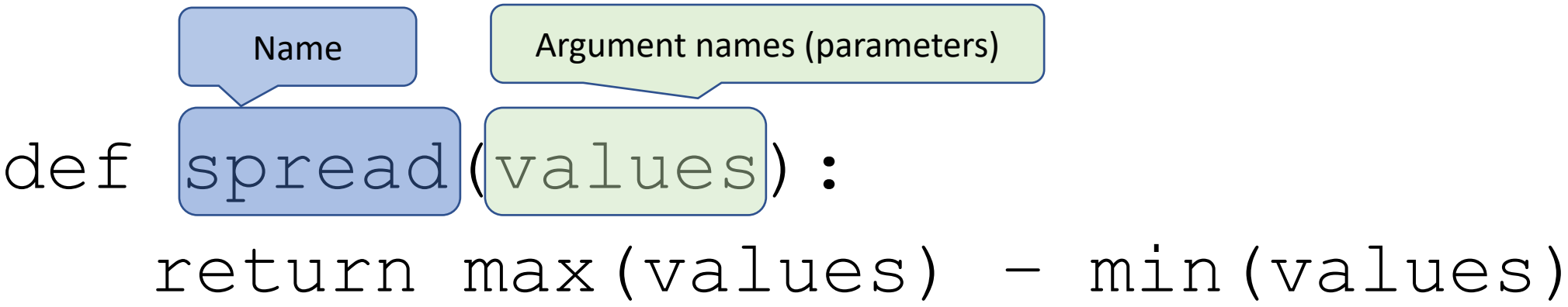
def Statements

User-defined functions give names to blocks of code



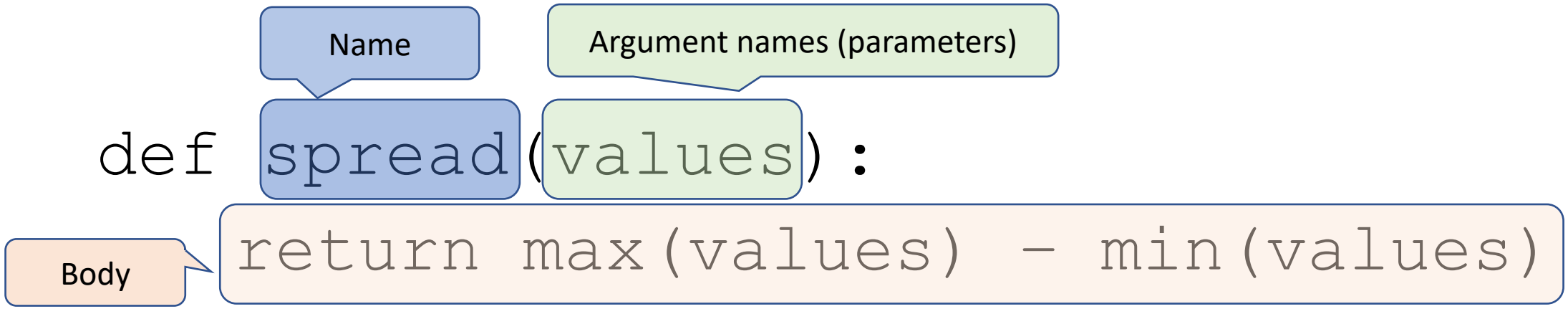
def Statements

User-defined functions give names to blocks of code



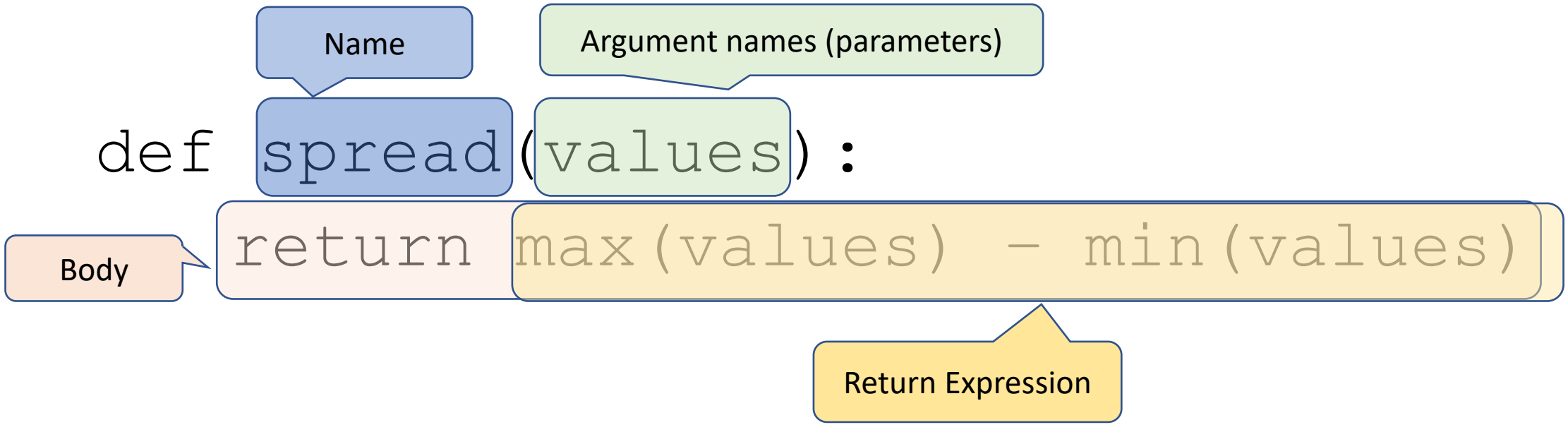
def Statements

User-defined functions give names to blocks of code



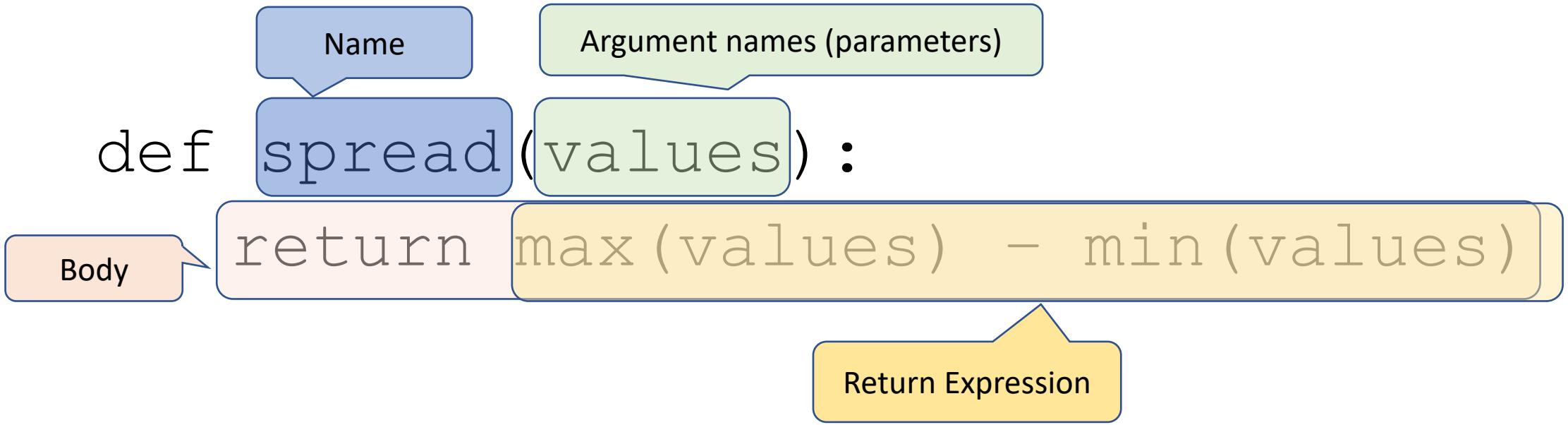
def Statements

User-defined functions give names to blocks of code



def Statements

User-defined functions give names to blocks of code



(Demo)

Discussion

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

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

Apply

- The `apply` method creates an array by calling a function n every element in input column(s)
  - First argument: function to apply
  - Other arguments: the input column(s)

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

Prediction: Sir Francis Galton

- 1822 – 1911 (knighted in 1909)
- A pioneer in making predictions
- Particular interest in heredity (father of eugenics)
- Charles Darwin's half-cousin