

Data 8, Lab 4

Plots and Functions

Hubert Luo

Fall 2019

20 September 2019

Agenda

1. Plots
2. Histograms
3. Functions

Plots

- Scatter plots:
 - Relationship between two numerical variables
- Line plot:
 - Sequence of one numerical variable (i.e., over time)
- Bar chart:
 - Distribution of categorical data
- Histogram:
 - Distribution of numerical data

Plots: Demo

- (Demo on Notebook)

Histograms

- Visualize the distribution of a numerical variable
- Bin one of the variables to get bars on the x axis
- Area of each bar corresponds the percentage of individuals in that bin
 - $\% \text{ in bin} = \text{Height of bin} \times \text{Width of bin}$
- Therefore, the y axis is the **density** ("Percent per unit on the x axis")
 - How "clustered" together the data is in a bin?
- Warning: Bins include the left endpoint, but not the right endpoint!

Bar Chart vs Histogram

- From lecture:

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

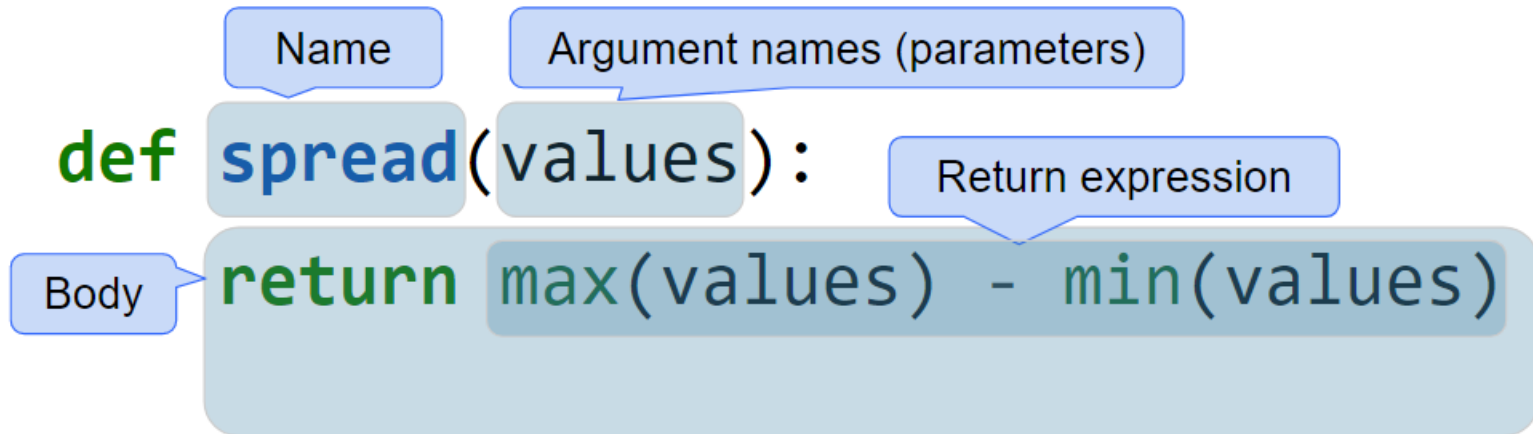
Histogram

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

Histogram: Demo

- (Demo on Notebook)

Functions



Functions: Apply

- Can apply a function to a column in a table
- Need name of function to call and column label

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

Functions: Demo

- (Demo on Notebook)

Lab Worksheet

Announcements

- Project 1 and HW4 getting released today
 - HW4 is pretty short!
- Everyone needs to fill out this form for partners:
 - Link: <https://tinyurl.com/data8-proj1-partners>
- We only have a set number of spots for tutoring sections – check the sign-up sheet periodically for availability
 - Don't email course staff!

Lab Notebook