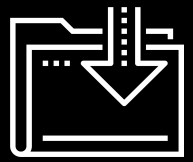
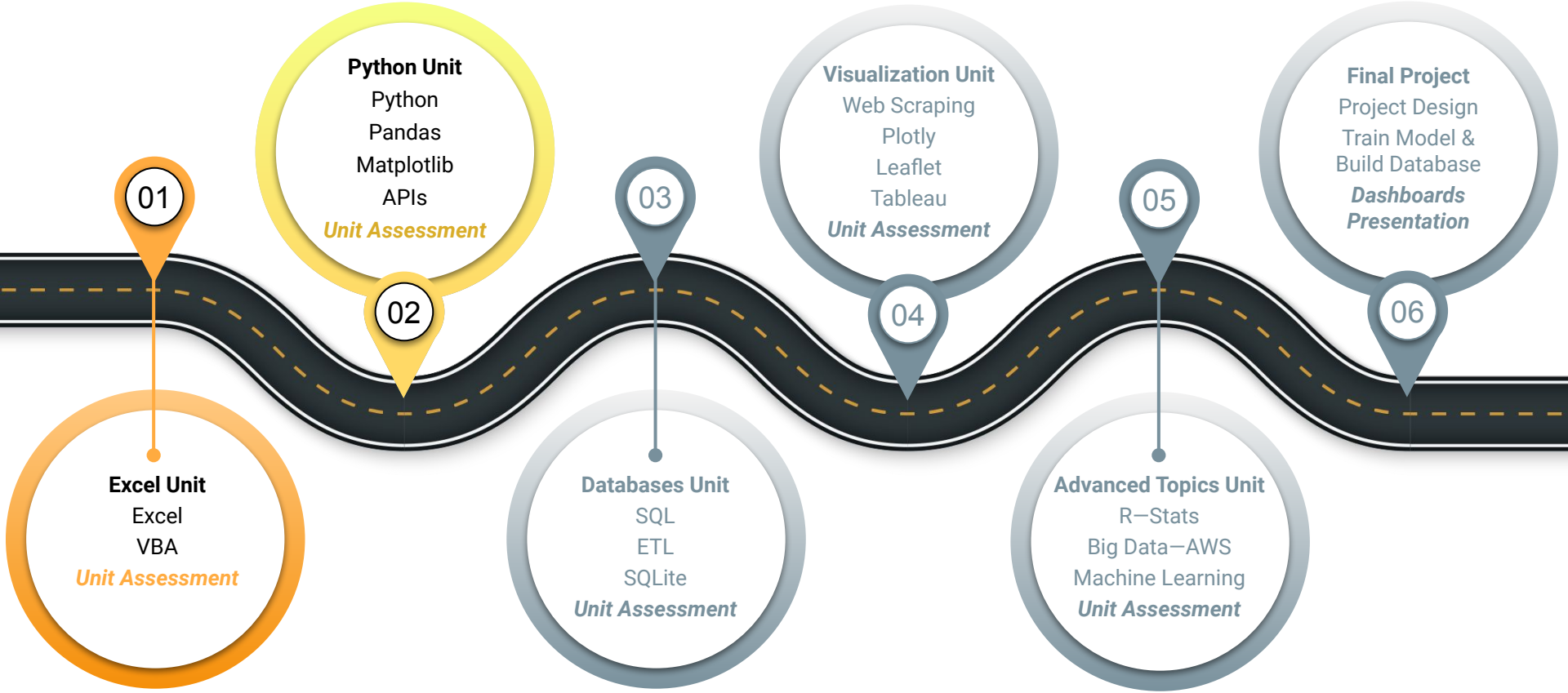
Matplotlib

Data Boot Camp

Lesson 5.1



The Big Picture



Pro Tip:

Using documentation to solve problems should become second nature to you!
Don’t stop practicing!

Module 5

This Week: Matplotlib

This Week: Matplotlib

By the end of this week, you'll know how to:

- Create line, bar, scatter, bubble, pie, and box-and-whisker plots using Matplotlib
- Add and modify features of Matplotlib charts
- Add error bars to line and bar charts
- Determine mean, median, and mode using Pandas, NumPy, and SciPy statistics

This Week's Challenge

Create a summary DataFrame of ride-sharing data by city type and a multiple-line graph showing weekly fares for each city type.

Career Connection

How will you use this module's content in your career?

Module 5

How to Succeed This Week

Pro Tip:

As we go through different plotting tools, remember to treat these challenging weeks as learning opportunities.

Module 5

Today’s Agenda

Today’s Agenda

By completing today’s activities, you’ll learn the following skills:

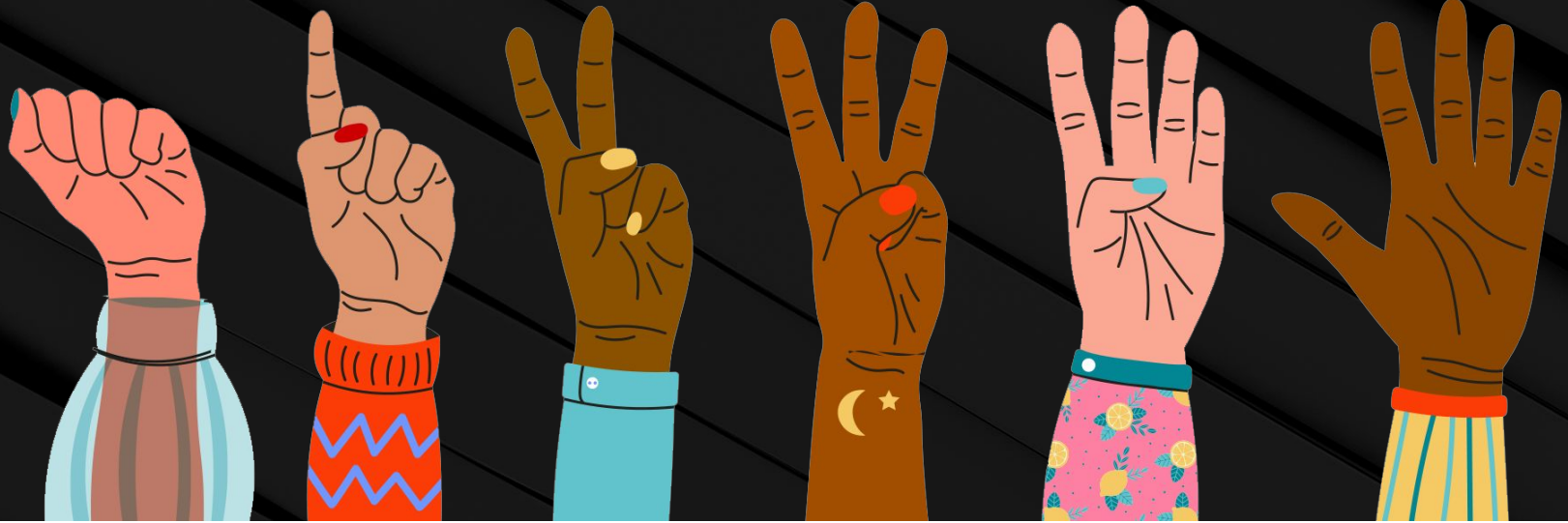
01 Create line, bar, pie, and scatter charts

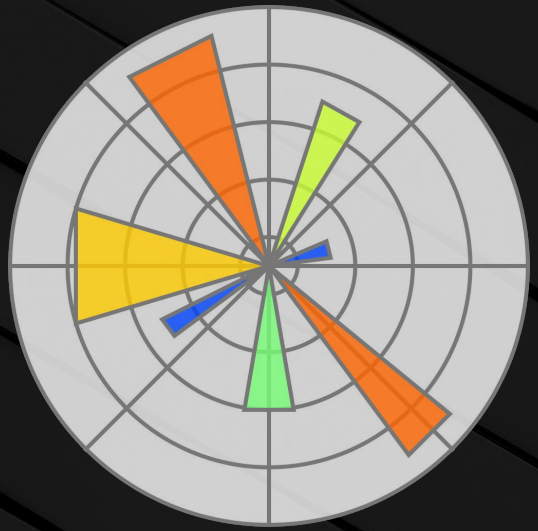
02 Add and modify chart features

Make sure you’ve downloaded any relevant class files!

FIST TO FIVE:

How comfortable do you feel with this topic?





What Is Matplotlib?

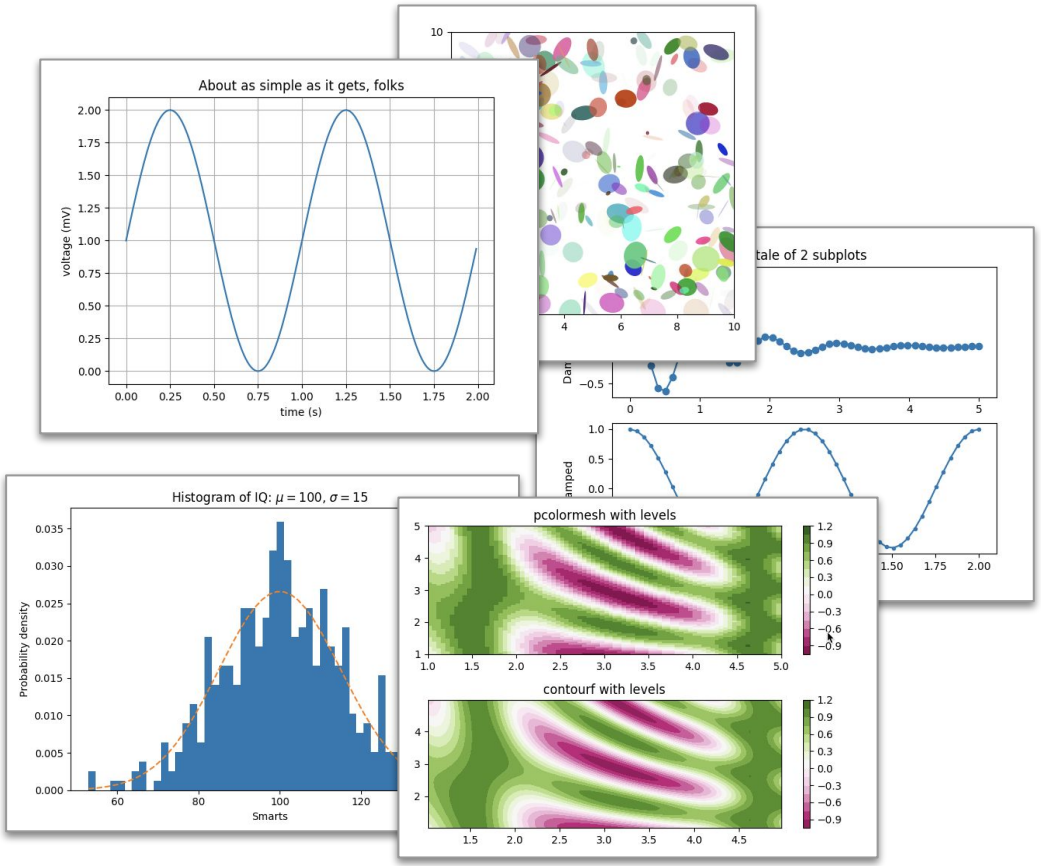
Matplotlib: A Python Library that Visualizes a Dataset

Types of datasets include:

- Pandas DataFrames
- Lists, tuples, and dictionaries
- NumPy arrays

Types of visualisations include:

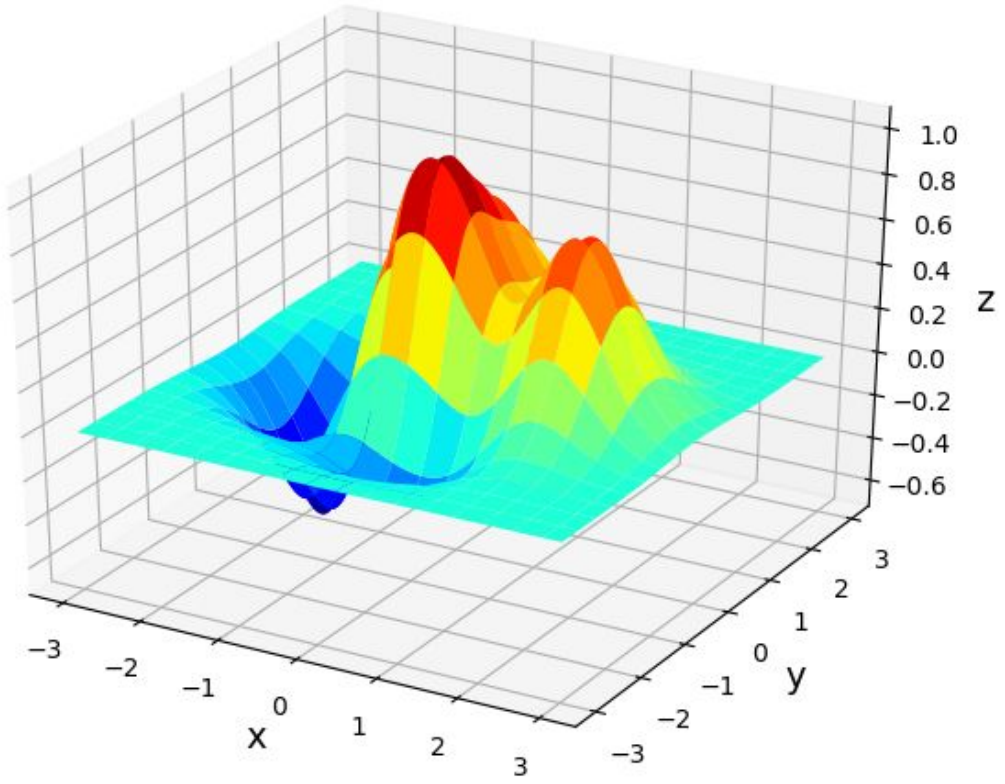
- Bar charts
- Pie charts
- Line charts
- Scatter plots
- And more!



The Pyplot Module = The Heart of Matplotlib

- Accepts many forms of input values
- Enables custom colours, shapes, labels, etc.
- Does most of the plotting logistics for us; we simply tell it which plot to make

🥰 Trust us: you'll love it!



General Plotting Process Using Pyplot

01

Create your dataset.
Data can be generated from functions, pulled from Pandas DataFrames, etc.

02

Generate your plot.
Use the `pyplot.plot()` function to tell Matplotlib what data to use and which plot to make.

03

Customize your plot.
Change the axes, label the figures, color the data points—make the plot as informative to the reader as possible.

Instructor Demonstration

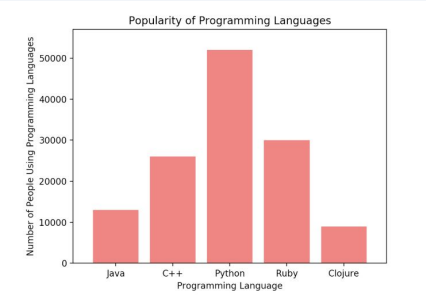
Different Plots

Questions?

Matplotlib: Not Just for Line Plots!

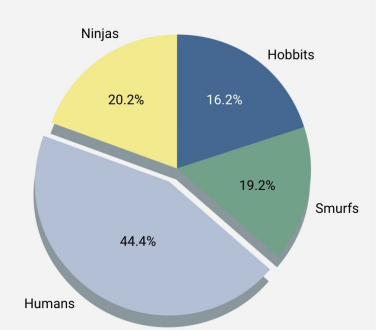
Bar Charts

Useful for comparing different entities with one another



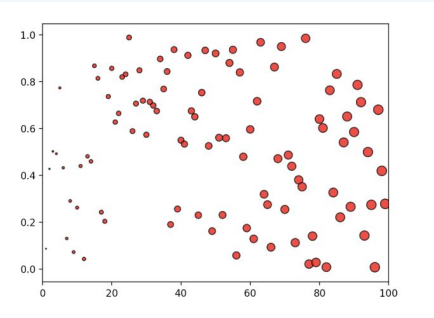
Pie Charts

Useful for demonstrating different elements of a complete dataset



Scatter Plots

Useful for displaying where values fall with respect to two factors



It’s very important to choose the right plot for a given dataset!

Instructor Demonstration

Bar Charts

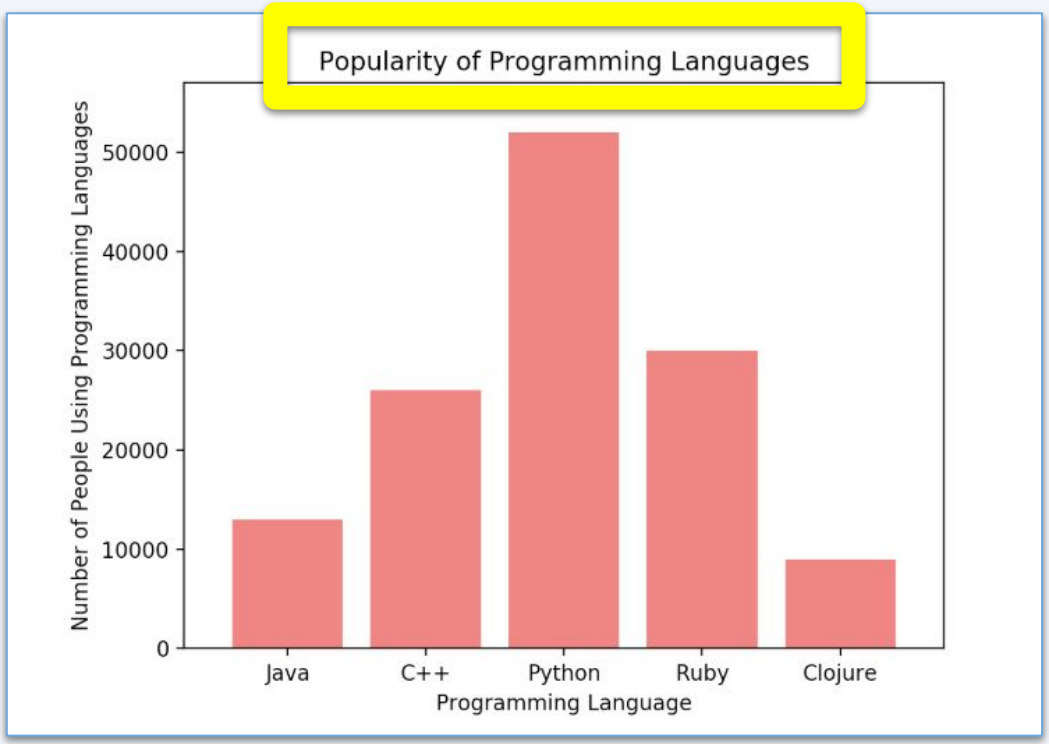
Bar Charts Help to Visualize Univariate Data

Univariate data refers to data with one variable, or one type of measurement.

Examples:

- Amount of rainfall, in millimetres
- Number of votes in a poll
- Number of people per category

Bar charts are particularly useful when a single variable is being counted multiple times.



Bar Charts

Bar charts are **NOT** effective for visualizing bivariate data.

Bivariate data refers to data with two variables. Anything you can plot as a line or scatter plot is bivariate data.

Example: A dataset comparing the number of ice cream bars sold versus daily temperature.

Think of other examples where a bar chart would be effective.

Activity: Bars Bar Chart

In this activity, you will create a bar chart that visualizes the density of bars within major US cities.

Suggested Time:

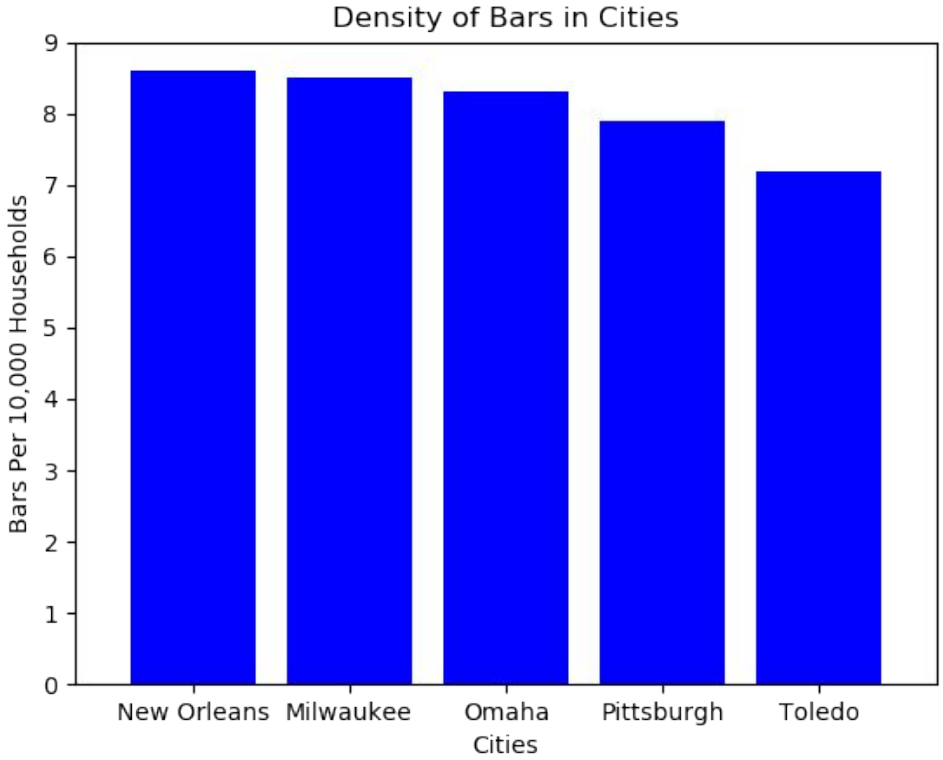
10 minutes

Bars Bar Chart Instructions

Using the starter code provided in your folders, recreate the figure as shown:

File:

`Unsolved/py_bars.ipynb`



Time’s Up! Let’s Review.

Instructor Demonstration

Pie Charts

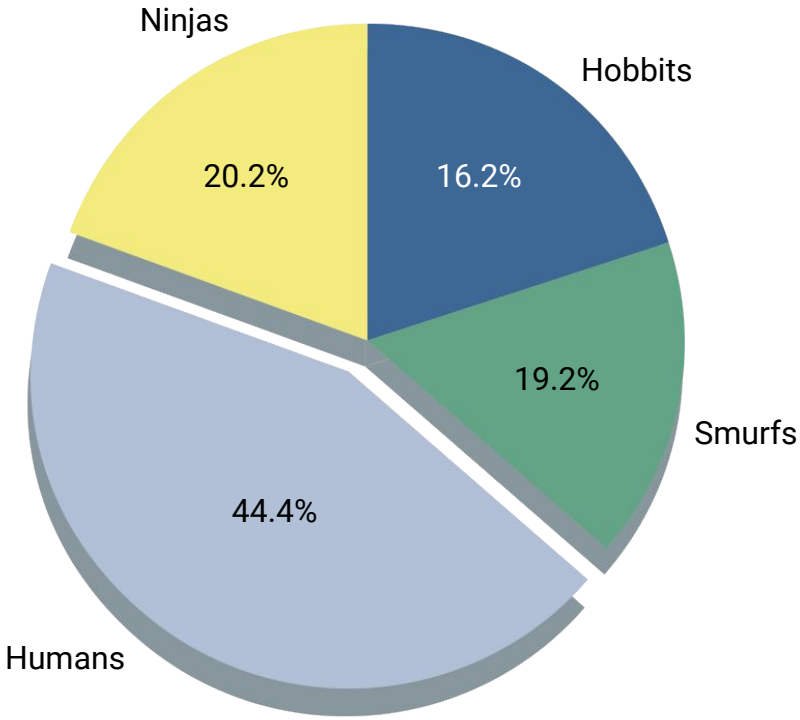
Pie Charts Help Visualize Simple Categorical Data

Pie charts are great for visualising data that is percentages or proportions.

Examples:

- Proportions of Democrat versus Republican versus independent voters
- Percentages of children's favorite story characters
- Distribution of left-handed versus right-handed pitchers in baseball

Fewer categories increase the effectiveness of a pie chart.



Pie Charts

Pie charts are **NOT** effective for large or multivariate datasets.

- With more than ~10 categories, pie charts become too crowded and lose effectiveness.
- Like bar charts, pie charts are only effective for visualizing univariate data.
- When in doubt, just use a bar chart.

Think of other examples where a pie chart would be effective.

Activity: Pies Pie Chart

In this activity, you will create a pie chart that visualizes the favorite pies of people in US.

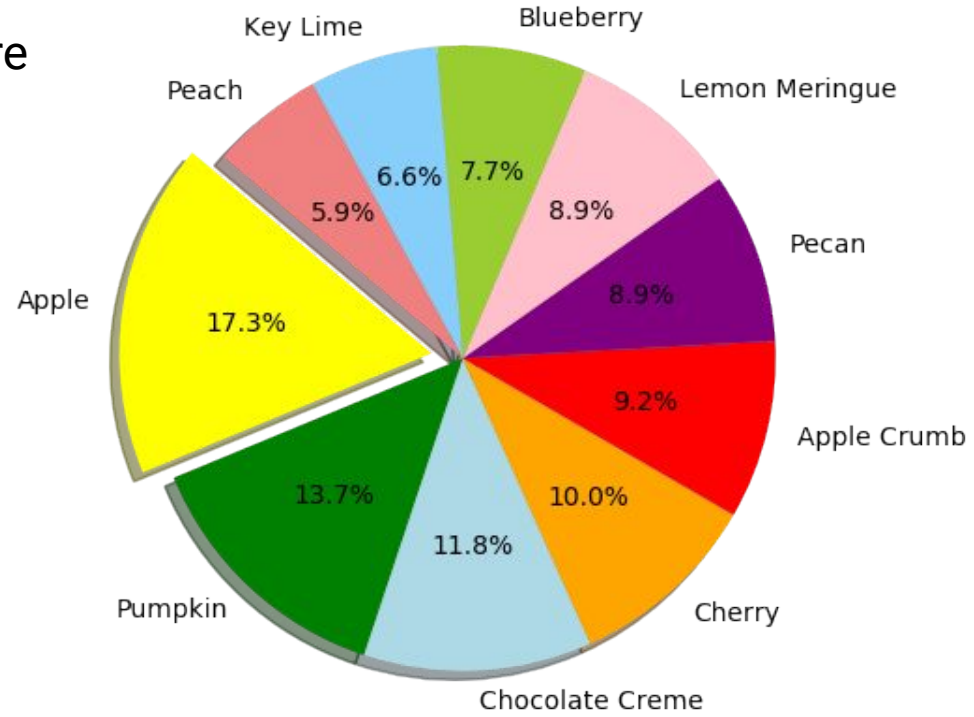
Suggested Time:

10 minutes

Pies Pie Chart Instructions

Using the provided starter code in your folders, recreate the figure as shown.

File: `iUnsolved/py_pie.ipynbi`



Time’s Up! Let’s Review.

Instructor Demonstration

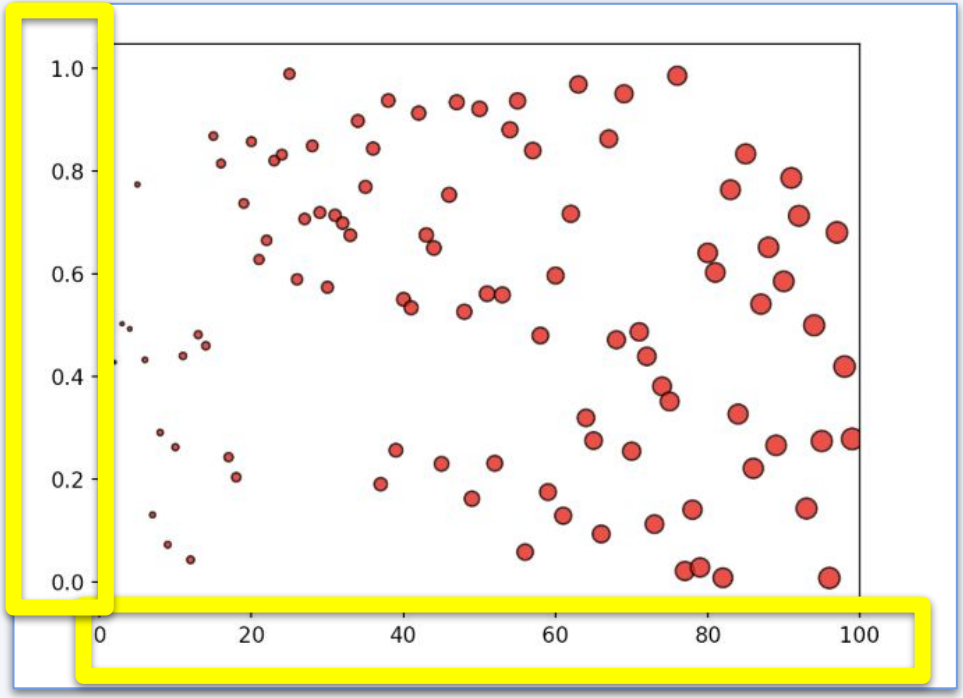
Scatter Plots

Scatter Plots = Powerful Visualisations for Bivariate Data

Bivariate data refers to data with two variables.

- Each data point is a combination of two variables
- Anything plotted on an x- and y-axis is bivariate data
- Example: The amount of ice cream sold per daily temperature
- Scatter plots are helpful for visualising large datasets (i.e., thousands of data points).

Scatter plots are frequently used to visualise clustering in a dataset.



Scatter Plots

Scatter plots are **NOT** effective for continuous measurements.

- When data is continuous, we'll often want to interpolate between measurements.
- The most common continuous data is time-series data.
- Scatter plots visualize "scattered" data, so interpolation is almost impossible.
- Line plots allow the audience to read between the data points.

Think of other examples where a scatter plot would be effective.

Activity: Scatter Py

In this activity, you will create a scatter plot that visualizes the relationship between ice cream sales and the increase in temperature.

Suggested Time:

10 minutes

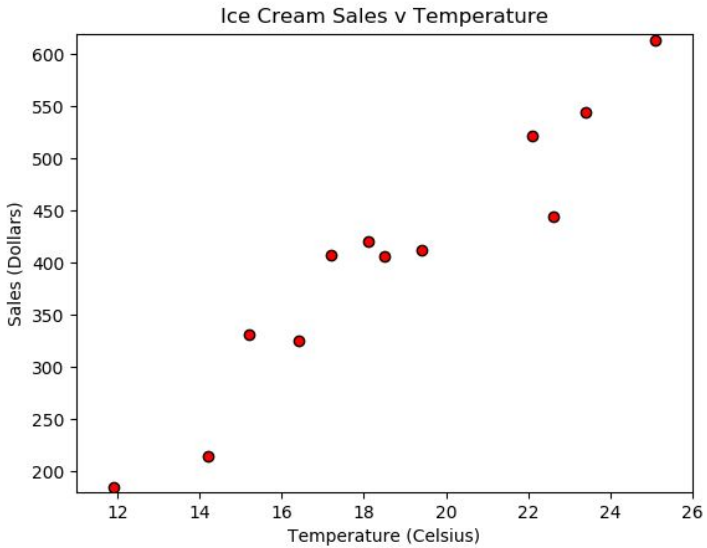
Scatter Py Instructions

Using the provided starter code in your folders, recreate the figure as shown.

File:
`iUnsolved/ice_cream_sales.ipynbi`

Bonus

Create a new list called scoop_price, fill it with values, and then set it so that the size of the dots are set according to those values.



Time’s Up! Let’s Review.

Questions?