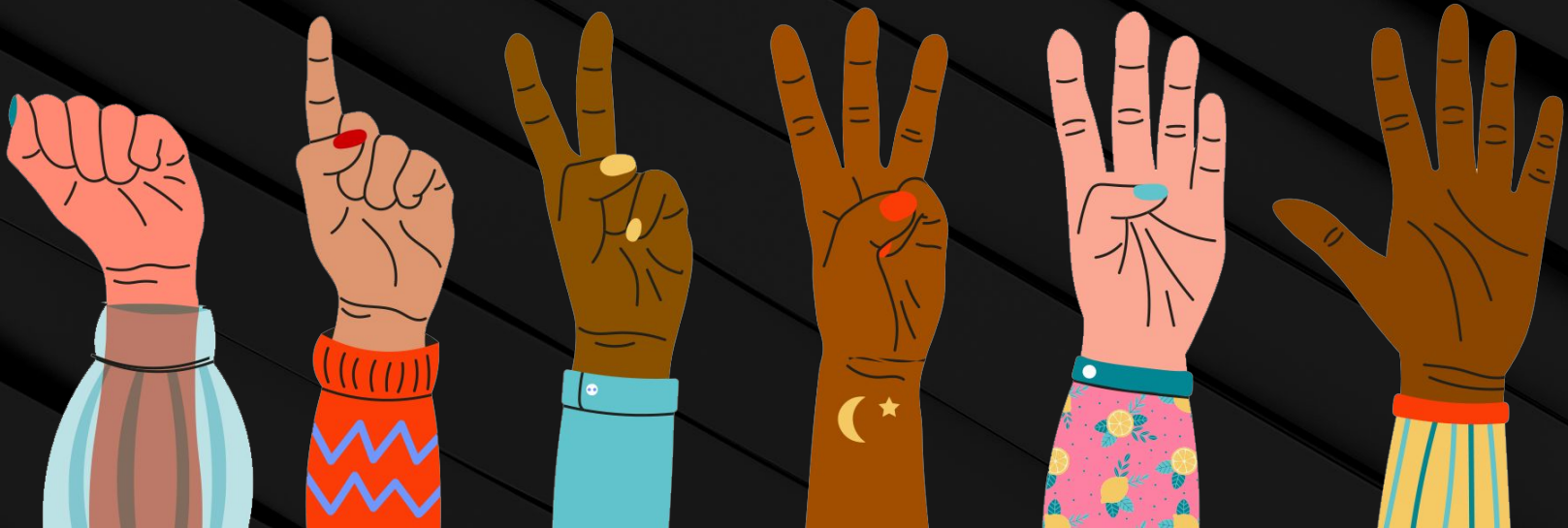# Plotting with Pandas

Data Boot Camp

Lesson 5.2

FIST TO FIVE:

How comfortable does everyone feel after the last class?

Warm-Up Activity: PyPlot

In this activity, you will use PyPlot to create the most effective visualisation for a variety of datasets.

Warm-Up Activity: PyPlot

Instructions	Open the unsolved activity file, and examine the starter code for each dataset.
	File: `Unsolved/plot_drills.ipynb`
	Determine what chart or plot best fits with the starter code for each dataset.
	Complete the code block to create a plot for each dataset.
	Be sure to provide each plot with a title and labels.
Hints	If you are unsure about what type of data is contained in the starter code, print it out!

Time's Up! Let's Review.

Instructor Demonstration

Plotting Pandas Data

Imaginary datasets are good to practise with...

...But we will deal with real-world data more often.

- Strange formats
- Messy
- Missing data
- Misleading headers

How to Work with Messy Data

Pandas enables us to quickly and easily:

Rename headers

Remove missing data

Convert and clean up column data

In most cases, we will work with real-world data in Pandas.

The Creators of Pandas Are Geniuses!

Most people who read datasets into Pandas will:

Clean and preprocess their data

Visualise their findings using Matplotlib

Pandas creators added Matplotlib functionality directly into Pandas, which:

Speeds up the process of creating lists and aesthetics

Still allows for PyPlot customizability

Best of both worlds

Time to <code>

Activity: Union Settlements

In this activity, you will create a bar chart that visualises the total number of major collective bargaining settlements by select unions in 1995.

Suggested Time:

20 minutes

Activity: Union Settlements

Instructions	Use Pandas to load the `union_settlements_1995.csv` dataset.
	Create a Series containing the number of settlements made by each union.
	Using this Series, create a bar chart with red bars.
	Adjust the figure size of the plot. Add a title and axis labels.
	Use your data to retrieve labels for your x-axis ticks, and rotate them 45 degrees.
	Display your plot. Which union agreed to the most settlements? The fewest?
Bonus	Find out how many national settlements were agreed to by each union, and plot the national totals alongside all settlement totals.

Time's Up! Let's Review.

Instructor Demonstration

Plotting Groups

How do we group data in Pandas?

Grouping and Summarizing in Pandas

The `dataframe.groupby()` function allows us to group data.

Data can be grouped by function or category.

The output of the function is a GroupBy object.

```
<pandas.core.groupby.groupby.DataFrameGroupBy object at 0x10cde6278>
```

	datetime	city	country	shape	duration (seconds)	duration (hours/min)	comments
state							
ak	311	311	311	311	311	311	311
al	629	629	629	629	629	629	629
ar	578	578	578	578	578	578	578
az	2362	2362	2362	2362	2362	2362	2362
ca	8683	8683	8683	8683	8683	8683	8683
co	1385	1385	1385	1385	1385	1385	1385
ct	865	865	865	865	865	865	865
dc	7	7	7	7	7	7	7
de	165	165	165	165	165	165	165
fl	3754	3754	3754	3754	3754	3754	3754

Grouping and Summarizing in Pandas

To return a DataFrame from a `GroupBy` Object

```
df.groupby('state').mean()
```

To return a Series from a `GroupBy` Object

```
states = df.groupby('state')
states = states['city'].mean()
```

Time to <code>

Activity: Library Usage

In this activity, you will create a pair of charts based on library usage collected from San Francisco.

Suggested Time:

20 minutes

Activity: Library Usage

Instructions

Open the starter file provided, and follow the prompts to import, split, and summarise the library usage dataset.

File: `Unsolved/library_usage.ipynb`

Create a bar chart, using Pandas and Matplotlib, that visualises how many library patrons checked out items from the library by patron type.

Create a pie graph, using Pandas and Matplotlib, that can be used to visualise the total items checked out by patron type of a single home branch, `'Home Library Definition'`.

Hint

Since there are quite a lot of patron types with minimal checkouts, you may like to include a filter to limit the minimum number of total checkouts by patron group.

Time's Up! Let's Review.

Countdown timer

15:00

(with alarm)

Activity: Litres per 100km

In this activity, you will create a scatter plot by using vehicle data, Pandas, and Matplotlib.
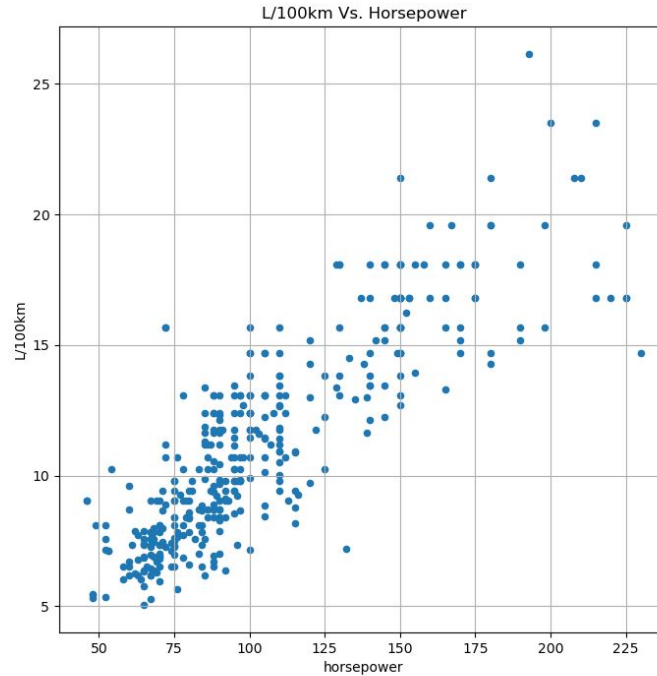
Suggested Time:

15 minutes

Activity: Litres per 100km

Create a scatter plot that compares the L/100km of a vehicle with its horsepower.

File: `Resources/Lp100km.csv`

L/100km Vs. Horsepower

Time's Up! Let's Review.

Instructor Demonstration

Plotting Multiple Lines

With Pandas, Adding Multiple Plots Is Easy!

We have learned how to use Pandas to:

Generate line, bar, pie, and scatter plots.

Plot summary figures by using `GroupBy` objects.

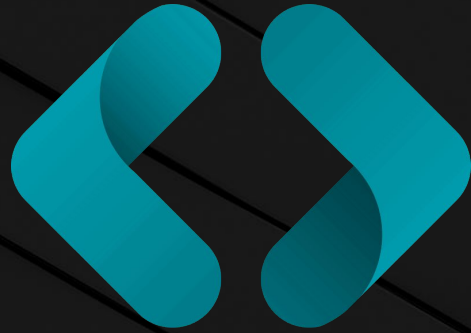But what if we want to combine multiple plots in the same figure?
Adding multiple plots to Pandas is the same as in PyPlot:

Add multiple plots in the same code block.

Show the combined figure by using `pyplot.show()`.

Time to <code>

Group Mini-Project: Travelling Companions

The remainder of class will consist of a group mini-project.
The mini-project will be broken up into three parts.
Every few minutes, we will review the previous part.

Group Mini-Project: Travelling Companions, Part 1

Follow the instructions for the activity in the unsolved Jupyter Notebook file.

File: `Unsolved/travelling_companions.ipynb`

Your output should align with the following table:

	COUNTRY OF NATIONALITY	2016 Alone	2016 With Spouse	2016 With Children	2016 With Family/Relatives	2016 Student Group	2016 With Friends	2016 With Business Associate	2016 With Incentive Group	2016 Others	...	2017 Others	2018 Alone	2018 With Spouse	2018 With Children	2018 With Family/Relative
0	SINGAPORE	18.6	35.5	3.9	23.7	0.1	14.7	2.2	0.4	0.8	...	0.5	16.9	34.6	5.0	25
1	THAILAND	30.1	16.7	0.9	12.0	3.1	30.2	4.8	1.0	1.2	...	1.8	20.6	6.7	7.2	33
2	INDONESIA	12.9	24.6	3.8	37.9	0.9	14.9	2.1	1.6	1.2	...	1.4	11.3	31.1	4.0	34
3	BRUNEI	10.2	24.9	2.2	44.5	0.0	16.2	1.1	0.1	0.7	...	0.0	17.5	28.1	2.0	40
4	PHILIPPINES	13.9	27.5	1.5	15.0	5.5	34.1	1.1	0.0	1.5	...	0.2	5.5	20.6	0.7	29

Time's Up! Let's Review.

Group Mini-Project: Travelling Companions, Part 2

Follow the instructions for the activity in the unsolved Jupyter Notebook file.

File: `Unsolved/travelling_companions.ipynb`

Your output should align with the following table:

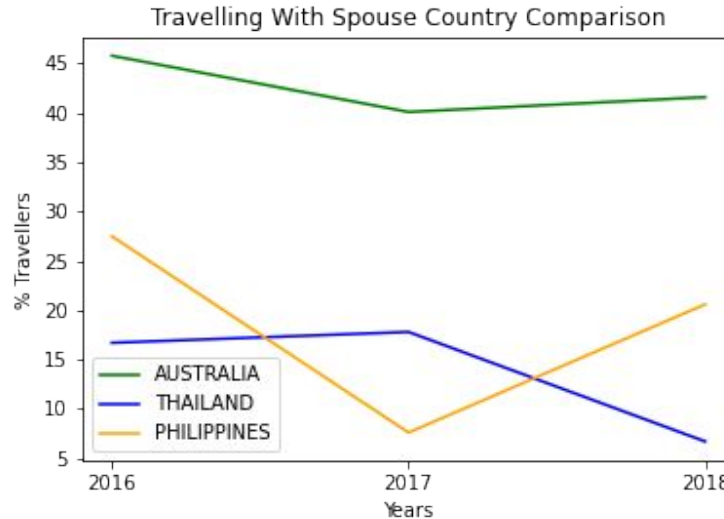| COUNTRY OF NATIONALITY	2016 Alone	2016 With Spouse	2016 With Children	2016 With Family/Relatives	2016 With Friends	2016 With Business Associate	2017 Alone	2017 With Spouse	2017 With Children	2017 With Family/Relatives	2017 With Friends	2017 With Business Associate	2018 Alone	2018 With Spouse
SINGAPORE	18.6	35.5	3.9	23.7	14.7	2.2	20.6	39.4	5.9	22.9	9.3	1.1	16.9	34.6
THAILAND	30.1	16.7	0.9	12.0	30.2	4.8	23.2	17.8	4.2	11.8	29.3	4.8	20.6	6.7
INDONESIA	12.9	24.6	3.8	37.9	14.9	2.1	10.4	28.0	3.9	34.9	15.5	2.7	11.3	31.1
BRUNEI	10.2	24.9	2.2	44.5	16.2	1.1	6.8	21.2	2.2	54.5	14.2	0.8	17.5	28.1
PHILIPPINES	13.9	27.5	1.5	15.0	34.1	1.1	3.5	7.6	0.2	40.0	46.4	1.9	5.5	20.6
VIETNAM	5.8	11.5	0.6	31.0	35.5	2.7	2.8	11.5	0.7	37.0	32.5	1.1	0.9	12.3
CHINA	4.2	30.9	0.7	37.8	18.4	1.3	1.9	30.4	0.7	42.0	20.4	0.5	1.9	23.9
JAPAN	10.9	26.5	0.6	16.4	18.4	25.6	8.4	30.3	0.0	20.0	25.4	8.5	4.2	22.7

Time's Up! Let's Review.

Group Mini-Project: Travelling Companions, Part 3

Follow the instructions for the activity in the unsolved Jupyter Notebook file.

File: `Unsolved/travelling_companions.ipynb`

Your output should align with the adjacent figure, depending on the user's input variable.

Time's Up! Let's Review.