

# Data Visualisation

# Overview

- Data Visualisation overview
- Chart types
- Jupyter Notebooks
- matplotlib

# Learning Objectives

- Understand what data visualisation is and when to use it
- Identify some data visualisation tools
- Discover some data visualisation techniques
- Gain insight into when to use Jupyter Notebook
- Implement visualisations using Jupyter and matplotlib

# What is Data Visualisation?

- A graphical representation of any data or information
- Visual elements such as charts, graphs, tables and maps are some tools that provide viewers with an easy and accessible way of understanding information
- Assists in understanding patterns or trends in data that otherwise wouldn't be spotted in raw data
- Used in all aspects of industries

# Why is it important?

- Easy to digest large quantities of information in a small space
- Establishes relationships between data
- Easy to share (image, web page, git repo etc.)
- Can be interactive with clicks, zoom and expanded areas
- Much more intuitive than raw data

# Data Viz Tools

There are many tools out there to help us create visualisations. Some popular ones include:

- Tableau: A business intelligence tool for visualisations
- Microsoft Power BI: Another business intelligence tool
- MatLab: A data analysis tool with an easy-to-use tool interface and graphical options

# Data Viz Tools

We'll be using two tools.

The first is called [Jupyter Notebook](#), an open-source web-based interactive computational environment.

The second is called [matplotlib](#), a plotting library in Python to generate visualisations from code.

You can use [matplotlib](#) on its own in Python or use it alongside Jupyter.

# Data Viz Techniques

- Know the audience / who is going to be viewing it
- Set goals for what you want to convey through the viz
- Choose a relevant chart type that best represents the data
- Use a colour scheme that represents different aspects of the data
- Use the best tools for the job

# Data Viz in Business

Data visualisations are not just for scientific research, we see them everywhere in life and business.

- Reports
- Displays
- Operation alerting
- Business growth / financial objectives
- Many more!

# Chart Types



Look at this graph! - Nickelback

# Graphs

- Used when you want to show patterns, trends and relationships in the numbers
- An excellent way to tell a story or to summarise something complex
- Can reveal insights that would otherwise be hidden
- Great for spotting errors or outliers in the data

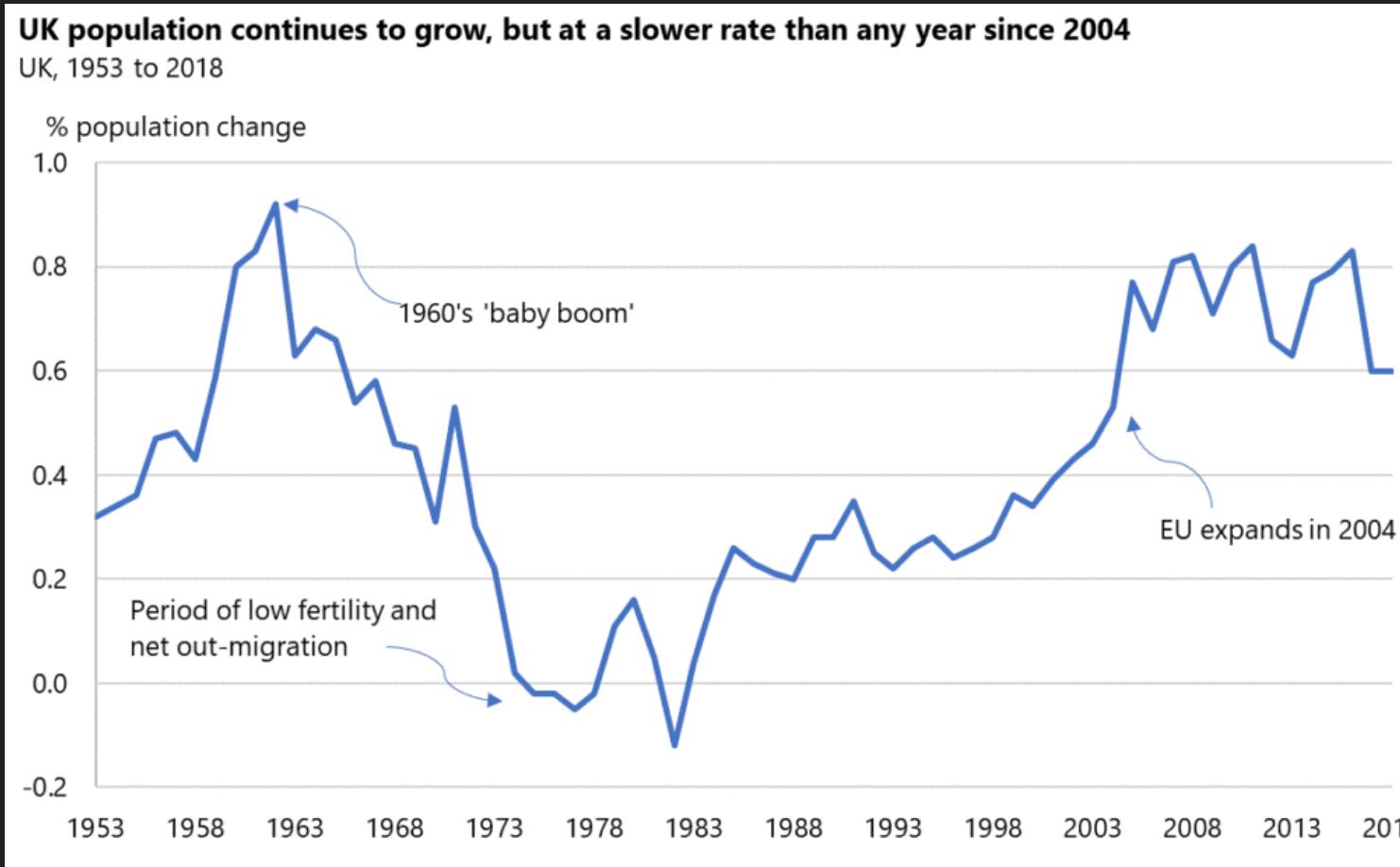
# Things to include in a graph

Title: Gives context to the viewer. Be concise and descriptive to summarise the main message.

Axes: Label each axis clearly to show what is plotted on each axes of the graph. You should include any relevant units. If the axis is obvious to the viewer, such as a range of year, a label may not be required.

Annotations: Add annotations to help further tell the story or highlight key areas.

# Example

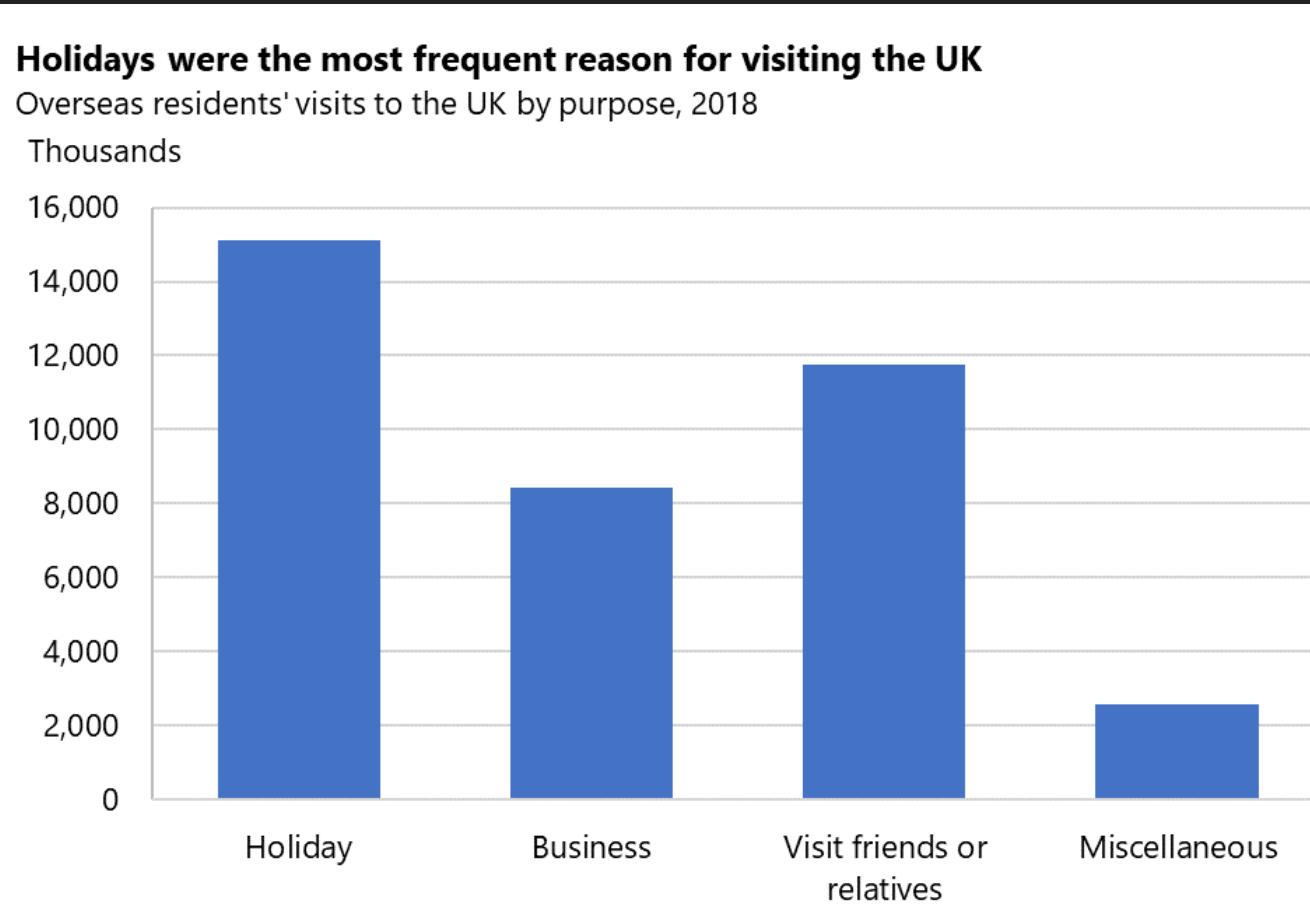


*Source: Office for National Statistics*

# Bar Graphs

- Work well for comparing the magnitude of different categories
- Can also be used to show time-series, deviation and distributions
- Have either horizontal or vertical bars

# Example

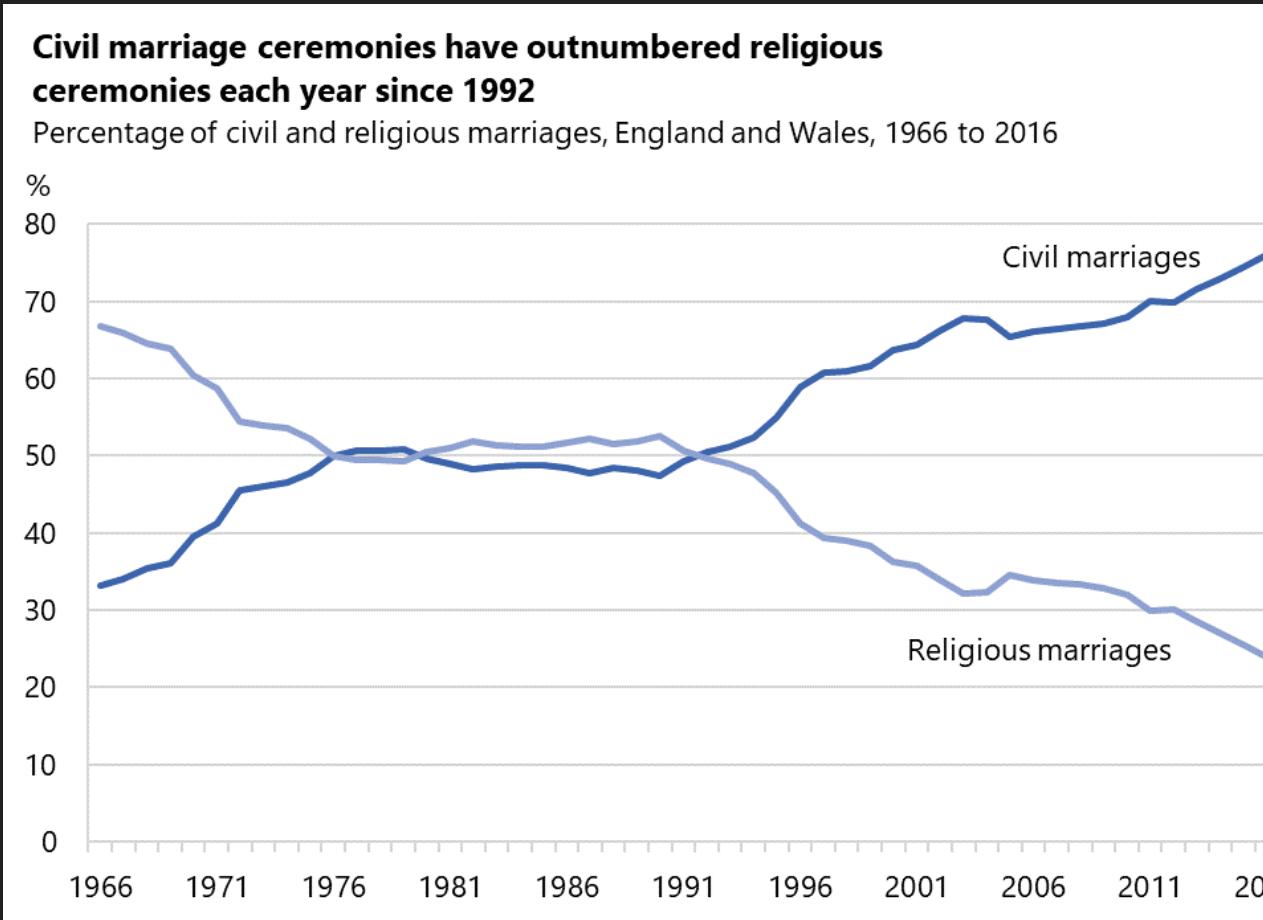


*Source: Office for National Statistics - International Passenger Survey*

# Line Graphs

- Displays information as a series of points connected by straight line segments
- Often used to visualise a trend in data over a time series

# Example



*Source: Office for National Statistics - Marriages in England and Wales*

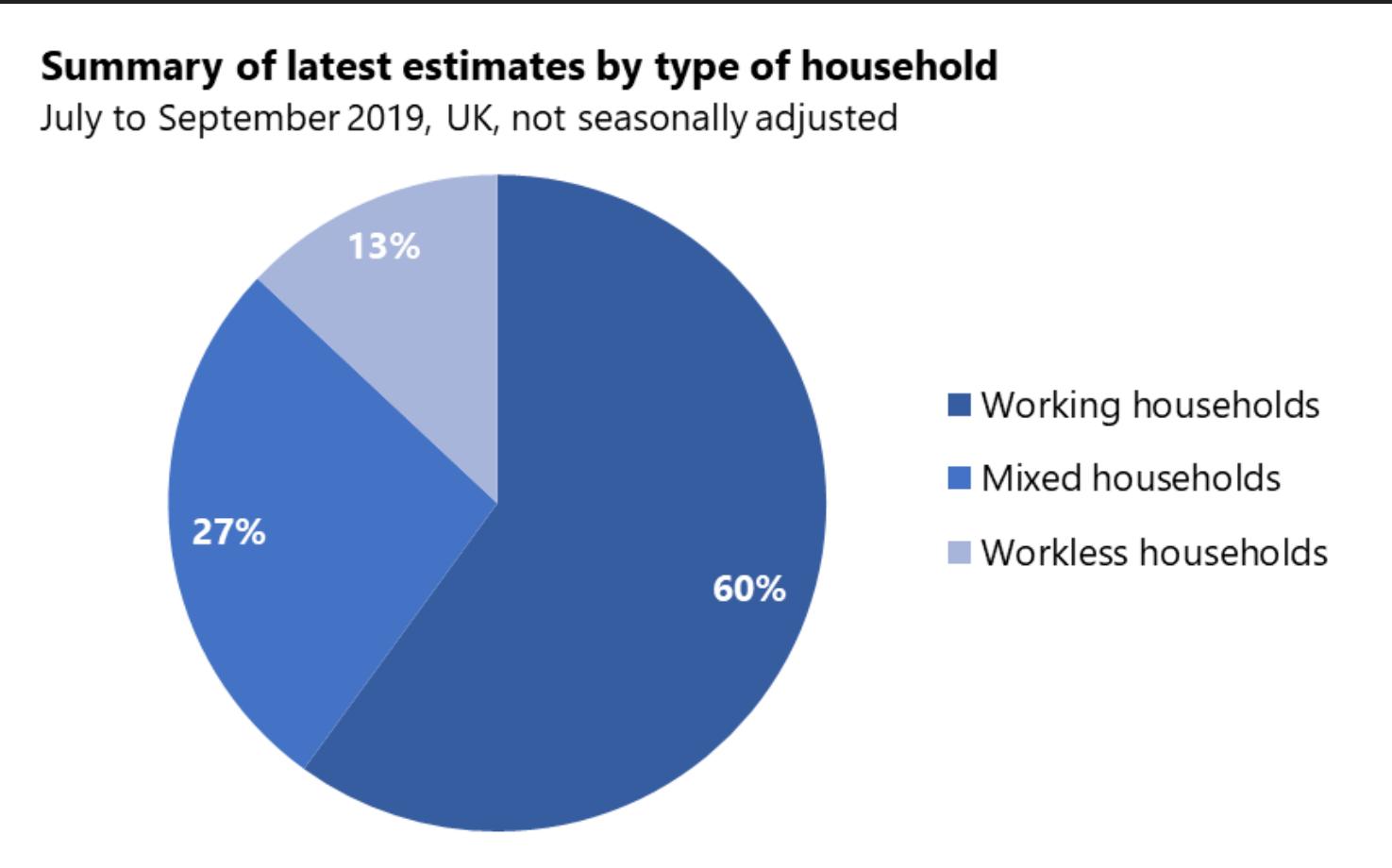
# Pie/Donut Chart

- Work well for showing part-to-whole relationships (how values can be split into parts)
- Clearly shows 'parts' add up to the 'whole'
- Donut charts differ slightly to pie - the central part is a convenient place to show the value of the total
- Bar charts can also show part-to-whole relationships, but are not as immediately obvious

Use one if:

- There are five or fewer categories
- The differences between categories are not significant
- You need to break up a page of bar graphs

# Example



*Source: Office for National Statistics - Household Labour Force Survey*

# Tables

- Used to present numbers in a clear and systematic way
- Harder for viewers to see patterns in a table than a graph
- Use a table to show multiple unrelated values at once. However, if the values are related then a graph would be more appropriate

A table is better if:

- You ask the viewer to compare individual values
- You want to include summary statistics such as means or totals
- You want to include values and measures such as percentages
- There is no trend / pattern / relationship between the values

# Example

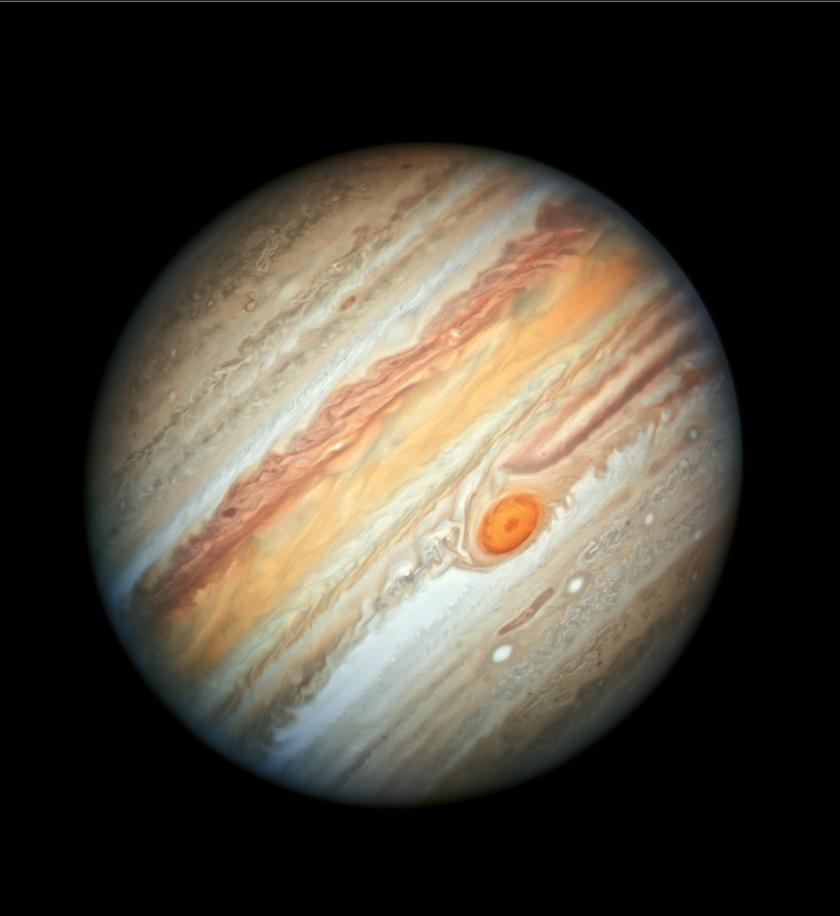
Table 1: Parliamentary constituencies with largest percentage change in electoral registrations, UK , December 2019 to March 2020

Code	Parliamentary constituency	Electoral registrations March 2020	Percent Change December 2019 to March 2020	Change December 2019 to March 2020	Change December 2018 to March 2020	Change December 2018 to March 2019
E14000832	Newcastle upon Tyne East	63,723	17.9%	9,681	8,264	-1,417
E14000602	Bristol West	99,859	13.7%	12,000	16,857	4,857
E14000919	Sheffield Central	89,266	12.4%	9,852	9,891	39
E14000780	Leeds North West	67,856	11.9%	7,215	9,303	2,088
E14000873	Oxford East	79,145	11%	7,841	8,105	264
E14000777	Leeds Central	91,069	10.8%	8,858	12,974	4,116
E14000549	Battersea	79,325	10%	7,210	9,980	2,770
E14000547	Bath	67,512	9.4%	5,824	7,790	1,966
E14000998	Tooting	76,986	7.9%	5,616	8,107	2,491
E14000831	Newcastle upon Tyne Central	58,302	7.9%	4,245	4,940	695

Source: Office for National Statistics, National Records of Scotland, Electoral Office for Northern Ireland – Electoral Statistics

*Source: Office for National Statistics - Parliamentary constituencies with largest percentage change in electoral registrations*

# Jupyter



No, not *that*!

# Jupyter Notebook

- An open-source web-based interactive computational environment that allows you to create live code, equations, visualisations, explanatory text and more

Can be used for:

- Data viz
- Data cleaning / transformation
- Statistical modelling
- Machine Learning
- And more!

# When should you use Jupyter Notebook?

Jupyter notebook runs python in the same way a python script runs python.  
So when would you use it over a simple python script?

- When you need to run your script on a computer that is not your laptop.  
If you needed to run the code on a server or in a docker container then jupyter will allow you to do that
- When you're presenting your data and want to show real-time data analysis
- When you want your code to be documented with mark-down
- When you want to use one of the features of jupyter notebooks e.g. magic commands like `%sql` (see example)
- when you're doing some exploratory coding and so only want to run your code one line at a time

# matplotlib

- An open-source plotting library for Python
- Provides an API for embedding plots into applications
- That's it!

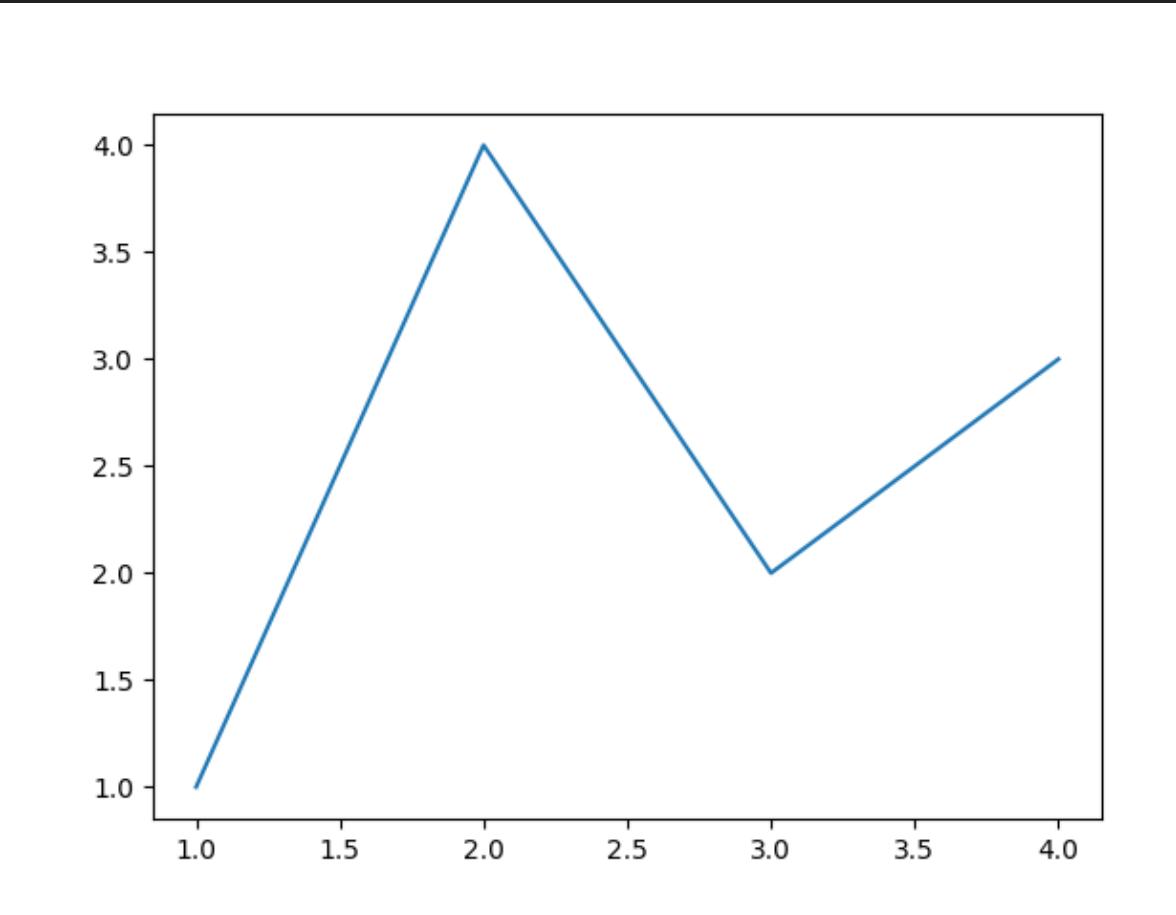
# Example

Matplotlib can plot your data on [figures](#). The simplest way of creating a figure is using `pyplot.plot` to plot a graph and `pyplot.show` to display the graph to the user as a GUI:

```
import matplotlib.pyplot as plt

# Create a figure containing a single axes
plt.plot([1, 2, 3, 4], [1, 4, 2, 3])
# Display the plot
plt.show()
```

# Output



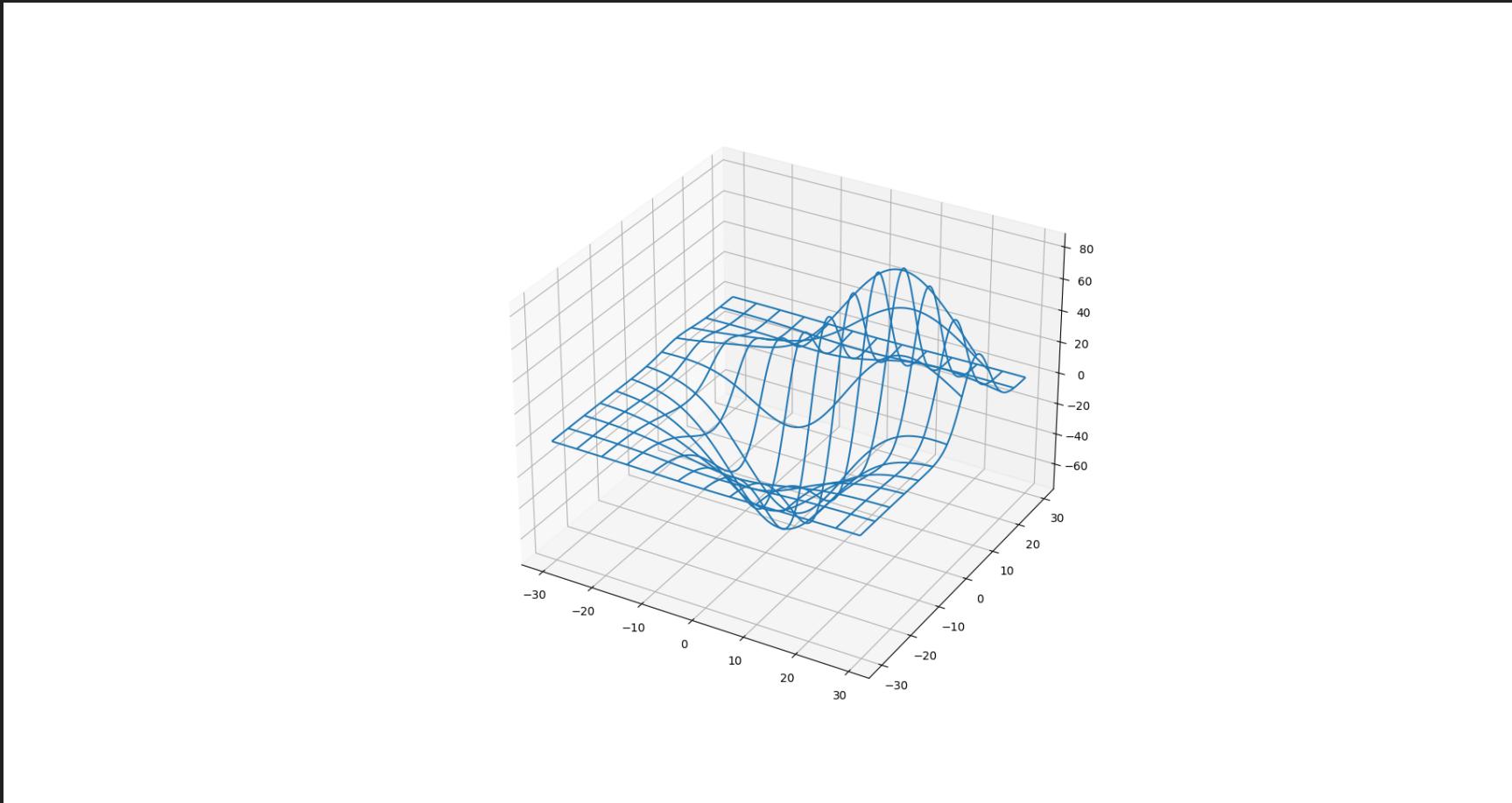
# Example 2

`pyplot` supports most data representations, including pie, bar, line, 3D, subplots, scattergrams and more. This example creates a 3D graph by adding a 3D projection sub plot with `add_subplot` and plotting to it with `plot_wireframe`.

```
from mpl_toolkits.mplot3d import axes3d
import matplotlib.pyplot as plt

# Create a new figure
fig = plt.figure()
# Add a 2D Axis onto the figure
ax = fig.add_subplot(111, projection="3d")
# Create some test data.
X, Y, Z = axes3d.get_test_data(0.05)
# Plot a the 3D wireframe. Stride defines the wire mesh size.
ax.plot_wireframe(X, Y, Z, rstride=10, cstride=10)
# Show the plot to the user.
plt.show()
```

# Output



Quiz Time! 😎

Which data visualisation would be best for showing a trend in data over a time series?

1. Bar chart
2. Line graph
3. Pie chart
4. Table

Answer: 2

Which data visualisation would be best for when there is no clear trend or pattern in the data?

1. Bar chart
2. Line graph
3. Pie chart
4. Table

Answer: 4

# Exercise

Follow the steps in the exercise handout.

# Learning Objectives Revisited

- Understand what data visualisation is and when to use it
- Identify some data visualisation tools
- Discover some data visualisation techniques
- Gain insight into when to use Jupyter Notebook
- Implement visualisations using Jupyter and matplotlib

# Terms and Definitions Recap

**Data Visualisation:** An interdisciplinary field that deals with the graphic representation of data.

**Chart:** A graphical representation of data, in which the data is represented by symbols, such as bars in a bar chart, lines in a line chart, or slices in a pie chart. A chart can represent tabular numeric data, functions or some kinds of quality structure and provides different information.

# Further Reading

- [Jupyter Official Website](#)
- [matplotlib Official Website](#)
- [matplotlib Guide](#)