



HACKWAGON
• ACADEMY •

DATA SCIENCE 102: VISUALISATION



AGENDA

- Data Visualisation
 - Use Cases
 - Types of Visualisations
- Matplotlib
 - Attributes
 - Design
 - Subplots
- Pandas
 - Coded Examples
- Good Practices

LEARNING OBJECTIVES



- Able to use Pandas to build intuitive visualisation
- Understand the good practices of data visualisation
- Learn

DATA VISUALISATION

- Applications of regression



MATPLOTLIB

- Intuition
- Interpretation
- Terminologies
- Performance Measures
- Ordinary Least Squares

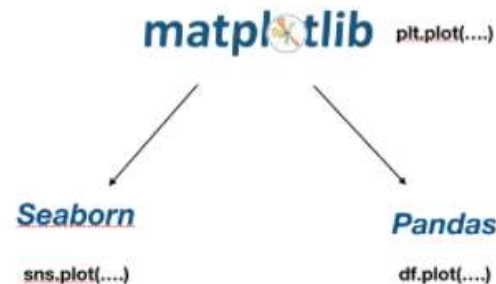


HACKWAGON
• ACADEMY •



Matplotlib Introduction

- Matplotlib is one of the first and many visualisation libraries
- Provides a standard for other visualisation libraries
- Leveraged by many other libraries
- Documentation is difficult to understand
- Restrictive syntax



Plotting with Pandas





Plotting with Pandas

- Using Pandas, many of Matplotlib's functions are sorted within `.plot` function
- Switch between different kinds of plots easily using `kind=` parameter

```
df.plot(kind= <chart>,  
        x = <column>,  
        y = <column>,  
        ... more matplotlib settings  
        )
```

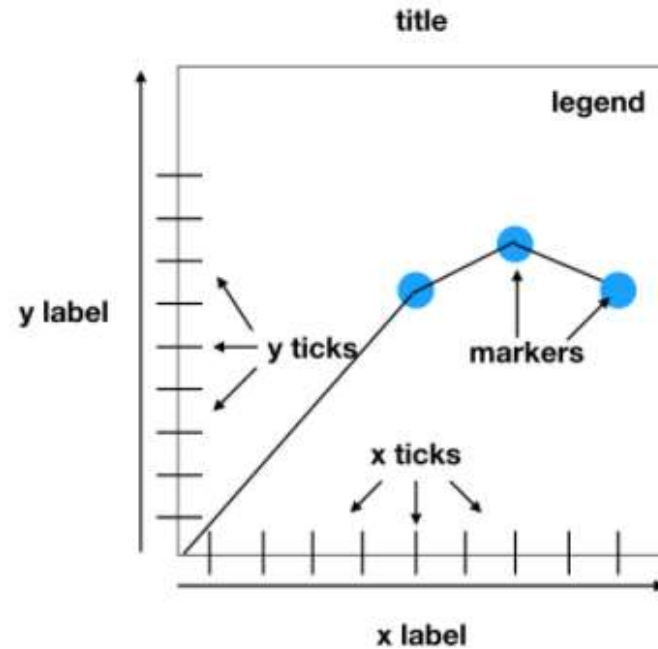

Anatomy of a Plot



Anatomy of a Plot



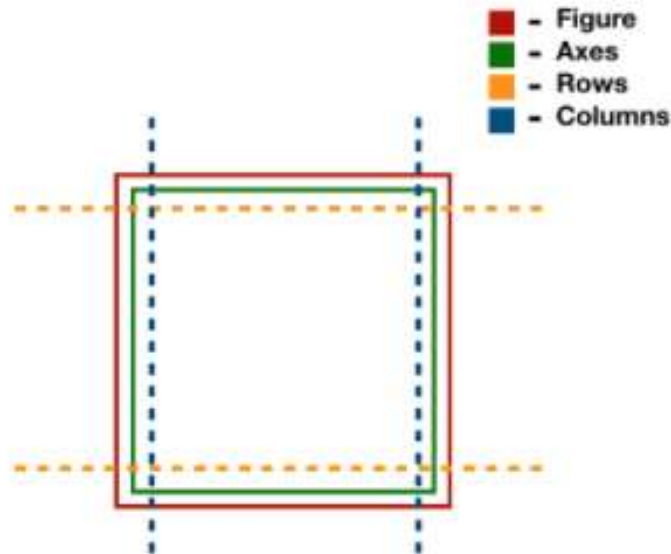
- Title
- Legend
- Axes Labels
- Tick Values



Anatomy of a Plot - Subplots



- To plot multiple plots within a single figure, need to understand columns, rows and index that identify them





DATA SCIENCE 102:

Text Mining

AGENDA



- Unstructured vs. Structured Data
- NLP & Definitions
- Text Mining Process
 - Simple Cleaning
 - Tokenisation
 - Stemming
 - Lemmatisation
 - Stop Word Removal
- Sentiment Analysis
 - VADER

LEARNING OBJECTIVES



- How to process textual data
- Applications of text mining libraries and tools

STRUCTURED VS. UNSTRUCTURED DATA



STRUCTURED DATA



- Structured data can be put into table format with rows and columns
- This kind of representation is called a schema representation
- Simple statistical analysis can be performed
 - such as mean, median, mode

| | country | points | price |
|---|-------------|--------|-------|
| 0 | France | 91 | 48.0 |
| 1 | Australia | 82 | 10.0 |
| 2 | US | 92 | 48.0 |
| 3 | Italy | 88 | 68.0 |
| 4 | New Zealand | 89 | 17.0 |
| 5 | US | 86 | 32.0 |
| 6 | Austria | 89 | 25.0 |
| 7 | US | 89 | 50.0 |
| 8 | US | 86 | 10.0 |
| 9 | US | 92 | 65.0 |

UNSTRUCTURED DATA



- Unstructured data cannot simply be stored in a table format and immediately analysed
- Examples of these includes images, videos and text
 - Images may vary in size and resolution
 - Text may vary in number of words and variety of words



TOP stories



Singapore
Court challenge filed on 377A arguing that gay sex law 'violates human dignity'



Singapore
13 fresh chicken distributors fined record S\$26.9m for price fixing, market sharing



Asia
Millions at risk as Philippines braces for strong Typhoon Mangkhut

LATEST news

Hide ▾

Sport

Usain Bolt toasts zero-gravity with champagne

10 minutes ago



Business

Takeda to pack up Illinois HQ, nearly 1,000 workers to be affected

10 minutes ago



World

Spain to go ahead with sale of bombs to Saudi - El Periodico

10 minutes ago



World

Germany's Merkel condemns far-right xenophobia, Nazi slogans

10 minutes ago



World

Libya closes Tripoli airport again after rocket fire

10 minutes ago



NLP & DEFINITIONS

- What is NLP
- Corpus
- Document
- Term



HACKWAGON
• ACADEMY •



WHAT IS NLP

- NLP is Natural Language Processing
- Finding useful insights from unstructured textual data
- Use cases include
 - Sentiment Analysis
 - News Categorisation
 - Title Inference
 - Search Engine Optimisation
 - Spam/Non-spam Filter
 - Questions & Answers

TEXT MINING PROCESS

- Simple Cleaning
- Tokenisation
- Stemming
- Lemmatisation
- Stop Word Removal



HACKWAGON
• ACADEMY •

SENTIMENT ANALYSIS

- Sentiment Analysis
- VADER



HACKWAGON
• ACADEMY •

THANK YOU!



HACKWAGON
• ACADEMY •