Data Science

Unit 2-01: Data Visualization

AKADEMI GA

GENERAL ASSEMBLY

COURSE CONTENT

Week 1 : Data Science Foundations

Installation and Github, Python fundamentals, Introduction to Pandas

Congratulations!

Week 2 : Working with Data

More pandas, basics of probability and statistics, Exploratory Data Analysis (EDA), working with data, use statistical analysis and visualisation

Week 3 : Data Science Modeling

Linear regression Train/Test/Split, Classification, Logistic Regression

Week 4 : Data Science Applications

Using APIs, Natural Language Processing, Time Series Analysis

Week 5: Final Presentation

Present your capstone project

Review of Week1: Data Science Foundations

Previously, we have covered:

- a review of Python fundamentals

- Introduction to Pandas

- Data Joining, cleaning, manipulation with Pandas

Week 1 Units
1-01 Installation and Github
1-02 Python Review and Practice
1-03 List Comprehension
1-04 Introduction to Pandas
1-05 Data Wrangling

Week 2: Working with Data

- *In Unit 2, we will continue to use Pandas for data visualization.*
- *We will also review basic statistical concepts such as probability, confidence intervals and hypothesis testing.*

Week 2 Units
2-01 Data Visualization
2-02 Data Transformation
2-03 Probability Distributions
2-04 Confidence Intervals
2-05 Hypothesis Testing

Schedule

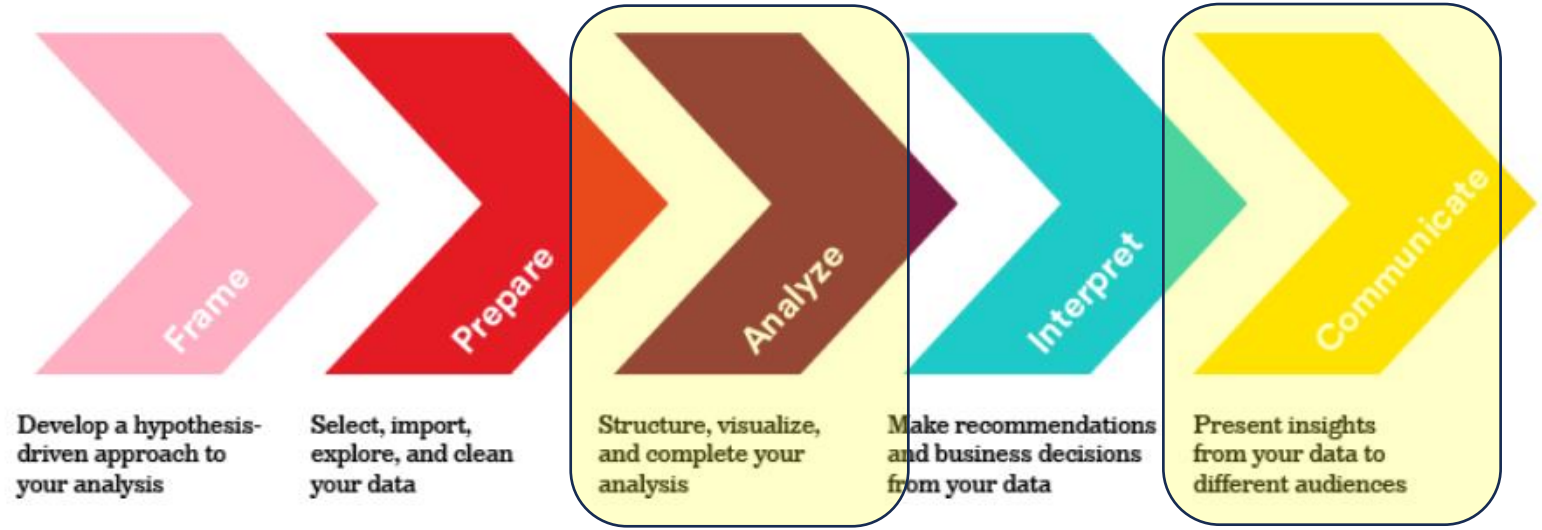
Time	Topics
5:00 - 6:30	Lesson 1: Basic plots with Matplotlib
6:30 - 6:45	Break
6:45 - 7:45	Lesson 2: Using Seaborn
7:45 - 8:00	Wrap up and Q&A

Our Learning Goals

In this lesson, we will learn how to:

- Create basic plots with Matplotlib
- Create plots with Seaborn

Data Visualization



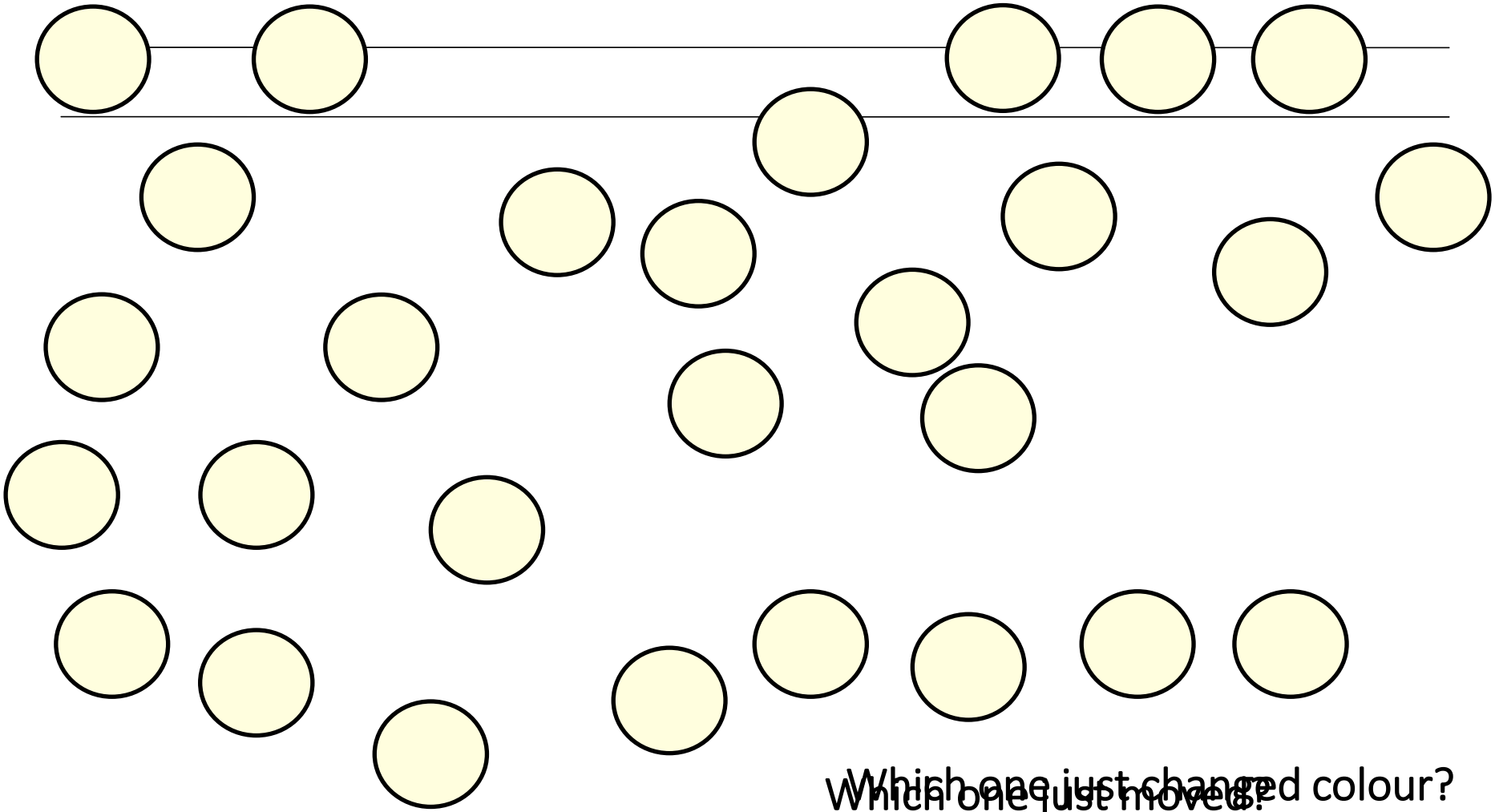
Data Visualization can help us :

- **Analyse the data** by visualizing the patterns quickly
- **Communicate** our results when telling our data story

Movement and shape changes

The next slide will be shown for only a few moments
Ready? 3… 2… 1…

Ready? 3…2…1…

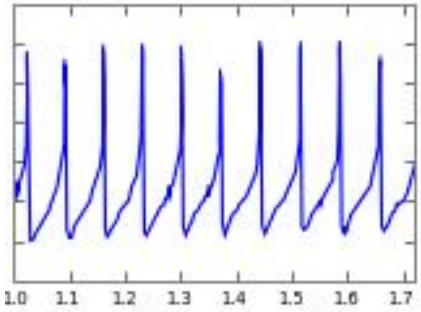

Which one just changed colour?
Which one just moved?

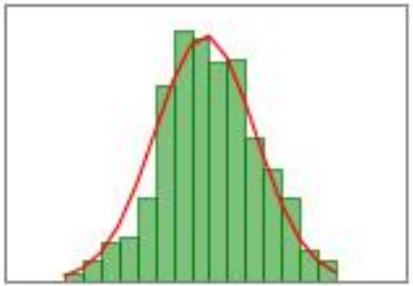
COMMON TOOLS FOR VISUALISATION

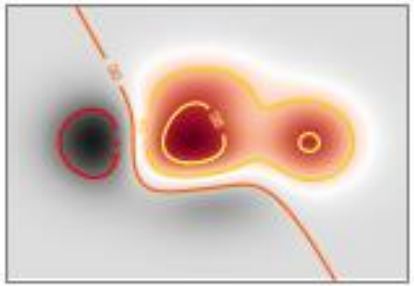
WHY MATPLOTLIB?

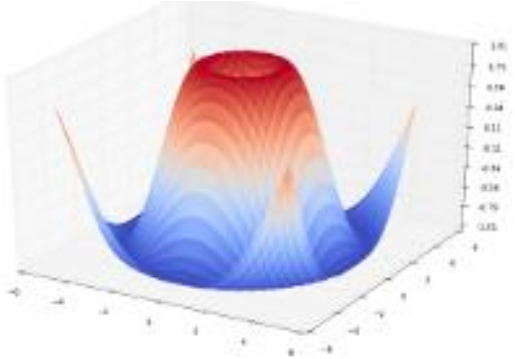
- **Python graphics plotting library**
- Plots numpy arrays
- Pandas DataFrames and Series behave like numpy arrays

matplotlib









Unit 1-05 Data Visualization

Lesson 1: Basic Plots with Matplotlib

AKADEMI GA

GENERAL ASSEMBLY

Pandas and Matplotlib

Pandas

- Library for wrangling data
- Pandas uses Matplotlib for creating plots
- Can use Matplotlib functions with Pandas to format plots after drawing

Matplotlib

- Library for creating plots
- Matplotlib uses Pandas DataFrames and Series as its data
- Many, many types of plots and functions

Plotting with Pandas and Matplotlib

- We will use the Matplotlib's pyplot module which contains functions for customizing our plots :

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # set the plots to display within the Jupyter notebook
5  %matplotlib inline
6
7  # set stylesheet
8  plt.style.use('ggplot')
9  # You can try other styles such as 'classic', 'ggplot', grayscale'
10
```

Plotting with Pandas and Matplotlib

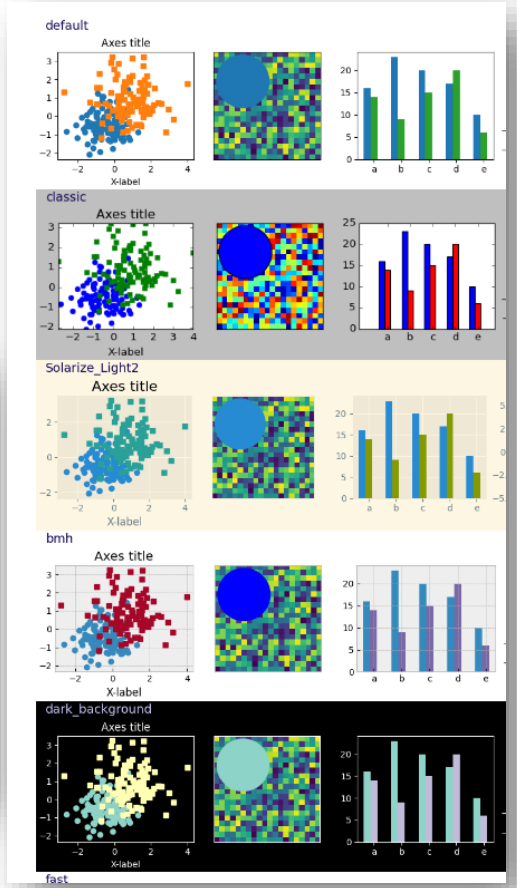
- Set the option to display the plots within the Jupyter notebook

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # set the plots to display within the Jupyter notebook
5  %matplotlib inline
6
7  # set stylesheet
8  plt.style.use('ggplot')
9  # You can try other styles such as 'classic', 'ggplot', grayscale'
10
```

Plotting with Pandas and Matplotlib

- Set the style to use for the plots

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # set the plots to display within the Jupyter notebook
5  %matplotlib inline
6
7  # set stylesheet
8  plt.style.use('ggplot')
9  # You can try other styles such as 'classic', 'ggplot', grayscale'
10
```

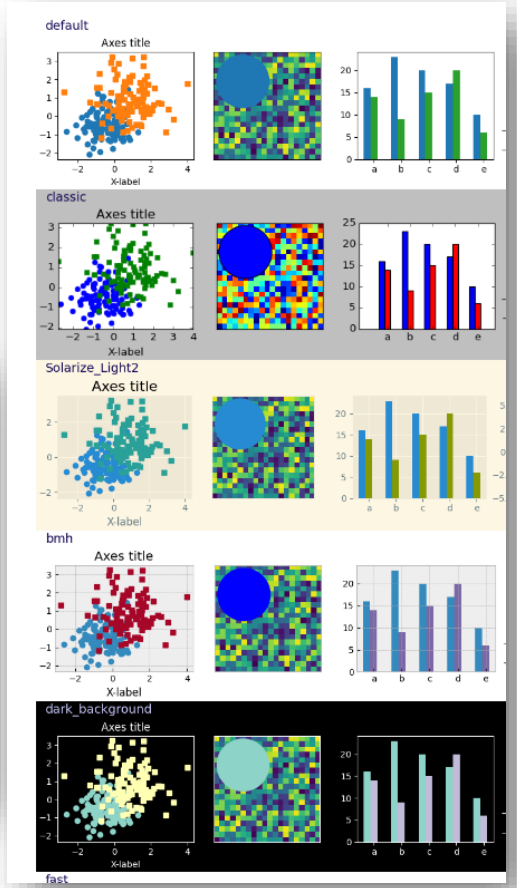


Sample Matplotlib Styleshee

Plotting with Pandas and Matplotlib

- Set the style to use for the plots

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # set the plots to display within the Jupyter notebook
5  %matplotlib inline
6
7  # set stylesheet
8  plt.style.use('ggplot')
9  # You can try other styles such as 'classic', 'ggplot', grayscale'
10
```

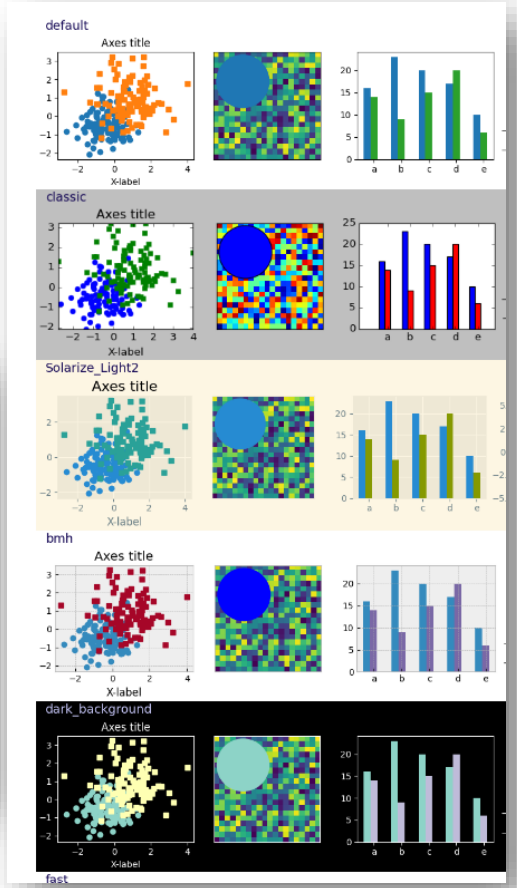


Sample Matplotlib Styleshee

Plotting with Pandas and Matplotlib

- Set the style to use for the plots

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3
4  # set the plots to display within the Jupyter notebook
5  %matplotlib inline
6
7  # set stylesheet
8  plt.style.use('ggplot')
9  # You can try other styles such as 'classic', 'ggplot', grayscale'
10
```



Sample Matplotlib Styleshee

Notebooks

- Unit 1-05 Lesson 1: Data Visualization
- Exercises
  - Plotting_lesson

Q&A

Schedule

Time	Topics
5:00 - 5:45	Lesson 1: Basic plots with Matplotlib
6:30 - 6:45	Break
6:45 - 7:45	Lesson 2: Using Seaborn
7:45 - 8:00	Wrap up and Q&A

Unit 1-05 Data Visualization

Lesson 2: Plotting with Seaborn

AKADEMI GA

GENERAL ASSEMBLY

seaborn

The Seaborn Library

- Seaborn is a library for making statistical graphics in Python.
- Advantages of Seaborn
  - Built on top of Matplotlib
  - Integrated with Pandas DataFrame objects
  - High level functions for more complex plots

Notebooks

- Plotting with Seaborn

Homework

- Complete the Seaborn Exercises

Recap

In this unit, we:

- Used Pandas and Matplotlib to create and format
  - Histograms
  - Scatter plots
  - Line plots
  - Bar plots

Looking Ahead

Homework:
Visualizing the Diamonds Dataset
Try out the other types of plots and share on the chat!

Next: Data and Transformation