Data Analytics Immersive

Data Visualization With pandas

Unit: Data analysis with Python

GENERAL ASSEMBLY

Learning Objectives

In this lesson, we'll:

- **Explain** the characteristics of a great data visualization.
- **Identify** when to use a bar chart, pie chart, line chart, scatterplot, or histogram.
- **Use** pandas to implement line charts, bar charts, scatterplots, and histograms.

Any questions?

Agenda

Data Visualization Best Practices (recap)

Using pandas To Visualize Data

Wrap Up and Q&A

Data Visualization With pandas

Data Visualization Best Practices (Recap)

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Change in average income since 1960 for American adults.

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Change in average income since 1960 for American adults.

Answer: Line charts are ideal for expressing change over time.

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Amount of sales per state.

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Amount of sales per state.

Answer: Bar charts are best for comparing numbers.

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Determine if there's a correlation between book length and sales.

Knowledge Check!

Which type of data visualization should we use for the scenario below?

Scenario: Determine if there's a correlation between book length and sales.

Answer: Scatterplots can compare the relationship between two variables.

Data Visualization With pandas

Using pandas To Visualize Data

pandas ❤ Matplotlib

pandas DataFrame objects use another library, known as Matplotlib, behind the scenes.

This means you can use Matplotlib functions in combination with pandas methods to alter plots after drawing them.

For example, you can use Matplotlib's `xlabel()` and `title()` functions to label the plot's x axis and title, respectively, after it is drawn.

Source: matplotlib

Our First Chart

Once you've loaded data into a pandas DataFrame, creating a chart is as simple as using the `.plot()` method.

```
import pandas as pd
import matplotlib.pyplot as plt
data_frame = pd.read_csv(file_address)
data_frame['column_name'].plot()
```

Plot Parameters

You may want to alter certain aspects of the chart, such as:

- The **kind** of plot you want (line, bar, scatter, etc.).
- The **style** of the lines, including color and line consistency.
- The size of the chart, or `figsize`.
- ... plus many other settings.

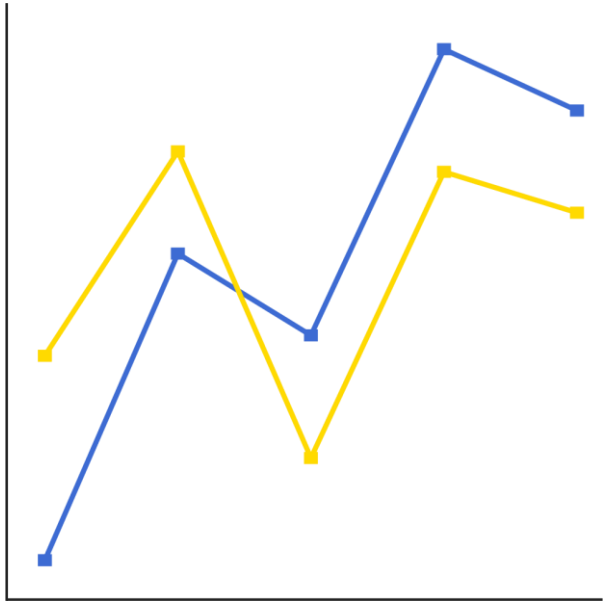
Customizations can be made using keyword parameters:

```
data_frame['column_name'].plot(style={'col1': 'r'},
figsize=(16,9))
```

Guided Walk-Through:
5.1 Line Charts in pandas

20 minutes

Let's practice creating line charts in Section 5.1 of the workbook.



Discussion:

Counting Games per Country

2 minutes

If we alter our chart from using the “year” column to using the “country” column, all of a sudden the line chart stops making sense.

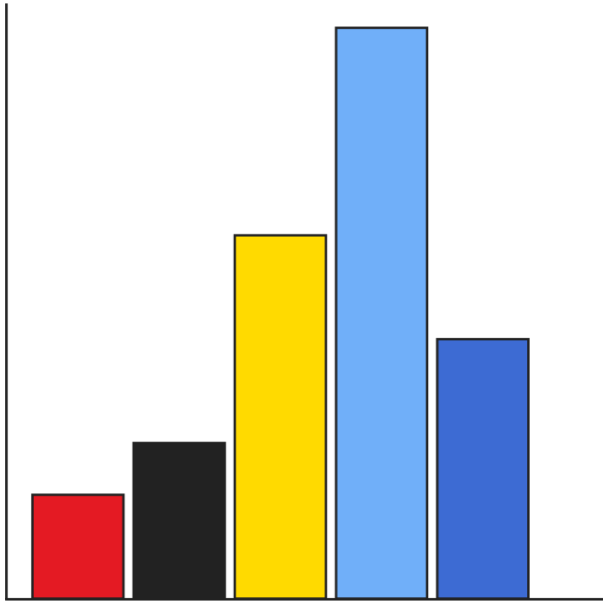
What chart should be used instead to compare the amount of games per country?

Partner Exercise:

5.2 Bar Charts

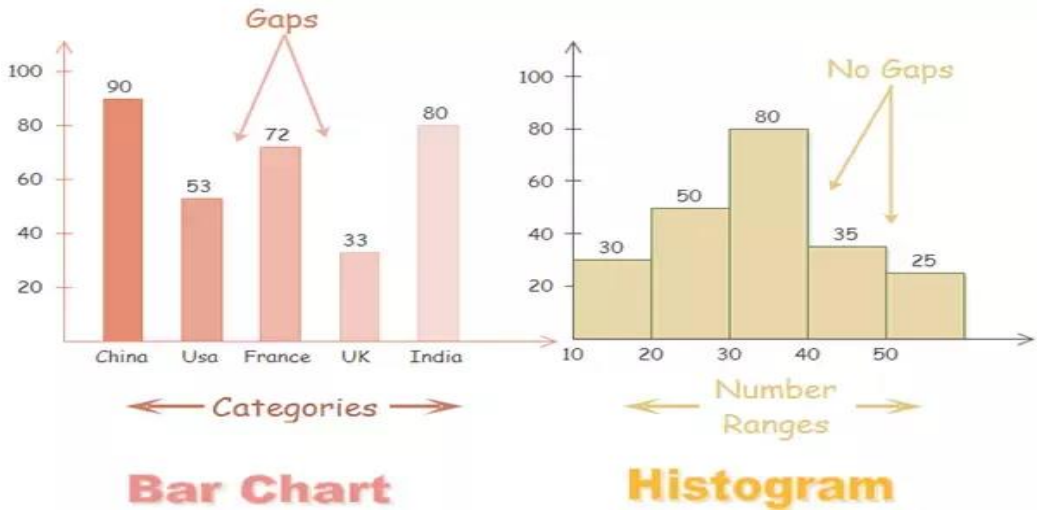
20 minutes

Let's use the same data set to start creating bar charts in Section 5.2.



Bar Charts vs. Histograms

Another common chart style is a **histogram**, which plots the distribution of values according to numerically defined groups rather than distinct categories.

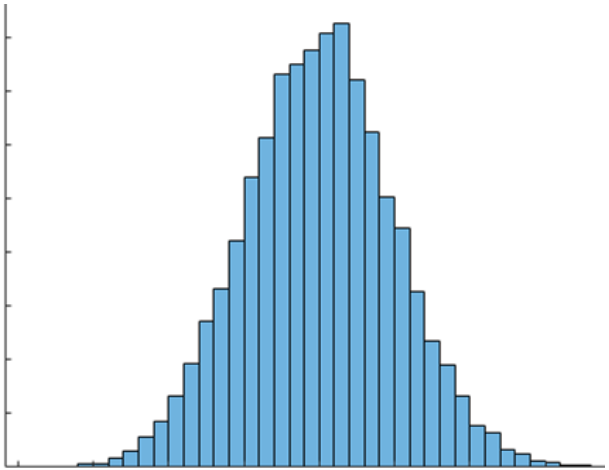


Source: Edrawsoft

Guided Walk-Through:
5.3 Histograms

20 minutes

Let's look at some of the challenges of histograms in Section 5.3.



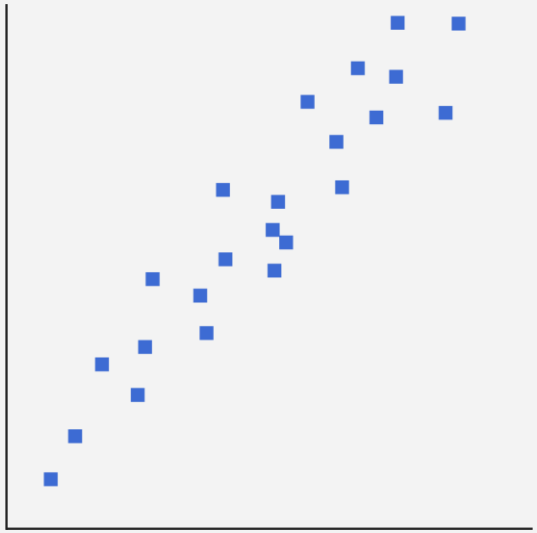
Source: Mathworks.com

Scatterplots

Scatterplots intend to demonstrate the correlation, or lack thereof, between different variables. Therefore, we have to specify which columns to compare:

```
data_frame.plot(kind='scatter',
x='column_a', y='column_b')
```

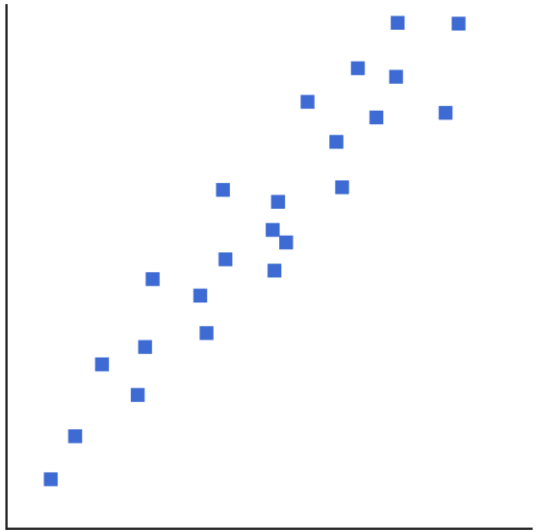
Scatterplots are most useful when values are **continuous**, rather than discrete with large gaps.



Guided Walk-Through:
5.4 Scatterplots

20 minutes

Let's practice using scatterplots to investigate correlations in Section 5.4.

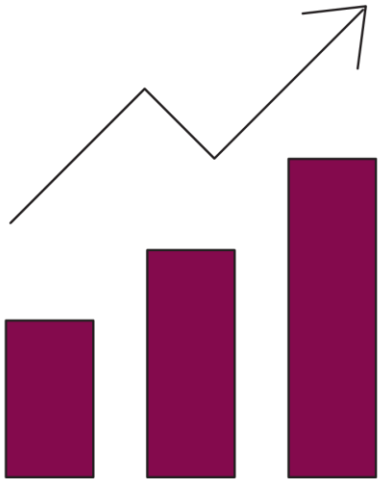


Group Exercise:
5.5 Visual Storytelling With Data

30 minutes

Working with the Superstore data set, use **exploratory data analysis methods** and **at least one data visualization** to communicate trends, outliers, and a hypothesis surrounding the data.

1. Be sure to include a title and a label on the x and y axes.
2. Share you visualization in slack when you are done.



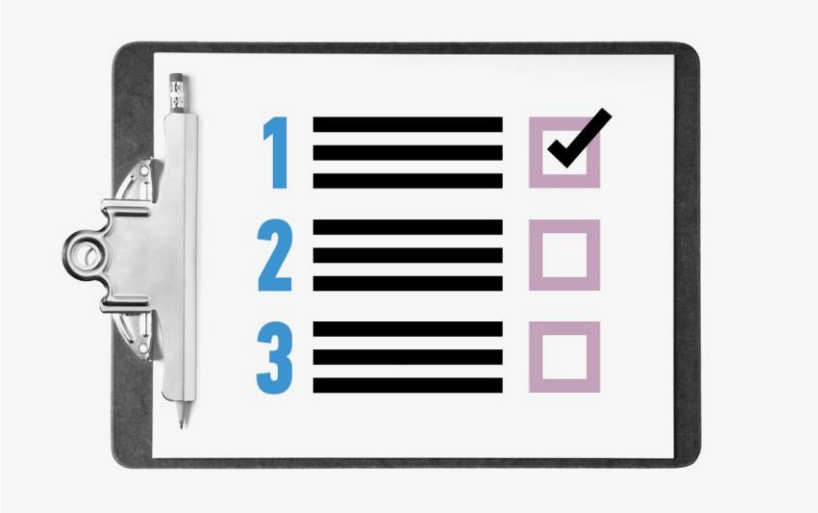
Data Visualization With pandas

Wrapping Up

Recap

Today, we:

- Explained the characteristics of a great data visualization.
- Identified when to use a bar chart, pie chart, line chart, scatterplot, or histogram.
- Used pandas to implement line charts, bar charts, scatterplots, and histograms.



Q&A


GA