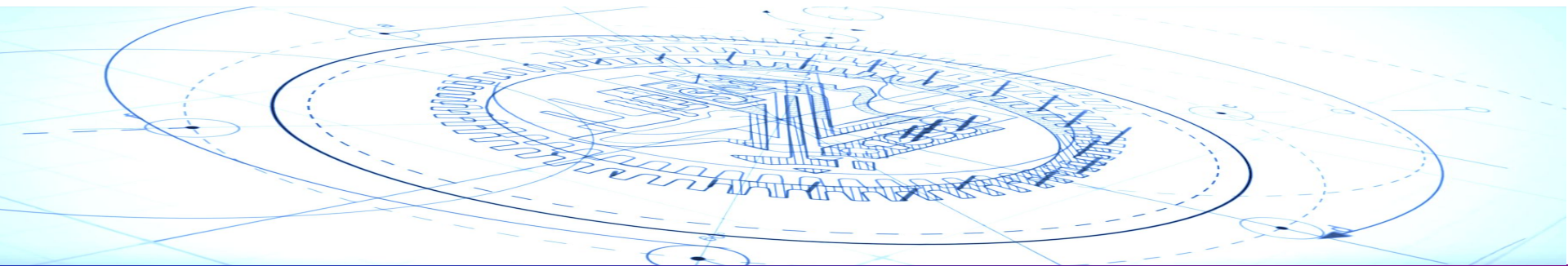# Pandas & Matplotlib

機器人碩士 藍裕棋

# Before We Start

- Install Pandas and Matplotlib

```
pip install pandas matplotlib
```

# Pandas

- A Python Package used for data manipulation and analysis
- Common data structures are DataFrames and Series
- Commonly used functions
  - pd.read_csv(file_path): Read a CSV file
  - pd.read_excel(file_path): Read an Excel file
  - pd.read_json(file_path): Read an JSON data
  - pd.Dataframe(data): Convert data to DataFrame
    - data is usually a dictionary
  - df.to_csv(file_path): Save the DataFrame to a CSV file

# Pandas

- DataFrame-related functions for data inspection:
  - df.head(n): shows the first n rows of data (default 5)
  - df.tail(n): shows the last n rows of data
  - df.info(): summarizes DataFrame structure
  - df.describe(): provides statistical summary

# Pandas

- DataFrame-related functions for data manipulation:
  - df.drop(columns = [col]): Removes specified columns
  - df.dropna(): Drop rows with missing values
  - df.fillna(value): Replace missing values with a specified value
  - df.sort_values(by = 'col'): Sorts df by a column
  - df.groupby('col'): Groups data by a column for aggregation
  - df.merge(df2, on = 'key'): Merges 2 df on a common column
  - df.pivot_table(): Creates a pivot table for summarizing data

# Pandas

- DataFrame-related functions for filtering and selction:
  - df.loc[]: Accesses rows/columns by labels
  - df.iloc[]: Accesses rows/columns by integer indices
  - df[df['col'] > value]: Filters rows based on a condition

# Matplotlib

- A Python library for visualization
- Most commonly used module is pyplot
  - import matplotlib.pyplot as plt
- Often used with Pandas and NumPy for visualizing data

# Matplotlib

- Commonly used plt functions for plot creation:
  - plt.plot(x, y): Creates a line plot
  - plt.scatter(x, y): Creates a scatter plot
  - plt.bar(x, height): Creates a bar chart
  - plt.hist(data): Plots a histogram
  - plt.pie(sizes): Creates a pie chart

# Matplotlib

- Commonly used plt functions for plot customization:
  - plt.xlabel(), plt.ylabel(): Set x- and y- axes labels
  - plt.title(): Set the plot title
  - plt.legend(): Displays a legend
    - Legend stands for names and their related icon in the plot
  - plt.grid(True): Add a grid to the plot
  - plt.xticks(), plt.yticks(): Customize axis ticks (unit)

# **Matplotlib**

- Commonly used plt functions for figure and subplot:
  - plt.figure(figsize = (w, h)): Creates a new figure with specified size
  - plt.subplot(rows, cols, index): Adds a subplot to a grid
  - plt.tightlayout(): Adjusts subplot spacing

# Matplotlib

- Commonly used plt functions for displaying and saving:
  - plt.show(): Displays the plot
    - press q to close the window if you run python file
    - in jupyter notebook, the plot will show under the code block
  - plt.savefig(file_path): Saves the plot to a file

# Today's Work

- Read a CSV file
- Make a analysis
- Ask TAs for demo!