

# Data Visualization with Matplotlib & Seaborn

KANN Bonpagna

*Data with Python, 2025*

## 1 Setup

Load the libraries and set theme:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme(style="whitegrid")
one_colour = sns.color_palette("crest", 1)[0]
# You can put any other color from the palette.
```

## 2 Saving a figure to your device.

Save a graph to a directory. If the directory does not exist, a new folder will be created automatically.

```
save_dir = "figures" # your folder path
file_name = "most_rated_movies.png"
full_path = os.path.join(save_dir, file_name)

os.makedirs(save_dir, exist_ok=True) # create folder tree if missing
plt.savefig(full_path, dpi=300, bbox_inches="tight")
# bbox_inches="tight" : trims extra whitespace around the axes.

print(f"Figure saved to: {full_path}")
```

### 3 Bar Chart (Implicit)

```
plt.figure(figsize=(15,8))
sns.barplot(data=top10_counts,
            x='title',
            y='num_ratings',
            color=one_colour)
plt.title('Most-Rated Movies', fontsize=20)
plt.ylabel('# ratings', fontsize=12)
plt.xlabel('Movie Title', fontsize=12)
plt.xticks(rotation=45, ha='right', fontsize=12)
plt.yticks(fontsize=12)
plt.tight_layout()
plt.show()
```

### 4 Multi-bar Chart

```
colors = sns.color_palette()
sns.barplot(data=sales_by_price, x='Day', y='Num_sales', hue='Price', palette=colors)
```

### 5 Line Graph

```
sns.lineplot(data=monthly, x='date', y='num_ratings', color=one_colour)
```

### 6 Scatter Plot

```
sns.scatterplot(data=lemonade, x='Flyers', y='Sales', color=one_colour)
```

### 7 Box Plot

```
sns.boxplot(data=lemonade, x='Month', y='Sales', color=one_colour, showmeans=True)
# Set showmeans=True to see the mean point on the box plot.
```

### 8 Histogram

```
sns.histplot(lemonade['Sales'], bins=10, color=one_colour, edgecolor='0.2', kde=True)
# "bin" is the number of bars, set "kde" to True to show the line representation.
```

## 9 Multiple Graphs in a Figure

```
categories = ['Rainfall', 'Sales', 'Flyers']
titles = ['Rainfall (mm)', 'Daily Sales', 'Flyers Distributed']

#Create three subplots with 1 row and 3 columns
# figsize is the size of the entire figure. Set sharey=True if you want to use only one y
# -axis label for all graphs
fig, axes = plt.subplots(1, 3, figsize=(15, 4), sharey=True)

#Draw histogram in each subplot
for ax, category, title in zip(axes, categories, titles):
    sns.histplot(data=lemonade, x=category, bins=10, color=one_colour, ax=ax)
    # ax=ax is to assign the figure
    ax.set_title(title, fontsize=14)
    ax.set_xlabel(category, fontsize=12)
    ax.set_ylabel('Count', fontsize=12)

fig.tight_layout()
plt.show()
```