



TEXAS ADVANCED COMPUTING CENTER

WWW.TACC.UTEXAS.EDU



TEXAS

The University of Texas at Austin

# Intro to Matplotlib

The Artist's Paintbrush

PRESENTED BY:

# ◆ Part 1: What Is Matplotlib?

## Why We Visualize Data

- Reveal patterns
- Spot outliers
- Communicate insights
- Find Errors in our Data

**In a nutshell: To see the Big Picture (hehehe)**

# ◆ 1. Basic Line Plot

```
import matplotlib.pyplot as plt
```

```
# Inline plotting
```

```
%matplotlib inline
```

```
# Basic line plot example
```

```
x = [1, 2, 3, 4]
```

```
y = [10, 20, 25, 30]
```

```
plt.plot(x, y)
```

```
plt.title("Simple Line Plot")
```

```
plt.xlabel("X")
```

```
plt.ylabel("Y")
```

```
plt.show()
```

## ◆ Part 2: Scatter Plots

Use Case: Visualizing Geographic Incidents (Latitude vs Longitude)

```
plt.figure(figsize=(10, 6))  
plt.scatter(df['Longitude'], df['Latitude'], alpha=0.3,  
            s=10)  
plt.title("Traffic Incidents in Austin")  
plt.xlabel("Longitude")  
plt.ylabel("Latitude")  
plt.grid(True)  
plt.show()
```

## ◆ Part 2: Scatter Plots

Use Case: Challenge!

✓ **Challenge:** There are obviously errors in our data that we've never taken care of. Drop all Longitude/Latitude rows where the data is wrong, either 0 or "not Austin"

## ◆ Part 2: Scatter Plots

Use Case: Challenge!

✓ **Challenge:** Try coloring points by the top 5 issues reported using `c=...` or using a different marker for each issue type.

## ◆ Part 2: Scatter Plots

Use Case: Challenge!

✓ **Challenge:** Try coloring points by the top 5 issues reported using `c=...` or using a different marker for each issue type.

```
# Get top 5 most common issues
### Your Code Here

# Filter dataset to only include these top 5
### Your Code Here

# Assign a color to each issue type
### Your Code Here

# Map color values
### Your Code Here
```

# ◆ Part 2: Scatter Plots

## Use Case: Challenge!

```
# Plot
plt.figure(figsize=(10, 6))
plt.scatter(
    df_top_issues['Longitude'],
    df_top_issues['Latitude'],
    c=df_top_issues['Color'],
    alpha=0.5,
    s=10
)
plt.title("Scatter Plot of Top 5 Traffic Issues by Location")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.grid(True)
plt.show()
```



# ◆ Part 2: Scatter Plots

## Use Case: Challenge!

```
import matplotlib.pyplot as plt

# Get top 5 most common issues
top_issues = df['Issue
Reported'].value_counts().head(5).index.tolist()

# Filter dataset to only include these top 5
df_top_issues = df[df['Issue
Reported'].isin(top_issues)].copy()
```

```
# Assign a color to each issue type
issue_colors = {
    top_issues[0]: 'red',
    top_issues[1]: 'orange',
    top_issues[2]: 'green',
    top_issues[3]: 'blue',
    top_issues[4]: 'purple',
}

# Map color values
df_top_issues['Color'] = df_top_issues['Issue
Reported'].map(issue_colors)
```

## ◆ Part 2: Scatter Plots

Optional: Add Legend for Readability

```
# Add a custom legend
import matplotlib.patches as mpatches

legend_handles = [mpatches.Patch(color=color, label=issue) for issue, color in
issue_colors.items()]

plt.legend(handles=legend_handles, title="Issue Reported", loc='lower left')
```

## ◆ Part 3: Histograms

Use Case: Distribution of incidents by hour of day

```
plt.hist(df['Hour'], bins=24, color='orange',  
edgecolor='black')  
plt.title("Incidents by Hour of Day")  
plt.xlabel("Hour")  
plt.ylabel("Number of Incidents")  
plt.grid(True)  
plt.show()
```

# ◆ Part 3: Histograms

Use Case: Challenge!

✓ **Challenge:** Create a histogram of incidents by month. What time of year is busiest?

# ◆ Part 3: Histograms

Use Case: Challenge!

```
plt.hist(df['Month'], bins=12, color='teal', edgecolor='black')  
plt.title("Incidents by Month")  
plt.xlabel("Month")  
plt.ylabel("Number of Incidents")  
plt.xticks(range(1, 13))  
plt.grid(True)  
plt.show()
```

## ◆ Part 4: Bar Charts

Use Case: Most common types of issues

```
top_issues = df['Issue Reported'].value_counts().head(5)
top_issues.plot(kind='bar', color='skyblue')
plt.title("Top 5 Traffic Issues Reported")
plt.xlabel("Issue")
plt.ylabel("Number of Reports")
plt.xticks(rotation=45)
plt.show()
```

# ◆ Part 4: Bar Charts

Use Case: Challenge

✓ **Challenge:** Plot the number of “Crash Urgent” reports per month.

# ◆ Part 4: Bar Charts

## Use Case: Challenge

```
# Filter just 'Crash Urgent' reports

crash_df = df[df['Issue Reported'] == 'Crash Urgent']

# Group by month and count

monthly_crashes = crash_df.groupby('Month').size()

monthly_crashes.plot(kind='bar', color='crimson', edgecolor='black')

plt.title("Crash Urgent Reports by Month")

plt.xlabel("Month")

plt.ylabel("Number of Reports")

plt.xticks(rotation=0)

plt.grid(axis='y')

plt.show()
```



# ◆ Part 5: Line Plots for Time Series

Use Case: Incident trends over time

```
daily_counts = df.groupby(df['Published  
Date'].dt.date).size()  
daily_counts.plot(figsize=(12, 4), title="Daily Incident  
Reports Over Time")  
plt.xlabel("Date")  
plt.ylabel("Number of Reports")  
plt.grid(True)  
plt.show()
```

# ◆ Part 5: Line Plots for Time Series

Use Case: Challenge

✓ **Challenge:** Plot incidents by week or month.

# ◆ Part 5: Line Plots for Time Series

## Use Case: Challenge

```
# Create a new column for week

df['Week'] = df['Published Date'].dt.to_period('W').apply(lambda r: r.start_time)


# Group by week and count

weekly_counts = df.groupby('Week').size()


# Plot

weekly_counts.plot(figsize=(12, 4), title="Weekly Incident Reports")

plt.xlabel("Week")

plt.ylabel("Number of Reports")

plt.grid(True)

plt.show()
```

# ◆ Part 6: Bonus – Customization & Subplots

```
## Add Style
plt.style.use('ggplot')

fig, axs = plt.subplots(1, 2, figsize=(12, 4))

axs[0].hist(df['Hour'], bins=24, color='purple')
axs[0].set_title("Hour of Day")

axs[1].bar(top_issues.index, top_issues.values, color='green')
axs[1].set_title("Top 5 Issues")

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

# ◆ Part 7: Final Project – Visual Data Storytelling

## 🎯 Challenge:

*Use 2–3 different plots to explore patterns in the data. What can you discover visually?*

Examples:

- What areas of the city have the most urgent crashes?
- Are there more traffic hazards during certain months?
- Do crash locations cluster around highways?