

2025

4Geeks Academy: data science cohort 12

DAY 4: DATA VISUALIZATION

TODO

DATA VISUALIZATION

Cover basics of plotting with Matplotlib and Seaborn

PANDAS EXERCISES

Submit data clean up assignment (Pandas module), if you haven't already

VISUALIZATION EXERCISES

Start data visualization exercises with matplotlib and seaborn (Data Visualization module), plan to finish before class Wednesday.

TOPICS

01 PLOTTING BEST PRACTICES

02 MATPLOTLIB

03 SEABORN

PLOTTING BEST PRACTICES

HAVE A CLEAR GOAL What do you want to know/convey?

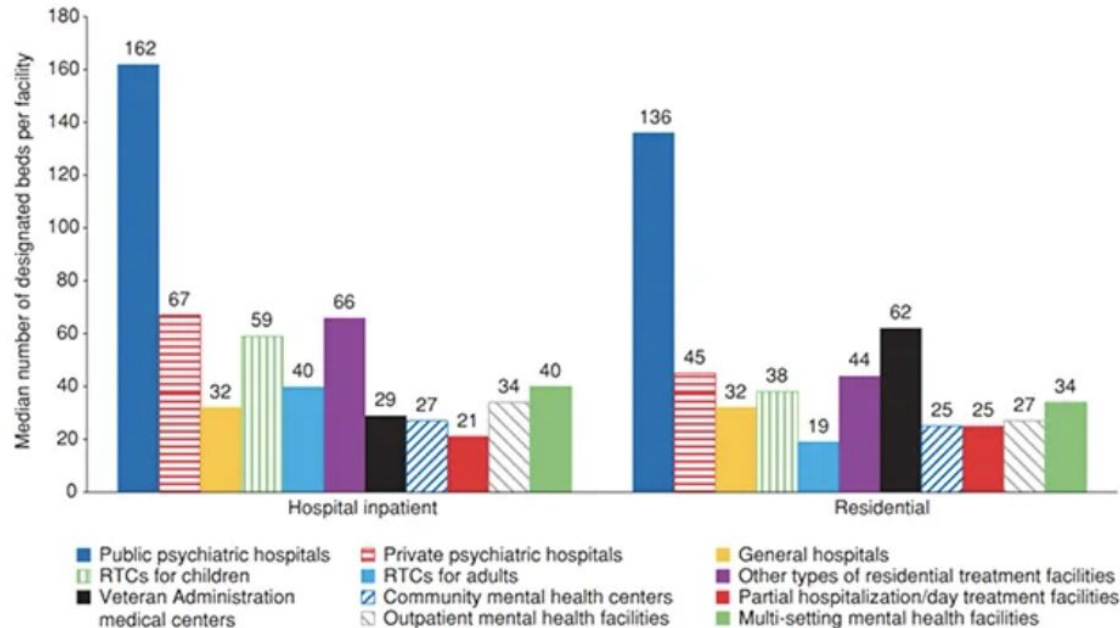
PICK THE RIGHT PLOT Bar, scatter, histogram, boxplot

KEEP IT SIMPLE Draw only what is needed to get the point across

LABEL EVERYTHING Title, legend (if needed) and axes

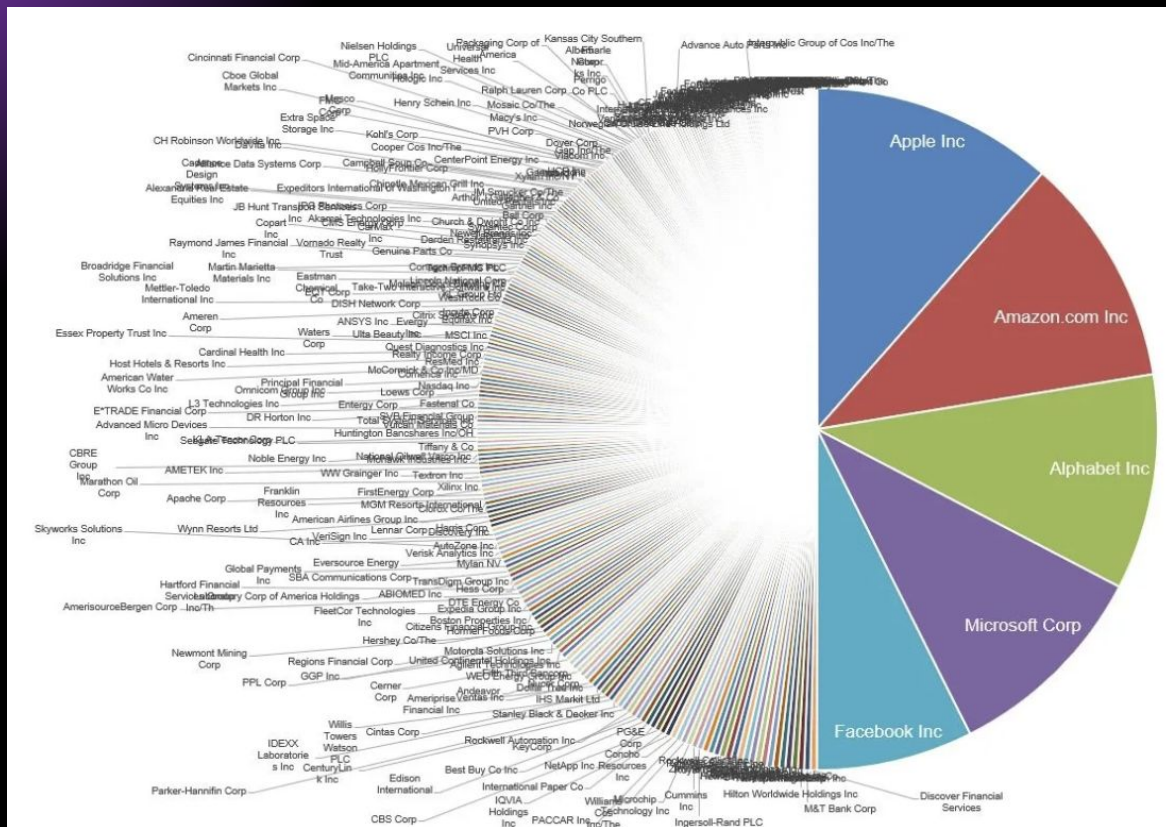
BAD PLOT EXAMPLES

Figure 2.6. Median number of designated beds for mental health treatment per facility, by service setting and facility type: April 29, 2016



SOURCE: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, National Mental Health Services Survey (N-MHSS), 2016.

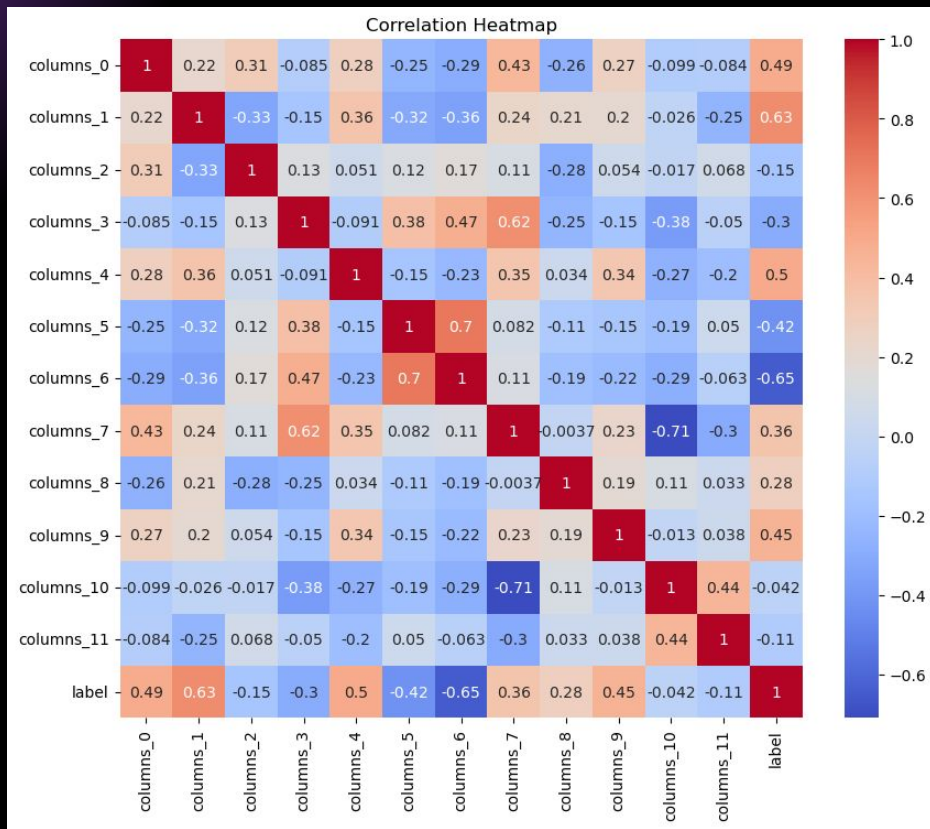
BAD PLOT EXAMPLES



BAD PLOT EXAMPLES



BAD PLOT EXAMPLES



MATPLOTLIB

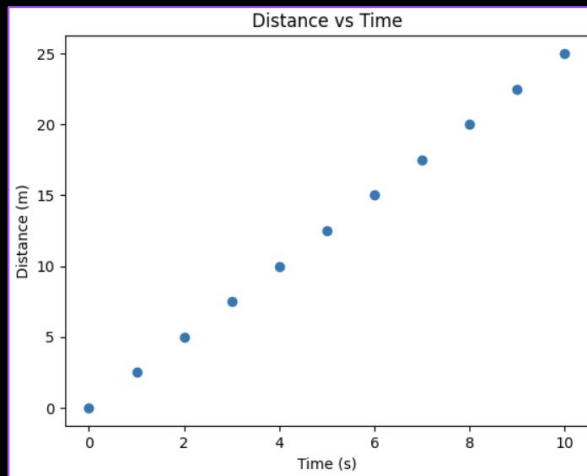
WHAT Plotting and data visualization library for Python with functions for many common plot types.

WHERE Documentation: matplotlib.org
GitHub: [matplotlib/matplotlib](https://github.com/matplotlib/matplotlib)
PyPI: [matplotlib](https://pypi.org/project/matplotlib)

```
import matplotlib.pyplot as plt

plt.scatter(times, distances)
plt.title('Distance vs Time')
plt.xlabel('Time (s)')
plt.ylabel('Distance (m)')
plt.show()
```

✓ 0.1s



MORE EXAMPLES [Matplotlib plot gallery](https://matplotlib.org/faq/plot_types.html)

SEABORN

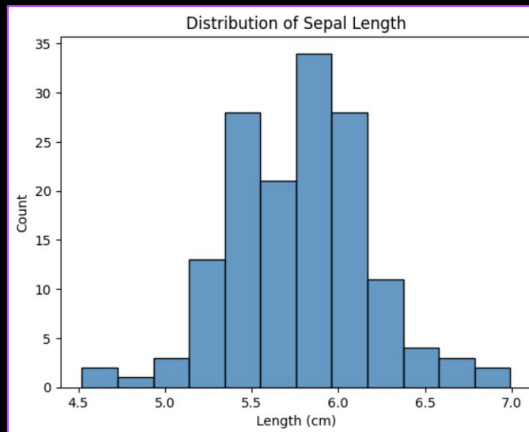
WHAT Statistical data visualization library based on Matplotlib. Easy to use interface for more complex plot types.

WHERE Documentation: seaborn.pydata.org
GitHub: [mwaskom/seaborn](https://github.com/mwaskom/seaborn)
PyPI: [seaborn](https://pypi.org/project/seaborn)

```
import seaborn as sns

sns.histplot(df_iris['sepal length'])
plt.title('Distribution of Sepal Length')
plt.xlabel('Length (cm)')
plt.ylabel('Count')
plt.show()
```

✓ 0.1s



MORE EXAMPLES [Seaborn example gallery](#)