

CSCI 491: Data Visualization

30- Visualizing Tools

Exploratory vs. Explanatory Visualization

- Data science is about helping humans understand the story behind the data, and visualizations provide a powerful tool for helping the analyst understand and communicate that story.
- Visualization can serve a dual purpose:
 - Data exploration
 - Data explanation

Exploratory Visualization

- Exploratory visualization aims to help analysts and researchers understand, discover, and extract insights from data. This type of visualization is typically used during the initial stages of data analysis, where the goal is to uncover patterns, trends, correlations, or anomalies in the data.
- Exploratory visualization is like opening 100 of oyster to find one or two pearls



Interactive Exploratory Visualization Tools

- Interactivity is a powerful tool to accomplish data exploration, whether it's just you, or your team huddled around your computer, allowing users to modify the visualization or data filters to explore different aspects of the data.
- <https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/#bysensitivity>

Explanatory Visualization

- Explanatory visualization focuses on presenting **specific insights** or conclusions derived from the data to a broader audience. These visualizations are generally more polished, as they aim to effectively communicate specific findings or results. Explanatory visualizations are often used in reports, presentations, or publications to support an argument or highlight key takeaways.



Visualizing Tools

- In this course, we delved into the theory of information visualization and explore its practical applications using visualization tools in Python like Matplotlib, Seaborn and Pandas.

Basic Exploratory Visualization Tools

Matplotlib:

- **Strength:** Versatile, low-level plotting library that allows for a wide range of customization.
- **Popular for:** General-purpose plotting, scientific computing, and creating static, publication-quality figures.

Seaborn:

- **Strength:** Built on top of Matplotlib, it provides a high-level interface for statistical graphics, easier to customize than matplotlib
- **Popular for:** Exploratory data analysis, producing attractive and informative statistical plots, and enhancing the aesthetics of Matplotlib plots.

Pandas:

- **Strength:** While primarily a data manipulation library, it has built-in visualization capabilities.
- **Popular for:** Quick visualizations during data analysis, such as bar plots, histograms, and scatter plots, directly from DataFrames.

Visualizing Tools

- Numerous other visualization libraries in Python, other programming languages, and various visualization software applications, serve to either exploratory or explanatory or both purposes.

Interactive Visualization Tools

Many effective explanatory visualizations give the viewer the ability to **interactively explore** the results themselves.

Interactivity in explanatory visualization should allow the viewer to track the narrative uncovered through the scientist's analysis.

Some Interactive Visualization Tools

Plotly:

- **Strength:** Interactive and responsive plots that can be easily integrated into web applications.
- **Popular for:** Creating interactive or animated, web-based visualizations for exploratory data analysis, dashboards, and scientific computing.

Bokeh:

- **Strength:** Offers interactive, web-based visualizations with a high level of customization.
- **Popular for:** Creating web-based dashboards, interactive data applications, and explanatory visualizations.

Altair:

- **Strength:** Declarative visualization library that emphasizes simplicity and provides a concise syntax.
- **Popular for:** Creating simple, interactive, and effective visualizations for data communication and explanatory purposes.

Some Explanatory Visualization Tools

ggplot (plotnine in Python):

- **Strength:** Based on the Grammar of Graphics, it offers a consistent, declarative interface for creating complex, multi-layered plots.
- **Popular for:** Creating publication-quality graphics, especially for those familiar with R's ggplot2.

Folium

- **Strength:** ability to easily visualize geospatial data by leveraging the capabilities of Leaflet, while providing a simple and intuitive Python interface
- **Popular for:** Create interactive maps

D3.js:

- **Strength:** Highly customizable and powerful library for creating web-based, interactive visualizations.
- **Popular for:** Advanced data-driven documents, interactive explanatory visualizations, and custom visualizations for the web.