



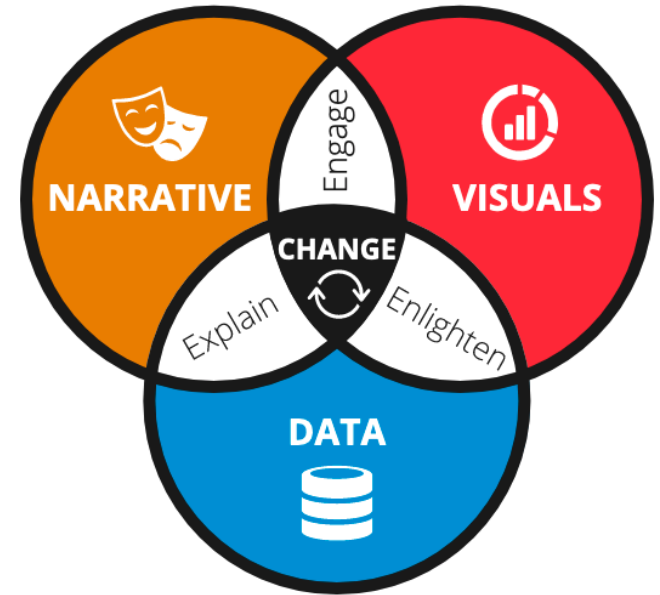
JOHNS HOPKINS
SCHOOL *of* MEDICINE

Module 2.2 - Essential Python Data Packages for Genomics: plotnine & The Grammar of Graphics

ME.440.825.FA23 9/15/23

Storytelling with Data in Python

- The power of Narrative in data visualization
- Data Storytelling: Combination of data, visuals, and narrative
- Why it matters: Making complex data relatable, memorable, and actionable
- Python's Role: Leveraging libraries to craft compelling data-driven narratives



Python has *many* packages for data visualization

- matplotlib is one of the original 'base' plotting packages in python
 - MOST python visualization packages use matplotlib under the hood
- Many packages have emerged for different types of visualization or different use cases for drawing plots
 - Each has its own syntax, style, strengths, and weaknesses

matplotlib

bokkeh



seaborn



plotly

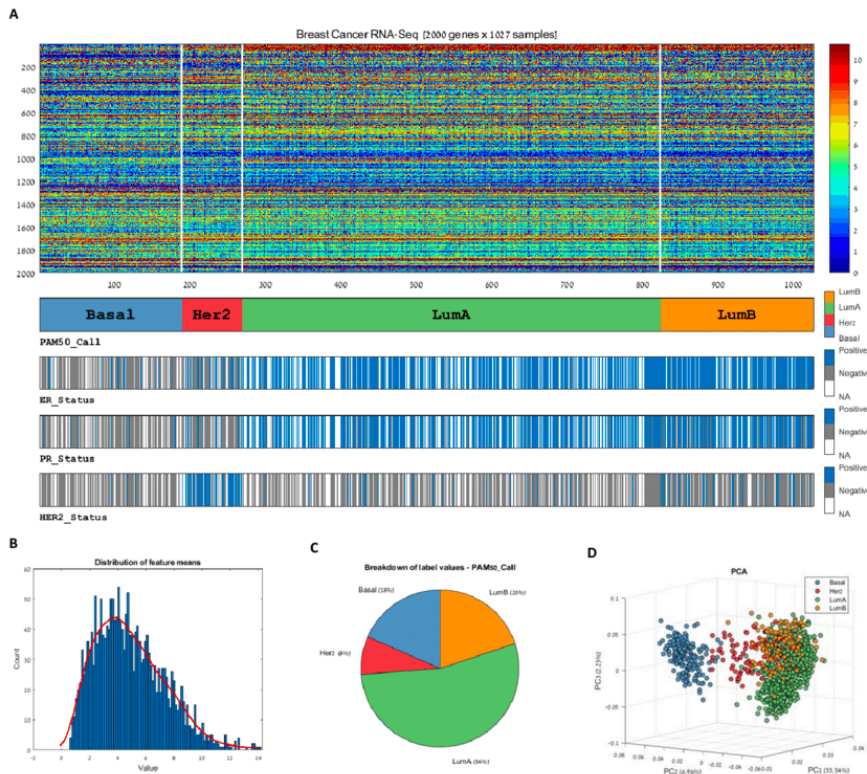


A Grammar of Graphics in Python

PLOTNINE

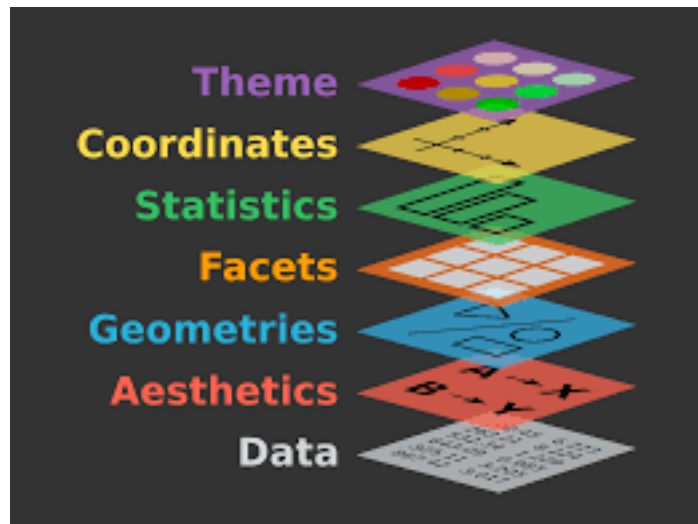
Why plotnine?

- Inspired by ggplot2
 - A python adaptation of R's ggplot2 - one of the most renowned visualization libraries
 - Inherits the strong foundation of the 'Grammar of Graphics'
- A Layered approach to visualization
 - Construct plots layer by layer - adding or modifying details as needed
 - Allows for high customizability without overwhelming complexity
- Flexibility & Integration
 - Easily integrates with the broader python data ecosystem, including numpy and pandas
 - If you can construct a plot in plotnine, you can construct the same plot in R with ggplot2
- Intuitive Data Storytelling
 - Makes complex multi-faceted visualizations more accessible



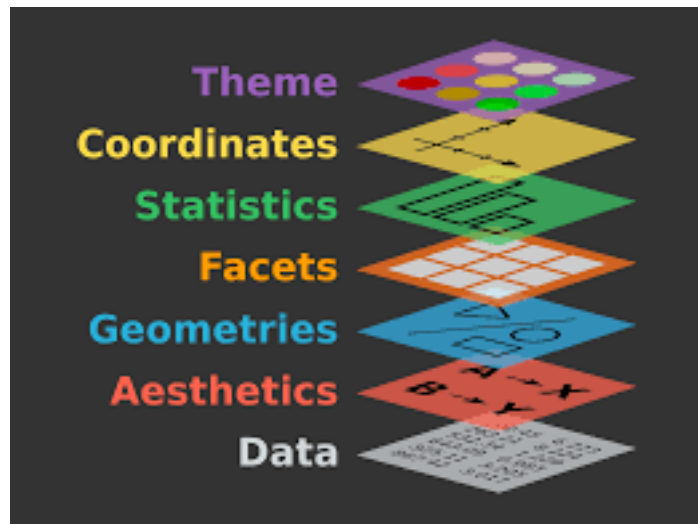
A Grammar of Graphics

- Foundational principles
 - Visualizations are built, layer by layer
 - **Aesthetics**: How data map to visual properties like position, size, color
 - **Geometries**: The visual elements used to represent data
- Core components
 - **Data**: The foundation upon which all visualizations are built.
 - **Transformations**: Modifying data before it's visualized, such as scaling or binning.
 - **Statistics**: Aggregating or summarizing data, like computing means or counts.
 - **Coordinates**: How data is positioned, like Cartesian or polar coordinates.



A Grammar of Graphics

- Enhancing the Visualization
 - **Faceting**: Dividing data into subsets and displaying them as a grid
 - **Themes**: Consistent and pleasing stylistic elements, fonts, and backgrounds
- Advantages
 - **Consistency**: A standardized methodology for diverse datasets
 - **Flexibility**: Adaptable to various data types and visualization goals
 - **Reproducibility**: Precise language leads to easily replicated visuals



Resources

- Plotnine Documentation
 - <https://plotnine.readthedocs.io/en/v0.12.3/>
- Plotnine Tutorials
 - <https://plotnine.readthedocs.io/en/v0.12.3/tutorials.html>
- Grammar of Graphics
 - https://byrneslab.net/classes/biol607/readings/wickham_layered-grammar.pdf
 - <http://stat405.had.co.nz/lectures/21-grammar-of-graphics.pdf>
 - https://www.academia.edu/54823545/The_grammar_of_graphics

LIVE CODING