

# Coffee, Cookies, and Coding (C<sup>3</sup>) Workshop Materials

## Data Visualization with ggplot2

March 27<sup>th</sup>, 2025

### Worked Through Example and Accessing the Code

Open the workshop webpage: <https://ysph-dsde.github.io/Data-Visualization-with-ggplot2/Worked-Through-Example>

From here, you will be able to see all the code involved in producing the worked through example the slides summarize, and you can download the code needed to participate. We suggest you have this page open and scroll through as the relevant slides are presented

### Layered Grammar of Graphics for ggplot2

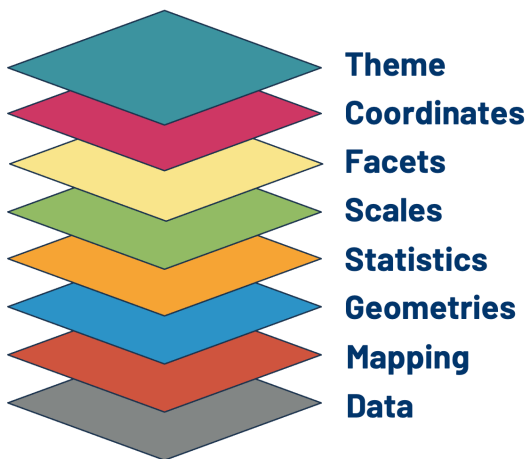


Figure from the "ggplot2 workshop part 1" by Thomas Lin Pedersen. Accessed from YouTube March 15<sup>th</sup>, 2025.

The Grammar of Graphics is the standard domain language used to describe components used in graphical representations of quantitative information. It was first formalized by Professor Leland Wilkinson in his book "The Grammar of Graphics", first published in 1999.

Professor Hadley Wickham adapted this framework into a layered system that he described in his 2010 article "A Layered Grammar of Graphics", published in the Journal of Computational and Graphical Statistics. The figure shown on the left was created by one of the ggplot2 developers, Thomas Lin Pedersen, and shows the different objects that constitute a complete graph.

### Discussions

During the workshop there will be three discussions that require coding to answer them. I have copied the code you need to consider each discussion question in "Discussion and Challenge Questions.R" file.

### Challenge Questions

Additional questions with suggested solutions are available for you to explore. We do not plan to cover them in the workshop, unless things move faster than expected. They can be found in the "Discussion and Challenge Questions.R" file.

# Coffee, Cookies, and Coding (C<sup>3</sup>)

## Workshop Materials

Students are encouraged to try these questions and review the suggested answers on their own. If you have additional questions, please feel free to reach out to me or set up an office hour with me to discuss further.

### Glossary

<b>Grammar of Graphics</b>	Definition for the distinct elements that make up all graphical representations of relational data in tabular form. First created by Professor Leland Wilkinson in 1999.
<b>Layered Grammar of Graphics</b>	The modified version of Grammar of Graphics that stores each element as an independent object. These objects get added together to generate a comprehensive plot.
<b>Layer: Data</b>	A “tidy” data frame with the necessary columns of information to generate the plot you intend.
<b>Layer: Mapping</b>	Assigns variables in the data frame to aesthetic features on the plot (i.e. shape, color, etc.).
<b>Layer: Geometry</b>	Engines (composite of operations) that process the data into a defined plot type (i.e. boxplot, histogram, line, etc.).
<b>Layer: Statistics</b>	Statistical transformations that generate a geometry. Sometimes interchangeable with <code>geom_*()</code> objects.
<b>Layer: Scales</b>	Interpret aesthetic Mappings into plottable values (i.e. axis scaling, color scaling, etc.).
<b>Layer: Facets</b>	Spreads out the same plot into new subplots, each showing distinct instantiations of a variable.
<b>Layer: Coordinates</b>	Defines the coordinate plane of the plot: i.e. Cartesian, polar, transformed, or a map projection.
<b>Layer: Theme</b>	Controls for the non-data elements of the plot.
<b>Map Projection</b>	Interpreting the curved surface of the earth into a flat plane for 2D plotting.
<b>Simple Feature (SF)</b>	Standard vector data produced by the Open Geospatial Consortium (OGC) that’s translates projection data into plottable polygons.