

Today: Grammar of Graphics
1-D Categorical
Friday: ggplot2, 1-D Categorical

September 7, 2016

ggplot2: Based on “The Grammar of Graphics” (Wilkinson, 2005)

Each plot can be broken down into core components.
Wilkinson defines core components in book.

Hadley Wickham puts this into practice in R via ggplot2.

R Package ggplot2 – Hadley Wickham

Core components of a plot:

1. **data**: in ggplot2, data must be stored as an R data frame
2. **coordinate system**: describes 2-D space that data is projected onto
e.g., Cartesian coordinates, polar coordinates, map projections, ...
3. **geometries**: describe type of geometric objects that represent data
e.g., points, lines, polygons, ...
4. **aesthetics**: describe visual characteristics that represent data
e.g., for example, position, size, color, shape, transparency, fill
5. **scales**: for each aesthetic, describe how it is converted into values
that are displayed on the actual graph
e.g., log scales, color scales, size scales, date scales, ...
6. **stats** : describe statistical transformations that help summarize data
e.g., counts, means, medians, regression lines, ...
7. **facets**: describe how data is split into subsets and displayed as
multiple small graphs (particularly important for categorical data!)

1-D Categorical Data

Recall: Data can be **categorical** or **continuous**

Categorical data can be **ordered** or **unordered / nominal**

1-D Categorical Data Structure:

How could we summarize this data?

What information would you report?

1-D Categorical Data

To show the differences among the categories, need to use *area plots*:

Examples of area plots?

1-D Categorical Data – Bar Charts

Bar Charts: rectangular bar is created for each unique categorical value. The area and height of the bar is proportional to % of observations with the categorical value. Bars usually have equal width.

1-D Categorical Data – Spine Charts

Spine Charts: rectangular bar is created for each unique categorical value. The height (or width) of all bars is equal, and the width (or height) of the bar corresponds to the proportion in that category.

1-D Categorical Data – Pie Charts

Pie Charts: circle divided up into sections (“pie slices”) such that the area of each section is proportional to the number of observations with each unique categorical value.

1-D Categorical Data – Rose Diagrams

Rose Diagrams: circle sections are created for each category. All sections have the same width/arc/angle. The radius of each section is proportional to the category frequency. Sections are called “petals”. Developed by Florence Nightingale (example posted to Blackboard).