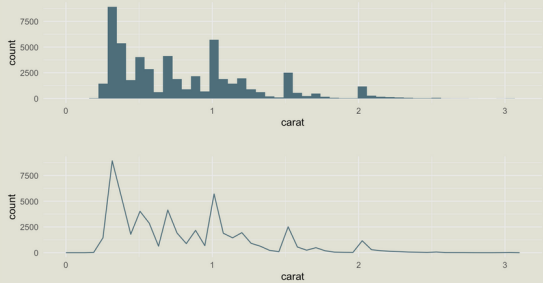
Andrew Gard - equitable.equations@gmail.com

What is the grammar of graphics?

The **grammar of graphics**, first described by Leland Wilkinson (2001) and popularized by Hadley Wickham (2007), describes every plot in terms of a small number of characteristics, the most important of which are

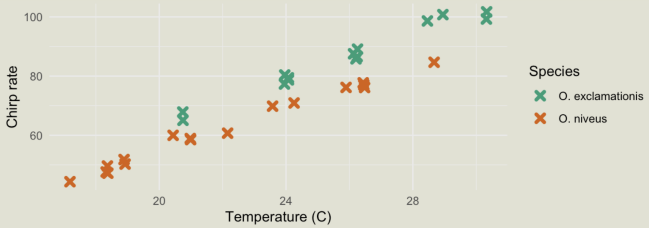
1. The variables represented and the dimensions used to represent them (x-axis, y-axis, color, etc)
2. The sort of geometry used to represent the data (points, bars, boxes, etc)
3. Stylistic elements not defined by the data, including labels, fonts, and color palettes,

Plots are primarily defined by their variables, not their geometries.



The histogram above shows exactly the same information as the frequency polygon.

Example. The following plot (based on the R data set `modeldata::crickets`) represents `temperature` on the x-axis, `chirp rate` on the y-axis, and ‘species‘ using color.



Non-data aspects of the plot include the choice of colors and the use of X’s to mark the observations.

The grammar of graphics is particularly popular among R users, where it's implemented using the `ggplot` function from the `ggplot2` package. The code for the previous plot is:

```
ggplot(crickets,
       aes(x = temp,
           y = rate,
           color = species)) +
  geom_jitter(shape = 4,
              stroke = 2) +
  labs(x = "Temperature (C)",
       y = "Chirp rate",
       color = "Species") +
  scale_color_brewer(palette = "Dark2") +
  theme(plot.background = element_rect(fill = "#E1E1D3"))
```

Many different dimension can be used to represent variables in a set. The most common are:

- x- and y-axes
- Boundary and interior colors
- Size
- Shape
- Line type (dotted, dashed, etc)
- Opacity

R users refer to these as **aesthetics** or **aesthetic mappings**.

Warming! Aesthetics describe how variables in a data set are represented in a plot. When a property of a graph (like color or shape) isn't determined by a variable, it is **not** and aesthetic but rather a stylistic choice.

Example. For each of the following plots, identify all aesthetic mappings (the variables and dimensions used to represent them). What are some non-data aspects of the plot?