What do I mean by `tidy` data?

Data are often stored in **tabular** (or matrix) form:

```
1 library(palmerpenguins)
2 penguins |> slice(1:5)
```

```
# A tibble: 5 × 8
  species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
1 Adelie  Torgersen           39.1          18.7               181        3750
2 Adelie  Torgersen           39.5          17.4               186        3800
3 Adelie  Torgersen           40.3          18                 195        3250
4 Adelie  Torgersen           NA            NA                  NA          NA
5 Adelie  Torgersen           36.7          19.3               193        3450
# ℹ 2 more variables: sex <fct>, year <int>
```

Each row == an observation

each column is variable / measurement

The Grammar of Graphics

Originally defined by Leland Wilkinson

1. **data**
2. **geometries**: type of geometric objects to represent data, e.g., points, lines
3. **aesthetics**: visual characteristics of geometric objects to represent data, e.g., position, size
4. **scales**: how each aesthetic is converted into values on the graph, e.g., color scales
5. **stats**: statistical transformations to summarize data, e.g., counts, means, regression lines
6. **facets**: split data and view as multiple graphs
7. **coordinate system**: 2D space the data are projected onto, e.g., Cartesian coordinates

Hadley Wickham created `ggplot2`

1. `data`
2. `geom_X` geom_bar geom_point
3. `aes`: mappings of columns to geometric objects
4. `scale`: one scale for each `aes` variable
5. `stat` scale_size_manual()
6. `facet`
7. `coord`
8. `labs`: labels/guides for each variable and other parts of the plot, e.g., title, subtitle, caption
9. `theme`: customization of plot layout