Data Visualization with R
Tidyverse tutorial

Ariane Ducellier

University of Washington - Fall 2023

What is tidyverse?

A collection of R packages designed for data science.

Basic packages:

- **ggplot2:** graphics
- **dplyr:** data manipulation
- **tidyr:** getting to tidy data
- **readr:** reading rectangular data (e.g. csv, tsv, fwf)
- **purrr:** working with functions and vectors
- **tibble:** a modern re-imagining of the data frame
- **stringr:** working with strings
- **forcats:** working with R factors to handle categorical variables

Additional packages associated to `tidyverse` need to be installed and loaded separately to import data, wrangle data, program and model.

Main concepts of data wrangling

- Understand.
- Format → Produce tidy data:
  - Every column is a variable.
  - Every row is an observation.
  - Every cell is a single value.
- Clean.
- Enrich.
- Validate.
- Analysis / Model → In our case, we are going to produce visuals to communicate information on the dataset to the viewer.

Tibbles versus Data frames

- Tibbles do not change input variable types by default.
- Tibbles can have lists as columns.
- Tibbles can have non-standard variable names.
- Tibbles return another Tibble when slicing (and not a vector).

The pipe operator

The `magrittr` package provides the $\% <> \%$ operator as a shortcut for modifying an object in place:

```
df_iris <- iris %>%
  group_by(Species) %>%
  summarize_if(is.numeric, mean) %>%
  ungroup() %>%
  gather(measure, value, -Species) %>%
  arrange(value)
```

=

```
df_iris <- group_by(iris, Species)
df_iris <- summarize_if(df_iris, is.numeric, mean)
df_iris <- ungroup(df_iris)
df_iris <- gather(df_iris, measure, value, -Species)
df_iris <- arrange(df_iris, value)
```