R Basics for DAV

Contents

Loading data

For this practical, we are going to use a dataset from Kaggle - the Google play store apps data by user lava18. We have downloaded it into the data folder already from https://www.kaggle.com/lava18/google-play-store-apps.

Tidyverse contains many data loading functions in the packages `readr` (default file types) and `haven` (external file types such as from SPSS or Stata), each for their own file type. The most common file type is `csv`, which is what we use here.

> Use the function `read_csv()` to import the file "data/googleplaystore.csv" and store it in a variable called `apps`.

```
apps <- read_csv("data/googleplaystore.csv")
```

```
## Parsed with column specification:
## cols(
##   App = col_character(),
##   Category = col_character(),
##   Rating = col_double(),
##   Reviews = col_integer(),
##   Size = col_character(),
##   Installs = col_character(),
##   Type = col_character(),
##   Price = col_character(),
##   `Content Rating` = col_character(),
##   Genres = col_character(),
##   `Last Updated` = col_character(),
##   `Current Ver` = col_character(),
##   `Android Ver` = col_character()
## )
```

These import functions are fast and safe: they display informative errors if anything goes wrong. In this case, we can also see what variable type each column gets.

> Use the function and `head()` to look at the first few rows of the `apps` dataset

```r
head(apps)
```

```
## # A tibble: 6 x 13
##   App   Category Rating Reviews Size  Installs Type  Price `Content Rating`
##   <chr> <chr>     <dbl>   <int> <chr> <chr>    <chr> <chr> <chr>
## 1 Phot~ ART_AND~    4.1     159 19M   10,000+  Free  0     Everyone
## 2 Colo~ ART_AND~    3.9     967 14M   500,000+ Free  0     Everyone
## 3 "U L~ ART_AND~    4.7   87510 8.7M  5,000,0~ Free  0     Everyone
## 4 Sket~ ART_AND~    4.5  215644 25M   50,000,~ Free  0     Teen
## 5 Pixe~ ART_AND~    4.3     967 2.8M  100,000+ Free  0     Everyone
## 6 Pape~ ART_AND~    4.4     167 5.6M  50,000+  Free  0     Everyone
## # ... with 4 more variables: Genres <chr>, `Last Updated` <chr>, `Current
## #   Ver` <chr>, `Android Ver` <chr>
```