Problem Set 1

Andreas Bender, Philipp Kopper, Philip Studener

26 October 2023

Ressources

Read chapter 2 & chapter 6.1 in The Art of R Programming. "extended examples" can generally be skipped. Paragraphs referencing matrices can also be skipped.

Exercises

Run this code before you start:

```
bundesliga <- c(
    "FC Bayern"           = 55L,
    "BVB"                 = 51L,
    "RB Leipzig"          = 50L,
    "Borussia MGB"        = 49L,
    "Bayer 04"            = 47L,
    "FC Schalke 04"       = 37L,
    "VfL Wolfsburg"       = 36L,
    "SC Freiburg"         = 36L,
    "TSG Hoffenheim"      = 35L,
    "1. FC Köln"          = 32L,
    "Union Berlin"        = 30L,
    "Eintracht Frankfurt" = 28L,
    "Hertha Berlin"       = 28L,
    "FC Augsburg"         = 27L,
    "Mainz 05"            = 26L,
    "Fortuna Duesseldorf" = 22L,
    "Werder Bremen"       = 18L,
    "SC Paderborn"        = 16L)
```

As you may already have guessed: Our topic is soccer. In particular, this code creates a vector with the point standings in the Bundesliga before the COVID-19 break during the first wave. Note: You don't need to use loops or to write functions to solve this exercise!

1. What is the data type (mode) of this vector? What is its scale (Skalenniveau)?
2. Extract the first three teams from the `bundesliga` vector. Then extract the last three teams.
3. Extract the points of the SC Freiburg, SC Paderborn and TSG Hoffenheim. Save the result in the vector `teams_selection`. Also compute the mean and sum of their points.
4. In the following we substract the mean of `teams_selection` from each value of `teams_selection` and divide the difference by `11.26943`. Is this transformation *allowed*? (*Hint*: think about the scale ("Skalenniveau") of the variable)?

```
(teams_selection - mean(teams_selection)) / 11.26943
```

5. Explain the code behaviour of the following chunks. *Hint*: Execute `?Comparison` in the console and read the help page.

```
bundesliga[1] == 55
```

```
## FC Bayern
##      TRUE
```

```
bundesliga[1] == "55"
```

```
## FC Bayern
##      TRUE
```

```
bundesliga[2] > 55
```

```
##   BVB
## FALSE
```

```
bundesliga[2] > "fifty-five"
```

```
##   BVB
## FALSE
```

```
bundesliga[2] < "fifty-five"
```

```
##  BVB
## TRUE
```

```
TRUE + TRUE
```

```
## [1] 2
```

```
typeof(TRUE + TRUE)
```

```
## [1] "integer"
```

```
TRUE > 3
```

```
## [1] FALSE
```

6. `wines` is a character vector of a selection of last year students' favorite wines/drinks.

```
wines <- c(
  "sauvignon blanc", "chardonnay", "merlot", "pinot grigio", "riesling",
  "white", "zinfadel", "sauvignon blanc", "rose", "cabernet sauvignon",
  "primitivo", "saint emilion", "pinot blanc", "riesling", "white", "white",
  "lugana", "red", "merlot", "white", "sauvignon blanc", "amarone",
  "pinot gris", "elbling", "blanc de noirs", "merlot", "beer", "rioja")
```

Categorize the reported wines/drinks to categories "red", "white" and "other". Use the internet to become familiar with the different wines/drinks. (This course takes its educational purpose very seriously and does not limit it to programming.) Store the results in a new vector `wines_cat`. Make sure that it is a `factor` variable. As starting help we give you the category of the first 19 wines/drinks: `c("white", "white", "red", "white", "white", "white", "red", "white", "other", "red", "red", "red", "white", "white", "white", "white", "red", "red")`

7. Explain the following output:

- As what type are the values stored as internally?
- Why are they sorted (e.g. `3` refers to "white") this way?

```r
str(wines_cat)
```

```
##  Factor w/ 3 levels "other","red",..: 3 3 2 3 3 3 2 3 1 2 ...
```

8. Count the number of different wine/drink categories.
9. Assume that we explicitly want to introduce the category `Rose wine`. Change the category of the 9th vector element from `other` to `Rose wine`. Why does this not work?

```r
wines_cat[9] <- "Rose wines"
```

```
## Warning in `[<-.factor`(`*tmp*`, 9, value = "Rose wines"): invalid factor level,
## NA generated
```

10. Try to solve the problem. *Hint*: `?levels`.
11. What is the scale (Skalenniveau) of the `wines_cat` variable?
12. Recreational sommelier P.K. has 8 wines in his cellar. The vector below gives their price range (in Euros).

```r
wine_prices <- c("0-20", "20-100", "20-100", "0-20", "0-20", "0-20", "0-20", "100-200")
```

What is the scale of the variable `wine_prices`? Transform the vector from class `character` to a class appropriate for its scale.

Session Info

```r
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.0
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.2.2  magrittr_2.0.3  fastmap_1.1.0   cli_3.4.1
##  [5] tools_4.2.2     htmltools_0.5.5 rstudioapi_0.14 yaml_2.3.6
##  [9] stringi_1.7.8   rmarkdown_2.25  knitr_1.40      stringr_1.4.1
## [13] xfun_0.40       digest_0.6.30   rlang_1.0.6     evaluate_0.17
```