Problem Set 6

Andreas Bender, Philipp Kopper, Sven Lorenz

09 December 2021

Read chapter 3 from *R for Data Science*. You can skip the exercises.

Exercise 1: Penguins

In the following, all solutions should be performed using R package **ggplot2**.

a) Consider the `penguins` data set that you can download from the Moodle page. See here for some information about the data. Create a histogram of the variable `body_mass_g`. In a second plot specify 20 bins and color the bars in yellow.

b) Create a scatter plot of the variables `flipper_length_mm` and `culmen_length_mm`. Create a second plot with transposed y and x axis.

c) Create a scatter plot of the variables `flipper_length_mm` and `culmen_length_mm`, in which the observations of different `species` are displayed in different colors.

d) Improve the plot from 3 by renaming the axis. Think of a title for the plot and add it to the plot. Additionally, change the size of the points to 3, fill them with color and frame them with black. Choose red for `Adelie`, green for `Chinstrap` and yellow for `Gentoo`. *Hint: Use* `scale_fill_manual`.

e) Create a boxplot of the variable `culmen_depth_mm`. Color the boxplot using the color code `#00FF00`.

f) Now, create a boxplot of the variable `culmen_depth_mm` for each `species`. The boxplots for each species should have the same colors as in 4).

g) Create a bar plot that shows the frequency of the different penguin species.

h) Create a stacked bar plot for the frequency of different penguin species on the different islands.

i) At the beginning of your document, after you load the **ggplot2** package, add the following line and re-knit the Rmarkdown document:

```
theme_set(theme_bw())
```

Exercise 2: Stock Market Visualisation

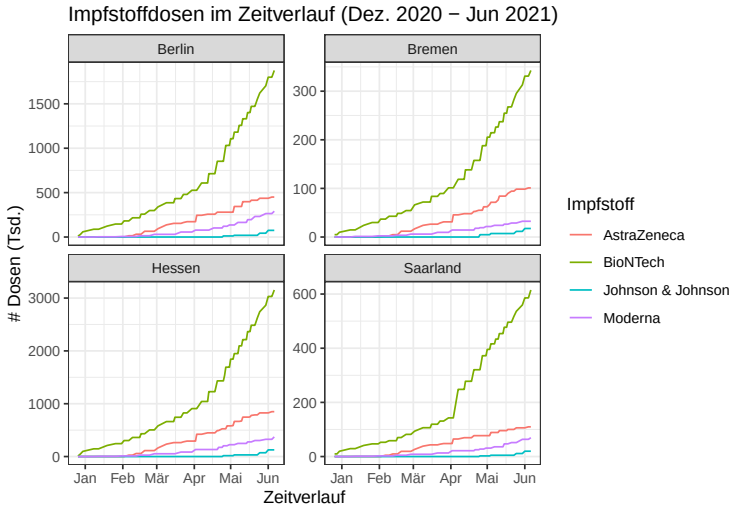
Line plots can be useful for the visualisation of developments over time. Use the `EuStockMarkets` data set (available directly in `R`) and create a plot of the time-related developments of all variables. Choose the color black for `DAX`, red for `SMI`, green for `CAC` and blue for `FTSE`. Use the WallStreetJournal theme of the package `ggthemes`. Use the following code to transform the `EuStockMarkets` object to a data frame with the information required for the plot.

```
stocks <- as.data.frame(EuStockMarkets)
stocks$time <- time(EuStockMarkets)
stocks$time <- as.numeric(stocks$time)
str(stocks)
```

```
## 'data.frame':    1860 obs. of  5 variables:
##  $ DAX : num  1629 1614 1607 1621 1618 ...
##  $ SMI : num  1678 1688 1679 1684 1687 ...
##  $ CAC : num  1773 1750 1718 1708 1723 ...
##  $ FTSE: num  2444 2460 2448 2470 2485 ...
##  $ time: num  1991 1992 1992 1992 1992 ...
```

Exercise 3: Plot Replication

You have data from June 2021 on the progress of the vaccination campaign in Germany. Use the `vaccs.Rds` data set from the Moodle page to replicate the plot below using the `ggplot2` package. No other packages may be used. You may need to preprocess the data first.



Session Info

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.3 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=de_DE.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=de_DE.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=de_DE.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=de_DE.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
```

```
## other attached packages:
## [1] ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
##  [1] pillar_1.6.4      compiler_4.1.2    tools_4.1.2       digest_0.6.28
##  [5] evaluate_0.14     lifecycle_1.0.1   tibble_3.1.6      gtable_0.3.0
##  [9] pkgconfig_2.0.3   rlang_0.99.0.9001 cli_3.1.0.9000    DBI_1.1.1
## [13] yaml_2.2.1        xfun_0.28         fastmap_1.1.0     withr_2.4.2
## [17] stringr_1.4.0     dplyr_1.0.7.9000  knitr_1.36        generics_0.1.1
## [21] vctrs_0.3.8       grid_4.1.2        tidyselect_1.1.1  glue_1.5.0
## [25] R6_2.5.1          fansi_0.5.0       rmarkdown_2.11    farver_2.1.0
## [29] purrr_0.3.4       magrittr_2.0.1    scales_1.1.1      ellipsis_0.3.2
## [33] htmltools_0.5.2   assertthat_0.2.1  colorspace_2.0-2  labeling_0.4.2
## [37] utf8_1.2.2        stringi_1.7.6     munsell_0.5.0     crayon_1.4.2
```