

# Journal (reproducible report)

Johannes Julius Halbe

2020-11-27

## Contents

Challenge 1	1
Challenge 2	4
Adding R stuff	6

This is an .Rmd file. It is plain text with special features. Any time you write just like this, it will be compiled to normal text in the website. If you put a # in front of your text, it will create a top level-header.

## Challenge 1

Last compiled: 2020-12-04

```
# Challenge 01 ----

# 1.0 Load libraries ----
library(tidyverse)

#Excel Files
library(readxl)

# 2.0 Importing Files ----
bikes_tbl      <- read_xlsx("docs/00_data/01_bike_sales/01_raw_data/bikes.xlsx")
orderlines_tbl <- read_xlsx("docs/00_data/01_bike_sales/01_raw_data/orderlines.xlsx")
bikeshops_tbl  <- read_xlsx("docs/00_data/01_bike_sales/01_raw_data/bikeshops.xlsx")

# 3.0 Examining Data ----

#orderlines_tbl

#glimpse(orderlines_tbl)

#view(orderlines_tbl)

# 4.0 Joining Data ----

bike_orderlines_joined_tbl <- orderlines_tbl %>%
```

```

left_join(bikes_tbl, by = c("product.id" = "bike.id")) %>%
left_join(bikeshops_tbl, by = c("customer.id" = "bikeshop.id"))

# 5.0 Wrangling Data ----
bike_orderlines_wrangled_tbl <- bike_orderlines_joined_tbl %>%
  select(-...1) %>%
  rename(bikeshop = name) %>%
  set_names(names(.) %>% str_replace_all("\\\\.", "_")) %>%
  separate(col = location,
            into = c("city", "state"),
            sep = ", ") %>%
  mutate(total_price = price * quantity)

# 6.0 Business Insights ----
# 6.1 Sales by location ----

# Step 1 - Manipulate

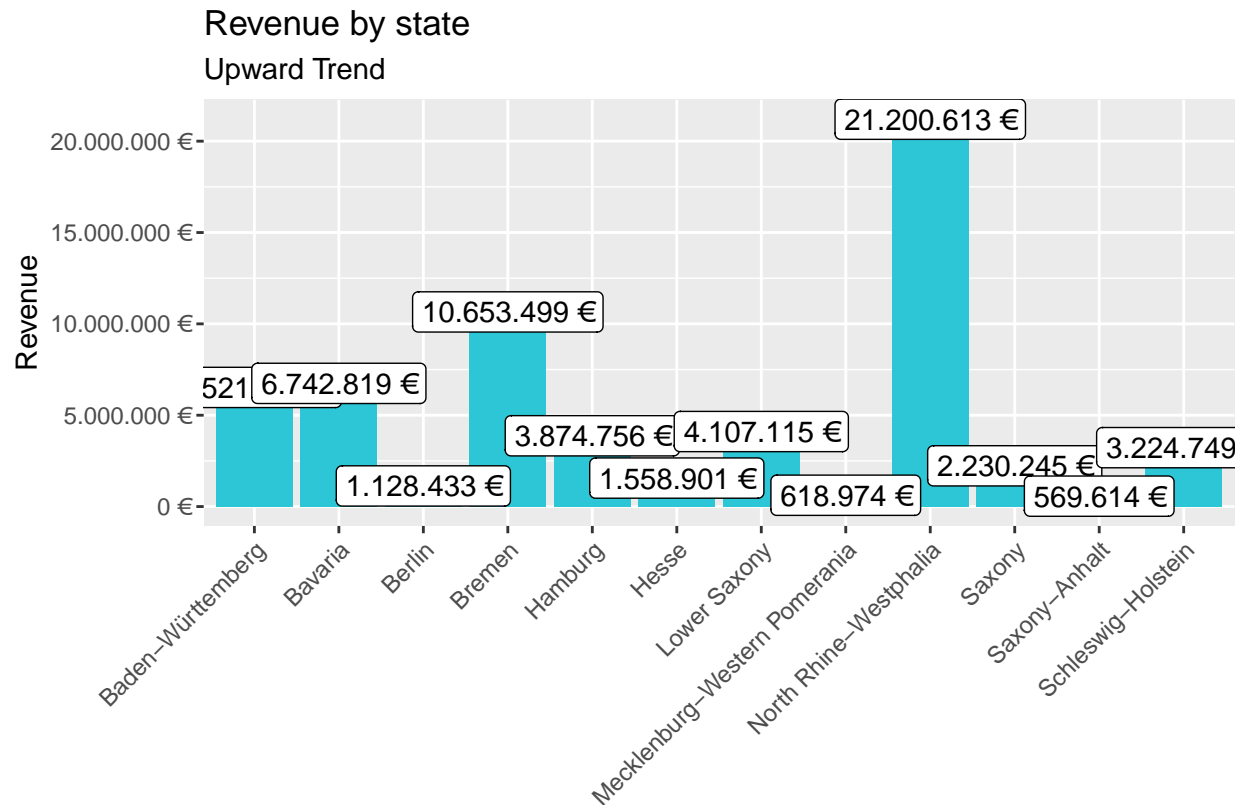
sales_by_location_tbl <- bike_orderlines_wrangled_tbl %>%
  select(state, total_price) %>%
  group_by(state) %>%
  summarize(sales = sum(total_price)) %>%
  mutate(sales_text = scales::dollar(sales, big.mark = ".",
                                     decimal.mark = ",",
                                     prefix = "",
                                     suffix = " €"))

# Step 2 - Visualize

sales_by_location_tbl %>%
  ggplot(aes(x = state, y = sales)) +
  geom_col(fill = "#2DC6D6") +
  geom_label(aes(label = sales_text)) +
  geom_smooth(method = "lm", se = FALSE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::dollar_format(big.mark = ".",
                                                    decimal.mark = ",",
                                                    prefix = "",
                                                    suffix = " €")) +

  labs(
    title = "Revenue by state",
    subtitle = "Upward Trend",
    x = "",
    y = "Revenue"
  )

```



```
# 6.2 Sales by location & year ----

# Step 1 - Manipulate
library(lubridate)

sales_by_location_year_tbl <- bike_orderlines_wrangled_tbl %>%

  select(state, total_price, order_date) %>%
  mutate(year = year(order_date)) %>%
  group_by(state, year) %>%
  summarise(sales = sum(total_price)) %>%
  ungroup() %>%

  mutate(sales_text = scales::dollar(sales, big.mark = ".",
    decimal.mark = ",",
    prefix = "",
    suffix = " €"))

# Step 2 - Visualize
sales_by_location_year_tbl %>%

  # Set up x, y, fill
  ggplot(aes(x = year, y = sales)) +

  # Geometries
  geom_col() + # Run up to here to get a stacked bar plot
```

```

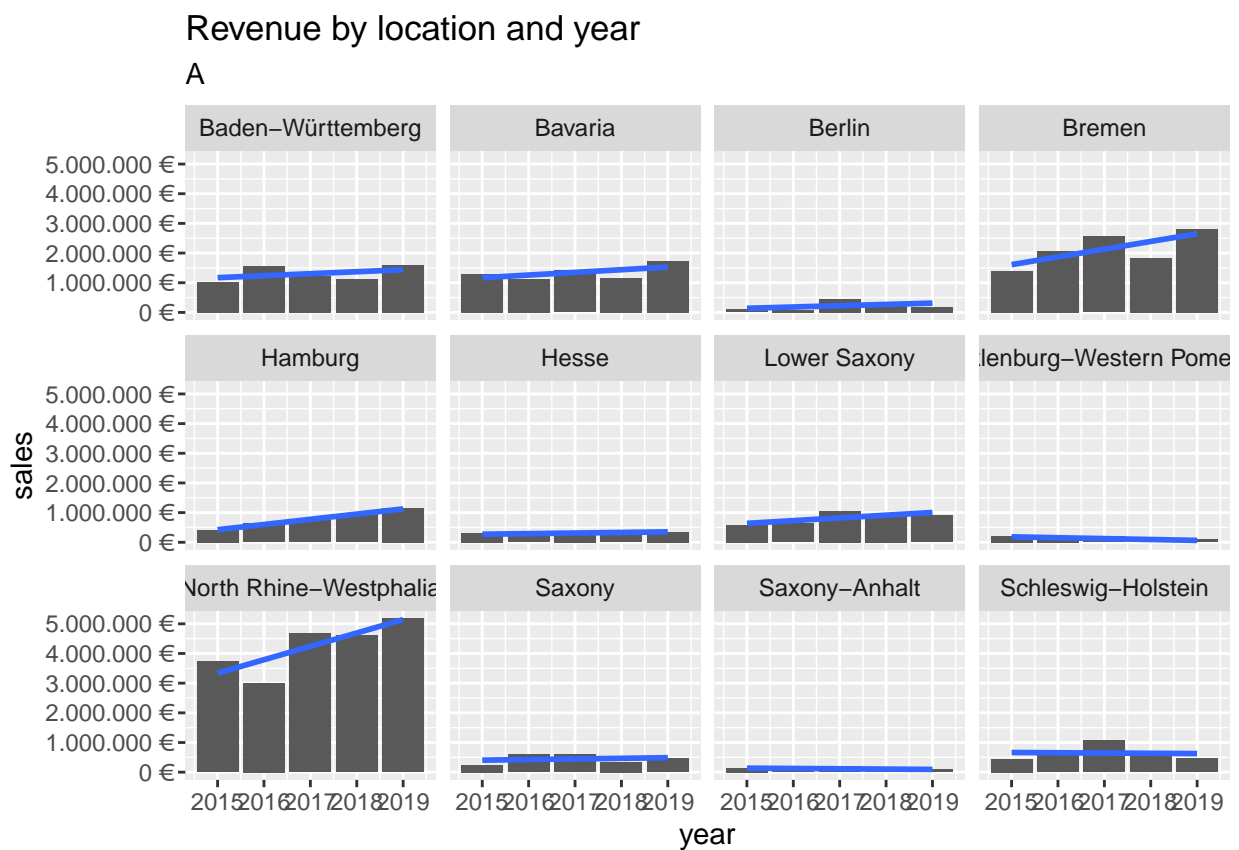
geom_smooth(method = "lm", se = FALSE) +

# Facet
facet_wrap(~ state) +

# Formatting
scale_y_continuous(labels = scales::dollar_format(big.mark = ".",
                                                    decimal.mark = ",",
                                                    prefix = "",
                                                    suffix = " €")) +

labs(
  title = "Revenue by location and year",
  subtitle = "A"
)

```



## Challange 2

Last compiled: 2020-12-04

### #1. API

```

library(tidyverse)
library(httr)
library(jsonlite)

```

```

library(tibble)
library(keyring)

keyring::key_set("token")

resp <- GET("https://www.ncdc.noaa.gov/cdo-web/api/v2/stations?limit=1000", add_headers(token = key_get("token")))

stations_tbl <- resp %>%
  .$content %>%
  rawToChar() %>%
  fromJSON() %>% .$results

head(stations_tbl,10)

```

```

##      elevation   mindate   maxdate latitude      name datacoverage
## 1      139.0 1948-01-01 2014-01-01 31.57020      ABBEVILLE, AL US      0.8813 COOP:010
## 2      249.3 1938-01-01 2015-11-01 34.25530      ADDISON, AL US      0.5059 COOP:010
## 3      302.1 1940-05-01 1962-03-01 34.41667      ADDISON CENTRAL TOWER, AL US      0.9658 COOP:010
## 4      172.2 1995-04-01 2015-11-01 33.17833      ALABASTER SHELBY CO AIRPORT, AL US      0.8064 COOP:010
## 5      183.8 1949-01-01 1949-12-01 34.68910      BELLE MINA 2 N, AL US      1.0000 COOP:010
## 6       34.1 1935-05-01 1936-11-01 31.13333      ALAGA, AL US      0.2624 COOP:010
## 7       53.3 1940-11-01 2014-12-01 32.23220      ALBERTA, AL US      0.9888 COOP:010
## 8      348.1 1931-01-01 1977-06-01 34.23333      ALBERTVILLE, AL US      0.9535 COOP:010
## 9      195.1 1969-10-01 2015-11-01 32.94520      ALEXANDER CITY, AL US      0.9946 COOP:010
## 10     200.9 1942-11-01 1969-10-01 32.98333      ALEXANDER CITY 6 NE, AL US      0.9629 COOP:010
##      elevationUnit longitude
## 1      METERS -85.24820
## 2      METERS -87.18140
## 3      METERS -87.31667
## 4      METERS -86.78167
## 5      METERS -86.88190
## 6      METERS -85.06667
## 7      METERS -87.41040
## 8      METERS -86.16667
## 9      METERS -85.94800
## 10     METERS -85.86667

```

## #2. Web scraping

### # LIBRARIES ----

```

library(tidyverse) # Main Package - Loads dplyr, purrr, etc.
library(rvest)     # HTML Hacking & Web Scraping
library(xopen)     # Quickly opening URLs
library(jsonlite)  # converts JSON files to R objects
library(glue)      # concatenate strings
library(stringi)   # character string/text processing

url <- "https://www.rosebikes.de/fahrr%C3%A4der/rennrad"
html <- url %>%
  read_html()

```

```

model_name <- html %>%
  html_nodes(".catalog-category-bikes__title > span") %>%
  html_text() %>%
  stringr::str_extract("(?<=\n).*(?=\n)")

model_price_cent <- html %>%
  html_nodes(".catalog-category-bikes__price-title") %>%
  html_text() %>%
  stringr::str_extract("(?<=ab\\s).*(?=\s€)") %>%
  str_replace_all(c("\\." = "", "," = "")) %>%
  as.numeric()

model_price_EUR = model_price_cent / 100

bikes_tbl <- tibble(model_name, model_price_EUR)
head(bikes_tbl, 10)

```

```

## # A tibble: 9 x 2
##   model_name      model_price_EUR
##   <chr>          <dbl>
## 1 PRO SL DISC      1599
## 2 PRO SL           1199
## 3 REVEAL FOUR DISC 2499
## 4 REVEAL FOUR      2099
## 5 REVEAL SIX DISC  3499
## 6 X-LITE FOUR DISC 2699
## 7 X-LITE FOUR      2199
## 8 X-LITE SIX DISC  3899
## 9 X-LITE SIX       3499

```

## Adding R stuff

So far this is just a blog where you can write in plain text and serve your writing to a webpage. One of the main purposes of this lab journal is to record your progress learning R. The reason I am asking you to use this process is because you can both make a website, and a lab journal, and learn R all in R-studio. This makes everything really convenient and in the same place.

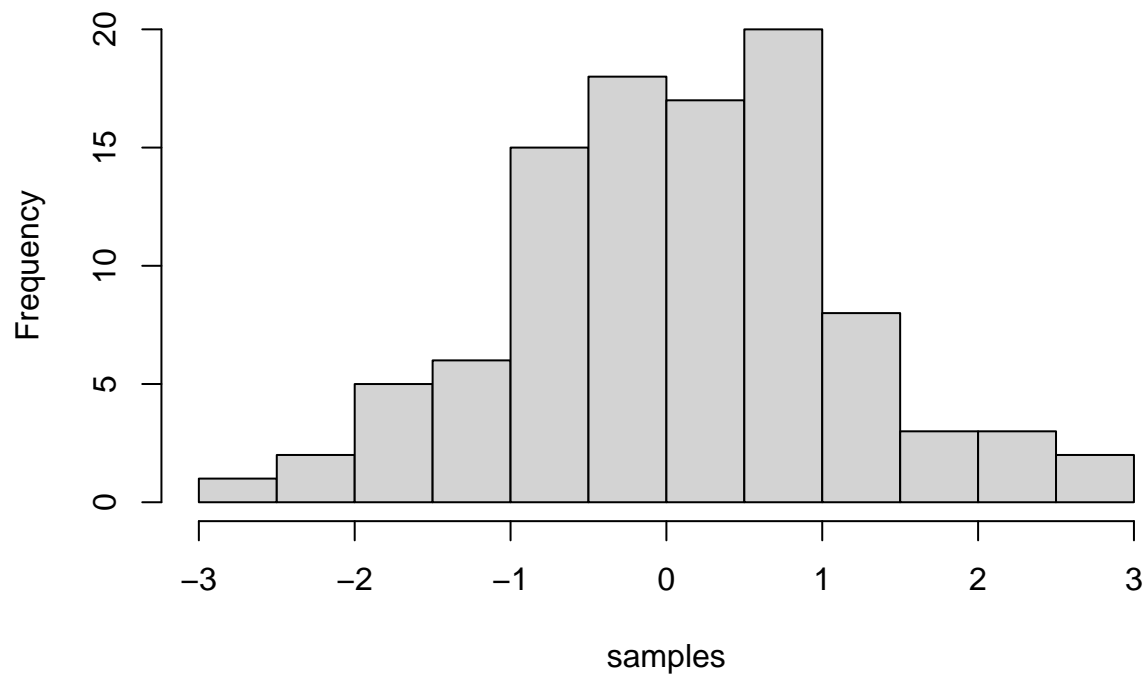
So, let's say you are learning how to make a histogram in R. For example, maybe you want to sample 100 numbers from a normal distribution with mean = 0, and standard deviation = 1, and then you want to plot a histogram. You can do this right here by using an r code block, like this:

```

samples <- rnorm(100, mean=0, sd=1)
hist(samples)

```

## Histogram of samples



```
numbers <- 1:1000
```

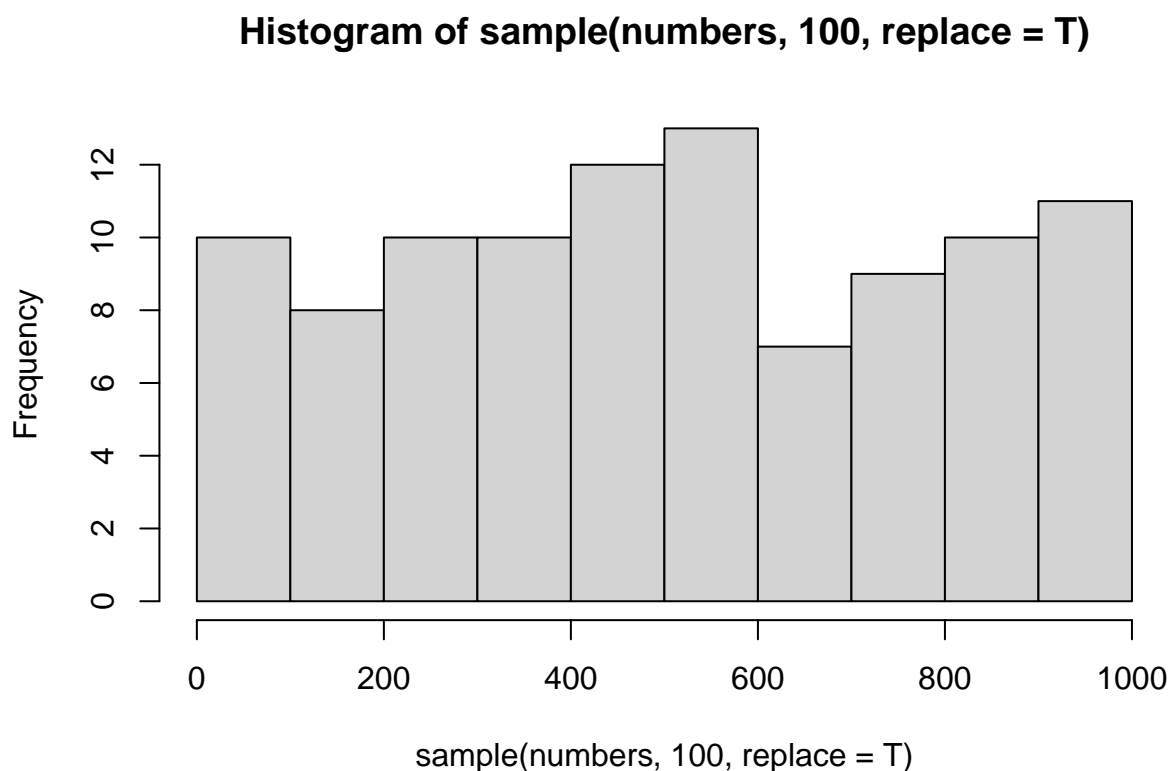
```
# This will print the first 10 elements of the vector numbers
```

```
numbers[1:10]
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

```
# This will plot a histogram of 100 random elements of the vector numbers
```

```
hist(sample(numbers, 100, replace = T))
```



When you knit this R Markdown document, you will see that the histogram is printed to the page, along with the R code. This document can be set up to hide the R code in the webpage, just delete the comment (hashtag) from the cold folding option in the yaml header up top. For purposes of letting yourself see the code, and me see the code, best to keep it the way that it is. You'll learn that all of these things and more can be customized in each R code block.