Exercise 1: Assignment

Philipp Kling

15.06.2020

Contents


```
library(knitr)

### Global options
options(max.print="150")
opts_chunk$set(echo=FALSE,
                 cache=FALSE,
               prompt=FALSE,
               tidy=TRUE,
               comment=NA,
               message=FALSE,
               warning=FALSE)
opts_knit$set(width=75)
rm(list = ls())
```

Assignment 1: RMarkdown

- Change the author of this file to your name.
- Change the settings of this file so that the code is by default shown.

Assignment 2: Import dataset

In the data folder under this assignment (data/ex2) you will find two datasets (twitter.Rda, guardian.Rda). These two datasets contain information retrieved from the respective APIs of the two platforms:

- **guardian.Rda**: articles of The Guardian obtained via its API. Contains information about an article, its author and its content.
- **twitter.Rda**: Tweets published by the account 'guardian' on Twitter. These Tweets often represent links to articles of The Guardian and contain information about the reactions to those articles (such as how many people favorited these statuses).

Assignment: Find a suitable variable to merge these two datasets on and then create a combined dataset that contains information about both the article characteristics (from the Guardian API) and the Twitter characteristics. Then use the **merge()** command to merge the two datasets. Report how many observations you lose from each original dataset.

Assignment 3: Insights

- Does the page number where the article occurred in the newspaper have a positive or negative correlation with the number of Retweets an article received?
- Do articles about music get more or less frequently liked ("favorited") than sport articles?