Supplement 1

We have seen the main data types of R, i.e. logical values, integers, doubles, and characters. R has some other types that build up on top of these. For instance, R represents dates as the number of days since 1907-01-01 and datetimes as the number of seconds since then. For the corona data we will only work with dates to keep it simple and thus drop any information about the daytime. It is possible to create dates either from numeric values or from strings. The default time zone is UTC.

```
tomorrow <- as.Date(1, origin = "2020-03-27")
newyear <- as.Date("2020-01-01")
as.integer(tomorrow)

# [1] 18349
```

So since 1970-01-01 18349 days passed by. Cause dates are represented as integers we can do math with them:

```
tomorrow - newyear

# Time difference of 87 days
```

R automatically considers leap years and even leap seconds. There are many useful functions to work with dates, e.g.

```
weekdays(newyear) # Language is system dependend.

# [1] "Mittwoch"
```

Unfortunately, dates are often not in the standard format given by `YYYY-mm-dd`. But we can specify the format for R using special placeholders.

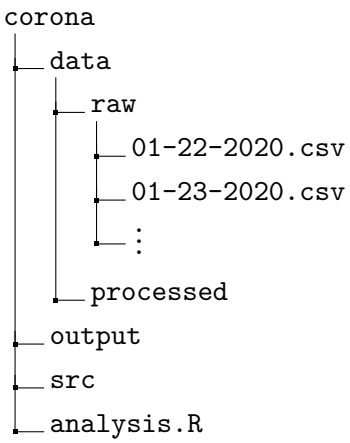
```
my_date <- "01/26/2015 07:24pm"
as.Date(my_date, format = "%m/%d/%Y") # Time is ignored.

# [1] "2015-01-26"
```

There are many letters with special meanings for declaring the proper format. Check the help page for `strptime`.

Exercise 1

In this exercise we will start to develop our corona project. Make sure to create a project folder called `corona` with a file `analysis.R` and a subfolder `data` as mentioned on the last exercise sheet. In addition create subfolders `output` and `src` and subfolders in `data` called `raw` and `processed`. The folder structure should look like the following. Put the data for daily reports in the `raw` folder. The project folder should look like this:



Now, we can start to work on our project. The code for the following should go to the `analysis.R` file.
For the following exercises make sure to always use relative paths. The root is always your top-level project folder.

(a) For now we will work with a single date. Import the data for March 18 into R. Make sure to not convert all strings to factors. Save the data to a variable called `data_corona`.

(b) Have a look on the data. Most columns should be clear. I guess longitude and latitude are the coordinates for presenting the numbers on a world map (but I don't know exactly).

(c) Convert the datetimes in `Last.Update` to dates using `as.Date`.

(d) Save the data frame as `processed_data.csv` in the `processed` folder. Look on the help page for `write.csv` and set the proper argument accordingly to omit row names. Save it also as an excel file for your favorite journal.

*Exercise 2

This exercise is for the upcoming lecture to prepare you for the content or to sensibilize you for a certain aspect.
Sometimes we have more data then we can deal with or that we need to answer our specific question. In this project for instance we have data for different regions in one country. It

is always a trade-off between how deeply we want to analyze our data and how complex the project will be. For this project for instance we will not consider each region separately but look at whole countries and we will ignore daytimes for the updates and only look at the date. This makes the analysis easier without loosing to much information.