T1.1 Data set of Covid-19 is taken from the website https://github.com/daenuprobst/covid19-cases-switzerland. Go further into the github and search for “covid19-cases-switzerland (or just download the raw data in here https://github.com/daenuprobst/covid19-cases-switzerland/blob/master/covid_19_data_switzerland.xlsx). Furthermore, I am interested how many persons are hospitalized in canton Zurich and in Switzerland. Therefore, I copy those data from spreadsheet “hospitalized”, named the header as “hospitalized_ZH” resp. “hospitalized_CH” and put it next to the column “CH” in the spreadsheet “Cases”.

After checking up the data on excel, I saved the file as data type .csv on my computer and named the data file as “Memeti_Nurdzane”.csv in the same folder as the following R-Script.

T2.1 Yes, the data is in a tidy format because each variable forms a column, each observation forms a row, and each type of observational unit forms a table.

T2.2 and T2.3 The first value is the date in data type “character. Further variables are the names of cantons in Switzerland, which are of data type “integer”. For my goal questions I added two more variables “Hospitalized_ZH” and “Hospitalized_CH”, which are also of data type “integer”.

T2.4 257 observations

T2.5 Output shows some missing data pattern in multivariate data. However, it illustrates blue pattern as observed data and red pattern as missing data. In my dataset it is not suprising that the column “Date”, “CH” and “Hospitalized_CH” have no missing data characterized. The cantons “Genf” and “Graubünden” have the lowest missing data values, whereas the cantons “AR” and “AI” have the highest missing data values.

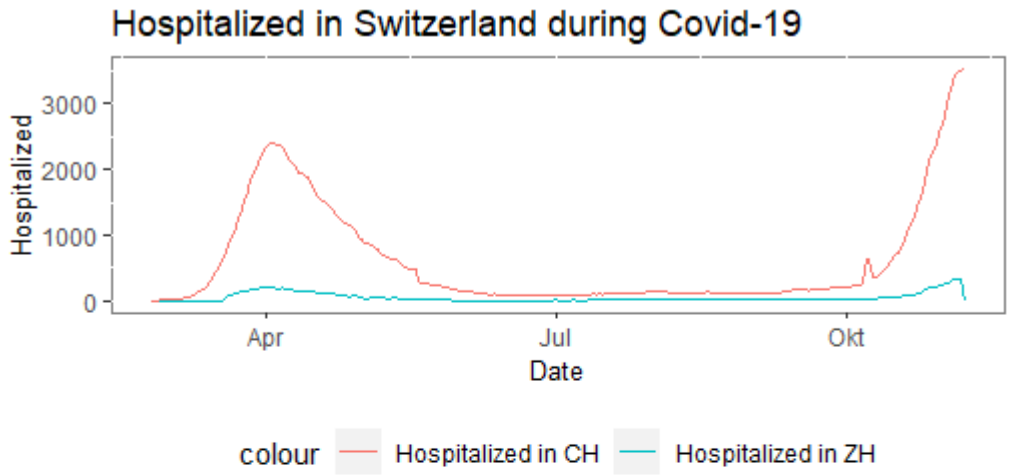
T3.1 My goal for this projects are answering the following questions: 1. How do the recorded new cases of infection between the canton Zurich and Switzerland over time? 2. In addition, how many hospitalized persons are in canton Zurich resp. in Switzerland over the time. As inspiration I took the existing visualisation “Veränderung in Hospitalisierungen“ from webpage: https://www.corona-data.ch/.

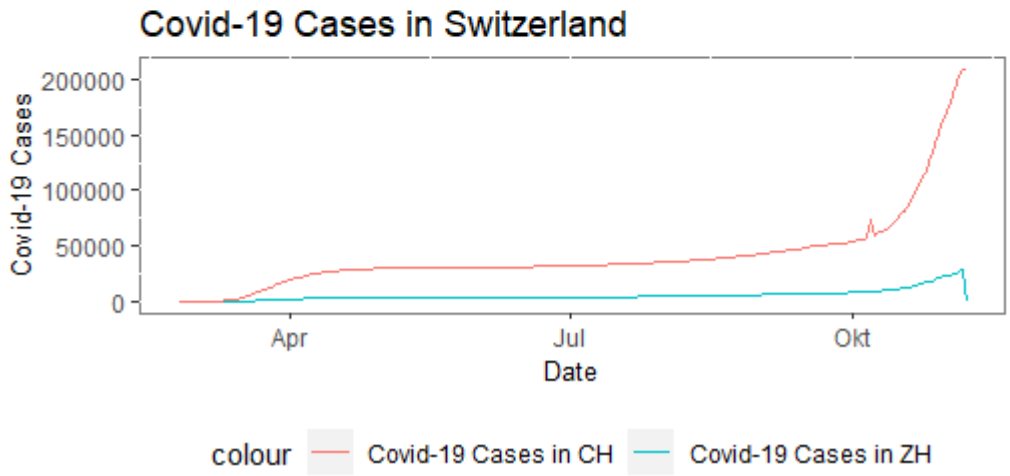
T4.1 In the dataset “mydata” I rename the the first column as “Date”. Furthermore, I converted the data type of the first variable “Date” from datatype "character" to "date". After that, I changed the missing values from “NA” to zero. There were also other options for missing data like “na.omit(mydata)”. But after I tried this function, I realized that they were less observations of the dataset. Although, it is not the best choice I made to exchange values “NA” to value “zero”, for this project I continued with this method. Further changes I already explained in the beginning, see T1.1.

5.1 and 5.2 As dataset I choose data values of hospitalized cases in canton Zurich. After using the Measures of central tendency, I wasn’t sure if the mean value is representative for this column. Finally, it was not representative, because I had many outlier who pool the dataset. Therefore, median is a better choice for outliers and for skeved data sets. My dataset is an asymmetric distribution.

T6.1 I construct a basic one layer plot with variable “Hospitalized_ZH”.

T6.2 Furthermore, I added a second layer with variable “Hospitalized_CH” for my first goal question. T6.3, T6.4 and T6.5: For the aesthetics I labelled with x-, y- axis, title and a subtitle. For the better plot, I changed the position of the legend. Furthermore I changed the size of my labelled title with the function “theme”. For my question 2 I repeated the tasks T6.2-T6.5.





In the following I will discuss the tufte's rules of data visualization in plot “Memeti_Nurdzane_plot.pdf”.

T7.1 I wanted to maximize the ratio of data to ink in my visualization and used just two different colours in plot lines. Furthermore, I deleted the grid of the origin basic plot and set the background of the plot “white”.

T7.2 In my plot I set minimum of visual elements and markings, because I wanted to show the course of hospitalized people in Switzerland aswell as new cases in Switzerland during Covid-19. I avoided chartjunk in my plot with deleting the dark grid lines. In addition, I set the legend table on the bottom to not take away the attention of the trend lines.

T7.3 I displayed my plots within a rectangular shape to follow tufte’s principle. I personally like to show time-series plot in rectangular shapes, because it shows better the curves of the dataset.

Hereby I affirm that I have read the Code of Conduct provided in the assignment, and that I have complied with it.