Importing data

**readxl package only for importing xlsx & xls file**

library("readxl")

data= read\_excel("G:/quarantine/data/orderlines.xlsx")

## New names:  
## \* `` -> ...1

data1= read\_xlsx("G:/quarantine/data/orderlines.xlsx")

## New names:  
## \* `` -> ...1

### readr package

**when a excel file of two column is separated by comma & semi-colon then we use csv and csv2 respectly**

library(dplyr)

library(readr)

data= read\_csv("G:/quarantine/data/Weather.csv")

## Parsed with column specification:  
## cols(  
## code = col\_double(),  
## province = col\_character(),  
## date = col\_date(format = ""),  
## avg\_temp = col\_double(),  
## min\_temp = col\_double(),  
## max\_temp = col\_double(),  
## precipitation = col\_double(),  
## max\_wind\_speed = col\_double(),  
## most\_wind\_direction = col\_double(),  
## avg\_relative\_humidity = col\_double()  
## )

dplyr::glimpse(data)

## Observations: 24,927  
## Variables: 10  
## $ code <dbl> 10000, 11000, 12000, 13000, 14000, 15000, 160...  
## $ province <chr> "Seoul", "Busan", "Daegu", "Gwangju", "Incheo...  
## $ date <date> 2016-01-01, 2016-01-01, 2016-01-01, 2016-01-...  
## $ avg\_temp <dbl> 1.2, 5.3, 1.7, 3.2, 3.1, 1.6, 4.0, 1.6, 5.1, ...  
## $ min\_temp <dbl> -3.3, 1.1, -4.0, -1.5, -0.4, -4.2, -1.6, -4.2...  
## $ max\_temp <dbl> 4.0, 10.9, 8.0, 8.1, 5.7, 7.7, 12.0, 5.7, 8.9...  
## $ precipitation <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...  
## $ max\_wind\_speed <dbl> 3.5, 7.4, 3.7, 2.7, 5.3, 4.4, 2.7, 2.1, 9.6, ...  
## $ most\_wind\_direction <dbl> 90, 340, 270, 230, 180, 320, 320, 180, 290, 1...  
## $ avg\_relative\_humidity <dbl> 73.0, 52.1, 70.5, 73.1, 83.9, 77.4, 53.3, 80....

data1= data %>% sample\_n(10, F)  
  
write\_csv(data1, "data1.csv")  
  
data= read\_csv("G:/quarantine/data/Weather.csv", col\_names = F)

## Parsed with column specification:  
## cols(  
## X1 = col\_character(),  
## X2 = col\_character(),  
## X3 = col\_character(),  
## X4 = col\_character(),  
## X5 = col\_character(),  
## X6 = col\_character(),  
## X7 = col\_character(),  
## X8 = col\_character(),  
## X9 = col\_character(),  
## X10 = col\_character()  
## )

data= read\_csv("G:/quarantine/data/Weather.csv", skip= 1, col\_names = F)

## Parsed with column specification:  
## cols(  
## X1 = col\_double(),  
## X2 = col\_character(),  
## X3 = col\_date(format = ""),  
## X4 = col\_double(),  
## X5 = col\_double(),  
## X6 = col\_double(),  
## X7 = col\_double(),  
## X8 = col\_double(),  
## X9 = col\_double(),  
## X10 = col\_double()  
## )

data= read\_csv("G:/quarantine/data/Weather.csv",  
 col\_names = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"))

## Parsed with column specification:  
## cols(  
## A = col\_character(),  
## B = col\_character(),  
## C = col\_character(),  
## D = col\_character(),  
## E = col\_character(),  
## F = col\_character(),  
## G = col\_character(),  
## H = col\_character(),  
## I = col\_character(),  
## J = col\_character()  
## )

data= read\_csv("G:/quarantine/data/Weather.csv",  
 skip = 1,   
 col\_names = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"))

## Parsed with column specification:  
## cols(  
## A = col\_double(),  
## B = col\_character(),  
## C = col\_date(format = ""),  
## D = col\_double(),  
## E = col\_double(),  
## F = col\_double(),  
## G = col\_double(),  
## H = col\_double(),  
## I = col\_double(),  
## J = col\_double()  
## )

data= read\_csv("G:/quarantine/data/Weather.csv", n\_max = 10)

## Parsed with column specification:  
## cols(  
## code = col\_double(),  
## province = col\_character(),  
## date = col\_date(format = ""),  
## avg\_temp = col\_double(),  
## min\_temp = col\_double(),  
## max\_temp = col\_double(),  
## precipitation = col\_double(),  
## max\_wind\_speed = col\_double(),  
## most\_wind\_direction = col\_double(),  
## avg\_relative\_humidity = col\_double()  
## )

data= read\_csv("G:/quarantine/data/test.csv", na= c("."))

## Parsed with column specification:  
## cols(  
## ForecastId = col\_double(),  
## Province\_State = col\_character(),  
## Country\_Region = col\_character(),  
## Date = col\_date(format = "")  
## )

data= read\_csv("G:/quarantine/data/test.csv")

## Parsed with column specification:  
## cols(  
## ForecastId = col\_double(),  
## Province\_State = col\_character(),  
## Country\_Region = col\_character(),  
## Date = col\_date(format = "")  
## )

data= read\_delim("G:/quarantine/data/ACWI.csv", delim= ",")

## Parsed with column specification:  
## cols(  
## Date = col\_date(format = ""),  
## Open = col\_double(),  
## High = col\_double(),  
## Low = col\_double(),  
## Close = col\_double(),  
## `Adj Close` = col\_double(),  
## Volume = col\_double()  
## )

read\_delim("G:/quarantine/data/ACWI.csv", delim= ",", col\_names = F)

## Parsed with column specification:  
## cols(  
## X1 = col\_character(),  
## X2 = col\_character(),  
## X3 = col\_character(),  
## X4 = col\_character(),  
## X5 = col\_character(),  
## X6 = col\_character(),  
## X7 = col\_character()  
## )

## # A tibble: 254 x 7  
## X1 X2 X3 X4 X5 X6 X7   
## <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Date Open High Low Close Adj Close Volume   
## 2 2019-03-25 71.279999 71.559998 71.019997 71.330002 69.599930 4008400  
## 3 2019-03-26 71.790001 72.070000 71.510002 71.820000 70.078041 2033200  
## 4 2019-03-27 71.820000 71.940002 71.089996 71.489998 69.756042 2407800  
## 5 2019-03-28 71.550003 71.760002 71.300003 71.699997 69.960945 1878100  
## 6 2019-03-29 72.110001 72.160004 71.769997 72.150002 70.400040 3205800  
## 7 2019-04-01 72.760002 73.070000 72.680000 73.029999 71.258690 2565200  
## 8 2019-04-02 73.029999 73.059998 72.769997 73.000000 71.229424 1889700  
## 9 2019-04-03 73.410004 73.639999 73.180000 73.339996 71.561165 3175500  
## 10 2019-04-04 73.300003 73.470001 73.199997 73.449997 71.668503 1991000  
## # ... with 244 more rows

read\_csv("G:/quarantine/data/ACWI.csv")

## Parsed with column specification:  
## cols(  
## Date = col\_date(format = ""),  
## Open = col\_double(),  
## High = col\_double(),  
## Low = col\_double(),  
## Close = col\_double(),  
## `Adj Close` = col\_double(),  
## Volume = col\_double()  
## )

## # A tibble: 253 x 7  
## Date Open High Low Close `Adj Close` Volume  
## <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2019-03-25 71.3 71.6 71.0 71.3 69.6 4008400  
## 2 2019-03-26 71.8 72.1 71.5 71.8 70.1 2033200  
## 3 2019-03-27 71.8 71.9 71.1 71.5 69.8 2407800  
## 4 2019-03-28 71.6 71.8 71.3 71.7 70.0 1878100  
## 5 2019-03-29 72.1 72.2 71.8 72.2 70.4 3205800  
## 6 2019-04-01 72.8 73.1 72.7 73.0 71.3 2565200  
## 7 2019-04-02 73.0 73.1 72.8 73 71.2 1889700  
## 8 2019-04-03 73.4 73.6 73.2 73.3 71.6 3175500  
## 9 2019-04-04 73.3 73.5 73.2 73.4 71.7 1991000  
## 10 2019-04-05 73.6 73.8 73.5 73.7 71.9 1064000  
## # ... with 243 more rows

**spss using haven package**

library(haven)

## Warning: package 'haven' was built under R version 3.6.3

data= read\_sav("G:/covid-19/regular\_case(covid-19)1.sav")  
data %>% class()

## [1] "tbl\_df" "tbl" "data.frame"

data= read\_dta("G:/covid-19/regular\_case(covid-19)1.dta")  
data %>% class()

## [1] "tbl\_df" "tbl" "data.frame"

**using foreign package**

library(foreign)  
  
data= read.csv("G:/quarantine/data/test.csv", header = T)  
data %>% class()

## [1] "data.frame"

data= read.spss("G:/covid-19/regular\_case(covid-19)1.sav",  
 to.data.frame = T)

## re-encoding from UTF-8

data %>% class()

## [1] "data.frame"

data= read.dta("G:/covid-19/regular\_case(covid-19)1.dta")  
data %>% class()

## [1] "data.frame"