**#IMPORTING FLAT FILES USING UTILS PACKAGE**

**-Format read.csv :**

**\*read.csv(file, header = TRUE, sep =”,”, dec = “.”)**

**\*read.csv2(file, header = TRUE, sep = “;”, dec =”,”)**

**-Read.CSV :**

**\*read.csv(“states.csv”, stringAsFactors=FALSE)**

\*by default stringAsFactors are TRUE, header = TRUE, sep =”,”

**-Format read.delim :**

**\*read.delim(dec =”.”)**

**\*read.delim2(dec =”,”)**

**-Read for delimited file : pemisahnya tab**

**\*read.delim(“states.txt”, stringAsFactors = FALSE)**

\*by default header =TRUE, sep =”\t”

**-Read any tabular file as a data frame :**

**\*read.table(“states2.txt”, header = TRUE, sep = “/”, stringAsFactors = FALSE)**

**\*by deafault header = FALSE, sep=””**

**\*bisa file CSV atau txt**

**-Jika tidak ada nama colomnya dan ingin menamai kolom tersebut :**

hotdogs <- read.table(path,

                      sep = "\t",

                      col.names = c("type", "calories", "sodium"))

**\*sep = “\t”, separator dengan tab**

**-Select nilai minimum dan maximum dari table :**

# Finish the read.delim() call

hotdogs <- read.delim("hotdogs.txt", header = FALSE, col.names = c("type", "calories", "sodium"))

# Select the hot dog with the least calories: lily

lily <- hotdogs[which.min(hotdogs$calories), ]

# Select the observation with the most sodium: tom

tom <- hotdogs[which.max(hotdogs$sodium), ]

**-Mengubah type data per column :**

# Previous call to import hotdogs.txt

hotdogs <- read.delim("hotdogs.txt", header = FALSE, col.names = c("type", "calories", "sodium"))

# Display structure of hotdogs

str(hotdogs)

# Edit the colClasses argument to import the data correctly: hotdogs2

hotdogs2 <- read.delim("hotdogs.txt", header = FALSE,

                       col.names = c("type", "calories", "sodium"),

                       colClasses = c("factor", "NULL", "numeric"))

# Display structure of hotdogs2

str(hotdogs2)

**\*jika colClasses nya NULL, maka data tersebut tidak akan ditampilkan dalam table**

**#READR & DATA.TABLE**

**-read\_csv**

# Load the readr package

library(readr)

# Import potatoes.csv with read\_csv(): potatoes

potatoes <- read\_csv("potatoes.csv")

potatoes

**-read\_tsv : tab separated values**

# readr is already loaded

# Column names

properties <- c("area", "temp", "size", "storage", "method",

                "texture", "flavor", "moistness")

# Import potatoes.txt: potatoes

potatoes <- read\_tsv("potatoes.txt", col\_names=properties)

# Call head() on potatoes

head(potatoes)

**-read\_delim :**

**\*read\_delim(“states2.txt”, delim = “/”)**

\*tidak perlu specify ke header, stringAsFactors

\*by Default header = TRUE

**-menambahkan nama colom :**

**\*read\_delim(“states2.txt”, delim = “/”, col\_names = c(“state”, “city”, “pop”, “area”))**

**-col\_types :**

**Col types will be guest first 30 row for each column**

**\*read\_delim(“states2.txt”, delim=”/”, col\_types = “cdil”)**

\*c for character, d for double or numeric (decimal), i for integer, l for logical, \_ for skip the column and the result will not show up in the data

**-skip and n\_max : untuk melihat data ke berapa dan maximum row yg ingin diambil**

**\*read\_delim(“states2.txt”, delim = “/”, skip =2, n\_max=3)**

\*skip 2 row, mengambil 3 row

**-contoh readr delim :**

# readr is already loaded

# Column names

properties <- c("area", "temp", "size", "storage", "method",

                "texture", "flavor", "moistness")

# Import potatoes.txt using read\_delim(): potatoes

potatoes <- read\_delim("potatoes.txt", delim="\t", col\_names=properties)

# Print out potatoes

potatoes

**-Contoh readr skip dan n\_max :**

# readr is already loaded

# Column names

properties <- c("area", "temp", "size", "storage", "method",

                "texture", "flavor", "moistness")

# Import 5 observations from potatoes.txt: potatoes\_fragment

potatoes\_fragment <- read\_tsv("potatoes.txt", skip = 6, n\_max = 5, col\_names = properties)

potatoes\_fragment

**-contoh mengubah col\_types :**

# readr is already loaded

# Column names

properties <- c("area", "temp", "size", "storage", "method",

                "texture", "flavor", "moistness")

# Import all data, but force all columns to be character: potatoes\_char

potatoes\_char <- read\_tsv("potatoes.txt", col\_types = "cccccccc", col\_names = properties)

# Print out structure of potatoes\_char

str(potatoes\_char)

**-example of another change col\_types**

# readr is already loaded

# Import without col\_types

hotdogs <- read\_tsv("hotdogs.txt", col\_names = c("type", "calories", "sodium"))

# Display the summary of hotdogs

summary(hotdogs)

# The collectors you will need to import the data

fac <- col\_factor(levels = c("Beef", "Meat", "Poultry"))

int <- col\_integer()

# Edit the col\_types argument to import the data correctly: hotdogs\_factor

hotdogs\_factor <- read\_tsv("hotdogs.txt",

                           col\_names = c("type", "calories", "sodium"),

                           col\_types = list(fac,int,int))

hotdogs\_factor

# Display the summary of hotdogs\_factor

summary(hotdogs\_factor)

**-fread :**

More faster, more simple. Function like data.table

**\*library(data.table)**

**\*potatoes <- fread(“potatoes.csv”)**

**-select column from fread :**

**Contoh ada column a-e**

**\*fread(“file.txt”, drop = 2:4) = membuang kolom 2-4**

**\*fread(“file.txt”, select = c(1,5)) = memilih kolom 1 dan 5**

**\*fread(“file.txt”,drop = c(“b”,”c”,”d”)) = membuang kolom b,c ,dan d**

**\*fread(“file.txt”,select = c(“a”, “e”)) = memilih kolom a dan e**

**#IMPORTING EXCEL DATA**

**-mengetahui file aktif :**

**\*dir()**

**-mengetahui sheet dalam file :**

**\*excel\_sheets(“nama\_file”)**

**-Import file :**

**\*read\_excel(“nama\_file”) : otomatis akan mengambil sheet pertama**

**\*read\_excel(“nama\_file”, sheets =”nama\_sheets”/index\_sheets)**

**-Import sheet dan menggabungkan sheet :**

# The readxl package is already loaded

# Read the sheets, one by one

pop\_1 <- read\_excel("urbanpop.xlsx", sheet = 1)

pop\_2 <- read\_excel("urbanpop.xlsx", sheet = 2)

pop\_3 <- read\_excel("urbanpop.xlsx", sheet = 3)

# Put pop\_1, pop\_2 and pop\_3 in a list: pop\_list

pop\_list <- list(pop\_1,pop\_2,pop\_3)

# Display the structure of pop\_list

str(pop\_list)

**-Import excel, with multiple sheet :**

**library(readxl)**

**setwd("D:")**

**test <- lapply(excel\_sheets("SSIT Data Nov QlikSense.xlsx"), read\_excel,**

**path=("SSIT Data Nov QlikSense.xlsx"))**

**str(test)**

**-import exel with col\_names and col\_types :**

**\*read\_excel(“nama\_file”, sheet =1, col\_names = TRUE, col\_types = NULL, skip =0)**

**By default read\_excel akan membaca nama kolom, dan type data akan otomatis membaca sesuai data, dan tanpa ada yg di skip.**

**-merubah column types :**

**\*read\_excel(“nama\_file”, sheet =1, col\_types = c(“text”, “text”))**

**Jika col\_types = “blank”, maka kolom tersebut akan diignore**

**-merubah column names :**

# The readxl package is already loaded

# Import the first Excel sheet of urbanpop\_nonames.xlsx (R gives names): pop\_a

pop\_a <- read\_excel("urbanpop\_nonames.xlsx", sheet =1, col\_names = FALSE)

# Import the first Excel sheet of urbanpop\_nonames.xlsx (specify col\_names): pop\_b

cols <- c("country", paste0("year\_", 1960:1966))

pop\_b <- read\_excel("urbanpop\_nonames.xlsx", sheet =1, col\_names = cols)

# Print the summary of pop\_a

summary(pop\_a)

# Print the summary of pop\_b

summary (pop\_b)

**-Skip Row :**

# The readxl package is already loaded

# Import the second sheet of urbanpop.xlsx, skipping the first 21 rows: urbanpop\_sel

urbanpop\_sel <- read\_excel("urbanpop.xlsx", sheet=2, col\_names=FALSE, skip=21)

# Print out the first observation from urbanpop\_sel

head(urbanpop\_sel, n=1)

**-gdata :**

**It works with two ways with perl. First convert to csv, and then reading it with read.csv**

**-read data :**

columns <- c("country", paste0("year\_", 1967:1974))

# Finish the read.xls call

urban\_pop <- read.xls("urbanpop.xls", sheet = 2,

                      skip = 50, header = FALSE, stringsAsFactors = FALSE,

                      col.names = columns)

**paste0 function : copy paste year and the 1967-1974**

**-bring it all gdata :**

# Add code to import data from all three sheets in urbanpop.xls

path <- "urbanpop.xls"

urban\_sheet1 <- read.xls(path, sheet = 1, stringsAsFactors = FALSE)

urban\_sheet2 <- read.xls(path, sheet = 2, stringsAsFactors = FALSE)

urban\_sheet3 <- read.xls(path, sheet = 3, stringsAsFactors = FALSE)

# Extend the cbind() call to include urban\_sheet3: urban

urban <- cbind(urban\_sheet1, urban\_sheet2[-1], urban\_sheet3[-1])

# Remove all rows with NAs from urban: urban\_clean

urban\_clean <- na.omit(urban)

# Print out a summary of urban\_clean

summary(urban\_clean)

is.na(urban\_clean)

**\*urban\_sheet2[-1] : menghapus column ke-1**

**\*na.omit(urban) : menghapus row yg NA**

**\*is.na(urban) : menguji ada nilai NA gak di row tersebut**

**#REPRODUCIBLE EXCEL WORK WITH XLCONNECT**

**-XL connect :**

**Memugkinkan untuk mengedit data di excel dengan R**

**-membaca file :**

# urbanpop.xlsx is available in your working directory

# Load the XLConnect package

library("XLConnect")

# Build connection to urbanpop.xlsx: my\_book

my\_book <- loadWorkbook("urbanpop.xlsx")

# Print out the class of my\_book

class(my\_book)

**\*class function : melihat jenis data. Data frmae, vector, atau matrix**

**-melihat sheet :**

**getSheets(my\_book)**

**-Combine from the sheet :**

# XLConnect is already available

# Build connection to urbanpop.xlsx

my\_book <- loadWorkbook("urbanpop.xlsx")

# Import columns 3, 4, and 5 from second sheet in my\_book: urbanpop\_sel

urbanpop\_sel <- readWorksheet(my\_book, sheet = 2, startCol=3, endCol=5)

# Import first column from second sheet in my\_book: countries

countries <- readWorksheet(my\_book, sheet=2, startCol=1, endCol=1)

# cbind() urbanpop\_sel and countries together: selection

selection <- cbind(countries, urbanpop\_sel)

**\*tambahan function : startRow & endRow. Sama dengan function startCol dan endCol**

**-add new sheet from R directly to Excel :**

**pop\_2010 <- data.frame(Capital = c("New York", "Berlin", "Madrid","Stockholm"),**

**Population = c(8191900, 3460725, 3273000, 1372565))**

**library(XLConnect)**

**setwd("D:")**

**my\_book <- loadWorkbook("SSIT Data Nov QlikSense.xlsx")**

**createSheet(my\_book, name="test")**

**writeWorksheet(my\_book, pop\_2010, sheet = "test")**

**saveWorkbook(my\_book , file = “cities2.xlsx”)**

**-rename\_sheet :**

**\*renameSheet (book, “file\_lama”, “file\_baru”)**

**\*saveWorkbook(book, file = “cities3”)**

**-removeSheet :**

**\*removeSheet(book, sheet = “nama\_sheet”)**

**\*saveWorkbook(book, file = “cities3”)**

**-modify sheet :**

# XLConnect is already available

# Build connection to urbanpop.xlsx

my\_book <- loadWorkbook("urbanpop.xlsx")

# Add a worksheet to my\_book, named "data\_summary"

createSheet(my\_book, "data\_summary")

# Create data frame: summ

sheets <- getSheets(my\_book)[1:3]

dims <- sapply(sheets, function(x) dim(readWorksheet(my\_book, sheet = x)), USE.NAMES = FALSE)

dims

summ <- data.frame(sheets = sheets,

                   nrows = dims[1, ],

                   ncols = dims[2, ])

summ

# Add data in summ to "data\_summary" sheet

writeWorksheet(my\_book, summ, sheet="data\_summary")

# Save workbook as summary.xlsx

saveWorkbook(my\_book, file="summary.xlsx")