

Data_Frame.R

```

#=====
# DataFrames.R
#=====
# Description
# The function data.frame() creates data frames, tightly coupled collections of
# variables which share many of the properties of matrices and of lists, used
# as the fundamental data structure by most of R's modeling software.
# A data frame is used for storing data tables. It is a list of vectors of
# equal length.
#-----

cars <- mtcars      # working with build-in data

# getting some info of the data frame cars
#-----
head(cars)          # looking at the beginning of the data
tail(cars)           # looking at the end of the data of cars

names(cars)          # finding the columns names of cars

nrow(cars)           # find the numbers of rows in the data frame cars
ncol(cars)           # find the numbers of columns in the data frame cars

#-----
# get the data from the first row sixth column
#-----
cars[1,6]            # using the index method
cars[1,'wt']         # using row number and column names

cars$wt[1]           # reference method

#-----
# get the data from the entire column(s) and row(s)
#-----
cars$wt              # get all the data from the 'wt' column

cars[1, 1:4]         # get the first four column of the first row
cars[3:7, 1:4]       # get the first four column from rows 3-7
cars[, 1:4]          # get the first four column

cars[c(1,5,10:15), c(2, 5, 7:10)] # more selective range of slective

#-----
# setting the data in data frame
#-----
cars2 <- cars[1:10, ] # making a copy of cars with only the first 10 rows

```

[illegible]

```

#-----
# Creating data frame from scratch
#-----
# creating a data.frame
df <- data.frame(firstname = 'Jeff',
                  lastname = 'Ward',
                  age = 22,
                  member = FALSE)

df

#-----
# add a row to the data.frame
#-----
df <- rbind(df, data.frame(firstname = 'Bill', lastname = 'Smith', age = 20, member =
TRUE))
df

#-----
# Doing math and plotting
#-----
cars2$mpgpercyl <- cars2$mpg / cars2$cyl # create new col. with math operation

cars2 # display cars2 on the console

hist(cars2$hp)

pairs(cars2[2:6], )

plot(cars$mpg, type = 'l')

#-----
# Doing simple OLS regression
#-----
fit <- lm(mpg ~ hp, data = cars2) # do a simple regression
summary(fit) # output regression results

plot(cars2$hp, cars2$mpg) # scatter plot data
abline(fit, col="lightblue") # plot OLS line

```

Project Code

```
#=====
# Data Frame Project
#=====
```

```
#-----
# Student Coding
#-----
```

```
# Create a Data frame with members
# YEAR, COUNTRYNAME, GDP, CPI
# 2016, USA, 57591.197652, 110.067
```

```
# Hint Code
df <- data.frame(YEAR = 2017 ,
                  COUNTRYNAME = "PUT SOME CODE HERE" ,
                  "PUT SOME CODE HERE" = "PUT SOME CODE HERE",
                  "PUT SOME CODE HERE" = "PUT SOME CODE HERE"
                )
```

```
#-----
# Student Code
#-----
```

```
# add other country data to the data frame
# DATE, COUNTRYNAME, GDP, CPI
# 2016, MEX, 2266350.4612, 122.7801
# hint: add a row
#-----
```

```
df <- rbind(df, "PUT SOME CODE HERE")
```

```
#=====
# Project Coding
#=====
```

```
canada.year <- c(2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
2011,
                2012, 2013, 2014, 2015, 2016)

canada.unemp <- c(6.816667, 7.216667, 7.675, 7.575, 7.191667, 6.758333, 6.333333,
6.05,
                6.141667, 8.35, 8.058333, 7.525, 7.291667, 7.083333, 6.908333,
6.908333,
                6.991667)
```

```
canada.cpi <- c(81.89038, 83.95821, 85.85432, 88.22266, 89.86119, 91.85031, 93.68918,
```

```
95.69262, 97.96079, 98.25415, 100, 102.9121, 104.472, 105.4522,
107.4628,
108.672, 110.2247)
```

```
canada.gdp <- c(898024.6847, 934754.5473, 967558.4379, 1019620.383, 1079987.408,
1167581.235,
1238000.732, 1297164.505, 1339060.553, 1304488.287, 1361136.052,
1427466.543,
1464565.336, 1550269.806, 1617564.49, 1599827.514, 1625361.347)
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```
# creating data frame with vectors
```

```
df.canada <- data.frame(YEAR      = "PUT SOME CODE HERE" ,
                        COUNTRYNAME = 'CAN' ,
                        GDP        = "PUT SOME CODE HERE",
                        CPI        = "PUT SOME CODE HERE"
                        )
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```
# add column 'INF' to data frame df.canada and set all entries to 'NA'
```

```
df.canada$"PUT SOME CODE HERE" <- "PUT SOME CODE HERE"
```

```
#=====
```

```
# Project Coding
```

```
#=====
```

```
# calculating inflation from CPI data and set INF column
```

```
# first for loop
```

```
# look at this code
```

```
for (i in 2:17){
df.canada$INF[i] <- ( ( df.canada$CPI[i] - df.canada$CPI[i-1] ) /df.canada$CPI[i-1]
) * 100
}
```

```
#=====
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```
# BUG IN THIS CODE
```

```
# HINT THINK OF THE COLUMNS of df and df.canada
```

```

# look at both data frame columns in the data viewer
#-----
# Code fix before this line
df <- rbind(df, df.canada)

#-----
# Student Code
#-----
# add column unemp and set its entries to NA

df$UNEMP <- "PUT SOME CODE HERE"

#-----
# Student Code
#-----
# no coding requested
# try to figure out what this for loop is doing
# hint: ran the code and look at the data viewer

for (i in 3:19){
  df$UNEMP[i] <- canada.unemp[i-2]
}

#=====
# Project Coding
# run code
#=====

canada.data <- subset(df, df$COUNTRYNAME == 'CAN')
old.par <- par(mfrow=c(2, 2))
plot(canada.data$YEAR, canada.data$GDP,
      main = 'Canada GDP', xlab = 'Year', ylab = 'GDP',
      type = 'l')

plot(canada.data$YEAR, canada.data$CPI,
      main = 'Canada CPI', xlab = 'Year', ylab = 'CPI',
      type = 'l')

plot(canada.data$YEAR, canada.data$UNEMP,
      main = 'Canada Unemployment', xlab = 'Year', ylab = 'UNEMP',
      type = 'l')

plot(canada.data$YEAR, canada.data$INF,
      main = 'Canada Inflation', xlab = 'Year', ylab = 'INF',
      type = 'l')

```



```
#=====
# Project End
#=====
```

Project Code Solution

```
#=====
# Data Frame Project Solution
#=====
```

```
#=====
# Student Coding Start
#-----
# Student Code
# Create a Data frame with members
# YEAR, COUNTRYNAME, GDP,          CPI
# 2016, USA,          57591.197652,    110.067
#-----
```

```
# Hint Code
df <- data.frame(YEAR      = 2016 ,
                  COUNTRYNAME = 'USA',
                  GDP        = 57591.197652,
                  CPI        = 110.067
                  )
```

```
#-----
# Student Code
#-----
# add other country data to the data frame
# DATE, COUNTRYNAME, GDP,          CPI
# 2016, MEX,          2266350.4612,    122.7801
# hint: add a row
#-----
# Hint Code
# df <- rbind(df, "PUT SOME CODE HERE")
```

```
df <- rbind(df, data.frame(YEAR      = 2016 ,
                           COUNTRYNAME = 'MEX',
                           GDP        = 2266350.4612,
                           CPI        = 122.7801 ))
```

```
#=====
# Project Coding
#=====
```

```
canada.year <- c(2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
2011,
                2012, 2013, 2014, 2015, 2016)

canada.unemp <- c(6.816667, 7.216667, 7.675, 7.575, 7.191667, 6.758333, 6.333333,
6.05,
                6.141667, 8.35, 8.058333, 7.525, 7.291667, 7.083333, 6.908333,
```

6.908333,

6.991667)

```
canada.cpi <- c(81.89038, 83.95821, 85.85432, 88.22266, 89.86119, 91.85031, 93.68918,
               95.69262, 97.96079, 98.25415, 100, 102.9121, 104.472, 105.4522,
               107.4628,
               108.672, 110.2247)
```

```
canada.gdp <- c(898024.6847, 934754.5473, 967558.4379, 1019620.383, 1079987.408,
               1167581.235,
               1238000.732, 1297164.505, 1339060.553, 1304488.287, 1361136.052,
               1427466.543,
               1464565.336, 1550269.806, 1617564.49, 1599827.514, 1625361.347)
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```
# creating data frame with vectors
```

```
df.canada <- data.frame(YEAR      = canada.year ,
                        COUNTRYNAME = 'CAN',
                        GDP        = canada.gdp,
                        CPI        = canada.cpi
                        )
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```
# add column 'INF' to data frame df.canada and set all entries to 'NA'
```

```
df.canada$INF <- NA
```

```
#=====
```

```
# Project Coding
```

```
#=====
```

```
# calculating inflation from CPI data and set INF column
```

```
# first for loop
```

```
# look at this code
```

```
for (i in 2:17){
df.canada$INF[i] <- ( ( df.canada$CPI[i] - df.canada$CPI[i-1] ) /df.canada$CPI[i-1]
) * 100
}
```

```
#=====
```

```
#-----
```

```
# Student Code
```

```
#-----
```

```

# BUG IN THIS CODE
# HINT THINK OF THE COLUMNS of df and df.canada
# look at both data frame columns in the data viewer
#-----
# Code fix before this line
# add INF column to data frame df and set all entries to 'NA'
df$INF <- NA

df <- rbind(df, df.canada)

#-----
# Student Code
#-----
# add column unemp and set its entries to NA

df$UNEMP <- NA

#-----
# Student Code
#-----
# no coding requested
# try to figure out what this for loop is doing
# hint: ran the code and look at the data viewer

for (i in 3:19){
  df$UNEMP[i] <- canada.unemp[i-2]
}

#=====
# Project Coding
# run code
#=====

# search in data frame df and find all CAN rows
canada.data <- subset(df, df$COUNTRYNAME == 'CAN')

# set up plot canvas for 2 rows 2 column plots
old.par <- par(mfrow=c(2, 2))

# plots with label look at these plots calls to see how it is done
plot(canada.data$YEAR, canada.data$GDP,
     main = 'Canada GDP', xlab = 'Year', ylab = 'GDP',
     type = 'l')

plot(canada.data$YEAR, canada.data$CPI,
     main = 'Canada CPI', xlab = 'Year', ylab = 'CPI',

```

```
type = 'l')

plot(canada.data$YEAR, canada.data$UNEMP,
     main = 'Canada Unemployment', xlab = 'Year', ylab = 'UNEMP',
     type = 'l')

plot(canada.data$YEAR, canada.data$INF,
     main = 'Canada Inflation', xlab = 'Year', ylab = 'INF',
     type = 'l')

#=====
# Project End
#=====
```