Hwk 03

Data in R, Fall 2022

# Question 1 - simple lists

## List1

Create list1 shown below:

## [[1]]  
## [1] "Soils 502"  
##   
## [[2]]  
## [1] 1 3 4 5  
##   
## [[3]]  
## [,1] [,2]  
## [1,] 1 2  
## [2,] 3 4  
## [3,] 5 6

Extract the following from list1:

## [,1] [,2]  
## [1,] 1 2  
## [2,] 3 4  
## [3,] 5 6

Extract the following from list1:

## [[1]]  
## [1] "Soils 502"  
##   
## [[2]]  
## [,1] [,2]  
## [1,] 1 2  
## [2,] 3 4  
## [3,] 5 6  
##   
## [[1]]  
## [1] "Soils 502"  
##   
## [[2]]  
## [,1] [,2]  
## [1,] 1 2  
## [2,] 3 4  
## [3,] 5 6

## (b) list2

Create the following list:

## $name  
## $name$Last  
## [1] "Williams"  
##   
## $name$First  
## [1] "Serena"  
##   
##   
## $age  
## [1] 36  
##   
## $record  
## wins losses   
## 801 136   
##   
## $grand.slam.wins  
## $grand.slam.wins$Australian.Open  
## [1] 2003 2005 2007 2009 2010 2015 2017  
##   
## $grand.slam.wins$French.Open  
## [1] 2002 2013 2015  
##   
## $grand.slam.wins$Wimbledon  
## [1] 2002 2003 2009 2010 2012 2015 2016  
##   
## $grand.slam.wins$US.Open  
## [1] 1999 2002 2008 2012 2013 2014

Why aren’t grand slam wins structured as a data frame or matrix?

# Question 2 - cbind, rbind

Create the following vectors and matrices.

## [1] 3 5 6 12

##  
## [1] TRUE FALSE TRUE TRUE

##  
## [1] "A" "B" "C" "D" "E"

##  
## [,1] [,2] [,3] [,4] [,5]  
## [1,] 2 14 26 5 17  
## [2,] 5 17 29 8 20  
## [3,] 8 20 32 11 23  
## [4,] 11 23 2 14 26

##  
## [,1] [,2] [,3] [,4]  
## [1,] 1 -1 1 0  
## [2,] 0 0 -1 1

## (a) cbind

Create the following three outputs using data objects created above and a single *cbind()* command (for each output).

## v2  
## [1,] 3 1  
## [2,] 5 0  
## [3,] 6 1  
## [4,] 12 1

## [,1] [,2]   
## [1,] "A" "TRUE"   
## [2,] "B" "FALSE"  
## [3,] "C" "TRUE"   
## [4,] "D" "TRUE"   
## [5,] "E" "TRUE"

## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] 2 14 26 5 17 3  
## [2,] 5 17 29 8 20 5  
## [3,] 8 20 32 11 23 6  
## [4,] 11 23 2 14 26 12

## (b) rbind

Create the following three outputs using data objects created above and a single *rbind()* command (for each output).

## [,1] [,2] [,3] [,4]  
## [1,] 2 14 26 5  
## [2,] 5 17 29 8  
## [3,] 8 20 32 11  
## [4,] 11 23 2 14  
## [5,] 1 -1 1 0  
## [6,] 0 0 -1 1

## [,1] [,2] [,3] [,4]  
## [1,] 3 5 6 12  
## [2,] 1 -1 1 0  
## [3,] 0 0 -1 1

## [,1] [,2] [,3] [,4] [,5]   
## v1 "3" "5" "6" "12" "3"   
## v2 "TRUE" "FALSE" "TRUE" "TRUE" "TRUE"  
## v3 "A" "B" "C" "D" "E"

# Question 3

## (a) Read in Iris data

Read in the csv Iris data directly from the UC-Irvine Machine Learning data repository into a dataframe (see iris.data link below).

* *Assign appropriate column names* (see metadate on website).
* Make sure columns have the correct data type.
* Use the **summary()** **head()** and/or **View()** functions to avoid dumping a large dataset into the console (and rmarkdown printout).

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris>

## (b) Present summaries (use summary function) of the following subsets

* Iris data with class equal to Iris-versicolor.
* Iris data with petal width greater than 1.5 cm OR sepal width greater than 3.0 cm.
  + Use | compute a logical “OR” . E.g. (x > 6 ⎮ y > 2) gives all observations w/ x greater than 6 OR y greater than 2.

Do you see a relationship between the width measurements and class of Iris?

# Question 4 - Cars

## (a) Read in car data

Read in the car data data provided in the *car.csv* file. These data consists entirely of categorical data, so when reading the file set the **stringsAsFactors** parameter appropriately.

Use the **summary()** **head()** and/or **View()** functions to avoid dumping a large dataset into the console (and rmarkdown printout).

## (b) remove all observations (rows) with missing values

Note that **is.na()** can be used to identify NA values. Show **dim** and **summary** of data with appropriate observations removed.

## (c) remove duplicate observations (rows)

Now remove all duplicate observations from NA-free car dataset. Show your results with the **dim** and **summary** functions.