Week 3 - Class Worksheet

Understand: Understanding Data and Data Structures

Dr. Anil Dolgun

Last updated: 18 June, 2020

Working in small groups or pairs, complete the following exercises.

Required Packages

The following packages will be required or may come in handy.

```
library(dplyr)
library(readr)
library(tidyr)
library(knitr)
```

Exercises

1 Use command `c()` to create vectors as listed below and check their class as you go. For factor class, check it's levels and label it.

i) Integers from 1 to 5 and name it `vect_int` .

ii) Double numeric variables from 0.5 to 3.5 incrementing it 1, and name it `vect_dbl` .

iii) Character variables using name of the colours red, green, blue, yellow, white and name it `vect_char` .

iv) Factor variables using very low, low, medium, high, very high and name it `vect_fact` . Order the levels and name it `vect_fact2` then check the levels again.

2 Use `ordered=TRUE` arguement for `vect_fact2` and name it `vect_fact3` . Type `vect_fact3` in the console to see the structure.

3 Combine `vect_int` and `vect_fact3` using `c()` command, name it as `vect_comb` . Guess the type of `vect_comb` 's class.

4 Use the vectors you created in the previous exercise and create a list and name it `vect_list` . Check the structure of `vect_list` . Add states of Australia as a vector to the list and name it `vect_list2` (Hint: Use `append()` function). Check the structure then name the elements of the list as comp1, comp2, …,comp8.

i) Select the third element of `comp5` .

ii) Select the second, fourth and eighth component of the list all together.

5

Create a 5×4 numeric matrix using `seq(0,36,by=2)` . Check out the warning message, notice that 5th row, 4th column is 0. Explain in a few words the reason of the warning and what this is called. (Hint: Refresh your memory with `swirl` package). Save this matrix as `mat1` , check the structure and attributes of it.

6

Create a matrix from `vect_char` and `vect_fact3` using row-bind and column-bind and name it `m1` and `m2` respectively. Pick a suitable bind function to add `m2` onto `mat1` to create 5×6 matrix, name it `mat2` , check the attributes and structure. Have you noticed that the columns don't have names?

7

Create a matrix with `vect_dbl` and `c(1,2,3,4)` , name it `m3` . Then combine `m2` and `m3` using column-bind. Explain in a few words what went wrong.

8

Add column names to the matrix `mat2` and name it seq1, seq2, seq3, seq4, colours, factor1. Add row names to the matrix `mat2` and name it x1, x2, x3, x4, x5. Check attributes.

9

Create a data frame using vectors `vect_int` and `vect_char` and name it `df1` . Check it's structure. As you can see when creating a data frame from existing vectors with different classes, the structures are carried to the new data frame. Remove the factors from the second column and rename the data frame as `df2` , check the structure and compare with `df1` 's structure.

10

Add `vect_fact3` onto `df1` as a third column and name it `df3` . Check the structure, use `stringsAsFactors` command to remove the factors. What could be the reason why it's not working? Now add `vect_dbl` to `df3` . Discuss the reason why we can't combine `vect_dbl` and `df3` .

11

Add column and row names to `df3` . Set the column names to numbers, colours, scale and row names to r1, r2, r3, r4, r5.

12

Subset `df3` by row numbers, only select the fourth and fifth rows. Then subset `df3` by column numbers, only select first and third columns. For both tasks use subsetting by row/column number and then the row/column name. Subset the third column using `$` operator.

13 Convert `df3` 's columns using `as` .

i) The numbers column into numeric,

ii) The colours and the scale column into character.

iii) Check the structure of `df3` .

14 Convert `mat2` into a dataframe and `df3` to a matrix. Use `is.matrix()` and `is.dataframe()` functions to check the type after you convert.

German General Social Survey Data

The following exercise is based on German General Society Survey germangss.csv (../data/germangss.csv) data set. This data set has 400 rows of categorical data which was used to study what affects political attitude in Germany through 1991-1992. This data set is taken from the book Analyzing Categorical Data by Jeffrey S. Simonoff (Simonoff, J. (2003). Analyzing Categorical Data. New York: Springer New York).

Variables in this data set contains:

`Political_system` : Political attitude

`Age` : Age categories

`Year` : Year that survey is recorded

`Schooling` : Education level

`Region` : Region name in Germany

`binaryClass` : Binary class, P=positive, N=Negative

15 Load germangss.csv (../data/germangss.csv) data set.

i) Find out the types of variables and the data structures. Rename the variables as Political Attitude, Age Category, Year, Education Level, Region and Binary Class.

ii) Check the class of each variable.

iii) Check the structure of the data set.

iv) Convert Political Attitude, Age Category and Education Level columns into factors and order the levels.

v) Convert Year column into numeric.

vi) Subset first, second and the fourth column and first 30 rows. Create a data frame with this subset and name it as `subgss` . Then check it with `is.data.frame()` .

Finished?

If you have finished the above tasks, work through the weekly list of tasks posted on the Canvas announcement page.

Return to Course Website (../index.html)