SAVEETHA SCHOOL OF ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

ITA 0443 - STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEM

DAY 2 – LAB EXERCISES

Reg No:192021033

Name:P.Bhavani

RESHAPE FUNCTION IN R

Exercise: 1

Construct the following data frame ‘country’.

Construct the following data frame ‘country’.

CODE:

country<-data.frame(c("A","B","C"),c(100,200,120),c(2000,7000,15000))

colnames(country)<- c("countries","population\_in\_million","gdp\_percapita")

country

OUTPUT:

> country<-data.frame(c("A","B","C"),c(100,200,120),c(2000,7000,15000))

> colnames(country)<- c("countries","population\_in\_million","gdp\_percapita")

> country

  countries population\_in\_million gdp\_percapita

1         A                   100          2000

2         B                   200          7000

3         C                   120         15000

>

a) Reshape in R from wide to long:

Reshape the above data frame from wide to long format in R.

 data frame “country” is passed to reshape function

 idvar is the variable which need to be left unaltered which is “countries”

 varying are the ones that needs to converted from wide to long

 v.names are the values that should be against the times in the resultant data frame.

 new.row.names is used to assign row names to the resultant dataset

 direction is, to which format the data needs to be transformed

CODE:

country\_w\_to\_L<- reshape(data=country, idvar="countries",

varying = c("population\_in\_million","gdp\_percapita"),

v.name=c("value"),

times=c("population\_in\_million","gdp\_percapita"),

new.row.names = 1:1000,

direction="long")

country\_w\_to\_L

OUTPUT:

> country\_w\_to\_L<- reshape(data=country, idvar="countries",

+ varying = c("population\_in\_million","gdp\_percapita"),

+ v.name=c("value"),

+ times=c("population\_in\_million","gdp\_percapita"),

+ new.row.names = 1:1000,

+ direction="long")

> country\_w\_to\_L

countries time value

1 A population\_in\_million 100

2 B population\_in\_million 200

3 C population\_in\_million 120

4 A gdp\_percapita 2000

5 B gdp\_percapita 7000

6 C gdp\_percapita 15000

b) Reshape in R from long to wide:

 data (country\_w\_to\_L) which is in long format,  is passed to reshape function

 idvar is the variable which need to be left unaltered, which is “countries”

 timevar are the variables that needs to converted to wide format

 v.names are the value variable

 direction is, to which format the data needs to be transformed

CODE:

library(tidyr)

data\_long = gather(country, detail, value, population\_in\_million:gdp\_percapita, factor\_key=TRUE)

data\_long

OUTPUT:

> library(tidyr)

>

> data\_long = gather(country, detail, value, population\_in\_million:gdp\_percapita, factor\_key=TRUE)

> data\_long

countries detail value

1 A population\_in\_million 100

2 B population\_in\_million 200

3 C population\_in\_million 120

4 A gdp\_percapita 2000

5 B gdp\_percapita 7000

6 C gdp\_percapita 15000

7. MELTING AND CASTING IN R

Exercises :

1. Melt airquality data set and display as a long – format data ?

CODE:

names(airquality) <- tolower(names(airquality))

head(airquality)

aql <- melt(airquality)

OUTPUT:

  ozone solar.r wind temp month day

1    41     190  7.4   67     5   1

2    36     118  8.0   72     5   2

3    12     149 12.6   74     5   3

4    18     313 11.5   62     5   4

5    NA      NA 14.3   56     5   5

6    28      NA 14.9   66     5   6

Error in melt(airquality) : could not find function "melt"

Execution halted

[Execution complete with exit code 1]

1. Melt airquality data and specify month and day to be “ID variables” ?
2. CODE:
3. aql <- melt(airquality, id.vars = c("month", "day"))
4. aqw <- dcast(aql, month + day ~ variable)
5. head(aqw)
6. OUTPUT:
7. month day ozone solar.r wind temp
8. 5   1    41     190  7.4   67
9. 2     5   2    36     118  8.0   72
10. 3     5   3    12     149 12.6   74
11. 4     5   4    18     313 11.5   62
12. 5     5   5    NA      NA 14.3   56
13. 6     5   6    28      NA 14.9   66

3. Cast the molten airquality data set .

CODE :

names(airquality) <- tolower(names(airquality))

aqm <- melt(airquality, id=c("month", "day"), na.rm=TRUE)

acast(aqm, day ~ month ~ variable)

acast(aqm, month ~ variable, mean)

acast(aqm, month ~ variable, mean, margins = TRUE)

dcast(aqm, month ~ variable, mean, margins = c("month", "variable"))

OUTPUT:

Error in melt(airquality, id = c("month", "day"), na.rm = TRUE) :

  could not find function "melt"

Execution halted

[Execution complete with exit code 1]

4. Use cast function appropriately and compute the average of Ozone, Solar.R , Wind

and temperature per month ?

CODE :

DataSet <- airquality

str(DataSet)

colnames(DataSet)

Na <- summary([is.na](http://is.na/)(DataSet))

print(Na)

OUTPUT:

'data.frame':  153 obs. of  6 variables:

$ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...

$ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...

$ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...

$ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...

$ Month  : int  5 5 5 5 5 5 5 5 5 5 ...

$ Day    : int  1 2 3 4 5 6 7 8 9 10 ...

[1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"

   Ozone          Solar.R           Wind            Temp

 Mode :logical   Mode :logical   Mode :logical   Mode :logical

 FALSE:116       FALSE:146       FALSE:153       FALSE:153

 TRUE :37        TRUE :7

   Month            Day

 Mode :logical   Mode :logical

 FALSE:153       FALSE:153

[Execution complete with exit code 0]

8 FILE MANUPULATION IN R

Exercise

1. Consider the following data present. Create this file using windows notepad . Save the

file as input.csv using the save As All files(\*.\*) option in notepad.

CODE :

id,name,salary,start\_date,dept

1,Rick,623.3,2012-01-01,IT

2,Dan,515.2,2013-09-23,Operations

3,Tusar,611,2014-11-15,IT

4,Ryan,729,2014-05-11,HR

5,Gary,843.25,2015-03-27,Finance

6,Rasmi,578,2013-05-21,IT

7,Pranab,632.8,2013-07-30,Operations

8,Guru,722.5,2014-06-17,Finance

CREATE IT IN NOTEPAD AND SAVE IT AS “.CSV”

OUTPUT:

id    name  salary  start\_date        dept

0   1    Rick  623.30  2012-01-01          IT

1   2     Dan  515.20  2013-09-23  Operations

2   3   Tusar  611.00  2014-11-15          IT

3   4    Ryan  729.00  2014-05-11          HR

4   5    Gary  843.25  2015-03-27     Finance

5   6   Rasmi  578.00  2013-05-21          IT

6   7  Pranab  632.80  2013-07-30  Operations

7   8    Guru  722.50  2014-06-17     Finance

2. Use appropriate R commands to read input.csv file.

CODE :

library(readr)

input <- read\_csv("input.csv")

OUTPUT:

> library(readr)

> input <- read\_csv("input.csv")

Rows: 8 Columns: 5

── Column specification ────────────────────────────────────────────────────────────

Delimiter: ","

chr  (2): name, dept

dbl  (2): id, salary

date (1): start\_date

3. Analyze the CSV File and compute the following.

a. Get the maximum salary

b. Get the details of the person with max salary

c. Get all the people working in IT department

d. Get the persons in IT department whose salary is greater than 600

e. Get the people who joined on or after 2014

4. Get the people who joined on or after 2014 and write the output onto a file called