



ACADGILD

SESSION 2: Introduction to working with R

Assignment 2

Submitted By:Uday Kumar R

5. Problem Statement :

1. Read multiple json files into a working directory for further converting into a dataset.
I have files text1, text2, text3 in the directory json.

Ans) The sample json files text1.json, text2.json and text3.json are present in the folder “json” E:\work space\JSON

Jsonlite packages is installed and then following commands are executed using R- studio:

```
setwd("E:/work space/JSON")
getwd()
library(jsonlite)
path=("E:/work space/JSON")
temp=list.files(path,pattern = "*.json")
s=sapply(temp, function(x) fromJSON(x))
x=data.frame(s)
View(x)
class(x)
```

Name	Type	Value
s	list [3]	List of length 3
example_1.json	list [3]	List of length 3
fruit	character [1]	'Apple'
size	character [1]	'Large'
color	character [1]	'Red'
example_2.json	list [1]	List of length 1
quiz	list [2]	List of length 2
sport	list [1]	List of length 1
q1	list [3]	List of length 3
maths	list [2]	List of length 2
q1	list [3]	List of length 3
q2	list [3]	List of length 3
example_9.json	list [3]	List of length 3
fruits	character [1]	'Apple'
size	character [1]	'Large'
color	character [1]	'Red'

Hence multiple JSON files are read into working directory and converted into datasets

example_1.json.fruit	example_1.json.size	example_1.json.color	example_2.json.quiz.sport.q1.question	example_2.json.quiz.sport.q1.options	example_9.json.fruits
1 Apple	Large	Red	Which one is correct team name in NBA?	New York Bulls	Huston R
2 Apple	Large	Red	Which one is correct team name in NBA?	Los Angeles Kings	Huston R
4 Apple	Large	Red	Which one is correct team name in NBA?	Huston Rocket	Huston R
3 Apple	Large	Red	Which one is correct team name in NBA?	Golden State Warriros	Huston R

- Data Frame

```
> x
example_1.json.fruit example_1.json.size example_1.json.color example_2.json.quiz.sport.q1.question
1 Apple Large Red Which one is correct team name in NBA?
2 Apple Large Red Which one is correct team name in NBA?
3 Apple Large Red Which one is correct team name in NBA?
4 Apple Large Red Which one is correct team name in NBA?
example_2.json.quiz.sport.q1.options example_2.json.quiz.sport.q1.answer example_2.json.quiz.maths.q1.question
1 New York Bulls Huston Rocket 5 + 7 = ?
2 Los Angeles Kings Huston Rocket 5 + 7 = ?
3 Golden State Warriros Huston Rocket 5 + 7 = ?
4 Huston Rocket Huston Rocket 5 + 7 = ?
example_2.json.quiz.maths.q1.options example_2.json.quiz.maths.q1.answer example_2.json.quiz.maths.q2.question
1 10 12 12 - 8 = ?
2 11 12 12 - 8 = ?
3 12 12 12 - 8 = ?
4 13 12 12 - 8 = ?
example_2.json.quiz.maths.q2.options example_2.json.quiz.maths.q2.answer example_9.json.fruits example_9.json.size
1 1 4 Apple Large
2 2 4 Apple Large
3 3 4 Apple Large
4 4 4 Apple Large
example_9.json.color
1 Red
2 Red
3 Red
4 Red
```

2. Parse the following JSON into a data frame

```
js<-'{ "name": null, "release_date_local": null, "title": "3 (2011)", "opening_weekend_take": 1234, "year": 2011, "release_date_wide": "2011-09-16", "gross": 59954 }'
```

Ans)

```
) js<-'{ "name": null, "release_date_local": null, "title": "3 (2011)", "opening_weekend_take": 1234, "year": 2011, "release_date_wide": "2011-09-16", "gross": 59954 }'
```

to parse the above json into data frame the following commands are executed :

```
library(jsonlite)
```

```
z=fromJSON(js)
```

```
as.data.frame(z)
```

```
View(z)
```

```
> z
$name
NULL

$release_date_local
NULL

$title
[1] "3 (2011)"

$opening_weekend_take
[1] 1234

$year
[1] 2011

$release_date_wide
[1] "2011-09-16"

$gross
[1] 59954
```

z	list [7]	List of length 7
name	NULL	Pairlist of length 0
release_date_local	NULL	Pairlist of length 0
title	character [1]	'3 (2011)'
opening_weekend_take	integer [1]	1234
year	integer [1]	2011
release_date_wide	character [1]	'2011-09-16'
gross	integer [1]	59954

'2011-09-16'

3) Write a script for Variable Binning using R.

Binning is the process of transforming numerical variables into categorical counterparts.

Writing binning() function for dividing the variable named age into 4 bins named as "group1-(1 to 25)", "group2-(26 to 50)", "group3-(51 to 75)", "group4-(76 to 90)"

```
age <- c(1:90)
age
```

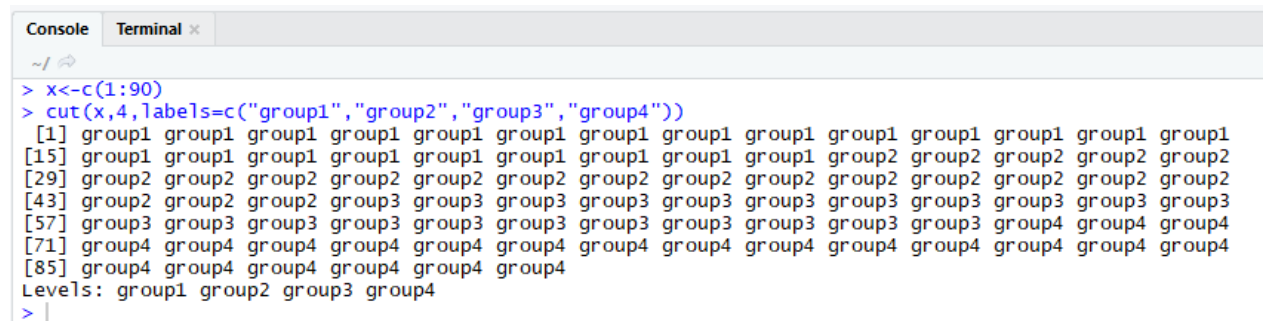
```
binning <- function(x)
{
  for(i in c(1:90))
  {
    ifelse(i <= 25, paste(i,"group1"),
           ifelse(i <= 50, paste(i,"group2"),
                  ifelse(i <= 75, paste(i,"group3"),
                         paste(i,"group4"))))
    break
  }
}
```

```
binning(age)
```

Example 1: Let us consider a vector consisting of values from 1 to 90 and we need to create 4 bins named "group1", "group2", "group3", "group4".

VARIABLE BINNING USING cut() function

```
x <- c(1:90)
cut(x, 4, labels = c("group1", "group2", "group3", "group4"))
```

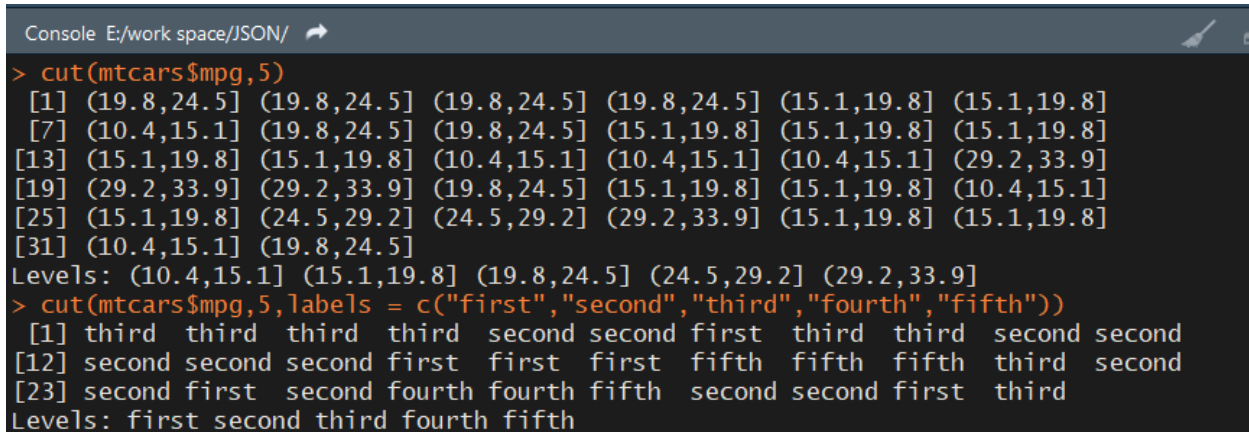


The screenshot shows an R console window with the following output:

```
> x <- c(1:90)
> cut(x, 4, labels = c("group1", "group2", "group3", "group4"))
 [1] group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group1
[15] group1 group1 group1 group1 group1 group1 group1 group1 group1 group1 group2 group2 group2 group2 group2
[29] group2 group2 group2 group2 group2 group2 group2 group2 group2 group2 group2 group2 group2 group2 group2
[43] group2 group2 group2 group2 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3
[57] group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3 group3
[71] group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4
[85] group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4 group4
Levels: group1 group2 group3 group4
>
```

Example 2: Import a mtcars.csv file into R-Studio and divide the variable named mpg into 5 bins named "FIRST", "SECOND", "THIRD", "FOURTH" and "FIFTH"

```
library(readr)
mtcars <- read_csv("E:/work space/JSON")
mpg <- mtcars$mpg
mpg
cut(mpg,5,labels=c("FIRST","SECOND","THIRD","FOURTH","FIFTH"))
```



```
Console E:/work space/JSON/
> cut(mtcars$mpg,5)
 [1] (19.8,24.5] (19.8,24.5] (19.8,24.5] (19.8,24.5] (15.1,19.8] (15.1,19.8]
 [7] (10.4,15.1] (19.8,24.5] (19.8,24.5] (15.1,19.8] (15.1,19.8] (15.1,19.8]
[13] (15.1,19.8] (15.1,19.8] (10.4,15.1] (10.4,15.1] (10.4,15.1] (29.2,33.9]
[19] (29.2,33.9] (29.2,33.9] (19.8,24.5] (15.1,19.8] (15.1,19.8] (10.4,15.1]
[25] (15.1,19.8] (24.5,29.2] (24.5,29.2] (29.2,33.9] (15.1,19.8] (15.1,19.8]
[31] (10.4,15.1] (19.8,24.5]
Levels: (10.4,15.1] (15.1,19.8] (19.8,24.5] (24.5,29.2] (29.2,33.9]
> cut(mtcars$mpg,5,labels = c("first","second","third","fourth","fifth"))
 [1] third  third  third  third  second second first  third  third  second second
[12] second second second first  first  first  fifth  fifth  fifth  third  second
[23] second first  second fourth fourth fifth  second second first  third
Levels: first second third fourth fifth
```

In the above example,the value of mtcars\$mpg ranges from 10.4 to 33.9.This range is divided into 5 bins with

Levels: (10.4,15.1] (15.1,19.8] (19.8,24.5] (24.5,29.2] (29.2,33.9]

Named as Levels: FIRST SECOND THIRD FOURTH FIFTH

Since the first value of mpg=21.0 lies in the THIRD bin ranging from 19.8 to 24.5; the output for the same is shown as (19.8,24.5] or THIRD.