# Problem Statement

1. Read multiple JSON files into a directory to convert into a dataset.

I have files text1, text2, text3 in the directory JSON.

|  |
| --- |
| > #For Json Files:  > install.packages("rjson")  Installing package into ‘C:/Users/raman/OneDrive/Documents/R/win-library/3.4’  (as ‘lib’ is unspecified)  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.4/rjson\_0.2.15.zip'  Content type 'application/zip' length 564436 bytes (551 KB)  downloaded 551 KB  package ‘rjson’ successfully unpacked and MD5 sums checked  Warning in install.packages :  cannot remove prior installation of package ‘rjson’  The downloaded binary packages are in  C:\Users\raman\AppData\Local\Temp\RtmpGeqjUb\downloaded\_packages  > library(rjson)  > filenames <- list.files("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON",  + pattern="\*.json",  + full.names=TRUE)  > filenames  [1] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text1.json"  [2] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text11.json"  [3] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text2.json"  [4] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text3.json"  > # this should give a character vector, with each file name represented by an entry  > myJSON <- lapply(filenames, function(x) fromJSON(file=x))  > myJSON  [[1]]  [[1]][[1]]  [[1]][[1]]$FIELD1  [1] "X"  [[1]][[1]]$FIELD2  [1] "Y"  [[1]][[1]]$FIELD3  [1] "Z"  [[1]][[2]]  [[1]][[2]]$FIELD1  [1] "1"  [[1]][[2]]$FIELD2  [1] "3"  [[1]][[2]]$FIELD3  [1] "5"  [[1]][[3]]  [[1]][[3]]$FIELD1  [1] "2"  [[1]][[3]]$FIELD2  [1] "4"  [[1]][[3]]$FIELD3  [1] "6"  [[1]][[4]]  [[1]][[4]]$FIELD1  [1] "3"  [[1]][[4]]$FIELD2  [1] "5"  [[1]][[4]]$FIELD3  [1] "7"  [[1]][[5]]  [[1]][[5]]$FIELD1  [1] "4"  [[1]][[5]]$FIELD2  [1] "6"  [[1]][[5]]$FIELD3  [1] "8"  [[1]][[6]]  [[1]][[6]]$FIELD1  [1] "5"  [[1]][[6]]$FIELD2  [1] "7"  [[1]][[6]]$FIELD3  [1] "9"  [[1]][[7]]  [[1]][[7]]$FIELD1  [1] "6"  [[1]][[7]]$FIELD2  [1] "8"  [[1]][[7]]$FIELD3  [1] "10"  [[2]]  [[2]][[1]]  [[2]][[1]]$FIELD1  [1] "X"  [[2]][[1]]$FIELD2  [1] "Y"  [[2]][[1]]$FIELD3  [1] "Z"  [[3]]  [[3]][[1]]  [[3]][[1]]$FIELD1  [1] "A"  [[3]][[1]]$FIELD2  [1] "B"  [[3]][[1]]$FIELD3  [1] "C"  [[3]][[2]]  [[3]][[2]]$FIELD1  [1] "11"  [[3]][[2]]$FIELD2  [1] "13"  [[3]][[2]]$FIELD3  [1] "15"  [[3]][[3]]  [[3]][[3]]$FIELD1  [1] "12"  [[3]][[3]]$FIELD2  [1] "14"  [[3]][[3]]$FIELD3  [1] "16"  [[3]][[4]]  [[3]][[4]]$FIELD1  [1] "13"  [[3]][[4]]$FIELD2  [1] "15"  [[3]][[4]]$FIELD3  [1] "17"  [[3]][[5]]  [[3]][[5]]$FIELD1  [1] "14"  [[3]][[5]]$FIELD2  [1] "16"  [[3]][[5]]$FIELD3  [1] "18"  [[3]][[6]]  [[3]][[6]]$FIELD1  [1] "15"  [[3]][[6]]$FIELD2  [1] "17"  [[3]][[6]]$FIELD3  [1] "19"  [[3]][[7]]  [[3]][[7]]$FIELD1  [1] "16"  [[3]][[7]]$FIELD2  [1] "18"  [[3]][[7]]$FIELD3  [1] "20"  [[4]]  [[4]][[1]]  [[4]][[1]]$FIELD1  [1] "X"  [[4]][[1]]$FIELD2  [1] "Y"  [[4]][[1]]$FIELD3  [1] "Z"  [[4]][[2]]  [[4]][[2]]$FIELD1  [1] "21"  [[4]][[2]]$FIELD2  [1] "27"  [[4]][[2]]$FIELD3  [1] "33"  [[4]][[3]]  [[4]][[3]]$FIELD1  [1] "22"  [[4]][[3]]$FIELD2  [1] "28"  [[4]][[3]]$FIELD3  [1] "34"  [[4]][[4]]  [[4]][[4]]$FIELD1  [1] "23"  [[4]][[4]]$FIELD2  [1] "29"  [[4]][[4]]$FIELD3  [1] "35"  [[4]][[5]]  [[4]][[5]]$FIELD1  [1] "24"  [[4]][[5]]$FIELD2  [1] "30"  [[4]][[5]]$FIELD3  [1] "36"  [[4]][[6]]  [[4]][[6]]$FIELD1  [1] "25"  [[4]][[6]]$FIELD2  [1] "31"  [[4]][[6]]$FIELD3  [1] "37"  [[4]][[7]]  [[4]][[7]]$FIELD1  [1] "26"  [[4]][[7]]$FIELD2  [1] "32"  [[4]][[7]]$FIELD3  [1] "38"  > # a list in which each element is one of your original JSON files  > temp <- list.files("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON", pattern="\*.json", full.names=TRUE)  > temp  [1] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text1.json"  [2] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text11.json"  [3] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text2.json"  [4] "D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\2 INTRODUCTION TO R\\JSON/text3.json"  > library(purrr)  > movies <- purrr::map\_df(temp, function(x) {  + purrr::map(jsonlite::fromJSON(x), function(y) ifelse(is.null(y), NA, y)) })  > movies  # A tibble: 4 x 3  FIELD1 FIELD2 FIELD3  *<chr>* *<chr>* *<chr>*  1 X Y Z  2 X Y Z  3 A B C  4 X Y Z |
|  |
| |  | | --- | | > | |

1. 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

}'

|  |
| --- |
| > ##2. Parse the following JSON into a data frame.  > library(jsonlite)  > library(purrr)  > js<-'{  + "name": null, "release\_date\_local": null, "title": "3 (2011)",  + "opening\_weekend\_take": 1234, "year": 2011,  + "release\_date\_wide": "2011-09-16", "gross": 59954  + }'  > js  [1] "{\n\"name\": null, \"release\_date\_local\": null, \"title\": \"3 (2011)\",\n\"opening\_weekend\_take\": 1234, \"year\": 2011,\n\"release\_date\_wide\": \"2011-09-16\", \"gross\": 59954\n}"  > movies <- purrr::map\_df(js, function(x) {  + purrr::map(jsonlite::fromJSON(x), function(y) ifelse(is.null(y), NA, y)) })  > movies  # A tibble: 1 x 7  name release\_date\_lo~ title opening\_weekend~ year release\_date\_wi~ gross  *<lgl>* *<lgl>* *<chr>* *<int>* *<int>* *<chr>* *<int>*  1 NA NA 3 (2~ 1234 2011 2011-09-16 59954 |
|  |
| |  | | --- | | > | |

1. Write a script for Variable Binning using R.

|  |
| --- |
| > #3. Write a script for Variable Binning using R.  > daf = data.frame(v=sample(1:90,1000,TRUE))  > daf$cat = cut(daf$v,c(-Inf,20,40,Inf))  > head(daf)  v cat  1 76 (40, Inf]  2 39 (20,40]  3 22 (20,40]  4 38 (20,40]  5 21 (20,40]  6 78 (40, Inf]  > table(daf$cat)  (-Inf,20] (20,40] (40, Inf]  208 215 577 |
|  |
| |  | | --- | | > | |