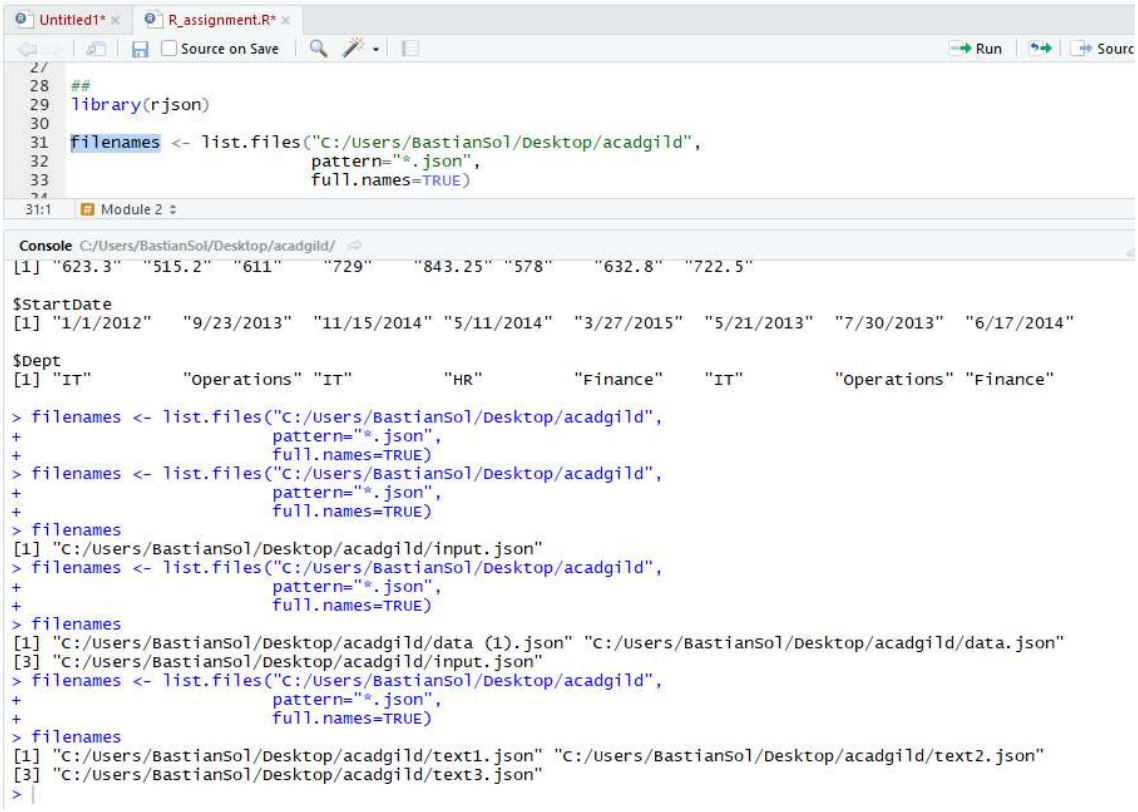
5. Problem Statement

1. Read multiple json files into a working directory for further converting into a dataset.

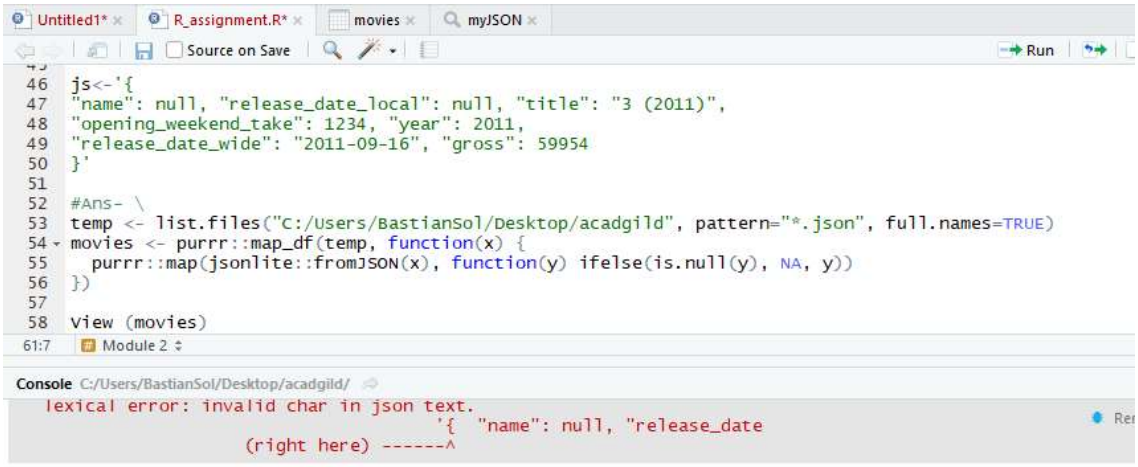
I have files text1, text2, text3 in the directory json.



```
Untitled1* | R_assignment.R*
Source on Save | Run | Sourc
27
28  ##
29  library(rjson)
30
31  filenames <- list.files("C:/Users/BastianSol/Desktop/acadgild",
32                          pattern="*.json",
33                          full.names=TRUE)
31:1  Module 2

Console C:/Users/BastianSol/Desktop/acadgild/
[1] "623.3"  "515.2"  "611"    "729"    "843.25" "578"    "632.8"  "722.5"

$StartDate
[1] "1/1/2012"   "9/23/2013"  "11/15/2014" "5/11/2014"  "3/27/2015"  "5/21/2013"  "7/30/2013"  "6/17/2014"

$Dept
[1] "IT"         "Operations" "IT"         "HR"         "Finance"    "IT"         "Operations" "Finance"

> filenames <- list.files("C:/Users/BastianSol/Desktop/acadgild",
+                         pattern="*.json",
+                         full.names=TRUE)
> filenames <- list.files("C:/Users/BastianSol/Desktop/acadgild",
+                         pattern="*.json",
+                         full.names=TRUE)
> filenames
[1] "C:/Users/BastianSol/Desktop/acadgild/input.json"
> filenames <- list.files("C:/Users/BastianSol/Desktop/acadgild",
+                         pattern="*.json",
+                         full.names=TRUE)
> filenames
[1] "C:/Users/BastianSol/Desktop/acadgild/data (1).json" "C:/Users/BastianSol/Desktop/acadgild/data.json"
[3] "C:/Users/BastianSol/Desktop/acadgild/input.json"
> filenames <- list.files("C:/Users/BastianSol/Desktop/acadgild",
+                         pattern="*.json",
+                         full.names=TRUE)
> filenames
[1] "C:/Users/BastianSol/Desktop/acadgild/text1.json" "C:/Users/BastianSol/Desktop/acadgild/text2.json"
[3] "C:/Users/BastianSol/Desktop/acadgild/text3.json"
>
```

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- \
> temp <- list.files("C:/Users/BastianSol/Desktop/acadgild", pattern="*.json", full.names=TRUE)
> temp
[1] "C:/Users/BastianSol/Desktop/acadgild/text1.json" "C:/Users/BastianSol/Desktop/acadgild/text2.json"
[3] "C:/Users/BastianSol/Desktop/acadgild/text3.json"
> movies <- purrr::map_df(temp, function(x) {
+   purrr::map(jsonlite::fromJSON(x), function(y) ifelse(is.null(y), NA, y))
+ })
> movies
# A tibble: 3 x 14
  ID    Name  Salary StartDate Dept      a b     name  release_date_lo~ title opening_weekend~  year
  <chr> <chr> <chr>  <chr>     <chr> <int> <chr> <lgl> <lgl>            <chr>             <int> <int>
1 1     Rick  623.3  1/1/2012  IT       NA NA    NA    NA               NA                   NA    NA
2 NA    NA    NA     NA        NA      123 4 5 6 NA    NA               NA                   NA    NA
3 NA    NA    NA     NA        NA       NA NA    NA    NA               3 (2~              1234  2011
# ... with 2 more variables: release_date_wide <chr>, gross <int>
> View (movies)
```

3. Write a script for variable binning using R.

```r
#Q6- Write a script for variable binning using R
# Ans-
df = data.frame(v=sample(1:60,1000,TRUE))
df$cat = cut(df$v,c(-Inf,6,12,Inf))
table(df$cat)
```

66:1 Module 2

Console C:/Users/BastianSol/Desktop/acadgild/

```
 [761] (12, Inf] (6,12]    (12, Inf] (-Inf,6]  (12, Inf] (
 [771] (6,12]    (6,12]    (12, Inf] (-Inf,6]  (-Inf,6]  (
 [781] (6,12]    (12, Inf] (12, Inf] (12, Inf] (-Inf,6]  (
 [791] (12, Inf] (12, Inf] (-Inf,6]  (12, Inf] (12, Inf] (
 [801] (12, Inf] (6,12]    (12, Inf] (6,12]    (12, Inf] (
 [811] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (6,12]    (
 [821] (-Inf,6]  (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [831] (12, Inf] (-Inf,6]  (12, Inf] (12, Inf] (12, Inf] (
 [841] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [851] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [861] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [871] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (6,12]    (
 [881] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [891] (-Inf,6]  (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [901] (12, Inf] (-Inf,6]  (12, Inf] (12, Inf] (12, Inf] (
 [911] (12, Inf] (12, Inf] (6,12]    (12, Inf] (-Inf,6]  (
 [921] (12, Inf] (12, Inf] (6,12]    (12, Inf] (12, Inf] (
 [931] (12, Inf] (12, Inf] (12, Inf] (6,12]    (12, Inf] (
 [941] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [951] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (-Inf,6]  (
 [961] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [971] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [981] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (
 [991] (12, Inf] (12, Inf] (12, Inf] (12, Inf] (-Inf,6]  (
Levels: (-Inf,6] (6,12] (12, Inf]
> table(df$cat)

 (-Inf,6]    (6,12] (12, Inf]
       99       102       799
```