

Assignment_02_KummarikuntaVidyasagar.R

12702

2020-09-12

```
# Assignment: ASSIGNMENT 2
# Name: Kummarikunta, Vidyasagar
# Date: 2020-09-11
```

```
## Check your current working directory using `getwd()`

getwd()
```

```
## [1] "C:/Users/12702/Desktop/MODatascience/DSC-520/week_2"
```

```
## List the contents of the working directory with the `dir()` function
dir()
```

```
## [1] "~$ch support .docx"
## [2] "Assignment_02_KummarikuntaVidyasagar.R"
## [3] "Assignment_02_KummarikuntaVidyasagar.spin.R"
## [4] "Assignment_02_KummarikuntaVidyasagar.spin.Rmd"
## [5] "Exercise-3.pdf"
## [6] "Exercise 3.Rmd"
## [7] "scores.csv"
## [8] "Tech support .docx"
## [9] "Week-2.2.html"
## [10] "Week-2.2.pdf"
## [11] "Week2.3.R"
```

```
## If the current directory does not contain the `data` directory, set the
## working directory to project root folder (the folder should contain the `data` directory
## Use `setwd()` if needed
setwd("/Users/12702/Desktop/MODatascience/DSC-520")
```

```
## Load the file `data/tidynomicon/person.csv` to `person_df1` using `read.csv`
## Examine the structure of `person_df1` using `str()`
person_df1 <- read.csv("data/tidynomicon/person.csv", header = TRUE)

head(person_df1)
```

```
##   person_id personal_name family_name
## 1      dyer      William      Dyer
## 2        pb        Frank    Pabodie
## 3       lake     Anderson      Lake
## 4        roe    Valentina    Roerich
## 5  danforth        Frank  Danforth
```

```
## R interpreted names as factors, which is not the behavior we want
## Load the same file to person_df2 using `read.csv` and setting `stringsAsFactors` to `FALSE`
## Examine the structure of `person_df2` using `str()`
person_df2 <- read.csv("data/tidynomicon/person.csv",
                      stringsAsFactors = FALSE, header = TRUE)

head(person_df2)
```

```
##   person_id personal_name family_name
## 1      dyer      William      Dyer
## 2       pb       Frank    Pabodie
## 3      lake    Anderson      Lake
## 4       roe    Valentina    Roerich
## 5 danforth      Frank    Danforth
```

```
## Read the file `data/scores.csv` to `scores_df`

scores_df <- read.csv("data/scores.csv" , stringsAsFactors = FALSE, header = TRUE)

head(scores_df)
```

```
##   Count Score Section
## 1    10   200  Sports
## 2    10   205  Sports
## 3    20   235  Sports
## 4    10   240  Sports
## 5    10   250  Sports
## 6    10   265 Regular
```

```
## Display summary statistics using the `summary()` function

summary(scores_df)
```

```
##      Count      Score      Section
## Min.   :10.00  Min.   :200.0  Length:38
## 1st Qu.:10.00  1st Qu.:300.0  Class :character
## Median :10.00  Median :322.5  Mode  :character
## Mean   :14.47  Mean   :317.5
## 3rd Qu.:20.00  3rd Qu.:357.5
## Max.   :30.00  Max.   :395.0
```

```
## Load the `readxl` library
library(readxl)

## Using the excel_sheets() function from the `readxl` package,
## List the worksheets from the file `data/G04ResultsDetail2004-11-02.xls`

data_result <- excel_sheets("data/G04ResultsDetail2004-11-02.xls")

head(data_result)
```

```
## [1] "Instructions" "Voter Turnout" "President" "House of Rep"
## [5] "Co Clerk" "Co Reg Deeds"
```

```
## Using the `read_excel` function, read the Voter Turnout sheet
## from the `data/G04ResultsDetail2004-11-02.xls`
## Assign the data to the `voter_turnout_df1`
## The header is in the second row, so make sure to skip the first row
## Examine the structure of `voter_turnout_df1` using `str()``
```

```
voter_turnout_df1 <- read_excel("data/G04ResultsDetail2004-11-02.xls", 'Voter Turnout', skip = 1)

head(voter_turnout_df1)
```

```
## # A tibble: 6 x 4
##   `Ward Precinct` `Ballots Cast` `Registered Voters` `Voter Turnout`
##   <chr>          <dbl>          <dbl>          <dbl>
## 1 01-01          421            678            0.621
## 2 01-02          443            691            0.641
## 3 01-03          705           1148            0.614
## 4 01-04          827           1308            0.632
## 5 01-05          527            978            0.539
## 6 01-06          323            574            0.563
```

```
## Using the `read_excel()` function, read the Voter Turnout sheet
## from `data/G04ResultsDetail2004-11-02.xls`
## Skip the first two rows and manually assign the columns using `col_names`
## Use the names "ward_precint", "ballots_cast", "registered_voters", "voter_turnout"
## Assign the data to the `voter_turnout_df2`
## Examine the structure of `voter_turnout_df2` using `str()``
```

```
voter_turnout_df2 <- read_excel("data/G04ResultsDetail2004-11-02.xls",
                               sheet = "Voter Turnout",
                               skip = 2,
                               col_names = c("ward_precint", "ballots_Cast", "registered_voters", "voter_
turnout"))
```

```
str(voter_turnout_df2)
```

```
## tibble [342 x 4] (S3: tbl_df/tbl/data.frame)
## $ ward_precint : chr [1:342] "01-01" "01-02" "01-03" "01-04" ...
## $ ballots_Cast : num [1:342] 421 443 705 827 527 323 358 410 440 500 ...
## $ registered_voters: num [1:342] 678 691 1148 1308 978 ...
## $ voter_turnout : num [1:342] 0.621 0.641 0.614 0.632 0.539 ...
```

```
#___

## Load the `DBI` library

library(DBI)
library(RSQLite)

## Create a database connection to `data/tidynomicon/example.db` using the dbConnect() function
## The first argument is the database driver which in this case is `RSQLite::SQLite()`
## The second argument is the path to the database file
## Assign the connection to `db` variable

db <- dbConnect(RSQLite::SQLite(), "data/tidynomicon/example.db")

## Query the Person table using the `dbGetQuery` function and the
## `SELECT * FROM PERSON;` SQL statement
## Assign the result to the `person_df` variable
## Use `head()` to look at the first few rows of the `person_df` dataframe

person_df <- dbGetQuery(db, "SELECT * FROM PERSON")

head(person_df)
```

```
##   person_id personal_name family_name
## 1      dyer      William      Dyer
## 2       pb       Frank    Pabodie
## 3      lake    Anderson      Lake
## 4       roe    Valentina    Roerich
## 5 danforth      Frank    Danforth
```

```
## List the tables using the `dbListTables()` function
## Assign the result to the `table_names` variable

table_names <- dbListTables(db)

head(table_names)
```

```
## [1] "Measurements" "Person"      "Site"      "Visited"
```

```
## Read all of the tables at once using the `lapply` function and assign the result to the `tables`
## variable
## Use `table_names`, `dbReadTable`, and `conn = db` as arguments
## Print out the tables

tables <- lapply(table_names, dbReadTable, conn = db)
```

```
## Warning in result_fetch(res@ptr, n = n): Column `reading`: mixed type, first
## seen values of type real, coercing other values of type string
```

```
print(tables)
```

```
## [[1]]
##   visit_id person_id quantity reading
## 1      619      dyer      rad    9.82
## 2      619      dyer      sal    0.13
## 3      622      dyer      rad    7.80
## 4      622      dyer      sal    0.09
## 5      734       pb      rad    8.41
## 6      734      lake      sal    0.05
## 7      734       pb      temp  -21.50
## 8      735       pb      rad    7.22
## 9      735      <NA>      sal    0.06
## 10     735      <NA>      temp  -26.00
## 11     751       pb      rad    4.35
## 12     751       pb      temp  -18.50
## 13     751      lake      sal    0.00
## 14     752      lake      rad    2.19
## 15     752      lake      sal    0.09
## 16     752      lake      temp  -16.00
## 17     752       roe      sal   41.60
## 18     837      lake      rad    1.46
## 19     837      lake      sal    0.21
## 20     837       roe      sal   22.50
## 21     844       roe      rad   11.25
##
## [[2]]
##   person_id personal_name family_name
## 1      dyer      William      Dyer
## 2       pb       Frank      Pabodie
## 3      lake      Anderson      Lake
## 4       roe      Valentina      Roerich
## 5  danforth       Frank      Danforth
##
## [[3]]
##   site_id latitude longitude
## 1    DR-1   -49.85   -128.57
## 2    DR-3   -47.15   -126.72
## 3   MSK-4   -48.87   -123.40
##
## [[4]]
##   visit_id site_id visit_date
## 1      619    DR-1 1927-02-08
## 2      622    DR-1 1927-02-10
## 3      734    DR-3 1930-01-07
## 4      735    DR-3 1930-01-12
## 5      751    DR-3 1930-02-26
## 6      752    DR-3      <NA>
## 7      837   MSK-4 1932-01-14
## 8      844    DR-1 1932-03-22
```

```
## Use the `dbDisconnect` function to disconnect from the database

dbDisconnect(db)

## Import the `jsonlite` library
library(jsonlite)

## Convert the scores_df dataframe to JSON using the `toJSON()` function

datajson <- toJSON(scores_df)

head(datajson)
```

```
## [1] "[{"Count\\":10,\"Score\\":200,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":205,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":235,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":240,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":250,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":265,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":275,\"Section\\\":\"Regular\\\"},{\"Count\\":30,\"Score\\":285,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":295,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":300,\"Section\\\":\"Regular\\\"},{\"Count\\":20,\"Score\\":300,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":305,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":305,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":310,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":310,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":320,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":305,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":315,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":320,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":325,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":325,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":330,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":330,\"Section\\\":\"Sports\\\"},{\"Count\\":30,\"Score\\":335,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":335,\"Section\\\":\"Regular\\\"},{\"Count\\":20,\"Score\\":340,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":340,\"Section\\\":\"Sports\\\"},{\"Count\\":30,\"Score\\":350,\"Section\\\":\"Regular\\\"},{\"Count\\":20,\"Score\\":360,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":360,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":365,\"Section\\\":\"Regular\\\"},{\"Count\\":20,\"Score\\":365,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":370,\"Section\\\":\"Sports\\\"},{\"Count\\":10,\"Score\\":370,\"Section\\\":\"Regular\\\"},{\"Count\\":20,\"Score\\":375,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":375,\"Section\\\":\"Sports\\\"},{\"Count\\":20,\"Score\\":380,\"Section\\\":\"Regular\\\"},{\"Count\\":10,\"Score\\":395,\"Section\\\":\"Sports\\\"}]"
```

```
## Convert the scores dataframe to JSON using the `toJSON()` function with the `pretty=TRUE` option
json_pretty <- toJSON(scores_df, pretty = TRUE)

head(json_pretty)
```

```
## [1] "[\n {\n  \"Count\": 10,\n  \"Score\": 200,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 205,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 235,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 240,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 250,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 265,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 275,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 30,\n  \"Score\": 285,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 295,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 300,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 300,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 305,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 305,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 310,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 310,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 320,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 305,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 315,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 320,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 325,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 325,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 330,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 330,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 30,\n  \"Score\": 335,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 335,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 340,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 340,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 30,\n  \"Score\": 350,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 360,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 360,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 365,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 365,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 370,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 370,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 375,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 375,\n  \"Section\": \"Sports\"\n },\n {\n  \"Count\": 20,\n  \"Score\": 380,\n  \"Section\": \"Regular\"\n },\n {\n  \"Count\": 10,\n  \"Score\": 395,\n  \"Section\": \"Sports\"\n }]\n"]
```