**> # Assignment: ASSIGNMENT 2**

**> # Name: Thangaraj, Arunkumar**

**> # Date: 2020-12-10**

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

getwd()

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

dir()

## 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("C:/Users/arun8/R/dsc520\_fork")

## Load the file `data/tidynomicon/person.csv` to `person\_df1` using `read.csv`

## Examine the structure of `person\_df1` using `str()`

## In R 4.0.0. default.stringsAsFactors() is False. So, creating person\_df11 to consider string as factors

person\_df1 <- read.csv(file = 'data/tidynomicon/person.csv')

str(person\_df1)

person\_df11 <- read.csv(file = 'data/tidynomicon/person.csv', stringsAsFactors=TRUE)

str(person\_df11)

## 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(file = 'data/tidynomicon/person.csv', stringsAsFactors=FALSE )

str(person\_df2)

## Read the file `data/scores.csv` to `scores\_df`

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

scores\_df <- read.csv(file = 'data/scores.csv', stringsAsFactors=TRUE)

summary(scores\_df)

## 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`

excel\_sheets('data/G04ResultsDetail2004-11-02.xls')

## 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', sheet='Voter Turnout', skip = 1)

str(voter\_turnout\_df1)

## 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)

## Load the `DBI` library

library(DBI)

## 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

library(RSQLite)

drv <- dbDriver('SQLite')

class(drv)

db <- dbConnect(drv, 'data/tidynomicon/example.db')

## 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", stringsAsFactors=FALSE)

head(person\_df)

## List the tables using the `dbListTables()` function

## Assign the result to the `table\_names` variable

table\_names <- dbListTables(db)

## 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)

tables

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

dbDisconnect(db)

db

## Import the `jsonlite` library

library(jsonlite)

## Convert the scores\_df dataframe to JSON using the `toJSON()` function

toJSON(scores\_df)

## Convert the scores dataframe to JSON using the `toJSON()` function with the `pretty=TRUE` option

toJSON(scores\_df, pretty=TRUE)

**OUTPUT:**

[1] "C:/Users/arun8/R/dsc520\_fork"

[1] "assignments" "completed" "data"

[4] "dsc520\_fork.Rproj" "LICENSE" "README.md"

[7] "RMarkdown.md"

'data.frame': 5 obs. of 3 variables:

$ person\_id : chr "dyer" "pb" "lake" "roe" ...

$ personal\_name: chr "William" "Frank" "Anderson" "Valentina" ...

$ family\_name : chr "Dyer" "Pabodie" "Lake" "Roerich" ...

'data.frame': 5 obs. of 3 variables:

$ person\_id : Factor w/ 5 levels "danforth","dyer",..: 2 4 3 5 1

$ personal\_name: Factor w/ 4 levels "Anderson","Frank",..: 4 2 1 3 2

$ family\_name : Factor w/ 5 levels "Danforth","Dyer",..: 2 4 3 5 1

'data.frame': 5 obs. of 3 variables:

$ person\_id : chr "dyer" "pb" "lake" "roe" ...

$ personal\_name: chr "William" "Frank" "Anderson" "Valentina" ...

$ family\_name : chr "Dyer" "Pabodie" "Lake" "Roerich" ...

Count Score Section

Min. :10.00 Min. :200.0 Regular:19

1st Qu.:10.00 1st Qu.:300.0 Sports :19

Median :10.00 Median :322.5

Mean :14.47 Mean :317.5

3rd Qu.:20.00 3rd Qu.:357.5

Max. :30.00 Max. :395.0

[1] "Instructions" "Voter Turnout" "President"

[4] "House of Rep" "Co Clerk" "Co Reg Deeds"

[7] "Co Public Defender" "Co Comm 1" "Co Comm 3"

[10] "Co Comm 5" "Co Comm 7" "St Bd of Ed 2"

[13] "St Bd of Ed 4" "Legislature 5" "Legislature 7"

[16] "Legislature 9" "Legislature 11" "Legislature 13"

[19] "Legislature 23" "Legislature 31" "Legislature 39"

[22] "MCC 1" "MCC 2" "MCC 3"

[25] "MCC 4" "OPPD" "MUD"

[28] "NRD 3" "NRD 5" "NRD 7"

[31] "NRD 9" "OPS 2" "OPS 4"

[34] "OPS 6" "OPS 8" "OPS 10"

[37] "OPS 11" "OPS 12" "ESU 2"

[40] "ESU 3" "Arlington Sch 24" "Bennington Sch 59"

[43] "Elkhorn Sch 10" "Fremont Sch 1" "Ft Calhoun Sch 3"

[46] "Gretna Sch 37" "Millard Sch 17" "Ralston Sch 54"

[49] "Valley Sch 33" "Waterloo Sch 11" "Bennington Mayor"

[52] "Elkhorn Mayor" "Valley Mayor" "Ralston Mayor"

[55] "Ralston Library Bd" "Bennington City Cnc 1" "Bennington City Cnc 2"

[58] "Elkhorn City Cnc A" "Elkhorn City Cnc B" "Elkhorn City Cnc C"

[61] "Ralston City Cnc 1" "Ralston City Cnc 2" "Ralston City Cnc 6"

[64] "Waterloo Bd Trustees" "Valley City Cnc" "Amendment 1"

[67] "Amendment 2" "Amendment 3" "Amendment 4"

[70] "Initiative 417" "Initiative 418" "Initiative 419"

[73] "Initiative 420"

tibble [342 x 4] (S3: tbl\_df/tbl/data.frame)

$ Ward Precinct : 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 ...

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 ...

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

[[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

[{"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"}]

[

{

"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"

}

]

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> qplot(Count,Score,data=dfs, main = "Sports Section")

> qplot(Count,Score,data=dfs, labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

+ )

Error: `mapping` must be created by `aes()`

Run `rlang::last\_error()` to see where the error occurred.

> qplot(Count,Score,data=dfs, main = "Sports Section")

> qplot(Count,Score,data=dfs, main = "Sports Section",xlab = "Student Count", ylab = "Total Score",)

> qplot(Count,Score,data=dfr, main = "Sports Section",xlab = "Student Count", ylab = "Total Score",)

> qplot(Count,Score,data=dfr, main = "Regular Section",xlab = "Student Count", ylab = "Total Score",)

> qplot(Count,Score,data=dfs, main = "Sports Section",xlab = "Student Count", ylab = "Total Score",)

> qplot(Count,Score,data=dfs, main = "Sports Section",xlab = "Student Count", ylab = "Total Score")

> qplot(Count,Score,data=dfr, main = "Regular Section",xlab = "Student Count", ylab = "Total Score")

> qplot(Count,Score,data=dfs, main = "Sports Section",xlab = "Student Count", ylab = "Total Score")

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

>

>

>

>

>

>

>

>

>

>

>

>

>

>

>

>

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

>

>

>

>

>

>

>

>

>

>

> qplot(Count,Score,data=dfs, main = "Sports Section", xlab = "Student Count", ylab = "Total Score")

> qplot(Count,Score1,data=dfs, main = "Sports Section", xlab = "Student Count", ylab = "Total Score")

Error in FUN(X[[i]], ...) : object 'Score1' not found

> qplot(Count,Score,data=dfs, main = "Sports Section", xlab = "Student Count", ylab = "Total Score")

> qplot(Count,Score,data=dfr, main = "Regular Section", xlab = "Student Count", ylab = "Total Score")

>

>

>

>

>

>

>

>

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

>

>

>

>

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

>

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity")

> ggplot(data=dfr, aes(x=Score, y=Count)) +

+ geom\_bar(stat="identity")

> ggplot(data=dfr, aes(x=Score, y=Count))

> ggplot(data=dfr, aes(x=Score, y=Count)) +

+ geom\_bar(stat="identity")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity") + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity") + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity") + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity") + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count,)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

>

>

>

>

> ggplot(data=dfs, aes(x=Score, y=Count,fill=Count)) + geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Sports Section Scores")

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> ggplot(data=dfr, aes(x=Count, y=Score, fill=Count)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(0,40) + ylim(175,400) +labs(x="Student Count",y="Student Score",title = "Regular Section Scores")

Warning message:

Removed 19 rows containing missing values (geom\_bar).

> ggplot(data=dfr, aes(x=Count, y=Score)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(0,40) + ylim(175,400) +labs(x="Student Count",y="Student Score",title = "Regular Section Scores")

Warning message:

Removed 19 rows containing missing values (geom\_bar).

> ggplot(data=dfr, aes(x=Score, y=Count,fill=Count)) +

+ geom\_bar(stat="identity",position = position\_dodge()) + xlim(175,400) + ylim(0,40) +labs(x="Student Score",y="Student Count",title = "Regular Section Scores")

> dfs

Count Score Section

1 10 200 Sports

2 10 205 Sports

3 20 235 Sports

4 10 240 Sports

5 10 250 Sports

8 30 285 Sports

11 20 300 Sports

12 10 305 Sports

15 10 310 Sports

18 10 315 Sports

21 10 325 Sports

23 10 330 Sports

24 30 335 Sports

27 10 340 Sports

30 10 360 Sports

32 20 365 Sports

33 10 370 Sports

36 10 375 Sports

38 10 395 Sports

> dfr

Count Score Section

6 10 265 Regular

7 10 275 Regular

9 10 295 Regular

10 10 300 Regular

13 10 305 Regular

14 10 310 Regular

16 20 320 Regular

17 10 305 Regular

19 20 320 Regular

20 10 325 Regular

22 20 330 Regular

25 10 335 Regular

26 20 340 Regular

28 30 350 Regular

29 20 360 Regular

31 20 365 Regular

34 10 370 Regular

35 20 375 Regular

37 20 380 Regular

> # Assignment: ASSIGNMENT 2

> # Name: Thangaraj, Arunkumar

> # Date: 2020-12-10

>

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

> getwd()

[1] "C:/Users/arun8/R/dsc520\_fork"

>

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

> dir()

[1] "assignments" "completed" "data" "dsc520\_fork.Rproj" "LICENSE"

[6] "README.md" "RMarkdown.md"

>

> ## 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("C:/Users/arun8/R/dsc520\_fork")

>

> ## Load the file `data/tidynomicon/person.csv` to `person\_df1` using `read.csv`

> ## Examine the structure of `person\_df1` using `str()`

> ## In R 4.0.0. default.stringsAsFactors() is False. So, creating person\_df11 to consider string as factors

> person\_df1 <- read.csv(file = 'data/tidynomicon/person.csv')

> str(person\_df1)

'data.frame': 5 obs. of 3 variables:

$ person\_id : chr "dyer" "pb" "lake" "roe" ...

$ personal\_name: chr "William" "Frank" "Anderson" "Valentina" ...

$ family\_name : chr "Dyer" "Pabodie" "Lake" "Roerich" ...

>

> person\_df11 <- read.csv(file = 'data/tidynomicon/person.csv', stringsAsFactors=TRUE)

> str(person\_df11)

'data.frame': 5 obs. of 3 variables:

$ person\_id : Factor w/ 5 levels "danforth","dyer",..: 2 4 3 5 1

$ personal\_name: Factor w/ 4 levels "Anderson","Frank",..: 4 2 1 3 2

$ family\_name : Factor w/ 5 levels "Danforth","Dyer",..: 2 4 3 5 1

>

> ## 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(file = 'data/tidynomicon/person.csv', stringsAsFactors=FALSE )

> str(person\_df2)

'data.frame': 5 obs. of 3 variables:

$ person\_id : chr "dyer" "pb" "lake" "roe" ...

$ personal\_name: chr "William" "Frank" "Anderson" "Valentina" ...

$ family\_name : chr "Dyer" "Pabodie" "Lake" "Roerich" ...

>

> ## Read the file `data/scores.csv` to `scores\_df`

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

> scores\_df <- read.csv(file = 'data/scores.csv', stringsAsFactors=TRUE)

> summary(scores\_df)

Count Score Section

Min. :10.00 Min. :200.0 Regular:19

1st Qu.:10.00 1st Qu.:300.0 Sports :19

Median :10.00 Median :322.5

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`

> excel\_sheets('data/G04ResultsDetail2004-11-02.xls')

[1] "Instructions" "Voter Turnout" "President" "House of Rep"

[5] "Co Clerk" "Co Reg Deeds" "Co Public Defender" "Co Comm 1"

[9] "Co Comm 3" "Co Comm 5" "Co Comm 7" "St Bd of Ed 2"

[13] "St Bd of Ed 4" "Legislature 5" "Legislature 7" "Legislature 9"

[17] "Legislature 11" "Legislature 13" "Legislature 23" "Legislature 31"

[21] "Legislature 39" "MCC 1" "MCC 2" "MCC 3"

[25] "MCC 4" "OPPD" "MUD" "NRD 3"

[29] "NRD 5" "NRD 7" "NRD 9" "OPS 2"

[33] "OPS 4" "OPS 6" "OPS 8" "OPS 10"

[37] "OPS 11" "OPS 12" "ESU 2" "ESU 3"

[41] "Arlington Sch 24" "Bennington Sch 59" "Elkhorn Sch 10" "Fremont Sch 1"

[45] "Ft Calhoun Sch 3" "Gretna Sch 37" "Millard Sch 17" "Ralston Sch 54"

[49] "Valley Sch 33" "Waterloo Sch 11" "Bennington Mayor" "Elkhorn Mayor"

[53] "Valley Mayor" "Ralston Mayor" "Ralston Library Bd" "Bennington City Cnc 1"

[57] "Bennington City Cnc 2" "Elkhorn City Cnc A" "Elkhorn City Cnc B" "Elkhorn City Cnc C"

[61] "Ralston City Cnc 1" "Ralston City Cnc 2" "Ralston City Cnc 6" "Waterloo Bd Trustees"

[65] "Valley City Cnc" "Amendment 1" "Amendment 2" "Amendment 3"

[69] "Amendment 4" "Initiative 417" "Initiative 418" "Initiative 419"

[73] "Initiative 420"

>

> ## 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', sheet='Voter Turnout', skip = 1)

> str(voter\_turnout\_df1)

tibble [342 x 4] (S3: tbl\_df/tbl/data.frame)

$ Ward Precinct : 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 ...

>

> ## 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)

>

> ## 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

> library(RSQLite)

> drv <- dbDriver('SQLite')

> class(drv)

[1] "SQLiteDriver"

attr(,"package")

[1] "RSQLite"

> db <- dbConnect(drv, 'data/tidynomicon/example.db')

> ## 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", stringsAsFactors=FALSE)

> 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)

>

> ## 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 message:

In result\_fetch(res@ptr, n = n) :

Column `reading`: mixed type, first seen values of type real, coercing other values of type string

> 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)

> db

<SQLiteConnection>

DISCONNECTED

>

> ## Import the `jsonlite` library

> library(jsonlite)

>

> ## Convert the scores\_df dataframe to JSON using the `toJSON()` function

> toJSON(scores\_df)

[{"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

> toJSON(scores\_df, pretty=TRUE)

[

{

"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"

}

]

>