

# STATS 506 HW 2

Anuraag Ramesh

October 27, 2022

## Contents

Q1. . . . .	1
Q2. . . . .	1
Q3. . . . .	2
Q4. . . . .	2
Q5. . . . .	3
Q6. . . . .	3
Q7. . . . .	3
Q8. . . . .	3
Q9. . . . .	3
Q10. . . . .	4
Q11. . . . .	4
Q12. . . . .	4
Q13. . . . .	4

Q1.

Q2.

```
df_func <- function(n1, n2){  
  headers = read.table('Data/2020_Business_Academic_QCQ.txt', sep = ',', nrow = 1,  
                       encoding = 'UTF-8')  
  df = read.table('Data/2020_Business_Academic_QCQ.txt', sep = ',', skip = n1, nrow = n2 - n1, fileEncoding = 'UTF-8')  
  colnames(df) = headers  
  df = df %>% select('State', 'County Code',  
                    'Employee Size (5) - Location',  
                    'Sales Volume (9) - Location',  
                    'Census Tract')  
  colnames(df) = gsub('-', '', colnames(df))  
  colnames(df) = gsub '\\(\\d\\)', '', colnames(df))  
  colnames(df) = gsub(' ', '', colnames(df))  
  colnames(df) = tolower(colnames(df))  
}
```

```
df <- df[complete.cases(df), ]

return(df)
}
```

### Q3.

```
df = data.frame()
for (i in seq(1, 300001, by = 20000)){
  if(i == 1){
    n1 = i
  }
  else{
    n2 = i
    df1 = df_func(n1, n2)
    n1 = n2
    df = rbind(df, df1)
  }
}

df
```

```
agg = df %>% group_by(censustract) %>%
  summarise(employeesize = sum(employeesizelocation),
            salesvolume = sum(salesvolumelocation))
df1 = data.frame(agg)

df1$censustract = as.character(df1$censustract)
```

### Q4.

Table creation in SQL:

```
CREATE TABLE df1 ( censustract VARCHAR(255), employeesize INT, salesvolume INT )
```

```
host_name = 'localhost'
port_no = 3306
db_name = 'Hw3db'
u_id = 'root'

pwd = Sys.getenv('sqlpwd')

conn = RMySQL::dbConnect(RMySQL::MySQL(), host = host_name, port = port_no, user = u_id,
                        password = pwd, dbname = db_name)

RMySQL::dbWriteTable(conn, name = 'df1', df1, overwrite = TRUE, row.names = FALSE)
```

```
[1] TRUE
```

**Q5.**

```
res = dbGetQuery(conn, 'SELECT censustract, salesvolume FROM df1 ORDER BY salesvolume DESC LIMIT 10')
res
```

**Q6.**

```
git branch new
git checkout new
```

**Q7.**

```
df2 = read.csv('Data/AL.csv')
df2 = df2[c('FIELD19', 'FIELD20', 'FIELD22', 'FIELD45', 'FIELD64', 'FIELD65')]

colnames(df2) = c('householdwealth', 'income', 'home_value', 'state', 'countycode', 'censustract')
df2 = df2[df2$home_value != 0, ]
df2 = df2 %>% group_by(censustract) %>%
  summarise(wealth = mean(householdwealth),
            income = mean(income),
            homevalue = mean(home_value))
df2$censustract = as.character(df2$censustract)
```

**Q8.**

```
RMySQL::dbWriteTable(conn, name = 'df2', df2, overwrite = TRUE, row.names = FALSE)
```

```
[1] TRUE
```

**Q9.**

commit d398fbb8fa29c92fc9e878a62b0932129cd3c963 (HEAD -> new, origin/new) Author: Anuraag Ramesh anuraagr@umich.edu Date: Fri Nov 25 22:55:19 2022 -0500

Question 8 Completed

commit 210c0f5c9c2219a076f8e70392aa9eb245843c7e (origin/main, main) Author: Anuraag Ramesh anuraagr@umich.edu Date: Fri Nov 25 21:51:26 2022 -0500

Q5 Completed

commit e9353aadc7e50f2f6c1bf456fce013fd6f9321b4 (origin/master) Author: Anuraag Ramesh anuraagr@umich.edu Date: Fri Nov 25 19:23:03 2022 -0500

Q5 Completed

HEAD means the current branch that is being checked out and points out the last commit.

Q10.

```
Sys.getenv("CENSUS_API_KEY")
```

```
[1] "c9c803849168d3116b2ba3cc7407cce3b36bae56"
```

```
census <- get_decennial(geography = 'tract', variables = c('H006001', 'H006002', 'H006003', 'H006004',  
year = 2010, state = '01')
```

Getting data from the 2010 decennial Census

Using Census Summary File 1

```
census = census %>% spread(variable, value)
```

```
census$Whitepercent = census$H006002/census$H006001
```

```
census$tract = gsub("[^0-9]", "", census$NAME)
```

```
RMySQL::dbWriteTable(conn, name = 'census', census, overwrite = TRUE, row.names = FALSE)
```

```
[1] TRUE
```

Q11.

```
res = dbSendQuery(conn, 'SELECT d2.censustract, d2.wealth, d2.income, d2.homevalue, d1.employeesize, d1  
FROM df2 d2 JOIN df1 d1 ON d2.censustract = d1.censustract  
JOIN census c ON c.tract = d1.censustract')  
combine = dbFetch(res)
```

Q12.

Q13.