Homework08a

Steve Simon

This file was created on 2020-07-12 and last modified on 2020-07-12.

Note: this solution uses R and SQLite. An alternate solution using SAS and Oracle is also available.

- For your homework, use the titanic database.
  - This is available in Oracle using schema='simons'.
  - In SQLite, it is a standalone file named titanic_db.sqlite.
- Run SQL queries to answer the following
  - Count the number of passengers with the title "Mr" somewhere in their name.
  - Run a query that counts the number of male and female children (Age <= 18)
  - Run a query that identifies the ages of the youngest and oldest patients in each passenger class.
  - The Survived field has values of 0 (died) and 1 (alive) and an average of this value provides a probability of survival.
  - Calculate this probability for six categories representing the combination of passenger class and sex.
  - Include only those groups where the survival probability is greater than 30%.
- Place the SQL code and the results of all your queries in a single PDF file.

Note: Some of the names used in this code are arbitrary and you can choose whatever names you want. To emphasize which names can be modified at your discretion, I am using names of famous statisticians.

The statistician being honored in this code is George W. Snedecor.

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Loading required package: RSQLite
```

```
snedecor <- dbConnect(SQLite(),
  dbname="../data/titanic_db.sqlite")
george1 <- dbGetQuery(conn=snedecor,  "
  select count(*) as number_of_misters
    from titanic_table
    where name like '% Mr %'
")

george1
```

```
##   number_of_misters
## 1               751
```

```
dbDisconnect(conn=snedecor)
```

```
library(sqldf)
snedecor <- dbConnect(SQLite(),
  dbname="../data/titanic_db.sqlite")
```

```
george2 <- dbGetQuery(conn=snedecor,  "
  select sex, count(*) as number_of_children
    from titanic_table
    where age < 18
    group by sex
")

george2
```

```
##      Sex number_of_children
## 1 female                 50
## 2   male                 46
```

```
dbDisconnect(conn=snedecor)
```

```
library(sqldf)
snedecor <- dbConnect(SQLite(),
  dbname="../data/titanic_db.sqlite")
george3 <- dbGetQuery(conn=snedecor,  "
  select sex, count(*) as number_of_children
    from titanic_table
    where age < 18
    group by sex
")

george3
```

```
##      Sex number_of_children
## 1 female                 50
## 2   male                 46
```

```
dbDisconnect(conn=snedecor)
```

```
library(sqldf)
snedecor <- dbConnect(SQLite(),
  dbname="../data/titanic_db.sqlite")
george4 <- dbGetQuery(conn=snedecor,  "
  select
    sex,
    pclass,
    avg(survived) as survival_probability
    from titanic_table
    group by sex, pclass
    having survival_probability > 0.30
")

george4
```

```
##      Sex PClass survival_probability
## 1 female    1st            0.9370629
## 2 female    2nd            0.8785047
## 3 female    3rd            0.3773585
## 4   male    1st            0.3296089
```

```
dbDisconnect(conn=snedecor)
```