Homework07a

Steve Simon

This file was created on 2020-07-06 and last modified on 2021-02-22.

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

Your homework

- Use the cigarettes table in the melange database. The variable weight_g is recorded to four decimal places. Round all the values to two decimal places and display them.
- Use the hospital table in the ehr database. The variable ACUTE_NONACUTE has the values Acute and Non-Acute. Use the substr function to convert these two values to the first letter (A or N). Print out the first ten values only.
- Use the patient_type table in the ehr database. List all the values of PAT_TYPE_DESC after converting the data to lower case.

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 John Tukey.

```
library(sqldf)
```

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

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

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

```
tukey1 <- dbConnect(SQLite(),
  dbname="../data/melange.sqlite")

john1 <- dbGetQuery(conn=tukey1, "
    select
      round(weight_g, 2) as rounded_weight
    from cigarettes
")

john1
```

```
##    rounded_weight
## 1            0.99
## 2            1.09
## 3            1.17
```

```
## 4            0.93
## 5            0.95
## 6            0.89
## 7            1.03
## 8            0.92
## 9            0.94
## 10           0.89
## 11           0.96
## 12           0.93
## 13           0.97
## 14           1.12
## 15           0.85
## 16           0.79
## 17           0.92
## 18           1.04
## 19           0.96
## 20           0.91
## 21           1.01
## 22           0.98
## 23           0.97
## 24           0.95
## 25           1.12
```

```
dbDisconnect(conn=tukey1)
```

```
library(sqldf)

tukey2 <- dbConnect(SQLite(),
  dbname="../data/ehr.sqlite")

john2 <- dbGetQuery(conn=tukey2, "
    select
      substr(ACUTE_NONACUTE, 1, 1) as single_letter_code
    from hospital
    limit 10
")

john2
```

```
##    single_letter_code
## 1                   A
## 2                   N
## 3                   A
## 4                   A
## 5                   A
## 6                   N
## 7                   A
## 8                   A
## 9                   N
## 10                  N
```

```
dbDisconnect(conn=tukey2)
```

```
library(sqldf)

tukey3 <- dbConnect(SQLite(),
  dbname="../data/ehr.sqlite")

john3 <- dbGetQuery(conn=tukey3, "
    select
      lower(PAT_TYPE_DESC) as description_lower_case
    from patient_type
    limit 10
")

john3
```

```
##    description_lower_case
## 1       unknown / invalid
## 2               community
## 3               emergency
## 4              laboratory
## 5             non-patient
## 6               inpatient
## 7         other specialty
## 8              not mapped
## 9                  clinic
## 10              recurring
```

```
dbDisconnect(conn=tukey3)
```