Homework14a

Suman Sahil and Steve Simon

This file was created on 2020-07-24 and last modified on 2021-04-10.

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

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 Helen Walker.

1. Compute the intersection of the ids from acupuncture_baseline_results and acupuncture_one_year_results. Display the first ten rows of data only.

```
library(sqldf)
```

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

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

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

```
walker <- dbConnect(SQLite(),
  dbname="../data/melange.sqlite")
helen1 <- dbGetQuery(conn=walker,  "
    select id
      from acupuncture_baseline_results
    intersect
    select id
      from acupuncture_one_year_results
    limit 10
")

helen1
```

```
##     id
## 1  104
## 2  108
## 3  112
## 4  113
## 5  114
## 6  126
## 7  130
## 8  131
## 9  135
## 10 137
```

```
dbDisconnect(conn=walker)
```

2. Compute the union of the ids from acupuncture_baseline_results and acupuncture_one_year_results. Display the first ten rows of data only.

```
library(sqldf)
walker <- dbConnect(SQLite(),
  dbname="../data/melange.sqlite")
helen2 <- dbGetQuery(conn=walker,  "
    select id
      from acupuncture_baseline_results
    union
    select id
      from acupuncture_one_year_results
    order by id
    limit 10
")

helen2
```

```
##     id
## 1  100
## 2  101
## 3  104
## 4  105
## 5  108
## 6  112
## 7  113
## 8  114
## 9  126
## 10 130
```

```
dbDisconnect(conn=walker)
```

3. In a previous module, you were asked to list the first ten ids that were in acupuncture_baseline_resuts but not in acupuncture_one_year_results. Use the set operator "minus" to achieve the same goal. Note: for SQLite, use "except" instead of "minus".

```
library(sqldf)
walker <- dbConnect(SQLite(),
  dbname="../data/melange.sqlite")
helen3 <- dbGetQuery(conn=walker,  "
    select id
      from acupuncture_baseline_results
    except
    select id
      from acupuncture_one_year_results
    limit 10
")

helen3
```

```
##     id
## 1  100
## 2  101
## 3  105
## 4  138
## 5  139
## 6  151
## 7  154
## 8  159
## 9  164
## 10 166
```

```
dbDisconnect(conn=walker)
```