Git & SQL Practical Assignment

Student Name

Assignment Overview

This assignment guides you through using Git and SQL in RStudio. You will clone a repository, work with a SQLite database, and use RStudio's Git interface to manage your work.

Step 1: Clone the Repository Using RStudio

1. Open RStudio.
2. Go to **File** → **New Project** → **Version Control** → **Git**.
3. Paste the repository URL provided by your instructor.
4. Choose a local folder and click **Create Project**.

Step 2: Copy the R Markdown Template

1. In the **Files** pane, locate `assignment_template.Rmd`.
2. Right-click and select **Copy**.
3. Rename the copy to `assignment_<your_student_id>.Rmd` (replace `<your_student_id>` with your actual ID).

Step 3: Open Your Rmd File

Double-click your new `.Rmd` file to open it in the editor.

Step 4: Connect to the SQLite Database

```
library(DBI)
library(RSQLite)
con <- dbConnect(RSQLite::SQLite(), "output_database.sqlite")
```

Step 5: Perform SQL Queries

a. List all tables in the database

```
dbListTables(con)
```

```
##  [1] "chemicals_clean"          "chemicals_unclean"
##  [3] "comments_clean"           "comments_unclean"
##  [5] "demographics_clean"       "demographics_unclean"
```

```
##  [7] "df_dictionary"            "df_medications_drug_info"
##  [9] "dietary_clean"            "dietary_unclean"
## [11] "medications_clean"        "medications_unclean"
## [13] "mortality_clean"          "mortality_unclean"
## [15] "occupation_clean"         "occupations_unclean"
## [17] "questionnaire_clean"      "questionnaire_unclean"
## [19] "response_clean"           "response_unclean"
## [21] "weights_clean"            "weights_unclean"
```

b. Show the first 5 rows o a table

```
dbGetQuery(con, "SELECT * FROM mortality_clean LIMIT 5;")
```

```
##   SEQN SEQN_new ELIGSTAT MORTSTAT UCOD_LEADING DIABETES HYPERTEN PERMTH_INT
## 1    1      C-1        2       NA           NA       NA       NA          NA
## 2    2      C-2        1        1            6        0        0         177
## 3    3      C-3        2       NA           NA       NA       NA          NA
## 4    4      C-4        2       NA           NA       NA       NA          NA
## 5    5      C-5        1        0           NA       NA       NA         244
##   PERMTH_EXM SDDSRVYR   VNELIGSTAT       VNMORTSTAT            VNUCOD_LEADING
## 1         NA        1 Under age 18             <NA>                      <NA>
## 2        177        1     Eligible Assumed deceased Alzheimer's disease (052)
## 3         NA        1 Under age 18             <NA>                      <NA>
## 4         NA        1 Under age 18             <NA>                      <NA>
## 5        244        1     Eligible    Assumed alive                      <NA>
##   VNDIABETES VNHYPERTEN
## 1       <NA>       <NA>
## 2         No         No
## 3       <NA>       <NA>
## 4       <NA>       <NA>
## 5       <NA>       <NA>
```

c. Count the number of records in another table

```
dbGetQuery(con, "SELECT COUNT(*) FROM demographics_clean;")
```

```
##   COUNT(*)
## 1   135310
```

d. Join two tables and display the first 5 results

```
# dbGetQuery(con, "
#   SELECT Patients.PatientID, Name, Diagnosis
#   FROM Patients
#   JOIN Diagnoses ON Patients.PatientID = Diagnoses.PatientID
#   LIMIT 5;
# ")
```

Step 6: Disconnect from the Database

```
dbDisconnect(con)
```

Step 7: Commit and Push Your work Using RStudio

1. In the Git pane, check the box next to your .Rmd file.
2. Click Commit.
3. Enter a commit message (e.g., “Completed assignment by ”).
4. Click Commit.
5. Click Push to upload your changes to the remote repository.