# Assignment 2 Exemplary Solutions

Exemplary: Avraham “Avi” Adler, Md Forhad Akbar, Murat “Anil” Akyildirim, Banu Boopalan, Jayanth Reddy Chintireddy and Saratchandra Palle, Steven Ellingson, Amber Ferger, Abdelmalek Hajjam, Jai Jeffryes, John Kellogg, Mikhail Kollontai, Donny Lofland, Devanshu Mehrotra, James Mundy, Mario Pena, Uliana Plotnikova, Scott Reed, Joe Rovalino, Farhana Zahir

Avraham “Avi” Adler [Section 1] Concise and superb.

Md Forhad Akbar [4], http://rpubs.com/forhadakbar/data607_assignment02
Great manipulation of temp password, using SurveyMonkey, odbc connection, and extras: exploratory analysis with gplot, dplyr, cleaning and aggregating data. Bravo for loading the .csv files into the local DB (on Windows).

Murat “Anil” Akyildirim [2], http://rpubs.com/anilak1978/movies-survey Superb. Asked and answered question, “Do movies have to cost more to be better?” which required loading additional information then combining with ratings; also—look at Anil’s RMarkdown style. Handled missing values gracefully, then assigned mean rating to missing values (perhaps not ideal, but *much* better than setting them to 0!!).

Banu Boopalan [4] http://rpubs.com/BanuB/526187 Great step-by-step explanations in rpubs. Using RMySQL for connection, reading DB from MySQL, joining and storing into R dataset, cleaning, visualization, and establishing the rating. Use of merge, Kable Tables and ggplot2 is a plus. Imported csv into MYSQL through the Excel Data Export Tool. [dbWriteTable to write the data frame back to MySql DB didn't work (error)]

*Team Project!* Jayanth Reddy Chintireddy [3] and Saratchandra Palle [3], http://rpubs.com/Jayanth0572/526307. Exemplary. Data sourced from SurveyMonkey; PostgreSQL; RMarkdown has good style, good use of text, good EDA.

Steven Ellingson [3], http://rpubs.com/StevenEllingson/525539. Handled all SQL inside of R code; good handling of NULLs

Amber Ferger [2], Superb work, http://rpubs.com/amberferger/DATA607_Assignment2, soup-to-nuts

Eunkyu Hahm [1],

Abdelmalek Hajjam [3], http://rpubs.com/ahajjam/Assignment2_AH_Data607 Created 3 table solution in SQL; combined data in R; analyzed and visualized resulting ratings

Jai Jeffryes [4], https://github.com/pnojai/cunyds/blob/master/DATA607/Week02/sql_r.md
Excellent explanations! Using DBI with library RMySQL to tackle the connection by a user different from root got the better of MySQL caching SHA2 pswd security.
Denormalizing the db via select with joins flattened the data set into a nice tidy data frame.
It would be interesting to avoid hard coding INSERTS into the DB

John Kellogg [2], http://rpubs.com/kelloggjohnd/526201 Used Google Cloud and movielens data.

# Assignment 2 Exemplary Solutions

Mikhail Kollontai [2], http://rpubs.com/mkollontai/DATA607_HW2 Compared user ratings to IMDB ratings. On SQL side, followed best practices by using three tables, and by providing separate scripts for creating and populating database tables.

Donny Lofland [1], http://rpubs.com/djlofland/DATA607-Homework2, Exemplary. Used surveymonkey, nice styles in .Rmd code, Secured information in RStudio .INI file.

Devanshu Mehrotra [4], https://github.com/DevMeh/607-Assgn-2/blob/master/MovieRatings.Rmd Nice step-by-step comments. Great use of MySQL Workbench csv import function.

James Mundy [4] http://rpubs.com/mutuelinvestor/526053 One of the best solutions:

1. Created simple (unTidy) moviesurvey Mysql database to persist my survey data
2. Created Shiny App to take survey and store results in Mysql table.
3. Used DBI package to connect to Mysql table and create a tibble
4. Used tidyr::gather to Tidy moviesurvey data
5. Produced plot using tidy data and ggplot2
6. standalone shiny app - https://github.com/MundyMSDS/DATA607

Mario Pena [1], http://rpubs.com/marioipena/525858

Uliana Plotnikova [4], http://rpubs.com/uplotnik/525880
Great HW2 submission: use of Postgres DB and RPostgreSQL odbc, loading into a dataframe, analysis and visualization of data using ggplot2.

Scott Reed [1], http://rpubs.com/yasth/Data607Week2 Wow. Scott hand-generated three SQL tables (may not be a best practice, but great for learning and for some data engineering work), then generated sample viewer and movie data from different sources, then randomly generated ratings… There's a sense of play in his work here. [Sometimes, in advanced data science work with networks or Bayesian analysis, you'll compare your results with the randomly generated case so this can be a good tool to have in your toolkit].

Joe Rovalino [3], http://rpubs.com/jrovalino/Data607-HW2, Used getPass package to manage password security. Joe also provided these helpful links as his citations:
https://stackoverflow.com/questions/50544230/connecting-to-mysql-from-r, and
https://www.rdocumentation.org/packages/getPass/versions/0.2-2/topics/getPass

Farhana Zahir [2], http://rpubs.com/zahirf/526165 Collected and analyzed (with nice boxplots) ratings by genre.