SQL DATABASES

There are four rules that apply to this project:
   a)   Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.
   b)   All code must be *scalable by sample size* unless specifically noted otherwise.
   c)   Always use standard headings unless otherwise specified: **R Studio API Code, Libraries, Data Import and Cleaning, Analysis, Visualization.** Delete any headings without code under them.
   d)   All code must be *refactored* to avoid future code rot (i.e., minimize unnecessary code complexity).

Project Context

- You have been given access to a database containing survey responses from a nationally representative sample of respondents.
- You know that in this survey, participants were among other things asked to respond to four questions in the following format: How acceptable, if at all, do you think it is for social media sites to use data about you and your online activities to… 1) recommend events in your area, 2) show you advertisements for products or services, 3) recommend someone you might want to know as a friend, 4) show you messages from political campaigns
- You also know that in this survey, participants were asked which social media platforms they use.
- Given this, you would like to ask the following **research question**: Do people that use more social media platforms (i.e., a larger count of platforms) have greater acceptance of privacy intrusions (i.e., a higher mean score across the 4-point scale), and do these differences vary by age?
- The database can be accessed at **tntlab.org**, port 3306, with user **rnlander_8960r** and password **rTCo.4vQ2vc-**
   o   The first time you access this database, you will need to email Richard your IPv4 address so that he can add your IPv4 address to the SQL whitelist. You can find this number at https://whatismyipaddress.com/ (be sure to send your IPv4, not your IPv6). You might need to do this again if your IP address changes (usually happens about once a week on cable).
- You have been given only this information about the database:
   o   There are three tables containing relevant data.
   o   At least one of these is a lookup table for social media use codes.

Project Instructions

1. This first step has an extra rule: **document your process for discovery within this database**. This means you should write SQL queries and immediately comment why you ran that SQL. If you go down an unproductive path, delete all code and comments down that path, but ensure your final path of discovery is documented. In a file called **week14.Rmd, which must be annotated to explain all steps taken,** import and restructure your dataset two ways. In **sql_tbl**, use only SQL code to import and combine tables. In **tidy_tbl**, import each table directly and combine using **tidyverse**. At the end of this step, **tidy_tbl** and **sql_tbl** should be essentially identical. Both should produce a final "tidy" dataset with all raw variables needed to be used in the next step.
2. Clean one of these datasets as necessary so that you can run the necessary analysis. You need these variables: **mean privacy score**, **age**, and **number of social media platforms**.
3. Run a simple moderation test to address the RQ above using ordinary least squares regression, visualize the relationship, and interpret what you find (i.e., no machine learning).