Final Exam

There are six rules that apply to this final exam.

a) Follow instructions precisely.
b) Do not use any functions or approaches to problems that we did not cover in this course.
c) All code must be scalable by sample size unless specifically noted otherwise.
d) Any code using magrittr should contain a max of one verb per line.
e) Always use the standard heading set up to this point in all files unless otherwise specified: **R Studio API Code, Libraries, Data Import and Cleaning, Analysis, Visualization**.
f) Ensure your file will run correctly from line 1 to the end with a fresh restart of R.

Part 1 – Set up a new R Studio Project with one R script called final.R
Part 2 – Data Import and Cleaning

1. Create a tibble called *imported_tbl* containing the last 2000 posts and metadata from this page: https://www.facebook.com/groups/1904560169871066/
2. Save this tibble in your project's output folder as a comma-separated data file called **original.dat**
3. Create a new tibble called *cutdown_tbl* that contains the name of each poster, the number of likes it received, the number of comments it received, and the post type (i.e., photo, status update, video, or link).
4. Create a new tibble called *cleaned_tbl* that replaces the name variable with a new variable called **first_name** that contains only the person's first name and change any variable types as needed.
5. Create a new tibble called *semifinal_tbl* using the following code, which requires the *gender* library. It will convert first names into their most likely gender, both, or NA if unmatched:
   semifinal_tbl <- left_join(cleaned_tbl,
   distinct(gender(cleaned_tbl$first_name, method="kantrowitz")),
   by=c("first_name"="name"))
6. Create a new tibble called *final_tbl* that drops the **first_name** variable from *semifinal_tbl* and listwise-deletes anyone whose gender was not identifiable as clearly male or female. Also ensure all final variable types are correct.
7. Save this tibble in your project's output folder as a tab-separated data file called **final.dat**

Part 3 – Visualization

8. Create an appropriate visualization of all univariate distributions and zero-order relationships.

Part 4 – Analysis

For this section, don't worry about testing or meeting assumptions. Write a ## comment after each analysis interpreting what you found and re-write results in APA format (e.g., F(1,30)=1.20, p = .76).

9. Conduct a t-test comparing the number of likes received as a function of gender. Calculate a Cohen's *d*. Display the results as a box plot.
10. Conduct a two-way Type III ANOVA examining comments resulting from the post type and gender. Display all post-hoc tests and calculate an overall effect size. Visualize with a marginal means plot.
11. Conduct a Poisson regression predicting like count from comment count with a gender moderator. Visualize as a scatterplot with two Poisson regression lines.