STA 104 Exam II Project, due
Monday, November 30th at 6:00pm CA time

Read the following instructions carefully:

- **You may work in a group of two, or by yourself.**

- **Unlike the homework, you are not allowed to discuss the questions with anyone other than the instructor or TA and your groupmate.**

- **Any outside help beyond that from the instructor or TA is considered plagiarism. This including asking a tutor, your classmates (for example, comparing answers), posting the questions to homework help sites, etc. Should we believe you have sought outside help, you will be reported to the Student Judicial Affairs office.**

- **You are allowed to use or modify your previous functions, or the instructor's functions that are posted online.**

- **Do not share answers, or specific values for calculations, particularly on Piazza**.

- You may ask clarifying questions about code and general approach on Piazza, but do not give away any numerical answers. If you are concerned you may be giving something away, email me or the TA's directly.

- Submit a single group submission on Canvas. This also means you and your partner will get the same grade (assuming you have a partner). You are responsible for your own work, and your partner's work.

Select two topics from the following three to analyze as a group (of one or two people).

Topic I

A student investigated just how effective washing with soap is in eliminating bacteria. To do this she tested four different methods—washing with water only, washing with regular soap, washing with antibacterial soap, and spraying hands with antibacterial spray. Each morning, she washed her hands according to a treatment randomly chosen. Then she placed her right hand on a sterile media plate designed to encourage bacteria growth. She incubated each plate for 2 days at 36°C, after which she counted the bacteria colonies. This procedure was replicated for nearly a month, and she gives you the data to analyze. This data is in `bacterial.csv`

Compare the washing methods, being as specific as you can.

Topic II

The data is found in the file `Mind.csv`, with the following columns:

Column 1: Treatment: The treatment the subject was given for a particular mental disorder (Medication, Therapy).

Column 2: Improve: The improvement the patient showed after 6 months (None, Mild, Moderate, Major).

The particular mental disorder was not disclosed for this data set. The goal is to determine if the variables are independent. If they are not, specify what the dependence is.

Topic III

You have a sample of failure times of lightbulbs in years. The data can be found in the file `failtimes.csv`. We are interested in estimating how different these times tend to be by measuring the interquartile range of the failure times, which is defined as 75th percentile - 25th percentile (of failure times).
Provide an answer using careful application of bootstrap methods.

3. The Report Format

Each question should be a short report. This means you write in full sentences, and have the following sections for each question, while being **as specific as you can** about your results. **There should not be any "copy and pasted" R code in this report. You must format the results you get from R.**. A format suggestion is as follows:

I. Introduction. State the question you are trying to answer, why it is a question of interest (why might we be interested in the answer), and what statistical technique you are going to use. **This must be a non-parametric technique**.

II. Summary of your data (and **only** the data you are using for the question). This should include things like plots (histograms, boxplots) including the interpretation of the plots, and summary values such as sample means and standard deviations. This is where you should justify which non-parametric technique you are using. An R handout is available online for graphing and summaries of various data types.

III. Analysis. Report back confidence intervals, test-statistics, and p-values, nulls and alternatives, etc. You may use tables here, but be sure that you organize your work. Remember to write your results in full sentences where possible.

IV. Interpretation. State your conclusion, and inference that you may draw from your corresponding tests or confidence intervals. These should all be in terms of your problem.

V. Conclusion. Summarize briefly your findings. Here you do not have to reiterate your numeric values, but summarize all relevant conclusions.

VI. Contributions. If you're in a group, briefly list the parts of the report that you individually contributed to.

4. Details

Your report should be the following format:

i. Typed.

ii. A title page/large header including your name/s, the name of the class, and the name of your instructor (me).

iii. Treat each question as a small, standalone report.

iv. An appendix of your R code used to produce the results. Do not include any R code in the body of your report.

For example, your project could be put together in the following order:

```
Cover Page
Parts I-VI for first question
Parts I-VI for second question
Code appendix
```