# Final Project Proposal

**Author: Jialun Li**

[ljlstudy@umich.edu](mailto:ljlstudy@umich.edu)

## Substantive Question

**For what portion of the day are different types of commercial buildings typically in use?**

Different buildings have different operation times. However, the operation time may vary from division to division and also from activities to activities. Some bulidings with specific use in specific divisions may have longer opeartion time than others.

Hence, I planned to carry out analysis to answer the question above.

## Data Set

2012 CBECS data.

**variables:**

- CBECS 2012: PUBID, CENDIV, ACT1, SQFT, WKHRS, WKHRSC, FINALWT, FINALWT1 - FINALWT197

## Proposed Data Analysis

1. Data cleaning and transformation: Select the variables I need and replace the code with labels using codebooks.

2. Select a subset of bulidings with specific first building activity. Some bulidings like warehouse and vacant bulidings are deselected from the whole data set because thier operation time won't vary much. In this analysis, I will mainly focus on important commercial bulidings.

3. Summarize the data by census division and first building activity. The summary statistic I want to use is **the mean of operation hours** and **the proportion of each operation hour categories** for buildings with different activities in different divisions.

The opeartion hours are weighted *also* by **the square footage (SQFT)** of buildings besides **final weights (FINALWT)** in order to compare with different divisions and activities. In other words, the square footage may serve as the *NWEIGHT* variable in the RECS data (*Note: I will also use FINALWT in this data set as weights*).

In a word, I planned to carry out analysis both on per-buliding mean / proportion and per square-footage mean / proportion.

To calculate the confidence intervals, replicate weights (**FINALWT1 - FINALWT197**) are also needed in computation.

4. Plot the summary statistics in a comprehensive way and find out some conclisions.

# Statistical Software

I will use **R (tidyverse and data.table)** only for this final project.