ST 517: Data Analytics I
Final Project

Instructions

This is an individual project, but please still feel free to use the Canvas Q&A discussions to ask questions. There is no single correct approach to this project. The purpose is to give you an opportunity to use some of the tools you've learned during the term.

There are two deadlines:

1. **Deadline 1: due end of week 9** This is a progress report. Please see the details of this submission posted as the homework for week 9.

2. **Deadline 2: due Monday of week 11 (finals week)** This is the completed project. The primary deliverable is a pdf report, but you should also submit an R script that we can execute to replicate your results if we need to (it is your responsibility to verify that your code is well-documented and completely self-contained).

Project Questions

You are provided the .csv file `or_acs_house.csv` which contains household level responses to the American Community Survey for households in Oregon.

Technically this is a Public Use Microdata Sample (PUMS) from the 2015 1-year survey. The data were obtained from http://www2.census.gov/programs-surveys/acs/data/pums/2015/1-Year/. You are provided a subset of variables and only households that have at least one person, pay for their electricity, and are not group accomodation. You may assume this is a random sample of all such households in Oregon.

The variable descriptions can be found in `PUMS_Data_Dictionary_2011_2015.txt`

The questions below are purposely a little vague. It is part of your task to add any additional definitions or other constraints that you find neccessary to make the task more precise. Make sure you justify any alterations or decisions you make.

1. **Explanatory Problem** Do people living in apartments pay less on electricity than those living in houses? How much? Make sure you adjust for (at least) the number of bedrooms and number of occupants in the household.

2. **Prediction Problem** Create a model that could be used to predict electricity costs for a household in Oregon.

3. **Compare and Contrast**

   Dicuss the differences in your approach to questions 1 and 2. Why are different approaches required? What challenges did you face, and how did they compare across the two tasks?