**Introduction**

As water becomes increasingly scarce, it will become increasingly important to design water management regimes that effectively and efficiently manage their water supply. In this analysis, I am trying to determine which group of water suppliers most efficiently manage their water supply: For-Profit companies, Public entities, or Mutual Water companies (i.e., companies that are owned by end-users). To investigate this, I use a dataset of Water Supplier Monthly Water Conservation reports from the California State Water Board. In 2014, the Board mandated that large urban water suppliers had to submit monthly water consumption figures and disclose what water conservation stage they were in. The compiled reports range from June 2014 to January 2021.

For this analysis, I focus on residential consumption in three counties in which all ownership groups can be found: Los Angeles, Kern, and Sacramento. I limited the scope to residential consumers to keep the analysis a little less complicated, but total monthly industrial and residential water consumption can be seen below.

A graph with green and blue lines

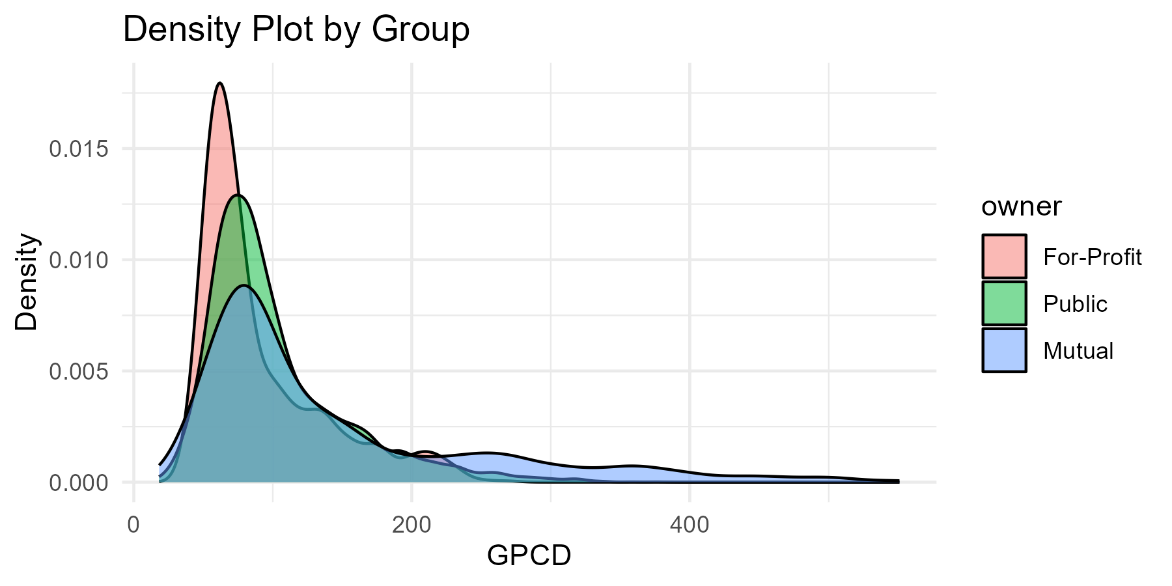
Description automatically generated

**Data Description**

In determining efficiency of water management, I focus on two variables: Residential Gallons Per Capita Day (GPCD) and water shortage stage.

GPCD: Water suppliers report estimates of the gallons of water per person per day by residential customers it serves. GPCD is calculated using the following formula:

where TMP = Total monthly potable water production; PRU = percent residential use; C = Unit conversion factor; TPS = total population served; and MD = number of days in the month. Unit conversion factor refers to methods of converting the water suppliers' potable water unit of choice to gallons [i.e., Million Gallons (MG) = 1,000,000 gallons (G)]. I focus on GPCD because it reflects water consumption habits at the user level.



Due to the long tails, I took the log and use that from now on:

A graph with different colored lines

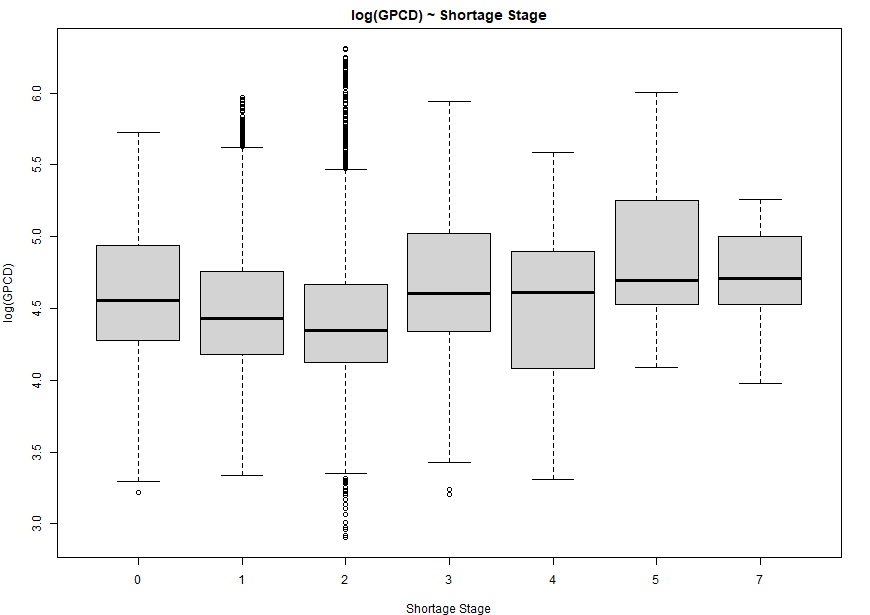
Description automatically generated

Upon first inspection, the box plot below suggests that Mutual water companies typically do have higher GPCD.

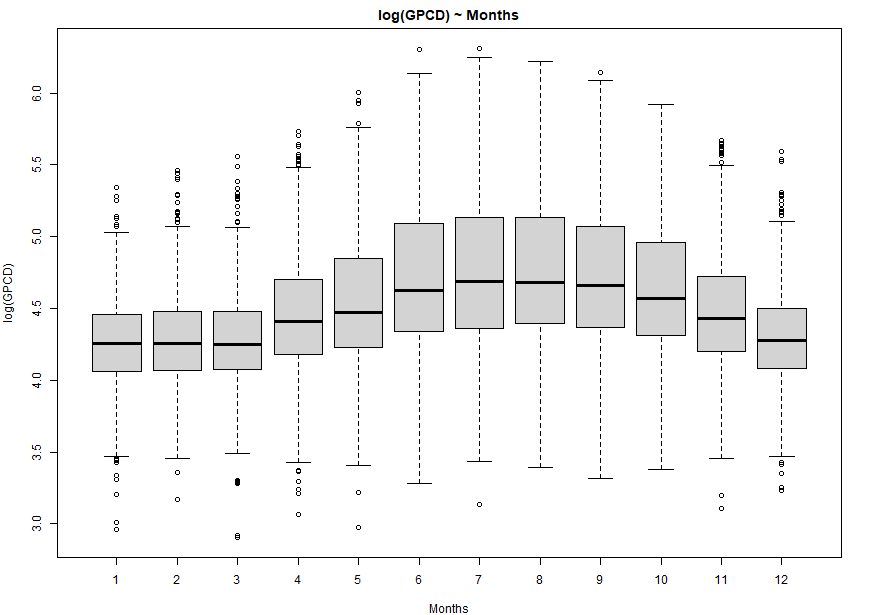
A diagram of a diagram

Description automatically generated

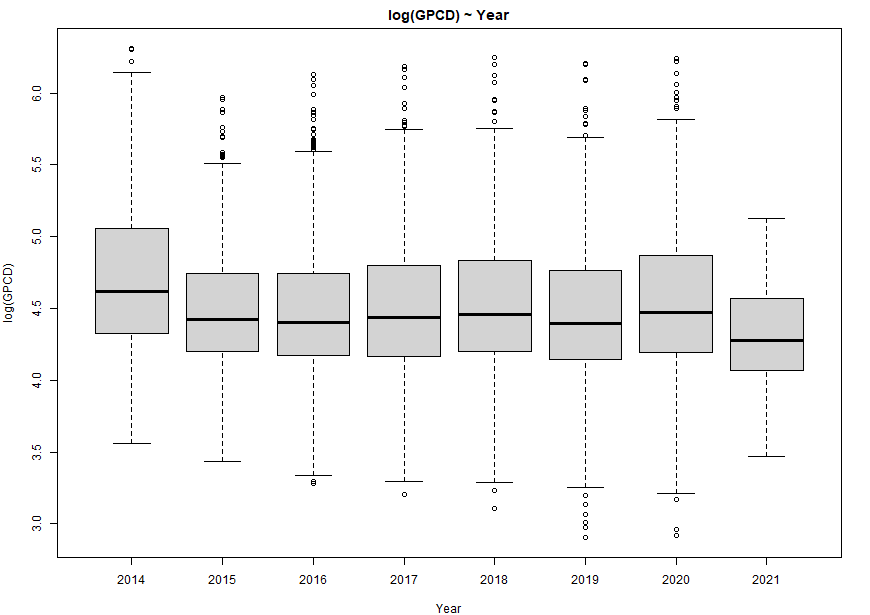
The water shortage stage tells us what water conservation stage the suppliers were in at the time of reporting. This required significant cleaning (done in Python) and still had some strange values after cleaning. For example, the stages range from 1-5 but certain utilities occasionally said that they were in “Stage 7” (see below). Later stages are higher because those suppliers tend to have a higher log(GPCD) value overall.



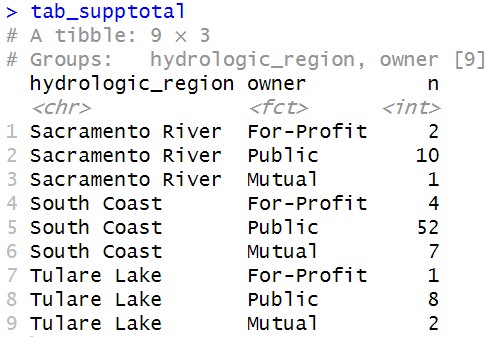
I account for monthly seasonality by using a “summer” dummy variable for June-September. I also include indicators for May and October because they still have higher consumption than is typical for the rest of the year.



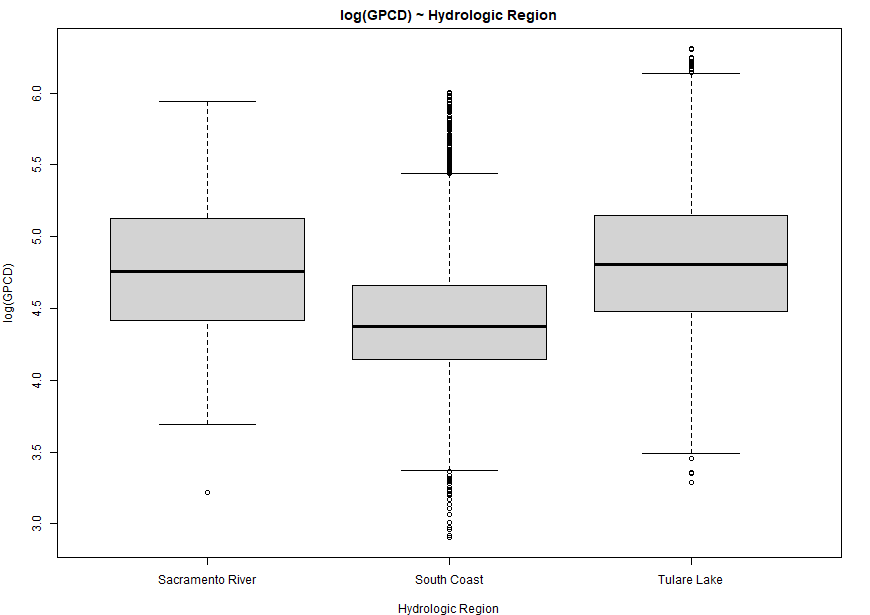
In aggregate, it doesn’t look like there is a significant annual trend, but I include year in my regressions as well to account for any time trends.



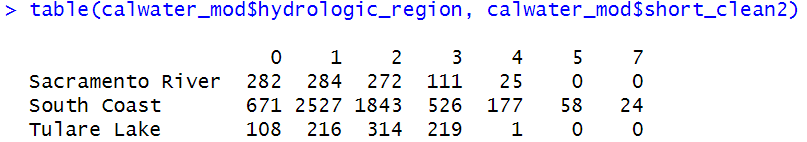
Although I selected which water suppliers to analyze by county, the hydrologic region variable better reflects the environmental constraints in which the suppliers find themselves. Moreover, each ownership group is still represented in each region.



Interestingly, the South Coast region typically consumes significantly less than the Sacramento River and Tulare Lake regions.



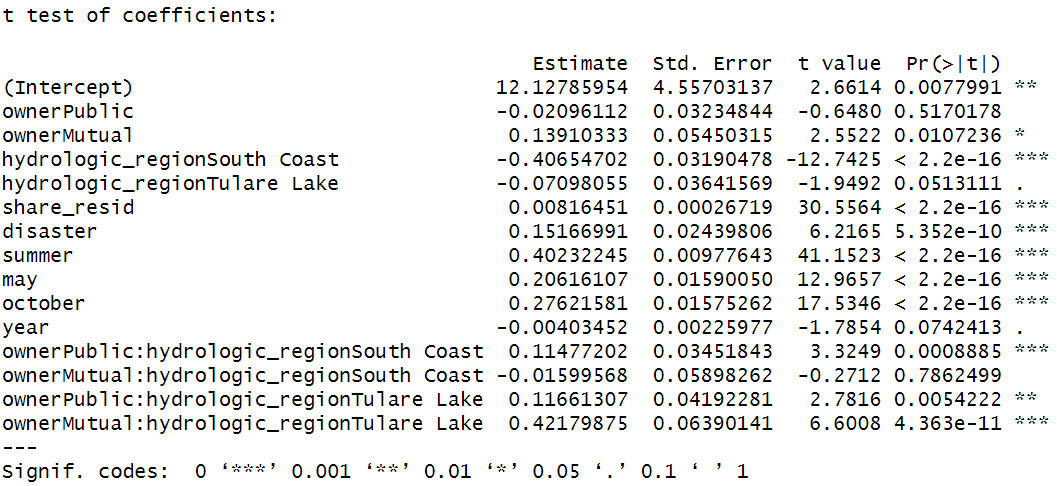
The suppliers most frequently find themselves in stages 1-3.



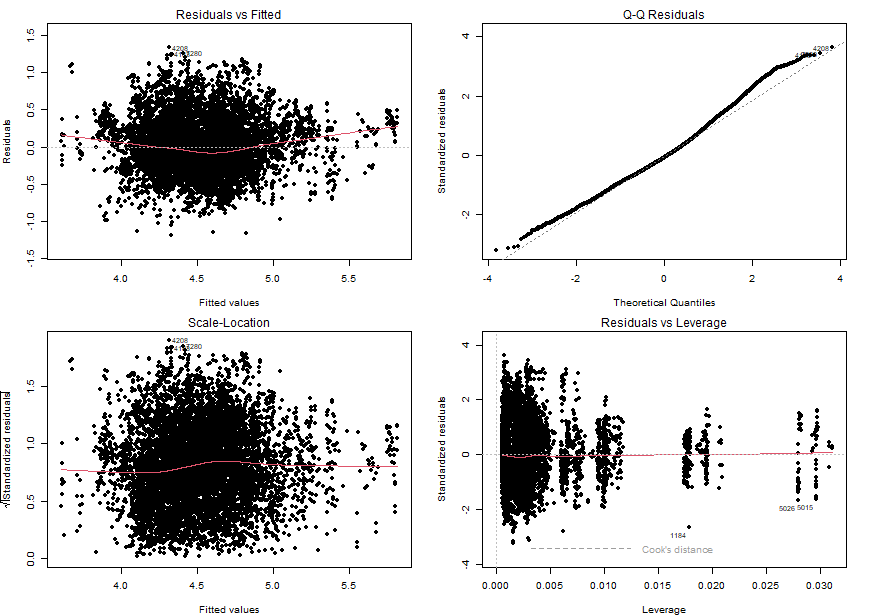
**Modeling Average Daily Water Consumption by Ownership Group via OLS**

The regression below tests whether or not consumers across ownership groups consume the same amount of water controlling for seasonality, share of residential consumers, and not being in a “disaster” (defined as being shortage stage 4 or higher). I’m also clustering standard errors around water supplier. Disaster likely has a positive coefficient because the suppliers that reached a ‘disaster’ stage tend to have higher log(GPCD) [non-disaster suppliers have a median log(GPCD) value of 4.44 compared to 4.55 for disaster suppliers like the city of Beverly Hills; see shortage stage box plot above].

Mutual companies’ consumers use 52.3% and 14.9% more water per day in Tulare Lake and Sacramento River hydrologic regions, respectively, than For-Profit entities and Public consumers use 12.4% and 12.2% more water per day in Tulare Lake and South Coast, respectively. We cannot reject the null hypothesis that consumption is the same in other regions.

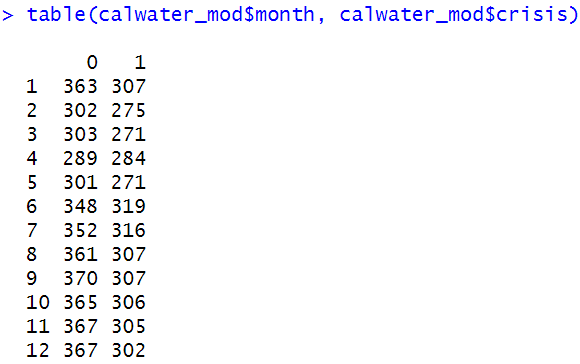


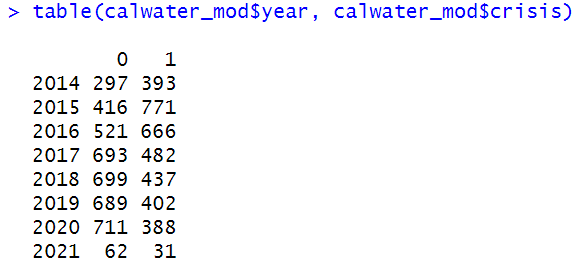
Looking at the diagnostics plot, there is some deviation from the 45-degree line suggesting that the model is missing something. The residual plot also suggests that the model may also violate the heteroskedasticity assumption.



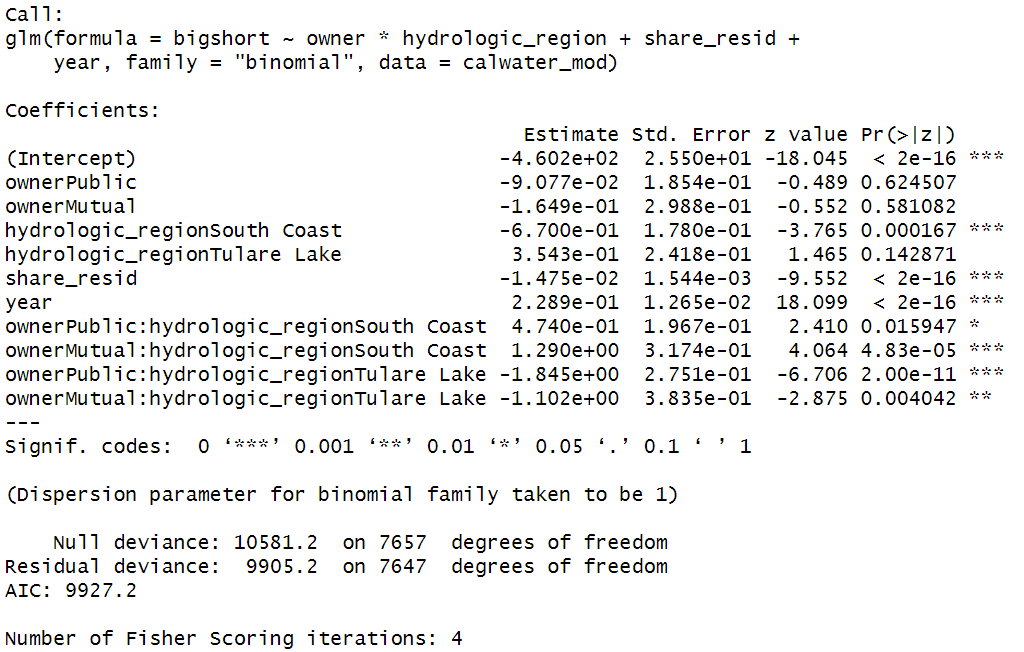
**Odds of Crisis**

I’ve defined crisis as being in a shortage stage of 2 or greater. The table below suggests that month is not a very important indicator for whether or not the utility company will be in a state of shortage while year does suggest that crises became less likely over time.





In the GLM model to predict likelihood of being in crisis, the odds of Mutual Water companies and Public entities being in a state of crisis are 263% and 60% higher than For-Profit companies in South Lake. While the results for Tulare Lake contradict the South Coast findings, we have a much more robust sample in South Lake for each group per hydrologic region vs. ownership table above. If the 1-2 For-Profit companies in Tulare Lake and Sacramento River are simply run not as well, then this would construe the results in favor of the other ownership groups.



Taken together, although the results are a little mixed, they suggest that Public entities and Mutual Water companies manage their water supply less efficiently if we put more weight on the South Coast hydrologic region because we have a more robust supplier sample in that region.