Statistics 133 Final Report:
California Drought
Team HTMN

Kirstian Hope Miranda, Tiffany Jann, Megan Lin, and Noosha Razavian

1. Introduction:

Even with the heavy rains from El Niño this past winter, California's drought continues to persist and deplete the state's water resources. 2016 marks the fifth straight year of this severe drought in California. Water depletion in California is a combination of many factors, including the depletion of snowpacks and groundwater, issues regarding the Bay Area Delta, and rising temperatures. With detrimental effects from climate change and global warming continuing to rise, substantial changes will be needed to mitigate California's drought.

Recognizing the need for serious change, California State Governor Jerry Brown declared a state of emergency on January 17, 2014 and directed state officials to begin to take appropriate actions for the drought. Attached to Governor Brown's State of Emergency was the enlistment of an expanded public awareness campaign, called "Save Our Water," to educate the public on the importance of water conservation, as well as a call for a 20 percent reduction in water use.

On April 25, 2014, Governor Brown issued a second executive order instructing the State Water Resources Control Board (the State Water Board) to create emergency regulations for urban water suppliers so that there would be a reduction in outdoor water use and in any additional inefficient water usages. The State Water Board responded in July 2014 by adopting the Emergency Water Conservation Regulation. Over 70 percent of California urban water suppliers subsequently enacted mandatory restrictions on outdoor irrigation within a few months of the State Water Board's regulation, and the monthly reporting of estimated gallons-per-capita-per-day of residential water users by urban water suppliers began on October 15, 2014.

However, the voluntary 20 percent reduction regulation from January 2014 proved insufficient as the drought raged on. This prompted Governor Brown to issue Executive Order B-29-15, on April 1, 2015, which became the first-ever statewide mandatory water restriction, mandating a 25 percent reduction of water consumption on the state's 400 urban water suppliers. These water suppliers provide water for over 90 percent of towns and cities in California. Owners of large farms were not included in this 25 percent reduction. EO B-29-15 was historic as this was the first time all California urban water users were to follow the conservation efforts and it instructed state agencies, like the State Water Board, to create measures that would protect the state's remaining, but depleting, potable urban water supplies. The State Water Board subsequently issued an emergency regulation in May 2015 to have a 25 percent reduction in potable urban water use. The State Water Board also began a "sliding scale" that placed each urban water supplier into one of eight tiers. These eight tiers are given conservation standards that range from 8 to 36 percent. These standards are relative to whether the respective cities and towns already began reducing their residential gallons-per-capita daily water use since the last major drought in 2013. It is up to the water agencies to find the most cost effective way to meet these standards. Agencies may choose to reward or punish their users through various pricing schemes in order to incentivize consumers to continuing to decrease their water usage.

With all of this in mind, we were curious about how California got to the point of five years of severe drought conditions. We first framed our research by asking ourselves, what were the drought conditions leading up to 2011 and how severe did the drought become from Fall 2011 to present? We were then curious to see how residential consumption patterns changed after the statewide mandatory restrictions by Executive Order B-29-15. Additionally, we examined whether penalties on water supplier agencies actually led to a reduction of residential water usage and if water supplier agencies met the conservation standards set by the State Water Board.

2. Data Collection:

We obtained our drought severity data from the *United States Drought Monitor* website. This resource provided data on drought levels for each week by any selected date range for a selected area and additional parameters. To download the data from the website we followed this pathway 'Maps and Data' > 'Maps and Data Services' > 'Statistical Data' > 'Basic Statistics'.

We then downloaded the CSV data set from January 1, 2000 to April 18, 2016 to build a comprehensive understanding of drought severity with the earliest and latest available data. This data set was valuable to us because it contained the following important variables: drought intensity broken into levels determined by the center (NONE, D0, D1, D2, D3, D4), date, and county of California. To begin our analysis, we read these csv files into R, selected the releaseDate, FIPS, and drought intensity variables (NONE, D0, D1, D2, D3, D4) and started data wrangling.

For our analysis on residential water usage, we utilized the California State Water Board data set. More specifically, on the website we followed this pathway 'Water Issues' > 'Programs' > 'Conservation Portal' > 'Conservation Reporting' > 'April 2016' and downloaded the `June 2014 - February 2016 Urban Water Supplier Report Dataset` which includes information on all the water supplier agencies, reported residential water usage, each supplier's conservation standards, total potable production, and penalties assessed each month on each supplier. The data begins from June 2014 and ends in February 2016. Although, this does seem like a limited timed period, there were no other official data sets that provided as much information on California residential water usage.

In addition, this data was available on the website for download as a CSV file. To use it for our research, we read the CSV file into R, selected appropriate variables for our respective analysis focus (depending on Restrictions, Penalties, or Conservation Standard) and then began data wrangling.

3. Drought in California:

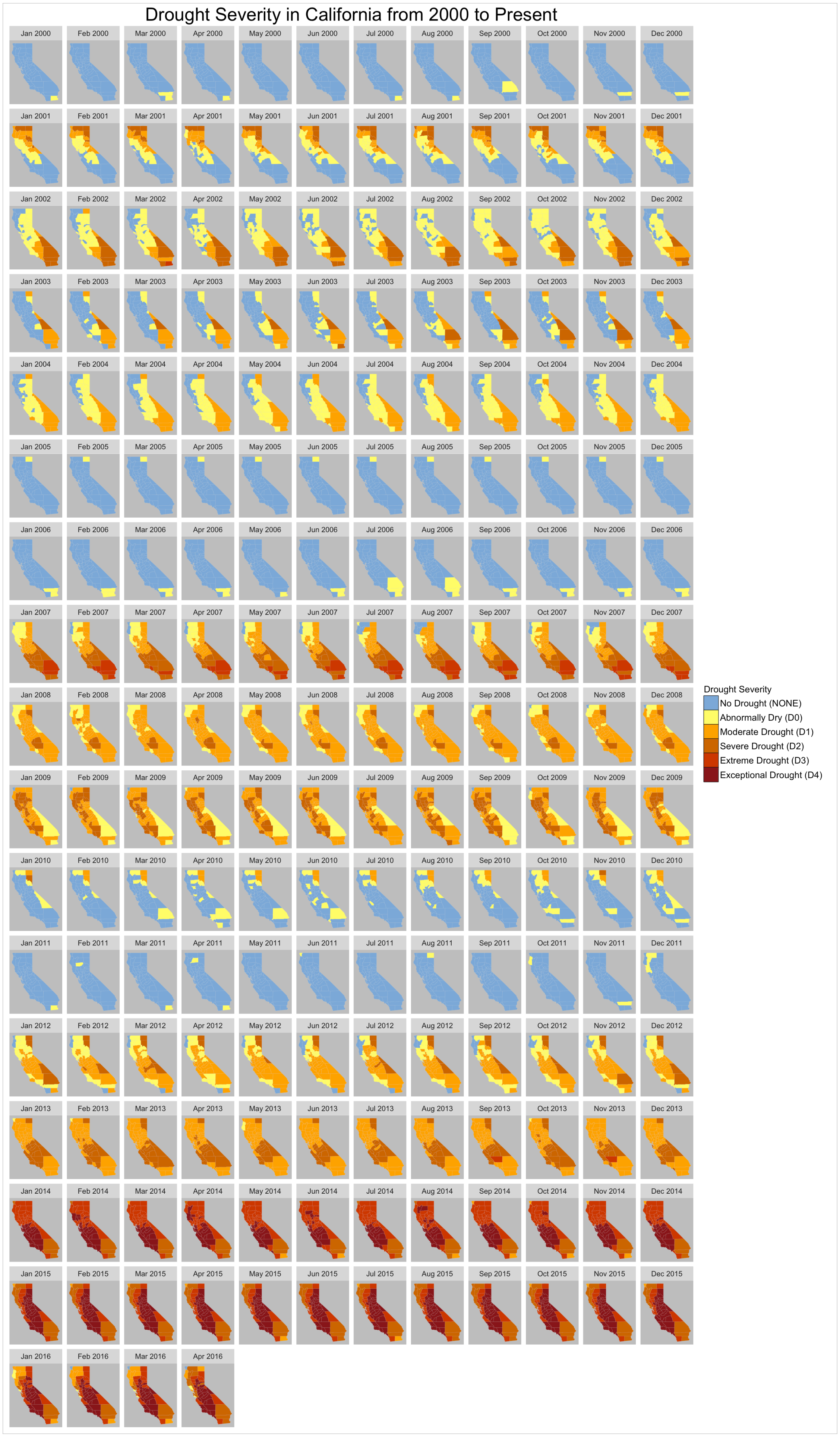
Just how bad is the drought? Over the past 15 years there have been two moderate droughts and one severe drought, which is still an ongoing concern. As a quick overview, we wanted to look at how the drought severity in California changed over time, even before the current drought, as well as the regions affected. Having a visual overview of the extent of the drought allowed us to have a better understanding as to what exactly prompted the recent state legislation.

3.1 Drought in California Overview:

We tidied and wrangled the data from the *United States Drought Monitor* website before joining that data set with a second data set including the coordinates for each California county using the mapdata package to properly map the data. We then took the weighted average of the drought severity for each county and each month so that the correct drought severity was represented for each county. We used geom_polygon from the ggplot2 package for plotting the maps, and faceted by the month.

Figure 3.1.1 shows which counties were affected as well as how quickly the conditions have changed just within the past 17 years. There have been two moderate droughts before our current one, and the contrast just between the previous drought and the current one is staggering.

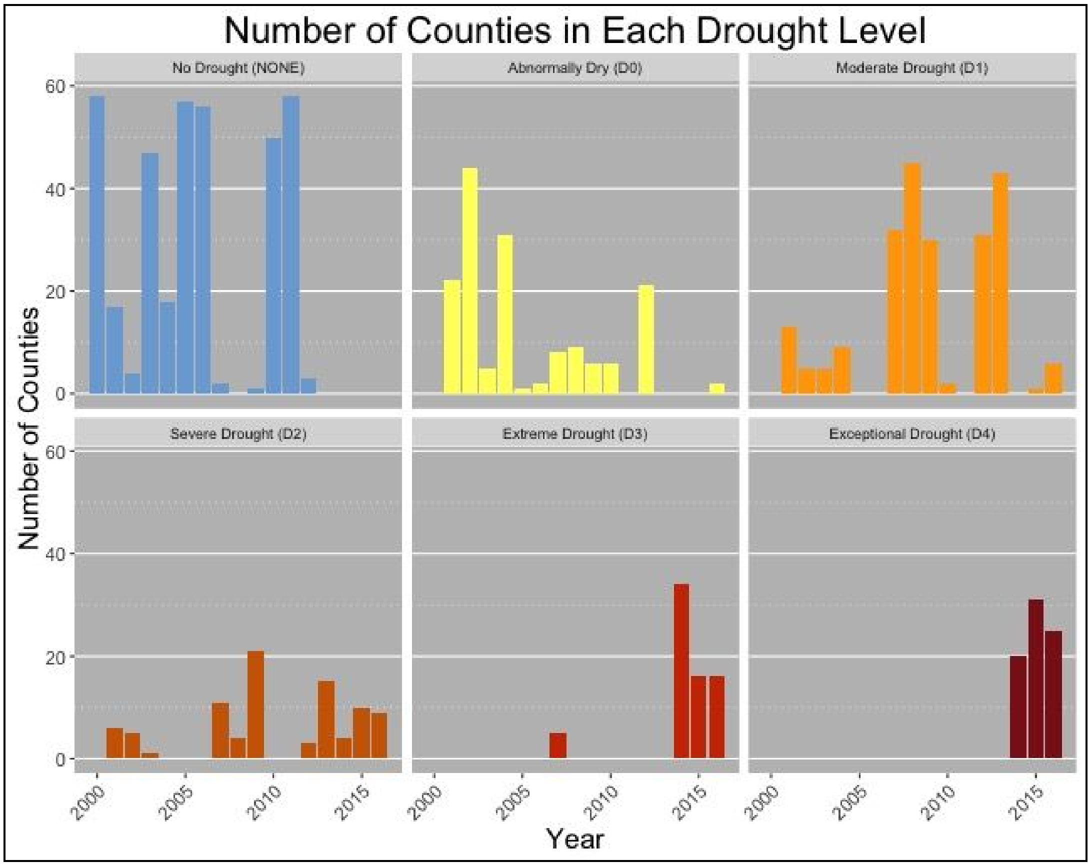
Figure 3.1.1:



Drought Severity in California from 2000 to Present
Jan 2000
Feb 2000
Mar 2000
Apr 2000
May 2000
Jun 2000
Jul 2000
Aug 2000
Sep 2000
Oct 2000
Nov 2000
Dec 2000
Jan 2016
Feb 2016
Mar 2016
Apr 2016
Drought Severity
No Drought (NONE)
Abnormally Dry (D0)
Moderate Drought (D1)
Severe Drought (D2)
Extreme Drought (D3)
Exceptional Drought (D4)

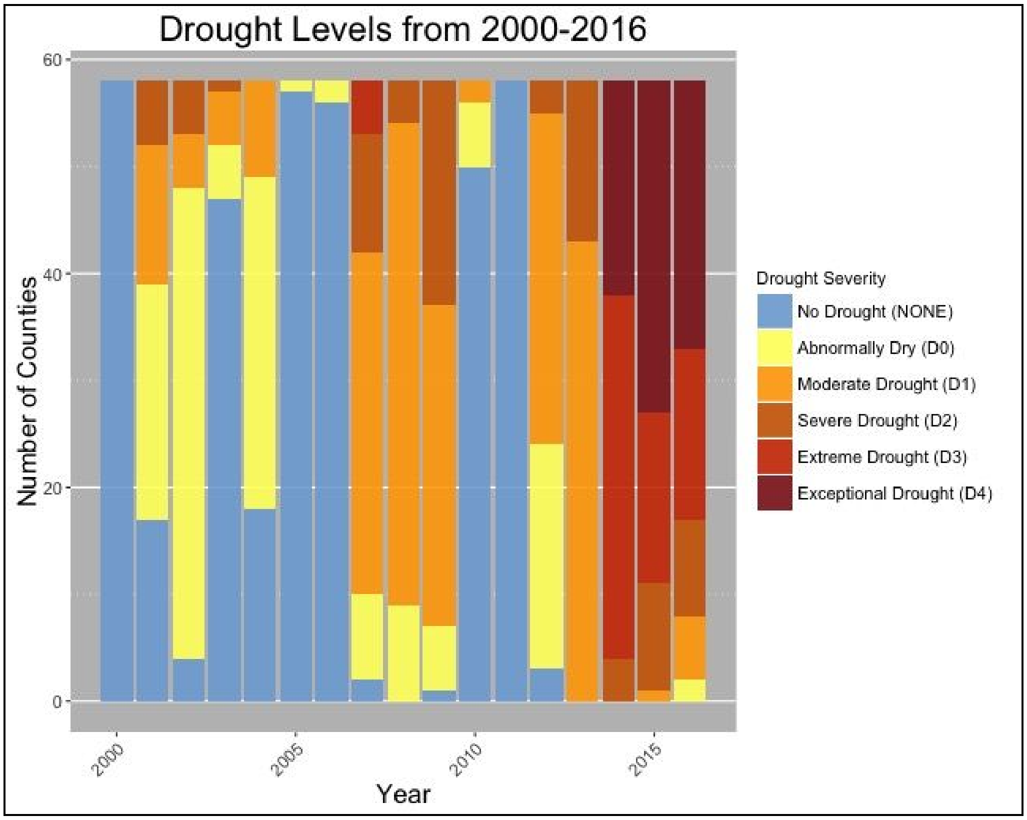
To look at the numerical trends of California's drought conditions, we also made a histogram comparing the number of counties in each drought level for years 2000-2016. The yearly drought levels were determined by weighted averages; the methodology can be found in the appendix at 3.1.2.

Figure 3.1.2:



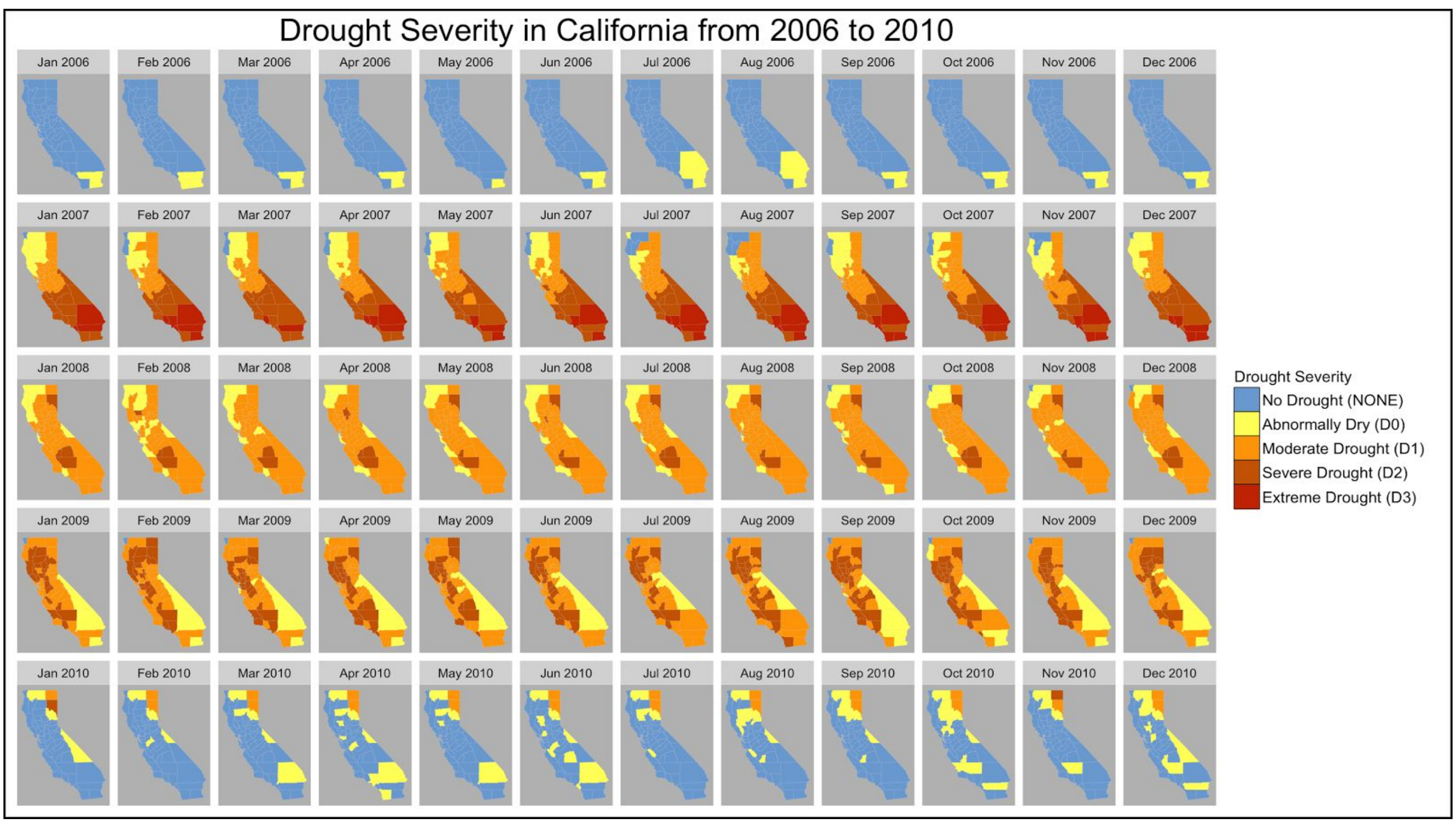
In Figure 3.1.2, we see fluctuations in number of counties for each drought level. We note how the three spikes in NONE – the years between the three drought – alternate with the three general spikes levels D0-D4. Participation in D3 and D4 in the later years show that this current drought is more severe than the previous ones. This is numerical evidence for California's drought developments in the last 15 years. To illustrate the makeup of counties across drought severity levels in each year, we stacked the previous histogram and attain Figure 3.1.3. The pattern of the three droughts mentioned beforehand is even more obvious now, especially if we group all drought levels, from D0 to D4 together, against NONE. Blue dominates 2000, 2005-2006, and 2010-2011, during the breaks between the three droughts. The progressively darker shading in the three droughts also indicate that the most recent drought is much more severe than before, supporting the suggestion that droughts may build on each other, as water supply levels have not been recovering even when we are not in states of drought emergency.

Figure 3.1.3



3.2 Taking a Closer Look:

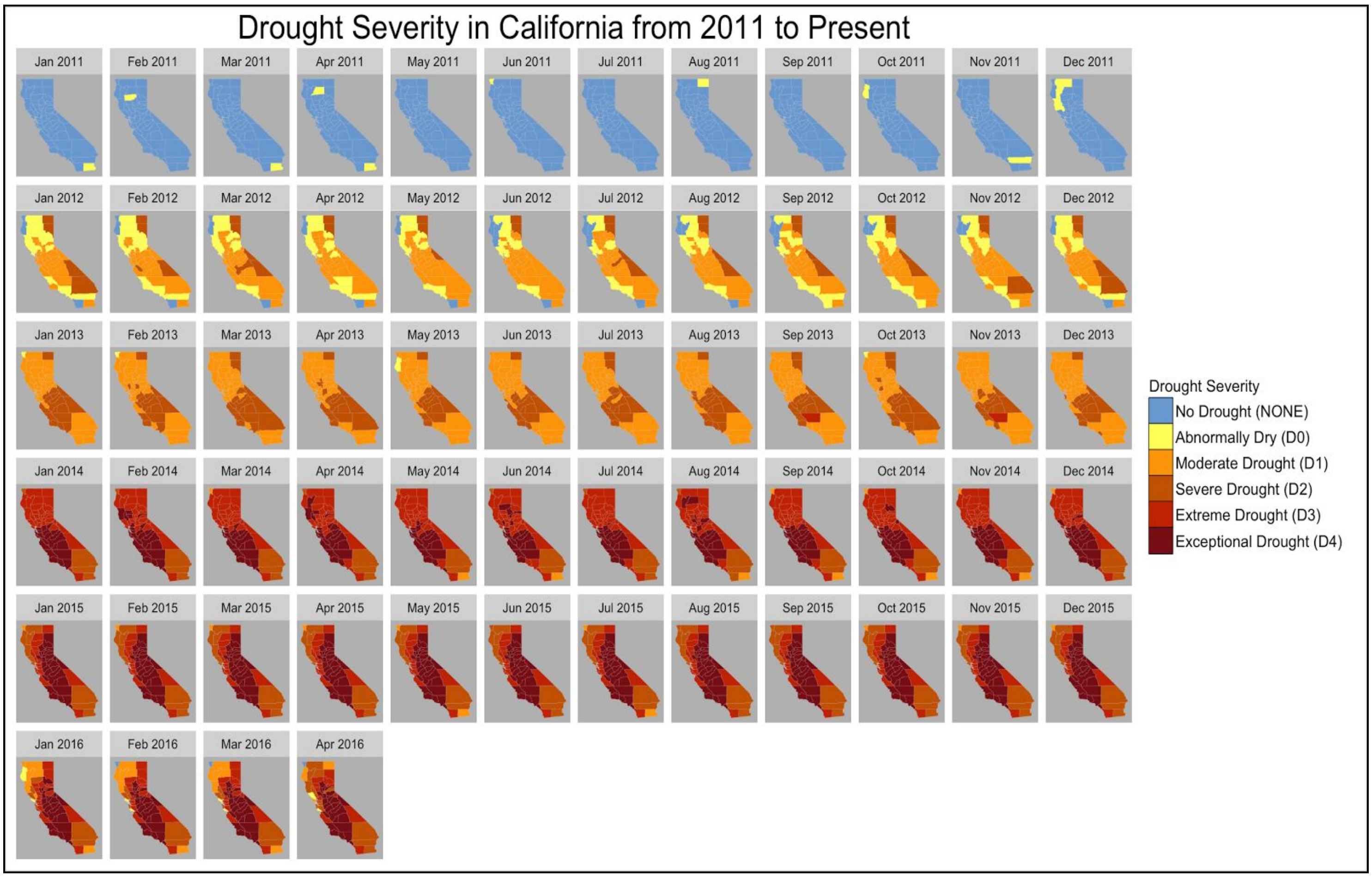
Figure 3.2.1- Drought Severity from 2006 to 2011:



Filtering our previous data set for just the previous drought from 2006 to 2010 allows us to take a closer look at what was going on. In 2007, when the drought severity was the worst for this timespan, Governor Schwarzenegger called for a special session to approve a comprehensive water plan to tackle the issue. The following year he declared a statewide drought and emergency in nine counties, and in 2009 declared a statewide emergency. Despite drought severity levels returning back to "No Drought" for many counties during 2010, as seen in Figure 3.2.1, water levels and water supply remained a great concern. In the beginning of 2010, Lester Snow, Director of California Department of Water Resources, testified in front of the US House of Representatives to call for an actionable plan regarding the water supply for Southern California. He stated, "even if we end this year with normal levels of snow and runoff, our water supply outlook will not improve significantly," which was true seeing as we find ourselves in a very severe drought currently just a few years later. So despite seeing drought severity levels decrease with the above average snowpack levels measured at the beginning of 2010, peaking at 143 percent of average at the end of April, water supply remained a concern.

Filtering the original data set again, but for the current drought beginning in 2011 to the present, allows us to take a closer look at the more immediate and recent concerns that plague Californians.

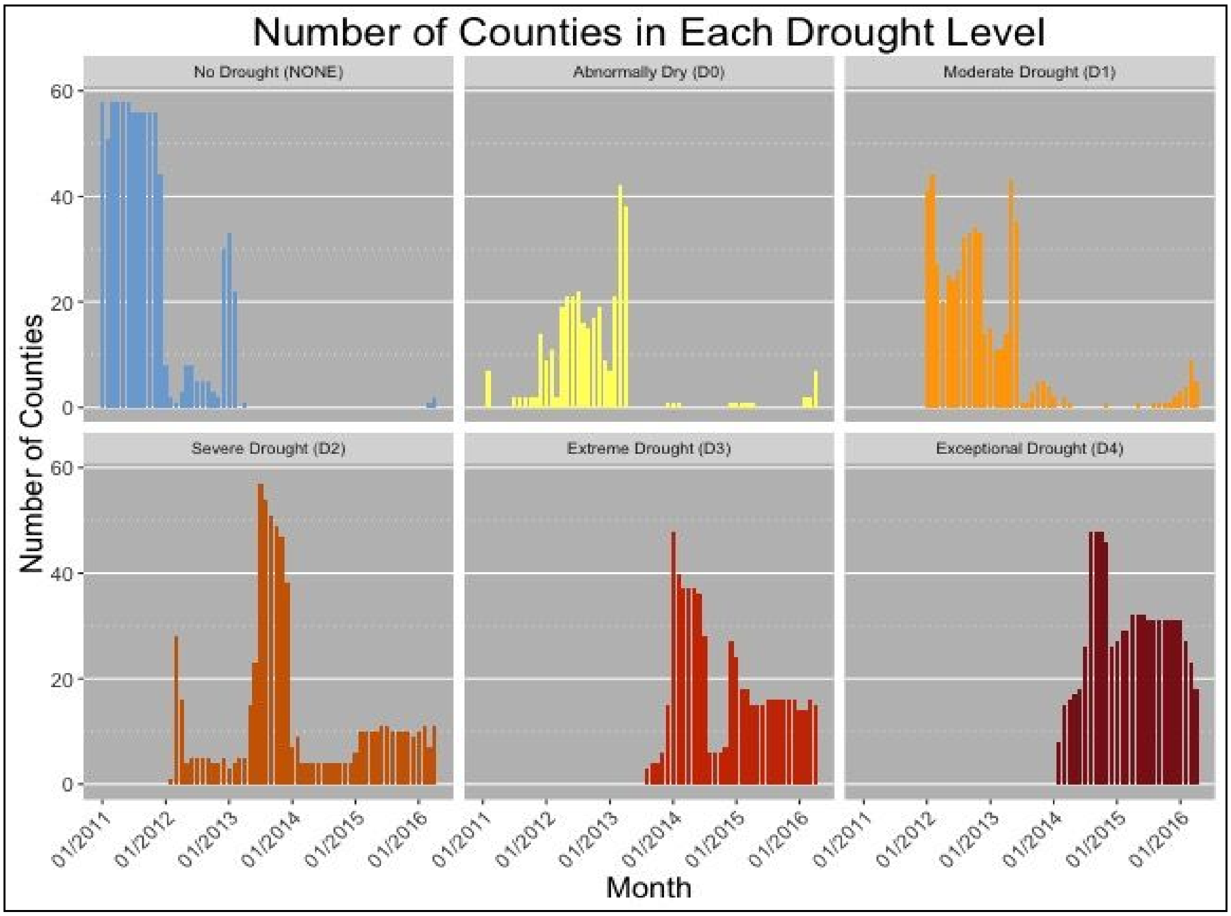
Figure 3.2.2 Current Drought 2011 to Present:



As seen in Figure 3.2.2, the current drought really started at the beginning of 2012. In January of 2012 snowpack levels were only at 19 percent of the average. After a full year of below average levels, and consistent readings of "Moderate Drought" severities in multiple counties, Governor Brown and President Obama's Administration produced an outline for the Bay Delta Conservation Plan. In 2013, California continued to have below average snowpack levels, and Governor Brown signed the first executive orders related to the drought – which streamlined the process for sending canal water to buyers throughout the state. Due to the increase in using groundwater, the withdrawals were the cause of about 1200 square miles of California's Central Valley to begin to sink. Finally in January of 2014, Governor Brown declared a state of emergency, and as we can see above this is when severity levels began to reach “Exceptional Drought” in multiple counties. That same month the California Natural Resources Agency, the California Environmental Protection Agency, and the California Department of Food and Agriculture released a Water Action Plan to address the drought. Studies at the end of 2014 showed that drought is the most severe on record for the past 1200 years, and 2014 was also the warmest year on record. Going into the fifth year of a severe drought a draft for the “Central Valley Project and State Water Project 2016 Drought Contingency Plan For Water Project Operations” was released in January.

Using the exact same data set and weighted average techniques that produced Figure 3.2.2, the following histogram examines California’s most recent drought conditions more numerically.

Figure 3.2.3:



Although we lose geographical insight with the histogram, from here we can quantitatively see the severity of the drought change over time. As we start off in 2011, almost all 58 counties are classified as “No Drought”, when we see a dramatic dip in number of counties classified as “No Drought,” we see increases in levels D0, especially D1, and slowly in D2, D3, and D4 as well. At the peak of the drought in 2014, we see no counties in NONE, as well as spikes in D3 and D4, which taper off but are together still the dominant levels today, as of May 2016. Note that from Figure 3.2.2, we saw colors grow increasingly intense from 2011-2016. From this histogram, the dates in which each drought level facet peaks tells the same story. We see NONE peak throughout the year in 2011, before the drought started in the beginning of 2012. D0 is the first to start picking up, beginning in 2011, and really picking up in 2012. D0 is soon followed by a fast spike in D1 starting in 2012, and we see D0 essentially die off in 2013. Next to peak is D2 in 2013, then D3 in early 2014, and finally D4 in mid-2014 in the heat of the drought. In this manner, the histogram really shows how the drought intensified through time, as well as how the 58 counties were distributed among the six drought classifications through time.

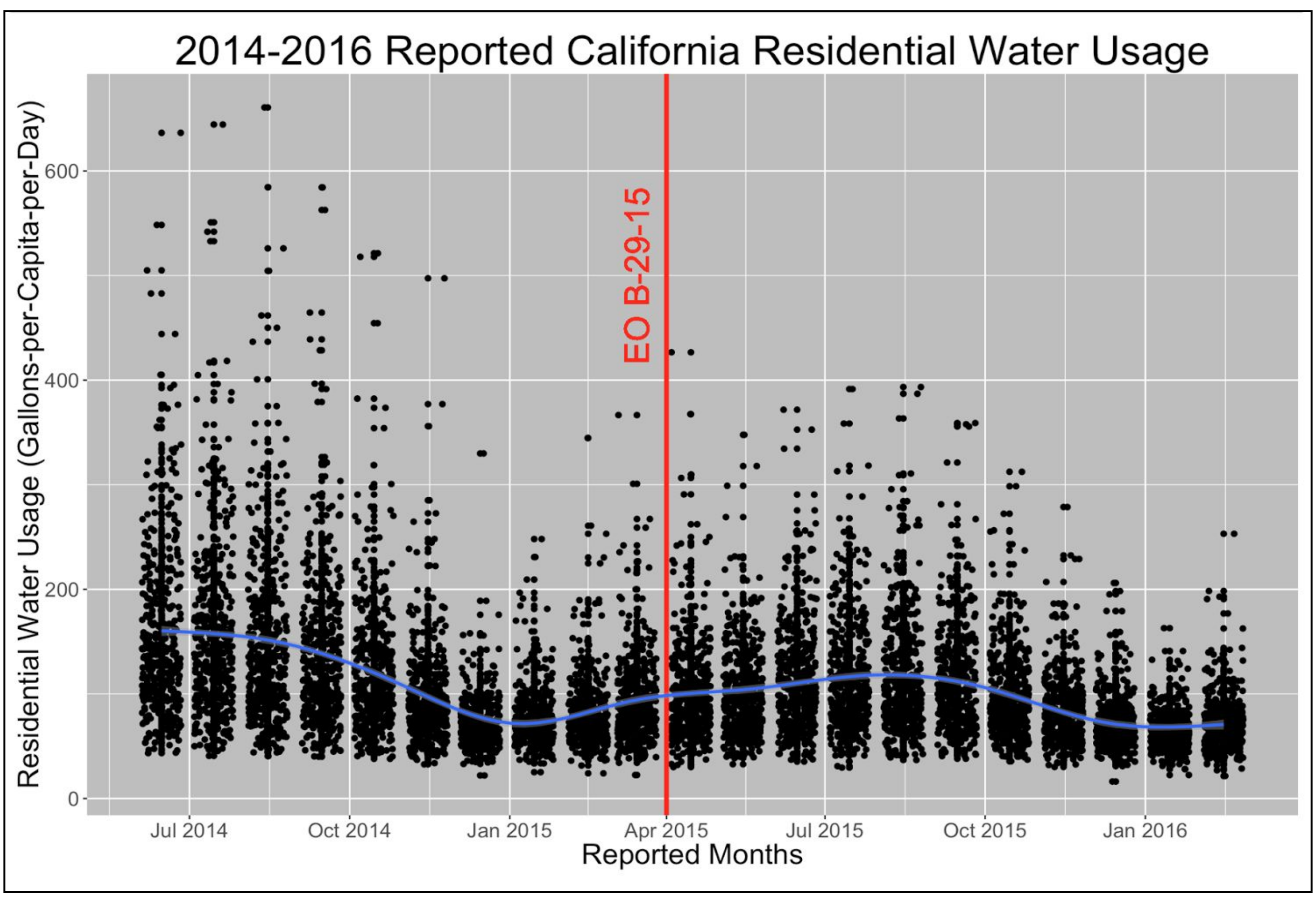
4. California Residential Water Usage:

The California State Water Resources Control Board (the State Water Board) wears many hats as a governmental agency. The State Water Board not only oversees the state's water quality by implementing the Clean Water Act in California and developing statewide policies, but it also is the only agency who reports statewide residential water usage. Urban water supplier agencies report their residential usage on the 15th of every month. For these reasons, we knew the State Water Board's data sets would be the best place to answer our research questions regarding urban water use. The State Water Board's data set includes data on the reported residential water usage and other related information on the supplier agencies, such as each supplier's conservation standards, total potable production, and penalties assessed each month on each supplier. The data begins from June 2014 and ends in February 2016. We recognize that this short time period limits our analysis, but no other data sets provided as much information on statewide residential water usage.

It is important to note that the variable we used to represent residential water usage was the calculated residential gallons-per-capita-per-day (R-GPCD) from the State Water Board's data set. We chose to use this variable instead of the variable indicating the initial reported R-GPCD from each water supplier agency, because the calculated R-GPCD variable consisted of a standard calculation that the State Water Board used for each water supplier agency in each reported month. The calculation of R-GPCD by the State Water Board was the following: R-GCPD = [(TMP*PRU*C) / TPS] / number of days in the month, where TMP represents the total monthly potable water production variable, PRU is the percent residential use variable, C is the unit conversion factor (since the initial reported usage was in different units, such as gallons and acre-feet), TPS represents the total population served variable. For simplicity, we will refer to the calculated residential gallons per-capita-per -day as just R-GPCD and will refer to the Governor's April 1, 2015 executive order, EO B-29-15, as just "executive order" (the executive order to should not be confused with any other executive orders prior to April 1, 2015).

4.1.1 Residential Water Use- An Overview:

Figure 4.1.1:



We began our research on California residential water usage by creating a simple, but very informative, scatter plot diagram, which is seen above in Figure 4.1.1. Each point represents each water supplier agency's R-GPCD for each reported month. We jittered the points so that we could see the data more clearly and better evaluate the scatter plot diagram. Additionally, we included the red vertical line to represent EO B-29-15, the first-ever mandated statewide residential water restrictions. Lastly, the blue curve illustrates the seasonal variation of residential water usage; residential water usage is higher in the spring-summer seasons, as compared to the winter season, because of several factors, including warmer temperatures and increased outdoor water use during spring-summer. Figure 4.1.1 reflects how the winter months of January and February have the lowest urban water usage.

Keeping seasonal variation in mind, it does appear that R-GPCD has decreased in each corresponding month (i.e. comparing June 2014 to June 2015, and so forth). If we take a closer look at the different slopes of the blue curve, we can see how this diagram might suggest that urban water usage has decreased during the spring-summer seasons. This is because the slope of the blue curve from the summer of 2014 to the winter of 2015 (before the executive order) is steeper than the slope of the curve from the summer of 2015 to the winter of 2016 (after the executive order).

Jittering the points also made it clear how the variation in the quantity of water used by each urban water supplier was much higher before the executive order as compared to after. For example, if we take a look at the summer months of June, July, and August of 2014 we can see how R-GPCD varies tremendously from around 50 R-GPCD to over 500 R-GPCD. This is in contrast to June, July, and August of 2015, where R-GPCD varied from 50 R-GPCD to less than 400 R-GPCD. The comparison of the summer months led us to want to examine how many water supplier agencies fell into different ranges of R-GPCD during each month. This is where our analysis turns to next.

4.1.2 Residential Water Usage Categories:

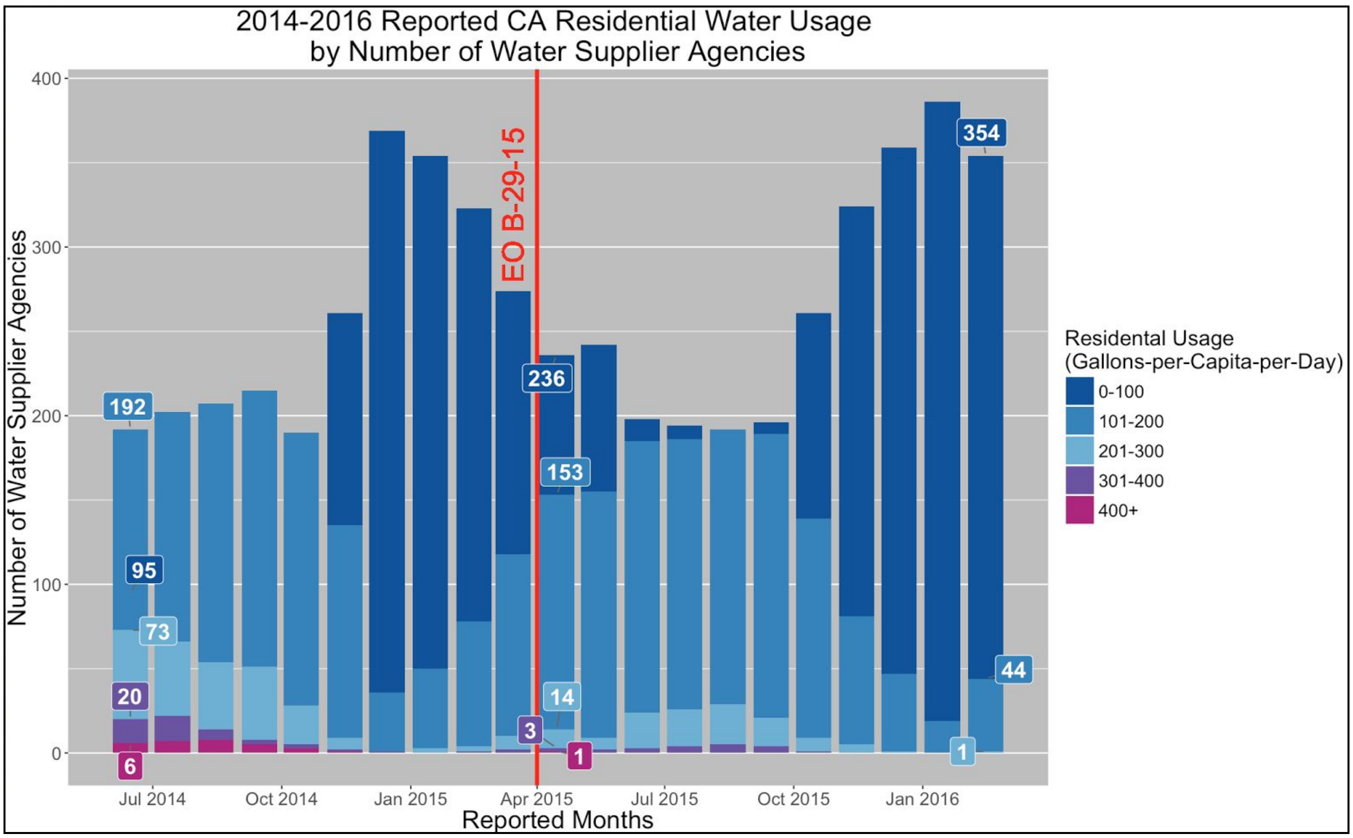
Before analyzing the different ranges of R-GPCD during each month, we performed some data wrangling by using mutate and if-else statements to create a new variable (named overtime_use) that divided the quantities reported in the R-GPCD variable into five ranges: 0-100, 101-200, 201-300, 301-400, and 400+ gallons-per-capita-per-day. Next, we grouped by the new variable and the reported months. Finally, we used summarise to calculate the total number of water supplier agencies in each of the five ranges during every reported month.

We decided the most fitting way to display the different R-GPCD ranges and the number of water supplier agencies over time was a bar graph. The graph is seen in Figure 4.1.2 below. A red vertical line was used once again to represent the Governor's April 1, 2015 executive order. Additionally, we utilized the ggrepel package to create helpful labels for the bar graph. The labels were intentionally used for June 2014, April 2015, and February 2016. The reasoning behind this is that we wanted to see how many water supplier agencies were in each range during the first reported month, the month of the mandatory statewide restrictions, as well as the last reported month in the dataset.

The results of Figure 4.1.2 indicate that the 201-300, 301-400, and 400+ R-GPCD ranges declined following the statewide restrictions. The number of water supplier agencies in the 400+ range is nonexistent following April 2015, while the number of agencies in the 301-400 range disappears after October 2015. For the 201-300 R-GPCD range, the number of agencies in that range was declining before the executive order due to the winter season and seasonal variation. However, if we compare corresponding months to one another, this figure makes it clear that there number of agencies in the 201-300 was less following the mandatory restrictions, with there being only one agency in this range in February 2016. Again, it is important to note that the number of water supplier agencies in the more than 200 R-GPCD ranges were declining before the executive order partially due to seasonal variation and the voluntary restrictions that were in place before April 2015.

In terms of the 0-100 and 101-200 R-GPCD ranges, it is not as clear on the bar graph if our initial hypothesis was correct because of the seasonal variation. The 0-100 range appears to be more common in the winter season, while the 101-200 range appears to be common during the summer-fall seasons. If we examine very closely, we can see how the number of water supplier agencies in the 101-200 range has decreased each month after the mandatory restrictions, as compared to the corresponding months before the mandate. This finding makes sense since the number of agencies in the 101-400+ ranges have declined, for the most part, after the mandatory restrictions in comparison to the prior corresponding months. This decline is because, as it appears in Figure 4.1.2, the number of agencies in the 0-100 range has increased each month after the mandatory restrictions, as compared to the corresponding months. Although we cannot make any definitive conclusions, this figure does suggest that the urban water supplier agencies and residential water users began to limit their water usage following the April 2015 executive order.

Figure 4.1.2:



4.1.3 Residential Water Usage Statistical Analysis

During our analyses of Figures 4.1.1 and 4.1.2, we were aware that the seasonal variation of residential water usage made it more difficult to better examine whether water usage had actually declined since the statewide mandatory restrictions. For these reasons, we chose to run a fixed effect regression. We chose to run the following fixed effect regression:

$$resid\_water_{it} = \beta_1\, executive\_order_{it} + supplier\_name_i + months_t + u_{it},$$ where

$resid\_water_{it}$ stands for the R-GPCD,

$executive\_order_{it}$ is a dummy variable where a 1 indicates during and after the April 2015 executive order and 0 indicates before,

$supplier\_name_i$ is a fixed effect for each of the 408 water supplier agencies,

$months_t$ is a categorical variable that ranges from 1-12 to reflect 12 months in a year, and

u_{it} is an error term.

Before we could run this regression, we did some data wrangling with mutate and if-else statements in order to create the $executive\_order_{it}$ and $months_t$ variables, which were not present in the data set before. To clarify, the $months_t$ variable is different from our previous month variable that represented the reported months. The $months_t$ variable in this case assigned a 1 through 12 to each reported month depending on what month it was. For example, "2014-06-15" and "2015-06-15" were both assigned 6's. An *i* and a *t* are placed on specific variables to denote the unit of observation (each water supplier agency) and the time (each month, 1-12), respectively.

Our main independent variable is $executive\_order_{it}$ and our dependent variable is $resid\_water_{it}$, because we wanted to examine how R-GPCD changed following the April 2015 executive order. We chose to run a fixed effect regression because we needed to control for each water supplier agency and for the seasonal variation, which provided a better estimated coefficient for $executive\_order_{it}$. The water supplier agency fixed effect variable can be interpreted as follows: $supplier\_name_i$ includes all the unobserved, time invariant factors of each water supplier agency, such as the agency's geographic location and demographic features, that affect $resid\_water_{it}$. Meanwhile, the months fixed effect can be interpreted as all the unobserved, invariant factors of each water supplier agency in each month that affect R-GPCD. We acknowledge that there are limitations to our regression. Other factors such as the price of water should be included in this regressions. However, because there is a thorough economic literature debating how to value water, we decided against including such a variable in our regression.

Using the lfe and stargazer packages, we ran the above fixed effect regression and found the results seen below in Table 4.1.3. For clarification purposes, we renamed the $executive\_order_{it}$ as "Statewide Mandatory Restrictions" and the $resid\_water_{it}$ as "Resident Water Use (R-GPCD)" in Table 4.1.3. Because there are 408 water supplier agencies in our data set and 12 months in a year, we chose to exclude the estimated coefficients of our fixed effects from Table 4.1.3. If we had included the results of the agencies and the months, we would have seen the results for 407 agencies and 11 months, since one observation of each fixed effect is omitted from the regression due to the third assumption of fixed effects: Each independent variable needs to vary across unit *i* and over time *t.* In other words, we did not want multicollinearity when running our regression.

Table 4.1.3:

	Dependent variable:
	Residential Water Use (R-GPCD)
Statewide Mandatory Restrictions	-19.982***
	(0.680)
Observations	8,533
R^2	0.772
Adjusted R^2	0.760
Residual Std. Error	29.060 (df = 8111)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

As stated before, the $executive\_order_{it}$ variable was our main independent variable of interest. Like many econometric hypotheses, the null hypothesis here was that the $\beta_1 = 0$ and the alternative hypothesis was that $\beta_1 \neq 0$. Since the estimated coefficient of the $executive\_order_{it}$ variable was statistically significant at the 1 percent significance level, we can reject our null in favor of the alternative hypothesis, and $\widehat{\beta}_1$ is statistically greater than zero at the 1 percent significance level. This suggests that the $executive\_order_{it}$ variable has a statistically significant and negative relationship with expected R-GPCD, holding all else constant. In other words, these results suggest, but do not definitively mean, that the enactment of the statewide mandatory restrictions led to a 19.982 gallons-per-capita-per-day decrease in residential water usage, holding all else constant. This result is promising because the mean residential usage over all the months is around 100 R-GPCD. In other words, the results suggest that residential water usage declined by about 20 percent, which is not far from Governor Brown's mandate of a 25 percent reduction of water consumption.

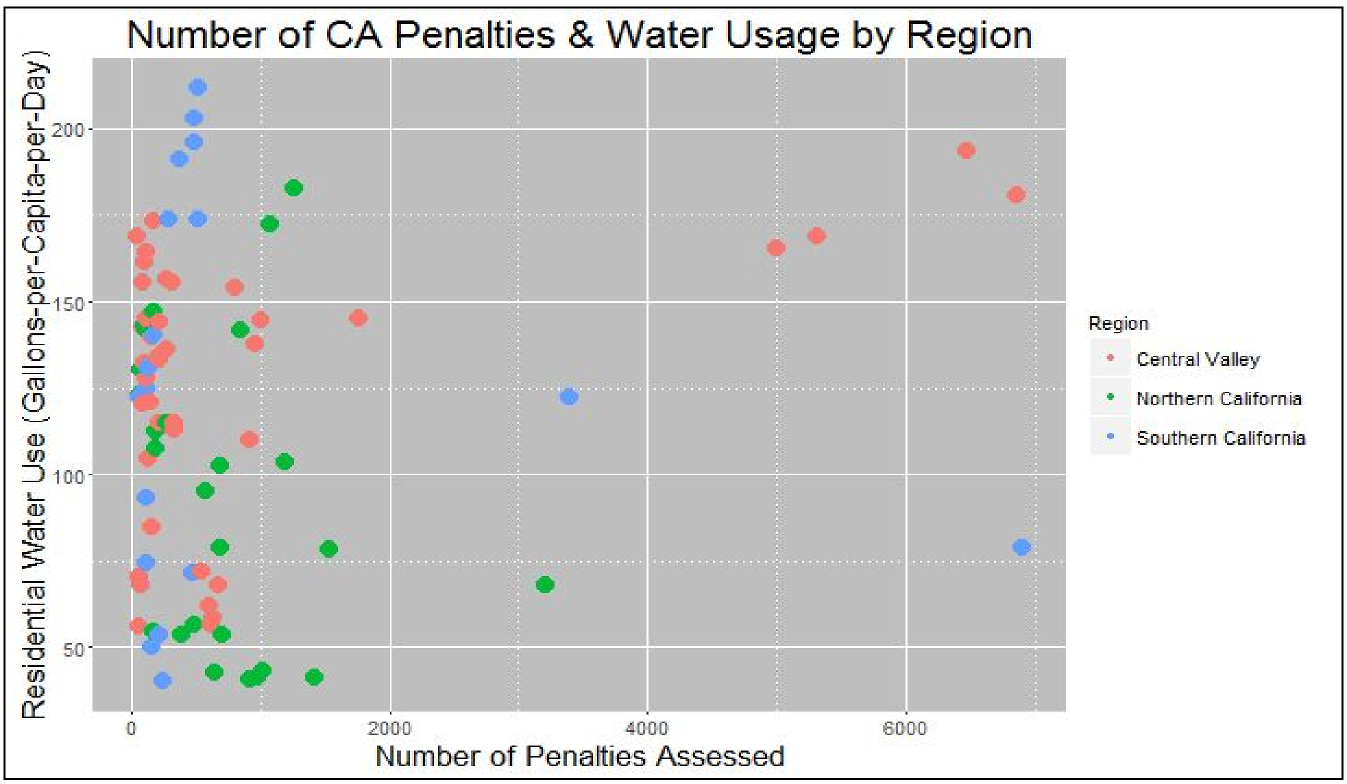
4.2.1 Impact of the Number of Penalties Assessed on Residential Water Usage

To tidy up the Supplier Usage CSV file, we selected our variables of interest (Water Supplier Agency, Residential Water Use, Penalties Assessed and Reporting Month). We believed that it would be more representative to look at the impact by region rather than as an entire state because we could control for potential water usage variations associated with geography. Moreover, we decided to examine the three most well-known geographic regions of California (Northern, Southern, and Central Valley).

To display the relationship between number of penalties and residential water usage, we created a scatter plot. We first grouped together all the Penalties and Water Supplier Agencies, then generated a new variable 'average' which was the mean Residential Water Usage for each category of penalty amounts, and then filtered out the Supplier Names for our regions of interest. More specifically, for each region, we selected eight water supplier agencies in the region that each had at least 100 total penalties assessed, since over 100 total penalties appeared to be a relatively substantial amount of penalties.

Next, in order to display our results well visually, we decided to look at penalties greater than or equal to an amount of 50. Figure 4.2.1 shows how each region of California was affected by the penalties.

Figure 4.2.1:



From Figure 4.2.1, we notice that in the Central Valley, for the most part, as the amount of penalties increases, the residential water usage begins to decrease. However, there are also several outliers from the range of 0-2000 penalties where residential water usage seems to be higher at higher penalty amounts than at a lower amount of penalties. On the other hand, it is interesting to notice that for Northern California, it seems that as the number of penalties increase within the 0-500 range, the residential water usage increases as well. However, after 1000 penalties, even as the number of penalties increase, the amount of residential usage for Northern California remains at levels all within the range of 150-200 R-GPCD. This could suggest that regardless of the amount of penalties, Northern California residents did not believe the conditions were as severe and thus despite the increase in penalties, they chose to maintain the same behavior.

Moreover, from Figure 4.2.1, Southern California is the only region that suggests a positive impact of the number of penalties on the reduction of residential water usage. More specifically, for Southern California, as the amount of penalties increased from range 0-1000, the R-GPCD also increased. However, for penalty amounts greater than 1000, residential water usage decreased. This suggests that as the severity of penalties increased to a certain threshold, residents in Southern California may have become more concerned with water usage and reversed their behavior. We also decided to make three separate plots (Figures 4.2.2-4.2.4) to show the relationship between number of penalties assessed and residential water usage for each region because it is useful to see the variation in water supplier agency.

Figure 4.2.2

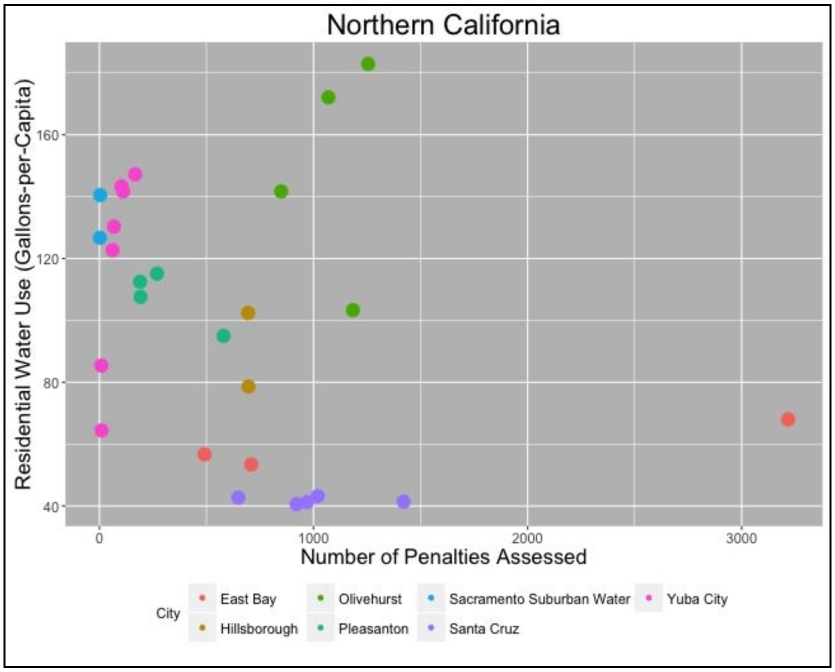


Figure 4.2.3

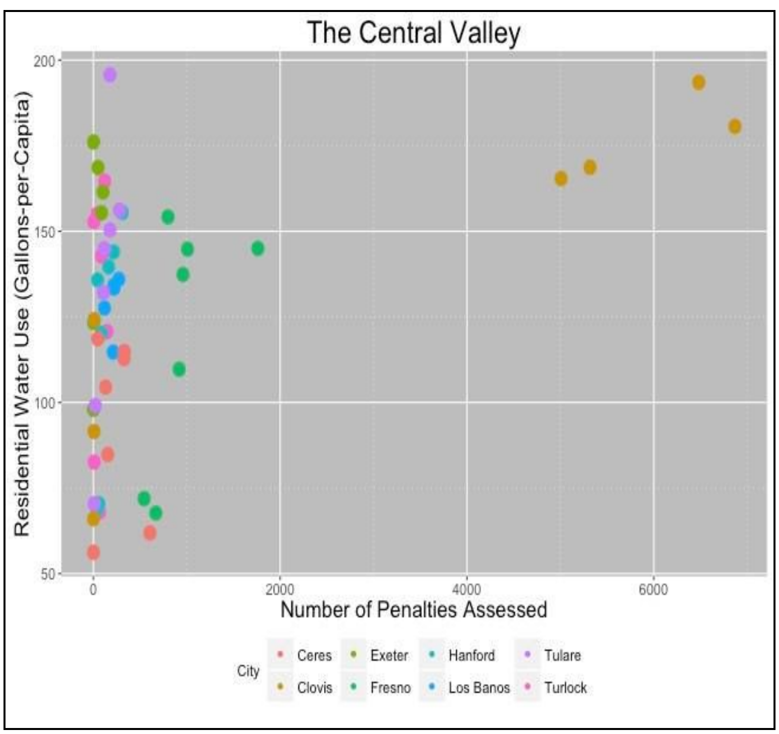
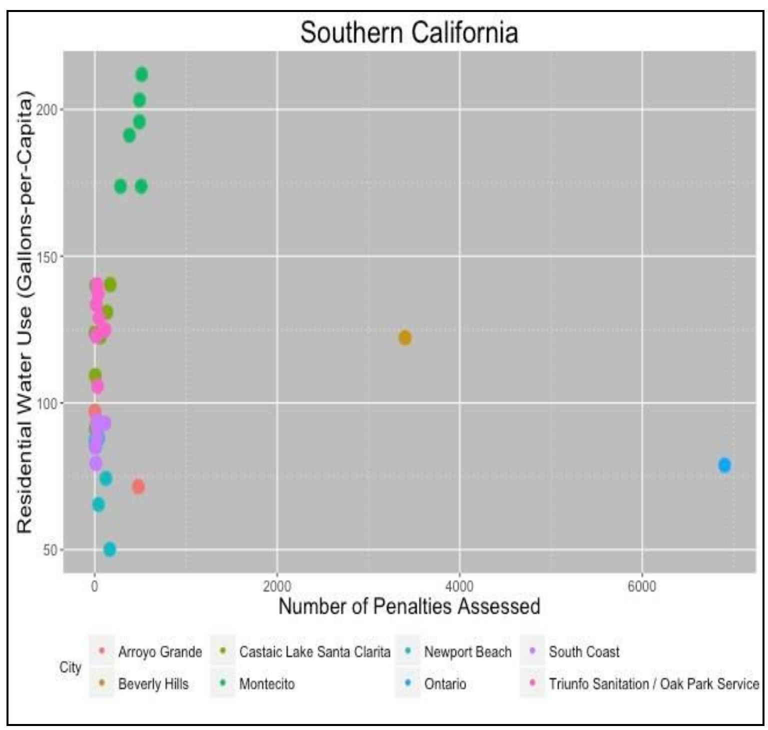


Figure 4.2.4



From these graphs, we do not see any clear behavioral pattern all cities follow which suggests that the response to penalties varies not only by region but also by city. However, we decided to run a regression to gain a better understanding of the impact. We chose to run the following fixed effect regression because we needed to control for each water supplier agency and for the seasonal variation. Here is the regression we applied to each region:

$$ResUse_{it} = \beta_1 Penalties\_Assessed_{it} + Supplier\_Name_i + Month_i + u_{it}, \text{ where}$$

$ResUse_{it}$ represents R-GPCD,
$Penalties\_Assessed_{it}$ represents the number of penalties received,
$Supplier\_Name_i$ is a fixed effect for each of 408 water supplier agencies
$Month_i$ is a categorical variable that ranges from 1-12 to reflect 12 months in a year, and
u_{it} is an error term

Our main independent variable was $Penalties\_Assessed_{it}$ and dependent variable was $ResUse_i$ because we were investigating whether or not the penalties impacted residential water usage.

Using the lfe and stargazer packages, we produced this table of our results:

Table 4.2.1:

	Dependent variable:		
	Residential Water Use(R-GPCD)		
	Northern CA	Central Valley	Southern CA
	(1)	(2)	(3)
Penalties Assessed	-0.017^{*}	0.005^{***}	0.009
	(0.008)	(0.002)	(0.019)
Observations	27	53	38
R^2	0.945	0.934	0.973
Adjusted R^2	0.891	0.910	0.957
Residual Std. Error	14.189 (df = 13)	11.005 (df = 38)	8.361 (df = 23)
Note:			*p<0.1;* ***p<0.05;*** p<0.01

Surprisingly, the results from the regressions differ from the inferences made from Figures 4.2.2-4.2.4. We see that the estimated coefficient of $Penalties\_Assessed_{it}$ is statistically significant for both Northern California and Central Valley. Since the coefficients are statistically significant, we are able to reject the null hypotheses in favor of the alternatives, and the estimated coefficients for number of penalties in both Northern California and Central Valley are statistically greater than zero. Furthermore, a coefficient of -0.017 for Northern California suggests that, holding all else constant, there was a negative relationship between penalties assessed and residential water usage for that region. On the other hand, a coefficient of 0.005 for Central Valley suggests that there was a positive relationship between penalties assessed and residential water usage for that region.

However, although these coefficients are statistically significant, it is important to note that the magnitude of each value is very small and only suggest a small relationship between number of penalties and residential water usage. Thus, we wanted to conduct another analysis of residential water usage. Based on the available data, we assessed water supplier compliance with the mandatory conservation standards to see if residents changed their behavior after the introduction of conservation standards.

4.2.5 Assessing Water Supplier Compliance to Conservation Standards

To determine if a water supplier is on track for meeting their conservation standard, the State Water Board compares the supplier's reported total production in 2015 with the supplier's total production for the same month in 2013. The compliance period conservation begins on June 1, 2015 and ends on February 29, 2016. Thus from the same data set, we selected water supplier, potable water production in 2013, potable water production in 2015, conservation standard and date variables.

The data set provided the expected conservation standard but we created a new variable that represented the actual conservation by calculating the percent change between 2013 and 2015 production, as well as a variable for the water usage prior to June 2015. We began our analysis, by running the following the regression:

$$Consumption\_Change = \beta_0 + \beta_1 Standard + pre\_June + u, \text{ where}$$

Consumption_Change represents the actual percent change between 2013 and 2015 production,
Standard represents the assigned conservation standard, and
pre_June represents the residential water usage before from before June 2015, and
u is an error term.

The results are shown in Table 4.2.2. The reasoning behind this regression is that we wanted to see if there was a correlation between the assigned conservation standards and the actual reduction percentage. We controlled for pre-June 2015 water usage in order to account for the fact that different water supplier agencies have different capacities to conserve. To put differently, water supplier agencies with higher usage before June 2015 may be able to conserve water more easily than agencies that already conserved a lot of water. The most important item we were concerned about in Table 4.2.2 was the R^2 , since the R^2 is the coefficient of determination and its square-root gives us the correlation coefficient, r. The square-root of .063 is 0.250998, thus $r = .251$. This result suggests that there is a slight positive relationship between the assigned conservation standard and the actual reduction percentage. Since this result appears to be relatively small, we could not make substantial conclusions regarding how the assigned conservation standards affect the actual reduction percentage.

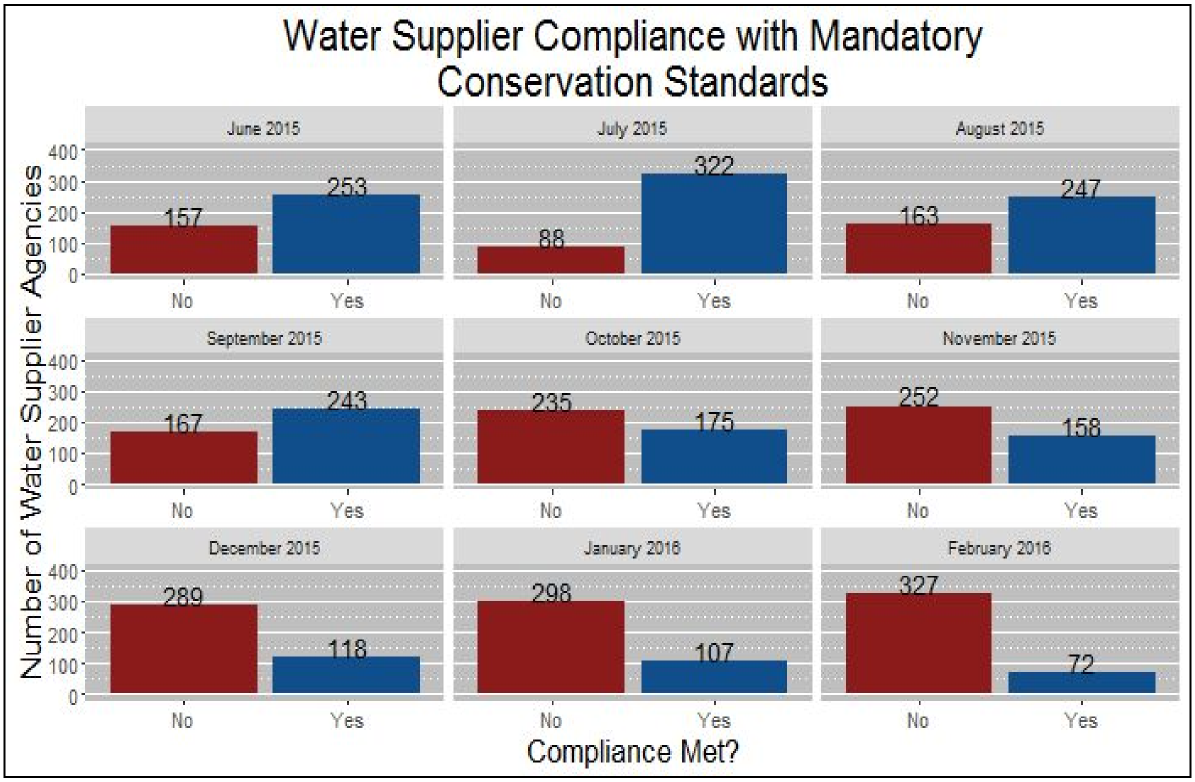
Table 4.2.2:

	Dependent variable:
	Percent Conserved
Assigned Percentage	0.343***
	(0.022)
Pre June 2015 Production	0.000***
	(0.000)
Constant	14.125***
	(0.585)
Observations	3,671
R^2	0.063
Adjusted R^2	0.062
Residual Std. Error	11.441 (df = 3668)
F Statistic	123.152*** (df = 2; 3668)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

To further enhance our analysis on conservation standards, we generated a new variable called *Met* that was a dummy variable for compliance. If the percent change was greater than the conservation standard, we considered that supplier to have complied to the standards and their *Met* value would be 1 and 0 otherwise. Afterwards, by summarizing the total number of cases corresponding with our dummy variable, we were able to retrieve a count for total number of water supplier agencies who complied.

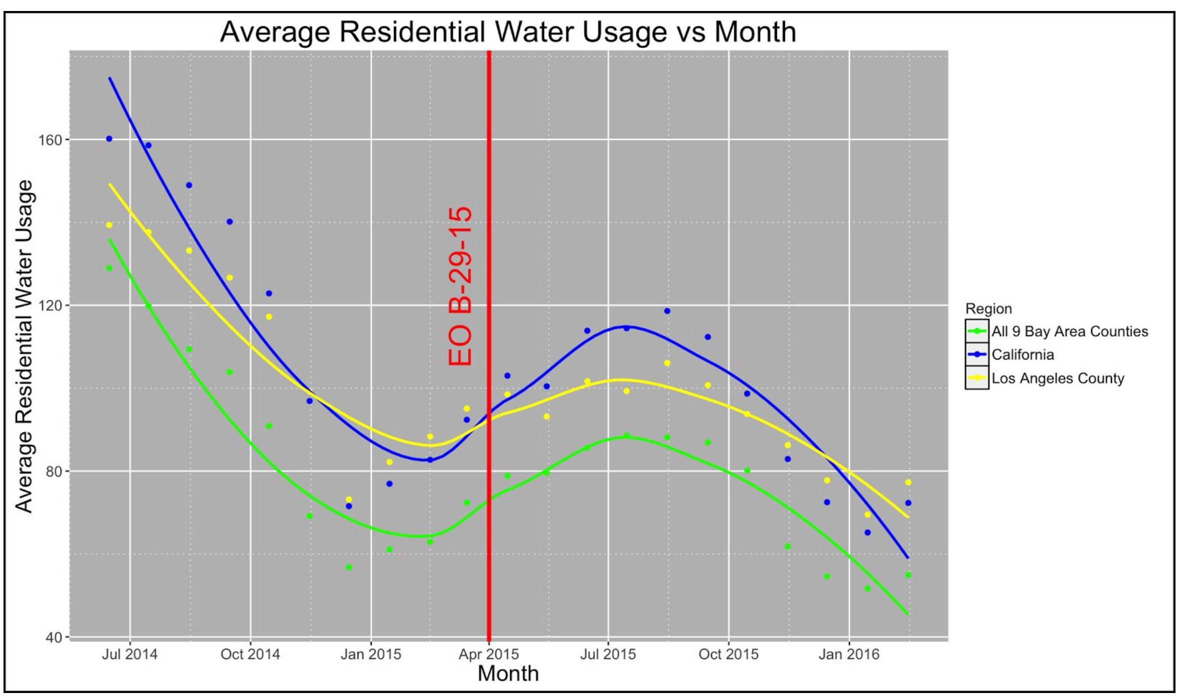
Our bar graph in Figure 4.2.2 shows how the majority of suppliers have complied, but from October 2015 to February 2016, the majority of suppliers have not met their conservation standard. This suggests that suppliers may have thought that the drought ended by October 2015. However, as we saw in our drought severity map (Figure 3.2.1) that is clearly not the case.

Figure 4.2.2:



4.3 Case Study: Who is doing their part?

Figure 4.3.1:



We wanted to directly compare our own county – Alameda County – with another major county in California, but since Alameda County is much smaller than Los Angeles County we decided to combine all of the Bay Area Counties to have a larger sample set to work from. As an extension of our examination of how residential consumption patterns changed after the statewide mandatory restrictions by Executive Order B-29-15, we wanted to explore whether geographical location made a difference in the change of water usage over time. We hypothesized that the type of person living in the Bay Area (simply from anecdotal evidence) would have lower water usage than the average resident living in Los Angeles. Once again we are using R-GPCD here, but since we had weekly values we took the average for each month for the plot. We also have overall California averages on our plot so that we can easily compare the two regions we picked to the state average.

Based on Figure 4.3.1, it seems as though in the last 1.5 years, Bay Area residents use less water than the average Californian, as well as less than the average Los Angeles County resident. Los Angeles County residents look to be on par with the state average, with a lower usage for both July of 2014 and of 2015 compared to the state. The red vertical line marks the executive order, and with only one full year to compare it does look as though overall usage for all regions we wanted to follow decreased after the executive order, however, with limited data we cannot definitively say that it is due to the executive order. Although, we can conclude that the average resident in the Bay Area does use less water overall when compared to the average resident in Los Angeles County.

5. Conclusion

With access to the National Drought Mitigation Center's *United States Drought Monitor* and the California State Water Board data sets we were able to create maps to easily visualize exactly where the drought is severe, how severe, the exact number of how many counties are affected, and how residential water usage changed over time with new information.

After calculating the weighted averages for drought severity levels for each county over time, we see that only a few counties currently have no measured level of drought. However, as the overview in Figure 3.1.1 suggests, drought severity levels in California are quick to change, and history is no indication of future levels. Regarding the impact of the average Californian in light of the drought, our regression results in *Section 4.1.3* suggest that overall residential water usage declined by about 20 percent, which is close to Governor Brown's mandate of a 25 percent reduction of water consumption. So moving away from higher maintenance plants, taking shorter showers, and other small actions really do add up.

Further studies should follow the current plans to address this issue in California by the "Central Valley Project and State Water Project 2016 Drought Contingency Plan For Water Project Operations." So how much longer will California be in a drought? We have no idea, because more research and more resources need to be devoted to solving the issue of a reliable water supply for California.

Appendix:

Sources:

"California Drought, Visualized with Open Data." *Center for Integrated Data Analytics.* U.S. Geological Survey, 17 May 2015. Web. 1 May 2016.

"Central Valley Project and State Water Project 2016 Drought Contingency Plan For Water Project Operations February - November 2016." *California Department of Water Resources.* State of California, 15 January 2016. Web. 1 May 2016.

"Final DWR Snow Survey: Indications of Good Water Supply for Agriculture." *Western Farm Press*. Penton Agriculture Market, 2 May 2011. Web. 1 May 2016.

"Governor Brown Declares Drought State of Emergency." *Office of Governor Edmund G. Brown Jr.* State of California, 17 January 2014. Web. 1 May 2016.

"Instructions for Estimating Residential Gallons Per Capita Day (R-GPCD) in Completing Monthly Urban Water Supplier Report." *California Water Boards.* State of California, n.d. PDF. 1 May 2016.

"Mandatory Water Conservation Regulation Go Into Effect." *California Water Boards.* State of California, 29 July 2014. PDF. 1 May 2016.

Nagourney, Adam. "California Imposes First Mandatory Water Restrictions to Deal With Drought." *The New York Times*. The New York Times, 1 April 2015. Web. 1 May 2016.

"Notice of Approval of Emergency Regulatory Action." *California Water Boards.* State of California, 28 July 2014. PDF. 1 May 2016.

Renwick, M. E., and R. D. Green (2000), Do residential water demand side management policies measure up? An analysis of eight California water agencies, *J. Environ. Econ. Manage.*, 40(1), 37–55, doi:10.1006/jeem.1999.1102

"State Reduces Water Use by Nearly 29 Percent in Advance of June Conservation Mandates." *California Water Boards.* State of California, 1 July 2015. PDF. 1 May 2016.

"Statement of Lester A. Snow Director, California Department of Water Resources Before the U. S. House of Representatives Committee on Resources Subcommittee on Water and Power Concerning 'Perspectives on California Water Supply: Challenges and Opportunities'." *California Department of Water Resources.* State of California, 25 January 2010. Web. 1 May 2016.

"Urban Water Suppliers Report Progress on Conservation Measures." *California Water Boards.* State of California, 9 September 2014. PDF. 1 May 2016.

R Code:

3.1.1

```
#To create overview maps of drought in california
#Site had a download file option so just load file and select relevant columns
timelapse <- "final_COUNTY2000-2016.csv"
timelapse <- timelapse  %>%
  read.file() %>%
  select(releaseDate, county, FIPS, NONE, D0, D1, D2, D3, D4)

bins <- timelapse %>% mutate(D0=D0-D1, D1=D1-D2, D2=D2-D3, D3=D3-D4)
bins$releaseDate <- as.Date(bins$releaseDate)
bins$month <- month(ymd(bins$releaseDate)+ months(0:11), label = TRUE, abbr = TRUE)
bins$year <- year(ymd(bins$releaseDate))

#Weighted average
bins <- bins %>% mutate( FIPS_avg = (1*D0 + 2*D1 + 3*D2 + 4*D3 + 5*D4) /100 - 1) %>% group_by(month,
year, FIPS, county) %>% summarise(monthly_avg = round(mean(FIPS_avg)))

#Cleaning up the county names to make it easier to join with CA county coordinates later
bins$county <- gsub(" County", "", bins$county)
bins$county <- sapply(bins$county, tolower)
names(bins)[names(bins) == "county"] <- "subregion"
bins<- data.frame(bins)

drought_levels <- function(x) {
        (
         if (x==-1) {
           "No Drought (NONE)"
         } else if (x==0) {
           "Abnormally Dry (D0)"
         } else if (x==1) {
           "Moderate Drought (D1)"
         } else if (x==2) {
           "Severe Drought (D2)"
         } else if (x==3) {
           "Extreme Drought (D3)"
         } else
           "Exceptional Drought (D4)"
        )
}

bins <- bins %>% mutate(monthly_avg=sapply(monthly_avg,drought_levels))

#Coordinates for the california counties
CAcounties <- map_data('county')%>%
  filter(region == "california")
CAcounties <- data.frame(CAcounties)

#Joining county coordinates with the data we have
binscounties <- bins%>%
  right_join(CAcounties, by = c("subregion"))
```

```
#Plot manipulation
binscounties$monthly_avg <- factor(binscounties$monthly_avg, levels = c("No Drought (NONE)", "Abnormally Dry (D0)", "Moderate Drought (D1)", "Severe Drought (D2)", "Extreme Drought (D3)", "Exceptional Drought (D4)"))

#Creating new date column
binscounties <- binscounties %>%
  group_by(year)%>%
  arrange(month)%>%
  mutate(completedate = (paste(month, year)))

#Create levels for all months/years
months <- c("Jan", "Feb", "Mar", "Apr", "May","Jun", "Jul", "Aug","Sep", "Oct", "Nov","Dec")
months <- rep(months, times = 17)
years <- rep(2000:2016, each = 12)
monthsyears <- paste(months, years)
monthsyears <- monthsyears[1:(length(monthsyears) - 8)]
binscounties$completedate <- factor(levels = c(monthsyears), binscounties$completedate)
binscounties$completedate <- as.factor(binscounties$completedate)
binscounties <- data.frame(binscounties)

#Figure 3.1.1
{r, fig.width=14, fig.height=24}
binscounties%>%
  na.omit()%>%
  ggplot() +
  geom_polygon(aes (x = long, y = lat, group = group, fill = monthly_avg), colour = "white", size = 0.02)+
  facet_wrap(~completedate, ncol = 12, drop = TRUE) +
  labs(title = "Drought Severity in California from 2000 to Present") +
  theme(axis.line=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_rect(colour = "black"),
        plot.background = element_rect(colour = "grey"),
        panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        panel.grid.minor = element_line(linetype = "dotted"),
        axis.title = element_blank(),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2))
        ) +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3", "firebrick4"))
```

3.1.2

```
#Selected data
timelapse <- "final_COUNTY2000-2016.csv"
timelapse <- timelapse %>%
  read.file() %>%
  select(releaseDate, FIPS, NONE, D0, D1, D2, D3, D4)
```

```
#Data manipulation
bins <- timelapse %>% mutate(D0=D0-D1, D1=D1-D2, D2=D2-D3, D3=D3-D4, releaseDate =
year(ymd(releaseDate)))
names(bins)[1] <- "Year"

#Weighted average
bins <- bins %>% mutate( FIPS_avg = (1*D0 + 2*D1 + 3*D2 + 4*D3 + 5*D4) /100 - 1) %>% group_by(Year,
FIPS) %>% summarise(yearly_avg = round(mean(FIPS_avg)))

#Renaming
drought_levels <- function(x) {
        (
         if (x==-1) {
           "No Drought (NONE)"
         } else if (x==0) {
           "Abnormally Dry (D0)"
         } else if (x==1) {
           "Moderate Drought (D1)"
         } else if (x==2) {
           "Severe Drought (D2)"
         } else if (x==3) {
           "Extreme Drought (D3)"
         } else
           "Exceptional Drought (D4)"
        )
}
bins <- bins %>% mutate(yearly_avg=sapply(yearly_avg,drought_levels))

#Plot manipulation
bins$yearly_avg <- factor(bins$yearly_avg, levels = c("No Drought (NONE)", "Abnormally Dry (D0)",
"Moderate Drought (D1)", "Severe Drought (D2)", "Extreme Drought (D3)", "Exceptional Drought (D4)"))

bins <- bins %>% select(Year, yearly_avg)  %>% arrange(Year)

#Figure 3.1.2
bins %>% ggplot(aes(x=Year)) + geom_bar(aes(color=Year, fill= yearly_avg)) +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3",
"firebrick4")) + facet_wrap(~ yearly_avg) + ggtitle("Number of Counties in Each Drought Level") +
ylab("Number of Counties") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + theme(panel.background
= element_rect(fill = "grey"),
       panel.background = element_rect(color = "black"),
       panel.grid.minor = element_line(linetype = "dotted"),
       axis.title = element_text(size = rel(1.5)),
       axis.text = element_text(size = rel(1.0)),
       legend.text = element_text(size = rel(1.0)),
       plot.title = element_text(size = rel(2)),
       panel.grid.major.x = element_blank(),
       panel.grid.minor.x = element_blank())+ guides(fill=F)
```

3.1.3

```
#Stacking the results from 3.1.2
info <- bins %>% group_by(Year, yearly_avg) %>% summarise(total = n())
```

```
#Figure 3.1.3
info %>% ggplot(aes(x=Year, y=total, fill=yearly_avg)) + geom_bar(stat="identity", alpha=0.9) +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3",
"firebrick4")) + ggtitle("Drought Levels from 2000-2016") + ylab("Number of Counties") + theme(axis.text.x =
element_text(angle = 45, hjust = 1)) + theme(panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        panel.grid.minor = element_line(linetype = "dotted"),
        axis.title = element_text(size = rel(1.5)),
        axis.text = element_text(size = rel(1.0)),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2)),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank())
```

3.2.1

```
#Figure 3.2.1
#Map from 2006-2010 to capture the previous drought in California
binscounties%>%
  na.omit()%>%
  filter(year >= 2006, year <= 2010)%>%
    ggplot() +
  geom_polygon(aes (x = long, y = lat, group = group, fill = monthly_avg), colour = "white", size = 0.02)+
  facet_wrap(~completedate, ncol = 12, drop = TRUE) +
  labs(title = "Drought Severity in California from 2006 to 2010") +
  theme(axis.line=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_rect(colour = "black"),
        plot.background = element_rect(colour = "grey"),
        panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        axis.title = element_blank(),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2))
        ) +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3",
"firebrick4"))
```

3.2.2

```
#Figure 3.2.2
#Map from 2011 to Present to show current drought:
{r, fig.width=14, fig.height=8}
binscounties%>%
  na.omit()%>%
  filter(year >= 2011)%>%
  ggplot() +
  geom_polygon(aes (x = long, y = lat, group = group, fill = monthly_avg), colour = "white", size = 0.02)+
  facet_wrap(~completedate, ncol = 12, drop = TRUE) +
  labs(title = "Drought Severity in California from 2011 to Present") +
  theme(axis.line=element_blank(),
```

```
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.key = element_rect(colour = "black"),
        plot.background = element_rect(colour = "grey"),
        panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        panel.grid.minor = element_line(linetype = "dotted"),
        axis.title = element_blank(),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2))
        ) +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3",
"firebrick4"))
```

3.2.3

```
#Selected data
timelapse <- "final_COUNTY2000-2016.csv"
timelapse <- timelapse  %>%
  read.file() %>%
  select(releaseDate, FIPS, NONE, D0, D1, D2, D3, D4)

#Data manipulation
bins <- timelapse %>% mutate(D0=D0-D1, D1=D1-D2, D2=D2-D3, D3=D3-D4, releaseDate =
floor_date(lubridate::ymd(releaseDate), "month")) %>% filter(year(releaseDate) > 2010)

names(bins)[1] <- "Month"

#Weighted average
bins <- bins %>% mutate( FIPS_avg = (1*D0 + 2*D1 + 3*D2 + 4*D3 + 5*D4) /100 - 1) %>% group_by(Month,
FIPS) %>% summarise(monthly_avg = round(mean(FIPS_avg)))

#Renaming
drought_levels <- function(x) {
        (
         if (x==-1) {
           "No Drought (NONE)"
         } else if (x==0) {
           "Abnormally Dry (D0)"
         } else if (x==1) {
           "Moderate Drought (D1)"
         } else if (x==2) {
           "Severe Drought (D2)"
         } else if (x==3) {
           "Extreme Drought (D3)"
         } else
           "Exceptional Drought (D4)"
        )
}

bins <- bins %>% mutate(monthly_avg=sapply(monthly_avg,drought_levels))

#Plot manipulation
```

```
bins$monthly_avg <- factor(bins$monthly_avg, levels = c("No Drought (NONE)", "Abnormally Dry (D0)", "Moderate Drought (D1)", "Severe Drought (D2)", "Extreme Drought (D3)", "Exceptional Drought (D4)"))

bins <- bins %>% select(Month, monthly_avg)  %>% arrange(Month)

#Figure 3.2.3
bins %>% ggplot(aes(x=as.Date(Month), fill=monthly_avg)) + geom_bar() +
 scale_fill_manual("Drought Severity", values=c("#7baad8", "#fffc67", "orange", "darkorange3", "orangered3", "firebrick4")) + facet_wrap(~ monthly_avg) + ggtitle("Number of Counties in Each Drought Level") +
ylab("Number of Counties") +xlab("Month") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
theme(panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        panel.grid.minor = element_line(linetype = "dotted"),
        axis.title = element_text(size = rel(1.5)),
        axis.text = element_text(size = rel(1.0)),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2)),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank()) +
        scale_x_date(labels = date_format("%m/%Y")) +
  guides(fill=F)
```

4.1.1

```
#Initial data wrangling
water_residential <- "uw_supplier_data040416.csv"
water_residential <- water_residential  %>%
  read.file() %>%
  select(Supplier.Name, Stage.Invoked, Mandatory.Restrictions,
         Reporting.Month,
CALCULATED.R.GPCD.Reporting.Month..Values.calculated.by.Water.Board.staff.using.methodology.available.at.http...www.waterboards.ca.gov.waterrights.water_issues.programs.drought.docs.ws_tools.guidance_estimate_res_gpcd.pdf.,
         Hydrologic.Region,
         Penalties.Assessed,
         X..Residential.Use,
         Water.Days.Allowed.Week,
         Conservation.Standard..starting.in.June.2015.) %>%
  mutate(reporting_month = lubridate::mdy(Reporting.Month),
         executive_order = ifelse(reporting_month < as.numeric(ymd("2015-04-15")), 0, 1),
    month_dummies = ifelse(reporting_month %in% ymd(c("2014-06-15", "2015-06-15")), 6,
                   ifelse(reporting_month %in% ymd(c("2014-07-15", "2015-07-15")), 7,
                   ifelse(reporting_month %in% ymd(c("2014-08-15", "2015-08-15")), 8,
                   ifelse(reporting_month %in% ymd(c("2014-09-15", "2015-09-15")), 9,
                   ifelse(reporting_month %in% ymd(c("2014-10-15", "2015-10-15")), 10,
                   ifelse(reporting_month %in% ymd(c("2014-11-15", "2015-11-15")), 11,
                   ifelse(reporting_month %in% ymd(c("2015-12-15", "2016-12-15")), 12,
                   ifelse(reporting_month %in% ymd(c("2015-01-15","2016-01-15")), 1,
                   ifelse(reporting_month %in% ymd(c("2015-02-15","2016-02-15")), 2,
ifelse(reporting_month == ymd(c("2015-03-15")), 3,                     ifelse(reporting_month == ymd(c("2015-04-15")), 4, 5)))))))))))

#Renaming variable
names(water_residential)[names(water_residential) ==
'CALCULATED.R.GPCD.Reporting.Month..Values.calculated.by.Water.Board.staff.using.methodology.availabl
```

```
e.at.http...www.waterboards.ca.gov.waterrights.water_issues.programs.drought.docs.ws_tools.guidance_estimate_res_gpcd.pdf.'] <- 'resid_use'

#Figure 4.1.1
water_residential %>%
  ggplot(aes(x=reporting_month, y= resid_use)) +
  geom_point() +
  geom_jitter() +
  geom_vline(aes(xintercept = as.numeric(ymd("2015-04-01"))), color = "red", size=1.5) +
  xlab("Reported Months") +
  ylab("Residential Water Usage (Gallons-per-Capita-per-Day)") +
  ggtitle("2014-2016 Reported California Residential Water Usage") +
  geom_smooth() +
  theme(panel.background = element_rect(fill = "grey"),
        axis.title = element_text(size = rel(1.75)),
        axis.text = element_text(size = rel(1.25)),
        plot.title = element_text(size = rel(2.5))) +
  geom_text(label = "EO B-29-15", aes(x = as.Date(ymd("2015-03-15"))), y = 500, size = 8, colour = "red", angle = 90)
```

4.1.2

```
#Data wrangling for Figure 4.1.2
month_water_residential <- water_residential  %>%
  mutate(overtime_use = ifelse(resid_use <= 100.00, "0-100",
                 ifelse(resid_use <= 200.00, "101-200",                      ifelse(resid_use <= 300.00, "201-300",
                  ifelse(resid_use <= 400.00, "301-400",
                    "400+"))))) %>%
  group_by(overtime_use, reporting_month, month_dummies) %>%
  summarise(total = n())

#Figure 4.1.2

month_water_residential %>%
  ggplot(aes(x=reporting_month, y = total, group =  overtime_use, fill=overtime_use)) +
  geom_bar(stat="identity", position="identity") +
  scale_fill_manual(name="Residental Usage\n(Gallons-per-Capita-per-Day)",
                 values=c("#08519C", "#3182BD", "#6BAED6", "#6a51a3", "#ae017e"),
                 breaks=c("0-100", "101-200", "201-300", "301-400", "400+"),
                 labels=c("0-100", "101-200", "201-300","301-400", "400+")) +
  geom_vline(aes(xintercept = as.numeric(ymd("2015-04-01"))), color = "red", size=1.5) +
  geom_label_repel(
    data = subset(month_water_residential, reporting_month %in% ymd(c("2014-06-15", "2015-04-15","2016-02-15"))),
    aes(label = total),
    size = 6, color = "white",
    fontface = 'bold',
    box.padding = unit(0.10, "lines"),
    point.padding = unit(.25, "lines"),
    show.legend=FALSE) +
  theme(panel.background = element_rect(fill = "grey"),
        axis.title = element_text(size = rel(1.75)),
```

```
        axis.text = element_text(size = rel(1.25)),
        legend.text = element_text(size = rel(1.25)),
        plot.title = element_text(size = rel(2.0)),
        panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        legend.title = element_text(size = rel(1.5))) +
  ylab("Number of Water Supplier Agencies") +
  xlab("Reported Months") +
  ggtitle("2014-2016 Reported CA Residential Water Usage \nby Number of Water Supplier Agencies")  +
  geom_text(label = "EO B-29-15", aes(x = as.Date(ymd("2015-03-15"))), y = 325, size = 8, colour = "red", angle
= 90)
```

4.1.3

```
#Mean R-GPCD during all the months
mean(water_residential$resid_use)

#Statistical Analysis:
fixed <- felm(resid_use ~ executive_order | Supplier.Name + month_dummies, data=water_residential)

#Stargazer package to create Table 4.1.3
stargazer(fixed, type="html", omit = NULL, covariate.labels = c("Statewide Mandatory Restrictions"),
dep.var.labels=c("Residential Water Use (R-GPCD)"), out="models.htm")
```

4.2.1

```
# Initial Data Wrangling
usage <- "C:\\Users\\Megan\\Desktop\\usagedataset.csv"
table <- usage %>%
  read.file() %>%
  select(Supplier_Name,ResUse, Penalties_Assessed, Reporting_Month, Percent_ResUse,
Mandatory_Restrictions,ReportedProd, Prod13, CS) %>%
  filter( CS != "NULL") %>%
  mutate(date = lubridate::mdy(Reporting_Month),  norcal = ifelse(Supplier_Name %in% c("Pleasanton  City
of","Santa Cruz  City of","East Bay Municipal Utilities District", "Yuba City  City of", "Hillsborough  Town of",
"Sacramento Suburban Water District", "Olivehurst Public Utility District", "Alco Water Service"), 1,0),  socal=
ifelse(Supplier_Name %in% c("Newport Beach  City of", "Beverly Hills  City of", "Castaic Lake Water Agency
Santa Clarita Water Division", "Montecito Water District", "Ontario  City of", "Triunfo Sanitation District / Oak
Park Water Service",  "Arroyo Grande  City of" , "South Coast Water District"), 1,0),
cv = ifelse(Supplier_Name %in% c("Clovis  City of",  "Fresno  City of" ,  "Ceres  City of", "Hanford  City of",
"Tulare, City of", "Turlock  City of ", "Los Banos, City of", "Exeter  City of"), 1, 0),
Month = month(date))
#Wrangling to prepare data frame for Figure 4.2.1
C <- c("Pleasanton  City of", "Santa Cruz  City of", "East Bay Municipal Utilities District","Yuba City  City of",
"Hillsborough  Town of", "Sacramento Suburban Water District", "Olivehurst Public Utility District", "Alco
Water Service")
CV <- c("Clovis  City of", "Fresno  City of", "Ceres  City of", "Hanford  City of", "Tulare, City of", "Turlock
City of", "Los Banos, City of", "Exeter  City of")
SC<- c("Newport Beach  City of", "Beverly Hills  City of","Castaic Lake Water Agency Santa Clarita Water
Division" ,"Montecito Water District", "Ontario  City of", "Triunfo Sanitation District / Oak Park Water Service",
"Arroyo Grande  City of", "South Coast Water District")

m2 <- table %>%
  group_by(Penalties_Assessed, Supplier_Name) %>%
```

```
  summarise(avg = mean(ResUse)) %>%
  filter(Supplier_Name %in% NC | Supplier_Name %in% CV |Supplier_Name %in% SC) %>%
  mutate(Region =
         ifelse(Supplier_Name %in% c("Pleasanton  City of","Santa Cruz  City of","East Bay Municipal Utilities District", "Yuba City  City of", "Hillsborough  Town of", "Sacramento Suburban Water District", "Olivehurst Public Utility District", "Alco Water Service"), "Northern California", ifelse(Supplier_Name %in% c("Newport Beach  City of", "Beverly Hills  City of", "Castaic Lake Water Agency Santa Clarita Water Division", "Montecito Water District", "Ontario  City of", "Triunfo Sanitation District / Oak Park Water Service",  "Arroyo Grande  City of" , "South Coast Water District"), "Southern California", ifelse(Supplier_Name %in% c("Clovis  City of", "Fresno  City of" ,  "Ceres  City of", "Hanford  City of",  "Tulare, City of", "Turlock  City of ", "Los Banos, City of", "Exeter  City of"), "Central Valley","Central Valley")))) %>%
  filter(Penalties_Assessed >= 50)

#Figure 4.2.1
plot2 <- m2 %>%
   ggplot(aes(x=Penalties_Assessed, y=avg)) + geom_point(aes(color=Region, size = 0.01)) + labs(x= "Number of Penalties Assessed", y= "Residential Water Use (Gallons-per-Capita-per-Day)", title= "Number of CA Penalties & Water Usage by Region") +  theme(panel.background = element_rect(fill = "grey"),
    panel.background = element_rect(color = "black"),
    panel.grid.minor = element_line(linetype = "dotted"),
    axis.title = element_text(size = rel(1.5)),
    axis.text = element_text(size = rel(1.0)),
    legend.text = element_text(size = rel(1.0)),
    plot.title = element_text(size = rel(2)))
plot2 + guides(size=FALSE)

#4.2.2
NC <- c("Pleasanton  City of", "Santa Cruz  City of", "East Bay Municipal Utilities District","Yuba City  City of", "Hillsborough  Town of", "Sacramento Suburban Water District", "Olivehurst Public Utility District", "Alco Water Service")

NCT <- table %>%
  filter(Supplier_Name %in% NC) %>%
  mutate(Year= year(date), Month = month(date)) %>%
  filter (Year == 2015) %>%
  filter (Penalties_Assessed > 50)

NCT$Supplier_Name <- gsub(" City of", "", NCT$Supplier_Name)
NCT$Supplier_Name <- gsub(" Town of", "", NCT$Supplier_Name)

#Figure 4.2.2
NCplot <- NCT %>%
  ggplot(aes(x=Penalties_Assessed, y=ResUse)) + geom_point(aes(color=Supplier_Name, size = 1)) + labs(title= "Northern California", x= "Number of Penalties Assessed", y= "Residential Water Use (Gallons-per-Capita-per-Day)", color = "City") +  theme(panel.background = element_rect(fill = "grey"),
    panel.background = element_rect(color = "black"),
    axis.title = element_text(size = rel(1.5)),
    axis.text = element_text(size = rel(1.0)),
    legend.text = element_text(size = rel(1.0)),
    plot.title = element_text(size = rel(2)))

NCplot + guides(size=FALSE)
```

```
#4.2.3
CV <- c("Clovis  City of", "Fresno  City of", "Ceres  City of", "Hanford  City of", "Tulare, City of", "Turlock
City of", "Los Banos, City of", "Exeter  City of")

CVT <- table %>%
  filter(Supplier_Name %in% CV) %>%
  mutate(Year= year(date)) %>%
  mutate(Month = month(date)) %>%
  filter (Year == 2015) %>%
  filter (Penalties_Assessed > 50)

CVT$Supplier_Name <- gsub(" City of", "", CVT$Supplier_Name)
CVT$Supplier_Name <- gsub(",", "", CVT$Supplier_Name)

# Figure 4.2.3
CVplot <- CVT %>%
  ggplot(aes(x=Penalties_Assessed, y=ResUse)) + geom_point(aes(color=Supplier_Name, size =1)) + labs(title=
"The Central Valley", x= "Number of Penalties Assessed", y= "Residential Water Use
(Gallons-per-Capita-per-Day)", color = "City") +
 theme(panel.background = element_rect(fill = "grey"),
    panel.background = element_rect(color = "black"),
    panel.grid.minor = element_line(linetype = "dotted"),
    axis.title = element_text(size = rel(1.5)),
    axis.text = element_text(size = rel(1.0)),
    legend.text = element_text(size = rel(1.0)),
    plot.title = element_text(size = rel(2)))

CVplot + guides(size=FALSE)

#4.2.4
SC<- c("Newport Beach  City of", "Beverly Hills  City of","Castaic Lake Water Agency Santa Clarita Water
Division" ,"Montecito Water District", "Ontario  City of", "Triunfo Sanitation District / Oak Park Water Service",
"Arroyo Grande  City of", "South Coast Water District")

SCT <- table %>%
  filter(Supplier_Name %in% SC) %>%
  mutate(Year= year(date)) %>%
  mutate(Month = month(date)) %>%
  filter (Year == 2015) %>%
  filter (Penalties_Assessed > 50)

SCT$Supplier_Name <- gsub(" City of", "", SCT$Supplier_Name)

# Figure 4.2.4
SCplot <- SCT %>%
  ggplot(aes(x=Penalties_Assessed, y=ResUse)) + geom_point(aes(color=Supplier_Name, size=1)) + labs(title=
"Southern California", x= "Number of Penalties Assessed", y= "Residential Water Use
(Gallons-per-Capita-per-Day)", color = "City") +  theme(panel.background = element_rect(fill = "grey"),
    panel.background = element_rect(color = "black"),
    panel.grid.minor = element_line(linetype = "dotted"),
    axis.title = element_text(size = rel(1.5)),
    axis.text = element_text(size = rel(1.0)),
    legend.text = element_text(size = rel(1.0)),
```

```
   plot.title = element_text(size = rel(2)))

SCplot + guides(size=FALSE)

#Regressions
fixed <- felm(ResUse ~ Penalties_Assessed | Supplier_Name + Month, data=NCT)
fixed1 <- felm(ResUse ~ Penalties_Assessed | Supplier_Name + Month, data=CVT)
fixed2 <- felm(ResUse ~ Penalties_Assessed | Supplier_Name + Month, data=SCT)

#Table 4.2.1:
stargazer(fixed, fixed1, fixed2, type="html", column.labels = c("Northern CA", "Central Valley", "Southern
CA"),
 dep.var.labels=c("Residential Water Use(R-GPCD)"),
 covariate.labels=c("Penalties Assessed","Water Supplier Agency Name"), out="models.htm")
```

4.2.5

```
table <- table %>%
  mutate(I_Prod = as.numeric(gsub(",", "", table$Prod13)),F_Prod = as.numeric(gsub(",", "",
table$ReportedProd)),
Standard = as.numeric(gsub("%", "", table$CS)))

CONS <- table %>%
  select(Supplier_Name, I_Prod, F_Prod, Standard, date) %>%
  mutate(Year= year(date)) %>%
  filter(Year == 2015 | Year == 2016) %>%
  filter(Standard != "NULL") %>%
  mutate(Month= month(date)) %>%
  mutate(Change = ((I_Prod-F_Prod)/(I_Prod)) *100,
         pc= ifelse(Change > Standard, 1, 0),
         met =ifelse(pc == 1, "Met", "Did not Meet")) %>%
  group_by(met, date) %>%
  summarise(total = n())

#Table 4.2.2
linear <- lm(Consumption_Change~Standard + pre_June,data = REGC)

stargazer(linear, type="html",
 dep.var.labels=c("Percent Conserved"),
 covariate.labels=c("Assigned Percentage","Pre June 2015 Production"), out="models.htm")

sqrt(.063)

#Creating Figure 4.2.5
month_named <- c(
  '2015-06-15'= "June 2015",
  '2015-07-15'= "July 2015",
  '2015-08-15'= "August 2015",
  '2015-09-15'= "September 2015",
  '2015-10-15'= "October 2015",
  '2015-11-15'= "November 2015",
  '2015-12-15'= "December 2015",
```

```
  '2016-01-15'= "January 2016",
  '2016-02-15'= "February 2016")

#Figure 4.2.5
ConP <- CONS %>%
  ggplot(aes(x=met, y=total, fill=met)) + geom_bar(stat="identity") + facet_wrap( ~ date, labeller =
as_labeller(month_named), ncol=3, scales= "free_x") + scale_x_discrete(labels=c("No", "Yes")) + labs(title=
"Water Supplier Compliance with Mandatory
Conservation Standards") +   ylab("Number of Water Supplier Agencies") + xlab("Compliance Met?") +
scale_fill_manual("Compliance", values=c("firebrick4","dodgerblue4")) +
geom_text(aes(label=total), vjust=-.01, color='black', size=5) +  ylim(0,400) +
theme(panel.background =  element_rect(fill = "grey"),
    panel.background = element_rect(color = "black"),
    panel.grid.minor = element_line(linetype = "dotted"),
    axis.title = element_text(size = rel(1.5)),
    axis.text = element_text(size = rel(1.0)),
    legend.text = element_text(size = rel(1.0)),
    plot.title = element_text(size = rel(2)),
    legend.position="none",
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank())
```

4.3.1

```
#Loading data set and selecting the columns we want
water_residential <- "uw_supplier_data040416.csv"
new_water_residential <- water_residential  %>%
  read.file() %>%
  select(Supplier.Name, Stage.Invoked, Mandatory.Restrictions,
        Reporting.Month,
CALCULATED.R.GPCD.Reporting.Month..Values.calculated.by.Water.Board.staff.using.methodology.available
.at.http...www.waterboards.ca.gov.waterrights.water_issues.programs.drought.docs.ws_tools.guidance_estimate_r
es_gpcd.pdf., Hydrologic.Region, Penalties.Assessed, X..Residential.Use)

names(new_water_residential)[names(new_water_residential) ==
'CALCULATED.R.GPCD.Reporting.Month..Values.calculated.by.Water.Board.staff.using.methodology.availabl
e.at.http...www.waterboards.ca.gov.waterrights.water_issues.programs.drought.docs.ws_tools.guidance_estimate_
res_gpcd.pdf.'] <- 'resid_use'

new_water_residential$Reporting.Month <- new_water_residential$Reporting.Month%>%
  mdy()

#Los angeles water suppliers
#Want all of los angeles county so any water supplier that matches the cities within LA county
#Some gsub to get this nice list of the cities in LA county

lacities <- "Agoura Hills Alhambra Arcadia Artesia Avalon Azusa Baldwin Park Bell Bell Gardens Bellflower
Beverly Hills Bradbury Burbank Calabasas Carson Cerritos Claremont Commerce Compton Covina Cudahy
Culver City Diamond Bar Downey Duarte El Monte El Segundo Gardena Glendale Glendora Hawaiian Gardens
Hawthorne Hermosa Beach Hidden Hills Huntington Park Industry Inglewood Irwindale La Cañada Flintridge La
Habra Heights La Mirada La Puente La Verne Lakewood Lancaster Lawndale Lomita Long Beach Los Angeles
Lynwood Malibu Manhattan Beach Maywood Monrovia Montebello Monterey Park Norwalk Palmdale Palos
Verdes Estates Paramount Pasadena Pico Rivera Pomona Rancho Palos Verdes Redondo Beach Rolling Hills
```

```
Rolling Hills Estates Rosemead San Dimas San Fernando San Gabriel San Marino Santa Clarita Santa Fe Springs
Santa Monica Sierra Madre Signal Hill South El Monte South Gate South Pasadena Temple City Torrance
Vernon Walnut West Covina West Hollywood Westlake Village Whittier"
lacities <- gsub(" " , ")|(" , lacities)

#NOT SHOWN:then manually put a . in between cities with more than one word in the name to account for
spaces
#Below is the data set that only has los angeles county levels of water usage

losangeles.avg <- new_water_residential%>%
filter(grepl("(Agoura.Hills)|(Alhambra)|(Arcadia)|(Artesia)|(Avalon)|(Azusa)|(Baldwin.Park)|(Bell)|(Bell.Gardens)
|(Bellflower)|(Beverly.Hills)|(Bradbury)|(Burbank)|(Calabasas)|(Carson)|(Cerritos)|(Claremont)|(Commerce)|(Com
pton)|(Covina)|(Cudahy)|(Culver.City)|(Diamond.Bar)|(Downey)|(Duarte)|(El.Monte)|(El.Segundo)|(Gardena)|(Gl
endale)|(Glendora)|(Hawaiian)|(Gardens)|(Hawthorne)|(Hermosa.Beach)|(Hidden.Hills)|(Huntington.Park)|(Industr
y)|(Inglewood)|(Irwindale)|(La.Cañada.Flintridge)|(La.Habra.Heights)|(La.Mirada)|(La.Puente)|(La.Verne)|(Lakew
ood)|(Lancaster)|(Lawndale)|(Lomita)|(Long.Beach)|(Los.Angeles)|(Lynwood)|(Malibu)|(Manhattan.Beach)|(May
wood)|(Monrovia)|(Montebello)|(Monterey.Park)|(Norwalk)|(Palmdale)|(Palos.Verdes.Estates)|(Paramount)|(Pasa
dena)|(Pico.Rivera)|(Pomona)|(Rancho)|(Palos.Verdes)|(Redondo.Beach)|(Rolling.Hills)|(Rolling.Hills.Estates)|(R
osemead)|(San.Dimas)|(San.Fernando)|(San.Gabriel)|(San.Marino)|(Santa.Clarita)|(Santa.Fe.Springs)|(Santa.Moni
ca)|(Sierra.Madre)|(Signal.Hill)|(South.El.Monte)|(South.Gate)|(South.Pasadena)|(Temple.City)|(Torrance)|(Verno
n)|(Walnut)|(West.Covina)|(West.Hollywood)|(Westlake)|(Village)|(Whittier)", Supplier.Name))%>%
  group_by(Reporting.Month)%>%
  summarise(la_use_avg = mean(resid_use))

#There are multiple counties surrounding the bay: SF, Marin, Sonoma, Napa, Solano, Contra Costa, Alameda,
Santa Clara, and San Mateo. Eastbay
#Copied list from website because there wasn't a nice list that was vector ready

sfbay <- "Alameda, California
+ Albany, California
+ American Canyon, California
+ Antioch, California
+ Atherton, California
+ B
+ Belmont, California
+ Belvedere, California
+ Benicia, California
+ Berkeley, California
+ Brentwood, California
+ Brisbane, California
+ Burlingame, California
+ C
+ Calistoga, California
+ Campbell, California
+ Clayton, California
+ Cloverdale, California
+ Colma, California
+ Concord, California
+ Corte Madera, California
+ Cotati, California
+ Cupertino, California
+ D
+ Daly City, California
```

+ Danville, California
+ Dixon, California
+ Dublin, California
+ E
+ East Palo Alto, California
+ El Cerrito, California
+ Emeryville, California
+ F
+ Fairfax, California
+ Foster City, California
+ Fremont, California
+ G
+ Gilroy, California
+ H
+ Half Moon Bay, California
+ Hayward, California
+ Healdsburg, California
+ Hercules, California
+ Hillsborough, California
+ L
+ Lafayette, California
+ Larkspur, California
+ Livermore, California
+ Los Altos, California
+ Los Altos Hills, California
+ Los Gatos, California
+ M
+ Martinez, California
+ Menlo Park, California
+ Mill Valley, California
+ Millbrae, California
+ Milpitas, California
+ Monte Sereno, California
+ Moraga, California
+ Morgan Hill, California
+ Mountain View, California
+ N
+ Napa, California
+ Newark, California
+ Novato, California
+ O
+ Oakland, California
+ Oakley, California
+ Orinda, California
+ P
+ Pacifica, California
+ Palo Alto, California
+ Petaluma, California
+ Piedmont, California
+ Pinole, California
+ Pittsburg, California
+ Pleasant Hill, California
+ Pleasanton, California

```
+ Portola Valley, California
+ R
+ Redwood City, California
+ Richmond, California
+ Rio Vista, California
+ Rohnert Park, California
+ Ross, California
+ S
+ St. Helena, California
+ San Anselmo, California
+ San Carlos, California
+ San Francisco
+ San Jose, California
+ San Leandro, California
+ San Mateo, California
+ San Pablo, California
+ San Rafael, California
+ San Ramon, California
+ Santa Clara, California
+ Santa Rosa, California
+ Saratoga, California
+ Sausalito, California
+ Sebastopol, California
+ Sonoma, California
+ South San Francisco, California
+ Suisun City, California
+ Sunnyvale, California
+ T
+ Tiburon, California
+ U
+ Union City, California
+ V
+ Vacaville, California
+ Vallejo, California
+ W
+ Walnut Creek, California
+ Windsor, California
+ Woodside, California
+ Y
+ Yountville, California"

#gsub to clean up the list
sfbay <- gsub("California", "", sfbay) #get rid of california
sfbay <- gsub("\n[A-Z]\n", "", sfbay) #get rid of the headers for each section of cities beginning with a certain letter
sfbay <- gsub("\n", "", sfbay) #getting rid of extra newlines
sfbay <- gsub(", " , ")|(", sfbay) #inputting the separations for when I use grepl later

#Had to look at it so I can copy/paste and make small edits for when I use it in grepl manually

bayavg <-  new_water_residential%>%
  filter(grepl(
"(Alameda)|(Albany)|(American.Canyon)|(Antioch)|(Atherton)|(Belmont)|(Belvedere)|(Benicia)|(Berkeley)|(Brent
```

```
wood)|(Brisbane)|(Burlingame)|(Calistoga)|(Campbell)|(Clayton)|(Cloverdale)|(Colma)|(Concord)|(Corte.Madera)|
(Cotati)|(Cupertino)|(Daly.City)|(Danville)|(Dixon)|(Dublin)|(East.Palo.Alto)|(El.Cerrito)|(Emeryville)|(Fairfax)|(F
oster.City)|(Fremont)|(Gilroy)|(Half.Moon.Bay)|(Hayward)|(Healdsburg)|(Hercules)|(Hillsborough)|(Lafayette)|(L
arkspur)|(Livermore)|(Los.Altos)|(Los.Altos.Hills)|(Los.Gatos)|(Martinez)|(Menlo.Park)|(Mill.Valley)|(Millbrae)|(
Milpitas)|(Monte.Sereno)|(Moraga)|(Morgan.Hill)|(Mountain.View)|(Napa)|(Newark)|(Novato)|(Oakland)|(Oakley
)|(Orinda)|(Pacifica)|(Palo.Alto)|(Petaluma)|(Piedmont)|(Pinole)|(Pittsburg)|(Pleasant.Hill)|(Pleasanton)|(Portola.V
alley)|(Redwood.City)|(Richmond)|(Rio.Vista)|(Rohnert.Park)|(Ross)|(St..Helena)|(San.Anselmo)|(San.Carlos)|(Sa
n.FranciscoSan.Jose)|(San.Leandro)|(San.Mateo)|(San.Pablo)|(San.Rafael)|(San.Ramon)|(Santa.Clara)|(Santa.Rosa
)|(Saratoga)|(Sausalito)|(Sebastopol)|(Sonoma)|(South.San.Francisco)|(Suisun.City)|(Sunnyvale)|(Tiburon)|(Union.
City)|(Vacaville)|(Vallejo)|(Walnut.Creek)|(Windsor)|(Woodside)|(Yountville)|(East.Bay)", Supplier.Name
))%>%
  group_by(Reporting.Month)%>%
  summarise(bay_use_avg = mean(resid_use))

#We want to compare to california as a whole so let's take the average for every month for the whole state!
state.avg <- new_water_residential%>%
  group_by(Reporting.Month)%>%
  summarise(state_use_avg = mean(resid_use))

#Now let's combine all three of these data sets so that we have just one data set we're plotting from
la_state <- state.avg%>%
  inner_join(losangeles.avg, by = c("Reporting.Month" = "Reporting.Month"))

alljoined <- la_state%>%
  inner_join(bayavg, by = c("Reporting.Month" = "Reporting.Month"))

#Now that I have everything joined, I have to make it tidy
#Using gather to make it narrow
narrowall <- alljoined%>%
  gather(key = boundary, value = avg_resid_use, state_use_avg, la_use_avg, bay_use_avg)

#using gsub to make the labels prettier
narrowall$boundary <- gsub("state_use_avg", "California", narrowall$boundary)
narrowall$boundary <- gsub("la_use_avg", "Los Angeles County", narrowall$boundary)
narrowall$boundary <- gsub("bay_use_avg", "All 9 Bay Area Counties", narrowall$boundary)

#Figure 4.3.1
  ggplot(narrowall, aes(x=Reporting.Month, y= avg_resid_use, color = boundary)) +
  geom_point() +
  stat_smooth(se=FALSE, method="loess") +
  labs(x = "Month", y = "Average Residential Water Usage", title = "Average Residential Water Usage vs Month")
+
  theme(legend.key = element_rect(colour = "black"),
        plot.background = element_rect(colour = "grey"),
        panel.background = element_rect(fill = "grey"),
        panel.background = element_rect(color = "black"),
        panel.grid.minor = element_line(linetype = "dotted"),
        axis.title = element_text(size = rel(1.5)),
        axis.text = element_text(size = rel(1.0)),
        legend.text = element_text(size = rel(1.0)),
        plot.title = element_text(size = rel(2))
        ) +
  geom_vline(aes(xintercept = as.numeric(ymd("2015-04-01"))), color = "red", size=1.5) +
```

```
scale_colour_manual("Region", values = c("green", "blue","yellow"))
```