Final Project: The Colorado River Basin

Wajeeh Khan
Email: wajeehk1@umbc.edu
GitHub: https://github.com/wajeehk1
Repository: https://github.com/wajeehk1/wajeehk.github.io
Portfolio: https://wajeehk1.github.io/wajeehk.github.io/
(the images were working until they didn't :/)

GES 486 Spring

Professor Mahmoudi

Abstract: Data found on 2010 municipal water allocation in the Colorado River Basin is visually presented, as well as analyzed. Correlation coefficients were calculated; mhni = 0.17, percent white = -0.3, total population = 0.97, county size = 0.07. Strongest correlation with domestic withdrawal was population, and the weakest was county size.

Introduction:

With the multiple heart attacks from the recent droughts in southern California, the lessened discharge that makes it to Mexico, and the political issues about allocating water resources, I decided to delve into some water consumption data to see which counties consumed the most, and what could potentially determine consumption rates. I mainly focused on relationships with water consumption between the variables median household income (mhni), percent white (per_white), total population (POP_2010 or Total_Pop), and the size of each county (Shape_Are). Using mapping/plotting and analytical techniques learnt during spring GES 486, analysis was done on water usage in the Colorado River Basin (CRB) (Thomas 2017; Walker 2023).

There is categorization when it comes to classifying who and what consumes water. I will be focusing on municipal data from 2010, otherwise known as domestic water withdrawal (Domestic_d).

Data + Methods

Data was obtained from US census data and these following links:

- Municipal Data 2010
- Water Use per State
- HUC Data

Using tidycensus, I obtained data from the counties of the states which bordered the hydrological boundary of the Colorado River Basin.

I imported points for major cities within the boundary, as well as the boundary itself. To be more aesthetically pleasing, I used st_boundary() for both the CRB boundary sf neighboring states sf to prevent the polygons from interfering with visual aesthetics.

Utilizing ggplot(), maps were made of the CRB counties, and then choropleth maps were used to show water use distribution across all the counties (Walker, 2023).

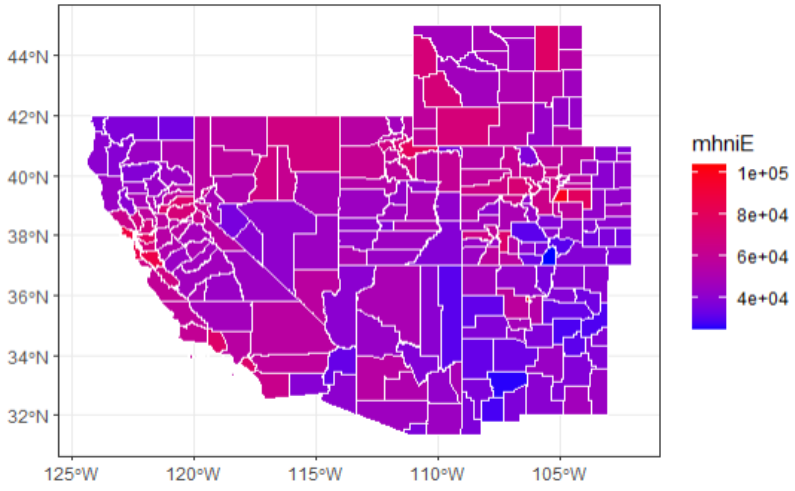


Figure 1: The initial plot from tidycensus (not polished)

Since I wanted only the counties which took from the Colorado Basin itself, I performed a st_join() between the census data and municipal data by matching GEOID values to

single out only the counties on which there were water consumption data on.

The primary function that I used to measure correlation was the cor() function (Thomas, 2017). Correlation with Domestic_d was tested for all variables listed in the introduction.

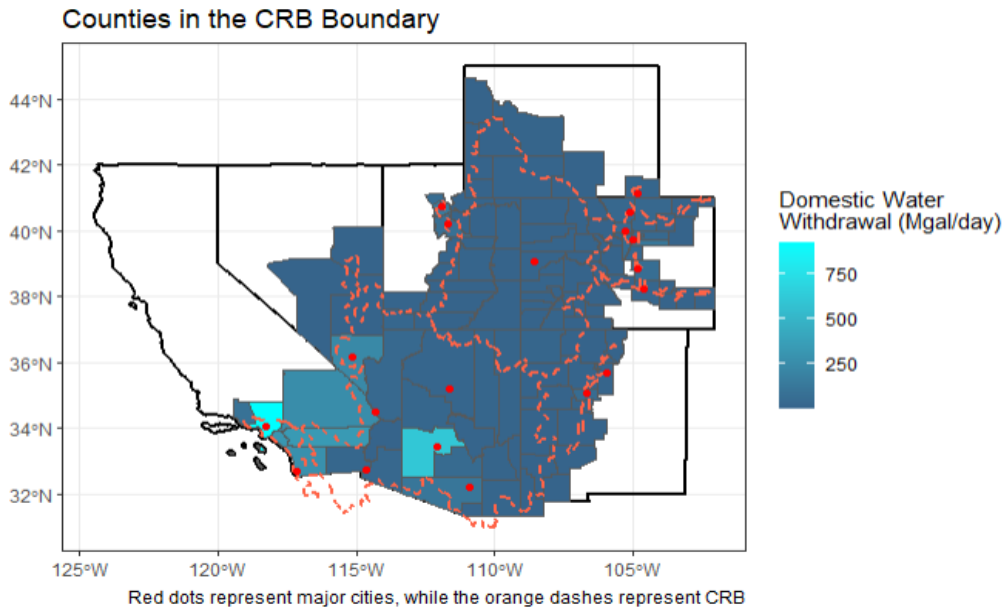
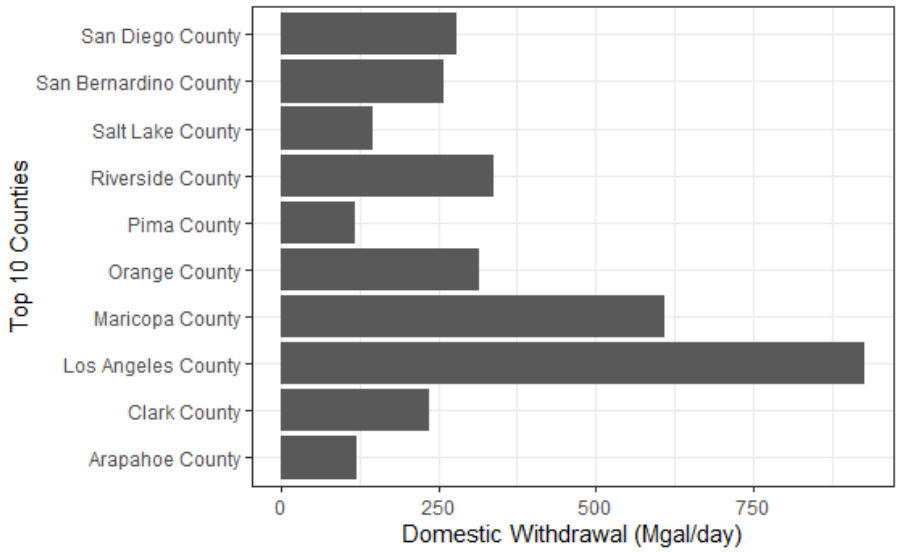


Figure 2: After performing a join and ridding rows containing NA values, the above was created.

Figure 3. A bar graph was created to show the top ten counties when it came to water consumption-

From the bar graph, it shows the top ten counties when it came to water consumption. Los Angeles is by far the worst offender.



From here on, I began to check for correlation between multiple values. The first being Median Household Income and Domestic Withdrawal-

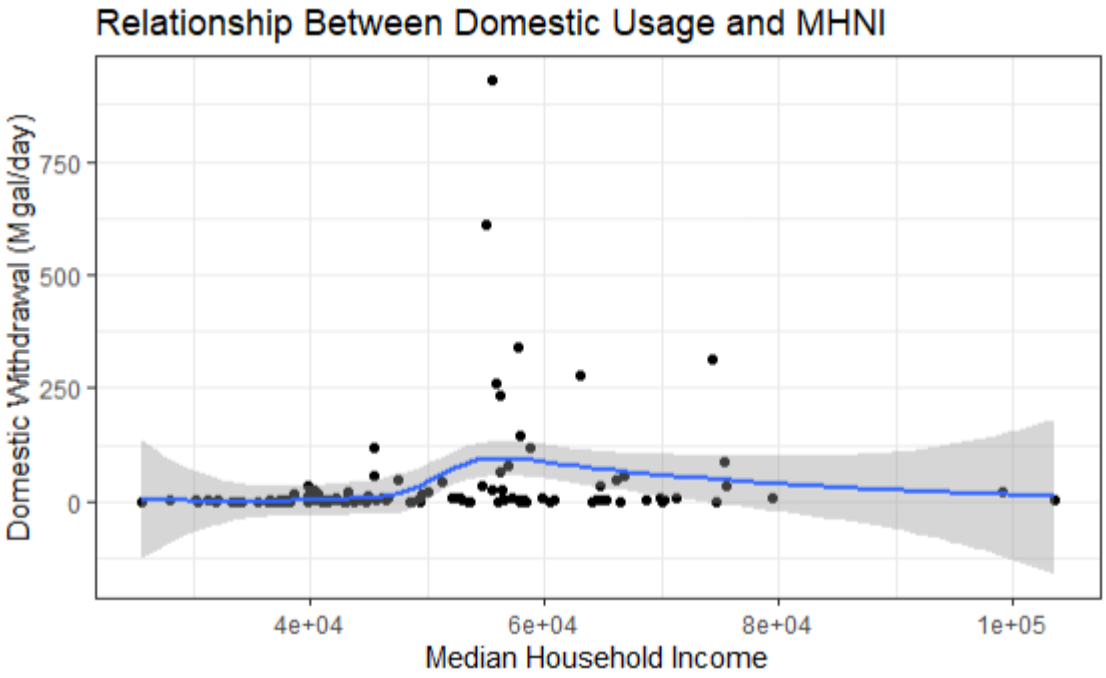


Figure 4: Los Angeles seems to be the reason for the hump in the geom_smooth() line. Corr = 0.17

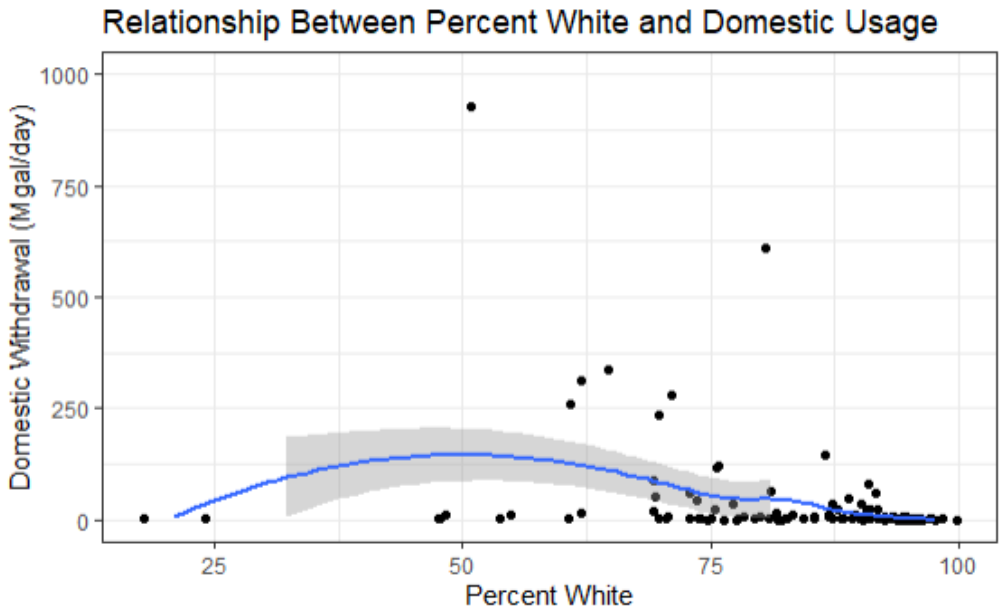


Figure 5: Percent white and domestic usage. Corr = -0.3

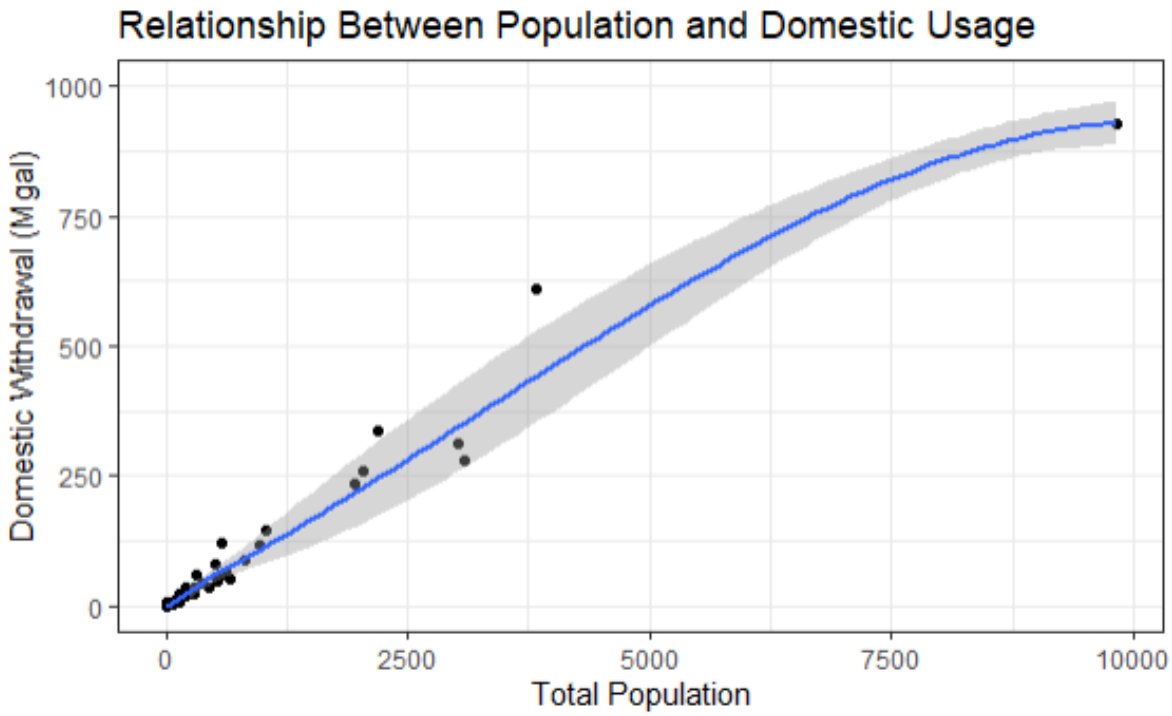


Figure 6: Population and domestic usage. Notable linear pattern. Corr = 0.97

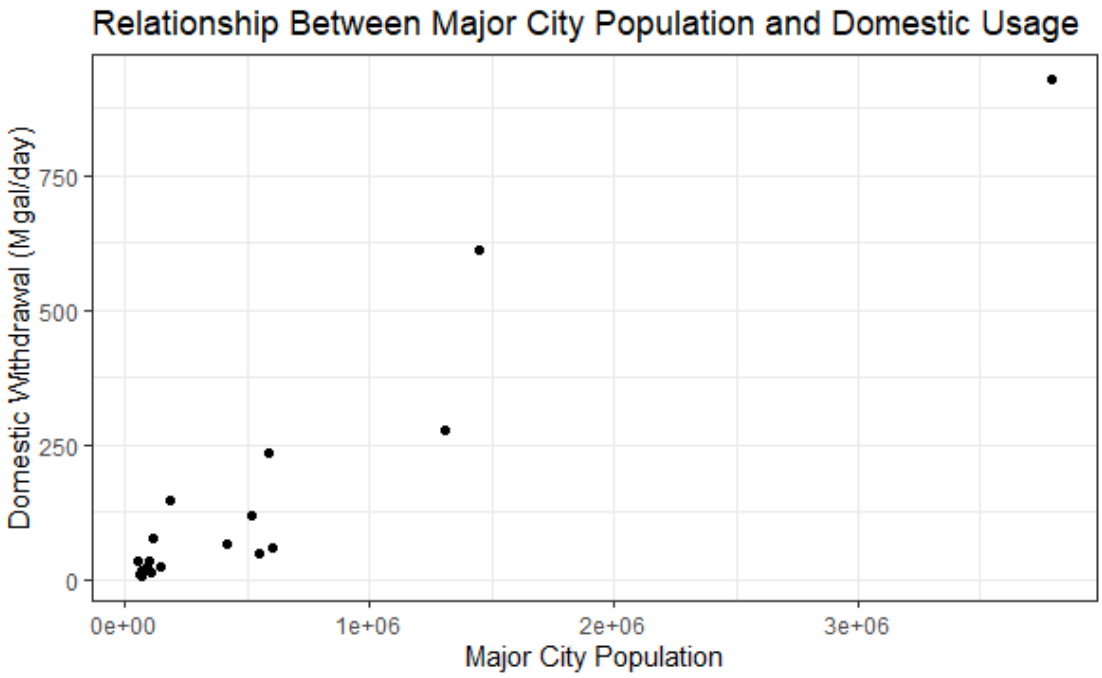


Figure 7: Major city population and domestic usage. Notable linear pattern.

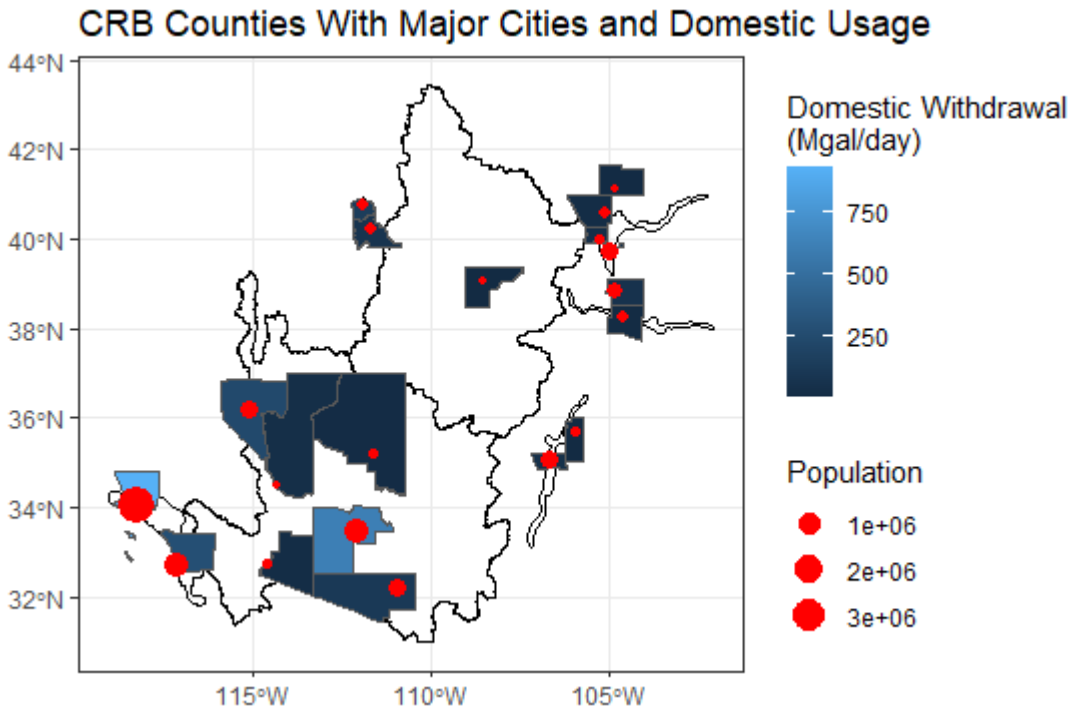


Figure 8. Visually representing major city population with domestic withdrawal-

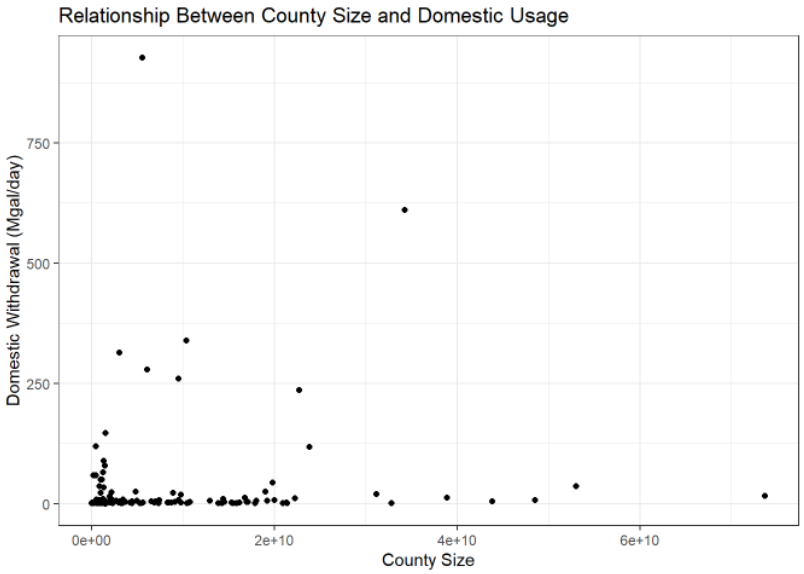


Figure 9. Plot for county size and domestic withdrawal. Corr = 0.07.

Analysis + Results

If one were to just look at the top ten consumers within the bar graph (see **Figure 3**), as well as the previous maps shown, one can assume Los Angeles County is unfairly allocating more resources towards themselves. This assumption can be nullified by looking at the relationship between population and water usage (see **Figure 6**). After using the cor() function for those two variables, it yielded a value of 0.97, indicating a very strong positive correlation. And because Los Angeles has a high population (see **Figure 3**), it then explains its greater usage compared to other neighboring counties within the CRB.

To test if there was an influence on water consumption with higher mhni values, I plotted it (see **Figure 4**) and then used the cor() once again. The resulting value of 0.17 infers a very weak relationship between median household income and water withdrawal. Given there is better infrastructure within wealthier areas, water usage can either be more efficient or wasteful. This is another reason as to why I wanted to test this relationship.

To test if there was any racial inequality when it came to using water from the CRB, I plotted percent white and usage (see **Figure 5**). The correlation value generated was -0.3, indicating a weak to moderate negative relationship. One could hypothesize from this value that the higher the percentage, the lower the domestic withdrawal.

Another variable that came to mind was the size of a county. Could the area of a county affect domestic uptake? The resulting correlation test for this scenario was 0.074. It can be said the correlation between the two is negligible.

Conclusion:

The strongest correlation to water uptake was total population. Graphically, it had a strong linear pattern, and a correlation value of 0.97. Other variables such as mhni (0.17), and the size of the county (0.07) held weak correlation values. One interesting find was the correlation of percent white and water. It held a value of -0.3, indicating a mild inversely proportional relationship.

Possible sources of error are the age and accuracy of the dataset, as data involving water can be inaccurate depending on the location as well as the level of dedication to document findings. The recent drought has to also be taken into account, as the drought is a recent issue (an issue whose severity has declined over the past couple of months) and could have greatly influenced domestic withdrawal.

Reflection:

I had some issues trying to flesh out the data in an interesting way. The data I had found was hard to implement, as it had used a HUC system instead of FIPS.

I wanted to compare usage across different consumers during the drought, which put the reservoirs under great stress. I wanted to see if uptake for agriculture or factories varied as much pre-drought and post-drought compared to municipal usage.

Work Cited

Khan, W. (2023). Final Project_ Colorado River Basin. “PDF.”

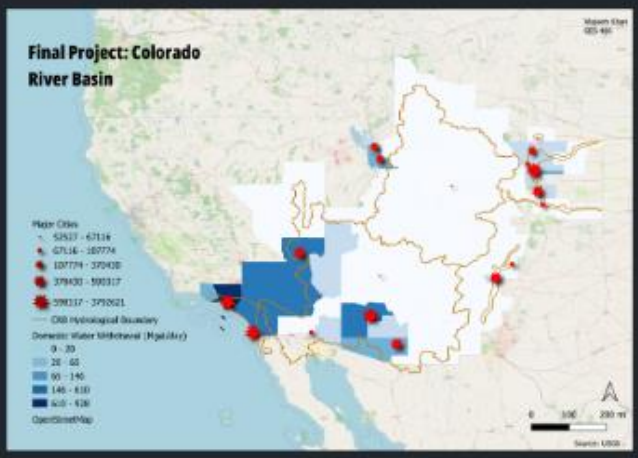
Thomas E. Love, Ph.D. “Data Science for Biological, Medical and Health Research: Notes for 431.” *Thomas E. Love*, thomaselove.github.io/431notes-2017/straight-line-models-and-correlation.html.

Walker, Kyle E. “Chapter 4 Exploring US Census Data with Visualization: Analyzing US Census Data.” *Chapter 4 Exploring US Census Data with Visualization | Analyzing US Census Data*, walker-data.com/census-r/exploring-us-census-data-with-visualization.html. Accessed 23 May 2023.

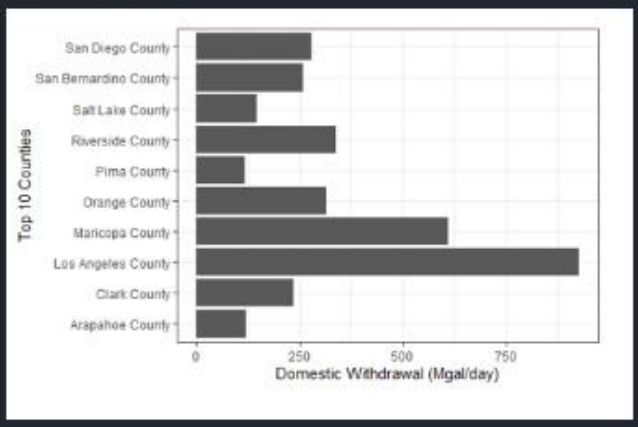
Final Project: Colorado River Basin

Given the multiple heart attacks from the recent droughts in southern California, the lessened discharge that makes it down to Mexico, and the political issues about allocating water resources, I decided to delve into some water consumption data to see which counties consumed the most, and what could potentially determine consumption rates. I mainly focused on relationships with water consumption between the variables median household income, percent white, total population, and the size of each county. Using mapping/plotting and analytical techniques learnt during the spring GEOG 486, analysis was done on domestic water usage in the Colorado River Basin (CRB).

Here is a map which shows the distribution of water consumption across the counties within the hydrological boundary



Top 10 resource hungry counties



Notable relationships found from plotting and analyzing

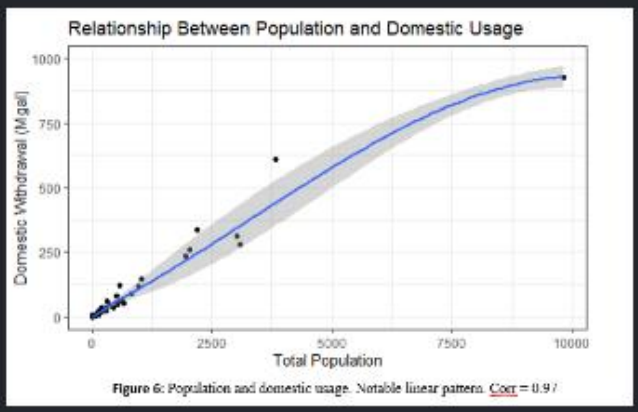


Figure 6: Population and domestic usage. Notable linear pattern. Corr = 0.97