Olivier Kwizera and Patrick Niyongabo

CSCI 385, Fall 2014

Final Project

Water Water Everywhere

I. Abstract

Fresh water is the limiting constraint for development in much of the world. Our aim was to model the water situation in the US and be able to come up with an effective, feasible, and cost-efficient water strategy that could be implemented in the real world. We approached this situation by first understanding the water situation in the US through applying some data analysis tools to the most recent water data published by the U.S. Geological Survey's National Water-Use Website[1]. Through analyzing these data in details and drawing some maps that show fresh water distribution across the country, we arrived at a strong foundation on which we based our conclusions. In particular, we were able to draw a map of US showing different clusters of states that tend to use water in the same way. We found that all 50 states can be grouped into 6 different clusters with each cluster having certain specific water strategies. As a matter of fact, the state of California stands apart in its own cluster because of how much water it uses; it accounts for about 11 percent of all total water withdrawals in the US and 10 percent of all fresh water, predominantly for irrigation. This result lead us to conclude that the use of water-saving sprinkler systems such as drip systems in irrigation in California would help save millions of gallons of water.

II. Introduction

Our main goal is to study the water problem and be able to at least provide a foundation for potential action plans. Since this water situation is specific to US, we believe that the best way to obtain a realistic model that can be implemented is to first understand the water problem in this particular country. The U.S. Geological Survey's National Water-Use program provides water data in files of different formats for users and those data can be accessed on their website[2]. Throughout this document, we will be extracting information from the file found on their website containing the most recent water use data.

Water use by category

There are about 8 different categories of water use:

- Public supply
- Domestic

---

[1] http://water.usgs.gov/watuse/data/2010/index.html

[2] ibid

- Industrial
- Irrigation
- Livestock
- Aquaculture
- Mining
- Thermoelectric power

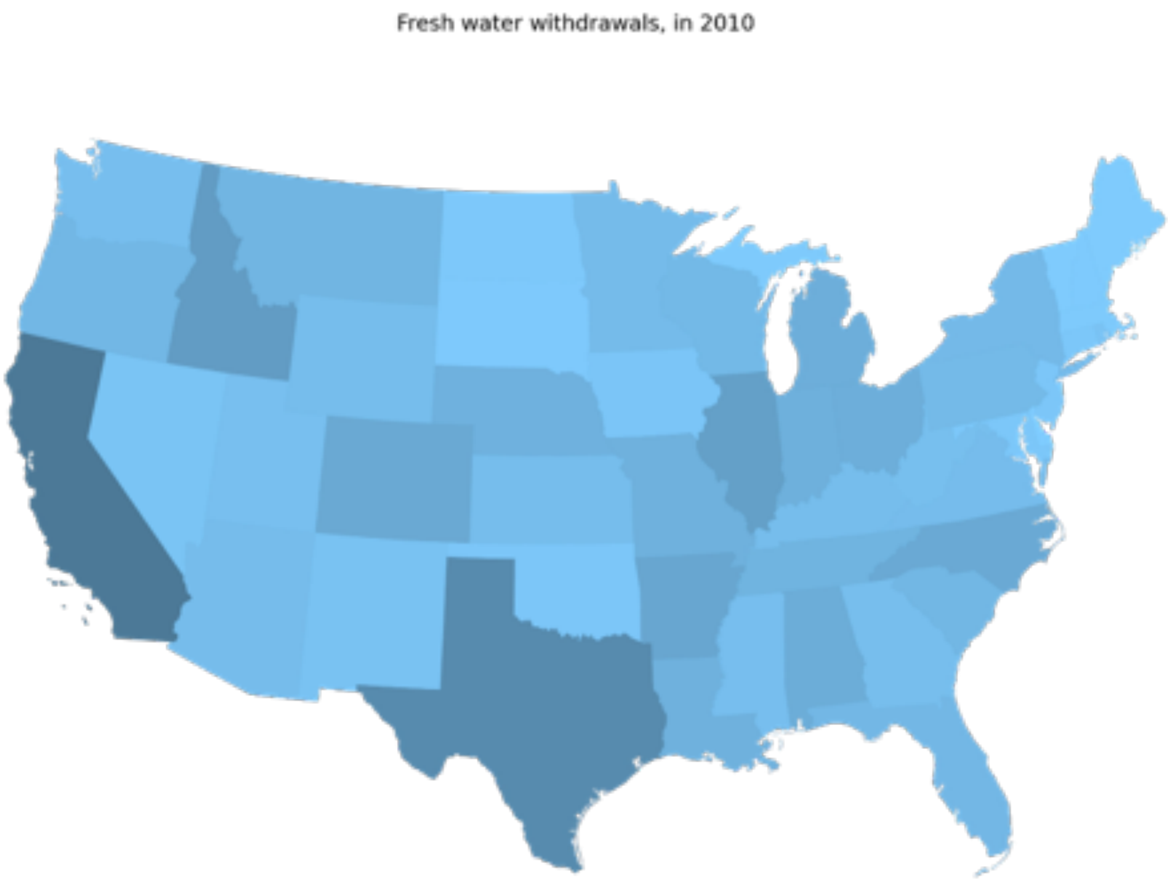
Water sources

Water sources described in this document come in types. Groundwater which refers to all subsurface water and surface-water sources which include streams and rivers, lakes and reservoirs, and oceans. For each water source, there are two types of water- fresh water and saline water.

We will be analyzing data in steps:

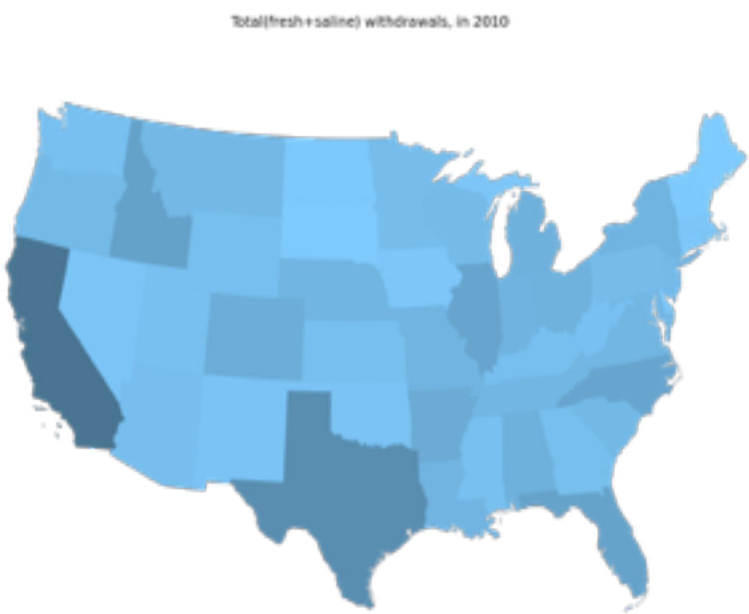
Step I: Overview

Since we have two main water categories (fresh water and saline water), we need to know exactly the total withdrawals for each category. To that end, we can make matters clearer by visualizing the distribution of total fresh water withdrawals and total saline water withdrawals across all the states. The following are three maps of US showing the distribution of water use across different states. **Map 1** below shows the total fresh water withdrawals in each state. It shows us that states like California, Texas, and Idaho are responsible for the highest fresh water withdrawals. In the same way, **Map 2** shows that states that touch oceans such as California and Florida use the most saline water. Finally, **Map 3** sums up the total water used in each state and this map helps to highlight that California uses more water than any other state on the overall.



Map 1: Map of USA illustrating the total fresh water withdrawals.

Map 2: Map of USA illustrating the total saline water withdrawals.



Map 3: Map of USA illustrating the total water (fresh + saline) withdrawals.

Step II: Description of Data types

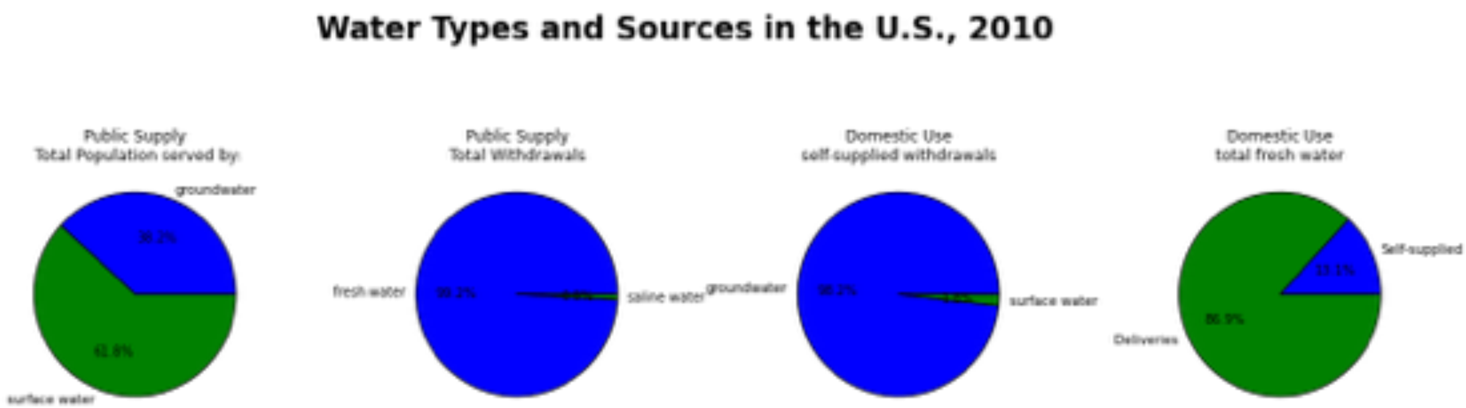
Since our main goal is to preserve as much fresh water as possible, our model is going to handle this task by dealing with each water use category individually. We have seen that there about eight different water use categories and some categories utilize both fresh water and saline water whereas other categories use almost exclusively fresh water only. The following section include self-explanatory pie charts and they show where each water-use category gets its water from and the quality of water used (saline or fresh).

Category 1: Public Supply Water Use

Public supply refers to water withdrawn by public and private water suppliers to at least 25 people or have a minimum of 15 connections. Public-supply water is delivered to users for domestic, commercial, and industrial purposes.

Category 2: Domestic Water Use

Domestic water use includes potable and non-potable water provided to households by a public water supplier or self-supplied water use.



As the above pie charts show, water used in these two categories (Public Supply and Domestic) can be from surface-water sources or groundwater. An efficient water strategy for this category would be to encourage people to use water at home more efficiently and water recycling.

Category 3: Industrial Water use

Industrial withdrawals provide water for different industrial purposes such as fabricating, processing, or for sanitation needs within the manufacturing ability. Conservation measures within industries and recycling systems are key to any water strategy in this category.

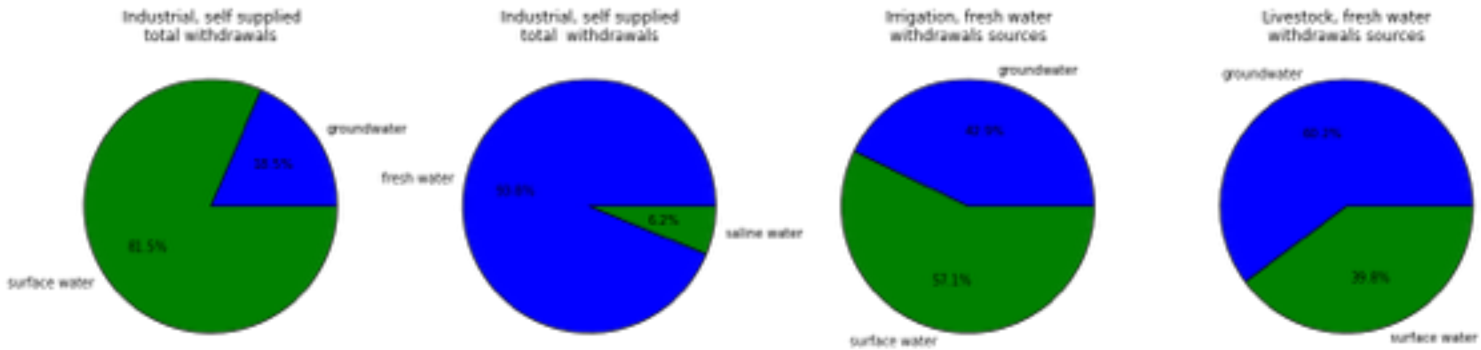
Category 4: Irrigation

Irrigation water use includes water that is applied by an irrigation system to sustain plant growth in all agricultural and horticultural practices. Withdrawals in this category also include water used in the irrigation of golf courses, parks, turfs, nurseries, cemeteries, and other self-supplied landscape-watering uses. Efficient water strategies in this category include the use of water-saving sprinkler systems such as drip systems and the use of efficient irrigation methods such as micro-irrigation and surface (flood) systems.

Category 5: Livestock

Livestock water use is water associated with livestock watering, feedlots, dairy operations, and other on-farm needs. Other livestock water uses include cooling of facilities for the animals and products, dairy sanitation, and wash down of facilities, animal waste-disposal systems, and incidental water losses.

The below pie charts show the water sources for each of the above three water use categories: industrial, irrigation, and livestock.



Also, industrial water use category utilize both saline and fresh water whereas irrigation and livestock categories use fresh water only.

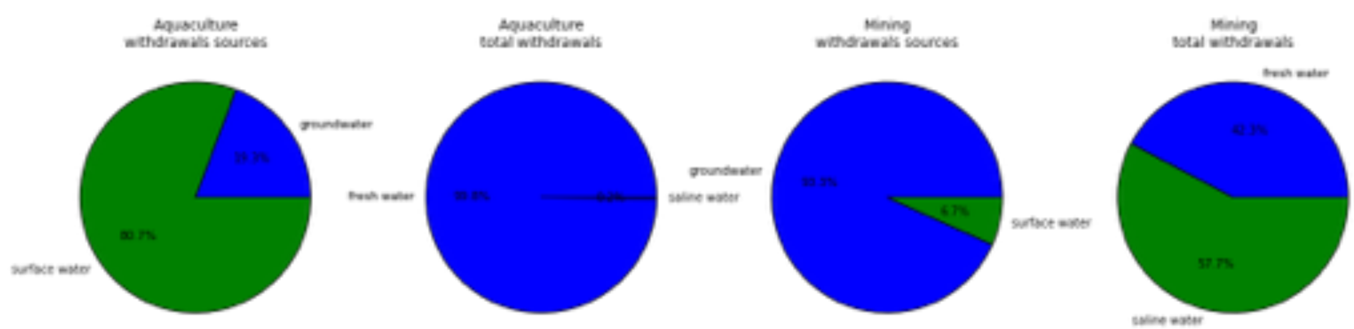
Category 6: Aquaculture water use

Aquaculture water use is water associated with raising organisms that live in water such as finfish and shellfish. Aquaculture production occurs under controlled feeding, sanitation, and harvest procedures.

Category 7: Mining Water use

Mining water use is water used for the extraction of minerals. The category includes milling of mined minerals, injection of water for secondary oil recovery or for unconventional oil or gas recovery (such as hydraulic fracturing), and other operations associated with mining activities.

The below pie charts show different water sources for the aquaculture and mining water categories.



From the above pie charts, it is clear that the mining water use category utilize a great deal of saline water more than any other category. Using more saline water from surface sources such as oceans can decrease the amount of water extracted from groundwater in states that practice mining more than others.

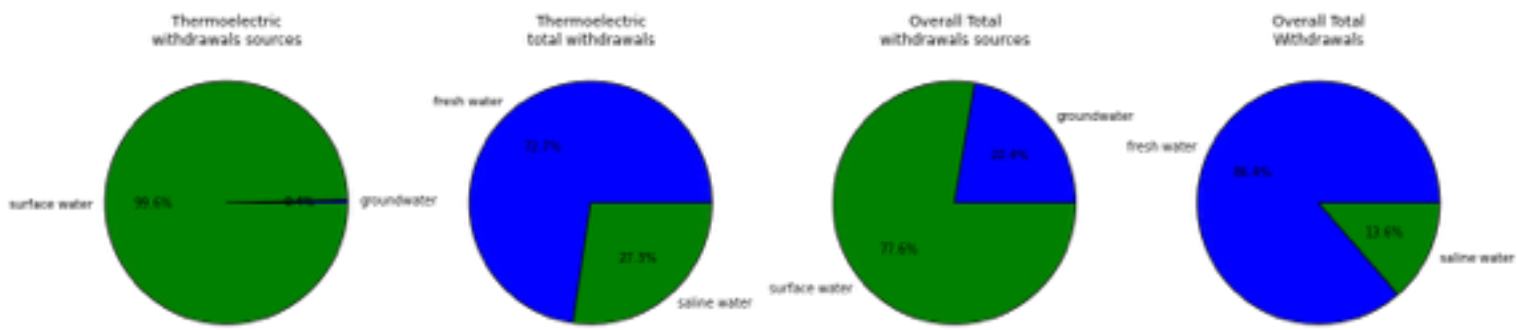
It would also help to decrease the cost since the process of gathering water from subsurface sources is expensive.

Category 8: Thermoelectric power

Water for thermoelectric power is used in generating electricity with steam-driven turbine generators. As the following series of pie charts below show, thermoelectric water use category accepts both saline and fresh water while this category almost uses water from surface sources only.

Overall Use of water in the U.S

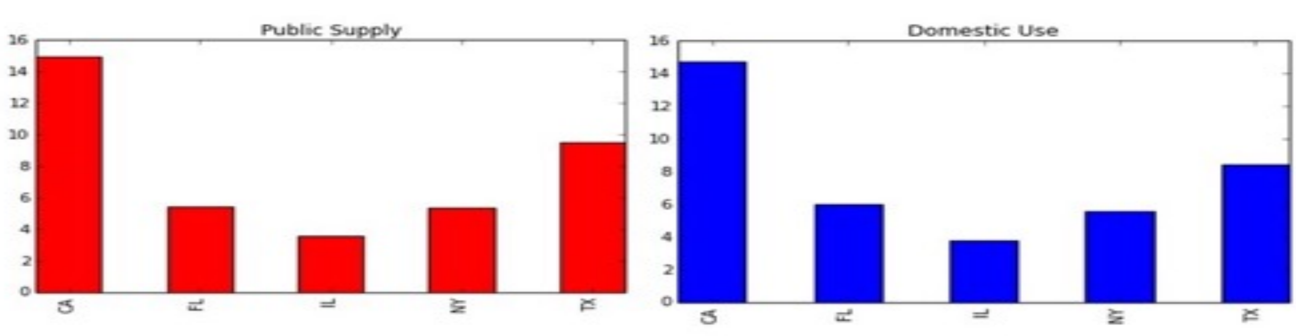
From each category data and the diagram below, one can see that fresh water is more used in the U.S. than saline water, and surface water withdrawals are far greater than groundwater withdrawals.



The above pie charts reveal that a water strategy involving desalination would be effective in states that use more water in the thermoelectric category since both saline and fresh water are used for this purpose.

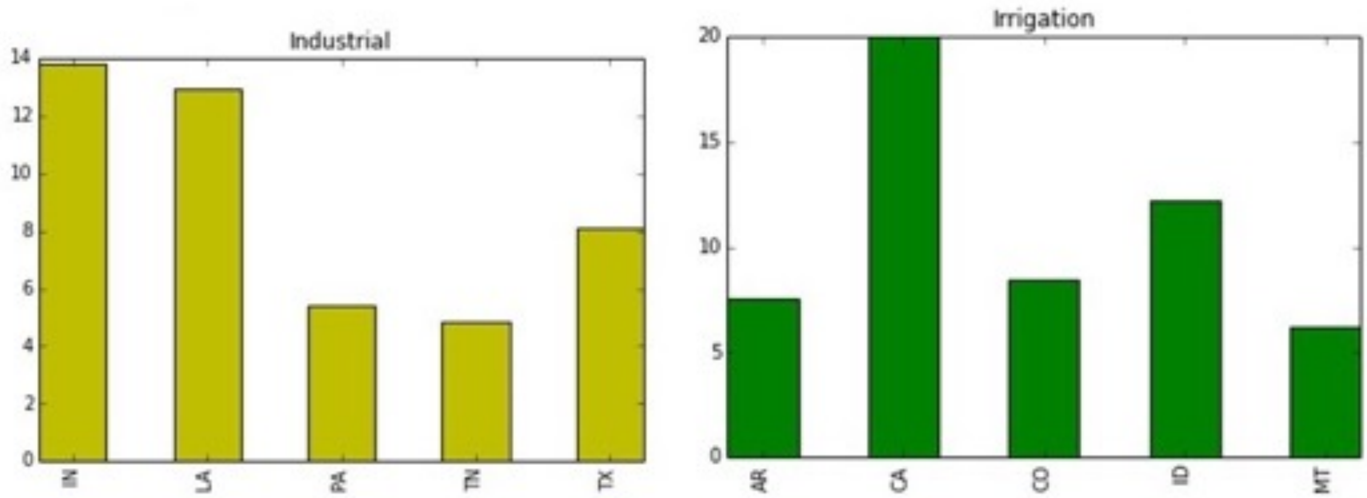
Furthermore, all the above pie charts help us to learn more about all eight water use categories and the potential water strategies for each water use category, but they don't tell us which states use most water in each category. To gain further understanding of the situation considering states individually, we could display the five states that use most water in each category. **Figure I** shows the top five states that use the most water in the Public supply and domestic water use category and the percentage that these states contribute in each category.

Figure I: Public Supply and Domestic use



What these two categories (public supply and domestic) have in common is that the amount of water used is proportional to the number of population, thus the result being that the most populous states[3] (California, Texas, Florida, New York, and Illinois) are the top five in these two categories. **Figure II** illustrates the trend of water use in both industrial and irrigation categories.

Figure II: Industrial use and Irrigation.



In the above bar charts, one can notice how California uses more than 20% of all the water withdrawals used for irrigation purposes.

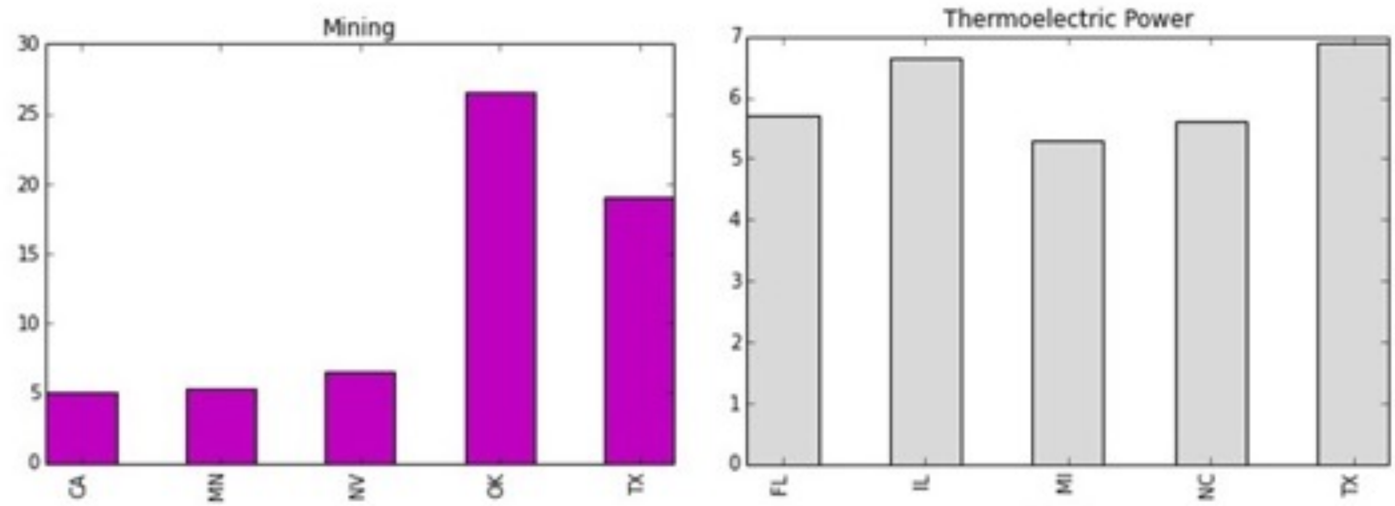
In the industrial use, Indiana and Louisiana lead. According to *usatoday.com*[4], those two states are among the top three states with many manufacturing factories in the U.S. This trend suggests that measures that emphasize water reuse and recycling in industrial facilities in Louisiana and Indiana could lead to a significant decrease in total withdrawals in the industrial water use category.

Figure III explains the same trend for mining and thermoelectric power categories.

[3] http://addictivelists.com/top-10-most-populated-states-in-us-2014/

[4] http://www.usatoday.com/story/money/business/2013/08/10/10-states-where-manufacturing-still-matters/2638363/

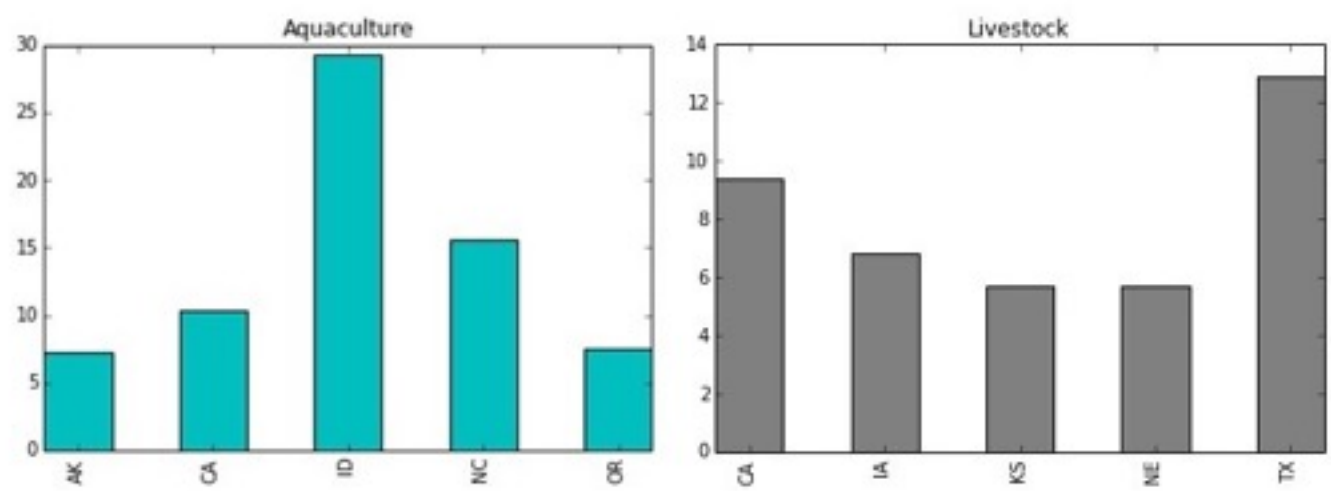
Figure III: Mining and Thermoelectric Power



The above bar chart on the left shows that Texas and Oklahoma are leading in mining water uses, because those two states' mining industries are among the top three in the U.S. based on GDP[5]. These two states are responsible for about more than 40 percent of all withdrawals in the mining category. On the thermoelectric power water uses bar chart, one can observe that states have almost equal percentages because all states have their own thermoelectric power plants.

Finally, **Figure IV** shows the states responsible for most withdrawals in the Aquaculture and livestock category.

Figure IV: Aquaculture and Livestock



It is not surprising to see Idaho being dominant in the aquaculture category, given that Idaho is the leading state in the nation when it comes to producing trouts [6](75%), and other domestic aquatic organisms.

---

[5] http://econpost.com/industry/mining-industry-top-10-states-gdp

[6] http://magicvalley.com/read-more-about-idaho-aquaculture-facts/article_e9c94c54-59c0-11e0-861e-001cc4c002e0.html

Texas and California lead in the livestock category, and part of this result is down to the large population of both states.

Step III: Hypothesis and Assumptions

At this point we have enough information about different water use categories, different water sources, and different water types used for each category. From now on we can begin our quest to finding feasible water strategies using the water use information that we have about each state. But first, we need to make some assumptions that will guide us through.

Assumptions

- Since different states, in fact different regions, use water mainly in different ways, one water strategy in one region might not be as effective in another region.

- All eight water use categories are not equally potent in each state, and therefore any potential water strategy is likely to be influenced more by some water use categories than others in different regions.

In addition to these assumptions, it is important that we make our hypotheses clear.

Hypotheses

- Depending on each state's main use of water, some states will be similar and therefore yield a particular water strategy. For instance, south states might converge to using much of their water in irrigation and livestock and therefore efficient irrigation systems might be at the forefront of any water strategy in these regions.

- Some water strategies work best to particular regions and not so much to others, but of course any efficient water strategy can be applied anywhere.

Step IV: Clustering

With these hypotheses and assumptions in mind, it seems logical to apply the unsupervised K-means clustering techniques to our large set of data in order to find different clusters, in other words what we can call different regions as far as determining water strategies is concerned. For each state, we will create an 8- dimensional vector based on the normalized fresh water use for each category. By using K-means clustering, we can find clusters that best represent the overall water use in the US. **Figure 1** shows an elbow chart for different number of clusters.

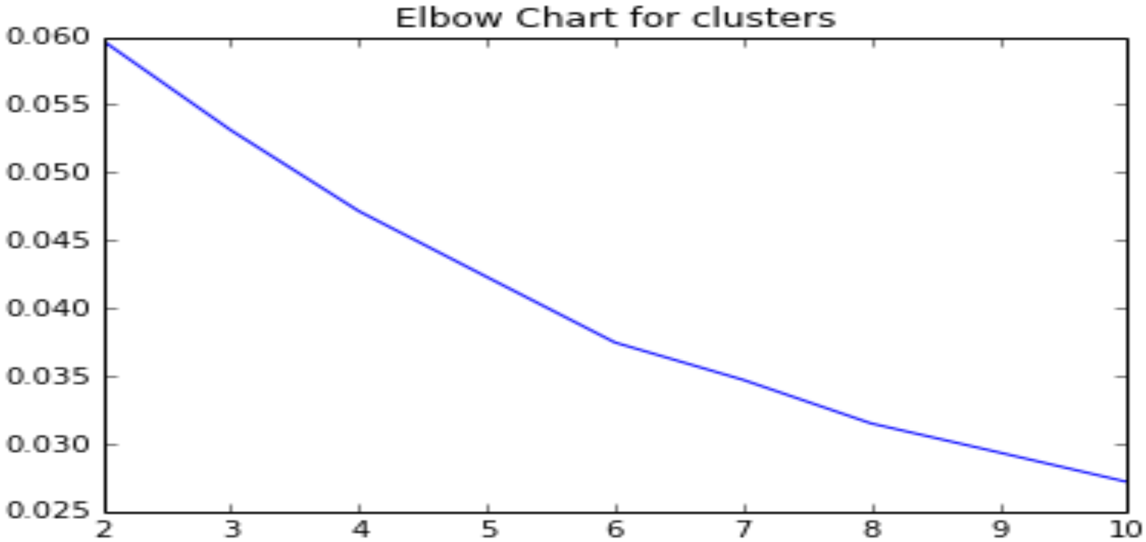
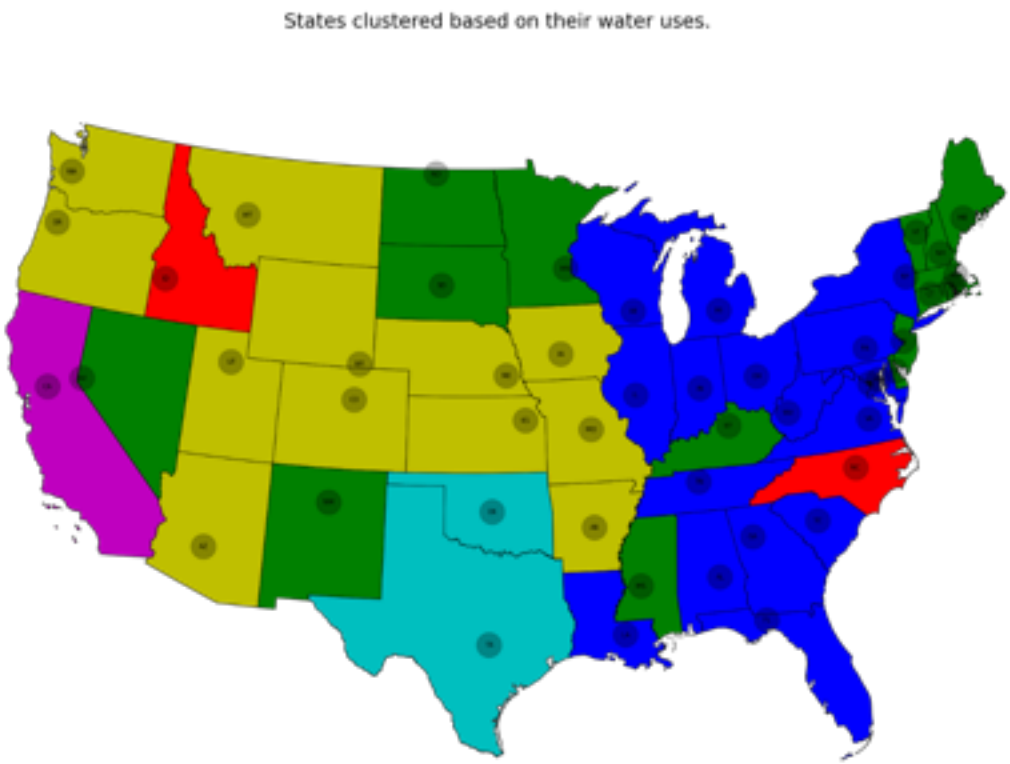


Figure 1: Elbow chart for different number of clusters

The above figure illustrates that the ideal number for states clusters is six.

Now, the question is what are the six different clusters and which states are in which clusters? **Map 4** illustrates the six clusters and the states that belong in each cluster. States with the same colors belong in the same cluster. After all, California uses way more water than any other state that it deserves to be in its own cluster.



Map 4: Map of USA showing the different clusters with different colors

Step V: Principal Component Analysis

After obtaining our clusters, we can perform the Principal Component Analysis (PCA) to the data vectors to determine the most important factors in our new transformed dimension. Consequently, we can determine different water strategies based on what the most important factors are and the correlation of water use categories in our states clusters. **Table A** shows the analysis of the first two principal components.

Table A: Water use categories and the first two principal components.

Variable (Water Use Category)	Principle Component	
	1	2
Public Supply	**-0.49986066**	-0.04048902
Domestic	**-0.49240917**	-0.01477053
Industrial	-0.19955013	**-0.46684823**
Irrigation	-0.33677691	**0.53790254**
Livestock	**-0.41177383**	0.06349924
Aquaculture	-0.14662314	**0.52114347**
Mining	-0.24317749	0.0596021
Thermoelectric	-0.32286571	-0.46008801

Included in the above table are the first two principal components. The bold values indicate what we would consider to be strong correlation coefficients to mean the furthest distance from zero in the positive or negative direction. The second column of the table (first principal component) shows the correlation between water use categories. Specifically, it implies a relatively moderate correlation between public supply and domestic water use categories. Similarly, the second principal component suggests a relatively strong correlation between the irrigation and aquaculture water use categories. Also, thermoelectric power and industrial water categories are fairly correlated.

Step VI: Conclusions

At this point, we can conclude our analysis with the main findings:

- California is responsible for the most water withdrawals; in fact, it deserves its particular water strategies. Since the total withdrawals in California are predominantly in irrigation (more than 20% percent of all irrigation withdrawals in the country) as **Figure II** shows, the priority in this state should be put in implementing efficient irrigation methods such as water-saving sprinkler systems and micro irrigation should be encouraged where possible.

- In order to reduce the total fresh water withdrawals that go toward mining purposes, Oklahoma and Texas should be closely monitored since they belong in the same cluster and between them they are responsible for about more than 40 percent of the total withdrawals in the mining category as **Figure III** indicates. In particular, an emphasis should be put into transporting water from the Gulf of Mexico at the south of Texas to the mining facilities in these two states since mining operations don't care much about the quality of water- saline or fresh.

- Any effort to reduce the water used in the aquaculture category should focus particularly on Idaho and North Carolina. It is important to note that the two states alone are responsible for about 45 percent of all the withdrawals in this category according to **Figure IV**. A potential solution would consist of implementing desalination plants in these two states that could produce fresh water from water containing minerals (both groundwater and surface water) through the process of desalination.

- Louisiana and Indiana are responsible for about 27% of all industrial water withdrawals as **Figure II** indicates, and therefore a significant reduction in water used in this particular category would encourage water reuse and recycling in industrial facilities in these two states. In addition, water reuse and recycling techniques should be at the heart of all strategies dealing with states in the blue color on **Map 4**; all these blue states share the trend of using the most water in industrial processes as well as thermoelectric power generation. Both these categories seem good fits for water reuse and recycling methods.

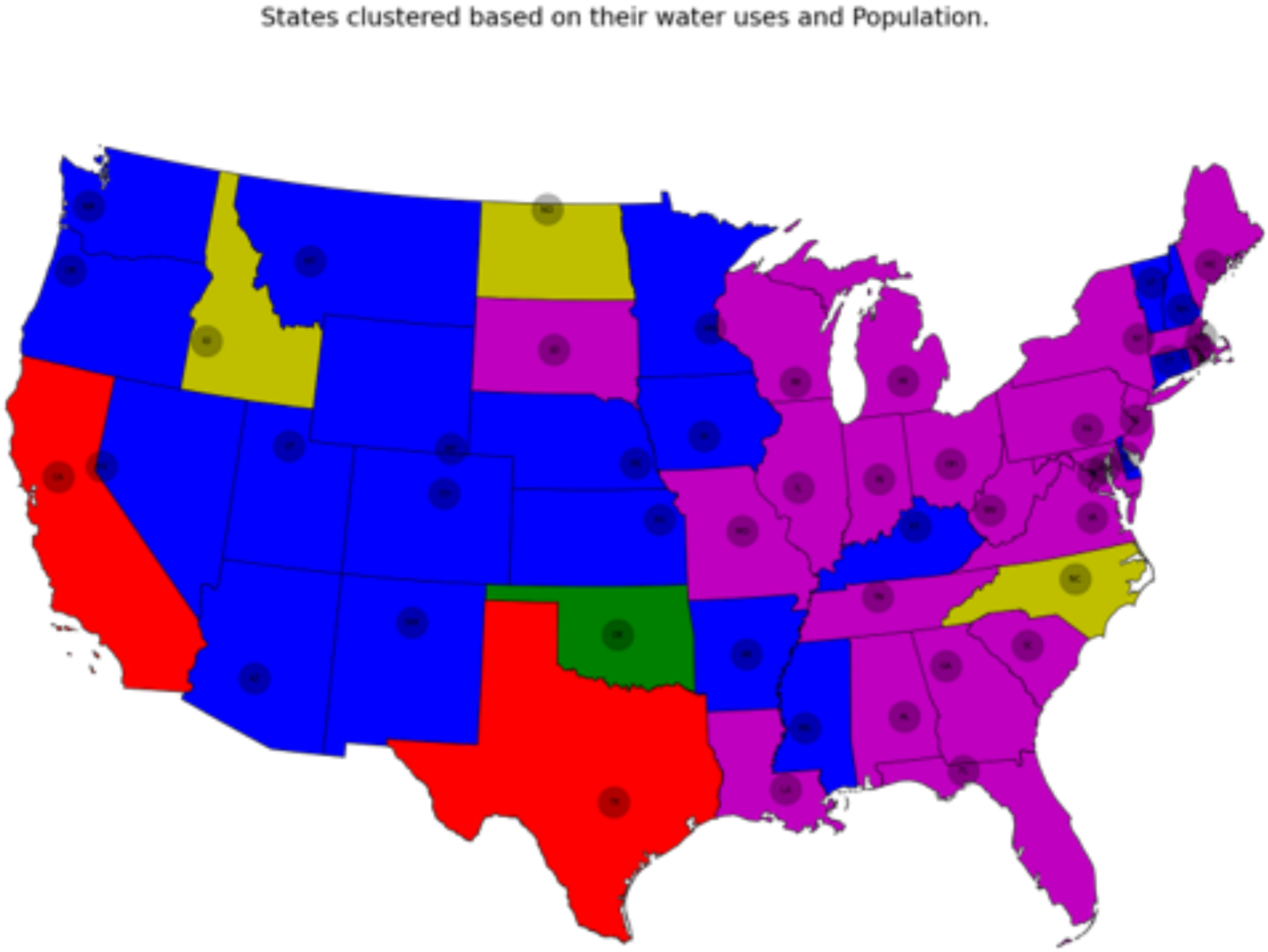
- Yellow states on **Map 4** tend to be mainly involved in the irrigation, livestock, and aquaculture water use categories. Potential water strategies include efficient irrigation methods such as the use of drip systems, micro irrigation, and so forth. Also, water reuse and recycling should be considered here.

- Finally, the green states on **Map 4** should consider measures that encourage more efficient water use domestically. In addition, water action plan in these states should put an emphasis on livestock and public supply water categories.

<u>Step VII: Sensitivity Testing</u>

To examine how our model can respond to different factors variation, we are going to introduce population in our state vectors. Does this change our clusters? If so, how significant does introducing population in our vectors alter our clusters and what are the implications of that change?

Map 5 shows the new clusters after introducing the population in each state's data vector.



Map 5: Map of USA after the population is taken into account

As the above map indicates, clusters tend to keep their shape on the overall. The merging of some states into one cluster can be explained by the fact that states that use water in a more fairly way become even similar when the population is factored into account. This can explain why Texas and California form their own cluster since they become more similar if we consider the population in these two states; they both have a huge population and they use significantly more water as **Map 3** indicates.

Model Strengths and Weaknesses

It is important to acknowledge what we believe are the strengths of our approach as well as its weaknesses.

Strengths

- The visualization part of our findings
- The results confirm our hypotheses.
- The results reflect the exact water situation in the US since our model relies on real data.

Weaknesses

- Our model does not account for other factors such as climate change and amount of wasted water.

- Our model doesn't guarantee a ready-to-be-implemented water strategy but it certainly points us in the right direction to look for potential solutions.
- Our model relies on the K-means clustering technique which is an expectation maximization algorithm; this implies that the clusters found are subject to change a little bit, but the core shape of clusters is retained.
- Finally, our model doesn't account for geographic location of states in determining our clusters. This geographic factor can influence the use of water between two or more neighbor states.

Speaking of limitations for our model, more work could be added so to make our model yield more precise results.

Future Work

Fresh water problems will always grab the attention of humans simply because people can't afford to live without enough fresh water. Consequently, modeling these water related situations in different countries and finding adequate solutions will always concern scientists and governments' leaders alike. In the future, one can imagine applying this model to a completely different country, one that we are less familiar with. It would be interesting to test our approach to a country like China or Egypt, and see what one can learn. Moreover, our model can provide a strong foundation to modeling more specific water strategies in the US such as modeling the best cost-efficient way to transport water across different states. That said, one more potential area of interest would be to study how water could be transported or managed from the nearest big surface water sources of each cluster since it seems that for each of the cluster found on **Map 4**, there is a big river that traverse the states in that cluster. For instance, the states in yellow on **Map 4** appear to be bounded by the Colorado River on the left as well as the Mississippi river on the right.

References Cited

E.M. Hess, Alexander, Michael B. Sauter, and Thomas C. Froehlich. "10 States Where Manufacturing Still Matters." *USA Today.* August 10, 2013. Accessed December 17, 2014. http://www.usatoday.com/story/money/business/2013/08/10/10-states-where-manufacturing-still-matters/2638363/.

"Estimated Use of Water in the United States County-Level Data for 2010." *U.S. Geological Survey.* Accessed December 1, 2014. http://water.usgs.gov/watuse/data/2010/index.html.

"Mining Industry Top 10 States by GDP." - Industries. Accessed December 17, 2014. http://econpost.com/industry/mining-industry-top-10-states-gdp.

Payne, Kevin. "Purple America." *Nifty.stanford.edu*. Accessed December 6, 2014. http://nifty.stanford.edu/2014/wayne-purple-america/ purple-america-data.zip.

"Read More about Idaho Aquaculture Facts." *Twin Falls Times-News*. March 29, 2011. Accessed December 17, 2014. http://magicvalley.com/read-more-about-idaho-aquaculture-facts/article_e9c94c54-59c0-11e0-861e-001cc4c002e0.html.

Ross, Dave. “State Table”. Accessed December 5, 2014. http://statetable.com/.

"Top 10 Most Populated States in US - 2014." *Addictive Lists*. April 26, 2014. Accessed December 17, 2014. http://addictivelists.com/top-10-most-populated-states-in-us-2014/.

"United States Census Bureau." 2010 FIPS Code Files for Counties. Accessed December 2, 2014. https://www.census.gov/geo/reference/codes/cou.html.