BUT THE DROUGHT CAME BACK? THE VERY NEXT YEAR!

Harvard University | CS171 |Final Project Process Book

Ben Steineman & Shiuh-Wuu (Victor) Liu

OVERVIEW AND MOTIVATION

Earlier this month, California Governor, Jerry Brown, issued new directives which aim to reduce water consumption which includes an unprecedented mandatory 25% cut in urban water use. These measures are intended to address the growing concerns and threats of the sustained drought over the last couple of years.

As proud Californians working for a renewable energy technology company, we are deeply concerned with sustainable living practices which will impact our friends and family as well as our posterity. As a result, we are very passionate about gathering insights into this topic which may lead to some innovative solutions that could help address this problem.

RELATED WORK

The following resources have inspired us and helped us drive forward a meaningful visualization for the purposes of conveying important insights into how the drought has affected California in the last few years.

1) ▶ MIT Residential Footprint - YouTube. (n.d.). Retrieved April 11, 2015, from https://www.youtube.com/watch?v=9-vl6AJ32fg
   a) We looked at some of the views presented in the MIT commute visualization and had considered a similar layout.

2) ▶ The Cat Came Back - Camp Songs - Kids Songs - Children's Songs by The Learning Station - YouTube. (n.d.). Retrieved April 11, 2015, from https://www.youtube.com/watch?v=LjMffHG1V_Q
   a) This song from our childhood helped inspired the title for our project.

3) California Drought Information | USGS California Water Science Center. (n.d.). Retrieved April 11, 2015, from http://ca.water.usgs.gov/data/drought/
   a) This website was referred to us by Eric Reichard egreich@usgs.gov who is our direct contact from USGS for any data questions that we may have.

4) California Land & Water Use. (n.d.). Retrieved April 11, 2015, from http://www.water.ca.gov/landwateruse/surveys.cfm

a) Data on agriculture uses of water if we decide to incorporate California crops and how crop selection affects the severity of the drought.

5) CDFA > STATISTICS. (n.d.). Retrieved April 11, 2015, from http://www.cdfa.ca.gov/statistics/
   a) Agriculture production data could be located here.

6) cida.usgs.gov/ca_drought/. (n.d.). Retrieved April 11, 2015, from http://cida.usgs.gov/ca_drought/
   a) A self-reported data visualization of the drought was created using freely available USHS data.

7) CIDA-Viz/ca_reservoirs.json at master · USGS-CIDA/CIDA-Viz · GitHub. (n.d.). Retrieved April 11, 2015, from https://github.com/USGS-CIDA/CIDA-Viz/blob/master/ca_reservoirs/Data/ca_reservoirs.json
   a) This is the reservoir capacity and utilization data broken out by date and by reservoir name.

8) Crop_Coeffients.pdf. (n.d.). Retrieved from http://www.cimis.water.ca.gov/Content/PDF/Crop_Coeffients.pdf
   a) The crop coefficients rates are multiplied by then evaporation-transpiration of the reference group. We were interested in using the data as an additional point of reference.

9) How to Estimate Water Useage Required for an Irrigation System. (n.d.). Retrieved April 11, 2015, from http://www.irrigationtutorials.com/how-to-estimate-water-useage-required-for-an-irrigation-system/
   a) We were considering the use of the 'formula to calculate the gallons of irrigation water needed per day' to derive water usage where other more precise data was not available.

10) Mapping the Spread of Drought Across the U.S. - NYTimes.com. (n.d.). Retrieved April 11, 2015, from http://www.nytimes.com/interactive/2014/upshot/mapping-the-spread-of-drought-across-the-us.html
   a) This data vis was created by Mike Bostock and it shows how the drought has affected the US at large. We wanted to look at California in particular, but it was definitely interesting to see how the drought has affected many other parts of the country.

11) Microsoft PowerPoint - Blaine-Hanson Water Forum complete.ppt - blaine-hanson_water_forum_complete.pdf. (n.d.). Retrieved from http://www.pge.com/includes/docs/pdfs/shared/edusafety/training/pec/water/blaine-hanson_water_forum_complete.pdf
   a) General information from PGE on Evapotranspiration based on the crop type. It was a possibility to use the data presented in the slides to infer water usage for certain crops grown in California.

12) Streaming through 1Channel.ch. (n.d.). Retrieved April 11, 2015, from https://add2ac80562d5288b8b87115bba350a041fd1663.googledrive.com/host/0B2kv7wOF5KquclBsZXlUR1hCNms/index.html

a) The list to scatter transition was of interest. It has no bearing on the California Drought, but the data vis involves a geomapping, list, and scatter plot which we are considering to include in our final project.

13) USGS Release: Data-driven Insights on the California Drought (12/8/2014 8:33:13 AM). (n.d.). Retrieved April 11, 2015, from http://www.usgs.gov/newsroom/article.asp?ID=4069#.VSmUZZPK5aZ

a) This is an example of a well-done data vis using D3 which focuses on the California Drought. It definitely a big inspiration for us in our design process. We believe that our reservoir capacity/current level vis could tell a similar story in a more compelling way.

14) Virtual Water - Discover how much WATER we EAT everyday. (n.d.). Retrieved April 11, 2015, from http://www.angelamorelli.com/water/

a) This data vis was not only informative, but also, it contains a downwards scrolling transition between visualizations. It would certainly be a nice to have feature for our final project.

15) waterfootprint.org. (n.d.). Retrieved April 11, 2015, from http://waterfootprint.org/en/

a) The various interactive tools are very inspiring for their application of geomapping methods. The animation during the loading of data is done in good taste. It is definitely a welcome distraction of a transition as it fits the water theme nicely. It would be a nice to have feature between our transitions.

16) Start Using Landsat on AWS | AWS Official Blog. (n.d.). Retrieved April 11, 2015, from https://aws.amazon.com/blogs/aws/start-using-landsat-on-aws/

a) 'Landsat on AWS' includes over 85,000 shots of the US West region. It is the first time that so much satellite imagery is made available to the public online via Amazon Web Services. We are considering using the images to correspond to the declining reservoir levels over time. Each selection would contain an actual satellite image corresponding to the time selection.

17) List of dams and reservoirs in California - Wikipedia, the free encyclopedia. (n.d.). Retrieved April 17, 2015, from http://en.wikipedia.org/wiki/List_of_dams_and_reservoirs_in_California

a) This Wikipedia entry includes California dam trivia that could be of interest to the data visualization consumer. They have only tangential connections to the main topic, but they may provide the personal touch that could engage the viewers while they derive insights from our main visualizations.

QUESTIONS

For our data visualization project, we are trying to answer the following questions:

1) What is the state of California's water reservoirs in terms of utilization, location, and changes over time?

2) Where is California water being used and how can those use cases be categorized and broken out by volume?

3) Bonus: How do the types of agriculture in California affect water usage over time?

DATA

In general, data that comes from USGS have already been formatted in CSV and JSON which means that they will require minimal processing as they are structured.

For the purposes of achieving Objective #1, we would need to create two tables from raw files which are 'Daily Reservoir Utilization' and 'Reservoir Meta'. The fields required for 'Daily Reservoir Utilization' include <Reservoir ID>, <Storage Level>, <Date Recorded>. The fields required for 'Reservoir Meta' include <Reservoir ID>, <Storage Capacity>, <Longitude>, <Latitude>, and <Reservoir Name>. The tables will be aggregated on by <Year-Month> using the <Date Recorded> field. The <Average Storage> and <Capacity %> fields will be aggregated by <Year-Month>. The two tables will be joined on <Reservoir ID> as the key. Any reservoir outside of the top 10 capacity reservoirs will be grouped into an Others category. The raw files that enable the above operations are:

storage.json – Fields: 3, Records: 437,881

Reservoir ID	Date Recorded	Storage Level
SHA	1/1/2015	315,000
SHA	2/1/2015	285,000
SHA	3/1/2015	245,000

reservoir.json – Fields: 10, Records: 91

Reservoir ID	Storage Capacity	Longitude	Latitude	Reservoir Name
SHA	524,000	-23.212	27.7142	Shasta
ORO	324,000	-25.212	29.7142	Oroville

For the purposes of achieving Objective #2, we would need to create a '2010 CA Water Withdrawal Data' table. The fields that we would use are <Year>, <Usage>, <Ground or Surface>, <Saline or Fresh>,<Daily Volume>. The raw files that enable the above operations are:

usco2010.xlsx – Fields: 117, Records: 3,225

Year	Source	Usage	Saline or Fresh	Daily Volume
2010	Ground	Agriculture	Fresh	323,000
2010	Surface	Mining	Saline	28,000

Sankey Chart

We choose to use the Sankey chart to visualize the flow of water withdrawals in the state of California. This allows us to visualize what the drought will affect the most.

The data structure for a Sankey Chart consists of Sources, Targets, Values and Nodes.

Sources, Targets and Values define where the water is coming from, where it's going and what quantity of water. Water travels through different categories and levels.

1) Water Source:
   a) Surface Water (Reservoirs, Rivers, Creeks, Streams and Lakes)
   b) Ground Water (Pumps and Aquifers)

2) Water Type:
   a) Fresh (Water with a low concentration of dissolved salt and solids, potable water and water that's not from the sea)
   b) Saline (Water containing salt, not potable)

3) Water Use:
   a) Public Supply
   b) Domestic
   c) Industrial
   d) Irrigation
   e) Irrigation Crop
   f) Irrigation Golf
   g) Livestock
   h) Aquaculture
   i) Mining
   j) Thermoelectric
   k) Thermoelectric once-through
   l) Thermoelectric recirculation

The quantity of water is measured in Millions of Gallons per Day (Mgal/d). By determining how many millions of gallons each reservoir holds you can easily calculate how much water every category is using.

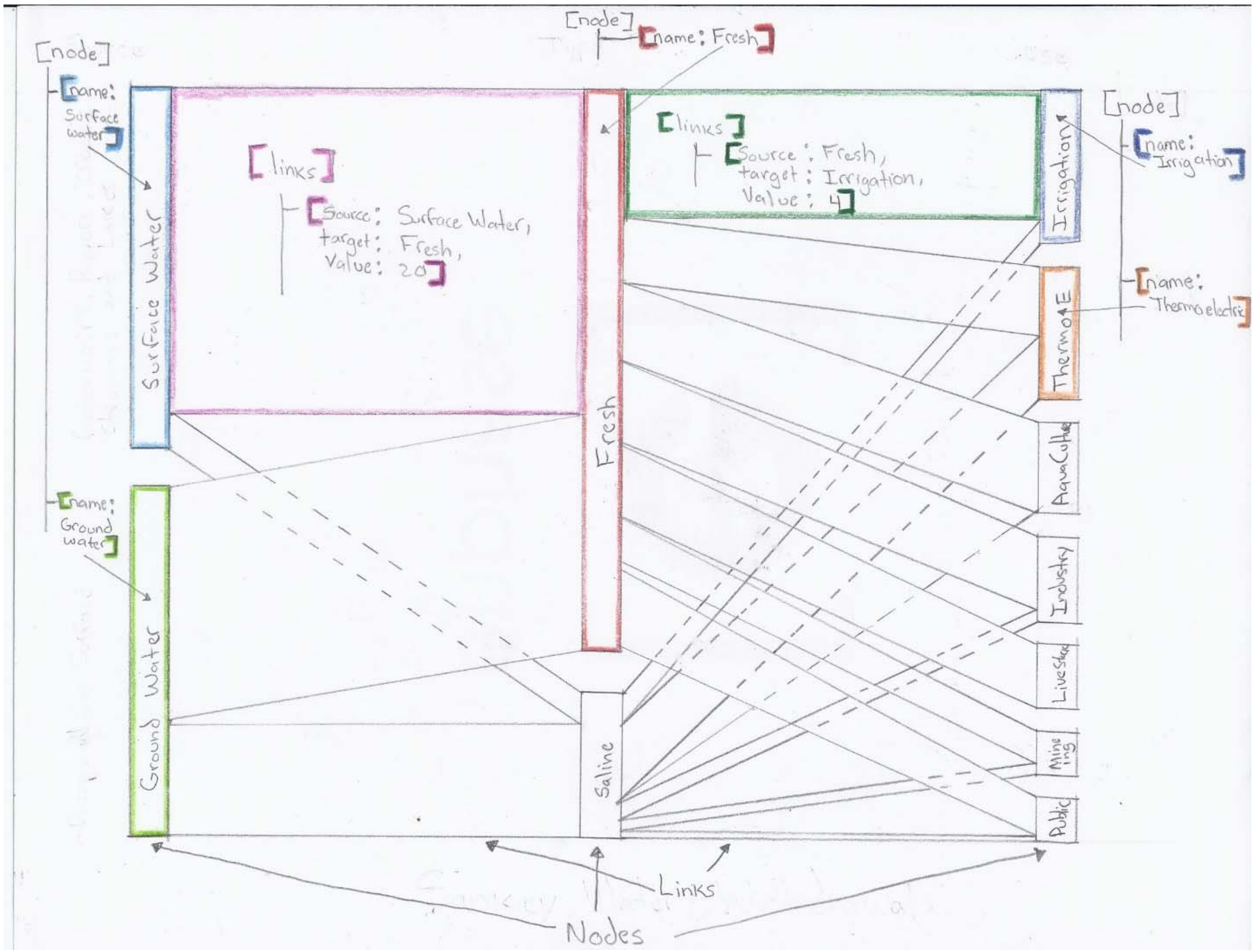
Links are defined as the segments between the nodes, the width of each link is determined by the Value (Mgal/d). Attributes of a link are: Source, Target, and Value.

Nodes are defined as the labeled blocks that are connected by the links. The height of each node is determined by the sum of all the values linking to the node. Attributes of a node are: Name.

An example of data layout for Sankey Chart:

{"links": [
{"source":"Surface Water","target":"Fresh","value":"20"},
{"source":"Fresh","target":"Irrigation","value":"4"},
] ,
"nodes": [
{"name":"Surface Water"},

{"name":"Fresh"},
] }

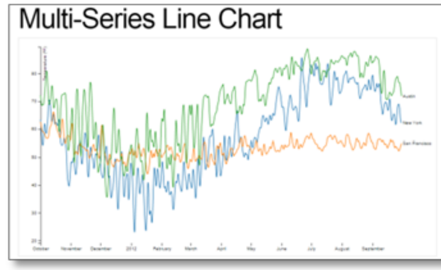


In order to visualize our data, we had to implement a large amount of data preparation, clean up, and wrangling. We began by looking for charts that would fit our models and designs. For the most intriguing charts, we would deconstructed them, and determine how data flows through from the raw to the visualization.

We came to the conclusion that preparing our data to fit the model of the chart would be more efficient and clean, rather than building visualizations from scratch around our source data structures.

1) Line Chart that displays reservoir capacity utilization over time
   a) Reference -- http://bl.ocks.org/mbostock/3884955

```
Data Structure from the Example
  {
    name: name,
    values: {
          date: d.date,
          temperature: +d[name]
          }
  }
```



b) Raw Data from USGS
   i) In our original data set, we contained dates that were nested alongside the storage level data. We also had multiple fields that were not relevant to our visualization.

```
Reservoir.json
[
        {
                "Station": "ANTELOPE LAKE",
                "ID": "ANT",
                "Elev": 4960,
                "Latitude": 40.18,
                "Longitude": -120.607,
                "County": "PLUMAS",
                "Nat_ID": "CA00037",
                "Year_Built": 1964,
                "Capacity": 22566,
                "Storage": {
                        "20000104": 15313.6428571429,
                        "20000111": 15240.7142857143,
                        "20000118": 15423,
                        "20000125": 15627.8571428571,
                        :
                        "20140826": 18292.1428571429,
                        "20140902": 17945.8571428571,
                        "20140909": 17598.4285714286,
                        "20140916": 17292.6666666667
                }
        },
        {
                "Station": "BEAR",
                "ID": "BAR",
                "Elev": 319,
                "Latitude": 37.367,
                "Longitude": -120.217,
                "County": "MARIPOSA",
```

Date was originally part of the storage label.

c) Processed Data
   i) After processing the data and cleaning it up. Were able to parse out Date and Storage into separate fields. We also renamed fields and removed fields that were not needed.

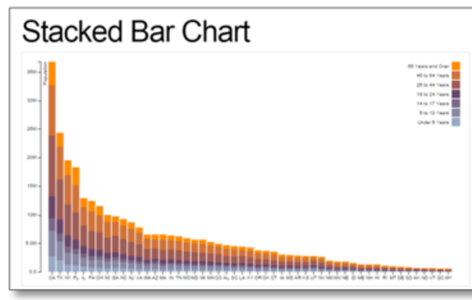
Reservoir_processed.json


```
[
    {
        "name": "ANTELOPE LAKE",
        "capacity": 22566,
        "latitude": 40.18,
        "longitude": -120.607,
        "values": [
            {
                "date": "20000104",
                "storage": 15313.6428571429
            },
            {
                "date": "20000118",
                "storage": 15423
            },
            {
                "date": "20140916",
                "storage": 17292.6666666667
            }
        ]
    },
    {
        "name": "BEAR",
        "capacity": 7700
```


Fields now have their own labels, "date", and "storage".

d) Stacked Bar Charts displays reservoir utilization

i) Reference -- http://bl.ocks.org/mbostock/3886208



Data Structure in Example

Each object in the data represents each bar.

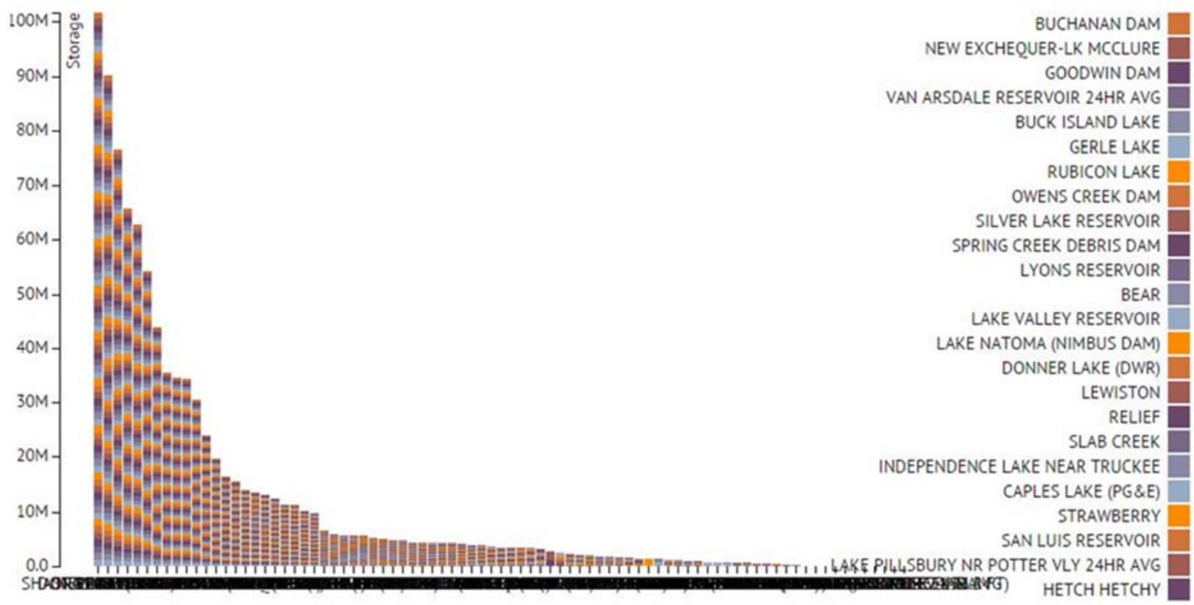
Each object represents a segment (rectangle in graph).
"y0" represents "y" coordinate at the bottom of the segment.
"y1" represents "y" coordinate at the top of the segment.

If there are 2 segments, with a values of 50 and 40.
The first segment will have a "y0"=0, and "y1"=50.
Then second segment will add the prior segment and have values of "y0"=50, and "y1"=90

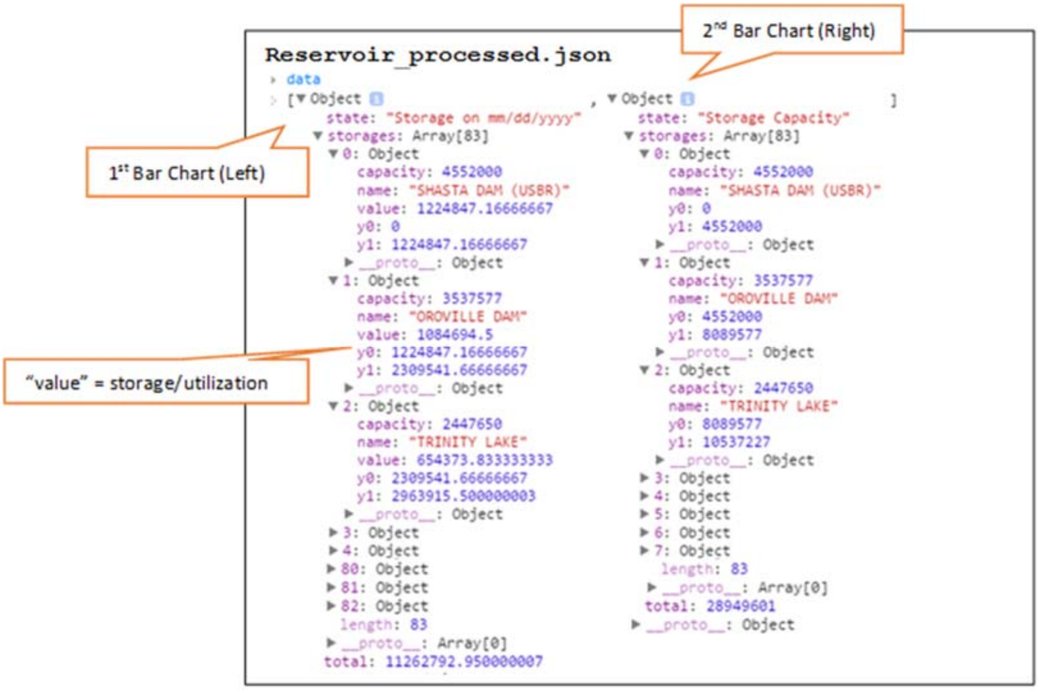
ii) Original Data (From file 1.Line Chart)

```
Reservoir_processed.json
[
	{
		"name": "ANTELOPE LAKE",
		"capacity": 22566,
		"latitude": 40.18,
		"longitude": -120.607,
		"values": [
			{
				"date": "20000104",
				"storage": 15313.6428571429
			},
			{
				"date": "20000118",
				"storage": 15423
			}
```


iii) When loading our unprocessed data into the bar chart, we received some un-expected results and proceeded to clean the data further and prepare it for the bar chart.

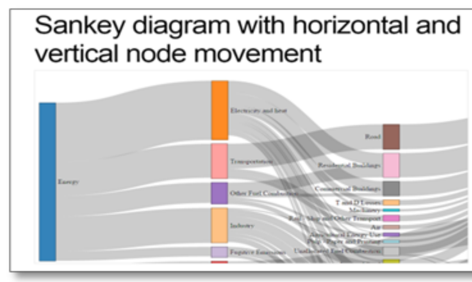


e) Processed Data



f) Water Withdrawal Sankey Chart

i) Reference -- http://bl.ocks.org/d3noob/5028304



(1) Future Enhancement: Display a drop down or Bar Chart (Bar Chart would be sorted by water usage by state) to allow users to select a state to view the water withdrawals for the selected state and compare it with California's water withdrawals.

Data Structure in Example


```
{
"links": [
{"source":"Agricultural Energy Use","target":"Carbon Dioxide","value":"1.4"},
{"source":"Agriculture","target":"Agriculture Soils","value":"5.2"},
{"source":"Agriculture","target":"Livestock and Manure","value":"5.4"},
{"source":"Agriculture","target":"Other Agriculture","value":"1.7"},

{"source":"Waste","target":"Waste water - Other Waste","value":"1.5"},
{"source":"Waste water - Other Waste","target":"Methane","value":"1.2"},
{"source":"Waste water - Other Waste","target":"Nitrous Oxide","value":"0.3"}
] ,
"nodes": [
{"name":"Energy"},
{"name":"Industrial Processes"},
{"name":"Electricity and heat"},
{"name":"Industry"},
```


ii) Original Data

(1) Usco2010.xlsx, "CountyData" saved as a CSV file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	STATE	STATEFIPS	COUNTY	COUNTYFIP	FIPS	YEAR	TP-TotPop	PS-GWPop	PS-SWPop	PS-TOPop	PS-WGWFr	PS-WGWSa	PS-WGWTo	PS-WSWFr	PS-WSWSa	PS-WSWTo	PS-WFrTo
2	AL	01	Autauga County	001	01001	2010	54.571			48.222	5.09	0.00	5.09	0.00	0.00	0.00	5.09
3	AL	01	Baldwin County	003	01003	2010	182.265			153.463	22.97	0.00	22.97	0.00	0.00	0.00	22.97
4	AL	01	Barbour County	005	01005	2010	27.457			25.555	4.15	0.00	4.15	0.00	0.00	0.00	4.15
5	AL	01	Bibb County	007	01007	2010	22.915			21.279	4.89	0.00	4.89	0.00	0.00	0.00	4.89
6	AL	01	Blount County	009	01009	2010	57.322			44.464	2.44	0.00	2.44	52.17	0.00	52.17	54.61
7	AL	01	Bullock County	011	01011	2010	10.914			10.176	2.30	0.00	2.30	0.00	0.00	0.00	2.30
8	AL	01	Butler County	013	01013	2010	20.947			17.599	2.70	0.00	2.70	0.00	0.00	0.00	2.70
9	AL	01	Calhoun County	015	01015	2010	118.572			112.390	20.83	0.00	20.83	2.47	0.00	2.47	23.30
10	AL	01	Chambers County	017	01017	2010	34.215			25.875	0.00	0.00	0.00	4.31	0.00	4.31	4.31
11	AL	01	Cherokee County	019	01019	2010	25.989			17.876	2.53	0.00	2.53	0.96	0.00	0.96	3.49

(2) Usco2010.xlsx, "DataDictionary" saved as a CSV file

	A	B	C	D	E
1	**Column Tag**	**Data Element**	Source	Type	Use
2	STATE	State postal abbreviation			
3	STATEFIPS	State FIPS code			
4	COUNTY	County name			
5	COUNTYFIPS	County FIPS code			
6	FIPS	Concatenated State-county FIPS code			
7	YEAR	Year of data=2010			
8	TP-TotPop	Total population of county, in thousands			
9	PS-GWPop	Public Supply, population served by groundwater, in thousands			
10	PS-SWPop	Public Supply, population served by surface water, in thousands			
11	PS-TOPop	Public Supply, total population served, in thousands			
12	PS-WGWFr	Public Supply, groundwater withdrawals, fresh, in Mgal/d	Ground	Fresh	Public Supply
13	PS-WGWSa	Public Supply, groundwater withdrawals, saline, in Mgal/d	Ground	Saline	Public Supply
14	PS-WGWTo	Public Supply, groundwater withdrawals, total, in Mgal/d			
15	PS-WSWFr	Public Supply, surface-water withdrawals, fresh, in Mgal/d	Surface	Fresh	Public Supply
16	PS-WSWSa	Public Supply, surface-water withdrawals, saline, in Mgal/d	Surface	Saline	Public Supply
17	PS-WSWTo	Public Supply, surface-water withdrawals, total, in Mgal/d			
18	PS-WFrTo	Public Supply, total withdrawals, fresh, in Mgal/d			
19	PS-WSaTo	Public Supply, total withdrawals, saline, in Mgal/d			

We added Source, Type, and Use.
Based on "Data Element" column.

iii) Processed Data (Step 1)

aggregatedUsageData_processed.json


```
{
        "PS-WGWFr": 2742.200000000001,
        "PS-WGWSa": 84.29,
        "PS-WSWFr": 3472.2299999999996,
        "PS-WSWSa": 0,
        "DO-WGWFr": 142.47000000000003,
        "DO-WSWFr": 29.449999999999999,
        "IN-WGWFr": 399.2699999999999,
        "IN-WGWSa": 0,
        "IN-WSWFr": 1.1300000000000001,
        "IN-WSWSa": 0,
        "IR-WGWFr": 8685.98,
        "IR-WSWFr": 14370.509999999995,
        "IR-IrSpr": 1792.5000000000005,
        "IR-IrMic": 2892.8500000000004,
        "IR-IrSur": 5665.98,
        "IC-WGWFr": 8553.369999999999,
        "IC-WSWFr": 14290.059999999992,
        "IC-IrSpr": 1701.2500000000002,
        "IC-IrMic": 2892.8500000000004,
```

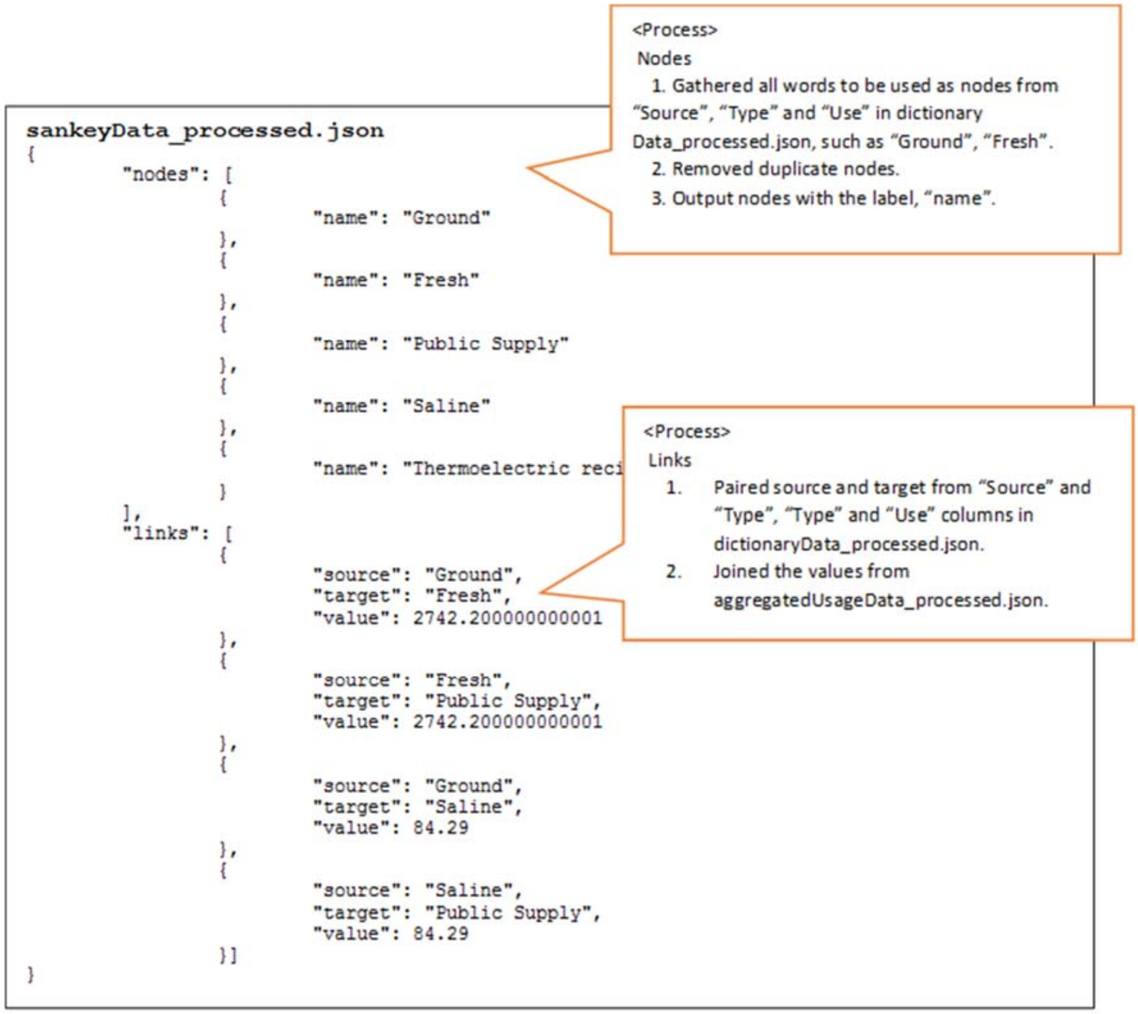

<Process>
1. Removed unused columns.
2. Filter to California (STATE="CA")
3. Aggregated all County rows together.

<Process>
Removed all unused columns.

dictionaryData_processed.json


```
[
        {
                "ColumnTag": "PS-WGWFr",
                "Data Element": "Public Supply,  groundwater withdrawals, fresh, in Mgal/d",
                "Source": "Ground",
                "Type": "Fresh",
                "Use": "Public Supply",
                "Lvl4": ""
        },
        {
                "ColumnTag": "PS-WGWSa",
                "Data Element": "Public Supply, groundwater withdrawals, saline, in Mgal/d",
                "Source": "Ground",
                "Type": "Saline",
                "Use": "Public Supply",
                "Lvl4": ""
        },
```


iv) Processed Data (Step 2)



EXPLORATORY DATA ANALYSIS

Figures 1-10 were created using screenshots of data visualizations created on the QlikView platform.

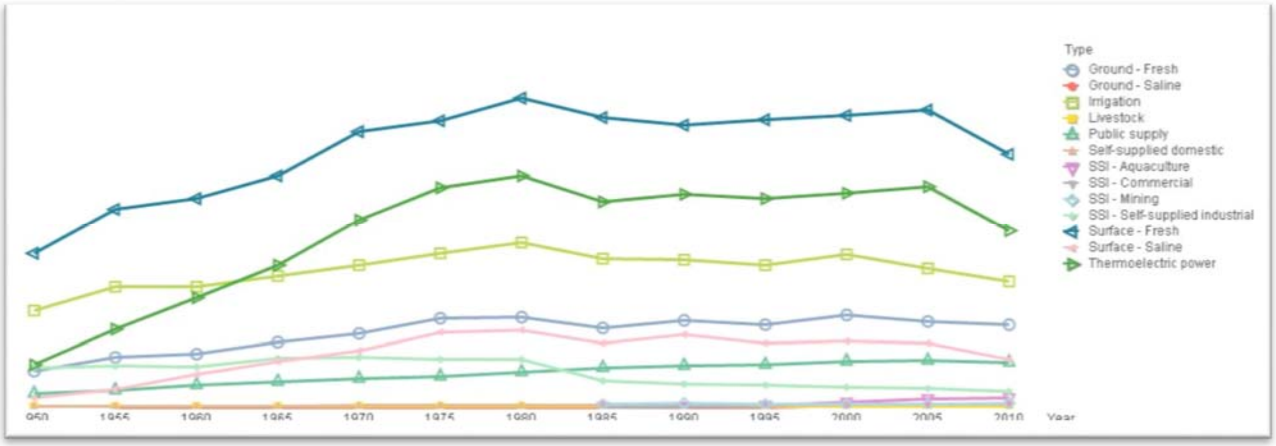


Figure 1

In Figure 1, we wanted to see how water withdrawal categories changed over time. The data was provided by USGS and was recorded every five years. It was interesting to see the rise of Thermoelectric power from the 1950's and plateauing in the 1970's.

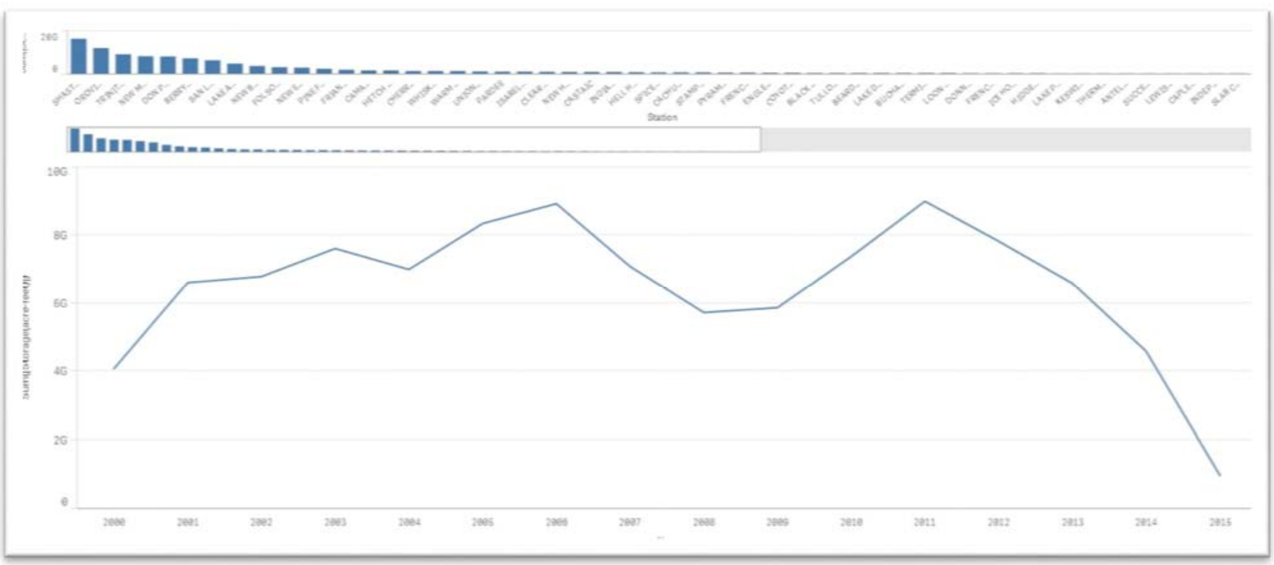


Figure 2

In Figure 2, we are looking at reservoir level data over time. Each reservoir in California is broken out in the bar chart while the line chart shows aggregate reservoir level using the brushing tool in between the two graphs. It is quite apparent that since 2011, the reservoir levels have dropped off dramatically to levels not seen since the early 2000's.

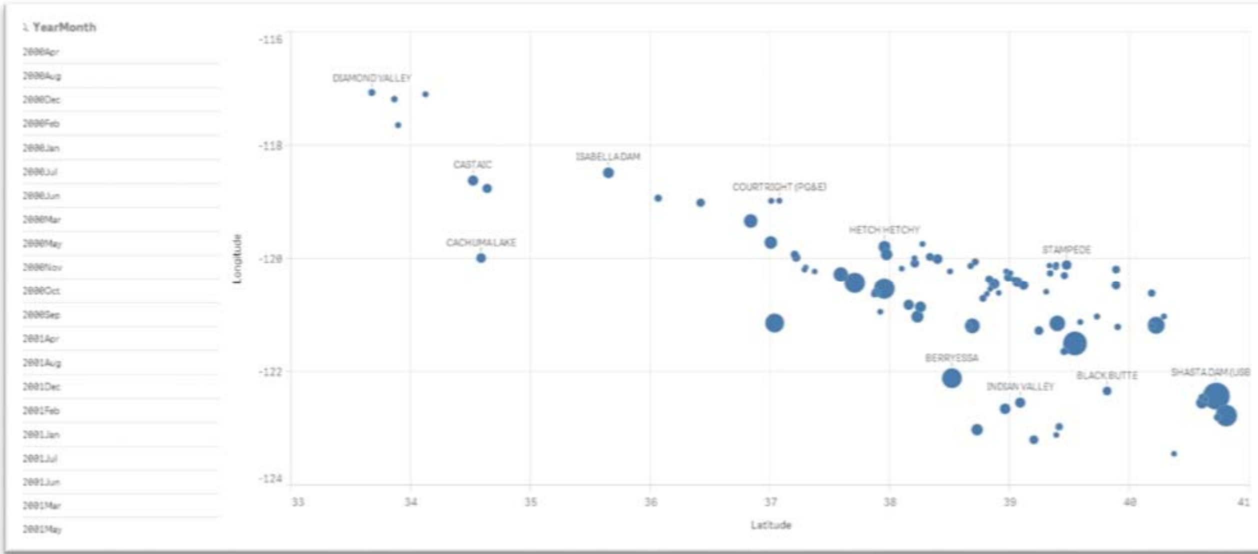


Figure 3

In Figure 3, a bubble chart is created to visualize relative locations based on geographical coordinates on a Cartesian plane of the reservoirs and their capacities. Shasta is the largest of the California reservoirs by volume.

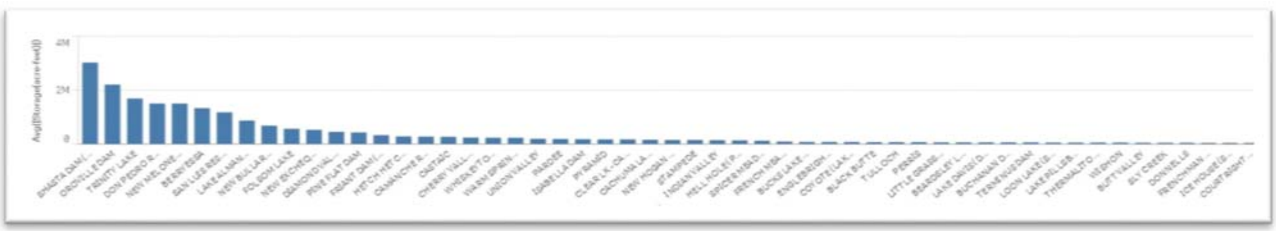


Figure 4

In Figure 4, we graphed the California reservoirs by name using the average water level over time as the dependent variable.

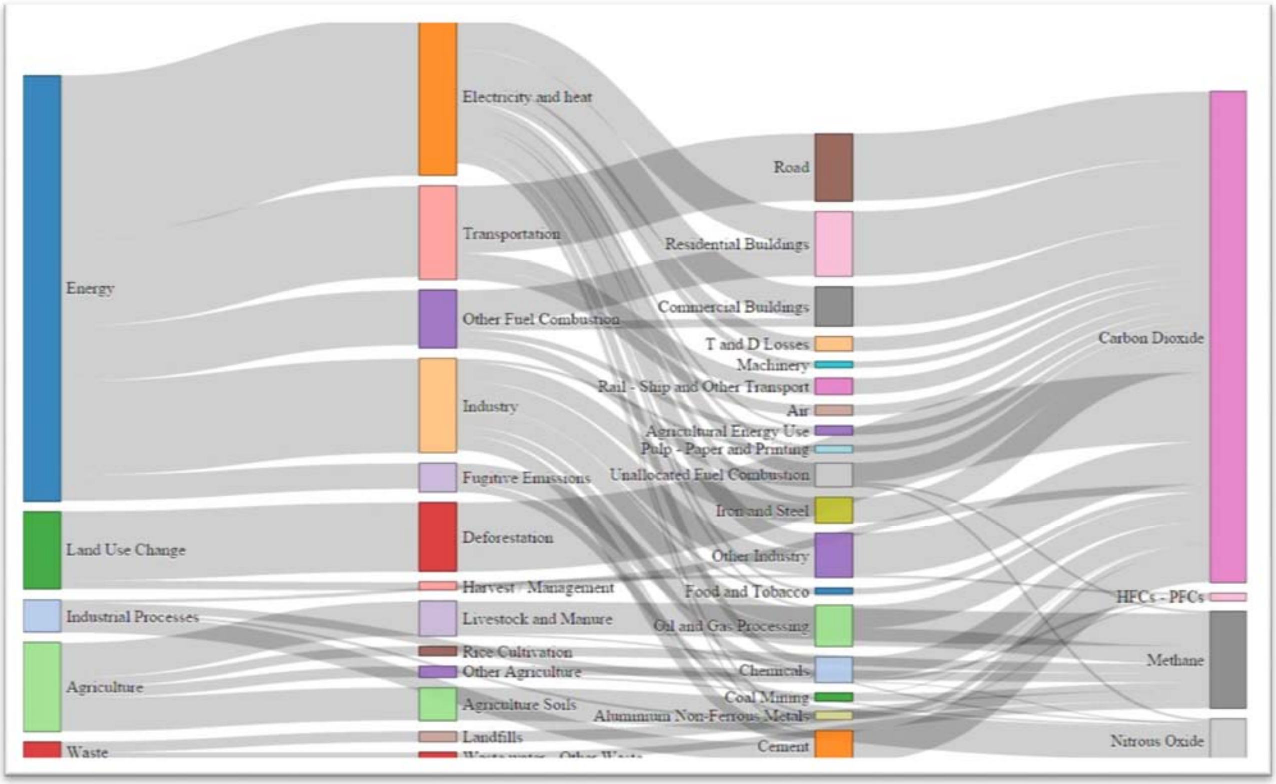


Figure 5

In Figure 5, we used a Sankey graph representation of the water withdrawal category data.

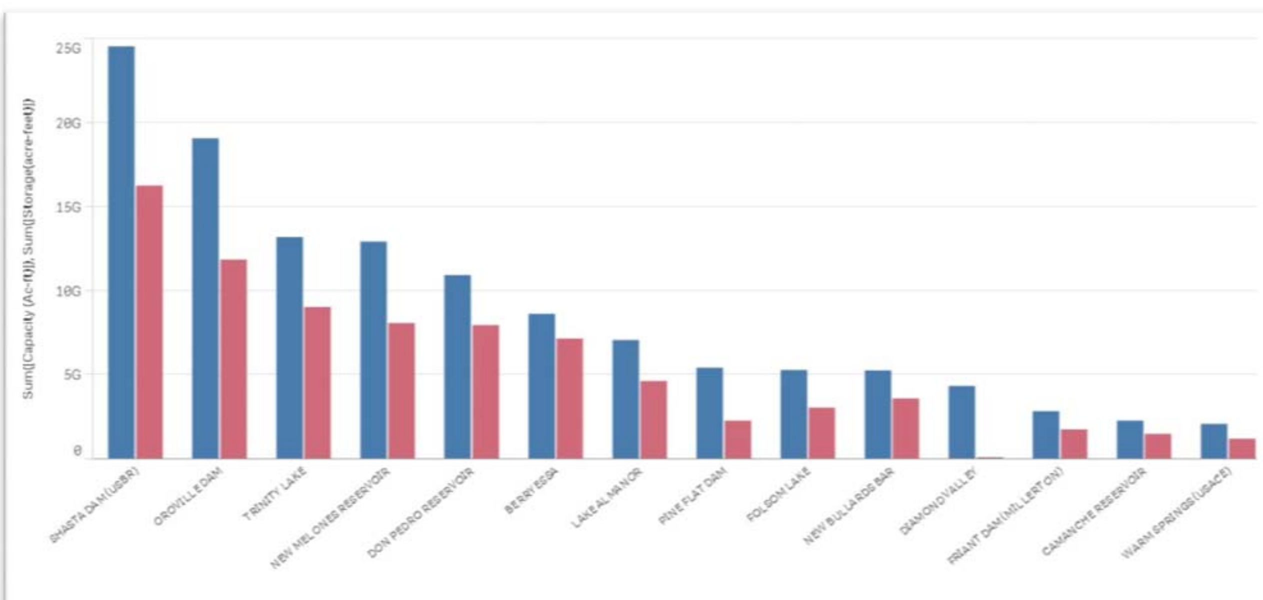


Figure 6

In Figure 6, we sought to show the total storage capacity versus total water levels. In retrospect, it may have made more sense to use an average as opposed to the sum because the capacity remains constant for each reservoir. Summing the capacities may be confusing for the end users.

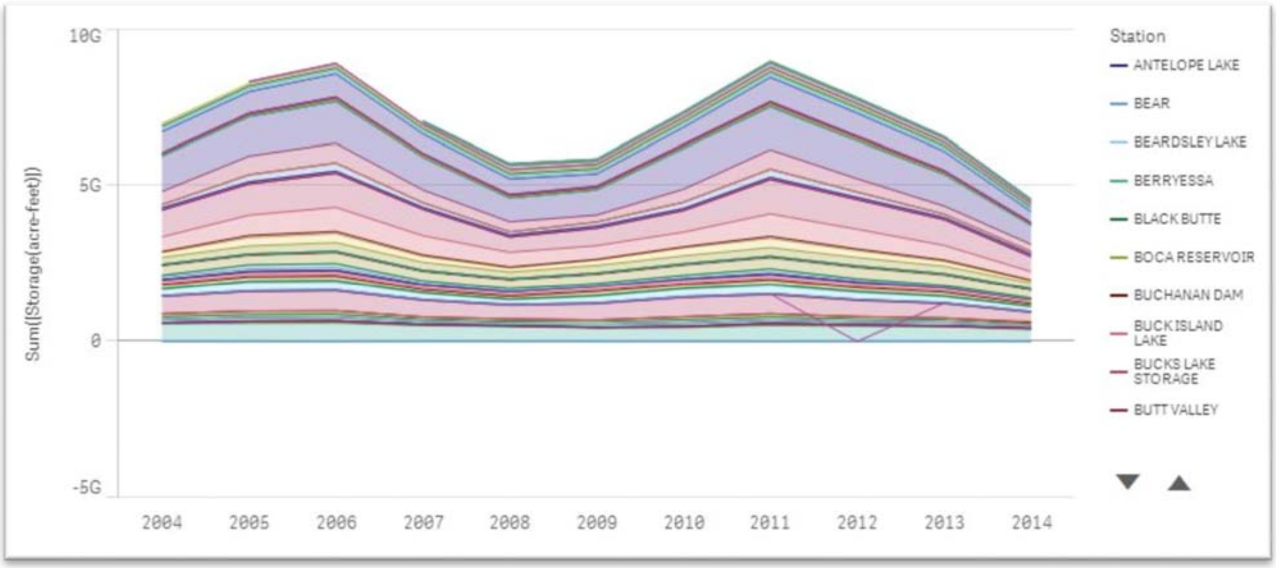


Figure 7

Figure 7 is a visualization of a stacked area chart displaying the contributions of each reservoir level over time.

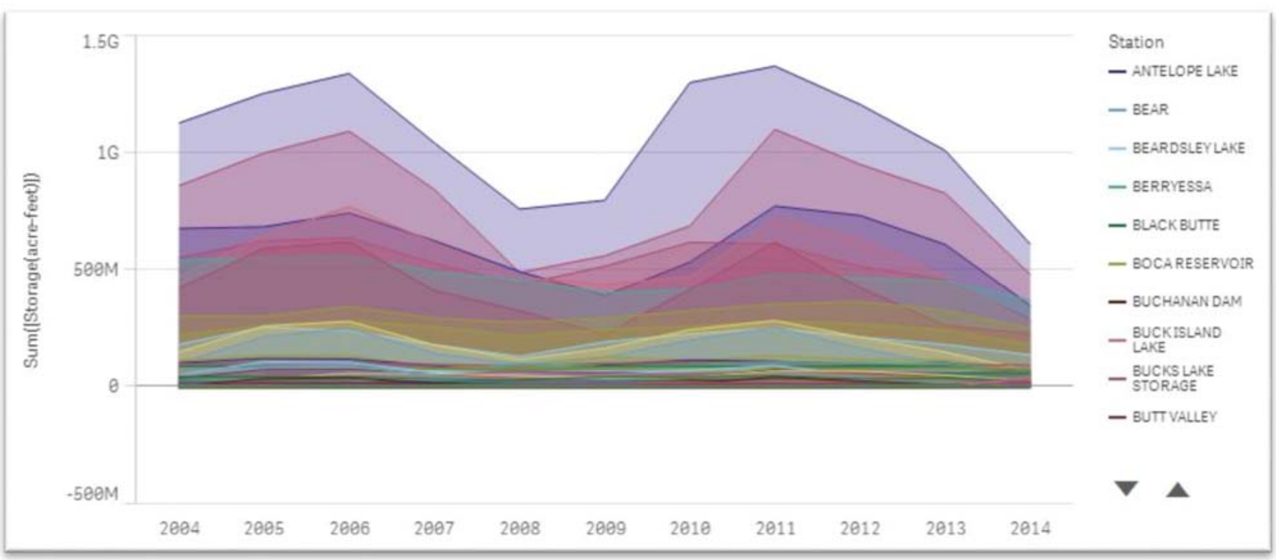


Figure 8

Figure 8 is an overlay of stacked area charts for each reservoir level over time.

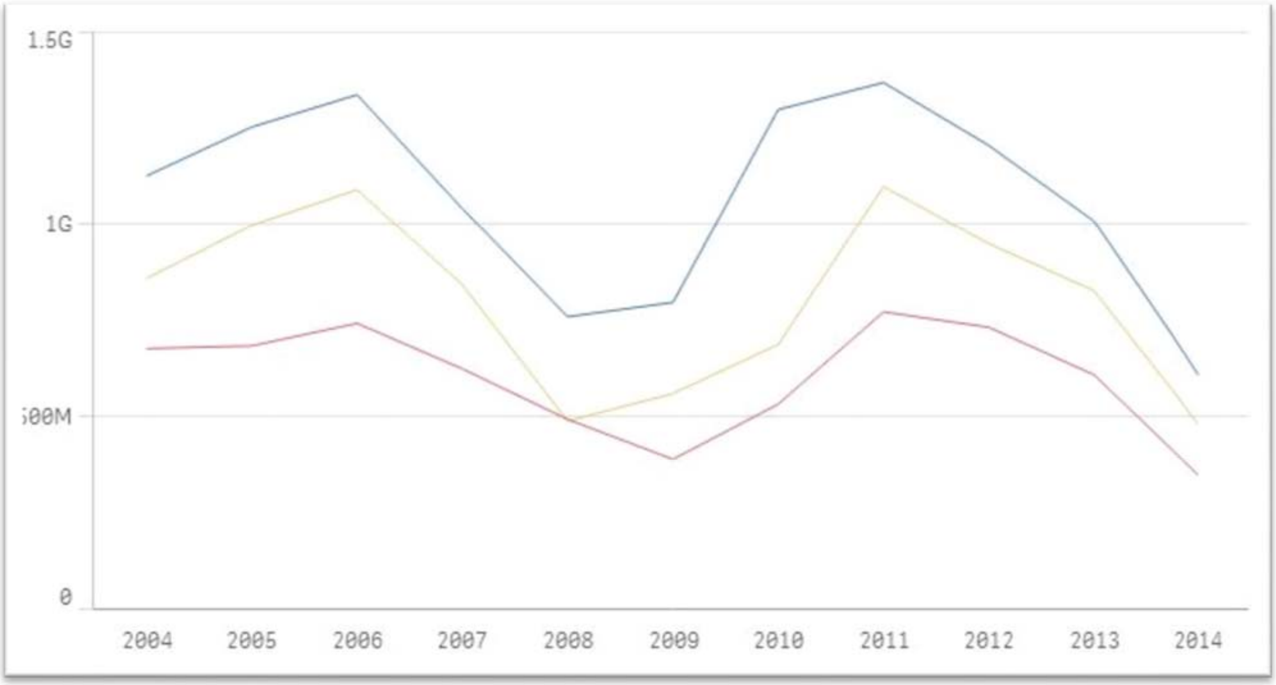


Figure 9

Figure 9 is an unshaded area chart of the top 3 reservoirs by sorted by greatest volume over time in descending order. It was an experiment to see if this would be a less cluttered view as opposed to a shaded area graph for every reservoir water level over time.

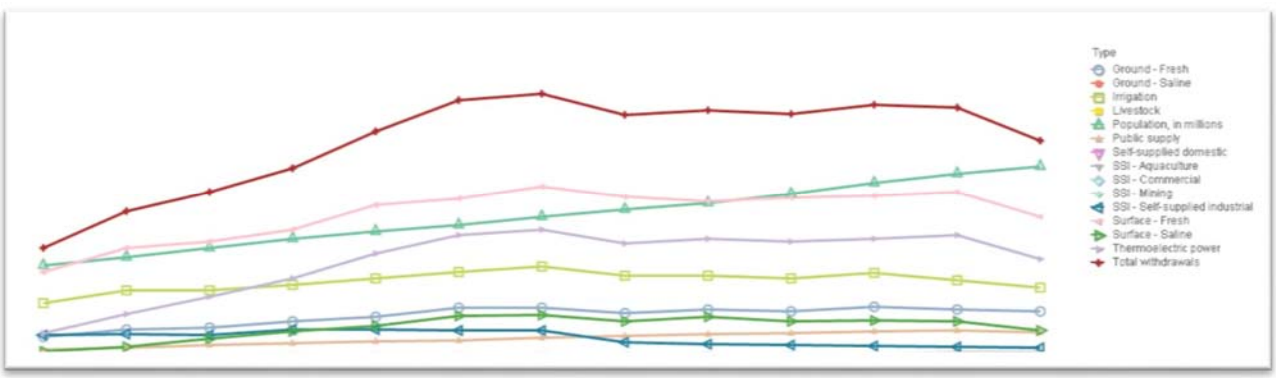


Figure 10

Figure 10 is a simple line graph that shows the trending of water withdrawal categories over time.

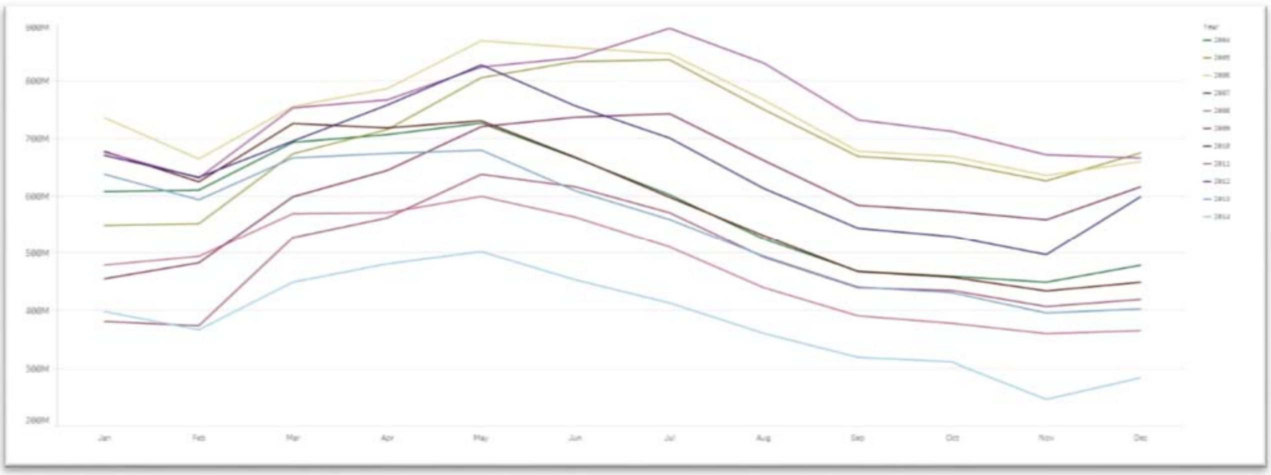
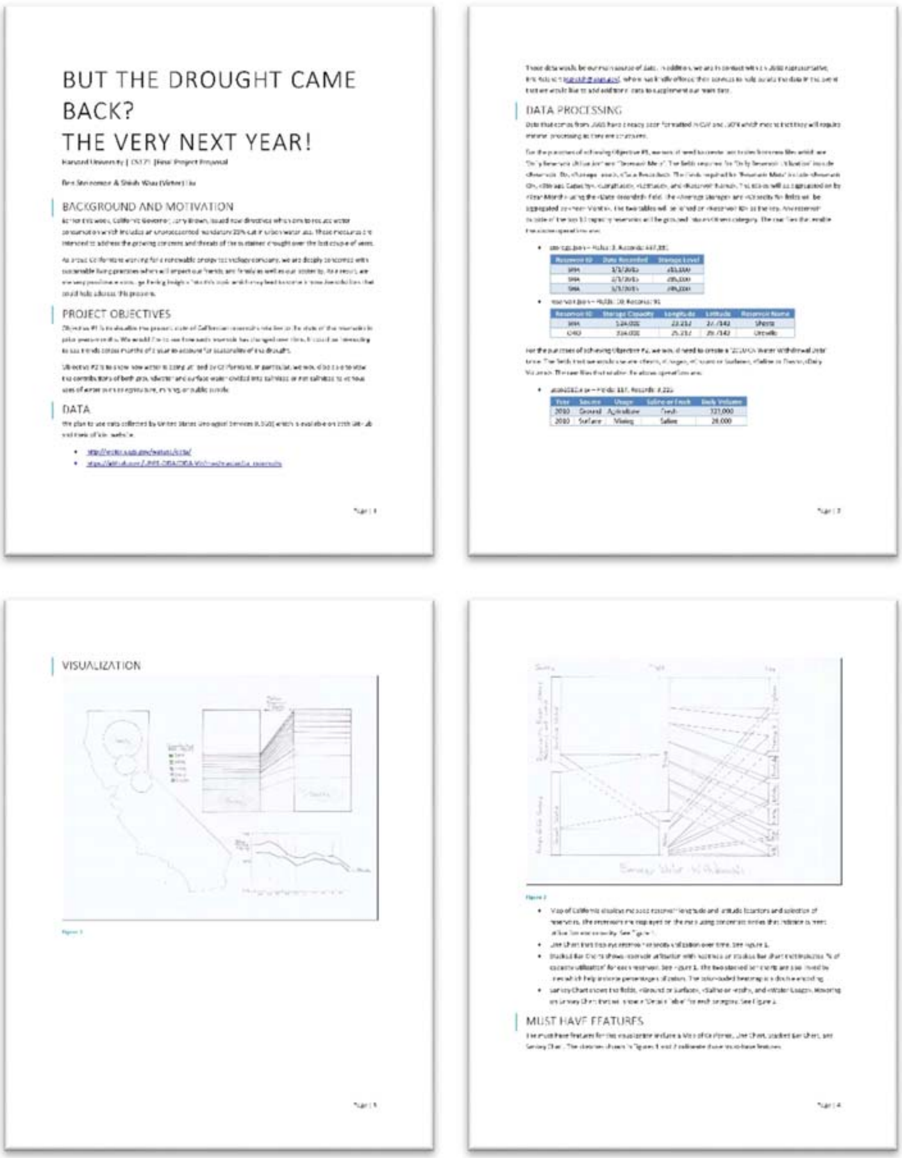


Figure 11

Figure 11 is a line chart comparing all reservoir capacities YoY for the past 10 years. The independent variable is Month while the dependent variable is Year.

DESIGN EVOLUTION

Our initial reasons for pursuing the California Drought as the subject matter and focus of our final project were delineated along with our initial designs and project component scheduling. This formed the majority of our content for the Project Proposal which is duplicated in Figures 11 and 12:



BUT THE DROUGHT CAME BACK?
THE VERY NEXT YEAR!

BACKGROUND AND MOTIVATION

PROJECT OBJECTIVES

DATA

Page | 1

DATA PROCESSING

Page | 2

VISUALIZATION

Page | 3

MUST HAVE FEATURES

Page | 4

Figure 11

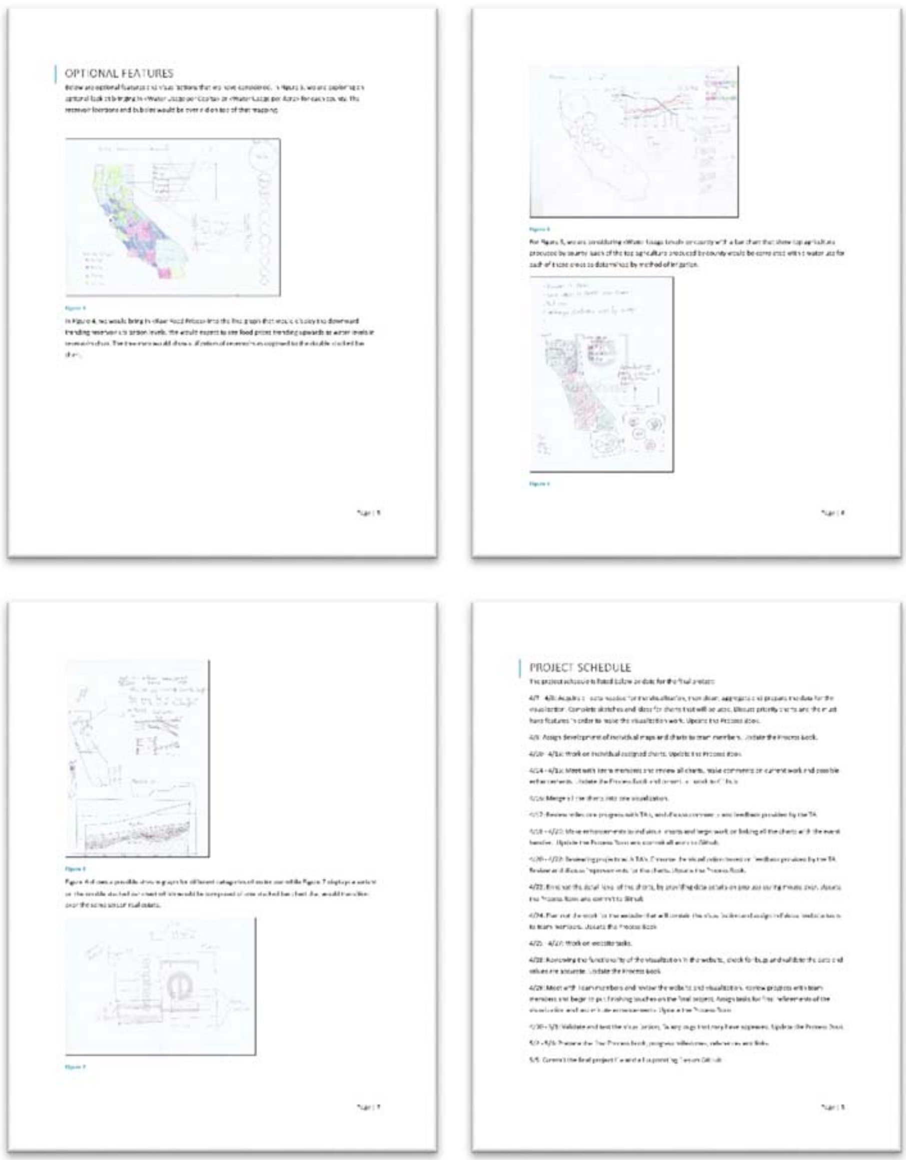


Figure 12

Our Final Project Proposal original design drawings are expanded below in Figures 14 and 15.

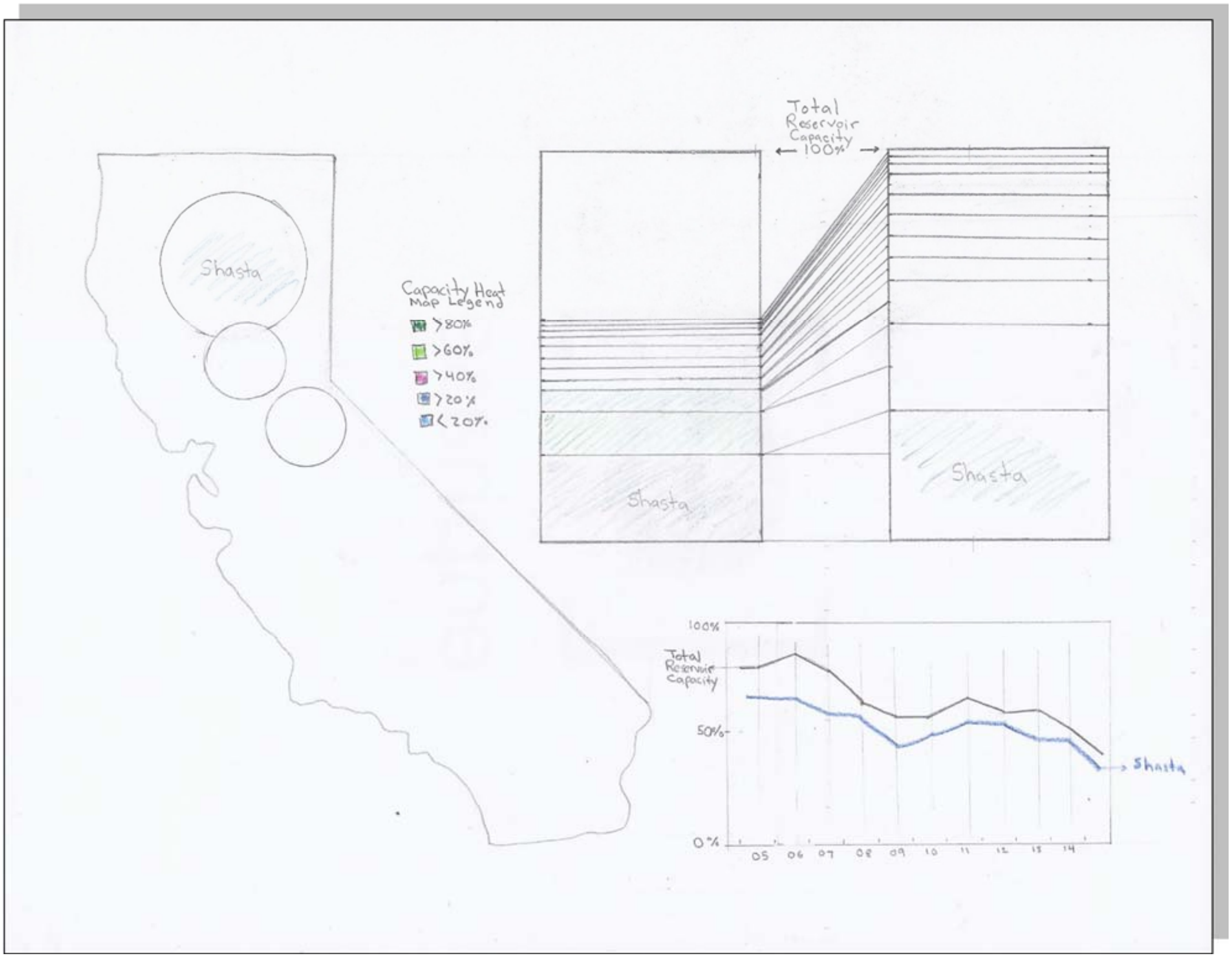


Figure 13

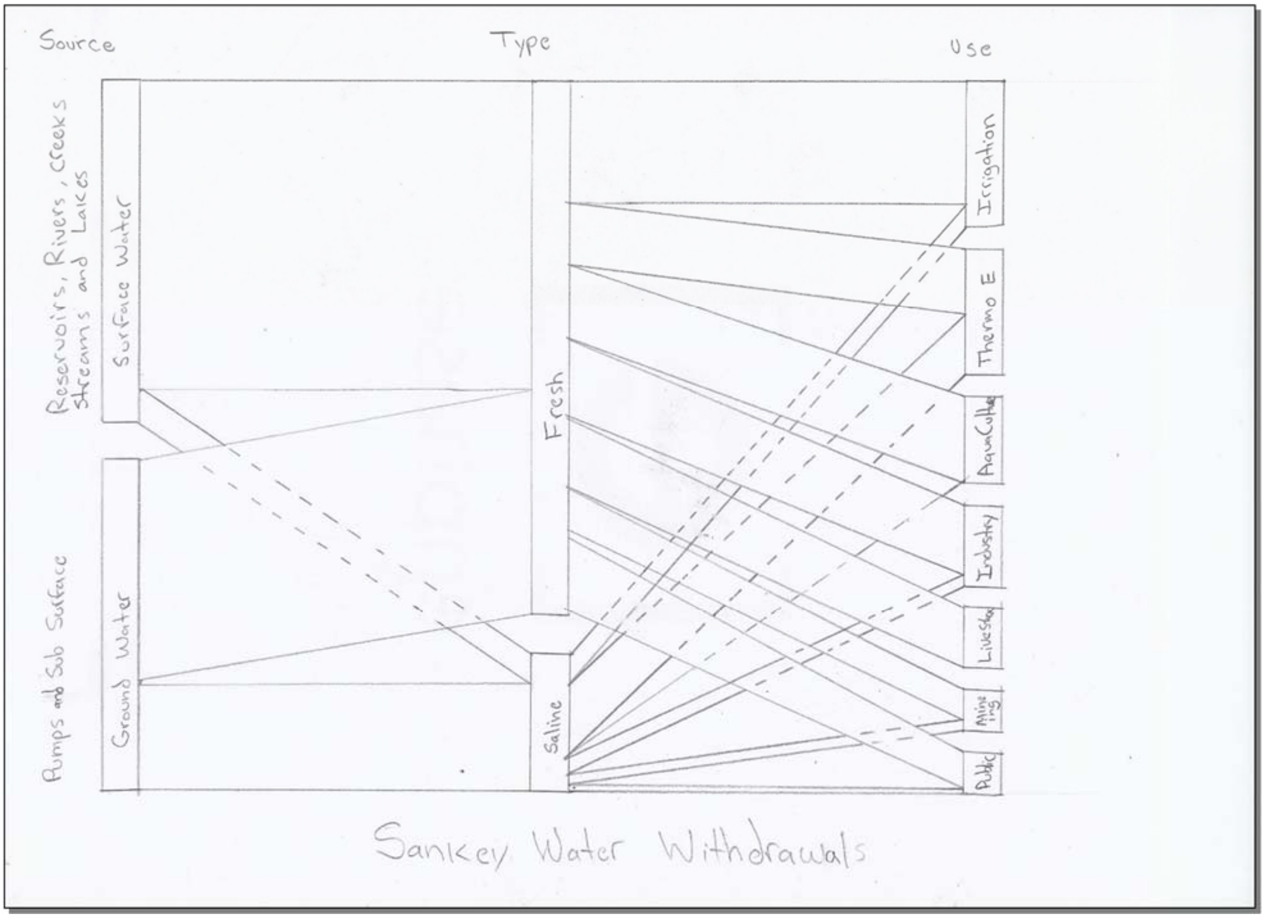
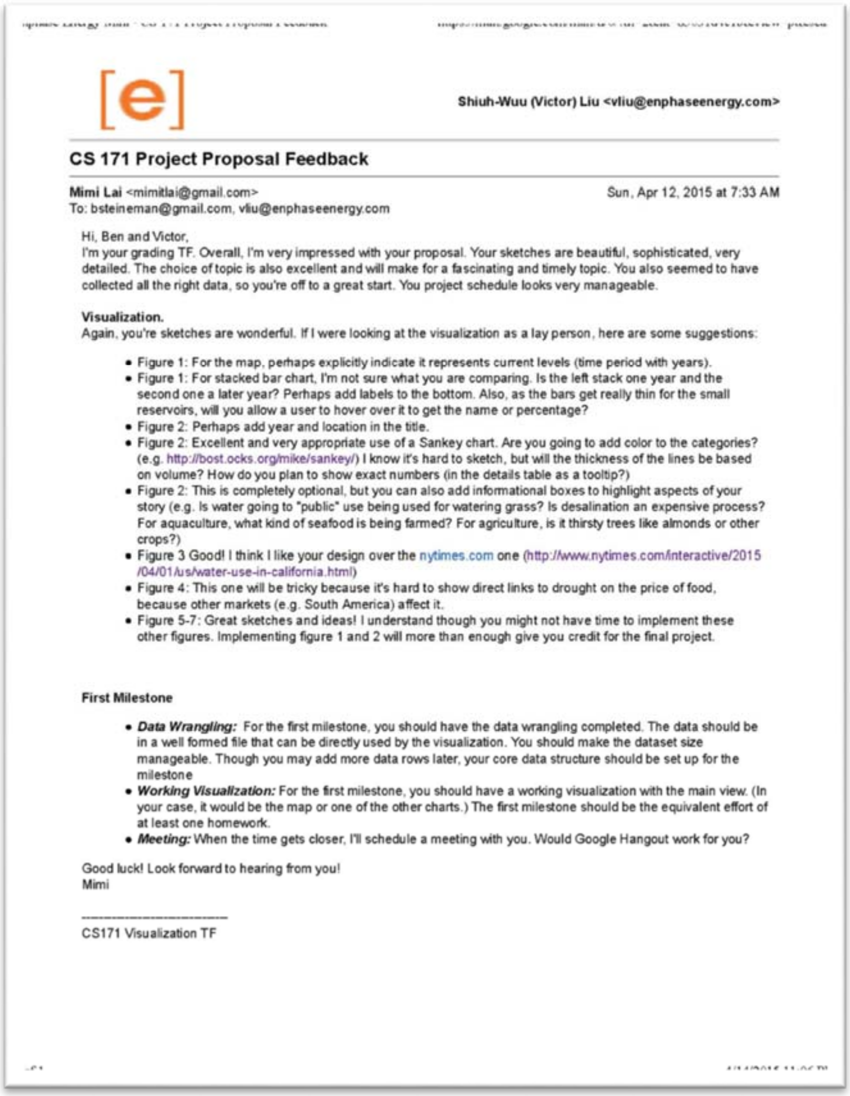


Figure 14

1) Map of California displays mapped reservoir longitude and latitude locations and selection of reservoirs. The reservoirs are displayed on the map using concentric circles that indicate current utilization and capacity. See Figure 13.
2) Line Chart that displays reservoir capacity utilization over time. See Figure 13.
3) Stacked Bar Charts shows reservoir utilization with heatmap on stacked bar chart that indicates '% of capacity utilization' for each reservoir. See Figure 13. The two stacked bar charts are also linked by lines which help indicate percentage utilization. The color-coded heatmap is a double encoding.
4) Sankey Chart shows the fields, <Ground or Surface>, <Saline or Fresh>, and <Water Usage>. Hovering on Sankey Chart that will show a 'Details Table' for each category. See Figure 14.

One of the TF's, Mimi Lai, reviewed our Project Proposal and gave us the following feedback as shown in Figure 15.



Shiuh-Wuu (Victor) Liu <vliu@enphaseenergy.com>

CS 171 Project Proposal Feedback

Mimi Lai <mimitlai@gmail.com> Sun, Apr 12, 2015 at 7:33 AM
To: bsteineman@gmail.com, vliu@enphaseenergy.com

Hi, Ben and Victor,
I'm your grading TF. Overall, I'm very impressed with your proposal. Your sketches are beautiful, sophisticated, very detailed. The choice of topic is also excellent and will make for a fascinating and timely topic. You also seemed to have collected all the right data, so you're off to a great start. You project schedule looks very manageable.

Visualization.
Again, you're sketches are wonderful. If I were looking at the visualization as a lay person, here are some suggestions:

- Figure 1: For the map, perhaps explicitly indicate it represents current levels (time period with years).
- Figure 1: For stacked bar chart, I'm not sure what you are comparing. Is the left stack one year and the second one a later year? Perhaps add labels to the bottom. Also, as the bars get really thin for the small reservoirs, will you allow a user to hover over it to get the name or percentage?
- Figure 2: Perhaps add year and location in the title.
- Figure 2: Excellent and very appropriate use of a Sankey chart. Are you going to add color to the categories? (e.g. http://bost.ocks.org/mike/sankey/) I know it's hard to sketch, but will the thickness of the lines be based on volume? How do you plan to show exact numbers (in the details table as a tooltip?)
- Figure 2: This is completely optional, but you can also add informational boxes to highlight aspects of your story (e.g. Is water going to "public" use being used for watering grass? Is desalination an expensive process? For aquaculture, what kind of seafood is being farmed? For agriculture, is it thirsty trees like almonds or other crops?)
- Figure 3 Good! I think I like your design over the nytimes.com one (http://www.nytimes.com/interactive/2015/04/01/us/water-use-in-california.html)
- Figure 4: This one will be tricky because it's hard to show direct links to drought on the price of food, because other markets (e.g. South America) affect it.
- Figure 5-7: Great sketches and ideas! I understand though you might not have time to implement these other figures. Implementing figure 1 and 2 will more than enough give you credit for the final project.

First Milestone

- ***Data Wrangling:*** For the first milestone, you should have the data wrangling completed. The data should be in a well formed file that can be directly used by the visualization. You should make the dataset size manageable. Though you may add more data rows later, your core data structure should be set up for the milestone
- ***Working Visualization:*** For the first milestone, you should have a working visualization with the main view. (In your case, it would be the map or one of the other charts.) The first milestone should be the equivalent effort of at least one homework.
- ***Meeting:*** When the time gets closer, I'll schedule a meeting with you. Would Google Hangout work for you?

Good luck! Look forward to hearing from you!
Mimi

CS171 Visualization TF

Figure 15

Additionally, we met with a representative, Joshua Darlington, from another project group to exchange feedback on our projects. Joshua was interested in our topic and, in fact, in our conversations it would appear that he was fairly well read up on the latest happenings of the California drought. He had some reservations in showing the declining reservoir levels as he believes that they are fairly self-evident.

We respectfully disagree with his assessment because reservoir levels are directly related to how much water is available for usage. As a result, we have decided to go ahead with our design as the intended audience is the general public whom may or may not be knowledgeable regarding this burning topic.

Joshua also mentioned that the drought is not only going to affect California, but also this will impact the entire Country. Per his suggestion we thought it would be of interest to our end users if we gave the option to show water usage filtered by all US states instead of only California. This would allow users to interact with our Sankey visualization and give some context on a national level by enabling the comparisons of water withdrawals trends in California to water withdrawal trends in other US states.

Our initial stab at the designs included all of the basic forms of our visualizations including the double stacked bar chart, line chart, and Sankey chart as shown in Figure 16.

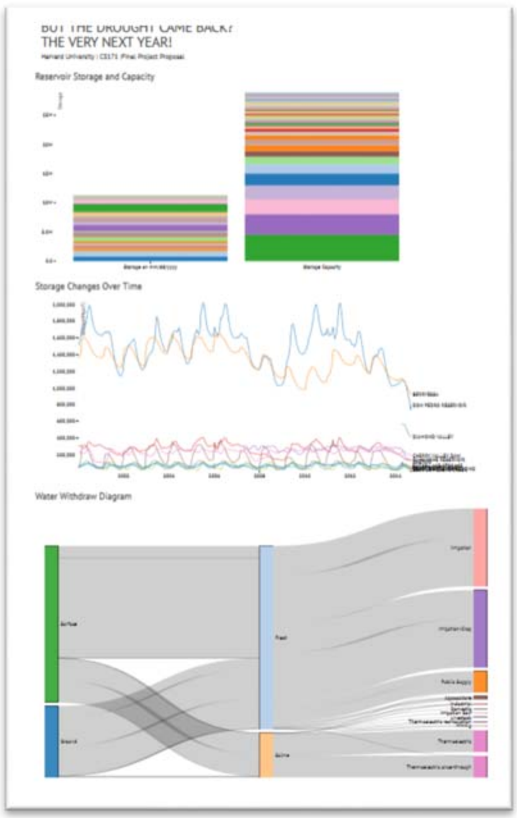


Figure 16

In Figure 17, we wanted to present a more high resolution view of the top graph in order to highlight the high number of reservoirs in California. We had considered displaying top 10 reservoirs by capacity and then grouping the remainder reservoirs into an 'Other' category, but after spending some time weighing the pros and cons, we have determined that showing all of the reservoirs with the given color selections.

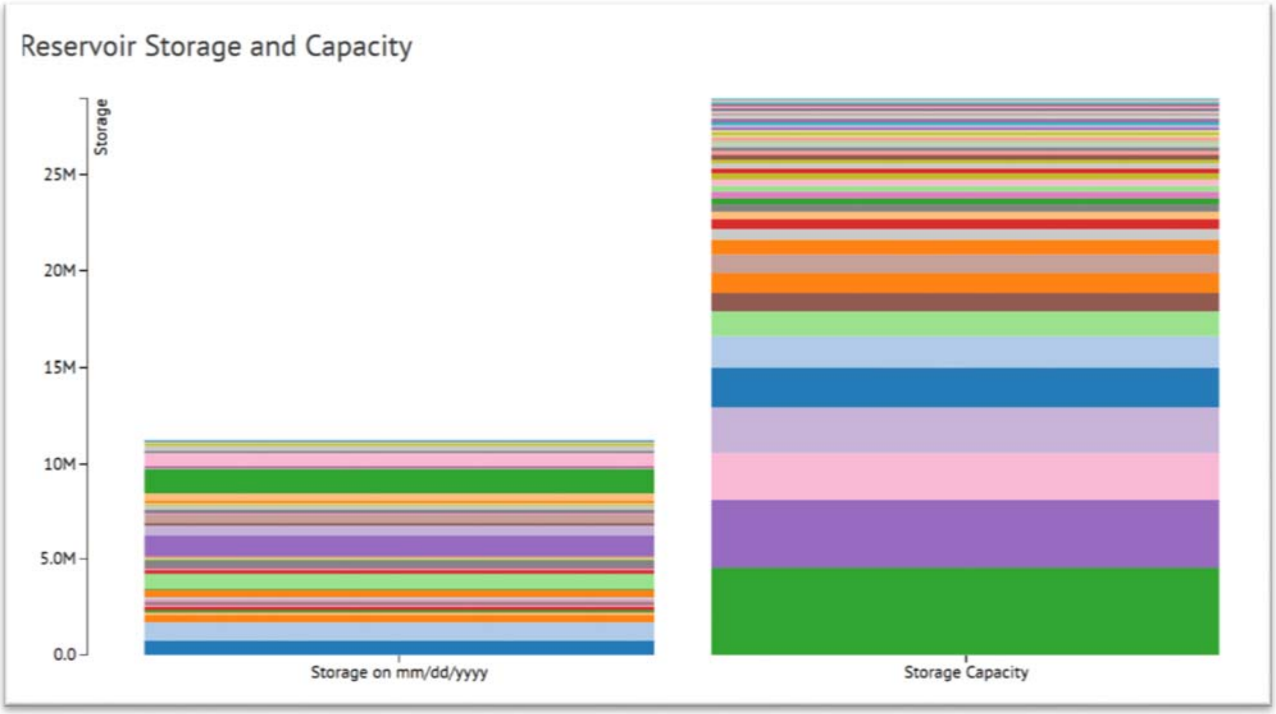


Figure 17

In Figure 18, we are showing the top 10 reservoirs by capacity. It is with the intention of reducing clutter in our visualization. In this particular version, we are experimenting with the omission of the aforementioned long tail reservoirs that are outside of the top 10. It makes more sense in this case as opposed to the stacked bar charts in Figure 17 because the reservoir capacities are superimposed on top of each other. In other words, the total reservoir capacities in California are not meant to be inferred from this chart.

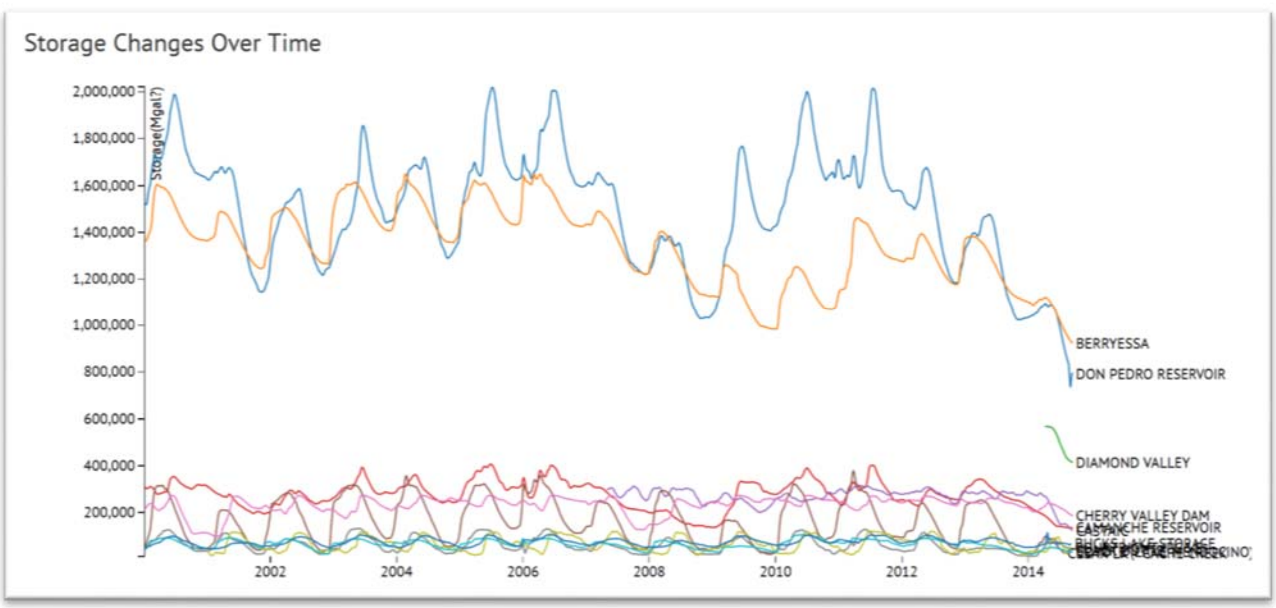


Figure 18

Figure 19, is shows how the Sankey chart could be manipulated to clear up relationships through the feature that allows for the rearrangement of each visual channel.

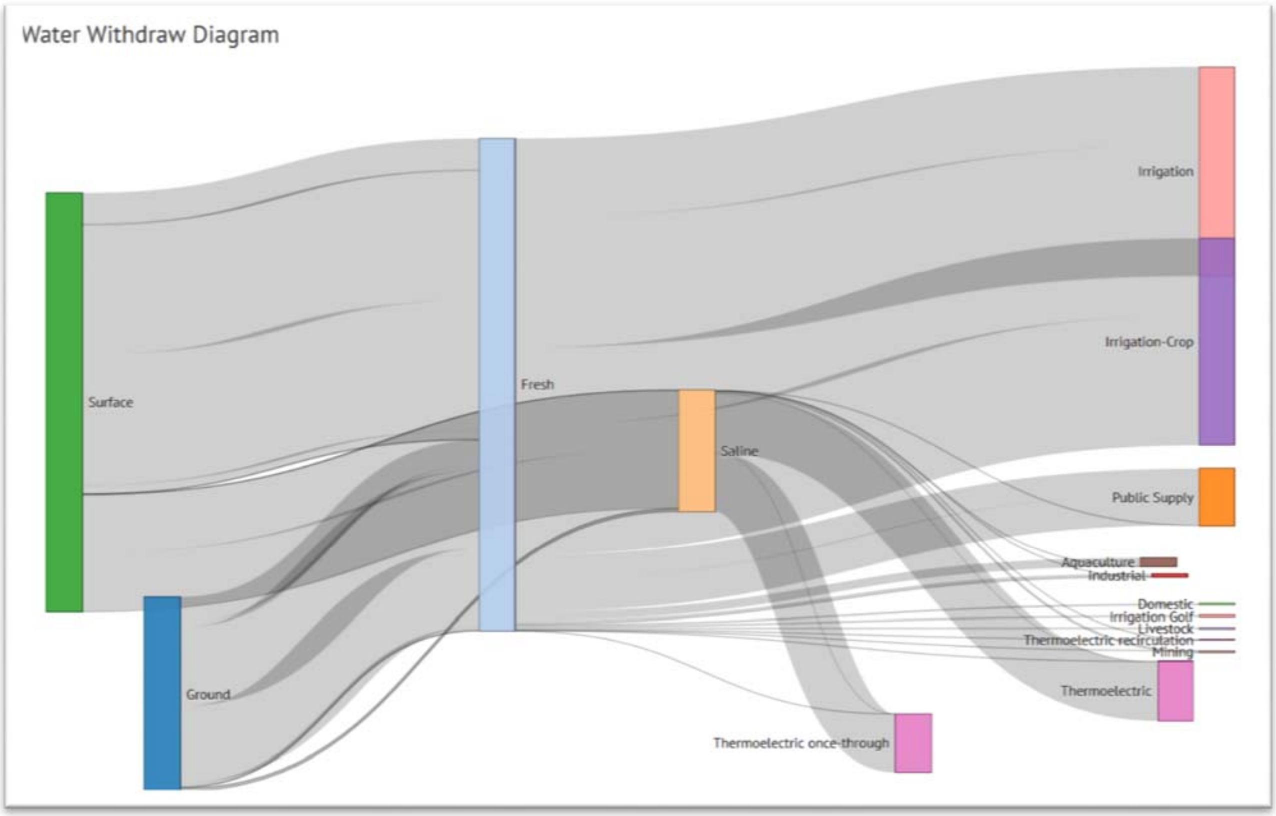


Figure 19

IMPLEMENTATION

To be continued…

EVALUATION

To be continued…