BUT THE DROUGHT CAME BACK?
THE VERY NEXT YEAR!

Harvard University | CS171 |Final Project Proposal

Ben Steineman & Shiuh-Wuu (Victor) Liu

BACKGROUND AND MOTIVATION

Earlier this week, 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.

PROJECT OBJECTIVES

Objective #1 is to visualize the present state of Californian reservoirs relative to the state of the reservoirs in prior years-months. We would like to see how each reservoir has changed over time. It could be interesting to see trends across months of a year to account for seasonality of the drought.

Objective #2 is to show how water is being utilized by Californians. In particular, we would be able to view the contributions of both groundwater and surface water divided into salinized or not salinized to various uses of water such as agriculture, mining, or public supply.

DATA

We plan to use data collected by United States Geological Services (USGS) which is available on both GitHub and their official website.

- http://water.usgs.gov/watuse/data/
- https://github.com/USGS-CIDA/CIDA-Viz/tree/master/ca_reservoirs

Those data would be our main source of data. In addition, we are in contact with an USGS representative, Eric Reichart (egreich@usgs.gov), whom has kindly offered their services to help curate the data in the event that we would like to add additional data to supplement our main data.

DATA PROCESSING

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

VISUALIZATION

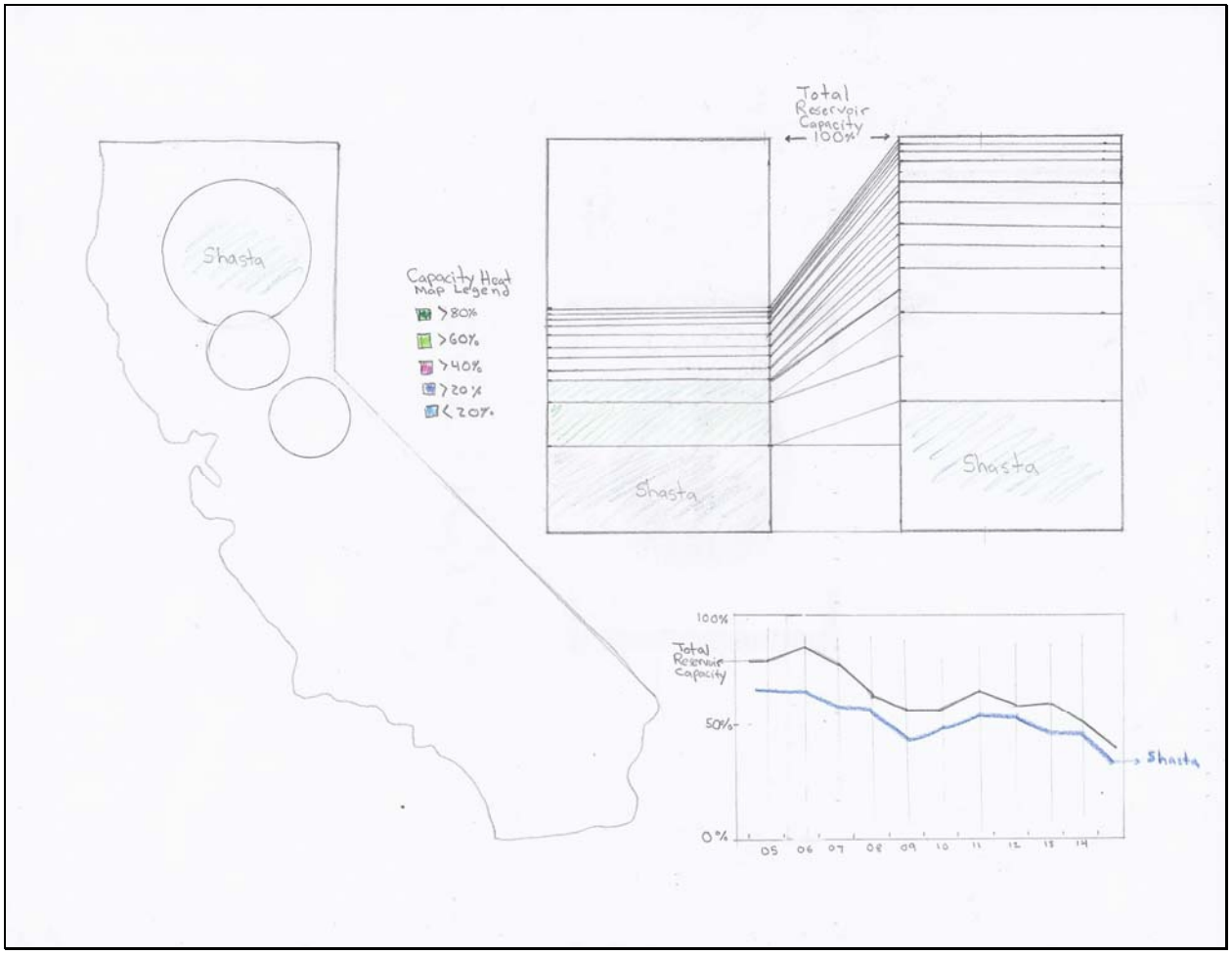

Figure 1

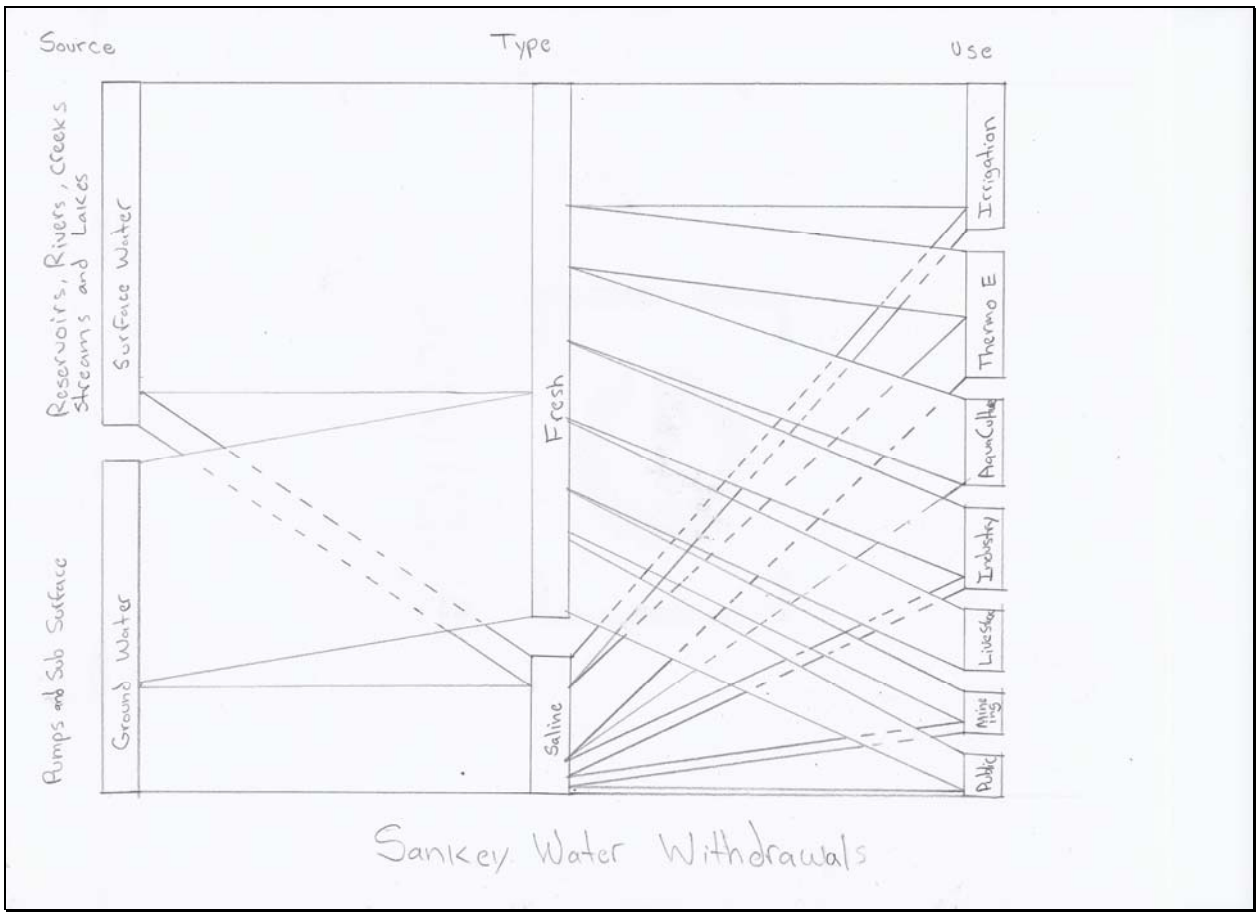

Figure 2

- 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 1.
- Line Chart that displays reservoir capacity utilization over time. See Figure 1.
- Stacked Bar Charts shows reservoir utilization with heatmap on stacked bar chart that indicates '% of capacity utilization' for each reservoir. See Figure 1. The two stacked bar charts are also linked by lines which help indicate percentage utilization. The color-coded heatmap is a double encoding.
- 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 2.

MUST HAVE FEATURES

The must-have features for this visualization include a Map of California, Line Chart, Stacked Bar Chart, and Sankey Chart. The sketches shown in Figures 1 and 2 delineate those must-have features.

OPTIONAL FEATURES

Below are optional features and visualizations that we have considered. In Figure 3, we are exploring an optional look at bringing in <Water Usage per Capita> or <Water Usage per Acre> for each county. The reservoir locations and bubbles would be overlaid on top of that mapping.

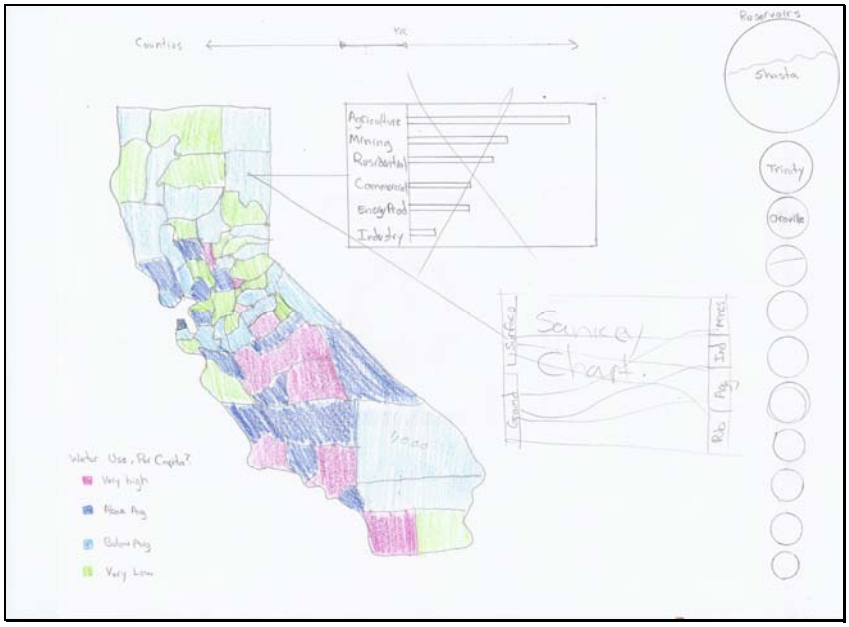

Figure 3

In Figure 4, we would bring in <Raw Food Prices> into the line graph that would display the downward trending reservoir utilization levels. We would expect to see food prices trending upwards as water levels in reservoirs drop. The tree map would show utilization of reservoirs as opposed to the double stacked bar chart.

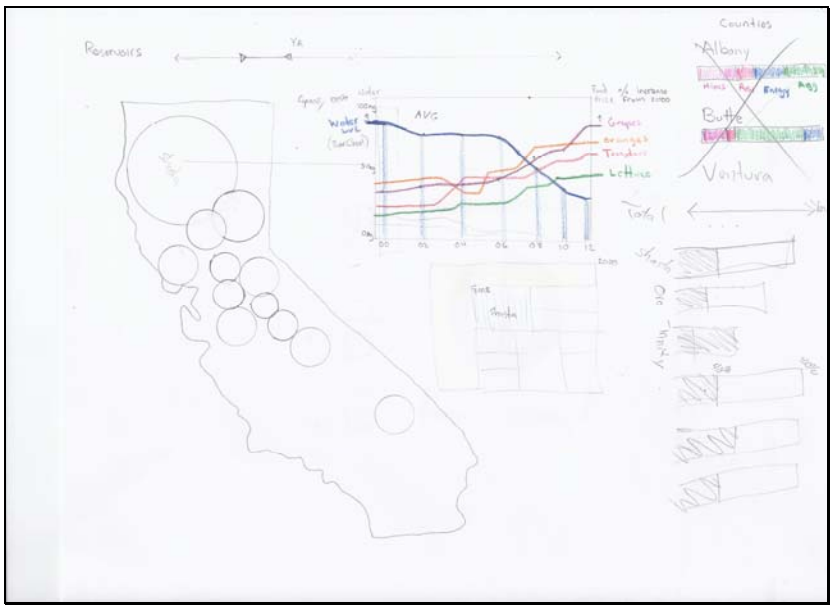

Figure 4

For Figure 5, we are considering <Water Usage Level> by county with a bar chart that show top agriculture produced by county. Each of the top agriculture produced by county would be correlated with a water use for each of those crops as determined by method of irrigation.

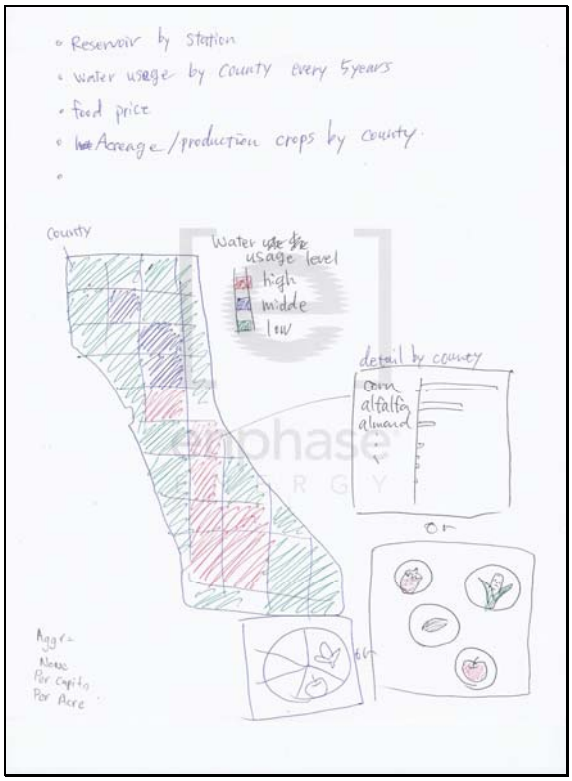

Figure 5

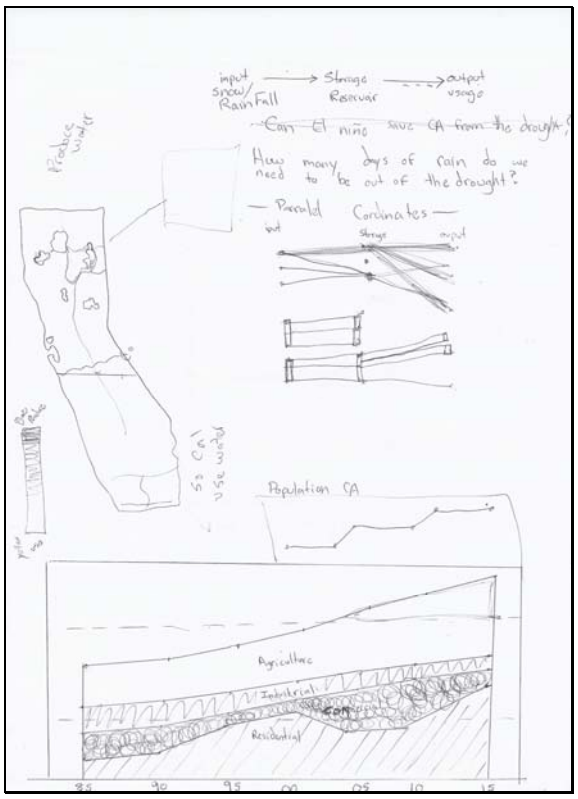

Figure 6

Figure 6 shows a possible stream graph for different categories of water use while Figure 7 displays a variant on the double stacked bar chart which would be composed of one stacked bar chart that would transition over the same screen real estate.

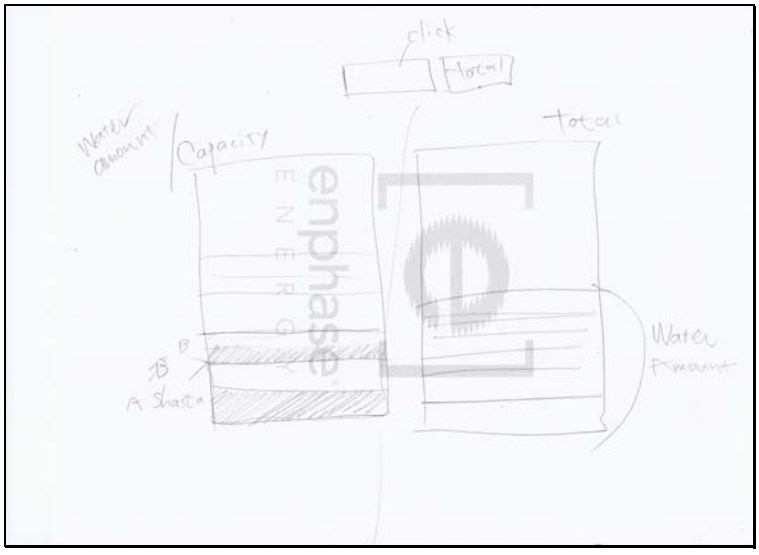

Figure 7

PROJECT SCHEDULE

The project schedule is listed below by date for the final project:

4/7 - 4/8: Acquire all data needed for the visualization, then clean, aggregate and prepare the data for the visualization. Complete sketches and ideas for charts that will be used. Discuss priority charts and the must have features in order to make the visualization work. Update the Process Book.

4/9: Assign development of individual maps and charts to team members. Update the Process Book.

4/10 - 4/13: Work on individual assigned charts. Update the Process Book.

4/14 - 4/15: Meet with Team members and review all charts, make comments on current work and possible enhancements. Update the Process Book and commit all work to Github.

4/16: Merge all the charts into one visualization.

4/17: Review milestone progress with TA's, and discuss comments and feedback provided by the TA.

4/18 - 4/20: Make enhancements to individual charts and begin work on linking all the charts with the event handler. Update the Process Book and commit all work to Github.

4/20 - 4/22: Reviewing projects with TA's. Enhance the visualization based on feedback provided by the TA. Review and discuss improvements for the charts. Update the Process Book.

4/23: Enhance the detail level of the charts, by providing data details on pop ups during mouse over. Update the Process Book and commit to Github

4/24: Plan out the work for the website that will contain the visualization and assign individual website tasks to team members. Update the Process Book

4/25 - 4/27: Work on website tasks.

4/28: Reviewing the functionality of the visualization in the website, check for bugs and validate the data and values are accurate. Update the Process Book

4/29: Meet with Team members and review the website and visualization. Review progress with team members and begin to put finishing touches on the final project. Assign tasks for final refinements of the visualization and last minute enhancements. Update the Process Book

4/30 - 5/1: Validate and test the visualization, fix any bugs that may have appeared. Update the Process Book

5/2 - 5/6: Prepare the final Process book, progress milestones, references and links.

5/5: Commit the final project file and all supporting files on GitHub