Final Project

Data Visualization

Jennifer Zhuang

Fall 2023

Abstract

This project utilizes data from the California Data Exchange Center (CDEC) to create data visualizations that examine historical trends and recent weather events. Spanning over 2000 stations with daily measurements, the dataset is transformed into Tableau dashboards that aim to address the challenge of interpreting the CDEC's extensive database. This project aims to aid in decision making regarding water resource management by emphasizing the need for adaptation and the sustainable use of water in the face of climate variability.

Background

Every autumn, both meteorologists and Californians turn their gaze to the skies, anticipating the much-needed fall of rain and snow. Unfortunately, over the past few years, their expectations have been met with disappointment. The state has been grappling with a decade-long drought and shifting climate patterns. The snowpack on the high Eastern Sierra mountains, a critical source of snowmelt for the watersheds to the west, typically accumulates during winter to support the rest of state through warmer seasons. However, this crucial water source has not been sufficiently replenished in years. Altogether, there has been a lack of precipitation statewide. These evolving climate patterns have posed challenges for California in effectively managing its water resources.

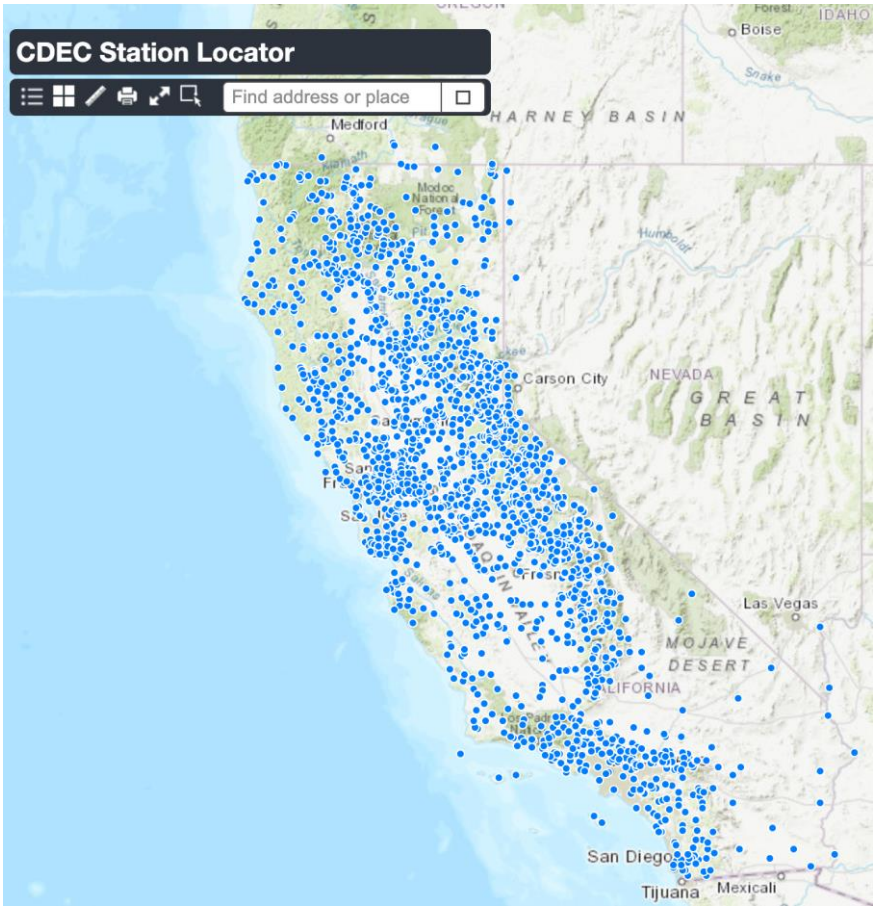
In a recent turn of events, California has undergone one of its wettest winters in a 70-year span, marked by a series of nine consecutive atmospheric rivers that brought unprecedented levels of rain and snow. This substantial influx of precipitation significantly alleviated the state's prolonged drought (Atmospheric Rivers Project, 2023). However, despite this shift in circumstances, the state is still not out of the woods yet. The challenge of climate change and water scarcity still persists. These extreme weather events, from prolonged droughts to sudden inundations of water, are just a peek into the state's future. We are still in need of adaptation and sustainable water use.

Considering the critical condition of California's water resources, this project aims to create data visualizations that examine historical trends as well as recent weather events. Using a wealth of information sourced from the California Department of Water Resources, these

visuals aim to offer a clear and accessible portrayal of California's hydrological conditions. Hopefully this analysis provides policymakers and stakeholders with a better understanding of the challenges posed by climate variability and supports informed decision making for water use in California.

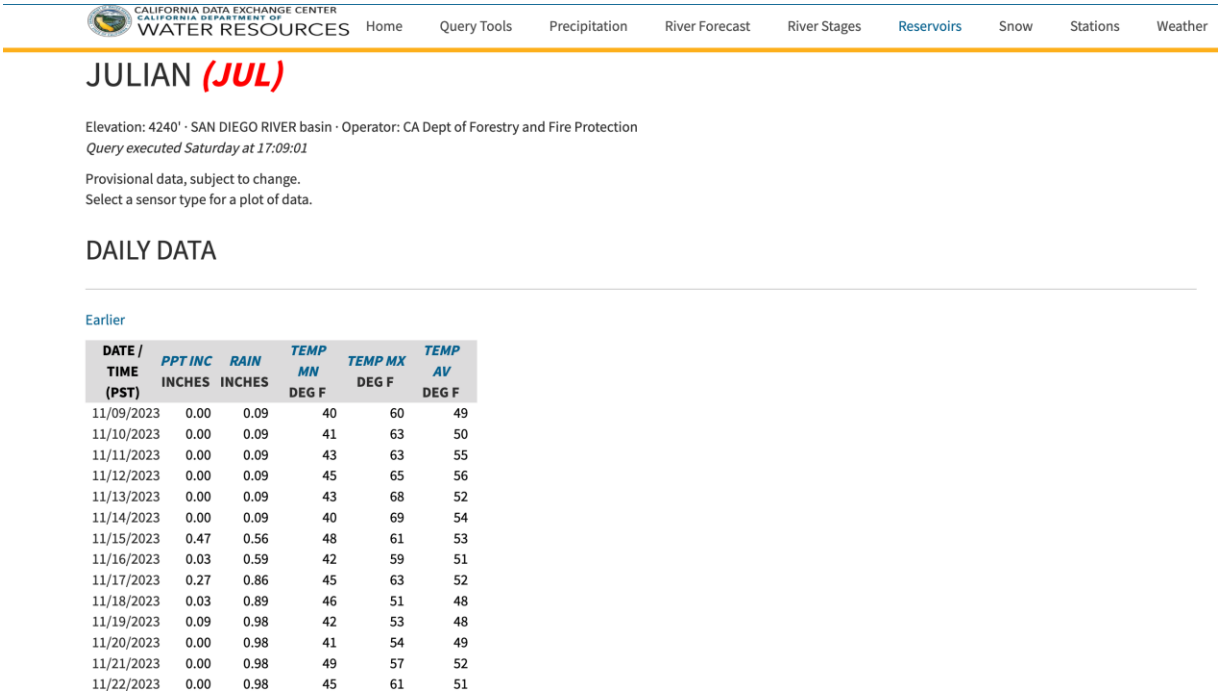
Introduction to the Dataset

The dataset we will be using to create visualizations comes from the California Data Exchange Center (CDEC) website. Maintained by the California Department of Water Resources (DWR), this extensive database provides users access to hydrologic and climate information. The department maintains numerous stations that span across the entire state. These stations record various kinds of information such as rainfall, reservoir levels, snowpack, weather, river flow, temperature, and more. Below is a photo of all the stations maintained by the CDEC (State of California, n.d.). There are over 2000 shown, and each consistently records information on an almost daily basis.



For this project, our focus will center on stations that record rainfall, snowfall, and temperature since these measures are critical indicators of climate change (CalEPA, n.d.). Rising temperatures, heightened fluctuations between drought and rainfall, and diminishing snowpack serve as symptomatic evidence of climate change.

To access information from CDEC database, users are required to query the website for data for each individual station. Stations record different measures, depending on what kinds of instruments are maintained. The data is presented in table form, and this pattern persists across all stations. Below is an example of a station called Julian which records rainfall and temperature (State of California, n.d.).



CALIFORNIA DATA EXCHANGE CENTER
CALIFORNIA DEPARTMENT OF WATER RESOURCES

Home Query Tools Precipitation River Forecast River Stages Reservoirs Snow Stations Weather

JULIAN *(JUL)*

Elevation: 4240' · SAN DIEGO RIVER basin · Operator: CA Dept of Forestry and Fire Protection
Query executed Saturday at 17:09:01

Provisional data, subject to change.
Select a sensor type for a plot of data.

DAILY DATA

Earlier

DATE / TIME (PST)	PPT INC INCHES	RAIN INCHES	TEMP MN DEG F	TEMP MX DEG F	TEMP AV DEG F
11/09/2023	0.00	0.09	40	60	49
11/10/2023	0.00	0.09	41	63	50
11/11/2023	0.00	0.09	43	63	55
11/12/2023	0.00	0.09	45	65	56
11/13/2023	0.00	0.09	43	68	52
11/14/2023	0.00	0.09	40	69	54
11/15/2023	0.47	0.56	48	61	53
11/16/2023	0.03	0.59	42	59	51
11/17/2023	0.27	0.86	45	63	52
11/18/2023	0.03	0.89	46	51	48
11/19/2023	0.09	0.98	42	53	48
11/20/2023	0.00	0.98	41	54	49
11/21/2023	0.00	0.98	49	57	52
11/22/2023	0.00	0.98	45	61	51

The database provides a wealth of information, yet accessing and interpreting its data can be challenging. The major issue lies in the intricate navigation required to retrieve information, limiting users to query the database station by station. This fragmented approach hinders the ability to comprehend broader patterns and gain a holistic view. Viewers cannot see the proverbial forest through the trees. Compounded by the sheer volume of data—exceeding 5 million data points—users may find the overwhelming number of numerical values difficult to interpret. To enhance usability and meaning, there is a clear need for data aggregation and visualization techniques that can distill this vast dataset into comprehensible trends and averages. This way, the CDEC's extensive water resource data can be transformed into a more accessible tool for readers and decision-makers alike.

Approach

The first phase in this project involves the fundamental extract, transform, and load (ETL) process of data science. This step combines data from multiple data sources into a single data store that is loaded into a target system. In our case, the multiple sources are the different stations and kinds of sensor data, the single data store would be CSV files, and the target system is Tableau. Given that the data is extensive, encompassing over 2000 stations and daily

measurements spanning decades, a strategic approach is required for querying and retrieving information. Manually clicking through the site pages and saving the information is pretty much unfeasible.

Therefore, I employed the Python library BeautifulSoup, a web scraping tool designed to pull data out of HTML files. It allowed me to parse over html code and search for the information I needed (Beautiful Soup, n.d.). For this project specifically, I parsed the html for the tables holding the data, iterated over the rows, and saved the columns that I needed. Used in conjunction with the Python Request library, I could easily iterate over a list of rainfall, snowfall, and temperature stations and grab the daily records. In essence, the library simplified the process of extracting relevant information and streamlined my data collection efforts.

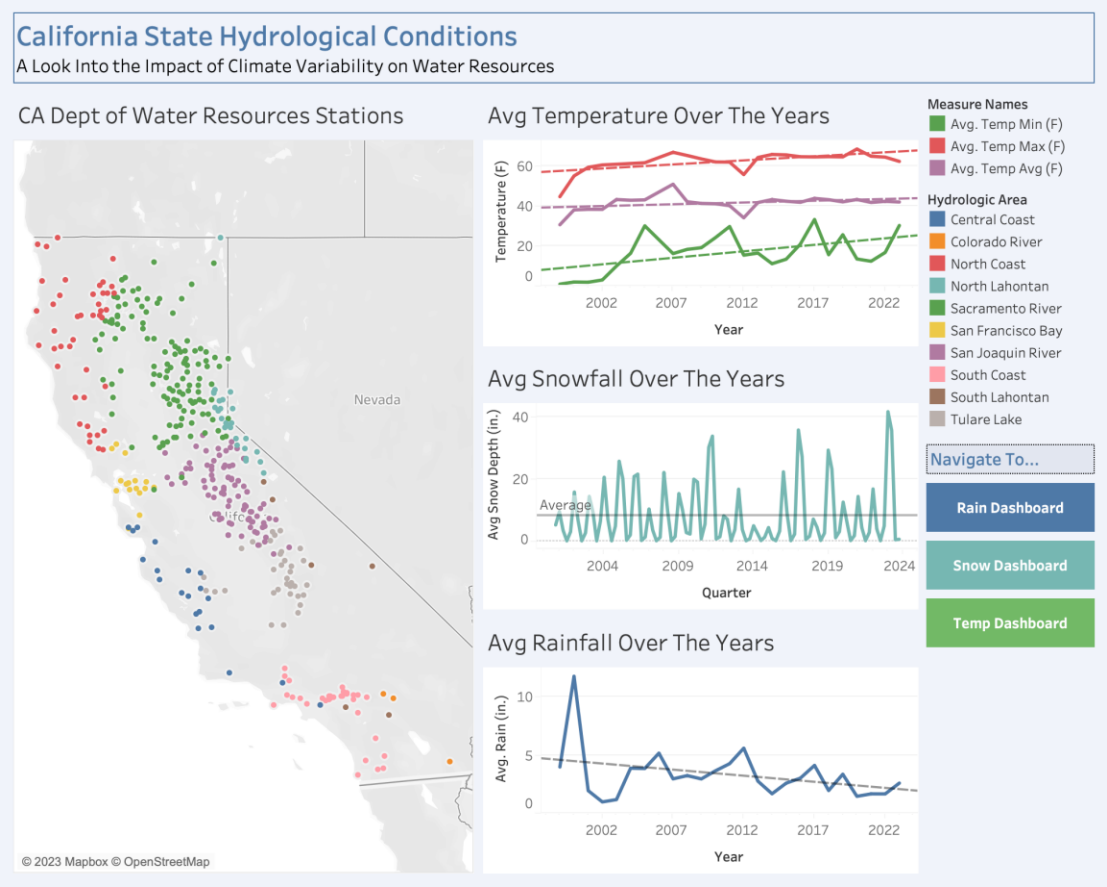
The last step of my approach was to find an effective way to approach the visualization process. Given that my goal was to illustrate how climate variability influences precipitation patterns and temperature in the State of California, I needed to anticipate the questions that viewers may have after viewing the data on the CDEC website. Putting myself in their shoes, I brainstormed key inquires such as the following:

- How has rainfall evolved over time? What are the historical patterns compared to the present?
- How has snowfall changed over? How do historical patterns compare to more recent data?
- How has temperature changed over time, and in what direction?
- Do these patterns reflect the negative impact we described above on how climate change affects precipitation patterns?

The goal is to not only present the raw data in a visually engaging manner but also to provide answers for these questions that facilitate a deeper understanding of the impact of climate change.

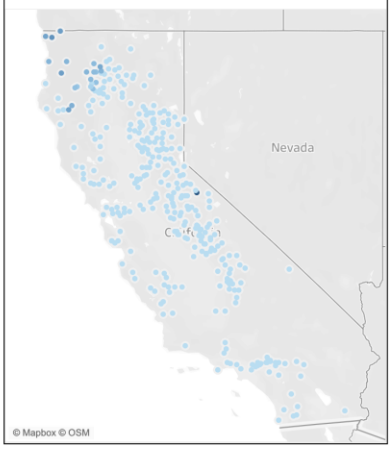
Results

Below are the results of importing our data into Tableau and creating a dashboard from it (Zhuang, 2023). I created a main dashboard that summarizes our findings, and then specific dashboards for our rainfall, snowfall, and temperature data. In the main dashboard, we can see that there has been a positive trend in temperature over the years while there has been a negative trend in snowfall and rainfall.



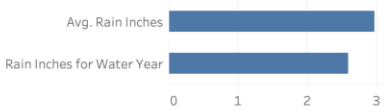
Rainfall Dashboard

CA Dept Water Resources Rain Stations

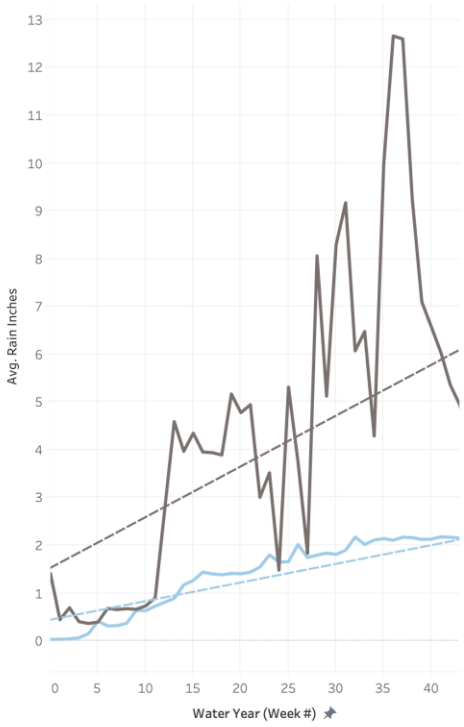


Rainfall Statistics
(2000-present)

Historical Avg Rain Inches : **2.972"**
Water Year to Date: **2.590"**
% of Historical Average: **0.8713**



Rainfall Year to Year Comparison



Avg. Rain Inches
0.0 105.4

Select Years to Compare Below:

Year of DATE / TIME (PST)..
Multiple values

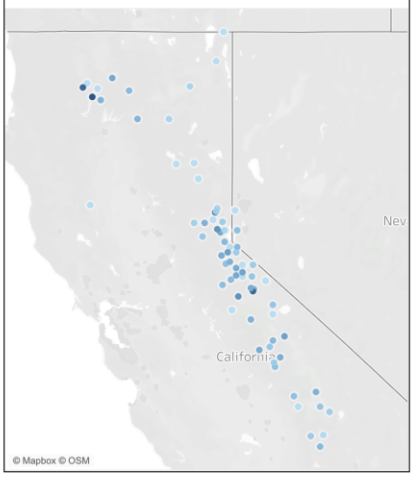
Year of DATE / TIME (PST)..
2011
2020

Navigate To...

Dashboard Home

Snowfall Dashboard

CA Dept. Water Resources Snow Stations



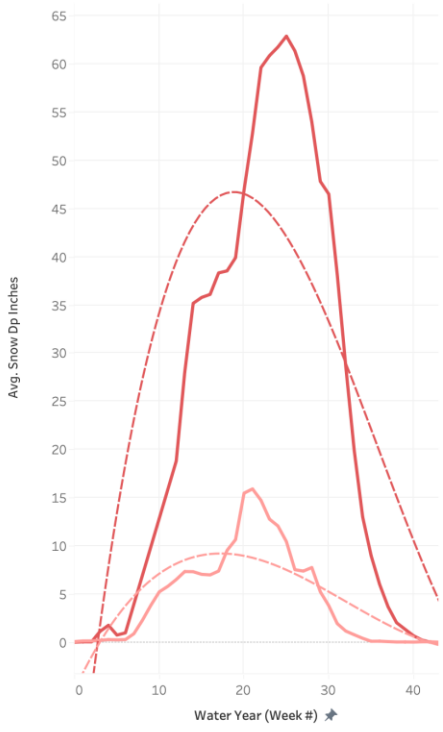
Snowfall Statistics (2000-present)

Historical Avg Snow Depth: **8.403"**
Historical Avg Snow Water Content: **8.231"**

Snow Depth Water YTD: **22.36"**
Percent of Historical Average: **2.660**

Snow Water Content YTD: **59.36"**
Percent of Historical Average: **7.212**

Snowfall Year to Year Comparison



Avg. Snow Dp Inches
1.34 26.43

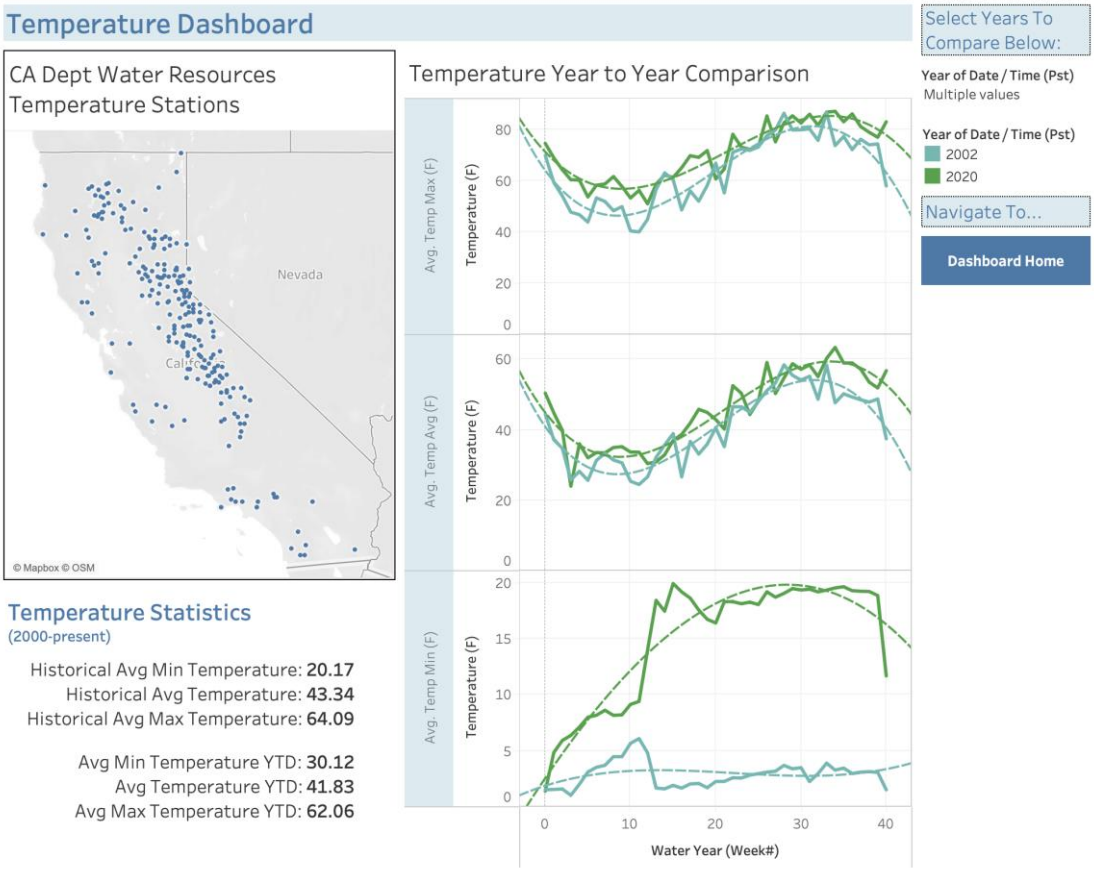
Select Years To Compare Below:

Year of DATE / TIME (PST)..
Multiple values

Year of DATE / TIME (PST)..
2007
2023

Navigate To

Dashboard Home



Within the data specific dashboards, I created line charts which allow viewers to compare water years to each other. To see how current data compare to historical trends, I set the years to compare as the early 2000s and then dates in the 2020s. After a few comparisons, we can see that there has been significantly less rainfall and snowfall in the state of California in more recent years. Further, the temperatures have been steadily increasing over two decades.

Conclusion

This project endeavors to shed light on the impact of climate change of California's water resources. Leveraging data from the CDEC, web scraping libraries from Python, and visualization tools from Tableau, the dashboards we created offer a comprehensive analysis of historical trends in rainfall, snowfall, and temperature. From this analysis, it is clear the specter of climate change and water scarcity persists, necessitating sustainable management of the state's water resources.

References

The Atmospheric Rivers Program. (2023, June 23). The Atmospheric Rivers Program. https://storymaps.arcgis.com/stories/109b3b7524034628918be55282ab578f

Beautiful Soup. (n.d.). *Beautiful Soup documentation¶*. Beautiful Soup 4.4.0 documentation. https://beautiful-soup-4.readthedocs.io/en/latest/

State of California. (n.d.-a). California Data Exchange Center. https://cdec.water.ca.gov/index.html

CalEPA. (n.d.). *Indicators of Climate Change* . ArcGIS StoryMaps. https://storymaps.arcgis.com/

State of California. (n.d.). *CDEC Web Applications*. California Data Exchange Center. https://cdec.water.ca.gov/dynamicapp/QueryDaily?s=JUL

Zhuang, J. (2023). *Final Project - Climate Variability & California Hydrological Conditions*. Tableau. https://public.tableau.com/app/profile/jennifer.zhuang/viz/FinalProject-ClimateVariabilityCaliforniaHydrologicalConditions/OverviewDashboard