# Data Sources

|  |
| --- |
|  |



<https://data.ca.gov/dataset/ground-water-water-quality-results>

<https://www.waterboards.ca.gov/water_issues/programs/water_quality_goals/docs/wq_goals_text.pdf>

<https://www.cityofwoodland.org/DocumentCenter/View/1067/Water-Quality-Report-PDF>



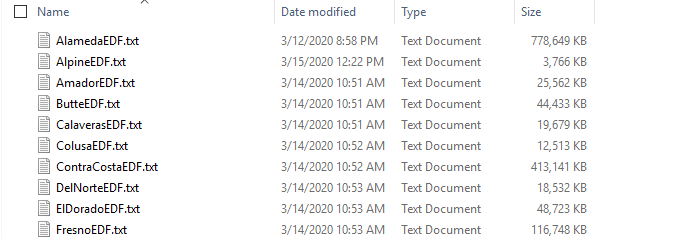
<https://www.census.gov/data/tables.html>



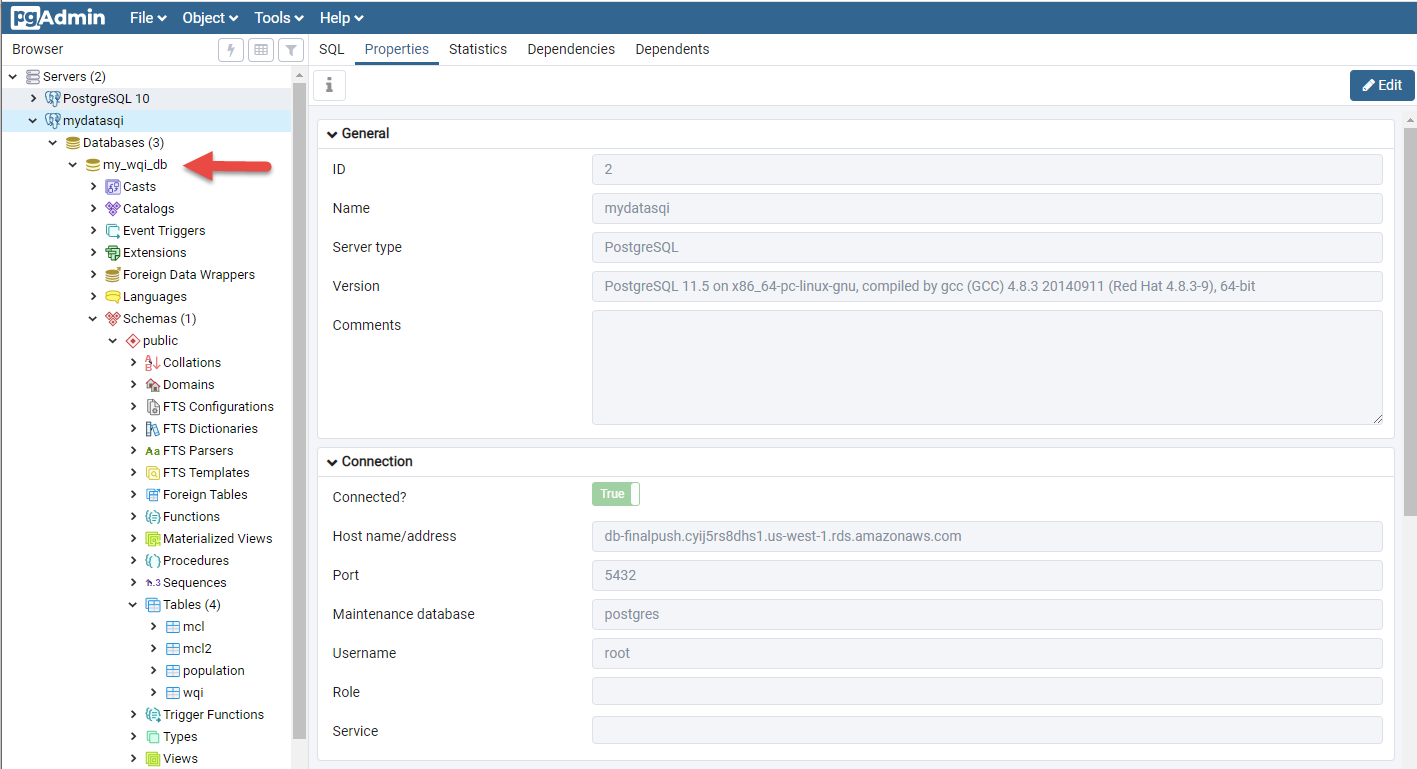
<https://droughtmonitor.unl.edu/Data/DataDownload/ComprehensiveStatistics.aspx>

# Database Creation

Initial download was 58 individual California counties .CSV files. These files were combined into four .CSV files.



We then created a PostgreSQL database instance on Amazon Web Services. We created the database schema in SQL and imported the CSV files into database. We also added the census information into it’s own table and added two additional tables for chemical contaminant levels, and one for drought conditions.



# Database Cleanup

We also inserted two new columns:

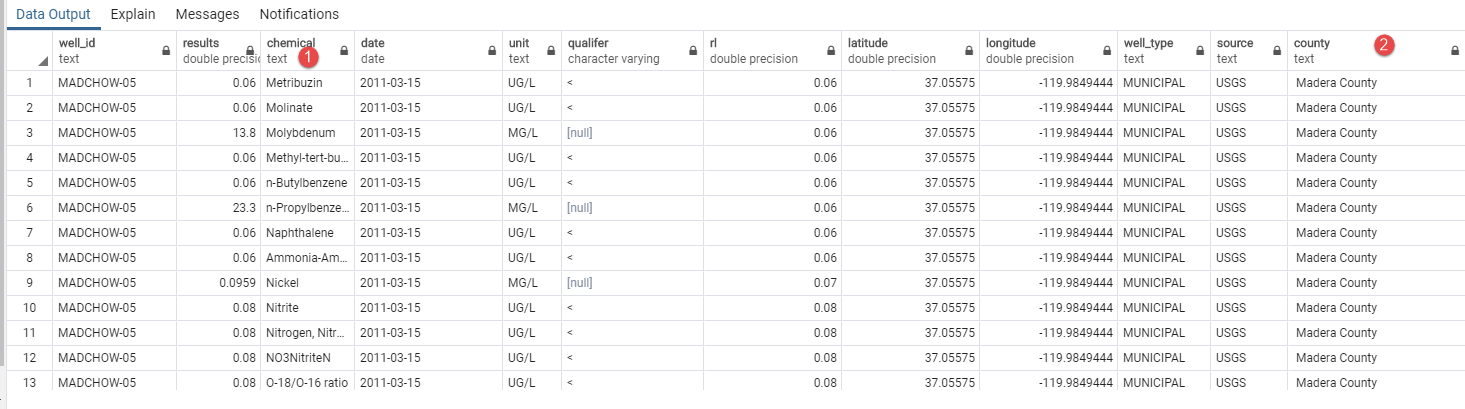
1 – Chemical - Updated to full Chemical name. We went to <https://pubchem.ncbi.nlm.nih.gov/> to find the chemical names then opened the .csv file in jupyter notebook and ran this command for each chemical.

for chem in data.chemical:   
print(chem)  
if chem == "MTBE":  
chem = "Methyl-tert-butyl ether"  
else if

2 – County – We added a county column. Originally all we had was latitude and longitude, but we needed county as well. We found geocoder that read the lat long in reverse called geopy.geocoder with an import called Nomanatim that we ran in Jypyter Notebook to accomplish this task.

Other cleanup tasks:

* Updating column names
* Removing needless columns
* Changing data types to allow the merge to occur
* Scrubbing Naan values and
* Ensuring equal amount of rows to each column.



# Tableau Connection

We connected our PostgreSQL database to Tableau. Here’s a screenshot of the data tab and tables. At this point, all project members had access to the AWS for our data reporting purposes.

