# Geovisualization:
# Integration and Visualization of Multiple Datasets Using Mapbox

Cecilia Cadenas

*California Polytechnic State University, San Luis Obispo*

**Abstract—The use of geovisualizations to analyze multiple datasets can ease the research and investigation to find correlations between these datasets. However, making these visualizations with multiple datasets can be challenging due to the lack of standardization within file formats and the software needed to render the visualizations.**

**This project investigates the making of these visualizations, maps in this case, using an open source platforms like Mapbox. Using Python scripts for the data parsing, CartoCSS for the designing of the maps, JavaScript to make the maps interactive, and HTML to combine all the layers together a set of maps were created. These maps contained data regarding tons of wine grape data crushed and yearly heat accumulation during the growing season that demonstrate how combining multiple data sets can lead to finding important correlations between these two separate yet related datasets.**

## I. BACKGROUND AND INTRODUCTION

### A. Geovisualization

According to the Commission on GeoVisualization of the International Cartographic Association, revisualization can be defined as the visualization of data with a geographical component. Often, these geographical data have a complex structure involving space, time, and a number of thematic attributes, which poses significant challenges to the visualization. The visualization of spatial data requires the use of maps or 3D displays where at least two display dimensions are utilized to represent the physical space, which is different from information visualization dealing with abstract data spaces. [3]

### B. Mapbox & TileMill

The creation of these visualizations can be challenging due to the complexity and differences between the geographical data types, and how easily they can be accessed and used. In order to help with the creation of these visualizations, many cartographers and software engineers have created many different tools that can make the creation of these visualizations easier and accessible to anyone with a computer.

For the creation of the maps in this project, one of these tools was used. After some research and testing different tools and platforms, Mapbox was selected as the best option for the customization level needed for the maps wanted. Mapbox is an online platform that makes the viewing of these maps simple across multiple platforms since it requires no extra software to be downloaded. Mapbox is open source and has many JavaScript libraries available to allow maps to be interactive. All of the needed libraries and tools needed to render a map are accessible within Mapbox. Mapbox also created TileMill, a desktop application that aids with the designing of maps using CartoCSS. It can help create the interactive maps that can then be uploaded to Mapbox and be accessed via the web.

Because Mapbox is open source, there are many examples within their site that can help with the creation of new maps. They also have many step-by-step tutorials on how to work with the data and how to use all their different tools to obtain the best possible maps. [2]

*C. Data Gathering & Parsing*

Within the last few years, obtaining data has become easily accessible thanks to the Internet. National organizations like the National Climatic Data Center and the National Agriculture Statistics Services keep open source records of the past few decades available for download as a csv or excel file.

One of the biggest challenges when working with any visualization tool is getting this data to work with the tools. Some of the files might have latitude and longitude that TileMill can recognize. However some data is relevant to more than just a single point on a map and is instead relevant to a whole region as the grape crush data is in this case. Also, the data obtained, might not be arranged in a way that can be meaningful in a map if presented that way. For example, the grape crush data is obtained in yearly summaries, each grape is its own row and its column is a district value. To make an interesting visualization combining data by grape rather than by year can be more meaningful.

There exist a file format (JSON) that can allow the creation of polygons given multiple latitude and longitudes that can then be used to make zone boundaries in maps. JSON files can also store other properties within each object like the average weather or the amount of grapes that were crushed in each zone.

In order to make this format type and obtain only the needed data, the original data needs to be parsed and modified or converted to a JSON file. To do this, the programming language, Python was used. Python allows the scripting of programs that can read files and create new ones in the proper format needed to create useful and informative maps.

## II. METHODOLOGY

Because of the number of different tools out there to create maps, the research of a couple of these tools was required before creating the final rendition and obtaining all the data in the right format. Finding the right tool meant finding something that allowed multiple data sets to be layered, and some sort of customizable, interactive tools for better understanding of the data. The first attempt of maps was used using openheatmap. This online tool allows users to create interactive maps in seconds by uploading csv files. With this tool, maps that could change by year were created fairly easy, but these maps were not very customizable and multiple data sets could not be added in one map. After emailing with the creator of openheatmap about the issues I had with it he suggested Mapbox because of the customizability it had.

Mapbox has all the tools that were needed, multiple data sets in one map, customizable, and interactive, but in order to take advantage of these tools, there was a learning curve to using this platform. Using Mapbox requires the knowledge of HTML, JavaScript, and CartoCSS in order to fully utilize all of its capabilities.

After many tutorials and getting familiar with the software, the creation of the maps began by making the data obtained online utilizable by Mapbox. The data gathered from the National Climatic Data Center consisted of one csv file containing monthly temperature averages from 2000 until 2013 in stations all over California. These averages, though important, on their own would be difficult to visualize in a static yearly map. To make these averages useful, the file was parsed and only the data during the grape growing season (April – October) were used. These values were used to compute the monthly average heat accumulation or also commonly known in agriculture as growing degree values that were then summed up into a yearly value for each station in California. [3,5]

In order to store the growing degree values calculated and be able to use within TileMill, the latitude and longitude of each station had to be converted to a JSON file format. To do this, a Google spreadsheet script was found and used that would do take the csv file and convert it to a JSON file. [1]

The grape data obtained from the National Agricultural Statistics Service did not require any more computations in order to used, but since the data had been split up by wine districts (Figure 1), this map needed to be created in a JSON format that TileMill could use. With the use of an available county map in JSON a new map of the 17 districts was created by manually making the borders using the boundaries of the counties that bordered each district. [4]

**Figure 1: California Grape Districts [4]**

1. Mendocino County
2. Lake County
3. Sonoma and Marin Counties
4. Napa County
5. Solano County
6. Alameda, Contra Costa, Santa Clara, San Francisco, San Mateo, and Santa Cruz Counties
7. Monterey and San Benito Counties
8. San Luis Obispo, Santa Barbara, and Ventura Counties
9. Yolo County north of Interstate 80 to the junction of Interstate 80 and U.S. 50 and north of U.S. 50; Sacramento County north of U.S. 50; Del Norte, Siskiyou, Modoc, Humboldt, Trinity, Shasta, Lassen, Tehama, Plumas, Glenn, Butte, Colusa, Sutter, Yuba, and Sierra Counties.
10. Nevada, Placer, El Dorado, Amador, Calaveras, Tuolumne and Mariposa Counties
11. San Joaquin County north of State Highway 4; and Sacramento County south of U.S. 50 and east of Interstate 5
12. San Joaquin County south of State Highway 4; Stanislaus and Merced Counties
13. Madera, Fresno, Alpine, Mono, Inyo Counties; and Kings and Tulare Counties north of Nevada Avenue (Avenue 192)
14. Kings and Tulare Counties south of Nevada Avenue (Avenue 192); and Kern County
15. Los Angeles and San Bernardino Counties
16. Orange, Riverside, San Diego, and Imperial Counties
17. Yolo County south of Interstate 80 from the Solano County line to the junction of Interstate 80 and U.S. 50 and south of U.S. 50 and Sacramento County south of U.S. 50 and west of Interstate 5

Once the district JSON maps were created, a script that parsed through the original csv files and added the grape data for that year to each object (district polygons) was developed. This script took in a minimum of four parameters, the original file grape, the year of the file, the district JSON file, and the grapes types wanted to be included. The script then creates a new JSON file with the district polygons and the grape crush data values. The new JSON data was then added to TileMill, and using CartoCSS a map layer was created for each year and each grape type.

## III. RESULTS

As Figure 2 shows, each data set individually has meaningful information regarding the weather of California and what district crushes more tons of Chardonnay. However, when these two datasets are combined, a new set of information can be discovered that can show more meaning as to why certain areas have more Chardonnay than others.

By combining these two datasets, and a timeline of 14 years, a trend can be seen and a relationship between the weather and the amount of grapes that were crushed that year. In the future this data can be used by viticulture specialist that can see what are actually is the best weather for growing each certain type of grape.

In order to make the maps interactive and allow anyone to switch between the years to see this relationship, the maps have a level of interactivity that allows the user to switch between the years as it is shown in figure 4. Users can also click on a specific district and a plot like figure 3 will show the whole time line of that district.

These maps were created for each of the seven most popular wine grapes grown in California, and a relationship between each grape and its main growing location can be seen for each grape type. The grapes are: Chardonnay, Sauvignon Blanc, Cabernet Sauvignon, Merlot, Pinot Noir, Syrah, and Zinfandel. These maps can be found at http://lupo.csc.calpoly.edu/wineviz



**Figure 2: Heat Accumulation and Tons of Chardonnay Grapes Crushed in 2012**
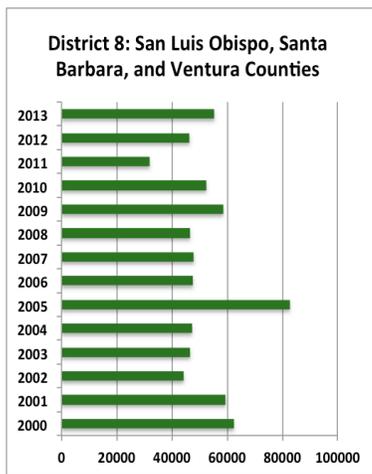
**Figure 3: Summary of Tons of Grapes Crushed in District 8**



**Figure 4: Final Version of Interactive Map**
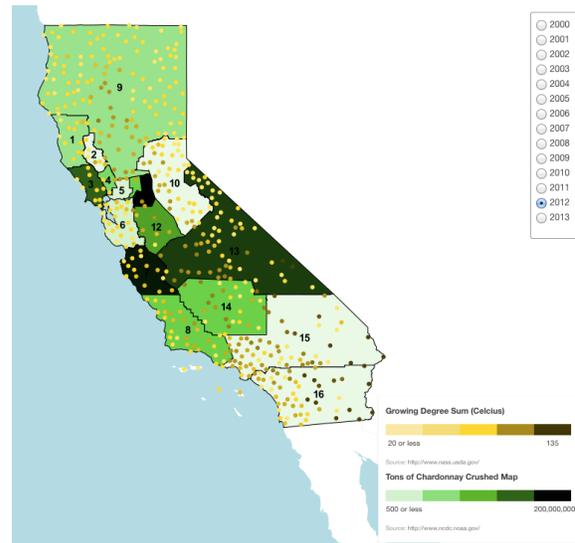
## IV.  FUTURE WORK

This project demonstrated that even though getting data to properly work with visualization software can be challenging, the end results could be very beneficial to the understanding of the relationship between different datasets. To improve the end results even more data like rain fall or wine awards could be added to the maps to see even more correlations between these datasets.  In order to make the data gathering process and also deal with getting more and more data, a database containing all of the data can be very beneficial. With the database even more of the map making process can be automated and the data will be guaranteed to be in the right format.

## V.  REFERENCES

[1] Cole, Dave and MacWright, Tom. "Geo for Google Docs"
<https://github.com/mapbox/geo-googledocs>
[2] Gundershon, Eric. Mapbox.com,
<https://www.mapbox.com>
[3]  International Cartographic Association,
<http://geoanalytics.net/ica>
[4] United States Department of Commerce. National Oceanic and Atmospheric Administration. <http://www.noaa.gov/>
[5] Washington State University. "Growing Degree Days"
<http://wine.wsu.edu/research-extension/weather/growing-degree-days/>