Weather or Not Wine is Fine

Michael Zhang, Clare Snyder, Shirley Burt

A. Mention any additional data that you used, such as shape files for maps. You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection. (10 pts)

For this project, it was required to use several datasets. We realized this presented an opportunity to find seemingly incongruous information and find interesting connections. We were drawn to the idea of visualizing wine ratings over time and region, and presenting them alongside weather information for the same areas and seasons. Although most people do not consider the weather of the year their wine was produced, the conditions in which the grapes which make the wine grow are very important to the quality of the grape; and by extension, the quality of the wine. We found the data to make our visualizations from Wine Spectator (for the rankings) and National Oceanic and Atmospheric Administration (for the weather data). The data can be found here:

- http://www.winespectator.com/wineratings
- https://www.ncdc.noaa.gov/cdo-web/search

N.B. Wine Spectator requires a subscription to access the data, but offers a 30-day free trial. We used this to get the data for this project.

From Wine Spectator, we recorded the rankings of three different wines: Chardonnay, Pinot Noir, and Rhone-style reds for each year and each wine-producing region in California. These wines were chosen because each region in California produced at least one of these wines, and these wines also had more years of data than others. We also noted the comments corresponding to each wine for each year. These comments in the original dataset did not describe each area, so by hand we rewrote each comment to describe the wine, the year, *and* the region. We also created an entry to note for each area which wines were produced there: an array of strings of length 1 to 3. Our final data was collected as an object with entries for each region. Each region's entry was an array of objects corresponding to the years for which we had data: 1997 - 2014. Within each year, we had the rating for each wine produced by the region, the array of strings indicating which wines these were, and a string with the comment for this year and region.

From the NOAA, we recorded the year, month, precipitation amount per month, number of days of precipitation per month, and average monthly temperature for each region, which we found using zip codes in each region. We realized that monthly data was too overwhelming to see, and not as relevant as general seasonal weather trends. So, we found the average amount of precipitation per day over each season, the total number of rainy days in each season, and the average temperature of each season. Seasons are defined as:

- Winter: January, February, December
- Spring: March, April, May
- Summer: June, July, August
- Fall: September, October, November

These were chosen because they are generally recognized as being the months corresponding to each season, and we wanted each season to be equal lengths. (Although we know in Ithaca this is not always

true!). We found the average amount of precipitation per day by dividing the amount of precipitation per month by the number of days of precipitation. We thought it was important to show not only the number of days of precipitation or the amount of precipitation per season, but the amount of precipitation for each day it precipitated, because there is a difference in a season with one day of torrential rain and a month of light rain. By calculating the data this way, we were able to show both scenarios. We organized the data as an object with object entries for each region, with objects for each year 1997 - 2014. These objects contained an array of objects for each season, with entries for season, temperature, daily precipitation, and number of days of precipitation.

Our data from each dataset was put in its own object, but is accessed in the same way because of its similar structure, so when a person selects a specific region from the map and a year from the slider, the code will easily access the information from both. The data is kept in different objects for clarity, as it is easier to read and edit the arrays of temperature without worrying about the ratings information and vice versa.

These SVG plotting files were found from online with sizes scaled accordingly when plotted and deserve the following attributions:

- Waterdrop icons made by Vectors Market from www.flaticon.com is licensed by CC 3.0 BY (https://www.flaticon.com/free-icon/drop_427112#term=rain%20drop&page=1&position=7)
- Rain cloud icons made by Freepik from www.flaticon.com is licensed by CC 3.0 BY (https://www.flaticon.com/free-icon/rain_704832#term=rain%20drops%20cloud&page=1&position=4)
- Farm icons made by Freepik from www.flaticon.com is licensed by CC 3.0 BY (https://www.flaticon.com/free-icon/farm_734190#term=barn&page=1&position=27) (Colors changed by us to show seasons)
- Snowflake icons made by Freepik from www.flaticon.com is licensed by CC 3.0 BY (https://www.flaticon.com/free-icon/snowflake_614116#term=snow&page=1&position=13)
- Snow cloud icons made by Freepik from www.flaticon.com is licensed by CC 3.0 BY (https://www.flaticon.com/free-icon/snow_414968#term=snowing&page=8&position=41)
- Wine glass stems borrowed from this wine class icon made by Smashicons from www.flaticon.com is licensed by CC 3.0 BY

In addition the the svg images, the text wrapping function we used, “wrap()”, was from an online implementation https://bl.ocks.org/mbostock/7555321 and not our original code, credit goes to that author.

An additional data set we used was a JSON file of California counties (https://github.com/scottpham/california-counties). From this data, we were able to create the map, with the hover-over textbox. Just like in class, we used d3’s geoJSON to create the base map from their built-in maps, then read in from the external JSON file to outline the counties.

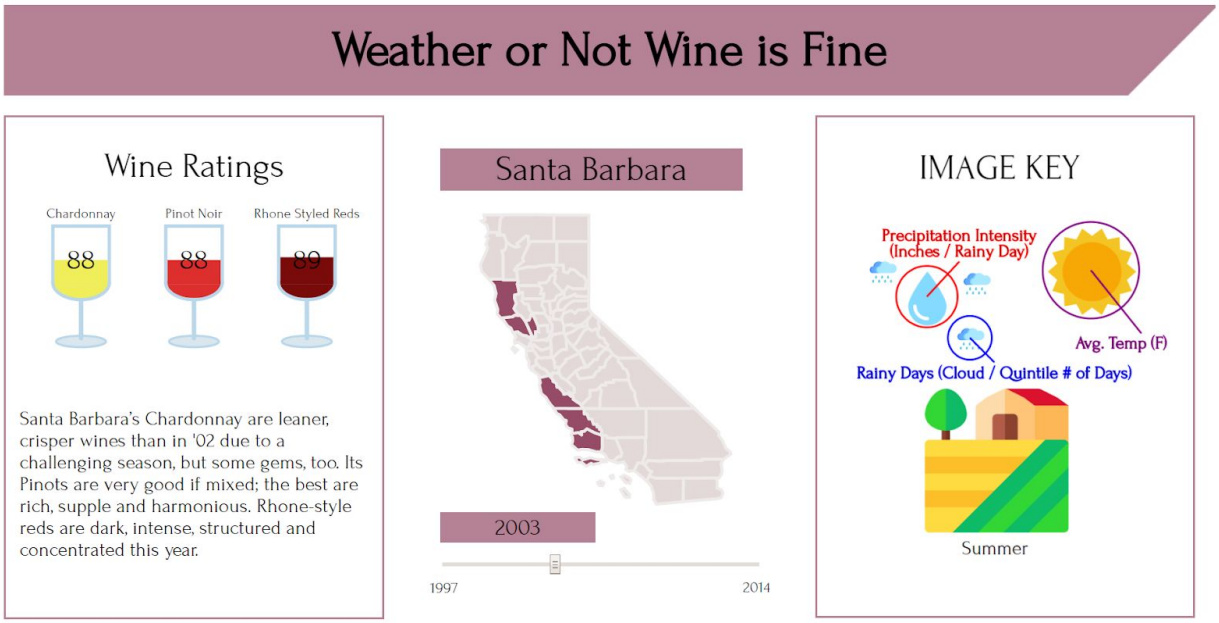
B. A description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales. (10 pts)

The data is graphed primarily in 2 visualizations, one displaying weather data, the other displaying wine vintage data.

- Weather plot:
    - All monthly data was grouped into seasonal data as described above
    - The svg is split into 4 portions to plot the data from each season in the order Winter, Spring, Summer, Fall with seasonal cues (snow and colors) to help readers distinguish seasons
    - There are 3 pieces of data plotted:
        - Days of precipitation is scaled based on a **quintile scale**, the number of rain/snow clouds represents the relative raininess of the season with the exception that if there was days of precipitation as define earlier by our data manipulation methods, there are no clouds (i.e. if there's 1 cloud, the season was within the 0th to 20th percentile for days of precipitation, if there are 5 clouds, the season was within 81st to 100th percentile for days of precipitation)
        - Rain intensity (average precipitation per rainy day) is scaled based on a **decile scale**, the size of the singular snowflake/raindrop represents the decile of the season's average precipitation per rainy day, this is a metric for intensity of rain when it does rain, as explained in your data above
        - Average temperature is scaled based on a **decile scale**, the size of the sun represents the decile of the season's average temperature
    - The rationale for decile scale: the goal of this project is to view relationships that require contrast to be visible. Graphing the data points on area scales is especially hard to view contrast because the radius of images grows logarithmically. Plus methods, such as temperatures are arbitrarily determined systems; if we measured our temperature in Celsius, Kelvin, etc. it would affect the scaling of our sun, but that shouldn't be the case that the system of measurement impacts the analytical power of our project. A relative comparison from one season to another provides enough digestible information to readers to allow them to see if it was relatively hot or cold that season and if it was relatively rainy that season, and is definitely a more applicable and a better scale. (***To show we understand what an area scale is, we have one that is implemented but unused and commented out***)
    - In order to give the reader a sense of numerical values, we've also given the 3 data values above the plot. While we could've made this interactive with a tooltip, we decided that would make viewing the data more challenging: in order to view the data, you'd need to mouse over the icon of interest, making the reader unable to compare all values at once. We believe our layout provided the most information in a non-distracting way.
    - Due to variability of screen viewing, the key for this map was created in an svg and then appended as an image to the project so that it fits on one page regardless of the zoom of your browser. The original function to create the svg key has been included in the file for your reference.

- Wine Vintage plot:
    - This plot shows all varietals that Wine Spectator rates for its annual review of a region's wine vintages
    - The svg positioning is dynamic to accommodate the varying number of varietals rated, spacing them evenly across the space
    - Coloring of the wine in the glasses and the varietal listed on top help readers determine which varietal is presented (colorings are attempted to be true to real colorings of the wine, this requires a little cultural background):
        - Chardonnay is represented by a vibrant yellow
        - Pinot Noir is represented by an light Earthy red
        - Rhone Styled Reds is represented by a deep ruby hue
    - This presents the rating of the vintage in two ways:
        - The vintage rating is posted within the wine glass in text
        - The height of the wine in the wine glass beyond the bottom curvature is increased for higher ratings and decreased for lower ratings. The amount is on a **linear scale from 80 to 100**. This scale does not begin at 0, and this is a conscious design choice we made. The reason for this is because although technically it is possible to assign ratings as low as zero, wine ratings really truly range from 80 to 100. Once again, the goal of this project is to make variation easy to read and a scale of 0 to 100 would highly diminish the visibility of variation just to be technically correct, unreasonably making comparisons challenging and consequently hurting the user experience of the project.

Note: We do realize our project has zoom and visualization differences among different computers, below is how we intended viewers to see it (view from our computer).



C. The story. What does your visualization tell us? What was surprising about it? (5 pts)

Sold for whopping $350,000, a single bottle of a 2015 Cabernet Sauvignon from California breaks the record for most expensive bottle of wine ever sold. With single bottle, just imagine how many years of Cornell tuition you could afford!

What differentiates a $5 bottle from Wegmans from this $350,000 bottle? Luckily, after paying four years of Cornell tuition and having taken Intro to Wines we know: "Vintage, vintage, vintage! Vintage is key!" Although there are many many factors that affect the price of wine, vintage variation is a key factor that drives the price of fine wines.
The vintage of the wine is the year the grapes to make the wine were harvested. Vintage is the amalgamation of a year's worth of sunshine, rainfall, frost, snow, and labor. Only from the most perfect weather conditions can we taste the fullest expression of the grapes in wine, but what are the perfect weather conditions that we are looking for? Is it the consistency of rain in Spring? The intensity of rain? Or is it the heat and sunshine the grapes receive in Summer? The science of viticulture is complex, but we hoped to offer some answers.

Our visualization is built for the avid wine enthusiast or curious reader to find the answer the vintage quality of a certain year. Looking up one's favorite wines by the county and vintage by the year, we set out to present data on the wine's vintage score and the weather that was so crucial to making the wine what it is, and it becomes possible to hypothesize the reasons behind an amazing vintage as a quick flip reveals weather patterns and the resulting vintage score.

We are not experts in grape growing, and there are many factors that affect the microclimates of viticulture, but as we each flipped through our own personal favorites, we saw that intense but infrequent Spring rain, a clear and dry Summer, and a warm growing season were vital to the success of the wine of our queries.

One thing that did surprise us was how much the ratings of different vintages changed from year to year. We expected more gradual changes between years, but instead, two consecutive years could have very different quality bottles! In part, this was because of the weather conditions from year to year that we observed in our plots.

Another thing we found surprising is that, although the wine ratings come from the same region that had the same weather conditions, some varietals would blossom while others would fall flat leading to variation among varietals within a vintage.

Overall, because of our enthusiasm about wine, we had an amazing time nerding out about the data, producing this project, and using our creation. It led to an hour or simply flipping through the data and talking about the tastes of our favorite bottles. Although this project doesn't cover enough variables to explain a $350,000 bottle, it does display the history and year in making of many California wines. It was an ambitious goal to attempt to explain the science behind all factors of viticulture, and although our project only glimpses into the field, it was still cool to design. Even if you enjoy wine casually are simply

curious about weather and wines, we hope you also enjoy searching up your favorite wines and seeing Mother Nature's part in producing it.

One takeaway is that because the weather and climate affect vintage quality so much, as climate change continues, the wines we are used to will change. Since there are many many factors to viticulture that our project cannot cover, it is difficult to predict how wines will be impacted, but we can expect that the ideal growing regions for wines will continue to evolve over the next generations.