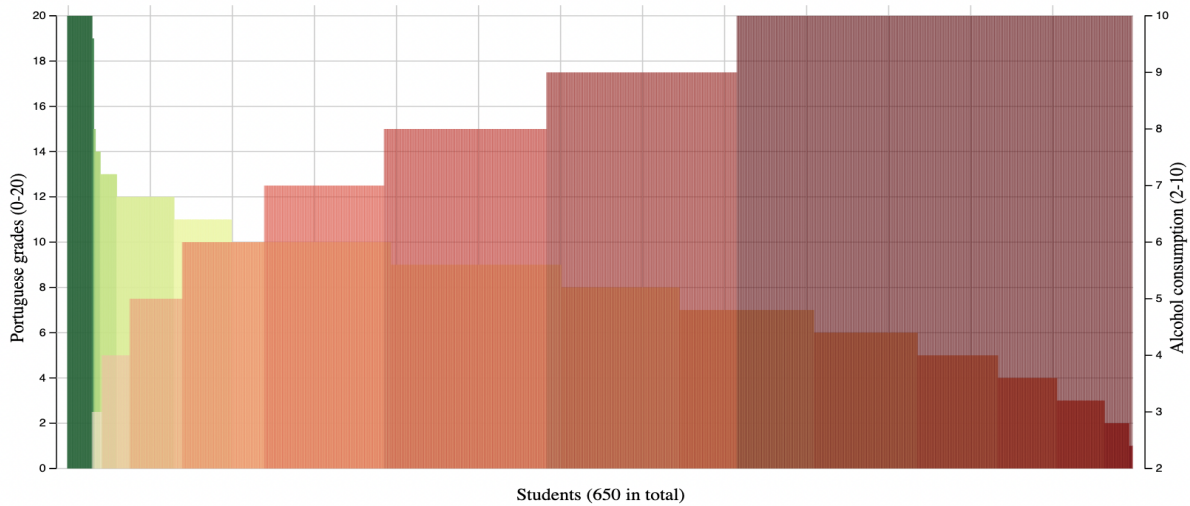


## Project 1 Final (Oct 7)

Kehui Guo (kg357), Zixian Jia (zj36), Yan Zhu(yz2477)

### Screenshots:

#### Relationship between Portuguese grade and total alcohol consumption

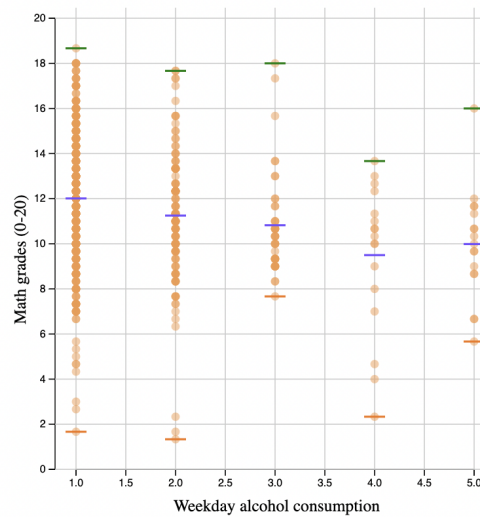
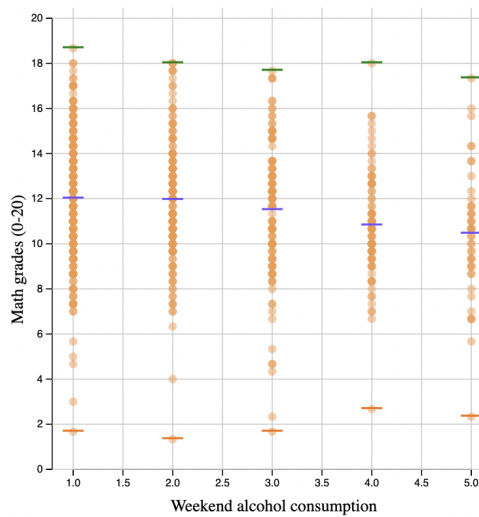


##### Color Legend

- Left: Portuguese grade. Red indicates lowest grade, neutral indicates average grade, green indicates highest grade
- Right: total alcohol consumption. Higher saturation of red indicates higher total consumption of alcohol.

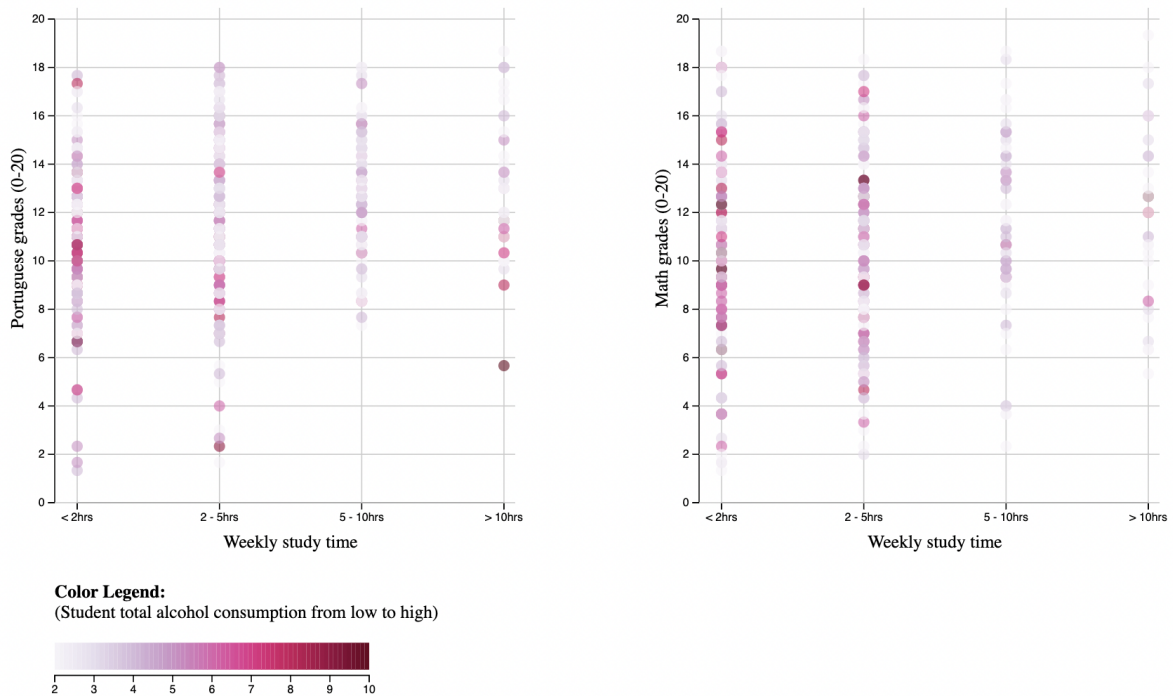


#### Correlation between Math grades and alcohol drinking on weekdays and weekends



- : Mean Math grade for a drinking level
- : Highest Math grade for a drinking level
- : Lowest Math grade for a drinking level

## Grade and study time for students with different total alcohol consumption



### Data Description:

The dataset was collected from [Kaggle](#). It delivers student achievement in Math and Portuguese classes from two Portuguese schools. The data variables include student grades, regular demographic attributes, family status, relationship status, previous academic failures, number of absences, school activities, and alcohol consumption. Each class subject has a separate dataset with the same attributes. In this project, only part of the data attributes and variables are selected for visualization. Demographic and personal information (family, relationship) are filtered out from the performance. The variables that are selected for the project based on the intended project goal are: G1, G2, G3 (grade at each period), Dalc and Walc (weekday and weekend alcohol consumption), and studytime (weekly study time) for both datasets. The selection is made based on the relevance to the project's curiosity and visualization priority. Finally, additional pre-processing has been done to the datasets. For each student represented in the datasets, the team added a new variable of Talc (total alcohol consumption) by adding Walc and Dalc of each student, and another variable of avgGrd (average grade) by taking the average of G1, G2, and G3 in order to get a better and balanced comprehension of the correlation between alcohol consumption and exam performance across different time periods.

## Overview of Design Rationale:

### Graph Set#1:

The graph describes the relationship between students' Portuguese grades and their total alcohol consumption. Each data entry is displayed with one thin colored rectangle, which creates bigger chunks of rectangles as a whole. The x-axis is used to indicate data from each student, which sum up to 650 students in total. Students' Portuguese grades are displayed in descending order from highest to lowest with indicators on the left y-axis using a scale of 0 to 20. A divergent color scale from red to green is used—red indicates the lowest grade with the connotation of worst academic performance, neutral color in the middle is average performance, and green indicates the highest grade and the most optimum performance. Students' total alcohol consumption level, which is the sum of their weekday alcohol consumption level (level 1 to 5) and weekend alcohol consumption level (level 1 to 5), is displayed in ascending order from lowest to highest on a scale from 2 to 10 with labels on the right y-axis. A sequential red color scale is used as the most saturated red represents the highest alcohol consumption and not ideal behavior, while the white/neutral red color represents the minimum consumption of alcohol. The gridlines follow the scale on the left vertical axis because it is more straightforward and simpler to understand, but both axes have labels and ticks to indicate the meaning and levels on the scale. The visual marks used in this visualization are colored rectangles. The visual channels employed are size, varying color hue, color saturation, vertical aligned length, and horizontal aligned positions of rectangles.

### Graph Set#2:

Both graphs describe the relationship between students' Math scores and their drinking activities. The first graph shows this relationship on Weekdays, and the second graph shows this relationship on Weekends. For each graph, the X-axis shows the level of alcohol consumption (level 1 to 5), and the Y-axis shows the average Math scores (0-20) of the students. Each orange circle within each graph represents 1 student. For each alcohol drinking level, the purple horizontal line in the middle represents the average Math grade of all students at that drinking level, the top orange horizontal line represents the highest grade of all students at that drinking level, and the bottom orange horizontal line represents the lowest grade of all students at that drinking level. The visual channels employed in this data visualization include colors, horizontal aligned positions, and vertical aligned positions. The marks employed include circles and lines.

### Graph Set#3:

The graphs describe the relationship between students' Portuguese(left)/Math(right) average exam score over the three exam periods and their weekly study time. The x-axis on each graph represents the weekly study time on an ordinal scale of 1-4 remapped into the actual number of hours defined by the dataset. The y-axis shows the average Portuguese/Math scores taken from three grades from 0-20. The color scale demonstrated by the legend is designed to be sequential in order to effectively show the total weekly alcohol consumption. The total weekly alcohol

consumption is calculated by adding the weekday alcohol consumption and the weekend alcohol consumption. Each circle on the graphs illustrates the total alcohol consumption (the higher the consumption the darker the circle color), his/her average test score of the subject, and his/her weekly study time. The visual channels selected for this visualization include color saturation variation, horizontal aligned position, and vertical aligned position. The marks for the data points are circles. The graph is designed in such a way that is able to present the distribution of each study time category clearly therefore I chose not to use a bar chart for the study time.

### **Visualization Insights:**

#### **Graph Set#1:**

From the visualization, it shows that only a small percentage of students receive the highest grades in Portuguese, and also there's a small percentage of students who consume a minimum level of alcohol. While this data visualization does not tell viewers what is the total alcohol consumption of a specific student, the general trend and the size of different levels indicate a loose pattern regarding the percentage of students getting high grades and consuming low levels of alcohol, and vice versa. The visualization of the large red areas of high alcohol consumption makes the viewers wonder whether the students need more support and guidance since a big portion of the student population is indicating extremely high levels of alcohol consumption.

#### **Graph Set#2:**

From the two graphs, it's interesting to find out that different drinking levels on both weekdays and weekends don't have a strong correlation with students' average Math grades. However, if we look at the highest scores (the orange horizontal bars) at each drinking level, we can see that more drinking on weekdays is correlated with lower max grades. That is, the top students with the highest Math grades don't drink a lot on weekdays. However, surprisingly, this trend is less prominent on weekend-drinking. We can reasonably say that the level of drinking on weekends is not highly correlated with Math grades, and students who drink a lot on weekends can still get very high average Math grades.

#### **Graph Set#3:**

Students who study the most (highest weekly study time) receive the highest average grade over the semester and are also the ones who drink the least (low total alcohol consumption) compared to other students. This finding from the visualization validated the general assumption that if students spend less time drinking overall and use that time to study they will excel in the class.

## Team Contributions

Kehui Guo:

- Designed and visualized graph#1 for the relationship between students' Portuguese grade and total alcohol consumption. (~ 6 hours)
- Brainstormed ideas and collected the dataset (~2 hours)
- Contribute to the project writeup (~ 2 hours)

Zixian Jia:

- Designed and visualized graph set#2 for correlation between students' Math scores and their drinking activities on Weekdays and Weekends. (~ 6 hours)
- Brainstormed ideas and collected the dataset (~1 hour)
- Contribute to the project writeup (~ 2 hours)
- Presented the project to other students during the project critique

Yan Zhu:

- Designed and visualized graph set#3 for correlation between students' average exam score for both subjects and their weekly study time (~ 6 hours)
- Brainstormed ideas and collected the dataset (~1 hour)
- Contribute to the project writeup (~ 2 hours)
- Polished and finalized visualization styling (~1 hour)