Student Grade Analysis based on Alcohol Consumption

The purpose behind my project was to gain more of an understanding of how alcohol effects a student's performance in school. I used a variety of visualizations to display student's performance in school, in terms of grades while filtering the data by student's alcohol consumption during the week and/or weekend. This way when a user uses my exploratory dashboard, he/she would be able to answer questions such as whether a student performs better or worst in school when consuming alcohol, whether different levels of alcohol consumption affect a student's performance similarly or differently, whether alcohol affects a student's performance in school in terms of study habits, and lastly whether alcohol affects a student's family relationship.

In terms of visualizations used in my dashboard, the first visualization put major emphasize on student's average grade received based on their weekday alcohol consumption. I chose this objective for my first visualization because I wanted users to know right away what the main topic my dashboard would be representing. Since, majority of effective dashboards capture their audience attention at the top of the dashboard design. This way users know right away what questions they we be able to answer, throughout the dashboard.

The types of questions the first visualization answered were what grades students receive on average depending on their alcohol consumption during the week. Consequently, the user would be able to see a range of values. Which would represent different levels of alcohol consumed by different segments of the student population who go to two different schools in Portugal named Gabriel Pereira and Mousinho da Silveira. The alcohol levels that would be displayed in this visualization ranged from levels of 1 (very low) to 5 (very high). A level of 1 meant the student did not drink alcoholic beverages that often during the week; at most one alcohol beverage a week. While a level of 5 meant that the student consumed alcoholic beverages everyday during the week.

The design chosen for this visualization was a square text chart. Mainly, to put specific emphasis on the comparison of final grade point averages between students of different alcohol consumption levels. Each grade point average were listed below the label which showed the respective level of alcohol consumed during the school week by each student population. The visualization also used the attribute of color intensity to show the spread of grade point averages in a visually more compelling way. The color intensity chosen was shades of blue to make it easier for a user to see who had the highest grade point average out of all alcohol student consumers. As well as who had the lowest grade point average out of all alcohol student consumers. With other ranges of alcohol consumption levels in relation to student grade point averages shown in between.

In this visualization I only wanted to show the final grade point averages and not midterm or quarterly grade point averages to not interfere with the main objective of this

visualization. Which one of Stephen Few's principles also addresses, "*display neither more or less than what is relevant to your message.*"

The next visualization I choose to use was a visualization that showed student's final grade point average based on weekend alcoholic consumption. I placed this visualization on the opposite side of my first visualization to show the comparison between grade point averages among the different segments of the student population. Since, I really wanted users to see if there was any similarities or differences in students' performance based on whether they consumed alcohol on the weekdays or weekends or both. In terms of design, this visualization was similar to the one I created as my first visualization. In that both visualizations used text box designs with labels of alcohol consumption on the top and their corresponding grade point averages on the bottom directly below the labels. Different shades of blue were also used as a preattentive attribute on the grade point average values. To show which students received higher final grades by the end of the school year. I also displayed this second visualization on the top of my dashboard, on the opposite side of visualization because I wanted both visualizations to act as my filters for this dashboard.

With the inclusion of filters from visualization one and two fields I was able to make the dashboard more dynamic and responsive. In terms of the data the dashboard would show its users. More specifically, the two filters would display a dropdown list of alcohol consumption levels that would allow a user to be able to partition the data. In order to show information correlated to a student's alcohol consumption level during the week, weekend, or both. This way users of the dashboard could get a more drill down result of statistical values pertaining to different segments of student population, with respect to alcoholic consumption level.

In terms of positioning, to make the two filters conform to the visual mapping principles, I placed both at the very top of my dashboard. I did this to make sure the two filters did not break the structure of my visualization design hierarchy in the dashboard. Also, with the two filters placed on the top opposite corners of the dashboard, users would be able to select different levels of alcohol consumption, before entering the main body of the dashboard. Thus, allowing the dashboard to dynamically change values at the beginning. Rather than waiting for the user to get to the bottom of the dashboard, then changing values. Which would happen if I placed the two filters at the bottom of the dashboard.

The third visualization I displayed in the dashboard, was a text table that showed the average final grade between student's gender. The two gender types collected in this dataset were female and male students. I color coded the female box silver and the male box blue. I could have chosen to leave the columns of the table one color hue, since there was not perceived quantitative difference between variable values. However, I wanted users to be able to depict differences between gender performances in school easier. And I felt with the display of two different color hues, users would be able to see which values pertained to which sex faster. From the third visualization questions users would be able to answer included, which sex scored higher final grade point averages during the conclusion of the school year.

In terms of making this visualization more interactive to users, the values listed in this visualization could be filtered two ways. One way was through the dropdown list of levels of alcohol consumption during the week and weekend consumed by the two genders of students. Another way was by filtering the data solely on one specific sex of the student population. This would allow the user to touch either sex icon displayed in the visualization and segment the data to show only male or female student grade point averages based on different alcohol consumption levels. Which allowed for a more intuitive approach at gather data about a student's performance in school.

The next visualization I chose to display was a column bar chart. This visualization displayed the amount of class failures a student accumulated during the school year. Which was then grouped by a student's age. The reason I included this visualization in the dashboard was to get a better understanding of why a student performed the way he/she did, during the duration of the school year. I felt from a statistical standpoint the number of failures a student obtained during the school year would directly correlate to their final grade point average received at the end of the year. As a result, questions a user would be able to answer with this chart would be which age groups faced more class failures and whether their grade point average decreased because of this choice of action.

In respect to chart design, I selected a bar chart to represent the data for this visualization because the data that was displayed was grouped in discrete qualitative categories. Which bar charts are best at showing and the comparison between different age groups performance could be displayed proportionally through different bar heights.

In conjunction to the third visualization described previously, the fourth visualization in could also be filtered two ways. The first way was through the two filters displayed on the top of the dashboard. That included the dropdown list of different alcohol consumption levels consumed by students throughout the week and weekend. While the second way the visualization could be filtered was through the chart itself. Which meant that a user could select an age group of their choosing and see how they performed in school, through the different statistical metrics displayed in other visualizations on the dashboard. For example, a user could compare students' performance between age groups of 15 and 16 year olds or 18 and 22 year olds by selecting their respective bar in the bar chart. Then the dashboard would change every visualization values throughout the dashboard, in relation to the age group that was selected.

Next, I color coded the bars of the different age groups with a diverging color scheme to quickly distinguish which age groups had the least amount of class failures. Consequently, being able to see if this age group faced less of a trouble obtaining a higher grade point average than age groups who faced higher amounts of class failures.

For my fifth visualization design I wanted to incorporate other statistical measures to predict a student's final grade result. And I thought the best way to do this was by including a student's quarter and midterm grades into a plot for comparison against their final grade. The

plot which I thought best handled this task was a scatter plot matrix. Since scatter plots are great when describing correlations between two or more variables of interest. The second variable of interest, I wanted to compare against different student's grade point averages was a student's absence total. Mainly because I wanted to see if the number of times a student missed class had a direct relationship with their quarterly, midterm, and final grade performance.

In the fifth visualization design, I included the number of absences a student accumulated on the y-axis and the different grade point averages a student scored on the x-axis. I also included trend lines for each plot in the scatterplot matrix. The reason for the inclusion of trend lines was to have statistical evidence (p-values) to back my hypothesis whether or not the number of absences a student obtained correlated and/or effected their grade point average throughout the school year. Other aspects of the visualization I used was different color hues to represent each selection of data points in each scatterplot in the matrix, as well as a background color to encompass the entire scatterplot.

The sixth visualization included in my dashboard incorporated a design choice of gauge charts. The use of gauge charts was necessary to show differences in the amount of study time needed between students to obtain above average final grade point averages. I used four gauge charts to represent each weekly study time groupings that showed the key metric of how much time a students spend studying to obtain final grade point averages of 80 or better. Since this metric was performance based, gauge charts were the most user-friendly choice to decipher proportional trends in data.

Questions users could answer from this visualization would include what different percentage of study times a given student population spent studying to receive grade point averages of 80 or better, which study time grouping was the most used among the students, and which study time grouping was the least used among students. In terms of filtering, I decided to only allow the data to be filtered from the top two filters of the dashboard; these two filters were related to levels of alcohol consumed by students during the week & weekend.

The seventh and eighth visualization used in my dashboard used a design of horizontal bar charts to display data pertaining to a student's family support system. The reasoning behind including a student's family support system in both visualizations was to show whether a student's mother or father education level played a role in determining a student's final grade. Since, I wanted to see if student's close family members such as their mother or father played an impactful role in their child's performance in school based on their education (1st grade-college) obtained.

Again, since the number of educational group levels was low, I felt bar charts would best display the comparison between associated groups. However, instead of using a diverging color scheme which was used in the other bar chart of the dashboard. I decided to use a sequential color scheme for these two visualizations, since I was not trying to discover which educational

grouping obtained by family members (father & mother) were more successful when compared to the average grade point average obtained by students.

In terms of interactive features for these last two visualizations, I decided to allow the two bar charts to be filtered only one way. Which was through the main two filters on the top of the dashboard, which displayed the list of alcohol consumption levels consumed by students.

The only other metrics included in my dashboard design were two static visualization charts. Which displayed supporting data to further back observations discovered from values displayed in visualization six and seven. More specifically the first static visualization had a metric that stated the total percentage of students who had a strong family relationship at home. I found this metric important to include because this way I knew if there was a chance for a student to show improvements in their grade point average. Since if there was a small percentage of students who had strong relationships with family members, there would probably would be a small selection of students who performed above average in school in visualization seven and eight. Mainly because these students would not have a great relationship with their mother or father. So, it would not matter what educational level their father or mother achieved. Then the second static metric visualization displayed in my dashboard stated the total percentage of students who had above average free time after school during the school year. I found this metric particularly important to include because students with above average free time should have no trouble spending time after school studying for test. Which could be a reason why students in my sixth visualization would have less percentages of people able to spend greater than 5 hours a day studying for tests during the school year.