
Student name: Navjot Singh Virk, Student Number: x13112406

REPORT OF ANALYSIS

Students Alcohol Consumption

DATASET 1: student-mat.csv
(Mathematics Students)

DATASET 2: student-por.csv
(Portuguese Language Students)
Currently available at:
https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION

Table of Contents

Introduction

The aim this report is present analysis results based on a series of analyses carried out on Dataset 1 and Dataset 2 using R language and the programming environment RStudio.

Formal Description of the Datasets

DATASET 1
The dataset 1 (student-mat.csv) used in this analysis is about Mathematics students in secondary school and contains 33 attributes which was composed by P.Cortez and A. Silva, at University of Minho in Portugal.

DATASET 2
The dataset 2 (student-por.csv) used in this analysis is about Portuguese subject students in secondary school and it also contains 33 attributes like Dataset 1 and is composed by the same author's as Dataset 1.

Below are the attributes of datasets (both the datasets contains same attributes which will be helpful during the analysis and comparing attributes results and draw meaningful results and predictions based on analysis) -

Attribute	Description
school	Students School (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira)
sex	Student's Sex (binary: "F" - female or "M" - male)
age	Student's Age (numeric: from 15 to 22)
address	Student's home address type (binary: "U" - urban or "R" - rural)
famsize	family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
Pstatus	parent's cohabitation status (binary: "T" - living together or "A" - apart)
Medu	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
Fedu	Father's education (numeric: 0 – 4)
Mjob	mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at home" or "other")
Fjob	father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at home" or "other")
reason	reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
guardian	student's guardian (nominal: "mother", "father" or "other")
traveltime	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
studytime	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
failures	number of past class failures (numeric: n if 1<=n<3, else 4)
schoolsup	extra educational support (binary: yes or no)
famsup	family educational support (binary: yes or no)
paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
internet	Internet access at home (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

freetime	free time after school (numeric: from 1 - very low to 5 - very high)
goout	going out with friends (numeric: from 1 - very low to 5 - very high)
Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
health	current health status (numeric: from 1 - very bad to 5 - very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20)

Objective of Analysis

The objective of the analysis of these 2 datasets is to draw meaningful results and predictions based on the analysis that will be carried out on these datasets. The datasets contains information like students age, study time, travel time, failures and more which will help us understand –

Data Processing Activities Carried out

Both selected datasets were well done already and required little to no cleaning

Data processing activities – the datasets were stored in R objects and from there the information was used to draw meaningful plots and predictions (which can be seen in the results section).

- Finding the weekend alcohol consumption based on guardian of the student to see if living with a particular parent or other guardian has any impact on drinking habits.
- Consumption based on gender to see if boys or girls are drinking more and predict a reason why ? and a possible solution
- Consumption based on age to check and possibly control the level of drinking in students of age that's drinking the most.
- School Absences and Alcohol Consumption – to see if not going to school has any impact on consumption levels
- Alcohol Consumption and Grades – to see how alcohol consumption is impacting the studies and grades of the students.
- Weekly Study times and grades – compare grades of students with more study hours to the students with less study hours.

For analysis of data combined from both datasets –

- Find out weekend and workday alcohol consumption based on age of students
- Consumption based on age and gender
- Consumption based on family relations
- What % of students have access to internet and other activities.
- Alcohol consumption and desire for higher education

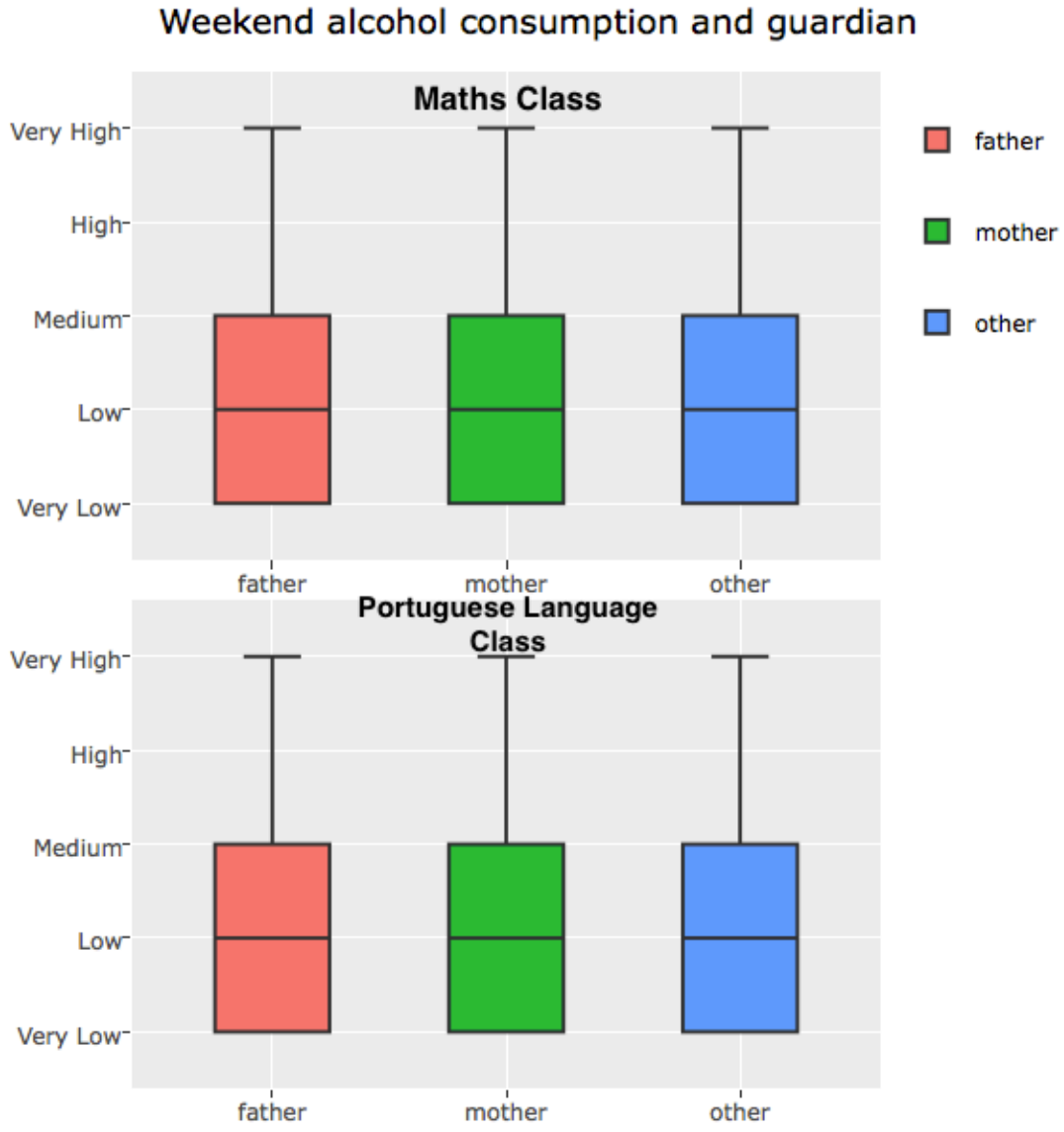
Final Analysis Results

Below are the results of individual analysis on each dataset and then analysis on the the data combined from both datasets.

Individual Analysis on Datasets (Results)

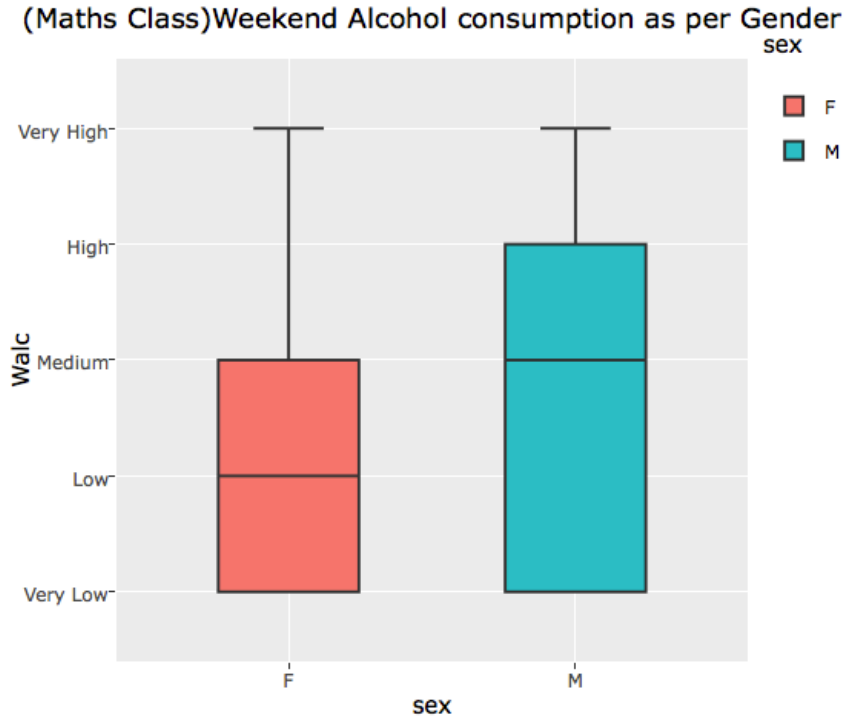
Weekend alcohol consumption based on if the student lives with their parents (mother or father) or other

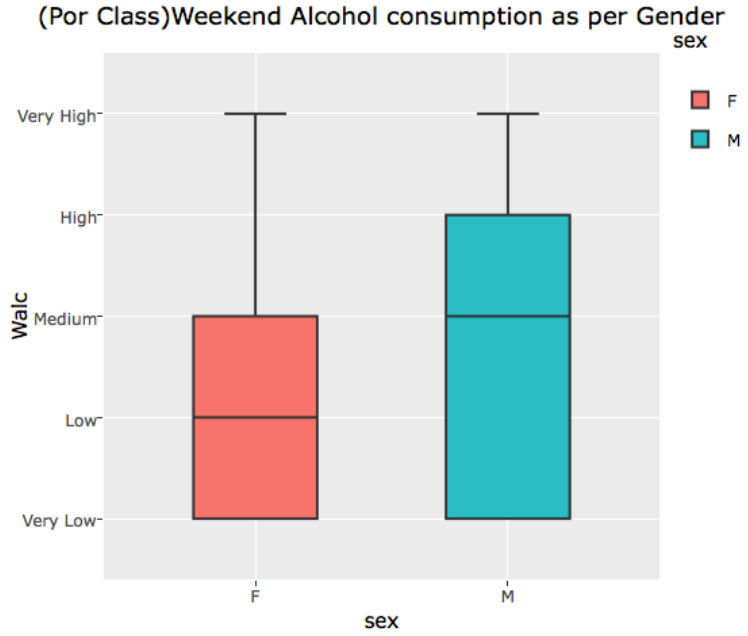
The results show that living with any of the guardian does not have huge impact on the alcohol consumption at least for these dataset (in the analysis its found very few students live with other person than there parents)



Weekend Alcohol Consumption based on Gender

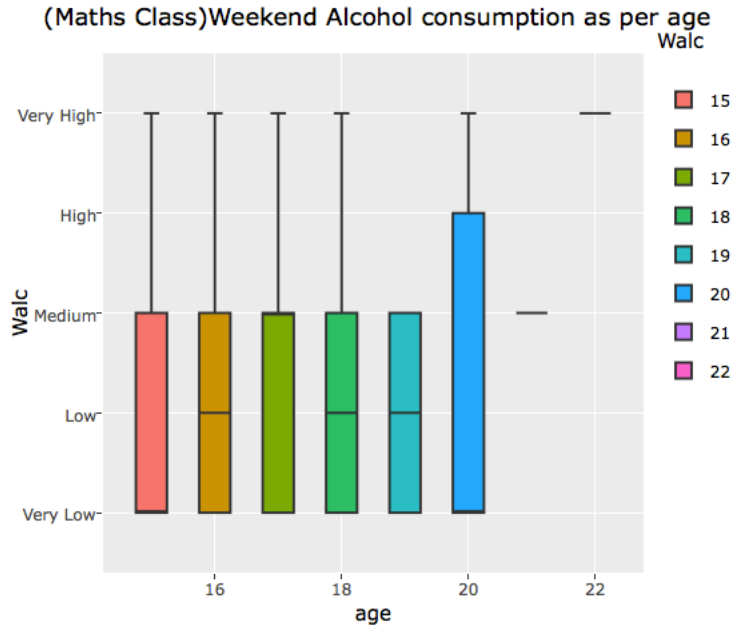
It can be easily seen from the two plots below that boys are drinking more than girls on weekends and girls highest is medium consumption level and on average low and for boys it high and on average it medium consumption level which shows boys are more prone hazards and are consuming a lot of alcohol.

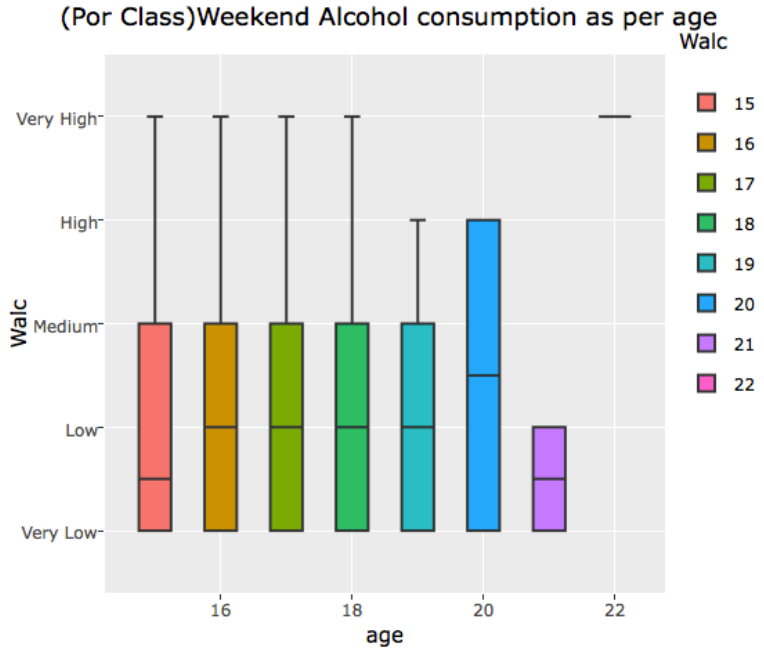




Weekend Alcohol based on age

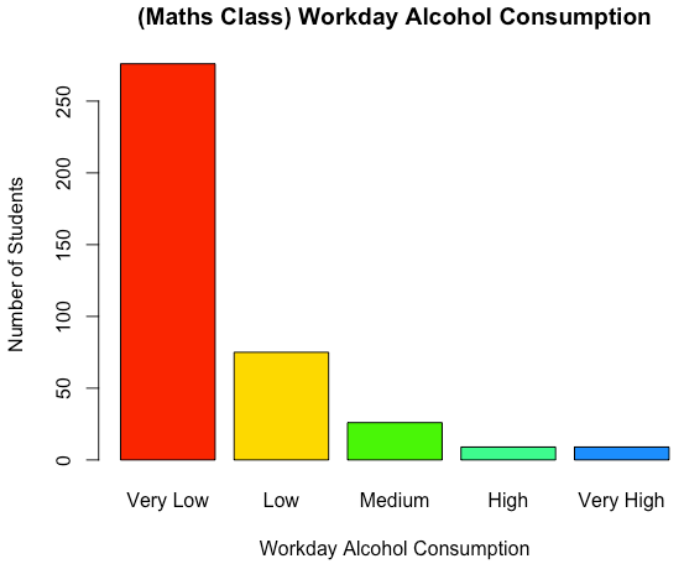
It can be seen in the results of both datasets that students of 20 years of age are consuming the highest level of alcohol compared to all students of other age.

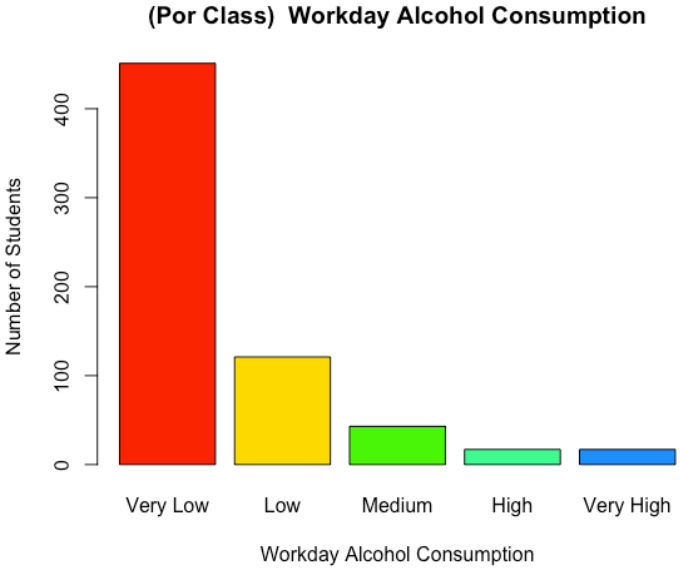




Workday Alcohol Consumption

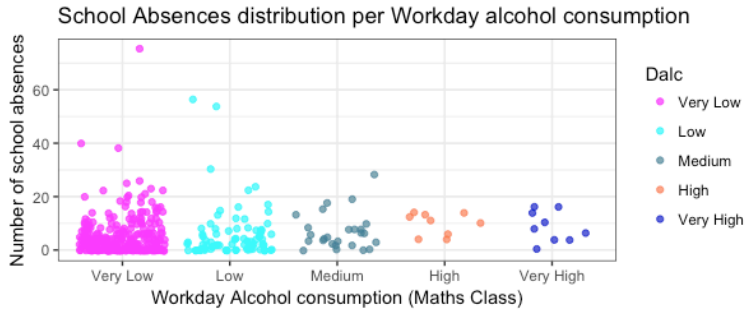
On workday the alcohol consumption of the majority is very low but still there are traces of medium and high consumption are found instead there should be no consumption at all as alcohol can ruin the life of these student.

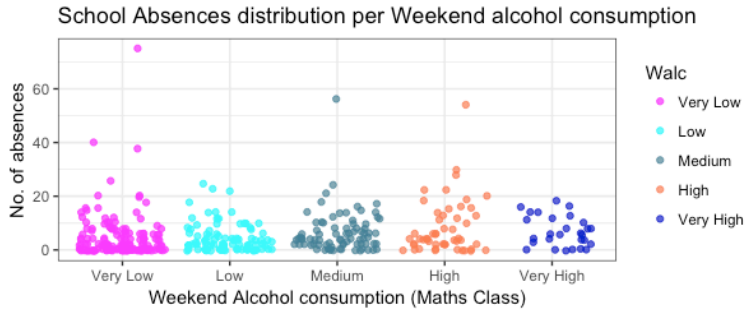


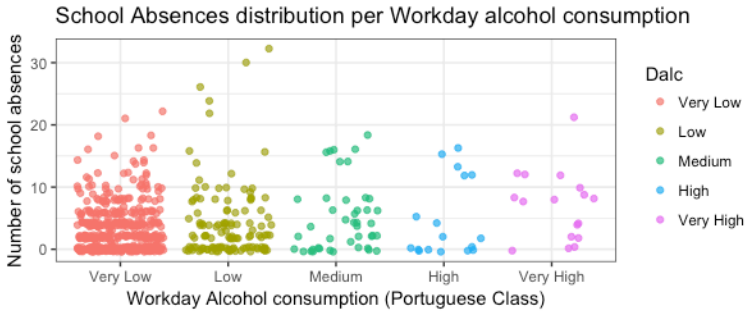


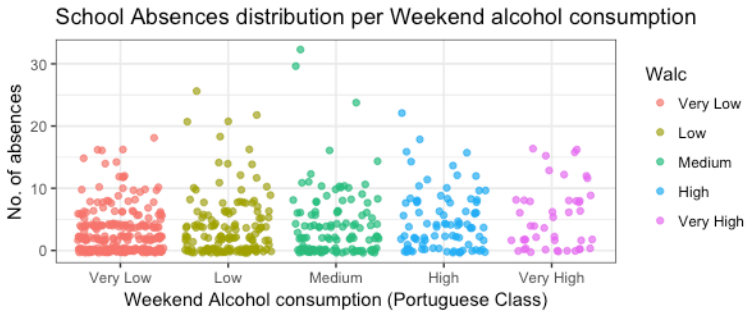
School Absences and Alcohol consumption

Getting absent in school can negatively impact students studies



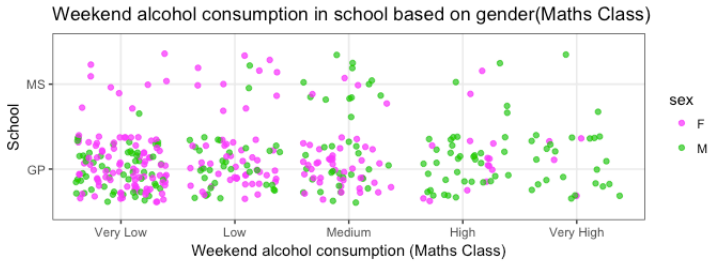


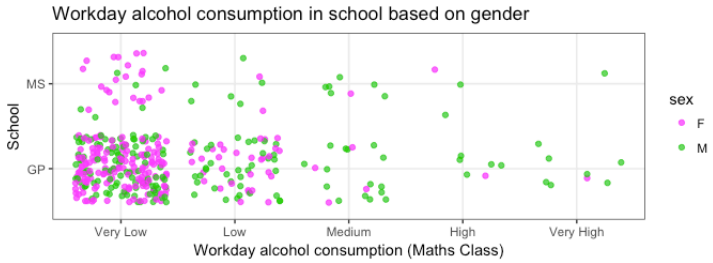


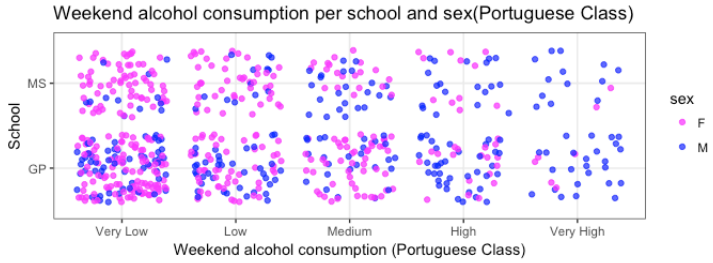


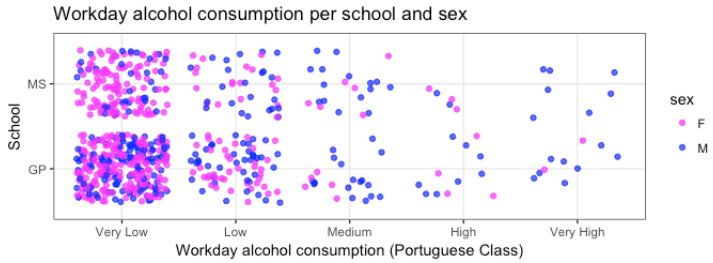
Weekend Alcohol Consumption based on gender (sex)

Majority of students are in very low consumption but In the plots it can be seen the very high level of alcohol consumption is done mainly by boys on weekends and there's also instance found many boys are consuming high level of alcohol on workdays which will have direct impact on their studies and the graphs below shows there is a good need for good direction for students of both schools.



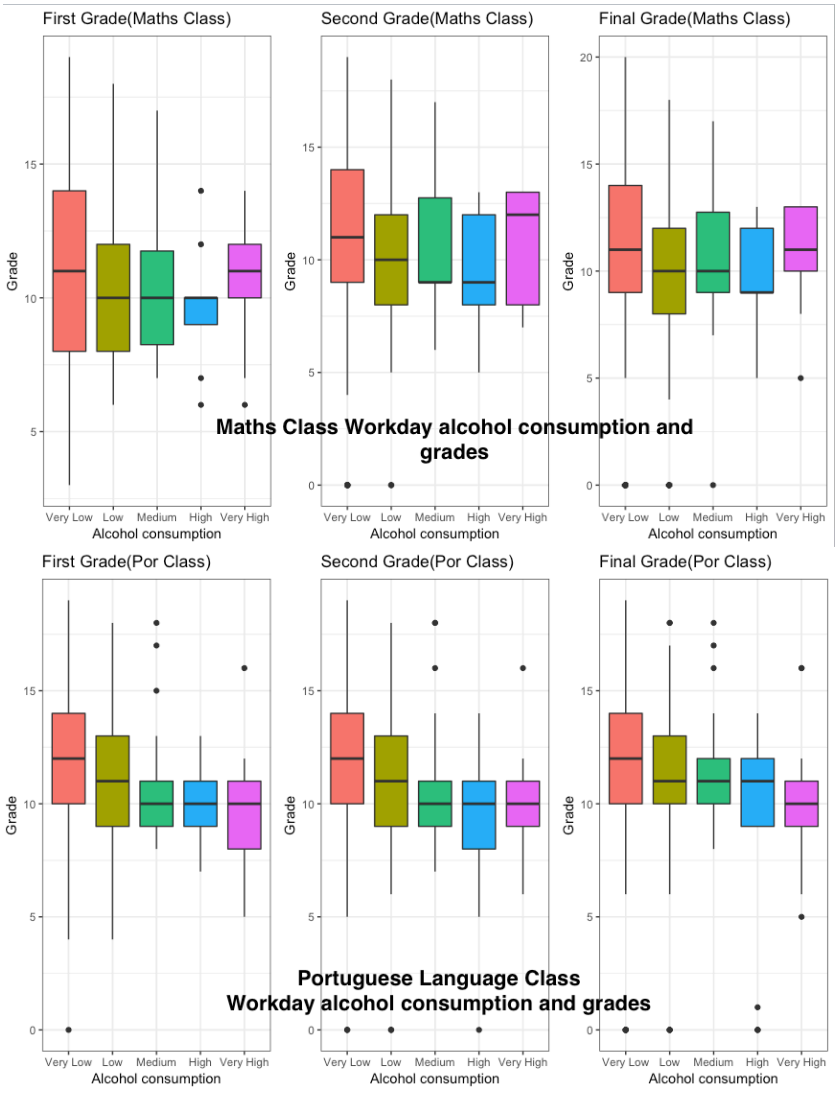






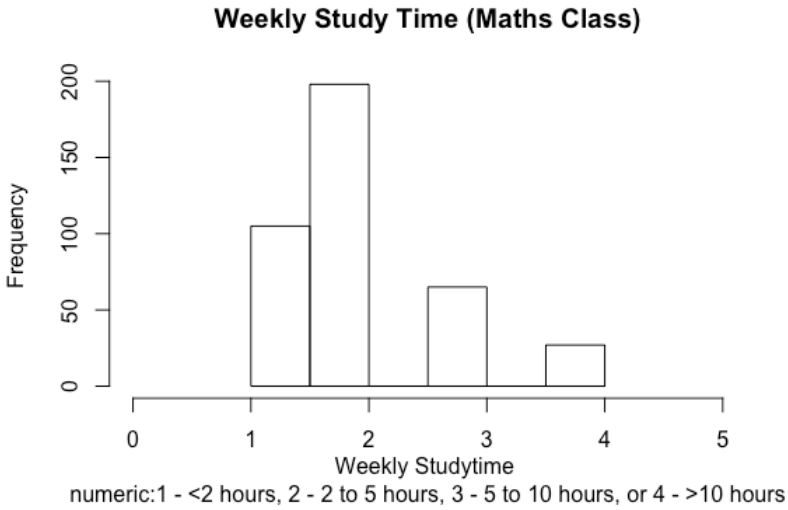
Workday Alcohol Consumption and Grades

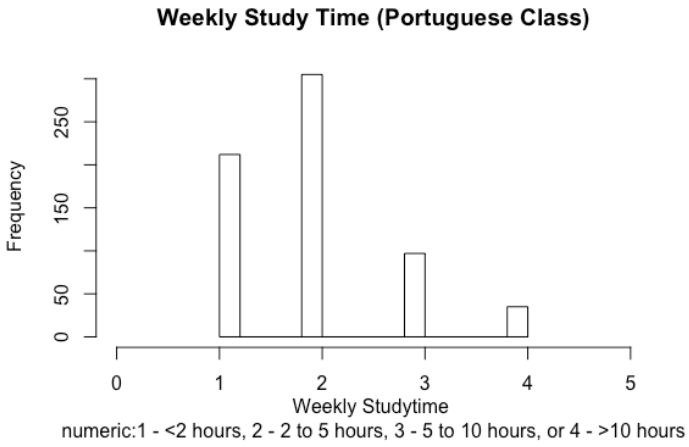
From the plot below it can be seen how alcohol has an impact on grades. The students consuming no alcohol / very low have better grades in all G1, G2, G3 and for both datasets



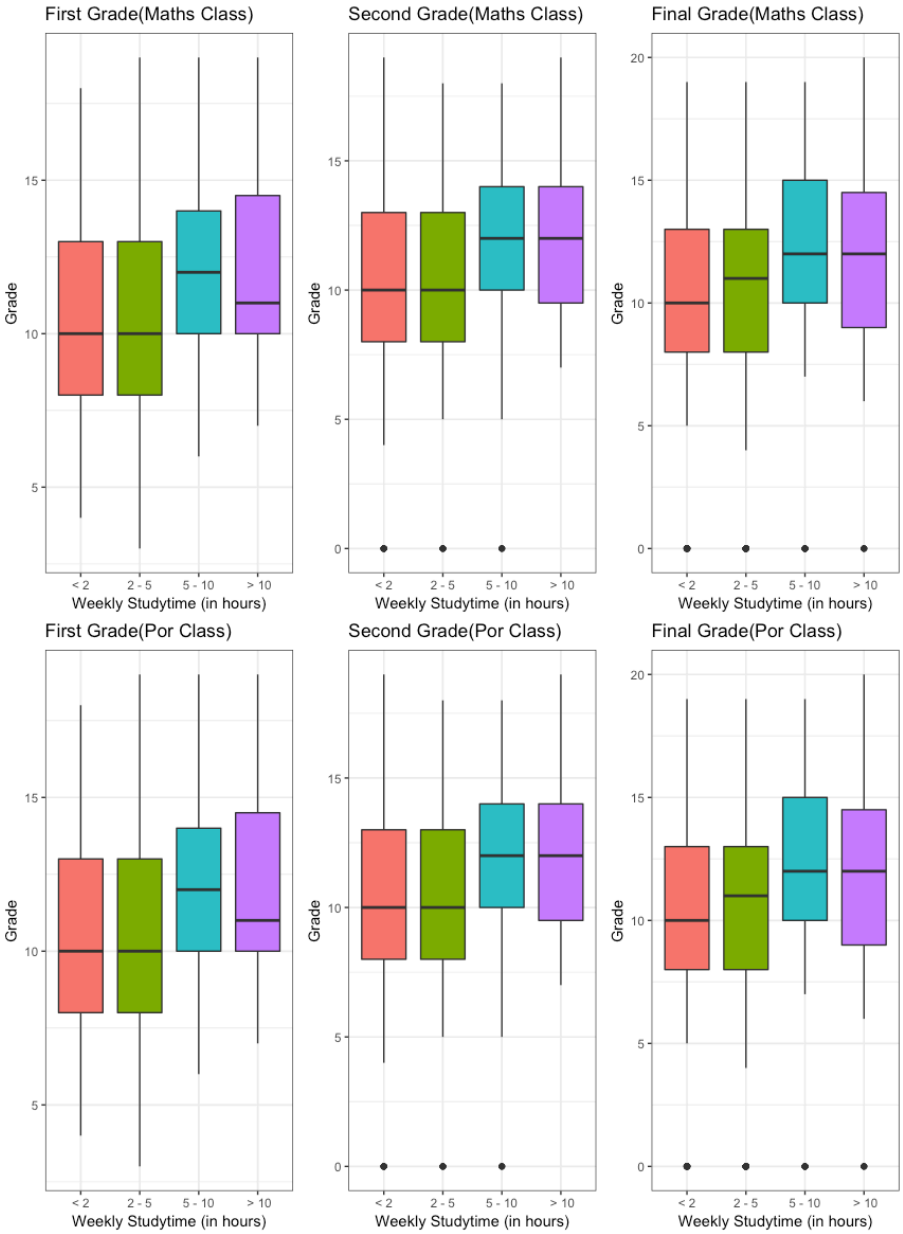
Weekly Study Time

The histograms below show weekly study time of student in both dataset (on average its 2 - 5 hours) more hours of study will result in better grades as we can see in plots below. So, the students should be suggested to study more and take part in activities and avoid alcohol



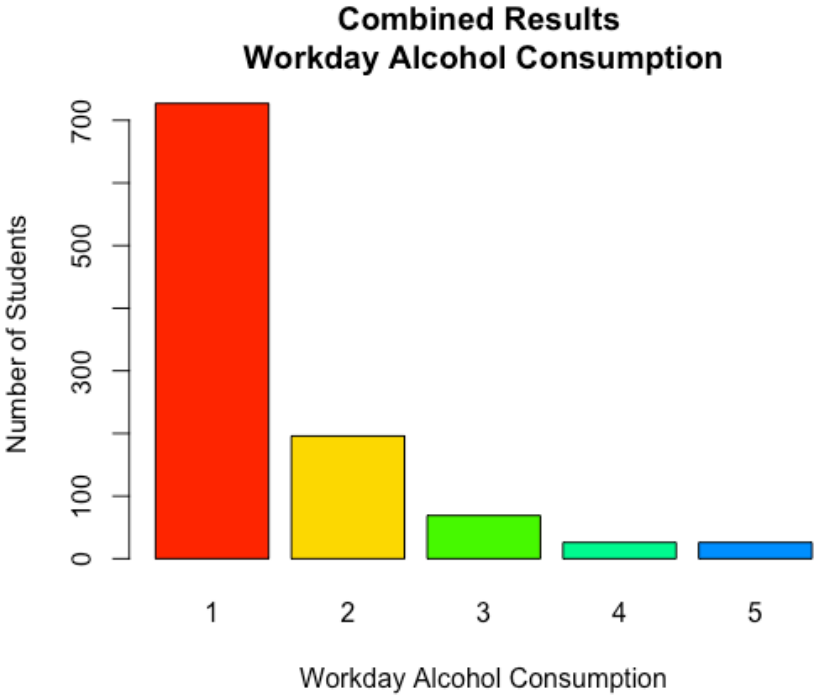


Better grades achieved by students when the weekly study hours are higher (for both datasets)



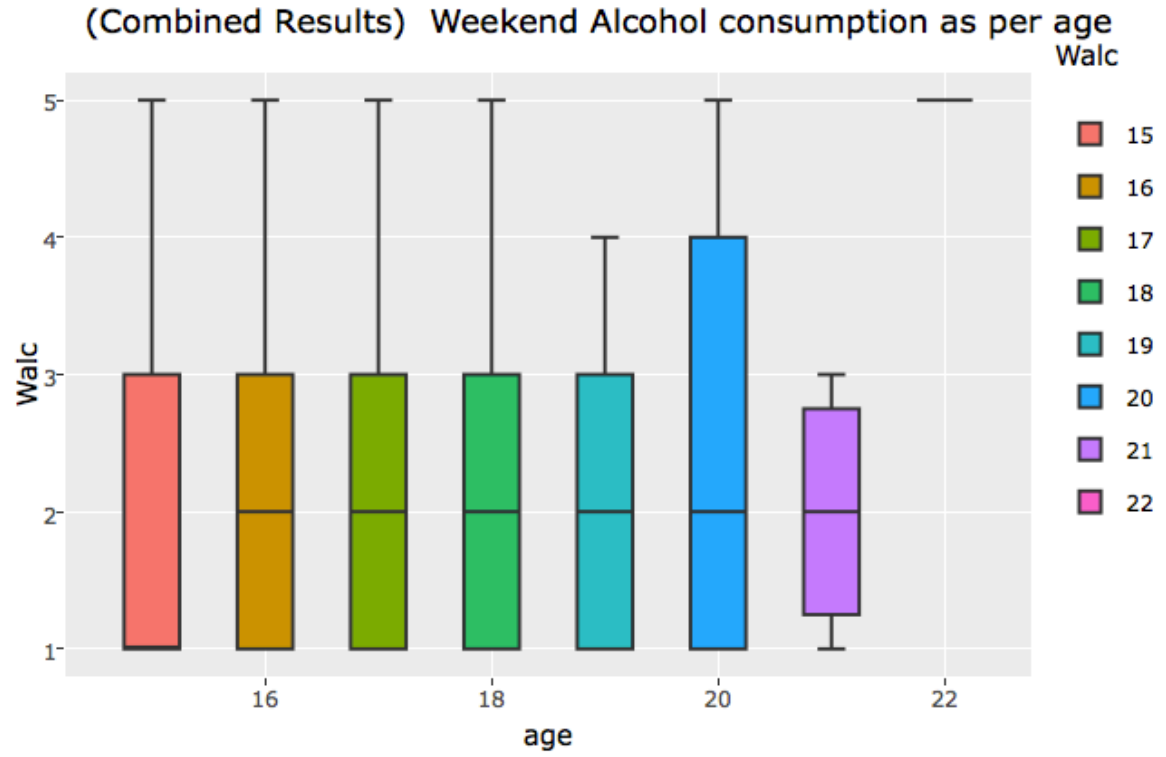
Combined Analysis (Results)

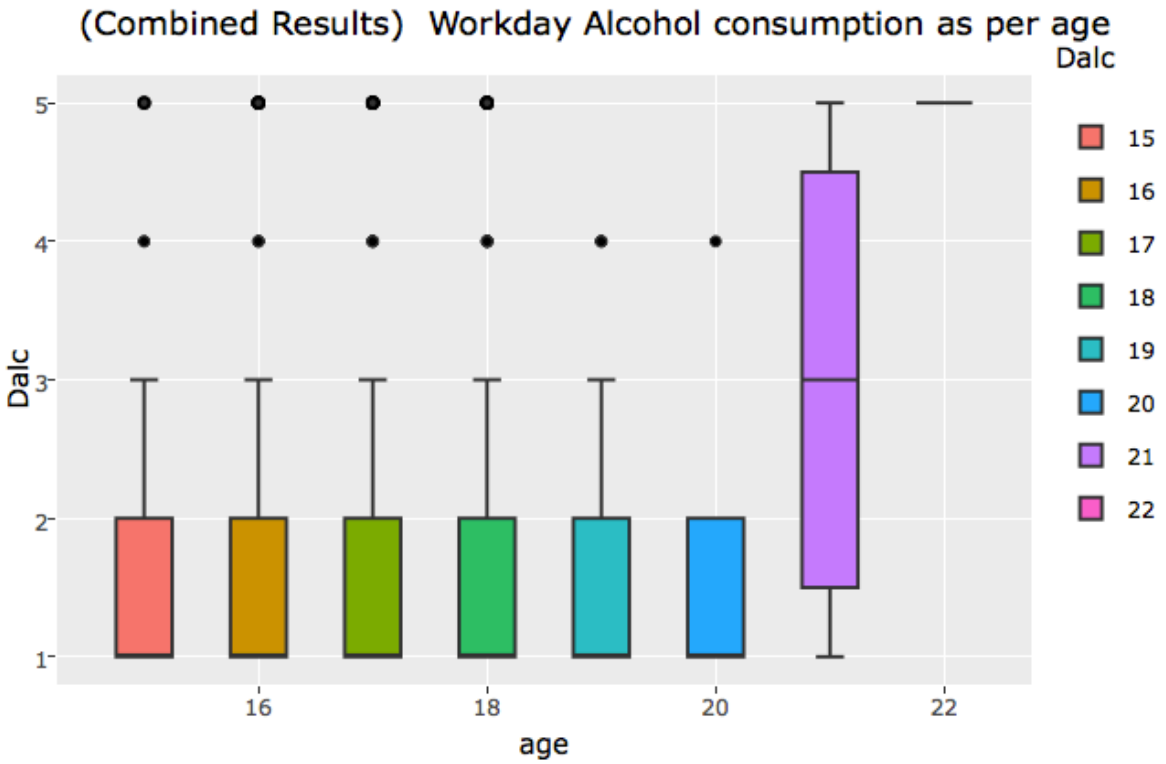
Workday Alcohol consumption histogram to show the alcohol consumption by the students of both courses.



Weekend and Workday alcohol consumption as per age

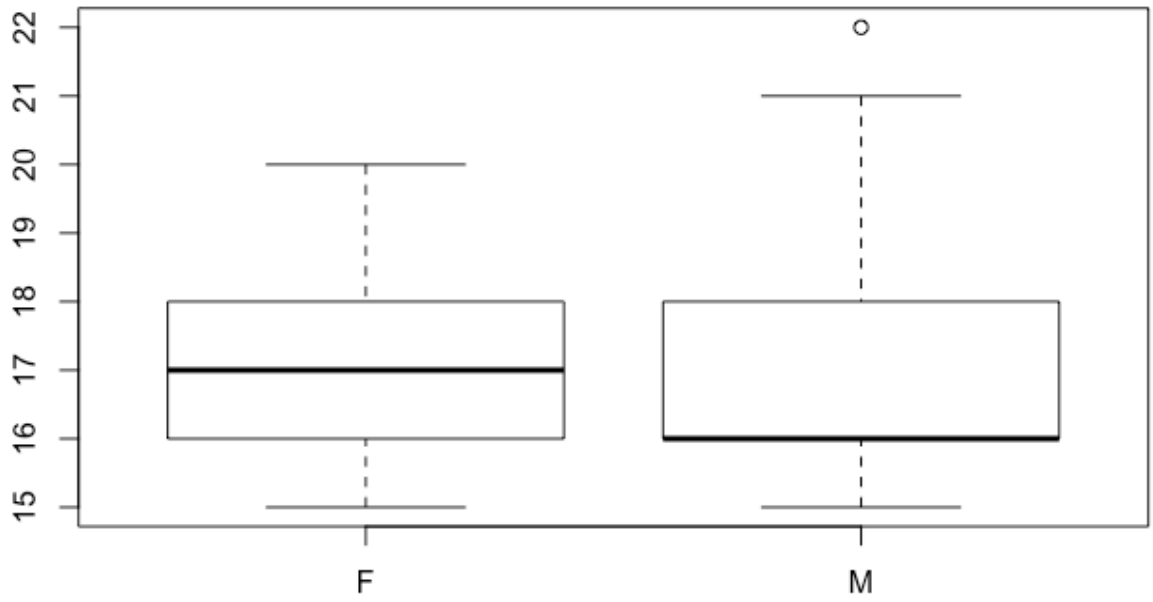
It can be seen on the weekends the 20 years old students are consuming high levels of alcohol and which is not a good sign as if they drink a lot on weekends their performance in school will definitely decrease for the coming week.





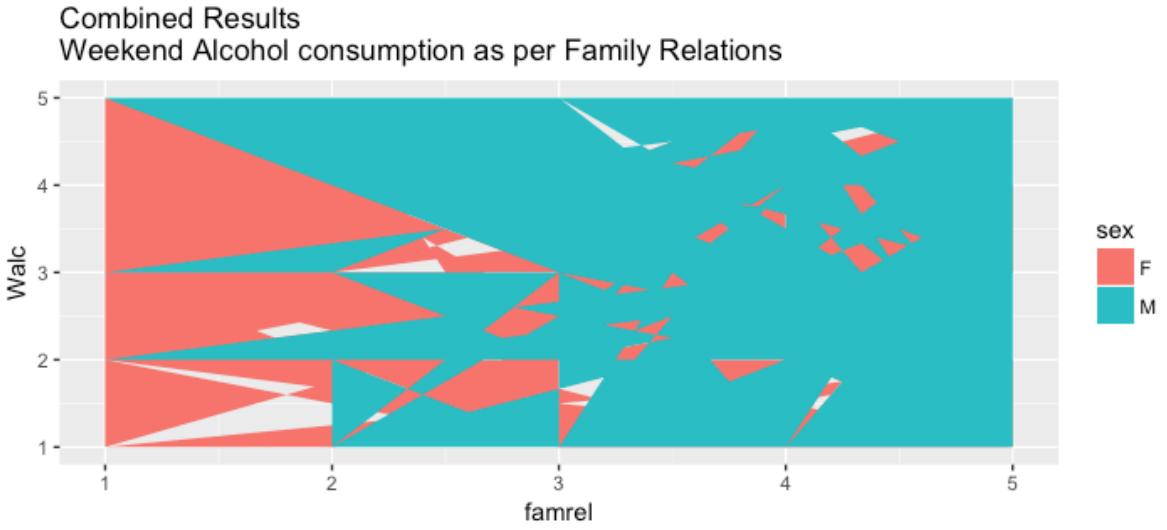
Age and Gender (Sex)

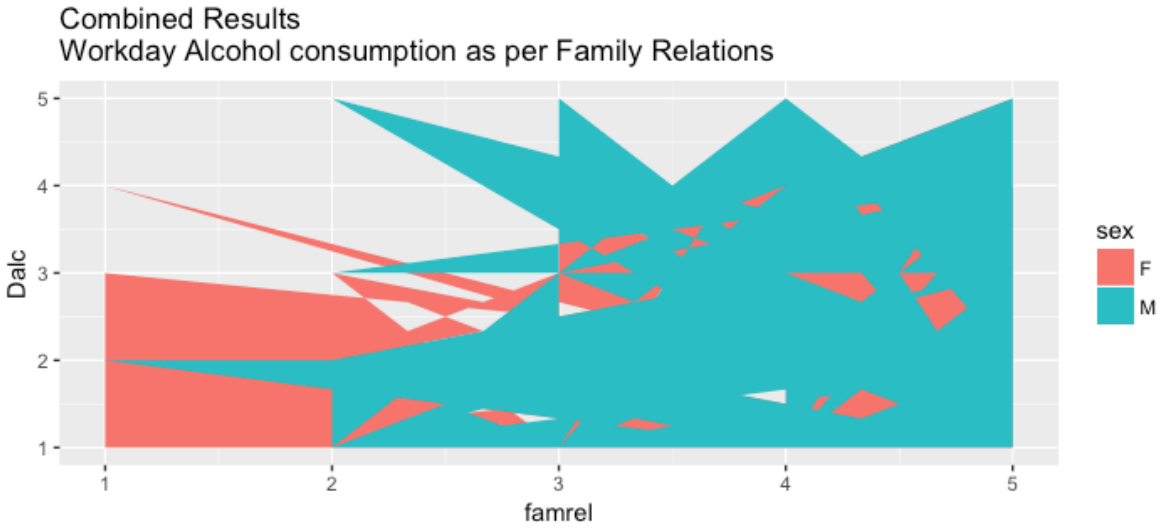
The plot below shows what is the age range of students in combined data and findings say on average the female students age is 17 years and for males its 16 years.



Alcohol Consumption distribution and family relations

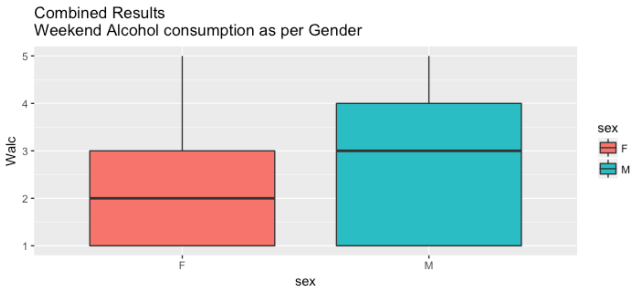
For the plot below we can see male students that have good family relations still do consume a lot of alcohol whereas for female students the results are much better than the other sex for both weekend and workday consumption distribution.

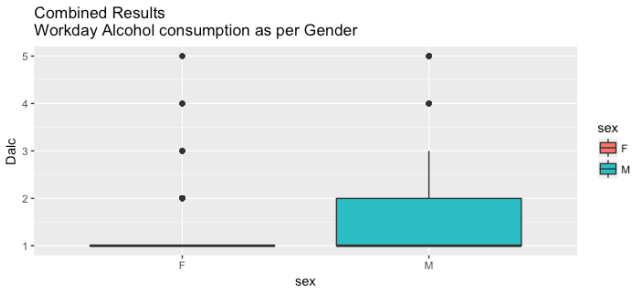




Alcohol consumption based on gender

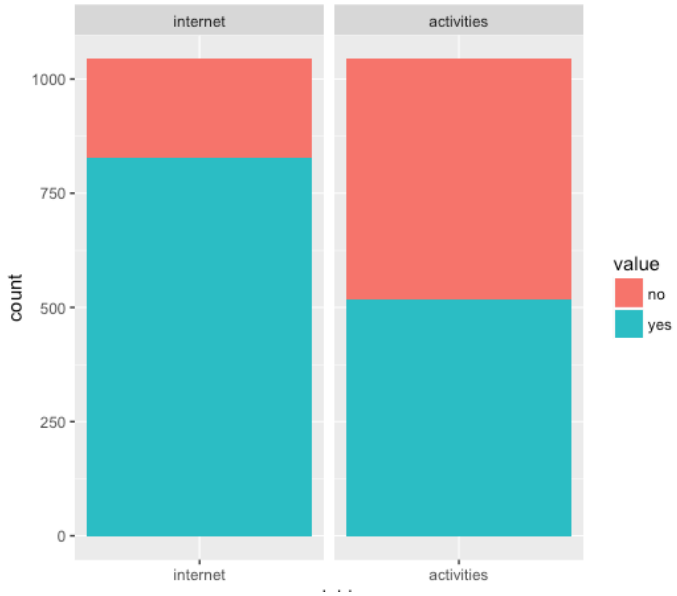
It can be seen for the plot below on weekends males are consuming the high level of alcohol and females on the other hand to medium range or less.





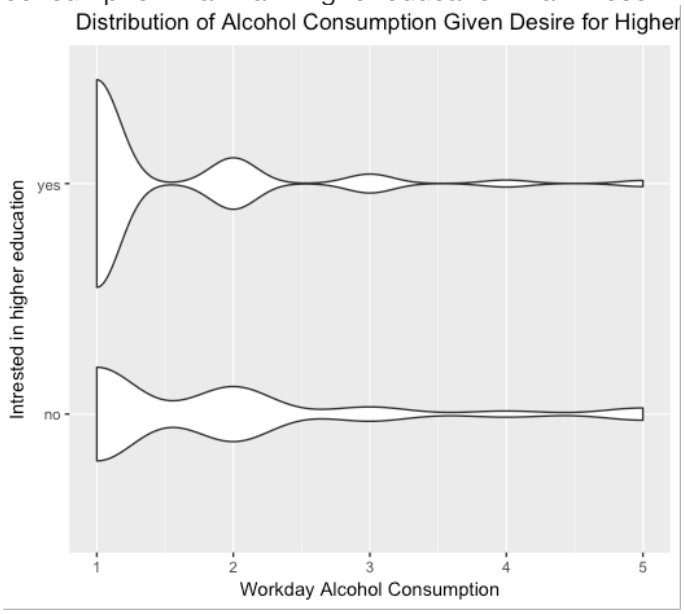
Plot to show to answer the question "How many students have access to internet and other activities

The results show that about 70-80% of the students have internet access which may be unsupervised and hence parental control on internet access should be strong so the students don't consume bad information from the internet and the level of activities is only about 50% which is somewhat less and students must be encouraged to take part in various activities.



Alcohol consumption based on desire for higher education

The plot shows the conditional distribution of workday alcohol consumption given the students desire for further higher education. There is a larger distribution of people in the very low alcohol consumption that want higher education than those who do not want higher education.



Reference: This plot has been done with the help of this site:
https://www.kaggle.com/interkf/d/uciml/student-alcohol-consumption/alcohol-consumption-from-portuguese-school