**Summary**

Data: student\_mat, student who studies math with data about their math grade, alcohol consumption, health, absenses etc.

Visualization of student alcohol consumption VS various other variable

Regression analysis and visualization such as histogram, scatter plot etc.

**Analysis**

Columns and its level

1. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2. sex - student's sex (binary: 'F' - female or 'M' - male)
3. age - student's age (numeric: from 15 to 22)
4. address - student's home address type (binary: 'U' - urban or 'R' - rural)
5. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
9. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
10. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
11. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13. 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)
14. studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15. failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16. schoolsup - extra educational support (binary: yes or no)
17. famsup - family educational support (binary: yes or no)
18. paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19. activities - extra-curricular activities (binary: yes or no)
20. nursery - attended nursery school (binary: yes or no)
21. higher - wants to take higher education (binary: yes or no)
22. internet - Internet access at home (binary: yes or no)
23. romantic - with a romantic relationship (binary: yes or no)
24. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25. freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26. goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29. health - current health status (numeric: from 1 - very bad to 5 - very good)
30. absences - number of school absences (numeric: from 0 to 93)

These grades are related with the course subject, Math or Portuguese:

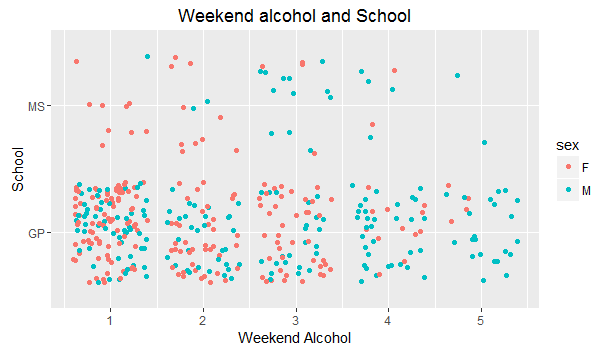
1. G1 - first period grade (numeric: from 0 to 20)
2. G2 - second period grade (numeric: from 0 to 20)
3. G3 - final grade (numeric: from 0 to 20, output target)

Added column during analysis:

1. Alcohol- sum of Dalc and Walc: Weekday and Workday consumption

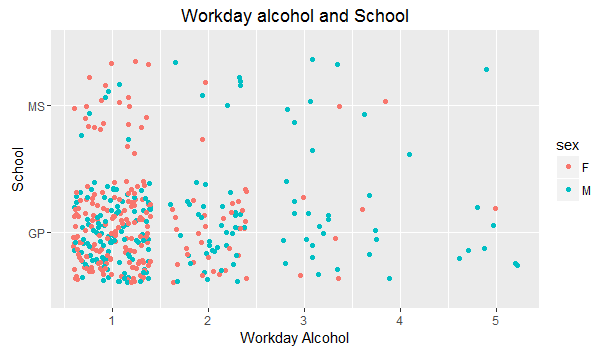
**Weekend Alcohol and School**

*viz1<- ggplot(math, aes(x= Walc, y= school, color= sex)) + geom\_jitter() + xlab("Weekend Alcohol") + ylab("School") + ggtitle("Weekend alcohol and School")*



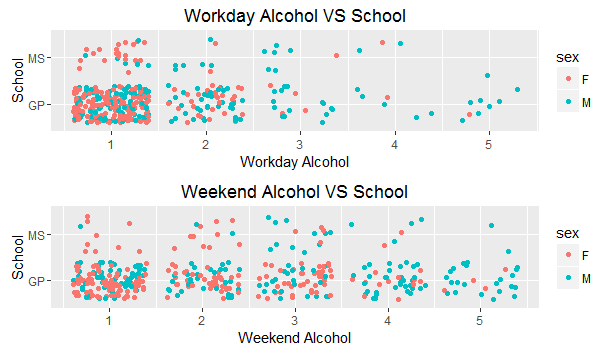
**Workday Alcohol and School**

*viz2<- ggplot(math, aes(x= Dalc, y= school, color= sex)) + geom\_jitter() + xlab("Workday Alcohol") + ylab("School") + ggtitle("Workday alcohol and School")*



**Weekday/Workday and School:** Alcohol consumption based on school for both sex

*grid.arrange(viz1, viz2)*

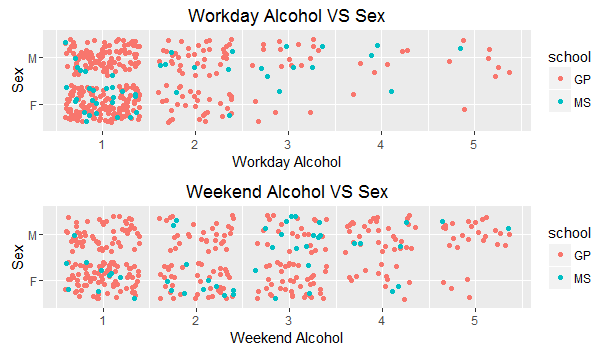
****

**Weekend/ Workday Alcohol and Sex:** Alcohol consumption based on sex on both schools

*viz3<- ggplot(math, aes(x= Walc, y= sex, color= school)) + geom\_jitter() + xlab("Weekend Alcohol") + ylab("Sex") + ggtitle("Weekend alcohol and Sex")*

*viz4<- ggplot(math, aes(x= Dalc, y= sex, color= school)) + geom\_jitter() + xlab("Workday Alcohol") + ylab("Sex") + ggtitle("Workday alcohol and Sex")*

*grid.arrange(viz3, viz4)*

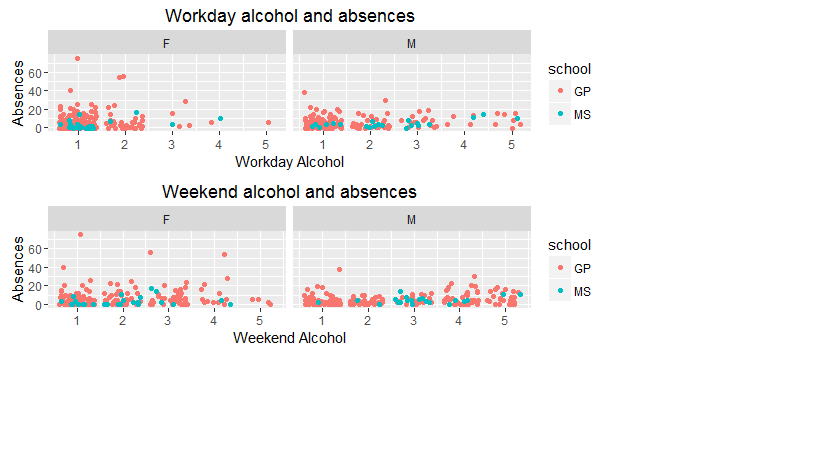
****

**Weekend/ Workday Alcohol and Absences:** How alcohol affects the absence from school

*viz9<-ggplot(math, aes(x= Dalc, y= absences, color= school)) + geom\_jitter() + xlab("Workday Alcohol") + ylab("Absences") + ggtitle("Workday alcohol and absences") + facet\_wrap(~sex)*

*viz10<-ggplot(math, aes(x= Walc, y= absences, color= school)) + geom\_jitter() + xlab("Weekend Alcohol") + ylab("Absences") + ggtitle("Weekend alcohol and absences") + facet\_wrap(~sex)*

grid.arrange(viz9, viz10)

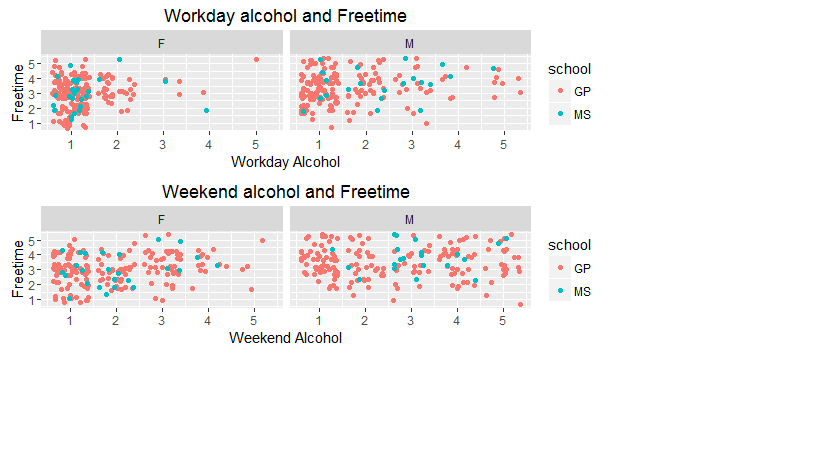
****

**Weekend/ Workday Alcohol and Freetime:** How freetime is used by girls and boys in both schools

*viz7<- ggplot(math, aes(x= Dalc, y= freetime, color= school)) + geom\_jitter() + xlab("Workday Alcohol") + ylab("Freetime") + ggtitle("Workday alcohol and Freetime") + facet\_wrap(~sex)*

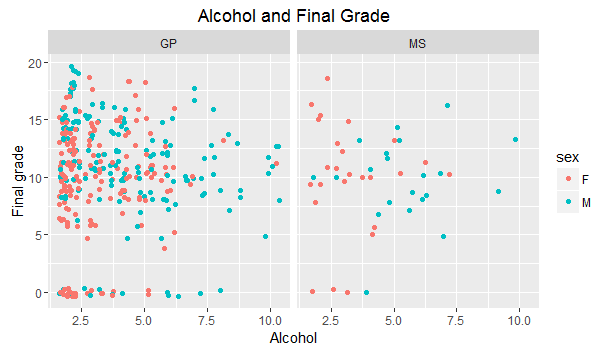
*viz8<- ggplot(math, aes(x= Walc, y= freetime, color= school)) + geom\_jitter() + xlab("Weekend Alcohol") + ylab("Freetime") + ggtitle("Weekend alcohol and Freetime") +facet\_wrap(~sex)*

grid.arrange(viz7, viz8)

****

**Alcohol and Final Grade:** How alcohol affects the final grade

viz11<-ggplot(math, aes(x= alcohol, y= G3, color= sex)) + geom\_jitter() + xlab("Alcohol") + ylab("Final grade") + ggtitle("Alcohol and Final Grade") +facet\_wrap(~school)

****

**Regression equation:** Regression analysis of final grade based on various factors

*lm(formula = G3 ~ freetime + failures + goout + Dalc + health + absences , data= math)*

Final\_grade\_math= 10.63 + 0.359 \* freetime -2.154 \* failures -0.51 \*gout +

0.27 \* studytime + 0.09 \* alcohol