

```

library(ggplot2)
library(corrplot)

##### 1 #####

ds = read.csv2("/Users/mattiaspiga/Dropbox/Marco-Mattia/Studio Universitario/04 - Metodi Statistici Data Science/03 -
Esercitazione/00 - ESAME FINALE/00 - Classwork 3Cat/student-mattt.csv")

ds = read.csv2("/Users/mattiaspiga/Dropbox/Marco-Mattia/Studio Universitario/04 - Metodi Statistici Data Science/03 -
Esercitazione/CLASSWORK01/02 - Analisi CW Marco Spiga/student-mattt.csv")

str(ds)

#correzione tipologia di variabili
ds$Medu = as.factor(ds$Medu)
ds$Fedu = as.factor(ds$Fedu)

View(ds)
summary(ds)

##### 2 #####
#hanno un esempio fatto a lezione di questo pezzo di codice
#e dovevano fare una cosa analoga nell'homework3
#GRAFICO1 - Corrplot
numeric_cols = sapply(ds, is.numeric)
corrplot(cor(ds[,numeric_cols]), method="pie")

#GRAFICO2
ggplot( ds ) +
  geom_bar( aes(G1) )

ggplot( ds ) +
  geom_bar( aes(G2) )

ggplot( ds ) +
  geom_bar( aes(G3) )

ggplot( ds ) +
  geom_bar( aes(studytime) )

#GRAFICO3
ggplot( ds, aes(x=studytime, fill=sex)) +
  geom_bar(position="dodge")

ggplot( ds, aes(x=G3, fill=sex)) +
  geom_bar(position="dodge")

#GRAFICO4

#bonus (qualche studente credo non avra' problemi a produrre qualcosa del genere)
temp = data.frame(G=rep("G1"), value=ds$G1)
temp = rbind(temp, data.frame(G=rep("G2"), value=ds$G2))
temp = rbind(temp, data.frame(G=rep("G3"), value=ds$G3))

#boxplot di G1, G2, G3
ggplot(temp, aes(x = G, y = value)) +
  geom_boxplot() + theme_bw()

#GRAFICO5
#densita' di G1, G2, G3
ggplot(temp, aes(x = value, fill = G)) +
  geom_density(alpha = 0.7) + theme_bw() +
  theme(legend.position = c(0.8, 0.8))

##### 3 #####
ds2 = ds[,-which(colnames(ds) == "G1" | colnames(ds) == "G2")]

##### 4 #####
lm1 = lm(G3~failures+sex, data=ds2)
summary(lm1)
plot(lm1)

##### 5 #####
G3_qual = ifelse(ds2$G3 < 12, "Insufficiente",
  ifelse(ds2$G3 < 15, "Buono", "Ottimo")
)

ds2$G3_qual = ordered(G3_qual, levels=c("Insufficiente", "Buono", "Ottimo") )
summary(ds2$G3_qual)

```

```
##### 6 #####
library(MASS)
set.seed(1)
train = sample(nrow(ds2), nrow(ds2)*0.7)

#Si ricorderanno che devono rimuovere G3 o l'lda non funziona causa collinearita'?
lda.fit = lda(G3_qual~.-G3, data=ds2[train,])
lda.fit

lda.pred = predict(lda.fit, ds2[-train,])

table(lda.pred$class, ds2[-train, "G3_qual"])
mean(lda.pred$class == as.character(ds2[-train, "G3_qual"]))

#GRAFICO6
#bonus (l'ho mostrato oggi a lezione)
lda.data = cbind(ds2[-train,], lda.pred$x)
ggplot(lda.data, aes(LD1, LD2)) +
  geom_point(aes(color = G3_qual))

##### 7 #####
library(boot)
set.seed(1)
#si ricorderanno di rimuovere G3_qual?
glm1 = glm(G3~failures+sex-G3_qual, data=ds2)
cv.err=cv.glm(ds2,glm1, K=10)
cv.err$delta
```