

Linear Regression

Part 1

```
df <- read.csv('student-mat.csv', sep=';')
head(df)
```

```
##  school sex age address famsize Pstatus Medu Fedu Mjob Fjob
## 1    GP  F  18      U    GT3      A    4    4 at_home teacher
## 2    GP  F  17      U    GT3      T    1    1 at_home  other
## 3    GP  F  15      U    LE3      T    1    1 at_home  other
## 4    GP  F  15      U    GT3      T    4    2 health services
## 5    GP  F  16      U    GT3      T    3    3  other  other
## 6    GP  M  16      U    LE3      T    4    3 services other
##
##      reason guardian traveltime studytime failures schoolsup famsup
paid
## 1    course    mother          2          2          0      yes    no
no
## 2    course    father          1          2          0      no    yes
no
## 3    other    mother          1          2          3      yes    no
yes
## 4    home    mother          1          3          0      no    yes
yes
## 5    home    father          1          2          0      no    yes
yes
## 6 reputation mother          1          2          0      no    yes
yes
##  activities nursery higher internet romantic famrel freetime goout
Dalc
## 1      no      yes    yes      no      no      4      3      4
1
## 2      no      no    yes    yes    no      5      3      3
1
## 3      no      yes    yes    yes    no      4      3      2
2
## 4      yes     yes    yes    yes    yes     3      2      2
1
## 5      no      yes    yes    no    no      4      3      2
1
## 6      yes     yes    yes    yes    no      5      4      2
1
##  Walc health absences G1 G2 G3
## 1    1      3          6  5  6  6
## 2    1      3          4  5  5  6
## 3    3      3         10  7  8 10
## 4    1      5          2 15 14 15
## 5    2      5          4  6 10 10
## 6    2      5         10 15 15 15
```

summary(df)

```
##  school  sex      age      address famsize  Pstatus      Medu
##  GP:349   F:208   Min.    :15.0   R: 88   GT3:281   A: 41   Min.    :
0.000
##  MS: 46   M:187   1st Qu.:16.0   U:307   LE3:114   T:354   1st
Qu.:2.000
##                      Median :17.0                      Median :
3.000
##                      Mean    :16.7                      Mean    :
2.749
##                      3rd Qu.:18.0                      3rd
Qu.:4.000
##                      Max.    :22.0                      Max.    :
4.000
##      Fedu      Mjob      Fjob      reason
##  Min.    :0.000   at_home : 59   at_home : 20   course    :145
##  1st Qu.:2.000   health  : 34   health  : 18   home      :109
##  Median :2.000   other   :141   other   :217   other     : 36
##  Mean    :2.522   services:103   services:111   reputation:105
##  3rd Qu.:3.000   teacher : 58   teacher : 29
##  Max.    :4.000
##  guardian  traveltime  studytime  failures
schoolsup
##  father: 90   Min.    :1.000   Min.    :1.000   Min.    :0.0000   no :
344
##  mother:273   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.0000   yes:
51
##  other : 32   Median :1.000   Median :2.000   Median :0.0000
##
##                      Mean    :1.448   Mean    :2.035   Mean    :0.3342
##
##                      3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:0.0000
##
##                      Max.    :4.000   Max.    :4.000   Max.    :3.0000
##
##  famsup      paid      activities nursery  higher  internet
romantic
##  no :153     no :214     no :194     no : 81   no : 20   no : 66   no :
263
##  yes:242     yes:181     yes:201     yes:314   yes:375   yes:329
yes:132
##
##
##
```

```
##
```

```
##      famrel      freetime      goout      Dalc
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:4.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:1.000
## Median :4.000   Median :3.000   Median :3.000   Median :1.000
## Mean   :3.944   Mean   :3.235   Mean   :3.109   Mean   :1.481
## 3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:2.000
## Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##      Walc      health      absences      G1
## Min.   :1.000   Min.   :1.000   Min.   : 0.000   Min.   : 3.00
## 1st Qu.:1.000   1st Qu.:3.000   1st Qu.: 0.000   1st Qu.: 8.00
## Median :2.000   Median :4.000   Median : 4.000   Median :11.00
## Mean   :2.291   Mean   :3.554   Mean   : 5.709   Mean   :10.91
## 3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.: 8.000   3rd Qu.:13.00
## Max.   :5.000   Max.   :5.000   Max.   :75.000   Max.   :19.00
##      G2      G3
## Min.   : 0.00   Min.   : 0.00
## 1st Qu.: 9.00   1st Qu.: 8.00
## Median :11.00   Median :11.00
## Mean   :10.71   Mean   :10.42
## 3rd Qu.:13.00   3rd Qu.:14.00
## Max.   :19.00   Max.   :20.00
```

```
str(df)
```

```
## 'data.frame':   395 obs. of  33 variables:
## $ school      : Factor w/ 2 levels "GP","MS": 1 1 1 1 1 1 1 1 1 1 ...
## $ sex         : Factor w/ 2 levels "F","M": 1 1 1 1 1 2 2 1 2 2 ...
## $ age         : int  18 17 15 15 16 16 16 17 15 15 ...
## $ address     : Factor w/ 2 levels "R","U": 2 2 2 2 2 2 2 2 2 2 ...
## $ famsize     : Factor w/ 2 levels "GT3","LE3": 1 1 2 1 1 2 2 1 2
1 ...
## $ Pstatus    : Factor w/ 2 levels "A","T": 1 2 2 2 2 2 2 1 1 2 ...
## $ Medu       : int   4 1 1 4 3 4 2 4 3 3 ...
## $ Fedu       : int   4 1 1 2 3 3 2 4 2 4 ...
## $ Mjob       : Factor w/ 5 levels "at_home","health",...: 1 1 1 2 3 4
3 3 4 3 ...
## $ Fjob       : Factor w/ 5 levels "at_home","health",...: 5 3 3 4 3 3
3 5 3 3 ...
## $ reason     : Factor w/ 4 levels "course","home",...: 1 1 3 2 2 4 2
2 2 2 ...
## $ guardian   : Factor w/ 3 levels "father","mother",...: 2 1 2 2 1 2
2 2 2 2 ...
## $ traveltime: int   2 1 1 1 1 1 1 2 1 1 ...
## $ studytime  : int   2 2 2 3 2 2 2 2 2 2 ...
## $ failures   : int   0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup  : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 2 1
1 ...
## $ famsup     : Factor w/ 2 levels "no","yes": 1 2 1 2 2 2 1 2 2
```

```

2 ...
## $ paid      : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 2
2 ...
## $ activities: Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 1 1 1
2 ...
## $ nursery   : Factor w/ 2 levels "no","yes": 2 1 2 2 2 2 2 2 2
2 ...
## $ higher    : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2
2 ...
## $ internet  : Factor w/ 2 levels "no","yes": 1 2 2 2 1 2 2 1 2
2 ...
## $ romantic  : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1
1 ...
## $ famrel    : int   4 5 4 3 4 5 4 4 4 5 ...
## $ freetime  : int   3 3 3 2 3 4 4 1 2 5 ...
## $ goout     : int   4 3 2 2 2 2 4 4 2 1 ...
## $ Dalc      : int   1 1 2 1 1 1 1 1 1 1 ...
## $ Walc      : int   1 1 3 1 2 2 1 1 1 1 ...
## $ health    : int   3 3 3 5 5 5 3 1 1 5 ...
## $ absences  : int   6 4 10 2 4 10 0 6 0 0 ...
## $ G1        : int   5 5 7 15 6 15 12 6 16 14 ...
## $ G2        : int   6 5 8 14 10 15 12 5 18 15 ...
## $ G3        : int   6 6 10 15 10 15 11 6 19 15 ...

```

Checking for Null Values

```
any(is.na(df))
```

```
## [1] FALSE
```

Loading libraries:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 3.3.2
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

library(corrgram)

## Warning: package 'corrgram' was built under R version 3.3.2

library(corrplot)
```

Num Only

```
num.cols <- sapply(df,is.numeric)
```

Filter

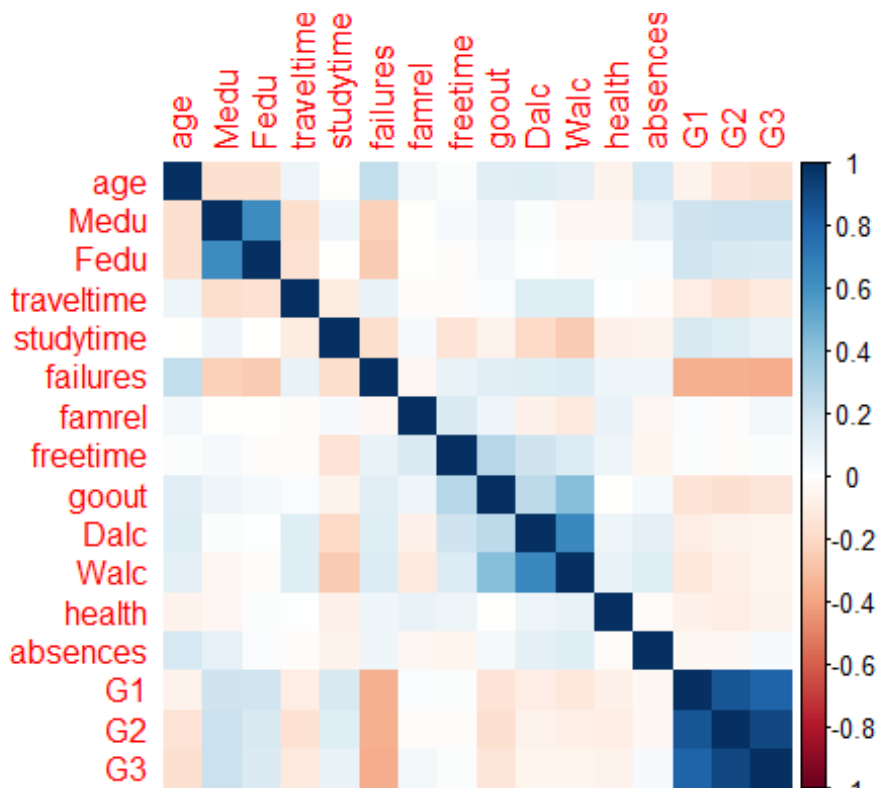
```
cor.data <- cor(df[,num.cols])
print(cor.data)
```

```
##           age           Medu           Fedu    traveltime
## age      1.000000000 -0.163658419 -0.163438069  0.070640721
## Medu     -0.163658419  1.000000000  0.623455112 -0.171639305
## Fedu     -0.163438069  0.623455112  1.000000000 -0.158194054
## traveltime 0.070640721 -0.171639305 -0.158194054  1.000000000
## studytime -0.004140037  0.064944137 -0.009174639 -0.100909119
## failures   0.243665377 -0.236679963 -0.250408444  0.092238746
## famrel     0.053940096 -0.003914458 -0.001369727 -0.016807986
## freetime   0.016434389  0.030890867 -0.012845528 -0.017024944
## goout      0.126963880  0.064094438  0.043104668  0.028539674
## Dalc       0.131124605  0.019834099  0.002386429  0.138325309
## Walc       0.117276052 -0.047123460 -0.012631018  0.134115752
## health     -0.062187369 -0.046877829  0.014741537  0.007500606
## absences   0.175230079  0.100284818  0.024472887 -0.012943775
## G1        -0.064081497  0.205340997  0.190269936 -0.093039992
## G2        -0.143474049  0.215527168  0.164893393 -0.153197963
## G3        -0.161579438  0.217147496  0.152456939 -0.117142053
##           studytime   failures           famrel   freetime
goout
## age      -0.004140037  0.24366538  0.053940096  0.01643439
0.126963880
## Medu      0.064944137 -0.23667996 -0.003914458  0.03089087
0.064094438
## Fedu     -0.009174639 -0.25040844 -0.001369727 -0.01284553
0.043104668
## traveltime -0.100909119  0.09223875 -0.016807986 -0.01702494
0.028539674
## studytime  1.000000000 -0.17356303  0.039730704 -0.14319841
-0.063903675
## failures  -0.173563031  1.00000000 -0.044336626  0.09198747
0.124560922
```

## famrel 0.064568411	0.039730704	-0.04433663	1.000000000	0.15070144
## freetime 0.285018715	-0.143198407	0.09198747	0.150701444	1.000000000
## goout 1.000000000	-0.063903675	0.12456092	0.064568411	0.28501871
## Dalc 0.266993848	-0.196019263	0.13604693	-0.077594357	0.20900085
## Walc 0.420385745	-0.253784731	0.14196203	-0.113397308	0.14782181
## health -0.009577254	-0.075615863	0.06582728	0.094055728	0.07573336
## absences 0.044302220	-0.062700175	0.06372583	-0.044354095	-0.05807792
## G1 -0.149103967	0.160611915	-0.35471761	0.022168316	0.01261293
## G2 -0.162250034	0.135879999	-0.35589563	-0.018281347	-0.01377714
## G3 -0.132791474	0.097819690	-0.36041494	0.051363429	0.01130724
##	Dalc	Walc	health	absences
G1				
## age -0.06408150	0.131124605	0.11727605	-0.062187369	0.17523008
## Medu 0.20534100	0.019834099	-0.04712346	-0.046877829	0.10028482
## Fedu 0.19026994	0.002386429	-0.01263102	0.014741537	0.02447289
## traveltime -0.09303999	0.138325309	0.13411575	0.007500606	-0.01294378
## studytime 0.16061192	-0.196019263	-0.25378473	-0.075615863	-0.06270018
## failures -0.35471761	0.136046931	0.14196203	0.065827282	0.06372583
## famrel 0.02216832	-0.077594357	-0.11339731	0.094055728	-0.04435409
## freetime 0.01261293	0.209000848	0.14782181	0.075733357	-0.05807792
## goout -0.14910397	0.266993848	0.42038575	-0.009577254	0.04430222
## Dalc -0.09415879	1.000000000	0.64754423	0.077179582	0.11190803
## Walc -0.12617921	0.647544230	1.000000000	0.092476317	0.13629110
## health -0.07317207	0.077179582	0.09247632	1.000000000	-0.02993671
## absences -0.03100290	0.111908026	0.13629110	-0.029936711	1.000000000
## G1 1.000000000	-0.094158792	-0.12617921	-0.073172073	-0.03100290

```
## G2          -0.064120183 -0.08492735 -0.097719866 -0.03177670
0.85211807
## G3          -0.054660041 -0.05193932 -0.061334605  0.03424732
0.80146793
##              G2          G3
## age          -0.14347405 -0.16157944
## Medu         0.21552717  0.21714750
## Fedu         0.16489339  0.15245694
## traveltime   -0.15319796 -0.11714205
## studytime    0.13588000  0.09781969
## failures     -0.35589563 -0.36041494
## famrel       -0.01828135  0.05136343
## freetime     -0.01377714  0.01130724
## goout        -0.16225003 -0.13279147
## Dalc         -0.06412018 -0.05466004
## Walc         -0.08492735 -0.05193932
## health       -0.09771987 -0.06133460
## absences     -0.03177670  0.03424732
## G1           0.85211807  0.80146793
## G2           1.00000000  0.90486799
## G3           0.90486799  1.00000000
```

```
print(corrplot(cor.data,method = 'color'))
```

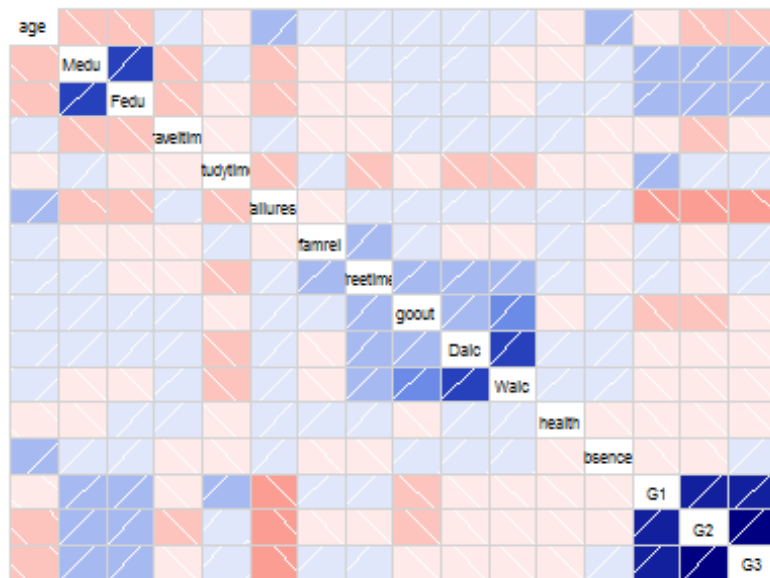


```
##              age          Medu          Fedu  traveltime
## age          1.000000000 -0.163658419 -0.163438069  0.070640721
```

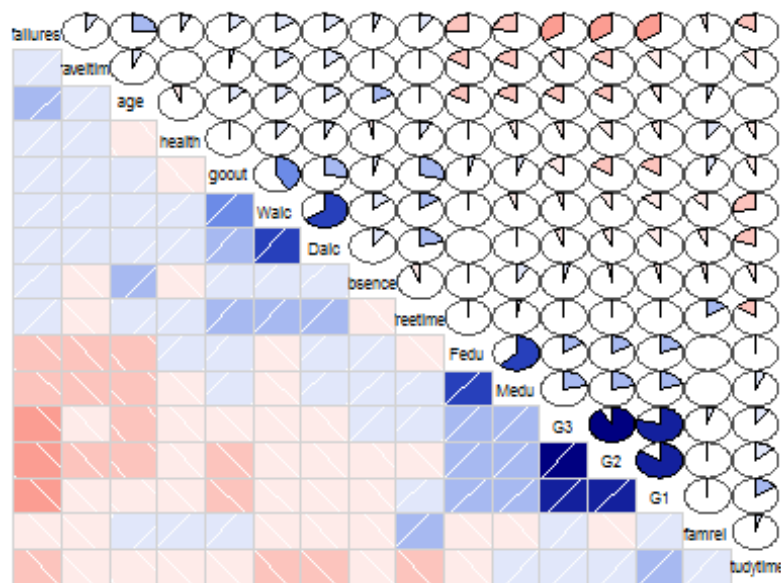
## Medu	-0.163658419	1.000000000	0.623455112	-0.171639305
## Fedu	-0.163438069	0.623455112	1.000000000	-0.158194054
## traveltime	0.070640721	-0.171639305	-0.158194054	1.000000000
## studytime	-0.004140037	0.064944137	-0.009174639	-0.100909119
## failures	0.243665377	-0.236679963	-0.250408444	0.092238746
## famrel	0.053940096	-0.003914458	-0.001369727	-0.016807986
## freetime	0.016434389	0.030890867	-0.012845528	-0.017024944
## goout	0.126963880	0.064094438	0.043104668	0.028539674
## Dalc	0.131124605	0.019834099	0.002386429	0.138325309
## Walc	0.117276052	-0.047123460	-0.012631018	0.134115752
## health	-0.062187369	-0.046877829	0.014741537	0.007500606
## absences	0.175230079	0.100284818	0.024472887	-0.012943775
## G1	-0.064081497	0.205340997	0.190269936	-0.093039992
## G2	-0.143474049	0.215527168	0.164893393	-0.153197963
## G3	-0.161579438	0.217147496	0.152456939	-0.117142053
##	studytime	failures	famrel	freetime
goout				
## age	-0.004140037	0.24366538	0.053940096	0.01643439
0.126963880				
## Medu	0.064944137	-0.23667996	-0.003914458	0.03089087
0.064094438				
## Fedu	-0.009174639	-0.25040844	-0.001369727	-0.01284553
0.043104668				
## traveltime	-0.100909119	0.09223875	-0.016807986	-0.01702494
0.028539674				
## studytime	1.000000000	-0.17356303	0.039730704	-0.14319841
-0.063903675				
## failures	-0.173563031	1.000000000	-0.044336626	0.09198747
0.124560922				
## famrel	0.039730704	-0.04433663	1.000000000	0.15070144
0.064568411				
## freetime	-0.143198407	0.09198747	0.150701444	1.000000000
0.285018715				
## goout	-0.063903675	0.12456092	0.064568411	0.28501871
1.000000000				
## Dalc	-0.196019263	0.13604693	-0.077594357	0.20900085
0.266993848				
## Walc	-0.253784731	0.14196203	-0.113397308	0.14782181
0.420385745				
## health	-0.075615863	0.06582728	0.094055728	0.07573336
-0.009577254				
## absences	-0.062700175	0.06372583	-0.044354095	-0.05807792
0.044302220				
## G1	0.160611915	-0.35471761	0.022168316	0.01261293
-0.149103967				
## G2	0.135879999	-0.35589563	-0.018281347	-0.01377714
-0.162250034				
## G3	0.097819690	-0.36041494	0.051363429	0.01130724
-0.132791474				
##	Dalc	Walc	health	absences

G1				
## age	0.131124605	0.11727605	-0.062187369	0.17523008
-0.06408150				
## Medu	0.019834099	-0.04712346	-0.046877829	0.10028482
0.20534100				
## Fedu	0.002386429	-0.01263102	0.014741537	0.02447289
0.19026994				
## traveltime	0.138325309	0.13411575	0.007500606	-0.01294378
-0.09303999				
## studytime	-0.196019263	-0.25378473	-0.075615863	-0.06270018
0.16061192				
## failures	0.136046931	0.14196203	0.065827282	0.06372583
-0.35471761				
## famrel	-0.077594357	-0.11339731	0.094055728	-0.04435409
0.02216832				
## freetime	0.209000848	0.14782181	0.075733357	-0.05807792
0.01261293				
## goout	0.266993848	0.42038575	-0.009577254	0.04430222
-0.14910397				
## Dalc	1.000000000	0.64754423	0.077179582	0.11190803
-0.09415879				
## Walc	0.647544230	1.000000000	0.092476317	0.13629110
-0.12617921				
## health	0.077179582	0.09247632	1.000000000	-0.02993671
-0.07317207				
## absences	0.111908026	0.13629110	-0.029936711	1.000000000
-0.03100290				
## G1	-0.094158792	-0.12617921	-0.073172073	-0.03100290
1.000000000				
## G2	-0.064120183	-0.08492735	-0.097719866	-0.03177670
0.85211807				
## G3	-0.054660041	-0.05193932	-0.061334605	0.03424732
0.80146793				
##	G2	G3		
## age	-0.14347405	-0.16157944		
## Medu	0.21552717	0.21714750		
## Fedu	0.16489339	0.15245694		
## traveltime	-0.15319796	-0.11714205		
## studytime	0.13588000	0.09781969		
## failures	-0.35589563	-0.36041494		
## famrel	-0.01828135	0.05136343		
## freetime	-0.01377714	0.01130724		
## goout	-0.16225003	-0.13279147		
## Dalc	-0.06412018	-0.05466004		
## Walc	-0.08492735	-0.05193932		
## health	-0.09771987	-0.06133460		
## absences	-0.03177670	0.03424732		
## G1	0.85211807	0.80146793		
## G2	1.000000000	0.90486799		
## G3	0.90486799	1.000000000		

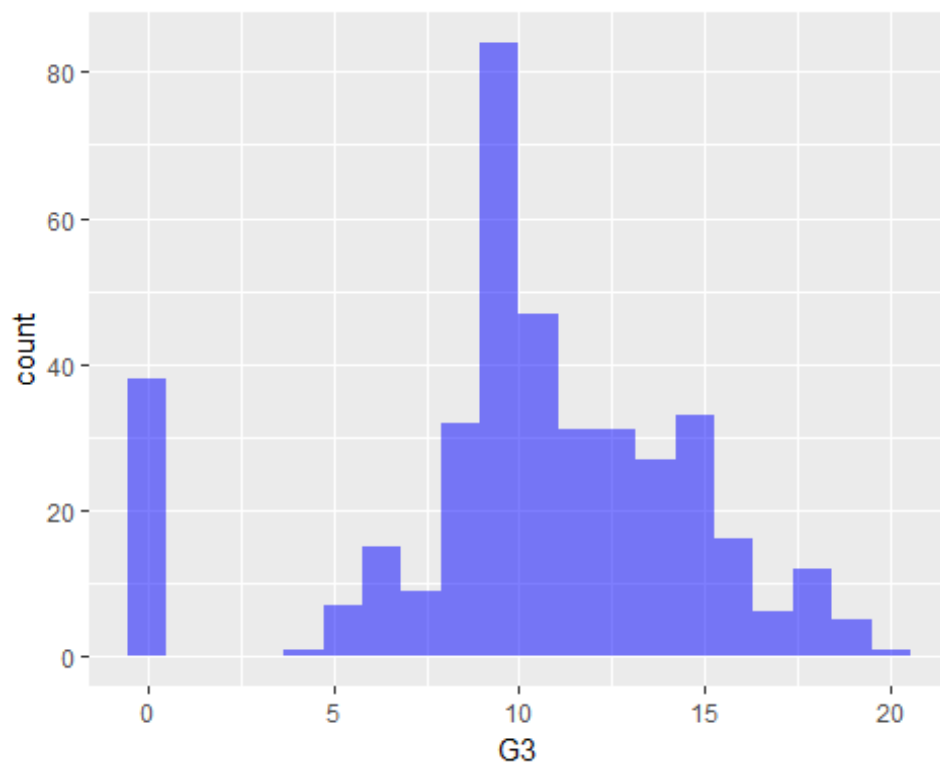
```
corrgram(df)
```



```
corrgram(df, order=TRUE, lower.panel=panel.shade,  
         upper.panel=panel.pie, text.panel=panel.txt)
```



```
ggplot(df, aes(x=G3)) + geom_histogram(bins=20, alpha=0.5, fill='blue')
```



Part 2

```
library(caTools)
```

Set a Seed

```
set.seed(101)
```

Split up Sample

```
sample <- sample.split(df$G3, SplitRatio = 0.7)
```

70% of data -> train

```
train <- subset(df, sample == TRUE)
```

30% will be test

```
test <- subset(df, sample == FALSE)
```

Train and Build Model

```
model <- lm(G3 ~ ., data = train)
```

Run Model

Interpret Model

```
summary(model)
```

```
##
## Call:
## lm(formula = G3 ~ ., data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4250 -0.6478  0.2844  1.0442  4.9840
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.70763    2.69488   1.376  0.17019
## schoolMS       0.66981    0.47436   1.412  0.15926
## sexM           0.25730    0.29257   0.879  0.38006
## age          -0.36163    0.12949  -2.793  0.00566 **
## addressU       0.08123    0.35652   0.228  0.81996
## famsizeLE3     0.12222    0.28709   0.426  0.67070
## PstatusT       0.06807    0.43032   0.158  0.87444
```

```

## Medu          0.11100      0.18757      0.592      0.55455
## Fedu          -0.16373      0.15928     -1.028      0.30503
## Mjobhealth    -0.63993      0.65314     -0.980      0.32820
## Mjobother     -0.15730      0.42323     -0.372      0.71048
## Mjobservices  -0.15872      0.46682     -0.340      0.73415
## Mjobteacher   -0.04930      0.62335     -0.079      0.93702
## Fjobhealth     0.17565      0.83034      0.212      0.83265
## Fjobother     -0.29559      0.56012     -0.528      0.59818
## Fjobservices  -0.76964      0.59476     -1.294      0.19692
## Fjobteacher   -0.27009      0.73824     -0.366      0.71480
## reasonhome    -0.41126      0.31857     -1.291      0.19799
## reasonother    0.06767      0.45323      0.149      0.88144
## reasonreputation 0.13478      0.34735      0.388      0.69834
## guardianmother -0.05442      0.31663     -0.172      0.86369
## guardianother  0.01588      0.58375      0.027      0.97832
## traveltime    -0.02353      0.19540     -0.120      0.90427
## studytime     -0.04294      0.16910     -0.254      0.79979
## failures      -0.17219      0.19668     -0.875      0.38220
## schoolsupyes   0.20742      0.42358      0.490      0.62481
## famsupyes     -0.05329      0.27753     -0.192      0.84789
## paidyes       0.31311      0.28284      1.107      0.26941
## activitiesyes  -0.26104      0.26687     -0.978      0.32901
## nurseryyes    -0.05345      0.31236     -0.171      0.86428
## higheryes     -0.94298      0.74005     -1.274      0.20385
## internetyes   -0.15834      0.37029     -0.428      0.66932
## romanticyes   -0.30048      0.28115     -1.069      0.28627
## famrel        0.36601      0.14609      2.505      0.01291 *
## freetime      0.08386      0.14247      0.589      0.55668
## goout         -0.12457      0.13306     -0.936      0.35015
## Dalc          -0.16995      0.20659     -0.823      0.41153
## Walc          0.21053      0.14963      1.407      0.16074
## health        0.07805      0.09341      0.836      0.40426
## absences      0.09547      0.02382      4.008      8.24e-05 ***
## G1            0.14259      0.07892      1.807      0.07206 .
## G2            0.98859      0.06929     14.267      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.962 on 235 degrees of freedom
## Multiple R-squared:  0.8456, Adjusted R-squared:  0.8187
## F-statistic: 31.39 on 41 and 235 DF,  p-value: < 2.2e-16

res <- residuals(model)
class(res)

## [1] "numeric"

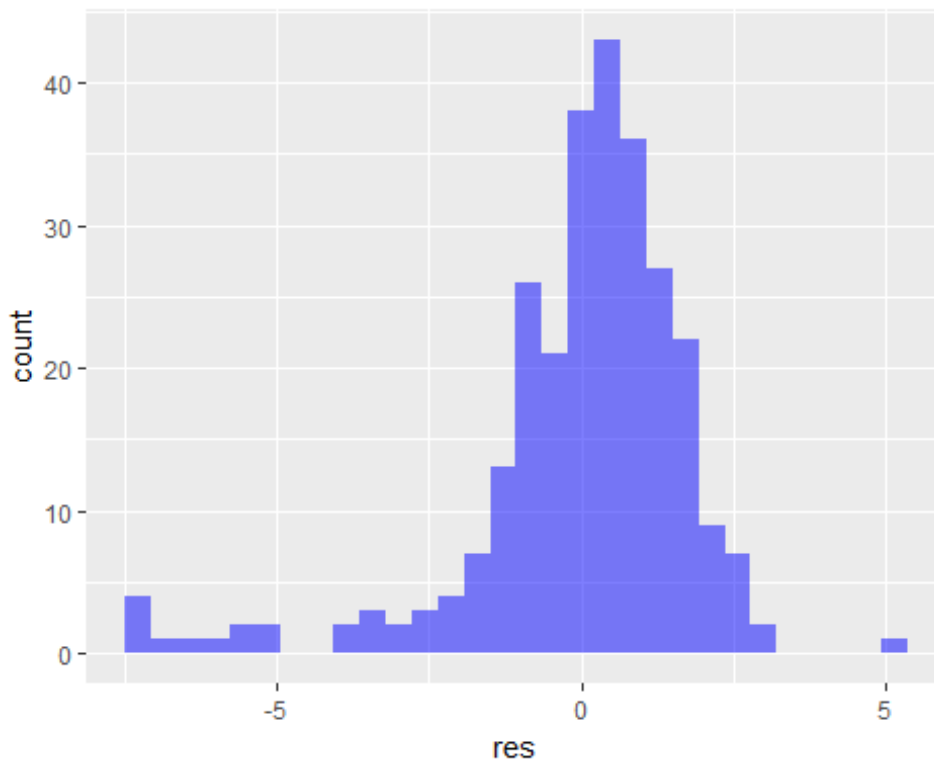
res <- as.data.frame(res)
head(res)

```

```
##          res
## 1  1.4684389
## 2  1.8826707
## 3  1.1866990
## 6 -2.2440152
## 9  0.5974865
## 11 0.8583062
```

```
ggplot(res, aes(res)) + geom_histogram(fill='blue',alpha=0.5)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Part 3

Predictions

```
G3.predictions <- predict(model,test)

results <- cbind(G3.predictions,test$G3)
colnames(results) <- c('predicted','actual')
results <- as.data.frame(results)
head(results)

##    predicted actual
## 4  12.682507     15
## 5   9.433677     10
```

```
## 7 11.312310 11
## 8 3.101530 6
## 10 15.564674 15
## 13 14.190360 14
```

Take Care of Neg Values

```
to_zero <- function(x){
  if(x < 0){
    return(0)
  }else{
    return(x)
  }
}
```

Apply Zero Function

```
results$predicted <- sapply(results$predicted, to_zero)
```

Mean Squared Error (MSE)

```
mse <- mean((results$actual - results$predicted)^2)
print('MSE')

## [1] "MSE"

print(mse)

## [1] 3.991675
```

RMSE

```
print(mse^0.5)

## [1] 1.997918
```

```
SSE <- sum((results$predicted - results$actual)^2)
SST <- sum((mean(df$G3) - results$actual)^2)

R2 <- 1 - SSE/SST
print('R2')

## [1] "R2"

print(R2)

## [1] 0.8044477
```