P7\_Data\_Analyst

Kyle Gibson

22/09/2022

library(readxl) # Reading in packages  
library(plotrix)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## 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(ggplot2)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(gridBase)  
library(gridGraphics)

## Loading required package: grid

getwd() # Checking working directory

## [1] "C:/AUT/Graphs R us"

d1 <- read\_excel("C:/P7/Manual\_field\_strata\_data.xlsx") # Importing raw data   
  
d2 <- read\_excel("C:/P7/Strata\_3Dmodel\_clast\_data.xlsx")  
  
d3 <- read\_excel("C:/P7/3D\_data\_strata\_2.xlsx")  
  
d4 <- read\_excel("C:/P7/clast\_data\_1,2,3.xlsx")   
  
head(d1) # Viewing first 6 rows of data

## # A tibble: 6 x 4  
## Strata ClastLength `Clast Width` ClastSize  
## <chr> <dbl> <dbl> <dbl>  
## 1 s2 18 4 72  
## 2 s2 12 5 60  
## 3 s2 10 5 50  
## 4 s2 8 3 24  
## 5 s2 7 4 28  
## 6 s2 6 2 12

View(d1) # Viewing entire dataset  
summary(d2) # Viewing summary statistics

## Strata clastlength clastwidth clastsize   
## Length:60 Min. : 3.00 Min. : 2.000 Min. : 6.00   
## Class :character 1st Qu.: 7.00 1st Qu.: 5.000 1st Qu.: 36.00   
## Mode :character Median : 9.00 Median : 7.000 Median : 61.50   
## Mean :10.48 Mean : 7.483 Mean : 95.47   
## 3rd Qu.:12.25 3rd Qu.: 8.000 3rd Qu.: 94.25   
## Max. :27.00 Max. :20.000 Max. :440.00

summary(d1)

## Strata ClastLength Clast Width ClastSize   
## Length:60 Min. : 2.00 Min. : 1.000 Min. : 2.0   
## Class :character 1st Qu.: 7.00 1st Qu.: 4.000 1st Qu.: 29.5   
## Mode :character Median :14.00 Median : 8.000 Median : 94.5   
## Mean :15.60 Mean : 9.267 Mean : 199.9   
## 3rd Qu.:19.25 3rd Qu.:13.000 3rd Qu.: 230.0   
## Max. :90.00 Max. :35.000 Max. :1575.0

# d1 = field data  
  
  
##### Using ggplot to assign plots to variables   
  
p1 <- qplot(d1$ClastLength, d1$`Clast Width`, col=d1$Strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)", ylim = c(0,40), xlim = c(0,80))+labs(colour="Strata")+ annotate(geom="text", x=40, y=40, label="A",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 1)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")   
  
   
  
  
p2 <- qplot(d1$Strata, d1$ClastSize, col=d1$Strata, xlab = "Strata", ylab = "Clast size (mm\u00B2)", ylim = c(0,1600))+labs(colour="Strata")+ annotate(geom="text", x=3.5, y=1600, label="B",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 1)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
# D2 = 3D model data  
  
  
p3 <- qplot(d2$Strata, d2$clastsize, col=d2$Strata, xlab = "Strata", ylab = "Clast size (mm\u00B2)", ylim = c(0,1600))+labs(colour="Strata")+ annotate(geom="text", x=3.5, y=1600, label="D",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
p4 <- qplot(d2$clastlength, d2$clastwidth, col=d2$Strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)", ylim = c(0,40), xlim = c(0,80))+labs(colour="Strata")+ annotate(geom="text", x=40, y=40, label="C",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
##### Creating plots with a best fit line   
  
bfd1 <- qplot(d1$ClastLength, d1$`Clast Width`, xlab = "Clast length (mm)", ylab = "Clast width (mm)", ylim = c(0,40), xlim = c(0,80))+labs(colour="Strata")+ annotate(geom="text", x=25, y=40, label=expression("A R-squared: 0.4195 P-value: 1.343"^"-8"), color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")  
  
  
bfd2 <- qplot(d3$clastlength, d3$clastwidth, xlab = "Clast length (mm)", ylab = "Clast width (mm)", ylim = c(0,40), xlim = c(0,80))+ annotate(geom="text", x=25, y=40, label=expression("B R-squared: 0.4519 P-value: 2.459"^"-9"),  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")  
  
  
grid.arrange(bfd1, bfd2, nrow = 2) # Viewing plots in plotting window

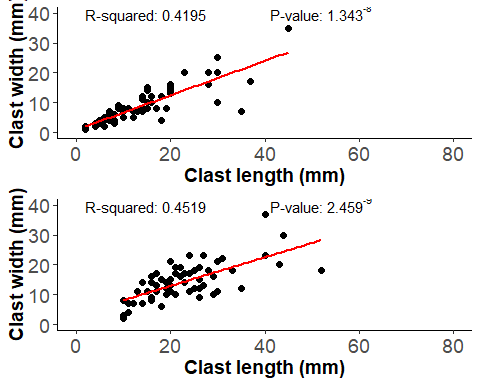
## Warning: Removed 1 rows containing non-finite values (stat\_smooth).

## Warning: Removed 1 rows containing missing values (geom\_point).

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'

## Warning: Removed 1 rows containing missing values (geom\_point).

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'



summary(lm(d1$ClastLength ~ d1$`Clast Width`)) # r-squared 0.4195 p-value 1.343e-08 # Checking values to ensure accuracy

##   
## Call:  
## lm(formula = d1$ClastLength ~ d1$`Clast Width`)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.461 -3.664 -2.007 0.279 66.539   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.8936 2.3325 1.241 0.22   
## d1$`Clast Width` 1.3712 0.2076 6.606 1.34e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.22 on 58 degrees of freedom  
## Multiple R-squared: 0.4293, Adjusted R-squared: 0.4195   
## F-statistic: 43.63 on 1 and 58 DF, p-value: 1.343e-08

summary(lm(d3$clastlength ~ d3$clastwidth)) # r-squared 0.4519 p-value 2.459^-9

##   
## Call:  
## lm(formula = d3$clastlength ~ d3$clastwidth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.264 -4.584 -1.504 3.136 25.616   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.1044 2.1199 4.295 6.75e-05 \*\*\*  
## d3$clastwidth 0.9600 0.1363 7.045 2.46e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.551 on 58 degrees of freedom  
## Multiple R-squared: 0.4612, Adjusted R-squared: 0.4519   
## F-statistic: 49.64 on 1 and 58 DF, p-value: 2.459e-09

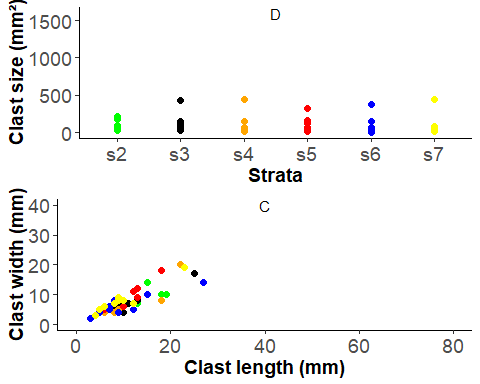
summary(d1)

## Strata ClastLength Clast Width ClastSize   
## Length:60 Min. : 2.00 Min. : 1.000 Min. : 2.0   
## Class :character 1st Qu.: 7.00 1st Qu.: 4.000 1st Qu.: 29.5   
## Mode :character Median :14.00 Median : 8.000 Median : 94.5   
## Mean :15.60 Mean : 9.267 Mean : 199.9   
## 3rd Qu.:19.25 3rd Qu.:13.000 3rd Qu.: 230.0   
## Max. :90.00 Max. :35.000 Max. :1575.0

summary(d3) # Viewing summary statistics

## strata clastlength clastwidth clastsize   
## Length:60 Min. :10.0 Min. : 2.00 Min. : 20.0   
## Class :character 1st Qu.:17.0 1st Qu.:10.75 1st Qu.: 194.5   
## Mode :character Median :21.0 Median :14.00 Median : 295.0   
## Mean :22.8 Mean :14.27 Mean : 362.3   
## 3rd Qu.:27.0 3rd Qu.:18.00 3rd Qu.: 427.5   
## Max. :52.0 Max. :37.00 Max. :1480.0

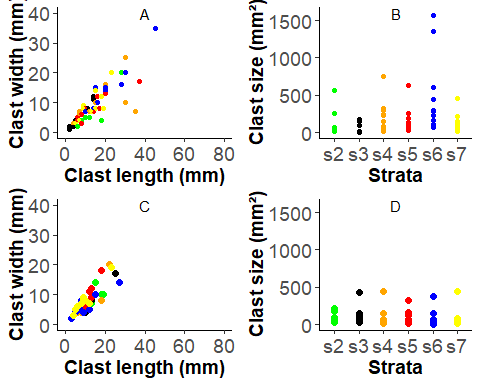
grid.arrange(p3, p4, nrow = 2) # Viewing plots in plotting window



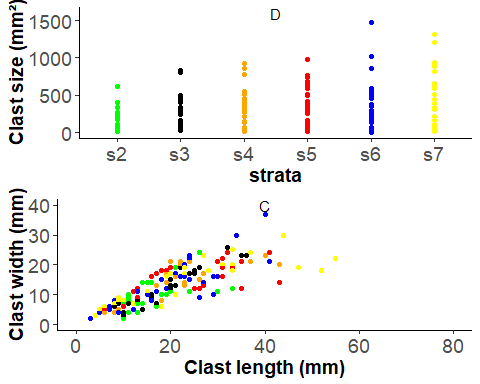
grid.arrange(p1, p2, p4, p3, nrow = 2)

## Warning: Removed 1 rows containing missing values (geom\_point).

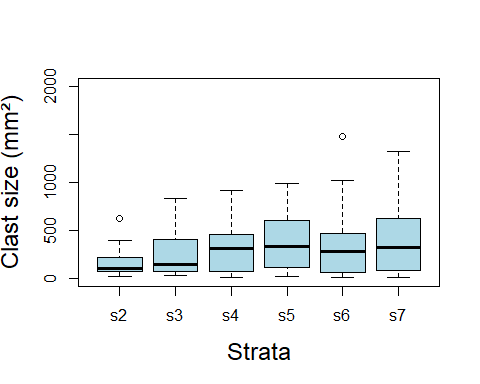
## Warning: Removed 1 rows containing missing values (geom\_point).



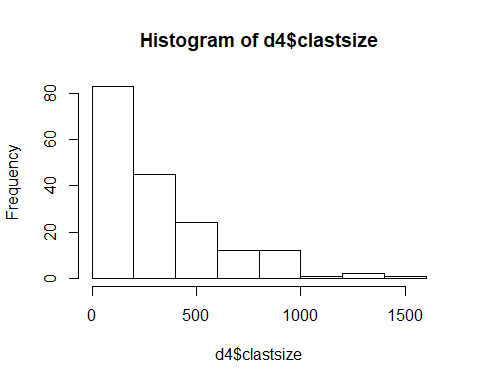
##### Creating plots using ggplot and assigning them to variables  
  
p5 <- qplot(d4$Strata, d4$clastsize, col=d4$Strata, xlab = "strata", ylab = "Clast size (mm\u00B2)", ylim = c(0,1600))+labs(colour="Strata")+ annotate(geom="text", x=3.5, y=1600, label="D",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 1)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
p6 <- qplot(d4$clastlength, d4$clastwidth, col=d4$Strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)", ylim = c(0,40), xlim = c(0,80))+labs(colour="Strata")+ annotate(geom="text", x=40, y=40, label="C",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 1)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
grid.arrange(p5, p6, nrow = 2) # Viewing plots



boxplot(d4$clastsize ~ d4$Strata, ylim = c(0,2000), col = "light blue", ylab = "Clast size (mm\u00B2)", xlab = "Strata", cex.lab = 1.5)

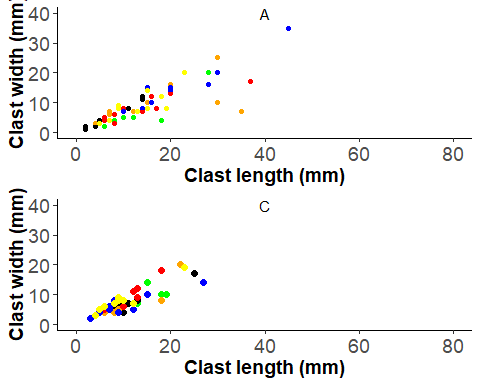


hist(d4$clastsize) # Quick plots for visual exploration of data

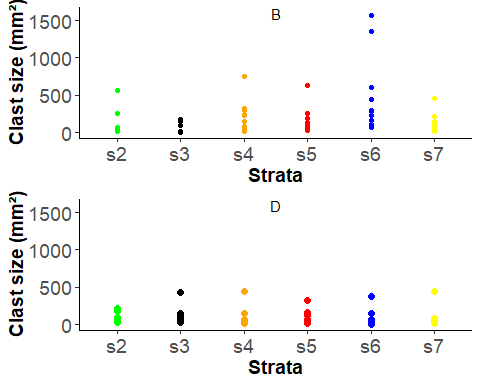


library("gridExtra", lib.loc="~/R/win-library/3.5") # Reading package  
grid.arrange(p1, p4, nrow = 2)

## Warning: Removed 1 rows containing missing values (geom\_point).  
  
## Warning: Removed 1 rows containing missing values (geom\_point).

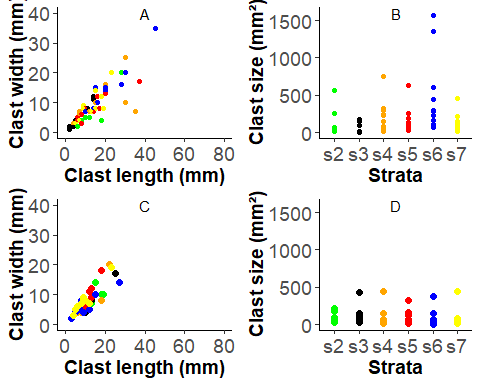


grid.arrange(p2, p3, nrow = 2) # Arranging plots



p5 <- grid.arrange(p1, p2, p4, p3, nrow = 2) # Viewing plots

## Warning: Removed 1 rows containing missing values (geom\_point).  
  
## Warning: Removed 1 rows containing missing values (geom\_point).



##### Viewing summary statistics of data for each stratum  
  
summary(d4)

## Strata clastlength clastwidth clastsize   
## Length:180 Min. : 3.00 Min. : 2.00 Min. : 6.0   
## Class :character 1st Qu.:10.75 1st Qu.: 7.00 1st Qu.: 77.0   
## Mode :character Median :18.50 Median :12.00 Median : 231.0   
## Mean :19.76 Mean :12.79 Mean : 309.5   
## 3rd Qu.:26.00 3rd Qu.:18.00 3rd Qu.: 442.5   
## Max. :55.00 Max. :37.00 Max. :1480.0

summary(d1$ClastSize[d1$Strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.0 25.0 41.0 111.3 69.0 560.0

summary(d1$ClastSize[d1$Strata=="s3"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.0 9.0 17.5 62.1 134.5 168.0

summary(d1$ClastSize[d1$Strata=="s4"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.00 57.75 187.50 217.70 286.25 750.00

summary(d1$ClastSize[d1$Strata=="s5"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 24.0 34.5 89.0 152.1 178.0 629.0

summary(d1$ClastSize[d1$Strata=="s6"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 70.0 176.2 290.0 512.0 562.0 1575.0

summary(d1$ClastSize[d1$Strata=="s7"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 15.00 74.25 105.50 144.50 195.50 460.00

summary(d4$clastsize[d4$Strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20.0 70.0 106.0 168.3 217.5 624.0

summary(d4$clastsize[d4$Strata=="s3"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 30.0 77.0 147.0 248.9 391.0 832.0

summary(d4$clastsize[d4$Strata=="s4"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.0 86.0 310.5 328.8 454.2 920.0

summary(d4$clastsize[d4$Strata=="s5"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20.0 120.8 333.0 367.5 598.8 984.0

summary(d4$clastsize[d4$Strata=="s6"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6.0 80.0 280.0 339.6 463.0 1480.0

summary(d4$clastsize[d4$Strata=="s7"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.00 81.75 322.00 404.13 620.75 1320.00

##### Calling mean, median and summary statistics  
  
mean(d1$ClastLength)

## [1] 15.6

mean(d2$clastlength)

## [1] 10.48333

mean(d1$`Clast Width`)

## [1] 9.266667

mean(d2$clastwidth)

## [1] 7.483333

mean(d1$ClastSize)

## [1] 199.95

mean(d2$clastsize)

## [1] 95.46667

median(d1$ClastSize)

## [1] 94.5

median(d2$clastsize)

## [1] 61.5

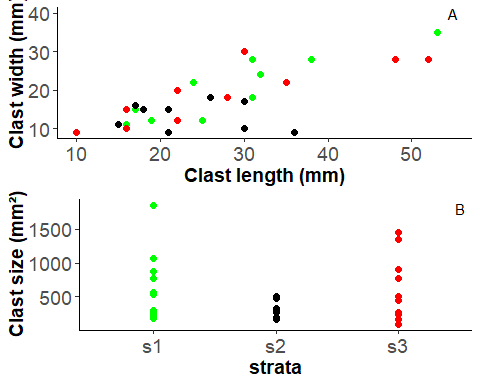
summary(d1)

## Strata ClastLength Clast Width ClastSize   
## Length:60 Min. : 2.00 Min. : 1.000 Min. : 2.0   
## Class :character 1st Qu.: 7.00 1st Qu.: 4.000 1st Qu.: 29.5   
## Mode :character Median :14.00 Median : 8.000 Median : 94.5   
## Mean :15.60 Mean : 9.267 Mean : 199.9   
## 3rd Qu.:19.25 3rd Qu.:13.000 3rd Qu.: 230.0   
## Max. :90.00 Max. :35.000 Max. :1575.0

summary(d2)

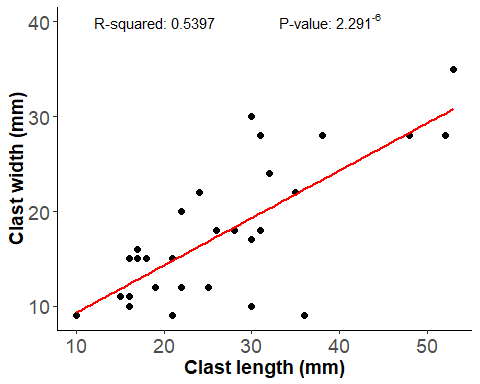
## Strata clastlength clastwidth clastsize   
## Length:60 Min. : 3.00 Min. : 2.000 Min. : 6.00   
## Class :character 1st Qu.: 7.00 1st Qu.: 5.000 1st Qu.: 36.00   
## Mode :character Median : 9.00 Median : 7.000 Median : 61.50   
## Mean :10.48 Mean : 7.483 Mean : 95.47   
## 3rd Qu.:12.25 3rd Qu.: 8.000 3rd Qu.: 94.25   
## Max. :27.00 Max. :20.000 Max. :440.00

d5 <- read\_excel("C:/P7/Longrock\_clast\_sizes.xlsx") # Importing raw data  
  
##### Assigning plots to variables using ggplot  
  
p7 <- qplot(d5$strata, d5$clastsize, col=d5$strata, xlab = "strata", ylab = "Clast size (mm\u00B2)")+labs(colour="Strata")+ annotate(geom="text", x=3.5, y=1800, label="B",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "red"))+  
 geom\_point(size = 2)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
p8 <- qplot(d5$clastlength, d5$clastwidth, col=d5$strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=55, y=40, label="A",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "red"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
grid.arrange(p8, p7, nrow = 2) # Viewing plots



##### Assigning plot   
  
bfd5 <- qplot(d5$clastlength, d5$clastwidth, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=25, y=40, label=expression(" R-squared: 0.5397 P-value: 2.291"^"-6"), color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")  
  
plot(bfd5) # Viewing plot

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'



summary(lm(d5$clastlength~d5$clastwidth)) # Viewing summary statistics

##   
## Call:  
## lm(formula = d5$clastlength ~ d5$clastwidth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.513 -6.608 -1.358 3.894 18.804   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.2029 3.5981 2.002 0.0551 .   
## d5$clastwidth 1.1103 0.1877 5.917 2.29e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.323 on 28 degrees of freedom  
## Multiple R-squared: 0.5556, Adjusted R-squared: 0.5397   
## F-statistic: 35.01 on 1 and 28 DF, p-value: 2.291e-06

summary(d5)

## strata clastlength clastwidth clastsize   
## Length:30 Min. :10.00 Min. : 9.0 Min. : 90.0   
## Class :character 1st Qu.:18.25 1st Qu.:12.0 1st Qu.: 257.2   
## Mode :character Median :25.50 Median :16.5 Median : 382.0   
## Mean :26.97 Mean :17.8 Mean : 536.4   
## 3rd Qu.:31.00 3rd Qu.:22.0 3rd Qu.: 715.5   
## Max. :53.00 Max. :35.0 Max. :1855.0

summary(d5$clastsize[d5$strata=="s1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 176.0 266.2 543.0 660.0 843.0 1855.0

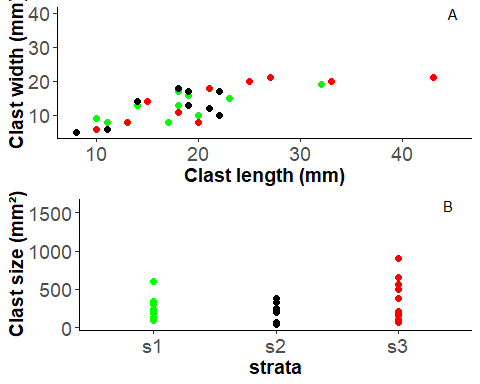
summary(d5$clastsize[d5$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 165.0 270.5 307.5 332.3 432.0 510.0

summary(d5$clastsize[d5$strata=="s3"])

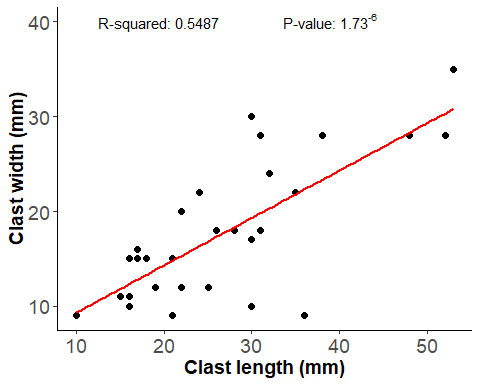
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 90.0 246.0 472.0 616.8 867.5 1456.0

d6 <- read\_excel("C:/P7/Smallrock\_clast\_sizes.xlsx") # Importing raw data  
  
##### Assigning plots  
  
p9 <- qplot(d6$strata, d6$clastsize, col=d6$strata, xlab = "strata", ylab = "Clast size (mm\u00B2)")+labs(colour="Strata")+ annotate(geom="text", x=3.4, y=1600, label="B",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "red"))+  
 geom\_point(size = 2)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
p10 <- qplot(d6$clastlength, d6$clastwidth, col=d6$strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=45, y=40, label="A",  
 color="black")+ scale\_color\_manual(values = c("green", "black", "red"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
grid.arrange(p10, p9, nrow = 2) # Viewing plots



##### Assigning plot  
  
bfd6 <- qplot(d5$clastlength, d5$clastwidth, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=25, y=40, label=expression(" R-squared: 0.5487 P-value: 1.73"^"-6"), color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")  
  
plot(bfd6) # Viewing plot

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'



summary(lm(d6$clastlength~d6$clastwidth)) # Viewing summary statistics

##   
## Call:  
## lm(formula = d6$clastlength ~ d6$clastwidth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.5404 -4.0372 -0.8918 2.9898 15.0138   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.8654 2.7257 1.418 0.167   
## d6$clastwidth 1.1486 0.1907 6.022 1.73e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.994 on 28 degrees of freedom  
## Multiple R-squared: 0.5643, Adjusted R-squared: 0.5487   
## F-statistic: 36.26 on 1 and 28 DF, p-value: 1.727e-06

summary(d6$clastsize[d6$strata=="s1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 88.0 147.5 217.0 249.3 305.5 608.0

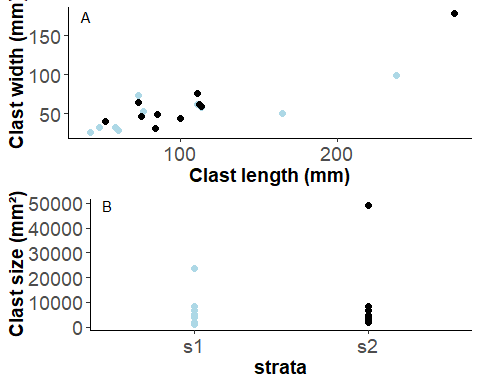
summary(d6$clastsize[d6$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 40.0 202.0 249.5 236.5 323.0 374.0

summary(d6$clastsize[d6$strata=="s3"])

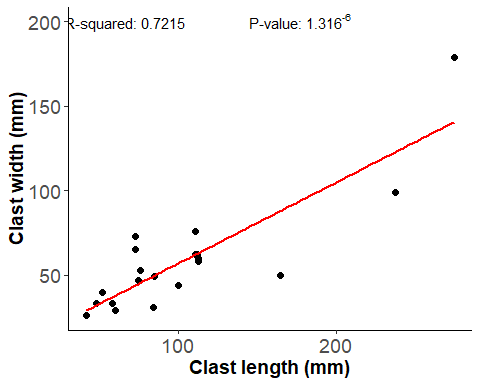
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 60.0 169.5 294.0 374.0 550.2 903.0

d7 <- read\_excel("C:/P7/Large\_model\_clast\_sizes.xlsx") # Importing raw data  
  
##### Assigning plots  
  
p11 <- qplot(d7$strata, d7$clastsize, col=d7$strata, xlab = "strata", ylab = "Clast size (mm\u00B2)")+labs(colour="Strata")+ annotate(geom="text", x=0.5, y=49000, label="B",  
 color="black")+ scale\_color\_manual(values = c("lightblue", "black"))+  
 geom\_point(size = 2)+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
  
  
p12 <- qplot(d7$clastlength, d7$clastwidth, col=d7$strata, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=40, y=175, label="A",  
 color="black")+ scale\_color\_manual(values = c("lightblue", "black"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")  
  
grid.arrange(p12, p11, nrow = 2) # Viewing plots



##### Assigning plot  
  
bfd1 <- qplot(d7$clastlength, d7$clastwidth, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+labs(colour="Strata")+ annotate(geom="text", x=100, y=200, label=expression(" R-squared: 0.7215 P-value: 1.316"^"-6"), color="black")+ scale\_color\_manual(values = c("green", "black", "orange", "red", "blue", "yellow"))+  
 geom\_point(size = 2)+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")  
  
plot(bfd1) # Viewing plot

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'



summary(lm(d7$clastlength~d7$clastwidth)) # Viewing summary statistics

##   
## Call:  
## lm(formula = d7$clastlength ~ d7$clastwidth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -52.555 -16.779 -4.889 8.187 74.783   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.3940 14.5386 0.921 0.369   
## d7$clastwidth 1.5365 0.2168 7.087 1.32e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31.87 on 18 degrees of freedom  
## Multiple R-squared: 0.7362, Adjusted R-squared: 0.7215   
## F-statistic: 50.22 on 1 and 18 DF, p-value: 1.316e-06

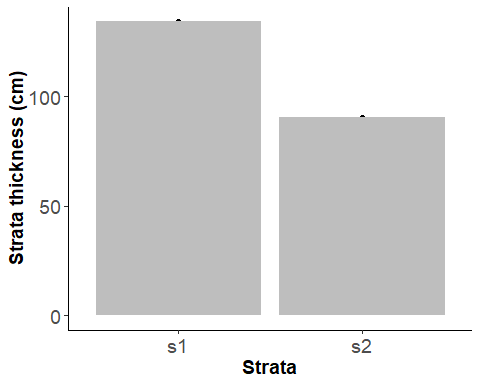
summary(d7$clastsize[d7$strata=="s1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1092 1784 4678 6094 6800 23562

summary(d7$clastsize[d7$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2080 3685 4572 9290 6903 49225

sw1 <- read\_excel("C:/P7/Outcrop\_1\_stratawidth.xlsx") # Importing raw data  
  
##### Viewing plots  
  
qplot(sw1$strata, sw1$width, ylab = "Strata thickness (cm)", xlab = "Strata")+ scale\_color\_manual(values = c("black", "lightblue", "darkgreen"))+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14))+geom\_point(size = 0)+labs(colour="Clast type")+geom\_bar(stat="identity", alpha = 1, fill = alpha(c("grey"), .3))+guides(color = guide\_legend(override.aes = list(alpha = 0.0)))



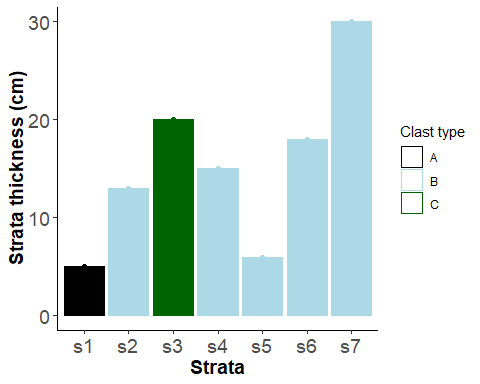
summary(sw1$width[sw1$strata=="s1"]) # Viewing summary statistics

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 134.1 134.1 134.1 134.1 134.1 134.1

summary(sw1$width[sw1$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 90.4 90.4 90.4 90.4 90.4 90.4

sw <- read\_excel("C:/P7/Strata\_width.xlsx") # Importing raw data  
  
qplot(sw$strata, sw$width, ylab = "Strata thickness (cm)", xlab = "Strata", col =sw$clasttype)+ labs(colour = "clasttype")+ scale\_color\_manual(values = c("black", "lightblue", "darkgreen"))+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14))+geom\_point(size = 0)+labs(colour="Clast type")+geom\_bar(stat="identity", alpha = 1, fill = alpha(c("black", "lightblue", "darkgreen", "lightblue", "lightblue", "lightblue", "lightblue"), .3))+guides(color = guide\_legend(override.aes = list(alpha = 0.0)))



summary(sw$width[sw$strata=="s1"]) # Viewing summary statistics

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5 5 5 5 5 5

summary(sw$width[sw$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 13 13 13 13 13 13

summary(sw$width[sw$strata=="s3"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 20 20 20 20 20 20

summary(sw$width[sw$strata=="s4"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 15 15 15 15 15 15

summary(sw$width[sw$strata=="s5"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 6 6 6 6 6 6

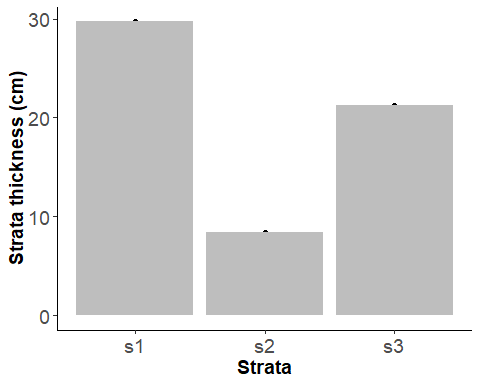
summary(sw$width[sw$strata=="s6"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 18 18 18 18 18 18

summary(sw$width[sw$strata=="s7"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 30 30 30 30 30 30

sw3 <- read\_excel("C:/P7/outcrop\_3\_stratawidth.xlsx") # Importing raw data  
  
##### Viewing plots  
  
qplot(sw3$strata, sw3$width, ylab = "Strata thickness (cm)", xlab = "Strata")+ scale\_color\_manual(values = c("black", "lightblue", "darkgreen"))+ theme\_bw()+ theme(panel.border = element\_blank(), panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14))+geom\_point(size = 0)+labs(colour="Clast type")+geom\_bar(stat="identity", alpha = 1, fill = alpha(c("grey"), .3))+guides(color = guide\_legend(override.aes = list(alpha = 0.0)))

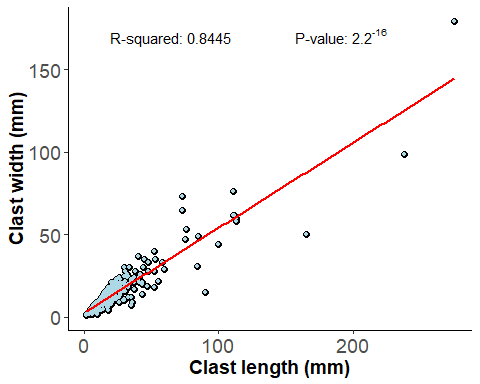


summary(sw3$width[sw3$strata=="s2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8.4 8.4 8.4 8.4 8.4 8.4

allclast <- read\_excel("C:/P7/All\_clast\_length\_width.xlsx") # Importing raw data  
  
##### Viewing plot  
  
qplot(allclast$clastlength, allclast$clastwidth, xlab = "Clast length (mm)", ylab = "Clast width (mm)")+ annotate(geom="text", x=100, y=170, label=expression(" R-squared: 0.8445 P-value: 2.2"^"-16"), color="black")+  
 geom\_point(size = 2, colour = "black")+geom\_point(size = 1, colour = "lightblue")+ theme\_bw() + theme(panel.border = element\_blank(), panel.grid.major = element\_blank(),  
panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "black"), axis.title=element\_text(size=14,face="bold"), axis.text.x = element\_text(size = 14), axis.text.y = element\_text(size = 14), legend.position="none")+ geom\_smooth(method = "lm", se = FALSE, col = "red")

## Warning in is.na(x): is.na() applied to non-(list or vector) of type  
## 'expression'



summary(lm(allclast$clastlength~allclast$clastwidth)) # Viewing summary statistics

##   
## Call:  
## lm(formula = allclast$clastlength ~ allclast$clastwidth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -46.038 -4.511 -0.789 2.739 83.656   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.60007 0.85818 -0.699 0.485   
## allclast$clastwidth 1.63888 0.03937 41.631 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.78 on 318 degrees of freedom  
## Multiple R-squared: 0.845, Adjusted R-squared: 0.8445   
## F-statistic: 1733 on 1 and 318 DF, p-value: < 2.2e-16

summary(allclast)

## clastlength clastwidth   
## Min. : 2.00 Min. : 1.00   
## 1st Qu.: 11.00 1st Qu.: 7.00   
## Median : 19.00 Median : 12.50   
## Mean : 24.83 Mean : 15.52   
## 3rd Qu.: 28.00 3rd Qu.: 18.25   
## Max. :275.00 Max. :179.00

mean(d7$clastsize) # d7 is outcrop one # Viewing mean values of particular variables

## [1] 7691.95

mean(d4$clastsize) #d4 is outcrop two

## [1] 309.5444

mean(d5$clastsize) # d5 is outcrop three

## [1] 536.3667

mean(d6$clastsize) # d6 is outcrop four

## [1] 286.6