Graphs

Steven Lam & Alexander Lee

December 6, 2015

Creates the charts for visualization of the data. Each chart shows the frequency of each grade (A+, A, A-, etc.) and lines indicating the average grade for that class. Classes are arranged by teacher.

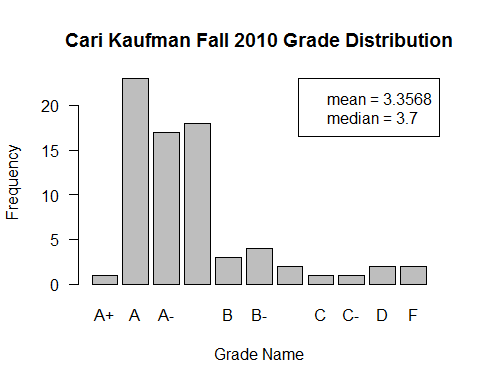
#re-import the data and whatnot so we can knit the .Rmd  
  
library(readr)  
setwd("../Data")  
  
stat2010f <- read\_csv("Grades2010f.csv")  
stat2011f <- read\_csv("Grades2011f.csv")  
stat2012f <- read\_csv("Grades2012f.csv")  
stat2013f <- read\_csv("Grades2013f.csv")  
stat2014f <- read\_csv("Grades2014f.csv")  
  
stat2010s <- read\_csv("Grades2010s.csv")  
stat2011s <- read\_csv("Grades2011s.csv")  
stat2012s <- read\_csv("Grades2012s.csv")  
stat2013s <- read\_csv("Grades2013s.csv")  
stat2014s <- read\_csv("Grades2014s.csv")  
stat2015s <- read\_csv("Grades2015s.csv")  
  
  
#fall 2012  
DeborahNolan2012f <- stat2012f[stat2012f$`Instructor Name` == "Deborah Nolan",]  
FletcherIbser2012f <- stat2012f[stat2012f$`Instructor Name` == "Fletcher Ibser",]  
  
#fall 2013  
Ingileif2013f <- stat2013f[stat2013f$`Instructor Name` == "Ingileif Hallgrimsdottir",]  
FletcherIbser2013f <- stat2013f[stat2013f$`Instructor Name` == "Fletcher Ibser",]  
  
#fall 2014  
BenoitDherin2014f <- stat2014f[stat2014f$`Instructor Name` == "Benoit Dherin",]  
FletcherIbser2014f <- stat2014f[stat2014f$`Instructor Name` == "Fletcher Ibser",]  
  
#spring 2012  
DeborahNolan2012s <- stat2012s[stat2012s$`Instructor Name` == "Deborah Nolan",]  
FletcherIbser2012s <- stat2012s[stat2012s$`Instructor Name` == "Fletcher Ibser",]  
  
#spring 2014  
BenoitDherin2014s <- stat2014s[stat2014s$`Instructor Name` == "Benoit Dherin",]  
CariKaufman2014s <- stat2014s[stat2014s$`Instructor Name` == "Cari Kaufman",]  
  
#spring 2015  
DeborahNolan2015s <- stat2015s[stat2015s$`Instructor Name` == "Deborah Nolan",]  
Ingileif2015s <- stat2015s[stat2015s$`Instructor Name` == "Ingileif Hallgrimsdottir",]  
  
  
#GPA calculations  
gpa <- c(12, 12, 11.1, 9.9, 9, 8.1, 6.9, 6, 5.1, 3.9, 3, 2.1, 0)  
grade\_names <- c("A+", "A", "A-", "B+", "B", "B-",   
 "C+", "C", "C-", "D+", "D", "D-", "F")  
names(gpa) <- grade\_names  
gpa <- gpa / 3  
  
  
#loopdy loop - find the total grade distribution for all classes  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
grades <- list(stat2010f, stat2011f, stat2012f,   
 stat2013f, stat2014f, stat2010s,   
 stat2011s, stat2012s, stat2013s,   
 stat2014s, stat2015s)  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],   
 "Enrollment.Cnt"])  
 }  
   
}  
  
  
  
#create a vector with the grade relative frequencies for sampling purposes  
grade\_samples <- rep(grade\_names, times = total\_sum)

# Data Visualization

## Cari Kaufman

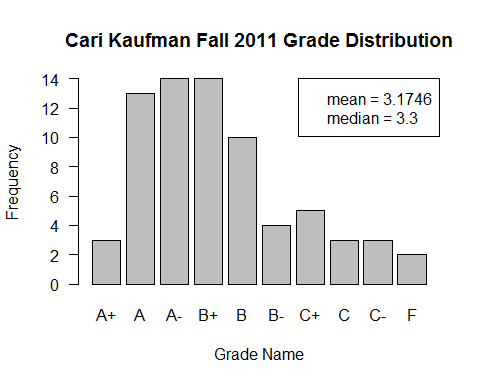
#### Cari Kaufman Fall 2010

#stat2010f  
  
  
barplot(height = stat2010f$`Enrollment Cnt`, names.arg = stat2010f$`Grade Nm`,  
 main = "Cari Kaufman Fall 2010 Grade Distribution",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(stat2010f$`Average Grade`, times = stat2010f$`Enrollment Cnt`)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



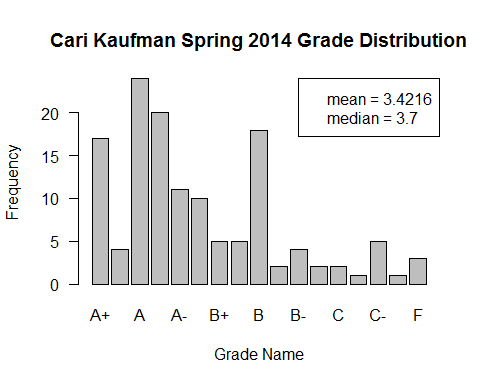
#### Cari Kaufman Fall 2011

#stat2011f  
  
barplot(height = stat2011f$`Enrollment Cnt`, names.arg = stat2011f$`Grade Nm`,  
 main = "Cari Kaufman Fall 2011 Grade Distribution",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(stat2011f$`Average Grade`, times = stat2011f$`Enrollment Cnt`)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



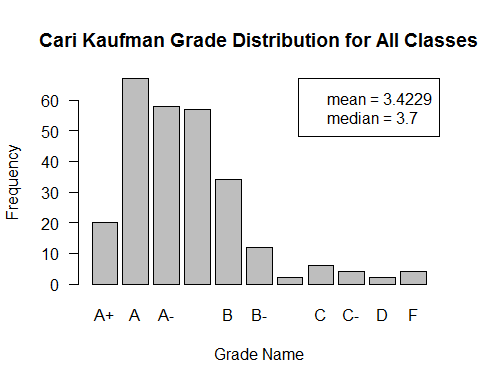
#### Cari Kaufman Spring 2014

#stat2014s  
  
barplot(height = stat2014s$`Enrollment Cnt`, names.arg = stat2014s$`Grade Nm`,  
 main = "Cari Kaufman Spring 2014 Grade Distribution",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(stat2014s$`Average Grade`, times = stat2014s$`Enrollment Cnt`)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))

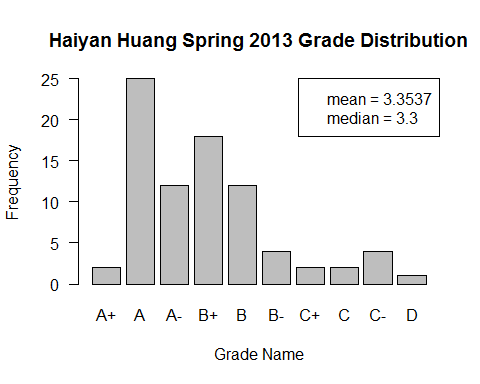


#### Cari Kaufman - Totals for All Years

#stat2010f  
#stat2011f  
#CariKaufman2014s  
  
  
#combine the grade data for all of  
#Cari Kaufman's classes  
  
grades <- list(DeborahNolan2012f, DeborahNolan2012s, DeborahNolan2015s)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Cari Kaufman Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))

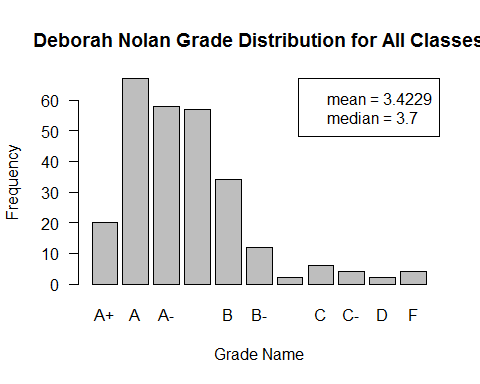


#2013s  
  
barplot(height = stat2013s$`Enrollment Cnt`, names.arg = stat2013s$`Grade Nm`,  
 main = "Haiyan Huang Spring 2013 Grade Distribution",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(stat2013s$`Average Grade`, times = stat2013s$`Enrollment Cnt`)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



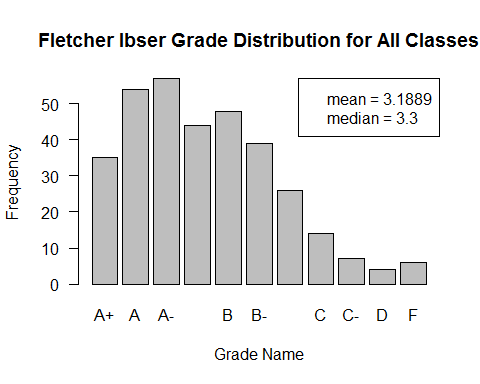
#### Deborah Nolan - All

grades <- list(DeborahNolan2012f, DeborahNolan2012s, DeborahNolan2015s)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Deborah Nolan Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



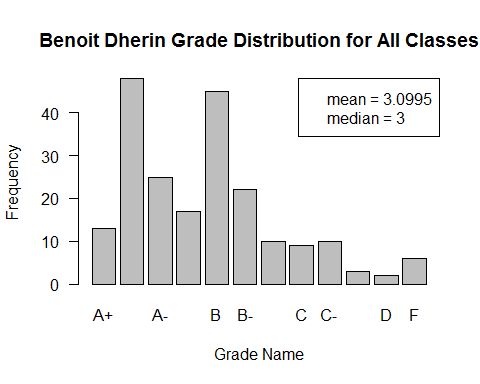
#### Fletcher Ibser - All

grades <- list(FletcherIbser2012f, FletcherIbser2012s,   
 FletcherIbser2013f, FletcherIbser2014f)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Fletcher Ibser Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



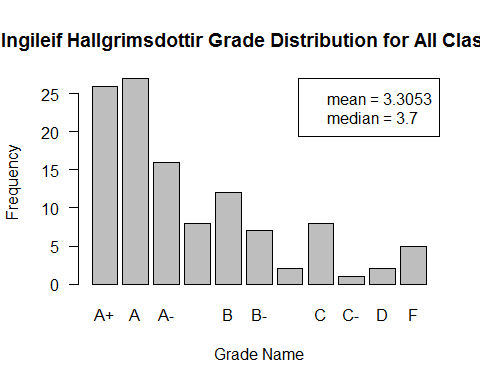
#### Benoit Dherin - All

grades <- list(stat2013f, BenoitDherin2014s, BenoitDherin2014f)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Benoit Dherin Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



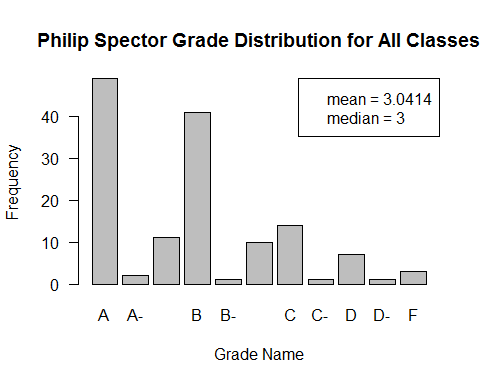
#### Ingileif Hallgrimsdottir - All

grades <- list(Ingileif2013f, Ingileif2015s)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Ingileif Hallgrimsdottir Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



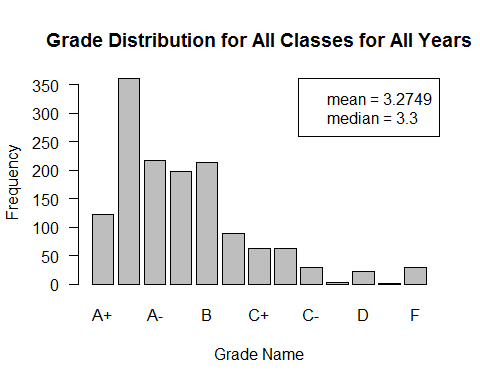
#### Philip Spector - All

grades <- list(stat2010s, stat2011s)   
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Philip Spector Grade Distribution for All Classes",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



#### All Classes

grades <- list(stat2010f, stat2011f, stat2012f,   
 stat2013f, stat2014f, stat2010s,   
 stat2011s, stat2012s, stat2013s,   
 stat2014s, stat2015s)  
  
total\_sum <- double(13)  
names(total\_sum) <- grade\_names  
  
for(i in 1:length(grades)){  
   
 #create a temporary variable for the data  
 temp\_data <- data.frame(grades[i])  
   
 for(j in 1:13){  
   
 total\_sum[j] <- total\_sum[j] + sum(temp\_data[temp\_data$Grade.Nm == grade\_names[j],  
 "Enrollment.Cnt"])  
   
 }  
   
}  
  
total\_sum <- total\_sum[total\_sum != 0]  
gpa\_values <- unname(gpa[intersect(names(gpa), names(total\_sum))])  
  
  
barplot(height = total\_sum, names.arg = names(total\_sum),  
 main = "Grade Distribution for All Classes for All Years",  
 xlab = "Grade Name", ylab = "Frequency", las = 1)  
  
#create the GPA distribution of the class based on its  
#`Enrollment Cnt`'s, used to find the mean/median  
grade\_samples <- rep(gpa\_values, times = total\_sum)  
  
#mean and median  
average\_grade <- mean(grade\_samples)  
median\_grade <- median(grade\_samples)  
  
#round the mean and median  
average\_grade <- round(average\_grade, 4)  
median\_grade <- round(median\_grade, 4)  
  
  
#create a legend that says what the lines mean  
legend("topright", legend = c(paste0("mean = ", as.character(average\_grade)),   
 paste0("median = ", as.character(median\_grade))))



total\_prop <- prop.table(total\_sum)  
sum(head(total\_prop, 3))

## [1] 0.497516

sum(total\_prop[4:6])

## [1] 0.3548616