Graphics in R

Dr Josh A Hodge

## Brief

Ornithologists from the capitol have taken body measurements of mockingjays present in the 3rd quarter quell. Values are averaged for birds captured more than once. The variables measured are:

* “BirdID” = tag for bird
* “KnownSex”= sex coded as F or M
* “Billdepth”= Thickness of the bill measured at the nostril (mm)
* “Billwidth”= width of bill (mm)
* “Billlength”= lenth of bill (mm)
* “Head”= distance from tip bill to back of head (mm)
* “Mass” = body mass (g)
* “Skull”= distance from base of bill to back of skull (mm)
* “Sex”= coded as 0 for females and 1 for males

Using this data, the ornithologists would like you to investigate the sexual dimorphisms of mockingjays and the bivariate relationships between the bill, head and mass traits. A series of exercises have been designed to help you tackles the problem. Sections one and two will be great practice for the base R and lattice graphics, whilst section three will focus on making graphics in ggplot2.

## Section One

This section would ease you into the functions used to make graphs in R. Take them at your own pace.

1. Set working directory

#setwd("")

1. Import “mockingjays.csv” datafile

data<-read.csv("mockingjays.csv")

1. Plot a histogram to examine the distribution of a trait of your choice.

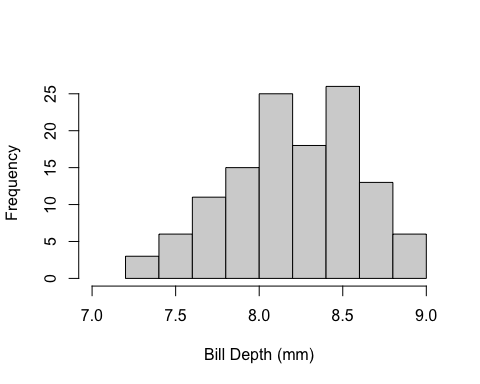
min(data$BillDepth)

## [1] 7.21

max(data$BillDepth)

## [1] 9

hist(data$BillDepth, xlab = "Bill Depth (mm)", ylab="Frequency", xlim = c(7, 9), main = "")



1. Plot two histograms (one for male and one for female) in the same window to examine the distribution of a trait of your choice.

males<- data%>%filter(KnownSex=="M")  
min(males$BillDepth)

## [1] 7.89

max(males$BillDepth)

## [1] 9

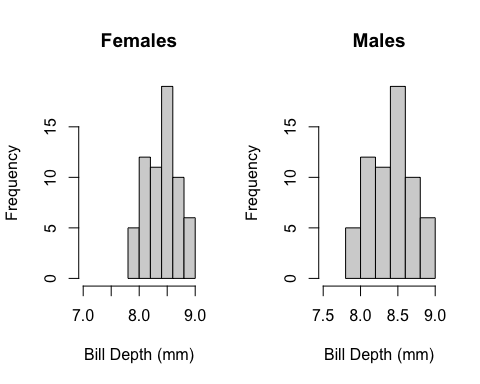
females<-data%>%filter(KnownSex=="F")  
min(females$BillDepth)

## [1] 7.21

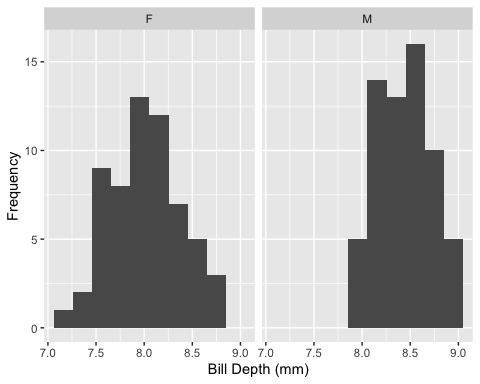
max(females$BillDepth)

## [1] 8.8

par(mfrow=c(1,2))  
hist(males$BillDepth, xlab = "Bill Depth (mm)", ylab="Frequency", xlim = c(7, 9), main = "Females")  
hist(males$BillDepth, xlab = "Bill Depth (mm)", ylab="Frequency", xlim = c(7.5, 9), main = "Males")



#or#  
ggplot(data, aes(x=BillDepth))+  
 geom\_histogram(bins=10)+  
 facet\_grid(.~KnownSex)+   
 labs(x="Bill Depth (mm)", y="Frequency")



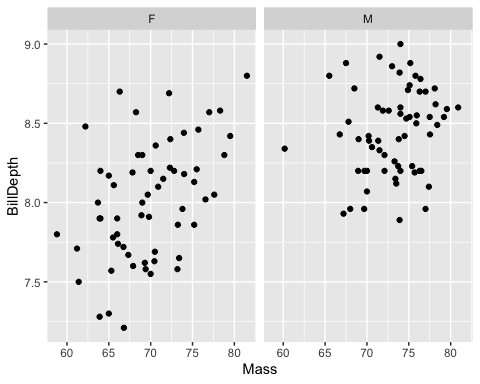
1. Choose two traits and examine their relationship using a scatterplot.

plot(data$BillDepth~data$Mass, xlab="Body Mass (g)", ylab="Bill Depth (mm)", main="")



1. Using the same two traits from question five, are these relationships different in males and females?

ggplot(data, aes(x=Mass, y=BillDepth))+  
 geom\_point()+  
 facet\_grid(.~KnownSex)

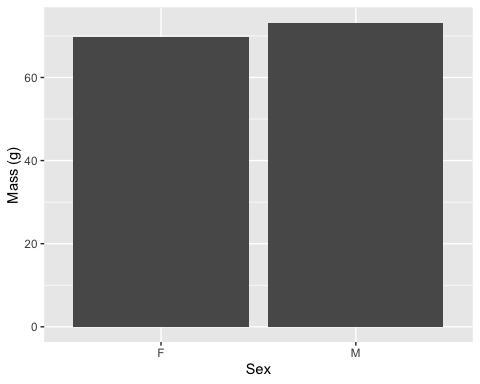


## Section Two

This section is designed to tie in some of the concepts taught in Thursday’s session on data wrangling and the graph making functions in base R and ggplot2 Some questions have been constructed for you to go off and do a little bit of research about new functions, primarily str\_detect(), but do not worry if you can’t figure them out and simply ask for some help from myself or the demonstrators.

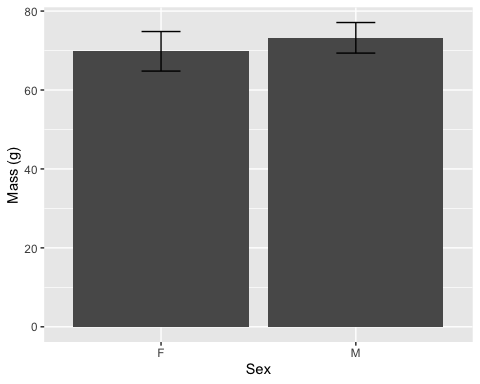
1. Choose a trait and make a summary table, including the standard deviations and means using the group\_by and summarise functions for each sex. Plot a barchart of the means.

summary\_stats <- data%>%group\_by(KnownSex)%>%summarise(Mean=mean(Mass, na.rm = T), SD=sd(Mass, na.rm = T))  
ggplot(summary\_stats, aes(x=KnownSex, y=Mean))+  
 geom\_bar(stat = "identity")+   
 labs(y="Mass (g)", x="Sex")



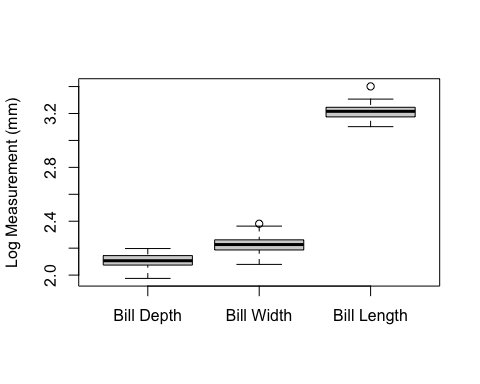
1. Use the standard deviation measures to add error bars to the barchart created in exercise 7. You need to look this up.

ggplot(summary\_stats, aes(x=KnownSex, y=Mean))+  
 geom\_bar(stat = "identity")+   
 geom\_errorbar(aes(ymin=Mean-SD, ymax=Mean+SD), width=0.2)+  
 labs(y="Mass (g)", x="Sex")



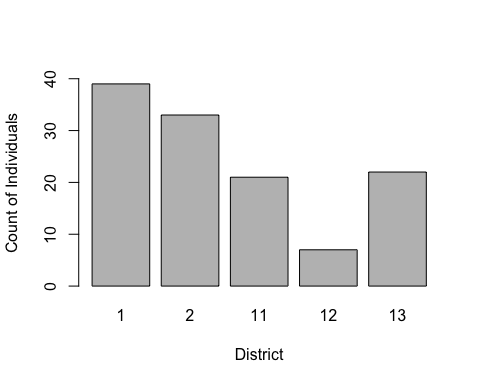
1. Write a code that will reproduce the plot below.

boxplot(log(data$BillDepth), log(data$BillWidth), log(data$BillLength), names=c("Bill Depth", "Bill Width", "Bill Length"), ylab="Log Measurement (mm)")

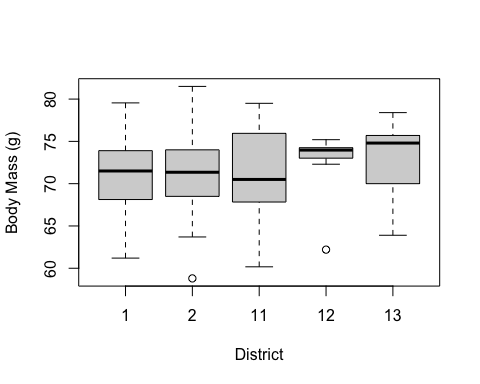


1. The BirdID column has a series of unique identifiers (usually a three or four digit number followed by a hyphen) that indicate the population the mockingjay was caught in. These are refer to the districts in Panem: 962 = District 1; 952 = District 2; 872 = District 11; 702 = District 12 and 1142 = District 13. The ornithologists would like to make a two plots, one that documents the total counts of mockingjays in each of these districts and another that shows the distribution of body mass across the districts. You may have to filter the BirdID column using a new function called str\_detect(), which is part of the “stringr” package so look up its usage with the filter function.

district1<- data%>%select(BirdID, Mass)%>%filter(str\_detect(BirdID, "962"))  
district1$District<- 1  
district2<- data%>%select(BirdID, Mass)%>%filter(str\_detect(BirdID, "952"))  
district2$District<- 2  
district11<- data%>%select(BirdID, Mass)%>%filter(str\_detect(BirdID, "872"))  
district11$District<- 11  
district12<- data%>%select(BirdID, Mass)%>%filter(str\_detect(BirdID, "702"))  
district12$District<- 12  
district13<- data%>%select(BirdID, Mass)%>%filter(str\_detect(BirdID, "1142"))  
district13$District<- 13  
districts<-rbind(district1, district2, district11, district12, district13)  
counts <- districts%>%group\_by(District)%>%summarise(Count=n())  
barplot(counts$Count~counts$District, xlab="District", ylab="Count of Individuals", ylim=c(0,40))



boxplot(districts$Mass~districts$District,xlab="District", ylab="Body Mass (g)")



## Section Three

This cover material outside of what was taught and requires you to do some research around using ggplot2 to reproduce the graphs above. ggplot2 is a one of the best graphing packages out there and mainly publications will use ggplot2.

1. Using ggplot2, recreate your plot from exercise 4. Try and overlay the histograms rather than making a faceted plot.

overlayed\_hist<- ggplot(data=data, aes(x=BillDepth, fill=KnownSex)) +  
 geom\_histogram(binwidth=.25, alpha=.5, position="identity")+   
 labs(x="Bill Depth (mm)", y="Count")

1. Using ggplot2, recreate your plot from exercises 7 and 8.

summary\_stats<- data%>%group\_by(KnownSex)%>%summarise(mean.billdepth=mean(BillDepth), sd.billdepth=sd(BillDepth))  
barchart\_w\_errors<- ggplot(data=summary\_stats, aes(x=KnownSex, y=mean.billdepth))+  
 geom\_bar(stat = "identity", fill=c("purple","orange"))+   
 geom\_errorbar(aes(ymin=mean.billdepth-sd.billdepth, ymax=mean.billdepth+sd.billdepth), width=0.2, position = position\_dodge(0.9))+  
 labs(y="Bill Depth (mm)", x="Sex")+  
 ylim(0, 10)