Assignment3

Assignment3

2022-10-12

# 1 start

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(Stat2Data)  
data("Hawks")  
c<-Hawks$Tail  
print(mean(c))

## [1] 198.8315

# 1.1&1.2

k<-Hawks %>% summarise(Wing\_mean=mean(Wing,na.rm=TRUE),Wing\_t\_mean=mean(Wing,na.rm=TRUE,trim=0.5),Wing\_med=median(Wing,na.rm=TRUE))  
print(k)

## Wing\_mean Wing\_t\_mean Wing\_med  
## 1 315.6375 370 370

kk<-Hawks %>%group\_by(Species)%>% summarise(Wing\_mean=mean(Wing,na.rm=TRUE),Wing\_t\_mean=mean(Wing,na.rm=TRUE,trim=0.5),Wing\_med=median(Wing,na.rm=TRUE))  
print(kk)

## # A tibble: 3 × 4  
## Species Wing\_mean Wing\_t\_mean Wing\_med  
## <fct> <dbl> <dbl> <dbl>  
## 1 CH 244. 240 240  
## 2 RT 383. 384 384  
## 3 SS 185. 191 191

# 1.4

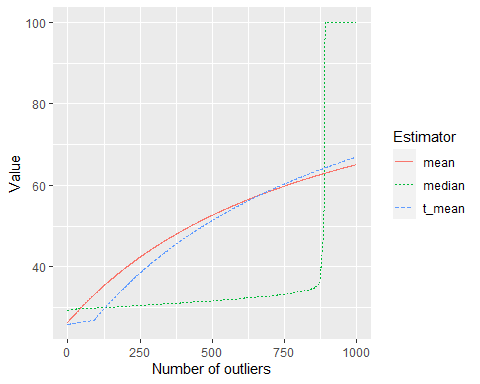
hal<-Hawks$Hallux  
hal<-hal[!is.na(hal)]  
outlier\_val<-100  
num\_outliers<-10  
corrupted\_hal<-c(hal,rep(outlier\_val,times=num\_outliers))  
mean(hal)

## [1] 26.41086

mean(corrupted\_hal)

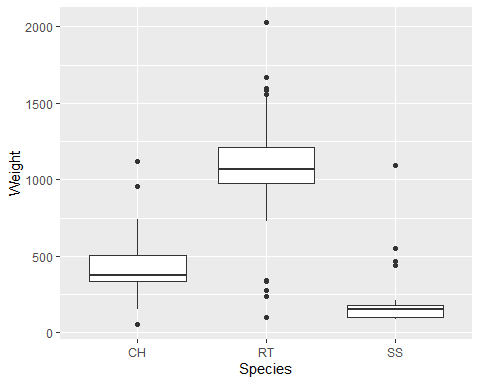
## [1] 27.21776

num\_outliers\_vect <- seq(0,1000)  
means\_vect <- c()  
for(num\_outliers in num\_outliers\_vect){  
corrupted\_hal <- c(hal,rep(outlier\_val,times=num\_outliers))  
means\_vect <- c(means\_vect, mean(corrupted\_hal))  
}  
medians\_vect<-c()  
for(num\_outliers in num\_outliers\_vect){  
corrupted\_hal <- c(hal,rep(outlier\_val,times=num\_outliers))  
medians\_vect <- c(medians\_vect, median(corrupted\_hal))  
}  
t\_means\_vect<-c()  
for(num\_outliers in num\_outliers\_vect){  
corrupted\_hal <- c(hal,rep(outlier\_val,times=num\_outliers))  
t\_means\_vect <- c(t\_means\_vect, mean(corrupted\_hal,trim=0.1))  
}  
df\_means\_medians <- data.frame(num\_outliers=num\_outliers\_vect, mean=means\_vect,  
t\_mean=t\_means\_vect, median=medians\_vect)  
df\_means\_medians %>%  
pivot\_longer(!num\_outliers, names\_to = "Estimator", values\_to = "Value") %>%  
ggplot(aes(x=num\_outliers,color=Estimator, linetype=Estimator,y=Value)) +  
geom\_line()+xlab("Number of outliers")

 # 1.5

ggplot(data=Hawks,aes(x=Species,y=Weight))+geom\_boxplot()+xlab('Species')+ylab('Weight')

## Warning: Removed 10 rows containing non-finite values (stat\_boxplot).



cc<-Hawks%>%select(Species,Weight)%>%group\_by(Species)  
summarise(cc,quantitile025=quantile(Weight,probs=0.25,na.rm=TRUE),quantitle050=quantile(Weight,probs=0.50,na.rm=TRUE),quantitle075=quantile(Weight,probs=0.75,na.rm=TRUE))

## # A tibble: 3 × 4  
## Species quantitile025 quantitle050 quantitle075  
## <fct> <dbl> <dbl> <dbl>  
## 1 CH 335 378. 505   
## 2 RT 980 1070 1210   
## 3 SS 100 155 178.

## 1.5 Q3

num\_outliers <-function(c){  
 min<-quantile(c,probs = 0.25,na.rm = TRUE)-1.5\*IQR(c,na.rm = TRUE)  
 max<-quantile(c,probs = 0.75,na.rm = TRUE)+1.5\*IQR(c,na.rm = TRUE)  
 sum<-0  
 for(i in seq(along=c)){  
 if(c[i]<min||c[i]>max){  
 sum<-sum+1  
 }  
 }  
 return(sum)  
}  
num\_outliers(c(0, 40,60,185))

## [1] 1

cc<-Hawks%>%select(Species,Weight)%>%group\_by(Species,na.rm=TRUE)  
summarise(cc, num\_outliers\_weight=num\_outliers(as.integer(na.omit(Weight))))

## `summarise()` has grouped output by 'Species'. You can override using the  
## `.groups` argument.

## # A tibble: 3 × 3  
## # Groups: Species [3]  
## Species na.rm num\_outliers\_weight  
## <fct> <lgl> <dbl>  
## 1 CH TRUE 3  
## 2 RT TRUE 13  
## 3 SS TRUE 4

# 3 Visualisation

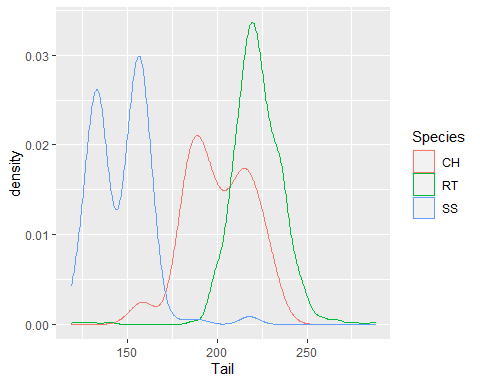
cov(Hawks$Weight,Hawks$Wing,use='complete.obs')

## [1] 41174.39

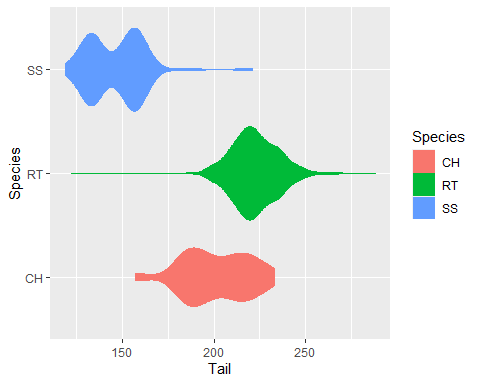
cor(Hawks$Weight,Hawks$Wing,use='complete.obs')

## [1] 0.9348575

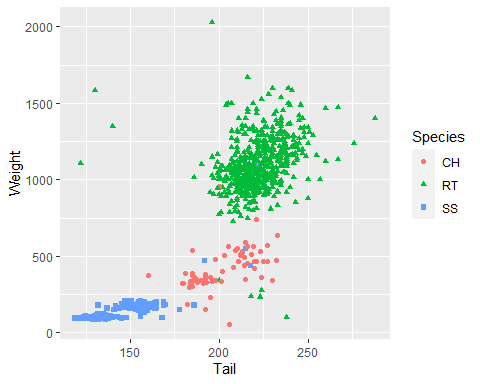
tt<-Hawks%>%select(Species,Tail)%>%group\_by(Species)  
  
ggplot(data=tt,aes(x=Tail,color=Species))+xlab("Tail")+geom\_density(na.rm = FALSE)+ylab("density")



ggplot(data=tt,aes(x=Tail,y=Species,fill=Species,color=Species))+xlab("Tail")+geom\_violin()

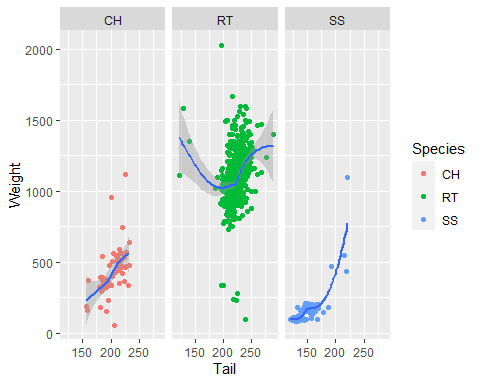


uu<-na.omit(Hawks%>%select(Species,Tail,Weight)%>%group\_by(Species))  
ggplot(data=uu,aes(x=Tail,y=Weight))+xlab("Tail")+ylab("Weight")+geom\_point(aes(color=Species,shape=Species))



ggplot(data=uu,aes(x=Tail,y=Weight))+xlab("Tail")+ylab("Weight")+geom\_point(aes(color=Species))+geom\_smooth()+facet\_wrap(~Species)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



d<-Hawks%>%filter(Weight>=2000&Tail>175)  
xx<-c(d$Tail);  
yy<-c(d$Weight);  
ggplot(data=uu,aes(x=Tail,y=Weight))+xlab("Tail")+ylab("Weight")+geom\_point(aes(color=Species,shape=Species))+geom\_curve(x=200,xend=xx[1],y=1800,yend=yy[1],arrow=arrow(length = unit(0.5,'cm')),curvature = 0.1)+geom\_text(x=225,y=2050,label="see this is outstanding")

