Assignment3

Assignment3

2022-10-12

1 start

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyve
rse 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_co
nflicts() ──
## ✖ 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=T
RUE),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

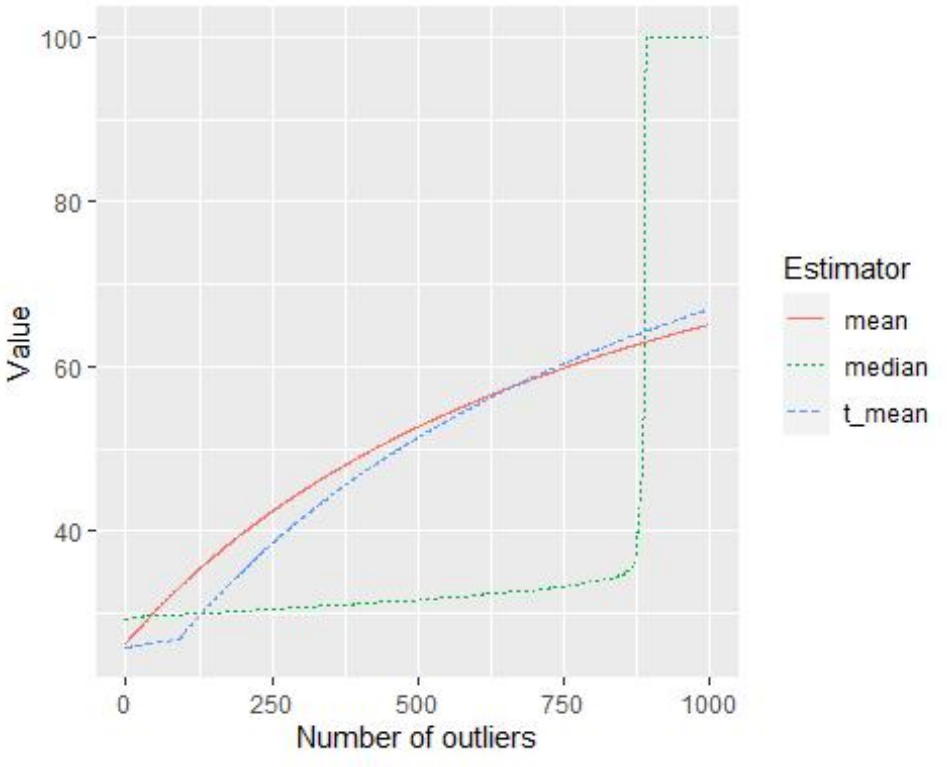
```
hal<-Hawks$Hallux
hal<-hal[!is.na(hal)]
outlier_val<-100
num_outliers<-10
corrupted_hal<-c(hal,rep(outlier_val,times=num_outliers))
mean(hal)
```

```
## [1] 26.41086
```

```
mean(corrupted_hal)
```

```
## [1] 27.21776
```

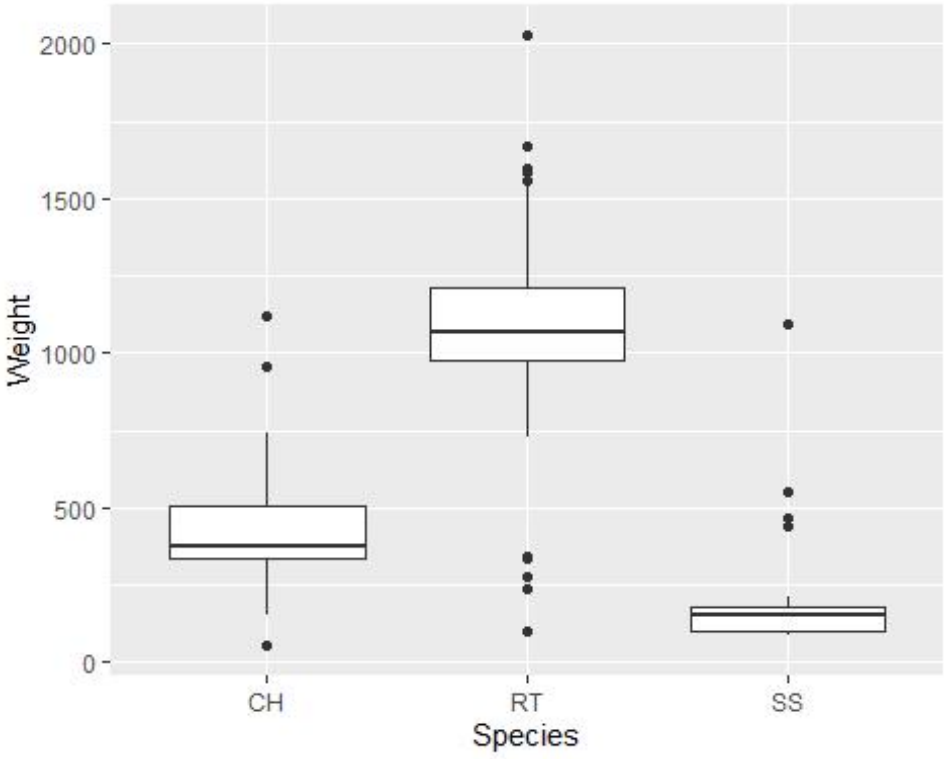
```
num_outliers_vect <- seq(0,1000)
means_vect <- c()
for(num_outliers in num_outliers_vect){
corrupted_hal <- c(hal,rep(outlier_val,times=num_outliers))
means_vect <- c(means_vect, mean(corrupted_hal))
}
medians_vect<-c()
for(num_outliers in num_outliers_vect){
corrupted_hal <- c(hal,rep(outlier_val,times=num_outliers))
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=mea
ns_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),quant
itle050=quantile(Weight,probs=0.50,na.rm=TRUE),quantitle075=quantile(We
ight,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(Weigh
t))))
```

```
## `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

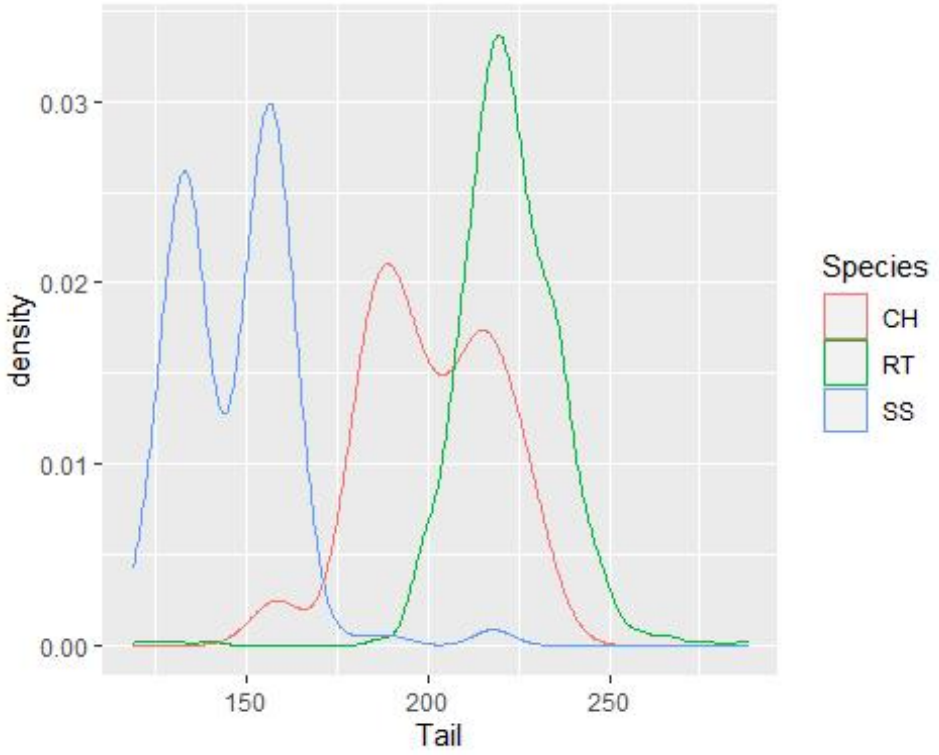
```
cov(Hawks$Weight,Hawks$Wing,use='complete.obs')
```

```
## [1] 41174.39
```

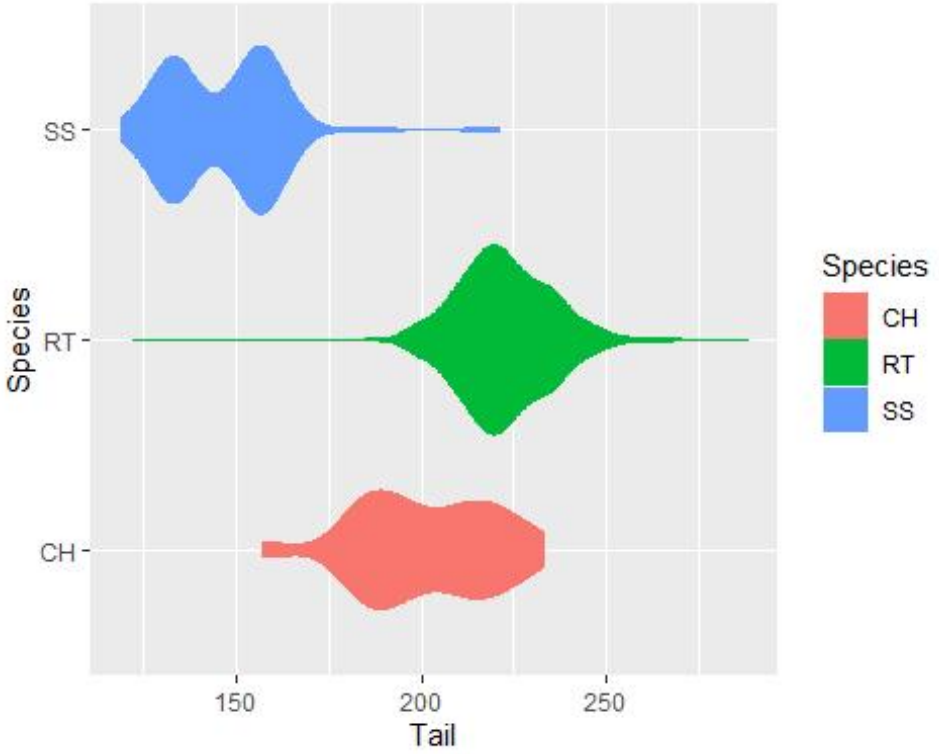
```
cor(Hawks$Weight,Hawks$Wing,use='complete.obs')
```

```
## [1] 0.9348575
```

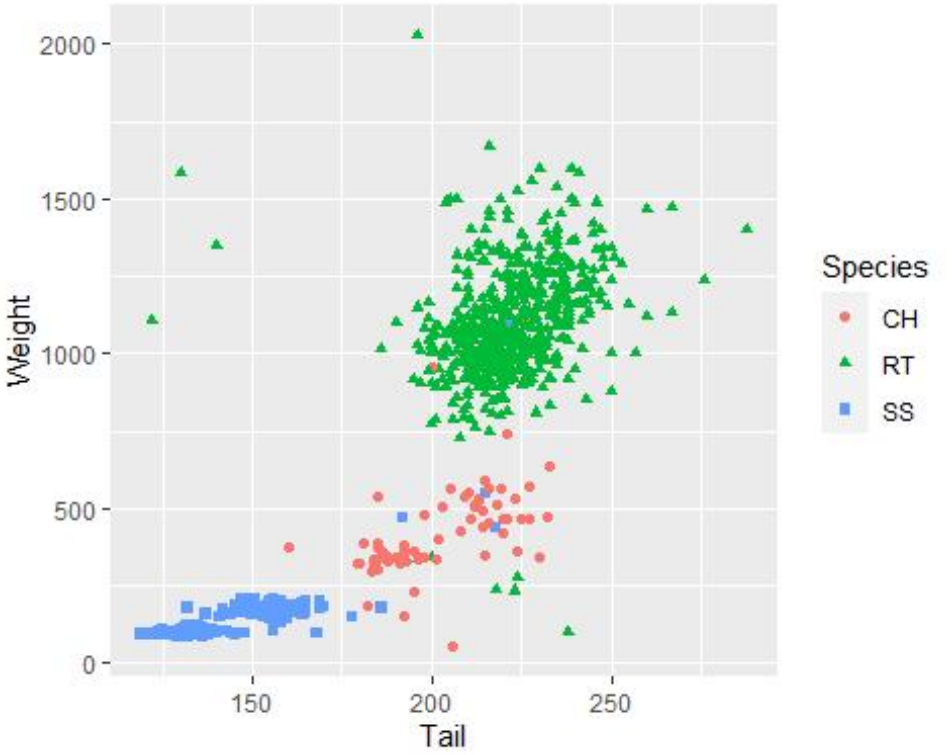
```
tt<-Hawks%>%select(Species,Tail)%>%group_by(Species)

ggplot(data=tt,aes(x=Tail,color=Species))+xlab("Tail")+geom_density(na.
rm = FALSE)+ylab("density")
```



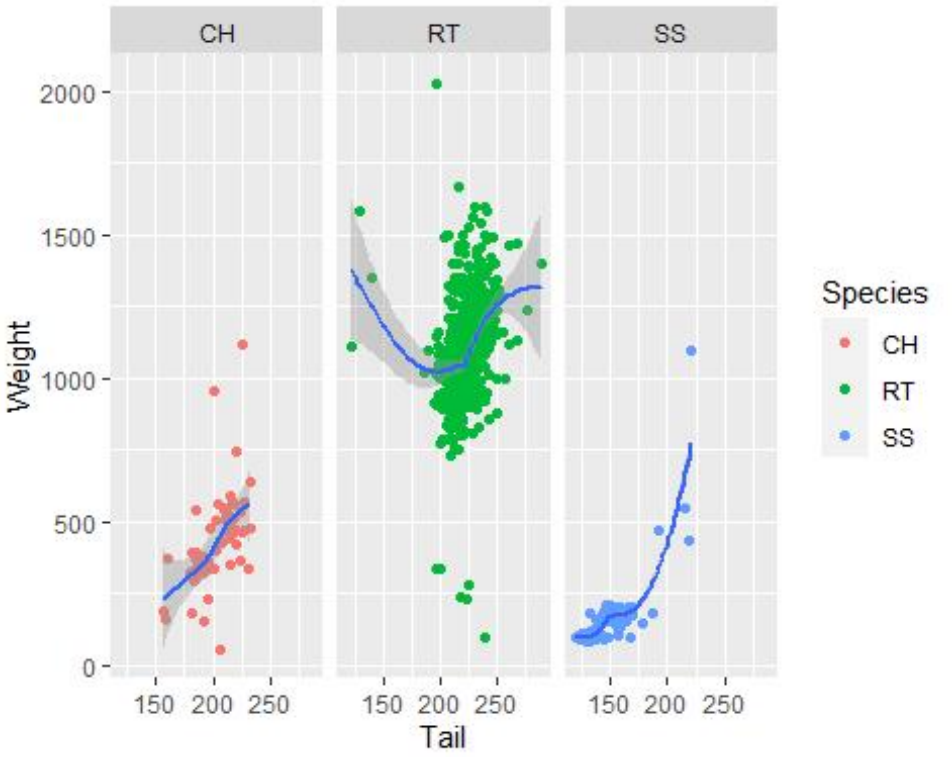

```
ggplot(data=tt,aes(x=Tail,y=Species,fill=Species,color=Species))+xlab("
Tail")+geom_violin()
```



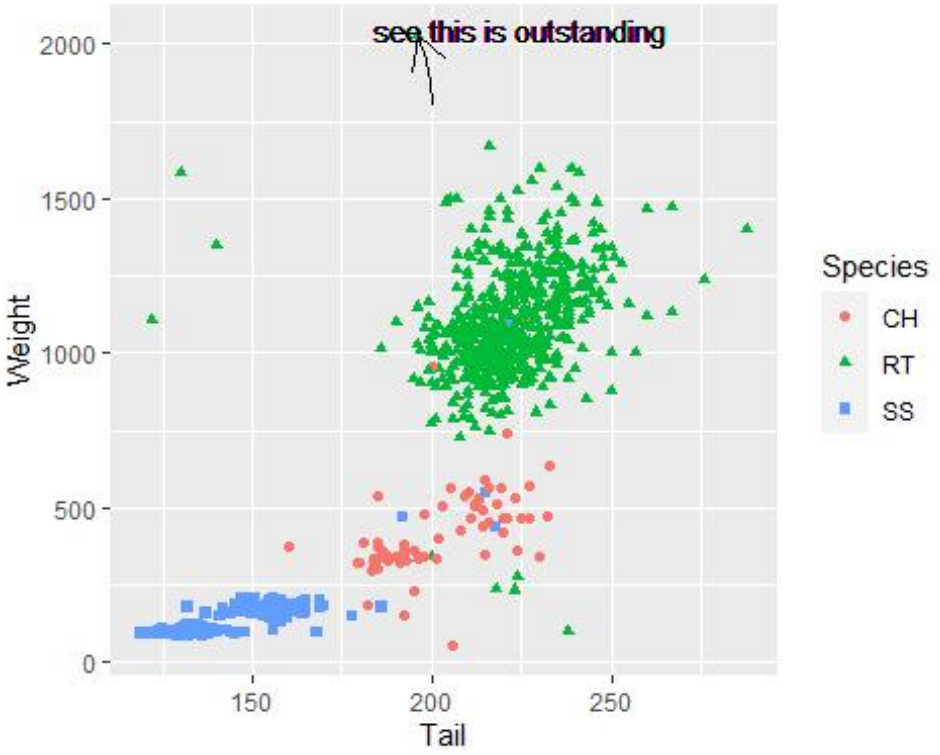
```
uu<-na.omit(Hawks%>%select(Species,Tail,Weight)%>%group_by(Species))
ggplot(data=uu,aes(x=Tail,y=Weight))+xlab("Tail")+ylab("Weight")+geom_p
oint(aes(color=Species,shape=Species))
```




```
ggplot(data=uu,aes(x=Tail,y=Weight))+xlab("Tail")+ylab("Weight")+geom_p
oint(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_p
oint(aes(color=Species,shape=Species))+geom_curve(x=200,xend=xx[1],y=18
00,yend=yy[1],arrow=arrow(length = unit(0.5,'cm')),curvature = 0.1)+geo
m_text(x=225,y=2050,label="see this is outstanding")
```


see this is outstanding
2000
1500
1000
500
0
Weight
150
200
250
Tail
Species
CH
RT
SS