통계의 활용 (2020 년 2 학기)			
담 당 교 수 : 김 태 수			
강좌번호		본인의 과제 자체 평가	

이름	김태형

제 출 일	2020 년 9 월 29 일
학 과	기초교육학부(교류학생)
학 번	

목차

1. 자료설명

DATA : IRIS(붓꽃)

-통계학자인 피셔[Fisher] 가 소개한 데이터,

-붓꽃의 3 가지 종(setosa, versicolor, virginica)에 대해 꽃받침[sepal]과 꽃잎[petal]의 길이를 정리 한 데이터. - 칼럼별 설명

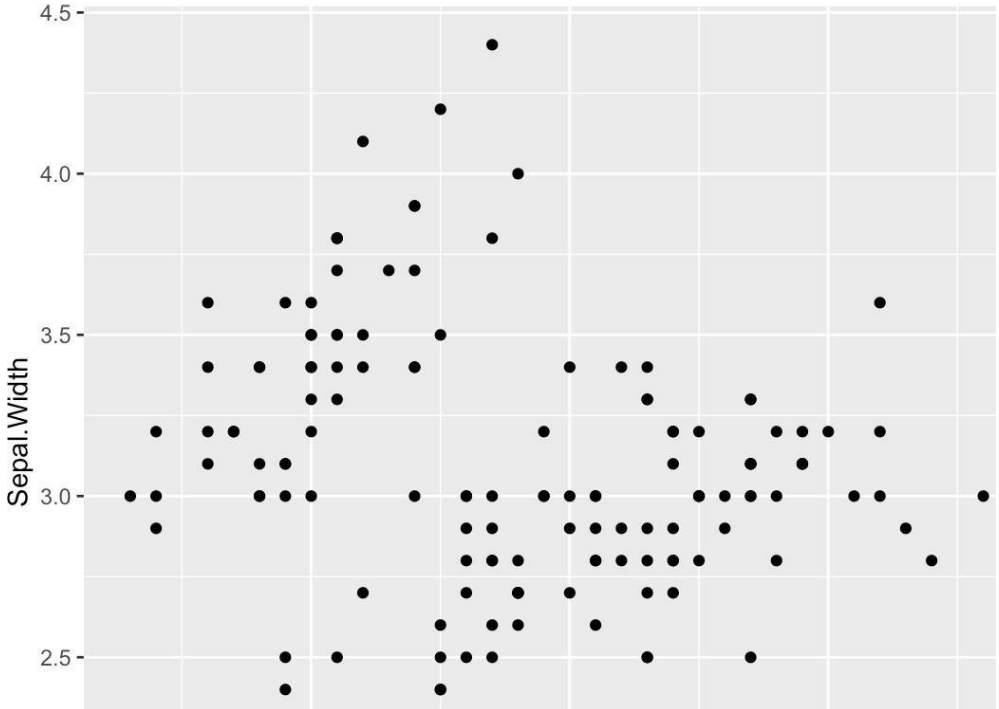
컬럼명	의미	데이터 타입
Species	붓꽃의 종. setosa, versicolor, virginica 세 가지 값 중 하나	Factor
Sepal.Width	꽃받침의 너비	Number
Sepal.Length	꽃받침의 길이	Number
Petal.Width	꽃잎의 너비	Number
Petal.Length	꽃잎의 길이	Number

iris 에는 붓꽃의 종별로 50 행씩, 총 150 개 행이 저장되어 있다.

2. 그래프 정리

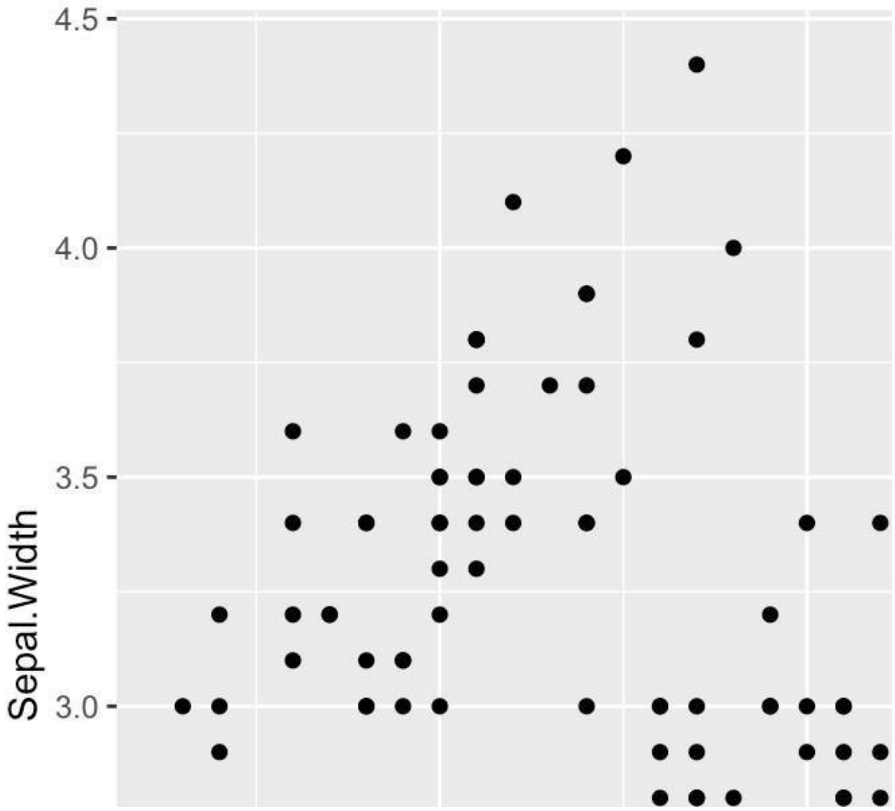
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point()

가로 축은 Sepal.Length 로, 세로축은 Sepal.Width 로 설정, geom_point 로



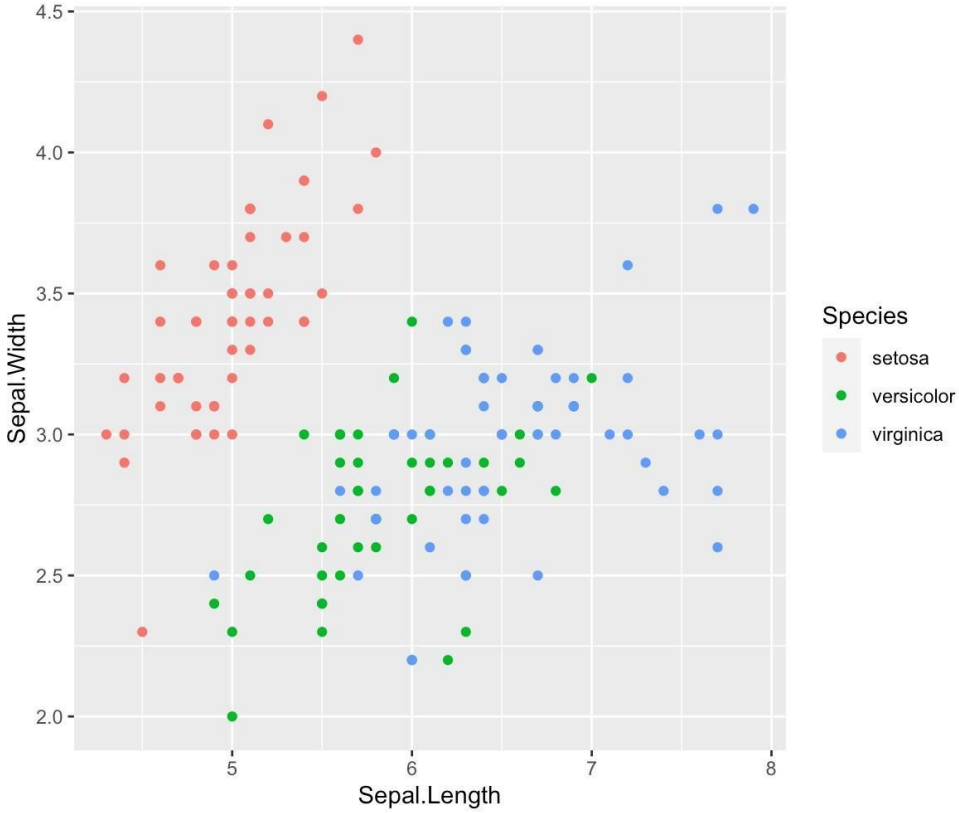
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point(aes(colour = Species, size=Petal.Width), alpha=I(0.7))

중복되어있는 점(겹쳐있는 점)을 표현



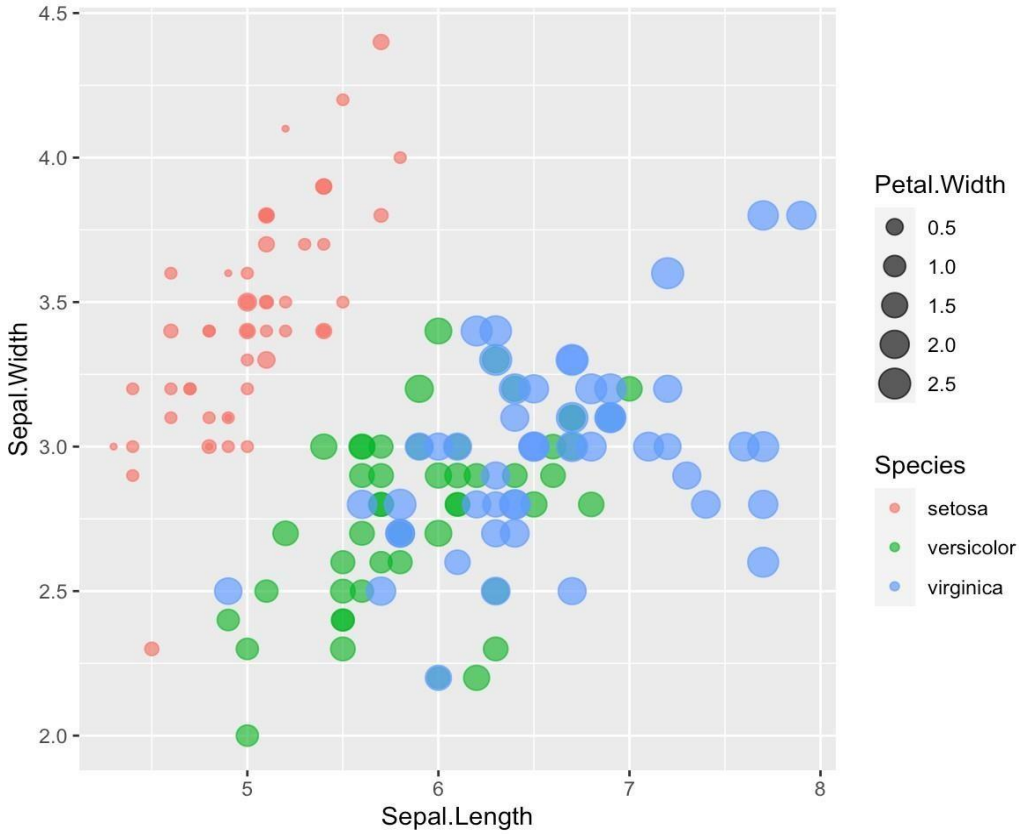
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point(aes(colour = Species))

색상추가



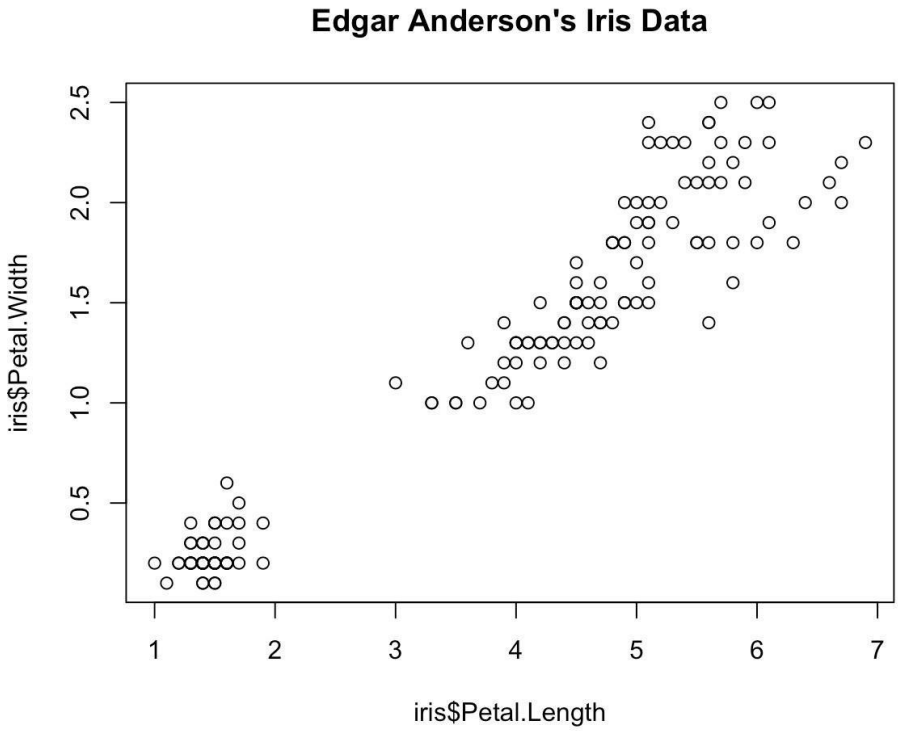
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point(aes(colour = Species, size=Petal.Width))

point 의 크기는 꽃잎의 넓이(Petal.Length)에 따라 설정

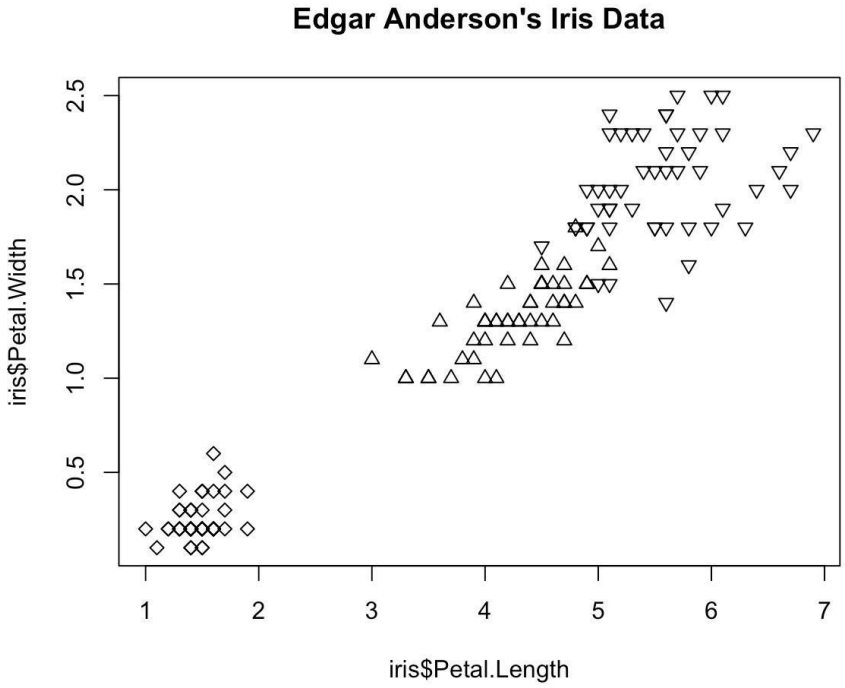


plot(iris$Petal.Length, iris$Petal.Width, main="Edgar Anderson's Iris Data")

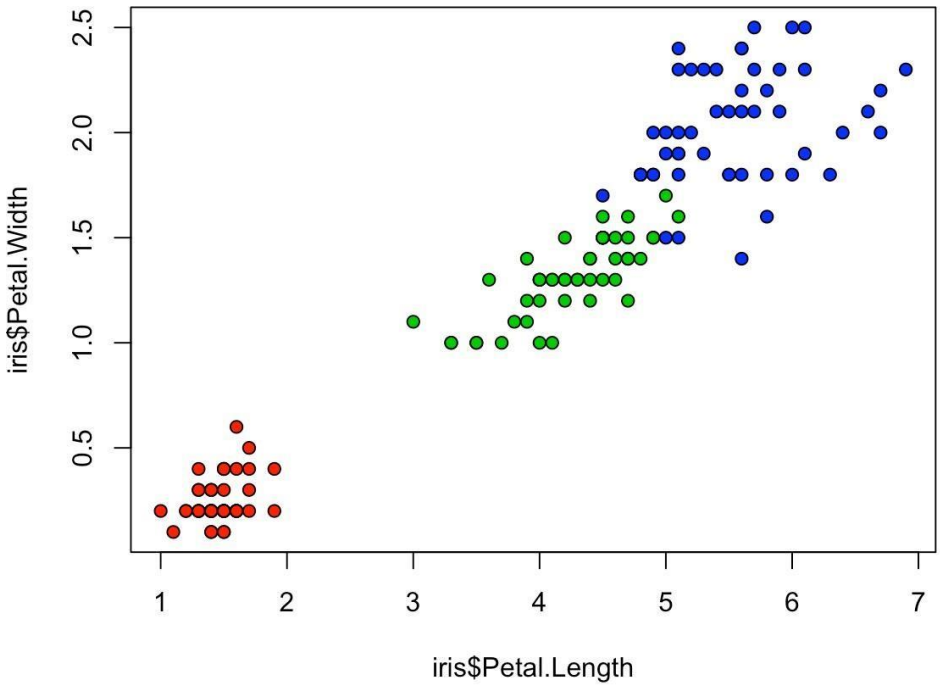
Simple Scatter Plots



plot(iris$Petal.Length, iris$Petal.Width, pch=c(23,24,25)[unclass(iris$Species)], main="Edgar Anderson's Iris Data")

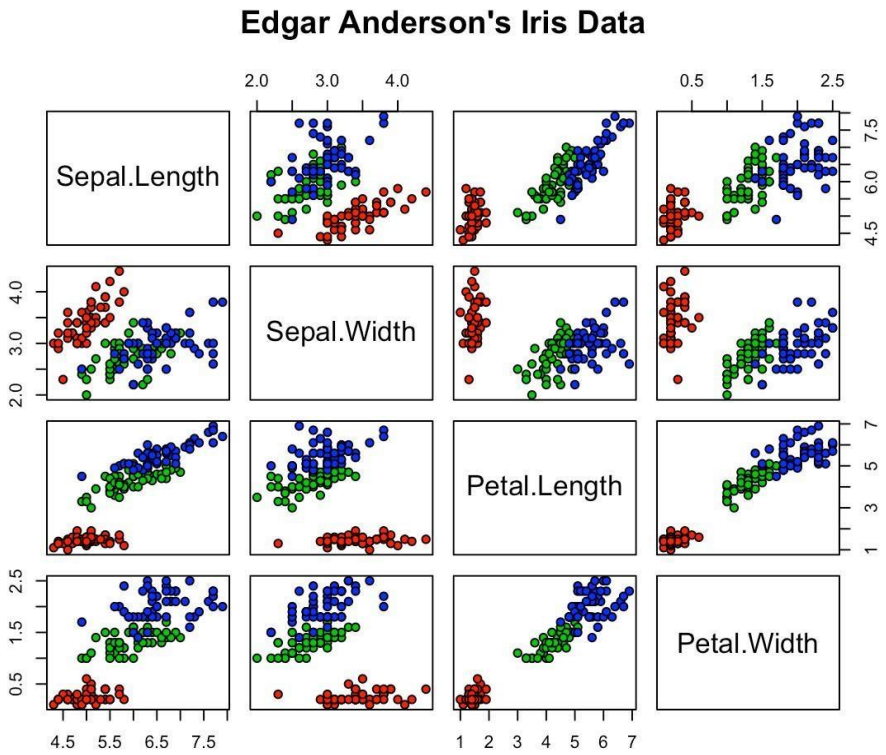


plot(iris$Petal.Length, iris$Petal.Width, pch=21, bg=c("red","green3","blue")[unclass(iris$Species)], main="Edgar Anderson's Iris Data")

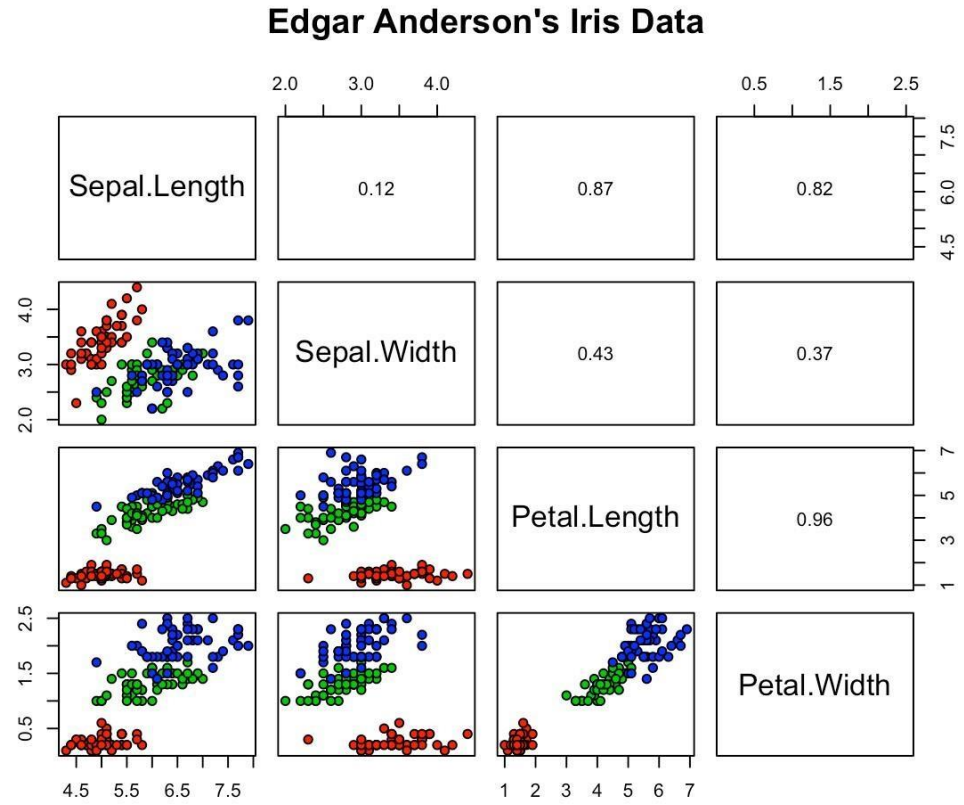


pairs(iris[1:4], main = "Edgar Anderson's Iris Data", pch = 21, bg = c("red","green3","blue")[unclass(iris$Species)], upper.panel=panel.pearson)

 # panel.pearson <-
function(x, y, ...) {

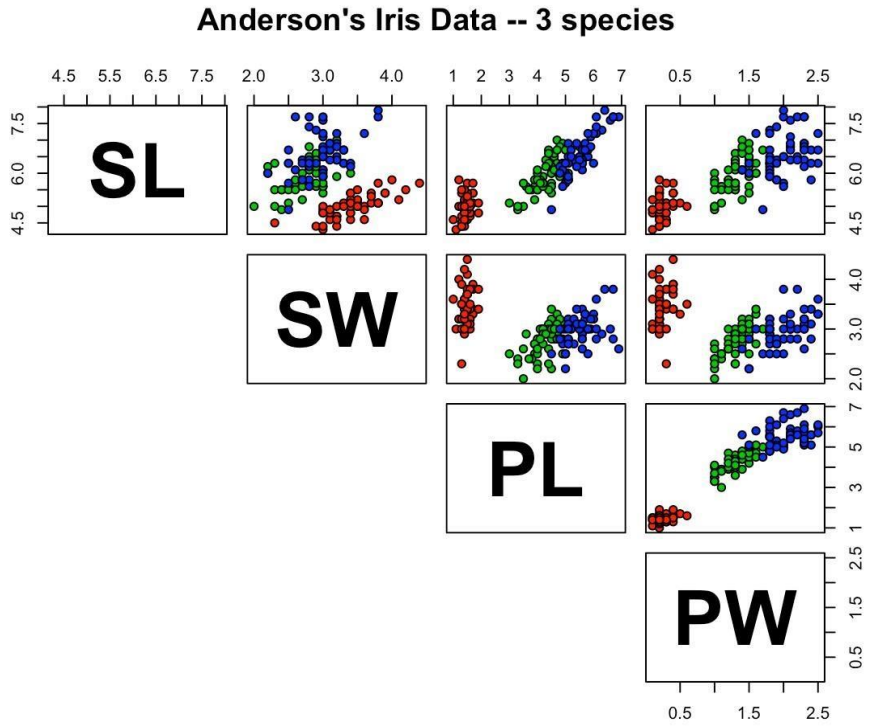



```
horizontal <- (par("usr")[1] + par("usr")[2]) / 2;  vertical <- (par("usr")[3] + par("usr")[4]) / 2; text(horizontal, vertical, format(abs(cor(x,y)), digits=2))
pairs(iris[1:4], main = "Edgar Anderson's Iris Data", pch = 21, bg = c("red","green3","blue")[unclass(iris$Species)],
```

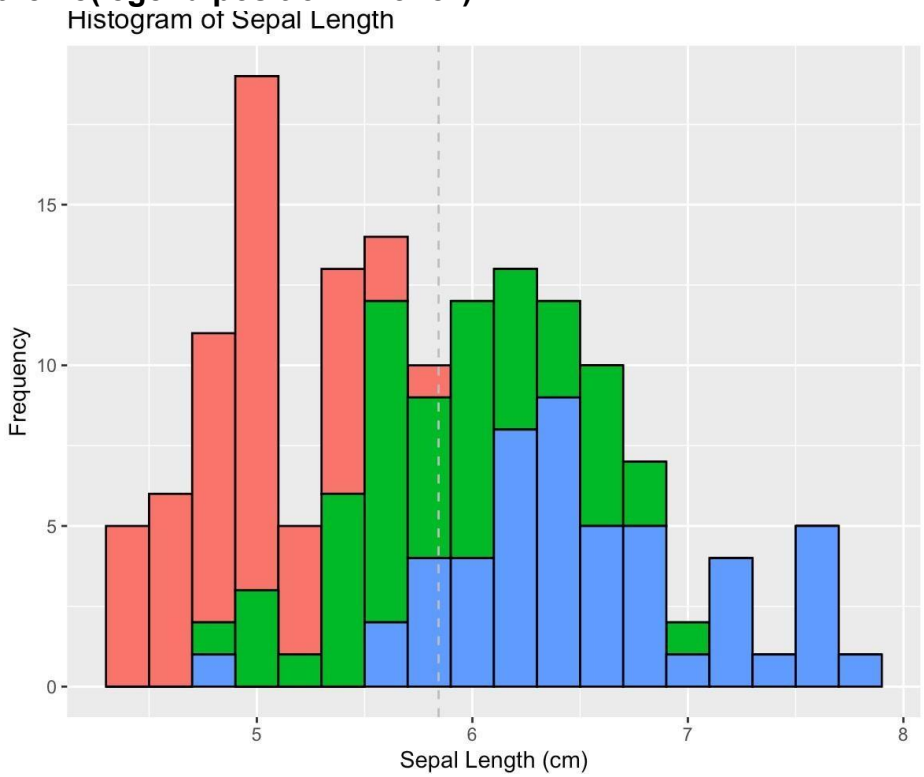



```
upper.panel=panel.pearson)
```

```
pairs(iris[1:4], main = "Anderson's Iris Data -- 3 species", pch = 21, bg = c("red", "green3",
"blue")
     [unclass(iris$Species)], lower.panel=NULL,
labels=c("SL","SW","PL","PW"), font.labels=2, cex.labels=4.5)
```

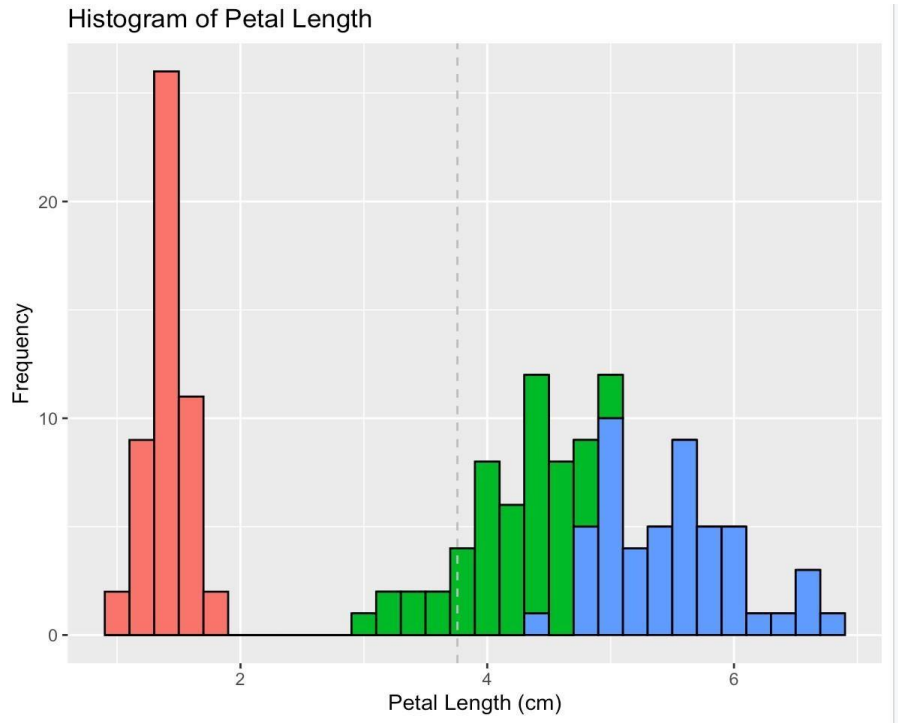



```
HisSl <- ggplot(data=iris, aes(x=Sepal.Length))+
  geom_histogram(binwidth=0.2, color="black", aes(fill=Species)) +
xlab("Sepal Length (cm)") +     ylab("Frequency") +
theme(legend.position="none")+
```



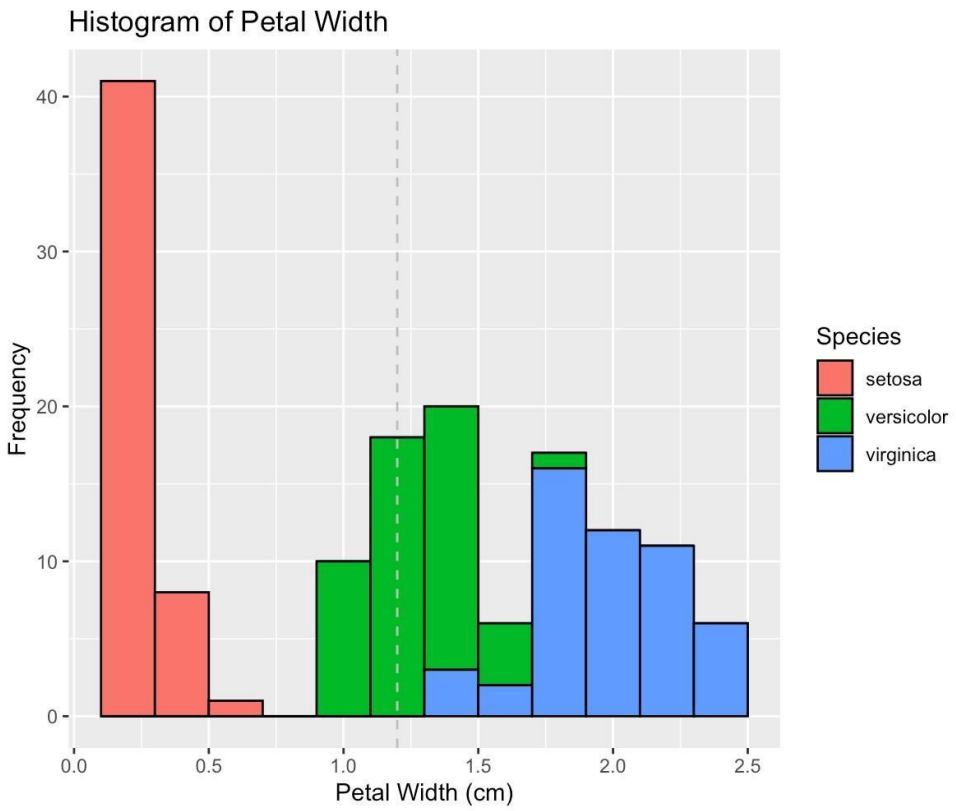

```
HistPl <- ggplot(data=iris, aes(x=Petal.Length))+
  geom_histogram(binwidth=0.2, color="black", aes(fill=Species))
+    xlab("Petal Length (cm)") +     ylab("Frequency") +
theme(legend.position="none")+   ggtitle("Histogram of Petal
```

Length")+ geom_vline(data=iris, aes(xintercept = mean(Petal.Length)), linetype="dashed",color="grey")



HistPw <- ggplot(data=iris, aes(x=Petal.Width))+ geom_histogram(binwidth=0.2, color="black", aes(fill=Species)) + xlab("Petal Width (cm)") + ylab("Frequency") + theme(legend.position="right")+ ggtitle("Histogram of Petal Width")+

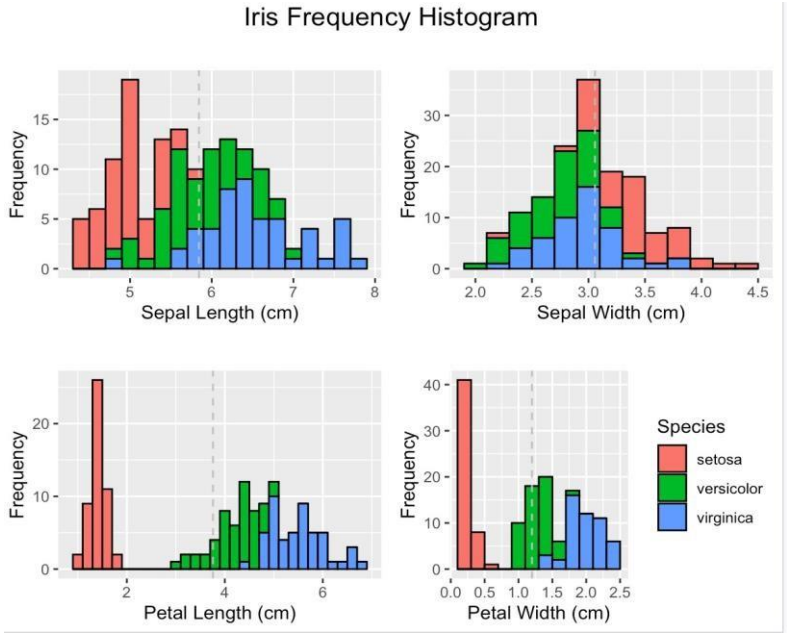
```
geom_vline(data=iris, aes(xintercept = mean(Petal.Width)),linetype="dashed",color="grey")
```



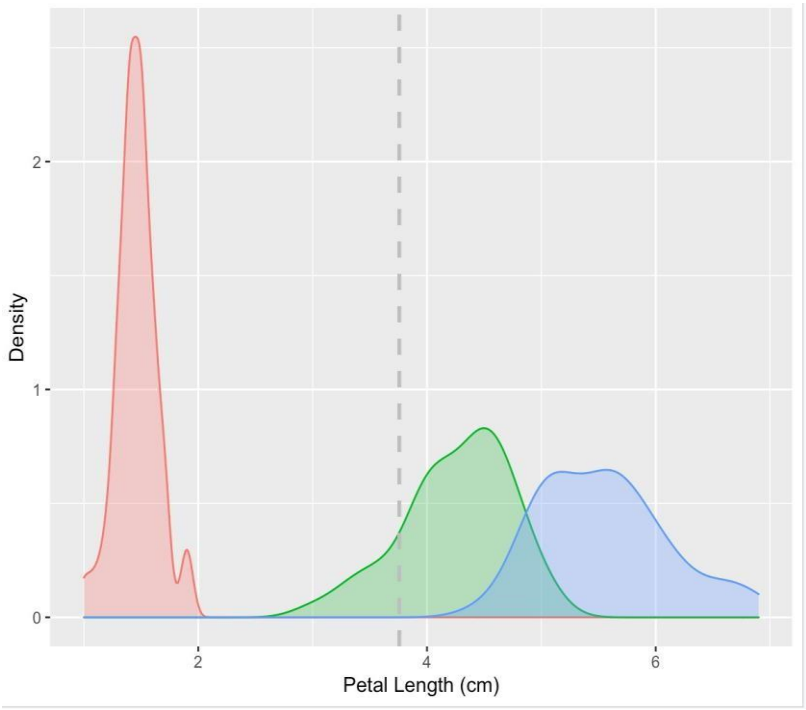

```
grid.arrange(HisSl
```

```
+ ggtitle(""),
        HistSw + ggtitle(""),
        HistPl + ggtitle(""),
HistPw          +       ggtitle(""),
nrow = 2,
        top = textGrob("Iris Frequency Histogram",
gp=gpar(fontsize=15))
```

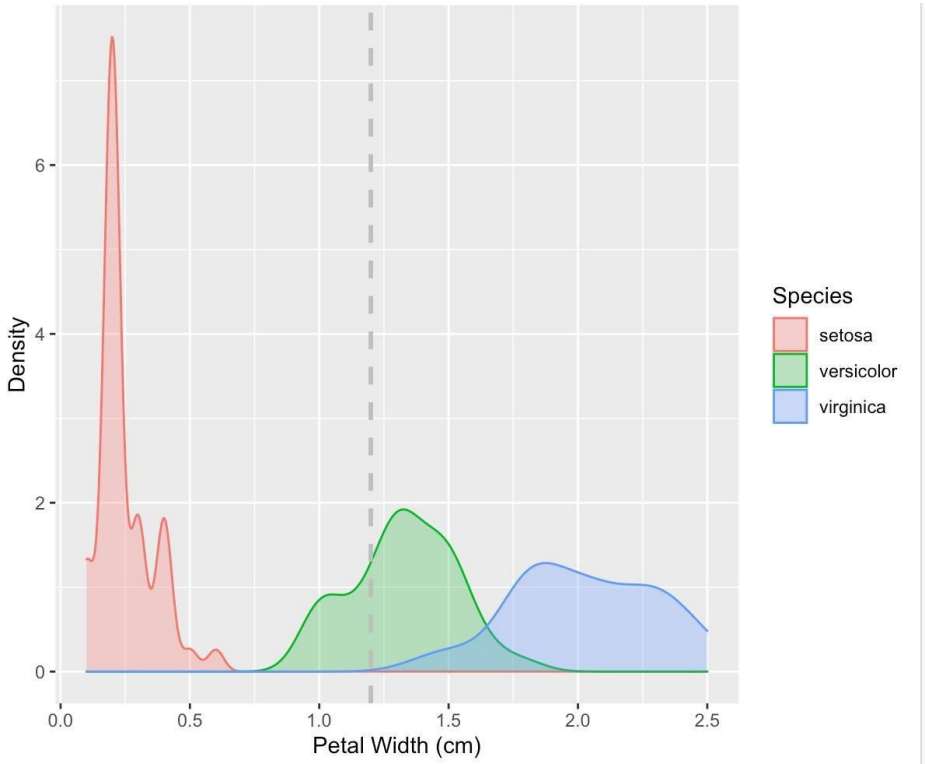
)



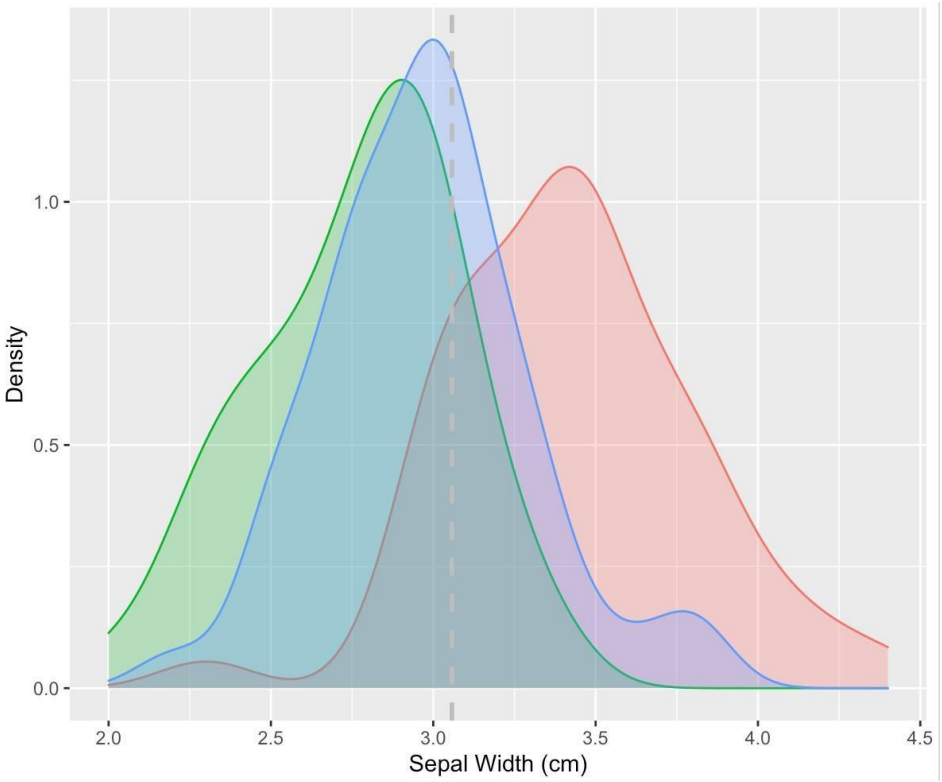

```
DhistPl <- ggplot(iris, aes(x=Petal.Length, colour=Species, fill=Species)) +
geom_density(alpha=.3) +
  geom_vline(aes(xintercept=mean(Petal.Length),
colour=Species),linetype="dashed",color="grey", size=1)+
xlab("Petal Length (cm)") +     ylab("Density")+
theme(legend.position="none")
```



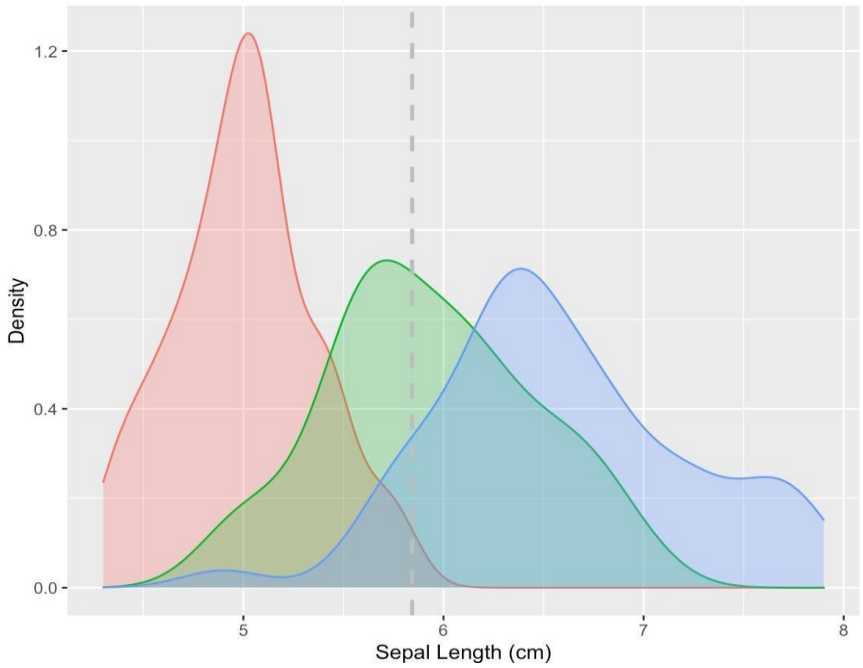

```
DhistPw <- ggplot(iris, aes(x=Petal.Width, colour=Species, fill=Species)) +   geom_density(alpha=.3)
+
  geom_vline(aes(xintercept=mean(Petal.Width),
colour=Species),linetype="dashed",color="grey", size=1)+
xlab("Petal Width (cm)") +     ylab("Density")
```



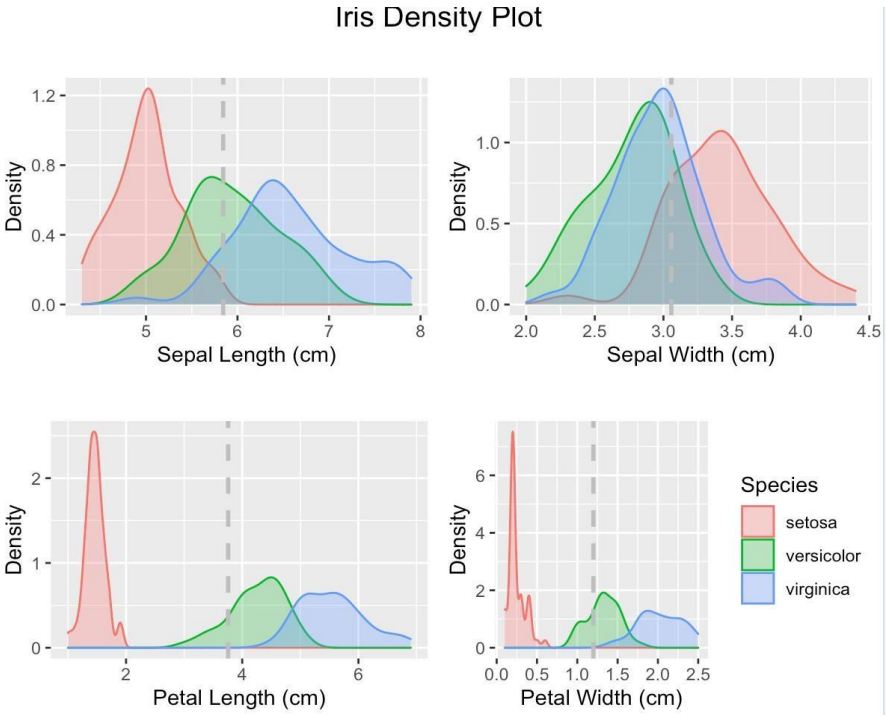

```
DhistSw <- ggplot(iris, aes(x=Sepal.Width, colour=Species, fill=Species)) +
geom_density(alpha=.3) +
  geom_vline(aes(xintercept=mean(Sepal.Width),  colour=Species),
linetype="dashed",color="grey", size=1)+   xlab("Sepal Width (cm)") +
ylab("Density")+   theme(legend.position="none")
```



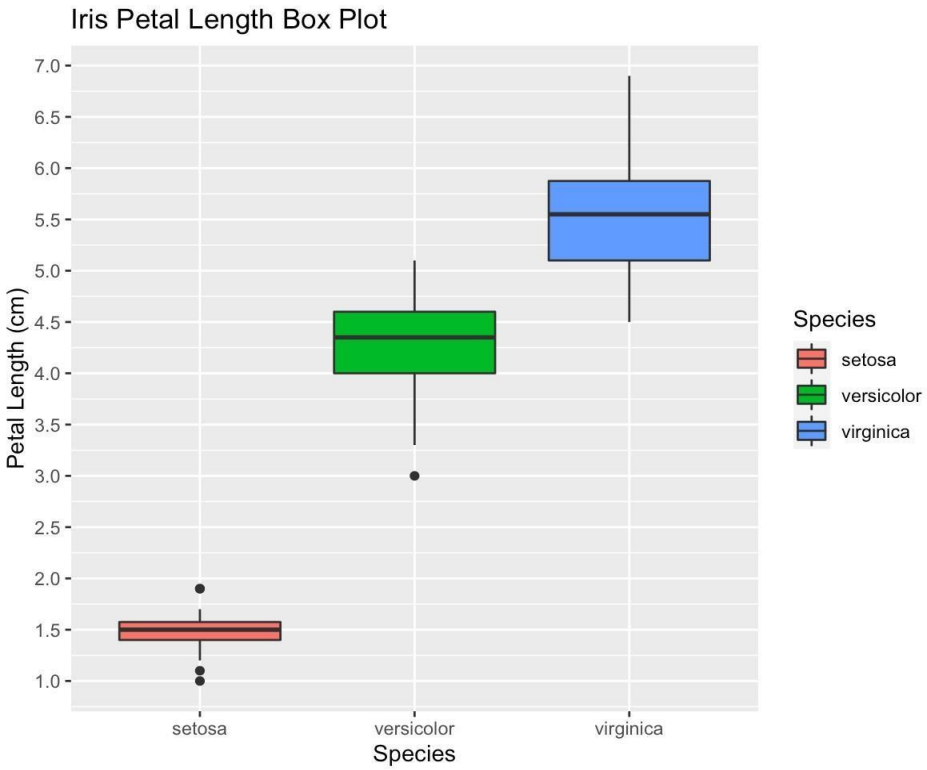

```
DhistSl <- ggplot(iris, aes(x=Sepal.Length, colour=Species, fill=Species)) +
geom_density(alpha=.3) +
  geom_vline(aes(xintercept=mean(Sepal.Length),  colour=Species),linetype="dashed",
color="grey", size=1)+   xlab("Sepal Length (cm)") +     ylab("Density")+
theme(legend.position="none")
```



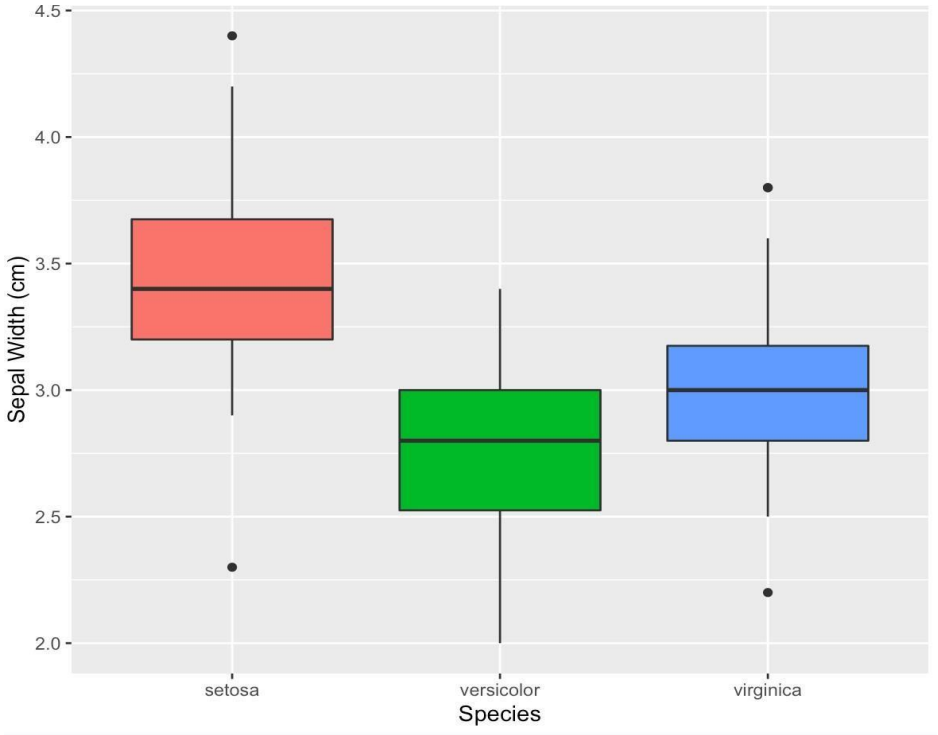

```
grid.arrange(DhistSl + ggtitle(""),
DhistSw  + ggtitle(""),
DhistPl          +          ggtitle(""),
DhistPw                     +          ggtitle(""),
nrow = 2,
        top = textGrob("Iris Density Plot",
                gp=gpar(fontsize=15))
        )
```



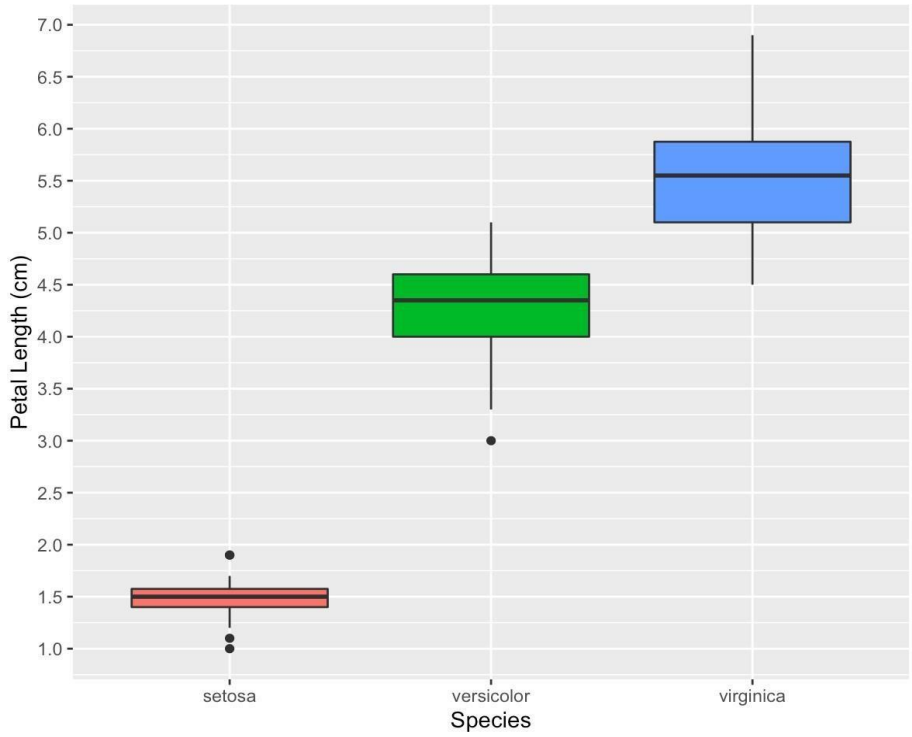

```
BpSl <- ggplot(iris, aes(Species, Sepal.Length, fill=Species)) + geom_boxplot()+
scale_y_continuous("Sepal Length (cm)", breaks= seq(0,30, by=.5))+
theme(legend.position="none")
```



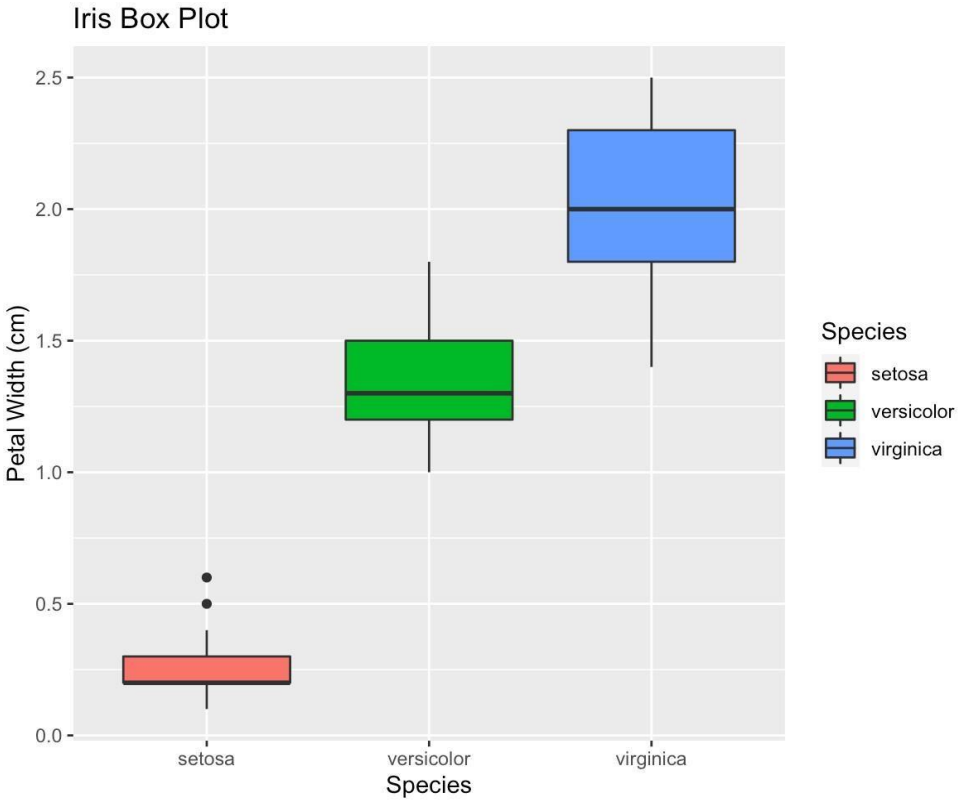
BpSw <- ggplot(iris, aes(Species, Sepal.Width, fill=Species)) + geom_boxplot()+ scale_y_continuous("Sepal Width (cm)", breaks= seq(0,30, by=.5))+ theme(legend.position="none")



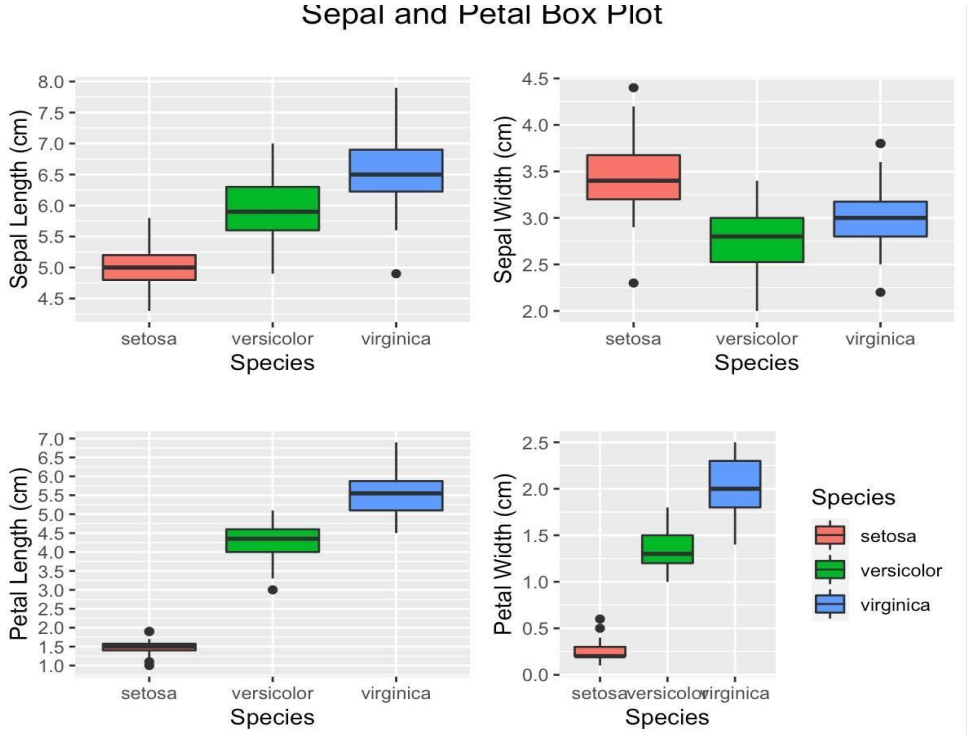
BpPl <- ggplot(iris, aes(Species, Petal.Length, fill=Species)) + geom_boxplot()+ scale_y_continuous("Petal Length (cm)", breaks= seq(0,30, by=.5))+ theme(legend.position="none")



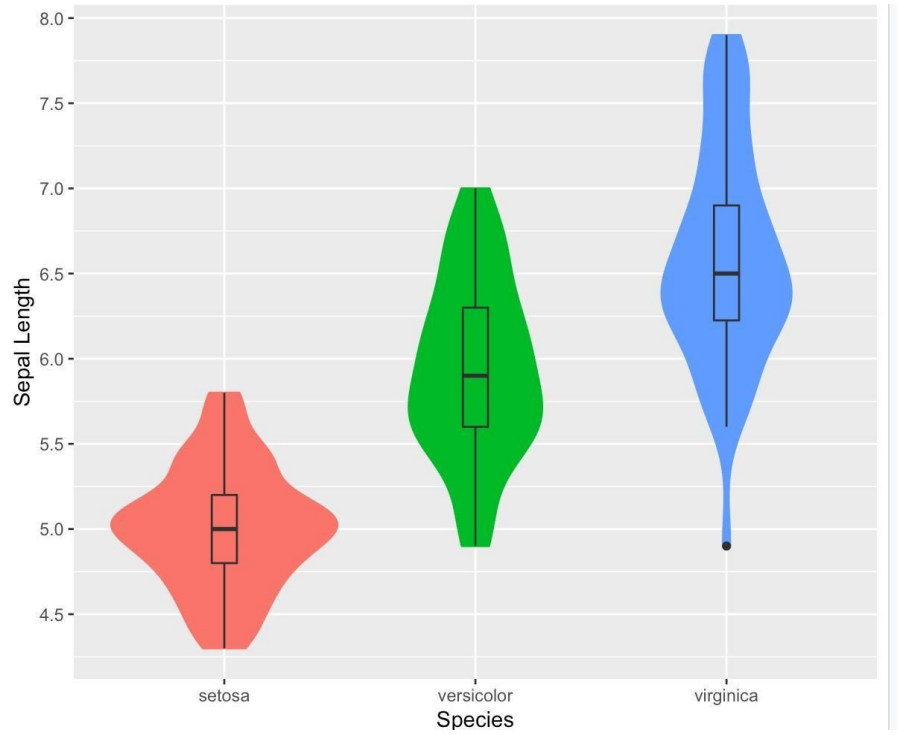

```
BpPw <-  ggplot(iris, aes(Species, Petal.Width, fill=Species)) +    geom_boxplot()+
  scale_y_continuous("Petal Width (cm)", breaks= seq(0,30, by=.5))+   labs(title
= "Iris Box Plot", x = "Species")
```



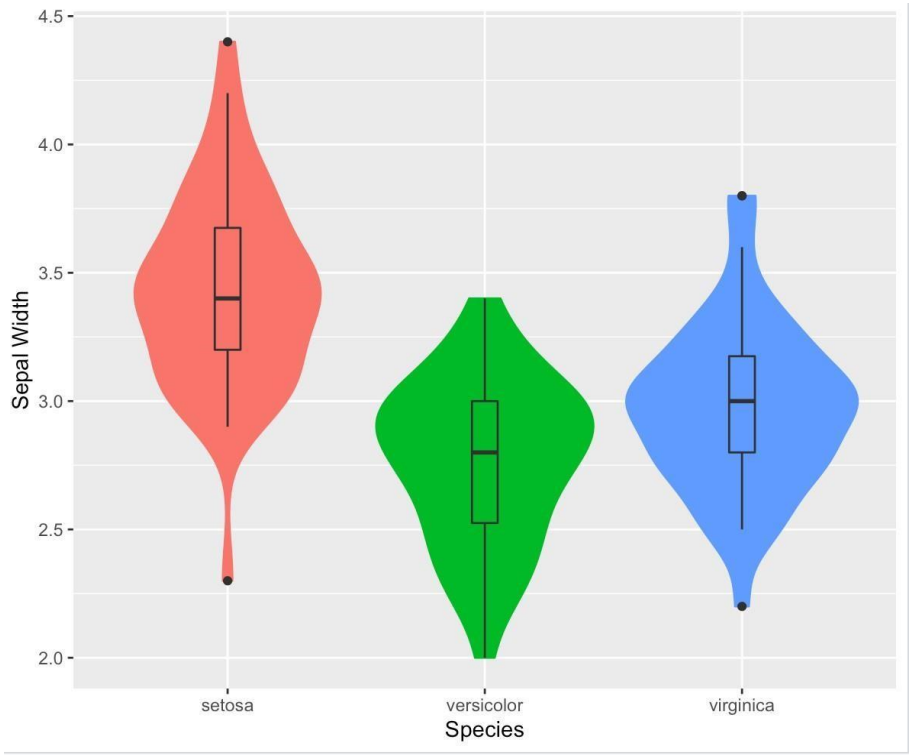

```
grid.arrange(BpSl  + ggtitle(""),
          BpSw  + ggtitle(""),
          BpPl + ggtitle(""),
BpPw + ggtitle(""),            nrow
= 2,
          top = textGrob("Sepal and Petal Box Plot",                    gp=gpar(fontsize=15))
          )
```



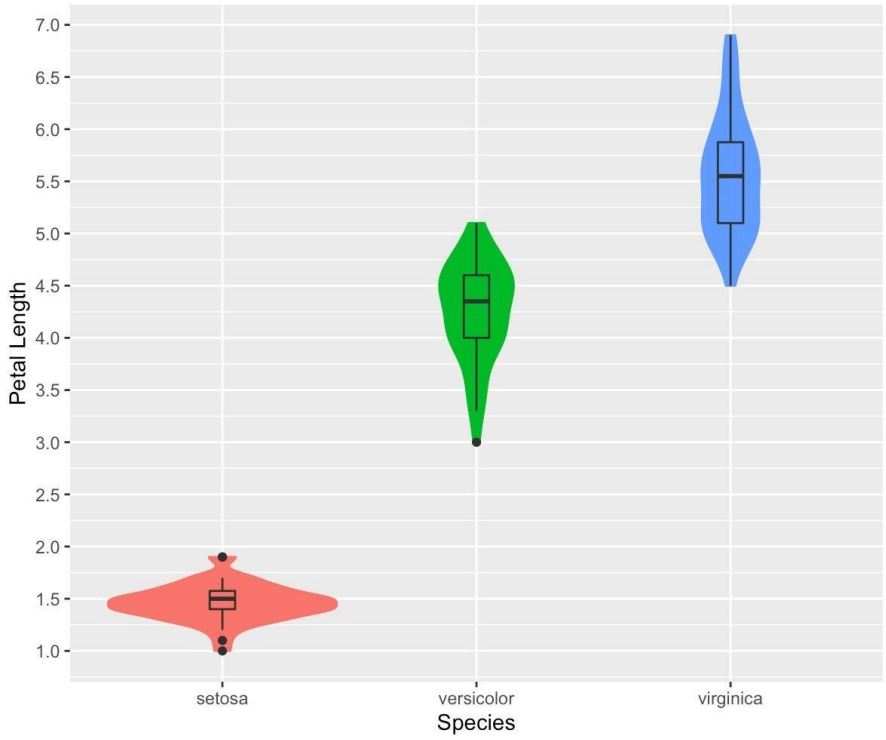
VpSl <- ggplot(iris, aes(Species, Sepal.Length, fill=Species)) + geom_violin(aes(color = Species), trim = T)+ scale_y_continuous("Sepal Length", breaks= seq(0,30, by=.5))+ geom_boxplot(width=0.1)+ theme(legend.position="none")



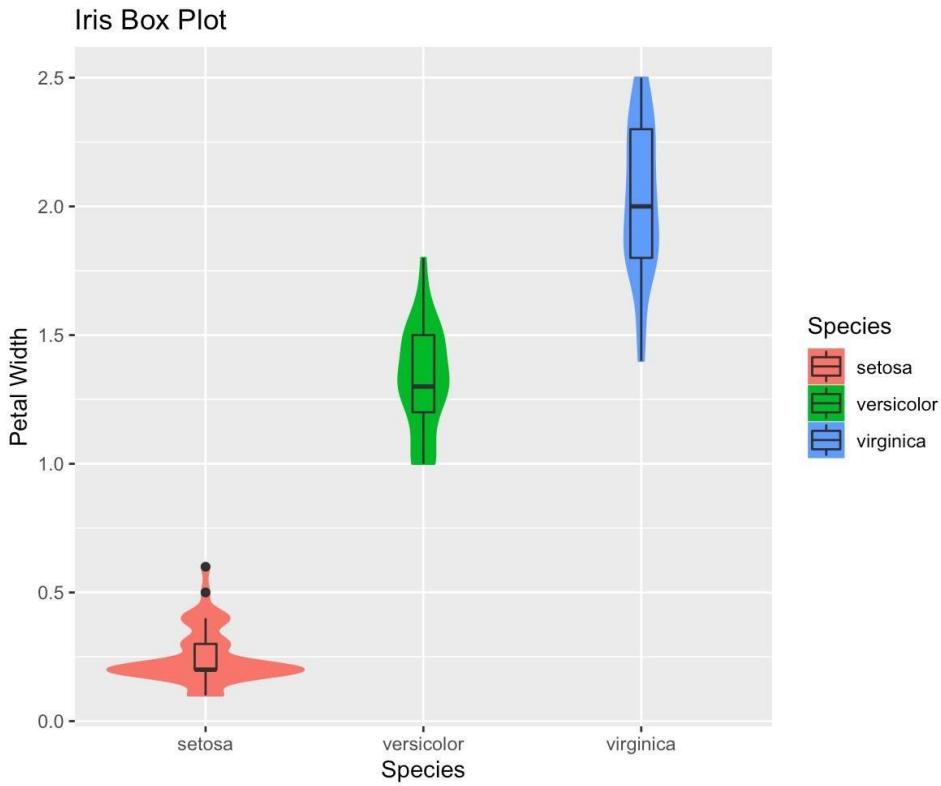
VpSw <- ggplot(iris, aes(Species, Sepal.Width, fill=Species)) + geom_violin(aes(color = Species), trim = T)+ scale_y_continuous("Sepal Width", breaks= seq(0,30, by=.5))+ geom_boxplot(width=0.1)+ theme(legend.position="none")




```
VpPl <-  ggplot(iris, aes(Species, Petal.Length, fill=Species)) +
geom_violin(aes(color = Species), trim = T)+
scale_y_continuous("Petal Length", breaks= seq(0,30, by=.5))+
geom_boxplot(width=0.1)+   theme(legend.position="none")
```

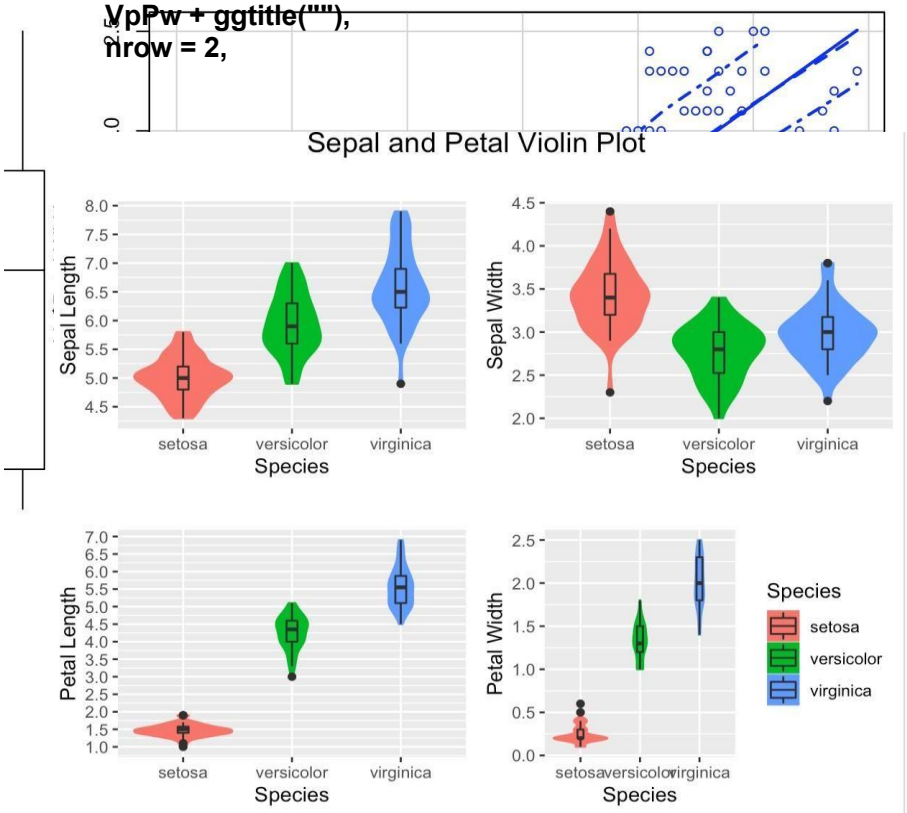



```
VpPw <-  ggplot(iris, aes(Species, Petal.Width, fill=Species)) +
geom_violin(aes(color = Species), trim = T)+
scale_y_continuous("Petal Width", breaks= seq(0,30, by=.5))+
geom_boxplot(width=0.1)+   labs(title = "Iris Box Plot", x = "Species")
```



grid.arrange(VpSl + ggtitle(""), VpSw + ggtitle(""),

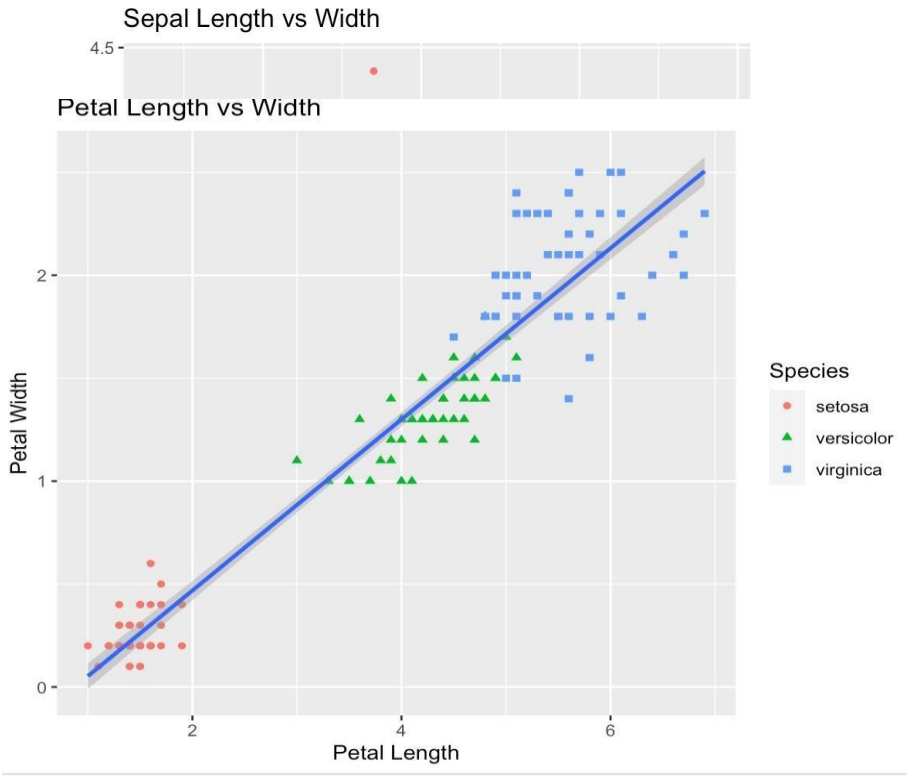
VpPw + ggtitle(""),
nrow = 2,



scatterplot(iris$Petal.Length,iris$Petal.Width)

VpPl + ggtitle(""),
ggplot(data=iris, aes(x = Sepal.Length, y = Sepal.Width)) +
geom_point(aes(color=Species, shape=Species)) +
xlab("Sepal Length") + ylab("Sepal Width") + ggtitle("Sepal Length vs Width") ggplot(data = iris, aes(x =

Petal.Length, y = Petal.Width))+

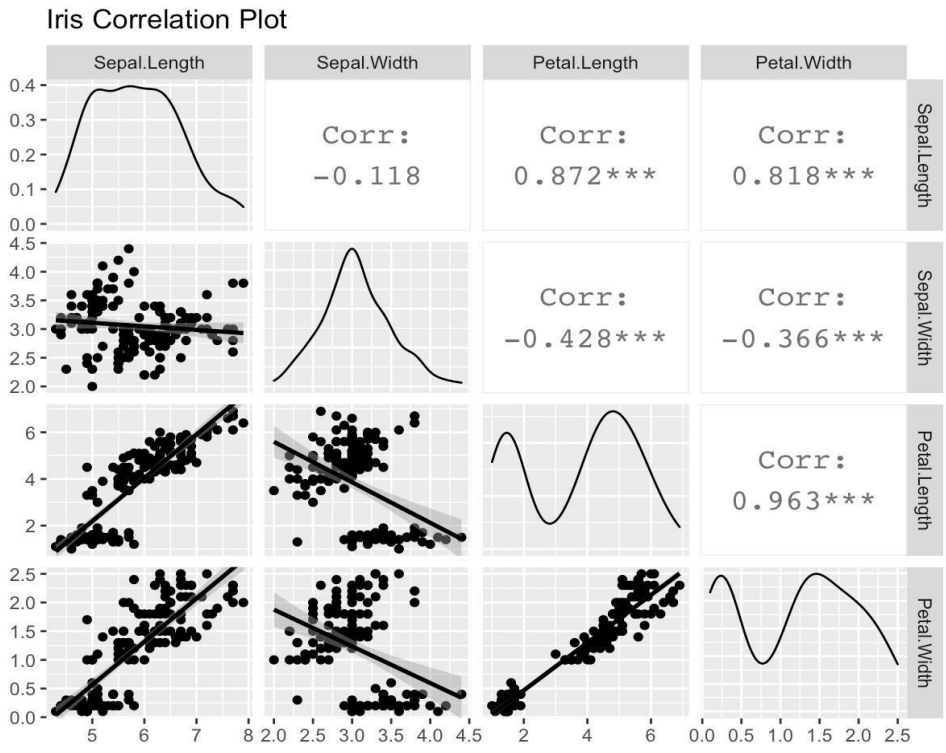


xlab("Petal Length")+

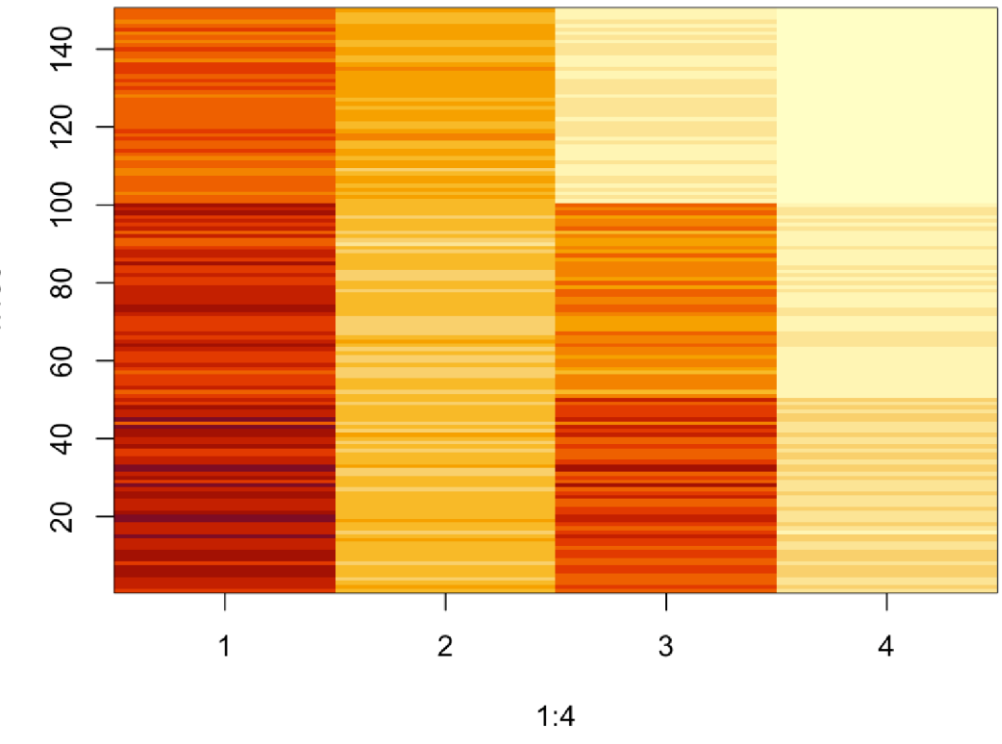
ggpairs(data = iris[1:4], title = "Iris Correlation Plot",

upper = list(continuous = wrap("cor", size = 5)), lower = list(continuous = "smooth")

)



irisMatix <- as.matrix(iris[1:150, 1:4])
irisTransposedMatrix <- t(irisMatix)[,nrow(irisMatix):1] image(1:4, 1:150, irisTransposedMatrix)



3. 결론

점 그래프에서 데이터의 전체적 분포를 확인할 수 있었으며, histogram 과 물결그래프에 서는 각 자료의 집합적인 분포를 볼 수 있었다 boxplot 과 vp 그래프로는 각 칼럼별 통계적 수치를 시각적으로 확인할 수 있었다. 또한 car 패키지의 활용은 칼럼간 비교에 유용하다는 점을 알 수 있었다. 시각화 분석 결과, 각 칼럼별 데이터 집도가 높으며, 크거나 작은 수치로 데이터가 집 경향을 보 으며 이를 토대로, 칼럼 내 데이터 간의 관계를 수치적으로 확인하기 위해서 anova 분석 등의 분산 분석법 사용이 유용할 것이라는 결론을 내릴 수 있었다.