lista 2.

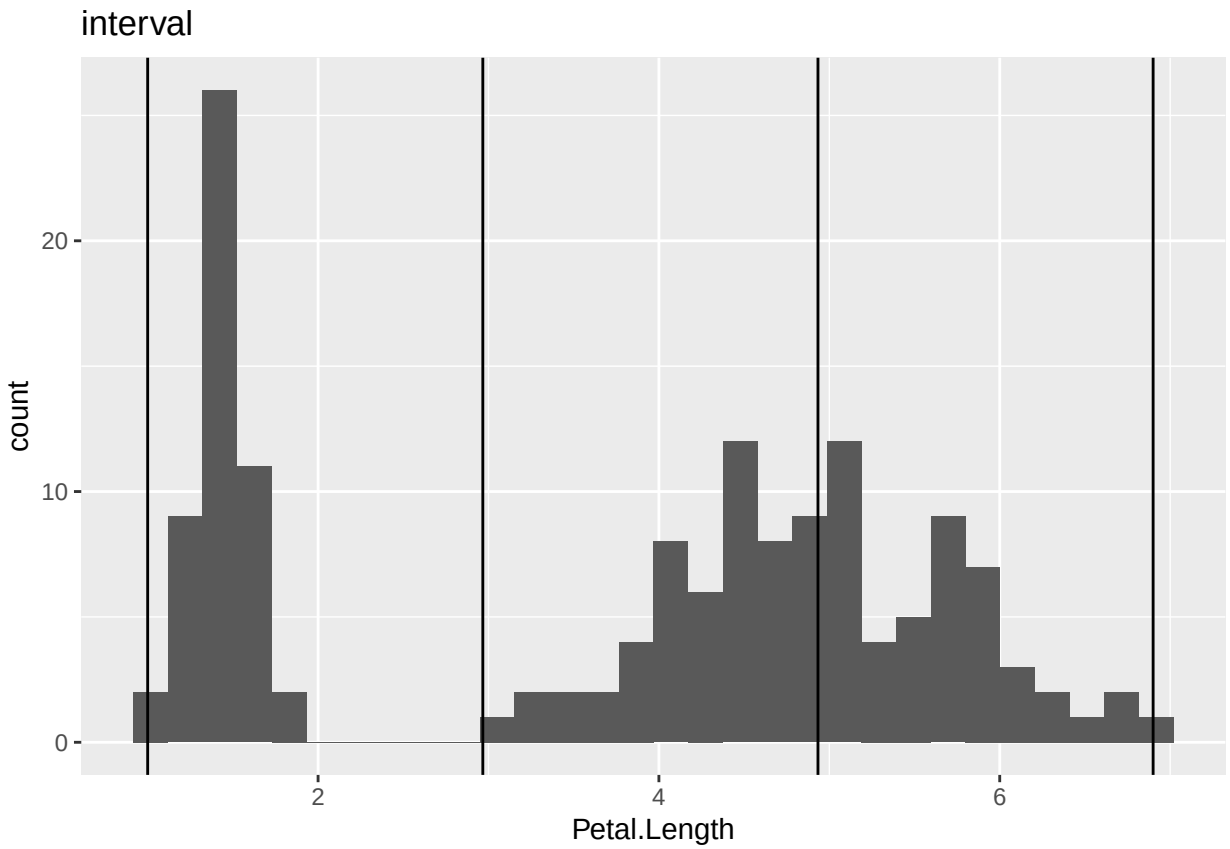
Mikołaj Langner, Marcin Kostrzewa

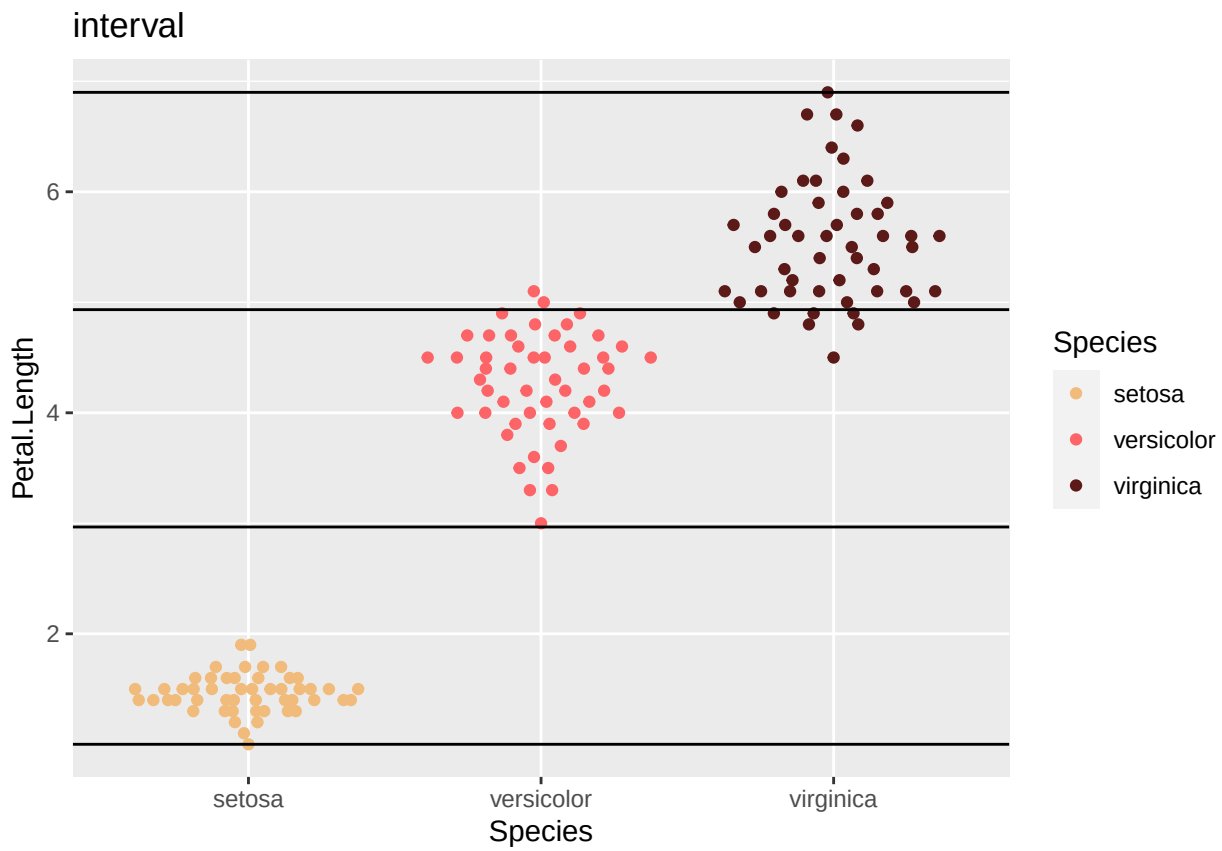
25.04.2021

ZADANIE 1.

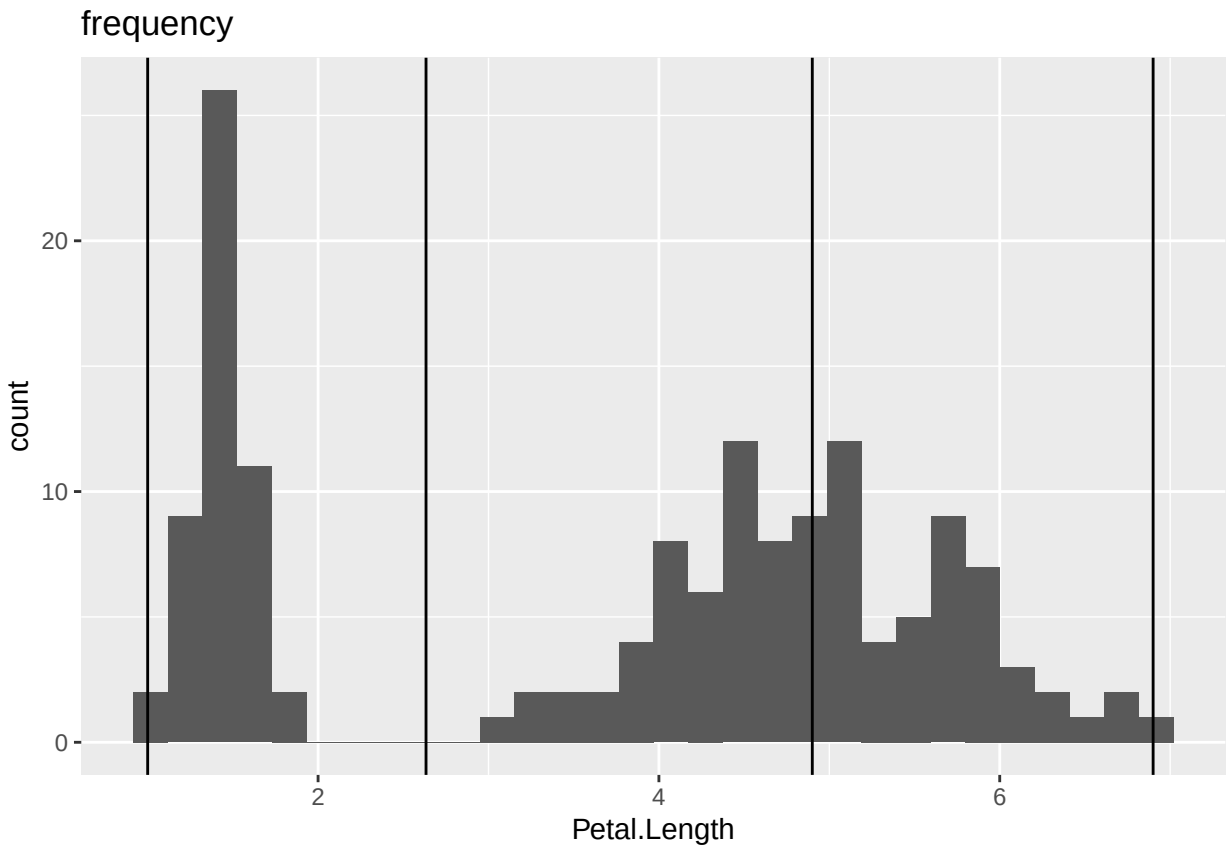
```
data(iris)
```

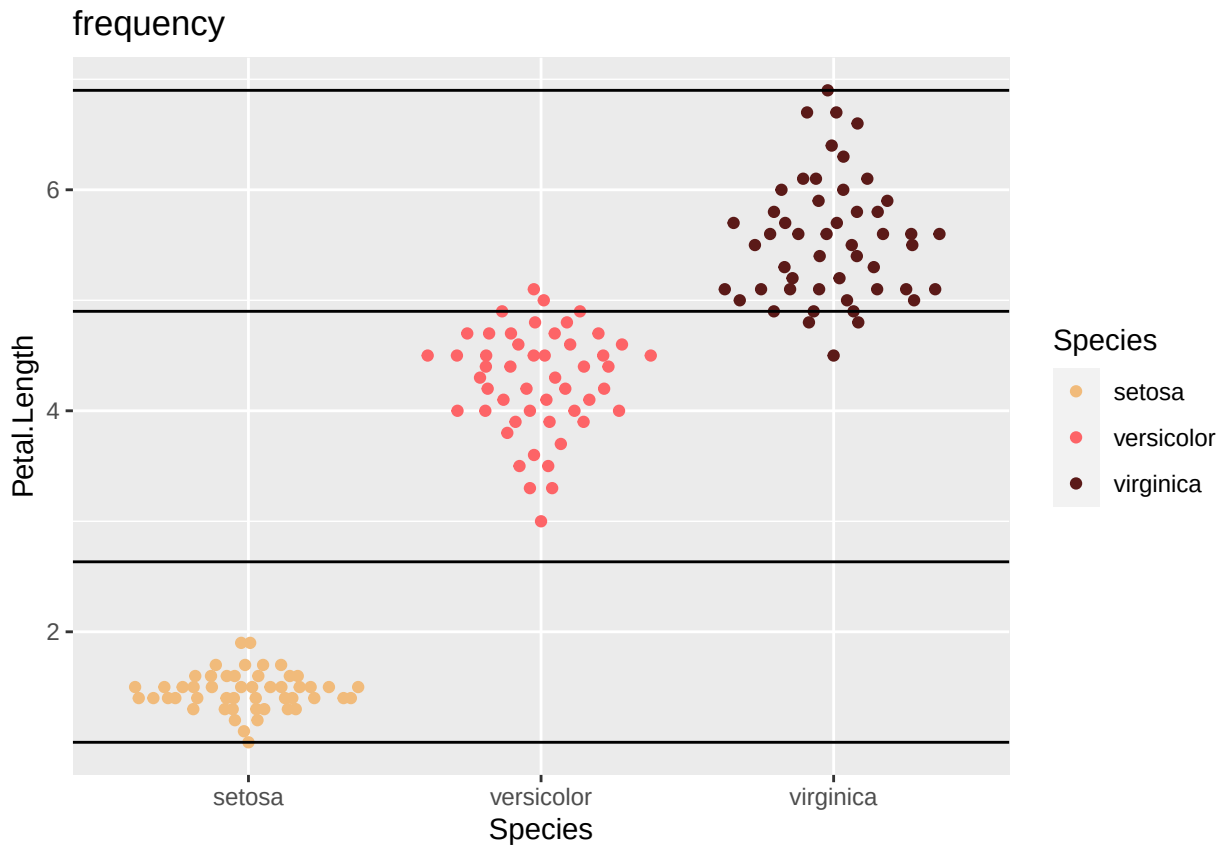
```
intervals <- c(min(iris$Petal.Length), 2, 5, max(iris$Petal.Length))
for (method in c("interval", "frequency", "cluster", "fixed")) {
  petal.length.discretized <- if (method != "fixed")
    discretize(iris$Petal.Length, method=method) else
    discretize(iris$Petal.Length, method=method, breaks=intervals)
  print(ggplot(iris, aes(Petal.Length)) +
          geom_histogram() +
          geom_vline(xintercept=attributes(petal.length.discretized)$"discretized:breaks") +
          ggtitle(method))
  print(ggplot(iris, aes(Species, Petal.Length)) +
    geom_quasirandom(aes(col=Species)) +
    scale_color_manual(values=wes_palette("GrandBudapest1", 3)) +
    geom_hline(yintercept=attributes(petal.length.discretized)$"discretized:breaks") +
    ggtitle(method))
  discretized.table <- table(petal.length.discretized, iris$Species)
  matchClasses(discretized.table)
}
```


interval
count
20
10
0
2
4
6
Petal.Length

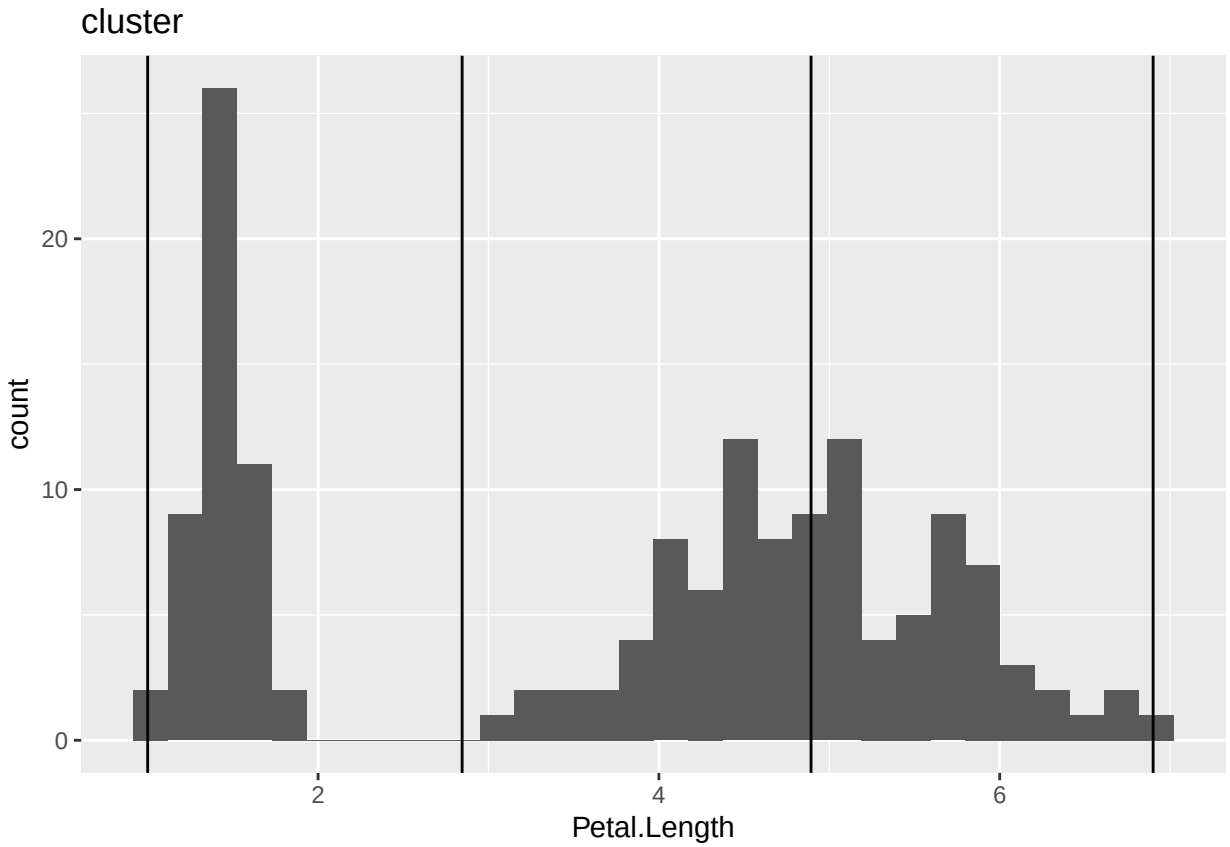


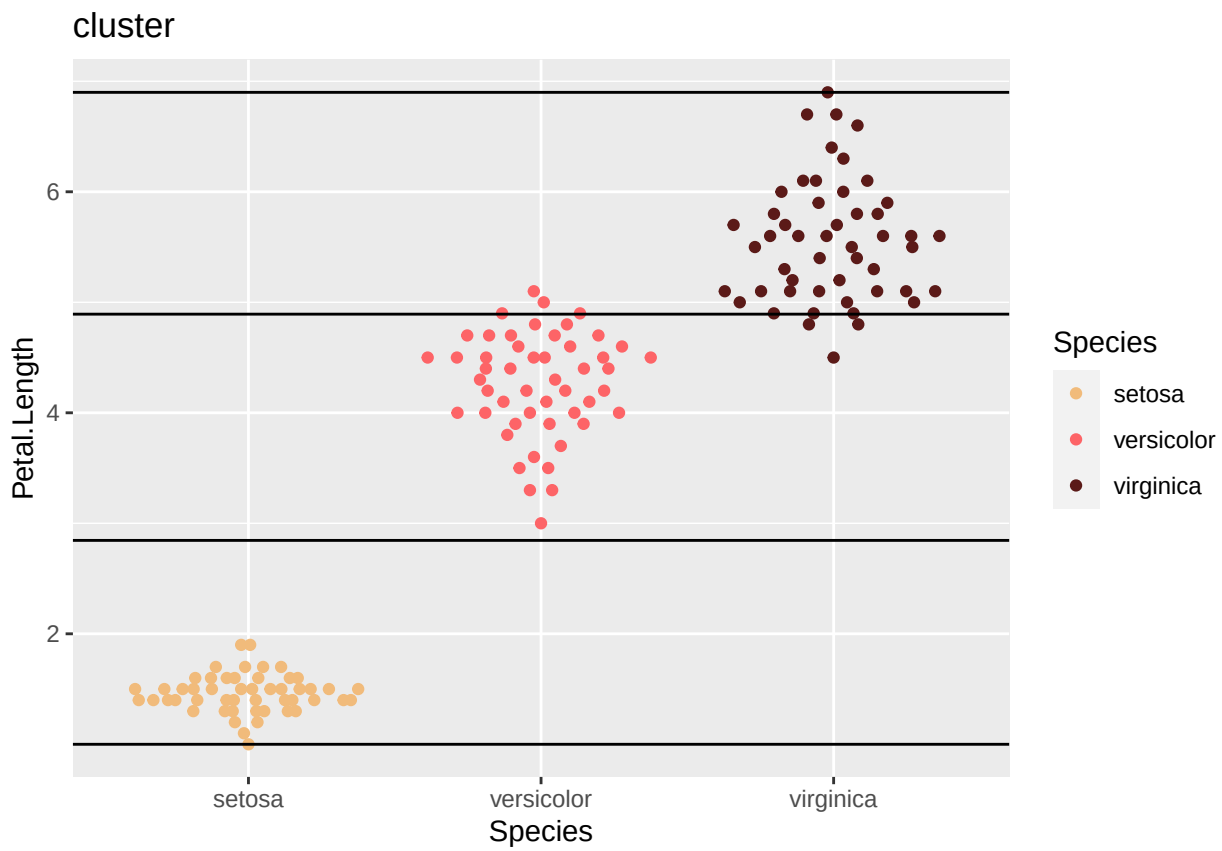

```
## Cases in matched pairs: 94.67 %
```


frequency
count
20
10
0
2
4
6
Petal.Length

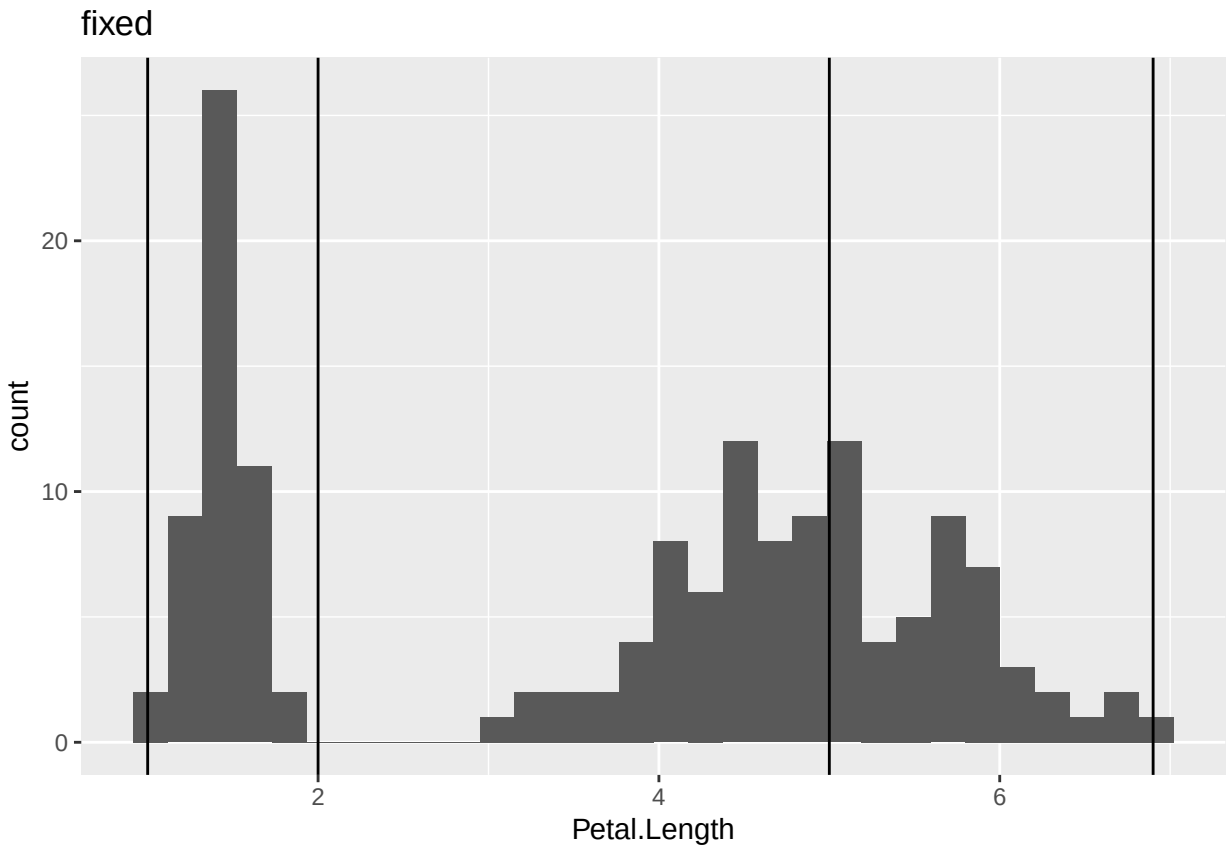


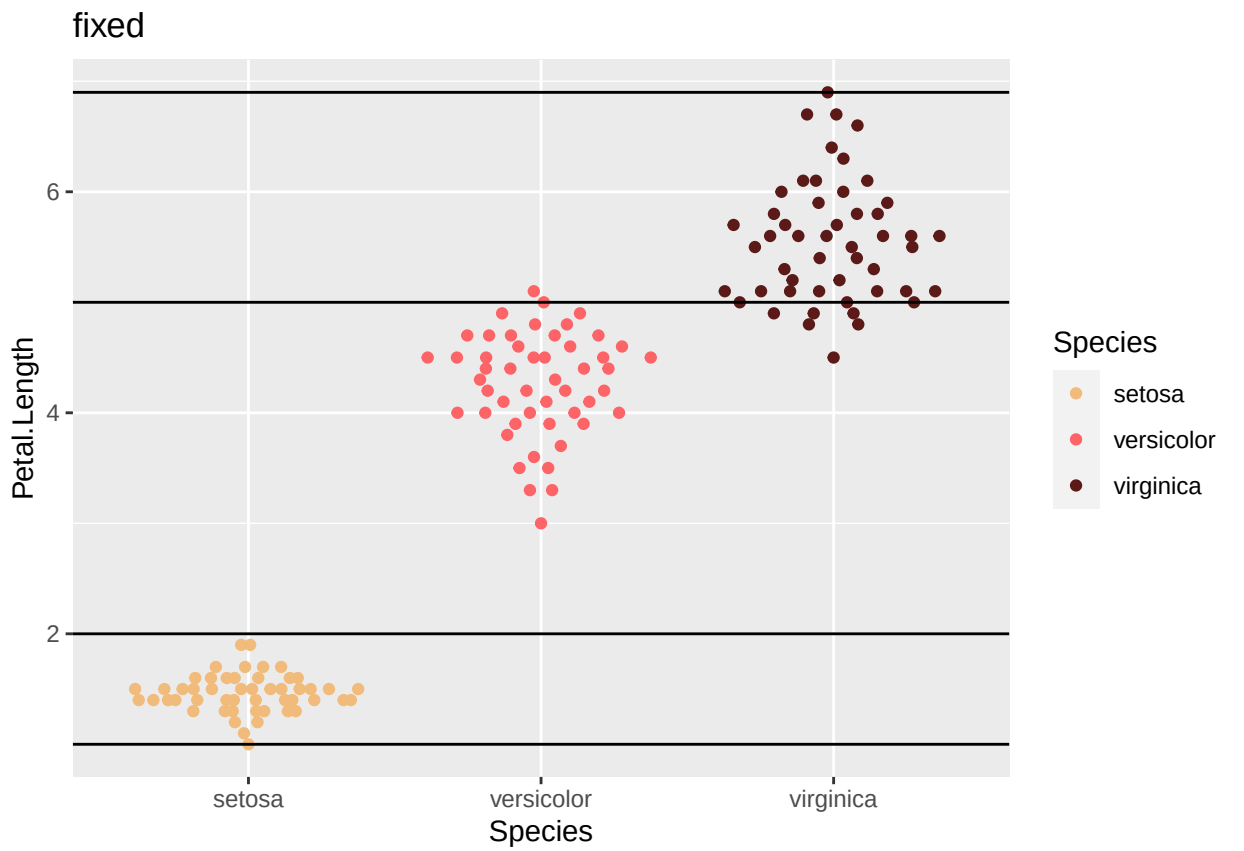

```
## Cases in matched pairs: 95.33 %
```


cluster
count
20
10
0
2
4
6
Petal.Length



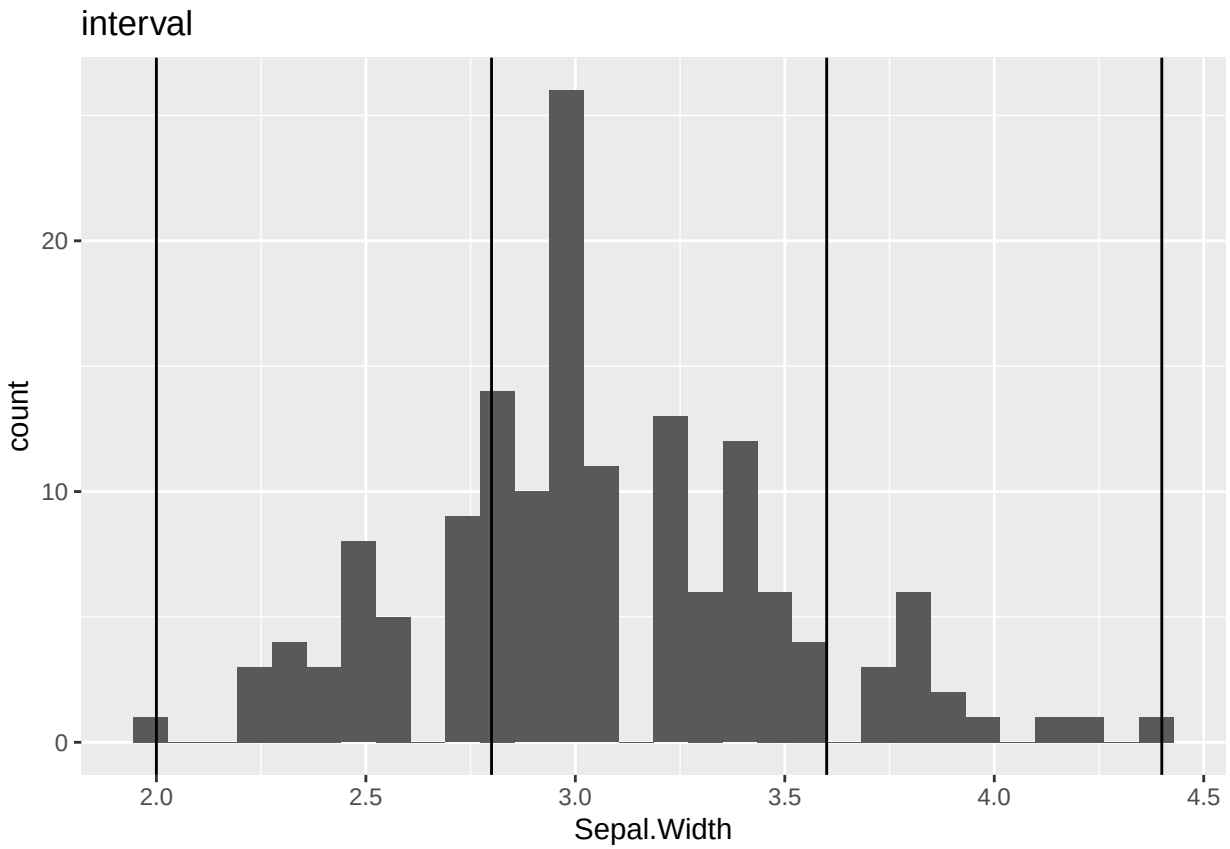

```
## Cases in matched pairs: 95.33 %
```

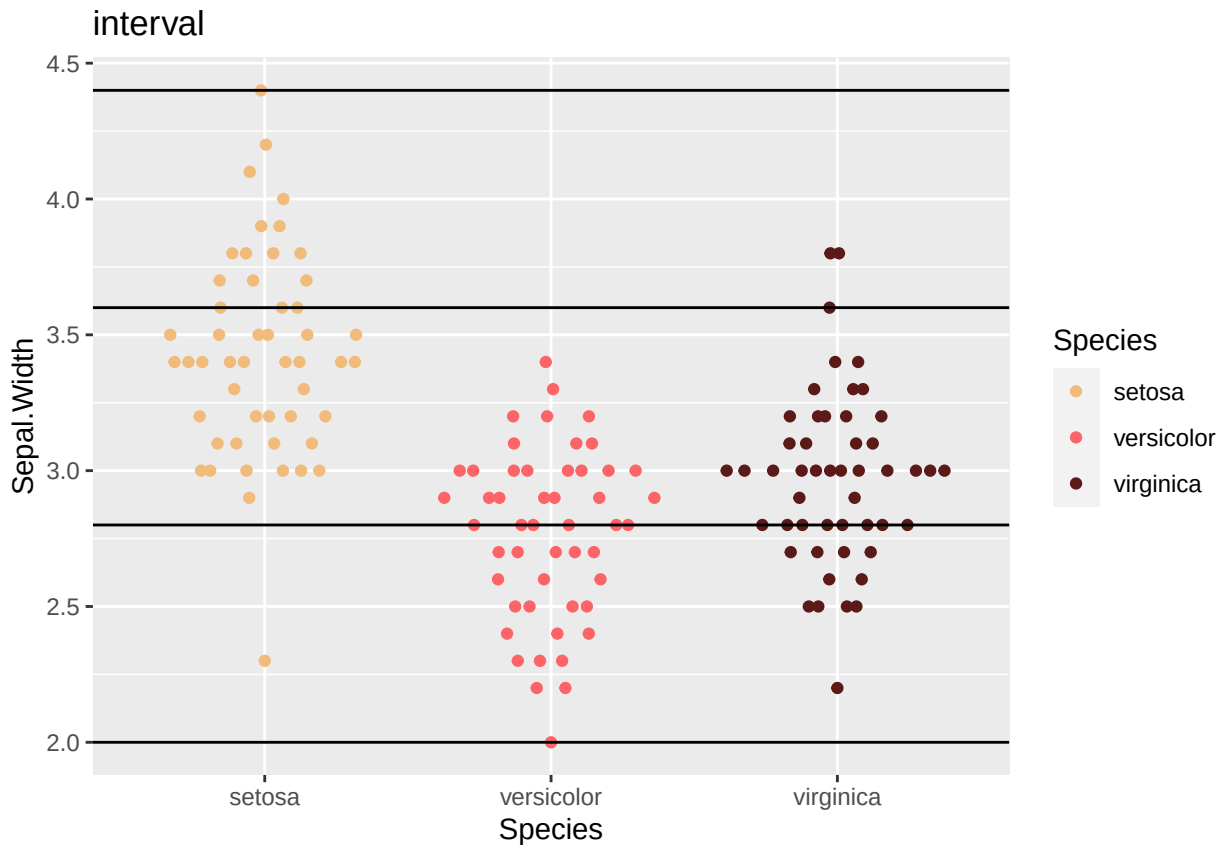

fixed
count
20
10
0
2
4
6
Petal.Length



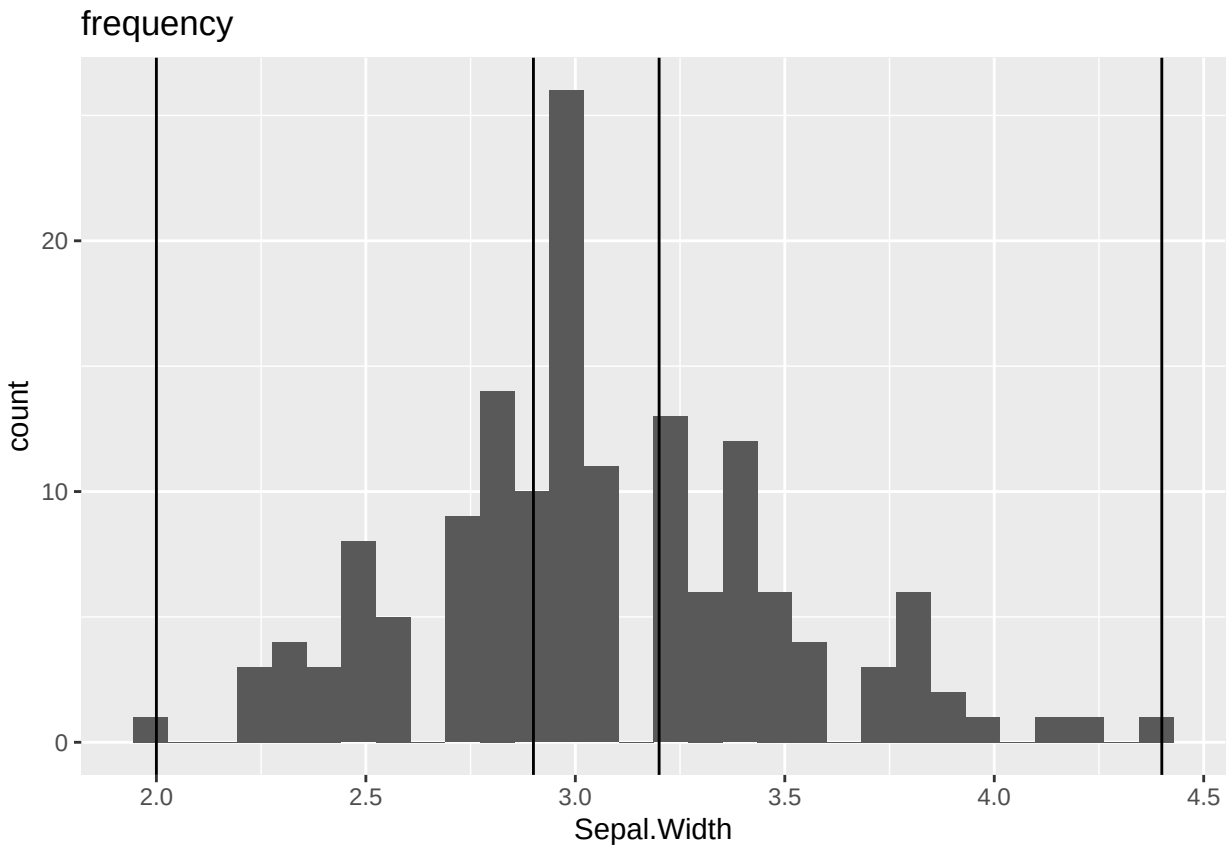

```
## Cases in matched pairs: 94.67 %
```

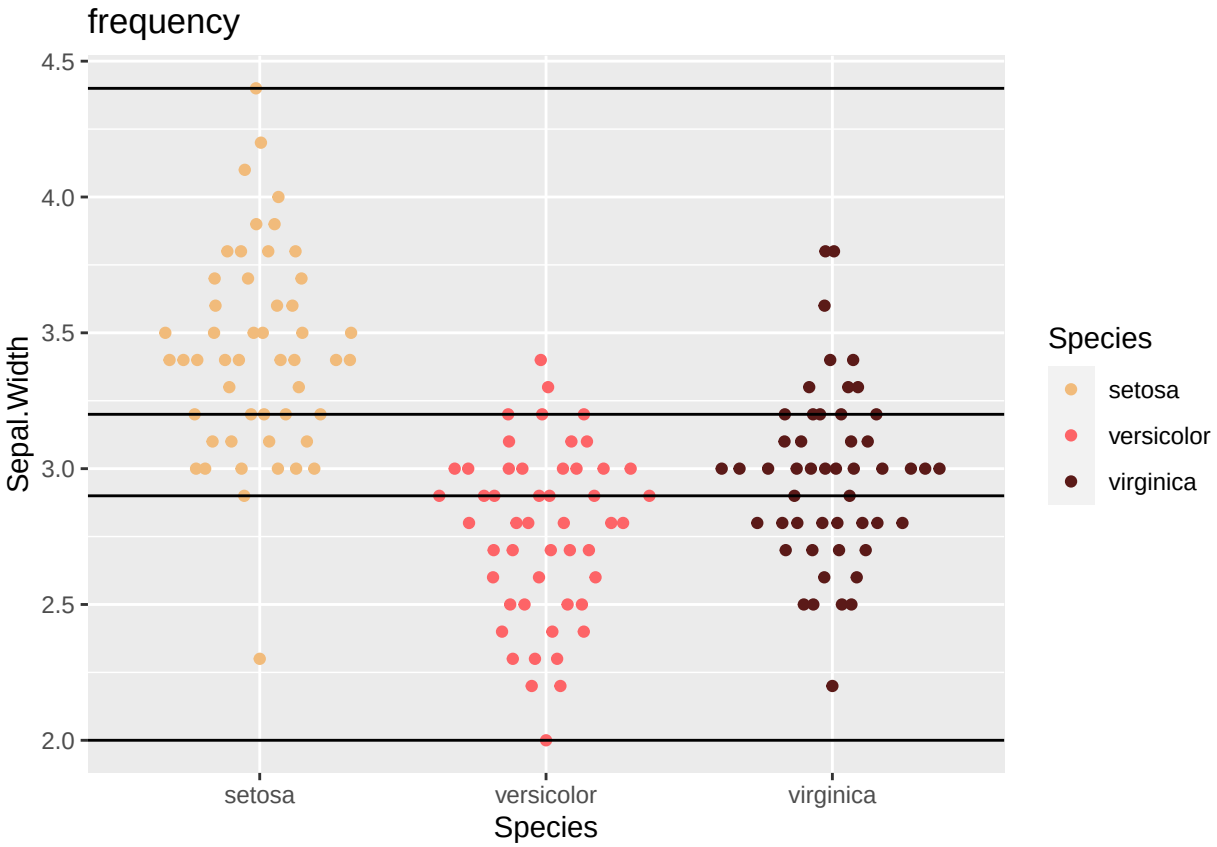
```
intervals <- c(min(iris$Sepal.Width), 2.5, 3, max(iris$Sepal.Width))
for (method in c("interval", "frequency", "cluster", "fixed")) {
  sepal.width.discretized <- if (method != "fixed")
    discretize(iris$Sepal.Width, method=method) else
    discretize(iris$Sepal.Width, method=method, breaks=intervals)
  print(ggplot(iris, aes(Sepal.Width)) +
          geom_histogram() +
          geom_vline(xintercept=attributes(sepal.width.discretized)$"discretized:breaks") +
          ggtitle(method))
  print(ggplot(iris, aes(Species, Sepal.Width)) +
    geom_quasirandom(aes(col=Species)) +
    scale_color_manual(values=wes_palette("GrandBudapest1", 3)) +
    geom_hline(yintercept=attributes(sepal.width.discretized)$"discretized:breaks") +
    ggtitle(method))
  discretized.table <- table(sepal.width.discretized, iris$Species)
  matchClasses(discretized.table)
}
```


interval
count
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5
Sepal.Width

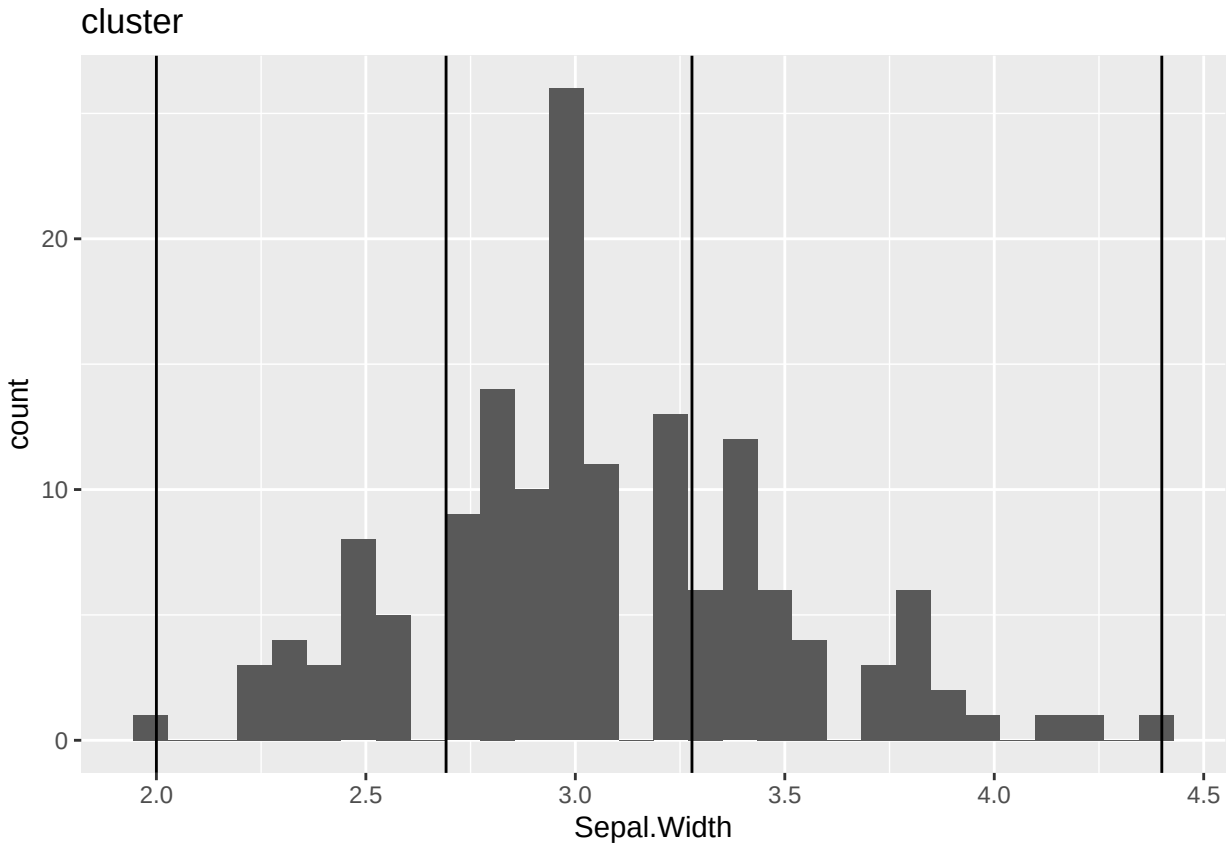


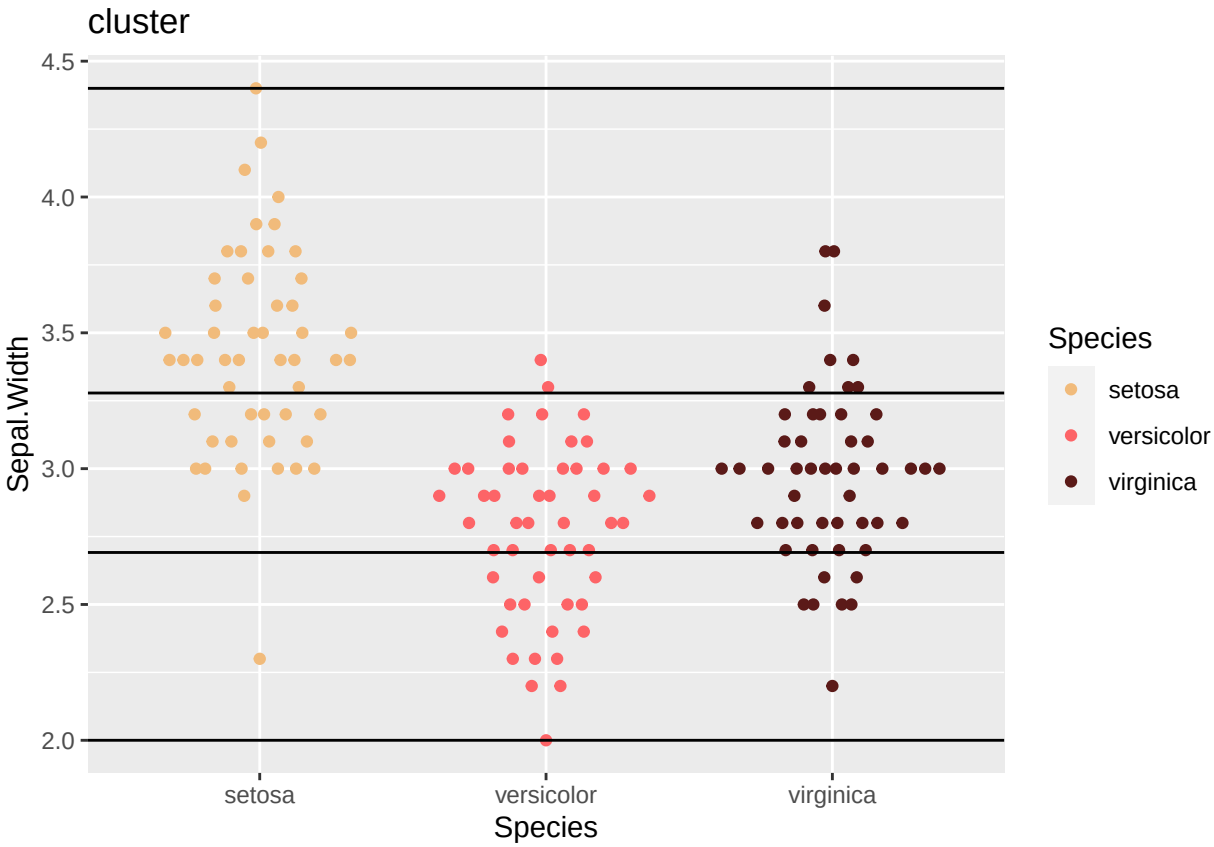

```
## Cases in matched pairs: 50.67 %
```


frequency
count
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5
Sepal.Width

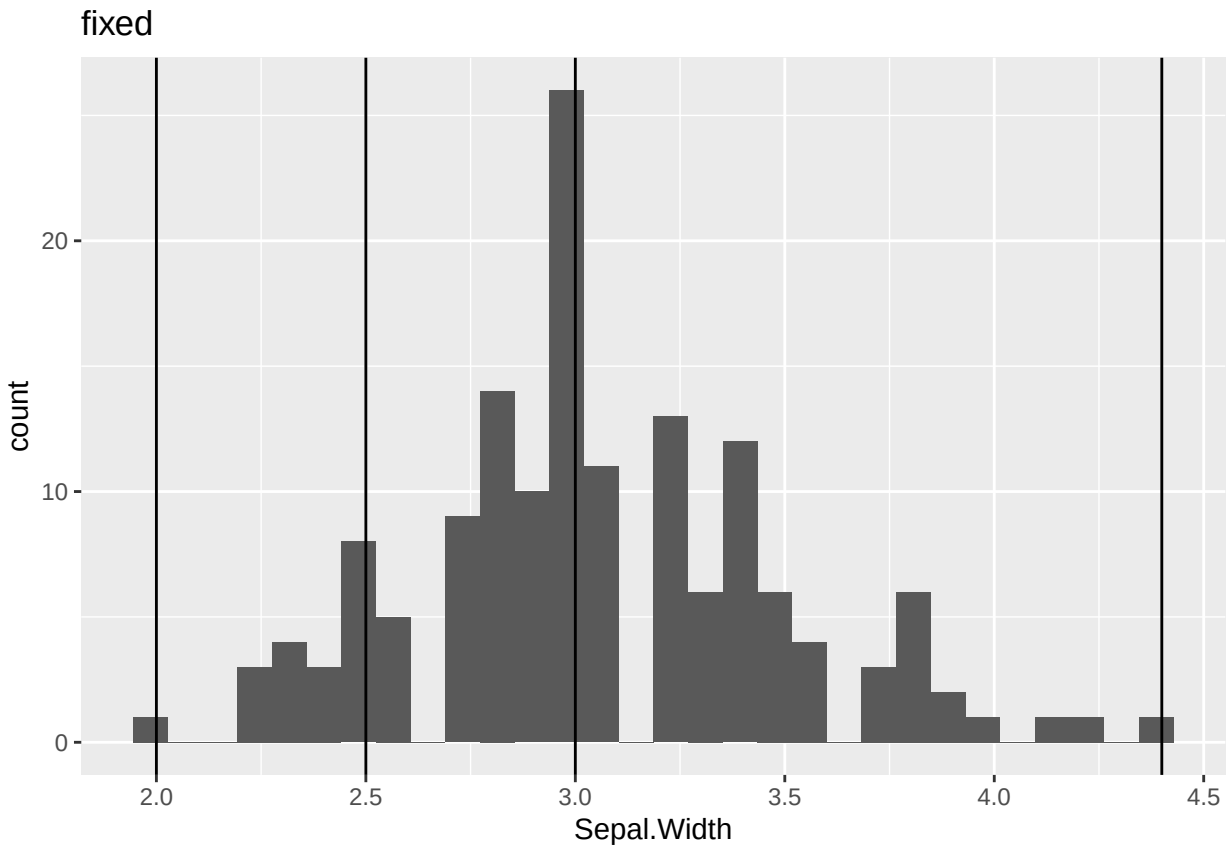


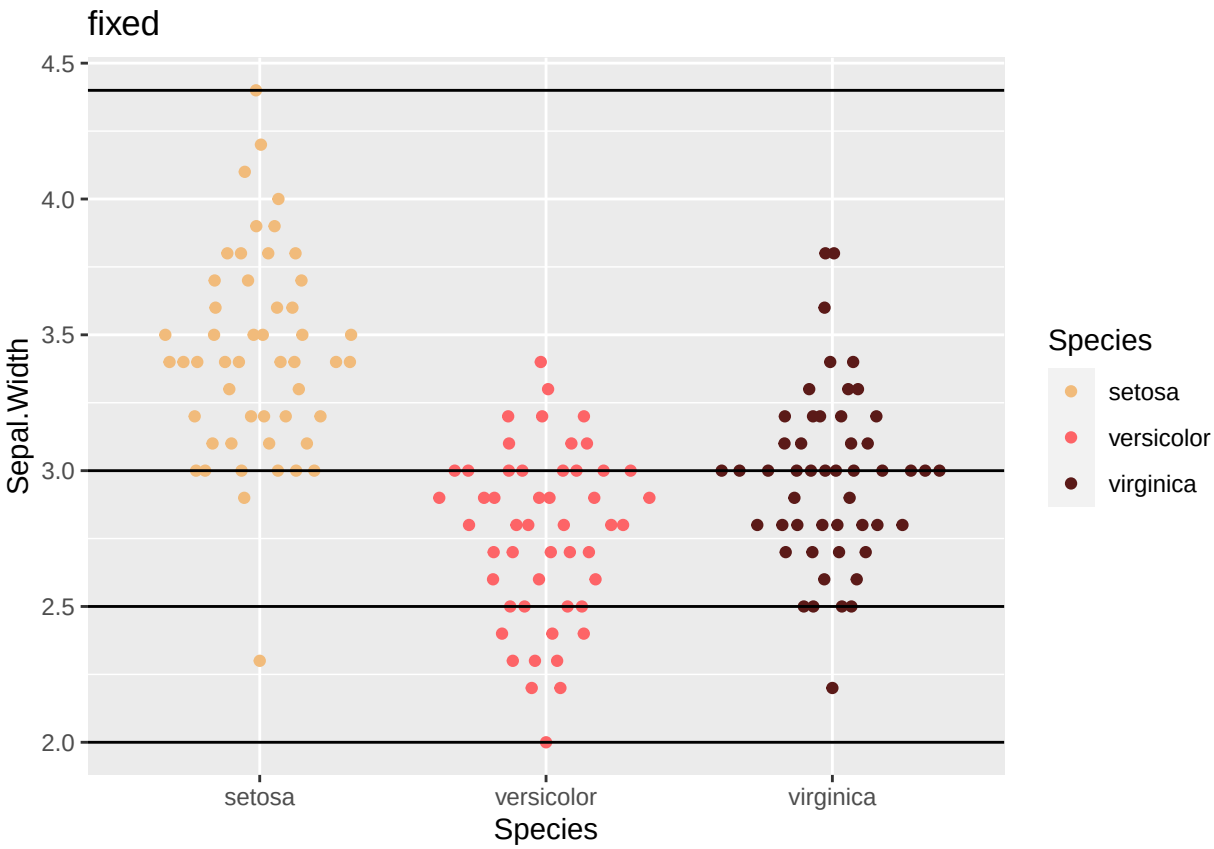

```
## Cases in matched pairs: 55.33 %
```


cluster
count
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5
Sepal.Width



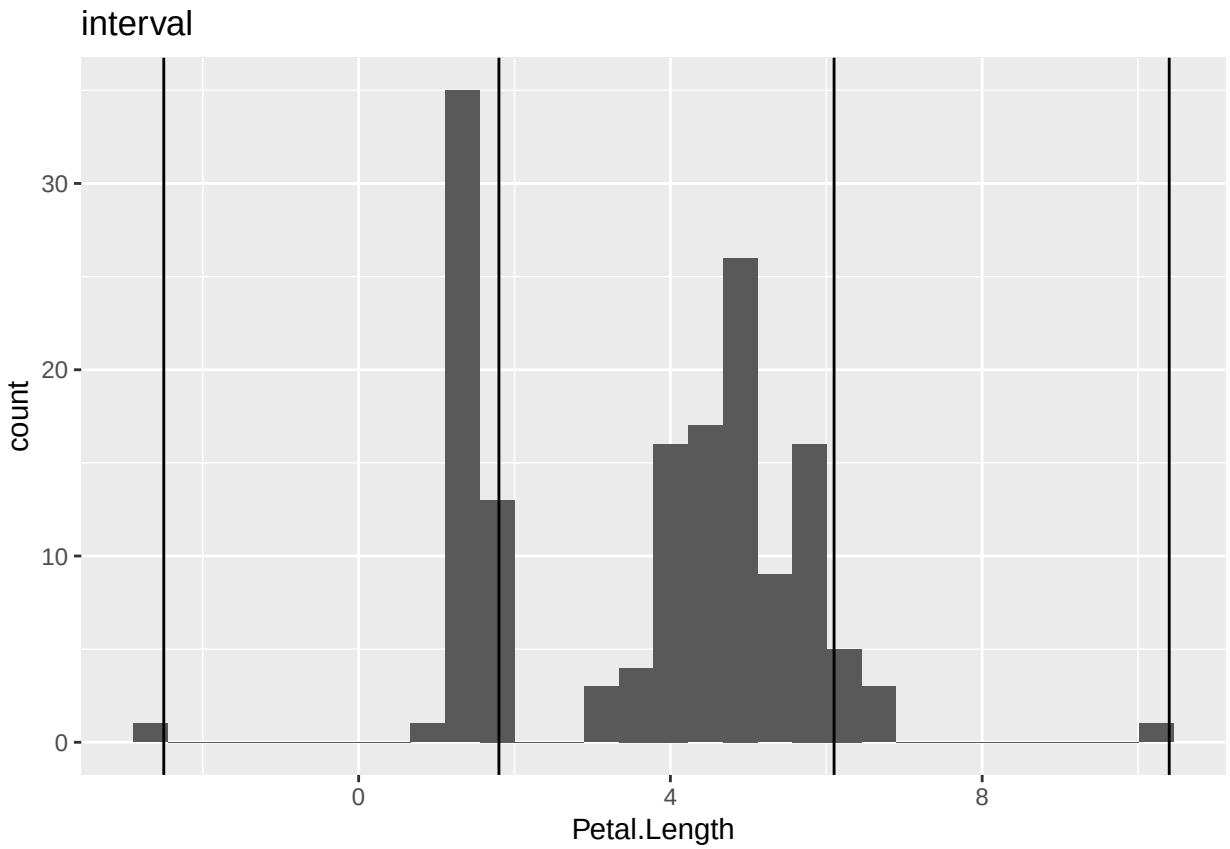

```
## Cases in matched pairs: 56 %
```

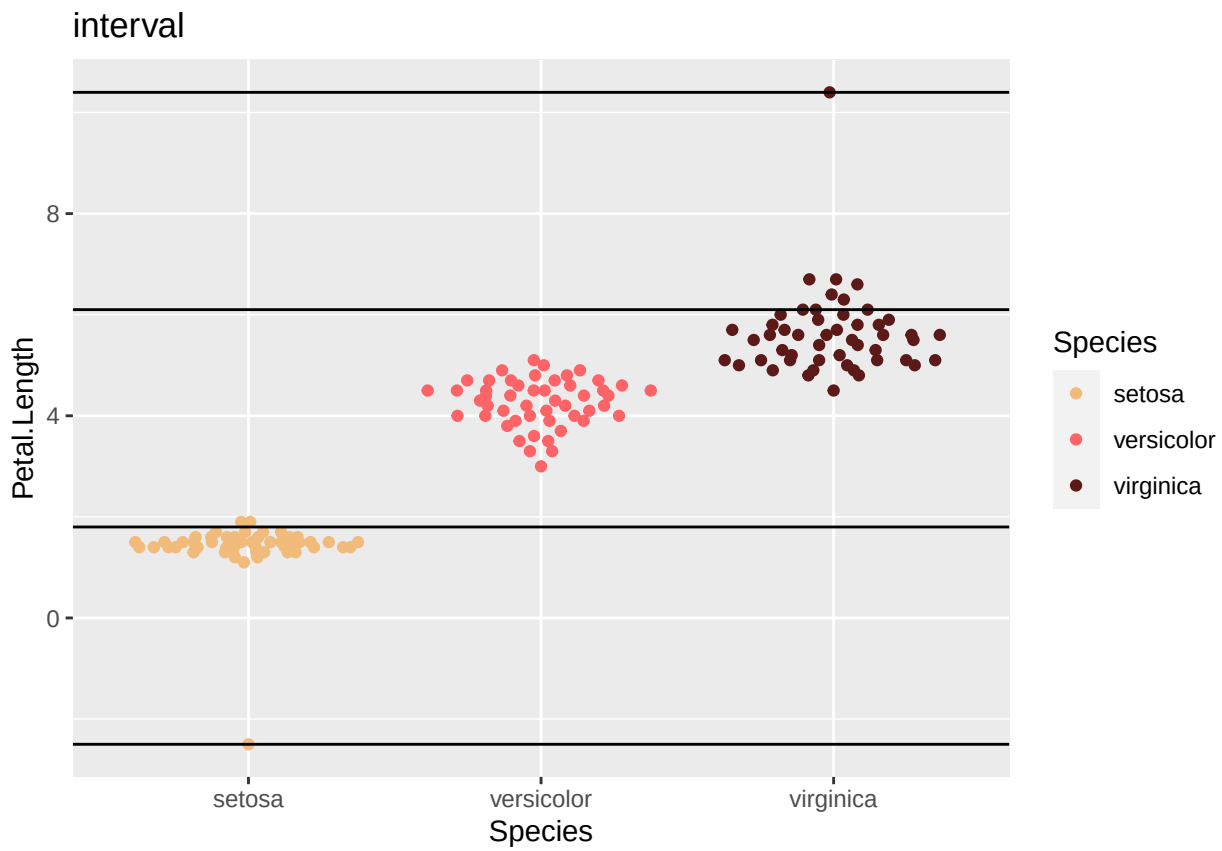

fixed
count
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5
Sepal.Width



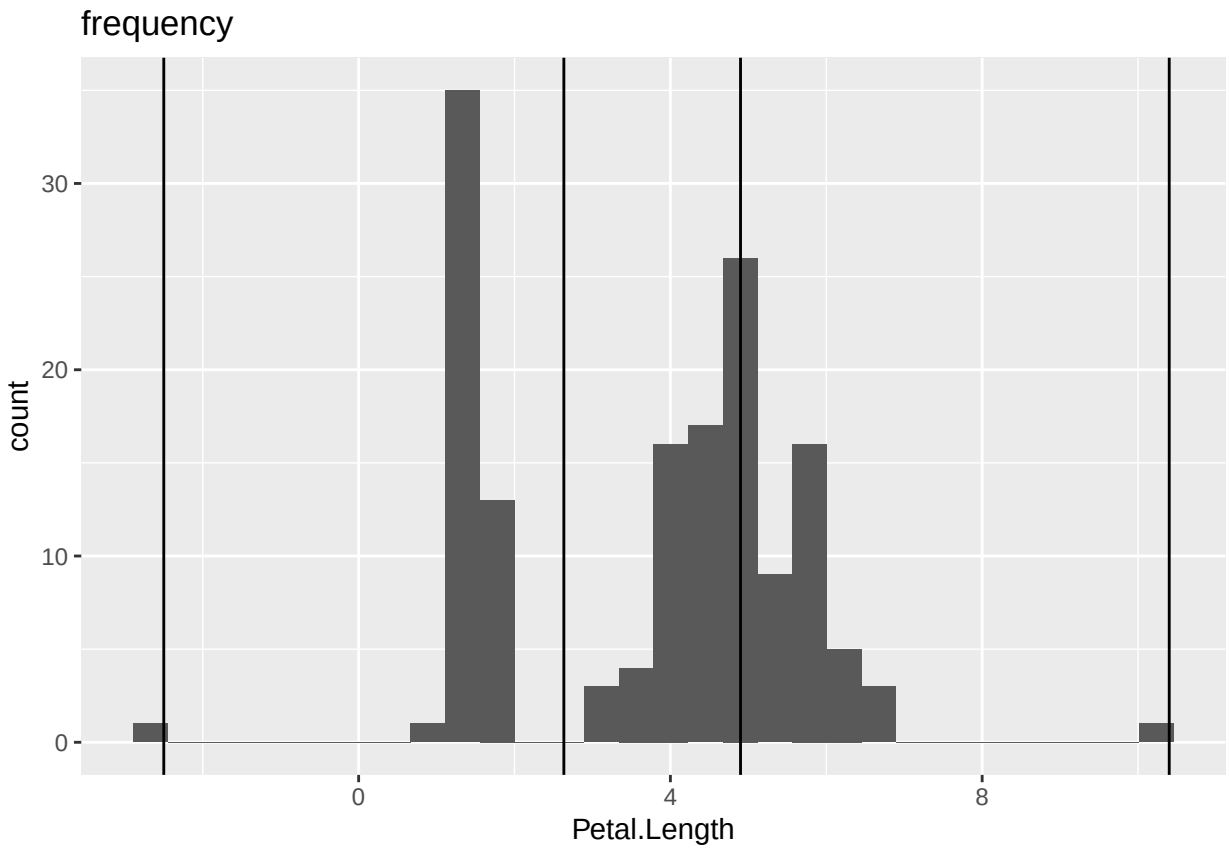

```
## Cases in matched pairs: 54.67 %
```

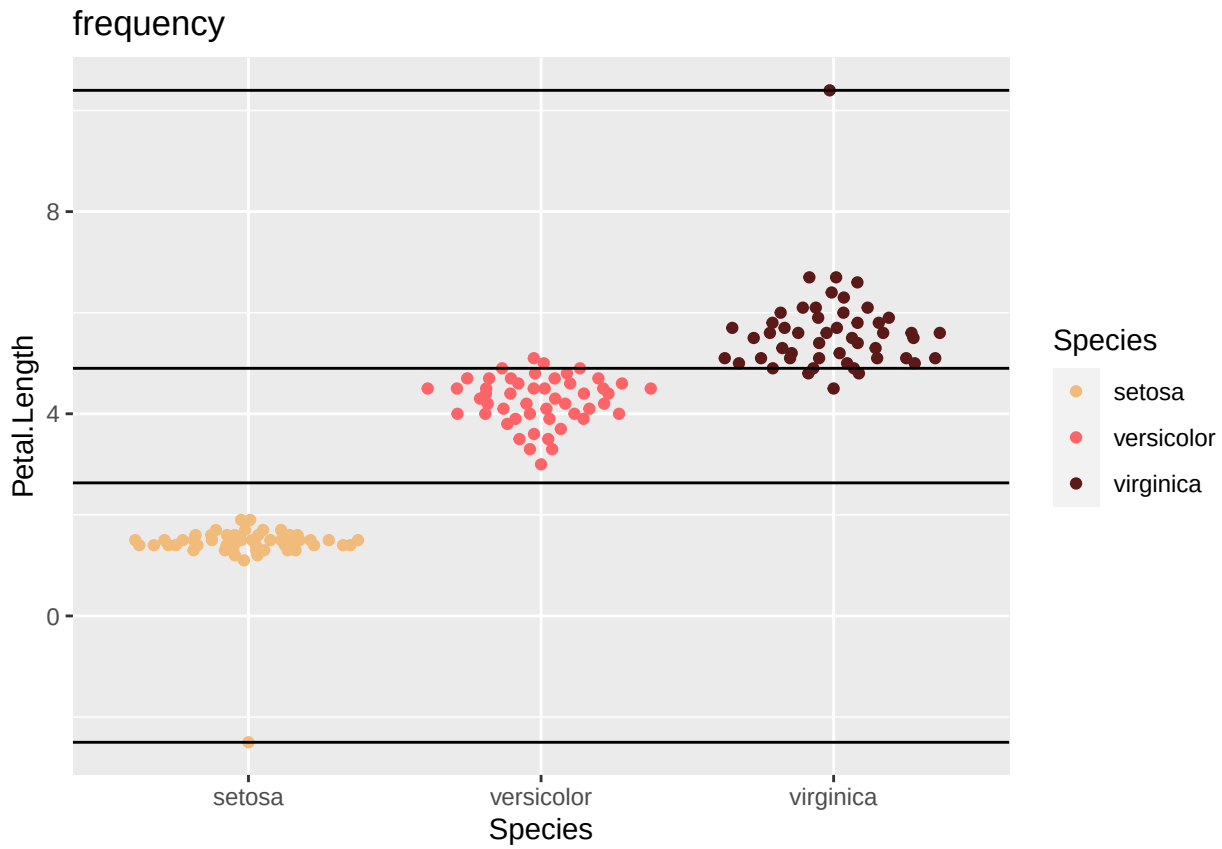
```
iris$Petal.Length[which.min(iris$Petal.Length)] <- min(iris$Petal.Length) - IQR(iris$Petal.Length)
iris$Petal.Length[which.max(iris$Petal.Length)] <- max(iris$Petal.Length) + IQR(iris$Petal.Length)
intervals <- c(min(iris$Petal.Length), 2, 5, max(iris$Petal.Length))
for (method in c("interval", "frequency", "cluster", "fixed")) {
  petal.length.discretized <- if (method != "fixed")
    discretize(iris$Petal.Length, method=method) else
    discretize(iris$Petal.Length, method=method, breaks=intervals)
  print(ggplot(iris, aes(Petal.Length)) +
          geom_histogram() +
          geom_vline(xintercept=attributes(petal.length.discretized)$"discretized:breaks") +
          ggtitle(method))
  print(ggplot(iris, aes(Species, Petal.Length)) +
    geom_quasirandom(aes(col=Species)) +
    scale_color_manual(values=wes_palette("GrandBudapest1", 3)) +
    geom_hline(yintercept=attributes(petal.length.discretized)$"discretized:breaks") +
    ggtitle(method))
  discretized.table <- table(petal.length.discretized, iris$Species)
  matchClasses(discretized.table)
}
```


interval
30
20
10
0
count
0
4
8
Petal.Length

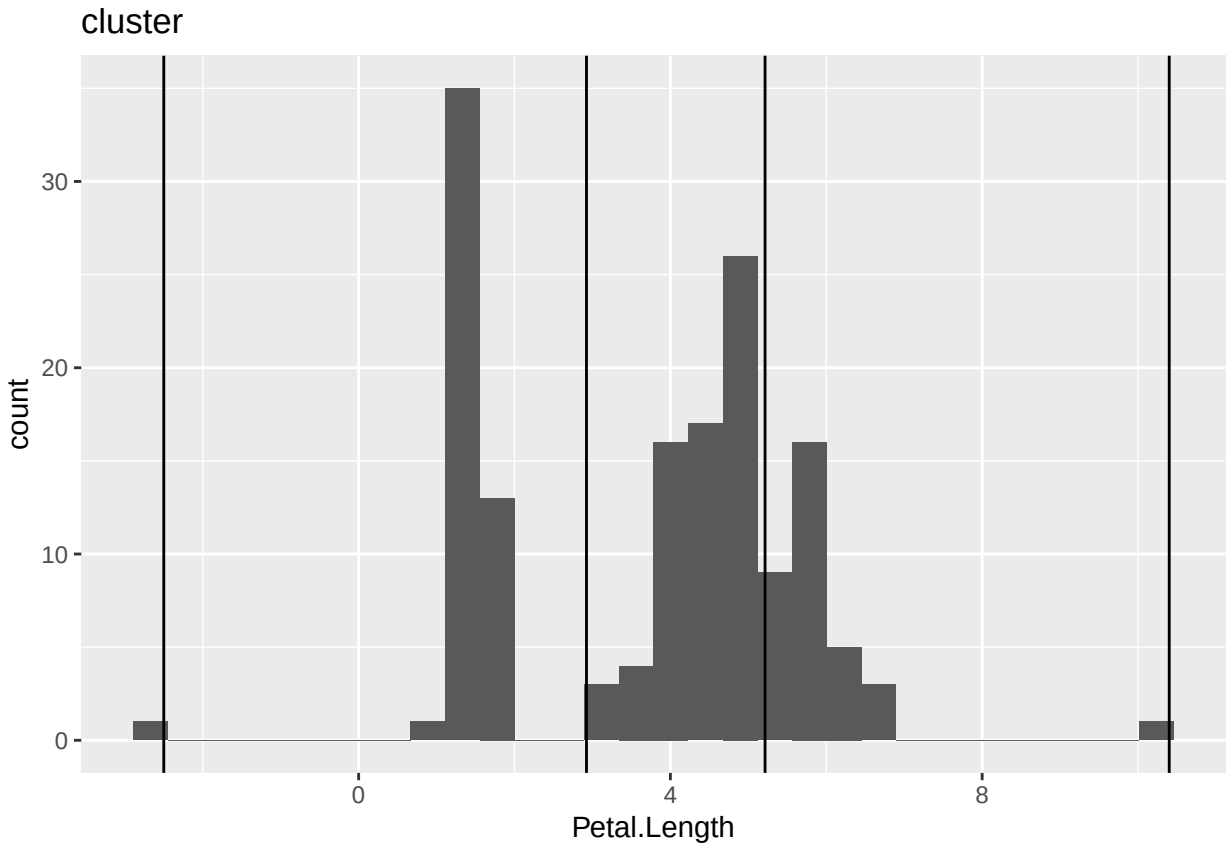


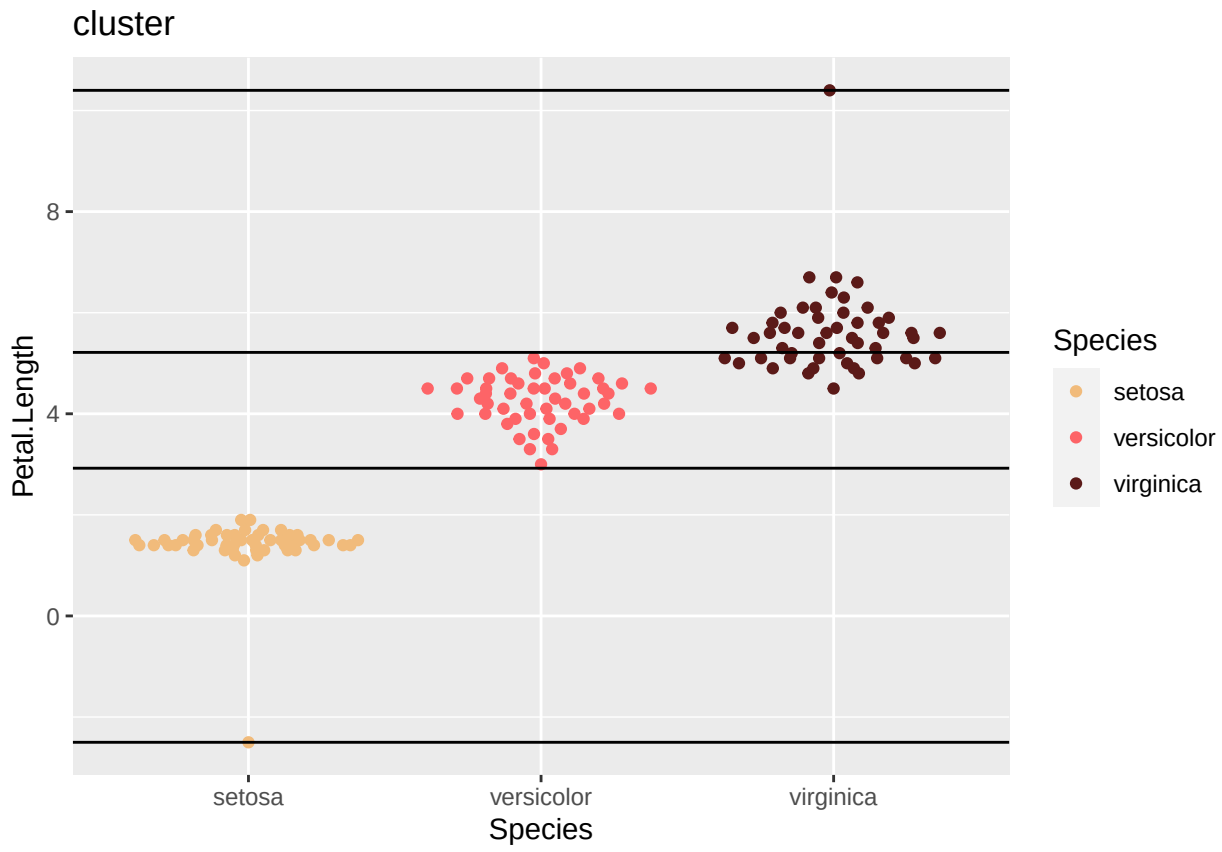

```
## Cases in matched pairs: 71.33 %
```


frequency
count
30
20
10
0
0
4
8
Petal.Length

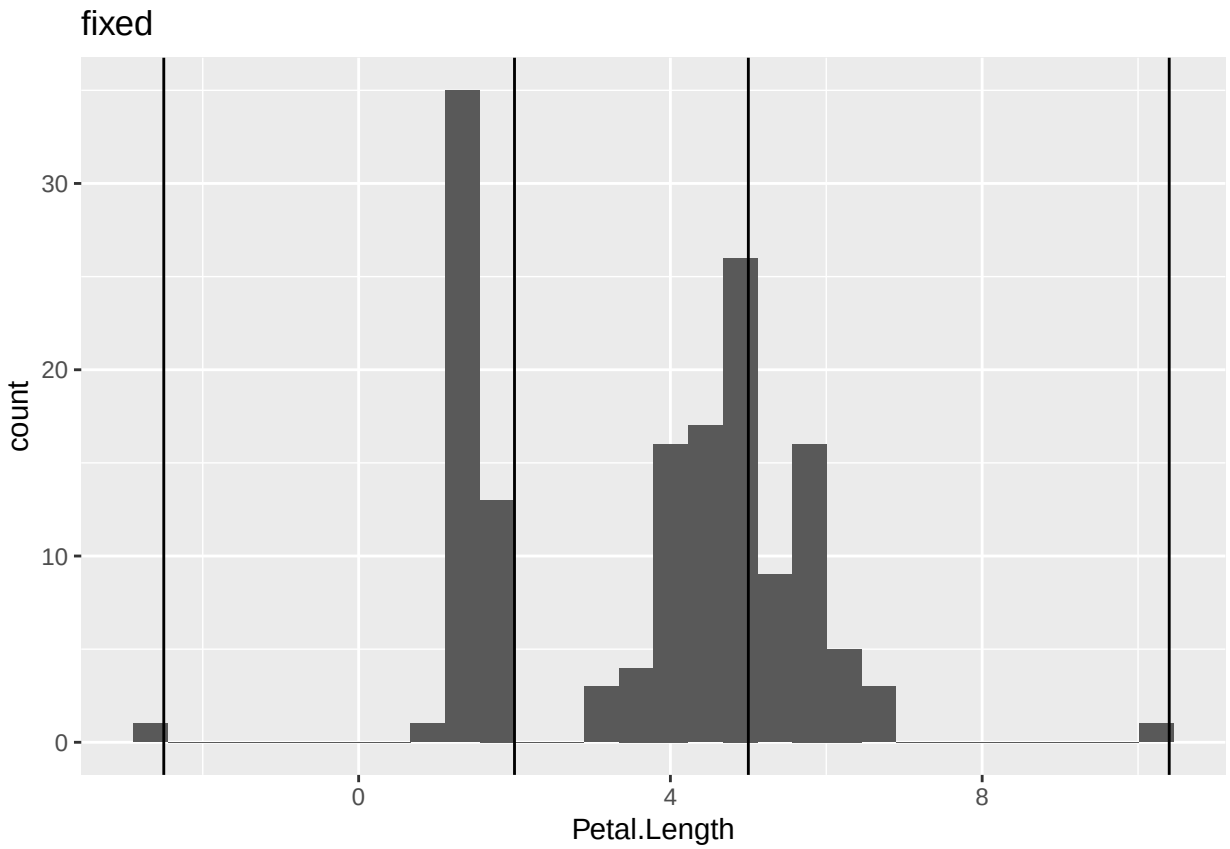


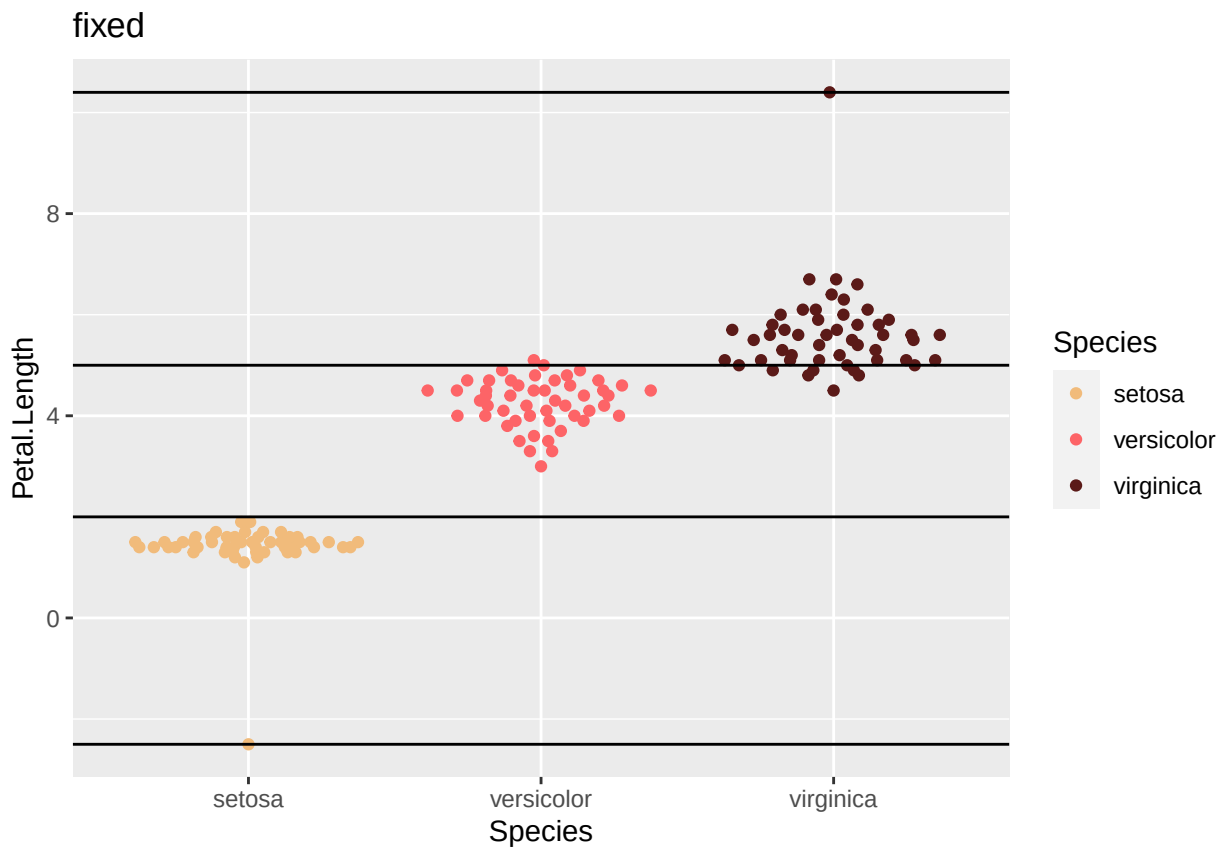

```
## Cases in matched pairs: 95.33 %
```


cluster
count
30
20
10
0
0
4
8
Petal.Length



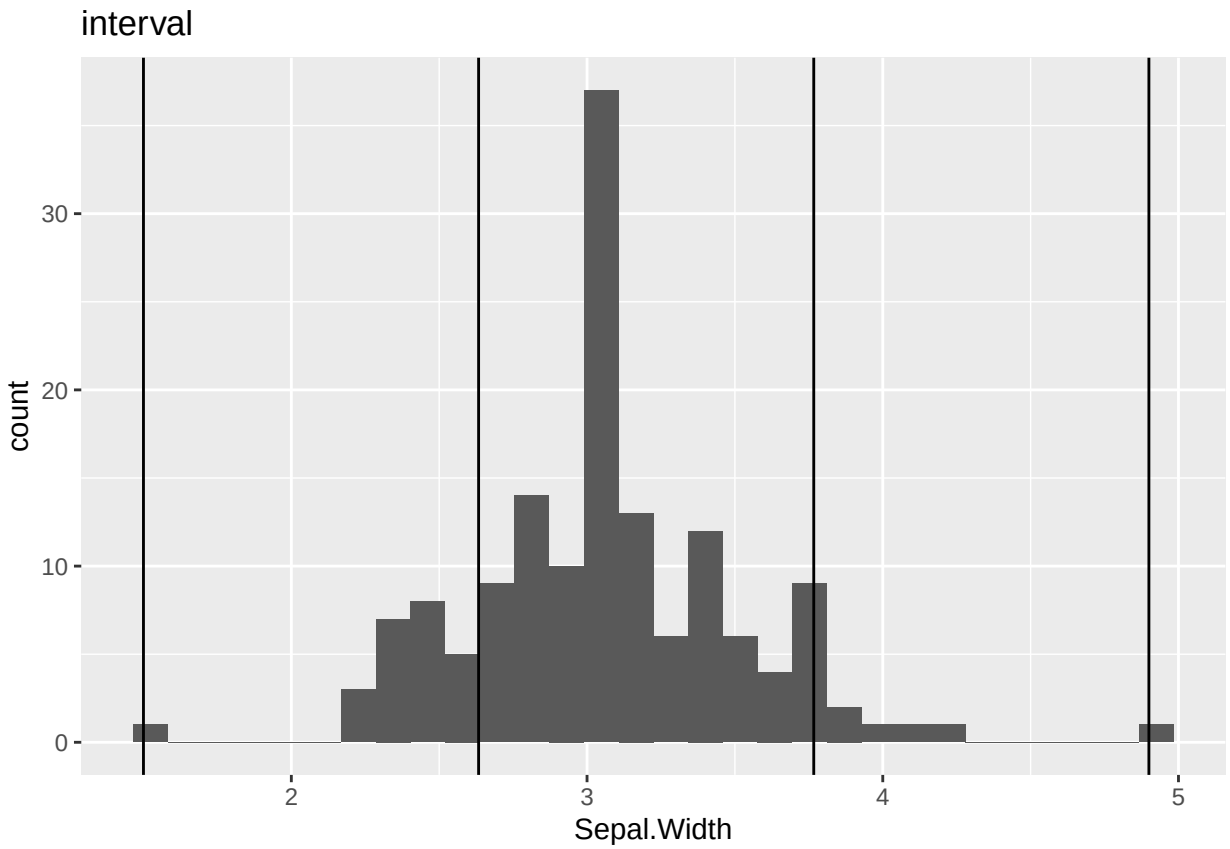

```
## Cases in matched pairs: 88 %
```

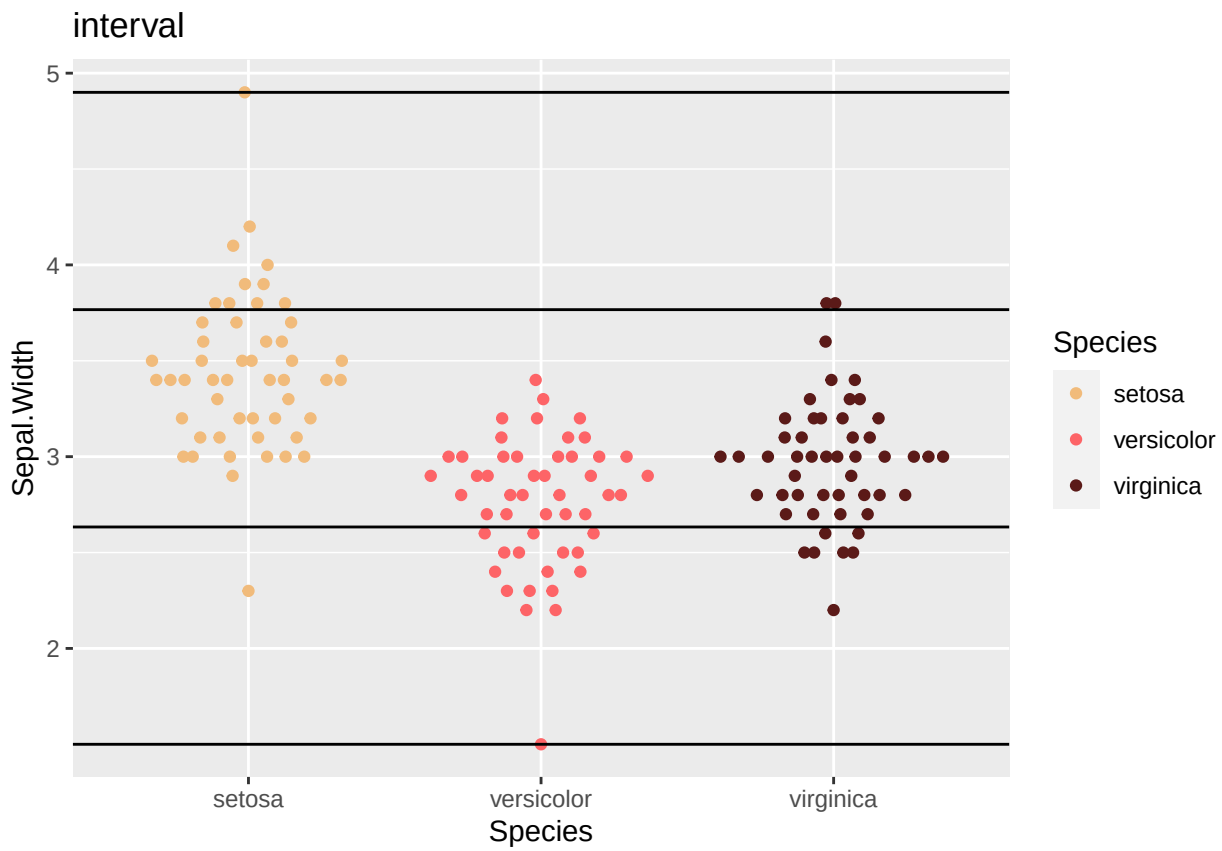

fixed
30
20
10
0
count
0
4
8
Petal.Length



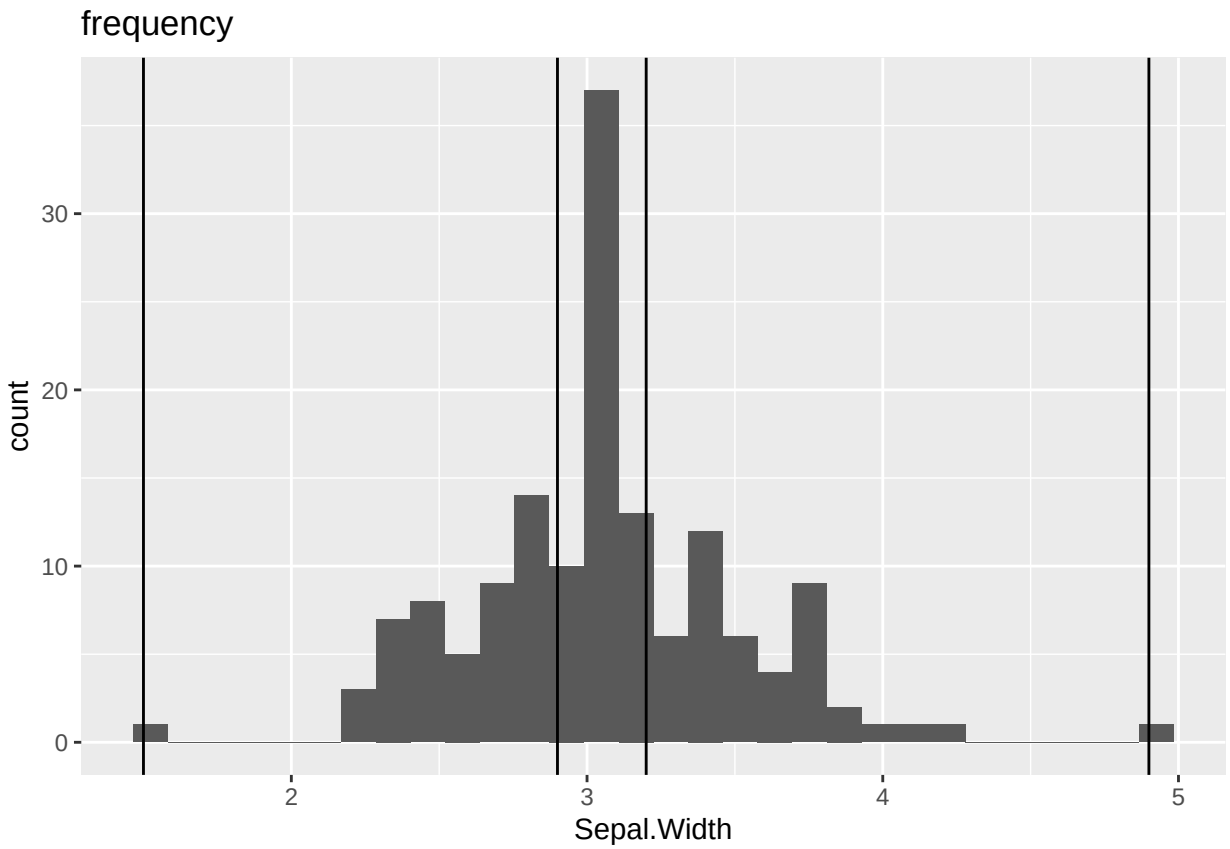

```
## Cases in matched pairs: 94.67 %
```

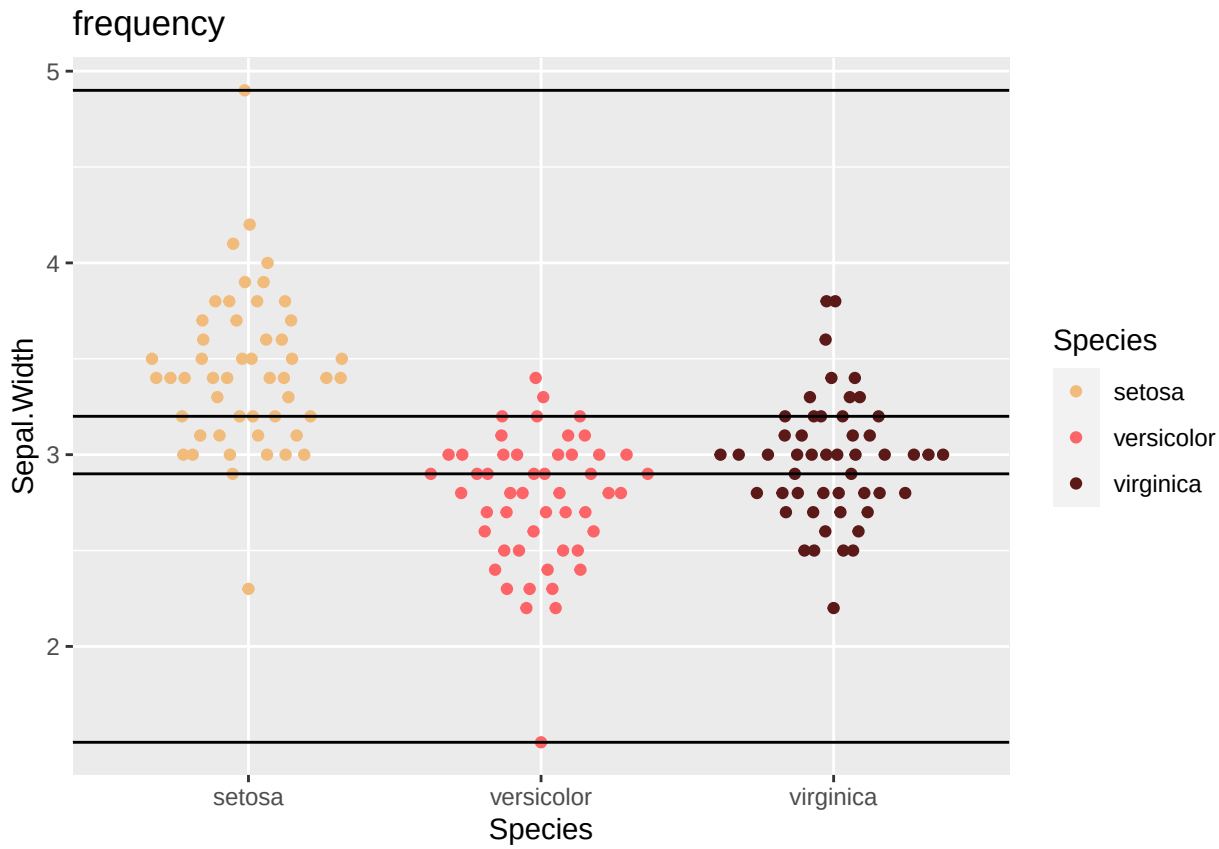
```
iris$Sepal.Width[which.min(iris$Sepal.Width)] <- min(iris$Sepal.Width) - IQR(iris$Sepal.Width)
iris$Sepal.Width[which.max(iris$Sepal.Width)] <- max(iris$Sepal.Width) + IQR(iris$Sepal.Width)
intervals <- c(min(iris$Sepal.Width), 2.5, 3, max(iris$Sepal.Width))
for (method in c("interval", "frequency", "cluster", "fixed")) {
  sepal.width.discretized <- if (method != "fixed")
    discretize(iris$Sepal.Width, method=method) else
    discretize(iris$Sepal.Width, method=method, breaks=intervals)
  print(ggplot(iris, aes(Sepal.Width)) +
          geom_histogram() +
          geom_vline(xintercept=attributes(sepal.width.discretized)$"discretized:breaks") +
          ggtitle(method))
  print(ggplot(iris, aes(Species, Sepal.Width)) +
    geom_quasirandom(aes(col=Species)) +
    scale_color_manual(values=wes_palette("GrandBudapest1", 3)) +
    geom_hline(yintercept=attributes(sepal.width.discretized)$"discretized:breaks") +
    ggtitle(method))
  discretized.table <- table(sepal.width.discretized, iris$Species)
  matchClasses(discretized.table)
}
```


interval
count
30
20
10
0
2
3
4
5
Sepal.Width

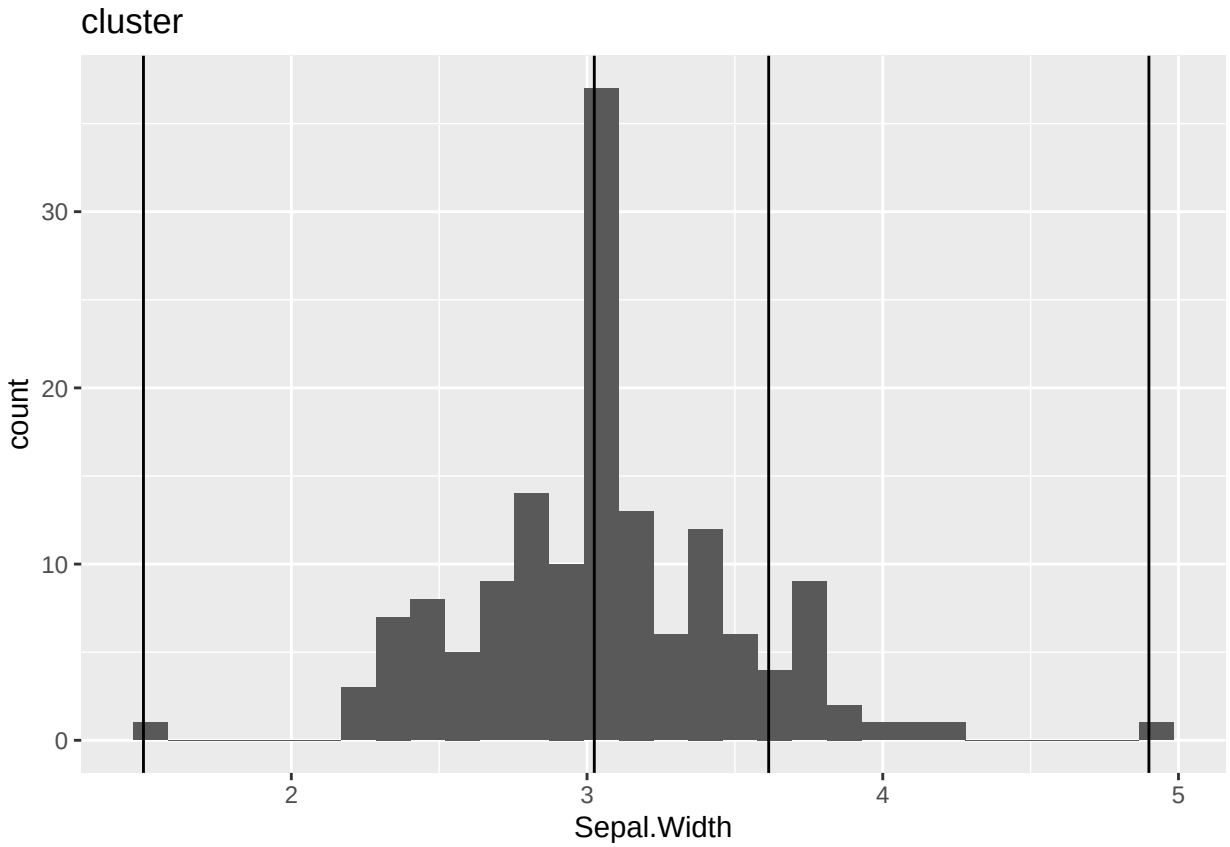


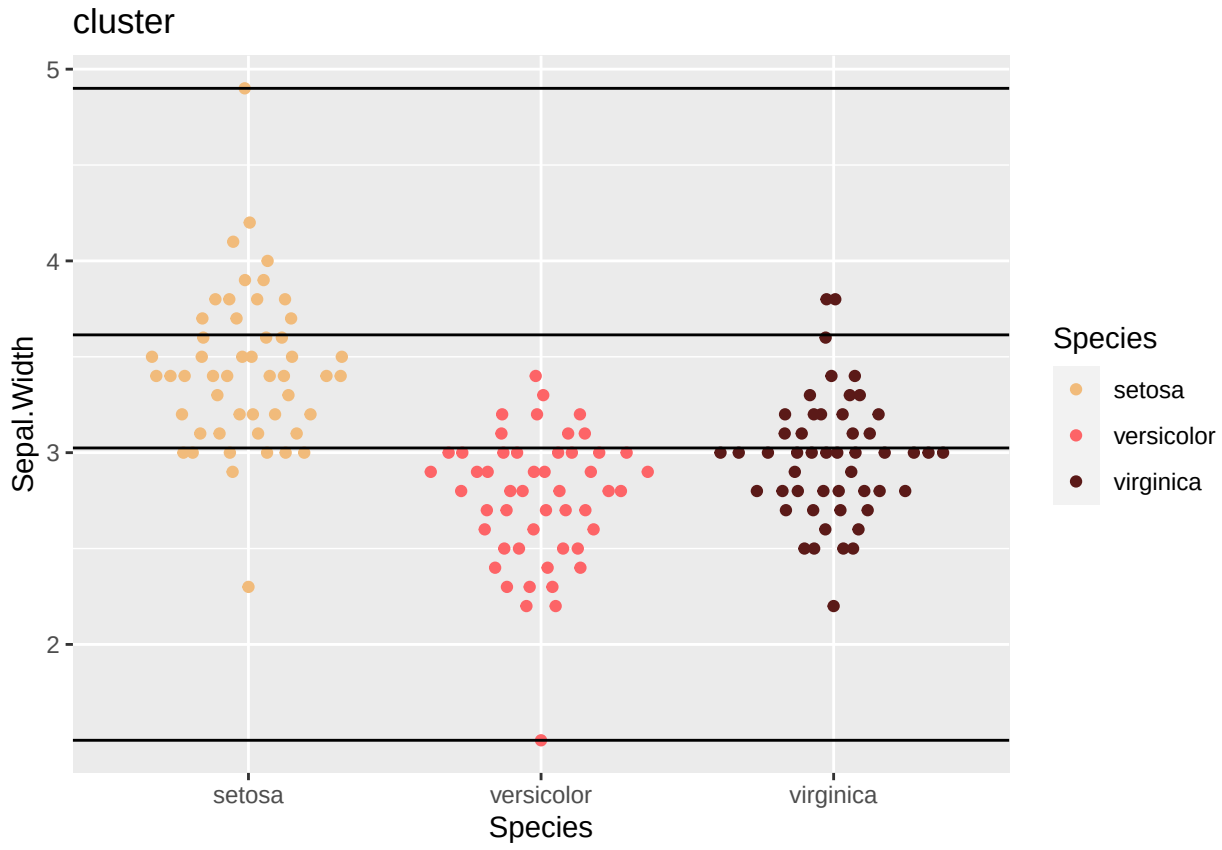

```
## Cases in matched pairs: 44.67 %
```


frequency
count
30
20
10
0
2
3
4
5
Sepal.Width

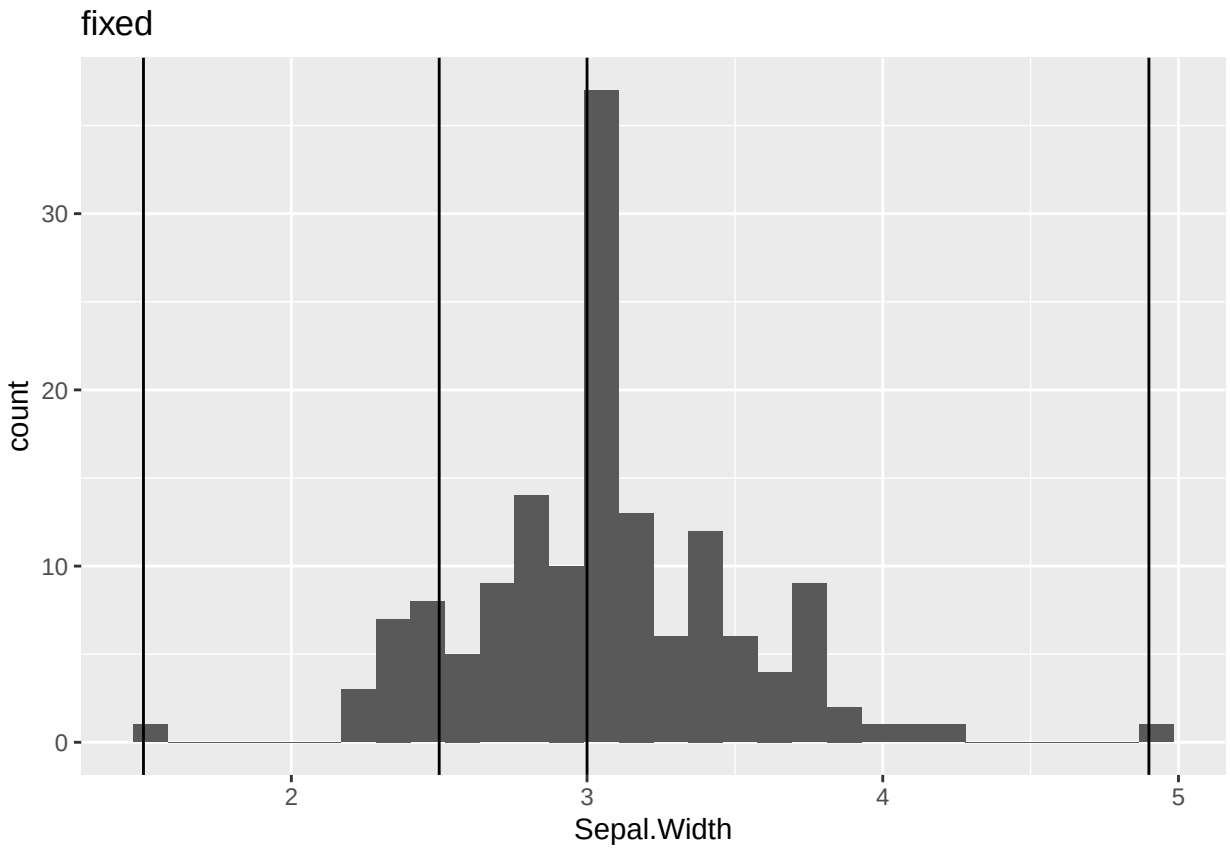


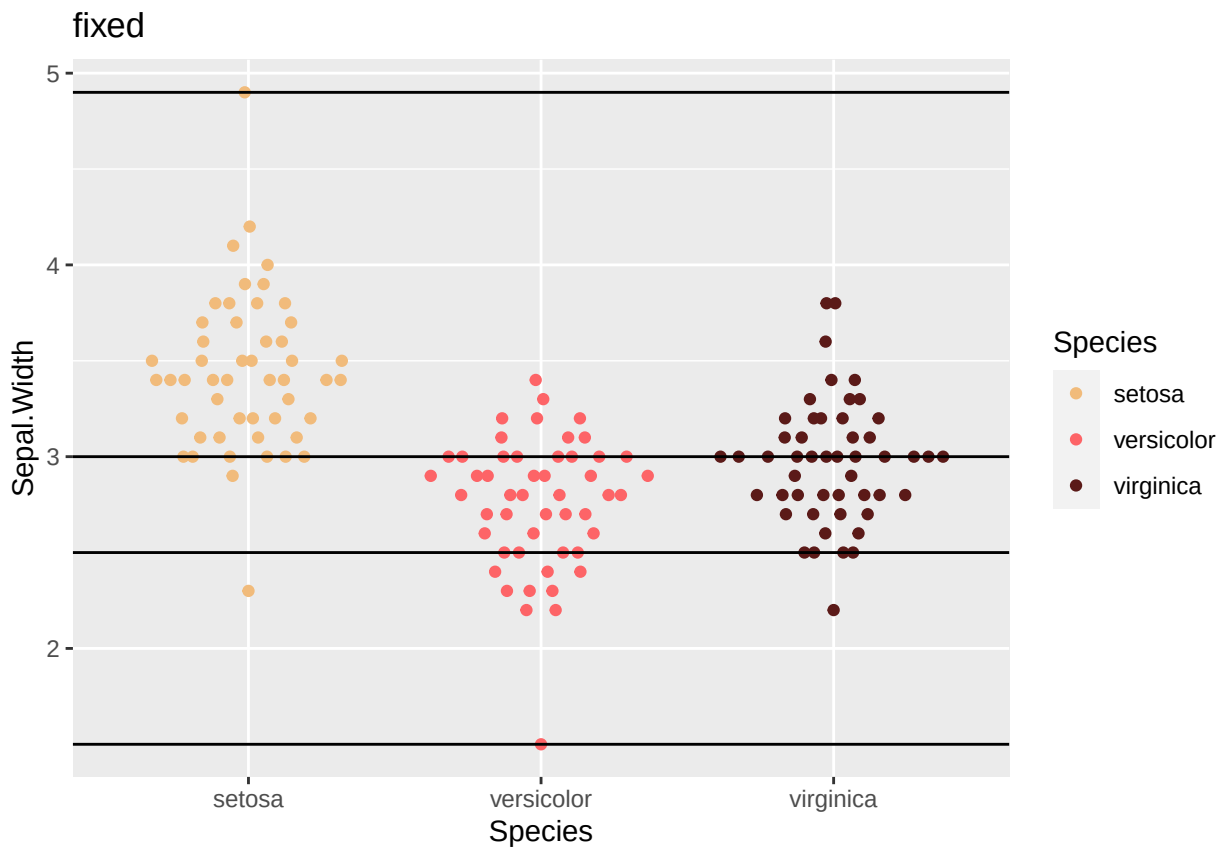

```
## Cases in matched pairs: 55.33 %
```


cluster
count
30
20
10
0
2
3
4
5
Sepal.Width




```
## Cases in matched pairs: 56 %
```


fixed
count
30
20
10
0
2
3
4
5
Sepal.Width




```
## Cases in matched pairs: 54.67 %
```