Raport 2

Eksploracja danych

Mikoaj Langner, Marcin Kostrzewa
nr albumów: 255716, 255749

2021-04-19

Spis treści

1 Wstp

Sprawozdanie zawiera rozwizanie zada z listy 2. Dotycz one zagadnie dyskretyzacji i redukcji wymiaru.

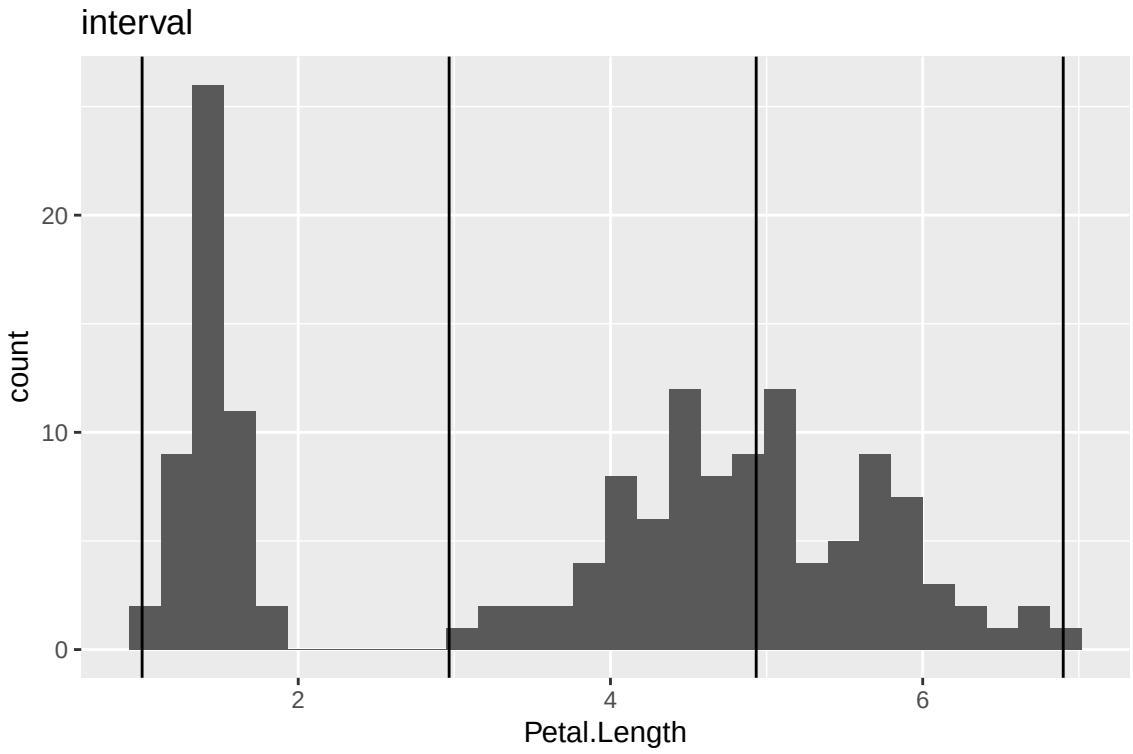
2 Zadanie 1

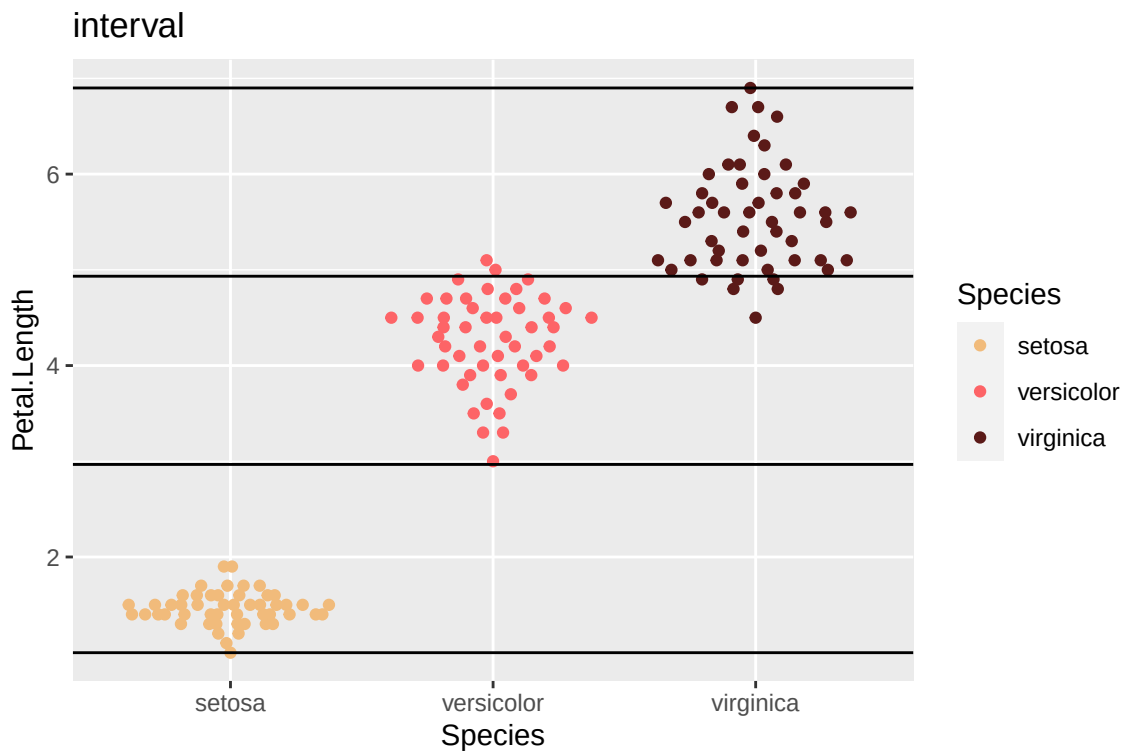
W pierwszym zadaniu mamy zbada

```
data(iris)
```

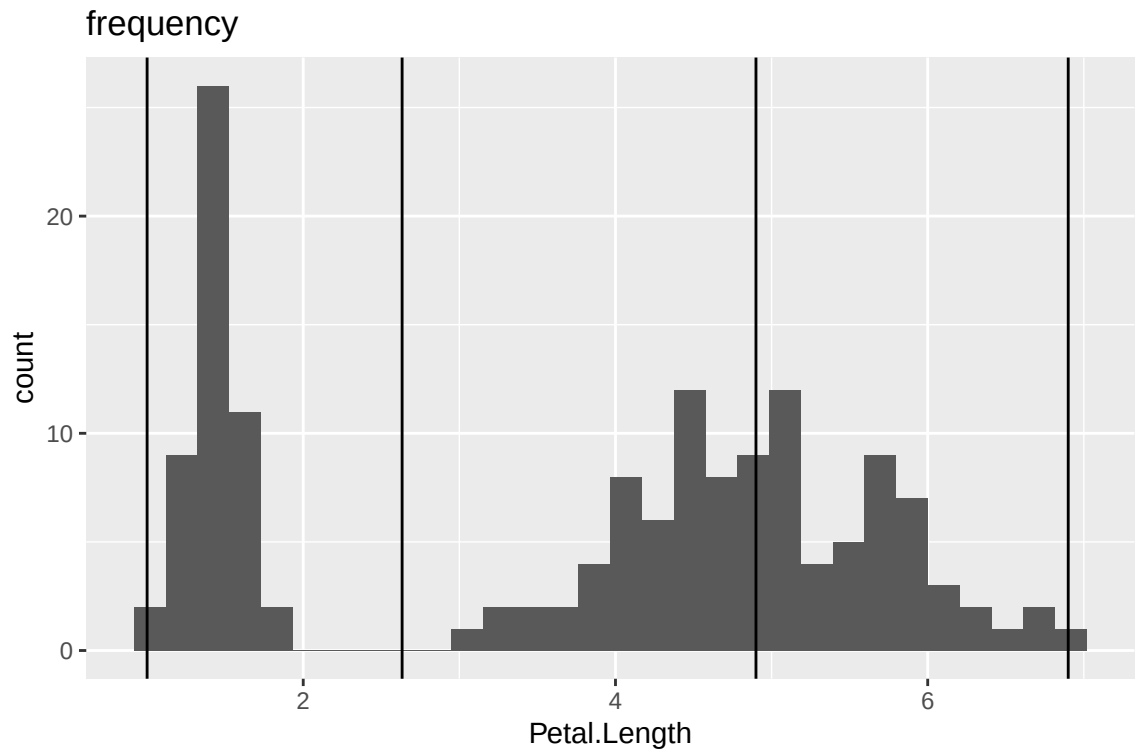
```
intervals <- c(min(iris$Petal.Length), 2, 5, max(iris$Petal.Length))
for (method in c("interval", "frequency", "cluster", "fixed")) {
  petal.length.discretized <- if (method != "fixed")
    discretize(iris$Petal.Length, method=method) else
    discretize(iris$Petal.Length, method=method, breaks=intervals)
  print(ggplot(iris, aes(Petal.Length)) +
          geom_histogram() +
          geom_vline(xintercept=attributes(petal.length.discretized)$"discretized:breaks
```

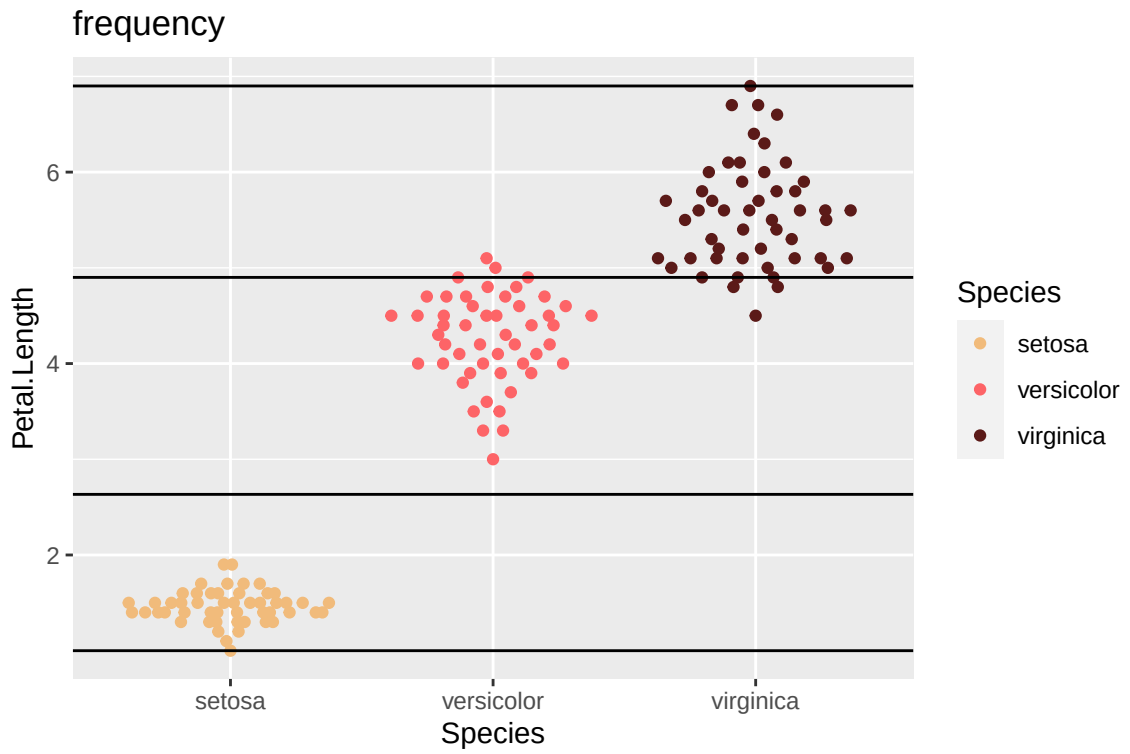
```
          ggtitle(method))
  print(ggplot(iris, aes(Species, Petal.Length)) +
    geom_quasirandom(aes(col=Species)) +
    scale_color_manual(values=wes_palette("GrandBudapest1", 3)) +
    geom_hline(yintercept=attributes(petal.length.discretized)$"discretized:breaks") +
    ggtitle(method))
  discretized.table <- table(petal.length.discretized, iris$Species)
  matchClasses(discretized.table)
}
```



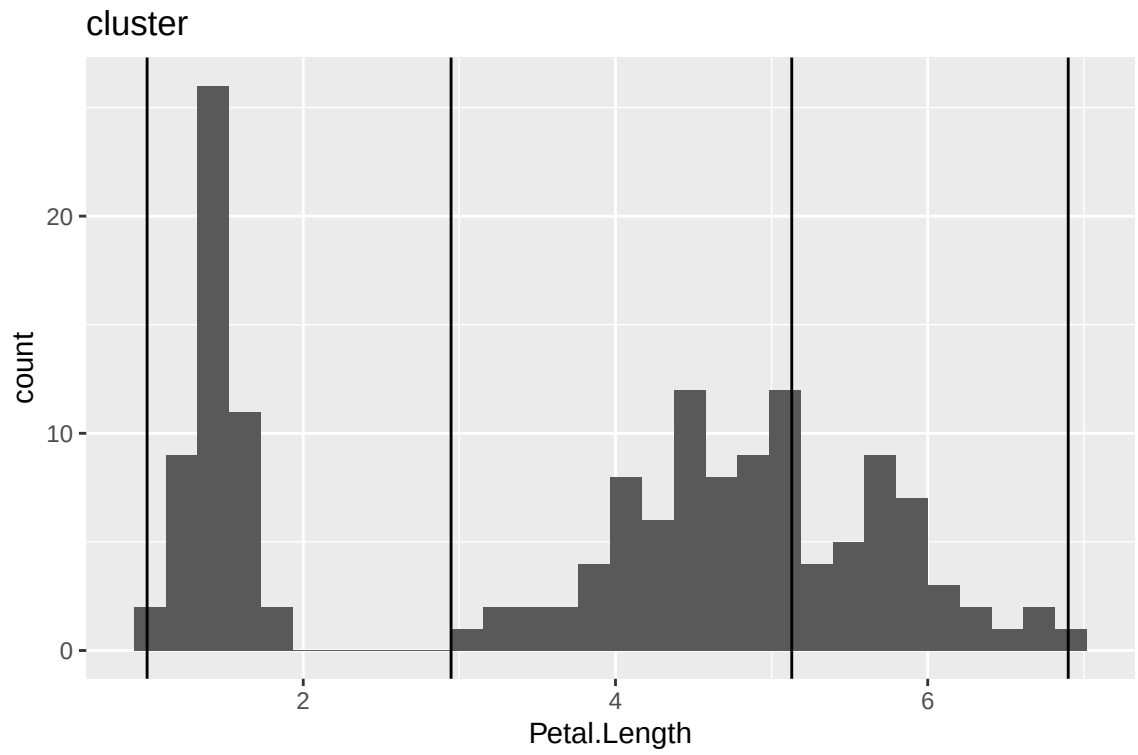


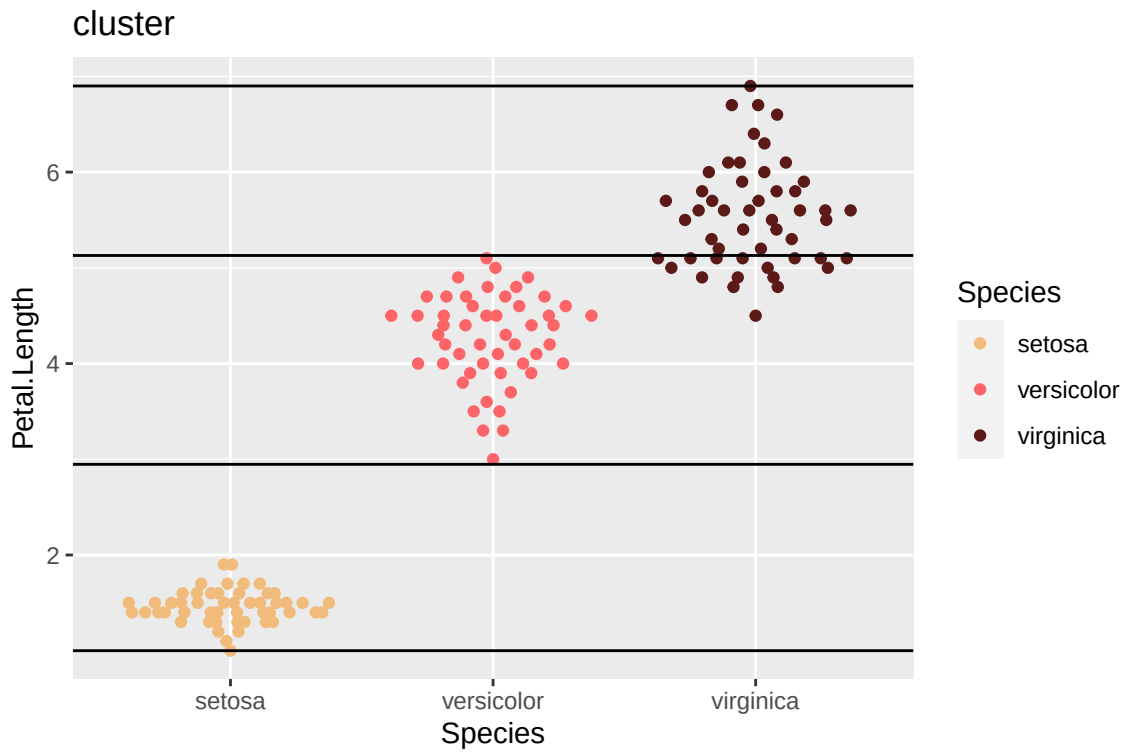

```
## Cases in matched pairs: 94.67 %
```



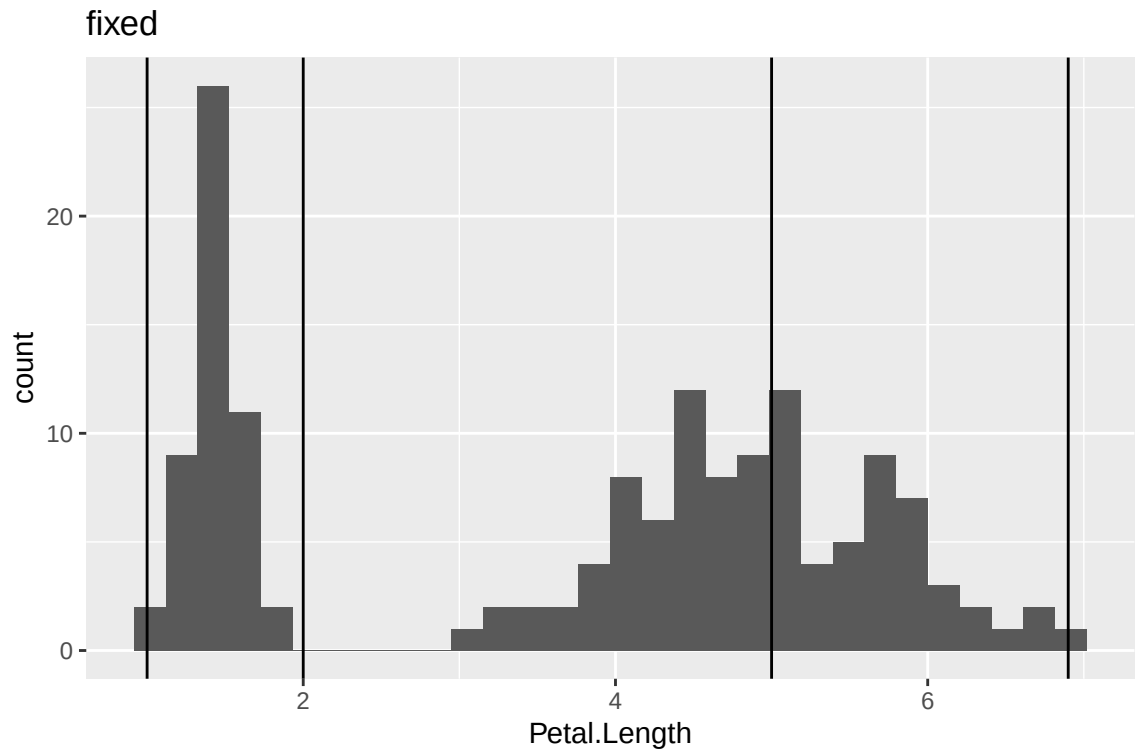


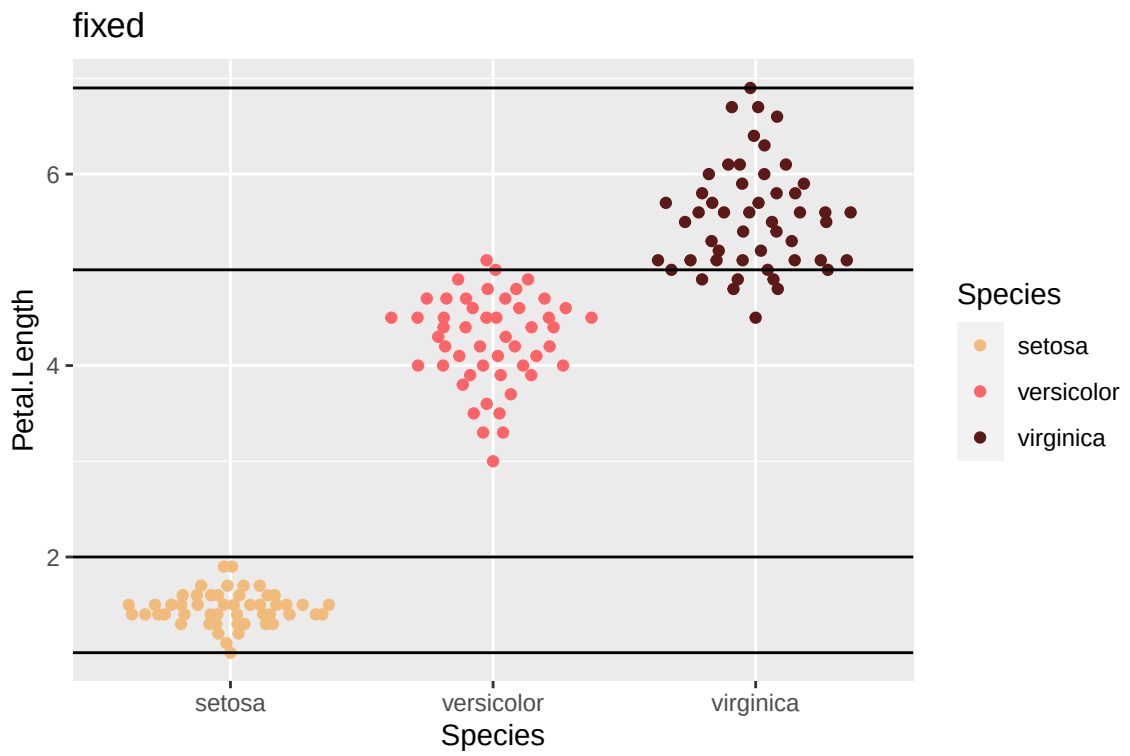

```
## Cases in matched pairs: 95.33 %
```






```
## Cases in matched pairs: 89.33 %
```






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

3 Zadanie 2

3.1 Wczytanie i przygotowanie danych

Teraz naszym zadaniem jest ...

Wczytajmy dane i uzupenijmy je o informacje geograficzne o wszytkich stanach.

```
data(state)
state <- as.data.frame(state.x77)
state_ <- state
state_$region <- state.region
state_$division <- state.division
```

By rozstrzygn, czy potrzebna jest normalizacja danych, przeanalizujemy wykresy pudekowe oraz wyznaczymy odchylenia standardowe i współczynniki zmiennoci.

```
plot_boxplot(data.frame(state, all="all"), by="all")
```

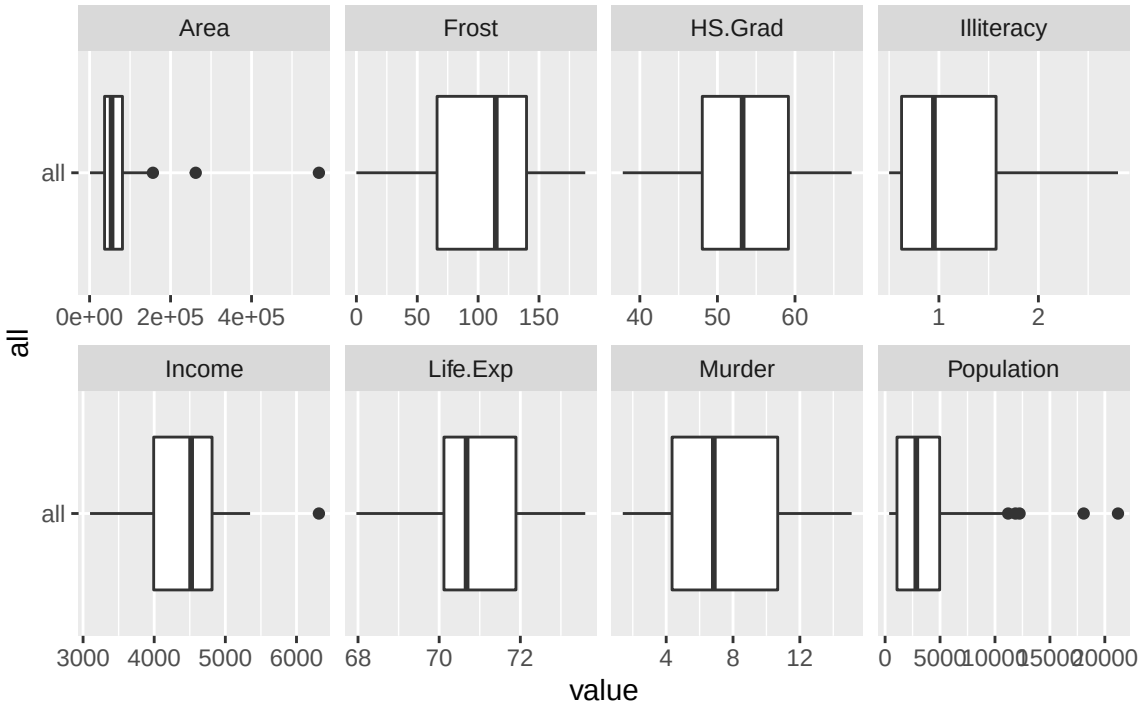
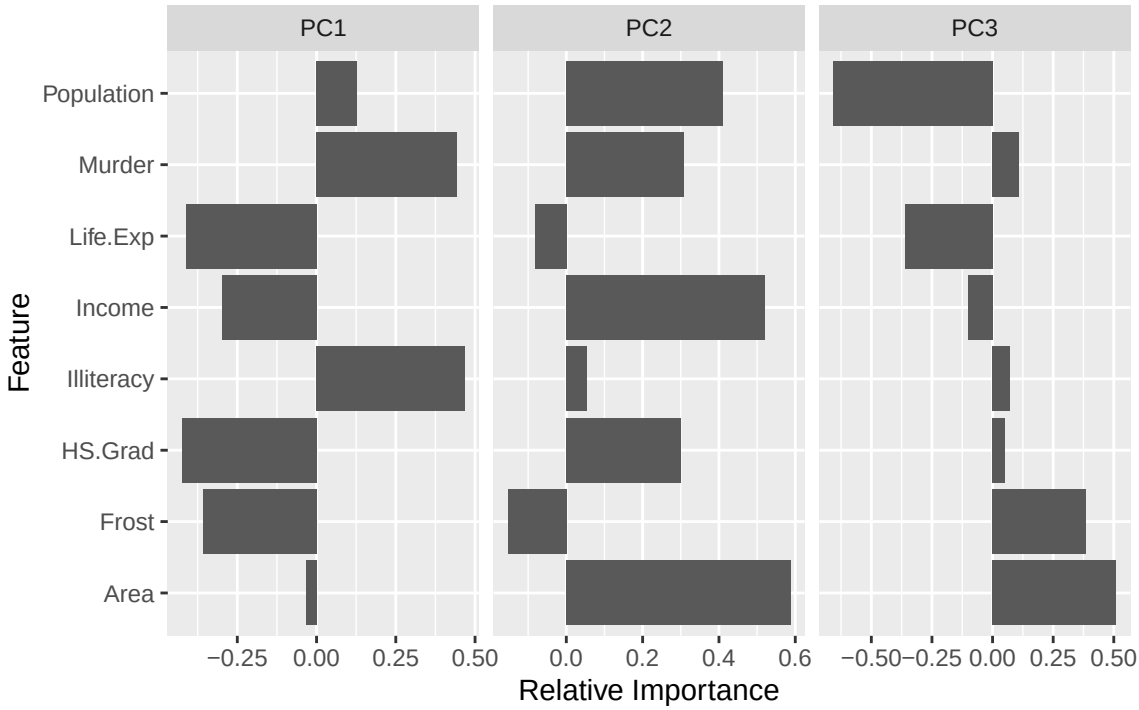


Tabela 1: Odchylenie standardowe i wspolczynnik zmiennosci dla zmienych

	Population	Income	Illiteracy	Life Exp	Murder	HS Grad	Frost	Area
Odchylenie standardowe	4464.491433	614.4699392	0.6095331	1.3423936	3.6915397	8.0769978	51.9808481	85327.29962
Wspolczynnik zmiennosci	1.051354	0.1385252	0.5209685	0.0189393	0.5003442	0.1520863	0.4976149	1.20628

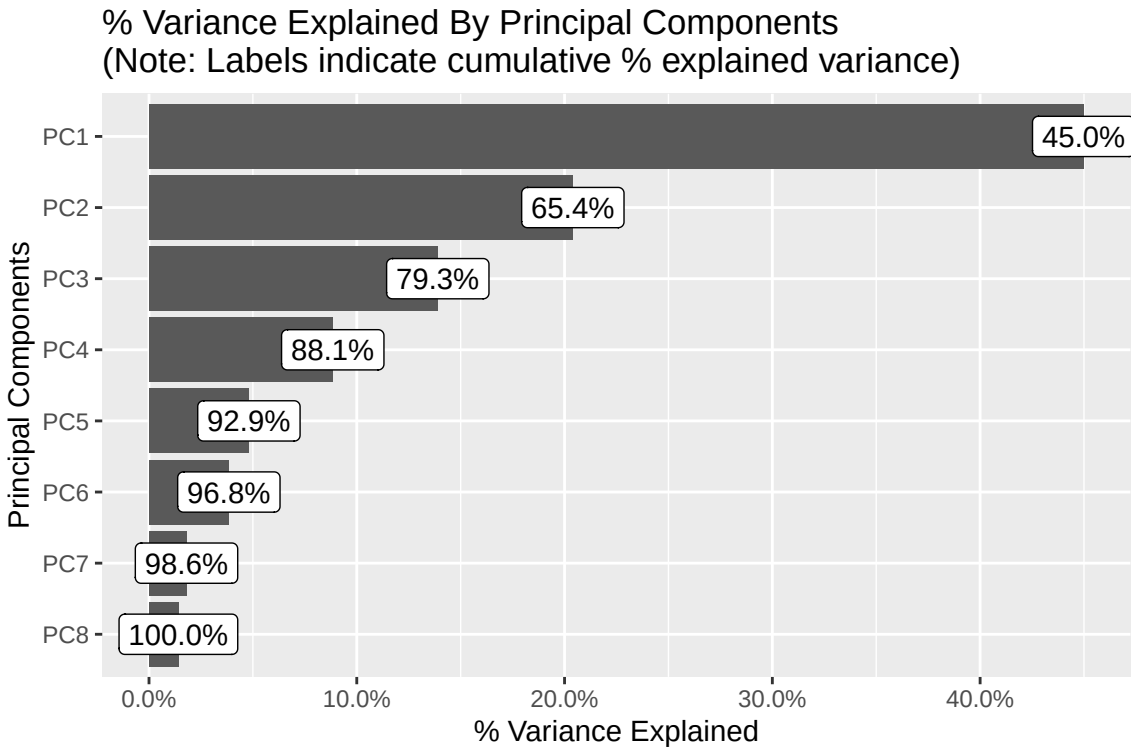
```
p <- plot_prcomp(state, prcomp_args=list(scale=TRUE, center=TRUE), variance_cap=0.8)[2]
```

```
print(p)
```




```
p <- plot_prcomp(state, prcomp_args=list(scale=TRUE, center=TRUE), variance_cap=1)[1]
```

```
print(p)
```



4 Zadanie 3