

Wojskowa Akademia Techniczna



Analiza i wizualizacja danych

Wizualizacja danych jedno i dwuczynnikowych

Prowadzący: xxxx

Wykonał: Konrad Prusaczyk

Grupa: WCY18IJ3S1

Data wykonania: 08/04/2021

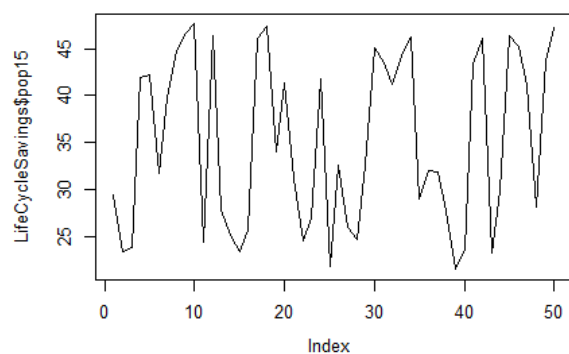
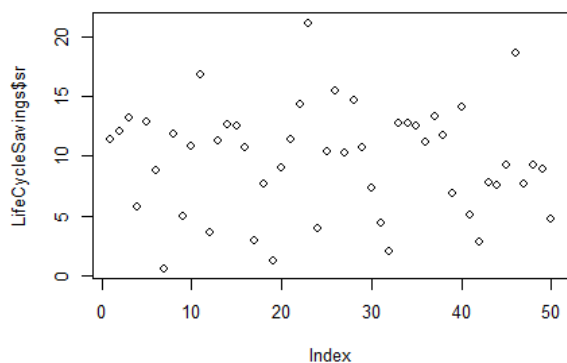
Rozpatrywanym zbiorem danych jest zbiór LifeCycleSavings dostępny w ramach pakietu datasets. Pierwsze 6 wierszy prezentuje się następująco:

```
> head(d1)
      sr pop15 pop75      dpi ddpi
1 11.43 29.35  2.87 2329.68 2.87
2 12.07 23.32  4.41 1507.99 3.93
3 13.17 23.80  4.43 2108.47 3.82
4  5.75 41.89  1.67  189.13 0.22
5 12.88 42.19  0.83  728.47 4.56
6  8.79 31.72  2.85 2982.88 2.43
> |
```

Podsumowanie poszczególnych kolumn w zbiorze danych:

```
> summary(d1)
      sr      pop15      pop75      dpi      ddpi
Min.   : 0.600   Min.   :21.44   Min.   :0.560   Min.   : 88.94   Min.   : 0.220
1st Qu.: 6.970   1st Qu.:26.21   1st Qu.:1.125   1st Qu.: 288.21   1st Qu.: 2.002
Median :10.510   Median :32.58   Median :2.175   Median : 695.66   Median : 3.000
Mean   : 9.671   Mean   :35.09   Mean   :2.293   Mean   :1106.76   Mean   : 3.758
3rd Qu.:12.617   3rd Qu.:44.06   3rd Qu.:3.325   3rd Qu.:1795.62   3rd Qu.: 4.478
Max.   :21.100   Max.   :47.64   Max.   :4.700   Max.   :4001.89   Max.   :16.710
> |
```

Przystępuję do obrazowania danych przy pomocy wykresów:



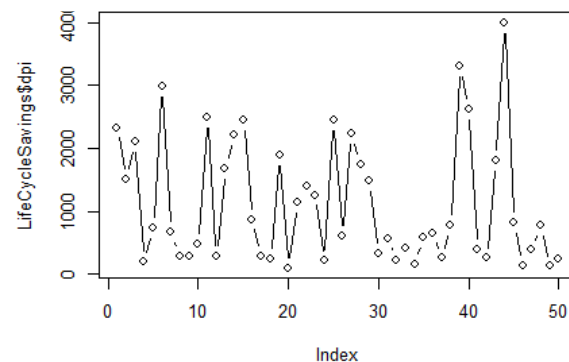
```
par(mfrow=c(2, 2), mar=c(4, 4, 1, 1))
```

```
plot(LifeCycleSavings$dpi, type="b")
```

```
plot(LifeCycleSavings$ddpi, type="h")
```

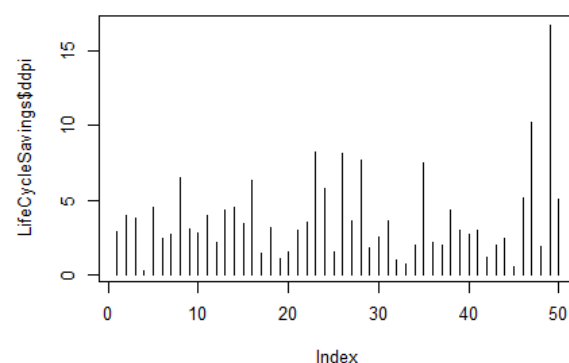
```
plot(LifeCycleSavings$sr, type="p")
```

```
plot(LifeCycleSavings$pop15, type="l")
```



Przedstawienie kolejno kolumn:

- *dpi* jako typ „both”
- *ddpi* jako typ „histogram”
- *sr* jako typ „points”
- *pop15* jako typ „lines”



```
rownames(d1) <- c()
```

```
d1
```

```
df<-d1 %>% select(4)
```

```
df
```

```
data(df, package="rattle")
```

```
df1<-scale(df)
```

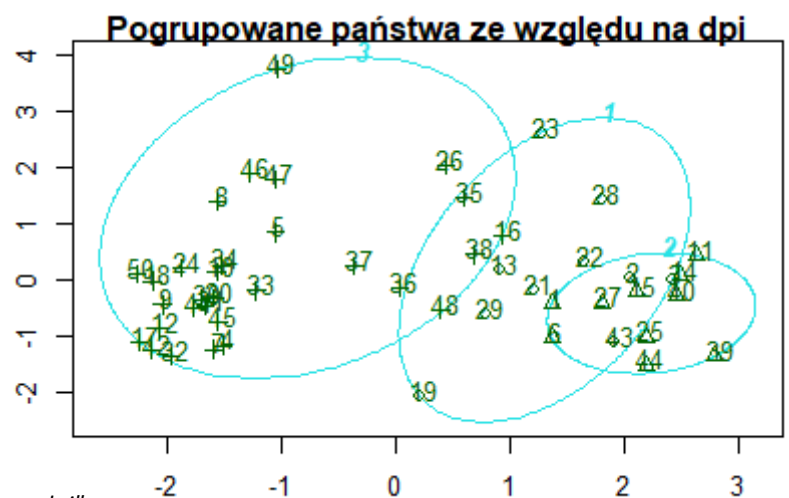
```
kmf<-kmeans(df1,3)
```

```
kmf$cluster
```

```
c1<-cbind(kmf$cluster)
```

```
c1
```

```
clusplot(d1, kmf$cluster,
          main="Pogrupowane państwa ze względu na dpi",
          labels=2, lines=0)
```



Starąłem się pogrupować dane kraje ze względu na dochód na mieszkańca, aby wykres był nieco bardziej czytelny zamieniłem nazwy państw na indeksy. Wykorzystałem metodę k-średnich z biblioteki „cluster”. Udało mi się wyodrębnić 3 grupy.

```
png(file = "dpi_percentage.jpg")
```

```
lbl<-c("Srednia pop15", "Srednia pop75")
```

```
piepercent<- round(100*x/sum(x), 1)
```

```
pie(x,
```

```
labels = piepercent,
```

```
main = "Srednia pop15 oraz pop75",
```

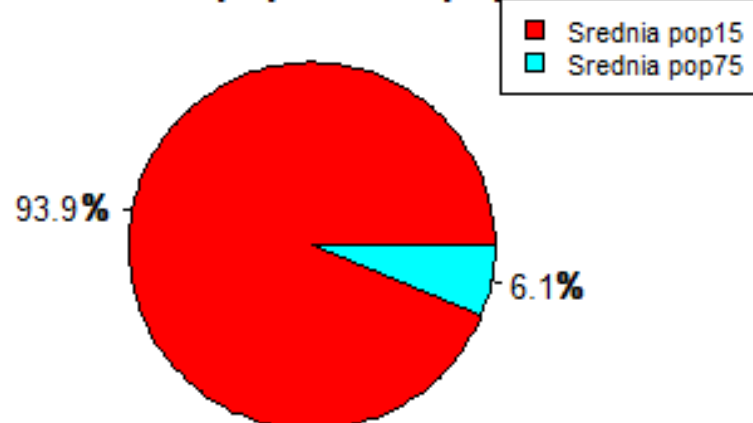
```
col = rainbow(length(x)))
```

```
legend("topright", c("Srednia pop15", "Srednia pop75"),
```

```
cex = 0.8, fill = rainbow(length(x)))
```

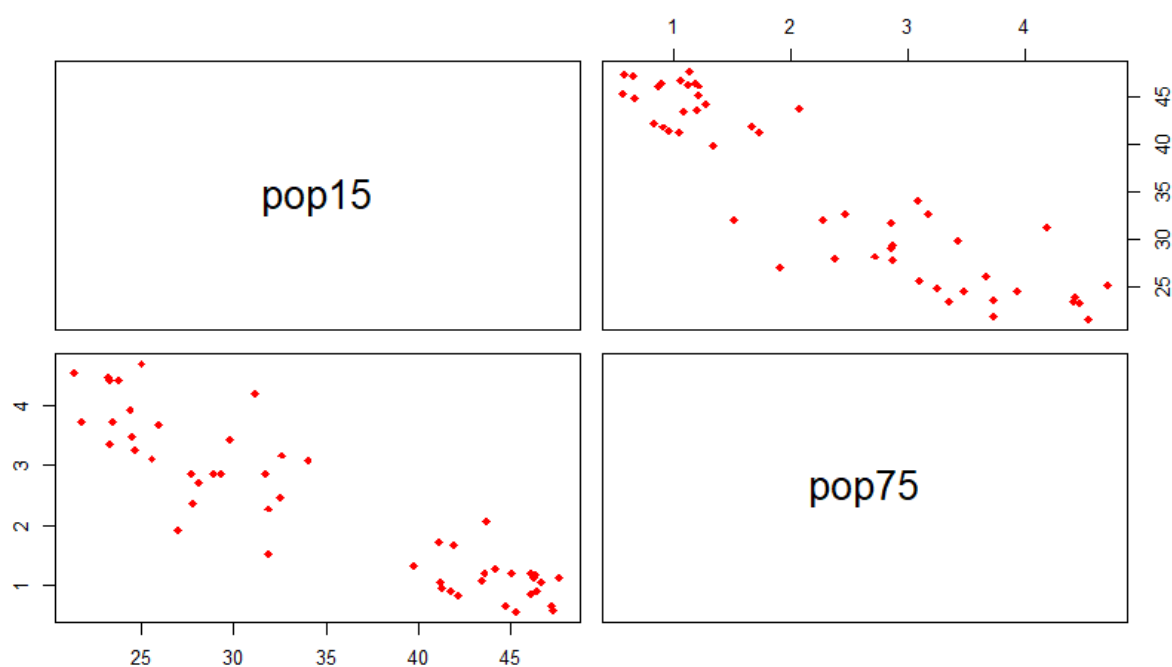
```
dev.off())
```

Index
Srednia pop15 oraz pop75



Wykres kołowy obrazujący zestawienie średniej wartości z kolumny „pop15” oraz „pop75” na tle obu wartości w postaci procentowej. W celu stworzenia wykresu skorzystałem z biblioteki plotrix.

Wykresy rozrzutu pop15 oraz pop75



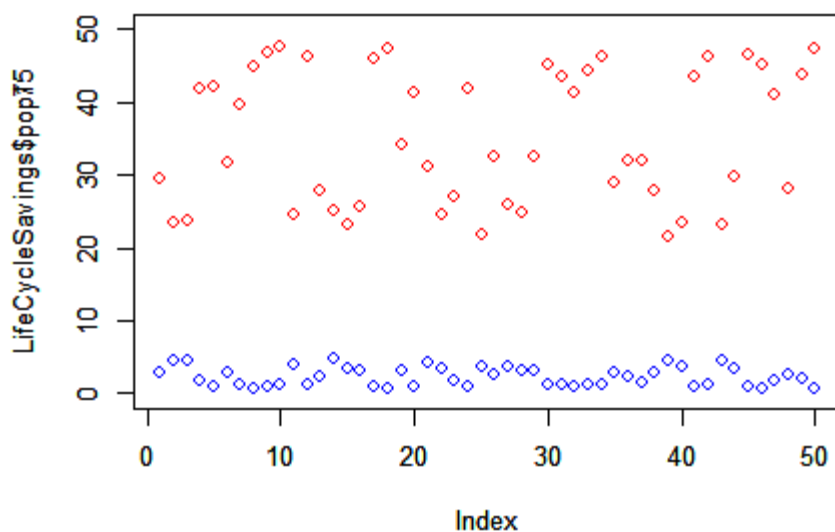
```
pairs(d1[, 2:3],
      col = "red",
      pch = 18,
      labels = c("pop15", "pop75"),
      main = "Wykresy rozrzutu pop15 oraz pop75")
```

Macierz wykresów, składającą się z wykresów rozrzutu dla kolumny pop15 oraz pop7. Gdyby zastosować w tym przypadku klasteryzację również dałoby się wyodrębnić poszczególne grupy państw.

```
plot(LifeCycleSavings$pop15,
     ylim=c(0, 50),
     type="p",col="red")

par(new=TRUE)

plot(LifeCycleSavings$pop75,
     ylim=c(0, 50),
     type="p",col="blue")
```



Powyższy wykres jest wynikiem sparowania dwóch wykresów (zależności zmiennej pop15 oraz pop75 względem państwa – w tym przypadku indeksu państwa).