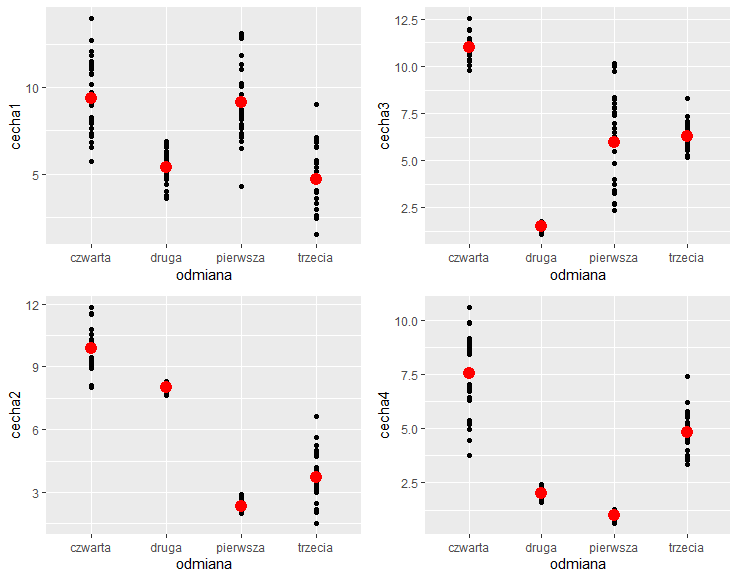
Karol Pichurski

321663

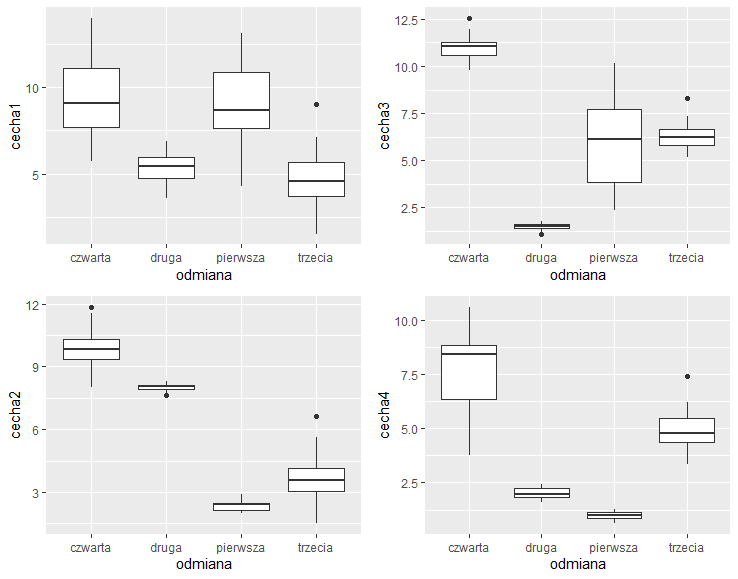
# Podstawy Analizy Danych – Lista 3

Wykres nr 1:



Na wykresie nr 1 przedstawiłem odmiany z podziałem na poszczególne cechy. Są przedstawione na nich poszczególne wartości oraz ich średnia wartość zaznaczona czerwonym punktem. Możemy zauważyć że niektóre cechy mają ogromny rozrzut w porównaniu do innych bardziej skoncentrowanych.

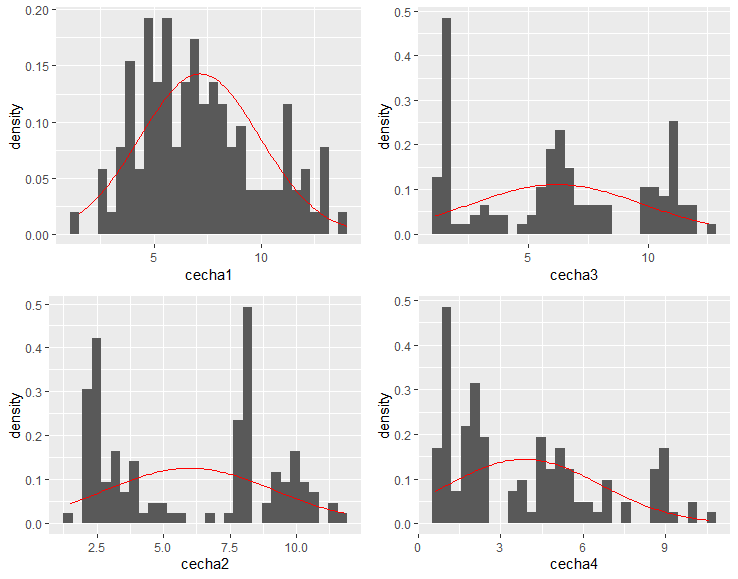
Wykres nr 2:



Na wykresie nr 2 przedstawiony jest typowy „boxplot” dla odmian podzielonych z poszczególnymi cechami.

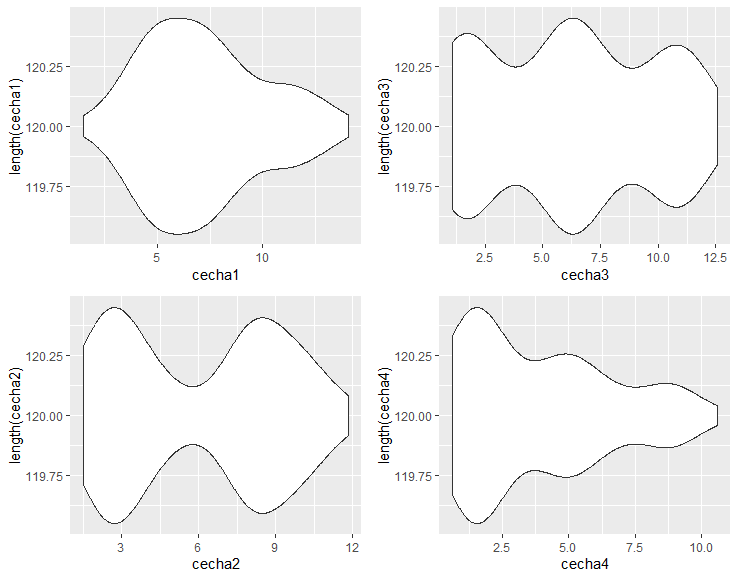
Wykres składa się z prostokąta (pudełka), osi współrzędnych i tak zwanych wąsów, czyli odległości wartości minimalnej i maksymalnej od pudełka. Wykres ma formę minimalistyczną, ponieważ przedstawia tylko 5 informacji i chociaż wydaję się, że jest to niewiele, wykres ramka-wąsy przekazuje nam na bardzo dużo wiadomości. Najczęściej do jego sporządzenia musimy znać dokładne wartości: pierwszego i trzeciego kwartyla, mediany oraz minimum i maksimum. Dzięki tym danym możemy badać poziom rozproszenia danej czy odległości mediany od minimum i maksimum.

Wykres nr 3:



Na wykresie nr 3 przedstawiony jest wykres cech jako histogram z czerwoną linią przedstawiającą rozkład hostogramu. Na wykresie możemy odczytać liczebność poszczególnych przedziałów, skośność rozkładu oraz przedziały odstające.

Wykres nr 4:



Na wykresie nr 4 przedstawiony jest wykres skrzypcowy, który jest często porównywany do wykresu boxplot.

Wykres skrzypcowy w porównaniu do wykresu ramka-wąsy:

* Daje prosty sposób porównania kilku cech na jednym wykresie.
* Przekazuje więcej informacji niż tylko wykres pudełkowy bądź wykres gęstości – łączy dwa wykresy w jednym.
* W przypadku obserwacji o specyficznym rozkładzie (np. bimodalnym – dwa pagórki) dzięki wyświetleniu informacji o rozkładzie przekaże więcej informacji niż sam wykres pudełkowy.
* Wykres skrzypcowy wciąż jest bardzo mało popularny, dlatego często trudniejszy do interpretacji. Nie jest na pierwszy rzut oka intuicyjny i należy się zorientować, co jest na nim przedstawione. Dlatego pod względem prostoty przegrywa zarówno z histogramem jak i z wykresem pudełkowym.

Kod:

1. multiplot <**-** function(..., plotlist**=**NULL, file, cols**=**1, layout**=**NULL) {
2. library(grid)
3. plots <**-** c(list(...), plotlist)
4. numPlots **=** length(plots)
6. **if** (**is**.null(layout)) {
7. layout <**-** matrix(seq(1, cols **\*** ceiling(numPlots**/**cols)),
8. ncol **=** cols, nrow **=** ceiling(numPlots**/**cols))
9. }
11. **if** (numPlots**==**1) {
12. print(plots[[1]])
14. } **else** {
15. grid.newpage()
16. pushViewport(viewport(layout **=** grid.layout(nrow(layout), ncol(layout))))
18. **for** (i **in** 1:numPlots) {
19. matchidx <**-** as.data.frame(which(layout **==** i, arr.ind **=** TRUE))
20. print(plots[[i]], vp **=** viewport(layout.pos.row **=** matchidx$row,
21. layout.pos.col **=** matchidx$col))
22. }
23. }
24. }
26. library(readxl)
27. library(ggplot2)
29. dane **=** read\_excel("Zadanie\_domowe\_nr\_3\_2022\_2023\_KP.xlsx")
30. odmiana **=** dane$odmiana
31. cecha1 **=** dane$cecha\_1
32. cecha2 **=** dane$cecha\_2
33. cecha3 **=** dane$cecha\_3
34. cecha4 **=** dane$cecha\_4
36. odmiana\_1\_cecha1 **=** cecha1[1:30]
37. odmiana\_2\_cecha1 **=** cecha1[31:60]
38. odmiana\_3\_cecha1 **=** cecha1[61:90]
39. odmiana\_4\_cecha1 **=** cecha1[91:120]
40. odmiana\_1\_cecha2 **=** cecha2[1:30]
41. odmiana\_2\_cecha2 **=** cecha2[31:60]
42. odmiana\_3\_cecha2 **=** cecha2[61:90]
43. odmiana\_4\_cecha2 **=** cecha2[91:120]
44. odmiana\_1\_cecha3 **=** cecha3[1:30]
45. odmiana\_2\_cecha3 **=** cecha3[31:60]
46. odmiana\_3\_cecha3 **=** cecha3[61:90]
47. odmiana\_4\_cecha3 **=** cecha3[91:120]
48. odmiana\_1\_cecha4 **=** cecha4[1:30]
49. odmiana\_2\_cecha4 **=** cecha4[31:60]
50. odmiana\_3\_cecha4 **=** cecha4[61:90]
51. odmiana\_4\_cecha4 **=** cecha4[91:120]
53. x1 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha1)) **+** geom\_point() **+**
54. geom\_point(aes(x**=**"pierwsza", y**=**mean(odmiana\_1\_cecha1)), col**=**'red', size**=**4) **+**
55. geom\_point(aes(x**=**"druga", y**=**mean(odmiana\_2\_cecha1)), col**=**'red', size**=**4) **+**
56. geom\_point(aes(x**=**"trzecia", y**=**mean(odmiana\_3\_cecha1)), col**=**'red', size**=**4) **+**
57. geom\_point(aes(x**=**"czwarta", y**=**mean(odmiana\_4\_cecha1)), col**=**'red', size**=**4)
59. x2 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha2)) **+** geom\_point() **+**
60. geom\_point(aes(x**=**"pierwsza", y**=**mean(odmiana\_1\_cecha2)), col**=**'red', size**=**4) **+**
61. geom\_point(aes(x**=**"druga", y**=**mean(odmiana\_2\_cecha2)), col**=**'red', size**=**4) **+**
62. geom\_point(aes(x**=**"trzecia", y**=**mean(odmiana\_3\_cecha2)), col**=**'red', size**=**4) **+**
63. geom\_point(aes(x**=**"czwarta", y**=**mean(odmiana\_4\_cecha2)), col**=**'red', size**=**4)
65. x3 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha3)) **+** geom\_point() **+**
66. geom\_point(aes(x**=**"pierwsza", y**=**mean(odmiana\_1\_cecha3)), col**=**'red', size**=**4) **+**
67. geom\_point(aes(x**=**"druga", y**=**mean(odmiana\_2\_cecha3)), col**=**'red', size**=**4) **+**
68. geom\_point(aes(x**=**"trzecia", y**=**mean(odmiana\_3\_cecha3)), col**=**'red', size**=**4) **+**
69. geom\_point(aes(x**=**"czwarta", y**=**mean(odmiana\_4\_cecha3)), col**=**'red', size**=**4)
71. x4 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha4)) **+** geom\_point() **+**
72. geom\_point(aes(x**=**"pierwsza", y**=**mean(odmiana\_1\_cecha4)), col**=**'red', size**=**4) **+**
73. geom\_point(aes(x**=**"druga", y**=**mean(odmiana\_2\_cecha4)), col**=**'red', size**=**4) **+**
74. geom\_point(aes(x**=**"trzecia", y**=**mean(odmiana\_3\_cecha4)), col**=**'red', size**=**4) **+**
75. geom\_point(aes(x**=**"czwarta", y**=**mean(odmiana\_4\_cecha4)), col**=**'red', size**=**4)
77. x **=** multiplot(x1, x2, x3, x4, cols**=**2)
78. print(x)
80. y1 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha1)) **+** geom\_boxplot()
81. y2 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha2)) **+** geom\_boxplot()
82. y3 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha3)) **+** geom\_boxplot()
83. y4 **=** ggplot(dane, aes(x**=**odmiana, y**=**cecha4)) **+** geom\_boxplot()
85. y **=** multiplot(y1, y2, y3, y4, cols**=**2)
86. print(y)
88. z1 **=** ggplot(dane, aes(x**=**cecha1)) **+** geom\_histogram(aes(y**=**..density..)) **+**
89. stat\_function(fun**=**dnorm, args**=**list(mean**=**mean(cecha1), sd**=**sd(cecha1)), col**=**"red")
91. z2 **=** ggplot(dane, aes(x**=**cecha2)) **+** geom\_histogram(aes(y**=**..density..)) **+**
92. stat\_function(fun**=**dnorm, args**=**list(mean**=**mean(cecha2), sd**=**sd(cecha2)), col**=**"red")
94. z3 **=** ggplot(dane, aes(x**=**cecha3)) **+** geom\_histogram(aes(y**=**..density..)) **+**
95. stat\_function(fun**=**dnorm, args**=**list(mean**=**mean(cecha3), sd**=**sd(cecha3)), col**=**"red")
97. z4 **=** ggplot(dane, aes(x**=**cecha4)) **+** geom\_histogram(aes(y**=**..density..)) **+**
98. stat\_function(fun**=**dnorm, args**=**list(mean**=**mean(cecha4), sd**=**sd(cecha4)), col**=**"red")
100. z **=** multiplot(z1, z2, z3, z4, cols**=**2)
101. print(z)
103. h1 **=** ggplot(dane, aes(x**=**cecha1, y**=**length(cecha1))) **+** geom\_violin()
104. h2 **=** ggplot(dane, aes(x**=**cecha2, y**=**length(cecha2))) **+** geom\_violin()
105. h3 **=** ggplot(dane, aes(x**=**cecha3, y**=**length(cecha3))) **+** geom\_violin()
106. h4 **=** ggplot(dane, aes(x**=**cecha4, y**=**length(cecha4))) **+** geom\_violin()
108. h **=** multiplot(h1, h2, h3, h4, cols**=**2)
109. print(h)