



**Vilniaus
universitetas**

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Jūnė Salickaitė
Simona Gelžinytė
Laineda Morkytė

3 laboratorinis darbas

II dalis

Darbo aprašas

Vilnius, 2022

Turiny

Įvadas	3
Tikslas	3
Uždaviniai	3
Tiriamajo darbo pagrindinė dalis	3
Pasirinkti duomenys	3
1 grafikas.....	3
2 grafikas.....	4
3 grafikas.....	5
Grafikų panelė.....	6
Šaltiniai	8
Naudota literatūra.....	8

Ivadas

Tai yra duomenų vizualizavimo laboratorinio darbo aprašas, rašytas Vilniaus Universiteto studentų Jūnės Salickaitės, Simonos Gelžinytės ir Lainedos Morkytės. Šiame darbe pateikta grafikų panelė, naudojant „R“ programos paketus – „ggplot2“, „patchwork“. Užduotis atlikta naudojant duomenis apie krūtų vėžį.

Tikslas

Turint duomenų apie krūties vėžio atvejus Viskonsino valstijoje, JAV palyginti piktybinio ir nepiktybinio krūtų vėžio auglio charakteristikas.

Uždaviniai

1. Išsiaiškinti koks ryšys yra tarp auglio ploto ir įdubimo.
2. Iširti auglio ploto pasiskirstymą.
3. Palyginti invazinių ir neinvazinių auglių perimetrą, spindulį bei tekstūrą.

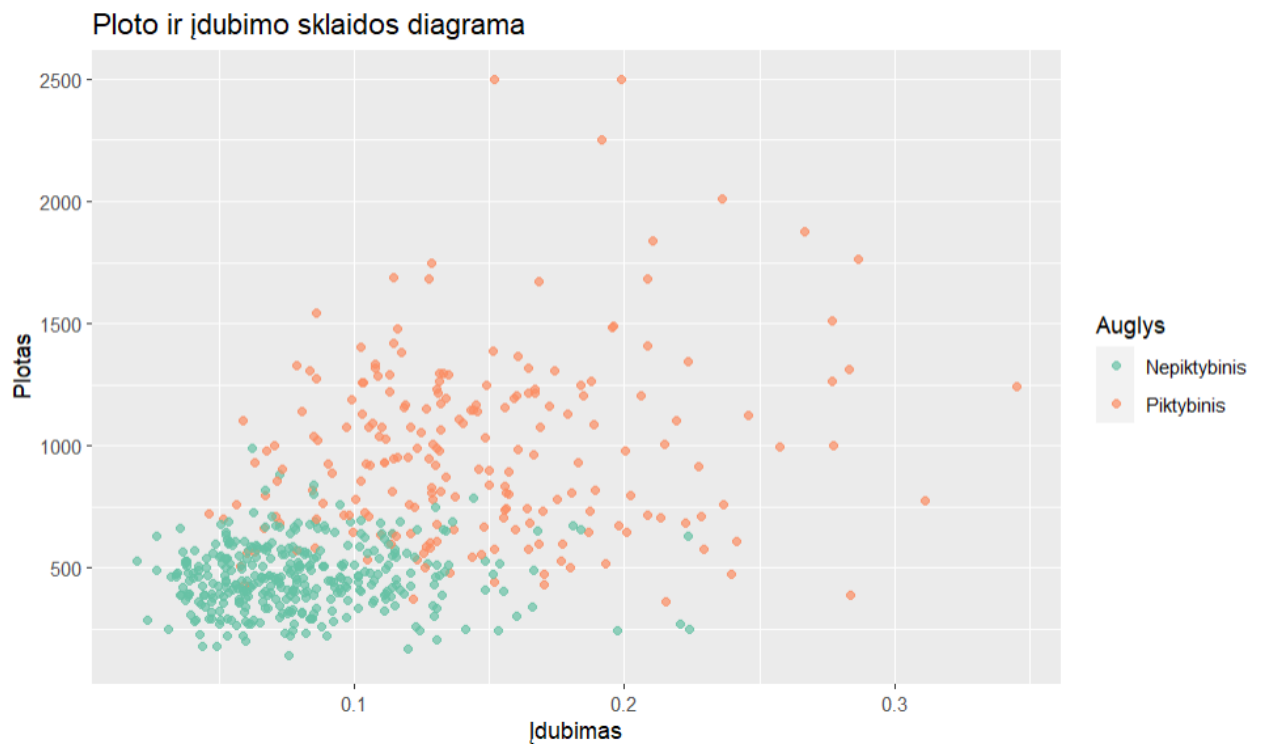
Tiriamąojo darbo pagrindinė dalis

Pasirinkti duomenys

Duomenis naudojome iš „UCI Machine Learning Repository“ apie krūtų vėžį Viskonsino valstijoje. Vertinta pagal auglio perimetrą, spindulį ir tekstūrą. Duomenys sugrupuoti į dvi grupes: piktybinis auglys ir nepiktybinis.

1 grafikas

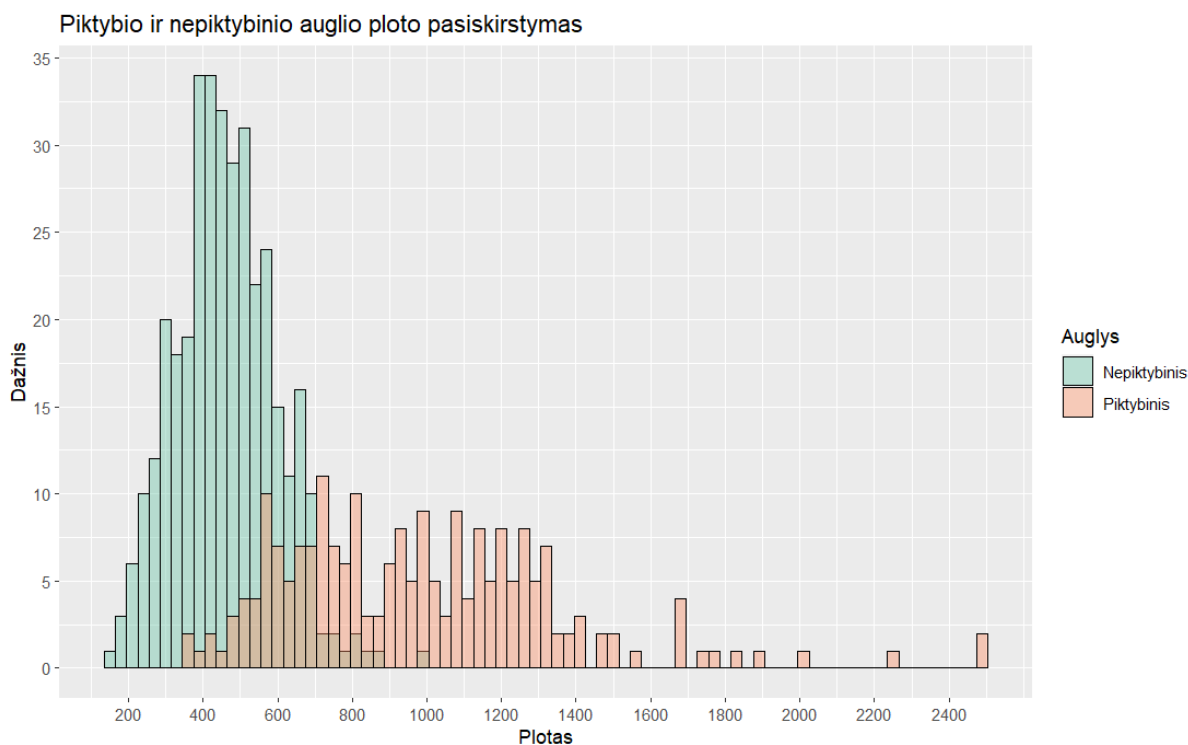
```
df1 <- data[c(2,6,8)]
colnames(df1) <- c('Auglys','Plotas','Įdubimas')
p1 <- ggplot(df1, aes(x=Įdubimas, y=Plotas, color = Auglys)) +
  geom_point(alpha=0.7) +
  ggtitle('Ploto ir įdubimo sklaidos diagrama') +
  scale_color_brewer(palette = "Set2", direction = 1)
p1
```



Išvada: Iš sklaidos diagramos galime matyti, jog kuo didesnis įdubimas, tuo didesnis ir auglio plotas. Taip pat galime pastebėti, jog didesnio ploto ir įdubimo esantys augliai dažniausiai yra piktybiniai, o mažesnio ploto ir įdubimo augliai – nepiktybiniai.

2 grafikas

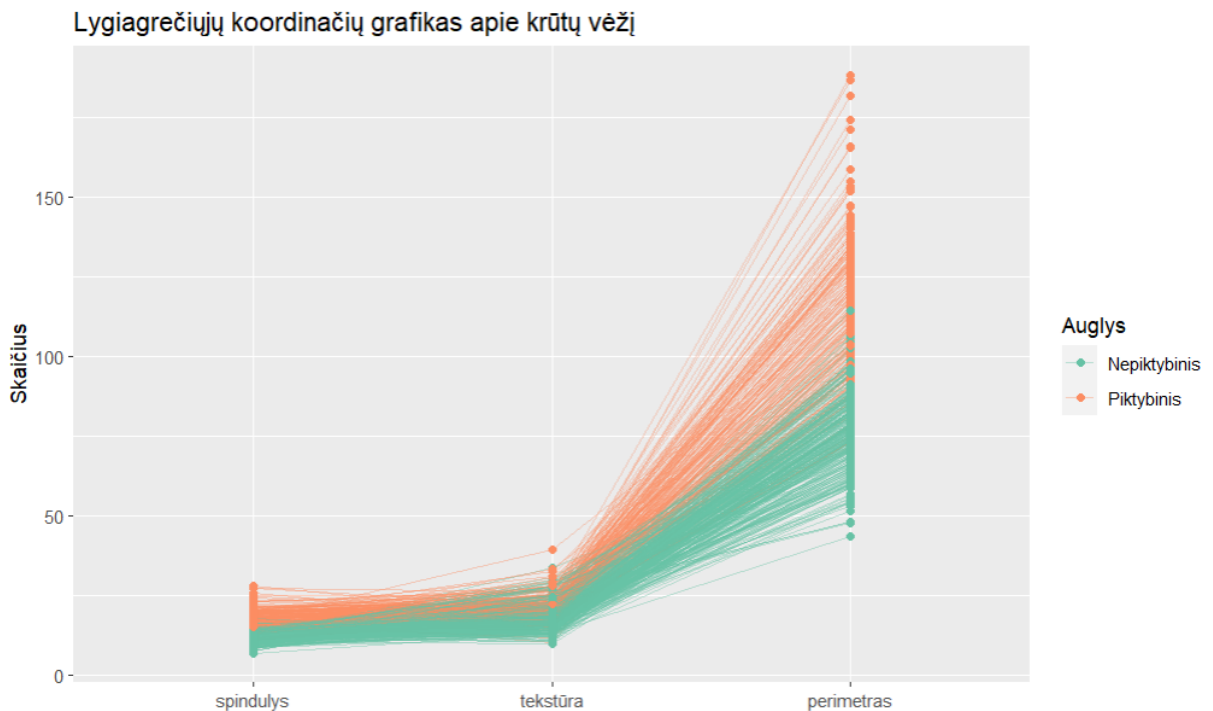
```
p2 <- ggplot(data, aes(x = area, fill = Diagnosis)) +
  geom_histogram(position = "identity", alpha = 0.4, binwidth = 30,
color="black") +
  scale_x_continuous(breaks = seq(200, 2500, 200)) +
  ggtitle("Piktybio ir nepiktybinio auglio ploto pasiskirstymas") +
  labs(y = 'Dažnis', x = 'Plotas', fill = 'Auglys') +
  scale_fill_brewer(palette = "Set2", direction = 1) +
  scale_y_continuous(breaks = seq(0, 40, 5))
p2
```



Išvada: Kaip matome iš grafiko nepiktybinio auglio plotas dažnai yra mažesnis, negu piktybinio, nes yra išsidėstęs kairėje histogramos pusėje, o piktybinis yra labiau pasiskirstęs per visą diagramos diapazoną. Todėl galime daryti išvadą, kad piktybinis auglys plotu dažnai būna didesnis negu nepiktybinio.

3 grafikas

```
df <- data[c(2,3,4,5)] #stulpeliu pavadinimai
colnames(df) <- c('Auglys','spindulys', 'tekstūra', 'perimetras')
# scale - globalminmax -> parodo tikrąsias reikšmes grafike
# alphaLines -> nustato linijų permatumą
# groupColumn -> pagal ką grupuosime duomenis
p3 <- ggparcoord(data = df, columns = 2:4, groupColumn = 1, scale =
  "globalminmax", showPoints = T, alphaLines = 0.4, title =
  "Lygiagrečiųjų koordinačių grafikas apie krūtų vėžį") +
  theme(plot.title = element_text(size=13)) +
  xlab(' ') + ylab('Skaičius') +
  scale_color_brewer(palette = "Set2")
p3
```



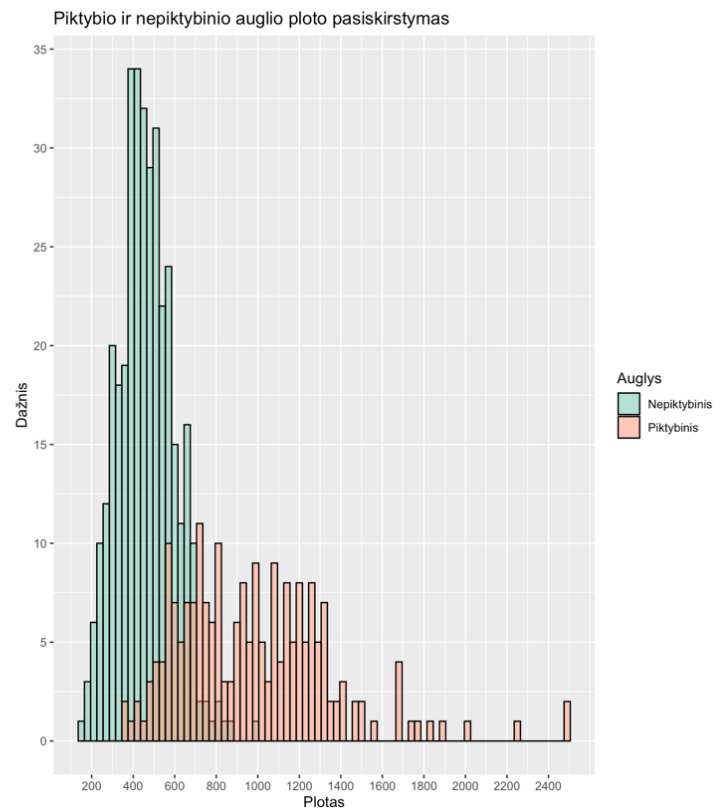
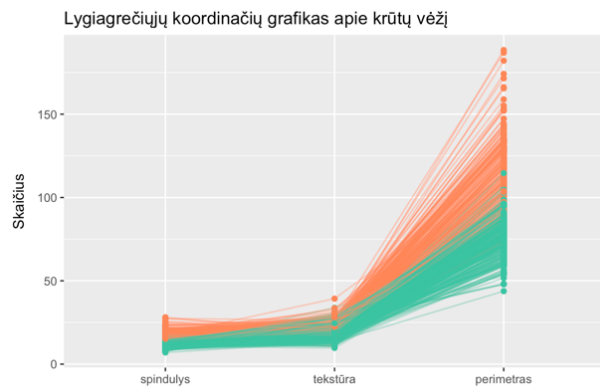
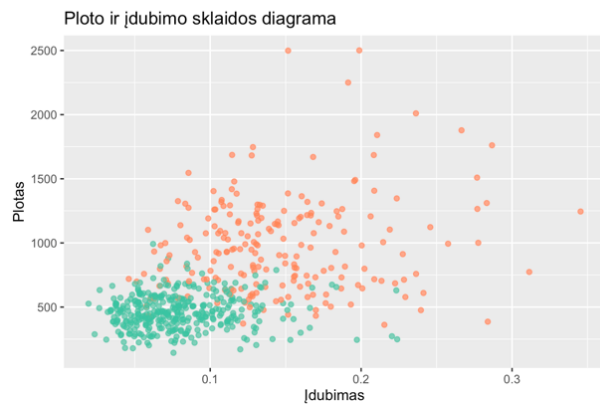
Išvada: iš grafiko galime matyti, jog invazinio auglio spindulys, tekstūra bei perimetras yra didesnis lyginant su neinvaziniu augliu.

Grafikų panelė

Grafikų panelė sudaryta iš 3 skirtingų grafikų. Pirmasis (kairys viršutinis kampas) vaizduoja ryšį tarp auglio ploto ir įdubimo. Antrasis (kairysis apatinis kampas) vaizduoja kiekvieno auglio atskirai spindulį, tekstūrą bei perimetrą. Trečiajame parodomas auglio plotų pasiskirstymas.

```
p <- (p1 / p3) | p2 + plot_annotation(tag_levels = "I")
p + plot_annotation(title = "Piktybinio ir nepiktybinio krūtų vėžio
  auglio charakteristikų palyginimas", theme = theme(plot.title =
  element_text(hjust = 0.5, size = 18)))
```

Piktybinio ir nepiktybinio krūtų vėžio auglio charakteristikų palyginimas



Šaltiniai

Naudota literatūra

- <https://cran.r-project.org/web/packages/egg/vignettes/Ecosystem.html>
- <https://ggplot2-book.org/> (9. Arranging plots)
- https://rpubs.com/Yael_K/344010
- <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>