

MATEMATIKOS IR INFORMATIKOS

FAKULTETAS

Jūnė Salickaitė

Simona Gelžinytė

Laineda Morkytė

**3 laboratorinis darbas**

**II dalis**

Darbo aprašas

Vilnius, 2022

Turinys

[**Įvadas** 3](#_Toc101379938)

[**Tikslas** 3](#_Toc101379939)

[**Uždaviniai** 3](#_Toc101379940)

[**Tiriamojo darbo pagrindinė dalis** 3](#_Toc101379941)

[**Pasirinkti duomenys** 3](#_Toc101379942)

[**1 grafikas** 3](#_Toc101379943)

[**2 grafikas** 4](#_Toc101379944)

[**3 grafikas** 5](#_Toc101379945)

[**Grafikų panelė** 6](#_Toc101379946)

[**Šaltiniai** 8](#_Toc101379947)

[**Naudota literatūra** 8](#_Toc101379948)

# **Įvadas**

Tai yra duomenų vizualizavimo laboratorinio darbo aprašas, rašytas Vilniaus Universiteto studenčių 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. Ištirti auglio ploto pasiskirstymą.
3. Palyginti invazinių ir neinvazinių auglių perimetrą, spindulį bei tekstūrą.

# **Tiriamojo 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'**)**

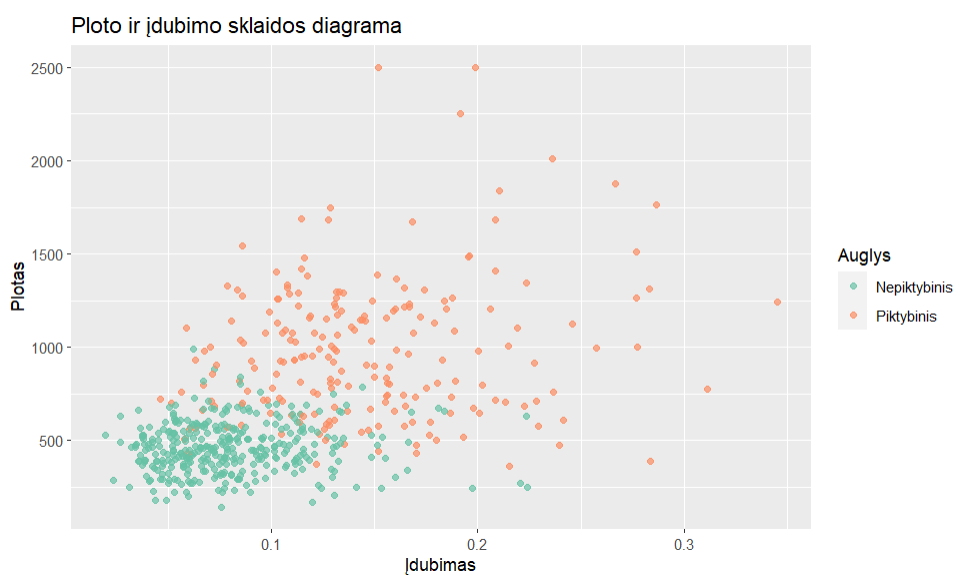
p1 **<-** ggplot**(**df1, aes**(**x**=**Įdubimas, y**=**Plotas, color **=** Auglys**))** **+**

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**))** **+**

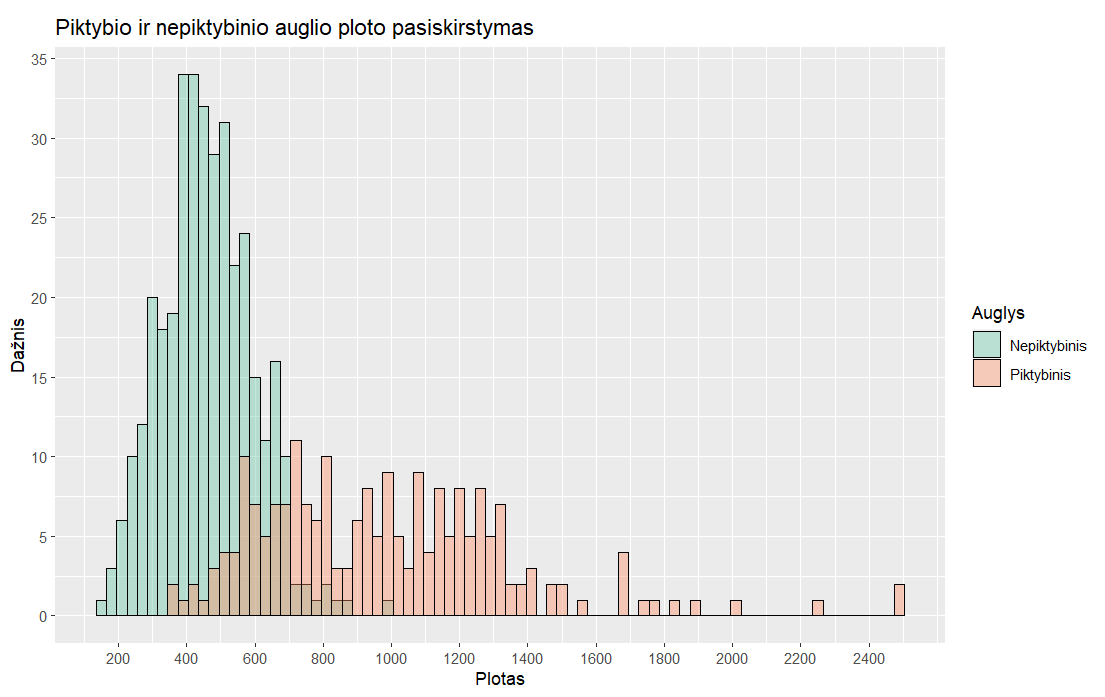
ggtitle**(**"Piktybio ir nepiktybinio auglio ploto pasiskirstymas"**)** **+**

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 permatomumą

# groupColumn -> pagal ką grupuosime duomenis

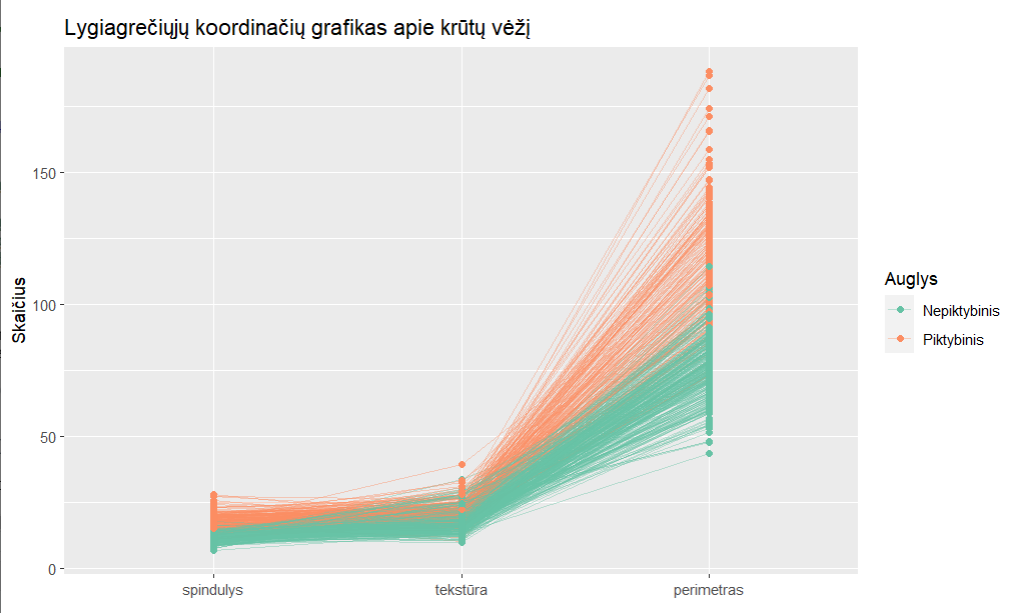
p3 **<-** ggparcoord**(**data **=** df, columns **=** 2**:**4, groupColumn **=** 1, scale **=** "globalminmax", showPoints **=** T , alphaLines **=** 0.4, title **=** "Lygiagrečiųjų koordinačių grafikas apie krūtų vėžį"**)** **+**

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, tektūrą 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**)))**

Chart

Description automatically generated

# **Šaltiniai**

## **Naudota literatūra**

* <https://cran.r-project.org/web/packages/egg/vignettes/Ecosystem.html>
* [https://ggplot2-book.org/](https://ggplot2-book.org/%20) (9. Arranging plots)
* <https://rpubs.com/Yael_K/344010>
* <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>