



**Vilniaus  
universitetas**

# MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Jūnė Salickaitė  
Simona Gelžinytė  
Laineda Morkytė

## **2 laboratorinis darbas**

Darbo aprašas

Vilnius, 2022

## **Turinys**

|  |           |
|--|-----------|
| <b>Įvadas.....</b>   | <b>3</b>  |
| <b>Tikslas.....</b>  | <b>3</b>  |
| <b>Uždaviniai.....</b>   | <b>3</b>  |
| <b>Tiriamąo darbo pagrindinė dalis .....</b>                               | <b>3</b>  |
| <b>Pasirinkti duomenys .....</b>   | <b>3</b>  |
| <b>Pirma užduotis .....</b>  | <b>4</b>  |
| <b>Vidutinė oro temperatūra Lietuvoje 2000 – 2013 metais .....</b>         | <b>4</b>  |
| <b>Vidutinė oro temperatūra Šiaurės Europoje 2000 metais.....</b>          | <b>4</b>  |
| <b>Vidutinė oro temperatūra šešiuose žemynuose 2000 m. ir 2013 m. ....</b> | <b>5</b>  |
| <b>Antra užduotis.....</b>   | <b>7</b>  |
| <b>Trečia užduotis.....</b>  | <b>8</b>  |
| <b>Ketvirta užduotis .....</b>   | <b>9</b>  |
| <b>Šaltiniai .....</b>   | <b>11</b> |

## **Ivadas**

Tai yra duomenų vizualizavimo laboratorinio darbo aprašas, rašytas Vilniaus Universiteto studentų Jūnės Salickaitės, Simonos Gelžinytės ir Lainedos Morkytės. Šiame darbe pateikti 3 tiesioginio vizualizavimo metodo grafikai, 1 lygiagrečių koordinačių, žvaigždžių bei koreliacijos koeficientų vizualizavimai, naudojant „R“ programos paketus – „ggplot2“, „corrplot“. Užduotys atliktos naudojant duomenis:

- 1) apie vidutinę oro temperatūrą šalyse bei žemynuose 2000-2013 metais;
- 2) apie krūtų vėžį.

## **Tikslas**

1. Turint duomenų rinkinį apie vidutinę oro temperatūrą išsiaiškinti kokia oro temperatūra yra skirtinguose žemynuose, Rytų Europoje bei Lietuvoje skirtingu laikotarpiu.
2. Turint duomenų apie krūtų vėžio atvejus Viskonsino valstijoje, JAV nustatyti bei palyginti invazinių ir neinvazinių auglių perimetrą, spindulį bei tekstūrą. Taip pat atvaizduoti grafike kaip skiriasi auglių spindulys, tekstūra ir perimetras tų asmenų, kurių auglys yra piktybinis. Bei pagal duotus kintamuosius, išsiaiškinti kaip kai kurie iš jų koreliuoja vienas su kitu.

## **Uždaviniai**

1. Nubrėžti 3 grafikus naudojantis tiesioginio vizualizavimo metodus.
2. Pavyzdžiu iliustruoti vieną iš geometrinio tiesioginio duomenų vizualizavimo metodą.
3. Pavyzdžiu iliustruoti pasirinktą simbolinį duomenų vizualizavimo metodą.
4. Nubrėžti 2 grafikus naudojantis koreliacijos koeficientų vizualizavimo paketą „corrplot“.

## **Tiriamąjo darbo pagrindinė dalis**

### **Pasirinkti duomenys**

Duomenys pasirinkti iš „kaggle“. Šiame duomenų rinkinyje turime informaciją apie vidutinę oro temperatūrą šalyse bei žemynuose 2000-2013 metais.

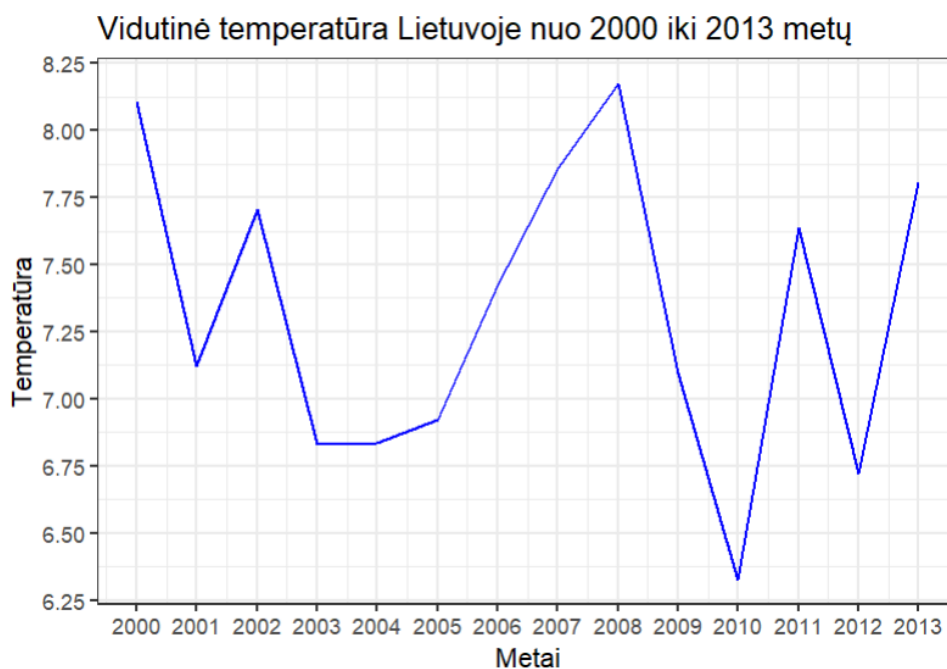
## Pirma užduotis

### Vidutinė oro temperatūra Lietuvoje 2000 – 2013 metais

Grafikas nubraižytas remiantis Lietuvos vidutinės temperatūros (2000 – 2013 metais) duomenimis. Kadangi duomenys kinta laike, pasirinkta linijinė diagrama.

```
#sudaroma lentelė
temperatūra <- c(data$Lithuania)
metai <- c(data$year)
lietuva <- data.frame(metai, temperatūra)

#nubraižoma diagrama
ggplot(lietuva, aes(x=metai, y=temperatūra)) +
  geom_line(color="blue") +
  scale_x_continuous(name="Metai", breaks=seq(2000,2013,1)) +
  scale_y_continuous(name="Temperatūra", breaks=seq(6,8.5,0.25)) +
  ggtitle("Vidutinė temperatūra Lietuvoje nuo 2000 iki 2013 metų") +
  theme_bw()
```



### Vidutinė oro temperatūra Šiaurės Europoje 2000 metais

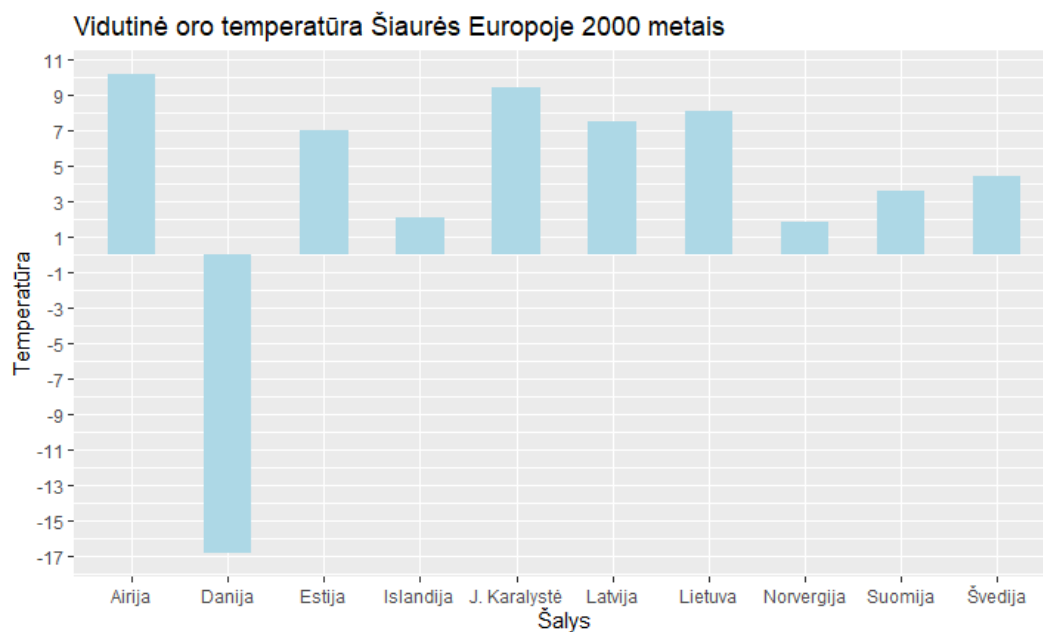
Grafikas nubraižytas naudojantis informaciją apie Šiaurės Europos šalių vidutinę oro temperatūrą 2000 metais.

```
Sys.setlocale("LC_ALL", "Lithuanian") #lietuviškos raidės
install.packages('ggplot2')
library(ggplot2)
```

```

#pasiimame 2000m. temp. eilutę
metai_2000 <- data[1,]
#išrenkame Šiaurės Europos šalis
temperatura <- c(metai_2000$Norway, metai_2000$Sweden, metai_2000$Denmark,
metai_2000$Finland, metai_2000$Iceland, metai_2000$United.Kingdom,
metai_2000$Ireland, metai_2000$Lithuania, metai_2000$Latvia,
metai_2000$Estonia)
#pavadiname duomenis
Salys <- c('Norvergija', 'Švedija', 'Danija', 'Suomija', 'Islandija',
           'J. Karalystė', 'Airija', 'Lietuva', 'Latvija', 'Estija')
North <- data.frame(Salys,temperatura) #sudedame reikiamus duomenis į
atskirą lentelę
ggplot(North, aes(x=Salys, y=temperatura)) + geom_bar(stat="identity", fill
= "lightblue", width = 0.5 ) +
  ggtitle("Vidutiniška oro temperatūra Šiaurės Europoje 2000 metais") +
  labs(y = 'Temperatūra', x = 'Šalys') +
  scale_y_continuous(breaks = seq(-17, 11,2))

```



### Vidutiniška oro temperatūra šešiuose žemynuose 2000 m. ir 2013 m.

Temperatūroms atvaizduoti yra pasirinkta grupuota stulpelinė diagrama, kuri leidžia lengviau pamatyti skirtumus, kaip per 13 metų pasikeitė vidutiniška oro temperatūra šešiuose skirtinguose pasaulio žemynuose.

```

Sys.setlocale("LC_ALL", "Lithuanian") #lietuviškos raidės
#pasiimame 2000m. ir 2013m. temperatūrų eilutes
metai_2000 <- data[1,]
metai_2013 <- data[14,]

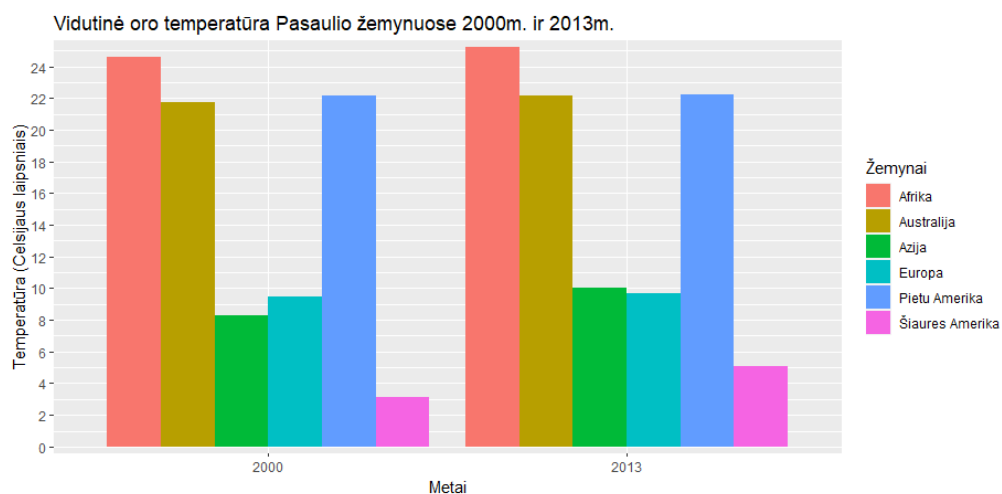
#išsirenkame žemynus (atsizvelgiant į metus)
temperatura <- c(metai_2000$Africa, metai_2000$Asia, metai_2000$Australia,
metai_2000$Europe, metai_2000$North.America, metai_2000$South.America,
metai_2013$Africa, metai_2013$Asia, metai_2013$Australia,
metai_2013$Europe, metai_2013$North.America, metai_2013$South.America)

metai <- c(rep('2000', 6), rep('2013', 6))
#metai kartojami po 6 kartus, nes yra 6 žemynų informacija apie temperatūrą.

#pavadiname išrinktus duomenis
zemynai <- c('Afrika', 'Azija', 'Australija', 'Europa', 'Šiaurės
Amerika', 'Pietų Amerika')

#naudojamus duomenis sudedame į atskirą lentelę
lentele <- data.frame(temperatura, metai, zemynai)
ggplot(lentele, aes(x=metai, y=temperatura, group=zemynai, fill=zemynai)) +
geom_bar(position='dodge', stat = 'identity') +
ggtitle("Vidutinė oro temperatūra Pasaulio žemynuose 2000m. ir 2013m.") +
labs(y = 'Temperatūra (Celsijaus laipsniais)', x = 'Metai') +
scale_fill_discrete(name = 'Žemynai')+
scale_y_continuous(expand=c(0,0.4), breaks = seq(0, 27, 2))

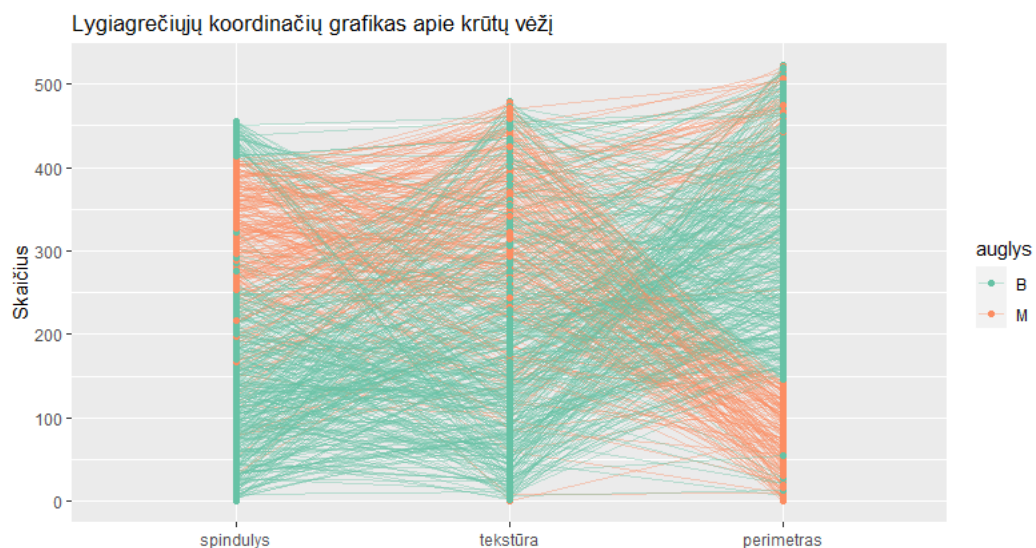
```



## Antra užduotis

Šioje užduotyje reikėjo pasirinkti vieną iš geometrinių vizualizavimo būdų. Pasirinkome lygiagrečių koordinatų metodą. 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 (malignant) auglys ir nepiktybinis (benign).

```
#stulpeliai, kuriuos naudosime
df <- data[c(2,3,4,5)]
#stulpeli pavadinimai
colnames(df) <- c('auglys','spindulys', 'tekstra', 'perimetras')
# scale - globalminmax -> parodo tikrąsias reikšmes grafike
# alphaLines -> nustato linijų permatomumą
# groupColumn -> pagal ką grupuosime duomenis
# M - piktybinis auglys (malignant)
# B - nepiktybinis auglys (benign)
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")
```



## Trečia užduotis

Simboliniam vaizdavimui pasirinkta naudoti žvaigždžių grafiką, kuriame yra vaizduojama auglio spindulys, tekstūra ir perimetras tarp asmenų turinčių piktybinį auglį.

Grafiko pavaizdavimui yra išsirinkti tik piktybinį auglį turintys žmonės, naudojantis „stars“ funkcija nubraižytas grafikas bei įjungtas parametras, vaizduojantis „žvaigždes“ pagal plotus.

```
library(ggplot2)
install.packages("dplyr")
library(dplyr)
library(GGally)
data$Diagnosis <- factor(data$Diagnosis,
                          levels = c("M", "B"),
                          labels = c("piktybinis", "nepiktybinis"))

#išsirenkami duomenų stulpeliai, kuriuos norime vaizduoti grafike
naudojami_duom <- data.frame(data[3], data[4], data[5], data[2])

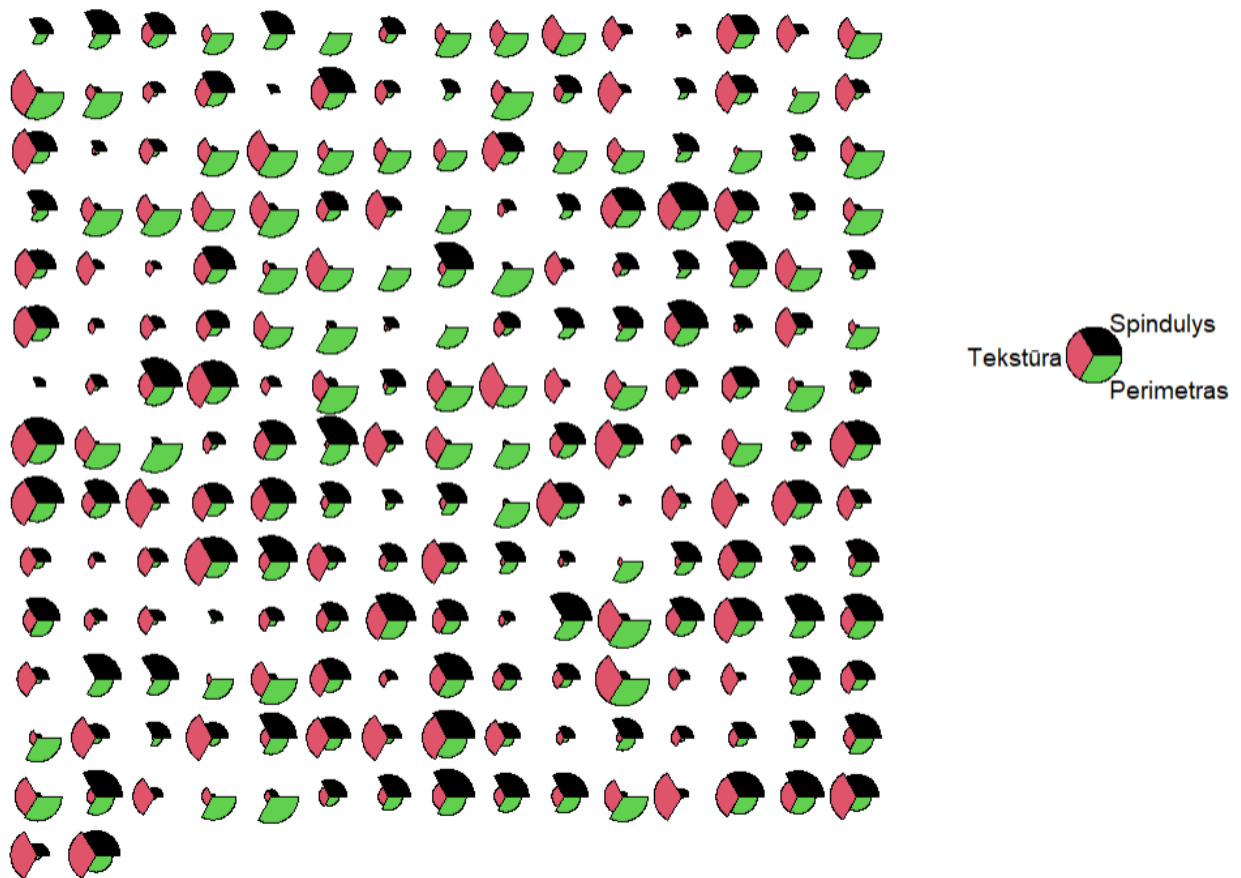
#%in% naudojama pasirinkti tik tuos duomenis, kur auglys yra piktybinis
piktybinis1 <- filter(naudojami_duom, data$Diagnosis %in% c("piktybinis"))

simbolinis <- data.frame(piktybinis1[1], piktybinis1[2], piktybinis1[3])

legenda <- c('Spindulys', 'Tekstūra', 'Perimetras')
stars(simbolinis, main = "Piktybinis auglys", key.labels = legenda,
draw.segments = T, key.loc=c(40,20))
```



## Piktybinis auglys



## Ketvirta užduotis

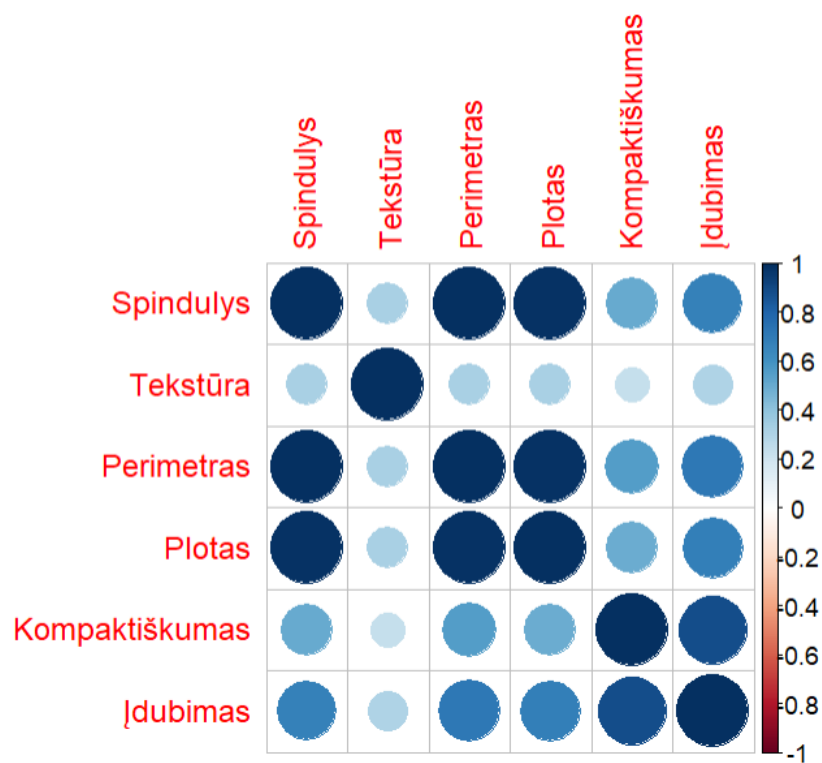
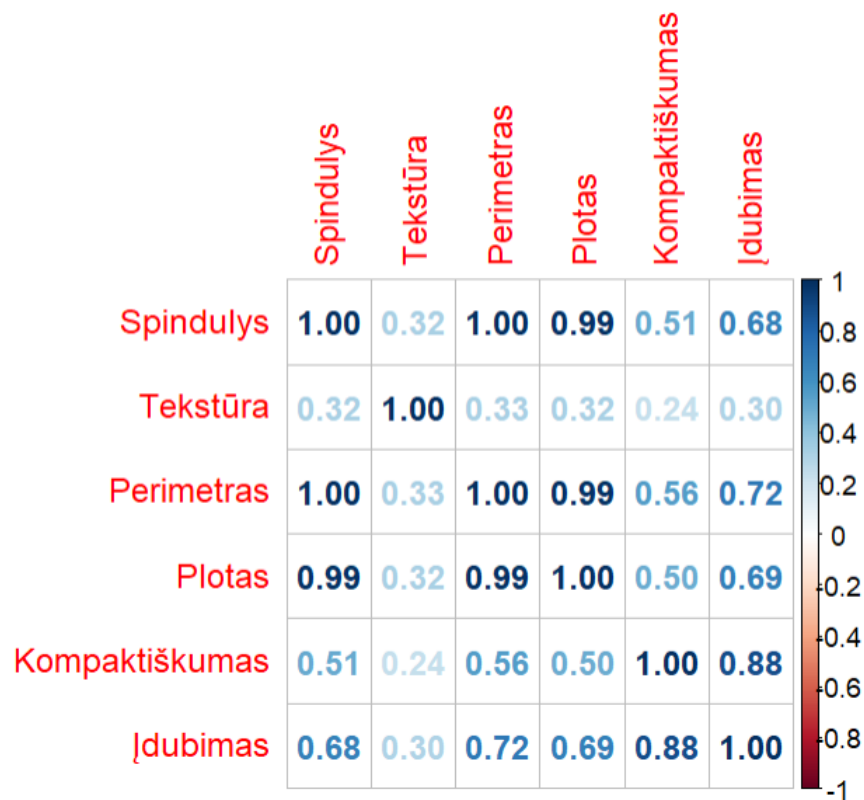
Šioje užduotyje reikėjo iliustruoti koreliacijos koeficientų vizualizavimą naudojant R paketą `corrplot`. Buvo panaudoti 2 vizualizavimo metodai: `'number'` ir `'circle'`. Jie atitinkamai rodo koreliacijos koeficientus skaičiais ir tam tikro dydžio apskritimais.

```
#sudaroma lentelė
parametr <- data.frame(data1$radius, data1$texture, data1$perimeter,
data1$area, data1$compactness, data1$concavity)
P = cor(parametr)
colnames(P) <- c("Spindulys", "Tekstūra", "Perimetras", "Plotas",
"Kompaktiškumas", "Įdubimas")
rownames(P) <- c("Spindulys", "Tekstūra", "Perimetras", "Plotas",
"Kompaktiškumas", "Įdubimas")
```

```
#braižome
```

```
corrplot(P, method = 'number')
```

```
corrplot(P, method = 'circle')
```



## Šaltiniai

### Naudota literatūra

- <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>
- <https://www.kaggle.com/akshaychavan/average-temperature-per-country-per-year/data>
- <https://www.r-graph-gallery.com/>
- <https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>