



**Matematikos ir
informatikos
fakultetas**

Tiesioginio duomenų vizualizavimo metodai

2-as laboratorinis darbas

Darbo tikslas ir uždaviniai

Tikslas:

Įvertinti ir grafiškai pavaizduoti Adelės, Papuasinio ir Antarktinio pingvinų išvaizdos ypatybes, naudojant skirtingus tiesioginio duomenų vizualizavimo metodus.

Uždaviniai:

1. Iširti 3-ų rūšių pingvinų masės ir pelekų ilgio priklausomybę
2. Naudojant **lygiagrečiąsias koordinates**, įvertinti pingvinų rūšis pagal jų savybių panašumą.
3. Palyginti skirtingas pingvinų rūšis, **žvaigždžių metodu** vizualizuojant pirmųjų penkių kiekvienos rūšies pingvinų duomenis.
4. Parodyti stebėtų kintamųjų tarpusavio sąryšį, vizualizuojant **koreliacijos koeficientus**.

Duomenys

```
install.packages("palmerpenguins")  
library(palmerpenguins)  
data(penguins)
```

Stebėti kintamieji:

- species – pingvino rūšys



Adelès pingvinas

Papuasinis pingvinas
(Gentoo penguin)

Antarktinis pingvinas
(Chinstrap penguin)



R version 4.1.3



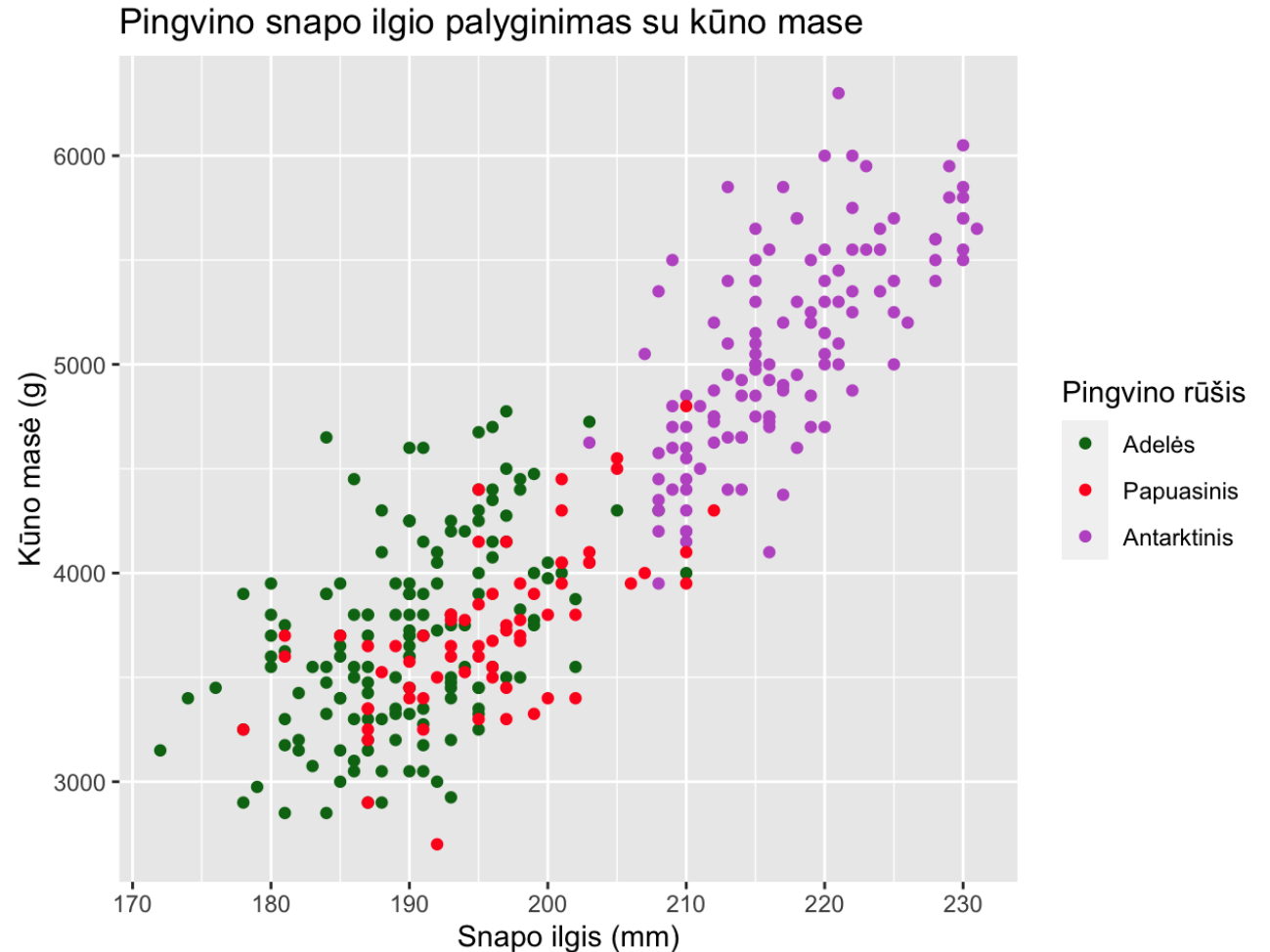
- island – sala Antarktidoje, kurioje matuoti pingvinai (Dream, Torgersen, Biscoe)
- bill_length_mm – snapo ilgis (mm)
- bill_depth_mm – snapo plotis (mm)
- flipper_length_mm – peleko ilgis (mm)
- body_mass_g – kūno masė (gramais)
- sex – pingvino lytis
- year – gimimo metai

1-oji užduotis. Sklaidos diagrama



```
library(ggplot2)
```

```
ggplot(data = penguins) +  
  geom_point(mapping = aes(x =  
    flipper_length_mm, y = body_mass_g, color =  
    species)) +  
  scale_color_manual(values = c('#057314',  
    '#ff0021', '#bf5ccb'), na.translate = FALSE,  
    name = "Pingvino rūšis", labels = c("Adelès",  
    "Papuasinis", "Antarktinis")) +  
  labs(title = " ", x = "Snapo ilgis (mm)",  
    y = "Kūno masė (g)", color = "Pingvino  
    rūšis") +  
  labs(title = "Pingvino snapo ilgio palyginimas  
    su kūno mase") +  
  annotate("text", x = 220, y = 3500, label = "  
    ", fontface = "italic", size = 4, angle = 30)
```



1-oji užduotis. Sklaidos diagramų matrica



```
library(ggplot2)
```

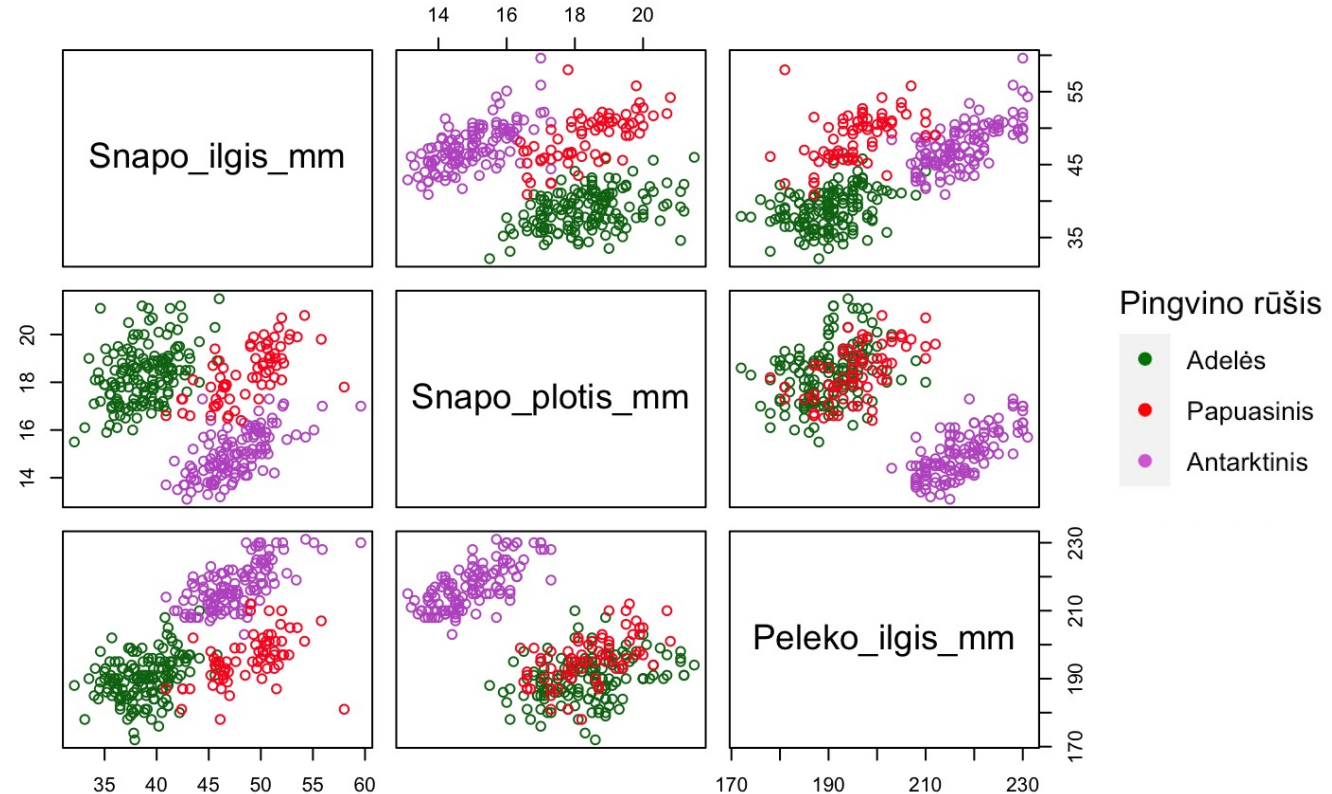
```
data(penguins)
pairs(penguins[,3:5], pch = 19)

df <- penguins[3:5]

species <- penguins[, 5]
pCol <- c('#057314', '#ff0021',
          '#bf5ccb')

colnames(df) <- c("Snapo_ilgis_mm",
                  "Snapo_plotis_mm",
                  "Peleko_ilgis_mm")

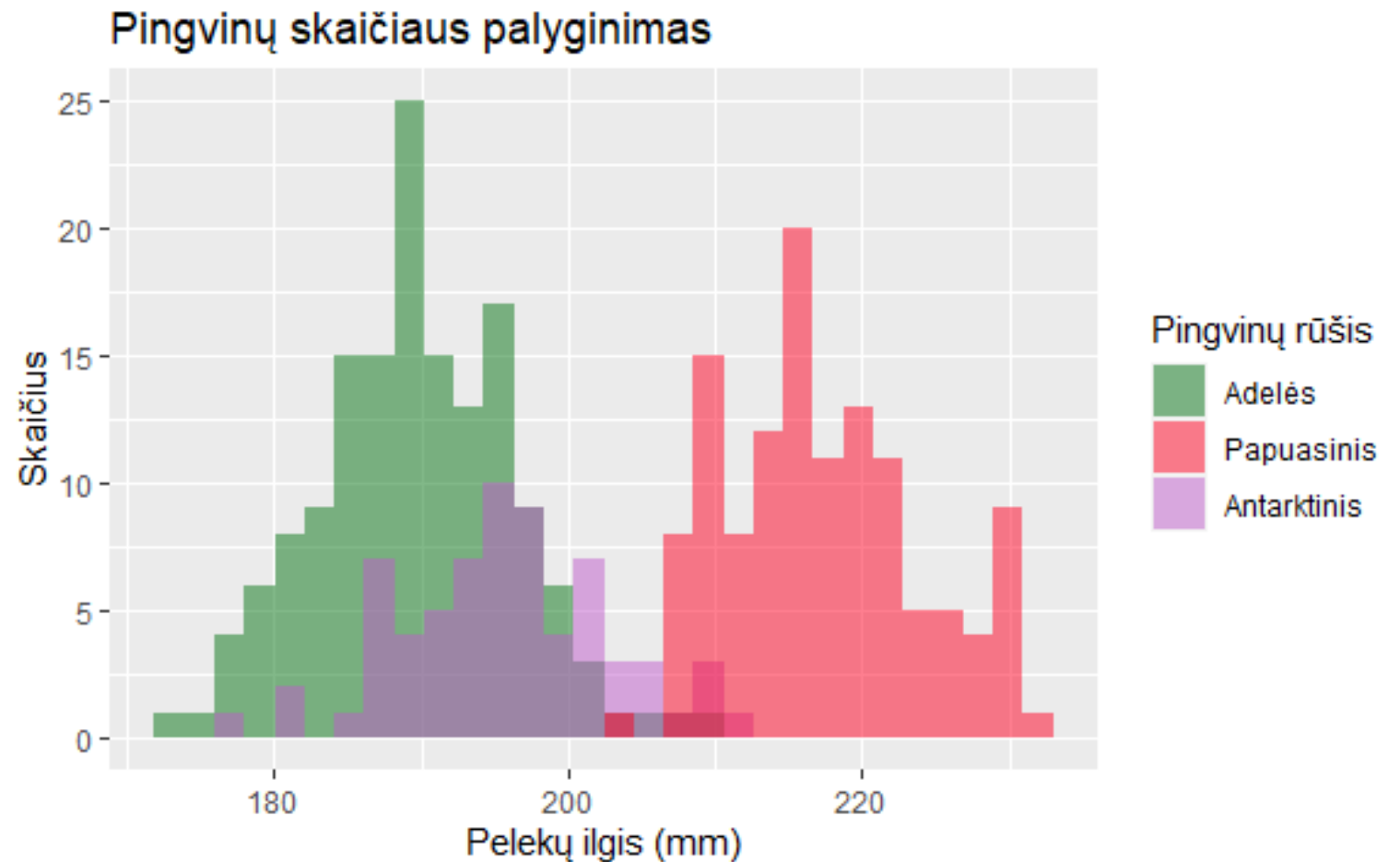
pairs(df, col =
pCol[penguins$species])
```



1-oji užduotis. Histograma

R Studio[®] `library(ggplot2)`

```
ggplot(data = penguins, aes(x =  
  flipper_length_mm)) +  
  geom_histogram(aes(fill = species),  
    alpha = 0.5, position = "identity") +  
  scale_fill_manual(values =  
    c('#057314', '#ff0021',  
      '#bf5ccb'), name = "Pingvinų rūšis",  
    breaks = c("Adelie",  
      "Gentoo", "Chinstrap"),  
    labels = c("Adelės",  
      "Papuasinis", "Antarktinis")) +  
  labs(x = "Pelekų ilgis (mm)", y =  
    "Skaičius", title = "Pingvinų  
    skaičiaus palyginimas")
```

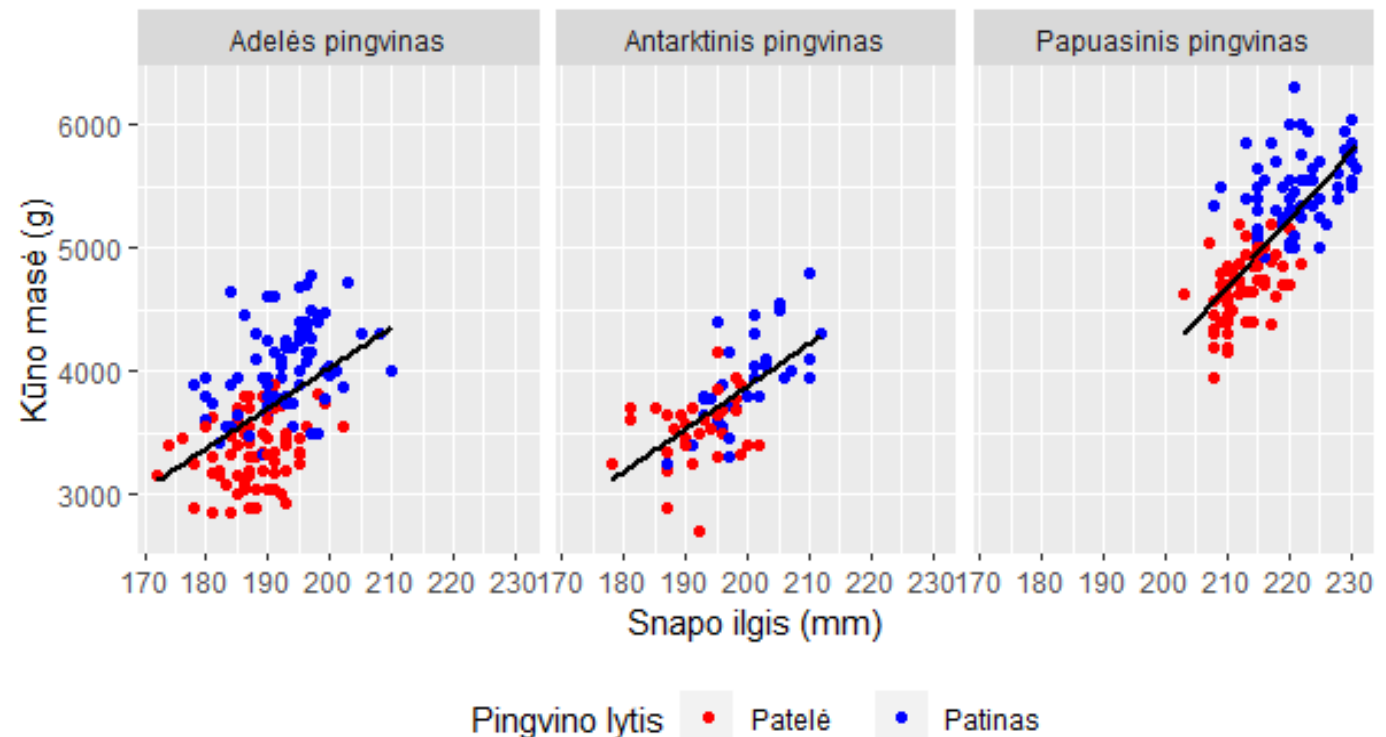


1-oji užduotis. Sklaidos diagrama

 Studio[®] `library(ggplot2)`

```
ggplot(penguins, aes(x = flipper_length_mm, y =  
body_mass_g)) +  
  geom_point(aes(color = sex)) +  
  scale_color_manual(values = c("red", "blue"),  
na.translate = FALSE, name = "Pingvino lytis",  
labels = c("Patelė", "Patinas")) +  
  labs(title = "Pingvino snapas ir kūno masė", x =  
"Snapo ilgis (mm)", y = "Kūno masė (g)",  
color = "Pingviniuko lytis") +  
  theme(legend.position = "bottom",  
plot.title.position = "plot", plot.caption =  
element_text(hjust = 0, face = "italic"),  
plot.caption.position = "plot") +  
  geom_smooth(method = "lm", se = FALSE, color =  
"black") +  
  facet_wrap(~species, labeller =  
labeller(species = c("Adelie" = "Adelės  
pingvinas", "Gentoo" = "Papuasinis pingvinas",  
"Chinstrap" = "Antarktinis pingvinas")))
```

Pingvino snapas ir kūno masė

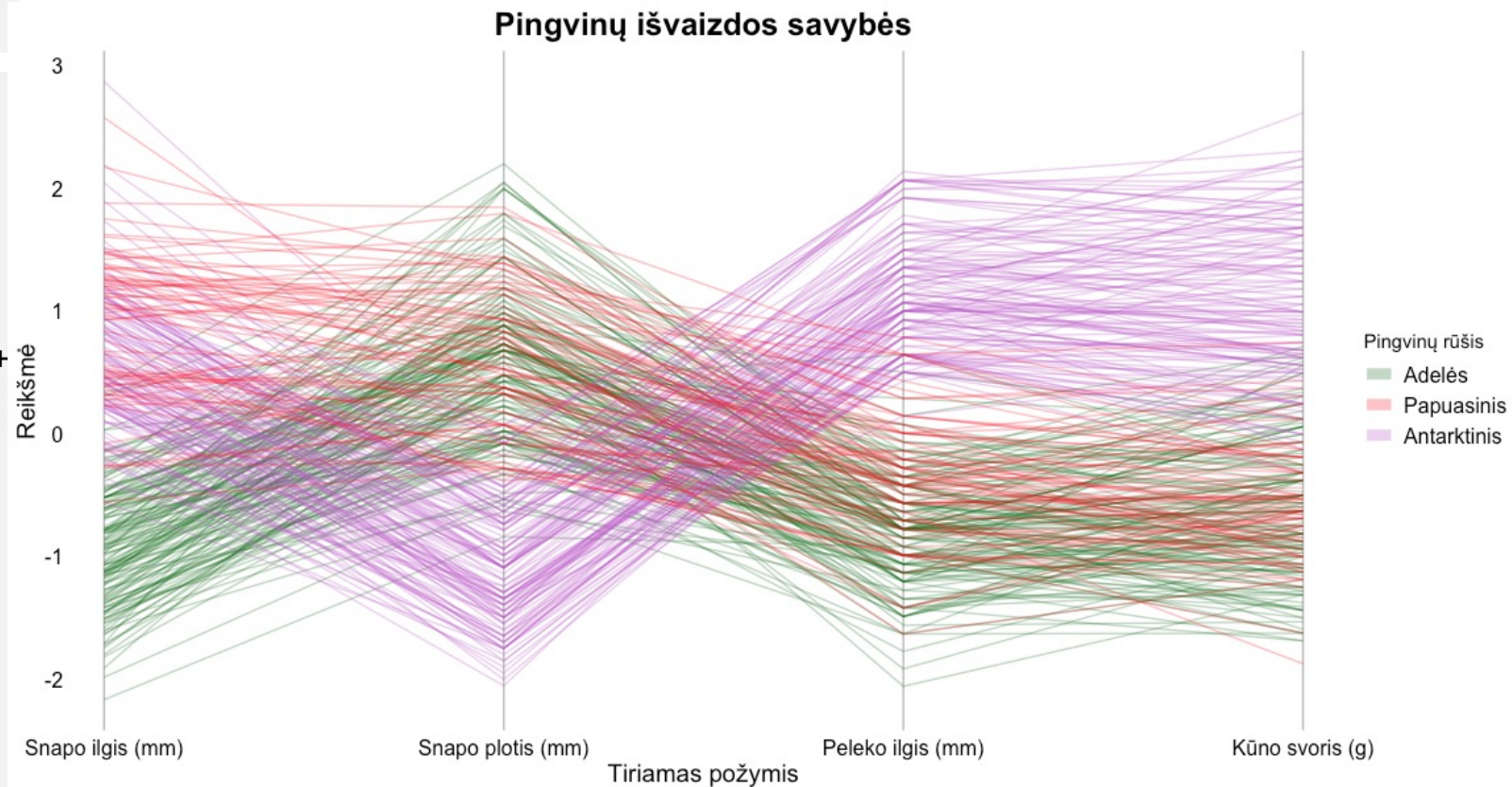


2-oji užduotis. Geometrinis metodas

Lygiagrečiosios koordinatės

R Studio® `library(GGally)`

```
...  
p <- ggparcoord(data = penguins,  
  groupColumn = 1,  
  scale = "std",  
  columns = c(3,4,5,6),  
  title = "Pingvinų  
išvaizdos savybės", alphaLines = 0.3)+  
  xlab('Tiriamas požymis') +  
  ylab('Reikšmė')  
p <- p + scale_color_manual(name =  
  "Rūšys",  
  values =  
  c('#057314', '#ff0021', '#bf5cccb'),  
  labels =  
  c("Adelės pingvinas", "Papuasinis  
pingvinas", "Antarktinis pingvinas"))  
...
```



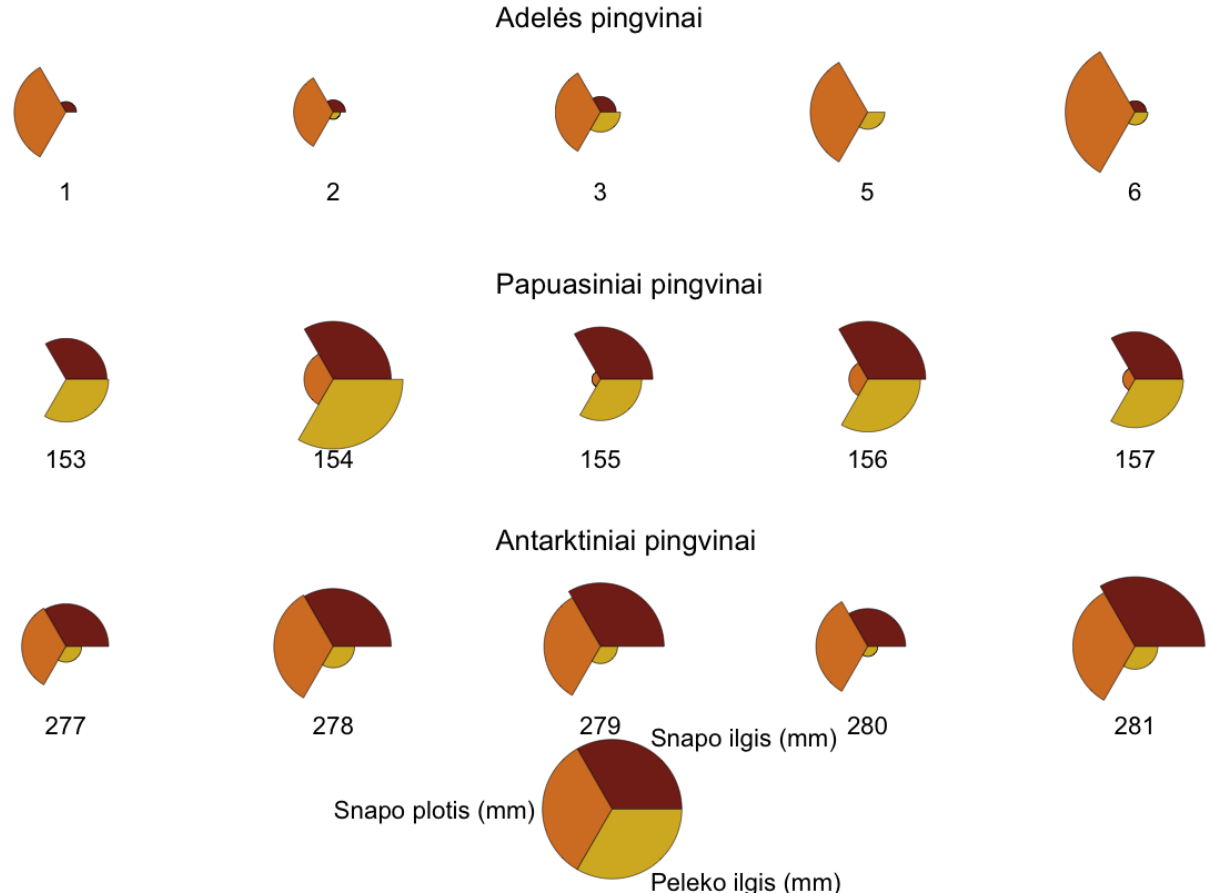
3-oji užduotis. Simbolinis metodas

Žvaigždžių metodas

Pingvinų rūšių išvaizdos savybės

R Studio `library(graphics)`

```
....  
colnames(penguins2)[1] <- "Snapo ilgis (mm)"  
colnames(penguins2)[2] <- "Snapo plotis (mm)"  
colnames(penguins2)[3] <- "Peleko ilgis (mm)"  
  
stars(penguins2, labels =  
  row.names(penguins2),  
      len = 0.6, key.loc = c(7, 0.9),  
      main = "Pingvinų rūšies išvaizdos  
savybės", sub = "subtitle", draw.segments =  
TRUE, frame.plot = F, nrow = 3, cex = .7,  
col.segments = c('#057076', '#ff8301',  
                  '#bf5ccb'))  
  
text(x = 6, y = 7.7, labels = "Adelės  
pingvinai", adj = 0, cex = 0.8)  
text(x = 6, y = 5.4, labels = "Papuasiniai  
pingvinai", adj = 0, cex = 0.8)  
text(x = 6, y = 3.2, labels = "Antarktiniai  
pingvinai", adj = 0, cex = 0.8)
```



4-oji užduotis (1). Koreliacijos koeficientų vizualizavimas

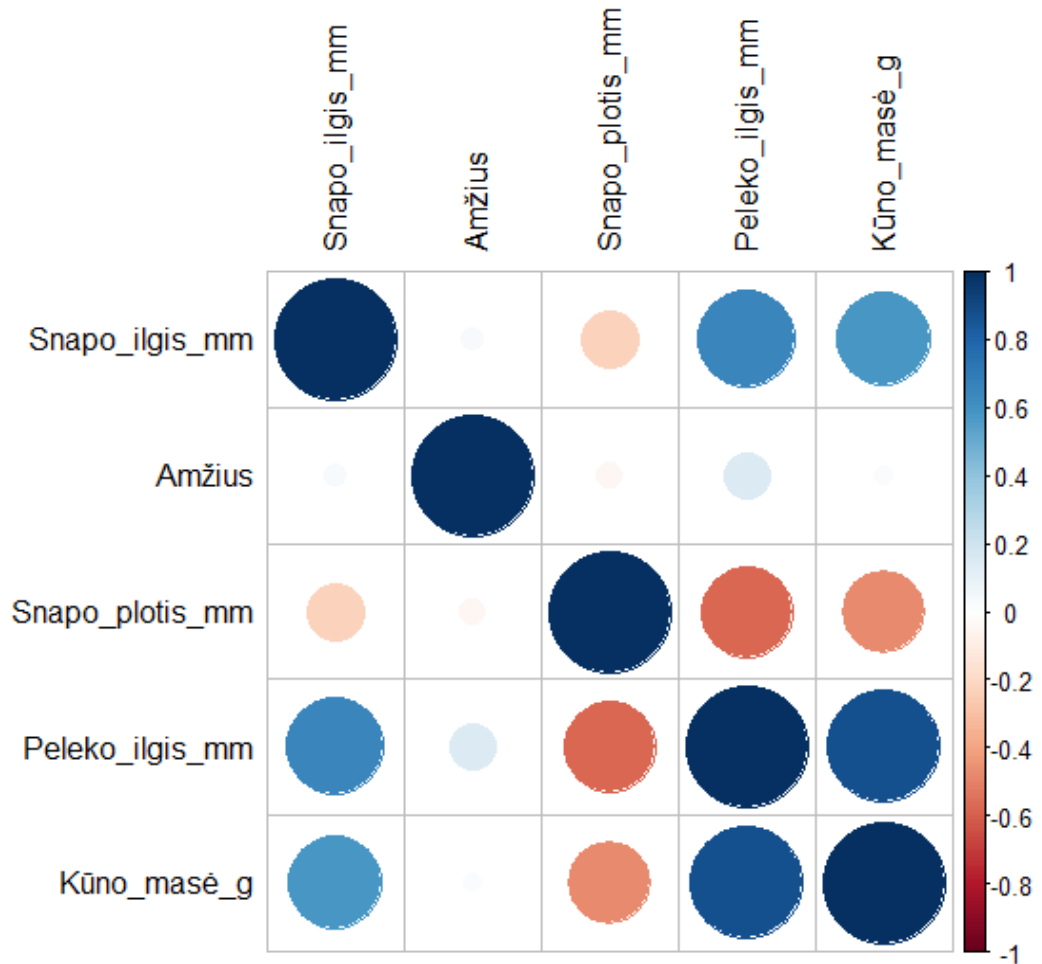


```
library(corrplot)
```

```
peng <- na.omit(penguins)

duomenys <- data.frame("Snapo_ilgis_mm" =
  peng$bill_length_mm,
  "Amžius" = peng$year,
  "Snapo_plotis_mm" = peng$bill_depth_mm,
  "Peleko_ilgis_mm" =
  peng$flipper_length_mm,
  "Kūno_masė_g" = peng$body_mass_g)

M <- cor(duomenys)
corrplot(M, tl.col = 'black')
```



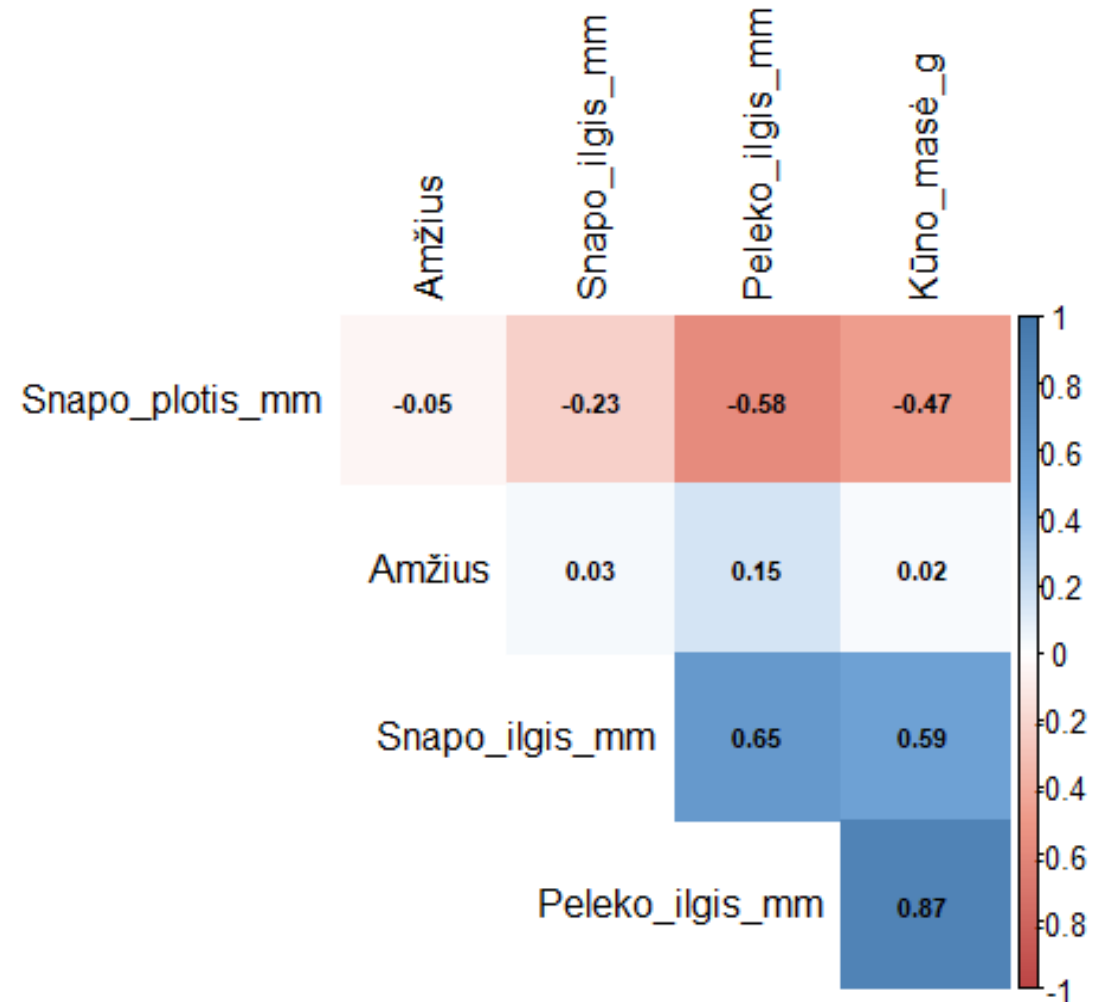
4-oji užduotis (2). Koreliacijos koeficientų vizualizavimas



```
library(corrplot)
```

```
peng <- na.omit(penguins)
duomenys <- data.frame("Snapo_ilgis_mm" =
  peng$bill_length_mm, "Amžius" = peng$year,
  "Snapo_plotis_mm" =
  peng$bill_depth_mm, "Peleko_ilgis_mm" =
  peng$flipper_length_mm, "Kūno_masė_g" =
  peng$body_mass_g)

M <- cor(duomenys)
col <- colorRampPalette(c("#BB4444", "#EE9988",
  "#FFFFFF", "#77AADD", "#4477AA"))
corrplot(M, method = 'color', col = col(200),
  type = 'upper', tl.col = 'black', cl.cex = 0.8,
  order = 'hclust', addCoef.col = 'black',
  number.cex = 0.7, number.digits = 2,
  diag = FALSE, tl.srt = 90, addColorbar = TRUE,
  colorbar = 'top',
  col.order = 'hclust')
```



Išvados

1. Norėdamos atvaizduoti vienoje plokštumoje dviejų dimensijų duomenis ir analizuoti ryšius tarp dviejų kintamųjų rinkomės **sklaidos diagramas**.
2. Norint palyginti pingvinų rūšių išvaizdos skirtumus, buvo pasirinkti atitinkami duomenų vizualizavimo metodai, tokie kaip: **geometrinis vaizdavimas lygiagrečiomis koordinatėmis**.
3. Vizualizuojant duomenis **žvaigždžių metodu**, parodėme, kad Adelės pingvinus atitinkančios žvaigždės yra mažesnės už kitų dviejų pingvinų rūšių žvaigždes.
4. Norint parodyti stiprumą tarp dviejų kintamųjų rinkomės duomenis atvaizduoti pasitelkiant **koreliacijos matricos** vaizdavimo būdą.

Literatūros ir šaltinių sąrašas

1. R duomenų rinkinys. Nuoroda: <https://cran.r-project.org/web/packages/palmerpenguins/index.html>
2. R paketas corrplot. Nuoroda: <https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>
3. R paketas GGally. Nuoroda: <https://cran.r-project.org/web/packages/GGally/index.html>



**Vilniaus
universitetas**

**Ačiū už dėmesį
KLAUSIMAI?**

