Data vizualisation with **R** (session 4)

Solution of the exercises

2021-09-28

Exercise 4.1

- Insert in a Markdown document the correlation matrix of the numeric variables of the **iris** data.

```r
library("kableExtra")
vs_dt <- as.data.frame(round(cor(iris[, 1:4]), 2))
vs_dt[1:4] <- lapply(vs_dt[1:4], function(x) {
    cell_spec(x, bold = T, color = spec_color(abs(x), scale_from = c(-1, 1)),
            font_size = spec_font_size(abs(x), scale_from = c(0, 1)))
})
kbl(vs_dt, escape = F, align = "c") %>% kable_classic("striped", full_width = F)
```

- Insert the table of results of a regression analysis of the **iris** data

```r
mod_full  <- lm(Sepal.Length ~ ., data = iris)
stargazer::stargazer(mod_full, type = "html",
  title = "Regression results", header = F)
```

Regression results

Dependent variable:

Sepal.Length

Sepal.Width

0.496***

(0.086)

Petal.Length

0.829***

(0.069)

Petal.Width

-0.315**

(0.151)

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1	-0.12	0.87	0.82
Sepal.Width	-0.12	1	-0.43	-0.37
Petal.Length	0.87	-0.43	1	0.96
Petal.Width	0.82	-0.37	0.96	1

Speciesversicolor

-0.724***

(0.240)

Speciesvirginica

-1.023***

(0.334)

Constant

2.171***

(0.280)

Observations

150

R2

0.867

Adjusted R2

0.863

Residual Std. Error

0.307 (df = 144)

F Statistic

188.251*** (df = 5; 144)

Note:

*p<0.1; **p<0.05;** p<0.01

Exercise 4.2

- Find the code which permits to obtain with **ggplot2** this figure:

```r
op <- par(oma = c(1, 1, 0, 1), las = 1)
boxplot(Sepal.Length ~ Species, data = iris)
points(as.numeric(iris$Species) + rnorm(150, 0, 0.1), iris$Sepal.Length)
points(c(1, 2, 3), tapply(iris$Sepal.Length, iris$Species, mean),
 col = "red", pch = 16, cex = 2)
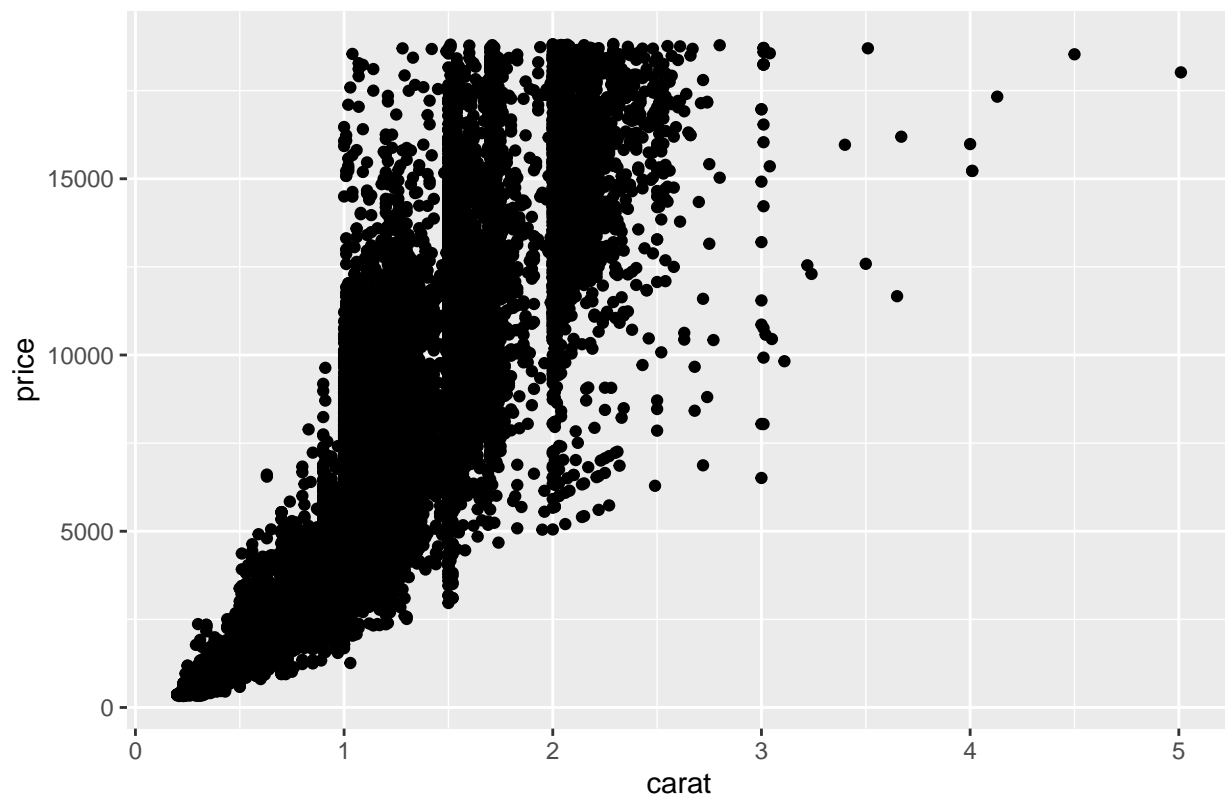```
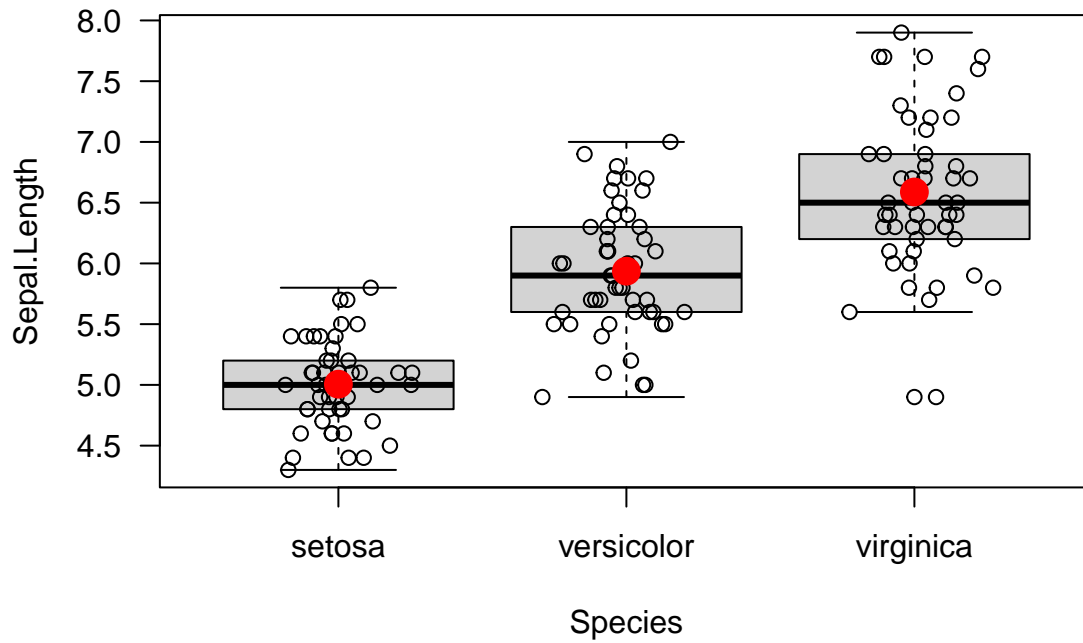
```
par(op)
```

```
library(ggplot2)
ggplot(iris, aes(x = Species, y = Sepal.Length)) +
  geom_boxplot()+
  geom_jitter(position=position_jitter(0.2)) +
  stat_summary(fun=mean, geom="point", shape=20, size=14, color="red", fill="red")
```
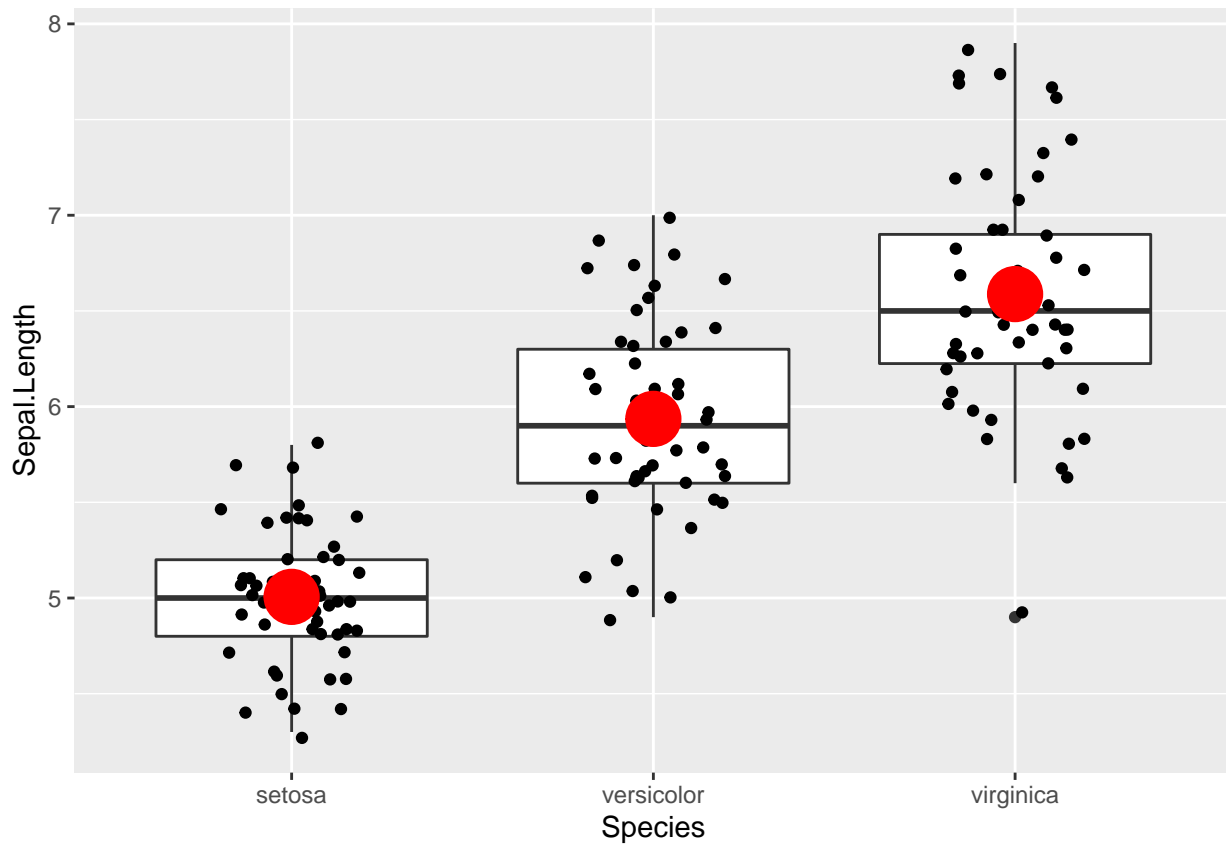
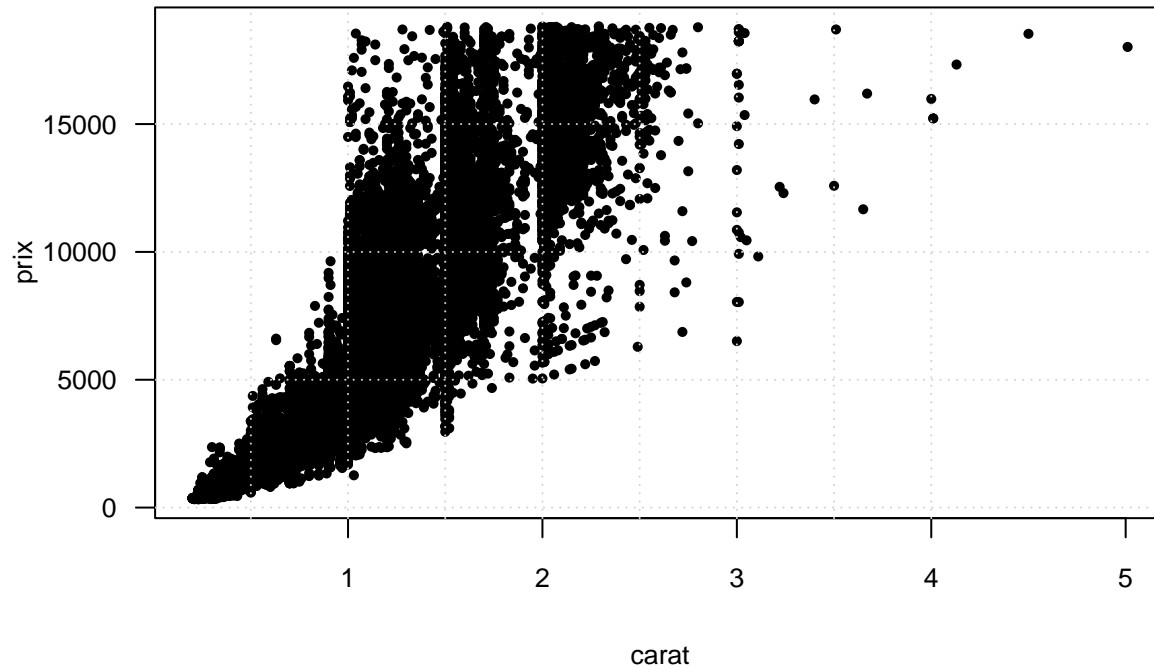- Find the code in **R** base code which permits to obtain this figure:

```r
data("diamonds")
ggplot(diamonds,
       aes(x = carat,
           y = price)) +
  geom_point() +
  ggtitle("My scatter plot")
```

My scatter plot


```r
par(las = 1, cex.axis = 0.8, cex.lab = 0.8)
plot(price ~ carat, data = diamonds, pch = 16, cex = 0.7, xlab = "carat",
    ylab= "prix", main = "Scatter plot")
abline(h = seq(0, 20000, by = 5000), v = seq(0, 4, by = 0.5), col = "lightgray", lty = "dotted")
```
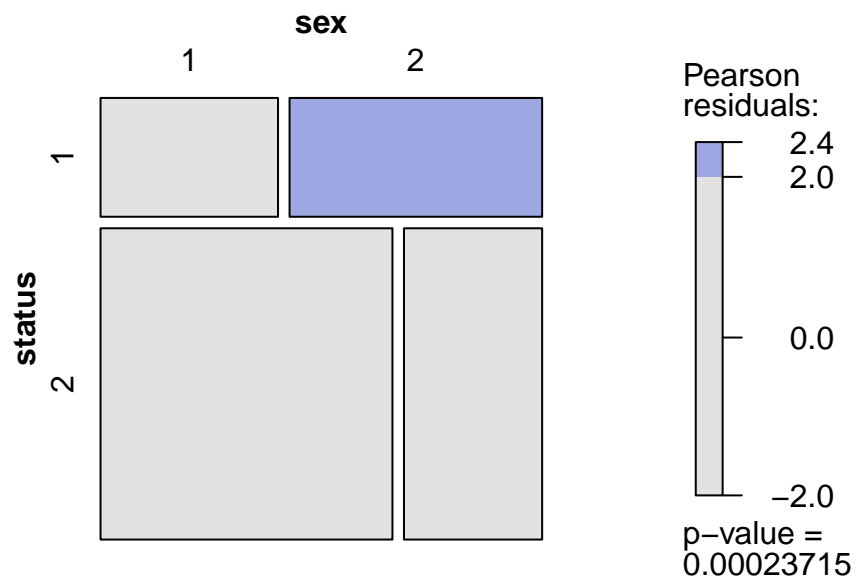
Scatter plot

Exercise 4.3

- On the **lung** data used previously, make a mosaic plot between **status** and **sex** variable.

```r
library("survival")
data(lung)
```

```
## Warning in data(lung): data set 'lung' not found
```

```r
tab <- xtabs(~ status + sex, lung)
vcd::mosaic(tab, shade = TRUE, legend = TRUE)
```
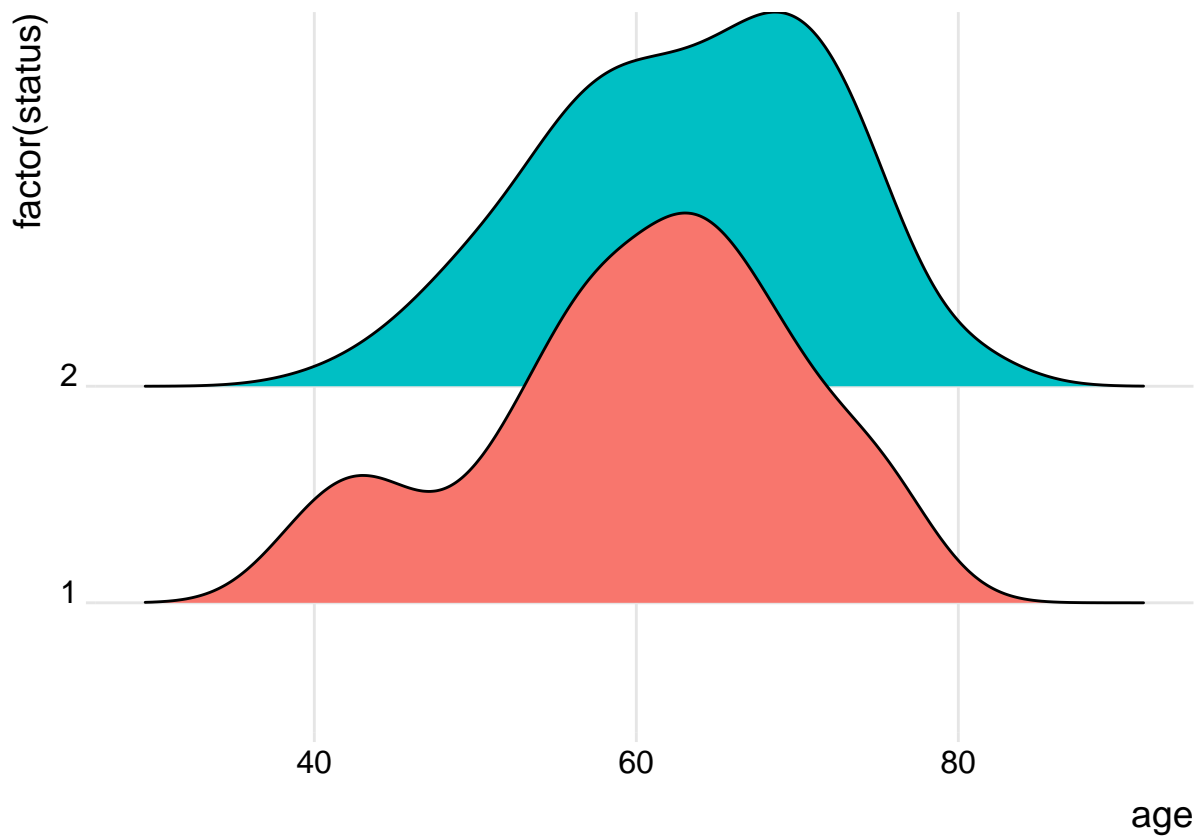

- On the **lung** data, make a ridge plot of variable **age** with respect to **status**.

```
library(ggridges)
ggplot(lung) +
  aes(x = age, y = factor(status), fill = factor(status)) +
  geom_density_ridges() +
  theme_ridges() +
  labs("Age by death/live") +
  theme(legend.position = "none")
```

Picking joint bandwidth of 3.17

- Make a correlation plot of variables **ph.karno**, **pat.karno**, **meal.cal**, **wt.loss** in the **lung** data.

```
library(ggcorrplot)
r <- cor(lung[, 7:10], use = "complete.obs")
ggcorrplot(r, hc.order = TRUE, type = "lower", lab = TRUE)
```