

Lab2

Завантаження пакетів

```
if (!require("pacman")) install.packages("pacman")

## Loading required package: pacman

pacman::p_load(pacman, ggplot2,
  plotly, rio, rmarkdown, moments, agricolae, corrplot, tidyverse, corrr)

library(pacman)
library(moments)
library(agricolae)
library(corrplot)
library(tidyverse)
library(corrr)
```

Завантаження даних

```
# CSV
data_csv <- import("/Users/victoria/Documents/StudentsPerformance.csv")
```

Опис даних

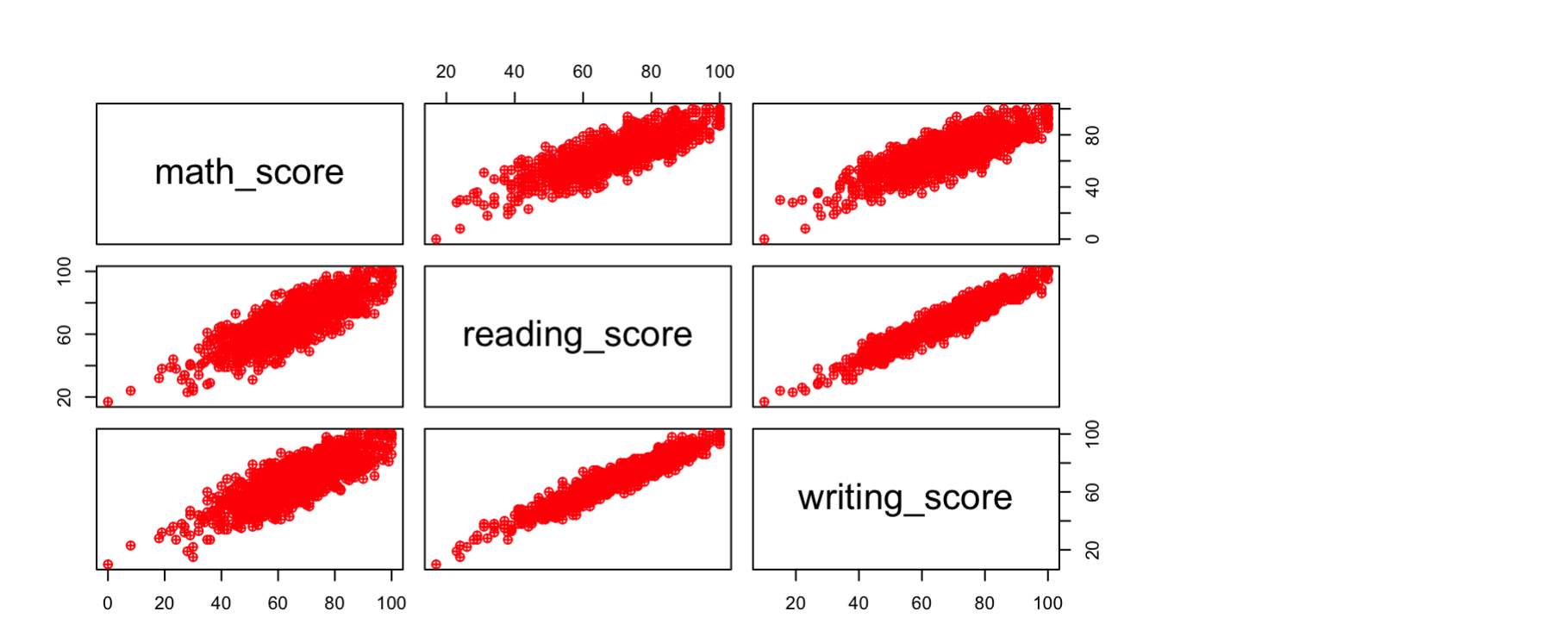
Датасет містить оцінки учнів старшої школи у США. Датасет створений штучно для досліджувальних цілей. Джерело: <https://www.kaggle.com/spscientist/students-performance-in-exams>

```
head(data_csv)

##   gender race/ethnicity parental_level_of_education      lunch
## 1 female      group B      bachelor's degree      standard
## 2 female      group C              some college      standard
## 3 female      group B      master's degree      standard
## 4 male        group A      associate's degree free/reduced
## 5 male        group C              some college      standard
## 6 female      group B      associate's degree      standard
##   test_preparation course math_score reading_score writing_score
## 1              none          72             72           74
## 2             completed          69             90           88
## 3              none          90             95           93
## 4              none          47             57           44
## 5              none          76             78           75
## 6              none          71             83           78
```

Для кількісних даних(math_score, reading_score, writing_score) зобразимо матричну діаграму розсіювання.

```
pairs(~math_score+reading_score+writing_score, data=data_csv, col="red", pch=10)
```



```
subset_data <- data_csv[,c("math_score", "reading_score", "writing_score")]
```

Виведемо карту кореляцій (кореляційна матриця).

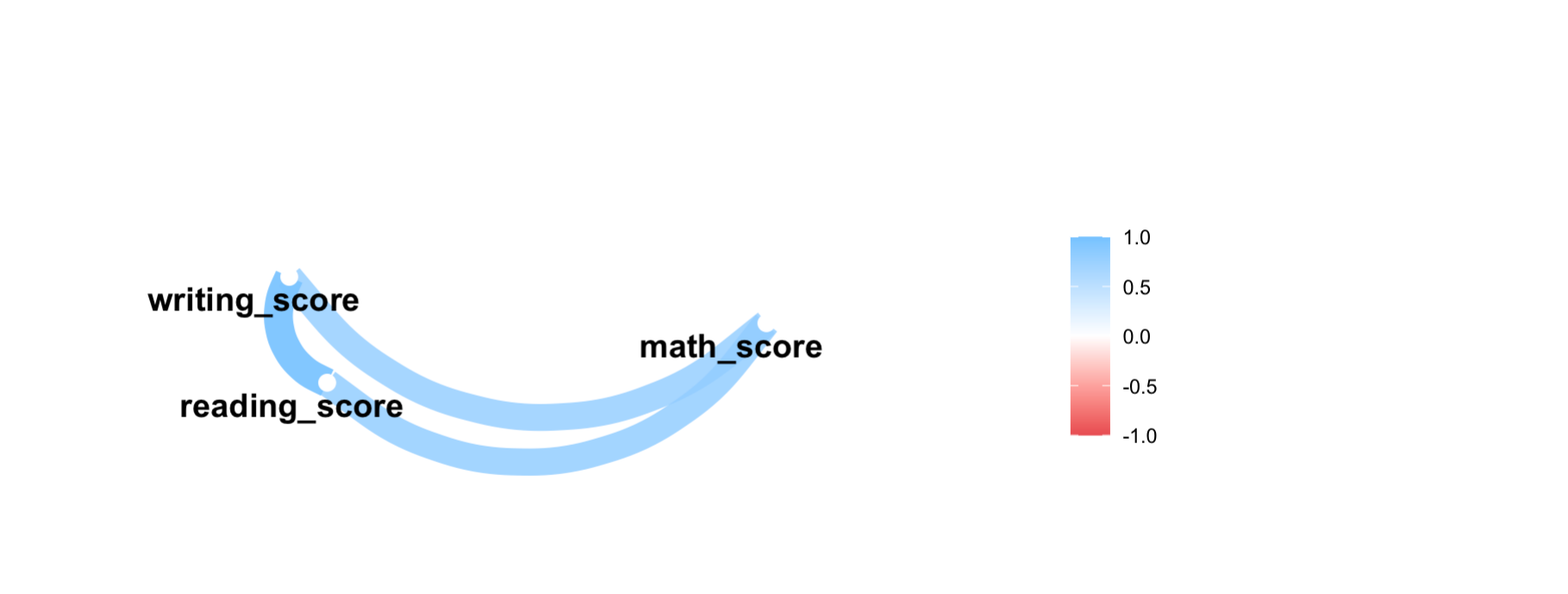
```
c <- cor(subset_data)
corrplot.mixed(c)
```



Граф кореляцій

```
subset_data %>% correlate() %>% network_plot(min_cor = .7)
```

```
##
## Correlation method: 'pearson'
## Missing treated using: 'pairwise.complete.obs'
```



Коефіцієнт кореляції Пірсона

```
cor.test(subset_data$reading_score, subset_data$writing_score, method="pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: subset_data$reading_score and subset_data$writing_score
## t = 101.23, df = 998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9487506 0.9597921
## sample estimates:
##      cor
## 0.9545981
```

Коефіцієнт кореляції Пірсона

```
cor.test(subset_data$reading_score, subset_data$math_score, method="pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: subset_data$reading_score and subset_data$math_score
## t = 44.855, df = 998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7959276 0.8371428
## sample estimates:
##      cor
## 0.8175797
```