

## Ćwiczenia 2.

Informacje ogólne dotyczące ćwiczeń.

R i RStudio powtórzenie

### Zadanie 1. Wektor

- Utwórz zmienną, która będzie przechowywać liczbę osób, które uczęszczają na zajęcia w danym dniu tygodnia. Nazwij zmienną *lista\_obecnosci* i nadaj nazwy poszczególnym wartościom. Wartości zmiennej: 25, 30, 16, 20, 10, 21, 12.
- Wyświetl, ile osób uczestniczyło w zajęciach w poniedziałek i piątek.
- Wyświetl dni z największą i najmniejszą liczbą uczestników.
- Oblicz, ile średnio osób uczestniczyło w zajęciach.
- Policz w ile dni frekwencja była wyższa lub równa 20 osób.
- Utwórz zmienną, która przechowuje liczbę osób, która powinna brać udział w zajęciach. Nazwij ją: *liczebosc\_grup* i przypisz jej wartości: 25, 30, 18, 20, 15, 21, 15.
- Sprawdź w jakie dni na zajęciach uczestniczyło 100% osób.
- Sprawdź w jakie dni na zajęciach uczestniczyło mniej niż 50% osób.

### Zadanie 2. Macierz

- Napisz kod, który da następujący wynik:

	<i>US</i>	<i>non-US</i>
<i>A New Hope</i>	460.998	314.4
<i>The Empire Strikes Back</i>	290.475	247.9
<i>Return of the Jedi</i>	309.306	165.8

- Wybierz z macierzy ostatni wiersz danych.
- Wybierz z macierzy wartość 314.4.

### Zadanie 3. Dataframe

- Napisz kod, który da następujący wynik:

	name	type	diameter	rotation	rings
1	Mercury	Terrestrial planet	0.382	58.64	FALSE
2	Venus	Terrestrial planet	0.949	-243.02	FALSE
3	Earth	Terrestrial planet	1.000	1.00	FALSE
4	Mars	Terrestrial planet	0.532	1.03	FALSE
5	Jupiter	Gas giant	11.209	0.41	TRUE
6	Saturn	Gas giant	9.449	0.43	TRUE
7	Uranus	Gas giant	4.007	-0.72	TRUE
8	Neptune	Gas giant	3.883	0.67	TRUE

- Wybierz średnice dla Marsa (diameter),
- Wybierz całość danych dla Uranus,
- Sprawdź strukturę utworzonej data frame
- Wyświetl informacje o pierścieniach (rings),
- Utwórz zmienną planets\_rings w której będą tylko te planety które mają pierścienie,
- Utwórz podzbiór zawierający planety o średnicy mniejszej niż 1,

#### Zadanie 4. Wektory czynnikowe

- Utwórz 2 wektory czynnikowe (factors), w tym jeden typu ordered.
- Sprawdź typ danych.
- Policz liczebność obserwacji należących do określonej klasy.

#### Zadanie 5. Funkcja

Napisz funkcję, która będzie wystawiać oceny na podstawie arkusza i Regulaminu Studiów AGH. Ocena może być wyliczona jeśli Student ma minimum 50% obecności. Jeśli nie spełnia tego warunku, zamiast oceny powinno wyświetlać się „brak zaliczenia”. Waga dla kolokwium to 0.9, dla każdej z pozostałych aktywności 0.05. W przypadku nie zaliczenia kolokwium ocena końcowa 2.0.

	liczba obecności	ocena kolokwium	ocena kartkowka	ocena aktywnosc
max	15	100	100	100

Proszę przetestować poprawność działania funkcji.

#### Zadanie 6. Pakiet dplyr

- Wykonaj filtrowanie danych gapminder dla Chin z roku 2002.
- Wykonaj filtrowanie danych gapminder dla roku 1957 i wykonaj sortowanie wyników według zmiennej pop.
- Dla danych z roku 2007, utwórz nową zmienną w obrębie gapminder
  - $\text{lifeExpMonths} = \text{lifeExp} * 12$
  - posortuj wyniki według nowo utworzonej zmiennej od najwyższych do najniższych wartości.
- Dla danych z roku 1957 oblicz medianę z lifeExp i umieść w zmiennej medianLifeExp oraz maksimum z gdpPercap i umieść w zmiennej maxGdpPercap.
- Dla pogrupowanych danych według kontynentu i roku, oblicz medianę z lifeExp i umieść w zmiennej medianLifeExp oraz wartość średnią z gdpPercap i umieść w zmiennej meanGdpPercap.

#### Zadanie 7. Import i export danych

- Zaimportuj pliki za pomocą funkcji i przypisz do odpowiadających im nazw zmiennych,
- Sprawdź poprawność danych,
- Wyeksportuj zmienne do pliku \*.csv,
- Powtórz z wykorzystaniem RStudio okna importu.

### Zadanie 8.

Dla zaimportowanych danych z zadania 7 wykonaj odpowiednie wykresy wykorzystując pakiet `graphics` i `ggplot2`. Wykresy do wykonania:

- histogram,
- wykres rozrzutu,
- wykres ramka-wąsy z uwzględnieniem zmiennej jakościowej,
- wykres liniowy.

### Zadanie 9.

RMarkdown pozwala na tworzenie dokumentów/raportów ze składnią Markdown. Tworzone są dokumenty zawierające nagłówki, grafiki, linki z elementami kodu w języku R. Po utworzeniu dokumentu `*.rmd` możemy go skompilować do np. `*.html` lub `*.pdf`

Aby korzystać z RMarkdown, należy zainstalować pakiet `'rmarkdown'`

Najważniejsze elementy języka:

1. Nagłówek dokumentu

```
---  
title: "DMinR_cwiczenia2"  
author: "MC"  
date: "06 03 2023"  
output: html_document  
---
```

2. Nagłówki w tekście, formatowanie tekstu:

```
# Nagłówek 1st  
## Nagłówek 2st  
*Italic*  
**Bold**  
  
Fragment 'kodu' w tekście  
* lista bez numeracji  
1. lista numerowana  
[Link] (https://www.google.com)
```

3. Code chunks

Wersja podstawowa:

```
` `{r}  
norm <- rnorm(100, mean = 0, sd = 1)  
```
```

Wersja z nazwaną code chunk

```
` `{r cars}  
summary(cars)  
```
```

Praca z pakietami R oraz nie wyświetlanie ostrzeżeń w kompilowanym dokumencie:

```
` `{r warning=FALSE}  
library(broom)
```

```
library(pander)
A <- c(20, 15, 10)
B <- c(1, 2, 3)
````
```

Nie wyświetli się kod, zostaną wykonane funkcje:

```
````{r, echo = FALSE}
A <- c("a", "a", "b", "b")
B <- c(5, 10, 15, 20)
dataframe <- data.frame(A, B)
print(dataframe)
````
```

Utwórz dokument Nazwisko.rmd. Po zakończeniu pracy z dokumentem wykonaj kompilację do Nazwisko.pdf.

Dokument powinien zawierać:

- Utworzony nagłówek dokumentu,
- Utworzony nagłówek 1 stopnia z tytułem „Data Mining- cw2”,
- 1 „code chunk” zawierający wczytanie danych „cars” i ich statystyki opisowe (funkcja summary), pierwsze i ostatnie obserwacje z pliku (funkcje head i tail),
- 2 „code chunk”: wykres dowolnej zmiennej (atrybutu dla danych cars) z wykorzystaniem pakietu graphics,
- 3 „code chunk”: histogram, z kodem nie wyświetlającym się po kompilacji,
- 4 „code chunk”: Utworzenie 3 zmiennych o różnych typach atomowych i ich wyświetlenie,
- Wyświetlone dane z pakietu gapminder dla roku 2002 dla zmiennej pop oraz w tym samym code chunk wykonany histogram dla zmiennej z punktu powyżej w pakiecie ggplot2,
- Wygenerowanie Nazwisko.pdf

Dodatkowe informacje: cheatsheet, <https://ourcodingclub.github.io/tutorials/rmarkdown/>

.