# # Programowanie w R – kurs podstawowy

## **# Zajęcia 4. Ramki danych (data frames)**

# **---------------------Co to jest ramka danych**

# Ramka danych może zawierać dane różnego typu, jak tekst, liczby i zmienne logiczne, ale jedna kolumna może zawierać tylko jeden typ danych

# Ramka danych przechowuje zmienne zbioru danych jako kolumny, a obserwacje jako wiersze

nazwisko typ wynik ocena zaliczenie

1 Burdzy dzienne 42 2.0 FALSE

2 Dyrkacz dzienne 85 4.5 TRUE

3 Grzeszczak zaoczne 49 2.0 FALSE

4 Kowalczyk dzienne 76 4.0 TRUE

5 Lubicki zaoczne 73 4.0 TRUE

6 Nowak zaoczne 36 2.0 FALSE

7 Nowicki dzienne 66 3.5 TRUE

8 Sachowicz zaoczne 91 5.0 TRUE

9 Walczak zaoczne 55 3.0 TRUE

10 Wrotek dzienne 89 4.5 TRUE

**# ---------------------Tworzenie ramki danych**

# Funkcja **data.frame()**pozwala utworzyć ramkę danych z istniejących wektorów o takiej samej długości

# Krok 1. Utwórzmy następujące wektory

nazwisko <- c("Burdzy", "Dyrkacz", "Grzeszczak", "Kowalczyk", "Lubicki", "Nowak", "Nowicki", "Sachowicz", "Walczak", "Wrotek")

typ <- c("dzienne", "dzienne", "zaoczne", "dzienne", "zaoczne", "zaoczne", "dzienne", "zaoczne", "zaoczne", "dzienne")

wynik <- c(42, 85, 49, 76, 73, 36, 66, 91, 55, 89)

ocena <- c(2, 4.5, 2, 4, 4, 2, 3.5, 5, 3, 4.5)

zaliczenie <- c(FALSE, TRUE, FALSE, TRUE, TRUE, FALSE, TRUE, TRUE, TRUE, TRUE)

# Krok 2. Utwórzmy ramkę danych “egzamin” z powyższych wektorów

egzamin <-data.frame(nazwisko, typ, wynik, ocena, zaliczenie)

**# -------------------Wybieranie elementów i podzbiorów ramki danych**

egzamin[1,3] # wybierze element z pierwszego wiersza i trzeciej kolumny

egzamin[1:3,2:4]# wybierze wiersze 1, 2, 3 oraz kolumny 2, 3, 4

egzamin[c(1,2,5,9),4:5] # jeśli nie są kolejne, to łączymy numery c()

egzamin[6,] # wybierze cały szósty wiersz

egzamin[,4] # wybierze całą czwartą kolumnę

egzamin[,4, drop=F] # zachowamy formę ramki

# Wybierając element ramki danych można również używać nazw kolumn

egzamin[1:3,"wynik"]

**# -------------------Użycie znaku $**

# W celu wybrania całej kolumny “wynik” możemy użyć dowolnej z poniższych wersji kodu

egzamin[,3]

egzamin[,"wynik"]

egzamin$wynik

# Wybraną kolumnę można od razu zapisać do nowego wektora

Wyniki\_egzaminu <- egzamin$wynik

# Wybór jednego elementu z kolumny „wynik”

egzamin[2, "wynik"]

egzamin$wynik[2]

**# ---------------------Funkcje statystyczne**

sum(egzamin$wynik)

mean(egzamin$wynik)

median(egzamin$wynik)

min(egzamin$wynik)

max(egzamin$wynik)

**# ---------------------Zmiana nazw zmiennych**

Metoda 1. Zmiana w już zaimportowanej ramce

colnames(egzamin)=c("Nazwisko","Typ\_studiów","Wynik\_egzaminu","Ocena","Zaliczenie”)

Metoda 2. Zmiana w trakcie importowania ramki

egzamin = data.frame(Nazwisko=nazwisko, Typ\_studiów=typ, Wynik\_egzaminu=wynik, Ocena=ocena, Zaliczenie=zaliczenie)

**# ----------------------- Sortowanie ramki danych**

# Funkcja **order()**

a = c(100, 10, 1000, 1)

order(a)

a[order(a)]

# Sortowanie po kolumnie Wynik\_egzaminu

ranking = order(egzamin$Wynik\_egzaminu)

egzamin[ranking,]

lub

egzamin[order(egzamin$Wynik\_egzaminu),]

# Uporządkujmy egzamin **malejąco** wg liczby punktów uzyskanych na egzaminie

ranking2 = order(egzamin$Wynik\_egzaminu, decreasing=T)

egzamin[ranking2,]

**# ---------------------Badanie struktury ramki danych**

iris # wbudowana ramka danych

# Funkcja nrow() sprawdza ile mamy wierszy

nrow(iris)

# Funkcja ncol() sprawdza ile mamy kolumn

ncol(iris)

# Funkcja length() też sprawdza ile mamy kolumn

length(iris)

# Funkcja dim() sprawdza ile mamy wierszy i kolumn

dim(iris)

# Funkcja head() - pokazuje początkowe obserwacje ramki danych

head(iris)

# Standardowo wyświetlanych jest 6 wierszy, ale można wywołać inną liczbę

head(iris, 10)

# Funkcja tail() – pokazuje ostatnie obserwacje ramki danych

tail(iris)

# Funkcja str() pozwala zbadać strukturę ramki danych.

str(iris)

# Funkcja summary()

summary(iris)

**# ---------------------Operacje na kolumnach**

iris$Petal.Width + 10

iris$Petal.Width - 10

iris$Petal.Width \* 4.2

iris$Petal.Width / 1000

iris$Sepal.Length + iris$Sepal.Width

iris$Sepal.Length \* iris$Sepal.Width

**# ---------------------Dodawanie nowej kolumny do ramki**

# Najpierw obliczmy sumę długości

Length = iris$Sepal.Length + iris$Petal.Length

# Dodajemy wektor jako nową kolumnę

iris$Length = Length

# Dodajemy od razu wynik działania jako nową kolumnę

iris$Length = iris$Sepal.Length + iris$Petal.Length

**# ---------------------Usuwanie kolumny z ramki**

iris$Petal.Width = NULL

**# ---------------------Pobranie danych**

kraje1=read.csv("http://www.kuligowska.com/uouw/kraje.csv", sep=";", dec=",", header=T)

kraje2=read.csv("http://www.kuligowska.com/uouw/kraje2.csv", sep=";", dec=",", header=T)

**# ---------------------Łączenie dwóch ramek danych**

# Przypomnijmy sobie jak wygląda ramka kraje

nrow(kraje1)

head(kraje1)

# Obejrzyjmy drugą ramkę kraje2

nrow(kraje2)

head(kraje2)

tail(kraje2)

str(kraje2)

summary(kraje2)

# Wyświetlmy head obu tabel, żeby je porównać

head(kraje1)

head(kraje2)

# Chcemy połączyć obie ramki w jedną. Najlepiej będzie połączyć po Kodzie, a nie po kraju (bo mogą być różnice w zapisie)

# Funkcja merge() – (ramka1, ramka2, nazwa kolumny z ramki1, nazwa kolumny z ramki2)

Kraje\_razem = merge(kraje1, kraje2, by.x=” Country.Code”, by.y=”Country.Code”)

# widzimy nową ramkę, ale ma o jedną zmienną mniej..

head(Kraje\_razem) # usunął nam jedną z tych kolumn, po których łączył, czyli został tylko jeden Kod kraju

**# ---------------------Usuwanie kolumny z ramki**

Kraje\_razem$Country = NULL

**# ---------------------Zmiana nazw zmiennych na polskie**

colnames(Kraje\_razem)=c("Kraj","Kod","PKB","Populacja","Dostęp\_do\_prądu","Wzrost\_populacji","Grupa\_dochodowa")

**# ---------------------Rozbicie tabeli wg kategorii (factor)**

Funkcja split() pozwala rozbić tabelę danych wg wskazanej kategorii

split(Kraje\_razem, Kraje\_razem$Grupa\_dochodowa)

**# -------------------Sortowanie kolumn**

# Dodajemy jedną dodatkową kolumnę

Studenci$LP = 1:10

# Przesuwamy ósmą kolumnę LP na pierwsze miejsce

Studenci[,c(8,1:7)]

# Chcemy zamienić pierwszą kolumnę z drugą

Studenci[,c(2,1, 3:8)]

Studenci[,c(2,1, 3:ncol(Studenci))]

# Przesuwamy ósmą kolumnę na drugie miejsce

Studenci[,c(1,8,2:7)]

Studenci[,c(ncol(Studenci),1:(ncol(Studenci)-1))]