# Eksploracja danych
## Labratorium 2

Sebastian Kuzara
KRUK S.A.
Statistical Methods Development Area
Wrocław, 2024

# Poprawność danych

Problemy z poprawnością danych mogą wystąpić:

Z poprzednich zajęć (patrz prezentacja):

1. Składowe zadłużenia
2. Relacja kapitał vs kwota pożyczki
3. Chronologia dat

Nowe:

4. Relacja zmiennych związanych z egzekucją komorniczą
5. Relacja wartości w kolumnach związanych z jednostką terytorialną (*Land*)
6. Relacja wartości w kolumnach związanych z ostatnią wpłatą
7. Kwoty wpłat z tabeli *events*
8. Wartości poza możliwym (realnym) zakresem, np. wartości ujemne
9. Relacja produktu z innymi cechami

# Wybrane dobre praktyki pisania kodu


1. Utworzenie projektu RStudio
2. Obiekty w R warto nazywać wg tego co zawierają lub robią – nie należy tworzyć “anonimowych” nazw np. a, b, c itp.
3. Warto pisać i używać własne funkcje
4. Podział skryptów na więcej niż 1 plik, np.
    a. Skrypt, który uruchamia program
    b. Funkcje w osobnym pliku
5. Nazwy funkcji rozpoczynamy od czasownika, obiekty przechowujące dane nazywamy rzeczownikiem
6. Nazwy angielskie
7. Funkcje powinny być krótkie i wykonywać jedną czynność (operację). Jako krótkie załóżmy “liberalnie”, że muszą mieścić się w całości na ekranie skryptu bez *scrollowania*.

Jeśli ktoś chciałby się więcej dowiedzieć o dobrych praktykach polecam stronę: https://refactoring.guru/pl

# Przykłady funkcji:


```
drawRandomNumber <- function() {
        random_number <- sample(1:100, 1)
        return(random_number)
}

drawRandomNumber()
```

```
add <- function(a, b=2) {
        result <- a+b
        return(result)
}
add(a=10)
add(a=10, b=20)
```

# R {ggplot2} - podstawy


Dokumentacja: https://ggplot2.tidyverse.org/index.html

Ściągawka: https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf

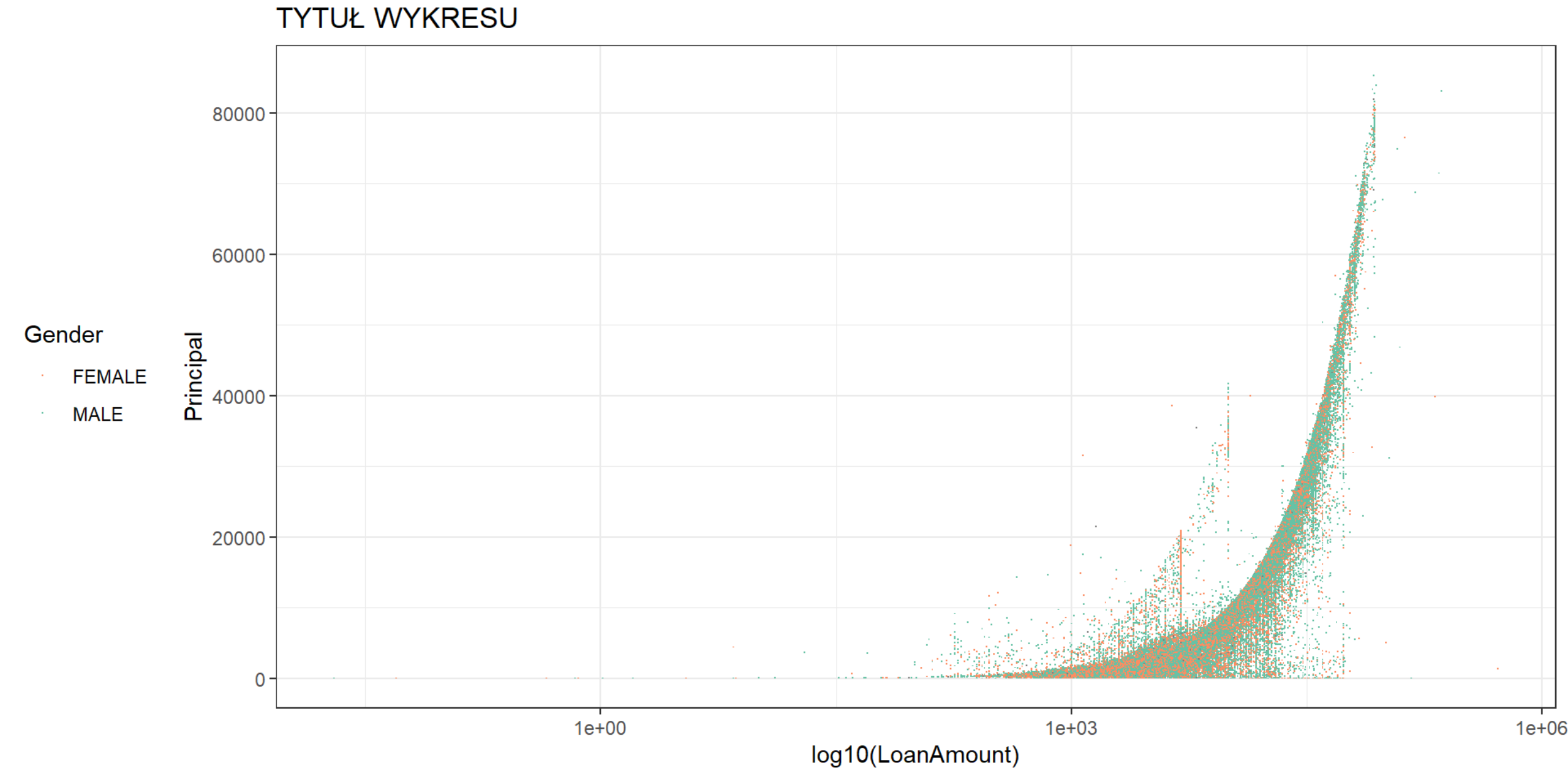
```
# Instalacja pakietu:

install.packages("ggplot2")
```

# ggplot2 - przykład


```
library(ggplot2)

# Tworzy nowy wykres ggplot
# elemnty wykresu (funkcje) łączymy za pomocą operatora +
ggplot(data = cases) +
# Layers (geoms): funkcje zaczynające się od geom_ # aes: mapowanie
parametrów
geom_point(aes(x = LoanAmount, y = Principal, color=Gender), shape=".") +
# Scales: przekształcenia parametrów
scale_color_manual(values = c("FEMALE" = "#fc8d62", "MALE" =
"#66c2a5")) +
scale_x_log10() + # Edycja elemntów wykrsu
ggtitle("TYTUŁ WYKRESU")
+ labs(x = "log10(LoanAmount)", y = "Principal") +
# Themes: kontrolują "wygląd" graficzny wykresu
theme_bw() +
theme(legend.position = "left")
```


TYTUŁ WYKRESU
Gender
FEMALE
MALE
Principal
80000
60000
40000
20000
0
1e+00
1e+03
1e+06
log10(LoanAmount)