



Uniwersytet Kazimierza Wielkiego

Instytut Informatyki

Zaawansowane metody eksploracji danych

Projekt w R Shiny

Liczba osób zameldowanych z podziałem na wiek i zamieszkanie

Piotr Kłopotek Główniczewski

Nr albumu: **89381**

Marek Tandecki

Nr albumu: **89530**

Bydgoszcz 2023 r.

Spis treści

1.	Wprowadzenie	3
2.	Zestaw danych	3
3.	Realizacja projektu	4
3.1.	Importowanie danych	4
3.2.	Biblioteki	4
3.3.	Interfejs	5
3.4.	Wykresy	6
3.4.2.	Wykres kołowy	6
3.4.3.	Wykres słupkowy pionowy	7
3.4.4.	Wykres słupkowy poziomy	7
3.4.5.	Wykres bąbelkowy	8
4.	Podsumowanie	8

1. Wprowadzenie

Celem projektu było stworzenie projektu, w którym to zostaną przedstawione dane statystyczne za pomocą różnych rodzajów wykresów. W tym celu wykorzystano aplikacje Shiny. Aplikacje shiny to strony internetowe tworzone i zasilane przez działającą na serwerze aplikację R. Użytkownicy aplikacji mogą poprzez stronę HTML wybierać parametry przetwarzania, przetwarzanie danych i parametrów ma miejsce na serwerze a jego wynik jest przedstawiany ponownie na stronie internetowej.

Aplikacje Shiny mają dwa komponenty - obiekt interfejsu użytkownika i funkcję serwera, które są przekazywane jako argumenty do shinyAppfunkcji, która tworzy obiekt aplikacji Shiny z tej pary interfejs użytkownika/serwer. Shiny web framework polega zasadniczo na ułatwieniu podłączania wartości wejściowych ze strony internetowej, dzięki czemu są one łatwo dostępne w języku R, a wyniki kodu R są zapisywane jako wartości wyjściowe z powrotem na stronie internetowej.

Ponieważ aplikacje internetowe Shiny są interaktywne, wartości wejściowe mogą ulec zmianie w dowolnym momencie, a wartości wyjściowe muszą zostać natychmiast zaktualizowane, aby odzwierciedlić te zmiany. Shiny jest dostarczany z reaktywną biblioteką programistyczną, której użyjesz do uporządkowania logiki aplikacji. Korzystając z tej biblioteki, zmiana wartości wejściowych w naturalny sposób spowoduje ponowne wykonanie właściwych części kodu R.

2. Zestaw danych

Zestawem danych wykorzystanym w projekcie R Shiny jest liczba osób zameldowanych na pobyt stały w podziale na poszczególne gminy. Przedmiotowe dane zostały opracowane na podstawie informacji zgromadzonych w rejestrze PESEL, dot. osób zameldowanych na pobyt stały tj. osób które na dzień wyboru z bazy nie miały w rejestrze PESEL zarejestrowanej daty zgonu lub daty znalezienia zwłok (osoby żyjące). Dane pochodzą z publicznej strony narodowej dane.gov.pl.

Portal danych jest źródłem wiarygodnych, na bieżąco aktualizowanych danych, udostępnianych bezpłatnie do ponownego wykorzystywania. W jednym miejscu znajdują się dane ponad 160 dostawców, tj. administracji publicznej oraz podmiotów prywatnych.

Portal został stworzony z myślą o obywatelach zainteresowanych działaniami państwa, firmach, które budują innowacyjne produkty i usługi oparte na danych, organizacjach pozarządowych, wykorzystujących dane w codziennej pracy, naukowcach prowadzących badania, urzędnikach przygotowujących raporty i analizy. Do korzystania ze strony nie jest wymagana rejestracja. Portal danych prowadzony jest przez ministra właściwego do spraw informatyzacji.

Link do danych wykorzystanych w projekcie:

<https://dane.gov.pl/pl/dataset/1082,liczba-osob-zameldowanych-na-pobyt-stay>

Kolumny znajdujące się w tabeli:

KOD_TERYT	WOJEWÓDZTWO	POWIAT	GMINA
WSZYSCY	KOBIETY_ZAMELDOWANE	MEZCZYJNI_ZAMELDOWANE	
PONIZEJ_18	KOBIETY_PONIZEJ_18	MEZCZYJNI_PONIZEJ_18	
POWYZEJ_18	KOBIETY_POWYZEJ_18	MEZCZYJNI_POWYZEJ_18	

3. Realizacja projektu

3.1. Importowanie danych

Do importu danych została wykorzystana biblioteka **library(openxlsx)** oraz serwis **GitHub**, na którym to został zawarty projekt wraz z plikiem Excel, który pochodzi ze strony dane.gov.pl.

Link do repozytorium GitHub: <https://github.com/pkglowczewski/zmed>

Funkcja w R:

```
8 df <-  
9   read.xlsx("https://github.com/pkglowczewski/zmed/raw/master/osoby.xlsx",  
10            sheet = 1)
```

3.2. Biblioteki

W projekcie zostały wykorzystane następujące biblioteki.

```
1 library(shiny)  
2 library(openxlsx)  
3 library(RColorBrewer)  
4 library(plotly)  
5 library(DT)
```

Szczegóły poniżej.

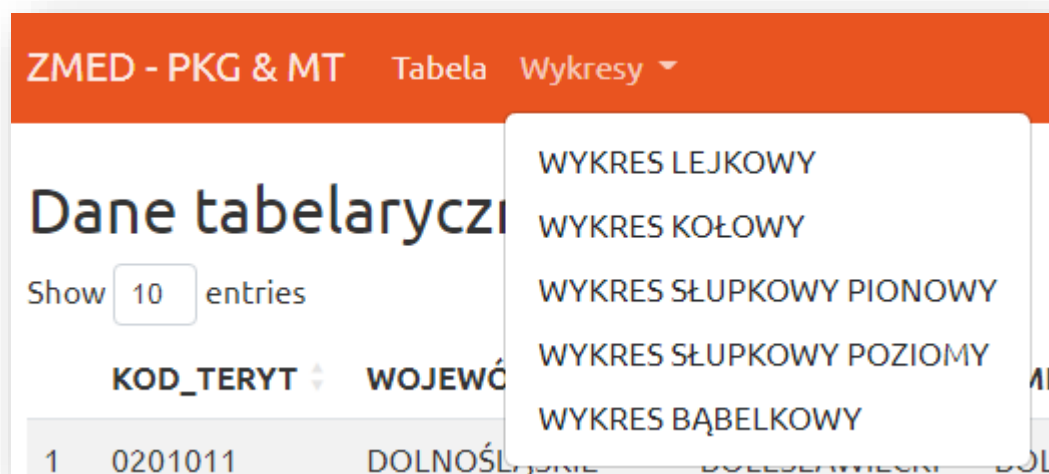
```
Matrix products: default  
  
locale:  
[1] LC_COLLATE=Polish_Poland.utf8 LC_CTYPE=Polish_Poland.utf8 LC_MONETARY=Polish_Poland.utf8  
[4] LC_NUMERIC=C LC_TIME=Polish_Poland.utf8  
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base  
  
other attached packages:  
[1] DT_0.26 plotly_4.10.1 ggplot2_3.4.0 RColorBrewer_1.1-3 openxlsx_4.2.5.1  
[6] shiny_1.7.4
```

3.3. Interfejs

W aplikacji R Shiny został utworzony responsywny oraz intuicyjny interfejs, który jest zrozumiały i prosto w obsłudze. Interfejs składa się z paska nawigacyjnego, w którym są dwie zakładki:

- Tabela – w której znajdują się pełne dane tabelaryczne z zaciągniętego pliku Excel.
- Wykresy – jest podmenu wykresów takich jak:
 - Wykres lejkowy
 - Wykres kołowy
 - Wykres słupkowy pionowy
 - Wykres słupkowy poziomy
 - Wykres bąbelkowy

Poniższy interfejs wygląda następująco:



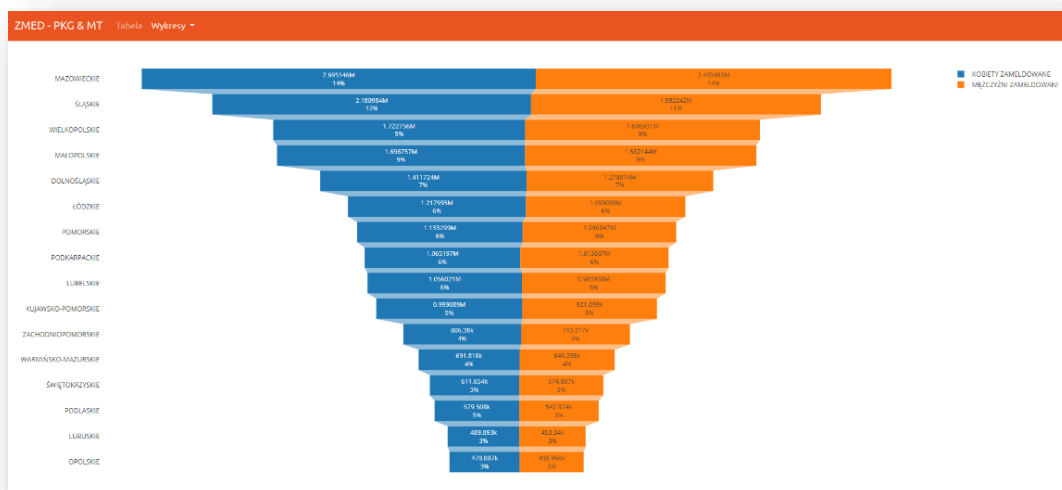
Po kliknięciu wybranego wykresu, strona załaduje interesujące nas przedstawienie danych.

3.4. Wykresy

W aplikacji Shiny zostały zaimplementowane wykresy dla których można definiować swoje dane wejściowe w oparciu o zestaw danych, które zostały pobrane z repozytorium Git.

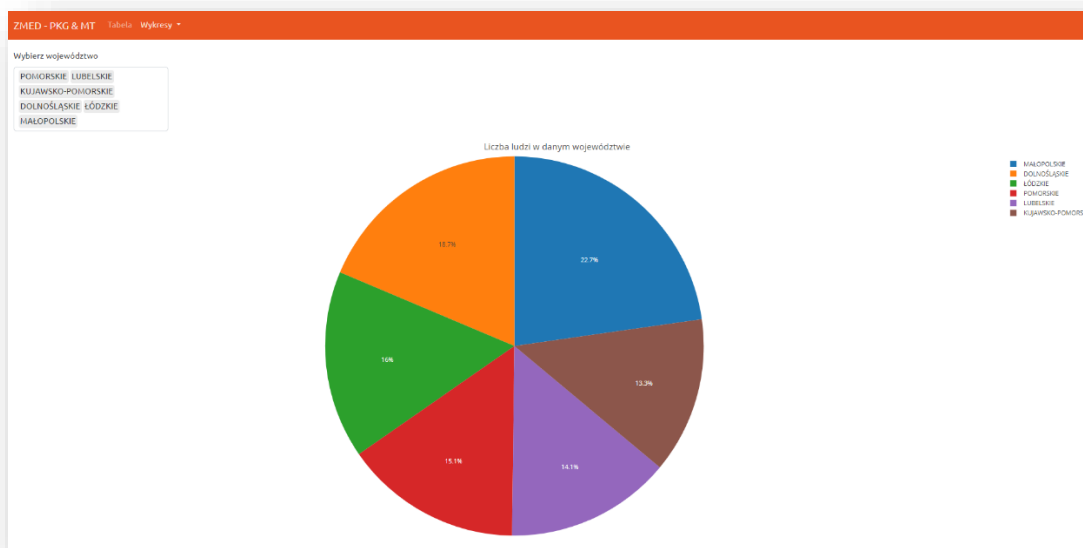
3.4.1. Wykresy to lejkowy

Wykres lejkowy jest zwykle używany w celu prezentacji wartości w przepływach i procesach. Pod względem wyświetlania jest on powiązany z wykresem kołowym. W tym przypadku wykres lejkowy przedstawia zależność pomiędzy województwami a kobietami zameldowanymi oraz mężczyznami zameldowanymi.



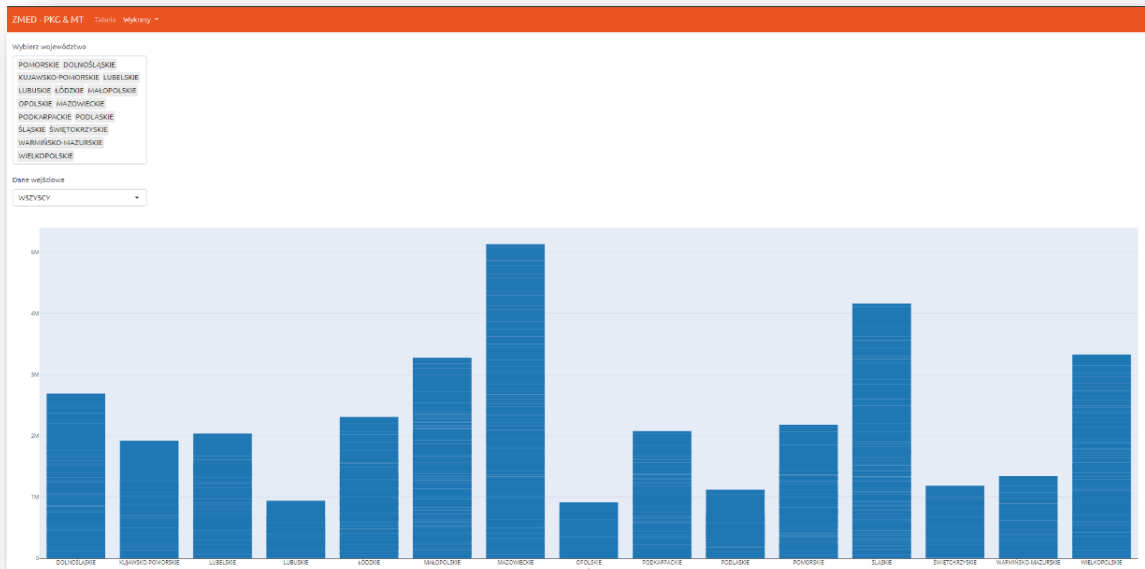
3.4.2. Wykres kołowy

Wykres kołowy podzielony jest na wycinki, obrazujące proporcje. Na diagramie kołowym długość łuku każdego wycinka jest proporcjonalna do ilości, jaką przedstawia. Wszystkie wycinki diagramu łącznie tworzą pełne koło. Wykres kołowy przedstawia liczbę ludzi w danym województwie. Należy dodać że ilość przedstawianych województw jest definiowalna przez użytkownika.



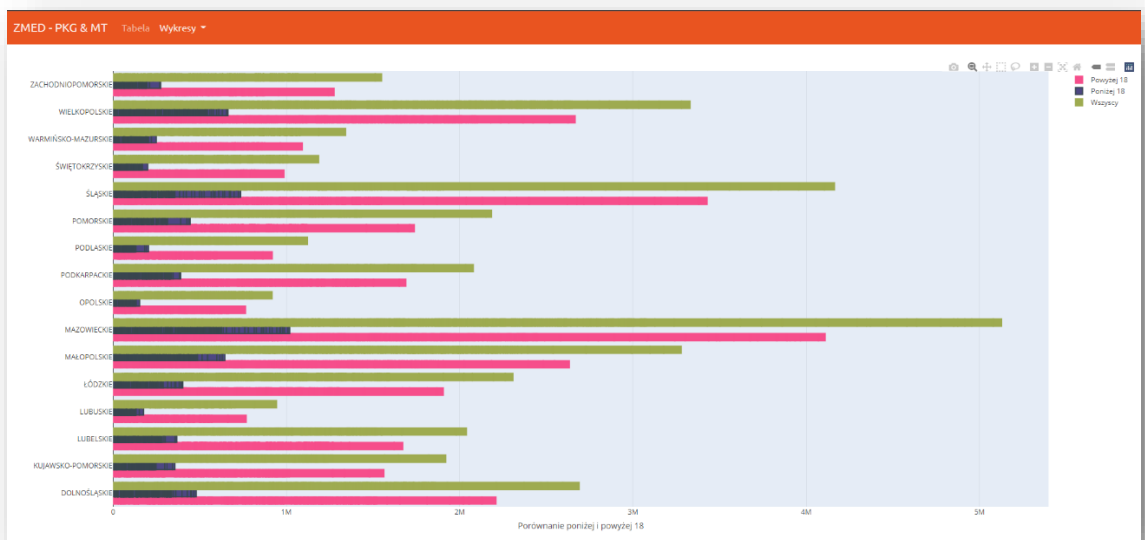
3.4.3. Wykres słupkowy pionowy

Wykres słupkowy przedstawia wyniki, w którym każda kategoria zmiennej jest reprezentowana przez słupek, którego długość pozostaje w proporcji do odpowiadającej jej wartości liczbowej. Tak jak w poprzednim przykładzie, tak również tutaj użytkownik może zdefiniować sobie interesujące dane wejściowe.



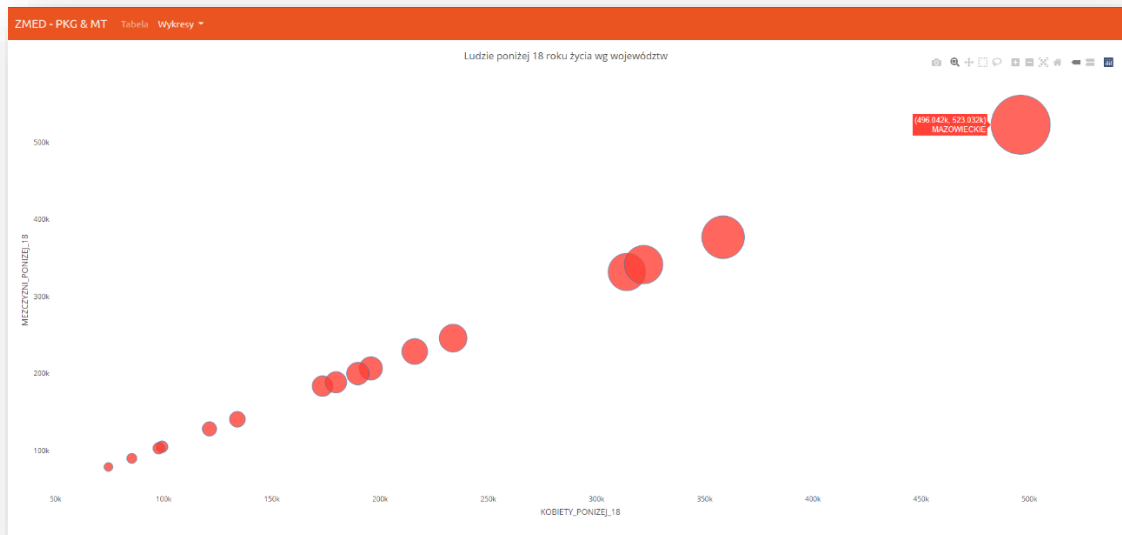
3.4.4. Wykres słupkowy poziomy

Diagram ten przedstawia w podobny sposób wartości jak wykres słupkowy pionowy, jednakże jest odwrócony. Zaletą tego typu diagramu jest możliwość wybrania interesujących nas słupków w celu przedstawienia danych. Wykres ten kategoryzuje dane dla wszystkich województw.



3.4.5. Wykres bąbelkowy

Wykres bąbelkowy to odmiana wykresu punktowego, na punkty danych którą zastępują bąbelki, a dodatkowy wymiar danych jest reprezentowany przez rozmiar bąbelków. Wykresy bąbelkowe i punktowe są doskonałym rozwiązaniem pozwalającym wyświetlać wiele pokrewnych danych na jednym wykresie. Wykres w programie przedstawia dane dotyczące osób poniżej 18 roku życia ze względu na województwa.



4. Podsumowanie

Prezentowany powyższy projekt spełnia wymagane kryteria dotyczące wyświetlania zestawu o dużej ilości danych jak i obszernym zasobie kolumn. Dzięki zastosowaniu aplikacji R Shiny jesteśmy w stanie wywnioskować jak wygląda aktualne zameldowanie w poszczególnych województwach, powiatach, gminach. Prezentacja danych pozwala na porównanie zameldowań pomiędzy kobietami oraz mężczyznami. Zestaw danych przedstawia również podział wiekowy ze względu na wiek. Dla kobiet oraz mężczyzn został uwzględniony podział na wiek powyżej oraz poniżej osiemnastu lat. Największą zaletą Shiny jest interaktywność wykresów jak i automatyczna aktualizacja danych podczas zmian danych wejściowych.