SPRAWOZDANIE

Zajęcia: Zbiory Big Data i Eksploracja Danych

Prowadząca: dr inż. Ruslana Ziubina

|  |  |
| --- | --- |
| Laboratorium nr 4  Data rozpoczęcia: 1.12.2023  Temat: Rattle – początki pracy | Rafał Klinowski  Informatyka  II stopień, stacjonarne,  Semestr 2, gr. a |

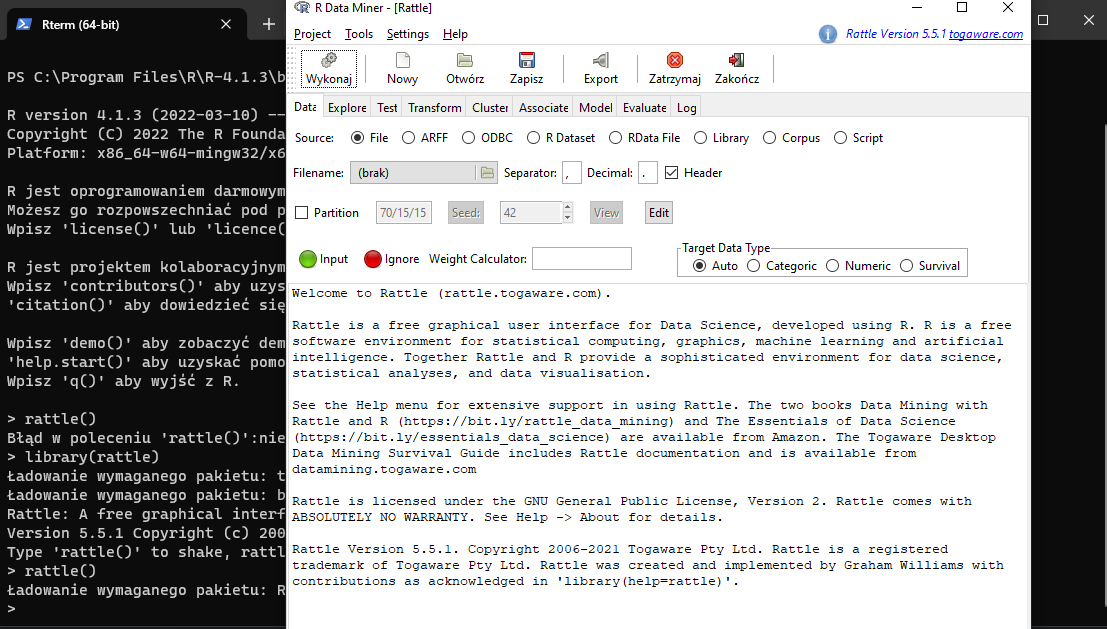
Poszczególne ćwiczenia będą wykonywane w pliku źródłowym edytowanym przy pomocy środowiska RStudio oraz Rattle, opisanego w poprzednich laboratoriach.

# Ćw. 1.

Pierwszym zadaniem była instalacja pakietu Rattle wraz z jego zależnościami. Podobnie jak w ramach Laboratorium 1, wymagane było skorzystanie z wcześniejszej wersji R (4.1.3) oraz pobranie pakietów z CRAN. W szczególności problemy występowały z pakietem RGtk2 oraz stringi.

# Ćw. 2.

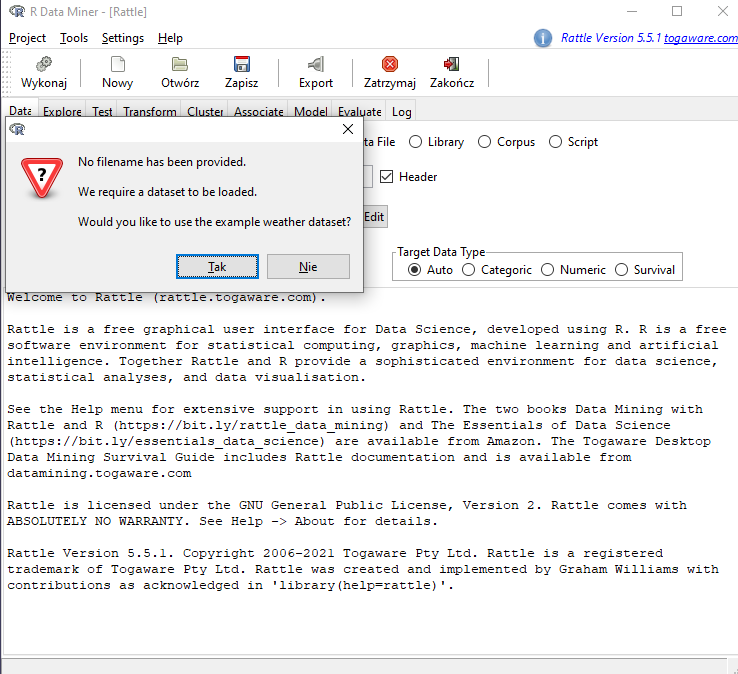
Po zainstalowaniu pakietów uruchomiono środowisko Rattle.



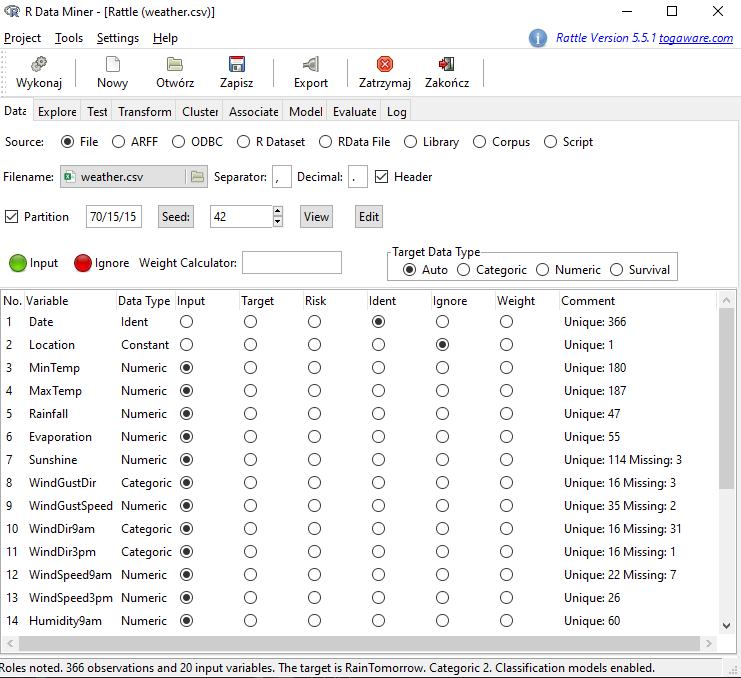
Rysunek . Środowisko Rattle.

# Ćw. 3.

Wczytanie przykładowego zbioru danych w Rattle.



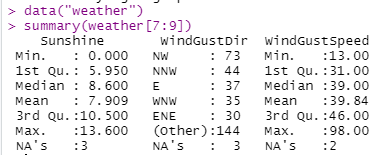
Rysunek . Ekran po kliknięciu Execute.



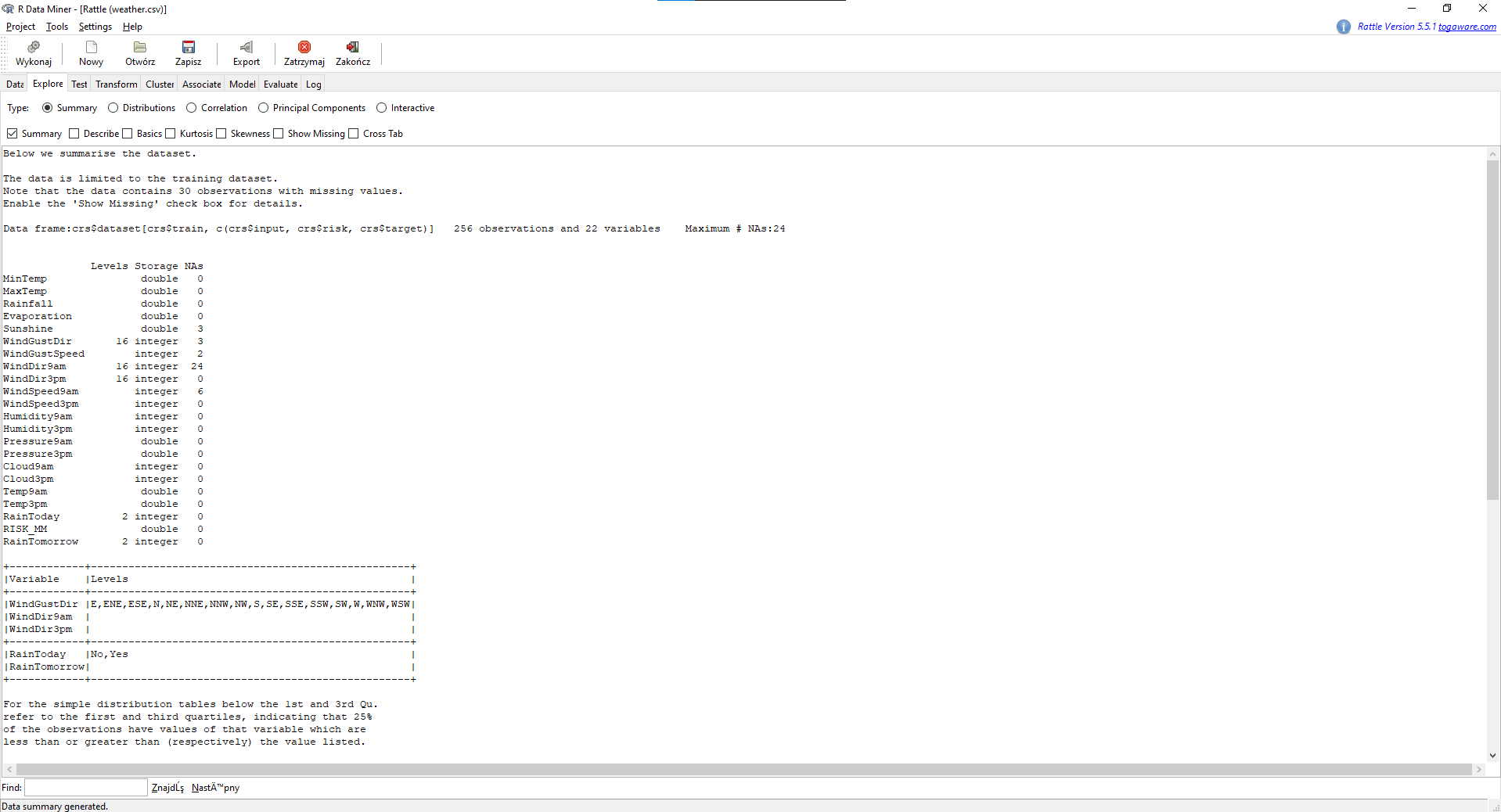
Rysunek . Wygląd wczytanego zbioru danych.

# Ćw. 4.

Proste podsumowanie tekstowe zbioru danych Weather.



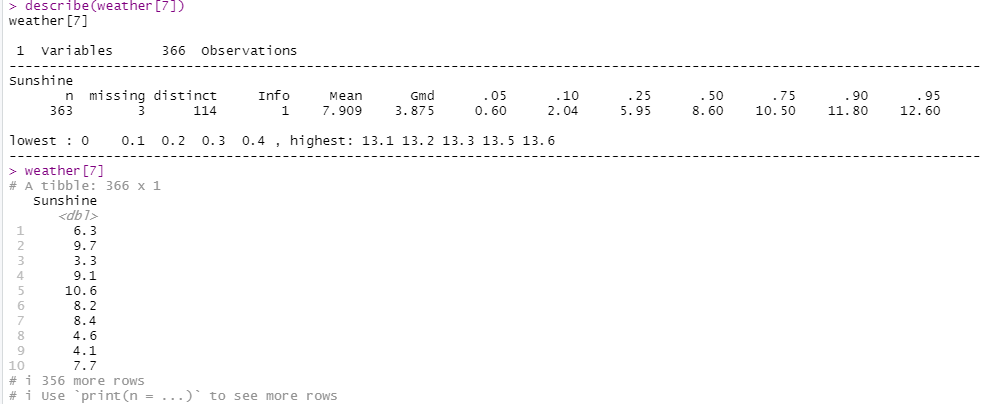
Rysunek . Podsumowanie zbioru danych w R.



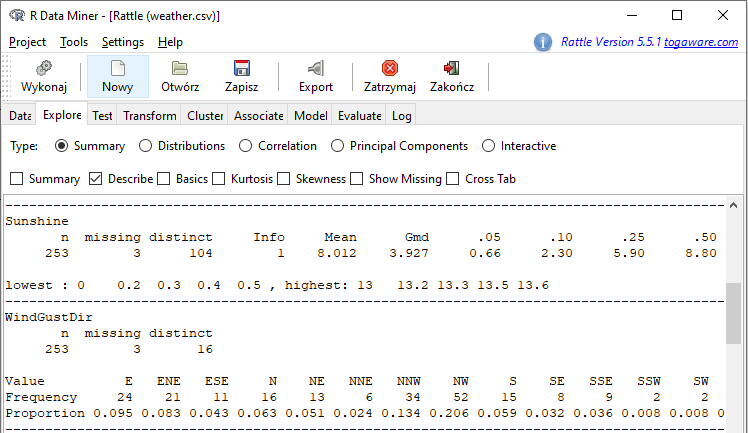
Rysunek . Podsumowanie zbioru danych w Rattle.

# Ćw. 5.

Podsumowanie danych za pomocą Hmisc. Konieczne było najpierw zainstalowanie pakietu.



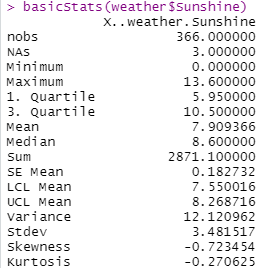
Rysunek . Podsumowanie danych przy pomocy Hmisc w R.



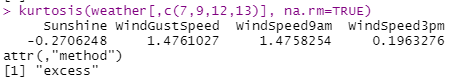
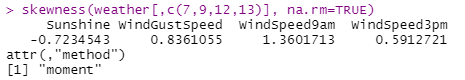
Rysunek . Podsumowanie w Rattle.

# Ćw. 6.

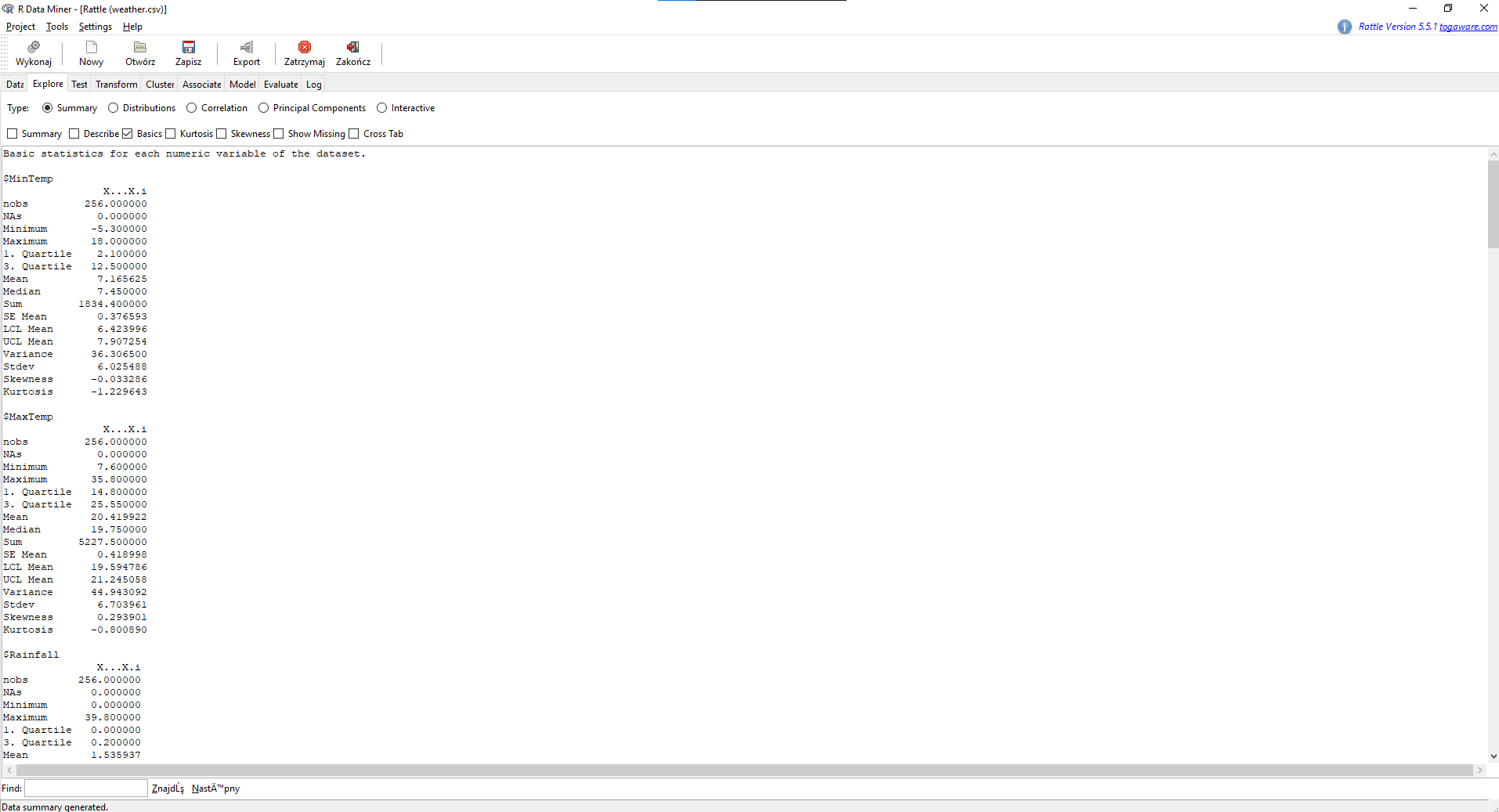
Podsumowanie numeryczne z wykorzystaniem pakietu fBasics.



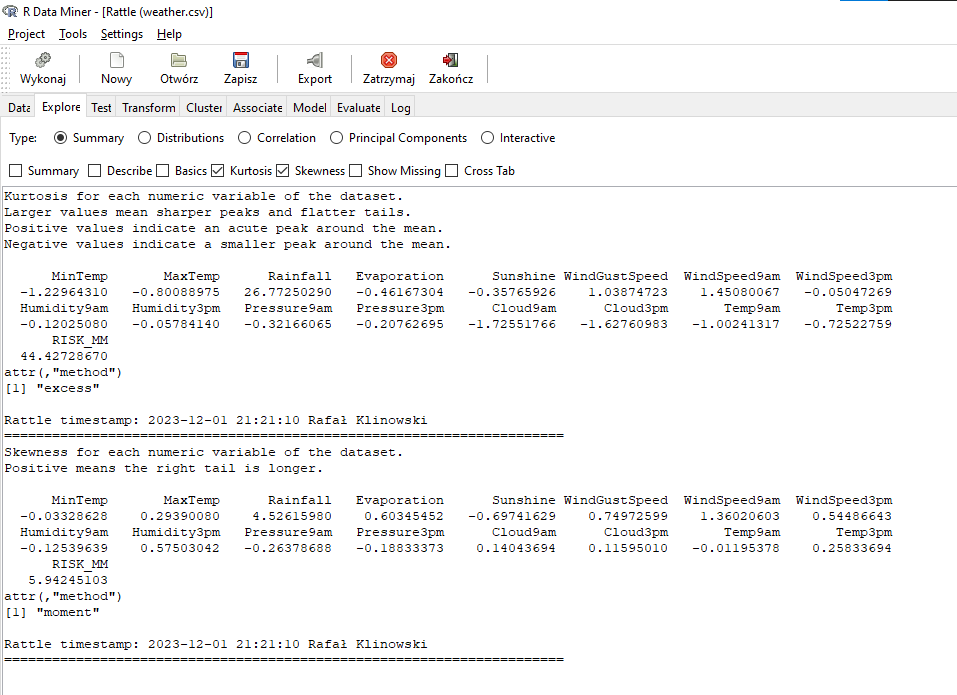
Rysunek . Podstawowe statystyki w R.



Rysunek . Skośność i kurtoza dla przykładowych danych w R.



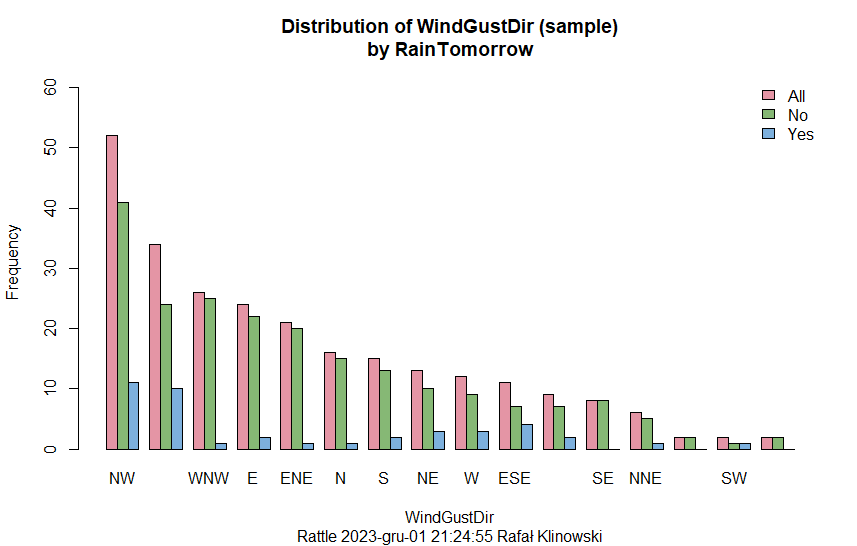
Rysunek . Podsumowanie numeryczne w Rattle.



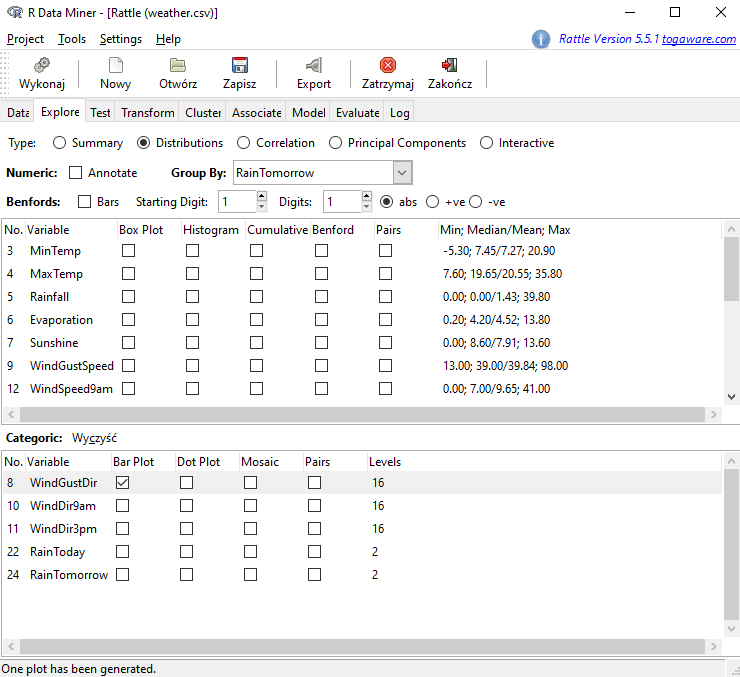
Rysunek . Skośność i kurtoza w Rattle.

# Ćw. 7.

Wykres słupkowy w Rattle.



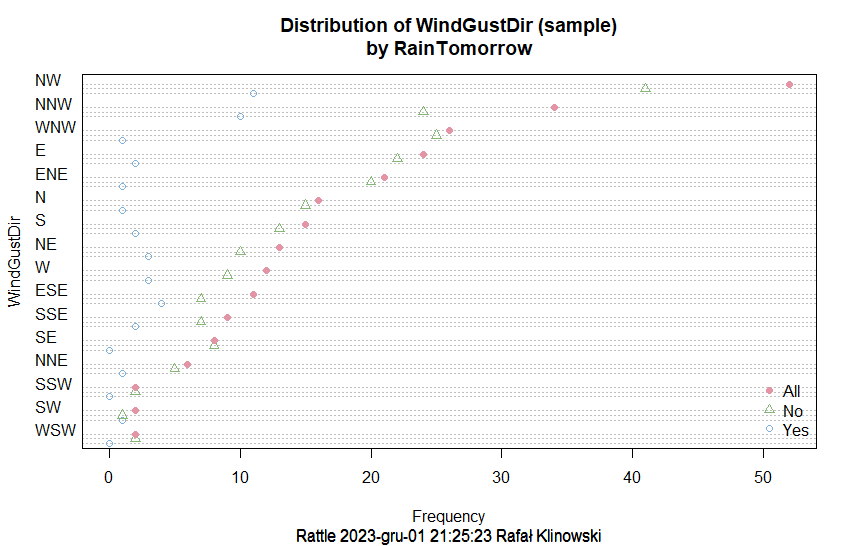
Rysunek . Wykres słupkowy w Rattle.



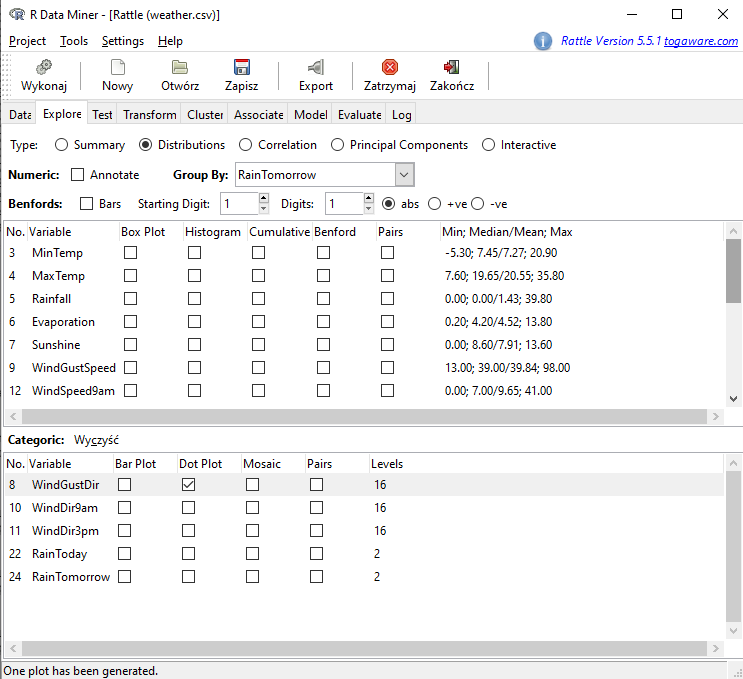
Rysunek . Parametry do utworzenia wykresu w Rattle.

# Ćw. 8.

Wykres kropkowy w Rattle.



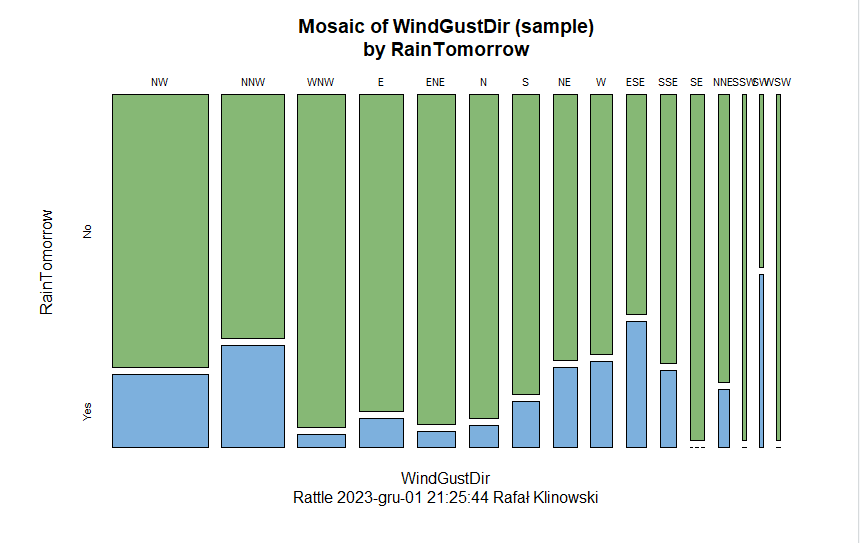
Rysunek . Wykres kropkowy utworzony w Rattle.



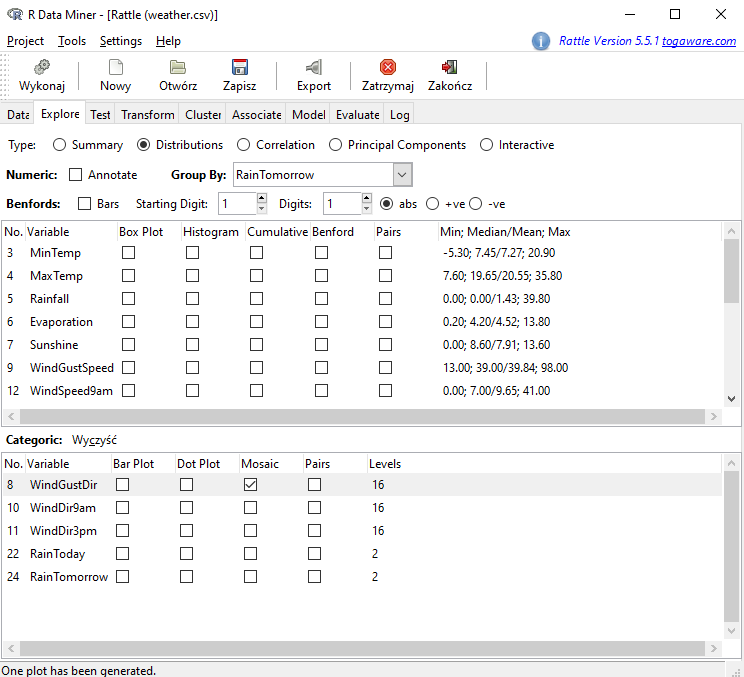
Rysunek . Ustawienia w Rattle do wygenerowania wykresu.

# Ćw. 9.

Wykres mozaikowy w Rattle.



Rysunek . Wykres mozaikowy utworzony w Rattle.



Rysunek . Ustawienia w Rattle do wygenerowania wykresu mozaikowego.

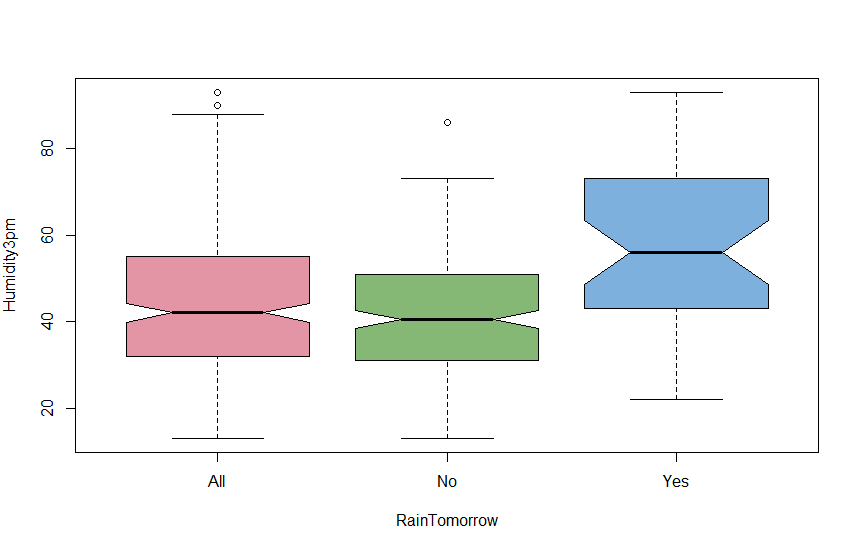
# Ćw. 10.

Utworzenie i edycja wykresu w R przy pomocy domyślnego zestawu danych Weather oraz doBy.

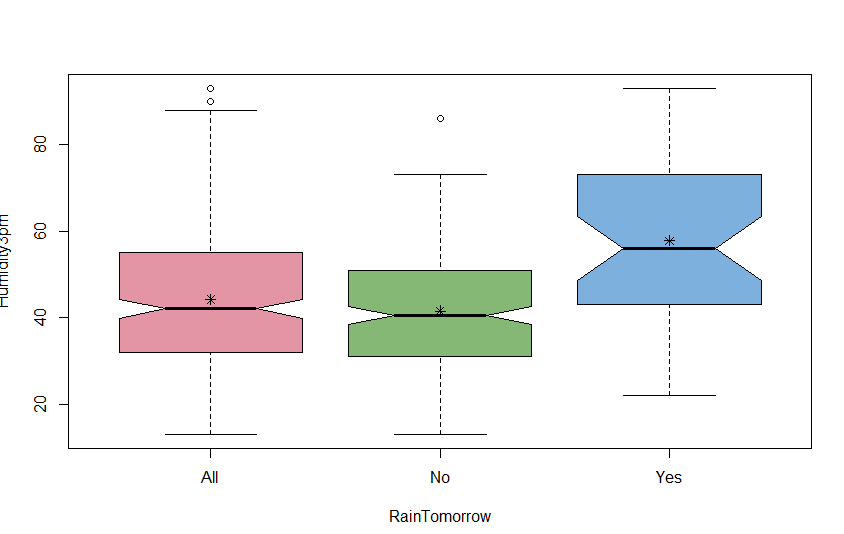
Na początku zgodnie z instrukcjami utworzono wykres dla kolumny Humidity3pm.



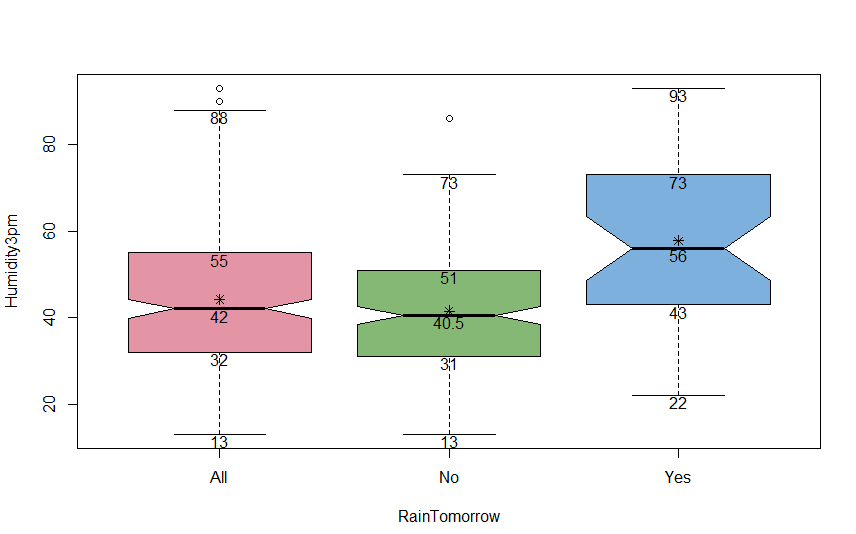
Rysunek . Utworzenie danych oraz prostego wykresu.



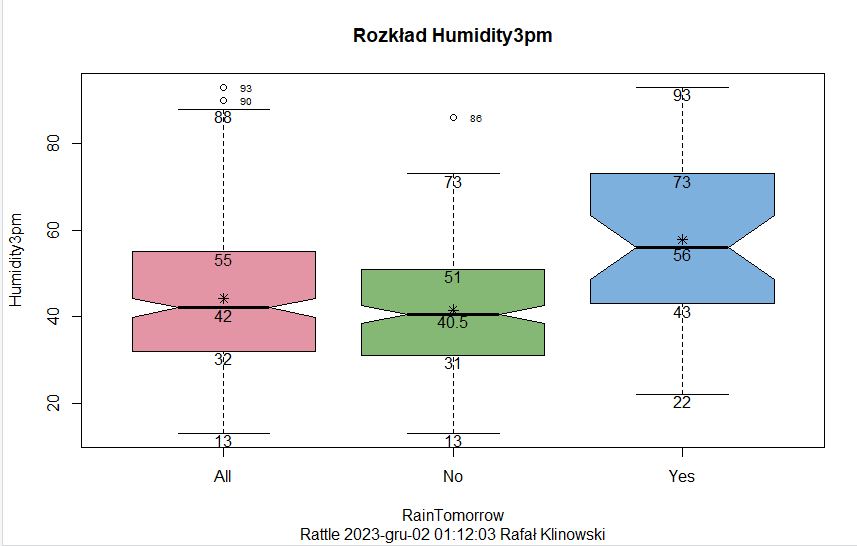
Rysunek . Wygląd podstawowego wykresu.



Rysunek . Wykres po dodaniu dodatkowych punktów danych.

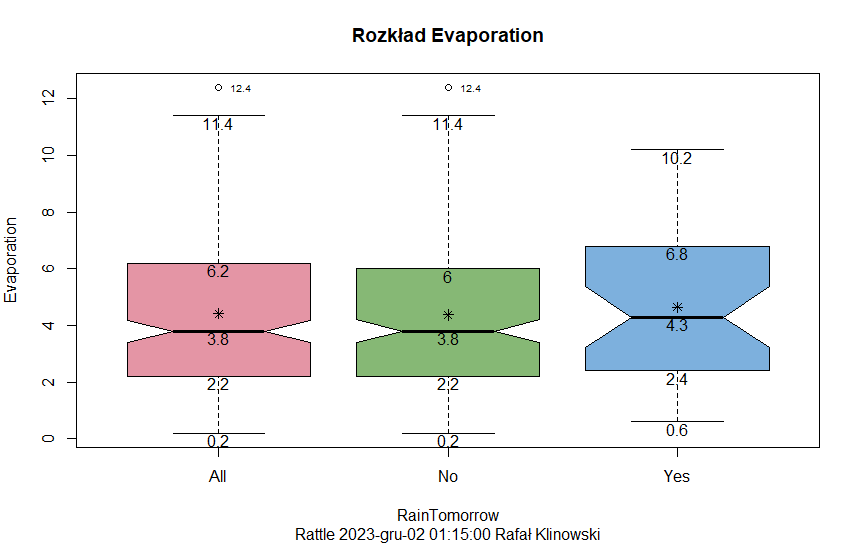


Rysunek . Wykres po dodaniu dodatkowego tekstu reprezentującego wartości.

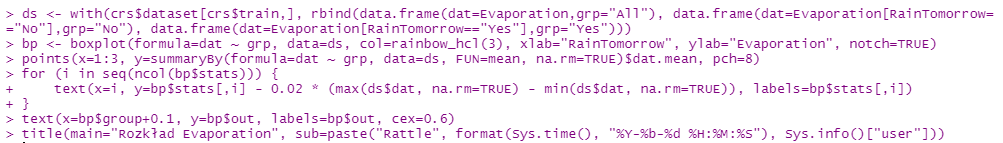


Rysunek . Gotowy wykres po dodaniu tytułu oraz wszystkich poprzednich elementów.

Proces powtórzono dla kolumny Evaporation.



Rysunek . Powtórzenie powyższych instrukcji analogicznie dla kolumny Evaporation.



Rysunek . Polecenia wykorzystane do utworzenia powyższego wykresu.

# Wnioski.

Największą trudnością podczas realizacji laboratorium było zainstalowanie pakietu Rattle oraz jego zależności. Wymagało to skorzystanie z systemu Windows (poprzez problemy z kompatybilnością na innych systemach, w tym korzystając z kontenera Docker) oraz ręczną instalację wielu z zależności (między innymi RGtk2 czy stringi), co zajęło znacznie więcej czasu niż sama realizacja laboratorium.

Środowisko Rattle umożliwia w prosty sposób pracę nad danymi, w tym tworzenie wykresów, eksplorację danych czy uzyskiwanie podsumowań. Rattle zawiera również sporą ilość narzędzi wycelowanych w uczenie maszynowe czy grupowanie danych. Środowisko współpracuje z wieloma dodatkowymi pakietami i zawiera przejrzysty interfejs, w którym dość łatwo znaleźć wszystkie interesujące nas opcje.