

Package ‘PogromcyDanych’

August 20, 2023

Type Package

Title DataCrunchers (PogromcyDanych) is the Massive Online Open Course that Brings R and Statistics to the People

Version 1.7.1

Date 2023-08-20?

Author Przemyslaw Biecek

Maintainer Przemyslaw Biecek <przemyslaw.biecek@gmail.com>

Description The data sets used in the online course „PogromcyDanych”. You can process data in many ways. The course Data Crunchers will introduce you to this variety. For this reason we will work on datasets of different size (from several to several hundred thousand rows), with various level of complexity (from two to two thousand columns) and prepared in different formats (text data, quantitative data and qualitative data). All of these data sets were gathered in a single big package called PogromcyDanych to facilitate access to them. It contains all sorts of data sets such as data about offer prices of cars, results of opinion polls, information about changes in stock market indices, data about names given to newborn babies, ski jumping results or information about outcomes of breast cancer patients treatment.

LazyLoad yes

LazyData yes

LazyDataCompression xz

License GPL-3

Depends R (>= 3.0), dplyr, SmarterPoland

NeedsCompilation no

Repository CRAN

Date/Publication 2023-08-20 21:22:41 UTC

R topics documented:

PogromcyDanych-package	2
auta2012	3
Diagnoza-Spoleczna	4

galton	5
imiona_warszawa	6
koty_ptaki	7
mandatySejmik2014	8
serieIMDB	9
setLang	10
skiJumps2013	11
TCGA_BRCA	13
WIG	14
Index	15

PogromcyDanych-package

PogromcyDanych / DataCrunchers is the Massive Online Open Course that Brings R and Statistics to the People

Description

The data sets used in the online course „PogromcyDanych”. The package includes both large data sets, such as auta2012 with prices for cars offered for sale in 2012, as well as small data sets for basic exercises such as cats_birds.

You can process data in many ways. The course Data Crunchers will introduce you to this variety. For this reason we will work on datasets of different size (from several to several hundred thousand rows), with various level of complexity (from two to two thousand columns) and prepared in different formats (text data, quantitative data and qualitative data).

All of these data sets were gathered in a single big package called PogromcyDanych to facilitate access to them. It contains all sorts of data sets such as data about offer prices of cars, results of opinion polls, information about changes in stock market indices, data about names given to newborn babies, ski jumping results or information about outcomes of breast cancer patients treatment.

[POL]

Praca z danymi może być bardzo różnorodna. Podczas kursu „Pogromcy Danych” będziemy tę różnorodność przybliżać. W tym celu będziemy pracować na zbiorach danych o różnej wielkości (od kilkunastu do kilkuset tysięcy wierszy), różnym stopniu złożoności (od dwóch do dwóch tysięcy kolumn), oraz o różnym formacie (dane tekstowe, ilościowe, jakościowe).

Aby ułatwić dostęp do tak różnych zbiorów danych zebraliśmy je wszystkie w jednym pakiecie o nazwie PogromcyDanych. Znaleźć można w nim najróżniejsze dane, w tym o cenach ofertowych aut, dane ankietowe dotyczące badań społecznych, informacje o zmianach indeksów giełdowych, imionach noworodków, wynikach skoków narciarskich czy o wynikach leczenia pacjentek z nowotworem piersi.

Author(s)

Author: Przemyslaw Biecek Maintainer: Przemyslaw Biecek <przemyslaw.biecek@gmail.com>

References

<http://www.PogromcyDanych.pl>

auta2012

Offer Prices of Second-hand Cars in 2012

Description

The data set auta2012 was designed with a view of creating opportunity to work on a big and interesting data set concerning transactions. Transaction data is usually rather extensive and it is processed in a very specific manner.

This data was extracted from the announcement service otomoto.pl in 2012. The set contains information about over 207 thousand car sales offers.

Each offer (each row) is described by 21 variables including:

- price, currency and information whether the given price is a net or a gross price,
- information on engine power expressed in horsepower and kW and its capacity,
- information about model, make and version of the car,
- information on the type of drive, year of manufacture and mileage (in km),
- information about color, number of doors, country of origin, country of registration and accessories.

Information about accessories is described by a very long line of text with values separated by comas.

[POL]

Ceny ofertowe używanych aut w roku 2012

Zbiór danych auta2012 przygotowaliśmy po to, by dać możliwość pracy z dużym i ciekawym zbiorem danych o transakcjach. Dane transakcyjne są zazwyczaj bardzo długie, a sposób ich przetwarzania jest specyficzny.

Dane zostały pobrane w roku 2012 z serwisu ogłoszeń otomoto.pl. Zbiór danych zawiera informacje o ofertach dla przeszło 207 tysięcy ogłoszeń sprzedaży auta.

Każda oferta (każdy wiersz) opisana jest przez 21 zmiennych, w tym:

- cenę, walutę i informację czy jest to cena brutto czy netto,
- informację o mocy silnika w koniach mechanicznych i kW, oraz jego pojemności,
- informację o modelu, marce, wersji auta,
- informację o rodzaju napędu, roku produkcji oraz przebiegu (w km),
- informację o kolorze, liczbie drzwi, kraju pochodzenia, kraju rejestracji oraz wyposażeniu.

Informacja o wyposażeniu jest opisana przez napis z wartościami rozdzielanymi przecinkiem, przez co sam napis może być bardzo długi.

Author(s)

Przemysław Biecek

Examples

```
head(auta2012)
```

Diagnoza-Spoleczna *A Subset of Polish Social Diagnosis Data*

Description

Social Diagnosis

Results of the panel research called Social Diagnosis (Diagnoza Społeczna) form a very interesting data set. The same people from a chosen group of households are polled every two years. The questions concern various issues and the answers of the participants allow for construction of a model of social changes taking place in Poland. You can find more information about this research, its results, as well as information about the data set on the project's website <http://diagnoza.com>.

The data set in the form processable by R is available on the website <https://github.com/pbiecek/Diagnoza>. You can install it using a command `install_github("pbiecek/Diagnoza")` after previous activation of the `library(devtools)` package.

The whole data set is large and small computers might have problems with it. For the purposes of this course I have prepared a subset of the data set from the Social Diagnosis research.

The subset is called `diagnoza` and it consists of 38461 rows. Each row presents answers of one person. The responses received in the pools are presented in 36 columns/variables. The names of the variables correspond to the questions asked in the poll http://diagnoza.com/pliki/kwestionariusze_instrukcje/kwestionariusze. The data set `diagnozaDict` gives full versions of all the questions.

The variables describe among other things:

- names of the respondents,
- analytical weights,
- number of years of study, gender, education, height, weight, income,
- answers to chosen questions concerning the worldview.

The data set called `diagnozaDict` describes names of columns from the `diagnoza` data set.

[POL]

Diagnoza społeczna

Ciekawym zbiorem danych jest wynik panelowego badania Diagnoza Społeczna. W ramach tego projektu co dwa lata ankietuje się osoby z wybranego zbioru gospodarstw domowych, za każdym razem tych samych gospodarstw. Podczas wywiadu członkowie gospodarstw są pytani o rozmaite zagadnienia, co pozwala na budowę obrazu przemian dziejących się w Polsce. Więcej o tym badaniu, wynikach jak i zbiorze danych można przeczytać na stronie internetowej projektu <http://diagnoza.com>.

Zbiór danych w postaci gotowej do przetwarzania w programie R, znajduje się na stronie <https://github.com/pbiecek/Diagnoza>. Można go zainstalować poleceniem `install_github("pbiecek/Diagnoza")` po wcześniejszym włączeniu pakietu `library(devtools)`.

Cały zbiór danych jest bardzo duży i mógłby sprawiać trudności na mniejszych komputerach. Dlatego na potrzeby tego kursu przygotowaliśmy podzbiór zbioru danych z badania Diagnoza Społeczna.

Podzbiór danych nazywa się diagnoza i zawiera 38461 wierszy. Każdy wiersz to odpowiedzi innej osoby. Odpowiedzi uzyskane w badaniu ankietowym zapisane są w 36 kolumnach / zmiennych. Nazwy tych zmiennych odpowiadają numerom pytań z kwestionariusza http://diagnoza.com/pliki/kwestionariusze_instrukcje. Opisy co znaczy które pytanie znajdują się w zbiorze danych diagnozaDict.

Wybrane zmienne opisują:

- imiona respondentów,
- wagi analityczne, wynikające ze sposobu losowania,
- liczbę lat nauki, płeć, wykształcenie, wzrost, wagę, dochody,
- odpowiedzi na wybrane pytania dotyczące światopoglądu.

Source: <http://diagnoza.com/> Full dataset: <https://github.com/pbiecek/Diagnoza>

Author(s)

Source: <http://diagnoza.com/>

galton

Galton's and Pearson's Height Data for Parents and Children

Description

Height of parents and children

The word regression comes from the pioneering studies carried out by Francis Galton and Karl Pearson concerning relationship between height of parents and their children's.

The original data sets created by the scientists are available in the variables galton and pearson. Using these data sets we can describe relationship between the height of a son and his father (pearson) and relationship between the weighted average height of the parents (galton) and their son's height.

This is a data set used by Galton in 1885 to study the relationship between a parent's height and their childrens. The second dataset was used by Pearson to investigate regression.

Inches were converted into centimeters.

There are 928 rows in the galton data set and 1078 rows in the pearson data set.

[POL]

Wzrost rodziców i dzieci

Słowo regresja wywodzi się z pionierskich badań Francis Galtona i Karla Pearsona nad zależnością wzrostu dzieci i rodziców.

Oryginalne zbiory danych obu tych badaczy są dostępne w zmiennych galton i pearson. Na bazie tych zbiorów danych można opisywać zależność pomiędzy wzrostem syna a ojca (pearson) oraz zależność pomiędzy ważoną średnią z wzrostu rodziców (galton) z wzrostem syna.

Poniżej przedstawiamy pierwszych 6 wierszy z każdego ze zbiorów danych. W zbiorze danych galton znajduje się 928 wierszy a w zbiorze danych pearson znajduje się 1078 wierszy.

Author(s)

Copy from other packages with inches transformed into centimeters.

Examples

```
head(galton)
head(pearson)
```

imiona_warszawa

Names of Infants Born in Warsaw

Description

Names of newborn kids [girls and boys] in Warsaw between 2004 and 2014.

The data set warsaw_names was prepared to present methods of trend analysis and methods of text analysis.

The data set contains information on the number of babies with certain names born in Warsaw in the subsequent months of the period between 2004 and 2014. This data was downloaded from the website <http://gorny.edu.pl/imiona/index.php> where you can also find information about the current number of births.

There are 84816 rows in this set and the data concerns different names, from very popular names to rather rare ones.

The columns of this data set present respectively:

- gender and name of the child,
- year and month in which the data was gathered,
- statistics of first names in a given year and month.

[POL]

Imiona noworodków w Warszawie

Zbiór danych imiona_warszawa przygotowaliśmy po to, by przedstawiać metody analizy trendów oraz metody analizy napisów.

W tym zbiorze danych zawarto informację o liczbie urodzonych noworodków o danym imieniu w Warszawie w kolejnych miesiącach w okresie od roku 2004 do 2014. Dane te są pobrane ze strony <http://gorny.edu.pl/imiona/index.php>, na której można znaleźć również informacje o aktualnej liczbie urodzin.

W tym zbiorze danych jest 84816 wierszy, dane dotyczą 696 różnych imion od tych najpopularniejszych do bardzo rzadkich.

Kolejne kolumny w tym zbiorze danych przedstawiają:

- imię i płeć dziecka,
- rok i miesiąc którego dotyczy ta statystyka,
- liczbę noworodków o danym imieniu urodzonych w Warszawie w danym roku i miesiącu.

Author(s)

Bartłomiej & Agata Gorny [keeping statistics at <http://gorny.edu.pl/imiona/index.php>]; Krzysiek Trajkowski [web scapping]; Przemysław Biecek [conversion to R]

Examples

```
head(imiona_warszawa)
```

koty_ptaki

The Fastest Cats and Birds in the World

Description

Size, habitat, speed and weight of big cats

The data set `cats_birds` was created as a so-called toy-example. It is a small set (13 rows and 7 columns) which can be fully displayed on the computer screen. You can use it to practice basic operations on the data.

The data set juxtaposes characteristics of 13 chosen species of cats and birds presented in separate rows. The chosen species are the fastest members of both groups. Each row contains information on a different species.

The columns of this data set present respectively:

- name of the species,
- maximum weight of an individual animal expressed in kilograms and maximum body length in meters,
- maximum speed (in the horizontal flight in case of birds, not during the dive) expressed in kilometers per hour,
- habitat and length of life in years,
- the last column specifies whether the species is a big cat or a bird.

[POL]

Najszybsze koty i ptaki na świecie

Zbiór danych `koty_ptaki` powstał jako tzw. „toy-example”. Jest to niewielki (13 wierszy 7 kolumn) zbiór, który można w całości wyświetlić na ekranie i na którym można przećwiczyć podstawowe operacje na danych.

W danych zestawiono charakterystyki dla 13 wybranych gatunków przedstawionych w kolejnych wierszach. Wybrane gatunki to najszybsi przedstawiciele kotów i ptaków. Każdy wiersz to informacja o innym gatunku.

Kolejne kolumny w tym zbiorze danych przedstawiają:

- nazwę gatunku,
- maksymalną osobniczą wagę w kilogramach i maksymalną długość ciała w metrach,
- maksymalną prędkość (dla ptaków w locie poziomym, nie w nurkowaniu) w kilometrach na godzinę,
- obszar zamieszkania oraz żywotność w latach,
- ostatnia kolumna określa czy gatunek jest dużym kotem czy ptakiem.

Author(s)

Przemyslaw Biecek

Examples

```
head(koty_ptaki)
```

mandatySejmik2014

Local Government Elections in Poland 2014

Description

The data set votes2014 / mandatySejmik2014 was prepared to show ways in which spatial data can be analyzed and visualized. The data was downloaded from the website of the National Electoral Commission (<http://wybory2014.pkw.gov.pl/pl/>).

The data presents information about the number of seats in the regional parliaments in Poland won in 2014. The set consists of 16 rows and 9 columns. Each row presents results concerning a different voivodeship.

The subsequent columns present respectively:

- name of the voivodeship,
- number of seats won by PSL, PiS, PO, SLD and other political parties,
- what is the percentage of valid votes of the total number of votes cast by the people entitled to vote,
- longitude and latitude of the geographical centre of the voivodeship.

[POL]

Wyniki wyborów samorządowych w 2014

Zbiór danych mandatySejmik2014 przygotowaliśmy po to, by pokazać jak można analizować lub wizualizować dane przestrzenne. Dane pobrano ze stron Państwowej Komisji Wyborczej (<http://wybory2014.pkw.gov.pl/pl/>).

Dane przedstawiają informacje o liczbie wygranych mandatów w sejmikach w wyborach samorządowych w Polsce w roku 2014. Zbiór danych zawiera 16 wierszy i 9 kolumn. Każdy wiersz przedstawia wyniki dla innego województwa.

Kolejne kolumny opisują:

- nazwę województwa,
- liczbę zdobytych mandatów przez PSL, PiS, PO, SLD i inne partie,
- jaki procent z uprawnionych do głosowania stanowi procent ważnych głosów,
- długość i szerokość geograficzną środka geograficznego województwa.

Author(s)

Data from <http://wybory2014.pkw.gov.pl/pl/>

Examples

```
head(mandatySejmik2014)
```


Description

The data set called serieIMDB contains information about popularity of particular episodes of TV series.

English version of this dataset is named seriesIMDB, use setLang() first.

The data about assessment and number of votes cast for each episode of 200 most popular TV series was downloaded from the website <http://www.imdb.com>. The data set also includes names of the series and titles of particular episodes.

Each of 20122 rows describes one episode. The subsequent variables/columns give the following information:

- title of the series and title of an episode,
- number of the season and number of the episode,
- average assessment of a given episode,
- number of votes,
- identification number of the series in IMDB database.

For example ratings for Ninjago series are available here <http://www.imdb.com/title/tt1871731>

[POL]

Oceny odcinków seriali filmowych.

W zbiorze danych serieIMDB zebraliśmy informacje o popularności odcinków seriali.

Z serwisu <http://www.imdb.com> pobraliśmy dane o ocenach oraz liczbie głosów oddanych na dany odcinek dla 200 najpopularniejszych seriali telewizyjnych. W zbiorze danych są również umieszczone nazwy seriali oraz nazwy poszczególnych odcinków.

Każdy wiersz opisuje jeden odcinek, wierszy w sumie jest 20122. Kolejne zmienne/kolumny opisują:

- nazwę serialu, nazwę odcinka,
- numer sezonu, numer odcinaka w sezonie,
- średnia ocena danego odcinka,
- liczba oddanych głosów,
- identyfikator serialu używany w bazie IMDB.

Author(s)

Przemysław Biecek [conversion to R]

Examples

```
head(serieIMDB)
```

`setLang`*Change Column Names / Datasets Names to Polish or English*

Description

The Data Crunchers course is available in two language versions – Polish and English. However, it was not necessary to create two language versions of data sets and thus double the data. For this reason your data loaded from the `PogromcyDanych` package will be in Polish.

If you want to switch to English use the command `setLang()` as in the example below. This function will translate the Polish names in the data sets (as well as names of columns and values) into English.

As a result of the function the following data sets will be copied into the namespaces:

Polish name -> English name

`koty_ptaki` -> `cats_birds`

`imiona_warszawa` -> `warsaw_names`

`mandatySejmik2014` -> `votes2014`

`auta2012` -> `auta2012`

`WIG` -> `WIG`

`pearson` -> `pearson`

`galton` -> `galton`

`skiJumps2013` -> `skiJumps2013`

`diagnoza` -> `diagnosis`

`diagnozaDict` -> `diagnosisDict`

`serieIMDB` -> `seriesIMDB`

`TCGA_BRCA` -> `TCGA_BRCA`

The `setLang('pol')` command will translate the names back into Polish.

[POL]

Dwie wersje językowe

Kurs „Pogromcy Danych” jest prowadzony w dwóch wersjach językowych, polskiej i angielskiej. Aby nie kopiować dwukrotnie danych dla obu wersji językowych po wczytaniu pakietu `PogromcyDanych` dostępne są dane w języku polskim.

Aby przełączyć się na wersję angielską należy użyć polecenia `setLang()` tak jak na poniższym przykładzie. Ta funkcja przetłumaczy polskie nazwy danych (też nazwy kolumn i wartości) na ich angielskojęzyczne odpowiedniki.

W wyniku jej działania do przestrzeni nazw wkopiowane będą następujące zbiory danych:

Nazwa polska -> Nazwa angielska

`koty_ptaki` -> `cats_birds`

`imiona_warszawa` -> `warsaw_names`

`mandatySejmik2014` -> `votes2014`

auta2012 -> auta2012
WIG -> WIG
pearson -> pearson
galton -> galton
skiJumps2013 -> skiJumps2013
diagnoza -> diagnosis
diagnozaDict -> diagnosisDict
serieIMDB -> seriesIMDB
TCGA_BRCA -> TCGA_BRCA
Instrukcja setLang("pol") tłumaczy z powrotem na polskie nazwy.

Usage

```
setLang(lang = "eng")
```

Arguments

lang A value 'eng' or 'pol'.

Value

The function setLang() changes column names in all datasets in the PogromcyDanych package to English or to Polish, depending on 'lang' argument.

Author(s)

Przemyslaw Biecek

Examples

```
## Not run:  
  setLang()  
  
## End(Not run)
```

Description

Results of ski jumping competitions

The data set skiJumps2013 contains results of ski jumping competitions season 2013/2014. It was created during construction of a diagram for a competition for the best data visualization during PAZUR conference.

The data set consists of 2130 rows. Each row describes one jump during the ski jumping competitions season 2013/2014.

Each jump is described by 16 columns. Their content is described in the data set skiJumps2013labels.

The skiJumps2013 data set contains the following information about each particular jump:

- information about competition – city, country, parameters of a ski jump,
- information about the ski jumper – name, surname, nationality, date of birth,
- information about the jumps themselves (two jumps if the jumper jumped twice or one jump if the jumper did not qualify for the second round) – speed, distance, points added to the classification.

[POL]

Wyniki skoków narciarskich

W zbiorze danych skiJumps2013 zebrane są wyniki skoków narciarskich z sezonu 2013/2014. Ten zbiór danych został przygotowany na potrzebę konkursu na wizualizację danych podczas konferencji PAZUR.

W zbiorze danych znajduje się 2130 wierszy, każdy wiersz opisuje jeden oddany skok w zawodach w skokach narciarskich z sezonu 2013/2014.

Każdy skok opisuje 16 kolumn, znaczenie poszczególnych kolumn jest opisane w zbiorze danych skiJumps2013labels.

W zbiorze danych skiJumps2013, dla każdego oddanego skoku znaleźć można informacje o:

- konkursach, takie jak: miejscowość, kraj, parametry skoczni,
- skoczkach, takie jak: imię, nazwisko, narodowość, data urodzin,
- skokach (dwóch, jeżeli skoczek skakał dwa razy lub jednego jeżeli nie zakwalifikował się do drugiej serii): prędkość, odległość, punkty do klasyfikacji.

Author(s)

Data from PAZUR <http://estymator.ue.poznan.pl/pazur/> with small transformations by Przemyslaw Biecek

Examples

```
head(skiJumps2013labels)
head(skiJumps2013)
```

TCGA_BRCA

Selected Variables from BReast CAncer Patients of The Cancer Genome Atlas Project

Description

A subset of clinical and genomic data concerning breast cancer patients (mainly female patients) from the database The Cancer Genome Atlas (TCGA) <http://cancergenome.nih.gov/>.

The Cancer Genome Atlas study collects clinical and genomic data from different tumors. This dataset incorporates selected variables from patients with BReast CAncer.

The data set TCGA_BRCA contains data about 999 patients. Each row corresponds to one patient and gives five basic pieces of information:

- information on TP53 gene mutation,
- patient's gender,
- information whether the patient lived for five years after the surgery,
- number of days between the surgery and death (if it occurred),
- information whether there was a relapse and a new tumor.

More details about TCGA project: <http://cancergenome.nih.gov/>

[POL]

Z bazy danych The Cancer Genome Atlas (TCGA) <http://cancergenome.nih.gov/> pobrano podzbiór danych klinicznych i genetycznych pacjentów (głównie pacjentek) z nowotworem piersi.

W zbiorze danych TCGA_BRCA zebrano wyniki dla 999 pacjentów. Każdy wiersz to jeden pacjent. Dla każdego pacjenta podane jest 5 cech:

- informacja o mutacji genu TP53,
- płeć pacjenta,
- informacja czy pacjent żyje pięć lat po zabiegu,
- liczba dni od operacji do zgonu (jeżeli wystąpił),
- informacja o tym czy doszło do wznowy i czy pojawił się nowy guz.

Author(s)

The Cancer Genome Atlas (data source)

Examples

```
head(TCGA_BRCA)
```

WIG

Warszawski Indeks Giełdowy (Warsaw Stock Exchange Index)

Description

WIG, index of the Warsaw Stock Exchange.

Originally an acronym for *Warszawski Indeks Giełdowy* (Warsaw Stock Exchange Index). WIG is the oldest index of the Warsaw Stock Exchange, introduced on the WSE's first trading session on 16 April 1991.

The data set WIG was prepared so that we could take a look at the analysis and visualization of data with longer time series. In this case the data concerns daily quotations on a stock exchange.

Information about the daily performance for WIG (Warsaw Stock Exchange Index) from the period December 2013-November 2014 were downloaded from the webpage of the Warsaw Stock Exchange.

The subsequent columns present respectively:

- data of quotation,
- opening and closing share price, minimum and maximum share price,
- change in share price and the turnover in thousands of PLN.

There are 248 rows in the data set and each row corresponds to one day of quotation. The Warsaw Stock Exchange is closed on Saturdays, Sundays and holidays and the number of quotation days is respectively smaller.

[POL]

Indeks WIG z Giełdy Papierów Wartościowych

Zbiór danych WIG przygotowaliśmy po to, by przyjrzeć się analizie i wizualizacji danych o dłuższych szeregach czasowych, w tym przypadku o dziennych notowaniach na giełdzie.

Ze strony Giełdy Papierów Wartościowych <http://www.gpwinfostrefa.pl/GPWIS2/pl/index/> pobrano dzienne notowania dla indeksu WIG (Warszawski Indeks Giełdowy) z okresu grudzień 2013 - listopad 2014.

W kolejnych kolumnach przedstawiono:

- datę, której dotyczą notowania,
- kurs otwarcia, zamknięcia, kurs minimalny i maksymalny,
- zmianą kursu oraz wartość obrotów w tysiącach złotych.

W zbiorze danych jest 248 wierszy, każdy wiersz odpowiada notowaniom z jednego dnia. W soboty, niedziele i święta giełda jest zamknięta, stąd ta liczba dni z notowaniami.

Author(s)

Przemysław Biecek

Examples

```
head(WIG)
```

Index

* database

setLang, [10](#)

* datasets

auta2012, [3](#)

galton, [5](#)

imiona_warszawa, [6](#)

koty_ptaki, [7](#)

mandatySejmik2014, [8](#)

serieIMDB, [9](#)

skiJumps2013, [11](#)

TCGA_BRCA, [13](#)

WIG, [14](#)

* package

PogromcyDanych-package, [2](#)

auta2012, [3](#)

auta2012_eng (auta2012), [3](#)

cats_birds (koty_ptaki), [7](#)

DataCrunchers (PogromcyDanych-package),
[2](#)

diagnosis (Diagnoza-Spoleczna), [4](#)

diagnosisDict (Diagnoza-Spoleczna), [4](#)

diagnoza (Diagnoza-Spoleczna), [4](#)

Diagnoza-Spoleczna, [4](#)

diagnozaDict (Diagnoza-Spoleczna), [4](#)

galton, [5](#)

imiona_warszawa, [6](#)

koty_ptaki, [7](#)

mandatySejmik2014, [8](#)

pearson (galton), [5](#)

PogromcyDanych

(PogromcyDanych-package), [2](#)

PogromcyDanych-package, [2](#)

serieIMDB, [9](#)

seriesIMDB (serieIMDB), [9](#)

setLang, [10](#)

skiJumps2013, [11](#)

skiJumps2013labels (skiJumps2013), [11](#)

TCGA_BRCA, [13](#)

votes2014 (mandatySejmik2014), [8](#)

warsaw_names (imiona_warszawa), [6](#)

WIG, [14](#)

WIG_eng (WIG), [14](#)