

# Package ‘RbyExample’

April 19, 2024

**Title** Data for the Book “R by Example”

**Version** 0.0.100

**Description** Data for the examples and exercises in the book “R by Example”.  
Jim Albert and Maria Rizzo (2012, ISBN 978-1-4614-1365-3).

**License** GPL (>= 2)

**Maintainer** Maria Rizzo <mrizzo@bgsu.edu>

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**Depends** R (>= 2.10)

**LazyData** true

**URL** <https://github.com/mariarizzo/RbyExample>

**NeedsCompilation** no

**Author** Maria Rizzo [aut, cre],  
Jim Albert [aut]

**Repository** CRAN

**Date/Publication** 2024-04-19 10:53:02 UTC

## R topics documented:

battinghistory . . . . .	2
batting_avg_2021 . . . . .	3
bball . . . . .	4
bball.men . . . . .	5
bball.women . . . . .	6
bgsu . . . . .	7
brainsize . . . . .	7
college . . . . .	8
CPUspeed . . . . .	9
draftlottery . . . . .	9
EtruscanItalian . . . . .	10
flicker . . . . .	11

four_players . . . . .	11
hubble . . . . .	12
lunatics . . . . .	13
nyc.marathon . . . . .	13
PATIENT . . . . .	14
peanuts . . . . .	15
poison . . . . .	15
rounding . . . . .	16
SiRstv . . . . .	16
snowfall . . . . .	17
statgrades . . . . .	17
twinIQ . . . . .	18
twins . . . . .	19
utley2006 . . . . .	20
wasterunup . . . . .	21
webhits . . . . .	21
world.record.mile . . . . .	22

<b>Index</b>	<b>23</b>
--------------	-----------

---

battinghistory	<i>Baseball Batting History Data</i>
----------------	--------------------------------------

---

### Description

Major League Baseball data on batting; number of hits, doubles, home runs by season. The data was extracted from baseball-reference.com website.

### Usage

battinghistory

### Format

140 obs. of 27 variables:

**Year** season

**Tms** number of teams

**N.Bat** number of players

**BatAge** batter's average age

**R** runs scored

**G** games played

**PA** plate appearances

**AB** at-bats

**H** hits

**X2B** doubles

- X3B** triples
- HR** home runs
- RBI** runs batted in
- SB** stolen bases
- CS** number caught stealing
- BB** walks
- SO** strikeouts
- BA** batting average
- OBP** on-base percentage
- SLG** slugging percentage
- OPS** OBP plus SLG
- TB** total bases
- GDP** ground into double plays
- HBP** hit by pitches
- SH** sacrifice hits
- SF** sacrifice flies
- IBB** intentional walks

**Note**

This version of the data is sorted in ascending order of Year. There are missing values, especially in early years.

**Source**

baseball-reference.com.

---

batting\_avg\_2021      *Batting Averages 2021*

---

**Description**

Batting data for all Major League players with at least 300 at-bats for the 2021 season. Data is from the Lahman database available through the Lahman package.

**Usage**

batting\_avg\_2021

**Format**

231 obs. of 5 variables:

**Player** Name of player

**lgID** League

**H** Hits

**AB** At bats

**AVG** Batting average

---

bball

*Men's and Women's NCAA Basketball Data*

---

**Description**

Description: Game averages for NCAA basketball

**Usage**

bball

**Format**

43 obs. of 20 variables:

**Season** season

**Teams** number of teams

**G** average number of games played

**FG** average number of field goals

**FGA** average number of field goal attempts

FG% field goal percentage

**3P** average number of three pointers

**3PA** average number of three point attempts

**3P%** three-point percentage

**FT** average number of free throws

**FTA** average number of free throw attempts

FT% free-throw percentage

**TRB** average number of total rebounds

**AST** average number of assists

**STL** average number of steals

**BLK** average number of blocks

**TOV** average number of turnovers

**PF** average number of personal fous

**PTS** average number of points scored

**Year** Year season started

**Gender** factor: "M" or "W" (men or women)

**Details**

The data is from Sports Reference <https://www.sports-reference.com/cbb/seasons/game-averages.html>

**Source**

Sports Reference

---

bball.men

*Men's NCAA Basketball Data*

---

**Description**

Description: Game averages for NCAA men basketball

**Usage**

bball.men

**Format**

77 obs. of 20 variables:

**Season** season

**Teams** number of teams

**G** average number of games played

**FG** average number of field goals

**FGA** average number of field goal attempts

FG% field goal percentage

**3P** average number of three pointers

**3PA** average number of three point attempts

**3P%** three-point percentage

**FT** average number of free throws

**FTA** average number of free throw attempts

FT% free-throw percentage

**TRB** average number of total rebounds

**AST** average number of assists

**STL** average number of steals

**BLK** average number of blocks

**TOV** average number of turnovers

**PF** average number of personal fouls

**PTS** average number of points scored

**Year** Year season started

**Details**

The data is from Sports Reference <https://www.sports-reference.com/cbb/seasons/game-averages.html>

**Source**

Sports Reference

---

bball.women

*Women's NCAA Basketball Data*

---

**Description**

Description: Game averages for NCAA women basketball

**Usage**

bball.women

**Format**

43 obs. of 20 variables:

**Season** season

**Teams** number of teams

**G** average number of games played

**FG** average number of field goals

**FGA** average number of field goal attempts

FG% field goal percentage

**3P** average number of three pointers

**3PA** average number of three point attempts

**3P%** three-point percentage

**FT** average number of free throws

**FTA** average number of free throw attempts

FT% free-throw percentage

**TRB** average number of total rebounds

**AST** average number of assists

**STL** average number of steals

**BLK** average number of blocks

**TOV** average number of turnovers

**PF** average number of personal fous

**PTS** average number of points scored

**Year** Year season started

**Details**

The data is from Sports Reference <https://www.sports-reference.com/cbb/seasons/game-averages.html>

**Source**

Sports Reference

---

bgsu	<i>BGSU Enrollment</i>
------	------------------------

---

**Description**

BGSU Enrollment

**Usage**

bgsu

**Format**

Data frame of selected BGSU enrollment data: 16 obs. of 2 variables

**Year** Year.

**Enrollment** Enrollment.

**Source**

J. Albert

---

brainsize	<i>Brain Size and Intelligence Data</i>
-----------	---

---

**Description**

Data from a study comparing brain size and intelligence.

**Usage**

brainsize

**Format**

40 obs. of 7 variables:

**Gender** Male or Female.

**FSIQ** Full Scale IQ scores based on four Wechsler (1981) subtests.

**VIQ** Verbal IQ scores based on four Wechsler (1981) subtests.

**PIQ** Performance IQ scores based on four Wechsler (1981) subtests.

**Weight** Body weight in pounds.

**Height** Height in inches.

**MRI\_Count** total pixel count from the 18 MRI scans.

**Note**

There are missing values in Weight (2) and Height (1).

**Source**

Willerman et al (1991).

---

college

*College Rating Data*

---

**Description**

College Rating Data

**Usage**

college

**Format**

260 obs. of 11 variables:

**School** Name of Institution.

**Enrollment** Enrollment of Institution.

**Tier** Ranking in tiers 1, 2, 3, 4.

**Retention** Pct. of freshmen who return the following year

**Grad.rate** Pct. of freshmen who graduate in six years

**Pct.20** Pct. of classes with 20 or fewer students

**Pct.50** Pct. of classes with 50 or fewer students

**Full.time** Pct. of faculty hired full-time

**Top.10** Pct. of incoming students who were in top 10% of high school class

**Accept.rate** Acceptance rate of students who apply

**Alumni.giving** Pct. of alumni who contribute financially



**Note**

There are missing values.

**Source**

US News and World Report "America's Best Colleges" 2009 report, National Universities.

---

CPUspeed

*CPU Speed Data*

---

**Description**

Maximum Intel CPU speed vs time from 1994 through 2004.

**Usage**

CPUspeed

**Format**

27 obs. of 6 variables:

**year** calendar year

**month** month

**day** day

**time** time in years

**speed** Max IA-32 Speed (GHz)

**log10speed** logarithm base 10 of speed

---

draftlottery

*Draft Lottery Data*

---

**Description**

Data from the 1970 military draft lottery. The lottery assigned numbers to potential draftees by their birth date. Those with lower draft numbers were drafted first.

**Usage**

draftlottery

**Format**

31 obs. of 13 variables

**Day** Day of month.

**Jan** Draft numbers for January birthdays by day of month.

**Feb** Draft numbers for February birthdays by day of month.

**Mar** Draft numbers for March birthdays by day of month.

**Apr** Draft numbers for April birthdays by day of month.

**May** Draft numbers for May birthdays by day of month.

**Jun** Draft numbers for June birthdays by day of month.

**Jul** Draft numbers for July birthdays by day of month.

**Aug** Draft numbers for August birthdays by day of month.

**Sep** Draft numbers for September] birthdays by day of month.

**Oct** Draft numbers for October birthdays by day of month.

**Nov** Draft numbers for November birthdays by day of month.

**Dec** Draft numbers for December birthdays by day of month.

**Note**

This is the data in "draft-lottery.txt".

**References**

Moore, David S. and George P. McCabe (1989). Introduction to the Practice of Statistics.

See Fienberg, S. E. (1971), Starr, N. (1997), and "Draft Lottery (1969)", Wikipedia.org for further discussion.

---

EtruscanItalian

*Etruscan-Italian Data*

---

**Description**

This data provides measurements of ancient Etruscan skulls and modern Italian skulls.

**Usage**

EtruscanItalian

**Format**

154 obs. of 2 variables:

**x** skull measurement

**group** character: Etruscan or Italian

---

flicker

*Flicker Data*

---

### Description

Critical flicker frequency and iris color of the eye for 19 individuals.

### Usage

flicker

### Format

19 obs. of 2 variables:

**Colour** Eye colour: Brown, Green, or Blue

**Flicker** Critical flicker frequency in cycles/sec.

### Details

Critical flicker frequency is the highest frequency at which the flicker in a flickering light source can be detected by the individual.

### Source

<http://www.statsci.org/data/general/flicker.txt>

<https://gksmyth.github.io/ozdasl/general/flicker.html>

### References

Smyth, Gordon K (2011). Australasian Data and Story Library (OzDASL). <https://gksmyth.github.io/ozdasl>.

---

four\_players

*Four Players Home Plate Statistics*

---

### Description

Grouped hit and home run data over regions over the zone for four players over the 2018-2023 baseball seasons. From Baseball Savant <https://baseballsavant.mlb.com/>

### Usage

four\_players

**Format**

64 obs. of 12 variables:

**PX** interval of values of plate\_x

**PZ** interval of values of plate\_z

**BIP** count of balls in play

**H** count of hits

**HR** count of home runs

**H\_Rate** hit rate

**HR\_Rate** home run rate

**Z\_H** z-score of hit rate

**Z\_HR** z-score of home run rate

**Player** chr: Player name

**px** midpoint of PX interval

**pz** midpoint of PZ interval

---

hubble

*Hubble Space Telescope Data*

---

**Description**

Distances and velocities measured for 24 galaxies containing Cepheid stars to measure the Hubble constant.

**Usage**

hubble

**Format**

24 obs. of 3 variables:

**Galaxy** A label to identify the galaxy (a factor)

**Velocity** Relative velocity in kilometers per second

**Distance** Distance in Mega parsecs

**Source**

Freedman et al. 2001. The Astrophysical Journal 553:47-72: Tables 4 and 5.

**References**

Freedman et al. (2001) Final results from the Hubble space telescope key project to measure the Hubble constant. The Astrophysical Journal (553), 47-72. Wood, S.N. (2017) Generalized Additive Models: An Introduction with R. CRC

---

lunatics

*Massachusetts Lunatics Data*

---

**Description**

Data from an 1854 survey by the Massachusetts Commission on Lunacy.

**Usage**

lunatics

**Format**

14 obs. of 6 variables:

**COUNTY** Name of county.

**NBR** Number of lunatics by county.

**DIST** Distance to nearest mental health center.

**POP** County population 1950 (thousands).

**PDEN** County population density per square mile.

**PHOME** Percent of lunatics cared for at home.

**References**

J.M. Hunter, "Need and Demand for Mental Health Care: Massachusetts 1854," *The Geographic Review*, 77:2 (April 1987), pp 139-156.

---

nyc.marathon

*New York City Marathon Data*

---

**Description**

Gender, age, and completion time (in minutes) for 276 people who completed the 2010 New York City Marathon.

**Usage**

nyc.marathon

**Format**

276 obs. of 3 variables:

**Gender** female or male

**Minutes** Time of runner in minutes

**Age** Age of runner

---

PATIENT

*Cancer Survival Times Data*

---

### Description

Survival times of cancer patients with advanced cancer of the stomach, bronchus, colon, ovary or breast, whose treatment included supplemental ascorbate.

### Usage

PATIENT

### Format

17 obs. of 5 variables:

**stomach** survival times for stomach cancer patients

**bronchus** survival times for bronchus cancer patients

**colon** survival times for colon cancer patients

**ovary** survival times for ovary cancer patients

**breast** survival times for breast cancer patients

### Details

See the text for details on how to input this data directly from the file PATIENT.DAT.

### Note

This is the data from "PATIENT.DAT" with column headings added. As input, the data is in wide format and should be stacked (long format) for a one-way ANOVA. See the text for details.

### Source

Hand et al. (1994).

### References

Cameron and Pauling (1978).

---

peanuts	<i>Peanuts Aflatoxin Data</i>
---------	-------------------------------

---

**Description**

The peanuts data records levels of a toxin (aflatoxin) in batches of peanuts.

**Usage**

peanuts

**Format**

34 obs. of 2 variables:

**Percent** percentage of non-contaminated peanuts in the batch

**Aflatoxin** average level of aflatoxin in parts per billion

**Source**

Hand et al. (1994)

---

poison	<i>Poison Survival Data</i>
--------	-----------------------------

---

**Description**

Survival times in units of 10 hours for animals exposed to different poisons.

**Usage**

poison

**Format**

48 obs. of 3 variables:

**Time** survival time in units of 10 hours

**Poison** poison: I, II, III

**Treatment** treatment: A, B, C, D

**Source**

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York.

---

rounding	<i>Rounding First Base Data</i>
----------	---------------------------------

---

**Description**

Times required to round first base for 22 baseball players using three styles: rounding out, a narrow angle and a wide angle. The goal is to determine if the method of rounding first base has a significant effect on times to round first base.

**Usage**

rounding

**Format**

66 obs. of 3 variables:

**times** time

**method** factor with 3 levels: NarrowAngle, RoundOut, WideAngle

**block** player ID (integer)

**Source**

Hollander and Wolfe (1999) Table 7.1, page 274.

---

SiRstv	<i>NIST SiRstv Data</i>
--------	-------------------------

---

**Description**

Measurements of bulk resistivity of silicon wafers made at NIST with 5 probing instruments on each of 5 days.

**Usage**

SiRstv

**Format**

25 obs. of 2 variables:

**Instrument** replicate

**Resistance** resistance

**Details**

[https://www.itl.nist.gov/div898/strd/anova/SiRstv\\_info.html](https://www.itl.nist.gov/div898/strd/anova/SiRstv_info.html)



**Source**

<https://www.itl.nist.gov/div898/strd/anova/SiRstv.html>

**References**

NIST Standard Reference Datasets: <https://www.itl.nist.gov/div898/strd/index.html>

---

snowfall	<i>Buffalo and Cleveland Snowfall Data</i>
----------	--

---

**Description**

Total snowfall in inches for the cities Buffalo and Cleveland for the seasons 1968-69 through 2008-09.

**Usage**

snowfall

**Format**

41 obs. of 3 variables:

**SEASON** character: winter season identified by years

**Cleveland** Cleveland snowfall

**Buffalo** Buffalo snowfall

---

statgrades	<i>Statistics Grades</i>
------------	--------------------------

---

**Description**

Grades from an undergraduate statistics class at BGSU.

**Usage**

statgrades

**Format**

23 obs. of 7 variables:

**ID** Student ID; integer 1:23

**Exam1** Percent grade on Exam 1

**Exam2** Percent grade on Exam 2

**HW** Percent grade on homework

**Final** Percent grade on Final Exam

**Major** Major coded 1, 2, 3

**Group** Group coded 1, 2

---

twinIQ

*Twins IQ Data*

---

**Description**

Twins IQ Data

**Usage**

twinIQ

**Format**

Data frame of Burt's IQ data for twins: 27 obs. of 3 variables

**Foster** IQ of twin raised with foster parents.

**Biological** IQ of twin raised with biological parents.

**Social** Social class of biological parents (high, low, middle)

**Source**

Burt, C. (1966). The genetic estimation of differences in intelligence: A study of monozygotic twins reared together and apart. *Br. J. Psych.*, 57, 147-153. Data is provided in R packages faraway and UsingR.

twins

*Twins Income and Education Levels Data***Description**

The data were collected at the 16th Annual Twins Day Festival in Twinsburg, Ohio, in August 1991. 495 adult twins were interviewed. The original study aimed to investigate 'By how much will another year of schooling most likely raise one's income?' Pairs of twins provide a control on confounding factors such as intelligence, family background, etc.

**Usage**

twins

**Format**

183 obs. of 16 variables:

**DLHRWAGE** the difference (twin 1 minus twin 2) in the logarithm of hourly wage, given in dollars.

**DEDUC1** the difference (twin 1 minus twin 2) in self-reported education, given in years.

**AGE** Age in years of twin 1.

**AGESQ** AGE squared.

**HRWAGEH** Hourly wage of twin 2.

**WHITEH** 1 if twin 2 is white, 0 otherwise.

**MALEH** 1 if twin 2 is male, 0 otherwise.

**EDUCH** Self-reported education (in years) of twin 2.

**HRWAGEL** Hourly wage of twin 1.

**WHITEL** 1 if twin 1 is white, 0 otherwise.

**MALEL** 1 if twin 1 is male, 0 otherwise.

**EDUCL** Self-reported education (in years) of twin 1.

**DEDUC2** the difference (twin 1 minus twin 2) in cross-reported education.

**DTEN** the difference (twin 1 minus twin 2) in tenure, or number of years at current job.

**DMARRIED** the difference (twin 1 minus twin 2) in marital status, where 1 signifies "married" and 0 signifies "unmarried".

**DUNCOV** the difference (twin 1 minus twin 2) in union coverage, where 1 signifies "covered" and 0 "uncovered".

**Note**

There are 183 cases; 147 complete cases. Twin 1's cross-reported education is the number of years of schooling completed by twin 1 as reported by twin 2. For data analysis, the logarithm of the hourly wage is typically used instead of hourly wage.

**Source**

Guido Imbens, PhD. UCLA, Department of Economics.

**References**

Ashenfelter, Orley and Krueger, Alan. "Estimates of the Economic Return to Schooling from a New Sample of Twins." *The American Economic Review* 84.5 (Dec. 1994) 1157-1173.

---

utley2006

*Chase Utley's Hitting Data for 2006*

---

**Description**

Chase Utley's Hitting Data for 2006

**Usage**

utley2006

**Format**

160 obs. of 6 variables:

**Game** game

**Date** date

**PA** plate appearances

**AB** at-bats

**R** home runs

**H** hits

**Details**

During the 2006 baseball season, Chase Utley of the Philadelphia Phillies had a hitting streak of 35 games, which is one of the best hitting streaks in baseball history.

**Source**

J. Albert

---

wasterunup	<i>Waste Run-up Data</i>
------------	--------------------------

---

**Description**

The 'Waste Run-up' data (Koopmans 1987, p. 86) reports weekly percentage waste of cloth by five different supplier plants of Levi-Strauss, relative to cutting from a computer pattern.

**Usage**

wasterunup

**Format**

22 obs. of 5 variables:

**PT1** weekly percentage waste of cloth for Plant 1

**PT2** weekly percentage waste of cloth for Plant 2

**PT3** weekly percentage waste of cloth for Plant 3

**PT4** weekly percentage waste of cloth for Plant 4

**PT5** weekly percentage waste of cloth for Plant 5

**Note**

There are missing values.

---

webhits	<i>Webpage Hits Data</i>
---------	--------------------------

---

**Description**

The number of daily visits to the author's website was obtained using Google Analytics. The data is summarized by week.

**Usage**

webhits

**Format**

35 obs. of 2 variables:

**Week** Week number

**Hits** Number of web hits

**Source**

J. Albert

---

world.record.mile      *World Record Mile Data*

---

**Description**

Mile run world record progression as recorded by the International Amateur Athletics Federation (IAAF). The dataset includes 32 world records for men ratified by the IAAF, and 29 world records for women both in the pre-IAAF and IAAF eras.

**Usage**

world.record.mile

**Format**

276 obs. of 3 variables:

**Gender** chr: female or male

**Time** chr: time as "mm:ss"

**mm** num: The whole minutes "mm" part of Time

**ss** num: The seconds "ss" part of Time

**seconds** num: time expressed in seconds

**Athlete** chr: Name

**Nationality** chr: nationality

**Date** chr: date

**Year** num: year

**Source**

Wikipedia page [https://en.wikipedia.org/wiki/Mile\\_run\\_world\\_record\\_progression](https://en.wikipedia.org/wiki/Mile_run_world_record_progression)

# Index

## \* datasets

- batting\_avg\_2021, 3
- battinghistory, 2
- bball, 4
- bball.men, 5
- bball.women, 6
- bgsu, 7
- brainsize, 7
- college, 8
- CPUSpeed, 9
- draftlottery, 9
- EtruscanItalian, 10
- flicker, 11
- four\_players, 11
- hubble, 12
- lunatics, 13
- nyc.marathon, 13
- PATIENT, 14
- peanuts, 15
- poison, 15
- rounding, 16
- SiRstv, 16
- snowfall, 17
- statgrades, 17
- twinIQ, 18
- twins, 19
- utley2006, 20
- wasterunup, 21
- webhits, 21
- world.record.mile, 22

  

- batting\_avg\_2021, 3
- battinghistory, 2
- bball, 4
- bball.men, 5
- bball.women, 6
- bgsu, 7
- brainsize, 7

  

- college, 8

  

- CPUSpeed, 9

  

- draftlottery, 9

  

- EtruscanItalian, 10

  

- flicker, 11
- four\_players, 11

  

- hubble, 12

  

- lunatics, 13

  

- nyc.marathon, 13

  

- PATIENT, 14
- peanuts, 15
- poison, 15

  

- rounding, 16

  

- SiRstv, 16
- snowfall, 17
- statgrades, 17

  

- twinIQ, 18
- twins, 19

  

- utley2006, 20

  

- wasterunup, 21
- webhits, 21
- world.record.mile, 22