Table of Information about Marathons

The data is called race_info.

Description

An interesting data set to see the effects of goals on what should be a unimodal distrubtion of finish times. The NYT had a good article - https://www.nytimes.com/2014/04/23/upshot/what-good-marathons-and-bad-investments-have-in-common.html?rref=upshot&_r=1

The source of this data is < https://faculty.chicagobooth.edu/george.wu/research/marathon/marathon_names.htm >

Data format

A data frame with columns:

variable	class	description
year	integer	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	integer	The number of finishers at the marathon
mean_time	numeric	The average finish time in minutes.

A resampled set of runners from all marathons with more 50 runners.

The data is called marathon_sample.

Description

Each marathon will have 100 runners (50 male, 50 female) per year. So any marathon with less than 50 runners in the group will have multiple resampled runners. This data set has over 500k runners. The original data had close to 10 million runners and a few more columns. The NYT had a good article - https://www.nytimes.com/2014/04/23/upshot/what-good-marathons-and-bad-investments-have-in-common.html?rref=upshot&_r=1

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner
year	integer	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held

variable	class	description
finishers	integer	The number of finishers at the marathon

A random sample of 50% of males and females for each year of runners for all years of the New York City marathon where gender is recorded.

The data is called marathon_nyc.

Description

This data set has just over 200k runners. The NYT had a good article - https://www.nytimes.com/2014/04/23/upshot/what-good-marathons-and-bad-investments-have-in-common.html?rref=upshot&_r=1. The NYC marathon website - https://www.nyrr.org/tcsnycmarathon

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner
year	numeric	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

The full set of runners for all races during 2010.

The data is called marathon_2010.

Description

This data set has 800k runners. The NYT had a good article - https://www.nytimes.com/2014/04/23/upshot/what-good-marathons-and-bad-investments-have-in-common.html?rref=upshot&_r=1.

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner

variable	class	description
year	numeric	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

The 50% sample of male/female runners for all years of the Berlin marathon that recorded gender.

The data is called marathon_berlin.

Description

This data set has ~200k observations. Marathon website - https://www.bmw-berlin-marathon.com/en/

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner
year	numeric	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

The full set of runners for the Big Sur marathon.

The data is called marathon_big_sur.

Description

This data set has ~40k observations. Marathon website - https://www.bigsurmarathon.org/

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner

variable	class	description
year	numeric	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

The full set of runners for the Jerusalem marathon.

The data is called marathon_jerusalem.

Description

This data set has ~2.5k observations. Marathon website - https://jerusalem-marathon.com/en/home-page/

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner
year	numeric	The year of the marathon
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

All of the runners for marathons with lat and long locations

The data is called marathon_location.

Description

This data set has ~150k observations.

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm >

Data format

A data frame with columns:

variable	class	description
age	numeric	The age of the runner
gender	character	The gender of the runner (M/F)
chiptime	numeric	The time in minutes for the runner
year	numeric	The year of the marathon

variable	class	description
marathon	character	The name of the marathon
country	character	The country where the marathon was held
finishers	numeric	The number of finishers at the marathon

The data is called race_location.

Description

This data set has ~2k observations.

The source of this data is < http://faculty.chicagobooth.edu/george.wu/research/marathon/data.htm and https://simplemaps.com/data/us-cities >

Data format

A data frame with columns:

variable	class	description
marathon	character	The name of the marathon that matches all other files
marathon_clean	character	A cleaned name of the marathon
state_id	character	The two letter ID for each US state
city	character	The name of the city where the race is held
finishers	integer	The number of finishers at the marathon
mean_time	numeric	The average finish time in minutes.
lat	numeric	The lattitude of the city as listed at https://simplemaps.com/data/us-cities
lng	numeric	The longitude of the city as listed at https://simplemaps.com/data/us-cities
elevation_m	numeric	The elevation in meters above sea level as estimated from the elevatr R package.
date	Date	The approximate date of the marathon. The year is correct but the month and day changes every year and we have marked it the same.
month	numeric	Approximate month of the marathon
day	numeric	Approximate day of the month of the marathon.
year	integer	The year of the marathon