Package ‘cherryblossom’

June 25, 2020

Title Cherry Blossom Run Race Results
Version 0.1.0
Description Race results of the Cherry Blossom Run, which is an annual road race that takes place in Washington, DC.
License GPL-3
Suggests ggplot2, testthat
Encoding UTF-8
LazyData true
RoxygenNote 7.1.0
URL https://github.com/OpenIntroStat/cherryblossom
BugReports https://github.com/OpenIntroStat/cherryblossom/issues
Depends R (>= 2.10)
NeedsCompilation no
Author Mine Çetinkaya-Rundel [aut, cre]
  (<https://orcid.org/0000-0001-6452-2420>)
Maintainer Mine Çetinkaya-Rundel <cetinkaya.mine@gmail.com>
Repository CRAN
Date/Publication 2020-06-25 10:10:02 UTC

R topics documented:

run09 ................................................................. 2
run12 ................................................................. 3
run17 ................................................................. 4

Index 6
Description

Details for all 14,974 runners in the 2009 Cherry Blossom Run, which is an annual road race that takes place in Washington, DC.

Usage

run09

Format

A data frame with 14,974 observations on the following 14 variables.

- **place**: Finishing position. Separate positions are provided for each gender.
- **time**: The total run time.
- **net_time**: The run time from the start line to the finish line.
- **pace**: Average time per mile, in minutes.
- **age**: Age.
- **gender**: Gender.
- **first**: First name.
- **last**: Last name.
- **city**: Hometown city.
- **state**: Hometown state.
- **country**: Hometown country.
- **div**: Running division (age group).
- **div_place**: Division place, also broken up by gender.
- **div_tot**: Total number of people in the division (again, also split by gender).

Source

Cherry Blossom Race Results

Examples

```r
clean5
library(ggplot2)

# Finishing times by gender
ggplot(run09, aes(x = time, y = gender)) +
geom_boxplot() +
labs(
```
title = "Finishing times for 2009 Cherry Blossom Run, by gender",
x = "Time to complete the race, in minutes",
y = "Gender"
)

# Pacing times by gender
ggplot(run09, aes(x = pace, y = gender)) +
geom_boxplot() +
labs(
    title = "Pacing for 2009 Cherry Blossom Run, by gender",
x = "Average time per mile, in minutes",
y = "Gender"
)

---

**run12**

*Cherry Blossom Run data, 2012*

**Description**

Details for all 16,924 runners in the 2012 Cherry Blossom Run, which is an annual road race that takes place in Washington, DC.

**Usage**

run12

**Format**

A data frame with 16,924 observations on the following 9 variables.

- **place** - Finishing position. Separate positions are provided for each gender.
- **time** - The total run time, in minutes.
- **pace** - Average time per mile, in minutes.
- **age** - Age.
- **gender** - Gender.
- **location** - Hometown city.
- **state** - Hometown state (if from the US) or country.
- **div_place** - Division place, also broken up by gender.
- **div_tot** - Total number of people in the division (again, also split by gender).

**Source**

Cherry Blossom Race Results
Examples

```r
library(ggplot2)

# Finishing times
ggplot(run12, aes(x = time)) +
  geom_histogram(binwidth = 5) +
  labs(
    title = "Finishing times for 2012 Cherry Blossom Run,",
    x = "Time to complete the race, in minutes",
    y = "Frequency"
  )

# Pacing
ggplot(run12, aes(x = pace)) +
  geom_histogram(binwidth = 0.5) +
  labs(
    title = "Pacing for 2012 Cherry Blossom Run",
    x = "Average time per mile, in minutes",
    y = "Frequency"
  )
```

Description

Details for all 19,961 runners in the 2017 Cherry Blossom Run, which is an annual road race that takes place in Washington, DC. Most runners participate in a 10-mile run while a smaller fraction take part in a 5k run or walk.

Usage

`run17`

Format

A data frame with 19,961 observations on the following 9 variables.

- **bib**  Number on the runner’s bib.
- **name** Name of the runner, with only the initial of their last name.
- **sex**  Gender of the runner.
- **age**  Age of the runner.
- **city** Home city of the runner.
- **net_sec** Time to complete the race, after accounting for the staggered starting time, in seconds.
- **clock_sec** Time to complete the race, ignoring the staggered starting time, in seconds.
- **pace_sec** Average time per mile, in seconds.
- **event** The event the racer participated in, either the “10 Mile” race or the “5K”.

Cherry Blossom Run data, 2017
Details

There was a time limit where all 10 Mile racers had to finish by. Can you figure out what that time is?

Source

Cherry Blossom Race Results

Examples

library(ggplot2)

# Finishing times
ggplot(run17, aes(x = net_sec)) +
   geom_histogram(binwidth = 300) +
   facet_wrap(~event, nrow = 2) +
   labs(
      title = "Finishing times for 2017 Cherry Blossom Run, by event",
      subtitle = "After accounting for the staggered starting time",
      x = "Time to complete the race, in seconds",
      y = "Frequency"
   )

# Pacing
ggplot(run17, aes(x = pace_sec)) +
   geom_histogram(binwidth = 100) +
   facet_wrap(~event, nrow = 2, scales = "free_y") +
   labs(
      title = "Pacing for 2017 Cherry Blossom Run, by event",
      x = "Average time per mile, in seconds",
      y = "Frequency"
   )
Index

+Topic datasets
  run09, 2
  run12, 3
  run17, 4

run09, 2
run12, 3
run17, 4