speedsR data package

A Comprehensive R Package Providing Sports-Specific Datasets for the AIS SPEEDS Project.

speedsR Overview

SPEEDS - **Sp**ort and **E**xercise Science **E**xcellence Through **D**ata **S**cience - is a project funded by the Australian Institute of Sport that aims to elevate data science literacy among Australian sports professionals. One of the key missions of the SPEEDS initiative is to curate a repository of de-identified, sport-specific datasets. These datasets are crucial for developing data science educational resources tailored to Australian sport and exercise science professionals.

The `speedsR` R data package, specifically designed for the SPEEDS project, offers an extensive collection of sports-specific datasets. It facilitates easy access to these datasets, enabling analysis and research within the sports science context. `speedsR` can be easily downloaded from the AIS SPEEDS repository and integrated as a library in any R-based Integrated Development Environment (IDE).

Synthetic Data

The `speedsR` package includes a combination of open-source and synthetic datasets. For certain datasets collected from sport and exercise science professionals, confidentiality constraints make it unsuitable to include the original data directly in the `speedsR` package. In these cases, we provide synthetic equivalents generated using the `Synthetic Data Vault` (SDV) and `synthcity` libraries in Python, as well as the `synthpop` package in R. These synthetic datasets are designed to reflect the structure and statistical properties of the original data, preserving the privacy and confidentiality of individuals involved.

Please note that these synthetic datasets are intended primarily for teaching purposes and may not be suitable for formal statistical analyses.

Installing `speedsR` from

`speedsR` is available for installation directly from , using the `remotes` package. If you have not already, please install `remotes`:

```
install.packages("remotes", repos="https://cloud.r-project.org/")
```

Once installed, load `remotes` into your R session:

```
library(remotes)
```

In order to install packages from , ensure is also set up on your machine. If it is not installed, [download and install Git](). This will allow `remotes` to clone and install packages from directly.

> 💡 Git Installation
>
> The easiest way to check if is installed on your computer is by using the terminal (on macOS or Linux) or the command prompt (on Windows).
>
> - **macOS and Linux**:
>   - Open the Terminal.
>   - Type **`git --version`** and press Return.
>   - If you see a version number, it means Git is installed. If you see a message like "command not found," then is not installed.
>
> - **Windows**:
>   - Open the Command Prompt.

- Type **git --version** and press Enter.
- If you see a version number, it means Git is installed. If you see a message like " 'git' is not recognized as an internal or external command", then is not installed.

Now, the `speedsR` package can be installed using the following command:

```
remotes::install_github("ais-speeds/speedsR")
```

Once installed, load `speedsR` as you would any R package::

```
library(speedsR)
```

List of Available Datasets

To explore the datasets included in the `speedsR` package, you can use several functions. Each offers a different level of detail about the datasets:

- **To view the list of datasets**: Use `data(package = "speedsR")`. This command displays a simple list of all datasets in the `speedsR` package.

- **For detailed documentation**: Use `help(package = "speedsR")`. This provides more comprehensive information, including documentation for each dataset.

- **For specific information**: Use `?speedsR` to access details about the `speedsR` package itself, and `?dataset_name` (replacing 'dataset_name' with your dataset of interest) for information on a specific dataset.

Data Format in `speedsR`

Datasets in `speedsR` are provided as tibbles. A tibble is a modern take on the dataframe, part of the `tidyverse` in R. It's similar to a dataframe and is particularly convenient for data analysis and manipulation in R, offering a user-friendly structure.

Here's how you can manipulate data from `speedsR`:

```
library(speedsR)

# Load and store a dataset with a descriptive variable name.
```

```r
# Here, we use the HbmassSynth dataset as an example.
descriptive_variable_name <- HbmassSynth

# View the first few rows of the dataset
head(descriptive_variable_name)

# Get a basic descriptive statistics summary of the dataset
summary(descriptive_variable_name)
```