

Script for generating dummy data

Dietmar H. and Fabian P.

Aug 2019

Introduction

Still to fill

General Settings

The `pacman` package is an R package management tool for loading and installing packages if necessary. The following packages are used for the data generation:

- **data.table:** For filtering, grouping and transforming the data as well as fast read and write operations. This package is particularly suitable for the fast processing of large amounts of data.
- **dplyr:** This package is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges. We need it especially for the use of the pipe operator.
- **rmarkdown:** R Markdown provides an authoring framework for data science. One can use a single R Markdown file to save and execute code and generate high quality reports that can be shared with an audience.
- **ggplot2:** ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. We suppress the warning for the `pacman` library because the message that the package was built under a different R-versions should not bother the user by being displayed on the console in red. Then the required packages are loaded and installed if necessary.

```
suppressWarnings(if (!require("pacman")) install.packages("pacman"))
pacman::p_load(data.table, dplyr, rmarkdown, ggplot2)
```

For the analysis of the date functions are used which work with random numbers. To make the results reproducible we initialize the random number generator with an arbitrary number (here 100) to ensure that the random numbers are the same at every program run.

```
set.seed(100)
```

The file is read whose data is to be used for an analysis.

```
raw_data <- fread("../data/insurance_portfolio.txt")
```

Visual inspection