

Additional Material for Plotting Histograms and Pie Chart

Sudhakar Kumar

5 May 2019

Data frame for analysis

A data frame (ÇETINKAYA-RUNDEL 2019) comprising information on movies has been used. It contains 651 observations (rows), each representing a movie, and 32 variables (columns). This data frame can be loaded in **R** either by using `read.csv` function or by `load` function.

```
load(url("https://stat.duke.edu/~mc301/data/movies.Rdata"))
dim(movies)
```

```
## [1] 651 32
```

Handling the missing values in data frame

There is an object named `runtime` in `movies` data frame. `runtime` denotes the runtime of movie (in minutes). We can find the range of `runtime` in `movies` by using `range` function. `range` function returns a vector containing the minimum and maximum of the given argument(s).

```
range(movies$runtime)
```

```
## [1] NA NA
```

R returns NA and NA as minimum and maximum runtime of `movies`. It means that `runtime` object contains NA.

NA stands for **not available**. NA (Jonge 2019) is a placeholder for a missing value. All basic operations in **R** handle NA without crashing and mostly return NA as an answer whenever one of the input arguments is NA. To know, the range of `runtime` in `movies`, we need to remove the **NAs**. For this, we set the argument `na.rm` to be TRUE.

```
minMax <- range(movies$runtime, na.rm = TRUE)
paste("Minimum runtime is", minMax[1], "minutes.")
```

```
## [1] "Minimum runtime is 39 minutes."
```

```
paste("Maximum runtime is", minMax[2], "minutes.")
```

```
## [1] "Maximum runtime is 267 minutes."
```

References

ÇETINKAYA-RUNDEL, MINE. 2019. "movies.RData – Mine Çetinkaya-Rundel." <http://www2.stat.duke.edu/~mc301/data/movies.html>.

Jonge, Edwin de. 2019. "An introduction to data cleaning with R." https://cran.r-project.org/doc/contrib/de_Jonge+van_der_Loo-Introduction_to_data_cleaning_with_R.pdf.