

In this set of data report assignments (D1 to D6), we will demonstrate all major topics in the class with a rich and multivariate dataset. This dataset stems from a replication attempt of published research (Wallisch & Whritner, 2017).

We will use the raw data from this project in these assignments. All data report assignments (D1 to D6) build on each other, so don't skip one.

We suggest you implement the different assignments in separate code segments (that you upload at the end). We do check progress with waypoints (in D1-D6 quizzes) on the way.

Description of dataset

This dataset features ratings data of 400 movies from 1097 research participants.

First row: Headers (Movie titles/questions)

Row 2-1098: Individual participants

Columns 1-400: These columns contain the ratings for the 400 movies (0 to 4, and missing data)

Columns 401-421: These columns contain self-assessments on sensation seeking behaviors (1-5)

Columns 422-464: These columns contain responses to personality questions (1-5)

Columns 465-474: These columns contain self-reported movie experience ratings (1-5)

Column 475: Gender identity (1 = female, 2 = male, 3 = self-described)

Column 476: Only child (1 = yes, 0 = no, -1 = no response)

Column 477: Movies are best enjoyed alone (1 = yes, 0 = no, -1 = no response)