# Movielens Project

Adele Taylor
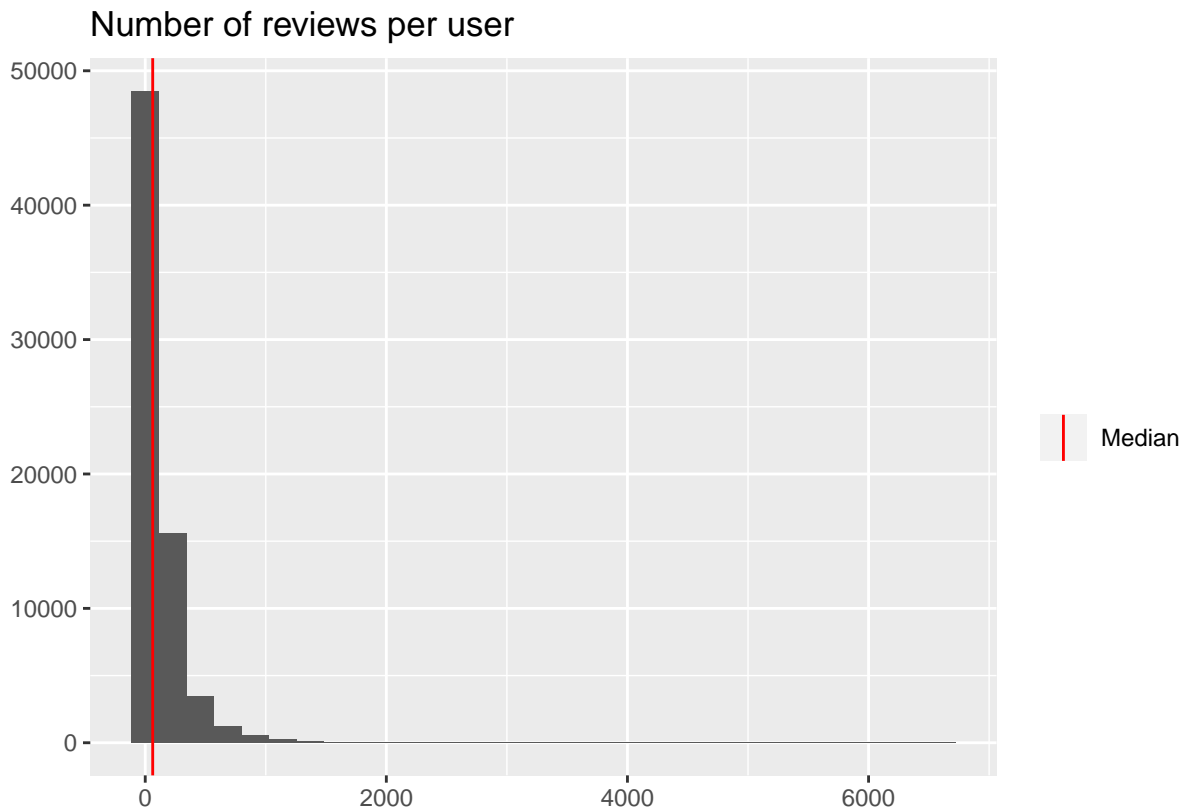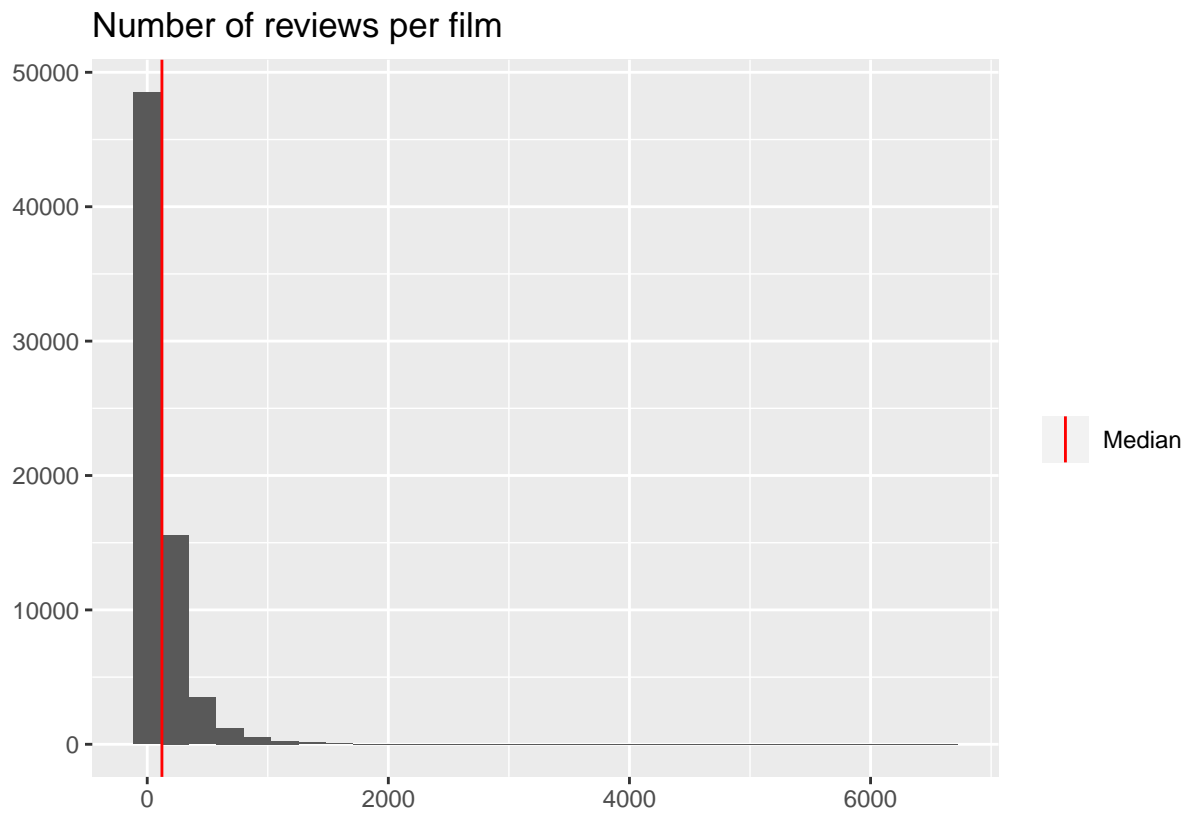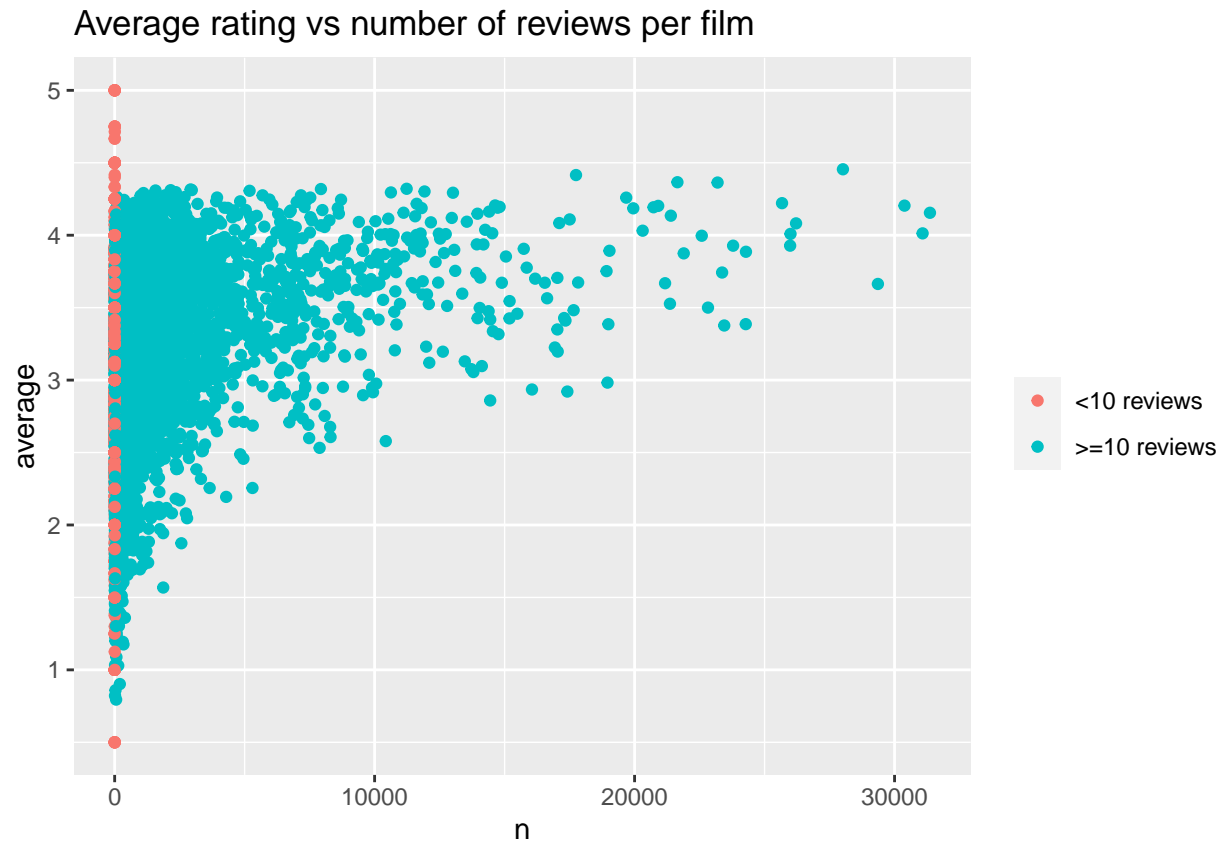
09/11/2020

## Introduction

MovieLens (movielens.org) is a non-commercial film rating and recommendation website run by the GroupLens research group at the University of Minnesota. The aim of this project is to use the 10M dataset to design and validate a movie recommendation system.

The dataset was initially divided into training and validation sets using code provided by the edX course "HarvardX PH125.9x Data Science: Capstone". All investigations and analysis was performed on the training set ("edx"), which comprised over 9 million observations consisting of the variables "userId", "movieId", "timestamp", "title" and "genres". All further reference to "the dataset" refer to this subset, unless explicitly stated otherwise.

The dataset contains ratings for movies by 'r length(unique(edx$userId)) %>% print()' users. To get an idea of how these ratings are distributed, several graphs were created as below.

## Number of reviews per user

# Number of reviews per film



Median

## Average rating vs number of reviews per film



**Analysis**

**Results**

**Conclusion**