IDS 2017
Assignment 1

Bogdan Petre (s3480941)
Low Daniel (s3120155)
Xu Teng Andrea (s3548120)
Group 7

September 9, 2017

1

1.1 Identify Data Types (10P)

- Brightness as measured by a light meter: continuous because each measurement obtains a distinct score[1], quantitative (ratio) because this device could have an absolute zero (i.e., absence of light).
- Brightness as measured by people's judgments: if you use a Lickert scale to measure people's judgment, then brightness would be discrete and qualitative (ordinal) becuase the measurements have a logical order but do not reflect numerical true values.
- Time in terms of AM or PM: binary, qualitative (nominal if one considers there is not a logical order between AM and PM or ordinal if one views PM coming after AM).
- Coat check number (certain places offer you to leave your coat to someone who, in turn, gives you a number tag that you need to claim it back when you leave): discrete, qualitative (ordinal or perhaps nominal if the coats aren't placed in the order of the integers).

1.2 Collect It... Link it! (50P)

In this exercise we write down enrich_script.R in order to collect additional data from the API of OMDB. How we developed the code and what we done are mostly written in the README.txt file.

1.3 Think About Types (20P)

- Title: discrete, qualitative (nominal) Discrete because there is a finite number of titles and qualitative nominal because it's like an ID
- ReleaseDate: discrete, quantitative (interval)

  Discrete for the same reason, and quantitative (interval) because they are calendar dates.
- Popularity: continuous, quantitative (ratio)

  It's continuos because there infinite numbers of popularity since they are real numbers and quantitative (ratio) because there is absolute zero
- Budget: continuous, quantitative (ratio)

  same reason as above
- Revenue: continuous, quantitative (ratio)

  same reason as above

- Genre: discrete, categorical (nominal)

  discrete because there are finite number of genres, qualitative (nominal) you cannot declare an order

- imdbRating: discrete, numeric (ratio)

  discrete because it has a precision to the first decimal number , it exists the absolute zero

- imdbVotes: continuos, quantitative (ratio)

  it's continuos because there is no finite number of imdbVotes, and ratio because we can have 0 votes.

- Director: discrete, categorical(nominal)

  discrete because there finite number of names, categorical (nominal) because they are names

- Country: discrete, qualitative (categorical)

  same reason as above

- PG rating: discrete, qualitative (ordinal)

  finite number of pg ratings, qualitative (ordinal) because you can order the ratings by the age the child can watch that movie

1.4

1.5 Bonus (+10P)

We used a scatter plot in order to visualize the data that we have and compare properly the IMDB ratings and Rotten Tomatoes ratings. Just see the png file imdb_vs_rottenT to see easily see that for some movie the ratings are pretty close for other there is some difference. That's because IMDB uses a weigthed mean [2] with all the users votes, in the other hand Rotten Tomatoes collects all the ratings from critical professionist from all over the world. There is also a Rotten Tomatoes rating from the users see [3].

References

[1] Field, A. (2009). Discovering statistics using SPSS. Sage publications.

[2] https://math.stackexchange.com/questions/169032/understanding-the-imdb-weighted-rating-function-for-usage-on-my-own-website

[3] https://www.rottentomatoes.com/about/