Data Exploration and Association Rules Mining Lab

Use this LBDE to gain experience doing data exploration, visualization, and association rules mining.

Ultimate Goals

- try to find something interesting about the datasets
- suggest further avenues of investigation by domain experts.

Datasets

2016collisionsfinal.csv, Transactions.csv

Tasks

Pre-Data Processing/Exploration:

a) determine what the data is reporting on / what it is about / create a "data dictionary" to explain the different fields and variables in the dataset.
b) develop a list of questions you would like to answer about the data.

1. Univariate Exploration: investigate individual variables (through charts, univariate distributions, etc.)

2. Bivariate Exploration: investigate relationship of pairs of variables (through charts, joint distributions, variable interactions (dep/indep), etc.)

3. Visualizations: create **5** multivariate visualizations (using ggplot2, etc.)

4. Questions:

a) Do you trust the data or not? Why? If you don't trust it, flag some potential issues with the data/specific entries.
b) What could we predict using this dataset / what type of questions could we try to answer?
c) How would you design a dashboard for this dataset?
d) What sort of story-telling could we carry out with this dataset?
e) Based on your initial analysis, can you suggest modifications to the data collection process?

5, Association Rules Mining: conduct an association rule mining analysis of the datasets. Using either the brute force approach or the *apriori* algorithm, determine 10-20 strong association rules. Visualize them, and interpret their results.