Assignment #1

Due Feb. 7th, 2019

electronically to smills4779@gmail.com

with Subject: StatCan 2019 Course

Data Visualization and Association Rule Mining

1. A static parallel co-ordinates plot can be done in R using information at https://rdrr.io/cran/GGally/man/ggparcoord.html or

   https://rdrr.io/cran/MASS/man/parcoord.html or https://plot.ly/r/parallel-coordinates-plot/

   We can also consider interactive parallel coordinate plots by using the information at http://www.buildingwidgets.com/blog/2015/1/30/week-04-interactive-parallel-coordinates-1 or https://www.rforge.net/doc/packages/iplots/ipcp.html

   Further information on categorical and discrete parallel co-ordinate plotting can be found at https://cran.r-project.org/web/packages/cdparcoord/vignettes/cdparcoord.html

   You may also use bubble plot to explore data – see https://towardsdatascience.com/exploring-the-census-income-dataset-using-bubble-plot-cfa1b366313b

   and here is more information on creating scatterplot matrices in R - https://www.statmethods.net/graphs/scatterplot.html

   a) *ggplot2* ships with a data set that records the carat size and the price of more than 50 thousand diamonds, from http://www.diamondse.info/ collected in 2008. Consider the analysis presented at https://www.r-bloggers.com/visualization-series-using-scatterplots-and-models-to-understand-the-diamond-market-so-you-dont-get-ripped-off/

      Use methods from Week 1 material and the above (scatterplot matrices, co-plotting, parallel coordinate plotting, ggobi representation, etc.) to explore this dataset and discover interesting relationships and patterns. Present your graphical results along with a written summary of findings.

OR

b) Try using various visualizations on the Census Income Data Set (also known as AdultUCI) Census Income Dataset to learn from this data and find interesting patterns and/or results. Present your graphical results together with a written summary of findings.

2. Refer to the dataset Census Income Data Set (also known as AdultUCI) Census Income Dataset at the UCI Machine Learning Repository with additional information on converting it to the Adult dataset for transaction processing at Adult Data Set information. Carry out association rule mining on this dataset by determining appropriate support and confidence levels and obtaining the top 10 rules.