Big Data analytics with RevoScaleR Exercises

In this set of exercise , you will explore how to handle bigdata with RevoscaleR package from Microsoft R (previously Revolution Analytics).It comes with Microsoft R client . You can get it from here . get the Credit card fraud data set from revolutionanalytics and lets get started
Answers to the exercises are available here.Please check the documentation before starting these exercise set

If you obtained a different (correct) answer than those listed on the solutions page, please feel free to post your answer as a comment on that page.

Exercise 1
The heart of RevoScaleR is the xdf file format , convert the creditcardfraud data set into xdf format .

Exercise 2

use the newly created xdf file to get information about the variables and print 10 rows to check the data .

- know how to import data and transform it for a specific moddeling or analysis goal,
- and much more.

Exercise 3
use rxSummary ,get the summary for variables gender, balance ,cardholder where numTrans is greater than 10

Exercise 4
use rxDataStep and create a variable avgbalpertran which is balance /numTran+numIntlTran .use rxGetInfo to check if your changes being reflected in the xdf data

Exercise 5
use rxCor and find the correlation between the newly created variable and fraudRisk

Exercise 6
use rxLinMod to construct the linear regression of fraudRisk on gender,balance and cardholder. Dont forget to check the summary of the model .

Exercise 7
Find the contingency table of fraudRisk and Gender , use rxCrossTab .Hint : Figure out how to include factors in the formula .

Exercise 8
use rxCube to find the mean balance for each of the two genders .

Exercise 9
Create a histogram from the xdf file on balance to show the relative frequency histogram .

Exercise 10
Create a two panel histogram with gender and fradurisk as explanatory variable to show the relative frequency of fraudrisk in two genders .