# Course Title

R and Data Mining

# Course Duration

16 hours

# Course Instructor

Dr. Yanchang Zhao

Senior Research Scientist, CSIRO, Australia

Adjunct Professor, University of Canberra

# Course Outline

This is a short course on data mining with R, presenting basics of the R programming language and popular data mining techniques, followed with examples on how to do data mining with R.

It consisting of 8 sessions below. Each session will be of two hours, composed of a 1.5hr tutorial and a 0.5hr exercise.

* R Programming  
  basics of R language and programming, parallel computing, and data import and export
* Data Exploration and Visualisation  
  summary, stats and various charts
* Regression and Classification  
  linear regression and logistic regression, decision trees and random forest
* Data Clustering  
  k-means clustering, k-medoids clustering, hierarchical clustering and density-based clustering
* Time Series Analysis  
  time series decomposition, forecasting, classification and clustering
* Association Rule Mining  
  mining and selecting interesting association rules, redundancy removal, and rule visualisation
* Text Mining  
  text mining, word cloud, topic modelling, and sentiment analysis,
* Network Analysis and Graph Mining  
  graph construction, graph query, centrality measures, and graph visualisation
* Big Data (optional)  
  Hadoop, Spark and R

# Prerequisite

* Knowledge and experience of R, or similar programming languages
* Basic knowledge of data mining and machine learning, such as
* Regression
* Classification
* Data clustering
* Association rules
* Text mining
* Time series analysis
* Social network analysis

# Requirement

You will need to bring your own laptop, if computers are not provided in the classroom. Please install the required software and R packages and download the datasets, slides and scripts below before coming to the course.

* Software and Packages
* R  
  http://www.r-project.org/
* RStudio (desktop edition)  
  http://www.rstudio.com/products/rstudio/download/
* R packages (please run the R script to install required R packages) http://www.rdatamining.com/books/rdm/code/Install-R-packages.R
* RStudio project archive [RDM-course.zip], which contains all datasets, slides and scripts below.
* Datasets
  + Titanic dataset  
    http://www.rdatamining.com/data/titanic.raw.rdata
  + Twitter dataset  
    http://www.rdatamining.com/data/RDataMining-Tweets-20160212.rds
  + Graph dataset  
    http://www.rdatamining.com/data/graph.rdata
* Slides
  + R Programming
  + Data Exploration and Visualisation with R
  + Regression and Classification with R
  + Data Clustering with R
  + Association Rule Mining with R
  + Text Mining with R
  + Time Series Analysis with R
  + Network Analysis and Graph Mining with R
  + Hadoop, Spark and R
  + R Reference Card for Data Mining
* R scripts