MSSC 6250 / Statistical Machine Learning

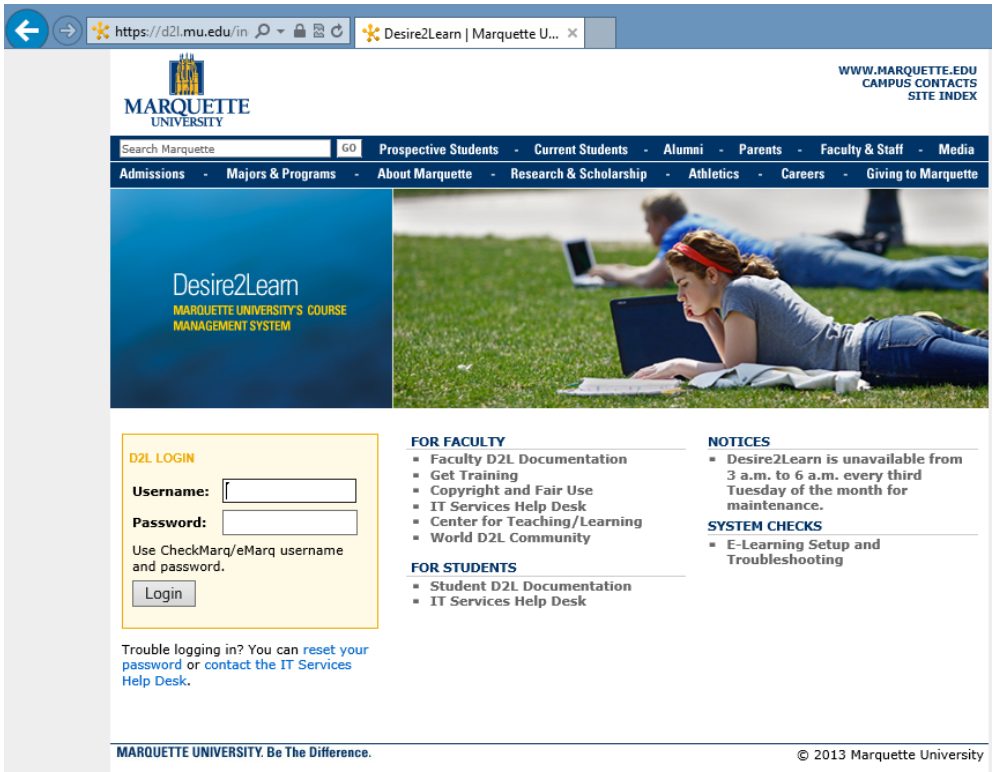
Instructor: Mehdi Maadooliat

Introduction

Department of Mathematics, Statistics and Computer Science

SYLLABUS - D2L - ONLINE MATERIALS

- Syllabus
- Go to **https://d2l.mu.edu/index.asp**
- Enter your account info and click on **Log in!**



R IS THE STATISTICAL SOFTWARE THAT WE MOSTLY USE IN THIS COURSE

- It is free!!

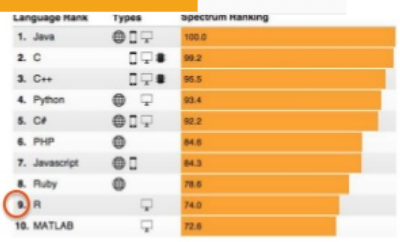
[IEEE Spectrum, July 2016]

Language Rank	Types	Spectrum Ranking
1. C		100.0
2. Java		98.1
3. Python		98.0
4. C++		95.9
5. R		87.9
6. C#		86.7
7. PHP		82.8
8. JavaScript		82.2
9. Ruby		74.5
10. Go		71.9

Language Rank	Types	
1. Java		
2. C		
3. C++		
4. Python		
5. C#		
6. R		
7. PHP		84.5
8. JavaScript		83.0
9. Ruby		76.2
10. Matlab		72.4

	84.5
	78.9
	74.3
	72.8

[IEEE Spectrum, July 2015]



- IEEE Spectrum, July 2014

- RedMonk Programming Language Rankings, 2013

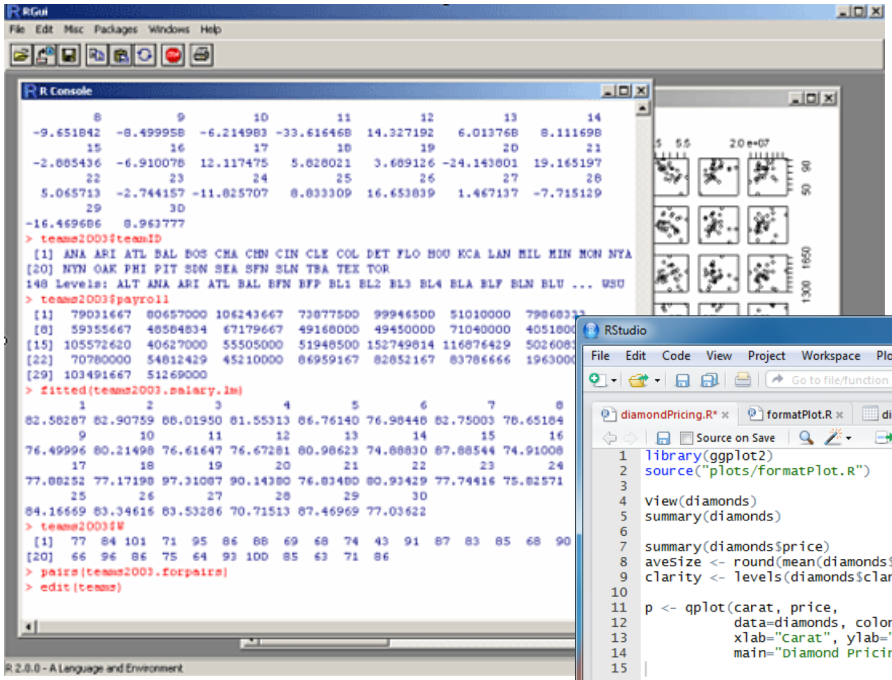
History of R (From Wikipedia)

- **R** is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years.
- R is an implementation of the S programming language combined with lexical scoping semantics inspired by Scheme. S was created by John Chambers while at Bell Labs. There are some important differences, but much of the code written for S runs unaltered.
- R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the *R Development Core Team*, of which Chambers is a member. R is named partly after the first names of the first two R authors and partly as a play on the name of S.
- R is a GNU project. The source code for the R software environment is written primarily in C, Fortran, and R. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems. R uses a command line interface; there are also several graphical front-ends for it.

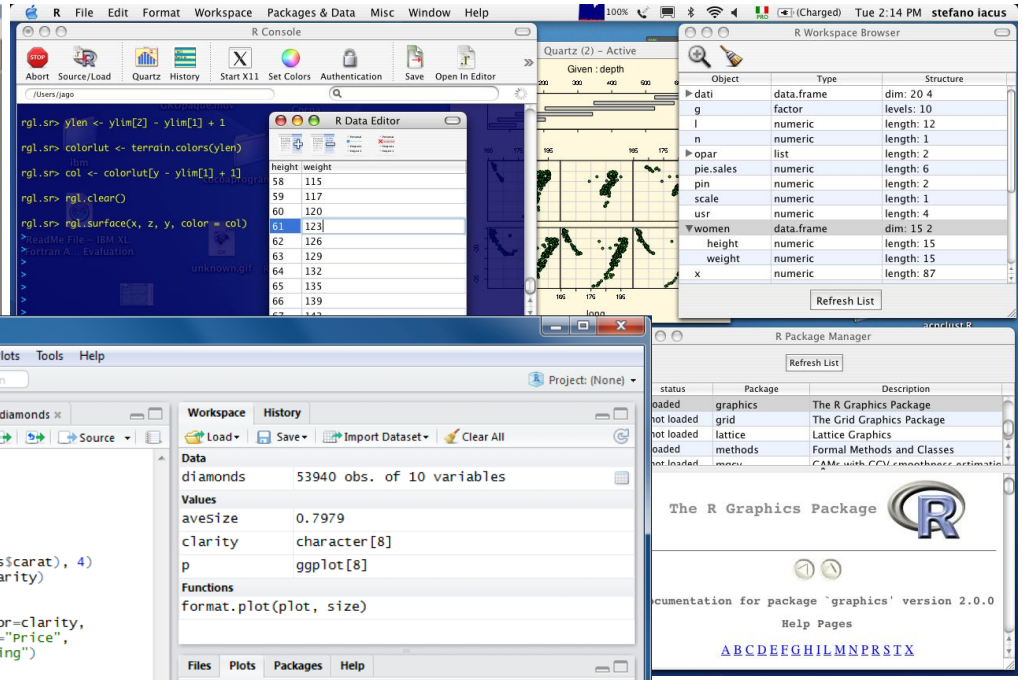
R : THE STATISTICAL SOFTWARE

- Download and Install: Comprehensive R Archive Network

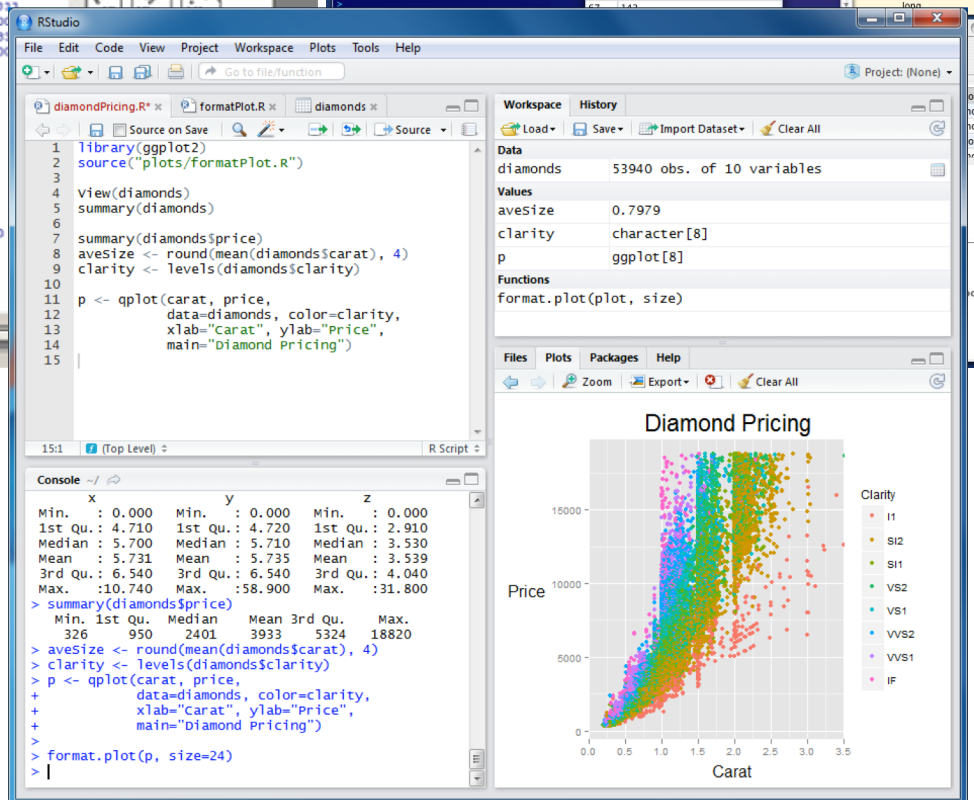
Windows



Mac

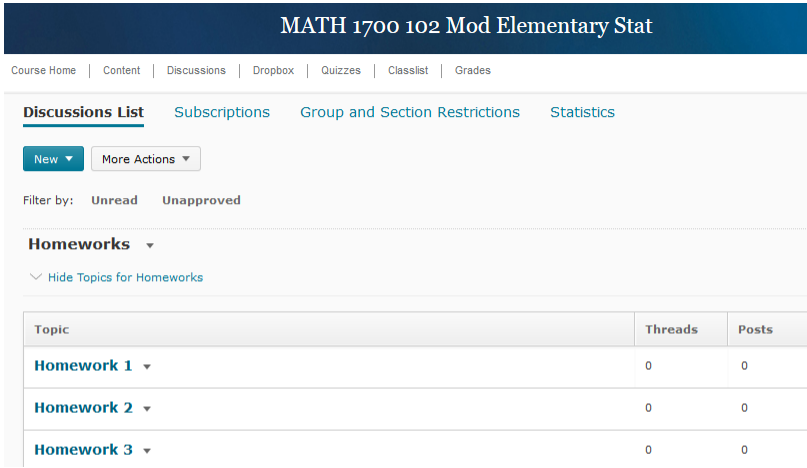


- RStudio:
  - Download



MATH 4780 CONT…

- Any General questions about Homework, **Projects** and Exams:
  - SHOULD be posted in d2l Discussion Board.
  - I will NOT answer general emails about Homework and Exams.



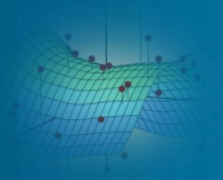
- **Homework, Projects and Exams:**
  - **Should be submitted as a PDF file (Otherwise you will get ZERO):**
    - **How to Combine Images into a PDF file [FREE & EASY + No Software] (Youtube)**
    - **Microsoft Word to PDF in 10 Seconds (Youtube)**
    - **How to: convert Images to PDF in Macbook/iMac (Youtube)**
      - http://apple.stackexchange.com/questions/11163/how-do-i-combine-two-or-more-images-to-get-a-single-pdf-file

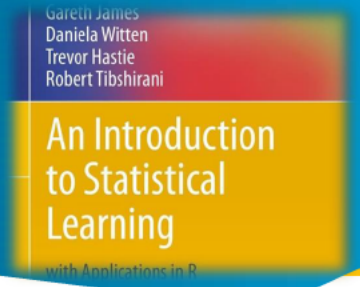
MARQUETTE UNIVERSITY

BE THE DIFFERENCE.

MSSC 6250 — SPRING 2020

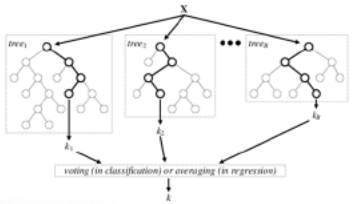
Statistical Machine Learning

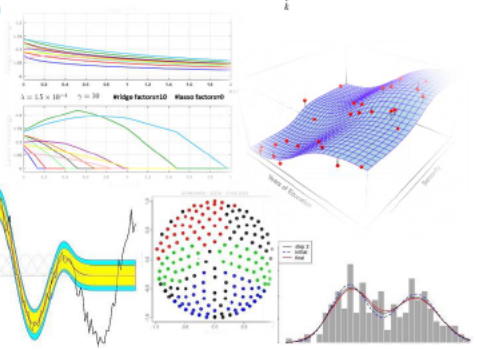


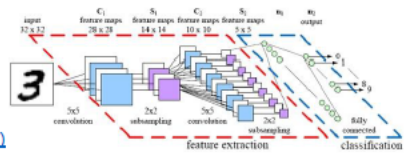


Tentative topics :

- **Multivariate Data and Exploratory Analysis**
- **Multivariate Normal Distribution**
- **Supervised Learning**:
  - Classification and Regression
    - Logistic Regression and Linear Discriminant Analysis
    - Ridge Regression and Lasso
  - Tree Based Methods
    - Classification tree and Regression tree
    - Bagging, Random Forests, Boosting
  - Support Vector Machine
- **Unsupervised Learning** (Dim. Reduction and Clustering)
  - Principal Component Analysis (PCA)
    - Human genetic clustering
    - Google PageRank
  - Independent Component Analysis (ICA)
    - Blind Source Separation
  - K-Means and Hierarchical Clustering
- **Moving Beyond Linearity** (Smoothing Splines)
- **Natural Language Processing** (NLP)
  - Bag of Words; TF-IDF; Word2Vec
- **Neural Networks**
  - Introduction to Deep Learning
    - TensorFlow and Keras, *or*
    - PyTorch and fast.ai
  - Convolutional Neural Networks (CNN)
  - Recurrent Neural Networks (RNN)







Prerequisites:

- A course in Statistical Methods
- A course in Linear Algebra

For more information, email the instructor:

- Dr. Mehdi Maadooliat (mehdi.maadooliat@marquette.edu)

R Resources: (From N. E. Helwig Homepage)

- The R Project for Statistical Computing
- The Comprehensive R Archive Network
- R Reference Card 2.0
- MATLAB ® / R Reference
- An Introduction to R
- A (Very) Short Introduction to R
- R for Beginners
- simpleR - Using R for Introductory Statistics
- Using R for Data Analysis and Graphics
- Practical Regression and Anova using R
- R Markdown Example (R code)
- Sweave Example (R code)

REVIEW

- Introduction to R and Programming (R code)
- Introduction to Linear Algebra (R code)
- Data, Covariance, and Correlation Matrix (R code)
- Multivariate Normal Distribution (R code)

QUESTIONS?

- ANY QUESTION?