

MSSC 6250 / Statistical Machine Learning

Instructor: Mehdi Maadooliat

Introduction



Department of Mathematics, Statistics and Computer Science

SYLLABUS - D2L - ONLINE MATERIALS

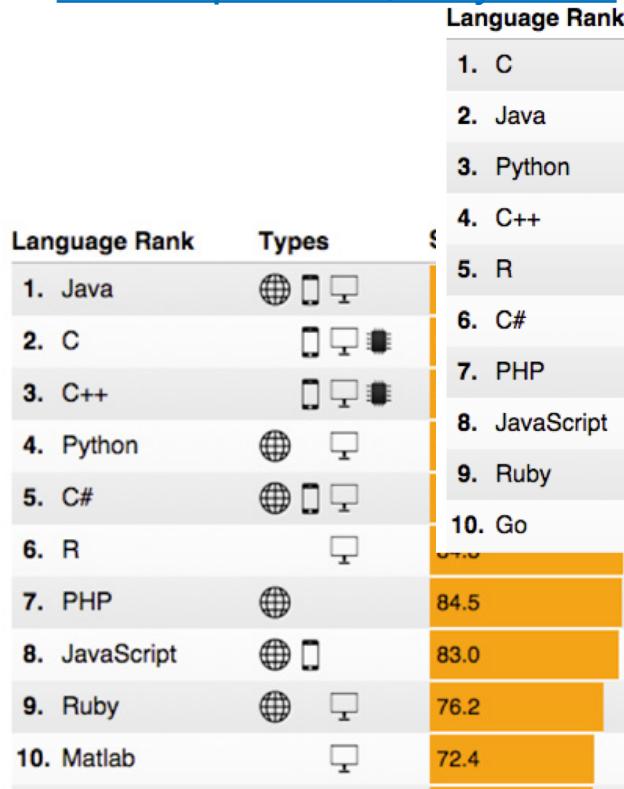
- Syllabus
- Go to <https://d2l.mu.edu/index.asp>
- Enter your account info and click on Log in!

The screenshot shows the Marquette University website with a blue header bar. The header includes the Marquette University logo, a search bar, and links for Prospective Students, Current Students, Alumni, Parents, Faculty & Staff, Media, Admissions, Majors & Programs, About Marquette, Research & Scholarship, Athletics, Careers, and Giving to Marquette. Below the header, there's a banner featuring two students working on laptops outdoors. On the left, a dark blue sidebar contains the text "Desire2Learn MARQUETTE UNIVERSITY'S COURSE MANAGEMENT SYSTEM". The main content area has a yellow "D2L LOGIN" box with fields for Username and Password, a "Login" button, and a note about using CheckMarq/eMarq credentials. To the right, sections for "FOR FACULTY" (Faculty D2L Documentation, Get Training, Copyright and Fair Use, IT Services Help Desk, Center for Teaching/Learning, World D2L Community) and "NOTICES" (Desire2Learn unavailable from 3 a.m. to 6 a.m. every third Tuesday) are visible. A "SYSTEM CHECKS" section lists E-Learning Setup and Troubleshooting. At the bottom, a note about password resets and contact information for IT Services is provided, along with the Marquette University logo and the tagline "MARQUETTE UNIVERSITY. Be The Difference."

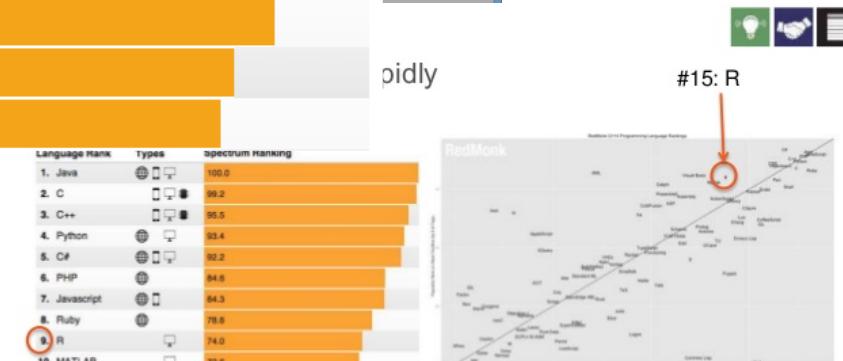
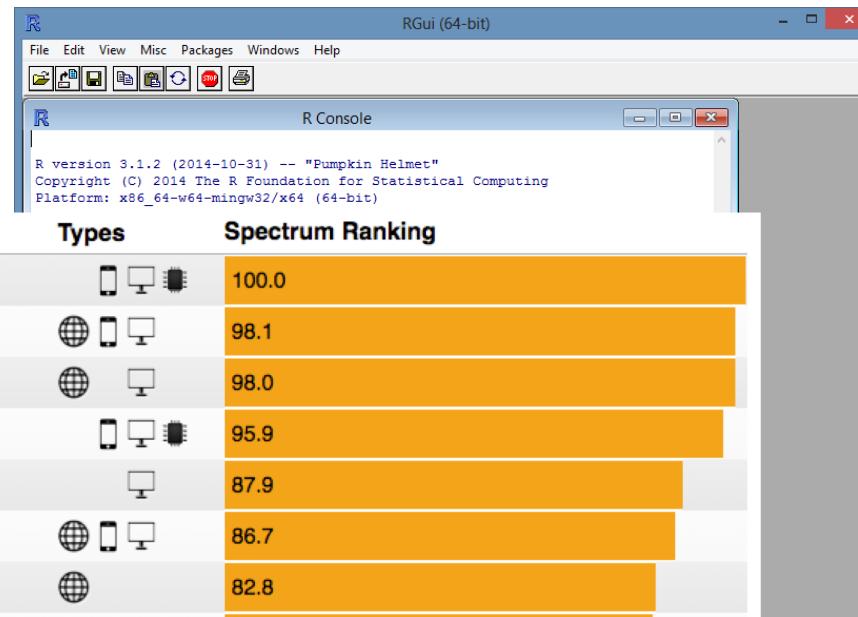
R IS THE STATISTICAL SOFTWARE THAT WE MOSTLY USE IN THIS COURSE

- It is free!!

[IEEE Spectrum, July 2016](#)



[IEEE Spectrum, July 2015](#)



- [IEEE Spectrum, July 2014](#)



- [RedMonk Programming Language Rankings, 2013](#)

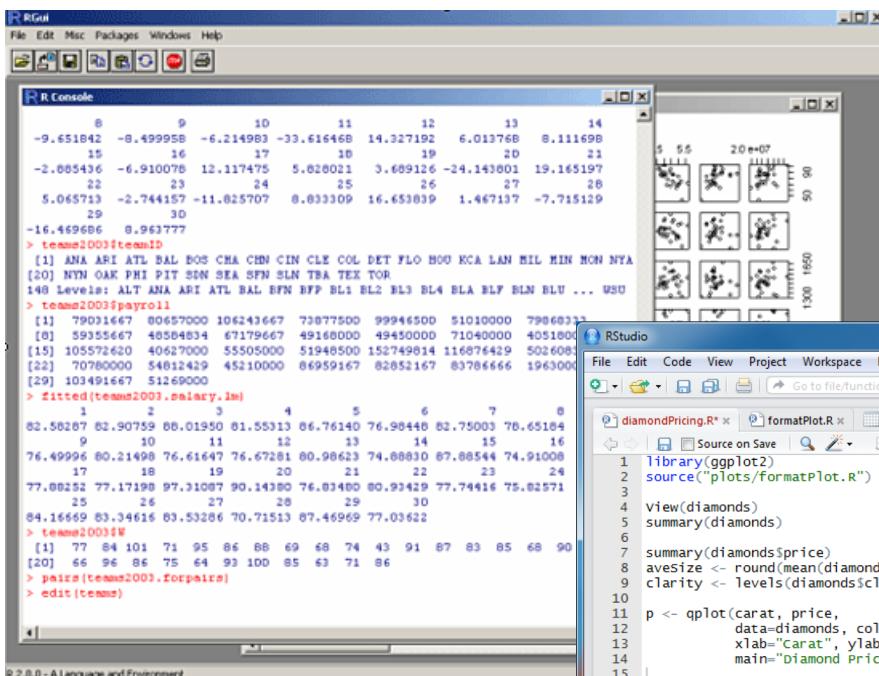
HISTORY OF R (FROM [WIKIPEDIA](#))

- 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:

Windows



Comprehensive R Archive Network

Mac

The screenshot displays the RStudio interface on a Mac. It includes a code editor with R script code, a workspace browser showing loaded packages and datasets, and a plot viewer displaying a scatter plot titled "Diamond Pricing".

Code Editor (RStudio):

```

library(ggplot2)
source("plots/formatPlot.R")
view(diamonds)
summary(diamonds$price)
avesize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)
p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")
format.plot(plot, size)
  
```

Workspace Browser:

Object	Type	Structure
dati	data.frame	dim: 20 4 levels: 10
I	numeric	length: 12
n	numeric	length: 1
opar	list	length: 2
pi.esales	numeric	length: 6
pin	numeric	length: 2
scale	numeric	length: 1
usr	numeric	length: 4
Twomen	data.frame	dim: 15 2
height	numeric	length: 15
weight	numeric	length: 15
x	numeric	length: 87

Plot Viewer:

Diamond Pricing

Clarity

Price

Carat

- 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.

MATH 1700 102 Mod Elementary Stat

Course Home | Content | Discussions | Dropbox | Quizzes | Classlist | Grades

Discussions List Subscriptions Group and Section Restrictions Statistics

New More Actions

Filter by: Unread Unapproved

Homeworks ▾

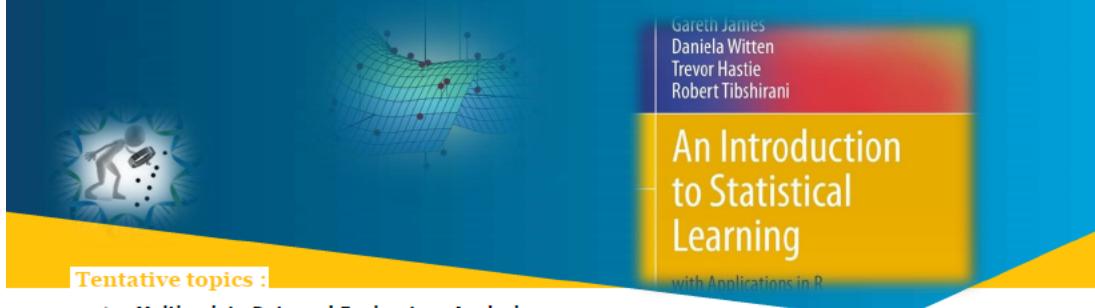
Hide Topics for Homeworks

Topic	Threads	Posts
Homework 1 ▾	0	0
Homework 2 ▾	0	0
Homework 3 ▾	0	0

- 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>

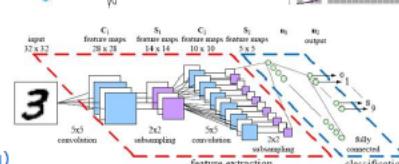
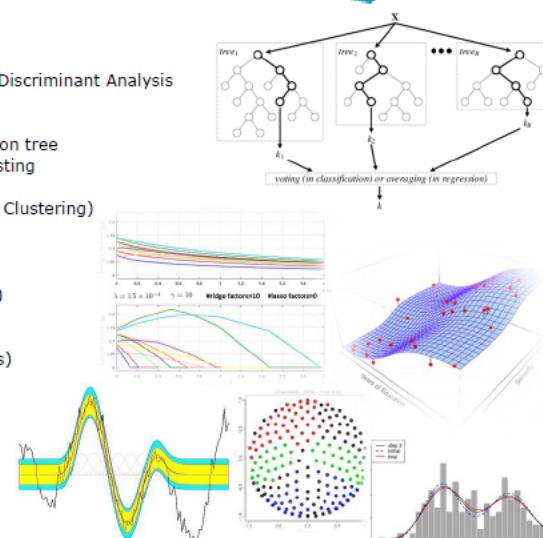
MSSC 6250 – SPRING 2020

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@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?