

MSSC 6250 / Statistical Machine Learning

Instructor: Mehdi Maadooliat

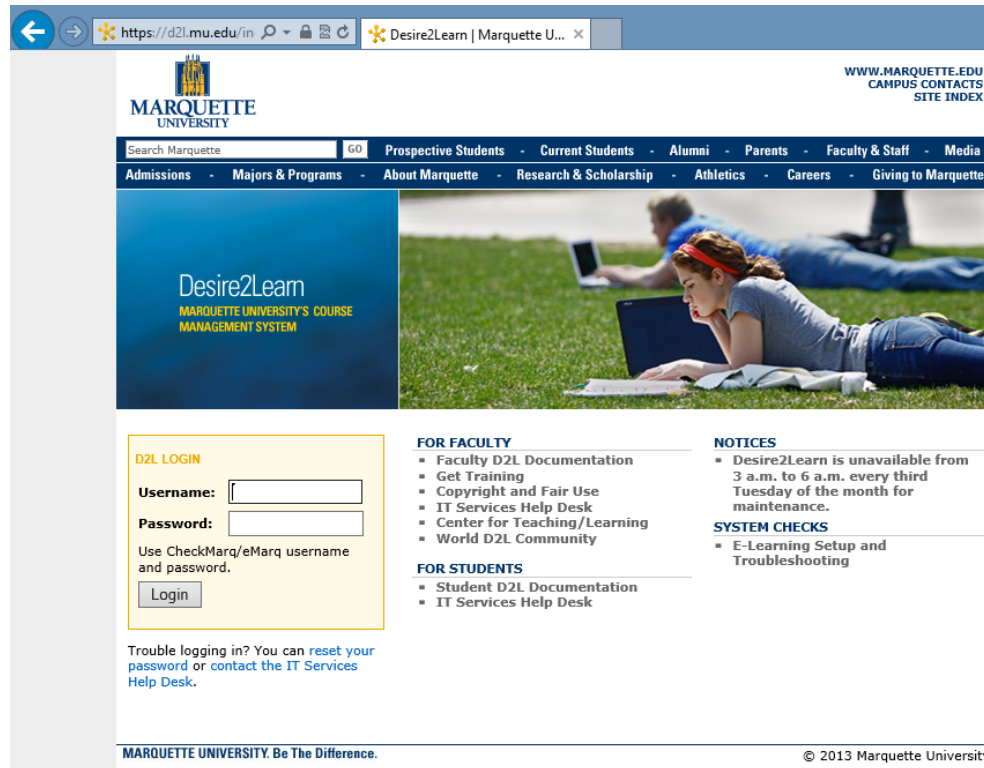
Introduction



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!**



Marquette University logo and navigation menu.

Desire2Learn
MARQUETTE UNIVERSITY'S COURSE
MANAGEMENT SYSTEM

D2L LOGIN

Username:

Password:

Use CheckMarq/eMarq username and password.

Login

Trouble logging in? You can [reset your password](#) or [contact the IT Services Help Desk](#).

FOR FACULTY

- Faculty D2L Documentation
- Get Training
- Copyright and Fair Use
- IT Services Help Desk
- Center for Teaching/Learning
- World D2L Community

FOR STUDENTS

- Student D2L Documentation
- IT Services Help Desk

NOTICES

- Desire2Learn is unavailable from 3 a.m. to 6 a.m. every third Tuesday of the month for maintenance.

SYSTEM CHECKS

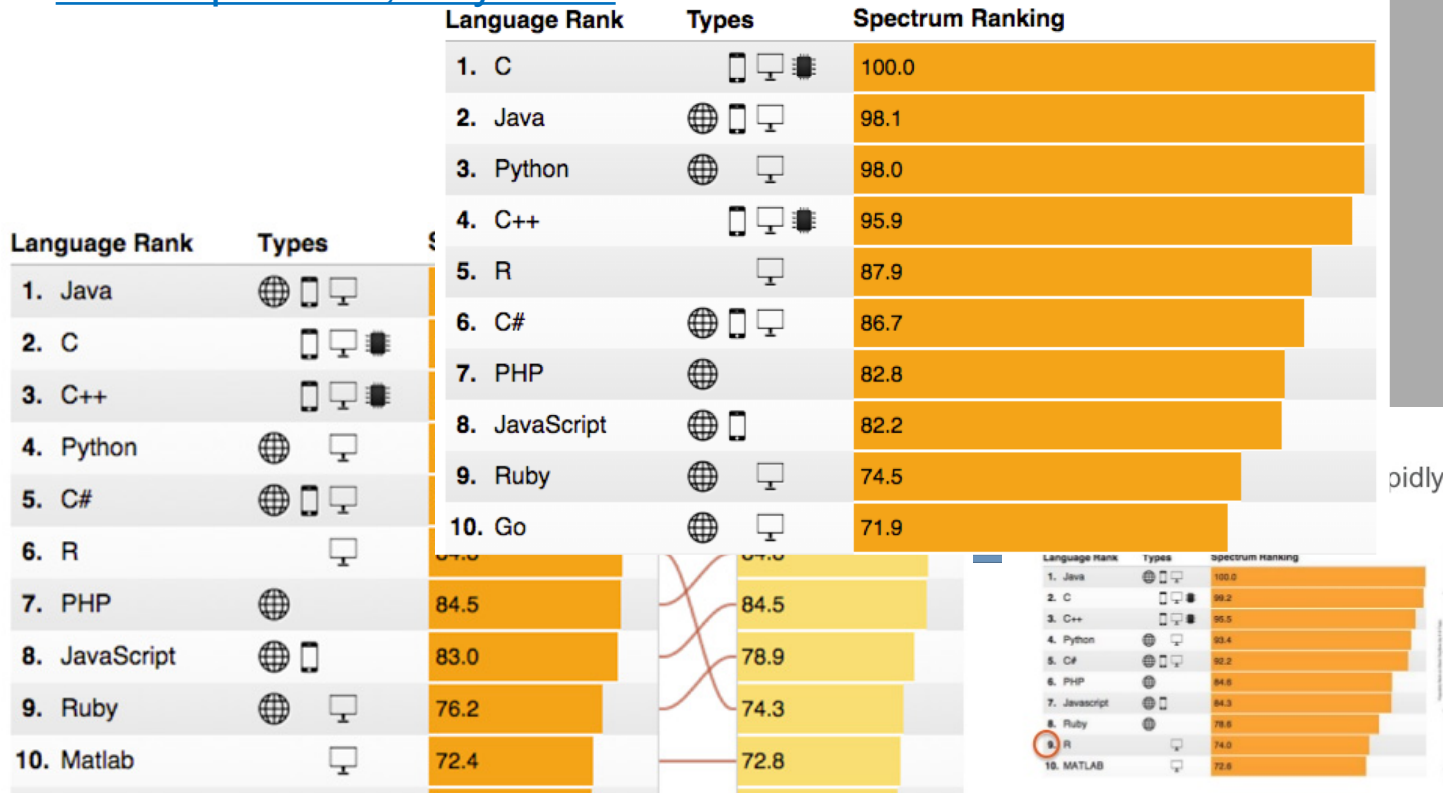
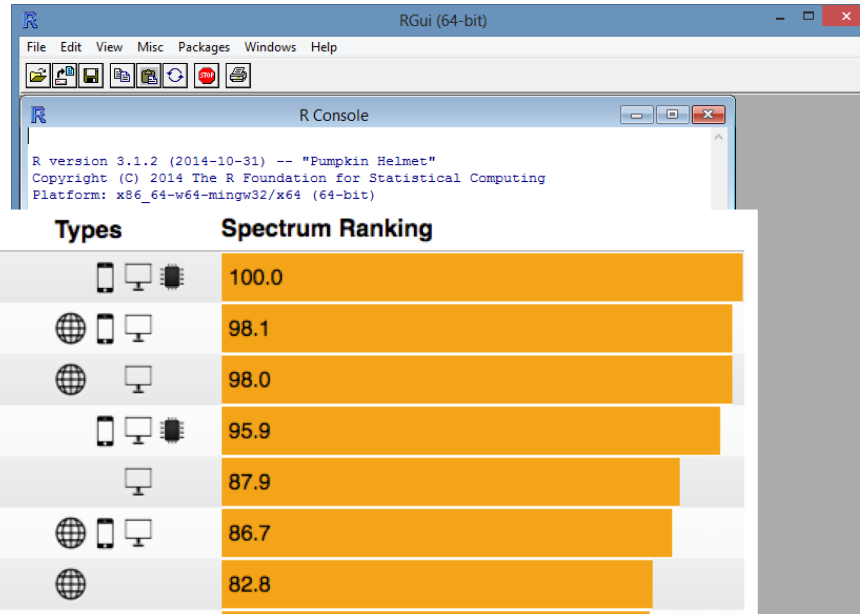
- E-Learning Setup and Troubleshooting

Marquette University logo and footer text: Be The Difference. © 2013 Marquette University

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

- It is free!!

[IEEE Spectrum, July 2016](#)



pidly

#15: R



[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 collage displays various R environments and outputs:

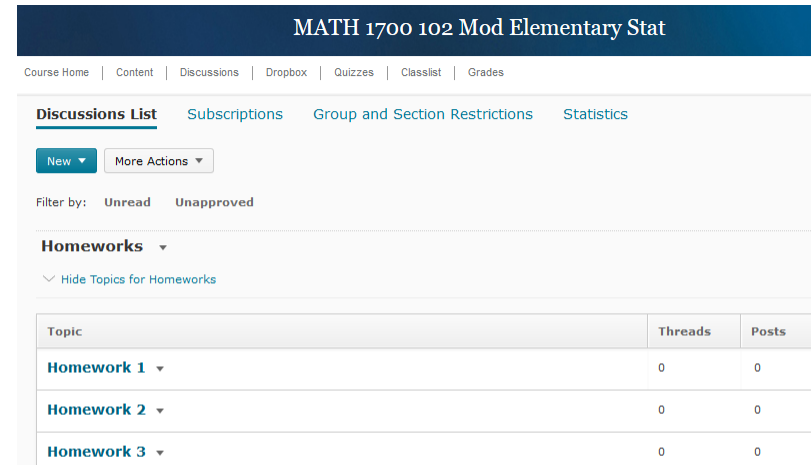
- R Console (Top Left):** Shows a payroll dataset with columns for employee ID, name, and salary. The output includes a list of employee names and their corresponding salaries.
- RStudio (Bottom Left):** The main IDE window showing a script for analyzing diamond pricing. The script includes loading the 'ggplot2' package, reading the 'diamonds' dataset, and creating a scatter plot of Price vs Carat, colored by Clarity. The console output shows the summary statistics for the diamonds dataset.
- R Console (Top Right):** Shows a script for creating a terrain plot using the 'raster' package. The output is a plot of a terrain surface.
- R Package Manager (Bottom Right):** A window showing the status of installed and available packages. It lists packages like 'graphics', 'grid', 'lattice', 'methods', and 'raster'.
- Scatter Plot (Bottom Right):** A plot titled 'Diamond Pricing' showing Price (Y-axis) versus Carat (X-axis). The points are colored by Clarity, with a legend on the right showing categories: I1, SI2, SI1, VS2, VVS2, VVS1, and IF.

- RStudio:
 - [Download](#)

MATH 4780 CONT...

- Any General questions about Homework, **Projects** and Exams:

- **SHOULD** be posted in d21 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>

Statistical Machine Learning

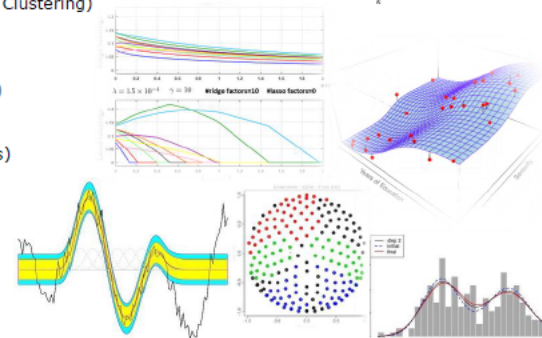
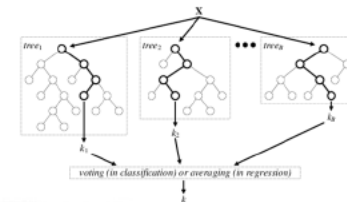
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning

with Applications in R

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)

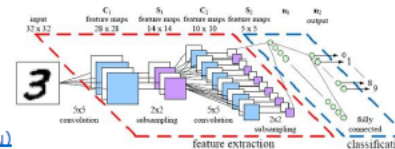


Prerequisites:

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

For more information, email the instructor:

- [Dr. Mehdi Maadooliat \(mehdi.maadooliat@michiganstate.edu\)](mailto:Dr_Mehdi_Maadooliat@michiganstate.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?