MSSC 6250/Stat. Mach. Learn.

Instructor: Mehdi Maadooliat

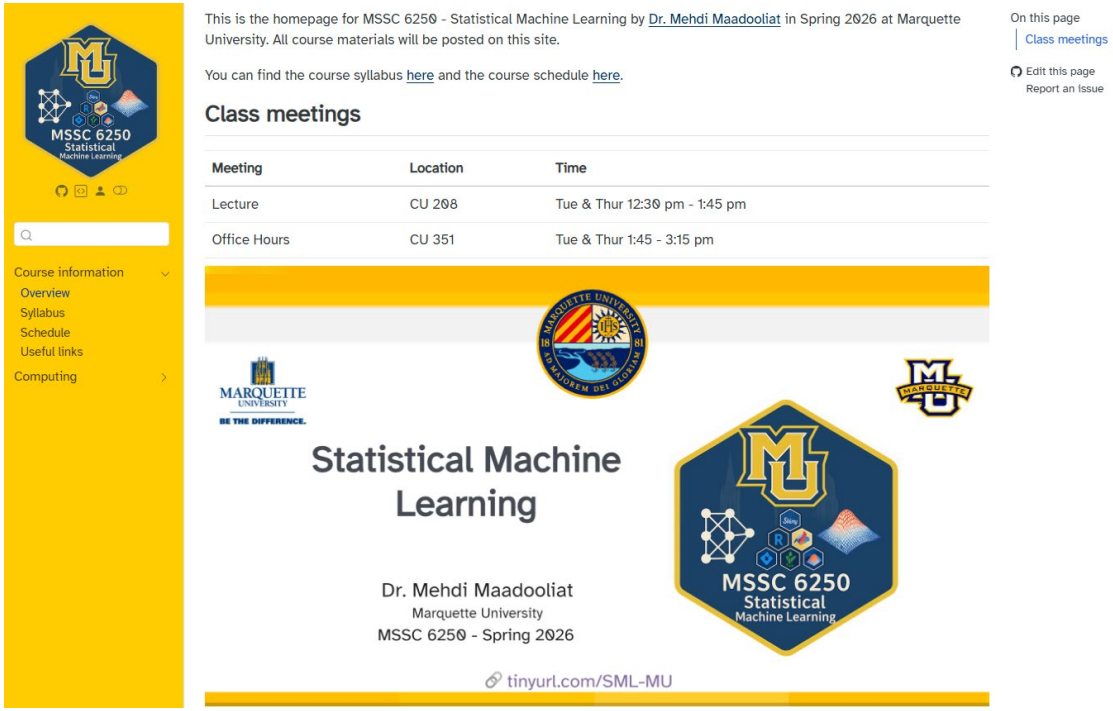
Introduction



Department of Mathematical and Statistical Sciences

SYLLABUS - D2L - ONLINE MATERIALS

- Syllabus
- Course website: http://tinyurl.com/SML-MU

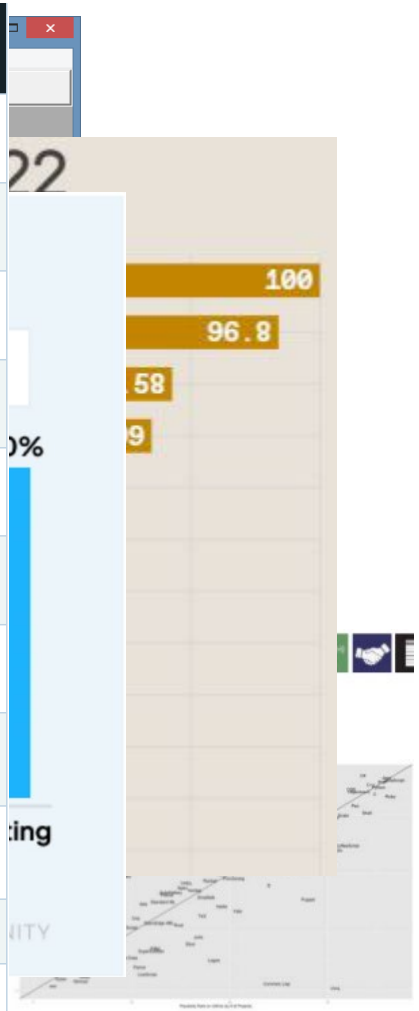


- Submissions (HW, Proj): https://d2l.mu.edu/index.asp

R VS PYTHON

MARQUETTE UNIVERSITY
Be The Difference.

Project Type	Recommended Language	Why
AI-Powered Applications	Python	Robust libraries like TensorFlow, PyTorch, and Scikit-learn simplify machine learning and AI development.
Statistical Analysis	R	Built-in statistical functions and libraries like Statsmodels and caret make R a top choice.
Web Applications	Python	Frameworks like Django and Flask enable efficient web app development and integration.
Big Data Processing	Python	Integrates with tools like Apache Spark and Dask for distributed data workflows
Academic Research	R	Focused on reproducibility, with strong support for hypothesis testing and data modeling
Real-Time Analytics	Python	Flexible for building pipelines with libraries like Pandas and integrating with streaming tools
Data Dashboards	R	Tools like Shiny and ggplot2 create interactive, publication-quality dashboards
Data Pipelines for E-commerce	Python	Handles large-scale data processing and real-time analytics for customer behavior
Healthcare Data Visualization	R	Creates detailed, professional-grade visualizations for healthcare analytics
Automation Scripts	Python	Libraries like Selenium and Beautiful Soup streamline automation tasks
Scientific Simulations	R	Offers statistical tools tailored for scientific modeling and research workflows
Interactive e-Learning Tools	Python	Supports web-based applications and AI-driven personalized learning systems



RedMonk Programming Language Rankings, 2013

http://tinyurl.com/SCiRT

HISTORY OF R (FROM WIKIPEDIA)

MARQUETTE UNIVERSITY
Be The Difference.

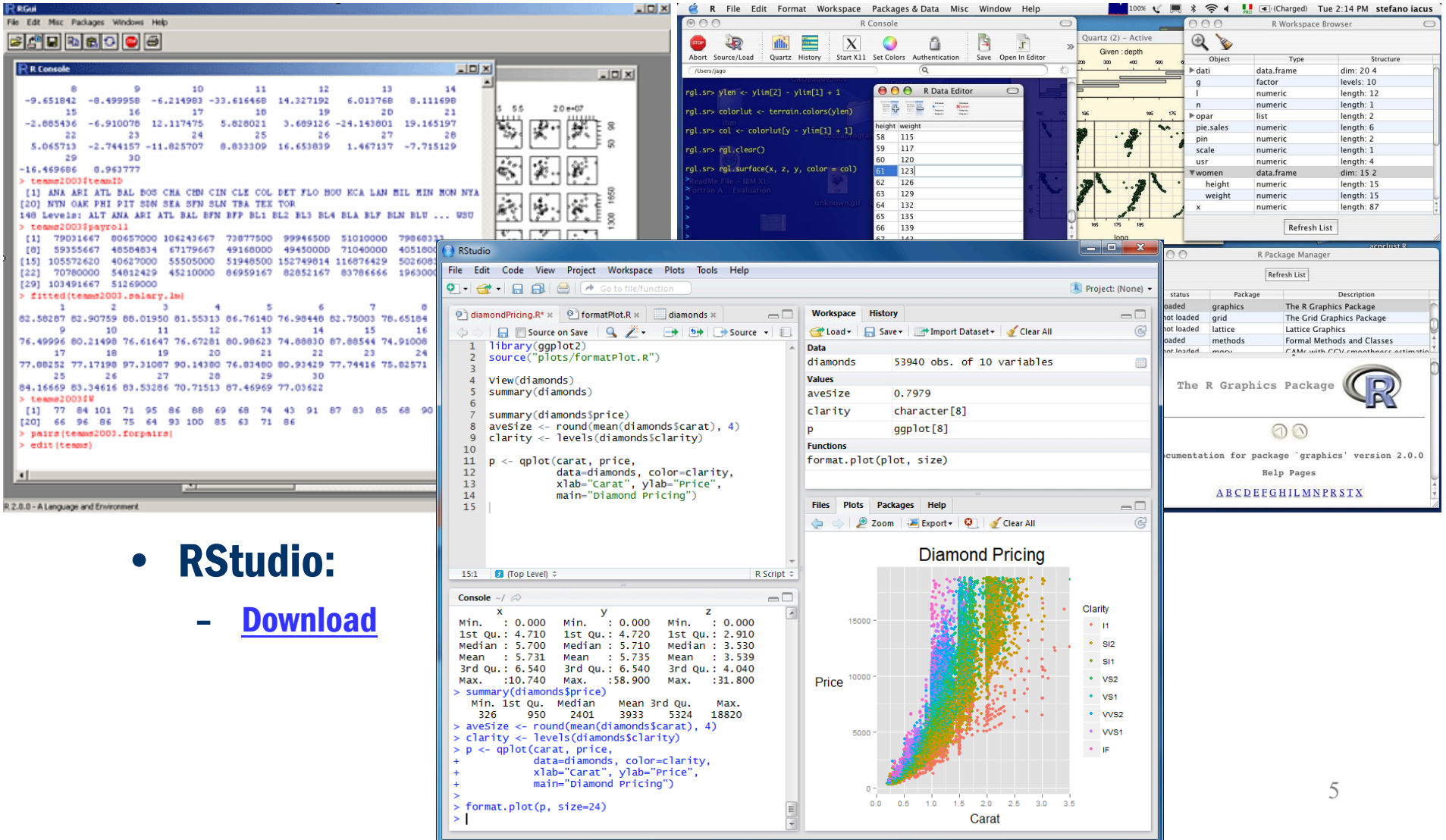
- R is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that 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

MARQUETTE UNIVERSITY
Be The Difference.

- **Download and Install:** [Comprehensive R Archive Network]

Windows **Mac**

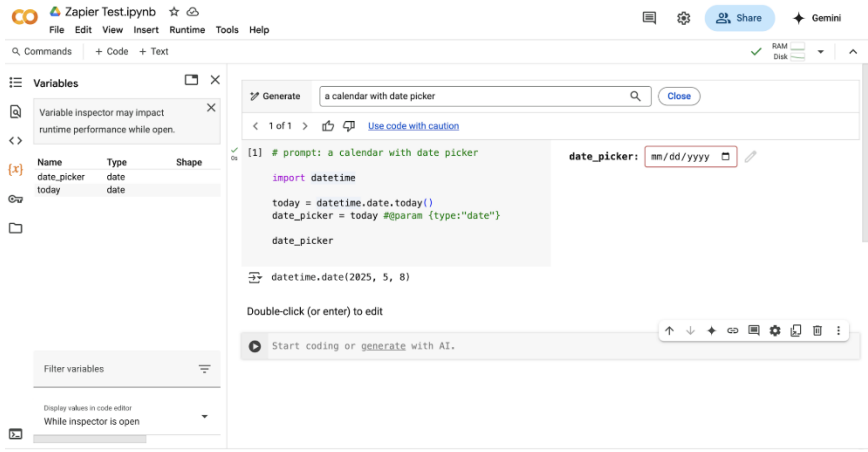


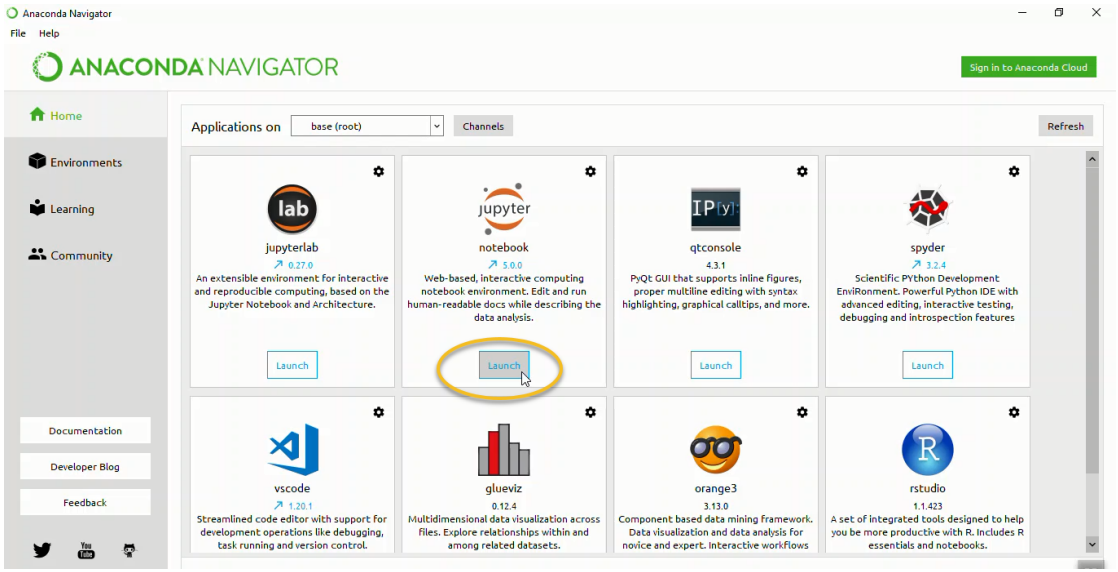
- **RStudio:**
  - [Download]

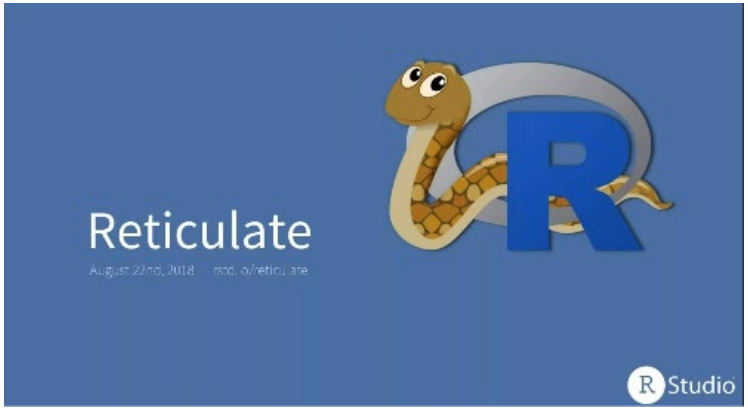
PYTHON

(ANACONDA – JUPYTER – COLAB - RETICULATE):





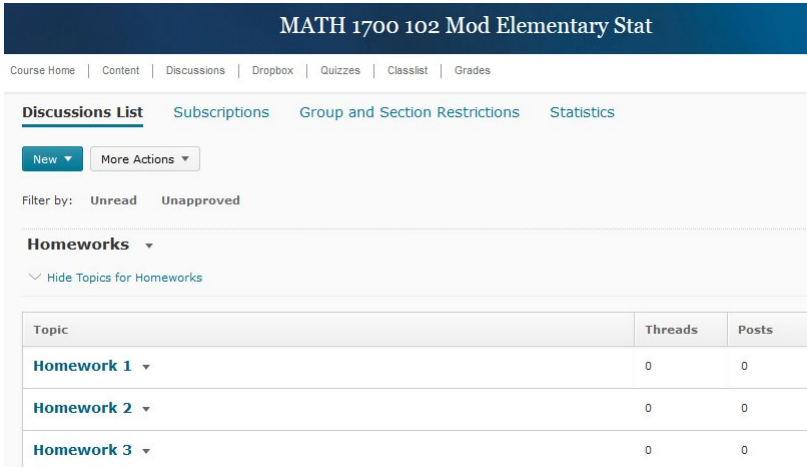




MSSC 6250 CONT…

MARQUETTE UNIVERSITY
Be The Difference.

- **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 2026

Statistical Machine Learning

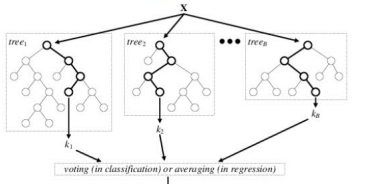
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning
with Applications in R
Second Edition

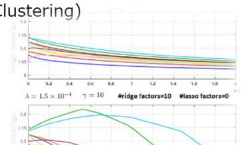
MSSC 6250
Statistical
Machine Learning

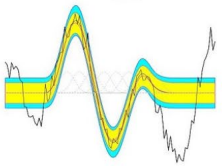
Tentative topics :

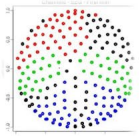
- **Multivariate Data and Exploratory Analysis**
- **Multivariate Normal Distribution**
- **Supervised Learning**:
  - Classification and Regression
    - Logistic Regression and Linear Discriminant Analysis
    - Ridge Regression and Lasso
  - Tree Based Methods
    - Classification tree and Regression 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)

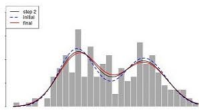






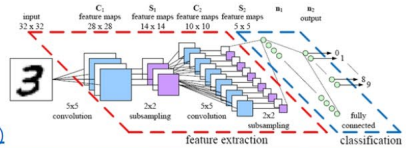






Prerequisites:

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



For more information, email the instructor:

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

MARQUETTE UNIVERSITY
Be The Difference.

SOME R RESOURCES

- 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
- Quarto

MARQUETTE UNIVERSITY
Be The Difference.

SOME PYTHON RESOURCES

- Python Software Foundation
- Python Package Index (PyPI)
- NumPy Reference
- Pandas API Reference
- The Python Tutorial
- Learn Python in Y minutes
- Python for Data Analysis
- SciPy Lecture Notes
- Scikit-learn User Guide
- Jupyter Notebook
- Matplotlib
- Reticulate (R–Python bridge)

REVIEW

- Introduction to R and Programming
- Introduction to Python (from ISLP)
- Introduction to Linear Algebra
- Multivariate Normal Distribution

QUESTIONS?

MARQUETTE UNIVERSITY
Be The Difference.

- **ANY QUESTION?**