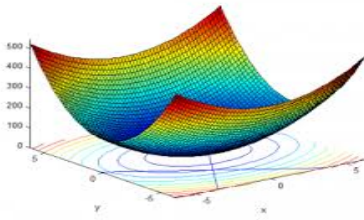# Point72

Data Driven Decision Making

Prof. Jake Feldman

jbfeldman@wustl.edu
jake-feldman.squarespace.com

Course Description: In this course, we will develop tools to analyze data, which we will then use to better guide our decision making process. Throughout the course, we will use the R programming language. The course will be broken down in to three sections, each of which is detailed below:

Data Manipulation and Plotting: In this section of the course, you will be given real-world data sets and asked to manipulate them in various ways in an effort to draw out important insights that could guide an estimation procedure or decision process. For the plotting portion we will use the R package ggplot. For the data manipulation sections we will use the R package sqldf, which will allow us to write SQL queries on R data frames. Examples of the data sets and exploratory questions we will consider are given below:

- Given the full set of results from the 2014-2015 NFL Season, how do I determine which team was the best when playing at home?

- Scoot is a electric scooter sharing service in San Fransisco similar to bikeshare services. Given usership data on rides throughout the day, can we determine the role that the charge of the scooter has on rides throughout the day?

- Swiftly is a transportation app that provides a user will all of the public transit they could need (walking, biking, bus, uber, etc ...) for traveling from A to B. Given usership data from the app, can we determine what factors influence how commuters choose their modes of transport?

Machine Learning: This portion of the course will be an introduction to machine learning. I will introduce a collection of supervised and unsupervised learning procedures, which we will apply to various data sets. We will also cover model selection and validation; which we will explore through both a qualitative and quantitative lens. The former refers to a more general procedure for assessing a model in which we ask high-level questions like: How do I assess the accuracy of my model? How and to what extent can I improve my model? The latter formalizes this general

framework by developing metrics to assess the accuracy of various models as well as control for overfitting. Another major theme of this section will be measuring and assessing the uncertainty surrounding the estimates we derive. In the final part of the course, we will study how to embed the estimation procedures we develop within an optimal decision making framework. Examples of problems we will consider are given below:

- Given Expedia booking data, can I predict which hotels a user will book?

- How can I use ticket/demographic info from passengers onboard the Titanic to predict if they survived or not?

- Can I use a player's past performance to predict fantasy football performance this week?

Simulation and Optimization: In this section, we will study optimal decision making under uncertainty. For the simulation component we will study Monte Carlo simulation. In the optimization portion of the course we will study linear and integer programming. We will begin with deterministic problems with known parameters and then build to problems where the parameters must be estimated and the problems have inherent randomness that must be accounted for. Examples of problems we will consider are given below:

- Given a metric to assess the current performance of each student in a class, what is the best (fairest?) way to form groups of 4 for an upcoming project?

- Given predictions for the fantasy football performance of all players, how do I choose the optimal one-day fantasy lineup?

- How should large music festivals develop the optimal schedule for the bands?

- Pappy's is a famous BBQ place in St. Louis. Given historical sales data of their ribs, how can they determine the optimal number of ribs to make each day?

Lecture #	Content: Data Manipulation and Plotting
Lecture 1	Session 1: • **Topic:** Introduction to R - console vs. editor, running code, data types, data structures Session 2: • **Topic:** setting working directory, reading data into R, slicing and indexing data frames Session 3: • **Topic:** Plotting using ggplot • **Dataset**: Executives data set **Homework 1:** Analyzing grades
Lecture 2	Session 1: • **Topic:** Basic SQL 1 • **Dataset**: Scoot data set Session 2: • **Topic:** Basic SQL 2 • **Dataset**: 2014/2015 NFL results Session 3: • **Topic:** Joins 1 **Homework 2:** Titanic Data Set
Lecture 3	Session 1: • **Topic:** Joins 2 • **Dataset**: Airlines data set Session 2: • **Topic:** Building Fantasy Football Tool • **Dataset**: Stats from 2014/2015 NFL Season **Homework 3:** Finish Fantasy Football Tool