HOGESCHOOL ROTTERDAM / CMI

Data Science: Visualisation and Analysis

CMIBOD022T

Number of study points: ???
Course owners: Abbey Waldron

Modulebeschrijving

Module name:	Data Science: Visualisation and Analysis
Module code:	CMIBOD022T
Study points and hours of effort for full-time students:	This module gives ???, in correspondance with 112 hours: • 8 x 3 hours lectures • 8 x 5 hours self-study for the practical assignments • 30 hours for the final project • 18 hours self-study and exam preparation
Examination:	Written examination and a final project
Course structure:	Lectures, practical assignments and presentations
Prerequisite knowledge:	Basic R Programming, Statistics.
Optional Learning tools:	• Swirl stats intro to R: http://swirlstats.com/students.html • Guide to base graphics in R: www.ling.upenn.edu/~joseff/rstudy/week4.html • Intro to ggplot2: varianceexplained.org/RData/code/code_lesson2/ • ggplot2 documentation: http://docs.ggplot2.org/current/
Learning objectives:	At the end of the course, the student can: • Learn how to explore data using visualisation and R • Learn how to apply statistics to draw conclusions • Understand how to ask good questions and design experiments • Learn how to explain your results using talks and visualisations

Content:	
	• Data Visualisation using R
	• How to make the right plot to answer the question
	• Experimental design and hypothesis statement
	• Applying statistics and mathematics to real problems
	• Presenting results in a clear way
Course owners:	Abbey Waldron
Date:	9 september 2015

1   General description

In this course you will learn about one of the most important skills of the data scientist: data visualisation. We will focus on using visualisation to understand problems and clearly explain your research to others. We will also be applying statistics and machine learning algorithms that you will learn in other parts of the course to conduct a number of small analysis projects.

1.1   Relationship with other teaching units

This course will rely heavily on material learned in the statistics module.

Form of course

One of the most important aspects of working with data is being able to communicate your results clearly with your colleagues and clients. As such, an integral part of this course will be explaining the results of each week's work to your classmates, and asking and answering each other questions.
Each week we will start with presentations of the previous week's work. After that new material and R code examples will be given, before we start work on the current week's work.
Some weeks the problems will require coding and then presenting your plots and results, and other weeks problems will be conceptual or mathematical in nature. Some weeks I will ask you to work in small groups and sometimes individually.

Workload

There will be assignments given each week. There will be time to work on them and get help during class. In addition I have scheduled 5 hours per week to work on them outside class. Please complete all the work.

2   Course Program

Lecture notes for each week will be uploaded after class. All of the problem sets will be handed out during class.

Week One

Introduction to data visualisation with R. We will go over the anatomy of a good plot, and learn how to make some standard types of plot in R. Exercises will require you to make a certain types of plots from a data set that will be provided.

Week Two

Making the right plot to answer the question. We will learn how to answer questions by making a plot. This week will involve lots of examples, some group work and then some individual coding tasks. The exact plot you need to make will no longer be given - you need to work it out.

Week Three

Functional forms, how to characterise different behaviour, correlations. Often when you plot two variables you can see some correlation between them. We will learn how to describe different types of functional relationship and suggest mathematical formulae that may be appropriate for fitting.

Week Four

Fitting data, interpolation and predictions. We will apply techniques from the machine learning lectures to a specific problem. We will discuss the importance of understanding the underlying data to guide the extent of our predictions.

Week Five

Showing distributions, averages, errors, combining errors. We will talk about performing measurements including their errors.

Week Six

Hypotheses and designing experiments. Now we will think about how to ask a good question, and then design an experiment to answer it. There will be some group work where you are asked to design experiments to test certain hypotheses.

Week Seven

Working on the final project. Everyone will complete their own project.

Week Eight

Present final projects and revision. You will be presenting the results from your final project to your classmates. Everyone will be presenting their work and answering questions.

3   Assessment

The course will be graded on the basis of an examination in week nine and a final project. The weighting of the grade will be 50-50. The deadline for handing in the final project report will be during class in week 8. As part of the final project students will be required to present their results in a short talk in week 8.