Automating Data Exploration with R

Introduction

As data scientists and analysts we face constant repetitive task when approaching new data sets. This class aims at automating a lot of these tasks in order to get to the actual analysis as quickly as possible. Of course, there will always be exceptions to the rule, some manual work and customization will be required. But overall a large swath of that work can be automated by building a smart pipeline. This is what we'll do here. This is especially important in the era of big data where handling variables by hand isn't always possible.

It is also a great learning strategy to think in terms of a processing pipeline and to understand, design and build each stage as separate and independent units.

- Reading Data
- Dates
- Categorical Data
- Raw Text Data
- Imputing Missing Data
- Basic Feature Engineering
- Basic Exploration
- Variable Importance
- Modeling - Supervised & Unsupervised