R Exploratory Data Analysis

R Brown Bag Series #3

Gary R Seamans

The MITRE Corporation

March 14, 2018

Agenda

- What is EDA?
- Demonstrations
- Questions

Overview -What is EDA?

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to:

- maximize insight into a data set
- uncover underlying structure
- extract important variables
- detect outliers and anomalies
- test underlying assumptions
- develop parsimonious models
- determine optimal factor settings

(*What Is EDA?*)

What is EDA contd.

The seminal work on EDA is by John Tukey (Tukey 1977). It is very readable and still relevant. Another excellent book is Understanding Robust and Exploratory Data Analysis (Hoaglin, Mosteller, and Tukey 1983)

Engineering Statistics Handbook (NIST) is another good reference.

EDA Examples

All of the examples demonstrate data loading, cleaning, exploratory data analysis, and a final analysis.

- Old Faithful
- Music
- On-time Arrivals
- Credit Card Default

Old Faithful

In this example we'll look at the eruptions of Old Faithful and see if there is a correlation between the length of the eruption and the time until the next eruption.

Music

This example takes music listeners preferences and does market basket analysis using association mining rules.

On-time arrivals

An example of analyzing on-time arrivals for a small number of airports.

Credit Card Default

Credit card default is a big problem for banks. This example shows how you might go about predicting who will default.

Questions

References I


Hoaglin, David C, Frederick Mosteller, and John Wilder Tukey (1983). *Understanding Robust and Exploratory Data Analysis*. Vol. 3. Wiley New York.


Tukey, John W (1977). *Exploratory Data Analysis*. Vol. 2. Reading, Mass.


What Is EDA? http://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm. (Visited on 03/14/2018).