R for Exploratory Data Analysis

Jim Crozier

Stratusfy: Data Science and Cloud Technology.

May 31, 2013

Overview I

Quick Notes

Code and documentation available at github/jimcrozier/r-meetup-rattle

Topics

Three topics for data exploration:

- rattle package
- knitr package
- shiny application

Bonus:

Bonus: R + javascript for geomapping

Overview II

Exploratory Data Analysis (EDA)

- EDA is the process of exploring, scrubing, and munging your data for analysis
- This presentation is a brief overview of how to use free and available tools in R for quickly exploring datasets
- Always work toward functionalizing repeatable task.
  - Much of EDA is a repeated on all of the datasets that you work with
  - Read between the lines in the following overviews to see ways for functionalizing and packaging your EDA steps
  - As a broad generalazation, use rattle to find tools that work for you, use knitr to make books of the techniques that work for you, and use shiny to separate the views from the models, and the controllers that move the models

rattle I

rattle Overview

rattle is a free gui available from Togaware for R

- I am not a rattle expert
- Loads of example available online
- Covering the very basics here

Getting started

Install and load gui

install.packages(rattle, dependencies=T)

- This takes some time due to the large number of dependencies

rattle II

rattle()

Make it work

Live demo

data(iris)

knitr I

Overview

knitr is a single best tool that you can learn

Overview

copy and paste rattle log into block

- Convert crs and crv to list()
- Make sure to load data that you sent to rattle (here data(iris))
- Still working out kinks with plots

knitr II

Overview

Tip: use cat() to write out .tex and looping to create integrated EDA

"Intelligence is the faculty of making artificial objects, especially tools to make tools." - Henri Bergson

shiny I

Overview

Interactivity is where data science is headed. shiny is easy way to interact your models and your views.

One of the things that I would like to see data science work towards is separating models, controllers, and views

Make it work

Live demo

Javascript with R I

Make it work

Use R, JSON and javascript (here the leaf package) to create interactive elements

Make it work

Live demo Rickshaw

Make it work

Live demo Leaf