Fundamentals in R: Getting Started

August 12, 2014

Module 2 - Getting Started

In this module we'll lay the ground work for getting started with R.

The objectives are

- get comfortable with command line, RGUI
- understand the role of packages, how to get help
- familiarize with the dataset for this course

Script Window & Command Line

Script Window

Symbol	Meaning
#	line of comments

Command Line

Prompt	Meaning
>	ready for a new command
+	awaiting the completion of an existing command

Getting help in R

From within the command line, for installed & loaded packages only:

```
?lm
help(lm)
```

Getting help in R

Some packages provide additional documentation via vignettes. List vignettes from all attached packages, for a specific package:

```
vignette()
vignette(package = "grid")
```

Packages & Functions

R is a functional programming language
http://adv-r.had.co.nz/Functional-programming.html

Functions ship in packages, along with other relevant functions, code and any
sample data helpful in getting you started

Packages & Functions

arg1 arg2 arg3 → retObj

Packages & Functions

```
myFunct <- function (arg1, arg2 = defaultVal2, arg3, ...) {
        code at the core
        of your function
        return(retObj)
        }
```

Packages & Functions

```
myFunct ( arg1 = x, arg3 = y )

myFunct ( x, arg3 = y )
```

Packages & Functions

The base packages comes installed by default and is loaded every time you start R. Additional packages that you have installed by default on start-up:

```
getOption("defaultPackages")
```

Additional packages can be installed using:

```
install.packages(c("pkg1", "pkg2"))
```

Packages & Functions

To ensure that all the packages on your system are up to date:

```
update.packages()
```

To remove a package:

```
remove.packages(c("pkg1", "pkg2")
```

To load a package:

```
library(pkg1)
```

Exercise: Get some help

- Look up the documentation for the lm function, which fits an OLS regression to a data set.
- Run one of the examples by copying and pasting the code under Examples section in the documentation.

Introducing our Data

In this course, we'll be working with a *subset* of the Million Song Dataset containing metadata and analysis of 10,000 contemporary tracks.

labrosa.ee.columbia.edu/MillionSong

Sample Track Descriptions

What's been done with the Data

- Popularity prediction by Genre using K-Means and SVM
  http://cs229.stanford.edu/proj2011/
  BorgHokkanen-WhatMakesForAHitPopSong.pdf

- Release date prediction by Timbre, Energy, Danceability, etc. http:
  //archive.ics.uci.edu/ml/datasets/YearPredictionMSD

- Identification of cover songs based on matching chord profiles
  http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/
  2013/09/67_Paper.pdf

- "Music hasn't changed much in the last 50 years" http://www.nature.
  com/srep/2012/120726/srep00521/full/srep00521.html

- Exploring for new acts that are on the verge of breaking out
  http://musicmachinery.com/2009/12/09/a-rising-star-or/

REVOLUTION
ANALYTICS

What we'll discover

Take a look at: Exploring for new acts that are on the verge of breaking out
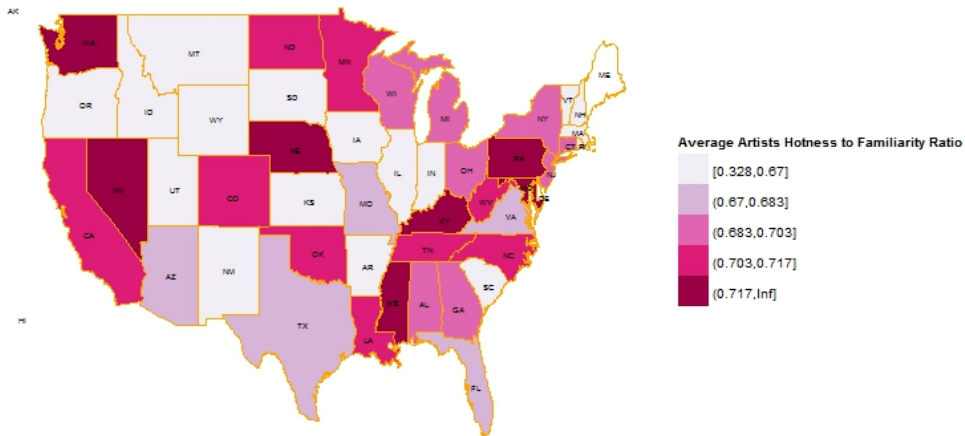`http://musicmachinery.com/2009/12/09/a-rising-star-or/`

Poll: Which state is associated with the highest artists break out scores (ratio of artist hotness to artist familiarity)?

What we'll discover

Think again!

Average Artists Hotness to Familiarity Ratio

- [0.328,0.67]
- (0.67,0.683]
- (0.683,0.703]
- (0.703,0.717]
- (0.717,Inf]

Thank you

Revolution Analytics is the leading commercial provider of software and support for the popular open source R statistics language.

www.revolutionanalytics.com, 1.855.GET.REVO, Twitter: @RevolutionR