

VISUAL ANALYTICS INTRODUCTION TO R TUTORIAL 1

Petra Isenberg

INSTALLATION

- <http://tinyurl.com/VisualAnalytics2015>



In this tutorial you build a basic R web scraper to download and process data. We will build the first part of the scraper together in class, and you will complete the second part on your own.

You should submit the completed assignment to us before 23:00 on Monday.

Getting Started

- Install R [from this website](#) or [from this website \(mirrors\)](#)
- Install RStudio [from its website](#)

The screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar labeled "Go to file/function". The main workspace is divided into several panes:

- Console:** Shows the R startup message: "R version 3.1.1 (2014-07-10) -- 'Sock it to Me' Copyright (C) 2014 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit)". It also displays the R license notice and instructions for using R. The console shows the workspace loaded from ~/.RData and the loading of required packages: RCurl and bitops. The prompt is currently at > |.
- Environment:** Shows the Global Environment, which is currently empty.
- Plots:** Empty.
- Packages:** Shows the installed packages, including RCurl and bitops.
- Help:** Shows the documentation for the R function `html_text`, which is used to extract attributes, text, and tag names from HTML. The description is: "Extract attributes, text and tag name from html."

A red arrow points from the bottom of the console to the `html_text` documentation in the Help pane.

R is an interpreted language. Type code here and have it executed

The screenshot displays the RStudio interface with three main panels. The left panel is the console, showing the R version (3.0.0) and the execution of several R commands. The top-right panel is the workspace, which lists two active objects, 'A' and 'B', both being 4x2 double matrices. The bottom-right panel is the files tab, showing a file explorer view of the current workspace directory, listing a file named '.Rhistory'.

Console Output:

```
R version 3.0.0 (2013-04-03) -- "Masked Marvel"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> getwd()
[1] "H:/MyData/RFiles"
> 5*5
[1] 25
> A <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2)
> A
  [,1] [,2]
[1,]  1   5
[2,]  2   6
[3,]  3   7
[4,]  4   8
> B <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2, byrow=TRUE)
> B
  [,1] [,2]
[1,]  1   2
[2,]  3   4
[3,]  5   6
[4,]  7   8
>
```

Workspace:

| Data | Size |
|------|-------------------|
| A | 4x2 double matrix |
| B | 4x2 double matrix |

Files:

| Name | Size | Modified |
|-----------|----------|-----------------------|
| .. | | |
| .Rhistory | 34 bytes | Aug 23, 2013, 1:26 PM |

The **workspace** tab shows all the active objects (see next slide). The **history** tab shows a list of commands used so far.

The **files** tab shows all the files and folders in your default workspace as if you were on a PC/Mac window. The **plots** tab will show all your graphs. The **packages** tab will list a series of packages or add-ons needed to run certain processes. For additional info see the **help** tab

The **console** is where you can type commands and see output

HELLO WORLD

- Type into your console

```
> print("Hello world!")
```

output:

```
[1] "Hello world!"
```

QUICK R TUTORIALS

Let's get you to work:

```
> install.packages("swirl")  
  
> library(swirl)  
> install_from_swirl("R Programming")  
> swirl()
```

Choose "R Programming"

If you are new to R complete the following lessons:

1, 2, 4, 7

If you are already a proficient R user pick a lesson that interests you

- | when you are at the R prompt (>): |
- Typing skip() allows you to skip the current question. |
- Typing play() lets you experiment with R on your own; swirl | will ignore what you do... |
- UNTIL you type nxt() which will regain swirl's attention. |
- Typing bye() causes swirl to exit. Your progress will be | saved. |
- Typing main() returns you to swirl's main menu. |
- Typing info() displays these options again.

DATA ANALYSIS

Challenge

MAYHEM AT DINO FUN WORLD

- DinoFun World is a typical modest-sized amusement park
- One event last year was a weekend tribute to Scott Jones, internationally renowned football star.

MAYHEM AT DINO FUN WORLD

- Scott was scheduled to appear in two stage shows each on Friday, Saturday, and Sunday to talk about his life and career.
- In addition, a show of memorabilia related to his illustrious career would be displayed in the park's Pavilion.
- However, the event did not go as planned. Scott's weekend was marred by crime and mayhem perpetrated by a poor, misguided and disgruntled figure from Scott's past.

MAYHEM AT DINO FUN WORLD

While the crimes were rapidly solved, park officials and law enforcement figures are interested in understanding just what happened during that weekend to better prepare themselves for future events. They are interested in understanding how people move and communicate in the park, as well as how patterns change and evolve over time, and what can be understood about motivations for changing patterns.

CHALLENGE DATA

- you will be given movement and communication data
- All visitors to the park (except for very young children) use a park app to check in to the park and rides and to communicate with fellow visitors. If visitors do not have compatible phones, they are provided with loaner devices. Visitors are assigned IDs and must use the app to check into rides and some other attractions.
- The park is equipped with sensor beacons that record movements within the park. Sensors are sensitive within a 5m x 5m grid cell. All pathways in the park are covered by these sensors, as are the ride check in locations.
- Locations are not recorded while people are on rides or inside attractions (including restaurants, stores, and rest rooms).
- App users may send text messages to anyone within their own designated group (for example, a family could have their own group).
- An app user may also make "a friend" at the park where they can send and receive texts, if both persons accept friend invitations.

CHALLENGE

- Can you identify movement and communication patterns in the park?
- Can you identify suspicious/unusual behavior?
- Specific questions will be supplied as we delve deeper into the data