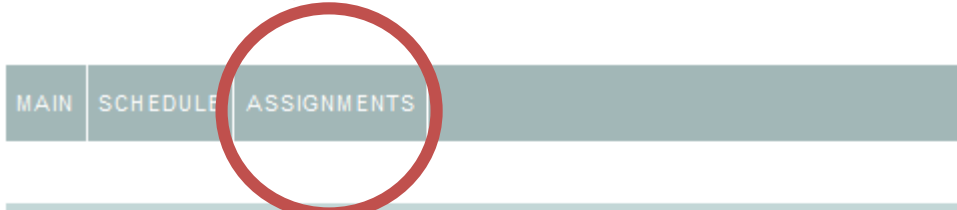


VISUAL ANALYTICS INTRODUCTION TO R TUTORIAL 1

Petra Isenberg

INSTALLATION

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



The screenshot shows a navigation bar with three items: 'MAIN', 'SCHEDULE', and 'ASSIGNMENTS'. The 'ASSIGNMENTS' item is highlighted with a red circle. Below the navigation bar is a section titled 'Tutorial 1 -- Building a Webscraper'. The text below the title describes the tutorial's goal and submission deadline. At the bottom, there is a 'Getting Started' section with two bullet points linking to R installation resources.

MAIN | SCHEDULE | **ASSIGNMENTS**

Tutorial 1 -- Building a Webscraper

In this tutorial you build a basic R web scraper to download and process data. We will build 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 window is divided into four panes. The left pane is the 'Console', which shows the R startup message: 'R version 3.1.1 (2014-07-10) -- "Sock it to Me"', copyright information, platform details, and instructions on how to use R. It also shows the workspace loaded from '~/.RData' and the loading of required packages 'RCurl' and 'bitops'. The right pane is the 'Environment' pane, which is currently empty. Below the Environment pane is the 'Files' pane, which shows the file 'R: Extract attributes, text and tag name from html.' and its documentation. A red arrow points from the bottom of the console to the prompt '> |'. An orange banner at the bottom of the window contains the text: 'R is an interpreted language. Type code here and have it executed'.

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

Console ~ /

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)

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.

[workspace loaded from ~/.RData]

Loading required package: RCurl
Loading required package: bitops
> |

Environment History

Import Dataset Clear

Global Environment

Environment is empty

Files Plots Packages Help Viewer

R: Extract attributes, text and tag name from html. Find in Topic

html_text {rvest} R Documentation

Extract attributes, text and tag name from html.

Description

Extract attributes, text and tag name from html.

html_text(x, ...)

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

The screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Project, Build, Tools, and Help. Below the menu is a toolbar with icons for running code, saving, and navigating. The main interface is divided into four panes: Console, Workspace, History, and Files. The Console pane on the left shows the R startup message and user commands. The Workspace pane on the top right lists active objects. The History pane on the top right shows a list of commands. The Files pane on the bottom right shows the file explorer for the current workspace.

Console: H:/MyData/RFiles/

```
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

Object	Type
A	4x2 double matrix
B	4x2 double matrix

History: (Empty)

Files: H: > MyData > RFiles

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

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

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

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 DINOFUN 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.

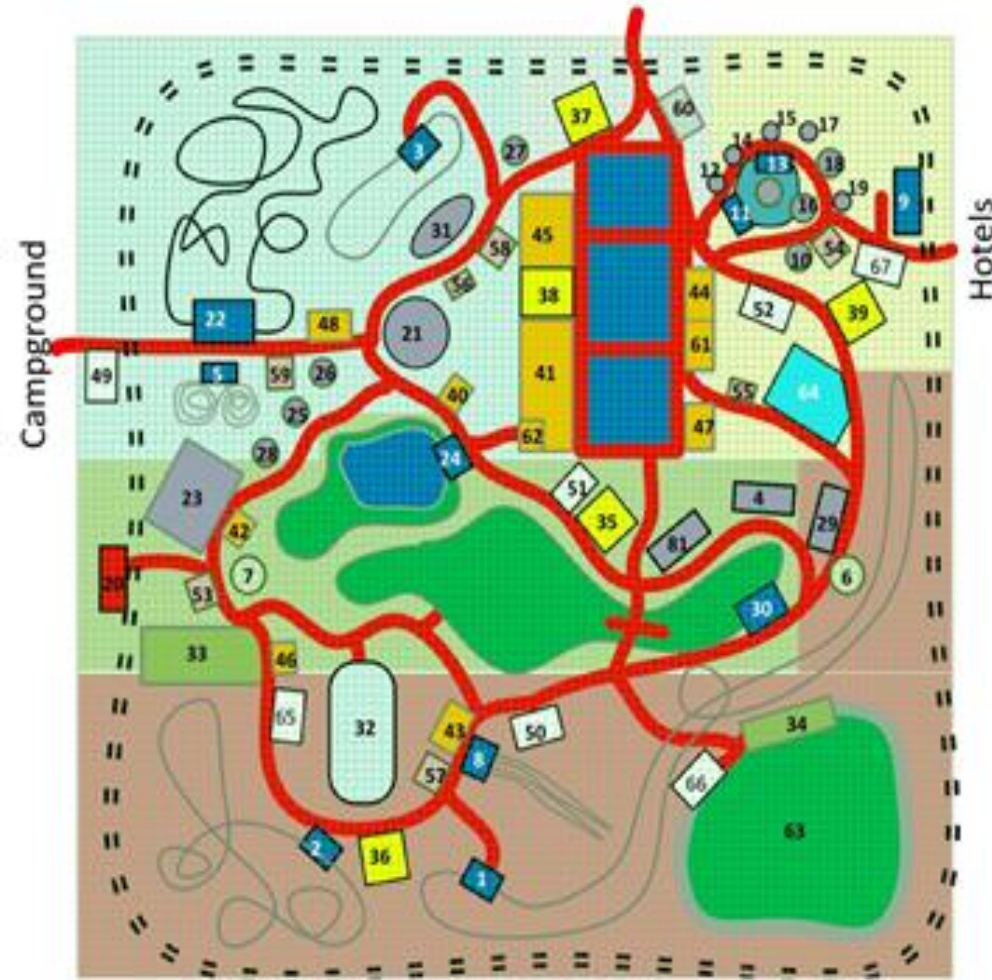
MAYHEM AT DINO FUN WORLD

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.

Park is 100 x 5 meters = 500 meters wide = 1640 feet
Each square is 5 meters on a side.



- The attractions are numbered and coded according to type (the index to the attractions can be found in the park website).
- The red line indicates the pathway through the park, although dark green areas are also areas where people can move.
- (Attraction 30 in the middle of the map is a water rapids ride, so people can watch from the "inside" of the ride boundaries.
- For other rides, people are not allowed to wander inside of the ride footprint.
- Attraction 63 is a show stage area, so people populate this area during performances).

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