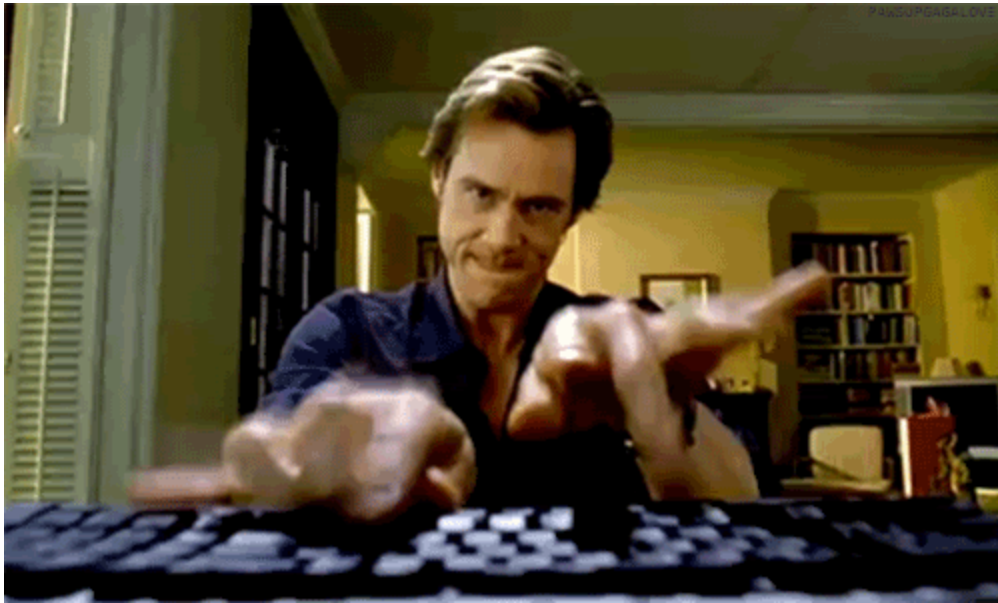


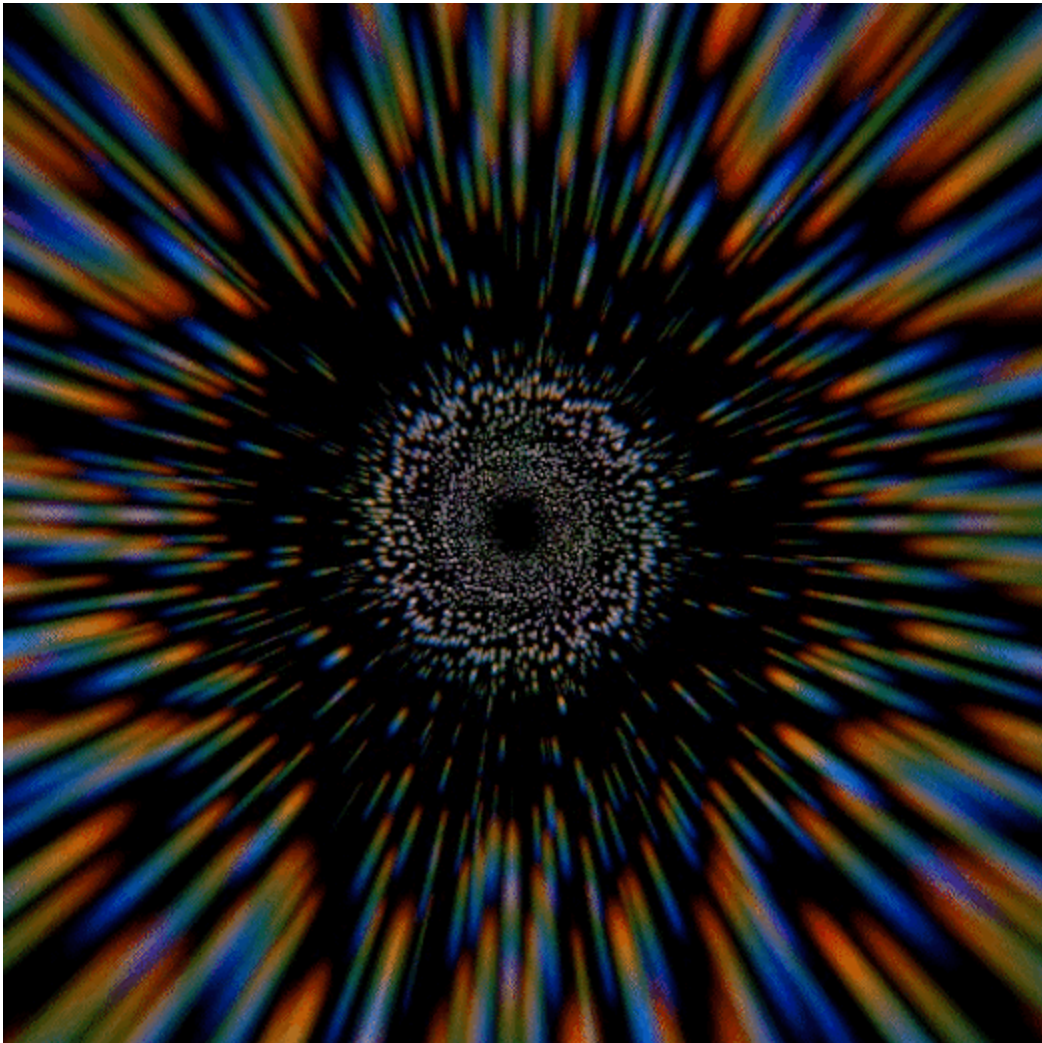
Getting data from the web - R packages

Andrew MacDonald

2014-11-23







Where is data on the internet?

Four ways to get web data

There are many ways to obtain data from the Internet; let's consider four categories:

- *click-and-download* on the internet as a "flat" file, such as .csv, .xls
- *install-and-play* published in a repository which has an API , which has been wrapped
- *API-query* published with an unwrapped API
- *Scraping* implicit in an html website

World values survey

7/16

Online data repositories

- [Figshare](#)
- [Dryad](#)
- [KNB](#)
- [Pangaea](#)

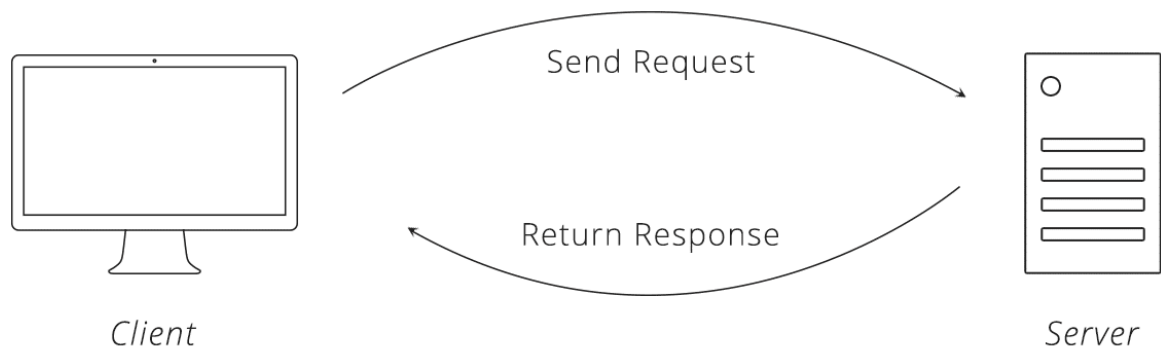
APIs

Application Programming Interfaces

The API specifies how software components should interact.

-wikipedia

There are **Thousands of APIs**



implicit within a website

- in tables [national sports](#)
- all over [IMDB](#) (hunger games example) [Star Trek Wiki](#)

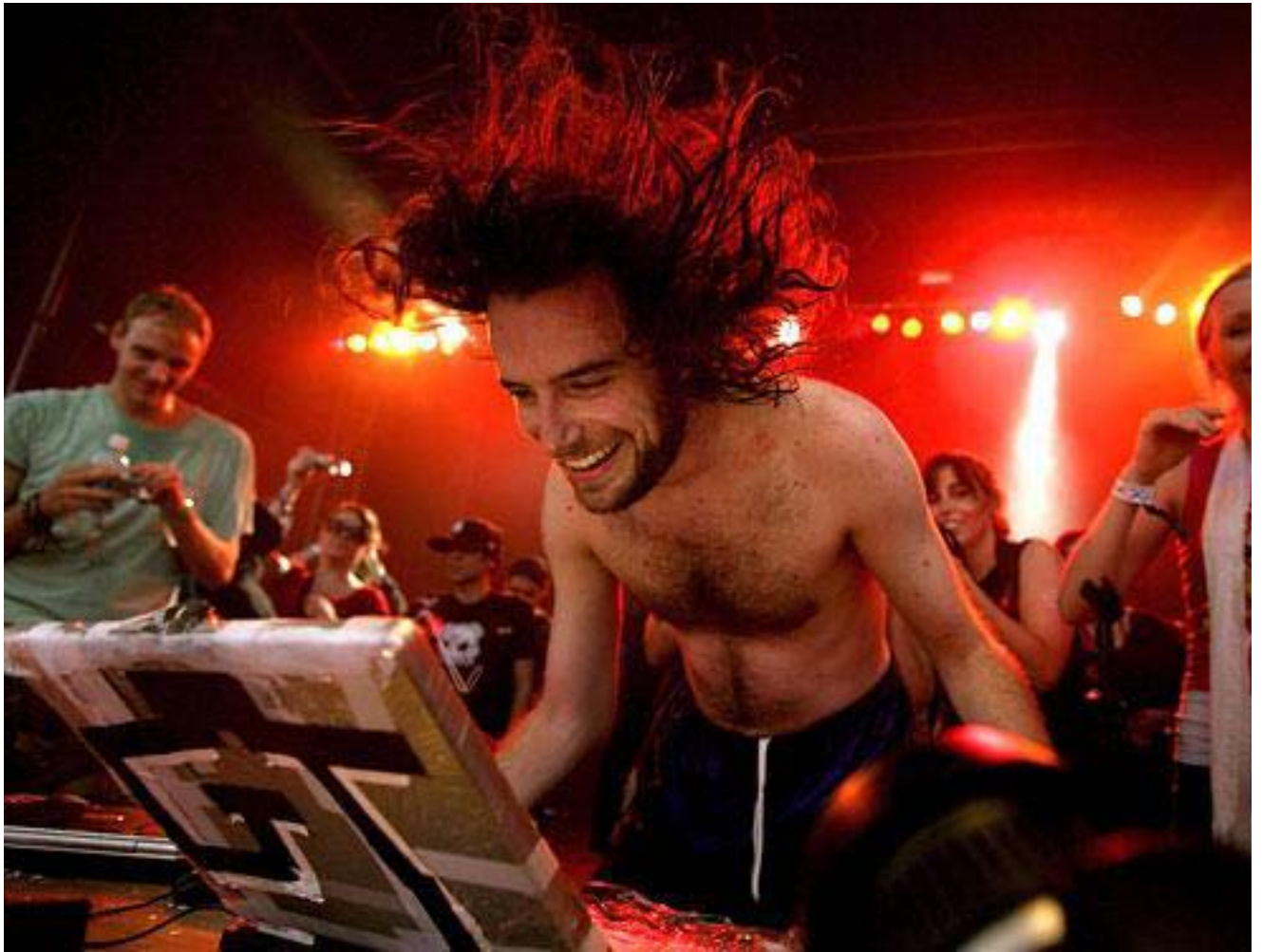
Today's goals

- explore some handy [ROpenSci](#) packages
- set up/learn basic authentication
- Take a look at API requests

Wait, but why?

Why do we want to do this?

- provenance
- reproducible
- updating
- ease
- scaling



ACTIVITY

Build an API!

Open Movie Database API

gameday !!

```
gday <- function(team="canucks") {  
  url <- paste0("http://live.nhle.com/GameData/GCScoreboard/",  
               Sys.Date(),  
               ".jsonp")  
  grepl(team, RCurl::getURL(url), ignore.case=TRUE)  
}  
  
library(httr)  
req <- GET("http://live.nhle.com/GameData/GCScoreboard/2014-11-24.jsonp")  
jsonp <- content(req, "text")  
json <- gsub('([a-zA-Z_0-9\\.]*\\(|\\)|;?$)', "", jsonp, perl = TRUE)  
data <- fromJSON(json)  
  
data$games %>%  
  kable
```