

# Getting Data from the Internet

---



Getting Data



HTTP



httr

R package



HTTP  
Hypertest Transfer  
Protocol

Allows messages to be sent on the Internet!



httr  
R package

Enables YOU to work with these data in R!

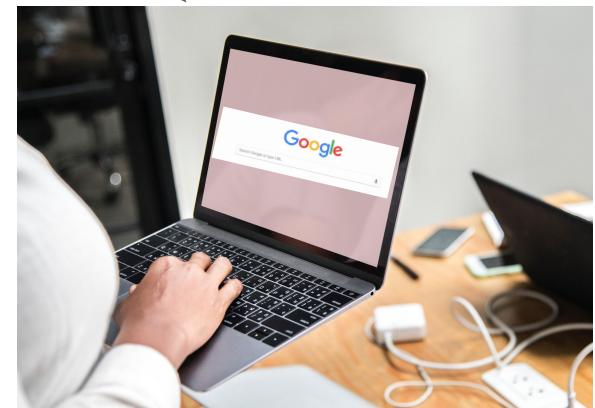


2. request goes  
to the website



Photo by [rawpixel](#) on [Unsplash](#)

1. type URL



3. website is  
displayed

Photo by [rawpixel](#) on [Unsplash](#)

## 2. API endpoint *interprets request*

GET(url = URL)

## 1. type API request

## 3. Get API response

```
> api_response
Response [https://raw.githubusercontent.com/fivethirtyeight/data/master/steak-survey/steak-risk-survey.csv]
Date: 2018-07-06 18:09
Status: 200
Content-Type: text/plain; charset=utf-8
Size: 62.7 kB
RespondentID,"Consider the ...
,Response,Response,Response...
3237565956,Lottery B,,,...,...
3234982343,Lottery A,No,Yes...
3234973379,Lottery A,No,Yes...
3234972383,Lottery B,Yes,Ye...
3234958833,Lottery B,No,Yes...
3234955240,Lottery A,No,No,...
3234955097,Lottery A,No,Yes...
3234955010,Lottery A,No,Yes...
...
```

## REST API v3

[Reference](#) [Guides](#) [Libraries](#)

# Overview

This describes the resources that make up the official GitHub REST API v3. If you have any problems or requests please contact [GitHub support](#).

- i. [Current version](#)
- ii. [Schema](#)
- iii. [Authentication](#)
- iv. [Parameters](#)
- v. [Root endpoint](#)
- vi. [GraphQL global node IDs](#)
- vii. [Client errors](#)
- viii. [HTTP redirects](#)
- ix. [HTTP verbs](#)
- x. [Hypermedia](#)
- xi. [Pagination](#)
- xii. [Rate limiting](#)
- xiii. [User agent required](#)
- xiv. [Conditional requests](#)

## ▼ Overview

[Media Types](#)[OAuth Authorizations API](#)[Other Authentication Methods](#)[Troubleshooting](#)[Pre-release Program](#)[API Previews](#)[Versions](#)[▶ Activity](#)[▶ Checks](#)[▶ Gists](#)[▶ Git Data](#)[▶ GitHub Apps](#)[▶ Issues](#)

```
#> ## load package
library(httr)
library(dplyr)

## Save GitHub username as variable
username <- 'janeeverydaydoe'                                API endpoint

## Save base endpoint as variable
url_git <- 'https://api.github.com/'                           API request

## Construct API request
api_response <- GET(url = paste0(url_git,
'users/', username, '/repos'))
```



```
# See variables in response  
names(api_response)
```

```
> names(api_response)  
[1] "url"          "status_code"   "headers"      "all_headers"  "cookies"      "content"  
[7] "date"         "times"        "request"     "handle"
```

What API request from  
httr returns

```
## Check Status Code of request  
api_response$status_code  
  
## Extract content from API response  
repo_content <- content(api_response)
```

```
> api_response$status_code  
[1] 200
```

'200' means request was successful!

```
> repo_content <- content(api_response)
```

content () extracts contents from API request

```
## function to get name and URL for each repo
lapply(repo_content, function(x) {
  df <- data_frame(repo = x$name,
                    address = x$html_url)) %>%
bind_rows()
```

```
> lapply(repo_content, function(x) {
+   df <- data_frame(repo = x$name,
+                     address = x$html_url)}) %>%
+   bind_rows()
# A tibble: 6 x 2
  repo                  address
  <chr>                <chr>
1 first_project        https://github.com/JaneEverydayDoe/first_project
2 hello-world           https://github.com/JaneEverydayDoe/hello-world
3 janeeverydaydoe.github.com https://github.com/JaneEverydayDoe/janeeverydaydoe.g...
4 my_first_project      https://github.com/JaneEverydayDoe/my_first_project
5 newproject             https://github.com/JaneEverydayDoe/newproject
6 Temporary_add_to_version_control https://github.com/JaneEverydayDoe/Temporary_add_to_...
```



# Our Data

We're sharing the data and code behind some of our articles and graphics.  
We hope you'll use it to check our work and to create stories and  
visualizations of your own.

• UPDATING

## DATA SET

## RELATED CONTENT

mlb-elo

**2018 MLB Predictions**

11 hours ago •

[info](#)

soccer-spi

**Club Soccer Predictions**

12 hours ago •

[info](#)

congress-generic-ballot

**Are Democrats Or Republicans Winning The Race For Congress?**

16 hours ago •

[info](#)

trump-approval-ratings

**How Popular Is Donald Trump?**

a day ago •

[info](#)

Code

Issues 5

Pull requests 4

Wiki

Insights

Branch: master ▾

data / steak-survey /

Create new file

Upload files

Find file

History



dmil update READMEs

Latest commit 9e6c424 on Feb 9

..



README.md

update READMEs

5 months ago



steak-risk-survey.csv

Rename steak-risk-survey to steak-risk-survey.csv

4 years ago



README.md

## Steak Survey

This folder contains data behind the stories:

- [How Americans Like Their Steak.](#)
- [How Americans Order Their Steak](#)

```
> ## Make API request  
> api_response <- GET(url = "https://raw.githubusercontent.com/fivethirtyeight/data/master/steak-survey/steak-risk-survey.csv")
```

```
>  
> ## Extract content from API response  
> df_steak <- content(api_response, type="text/csv")
```

No encoding supplied: defaulting to UTF-8.

Parsed with column specification:

```
cols(
```

```
  RespondentID = col_double(),
```

`Consider the following hypothetical situations: <br>In Lottery A, you have a 50% chance of success, with a payout of \$100. <br>In Lottery B, you have a 90% chance of success, with a pay out of \$20. <br><br>Assuming you have \$10 to bet, would you play Lottery A or Lottery B?` = col\_character(),

```
  `Do you ever smoke cigarettes?` = col_character(),
```

```
  `Do you ever drink alcohol?` = col_character(),
```

```
  `Do you ever gamble?` = col_character(),
```

```
  `Have you ever been skydiving?` = col_character(),
```

```
  `Do you ever drive above the speed limit?` = col_character(),
```

```
  `Have you ever cheated on your significant other?` = col_character(),
```

```
  `Do you eat steak?` = col_character(),
```

```
  `How do you like your steak prepared?` = col_character(),
```

```
  Gender = col_character(),
```

```
  Age = col_character(),
```

```
  `Household Income` = col_character(),
```

```
  Education = col_character(),
```

```
  `Location (Census Region)` = col_character()
```

```
)
```

GET () makes API request

content () extracts information

```
myapp = oauth_app("twitter",
                  key = "yourConsumerKeyHere",
                  secret = "yourConsumerSecretHere")
sig = sign_oauth1.0(myapp,
                     token = "yourTokenHere",
                     token_secret = "yourTokenSecretHere")
homeTL =
GET("https://api.twitter.com/1.1/statuses/home_timeline.json", sig)
```







<https://github.com/hadley/rvest>

# Chromebook Data Science Curriculum

There are currently 12 courses that are offered in the Chromebook Data Science Curriculum.

Class	Course Description	Leanpub Link
<b>Introduction to Chromebook Data Science</b>	This is the first class in the Chromebook Data Science series. Data science is one of the most exciting and fastest growing careers in the world. The goal of this series is to help people with no background and limited resources transition into data science. The only pre-requisites are a computer with a web browser and the ability to type and follow instructions. We guide you through the rest.	<a href="#">Course 0</a>
<b>How to Use A Chromebook</b>	This course will introduce you to using a Chromebook. The Introduction and Setup course might sound simple, but it will set up the infrastructure for success with the later, more challenging courses.	<a href="#">Course 1</a>
<b>Google and the Cloud</b>	The Google and the Cloud course introduces using Google's in-built apps, which form the fundamental backbone of a Chromebook. We'll go step by step through the process to integrating these apps together to form your productivity workflow.	<a href="#">Course 2</a>
<b>Organizing Data Science</b>	Projects are central to the role of any data scientist. These lessons will	<a href="#">Course 3</a>

Different parts of the webpage have different HTML tags

tags

tags



# SelectorGadget

offered by [selectorgadget.com](http://selectorgadget.com)

★★★★★ (68)

[Developer Tools](#)

78,479 users

+ ADD TO CHROME



OVERVIEW

REVIEWS

SUPPORT

RELATED

Hacker News | New | threads | comments | ask | jobs | submit

1. John Carmack Joins Oculus as CEO ([oculusrvr.com](http://oculusrvr.com))
2. + Android ADSP maintainer quits ([plus.google.com](http://plus.google.com))  
73 points by [w00t00](#) 1 hour ago | [Flag](#) | 44 comments
3. + Steps To \$5,000 In Monthly Recurring Revenue ([statuspage.io](http://statuspage.io))  
101 points by [kennethreitz](#) 1 hour ago | [Flag](#) | 28 comments
4. + Slashdot founder Rob Malda on why there won't be another Hacker News ([washingtonpost.com](http://washingtonpost.com))  
134 points by [liberate00](#) 2 hours ago | [Flag](#) | 1330 comments
5. + Thoughts on Twitter's new Two-Factor Authentication ([authy.com](http://authy.com))  
48 points by [craigj](#) 3 hours ago | [Flag](#) | 114 comments
6. + Staring at the Sun: Dalvik vs. Asm.js vs. Native ([mozilla.org](http://mozilla.org))  
74 points by [evelin](#) 3 hours ago | [Flag](#) | 16 comments
7. MakeGamesWithUs (YC W12) Needs a Graphic/Game Designer  
8 minutes ago
8. + DigitalOcean raises \$3.2M Seed Round ([techcrunch.com](http://techcrunch.com))  
138 points by [bogdanski](#) 5 hours ago | [Flag](#) | 172 comments
9. + EFF: "Parallel construction" is really intelligence laundering ([eff.org](http://eff.org))  
294 points by [r00t3r](#) 7 hours ago | [Flag](#) | 135 comments
10. + Programming languages to watch: LiveScript, Julie, Elixir ([adambard.com](http://adambard.com))  
162 points by [probz00](#) 8 hours ago | [Flag](#) | 73 comments
11. + Envelopes - Python email for humans ([lasmekwojcie.github.io](http://lasmekwojcie.github.io))  
34 points by [PekkaPekka](#) 2 hours ago | [Flag](#) | 3 comments
12. + App.net Response to Brennan Novak, part II ([app.net](http://app.net))  
18 points by [aaronpk](#) 1 hour ago | [Flag](#) | 12 comments
13. + Pixel-perfect timing attacks with HTML5 ([contextis.co.uk](http://contextis.co.uk))  
118 points by [bentov](#) 6 hours ago | [Flag](#) | 133 comments
14. + Restoring Trust in Government and the Internet ([schneier.com](http://schneier.com))  
148 points by [hatchen](#) 7 hours ago | [Flag](#) | 72 comments
15. + JS-Git ([github.com](http://github.com))  
93 points by [eliasj](#) 8 hours ago | [Flag](#) | 111 comments
16. + Julia: IPython Notebook for Julia ([github.com](http://github.com))  
53 points by [funkier](#) 8 hours ago | [Flag](#) | 5 comments
17. + Reactive.coffee: reactive programming and declarative UIs in CoffeeScript ([mit.edu](http://mit.edu))  
47 points by [jacksonk](#) 8 hours ago | [Flag](#) | 14 comments
18. + Keep Your Identity Small (2009) ([nugraham.com](http://nugraham.com))  
18 points by [CaleWright](#) 8 hours ago | [Flag](#) | 16 comments
19. + Streak's New iPhone App Is A CRM Service With Gmail Baked In (YC S11) ([techcrunch.com](http://techcrunch.com))  
8 points by [mikematt](#) 10 minutes ago | [Flag](#) | 14 comments
20. + Stealth Startups Are Stupid ([alexstechthoughts.com](http://alexstechthoughts.com))  
65 points by [AlasdairK](#) 3 hours ago | [Flag](#) | 73 comments
21. + OpenMailBox: An alternative to Gmail based on free software ([perseosblog.com](http://perseosblog.com))  
7 points by [simeonstevens](#) 54 minutes ago | [Flag](#) | 1 comment
22. + This Startup's Cheap Sensors Could Create an OS for Everyday Life ([wired.com](http://wired.com))  
33 points by [jordan](#) 4 hours ago | [Flag](#) | 136 comments
23. + Adding Interval Sets to Redis ([redis.io](http://redis.io))  
14 points by [julianmarcus](#) 4 hour ago | [Flag](#) | 10 comments

.title~ .title a

Clear (30)

Toggle Position

XPath

?

X



Compatible with your device

Easy, powerful CSS Selector generation.

Selector Gadget is an open source Chrome Extension that makes CSS selector generation and discovery on complicated sites a breeze.

After having installed the extension, go to any page and launch it. A box will open in the bottom right of the website. Click on a page element that you would like your selector to match (it will turn green). SelectorGadget will then generate a minimal CSS selector for that element, and will highlight *and* *all* *ancestor* *elements* that are matched.

[Website](#)

[Report Abuse](#)

Additional Information

Version: 1.1

Updated: March 20, 2015

Size: 79.7KIB

Language: English



## Add "SelectorGadget"?



It can:

Read and change all your data on the websites you visit

Cancel

Add extension





X



SelectorGadget has been added to  
Chrome

Use this extension by clicking on this icon.

Manage your extensions by clicking Extensions in  
the Window menu.



# Chromebook Data Science Curriculum

! [SelectorGadget icon] (images/03\_internet\_03\_data\_internet-20.png) There are currently 12 courses that are offered in the Chromebook Data Science Curriculum.

Class	Course Description	Leanpub Link
<b>Introduction to Chromebook Data Science</b>	This is the first class in the Chromebook Data Science series. Data science is one of the most exciting and fastest growing careers in the world. The goal of this series is to help people with no background and limited resources transition into data science. The only pre-requisites are a computer with a web browser and the ability to type and follow instructions. We guide you through the rest.	<a href="#">Course 0</a>
<b>How to Use A Chromebook</b>	This course will introduce you to using a Chromebook. The Introduction and Setup course might sound simple, but it will set up the infrastructure for success with the later, more challenging courses.	<a href="#">Course 1</a>
<b>Google and the Cloud</b>	The Google and the Cloud course introduces using Google's in-built apps, which form the fundamental backbone of a Chromebook. We'll go step by step through the process to integrating these apps together to form your productivity workflow.	<a href="#">Course 2</a>
<b>Organizing Data Science Projects</b>	Projects are central to the role of any data scientist. These lessons will discuss how to organize projects and how to use them effectively.	<a href="#">Course 3</a>

No valid path found.

Clear

Toggle Position

XPath

?

X

# Chromebook Data Science Curriculum

There are currently 12 courses that are offered in the Chromebook Data Science Curriculum.

Class	Course Description	Leanpub Link
<b>Introduction to Chromebook</b> <b>Data Science</b> td strong	This is the first class in the Chromebook Data Science series. Data science is one of the most exciting and fastest growing careers in the world. The goal of this series is to help people with no background and limited resources transition into data science. The only pre-requisites are a computer with a web browser and the ability to type and follow instructions. We guide you through the rest.	<a href="#">Course 0</a>
<b>How to Use A Chromebook</b>	This course will introduce you to using a Chromebook. The Introduction and Setup course might sound simple, but it will set up the infrastructure for success with the later, more challenging courses.	<a href="#">Course 1</a>
<b>Google and the Cloud</b>	The Google and the Cloud course introduces using Google's in-built apps, which form the fundamental backbone of a Chromebook. We'll go step by step through the process to integrating these apps together to form your productivity workflow.	<a href="#">Course 2</a>
<b>Organizing Data Science</b> <b>Projects</b>	Projects are central to the role of any data scientist. These lessons will discuss how to strong	<a href="#">Course 3</a> Clear (12) Toggle Position XPath ? X

Text you'll use to specify which part of the webpage you'd like to scrape

```
## load package  
library(rvest)  
  
## provide URL  
courses <-  
read_html("http://jhudatascience.org/chromebook  
datascience/curriculum.html")  
  
## Get Courses  
courses %>%  
  html_nodes("strong") %>%  
  html_text()
```

Text from  
SelectorGadget

```
> courses %>%  
+     html_nodes("strong") %>%  
+     html_text()  
[1] "Introduction to Chromebook Data Science"  
[2] "How to Use A Chromebook"  
[3] "Google and the Cloud"  
[4] "Organizing Data Science Projects"  
[5] "Version Control"  
[6] "Introduction to R"  
[7] "Data: Tidying"  
[8] "Data Visualization"  
[9] "Getting Data"  
[10] "Basics of Data Analysis"  
[11] "Written and Oral Communication in DS"  
[12] "Getting a job in data science"
```



# SelectorGadget: point and click CSS selectors



The screenshot shows a video player interface for a 'SelectorGadget Screencast' from Andrew Cantino. The video is 1:36 long and is currently at the beginning. The title bar says 'SelectorGadget Screencast from Andrew Cantino'. The video content displays a list of news items from Hacker News, with the 17th item highlighted in yellow. The 17th item is: 'Stanford grad's site nets Southwest 'cease and desist'' (paloaltoonline.com). The video player includes standard controls like play/pause, volume, and a progress bar. Below the video, the URL 'news.ycombinator.com/item?id=4689308' and the source 'E-BOOKS and Others (wired.com)' are visible.

1. ▾ AnandTech, Microsoft Surface Review (anandtech.com)  
77 points by barista 2 hours ago | 37 comments  
2. ▾ Wired's Review of this Microsoft Surface (wired.com)  
42 points by colinleehordon 2 hours ago | 16 comments  
3. ▾ Zynga May Have Just Laid Off 100+ Employees From Its Austin Office (techcrunch.com)  
384 points by haribabu 10 hours ago | 1 comment  
4. ▾ The Hardware Renaissance (techcrunch.com)  
366 points by niquemil 11 hours ago | 171 comments  
5. ▾ Don't Call The New Microsoft Surface RT A Tablet, This Is A PC (techcrunch.com)  
23 points by vyrotek 2 hours ago | 16 comments  
6. ▾ Why we buy into ideas: how to convince others of our thoughts (bufferapp.com)  
6 points by sunik34 23 minutes ago | 1 discussion  
7. ▾ The rise of the "successful" unsustainable company (asmartbear.com)  
281 points by yanickmabe 12 hours ago | 103 comments  
8. ▾ Under the hood of Windows 8, or why desktop users should upgrade from Windows 7 (extremetech.com)  
281 points by rvoe 9-12 hours ago | 170 comments  
9. ▾ Marc Andreessen's Productivity Trick to Feeling Marvelously Efficient (idonethis.com)  
106 points by mikusen 7 hours ago | 34 comments  
10. ▾ Show HN: Taurus.io - Create a product tour for your web app in 15 minutes (taurus.io)  
31 points by etzlo 3 hours ago | 30 comments  
11. ▾ The PC isn't dead (dondery.net)  
9 points by dendry 1 hour ago | 6 comments  
12. ▾ Ceefax Final Broadcast: 'Goodbye, cruel world.' (h4ck.in)  
76 points by leumasr 7 hours ago | 124 comments  
13. ▾ Show HN: Fact check last night's Presidential debate with Quip (quipvideo.com)  
32 points by dmedvedman 4 hours ago | 12 comments  
14. ▾ Increasing wireless network speed by 1000%, by replacing packets with algebra (extremetech.com)  
98 points by oliver 7 hours ago | 30 comments  
15. ▾ Amazon reopens wiped Kindle account (translate.google.com)  
258 points by EwanT0n 15 hours ago | 137 comments  
16. ▾ Zynga CEO Mark Pincus Confirms Layoffs: 5% of Workforce (techcrunch.com)  
47 points by nikunjk 6 hours ago | 11 comments  
17. ▾ Stanford grad's site nets Southwest 'cease and desist' (paloaltoonline.com)  
21 points by c533 4 hours ago | 18 comments  
18. ▾ OrderAhead is hiring a Marketing Associate  
2 hours ago  
19. ▾ New theory may explain the notorious cold fusion experiment from two decades ago (discovermagazine.com)  
106 points by supregek 10 hours ago | 25 comments  
20. ▾ Digging Into Apple's Fusion Drive Details (macobserver.com)  
16 points by barista 2 hours ago | 17 comments

[SelectorGadget Screencast](#) from Andrew Cantino on [Vimeo](#).

## JSON: key-value pairs

---

```
{"Name": "Isabela"}
```

key

value

```
library(jsonlite)

## generate a JSON object
json <-
'[
  {"Name" : "Woody", "Age" : 40, "Occupation" :
"Sherriff"},  

  {"Name" : "Buzz Lightyear", "Age" : 34,  

"Occupation" : "Space Ranger"},  

  {"Name" : "Andy", "Occupation" : "Toy Owner"}  

]'

## take a look
json
```

> json

```
[{"Name": "Woody", "Age": 40, "Occupation": "Sherriff"}, {"Name": "Buzz Lightyear", "Age": 34, "Occupation": "Space Ranger"}, {"Name": "Andy", "Occupation": "Toy Owner"}]
```

```
## take JSON object and convert to a  
data frame  
mydf <- fromJSON(json)
```

```
## take a look
```

```
mydf > mydf
```

	Name	Age	Occupation
1	Woody	40	Sherriff
2	Buzz Lightyear	34	Space Ranger
3	Andy	NA	Toy Owner

```
## take JSON object and convert to a  
data frame  
json <- toJSON(mydf)
```

```
> json  
[{"Name": "Woody", "Age": 40, "Occupation": "Sherriff"}, {"Name": "Buzz Lightyear", "Age": 34, "Occupati  
on": "Space Ranger"}, {"Name": "Andy", "Occupation": "Toy Owner"}]
```



Photo by [rawpixel](#) on [Unsplash](#)