Week 2 Notes - Getting and Cleaning Data

Reading from MySQL

Data in SQL are structured into databases -> databases consist of tables with fields -> tables contain entries as rows. The tables themselves often represent specific aspects of the data which are interlinked within the database - say a table for the salaries of employees, another of the annnual leave, another table for their personal details and so on.

Let's install MySQL

In R `install.packages("RMySQL")`

In julia

```
In [1]: using Pkg; Pkg.add("MySQL") ; using MySQL

         Resolving package versions...
         No Changes to `~/.julia/environments/v1.10/Project.toml`
         No Changes to `~/.julia/environments/v1.10/Manifest.toml`

In [2]: Pkg.add("DataFrames") ; using DataFrames

         Resolving package versions...
         No Changes to `~/.julia/environments/v1.10/Project.toml`
         No Changes to `~/.julia/environments/v1.10/Manifest.toml`
```

Connecting to databases - UCSC genome browser example

Connecting to the UCSC MySQL server and pulling the databases on UCSC that are available to us -- let's do it in R. This will establish a connection to the server - a stream, and then using this stream we will execute a MySQL command `show databases;` to retrieve the available databases, and then disconnect from the stream.

```
ucscDB <- dbConnect(MySQL(), user="genome", host="genome-
mysql.cse.ucsc.edu")
result <- dbGetQuery(ucscDB, "show databases;"); dbDisconnect(ucscDB);
```

In Julia we can do this by employing the MySQL.jl package - part of the Databases.jl family;

```
In [3]: # Connecting https://mysql.juliadatabases.org/dev/
        ucscDB = DBInterface.connect(MySQL.Connection, "genome-mysql.soe.ucsc.edu", "genome")

Out[3]: MySQL.Connection(host="genome-mysql.soe.ucsc.edu", user="genome", port="3306", db="")

In [4]: # Query the server and store the query in a dataframe - or a csv etc.
        result = DBInterface.execute(ucscDB, "show databases") |> DataFrame;

In [5]: # Lets view the result - we can see that it lists all the genomes stored on UCSC
        result
```

340×1 DataFrame   *315 rows omitted*

Row	Database
	String
1	acaChl1
2	ailMel1
3	allMis1
4	allSin1
5	amaVit1
6	anaPla1
7	ancCey1
8	angJap1
9	anoCar1
10	anoCar2
11	anoGam1
12	anoGam3
13	apaSpi1
⋮	⋮
329	vicPac1
330	vicPac2
331	visiGene
332	wuhCor1
333	xenLae2
334	xenTro1
335	xenTro10
336	xenTro2
337	xenTro3
338	xenTro7
339	xenTro9
340	zonAlb1

In [6]:
```
"hg38" in result.Database
```

Out[6]: true

In [7]:
```
# Close the connection stream
DBInterface.close!(ucscDB)
```

Now that we've connected to the MySQL server, we will connect to a specific database and perform some queries.

In R - we'll connect, retrieve all of the table associated with the db, and then execute a function to see how many table are stored

```
hg38 <- dbConnect(MySQL(), user="genome", db="hg38", host="genome-mysql.soe.ucsc.edu")
```

```
allTables <- dbListTables(hg38)
length(allTables)
```

Let's get cracking on Julia

In [8]: `hg38 = DBInterface.connect(MySQL.Connection, "genome-mysql.soe.ucsc.edu", "genome", d`

Out[8]: MySQL.Connection(host="genome-mysql.soe.ucsc.edu", user="genome", port="3306", db="hg38")

In [9]: `hg38Tables = DBInterface.execute(hg38, "show tables") |> DataFrame;`

In [10]: `hg38Tables[1:10, :]`

Out[10]: 10×1 DataFrame

Row	Tables_in_hg38
	String
1	affyGnf1h
2	affyU133
3	affyU95
4	all_est
5	all_mrna
6	all_sts_primer
7	all_sts_seq
8	altLocations
9	altSeqLiftOverPsl
10	altSeqLiftOverPslP3

A whopping 2835 different tables ! UCSC is extreeeeemly information rich, as we can see, there are many many many data sources we can pull from.

Now to investigate the specific fields within a specific table

In R; get a table and perform a basic SQL function to show how many fields are in the table

```
dbListFields(hg38, "all_mrna")
dbGetQuery(hg38, "select count(*) from all_mrna")
```

In Julia

In [11]: `countsql = DBInterface.execute(hg38, "select count(*) from all_mrna") |> DataFrame;`

In [12]: `countsql`

Out[12]: 1×1 DataFrame

Row	count(*)
	Int64
1	10489979

Now lets play with the contents of the fields in R

```
mrnaData <- dbReadTable(hg38, "all_mrna")
head(mrnaData)
```

Julia !

In [30]:
```
mrnaData = DBInterface.execute(hg38, "select * from all_mrna;") |> DataFrame
```

10489979×22 DataFrame

Row	bin	matches	misMatches	repMatches	nCount	qNumInsert	qBaseInsert	tNumInsert
	UInt16	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32
1	585	1579	25	0	0	0	0	2
2	585	1419	21	0	0	0	0	2
3	585	1533	12	0	0	0	0	4
4	585	1578	27	0	0	0	0	2

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
5	585	1652	0	0	0	0	0	2
6	585	1650	2	0	0	0	0	2
7	585	1648	4	0	0	0	0	2
8	585	1485	3	0	0	0	0	2

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
9	585	1631	8	0	0	0	0	4
10	585	1736	4	0	0	0	0	3
11	585	402	0	0	0	0	0	2
12	585	1241	7	0	0	0	0	3

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
13	585	23	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10489968	585	284	3	0	0	0	0	0
10489969	586	252	6	0	0	1	10	1
10489970	586	287	0	0	0	0	0	0
10489971	586	285	0	0	0	0	0	0
10489972	586	279	10	0	0	0	0	0
10489973	586	285	1	0	0	0	0	1
10489974	586	284	2	0	0	0	0	0

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
10489975	587	298	4	0	0	0	0	0
10489976	587	272	4	0	0	0	0	0
10489977	588	280	7	0	0	0	0	0
10489978	588	283	3	0	0	0	0	0
10489979	703	2427	0	0	0	0	0	10

In [ ]:

In [29]: `Pkg.add("CSV") ; using CSV`

```
 Resolving package versions...
No Changes to `~/.julia/environments/v1.10/Project.toml`
No Changes to `~/.julia/environments/v1.10/Manifest.toml`
```

Refined queries

MySQL has a wide range of query options which permit the extration of virtually any aspect of the data, with conditionals, ranges, mismatches and so on. These statements must simply be crafted according to the query structure and provided to the julia functions in order to get what you're looking for.

For instance we can extract entries from the table which have values in the 'mismatches' column between 1 and 3;

In R

```
query <- dbSendQuery(hg38, "select * from all_mrna where misMatches between
1 and 3")
mrnas <- fetch(query)
```

In Julia

In [32]:
```
query = DBInterface.execute(hg38, "select * from all_mrna where misMatches between 1
```

```
query <- dbSendQuery(hg38, "select * from all_mrna where misMatches between
1 and 3")
mrnas <- fetch(query)
```

In Julia

In [32]:
```
query = DBInterface.execute(hg38, "select * from all_mrna where misMatches between 1
```

`Out[32]:` 403833×22 DataFrame

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
1	585	1650	2	0	0	0	0	2
2	585	1485	3	0	0	0	0	2
3	585	925	2	0	0	1	3	7
4	585	986	1	0	0	0	0	8

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
5	585	970	3	0	0	0	0	4
6	585	953	2	0	0	0	0	8
7	585	775	1	0	0	0	0	6
8	585	974	2	0	0	0	0	0

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
9	585	1123	1	0	0	0	0	2
10	585	1122	2	0	0	0	0	2
11	586	1381	1	0	0	0	0	0
12	586	1625	2	0	0	0	0	0
13	586	29	1	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
403822	1270	283	3	0	0	0	0	0

Row	bin UInt16	matches UInt32	misMatches UInt32	repMatches UInt32	nCount UInt32	qNumInsert UInt32	qBaseInsert UInt32	tNumInsert UInt32
403823	755	294	2	0	0	0	0	0
403824	755	292	3	0	0	0	0	0
403825	94	321	3	0	0	0	0	2
403826	94	319	3	0	0	0	0	2
403827	585	294	2	0	0	0	0	0
403828	585	292	3	0	0	0	0	0
403829	585	294	3	0	0	0	0	0

Row	bin	matches	misMatches	repMatches	nCount	qNumInsert	qBaseInsert	tNumInsert
	UInt16	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32
403830	585	284	3	0	0	0	0	0
403831	586	285	1	0	0	0	0	1
403832	586	284	2	0	0	0	0	0
403833	588	283	3	0	0	0	0	0

In [34]: `first(query)`

Out[34]: DataFrameRow (22 columns)

Row	bin	matches	misMatches	repMatches	nCount	qNumInsert	qBaseInsert	tNumInsert	tBa
	UInt16	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32	UInt32	UInt
1	585	1650	2	0	0	0	0	2	

In [ ]: `size(query)`

Close the connection!

In [37]: `DBInterface.close!(hg38)`

HDF5 - Reading and Handling this data type

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

HDF5 is a data format used for storing large datasets - FAST5 uses a HDF5 backbone, and we know how large FAST5 files are!!! It stands for **H**eirarchical **D**ata **F**ormat.

As expected, HDF5 support for R comes in a library, this time it is downloaded from bioconductor

```
source("url")
biocLite("rhdf5")
library(rhdf5)
```
We can create an example file using the hdf5 functions `created = h5createFile("example.h5")`

Julia too has a package for working with HDF5 formats, can you guess the original name? **HDF5.jl**. Remarkable indeed. https://juliaio.github.io/HDF5.jl/stable/ - A nice explaner of HDF5 from the julia software page

"HDF5 stands for Hierarchical Data Format v5 and is closely modeled on file systems. In HDF5, a "group" is analogous to a directory, a "dataset" is like a file. HDF5 also uses "attributes" to associate metadata with a particular group or dataset. HDF5 uses ASCII names for these different objects, and objects can be accessed by Unix-like pathnames, e.g., "/sample1/tempsensor/firsttrial" for a top-level group "sample1", a subgroup "tempsensor", and a dataset "firsttrial"."

(Datasets cannot have child datasets, but groups can have either.)

Groups/subgroups/datasets NOT Groups/datasets/datasets/groups

In [39]:
```julia
Pkg.add("HDF5") ; using HDF5
```

```
  Resolving package versions...
  Installed MPIPreferences — v0.1.10
  Installed HDF5_jll ———————— v1.12.2+2
  Installed Requires ———————— v1.3.0
  Installed HDF5 ———————————— v0.17.1
   Updating `~/.julia/environments/v1.10/Project.toml`
  [f67ccb44] + HDF5 v0.17.1
   Updating `~/.julia/environments/v1.10/Manifest.toml`
  [f67ccb44] + HDF5 v0.17.1
  [3da0fdf6] + MPIPreferences v0.1.10
  [ae029012] + Requires v1.3.0
^ [0234f1f7] + HDF5_jll v1.12.2+2
        Info Packages marked with ^ have new versions available and may be upgradable.
Precompiling project...
  ✓ MPIPreferences
  ✓ Requires
  ✓ HDF5_jll
  ✓ HDF5
  4 dependencies successfully precompiled in 10 seconds. 60 already precompiled.
```

To write an example file in julia

In [41]:
```julia
example_hd = h5open("example.h5", "cw")
```

Out[41]: 📁 HDF5.File: (read-write) example.h5

In [43]:
```julia
close(example_hd)
```

In [45]:
```julia
example_hd = h5open("example.h5", "r+")
```

Out[45]: 📁 HDF5.File: (read-write) example.h5

Since HDF5 files are hierarchical and based upon a file-system like structure, we create groups

Once we have groups, we write to the groups, and perhaps subgroups, subsubgroups etc. - remember, HDF5 is akin to a filesystem

We'll create some matrix data which we'll write to a group called foo/A in R

```r
A = matrix(1:10, nr=5, nc=2)
h5write(A, "example.h5", "foo/A") #group is the third argument
B = array(seq(0.1,2.0, by=0.1), dim=c(5,2,2))
attr(B, "scale") <- "liter"
h5write(B, "example.h5", "foo/foobaa/B")
h5ls("example.h5") # h5 ls view
```

In Julia let's work with the HDF5 package

Create a group called "foo"

In [46]:
```julia
create_group(example_hd, "foo")
```

Out[46]: 📁 HDF5.Group: /foo (file: example.h5)

Create some mock data

In [49]:
```julia
samp = Array(rand(2, 4))
```

Out[49]:
```
2×4 Matrix{Float64}:
 0.453924  0.774484  0.896302  0.714723
 0.141766  0.911775  0.745648  0.932475
```

Write it to the group foo

In [51]:
```julia
# If a group doesn't already exist we can write to it by indexing
example_hd["newgroup"] = "yes"
```

Out[51]: "yes"

In [66]:
```julia
# To write to a pre-existing group
# First initialize the group and load it into a variable
g = example_hd["foo"]
# Write to it by indexing directly
g["mydataset"] = samp
# Write to it using the create_dataset() function
create_dataset(g, "simplestring", zeros(1, 2))
```

Out[66]: (HDF5.Dataset: /foo/simplestring (file: example.h5 xfer_mode: 0), HDF5.Datatype: H5T_IEEE_F64LE)

In [60]:
```julia
example_hd["foo/mydataset"]
```

Out[60]: 🔢 HDF5.Dataset: /foo/mydataset (file: example.h5 xfer_mode: 0)

Read the contents of the dataset and its groups using the **read** function

In [61]:
```julia
read(example_hd["foo/mydataset"])
```

Out[61]:
```
2×4 Matrix{Float64}:
 0.453924  0.774484  0.896302  0.714723
 0.141766  0.911775  0.745648  0.932475
```

In [67]:
```julia
read(example_hd["foo/simplestring"])
```

Out[67]: 1×2 Matrix{Float64}:

```
In [68]:  example_hd
```

Out[68]: 📁 HDF5.File: (read-write) example.h5
          ├─ 📁 foo
          │ ├─ 🔢 mydataset
          │ └─ 🔢 simplestring
          └─ 🔢 newgroup

```
In [71]:  close(example_hd)
```

For convience and consistency we can also use the **do** block conventions, which will take care of closing the stream for us

```
In [73]:  h5open("example.h5", "r+") do stream
              group = create_group(stream, "dogroup")
              dataset = create_dataset(group, "thisdata", Float64, (10,10))
              write(dataset, rand(10,10))
          end
```

Basic notes; Datasets can be create by

```
g["mydataset"] = rand(3,5)
# or
write(g, "mydataset", rand(3,5))
```

Reading specific parts of the data

In R;

```
h5read("example.h5", "foo/A")
h5read("example.h5", "foo/new/dataset")
```

In Julia

```
In [89]:  openh5 = h5open("example.h5", "r+")
```

Out[89]: 📁 HDF5.File: (read-write) example.h5
          ├─ 📁 dogroup
          │ └─ 🔢 thisdata
          ├─ 📁 foo
          │ ├─ 🔢 mydataset
          │ └─ 🔢 simplestring
          └─ 🔢 newgroup

```
In [91]:  read(openh5,"foo/mydataset")
```

Out[91]: 2×4 Matrix{Float64}:
           0.453924 0.774484 0.896302 0.714723
           0.141766 0.911775 0.745648 0.932475

Chunking and Compression

The section on the Julia HDF5.jl package site does a nice job of explaining this !
https://juliaio.github.io/HDF5.jl/stable/

Webscraping HTML

Scraping information from the internet is a fun endevour, which based upon the virility of ones creativity, can lead to a large number of exciting avenues of adventure. As the web has been around for decades,
```

almost every high level programming language has its own packages and libraries for working with web HTML and XML files. The most important part of web scraping is to avoid excessive request to websites as this will likely get your IP address blocked. We can safely assume that large companies such as amazon are interested in protecting themselves from massive data trawling operations, even though they are doing it on a massive scale.

## Examples using google scholar

In R first;

We'll store a url link, and then simply read the lines of the link, and then close the link

```
con = url("https://scholar.google.com/citations?user=kzht3-0AAAAJ&hl=en")
htmlCode = readLines(con)
close(con)
htmlCode
```

Now in Julia - we can use the HTTP.jl package - or Requests.jl

In [105... `Pkg.add("HTTP") ; using HTTP`

```
  Resolving package versions...
  Installed ExceptionUnwrapping — v0.1.10
  Installed SimpleBufferStream —— v1.1.0
  Installed ConcurrentUtilities — v2.3.0
  Installed BitFlags ───────────── v0.1.8
  Installed OpenSSL ───────────── v1.4.1
  Installed LoggingExtras ─────── v1.0.3
  Installed URIs ───────────────── v1.5.1
  Installed HTTP ───────────────── v1.10.1
   Updating `~/.julia/environments/v1.10/Project.toml`
  [cd3eb016] + HTTP v1.10.1
   Updating `~/.julia/environments/v1.10/Manifest.toml`
  [d1d4a3ce] + BitFlags v0.1.8
  [f0e56b4a] + ConcurrentUtilities v2.3.0
  [460bff9d] + ExceptionUnwrapping v0.1.10
  [cd3eb016] + HTTP v1.10.1
  [e6f89c97] + LoggingExtras v1.0.3
  [4d8831e6] + OpenSSL v1.4.1
  [777ac1f9] + SimpleBufferStream v1.1.0
  [5c2747f8] + URIs v1.5.1
Precompiling project...
  ✓ BitFlags
  ✓ SimpleBufferStream
  ✓ ExceptionUnwrapping
  ✓ LoggingExtras
  ✓ URIs
  ✓ ConcurrentUtilities
  ✓ OpenSSL
  ✓ HTTP
  8 dependencies successfully precompiled in 7 seconds. 64 already precompiled.
```

In [94]: `url = "https://scholar.google.com/citations?user=kzht3-0AAAAJ&hl=en"`

Out[94]: `"https://scholar.google.com/citations?user=kzht3-0AAAAJ&hl=en"`

In [113... `resp = HTTP.get(url)`

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
Out[113]: HTTP.Messages.Response:
         """
         HTTP/1.1 200 OK
         Date: Wed, 31 Jan 2024 02:32:23 GMT
         Expires: Fri, 01 Jan 1990 00:00:00 GMT
         Cache-Control: no-cache, must-revalidate
         Pragma: no-cache
         Content-Type: text/html; charset=ISO-8859-1
         X-Content-Type-Options: nosniff
         Content-Encoding: gzip
         Server: citations
         X-XSS-Protection: 0
         X-Frame-Options: SAMEORIGIN
         Alt-Svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000
         Transfer-Encoding: chunked

         <!doctype html><html><head><title>Michael Lynch - Google Scholar</title><meta http-e
         quiv="Content-Type" content="text/html;charset=ISO-8859-1"><meta http-equiv="X-UA-Co
         mpatible" content="IE=Edge"><meta name="referrer" content="always"><meta name="viewp
         ort" content="width=device-width,initial-scale=1,minimum-scale=1,maximum-scale=2"><m
         eta name="format-detection" content="telephone=no"><link rel="shortcut icon" href="/
         favicon.ico"><link rel="canonical" href="https://scholar.google.com/citations?user=k
         zht3-0AAAAJ&amp;hl=en"><meta name="description" content="Arizona State University -
         Cited by 79,474 - evolution - genetics - cell biology"><meta property="og:descriptio
         n" content="Arizona State University - Cited by 79,474 - evolution - genetics - cell
         biology"><meta property="og:title" content="Michael Lynch"><meta property="og:image"
         content="https://scholar.google.com/citations/images/avatar_scholar_256.png"><meta p
         roperty="og:type" content="website"><meta name="twitter:card" content="sum
          ⋮
         152364-byte body
         """
```

A note on GET vs POST requests

GET

- parameters are in the URL
- used for fetching and *GETTING* documents
- maximum URL length
- OK to cache
- won't change the server

POST

- parameters are in the body
- used for updating and *POSTING* data
- not ok to cache
- can change the server

Parsing HTML file as a XML - conversion

Depending on the structure of the HTML, we may be able to translate it as a XML and interpret it in the
same structure - let's give it a go and see if it works.

In R;

```r
library(XML)
url <- "https://scholar.google.com/citations?user=kzht3-0AAAAJ&hl=en"
html <- htmlTreeParse(url, useInternalNodes=T)
Using xpath language queries
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
xpathSApply(html, "//title", xmlValue)
```
Now let's look at the citation counts

```
xpathSApply(html, "//td[@id='col-citedby']", xmlValue)
```

Give it a crack in Julia using the EzXML package - it will take a few more steps, and perhaps it can get compressed down with experience into some simpler code.

The basic steps are

1. Load HTTP package, perform a HTTP get request on the url
2. Parse the URL as a string by accessing the requests body
3. Using the EzXML package, parse the html string using parsehtml()
4. Use the xpath query language with findall() and nodecontent() to extract the relevant information from the file

In [114… `Pkg.add("EzXML") ; using EzXML`

```
  Resolving package versions...
  Installed XML2_jll — v2.12.2+0
  Installed EzXML ——— v1.2.0
   Updating `~/.julia/environments/v1.10/Project.toml`
  [8f5d6c58] + EzXML v1.2.0
   Updating `~/.julia/environments/v1.10/Manifest.toml`
  [8f5d6c58] + EzXML v1.2.0
  [02c8fc9c] + XML2_jll v2.12.2+0
Precompiling project...
  ✓ XML2_jll
  ✓ EzXML
  2 dependencies successfully precompiled in 2 seconds. 72 already precompiled.
```

Open a HTTP request, and then parse the HTML file as a string which can be read by the EzXML package using the **parsehtml()** function.

In [144…
```
samp = HTTP.get(url, cookies=true);
data = String(samp.body)
```

Out[144]: "<!doctype html><html><head><title>Michael Lynch - Google Scholar</title><meta http-equiv=\"Content-Type\" content=\"text/html;charset=ISO-8859-1\"><meta http-equiv=\"X-UA-Compatible\" content=\"IE=Edge\"><meta name=\"referrer\" content=\"always\"><meta name=\"viewport\" content=\"wid" ⋯ **151840 bytes** ⋯ "le=\"menuitem\" href=\"/intl/en/scholar/about.html\" tabindex=\"-1\" class=\"gs_md_li\">About ScholarSearch help</div></div></div></body></html>"

Parse the html file now and store the root as a variable

In [154…
```
q = parsehtml(data)
scholar_root = root(q)
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Out[154]: EzXML.Node(<ELEMENT_NODE[html]@0x0000000005486f00>)

Using the **nodecontent.()** function of EzXML, find your query using the xpath language

In [160…
```julia
nodecontent.(findall("//title", scholar_root))
```

Out[160]: 1-element Vector{String}:
    "Michael Lynch - Google Scholar"

Now for the more elaborate Xpath query

In [164…
```julia
for citation in nodecontent.(findall("//td[@id='col-citedby']", scholar_root))
    println(citation)
end
```

Using the httr library for R

httr makes doing much of this a bit easier - as we can see the workflow is very similar to that of HTTP.jl

```r
library(httr); html2 = GET(url)
content2 = content(html2, as="text")
parsedHtml = htmlParse(content2, asText=True)
xpathSapply(parsedHtml, "//title", xmlValue)
```

Websites with user and password authentication

In order to access websites which request a user and password, in R we can include the information as an argument

```r
pp2 = GET("url", authenticate("user","passwd"))
```
In Julia we prefix the server address as such;

In [168…
```julia
pp2 = HTTP.get("https://user:passwd@httpbin.org/basic-auth/user/passwd")
```

Out[168]: HTTP.Messages.Response:
          """
          HTTP/1.1 200 OK
          Date: Thu, 01 Feb 2024 02:37:55 GMT
          Content-Type: application/json
          Content-Length: 47
          Connection: keep-alive
          Server: gunicorn/19.9.0
          Access-Control-Allow-Origin: *
          Access-Control-Allow-Credentials: true

          {
            "authenticated": true,
            "user": "user"
          }
          """

Quiz

1. Read up on the github API

https://github.com/settings/applications . Access the API to get information on your instructors repositories (hint: this is the url you want "https://api.github.com/users/jtleek/repos"). Use this data to find the time that the datasharing repo was created. What time was it created?

This tutorial may be useful (https://github.com/hadley/httr/blob/master/demo/oauth2-github.r). You may also need to run the code in the base R package and not R studio.

In Julia, we'll use the github.jl package from https://github.com/JuliaWeb/GitHub.jl

```julia
In [174… Pkg.add("GitHub") ; using GitHub
```

```
  Resolving package versions...
No Changes to `~/.julia/environments/v1.10/Project.toml`
No Changes to `~/.julia/environments/v1.10/Manifest.toml`
```

Using julias github package, let's take a look at all of jtleeks repos, using the simple **repos** function with his username as the argument

```julia
In [197… leeks_repos = repos("jtleek")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
Out[197]: (Repo[Repo (all fields are Union{Nothing, T}):
           name: "2018"
           full_name: "jtleek/2018"
           description: "Fall 2018 repository with course materials for JHU Advanced Data Sci
         ence"
           language: "HTML"
           default_branch: "master"
           owner: Owner("jtleek")
           id: 155565363
           size: 60855
           forks_count: 3
           stargazers_count: 1
           watchers_count: 1
           open_issues_count: 0
           url: URI("https://api.github.com/repos/jtleek/2018")
           html_url: URI("https://github.com/jtleek/2018")
           clone_url: URI("https://github.com/jtleek/2018.git")
           ssh_url: URI("git@github.com:jtleek/2018.git")
           homepage: URI("https://jhu-advdatasci.github.io/2018/")
           pushed_at: DateTime("2018-10-30T18:13:41")
           created_at: DateTime("2018-10-31T13:50:37")
           updated_at: DateTime("2021-12-05T09:30:00")
           has_issues: false
           has_wiki: true
           has_downloads: true
           has_pages: false
           private: false
           fork: true, Repo (all fields are Union{Nothing, T}):
           name: "ads2020"
           full_name: "jtleek/ads2020"
           description: "Advanced Data Science 2020 Edition"
           language: "CSS"
           default_branch: "master"
           owner: Owner("jtleek")
           id: 264786491
           size: 75774
           forks_count: 22
           stargazers_count: 73
           watchers_count: 73
           open_issues_count: 10
           url: URI("https://api.github.com/repos/jtleek/ads2020")
           html_url: URI("https://github.com/jtleek/ads2020")
           clone_url: URI("https://github.com/jtleek/ads2020.git")
           ssh_url: URI("git@github.com:jtleek/ads2020.git")
           pushed_at: DateTime("2023-10-10T14:55:03")
           created_at: DateTime("2020-05-18T00:45:46")
           updated_at: DateTime("2023-11-24T17:22:10")
           has_issues: true
           has_wiki: true
           has_downloads: true
           has_pages: true
           private: false
           fork: false, Repo (all fields are Union{Nothing, T}):
           name: "advdatasci"
           full_name: "jtleek/advdatasci"
           language: "HTML"
           default_branch: "master"
           owner: Owner("jtleek")
           id: 101394164
           size: 172184
           forks_count: 12
           stargazers_count: 23
           watchers_count: 23
           open_issues_count: 0
```
.com/repos/jtleek/advdatasci")

```
html_url: URI("https://github.com/jtleek/advdatasci")
clone_url: URI("https://github.com/jtleek/advdatasci.git")
ssh_url: URI("git@github.com:jtleek/advdatasci.git")
pushed_at: DateTime("2020-05-17T23:06:04")
created_at: DateTime("2017-08-25T10:38:16")
updated_at: DateTime("2023-09-08T17:29:07")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "advdatasci-project"
full_name: "jtleek/advdatasci-project"
description: "Awesome project! "
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 111447948
size: 680
forks_count: 1
stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci-project")
html_url: URI("https://github.com/jtleek/advdatasci-project")
clone_url: URI("https://github.com/jtleek/advdatasci-project.git")
ssh_url: URI("git@github.com:jtleek/advdatasci-project.git")
pushed_at: DateTime("2017-11-20T18:40:45")
created_at: DateTime("2017-11-20T18:33:29")
updated_at: DateTime("2023-09-08T17:32:51")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci-swirl"
full_name: "jtleek/advdatasci-swirl"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 47568815
size: 188
forks_count: 2
stargazers_count: 4
watchers_count: 4
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/advdatasci-swirl")
html_url: URI("https://github.com/jtleek/advdatasci-swirl")
clone_url: URI("https://github.com/jtleek/advdatasci-swirl.git")
ssh_url: URI("git@github.com:jtleek/advdatasci-swirl.git")
pushed_at: DateTime("2015-12-09T20:21:13")
created_at: DateTime("2015-12-07T17:43:53")
updated_at: DateTime("2018-04-30T21:02:30")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci15"
full_name: "jtleek/advdatasci15"
description: "Advanced Data Science @ JHU Biostats"
language: "HTML"
```

```
owner: Owner("jtleek")
id: 41645119
size: 200992
forks_count: 7
stargazers_count: 16
watchers_count: 16
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci15")
html_url: URI("https://github.com/jtleek/advdatasci15")
clone_url: URI("https://github.com/jtleek/advdatasci15.git")
ssh_url: URI("git@github.com:jtleek/advdatasci15.git")
pushed_at: DateTime("2015-12-16T18:12:31")
created_at: DateTime("2015-08-30T22:27:26")
updated_at: DateTime("2023-02-14T22:06:27")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci16"
full_name: "jtleek/advdatasci16"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 65922328
size: 3427
forks_count: 8
stargazers_count: 7
watchers_count: 7
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci16")
html_url: URI("https://github.com/jtleek/advdatasci16")
clone_url: URI("https://github.com/jtleek/advdatasci16.git")
ssh_url: URI("git@github.com:jtleek/advdatasci16.git")
pushed_at: DateTime("2017-12-06T17:51:23")
created_at: DateTime("2016-08-17T16:08:20")
updated_at: DateTime("2023-09-08T17:13:47")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci_swirl"
full_name: "jtleek/advdatasci_swirl"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 66415014
size: 4683
forks_count: 5
stargazers_count: 5
watchers_count: 5
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/advdatasci_swirl")
html_url: URI("https://github.com/jtleek/advdatasci_swirl")
clone_url: URI("https://github.com/jtleek/advdatasci_swirl.git")
ssh_url: URI("git@github.com:jtleek/advdatasci_swirl.git")
pushed_at: DateTime("2016-10-26T14:05:49")
created_at: DateTime("2016-08-24T00:44:35")
updated_at: DateTime("2023-09-08T17:14:02")
has_issues: true
has_wiki: true
has_downloads: true
```

```
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "ai"
full_name: "jtleek/ai"
description: "A few AI resources that I've found interesting or that we are workin
g on"
default_branch: "main"
owner: Owner("jtleek")
id: 650622901
size: 10
forks_count: 1
stargazers_count: 5
watchers_count: 5
open_issues_count: 3
url: URI("https://api.github.com/repos/jtleek/ai")
html_url: URI("https://github.com/jtleek/ai")
clone_url: URI("https://github.com/jtleek/ai.git")
ssh_url: URI("git@github.com:jtleek/ai.git")
pushed_at: DateTime("2023-06-07T18:43:02")
created_at: DateTime("2023-06-07T13:06:23")
updated_at: DateTime("2023-06-15T12:26:58")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "ballgown"
full_name: "jtleek/ballgown"
description: "code for manipulating ballgown output in R"
language: "C++"
default_branch: "master"
owner: Owner("jtleek")
id: 12441219
size: 10421
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/ballgown")
html_url: URI("https://github.com/jtleek/ballgown")
clone_url: URI("https://github.com/jtleek/ballgown.git")
ssh_url: URI("git@github.com:jtleek/ballgown.git")
pushed_at: DateTime("2014-03-27T15:06:21")
created_at: DateTime("2013-08-28T18:18:50")
updated_at: DateTime("2023-09-08T16:41:42")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true  …  Repo (all fields are Union{Nothing, T}):
name: "talks"
full_name: "jtleek/talks"
description: "Slides from presentations"
default_branch: "master"
owner: Owner("jtleek")
id: 12133281
size: 23141
forks_count: 6
stargazers_count: 14
watchers_count: 14
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/talks")
html_url: URI("https://github.com/jtleek/talks")
ub.com/jtleek/talks.git")
```

```
ssh_url: URI("git@github.com:jtleek/talks.git")
pushed_at: DateTime("2024-01-31T14:09:44")
created_at: DateTime("2013-08-15T11:52:51")
updated_at: DateTime("2023-11-24T17:27:06")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testdir"
full_name: "jtleek/testdir"
description: "this is a test"
default_branch: "master"
owner: Owner("jtleek")
id: 38569111
size: 112
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testdir")
html_url: URI("https://github.com/jtleek/testdir")
clone_url: URI("https://github.com/jtleek/testdir.git")
ssh_url: URI("git@github.com:jtleek/testdir.git")
pushed_at: DateTime("2015-07-05T13:07:11")
created_at: DateTime("2015-07-05T12:57:04")
updated_at: DateTime("2015-07-05T12:57:04")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testproject"
full_name: "jtleek/testproject"
default_branch: "master"
owner: Owner("jtleek")
id: 38634839
size: 112
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testproject")
html_url: URI("https://github.com/jtleek/testproject")
clone_url: URI("https://github.com/jtleek/testproject.git")
ssh_url: URI("git@github.com:jtleek/testproject.git")
pushed_at: DateTime("2015-07-06T17:37:08")
created_at: DateTime("2015-07-06T17:34:35")
updated_at: DateTime("2015-07-06T17:34:35")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testrepository"
full_name: "jtleek/testrepository"
description: "testrepository"
default_branch: "master"
owner: Owner("jtleek")
id: 4729869
size: 48
```

```
stargazers_count: 3
watchers_count: 3
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testrepository")
html_url: URI("https://github.com/jtleek/testrepository")
clone_url: URI("https://github.com/jtleek/testrepository.git")
ssh_url: URI("git@github.com:jtleek/testrepository.git")
pushed_at: DateTime("2012-06-20T18:39:07")
created_at: DateTime("2012-06-20T18:39:06")
updated_at: DateTime("2017-11-13T14:02:46")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "tidypvals"
full_name: "jtleek/tidypvals"
description: "An R package with several million published p-values in tidy data se
ts. "
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 98407699
size: 55034
forks_count: 4
stargazers_count: 74
watchers_count: 74
open_issues_count: 2
url: URI("https://api.github.com/repos/jtleek/tidypvals")
html_url: URI("https://github.com/jtleek/tidypvals")
clone_url: URI("https://github.com/jtleek/tidypvals.git")
ssh_url: URI("git@github.com:jtleek/tidypvals.git")
homepage: URI("")
pushed_at: DateTime("2017-08-04T02:21:47")
created_at: DateTime("2017-07-26T09:51:06")
updated_at: DateTime("2023-09-08T17:27:54")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "tornado"
full_name: "jtleek/tornado"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 6374724
size: 484
forks_count: 2
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/tornado")
html_url: URI("https://github.com/jtleek/tornado")
clone_url: URI("https://github.com/jtleek/tornado.git")
ssh_url: URI("git@github.com:jtleek/tornado.git")
pushed_at: DateTime("2012-09-26T19:49:17")
created_at: DateTime("2012-10-24T17:50:09")
updated_at: DateTime("2014-03-04T14:58:03")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
```

```
fork: false, Repo (all fields are Union{Nothing, T}):
name: "tsp-devel"
full_name: "jtleek/tsp-devel"
description: "TSP Development Software"
language: "C++"
default_branch: "master"
owner: Owner("jtleek")
id: 12784874
size: 112
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/tsp-devel")
html_url: URI("https://github.com/jtleek/tsp-devel")
clone_url: URI("https://github.com/jtleek/tsp-devel.git")
ssh_url: URI("git@github.com:jtleek/tsp-devel.git")
pushed_at: DateTime("2013-09-20T19:58:53")
created_at: DateTime("2013-09-12T13:49:18")
updated_at: DateTime("2013-11-25T18:49:53")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "tspreg"
full_name: "jtleek/tspreg"
description: "An R package for performing top-scoring pairs regression."
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 12134822
size: 190
forks_count: 8
stargazers_count: 4
watchers_count: 4
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/tspreg")
html_url: URI("https://github.com/jtleek/tspreg")
clone_url: URI("https://github.com/jtleek/tspreg.git")
ssh_url: URI("git@github.com:jtleek/tspreg.git")
pushed_at: DateTime("2015-03-31T01:29:50")
created_at: DateTime("2013-08-15T13:20:30")
updated_at: DateTime("2020-10-23T14:27:36")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "which_house_races_are_close"
full_name: "jtleek/which_house_races_are_close"
description: "Join together FiveThirtyEight + Cook ratings on House races and put
into easily sortable table. "
language: "Jupyter Notebook"
default_branch: "master"
owner: Owner("jtleek")
id: 145731567
size: 885
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/which_house_races_are_close")
```
b.com/jtleek/which_house_races_are_close")

```
    clone_url: URI("https://github.com/jtleek/which_house_races_are_close.git")
    ssh_url: URI("git@github.com:jtleek/which_house_races_are_close.git")
    pushed_at: DateTime("2018-08-22T17:10:59")
    created_at: DateTime("2018-08-22T15:56:55")
    updated_at: DateTime("2018-08-22T17:11:00")
    has_issues: false
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: true, Repo (all fields are Union{Nothing, T}):
    name: "who-is-who"
    full_name: "jtleek/who-is-who"
    description: "A repo to create a who's who website for the Fred Hutch data communi
ty"
    default_branch: "main"
    owner: Owner("jtleek")
    id: 571676878
    size: 367
    forks_count: 0
    stargazers_count: 0
    watchers_count: 0
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/who-is-who")
    html_url: URI("https://github.com/jtleek/who-is-who")
    clone_url: URI("https://github.com/jtleek/who-is-who.git")
    ssh_url: URI("git@github.com:jtleek/who-is-who.git")
    homepage: URI("https://whoiswho.fredhutch.org")
    pushed_at: DateTime("2022-11-28T16:43:46")
    created_at: DateTime("2022-11-28T16:39:50")
    updated_at: DateTime("2023-09-08T18:39:26")
    has_issues: false
    has_wiki: false
    has_downloads: true
    has_pages: false
    private: false
    fork: true], Dict("prev" => "https://api.github.com/user/1571674/repos?page=3", "f
irst" => "https://api.github.com/user/1571674/repos?page=1"))
```

In [210…
```
# Doesn't work - I can't seem to get an iterator going
for f in leeks_repos
        print(f)
    end
end
```

```
Repo[Repo (all fields are Union{Nothing, T}):
  name: "2018"
  full_name: "jtleek/2018"
  description: "Fall 2018 repository with course materials for JHU Advanced Data Scien
ce"
  language: "HTML"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 155565363
  size: 60855
  forks_count: 3
  stargazers_count: 1
  watchers_count: 1
  open_issues_count: 0
  url: URI("https://api.github.com/repos/jtleek/2018")
  html_url: URI("https://github.com/jtleek/2018")
  clone_url: URI("https://github.com/jtleek/2018.git")
  ssh_url: URI("git@github.com:jtleek/2018.git")
  homepage: URI("https://jhu-advdatasci.github.io/2018/")
  pushed_at: DateTime("2018-10-30T18:13:41")
  created_at: DateTime("2018-10-31T13:50:37")
  updated_at: DateTime("2021-12-05T09:30:00")
  has_issues: false
  has_wiki: true
  has_downloads: true
  has_pages: false
  private: false
  fork: true, Repo (all fields are Union{Nothing, T}):
  name: "ads2020"
  full_name: "jtleek/ads2020"
  description: "Advanced Data Science 2020 Edition"
  language: "CSS"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 264786491
  size: 75774
  forks_count: 22
  stargazers_count: 73
  watchers_count: 73
  open_issues_count: 10
  url: URI("https://api.github.com/repos/jtleek/ads2020")
  html_url: URI("https://github.com/jtleek/ads2020")
  clone_url: URI("https://github.com/jtleek/ads2020.git")
  ssh_url: URI("git@github.com:jtleek/ads2020.git")
  pushed_at: DateTime("2023-10-10T14:55:03")
  created_at: DateTime("2020-05-18T00:45:46")
  updated_at: DateTime("2023-11-24T17:22:10")
  has_issues: true
  has_wiki: true
  has_downloads: true
  has_pages: true
  private: false
  fork: false, Repo (all fields are Union{Nothing, T}):
  name: "advdatasci"
  full_name: "jtleek/advdatasci"
  language: "HTML"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 101394164
  size: 172184
  forks_count: 12
  stargazers_count: 23
  watchers_count: 23
  open_issues_count: 0
  url: URI("https://api.github.com/repos/jtleek/advdatasci")
```
```
com/jtleek/advdatasci")
```

```
clone_url: URI("https://github.com/jtleek/advdatasci.git")
ssh_url: URI("git@github.com:jtleek/advdatasci.git")
pushed_at: DateTime("2020-05-17T23:06:04")
created_at: DateTime("2017-08-25T10:38:16")
updated_at: DateTime("2023-09-08T17:29:07")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "advdatasci-project"
full_name: "jtleek/advdatasci-project"
description: "Awesome project! "
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 111447948
size: 680
forks_count: 1
stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci-project")
html_url: URI("https://github.com/jtleek/advdatasci-project")
clone_url: URI("https://github.com/jtleek/advdatasci-project.git")
ssh_url: URI("git@github.com:jtleek/advdatasci-project.git")
pushed_at: DateTime("2017-11-20T18:40:45")
created_at: DateTime("2017-11-20T18:33:29")
updated_at: DateTime("2023-09-08T17:32:51")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci-swirl"
full_name: "jtleek/advdatasci-swirl"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 47568815
size: 188
forks_count: 2
stargazers_count: 4
watchers_count: 4
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/advdatasci-swirl")
html_url: URI("https://github.com/jtleek/advdatasci-swirl")
clone_url: URI("https://github.com/jtleek/advdatasci-swirl.git")
ssh_url: URI("git@github.com:jtleek/advdatasci-swirl.git")
pushed_at: DateTime("2015-12-09T20:21:13")
created_at: DateTime("2015-12-07T17:43:53")
updated_at: DateTime("2018-04-30T21:02:30")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci15"
full_name: "jtleek/advdatasci15"
description: "Advanced Data Science @ JHU Biostats"
language: "HTML"
default_branch: "gh-pages"
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

id: 41645119
size: 200992
forks_count: 7
stargazers_count: 16
watchers_count: 16
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci15")
html_url: URI("https://github.com/jtleek/advdatasci15")
clone_url: URI("https://github.com/jtleek/advdatasci15.git")
ssh_url: URI("git@github.com:jtleek/advdatasci15.git")
pushed_at: DateTime("2015-12-16T18:12:31")
created_at: DateTime("2015-08-30T22:27:26")
updated_at: DateTime("2023-02-14T22:06:27")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci16"
full_name: "jtleek/advdatasci16"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 65922328
size: 3427
forks_count: 8
stargazers_count: 7
watchers_count: 7
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/advdatasci16")
html_url: URI("https://github.com/jtleek/advdatasci16")
clone_url: URI("https://github.com/jtleek/advdatasci16.git")
ssh_url: URI("git@github.com:jtleek/advdatasci16.git")
pushed_at: DateTime("2017-12-06T17:51:23")
created_at: DateTime("2016-08-17T16:08:20")
updated_at: DateTime("2023-09-08T17:13:47")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "advdatasci_swirl"
full_name: "jtleek/advdatasci_swirl"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 66415014
size: 4683
forks_count: 5
stargazers_count: 5
watchers_count: 5
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/advdatasci_swirl")
html_url: URI("https://github.com/jtleek/advdatasci_swirl")
clone_url: URI("https://github.com/jtleek/advdatasci_swirl.git")
ssh_url: URI("git@github.com:jtleek/advdatasci_swirl.git")
pushed_at: DateTime("2016-10-26T14:05:49")
created_at: DateTime("2016-08-24T00:44:35")
updated_at: DateTime("2023-09-08T17:14:02")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false

```
fork: false, Repo (all fields are Union{Nothing, T}):
name: "ai"
full_name: "jtleek/ai"
description: "A few AI resources that I've found interesting or that we are working
on"
default_branch: "main"
owner: Owner("jtleek")
id: 650622901
size: 10
forks_count: 1
stargazers_count: 5
watchers_count: 5
open_issues_count: 3
url: URI("https://api.github.com/repos/jtleek/ai")
html_url: URI("https://github.com/jtleek/ai")
clone_url: URI("https://github.com/jtleek/ai.git")
ssh_url: URI("git@github.com:jtleek/ai.git")
pushed_at: DateTime("2023-06-07T18:43:02")
created_at: DateTime("2023-06-07T13:06:23")
updated_at: DateTime("2023-06-15T12:26:58")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "ballgown"
full_name: "jtleek/ballgown"
description: "code for manipulating ballgown output in R"
language: "C++"
default_branch: "master"
owner: Owner("jtleek")
id: 12441219
size: 10421
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/ballgown")
html_url: URI("https://github.com/jtleek/ballgown")
clone_url: URI("https://github.com/jtleek/ballgown.git")
ssh_url: URI("git@github.com:jtleek/ballgown.git")
pushed_at: DateTime("2014-03-27T15:06:21")
created_at: DateTime("2013-08-28T18:18:50")
updated_at: DateTime("2023-09-08T16:41:42")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "big_course"
full_name: "jtleek/big_course"
description: "Repo for the BIG Experiment MOOC"
default_branch: "master"
owner: Owner("jtleek")
id: 258010045
size: 0
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/big_course")
html_url: URI("https://github.com/jtleek/big_course")
clone_url: URI("https://github.com/jtleek/big_course.git")
```
tleek/big_course.git")

```
pushed_at: DateTime("2020-04-22T20:16:03")
created_at: DateTime("2020-04-22T20:16:02")
updated_at: DateTime("2020-08-22T00:48:38")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "bookdown-start"
full_name: "jtleek/bookdown-start"
language: "TeX"
default_branch: "master"
owner: Owner("jtleek")
id: 75141158
size: 237
forks_count: 2
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/bookdown-start")
html_url: URI("https://github.com/jtleek/bookdown-start")
clone_url: URI("https://github.com/jtleek/bookdown-start.git")
ssh_url: URI("git@github.com:jtleek/bookdown-start.git")
pushed_at: DateTime("2016-11-18T20:07:47")
created_at: DateTime("2016-11-30T02:05:00")
updated_at: DateTime("2023-09-08T17:17:44")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "books"
full_name: "jtleek/books"
description: "Books by Jeff Leek"
default_branch: "main"
owner: Owner("jtleek")
id: 320396846
size: 4
forks_count: 2
stargazers_count: 11
watchers_count: 11
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/books")
html_url: URI("https://github.com/jtleek/books")
clone_url: URI("https://github.com/jtleek/books.git")
ssh_url: URI("git@github.com:jtleek/books.git")
pushed_at: DateTime("2020-12-10T21:42:39")
created_at: DateTime("2020-12-10T21:42:01")
updated_at: DateTime("2023-12-06T04:23:45")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "capitalIn21stCenturyinR"
full_name: "jtleek/capitalIn21stCenturyinR"
description: "Piketty in R"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 20234724
size: 374812
```

stargazers_count: 213
watchers_count: 213
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/capitalIn21stCenturyinR")
html_url: URI("https://github.com/jtleek/capitalIn21stCenturyinR")
clone_url: URI("https://github.com/jtleek/capitalIn21stCenturyinR.git")
ssh_url: URI("git@github.com:jtleek/capitalIn21stCenturyinR.git")
pushed_at: DateTime("2016-07-18T17:22:51")
created_at: DateTime("2014-05-27T20:38:31")
updated_at: DateTime("2024-01-04T22:12:27")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "careerplanning"
full_name: "jtleek/careerplanning"
description: "A career planning guide."
default_branch: "master"
owner: Owner("jtleek")
id: 36437287
size: 128
forks_count: 23
stargazers_count: 119
watchers_count: 119
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/careerplanning")
html_url: URI("https://github.com/jtleek/careerplanning")
clone_url: URI("https://github.com/jtleek/careerplanning.git")
ssh_url: URI("git@github.com:jtleek/careerplanning.git")
pushed_at: DateTime("2015-05-28T13:51:02")
created_at: DateTime("2015-05-28T12:37:04")
updated_at: DateTime("2024-01-31T21:52:19")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "coc"
full_name: "jtleek/coc"
description: "Leek Group Code of Conduct"
default_branch: "main"
owner: Owner("jtleek")
id: 320457870
size: 6
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/coc")
html_url: URI("https://github.com/jtleek/coc")
clone_url: URI("https://github.com/jtleek/coc.git")
ssh_url: URI("git@github.com:jtleek/coc.git")
pushed_at: DateTime("2020-12-11T03:50:54")
created_at: DateTime("2020-12-11T03:34:56")
updated_at: DateTime("2020-12-11T03:50:57")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "courses"

```
description: "Courses taught by Jeff"
default_branch: "main"
owner: Owner("jtleek")
id: 320462039
size: 2
forks_count: 4
stargazers_count: 14
watchers_count: 14
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/courses")
html_url: URI("https://github.com/jtleek/courses")
clone_url: URI("https://github.com/jtleek/courses.git")
ssh_url: URI("git@github.com:jtleek/courses.git")
pushed_at: DateTime("2020-12-11T04:04:52")
created_at: DateTime("2020-12-11T03:58:32")
updated_at: DateTime("2023-10-31T19:28:52")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "COVID-19"
full_name: "jtleek/COVID-19"
description: "Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE"
default_branch: "master"
owner: Owner("jtleek")
id: 250507642
size: 65648
forks_count: 2
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/COVID-19")
html_url: URI("https://github.com/jtleek/COVID-19")
clone_url: URI("https://github.com/jtleek/COVID-19.git")
ssh_url: URI("git@github.com:jtleek/COVID-19.git")
homepage: URI("https://systems.jhu.edu/research/public-health/ncov/")
pushed_at: DateTime("2020-03-27T10:39:34")
created_at: DateTime("2020-03-27T10:43:35")
updated_at: DateTime("2020-06-01T13:44:00")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "crsra"
full_name: "jtleek/crsra"
description: "A package for cleaning and analyzing Coursera research data exports"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 136021897
size: 10594
forks_count: 4
stargazers_count: 4
watchers_count: 4
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/crsra")
html_url: URI("https://github.com/jtleek/crsra")
clone_url: URI("https://github.com/jtleek/crsra.git")
ssh_url: URI("git@github.com:jtleek/crsra.git")
homepage: URI("")
pushed_at: DateTime("2018-05-07T15:04:20")
```
04T12:29:13")

updated_at: DateTime("2023-09-08T17:41:19")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "cshlcg-labs"
full_name: "jtleek/cshlcg-labs"
description: "Cold Spring Harbor Labs Computational Genomics"
default_branch: "master"
owner: Owner("jtleek")
id: 112599365
size: 20
forks_count: 7
stargazers_count: 6
watchers_count: 6
open_issues_count: 2
url: URI("https://api.github.com/repos/jtleek/cshlcg-labs")
html_url: URI("https://github.com/jtleek/cshlcg-labs")
clone_url: URI("https://github.com/jtleek/cshlcg-labs.git")
ssh_url: URI("git@github.com:jtleek/cshlcg-labs.git")
pushed_at: DateTime("2019-12-06T15:01:23")
created_at: DateTime("2017-11-30T10:41:31")
updated_at: DateTime("2023-05-30T15:28:11")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "data"
full_name: "jtleek/data"
description: "Data resources created by the Leek group"
default_branch: "main"
owner: Owner("jtleek")
id: 320396454
size: 2
forks_count: 4
stargazers_count: 11
watchers_count: 11
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/data")
html_url: URI("https://github.com/jtleek/data")
clone_url: URI("https://github.com/jtleek/data.git")
ssh_url: URI("git@github.com:jtleek/data.git")
pushed_at: DateTime("2020-12-10T21:40:55")
created_at: DateTime("2020-12-10T21:39:54")
updated_at: DateTime("2023-09-08T18:15:49")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "dataanalysis"
full_name: "jtleek/dataanalysis"
description: "The lecture slides for Coursera's Data Analysis class"
language: "JavaScript"
default_branch: "master"
owner: Owner("jtleek")
id: 7751816
size: 156389
forks_count: 730
stargazers_count: 756

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

open_issues_count: 5
url: URI("https://api.github.com/repos/jtleek/dataanalysis")
html_url: URI("https://github.com/jtleek/dataanalysis")
clone_url: URI("https://github.com/jtleek/dataanalysis.git")
ssh_url: URI("git@github.com:jtleek/dataanalysis.git")
pushed_at: DateTime("2015-09-22T09:38:52")
created_at: DateTime("2013-01-22T12:16:41")
updated_at: DateTime("2024-01-21T14:08:49")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "datascientist"
full_name: "jtleek/datascientist"
description: "datascientist"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 4772877
size: 98
forks_count: 40
stargazers_count: 10
watchers_count: 10
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/datascientist")
html_url: URI("https://github.com/jtleek/datascientist")
clone_url: URI("https://github.com/jtleek/datascientist.git")
ssh_url: URI("git@github.com:jtleek/datascientist.git")
pushed_at: DateTime("2012-06-24T14:38:18")
created_at: DateTime("2012-06-24T14:36:20")
updated_at: DateTime("2023-09-08T16:34:02")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "datasharing"
full_name: "jtleek/datasharing"
description: "The Leek group guide to data sharing "
default_branch: "master"
owner: Owner("jtleek")
id: 14204342
size: 590
forks_count: 243587
stargazers_count: 6434
watchers_count: 6434
open_issues_count: 892
url: URI("https://api.github.com/repos/jtleek/datasharing")
html_url: URI("https://github.com/jtleek/datasharing")
clone_url: URI("https://github.com/jtleek/datasharing.git")
ssh_url: URI("git@github.com:jtleek/datasharing.git")
pushed_at: DateTime("2024-01-05T04:49:32")
created_at: DateTime("2013-11-07T13:25:07")
updated_at: DateTime("2024-02-01T12:03:04")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "datawomenontwitter"
full_name: "jtleek/datawomenontwitter"

p://simplystatistics.org/2014/09/09/a-non-comprehensive-list-of-awesome-female-data-pe
ople-on-twitter/)"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 23840078
  size: 112
  forks_count: 5
  stargazers_count: 7
  watchers_count: 7
  open_issues_count: 1
  url: URI("https://api.github.com/repos/jtleek/datawomenontwitter")
  html_url: URI("https://github.com/jtleek/datawomenontwitter")
  clone_url: URI("https://github.com/jtleek/datawomenontwitter.git")
  ssh_url: URI("git@github.com:jtleek/datawomenontwitter.git")
  pushed_at: DateTime("2014-09-09T15:50:55")
  created_at: DateTime("2014-09-09T15:49:57")
  updated_at: DateTime("2022-08-22T12:21:19")
  has_issues: true
  has_wiki: true
  has_downloads: true
  has_pages: false
  private: false
  fork: false, Repo (all fields are Union{Nothing, T}):
  name: "day1"
  full_name: "jtleek/day1"
  description: "This is for SISBID"
  language: "HTML"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 96824057
  size: 274
  forks_count: 1
  stargazers_count: 0
  watchers_count: 0
  open_issues_count: 0
  url: URI("https://api.github.com/repos/jtleek/day1")
  html_url: URI("https://github.com/jtleek/day1")
  clone_url: URI("https://github.com/jtleek/day1.git")
  ssh_url: URI("git@github.com:jtleek/day1.git")
  pushed_at: DateTime("2017-07-11T15:25:20")
  created_at: DateTime("2017-07-10T21:44:28")
  updated_at: DateTime("2017-07-11T14:13:55")
  has_issues: true
  has_wiki: true
  has_downloads: true
  has_pages: false
  private: false
  fork: false, Repo (all fields are Union{Nothing, T}):
  name: "derfinder"
  full_name: "jtleek/derfinder"
  language: "R"
  default_branch: "master"
  owner: Owner("jtleek")
  id: 11549405
  size: 388
  forks_count: 0
  stargazers_count: 0
  watchers_count: 0
  open_issues_count: 0
  url: URI("https://api.github.com/repos/jtleek/derfinder")
  html_url: URI("https://github.com/jtleek/derfinder")
  clone_url: URI("https://github.com/jtleek/derfinder.git")
  ssh_url: URI("git@github.com:jtleek/derfinder.git")
  pushed_at: DateTime("2013-06-24T21:17:27")
  created_at: DateTime("2013-07-20T16:05:59")
  updated_at: DateTime("...08T16:40:51")

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "derfinder-1"
full_name: "jtleek/derfinder-1"
description: "Fast differential expression analysis of RNA-seq data at base-pair res
olution"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 14240696
size: 367
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/derfinder-1")
html_url: URI("https://github.com/jtleek/derfinder-1")
clone_url: URI("https://github.com/jtleek/derfinder-1.git")
ssh_url: URI("git@github.com:jtleek/derfinder-1.git")
homepage: URI("")
pushed_at: DateTime("2013-11-08T18:19:42")
created_at: DateTime("2013-11-08T18:16:27")
updated_at: DateTime("2013-11-08T20:39:32")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "DSM"
full_name: "jtleek/DSM"
description: "Deterministic Statistical Machine"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 11976743
size: 50000
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/DSM")
html_url: URI("https://github.com/jtleek/DSM")
clone_url: URI("https://github.com/jtleek/DSM.git")
ssh_url: URI("git@github.com:jtleek/DSM.git")
pushed_at: DateTime("2013-08-06T17:36:58")
created_at: DateTime("2013-08-08T13:20:40")
updated_at: DateTime("2023-09-08T16:41:19")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "EDA-Project"
full_name: "jtleek/EDA-Project"
language: "JavaScript"
default_branch: "master"
owner: Owner("jtleek")
id: 8730097
size: 91008
```

stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/EDA-Project")
html_url: URI("https://github.com/jtleek/EDA-Project")
clone_url: URI("https://github.com/jtleek/EDA-Project.git")
ssh_url: URI("git@github.com:jtleek/EDA-Project.git")
pushed_at: DateTime("2013-03-10T23:17:30")
created_at: DateTime("2013-03-12T14:46:24")
updated_at: DateTime("2022-05-26T23:29:44")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "escalatr"
full_name: "jtleek/escalatr"
description: "A package for making R markdown websites. "
default_branch: "master"
owner: Owner("jtleek")
id: 266150752
size: 0
forks_count: 0
stargazers_count: 7
watchers_count: 7
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/escalatr")
html_url: URI("https://github.com/jtleek/escalatr")
clone_url: URI("https://github.com/jtleek/escalatr.git")
ssh_url: URI("git@github.com:jtleek/escalatr.git")
pushed_at: DateTime("2020-05-22T15:59:12")
created_at: DateTime("2020-05-22T15:59:10")
updated_at: DateTime("2023-09-08T18:07:50")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "euclideo"
full_name: "jtleek/euclideo"
description: "Game for my kids!"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 262636504
size: 106
forks_count: 3
stargazers_count: 0
watchers_count: 0
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/euclideo")
html_url: URI("https://github.com/jtleek/euclideo")
clone_url: URI("https://github.com/jtleek/euclideo.git")
ssh_url: URI("git@github.com:jtleek/euclideo.git")
pushed_at: DateTime("2020-05-17T19:33:13")
created_at: DateTime("2020-05-09T18:43:24")
updated_at: DateTime("2021-08-03T19:09:57")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
are Union{Nothing, T}):

name: "firstpaper"
full_name: "jtleek/firstpaper"
default_branch: "master"
owner: Owner("jtleek")
id: 42663787
size: 45
forks_count: 60
stargazers_count: 287
watchers_count: 287
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/firstpaper")
html_url: URI("https://github.com/jtleek/firstpaper")
clone_url: URI("https://github.com/jtleek/firstpaper.git")
ssh_url: URI("git@github.com:jtleek/firstpaper.git")
pushed_at: DateTime("2016-08-03T17:42:23")
created_at: DateTime("2015-09-17T15:17:01")
updated_at: DateTime("2024-01-08T17:11:51")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("GPL-2.0")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "futureofstats"
full_name: "jtleek/futureofstats"
description: "Take Homes from the Unconference on the Future of Statistics #futureof
stats"
default_branch: "master"
owner: Owner("jtleek")
id: 14590772
size: 147
forks_count: 22
stargazers_count: 33
watchers_count: 33
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/futureofstats")
html_url: URI("https://github.com/jtleek/futureofstats")
clone_url: URI("https://github.com/jtleek/futureofstats.git")
ssh_url: URI("git@github.com:jtleek/futureofstats.git")
pushed_at: DateTime("2013-11-22T03:06:56")
created_at: DateTime("2013-11-21T15:07:59")
updated_at: DateTime("2022-06-16T21:35:54")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "gcd"
full_name: "jtleek/gcd"
description: "Getting and cleaning data reboot"
default_branch: "master"
owner: Owner("jtleek")
id: 136025738
size: 0
forks_count: 1
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/gcd")
html_url: URI("https://github.com/jtleek/gcd")
clone_url: URI("https://github.com/jtleek/gcd.git")
ssh_url: URI("git@github.com:jtleek/gcd.git")
pushed_at: DateTime("2018-06-04T13:01:40")
_at: DateTime("   )4T13:01:18")

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

updated_at: DateTime("2023-09-08T17:41:19")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "gdspi"
full_name: "jtleek/gdspi"
description: "Genomic Data Science for PIs Curriculum Outline"
default_branch: "master"
owner: Owner("jtleek")
id: 213703490
size: 2
forks_count: 4
stargazers_count: 9
watchers_count: 9
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/gdspi")
html_url: URI("https://github.com/jtleek/gdspi")
clone_url: URI("https://github.com/jtleek/gdspi.git")
ssh_url: URI("git@github.com:jtleek/gdspi.git")
pushed_at: DateTime("2019-10-08T17:26:27")
created_at: DateTime("2019-10-08T17:06:34")
updated_at: DateTime("2022-10-01T16:01:09")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "genomicspapers"
full_name: "jtleek/genomicspapers"
description: "The Leek group guide to genomics papers"
default_branch: "master"
owner: Owner("jtleek")
id: 24905212
size: 9
forks_count: 173
stargazers_count: 454
watchers_count: 454
open_issues_count: 3
url: URI("https://api.github.com/repos/jtleek/genomicspapers")
html_url: URI("https://github.com/jtleek/genomicspapers")
clone_url: URI("https://github.com/jtleek/genomicspapers.git")
ssh_url: URI("git@github.com:jtleek/genomicspapers.git")
pushed_at: DateTime("2018-11-05T18:49:23")
created_at: DateTime("2014-10-07T18:32:10")
updated_at: DateTime("2024-01-23T15:35:11")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "genstats"
full_name: "jtleek/genstats"
description: "Statistics course for JHU Genomic Data Science Sequence"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 39025267
size: 15870
forks_count: 216
stargazers_count: 142

```
open_issues_count: 4
url: URI("https://api.github.com/repos/jtleek/genstats")
html_url: URI("https://github.com/jtleek/genstats")
clone_url: URI("https://github.com/jtleek/genstats.git")
ssh_url: URI("git@github.com:jtleek/genstats.git")
pushed_at: DateTime("2015-09-19T13:32:57")
created_at: DateTime("2015-07-13T16:55:21")
updated_at: DateTime("2023-10-26T02:53:20")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
license: License("NOASSERTION")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "genstats_site"
full_name: "jtleek/genstats_site"
description: "Site for Genomic Data Science Class"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 41094217
size: 41702
forks_count: 11
stargazers_count: 16
watchers_count: 16
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/genstats_site")
html_url: URI("https://github.com/jtleek/genstats_site")
clone_url: URI("https://github.com/jtleek/genstats_site.git")
ssh_url: URI("git@github.com:jtleek/genstats_site.git")
pushed_at: DateTime("2015-09-07T13:54:52")
created_at: DateTime("2015-08-20T12:23:21")
updated_at: DateTime("2023-10-05T18:12:28")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "github-slideshow"
full_name: "jtleek/github-slideshow"
description: "A robot powered training repository :robot:"
language: "Ruby"
default_branch: "master"
owner: Owner("jtleek")
id: 289258980
size: 3515
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/github-slideshow")
html_url: URI("https://github.com/jtleek/github-slideshow")
clone_url: URI("https://github.com/jtleek/github-slideshow.git")
ssh_url: URI("git@github.com:jtleek/github-slideshow.git")
homepage: URI("https://lab.github.com/githubtraining/introduction-to-github")
pushed_at: DateTime("2020-08-21T11:59:29")
created_at: DateTime("2020-08-21T11:59:23")
updated_at: DateTime("2020-08-21T11:59:33")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("MIT")
```

      fork: false, Repo (all fields are Union{Nothing, T}):
      name: "googleCite"
      full_name: "jtleek/googleCite"
      description: "googleCite is a function for creating a wordcloud of your google schol
ar citations page. "
      default_branch: "master"
      owner: Owner("jtleek")
      id: 12563551
      size: 116
      forks_count: 6
      stargazers_count: 4
      watchers_count: 4
      open_issues_count: 1
      url: URI("https://api.github.com/repos/jtleek/googleCite")
      html_url: URI("https://github.com/jtleek/googleCite")
      clone_url: URI("https://github.com/jtleek/googleCite.git")
      ssh_url: URI("git@github.com:jtleek/googleCite.git")
      pushed_at: DateTime("2013-09-03T14:10:06")
      created_at: DateTime("2013-09-03T12:48:28")
      updated_at: DateTime("2023-09-08T16:41:50")
      has_issues: true
      has_wiki: true
      has_downloads: true
      has_pages: false
      private: false
      fork: false, Repo (all fields are Union{Nothing, T}):
      name: "graduate"
      full_name: "jtleek/graduate"
      default_branch: "master"
      owner: Owner("jtleek")
      id: 6582536
      size: 1199
      forks_count: 6
      stargazers_count: 3
      watchers_count: 3
      open_issues_count: 0
      url: URI("https://api.github.com/repos/jtleek/graduate")
      html_url: URI("https://github.com/jtleek/graduate")
      clone_url: URI("https://github.com/jtleek/graduate.git")
      ssh_url: URI("git@github.com:jtleek/graduate.git")
      pushed_at: DateTime("2012-11-07T05:43:21")
      created_at: DateTime("2012-11-07T16:43:18")
      updated_at: DateTime("2021-11-16T11:47:23")
      has_issues: false
      has_wiki: true
      has_downloads: true
      has_pages: false
      private: false
      fork: false, Repo (all fields are Union{Nothing, T}):
      name: "healthvis"
      full_name: "jtleek/healthvis"
      description: "An Interactive Health Visualization Package"
      language: "Python"
      default_branch: "master"
      owner: Owner("jtleek")
      id: 6661008
      size: 1067
      forks_count: 8
      stargazers_count: 8
      watchers_count: 8
      open_issues_count: 0
      url: URI("https://api.github.com/repos/jtleek/healthvis")
      html_url: URI("https://github.com/jtleek/healthvis")
      clone_url: URI("https://github.com/jtleek/healthvis.git")
      ssh_url: URI("git@github.com:jtleek/healthvis.git")

created_at: DateTime("2012-11-12T21:35:30")
updated_at: DateTime("2017-01-17T21:09:20")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "hr-in-ds"
full_name: "jtleek/hr-in-ds"
description: "A collaborative white paper on challenges and opportunities with human
resources for data science positions"
default_branch: "main"
owner: Owner("jtleek")
id: 346764495
size: 0
forks_count: 0
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/hr-in-ds")
html_url: URI("https://github.com/jtleek/hr-in-ds")
clone_url: URI("https://github.com/jtleek/hr-in-ds.git")
ssh_url: URI("git@github.com:jtleek/hr-in-ds.git")
pushed_at: DateTime("2021-03-11T16:22:19")
created_at: DateTime("2021-03-11T16:22:19")
updated_at: DateTime("2023-09-08T18:19:39")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "hyde"
full_name: "jtleek/hyde"
description: "A brazen two-column theme for Jekyll."
language: "CSS"
default_branch: "master"
owner: Owner("jtleek")
id: 19133476
size: 447
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/hyde")
html_url: URI("https://github.com/jtleek/hyde")
clone_url: URI("https://github.com/jtleek/hyde.git")
ssh_url: URI("git@github.com:jtleek/hyde.git")
homepage: URI("http://hyde.getpoole.com")
pushed_at: DateTime("2014-03-16T03:23:52")
created_at: DateTime("2014-04-25T03:59:31")
updated_at: DateTime("2014-04-25T03:59:31")
has_issues: false
has_wiki: false
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "iap"
full_name: "jtleek/iap"
description: "This is the repository for the inference after prediction package"
language: "R"
default_branch: "master"

id: 198452807
size: 2
forks_count: 1
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/iap")
html_url: URI("https://github.com/jtleek/iap")
clone_url: URI("https://github.com/jtleek/iap.git")
ssh_url: URI("git@github.com:jtleek/iap.git")
pushed_at: DateTime("2019-07-23T21:01:02")
created_at: DateTime("2019-07-23T14:58:51")
updated_at: DateTime("2021-01-23T21:13:50")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "inclassfeb62014"
full_name: "jtleek/inclassfeb62014"
description: "In class project repo"
language: "Shell"
default_branch: "master"
owner: Owner("jtleek")
id: 16584923
size: 6248
forks_count: 5
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/inclassfeb62014")
html_url: URI("https://github.com/jtleek/inclassfeb62014")
clone_url: URI("https://github.com/jtleek/inclassfeb62014.git")
ssh_url: URI("git@github.com:jtleek/inclassfeb62014.git")
pushed_at: DateTime("2014-02-18T10:34:32")
created_at: DateTime("2014-02-06T16:13:11")
updated_at: DateTime("2014-06-05T17:59:46")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "intro-ml-2018"
full_name: "jtleek/intro-ml-2018"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 155591401
size: 10172
forks_count: 4
stargazers_count: 8
watchers_count: 8
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/intro-ml-2018")
html_url: URI("https://github.com/jtleek/intro-ml-2018")
clone_url: URI("https://github.com/jtleek/intro-ml-2018.git")
ssh_url: URI("git@github.com:jtleek/intro-ml-2018.git")
pushed_at: DateTime("2018-11-05T18:07:02")
created_at: DateTime("2018-10-31T16:42:13")
updated_at: DateTime("2023-09-08T17:46:44")
has_issues: true
has_wiki: true

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jhsph-irb-research-plan-template"
full_name: "jtleek/jhsph-irb-research-plan-template"
description: "JHSPH IRB form "
default_branch: "master"
owner: Owner("jtleek")
id: 257074313
size: 43
forks_count: 1
stargazers_count: 4
watchers_count: 4
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/jhsph-irb-research-plan-template")
html_url: URI("https://github.com/jtleek/jhsph-irb-research-plan-template")
clone_url: URI("https://github.com/jtleek/jhsph-irb-research-plan-template.git")
ssh_url: URI("git@github.com:jtleek/jhsph-irb-research-plan-template.git")
pushed_at: DateTime("2020-09-14T15:50:42")
created_at: DateTime("2020-04-19T18:34:51")
updated_at: DateTime("2023-09-08T18:06:12")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jhsph753"
full_name: "jtleek/jhsph753"
description: "Web page for JHSPH Advanced Methods/Applied Statistics"
language: "JavaScript"
default_branch: "master"
owner: Owner("jtleek")
id: 7745123
size: 27878
forks_count: 5
stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/jhsph753")
html_url: URI("https://github.com/jtleek/jhsph753")
clone_url: URI("https://github.com/jtleek/jhsph753.git")
ssh_url: URI("git@github.com:jtleek/jhsph753.git")
pushed_at: DateTime("2013-05-16T11:43:34")
created_at: DateTime("2013-01-22T03:06:39")
updated_at: DateTime("2020-01-22T14:39:20")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jhsph753and4"
full_name: "jtleek/jhsph753and4"
description: "Class github repository for 751 and 2; doctoral classes in the Departm
ent of Biostatistics at Johns Hopkins"
language: "JavaScript"
default_branch: "master"
owner: Owner("jtleek")
id: 15639612
size: 185443
forks_count: 33
stargazers_count: 14
watchers_count: 14
open_issues_count: 0
```
```
om/repos/jtleek/jhsph753and4")
```

html_url: URI("https://github.com/jtleek/jhsph753and4")
clone_url: URI("https://github.com/jtleek/jhsph753and4.git")
ssh_url: URI("git@github.com:jtleek/jhsph753and4.git")
pushed_at: DateTime("2014-05-15T14:22:41")
created_at: DateTime("2014-01-04T21:06:44")
updated_at: DateTime("2022-04-14T23:52:19")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jhudash"
full_name: "jtleek/jhudash"
description: "A repository for all things DaSH"
default_branch: "master"
owner: Owner("jtleek")
id: 42834789
size: 381
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/jhudash")
html_url: URI("https://github.com/jtleek/jhudash")
clone_url: URI("https://github.com/jtleek/jhudash.git")
ssh_url: URI("git@github.com:jtleek/jhudash.git")
pushed_at: DateTime("2015-09-21T10:35:40")
created_at: DateTime("2015-09-21T00:11:42")
updated_at: DateTime("2023-09-08T17:01:53")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
license: License("CC0-1.0")
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "jhudash-refugee"
full_name: "jtleek/jhudash-refugee"
description: "Code to collect data for the #jhudash refugee project"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 42873969
size: 23206
forks_count: 3
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/jhudash-refugee")
html_url: URI("https://github.com/jtleek/jhudash-refugee")
clone_url: URI("https://github.com/jtleek/jhudash-refugee.git")
ssh_url: URI("git@github.com:jtleek/jhudash-refugee.git")
pushed_at: DateTime("2015-11-09T20:28:11")
created_at: DateTime("2015-09-21T15:08:47")
updated_at: DateTime("2021-02-12T13:44:44")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jobs"
full_name: "jtleek/jobs"
description: "Jobs"

```
owner: Owner("jtleek")
id: 320397498
size: 24
forks_count: 2
stargazers_count: 10
watchers_count: 10
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/jobs")
html_url: URI("https://github.com/jtleek/jobs")
clone_url: URI("https://github.com/jtleek/jobs.git")
ssh_url: URI("git@github.com:jtleek/jobs.git")
pushed_at: DateTime("2022-10-01T01:58:12")
created_at: DateTime("2020-12-10T21:45:41")
updated_at: DateTime("2022-09-04T00:45:24")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "jtleek.github.io"
full_name: "jtleek/jtleek.github.io"
description: "Website"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 19133794
size: 12721
forks_count: 68
stargazers_count: 15
watchers_count: 15
open_issues_count: 6
url: URI("https://api.github.com/repos/jtleek/jtleek.github.io")
html_url: URI("https://github.com/jtleek/jtleek.github.io")
clone_url: URI("https://github.com/jtleek/jtleek.github.io.git")
ssh_url: URI("git@github.com:jtleek/jtleek.github.io.git")
pushed_at: DateTime("2022-05-15T17:00:27")
created_at: DateTime("2014-04-25T04:14:45")
updated_at: DateTime("2021-11-21T14:36:43")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "langmead-lab"
full_name: "jtleek/langmead-lab"
description: "Publicly-visible Langmead Lab resources"
default_branch: "master"
owner: Owner("jtleek")
id: 36437212
size: 61
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/langmead-lab")
html_url: URI("https://github.com/jtleek/langmead-lab")
clone_url: URI("https://github.com/jtleek/langmead-lab.git")
ssh_url: URI("git@github.com:jtleek/langmead-lab.git")
pushed_at: DateTime("2015-05-28T01:33:53")
created_at: DateTime("2015-05-28T12:35:48")
updated_at: DateTime("2015-05-28T12:24:25")
has_issues: false
has_wiki: true
```

```
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "leah"
full_name: "jtleek/leah"
description: "blah de blah"
default_branch: "master"
owner: Owner("jtleek")
id: 17446438
size: 136
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/leah")
html_url: URI("https://github.com/jtleek/leah")
clone_url: URI("https://github.com/jtleek/leah.git")
ssh_url: URI("git@github.com:jtleek/leah.git")
pushed_at: DateTime("2014-03-05T16:18:30")
created_at: DateTime("2014-03-05T16:11:46")
updated_at: DateTime("2014-03-05T16:18:30")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "leanpubdemorepo"
full_name: "jtleek/leanpubdemorepo"
default_branch: "master"
owner: Owner("jtleek")
id: 177818426
size: 0
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/leanpubdemorepo")
html_url: URI("https://github.com/jtleek/leanpubdemorepo")
clone_url: URI("https://github.com/jtleek/leanpubdemorepo.git")
ssh_url: URI("git@github.com:jtleek/leanpubdemorepo.git")
pushed_at: DateTime("2019-03-26T15:36:30")
created_at: DateTime("2019-03-26T15:36:29")
updated_at: DateTime("2019-05-21T15:01:58")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "leekasso"
full_name: "jtleek/leekasso"
description: "Code for comparing the top 10 predictors to the lasso/debiased lasso"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 15723485
size: 357
forks_count: 9
stargazers_count: 12
watchers_count: 12
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/leekasso")
html_url: URI("https://github.com/jtleek/leekasso")
clone_url: URI("https://github.com/jtleek/leekasso.git")
tleek/leekasso.git")
```

```
pushed_at: DateTime("2014-01-23T11:49:37")
created_at: DateTime("2014-01-08T02:40:21")
updated_at: DateTime("2024-01-03T14:11:47")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "Module1"
full_name: "jtleek/Module1"
description: "Teaching material for Summer Institute in Statistics for Big Data Modu
le 1."
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 38310798
size: 3279
forks_count: 0
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/Module1")
html_url: URI("https://github.com/jtleek/Module1")
clone_url: URI("https://github.com/jtleek/Module1.git")
ssh_url: URI("git@github.com:jtleek/Module1.git")
homepage: URI("")
pushed_at: DateTime("2015-07-06T14:33:00")
created_at: DateTime("2015-06-30T13:28:23")
updated_at: DateTime("2022-05-26T23:30:29")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: true
license: License("NOASSERTION")
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "modules"
full_name: "jtleek/modules"
language: "JavaScript"
default_branch: "master"
owner: Owner("jtleek")
id: 11378145
size: 513853
forks_count: 899
stargazers_count: 96
watchers_count: 96
open_issues_count: 7
url: URI("https://api.github.com/repos/jtleek/modules")
html_url: URI("https://github.com/jtleek/modules")
clone_url: URI("https://github.com/jtleek/modules.git")
ssh_url: URI("git@github.com:jtleek/modules.git")
pushed_at: DateTime("2017-01-16T03:10:38")
created_at: DateTime("2013-07-12T21:09:41")
updated_at: DateTime("2023-07-23T05:36:49")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "MOOCtalk"
full_name: "jtleek/MOOCtalk"
language: "JavaScript"
default_branch: "master"
```

```
id: 17711648
size: 148575
forks_count: 1
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/MOOCtalk")
html_url: URI("https://github.com/jtleek/MOOCtalk")
clone_url: URI("https://github.com/jtleek/MOOCtalk.git")
ssh_url: URI("git@github.com:jtleek/MOOCtalk.git")
pushed_at: DateTime("2014-05-20T18:04:34")
created_at: DateTime("2014-03-13T13:53:05")
updated_at: DateTime("2014-06-05T17:59:59")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "mypkg"
full_name: "jtleek/mypkg"
description: "This is a test package for advdatasci"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 47994844
size: 5
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/mypkg")
html_url: URI("https://github.com/jtleek/mypkg")
clone_url: URI("https://github.com/jtleek/mypkg.git")
ssh_url: URI("git@github.com:jtleek/mypkg.git")
pushed_at: DateTime("2015-12-16T19:10:56")
created_at: DateTime("2015-12-14T18:56:54")
updated_at: DateTime("2015-12-14T19:21:37")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "newproject"
full_name: "jtleek/newproject"
description: "This is my new project."
default_branch: "master"
owner: Owner("jtleek")
id: 101914268
size: 2
forks_count: 94
stargazers_count: 1
watchers_count: 1
open_issues_count: 73
url: URI("https://api.github.com/repos/jtleek/newproject")
html_url: URI("https://github.com/jtleek/newproject")
clone_url: URI("https://github.com/jtleek/newproject.git")
ssh_url: URI("git@github.com:jtleek/newproject.git")
pushed_at: DateTime("2023-08-08T04:37:22")
created_at: DateTime("2017-08-30T18:14:23")
updated_at: DateTime("2021-11-21T14:36:49")
has_issues: true
has_wiki: true
has_downloads: true
```

```
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "new_project"
full_name: "jtleek/new_project"
description: "This is my new project"
default_branch: "master"
owner: Owner("jtleek")
id: 101907019
size: 0
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/new_project")
html_url: URI("https://github.com/jtleek/new_project")
clone_url: URI("https://github.com/jtleek/new_project.git")
ssh_url: URI("git@github.com:jtleek/new_project.git")
pushed_at: DateTime("2017-08-30T16:58:22")
created_at: DateTime("2017-08-30T16:58:22")
updated_at: DateTime("2017-08-30T16:58:22")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "papr"
full_name: "jtleek/papr"
description: "Paper app"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 69924967
size: 24050
forks_count: 16
stargazers_count: 19
watchers_count: 19
open_issues_count: 3
url: URI("https://api.github.com/repos/jtleek/papr")
html_url: URI("https://github.com/jtleek/papr")
clone_url: URI("https://github.com/jtleek/papr.git")
ssh_url: URI("git@github.com:jtleek/papr.git")
pushed_at: DateTime("2019-05-15T17:37:12")
created_at: DateTime("2016-10-04T01:22:46")
updated_at: DateTime("2023-11-24T15:34:03")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "polyester"
full_name: "jtleek/polyester"
description: "RNA-seq read simulator"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 20548045
size: 877
forks_count: 2
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/polyester")
html_url: URI("https://github.com/jtleek/polyester")
.com/jtleek/polyester.git")
```

ssh_url: URI("git@github.com:jtleek/polyester.git")
pushed_at: DateTime("2014-06-18T23:21:50")
created_at: DateTime("2014-06-06T01:57:33")
updated_at: DateTime("2014-06-19T18:06:42")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "portfolio"
full_name: "jtleek/portfolio"
description: "This is my Data Science Specialization Portfolio"
default_branch: "master"
owner: Owner("jtleek")
id: 16103392
size: 120
forks_count: 2
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/portfolio")
html_url: URI("https://github.com/jtleek/portfolio")
clone_url: URI("https://github.com/jtleek/portfolio.git")
ssh_url: URI("git@github.com:jtleek/portfolio.git")
pushed_at: DateTime("2014-01-21T13:35:40")
created_at: DateTime("2014-01-21T13:35:16")
updated_at: DateTime("2017-08-27T17:58:39")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "practicecourse"
full_name: "jtleek/practicecourse"
description: "Practice course for CDS"
default_branch: "master"
owner: Owner("jtleek")
id: 113584733
size: 33
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/practicecourse")
html_url: URI("https://github.com/jtleek/practicecourse")
clone_url: URI("https://github.com/jtleek/practicecourse.git")
ssh_url: URI("git@github.com:jtleek/practicecourse.git")
pushed_at: DateTime("2017-12-08T14:52:26")
created_at: DateTime("2017-12-08T14:48:09")
updated_at: DateTime("2023-09-08T17:33:42")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "protocols"
full_name: "jtleek/protocols"
description: "This will be a directory of lab analysis protocols."
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 23202748

```
forks_count: 5
stargazers_count: 13
watchers_count: 13
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/protocols")
html_url: URI("https://github.com/jtleek/protocols")
clone_url: URI("https://github.com/jtleek/protocols.git")
ssh_url: URI("git@github.com:jtleek/protocols.git")
pushed_at: DateTime("2015-07-23T14:54:03")
created_at: DateTime("2014-08-21T20:55:43")
updated_at: DateTime("2023-09-08T16:49:59")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "R-Markdown-Parallax"
full_name: "jtleek/R-Markdown-Parallax"
description: "How to use parallax on your R Markdown output"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 192386198
size: 2113
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/R-Markdown-Parallax")
html_url: URI("https://github.com/jtleek/R-Markdown-Parallax")
clone_url: URI("https://github.com/jtleek/R-Markdown-Parallax.git")
ssh_url: URI("git@github.com:jtleek/R-Markdown-Parallax.git")
pushed_at: DateTime("2018-04-26T09:29:26")
created_at: DateTime("2019-06-17T16:58:00")
updated_at: DateTime("2023-09-08T17:54:56")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "rdsmGeneSig"
full_name: "jtleek/rdsmGeneSig"
description: "A deterministic statistical machine (http://simplystatistics.org/2012/
08/27/a-deterministic-statistical-machine/) for calculating and validating a gene sign
ature."
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 12134722
size: 116
forks_count: 3
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/rdsmGeneSig")
html_url: URI("https://github.com/jtleek/rdsmGeneSig")
clone_url: URI("https://github.com/jtleek/rdsmGeneSig.git")
ssh_url: URI("git@github.com:jtleek/rdsmGeneSig.git")
pushed_at: DateTime("2013-08-15T15:03:23")
created_at: DateTime("2013-08-15T13:16:01")
updated_at: DateTime("2022-05-26T23:29:12")
has_issues: true
has_wiki: true
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "readingpapers"
full_name: "jtleek/readingpapers"
description: "A guide to reading scientific papers"
default_branch: "master"
owner: Owner("jtleek")
id: 51390717
size: 4
forks_count: 89
stargazers_count: 445
watchers_count: 445
open_issues_count: 2
url: URI("https://api.github.com/repos/jtleek/readingpapers")
html_url: URI("https://github.com/jtleek/readingpapers")
clone_url: URI("https://github.com/jtleek/readingpapers.git")
ssh_url: URI("git@github.com:jtleek/readingpapers.git")
pushed_at: DateTime("2020-02-02T19:53:31")
created_at: DateTime("2016-02-09T18:56:30")
updated_at: DateTime("2024-01-24T00:15:20")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "replication_paper"
full_name: "jtleek/replication_paper"
description: "Replication paper"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 41891336
size: 1262
forks_count: 7
stargazers_count: 3
watchers_count: 3
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/replication_paper")
html_url: URI("https://github.com/jtleek/replication_paper")
clone_url: URI("https://github.com/jtleek/replication_paper.git")
ssh_url: URI("git@github.com:jtleek/replication_paper.git")
pushed_at: DateTime("2016-02-08T20:44:17")
created_at: DateTime("2015-09-04T00:59:36")
updated_at: DateTime("2017-06-24T17:13:58")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "reviews"
full_name: "jtleek/reviews"
description: "Writing reviews of academic papers"
default_branch: "master"
owner: Owner("jtleek")
id: 13788992
size: 154
forks_count: 105
stargazers_count: 447
watchers_count: 447
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/reviews")
html_url: URI("https://github.com/jtleek/reviews")
.com/jtleek/reviews.git")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
ssh_url: URI("git@github.com:jtleek/reviews.git")
pushed_at: DateTime("2015-08-12T15:10:48")
created_at: DateTime("2013-10-22T23:47:38")
updated_at: DateTime("2024-01-24T00:15:21")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("GPL-2.0")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "rfitbit"
full_name: "jtleek/rfitbit"
description: "An R package to download and play with fitbit data"
default_branch: "master"
owner: Owner("jtleek")
id: 15532926
size: 120
forks_count: 4
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/rfitbit")
html_url: URI("https://github.com/jtleek/rfitbit")
clone_url: URI("https://github.com/jtleek/rfitbit.git")
ssh_url: URI("git@github.com:jtleek/rfitbit.git")
pushed_at: DateTime("2013-12-30T18:18:18")
created_at: DateTime("2013-12-30T17:43:29")
updated_at: DateTime("2014-06-05T17:56:17")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "rmd4edu"
full_name: "jtleek/rmd4edu"
description: "A fresh batch of R Markdown templates"
default_branch: "master"
owner: Owner("jtleek")
id: 273780186
size: 12873
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/rmd4edu")
html_url: URI("https://github.com/jtleek/rmd4edu")
clone_url: URI("https://github.com/jtleek/rmd4edu.git")
ssh_url: URI("git@github.com:jtleek/rmd4edu.git")
pushed_at: DateTime("2019-09-03T14:08:12")
created_at: DateTime("2020-06-20T20:33:17")
updated_at: DateTime("2023-09-08T18:09:02")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "robotjeff"
full_name: "jtleek/robotjeff"
description: "This is the Shiny app to make robot Jeff talk"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

size: 2
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/robotjeff")
html_url: URI("https://github.com/jtleek/robotjeff")
clone_url: URI("https://github.com/jtleek/robotjeff.git")
ssh_url: URI("git@github.com:jtleek/robotjeff.git")
pushed_at: DateTime("2017-06-13T04:34:32")
created_at: DateTime("2017-06-13T04:17:05")
updated_at: DateTime("2017-06-13T04:20:10")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "rpackages"
full_name: "jtleek/rpackages"
description: "R package development - the Leek group way!"
default_branch: "master"
owner: Owner("jtleek")
id: 12931390
size: 958
forks_count: 285
stargazers_count: 503
watchers_count: 503
open_issues_count: 4
url: URI("https://api.github.com/repos/jtleek/rpackages")
html_url: URI("https://github.com/jtleek/rpackages")
clone_url: URI("https://github.com/jtleek/rpackages.git")
ssh_url: URI("git@github.com:jtleek/rpackages.git")
pushed_at: DateTime("2024-01-10T21:20:39")
created_at: DateTime("2013-09-18T18:38:33")
updated_at: DateTime("2024-01-22T08:50:03")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "runs"
full_name: "jtleek/runs"
description: "Scripts for reproducing analyses of large RNA-seq datasets"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 48818093
size: 20409
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/runs")
html_url: URI("https://github.com/jtleek/runs")
clone_url: URI("https://github.com/jtleek/runs.git")
ssh_url: URI("git@github.com:jtleek/runs.git")
homepage: URI("")
pushed_at: DateTime("2015-12-29T21:32:24")
created_at: DateTime("2015-12-30T20:22:37")
updated_at: DateTime("2015-12-30T20:22:40")
has_issues: false
has_wiki: true
has_downloads: true

```
license: License("MIT")
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "simplystats"
full_name: "jtleek/simplystats"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 28644442
size: 175
forks_count: 7
stargazers_count: 6
watchers_count: 6
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/simplystats")
html_url: URI("https://github.com/jtleek/simplystats")
clone_url: URI("https://github.com/jtleek/simplystats.git")
ssh_url: URI("git@github.com:jtleek/simplystats.git")
pushed_at: DateTime("2015-01-16T04:43:03")
created_at: DateTime("2014-12-30T20:32:06")
updated_at: DateTime("2021-11-22T14:07:37")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "simplystats_analysis"
full_name: "jtleek/simplystats_analysis"
description: "Wrapping up!"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 140899822
size: 3
forks_count: 0
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/simplystats_analysis")
html_url: URI("https://github.com/jtleek/simplystats_analysis")
clone_url: URI("https://github.com/jtleek/simplystats_analysis.git")
ssh_url: URI("git@github.com:jtleek/simplystats_analysis.git")
pushed_at: DateTime("2018-07-13T23:44:21")
created_at: DateTime("2018-07-13T23:03:03")
updated_at: DateTime("2022-03-15T19:32:33")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "simplystats_posts_sisbid"
full_name: "jtleek/simplystats_posts_sisbid"
description: "This is the simply stats posts for SISBID"
default_branch: "master"
owner: Owner("jtleek")
id: 140879448
size: 1568
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/simplystats_posts_sisbid")
html_url: URI("https://github.com/jtleek/simplystats_posts_sisbid")
```
.com/jtleek/simplystats_posts_sisbid.git")

```
ssh_url: URI("git@github.com:jtleek/simplystats_posts_sisbid.git")
pushed_at: DateTime("2018-07-13T18:20:28")
created_at: DateTime("2018-07-13T18:15:39")
updated_at: DateTime("2018-07-13T18:20:33")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sisbid-rstudio"
full_name: "jtleek/sisbid-rstudio"
default_branch: "master"
owner: Owner("jtleek")
id: 63102010
size: 0
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sisbid-rstudio")
html_url: URI("https://github.com/jtleek/sisbid-rstudio")
clone_url: URI("https://github.com/jtleek/sisbid-rstudio.git")
ssh_url: URI("git@github.com:jtleek/sisbid-rstudio.git")
pushed_at: DateTime("2016-07-11T21:07:09")
created_at: DateTime("2016-07-11T21:03:02")
updated_at: DateTime("2023-09-08T17:12:32")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sisbidday2"
full_name: "jtleek/sisbidday2"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 63170038
size: 1268
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sisbidday2")
html_url: URI("https://github.com/jtleek/sisbidday2")
clone_url: URI("https://github.com/jtleek/sisbidday2.git")
ssh_url: URI("git@github.com:jtleek/sisbidday2.git")
pushed_at: DateTime("2016-07-13T15:25:33")
created_at: DateTime("2016-07-12T15:33:04")
updated_at: DateTime("2016-07-12T15:35:55")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sisbid_2019_practice"
full_name: "jtleek/sisbid_2019_practice"
description: "This is my practice repo "
default_branch: "master"
owner: Owner("jtleek")
id: 197073776
size: 2
forks_count: 0
```

watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sisbid_2019_practice")
html_url: URI("https://github.com/jtleek/sisbid_2019_practice")
clone_url: URI("https://github.com/jtleek/sisbid_2019_practice.git")
ssh_url: URI("git@github.com:jtleek/sisbid_2019_practice.git")
pushed_at: DateTime("2019-07-15T21:53:01")
created_at: DateTime("2019-07-15T21:17:55")
updated_at: DateTime("2019-07-15T21:53:03")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sisbid_day2"
full_name: "jtleek/sisbid_day2"
description: "This is my first Github project."
default_branch: "master"
owner: Owner("jtleek")
id: 140734891
size: 4
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/sisbid_day2")
html_url: URI("https://github.com/jtleek/sisbid_day2")
clone_url: URI("https://github.com/jtleek/sisbid_day2.git")
ssh_url: URI("git@github.com:jtleek/sisbid_day2.git")
pushed_at: DateTime("2018-07-12T17:01:21")
created_at: DateTime("2018-07-12T15:53:27")
updated_at: DateTime("2018-07-12T17:01:23")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sisbid_day2_2018"
full_name: "jtleek/sisbid_day2_2018"
description: "sisbid"
language: "HTML"
default_branch: "master"
owner: Owner("jtleek")
id: 140739794
size: 244
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sisbid_day2_2018")
html_url: URI("https://github.com/jtleek/sisbid_day2_2018")
clone_url: URI("https://github.com/jtleek/sisbid_day2_2018.git")
ssh_url: URI("git@github.com:jtleek/sisbid_day2_2018.git")
pushed_at: DateTime("2018-07-12T16:56:01")
created_at: DateTime("2018-07-12T16:38:13")
updated_at: DateTime("2018-07-12T16:56:02")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "sisg"

description: "SISG Module 6"
language: "HTML"
default_branch: "gh-pages"
owner: Owner("jtleek")
id: 38813928
size: 963
forks_count: 1
stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sisg")
html_url: URI("https://github.com/jtleek/sisg")
clone_url: URI("https://github.com/jtleek/sisg.git")
ssh_url: URI("git@github.com:jtleek/sisg.git")
pushed_at: DateTime("2015-07-10T15:26:38")
created_at: DateTime("2015-07-09T10:24:43")
updated_at: DateTime("2015-07-09T21:05:59")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
license: License("NOASSERTION")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "slipper"
full_name: "jtleek/slipper"
description: "Tidy and easy bootstrapping"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 75975252
size: 74
forks_count: 12
stargazers_count: 119
watchers_count: 119
open_issues_count: 2
url: URI("https://api.github.com/repos/jtleek/slipper")
html_url: URI("https://github.com/jtleek/slipper")
clone_url: URI("https://github.com/jtleek/slipper.git")
ssh_url: URI("git@github.com:jtleek/slipper.git")
pushed_at: DateTime("2017-10-05T18:12:29")
created_at: DateTime("2016-12-08T21:08:11")
updated_at: DateTime("2024-01-24T22:26:21")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "software"
full_name: "jtleek/software"
description: "Leek group software"
default_branch: "main"
owner: Owner("jtleek")
id: 320461596
size: 3
forks_count: 2
stargazers_count: 4
watchers_count: 4
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/software")
html_url: URI("https://github.com/jtleek/software")
clone_url: URI("https://github.com/jtleek/software.git")
ssh_url: URI("git@github.com:jtleek/software.git")
pushed_at: DateTime("2020-12-11T03:57:04")
11T03:55:48")

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
    updated_at: DateTime("2023-09-08T18:15:50")
    has_issues: true
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "sra-gtex"
    full_name: "jtleek/sra-gtex"
    description: "Some code I'm messing with"
    language: "R"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 57915259
    size: 3
    forks_count: 2
    stargazers_count: 0
    watchers_count: 0
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/sra-gtex")
    html_url: URI("https://github.com/jtleek/sra-gtex")
    clone_url: URI("https://github.com/jtleek/sra-gtex.git")
    ssh_url: URI("git@github.com:jtleek/sra-gtex.git")
    pushed_at: DateTime("2016-05-02T19:26:21")
    created_at: DateTime("2016-05-02T19:23:09")
    updated_at: DateTime("2016-05-02T19:23:28")
    has_issues: true
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "statements"
    full_name: "jtleek/statements"
    description: "Research & Teaching Statements in Evolution/Ecology"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 42584081
    size: 10776
    forks_count: 1
    stargazers_count: 5
    watchers_count: 5
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/statements")
    html_url: URI("https://github.com/jtleek/statements")
    clone_url: URI("https://github.com/jtleek/statements.git")
    ssh_url: URI("git@github.com:jtleek/statements.git")
    pushed_at: DateTime("2015-09-16T11:57:45")
    created_at: DateTime("2015-09-16T11:51:52")
    updated_at: DateTime("2015-09-18T15:05:44")
    has_issues: false
    has_wiki: false
    has_downloads: true
    has_pages: false
    license: License("NOASSERTION")
    private: false
    fork: true, Repo (all fields are Union{Nothing, T}):
    name: "sva"
    full_name: "jtleek/sva"
    description: "This is a read-only mirror of the Bioconductor SVN repository. Packag
  e Homepage: http://bioconductor.org/packages/devel/bioc/html/sva.html Bug Reports: htt
  ps://support.bioconductor.org/p/new/post/?tag_val=sva."
    language: "R"
    default_branch: "master"
    owner: Owner("jtleek")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

size: 210
forks_count: 3
stargazers_count: 3
watchers_count: 3
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/sva")
html_url: URI("https://github.com/jtleek/sva")
clone_url: URI("https://github.com/jtleek/sva.git")
ssh_url: URI("git@github.com:jtleek/sva.git")
homepage: URI("")
pushed_at: DateTime("2015-10-19T20:17:50")
created_at: DateTime("2015-12-30T18:18:32")
updated_at: DateTime("2022-07-17T02:29:11")
has_issues: false
has_wiki: false
has_downloads: false
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "sva-devel"
full_name: "jtleek/sva-devel"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 20508464
size: 395
forks_count: 47
stargazers_count: 28
watchers_count: 28
open_issues_count: 28
url: URI("https://api.github.com/repos/jtleek/sva-devel")
html_url: URI("https://github.com/jtleek/sva-devel")
clone_url: URI("https://github.com/jtleek/sva-devel.git")
ssh_url: URI("git@github.com:jtleek/sva-devel.git")
pushed_at: DateTime("2023-03-02T14:06:03")
created_at: DateTime("2014-06-05T02:09:23")
updated_at: DateTime("2023-07-26T19:46:16")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "svaruv"
full_name: "jtleek/svaruv"
default_branch: "master"
owner: Owner("jtleek")
id: 20369329
size: 308
forks_count: 1
stargazers_count: 2
watchers_count: 2
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/svaruv")
html_url: URI("https://github.com/jtleek/svaruv")
clone_url: URI("https://github.com/jtleek/svaruv.git")
ssh_url: URI("git@github.com:jtleek/svaruv.git")
pushed_at: DateTime("2014-06-01T10:58:10")
created_at: DateTime("2014-06-01T02:18:13")
updated_at: DateTime("2020-04-03T03:49:49")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
are Union{Nothing, T}):

```
name: "svaseq"
full_name: "jtleek/svaseq"
description: "Analysis for svaseq paper"
default_branch: "master"
owner: Owner("jtleek")
id: 21147003
size: 11451
forks_count: 11
stargazers_count: 17
watchers_count: 17
open_issues_count: 1
url: URI("https://api.github.com/repos/jtleek/svaseq")
html_url: URI("https://github.com/jtleek/svaseq")
clone_url: URI("https://github.com/jtleek/svaseq.git")
ssh_url: URI("git@github.com:jtleek/svaseq.git")
pushed_at: DateTime("2014-08-20T19:11:30")
created_at: DateTime("2014-06-24T00:31:35")
updated_at: DateTime("2023-03-05T17:32:34")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: true
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "swfdr"
full_name: "jtleek/swfdr"
description: "R code for calculating the Science-wise False Discovery Rate"
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 7056188
size: 5228
forks_count: 32
stargazers_count: 26
watchers_count: 26
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/swfdr")
html_url: URI("https://github.com/jtleek/swfdr")
clone_url: URI("https://github.com/jtleek/swfdr.git")
ssh_url: URI("git@github.com:jtleek/swfdr.git")
homepage: URI("")
pushed_at: DateTime("2017-09-04T11:12:27")
created_at: DateTime("2012-12-07T16:25:51")
updated_at: DateTime("2023-09-08T16:36:36")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "swirl"
full_name: "jtleek/swirl"
description: ":cyclone: Learn R, in R."
language: "R"
default_branch: "master"
owner: Owner("jtleek")
id: 148186170
size: 1551
forks_count: 1
stargazers_count: 1
watchers_count: 1
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/swirl")
html_url: URI("https://github.com/jtleek/swirl")
clone_url: URI("https://github.com/jtleek/swirl.git")
ssh_url: URI("git@github.com:jtleek/swirl.git")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
homepage: URI("http://swirlstats.com")
pushed_at: DateTime("2018-09-21T19:25:07")
created_at: DateTime("2018-09-10T16:32:42")
updated_at: DateTime("2023-09-08T17:44:54")
has_issues: false
has_wiki: true
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
fork: true, Repo (all fields are Union{Nothing, T}):
name: "talkguide"
full_name: "jtleek/talkguide"
description: "The Leek Group Guide to Giving Talks"
default_branch: "master"
owner: Owner("jtleek")
id: 18328460
size: 10
forks_count: 59
stargazers_count: 255
watchers_count: 255
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/talkguide")
html_url: URI("https://github.com/jtleek/talkguide")
clone_url: URI("https://github.com/jtleek/talkguide.git")
ssh_url: URI("git@github.com:jtleek/talkguide.git")
pushed_at: DateTime("2016-12-15T17:52:21")
created_at: DateTime("2014-04-01T12:22:06")
updated_at: DateTime("2024-01-21T06:28:47")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "talks"
full_name: "jtleek/talks"
description: "Slides from presentations"
default_branch: "master"
owner: Owner("jtleek")
id: 12133281
size: 23141
forks_count: 6
stargazers_count: 14
watchers_count: 14
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/talks")
html_url: URI("https://github.com/jtleek/talks")
clone_url: URI("https://github.com/jtleek/talks.git")
ssh_url: URI("git@github.com:jtleek/talks.git")
pushed_at: DateTime("2024-01-31T14:09:44")
created_at: DateTime("2013-08-15T11:52:51")
updated_at: DateTime("2023-11-24T17:27:06")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
license: License("NOASSERTION")
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testdir"
full_name: "jtleek/testdir"
description: "this is a test"
default_branch: "master"
owner: Owner("jtleek")
```

```
size: 112
forks_count: 1
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testdir")
html_url: URI("https://github.com/jtleek/testdir")
clone_url: URI("https://github.com/jtleek/testdir.git")
ssh_url: URI("git@github.com:jtleek/testdir.git")
pushed_at: DateTime("2015-07-05T13:07:11")
created_at: DateTime("2015-07-05T12:57:04")
updated_at: DateTime("2015-07-05T12:57:04")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testproject"
full_name: "jtleek/testproject"
default_branch: "master"
owner: Owner("jtleek")
id: 38634839
size: 112
forks_count: 0
stargazers_count: 0
watchers_count: 0
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testproject")
html_url: URI("https://github.com/jtleek/testproject")
clone_url: URI("https://github.com/jtleek/testproject.git")
ssh_url: URI("git@github.com:jtleek/testproject.git")
pushed_at: DateTime("2015-07-06T17:37:08")
created_at: DateTime("2015-07-06T17:34:35")
updated_at: DateTime("2015-07-06T17:34:35")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
name: "testrepository"
full_name: "jtleek/testrepository"
description: "testrepository"
default_branch: "master"
owner: Owner("jtleek")
id: 4729869
size: 48
forks_count: 0
stargazers_count: 3
watchers_count: 3
open_issues_count: 0
url: URI("https://api.github.com/repos/jtleek/testrepository")
html_url: URI("https://github.com/jtleek/testrepository")
clone_url: URI("https://github.com/jtleek/testrepository.git")
ssh_url: URI("git@github.com:jtleek/testrepository.git")
pushed_at: DateTime("2012-06-20T18:39:07")
created_at: DateTime("2012-06-20T18:39:06")
updated_at: DateTime("2017-11-13T14:02:46")
has_issues: true
has_wiki: true
has_downloads: true
has_pages: false
private: false
fork: false, Repo (all fields are Union{Nothing, T}):
```

```
    full_name: "jtleek/tidypvals"
    description: "An R package with several million published p-values in tidy data set
s. "
    language: "HTML"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 98407699
    size: 55034
    forks_count: 4
    stargazers_count: 74
    watchers_count: 74
    open_issues_count: 2
    url: URI("https://api.github.com/repos/jtleek/tidypvals")
    html_url: URI("https://github.com/jtleek/tidypvals")
    clone_url: URI("https://github.com/jtleek/tidypvals.git")
    ssh_url: URI("git@github.com:jtleek/tidypvals.git")
    homepage: URI("")
    pushed_at: DateTime("2017-08-04T02:21:47")
    created_at: DateTime("2017-07-26T09:51:06")
    updated_at: DateTime("2023-09-08T17:27:54")
    has_issues: true
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "tornado"
    full_name: "jtleek/tornado"
    language: "R"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 6374724
    size: 484
    forks_count: 2
    stargazers_count: 0
    watchers_count: 0
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/tornado")
    html_url: URI("https://github.com/jtleek/tornado")
    clone_url: URI("https://github.com/jtleek/tornado.git")
    ssh_url: URI("git@github.com:jtleek/tornado.git")
    pushed_at: DateTime("2012-09-26T19:49:17")
    created_at: DateTime("2012-10-24T17:50:09")
    updated_at: DateTime("2014-03-04T14:58:03")
    has_issues: false
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "tsp-devel"
    full_name: "jtleek/tsp-devel"
    description: "TSP Development Software"
    language: "C++"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 12784874
    size: 112
    forks_count: 1
    stargazers_count: 0
    watchers_count: 0
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/tsp-devel")
    html_url: URI("https://github.com/jtleek/tsp-devel")
    clone_url: URI("https://github.com/jtleek/tsp-devel.git")
    tleek/tsp-devel.git")
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

    pushed_at: DateTime("2013-09-20T19:58:53")
    created_at: DateTime("2013-09-12T13:49:18")
    updated_at: DateTime("2013-11-25T18:49:53")
    has_issues: true
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "tspreg"
    full_name: "jtleek/tspreg"
    description: "An R package for performing top-scoring pairs regression."
    language: "R"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 12134822
    size: 190
    forks_count: 8
    stargazers_count: 4
    watchers_count: 4
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/tspreg")
    html_url: URI("https://github.com/jtleek/tspreg")
    clone_url: URI("https://github.com/jtleek/tspreg.git")
    ssh_url: URI("git@github.com:jtleek/tspreg.git")
    pushed_at: DateTime("2015-03-31T01:29:50")
    created_at: DateTime("2013-08-15T13:20:30")
    updated_at: DateTime("2020-10-23T14:27:36")
    has_issues: true
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: false, Repo (all fields are Union{Nothing, T}):
    name: "which_house_races_are_close"
    full_name: "jtleek/which_house_races_are_close"
    description: "Join together FiveThirtyEight + Cook ratings on House races and put in
to easily sortable table. "
    language: "Jupyter Notebook"
    default_branch: "master"
    owner: Owner("jtleek")
    id: 145731567
    size: 885
    forks_count: 0
    stargazers_count: 0
    watchers_count: 0
    open_issues_count: 0
    url: URI("https://api.github.com/repos/jtleek/which_house_races_are_close")
    html_url: URI("https://github.com/jtleek/which_house_races_are_close")
    clone_url: URI("https://github.com/jtleek/which_house_races_are_close.git")
    ssh_url: URI("git@github.com:jtleek/which_house_races_are_close.git")
    pushed_at: DateTime("2018-08-22T17:10:59")
    created_at: DateTime("2018-08-22T15:56:55")
    updated_at: DateTime("2018-08-22T17:11:00")
    has_issues: false
    has_wiki: true
    has_downloads: true
    has_pages: false
    private: false
    fork: true, Repo (all fields are Union{Nothing, T}):
    name: "who-is-who"
    full_name: "jtleek/who-is-who"
    description: "A repo to create a who's who website for the Fred Hutch data communit
y"
    default_branch: "main"

```
  id: 571676878
  size: 367
  forks_count: 0
  stargazers_count: 0
  watchers_count: 0
  open_issues_count: 0
  url: URI("https://api.github.com/repos/jtleek/who-is-who")
  html_url: URI("https://github.com/jtleek/who-is-who")
  clone_url: URI("https://github.com/jtleek/who-is-who.git")
  ssh_url: URI("git@github.com:jtleek/who-is-who.git")
  homepage: URI("https://whoiswho.fredhutch.org")
  pushed_at: DateTime("2022-11-28T16:43:46")
  created_at: DateTime("2022-11-28T16:39:50")
  updated_at: DateTime("2023-09-08T18:39:26")
  has_issues: false
  has_wiki: false
  has_downloads: true
  has_pages: false
  private: false
  fork: true]Dict("prev" => "https://api.github.com/user/1571674/repos?page=3", "firs
t" => "https://api.github.com/user/1571674/repos?page=1")
```

ParseError:
Error @]8;;file:///home/number25/MEGA/Computational-Bio/Kobe-n-Pascal/Computationa
lBiology-for-Autodidacts/Course_notes/Coursera_Data_Science_Track_JHOP/Mod3_Data_Clean
ing/Week_2/In[210]#5:1 \In[210]:5:1]8;; \
    end
end
⌴ — invalid identifier

Stacktrace:
 [1] top-level scope
   @ In[210]:5

Search the output for the repo "datasharing" - hmmm this seems more difficult than I first assumed, as the output type is a Repo-Union custom type it seems

Let's try honing in a specific repo, so use the **repo** function with the exact repo as the argument (including his username: username/repo)

```
In [184… repo("jtleek/datasharing")
```

```
Out[184]:  Repo (all fields are Union{Nothing, T}):
               name: "datasharing"
               full_name: "jtleek/datasharing"
               description: "The Leek group guide to data sharing "
               default_branch: "master"
               owner: Owner("jtleek")
               id: 14204342
               size: 590
               subscribers_count: 561
               forks_count: 243587
               stargazers_count: 6434
               watchers_count: 6434
               open_issues_count: 892
               url: URI("https://api.github.com/repos/jtleek/datasharing")
               html_url: URI("https://github.com/jtleek/datasharing")
               clone_url: URI("https://github.com/jtleek/datasharing.git")
               ssh_url: URI("git@github.com:jtleek/datasharing.git")
               pushed_at: DateTime("2024-01-05T04:49:32")
               created_at: DateTime("2013-11-07T13:25:07")
               updated_at: DateTime("2024-02-01T12:03:04")
               has_issues: true
               has_wiki: true
               has_downloads: true
               has_pages: false
               private: false
               fork: false
```

Answer to the question is **created_at: DateTime("2013-11-07T13:25:07")**

2.

The sqldf package allows for execution of SQL commands on R data frames. We will use the sqldf package to practice the queries we might send with the dbSendQuery command in RMySQL.

Download the American Community Survey data and load it into an R object called

https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv

Which of the following commands will select only the data for the probability weights pwgtp1 with ages less than 50?

```
In [211…  using CSV
```

Download the csv and port it into a dataframe

```
In [213…  download("https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv"
          , "quiz2.csv")
```

```
Out[213]:  "quiz2.csv"
```

```
In [215…  quiz2_dataframe = CSV.File("quiz2.csv") |> DataFrame
```

14931×239 DataFrame

Row	RT	SERIALNO	SPORDER	PUMA	ST	ADJUST	PWGTP	AGEP	CIT	COW	DD
	String1	Int64	Int64	Int64	Int64	Int64	Int64	Int64	Int64	Int64?	Int
1	P	186	1	700	16	1015675	89	43	1	7	
2	P	186	2	700	16	1015675	92	42	1	4	
3	P	186	3	700	16	1015675	107	16	1	1	
4	P	186	4	700	16	1015675	91	14	1	*missing*	
5	P	306	1	700	16	1015675	309	29	1	5	
6	P	395	1	100	16	1015675	108	40	1	8	
7	P	395	2	100	16	1015675	90	15	1	*missing*	
8	P	506	1	700	16	1015675	239	28	1	1	
9	P	506	2	700	16	1015675	213	30	1	1	
10	P	506	3	700	16	1015675	219	4	1	*missing*	*mi.*
11	P	506	4	700	16	1015675	226	1	1	*missing*	*mi.*
12	P	835	1	800	16	1015675	121	40	1	1	
13	P	835	2	800	16	1015675	133	42	1	6	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14920	P	1357182	1	700	16	1015675	24	19	1	6	
14921	P	1357182	2	700	16	1015675	24	2	1	*missing*	*mi.*
14922	P	1357238	1	300	16	1015675	111	24	1	1	
14923	P	1357238	2	300	16	1015675	74	21	1	4	
14924	P	1357340	1	400	16	1015675	57	85	1	*missing*	
14925	P	1357340	2	400	16	1015675	69	79	1	*missing*	
14926	P	1357874	1	900	16	1015675	32	84	1	*missing*	
14927	P	1357874	2	900	16	1015675	28	74	1	2	
14928	P	1357880	1	500	16	1015675	121	22	1	1	
14929	P	1357880	2	500	16	1015675	112	22	1	1	
14930	P	1358490	1	700	16	1015675	353	28	1	1	
14931	P	1358490	2	700	16	1015675	386	23	1	1	

Since this quiz question is asking us to use a library which allows SQL quiries on DataFrames, we have to import it here

In [216… `Pkg.add("SQLdf") ; using SQLdf`

```
    Resolving package versions...
    Installed Calculus ————————————— v0.5.1
    Installed Rmath ———————————————— v0.7.1
    Installed DualNumbers ——————————— v0.6.8
    Installed WinReg ———————————————— v1.0.0
    Installed SQLite_jll ———————————— v3.45.0+0
    Installed HypergeometricFunctions — v0.3.23
    Installed StatsFuns ————————————— v1.3.0
    Installed Rmath_jll ————————————— v0.4.0+0
    Installed NaNMath ——————————————— v1.0.2
    Installed SQLite ———————————————— v1.6.0
    Installed StatsAPI —————————————— v1.7.0
    Installed ShiftedArrays ————————— v2.0.0
    Installed CategoricalArrays ————— v0.10.8
    Installed StatsModels ——————————— v0.7.3
    Installed RCall ————————————————— v0.13.18
    Installed StatsBase ————————————— v0.34.2
    Installed SQLdf ————————————————— v0.2.0
    Updating `~/.julia/environments/v1.10/Project.toml`
  [1d9d1391] + SQLdf v0.2.0
    Updating `~/.julia/environments/v1.10/Manifest.toml`
  [49dc2e85] + Calculus v0.5.1
  [324d7699] + CategoricalArrays v0.10.8
  [fa6b7ba4] + DualNumbers v0.6.8
  [34004b35] + HypergeometricFunctions v0.3.23
  [77ba4419] + NaNMath v1.0.2
⌃ [6f49c342] + RCall v0.13.18
  [79098fc4] + Rmath v0.7.1
  [1d9d1391] + SQLdf v0.2.0
  [0aa819cd] + SQLite v1.6.0
  [1277b4bf] + ShiftedArrays v2.0.0
  [82ae8749] + StatsAPI v1.7.0
  [2913bbd2] + StatsBase v0.34.2
  [4c63d2b9] + StatsFuns v1.3.0
  [3eaba693] + StatsModels v0.7.3
  [1b915085] + WinReg v1.0.0
  [f50d1b31] + Rmath_jll v0.4.0+0
  [76ed43ae] + SQLite_jll v3.45.0+0
        Info Packages marked with ⌃ have new versions available but compatibility cons
traints restrict them from upgrading. To see why use `status --outdated -m`
    Building RCall → `~/.julia/scratchspaces/44cfe95a-1eb2-52ea-b672-e2afdf69b78f/3084
689b18f9e5e817a6ce9a83a7654d8ad0f2f6/build.log`
Precompiling project...
  ✓ StatsAPI
  ✓ ShiftedArrays
  ✓ NaNMath
  ✓ WinReg
  ✓ Calculus
  ✓ Rmath_jll
  ✓ SQLite_jll
  ✓ Rmath
  ✓ CategoricalArrays
  ✓ DualNumbers
  ✓ CategoricalArrays → CategoricalArraysSentinelArraysExt
  ✓ SQLite
  ✓ HypergeometricFunctions
  ✓ StatsBase
  ✓ CategoricalArrays → CategoricalArraysJSONExt
  ✓ StatsFuns
  ✓ StatsModels
  ✓ RCall
  ✓ SQLdf
  19 dependencies successfully precompiled in 23 seconds. 76 already precompiled.
```

Perform the query using the SQLdf macro - looks pretty damn simply, albeit a bit slow

`@sqldf "select pwgtp1 from quiz2_dataframe where AGEP < 50"`

Out[218]: 10093×1 DataFrame   *10068 rows omitted*

Row	pwgtp1
	Int64
1	87
2	88
3	94
4	91
5	539
6	192
7	153
8	232
9	205
10	226
11	225
12	109
13	129
⋮	⋮
10082	10
10083	22
10084	11
10085	23
10086	20
10087	24
10088	160
10089	104
10090	39
10091	28
10092	397
10093	481

3.

Using the same data frame you created in the previous problem, what is the equivalent function to unique(acs$AGEP)

In [220… `@sqldf "select distinct AGEP from quiz2_dataframe"`

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Row	AGEP
	Int64
1	43
2	42
3	16
4	14
5	29
6	40
7	15
8	28
9	30
10	4
11	1
12	18
13	37
⋮	⋮
80	66
81	25
82	93
83	8
84	36
85	41
86	44
87	84
88	72
89	73
90	85
91	89

4.

How many characters are in the 10th, 20th, 30th and 100th lines of HTML from this page:

http://biostat.jhsph.edu/~jleek/contact.html

```
In [261… url = "http://biostat.jhsph.edu/~jleek/contact.html"
```

Out[261]: "http://biostat.jhsph.edu/~jleek/contact.html"

```
In [262… download(url, "contact.html")
```

Out[262]: "contact.html"

```
In [248… #samp = HTTP.get(url, cookies=true);
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
#data = String(samp.body)
#samp
#data
```

```
MethodError: no method matching open(::String)
You may have intended to import Base.open

Closest candidates are:
  open(::Function, ::Union{String, Symbol}, ::Any, ::Any; kw...)
   @ HTTP ~/.julia/packages/HTTP/bDoga/src/HTTP.jl:579
  open(::Function, ::Union{String, Symbol}, ::Any; ...)
   @ HTTP ~/.julia/packages/HTTP/bDoga/src/HTTP.jl:579
  open(::Function, ::Any; suppress_close_error, verbose, headers, maxframesize, maxfra
gmentation, kw...)
   @ HTTP ~/.julia/packages/HTTP/bDoga/src/WebSockets.jl:351


Stacktrace:
 [1] top-level scope
   @ In[248]:2
```

In [241…
```julia
counter = 0
for line in eachline(samp) #(open("contact.html"))
        if counter == 10
            println(line)
        else counter += 1
    end
end
```

```
MethodError: no method matching eachline(::HTTP.Messages.Response)

Closest candidates are:
  eachline(; ...)
   @ Base io.jl:1063
  eachline(::Base.AbstractCmd; keep)
   @ Base process.jl:338
  eachline(::AbstractString; keep)
   @ Base io.jl:1067
  ...


Stacktrace:
 [1] top-level scope
   @ ./In[241]:2
```

After some trial and error, I've discovered that it's likely easiest to use the Gumbo.jl package for parsing HTML files https://github.com/JuliaWeb/Gumbo.jl . Iterating through a HTML using the previous methods was inadequate and clunky

In [252…
```julia
Pkg.add("Gumbo") ; using Gumbo
```

```
  Resolving package versions...
 No Changes to `~/.julia/environments/v1.10/Project.toml`
 No Changes to `~/.julia/environments/v1.10/Manifest.toml`
WARNING: using Gumbo.parsehtml in module Main conflicts with an existing identifier.
```

In [283…
```julia
leekhtml = Gumbo.parsehtml(read("contact.html", String), preserve_whitespace=true)
```

```
Out[283]: HTML Document:
          <!DOCTYPE html>
          HTMLElement{:HTML}:<HTML lang="en" xml:lang="en" xmlns="http://www.w3.org/1999/xhtm
          l">
            <head>
              <meta content="Welcome to Jeff Leek&#39;s Research Group" name="Description"/>
              <meta content="Johns Hopkins University, Bloomberg School of Public Health, Bios
          tatistics, Surrogate Variable Analysis, sva, expression heterogeneity, clinical geno
          mics, reproducibility, high-dimensional data, Top Scoring Pair, tspair " name="Keywo
          rds"/>
              <meta content="text/html; charset=iso-8859-1" http-equiv="Content-Type"/>
              <meta content="Global" name="Distribution"/>
              <meta content="index,follow" name="Robots"/>
              <link href="images/favicon.ico" rel="icon" type="image/vnd.microsoft.icon"/>
              <link href="images/PixelGreen.css" rel="stylesheet" type="text/css"/>
              <title>
                jeffrey leek contact
              </title>
              <script type="text/javascript">

          var _gaq = _gaq || [];
          _gaq.push(['_setAccount', 'UA-20898652-1']);
          _gaq.push(['_trackPageview']);

          (function() {
            var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async
          = true;
          ...
```

```
In [289…  counter = 0
          for elem in PreOrderDFS(leekhtml.root)
              if counter == 10
                      println(attrs(elem))
              return
              else counter += 1
              end
              println(elem)
          end
```

```html
<HTML lang="en" xml:lang="en" xmlns="http://www.w3.org/1999/xhtml"><head>

<meta content="Welcome to Jeff Leek&#39;s Research Group" name="Description"/>
<meta content="Johns Hopkins University, Bloomberg School of Public Health, Biostatist
ics, Surrogate Variable Analysis, sva, expression heterogeneity, clinical genomics, re
producibility, high-dimensional data, Top Scoring Pair, tspair " name="Keywords"/>
<meta content="text/html; charset=iso-8859-1" http-equiv="Content-Type"/>
<meta content="Global" name="Distribution"/>
<meta content="index,follow" name="Robots"/>

<link href="images/favicon.ico" rel="icon" type="image/vnd.microsoft.icon"/>
<link href="images/PixelGreen.css" rel="stylesheet" type="text/css"/>



<title>jeffrey leek contact</title>

<script type="text/javascript">

  var _gaq = _gaq || [];
  _gaq.push(['_setAccount', 'UA-20898652-1']);
  _gaq.push(['_trackPageview']);

  (function() {
    var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async =
true;
    ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') +
'.google-analytics.com/ga.js';
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga,
s);
  })();

</script>
</head>

<body>

<div id="wrap">

        <div id="header"><div id="header-content">

                <h1 id="logo"><a href="research.html" title="">jeffrey<span class="gra
y">leek</span></a></h1>
                <h2 id="slogan">High Dimensional Data and Genomics</h2>


                <ul>
                        <li><a href="research.html">Research</a></li>
                        <li><a href="publications.html">Publications</a></li>
                        <li><a href="software.html">Software and Data</a></li>
                        <li><a href="about.html">About</a></li>
                        <li><a href="contact.html" id="current">Contact</a></li>
                </ul>


        </div></div>

        <div class="headerphoto-cont"></div>


        <div id="content-wrap"><div id="content">

                <div id="sidebar">

                        <div class="sidebox">
```
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```html
                                        <h1>About</h1>

                                        <p>Jeff is an assistant professor in the Biostatistics
Department of the Johns Hopkins Bloomberg School of Public Health. His work focuses on
statistical
                                        methods for high-dimensional data and genomics. For more inf
o check out his <a href="about.html">Bio</a>.</p>

                                </div>


                                        <div class="sidebox">
                                        <h1>Media</h1>
                                        <ul class="sidemenu">
                                                <li><a href="http://simplystatistics.tumblr.co
m/">Simply Statistics (Blog)</a></li>
                                                <li><a href="http://www.twitter.com/leekgrou
p">@leekgroup</a></li>
                                        </ul>

                                </div>


                                <div class="sidebox">


                                        <h1>Affiliations</h1>
                                        <ul class="sidemenu">
                                                <li><a class="top" href="http://www.jhsph.ed
u/">JHSPH</a></li>
                                                <li><a href="http://www.biostat.jhsph.edu/">Bi
ostatistics Department</a></li>
                                                <li><a href="http://www.biostat.jhsph.edu/geno
mics/">Genomics @ Biostat</a></li>
                                                <li><a href="http://genomics.jhu.edu/">Computa
tional Genomics</a></li>
                                        </ul>

                                </div>

                        </div>

                        <div id="main">

                                <div class="post">

                                        <a name="TemplateInfo"></a>
                                        <h1>Contact Information</h1>

                                  <h3>Address </h3>
                                  <p>
                                  Johns Hopkins University <br/>
                                  Bloomberg School of Public Health <br/>
                                  615 North Wolfe Street <br/>
                                  Baltimore, MD 21205-2179 <br/>
                                  </p>
                                  <h3>Phone</h3>
                                  <p>410-955-1166 (I am <b>much</b> easier to reach by email)
</p>
                                  <h3>Fax</h3>
                                  <p>410-955-0958</p>
                                  <h3>Email</h3>
                                  <p> jleek || jhsph dot edu </p>
                                  itter</h3>
```

```html
                                                <p> <a href="http://www.twitter.com/leekgroup">@leekgroup</a
></p>

                                                <h3>Blog</h3>
                                                <p> <a href="http://simplystatistics.tumblr.com/">Simply Sta
tistics</a></p>


                                </div>


                </div>


        </div></div>



<div id="footer"><div id="footer-content">


                <div class="col float-left">
                        <h1>Links</h1>
                        <ul>
                                <li><a href="http://groups.google.com/group/edge-softw
are?pli=1">EDGE Discussion Group</a></li>
                                <li><a href="http://www.biostat.jhsph.edu/newsEvent/ev
ent/seminar/seminars.shtml">Biostat Seminars</a></li>
                        </ul>
                </div>


                <div class="col2 float-right">
                <p>
                © copyright 2009 <strong>Jeffrey Leek</strong><br/>
                Artwork © copyright 2009 <a href="mailto:codyandanna@gmail.com">Cody L
ail</a> <br/>
                Photo <a href="attributions.html">Attributions</a><br/>
                Design by: <a href="http://www.styleshout.com/"><strong>styleshout</st
rong></a>
                Valid <a href="http://jigsaw.w3.org/css-validator/check/referer"><stro
ng>CSS</strong></a> |
                        <a href="http://validator.w3.org/check/referer"><strong>XHTML</s
trong></a>
                </p>


                </div>

</div></div>



</div>



</body></HTML>
<head>

<meta content="Welcome to Jeff Leek&#39;s Research Group" name="Description"/>
<meta content="Johns Hopkins University, Bloomberg School of Public Health, Biostatist
ics, Surrogate Variable Analysis, sva, expression heterogeneity, clinical genomics, re
producibility, high-dimensional data, Top Scoring Pair, tspair " name="Keywords"/>
<meta content="text/html; charset=iso-8859-1" http-equiv="Content-Type"/>
<meta content="Global" name="Distribution"/>
```
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js ="Robots"/>

```
<link href="images/favicon.ico" rel="icon" type="image/vnd.microsoft.icon"/>
<link href="images/PixelGreen.css" rel="stylesheet" type="text/css"/>


<title>jeffrey leek contact</title>

<script type="text/javascript">

  var _gaq = _gaq || [];
  _gaq.push(['_setAccount', 'UA-20898652-1']);
  _gaq.push(['_trackPageview']);

  (function() {
    var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async =
true;
    ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') +
'.google-analytics.com/ga.js';
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga,
s);
  })();

</script>
</head>



<meta content="Welcome to Jeff Leek&#39;s Research Group" name="Description"/>


<meta content="Johns Hopkins University, Bloomberg School of Public Health, Biostatist
ics, Surrogate Variable Analysis, sva, expression heterogeneity, clinical genomics, re
producibility, high-dimensional data, Top Scoring Pair, tspair " name="Keywords"/>


<meta content="text/html; charset=iso-8859-1" http-equiv="Content-Type"/>


<meta content="Global" name="Distribution"/>
```

```
UndefVarError: `attrs` not defined

Stacktrace:
 [1] top-level scope
   @ ./In[289]:4
```

After a string of unsuccessful attempts, I have opted to simply used some bash magic, to extract a specific line using sed from within julia.

We can execute commands within julia by enclosing the command in back ticks `command`

In [310… `line_ten = read(`sed -e 10p -n contact.html`, String)`

Out[310]: `"<meta name=\"Distribution\" content=\"Global\" />\n"`

As we can see, the newline character is also present and will therefore be counted - so let's use the **chomp** function to remove this trailing character

In [311… `chomp(line_ten)`

Out[311]: `"<meta name=\"Distribution\" content=\"Global\" />"`

Now take the length of the line

```
In [312... length(chomp(line_ten))
```

Out[312]: 45

45. Do the remainder of the tasks

```
In [314... line_20 = read(`sed -e 20p -n contact.html`, String)
```

Out[314]: "<script type=\"text/javascript\">\n"

```
In [315... length(chomp(line_20))
```

Out[315]: 31

```
In [316... line_30 = read(`sed -e 30p -n contact.html`, String)
```

Out[316]: " })();\n"

```
In [317... length(chomp(line_30))
```

Out[317]: 7

```
In [318... line_100 = read(`sed -e 100p -n contact.html`, String)
```

Out[318]: "\t\t\t\t<ul class=\"sidemenu\">\n"

```
In [319... length(chomp(line_100))
```

Out[319]: 25

Answer is 45, 31, 7 and 25

5.

Read this data set into R and report the sum of the numbers in the fourth of the nine columns.

https://d396qusza40orc.cloudfront.net/getdata%2Fwksst8110.for

Original source of the data: http://www.cpc.ncep.noaa.gov/data/indices/wksst8110.for

(Hint this is a fixed width file format)

```
In [320... download("https://d396qusza40orc.cloudfront.net/getdata%2Fwksst8110.for", "question5.
```

Out[320]: "question5.for"

```
In [333... q = readlines(open("question5.for"))
```

```
Out[333]:  1258-element Vector{String}:
           " Weekly SST data starts week centered on 3Jan1990"
           ""
           "                Nino1+2      Nino3       Nino34       Nino4"
           " Week          SST SSTA    SST SSTA    SST SSTA    SST SSTA"
           " 03JAN1990     23.4-0.4    25.1-0.3    26.6 0.0    28.6 0.3"
           " 10JAN1990     23.4-0.8    25.2-0.3    26.6 0.1    28.6 0.3"
           " 17JAN1990     24.2-0.3    25.3-0.3    26.5-0.1    28.6 0.3"
           " 24JAN1990     24.4-0.5    25.5-0.4    26.5-0.1    28.4 0.2"
           " 31JAN1990     25.1-0.2    25.8-0.2    26.7 0.1    28.4 0.2"
           " 07FEB1990     25.8 0.2    26.1-0.1    26.8 0.1    28.4 0.3"
           " 14FEB1990     25.9-0.1    26.4 0.0    26.9 0.2    28.5 0.4"
           " 21FEB1990     26.1-0.1    26.7 0.2    27.1 0.3    28.9 0.8"
           " 28FEB1990     26.1-0.2    26.7-0.1    27.2 0.3    29.0 0.8"
           ⋮
           " 23OCT2013     20.5-0.4    24.7-0.2    26.3-0.4    28.7 0.1"
           " 30OCT2013     20.5-0.7    24.8-0.1    26.5-0.2    28.8 0.2"
           " 06NOV2013     21.0-0.4    24.8-0.1    26.6 0.0    28.9 0.3"
           " 13NOV2013     21.1-0.4    24.7-0.2    26.6 0.0    28.9 0.3"
           " 20NOV2013     21.3-0.5    24.8-0.2    26.7 0.1    28.9 0.3"
           " 27NOV2013     21.7-0.4    24.9-0.1    26.7 0.1    28.8 0.2"
           " 04DEC2013     22.0-0.3    24.9-0.2    26.8 0.2    28.8 0.3"
           " 11DEC2013     22.4-0.2    25.1 0.0    26.6 0.0    28.7 0.2"
           " 18DEC2013     22.4-0.5    25.2 0.0    26.5-0.1    28.6 0.1"
           " 25DEC2013     22.8-0.4    25.2-0.1    26.4-0.2    28.5 0.1"
           " 01JAN2014     23.7 0.0    25.2-0.2    26.3-0.3    28.2-0.2"
           " 08JAN2014     24.2 0.1    25.1-0.5    26.0-0.5    28.2-0.2"
```

Toying with a function which will take the Vector{String} formatted dataset and input it to a DataFrame - more difficult than it first appears

In [506…
```julia
function vectodf(q)
    #mymatrix = Array(1300, 5){undef}
    #mydata = DataFrame(Matrix{Any}(undef, 1300, 5), ["one", "two", "three", "four",
    mydata = DataFrame(Matrix(zeros(1300, 5)), ["one", "two", "three", "four", "five"
    @show mydata
    dfsize = length(q[5:end])
    col, row = 1,1
    while row <= dfsize
        if col == 6
            col = 1
            row += 1
        end
        for f in collect(q[5:end])
            z = split(f)
            for line in z
            @show String(line)
            push!(mydata[row,col], String(line)) #[row,col]
            col += 1
            end
        #println(f)
        end
    end
    mydata
end
```

Out[506]: vectodf (generic function with 2 methods)

In [501…
```julia
mydata = DataFrame(Matrix(zeros(1300, 5)), ["one", "two", "three", "four", "five"])
```

Row	one	two	three	four	five
	Float64	Float64	Float64	Float64	Float64
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0
⋮	⋮	⋮	⋮	⋮	⋮
1289	0.0	0.0	0.0	0.0	0.0
1290	0.0	0.0	0.0	0.0	0.0
1291	0.0	0.0	0.0	0.0	0.0
1292	0.0	0.0	0.0	0.0	0.0
1293	0.0	0.0	0.0	0.0	0.0
1294	0.0	0.0	0.0	0.0	0.0
1295	0.0	0.0	0.0	0.0	0.0
1296	0.0	0.0	0.0	0.0	0.0
1297	0.0	0.0	0.0	0.0	0.0
1298	0.0	0.0	0.0	0.0	0.0
1299	0.0	0.0	0.0	0.0	0.0
1300	0.0	0.0	0.0	0.0	0.0

In [414…
```julia
q_header = split(q[3])
for f in q_header
    parse(f, String)
end
#new_header = "Week"
#push!(new_header, x -> String(x) q_header)
#new_header[1:end]
```

```
MethodError: no method matching parse(::SubString{String}, ::Type{String})

Closest candidates are:
  parse(::Type{Union{}}, ::Any...; kwargs...)
   @ Base parse.jl:39


Stacktrace:
 [1] top-level scope
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Just trying out how I would create a DataFrame using the correct header

In [400… `q5_dataframe = DataFrame(zeros(10, 5), new_header)`

Out[400]: 10×5 DataFrame

Row	Nino1+2	Nino3	Nino34	Nino4	Week
	Float64	Float64	Float64	Float64	Float64
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0

Restructing the file using bash commands - removing the first 4 rows, then cleaning up the inconsistent columns - some are separated by hyphens and others are spaced apart. This confused me for hours as I assumed that there were only 5 columns, whereby the hyphens stood for measurement error, but in fact there are 9 columns in total.

```
tail -n +5 question5.for | tr - " "  | /
awk '{print $1, $2, $3, $4, $5, $6,
$7, $8, $9}' /
| tr " " '\t' > question5_cleaned.tsv
```

Import the re-shaped data into Julia

In [529… `cleaned_dataframe = DataFrame(CSV.File("question5_cleaned.tsv", header=false))`

`Out[529]:` 1254×9 DataFrame   *1229 rows omitt*

Row	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column
	String15	Float64	Float64	Float64	Float64	Float64	Float64	Float64	Float64
1	03JAN1990	23.4	0.4	25.1	0.3	26.6	0.0	28.6	0.
2	10JAN1990	23.4	0.8	25.2	0.3	26.6	0.1	28.6	0.
3	17JAN1990	24.2	0.3	25.3	0.3	26.5	0.1	28.6	0.
4	24JAN1990	24.4	0.5	25.5	0.4	26.5	0.1	28.4	0.
5	31JAN1990	25.1	0.2	25.8	0.2	26.7	0.1	28.4	0.
6	07FEB1990	25.8	0.2	26.1	0.1	26.8	0.1	28.4	0.
7	14FEB1990	25.9	0.1	26.4	0.0	26.9	0.2	28.5	0.
8	21FEB1990	26.1	0.1	26.7	0.2	27.1	0.3	28.9	0.
9	28FEB1990	26.1	0.2	26.7	0.1	27.2	0.3	29.0	0.
10	07MAR1990	26.7	0.3	26.7	0.2	27.3	0.2	28.9	0.
11	14MAR1990	26.1	0.4	26.9	0.2	27.3	0.1	28.6	0.
12	21MAR1990	26.1	0.2	27.2	0.0	27.6	0.3	28.7	0.
13	28MAR1990	25.7	0.4	27.5	0.2	27.8	0.3	28.8	0.
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
1243	23OCT2013	20.5	0.4	24.7	0.2	26.3	0.4	28.7	0.
1244	30OCT2013	20.5	0.7	24.8	0.1	26.5	0.2	28.8	0.
1245	06NOV2013	21.0	0.4	24.8	0.1	26.6	0.0	28.9	0.
1246	13NOV2013	21.1	0.4	24.7	0.2	26.6	0.0	28.9	0.
1247	20NOV2013	21.3	0.5	24.8	0.2	26.7	0.1	28.9	0.
1248	27NOV2013	21.7	0.4	24.9	0.1	26.7	0.1	28.8	0.
1249	04DEC2013	22.0	0.3	24.9	0.2	26.8	0.2	28.8	0.
1250	11DEC2013	22.4	0.2	25.1	0.0	26.6	0.0	28.7	0.
1251	18DEC2013	22.4	0.5	25.2	0.0	26.5	0.1	28.6	0.
1252	25DEC2013	22.8	0.4	25.2	0.1	26.4	0.2	28.5	0.
1253	01JAN2014	23.7	0.0	25.2	0.2	26.3	0.3	28.2	0.
1254	08JAN2014	24.2	0.1	25.1	0.5	26.0	0.5	28.2	0.

`In [530…`
```
sum(cleaned_dataframe.Column4)
```

`Out[530]:` 32426.699999999997

The answer is 32426.7

This took hours longer than I had hoped for, and although the tinkering and trial and error is stimulating and fruitful, knowing where the problem lays is importpant for future encounters - the idea behind reading "fixed-width formats" was new to me, so this unfamiliatiry is partly responsible for the pace of this question. These discussions provide some stimulating notes and hints

- https://discourse.julialang.org/t/reading-fixed-width-files/30163
- https://github.com/JuliaData/WeakRefStrings.jl
- https://gist.github.com/vaclavdekanovsky/245bbb0158f3480295b817ba88b77702

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

It took me agggggggges to notice that the columns for each entry were inconsistently formatted - in future I must peruse and inspect the dataset much more rigorously so avoid these relatively bening problems

We could try to **replace** the "-" delimiter with a space and thus even out the differences - and load this into a DataFrame in a single command - the first few lines of the fill will still cause issues in my guess

```
df = CSV.read(IOBuffer(replace(read("), UInt8('\t') => UInt8(' '))),
header=1, delim=" ", ignorerepeated=true, type=Int64, DataFrame)
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js