Webscraping

Hisam Sabouni

Goal

In this lecture we will explor a variety of data sources. I'll do my best to include some data sources that are of interest to all subjects and cover methods to extract data from websites. However, given my background we will first introduce a variety of data sources that may be of use to economists.

Economic Data Sources

The package *quantmod* in R "is designed to assist the quantitative trader in the development, testing, and deployment of statistically based trading models." To use a package in R (and Python) you have to first install the package. After the package is installed you have to load the package in to every new R (Python) session. That is, if you close R (Python) you have to reload in the package in order to use the functions within the package.

In R we can install packages through the command *install.packages('name of package')*. For example, below we will install the package quantmod:

```
#Delete the hashtag/pound symbol below to install the pacakge on your machine
#install.packages('quantmod')
```

Again, the installation only needs to happen once. From here on out to use the package you can simply load it in throught the command *library(package name)*, notice that no quotations are needed.

```
library(quantmod)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo
```

```
## Loading required package: TTR
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```
?quantmod
```

In the package documentation we can see there are many built in functions tailored for applications in quantitative finance. Given that this lecture is focused on gathering data we will give particular attention to the function *getSymbols()*.

Figure 1: Real GDP Growth from FRED website https://fred.stlouisfed.org/series/A191RL1Q225SBEA. Symbol for Real GDP is shown in the red box.

```
?getSymbols
```

The getSymbols function in the quantmod package can pull data from the following sources: Yahoo Finance, Google Finance (discontinued), MySQL, FRED, csv, RData, oanda (a foreign exchange broker), and av (alpha vantage).

Federal Reserve Economic Data (FRED)

For conducting research in Macroeconomics, data from FRED is extremely useful. We can pull up-to-date data on unemployment, inflation, GDP growth, etc. FRED has excellent coverage of data for the United States and other developed economies but is lacking in data for developing/emerging markets.

For example, we can pull data on Real GDP growth for the United States quite easily. The data the FED has on record can be found here: https://fred.stlouisfed.org/series/A191RL1Q225SBEA

On every page of the FRED website there are symbols for economic variables. These symbols can be used in conjunction with the quantmod function *getSymbols* to pull data directly into R. The default option for the *getSymbols* function is to pull data from Yahoo! Finance, we can tell the *getSymbols* function that we are interested in collecting data from FRED through the option *src*.

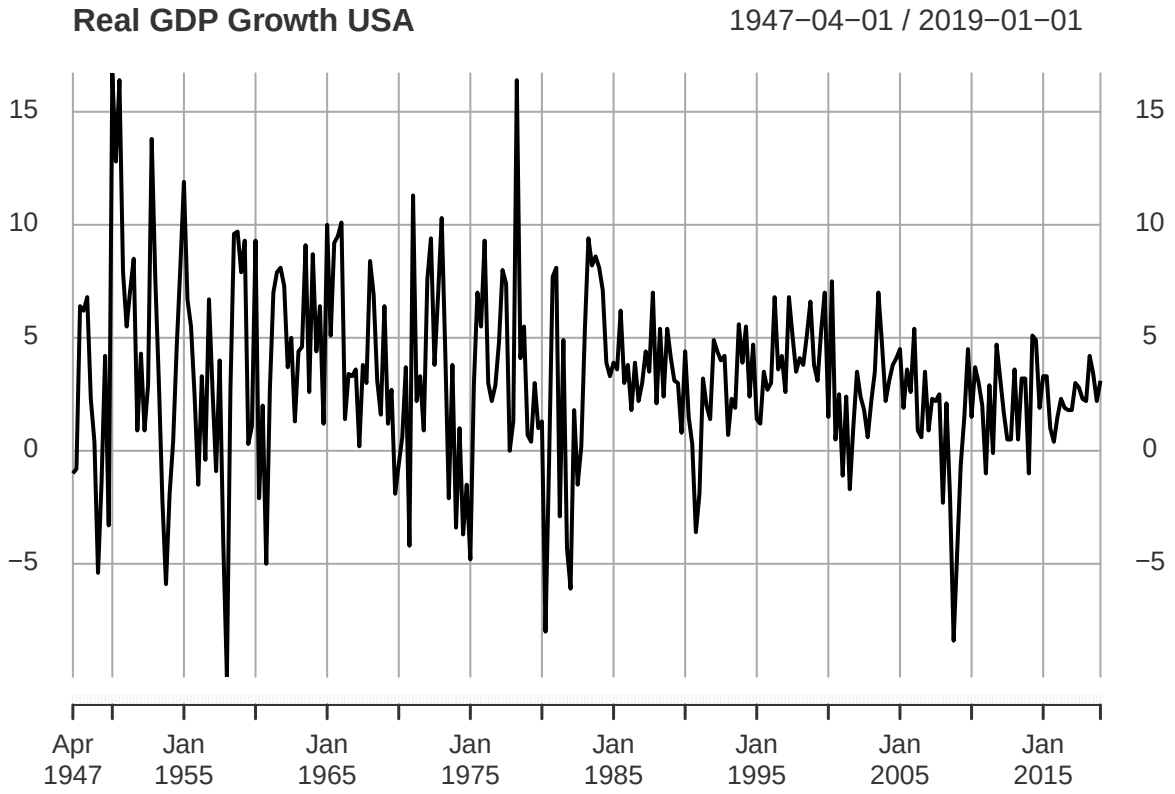
```
getSymbols('A191RL1Q225SBEA',src = 'FRED')
```

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
```

```
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
```

```
## [1] "A191RL1Q225SBEA"
```

```
plot(A191RL1Q225SBEA,ylab='%',xlab='',main='Real GDP Growth USA')
```



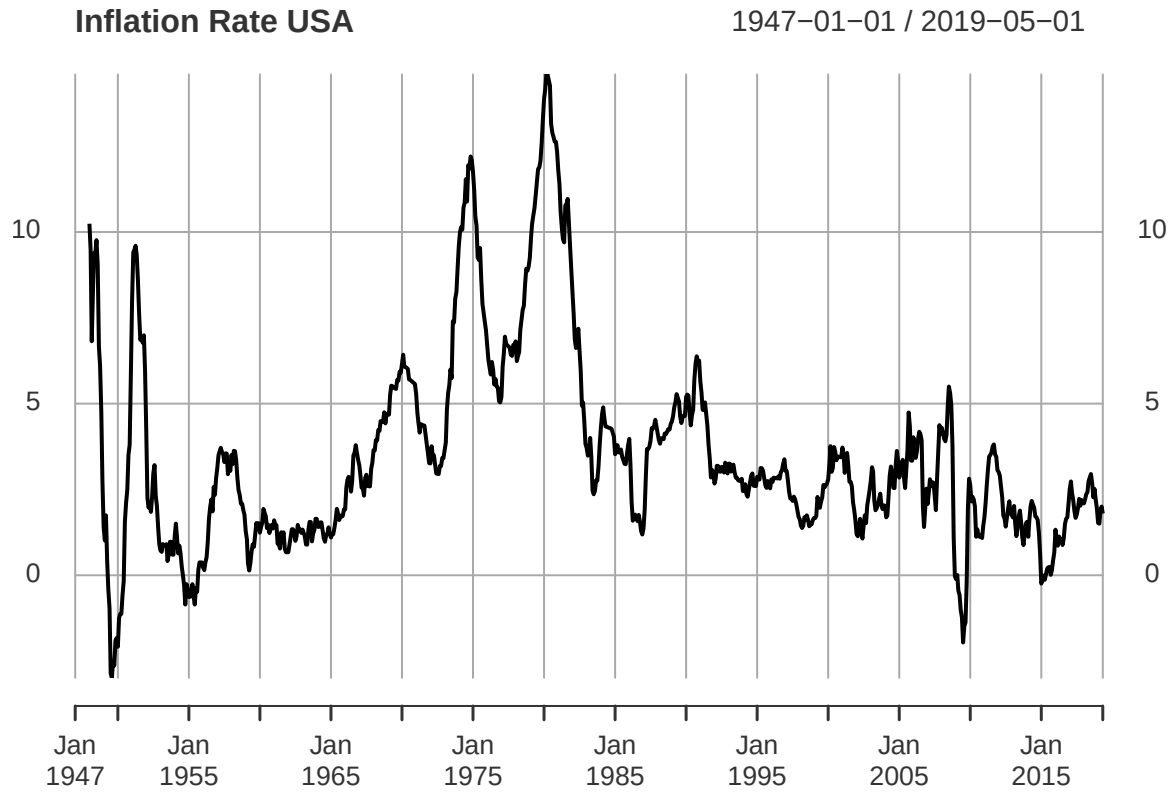
Every time we run the command *getSymbols('A191RL1Q225SBEA',src = 'FRED')* R will get the most recent data that is published by the Federal Reserve. Notice that although we used the base plot command we got a nicely formatted plot that contains dates along the x-axis. If you pay careful attention to your global enviornment, you'll see that the data downloaded by quantmod is actually coded as a time-series object (xts). Time-series objects contain dates and a series of data.

In the example above, we passed the *getSymbols* function a string containing they symbol for Real GDP growth. We can pull data on a variety of series at once by passing the *getSymbols* a vector of symbols. Below we will pull data on inflation (as measured by CPI), unemployment, the National Bureau of Economic Research (NBER) recession indicator, and the ten year treasury rate (the cost of borrowing for the US Government for 10 years):

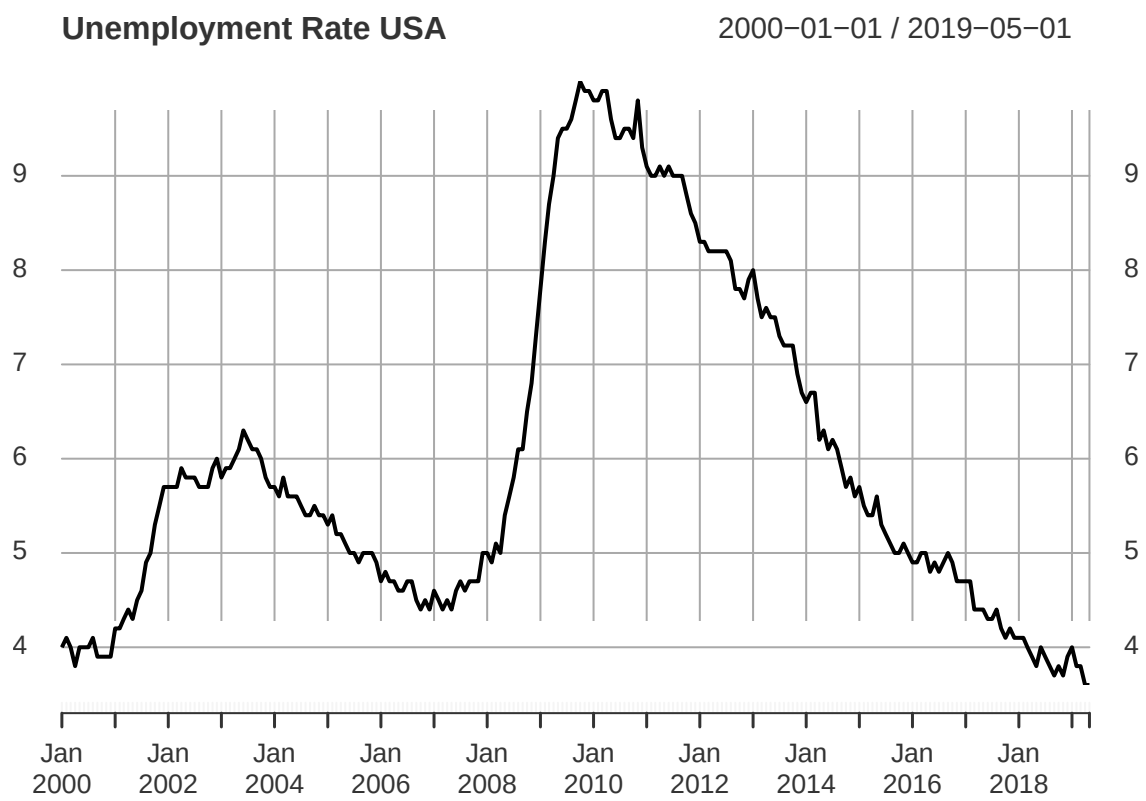
```
#Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)
#Civilian Unemployment Rate (UNRATE)
# NBER based Recession Indicators for the United States from the Period following the Peak through the
#10-Year Treasury Constant Maturity Rate (DGS10)
getSymbols(c('CPIAUCSL','UNRATE','USREC','DGS10'),src = 'FRED')
```

```
## [1] "CPIAUCSL" "UNRATE"   "USREC"    "DGS10"
```

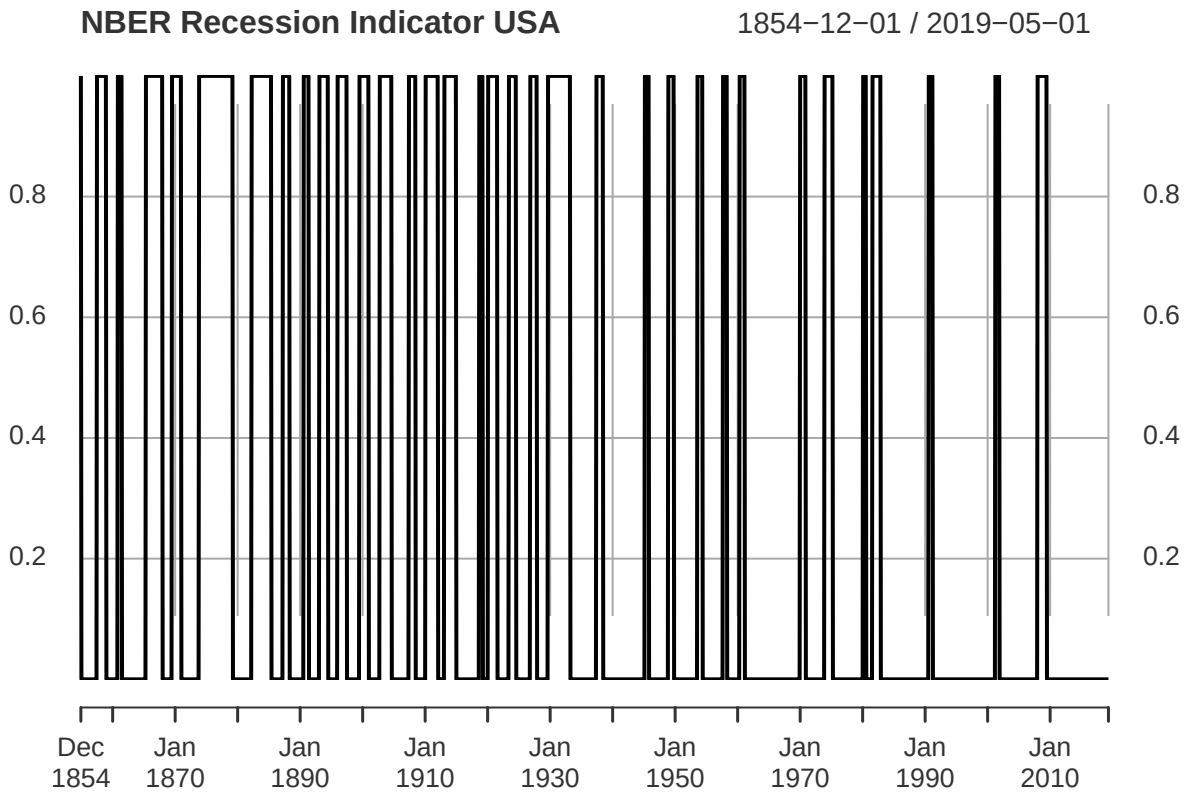
```
#The CPI data is an index; we can convert it to percentages using the command Delt
  #12 period differences are used to annualize the inflation rate, given that the data is monthly
cpi_percentage <- Delt(CPIAUCSL,k=12)*100
plot(cpi_percentage,ylab='%',xlab='',main='Inflation Rate USA')
```



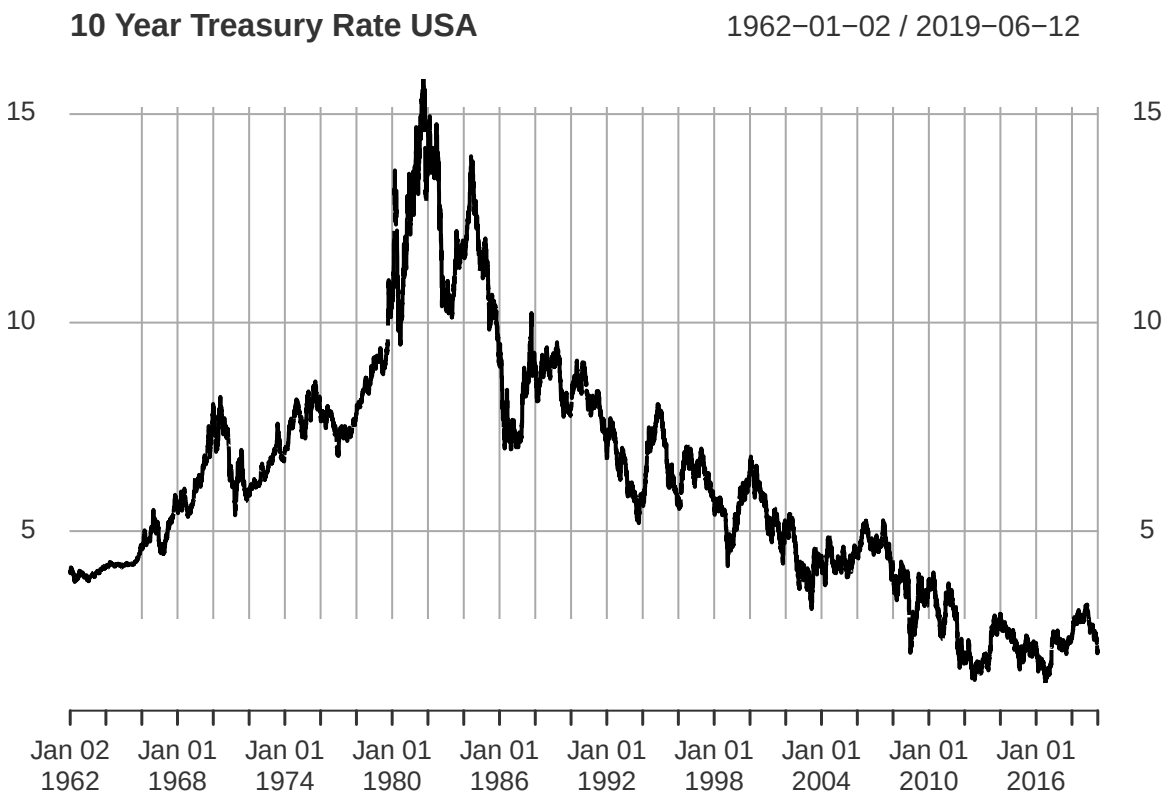

```
plot(UNRATE[index(UNRATE) >= as.Date('2000-01-01')],ylab='%',xlab='',main='Unemployment Rate USA')
```



```
plot(USREC,ylab='',xlab='',main='NBER Recession Indicator USA')
```




```
plot(DGS10,ylab='%',xlab='',main='10 Year Treasury Rate USA')
```



```
head(DGS10)
```

```
##            DGS10
## 1962-01-02  4.06
## 1962-01-03  4.03
## 1962-01-04  3.99
## 1962-01-05  4.02
## 1962-01-08  4.03
## 1962-01-09  4.05
```

At this point we should have six objects in our global envionment, all of which are 'xts' objects. We can combine all of our data into one data.frame through the command *merge*. R will recognize that time-series objects should be merged together by dates automatically.

```
#Convert the daily interst rates into quarterly data
quarterly_interest <- to.quarterly(DGS10)
```

```
## Warning in to.period(x, "quarters", indexAt = indexAt, name = name, ...):
## missing values removed from data
```

```
head(quarterly_interest) #OHLC data (open high low close of the quarter).
```

```
##         DGS10.Open DGS10.High DGS10.Low DGS10.Close
## 1962 Q1       4.06       4.13      3.83        3.86
## 1962 Q2       3.86       4.00      3.78        4.00
## 1962 Q3       4.00       4.05      3.94        3.94
## 1962 Q4       3.93       3.95      3.79        3.85
## 1963 Q1       3.82       3.96      3.80        3.95
## 1963 Q2       3.95       4.01      3.90        4.00
```

```
economic_data <- merge(A191RL1Q225SBEA,cpi_percentage,UNRATE,USREC,quarterly_interest[,4],all=F)
#We can change the column names:
colnames(economic_data) <- c('realGdp','cpi','unrate','recession','tenYearRate')
head(economic_data)
```

```
##            realGdp       cpi unrate recession tenYearRate
## 1962-01-01     7.3 0.6702413    5.8         0        3.86
## 1962-04-01     3.7 1.3418316    5.6         0        4.00
## 1962-07-01     5.0 1.0026738    5.4         0        3.94
## 1962-10-01     1.3 1.3342228    5.4         0        3.85
## 1963-01-01     4.4 1.3315579    5.7         0        3.95
## 1963-04-01     4.6 0.8937438    5.7         0        4.00
```

```
tail(economic_data)
```

```
##            realGdp      cpi unrate recession tenYearRate
## 2017-10-01     2.3 2.025727    4.1         0        2.40
## 2018-01-01     2.2 2.093691    4.1         0        2.74
## 2018-04-01     4.2 2.419576    3.9         0        2.85
## 2018-07-01     3.4 2.948975    3.9         0        3.05
## 2018-10-01     2.2 2.517164    3.8         0        2.69
## 2019-01-01     3.1 1.522396    4.0         0        2.41
```

Given the structure of our data we can now easily run a vareity of calculations, such as trying to build a model to forecast GDP, Unemployment, or simply see how different economic variables behave during expansionary versus recessionary times. The NBER Recession indicator takes on a value of 1 during recession and a value of 0 during expansionary periods.

```
#Tapply applies to calculations to a set of ids; here will summarize the data in the first argument, in
tapply(economic_data$realGdp,economic_data$recession,summary)
```

```
## $`0`
##      Index               realGdp
##  Min.   :1962-01-01   Min.   :-2.900
##  1st Qu.:1976-10-24   1st Qu.: 2.000
##  Median :1991-08-16   Median : 3.300
##  Mean   :1990-11-22   Mean   : 3.734
##  3rd Qu.:2004-12-09   3rd Qu.: 4.900
##  Max.   :2019-01-01   Max.   :16.400
##
## $`1`
##      Index               realGdp
##  Min.   :1970-01-01   Min.   :-8.400
##  1st Qu.:1974-08-16   1st Qu.:-3.950
##  Median :1982-04-01   Median :-1.700
##  Mean   :1987-07-22   Mean   :-1.878
##  3rd Qu.:2001-08-16   3rd Qu.: 0.400
##  Max.   :2009-04-01   Max.   : 3.700
```

```
tapply(economic_data$cpi,economic_data$recession,summary)
```

```
## $`0`
##      Index                 cpi
##  Min.   :1962-01-01   Min.   :-1.959
##  1st Qu.:1976-10-24   1st Qu.: 1.894
##  Median :1991-08-16   Median : 2.877
##  Mean   :1990-11-22   Mean   : 3.453
##  3rd Qu.:2004-12-09   3rd Qu.: 4.235
##  Max.   :2019-01-01   Max.   :13.869
##
## $`1`
##      Index                 cpi
##  Min.   :1970-01-01   Min.   :-0.5763
##  1st Qu.:1974-08-16   1st Qu.: 4.0992
##  Median :1982-04-01   Median : 6.0606
##  Mean   :1987-07-22   Mean   : 6.6546
##  3rd Qu.:2001-08-16   3rd Qu.: 9.8353
##  Max.   :2009-04-01   Max.   :14.5892
```

```
tapply(economic_data$unrate,economic_data$recession,summary)
```

```
## $`0`
##      Index               unrate
##  Min.   :1962-01-01   Min.   : 3.400
##  1st Qu.:1976-10-24   1st Qu.: 4.800
##  Median :1991-08-16   Median : 5.700
##  Mean   :1990-11-22   Mean   : 5.925
##  3rd Qu.:2004-12-09   3rd Qu.: 7.000
##  Max.   :2019-01-01   Max.   :10.400
##
## $`1`
##      Index               unrate
##  Min.   :1970-01-01   Min.   : 3.900
```

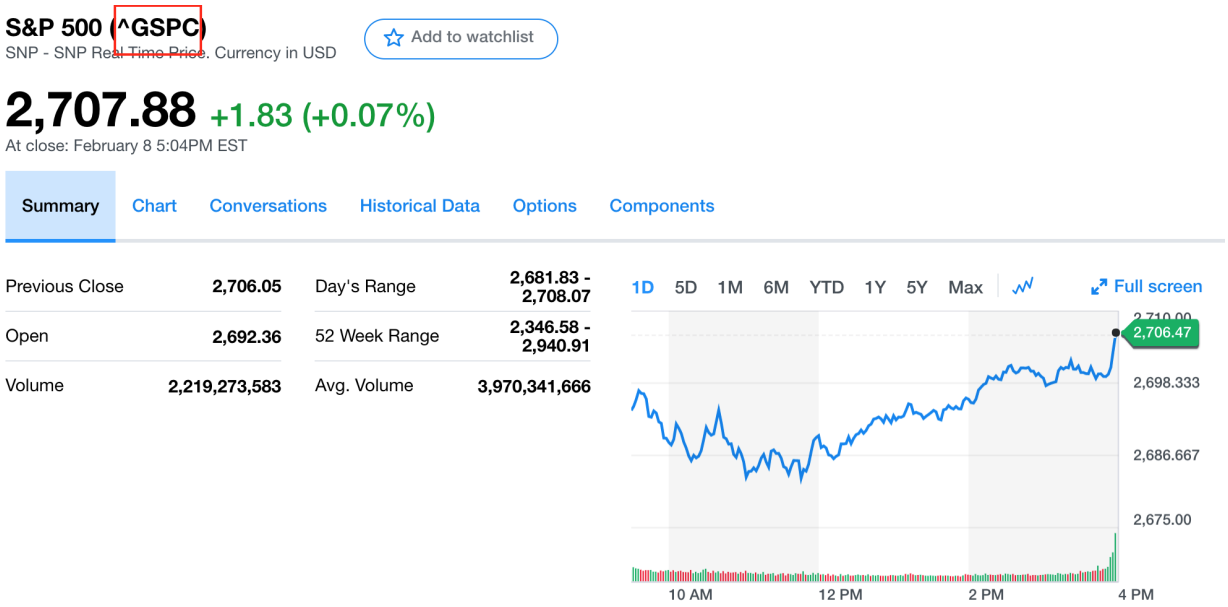


Figure 2: S&P 500 data from Yahoo! Finance. Symbol for S&P 500 is shown in the red box.

```
##  1st Qu.:1974-08-16   1st Qu.: 5.050
##  Median :1982-04-01   Median : 5.900
##  Mean   :1987-07-22   Mean   : 6.489
##  3rd Qu.:2001-08-16   3rd Qu.: 7.850
##  Max.   :2009-04-01   Max.   :10.400
```

```
tapply(economic_data$tenYearRate,economic_data$recession,mean)
```

```
##        0        1
## 5.971337 7.577407
```

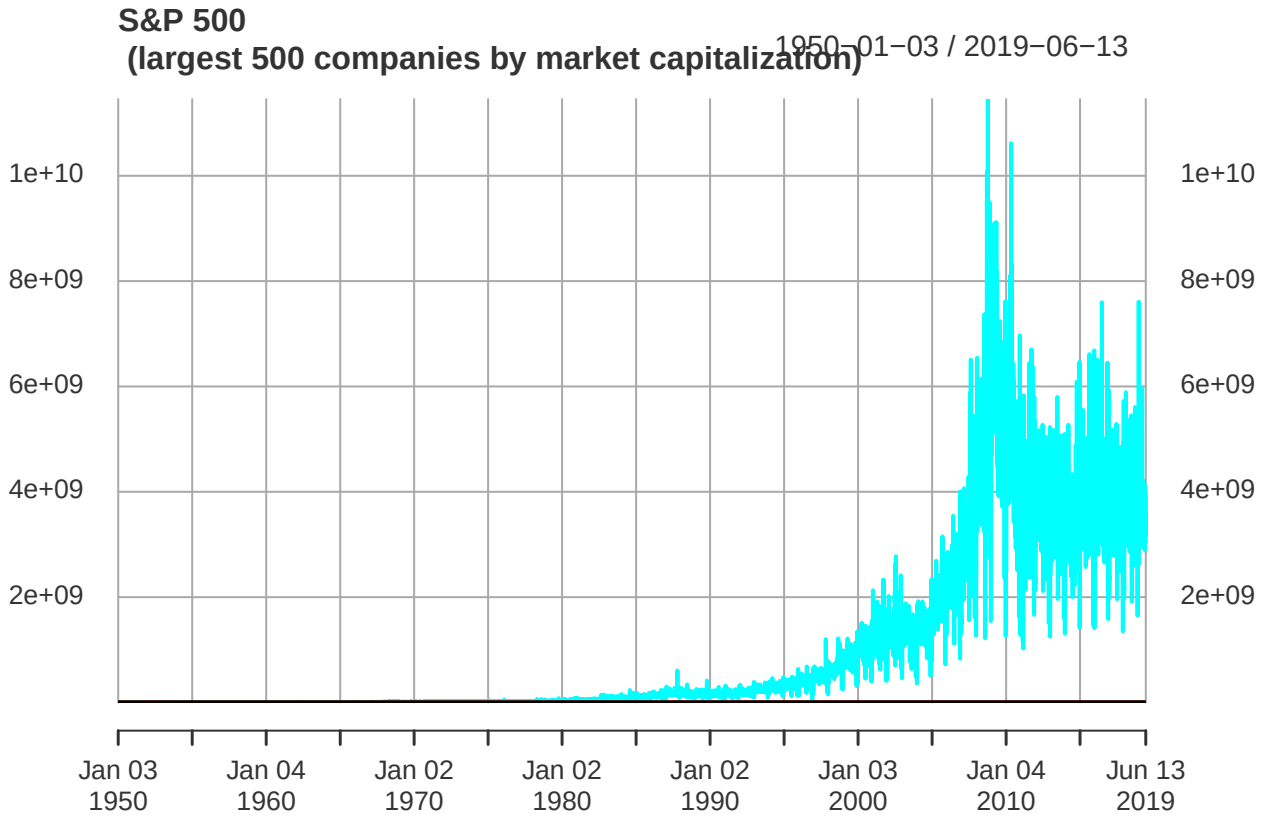
Yahoo Finance

The ability to pull data from Yahoo! Finance makes it easy to prototype ideas in financial research or to build portfolio managment software for your own financial endeavors. One can pull data form Yahoo! Finance by again extracting the symbols on the Yahoo! Finance website and passing them through the getSymbols function.

```
getSymbols('^GSPC',from ='1950-01-01')
```

```
## [1] "^GSPC"
```

```
plot(GSPC,main='S&P 500\n (largest 500 companies by market capitalization)')
```



Alpha Vantage

Alpha Vantage is a new data aggregation company that provides free access to intra-day stock market data, foreign exchange data, and cryptocurency data. They are free to use but have a rate limit of up to 5 API requests per minute and 500 requests per day. They offer a paid service for higher frequency data requests.

All of there documentation can be found here: https://www.alphavantage.co/documentation/

```
#getSymbols(c('INTC','ORCL'), src = 'av', api.key = '[your key]')
```

World Bank

The World Bank was formed in 1944 to help provide funding for poor/developing countries that were unable to recieve funding from commercial/institutional lenders. The World Bank collects data on countries from across the globe and their datasets are used for lots of reserach in development economics. Their website is:

https://databank.worldbank.org/data/home.aspx

Similar to *quantmod*, there is a package in R called *wbstats* that access the online API (application programming interface) the World Bank has created.

```
#install.packages('wbstats')
library(wbstats)
```

The *wbstats* library contains a helpful search feature to find the symbols the World Bank uses to code up their datasets. For instance if we wanted to find data on unemployment on all countries we could use the command *wbsearch* to view all of the available

```
unemp_search <- wbsearch(pattern = "unemployment")
head(unemp_search)
```

```
##    indicatorID
## 35   WP15177.9
## 36   WP15177.8
## 37   WP15177.7
## 38   WP15177.6
## 39   WP15177.5
## 40   WP15177.4
##                                                                                indicator
## 35         Received government transfers in the past year, income, richest 60% (% ages 15+) [w2]
## 36         Received government transfers in the past year, income, poorest 40% (% ages 15+) [w2]
## 37 Received government transfers in the past year, secondary education or more (% ages 15+) [w2]
## 38   Received government transfers in the past year, primary education or less (% ages 15+) [w2]
## 39                Received government transfers in the past year, older adults (% ages 25+) [w2]
## 40              Received government transfers in the past year, young adults (% ages 15-24) [w2]
```

We can pick one of the available symbols and download all of the data for that symbol through the *wb* command:

```
#Download all unemployment data for everything in the world bank database
unemployment_data <- wb(country = 'all',indicator = "UNEMPSA_")
head(unemployment_data)# First 6 rows of data
```

```
##   iso3c date    value indicatorID                    indicator iso2c
## 2  <NA> 2017 6.002274    UNEMPSA_ Unemployment rate,Percent,,,   AME
## 3  <NA> 2016 6.470872    UNEMPSA_ Unemployment rate,Percent,,,   AME
## 4  <NA> 2015 6.950939    UNEMPSA_ Unemployment rate,Percent,,,   AME
## 5  <NA> 2014 7.556938    UNEMPSA_ Unemployment rate,Percent,,,   AME
## 6  <NA> 2013 8.177776    UNEMPSA_ Unemployment rate,Percent,,,   AME
## 7  <NA> 2012 8.216599    UNEMPSA_ Unemployment rate,Percent,,,   AME
##              country
## 2 Advanced Economies
## 3 Advanced Economies
## 4 Advanced Economies
## 5 Advanced Economies
## 6 Advanced Economies
## 7 Advanced Economies
```

```
tail(unemployment_data)# Last 6 rows of data
```

```
##      iso3c date value indicatorID                    indicator iso2c
## 6898   VNM 2001  6.28    UNEMPSA_ Unemployment rate,Percent,,,    VN
## 6899   VNM 2000  6.42    UNEMPSA_ Unemployment rate,Percent,,,    VN
## 6900   VNM 1999  6.74    UNEMPSA_ Unemployment rate,Percent,,,    VN
## 6901   VNM 1998  6.85    UNEMPSA_ Unemployment rate,Percent,,,    VN
## 6902   VNM 1997  6.01    UNEMPSA_ Unemployment rate,Percent,,,    VN
## 6903   VNM 1996  5.88    UNEMPSA_ Unemployment rate,Percent,,,    VN
##      country
## 6898 Vietnam
## 6899 Vietnam
## 6900 Vietnam
## 6901 Vietnam
## 6902 Vietnam
## 6903 Vietnam
```

Note that the data structure stacks each countries data on top of each other. You have to be careful with such data when calculating things like returns or changes.

See here for more: https://cran.r-project.org/web/packages/wbstats/vignettes/Using_the_wbstats_package.html

International Monetary Fund

The IMF was created alongside the World Bank in 1944 at the Bretton Woods Conference. "The International Monetary Fund (IMF) is an organization of 189 countries, working to foster global monetary cooperation, secure financial stability, facilitate international trade, promote high employment and sustainable economic growth, and reduce poverty around the world." The IMF has also has a publicly accessible dataset that can be loaded into R:

```
#install.packages('imfr')
library(imfr)
```

Each R package has a vignette that summarizes all of the features of the pacakge

Check out this pacakges documentation at: https://cran.r-project.org/web/packages/imfr/imfr.pdf

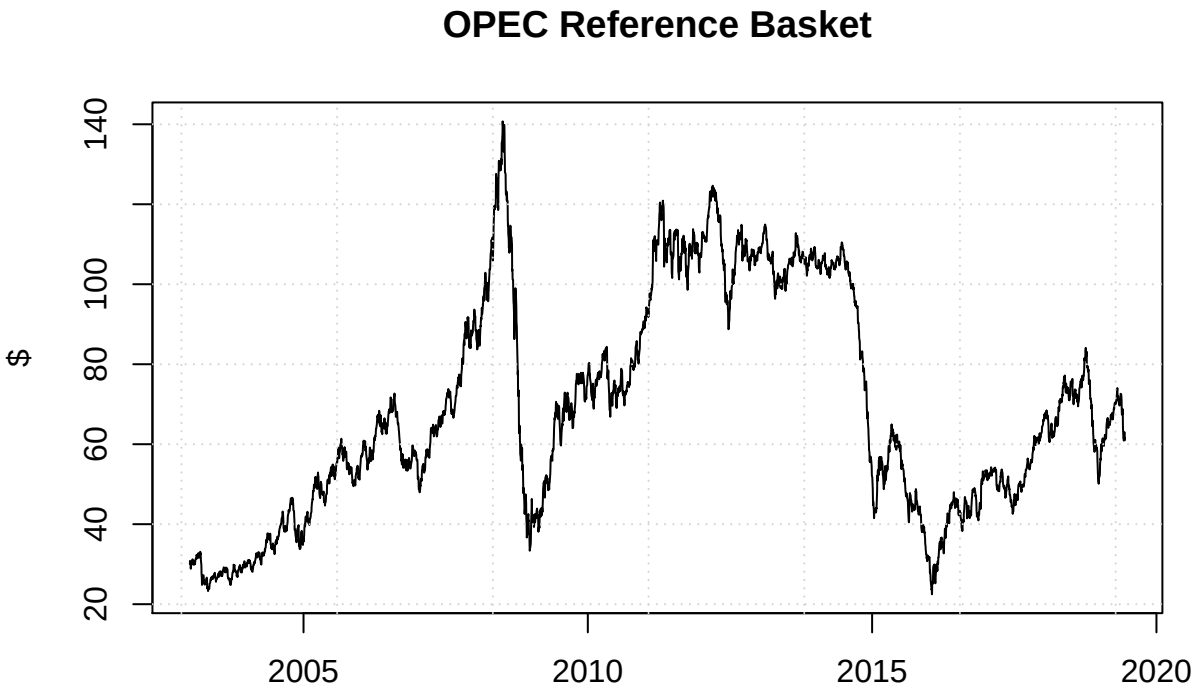
Examples use: https://cran.r-project.org/web/packages/imfr/README.html

Quandl

Quandl is a large data aggregator that collects data from publicly accessible sources (eg. FRED) as well as contains private *alternative* data that is available for purchase (https://www.quandl.com/). To use their R package (also have a python package) you need to create a free API key.

See here for more information: https://www.quandl.com/tools/r

```
#install.packages('Quandl')
library(Quandl)
Quandl.api_key("1zm1xSnnoqFeAGksg3S1")
oil_prices = Quandl("OPEC/ORB")
plot(oil_prices$Date,oil_prices$Value,typ='l',xlab='',ylab='$',main='OPEC Reference Basket')
grid()
```



Generic Web Scraping

IEX exchange (A simple API)

The investors exchange was founded in 2012 to ban the use of high frequency trading. As they are new exchange they have been providing a variety of resources for free to entice investors to trade more on their exchange.

https://iextrading.com/developer/docs/

They have a web api that we can use to load data into R. For instance at the following URL the IEX exchange has the most recent trading day of data for Apple stock sumarized in one-minute quotations:

https://api.iextrading.com/1.0/stock/ibm/chart/date/20190129

This data is known as JSON (JavaScript Object Notation). In essence the data is formatted as combinations of keys and values for a given time stamp. We can load this type of data directly with the help of the *jsonlite* package into R as follows:

```
url <- 'https://api.iextrading.com/1.0/stock/aapl/chart/date/20190129'
#install.packages('jsonlite')
library(jsonlite)
data <- fromJSON(url)
head(data)
```

```
## list()
```

Collect data from Wikipedia tables and lists

Occasionally it is quite helpful to pull data that are in structured tables on websites. An example that I'm sure cuts across a wide variety of disciplines are the awesome tables on Wikipedia pages. For example, suppose I want to pull stock data on all of the companies that are in the S&P 500. First I would need to

know all of the 500 companies in the S&P 500 (duh). Fortunately, on this Wikipedia page there are listed the most recent companies in the S&P 500 (they change over time): https://en.wikipedia.org/wiki/List_of_S%26P_500_companies

First we need to get the HTML from the wikipedia page through the command *GET* in the *httr* package, then we can load in tables on *most* websites through the command *readHTMLTables* that comes in the *XML* package. We need the httr package to pull data from all URLs that have an HTTPS address.

```
#install.packages(c('XML','httr'))
library(XML)
library(httr)
url <- 'https://en.wikipedia.org/wiki/List_of_S%26P_500_companies'
?GET
html_data <- GET(url)
html_data$status_code
```

```
## [1] 200
```

```
html_data$status_code
```

```
## [1] 200
```

```
html_data$url
```

```
## [1] "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies"
```

```
html_text <- content(html_data,as = 'text')
wiki_tables <- readHTMLTable(html_text)
#Got it!

wiki_tables[[2]]
```

```
##                    V1                              V2
## 1                Date                           Added
## 2              Ticker                        Security
## 3       June 7, 2019                             BMS
## 4       June 3, 2019                            CTVA
## 5      April 2, 2019                             DOW
## 6   February 27, 2019                             WAB
## 7   February 15, 2019                             ATO
## 8    January 18, 2019                             TFX
## 9     January 2, 2019                             FRC
## 10  December 24, 2018                              CE
## 11   December 3, 2018                              LW
## 12               MXIM    Maxim Integrated Products Inc
## 13               FANG              Diamondback Energy
## 14  November 13, 2018                            JKHY
## 15   November 6, 2018                            KEYS
## 16   October 11, 2018                            FTNT
## 17    October 1, 2018                             ROL
## 18 September 14, 2018                             WCG
## 19    August 28, 2018                            ANET
## 20       July 2, 2018                            CPRT
## 21      June 20, 2018                             FLT
## 22      June 18, 2018                              BR
## 23               HFC                    HollyFrontier
## 24       June 7, 2018                            TWTR
## 25       June 5, 2018                            EVRG
```

```
## 26        May 31, 2018                            ABMD
## 27       April 4, 2018                            MSCI
## 28      March 19, 2018                            TTWO
## 29                SIVB                  SVB Financial
## 30                NKTR             Nektar Therapeutics
## 31       March 7, 2018                            IPGP
## 32     January 3, 2018                             HII
## 33    October 13, 2017                            NCLH
## 34  September 18, 2017                            CDNS
## 35   September 1, 2017                            SBAC
## 36     August 29, 2017                               Q
## 37      August 8, 2017
## 38      August 7, 2017                             BHF
## 39       July 26, 2017                             DRE
## 40                 AOS                       A.O. Smith
## 41                 PKG Packaging Corporation of America
## 42                 RMD                           ResMed
## 43                 MGM       MGM Resorts International
## 44       June 19, 2017                             HLT
## 45                ALGN           Align Technology Inc.
## 46                ANSS                       ANSYS Inc.
## 47                  RE                Everest Re Group
## 48        June 2, 2017                            INFO
## 49       April 5, 2017                              IT
## 50       April 4, 2017                             DXC
## 51      March 20, 2017                             AMD
## 52                 RJF          Raymond James Financial
## 53                 ARE  Alexandria Real Estate Equities
## 54      March 16, 2017                            SNPS
## 55      March 13, 2017                            DISH
## 56       March 2, 2017                             REG
## 57       March 1, 2017                            CBOE
## 58   February 28, 2017                            INCY
## 59     January 5, 2017                            IDXX
## 60    December 2, 2016                             MAA
## 61                EVHC             Envision Healthcare
## 62  September 30, 2016                            COTY
## 63  September 22, 2016                             COO
## 64   September 8, 2016                            CHTR
## 65   September 6, 2016                             MTD
## 66        July 5, 2016                             FTV
## 67        July 1, 2016                             LNT
## 68        July 1, 2016                             ALB
## 69       June 22, 2016                            FBHS
## 70       June 20, 2016                            UA-C
## 71        June 3, 2016                             TDG
## 72        May 31, 2016                             AJG
## 73        May 23, 2016                             LKQ
## 74        May 18, 2016                             DLR
## 75        May 13, 2016                             ALK
## 76         May 3, 2016                             AYI
## 77      April 25, 2016                             GPN
## 78      April 18, 2016                            ULTA
## 79       April 4, 2016                              FL
```

```
## 80      March 30, 2016                             HOLX
## 81      March 30, 2016                              CNC
## 82       March 7, 2016                              UDR
## 83       March 4, 2016                              AWK
## 84   February 22, 2016                              CXO
## 85    February 1, 2016                              CFG
## 86    February 1, 2016                              FRT
## 87    January 19, 2016                              EXR
## 88     January 5, 2016                             WLTW
## 89   December 29, 2015                              CHD
## 90   December 15, 2015
## 91    December 1, 2015                             CSRA
## 92   November 19, 2015                             ILMN
## 93   November 18, 2015                              SYF
## 94    November 2, 2015                              HPE
## 95     October 7, 2015                             VRSK
## 96  September 18, 2015                            CMCSK
## 97                FOX Twenty-First Century Fox Class B
## 98                NWS         News Corporation Class B
## 99   September 2, 2015                              UAL
## 100    August 28, 2015                             ATVI
## 101      July 29, 2015                              SIG
## 102      July 20, 2015                             PYPL
## 103       July 8, 2015                              AAP
## 104       July 6, 2015                              KHC
## 105       July 2, 2015                             CPGX
## 106       July 1, 2015                             JBHT
## 107       July 1, 2015                             BXLT
## 108      June 11, 2015                             QRVO
## 109      April 7, 2015                                O
## 110     March 23, 2015                              AAL
## 111     March 23, 2015                             EQIX
## 112                SLG                 SL Green Realty
## 113                HBI                     Hanesbrands
## 114     March 18, 2015                             HSIC
## 115     March 12, 2015                             SWKS
## 116   January 27, 2015                              HCA
## 117   January 27, 2015                             ENDP
## 118   December 5, 2014                              RCL
## 119   November 5, 2014                             LVLT
## 120 September 20, 2014                              URI
## 121                UHS       Universal Health Services
## 122    August 18, 2014                              MNK
## 123     August 6, 2014                            DISCK
## 124       July 2, 2014                              MLM
## 125       July 1, 2014                              AMG
## 126      June 20, 2014                              XEC
## 127        May 8, 2014                             AVGO
## 128        May 1, 2014                               UA
## 129        May 1, 2014                             NAVI
## 130      April 3, 2014                            GOOGL
## 131      April 2, 2014                              ESS
## 132     March 21, 2014                             GMCR
## 133   January 24, 2014                             TSCO
```

```
## 134  December 23, 2013                        ADS
## 135                 MHK          Mohawk Industries
## 136                  FB                   Facebook
## 137  December 10, 2013                         GGP
## 138   December 2, 2013                        ALLE
## 139  November 13, 2013                        KORS
## 140   October 29, 2013                         RIG
## 141 September 20, 2013                        VRTX
## 142                 AME                      Ametek
## 143 September 10, 2013                         DAL
## 144       July 8, 2013                        NLSN
## 145      June 28, 2013                        FOXA
## 146      June 21, 2013                         ZTS
## 147       June 6, 2013                          GM
## 148       May 23, 2013                         KSU
## 149        May 8, 2013                         MAC
## 150     April 30, 2013                        REGN
## 151  February 15, 2013                         PVH
## 152  December 31, 2012                        ABBV
## 153  December 21, 2012                        DLPH
## 154  December 11, 2012                        GRMN
## 155   December 3, 2012                          DG
## 156   October 10, 2012                        PETM
## 157    October 2, 2012                        KRFT
## 158    October 1, 2012                         ADT
## 159    October 1, 2012                         PNR
## 160  September 5, 2012                         LYB
## 161      July 31, 2012                         ESV
## 162       July 2, 2012                         STX
## 163      June 29, 2012                        MNST
## 164       June 5, 2012                        LRCX
## 165       May 21, 2012                        ALXN
## 166       May 17, 2012                         KMI
## 167     April 23, 2012                         PSX
## 168      April 3, 2012                        FOSL
## 169     March 13, 2012                         CCI
## 170  December 31, 2011                         WPX
## 171  December 20, 2011                        TRIP
## 172  December 16, 2011                         BWA
## 173                PRGO               Perrigo Co.
## 174                DLTR         Dollar Tree Inc.
## 175  December 12, 2011                         GAS
## 176  November 18, 2011                         CBE
## 177   October 31, 2011                         XYL
## 178   October 14, 2011                         TEL
## 179 September 23, 2011                         MOS
## 180       July 5, 2011                         ACN
## 181      June 30, 2011                         MPC
## 182       June 1, 2011                         ANR
## 183     April 27, 2011                         CMG
## 184      April 1, 2011                         BLK
## 185     March 31, 2011                          EW
## 186  February 28, 2011                         COV
## 187  February 25, 2011                         JOY
```

```
## 188  December 17, 2010                          CVC
## 189  December 17, 2010                         FFIV
## 190               NFLX                Netflix Inc.
## 191                NFX          Newfield Exploration
## 192    August 26, 2010                          TYC
## 193      July 14, 2010                           CB
## 194      June 30, 2010                          QEP
## 195      June 28, 2010                          KMX
## 196     April 29, 2010                         CERN
## 197  February 26, 2010                           HP
## 198   November 3, 2009                         PCLN
## 199 September 28, 2009                          ARG
## 200      March 3, 2009                          HRL
## 201                VTR                  Ventas Inc.
## 202      June 10, 2008                           LO
## 203  December 20, 2007                          RRC
## 204  December 13, 2007                          GME
## 205   October 25, 2007                          JEC
## 206    August 24, 2007                          LUK
## 207     March 30, 2007                          KFT
## 208   January 10, 2007                          AVB
## 209       July 1, 2005                          STZ
## 210 September 25, 2003                         ESRX
## 211   December 5, 2000                         INTU
## 212                SBL          Symbol Technologies
## 213                AYE             Allegheny Energy
## 214                ABK              Ambac Financial
##                                    V3                          V4
## 1                             Removed                      Reason
## 2                              Ticker                    Security
## 3                       Bemis Co Inc                         MAT
## 4                            Corteva                         FLR
## 5                          Dow Inc.                          BHF
## 6                 Wabtec Corporation                          GT
## 7                  Atmos Energy Corp                         NFX
## 8                           Teleflex                         PCG
## 9                First Republic Bank                         SCG
## 10                          Celanese                        ESRX
## 11          Lamb Weston Holdings Inc                         COL
## 12                               AET                  Aetna Inc
## 13                              SRCL                Stericycle
## 14            Jack Henry & Associates                         EQT
## 15              Keysight Technologies                          CA
## 16                          Fortinet                        EVHC
## 17                       Rollins Inc.                       ANDV
## 18                          WellCare                          XL
## 19                 Arista Networks                          GGP
## 20                           Copart                         DPS
## 21             FleetCor Technologies                         TWX
## 22     Broadridge Financial Solutions                         RRC
## 23                               AYI              Acuity Brands
## 24                           Twitter                         MON
## 25                            Evergy                        NAVI
## 26                           Abiomed                         WYN
```

```
## 27                                MSCI                        CSRA
## 28                 Take-Two Interactive                         SIG
## 29                                PDCO           Patterson Companies
## 30                                 CHK             Chesapeake Energy
## 31                         IPG Photonics                         SNI
## 32          Huntington Ingalls Industries                         BCR
## 33         Norwegian Cruise Line Holdings                        LVLT
## 34                 Cadence Design Systems                        SPLS
## 35                     SBA Communications                          DD
## 36                           QuintilesIMS                         WFM
## 37                                                                AN
## 38              Brighthouse Financial Inc
## 39                        Duke Realty Corp                         RIG
## 40                                BBBY           Bed Bath & Beyond
## 41                                 MUR                  Murphy Oil
## 42                                 MNK                Mallinckrodt
## 43                                 RAI             Reynolds American
## 44         Hilton Worldwide Holdings Inc.                        YHOO
## 45                                 TDC              Teradata Corp.
## 46                                   R           Ryder Systems Inc.
## 47                                 MJN                Mead Johnson
## 48                       IHS Markit Ltd.                        TGNA
## 49                           Gartner Inc                         DNB
## 50                        DXC Technology                         SWN
## 51                Advanced Micro Devices                        URBN
## 52                                 FTR       Frontier Communications
## 53                                FSLR                First Solar
## 54                             Synopsys                         HAR
## 55                          Dish Network                        LLTC
## 56            Regency Centers Corporation                        ENDP
## 57                          CBOE Holdings                         PBI
## 58                                Incyte                          SE
## 59                     IDEXX Laboratories                         STJ
## 60                 Mid-America Apartments                          OI
## 61                                   LM                 Legg Mason
## 62                             Coty, Inc.                          DO
## 63                   The Cooper Companies                         HOT
## 64                 Charter Communications                         EMC
## 65                         Mettler Toledo                         JCI
## 66                           Fortive Corp                        CPGX
## 67                     Alliant Energy Corp                         GAS
## 68                          Albemarle Corp                          TE
## 69         Fortune Brands Home & Security                         CVC
## 70                   Under Armour Class C
## 71                        TransDigm Group                        BXLT
## 72              Arthur J. Gallagher & Co.                         CCE
## 73                        LKQ Corporation                         ARG
## 74               Digital Realty Trust Inc                         TWC
## 75                    Alaska Air Group Inc                        SNDK
## 76                      Acuity Brands Inc                         ADT
## 77                    Global Payments Inc.                         GME
## 78  Ulta Salon, Cosmetics & Fragrance Inc                         THC
## 79                          Foot Locker Inc                         CAM
## 80                              Hologic Inc                         POM
```

```
## 81                  Centene Corporation                      ESV
## 82                          UDR Inc                         GMCR
## 83      American Water Works Company Inc                      CNX
## 84                  Concho Resources                         PCL
## 85           Citizens Financial Group                        PCP
## 86       Federal Realty Investment Trust                     BRCM
## 87                  Extra Space Storage                       CB
## 88                Willis Towers Watson                      FOSL
## 89                     Church & Dwight                      ALTR
## 90                                                         CMCSK
## 91                          CSRA Inc                        CSC
## 92                       Illumina Inc                       SIAL
## 93                  Synchrony Financial                     GNW
## 94            Hewlett Packard Enterprise                    HCBK
## 95                    Verisk Analytics                       JOY
## 96               Comcast Class K Special
## 97
## 98
## 99           United Continental Holdings                     HSP
## 100                 Activision Blizzard                      PLL
## 101                     Signet Jewelers                      DTV
## 102                             PayPal                        NE
## 103                  Advance Auto Parts                      FDO
## 104                        Kraft Heinz                      KRFT
## 105             Columbia Pipeline Group                      ATI
## 106                          J. B. Hunt                      TEG
## 107                             Baxalta                      QEP
## 108                               Qorvo                       LO
## 109             Realty Income Corporation                    WIN
## 110               American Airlines Group                    AGN
## 111                             Equinix                      DNR
## 112                                 NBR          Nabors Industries
## 113                                 AVP              Avon Products
## 114                        Henry Schein                       CFN
## 115             Skyworks Solutions Inc.                     PETM
## 116                       HCA Holdings                      SWY
## 117                    Endo International                     COV
## 118             Royal Caribbean Cruises                      BMS
## 119              Level 3 Communications                      JBL
## 120                      United Rentals                      BTU
## 121                                 GHC           Graham Holdings
## 122                     Mallinckrodt Plc                     RDC
## 123             Discovery Communications
## 124            Martin Marietta Materials                      X
## 125            Affiliated Managers Group                    FRX
## 126                       Cimarex Energy                    IGT
## 127                     Avago Technologies                   LSI
## 128                           Under Armour                   BEAM
## 129                                Navient                   SLM
## 130                          Google Inc.
## 131              Essex Property Trust Inc                    CLF
## 132                 Keurig Green Mountain                    WPX
## 133               Tractor Supply Company                     LIFE
## 134                Alliance Data Systems                     ANF
```

```
## 135                              JDSU                    JDS Uniphase
## 136                               TER                        Teradyne
## 137            General Growth Properties                            MOLX
## 138                          Allegion                             JCP
## 139                       Michael Kors                             NYX
## 140                         Transocean                            DELL
## 141                 Vertex Pharmaceuticals                            AMD
## 142                               SAI                            SAIC
## 143                      Delta Air Lines                             BMC
## 144                     Nielsen Holdings                               S
## 145                     21st Century Fox                            APOL
## 146                              Zoetis                             FHN
## 147                       General Motors                             HNZ
## 148                  Kansas City Southern                              DF
## 149                             Macerich                             CVH
## 150                            Regeneron                             PCS
## 151                            PVH Corp.                             BIG
## 152                                AbbVie                             FII
## 153                     Delphi Automotive                             TIE
## 154                           Garmin Ltd.                             RRD
## 155                        Dollar General                             CBE
## 156                        PetSmart, Inc.                             SUN
## 157                     Kraft Foods Group                             ANR
## 158                              ADT Corp                             LXK
## 159                           Pentair Ltd.                              DV
## 160                        LyondellBasell                            SHLD
## 161                             EnscoRowan                              GR
## 162                     Seagate Technology                             PGN
## 163                       Monster Beverage                             SLE
## 164                           Lam Research                            NVLS
## 165                 Alexion Pharmaceuticals                             MMI
## 166                           Kinder Morgan                              EP
## 167                            Phillips 66                             SVU
## 168                            Fossil, Inc.                             MHS
## 169      Crown Castle International Corp.                             CEG
## 170                      WPX Energy, Inc.                            CPWR
## 171                      TripAdvisor Inc.                            TLAB
## 172                       BorgWarner Inc.                             AKS
## 173                                  MWW         Monster Worldwide Inc.
## 174                                  WFR MEMC Electronic Materials Inc.
## 175                    AGL Resources Inc.                             GAS
## 176                     Cooper Industries                             JNS
## 177                             Xylem Inc.                             ITT
## 178                   TE Connectivity Ltd.                            CEPH
## 179                     The Mosaic Company                             NSM
## 180                          Accenture plc                              MI
## 181                Marathon Petroleum Corp                             RSH
## 182           Alpha Natural Resources, Inc.                             MEE
## 183                  Chipotle Mexican Grill                            NOVL
## 184                               BlackRock                            GENZ
## 185                   Edwards Lifesciences                               Q
## 186                           Covidien Plc                             MFE
## 187                        Joy Global Inc.                             AYE
## 188               Cablevision Systems Corp.                              KG
```

```
## 189                    F5 Networks Inc.                         EK
## 190                                ODP            Office Depot Inc.
## 191                                NYT         The New York Times Co.
## 192             Tyco International Ltd.                        SII
## 193                       Chubb Limited                        MIL
## 194                       QEP Resources                        STR
## 195                        CarMax, Inc.                        XTO
## 196                        Cerner Corp.                        BJS
## 197                   Helmerich & Payne                         RX
## 198                       Priceline.com                        SGP
## 199                          Airgas Inc                        CBE
## 200                       Hormel Foods                       ACAS
## 201                                JNY           Jones Apparel Group
## 202                     Lorillard Inc.                        ABK
## 203                    Range Resources                        TRB
## 204                          GameStop                         DJ
## 205             Jacobs Engineering Group                         AV
## 206                   Leucadia National                        KSE
## 207                        Kraft Foods                        TSG
## 208          AvalonBay Communities, Inc.                        SBL
## 209                 Constellation Brands                        GLK
## 210                       Express Scripts                       QTRN
## 211                               Intuit                         BS
## 212                                   OI            Owens-Illinois
## 213                                GRA                W.R. Grace
## 214                                CCK            Crown Holdings
##                                                                 V5
## 1                                                             <NA>
## 2                                                             <NA>
## 3                                                       Mattel Inc
## 4                                                  Fluor Corporation
## 5                                               Brighthouse Financial
## 6                                 The Goodyear Tire & Rubber Company
## 7                                            Newfield Exploration
## 8                                                    PG&E Corp
## 9                                                        SCANA
## 10                                                 Express Scripts
## 11                                                 Rockwell Collins
## 12                                          CVS acquires Aetna[15]
## 13                                  Market Capitalization change[15]
## 14                                                     EQT Corp
## 15                                                      CA Inc.
## 16                                           Envision Healthcare
## 17                                                     Andeavor
## 18                                                     XL Group
## 19                                                     GGP Inc.
## 20                                          Dr Pepper Snapple Group
## 21                                                  Time Warner
## 22                                                 Range Resources
## 23                                                            <NA>
## 24                                                      Monsanto
## 25                                                   Navient Corp
## 26                                                 Wyndham Worldwide
## 27                                                      CSRA Inc.
```

```
## 28                                                                 Signet Jewelers
## 29                                                                            <NA>
## 30                                                                            <NA>
## 31                                                  Scripps Networks Interactive
## 32                                                                      CR Bard
## 33                                                    Level 3 Communications
## 34                                                               Staples Inc.
## 35                                                                     DuPont
## 36                                                          Whole Foods Market
## 37                                                              AutoNation Inc
## 38
## 39                                                                  Transocean
## 40                                                                            <NA>
## 41                                                                            <NA>
## 42                                                                            <NA>
## 43  British American Tobacco plc (NYSE MKT:BTI) acquired Reynolds American.[36]
## 44                                                                  Yahoo! Inc.
## 45                                             Market capitalization changes.[37]
## 46                                                                            <NA>
## 47            Reckitt Benckiser Group plc acquired Mead Johnson Nutrition.[38]
## 48                                                                  Tegna, Inc.
## 49                                                            Dun & Bradstreet
## 50                                                          Southwestern Energy
## 51                                                             Urban Outfitters
## 52                                                                            <NA>
## 53                                                                            <NA>
## 54                                                        Harman Int'l Industries
## 55                                                              Linear Technology
## 56                                                         Endo International plc
## 57                                                               Pitney Bowes Inc
## 58                                                          Spectra Energy Corp
## 59                                                             St. Jude Medical
## 60                                                               Owens-Illinois
## 61                                                                            <NA>
## 62                                                      Diamond Offshore Drilling
## 63                                        Starwood Hotels & Resorts Worldwide Inc
## 64                                                                EMC Corporation
## 65                                                           Johnson Controls Inc
## 66                                                      Columbia Pipeline Group
## 67                                                                AGL Resources
## 68                                                                  TECO Energy
## 69                                                          Cablevision Systems
## 70
## 71                                                                  Baxalta Inc
## 72                                                        Coca-Cola Enterprises
## 73                                                                   Airgas Inc
## 74                                                             Time Warner Cable
## 75                                                          SanDisk Corporation
## 76                                                                    ADT Corp
## 77                                                                   GameStop
## 78                                                               Tenet Healthcare
## 79                                                          Cameron International
## 80                                                             Pepco Holdings Inc
## 81                                                                  EnscoRowan
```

```
## 82                                  Keurig Green Mountain
## 83                                          Consol Energy
## 84                                     Plum Creek Timber
## 85                                Precision Castparts Corp.
## 86                                     Broadcom Corporation
## 87                                             Chubb Corp
## 88                                           Fossil Group
## 89                                            Altera Corp
## 90                                          Comcast K Corp
## 91                                  Computer Sciences Corp
## 92                                      Sigma-Aldrich Corp
## 93                                     Genworth Financial
## 94                                   Hudson City Bancorp Inc
## 95                                            Joy Global
## 96
## 97                                                   <NA>
## 98                                                   <NA>
## 99                                           Hospira Inc
## 100                                            Pall Corp
## 101                                              DirecTV
## 102                                           Noble Corp
## 103                                Family Dollar Stores Inc.
## 104                                    Kraft Foods Group
## 105                                Allegheny Technologies
## 106                               Integrys Energy Group Inc.
## 107                                        QEP Resources
## 108                                       Lorillard Inc.
## 109                                  Windstream Holdings Inc
## 110                                        Allergan, Inc
## 111                                      Denbury Resources
## 112                                                  <NA>
## 113                                                  <NA>
## 114                                           Carefusion
## 115                                         PetSmart Inc.
## 116                                           Safeway Inc
## 117                                              Covidien
## 118                                         Bemis Company
## 119                                         Jabil Circuit
## 120                                        Peabody Energy
## 121                                                  <NA>
## 122                                  Rowan Companies plc
## 123
## 124                          United States Steel Corporation
## 125                                    Forest Laboratories
## 126                          International Game Technology
## 127                                       LSI Corporation
## 128                                             Beam Inc.
## 129                                       SLM Corporation
## 130
## 131                                Cliffs Natural Resources
## 132                                       WPX Energy, Inc.
## 133                                     Life Technologies
## 134                                  Abercrombie & Fitch
## 135                                                  <NA>
```

```
## 136                                    <NA>
## 137                              Molex Inc.
## 138                             J.C. Penney
## 139                          NYSE Euronext
## 140                              Dell, Inc.
## 141                  Advanced Micro Devices
## 142                                    <NA>
## 143                             BMC Software
## 144                        Sprint Nextel Corp.
## 145                          Apollo Group Inc.
## 146                              First Horizon
## 147                        H. J. Heinz Company
## 148                                Dean Foods
## 149                      Coventry Health Care
## 150                                 MetroPCS
## 151                              Big Lots Inc.
## 152                       Federated Investors
## 153                           Titanium Metals
## 154                            R.R. Donnelley
## 155                         Cooper Industries
## 156                               Sunoco Inc.
## 157                    Alpha Natural Resources
## 158                          Lexmark Int'l Inc
## 159                               DeVry, Inc.
## 160                 Sears Holding Corporation
## 161                      Goodrich Corporation
## 162                       Progress Energy Inc
## 163                            Sara Lee Corp.
## 164                          Novellus Systems
## 165                        Motorola Mobility
## 166                            El Paso Corp.
## 167                          Supervalu Inc.
## 168               Medco Health Solutions Inc.
## 169               Constellation Energy Group
## 170                                Compuware
## 171                           Tellabs Inc.
## 172                 AK Steel Holding Corp.
## 173                                    <NA>
## 174                                    <NA>
## 175                            Nicor Inc.
## 176                       Janus Capital Group
## 177                                ITT Corp.
## 178                           Cephalon Inc.
## 179            National Semiconductor Corp.
## 180                 Marshall & Iisley Corp.
## 181                        RadioShack Corp.
## 182                      Massey Energy Company
## 183                                Novell, Inc.
## 184                           Genzyme Corp.
## 185                      Qwest Communications
## 186                             McAfee Inc.
## 187                   Allegheny Energy Inc.
## 188                 King Pharmaceuticals Inc.
## 189                        Eastman Kodak Co.
```

```
## 190                                                                <NA>
## 191                                                                <NA>
## 192                                             Smith International Inc.
## 193                                                       Millipore Inc.
## 194                                                        Questar Corp.
## 195                                                       XTO Energy Inc.
## 196                                                   BJ Services Company
## 197                                                           IMS Health
## 198                                                 Schering-Plough Corp.
## 199                                               Cooper Industries Ltd.
## 200                                                     American Capital
## 201                                                                <NA>
## 202                                                       Ambac Financial
## 203                                                           Tribune Co.
## 204                                                             Dow Jones
## 205                                                           Avaya Inc.
## 206                                                              KeySpan
## 207                                                     Sabre Holdings
## 208                                                  Symbol Technologies
## 209                                                 Great Lakes Chemical
## 210                                              Quintiles Transnational
## 211                                                     Bethlehem Steel
## 212                                                                <NA>
## 213                                                                <NA>
## 214                                                                <NA>
##
## 1
## 2
## 3                                                                                                  Bemis r
## 4                                                                                 CTVA spun off from D
## 5
## 6                                                                                                    WAB
## 7
## 8
## 9                                                                                                  Dominio
## 10                                                                                S&P 500 constit
## 11
## 12
## 13
## 14
## 15
## 16                                                                                                 EVHC
## 17
## 18
## 19                                                                                               GGP acqui
## 20                                                                                                     DI
## 21
## 22
## 23
## 24
## 25                                                 Westar Energy (NYSE: WR) acquired Great Plains Energy (NY
## 26                                                                  Wyndham Worldwide spun off Wynn
## 27                                                                  S&P 500 constituent General Dyna
## 28
```

```
## 29
## 30
## 31                                                          Di
## 32
## 33
## 34
## 35                                                          The
## 36                                                          Amazon.
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44                   VZ acquired YHOO operations; remainder of YHOO converted to
## 45
## 46
## 47
## 48
## 49
## 50                                              HPE spins off Everett
## 51
## 52
## 53
## 54                                  Samsung Electronics Co. Ltd. acqu
## 55                              S&P 500 constituent Analog Devices Inc. (
## 56
## 57
## 58
## 59                         S&P 100 & 500 constituent Abbott Laboratories
## 60
## 61
## 62
## 63
## 64
## 65                                                          TY
## 66
## 67
## 68
## 69
## 70                                                   Under Arm
## 71
## 72
## 73
## 74
## 75
## 76
## 77                                              Global Payments i
## 78
## 79
## 80
## 81
## 82                                                JAB Holding C
```

```
## 83
## 84
## 85
## 86
## 87
## 88                                                                                               WSH
## 89
## 90
## 91
## 92                                                                                           Sigma-Al
## 93
## 94
## 95
## 96
## 97
## 98
## 99
## 100
## 101
## 102
## 103
## 104
## 105
## 106
## 107
## 108
## 109
## 110                                                                    Allergan acquired by Actavis pl
## 111
## 112
## 113
## 114                                                                                                Ca
## 115                                                                                         PetSmart a
## 116                                                                                          Safeway a
## 117
## 118
## 119
## 120
## 121
## 122
## 123
## 124
## 125                                                                                                  A
## 126
## 127
## 128
## 129
## 130
## 131
## 132
## 133                                                                            Life Technologies acqu
## 134
## 135
## 136
```

```
## 137
## 138
## 139
## 140                                                                              Founder Michael Dell and
## 141
## 142
## 143
## 144                                                                            Softbank consortium purchas
## 145                                                                               Apollo Group's market
## 146
## 147
## 148                                                                                         DF too smal
## 149
## 150                                                                                       A majority
## 151
## 152                                                                                                 A
## 153                                                                                            TIE acc
## 154
## 155
## 156                                                                                            Acquire
## 157
## 158
## 159
## 160
## 161                                                                                                Ac
## 162
## 163
## 164
## 165
## 166
## 167
## 168
## 169
## 170
## 171                                                                                               Exp
## 172
## 173
## 174
## 175                                                                                 Nicor acquired by
## 176 Janus Capital Group’s market capitalization is less than $1.2 billion and is no longer representa
## 177
## 178                                                                                    Acquired by Tev
## 179                                                                                                 A
## 180                                                                                 Marshall & Iisley
## 181
## 182  Alpha Natural Resources is acquiring Massey Energy in a deal expected to becompleted on or about
## 183
## 184
## 185
## 186
## 187
## 188
## 189
## 190
```

```
## 191
## 192
## 193
## 194                                                                    Company split
## 195
## 196
## 197
## 198
## 199
## 200
## 201
## 202
## 203
## 204
## 205
## 206
## 207
## 208
## 209
## 210
## 211
## 212
## 213
## 214
```

```
head(wiki_tables[[1]])
```

```
##       V1                   V2          V3                     V4
## 1 Symbol             Security SEC filings             GICS Sector
## 2    MMM           3M Company     reports             Industrials
## 3    ABT Abbott Laboratories     reports             Health Care
## 4   ABBV          AbbVie Inc.     reports             Health Care
## 5   ABMD          ABIOMED Inc     reports             Health Care
## 6    ACN        Accenture plc     reports Information Technology
##                                V5                       V6               V7
## 1              GICS Sub Industry   Headquarters Location Date first added
## 2       Industrial Conglomerates     St. Paul, Minnesota
## 3          Health Care Equipment North Chicago, Illinois       1964-03-31
## 4                Pharmaceuticals North Chicago, Illinois       2012-12-31
## 5          Health Care Equipment  Danvers, Massachusetts       2018-05-31
## 6 IT Consulting & Other Services         Dublin, Ireland       2011-07-06
##           V8          V9
## 1        CIK     Founded
## 2 0000066740        1902
## 3 0000001800        1888
## 4 0001551152 2013 (1888)
## 5 0000815094        1981
## 6 0001467373        1989
```

```
#Now we can do some formatting specifically to this example#
sp_companies <- wiki_tables[[1]]
sp_companies$V2
```

```
##   [1] Security
##   [2] 3M Company
```

```
##   [3] Abbott Laboratories
##   [4] AbbVie Inc.
##   [5] ABIOMED Inc
##   [6] Accenture plc
##   [7] Activision Blizzard
##   [8] Adobe Systems Inc
##   [9] Advanced Micro Devices Inc
##  [10] Advance Auto Parts
##  [11] AES Corp
##  [12] Affiliated Managers Group Inc
##  [13] AFLAC Inc
##  [14] Agilent Technologies Inc
##  [15] Air Products & Chemicals Inc
##  [16] Akamai Technologies Inc
##  [17] Alaska Air Group Inc
##  [18] Albemarle Corp
##  [19] Alexandria Real Estate Equities
##  [20] Alexion Pharmaceuticals
##  [21] Align Technology
##  [22] Allegion
##  [23] Allergan, Plc
##  [24] Alliance Data Systems
##  [25] Alliant Energy Corp
##  [26] Allstate Corp
##  [27] Alphabet Inc Class A
##  [28] Alphabet Inc Class C
##  [29] Altria Group Inc
##  [30] Amazon.com Inc.
##  [31] Amcor plc
##  [32] Ameren Corp
##  [33] American Airlines Group
##  [34] American Electric Power
##  [35] American Express Co
##  [36] American International Group
##  [37] American Tower Corp.
##  [38] American Water Works Company Inc
##  [39] Ameriprise Financial
##  [40] AmerisourceBergen Corp
##  [41] AMETEK Inc.
##  [42] Amgen Inc.
##  [43] Amphenol Corp
##  [44] Anadarko Petroleum Corp
##  [45] Analog Devices, Inc.
##  [46] ANSYS
##  [47] Anthem Inc.
##  [48] Aon plc
##  [49] A.O. Smith Corp
##  [50] Apache Corporation
##  [51] Apartment Investment & Management
##  [52] Apple Inc.
##  [53] Applied Materials Inc.
##  [54] Aptiv Plc
##  [55] Archer-Daniels-Midland Co
##  [56] Arconic Inc.
```

```
##  [57] Arista Networks
##  [58] Arthur J. Gallagher & Co.
##  [59] Assurant
##  [60] Atmos Energy Corp
##  [61] AT&T Inc.
##  [62] Autodesk Inc.
##  [63] Automatic Data Processing
##  [64] AutoZone Inc
##  [65] AvalonBay Communities, Inc.
##  [66] Avery Dennison Corp
##  [67] Baker Hughes, a GE Company
##  [68] Ball Corp
##  [69] Bank of America Corp
##  [70] The Bank of New York Mellon Corp.
##  [71] Baxter International Inc.
##  [72] BB&T Corporation
##  [73] Becton Dickinson
##  [74] Berkshire Hathaway
##  [75] Best Buy Co. Inc.
##  [76] Biogen Inc.
##  [77] BlackRock
##  [78] Block H&R
##  [79] Boeing Company
##  [80] Booking Holdings Inc
##  [81] BorgWarner
##  [82] Boston Properties
##  [83] Boston Scientific
##  [84] Bristol-Myers Squibb
##  [85] Broadcom Inc.
##  [86] Broadridge Financial Solutions
##  [87] Brown-Forman Corp.
##  [88] C. H. Robinson Worldwide
##  [89] Cabot Oil & Gas
##  [90] Cadence Design Systems
##  [91] Campbell Soup
##  [92] Capital One Financial
##  [93] Capri Holdings
##  [94] Cardinal Health Inc.
##  [95] Carmax Inc
##  [96] Carnival Corp.
##  [97] Caterpillar Inc.
##  [98] Cboe Global Markets
##  [99] CBRE Group
## [100] CBS Corp.
## [101] Celanese
## [102] Celgene Corp.
## [103] Centene Corporation
## [104] CenterPoint Energy
## [105] CenturyLink Inc
## [106] Cerner
## [107] CF Industries Holdings Inc
## [108] Charles Schwab Corporation
## [109] Charter Communications
## [110] Chevron Corp.
```

```
## [111] Chipotle Mexican Grill
## [112] Chubb Limited
## [113] Church & Dwight
## [114] CIGNA Corp.
## [115] Cimarex Energy
## [116] Cincinnati Financial
## [117] Cintas Corporation
## [118] Cisco Systems
## [119] Citigroup Inc.
## [120] Citizens Financial Group
## [121] Citrix Systems
## [122] The Clorox Company
## [123] CME Group Inc.
## [124] CMS Energy
## [125] Coca-Cola Company
## [126] Cognizant Technology Solutions
## [127] Colgate-Palmolive
## [128] Comcast Corp.
## [129] Comerica Inc.
## [130] Conagra Brands
## [131] Concho Resources
## [132] ConocoPhillips
## [133] Consolidated Edison
## [134] Constellation Brands
## [135] The Cooper Companies
## [136] Copart Inc
## [137] Corning Inc.
## [138] Corteva
## [139] Costco Wholesale Corp.
## [140] Coty, Inc
## [141] Crown Castle International Corp.
## [142] CSX Corp.
## [143] Cummins Inc.
## [144] CVS Health
## [145] D. R. Horton
## [146] Danaher Corp.
## [147] Darden Restaurants
## [148] DaVita Inc.
## [149] Deere & Co.
## [150] Delta Air Lines Inc.
## [151] Dentsply Sirona
## [152] Devon Energy
## [153] Diamondback Energy
## [154] Digital Realty Trust Inc
## [155] Discover Financial Services
## [156] Discovery Inc. Class A
## [157] Discovery Inc. Class C
## [158] Dish Network
## [159] Dollar General
## [160] Dollar Tree
## [161] Dominion Energy
## [162] Dover Corp.
## [163] Dow Inc.
## [164] DuPont de Nemours Inc
```

```
## [165] DTE Energy Co.
## [166] Duke Realty Corp
## [167] Duke Energy
## [168] DXC Technology
## [169] E*Trade
## [170] Eastman Chemical
## [171] Eaton Corporation
## [172] eBay Inc.
## [173] Ecolab Inc.
## [174] Edison Int'l
## [175] Edwards Lifesciences
## [176] Electronic Arts
## [177] Emerson Electric Company
## [178] Entergy Corp.
## [179] EOG Resources
## [180] Equifax Inc.
## [181] Equinix
## [182] Equity Residential
## [183] Essex Property Trust, Inc.
## [184] Estee Lauder Cos.
## [185] Evergy
## [186] Eversource Energy
## [187] Everest Re Group Ltd.
## [188] Exelon Corp.
## [189] Expedia Group
## [190] Expeditors
## [191] Extra Space Storage
## [192] Exxon Mobil Corp.
## [193] F5 Networks
## [194] Facebook, Inc.
## [195] Fastenal Co
## [196] Federal Realty Investment Trust
## [197] FedEx Corporation
## [198] Fidelity National Information Services
## [199] Fifth Third Bancorp
## [200] FirstEnergy Corp
## [201] First Republic Bank
## [202] Fiserv Inc
## [203] FleetCor Technologies Inc
## [204] FLIR Systems
## [205] Flowserve Corporation
## [206] FMC Corporation
## [207] Foot Locker Inc
## [208] Ford Motor
## [209] Fortinet
## [210] Fortive Corp
## [211] Fortune Brands Home & Security
## [212] Fox Corporation Class A
## [213] Fox Corporation Class B
## [214] Franklin Resources
## [215] Freeport-McMoRan Inc.
## [216] Gap Inc.
## [217] Garmin Ltd.
## [218] Gartner Inc
```

```
## [219] General Dynamics
## [220] General Electric
## [221] General Mills
## [222] General Motors
## [223] Genuine Parts
## [224] Gilead Sciences
## [225] Global Payments Inc.
## [226] Goldman Sachs Group
## [227] Grainger (W.W.) Inc.
## [228] Halliburton Co.
## [229] Hanesbrands Inc
## [230] Harley-Davidson
## [231] Harris Corporation
## [232] Hartford Financial Svc.Gp.
## [233] Hasbro Inc.
## [234] HCA Healthcare
## [235] HCP Inc.
## [236] Helmerich & Payne
## [237] Henry Schein
## [238] The Hershey Company
## [239] Hess Corporation
## [240] Hewlett Packard Enterprise
## [241] Hilton Worldwide Holdings Inc
## [242] HollyFrontier Corp
## [243] Hologic
## [244] Home Depot
## [245] Honeywell Int'l Inc.
## [246] Hormel Foods Corp.
## [247] Host Hotels & Resorts
## [248] HP Inc.
## [249] Humana Inc.
## [250] Huntington Bancshares
## [251] Huntington Ingalls Industries
## [252] IDEXX Laboratories
## [253] IHS Markit Ltd.
## [254] Illinois Tool Works
## [255] Illumina Inc
## [256] Ingersoll-Rand PLC
## [257] Intel Corp.
## [258] Intercontinental Exchange
## [259] International Business Machines
## [260] Incyte
## [261] International Paper
## [262] Interpublic Group
## [263] Intl Flavors & Fragrances
## [264] Intuit Inc.
## [265] Intuitive Surgical Inc.
## [266] Invesco Ltd.
## [267] IPG Photonics Corp.
## [268] IQVIA Holdings Inc.
## [269] Iron Mountain Incorporated
## [270] Jack Henry & Associates
## [271] Jacobs Engineering Group
## [272] J. B. Hunt Transport Services
```

```
## [273] Jefferies Financial Group
## [274] JM Smucker
## [275] Johnson & Johnson
## [276] Johnson Controls International
## [277] JPMorgan Chase & Co.
## [278] Juniper Networks
## [279] Kansas City Southern
## [280] Kellogg Co.
## [281] KeyCorp
## [282] Keysight Technologies
## [283] Kimberly-Clark
## [284] Kimco Realty
## [285] Kinder Morgan
## [286] KLA-Tencor Corp.
## [287] Kohl's Corp.
## [288] Kraft Heinz Co
## [289] Kroger Co.
## [290] L Brands Inc.
## [291] L-3 Communications Holdings
## [292] Laboratory Corp. of America Holding
## [293] Lam Research
## [294] Lamb Weston Holdings Inc
## [295] Leggett & Platt
## [296] Lennar Corp.
## [297] Lilly (Eli) & Co.
## [298] Lincoln National
## [299] Linde plc
## [300] LKQ Corporation
## [301] Lockheed Martin Corp.
## [302] Loews Corp.
## [303] Lowe's Cos.
## [304] LyondellBasell
## [305] M&T Bank Corp.
## [306] Macerich
## [307] Macy's Inc.
## [308] Marathon Oil Corp.
## [309] Marathon Petroleum
## [310] Marriott Int'l.
## [311] Marsh & McLennan
## [312] Martin Marietta Materials
## [313] Masco Corp.
## [314] Mastercard Inc.
## [315] McCormick & Co.
## [316] Maxim Integrated Products Inc
## [317] McDonald's Corp.
## [318] McKesson Corp.
## [319] Medtronic plc
## [320] Merck & Co.
## [321] MetLife Inc.
## [322] Mettler Toledo
## [323] MGM Resorts International
## [324] Microchip Technology
## [325] Micron Technology
## [326] Microsoft Corp.
```

```
## [327] Mid-America Apartments
## [328] Mohawk Industries
## [329] Molson Coors Brewing Company
## [330] Mondelez International
## [331] Monster Beverage
## [332] Moody's Corp
## [333] Morgan Stanley
## [334] The Mosaic Company
## [335] Motorola Solutions Inc.
## [336] MSCI Inc
## [337] Mylan N.V.
## [338] Nasdaq, Inc.
## [339] National Oilwell Varco Inc.
## [340] Nektar Therapeutics
## [341] NetApp
## [342] Netflix Inc.
## [343] Newell Brands
## [344] Newmont Goldcorp
## [345] News Corp. Class A
## [346] News Corp. Class B
## [347] NextEra Energy
## [348] Nielsen Holdings
## [349] Nike
## [350] NiSource Inc.
## [351] Noble Energy Inc
## [352] Nordstrom
## [353] Norfolk Southern Corp.
## [354] Northern Trust Corp.
## [355] Northrop Grumman
## [356] Norwegian Cruise Line Holdings
## [357] NRG Energy
## [358] Nucor Corp.
## [359] Nvidia Corporation
## [360] O'Reilly Automotive
## [361] Occidental Petroleum
## [362] Omnicom Group
## [363] ONEOK
## [364] Oracle Corp.
## [365] PACCAR Inc.
## [366] Packaging Corporation of America
## [367] Parker-Hannifin
## [368] Paychex Inc.
## [369] PayPal
## [370] Pentair plc
## [371] People's United Financial
## [372] PepsiCo Inc.
## [373] PerkinElmer
## [374] Perrigo
## [375] Pfizer Inc.
## [376] Philip Morris International
## [377] Phillips 66
## [378] Pinnacle West Capital
## [379] Pioneer Natural Resources
## [380] PNC Financial Services
```

```
## [381] PPG Industries
## [382] PPL Corp.
## [383] Principal Financial Group
## [384] Procter & Gamble
## [385] Progressive Corp.
## [386] Prologis
## [387] Prudential Financial
## [388] Public Serv. Enterprise Inc.
## [389] Public Storage
## [390] Pulte Homes Inc.
## [391] PVH Corp.
## [392] Qorvo
## [393] Quanta Services Inc.
## [394] QUALCOMM Inc.
## [395] Quest Diagnostics
## [396] Ralph Lauren Corporation
## [397] Raymond James Financial Inc.
## [398] Raytheon Co.
## [399] Realty Income Corporation
## [400] Red Hat Inc.
## [401] Regency Centers Corporation
## [402] Regeneron
## [403] Regions Financial Corp.
## [404] Republic Services Inc
## [405] ResMed
## [406] Robert Half International
## [407] Rockwell Automation Inc.
## [408] Rollins Inc.
## [409] Roper Technologies
## [410] Ross Stores
## [411] Royal Caribbean Cruises Ltd
## [412] Salesforce.com
## [413] SBA Communications
## [414] Schlumberger Ltd.
## [415] Seagate Technology
## [416] Sealed Air
## [417] Sempra Energy
## [418] Sherwin-Williams
## [419] Simon Property Group Inc
## [420] Skyworks Solutions
## [421] SL Green Realty
## [422] Snap-on
## [423] Southern Co.
## [424] Southwest Airlines
## [425] S&P Global, Inc.
## [426] Stanley Black & Decker
## [427] Starbucks Corp.
## [428] State Street Corp.
## [429] Stryker Corp.
## [430] SunTrust Banks
## [431] SVB Financial
## [432] Symantec Corp.
## [433] Synchrony Financial
## [434] Synopsys Inc.
```

```
## [435] Sysco Corp.
## [436] T. Rowe Price Group
## [437] Take-Two Interactive
## [438] Tapestry, Inc.
## [439] Target Corp.
## [440] TE Connectivity Ltd.
## [441] TechnipFMC
## [442] Teleflex
## [443] Texas Instruments
## [444] Textron Inc.
## [445] Thermo Fisher Scientific
## [446] Tiffany & Co.
## [447] Twitter, Inc.
## [448] TJX Companies Inc.
## [449] Torchmark Corp.
## [450] Total System Services
## [451] Tractor Supply Company
## [452] TransDigm Group
## [453] The Travelers Companies Inc.
## [454] TripAdvisor
## [455] Tyson Foods
## [456] UDR Inc
## [457] Ulta Beauty
## [458] U.S. Bancorp
## [459] Under Armour Class A
## [460] Under Armour Class C
## [461] Union Pacific
## [462] United Continental Holdings
## [463] United Health Group Inc.
## [464] United Parcel Service
## [465] United Rentals, Inc.
## [466] United Technologies
## [467] Universal Health Services, Inc.
## [468] Unum Group
## [469] V.F. Corp.
## [470] Valero Energy
## [471] Varian Medical Systems
## [472] Ventas Inc
## [473] Verisign Inc.
## [474] Verisk Analytics
## [475] Verizon Communications
## [476] Vertex Pharmaceuticals Inc
## [477] Viacom Inc.
## [478] Visa Inc.
## [479] Vornado Realty Trust
## [480] Vulcan Materials
## [481] Wabtec Corporation
## [482] Walmart
## [483] Walgreens Boots Alliance
## [484] The Walt Disney Company
## [485] Waste Management Inc.
## [486] Waters Corporation
## [487] Wec Energy Group Inc
## [488] WellCare
```

```
## [489] Wells Fargo
## [490] Welltower Inc.
## [491] Western Digital
## [492] Western Union Co
## [493] WestRock
## [494] Weyerhaeuser
## [495] Whirlpool Corp.
## [496] Williams Cos.
## [497] Willis Towers Watson
## [498] Wynn Resorts Ltd
## [499] Xcel Energy Inc
## [500] Xerox
## [501] Xilinx
## [502] Xylem Inc.
## [503] Yum! Brands Inc
## [504] Zimmer Biomet Holdings
## [505] Zions Bancorp
## [506] Zoetis
## 506 Levels: 3M Company A.O. Smith Corp Abbott Laboratories ... Zoetis
```

```
sp_companies <- apply(sp_companies,2,as.character)

colnames(sp_companies) <- as.character(sp_companies[1,]) #Set colnames to first row
sp_companies <- data.frame(sp_companies,stringsAsFactors = F)
sp_companies <- sp_companies[-1,] #Remove the first row
head(sp_companies)
```

```
##   Symbol            Security SEC.filings            GICS.Sector
## 2    MMM          3M Company     reports            Industrials
## 3    ABT Abbott Laboratories     reports            Health Care
## 4   ABBV         AbbVie Inc.     reports            Health Care
## 5   ABMD         ABIOMED Inc     reports            Health Care
## 6    ACN       Accenture plc     reports Information Technology
## 7   ATVI Activision Blizzard     reports Communication Services
##                GICS.Sub.Industry    Headquarters.Location Date.first.added
## 2       Industrial Conglomerates      St. Paul, Minnesota
## 3          Health Care Equipment  North Chicago, Illinois       1964-03-31
## 4                Pharmaceuticals  North Chicago, Illinois       2012-12-31
## 5          Health Care Equipment   Danvers, Massachusetts       2018-05-31
## 6 IT Consulting & Other Services          Dublin, Ireland       2011-07-06
## 7 Interactive Home Entertainment Santa Monica, California       2015-08-31
##          CIK     Founded
## 2 0000066740        1902
## 3 0000001800        1888
## 4 0001551152 2013 (1888)
## 5 0000815094        1981
## 6 0001467373        1989
## 7 0000718877        2008
```

How might we turn the above code into a function that we can pass in a url and extract all of the html tables on the page?

```
#Make a function to get all html tables
```

Now lets take a look at extracting lists from Wikipedia pages, to keep our Psych friends happy lets pull data on Mental Disorders:

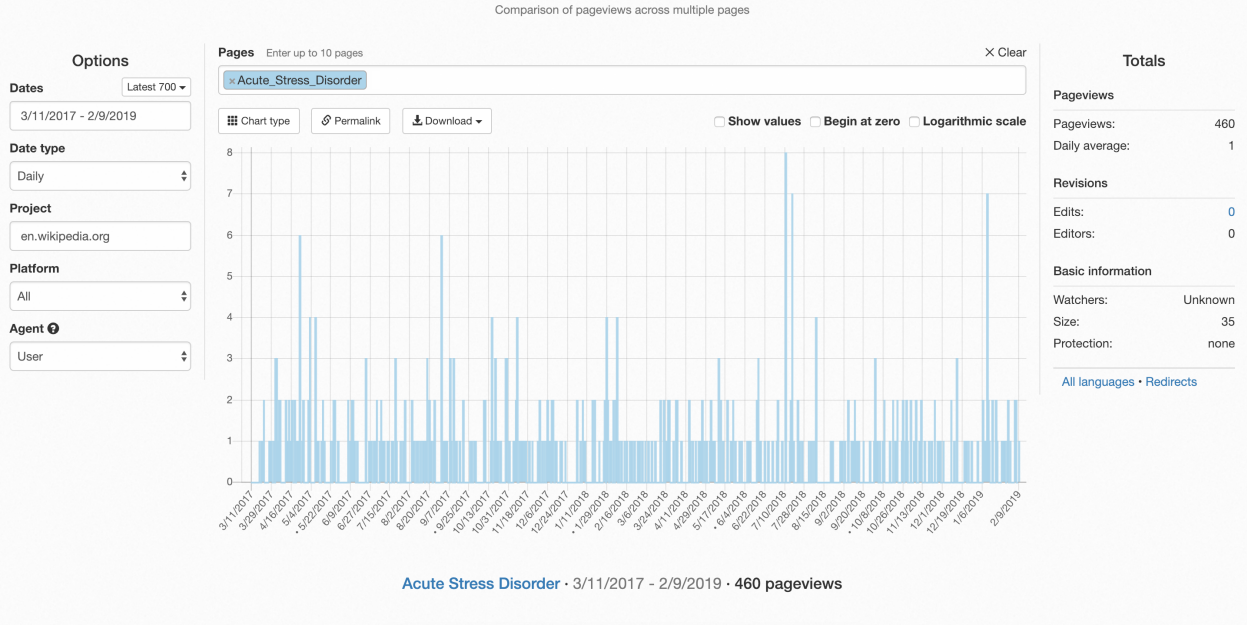


Figure 3: Accute Stress Disorder page views from 03/11/2017 to 02/09/2017.

https://en.wikipedia.org/wiki/List_of_mental_disorders

Similar to the command *readHTMLTables* there is a command called *readHTMLList* in the *XML* package that we can use to collect lists:

```
url <- 'https://en.wikipedia.org/wiki/List_of_mental_disorders'
html_data <- GET(url) #httr
html_text <- content(html_data,as = 'text')#httr
wiki_lists <- readHTMLList(html_text) #XML
length(wiki_lists)
```

```
## [1] 44
```

```
#wiki_lists

main_disorders <- unlist(wiki_lists[2:20])
```

Collect Wikipedia Page Views

An excellent tool by Wikipedia provides data on how frequently individuals view a particular wikipedia page:

https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics#Pageviews_Analysis

https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&range=latest-20&pages=Cat%7CDog

We have collected quite a long list of 190 mental disorders listed on Wikipedia. We can use the Wikipedia page views tool along with our list of disorders to see how frequently individuals look up specific disorders in the past two years.

For example, the accute stress disorder page has recieved 460 page views:

In most web browsers you have *Development tools*. Usually these tools can be accessed by right clicking on a web page and selecting 'Inspect' or 'Inspect Element'. Once the developer tools are open click on 'Console'.

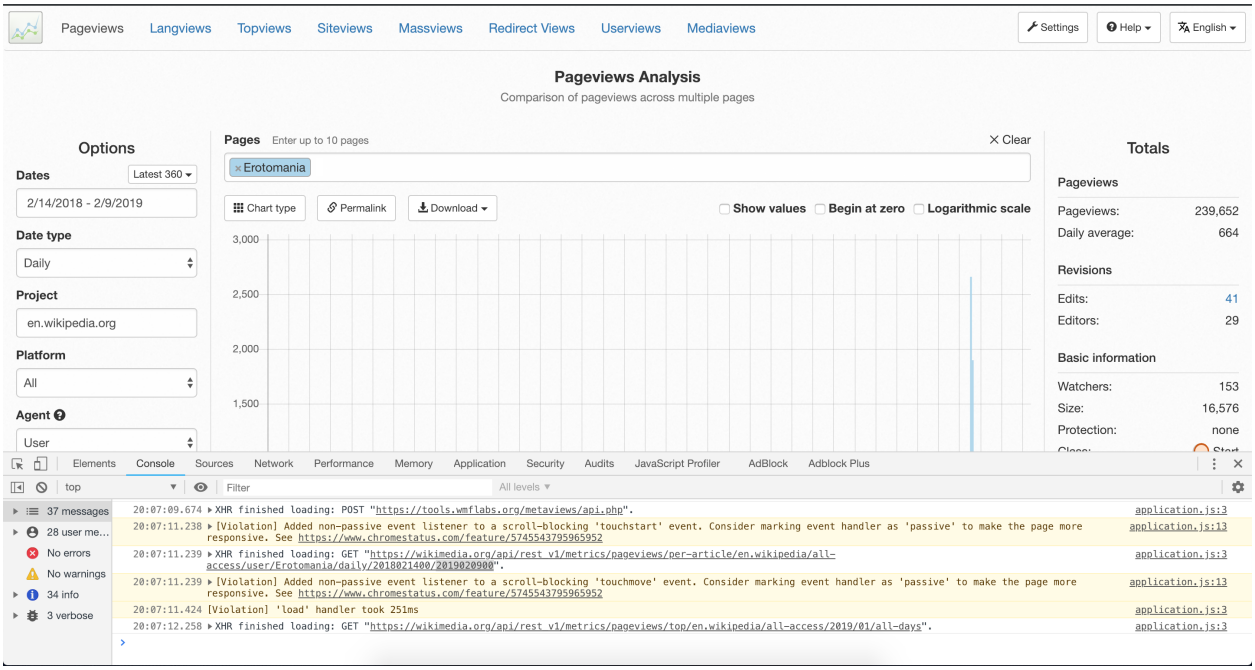


Figure 4: Erotomania page views with Google Chrome developer tools. Here we see that the website is making calls to a hidden API! We can use the API to download data directly into R.

This will show you all JavaScript requests that the website is making. Your console should look like the following:

```
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:XML':
##
##     xml
```

```
#From the developer console:
  #https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Erotoma
all_urls <- paste('https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-acc
head(all_urls)
```

```
## [1] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Acu
## [2] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ad
## [3] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ado
## [4] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Adu
## [5] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ag
## [6] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Al
```

```
all_urls <- gsub(' ','_',all_urls)
head(all_urls)
```

```
## [1] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Acu
## [2] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ad
## [3] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ado
```

```
## [4] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Adu
## [5] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Ag
## [6] "https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/user/Al
```

```
data <- fromJSON(all_urls[1])
data <- data$items
head(data)
```

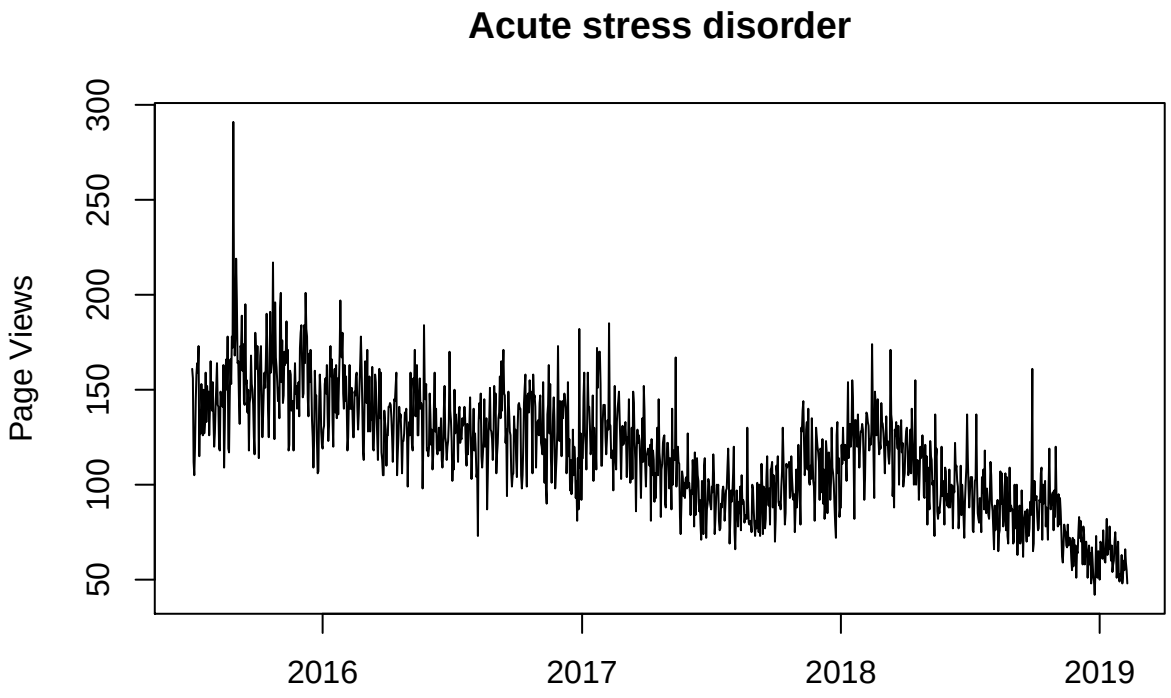
```
##        project               article granularity  timestamp     access
## 1 en.wikipedia Acute_stress_disorder       daily 2015070100 all-access
## 2 en.wikipedia Acute_stress_disorder       daily 2015070200 all-access
## 3 en.wikipedia Acute_stress_disorder       daily 2015070300 all-access
## 4 en.wikipedia Acute_stress_disorder       daily 2015070400 all-access
## 5 en.wikipedia Acute_stress_disorder       daily 2015070500 all-access
## 6 en.wikipedia Acute_stress_disorder       daily 2015070600 all-access
##   agent views
## 1  user   161
## 2  user   156
## 3  user   112
## 4  user   105
## 5  user   120
## 6  user   144
```

Notice the timestamps column is recorded as a character. We should let R know that this column is actually a date:

```
#Remove the trailing zeros
substr(data$timestamp[1],1,8)
```

```
## [1] "20150701"
```

```
data$timestamp <- substr(data$timestamp,1,8)
data$timestamp <- as.Date(data$timestamp,'%Y%m%d')
plot(data$timestamp,data$views,typ='l',main = main_disorders[1],ylab='Page Views',xlab='')
```



How might we want to extract the data for all of the urls?

```
# hold_all_data <- list()
#How much time do we expect the loop below to take to run?
# for(i in 1:length(all_urls)){
#   tryCatch({
#     data <- fromJSON(all_urls[i])
#     data <- data$items
#     data$timestamp <- substr(data$timestamp,1,8)
#     data$timestamp <- as.Date(data$timestamp,'%Y%m%d')
#     hold_all_data[[i]] <- data
#     names(hold_all_data)[i] <- main_disorders[i]
#     Sys.sleep(runif(1,1,8))
#     print(i/length(all_urls))
#   },error = function(e) print(e))
# }
#saveRDS(hold_all_data,'all_disorder_wiki_page_views.RDS')
hold_all_data <- readRDS('all_disorder_wiki_page_views.RDS')
length(hold_all_data)
```

```
## [1] 189
```

```
head(hold_all_data[[1]])
```

```
##        project               article granularity  timestamp     access
## 1 en.wikipedia Acute_stress_disorder       daily 2015-07-01 all-access
## 2 en.wikipedia Acute_stress_disorder       daily 2015-07-02 all-access
## 3 en.wikipedia Acute_stress_disorder       daily 2015-07-03 all-access
## 4 en.wikipedia Acute_stress_disorder       daily 2015-07-04 all-access
## 5 en.wikipedia Acute_stress_disorder       daily 2015-07-05 all-access
## 6 en.wikipedia Acute_stress_disorder       daily 2015-07-06 all-access
##   agent views
## 1  user   161
## 2  user   156
## 3  user   112
## 4  user   105
## 5  user   120
## 6  user   144
```

```
head(hold_all_data[[189]])
```

```
##        project          article granularity  timestamp     access agent
## 1 en.wikipedia Trichotillomania       daily 2015-07-01 all-access  user
## 2 en.wikipedia Trichotillomania       daily 2015-07-02 all-access  user
## 3 en.wikipedia Trichotillomania       daily 2015-07-03 all-access  user
## 4 en.wikipedia Trichotillomania       daily 2015-07-04 all-access  user
## 5 en.wikipedia Trichotillomania       daily 2015-07-05 all-access  user
## 6 en.wikipedia Trichotillomania       daily 2015-07-06 all-access  user
##   views
## 1  2195
## 2  2149
## 3  2015
## 4  1890
## 5  1690
## 6  2432
```

```r
all_disorder_views <- do.call('rbind',hold_all_data)
summary_stats <- tapply(all_disorder_views$views,all_disorder_views$article,summary)
summary_stats <- do.call('rbind',summary_stats)
head(summary_stats)
```

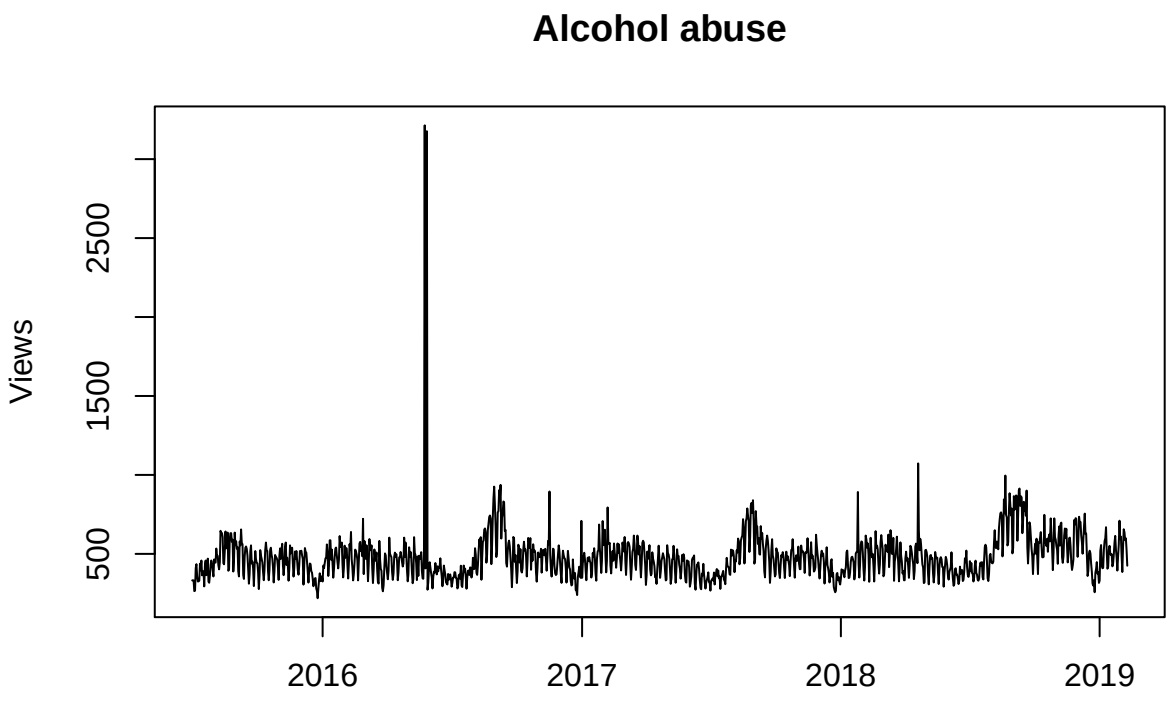
```
##                       Min. 1st Qu. Median      Mean 3rd Qu.  Max.
## Acute_stress_disorder   42      94  117.0  116.5144     138   291
## Adjustment_disorder    201     478  596.0  591.8538     694  1334
## Agoraphobia           2499    3307 3797.5 3921.5811    4300 13834
## Alcohol_abuse          219     385  465.0  478.6992     536  3214
## Alcohol_dependence     134     250  285.0  293.5795     322  1449
## Alcohol_withdrawal      52      93  106.0  109.9773     123   427
```

```r
summary_stats <- summary_stats[order(summary_stats[,3],decreasing = T),]
head(summary_stats,10)
```

```
##                                       Min. 1st Qu.  Median      Mean
## Asperger_syndrome                     6573 9019.75 10200.5 10998.523
## Bipolar_disorder                      5600 8024.75  8834.0  9134.618
## Schizophrenia                         4455 7635.00  8619.0  8770.933
## Borderline_personality_disorder       5594 7490.50  8199.0  8411.958
## Narcissistic_personality_disorder     2872 5051.75  7995.0  7638.295
## Autism                                3468 5271.50  5883.5  6114.971
## Sleep_paralysis                       2823 4610.00  5648.0  6077.911
## Alzheimer's_disease                   3195 4392.00  4840.0  4997.688
## Stockholm_syndrome                    3317 4214.50  4581.5  4986.642
## Attention_deficit_hyperactivity_disorder 2234 3635.75  4216.0  4345.894
##                                       3rd Qu.   Max.
## Asperger_syndrome                    11932.25  55405
## Bipolar_disorder                      9820.25  49638
## Schizophrenia                         9629.75 199910
## Borderline_personality_disorder       8988.75  33347
## Narcissistic_personality_disorder     9538.50  53572
## Autism                                6615.00  20843
## Sleep_paralysis                       6892.50  40558
## Alzheimer's_disease                   5396.25  39914
## Stockholm_syndrome                    5206.25  25857
## Attention_deficit_hyperactivity_disorder  4896.25  14798
```

```r
write.csv(summary_stats,'wikipedia_page_views_summary_statistics.csv')

plot(hold_all_data$`Alcohol abuse`$timestamp,
     hold_all_data$`Alcohol abuse`$views,typ = 'l',
     xlab='',ylab='Views', main = 'Alcohol abuse')
```




```
plot(hold_all_data$`Antisocial personality disorder`$timestamp,
     hold_all_data$`Antisocial personality disorder`$views,typ = 'l',
     xlab='',ylab='Views', main = 'Antisocial personality disorder')
```