**Question 1**

Register an application with the Github API here https://github.com/settings/applications. Access the API to get information on your instructors repositories (hint: this is the url you want "https://api.github.com/users/jtleek/repos"). Use this data to find the time that the datasharing repo was created. What time was it created? This tutorial may be useful (https://github.com/hadley/httr/blob/master/demo/oauth2-github.r). You may also need to run the code in the base R package and not R studio.

2012-06-21T17:28:38Z

2014-01-04T21:06:44Z

2014-02-06T16:13:11Z

2013-11-07T13:25:07Z

Support Forum: <https://class.coursera.org/getdata-015/forum/thread?thread_id=22>

Potential Answers:

* <https://github.com/bquast/Getting-Cleaning-Data/blob/master/Quiz2-Question1.R>
* <https://github.com/pachamaltese/getting-and-cleaning-data/blob/master/quiz2/desarrollo-quiz2.R>

library(httr)

oauth\_endpoints("github")

myapp <- oauth\_app("github", "3ab313538c301fa31716", "233ec0ebcdfb48f0bb0027f5c2ffef4df7876024")

#Use http://localhost:1410 as the callback url

github\_token <- oauth2.0\_token(oauth\_endpoints("github"), myapp)

req <- GET("https://api.github.com/rate\_limit", config(token = github\_token))

stop\_for\_status(req)

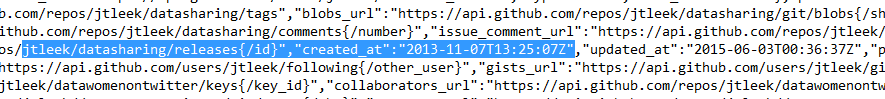
content(req)

# curl -u Access Token:x-oauth-basic "https://api.github.com/users/jtleek/repos"

BROWSE("https://api.github.com/users/jtleek/repos",authenticate("Access Token","x-oauth-basic","basic"))

# 2013-11-07T13:25:07Z

Cliff’s Work:



**Question 2**

The sqldf package allows for execution of SQL commands on R data frames. We will use the sqldf package to practice the queries we might send with the dbSendQuery command in RMySQL. Download the American Community Survey data and load it into an R object called

acs

<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv>   
  
Which of the following commands will select only the data for the probability weights pwgtp1 with ages less than 50?

sqldf("select \* from acs where AGEP < 50 and pwgtp1")

sqldf("select pwgtp1 from acs where AGEP < 50")

sqldf("select \* from acs where AGEP < 50")

sqldf("select pwgtp1 from acs")

Support Forum: <https://class.coursera.org/getdata-015/forum/thread?thread_id=23>

install.packages("sqldf")

library(sqldf)

acs <- read.csv("getdata\_data\_ss06pid.csv", header=T, sep=",")

head(acs)

sqldf("select pwgtp1 from acs where AGEP < 50")

**Question 3**

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

sqldf("select AGEP where unique from acs")

sqldf("select distinct AGEP from acs")

sqldf("select unique \* from acs")

sqldf("select unique AGEP from acs")

Support Forum:

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

sqldf("select distinct AGEP from acs")

length(unique(acs$AGEP))

#91

**Question 4**

How many characters are in the 10th, 20th, 30th and 100th lines of HTML from this page:   
<http://biostat.jhsph.edu/~jleek/contact.html>

(Hint: the nchar() function in R may be helpful)

45 92 7 2

43 99 8 6

45 31 2 25

45 0 2 2

43 99 7 25

45 31 7 25

45 31 7 31

Support Forum: <https://class.coursera.org/getdata-015/forum/thread?thread_id=24>

nchar takes a character vector as an argument and returns a vector whose elements contain the sizes of the corresponding elements of x.

hurl <- "http://biostat.jhsph.edu/~jleek/contact.html"

con <- url(hurl)

htmlCode <- readLines(con)

close(con)

sapply(htmlCode[c(10, 20, 30, 100)], nchar)

#45 31 7 25

**Question 5**

Read this data set into R and report the sum of the numbers in the fourth of the nine columns.   
  
<https://d396qusza40orc.cloudfront.net/getdata%2Fwksst8110.for>   
  
Original source of the data: <http://www.cpc.ncep.noaa.gov/data/indices/wksst8110.for>   
  
(Hint this is a fixed width file format)

28893.3

32426.7

222243.1

36.5

101.83

35824.9

Support Forum: <https://class.coursera.org/getdata-015/forum/thread?thread_id=25>

Set working directory to downloaded text file:

See R Cookbook on Kindle RE: Fixed-Width txt files

setwd("F:\\Skydrive\\Certification\\Data Science Specialization\\Class 3 - Getting and Cleaning Data\\Week 2")

# write the file url and file destination to an object  
file.url <- 'https://d396qusza40orc.cloudfront.net/getdata%2Fwksst8110.for'

file.dest <- 'getdata.for'

# download from the URL  
download.file(file.url, file.dest)

# load the data  
getdata <- read.fwf('getdata.for', skip=4, widths=c(12, 7,4, 9,4, 9,4, 9,4))



# inspect data  
head(getdata)

# calculate sum for column 4  
sum(getdata$V4)