

Question 1

Question

- 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.

Options

- (i) 2012-06-20T18:39:06Z
- (ii) 2014-03-05T16:11:46Z
- (iii) 2014-01-04T21:06:44Z
- (iv) 2013-11-07T13:25:07Z

The sqldf package

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?

- (i) `sqldf("select * from acs where AGE < 50 and pwgtp1")`
- (ii) `sqldf("select * from acs")`
- (iii) `sqldf("select * from acs where AGE < 50")`
- (iv) `sqldf("select pwgtp1 from acs where AGE < 50")`

`names(acs)`

Question 3

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

- (i) `sqldf("select unique AGEP from acs")`
- (ii) `sqldf("select distinct pwgtp1 from acs")`
- (iii) `sqldf("select AGEP where unique from acs")`
- (iv) `sqldf("select distinct AGEP from acs")`

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)

- (i) 43 99 8 6
- (ii) 45 31 7 31
- (iii) 43 99 7 25
- (iv) 45 31 7 25
- (v) 45 0 2 2
- (vi) 45 31 2 25
- (vii) 45 92 7 2

Question 5

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

<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 (fwf) format)

- 32426.7
- 35824.9
- 222243.1
- 36.5
- 28893.3
- 101.83

```
help(read.fwf)
```