Getting and Cleaning Data

Quiz 2

Question 1.)

library(httr)

require(httpuv)

require(jsonlite)

# 1. Find OAuth settings for github:

# http://developer.github.com/v3/oauth/

oauth\_endpoints("github")

# Must create an app on GitHub to call APIs from

# use client ID as key, client secret as secret

# 2. Register an application at https://github.com/settings/applications

# Insert your values below - if secret is omitted, it will look it up in

# the GITHUB\_CONSUMER\_SECRET environmental variable.

#

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

myapp <- oauth\_app("quiz2", "ddb0d599de51ccd02f4b", secret="6af1109f6ecf442d292425087d49bb13d9bbe9c8")

# 3. Get OAuth credentials

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

# 4. Use API

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

stop\_for\_status(req)

output <- content(req)

list(output[[4]]$name, output[[4]]$created\_at)

Question 2.)

> url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06pid.csv"

> setwd("~/datasciencecoursera/Getting\_and\_Cleaning\_Data")

> download.file(url, f)

Error in download.file(url, f) : unsupported URL scheme

> download.file(url, f, method = "curl")

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

0 0 0 0 0 0 0 0 --:--:-- --:--:-- --:--:-- 0 0 10.9M 0 15051 0 0 17732 0 0:10:46 --:--:-- 0:10:46 17727 10 10.9M 10 1199k 0 0 662k 0 0:00:16 0:00:01 0:00:15 661k 19 10.9M 19 2191k 0 0 779k 0 0:00:14 0:00:02 0:00:12 779k 25 10.9M 25 2847k 0 0 741k 0 0:00:15 0:00:03 0:00:12 741k 28 10.9M 28 3231k 0 0 670k 0 0:00:16 0:00:04 0:00:12 670k 32 10.9M 32 3599k 0 0 616k 0 0:00:18 0:00:05 0:00:13 718k 37 10.9M 37 4159k 0 0 608k 0 0:00:18 0:00:06 0:00:12 589k 42 10.9M 42 4767k 0 0 609k 0 0:00:18 0:00:07 0:00:11 514k 47 10.9M 47 5343k 0 0 604k 0 0:00:18 0:00:08 0:00:10 499k 53 10.9M 53 5967k 0 0 606k 0 0:00:18 0:00:09 0:00:09 544k 58 10.9M 58 6591k 0 0 609k 0 0:00:18 0:00:10 0:00:08 601k 64 10.9M 64 7231k 0 0 612k 0 0:00:18 0:00:11 0:00:07 616k 70 10.9M 70 7855k 0 0 613k 0 0:00:18 0:00:12 0:00:06 618k 75 10.9M 75 8479k 0 0 613k 0 0:00:18 0:00:13 0:00:05 629k 80 10.9M 80 9055k 0 0 610k 0 0:00:18 0:00:14 0:00:04 618k 85 10.9M 85 9551k 0 0 604k 0 0:00:18 0:00:15 0:00:03 592k 87 10.9M 87 9839k 0 0 585k 0 0:00:19 0:00:16 0:00:03 522k 90 10.9M 90 9.8M 0 0 566k 0 0:00:19 0:00:17 0:00:02 447k 93 10.9M 93 10.1M 0 0 553k 0 0:00:20 0:00:18 0:00:02 387k 95 10.9M 95 10.4M 0 0 538k 0 0:00:20 0:00:19 0:00:01 324k 98 10.9M 98 10.7M 0 0 527k 0 0:00:21 0:00:20 0:00:01 284k100 10.9M 100 10.9M 0 0 521k 0 0:00:21 0:00:21 --:--:-- 290k

> acs <- data.table(read.csv(f))

> query1 <- sqldf("select pwgtp1 from acs where AGEP < 50")

> query2 <- sqldf("select pwgtp1 from acs")

> query3 <- sqldf("select \* from acs where AGEP < 50 and pwgtp1")

> query4 <- sqldf("select \* from acs where AGEP < 50")

Question 3.)

> unique(acs$AGEP)

[1] 43 42 16 14 29 40 15 28 30 4 1 18 37 39 3 87 70 49 45 50 60 59 61 64 35 12 19 31 9

[30] 0 33 32 20 88 53 58 69 68 48 24 27 74 56 75 17 38 55 26 23 86 81 77 7 51 13 11 82 47

[59] 46 80 21 54 78 67 22 2 76 6 71 34 10 5 65 62 63 57 52 79 83 66 25 93 8 36 41 44 84

[88] 72 73 85 89

> length( unique(acs$AGEP))

[1] 91

> query1 <- sqldf("select distinct AGEP from acs")

> length(query1)

[1] 1

> query1

AGEP

1 43

2 42

3 16

4 14

5 29

6 40

7 15

8 28

9 30

10 4

11 1

12 18

13 37

14 39

15 3

16 87

17 70

18 49

19 45

20 50

21 60

22 59

23 61

24 64

25 35

26 12

27 19

28 31

29 9

30 0

31 33

32 32

33 20

34 88

35 53

36 58

37 69

38 68

39 48

40 24

41 27

42 74

43 56

44 75

45 17

46 38

47 55

48 26

49 23

50 86

51 81

52 77

53 7

54 51

55 13

56 11

57 82

58 47

59 46

60 80

61 21

62 54

63 78

64 67

65 22

66 2

67 76

68 6

69 71

70 34

71 10

72 5

73 65

74 62

75 63

76 57

77 52

78 79

79 83

80 66

81 25

82 93

83 8

84 36

85 41

86 44

87 84

88 72

89 73

90 85

91 89

Question 4.)

> nchar(htmlCode[10])

[1] 45

> nchar(htmlCode[20])

[1] 31

> nchar(htmlCode[30])

[1] 7

> nchar(htmlCode[100])

[1] 25

Questions 5.)

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

tab <- read.fwf("getdata-wksst8110", widths = c(10, 9, 4, 9, 4, 9, 4, 9, 4 ), skip = 4)

> tail(tab["V4"])

V4

1249 24.9

1250 25.1

1251 25.2

1252 25.2

1253 25.2

1254 25.1

> sum(tab["V4"])

[1] 32426.7