

thoughtfulbloke aka David Hood

Getting and Cleaning week 4

The last week, and another week of cleaning/ manipulating data

- Part 1 : General Advice (<https://thoughtfulbloke.wordpress.com/2015/08/31/hello-world/>)
- Part 2 : Getting and Cleaning Week 1
(<https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-week-1/>)
- Part 3 : Getting and Cleaning Week 2
(<https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-week-2/>)
- Part 4 : Getting and Cleaning Week 3
(<https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-week-3/>)
- Part 5 : Getting and Cleaning Week 4
(<https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-week-4/>)
- Part 6 : Getting and Cleaning the Assignment
(<https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-the-assignment/>)

Quiz 4 advice

Question 1

The commonest mistake people make with this question is people miss that we are using the hid file (other questions have used the pid file, so be a bit careful about just reusing files from previous questions).

The second commonest mistake is people not noticing the question says the names of the data frame, but having drawn that to your attention, I hope that your experience with column names in the assignment (due the week before this quiz) should suggest some directions to explore with names()

Question 2

For this question, the kind of issues you dealt with in quiz 3 are still relevant (things like checking your data at each stage), so if you saved your work from there you should be in a good starting point. The commonest error people make here is mixing up GDP and GDP rank, if you are having problems check the information you are working with against what the question asks for.

Question 3

For this question, a few people miss that there are two questions (like Quiz 3 question 3)- what is the pattern, and how many answers. This is why the answers are both a pattern and a number.

Question 4

My interpretation is the learning objective is to get used to picking out information by search through a data source. To get there, you are probably going to want to see what is in various columns of your data by exploring what is contained in the various columns with regex, gradually refining it over several tries to pull together the ones that match. Zeroing in on the result in one go is hard, starting from “search the entire data frame for the name of the month” and refining is much easier.

Question 5

My initial advice is for Quiz 4 Question 5 is the learning objective is to get used to working with date information. As per the lecture there are a huge range of potential approaches you can choose from to approach this one (not to mention a swirl tutorial).

Also, it is `quantmod`, not `quantmode` nor `quantmood` nor `quant_mod` (a few people in every course make that mistake)

September 9, 2015
Cleaning, R

thoughtfulbloke

Coursera, Getting and

Blog at WordPress.com.