MITx: 15.071x The Analytics Edge

sandipan_dey (/dashboard)

Courseware (/courses/MITx/15.071x/1T2014/courseware)

Course Info (/courses/MITx/15.071x/1T2014/info)

Discussion (/courses/MITx/15.071x/1T2014/discussion/forum)

Progress (/courses/MITx/15.071x/1T2014/progress)

/llabus (/courses/MITx/15.071x/1T2014/4264e68418f34d839cf0b33a5da644b2/)

chedule (/courses/MITx/15.071x/1T2014/2891f8bf120945b9aa12e6601739c3e6/)

QUICK QUESTION 5 (2/2 points)

Given a corpus in R, how many commands do you need to run in R to clean up the irregularities (removing capital letters and punctuation)?

2

 $\mathbf{2}$

Help

Answer: 2

How many commands do you need to run to stem the document?

1

1

Answer: 1

EXPLANATION

In R, you can clean up the irregularities with two lines:

corpus = tm_map(corpus, tolower)

corpus = tm_map(corpus, removePunctuation)

And you can stem the document with one line:

corpus = tm_map(corpus, stemDocument)

Check

Save

Hide Answer

You have used 1 of 3 submissions



EdX is a non-profit created by founding partners Harvard and MIT whose mission is to bring the best of higher education to students of all ages anywhere in the world, wherever there is Internet access. EdX's free online MOOCs are interactive and subjects include computer science, public health, and artificial intelligence.



(http://www.meetup.com/edX-Global-Community/)



(http://www.facebook.com/EdxOnline)



(https://twitter.com/edXOnline)





(http://youtube.com/user/edxonline) © 2014 edX, some rights reserved.

Terms of Service and Honor Code - Privacy Policy (https://www.edx.org/edx-privacy-policy)