CME 195/Stats 195 Alternate HW
Due ***

Instructions:
1. Upload R script and any other files to Coursework dropbox using the filename <your SUNetid>stats195hw***.R. E.g., SUNetid "hgm7" gives filename "hgm7stats195hw***.R"
2. Follow any naming indications below.

Problem 1

It has been noted that if you start on any given wikipedia page, then click the first linked article, and repeat, you will soon arrive at the wikipedia article on philosophy. Read more at

`https://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy`.

In this assignment you will programmatically test this idea. Specifically, write a function `getting.to.philosophy(start.url,nlines,max.pages)` that, given a starting wikipedia page, will click on the initial linked article until it arrives at the philosophy article, or `max.pages` links have been pursued. The `nlines` argument specifies the maximum number of lines of each article to be read. Since you are only looking for the first link, you will typically not need to download entire articles.

There are many provisos to the "getting to philosophy" process described above. Four that you should handle are:

1. Your article should ignore links preceding the article text. E.g., the links to disambiguation pages or information about the article itself (e.g., it is a biography of a living person) usually precede the start of the article.

2. You should ignore links contained in parentheses, but take care not to exclude links to articles that themselves contain parentheses (see examples below).

3. You should ignore wikipedia's "meta" links, articles with prefixes `Wikipedia:`, `Help:`, `Template:`, and the like (see, e.g., the form of the url given above).

4. If you find a loop your function should immediately exit with an appropriate message.

Finally, you should pause for 2-3 seconds between each article pull.

For the minimal credit, your function should be able to reproduce the examples further below.

Extra Credit

Extend your function to check whether you can get to a user-specified `target.page` using the same rules as above. I.e., don't hard code the philosophy article. Can you estimate how special or not the philosophy article is among wikipedia articles, in being accessible in the way described above?

```
-->start:  Donna_Summer
[1] "Disco"
[1] "Music_genre"
[1] "Music"
[1] "Art"
```

```
[1] "Human_behavior"
[1] "Motion_(physics)"
[1] "Physics"
[1] "Natural_science"
[1] "Science"
[1] "Knowledge"
[1] "Awareness"
[1] "Conscious"
[1] "Quality_(philosophy)"
[1] "Attribute"
[1] "Property_(philosophy)"
[1] "Philosophy"

-->start:  Dorchester_Avenue_(Boston)
[1] "Boston,_Massachusetts"
[1] "List_of_capitals_in_the_United_States"
[1] "Washington_D.C."
[1] "Capital_city"
[1] "Municipality"
[1] "Administrative_division"
[1] "Country"
[1] "Political_geography"
[1] "Politics"
[1] "Power_(social_and_political)"
[1] "Social_science"
[1] "List_of_academic_disciplines_and_sub-disciplines"
[1] "Outline_(list)"
[1] "Hierarchical"
[1] "Path_(graph_theory)"
[1] "Graph_theory"
[1] "Mathematics"
[1] "Quantity"
[1] "Property_(philosophy)"
[1] "Philosophy"

> -->start:  Statistics
[1] "Analysis"
[1] "Complexity"
[1] "Model_(disambiguation)"
[1] "Conceptual_model"
[1] "Concept"
[1] "Generalization"
[1] "Concept"
[1] "-->loop"
```