Please do this quiz using an R script and submit it via Moodle. All questions should be answered using R. Make sure your code is well-organized by each question, commenting the question numbers. Avoid hardcoding. **NOTE**: You may use any functions I have provided in the lectures.

1. Download the file takotsubo.txt from Moodle and put it in your tmData folder, then create a corpus containing the documents in this file.
   1. How many documents are there in this corpus?
   2. Create a corpus meta field called search and set it to “Takotsubo”. Also create a corpus meta field called alternateName and set it to “Broken Heart Syndrome”. View the corpus meta field afterwards.
   3. Create a phraseDoc object with phrases that have a minimum frequency of 10 (this will take a few minutes). Then create a term-document matrix from this object. What is the dimension of the term-document matrix?
   4. Show the 6 phrases with the highest frequency over all documents in the corpus, and store this information in a variable called fp.
   5. For the phrase “chest pain” find all the documents that contain this phrase, and save this information with their frequencies in an R object called gd. Then find the document that has the highest frequency of this phrase. Store the index of this document in a variable called id. Note that you want the index with respect to the original term-document matrix (this will be equal to the document name)!
   6. Find all the principal phrases contained in the document you found in part e).
   7. Inspect the document you found in part e), and find its PMID (it is in the id field).
   8. Create a new phraseDoc object with the phrases in fp removed (you created fp in part d).
   9. Select the 100 most frequent phrases from the phraseDoc object created in h) and display them in a word cloud (either one). Make sure you add color!