

Name: _____

Directions: **Work only on this sheet** (on both sides, if needed); do not turn in any supplementary sheets of paper. There is actually plenty of room for your answers, as long as you organize yourself BEFORE starting writing.

10-POINT BONUS: See instructions on the blackboard.

You will write an R class, **"tfile"**, whose basis is similar to the contents of our handout **text.R**.

An object of this class will consist of three components:

- **story**, a vector of character strings, consisting of the words in the input file, in the same sequence as the file
- **distinctwords**, a vector of character strings, consisting of all the distinct words in the file
- **places**, an R list, one element per distinct word in the file, with that element being a vector of integers showing where the word occurs in the file

You will write three functions:

- **newtfile()**: the constructor, with the input file name as argument
- **places()**: argument is an object of class **"tfile"**, output is an R list as described in the **places** component above; in fact, you must have your **newtfile()** call this function
- **ctxt()**: prints the context of any word in a **"tfile"** object, printing out (on the same line) the given word and the words immediately preceding and following it in all cases in which the given word appears (if you wish, you can be sloppy and assume you'll never have a case at one end or the other of the file, which is why I got NAs below)

For convenience, assume that the class is case-insensitive and that there is no punctuation in the file. **For full credit, your code must be loop-free.**

For instance, say the input file **infile** consists of

```
how much wood
could a woodchuck chuck
if a woodchuck could chuck wood
```

Here is usage on that file:

```
> howmuch <- newtfile("infile")
> howmuch$story
[1] "how"      "much"      "wood"
"could"     "a"         "woodchuck"
[7] "chuck"    "if"        "a"
"woodchuck" "could"     "chuck"
[13] "wood"
```

```
> howmuch$distinctwords
[1] "how"      "much"      "wood"
"could"     "a"         "woodchuck"
[7] "chuck"    "if"
> howmuch$places[["could"]]
[1] 4 11
> ctxt(howmuch,"wood")
[1] "much wood could"
[1] "chuck wood NA"
```

You may find the R function **unique()** useful:

```
> unique(c(5,12,13,12,5))
[1] 5 12 13
```

Solutions:

```
# constructor
newfile <- function(tfilename) {
  tmp <- list()
  # vector of all the words, in sequence of the original text
  tmp$story <- scan(tfilename, "")
  # vector of the distinct words in the text
  tmp$distinctwords <- unique(tmp$story)
  # for each word, its various positions within the text
  tmp$places <- places(tmp)
  class(tmp) <- "tfile"
  return(tmp)
}

places <- function(tfileobj) {
  s <- tfileobj$story
  split(1:length(s), s)
}

# prints the context of any word, printing the word before and the word after
# the given one; prints on the same line; assumes not at an end
ctxt <- function(tfileobj, word) {
  pts <- tfileobj$places[[word]]
  story <- tfileobj$story
  pct <- function(i) {
    print(story[(i-1):(i+1)])
  }
  lapply(pts, pct)
}
```