HW1

Audrey Bahr

March 10, 2021

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 3.6.2

## Warning: package 'tibble' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'readr' was built under R version 3.6.2

## Warning: package 'purrr' was built under R version 3.6.2

## Warning: package 'dplyr' was built under R version 3.6.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Challenge 1

This challenge involves string manipulation on the last chapter of Charles Darwin's "On the Origin of Species."

```r
darwin <- read.delim("darwin.txt", comment.char="#", stringsAsFactors=FALSE)
```

```r
darwin_vec <- darwin$CHAPTER.XV..RECAPITULATION.AND.CONCLUSION.

# print number of paragraphs
print(length(darwin_vec))
```

```
## [1] 56
```

```r
# print 34th paragraph
print(darwin_vec[34])
```

```
## [1] "The fact, as we have seen, that all past and present organic beings can be arranged within a few
```

```r
# split vector of paragraphs into a vector of words
darwin_split <- strsplit(darwin_vec, " ") %>% unlist()
# remove punctuation
darwin_splitpunct <- gsub("[[:punct:]]","", darwin_split)

# count number of different words
length(unique(darwin_splitpunct))
```

```
## [1] 2058
```

```r
# find most common word
row <- which.max(data.frame(table(darwin_splitpunct))$Freq)
data.frame(table(darwin_splitpunct))[row,]
```

```
##      darwin_splitpunct Freq
## 1827               the  718
```

```r
# words that appear a certain number of times
sum(data.frame(table(darwin_splitpunct))$Freq == 1)
```

```
## [1] 1071
```

```r
sum(data.frame(table(darwin_splitpunct))$Freq >= 5)
```

```
## [1] 369
```

```r
# last paragraph
final_quote <- darwin_vec[56]
split_quote <- strsplit(final_quote, " ") %>% unlist()

# extract every third word
every_third <- split_quote[seq(3, length(split_quote), by = 3)]
print(every_third)
```

```
##  [1] "interesting"   "a"             "clothed"       "plants"
##  [5] "kinds,"        "singing"       "bushes,"       "insects"
##  [9] "and"           "crawling"      "damp"          "to"
## [13] "these"         "forms,"        "from"          "and"
## [17] "each"          "so"            "manner,"       "been"
## [21] "laws"          "us."           "taken"         "largest"
## [25] "Growth"        "Inheritance"   "almost"        "reproduction;"
## [29] "the"           "direct"        "the"           "life,"
## [33] "use"           "a"             "Increase"      "as"
## [37] "to"            "for"           "as"            "to"
```

```
## [41] "entailing"      "Character"      "Extinction"    "improved"
## [45] "from"           "of"             "famine"        "the"
## [49] "object"         "are"            "conceiving,"   "production"
## [53] "higher"         "follows."       "grandeur"      "view"
## [57] "with"           "powers,"        "originally"    "the"
## [61] "a"              "or"             "and"           "this"
## [65] "gone"           "according"      "fixed"         "gravity,"
## [69] "simple"         "endless"        "beautiful"     "wonderful"
## [73] "and"            "evolved."
```

```r
# sort by reverse alphabetical order
every_third <- gsub("[[:punct:]]","", every_third)
every_third <- sort(every_third, TRUE)

print(every_third)
```

```
##  [1] "wonderful"    "with"         "view"        "use"          "us"
##  [6] "to"           "to"           "to"          "this"         "these"
## [11] "the"          "the"          "the"         "the"          "taken"
## [16] "so"           "singing"      "simple"      "reproduction" "production"
## [21] "powers"       "plants"       "originally"  "or"           "of"
## [26] "object"       "manner"       "life"        "laws"         "largest"
## [31] "kinds"        "interesting"  "insects"     "Inheritance"  "Increase"
## [36] "improved"     "higher"       "Growth"      "gravity"      "grandeur"
## [41] "gone"         "from"         "from"        "forms"        "for"
## [46] "follows"      "fixed"        "famine"      "Extinction"   "evolved"
## [51] "entailing"    "endless"      "each"        "direct"       "damp"
## [56] "crawling"     "conceiving"   "clothed"     "Character"    "bushes"
## [61] "been"         "beautiful"    "as"          "as"           "are"
## [66] "and"          "and"          "and"         "and"          "almost"
## [71] "according"    "a"            "a"           "a"
```

There are 56 paragraphs and 2,058 unique words in this vector. "The" is the most common word in this chapter, appearing 718 times. There are 1071 words that appear only once, and 369 words that appear five times or more. Punctuation was removed when the frequency of words were counted, because otherwise words with punctuation attached are counted as different (e.g. condition != condition,).

Challenge 2

Average high temperatures in January for six cities.

```r
t <- c(35, 88, 42, 84, 81, 30)
city <- c("Beijing", "Lagos", "Paris", "Rio de Janeiro", "San Juan", "Toronto")

names(t) <- city

t[1:3]
```

```
## Beijing   Lagos   Paris
##      35      88      42
```

```r
t[c("Paris", "San Juan")]
```

```
##     Paris San Juan
##       42      81
```

Challenge 3

Matrix manipulation.

```
m1 <- matrix(data = sort(c(0:159), TRUE), nrow = 8, ncol = 20)
print(m1[5,2])
```

```
## [1] 147
```

```
print(m1[5:7,])
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## [1,]  155  147  139  131  123  115  107   99   91    83    75    67    59    51
## [2,]  154  146  138  130  122  114  106   98   90    82    74    66    58    50
## [3,]  153  145  137  129  121  113  105   97   89    81    73    65    57    49
##      [,15] [,16] [,17] [,18] [,19] [,20]
## [1,]    43    35    27    19    11     3
## [2,]    42    34    26    18    10     2
## [3,]    41    33    25    17     9     1
```

```
m2 <- m1[3:6,4:9]
class(m2)
```

```
## [1] "matrix"
```

```
mode(m2)
```

```
## [1] "numeric"
```

The new object, m2, is a matrix of mode numeric.

Challenge 4

Array manipulation.

```
a <- array(sort(seq(2, 800, 2), TRUE), dim = c(5, 5, 4, 4))

a[1, 1, 1, 2]
```

```
## [1] 600
```

```
a[2, 3, 2, ]
```

```
## [1] 728 528 328 128
```

```
a[1:5, 1:5, 3, 3]
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  300  290  280  270  260
## [2,]  298  288  278  268  258
## [3,]  296  286  276  266  256
## [4,]  294  284  274  264  254
## [5,]  292  282  272  262  252
```

Challenge 5

Simplified primate taxonomy represented by a multidimensional list.

```
# superfamilies
Lorisoidea <- c("Lorisidae", "Galagidae")
Lemuroidea <- c("Cheirogaleidae", "Lepilemuridae", "Indriidae", "Lemuridae", "Daubentoniidae")
Tarsioidea <- c("Tarsiidae")
Ceboidea <- c("Cebidae", "Atelidae", "Pitheciidae")
Hominoidea <- c("Hylobatidae", "Hominidae")
Cercopithecoidea <- c("Cercopithecidae")

# parvorders
Catarrhini <- list("Superfamily Hominoidea" = Hominoidea, "Superfamily Cercopithecoidea" = Cercopitheco:
Platyrrhini = list("Superfamily Ceboidea" = c(Ceboidea))

# infraorders
Lorisiformes = list("Superfamily Lorisoidea" = (Lorisoidea))
Lemuriformes = list("Superfamily Lemuroidea" = (Lemuroidea))
Tarsiiformes = list("Superfamily Tarsioidea" = (Tarsioidea))
Simiiformes = list("Parvorder: Platyrrhini" = Platyrrhini, "Parvorder: Catarrhini" = Catarrhini)

# suborders
Strepsirhini = list("Infraorder: Lorisiformes" = Lorisiformes, "Infraorder: Lemuriformes" = Lemuriformes
Haplorhini = list("Infraorder: Tarsiiformes" = (Tarsiiformes), "Infraorder: Simiiformes" = Simiiformes)

# all together
Primates = list("Suborder: Strepsirhini" = Strepsirhini, "Suborder: Haplorhini" = Haplorhini)

# extract elements
platyrrhines <- Primates[[2]][[2]][1]
class(platyrrhines)
```

```
## [1] "list"
```

```
mode(platyrrhines)
```

```
## [1] "list"
```

```
Primates$`Suborder: Haplorhini`$`Infraorder: Tarsiiformes`$`Superfamily Tarsioidea`
```

```
## [1] "Tarsiidae"
```

The class and mode of platyrrhines are both list.