Assignment 3

Irene Jacob

2020-09-07

1.

Using the 173 majors listed in fivethirtyeight.com's College Majors dataset [https://fivethirtyeight.com/features/the-economic-guide-to-picking-a-college-major/], provide code that identifies the majors that contain either "DATA" or "STATISTICS"

```
majors <- read.csv("https://raw.githubusercontent.com/datasets/five-thirty-eight-datasets/master/datase

temp <- majors$major

dt<-grepl("data", temp)       #identify majors containing DATA

temp[dt]
```

Load the data from csv file and identify the majors that contain DATA and STATISTICS

```
## [1] "computer programming and data processing"
```

```
st<-grepl("statistics", temp)   #identify majors containing STATISTICS

temp[st]
```

```
## [1] "management information systems and statistics"
## [2] "statistics and decision science"
```

2.

Write code that transforms the data below:

[1] “bell pepper” “bilberry” “blackberry” “blood orange”

[5] “blueberry” “cantaloupe” “chili pepper” “cloudberry”

[9] “elderberry” “lime” “lychee” “mulberry”

[13] “olive” “salal berry”

Into a format like this:

c(“bell pepper”, “bilberry”, “blackberry”, “blood orange”, “blueberry”, “cantaloupe”, “chili pepper”, “cloudberry”, “elderberry”, “lime”, “lychee”, “mulberry”, “olive”, “salal berry”)

```r
s1 = c("bell pepper", "bilberry", "blackberry","blood orange")
s2 = c("blueberry","cantalope","chili pepper","cloudberry")
s3 = c("elderberry","lime","lychee","mulberry")
s4 = c("olive","salal berry")

berries <- c(s1,s2,s3,s4)
print(berries)
```

```
##  [1] "bell pepper"  "bilberry"     "blackberry"   "blood orange" "blueberry"
##  [6] "cantalope"    "chili pepper" "cloudberry"   "elderberry"   "lime"
## [11] "lychee"       "mulberry"     "olive"        "salal berry"
```

3.

Describe, in words, what these expressions will match:

1. (.)\1\1 A character will be repeated thrice. Example iii.

2. “(.)(.)\2\1” One pair of characters will be repeated twice but the second time will be the reverse of the first. Example acca.

3. (..)\1 2 characters will be repeated. It can be any 2 characters. Example 0101.

4. “(.).\1.\1” Example mpmnm. A character followed by another character then the first character again, then a new character followed by the first character again.

5. "(.)(.)(.).*\3\2\1" In the beginning there are 3 characters and the same 3 characters are there in the end of the string but in the reverse order. In the middle there are zero or more characters. Example i123456721i or i1221i

4.

Construct regular expressions to match words that:

1. Start and end with the same character. "^(.).*\1$"

2. Contain a repeated pair of letters (e.g. “church” contains “ch” repeated twice.) "(.)(.).*\1\2"

3. Contain one letter repeated in at least three places (e.g. “eleven” contains three “e”s.) “(.).\ *1.*\1”