DATA 607: Assignment Three

Zachary Safir

2/17/2021

#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(url("https://raw.githubusercontent.com/fivethirtyeight/data/master/college-majors/ma
```

```
grep(pattern = 'Data|Statistics', majors$Major, value = TRUE, ignore.case = TRUE)
```

```
## [1] "MANAGEMENT INFORMATION SYSTEMS AND STATISTICS"
## [2] "COMPUTER PROGRAMMING AND DATA PROCESSING"
## [3] "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")

#3 Describe, in words, what these expressions will match:

```
(.)\1\1
"(.)(.)\\2\\1"
(..)\1
"(.).\\1.\\1"
"(.)(.)(.).*\\3\\2\\1"
```

#4 Construct regular expressions to match words that:

- Start and end with the same character.
- Contain a repeated pair of letters (e.g. "church" contains "ch" repeated twice.)
- Contain one letter repeated in at least three places (e.g. "eleven" contains three "e"s.)