[**Week 3 assignment**](https://bbhosted.cuny.edu/webapps/assignment/uploadAssignment?content_id=_54556570_1&course_id=_1979778_1&group_id=&mode=view)

*Please deliver links to an R Markdown file (in GitHub and rpubs.com) with solutions to the problems below.  You may work in a small group, but please submit separately with names of all group participants in your submission.*

#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"

#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")

The two exercises below are taken from*R for Data Science*, 14.3.5.1 in the on-line version:

#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.)