Homework 1

Direct to consumer marketing is an effective strategy to distribute agricultural and farm products to consumers. Farmers market forms an important link between farmers and consumers that helps foster farmer consumer relationships. The United States Department of Agriculture (USDA) has recognized the importance of farmers markets. Through its many programs, USDA has helped the growth of farmers markets across the country.

The data file contains the following details

a. Variables indicating the geographical location of the farmers market (lat, long, street, county, state etc.)
b. Variables indicating types of products (herbs, vegetables, seafood etc.)
c. Variables indicating type of payment accepted (cash, WIC, SNAP, SFMNP etc.)
d. Variables indicating online social media presence
e. Variables indicating date and time

The directory of farmers market across the US is given in the file Answer the following questions from the dataset (***fm.csv***) **For questions 1.2-1.3 use for loop**

1.1 **(10 points)** Compute the number of farmers market in the United States

1.2 **(10 Points)** Write a code to compute the number of farmers markets by state and arrange them in descending order of number of farmers market.

1.3 **(10 Points)** Write a code to compute the number of farmers market by cities in Massachusetts and display the top five cities.

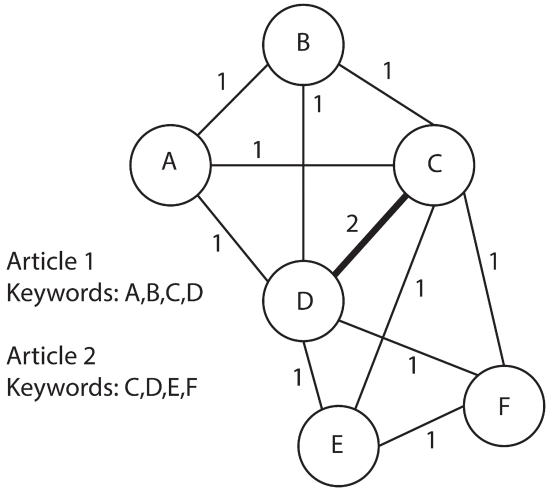
1.4 (**10 Points**) Write a code to show the top 5 states by number of farmers market that offers coffee

From the "***wine_data.csv***" answer the following question

2.1 (**20 points**) Use the "*designation*" variable and calculate the number of 20 year old wine in the dataset

In this problem, the students are expected to create a keyword co-occurrence network (KCN). Any article be it academic or otherwise is tagged with a bunch of keywords to help the article popup in database searches. Not all articles have the same type or an equal number of keywords. They vary according to the content of the article.

A KCN is created by treating each keyword as a node and each co-occurrence of a pair of words as a link between those two words (see Fig below). The number of times that a pair of words co-occurs constitutes the weight of the link connecting these two keywords. The network constructed in this manner represents a weighted network.



3.1 (**40 points**) Create a KCN network from a sample data

1. Download the dataset (https://docs.google.com/spreadsheets/d/1GTwv07i98vL7S-J9eeP8NV1fJVnymm1eJ31RDyt4Mxw/edit?usp=sharing)
2. Write an R code to extract keyword data from the above file and convert it to a weighted adjacency matrix. See the figure below to understand the process

Article 1: NEU, MIT, Harvard
Article 2: NEU, MIT, Harvard, Tufts ⟶ R Code ⟶

	NEU	MIT	Harvard	Tufts
NEU	0	2	2	1
MIT	2	0	2	1
Harvard	2	2	0	1
Tufts	1	1	1	0

Submission Format

1. Each group should submit only one homework copy
2. Submit the R markdown file
3. Each question should be in one code block
4. The codes for reading the input data must be included in the code block
5. Make sure the code follows the R coding standards (Link)