Slow Trend Report

Potential Approach:

I could read each monthly report and conduct dtm to analyze the frequent text and combine 60 times. Then, I could find the frequently associated words. However, it would be a more complicated and time-consuming approach. Thus, I chose to do the steps below and find out the trend.

Actual Approach:

In order to discover the slow trend with the given Facebook posts, I came up with two steps to detect cauliflower rice.

Step one: Co-occurrence Analysis
I used the method that I learned from homework 2, which is TCM to see the association between cauliflower and rice. I took 2014 and 2015's posts and read in corpus and created a vocabulary "ingredient" as a dictionary so that I can further create an object defining how to transform lists to token into vector space. This maps words to indices. Also, I set context window size to 3 to further calculate the cosim between rice and cauliflower.

Target 1: cauliflower
Target 2: rice

2013:

```
> sort(tcm[target1,], decreasing=TRUE)[1:5]
     soup      rice    cheese    potato     salad
149.16667 125.00002  69.66665  66.66666  61.83332
> cosim(tcm[target1,],tcm[target2,]) #0.7035128
[1] 0.5748584
>
```

2014:

```
> sort(tcm[target1,], decreasing=TRUE)[1:5]
  rice   soup cheese  pizza  salad
 272.0  164.5  114.0  102.5   93.5
> cosim(tcm[target1,],tcm[target2,]) #0.7035128
[1] 0.5082507
```

2015:

```
> sort(tcm[target1,], decreasing=TRUE)[1:5]
    rice    salad     soup   cheese    pizza
523.1669 231.4999 210.4998 193.4999 192.8333
> cosim(tcm[target1,],tcm[target2,]) #0.7035128
[1] 0.5968304
```

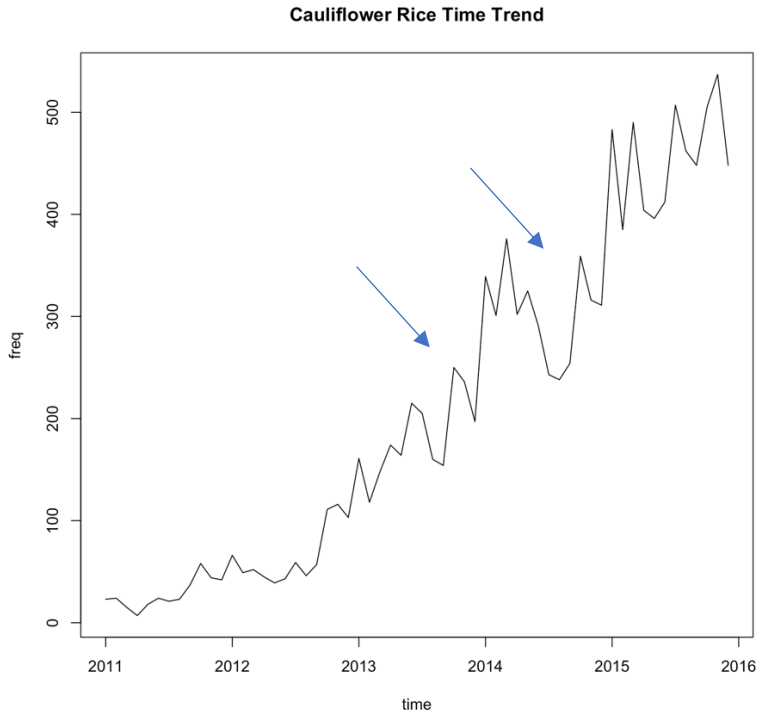
Based on the results, the most related word to cauliflower in 2014 and 2015 is rice; and in 2013, rice is the second highest related word, which indicated that there is probably a trend. Now, I know that there is an association between cauliflower and rice, so I will start to find the trend.

Step Two: Find Frequency and Plot "Cauliflower Rice"

I converted all documents to corpus and construct all the documents to Document Term Matrix as called "FoodDTM".

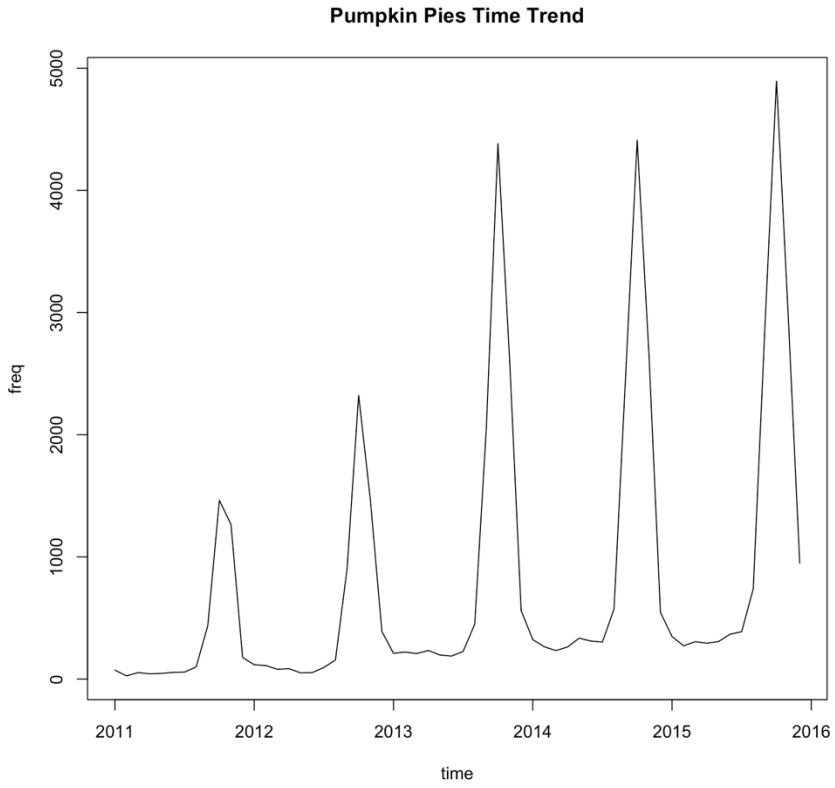
Because we need to find the time series, I create a column to indicate time and sorted as it is supposed to be. Then, I made a new dataframe with the word frequency and time series, called "data", and conduct the plot.

Therefore, we can conclude that there is a time trend for cauliflower rice. Additionally, the early stage that we are able to detect the trend is around third quarter in 2013.



Step Three: Validation Test with Pumpkin Pie

In order to ensure the accuracy of time trend model, I use “pumpkin pies” as a validation test, which is shown as below.



Thus, I am confident that my model of finding time trend with cauliflower rice is correct.