

2011-2015 Food Trend Analysis Report

Xueqing Hou (MSMA)

1. Method

I used two method to detect the number of specific ingredients and frequency of specific food.

First, I create document-term matrix to count how many times the word “cauliflower”, “zucchini” and “pumpkin”. it is important to create the document-term matrix of the corpus we are interested in. A DTM is basically a matrix, with documents designated by rows and words by columns, that the elements are the counts or the weights (usually by tf-idf). Subsequent analysis is usually based creatively on DTM.

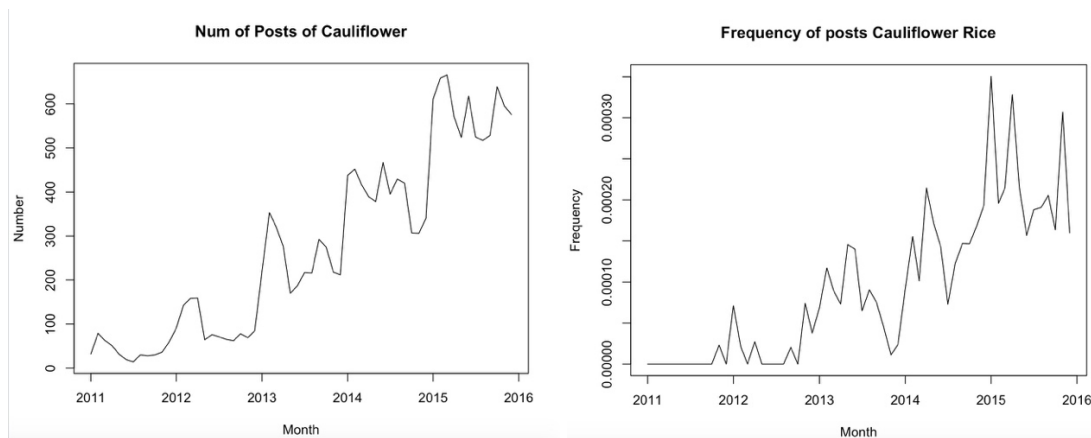
Then, I use ‘readline’ and ‘grep’ to analysis the frequency of “cauliflower rice” and “vegetable noodle” and do validation with “pumpkin pie”.

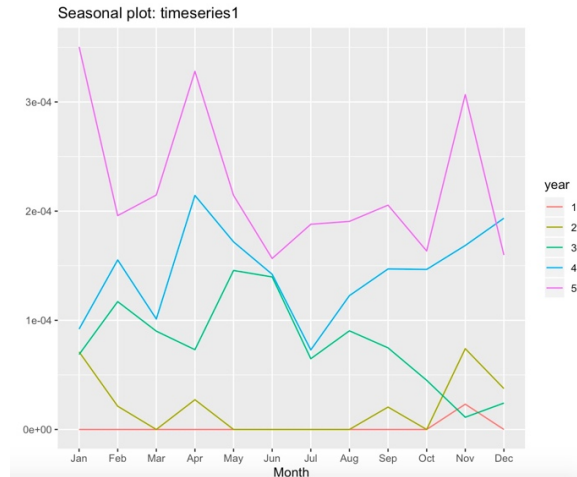
Family, I plot the seasonal trend of the three foods.

The results are showed below.

2. Cauliflower Rice Trend

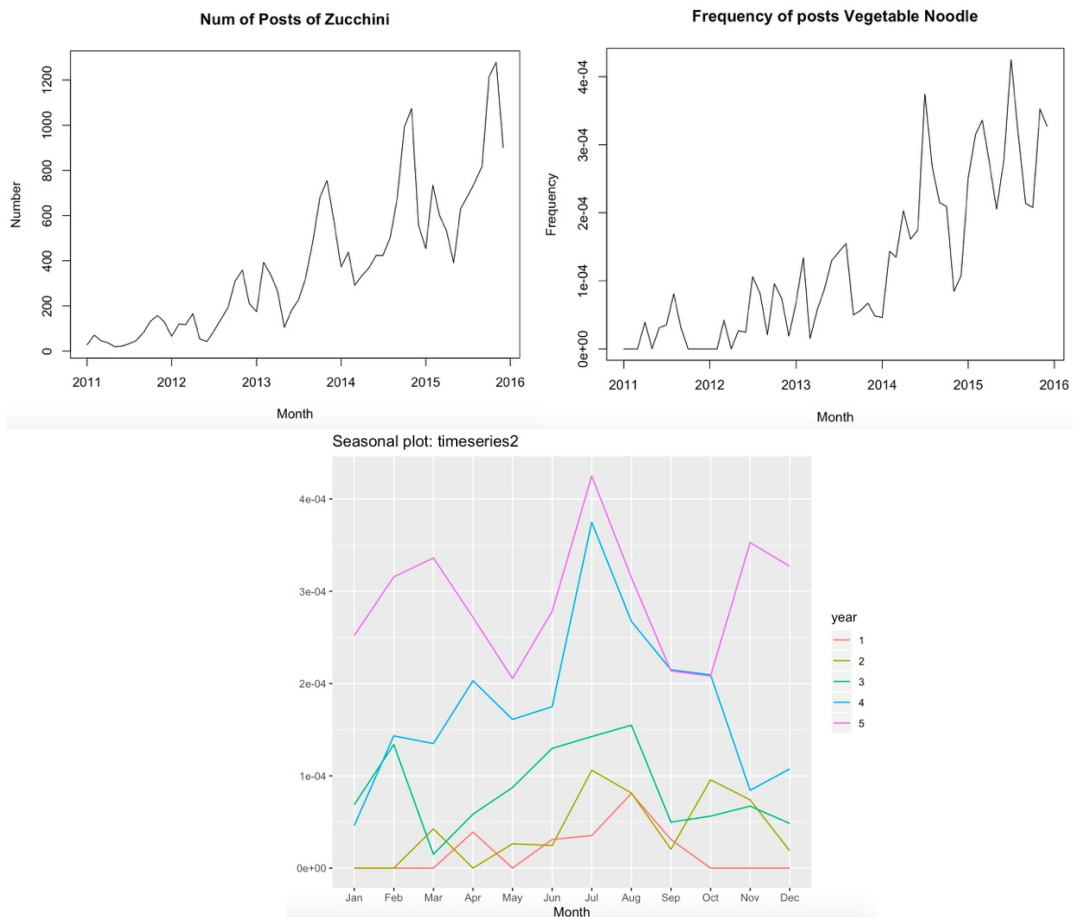
We can see that the main ingredient of the Cauliflower Rice : cauliflower, shows an increasing trend during 2011-2015 as well as the trend of the food itself. The seasonal trend shows that people posted more Cauliflower Rice in April, this may be due to the harvest time.





3. Vegetable Noodle Trend

We can see that the main ingredient of the Vegetable Noodle: zucchini, shows an increasing trend during 2011-2015 as well as the trend of the food itself. The seasonal trend shows that people posted more Vegetable Noodle in June, this may because of people like eating light in summer.



4. Pumpkin Pie trend detection

We can see the seasonal trend of pumpkin pie shows that people posted most of Pumpkin pie in November, this prove that our methods are basically correct.

