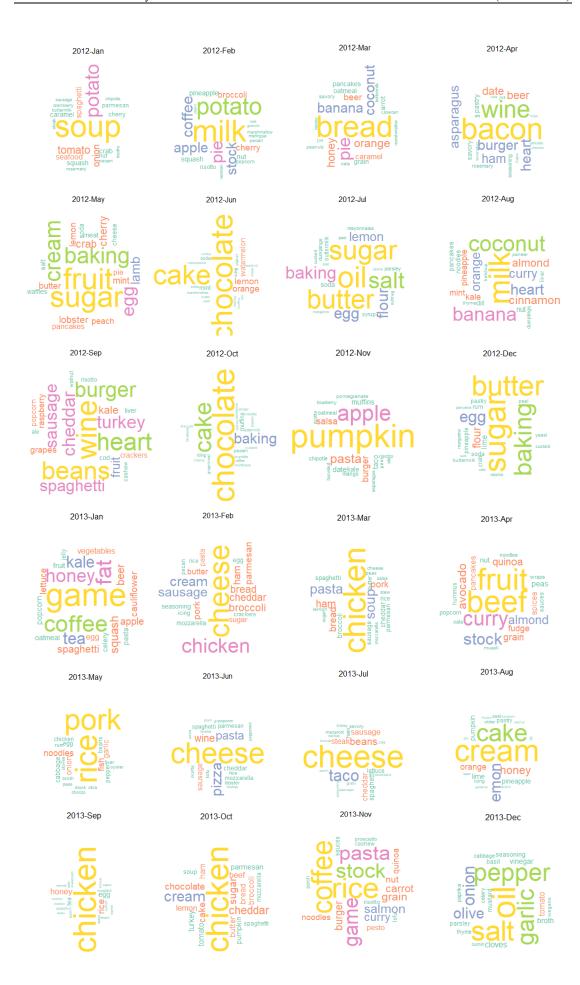# Slow Trend Project Report

In the attempt to detect emerging trend of food consumption from social media data, we'll need to extract the top food ingredients that appear the most in Facebook posts. A large proportion of mentioning depicts the emerging trend of consumers' tastes and preferences in that time period, and launching the new products using ingredients that most consumers like can be a big plus for the supermarket's profitability. The potential approaches to get the trend including calculating the percentage of occurrence of ingredients words in the Facebook posts and extracting the top trending topics through LDA model. For the first approach, we can convert the Facebook posts into document terms that shows the occurrence of each ingredient in the posts, and calculate the percentage that the ingredient is mentioned in that time period. The second approach adapts a topic model to extract the trending topics from documents. I used the second approach to depict the food trend this time.

Firstly, after the five-year monthly Facebook post data are cleaned, I converted the documents into document term matrixes and applied LDA model to get the terms. After sorting the terms based on their probabilities in each topic, I used them to generate a wordcloud that shows the top 20 ingredients mentioned in the corresponding period of time. The following shows the wordclouds I got from the model (per month for five years).
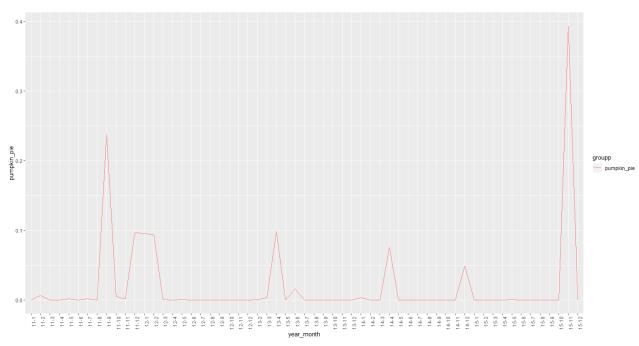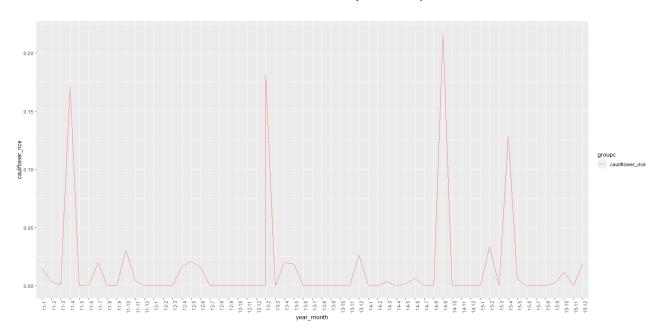
In addition to the wordclouds that show the top ingredients mentioned in Facebook posts in each month in 2011-2015, I plotted the trend of pumpkin pies and cauliflower rice in the five-year interval. The line charts of the trends are shown below:

Pumpkin Pie Trend (2011-2015)

As shown in the line chart, pumpkin pies are mentioned frequently during the months that are closed to Thanksgiving, and the topic starts to arise 1-2 month before the Thanksgiving Day.

Cauliflower Rice Trend (2011-2015)

As shown in the line chart, the tallest spike of cauliflower rice arose around Aug/Sep 2014. It could be a signal of cauliflower rice's popularization in the following years.

```
#==================CIS434 Social Media Analytics=================#
#=====================HW4 Question 2=============================#
#==================Yukun Gao (ID: 31616027)=====================#
rm(list=ls())
library(tm)
library(wordcloud)
library(topicmodels)
library(tidyr)
library(readr)
require(ggplot2)
#==================Data Cleaning=================#
#==============Facebook Posts 2011===============#
setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
Q2/fb2011")

temp = list.files(pattern="*.csv")
for (i in 1:length(temp)){
  assign(temp[i], read.csv(temp[i],sep = ',', quote = '"', header = FALSE))
}

clean_data <- function(filename,month_index){
  filename = unite(filename, text, c(1:ncol(filename)), sep = " ", remove = TRUE,
na.rm = FALSE)
  filename$doc_id = month_index
  filename = filename[c("doc_id", "text")]
}
`fpost-2011-1.csv` <- clean_data(`fpost-2011-1.csv`,1)
`fpost-2011-2.csv` <- clean_data(`fpost-2011-2.csv`,2)
`fpost-2011-3.csv` <- clean_data(`fpost-2011-3.csv`,3)
`fpost-2011-4.csv` <- clean_data(`fpost-2011-4.csv`,4)
`fpost-2011-5.csv` <- clean_data(`fpost-2011-5.csv`,5)
`fpost-2011-6.csv` <- clean_data(`fpost-2011-6.csv`,6)
`fpost-2011-7.csv` <- clean_data(`fpost-2011-7.csv`,7)
`fpost-2011-8.csv` <- clean_data(`fpost-2011-8.csv`,8)
`fpost-2011-9.csv` <- clean_data(`fpost-2011-9.csv`,9)
`fpost-2011-10.csv` <- clean_data(`fpost-2011-10.csv`,10)
`fpost-2011-11.csv` <- clean_data(`fpost-2011-11.csv`,11)
`fpost-2011-12.csv` <- clean_data(`fpost-2011-12.csv`,12)


#==============Facebook Posts 2012===============#
setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
Q2/fb2012")
```

```
temp = list.files(pattern="*.csv")
for (i in 1:length(temp)){
  assign(temp[i], read.csv(temp[i],sep = ',', quote = '"', header = FALSE))
}


`fpost-2012-1.csv` <- clean_data(`fpost-2012-1.csv`,1)
`fpost-2012-2.csv` <- clean_data(`fpost-2012-2.csv`,2)
`fpost-2012-3.csv` <- clean_data(`fpost-2012-3.csv`,3)
`fpost-2012-4.csv` <- clean_data(`fpost-2012-4.csv`,4)
`fpost-2012-5.csv` <- clean_data(`fpost-2012-5.csv`,5)
`fpost-2012-6.csv` <- clean_data(`fpost-2012-6.csv`,6)
`fpost-2012-7.csv` <- clean_data(`fpost-2012-7.csv`,7)
`fpost-2012-8.csv` <- clean_data(`fpost-2012-8.csv`,8)
`fpost-2012-9.csv` <- clean_data(`fpost-2012-9.csv`,9)
`fpost-2012-10.csv` <- clean_data(`fpost-2012-10.csv`,10)
`fpost-2012-11.csv` <- clean_data(`fpost-2012-11.csv`,11)
`fpost-2012-12.csv` <- clean_data(`fpost-2012-12.csv`,12)


#===============Facebook Posts 2013================#
setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
Q2/fb2013")

temp = list.files(pattern="*.csv")
for (i in 1:length(temp)){
  assign(temp[i], read.csv(temp[i],sep = ',', quote = '"', header = FALSE))
}


`fpost-2013-1.csv` <- clean_data(`fpost-2013-1.csv`,1)
`fpost-2013-2.csv` <- clean_data(`fpost-2013-2.csv`,2)
`fpost-2013-3.csv` <- clean_data(`fpost-2013-3.csv`,3)
`fpost-2013-4.csv` <- clean_data(`fpost-2013-4.csv`,4)
`fpost-2013-5.csv` <- clean_data(`fpost-2013-5.csv`,5)
`fpost-2013-6.csv` <- clean_data(`fpost-2013-6.csv`,6)
`fpost-2013-7.csv` <- clean_data(`fpost-2013-7.csv`,7)
`fpost-2013-8.csv` <- clean_data(`fpost-2013-8.csv`,8)
`fpost-2013-9.csv` <- clean_data(`fpost-2013-9.csv`,9)
`fpost-2013-10.csv` <- clean_data(`fpost-2013-10.csv`,10)
`fpost-2013-11.csv` <- clean_data(`fpost-2013-11.csv`,11)
`fpost-2013-12.csv` <- clean_data(`fpost-2013-12.csv`,12)


#===============Facebook Posts 2014================#
setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
Q2/fb2014")
```

```
temp = list.files(pattern="*.csv")
for (i in 1:length(temp)){
  assign(temp[i], read.csv(temp[i],sep = ',', quote = '"', header = FALSE))
}


`fpost-2014-1.csv` <- clean_data(`fpost-2014-1.csv`,1)
`fpost-2014-2.csv` <- clean_data(`fpost-2014-2.csv`,2)
`fpost-2014-3.csv` <- clean_data(`fpost-2014-3.csv`,3)
`fpost-2014-4.csv` <- clean_data(`fpost-2014-4.csv`,4)
`fpost-2014-5.csv` <- clean_data(`fpost-2014-5.csv`,5)
`fpost-2014-6.csv` <- clean_data(`fpost-2014-6.csv`,6)
`fpost-2014-7.csv` <- clean_data(`fpost-2014-7.csv`,7)
`fpost-2014-8.csv` <- clean_data(`fpost-2014-8.csv`,8)
`fpost-2014-9.csv` <- clean_data(`fpost-2014-9.csv`,9)
`fpost-2014-10.csv` <- clean_data(`fpost-2014-10.csv`,10)
`fpost-2014-11.csv` <- clean_data(`fpost-2014-11.csv`,11)
`fpost-2014-12.csv` <- clean_data(`fpost-2014-12.csv`,12)


#===============Facebook Posts 2015================#
setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
Q2/fb2015")

temp = list.files(pattern="*.csv")
for (i in 1:length(temp)){
  assign(temp[i], read.csv(temp[i],sep = ',', quote = '"', header = FALSE))
}


`fpost-2015-1.csv` <- clean_data(`fpost-2015-1.csv`,1)
`fpost-2015-2.csv` <- clean_data(`fpost-2015-2.csv`,2)
`fpost-2015-3.csv` <- clean_data(`fpost-2015-3.csv`,3)
`fpost-2015-4.csv` <- clean_data(`fpost-2015-4.csv`,4)
`fpost-2015-5.csv` <- clean_data(`fpost-2015-5.csv`,5)
`fpost-2015-6.csv` <- clean_data(`fpost-2015-6.csv`,6)
`fpost-2015-7.csv` <- clean_data(`fpost-2015-7.csv`,7)
`fpost-2015-8.csv` <- clean_data(`fpost-2015-8.csv`,8)
`fpost-2015-9.csv` <- clean_data(`fpost-2015-9.csv`,9)
`fpost-2015-10.csv` <- clean_data(`fpost-2015-10.csv`,10)
`fpost-2015-11.csv` <- clean_data(`fpost-2015-11.csv`,11)
`fpost-2015-12.csv` <- clean_data(`fpost-2015-12.csv`,12)



#===============Get the Top Food Trend for Each Month================#

setwd("C:/Users/yukun/OneDrive/UR Graduate/Fall B/CIS434 Social Media Analytics/HW4
```

```
Q2")
mydic <- tolower(scan('ingredients.txt', character(), quote = "",sep = "\n"))



Trendselector <- function(file,year,month){
  docs <- Corpus(DataframeSource(file))
  dtm <- DocumentTermMatrix(docs, control = list(dictionary=mydic,tolower=T,
stopwords = c('and',stopwords('english'))))
  idx <- rowSums(as.matrix(dtm))>0
  newdocs <- docs[idx]
  dtm = dtm[idx,]

  lda.model = LDA(dtm, 12)

  myposterior <- posterior(lda.model) # get the posterior of the model
  coins = myposterior$topics
  dices = myposterior$terms
  tid <- 2
  dice <- dices[tid, ]

  layout(matrix(c(1, 2), nrow=2), heights=c(1, 4))
  par(mar=rep(0, 4))
  plot.new()
  title = paste(year,month,sep = '-')
  text(x=0.5, y=0.1, title)
  wordcloud(names(dice), dice, max.words=20, colors=brewer.pal(6,"Set2"),
scale=c(4,.4),
          random.order=FALSE, rot.per=0.35,)
}
#===============Food Trend 2011================#
Trendselector(`fpost-2011-1.csv`,2011,'Jan')
Trendselector(`fpost-2011-2.csv`,2011,'Feb')
Trendselector(`fpost-2011-3.csv`,2011,'Mar')
Trendselector(`fpost-2011-4.csv`,2011,'Apr')
Trendselector(`fpost-2011-5.csv`,2011,'May')
Trendselector(`fpost-2011-6.csv`,2011,'Jun')
Trendselector(`fpost-2011-7.csv`,2011,'Jul')
Trendselector(`fpost-2011-8.csv`,2011,'Aug')
Trendselector(`fpost-2011-9.csv`,2011,'Sep')
Trendselector(`fpost-2011-10.csv`,2011,'Oct')
Trendselector(`fpost-2011-11.csv`,2011,'Nov')
Trendselector(`fpost-2011-12.csv`,2011,'Dec')

#===============Food Trend 2012================#
```

```
Trendselector(`fpost-2012-1.csv`,2012,'Jan')
Trendselector(`fpost-2012-2.csv`,2012,'Feb')
Trendselector(`fpost-2012-3.csv`,2012,'Mar')
Trendselector(`fpost-2012-4.csv`,2012,'Apr')
Trendselector(`fpost-2012-5.csv`,2012,'May')
Trendselector(`fpost-2012-6.csv`,2012,'Jun')
Trendselector(`fpost-2012-7.csv`,2012,'Jul')
Trendselector(`fpost-2012-8.csv`,2012,'Aug')
Trendselector(`fpost-2012-9.csv`,2012,'Sep')
Trendselector(`fpost-2012-10.csv`,2012,'Oct')
Trendselector(`fpost-2012-11.csv`,2012,'Nov')
Trendselector(`fpost-2012-12.csv`,2012,'Dec')


#===============Food Trend 2013===============#
Trendselector(`fpost-2013-1.csv`,2013,'Jan')
Trendselector(`fpost-2013-2.csv`,2013,'Feb')
Trendselector(`fpost-2013-3.csv`,2013,'Mar')
Trendselector(`fpost-2013-4.csv`,2013,'Apr')
Trendselector(`fpost-2013-5.csv`,2013,'May')
Trendselector(`fpost-2013-6.csv`,2013,'Jun')
Trendselector(`fpost-2013-7.csv`,2013,'Jul')
Trendselector(`fpost-2013-8.csv`,2013,'Aug')
Trendselector(`fpost-2013-9.csv`,2013,'Sep')
Trendselector(`fpost-2013-10.csv`,2013,'Oct')
Trendselector(`fpost-2013-11.csv`,2013,'Nov')
Trendselector(`fpost-2013-12.csv`,2013,'Dec')


#===============Food Trend 2014===============#
Trendselector(`fpost-2014-1.csv`,2014,'Jan')
Trendselector(`fpost-2014-2.csv`,2014,'Feb')
Trendselector(`fpost-2014-3.csv`,2014,'Mar')
Trendselector(`fpost-2014-4.csv`,2014,'Apr')
Trendselector(`fpost-2014-5.csv`,2014,'May')
Trendselector(`fpost-2014-6.csv`,2014,'Jun')
Trendselector(`fpost-2014-7.csv`,2014,'Jul')
Trendselector(`fpost-2014-8.csv`,2014,'Aug')
Trendselector(`fpost-2014-9.csv`,2014,'Sep')
Trendselector(`fpost-2014-10.csv`,2014,'Oct')
Trendselector(`fpost-2014-11.csv`,2014,'Nov')
Trendselector(`fpost-2014-12.csv`,2014,'Dec')


#===============Food Trend 2015===============#
Trendselector(`fpost-2015-1.csv`,2015,'Jan')
Trendselector(`fpost-2015-2.csv`,2015,'Feb')
```

```
Trendselector(`fpost-2015-3.csv`,2015,'Mar')
Trendselector(`fpost-2015-4.csv`,2015,'Apr')
Trendselector(`fpost-2015-5.csv`,2015,'May')
Trendselector(`fpost-2015-6.csv`,2015,'Jun')
Trendselector(`fpost-2015-7.csv`,2015,'Jul')
Trendselector(`fpost-2015-8.csv`,2015,'Aug')
Trendselector(`fpost-2015-9.csv`,2015,'Sep')
Trendselector(`fpost-2015-10.csv`,2015,'Oct')
Trendselector(`fpost-2015-11.csv`,2015,'Nov')
Trendselector(`fpost-2015-12.csv`,2015,'Dec')


#===============Show Trends on Particular Ingredients given Time
Index================#
docs <- Corpus(DataframeSource(`fpost-2011-1.csv` ))
dtm <- DocumentTermMatrix(docs, control = list(dictionary=mydic,tolower=T,
stopwords = c('and',stopwords('english'))))
idx <- rowSums(as.matrix(dtm))>0
newdocs <- docs[idx]
dtm = dtm[idx,]


lda.model = LDA(dtm, 12)


myposterior <- posterior(lda.model) # get the posterior of the model
coins = myposterior$topics
dices = myposterior$terms
tid <- 2
dice <- dices[tid, ]
freqterms = sort( dice, decreasing=TRUE )
p = freqterms['pumpkin'] + freqterms['pie']
c = freqterms['cauliflower'] + freqterms['rice']


trendtable <- data.frame(year_month = '11-1', pumpkin_pie = p, cauliflower_rice =
c)


get_trend <- function(filename,time){
  docs <- Corpus(DataframeSource(filename))
  dtm <- DocumentTermMatrix(docs, control = list(dictionary=mydic,tolower=T,
stopwords = c('and',stopwords('english'))))
  idx <- rowSums(as.matrix(dtm))>0
  newdocs <- docs[idx]
  dtm = dtm[idx,]


  lda.model = LDA(dtm, 12)
```

```r
  myposterior <- posterior(lda.model) # get the posterior of the model
  coins = myposterior$topics
  dices = myposterior$terms
  tid <- 2
  dice <- dices[tid, ]
  freqterms = sort( dice, decreasing=TRUE )
  p = freqterms['pumpkin'] + freqterms['pie']
  c = freqterms['cauliflower'] + freqterms['rice']
  newtable <- data.frame(year_month = time, pumpkin_pie = p, cauliflower_rice = c)
  trendtable <- rbind(trendtable, newtable)
}
#==============2011==============#
trendtable <- get_trend(`fpost-2011-2.csv`,'11-2' )
trendtable <- get_trend(`fpost-2011-3.csv`,'11-3' )
trendtable <- get_trend(`fpost-2011-4.csv`,'11-4' )
trendtable <- get_trend(`fpost-2011-5.csv`,'11-5' )
trendtable <- get_trend(`fpost-2011-6.csv`,'11-6' )
trendtable <- get_trend(`fpost-2011-7.csv`,'11-7' )
trendtable <- get_trend(`fpost-2011-8.csv`,'11-8' )
trendtable <- get_trend(`fpost-2011-9.csv`,'11-9' )
trendtable <- get_trend(`fpost-2011-10.csv`,'11-10' )
trendtable <- get_trend(`fpost-2011-11.csv`,'11-11' )
trendtable <- get_trend(`fpost-2011-12.csv`,'11-12' )
#==============2012==============#
trendtable <- get_trend(`fpost-2012-1.csv`,'12-1' )
trendtable <- get_trend(`fpost-2012-2.csv`,'12-2' )
trendtable <- get_trend(`fpost-2012-3.csv`,'12-3' )
trendtable <- get_trend(`fpost-2012-4.csv`,'12-4' )
trendtable <- get_trend(`fpost-2012-5.csv`,'12-5' )
trendtable <- get_trend(`fpost-2012-6.csv`,'12-6' )
trendtable <- get_trend(`fpost-2012-7.csv`,'12-7' )
trendtable <- get_trend(`fpost-2012-8.csv`,'12-8' )
trendtable <- get_trend(`fpost-2012-9.csv`,'12-9' )
trendtable <- get_trend(`fpost-2012-10.csv`,'12-10' )
trendtable <- get_trend(`fpost-2012-11.csv`,'12-11' )
trendtable <- get_trend(`fpost-2012-12.csv`,'12-12' )
#==============2013==============#
trendtable <- get_trend(`fpost-2013-1.csv`,'13-2' )
trendtable <- get_trend(`fpost-2013-2.csv`,'13-2' )
trendtable <- get_trend(`fpost-2013-3.csv`,'13-3' )
trendtable <- get_trend(`fpost-2013-4.csv`,'13-4' )
trendtable <- get_trend(`fpost-2013-5.csv`,'13-5' )
trendtable <- get_trend(`fpost-2013-6.csv`,'13-6' )
trendtable <- get_trend(`fpost-2013-7.csv`,'13-7' )
```

```
trendtable <- get_trend(`fpost-2013-8.csv`,'13-8' )
trendtable <- get_trend(`fpost-2013-9.csv`,'13-9' )
trendtable <- get_trend(`fpost-2013-10.csv`,'13-10' )
trendtable <- get_trend(`fpost-2013-11.csv`,'13-11' )
trendtable <- get_trend(`fpost-2013-12.csv`,'13-12' )
#==============2014===============#
trendtable <- get_trend(`fpost-2014-1.csv`,'14-1' )
trendtable <- get_trend(`fpost-2014-2.csv`,'14-2' )
trendtable <- get_trend(`fpost-2014-3.csv`,'14-3' )
trendtable <- get_trend(`fpost-2014-4.csv`,'14-4' )
trendtable <- get_trend(`fpost-2014-5.csv`,'14-5' )
trendtable <- get_trend(`fpost-2014-6.csv`,'14-6' )
trendtable <- get_trend(`fpost-2014-7.csv`,'14-7' )
trendtable <- get_trend(`fpost-2014-8.csv`,'14-8' )
trendtable <- get_trend(`fpost-2014-9.csv`,'14-9' )
trendtable <- get_trend(`fpost-2014-10.csv`,'14-10' )
trendtable <- get_trend(`fpost-2014-11.csv`,'14-11' )
trendtable <- get_trend(`fpost-2014-12.csv`,'14-12' )
#==============2015===============#
trendtable <- get_trend(`fpost-2015-1.csv`,'15-1' )
trendtable <- get_trend(`fpost-2015-2.csv`,'15-2' )
trendtable <- get_trend(`fpost-2015-3.csv`,'15-3' )
trendtable <- get_trend(`fpost-2015-4.csv`,'15-4' )
trendtable <- get_trend(`fpost-2015-5.csv`,'15-5' )
trendtable <- get_trend(`fpost-2015-6.csv`,'15-6' )
trendtable <- get_trend(`fpost-2015-7.csv`,'15-7' )
trendtable <- get_trend(`fpost-2015-8.csv`,'15-8' )
trendtable <- get_trend(`fpost-2015-9.csv`,'15-9' )
trendtable <- get_trend(`fpost-2015-10.csv`,'15-10' )
trendtable <- get_trend(`fpost-2015-11.csv`,'15-11' )
trendtable <- get_trend(`fpost-2015-12.csv`,'15-12' )


row.names(trendtable) <- NULL
trendtable$groupp = 'pumpkin_pie'
trendtable$groupc = 'cauliflower_rice'
pumpkin_trend = trendtable[,c("year_month","pumpkin_pie","groupp")]
cauliflower_trend = trendtable[,c("year_month","cauliflower_rice","groupc")]



pplot = ggplot(pumpkin_trend, aes(x=year_month, y=pumpkin_pie, color = groupp,
group = groupp)) + geom_line()
pplot + theme(axis.text.x = element_text(angle = 90, hjust = 1))
cplot = ggplot(cauliflower_trend, aes(x=year_month, y=cauliflower_rice,color =
```

```
groupc, group = groupc)) + geom_line()
cplot + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

```
#==============================END=====================================#
```