PCA and Clustering on lettuce farm and retail prices

```
# Loadpackages
library(fpp3)
library(tsibble)
library(plyr)
library(dplyr)
library(tidyverse)
library(ggplot2)
```

```
# read data with phoenix -> chicago reefer lane, trucking volume, and yuma weather
data_raw <- readr::read_csv(file = "data/data_phoenix_with_yuma_weather_and_volume_and_lags.csv")  %>%
  mutate(yw = yearweek(yw)) %>%
  select(-X1) %>%
  as_tsibble(key = c(Mode,ORegionDAT, DRegionDAT), index = yw) %>%
  relocate(yw, Mode, ORegionDAT, DRegionDAT, approx_cost, approx_vol, tmax, prcp)
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   yw = col_character(),
##   Mode = col_character(),
##   ORegionDAT = col_character(),
##   DRegionDAT = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```
data_raw
```

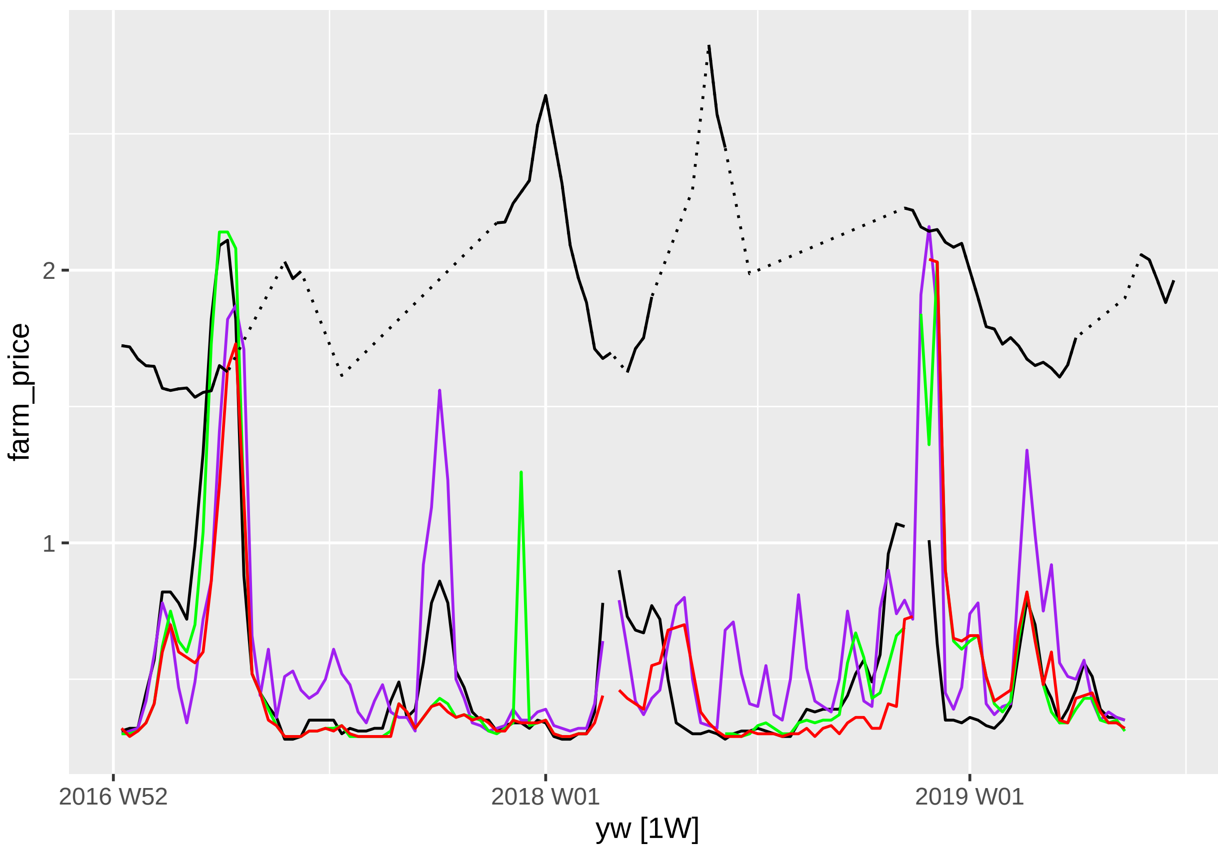
```
## # A tsibble: 237 x 21 [1W]
## # Key:       Mode, ORegionDAT, DRegionDAT [1]
##          yw Mode  ORegionDAT DRegionDAT approx_cost approx_vol  tmax    prcp
##      <week> <chr> <chr>      <chr>            <dbl>      <dbl> <dbl>   <dbl>
##  1 2017 W01 R     AZ_PHO     IL_CHI            1.72       16.4  64.1 0
##  2 2017 W02 R     AZ_PHO     IL_CHI            1.72       13.1  68.1 0
##  3 2017 W03 R     AZ_PHO     IL_CHI            1.67       11.3  66   0.0257
##  4 2017 W04 R     AZ_PHO     IL_CHI            1.65       16.1  63.1 0
##  5 2017 W05 R     AZ_PHO     IL_CHI            1.65       13.7  74.9 0
##  6 2017 W06 R     AZ_PHO     IL_CHI            1.57       11.1  77.4 0.00286
##  7 2017 W07 R     AZ_PHO     IL_CHI            1.56       11.1  74.1 0.124
##  8 2017 W08 R     AZ_PHO     IL_CHI            1.56       14.9  69.7 0
##  9 2017 W09 R     AZ_PHO     IL_CHI            1.57       14.3  71.9 0.0471
## 10 2017 W10 R     AZ_PHO     IL_CHI            1.53       16.9  83.1 0
## # ... with 227 more rows, and 13 more variables: sanitized_cost <dbl>,
## #   sanitized_vol <dbl>, tmax_lag_12 <dbl>, tmax_lag_8 <dbl>, tmax_lag_4 <dbl>,
## #   tmax_lag_2 <dbl>, prcp_lag_12 <dbl>, prcp_lag_8 <dbl>, prcp_lag_4 <dbl>,
## #   prcp_lag_2 <dbl>, cluster_1 <dbl>, cluster_2 <dbl>, cluster_3 <dbl>
```

```r
# Import lettuce data
lettuce <- readr::read_csv(file = 'data/lettuce_wholesale/all_lettuce.csv') %>%
  dplyr::select(-X1) %>%
  mutate(yw = yearweek(yw)) %>%
  relocate(commodity,yw)%>%
  as_tsibble(key = commodity,
             index = yw)

romaine <- lettuce %>% filter(commodity=="RO")
iceberg <- lettuce %>% filter(commodity=="IC")
green_leaf <- lettuce %>% filter(commodity=="GL")
red_leaf <- lettuce %>% filter(commodity=="RL")

print(head(romaine))
```

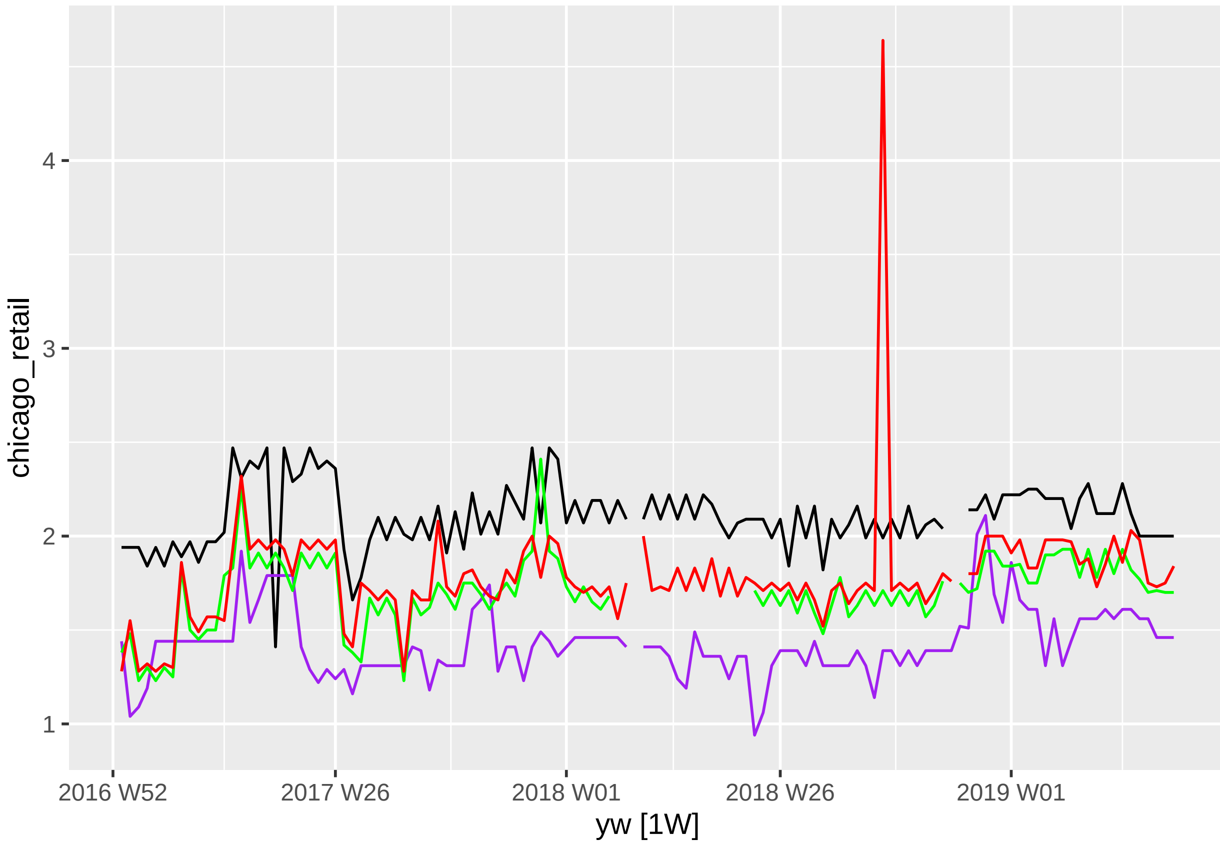
```
## # A tsibble: 6 x 4 [1W]
## # Key:       commodity [1]
##   commodity       yw farm_price chicago_retail
##   <chr>       <week>      <dbl>          <dbl>
## 1 RO        2017 W01       0.31           1.94
## 2 RO        2017 W02       0.32           1.94
## 3 RO        2017 W03       0.32           1.94
## 4 RO        2017 W04       0.45           1.84
## 5 RO        2017 W05       0.57           1.94
## 6 RO        2017 W06       0.82           1.84
```

Visualize what the 4 lettuce price series look like.

```r
#visualize farm prices
autoplot(romaine, farm_price, color = "black") +
  autolayer(iceberg, farm_price, color="purple") +
  autolayer(green_leaf, farm_price, color="green") +
  autolayer(red_leaf, farm_price, color="red") +
  autolayer(data_raw %>% filter_index(~"2019 W26"), approx_cost, color="black", linetype="dotted") +
  autolayer(data_raw %>% filter_index(~"2019 W26"), sanitized_cost, color ="black")
```




```
#visualize chciago retail prices
autoplot(romaine, chicago_retail, color = "black") +
  autolayer(iceberg, chicago_retail, color="purple") +
  autolayer(green_leaf, chicago_retail, color="green") +
  autolayer(red_leaf, chicago_retail, color="red")
```



Get data into right format for PCA and clustering

```
romaine_clust <- as_tibble(romaine) %>%
  mutate(ro_farm_price = farm_price, ro_chicago_retail = chicago_retail) %>%
  select(yw, ro_farm_price, ro_chicago_retail)

iceberg_clust <- as_tibble(iceberg) %>%
  mutate(ic_farm_price = farm_price, ic_chicago_retail = chicago_retail) %>%
  select(yw, ic_farm_price, ic_chicago_retail)

green_leaf_clust <- as_tibble(green_leaf) %>%
  mutate(gl_farm_price = farm_price, gl_chicago_retail = chicago_retail) %>%
  select(yw, gl_farm_price, gl_chicago_retail)

red_leaf_clust <- as_tibble(red_leaf) %>%
  mutate(rl_farm_price = farm_price, rl_chicago_retail = chicago_retail) %>%
  select(yw, rl_farm_price, rl_chicago_retail)

data2 <- join(romaine_clust, iceberg_clust, by = "yw", type = "full")
data3 <- join(data2, green_leaf_clust, by = "yw", type = "full")
data <- join(data3, red_leaf_clust, by = "yw", type = "full")


row.names(data) <- as.character(data$yw)

data <- data %>%
  select(-yw)
```

```
data
```

```
##          ro_farm_price ro_chicago_retail ic_farm_price ic_chicago_retail
## 2017 W01          0.31              1.94          0.30              1.44
## 2017 W02          0.32              1.94          0.31              1.04
## 2017 W03          0.32              1.94          0.32              1.09
## 2017 W04          0.45              1.84          0.42              1.19
## 2017 W05          0.57              1.94          0.59              1.44
## 2017 W06          0.82              1.84          0.78              1.44
## 2017 W07          0.82              1.97          0.69              1.44
## 2017 W08          0.78              1.89          0.47              1.44
## 2017 W09          0.72              1.97          0.34              1.44
## 2017 W10          0.99              1.86          0.49              1.44
## 2017 W11          1.33              1.97          0.72              1.44
## 2017 W12          1.82              1.97          0.86              1.44
## 2017 W13          2.09              2.02          1.41              1.44
## 2017 W14          2.11              2.47          1.82              1.44
## 2017 W15          1.82              2.31          1.87              1.92
## 2017 W16          0.88              2.40          1.71              1.54
## 2017 W17          0.52              2.36          0.66              1.66
## 2017 W18          0.45              2.47          0.44              1.79
## 2017 W19          0.40              1.41          0.61              1.79
## 2017 W20          0.36              2.47          0.36              1.79
## 2017 W21          0.28              2.29          0.51              1.79
## 2017 W22          0.28              2.33          0.53              1.41
## 2017 W23          0.29              2.47          0.46              1.29
## 2017 W24          0.35              2.36          0.43              1.22
## 2017 W25          0.35              2.40          0.45              1.29
## 2017 W26          0.35              2.36          0.50              1.24
## 2017 W27          0.35              1.93          0.61              1.29
## 2017 W28          0.30              1.66          0.52              1.16
## 2017 W29          0.32              1.78          0.48              1.31
## 2017 W30          0.31              1.98          0.38              1.31
## 2017 W31          0.31              2.10          0.34              1.31
## 2017 W32          0.32              1.98          0.42              1.31
## 2017 W33          0.32              2.10          0.48              1.31
## 2017 W34          0.42              2.01          0.38              1.31
## 2017 W35          0.49              1.98          0.36              1.41
## 2017 W36          0.36              2.10          0.36              1.39
## 2017 W37          0.39              1.98          0.31              1.18
## 2017 W38          0.56              2.16          0.92              1.34
## 2017 W39          0.78              1.91          1.13              1.31
## 2017 W40          0.86              2.13          1.56              1.31
## 2017 W41          0.78              1.93          1.23              1.31
## 2017 W42          0.53              2.23          0.50              1.61
## 2017 W43          0.47              2.01          0.43              1.66
## 2017 W44          0.38              2.13          0.34              1.74
## 2017 W45          0.35              2.01          0.33              1.28
## 2017 W46          0.35              2.27          0.31              1.41
## 2017 W47          0.31              2.18          0.32              1.41
## 2017 W48          0.33              2.09          0.33              1.23
## 2017 W49          0.34              2.47          0.39              1.41
## 2017 W50          0.34              2.07          0.35              1.49
```

```
## 2017 W51          0.32             2.47         0.35             1.44
## 2017 W52          0.35             2.41         0.38             1.36
## 2018 W01          0.34             2.07         0.39             1.41
## 2018 W02          0.29             2.19         0.33             1.46
## 2018 W03          0.28             2.07         0.32             1.46
## 2018 W04          0.28             2.19         0.31             1.46
## 2018 W05          0.30             2.19         0.32             1.46
## 2018 W06          0.30             2.07         0.32             1.46
## 2018 W07          0.38             2.19         0.41             1.46
## 2018 W08          0.78             2.09         0.64             1.41
## 2018 W09            NA               NA           NA               NA
## 2018 W10          0.90             2.09         0.79             1.41
## 2018 W11          0.73             2.22         0.61             1.41
## 2018 W12          0.68             2.09         0.42             1.41
## 2018 W13          0.67             2.22         0.37             1.36
## 2018 W14          0.77             2.09         0.43             1.24
## 2018 W15          0.72             2.22         0.46             1.19
## 2018 W16          0.50             2.09         0.63             1.49
## 2018 W17          0.34             2.22         0.77             1.36
## 2018 W18          0.32             2.17         0.80             1.36
## 2018 W19          0.30             2.07         0.50             1.36
## 2018 W20          0.30             1.99         0.34             1.24
## 2018 W21          0.31             2.07         0.33             1.36
## 2018 W22          0.30             2.09         0.32             1.36
## 2018 W23          0.28             2.09         0.68             0.94
## 2018 W24          0.30             2.09         0.71             1.06
## 2018 W25          0.31             1.99         0.52             1.31
## 2018 W26          0.31             2.09         0.41             1.39
## 2018 W27          0.32             1.84         0.40             1.39
## 2018 W28          0.31             2.16         0.55             1.39
## 2018 W29          0.30             1.99         0.37             1.31
## 2018 W30          0.29             2.16         0.35             1.44
## 2018 W31          0.29             1.82         0.50             1.31
## 2018 W32          0.34             2.09         0.81             1.31
## 2018 W33          0.39             1.99         0.54             1.31
## 2018 W34          0.38             2.06         0.42             1.31
## 2018 W35          0.39             2.16         0.40             1.39
## 2018 W36          0.39             1.99         0.38             1.31
## 2018 W37          0.39             2.09         0.50             1.14
## 2018 W38          0.44             1.99         0.75             1.39
## 2018 W39          0.52             2.09         0.58             1.39
## 2018 W40          0.57             1.99         0.42             1.31
## 2018 W41          0.49             2.16         0.40             1.39
## 2018 W42          0.59             1.99         0.76             1.31
## 2018 W43          0.96             2.06         0.90             1.39
## 2018 W44          1.07             2.09         0.74             1.39
## 2018 W45          1.06             2.04         0.79             1.39
## 2018 W46            NA               NA         0.72             1.39
## 2018 W47            NA               NA         1.91             1.52
## 2018 W48          1.01             2.14         2.16             1.51
## 2018 W49          0.63             2.14         1.84             2.01
## 2018 W50          0.35             2.22         0.45             2.11
## 2018 W51          0.35             2.09         0.39             1.69
## 2018 W52          0.34             2.22         0.47             1.54
```

```
## 2019 W01          0.36               2.22          0.74               1.86
## 2019 W02          0.35               2.22          0.78               1.66
## 2019 W03          0.33               2.25          0.41               1.61
## 2019 W04          0.32               2.25          0.37               1.61
## 2019 W05          0.35               2.20          0.40               1.31
## 2019 W06          0.40               2.20          0.41               1.56
## 2019 W07          0.60               2.20          0.88               1.31
## 2019 W08          0.79               2.04          1.34               1.44
## 2019 W09          0.70               2.20          1.03               1.56
## 2019 W10          0.49               2.28          0.75               1.56
## 2019 W11          0.43               2.12          0.92               1.56
## 2019 W12          0.34               2.12          0.56               1.61
## 2019 W13          0.39               2.12          0.51               1.56
## 2019 W14          0.46               2.28          0.50               1.61
## 2019 W15          0.56               2.12          0.57               1.61
## 2019 W16          0.51               2.00          0.42               1.56
## 2019 W17          0.39               2.00          0.35               1.56
## 2019 W18          0.36               2.00          0.38               1.46
## 2019 W19          0.36               2.00          0.36               1.46
## 2019 W20          0.35               2.00          0.35               1.46
##          gl_farm_price gl_chicago_retail rl_farm_price rl_chicago_retail
## 2017 W01          0.30               1.38          0.32               1.28
## 2017 W02          0.30               1.48          0.29               1.55
## 2017 W03          0.31               1.23          0.31               1.28
## 2017 W04          0.34               1.30          0.34               1.32
## 2017 W05          0.41               1.23          0.41               1.28
## 2017 W06          0.62               1.30          0.60               1.32
## 2017 W07          0.75               1.25          0.70               1.30
## 2017 W08          0.64               1.83          0.60               1.86
## 2017 W09          0.60               1.50          0.58               1.57
## 2017 W10          0.70               1.45          0.56               1.49
## 2017 W11          1.04               1.50          0.60               1.57
## 2017 W12          1.73               1.50          0.86               1.57
## 2017 W13          2.14               1.79          1.21               1.55
## 2017 W14          2.14               1.83          1.64               1.93
## 2017 W15          2.08               2.26          1.73               2.32
## 2017 W16          1.14               1.83          1.16               1.93
## 2017 W17          0.52               1.91          0.52               1.98
## 2017 W18          0.45               1.83          0.45               1.93
## 2017 W19          0.39               1.91          0.35               1.98
## 2017 W20          0.33               1.83          0.33               1.93
## 2017 W21          0.29               1.71          0.29               1.79
## 2017 W22          0.29               1.91          0.29               1.98
## 2017 W23          0.29               1.83          0.29               1.93
## 2017 W24          0.31               1.91          0.31               1.98
## 2017 W25          0.31               1.83          0.31               1.93
## 2017 W26          0.32               1.91          0.32               1.98
## 2017 W27          0.32               1.42          0.31               1.48
## 2017 W28          0.33               1.38          0.33               1.41
## 2017 W29          0.29               1.33          0.30               1.75
## 2017 W30          0.29               1.67          0.29               1.71
## 2017 W31          0.29               1.58          0.29               1.66
## 2017 W32          0.29               1.67          0.29               1.71
## 2017 W33          0.29               1.58          0.29               1.66
```

```
## 2017 W34          0.31              1.23          0.29              1.28
## 2017 W35          0.41              1.67          0.41              1.71
## 2017 W36          0.38              1.58          0.38              1.66
## 2017 W37          0.32              1.62          0.32              1.66
## 2017 W38          0.36              1.75          0.36              2.08
## 2017 W39          0.40              1.69          0.40              1.73
## 2017 W40          0.43              1.61          0.41              1.68
## 2017 W41          0.41              1.75          0.38              1.80
## 2017 W42          0.36              1.75          0.36              1.82
## 2017 W43          0.37              1.69          0.37              1.73
## 2017 W44          0.36              1.61          0.35              1.68
## 2017 W45          0.35              1.69          0.36              1.66
## 2017 W46          0.31              1.75          0.34              1.82
## 2017 W47          0.30              1.68          0.31              1.75
## 2017 W48          0.32              1.87          0.31              1.92
## 2017 W49          0.34              1.92          0.35              2.00
## 2017 W50          1.26              2.41          0.34              1.78
## 2017 W51          0.33              1.92          0.34              2.00
## 2017 W52          0.34              1.88          0.34              1.96
## 2018 W01          0.35              1.73          0.35              1.78
## 2018 W02          0.30              1.65          0.30              1.73
## 2018 W03          0.29              1.73          0.29              1.70
## 2018 W04          0.29              1.65          0.29              1.73
## 2018 W05          0.30              1.61          0.30              1.68
## 2018 W06          0.30              1.68          0.30              1.73
## 2018 W07            NA                NA          0.34              1.56
## 2018 W08            NA                NA          0.44              1.75
## 2018 W09            NA                NA            NA                NA
## 2018 W10            NA                NA          0.46              2.00
## 2018 W11            NA                NA          0.43              1.71
## 2018 W12            NA                NA          0.41              1.73
## 2018 W13            NA                NA          0.39              1.71
## 2018 W14            NA                NA          0.55              1.83
## 2018 W15            NA                NA          0.56              1.71
## 2018 W16            NA                NA          0.68              1.83
## 2018 W17            NA                NA          0.69              1.71
## 2018 W18            NA                NA          0.70              1.88
## 2018 W19            NA                NA          0.54              1.68
## 2018 W20            NA                NA          0.38              1.83
## 2018 W21            NA                NA          0.34              1.68
## 2018 W22            NA                NA          0.31              1.78
## 2018 W23          0.30              1.71          0.29              1.75
## 2018 W24          0.30              1.63          0.29              1.71
## 2018 W25          0.29              1.71          0.29              1.75
## 2018 W26          0.30              1.63          0.31              1.71
## 2018 W27          0.33              1.71          0.30              1.75
## 2018 W28          0.34              1.59          0.30              1.66
## 2018 W29          0.32              1.71          0.30              1.75
## 2018 W30          0.30              1.59          0.29              1.66
## 2018 W31          0.30              1.48          0.30              1.52
## 2018 W32          0.34              1.63          0.30              1.71
## 2018 W33          0.35              1.78          0.32              1.75
## 2018 W34          0.34              1.57          0.29              1.64
## 2018 W35          0.35              1.63          0.32              1.71
```

```
## 2018 W36          0.35              1.71          0.33              1.75
## 2018 W37          0.37              1.63          0.30              1.71
## 2018 W38          0.56              1.71          0.34              4.64
## 2018 W39          0.67              1.63          0.36              1.71
## 2018 W40          0.58              1.71          0.36              1.75
## 2018 W41          0.43              1.63          0.32              1.71
## 2018 W42          0.45              1.71          0.32              1.75
## 2018 W43          0.55              1.57          0.41              1.64
## 2018 W44          0.66              1.63          0.40              1.71
## 2018 W45          0.69              1.76          0.72              1.80
## 2018 W46            NA                NA          0.73              1.76
## 2018 W47          1.84              1.75            NA                NA
## 2018 W48          1.36              1.70          2.04              1.80
## 2018 W49          2.03              1.72          2.03              1.80
## 2018 W50          0.90              1.92          0.90              2.00
## 2018 W51          0.64              1.92          0.65              2.00
## 2018 W52          0.61              1.84          0.64              2.00
## 2019 W01          0.64              1.84          0.66              1.91
## 2019 W02          0.66              1.85          0.66              1.98
## 2019 W03          0.51              1.75          0.51              1.83
## 2019 W04          0.41              1.75          0.42              1.83
## 2019 W05          0.38              1.90          0.44              1.98
## 2019 W06          0.42              1.90          0.46              1.98
## 2019 W07          0.65              1.93          0.68              1.98
## 2019 W08          0.82              1.93          0.82              1.97
## 2019 W09          0.65              1.78          0.64              1.85
## 2019 W10          0.48              1.93          0.48              1.88
## 2019 W11          0.38              1.78          0.60              1.73
## 2019 W12          0.34              1.93          0.35              1.85
## 2019 W13          0.34              1.80          0.34              2.00
## 2019 W14          0.39              1.93          0.43              1.86
## 2019 W15          0.43              1.82          0.44              2.03
## 2019 W16          0.43              1.77          0.45              1.98
## 2019 W17          0.35              1.70          0.38              1.75
## 2019 W18          0.34              1.71          0.34              1.73
## 2019 W19          0.35              1.70          0.34              1.75
## 2019 W20          0.31              1.70          0.32              1.84
```

Green leaf farm and retail prices actually have a lot of missing values in 2018. Rather than trying to interpolate, I'm goign to just drop those columns. Green leaf and red leaf prices look closely correlated, so the PCA would essentially take one out anyways.

There is at the end of 2017 a big spike in green leaf farm, not reflected in red leaf farm. Also in the middle of 2018 a big spike in red leaf retail not reflected in green leaf retail. Not sure what that's about. But again, I'm jsut dorpping green leaf data for the rest of this analysis

```
data <- data %>%
  select(-gl_farm_price, -gl_chicago_retail)
```

Interactive visualizations to play around with to see how the prices of different lettuce types relate to each other.

```r
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following objects are masked from 'package:plyr':
##
##     arrange, mutate, rename, summarise

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```r
data_plot <- data%>%mutate(yw = row.names(data))

data_plot
```

```
##          ro_farm_price ro_chicago_retail ic_farm_price ic_chicago_retail
## 2017 W01          0.31              1.94          0.30              1.44
## 2017 W02          0.32              1.94          0.31              1.04
## 2017 W03          0.32              1.94          0.32              1.09
## 2017 W04          0.45              1.84          0.42              1.19
## 2017 W05          0.57              1.94          0.59              1.44
## 2017 W06          0.82              1.84          0.78              1.44
## 2017 W07          0.82              1.97          0.69              1.44
## 2017 W08          0.78              1.89          0.47              1.44
## 2017 W09          0.72              1.97          0.34              1.44
## 2017 W10          0.99              1.86          0.49              1.44
## 2017 W11          1.33              1.97          0.72              1.44
## 2017 W12          1.82              1.97          0.86              1.44
## 2017 W13          2.09              2.02          1.41              1.44
## 2017 W14          2.11              2.47          1.82              1.44
## 2017 W15          1.82              2.31          1.87              1.92
## 2017 W16          0.88              2.40          1.71              1.54
## 2017 W17          0.52              2.36          0.66              1.66
## 2017 W18          0.45              2.47          0.44              1.79
## 2017 W19          0.40              1.41          0.61              1.79
## 2017 W20          0.36              2.47          0.36              1.79
## 2017 W21          0.28              2.29          0.51              1.79
## 2017 W22          0.28              2.33          0.53              1.41
## 2017 W23          0.29              2.47          0.46              1.29
## 2017 W24          0.35              2.36          0.43              1.22
## 2017 W25          0.35              2.40          0.45              1.29
```

```
## 2017 W26           0.35                 2.36           0.50                 1.24
## 2017 W27           0.35                 1.93           0.61                 1.29
## 2017 W28           0.30                 1.66           0.52                 1.16
## 2017 W29           0.32                 1.78           0.48                 1.31
## 2017 W30           0.31                 1.98           0.38                 1.31
## 2017 W31           0.31                 2.10           0.34                 1.31
## 2017 W32           0.32                 1.98           0.42                 1.31
## 2017 W33           0.32                 2.10           0.48                 1.31
## 2017 W34           0.42                 2.01           0.38                 1.31
## 2017 W35           0.49                 1.98           0.36                 1.41
## 2017 W36           0.36                 2.10           0.36                 1.39
## 2017 W37           0.39                 1.98           0.31                 1.18
## 2017 W38           0.56                 2.16           0.92                 1.34
## 2017 W39           0.78                 1.91           1.13                 1.31
## 2017 W40           0.86                 2.13           1.56                 1.31
## 2017 W41           0.78                 1.93           1.23                 1.31
## 2017 W42           0.53                 2.23           0.50                 1.61
## 2017 W43           0.47                 2.01           0.43                 1.66
## 2017 W44           0.38                 2.13           0.34                 1.74
## 2017 W45           0.35                 2.01           0.33                 1.28
## 2017 W46           0.35                 2.27           0.31                 1.41
## 2017 W47           0.31                 2.18           0.32                 1.41
## 2017 W48           0.33                 2.09           0.33                 1.23
## 2017 W49           0.34                 2.47           0.39                 1.41
## 2017 W50           0.34                 2.07           0.35                 1.49
## 2017 W51           0.32                 2.47           0.35                 1.44
## 2017 W52           0.35                 2.41           0.38                 1.36
## 2018 W01           0.34                 2.07           0.39                 1.41
## 2018 W02           0.29                 2.19           0.33                 1.46
## 2018 W03           0.28                 2.07           0.32                 1.46
## 2018 W04           0.28                 2.19           0.31                 1.46
## 2018 W05           0.30                 2.19           0.32                 1.46
## 2018 W06           0.30                 2.07           0.32                 1.46
## 2018 W07           0.38                 2.19           0.41                 1.46
## 2018 W08           0.78                 2.09           0.64                 1.41
## 2018 W09             NA                   NA             NA                   NA
## 2018 W10           0.90                 2.09           0.79                 1.41
## 2018 W11           0.73                 2.22           0.61                 1.41
## 2018 W12           0.68                 2.09           0.42                 1.41
## 2018 W13           0.67                 2.22           0.37                 1.36
## 2018 W14           0.77                 2.09           0.43                 1.24
## 2018 W15           0.72                 2.22           0.46                 1.19
## 2018 W16           0.50                 2.09           0.63                 1.49
## 2018 W17           0.34                 2.22           0.77                 1.36
## 2018 W18           0.32                 2.17           0.80                 1.36
## 2018 W19           0.30                 2.07           0.50                 1.36
## 2018 W20           0.30                 1.99           0.34                 1.24
## 2018 W21           0.31                 2.07           0.33                 1.36
## 2018 W22           0.30                 2.09           0.32                 1.36
## 2018 W23           0.28                 2.09           0.68                 0.94
## 2018 W24           0.30                 2.09           0.71                 1.06
## 2018 W25           0.31                 1.99           0.52                 1.31
## 2018 W26           0.31                 2.09           0.41                 1.39
## 2018 W27           0.32                 1.84           0.40                 1.39
```

```
## 2018 W28          0.31              2.16          0.55              1.39
## 2018 W29          0.30              1.99          0.37              1.31
## 2018 W30          0.29              2.16          0.35              1.44
## 2018 W31          0.29              1.82          0.50              1.31
## 2018 W32          0.34              2.09          0.81              1.31
## 2018 W33          0.39              1.99          0.54              1.31
## 2018 W34          0.38              2.06          0.42              1.31
## 2018 W35          0.39              2.16          0.40              1.39
## 2018 W36          0.39              1.99          0.38              1.31
## 2018 W37          0.39              2.09          0.50              1.14
## 2018 W38          0.44              1.99          0.75              1.39
## 2018 W39          0.52              2.09          0.58              1.39
## 2018 W40          0.57              1.99          0.42              1.31
## 2018 W41          0.49              2.16          0.40              1.39
## 2018 W42          0.59              1.99          0.76              1.31
## 2018 W43          0.96              2.06          0.90              1.39
## 2018 W44          1.07              2.09          0.74              1.39
## 2018 W45          1.06              2.04          0.79              1.39
## 2018 W46            NA                NA          0.72              1.39
## 2018 W47            NA                NA          1.91              1.52
## 2018 W48          1.01              2.14          2.16              1.51
## 2018 W49          0.63              2.14          1.84              2.01
## 2018 W50          0.35              2.22          0.45              2.11
## 2018 W51          0.35              2.09          0.39              1.69
## 2018 W52          0.34              2.22          0.47              1.54
## 2019 W01          0.36              2.22          0.74              1.86
## 2019 W02          0.35              2.22          0.78              1.66
## 2019 W03          0.33              2.25          0.41              1.61
## 2019 W04          0.32              2.25          0.37              1.61
## 2019 W05          0.35              2.20          0.40              1.31
## 2019 W06          0.40              2.20          0.41              1.56
## 2019 W07          0.60              2.20          0.88              1.31
## 2019 W08          0.79              2.04          1.34              1.44
## 2019 W09          0.70              2.20          1.03              1.56
## 2019 W10          0.49              2.28          0.75              1.56
## 2019 W11          0.43              2.12          0.92              1.56
## 2019 W12          0.34              2.12          0.56              1.61
## 2019 W13          0.39              2.12          0.51              1.56
## 2019 W14          0.46              2.28          0.50              1.61
## 2019 W15          0.56              2.12          0.57              1.61
## 2019 W16          0.51              2.00          0.42              1.56
## 2019 W17          0.39              2.00          0.35              1.56
## 2019 W18          0.36              2.00          0.38              1.46
## 2019 W19          0.36              2.00          0.36              1.46
## 2019 W20          0.35              2.00          0.35              1.46
##          rl_farm_price rl_chicago_retail       yw
## 2017 W01          0.32              1.28 2017 W01
## 2017 W02          0.29              1.55 2017 W02
## 2017 W03          0.31              1.28 2017 W03
## 2017 W04          0.34              1.32 2017 W04
## 2017 W05          0.41              1.28 2017 W05
## 2017 W06          0.60              1.32 2017 W06
## 2017 W07          0.70              1.30 2017 W07
## 2017 W08          0.60              1.86 2017 W08
```

```
## 2017 W09          0.58              1.57 2017 W09
## 2017 W10          0.56              1.49 2017 W10
## 2017 W11          0.60              1.57 2017 W11
## 2017 W12          0.86              1.57 2017 W12
## 2017 W13          1.21              1.55 2017 W13
## 2017 W14          1.64              1.93 2017 W14
## 2017 W15          1.73              2.32 2017 W15
## 2017 W16          1.16              1.93 2017 W16
## 2017 W17          0.52              1.98 2017 W17
## 2017 W18          0.45              1.93 2017 W18
## 2017 W19          0.35              1.98 2017 W19
## 2017 W20          0.33              1.93 2017 W20
## 2017 W21          0.29              1.79 2017 W21
## 2017 W22          0.29              1.98 2017 W22
## 2017 W23          0.29              1.93 2017 W23
## 2017 W24          0.31              1.98 2017 W24
## 2017 W25          0.31              1.93 2017 W25
## 2017 W26          0.32              1.98 2017 W26
## 2017 W27          0.31              1.48 2017 W27
## 2017 W28          0.33              1.41 2017 W28
## 2017 W29          0.30              1.75 2017 W29
## 2017 W30          0.29              1.71 2017 W30
## 2017 W31          0.29              1.66 2017 W31
## 2017 W32          0.29              1.71 2017 W32
## 2017 W33          0.29              1.66 2017 W33
## 2017 W34          0.29              1.28 2017 W34
## 2017 W35          0.41              1.71 2017 W35
## 2017 W36          0.38              1.66 2017 W36
## 2017 W37          0.32              1.66 2017 W37
## 2017 W38          0.36              2.08 2017 W38
## 2017 W39          0.40              1.73 2017 W39
## 2017 W40          0.41              1.68 2017 W40
## 2017 W41          0.38              1.80 2017 W41
## 2017 W42          0.36              1.82 2017 W42
## 2017 W43          0.37              1.73 2017 W43
## 2017 W44          0.35              1.68 2017 W44
## 2017 W45          0.36              1.66 2017 W45
## 2017 W46          0.34              1.82 2017 W46
## 2017 W47          0.31              1.75 2017 W47
## 2017 W48          0.31              1.92 2017 W48
## 2017 W49          0.35              2.00 2017 W49
## 2017 W50          0.34              1.78 2017 W50
## 2017 W51          0.34              2.00 2017 W51
## 2017 W52          0.34              1.96 2017 W52
## 2018 W01          0.35              1.78 2018 W01
## 2018 W02          0.30              1.73 2018 W02
## 2018 W03          0.29              1.70 2018 W03
## 2018 W04          0.29              1.73 2018 W04
## 2018 W05          0.30              1.68 2018 W05
## 2018 W06          0.30              1.73 2018 W06
## 2018 W07          0.34              1.56 2018 W07
## 2018 W08          0.44              1.75 2018 W08
## 2018 W09            NA                NA 2018 W09
## 2018 W10          0.46              2.00 2018 W10
```

```
## 2018 W11          0.43            1.71 2018 W11
## 2018 W12          0.41            1.73 2018 W12
## 2018 W13          0.39            1.71 2018 W13
## 2018 W14          0.55            1.83 2018 W14
## 2018 W15          0.56            1.71 2018 W15
## 2018 W16          0.68            1.83 2018 W16
## 2018 W17          0.69            1.71 2018 W17
## 2018 W18          0.70            1.88 2018 W18
## 2018 W19          0.54            1.68 2018 W19
## 2018 W20          0.38            1.83 2018 W20
## 2018 W21          0.34            1.68 2018 W21
## 2018 W22          0.31            1.78 2018 W22
## 2018 W23          0.29            1.75 2018 W23
## 2018 W24          0.29            1.71 2018 W24
## 2018 W25          0.29            1.75 2018 W25
## 2018 W26          0.31            1.71 2018 W26
## 2018 W27          0.30            1.75 2018 W27
## 2018 W28          0.30            1.66 2018 W28
## 2018 W29          0.30            1.75 2018 W29
## 2018 W30          0.29            1.66 2018 W30
## 2018 W31          0.30            1.52 2018 W31
## 2018 W32          0.30            1.71 2018 W32
## 2018 W33          0.32            1.75 2018 W33
## 2018 W34          0.29            1.64 2018 W34
## 2018 W35          0.32            1.71 2018 W35
## 2018 W36          0.33            1.75 2018 W36
## 2018 W37          0.30            1.71 2018 W37
## 2018 W38          0.34            4.64 2018 W38
## 2018 W39          0.36            1.71 2018 W39
## 2018 W40          0.36            1.75 2018 W40
## 2018 W41          0.32            1.71 2018 W41
## 2018 W42          0.32            1.75 2018 W42
## 2018 W43          0.41            1.64 2018 W43
## 2018 W44          0.40            1.71 2018 W44
## 2018 W45          0.72            1.80 2018 W45
## 2018 W46          0.73            1.76 2018 W46
## 2018 W47            NA              NA 2018 W47
## 2018 W48          2.04            1.80 2018 W48
## 2018 W49          2.03            1.80 2018 W49
## 2018 W50          0.90            2.00 2018 W50
## 2018 W51          0.65            2.00 2018 W51
## 2018 W52          0.64            2.00 2018 W52
## 2019 W01          0.66            1.91 2019 W01
## 2019 W02          0.66            1.98 2019 W02
## 2019 W03          0.51            1.83 2019 W03
## 2019 W04          0.42            1.83 2019 W04
## 2019 W05          0.44            1.98 2019 W05
## 2019 W06          0.46            1.98 2019 W06
## 2019 W07          0.68            1.98 2019 W07
## 2019 W08          0.82            1.97 2019 W08
## 2019 W09          0.64            1.85 2019 W09
## 2019 W10          0.48            1.88 2019 W10
## 2019 W11          0.60            1.73 2019 W11
## 2019 W12          0.35            1.85 2019 W12
```

```
## 2019 W13          0.34             2.00 2019 W13
## 2019 W14          0.43             1.86 2019 W14
## 2019 W15          0.44             2.03 2019 W15
## 2019 W16          0.45             1.98 2019 W16
## 2019 W17          0.38             1.75 2019 W17
## 2019 W18          0.34             1.73 2019 W18
## 2019 W19          0.34             1.75 2019 W19
## 2019 W20          0.32             1.84 2019 W20
```

```
p <- ggplot(data_plot, aes(x=log(ro_farm_price), y=log(rl_farm_price),
                       color = format(as.Date(yearweek(yw)),format = "%Y"),
                       text=yw)) +
  geom_point(alpha=0.75)

#ggplotly(p, tooltip="yw")
```

library for PCA and clustering visualizations to come.

```
library(factoextra)
```

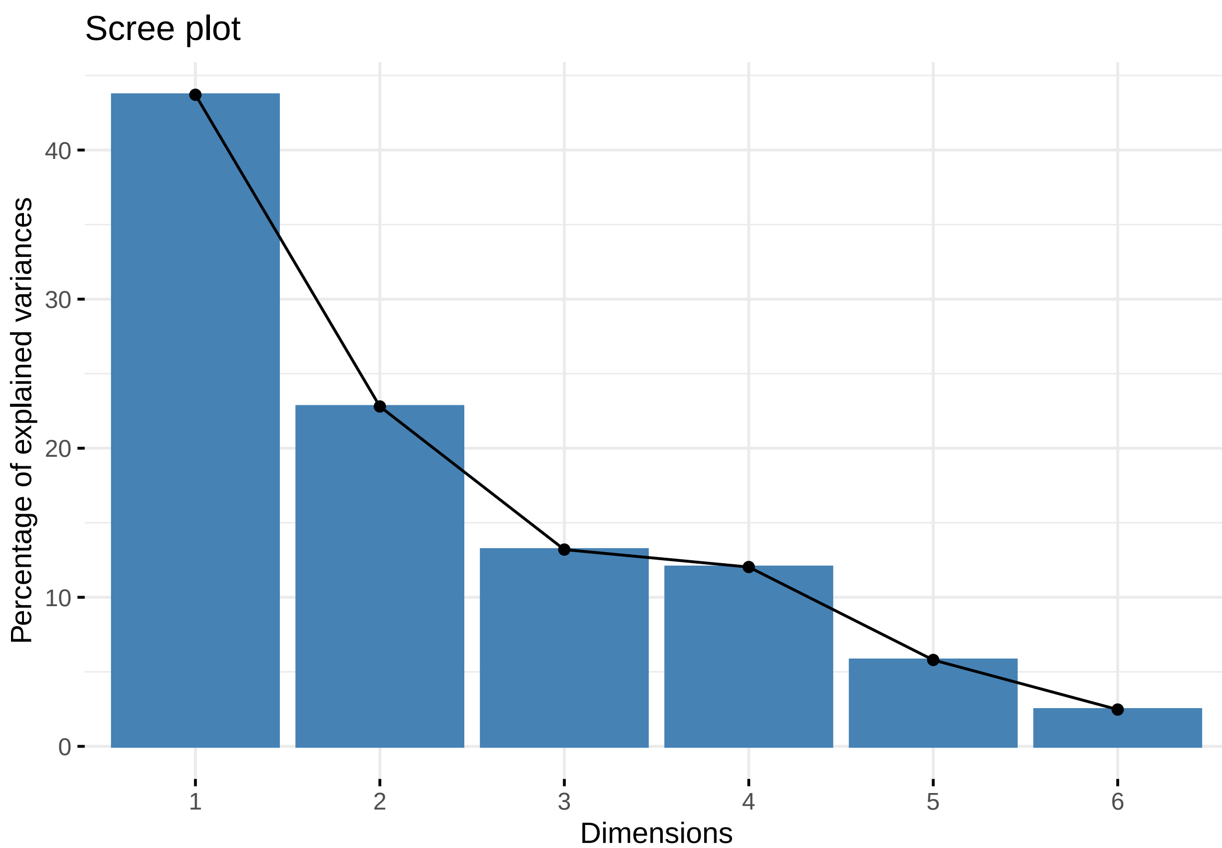
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

PCA

There are still a few sporadic na values in the data. There's not many, so I will drop those rows.

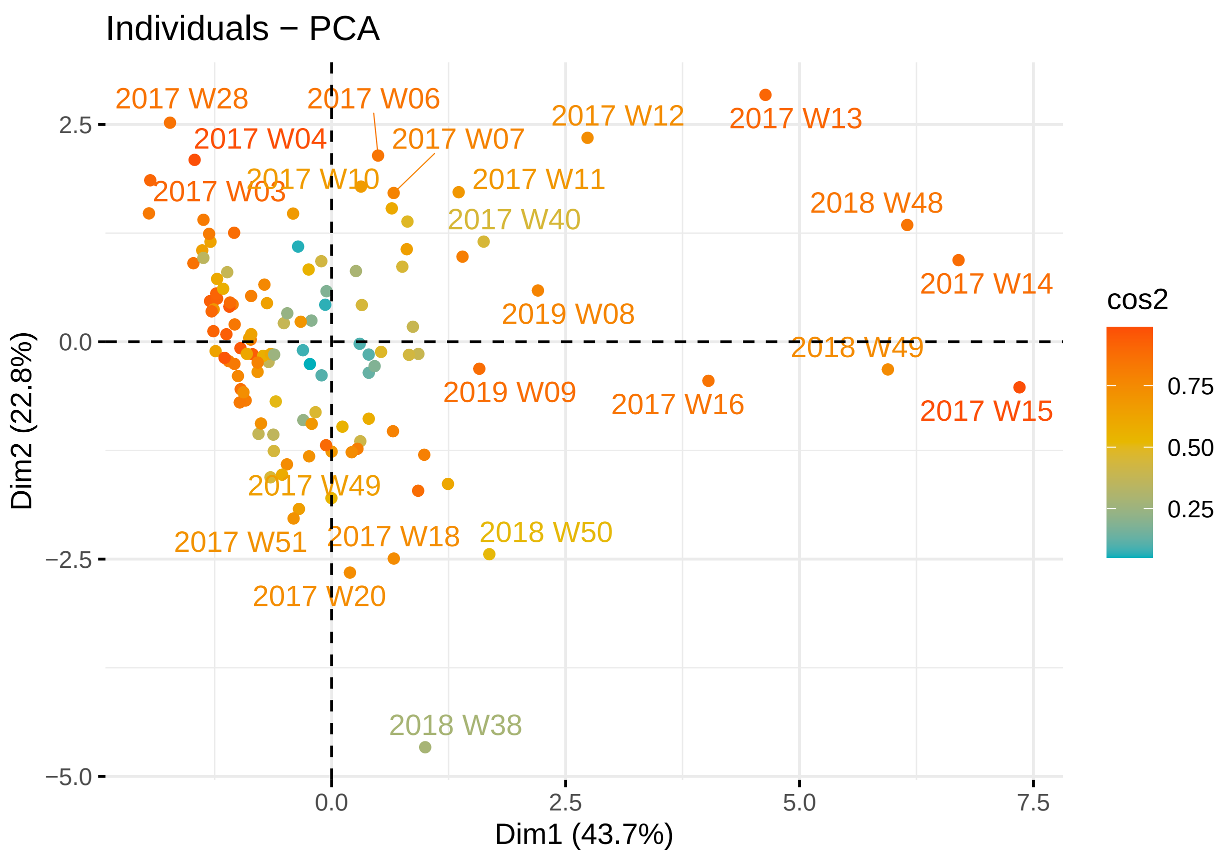
```
#drop rows containing any NA values
data_drop_na <- drop_na(data)
```

```
#compute PCA
pca <- prcomp(data_drop_na, scale = TRUE)

#visualize eigenvalues
fviz_eig(pca)
```

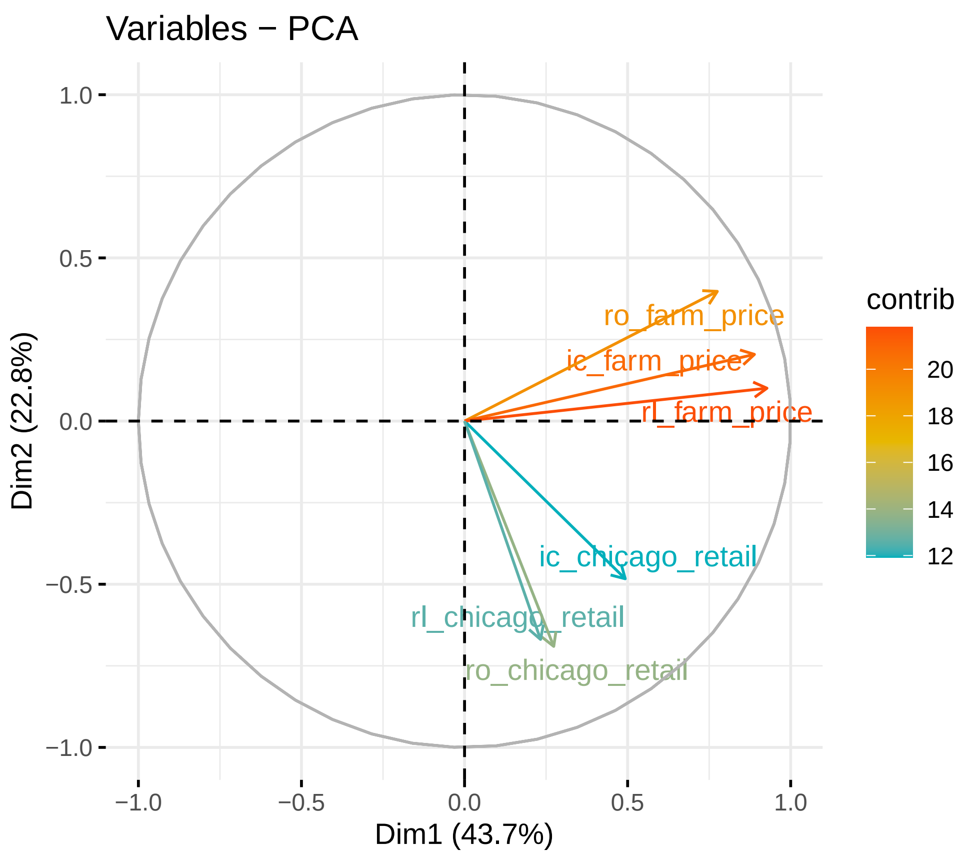



```r
#graph of observations
fviz_pca_ind(pca,
             col.ind = "cos2", # Color by the quality of representation
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE     # Avoid text overlapping
             )
```

```
## Warning: ggrepel: 98 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

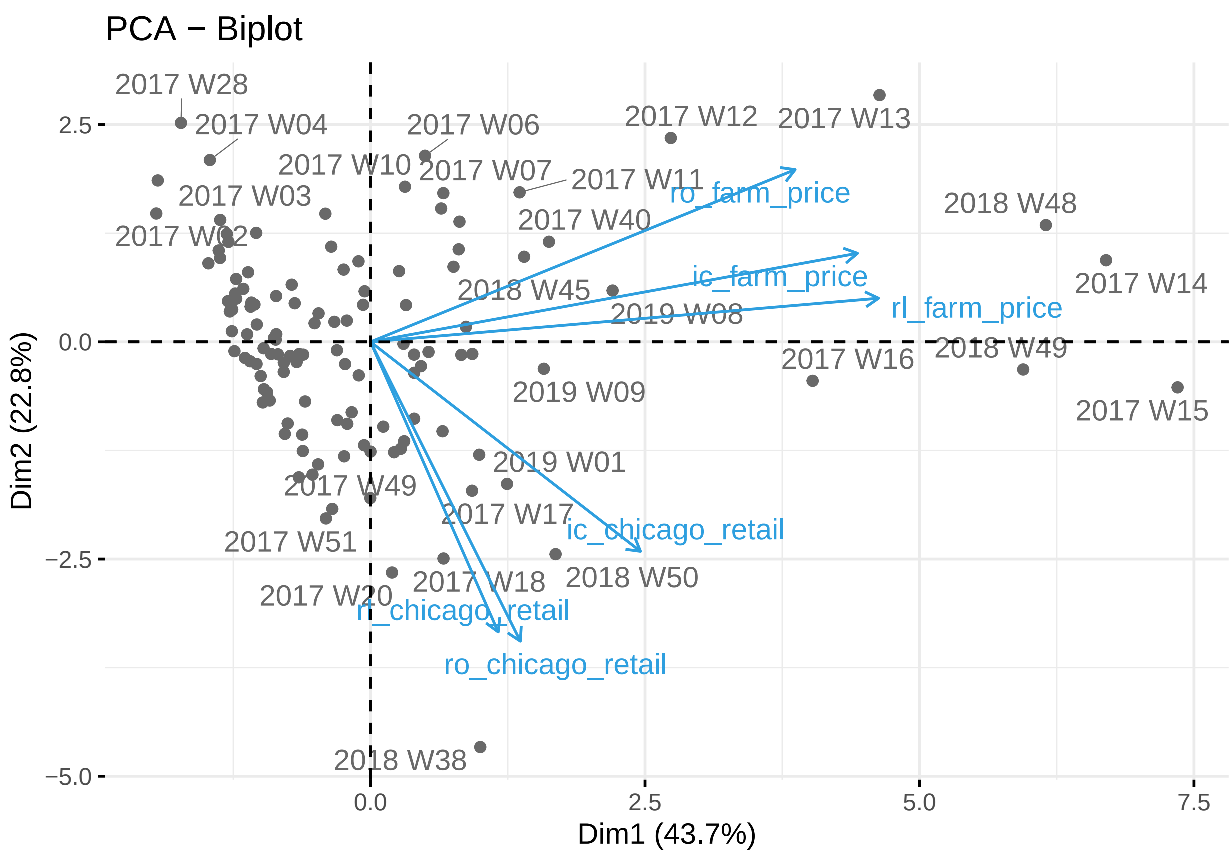



```
#graph of variables
fviz_pca_var(pca,
             col.var = "contrib", # Color by contributions to the PC
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE     # Avoid text overlapping
             )
```




```
#biplot of observations and variables
fviz_pca_biplot(pca, repel = TRUE,
                col.var = "#2E9FDF", # Variables color
                col.ind = "#696969"  # Individuals color
                )
```

```
## Warning: ggrepel: 94 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



Scree plot suggests we should keep the first 3 dimensions of the PCA - because there is an "elbow" in the scree plot there, and also because the first 3 components added up explain about 85% of the total variance.

I think maybe we should do 4 because it makes intuitive sense to me. The market is different for romaine/non-romaine lettuces during e. coli outbreaks, so we should need 4 dims to describe retail/farm prics for romainee/non-romaine lettuces.

```
# First 4 principal components
comps <- data.frame(pca$x[,1:4])
```

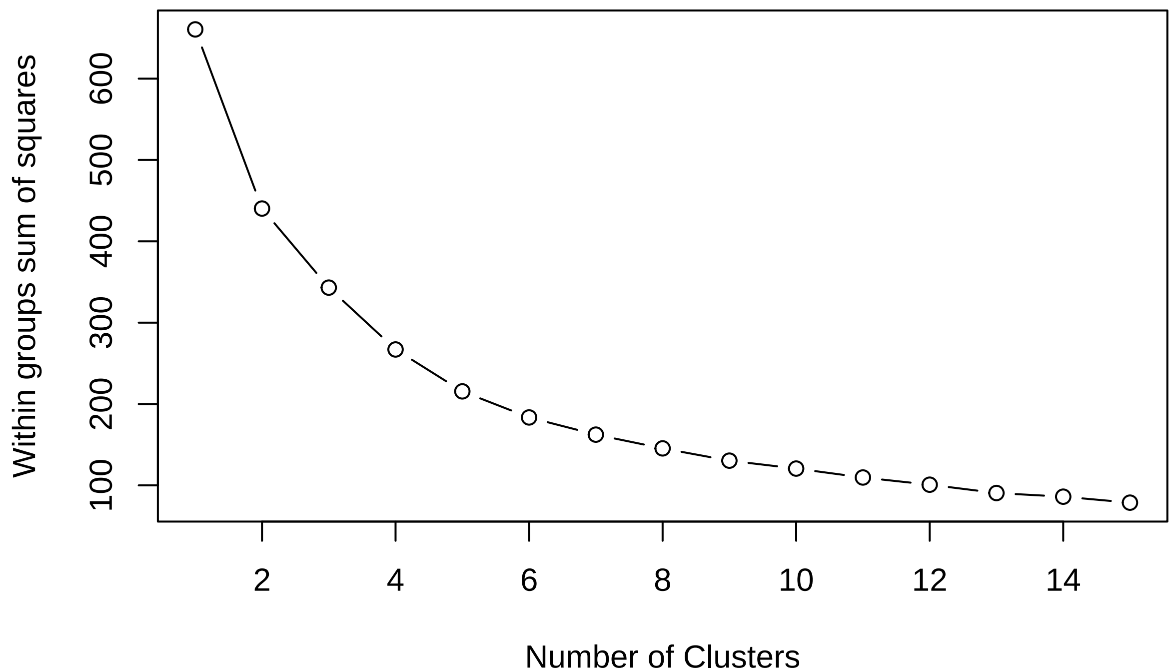
We can make a cool 3D interactive plot

```
#library(rgl)
# 3D plot (window pops out)

#plot3d(comps$PC1, comps$PC2, comps$PC3)
```

K-Means clustering

```
# Determine number of clusters
wss <- (nrow(comps)-1)*sum(apply(comps,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(comps,
                                     nstart = 25,
                                     iter.max=1000,
```

```
                                          centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
```



I don't see an obvious elbow bend in the plot, so it's not clear to me how many clusters to do. I'll try a few possibilities.

First, I'll do 2 clusters.

```
set.seed(123) # set seed for reproducibility

#do the clustering
compsk2 <- kmeans(comps, 2, nstart = 25, iter.max=1000)

# 3D plot
#plot3d(comps$PC1, comps$PC2, comps$PC3, col=compsk2$clust)
```

```
# Cluster sizes
sort(table(compsk2$clust))
```

```
##
##   2   1
##   7 114
```

```
clust <- names(sort(table(compsk2$clust)))
# First cluster
cluster1_of_2 = row.names(data[compsk2$clust==clust[1],])
# Second Cluster
cluster2_of_2 = row.names(data[compsk2$clust==clust[2],])

# cluster members
cluster1_of_2
```

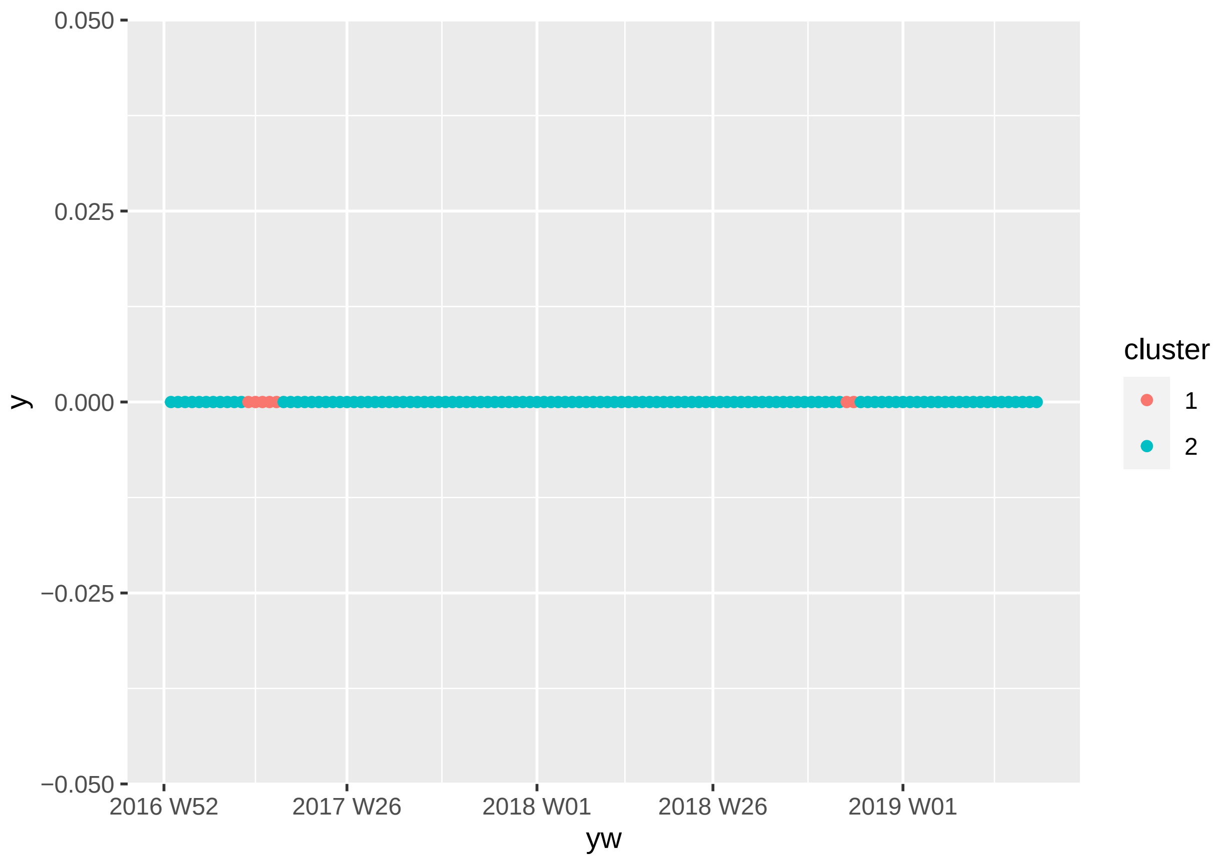
```
## [1] "2017 W12" "2017 W13" "2017 W14" "2017 W15" "2017 W16" "2018 W45" "2018 W46"
```

So 2017 W12-16 and 2018 W45-46 were picked out as one cluster. Everything else was the other cluster.

```
#visualize the clusters on a timeline

cluster1_of_2_df = data.frame(cluster1_of_2) %>%
  rename(yw = cluster1_of_2) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 1)

cluster2_of_2_df = data.frame(cluster2_of_2) %>%
  rename(yw = cluster2_of_2) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 2)

clusterk2_df = join(cluster1_of_2_df, cluster2_of_2_df, type="full") %>%
  mutate(cluster = as.character(cluster)) %>%
  as_tsibble(index = yw)
```

```
## Joining by: yw, cluster
```

```
ggplot(clusterk2_df, aes(x=yw, y=0, color = cluster)) + geom_point()
```




```
unique(clusterk2_df$cluster)
```

```
## [1] "2" "1"
```

Next, let's try 3 clusters.

```
set.seed(123) # set seed for reproducibility

#do the clustering
compsk3 <- kmeans(comps, 3, nstart = 25, iter.max=1000)

# 3D plot
#plot3d(comps$PC1, comps$PC2, comps$PC3, col=compsk3$clust)
```

```
# Cluster sizes
sort(table(compsk3$clust))
```

```
##
##  2  3  1
##  7 44 70
```

```
clust <- names(sort(table(compsk3$clust)))
# First cluster
cluster1_of_3 = row.names(data[compsk3$clust==clust[1],])
```

```r
# Second Cluster
cluster2_of_3 = row.names(data[compsk3$clust==clust[2],])
# Third Cluster
cluster3_of_3 = row.names(data[compsk3$clust==clust[3],])
```

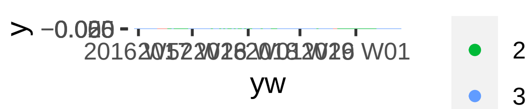
```r
#visualize the clusters on a timeline

cluster1_of_3_df = data.frame(cluster1_of_3) %>%
  rename(yw = cluster1_of_3) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 1)

cluster2_of_3_df = data.frame(cluster2_of_3) %>%
  rename(yw = cluster2_of_3) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 2)

cluster3_of_3_df = data.frame(cluster3_of_3) %>%
  rename(yw = cluster3_of_3) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 3)

clusterk3_df_temp = join(cluster1_of_3_df, cluster2_of_3_df, type="full")
```

```
## Joining by: yw, cluster
```

```r
clusterk3_df = join(clusterk3_df_temp, cluster3_of_3_df, type = "full") %>%
  mutate(cluster = as.character(cluster)) %>%
  as_tsibble(index = yw)
```

```
## Joining by: yw, cluster
```

```r
ggplot(clusterk3_df) + geom_point(aes(x=yw, y=0, color = cluster))
```



Next, let's try 5 clusters.

```r
set.seed(123) # set seed for reproducibility

#do the clustering
compsk4 <- kmeans(comps, 5, nstart = 25, iter.max=1000)

# 3D plot
#plot3d(comps$PC1, comps$PC2, comps$PC3, col=compsk4$clust)
```

```r
# Cluster sizes
sort(table(compsk4$clust))
```

```
##
##   1   2   3   4   5
##   1   6  25  32  57
```

```r
clust <- names(sort(table(compsk4$clust)))
# First cluster
cluster1_of_4 = row.names(data[compsk4$clust==clust[1],])
# Second Cluster
cluster2_of_4 = row.names(data[compsk4$clust==clust[2],])
# Third Cluster
cluster3_of_4 = row.names(data[compsk4$clust==clust[3],])
# Fourth Cluster
cluster4_of_4 = row.names(data[compsk4$clust==clust[4],])
# Fifth Cluster
cluster5_of_4 = row.names(data[compsk4$clust==clust[5],])
```

```r
#visualize the clusters on a timeline

cluster1_of_4_df = data.frame(cluster1_of_4) %>%
  rename(yw = cluster1_of_4) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 1)

cluster2_of_4_df = data.frame(cluster2_of_4) %>%
  rename(yw = cluster2_of_4) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 2)

cluster3_of_4_df = data.frame(cluster3_of_4) %>%
  rename(yw = cluster3_of_4) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 3)

cluster4_of_4_df = data.frame(cluster4_of_4) %>%
  rename(yw = cluster4_of_4) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 4)

cluster5_of_4_df = data.frame(cluster5_of_4) %>%
  rename(yw = cluster5_of_4) %>%
  mutate(yw = yearweek(yw)) %>%
  mutate(cluster = 5)

clusterk4_df_temp = join(cluster1_of_4_df, cluster2_of_4_df, type="full")
```

```
## Joining by: yw, cluster
```

```r
clusterk4_df_temp_2 = join(clusterk4_df_temp, cluster3_of_4_df, type="full")
```

```
## Joining by: yw, cluster
```

```r
clusterk4_df_temp_3 = join(clusterk4_df_temp_2, cluster4_of_4_df, type="full")
```

```
## Joining by: yw, cluster
```

```r
clusterk4_df = join(clusterk4_df_temp_3, cluster5_of_4_df, type = "full") %>%
  mutate(cluster = as.character(cluster)) %>%
  as_tsibble(index = yw)
```

```
## Joining by: yw, cluster
```

```r
ggplot(clusterk4_df) + geom_point(aes(x=yw, y=0, color = cluster))
```