Midterm

Riley Payung

2019-10-08

Question 1

```
avocado %>%
select(region) %>%
unique()
```

region
Albany
Atlanta
BaltimoreWashington
Boise
Boston
BuffaloRochester
California
Charlotte
Chicago
CincinnatiDayton
Columbus
DallasFtWorth
Denver
Detroit
GrandRapids
GreatLakes
HarrisburgScranton
HartfordSpringfield
Houston
Indianapolis
Jacksonville
LasVegas
LosAngeles
Louisville
MiamiFtLauderdale
Midsouth
Nashville
NewOrleansMobile
NewYork
Northeast
NorthernNewEngland
Orlando
Philadelphia
PhoenixTucson

region
Pittsburgh
Plains
Portland
RaleighGreensboro
RichmondNorfolk
Roanoke
Sacramento
SanDiego
SanFrancisco
Seattle
SouthCarolina
SouthCentral
Southeast
Spokane
StLouis
Syracuse
Tampa
TotalUS
West
WestTexNewMexico

   i. BaltimoreWashington

  ii. TotalUS

Question 2

```r
avocado2 <- avocado %>%
  filter(region=='TotalUS',type=='conventional')
head(avocado2)
```

week_number	Date	AveragePrice	Total Volume	4046	4225	4770	Total Bags	Sma
52	2015-12-27	0.95	27297984	9626901	10197890	1184340.1	6288852	4
51	2015-12-20	0.98	25083647	8710022	9329862	1201020.0	5842744	4
50	2015-12-13	0.93	28041335	9855054	10805839	1016163.2	6364280	4
49	2015-12-06	0.89	28800397	9405464	12160839	931830.6	6302263	5
48	2015-11-29	0.99	22617999	8094804	9003178	731008.4	4789009	3
47	2015-11-22	0.96	25114228	8571850	10389880	804662.8	5347836	4

   i. 169 rows.

  ii. The time period of each row corresponds to the week that avacados were sold over the entirety of the US between 2015 and 2017 and the end of 2018. There is missing data from week 1 of 2018 to week 40 of 2018. Data begins from 2018 in week 41.

Question 3

```
avocado2 %>%
  ggplot(mapping = aes(x = Date, y = AveragePrice)) +
    geom_smooth() +
    geom_line()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```
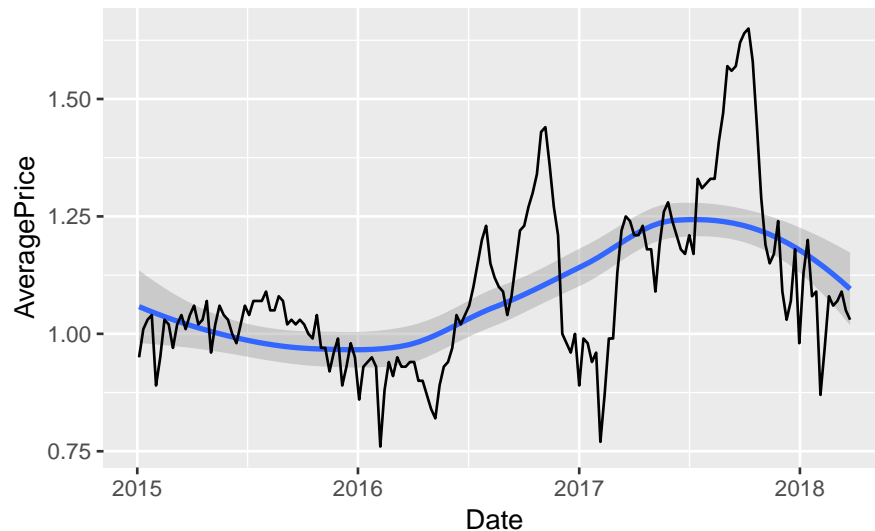

There are some seasonal trends in the graph; during the winter months, there tends to be less demand for avocados, so the price goes down, and in the summer months, the price tends to go up due to increased demand. The price of avocados goes up quite a lot in recent years due to demand, drought, fire, trade, etc. The price began to go down again at the end of 2018.

Question 4

```
avocado3 <- avocado %>%
  filter(region=="TotalUS" | region=="BaltimoreWashington", type=="conventional") %>%
  group_by(region)
# Graphical representation:
avocado3 %>%
  ggplot(mapping = aes(x = Date, y = AveragePrice, color = region)) +
    geom_smooth() +
    geom_line()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```
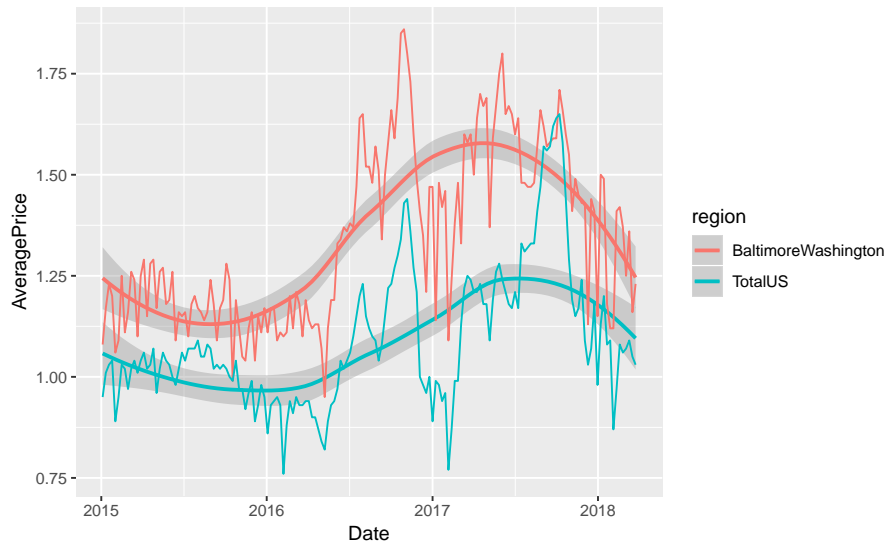
I would say that we are in fact paying FAIR prices for avocodos compared to the national average. Our prices did not increase nearly as much in the recent years as across the country. This could be in part from shipping prices to get avocados to certain regions of the United States, or maybe our avocados come from some other source than most of the entire country.

Question 5

```
avocado4 <- avocado %>%
   group_by(region) %>%
   filter(type=="conventional") %>%
   select(AveragePrice,region) %>%
   summarize(AverageByRegion = mean(AveragePrice))
 print(filter(avocado4,region =='HartfordSpringfield' | region=='PhoenixTucson'))
```

```
## # A tibble: 2 x 2
##   region            AverageByRegion
##   <chr>                       <dbl>
## 1 HartfordSpringfield          1.41
## 2 PhoenixTucson                0.728
```

The best region to get avocados in is PhoenixTucson, which had the lowest average cost of avocados at $0.73 (rounded up from $0.728). The worst region to get avocados in is HartfordSpringfield, which had the highest average cost of avocados at $1.41 (rounded up from $1.408)
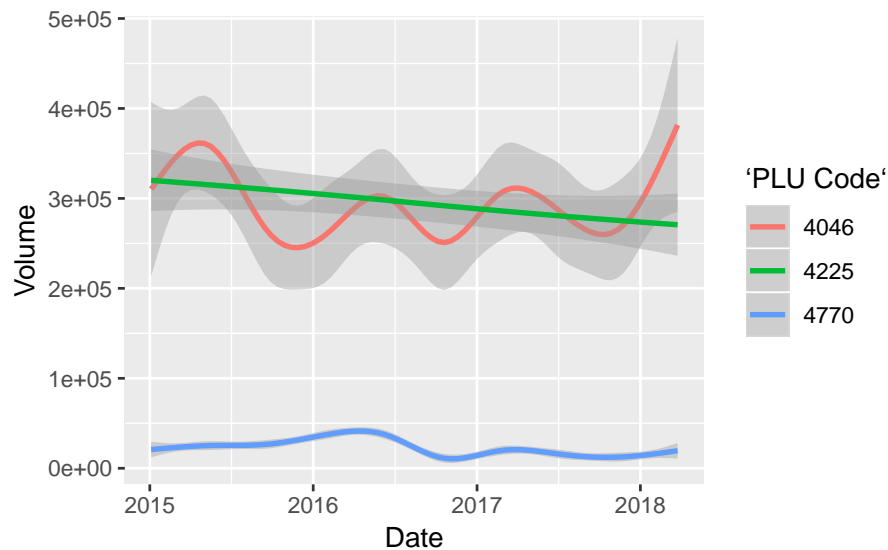
Question 6

```
  avocado5 <- avocado %>%
    gather(
      key = `PLU Code`,
      value = `Volume`,
      `4046`:`4770`
    )
```

```r
head(select(avocado5,type:`Volume`))
```

type	year	region	PLU Code	Volume
conventional	2015	Albany	4046	1036.74
conventional	2015	Albany	4046	674.28
conventional	2015	Albany	4046	794.70
conventional	2015	Albany	4046	1132.00
conventional	2015	Albany	4046	941.48
conventional	2015	Albany	4046	1184.27

```r
avocado5 %>%
  ggplot(mapping = aes(x = Date, y = Volume, color=`PLU Code`)) +
    geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


Question 7

```r
avocado6 <- avocado %>%
  separate(
    col = Date,
    into = c("year","month","day"),
    sep = "-",
    convert = FALSE
  )
head(avocado6)
```

week_number	year	month	day	AveragePrice	Total Volume	4046	4225	4770	Total Bags
52	2015	12	27	1.33	64236.62	1036.74	54454.85	48.16	8696.87
51	2015	12	20	1.35	54876.98	674.28	44638.81	58.33	9505.56

week_number	year	month	day	AveragePrice	Total Volume	4046	4225	4770	Total Bags
50	2015	12	13	0.93	118220.22	794.70	109149.67	130.50	8145.35
49	2015	12	06	1.08	78992.15	1132.00	71976.41	72.58	5811.16
48	2015	11	29	1.28	51039.60	941.48	43838.39	75.78	6183.95
47	2015	11	22	1.26	55979.78	1184.27	48067.99	43.61	6683.91