Price and sales - Avocados

April 15, 2021

1 Avocado Prices

Historical data on avocado prices and sales volume in multiple US markets

This dataset is available on Kaggle: https://www.kaggle.com/neuromusic/avocado-prices

To get a better picture of what is going, we will answer the following questions:

- What are the regions which the avocado is most and least expensive?
- Has the volume sales of avocado increased between 2015 to 2018?
- Has the price of avocados increased between 2015 to 2018?
- How do organic vs conventional avocados vary in prices?
- What is the annual average price by region?

```
[43]: Image('avocadopic.jpg')
```

[43]:

```
[50]: #Import useful libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statistics as st
import numpy as np
from IPython.display import Image
```

```
[9]: pwd
```

```
[9]: 'C:\\Users\\ginna'
```

1 - Data Importing

```
[13]: avocado = pd.read_csv("avocado.csv")
      #option 1 in order to visualize the data:
      avocado.head()
```

```
[13]:    Unnamed: 0        Date  AveragePrice  Total Volume     4046       4225  \
      0           0  2015-12-27          1.33      64236.62  1036.74   54454.85
      1           1  2015-12-20          1.35      54876.98   674.28   44638.81
      2           2  2015-12-13          0.93     118220.22   794.70  109149.67
      3           3  2015-12-06          1.08      78992.15  1132.00   71976.41
      4           4  2015-11-29          1.28      51039.60   941.48   43838.39

           4770  Total Bags  Small Bags  Large Bags  XLarge Bags         type  \
      0   48.16     8696.87     8603.62       93.25          0.0  conventional
      1   58.33     9505.56     9408.07       97.49          0.0  conventional
      2  130.50     8145.35     8042.21      103.14          0.0  conventional
      3   72.58     5811.16     5677.40      133.76          0.0  conventional
      4   75.78     6183.95     5986.26      197.69          0.0  conventional

         year  region
      0  2015  Albany
      1  2015  Albany
      2  2015  Albany
      3  2015  Albany
      4  2015  Albany
```

```
[81]: #option 2 in order to visualize the data set
      display(avocado)
```

```
             Unnamed: 0        Date  AveragePrice  Total Volume     4046       4225  \
      0               0  2015-12-27          1.33      64236.62  1036.74   54454.85
      1               1  2015-12-20          1.35      54876.98   674.28   44638.81
      2               2  2015-12-13          0.93     118220.22   794.70  109149.67
      3               3  2015-12-06          1.08      78992.15  1132.00   71976.41
      4               4  2015-11-29          1.28      51039.60   941.48   43838.39
      …             …         …            …            …         …          …
      18244           7  2018-02-04          1.63      17074.83  2046.96    1529.20
      18245           8  2018-01-28          1.71      13888.04  1191.70    3431.50
      18246           9  2018-01-21          1.87      13766.76  1191.92    2452.79
      18247          10  2018-01-14          1.93      16205.22  1527.63    2981.04
      18248          11  2018-01-07          1.62      17489.58  2894.77    2356.13

               4770  Total Bags  Small Bags  Large Bags  XLarge Bags         type  \
      0       48.16     8696.87     8603.62       93.25          0.0  conventional
      1       58.33     9505.56     9408.07       97.49          0.0  conventional
      2      130.50     8145.35     8042.21      103.14          0.0  conventional
      3       72.58     5811.16     5677.40      133.76          0.0  conventional
```

```
4        75.78     6183.95     5986.26      197.69           0.0  conventional
…        …         …           …           …            …           …
18244    0.00     13498.67    13066.82      431.85           0.0       organic
18245    0.00      9264.84     8940.04      324.80           0.0       organic
18246  727.94      9394.11     9351.80       42.31           0.0       organic
18247  727.01     10969.54    10919.54       50.00           0.0       organic
18248  224.53     12014.15    11988.14       26.01           0.0       organic

       year            region
0      2015            Albany
1      2015            Albany
2      2015            Albany
3      2015            Albany
4      2015            Albany
…     …               …
18244  2018  WestTexNewMexico
18245  2018  WestTexNewMexico
18246  2018  WestTexNewMexico
18247  2018  WestTexNewMexico
18248  2018  WestTexNewMexico

[18249 rows x 14 columns]
```

-For the data shape we got 18249 rows and 14 columns -I also can find that information throught: avocado.shape

[19]:
```
avocado.columns
```

[19]:
```
Index(['Unnamed: 0', 'Date', 'AveragePrice', 'Total Volume', '4046', '4225',
       '4770', 'Total Bags', 'Small Bags', 'Large Bags', 'XLarge Bags', 'type',
       'year', 'region'],
      dtype='object')
```

[20]:
```
#check if there is nul values
avocado.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18249 entries, 0 to 18248
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Unnamed: 0    18249 non-null  int64
 1   Date          18249 non-null  object
 2   AveragePrice  18249 non-null  float64
 3   Total Volume  18249 non-null  float64
 4   4046          18249 non-null  float64
 5   4225          18249 non-null  float64
 6   4770          18249 non-null  float64
 7   Total Bags    18249 non-null  float64
 8   Small Bags    18249 non-null  float64
 9   Large Bags    18249 non-null  float64
 10  XLarge Bags   18249 non-null  float64
 11  type          18249 non-null  object
 12  year          18249 non-null  int64
 13  region        18249 non-null  object
dtypes: float64(9), int64(2), object(3)
memory usage: 1.9+ MB
```

[21]:

```
avocado.describe()
```

[21]:

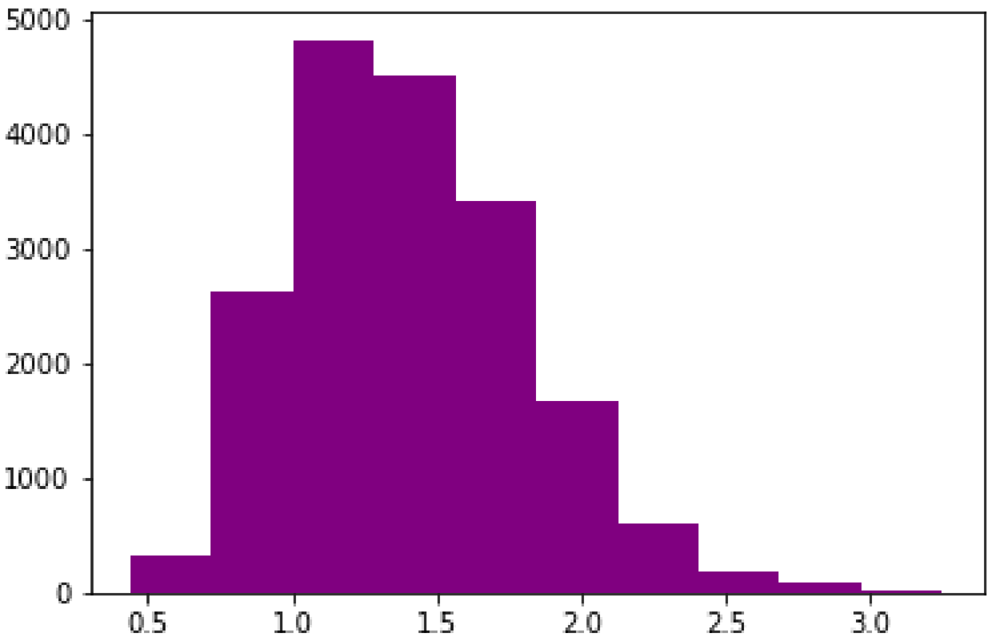
```
         Unnamed: 0  AveragePrice  Total Volume          4046          4225  \
count  18249.000000  18249.000000  1.824900e+04  1.824900e+04  1.824900e+04
mean      24.232232      1.405978  8.506440e+05  2.930084e+05  2.951546e+05
std       15.481045      0.402677  3.453545e+06  1.264989e+06  1.204120e+06
min        0.000000      0.440000  8.456000e+01  0.000000e+00  0.000000e+00
25%       10.000000      1.100000  1.083858e+04  8.540700e+02  3.008780e+03
50%       24.000000      1.370000  1.073768e+05  8.645300e+03  2.906102e+04
75%       38.000000      1.660000  4.329623e+05  1.110202e+05  1.502069e+05
max       52.000000      3.250000  6.250565e+07  2.274362e+07  2.047057e+07

               4770    Total Bags    Small Bags    Large Bags    XLarge Bags  \
count  1.824900e+04  1.824900e+04  1.824900e+04  1.824900e+04   18249.000000
mean   2.283974e+04  2.396392e+05  1.821947e+05  5.433809e+04    3106.426507
std    1.074641e+05  9.862424e+05  7.461785e+05  2.439660e+05   17692.894652
min    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00       0.000000
25%    0.000000e+00  5.088640e+03  2.849420e+03  1.274700e+02       0.000000
50%    1.849900e+02  3.974383e+04  2.636282e+04  2.647710e+03       0.000000
75%    6.243420e+03  1.107834e+05  8.333767e+04  2.202925e+04     132.500000
max    2.546439e+06  1.937313e+07  1.338459e+07  5.719097e+06  551693.650000

               year
count  18249.000000
mean    2016.147899
std        0.939938
min     2015.000000
25%     2015.000000
50%     2016.000000
75%     2017.000000
max     2018.000000
```

[22]:

```
#Average Price
plt.hist(avocado['AveragePrice'],bins=10, histtype='bar', color='purple')
```

[22]:

```
(array([ 331., 2632., 4824., 4506., 3412., 1672.,  598.,  177.,   86.,
          11.]),
 array([0.44 , 0.721, 1.002, 1.283, 1.564, 1.845, 2.126, 2.407, 2.688,
        2.969, 3.25 ]),
 <BarContainer object of 10 artists>)
```



```
[23]: prices = avocado['AveragePrice']
      st.mean(prices)
```

```
[23]: 1.405978409775878
```

```
[24]: avocado['Total Volume'].sum(), avocado['4046'].sum() + avocado['4225'].sum() +␣
       ↪avocado['4770'].sum()
```

```
[24]: (15523402593.400002, 11150188799.32)
```

```
[25]: #Conventional X Organic percentage
      type_data = (avocado['type'].value_counts()/avocado.shape[0])*100
      display(round(type_data,4))
```

```
conventional    50.0082
organic         49.9918
Name: type, dtype: float64
```

2 Question 1:

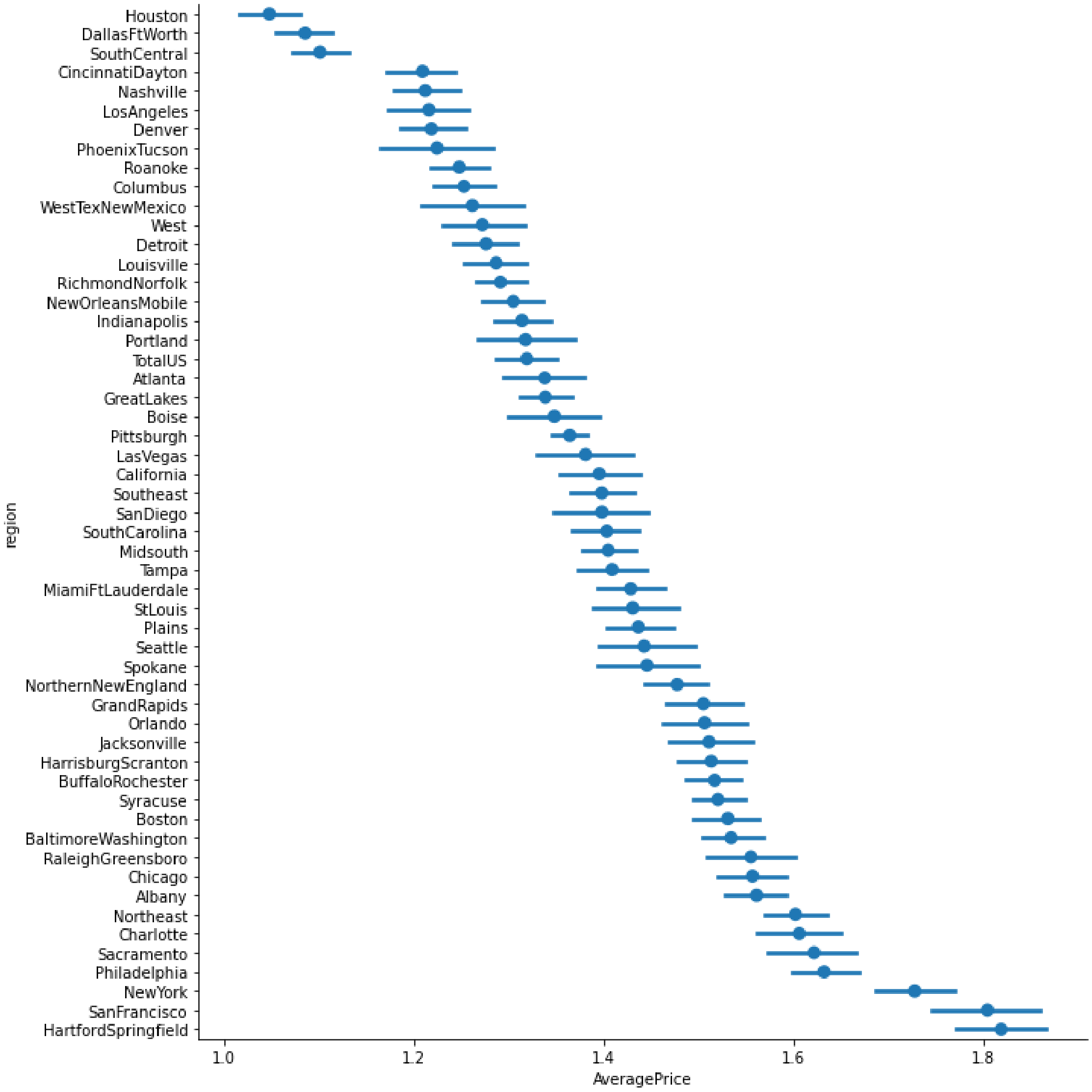
What are the regions which the avocado is most and least expensive?

```
[35]: #Ordering graph by region & average price
      order = (
          avocado.groupby('region')['AveragePrice']
          .mean()
          .sort_values()
          .index)
```

[27]:

```
#Graph comparing all regions with their mean/IQR prices
graph = sns.factorplot('AveragePrice','region', data=avocado,
                        size=10,
                        order=order,
                       join=False,)
```

```
C:\Users\ginna\anaconda3\lib\site-packages\seaborn\categorical.py:3704:
UserWarning: The `factorplot` function has been renamed to `catplot`. The
original name will be removed in a future release. Please update your code. Note
that the default `kind` in `factorplot` (`'point'`) has changed `'strip'` in
`catplot`.
  warnings.warn(msg)
C:\Users\ginna\anaconda3\lib\site-packages\seaborn\categorical.py:3710:
UserWarning: The `size` parameter has been renamed to `height`; please update
your code.
  warnings.warn(msg, UserWarning)
C:\Users\ginna\anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variables as keyword args: x, y. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or
misinterpretation.
  warnings.warn(
```



#According to the result, it shows that the TOP 3 most expensive regions for avocados across all years are:

1-Hartford Springfield, 2 -San Francisco and 3 -New York

#And the TOP 3 least expensive regions for avocados are:

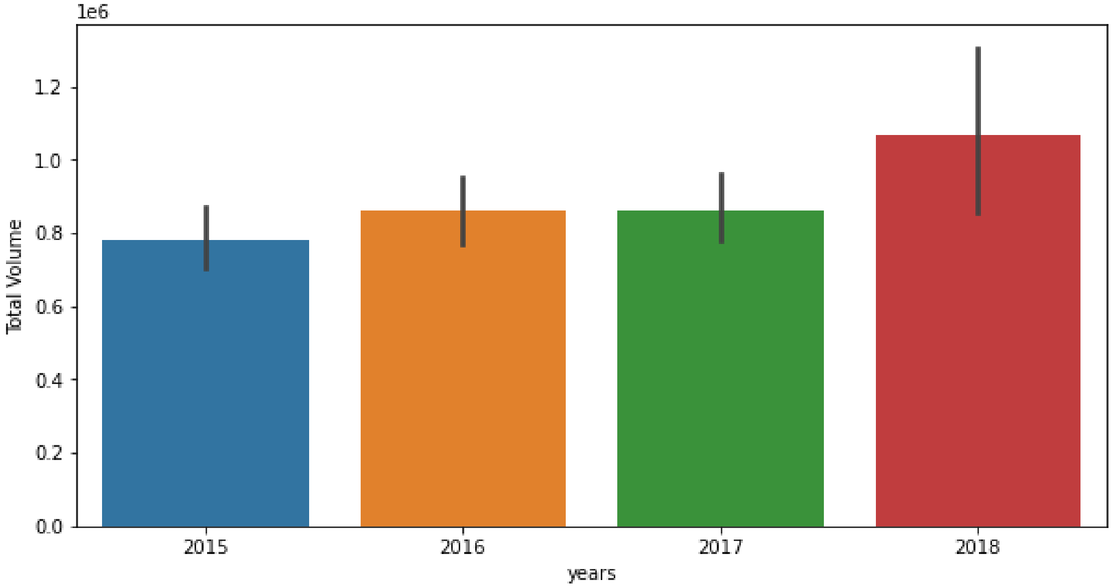
1- Houston, 2- Dallas Fort Worth and 3- South Central.

3 Question 2:

Has the volume sales of avocado increased between 2015 to 2018?

[28]:

```
# ploting total volume x years
plt.figure(figsize=(10,5))
sns.barplot(x=avocado['year'],y=avocado['Total Volume'])
plt.xlabel('years')
plt.ylabel('Total Volume')
plt.show()
```

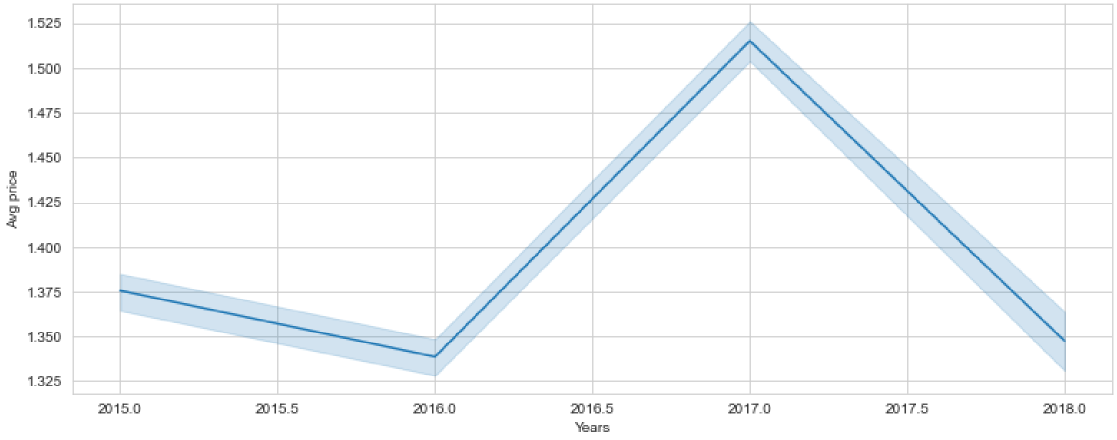


4 Question 3:

Has the price of avocados increased between 2015 to 2018?

[40]:

```
# ploting avg price x years
plt.figure(figsize=(13,5))
sns.lineplot(x=avocado['year'],y=avocado['AveragePrice'])
plt.xlabel('Years')
plt.ylabel('Avg price')
plt.show()
```



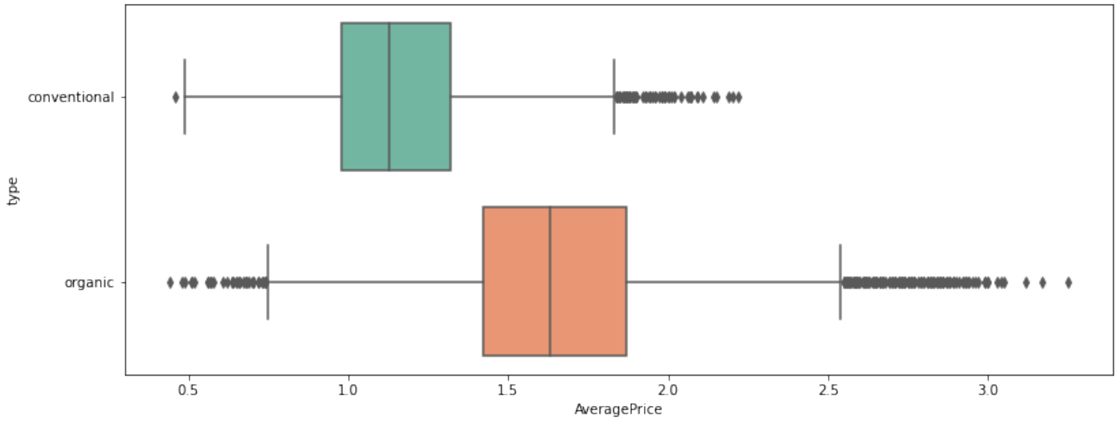
Avocados Average Price had a peak of increase around the start of 2017, reaching 1.52 dollar.

5 Question 4:

How do organic vs conventional avocados vary in prices?

[30]:
```
# Analysing Type of avocado X Avg price
plt.figure(figsize=(13,5))
sns.boxplot(y="type", x="AveragePrice", data=avocado, palette = 'Set2')
```

[30]:
```
<AxesSubplot:xlabel='AveragePrice', ylabel='type'>
```



Organic avocado is 0.5 dollar more expensive than conventional avocado.

6 Question 5:

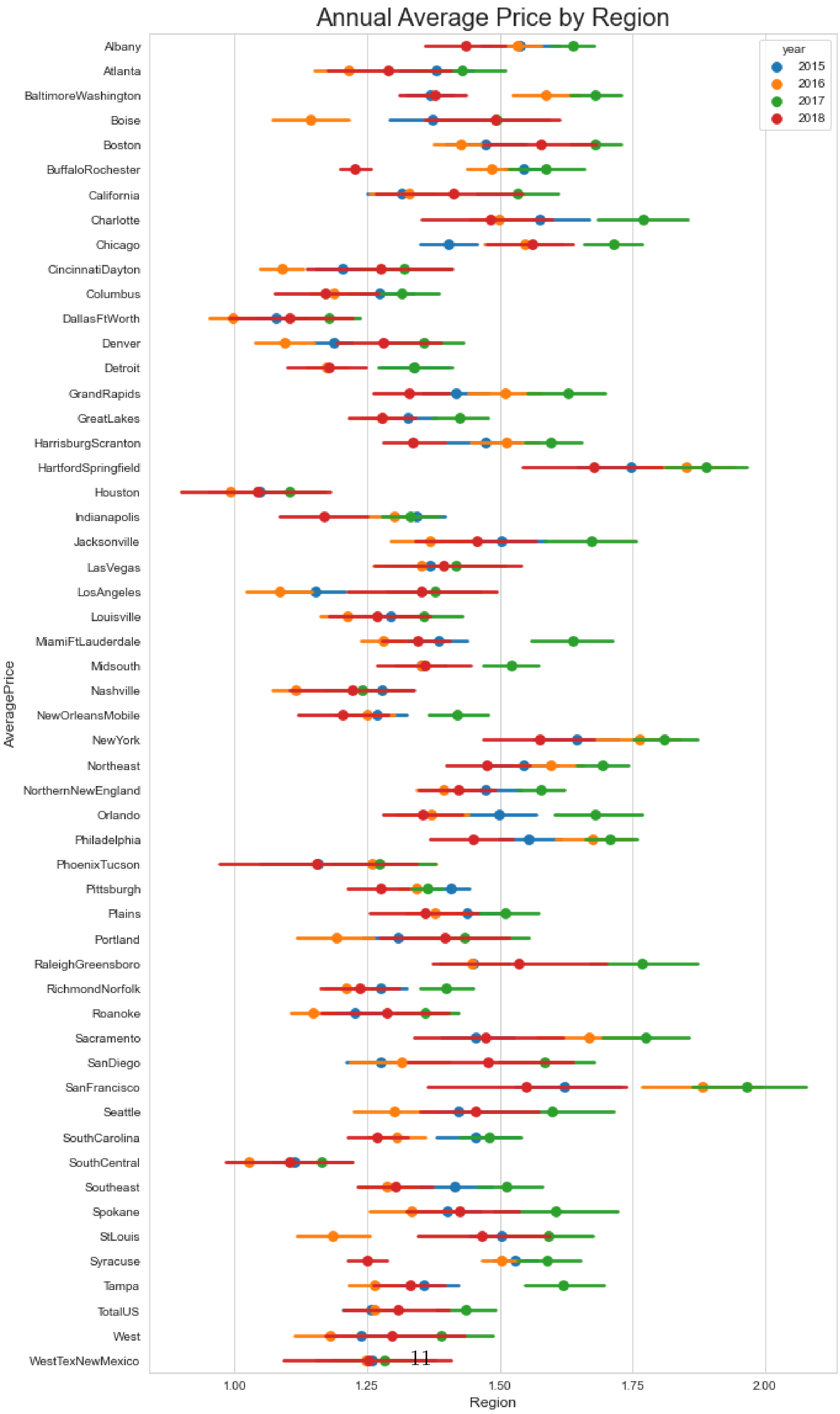
What is the annual average price by region?

[39]:

```
# Annual Average price by region

plt.figure(figsize=(10,20))
sns.set_style('whitegrid')
sns.pointplot(x='AveragePrice',y='region',data=avocado, hue='year',join=False)
plt.xticks(np.linspace(1,2,5))
plt.xlabel('Region',{'fontsize' : 'large'})
plt.ylabel('AveragePrice',{'fontsize':'large'})
plt.title("Annual Average Price by Region",{'fontsize':20})
```

[39]:

```
Text(0.5, 1.0, 'Annual Average Price by Region')
```



Annual Average Price by Region
year
● 2015
● 2016
● 2017
● 2018
AveragePrice
Region
1.00
1.25
1.50
1.75
2.00
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
Pittsburgh
Plains
Portland
RaleighGreensboro
RichmondNorfolk
Roanoke
Sacramento
SanDiego
SanFrancisco
Seattle
SouthCarolina
SouthCentral
Southeast
Spokane
StLouis
Syracuse
Tampa
TotalUS
West
WestTexNewMexico

-The green line shows that in 2017 Average Price was most expensive in almost all regions.

```
[52]: Image('avocadopic2.jpg')
```

[52]:

7 Conclusions:

- Average price is 1,40
- Avocados Average price had an increase peak in the begining of 2017
- Top 3 most expensive regions for avocados across all years are:Hartford Springfield, San Francisco and New York
- Top 3 least expensive regions for avocados are: Houston, Dallas Fort Worth and South Central
- Organic avocado is 0.5 dollar more expensive than conventional avocado
- Average Price was more expensive in almost all regions in 2017
- The ideal region for millenial to live would be Houston, the region on USA where the average price was least expensive
- When it comes to Total Volume, 2018 holds the biggest volume.
- Avocado has been more purchased over time.