

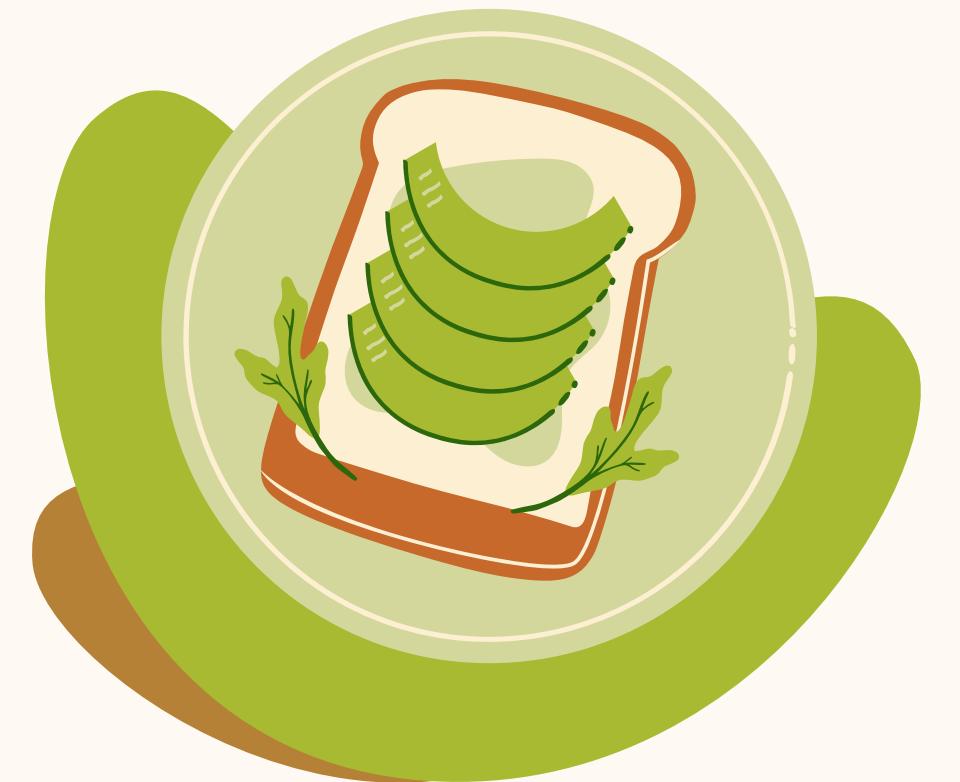


Where to Live + Afford

AVOCADO TOAST



By Brit Quamberg



Background

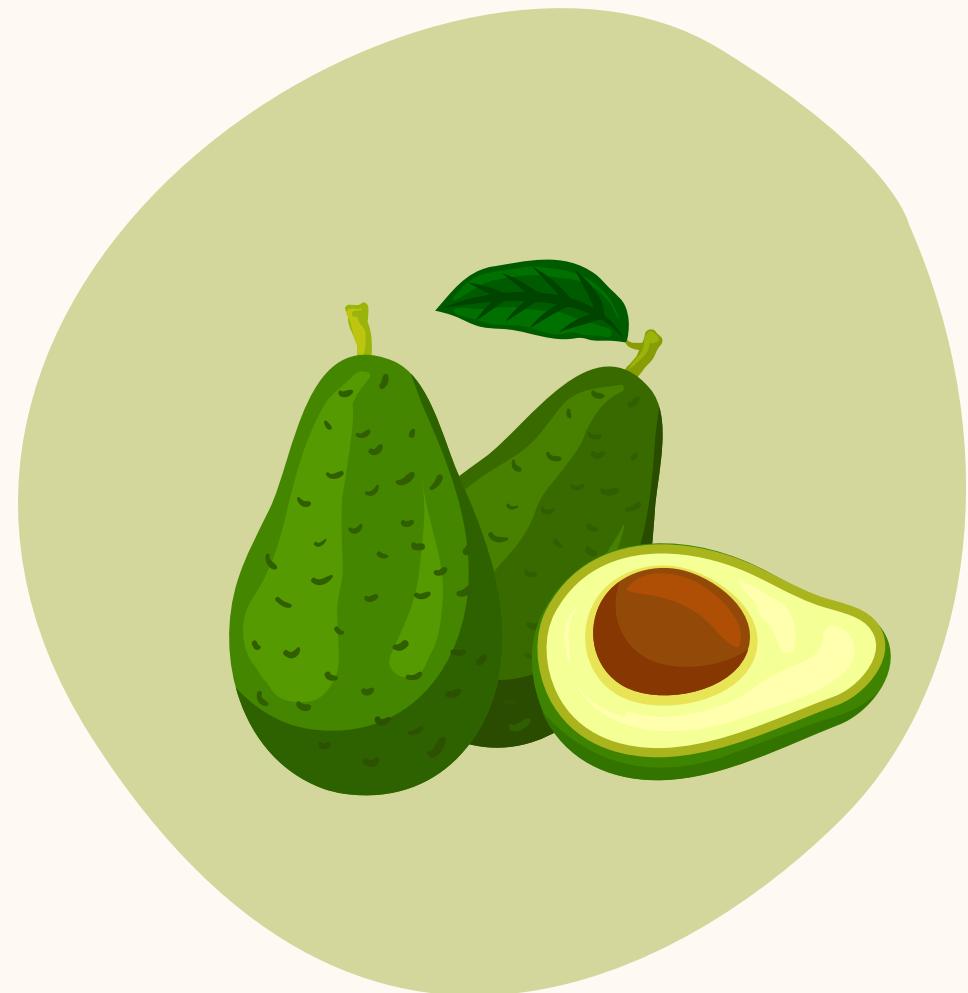
BRIT QUARNBERG

Data science is an adjacent yet new field for me. For the past 19 years, I've taught students mathematics (including statistics) in-person, online, and internationally. I'm a data scientist, teacher, mathematician, mother, wife, traveller, and solo artist on this project.



Project: Avocado Toast Dilemma

BEYOND BASEMENTS?



"It is a well known fact that Millennials LOVE Avocado Toast. It is also a well known fact that Millennials live in their parents' basements.

Clearly, they aren't buying houses because they are buying too much Avocado Toast. "



My Focus

WHICH CITY HAS AFFORDABLE AVOCADOS?

I looked at these main questions:

- Does the price and volume of avocados differ between 3 major cities?
- Is the size of bag avocados are sold in more common in those 3 cities and does the bag size effect the price of avocados?



Comparison of 3 CITIES

One from the MidWest,

one from the mountains,

& one by the ocean

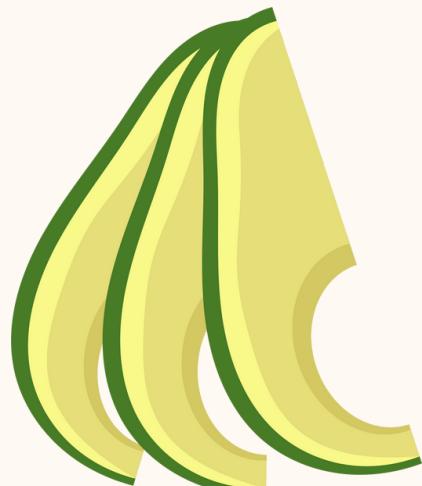




Analysis METHODS



To answer if the price and volume of avocados differ between Chicago, Denver and Los Angeles, I ran an ANOVA (analysis of variance) of the average price vs the cities and the volume of avocados vs the cities. This showed me how much influence each average price and volume have on the cost in those cities.



Analysis METHODS

-  To answer if the size of bag avocados sold is more common in Chicago, Denver or Los Angeles, I ran an Independent Chi-Square.
-  To answer if the price is effected by the region even when considering the bag size in those cities, I ran an ANCOVA, or analysis of covariance.

Results

AVERAGE PRICE

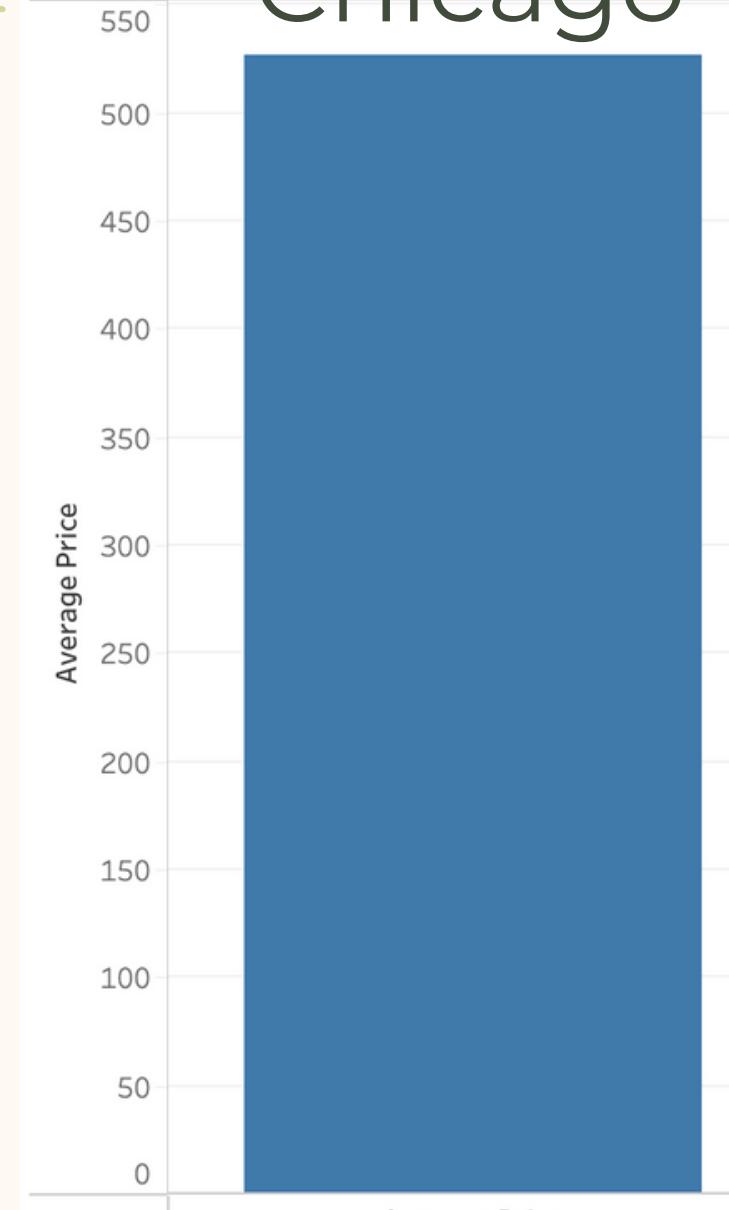
Overall, you can see that the highest average price, was Chicago between the 3 cities.

Statistically, the ANOVA revealed there is indeed a significant difference between these 3 cities.

Chicago has a 28% higher average price of avocados than both Denver and Los Angeles.

Average Price by City

Chicago



Region
Denver

Los Angeles

Denver

Los Angeles

Average Price

Average Price

Average Price

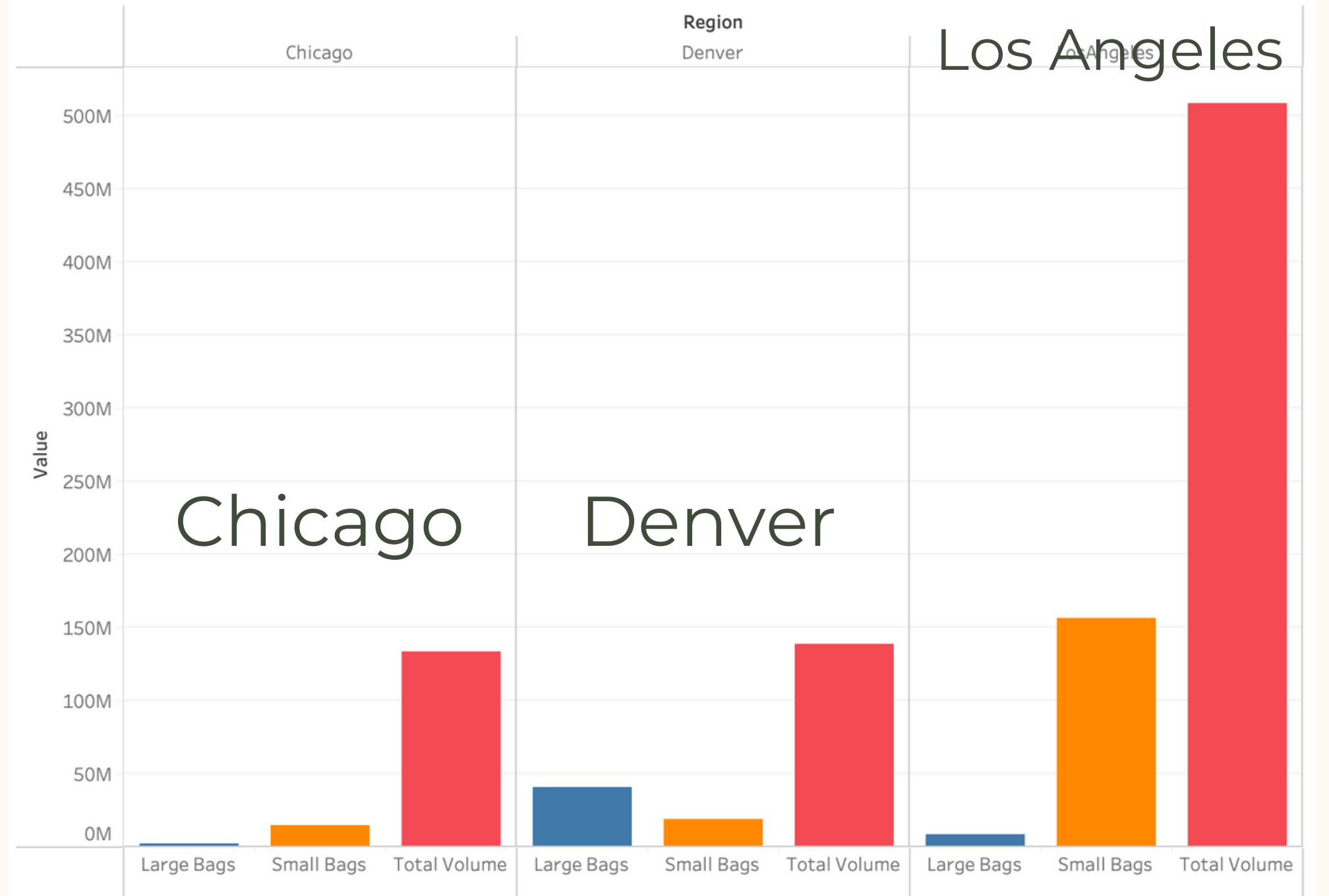


Results

TOTAL VOLUME



Large bags, Small Bags & Total Volume



Here you can see that Los Angeles sold more avocados altogether.

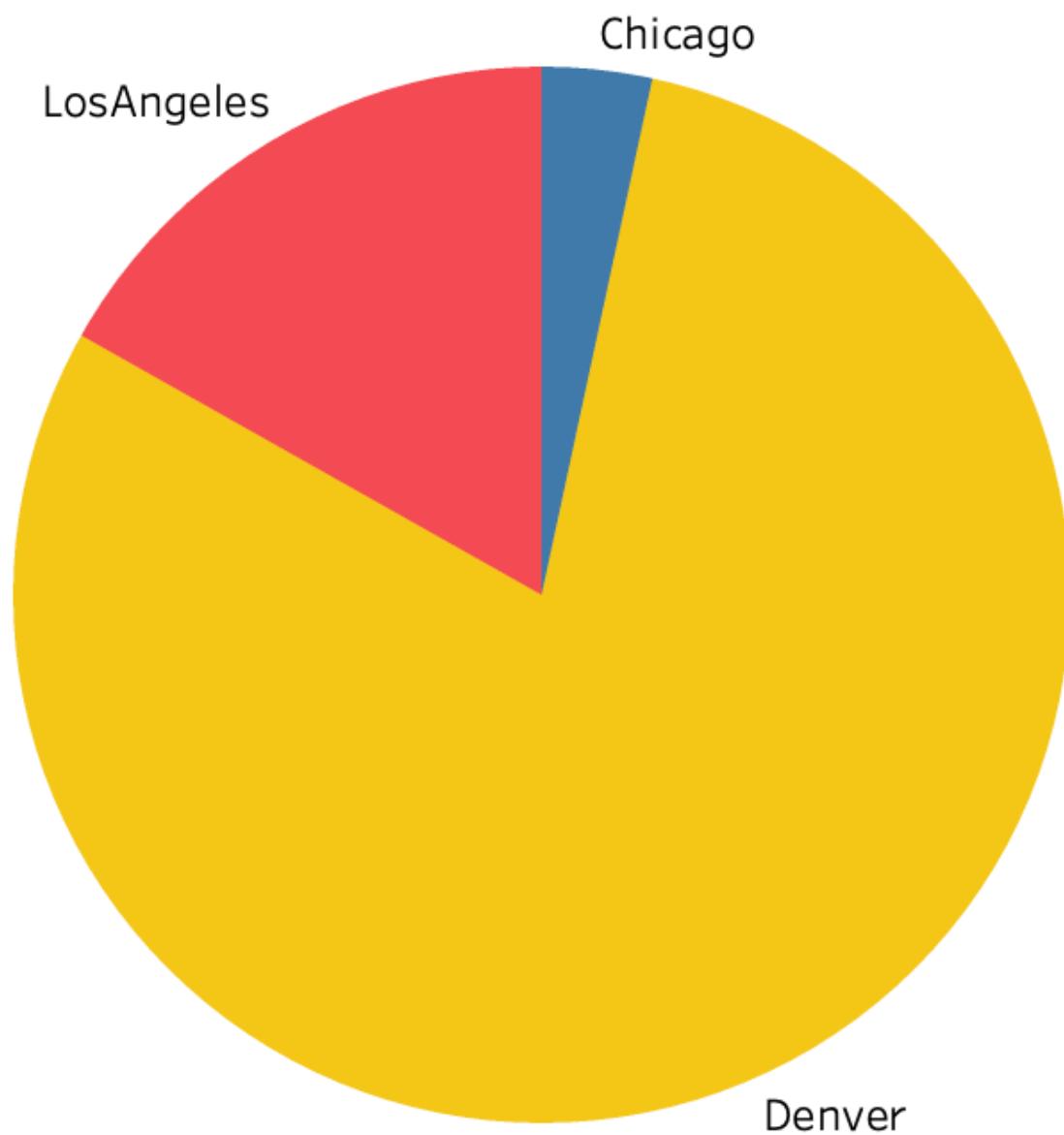
Statistically, the ANOVA revealed the same thing, and after further analysis it revealed

Los Angeles sold nearly 4 times more volume of avocados than both Chicago and Denver.

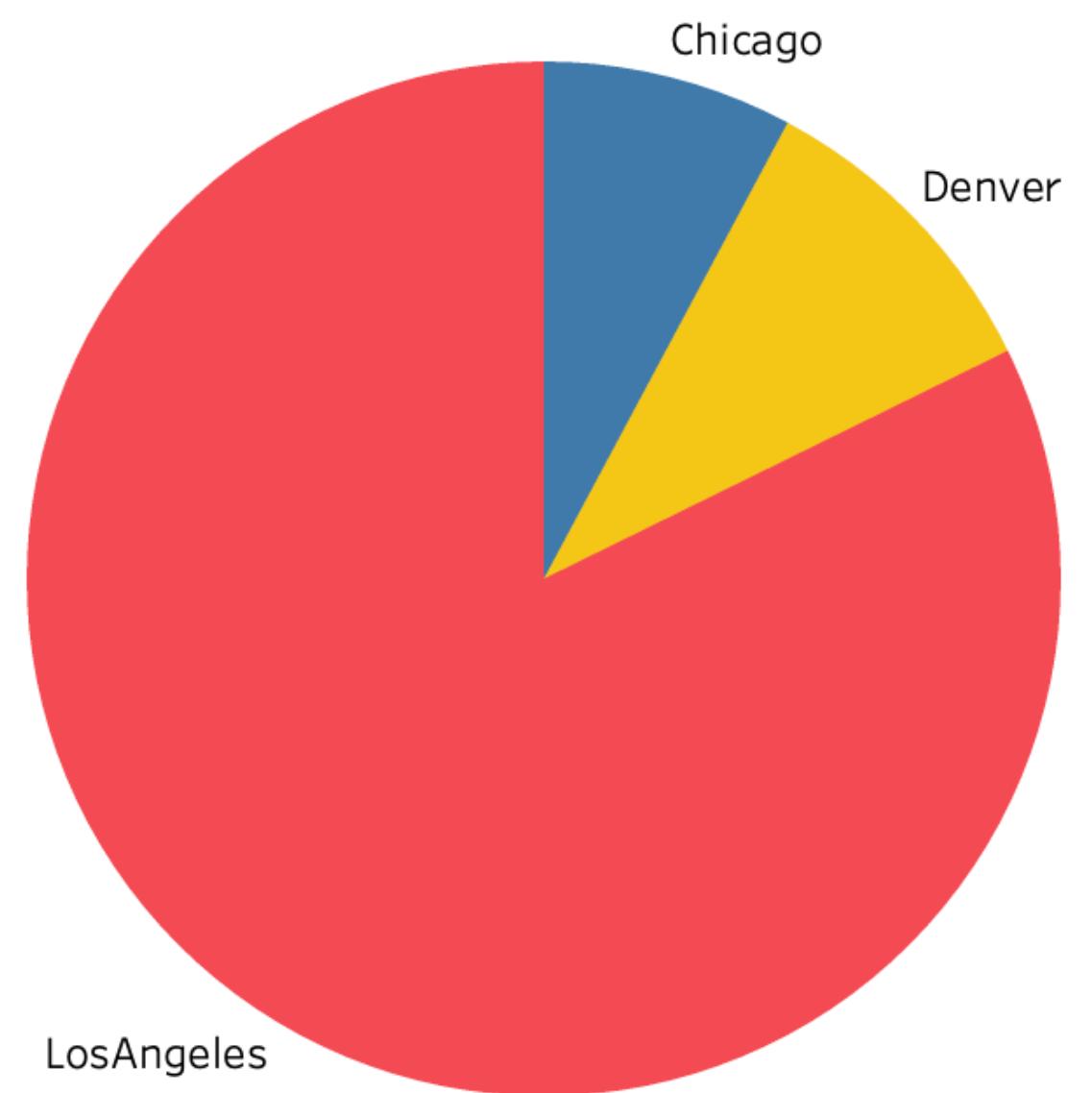
Results

BAG SIZE

I couldn't help but notice the size of bags. Look at this visual comparison of bag size and



Denver: greatest Large Bag Sales



Los Angeles: more Small Bag Sales



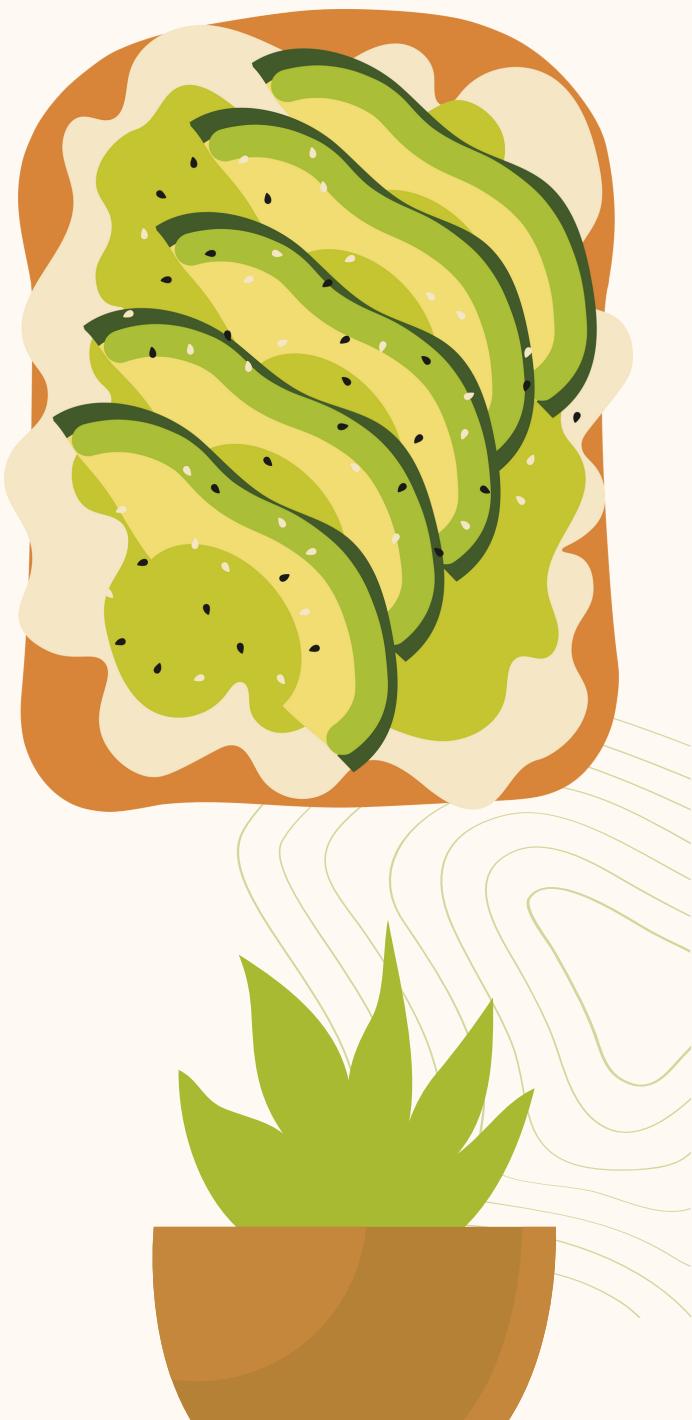
Results

DO CERTAIN REGIONS SELL MORE BAGS OF A CERTAIN SIZE?

I ran an Independent Chi-Square and found that

Denver sold 223 times more large bags than Chicago and 45 times more large bags than Los Angeles*,

but, Denver sold 1/3 of the number of small bags than both Chicago and Los Angeles.*



Results

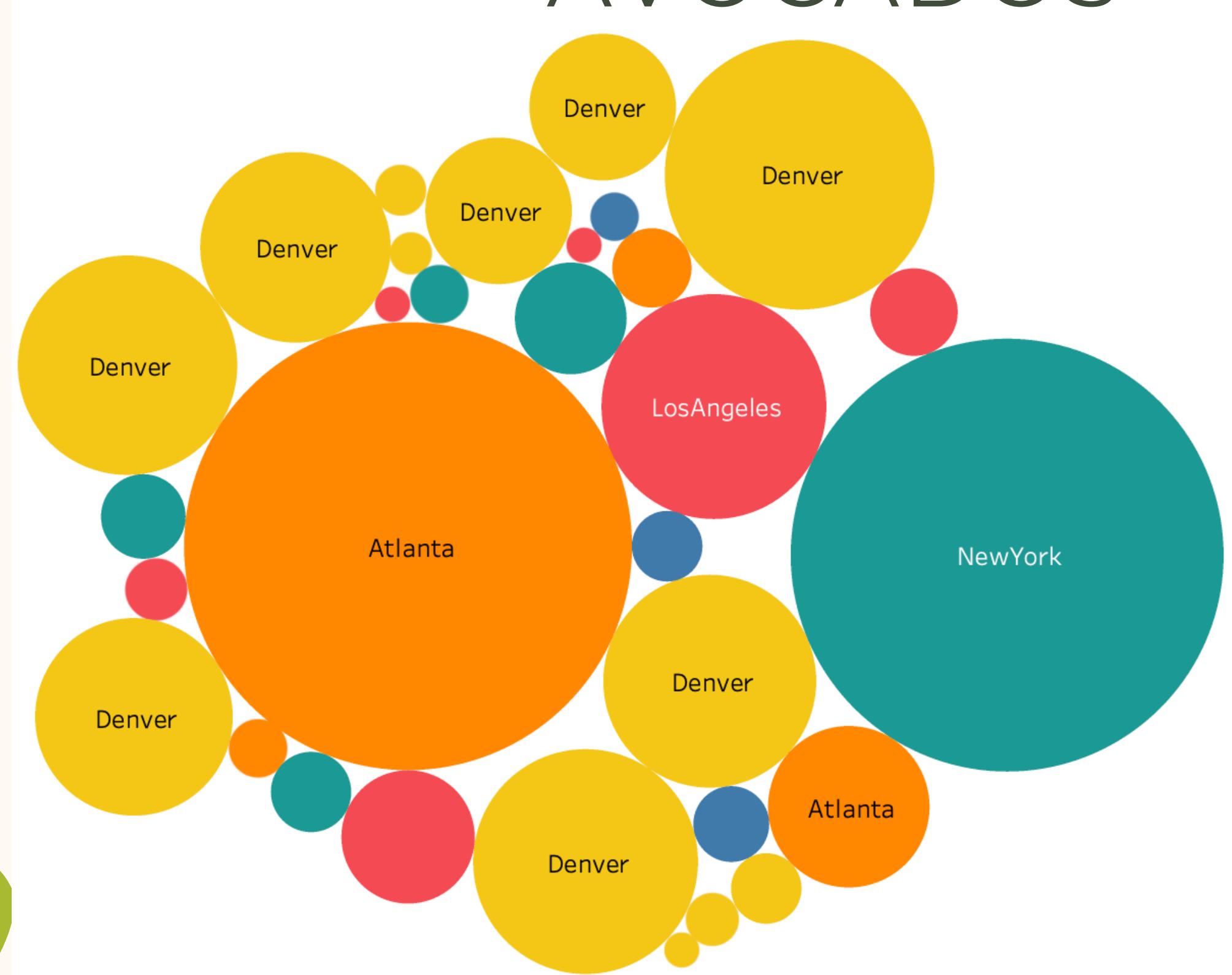
IS THE PRICE OF AVOCADOS INFLUENCED BY CITY & BAG SIZE?

YES! Statistically, the ANCOVA revealed that bag size and the city influenced the average price of avocados.

Check out the visualizations of this on the next slides.

Results

LARGE BAGS & PRICE OF AVOCADOS



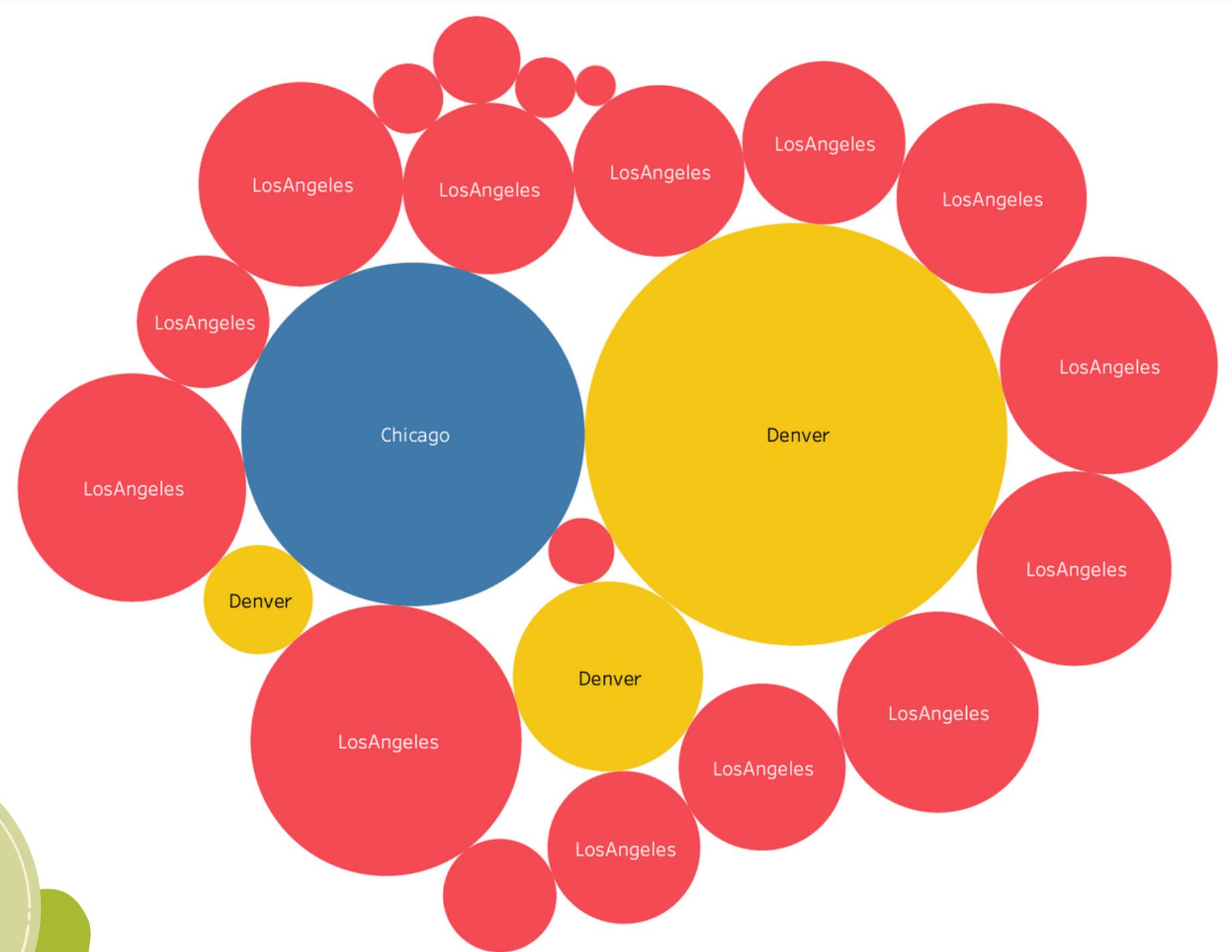
For this visualization, I threw in data for Atlanta and New York.

The larger the size of circle, the greater the price.

You can see that Denver (yellow) still sold more large bags of avocados but the average price was smaller than other cities, but Los Angeles had the lowest price.

Results

SMALL BAGS & PRICE OF AVOCADOS



The greater the price, the larger the circle.

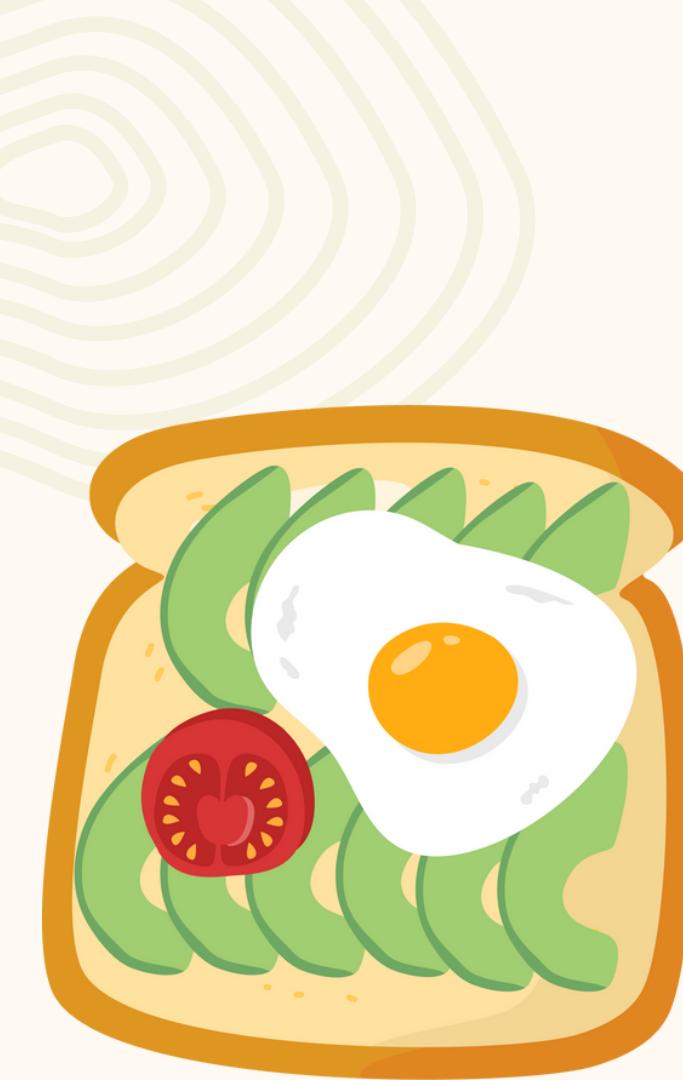
You can see that Chicago (blue) sold fewer small bags than Denver and Los Angeles (only one circle)

Los Angeles (red) had a smaller average price than the other cities, but sold more small bags.

Denver, in this representation had the most expensive avocados but sold fewer small bags.

Summary + Conclusions

- Avocados are significantly higher priced in Chicago. It would not be an avocado toaster's dream to live there as they would have to decrease their avocado toast consumption.
- Los Angeles sold more avocados than the other cities and had the lowest price. Though this is appealing, it is also a well known fact that California is more expensive (an analysis for another day) and would not be ideal.
- If you're into buying avocados in bulk, and want a city where you could afford the Avocado Toast lifestyle, Denver seems to be the place for you.



Questions?



Personally, I would like to analyze more cities in this analysis.

I would also like to involve the cost of housing in this analysis to have it be more comprehensive and be able to actually answer the avocado toast dilemma.

How I coded the bag size seems to have altered the results between what I got statistically with the Independent Chi-Squared and the ANCOVA according to the Tableau representations.

Technical Details

In order to run the ANOVA for both average price and volume, I had to perform the following data wrangling:

- Subset my data to only include Chicago, Denver and Los Angeles
- Recode the cities as a number



Technical Details

In order to run the ANOVA for both average price and volume, I had to run the **tests of assumptions** necessary to run an ANOVA.

- The sample size was sufficiently large (>20)
- There was no possible overlap in the data as I selected cities, not regions and cities that could be included in the region.
- To test for normality for the price, I had to take the square root of the price. Even then, though, the data was almost split or bimodal.
- To test for normality for the volume, I eventually used the log of the volume, though the result was totally bimodal. I am not sure how to deal with that, so I left it.
- The assumption for homogeneity of variance on both the price and the volume was not, so technically we should proceed with caution/consider subsetting the data at the split?



Technical Details



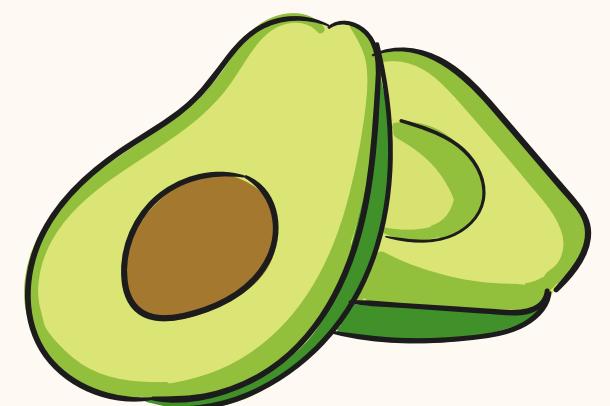
The Post Hoc ran after the ANOVA confirmed what we can see in the Tableau charts:

- For price, comparing the mean average price for the 3 cities and then performing simple calculations revealed that Chicago was region 2, and had a significantly higher average price than Denver and Los Angeles by about 0.34, or a 28% higher price. For this problem, that means it would be more difficult to live the Avocado Toast Live in Chicago than in Denver or Los Angeles according to the price of Avocados.
- For the volume, comparing the mean total volume for the 3 cities, and then performing simple calculations revealed that Los Angeles sold a mean volume of avocados that is nearly 4 times bigger than both Denver and Chicago, and that Denver and Chicago sold a similar volume of avocados.

Technical Details

In order to run the **Independent Chi-Square** for telling if bag sizes were statistically more common in one city over another, I had to perform the following data wrangling:

- Subset my data to only include Chicago, Denver and Los Angeles
- Recode the cities as a number (Denver = 0, Los Angeles = 1, Chicago = 2)
- Combine the bag size columns using a condition that whichever bag size was sold the most in the city was the new label for that cell. For example, if a city sold 4 small bags, 9 large bags and 2 extra large bags, the label for that city was "large".



Technical Details

In order to run the **Independent Chi-Square** for telling if bag sizes were statistically more common in one city over another, I had to test for the following assumptions:

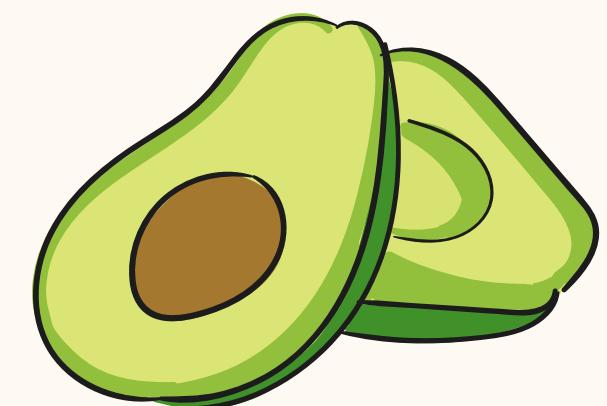
- I made a contingency table in order to run the Chi-Squared. All cells are mutually exclusive.
- After running the test, I got the expected values. All expected values were greater than 5.
- Bag sizes as far as I can tell are independent of the price.



Technical Details

The Post Hoc after the Independent Chi-Squared revealed something different than what I see in the Tableau Infographic, probably due to me NOT recoding for bag size in Tableau.

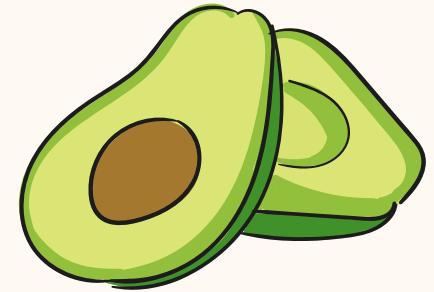
- What I got was that the number of large bags sold in Denver was 223 times larger than the number of large bags sold in Los Angeles, and 45 time more than the number of large bags sold in Chicago.
- The pie chart from Tableau doesn't match this result, so I edited my results on the presentation. I assume why this happened is because of the way I recoded the bag size.
- I used BINS in Tableau. This makes me wonder if I could perform an Independent Chi-Square with the bag sizes in bins with multiple variables for the analysis. Something to consider...



Technical Details

In order to run the ANCOVA, I had to perform the following data wrangling

- Subset my data to only include Chicago, Denver and Los Angeles
- Recode the cities as a number
- Combine the bag size columns using a condition that whichever bag size was sold the most in the city was the new label for that cell. For example, if a city sold 4 small bags, 9 large bags and 2 extra large bags, the label for that city was "large".
- Recode the bag size as a number



Technical Details

Then I ran the ANCOVA with the bag size as a covariate, comparing average price between the region. It showed even accounting for the size of bags sold, there IS a difference in the region and the price.

** However, when I ran the ANCOVA with the region as the covariate, comparing between the average price and the bag size, the test was not significant, specifically for bag size. This would mean if we looked at it that way, there is NOT a significant difference in bag size between cities when you take the price into consideration**