MODULE 4 ASSIGNMENT: VISUALIZATION

SUBMITTED BY:

PRATUL SUNIL AGARWAL
SHIFA CHAUDHARY
GAURAV GIRISH KULKARNI
SHRIRENGAPRIYA SREETHARAN
VAGEESHA MAIYA

VIZ Narrative: Airlines and Jet Fuel Costs

The purpose of this Visualization experiment is to represent the influence of ups and downs in the cost of jet fuel on the price of four airline company's stocks – AAL, DAL, UAL and LUV. The narrative revolves around basic notion <u>Does the cost of jet fuel affect the price of airline stocks</u>?

1. Datasets:

To start on a preliminary analysis, the first step was to grab some data. First, Yahoo Finance had the stock trading data for some of the bigger airlines. The individual stocks were group together in a star schema with data and company dimensions, mirroring the S&P 500 data used as an example. We have used this data <u>airline_stocks.csv</u> without any further manipulations.

The 4 airline stocks, we are demonstrating are:

AAL – American Airlines Group INC.

DAL – Delta Airlines

LUV – Southwest Airlines Co.

UAL – United Airlines Co.

The jet fuel prices come from the government site <u>US Energy Information Administration (EIA)</u>. The dataset we worked on is <u>jet_fuel_prices.csv</u>

2. The Columns in the airline_stocks file are as below:

TRADE_DATE

TRADE_JDN

OPEN_PRICE

HIGH_PRICE

LOW_PRICE

CLOSE_PRICE

SHARE_VOLUME

ADJUSTED_CLOSE

TICKER_SYMBOL

State

City

Out of the above mentioned columns, the 'State' and 'City' are the two which we have added explicitly to the airline_stocks file so as to come up with some interesting geographical visualizations.

3. The Columns in the jet_fuel_prices file are as below:

SPOT_PRICE

TRADE_DATE

TRADE_JDN

The SPOT_PRICE is the most important variable in the jet_fuel_prices file which will be used an a measure in demonstrating the dimensions from the airline_stocks csv. We have taken the left outer join between airline_stocks and jet_fuel_prices joining them on TRADE_DATE variable.
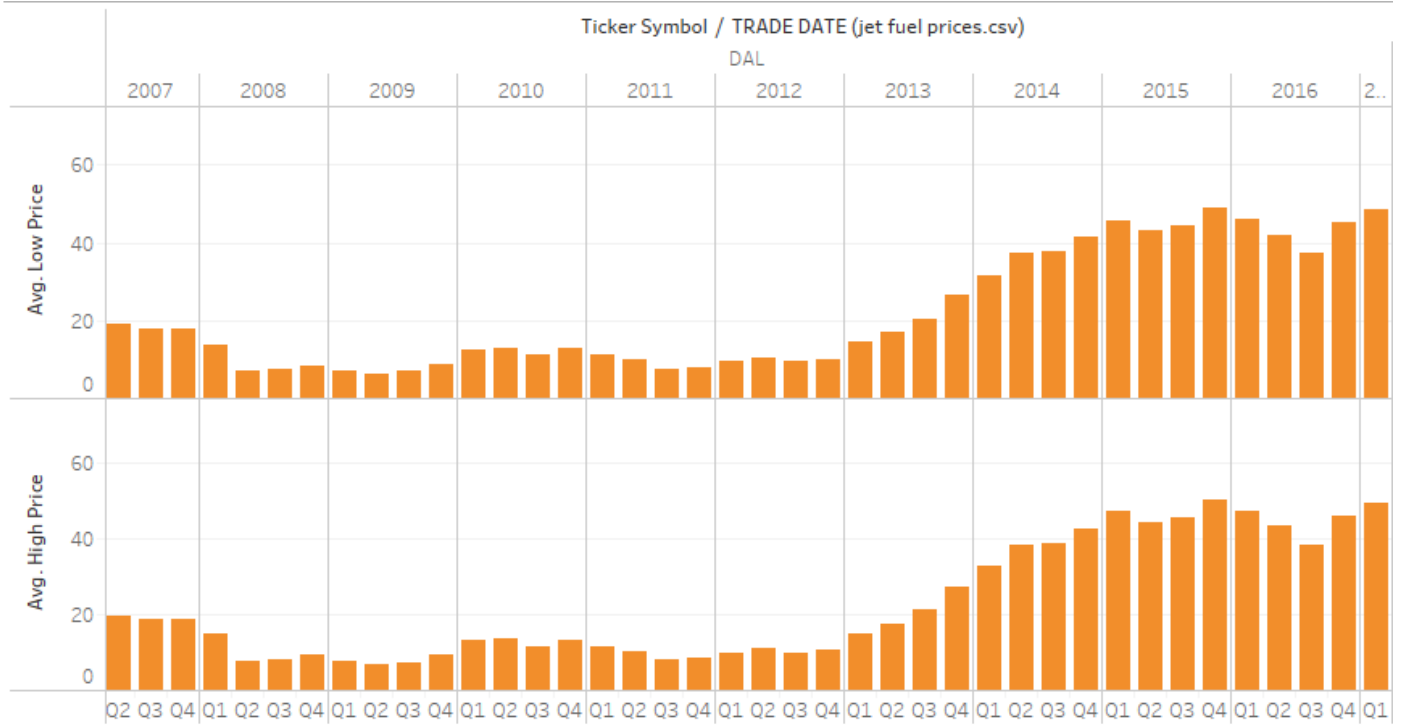
Let us take a look at some interesting visualizations to start the narrative with:

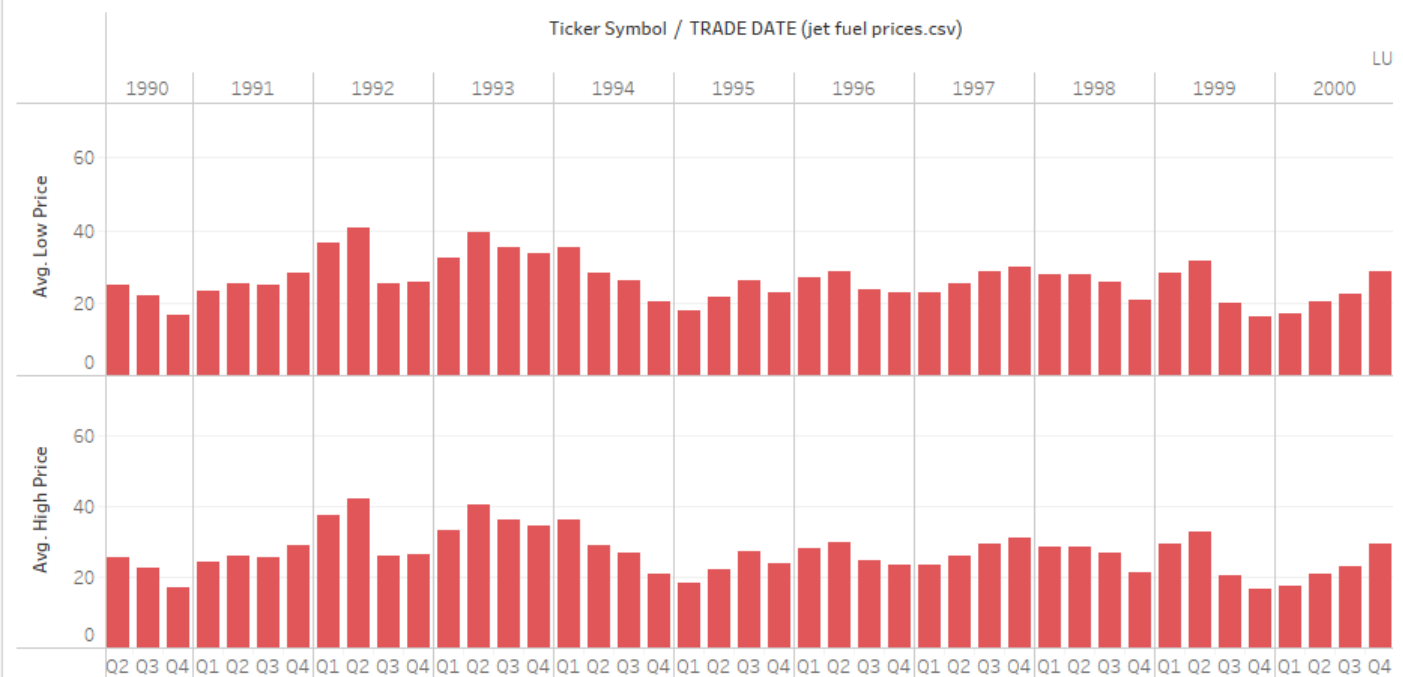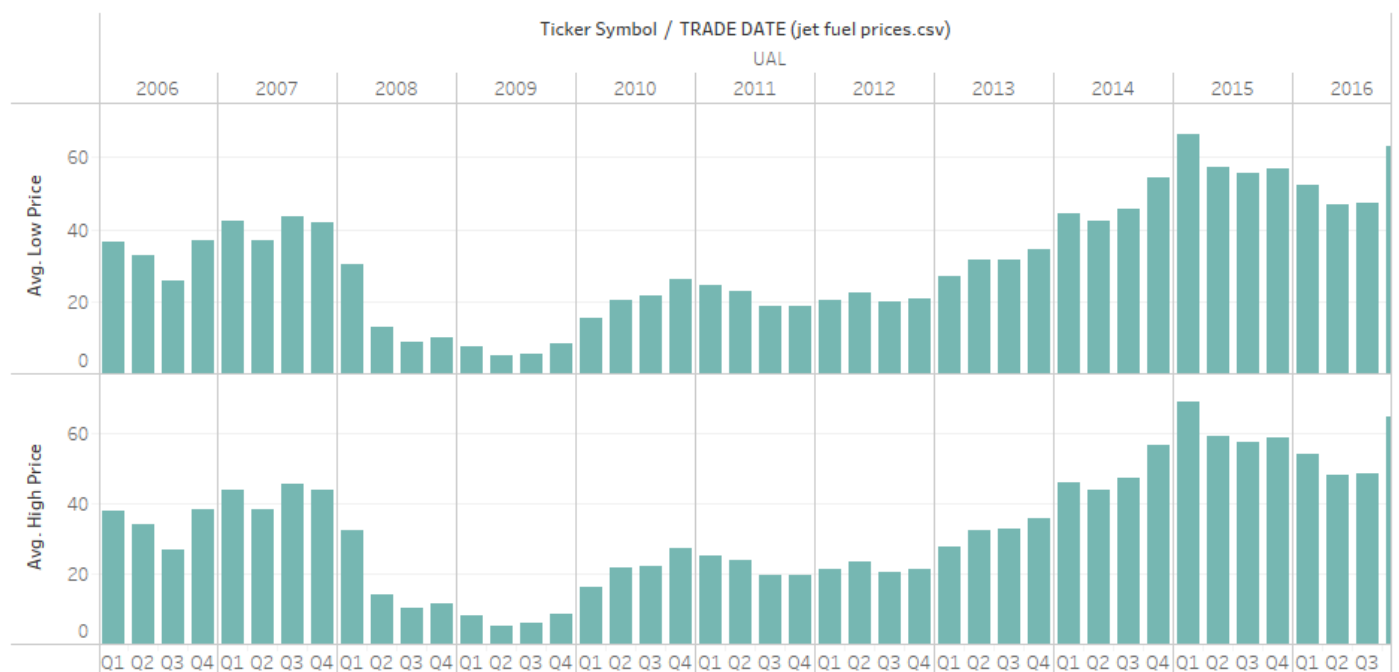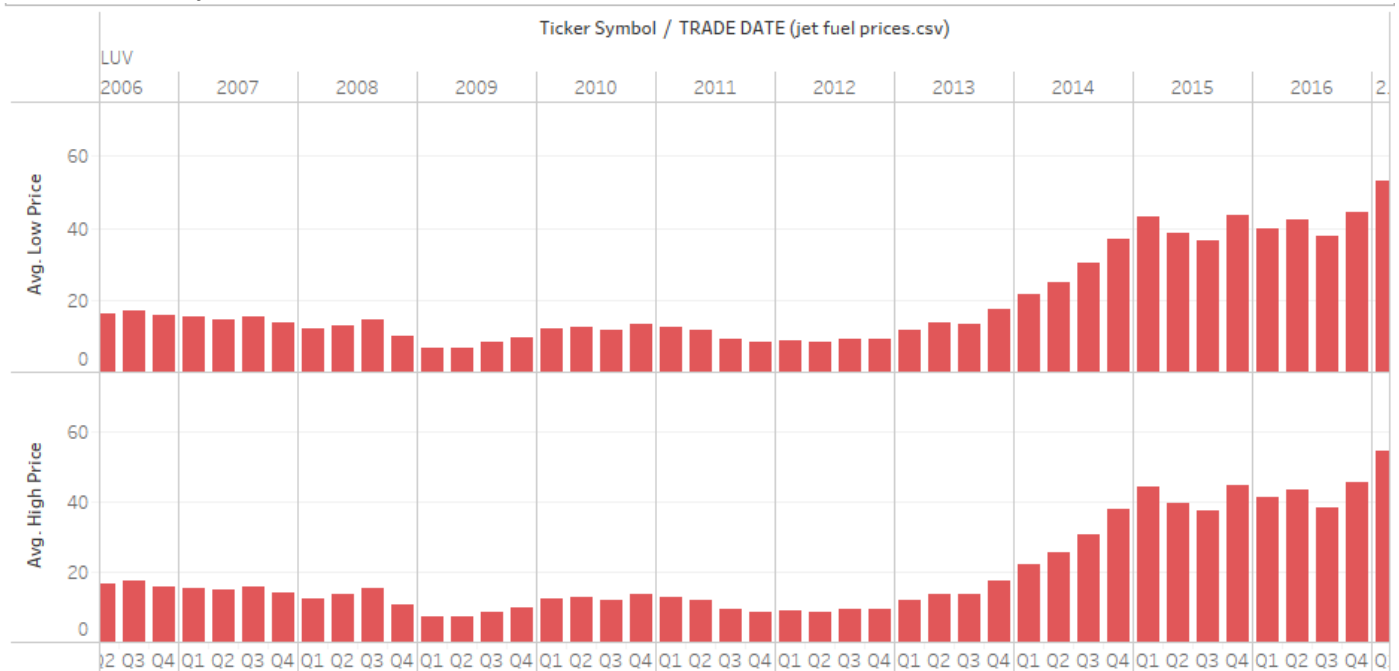1. **The distribution of Low Price and High price over the 4 quarters in each year**:

Ticker Symbol / TRADE DATE (jet fuel prices.csv)

DAL

## Sheet 1

**Ticker Symbol / TRADE DATE (jet fuel prices.csv)**

LU
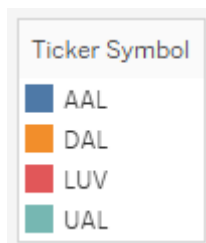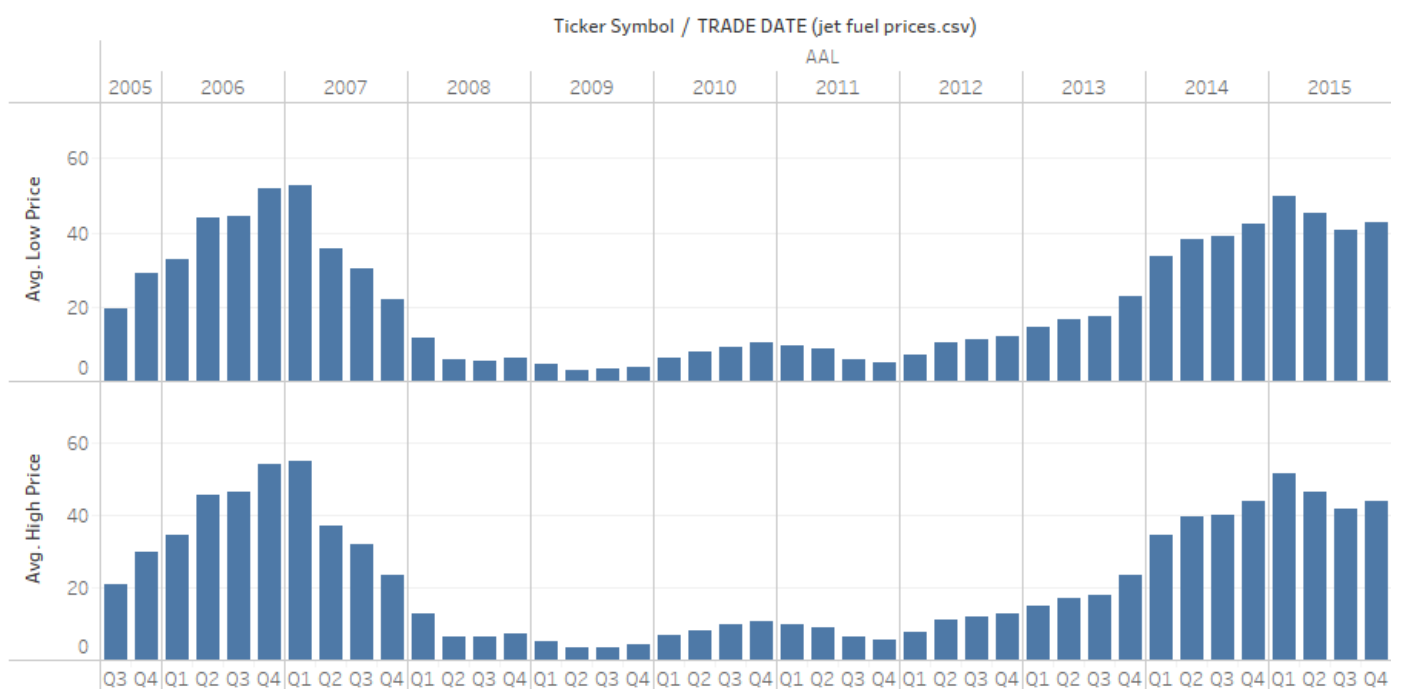
The Bar charts for the four airlines clearly represent that the Average high and average low values of stocks over each quarter significantly dropped from the Q2 of year 2007 and they remained very low till the Q4 of year 2012. This is the result of the Financial Crisis of 2007-08 which is known as The Great Recession. The stock high and low prices were decreased due to reluctance in the stock market however the same prices seem to start gradually increasing and taking grip in the Q1 of 2013. This applies to all the 4 airline stocks which defines the exactly same pattern for all of them.

**2.   Distribution of Average Spot Price of Jet Fuel over Trade Years:**
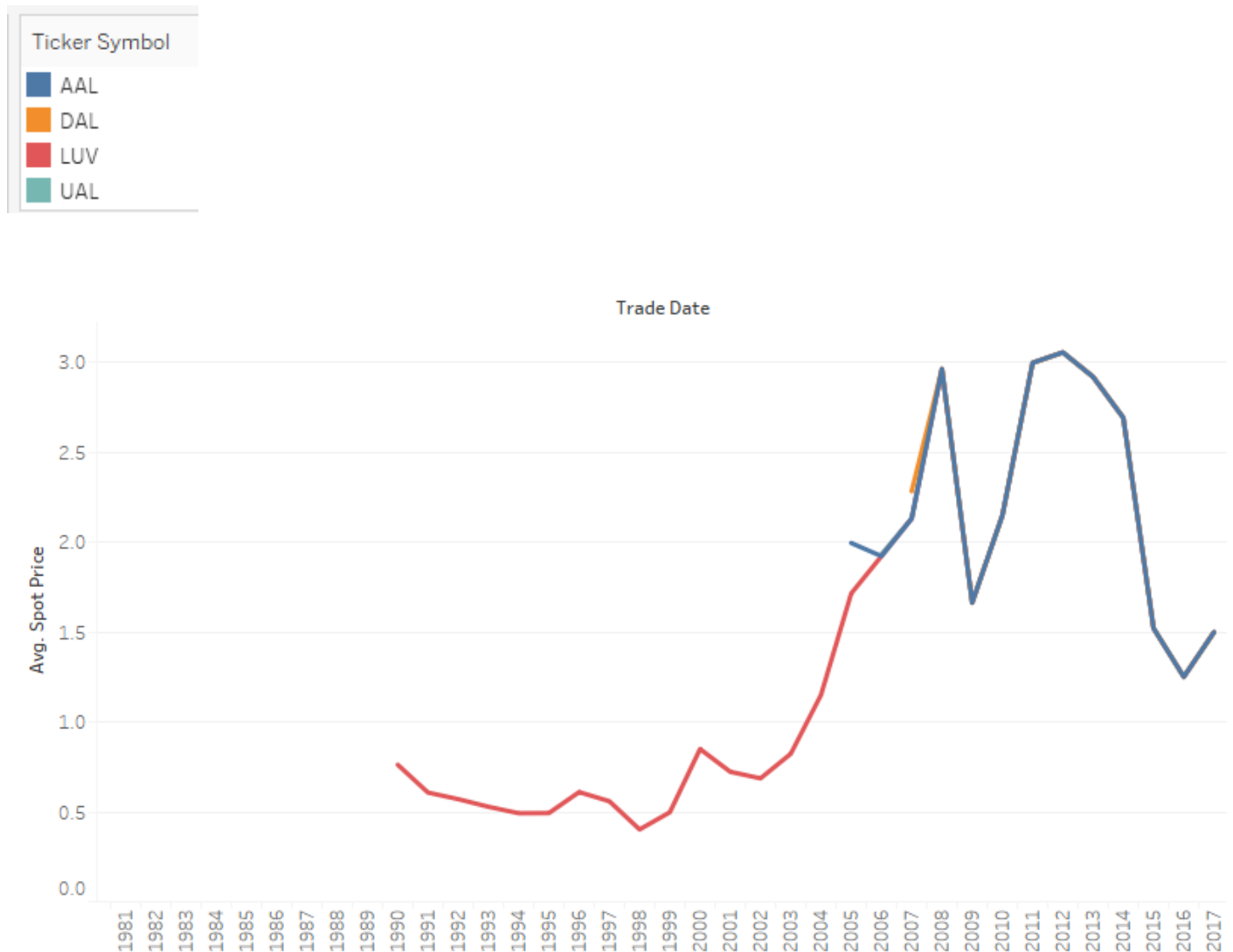
From the above visualization, we conclude that there is no fixed pattern for the distribution of stock price of Southwest airline over the years' range 1990-2003. After 2003, the stocks seem to be soaring up. On the same note, Delta Airlines have shown a sudden rise in the stock between 2007-2008 before the recession period. The American Airlines stocks have very logical behaviour with respect to jet fuel prices i.e. they have seen tremendous growth within window 2006-08 and then drop within the window 2008-2011 which falls within the great recession. After 2012, it again shown the inverse relation with avg. spot prices.

**3. Distribution of Average Adjusted Closing Price of Stocks over Trade Years:**
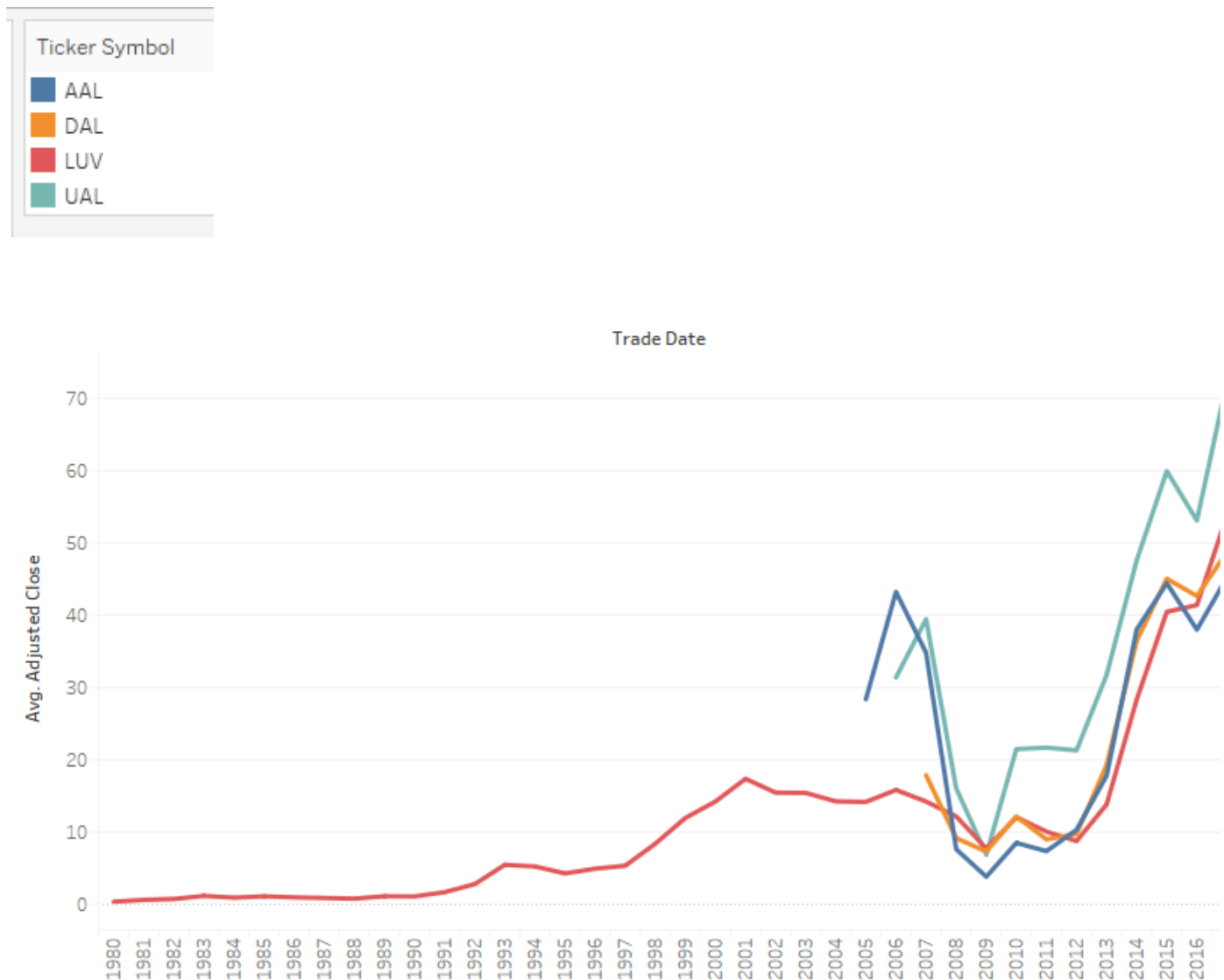
The Southwest airline has shown a pretty steady growth pattern with a negligible depreciation during 2006-2012. The remaining three airline stocks have a very expected behaviour during the great recession followed by a exponential growth in adjusted close prices.

4. **The distribution of Average Share volume of Airline Stocks over the 4 quarters in three years (2015,16,17)**:
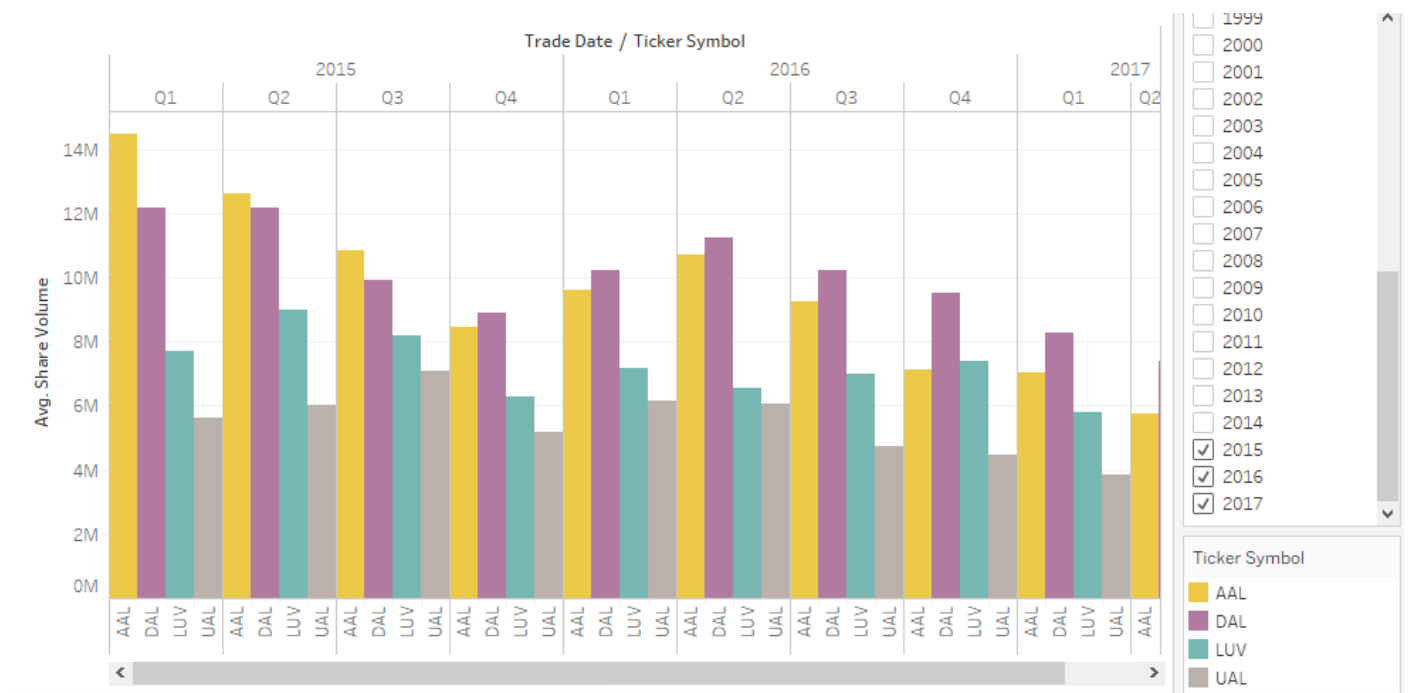
The distribution pattern of Average Share volume of Airline Stocks over the 4 quarters in three years (2015,16,17) are quite identical with only one exception. In first three quarters of 2015, the share volume of American Airlines traded used to be consistently higher than Delta Airlines. From the Q4 onwards, Delta Airlines stock were traded in a greater quantity than American Airline. The United stocks were traded in the least quantity followed by Southwest Airlines. This concludes that AAL and DAL have been ahead in terms of their business which, in turn, results in more stock market trading in terms of no. of shares.

5. **The distribution of Average Share volume of major airline stocks over a Recession period:**

Major US Carriers Shares Performance - Recession Years
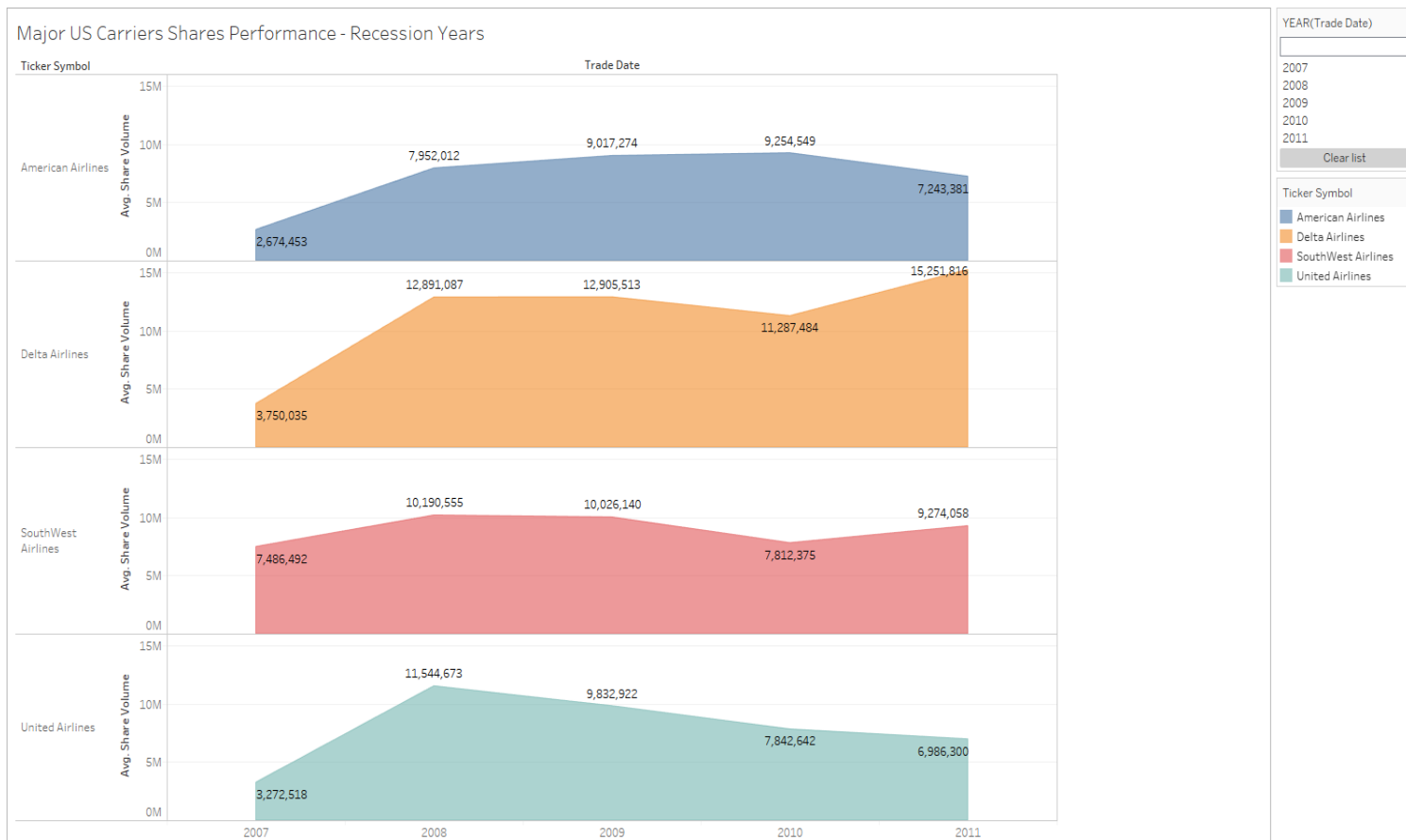
The visualization above displays the volume of shares traded over the years 2007 to 2011. We are particularly interested in this time since the global economy experienced a recession and stock prices dipped low. If we examine the performance of carriers we notice that American Airlines faced difficulty in the year 20007 as compared to other carriers. Their stock volume was the least at 2,674,453 and Southwest Airlines did well during that time. If we consider Southwest's strategy, we find that they operate in smaller towns and cities domestically in the US where other major carriers do not have a hold over the market which has helped them to be the preferred choice of customers since there were no better alternatives. The assumption in this pattern is that the revenue of these carriers is directly proportional to the volume of the stocks traded.

Also, another interesting discovery is that Delta was the fastest to revive from the tough period. Despite being at a low of 3,750,035 it could climb up to over 15,000,000 a remarkable recovery story.

6. <u>The variation of average adjusted close price and average spot price over the years:</u>

The above visualization helps us understand the trends followed by adjusted close price and compare that to the spot price over the years. If we take a close look we observe that there is an inverse relation between Average Adjusted Close price of stocks and Average Spot price from the year 2012 to 2016. A similar trend is observed in the years 2006 to 2007. As the adjusted price dropped the spot price increased. However, the years 2008 to 2011 show a spike in the spot price and a very steep drop in the adjusted price, so much so that the plots appear to be volatile. The affect could be a by-product of the volatility that the market experienced that time. We can see that the scatterplots help us to pinpoint the exact values and drops in prices and determine the average adjusted change in close price.

7. **Distribution of Avg. high price in accordance to Year trade date with quick table calculation (year over year growth + tooltip) in Tableau**:
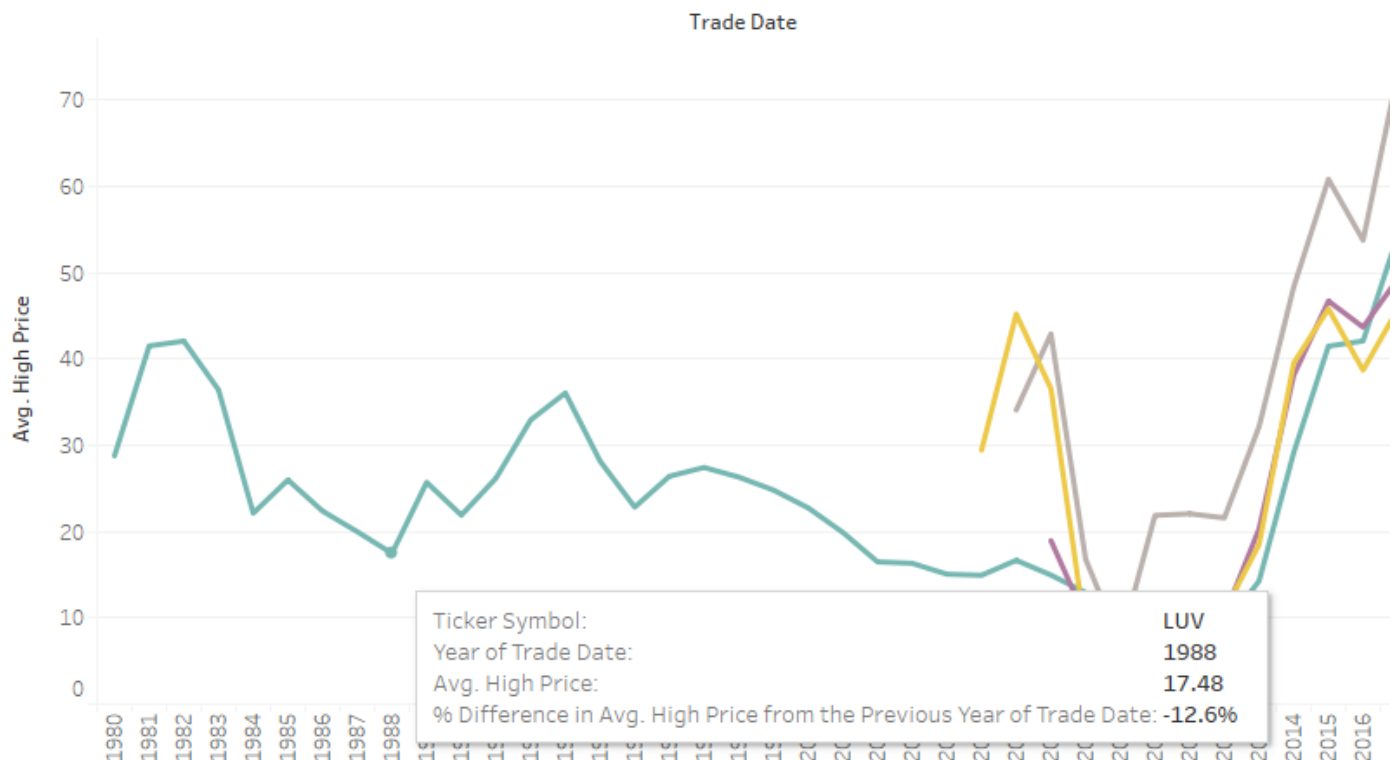
We have used a tooltip representation option to show the % difference in Avg. High Price of stocks from the previous year of trade date.

8. **A split up of Volume of shares by state for the entire time period based upon volume buckets:**
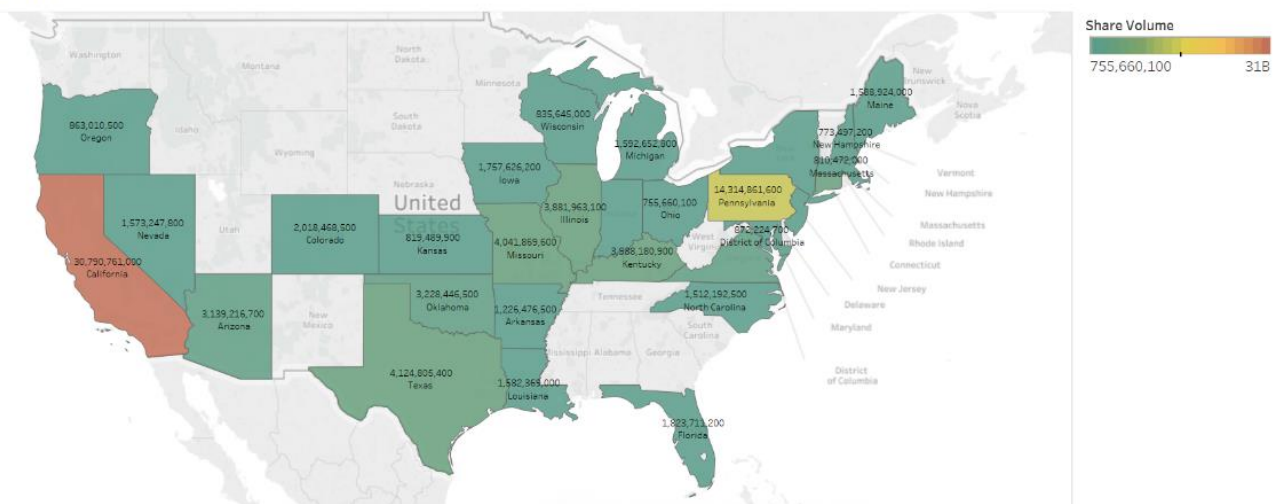
With California amounting for the highest - 31 Billion. and Ohio the least with 755,660,100. An improvisation would be filtering this visualization over a specified time period, say last 1 year. This would bring out a better picture as to which state has higher share volume density.

8 (A). **Pie Charts for each Airline Company state wise:**

American Airlines:
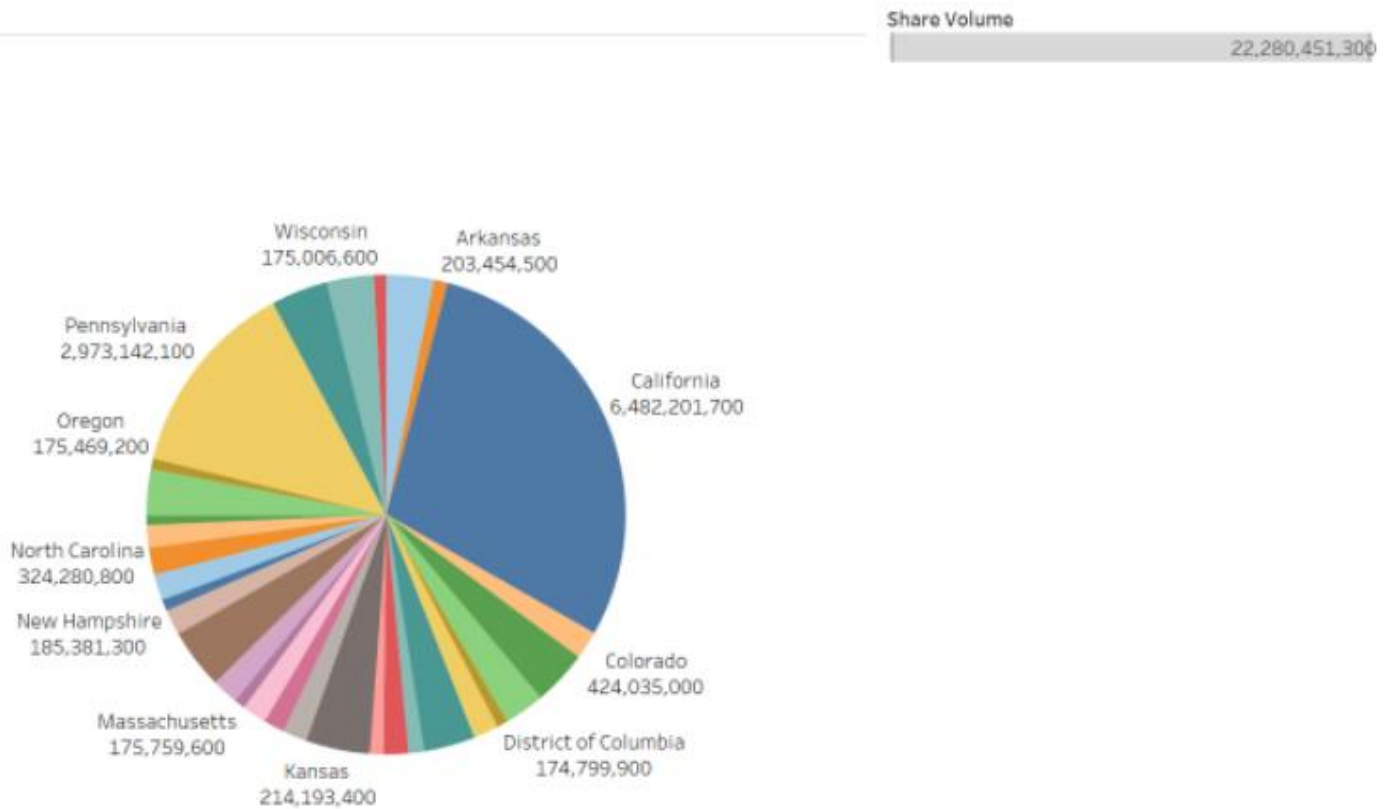
Delta Airlines:

Share Volume
29,290,414,400
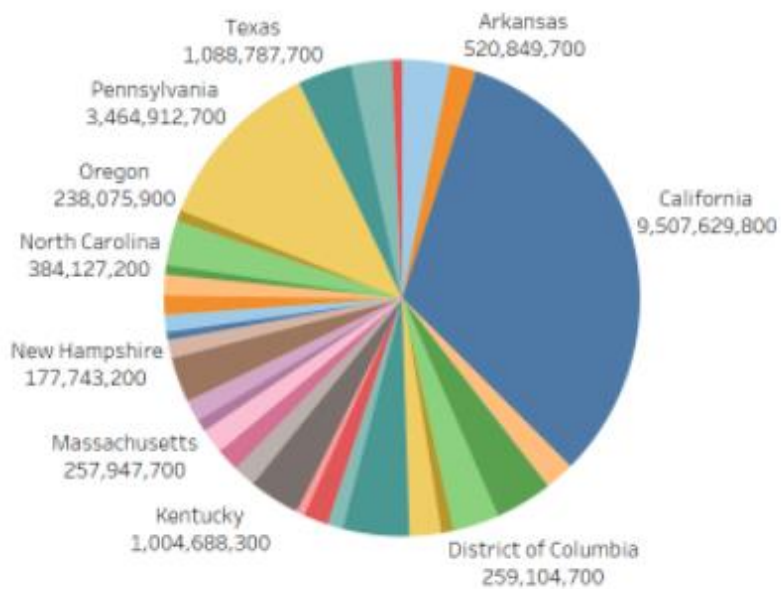
Southwest Airlines:

Share Volume

34,570,545,700

Virginia
1,209,211,300

Arkansas
342,682,000

Pennsylvania
5,178,425,700

California
9,872,797,200

Oregon
305,718,200

North Carolina
557,768,800

New Hampshire
283,781,000

Colorado
677,782,800

Massachusetts
255,926,700

District of Columbia
274,924,300

Kansas
279,625,000

United Airlines: Page **15** of
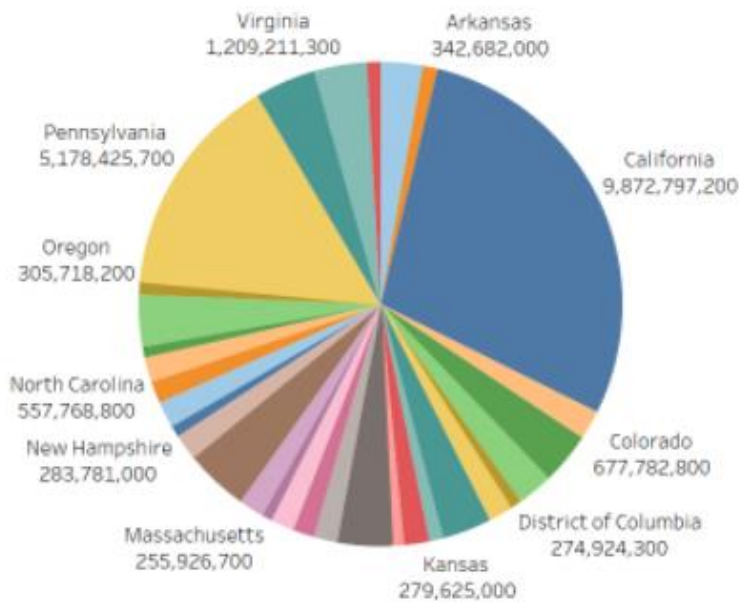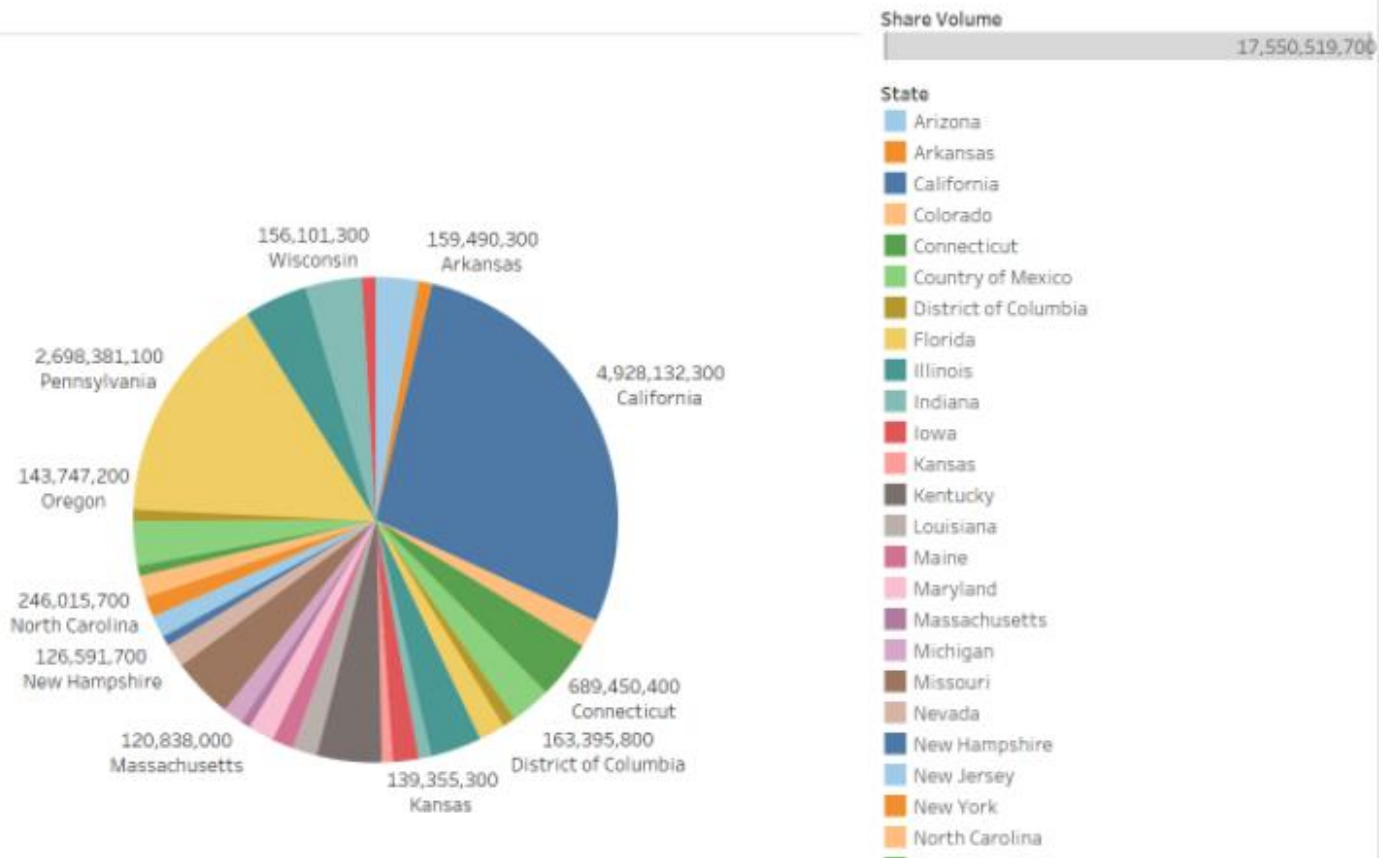
8 (B). **Distribution of share volume over different states during recession & last five years**:

Share of Volume - Recession
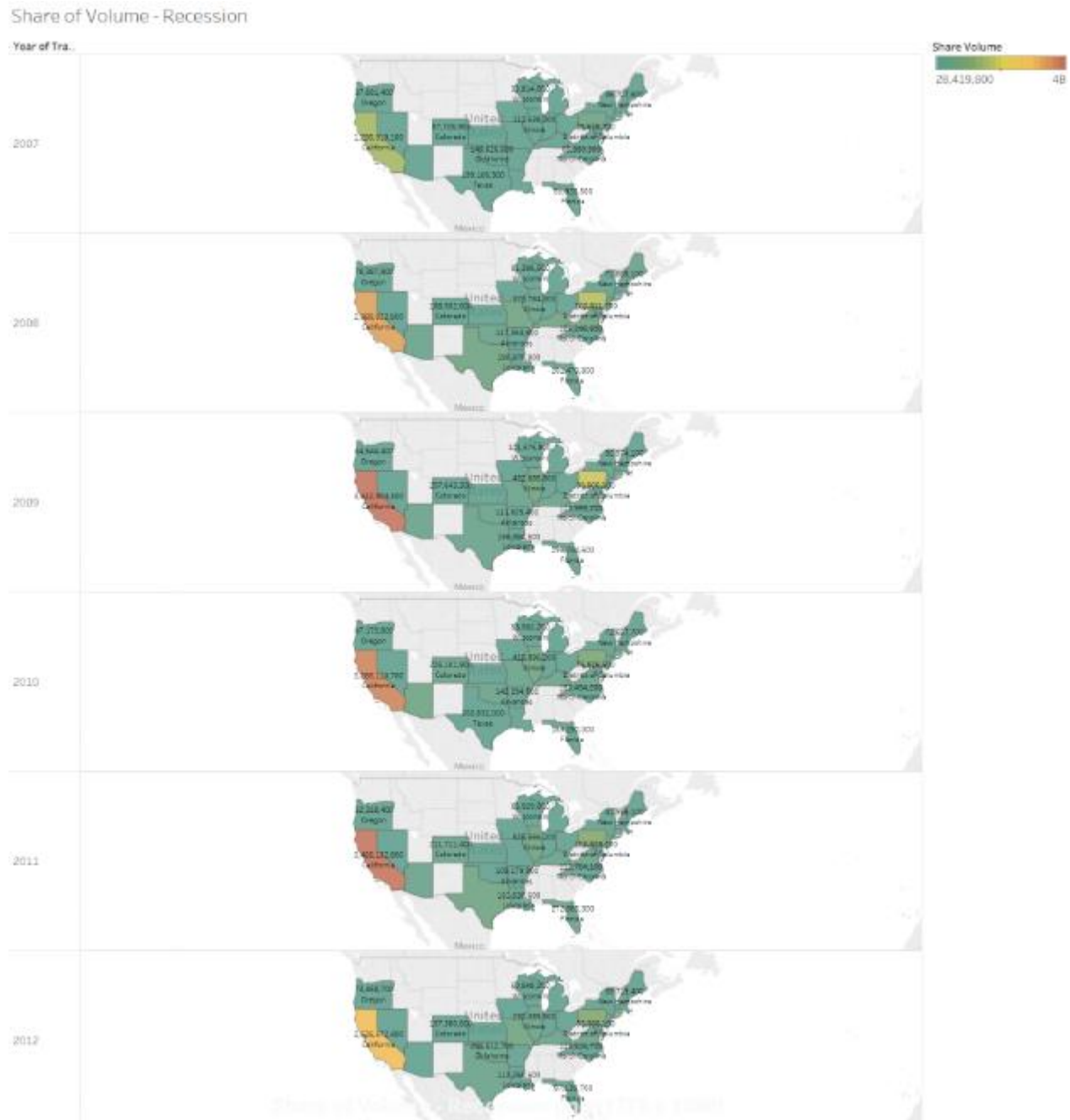
Share of Volume - Last 5 years

9. **Distribution of Share Volume for State of California**:
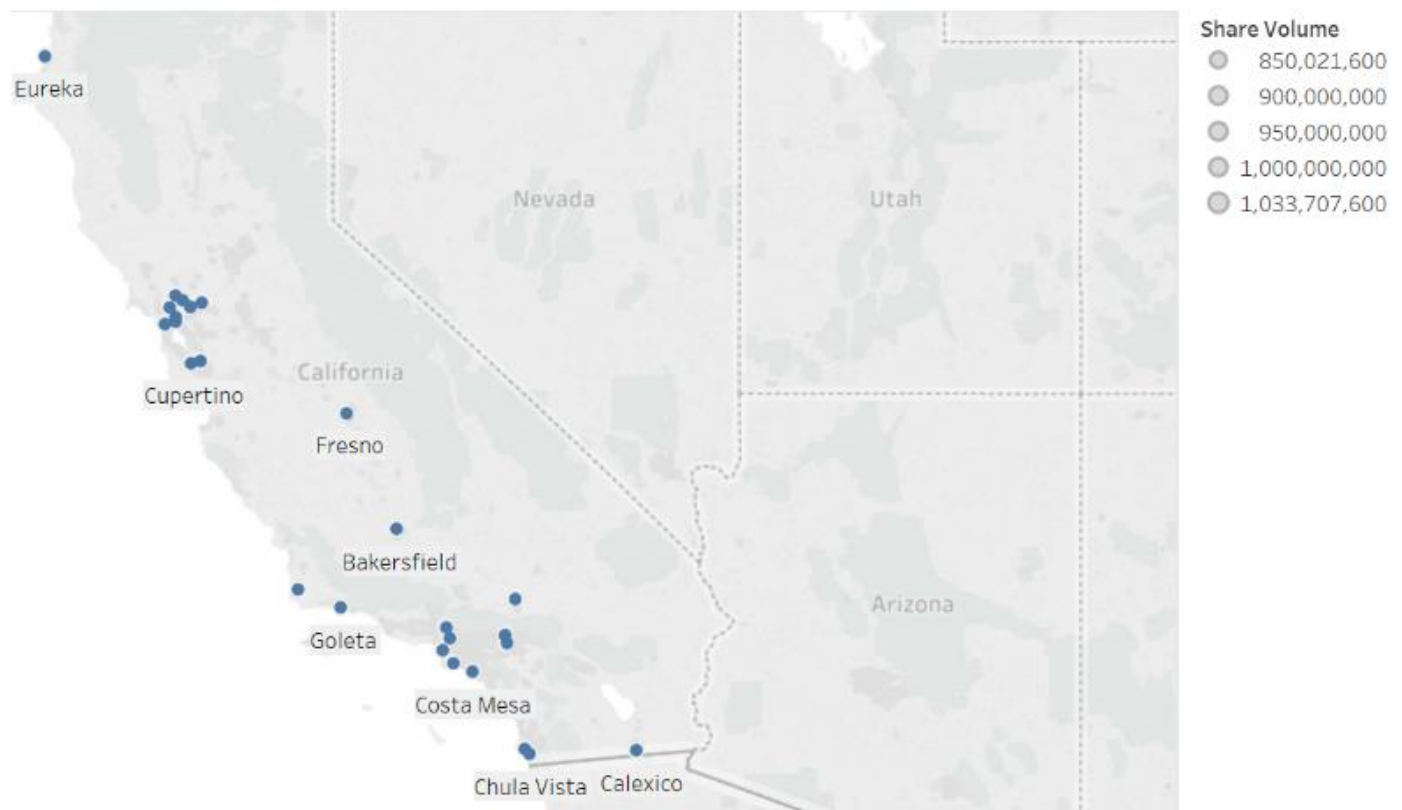
An earlier visualization brought back highest volume of hares for state of California. Here we notch up our analysis one granular scale higher by measuring the cities that add up to the such volume. This would benefit potential investor to choose which city comes back with higher trading activity and an impetus to closely watch/invest in companies from that city.

10. **Distribution of shares' volume in cities of California for 5 recession and 5 latest years + single average point with 95% Confidence Interval:**

In this case, we have considered distribution of share volume of airline stocks across all cities in the California state. Also, we have modelled the Average distribution of these volumes over the Average value assuming 95% Confidence Interval.

Figure (A)

In the diagram (A), we are observing the distribution over the period of 2007-2012 when the United States' Airline businesses and concerned Stock market was heavily hit by the economic breakdown. The average value of available data with 95% Confidence Interval is achieved to be ~296 M. The cities in CA where the collective share volume of 4 airline companies was shown up as greater than average value (within 2008, 09, 11) are predominantly Arden-Arcade, Bakersfield, Benicia, Berkeley, Bethel Island and Burbank.

Sheet 1
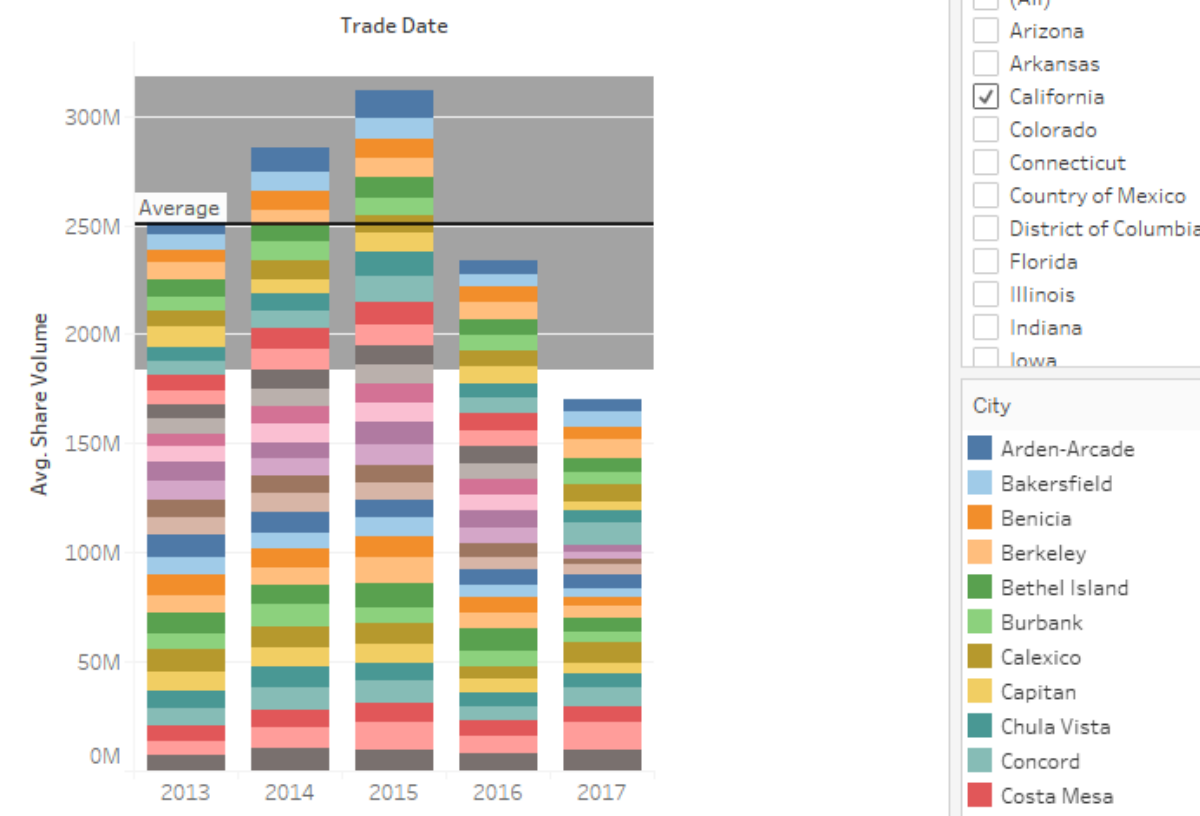
Figure (B)

When we visualize the exactly same scenario for years 2013-2017, we find that average value of available data is ~250M. In 2014, the cities in CA where the collective share volume of 4 airline companies was shown up as greater than average value are predominantly Arden-Arcade, Bakersfield, Benicia and Berkeley. In 2015, the cities in CA where the collective share volume of 4 airline companies was shown up as greater than average value are Arden-Arcade, Bakersfield, Benicia, Berkeley, Bethel Island and Burbank. Otherwise, for rest of the latest years, the collective share volume of 4 airline companies is seemed to be consistently below the average share volume value.

During the drastic influences of economic breakdown, the trading in the cities in CA is observed to be surged in terms of number of shares, not in terms of high and adjusted close values of stocks. On the contrary, it seems to be dropped down in the recent years till 2017. Thus, the inverse proportion between the trading years and average share volume is concluded.

11. **High Price Vs Year Vs State:**

Year of TRADE DATE (jet fuel prices.csv)
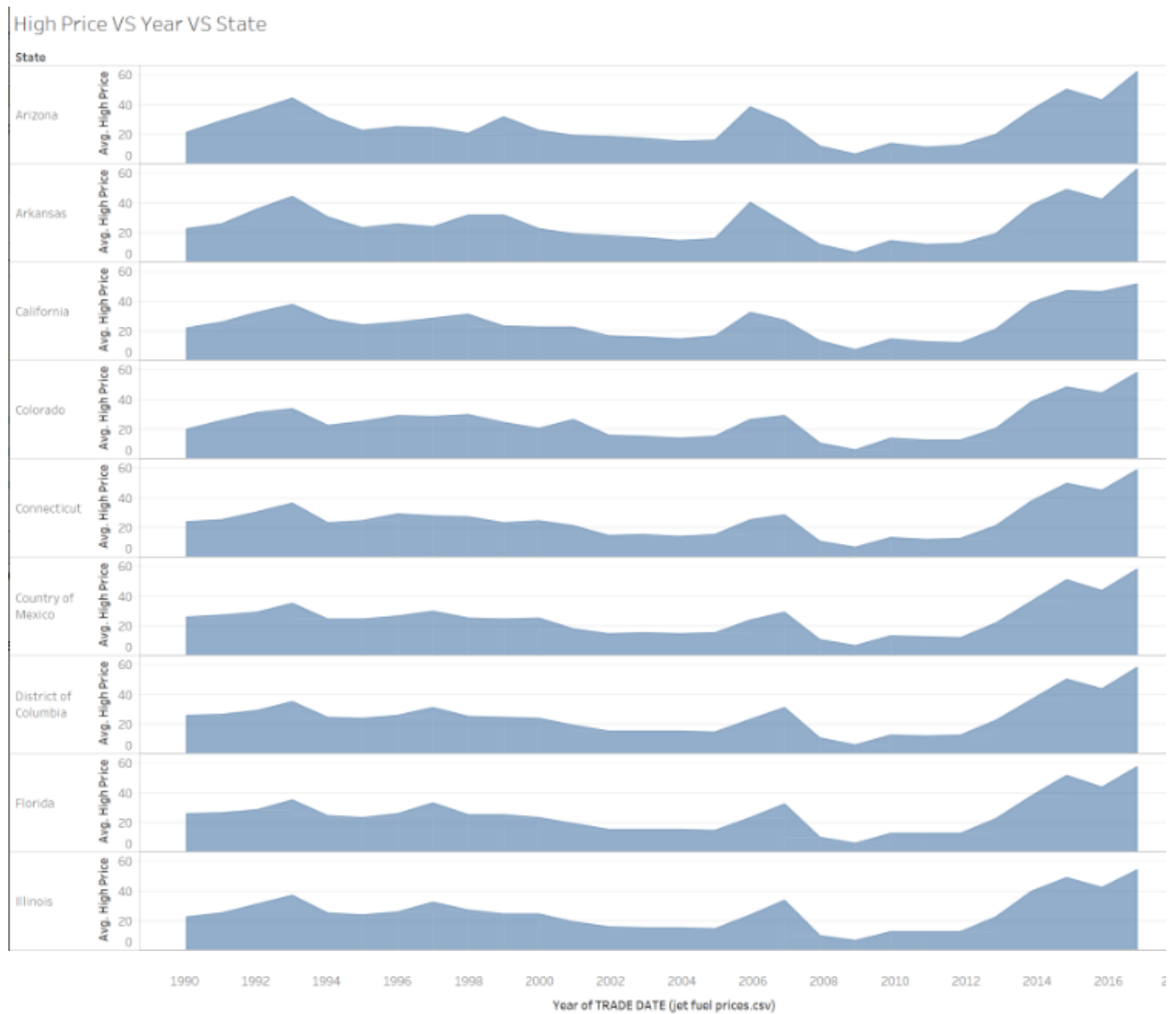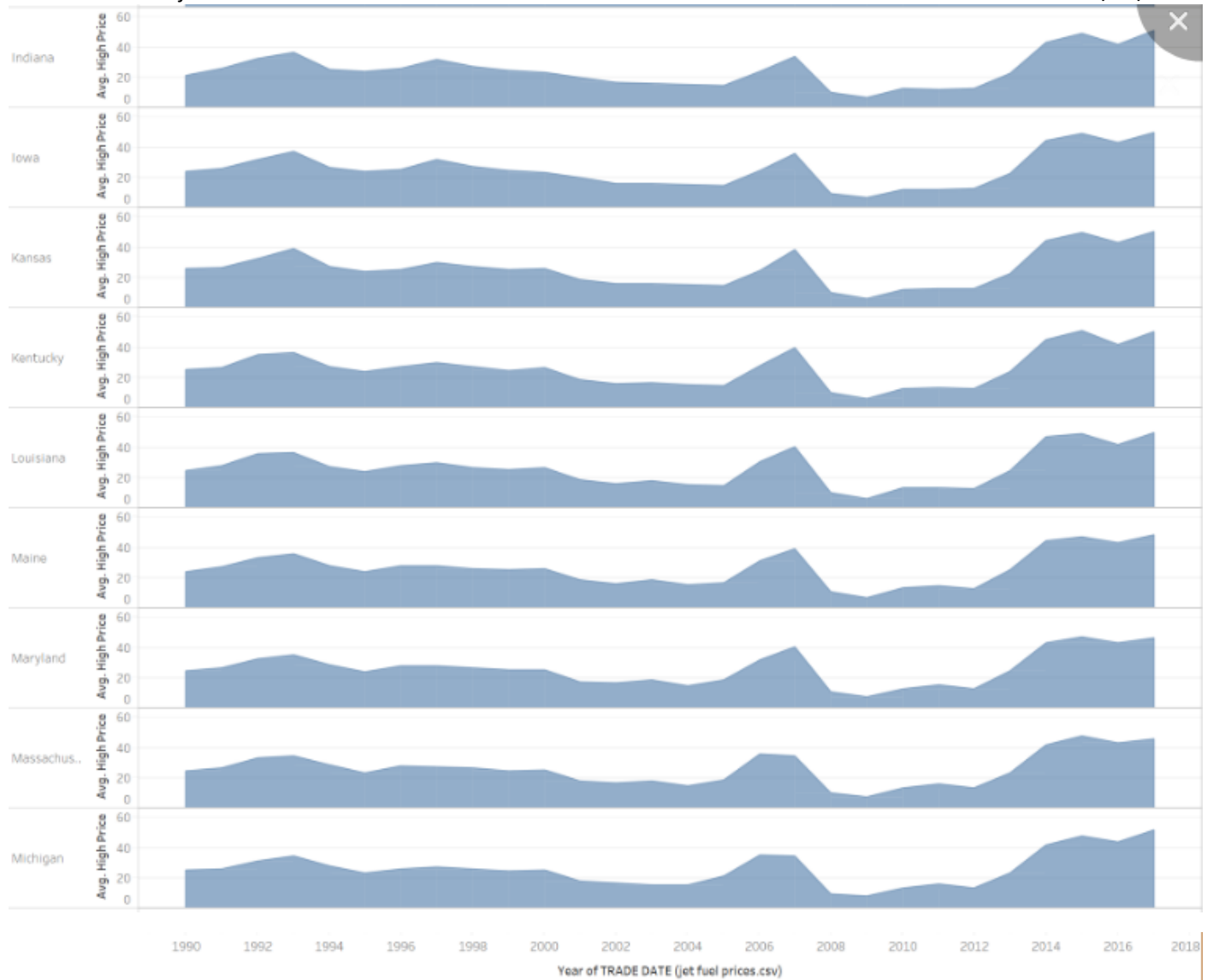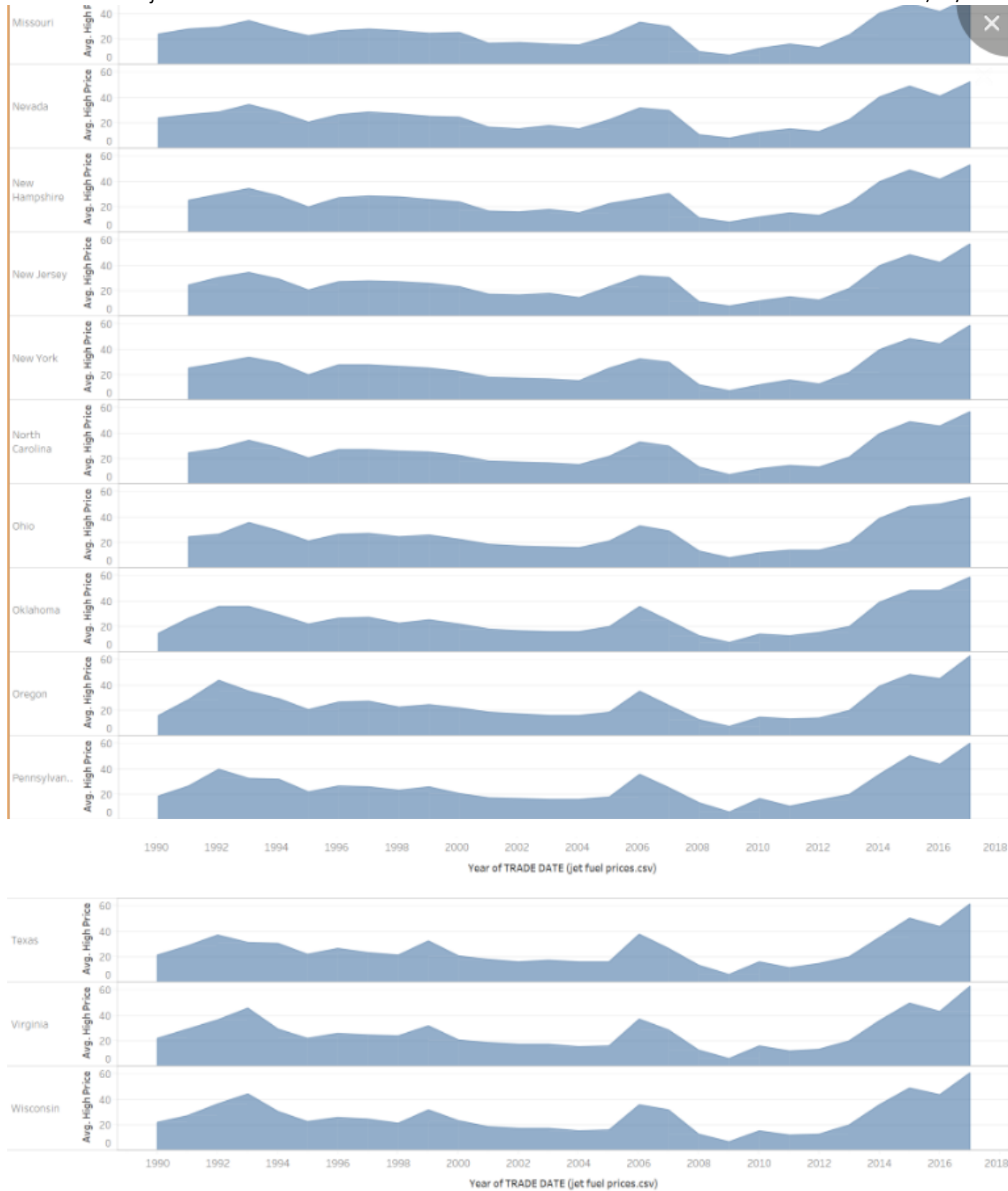
A trend map marking the avg. high price/ state through the years. To naked eye there are very few anomalies. On a closer look, we can see how few states have survived through market crashes and recession during the year 2008-2009 with tiny bumps in comparison to other states. Certain states have managed to carry on and not be too affected by such crashes in comparison.

12. **Avg Open Price, Avg Close Price Vs State**:

Avg Open Price VS Avg Close Price
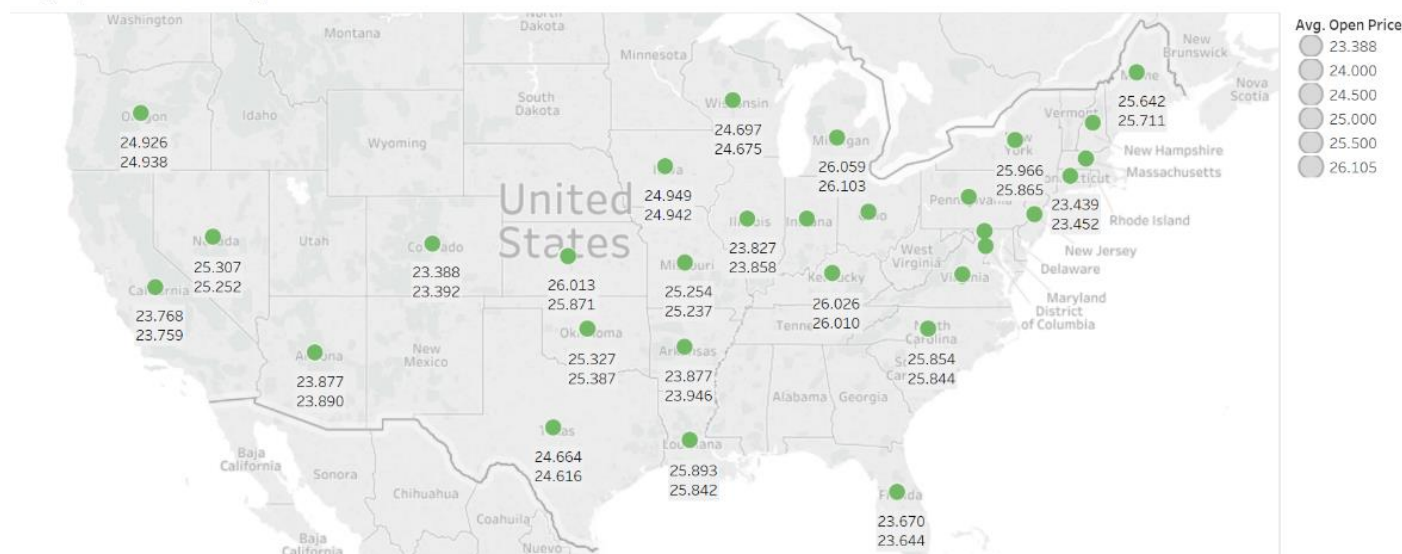
It's interesting to see how few states have a higher avg. opening price in comparison to their avg. closing price. That shows how markets have started out strongly but have managed to dip through the day. State of Florida could be one such example. On the contrary states of Illinois, Arkansas have done better through the day and closed stronger each time.

13. **Distribution of the difference between the Significant Factor (close price - adjusted close price) of stocks**:

Here, we have considered the difference between the close price and adjusted close price each day as Significant factor. Generally, the closing price of a stock is exactly that: the price of that stock at the close of the trading day. The adjusted closing price uses the closing price as a starting point, but it takes into account factors such as dividends, stock splits and new stock offerings. This will help us derive some conclusions regarding latest stock affairs in market. For this, we added one columns in stocks file which calculates the difference between above mentioned entities.
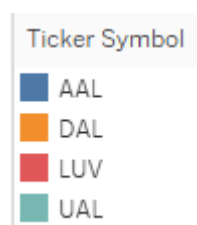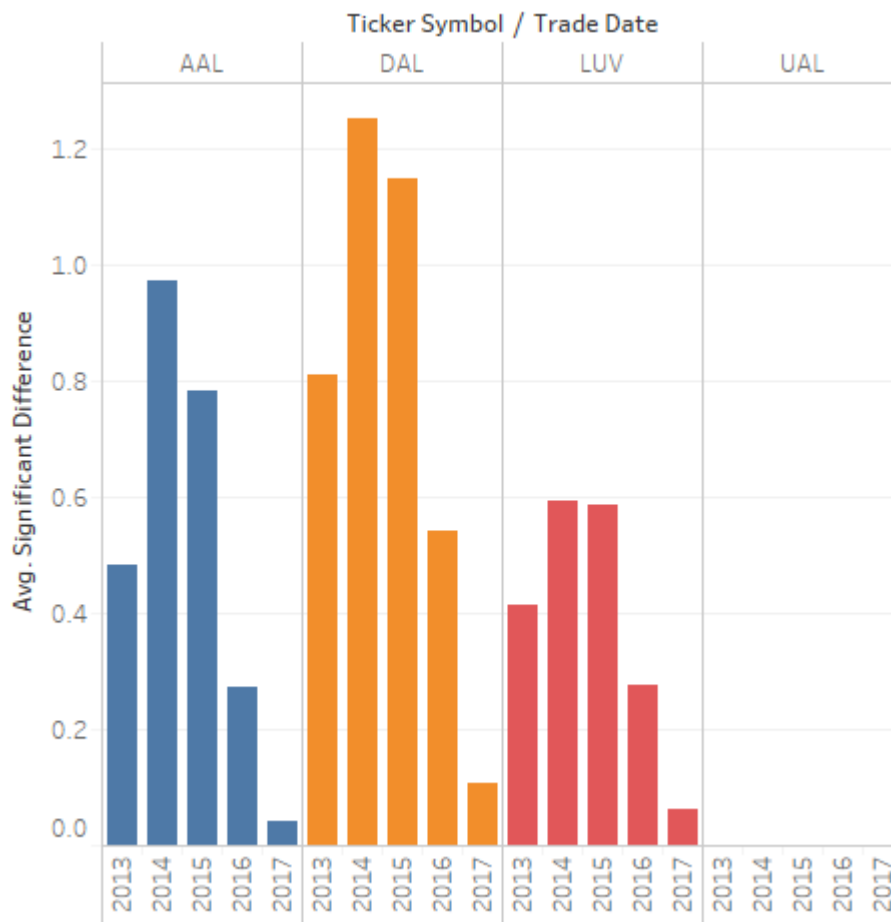
Fig. (A)

In Fig. (A), the distribution of average significant factor for 4 airline companies' stocks over the period of latest 5 years (2013-2017) is modelled. For American, Southwest and Delta airlines, the average significant factor i.e. difference between the close price and adjusted close price, is maximum over 2014 and minimum over 2017. This simply represents a more accurate reflection of a stock's value, since distributions and new offerings have altered the closing price.

For United Airlines, the Significant factor has consistently been 0 i.e. either none of the market activities like dividends, stock splits and new stock offerings have influenced the adjusted close price or such market activities did not occur.
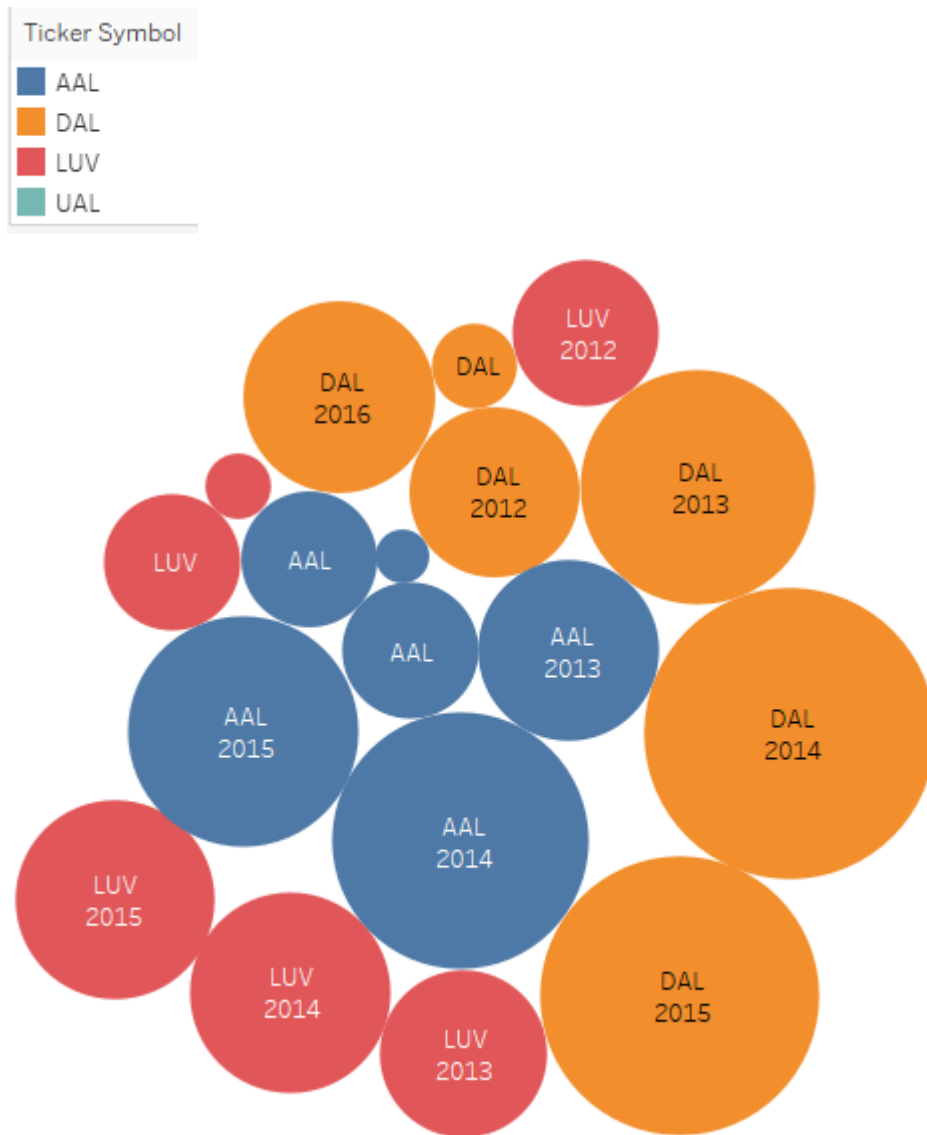
Fig. (B)

Fig. (B), size of **PACKED BUBBLES** is directly proportional to the value of average significant factor over the respective year. Delta airlines had maximum average significant factor (1.252) over 2014 while American Airlines had the least (0.043).

Analytical SQL Queries:

1. Retrieve a simple Pearson correlation coefficient between Adjusted Close Price and Spot Price

```sql
SELECT
(Avg(a.ADJUSTED_CLOSE * b.SPOT_PRICE) - Avg(a.ADJUSTED_CLOSE) * Avg(b.SPOT_PRICE)) /
(StDevP(a.ADJUSTED_CLOSE) * StDevP(b.SPOT_PRICE)) AS Correlation
FROM
dbo.airline_stocks a JOIN
dbo.jet_fuel_prices b ON a.TRADE_DATE = b.TRADE_DATE
```
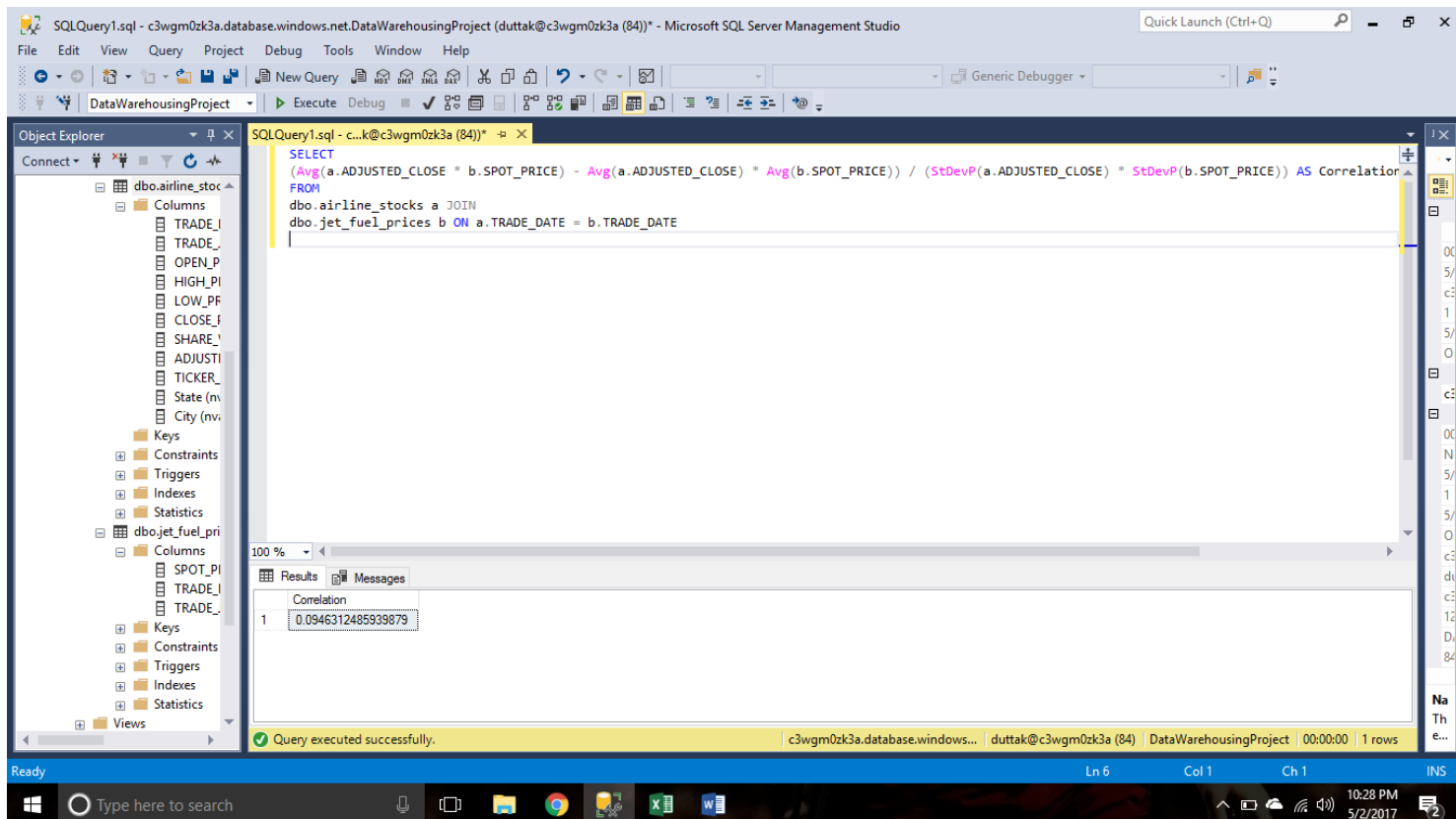

2. Compute the N-point Moving Average of Volume of shares using the last 7 trade days and display the result sorted by most recent days

```sql
SELECT
 a.TRADE_DATE,
 YEAR(a.TRADE_DATE) DT_YEAR,
 SUM(a.SHARE_VOLUME) TOTAL_VOLUME_SHARE,
ROUND(AVG(a.SHARE_VOLUME) OVER (ORDER BY TRADE_DATE ASC),2) MOVING_AVG_VOLUME_SHARE,
ROUND(AVG(a.SHARE_VOLUME) OVER (ORDER BY TRADE_DATE ASC ROWS BETWEEN 6
PRECEDING AND CURRENT ROW),2) MOVING_AVG_VOLUME_SHARE_LAST7DAYS
FROM
dbo.AIRLINE_STOCKS a
GROUP BY a.TRADE_DATE,
YEAR(a.TRADE_DATE),
a.SHARE_VOLUME
ORDER BY a.TRADE_DATE DESC
```

3. Calculate the overall average value of closing stock and spot price aggregated on a hierarchical fashion (using ROLLUP)

```sql
SELECT
d.TRADE_DATE,
ROUND(AVG(d.CLOSE_PRICE), 2) AS AVG_CLOSE_PRICE,
ROUND(AVG(j.SPOT_PRICE), 2) AS AVG_SPOT_PRICE
FROM
DBO.AIRLINE_STOCKS d
JOIN DBO.JET_FUEL_PRICES j ON d.TRADE_DATE=j.TRADE_DATE
WHERE d.TRADE_DATE IS NOT NULL
GROUP BY
ROLLUP (d.TRADE_DATE, j.TRADE_DATE)
ORDER BY d.TRADE_DATE DESC
```

4.A) Get the Max and Min of Volume of shares for each and every trade day; Make use of this window when you bin the volume of shares in the next query

```sql
SELECT a.TRADE_DATE,
min(a.SHARE_VOLUME) MIN_VOL,
max(a.SHARE_VOLUME) MAX_VOL
FROM
dbo.airline_stocks a
group by a.TRADE_DATE
```

Conclusion:

At the completion of this Visualization project, we could learn how to prepare and understand the data to tell the story and prepare the corresponding visualizations. We used the Tableau Public Desktop version 10.2 to draw meaningful trends, bar charts etc. to support the story based upon important dimensions and measures of the business data. We also learnt to manipulate the measures to make sense to the story instance along with preparing the quick table calculations to highlight some of the fiscal patterns. We added the geographic parameters (state and city) to our csv file to extend our project to geographical visualizations. This helped us elaborate on our story with the representations of share volumes and average prices spread across different state and specific cities within those states. The work Analytical SQL Queries represent moving averages, bins, correlations and aggregation within the important variables of dataset.