**Group 16**

**Data Visualization Case Study 2**

Data Source

**Primary Dataset:** Startup Venture Funding - Information about startup companies, investments, and acquisitions via Crunchbase

Link: <https://public.tableau.com/s/resources>

Once you are on the above provided web URL, locate the *Startup Venture Funding* Sample Data Set under the *Technology* subheading.

The dataset was extracted from Crunchbase in Dec’14 and contains multiple workbooks pertaining to the Startup companies, their investment rounds, acquisitions history and other details. In our analysis, we have made use of the Companies, Investments and Rounds workbooks and our concentration has been towards the US venture capital investments for Startups.

The following attributes have been used to bring more relevance to our analysis:

Investments -

|  |
| --- |
| company\_permalink |
| company\_name |
| company\_category\_list |
| company\_market |
| company\_country\_code |
| company\_state\_code |
| company\_region |
| company\_city |
| investor\_permalink |
| investor\_name |
| investor\_category\_list |
| investor\_market |
| investor\_country\_code |
| investor\_state\_code |
| investor\_region |
| investor\_city |
| funding\_round\_permalink |
| funding\_round\_type |
| funding\_round\_code |
| funded\_at |
| funded\_month |
| funded\_quarter |
| funded\_year |
| raised\_amount\_usd |

Rounds –

|  |
| --- |
| company\_permalink |
| company\_name |
| company\_category\_list |
| company\_market |
| company\_country\_code |
| company\_state\_code |
| company\_region |
| company\_city |
| funding\_round\_permalink |
| funding\_round\_type |
| funding\_round\_code |
| funded\_at |
| funded\_month |
| funded\_quarter |
| funded\_year |
| raised\_amount\_usd |

**Secondary Dataset:** US House Price Index

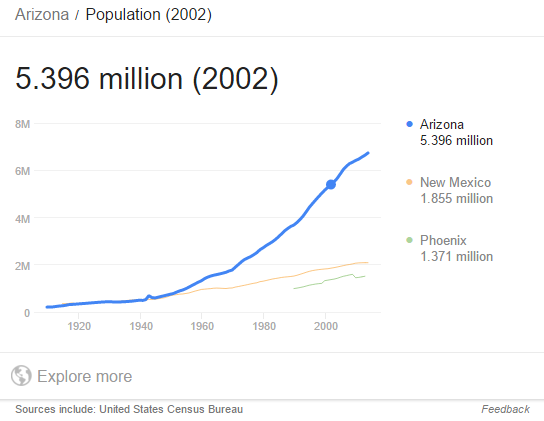
Link: <http://data.okfn.org/data/core/house-prices-us#data>

The data comes from Standard & Poors’ Case-Shiller data and includes both the National Home Price (single-family) Index and indices for 20 metropolitan regions. In our analysis, we have made use of the indices for the metropolitan regions but not the National and Composite Indices. This seeks to measure changes in the total value of all existing single-family housing stock.

The following attributes have been used to bring more relevance to our analysis:

|  |
| --- |
| Date |
| AZ-Phoenix |
| CA-Los Angeles |
| CA-San Diego |
| CA-San Francisco |
| CO-Denver |
| DC-Washington |
| FL-Miami |
| FL-Tampa |
| GA-Atlanta |
| IL-Chicago |
| MA-Boston |
| MI-Detroit |
| MN-Minneapolis |
| NC-Charlotte |
| NV-Las Vegas |
| NY-New York |
| OH-Cleveland |
| OR-Portland |
| TX-Dallas |
| WA-Seattle |
| ~~Composite-10~~ |
| ~~Composite-20~~ |
| ~~National-US~~ |

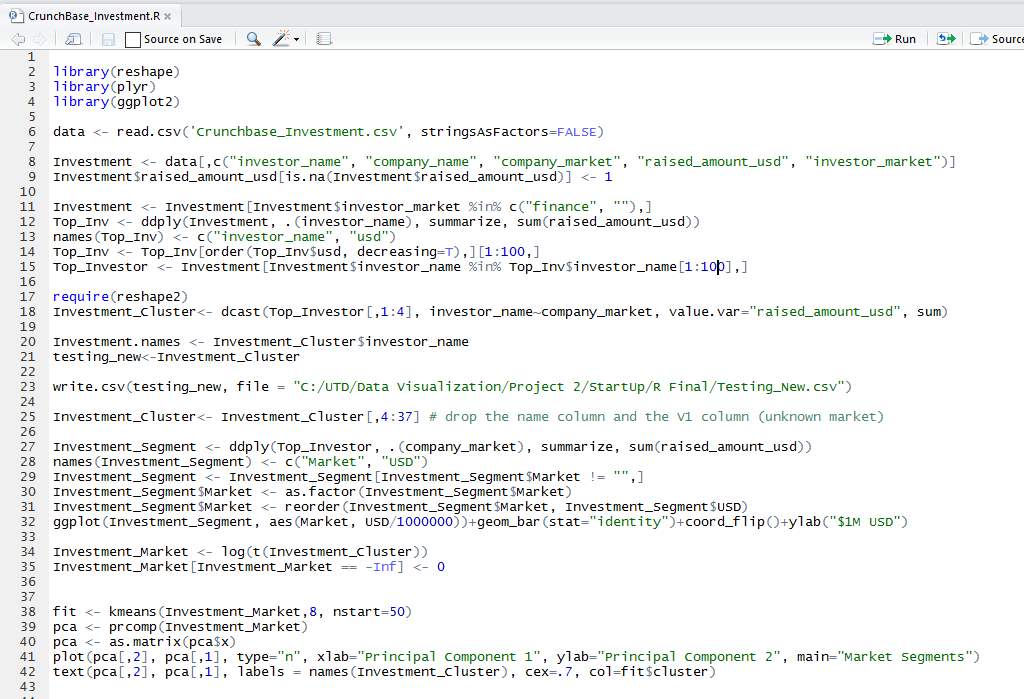
We also made use of the US Statewise Population data that we gathered from <http://www.census.gov/>. For some back dated data, we picked records from [www.google.com](http://www.google.com)/ that were not available on the former URL.



R Analysis

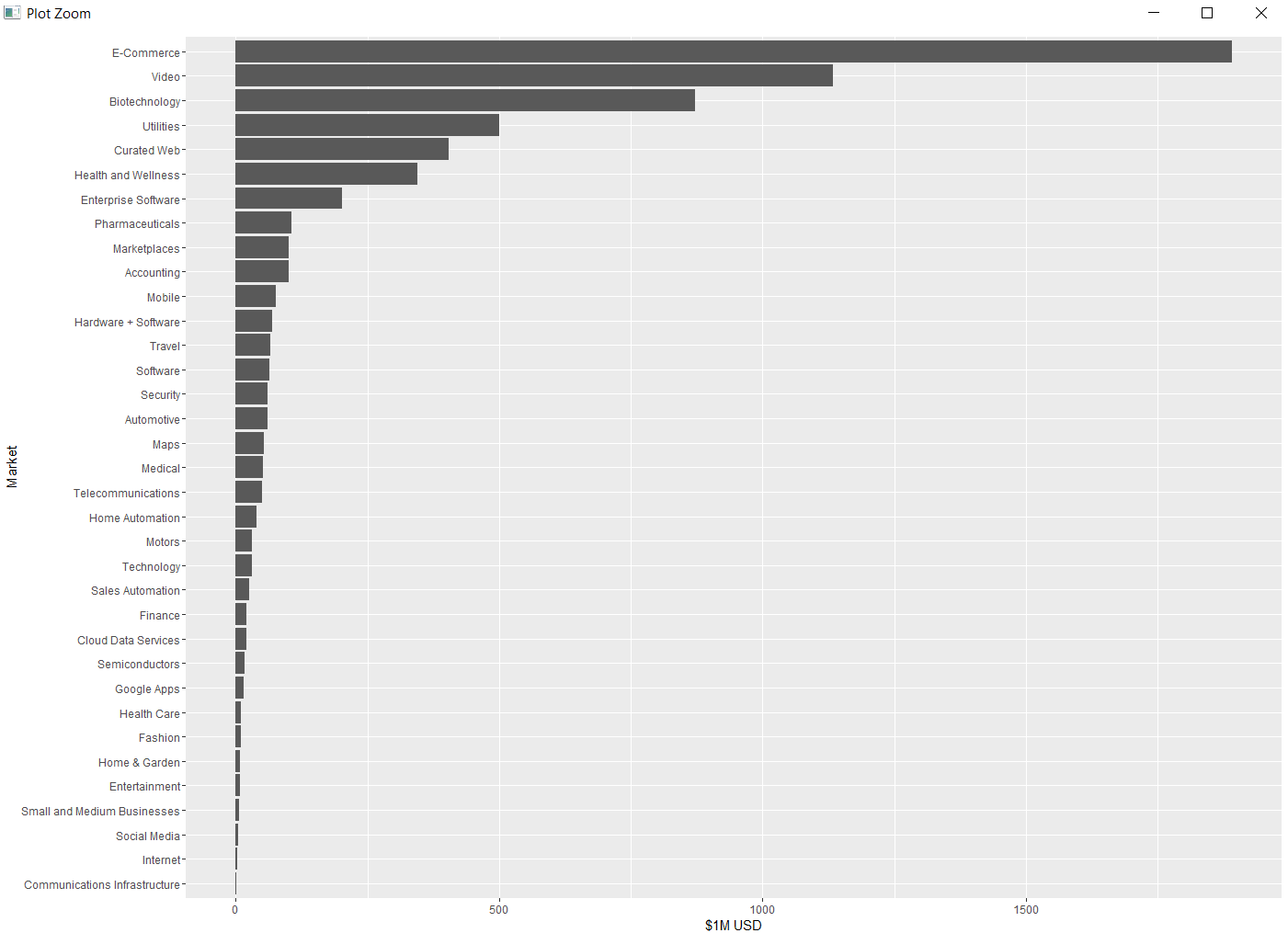
In our project, we have used K-Means Cluster Analysis using R to group records, i.e. the Market Segments, based on the Amount Raised by Venture Investments (top 100).

R Script:



*Note:* The below 2 screenshots have been captured using data for the Top 20 investors (with changes in R script) and data for Quarter 4 of the year 2014. This has been done in order to provide a cleaner visual of the graphs.

This GGPLOT provides us the leading market segments (also the most important) based on the amount of venture funding raised by the top 20 investors.



This plot provides us a visual of the similar Market Segment clusters along the first 2 Principal Components which account for maximum variance in the data. This helps us in analyzing which market areas are related to one another in terms of Venture Funding raised.

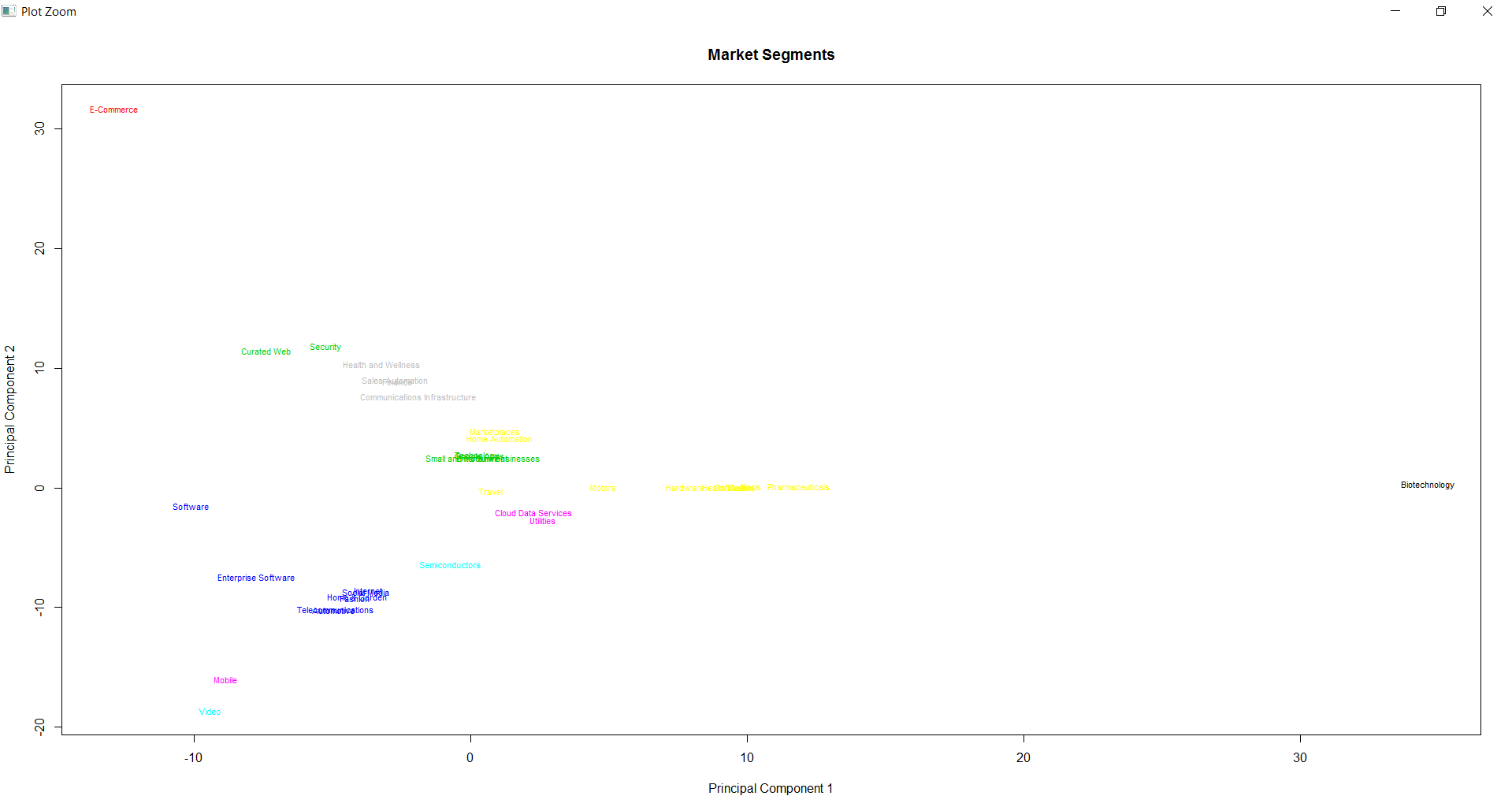
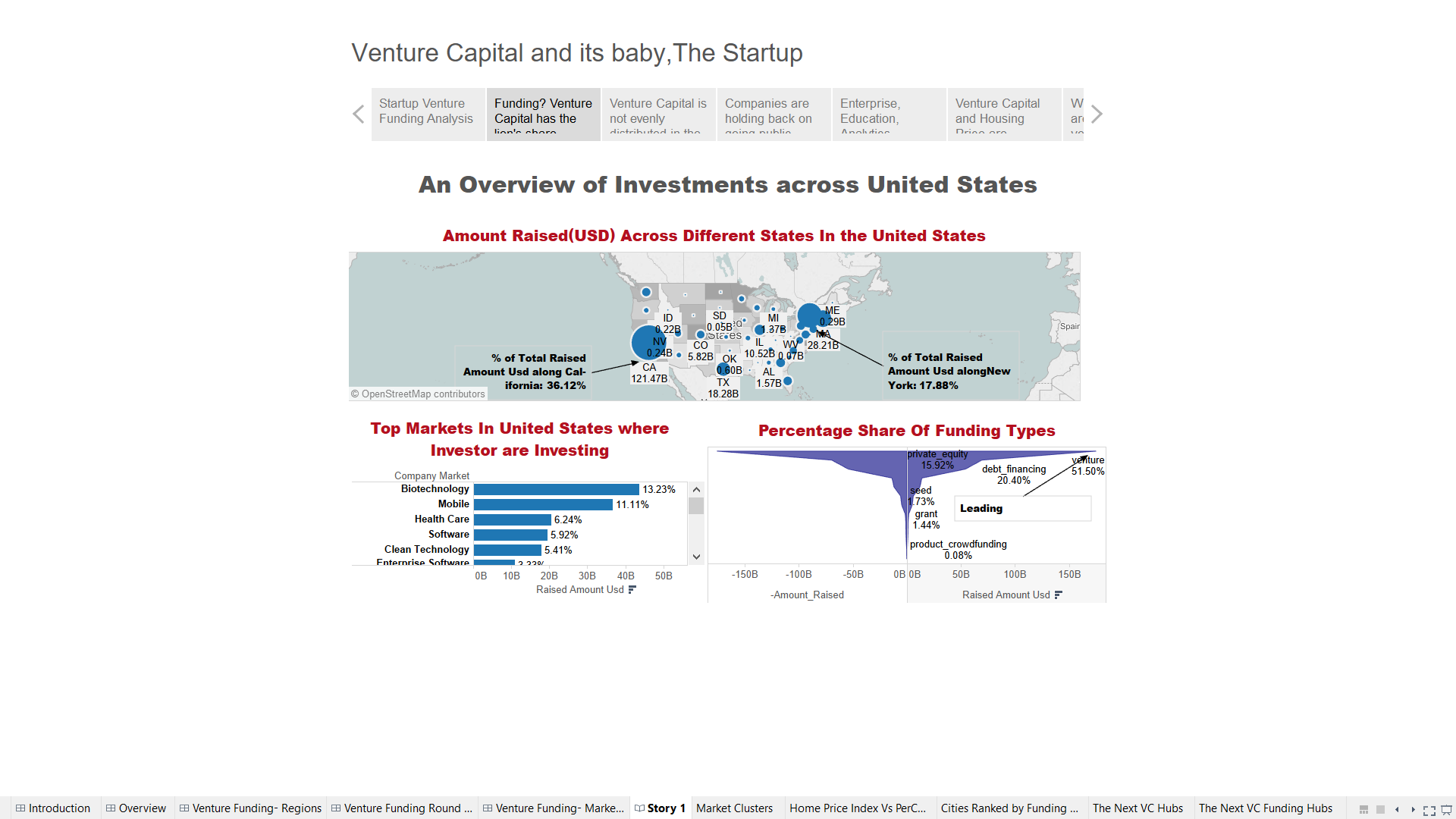


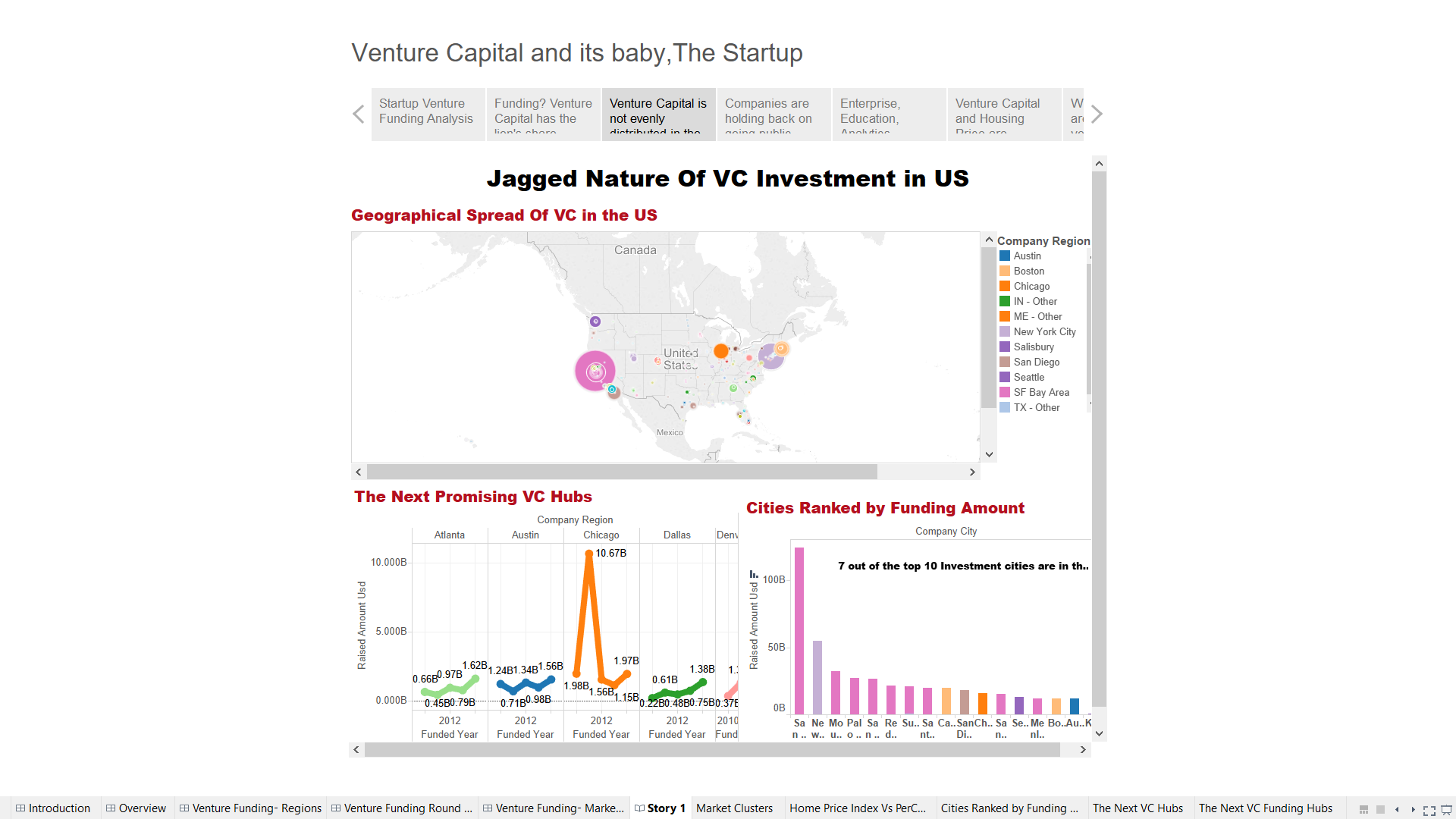
Tableau Visualization

1.



This visualization provides us an overview of the total amount raised (USD) through various funding types across the different states in the United States. The data being accounted is for years 2010 to 2015 with the highest raised amount being observed in states of California, New York, Massachusetts etc. Visualization on the Top Markets in the US provides us the percentage share of the top Market Segments based on the amount raised (USD) through various funding types. As observed, Biotechnology is leading the way in receiving investments by Mobile and Healthcare. The last visualization gives us the percentage share of different funding types across the United States. As observed, amount raised by the Venture Investments (>50%) is the greatest followed by Debt Financing and Private Equity.

2.

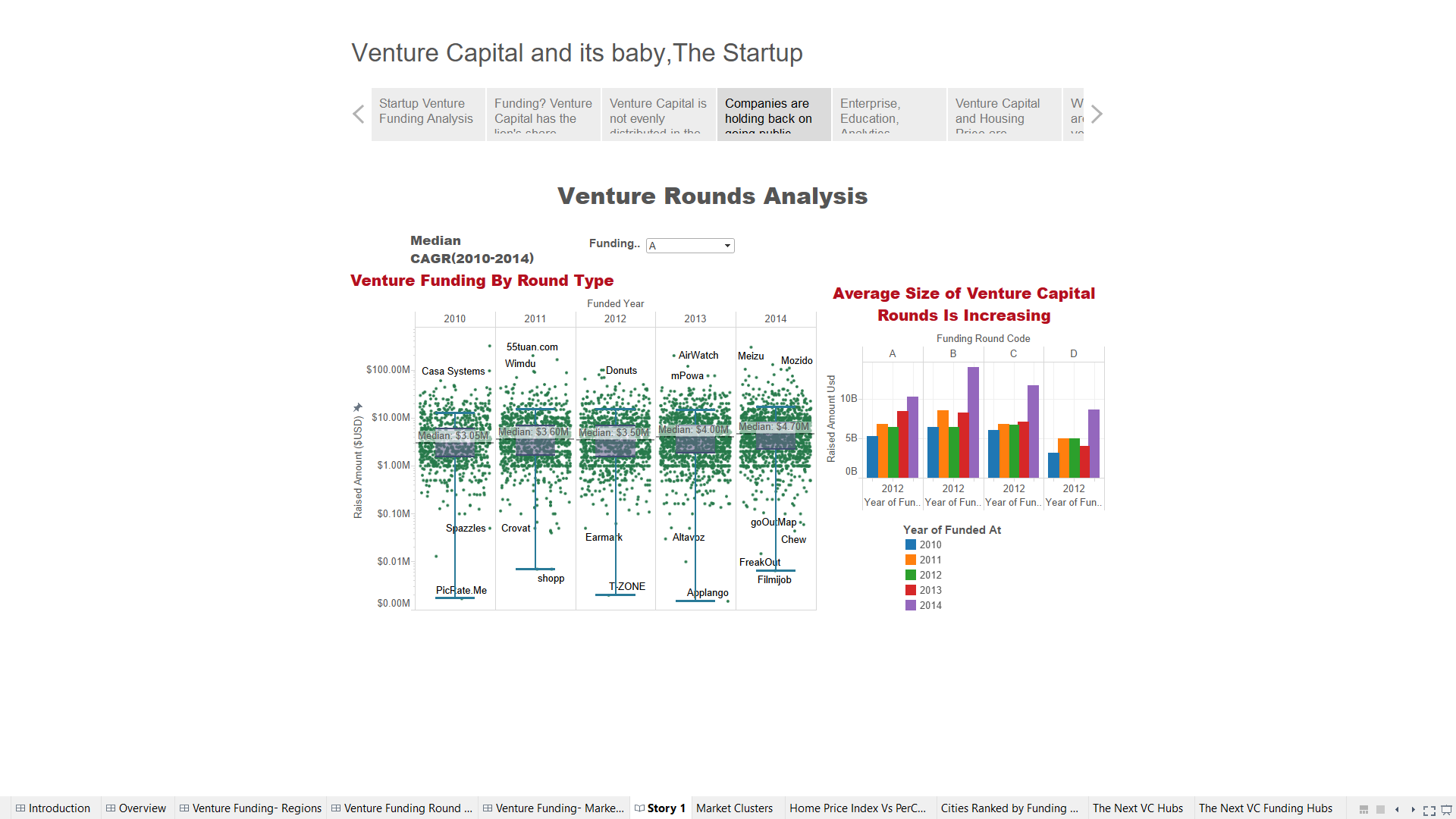


The top visualization provides us a geographical spread of the Venture Investments across the United States. It is evident that the investment pattern is uneven in the country. The top 10 metros account for more than 75% of the investments. San Francisco Bay Area has the lion’s share with about one-third of the total VC investment. It is followed by the Northeast Corridor (New York-Boston-Washington D.C.) and Southern California Corridor (LA-San Diego-Orange County and Santa Barbara). These are the 3 major clusters of the US.

The bottom left visualization adds to the jagged pattern of VC investment in the United States. 7 out of the top 10 locations for VC investment are in the SF Bay Area. The distribution of VC investment in not evenly distributed.

The bottom right viz. provides with the cities which come in the second tier in terms of venture capital investment amount. They also show a rising trend in venture capital invested year.

3.

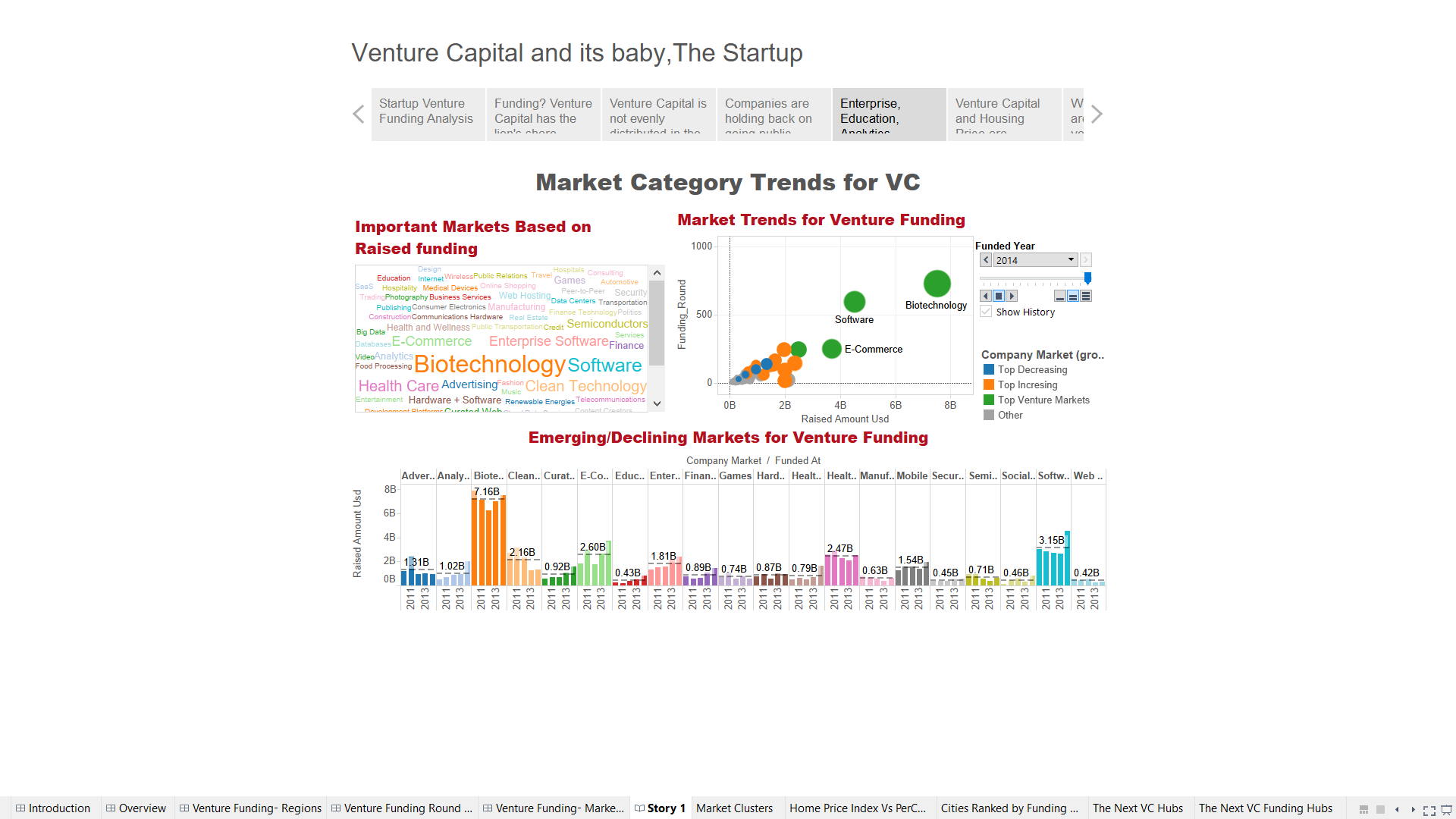


The Median CAGR (2010 - 2014) gives us an average of the Median Venture Funding for the different round types.

Insight

As observed from the box plot for 2014, the median of Series D round was $26 million, approximately 63% higher than that in 2013. *The noticeable increase at the later stages indicates to the deep pipeline of companies that are choosing to stay private for longer by delaying IPO plans.*

4.



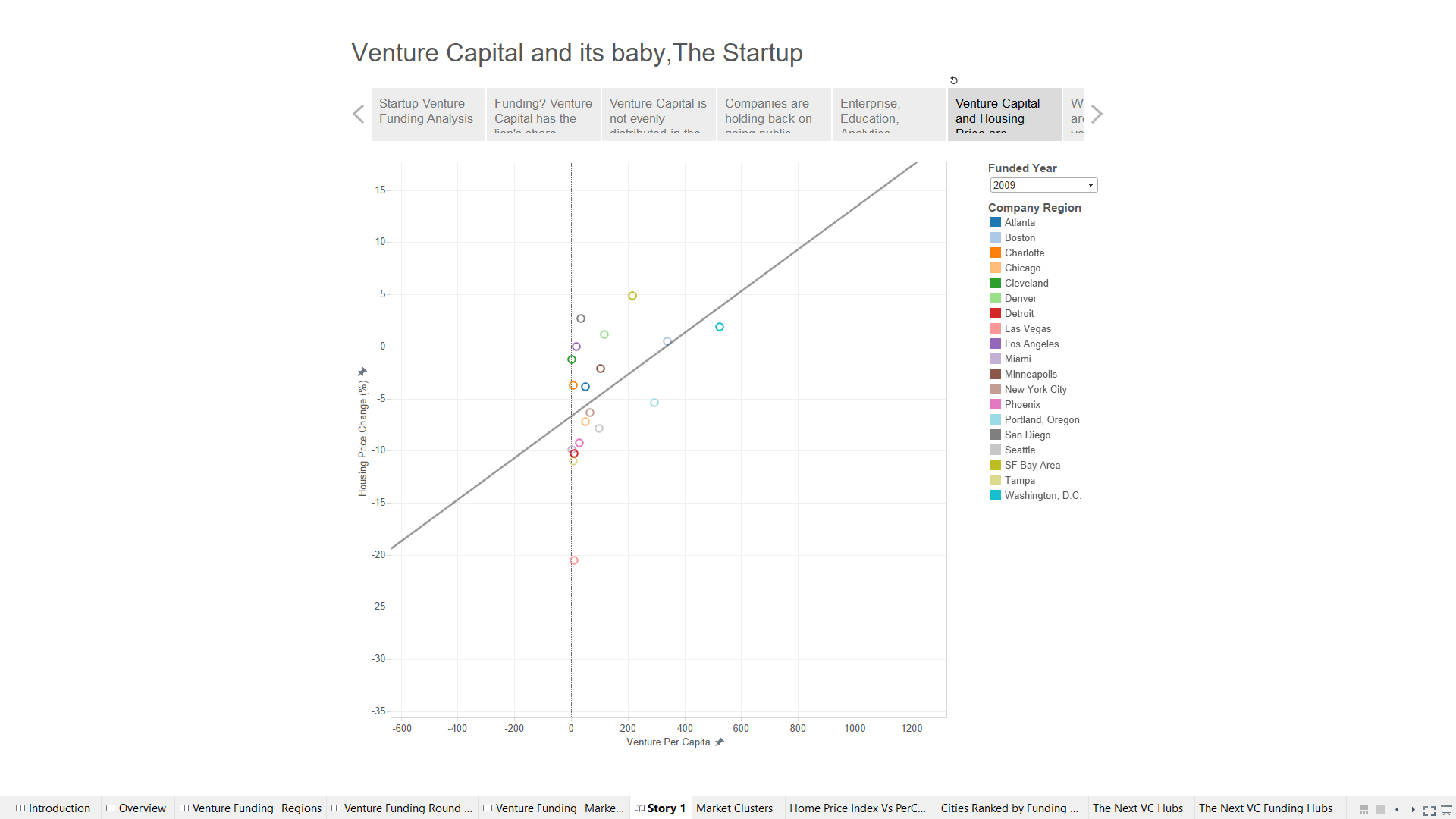
Top left viz. provides with the top markets based on amount of VC investment.

Top right viz. provides a yearly update on the emerging market areas for venture investment.

The bottom viz. provides with the changing trends in the markets. Though Biotech still remains a leader in receiving the funding, other markets such as Enterprise, Hardware, Security, Education, Finance and Analytics have received increased funding year on year.

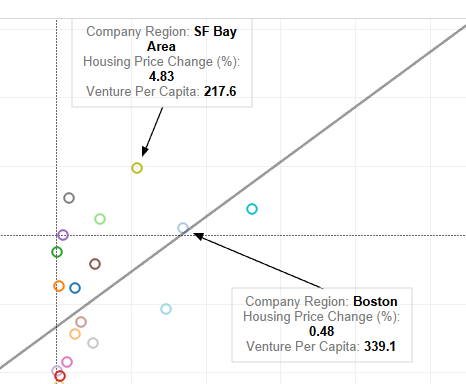
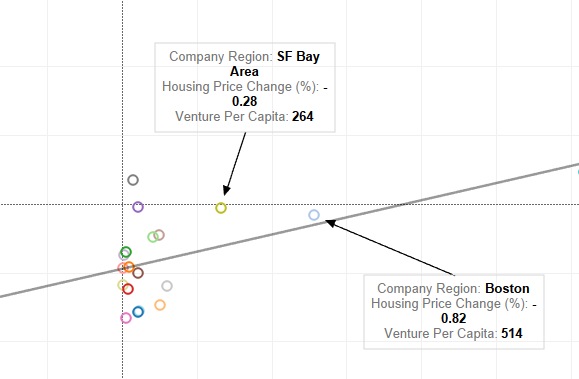
Markets such as Mobile, Clean Technology, Social, Gaming, and Advertising are on the downhill.

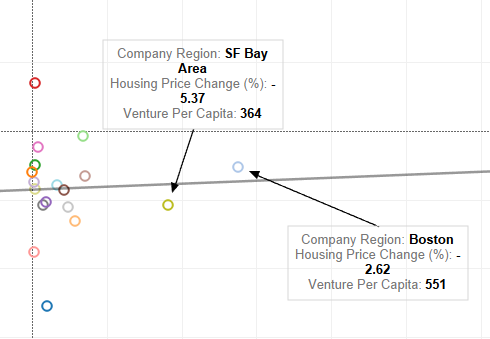
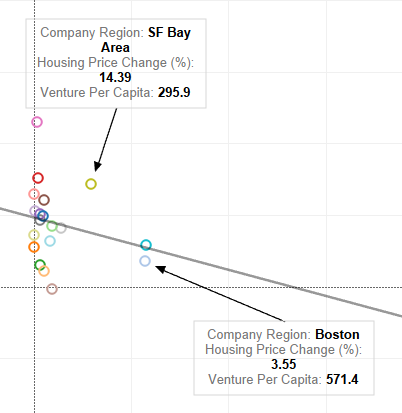
5.

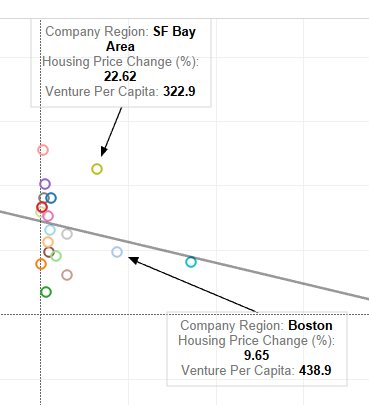
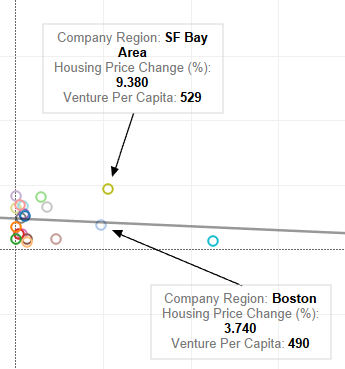


Insights

* *The Trend Line in the visualization can be attributed for developing a correlation between the Home Price Index (1 year change) and the Per Capita Venture Investment of the State. As can be observed in the viz. above, the trend line is rising with the increase in per capita venture investment along with a rising Home Price index.*

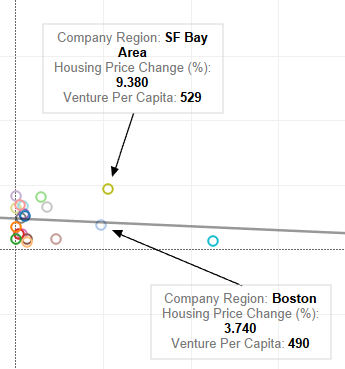
2009  2010 

2011 2012 

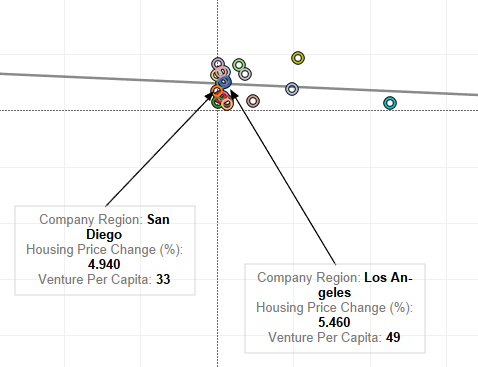
2013 2014 

* *Also, due to California’s large overall population, Massachusetts outperformed the former in terms of per capita venture investment for years before 2014. Though, the increasing investment in other areas of California like Los Angeles and San Diego the per capita for the state has dominated over the other. (*Observe the screenshots below for the per capita investments for SF Bay area and Boston)

Los Angeles/San Diego Per Capita - 2013

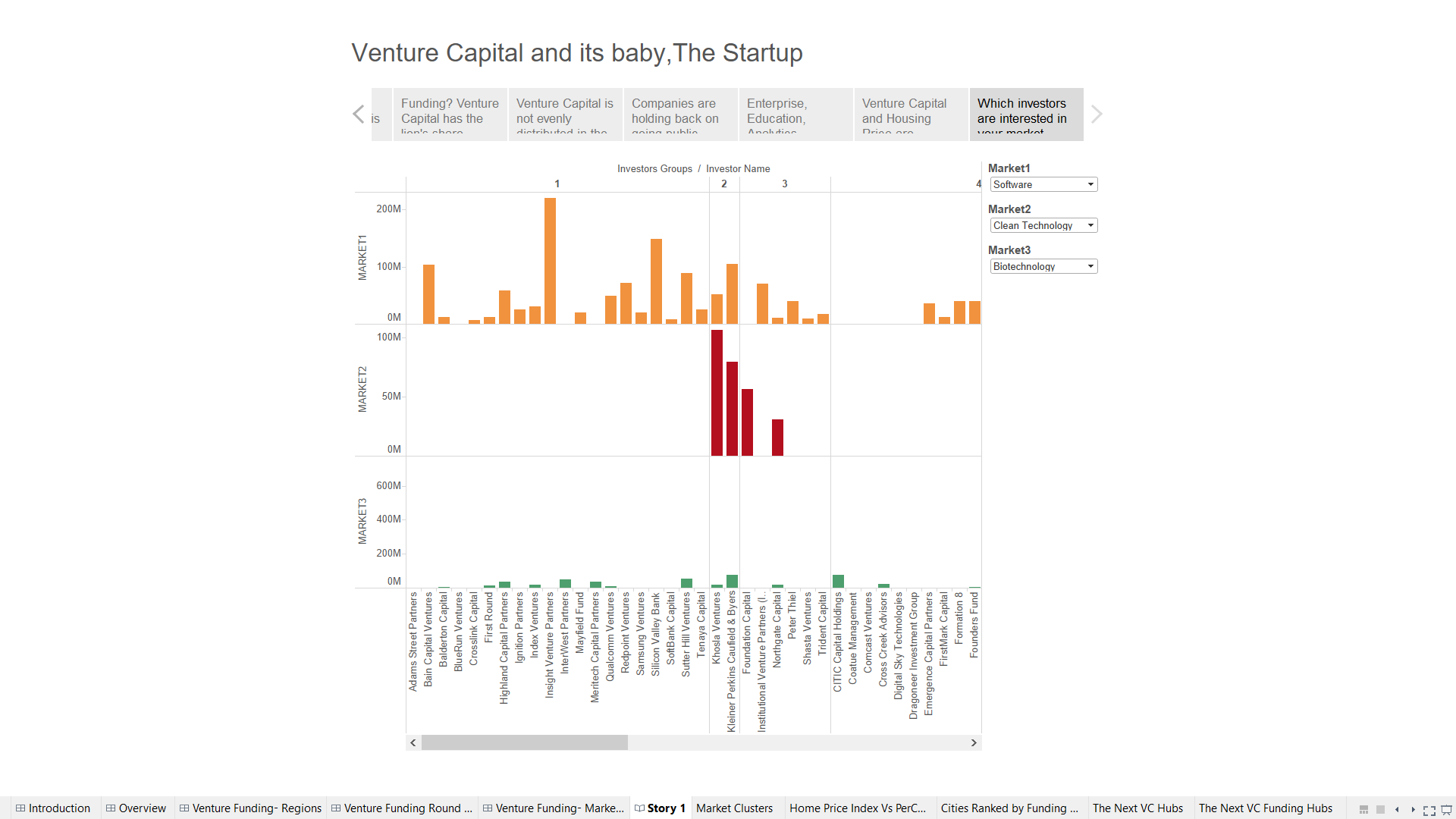


LA/San Diego Per Capita - 2014



There has been an increase in the per capita of LA from 2013 to 2014.

6.



Insights

*This visualization provides us with a relation between different market segments in terms of their investments. It can also be taken into account that the Venture Capitalists who are investing in one or more market segments are also investing in market areas belonging to the same clusters.*