

# **Investors & Startups:**

## **Crowd Sourced Data on Investors and Startup Success**

**Hosted:** <http://unicornviz.us>

GitHub: <https://github.com/gracenbrilz/InvestorsStartups/tree/master/UnicornsApp>

Gracen Brilmyer

Daniel Chen

Pratik Nadagouda

Carlo Liquido

# Table of Contents

1 Project Goals

2 Related Work

    Animated Map

    Stacked Bar Charts

    Shaded Area Chart

    Linked Bar & Line Charts

3 Visualization Descriptions (screenshots at the end of each section)

    Overall Design & Layout

    Animated Map

    Stacked Bar Charts

    Shaded Area Chart

    Linked Bar & Line Charts

4 Data Used

5 Tools Used

6 Accomplishing Our Goals

7 Feedback & Future Work

9 Table of Responsibilities

# 1 Project goals

The technology world is a mythical place. It's without a doubt the place where a lot of current excitement is happening. But the ever-changing nature of the technology sector makes it difficult for people to wrap their heads around on what exactly is going on inside this world. We have chosen to apply a data-driven storytelling approach to illuminate some trends as we drill down into the CrunchBase dataset.

Our interface is designed to allow users to see different high-level views of the data. In the interface we are asking many questions about the data, such as where the majority of VC funding is coming from and into what startup markets, however we aim to allow users to explore their own angles of inquiry. Although we are not providing any answers as to *why* different VCs and startups fail or succeed, we hope to allow users to compare different aspects of the data and compare overall trends. They can see amounts of funding reported and received from different countries, compare funding between markets and against the total and averages of all markets, and contrast funding reported in all startup markets with those valued at \$1 billion (Unicorns) which is one metric of startup success. We also cleaned and selected the data from the past ten years, which seemed to be the richest reporting of funding. By creating these visualizations we are helping the user make sense of the data at a high level, where they might want to pursue further, more detailed, inquiries after looking at the bigger picture or if this dataset mimics or differs from other resources related to VCs and startups.

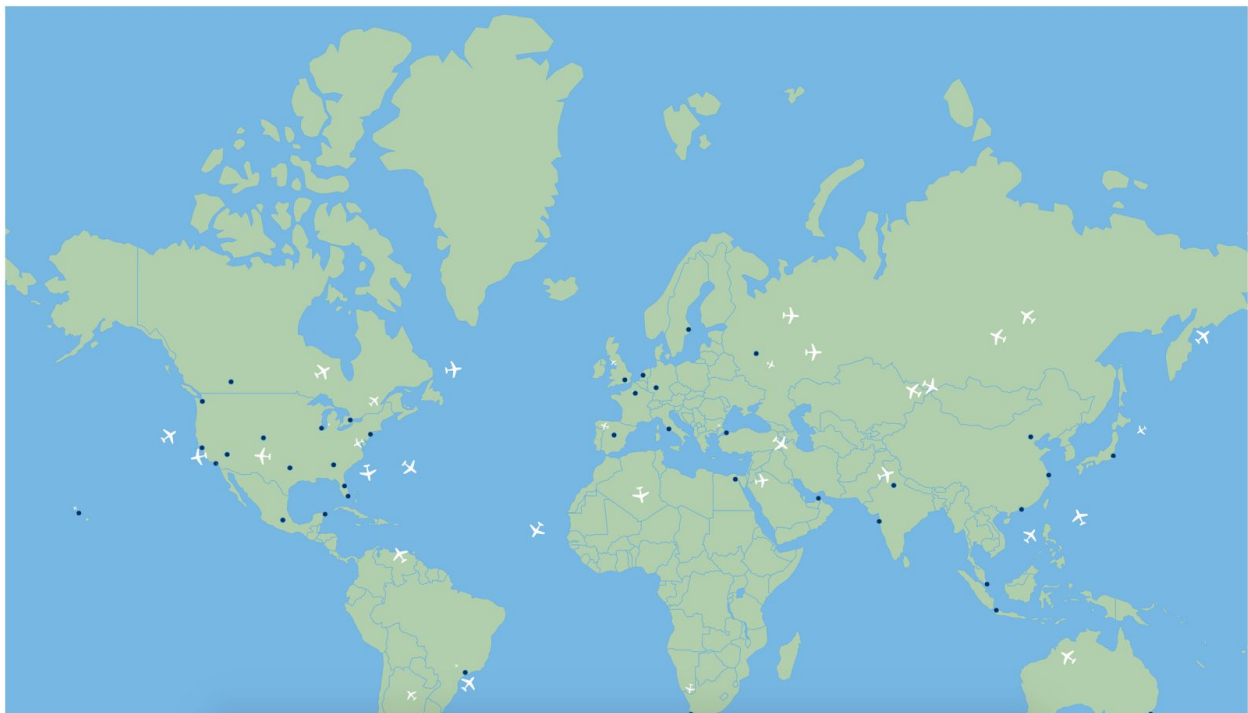
Lastly, each of the team members wanted to further develop their skills of the JavaScript library, D3. Challenging ourselves both in designing and executing interactive visualizations was a priority was important as we asked questions about a dataset.

## 2 Related Work

There are four main pieces of visualizations in our project – Animated Map, Animated Stacked Bar Charts, Area Chart, Shaded Area Chart, and Interactive Bar/Line Chart. We have utilized these give interactive graphs to design a story where a user is first exposed to the overall trends of our data and then later provided tools where they can interactive with to gain more details about the data. Each of the visualization is inspired by some work that have previously been done, and we have apply some creative spin to create our own style of graphs that best suit our story.

### **Animated Map**

Related Work 1:



*Fig 2.1 Animated Map Inspiration*

<http://www.tnoda.com/blog/2014-04-02>

This animated map where airplanes are flying from one destination to another is what inspired and led to the creation of our animated map. This map has led our team to realize how interesting a map can be by applying animation. We also believe that an animated map

can not only grab a user's attention, but also provided the user a high-level overview of the data, thus explains why we used the map in the section of data exploration

Related Work 2:



Fig 2.2 Animated Map Inspiration from "How to make a map"

<https://bost.ocks.org/mike/map/>

The above map is made by the creator of D3 and has gone over the basics of how to make a map with D3. Tutorials cover concept such as map project, map scale, data format, installation...etc. It was the main tutorial I refer to when I made the map.

Related Work 3:

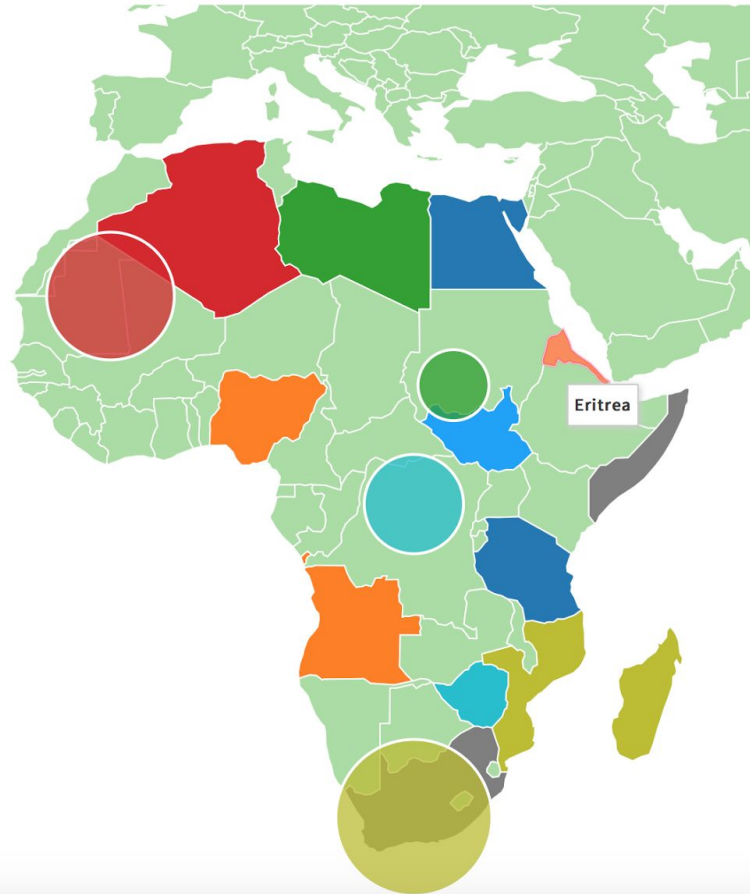


Fig 2.3 Animated Map Inspiration from "Data Map"

<http://datamaps.github.io>

This map made by data map has provided the primary source of inspiration as to why we chose expanding circles to represent our data point. We understand the flaw of circles is that it's very hard to compare since areas of circles do not provide precise measurement. However, as we can see from this above image, the snapshot provided a very easy comparison and allowed user to tell which circles are bigger than the smaller ones. The ability to color code the circle is also powerful and we do want to use it in the future to enhance the map.

### **Stacked Bar Chart**

Related Work 1:

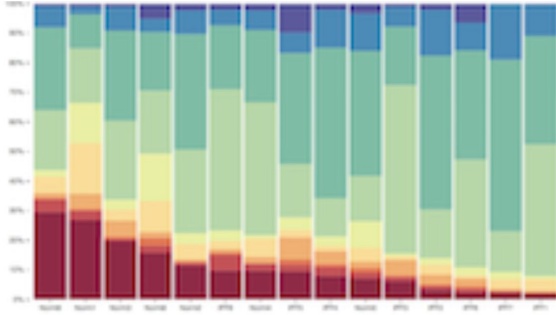


Fig 2.4 Animated Stacked Bar Chart Inspiration

<https://gist.github.com/stylecoder/6a8f6ef464bae17f97cf>

There are two stacked bar charts in section two. The first bar chart display the total amount of funding for the first top 10 countries, and the second bar chart displays the percentage of a country’s investment allocation. We have found this very interactive D3 animation online and thought it would be perfect for the job. Since one of our chart is showing how each country invests in each other or itself different, the information is highly loaded and is hard to be displayed at one glimpses, having this interactivity where user can click on a country and the chart would sort according to the country clicked would enable the users to dig into the data themselves.

**Shaded Area Chart**

Related Work 1:

Relevant Source 1 provides major inspiration to why we have decided to used area shaded chart. The source visualize the number of people killed versus the supposed “stolen years” from the murders as to compare how much time could these people lived if they weren’t killed. The highly informative design inspired our application to our market segment and funding data since we want to compare the funding received in each market and compare them against the number of startups.

Related Work 1:

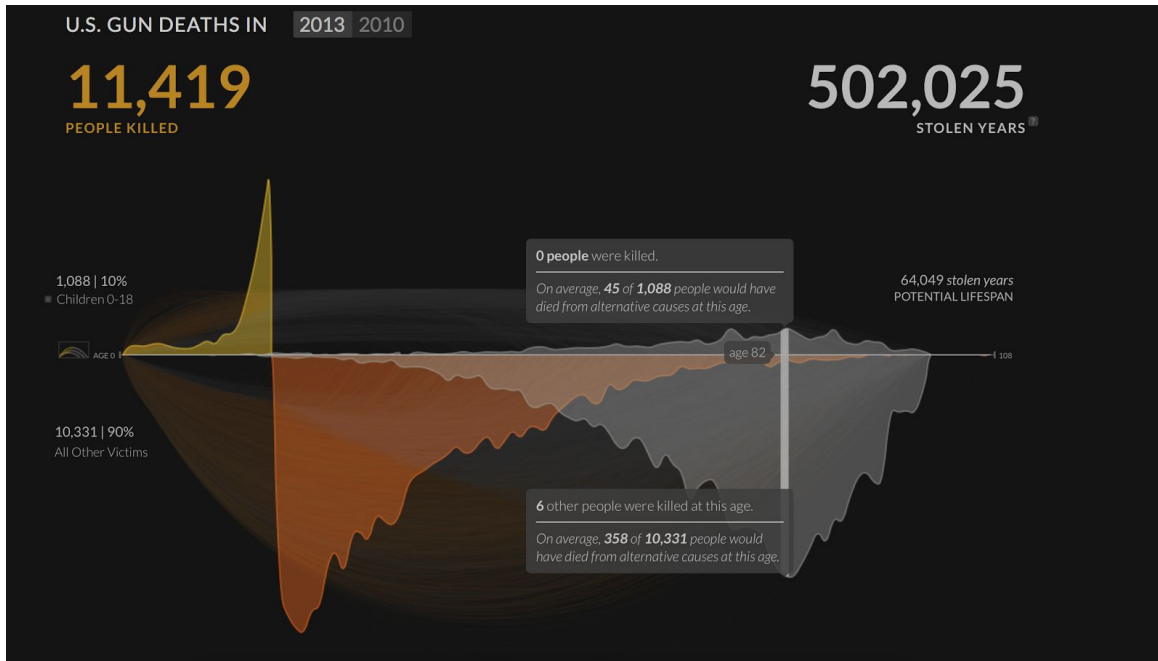


Fig 2.5 Shaded Area Chart Inspiration

Our area bar chart is intended to allow comparison easy and interactive. Since there are two groups of data we wanted to display, we have used the double-sided area shaded chart to show our data. Relevant source one is a source we have referred to in order to make tooltip very useful for the user.

Relevant Work 2:

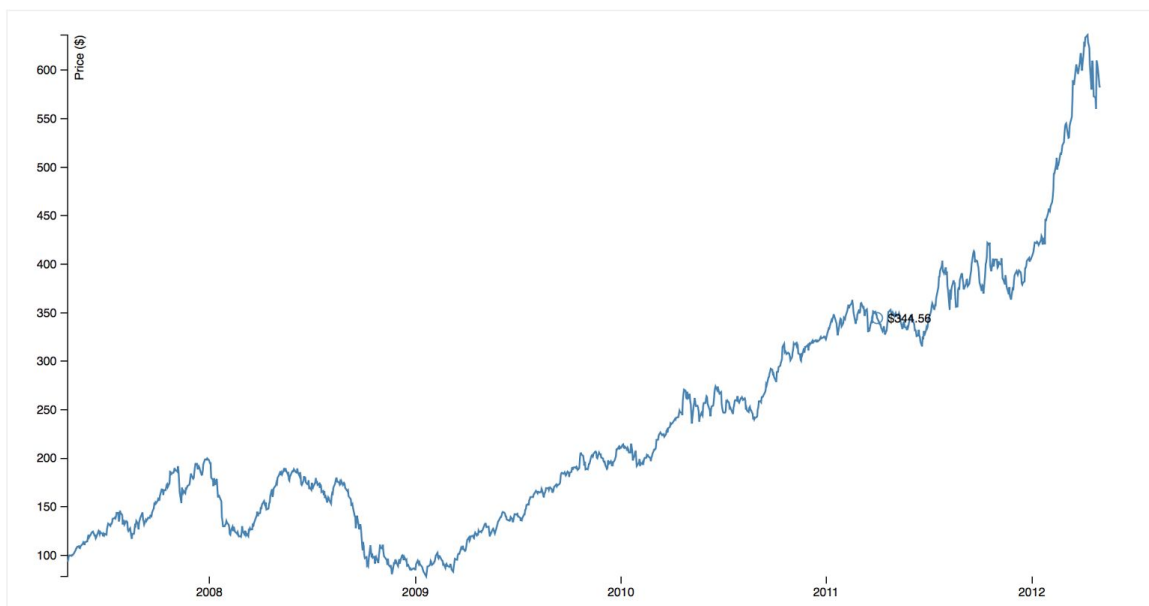




Fig 2.6 Shaded Area Chart Reference

<https://bl.ocks.org/mbostock/3902569>

Related Work 3:

Lastly, we used relevant source 3 to make our transitioning of charts smooth and elegant.

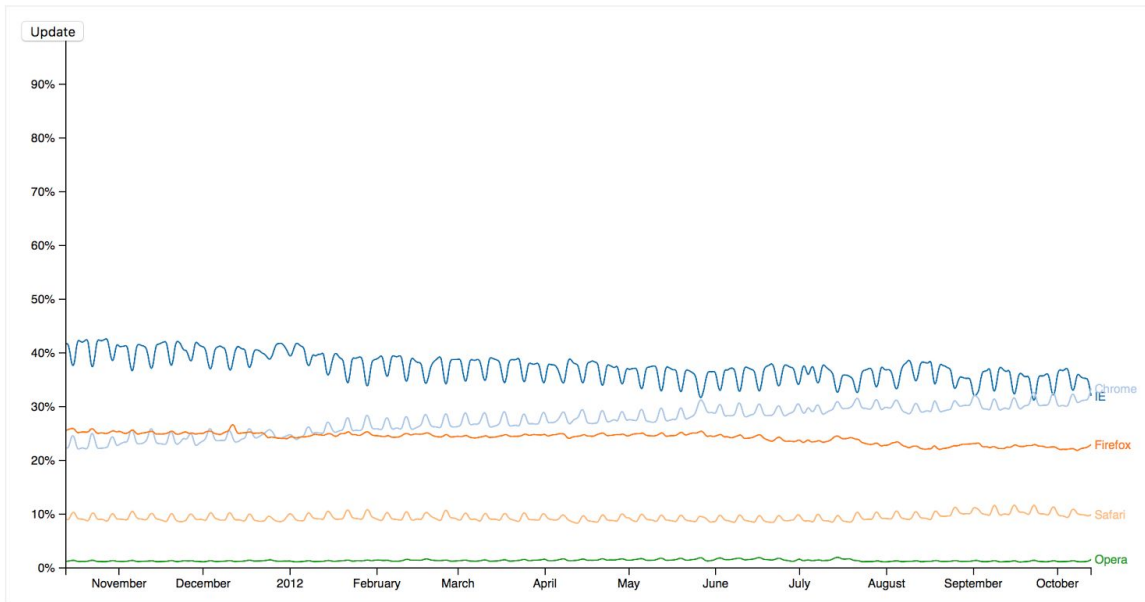


Fig 2.7 Shaded Area Chart Reference

### **Linked Bar and Line Charts**

Related Work 1:

Our last visualization focuses on the most interesting part of our data, the subset of unicorns and their funding information. The bar chart is linked to the line chart and as user clicked on the bar of each market, the line chart will show the funding history of each of the unicorns in the market. The bar chart will show user information about the valuation of the unicorns in the respective market comparing to the number of startup (shown in another bar chart). The line chart links together each circle and the size of the circle is determined by the size of funding. We have referred to relevant work 1 as an inspiration of using bar charts to display the metadata about unicorns.

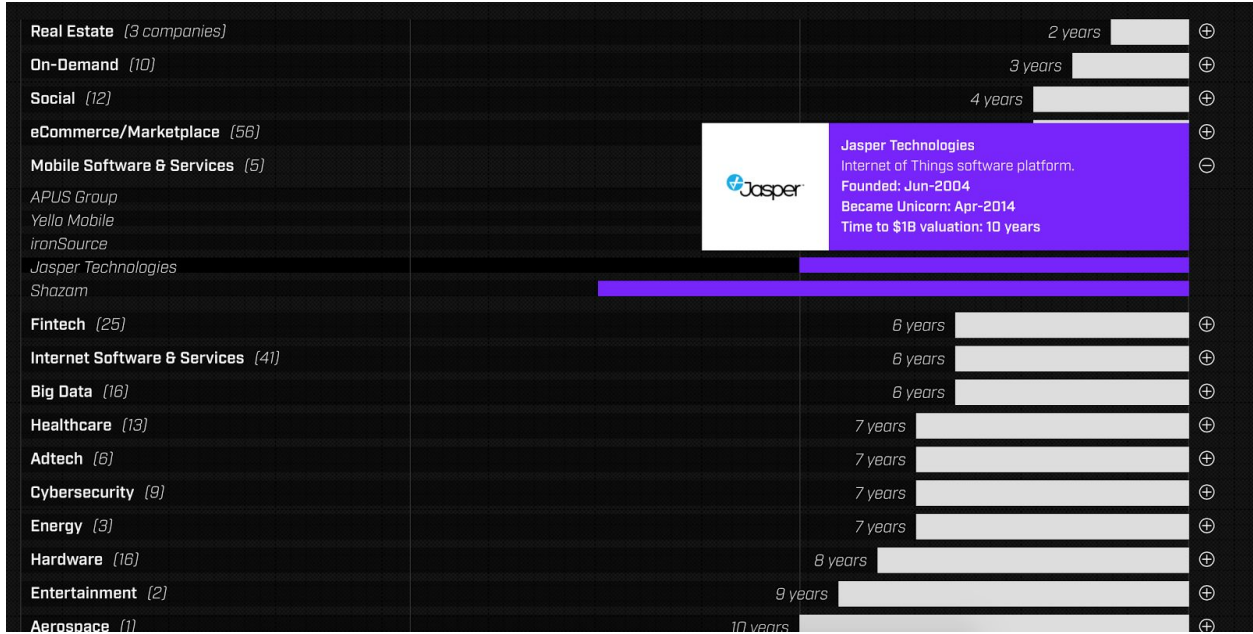


Fig 2.8 Linked Bar Chart and Line Chart Inspiration

<https://fleximize.com/unicorns/>

Related Work 2:

As discussed, we used the size of circle to determine funding size, and we used relevant work 2 as a reference to build the visualization.

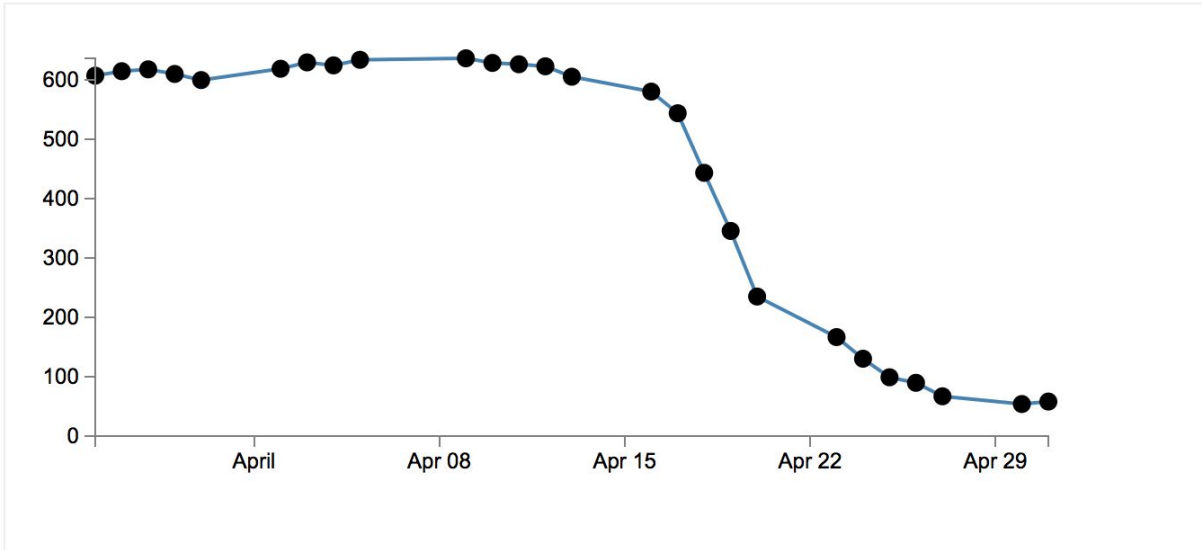


Fig 2.9 Linked Bar Chart and Line Chart Reference

<https://gist.github.com/stylecoder/6a8f6ef464bae17f97cf>

### Related Work 3:

At last, we used relevant work 3 in our last visualization to make our line chart interactive and multi-dimensional. The below work provides source code where we can refer to during the process of development.

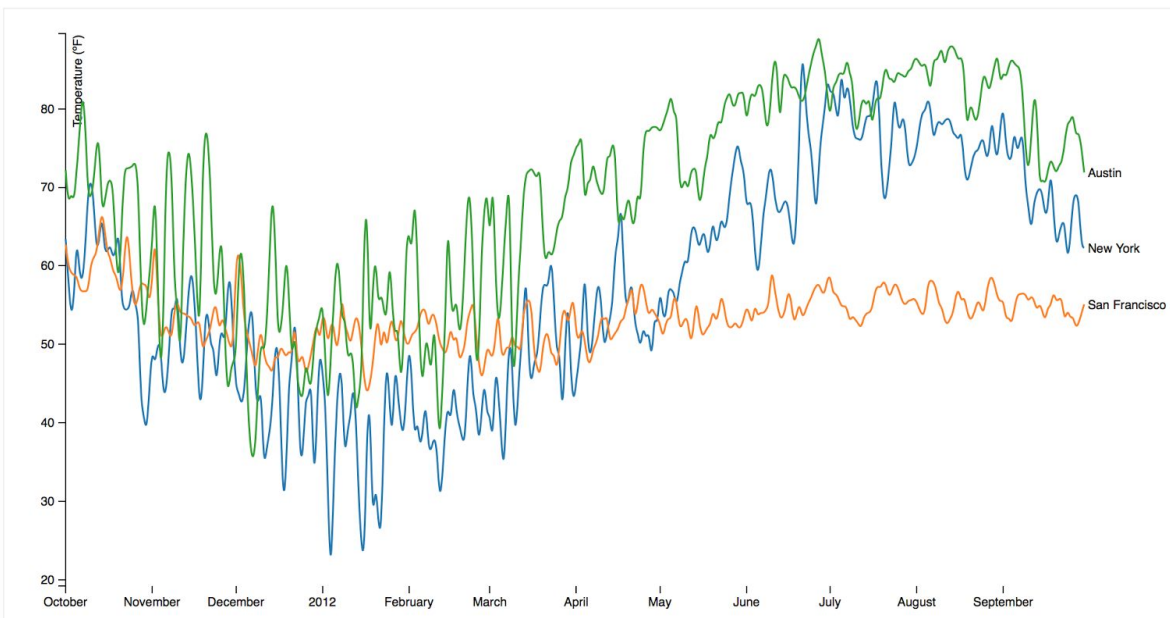


Fig 2.10 Linked Bar Chart and Line Chart Reference

<https://bl.ocks.org/mbostock/3884955>

## 3 Visualization Descriptions

### Overall Design and Layout

As we have set out to do, our goal was to leverage D3's power as an information visualization tool that provides very customizable ways of designing our story. The design for the visualization was inspired by the current aesthetics in startup and VC culture, particularly that of financial related media such as Forbes (Fig 3.1). By using serif fonts for the title and body text and sans serif fonts for the section title as well as bordered boxes for containing and connecting different pieces of information, the interface took a contemporary spin on classic business visual styles. Using a single page site for layering and displaying the different information and visualizations was another design choice to keep an editorial look. Additionally, we utilized a floating side menu of our questions as the dominant navigation for the page, which also mimics the covers' use of side text to ask questions that tell the narrative of the content.



Fig 3.1 Magazine Cover Examples for Aesthetic Inspiration

### Animated Map

The initial visualization that the user is presented with is a map that depicts where funders are located. The map was designed as the first visualization to be displayed to the audience's eyes when they visit our website, so we want this visualization to intrigue the audiences into exploring the rest of interactive graphs. That being said, the map should accomplish two things: For one the map should display the locations of VC funding by answering the question "Where are startup investor located?" and for two the map should show the general trends of the sizes of the rounded being given by investors across the

world through the timeframe of 2003-2014. To accomplish these two goals, we designed the animated circles that would expand from a radius of zero to radius of (number of rounding given out by all VCs in a given city) \* 4 for every months in the course of 10 years. As the map plays chronologically, semi-translucent green circles are sized amount of funding and burst in the location of that funder. Displayed at the bottom, to help orient the user, is a progressing year and month, the top funding city from that time, as well as how many funding rounds occurred. Through this map, the user can see that reported funding dramatically increases as time passes, as well as the US, in particular Silicon Valley, play a dominant role in funding given. We decided to animate through all months because displaying all funding across the world in a year will yield to too many data points in one snapshot for the user to absorb and will also impose a very challenging technical task due to the limited performances of browsers' ability to render that many data points. The trade-off was being made between the entire animation being too long and results in audience losing interests versus every snapshot of animation (a month) is animated too quickly resulting in the audience feeling lost and not being able to interpret the information.

We want to provide some facts on the side along the approximately 3 minutes to ensure the user is also absorbing other useful information at the same time. In terms of how we achieved this animated map, we first have to display the locations of cities accurately. We have used python's package named geopy and requested 3000 unique cities' latitude and longitude that are in our data and map them individually in pandas and Vlookup in excel. We have experimented the animation with both SVG and Canvas and ended up using SVG. The pros of SVG are that it's interactive and allows the user to hover over the map to look over tooltips and the quality of display is much higher than Canvas. However, when we tried to display more than 200 cities within one second, the screen became laggy and the rest of visualization is affected. Then we solved the laggy problem by using D3's data group function that will allow us to render more than 300 data points at once. We spent a lot of time trying to optimize the animation in SVG because we weren't satisfied with Canvas' image resolution and ability to do more customizable animation that would be possible in the future even though that we were able to display more than 1000 data points with Canvas without lagging. To display the data points, we converted csv files downloaded from Crunchbase and I aggregated and transform the data to give columns, which are "investor\_city", "number\_of\_round", "latitude", "longitude", and "dates". These five columns are stored in a dictionary with dates as the key and inside each key there are key-value pairs of city, rounds, latitude and longitude. The javascript will read each date and wrap the key value pairs in a data group and send it for d3 to render all cities at once. We used requestAnimationFrame to time the animation better to make sure there wouldn't be overlapping because javascript's setTimeout is not as accurate. In terms of map projection, I used d3's mercator projection to project the shape of the map, as it's the most suitable for

the task. In terms of future work for the map, we plan to provide more interactivity with the users as the animation is displaying, we want to show the tooltips on the maps so the user can hover over and get more detailed information about each city. Also, we want to color the circles by the cities that are being invested by the VC so the user would be able to see another layer of information which is the destination of the money.

Having a good understanding of the data, we think the users will have a few takeaways from watching the animated visualization. We think the users will easily understand that most of the VC funding in the world are given out in the United States and the size of VC funding has grown significantly in the period of 2007 – 2013 comparing to 2003 – 2007. Below are two snapshots that will provide the comparisons of the sizes of funding. As we can see from the screenshot below, in 2004, there are some actions going on, but as we compare to the second screenshot, the differences are easily perceived.



Fig 3.2 Animated Map Screenshot





*Fig 3.3 Animated Map Screenshot*

As the user goes through the animation, we expect the user to notice the growth of the funding and generate more questions about the data. Another thing we expect the user to notice in the second screenshot is the abundance of cities that are generating circles and the varying size of the circles across different cities. Even though United States dominated most of the funding, the user will still be able to observe many other actions that are going on around the globe.

### **Stacked Bar Charts**

Below the map, a stacked bar chart also updates in correlation with the map, showing the the amount of funding coming from each country (x axis) and in what countries they invest (y axis). Early in the development of these two charts we realized that using all of the data was extremely overwhelming for the user - not only did it make each bar chart very thin (to fit in the space provided) but also used so many colors (since the number of countries invested in were greater) that it became difficult to tell certain colors apart. To make the pertinent information more valuable to our users, we decided to cut the data down to only the top 10 countries that invested the most money.

The two stacked bar charts are made with different JavaScript files, both using D3. The initial stacked bar chart (Fig 3.4) accrues data simultaneous with the map. It clearly reinforces and more clearly illustrates the accruing data from the map, depicting the US and dominating VC funding. The raw data shows that the US dominates both in the most funding given and the most received, making it difficult for a user to see what the data

shows for other countries. As the reported funding amounts add up, the US bar chart towers over the other 9 countries, making it difficult to view the details of those countries. In an attempt to counteract this minimizing of the other countries' data, we added a hovering tooltip, so a user can still see the country invested in and the amount.

In order offset the compression of all non-US countries on the y-axis, we also made a normalized version, where a user can see the same data by percentile (Fig 3.5 & 6). Any country funded in this stacked bar chart can be clicked on which results in it being dropped to the bottom and sorted numerically for easy comparison. This feature was chosen to address as aspect of stacked bar charts that isn't a usable: that it is difficult for a user to compare the same variable across different bars if it is not at the bottom. As we have learned in class, it is hard both for users to quickly and accurately compare multiple areas, as well as compare more than 2 row numbers. The dropping to the bottom helps the values be compared and the sorting by greatest to least makes comparing them even easier where a user doesn't have to memorize the numbers in order to compare them (Fig 3.6). In addition to these features, we also added a tooltip that shows the country invested in as well as the percentage invested to help make the data easier to understand (Fig 3.6).

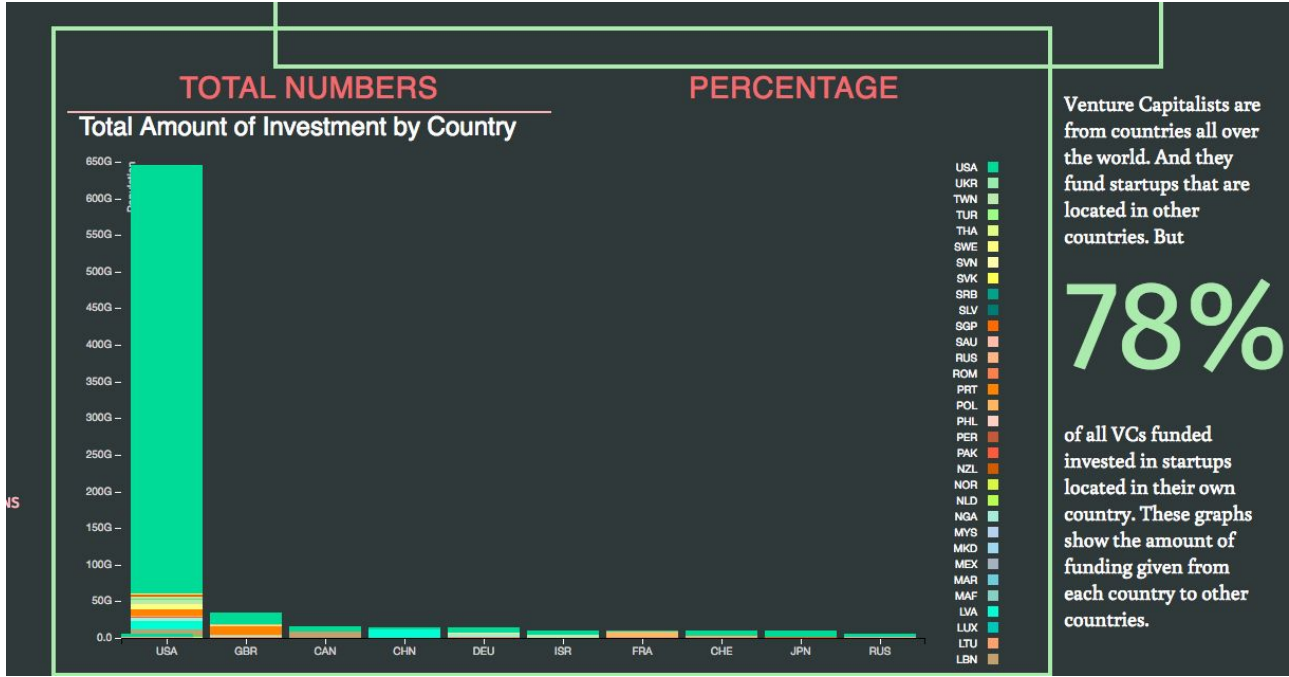


Fig 3.4 Stacked Bar Chart: Raw Data Loaded Screenshot



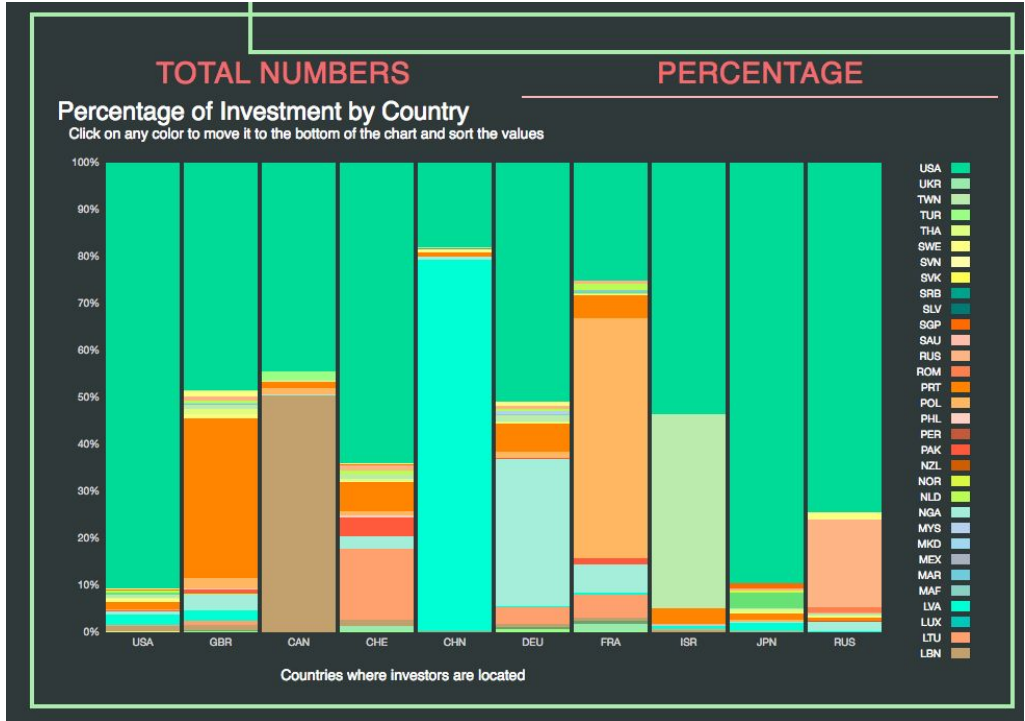


Fig 3.5 Stacked Bar Chart: Normalized Screenshot

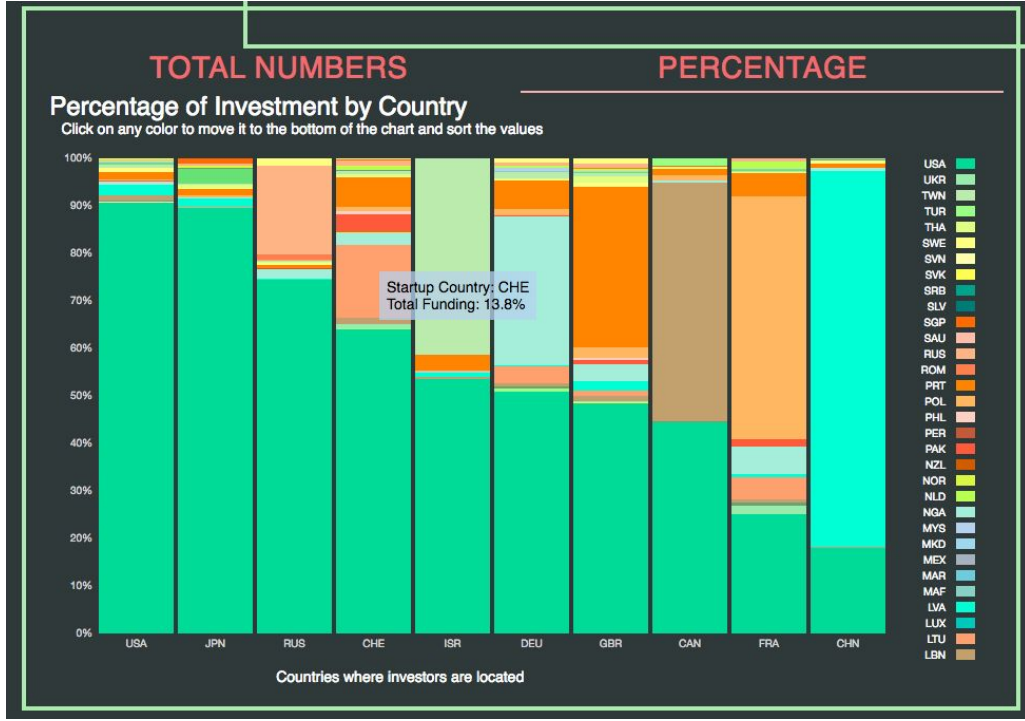


Fig 3.5 Stacked Bar Chart: Normalized & Sorted with tooltip Screenshot

## Shaded Area Charts

After exploring location data in the first two graphs, we wanted to look into the various markets for startups and investors (Fig 3.6). Each startup was assigned a market based on its field of operation. We extracted a subset of the data with the top 15 markets and analyzed different measures for each of them.

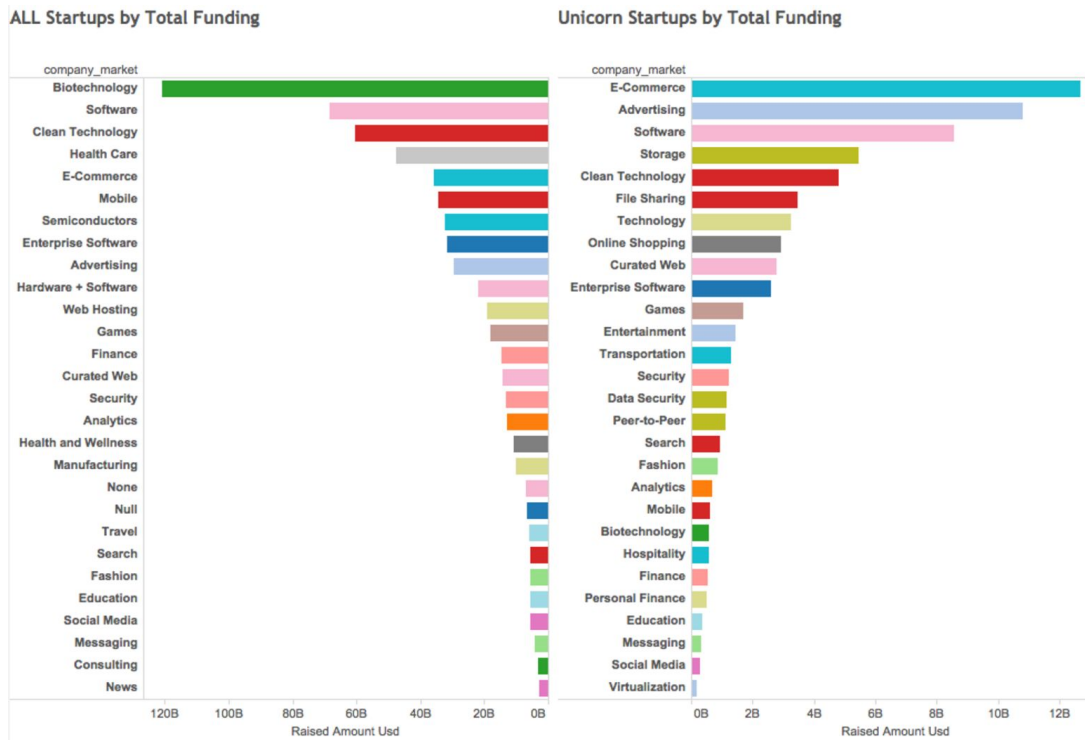


Fig 3.6 Bar Charts comparing total funding by market by unicorns and all startups

We decided to focus on the total amount of funding for each market along with the number of startups in each market over a time period of ten years from 2004 to 2014. We initially had a larger data range, but we received the concern that older data may not have the same accuracy and completeness as more recent data. Showing each of the points over time did not provide a good view of the information. So instead we aggregated each month by the median amount. We used median instead of average in order to not be affected too much by outliers. This was important in this case because while most funding amounts were in the low millions, there would be a tiny amount of irregular amounts of around \$1 billion (such as Alibaba in the E-commerce market).

We depicted the actual data in two area charts that followed along the same X-axis. This was in order to observe patterns with the relationship between number of startups being funded and amounts they were funded over time. It would have been too crowded to display all the market areas at once, so we included a toggle list on the left, which can be used to

display one market at a time compared with the median of all markets together. The list is ordered by the total funding given to all startups in each market (Fig 3.7).

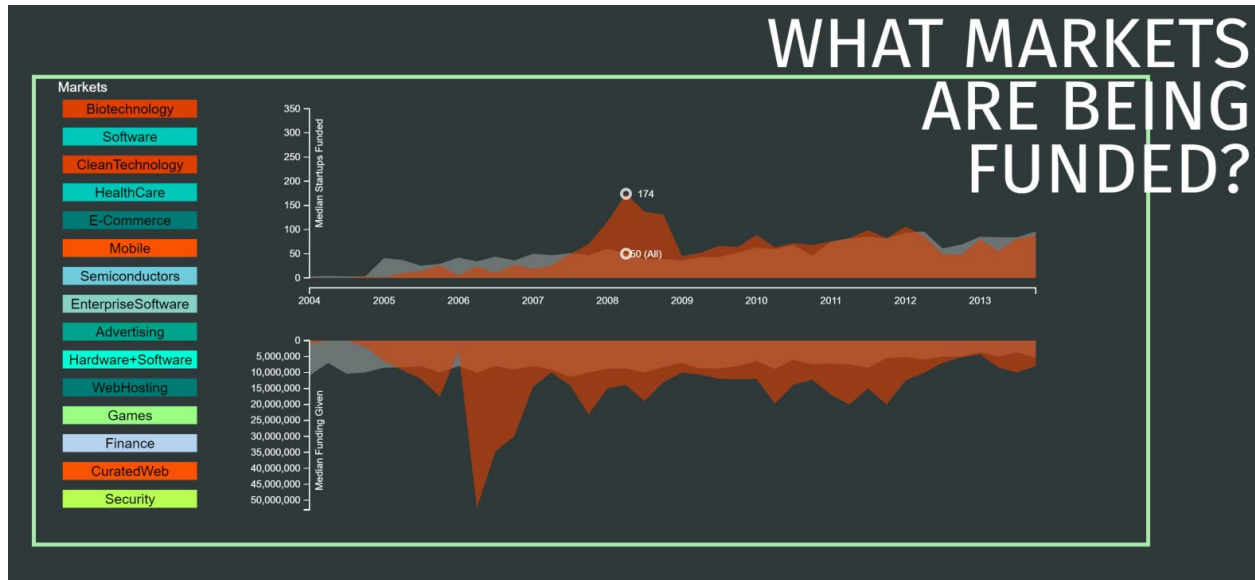


Fig 3.7 Area charts depicting median funding and median count of all startups by market

The opacity was set low in order to view both areas overlapping each other. We also added a mouse hover that showed the values at each point in time for the chart. The transitions between market selections are animated with the help of D3.

Through initial feedback, we were told that the bottom graph may give the impression of negative data. So we attempted to flip the charts on their axes so that time flows vertically (Fig 3.8). After looking at it, we felt that time as a scale is read better horizontally. But we included a screenshot of that version below for documentation.

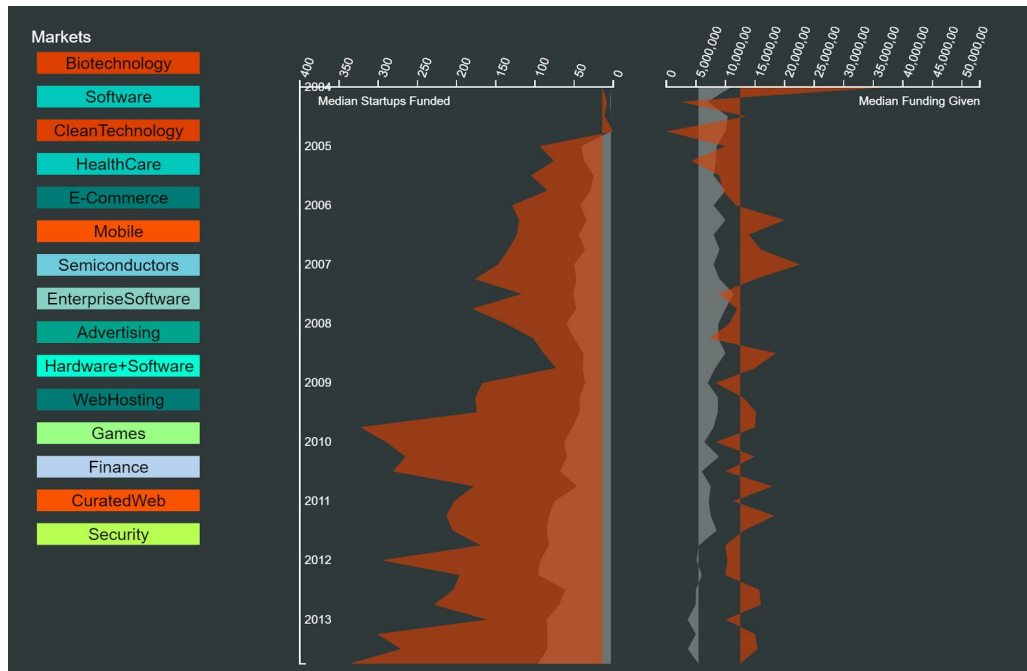


Fig 3.8 Replicated area chart from Fig 3.7 flipped vertically

## Linked Bar and Line charts

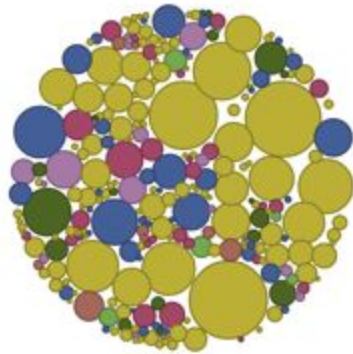
The previous visualization speaks to which markets are being funded overall. It gives you a general sense of each industry’s performance over time. However, it does not tell you about the differences between successful startups and unsuccessful startups and the how the landscape differs when viewing them separately. We decided to define a successful startup as one that has become a unicorn—that is, a private or public company valued at over \$1 billion. By scraping the TechCrunch website for current unicorns, exited unicorns, and emerging unicorns, we were able to retrieve a comprehensive list of relevant unicorns for our dataset. There were roughly 200 unicorns that we could use to compare against all startups.

For this last visualization piece, we went through a number of iterations. Our goal was to present the data as granular as possible without losing sight of important trends among startups. We first created a bubble chart (Fig 3.9) that arranged all the startups and venture capitalists into a large circle, where color signified market and size signified amount funded in total. When the user clicked on the market, the bubbles would rearrange themselves into their respective markets. This gave the user a sense of which markets were largest and who were the big players in each. However, after some additional analysis and feedback from

user testing, the results were quite ambiguous. Many users felt inundated with the vast amount of data. There were also some who were curious into the individual rounds of each startup over time. This created a very difficult problem—how to show even more granular data without overwhelming the user.

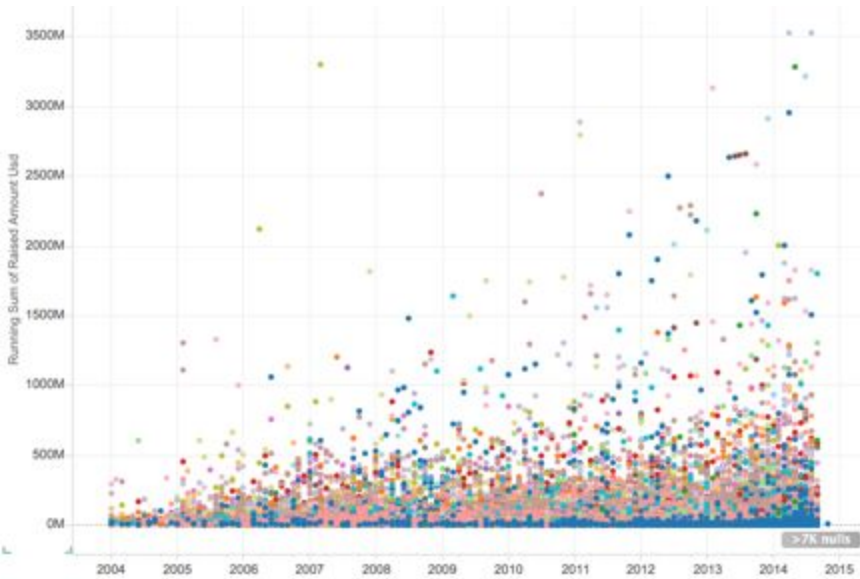
### Top Eight Markets by VC in 2013

All VCs VCs By Market



*Fig 3.9 Bubble Chart displaying total funding by market*

For the second iteration on this visualization we decided to use a scatterplot of each round over time for all startups colored by their market (Fig 3.10). Users would be able to filter on unicorns and non-unicorns in order to get a sense of how these startups differed. We imagined that the scatterplot would show in aggregate interesting trends. However, we found that the vast majority of startups are funded similar amounts, which skew the axes to the point of total ambiguity. We realized at this point that in order to represent all the data, we needed to create two separate linked graphs—the aggregated graph with the ability to filter individual startups and their respective funding rounds.



*Fig 3.10 Scatter plot of each startup round by year and funding*

The final iteration of the visualization (Figure 3.11) was thus a medley of aggregated bar charts and disaggregated line charts. The bar charts aggregated the data by displaying the total amount of funding for each of the top 35 markets—the first bar chart was subsetted by only unicorns while the second demonstrated data for all startups. Here, you are able to see some very interesting trends. There is a stark discontinuity between the amount of funding received overall and the amount of funding received by just unicorns. For instance, Biotechnology is heavily funded overall and in fact receives the most funding. However, when you look at just unicorns, you see that Biotechnology receives very little funding. These bar charts are then linked to the line chart. By hovering over each market, the line chart is subsetted by that market according to all unicorns. Here, you are able to see the lifespan of a unicorn and the amount of funding it received at each round. The tooltip also allows for further drilldown.

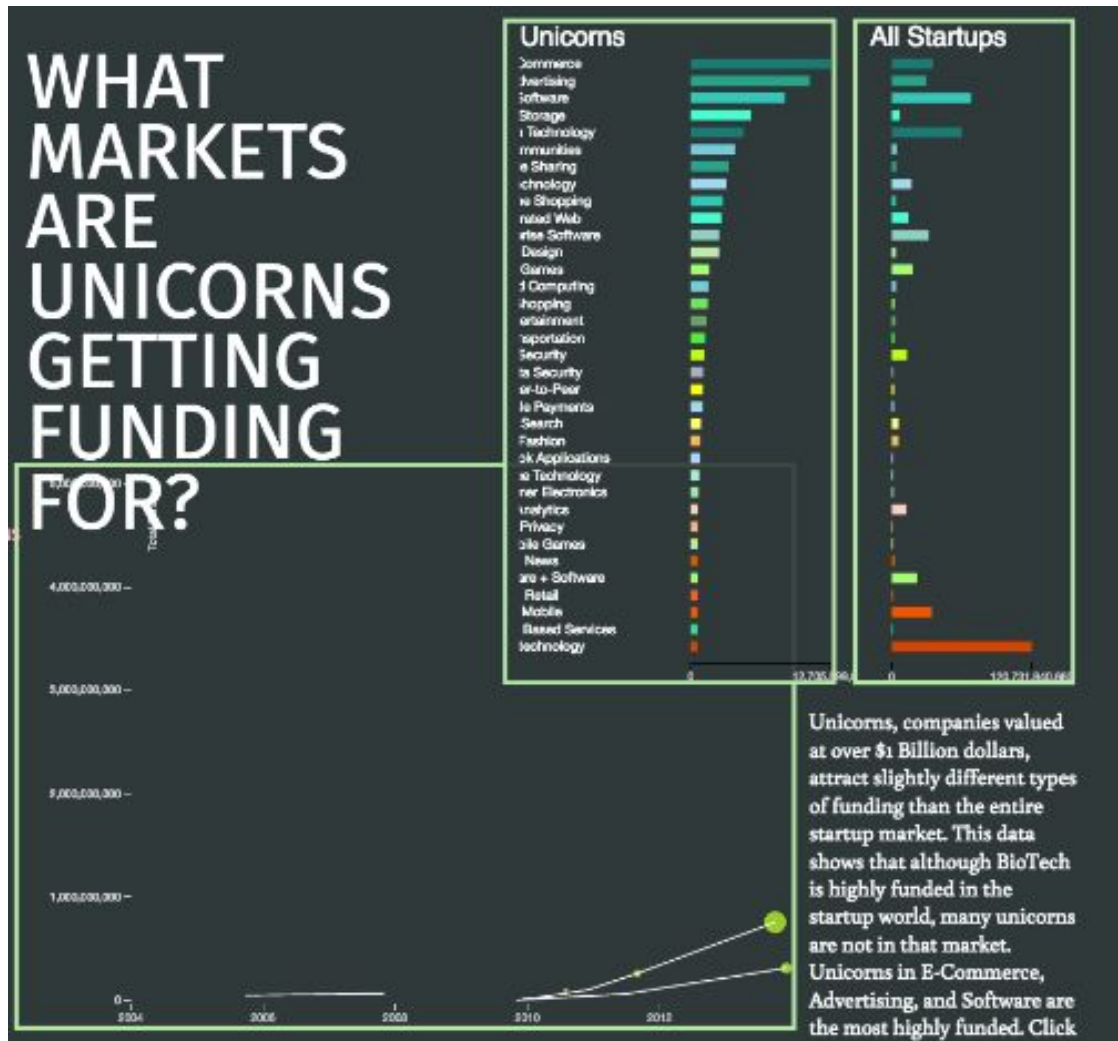


Fig 3.11 Aggregated funding bar charts that drilldown on unicorn startups and their respective rounds

## 4 Data Used

Our primary data source was Crunchbase, an online curated data platform, hosted by Techcrunch. The data present there has been collected by members of the public, by making submissions to the database. All changes and additions are subject to review by moderators before being accepted. Live data is also monitored by editors, who make updates accordingly. This makes Crunchbase similar to a Wikipedia for Startup data. There are currently 50,000 active contributors to the dataset. We extracted the fields for company name, market, location, funding received, investor name, investor location, and date. These rows each pointed to a single instance of funding from an investor to a startup at a particular point in time. We divided the data into subsets based on how we wanted to use in our charts.

The first chart animation used investor location and an aggregation of number of rounds for each 'money bubble'. To make the chart in d3, we required the use of an external python package, Geopy, to retrieve latitude and longitude data for each city and country. For the stacked bar charts, we used investor country and startup country to depict the source and destination of the money. The area charts required an aggregation of the existing data, to find the median number of startups in each market and the median funded amount in each market. We cut a subset of the data here to include only the top 15 markets by amounts funded.

Unicorns were the highlight of the last part of the visualization. Unicorn data was not available in the initial dataset, so we had to acquire it through another source. Techcrunch maintains an online list of current Unicorns at <http://techcrunch.com/unicorn-leaderboard/>. Using the beautifulsoup and urllib python libraries, we scraped the data from the website. In addition to current unicorns, we also retrieved emerging unicorns and exited unicorns. Thus, by adhering to their Terms of Service, we gathered a list of company names identifying around 200 unicorns.



## 5 Tools Used

- Tableau
  - For the exploratory data analysis phase we extensively used Tableau. It was not only used to explore the data but also to validate our assumptions and design prototypes. We found that many times the ideas and visualizations we drew on paper did not reflect as well using real data.
- Python
  - Pandas
    - Since we were aware that the data was dirty and quite intricate, we decided to use python (specifically pandas) for data cleaning. We created and streamlined cleaning pipelines for our different datasets and varying visualizations. At many points in the process, we discovered that we wanted to shape our data in one way rather than another (e.g. grouping by the year instead of month). Just using excel would eventually mean more work and open the project to possible mistakes.
  - Geopy
    - Python package used to handle geo data such as finding coordinates, distances, etc. We used geopy for our world map in which we retrieved the latitude and longitude of over 3,000 cities.
- D3
  - For our final visualization we decided to use D3 exclusively for all four of our pieces. This included a geolocation map, bar charts, stacked bar charts, line charts, and area charts. We decided to use D3 for it's flexibility in terms of animation and linking.
- Materialize
  - We used materialize as a way to structure the front-end content. It allowed us to make our gridding system more easily and provided the scrolling interactivity.
- Github
  - All of our code is hosted on github. Because we had a fairly large group of four people it was necessary to have clean version control. We applied best practices in keeping our master branch clean and using controlled branches for temporary work.

## 6 Accomplishing Our Goals

We pursued our data analysis efforts with the ultimate goal in mind that our visualizations should insight curiosity in people interested in the venture capitalist and startup world. First we found that there great disparity in the amount different markets are funded from 2004-2014 (Fig 6.1). Biotechnology receives a great amount of funding as compared to many markets. Upon further personal reading, we found that this made sense, as the startup costs for biotechnology companies are generally very high. You can also recognize that most of the funding is concentrated in the top ten markets. From a very early point in our exploratory phase, we recognized that diving into market distribution would be both critical and interesting to understanding the startup world.

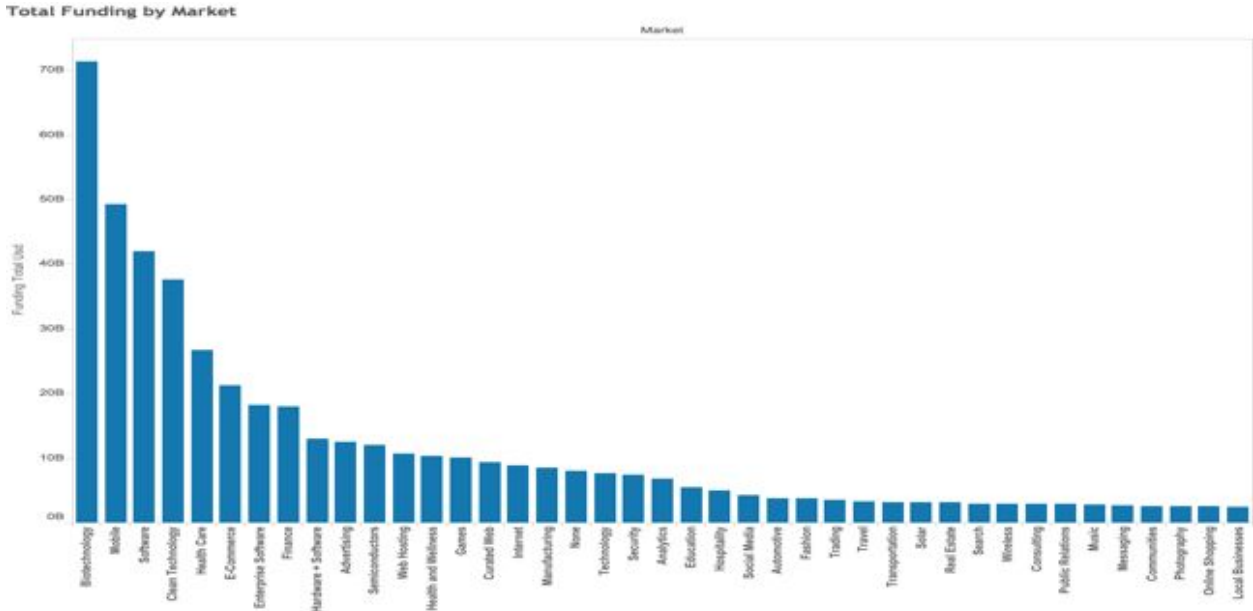


Fig 6.1 Total Funding by market shows clear trends

After looking at the distribution of funds across markets, we wanted to see how funds progressed over time and market. By creating a line chart of cumulative total funding over time colored by market we were able to discover and validate a few key findings (Fig 6.2). Similar to the previous graph, it was apparent that there were key markets which dominated the startup world scene—for example, there is Biotechnology in dark blue with mobile in light blue, software in dark orange, and clean technology in light orange all not too far behind. Secondly, the amount of funding really starts to pick up around 2003. We did further investigation into this stark difference in funding and decided to subset our data from 2003 and on. Finally, the grades of the lines provide insight into the relative importance of each market.

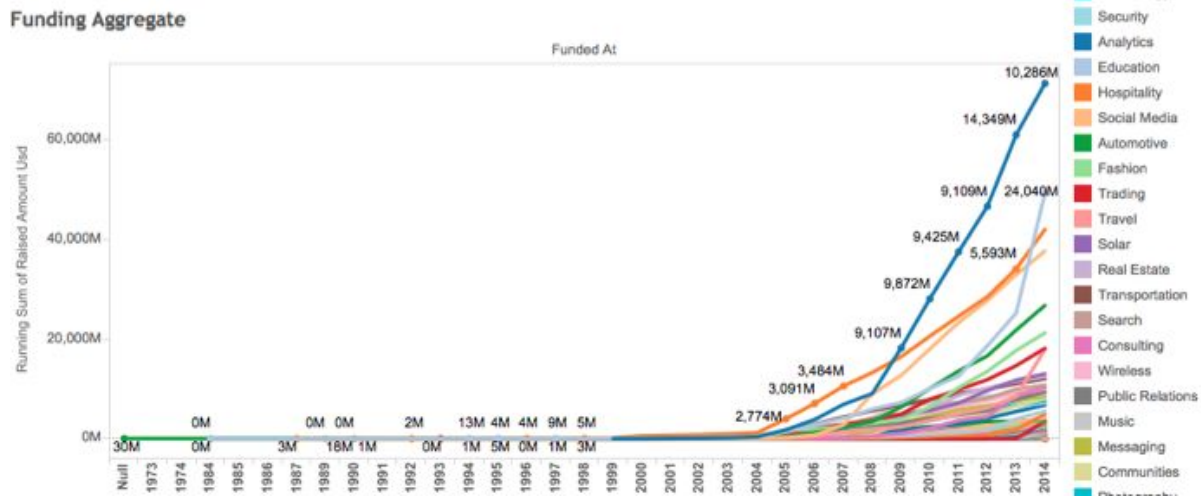


Fig 6.2 Line chart of aggregated total funding by market over time

We also disaggregated the funding by looking at the amount of funding for each specific year over time (Fig 6.3). While the previous visualization gives the impression that Biotechnology continues to dwarf the remaining markets through 2014, we see that disaggregating data tells a different story. In 2014, there is a noticeable drop in the amount of funding that Biotechnology receives. Moreover, since Biotechnology has dominated the scene since 2008, it is difficult to ascertain the relative amount of funding when the data is cumulative. In fact, we can see that the amount of funding given to Biotechnology from 2008 until 2012 completely plateaus. Mobile, on the other hand, receives a huge bump in funding while Software almost catches up with Biotechnology by 2014. This reinforced our thought process in that representing this interplay between market and time would be very important.

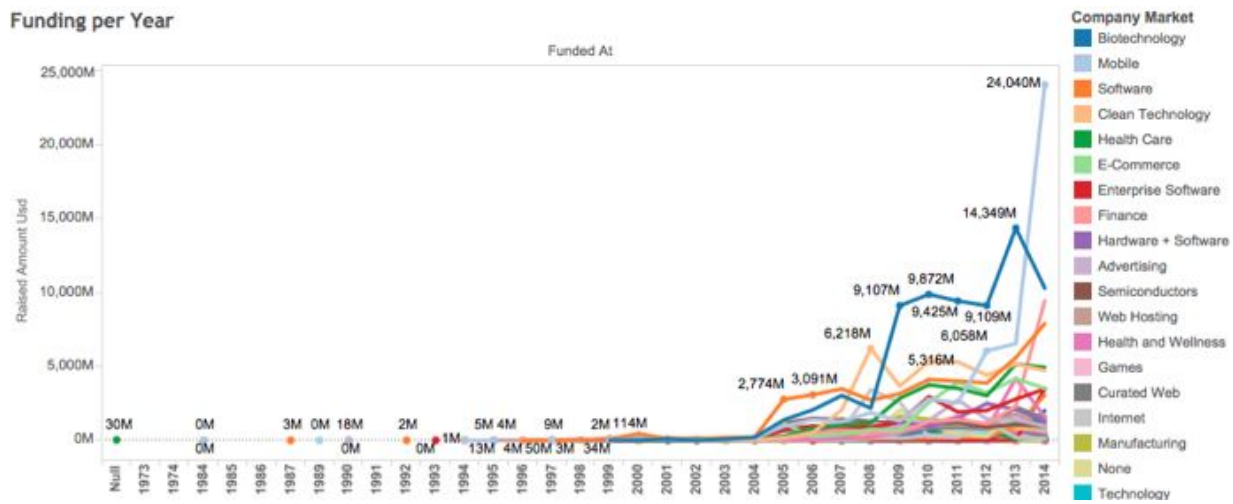


Fig 6.3 Line chart of disaggregated total funding by market over time

After looking at each of the markets, we analyzed funding for startups and venture capitalists over space as well. As we had initially assumed, the USA heavily dominated the funding scene in terms of amount of funding given. The visualization on the left clearly validates this assumption. However, we also wanted to analyze which countries were be investing in. In order to make a clearer deduction, we normalized the data by converting raw funds by country into a percent total by country. These two visualizations directly inspired the second section in our final presentation (Fig 6.4).

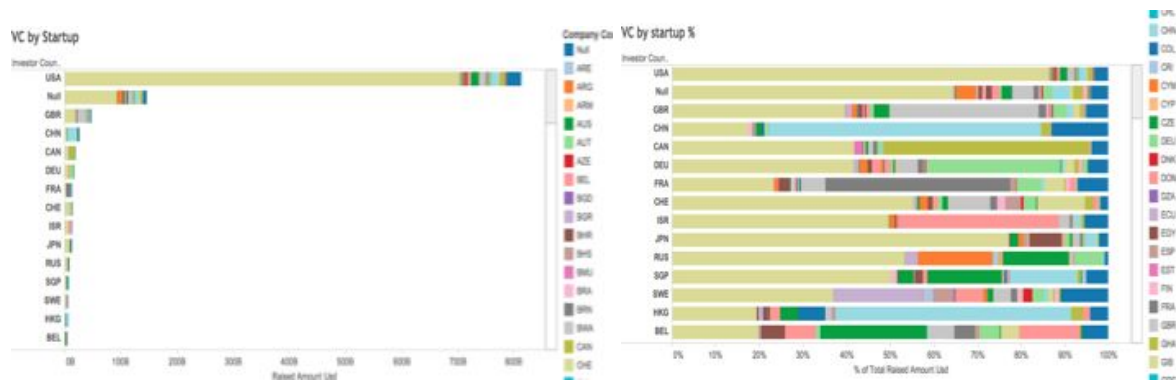


Fig 6.4 Bar chart of raw total funding as compared to normalized total funding

We decided to extend this analysis of space by adding the element of time and market into the picture. Similar to the visualization which analyzed amount of cumulative funding by market, we created small multiples by country. However, the scale of funding greatly distorted the visualizations, as USA heavily dominated the scene. Instead, we decided to normalize the data by making each axis relative to the relative maximum value of each country (Fig 6.5). It is interesting to see which markets different countries choose to fund and in what distribution. USA and Great Britain, for instance, heavily invest in Biotechnology. China, on the other hand, is clearly dominated by E-commerce, while Canada has a more even distribution of funds allotted to various markets.

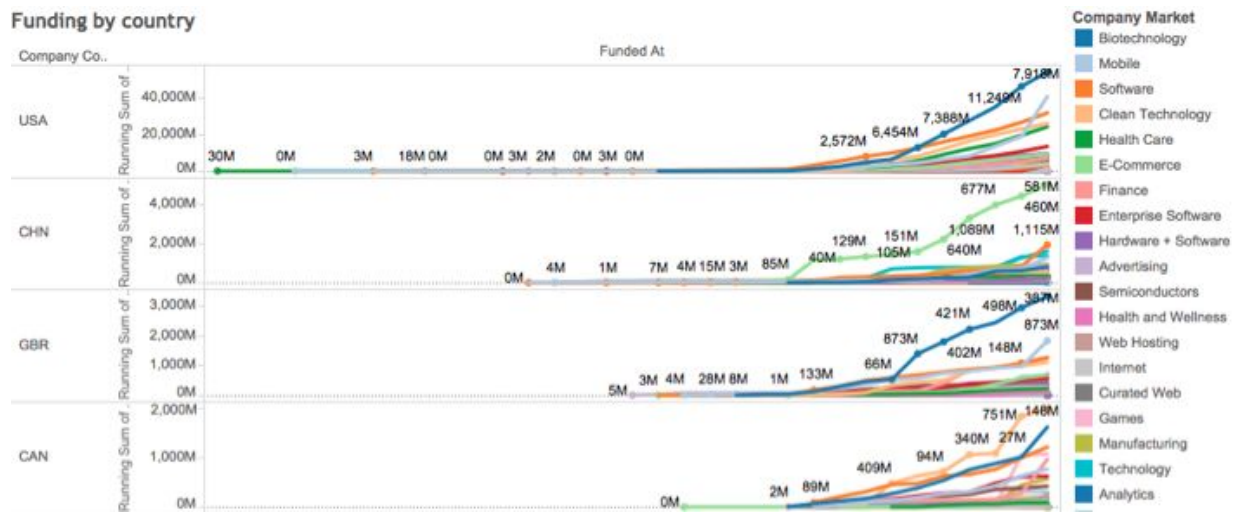


Fig 6.5 Line chart of total funding by country and market over time

The following graph, which we decided not to present in our final visualization, explored the types of funding rounds among startups. For each funding round, we had data on whether the startup was currently operating, closed, or had been acquired. Startups were all funded at least once and at most 15 times. The graph below shows the distribution of the round status among those different funding rounds (Fig 6.6). In the end you do not see any clear bias. Startups who were only funded once versus those which were funded nine times generally have the same status distribution.

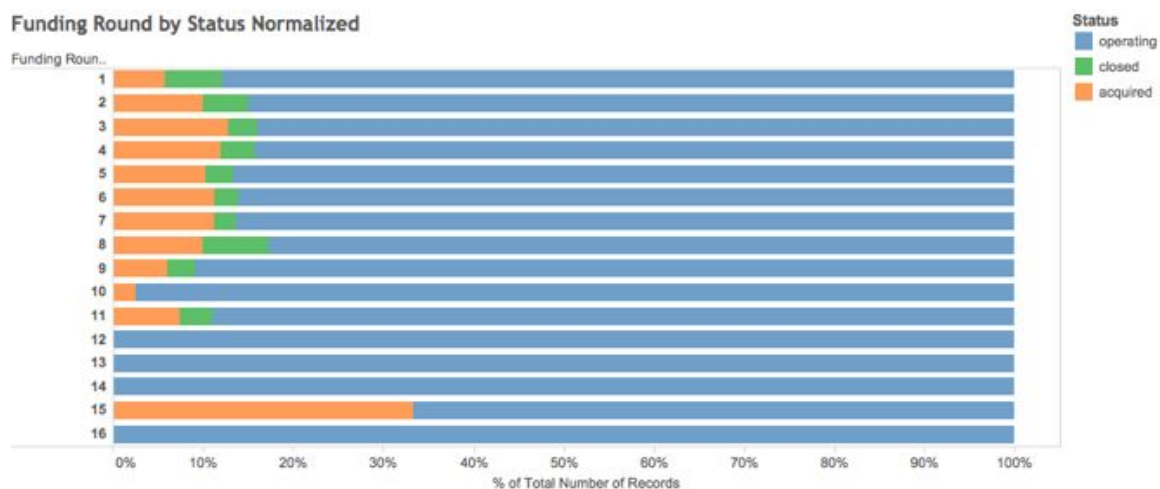


Fig 6.6 Normalized bar chart of total count of funding round by status

This exploratory data analysis was crucial for the development of our final visualization. We decided to focus on the interplay between market, time and space. From our user studies, we deduced that users were most interested in this combination. Showing the multitude of intricate layers behind the dataset and the venture capitalist and startup scene made this problem even more difficult. Thus the final visualization starts very general (as can be seen from the world map) and becomes progressively more detailed, aiming to show different and interesting angles of the data. This initial analysis also demonstrated the nuances of the dataset—namely the concentrated distribution in both market and country. In each graph we had to normalize the data in one way or another.

## 7 Feedback & Future Work

Our usability tests with fellow MIMS students focused on the clarity of the visualization. Our initial charts and premise resulted in complex combinations of graphs and we wanted to make sure that they were understandable by our audience. This involved the correct spacing between elements to understand grouping and sufficient labeling to support comprehension. During the initial design phase, we prototyped a world map with money bubbles and another bubble chart divided by market. The feedback we received was regarding the usefulness of these graphs and to tweak them accordingly.

During the development process, we received feedback from our testers such as:

- Switch the colors of the axis labels to white.
- Move the Unicorn title closer to the graph it represents.
- Shift the axis lines towards the bars or else it looks independent.
- There were too many countries in the percentage stack bar chart, making it difficult to differentiate and follow patterns.
- The different positioning for the bars of all startups and unicorns makes it confusing to discern which markets the unicorns belong to. Users tended to trace a path horizontally to the market name, whereas our intention was different.

In addition, we gathered some additional feedback from viewers at the project showcase event:

- Use median number of startups for the area chart of all markets so that you have a consistent aggregation that provides more accurate comparisons.
- The Y-axis going down on the area chart gave a misleading representation of negative data, when in fact it was median funding amount.
- The colors and order for the list of markets in the area chart did not appear to follow any specific order or pattern.

- In the world map, the word 'rounds' sounds like guns and should be changed to something more peaceful. This is to not give the impression of violence with money bombs as well.
- Tooltips and labels for the countries in the bar chart and dots in the line chart were a natural expectation.

Despite our efforts to make this visualization a deep drilldown into each startup without overwhelming users with too much data, many had difficulties in interpreting the visualization. After careful description of the two bar graphs, it became apparent how these visualizations functioned and why they were important. However, it was not easily ascertained without a comprehensive description. For instance, many users could not deduce which market in the 'All Startups' column matched to which market in the 'Unicorn Startups' column. We had decided to normalize the scales by relative maximum value, as using the raw number would dwarf the column with only unicorns. In this way, the user could compare and contrast how unicorns compared to all startups on a relative scale. However, this again confused the user as to how these values were represented. In the future, we would make greater efforts into making this more readable. An easy solution would be to annotate our graphs much more. A second option would be to create a scatterplot by market with total funding on one axis and unicorn funding on another axis. While we were able to make some configurations for this submission, we will consider the other suggestions as implementations of future work.

## 7 Table of Responsibilities

Project Element	Gracen	Daniel	Pratik	Carlo
<b>Data extraction</b>	10% Csv formatting	30% Lat/Long processing	30% Crunchbase retrieval	30% Data cleaning and shaping
<b>EDA</b>	0%	15% Initial exploration	45% Initial exploration and pattern finding	55% Exploration into other specific questions
<b>Layout Design</b>	80% Conceptualization, color choice, overall layout decisions, text on page	5% Edits & input. Formatted map to fit aesthetics	10% Edits & input. Formatted area charts to fit aesthetics, text on page	5% Edits & input. Formatted linked bar and line charts to fit aesthetics
<b>Questions of exploration</b>	25% All developed together	25% All developed together	25% All developed together	25% All developed together
<b>Animated Map</b>	0%	90% Main implementation	0%	10% Debugging
<b>Stacked bar charts</b>	70% Main implementation	0%	0%	30% Connecting to map data
<b>Shaded Area Charts</b>	0%	0%	90% Main implementation	10% Animation
<b>Linked bar and line charts</b>	0%	0%	10% Debugging	0% Main implementation
<b>Front End and Integration</b>	70% Main implementation - CSS JS integration & debugging	10% Debugging and updates	20% JS variable debugging	10% Debugging and adding updates after presentation edits
<b>Report</b>	25% Project Goals, Overall Design & Layout, Stacked Bar Charts description, Table of Responsibilities, Edits	25% Shaded Area Chart, Data Used, Future Work, Table of Responsibilities, Edits	25% Related Work, Animated Map description, Table of Responsibilities, hosting	25% Linked Bar & Line Charts description, Accomplishing our Goals, Tools Used, Table of Responsibilities