

Startup Investments Analysis (2000-2014)

Final Project Report

BA 631: Visual Reporting and Communication

Submitted to: Shakeel Khan Dated December 16 2022

Group Members

Mario Alcaraz, Andrea Garcia Fernandez

Contents

Introduction.....	3
Business Problem and Questions	4
Data Gathering.....	5
Data Dictionary.....	5
Techniques and Tools	8
Data Cleaning	8
Descriptive Analytics.....	10
Startups Investments Dashboard Key Points	10
Market Investments Dashboard Key Points	10
Investment Types Dashboard Key Points	11
Exploratory Data Analysis/Key Findings	12
Trends	12
Top Locations	13
Top Markets After 2008	14
Top 5 Markets Average Amount of Investments	15
Design	15
Conclusion	16
References	17
Appendix	18

Introduction

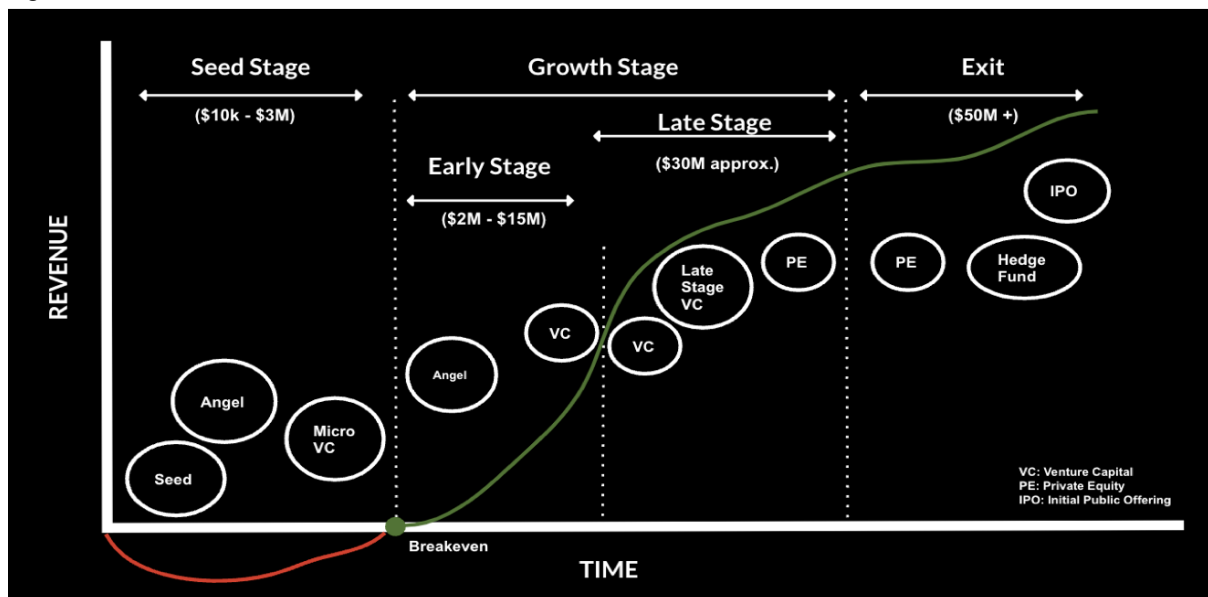
A startup is an emerging company in the first stages of operation. The founder's objective is bringing a new service or product to the market that is believed to have demand. These companies usually do not have an established or fully developed business model. Hence, the first period of a startup is the most challenging. At the beginning, the startup is in the idea validation stage. This is translated to very high costs of operation and small or non significant revenue. This stage is often referred to as 'the death valley'. Therefore, the startup world is known as a high risk high reward when it comes to investors. Different entities and investment firms constantly analyze and seek the next great idea to fund at an early stage. To fully comprehend the complexity of the startup funding cycle and the roles that different investments/investors play in its development, it is crucial to explain the capital power from each investor type and the timing to jump in.

- *Seed Investment:* Can come from relatives, friends or wealthy individuals willing to invest upfront money on the development of the idea/product. The funding amount received on this stage varies but is usually approximate to \$50k.
- *Angel Investor:* Wealthy individual who seeks to fund innovation or new business ideas. An angel investor invests from his own capital and is more likely to invest as a seed investment. However, the range is wider as they sometimes also prefer some proof of success of the business. There is no minimum amount to invest as an angel, but it usually starts at \$10k and can be as high as a few million in some cases.
- *Venture Capital (VC):* A Venture Capital is the type of investor who seeks funding promising companies. Unlike Angel Investors, venture capital is run by managers who collect and invest other people's money along with their own. As a result, venture capitalists seek more proof of potential of the startup and revenue created as demand. Depending on the stage of the startup the investment can be from \$10M to \$30M (Early Stage) or \$30M+ (Late Stage).
- *Funding Rounds:* These rounds happen when a startup has already received initial investment to impulse the idea. The rounds are categorized as A, B and C as the startup develops. Unlike an Angel or an investor this is a funding type that can involve both. The company gives up a percentage of equity in exchange of money based on the startup valuation and other factors.
- *Private Equity (PE):* PE firms invest when a company has built profitable margins, stable cash flow and is able to service a sizable amount of debt. PE looks for mature companies that are struggling due to leadership or

operations, restructures them to improve their performance, and sells them at a profit. Deals can vary wildly depending on the type of business, generally going from \$5 million to billion-dollar deals.

Having a clear understanding of the startup funding types and the roles of the most common investor types. A visual was created for an easier interpretation for anyone interested on this topic. The visual explains the role of each investor and the timeline of the funding cycle, as revenue and time increase different investors join the project. It is also important to highlight that as the startup grows and Private Equity is a potential investor at the Late Stage, the company's growth rate is not as aggressive as on the early stage which is an area more targeted for Angel Investors and small to medium Venture Capitals (Fig. 1).

Fig. 1



Business Problem and Questions

It is not an easy task to keep track of the number of startup companies in the world and even less to find the right startup to invest in. In order to find your next investment, it is important to know what kind of investor you are (Angel, VC, PE), to narrow down your search to a market niche and to understand the numbers around your investment. The analysis done on this paper aims to make this process easier for both investors and entrepreneurs by identifying trends, upcoming products/services, ideal locations and amount of money to invest.

Some of the questions that we have tried to answer are:

- Which markets are receiving the most funding?
- Are there any “hot” locations for each market?
- Where is the money from Angel, Seed, Venture Capital and Private Equity Investors going in terms of market and location?

The analysis is intended to be used by entrepreneurs and investors who are looking for meaningful insights as ‘hot’ markets/locations or trends that enable them to make data driven decisions.

Data Gathering

The data used in this project is a dataset generated by Crunchbase (1). Crunchbase is a data analytics company that works with multiple partners to gather and provide data for specific investment purposes. Their focus is to provide data and/or deliver an analysis of the same. However, startup funding and investor’s transactions are sensitive data and are kept private unless you have professional purposes and a certified investment firm or startup entrepreneur. However, past data is available for exploratory analysis purposes and more. Our Crunchbase dataset was downloaded from Kaggle (2). Kaggle is an online community that allows users to find and publish data sets, explore and develop models in a web-based data science environment, and collaborate with other data scientists and machine learning experts.

The dataset contains information about startup companies around the world and its investments, location, date funded among other characteristics. The dimensions of the initial dataset were 54,294 rows by 39 columns representing different data about each startup.

Data Dictionary

Variable	Data Type	Description
permalink	Nominal	Unique link identifier for the type of entity and name
name	Nominal	Unique identifier for company’s name
homepage_url	Nominal	Unique identifier for company’s website
category_list	Nominal	Startup market category

market	Nominal	Market category which the startup falls in.
funding_total_usd	Numeric	Amount of total funding received to date
status	Nominal	Category of 3 status of the company (operating, acquired, closed)
country_code	Nominal	Country where the startup is located
state_code	Nominal	State code where the startup is located
region	Nominal	Region where the startup is located
funding_rounds	Numeric	Amount of funding rounds that has been through
founded_at	Date	Startup founding date
founded_month	Date	Startup founding year and month
founded_quarter	Ordinal	Founding month corresponding to combination of year founded and quarter
founded_year	Ordinal	Founding year from an interval (1905-2014)
first_funding_at	Date	First funding record date
last_funding_at	Date	Last funding record date
seed	Numeric	Amount of seed investment received
venture	Numeric	Amount of venture capital investment received
equity_crowdfunding	Numeric	Amount of equity crowdfunding investment
undisclosed	Numeric	Amount of undisclosed investment received
convertible_note	Numeric	Amount of convertible note investment received
debt_financing	Numeric	Amount of debt financing

angel	Numeric	Amount of angel investment received
grant	Numeric	Amount of grant received
private_equity	Numeric	Amount of private equity investment received
post_ipo_equity	Numeric	Amount of post IPO equity
post_ipo_debt	Numeric	Amount of post IPO debt
secondary_market	Numeric	Amount of investment generated by secondary market (stock market)
product_crowdfunding	Numeric	Amount of investment raised by product crowdfunding
round_A	Numeric	Amount of money raised on round A
round_B	Numeric	Amount of money raised on round B
round_C	Numeric	Amount of money raised on round C
round_D	Numeric	Amount of money raised on round D
round_E	Numeric	Amount of money raised on round E
round_F	Numeric	Amount of money raised on round F
round_G	Numeric	Amount of money raised on round G
round_H	Numeric	Amount of money raised on round H

Techniques and Tools

The programs that were used to analyze this project are as follows:

1. *Python*: Understanding data dimensions, data types, data cleaning and null values handling to bring a clean dataset to PowerBI.
2. *Google Colab*: This cloud based notebook allowed us to contribute to the data cleaning as a team and see each other's work during data cleaning.
3. *Power BI*: Descriptive and Exploratory Analysis through creation of measures and different visuals to help us understand better our data in terms of distribution.
4. *Tableau*: Creation of charts of key findings and storytelling.

Data Cleaning

1. Column Renaming and Selection

For this step, python and Google Colab were used to create a collaborative environment. First a statistical analysis was run to understand the distribution of the numerical fields. Based on our research question, we noticed how several variables contained irrelevant information. Therefore, we decided to drop irrelevant columns for our analysis purposes. However, renaming some columns (market, total_funding_usd) was necessary as they contained spaces before and after and that caused issues when creating a new dataframe with the specific columns we needed. Once this was completed, the dimensions were 54,294 rows by 19 columns.

2. Data Dimensionality Reduction

Next, an analysis of the amount of *null* values was run to identify the quality of the data. On this step, we found 4856 rows that contained *null* for every column. Hence, we proceeded to drop them and keep 49,438 records.

3. Data Normalization

1. The next step was to transform all of our data to a standardized way as some columns were not formatted in a proper manner. First, we transformed the country_code column null values to nan. Initially, null values were referred to as ' - ' which did not match other columns.

2. Next, we had to transform the data in the `funding_total_usd` column. The values were not written in a way that would make sense (ex. ' 17,00,000 '). Noticing that it had spaces and misplaced commas we had to remove spaces, and commas to transform to integer values and then add back the proper format for currency.
3. For the next cleaning step, we adjusted the funded year to the appropriate data type. Initially, years were represented as floats (2012.0, 1905.0, etc.) which did not make sense as a year must be an integer.

Finally, after running all these steps we were left with a standardized data frame representing null values in a proper manner and correct data types for each column. Then we proceeded to export the clean data frame as a new csv to pass to PowerBI for the next steps.

The remaining columns used for our analysis are found below:

- Name
- Market
- Country_code
- Funding_total_usd
- Status
- State_code
- Funding_rounds
- Funded_at
- Founded_year
- First_funding_at
- Last_funding_at
- Seed
- Angel
- Venture
- Private_equity
- Debt_financing
- Round_A
- Round_B
- Round_C

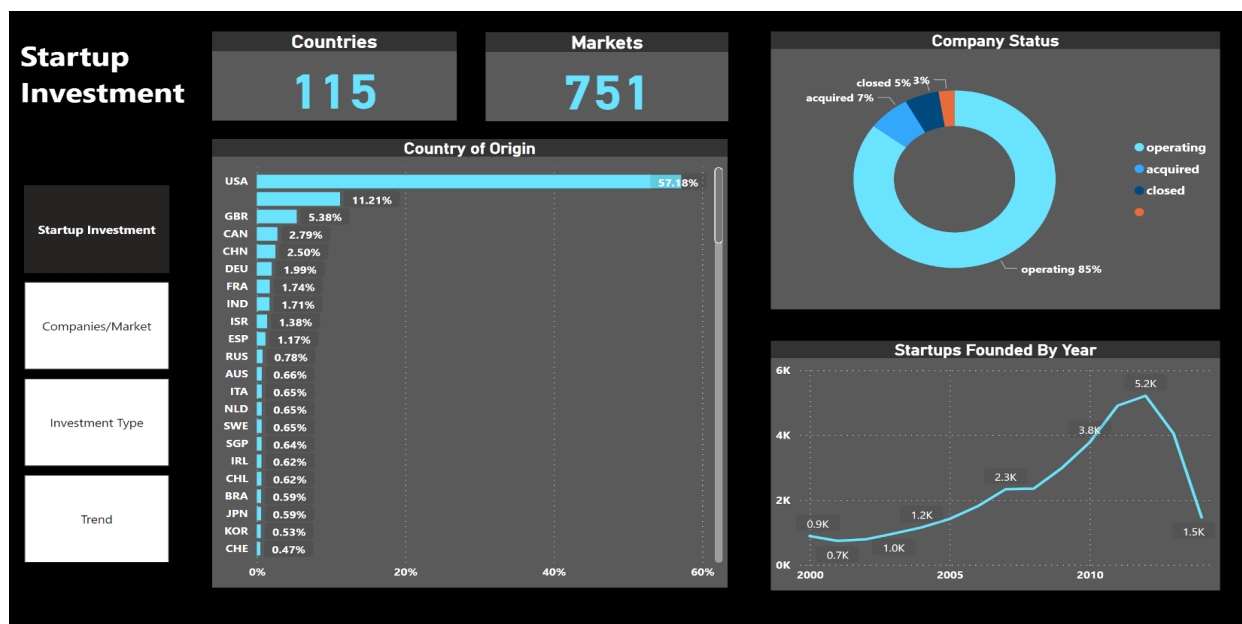
For a more detailed view on the process, please refer to our Google Colab Notebook (3).

Descriptive Analytics (PowerBI)

To make the analysis more relevant, we decided to only keep startup information that went from 2000 to 2014. This was done through data modeling in PowerBI. Multiple visuals were created and divided on 3 dashboards to explain 3 different topics: Startup distribution, Market analysis and investment types analysis.

Startups Investments Dashboard Key Points

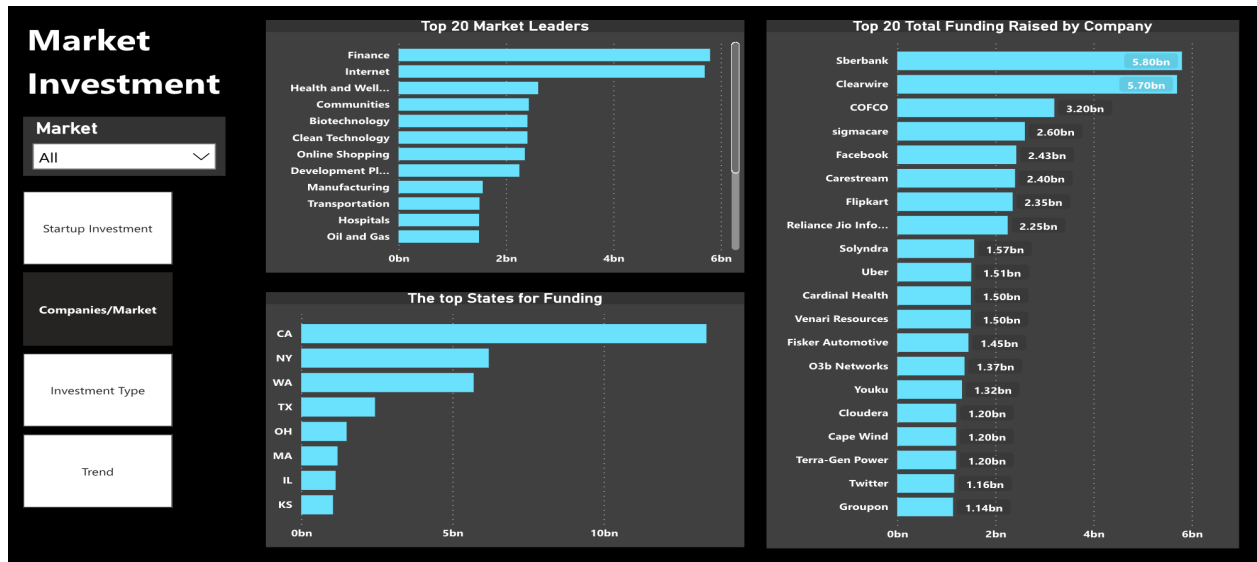
- Data contains 115 unique countries and 751 markets.
- As of 2014, 85% of the companies are still operating, 7% were acquired, and only 5% were closed. Unknown status accounts for 3%.
- Most of the startups are based in the USA(57.18%), followed by Great Britain(5.38%), and Canada(2.79%).
- A slow positive trend can be observed on the “Startups Founded by Year” chart until 2008, where the line flattens out, and then spikes in 2014.



Market Investments Dashboard Key Points:

- The top three leading markets are finance, followed by internet, and health and wellbeing.
- Sberbank(Russian banking and financial services) is the company with the largest investment(\$5.8 billion). Other well established companies such as Facebook, Twitter, and Uber can be observed in the top 20.

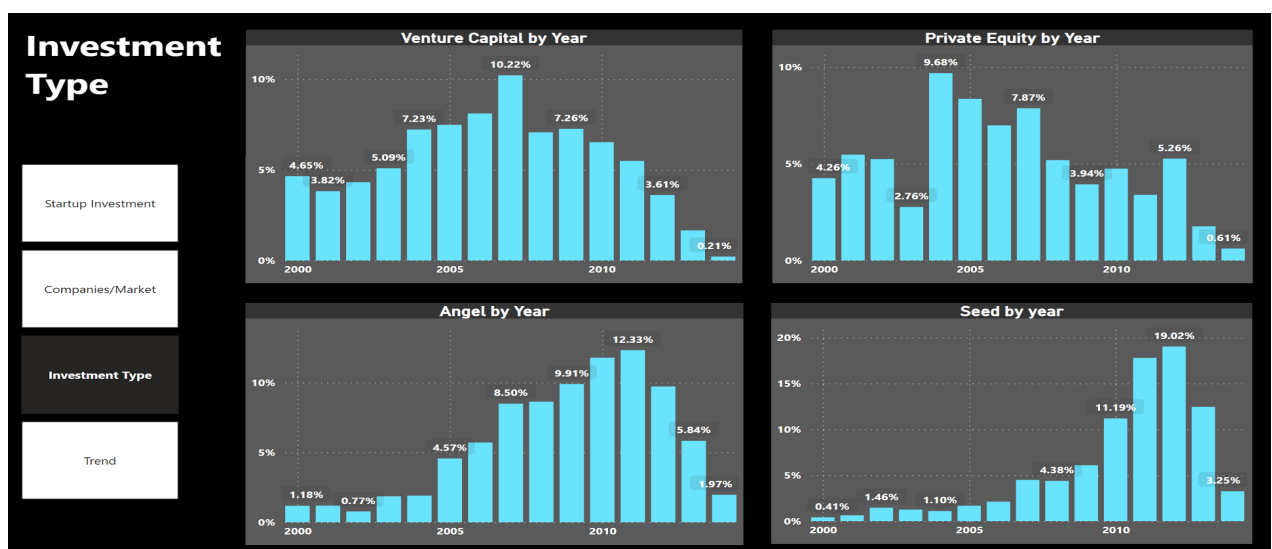
- Most of these companies are located in California, New York, and Washington, seeing how the USA dominates startup investment.
- The dashboard is interactive so that the user can choose the market of preference to see insights.



Investment Types Dashboard Key Points:

This dashboard shows the distribution of the different types of investments. We decided to focus mostly on 4 investment types: Seed, Angel Investor, Venture Capital and Private Equity.

- Both VC and PE follow an overall similar pattern. A positive trend until 2008 and a peak before 2008, followed by a negative one.
- The opposite is observed on both angel and seed investments. The trend started going up after 2007/2008 with a peak after 2008.



Exploratory Data Analysis/ Key Findings (Tableau)

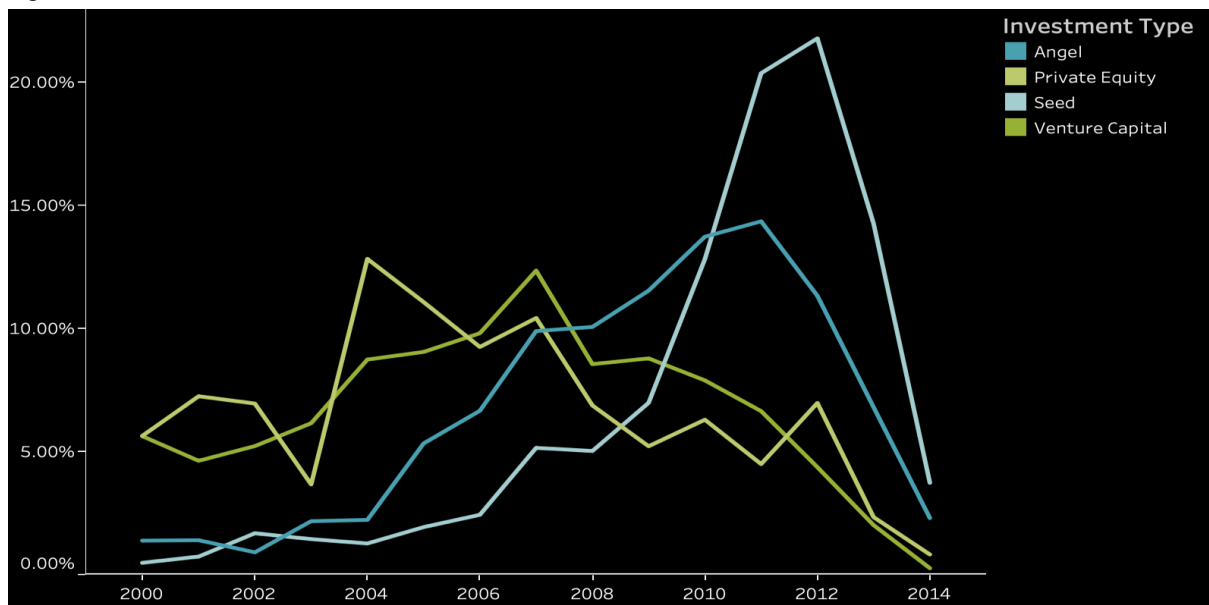
For the final step of our analysis, different visuals were considered until getting to the right one to represent the best according to the values, measures and dimensions.

Trends

Investments trends (2000-2014)

The idea with this visual was to easily identify the different trends followed by Seed, Angel, VC and Private Equity investments on the same page. The decided best visual was a line chart comparing the different investments side by side. Additionally to it, a distinct color was assigned to Angel/Seed and VC/Private Equity. The reasoning behind this is that the first two mentioned have an early stage and seed focus and the last two tend to get involved when there is proof of success or is a mature business. The line chart showed a similar trend line for the grouped investment types, with opposite directions after 2008 (Fig. 2). Further research was made on the economic background to explain this switch. Being the 2008 housing market crash and post financial crisis the reason of the decrease in investment activity for bigger firms and more impulse to new companies (4)

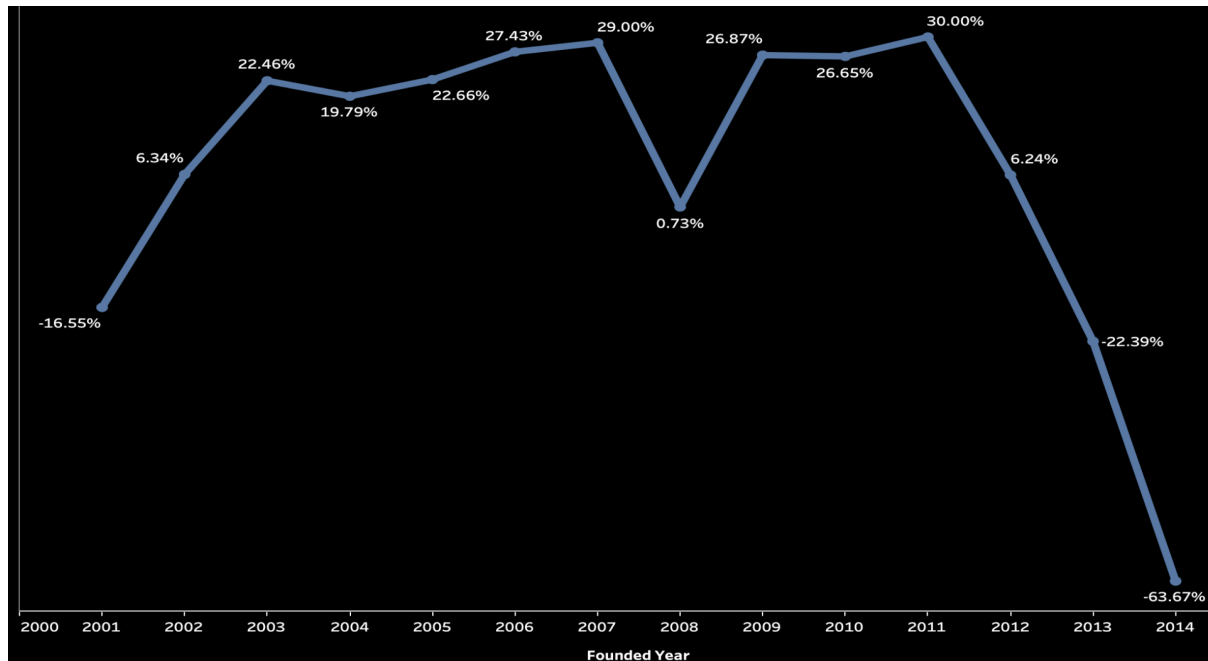
Fig. 2



Startups Creation Rate

Before jumping into conclusions of more startups being created after 2008. Another line chart was created to analyze the start up creation rate compared to the previous year (Fig. 3). Indeed, we were able to verify that after 2008 the rate increased. Giving a good explanation to the increase in Angel and Seed investments in the same time period.

Fig. 3



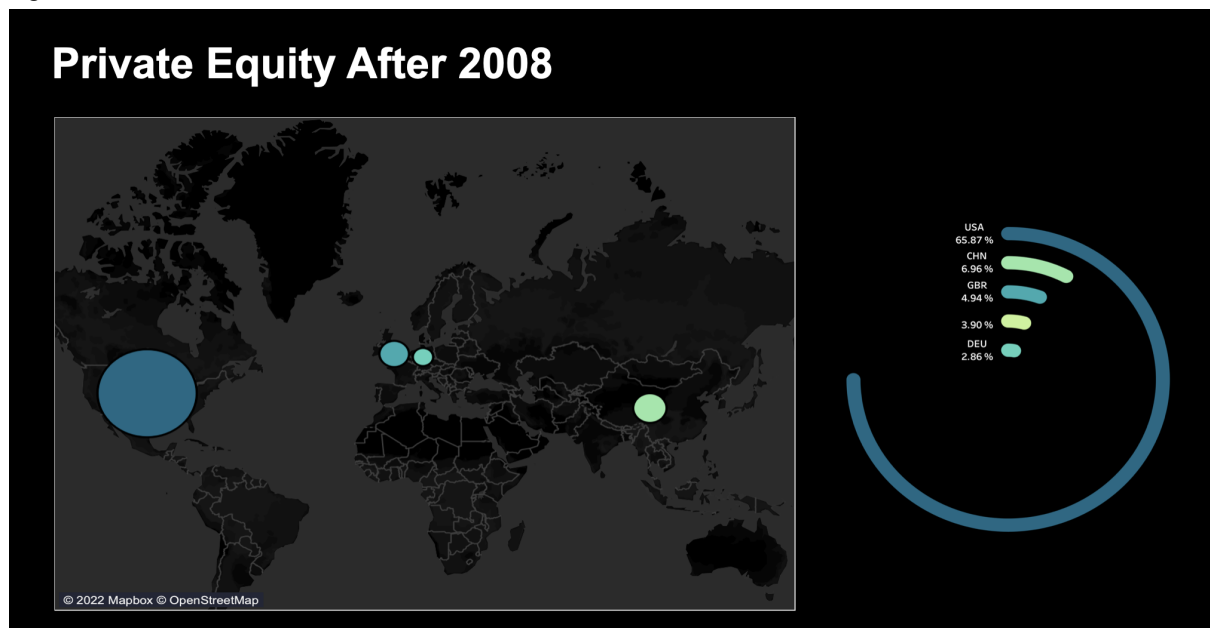
Top Locations

For the next visual, a filter was applied to display the top 5 countries for each Investment type to plot in a map using country_code. Important to mention that this was filtered to be after 2008, with the purpose of seeing where the money was going for each investment type. Additionally, we wanted to show the actual percentage from total for each of the countries. In order to do that, we created a percentage ring visualization.

The percentage ring visualization was created through new measures which assign the percentage value to the percentage value of a circle circumference and calculate for the x and y axis (5). Then we made sure to match the palette colors to the corresponding country to the map view. Finally, the percentage ring and map are placed side by side for an easy understanding of the public (Fig 4). This process was repeated for the four investment types that we were focused on.

Through it, we were able to identify the United States, China, and United Kingdom as the most frequent locations for each investment type. (Rest of visuals in appendix)

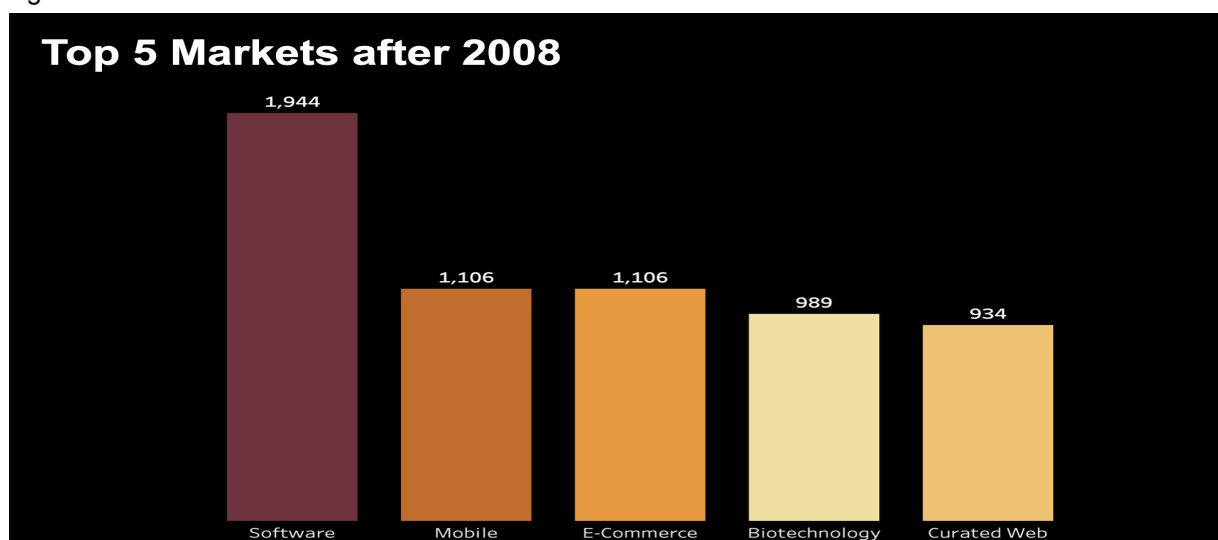
Fig. 4



Top Markets After 2008

As seen before, a distinct trend was shown after 2008. Therefore, a bar graph showing the top 5 markets after 2008 was created. The motive of this visual was to compare the top markets after 2008 to the top markets shown in the descriptive analysis. A distinct color palette from the country's analysis was used to avoid color confusion. The markets with the most startups created after 2008 were Software, Mobile, E-commerce, Biotechnology and Curated Web.

Fig. 5



Top 5 Markets Average Amount of Investments

In our thought process, we realized that the top markets were shown as the amount of startups created. This did not mean that the money distribution was happening in the same order. Therefore, we decided to calculate the average amount of money invested on each of these markets by investor type. The results show that early stage investors such as Angel/Seed were leaning towards a more unexplored market with a higher risk (Biotechnology). On the other hand, later stage investors VC/Private Equity were leaning toward more mature markets such as E-commerce (Fig. 6).

Now, if we stop and think this was very reasonable. E-commerce had been around for longer as the internet started rising in the early 2000's. On the other hand, Biotechnology is still to the date a market that is not considered very mature based on the number of companies that are focused on it.

Fig. 6

Average Investment for Top 5 Markets After 2008					
	Market				
	Biotechnology	Curated Web	E-Commerce	Mobile	Software
Avg. Angel	\$1,547,206	\$771,609	\$1,110,574	\$1,388,220	\$758,761
Avg. Private Equity	\$17,996,732	\$76,692,795	\$78,061,753	\$32,235,275	\$17,603,424
Avg. Seed	\$1,113,696	\$742,655	\$712,753	\$787,533	\$808,597
Avg. Venture	\$13,455,335	\$12,684,904	\$17,749,651	\$9,621,324	\$8,402,183

Design

Lastly, design was applied to all charts. We wanted to go with simple colors that allowed us to highlight important information with colors light green and blue. For that reason, we kept the background color along all dashboards (PowewrBI/Tableau) as black and white color for text. This created an amazing contrast giving us more options of colors to make emphasis on the key points we wanted to transmit and being eye catching to whoever looks at the visual.

Conclusion

To summarize the key findings for entrepreneurs or investors to use.

2000-2014

- Top 3 most funded markets 2000-2014: Finance, Internet, Health-Wellness.
- Most funded companies: Sberbank, Clearwire, COFCO.
- Financial Crisis in 2008.

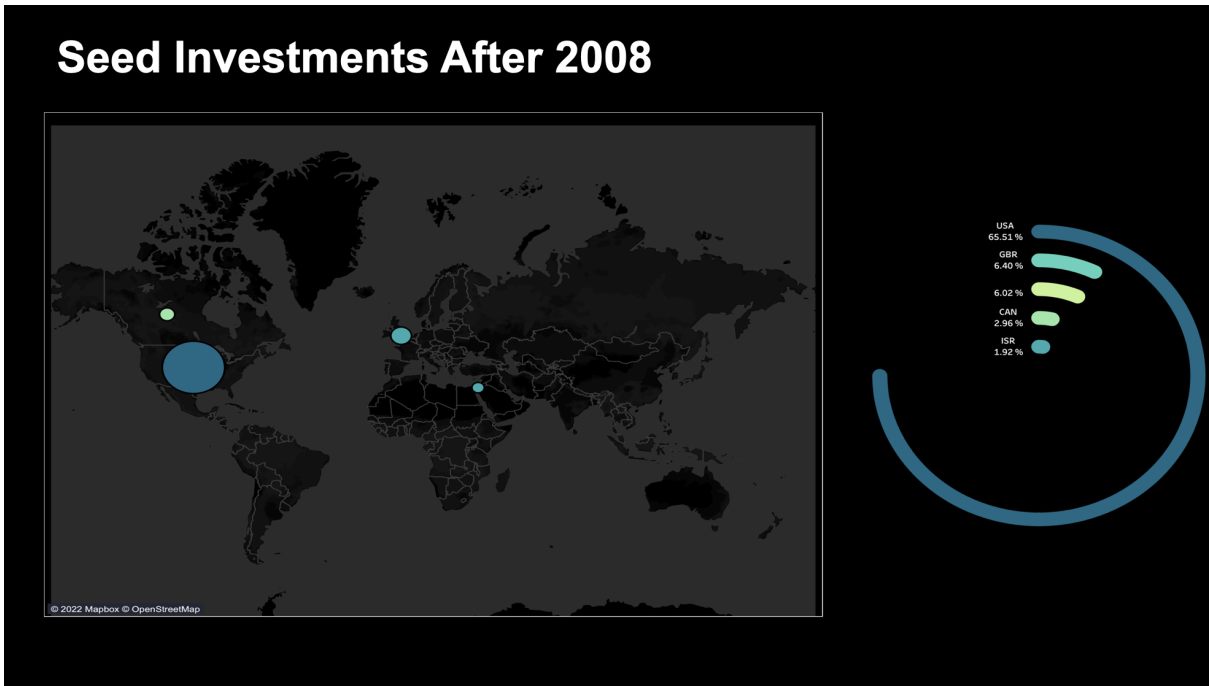
Post 2008 Market Crash Behavior

- Top 5 markets: Software, Mobile, Biotechnology, E-commerce, Curated Web.
- Angel Investors financing Biotechnology.
- Private Equity/Venture Capital financing E-commerce.
- The United States, China and the United Kingdom with the most investment presence in all investments.

References

- 1.- <https://www.crunchbase.com/>
- 2.- <https://www.kaggle.com/datasets/arindam235/startup-investments-crunchbase>
- 3.-<https://colab.research.google.com/drive/1Hrg2xMVRFQJGW7iwkMPtknzBZRxumUJb?usp=sharing>
- 4.-<https://news.crunchbase.com/startups/lessons-from-2008-how-the-downturn-impacted-funding-two-to-four-years-out/>
- 5.-<https://vizzingdata.com/a-different-approach-to-represent-percent-values-tableau-rings/>

Appendix



Angel Investments After 2008

