Investments in Startups: What Factors are Involved?

André Bastos
Nº 56969

Carolina Goldstein
Nº 57213

Rafaela Cruz
Nº 56926

Many say startups are now the change engines of the world, mostly because of the speed of change triggered by technological advances. But have we always paid this much attention to startups? And is this attention reserved to technological startups? It certainly seems like the interest in development and investment in startups has been growing over the years. Thus, in this project, we will study and try to understand what factors are involved in the investment in these companies, how these factors have been changing over the years and see if we can find indicators that explain this change.

I. Introduction

A startup is a company or project initiated by an entrepreneur to seek, effectively develop, and validate a scalable business model. [1]

The evolution from a startup to a successful company usually goes through three basic stages: idea, launch and growth. Throughout these stages, there is a need for external funding, i.e., money to be used for anything business-related, from product development and manufacturing to marketing campaigns and office equipment. Funding is, thus, a major factor that influences the success of a startup, which is supported by the fact that 29% of startup failures are due to lack of finance. [2]

Startups usually raise capital for their businesses through series of investments called funding rounds. These rounds usually begin with an initial seed round, followed by Series A, B, C and beyond. Funds are offered by investors, in exchange for partial ownership of the company or some equity, which is the value of the shares issued by a company. There are several types of investors, including angel investors, venture capital firms, private equity firms and crowdfunding. [3]

Angel investors are individuals who have a net worth of at least one million dollars and use their money to invest in companies. They may invest individually or pool their money with a group and may choose to play a larger role in the startup, requesting a seat on the board of directors. [4]

Venture capital (VC) firms are firms that invest in companies in exchange for equity stakes in the business. VC firms usually play a more active role in startups and receive their return on investment (ROI) from a percentage of profits or private equity. On the other hand, **private equity firms** are firms that invest in startups by purchasing shares for partial or total ownership in the company (i.e., equity). A private equity firm has the capability to buy out a public company, turning it into a private business. [3]

Crowdfunding consists in presenting a startup in a crowdfunding platform and letting people around the world donate money, in exchange for, for example, equity. Thus, crowdfunding opens up the opportunity for investment to the entire public, being significantly different from the private investors described before. [3]

As stated before, these investors usually raise funds for startups in specific funding rounds, which often start with the seed funding round. The **seed funding round** is when investors provide funds before a startup becomes operational. It is used to take a startup from an idea to the first steps, such as product development or market research. Seed funding may be raised from family and friends, crowdfunding, angel investors and venture capital firms that focus on early-stage startups. [5]

The seed funding round is usually followed by **Round A**. This round is focused on startups that have an actual business model that will elicit an immediate profit. In Round A, startups are expected to have a strategy for taking investments and turning them into long-term growth. Round A investors are typically venture capital firms. [3]

If a startup reaches the next funding round, **Round B**, then it has already finished the development stage, i.e., it found its product/market fit and needs help expanding. Round B allows startups to grow so that they can meet the various demands of their customers and compete in tight markets in terms of competition, helping them in the transition to well-established companies. This round's funding usually comes from venture capital firms, often the same investors who led the previous round. [6]

When startups move on to Funding **Round C and beyond**, it is to continue expansion at a higher level. These startups are now highly successful and want to further their success by creating new products, acquiring companies or expanding their reach to new markets or locations. In Round C, private equity firms, investment banks and large venture capital firms tend to be the lead investors. From this stage on, the outcome tends to be an Initial Public Offering (IPO) [1] or to get acquired by a much bigger company. [7]

Note that the studied dataset contains many more types of investments beyond those mentioned in this Section. Thus, for a better understanding of our work, we present a brief explanation of each concept in Section A.

[1]IPO is the process by which a private company can go public by sale of its stocks to general public. Companies can raise equity capital with the help of an IPO by issuing new shares to the public or the existing shareholders can sell their shares to the public without raising any fresh capital. [8]

The interest in startups has been growing over the years. In fact, according to dealroom.co, a comprehensive source of data on innovative companies and venture capital investment, a record of 28 billion euros were invested in European and Israeli startups in 2018. [9]

With that in mind, we decided to study the evolution of startup investments over the years, in order to find some patterns correlating these investments to the type of startup or country in which that startup was born.

We start by listing questions we intend to answer throughout our study, followed by the description of the datasets used to answer these questions. We then look into some studies and visualizations that have been done in this area. In the end, we propose some data visualizations that allow us to answer our research questions.

II. Research questions

In order to study the investments made in startups, we propose several research questions.

- How is the distribution of the accumulated number of startups by market and country?
- How does the invested amount vary between market areas? In which market areas do startups get a larger investment? Does this tendency vary over the years or according to country?
- In what markets are startups more acquired or closed? Does this tendency vary according to founding year or country?
- What factors (market or year) seem to influence the type of investment (whether it is crowdfunding, angel investment, private equity or another)?
- Is there a relation between the number of funding rounds and the success of a startup?

Because the world is a complex system and no event can be seen isolated, we also have some research questions related to other datasets.

- How are investments in startups in a given country influenced by the annual GDP growth? Are expansion or recession years reflected in the total invested amount? What about in the outcome of the startup, that is, whether it is acquired, closed or continues operating?
- Are the different startup market areas affected differently by the economic growth of the country where they were founded?
- How are investments in different startup areas related to trends in those areas? Are there startup areas more influenced by general trends than others? What happens to investment in these areas when the hype is gone?
- What usually happens first, the increase in the number of startups in a given market area or the general interest in this area? That is, does the interest in a certain topic lead to more startups being created around it, or is it the other way around and the appearance of new startups sparks the interest of the general public in a certain area? What comes first - the chicken or the egg?

III. Datasets

A. Startup Investments

The main dataset, available in [10], contains information about startup investments, where each row corresponds to a startup. Its file format is .xlsx and this file contains 49438 rows, each having a value for every 39 of its columns. There are many significant variables in this dataset, ranging from domain-specific to temporal and geo-spatial variables.

Domain-specific variables include the market (whether it is news, games, electronics, tourism, health and wellness, or other markets), status (if the startup has been acquired, is operating, or closed), total funding in US dollars, type of funding (seed investment, venture, crowdfunding, other or even if this information is undisclosed) and how many funding rounds the company has gone through.

Regarding geo-spatial variables, they include country code, state code, region and city. As for temporal variables, the dataset includes founding day, month, quarter and year and day of the first and last funding.

Note that some of the columns were not useful, as they did not provide any additional information to the remaining ones, so we removed them. Thus, our final dataset ended up with 30 columns, the description of which is in Section A.

To enrich our research, we decided to complement this study with other datasets.

B. GDP Growth by Country

This dataset was obtained from the World Bank, at [11]. It contains the percentage of the annual gross domestic product (GDP) growth over the years, by country. Accordingly, the dataset is a .xlsx file with 265 rows, where each row corresponds to a country and each column to a year, ranging from 1960 to 2018.

Since the GDP is a monetary measure of the market value of all the final goods and services produced within a country during a period of time, we will use this dataset to try to explain some patterns found in the main dataset. Namely, we will investigate if the investments in startups in a given country are influenced by its GDP growth and if different startup areas are affected differently by the economic growth.

We will integrate this dataset using the country and year information, so that, for each country in which a given startup was created and each year, we match the economic growth in that year.

C. Google Trends

Google Trends is a large dataset sample from Google's search data. It allows us to measure interest in a particular topic across search, from around the globe, right down to city-level geography. [12]

We can select a topic and obtain a normalized measure of the number of times that topic was searched using Google's search engine. Google Trends provides data for every month, from 2004 to present day, at country or global level.

Thus, we pretend to use Google Trends to try to explain some patterns in the startup investments dataset. For example,

if the investments in some startup market had a growth in some years, we can try to relate that growth with the public interest in that market.

We will integrate this dataset using market and temporal information (month and year), so we can relate the trends in some topic with the evolution of investments in startups of that topic over time.

IV. State of the art

In this Section we will present some visualizations and claims made in other studies related to this subject. This is important to help us understand what studies have already been done in this field and what visualizations work best to understand our data.

This Section will be divided into three subsections: in the first subsection, we will present claims and visualizations regarding startups in general; in the second subsection, we will analyse some studies about startup failures, as this is a common subject in this field; lastly, we will focus on some studies that correlate startups and entrepreneurship with the economic growth in a given country.

A. General Startup Visualizations

In this subsection, we will show some visualizations and claims made about startups in general and startup funding. Note that we could not find interesting scientific papers about this subject, as most of them focus on very specific and complicated details about startups and funding, which is not the focus of our work. However, we did find many claims and visualizations in websites, that would be interesting to corroborate with our data.

Using a dataset similar to ours, Susan Li [13] proposed some simple visualizations to study investments in startups by market, location and type of funding (Fig. 1, 2 and 3).

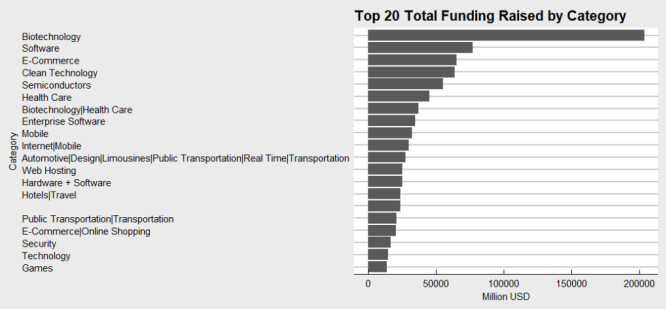


Fig. 1: Capital invested in startups by market category. Obtained from [13]

These visualizations allow us to draw some conclusions. For example, we see that the market where investors provide more funding is Biotechnology, with a large difference from the other markets. We can also see that the region where startups get more funding is San Francisco Bay Area and that venture capital firms are the investors that provide more capital. However, these visualizations are very simple and do not allow us to answer more complicated questions like the ones proposed in Section II.

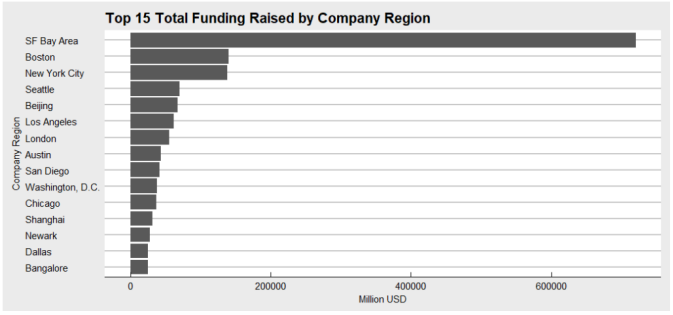


Fig. 2: Capital invested in startups by region. Obtained from [13]

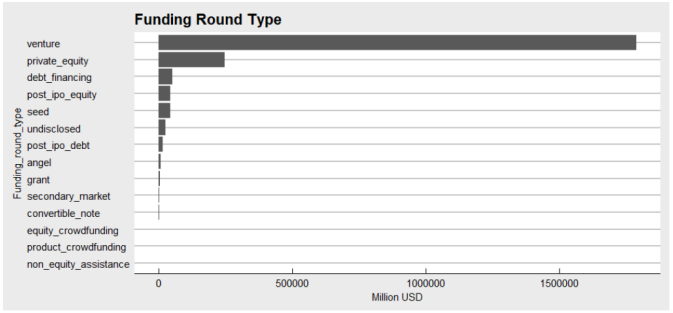


Fig. 3: Capital invested in startups by type of funding. Obtained from [13]

The same author also published a visualization in [14], where we can see the capital raised by funding for each region throughout the years (Fig. 4). The user can visualize the years, from 1982 to 2015, in the columns and the regions in the rows. The user can also filter the regions, showing only the ones belonging to a desired country. For each combination of city and year there is a marker, representing the amount of raised capital, that varies in size proportionally to its value. When hovered, the absolute value of the raised amount is shown.

This visualization allows us to study the evolution of funding throughout the years in a given region. The fact that it is interactive lets us draw more complex conclusions than the previous visualizations.

The website Fundz [15] also made available some visualizations about a startup funding trends dataset, containing 17,459 funding rounds across the U.S. between 2015 and 2017, categorized by city, state and industry. Although the data is about the U.S. only, the type of information to be shown is similar to ours, giving us some insight about what we can expect to take from our own dataset.

Fundz started by representing to total capital raised in funding rounds by state. In order to normalize for the inherent bias towards states with larger populations and big cities, they calculated startup funding per person in each state by dividing the total money raised by the state's population, obtaining a map with the startup funding per person by state (Fig. 5).

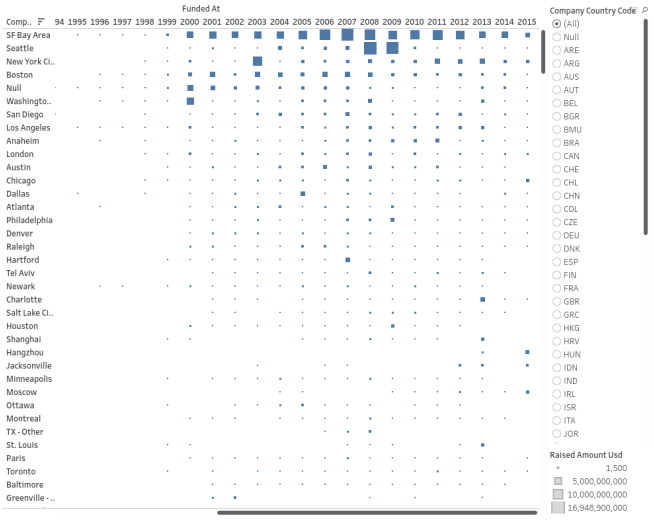


Fig. 4: Capital raised by funding per year and region. Obtained from [14]

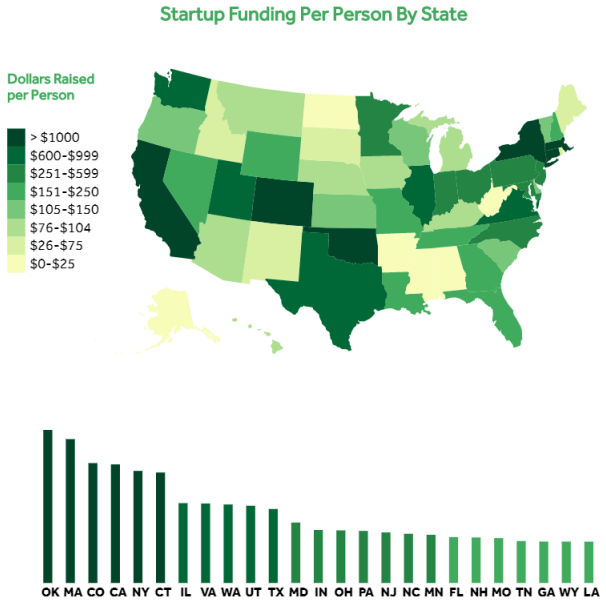


Fig. 5: Capital raised by funding per person in each state of the U.S. Obtained from [15].

Observing Fig. 5, we see that the states where startups get more funding by person are Oklahoma, Massachusetts, Colorado, California, New York and Connecticut. However, this type of visualization (the choropleth map) has some problems, because the most interesting values may be concentrated in densely populated areas and less interesting values may be by spread out over sparsely populated areas. For example, California has a high startup funding per person, but this value may be very high only in certain regions, like San Francisco Bay Area and Los Angeles, and very low in the remaining regions.

Another visualization that held our interest was a map of the U.S. that shows circles for every city where money was raised (Fig. 6). The size of the marker is proportional to the amount of funding raised. The hue of the circle can be red or blue, for non-tech and tech categories respectively. As the user hovers over the map, a toolbox shows the total amount raised in the location, the amount for tech startups and also the amount for the non tech ones.

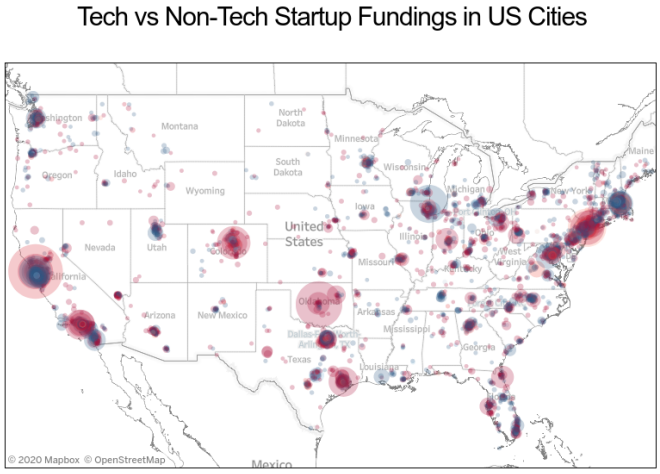


Fig. 6: Capital raised for tech and non-tech startups by city. Obtained from [15].

Observing Fig. 6, we notice that there is a lot of overlapping between the markers, which could be a problem as it could be hard to click on some of them. Nonetheless, the fact that the hue is not totally opaque allows us to understand the data and shows us the states where there are a higher concentration of cities where there is funding for startups.We can see that while tech startups dominate in some major cities, a large portion of the rest of the country is pursuing startup projects outside of the technology sector.

Another visualization that caught our attention was one that related the amount of raised funds for each startup (Fig. 7 and 8). Having the amount of raised money in the vertical axis (that does not seem to be linear) and the companies alongside the horizontal axis, every round marker represents a startup. These are chained together and form a graphic line, showing the evolution of the startups and their funding throughout the years.

When hovered, the markers also give us the name of the company, how much money it raised in funding and, if it is for sale, for how much. These can be filtered with the buttons above the visualization. The buttons represent each market area and once the user selects an area, the visualization only shows companies for that specific market, adding to the interactivity aspect. In Fig. 7 and 8, we present an example for the Computer and the Biotechnology market, respectively.

This visualization is very useful, giving us different kinds of information, including how many companies were funded in a given market, how was the distribution of funded companies

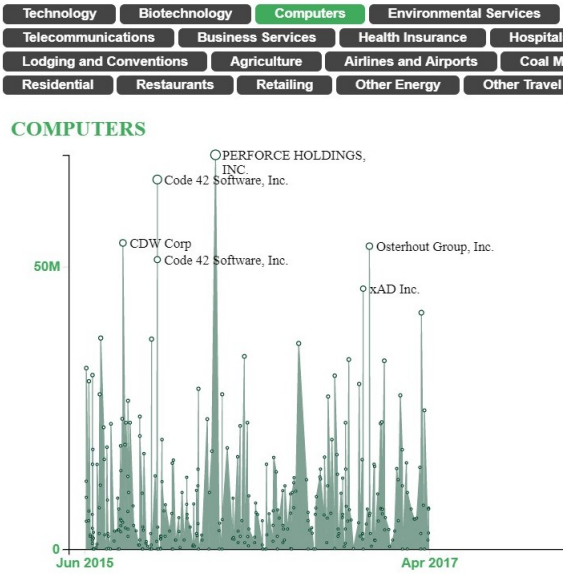


Fig. 7: Amount of money raised for each company on the Computer market. Obtained from [15].

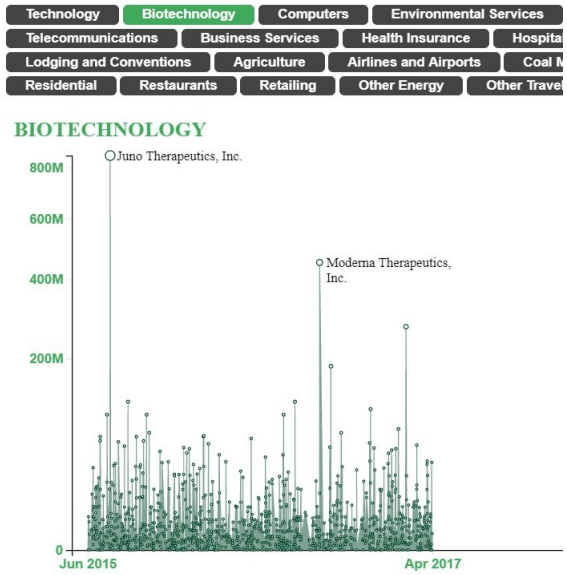


Fig. 8: Amount of money raised for each company on the Biotechnology market. Obtained from [15].

over time and how much money each company received in funding.

In many of the previous visualizations, we saw that the total capital invested in startups can reach several billion dollars. But what about the amount invested in each funding round mentioned in Section I? And by each investor? According to [3], angel investors usually provide between $25,000 and $100,000 in a funding round, and venture capital firms $7 million to $10 million. The average funding amount for a Series C round is $50 million, and while the average Series A in 2010 was $4.9 million, by 2017, it reached $12.1 million. On the other hand, Series B can generate $7 million to $10 million. [16] [17] This is also something that we can confirm using our dataset.

B. Startup Failures

Our dataset contains information about the status of a startup, i.e., if the startup is still operating, was acquired or already closed. It would be interesting to understand what factors are involved in these outcomes.

Many entrepreneurs and business gurus have tried to identify the most crucial factors that determine whether a startup will be successful or not. Although it is difficult to accept most of the found information as general truths, especially as there are around 137,000 businesses being brought to life everyday, it can give us a glimpse into what to expect. [18]

The funding strategy often comes as number one concern, but it is not alone. The timing of a startup's fundraising is one important factor. In fact, Forbes names the time of year of the raising as one of the key factors, together with location, market trends, availability of capital and appetite for startup investments. [19]

Regarding startup failure, about 90% of startups fail and 10% even within the first year. [20] Interestingly, it seems that across different industries, startup failure rates are not very different, with information startups leading the failure rates. [21]

The website Failory created an interesting dashboard made especially to analyze startup failure data. [20] To do this, the data about failed startups collected in the Startup cemetery [22] was loaded into a visual Cumul.io dashboard (for more information on this check [23]). The data used is about startups that were created between 1994 and 2016, from various market areas. The user can choose to see the visualizations only for certain years of creation or failure or for certain markets.

To analyse the distribution of startups per country, Failory presented a map with the number of startups mapped to color (Fig. 9).

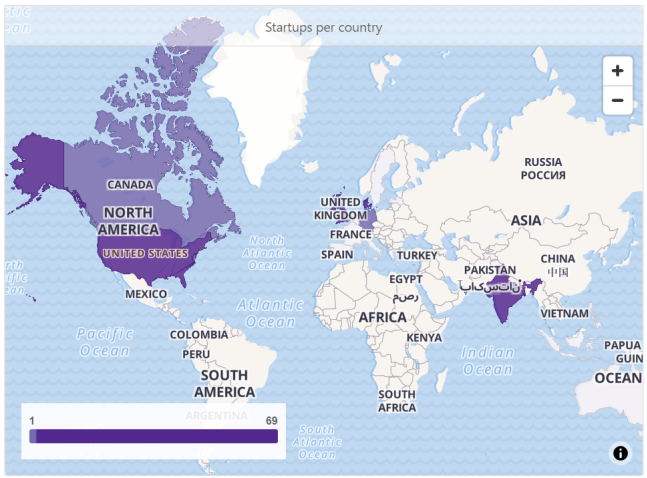


Fig. 9: Number of failed startups per country. Obtained fom [20].

Observing Fig. 9, we see that the legend explaining the mapping to color is not very explicit. However, this map is interactive, so it shows the actual number of startups when we hover over a country. We see that, in the dataset, the U.S. is the country with the highest number of failed startups.

Failory also makes available the visualization of the number of failed startups for each market area (Fig. 10). Each circle represents a market area, showing inside its corresponding name and quantity of startups. When hovered, a toolbox appears detailing the already mentioned information and the percentage related to the total number of startups, instead of the already shown absolute value.

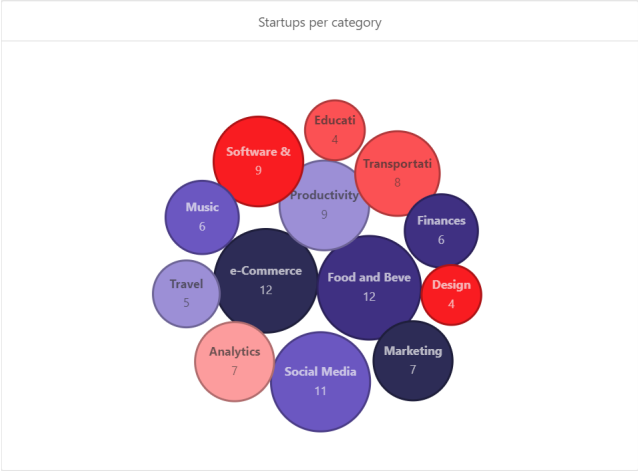


Fig. 10: Number of failed startups per market category. Obtained from [20].

We see that each market has a color, but no information is given to what this coloring may be referring to, as it does not seem to be about the shown values at all.

What also seems to illustrate the absolute value of the quantities is the size of the circle, giving the user the opportunity of a very effective comparison between different areas.

This visualization has a quite high effectiveness overall, but does not seem to be a highly expressive one. This may be due to the fact that few information is revealed once the user hovers over an item, but may also be due to the apparent non existing usefulness of the coloring.

Lastly, Failory also presents a visualization showing the cause of failure for the startups in the dataset (Fig. 11). Again, the color of the rectangles in this visualization does not seem to have a specific meaning. The visualization allows us to conclude that competition and a bad business model were the main causes for the failure of the studied startups.

Note, however, that Fig. 11, as well as the other visualizations presented in this subsection, is very simple, allowing us to answer only very simple questions.

C. Startups and Economic Growth

Using a second dataset, containing the GDP growth throughout the years by country, we plan to study how investments in startups are influenced by this factor. To our knowledge,

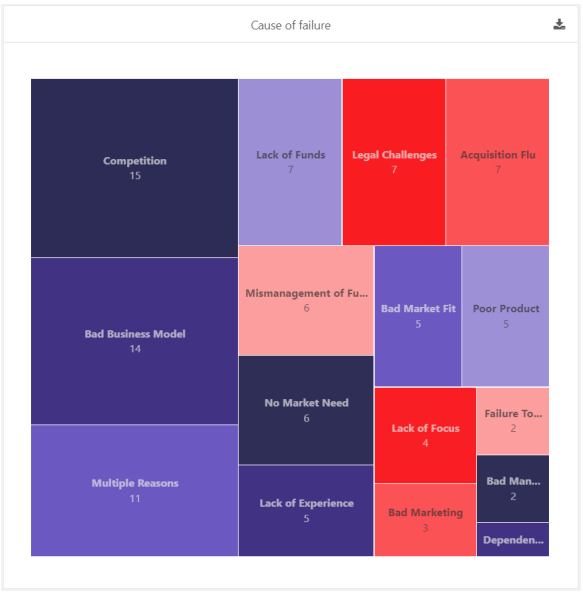


Fig. 11: Cause of startup failure. Obtained from [20].

this has not been done yet. However, the opposite, i.e., the influence of startups in economic growth, has been approached by several studies.

The authors of [24], [25] and [26] all reached the conclusion that entrepreneurial activity (including the development of startups) stimulated GDP *per capita* in high-income countries while having a negative effect in middle/low-income economies.

This may be explained by the fact that, in most high-income countries, significant support is provided to entrepreneurs in the form of business development courses through universities and other training systems, government grants and policies that promote and support entrepreneurship, and also through established institutions which can provide funding, such as venture capital firms. These training facilities, grants and institutions are typically lacking in middle/low-income countries, which may negatively impact on entrepreneurial attitudes. Furthermore, in middle/low-income countries, the entrepreneurial activity is dominated by individuals who are forced into starting their own business due to a lack of employment, and is typically not highly productive. Thus, entrepreneurs in high-income economies may be more innovative, internationalised and growth focused than entrepreneurs in middle/low-income economies, which may cause the observed effect. [26]

These conclusions will be useful to explain some of the results obtained with our dataset. Note, however, that the authors studied entrepreneurship activity in general, which includes other aspects beyond the development of startups. Furthermore, the startups contained in our dataset do not correspond to all startups created in a given year. Thus, our results may not always be in accordance with the results obtained by the authors.

V. Proposal

A. Startups

In this Section, we will try to answer the research questions proposed in Section II, as well as confirm some of the statements presented in Section IV.

Before answering our research questions, we decided to study the seasonality of the startup distribution, that is, how the distribution of the number of startups founded changes according to month and year. For this purpose, we created the dashboard presented in Fig. 12. This dashboard shows, for each row, the founding years, varying from 1998 and 2014, and for each column the founding month. On top of each month, one can see the total number of startups founded in each, and at the end of each row there is the total number of startups founded in the corresponding year.

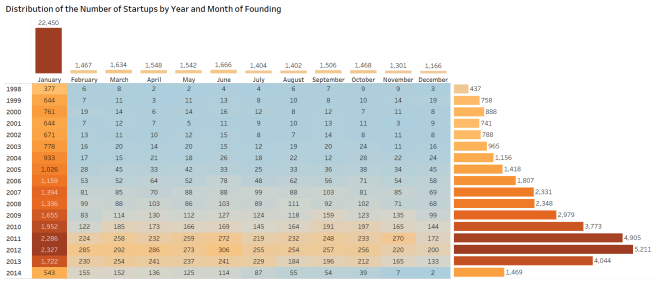


Fig. 12: Dashboard 1: Distribution of the number of startups by founding year and month.

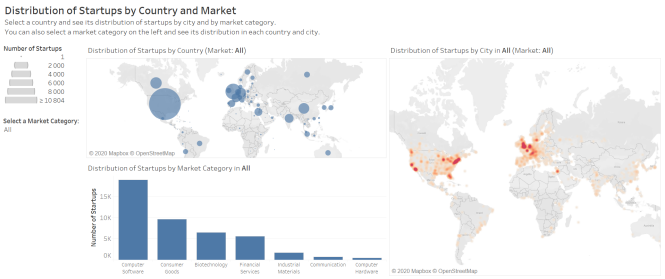
The total number of startups seems to increase throughout the years, as expected. There seems to be quite a drastic decrease in 2014, but this is more likely to have been caused by the lack of records as this is the last year present in the dataset, than an actual decrease in the number of startups.

To our surprise, the majority of startups seem to have first been founded in January, across all years. After some research, we found that there is in fact a general bigger impulse for entrepreneurs to start a new business at the beginning of the year, in relation to the "New Year new beginnings" atmosphere.

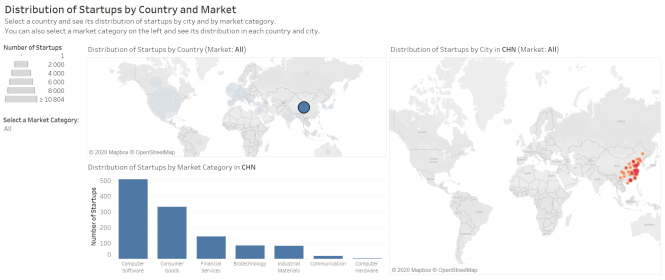
However, this could also be due to some errors in the creation of the dataset, rather than a real phenomenon. The creator could have chosen to register a startup for which only the year is present as having been founded in the first day and month of that year, possibly explaining the discrepancy between the records registered in January and the other months.

In order to study the distribution of the number of startups by market and country, we started by grouping the markets in the dataset into seven larger groups: Consumer Goods, Financial Services, Communication, Industrial Materials, Biotechnology, Computer Software and Computer Hardware. The markets that did not fall in any of this categories were placed in a category named Miscellaneous.

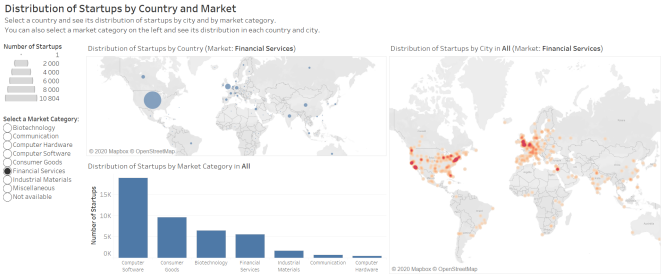
Then, we created a dashboard that shows the total number of startups by country, on the upper left panel, by city, on the right panel, and by market category, on the bottom left panel (Fig. 13). Note that, in the right panel, we chose to represent the startups in each city using a density map, in order to better understand in what city there is a larger concentration of startups.



(a) Without filters.



(b) Filtered by country: China.



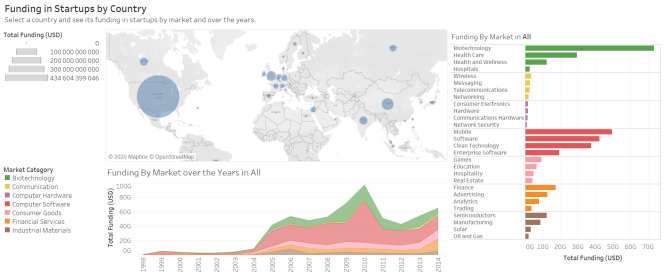
(c) Filtered by market category: Financial Services.

Fig. 13: Dashboard 2: Distribution of the number of startups by country, city and market category.

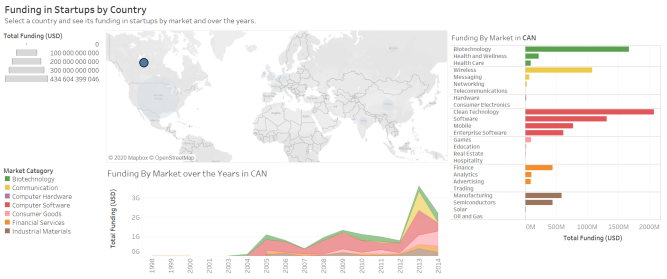
Observing Fig. 13a, we see that the market category with the highest number of startups is Computer Software, followed by Consumer Goods and Biotechnology. The U.S. is the country with the highest number of startups, which seem to be more concentrated in the states of California, New York and Massachusetts.

This dashboard is interactive, allowing the use of two different kinds of filters. By selecting a country on the upper left panel, the user can choose to see the distribution of the number of startups in that country by market and city. As an example, in Fig. 13b, we selected China and saw that startups are concentrated in cities on the east of the country. Moreover, the user can select a market category by using the filter on the left and see the number of startups of that market in each country and city. In Fig. 13c, we show an example, where we selected the market category Financial Services.

We then created a dashboard to study the total funding by

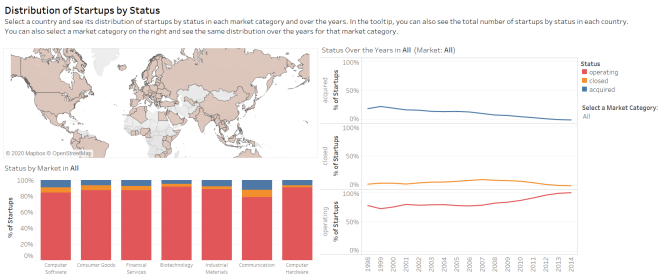


(a) Without filters.

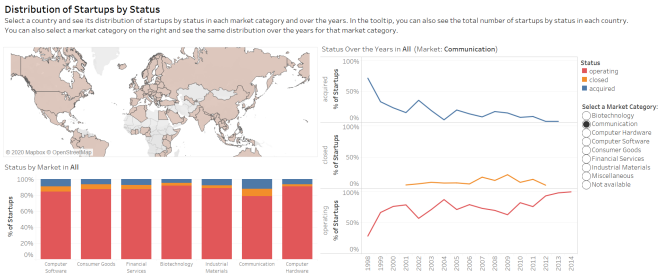


(b) Filtered by country: Canada.

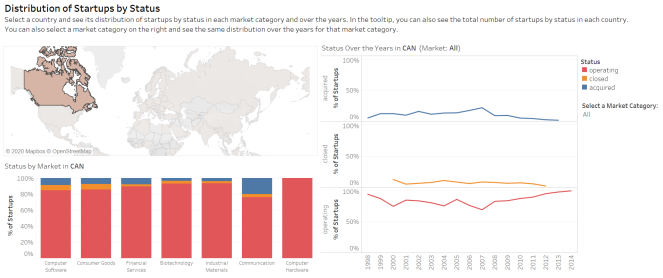
Fig. 14: Dashboard 3: Total funding in USD by market category and country.



(a) Without filters.



(b) Filtered by market category: Communication.



(c) Filtered by country: Canada.

Fig. 15: Dashboard 4: Percentage of operating, acquired and closed startups by country and market category.

market in each country and over the years (Fig. 14a). Each market category is represented by a color, the caption of which is shown on the left.

On the upper left panel, we have the total funding in startups in each country. We can see that the U.S. is the country where more money is invested in startups, although this is also because the number of records is higher in this country.

On the bottom left panel, we have the distribution of funding over the years for each market category. We can see that Computer Software is the market category that received more funding throughout the years, having a peak in 2010.

To compare to investments made in different market categories, we also show these categories decomposed in specific market areas on the right panel. We can see that, even though the category where the most money is invested altogether is Computer Software, the Biotechnology market (included in the category Biotechnology) is the specific market that receives more funding.

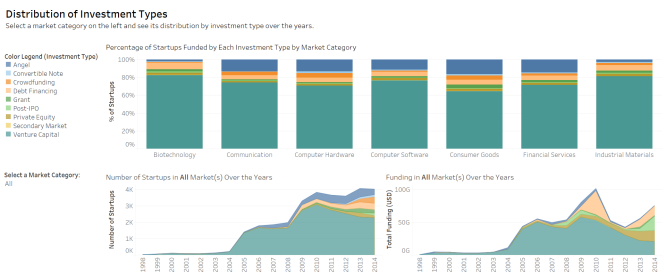
This dashboard is also interactive: we can select a country and check the total funding in each market and over the years. For example, if we select Canada (Fig. 14b), we see that the market where the most money is invested is clean technology. The market category Computer Software has been the category that receives more funding since 2004, with several peaks throughout the years (e.g. 2005, 2009 and 2013). However the funding in Communication also reached a peak in 2013, with the total funding being close to the leading category.

To study in what markets startups are more acquired or closed, and if that tendency varies according to founding year or country, we propose the dashboard presented in Fig 15a. In the bottom left panel, we can see that, overall, there does not seem to exist a relation between the status of a startup and its market category. Communication seems to be the market category where more startups are acquired and closed, but the difference between this market category and the others is not significant. This corroborates with the claim mentioned in Section IV and detailed in [21], stating that startup failure rates do not seem to vary across different industries.

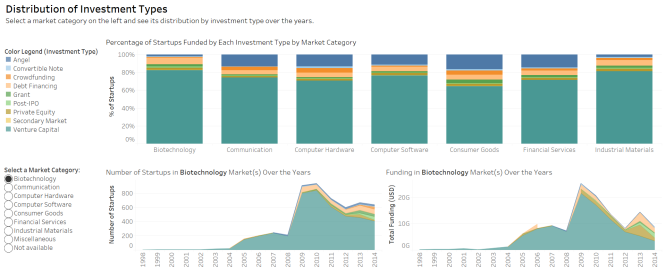
Moreover, we see that the majority of the startups in the dataset are still operating. This goes against the statement presented in Section IV that 90% of the created startups fail. One explanation for this is that our dataset might not have been created using a random and representative sample of startups. Instead, the author of the dataset might have focused on startups that were successful and are still operating.

By selecting a market category on the right, we can see how this tendency varied throughout the years for a given market. In Fig. 15b, we can see that a lot of startups of Communication founded in 2002, 2005 and 2008 were acquired and a high percentage of startups of that market category founded in 2007, 2009 and 2011 were closed.

We can also select a country on the upper left panel and

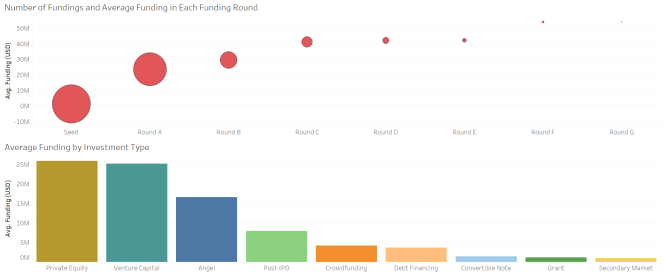


(a) Without filters.

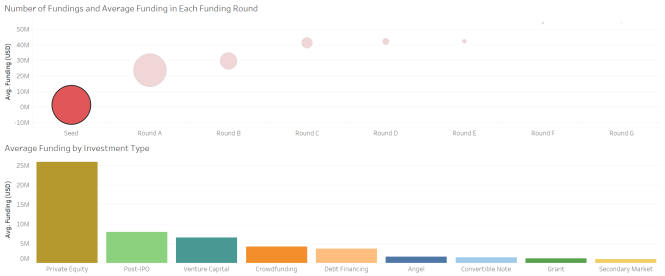


(b) Filtered by market category: Biotechnology.

Fig. 16: Dashboard 5: Investment types by market category and throughout the years.



(a) Without filters.



(b) Filtered by round: seed.

Fig. 17: Dashboard 6: Average funding by funding round and by investor type.

see the status of startups by market category and year. For example, in Fig. 15c, we selected Canada and saw that the most acquired startups are startups in Communication and startups founded in 2007. For the startups that closed, there does not seem to be a significant variation throughout the years or according to market category.

To study the factors that influence the type of investment in startups, we started by creating a calculated field with types of investment (venture capital, debt financing, private equity, grant, angel, post-IPO, crowdfunding, convertible note and secondary market). The meaning of these investment types is explained in Section A. Using this calculated field, we created a dashboard (Fig. 16) that focuses on this aspect of the data. On the left, there is the caption for the color meaning, each of them representing a different type of investment. This is important as all three visualizations in this dashboard depend on these.

On the upper panel of the dashboard, we placed a bar chart which shows, for each market category, the percentage of the number of startups that were funded by each type of investment. Thus, this gives us the relative distribution of investment types for each market category.

On Fig. 16a, we can observe that, for all market categories, the main investment type was venture capital, with more than half of the investments made in every category being of this type. There seems to exist a big disinterest by angels for the Biotechnology and Industrial fields, but the other types of investment seem to remain more or less constant over all the market categories.

On the bottom panels, we represented the number of startups and the funding over the years, from 1998 to 2014. We used stacked lines to represent the types of investment. This way, we can study what kinds of investors have a bigger role in the funding of startups in each year.

On Fig. 16a, we see that the venture capital was the main investment type in all years, even though other types of investment have been increasing in the most recent years, creating an overall bigger variety in investment types. Moreover, although debt financing, post-IPO and private equity seem to be used as investment types only by a few startups (bottom left panel), they manage to raise a lot of money, especially from 2009 on (bottom right panel).

Note that, on the left, we added a filter in which the user can select a market category. This can limit the records used for the bottom visualizations, deepening both studies into market specification. One example of this filtering can be visualized in Fig. 16b, showing the dashboard when a market category, in this case Biotechnology, is selected. We can see that, accordingly with what is shown in the upper panel, the investments made by angel investors dropped significantly, in comparison to 16a. Moreover, there was a peak in investments made by venture capital firms in 2009.

Even though it was not a research question, we thought it would be interesting to try to confirm the statements made in the last paragraph of Subsection IV-A. To do that, we created a calculated field with the funding rounds (seed and rounds A to G) and created the dashboard presented in Fig. 17. On the upper panel, we represented the average funding made in each round using circles as marks, with the size of each circle representing the number of startups that received funding in that round. On the bottom panel, we represented the average funding by each investor type.

Analysing the upper panel, we see that the funding round with the highest number of records is the seed funding round, because it is the first funding round. This number decreases as we go through the funding rounds, because startups usually only go up to funding rounds B or C. Moreover, we see that the seed is the funding round with the smallest average funding, followed by round A. The latter has an average funding of around $23 million, which is a very different value from the ones presented in Subsection IV-A ($4.9 million in 2010 and $12.1 million in 2017). For rounds B and C, the average funding is around $30 million and $41 million respectively. In Subsection IV-A, we stated that round B can generate up to $10 million and round C generates on average $50 million. Thus, these values are also a bit different from the ones obtained with this dataset.

On the bottom panel, we see that the average funding made by angel investors and venture capital firms is around $8 million and $18 million, respectively. These values are also higher than the ones presented in Subsection IV-A.

Like our previous dashboards, Fig. 17 is also interactive: we can select a funding round in the upper panel and see the types of investments made in that round in the bottom panel. As an example, we selected the seed funding round in Fig. 17b, where we can see that the main type of investors are the private equity firms. Even though it cannot be seen in this paper, for all rounds that the user can select except for this one (seed), the investment types used are just angel investors and venture capital firms.

Lastly, for the last question of the main dataset, we studied the relationship between the number of funding rounds and the success of startups. We considered that a startup has succeeded if it is still operating or was acquired. On the other hand, a startup has failed if its status is closed. Thus, we created a calculated field with this new variable, obtained from the variable status.

In Fig. 18, we defined as columns the different states of Failure and Success and as rows the different numbers of funding rounds that startups can go through. The visualization shows the percentage of the number of rounds for each category (failure or success). This notion is also accompanied by a hue scale.

Observing Fig. 18, we see that around 74% of the startups that failed had only one funding round. However, that percentage is not very different from the one obtained for the startups that succeeded (around 64%). This difference becomes even smaller as we go through the the number of funding rounds. Thus, there does not seem to exist a correlation between the success of a startup and the number of funding rounds.

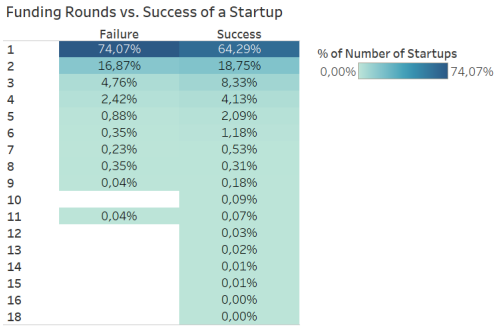


Fig. 18: Dashboard 7: Visualization of the percentage of the number of funding rounds in each outcome: success or failure.

B. Startups and Economic Growth

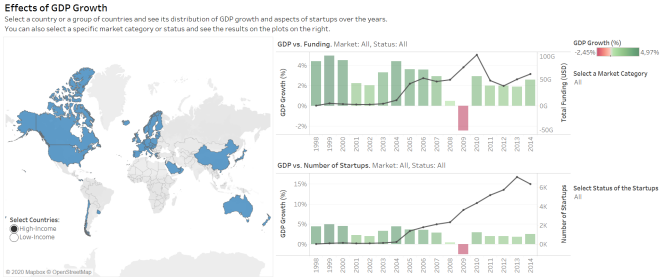
In this subsection, we want to study the influence that the GDP growth in a given country might have in the number of startups founded or their funding. As we mentioned in Subsection IV-C, several articles have shown that the opposite can happen: the number of startups can influence the GDP growth. We can thus explain some of our results using these conclusions.

In order to understand how the GDP relates to startups, we started by modifying our GDP growth dataset to have only four columns: country name, country code, year and respective GDP growth percentage. We then joined this dataset with the main one by country code and year, so that, for each startup in a given country, funded in a given year, we have the GDP growth for that country in that year.

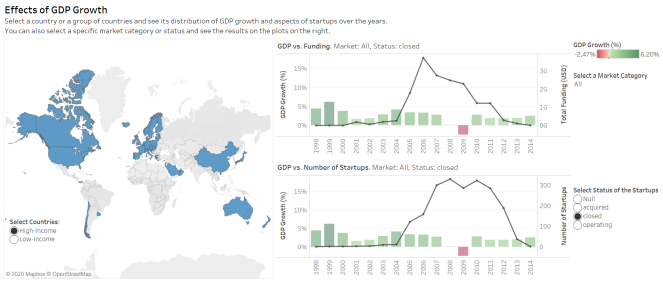
Inspired by Subsection IV-C, we created a calculated field in order to group the countries in two large groups: high-income countries and middle/low-income countries. This grouping was based in the Gross National Income (GNI) *per capita*, using the lists presented in [27].

We then propose a dashboard that allows us to answer several different questions (Fig. 19). This dashboard is quite interactive, allowing us to apply multiple filters that help us in the drill down of the data. The user can choose to select only the high or low-income countries on the panel on the left and see the GDP growth, the funding in startups and the number of startups over the years for those countries in the panels on the right. In addition to being able to choose between these two groups of countries, the user can also choose only one country by selecting it on the map.

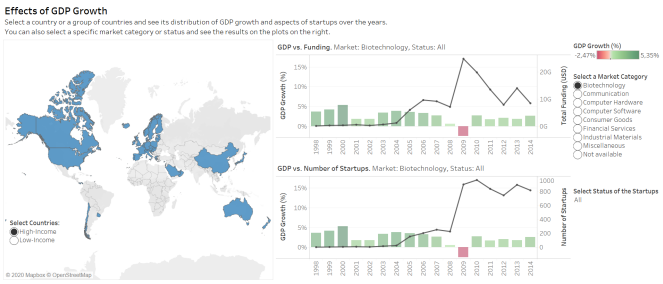
As an example, in Fig. 19a, we selected the high-income countries, in order to understand how investments in startups in this group are influenced by the annual GDP growth. In the right panels, we can see that in 2009 there was a recession in the GDP growth. It is interesting to see that this recession does not seem to influence the total funding in startups in these countries. In fact, in 2010, there is a peak in the money invested. Moreover, the number of startups also increased from 3500 in 2009 to almost double in 2013. This increase could be explained by the fact that, in the years following the recession, people were more encouraged to start businesses in order to improve their economic situation. The increase in funding of startups could be due to the fact that firms were more encouraged to invest in these kind of businesses in order to improve the GDP growth (because, as stated in Subsection



(a) Filtered by high-income countries.



(b) Filtered by high-income countries and by status: closed.



(c) Filtered by high-income countries and by market category: Biotechnology.

Fig. 19: Dashboard 8: Effects of GDP growth (%) on the total funding in startups (USD) and the number of startups throughout the years.

IV-C, entrepreneurial activity stimulates the GDP growth in high-income countries).

However, one could also argue that, in the years following 2010, there was a decrease in the total funding that could still have been influenced by delayed effects of the recession in 2009.

This dashboard is also useful to study if the outcome of a startup was influenced by the GDP growth. To do this, we simply select a status of startups and observe the number of startups in that situation over the years. In Fig. 19b, for example, we kept the filter in high-income countries and selected closed startups. In the bottom right panel, we see that the recession year did not seem to influence the number of startups that closed, which remained approximately constant in the following years. Note that the year used in this visualization corresponds to the year when the startup was first funded and not the year of a status change, for instance. In this particular context it would have been interesting to know the year when the startups were closed, in order to see a more direct relation between a recession and the closing of startups.

We can also study if the different startup market areas are affected differently by the economic growth of the country where they were founded. To do that, we can select a market category in the filter on the right and see if the GDP growth had any influence in the funding or number of startups in that market category.

In Fig. 19c, we selected the market category Biotechnology. We can see that, after the recession year (2009), the total funding in startups in this market started decreasing. However, the number of startups seems to increase from the recession year.

Therefore, this dashboard is very useful to study the relation of GDP growth with the funding and number of startups over the years from multiple perspectives and levels of detail. Note, however, that the dashboard is not useful to make comparisons between different uses of filters, unless we draw some conclusions first.

C. Startups and Market Trends

Finally, we studied if there is a correlation between the number of startups or the invested amount in a given market and the general interest of the public in topics related to this market. To measure this interest, we used the search for that market in Google Trends. For that, we selected a few market areas that we thought would be interesting to study, like Mobile, E-Commerce and Clean Technology, and obtained a measure of how many times those topics were searched on Google. We ended up with a dataset containing three columns: market, date and measure of the trend. This measure represents the interest of search of a term normalized with respect to the highest point in the plot, i.e., a value of 100 is the peak of popularity of a term.

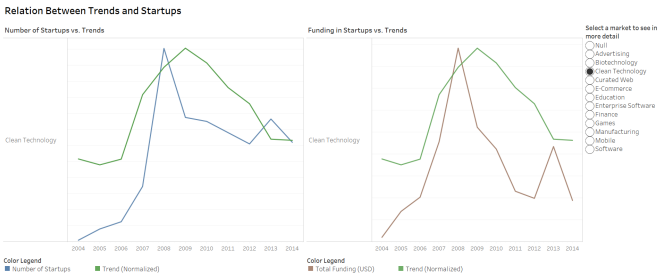
We joined this dataset with the main one by market and by month and year. With this joined dataset, we tried to answer the last two research questions in Section II, studying how the number of startups and investments in different markets are related to trends in those markets.

To do that, we created a dashboard with sparklines, as shown in Fig. 20a. This dashboard contains, for each market, the comparison between the trends and the number of startups in that market on the left panel, and the comparison between the trends and the funding on the right panel. We can thus check if there is a relation between the trend in a given market area and the number of startups and total funding in that market area.

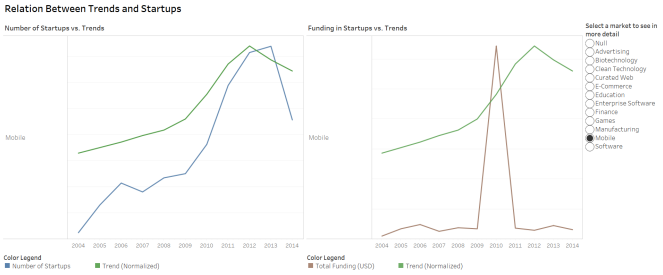
Observing Fig. 20a, we can easily see that, in most markets, there does not seem to exist a correlation between the studied variables. In fact, some markets like E-Commerce and Software seem to have an inverse relation: the trend decreases throughout the years, but the number of startups and the total funding increase.

To study each market in more detail, the user can also select a market in the filter on the right. In Fig. 20b, we selected the Clean Technology market. We can see that there is a peak in the studied variables around the same time, which

(a) Without filters (all markets).



(b) Filtered to show trends in clean technology compared with the number of startups and the total funding in that market throughout the years.



(c) Filtered to show trends in mobile compared with the number of startups and the total funding in that market throughout the years.

Fig. 20: Dashboard 9: Trends in some markets compared with the number of startups and total funding in those markets throughout the years.

could be a sign that they are correlated. As the interest in this market started growing, the same happened to the number of startups and the investments. Perhaps, people started to get more interested in this topic, searching more about it, and that caused them to want to start companies or invest in this market.

In Fig. 20c, we selected the Mobile market, since it seems to be one of the only markets with a clear correlation between the number of startups and the trend. We can see that the number of startups in mobile seems to grow in the same way as the trend in that topic. However, the investments made in this market do not seem to be influenced by the trend of this topic, as there was a single peak in investments in 2010, unlike the trend, that increases steadily until 2012.

VI. Conclusions and Final Remarks

The visualizations proposed in this paper allow us to more easily answer complex questions like the ones presented in Section II, when compared to most of the visualizations shared in Section IV. These were very simple, answering only very straightforward questions and lacking in diversity that could be attained using filters or dashboards.

Most of our visualizations have filters, which improve the interactivity and the comprehension depth of the user. Filters allow any user to select for instance a country, market category or investment type, to better understand the dynamics of the data in these specific dimensions. Filters also allow for a different perspective over the dimensions, by comparing any two or more categories in which the user may be interested.

The use of dashboards is powerful when we have complex questions to answer, as seeing these visualizations side by side lets the user efficiently perceive intertwined behaviours that could otherwise remain unnoticed.

However, one should be aware that our dataset contains a lot of null values, especially on the Investment Types and Funding Rounds columns. We filtered these null values for our visualizations, as well as funding and founding years before 1998, as there were very few startups with funding and founding years before that one. This filtering caused the number of records to be significantly smaller than in the original dataset for some of our visualizations. Therefore, some of our results may not represent reality as well as we would like.

All in all, our visualizations seem to achieve all our goals, allowing us to answer our research questions and draw multiple conclusions from the data with both expressiveness and effectiveness. In the future, some visualizations could be improved by crossing our dataset with more data sources, enriching our data, so that we could take more robust conclusions and answer these and more questions with a bigger certainty.

References

[1] Startup company. Retrieved from https://en.wikipedia.org/wiki/Startup_company. Last accessed in April 23, 2020.

[2] Why Startups Fail: Top 20 Reasons | CB Insights Retrieved from https://www.cbinsights.com/research/startup-failure-reasons-top/. Last accessed in May 7, 2020.

[3] Denise Stephan, How to Raise Startup Capital Through Funding Rounds. Retrieved from https://about.crunchbase.com/blog/raising-startup-capital/. Last accessed in April 23, 2020.

[4] Hai Truong, 5 Stages of Startup Funding. Retrieved from https://techdayhq.com/community/articles/5-stages-of-startup-funding. Last accessed in May 8, 2020.

[5] The Startups Team, How Startup Funding Stages Work. Retrieved from https://www.startups.com/library/expert-advice/startup-funding-stages. Last accessed in May 8, 2020.

[6] Arsalan Sajid, Startup Funding Stages You Should Know About. Retrieved from https://www.cloudways.com/blog/startup-funding-stages/#seed. Last accessed in May 8, 2020.

[7] Understanding differences in startup financing stages. Retrieved from https://startupxplore.com/en/blog/types-startup-investing/. Last accessed in May 8, 2020.

[8] The Economic Times: What is Ipo? Definition of Ipo, Ipo Meaning. Retrieved from https://economictimes.indiatimes.com/definition/ipo. Last accessed in May 8, 2020.

[9] Ivan Draganov, Full 2018 report: €28B invested in European & Israeli startups - Blog: Dealroom. Retrieved from https://blog.dealroom.co/full-2018-report-28b-invested-in-european-israeli-startups/. Last accessed in April 22, 2020.
[10] StartUp Investments (Crunchbase). Retrieved from https://www.kaggle.com/arindam235/startup-investments-crunchbase. Last accessed in April 13, 2020.
[11] GDP growth (annual %). Retrieved from https://data.worldbank.org/indicator/ny.gdp.mktp.kd.zg?name_desc=false. Last accessed in April 24, 2020.
[12] Simon Rogers, What is Google Trends data — and what does it mean? Retrieved from https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8. Last accessed in April 24, 2020.
[13] Susan Li, Tale of 1000 Crunchbase Startups. Retrieved from https://towardsdatascience.com/tale-of-1000-crunchbase-startups-6de0ff97f60e. Last accessed in May 10, 2020.
[14] Susan Li, Tale of 1000 Crunchbase Startups (Tableau Story). Retrieved from https://public.tableau.com/profile/susan.li#!/vizhome/CrunchbaseData_2/Taleof1000CrunchbaseStartups?publish=yese. Last accessed in May 10, 2020.
[15] Startup Database Funding Trends: Where Will We See Tomorrow's Best Startup Companies?. Retrieved from https://www.fundz.net/startup-database-funding-trends. Last accessed in May 10, 2020.
[16] Series A, B, C Funding - The ultimate guide. Retrieved from https://www.fundz.net/what-is-series-a-funding-series-b-funding-and-more. Last accessed in May 8, 2020.
[17] New numbers illustrate how fast fundraising has changed for young startups. Retrieved from https://techcrunch.com/2018/04/24/new-numbers-illustrate-how-fast-fundraising-has-changed-for-young-startups/. Last accessed in May 8, 2020.
[18] Worldwide Business Start-Ups. Retrieved from http://www.moyak.com/papers/business-startups-entrepreneurs.html. Last accessed in May 8, 2020.
[19] How Long It Takes To Raise Capital For A Startup. Retrieved from https://www.forbes.com/sites/alejandrocremades/2019/01/03/how-long-it-takes-to-raise-capital-for-a-startup/#10b4bd057a41. Last accessed in May 8, 2020.
[20] The Ultimate Startup Failure Rate Report [2020]. Retrieved from https://www.failory.com/blog/startup-failure-rate. Last accessed in May 8, 2020.
[21] Do economic or industry factors affect business survival?. Retrieved from https://www.sba.gov/sites/default/files/Business-Survival.pdf. Last accessed in May 8, 2020.
[22] Startup Cemetery. Retrieved from https://www.failory.com/cemetery. Last accessed in May 8, 2020.
[23] Cumul.io dashboards. Retrieved from https://cumul.io/main?utm_source=failory&utm_medium=referral&utm_campaign=failory_guestpost. Last accessed in May 8, 2020.
[24] André van Stel, Martin Carree & Roy Thurik (2004) The effect of entrepreneurship on national economic growth: an analysis using the GEM database, *Discussion Papers on Entrepreneurship, Growth and Public Policy*.
[25] Dave Valliere & Rein Peterson (2009) Entrepreneurship and economic growth: Evidence from emerging and developed countries, *Entrepreneurship and Regional Development*, 21(5):459-480, DOI: 10.1080/08985620802332723.
[26] Justin Doran, Nóirín McCarthy & Marie O'Connor (2018) The role of entrepreneurship in stimulating economic growth in developed and developing countries, *Cogent Economics & Finance*, 6:1, 1442093, DOI: 10.1080/23322039.2018.1442093.
[27] List of countries by GNI (nominal) per capita. Retrieved from https://en.wikipedia.org/wiki/List_of_countries_by_GNI_(nominal)_per_capita. Last accessed in May 29, 2020.

APPENDIX

In this Section, we present a glossary explaining each column in the main dataset.

- **name -** Name of the startup company.
- **market_category -** Broad market category in which the company operates (Consumer Goods, Financial Services, Communication, Industrial Materials, Biotechnology, Computer Software and Computer Hardware).
- **market -** Specific market segment on which the company focuses.
- **status -** Whether the startup is still operating, was acquired or closed.
- **founded_at -** Temporal information about when the startup was founded.
- **country_code, state_code, region, city -** Spatial information.
- **first_funding_at, last_funding_at -** Temporal information about when were the first and the last funding rounds.
- **funding_rounds -** Total number of funding rounds.
- **funding_total_usd -** Capital raised through the funding rounds.
- **angel -** Capital invested by angel investors.
- **venture -** Capital invested by venture capital firms.
- **private_equity -** Capital invested by a private equity firm, when a startup is more firmly established.
- **equity_crowdfunding -** Capital invested through equity crowdfunding platforms, which allow individual users to invest in companies in exchange for equity.
- **product_crowdfunding -** Capital raised when a startup provides its product, which is often still in development, in exchange for money.
- **convertible_note -** Capital raised in an 'in-between' round funding to help startups hold over until they want to raise their next round of funding (e.g. from Round A to Round B). When they raise the next round, this note 'converts' with a discount at the price of the new round.
- **grant -** When a company, investor or government agency provides capital to a startup without taking an equity stake in that startup.
- **debt_financing -** Capital lent by an investor to a startup, which promises to repay the debt with added interest.
- **post_ipo_equity -** Capital invested by firms after the startup has already gone public.
- **post_ipo_debt -** Capital loaned by firms after the startup has already gone public. Similar to debt financing, a company will promise to repay the principal as well as added interest on the debt.
- **secondary_market -** Capital raised in a fundraising event in which one investor purchases shares of stock in a company from other existing shareholders rather than from the company directly.
- **undisclosed -** Investor information was not disclosed.
- **seed -** First funding round, in which investors provide funds before a startup becomes operational.
- **round_A, round_B, round_C, round_D, round_E, round_F, round_G, round_H** - Following funding rounds.