Politecnico di Torino
Collegio di Ingegneria Gestionale, Classe LM-31
Corso di Laurea in Engineering and Management

Tesi di laurea di II livello

Social Impact Investments: how investors characteristics impact on social investments choice

Relatore
Prof.ssa Elisa Ughetto

Candidata
Giorgia Pierotti
matricola 258312

Anno accademico 2019-2020

“È difficile coincidere con lo spazio nel quale proviamo ad inserirci:

solo chi manca può riempire il vuoto che ha lasciato.”

Giulia Carcasi

A mamma Ro, alla sua dolce e granitica forza

A papà Pino, alla sua insaziabile voglia di vivere

Ad Ale, al mio faro e mio alter ego

A Luca e Maria Antonietta, alle mie radici

A Gaetano, la mia casa lontano da casa

INDEX

1 Introduction

The main aim of this research is to analyze the role played by investors characteristics in the social investment choice.

The research will start from theory about social impact investments, that are investments, usually with a return, that are characterized by the "intentionality" of investors who actively seek an opportunity to make a social impact (Oxford, s.d.). After explaining what differentiates these activities from financial investments and philanthropic activities, the literature review will highlight the role played by VC teams in the investment decision choice.

Attention will then be shifted to the methodology used for the collection of a dataset, from Crunchbase website that contains information of both financial and social investments, whose analysis represents the fulcrum of this research. The use of further websites, mainly Impact Asset, Impact Base, Impact Space, will help in the identification of social funds. To fill missing values, LinkedIn was used too.

At this time, the dataset will be analyzed, descriptive statistics will be highlighted. The starting point will be represented by a general overview, followed by more specific ones, like the statistics related to education and professional background, investment and mixed statics and the ones related to partners who will be consider expert in the social field.

To go more in dept, the last part of this research is represented by the construction of a correlation matrix that will be drawn, taking into consideration the more relevant variables of the dataset. This matrix will help to highlighted further statics and to build some indexes, that will be useful also to understand the level of homophily present in the dataset.

The results will show that the investments choice are driven by culture affinity due to geography, that investment field is influenced by university career and that in case of expert investors the choice is driven by the context of previous social experiences.

Even if in general terms, the results can be considered in line with literature evidence, while in some other cases some dummy variables that were created in this analysis will not support the literature results (like in case of top colleges alumni, for whom having studied in these universities doesn't guarantee to cover a more relevant role inside the investment fund).

2 Venture Capital partners role in the context of social impact investments, literature review

2.1 What is a social impact fund?

First of all, to define what a social impact fund is there are two aspects that should be underlined, namely the social sphere and the investment fund sphere. Thus, social impact investments (SIIs) can be defined as a channel through which generates social and/or environmental impact (Freireich & Fulton, 2009). The term “investment fund” refers to any collective investment vehicle that pools capital from many investors with the purpose of challenging them toward investee organizations (Russell, 2007).
When the investment is made to obtain two different impacts, and so both the financial return and the social and/or environmental impact, then we can refer to them as impact-oriented funds (Chiappini, Social Impact Funds, 2017).
The main elements of SIIs focus on:

- A demonstrated aim of obtain measurable social or environmental impact and
- The realization of a financial goal, whether it is considered as essential or secondary aim (Clark, 2012).

For this reason, such enterprises are also called “multidimensional enterprises”, as they are double (even triple sometimes) bottom-lined (Barman, 2015): they pursue a social or environmental mission, or another non-profit-oriented mission, and a profit-oriented mission (Emerson, 1996).

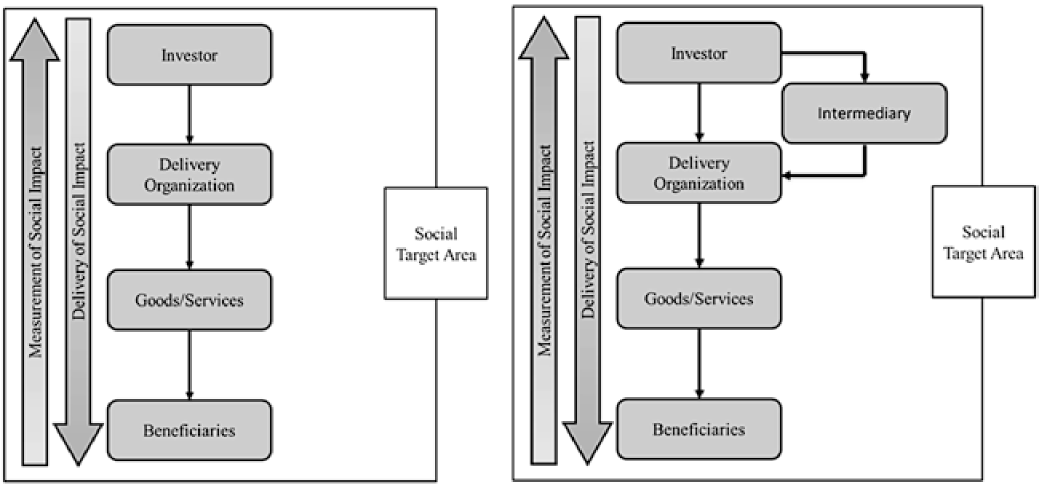
Moreover, differently from the other investments, SII are characterized by the fact that (Calderini, Chiodo, & Michelucci, 2016):

- Social and environmental returns are not accidental but a priori defined and ex post measured
- Proactive approach is used in the search of social impact
- The expectation of at least the repayment of the capital sets it apart from philanthropic activities.

The differences go beyond simply adding a financial goal to the social mission. Hybridity of the investee constitutes the heart of value creation in impact investing (Pache & Santos, 2013). The multidimensional enterprise also creates financial synergies thanks to hybridity. Diversifying activities can create value through coinsurance effect (decrease in default risk and consequently in bankruptcy cost) (Lewellen, 1971). So, contrary to shareholder value creation (measured through financial return and risk), social value creation should be measured according to three dimensions: financial return, social impact and risk.

An impact-oriented fund acts as a collective investor who finances delivery organizations, directly or through an intermediary, to achieve a specific social intent (figure 2.1).

Figure 2.1 *How does SII work?*



Delivery organizations provide services or goods to target beneficiaries. The investment is realized with a specific social or environmental social area of the investment (i.e. health, education, financial exclusion). Impact funds collectively measure and report the social impact realized by delivery organizations.

The establishment of impact-oriented funds can be promoted through a bottom-up initiative or a top-down initiative:

- In the *bottom-up* initiative, the pressure for establishing an impact-oriented fund comes from the bottom, thus institutions involved in addressing specific social needs recognized a need for creating a financial vehicle in order to pool funds and invest money in social or environmental projects or in delivery organizations strictly involved in alleviating social constraints.
- In the *top-down* initiative, the creation of impact-oriented funds can be promoted by enlightened fund managers who recognize an evident and increasing demand of investments generating financial return alongside social impact.

There are mainly two different capital structures:

- *Plain vanilla funds*, in which any owner of share has the same rights of other owners in terms of participation to losses and income and
- *Structured fund*, in which asset owners buy shares with different risk, return and impact profiles and exit requirements.

The structure fund is the type of capital structure that is particularly useful in impact investing because it can permit the attraction of public and private funds.

Moreover, impact-oriented funds can be classified in three groups (Chiappini, Social Impact Funds, 2017):

- Commercial impact-oriented funds
- Non-commercial impact-oriented funds
- Quasi-commercial impact-oriented funds.

According to Mulgan (Mulgan, 2010), even if there is no official definition of social value, it can refer to "wider non-financial impacts of programmes, organizations and interventions, including the wellbeing of individuals and communities, social capital and the environment".

These enterprises need to have formal measures of performance and estimate the outcomes of their activity.

If in the case of an investment that involves a for-profit organization some standard metrics are used to evaluate its performance, like Return On Investment (ROI), the same tool is not valid for the investments that involve social impact funds. The main challenge of the measure of social value is to ensure comparability between activities or entities (Kroeger & Weber, 2014). For example, to address this problem there is a tool that could be used, in particular Social Return On Investment (SROI), a methodology that estimates the monetary equivalent of social value created and compare this equivalent to input used.

SROI methods combine a set of principles that acknowledge power relations by connecting materiality judgements on what outcomes to include with stakeholder consultation via a cost benefit model that uses the principles of net present value discounting to future blended value using monetary proxies (Nicholls, A general theory of Social Impact Accounting, 2018).

The diversity of supports for these firms, from foundations to private investors, makes it necessary to provide for *transparency* and *accountability* (Nicholls, We do good things, don't we? Blended value accounting. Social entrepreneurship , 2009), so that these diversified providers of funds can compare targets and choose to finance the most effective ones (Kroeger & Weber, 2014).

The business model of impact investee is often new and this makes difficult to apply statistical tools and traditional financial analysis (Chell, 2007), but there are some challenges and drivers that can be taken into consideration like transparency, people and partnership (Chiappini, Social Impact Funds, 2017).

- *Transparency* is an essential driver to attract investors. Potential investors should be able to compare investment from either a financial and either social perspective. This means adopting standardized and recognized metrics. Thus, the lack of periodic reports assessing impact funds' performance makes it difficult for organizations or people interested in financial return to invest in impact funds.
- *People* working with impact funds can represent a driver of success, similarly to any other company or organization who hires key talent.
- *Partnership* between different categories of investment can be also essential for the market growth and the success of impact funds.

Moreover, Clark (Clark, 2012) identify four main features capable of driving the success of investment funds.

Outstanding impact funds:

- Operate with the financial and regulatory support of governments,
- Are managed by people with financial and non-financial skills,
- Provide the same priority for social and financial objectives and
- Understand the role of aligning fund investors' objective not only to financial aim, but also to social strategies (Clark, 2012).

A little variation is the one proposed by Brown and Swersky, because, according to them, there should not be same weight, same priority but social objectives must be over profit maximizations and profit distributions in favor of the social mission (Brown & Swersky, 2012).

Furthermore, Drexler and Noble (Drexler & Nobel, 2013) recognize the central role played by funds within the social impact investments framework and outline three recommendations to foster the market:

- The attention to transparency and the provision of financial information to stakeholders,
- A reliable social impact measurement framework useful to make third parties confident of the impact fund's benefit and
- Appropriate strategies to attract financial resources from large-scale investors.

2.2 How do Venture Capital Partners match with startup founders?

The venture capital market is characterized by personal interactions between VC firms and the startups they finance. A central feature is the matching between borrower and investor.

On financial capital supply side, selection is difficult because startup quality is widely dispersed and hidden quality problems can be severe (Gompers & Lerner, Gompers, Lerner, 2001, The venture capital revolution, 2001) (Cochrane, 2006). On demand side, entrepreneur must make a decision on which VCs will most reliably provide not only financial capital but also professional services which can spur startup corporate development (Macmillan, 1989) (Gorman & Sahlman, 1990) (Hellmann & Puri, The interaction between product market and financing strategies, 2000) (Brander, 2002).

The literature suggests that social networks and trusted referrals are important in explaining the matching process (Fried & Hisrich, 1994) and a second literature examines also the geographical proximity between VCs and startups.

Before an investment is made, the counterparties have to know about and assess one another, both of which are facilitated by geographical proximity (Stuart & Sorensen, 2001). After an investment is made co-location can facilitate VCs' ability to both monitor entrepreneurs (Lerner, Venture Capitalists and the oversight of Private Firms, 1995) and add value to them.

Founders and VC partners engage in frequent face-to-face interactions during pitching, screening, contract negotiations, monitoring and post-investment interactions.

Similarity in founder's and partner's ethnicity and education strongly predicts matching, whereas only professional operational experience complementarity predicts VC-entrepreneur matching (Bengtsson & Hsu, How do venture capital partners match with startup founders?, 2010).

Personal similarity matters in the VC matching market. These linkages are significant only for early stage investment in industries with higher level of intangible assets, for which information costs are likely to be more pronounced. These linkages are also more important when the distance between VC and company is greater.

These suggest that the economic role of similarity is reduce *information costs*.

Another important aspect is the VC syndication, that takes place to facilitate due diligence and assessment processes.

2.3 Which is the influenced played by team? (Literature review)

Intellectual capital (IC) is considered an appropriate source of sustainable competitive advantage and it is currently identified as an essential intangible asset in business, especially in those sectors of industry characterized by their highly intensive knowledge capital and advanced technology. IC can be used to enhance an organization's success and to encourage organizational benefits, such as innovativeness, creativity, competitive edge and value creation. Enterprises could gain and retain a competitive advantage when they have great human talents, major capabilities, and boundless innovation and creativity. Indeed, the argument based on the influences that IC has on entrepreneurial performance is consistent with the Resource-Based theory, which advocates that an organization should identify and manage their resources effectively to attain higher performance (Kristandl & Bontis, 2007).

Organization competitiveness requires the skills, knowledge and capabilities of top managers. On the one hand, the great diversity might lead to more internal conflict, which implies less effective decision-making and lower firm outcomes. On the other hand, diversity enables top managers to identify the environmental opportunities and threats in order to formulate the most accurate strategies thereby boosting firm performance (Allen, Dawson, Whetly, & White, 2008).

Then, it's important to say that capital is a critical resource for the establishment and growth of mission-driven companies and impact investors are considered a new type of financial institution with a capacity to support such enterprises (Ozkazanc-Pan & Cetindamar, 2017). There are different forms that invest for impact: social Venture Capital, venture philanthropy, crowdfunding, microfinance and so on (Revelli & Viviani, 2015).

To go deeply in the analysis of VC investments, the first reasonable question to be answered is: when do venture capitalists collaborate?

In VC financing the involvement of a partner VC is a common means to access new financial and managerial resources.

Teaming up with a partner is referred to as *syndication* (Hopp, 2009). In this case, funded firms are thus backed by more than a single VC.

In more-uncertain environments, for example when there is a higher asset specificity, VCs would benefit more from involving partner VCs and their corresponding managerial expertise and financial endowments to either improve selection of funded firms (Lerner, The syndication of venture Capital investments, 1994) or diminish individual capital contributions, thereby allowing participating VCs to invest smaller amounts of capital into a larger number of funded firms to diversify risks (Lockett & Wright, 1999).

VCs that are more open to syndication enjoy more favorable network positions that enable them to benefit from high-quality relationship. Given the underlying uncertainty associated with asymmetric information in partner selection, industry expertise can serve as a positive signal and lend legitimacy to lead VCs (Hopp, 2009).

Moreover, VCs have a strong tendency to collaborate with other venture capitalist because of *affinity*.

The principle of *homophily* shapes group formation and social connection in a wide variety of settings in which similarity between dyad or group member is observed across a broad range of characteristics. But a drawback of homophily is that it may induce social conformity and groupthink that may lead to inefficient decision-making (Ishii & Xuan, Acquirer-target social ties and merger outcomes, 2014). In fact, individuals in homophilic relationships often have an enhanced desire for unanimity and ignore the disadvantages of the favored decision as well as the advice from experts outside the group.

Furthermore, individuals may lower the expected return hurdle and due diligence standards on a project for the opportunity to work with similar others because they derive personal utility from the collaboration. Consequently, collaborations based on characteristics unrelated to ability might suffer from a *cost of friendship* (Gompers, Mukharlyamov, & Xuan, The cost of friendship, 2016) and induce a negative relationship between affinity-based similarities and performance.

Thus, similarities between venture capitalists based on affinity-related characteristics may worsen the performance of their common investments.

Affinity-based similarity not only determines people's attractions to work together for the first time, but also increases their frequency of repeated collaborations.

The inferior performance of investments undertaken by VCs with a high level of affinity between them may be attributed either to *selection* or *treatment*:

- Selection, as collaboration with similar others may have a value in itself (e.g. a venture capital may derive a personal utility from the collaboration),
- Treatment, in fact the negative aspect of affinity may be due to treatment effects after the investment is made (the dark side of homophily can lead to poor decision making, while differences in knowledge, skills and perspectives among team members with different backgrounds may enhance creativity and innovation and elicit a multiplicity of views, adding dimensions to problem-solving and decision-making processes as well as eventually improving performance (Williams & O'Reilly, 1998)).

VC investors provide significant value-add to their portfolio companies. Post-investment, they make important decisions like hiring, firing CEO, identifying customers or partnering opportunities etc, thus, any inefficient decision-making post-investment induced by homophily among high-affinity venture capitalists will negatively impact the success of the portfolio company that they oversee (in other words, the lower likelihood of success of co-investments between venture capitalists that share similar characteristics is triggered by them making inefficient decision or even mistakes that they would otherwise avoid).

So, in homophily, the attraction to each other based on affinity that venture capitalists exhibit is costly and the detrimental nature of affinity is especially prominent for early-stage investments.

When talking about homophily, so the tendency to associate with socially similar others (because of ethnicity, gender, working background, studies, etc) another aspect is the one about coethnicity between the investing VC and funded start-up, that indicates whether the VC and the company have top-level personnel of the same ethnicity. VCs in fact are systematically more likely to invest in a start-up when the VC and the company share the same ethnicity (Hedge & Tumlinson, 2014). A shared ethnicity increases the likelihood that a VC firm invest in a start-up, strengthens the degree of the VC firm's involvement, increases the size and scope of the investment and makes the financial contract more friendly. But, at the same time, VC partner may overestimate the benefits of investing in a founder from the same ethnic group (Bengtsson & Hsu, Ethnic matching in the U.S. venture capital market, 2015).

Proximity gives the possibility to improve performances by reducing post-selection coordination costs. Beside the geographical distance, another role is played by the industry distance, in fact VCs find it easier to make and monitor investment in industries in which they have prior experience (Hellmann, Venture capitalists: the coaches of Silicon Valley, 2000).

The match should work properly, in fact a high-quality match can arguably improve the startup's success chances, whereas a low-quality match can create tensions that impede value-creation. In this sense, a shared ethnicity predicts the existence and intensity of the match between the VC and the startup.

Thus, five evidence come from the literature:

- A shared ethnicity is associated with a *higher probability* that the VC firm invests in the startup,
- A shared ethnicity is associated with *a higher probability* that the VC firm takes a *board seat*,
- A shared ethnicity is associated with the VC firm being more likely to invest at an *early stage* of the startup,
- A shared ethnicity is associated with the VC firm investing *more capital* and across more rounds in the startup,
- A shared ethnicity is associated with the VC firm offering higher pre-money valuation and receive *fewer investor-friendly cash flow contingency* rights in the financial contract.

Some final considerations should mention also the difference between Philanthropic venture capital (PhVC) and traditional venture capital (TVC) because, as said at the beginning of this chapter, social venture capital can fairly be considered in the middle.

Although TVC and PhVC firms have a similar expectation for the cycling of capital – raising, investing, and scaling their investments- they have different organizational objectives. While TVC firms have a singular focus of their on maximizing the economic return of their investments, PhVc have both the objective to maximize the social and economic returns.

While both TVC and PhVC firm founders have high levels of commercial experience, TVC firm founders tend to hold degrees in science, engineering, business, and law more frequently than PhVC firm founders. PhVC founders also differ from TVC founders by having greater work experience in the social sector (Scarlata, Walske, & Zacharakis, 2017).

Finally, according to Becker, human capital can be categorized as general or specific:

- *General* human capital is defined as the acquisition of knowledge and skills through formal education (Becker, 1964),
- *Specific* human capital is typically acquired through work experience (Polanyi, 1967). As such, specific human capital represents the knowledge and the skills that make individual actions and decisions difficult to replicate, as they are often contextually derived.

3 Methodology

3.1 Data collection

The data that will be analyzed in this research were downloaded from Crunchbase, which gives information related to investment funds, the investments received from startups, the partners of the venture capitalist funds that made the investment and the information on individual characteristics of partners. The first types of information are contained inside the dataset called “Investmnet_partners”, while the other are inside “Job+title_partner_list”.

The first step of this research was represented by the analysis of the structured dataset called “Job+title_partner_list”. The analysis has mainly three aims:

- Firstly, identify the variables already contained that can be used to pursue the purpose of the research,
- Secondly, to identify if there were missing values in the relevant variables that should be filled,
- Thirdly, to decide which further variables should be created to enrich the research

The original columns contained are:

- Partner_uuid
- Partner_name
- Org_uuid
- Orgn_name
- Started_on
- Ended_on
- Is_current
- Title
- Job_type
- First_name
- Last_name
- Gender
- Country_code
- State_code
- Region
- City
- Feautured_job_organization_uuid
- Feautured_job_organization_name

- Feautured_job_title

“Partner_name” contains the partner’s name and surname and the same information is then present separately in the columns “First_name” and “Last_name”. To each partner is associated also a unique alphanumerical code contained in the column “partner_uuid”. The same is true for the columns “org_uuid” and “org_name” that refer to investment funds or corporations.

The variables “Started_on” and “ended_on” contains the temporal information about the starting date of working and the ending date of working of a partner in a specific fund or corporation. Moreover, the variable “ended_on” contains empty values in case the partner is still working in that fund or corporation. In fact, in this case, the variable “is_current” will contain a value equal to “true” in case the partner is still working there, “false” otherwise.

The variables “title” and “Job_type” tell in a general and in a more specific way which is the role covered by the partner in the fund or corporation.

There is then the information about the gender, contained in the variable “gender”, and information related the country, state, region and city in which the partner operates, contained in “country_code”, “state_code”, “region”, “city”.

The last variables contain information related to “Feautured_job_organization_uuid”, “Feautured_job_organization_name”, “Feautured_job_title”. These columns contain information related to the main organization in which the investor works. “Feautured_job_organization_uuid”, concerns the unique company code, “Feautured_job_organization_name”, concerns the company name and “Feautured_job_title” concerns the role played within the company.

The number of total rows contained in the dataset is equal to 92415.

As this research would like to discriminate social and traditional investors, it is important to firstly identify which partners invested at least once in a social fund, to be considered social investors.

A research was done to identify social funds. In this case the starting point was the use of the dataset “VC_crunchbase”. In fact, a big dataset was downloaded by *CrunchBase*, that is a leading platform for professionals to discover innovative companies and connect people to pursue new opportunities. Differently from the other websites, it contains information on social investors but also not-social ones.

The dimension of such dataset accounts to 39639 different investments funds and it contains information like investor name, investor uuid, investor continent and country, but without saying if the fund is social or traditional.

Thus, the further step was represented by the collection of data regarding investors in the Social impact investments sector.

For this purpose, three main websites were considered as a guide, respectively *Impact Asset*, *Impact Base*, *Impact Space* (https://www.impactassets.org/, s.d.) (https://impactdatabase.eu/, s.d.) (https://impactspace.com/, s.d.) (https://www.crunchbase.com/, s.d.).

Impact Asset is an online platform that originally belonged to Calvert Impact Capital (CIC) and only in 2010 was spun out of CIC in recognition of the growing need to increase flows of capital to the world's greatest challenges. It connects donors to a rotating offering of private impact funds. Its aim is to provide unparalleled access to investment into top entrepreneurs and fund managers best positioned to tackle these challenges.

The information attached to each fund mainly regards:

- Investor type
- years of operation
- number of investors
- fields in which they operate
- geographic target.

Impact Base is an online platform developed by GIIN (Global Impact Investing Network) in 2010 with the purpose of allowing fund managers to expose their funds and investors to recognize the best investment opportunities. It is defined by GIIN as "a searchable, online database of impact investments funds and products designed for investors". Information collected is grouped into three sections:

- An overview including backgrounds information, target geographies, fund status, fund exit, etc.
- A financial section describing financial investment strategy of funds,
- An impact section identifying the impact strategy.

Impact Space is an open data platform that powers the global impact marketplace. Together with its sister site, Impact Alpha, it provides stories and data to investors, entrepreneurs and other market participants driving business advantage with social and environmental impact.

To each impact investor there are different related characteristics, like:

- Investor type
- Legal structure
- Operating status
- Field in which they operate
- Geographic target.

Information collected from each website were then reported on Excel sheets.

The number of impact investors reported on these websites was respectively of:

- Impact Asset: 139 impact investors,
- Impact Base: 287 impact investors,
- Impact Space: 3343 impact investors.

As the purpose was to create a bigger dataset containing all the impact investors founded, the three datasets were merged in a bigger one. It’s important to say that they were not simply added, as there were different funds that belonged to two or sometimes to all these three datasets. In detail:

- 47 funds belonged both to Impact Asset and Impact Base,
- 108 funds belonged both to Impact Asset and Impact Space,
- 86 funds belonged both to Impact Base and Impact Space,
- 40 funds belonged to all the three datasets.

3.2 Identification of social funds

The next step then was represented by the comparison of investors present in the merged dataset (that contains data from Impact Asset, Impact Base and Impact Space) and the ones downloaded from Crunchbase.

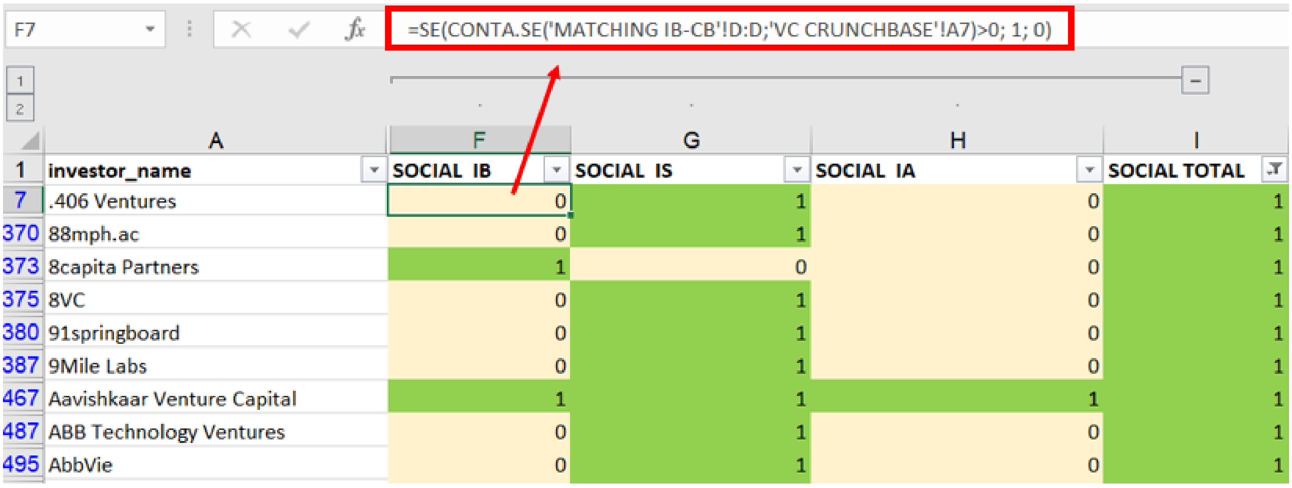
A three further columns were added in the sheet containing Crunchbase data and their cells were filled with value equal to 1 or 0. To do this, three dummy variables called “social IA”, “social IB” and “social IS” were introduced, their value is equal to 1 in case such fund is social, 0 otherwise.

At first *vlookup* Excel function was used to find out if fund can be considered social. Its Excel syntax is:

=Vlookup(lookup_value, table_array, col_index_num, [range_lookup])

The following image shows a little output taken from dataset using Vlookup function *(*Figure 3.1*).*

Figure 3.1 *Vlookup function*



F7 =SE(CONTA.SE('MATCHING IB-CB'!D:D;'VC CRUNCHBASE'!A7)>0; 1; 0)

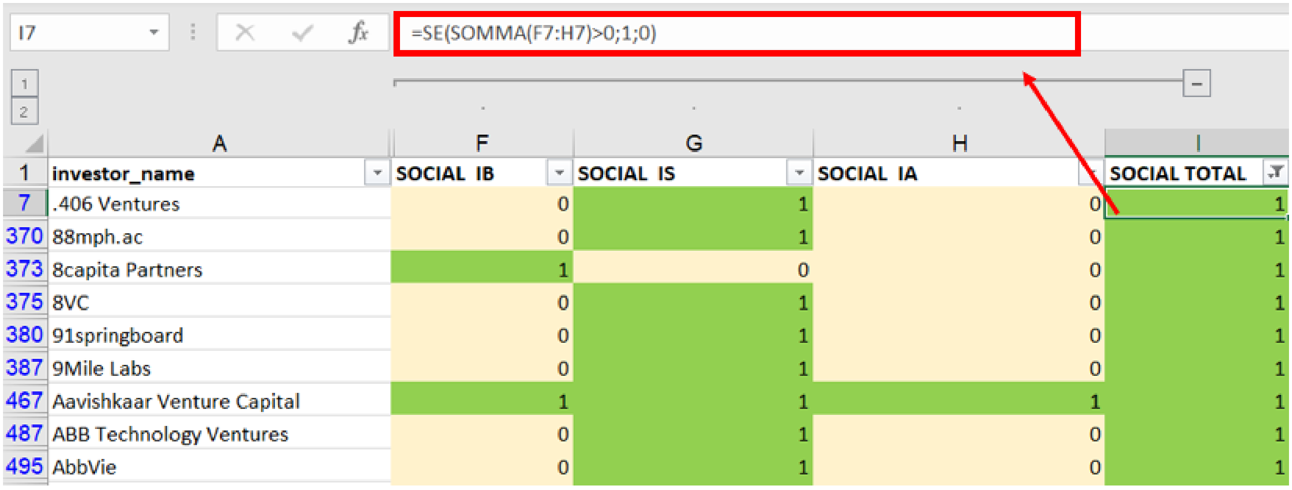
	A	F	G	H	I
1	investor_name	SOCIAL IB	SOCIAL IS	SOCIAL IA	SOCIAL TOTAL
7	.406 Ventures	0	1	0	1
370	88mph.ac	0	1	0	1
373	8capita Partners	1	0	0	1
375	8VC	0	1	0	1
380	91springboard	0	1	0	1
387	9Mile Labs	0	1	0	1
467	Aavishkaar Venture Capital	1	1	1	1
487	ABB Technology Ventures	0	1	0	1
495	AbbVie	0	1	0	1

The fourth value of such function, so [range_lookup], was set equal to 0 (that means setting the value equal to “false”), and this allowed to spot out only the *exact match* between data contained in Crunchbase dataset and in the merged dataset.

As further step, a manual check was made. This step was useful to spot out more funds, that was excluded by using vlookup function, due to differences in the punctuation (there was sometimes an extra comma) or due to the presence of some acronyms that replaced the whole text (like LLP that stands for Limited Liability Partnership). In this case to be surer that the issue fund was the same, the check was made also taking into consideration other characteristics (like the geographic position and the investor type).

A final dummy variable called “social total” was created using *If* function. Its value is equal to 1 if the sum of the three variables “social IA”, “social IB” and “social IS” was higher than 0, so if at least one of these datasets consider the fund as social, 0 otherwise, as shown in the following image *(*figure 3.2*).*

Figure 3.2 *If function*



I7 =SE(SOMMA(F7:H7)>0;1;0)

	A	F	G	H	I
1	investor_name	SOCIAL IB	SOCIAL IS	SOCIAL IA	SOCIAL TOTAL
7	.406 Ventures	0	1	0	1
370	88mph.ac	0	1	0	1
373	8capita Partners	1	0	0	1
375	8VC	0	1	0	1
380	91springboard	0	1	0	1
387	9Mile Labs	0	1	0	1
467	Aavishkaar Venture Capital	1	1	1	1
487	ABB Technology Ventures	0	1	0	1
495	AbbVie	0	1	0	1

In the end, the number of social funds present either in the merged dataset and either in Crunchbase dataset amounts to 2474 (that more or less represents one third of the data present in the merged one). Moreover:

- 130 from Impact Asset
- 153 from Impact Base
- 1814 from Impact Space

About social funds, it is dutiful to specify that the following analysis will consider as social only those funds matched with Crunchbase dataset, while the others that resulted to be not matched will be considered as traditional.

At this time, it was possible to make a check between “Job+title_partner_list” and the other datasets.

The column “org_uuid” of “Job+title_partner_list” was compared to the column “investor_uuid” of “VC_crunchbase”. This operation was useful to consider 1638 rows present in “Job+title_partner_list” as social.

The column “org_name” of “Job+title_partner_list” was compared to the column “investor_name” of “Impact_space”. This operation was useful to consider 4299 rows present in “Job+title_partner_list” as social.

The column “org_name” of “Job+title_partner_list” was compared to the column “investor_name” of “Impact_asset”. This operation was useful to consider 106 rows present in “Job+title_partner_list” as social.

A manual search was also done for some funds that sound to be social, for example from the name.

In the end, aggregating all the result obtained considering the several checks and the manual check, the number of rows that will be consider social amounts to 6918. It’s possible to see this information in the column “socialtotal”. It’s important to specify that the other rows will be considered traditional. In case of traditional, a further difference was considered:

- The rows that were inside the dataset “VC_crunchbase” are considered traditional funds
- The further rows contained in “Job+title_partner_list” that aren’t consider social funds and neither as traditional funds will be considered as traditional corporation.

This information is contained in the column “1_0_org” created to highlight this difference.

To go on with the analysis to spot the missing values, it’s important to say that the more relevant rows will be the ones whose value in the column “1_0_org” is equal to 1.

3.3 Data entry

A data entry activity was then necessary to fill many missing values and to create further variables.

Looking at the original columns, the variable “gender” had quite not significant values 516 times: in particular 324 times the cells were empty, 86 times the value contained was “not_provided”. For all these 516 rows, this information was searched to fill the cell with the value “female” or “male”. All these values were filled. This operation was done searching the information related to the partner name firstly on Crunchbase and later on LinkedIn.

Regarding the column “country_code”, it’s possible to say that there were 672 missing values related to the social funds. All these values were filled. In addition, a further column called “continent_code” was created. Its values were derived from the ones contained inside “country_code”, and it contains seven macro subsets: Europe, Africa, Asia, Oceania, North America, Central America and South America.

It's true that the dataset sometimes contains also more detailed geographical information, that are related to state code, region and city, but for these columns no additional detailed information has been added, because they are not considered as relevant variables for this research.

Till this moment, the large majority of the information clearly seems to be related to work experience background, so maybe this will allow to make a good analysis from job perspective but clearly lack of a component related to the educational background of the partners.

Following this reason, it was decided to look for educational information. In particular, again due to the big numbers, this was done for social male investors and for female investors, either social, either traditional.

Several educational variables were created, they referred to:

- The university in which they studied
- The country in which the university is located
- The continent in which the university is located
- The title they achieved
- The field of study.

In the end as the different number of fields was big, they were then aggregated into bigger macro categories.

The data were collected separately for men and women, and for this reason the dataset contains for example either the variable “university_name_female” either “university_name_male”, either “country_education_male” either “country_education_female”, but at the end these categories were aggregated into a more generic variable, that contains either the information about male educational background, either the ones about female educational background, and are the ones contained in the columns “University_name_total” or “country_education_total” and so on.

In the end, the matrix obtained is 71 for 92415 where 71 is the number of the total variables contained in the dataset, and 92415 is the number of rows. Each row contains information of an investment partner, related to the investment funds in which he operates, or he has operated

A little sample is showed in the image below (figure 3.3).

Figure 3.3 *Final dataset sample*

partner_name	org_name	1_0_org	Investment_rou	Industry	is_current	title	job_type	gende	country_cod
Ã…Âukasz Obuchowicz		org						male	
Aadil Mamujee	Musha Ventures	0	Pre Seed Round		True	Founder	executive	male	USA
Aadil Mamujee	Segovia	org			True	Product & Deployment	employee	male	USA
Aadil Mamujee	Spektra Inc.	org			True	Advisor	advisor	male	USA
Aakanksha Sharma		org						female	IND
Aakash Goel	mer Venture Par	1	Series A	Healthcare	False	Vice President	executive	male	IND
Aakash Goel	cKinsey & Compa	1		Social	False	Consultant	employee	male	IND
Aakash Goel	Sequoia Capital	1		Social	False	Associate	executive	male	IND
Aakash Goel	Infosys	0			True	Software Engineer	employee	male	IND
Aakash Sachdev		org						male	
Aakrit Vaish	Times Internet	1		Impact-tech	True	Co-Founder & CEO of Haptil	executive	male	IND
Aakrit Vaish	Dataweave	org			True	Advisor	advisor	male	IND
Aakrit Vaish	Flat.to	org			True	Founder & Chairperson	executive	male	IND
Aakrit Vaish	Flurry	org			False	Director, Flurry India	executive	male	IND
Aakrit Vaish	Haptik	org			True	Chief Executive Officer	executive	male	IND
Aanand Radia	niversity Venture	1	Seed Round	Education	True	Managing Director	executive	male	USA
Aanand Radia	Always Hired	org			True	Board Member	board_member	male	USA
Aanand Radia	Galvanize	0			True	Board Observer	board_observer	male	USA
Aanand Radia	GSVlabs	0			True	Mentor	advisor	male	USA
Aanand Radia	Vemo Education	org			True	Director	board_member	male	USA
Aapo Bovellan	Festicket	org			True	Board Member	board_member	male	GBR
Aapo Bovellan	Paptic	org			True	Board Member	board_member	male	GBR
Aapo Bovellan	Proxy Ventures	0	Seed Round		True	Founding Partner	executive	male	GBR
Aapo Bovellan	Uplause	org			True	Board Member	board_member	male	GBR
Aapo Bovellan	Zen-Me	org			True	Board Member	board_member	male	GBR
Aaref Hilaly	Sequoia Capital	1	Series C	Social	False	Partner	executive	male	USA
Aaref Hilaly	CenterRun	org			False	Founder, CEO	executive	male	USA
Aaref Hilaly	Clari	org			True	Board Member	board_member	male	USA

Once finished the analysis of the first dataset, it was necessary to focus on the second one, called “Investment_partners”. The main variables contained are:

- Investor_name
- Investor_uuid
- Partner_name
- Partner_uuid
- Founding_round_uuid
- Founding_round_name

Thus, it was quite clear that the variable “founding_round_name” is interesting and it may enrich this research. It should be mentioned that this variable contains at the same time two information:

- The first information is related to the investment round (like “Series A”, “Seed Round”, “Venture Round” and so on)
- The second information is related to the startup that received the social investments.

Once more, to enrich the research, additional information was searched.

In fact, for all the investment considered as social, thus exploiting the dummy variable “social” already built in the previous step, the information related to the industrial specialization in which the startup that received the investment operates was reached. This information is contained in the column “Industrial_field”, that can have many values, like “Agriculture & nutrition”, “Education”, “Empower people”, “Energy” and “Healthcare”.

4 Descriptive Statistics

To derive the following statistics, two matrixes will be used. They are contained inside the files "Job+people_partner_list" and "investment_partners". While the first one mainly refers to information related to the investment fund, investment partner, job type covered, educational and geographical information, the second file mostly refers to investment round and field in which the startup that had received the investment operates.

With reference to both files, each row contains information of an investment partner, related to the investment funds in which he operates, or he has operated.

The aim of this chapter is to highlight the descriptive statistics, thus, to create a summary statistic that quantitatively describes and summarize features (https://en.wikipedia.org/wiki/Descriptive_statistics, s.d.) from the data contained in the aforementioned matrix.

The analysis will start from a general perspective, then it will highlight the educational and professional background, the specific information related to the investment, there will be then a more complex analysis that will consider more than one of these different aspects at the same time and a distinction between the different social investors will be made, based on the time they invested in a social startup.

4.1 General Statistics

To explain the general statistics, the used file was "Job+people_partner_list" and the relevant variables that were considered are:

- Org_uuid
- Partner_uuid
- Gender
- _0_org

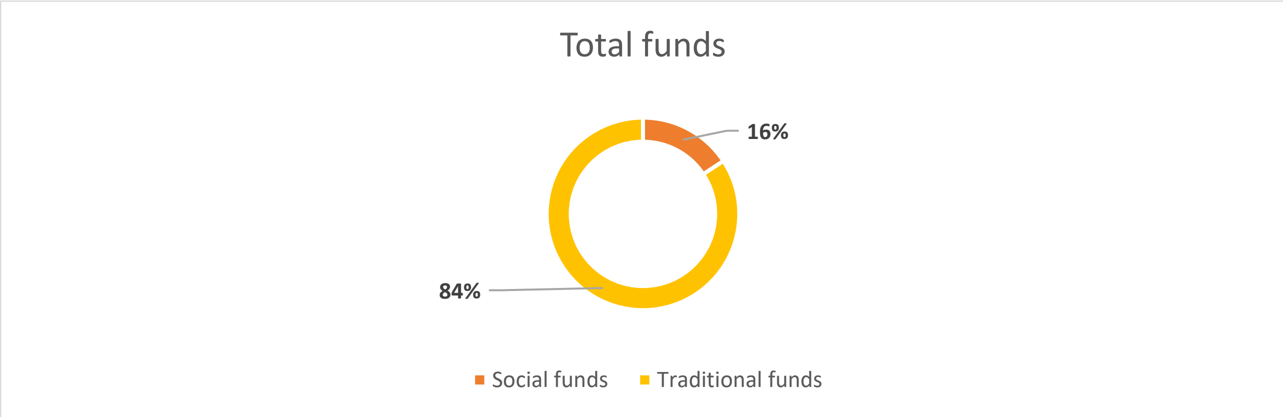
In particular, the variable "_0_org" is equal to 1 in case the organization is considered as social fund, 0 in case of a traditional fund, "org" in case the organization is classified as a traditional corporation.

Coding on Stata made possible to obtain the frequencies.

First of all, the number of rows that contains information about the organization code (org_uuid) is equal to 80971, of which 6992 are classified as social funds, 22018 as traditional funds, 51961 as traditional corporation. As the focus on this research is about investment funds, the rows that refer to corporation will be neglected in this analysis.

Moreover, it's possible to say that the dataset contains univocally 8378 different funds, of which 15,74% are considered as social and 84,26% are considered as traditional (figure 4.1).

***Figure 4.1** Total investment funds*

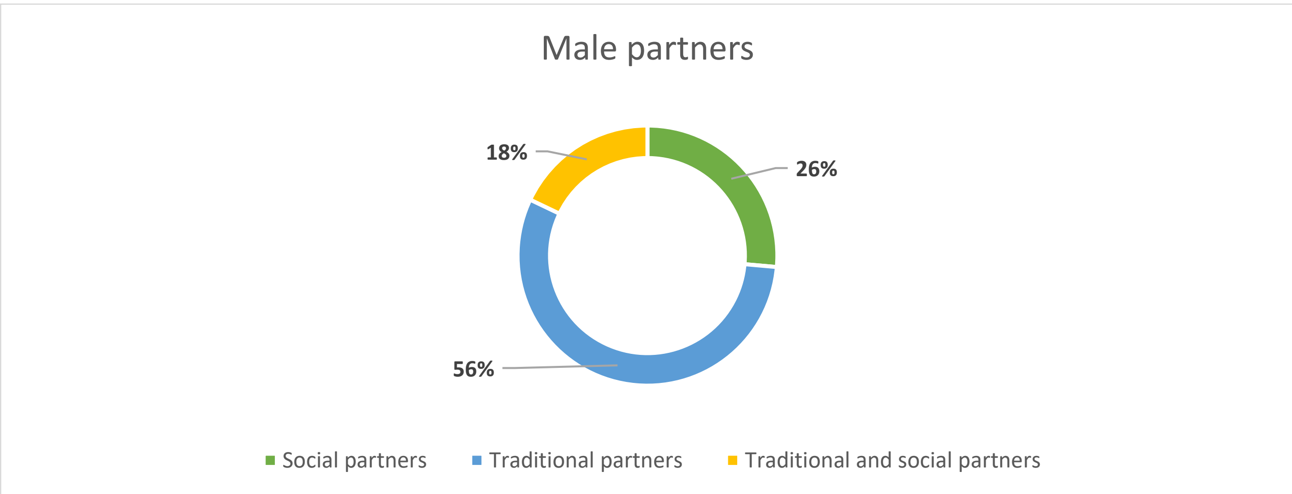


Moving from investment funds to investment partners, it's possible to find out that the dataset contains univocally 11700 different partners, of which 10471 (thus the 89,50%) are men, while 1229 (thus the 10,50%) are women.

Now it's time to look deeply, considering at the same time if the fund is social or traditional and considering the gender. In case of male partners, 4309 is the number of male partners that invested at least once in a social fund, 9087 is the number of male partners that invested at least once in a traditional fund and 2925 is the number of male partners that invested at least once either in a traditional either in a social fund.

The percentages are shown in the following graph (figure 4.2).

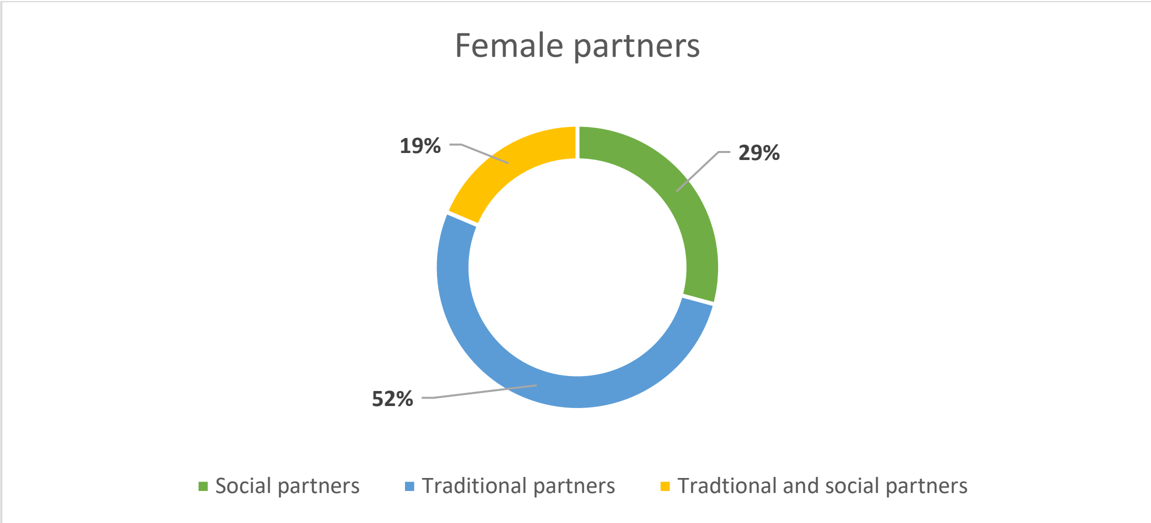
***Figure 4.2** Male partners*



In case of female partners, 572 is the number of female partners that invested at least once in a social fund, 1022 is the number of female partners that invested at least once in a traditional fund and 365 is the number of female partners that invested at least once either in a traditional either in a social fund.

The percentages are shown in the following graph (figure 4.3).

Figure 4.3 *Female partners*



Looking at these initial percentages, there seems to be no difference in how partners invest in social and/or traditional funds if the variable related to gender is taken into consideration.

To take a further step forward, it can be interesting to create a variable to discriminate whether a partner invested only once in a social fund or more than once. In this second case the partner may be considered as "social expert investor" while in the first case he will be considered simply as a "social investor". This part will be explained inside the paragraph "Social expert statistics", at the end of this chapter.

4.2 Educational Statistics

To enrich the research, information about educational backgrounds were searched and added inside the file “Job+people_partner_list”.

As this research was made in a manual way, thus looking at one investor at a time, this operation was performed for all the female investors, consider both when social and when traditional (they totally amount to 1229), while due to the big amount of male investors, in this case the research was fulfilled only for those who are considered social (that amount to 4309).

To be considered fair, the comparison in this case will be made between social male partners and social female partners.

For each of them a research about their educational backgrounds was done to identify:

- The university in which they studied
- The country in which the university is located
- The continent in which the university is located
- The title they achieved
- The field of study.

In the end as the different number of fields amounts to more than 150 categories, they were then aggregated into 28 bigger macro categories (that will be called “macro_field_education”).

The variable “degree” contains the most important title obtained by investor, in accordance to the International Standard Classification of Education (ISCED), that is a standard created by UNESCO, that considers the highest title as the PhD, followed by MBA, Master, Bachelor, High school diploma. Thus, for example, if the investor obtained both a Bachelor and an MBA title, in the “degree” there will be the reference to the MBA.

The website used to extract this information was Crunchbase.

In most cases, it was possible to obtain all the information digiting the name of the investor in the search bar. When there were missing values, an additional research was done on LinkedIn to combine the information. When there were still missing values or when there was too much uncertainty (like in the case of an investor with a common name and surname), the values of these variables were kept empty. It also seemed that the older is the title or the longest is the work experience, the higher is the probability to find empty values.

In the end, the analysis will compare 3311 social male partners and 477 social female partners for whom all the above variables are available.

To explain the educational statistics, the relevant variables that will be considered are:

- Partner_uuid
- Gender
- _0_org
- University_name_tot
- Country_code_education_tot
- Continent_education_tot
- Degree_tot
- Field_education_tot
- Macro_field_education_tot

First of all, considering the variable "University_name", the titles were obtained in 749 different universities. 2876 represents the number of titles obtained in the American universities (78,19%). The more relevant are then the British that amount to 222 (6,04%), 113 are French (3,07%), the Canadian are 94 (2,56%) and the Indian ones are 81 (2,20%).

Thus, although in the dataset there are 42 countries, the cumulative of the five most frequent amounts to 92,06%.

All the values are shown in the following table (table4.1).

Table 4.1 *Total titles by country*

Education country code	*Frequency*	*Percentage*	*Cumulate*
ARG	5	0,14	0,14
AUS	23	0,63	0,76
AUT	7	0,19	0,95
BEL	2	0,05	1,01
BRA	5	0,14	1,14
CAN	94	2,56	3,70
CHE	17	0,46	4,16
CHN	16	0,44	4,59
COL	1	0,03	4,62
CZE	3	0,08	4,70
DEU	43	1,17	5,87
DNK	5	0,14	6,01
EGY	2	0,05	6,06
ESP	2	0,05	6,12
EST	2	0,05	6,17
FIN	5	0,14	6,31
FRA	113	3,07	9,38
GBR	222	6,04	15,42
HKG	1	0,03	15,44
IDN	2	0,05	15,50
IND	81	2,20	17,70
IRL	7	0,19	17,89
ISR	39	1,06	18,95
ITA	15	0,41	19,36
JPN	11	0,30	19,66
KOR	1	0,03	19,68
MEX	3	0,08	19,77
NLD	13	0,35	20,12
NOR	5	0,14	20,26
PHL	1	0,03	20,28
POL	2	0,05	20,34
PRT	2	0,05	20,39
ROU	2	0,05	20,42
RUS	1	0,03	20,69
SGP	10	0,27	20,77
SPA	3	0,08	21,10
SWE	12	0,33	21,53
THA	16	0,44	21,56
TUR	1	0,03	21,59
TWA	1	0,03	21,68
USA	2876	78,19	99,81
ZAF	7	0,19	100
Total	*3678*	*100*	

Then, the following table shows the results if the continents are taken into consideration (table4.2), showing how the ones obtained in Africa, Oceania, and South America seems to be quite negligible (the cumulate sum is equal to 1,2%).

Table 4.2 *Total titles by continent*

Educational continent	*Frequency*	*Percentage*	*Cumulate*
Africa	10	0,27	0,27
Asia	157	4,27	4,54
Europe	504	13,70	18,24
North America	2973	80,83	99,08
Oceania	23	0,63	99,70
South America	11	0,30	100
Total	*3678*	*100*	

The next table shows that if the education level is taken into consideration, the most frequent title is MBA, followed by Bachelor, Master and PhD (table 4.3).

Table 4.3 *Titles*

Degree	*Frequency*	*Percentage*	*Cumulate*
Bachelor	1209	32,87	32,87
MBA	1504	40,89	73,76
Master	669	18,19	91,95
PhD	296	8,05	100
Total	*3678*	*100*	

This ranking is valid either for female partners either for male partners, but in case of female investors it can be highlighted that the percentage of women that obtained the highest title, PhD, it is almost twice as many as men (13,21% vs 7,28%), and that the percentage in case of Master don't differ too much (14,26% vs 18,78%).

In the end it will be correct to say that female partners are more educated than male partners (in fact, remembering ISCED ranking, in that case the Bachelor is the lowest title, and it represents the 21,80 of female titles and the 34,52% of male titles).

All the results are shown in the following tables (table4.4, table4.5).

Table 4.4 *Female titles*

Degree	*Frequency*	*Percentage*	*Cumulate*
Bachelor	104	21,80	21,80
MBA	242	50,73	72,54
Master	68	14,26	86,79
PhD	63	13,21	100
Total	*477*	*100*	

Table 4.5 *Male titles*

Degree	*Frequency*	*Percentage*	*Cumulate*
Bachelor	1105	34,52	34,52
MBA	1262	39,43	73,95
Master	601	18,78	92,72
PhD	233	7,28	100
Total	*3201*	*100*	

Finally, considering the 28 macro aggregations, the related titles fields are shown in the following table (table 4.6), irrespectively to the titles.

Table 4.6 *Macro categories of field*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Entrepreneurship	1	0,03	0,03
Biology	63	1,71	1,74
Business Administration	556	15,12	16,86
Chemistry	75	2,04	18,90
Design	9	0,24	19,14
Economics & Commerce	266	7,23	26,37
Education	5	0,14	26,51
Energy	4	0,11	26,62
Engineering	379	10,20	36,92
Entrepreneurship	79	2,15	39,07
Finance	425	11,56	50,63
Healthcare	31	0,84	51,47
Humanistic sciences	132	3,59	55,06
Languages	19	0,52	55,57
Law	181	4,92	60,49
Leadership	2	0,05	60,55
Management	685	18,62	79,17
Marketing	112	3,05	82,22
Mathematics	55	1,50	83,71
Medicine	81	2,20	85,92
Others	8	0,22	86,13
Pharmacy	4	0,11	86,24
Physics	55	1,50	87,74

Political sciences & International relations	110	2,99	90,73
Science	33	0,90	91,63
Social sciences	35	0,95	92,58
Strategy	25	0,68	93,26
Technology	248	6,74	100
Total	*3678*	*100*	

It is quite clear that the great majority of titles are related to managerial and economical subjects, in fact the category "Management" represents 18,62%, "Business Administration" amounts to 15,12%, "Finance" amounts to 11,56% and "Economics & Commerce" amounts to 7,23%.

The cumulative sum of these categories represents 52,53%, thus it is possible to say that one in two partners has a managerial or economical educational background.

Looking at the data in a more accurate way, there are two clear difference:

- The cumulative sum of "Management", "Business Administration", "Finance" and "Economics and Commerce" is quite high for both categories, but while for male partners it is possible to say that one investor out of two has a managerial or educational background (51,04%), in case of female partners it is possible to say that almost two investors out of three have this kind of background (62,48%);
- Although the values of the cumulative sums don't differ too much, this is not true if looking at micro categories, in fact "Business Administration" represents the 36,06% for female partners and the 12% for male partners, while "Management" represents the 20,74% for male partners and only 4,40% for female partners.

All the details are contained in the following tables (table 4.7, table 4.8).

Table 4.7 *Macro categories of field for male partners*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Biology	51	1,59	1,59
Business Administration	384	12,00	13,59
Chemistry	67	2,09	15,68
Design	9	0,28	15,96
Economics & Commerce	228	7,12	23,09
Education	3	0,09	23,18
Energy	4	0,12	23,31
Engineering	353	11,03	34,33
Entrepreneurship	70	2,19	36,52
Finance	358	11,18	47,70
Healthcare	24	0,75	48,85
Humanistic sciences	125	3,91	52,36
Languages	12	0,37	52,73
Law	160	5,00	57,73
Leadership	2	0,06	57,79
Management	664	20,74	78,54
Marketing	100	3,12	81,66
Mathematics	51	1,59	83,26
Medicine	62	1,94	85,19
Others	5	0,16	85,35
Pharmacy	3	0,09	85,44
Physics	54	1,69	87,13
Political sciences & International relations	92	2,87	90,00
Science	27		90,85
Social sciences	33	1,03	91,88
Strategy	21	0,66	92,53
Technology	239	7,47	100
Total	*3201*	*100*	

Table 4.8 *Macro categories of field for female partners*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Entrepreneurship	1	0,21	0,21
Biology	12	2,52	2,73
Business Administration	172	36,06	38,78
Chemistry	8	1,68	40,46
Economics & Commerce	38	7,97	48,43
Education	2	0,42	48,85
Engineering	26	5,45	54,30
Entrepreneurship	9	1,89	56,18
Finance	67	14,05	70,23
Healthcare	7	1,47	71,70
Humanistic sciences	7	1,47	73,17
Languages	7	1,47	74,63
Law	21	4,40	79,04
Management	21	4,40	83,44

Marketing	12	2,52	85,95
Mathematics	4	0,84	86,79
Medicine	19	3,98	90,78
Others	3	0,63	91,40
Pharmacy	1	0,21	91,61
Physics	1	0,21	91,82
Political sciences & International relations	18	3,77	95,60
Science	6	1,26	96,86
Social sciences	2	0,42	97,27
Strategy	4	0,84	98,11
Technology	9	1,89	100
Total	*477*	*100*	

With regard to technical profiles for all the partners, either female and male, the category "Engineering", that represents the 10,30% seems to deserve a mention. Moreover, this macro category contains 33 micro branches, as shown in the following table (table 4.9).

Table 4.9 *Engineering fields*

Education field	*Frequency*	*Percentage*	*Cumulate*
Bioengineering	1	0,26	0,26
Computer Engineering	1	0,26	0,53
Aeronautical Engineering	6	1,58	2,11
Aerospace engineering	3	0,79	2,90
Applied Engineering	1	0,26	3,17
Astronautical engineering	1	0,26	3,43
Biomechanical engineering	1	0,26	3,69
Biochemical engineering	1	0,26	3,96
Bioengineering	4	1,06	5,01
Biological engineering	1	0,26	5,28
Biomedical engineering	9	2,37	7,65
Chemical engineering	20	5,28	12,93
Civil engineering	7	1,85	14,78
Communication engineering	3	0,79	15,57
Electrical engineering	143	37,73	53,30
Electrical tech engineering	1	0,26	53,56
Electronic engineering	1	0,26	53,83
Engineering	77	20,32	74,14

Environmental engineering	2	0,53	74,67
Industrial Engineering	29	7,65	82,32
Industrial & mang engineering	1	0,26	82,59
Management engineering	10	2,64	85,22
Manufacturing engineering	1	0,26	85,49
Material engineering	2	0,53	86,02
Mechanical engineering	39	10,29	96,31
Mechanical ec engineering	1	0,26	96,57
Mineral engineering	1	0,26	96,83
Naval engineering	1	0,26	97,10
Nuclear engineering	4	1,06	98,15
Polymer engineering	1	0,26	98,42
Production engineering	1	0,26	98,68
Software engineering	2	0,53	99,21
Systems engineering	3	0,79	100
Total	*379*	*100*	

Moreover, looking at the data is it possible to say that one engineer out of two is an electrical engineer (37,73%) or a mechanical engineer (10,29%), as the cumulative sum of these subcategories amounts to 48,02%.

Now it's time to look deeper, considering more than one variable at a time.

Let's start from considering "degree" and "macro_field_education".

If the variables "degree" and "macro_field_education" are considered together, there is then a significant outcome. In fact, setting "degree" equal to MBA, Management represents the 41,69% followed by Business Administration (28,32%) and Finance (17,22%), with the cumulative sum equal to 87,23%. Thus, the results show that the managerial and economical fields prevail largely over the others in case of MBA, as shown in the following table (table 4.10).

Table 4.10 *Fields if degree = MBA*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Entrepreneurship	1	0,07	0,07
Biology	1	0,07	0,13
Business Administration	426	28,32	28,46
Economics & Commerce	4	0,27	28,72
Energy	2	0,13	28,86
Engineering	1	0,07	28,92
Entrepreneurship	59	3,92	32,85
Finance	259	17,22	50,07
Healthcare	19	1,26	51,33
Humanistic sciences	1	0,07	51,40
Law	3	0,20	51,60
Leadership	1	0,07	51,66
Management	627	41,69	93,35
Marketing	66	4,39	97,74
Mathematics	1	0,07	97,81
Medicine	3	0,20	98,01
Others	3	0,20	98,20
Pharmacy	1	0,07	98,27
Physics	1	0,07	98,34
Science	2	0,13	98,47
Strategy	18	1,20	99,67
Technology	5	0,33	100
Total	*1504*	*100*	

Now setting “degree” equal to “PhD” it is really clear that technical and scientific categories prevail over the managerial and economical ones. In this case the higher frequency is represented by the macro field Engineering that amounts to 22,30%, followed by Chemistry that represents 14,86%, Technology that amounts to 12,84%, Medicine that amounts to 11,15%, and then Biology 8,78% and Physics 7,77%.

If now the three most frequent categories of MBA are aggregated (in that case they represented 87,23%), it can be seen that the cumulative sum is equal to 5,06% (Management 1,01%, Business Administration 3,04%, Finance 1,01%).

All the details are contained in the following table (table 4.11).

Table 4.11 *Fields if degree = PhD*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Biology	26	8,78	8,78
Business Administration	9	3,04	11,82
Chemistry	44	14,86	26,69
Economics & Commerce	6	2,03	28,72
Engineering	66	22,30	51,01
Entrepreneurship	2	0,68	51,69
Finance	3	1,01	52,70
Healthcare	2	0,68	53,38
Humanistic sciences	3	1,01	54,39
Law	18	6,08	60,47
Management	3	1,01	61,49
Marketing	2	0,68	62,16
Mathematics	6	2,03	64,19
Medicine	33	11,15	75,34
Others	1	0,34	75,68
Physics	23		83,45
Political science & international relations	1	0,34	86,49
Science	8	2,80	86,49
Social science	1	0,34	86,82
Strategy	1	0,34	87,16
Technology	38	12,84	100
Total	*296*	*100*	

Finally setting “degree” equal to “Bachelor” and then equal to “Master”, the results that can be obtained are shown in the tables below (table12, table13).

The analysis of Bachelor title highlights that there is not a unique trend, but that there is a great variety that goes from “Economics & commerce” (17,29%) to Engineering (15,30%), from “Technology” (11,83%) to “Finance” (10,42%).

All the details are shown in the following table (table 4.12).

Table 4.12 *Fields if degree = Bachelor*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Biology	33	2,73	2,73
Business Administration	74	6,12	8,85
Chemistry	21	1,74	10,59
Design	6	0,50	11,08
Economics & Commerce	209	17,29	28,37
Education	2	0,17	28,54
Energy	1	0,08	28,62
Engineering	185	15,30	43,92
Entrepreneurship	10	0,83	44,75
Finance	126	10,42	55,17
Healthcare	2	0,17	55,33
Humanistic sciences	100	8,27	63,61
Languages	15	1,24	64,85
Law	26	2,15	67,00
Management	23	1,90	68,90
Marketing	39	3,23	72,13
Mathematics	39	3,23	75,35
Medicine	3	0,25	75,60
Others	1	0,08	75,68
Physics	21	1,74	77,42
Political science & international relations	89	7,36	84,78
Science	13	1,08	85,86
Social science	27	2,23	88,09
Strategy	1	0,08	88,17
Technology	143	11,83	100
Total	*1209*	*100*	

The output of the analysis of Master title is quite similar to the one obtained while considering Bachelor title.

An aspect that can be highlighted in this case is the presence of the category “Law”, whose percentage seems to be significant (20,03%).

All the details are contained in the table below (table 4.13).

***Table 4.13** Fields if degree = Master*

Macro education field	*Frequency*	*Percentage*	*Cumulate*
Biology	3	0,45	0,45
Business Administration	47	7,03	7,47
Chemistry	10	1,49	8,97
Design	3	0,45	9,42
Economics & Commerce	47	7,03	16,44
Education	3	0,45	16,89
Energy	1	0,15	17,04
Engineering	127	18,98	36,02
Entrepreneurship	8	1,20	37,22
Finance	37	5,53	42,75
Healthcare	8	1,20	43,95
Humanistic sciences	28	4,19	48,13
Languages	4	0,60	48,73
Law	134	20,03	68,76
Leadership	1	0,15	68,91
Management	32	4,78	73,69
Marketing	5	0,75	74,44
Mathematics	9	1,35	75,78
Medicine	42	6,28	82,06
Others	3	0,45	82,51
Pharmacy	3	0,45	82,96
Physics	10	1,49	84,45
Political science & international relations	20	2,99	87,44
Science	10	1,49	88,94
Social science	7	1,05	89,99
Strategy	5	0,75	90,73
Technology	62	9,27	100
Total	*669*	*100*	

As stated at the beginning of this paragraph, the most involved country when considering educational background is represented by the United States, that represents, more or less, the 80%.

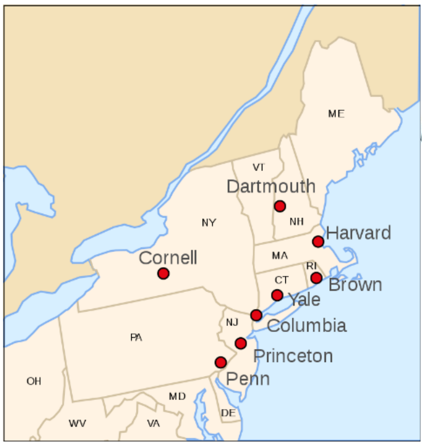
So, to go even deeper, an interesting thing could be represented by the creation of a further variable that, inside the American environment, will consider if those investors have studied in the best American universities or if they hadn't.

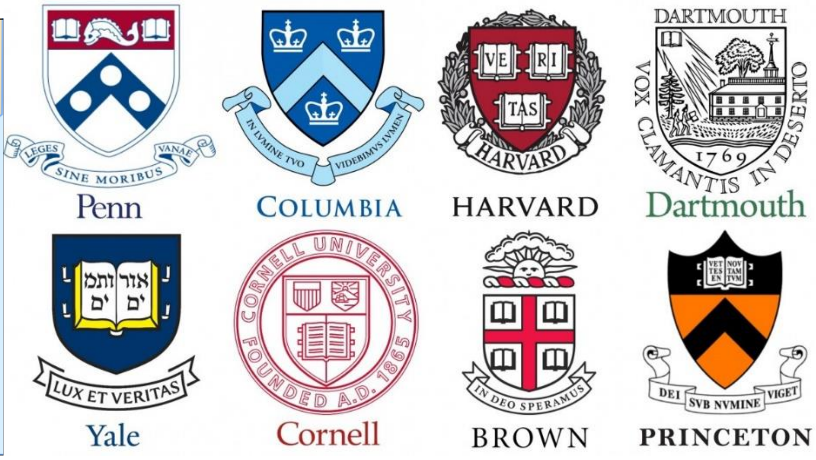
The creation of this variable will not be simply considered in case of educational background, but it can also be used in correlation with other variables regarding the work experience to see different thing like, for example, if those investors have an higher probability to cover a type of role in the fund or not, in which industry they prefer to invest and so on.

In detail, these universities are the ones that belong to the Ivy League, that is a conference comprising eight private universities in the Northeastern United States.

***Figure 4.4** Geographic position of Ivy League universities*

***Figure 4.5** University arms*





As also shown in the pictures above (figure 4.4 and figure 4.5), its members in alphabetic order are:

- Brown University
- Columbia University
- Cornell University
- Dartmouth College
- Harvard University
- University of Pennsylvania
- Princeton University
- Yale University

Numerically, the number of titles achieved in these universities is equal to 962. The details can be seen in the following graphs (figure 4.6, figure 4.7, figure 4.8, figure 4.9, figure 4.10, figure 4.11, figure 4.12, figure 4.13, figure 4.14).

Figure 4.6 Titles at Brown University

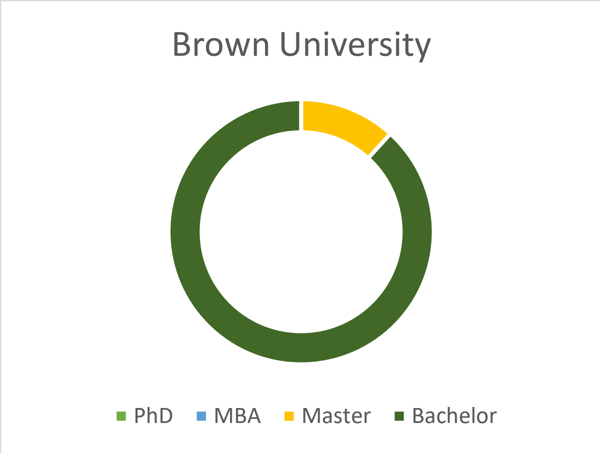


Figure 4.7 Titles at Columbia University

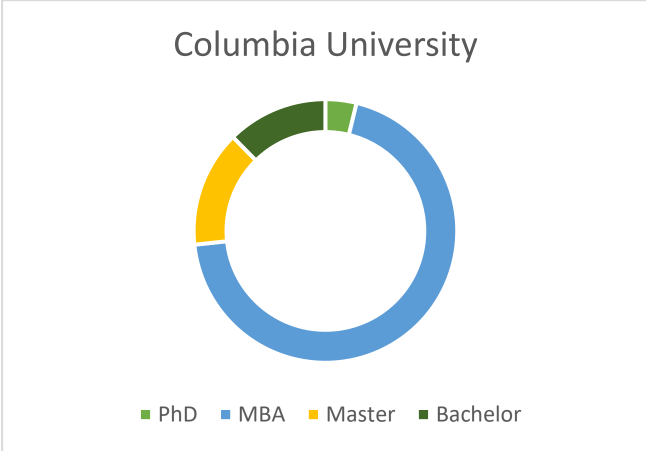


Figure 4.8 Titles at Cornell University

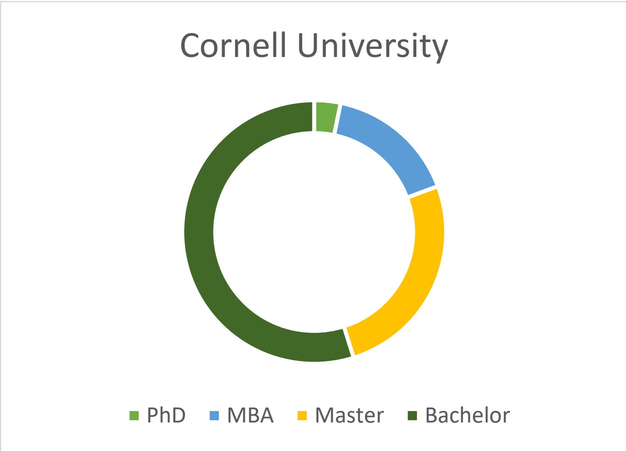


Figure 4.9 Titles at Dartmouth College

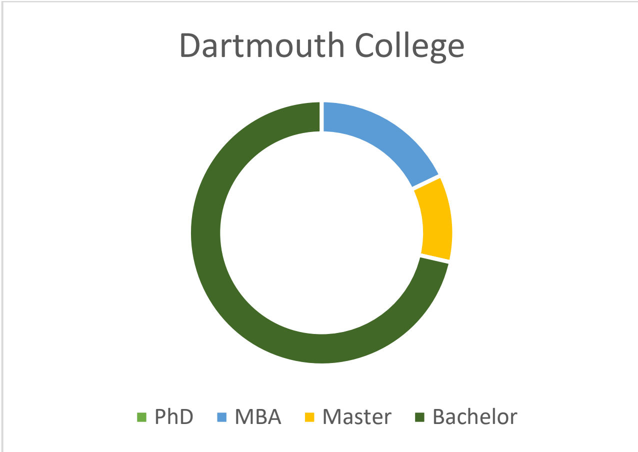


Figure 4.10 Titles at Harvard University

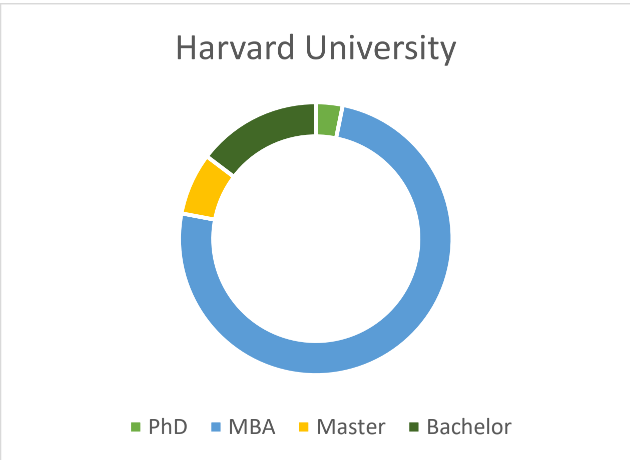


Figure 4.11 Titles at University of Pennsylvania

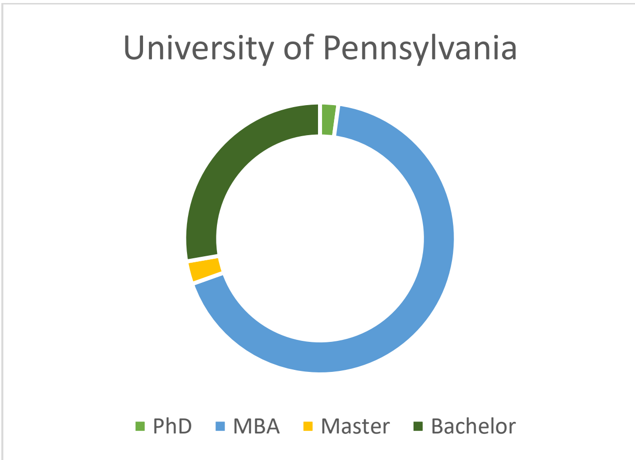
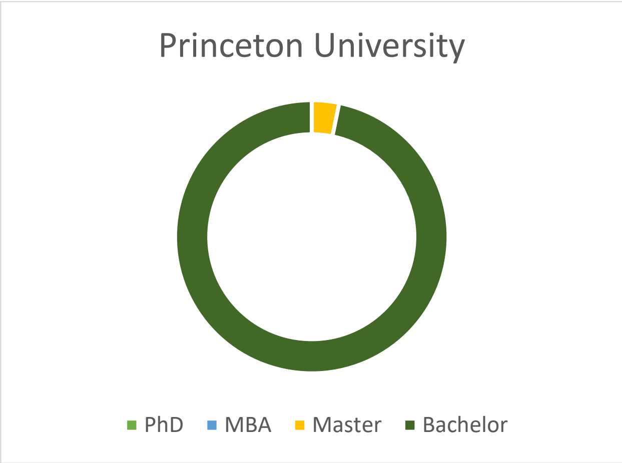


Figure 4.12 *Titles at Princeton University* ***Figure 4.13*** *Titles at Yale University*



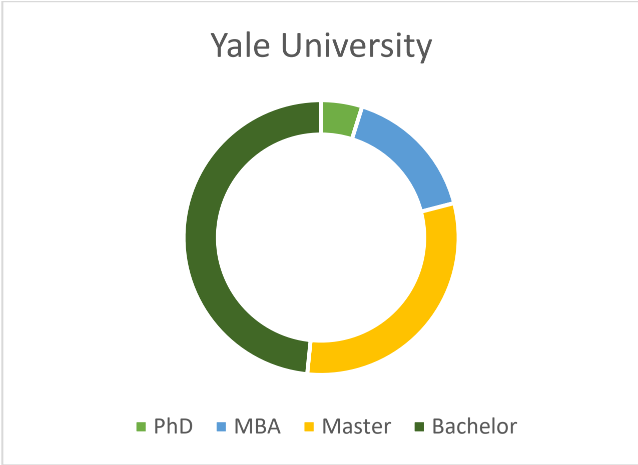
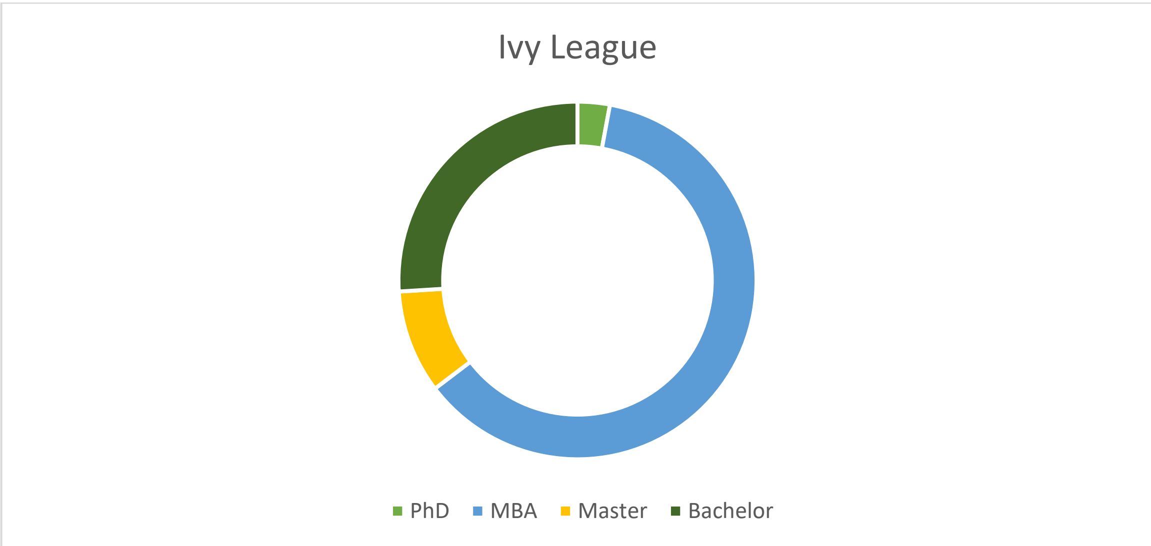


Figure 4.14 *Titles at Ivy League*



In the end it's possible to say that the 61,75% of the titles achieved at Ivy League is represented by a MBA, 25,98% of the time by a Bachelor, 9,36% by a Master and only 2,91% of the time by a PhD.

So, a dummy variable was created, and it was called "Ivy_league": its value is equal to 1 if the investor studied in one of these eight universities, irrespectively to the title achieved, 0 otherwise. This was done using the function *gen*:

```
gen Ivy_League = education == "Brown University" | education == "Columbia University" | education == "Cornell University" | education == "Dartmouth College" | education == "Harvard University" | education == "University of Pennsylvania" |education == "Princeton University" |education == "Yale University"
```

4.3 Professional Background

To explain the professional background statistics, the relevant variables that will be considered from the file "Job+people_partner_list" are:

- Partner_uuid
- Gender
- _0_org
- Job_type
- Country_code
- Continent
- Featured_job_type

This analysis will consider either the type of role covered, either the country in which the partners work. The comparison will be made considering the investment funds, either social either traditional, neglecting the corporations.

From a geographic perspective, investments are made in a very vast way, covering 90 distinct countries. Moreover, it's possible to say that two investments out of three are made in United States, its percentage amounts to 67,43%.

To look at the picture from a more general perspective, it can be observed what happen in the continents. Obviously, the role played by North America is the predominant one, but, at an aggregate level, Europe and Asia are relevant too. The cumulate sum of these continents amounts to 98,2%. All the details are contained in the following table (table 4.14).

Table 4.14 *Investments by continent*

Continent	*Frequency*	*Percentage*	*Cumulate*
Africa	157	0,54	0,54
Asia	2874	9,91	10,45
Central America	9	0,03	10,48
Europe	5370	18,51	28,99
North America	20242	69,78	98,77
Oceania	221	0,76	99,53
South America	137	0,47	100
Total	*29010*	*100*	

Now it's time to look at the job covered by partners. There are two main variables that contain this information and they are called "featured_job_type", that tells accurately the main role played within the company , and "job_type" that can be considered as a variable that aggregates the values contained in "featured_job_type".

For example, the value "advisor" of the variable "job_type" contains 325 different values the variables "featured_job" like "advisor directory", "board member advisor", "financial analyst" and so on.

The distinct values of "featured_job_type" totally amount to 2295, while there are 5 distinct values related to "job_type", that will be used in the following analysis.

All the details related to job type are contained in the table below (table 4.15).

Table 4.15 *Job type*

Job type	*Frequency*	*Percentage*	*Cumulate*
Advisor	1367	4,71	4,71
Board member	3965	13,67	18,38
Board observer	209	0,72	19,10
Employee	5849	20,16	39,26
Executive	17620	60,74	100
Total	*29010*	*100*	

The table tells that the most frequent job is represented by executive, followed by employee, board member, advisor and then board observer. It will be shown in a while how the percentage and the ranking will change if the variables "gender" and "_0_org" are considered.

Let's see if there is any difference in the role covered by partners in case the investment is made in a social or in a traditional fund. Although the ranking is the same in both cases (thus executive, followed by employee, board member, advisor and then board observer), it is possible to see that the likelihood to cover the "board member" role is more or less double in case of traditional fund rather than in case of social fund (15,36% vs 8,34%).

The details are all contained in the following tables (table 4.16, table 4.17).

Table 4.16 *Job type for traditional funds*

Job type	*Frequency*	*Percentage*	*Cumulate*
Advisor	983	4,46	4,46
Board member	3382	15,36	19,82
Board observer	195	0,89	20,71
Employee	4437	20,15	40,86
Executive	13021	59,14	100
Total	*22018*	*100*	

Table 4.17 *Job type for social funds*

Job type	*Frequency*	*Percentage*	*Cumulate*
Advisor	384	5,49	5,49
Board member	583	8,34	13,83
Board observer	14	0,20	14,03
Employee	1412	20,19	34,22
Executive	4599	65,78	100
Total	*6992*	*100*	

Let's see if there is any difference in the role covered by partners in case the investment is made by a male partner or by a female partner. Although the ranking is the same in both cases (thus executive, followed by employee, board member, advisor and then board observer), unsurprisingly male partners cover more relevant roles rather than female partners. This is clear from the data in fact the percentage of "board member" for male partners is 14,01% against the 10,62% for female, the percentage of "employee is 19,91% for male partners against 22,39% for female partners.

All the details are contained in the following tables (table 4.18, table 4.19).

Table 4.18 *Job type for male partners*

Job type	*Frequency*	*Percentage*	*Cumulate*
Advisor	1215	4,66	4,66
Board member	3651	14,01	18,68
Board observer	199	0,76	19,44
Employee	5187	19,91	39,35
Executive	15801	60,65	100
Total	*26053*	*100*	

Table 4.19 *Job type for female partners*

Job type	*Frequency*	*Percentage*	*Cumulate*
Advisor	152	5,14	5,14
Board member	314	10,62	15,76
Board observer	10	0,34	16,10
Employee	662	22,39	38,48
Executive	1819	61,52	100
Total	*2957*	*100*	

While talking about job type, finally let's see how results change if considering at the same time the gender and the kind of fund. In all the case, thus for traditional male partners, for social male partners, for traditional female partners and for social female partners, the ranking is always the same (executive, employee, board member, advisor, board observer).

In particular, considering the job type "board member", an interesting thing is that the female partners have an high likelihood to cover this role in case of traditional fund (11,30%), rather than in case of a social fund (8,9%).

All the details are contained in the tables below (table 4.20, table 4.21).

***Table 4.20** Job type for male partners, in case of social and traditional fund*

Job type	*Traditional*	*Social*	*Total*
Advisor	891	324	1215
Board member	3144	507	3651
Board observer	188	11	199
Employee	3944	1243	5187
Executive	11748	4053	15801
Total	*19915*	*6138*	26053

***Table 4.21** Job type for female partners, in case of social and traditional fund*

Job type	*Traditional*	*Social*	*Total*
Advisor	92	60	152
Board member	238	76	314
Board observer	7	3	10
Employee	493	169	662
Executive	1273	546	1819
Total	*2103*	*854*	2957

To end this paragraph, it's interesting to consider more than two variables at the same time, thus relating the kind of job to the kind of fund, to the gender and to the geographical information.

As stated before, the more relevant continents are North America and Europe. Looking at data, it's possible to say that a female partner has an higher probability to cover a board set in case she is working for an American investment fund rather than in the case in which she is working for an European investment fund (11,59% vs 6,56%) and the same is true for male partners (15,33% for Americans vs 10,17% for Europeans).

All the details are contained in the following tables (table 4.22, table 4.23).

Table 4.22 *Job type for American partners, by gender*

Job type	*Female*	*Male*	*Total*
Advisor	127	976	1103
Board member	247	2778	3025
Board observer	8	173	181
Employee	468	3586	4054
Executive	1280	10599	11879
Total	*2130*	*18112*	20242

Table 4.23 *Job type for European partners, by gender*

Job type	*Traditional*	*Social*	*Total*
Advisor	13	147	160
Board member	35	492	527
Board observer	1	19	20
Employee	141	975	1116
Executive	343	3204	3547
Total	*533*	*4837*	5370

4.4 Investment statistics

This one will be the more technical paragraph of this chapter, as it will analyze deeply which is the investment round, which is the decade in which the investment partner started to work in a social investment fund and which is the field in which operates the startup that received the investment. So, to explain the investment statistics, both the files were used and the relevant variables that will be considered are:

- Investor_uuid
- Partner_uuid
- Gender
- _0_org
- Cluster_started_on
- Fundinground
- Fundfield

Depending on the type of investor, the investment round may change. For example looking at the differences between Business Angels and Venture Capitalists, the literature evidence (Ughetto, 2019) tells that while for the former one the “early-stage” represents the “stage focus”, for the latter the stage focus is more on “expansion” and “later-stage”, for different reasons like the one related to the investment riskiness.

There is a technical glossary related to “stage focus”.

The main categories are (https://support.crunchbase.com/hc/en-us/articles/115010458467-Glossary-of-Funding-Types, s.d.):

- **Angel**: An angel round is typically a small round designed to get a new company off the ground. Investors in an angel round include individual angel investors, angel investor groups, friends, and family.
- **Pre-Seed**: A Pre-Seed round is a pre-institutional seed round that either has no institutional investors or is a very low amount, often below $150k
- **Seed:** Seed rounds are among the first rounds of funding a company will receive, generally while the company is young and working to gain traction. Round sizes range between $10k–$2M. A seed round typically comes after an angel round (if applicable) and before a company’s Series A round.
- **Venture Round**: Venture funding refers to an investment that comes from a venture capital firm and describes Series A, Series B, and later rounds. This funding type is used for any funding round that is clearly a venture round but where the series has not been specified.
- **Series A and Series B:** rounds are funding rounds for earlier stage companies and range on average between $1M–$30M.

- **Series C:** rounds and onwards are for later stage and more established companies. These rounds are usually $10M+ and are often much larger.
- **Equity Crowdfunding:** Equity crowdfunding platforms allow individual users to invest in companies in exchange for equity. Typically, on these platforms the investors invest small amounts of money, though syndicates are formed to allow an individual to take a lead on evaluating an investment and pooling funding from a group of individual investors.
- **Private equity:** A private equity round is led by a private equity firm or a hedge fund and is a late stage round. It is a less risky investment because the company is more firmly established, and the rounds are typically upwards of $50M.
- **Convertible note:** A convertible note is an 'in-between' round funding to help companies hold over until they want to raise their next round of funding. When they raise the next round, this note 'converts' with a discount at the price of the new round.
- **Debt Financing**: In a debt round, an investor lends money to a company, and the company promises to repay the debt with added interest.

Of course, these are the main categories in general terms, but it's important to say that there are also other ways to invest in a fund (for example, there is also Series D, Series E, Series F, Grants and so on).

Thus, it is clear that what differentiates one category from another one it's not merely a temporal consideration but it's also in terms of amount of money invested.

Now let's consider the current dataset.

The main categories are represented by "Series B" (21,50%), followed by "Series B" (21,15%), "Seed Round" (13,80%), "Series C" (13,56%), "Venture Round" (10,73%). Thus, it's clear that four investments out of five are made during one of these five rounds (the cumulate sum is equal to 80,74%), while other categories present in this dataset, like "Angel" or all the other Series, seem to be more negligible.

All the details about the statistics on the investment rounds based on the data present in the current dataset are shown in the following table (table 4.24).

Table 4.24 *Investment Round*

Funding round	*Frequency*	*Percentage*	*Cumulate*
Angel Round	35	0,58	0,58
Convertibile note	181	3,02	3,61
Corporate round	6	0,10	3,71
Debt financing	88	1,47	5,18
Equity crowdfunding	2	0,03	5,21
Funding round	15	0,25	5,46
Grant	13	0,22	5,68
Pre Seed round	30	0,50	6,18
Private equity round	148	2,47	8,65
Secondary market	7	0,12	8,77
Seed round	826	13,80	22,57
Series A	1266	21,15	43,72
Series B	1287	21,50	65,22
Series C	812	13,56	78,78
Series D	396	6,62	85,40
Series E	147	2,46	87,85
Series F	59	0,99	88,84
Series G	25	0,42	89,26
Series H	1	0,02	89,27
Venture round	642	10,73	100
Total	5986	100	

On the other side, to know exactly when the investment partner started working for a social investment fund, the variable "Cluster_started_on" should be used. Starting from 1981, it's possible to see that the investments increased overtime, as shown in the following table (table 4.25).

Table 4.25 *Number of investments by decade*

Decade	***Frequency***	***Percentage***	***Cumulate***
1981-1990	797	4,93	4,93
1991-2000	2755	17,05	21,99
2011-2010	5461	33,80	55,79
2011-2020	7143	44,21	100
Total	*16156*	*100*	

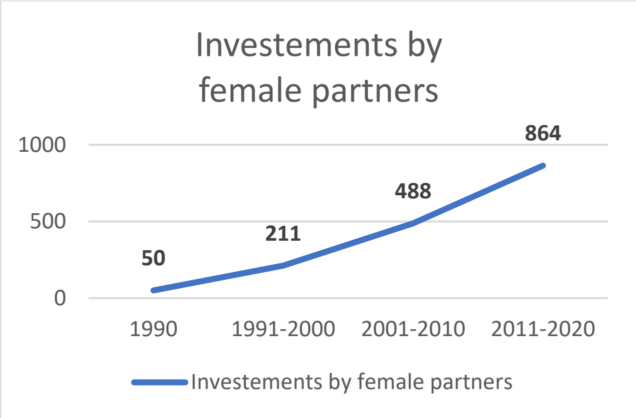
How will change the result if the decade and the type of fund are considered at the same time?

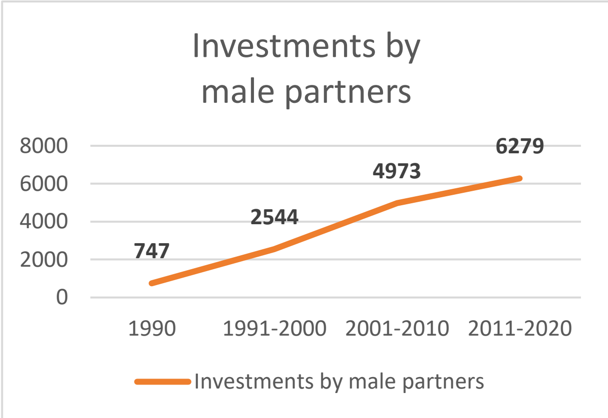
The answer is shown in the following graphs (figure 4.15, figure 4.16).

Even if the order of magnitude of female and male partners is different, the growth pattern in the last four decades is different, in fact one can see that the growth curve is more than linear for female partners (on the left) while the growth rate (so the time derivative of growth) is going down for male partners (on the right).

***Figure 4.15** , Investments by female partners, by decade*

***Figure 4.16** , Investments by male partners, by decade*

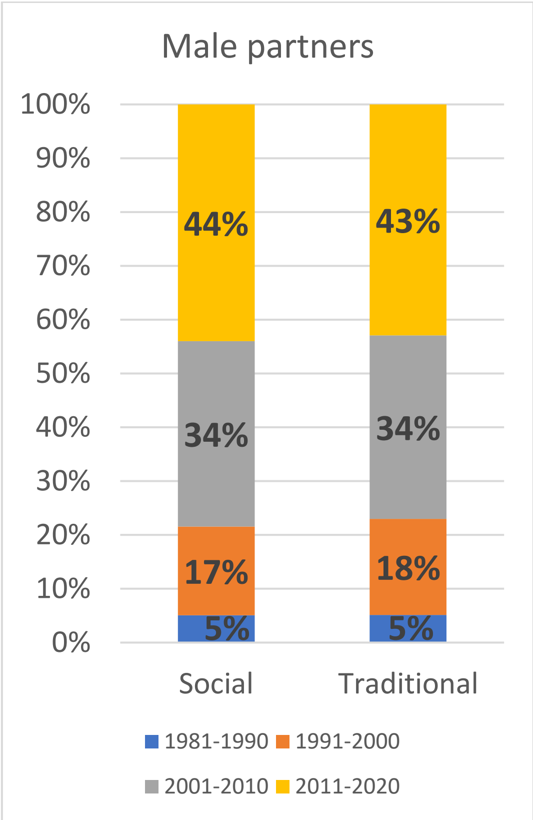


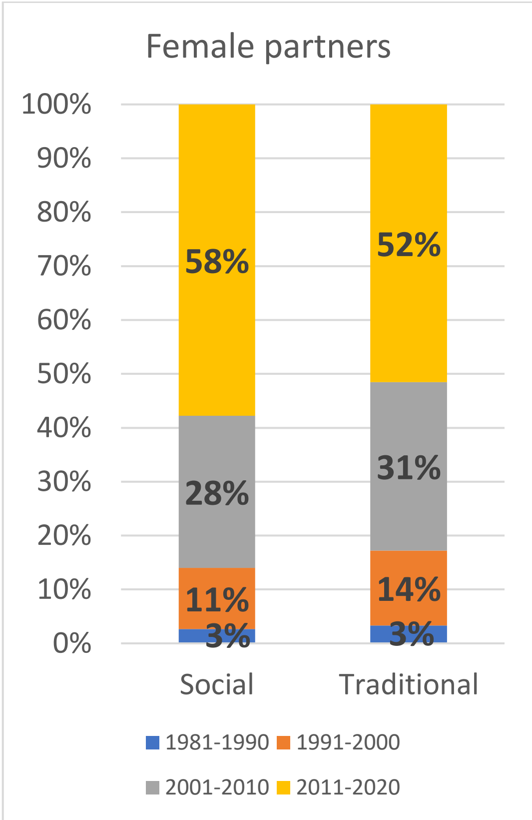


In addition to seeing how much social investments have grown over last decades, it would be interesting to understand if the attention that men or women pay to social or traditional funds is the same over years. In this regard, the following two graphs are useful (figure 4.17, figure 4.18). In fact, it's possible to see that, while the male partners show the same percentage of interest in social and traditional fund, over the years the interest of female partners to social funds increased more than the one related to traditional funds.

Figure 4.17 *, Investments by female partners over decades, percentage*

Figure 4.18 *Investments by male partners over decades, percentage*



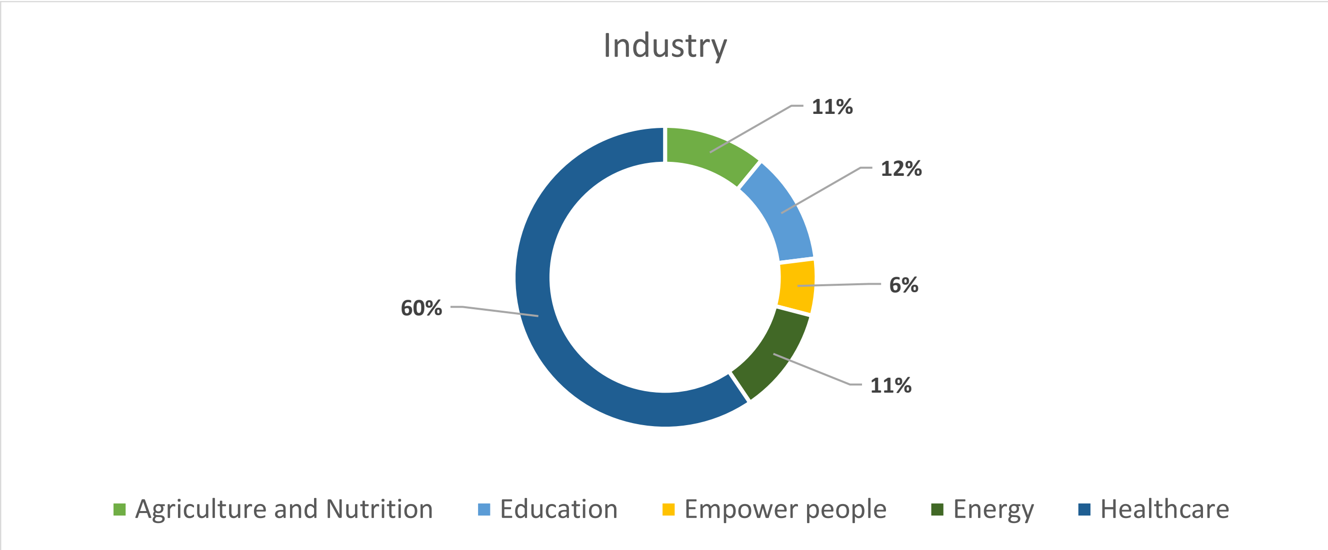


The analysis will now take into consideration the field in which the startup operates. As this information is available for the social male partners and the female partners, the comparison to be fair will be between social male and social female partners.

In general terms, the results show that there are 5 macro categories, that follow the following ranking: Healthcare, Education, Energy, Agriculture and nutrition, Empower people.

The percentages are shown in the following graph (figure 4.19).

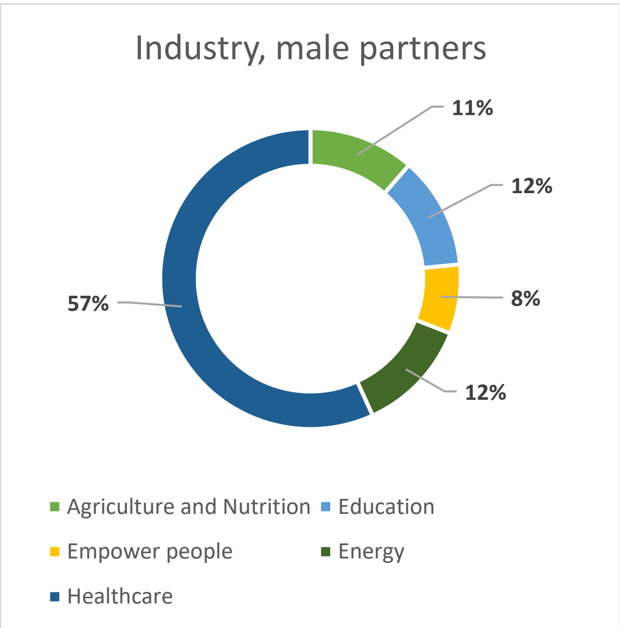
__Figure 4.19__ Industry

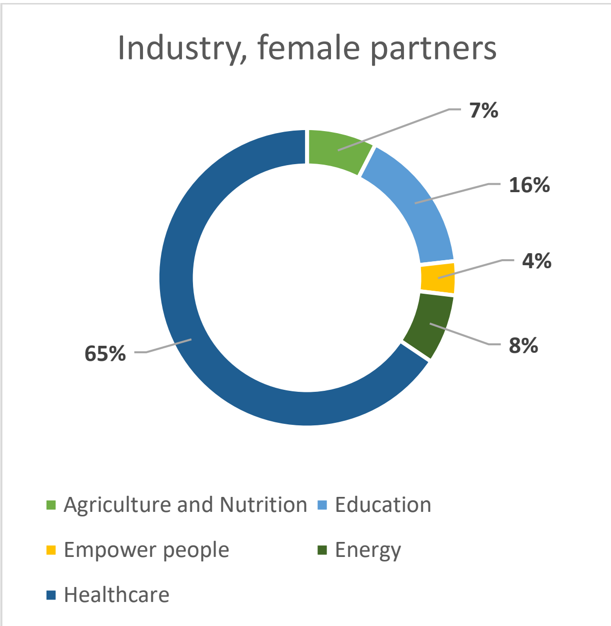


The following graphs show the percentages when the industry and the gender are considered at the same time. Looking at the percentages, it seems that while male partners invest more in Energy and Agriculture and Nutrition, female partners invest more in Education, Empower people, and Healthcare.

All the details are shown in the following graphs (figure 4.20, figure 4.21).

__Figure 4.20__ Industry for male partners *__Figure 4.21__ Industry for female partners*





Another step that may be done is to consider industry and investment round. About investment round, there are 23 different categories but the cumulate sum of the four most frequent categories (Series A, Series B, Series C, Seed round, Venture round) is about 80,74%. So, this step will take into consideration only them.

The number of investment made during “Series A” totally amounts to 1266, while the ones during “Series B” to 1287, during “Series C” to 812, during “Seed Round” to 826 and then during “Venture Round” to 642.

The category “Agriculture and Food” goes from a value of 7,86% in case of “Series C” to a more significant 28% in case of “Seed Round”.

The industrial category “Education” goes from a value of 9,1% in case of “Venture Round” to a more significant 29% in case of “Series A”.

The industrial category “Empowering people” goes from a value of 9,9% in case of “Seed Round” to a more significant 22,17% in case of “Series A”.

The industrial category “Energy” goes from a value of 7,33% in case of “Seed Round” to a more significant 21,43% in case of “Series B”.

The industrial category “Healthcare” goes from a value of 10,58% in case of “Venture Round” to a more significant 23,47% in case of “Series B”.

4.5 Mixed Statistics

Now that there is a clearer picture of the information in the dataset, it's possible to go even deeper with the analysis. While in the previous paragraphs the variables were analyzed dividing them into macro topics (general, educational, professional and investment information), at this time of the analysis many variables that belong to different macro topics will be consider at the same time, like:

- Partner_uuid
- Gender
- _0_org
- Degree_tot
- Ivy_League
- Job_type

How will the type of title influence the type of working role?

To answer, it's possible to consider also the a third variable related to gender. If it's true that partners holding a MBA have higher probability to cover a seat in the board, the situation is different in case the partner is male or female. In fact, while a male partner can be considered a board member 8,29% of the times, a female partner will do the same but 10,64% of time.

About PhD, that is consider as the highest title by ISCED (as stated in the paragraph 3.2), either male either female partners will cover more likely the executive role, showing also the same percentage (65,87% for male and 66,66% for female).

All the details are shown in the following tables (table 4.26, table 4.27).

Table 4.26 *Type of role by titles, male partners*

Job type	*Bachelor*	*MBA*	*Master*	*PhD*	*Total*
Advisor	95	85	64	20	264
Board member	140	160	86	23	409
Board observer	3	3	4	0	10
Employee	343	425	156	72	996
Executive	1075	1256	559	222	3112
Total	*1656*	*1929*	869	337	4791

Table 4.27 *Type of role by titles, female partners*

Job type	*Bachelor*	*MBA*	*Master*	*PhD*	*Total*
Advisor	12	21	10	8	51
Board member	16	41	10	1	68
Board observer	1	1	1	0	3
Employee	31	81	20	20	152
Executive	101	241	60	58	460
Total	*161*	*385*	101	87	734

In the paragraph related to educational statistics, a new variable was introduced and called "Ivy_Leaugue". Its value is equal to 1 if the title has been obtained in one of the most prestigious American university, 0 otherwise. Using this variable now it can be useful in understanding if partners graduated there have higher probability to cover relevant role, or if there is a particular field in which they prefer investing.

Let's consider the job irrespectively to the level of title obtained. The probability to be a board member seems to be the same (9,31% for Ivy League alumni, 8,37% for the others), and so does the probability to cover executive role, in fact it's equal to 62,60% for Ivy League alumni against 65,43% for the others. All the details are shown in the following table (table 4.28).

Table 4.28 *Type of role, Ivy League alumni*

Job type	*Ivy League alumni*	*Other alumni*	*Total*
Advisor	243	72	315
Board member	335	142	477
Board observer	9	4	13
Employee	796	352	1148
Executive	2618	954	3572
Total	*4001*	*1524*	5525

Finally, looking at Ivy League alumni and not Ivy League alumni and the industry in which they invested is possible to highlight some differences. In fact, although the percentages are quite similar, the ranking related to their interest is different. Ivy League alumni prefer investing in healthcare, followed by education, energy, agriculture & food and then by empower people. Non-Ivy League alumni prefer investing in healthcare, followed by agriculture & food, education, energy and then by empower people.

All the details are shown in the following tables (table 4.29, table 4.30).

Table 4.29 *Fund field, Ivy League alumni*

Fundfield	*Frequency*	*Percentage*	*Cumulate*
Agriculture & nutrition	129	8,26	8,26
Education	220	14,09	22,36
Empower people	125	8,01	30,37
Energy	208	13,32	43,69
Healthcare	879	56,31	100
Total	*1561*	*100*	

Table 4.30 *Fund field, not-Ivy League alumni*

Fundfield	*Frequency*	*Percentage*	*Cumulate*
Agriculture & nutrition	532	12,02	12,02
Education	525	11,86	23,89
Empower people	299	6,76	30,64
Energy	501	11,32	41,97
Healthcare	2568	58,03	100
Total	*4425*	*100*	

4.6 Social expert partners statistics

As stated at the beginning of this chapter, while talking about the general statistics, it was stated that an interesting thing can be represented by diving into few groups the social partners based on the times they have invested in a social funds. Having used the Stata function "`duplicates report`", it was possible to find out this information, that is synthetized in the following table (table 4.31).

Table 4.31 *Number of social investments related to number of partners*

Number of investments	Total partners	Percentage
1	584	40,46
2	248	17,21
3	167	11,59
4	103	7,15
5	60	4,16
6	48	3,33
7	37	2,57
8	29	2,01
9	30	2,08
10	20	1,39
11	16	1,11
12	12	0,83
13	9	0,62
14	10	0,69
15	5	0,35
16	10	0,69
17	3	0,21
18	2	0,14
19	4	0,28
20	10	0,69
21	6	0,42
22	1	0,07
23	1	0,07
25	3	0,21
26	1	0,07
27	3	0,21
28	1	0,07
29	1	0,07
30	3	0,21
31	2	0,14
32	2	0,14
37	1	0,07
40	2	0,14
41	1	0,07
43	2	0,14
68	1	0,07
69	1	0,07
76	1	0,07
81	1	0,07
105	1	0,07

Total	1441	100

Although the great majority is represented by partners who made a single investment, there is also another important percentage related to the ones that invested at least twice, they represent the 59% of social partners.

The main categories identified are:

- Basic social partner
- Intermediate social partner
- Expert social partner

According to the identification of these categories, three different dummy variables were created. They are called "basic_social", "intermediate_social" and "expert_social".

The value of "basic_social" is equal to 1 if the partner invested only once in a social fund, 0 otherwise.

The value of "intermediate_social" is equal to 1 if the partner invested more than once and less than 10 times, 0 otherwise.

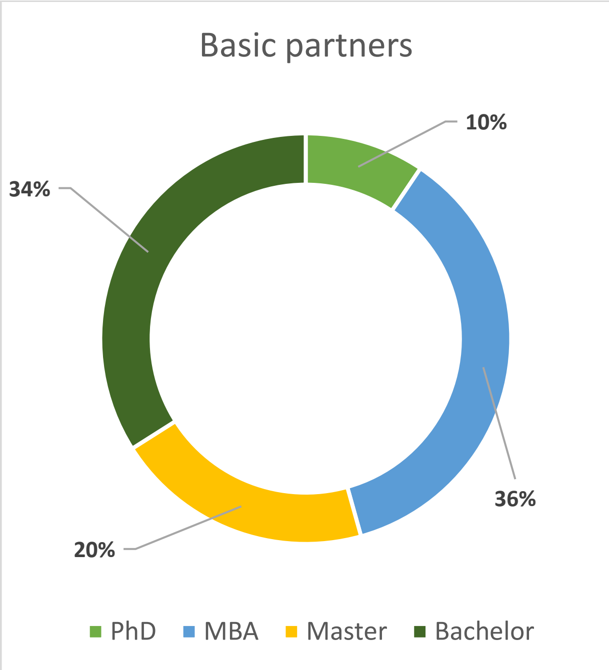
The value of "expert_social" is equal to 1 if the partner invested at least ten times in a social fund, 0 otherwise.

Once created these new variables, they can be combined with variables belonging to other categories (education, profession, investment) to see if there are some trends.

From an educational perspective, it's quite clear that the social expert investors are the ones that have a higher education level, in fact while the percentage related to Bachelor decreases, the ones related to PhD and MBA increases. All the details about this are shown in the following graphs (figure 4.22, figure 4.23, figure 4.24).

Figure 4.22 *Education, basic partners*

Figure 4.23 *Education, intermediate partners*



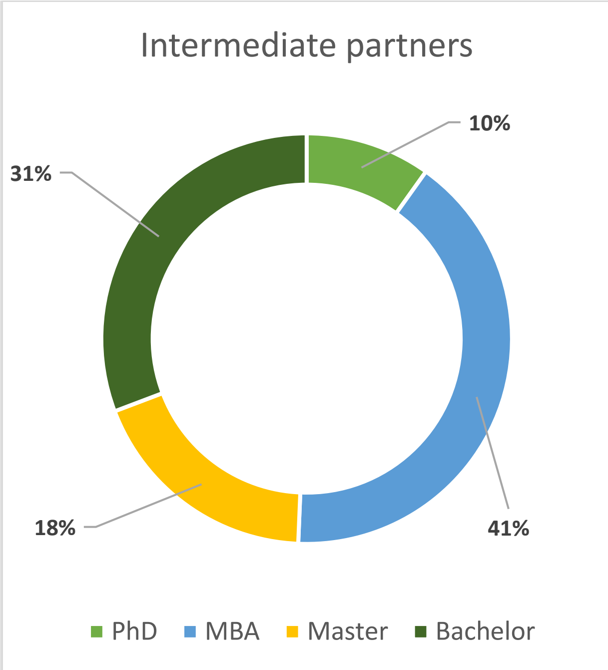
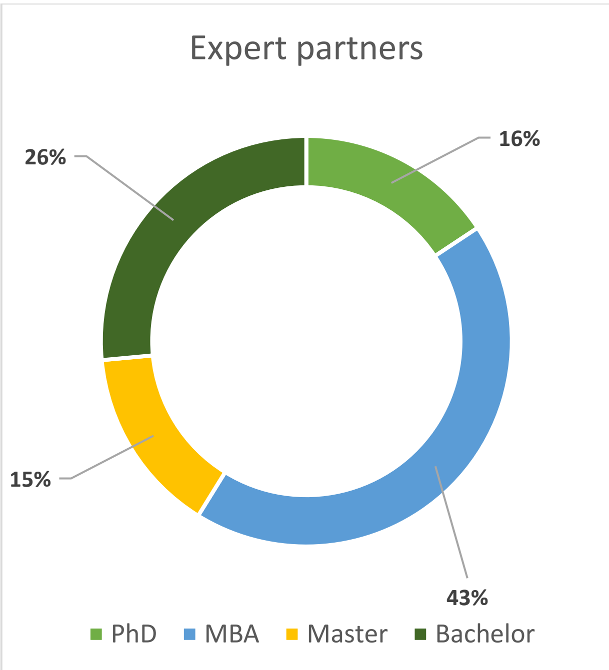


Figure 4.24 *Education, expert partners*



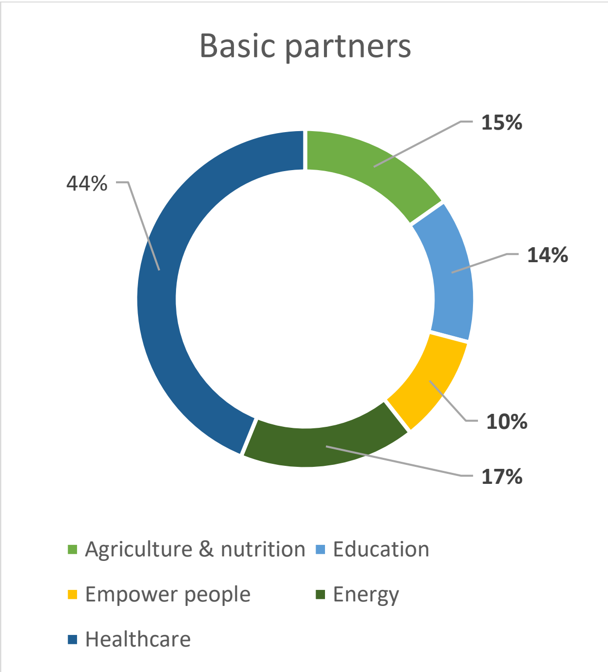
An interesting issue point could be to see, if the personal investment history of partner is able to drive him in his investment choice.

From an investment perspective, it’s quite clear that the field in which the three categories (basic, intermediate and expert partners) invest differs. If the percentages are quite similar for basic and intermediate partners, they change a lot for the expert partners, that, in particular, prefer a lot investing in healthcare field.

All the details are shown in the following graphs (figure 4.25, figure 4.26, figure 4.27).

Figure 4.25 *Industry, basic partners*

Figure 4.26 *Industry, intermediate partners*



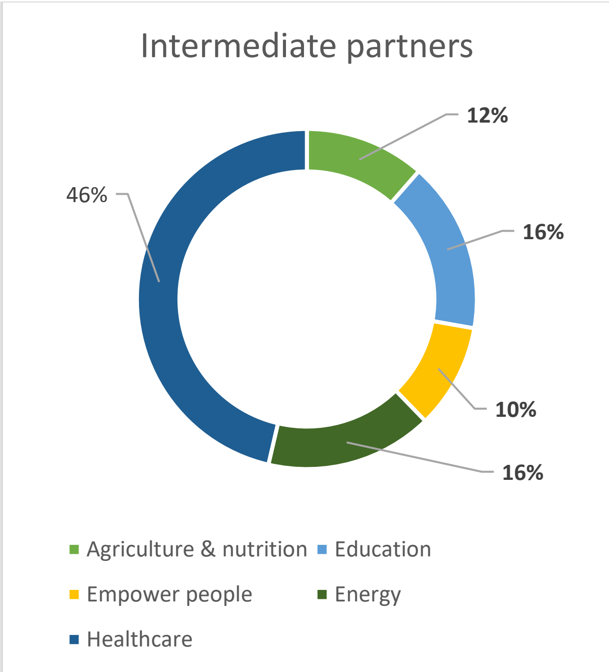
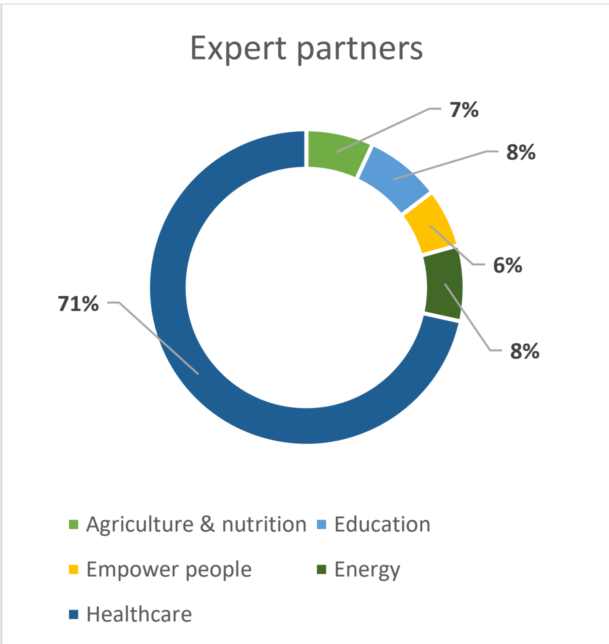


Figure 4.27 *Industry, expert partners*



5 Correlation matrix and final statistics

5.1 Correlation matrix

Now, to see if there are other mixed statistics that should be highlighted, it's time to draw the correlation matrix, that is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. The correlation value present on the diagonal is always equal to 1, as it shows that each variable always perfectly correlates with itself (https://www.displayr.com/what-is-a-correlation-matrix/#:~:text=A%20correlation%20matrix%20is%20a,a%20diagnostic%20for%20advanced%20analyses, s.d.).

In this case, the correlation matrix will take into consideration only the social investments and social partners, that's why there won't be the dummy variable "social" in this matrix, because it's setting by default equal to 1. The chosen variables amount to twelve and, in particular, they are:

- Sex, related to gender,
- Funding round, called FuRound,
- Job type, called job_type,
- Fund field, called FuField,
- Education field, called EduField,
- Type of title, called EduTitle,
- Ivy league, called Ivy_League,
- Education country code, called EduConCode,
- Investment country code, called InConCode,
- The dummy variable basic social investor, called basic_social,
- The dummy variable intermediate social investor, called intermediate_social,
- The dummy variable expert social investor, called expert_social.

Setting the value of α equal to 0,11, there are around ten values that suggests the existence of a correlation. Moreover, while there are seven variables that seem to be correlated (fund field, education field, education title, Ivy league, education country code, investment country code and expert social investor), the remaining ones seem to be not correlated (funding round, partners' gender, job type, basic and intermediate social investors).

All the details are shown in the following table (table 5.1).

Table 5.1 *Correlation matrix*

	FuRound	sex	Job_type	FuField	EduField	EduTitle	Ivy_Le~e
FuRound	1.0000						
sex	0.0616	1.0000					
Job_type	0.0465	0.0455	1.0000				
FuField	0.0830	-0.0352	-0.0405	1.0000			
EduField	0.0449	0.0204	0.0170	0.1327	1.0000		
EduTitle	0.0452	-0.1428	0.0004	0.1306	0.5402	1.0000	
Ivy_League	-0.0050	-0.0314	-0.0166	0.0217	0.1794	0.1158	1.0000
EduConCode	0.0513	-0.1200	0.0436	0.0546	0.5335	0.5417	0.3814
InConCode	-0.0048	-0.0224	0.0117	0.0824	0.1062	0.0670	0.1137
basic_social	-0.0841	-0.0650	-0.0313	-0.0575	-0.1039	-0.1201	-0.0795
intermedia~l	0.0607	0.0438	0.0125	-0.0073	0.0545	0.0740	0.0535
expert_soc~l	0.0403	0.0366	0.0336	0.1172	0.0874	0.0808	0.0452

	EduCon~e	InConC~e	basic_~l	interm~l	expert~l
EduConCode	1.0000				
InConCode	0.2615	1.0000			
basic_social	-0.1572	-0.0970	1.0000		
intermedia~l	0.1070	0.0269	-0.8494	1.0000	
expert_soc~l	0.0871	0.1257	-0.2440	-0.3045	1.0000

As some of the correlations shown here were already spotted in the previous chapter, this chapter will mostly try to see if:

1. Culture affinity due to geography forces partners to prefer investing in the country in which they studied,
2. The choice related to investment field is based on a principle that favors fields linked to the partner's university career,
3. In case of social investors who have already invested more than once in a social startup, the choice is driven by the context of previous social experiences.

5.1 Final statistics

1. **Culture affinity due to geography forces partners to prefer investing in the country in which they studied.**

To evaluate the validity of this statement, variables related to geography were taken into consideration. Moreover, as the number of different countries exceeds 90 countries, this analysis will consider the information related to investment continent and to education continent.

Let's start from the analysis of the most significant continents of this dataset, thus Asia, Europe and North America.

Although there is not a unique trend for all the continents, it's possible to say that culture affinity plays an important role in case of Asian and North American alumni, in fact:

- In case of Asian alumni, they invested 80% of the time in Asia, 17% in North America and 3% of the time in Oceania,
- In case of North American alumni, they invested 87% of the time in their own continent and the remaining 13% in Asia, Europe, Oceania and South America.

The situation is different if European alumni are considered. As it's possible to say, only one European over three invested in his own continent (35%) while the greatest majority preferred investing in North America (56%). In this case Asia represents 6,7%, while the cumulate of Africa, Central America and Oceania totally amounts to 1%.

Are these results still valid if also the gender of the partner is taken into consideration? Not entirely. Although the results won't change in case of Asian and North American alumni, the same it's not valid for female European alumni: in this case two investors over three will prefer investing in Europe and only one investor over three will prefer investing in North America, contrary to the result previously obtained.

Now let's continue the analysis considering the other continents present in the dataset, and so Africa, Oceania and South America. In this case the results obtained differ one to each other in fact:

- In case of South American alumni, the totality of them invested in their own continent
- In case of African alumni, the totality of them invested in North America. In this case it's important to add that all these alumni graduated in South Africa, so maybe this choice is driven by language affinity,
- In case of Oceanian alumni, the investment choice is more various, in fact they invest either in their own country, either in others like Asia, Africa and North America.

All the details are contained in the following graphs (figure 5.1, figure 5.2). The horizontal axis refers to the education continent.

Figure 5.1 *Correlation between education and investment continents pt1*

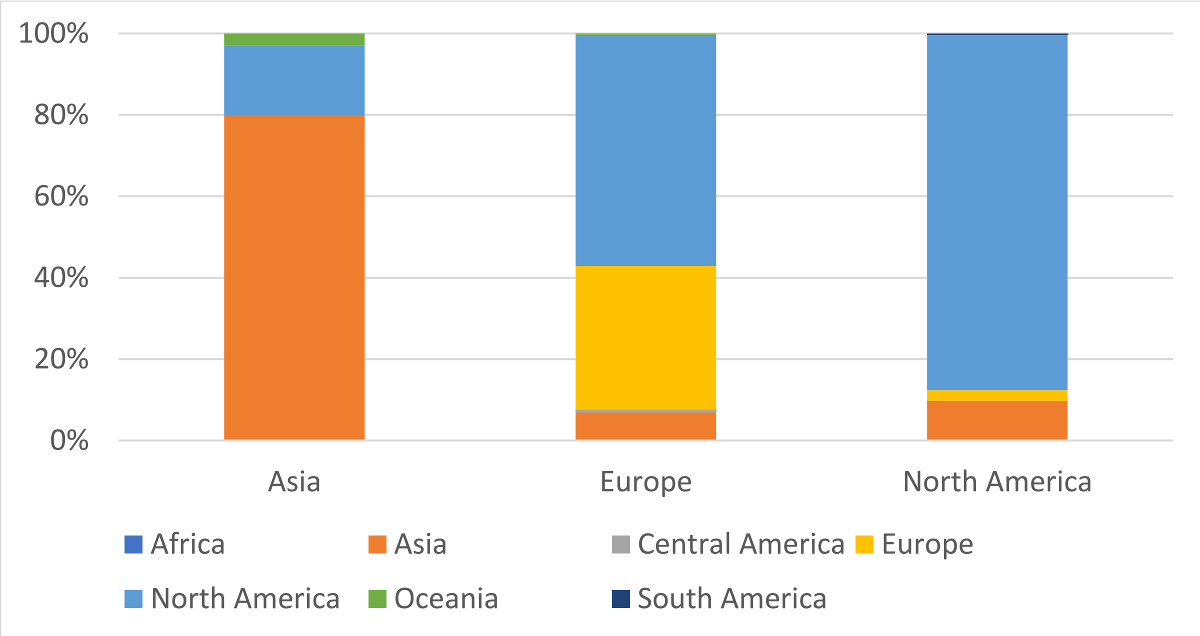
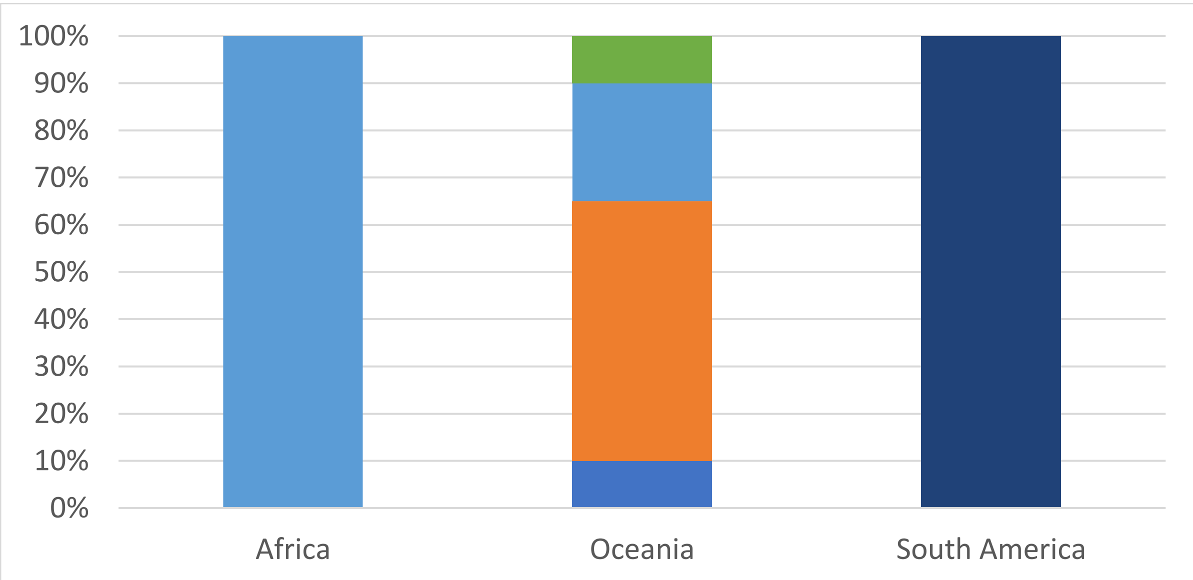


Figure 5.2 *Correlation between education and investment continents pt2*



2. **The choice related to investment field is based on a principle that favors fields linked to the partner's university career.**

In this case the aim is to understand if there is a relationship between fund and educational fields for all the investors for whom this information is available. The relationship was searched for the macro educational categories "Business Administration", "Economics & commerce", "Engineering", "Finance", "Humanistic studies", "Languages", "Law", "Marketing", "Political science & international relations" and "Scientific studies".

Some of the obtained results are not surprising. In fact, it's quite clear that for the alumni with a scientific background there is a high tendency to invest in fund committed in the scientific sector. Furthermore, the category "Healthcare" represents 75% in case of "Scientific studies" (a category that includes subjects like Biology, Chemistry, Physics, Medicine and so on) and a still significant 58% in case of "Engineering". It's also necessary to highlight that the fund category "Energy" reaches is maximum, that amount to 17%, for investors who studied engineering (in this case it's also important to remember that the 40% of the engineers present in the dataset studied electrical engineering). So, in case of scientific studies it's possible to state that there is a strong relationship with the investment field.

Now let's consider what happens in case of economic and managerial background. Differently from the result obtained before, these categories show a higher variety of interest. Moreover, although the percentage may change, the ranking is quit always the same: the greatest interest is represented by "Healthcare", followed by "Education" an "Agriculture and Nutrition", while the minor categories are represented by "Empower people" and "Energy".

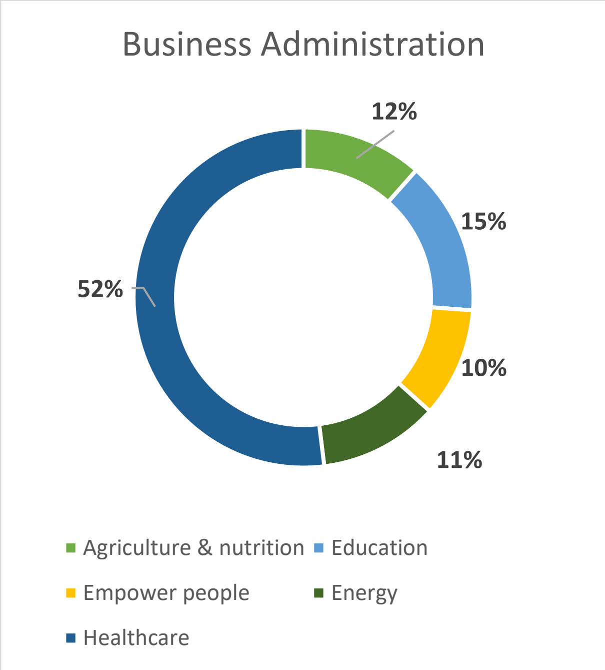
Considering now the remaining categories, it should be highlighted that the category "Education" has a high percentage both in case of "Law" and of "Humanistic studies", but it's precisely for the latter that it reaches its maximum value (25%), while the categories "Empower people" reaches its maximum value in case of "Marketing" alumni (16%). The surprising result is the one that involved investors who studied "Languages", in fact four in five investors invested in "Healthcare" and none of them invested in "Agriculture and nutrition" and "Energy".

In the end it's possible to say that, although there are some surprises, these are mostly related to the minor categories. From a general view it's possible to state that there is a quit clear relationship between education and investment field, that becomes even stronger in case of scientific backgrounds.

All the details are contained in the following graphs (figure 5.3, figure 5.4, figure 5.5, figure 5.6, figure 5.7, figure 5.8, figure 5.9, figure 5.10, figure 5.11, figure 5.12, figure 5.13).

Figure 5.3 *Investment field, "Business Administration" alumni*

Figure 5.4 *"Economics & Commerce" alumni*



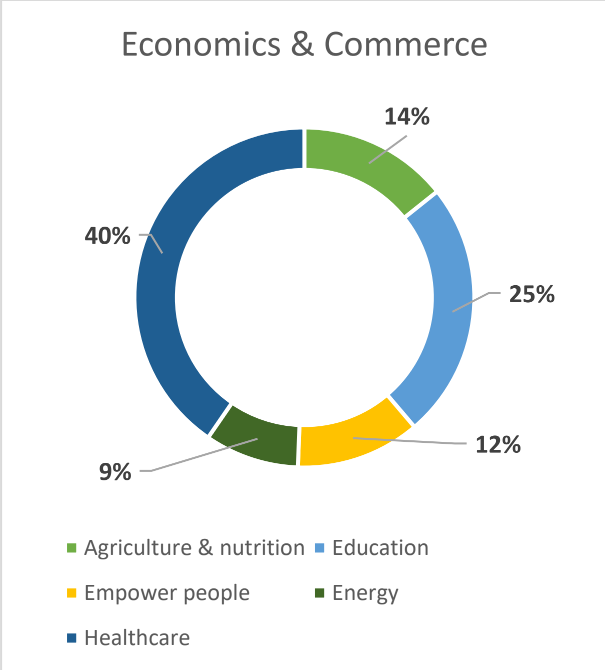
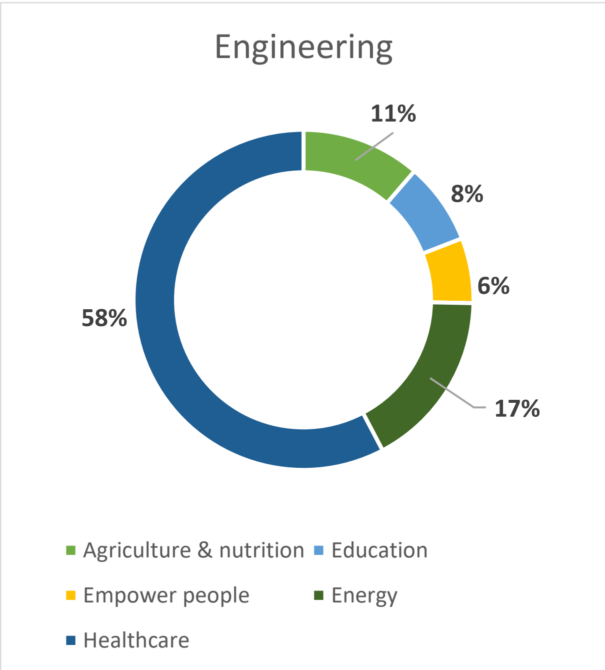


Figure 5.5 *Investment field, "Engineering" alumni*

Figure 5.6 *Investment field, "Finance" alumni*



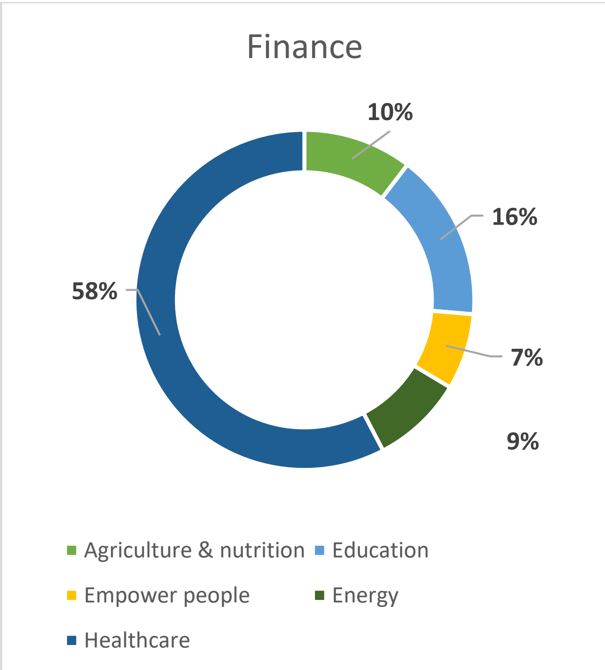
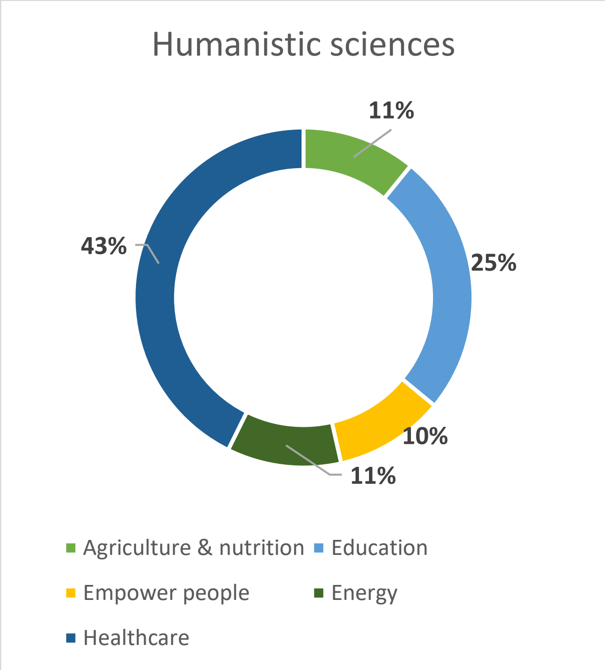


Figure 5.7 *Investment field, "Humanistic sciences" alumni*

Figure 5.8 *Investment field, "Languages" alumni*



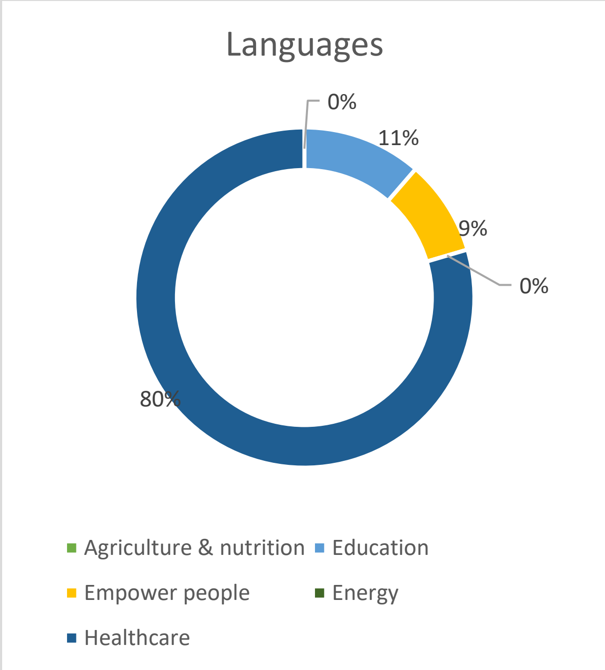
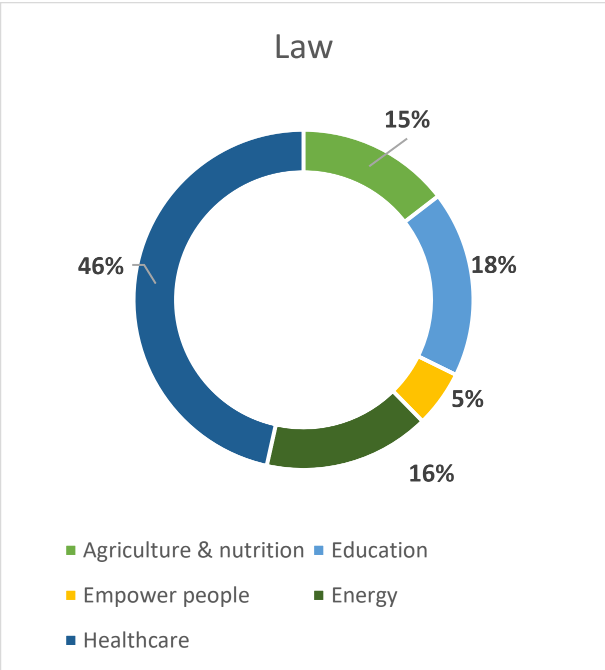


Figure 5.9 *Investment field, "Law" alumni*

Figure 5.10 *Investment field, "Management" alumni*



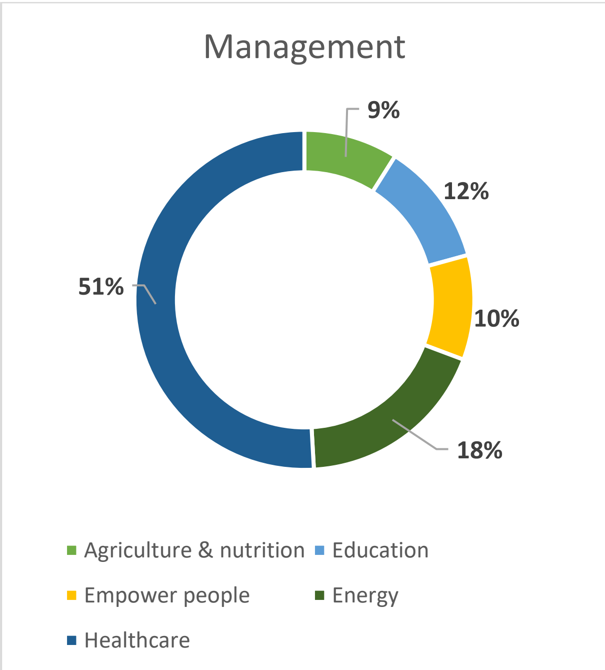
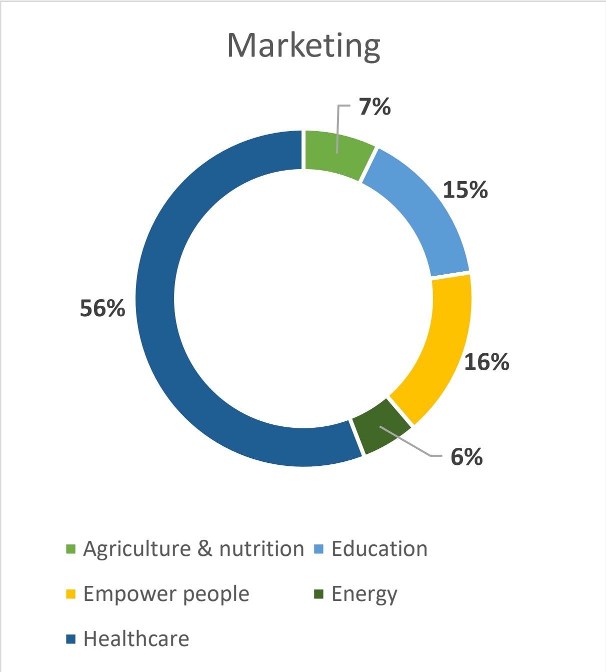


Figure 5.11 *Investment field, "Marketing" alumni*

Figure 5.12 *Graph 4.12, Investment field, "Political sciences & international sciences" alumni*



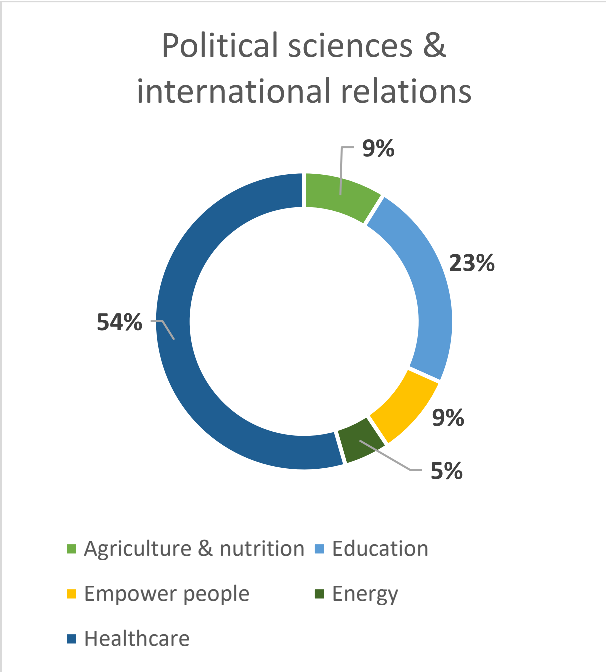
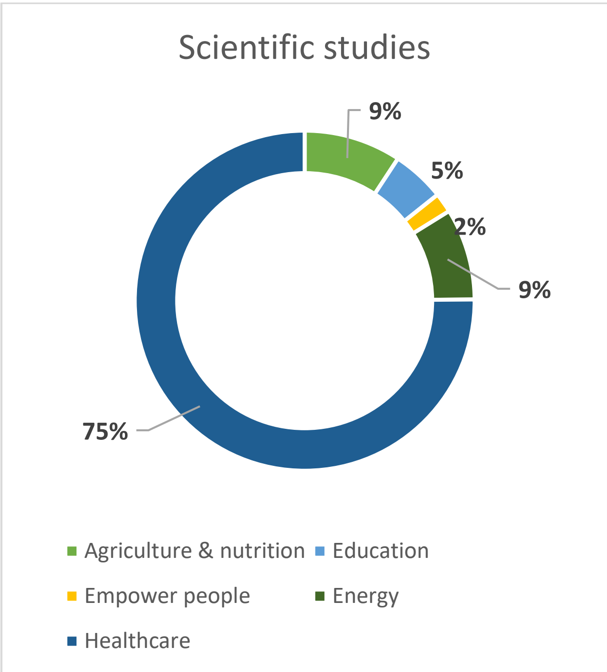


Figure 5.13 *Investment field, "Scientific studies" alumni*



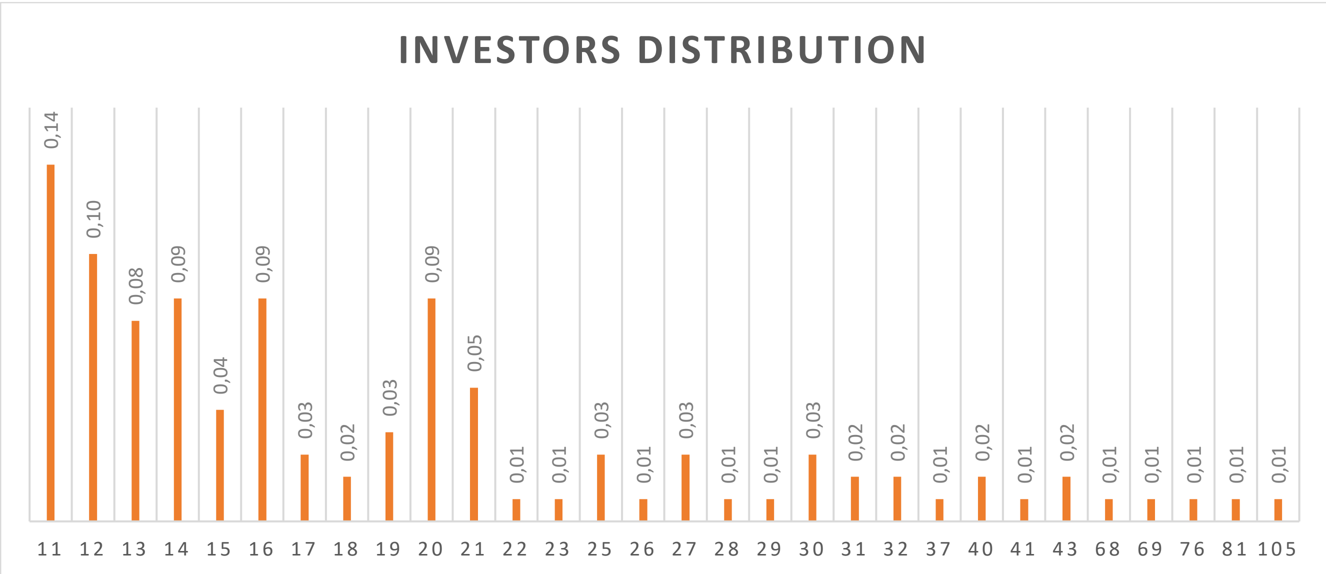
3. **In case of social investors who have already invested more than once in a social stratup, the choice is driven by the context of previous social experiences.**

To evaluate the validity of this statement, first of all it's mandatory to say that only the social partners considered expert will be taken into consideration, while the dummy variables basic and intermediate will be neglected. It's also important to remind that the social expert partners are the ones that invested more than ten times in a social startup and that the label "Investment field" can assume five different values: "Agriculture & Food", "Education", "Empower people", "Energy" and "Healthcare".

First of all, let's see how these investors are distributed. The following graph shows that the number of total investors tend to decrease while increasing the number of investments made. Moreover, 70% of investors made less than 21 investments, while the remaining 30% made more than 21 investments.

All the details are shown in the following graph (figure 5.14).

Figure 5.14 *Distribution of investors related to number of investments*



The following step was to understand whether the investment choice was driven by the context of previous social experience.

The first idea was to compute for each investor the variance of his investments. This idea implies to assign to each field a numerical value and then to compute the variance. The values were assigned following an alphabetic order, so "Agriculture & nutrition" = 1, "Education" =2, "Empower people" =3, "Energy" =4 "Healthcare" = 5, but the limitations of such methodology became apparent early on. In fact, in case two different partners who made the same number of investments are considered, although they invested in only two categories, their variance could be greatly different due to this values assignment (for example the variance of an investor that invested in Agriculture and

Healthcare would be greater than the variance of an investor who invested in Energy and Healthcare (1 and 5 vs 4 and 5).

Of course, the weakness of this method is the arbitrary assignment of the values: the scale in question is categorical and more specifically it's nominal (Franceschini, Galetto, & Maisano), thus the basic relations among objects is only about equality (we are only able to say if the fields are or not the same), but there is not an order.

To avoid dealing with the choice of assigning values, another index was built. It would be called "Experience index".

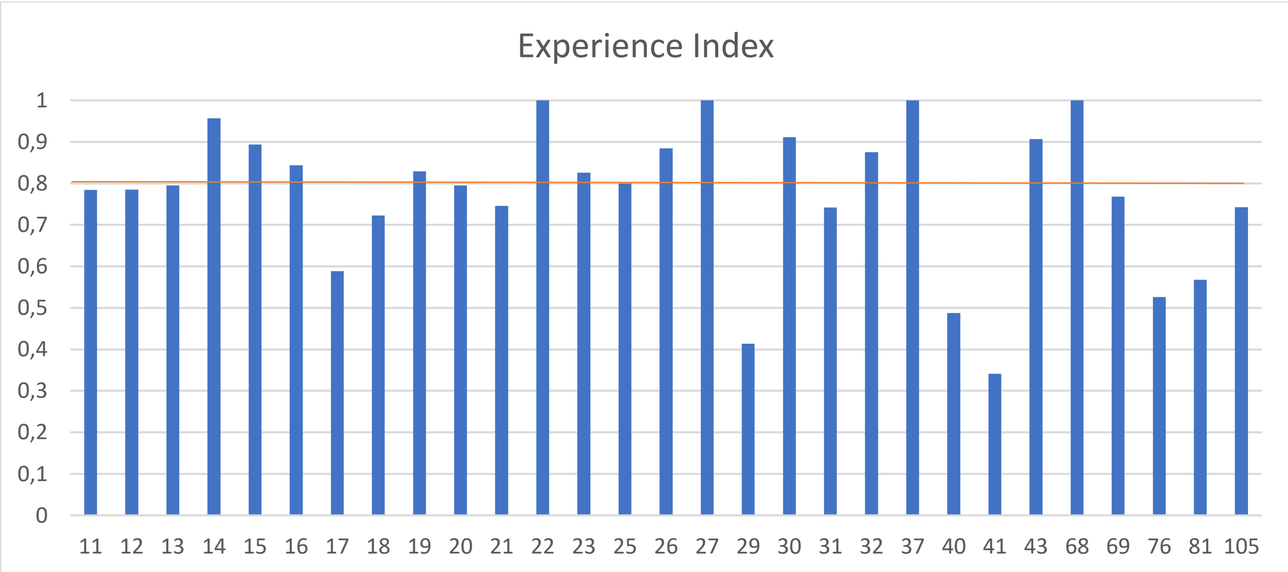
For each investor, the number of times invested in a given sector are counted and the sector that shows the highest value is identified. The "experience index" is then defined as the ratio between the sector where most of the investments have been made and the total number of investments. Of course, the higher the value of "experience index", the stronger the effect played by previous choices on future ones.

Moreover, is it possible to say that, considering all the social expert partners, on average, the "experience index" is equal to 0,80 thus the largest majority of investors are driven by their own previous experience. This statement is valid both for female partners and male partners.

Furthermore, 50% of the time the experience index is equal to 1; that means that one in two investors invested systematically in the same field.

All the details are shown in the following graph (figure 5.15).

Figure 5.15 *Experience index*



Finally, let's see if it's possible to link the experience index and the educational background. Looking at the values assumed by the "experience index" and at the educational background, it's possible to say that the value of the index is generally really high, often close to 1, in case the investor held a scientific title, and so it's possible to say that this investor is much more sectorial and he tends to favor always the same field. On the contrary, investors holding economic and managerial backgrounds tend to diversify more, investing in several fields. The same is true for categories like "humanistic studies", "law" and "marketing".

The values of the "experience index" related to the most significant education fields are shown in the following table (table5.2).

Table 5.2 *Experience index, education background*

Education field	Experience index
Biology	1
Business Administration	0,70
Chemistry	0,97
Economics & commerce	0,70
Finance	0,76
Healthcare	1
Humanistic studies	0,775
Law	0,60
Management	0,78
Marketing	0,68
Medicine	1

5.2 Homophily

As stated in the first chapter, homophily can be considered as the tendency to associate with socially similar others, and this may happen due to ethnicity, gender, studies or working backgrounds. Even if the positive aspects are often highlighted, on the other side there is the drawback related to social conformity and groupthink that may lead to inefficient decision making (Ishii & Xuan, Acquirer-target social ties and merger outcomes, 2014).

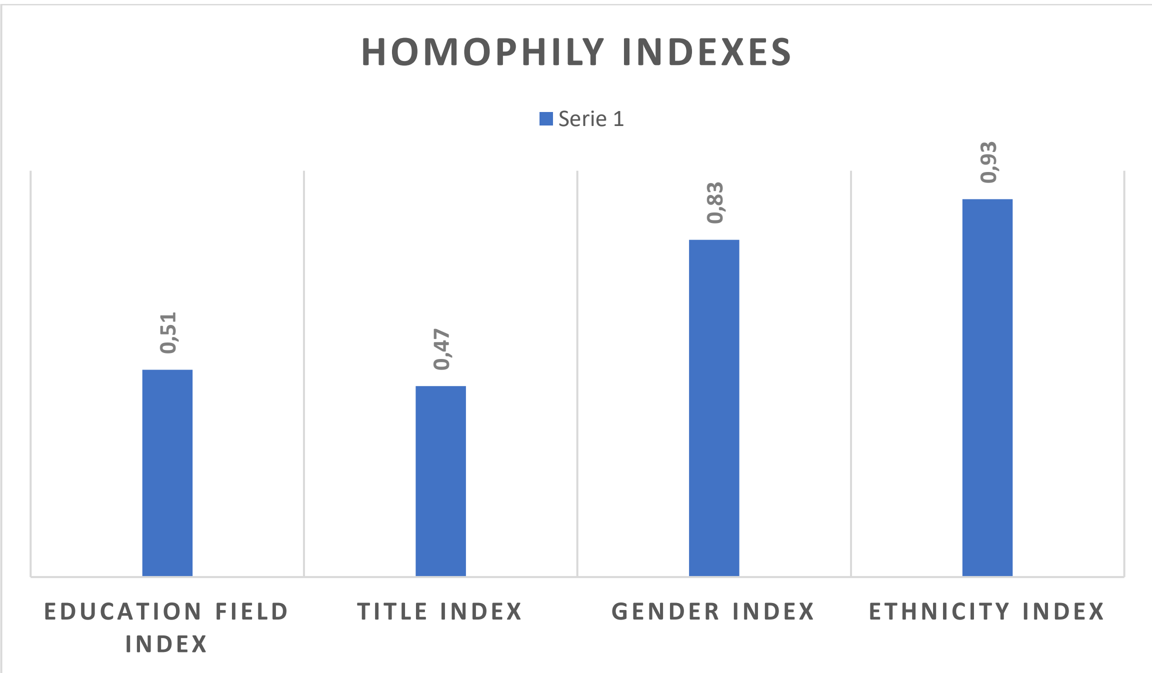
To understand if there is a high level of homophily also in this dataset, the first step was to select the investment funds that count on several investors (the threshold was chosen equal to 10), and then homophily indexes was constructed. These indexes, that will be called "Education homophily", "Title homophily", "Ethnicity homophily" and "gender homophily", were built following the same procedure of the "experience index": a ratio between the most frequent education field (or tile, ethnicity, gender) of the investors that work in an investment fund was divided by the total number of different education fields (or title, ethnicity, gender).

Of course, the higher the value of these indexes, the stronger the homophily effect.

In this dataset, the effect played by homophily seems to be stronger both in case of gender and ethnicity and weaker in case of education field and title.

The values of these indexes are shown in the following graph (figure 5.16).

Figure 5.16 *Homophily Indexes*



6 Conclusions

As stated at the beginning of this research, in the context of social impact investments the aim was to understand whether there is a relationship between the investors' characteristics and the investment choices. The purpose of this last chapter is to highlight the most significant results obtained from the analysis of Crunchbase dataset.

The results tell that the majority of social investors have a high level of education and, even if the greatest majority is represented by male partners, the female ones are the more educated. Inside the dataset there is an interesting variety of educational fields, with a strong component related to managerial and economic backgrounds, but the scientific titles seem to be relevant too. Furthermore, it seems also that there is a relationship between the type of title and the education field (for example PhD is more related to scientific studies, MBA more to managerial studies).

When the type of job is taken into consideration, the result, unsurprisingly, tells that a male partner has a higher probability to cover a relevant role, especially in case of board member. On the other side, for a female partner the percentage to be set in the board member is a little bit higher in case of traditional fund rather than in case of a social one and moreover the nationality may also play a role: an American female partner has an higher percentage than a European one.

Furthermore, it's true that over last decades the investment in social funds has increased, but while male interest for social sector doesn't seem to differ from that for traditional funds, female partners seem to be more interested in social funds.

Another interesting result was obtained after having divided into classes the social investors based on the times they invested in a social fund. In this case the statistics that there are differences not only in the education level (the more expert in the social investments are also the more educated), but that they are also interested in different social sectors.

At geographic level, it's possible to state that there is a relationship between the continent in which the partner obtained the university title and the continent in which operates the social startup that received the investments. Even if this statement is valid in general terms (like in case of North American, Asian and European female investors), there are also cases in which is not valid anymore (like in case of European male investors).

From the analysis of this dataset it was also possible, in general, to state that there is a relationship between the educational field and the investment field. In some cases, the effect is stronger (like in case of Biology, Chemistry, Medicine degree), in some others is weaker (like in case of Law, Marketing degree).

In particular, for the investors considered as social expert, it's possible to state that there is a quite strong relationship between the total fields in which they invested, and so that the investment choice is driven by the previous social experiences. Combining the investment field to the educational, it's possible to state that almost the totality of social expert partners holding a scientific title invests systematically in the same industrial field.

Following some inputs taken from the literature, a further variable was built to take into consideration if there were some shared characteristics among partners who are alumni of top colleges, in particular of Ivy League. In this case, unsurprisingly, it was quite evident that they have a much higher level of education but that this in reality doesn't translate into a greater chances of covering a role within the fund or that they prefer investing in a specific sector rather than in another one.

Finally, some considerations about homophily. To draw conclusions, a smaller dataset than the original one was selected. In fact, the limitation of the original one was to contain social investment funds in which there was only one partner or a number of partners that is too small to come up with significant statistics. The result of the selected dataset tells that the homophily effect plays a role for these data, sometimes a stronger effect (like in case of ethnicity) sometimes weaker (like in the case of education field. Therefore, the obtained results can be considered in line with the available literature.

Finally, some ideas that could further enrich the research:

- Try to personally contact the social partners who were left out of the analysis, because their educational information wasn't available, to collect the missing data. This may be interesting, as stated in the third chapter, because they seem to have long work experience that may enrich the results.
- Try to classify the VC investment funds the four different categories (Independent VC, Corporate VC, Bank-controlled VC and Governmental VC) to see if relevant statistics vary for each category,
- Try to collect personal information of the employees of the startup that had received the investment from the venture capitalists to see if homophily is present between these two actors.

7 Bibliography and sitography

Allen, Dawson, Whetly, & White. (2008). *Perceived diversity and organizational performance.*

Barman. (2015). *Of principle and principal: value plurality in the market of impact investing.*

Becker. (1964). *Human capital.*

Bengtsson, & Hsu. (2010). *How do venture capital partners match with startup founders?*

Bengtsson, & Hsu. (2015). *Ethnic matching in the U.S. venture capital market.*

Brander. (2002). *Venture-capital syndication .*

Brown, & Swersky. (2012). *The First Billion. A Forecast of Social Investment Demand.*

Calderini, Chiodo, & Michelucci. (2016). *The social impact investment race: toward an interpretative framework.*

Cetindamar, & Ozkazanc-Pan. (2017). *Assessing mission drift at venture capital impact investors.*

Cetindamar, & Ozkazanc-Pan. (2017). *Assessing mission drift at venture capital impact investors.*

Chell. (2007). *Social enterprise and entrepreneurship towards a convergent theory of the entrepreneurial process.*

Chiappini. (2017). *Social Impact Funds.*

Chiappini. (2017). *Social Impact Funds.*

Clark. (2012).

Cochrane. (2006). *The risk and return of venture capital.*

Drexler, & Nobel. (2013). *Drexler and Nobel, 2013, Impact Investing: putting ideas into practice.*

Emerson. (1996). *New Social Entrepreneurs: the success, challenge, and lesson of non-profit enterprise creation.*

Franceschini, Galetto, & Maisano. (s.d.). *Designing performance measurement systems.*

Freireich, & Fulton. (2009). *Investing for social and environmental impact.*

Fried, & Hisrich. (1994). *Toward a model of venture capital investment decision making.*

Gompers, & Lerner. (2001). *Gompers, Lerner, 2001, The venture capital revolution.*

Gompers, Mukharlyamov, & Xuan. (2016). *The cost of friendship.*

Gorman, & Sahlman. (1990). *What do venture capitalists do?*

Hedge, & Tumlinson. (2014). *Does social proximity enhance business partnership? Theory and evidence from ethnicity's role in U.S. venture capital.*

Hellmann. (2000). *Venture capitalists: the coaches of Silicon Valley.*

Hellmann, & Puri. (2000). *The interaction between product market and financing strategies.*

Hopp. (2009). *When do venture capitalist collaborate? Evidence on the driving forces of venture capital syndication.*

https://en.wikipedia.org/wiki/Descriptive_statistics. (s.d.).

https://impactdatabase.eu/. (s.d.).

https://impactspace.com/. (s.d.).

https://support.crunchbase.com/hc/en-us/articles/115010458467-Glossary-of-Funding-Types. (s.d.).

https://www.crunchbase.com/. (s.d.).

https://www.displayr.com/what-is-a-correlation-matrix/#:~:text=A%20correlation%20matrix%20is%20a,a%20diagnostic%20for%20advanced%20analyses. (s.d.).

https://www.impactassets.org/. (s.d.).

Ishii, & Xuan. (2014). *Acquirer-target social ties and merger outcomes.*

Ishii, & Xuan. (2014). *Acquirer-target social ties and merger outcomes.*

Kristandl, & Bontis. (2007). *Management decision.*

Kroeger, & Weber. (2014). *Developing a conceptual framework for comparing social value creation.*

Lerner. (1994). *The syndication of venture Capital investments.*

Lerner. (1995). *Venture Capitalists and the oversight of Private Firms.*

Lewellen. (1971). *A pure financial rationale for conglomerate merger.*

Lockett, & Wright. (1999). *Lockett, Wright, 1999, The syndication of private equity: evidence from the UK.*

Macmillan. (1989). *Venture capitalists' involvement in their investments.*

Mulgan. (2010). *Measuring social value.*

Nicholls. (2009). *We do good things, don't we? Blended value accounting. Social entrepreneurship .*

Nicholls. (2018). *A general theory of Social Impact Accounting.*

Oxford, U. o. (s.d.). *https://golab.bsg.ox.ac.uk/the-basics/social-impact-investing/#chapter_1_what-is-social-impact-investment__title*.

Ozkazanc-Pan, & Cetindamar. (2017). *Assessing mission drift at venture capital impact investors.*

Pache, & Santos. (2013). *Inside the hybrid organization: selective coupling and as a response to competing institutional logics.*

Polanyi. (1967). *The growth of science in society.*

Revelli, & Viviani. (2015). *Financial performance of socially responsible investing.*

Russell. (2007).

Scarlata, Walske, & Zacharakis. (2017). *How the human capital of philanthropic and traditional venture capital differs.*

Stuart, & Sorensen. (2001). *Syndication networks and the spatial distribution of Venture Capital Investments.*

Ughetto. (2019). *Disintermediate Finance: Business Angels.*

Williams, & O'Reilly. (1998). *Demography and diversity in organizations.*

8 Figure and table index

9 Appendix, Stata Code

Code for General Statistics

```
drop if missing(org_uuid)
drop if missing(is_current)
drop if missing(title)
drop if missing(job_type)
drop if missing(gender)
drop if missing(country_code)
drop if missing(country)
drop if missing(continent)
drop if missing(featured_job_organization_uuid)
drop if missing(featured_job_title)

tab _0_org

drop if _0_org == ”org”

tab _0_org

codebook org_uuid
codebook org_uuid if _0_org == ”1”

codebook partner_uuid

codebook partner_uuid  if gender == ”male”
codebook partner_uuid  if gender == ”male” & _0_org == ”1”
codebook partner_uuid  if gender == ”male” & _0_org == ”0”

codebook partner_uuid  if gender == ”female”
codebook partner_uuid  if gender == ”female” & _0_org == ”1”
codebook partner_uuid  if gender == ”female” & _0_org == ”0”

duplicates report partner_uuid if gender == ”male” & _0_org == ”1”
duplicates report partner_uuid if gender == ”female” & _0_org == ”1”
```

3.6.2 Code for Educational Statistics

```
drop if missing(org_uuid)
drop if missing(is_current)
drop if missing(title)
drop if missing(job_type)
drop if missing(gender)
drop if missing(country_code)
drop if missing(country)
drop if missing(continent)
drop if missing(featured_job_organization_uuid)
drop if missing(featured_job_title)
drop if missing(university_name_tot)
drop if missing(country_code_education_tot)
drop if missing(continent_education_tot)
drop if missing(degree_tot)
drop if missing(field_education_tot)
drop if missing(macro_field_education_tot)

drop if country_code_education_tot =="#N/D"
drop if continent_education_tot =="#N/D"
drop if degree_tot =="#N/D"
drop if field_education_tot =="#N/D"
drop if macro_field_education_tot =="#N/D"

drop if _0_org == "org"
drop if _0_org == "0"
tab _0_org

codebook partner_uuid
codebook partner_uuid  if gender == "male"
codebook partner_uuid  if gender == "female"
duplicates drop partner_uuid , force
codebook partner_uuid  if gender == "male"
codebook partner_uuid  if gender == "female"

replace degree_tot="Master" if degree_tot=="ms"
replace degree_tot="Bachelor" if degree_tot=="Other"
replace industry="Energy" if industry=="Energy "
replace macro_field_education_tot="Finance" if
macro_field_education_tot=="Finance & Entrepreneurship"
```

```
replace macro_field_education_tot="Humanistic sciences" if
macro_field_education_tot=="Humanistic studies"

replace macro_field_education_tot="Political Science & International Relations"
if macro_field_education_male=="International Relations and Political Science"

replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical Engineering "

replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical Engineering Technologies/Tec"

replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical and Electronic Engineering"

replace field_education_tot="Management Engineering" if
field_education_tot=="Management Science & Engineering"

replace field_education_tot="Management Engineering" if
field_education_tot=="Management Science and Engineering"

replace field_education_tot="Mechanical Engineering" if
field_education_tot=="Mechanical Engineering / Economics"

replace field_education_tot="Systems and Information Engineering" if
field_education_tot=="systems engineering"

replace field_education_tot="Chemical Engineering" if
field_education_tot=="Chemical Engineering "

replace field_education_tot="Aeronautical Engineering" if
field_education_tot=="Aeronautics & Astronautics Engineering"

replace field_education_tot="Aeronautical Engineering" if
field_education_tot=="Aeronautics Engineering"

replace field_education_tot="Software Engineering" if
field_education_tot=="Applied Science (Software Engineering)"

replace field_education_tot="Material Engineering" if
field_education_tot=="Materials Science & Engineering"

replace field_education_tot="Management Engineering" if
field_education_tot=="Engineering Management"

replace field_education_tot="Systems and Information Engineering" if
field_education_tot=="Engineering Systems Division"



codebook university_name_tot

codebook country_code_education_tot


tab country_code_education_tot

tab country_code_education_tot  if gender =="male"

tab country_code_education_tot  if gender =="female"


tab continent_education_tot

tab continent_education_tot  if gender =="male"

tab continent_education_tot  if gender =="female"


tab degree_tot

tab degree_tot if gender =="male"
```

```
tab degree_tot if gender =="female"

codebook macro_field_education_tot
tab macro_field_education_tot
tab macro_field_education_tot if gender =="male"
tab macro_field_education_tot if gender =="female"

codebook field_education_tot if macro_field_education_tot =="Engineering"
tab field_education_tot if macro_field_education_tot =="Engineering"

tab macro_field_education_tot if degree_tot=="Bachelor"
tab macro_field_education_tot if degree_tot=="Master"
tab macro_field_education_tot if degree_tot=="MBA"
tab macro_field_education_tot if degree_tot=="PhD"

gen Ivy_League = university_name_tot == "Brown University" | university_name_tot
== "Columbia University" | university_name_tot == "Cornell University" |
university_name_tot == "Dartmouth College" | university_name_tot == "Harvard
University" | university_name_tot == "University of Pennsylvania"
|university_name_tot == "Princeton University" |university_name_tot == "Yale
University"

tab degree_tot if university_name_tot =="Brown University"
tab degree_tot if university_name_tot =="Columbia University"
tab degree_tot if university_name_tot =="Cornell University"
tab degree_tot if university_name_tot =="Dartmouth College"
tab degree_tot if university_name_tot =="Harvard University"
tab degree_tot if university_name_tot =="University of Pennsylvania"
tab degree_tot if university_name_tot =="Princeton University"
tab degree_tot if university_name_tot =="Yale University"

tab degree_tot if Ivy_League==1
```

3.6.3 Code for Professional Statistics

```
drop if missing(org_uuid)
drop if missing(is_current)
drop if missing(title)
drop if missing(job_type)
drop if missing(gender)
drop if missing(country_code)
drop if missing(country)
```

```
drop if missing(continent)
drop if missing(featured_job_organization_uuid)
drop if missing(featured_job_title)

tab _0_org
drop if _0_org == "org"
tab _0_org

codebook country_code
tab country_code
tab continent

codebook job_type
tab job_type

tab job_type if _0_org=="1"
tab job_type if _0_org=="0"

tab job_type if gender== "male"
tab job_type _0_org if gender== "male"

tab job_type if gender== "female"
tab job_type _0_org if gender== "female"

tab job_type
tab job_type continent
tab job_type continent, row column
tab job_type if continent=="North America"
tab job_type if continent=="Europe"
tab job_type gender if continent=="North America"
tab job_type gender if continent=="Europe"

tab job_type continent if gender=="male", row column
tab job_type continent if gender=="female", row column

tab job_type if _0_org=="0" & gender=="male"
tab job_type if _0_org=="1" & gender=="male"
tab job_type if _0_org=="0" & gender=="female"
tab job_type if _0_org=="1" & gender=="female"
```

```
tab job_type continent if _0_org=="0" & gender=="male"
tab job_type continent if _0_org=="1" & gender=="male"
tab job_type continent if _0_org=="0" & gender=="female"
tab job_type continent if _0_org=="1" & gender=="female"

codebook featured_job_title
codebook featured_job_title if job_type == "advisor"
tab featured_job_title if job_type == "advisor"
```

3.6.4 Code for Investment statistics

```
drop if social==0
drop if missing(fundfield)

replace macro_field_education="Scientific studies" if macro_field_education=="Biology"
replace macro_field_education="Scientific studies" if macro_field_education=="Chemistry"
replace macro_field_education="Scientific studies" if macro_field_education=="Energy"
replace macro_field_education="Scientific studies" if macro_field_education=="Healthcare"
replace macro_field_education="Scientific studies" if macro_field_education=="Mathematics"

replace macro_field_education="Scientific studies" if macro_field_education=="Medicine"
replace macro_field_education="Scientific studies" if macro_field_education=="Pharmacy"
replace macro_field_education="Scientific studies" if macro_field_education=="Physics"
replace macro_field_education="Scientific studies" if macro_field_education=="Science"
replace macro_field_education="Scientific studies" if macro_field_education=="Technology"
replace macro_field_education="Scientific studies" if macro_field_education=="Design"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Entrepreneurship "
replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Entrepreneurship"
replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Leadership"
replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Strategy"
```

```
replace macro_field_education="Humanistic sciences" if
macro_field_education=="Social Science"

replace macro_field_education="Humanistic sciences" if
macro_field_education=="Education"


gen Ivy_League = university_name == "Brown University" | university_name ==
"Columbia University" | university_name == "Cornell University" |
university_name == "Dartmouth College" | university_name == "Harvard University"
| university_name == "University of Pennsylvania" |university_name == "Princeton
University" |university_name == "Yale University"


by partner_uuid, sort : egen count_social_inv=sum(social)

tab count_social_inv

rename count_social_inv count_social_invper_partner

preserve

duplicates drop partner_uuid, force

tab count_social_invper_partner



gen basic_social = count_social_invper_partner==1

tab basic_social


gen intermediate_social = count_social_invper_partner>1 &
count_social_invper_partner<11

tab intermediate_social


gen expert_social = count_social_invper_partner>10

tab expert_social
```

3.5 Code for Mixed Statistics

```
drop if missing(org_uuid)

drop if missing(is_current)

drop if missing(title)

drop if missing(job_type)

drop if missing(gender)

drop if missing(country_code)

drop if missing(country)

drop if missing(continent)

drop if missing(featured_job_organization_uuid)

drop if missing(featured_job_title)

drop if missing(university_name_tot)

drop if missing(country_code_education_tot)

drop if missing(continent_education_tot)
```

```
drop if missing(degree_tot)
drop if missing(field_education_tot)
drop if missing(macro_field_education_tot)

drop if country_code_education_tot =="#N/D"
drop if continent_education_tot =="#N/D"
drop if degree_tot =="#N/D"
drop if field_education_tot =="#N/D"
drop if macro_field_education_tot =="#N/D"

drop if _0_org == "org"
drop if _0_org == "0"
tab _0_org

replace degree_tot="Master" if degree_tot=="ms"
replace degree_tot="Bachelor" if degree_tot=="Other"
replace industry="Energy" if industry=="Energy "
replace macro_field_education_tot="Finance" if
macro_field_education_tot=="Finance & Entrepreneurship"
replace macro_field_education_tot="Humanistic sciences" if
macro_field_education_tot=="Humanistic studies"
replace macro_field_education_tot="Political Science & International Relations"
if macro_field_education_male=="International Relations and Political Science"
replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical Engineering "
replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical Engineering Technologies/Tec"
replace field_education_tot="Electrical Engineering" if
field_education_tot=="Electrical and Electronic Engineering"
replace field_education_tot="Management Engineering" if
field_education_tot=="Management Science & Engineering"
replace field_education_tot="Management Engineering" if
field_education_tot=="Management Science and Engineering"
replace field_education_tot="Mechanical Engineering" if
field_education_tot=="Mechanical Engineering / Economics"
replace field_education_tot="Systems and Information Engineering" if
field_education_tot=="systems engineering"
replace field_education_tot="Chemical Engineering" if
field_education_tot=="Chemical Engineering "
replace field_education_tot="Aeronautical Engineering" if
field_education_tot=="Aeronautics & Astronautics Engineering"
replace field_education_tot="Aeronautical Engineering" if
field_education_tot=="Aeronautics Engineering"
replace field_education_tot="Software Engineering" if
field_education_tot=="Applied Science (Software Engineering)"
replace field_education_tot="Material Engineering" if
field_education_tot=="Materials Science & Engineering"
```

```
replace field_education_tot="Management Engineering" if
field_education_tot=="Engineering Management"

replace field_education_tot="Systems and Information Engineering" if
field_education_tot=="Engineering Systems Division"


replace industry="Energy" if industry=="Energy "

replace industry="Impact-tech" if industry=="impact-tech"


tab degree_tot industry

tab macro_field_education_tot industry


replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Biology"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Chemistry"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Energy"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Healthcare"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Mathematics"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Medicine"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Pharmacy"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Physics"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Science"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Technology"

replace macro_field_education_tot="Scientific studies" if
macro_field_education_tot=="Design"


replace macro_field_education_tot="Entrepreneurship, Strategy & Leadership" if
macro_field_education_tot=="Entrepreneurship "

replace macro_field_education_tot="Entrepreneurship, Strategy & Leadership" if
macro_field_education_tot=="Entrepreneurship"

replace macro_field_education_tot="Entrepreneurship, Strategy & Leadership" if
macro_field_education_tot=="Leadership"

replace macro_field_education_tot="Entrepreneurship, Strategy & Leadership" if
macro_field_education_tot=="Strategy"


replace macro_field_education_tot="Humanistic sciences" if
macro_field_education_tot=="Social Science"

replace macro_field_education_tot="Humanistic sciences" if
macro_field_education_tot=="Education"
```

```
tab macro_field_education_tot
tab macro_field_education_tot industry

tab investment_round degree_tot

tab job_type degree_tot
tab job_type degree_tot if gender=="male"
tab job_type degree_tot if gender=="female"

gen Ivy_League = university_name_tot == "Brown University" | university_name_tot
== "Columbia University" | university_name_tot == "Cornell University" |
university_name_tot == "Dartmouth College" | university_name_tot == "Harvard
University" | university_name_tot == "University of Pennsylvania"
|university_name_tot == "Princeton University" |university_name_tot == "Yale
University"

tab job_type Ivy_League
tab job_type if Ivy_League==1
tab job_type if Ivy_League==1 & gender=="male"
tab job_type if Ivy_League==1 & gender=="female"

tab job_type Ivy_League if degree_tot == "MBA"
tab industry if Ivy_League==1
tab industry if Ivy_League==0
```

Expert investors

```
drop if missing(org_uuid)
drop if missing(is_current)
drop if missing(title)
drop if missing(job_type)
drop if missing(gender)
drop if missing(country_code)
drop if missing(country)
drop if missing(continent)
drop if missing(featured_job_organization_uuid)
drop if missing(featured_job_title)

drop if degree_tot =="#N/D"

replace degree_tot="Master" if degree_tot=="ms"
replace degree_tot="Bachelor" if degree_tot=="Other"
replace industry="Energy" if industry=="Energy "
```

```
gen basic_social = social_level==1

gen intermediate_social = social_level==2 | social_level==3 | social_level==4

gen expert_social = social_level>4

gen Ivy_League = university_name_tot == "Brown University" | university_name_tot
== "Columbia University" | university_name_tot == "Cornell University" |
university_name_tot == "Dartmouth College" | university_name_tot == "Harvard
University" | university_name_tot == "University of Pennsylvania"
|university_name_tot == "Princeton University" |university_name_tot == "Yale
University"



tab _0_org

drop if _0_org == "org"

drop if _0_org == "0"

tab _0_org


tab gender if basic_social==1

tab gender if intermediate_social==1

tab gender if expert_social==1


tab job_type if basic_social==1

tab job_type if intermediate_social==1

tab job_type if expert_social==1


tab degree_tot if basic_social==1

tab degree_tot if intermediate_social==1

tab degree_tot if expert_social==1


replace industry="Energy" if industry=="Energy "

replace industry="Impact-tech" if industry=="impact-tech"


tab industry if basic_social==1

tab industry if intermediate_social==1

tab industry if expert_social==1


tab investment_round

tab investment_round if basic_social==1

tab investment_round if intermediate_social==1

tab investment_round if expert_social==1
```

```
drop if social==0
drop if missing(fundfield)
drop if degree=="0"
drop if degree=="ms"
drop if degree=="Other"

by partner_uuid, sort : egen count_social_inv=sum(social)
tab count_social_inv
rename count_social_inv count_social_invper_partner
preserve
duplicates drop partner_uuid, force
tab count_social_invper_partner

gen basic_social = count_social_invper_partner==1
tab basic_social

gen intermediate_social = count_social_invper_partner>1 & count_social_invper_partner<11
tab intermediate_social

gen expert_social = count_social_invper_partner>10
tab expert_social

duplicates drop partner_uuid , force
tab degree if expert_social ==1
tab degree if intermediate_social ==1
tab degree if basic_social ==1
```

Correlation matrix

```
drop if social==0
drop if missing(fundfield)

replace macro_field_education="Scientific studies" if macro_field_education=="Biology"
replace macro_field_education="Scientific studies" if macro_field_education=="Chemistry"
replace macro_field_education="Scientific studies" if macro_field_education=="Energy"
replace macro_field_education="Scientific studies" if macro_field_education=="Healthcare"
```

```
replace macro_field_education="Scientific studies" if
macro_field_education=="Mathematics"

replace macro_field_education="Scientific studies" if
macro_field_education=="Medicine"

replace macro_field_education="Scientific studies" if
macro_field_education=="Pharmacy"

replace macro_field_education="Scientific studies" if
macro_field_education=="Physics"

replace macro_field_education="Scientific studies" if
macro_field_education=="Science"

replace macro_field_education="Scientific studies" if
macro_field_education=="Technology"

replace macro_field_education="Scientific studies" if
macro_field_education=="Design"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if
macro_field_education=="Entrepreneurship "

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if
macro_field_education=="Entrepreneurship"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if
macro_field_education=="Leadership"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if
macro_field_education=="Strategy"

replace macro_field_education="Humanistic sciences" if
macro_field_education=="Social Science"

replace macro_field_education="Humanistic sciences" if
macro_field_education=="Education"


gen Ivy_League = university_name == "Brown University" | university_name ==
"Columbia University" | university_name == "Cornell University" |
university_name == "Dartmouth College" | university_name == "Harvard University"
| university_name == "University of Pennsylvania" |university_name == "Princeton
University" |university_name == "Yale University"


by partner_uuid, sort : egen count_social_inv=sum(social)

tab count_social_inv

rename count_social_inv count_social_invper_partner

preserve

duplicates drop partner_uuid, force

tab count_social_invper_partner



gen basic_social = count_social_invper_partner==1

tab basic_social


gen intermediate_social = count_social_invper_partner>1 &
count_social_invper_partner<11

tab intermediate_social
```

```
gen expert_social = count_social_invper_partner>10
tab expert_social

encode fundinground, gen(FuRound)
encode gender, gen(sex)
encode jobtype, gen(Job_type)
encode fundfield, gen(FuField)
encode degree, gen(EduTitle)
encode macro_field_education, gen(EduField)
encode education_country_code, gen(EduConCode)
encode investor_country_code, gen(InConCode)

pwcorr FuRound sex Job_type FuField EduField EduTitle Ivy_League EduConCode
InConCode basic_social intermediate_social expert_social
```

2nd Statement code

```
tab social

drop if social==0

tab fundfield

drop if missing(fundfield)

tab macro_field_education

drop if macro_field_education==”0”

replace macro_field_education=”Scientific studies” if macro_field_education==”Biology”

replace macro_field_education=”Scientific studies” if macro_field_education==”Chemistry”

replace macro_field_education=”Scientific studies” if macro_field_education==”Energy”

replace macro_field_education=”Scientific studies” if macro_field_education==”Healthcare”

replace macro_field_education=”Scientific studies” if macro_field_education==”Mathematics”

replace macro_field_education=”Scientific studies” if macro_field_education==”Medicine”

replace macro_field_education=”Scientific studies” if macro_field_education==”Pharmacy”

replace macro_field_education=”Scientific studies” if macro_field_education==”Physics”
```

```
replace macro_field_education="Scientific studies" if macro_field_education=="Science"

replace macro_field_education="Scientific studies" if macro_field_education=="Technology"

replace macro_field_education="Scientific studies" if macro_field_education=="Design"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Entrepreneurship "

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Entrepreneurship"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Leadership"

replace macro_field_education="Entrepreneurship, Strategy & Leadership" if macro_field_education=="Strategy"

replace macro_field_education="Finance" if macro_field_education=="Entrepreneurship, Strategy & Leadership"

replace macro_field_education="Humanistic sciences" if macro_field_education=="Social Science"

replace macro_field_education="Humanistic sciences" if macro_field_education=="Education"

tab macro_field_education

tab fundfield if macro_field_education=="Business Administration"

tab fundfield if macro_field_education=="Economics & Commerce"

tab fundfield if macro_field_education=="Engineering"

tab fundfield if macro_field_education=="Finance"

tab fundfield if macro_field_education=="Humanistic sciences"

tab fundfield if macro_field_education=="Languages"

tab fundfield if macro_field_education=="Law"

tab fundfield if macro_field_education=="Management"

tab fundfield if macro_field_education=="Marketing"

tab fundfield if macro_field_education=="Political Science & International Relations"

tab fundfield if macro_field_education=="Scientific studies"
```