How do the founders' educational backgrounds impact startup funding?

Minh Quang Khuat Duy

University of Toronto

ECO225

Abstract

This paper assesses the influence of founders' educational background on startup funding based on the human capital theory. The Crunchbase dataset containing information about 5311 startups in America and their founders was utilized to examine the impact of founders' educational background on startup funding. The main findings are: 1) Founders that have a higher education level are associated with higher startup funding. 2) Founders who studied a STEM field for their highest academic qualification are associated with higher startup funding. 3) Founders who graduated from higher quality schools are associated with higher startup funding. 4) State income and R&D expenditure, founders' level of education and founders' schools' research environment are the most important factors in determining startup funding. These findings enhance the understanding of education's role in entrepreneurial fund-raising endeavors. Aspiring entrepreneurs can consider these factors in the tertiary education decision making process.

Keywords: Startups, Founders, Education, Startup Funding, Fund-raising, Investment.

INTRODUCTION

This paper provides in-depth analysis into the question "How does the educational background of founders impact the amount of funding that American startups receive?". Founders have always been the driving force of startups and often a crucial determinant of the level of success that corporations can achieve. The human capital theory states that education drives human capital, which, in theory, proceeds to drive startup success, measured by startup funding. This project aims to explore this assumption by examining key educational factors of founders against the total amount of fundings in USD.

Current research has pointed to certain factors of founders' education, such as education qualifications and subjects of study, that may influence funding. A paper that analyzed 4953 digital startups stated that founders with a technical education are more likely to raise equity investment although at higher levels of education the effect of a technical education reduces (Ratzinger et al., 2018). Another paper also provides some support to this notion, underlining that startup founders that have higher academic qualifications tend to receive improved funding for their companies, under the conditions that they do not study in different fields (Franco et al., 2021). Not only limited to the US, but there is also some evidence to indicate that these findings can be replicated in other countries as well. In Indonesia, a paper analyzed 207 digital startup entrepreneurs to find that entrepreneurs who pursued further education notably improved their likelihood of attaining increased levels of investment (Furqon et al., 2023). Human Capital Theory can offer potential explanations for these trends. The theory states that education is a major contributory factor to human capital, which dictates the productivity and economic influence of an individual (Ross, 2023). Despite certain limitations, this theory still holds weight in the modern business world and has been supported by research papers. There has also been specific evidence for how higher education and STEM-education specifically have a significant impact on increasing human capital. According to Adedeji and Campbell (2013), higher education is a must-have factor for the development of human

capital and economic growth. In addition, Winters (2014) stated in his paper that STEM-education has a significant positive impact on an individual's level of productivity. Even further, STEM-graduates have been found to create positive externalities, and indirectly increase the wage of people in their environment. This can be the result of the substantial impact on productivity that STEM-educated people have on their team settings, leading to higher wages for members, or other externalities. At any rate, research has shown that having a STEM education and higher levels of education propel productivity and economic growth (human capital) significantly. Based on this information, it is safe to assume that investors would prefer individuals with these educational characteristics as the leaders of companies that they invest in.

This paper reaffirms the relationships between the impact of the founders' level of education and STEM education on startup funding using a dataset containing information about startups in the US. Beyond this, these relationships were investigated on a geographical level. Different state-level maps provide support to the main findings; there were many inconsistencies due to the different distributions of startups in each state. In addition, this research paper explores the relationship between specific characteristics of founders' schools and schools 'overall quality with startup fundings. Crucially, this paper applied Machine Learning methods to determine the importance of educational factors, external state factors and schools-related factors of founders when it comes to startup funding.

To answer the research question, the following sections will explore the dataset, provide interpretations and visualizations for each finding and perform regression analysis using linear, decision tree and random forest methods to assess the importance of different factors.

DATA EXPLORATION

The dataset that will be explored in the project is the Startup Investments dataset retrieved from Kaggle, attributed to Crunchbase's 2013 Snapshot dataset. Three main tables from this dataset will be reviewed extensively: objects.csv, degrees.csv and relationships.csv. The objects.csv table contains the overall information of the companies and employees; the main columns of interest are the company id and the total amount of funding columns. The degrees.csv table contains information about the educational backgrounds of employees and the relationships.csv has information about each employee and their connections to and positions in companies. These three tables will be merged to create the main data frame, which will then be filtered for only educational information concerning the founders of startups, resulting in a data frame of roughly 10,600 observations. This data frame will be the foundation of the project on which variable analysis will be conducted. By the end of the paper, with further adjustments and suitable merging of other observations, the data frame used for regression analysis will contain roughly 5311 startups and their founders' educational backgrounds.

SUMMARY STATISTICS

There are fundamental aspects of educational background that will be examined in this paper, including the number of degrees, institutions attended, the level of education (depth) achieved and the types of fields that startups founders studied. These variables are picked based on the assumption that they can be factors of interest for investors as they conduct background checks. These three aspects will be divided into 5 independent variables to assess against the dependent variable of total funds raised. The 5 independent variables are:

1) The number of Art (BA + MA degrees) and Science (BS + MS degrees) degrees of the founders
2) Level of education (Bachelor, Master and PhD) of the founders

3) The average number of degrees that founders of startups hold (The average is taken since start-ups can have multiple co-founders)
4) The STEM designation for founders' highest level of education (whether the founders' highest level of education is in a STEM field or not)
5) The Ivy League designation for founders' (whether the founders have studied at an Ivy League school or not)

The total funding raised is a crucial indicator of a company's success. Having significant funding allows startups to grow at an accelerated rate, invest in research and development projects and enhance marketing and branding opportunity. At the beginning of their lifespan where startups are struggling to break even in their business, funding is sometimes the main factor keeping the company afloat in the market. Raising funds is one of the first major issues that startups owners have to tackle in order to maintain longevity. The company must have an overall decent business prospect in order to attract funding from investors. In this light, the total amount of funding raised can be viewed as a significant signal of success. The importance of funding to startups cannot be overstated, which is why this research paper aims to provide deeper insights by examining funding as the dependent variable. Data manipulation and aggregation have been performed on the original dataset to produce the main summary statistics table containing descriptive statistics about variables of interest.

Figure 1*: Main Summary Statistics Table*

	attended_ivy_league	funding_in_millions	studied_STEM_field	founders_avg_no_of_degrees	highest_education_index	founders_no_of_art_degrees	founders_no_of_sciences_degrees
count	5311.000000	5311.000000	5311.000000	5311.000000	5311.000000	5311.000000	5311.000000
mean	0.181698	24.230305	0.309358	2.211134	1.595933	0.567501	1.224816
std	0.385632	83.957226	0.462272	1.926570	0.638217	0.759813	1.008399
min	0.000000	0.001003	0.000000	1.000000	1.000000	0.000000	0.000000
25%	0.000000	0.961000	0.000000	1.000000	1.000000	0.000000	1.000000
50%	0.000000	4.600000	0.000000	2.000000	2.000000	0.000000	1.000000
75%	0.000000	19.714177	1.000000	2.500000	2.000000	1.000000	2.000000
max	1.000000	2425.700000	1.000000	28.000000	3.000000	4.000000	4.000000

The main summary statistic table reveals interesting information about the independent and dependent variables. For the first independent variables (the count of art and science degrees), it seems that founders of startups on average have more science degrees rather than art degrees. Regarding the average number of degrees held by founders, the summary statistics table reveals that on average founders of startups have more than two academic qualifications. However, it is worth noting that the maximum number of degrees of 28 is an extreme outlier that is impossible, presumably errors caused by the original dataset's inconsistencies, indicating that the average number of degrees held by founders has been skewed by outliers and should be less than 2.2.

The education level index (an index that assigns the values 1, 2, 3 to Bachelor, Masters' and PhD degrees, respectively) has an average of 1.64, indicating that a number of startup founders have a higher degree than the Bachelor level. Specifically, founders who have Bachelor and Master as their highest education level each take up roughly 45% (~2700 founders) of the dataset; only around 10% (535 founders) of the founders have achieved a PhD-level education. Additionally, around 30% of founders studied a STEM field for their highest level of education and 16.5% of founders studied at an Ivy League school.

Lastly, the independent variable of funding shows that the average funding of a startup is $25 million, which is once again somewhat skewed by the outliers as can be seen from the maximum amount of funding that a startup in the dataset received: $2.5 billion. It is important to note that all of these variables are only estimations due to the nature of the dataset, having founders who started multiple companies and companies started by multiple co-founders.

Initial Analysis of Educational Factors

Different data aggregation and correlation analysis methods have been performed to gain an initial assessment of the relationships between the five independent variables of interest with startup funding.

1. Art and Science Degrees

There appears to be no apparent correlation between the number of arts or sciences the founders have degrees with the amount of fundings that companies received. In other words, the differentiation between art degrees and science degrees has no impact on how much fundings companies receive. Similarly, the average number of degrees between founders of companies also hold no relationship with startup funding.

2. Highest Level of Education

Analyzing the highest level of education of founders, the result indicates that, for different levels of Bachelor, Masters and PhD, the higher education level are associated with higher average funding (the average funding that founders with this qualification as their highest level of education receive). In detail, although the discrepancy between the Masters and the Bachelor level for average funding is marginal, the average funding for the PhD-level founders is almost double that of Masters-level founders. Having a higher-level education can be translated to having a higher level of expertise in the field, which may be an important factor for investors to consider in investment decisions, therefore potentially impacting funding.

3. STEM Founders

Whether founders had their highest academic qualification in a STEM field is also one of the variables of interest. A dummy variable was created with 1 indicating startups with founders who studied a STEM field for their highest qualification (STEM-founders) and 0 indicating startups with founders who did not (non-STEM-founders). For startups with non-STEM founders, the median amount of funding is around $4 million while this figure is around $6 million for STEM founders. This means that, on average, startups with STEM-founders typically receive slightly higher funding than companies with non-STEM founders in this dataset. In addition, 75% of the startups that have STEM founders have received total funding below $26 million while 75% of the startups by non-STEM founders only received total funding below the $17 million mark. This demonstrates that the majority of companies founded by STEM-

educated entrepreneurs have a wider range of funding and that some companies can reach higher amounts of funding than their non-STEM counterparts.

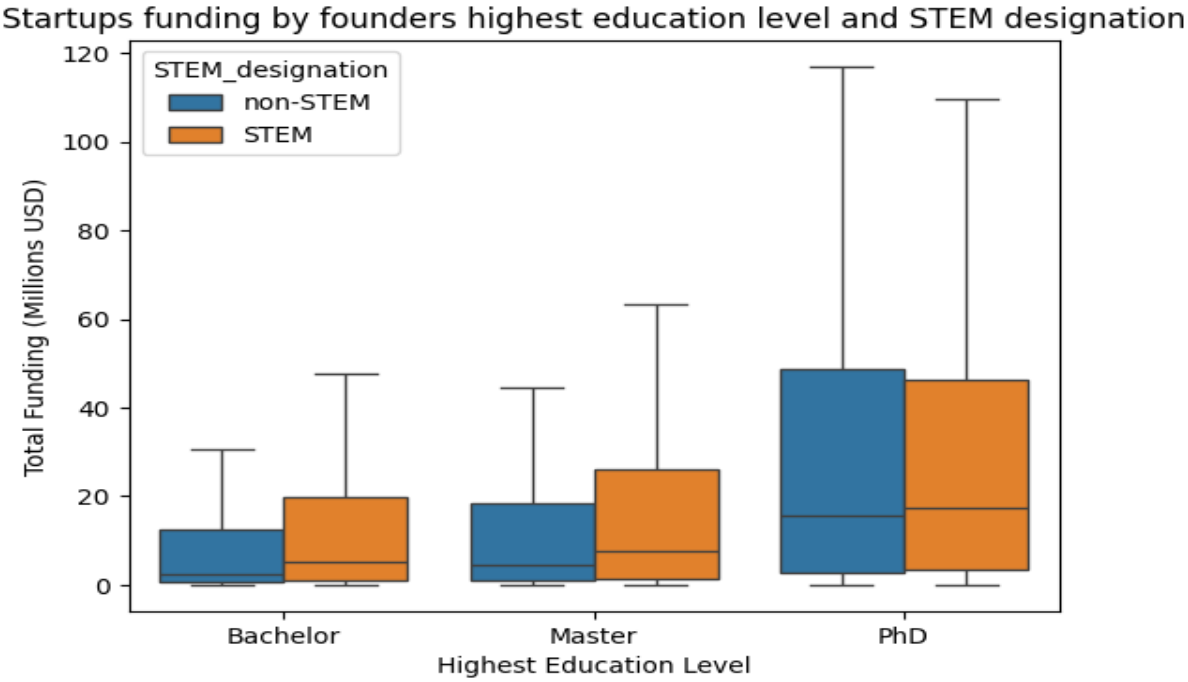
4. Ivy League Founders

Lastly, the quality of institutions which founders attended is also taken into consideration. Particularly, Ivy League schools, which are known as the best institutions in the world, can be an interesting measure of this factor. A dummy variable was created with 1 indicating startups with founders who attended an Ivy League school (Ivy-founders) and 0 indicating startups with founders who did not (non-Ivy-founders). From the analysis, it seems that on average startups with Ivy-founders perform better than those with non-Ivy-founders. 75% of the startups that have Ivy-founders have received total funding below $29 million while 75% of the startups by non-STEM founders only received total funding below the $19 million mark. This demonstrates that the majority of companies founded by Ivy League-educated entrepreneurs have a wider range of funding and that some companies founded by Ivy founders can reach higher amounts of funding, as opposed to those founded by non-Ivy founders. Some possible reasons for this discrepancy may include the quality of education, connections and resources that are better for startups development at Ivy League schools. Future projects should explore further to discover exactly what values that Ivy League schools offer the founders which led to these startups to receive higher funding.

Main Finding from Initial Assessment

Higher educational qualifications and STEM education for founders is associated with higher total funding for startups. There is a drastic difference between fundings for startups that have STEM PhD-educated founders as opposed to STEM Masters-educated founders, while the funding discrepancy for STEM Master-educated founders and STEM Bachelor-educated founders is less significant.

Figure 2: *Startups Funding by Founders Highest Education Level and STEM Designation*



For the majority of the startups, companies with Master-educated founders average around $5 million in fundings, slightly higher than those founded by Bachelor's degrees holders. In comparison, the most significant proportion of startups with PhD-level founders received around $18 million in funding. This implies that, for most of the startups in this dataset, higher educational level of founders is associated with higher funding for the startups, suggesting that the educational level of founders has the potential to be one of the deciding factors for investors when it comes to providing funding for the company.

In addition, the spread of total funding for startups with PhD-educated founders is also more wide than their 2 counterparts. 75% of the observations for startups with Bachelor-educated founders, Master-educated founders and PhD-educated founders can reach up to approximately $19 million, $21 million, and $47 million respectively. This finding indicates that, for a significant number of companies, the amount of funding that startups with PhD-educated founders can receive is more than double of what the startups with less academically qualified founders can receive.

It is noticeable that the third quartile of the fundings for startups with founders who studied STEM subjects is markedly higher than non-STEM subjects for the Bachelor and Master level. In contrast, for the PhD level the opposite is true, as the third quartile for non-STEM subjects is marginally higher than STEM subjects, this difference for the PhD level is a topic that should be investigated further. In this light, startups with STEM-educated founders have the potential to reach higher level of funding than those with non-STEM-educated founders for the Bachelor and Master level, implying that, to a certain degree, investors may possibly prefer STEM-educated founders when it comes to funding decisions.

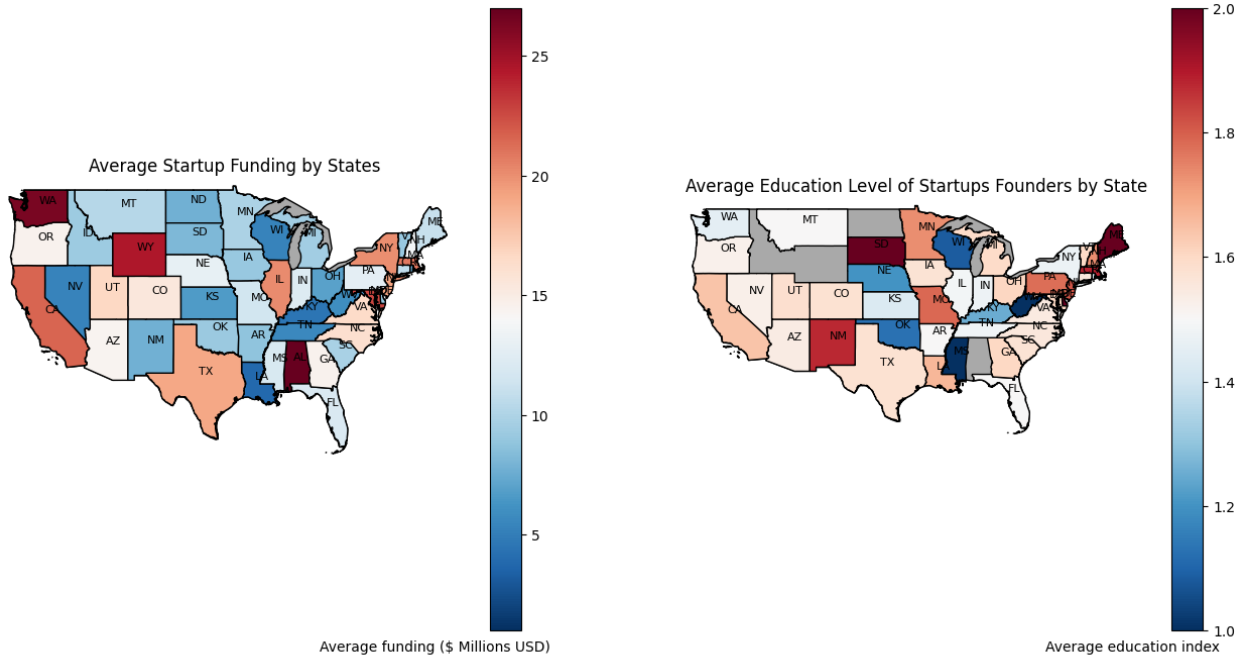
It is important to note that this visualization does not account for outliers, as it is assumed that the emphasis should be placed on the bulk of the data instead of the exceptions. Accounting for the outliers, the companies that achieved the highest fundings, reaching around $2.5 billion, had Bachelor-educated founders. Nevertheless, considering the small size of these outliers compared to the total volume of the dataset, it is assumed that these abnormal observations are not representative of the nature of the dataset.

These trends can potentially be explained by the Human Capital Theory, which states that education contributes significantly to an individual's domain knowledge, productivity, and innovative capability. In this light, investors may assess that founders who have a higher level of education have higher human capital and thus more capable of effectively leading successful startups. Similarly, STEM-educated founders can also be perceived to have higher human capital since their domain knowledge achieved from education can be more directly translated to real world application. Other research (Ratzinger et al., 2018), provides more support for this assumption. Following this logic, the study of education's impact on human capital is a topic that warrants further investigation in the entrepreneurial world.

Geographical Analysis

To gain a deeper understanding of the findings above, different types of maps have been plotted on a state level. Although the information extracted from these maps provide support to the initial findings to a certain extent, there are many inconsistencies throughout the states, making the evidence for the initial finding on a geographic level relatively weak in strength. A map of the average startup funding by state will be used as a baseline to compare with maps of other factors of interest.

Figure 3: *Map of the Average Education Level of Startup Founders by States vs Map of Average Startup Funding by States*



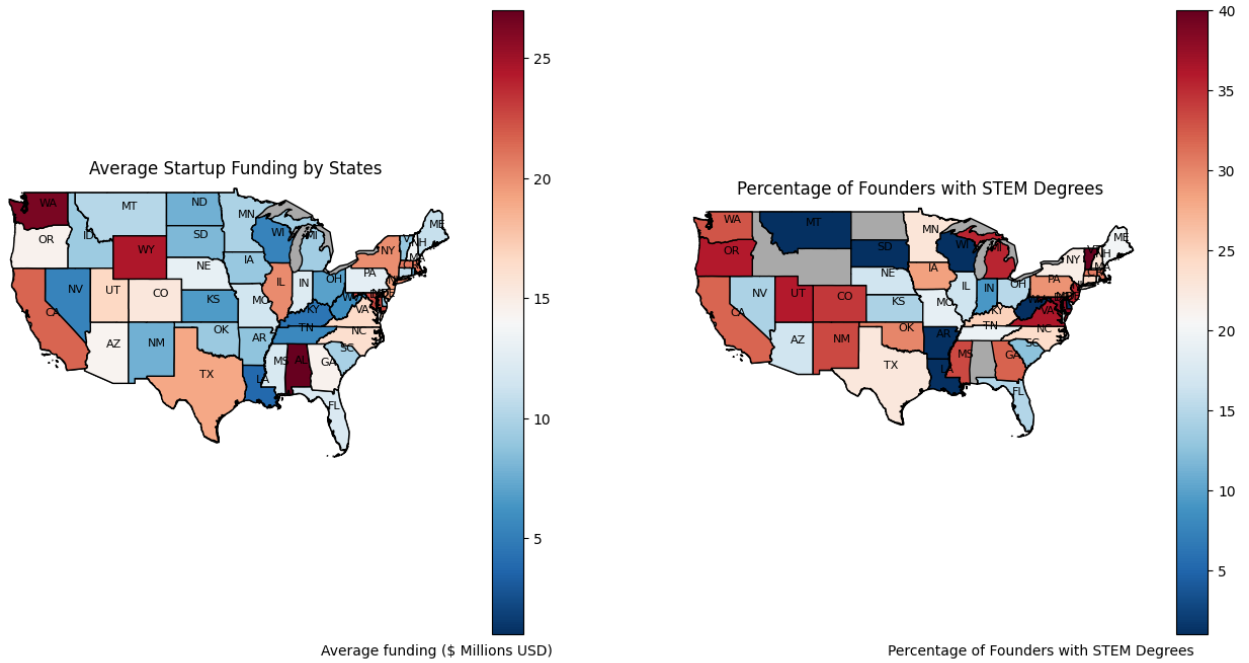
This map depicts the average education level of founders for each state, calculated by an "education index". This education index hierarchically ranks the education level of founders, assuming values of 1 for a Bachelor's degree, 2 for a Master's degree and 3 for a PhD degree. The assumption is that this map will appear similar to the Average Startup Funding by State map to provide more support to the message

that startups by founders that have higher academic qualifications have the potential to receive higher funding.

Due to filtering and the lack of data about founders for certain states in the original dataset, information for certain states such as North Dakota, Wyoming and Idaho are missing. In addition, some states, such as SD, NM and ME have very few startups reported in the dataset (<5) and thus the average education index may not be reflective of the true average education level of founders for the state.

Overall, very few states in the US have an average education index of under 1.3, meaning that for the majority of states, there is a significant proportion of startup founders who have a Master's degree or higher. The West and Northeast are the two regions with consistently above average education index (1.5) across all states while the Southeast and Midwest regions have some major states below this benchmark (a majority of founders from these regions' highest education level is a Bachelor's degree). In comparison with the Average Startup Funding by State map, this map is considerably different, meaning that states that have a higher average educational level among founders do not consistently have higher average funding. However, the average education index for some states is skewed by a low sample size in startups (<5) reported. For the states that have a higher startup sample size (>45), such as California (CA), Texas (TX) and Florida (FL), the correlation between average educational level and average startup funding per state can be supported.

Figure 4: *Map of the Percentage of STEM Startup Founders by States vs Map of Average Startup Funding by States*



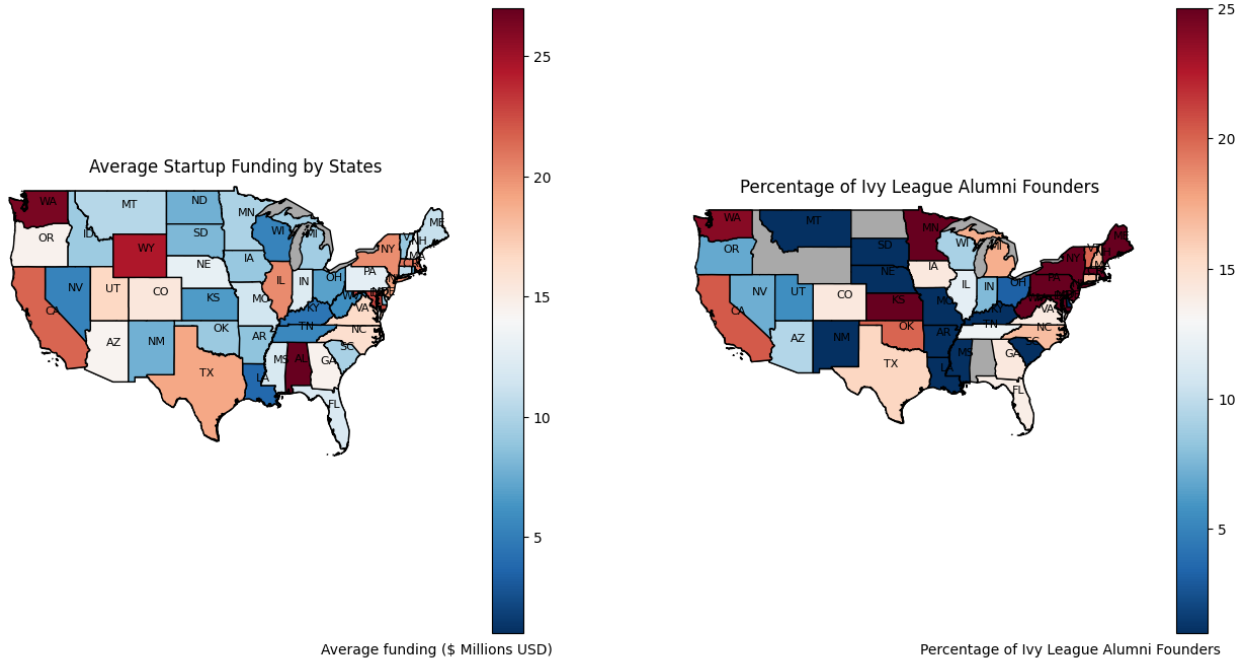
This map depicts the percentage of founders that pursued a STEM field for their highest education level (STEM founders) for each state. The assumption is that this map will appear similar to the Average Startup Funding by State map to provide more support to the message that startups by STEM founders perform better in fund raising then those of non-STEM founders.

Due to filtering and the lack of data about founders for certain states in the original dataset, information for certain states such as North Dakota, Wyoming, and Idaho are missing. At least one fifth of the startups across all states in the Northern region of the US have founders who pursued a STEM subject as their highest level of education. More than 25% of startups founders in 6 major states (WA, OR, CA, UT, CO, NM) of the Western region studied a STEM Subject as their highest education level. Across the US, the majority of US states have at least 15% of startups that have STEM-background founders. Overall

however, most of the startups in the US from this dataset had their highest level of education in a non-STEM subject.

In comparison with the Average Startup Funding by State map, this map is relatively different, despite many similarities. In other words, although many states that have a higher STEM founders percentage have a higher average startup funding, this is not the case for all states. However, it is worth noting that the states that do have an above medium average startup funding (>$15 million) most often have a higher STEM founders' percentage (>25%).

Figure 5: *Map of the Percentage of Ivy League Alumni Startup Founders by States vs Map of Average Startup Funding by States*



This map showcases the proportion of founders who studied at an Ivy League institution for each state based on the office locations. The assumption is that this map will appear similar to the Average Startup Funding by State map to provide more support to the message that startups by founders who attended an Ivy League school perform better in fund raising then those of non-Ivy founders. The Northeast region of

the US, being home to the Ivy League schools, is unsurprisingly the region with the highest Ivy startup founders percentage. It is worth noting that this map indicates that many startups from Ivy League alumni are based near the location of the institutions (as evident from the high Ivy-founders percentage of the Northeast region and California (where Yale is located).

In comparison with the Average Startup Funding per state map, although many states' average funding show correlation with the proportion of Ivy Alumni founders, there are also major inconsistencies with certain states, such as KS, OK and NM, which have high percentage of Ivy alumni founders yet low average startup fundings. Therefore, from a geographical standpoint, the evidence supporting a correlation between the founders graduating from an Ivy League school and better funding performance is relatively weak.

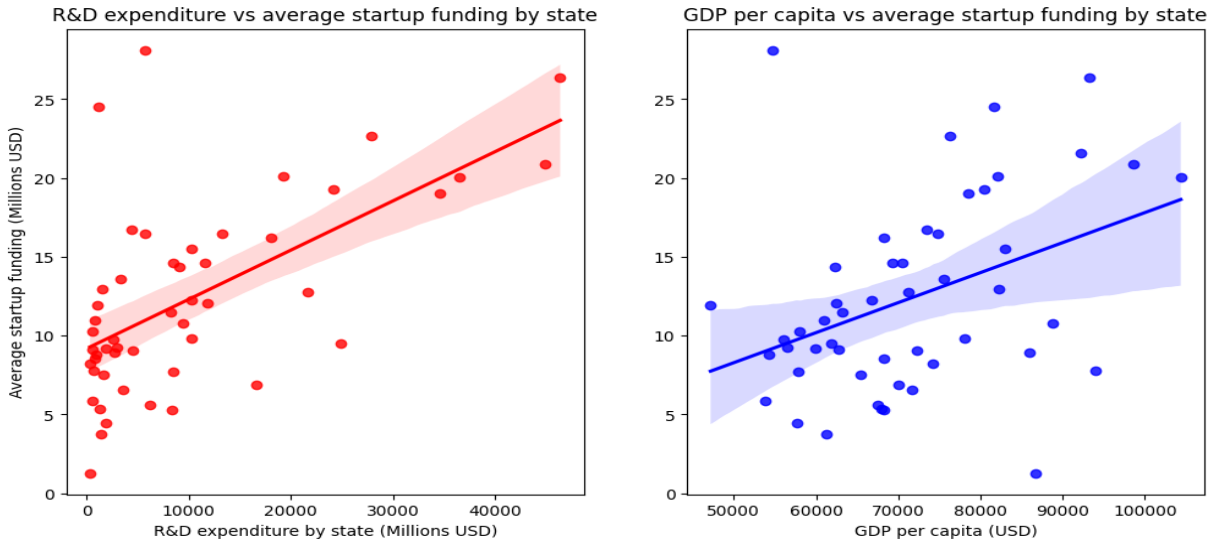
External State Factors

Although a correlation has been established between the founders' education and the funding that startups received on a state level, it is important to note that this relationship may be coincidental. In other words, there are other factors other than founders' education that impact the average startup funding in that state, such as startup resources available, skilled workforce and industry clusters in these states. Wikipedia can provide access to two such factors: the state GDP per capita and R&D expenditure to explore other elements of the states that also potentially contribute to higher average startups funding.

These datasets will be scraped from two different Wikipedia websites: List of U.S. states by research and development spending and List of U.S. states and territories by GDP. This information will be scraped, converted into a data frame, and merged with the average funding of each state data frame. The average GDP per capita for each state is roughly $71,500 and the standard deviation is $12,812, meaning that the amount of GDP per capita by different states do not vary significantly from the average. On the other hand, the average R&D expenditure for each state is roughly $14 billion and the standard deviation is $31

billion, meaning that the amount of R&D expenditure by different states varies quite substantially from the average.

Figure 6: *Correlation between R&D Expenditure and GDP per capita and Average Startup Funding by State*

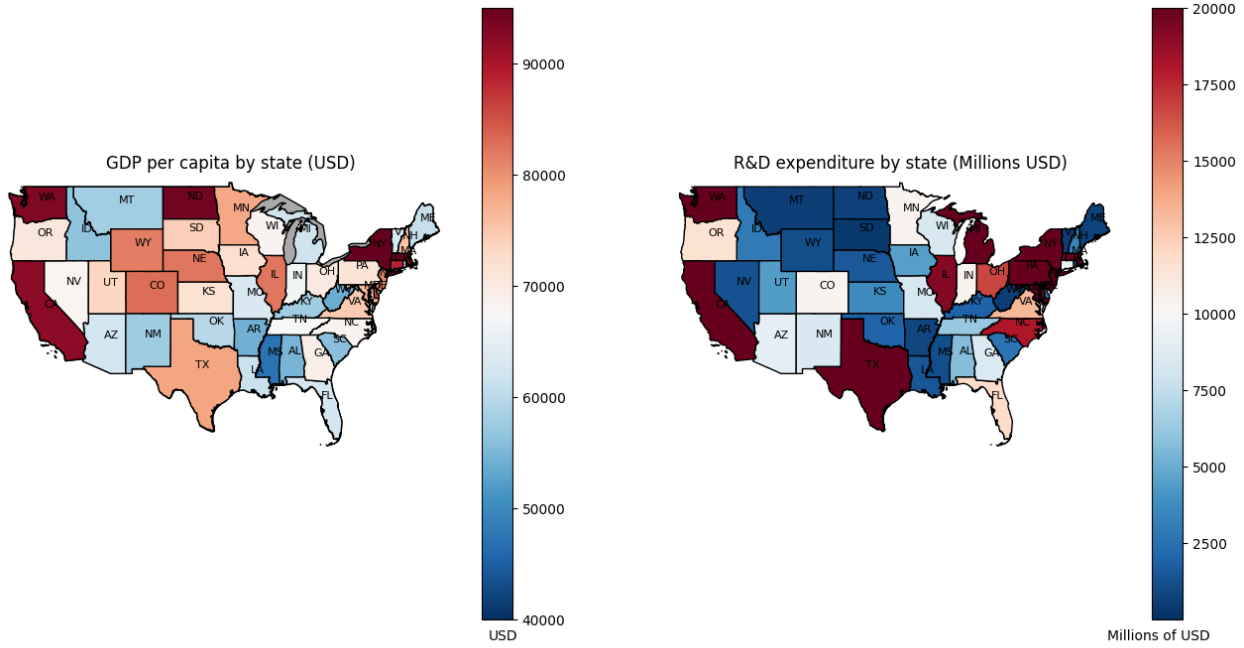


From this visualization it is evident that there are relatively strong correlation relationships between state R&D expenditure, state GDP per capita and the average startup funding by state. In other words, companies in states that expend more on Research and Development, a crucial aspect of startups development, typically receive higher funding for startups, most likely as a result of the impeccable quality of the product produced from the R&D development investment. Although it is worth noting that a significant proportion of funding is reinvested in R&D, so this relationship is reciprocal. Nevertheless, it is likely that companies that have shown strong monetary commitment to R&D have the potential to appear more appealing to investors. In addition, states that have a higher GDP per capita, also tend to have higher startup funding. It is assumed that these states have higher-income investors that are more willing invest more into startups.

All in all, both state R&D expenditure and state GDP per capita may have influence over average startup funding, revealing the possible need to closely inspect the impact of founders' education to access the strength of this factor as a determinant for investors' funding decisions.

Once again, it is important to state that the state of California has been dropped for the left graph and the District of Columbia has been dropped for the right graph due to extreme values. This serves to make the message from the visualization clearer and these datapoints still follow the graph's trends.

Figure 7: *Map of R&D Expenditure and GDP per capita by State*



This visualization provides further information that supports the idea that the correlation between founders' education and the average funding for each state need to be scrutinized. In the two maps, it is clear that certain states such as the West Coast states, Texas and the Northeast states have high levels of GDP per capita and R&D expenditure. These are the states that have high average startup funding, but also relatively high percentage of schools with alumni founders, Ivy-league founders and high average education levels.

In short, the main takeaway is that the positive correlation between educational factors of founders and startup funding is not clear cut, as the state-level data that provides support to this relationship also have other characteristics such as high GDP per capita and R&D expenditure which could also be reasons why startups in these states receive higher funding. This emphasizes the importance of determining the strength of the relationship between educational background of founders and startup funding specifically, in regards of other contributing variables.

Relevant characteristics of founders' educational institutions

Although it has been established that having founders who studied at higher quality educational institutions (Ivy League schools) can lead to higher startup funding, the specific characteristics of these institutions that contribute to startups' success have yet to be determined. The QS University Ranking is one of the most popular rankings for universities, renowned for its impeccable ratings for universities crafted by synthesizing numerous research. The ranking is based on ratings of specific elements of an educational institution. These ratings can provide deeper insights into what specific factors of an educational institution can help founders secure higher funding for startups.

These different factors all contribute to the founders' human capital in different ways, which, as aforementioned, can be linked to higher startup funding.

1. The Employer Reputation ratings can serve as a potential identifier of competent, innovative and skilled individuals, making them more trustworthy by investors.
2. The Citations per Faculty rating is a powerful indicator of an institution's research environment, which can provide substantial value to startups in the forms of resources, connections and inventions (Singha, 2023).

3. The Faculty Student Ratio rating can reveal information such as which schools provide better expert networking opportunities and training for students given the more accessible faculty.
4. The International Faculty and International Students rating shows schools that can provide future founders with the diverse international experience of being taught by and working with individuals from different backgrounds.
5. The Sustainability rating depicts the commitment of institutions to sustainability, and how this may influence young entrepreneurs to incorporate values of sustainability into their startups, which may be highly sought after by investors.

An assumption can be made that all of these characteristics help add to founders' human capital and can increase the funding chances for startups. By examining the QS University Ranking dataset further, it is possible to identify the specific factors from an educational institution that have a positive effect on startup funding. This is very informative for aspiring entrepreneurs who are currently researching tertiary education prospects.

Figure 8: *Summary Statistics for the QS University Ranking dataset*

	Employer Reputation	Citations per Faculty	Faculty Student	International Faculty	International Students	Sustainability
count	190.000000	189.000000	190.000000	185.000000	187.000000	166.000000
mean	20.882368	32.196892	32.399276	25.440541	22.647928	41.214383
std	26.638021	26.699821	30.805575	25.345084	26.688372	33.106535
min	1.400000	1.400000	2.800000	1.100000	1.400000	1.000000
25%	4.300000	12.300000	8.900000	6.000000	4.550000	8.600000
50%	8.500000	21.500000	18.900000	15.900000	10.400000	34.450000
75%	24.350000	43.250000	49.775000	36.400000	29.850000	70.400000
max	100.000000	100.000000	100.000000	100.000000	98.200000	99.900000

The dataset contains ratings for up to 190 different schools, with some ratings missing for certain institutions. On average, Sustainability rating for schools was the highest while employer reputation was the lowest average rating among schools on the dataset, at 41.2 and 20.8 points out of 100. Sustainability rating being higher can be explained by the fact that it is missing ratings for around 24 schools.

Sustainability is also the rating with the highest standard deviation, meaning that individual school's sustainability varies quite substantially for the mean. From an economic standpoint, the main takeaway is that the average rating for Citations per Faculty, Faculty Student and Sustainability for schools in this dataset are higher, meaning that these characteristics of schools are assessed to be moderately better than the other characteristics (International Faculty, International Students and Employer Reputation).

All of the rating distributions are left-skewed, meaning that higher ratings for these characteristics are rarer than lower ratings, specifically for the employer reputation and the international student ratio ratings. A very important observation to note that these characteristics all appear to be positively correlated. This indicates that the dataset is segregated holistically, meaning that there are higher quality schools that score well on all aspects and lower quality schools that score more poorly on all aspects. This will be important to consider when examining the relationship of each characteristic individually on funding. It is assumed that schools that are not rated are schools of lower quality, based on the research-based and expert-reviewed QS assessment. Observing how startups by founders from these non-Rated schools performed compared to Rated schools will also be a potentially interesting observation.

Figure 9: *Correlation between Employer Reputation, Citation per Faculty and Faculty Student ratio ratings versus Average Startup Funding*

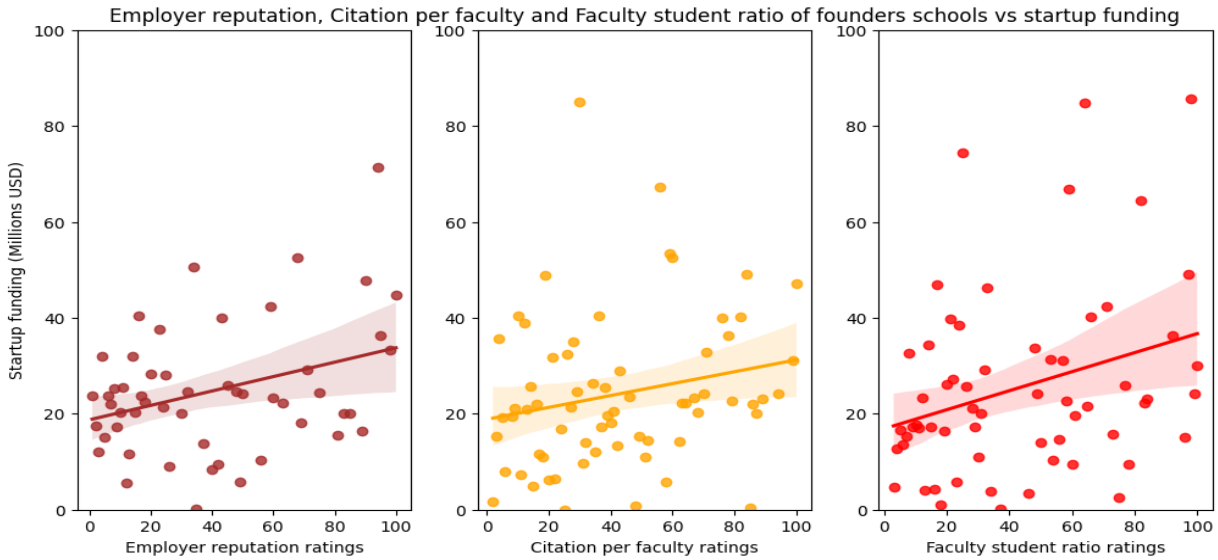
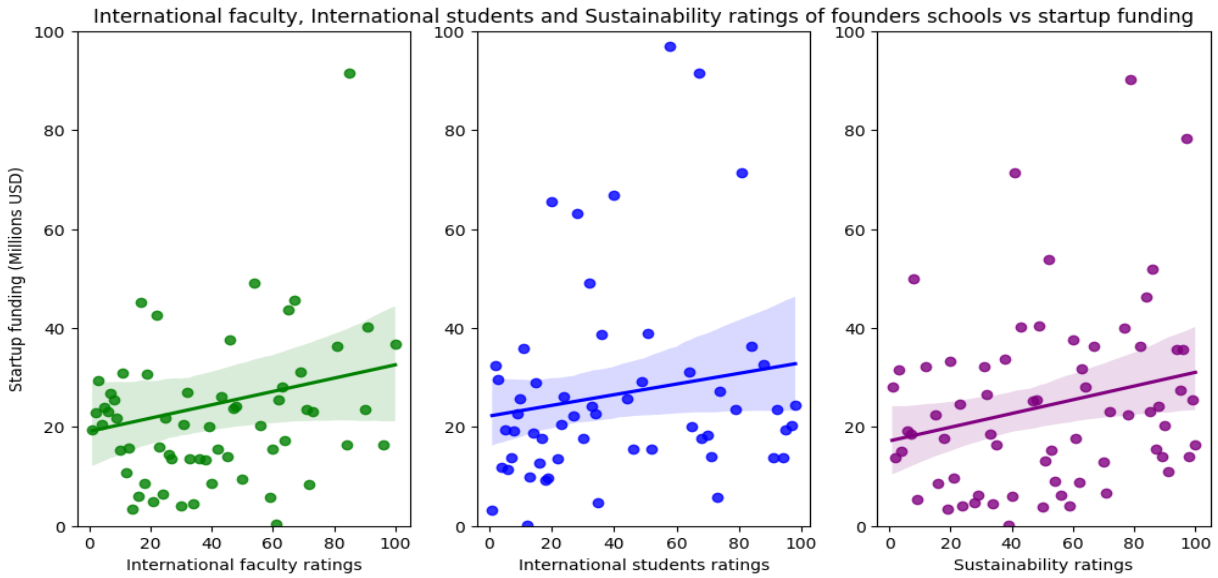


Figure 10: *Correlation between International Faculty Ratio, International Student Ratio and Sustainability ratings versus Average Startup Funding*



All in all, different aspects of founders' schools such as employer reputation, citation per faculty, faculty-student ratio, international faculty, international students' ratio and sustainability all have a marginal positive correlation with startup funding of founders. This means that they appear to be minor factors in the investment decision making process for investors. The similarity between the correlations of all the characteristics with average startup fundings can be explained by the fact that these factors are all

positively correlated with each other to a certain degree. In other words, higher quality schools tend to score well across all areas and train founders that receive higher startup funding.

Based on these graphs, employer reputation and international student ratio have slightly stronger positive correlation with funding than other factors. In other words, employer reputation and international student ratio are potentially more influential factors to higher startup funding. One possible explanation of this trend is that founders benefit from the diverse studying experience with peers around the globe in schools that have a higher international student ratio, and this experience is later translated to them building better startups. Furthermore, founders from schools that have a stronger employer reputation are more likely to be recognized by investors as more competent, which helps with the funding decision process. Multiple regression must be implemented to investigate precisely how each characteristic of the founders' educational institution impacts startup funding. On average, higher quality institutions (using the QS ranking as a reliable measure of quality), tend to produce startup founders that perform marginally better in fundraising. Startups founders from rated schools also have the potential to reach higher levels of funding. This provides further support for the assumption that higher quality overall scores higher in all characteristics, and founders from these institutions receive higher startup funding as a result.

However, the assumption is that all of these x-variables have a weak linear relationship with the y-variable based on the individual analysis of each of these x-variables with startup funding performed above. In all of the correlation graphs visualized above, a linear relationship between the chosen independent variables and startup funding is relatively weak in strength with a large amount of scatter in every graph. These factors showed that a linear relationship is very challenging to establish firmly. This observation can be explained by the fact that different aspects of founders' education, although may play some indirect roles in influencing factors, are not the main factors of consideration for startup investors.

Conceptually, relating to the Human Capital theory, these variables all contribute to improving human capital, which is assumed to be an indication of competency, a crucial factor for investors to consider, impacting startup funding. However, founders' education may only have a very weak relationship with startup funding due to other prominent factors that take priority when investors enter the decision making process. Characteristics relating to the startups themselves are likely more influential towards funding, for example, the business model, profitability, and the skill level of the companies' workforce are assumably more significant elements to consider. In addition, education is also not the only indicator of a founders' competency, as this can be judged from many other aspects such as innovation, charisma and other personal qualities. In short, based on the significant weight that other factors carry in the funding decision making process, only weak linear relationships can be established between different aspects of founders' education and startup funding.

RESULTS

Performing regressions on the datasets helps provide deeper insights into how different aspects of founders' education relate to startup funding. This paper will use Linear Regressions, Decision Tree and Random Forest Models to re-examine the findings made above based on correlation.

12 independent variables (X-variables) have been chosen to measure their influence on startup funding (y-variable). These variables include:

- Highest_education_index (this indicate the founder's level of education with integers Bachelor as 1, Masters as 2 and PhD as 3)
- studied_STEM_field (this indicates whether the founders studied STEM for his highest qualification)
- Attended_ivy_league (this indicates whether the founders studied at an Ivy League school for his highest qualification)

- state_GDP_per_cap_22 (this indicates the GDP per capita in 2022 of the state in which the office of the startup is located in)
- state_R&D_expenditure_in_millions (this indicates the R&D expenditure of the state in which the office of the startup is located)
- Employer_reputation (this indicates the employer reputation rating of the founder's school)
- Citations_per_faculty (this indicates the research environment rating of the founder's school)
- Faculty_student (this indicate the faculty student ratio rating of the founder's school)
- International_faculty (this indicates the international faculty ratio rating of the founder's school)
- International_students (this indicates the international student ratio rating of the founder's school)
- Sustainability (this indicates the sustainability rating of the founder's school)
- Rated_dummy (this indicates whether the founder's school was rated)

OLS Regressions

The analysis from the above sections suggests a relationship between founders' level of education, STEM education and Ivy League education with startup funding. The first three regressions will investigate the individual effects of each variable against startup funding. Regression models (4) to (6) are multivariate regressions which each examine different scenarios where one of the variables are excluded. Lastly, regression model (7) considers the effect of all three variables on startup funding.

Objective functions for OLS Regressions in Table 1:

Model (1): $log(funding) = \beta_0 + \beta_1\, highest\_education\_index$

Model (2): $log(funding) = \beta_0 + \beta_1\, studied\_STEM\_field$

Model (3): $log(funding) = \beta_0 + \beta_1\, attended\_Ivy\_League$

Model (4): $log(funding) = \beta_0 + \beta_1\, highest\_education\_index + \beta_2\, studied\_STEM\_field$

Model (5): $log(funding) = \beta_0 + \beta_1\, studied\_STEM\_field + \beta_2\, attended\_Ivy\_League$

Model (6): $log(funding) = \beta_0 + \beta_1\, highest\_education\_index + \beta_2\, attended\_Ivy\_League$

Model (7): $log(funding) = \beta_0 + \beta_1\, highest\_education\_index + \beta_2\, attended\_Ivy\_League + \beta_3\, studied\_STEM\_field$

***Table 1**: Estimates of Founders' Individual Education Characteristics Against Startup Funding*

							Dependent variable: funding_in_millions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
attended_ivy_league			9.451***		10.262***	9.444***	10.295***
			(3.360)		(3.365)	(3.376)	(3.383)
const	28.261***	26.482***	27.377***	25.796***	24.259***	27.307***	24.579***
	(3.620)	(1.599)	(1.495)	(3.700)	(1.756)	(3.633)	(3.719)
highest_education_index	0.608			0.426		0.043	-0.203
	(2.070)			(2.069)		(2.079)	(2.078)
studied_STEM_field		9.252***		9.235***	9.890***		9.900***
		(2.925)		(2.926)	(2.930)		(2.932)
Observations	5649	5649	5649	5649	5649	5649	5649
R^2	0.000	0.002	0.001	0.002	0.003	0.001	0.003
Adjusted R^2	-0.000	0.002	0.001	0.001	0.003	0.001	0.003
Residual Std. Error	100.719 (df=5647)	100.631 (df=5647)	100.650 (df=5647)	100.640 (df=5646)	100.557 (df=5646)	100.659 (df=5646)	100.566 (df=5645)
F Statistic	0.086 (df=1; 5647)	10.009*** (df=1; 5647)	7.912*** (df=1; 5647)	5.025*** (df=2; 5646)	9.661*** (df=2; 5646)	3.956** (df=2; 5646)	6.442*** (df=3; 5645)
Note:							*p<0.1; **p<0.05; ***p<0.01

The dummy variables for whether the founders attended an Ivy League school and whether a founder studied a STEM field maintain a statistically significant positive relationship with startup funding, after accounting for the remaining two variables. This indicates that these two variables are potentially good predictors of startup funding. This can be supported by model (5), which excludes the educational level of founders and is the most statistically significant model according to the F statistics. According to model (7), holding fixed the education level and whether the founders studied a STEM field, founders that attended an Ivy League school on average raised around $10.3 million more in startup funding than those who did not attend these institutions. Similarly Holding fixed the education level and whether the founders studied at an Ivy League school, founders that studied a STEM field for their highest educational qualification on average raised around $9.9 million more in startup funding than those who did not attend these institutions. On the other hand, according to all the regressions ran in the table, the level of

education of founders appears to share no relationship with startup funding, which is a major contradiction to previous findings. This is further supported by the fact that in every model that includes this variable, the explanatory power of the model (measured by the F-statistic) is lower. In short, the level of education of founders is likely to be a poor predictor of startup funding.

Going forward, **model (7) from table 1** will be chosen as the **preferred specification**. This regression model factors in the 3 different individual educational background characteristics for founders, consisting of the founders' level of education, whether they attended an Ivy League school and whether they studied a STEM field for their highest qualification. Out of these three variables, the coefficient for whether they studied a STEM field and whether they studied at an Ivy League institution are statistically significant. Ultimately, this model is chosen as the preferred specification since it is the most relevant to the research question, to examine the specific aspect of founders' education that can potentially be leveraged to influence startup funding. Machine learning techniques can be utilized to develop a deeper understanding of how these variables correlate with startup funding.

Objective functions for OLS Regressions in Table 1:

Model (1): $log(funding) = \beta_0 + \beta_1\, citation\_per\_faculty$

Model (2): $log(funding) = \beta_0 + \beta_1\, employer\_reputation$

Model (3): $log(funding) = \beta_0 + \beta_1\, citation\_per\_faculty + \beta_2\, employer\_reputation$

Model (4): $log(funding) = \beta_0 + \beta_1\, employer\_reputation + \beta_2\, internation\_faculty$

Model (5): $log(funding) = \beta_0 + \beta_1\, citation\_per\_faculty + \beta_2\, internation\_faculty$

Model (6): $log(funding) = \beta_0 + \beta_1\, employer\_reputation + \beta_2\, citation\_per\_faculty + \beta_3\, internation\_faculty$

***Table 2**: Estimates of Founders' Schools' Characteristics Against Startup Funding*

	Dependent variable: funding_in_millions					
	(1)	(2)	(3)	(4)	(5)	(6)
citations_per_faculty	0.303***		0.239***		0.309***	0.263***
	(0.066)		(0.091)		(0.093)	(0.102)
const	12.628***	18.830***	11.894**	17.979***	12.439**	11.784**
	(4.815)	(4.085)	(4.869)	(4.353)	(4.927)	(4.962)
employer_reputation		0.223***	0.080	0.188**		0.096
		(0.057)	(0.079)	(0.079)		(0.087)
international_faculty				0.054	-0.005	-0.045
				(0.083)	(0.084)	(0.091)
Observations	3112	3112	3112	3092	3092	3092
R^2	0.007	0.005	0.007	0.005	0.007	0.007
Adjusted R^2	0.006	0.005	0.006	0.004	0.006	0.006
Residual Std. Error	114.168 (df=3110)	114.274 (df=3110)	114.168 (df=3109)	114.633 (df=3089)	114.532 (df=3089)	114.527 (df=3088)
F Statistic	20.939*** (df=1; 3110)	15.114*** (df=1; 3110)	10.984*** (df=2; 3109)	7.596*** (df=2; 3089)	10.356*** (df=2; 3089)	7.315*** (df=3; 3088)
Note:						*p<0.1; **p<0.05; ***p<0.01

The rating on different characteristics of schools where founders studied may indicate what specific aspects of these institutions help influence startup funding. From the analysis performed in the previous sections, the ratings for schools' reputation with employers, citations per faculty and international faculty ratio have the most positive relationships with startup funding. Running regressions using these variables will provide more information on the nature of their relationship.

Out of the three independent variables, the citations per faculty rating for schools appears to be the only variable that maintains a statistically significant positive relationship with startup funding, after accounting for the remaining two variables. This indicates that the citations per faculty rating of school is potentially a good predictor variable of startup funding. In other words, the results from the regressions performed (especially model (3) and (5)) provided support for the claim that founders from schools that have a stronger research environment tend to receive higher funding. Holding fixed the employer reputation rating and the international faculty rating, founders that attended school with 1 unit of citations per faculty rating higher on average raised $263,000 more in startup funding.

From the regression model (6), it is evident that the employer reputation rating and the international faculty rating barely have any relationship with startup funding once accounting for the other two variables. This means that the correlation with funding previously observed from these variables are attributable to the citation per faculty rating rather than their own individual impact. This is further validated by regression (4), which did not include citation per faculty rating and had a poor F-statistic, meaning that the model was statistically insignificant.

Machine Learning: Decision Tree and Random Forest Models

Objective Function:

$$\boldsymbol{min} \sum (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y})^{\boldsymbol{2}} + \propto |\boldsymbol{terminal\ nodes\ in\ tree}|$$

Regularization Parameter

In this model there are 3 Regularization Parameters: The Minimum Leaf Size, Maximum tree depth and $\propto$. The Minimum Leaf Size dictates the minimum number of observations taken into consideration at each leaf note in order for a split to happen, and choosing a Maximum tree depth decides The Minimum Leaf Size. A higher Minimum Leaf Size will reduce the depths of the tree and result in a simpler model at the cost of underfitting, where the model cannot capture all the detail to predict startup funding. Following this logic, the Maximum Tree Depth dictates the number of leaf notes and the simplicity of the model, setting these parameters too low will result in underfitting. Lastly, $\propto$ also regulates the complexity of the tree, choosing an $\propto$ will decide in the pruning process whether to focus on the model's ability to explain the data given or on the simplicity of the model. In short, all three of these parameters control the complexity of the model, setting these parameters too high will result in overfitting (where the model learns the data too well and give poor predictions) and too low will result in underfitting (where the model learns the data not well enough and give poor predictions).

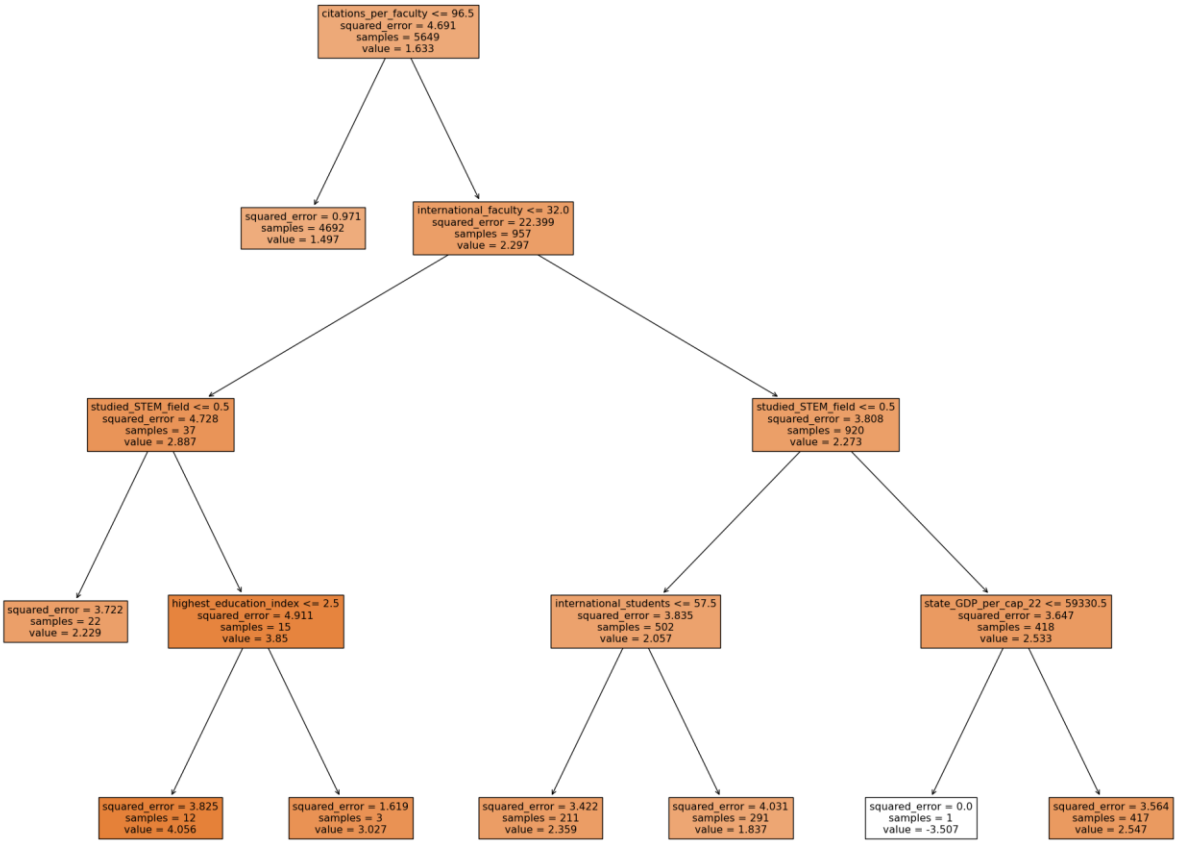
Explanation of the objective function

The Decision Tree model operates based on the objective function above. The model determines the most important features and the threshold of these features to split the data into subregions based on the startup funding associated with these splits. The model then makes predictions about startup funding based on the subregions and calculates the sum of the errors between its predictions and the actual startup funding, which makes up the first part of the objective function. For the second part of the objective function, the $\propto$ determines the degree to which the complexity of the model will be punished to prevent overfitting. The number of terminal nodes in the tree represents the complexity and accuracy of the model. There is a tradeoff between overfitting the model, represented by the first part of the objective function and underfitting, represented by the second part of the model.

The maximum depth is set to 4 for the Decision Tree Model that contains all the independent variables. After tuning, this depth is assumed to be appropriate as it balances the model's complexity and minimizes the standard error. Increasing the depth will reduce the mean squared error at the increased risk of overfitting the data. In contrast, reducing this depth will reduce the risk of overfitting the data at the cost of providing less accurate predictions. A max depth of 3 is set for the model of the preferred specification as there's only 3 explanatory variables.

1. **Decision Tree Model for all independent variables**

Figure 11: *Decision Tree Model for all independent variables*



This Decision Tree model contains all explanatory variables in the main data frame. Institutions' citations per faculty rating is the top node, which indicates that this dependent variable is the most important in determining the startup funding. The model separated the dataset into two main groups: startups that have founders graduating from schools with a citation per faculty rating above 96.5 points and the remaining startups. Based on the value provided, it seems that average startup funding is considerably higher for founders who studied at schools with a citation per faculty rating higher than 96.5 points. After institutions' citations per faculty rating, the model determines the international faculty ratio rating of schools to be the second most important explanatory factor for startup funding, as it split the dataset based

on schools which received an international faculty rating of less than 32. However, since the sample size for schools with this rating range is very limited (37 observations), the results provided from this branch will likely not be reflective of the dataset and will be neglected. For the other side of the dataset, the next most important variable is whether the founders studied STEM for their highest level of education. For non-STEM founders, the average funding of their startup is relatively lower than STEM founders, and those who studied at schools with less international students performed better in fundraising. Lastly, for STEM founders, the next important criteria to consider is whether the state GDP per capita of that startup is higher than $59,330. According to the Decision Tree model, the four most important predictors of startup funding are, in order, the founders' schools' research environment (citation per faculty), international faculty ratio, whether the founders studied STEM and the GDP per capita of the state where the startup is located. In short, STEM founders who studied at schools with higher ratings for citations per faculty, international faculty and have startups located in states with higher GDP per capita, on average, performed better in fundraising.

2. Decision Tree Model for preferred specification

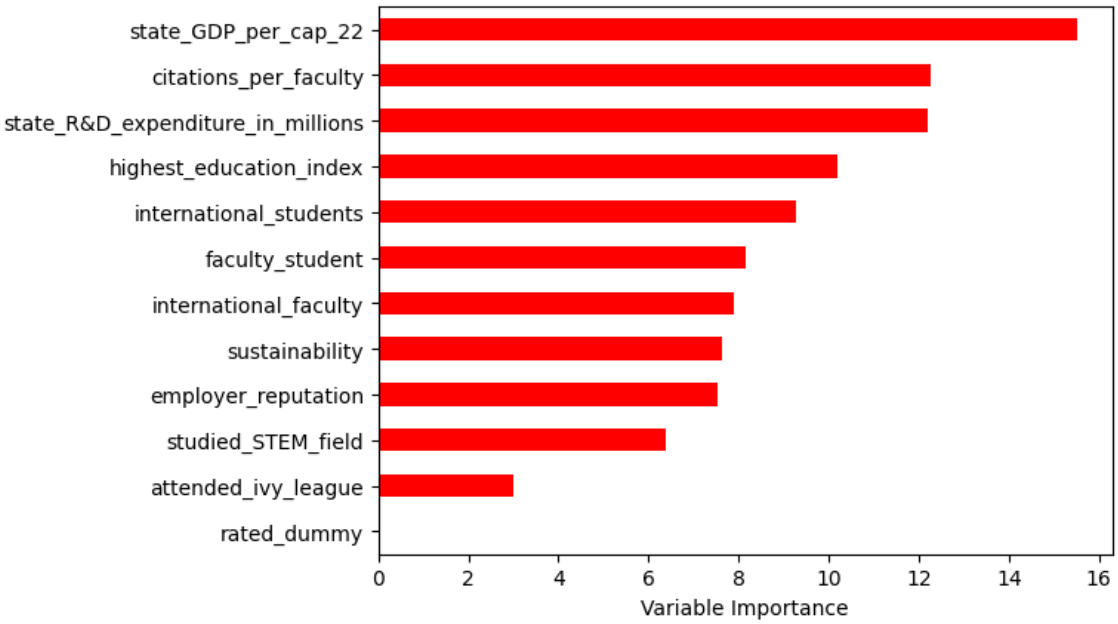
This Decision Tree model contains 3 explanatory variables from the preferred specification: founders' highest level of education, whether the founders studied a STEM field for their highest qualification and whether the founders studied at an Ivy League school. Founders' highest level of education is the top node, which suggests that this dependent variable is the most important in determining the startup funding. The model determines that startups by founders who studied up to the PhD level (having a highest_education index > 2.5) perform considerably better in fundraising, which is consistent with previous findings. For PhD-level founders, whether they attended an Ivy League institution is the next most important factor in determining startup funding, as Ivy League alumni have more success in fundraising. In contrast, for non-PhD founders, whether they studied a STEM field for their highest qualification is a factor of higher concern than whether they attended an Ivy League school. Overall, the trend is that founders with higher levels of education who graduated from an Ivy League school and study

STEM at their highest level of education receive higher funding in general. This is contradictory to the results of the Linear Regression, which determined that, holding fixed the STEM education and Ivy Education factors, the highest education level of the founders share a very statistically insignificant relationship with startup funding. However, both models concur that whether the founders studied a STEM field for their highest qualification and whether they attended an Ivy League school are important determinants of startup funding.

The Decision Tree Model for the preferred specification containing the independent variables about the individual educational characteristic of founders performed slightly worse than the model using all the explanatory variables in the dataset, based on the Mean Squared Error. This means that, under a Decision Tree system, using the level of education, STEM education and Ivy League education factors to predict startup funding is only marginally less accurate than using all the mentioned variables combined with external state factors and relevant educational school ratings. All in all, the Mean Squared Error for both models were still relatively high, indicating that utilizing the Decision Tree technique may not be suitable for this dataset.

3. Random Forest Model for all independent variables

***Figure 12**: Importance Matrix*



The results provided by the Random Forest model utilizing all independent variables are considerably different from the Decision Tree model. According to the former, external factors such as State GDP per capita and State R&D expenditure are more effective determinants of startup funding than most aspects of founders' education, ranking 1st and 3rd in terms of their explanatory power of startup funding. The citations per faculty rating for founders' schools are the second most important predictor of startup funding, and the most important factor of founders' education that may have an impact on startup funding. The founders' level of education ranks 4th in terms of explanatory power for startup funding, contradicting the results by the Linear Regression and the Decision Tree. Out of all the independent variables considered, whether the founders studied a STEM field for their highest qualification or attended an Ivy League institution are the least significant predictors of startup funding, ranking behind other types of school ratings.

4. Comparing Different Models

Compared to the Linear Regression model, the Decision Tree is better at capturing non-linear relationships, which is more suitable for the independent variables in this dataset. The Decision Tree model splits data according to features to maximize the accuracy of the model, therefore better capturing the non-linear nuances presented in the data, whereas the Linear Regression model neglect this factor and falsely assumes a firm linear relationship between explanatory variables and startup funding. Additionally, the Decision Tree model provides information about the importance of the specific factors that can impact startup funding and the important thresholds for each factor, which cannot be achieved with a Linear Regression model. These features of the model help answer the research question in terms of comparing the significance of different independent variables to startup funding.

Judging the models' performances based on R-squared, the Random Forest model has the highest accuracy in using factors of founders' education to predict startup funding, followed by the Decision Tree model and the Linear Regression model. Accounting for the variations in all of the independent variables of founders' education, the Random Forest model can explain roughly 47% of the variation in startup funding, which is significantly better than the 9% and 1% of Linear Regression and Decision Tree, respectively. Similarly, for the models using only explanatory variables from the preferred specification, the Random Forest model still outperformed the Decision Tree model and the Linear Regression models, although to a less significant extent.

In short, the results provided by the Random Forest models are likely the most accurate out of the 3 types of models. This is expected, as the Random Forest model is better the other 2 modes at capturing non-linear complexities between the independent variables and dependent variables. It is also more capable of handling multicollinearity (a big issue in the dataset) and it is capable of providing a measure of variable importance. The Mean Squared Error of the Random Forest Model is also significantly smaller than that of the Decision Tree Model, indicating superior accuracy in predicting startup funding. In short, based on

the most accurate model used in this paper, the factors of founders' education that most impact startup funding is their school's research environment, their level of education and their schools' international student's ratio, in that order. All of these factors contribute to building the founders' human capital in different ways, which may potentially play a part in building better startups, leading to higher funding.

CONCLUSION

As startups navigate the multitude of challenges in staying operational in the current business world, having a deep understanding of human capital and their impact is crucial. This research project has explored the founders' educational background (defined by certain variables) and the amount of funding raised by companies, a pivotal indicator of their success. The Startup Investments dataset provided by Crunchbase has allowed a detailed look into these elements and infer several key insights.

Firstly, from the dataset, it is assumed that the number of art and science degrees attained by founders of a company is not a contributory factor to attract funding. A higher number of Art degrees will not contribute to a company's fundraising endeavors although having more than one Science degree can have a positive impact. Secondly, exploring the first variable further, the average number of degrees held by the founders of startups is also not a clear determinant in the process of securing fundings. In this light, the quantity of degrees of founders does not appear to be relevant to the total amount of funding received by the startups.

On the other hand, the quality of the degrees might play an important role. From the analysis performed above, a higher-level academic qualification (a PhD degree as opposed to a Bachelor degree) has a higher average funding amount per degree for the average startup. In general, startups that have founders with higher academic qualifications tend to have higher funding. Another area to consider is whether founders studied in a STEM or non-STEM field, as founders with a STEM background on average secure higher

funding than those without. Furthermore, the quality of institutions that founders graduated from is also considered, as the analysis indicates that founders that studied at Ivy League schools are associated with higher-funded startups compared to those who studied at other institutions. This is further consolidated by the finding that schools with top rankings tend to attract slightly higher funding compared to those not ranked. Additionally, the characteristics of founders' schools play a role in influencing startup funding. This paper found that the employer reputation, citation per faculty, faculty-student ratio, international faculty, international students' ratios and sustainability of founders' schools could positively impact fundraising for startups to a certain extent.

Various state-level maps provide more support for these findings, despite some inconsistencies arising from the differing startup distributions across states. Essentially, states with higher average education levels and a greater proportion of STEM founders tend to exhibit higher average funding for startups, although this pattern is not consistent across all states. Another interesting finding is that the West Coast and Eastern regions of the US, particularly the Northeastern states, are densely populated with startups offices and have a high percentage of schools that produce startup founders. New York, Pennsylvania, and Texas consistently produce founders securing medium to high levels of funding for startups.

Using different Regression and Machine Learning models, this paper was able to determine the factors that have the most significant relationship with startup funding. Firstly, the external state factors such as state GDP per capita and State R&D Expenditure are two of the most influential elements to determining the amount of startup funding. Higher funded startups tend to be located in higher-income and R&D-centric states. The research environment of schools where founders studied is determined to be the most important factor in predicting startup funding out of all the elements of the founders' educational background. Other schools related factors that also have a significant relationship with startup funding are the international student ratio and the faculty-student ratio. In terms of individual educational characteristics, the founders' level of education is the important factor in predicting startup funding,

higher educated founders are associated with higher funded startups. Lastly, the founders' STEM education and Ivy League education are the least influential variables when it comes to determining startup funding.

It is important to note that this project has several limitations and areas where future research can build on. Firstly, the original dataset has been modified through a process of data cleaning, filtering, and aggregating. As a result, the remaining sample size has been reduced significantly. Second, the original dataset had many missing values and inconsistencies that cannot be thoroughly addressed by the data cleaning process, leading to possible inaccuracies in the analysis. On the geographical level, there is missing data for many of the states, this is an aspect that could be investigated further to provide a more well-rounded picture. Additionally, future research can explore more complex models that better fit the data than the Random Forest Model, which can provide a more accurate assessment of the degree of impact that educational background factors have on startup funding.

From this analysis, future research should focus on determining causation in areas where patterns between academic background and funding have been found. Although this paper has established the ground for which potential educational factors can impact startup funding, future papers can aim to explain why these relationships exist and build on the theory and assumptions made in this paper.

REFERENCES

Adedeji, Olugbenga S., and Omolara A. Campbell (2013). "The Role of Higher Education in Human Capital Development." Social Science Research Network, January. https://doi.org/10.2139/ssrn.2380878.

Croak, Mallory (2018). "The Effects of STEM Education on Economic Growth." Union | Digital Works. https://digitalworks.union.edu/theses/1705/.

Degree-granting postsecondary institutions, by control and classification of institution and state or jurisdiction: (2016). US National Center for Education Statistics. https://nces.ed.gov/programs/digest/d17/tables/dt17_317.20.asp

Franco, S, et al. (2021). Founder Education and start-up funds raised. IEEE Engineering Management Review. https://ieeexplore.ieee.org/document/9428508

Furqon D. A., Permana M. Y., & Komarudin N. A. (2023). The Influence of The Entrepreneurs' Background on Digital Startups Attaining Investment Milestones in Indonesia. Indonesian Journal of Business and Entrepreneurship (IJBE), 9(3), 502. https://doi.org/10.17358/ijbe.9.3.502

Location of postsecondary institutions by states (Datafile: hd2022) . US National Center for Education Statistics. https://nces.ed.gov/ipeds/datacenter/datafiles.aspx?sid=e0c7fe2c-203f-4077-ad58-99e7316bffcf&rtid=1

Ratzinger, D, et al. (2018). The impact of digital start-up founders' Higher Education on reaching Equity Investment Milestones. The Journal of technology transfer. https://pubmed.ncbi.nlm.nih.gov/31258242/

Ross, Sean. 2023. "What Is the Human Capital Theory and How Is It Used?" Investopedia. September 17, 2023. https://www.investopedia.com/ask/answers/032715/what-human-capital-and-how-it-used.asp.

Singha, Rahul. (2023). "QS World University Rankings Methodology 2023." Shiksha.Com. June 9, 2023. https://www.shiksha.com/studyabroad/qs-world-university-rankings-methodology-articlepage-2777.

Winters, John V. 2014. "STEM Graduates, Human Capital Externalities, and Wages in the U.S." Regional Science and Urban Economics 48 (September): 190–98. https://doi.org/10.1016/j.regsciurbeco.2014.07.003.