Do Your Startup's Location and Sector of Operation Determine Its Venture Capital Funding?

Parmis Mehdiyar

Department of Economics, The University of Toronto

Introduction

The startup world, which emerged primarily in the United States with the rise of Silicon Valley in 1970s, has seen rapid growth in the past three decades, particularly after the dot.com boom in 1990s (Patel, 2022). While the US remains home to the largest number of startups, today, entrepreneurs across the world are inspired to launch startups that take over traditional practices in a wide variety of industries. In fact, many have already done so, and advanced their startups to witness M&As and IPOs. However, this is only the case for a very small fraction of startups – only 1 out 10 startups succeed, with more than 20% of startups not surviving longer than a year (Patel, 2022).

While success is primarily attributable to precise marketing, product development, successful R&D and strong management, access to sufficient funding is perhaps the most vital factor determining whether a young startup survives. Venture capital funding in particular is a major source of funding in the early stages, and their yearly investments exceed $100B in the United States alone. Although the aforementioned signals of success are theoretically the most important factors in attracting funding for startups, systematic differences in fundings could also exist amongst different classifications of startups. Particularly, given the distinct historical developments and diverse economic structures of different sectors and countries, a startup could be subject to systematic differences in funding levels, depending on the country in which it is established in, and the industry that it operates in. The prospect of such systematic differences in funding levels is the main motivation of this study.

So far, literature on venture capital funding for startups has had a much stronger focus on how beneficial access to venture capital financing is to early-stage startups. For example, Krishna et al developed a precise model using machine learning tools, namely Random Forest, ADTrees and Bayesian Networks to predict the successfulness of startups, as measured by the milestones they achieve. Their models consistently discovered that venture capital funding amounts are of significant importance in determining a young startup's success (Krishna, 2016). Similarly, a study by Davila et al suggests that startups that have venture capital support grow much faster in relevant measures, such as sales and revenue, human capital, steady cashflows and supplier network (Davila, 2003). The Wall Street Journal authors also seem to agree that none venture-backed companies fail more often than venture-backed companies (Gage, 2012). Venture capital funding is thus, undoubtedly beneficial to young startups.

What most literature fail to address beyond proving these benefits, is which startups actually receive venture capital financing, and what explains the differences in funding levels between them. This is perhaps because it is generally believed that venture capital financing, and its magnitude, are based on unbiased financial valuations of anticipated potentials of a startup. However, this is not always the case. An important study that gives insight into this issue was conducted by Miloud et al, who agreed that in their valuation, venture capital firms do consider factors that are important to firm performance, but only when that information is available and is reliable. However, for new ventures in early-stages, lack of historical accounting information, and potential of asymmetry of information leads them into using alternative models to evaluate the growth prospects of the startup, and one of the factors that they consider in this process is their own assessments of industry attractiveness (Miloud, 2012).

In a different study by Bruton et al, the researchers used survey results to identify that venture capital investment decisions are made much differently in East Asia than they do in the West, as East Asian venture capital firms consider very different factors in evaluating funding opportunities. Precisely, due to developed regulatory institutions, it is much harder for venture capital firms to find valid and reliable information on startups. Therefore, they tend to base decisions much more strongly on established rapports, networks and relationships with startups when considering funding (Bruton, 2004).

The first set of reviewed literatures provide strong reasons for developing a framework to determine whether startups, based on the country in which they are established and the sector they fall under, receive different magnitudes of venture capital funding, as magnitudes of venture capital funding are obviously major factors of success. Furthermore, Bruton and Miloud's studies provide a basis for my skepticism of a startup's funding being independent of its country and sector of operation: industry growth forecasts may play a role in determining startup fundings in that given industry (Miloud, 2012), and venture capital firms in East Asia and the West have different approaches to making financing decisions for startups. To the extent of my knowledge, this paper is the first to investigate systematic differences in funding levels for startups in different countries and industries. This will consequently increase entrepreneur's likelihood of establishing successful startups by advancing their knowledge beyond the importance of venture capital financing, demonstrating factors that enhance their chances of actually securing that venture capital investment.

This paper therefore aims to discover whether the location and the sector (industrial or service) of a startup, separately or in conjugation with one another, correspond to systematic differences in venture capital fundings (hereby fundings) of startups, and if so, what are these differences. More specifically, I look at whether the income level of a startup's country of origin, as described by its GNI per capita, explains variations in fundings across startups in different countries. I also explore whether the sectoral income levels, as described by the GDP of a particular sector in a particular country, explains variations in fundings across startups in different sectors, i.e., industrial or service sector, of different countries. Finally, I explore how the interactions between country-wide income and sectoral income levels affect funding levels of the startups they contain. The theoretical framework supporting the choice and combinations of these two variables and the measures used will be justified using economic theories in detail under the theoretical framework section. Furthermore, the original hypotheses that guide the research question will be reviewed and the theories that support these hypotheses will be discussed in great detail.

Theoretical framework and hypotheses

In macroeconomics, investment (I) is a function of change in income (ΔY) and interest rates (r), where income affects capital investment positively, as a rise in GDP corresponds to increased production and economic transactions, signaling optimistic investment opportunities.

$$I = f(r, \Delta Y)$$

On this basis, I hypothesize that countries with higher income levels should have higher startup funding levels on average (H1). Applying the same logic to sectoral income, I also hypothesize that for a given sector in a given country, higher sectoral income should correspond to higher funding for startups in that sector in that country (H2). This is consistent with Miloud et al's research, which had suggested that industry forecasts tend to be used as an alternative tool for evaluating young startups with no previous accounting data (Miloud, 2012). This simply means that if the venture capital analysts predict an increase in revenue in a given industry, as profit-maximizing investors, they are more likely to fund a startup in that industry.

My last hypothesis is quite abstract: that the described relationship may not apply uniformly to all countries (H3). Based on Bruton's findings, I hypothesize that funding patterns differ in East Asian countries, particularly expecting higher funding levels on average in startups where funding does exist, as the strong rapport and relationships between investors and entrepreneurs could mean that investors trust the startups that they fund and the information that they receive from them, hence increasing transparency and reducing barriers to higher funding levels on average. For East Asian countries, I also expect the funding levels to be higher on average in traditionally offshore-outsourced sectors and industries. This is because they likely have deeper and more historical roots in the East Asian economies, possibly leading to improved networks and rapports amongst venture capital investors and their clients, which would consequently reduce barriers to higher average funding levels.

Data and research methodology

The main data used in this study comes from a diverse dataset containing published information about the startup ecosystem across the globe until 2013, including the industry, funding amounts and geographical location of startups. This dataset contains 196,553 observations of startups, and is publicly available on Kaggle.com. For the purposes of this analysis, I mostly focus on 3 columns: Total Startup Funding in USD (continuous variable), startup country (categorical variable) and startup category of operation (categorical variable). Note that throughout this paper, I often use the mean average of startup fundings in USD in a given country, or a given sector of a given country as my dependent variable for improved analysis and visualizations. I also use average fundings for OLS regressions.

I later merge this dataset with 2020 Gross National Income (GNI) per capita (USD) provided by the World Bank. I chose the newer data set to keep the research more relevant to the current world economic hierarchies. I believe that for the purposes of this analysis, this inconsistency should not matter as the general differences in income levels, which is what I analyze, is still preserved. Finally, I scrape Wikipedia, while fully abiding by their terms of use, for the most recent (2017) data on each country's sectoral GDP (USD), particularly for service and industrial sector's GDPs, as startups do not fall under the agriculture sector by nature. For consistency, all monetary values (GDP, GNI and funding) are presented in USD.

Note that I use GNI per capita as a representation for income level in each country. I prefer this measure over the GDP, or GDP per capita, as there are standard universal thresholds for this measure that allow for dividing countries into income groups (i.e., low-, middle- and high-income groups). According to the World Bank, low-income economies are those with a GNI per capita of 1,045 USD or less in 2020; lower middle-income economies are those with a GNI per capita between 1,046 USD and 4,095 USD; upper middle-income economies are those with a GNI per capita between 4,096 USD and 12,695 USD and high-income economies are those with a GNI per capita of 12,696 USD or more (World Bank Data Help Desk, 2020). For the purpose of this analysis, I categorized upper middle-income countries to simply middle-income, and joined lower middle-income and low-income economies together as low income, as there were not enough startup funding data on lower middle-income and low-income economies separately. The constructed dataset, being the income level in a startup's home country and measured by GNI per capita, constitutes one of the independent variables.

Next, it is important to note that for this study, I categorized startups based on their category of operation into services and industrial functions. In doing so, I use startups in the top 15 most popular (with most observations) categories only, as opposed to all 42 categories, since this captures a large enough sample size of 91,632 observations without missing variables, and classifying the remaining categories into sectors would not only be time consuming, but also not very informative (appendix A). Next, I used my own judgement and common knowledge to categorize these fields (Appendix B). Notice that there may be some sectoral overlap for certain sectors, the analysis and determination of which is beyond the scope of this paper. Future replications could classify these categories more carefully and more scientifically. Alternatively, data specific to individual industries (rather than sectors) would allow for both of these steps to be skipped, and thus would be preferred for a more detailed analysis. However, the available data would still allow for sufficient and convenient analysis to determine whether income levels can explain variations in funding across countries and more specific fields. By matching the constructed data with sectoral income datasets, the second independent variable is obtained: income level in a startup's sector (within its home country), as measured by GDP of that sector

.

Summary statistics

First, in describing the main dependent variable, the total funding that a startup receives, summary statistics suggests that funding is overall extremely positively skewed, such that even at the 75% percentile, startups have no funding. Furthermore, the standard deviation ($26,034,848) is much larger than the mean ($2,101,193), which further emphasizes the skew of fundings. This high skew simply means that most startups received no funding, while a few received increasingly large sums of funding, with the largest venture capital funding for a single startup being $5.7B. This could be indicative of the high degree of competition between startups to attract venture capital funds (Appendix C).

When categorized by the country in which they were established, it is observed that the largest number of startups, 51,637 of them particularly, belong in the United States, followed by Britain, India, Canada and Germany. However, the US ranked 7th in highest average fundings, with Bermuda, Luxemburg, China, Cayman Islands, Malaysia and Switzerland having higher average fundings. According to figure 1, Other than in China, Singapore and Malaysia, no other East Asian country has noticeably high startup fundings on average. This does not contradict my hypothesis based on Miloud et al's findings (H3) as the researchers also primarily focus on startups from these East Asian countries for their observations (Miloud, 2012). This is because, as opposed to the west, startup culture is not as strong in smaller East Asian economies, especially up until 2013. In fact, China is the only East Asian country that stands amongst the top 15 countries with most startups. It is important to note that many countries have an average funding of 0, which indicates that there is venture capital lending in that country. This phenomenon is mainly visible in Antarctica, sub-Saharan Africa and few countries in Central Asia and Central America (see figure 1). Finally, countries with high fundings do not share any significant geographical characteristics, such as region, access to sea or land area, hence the importance of classifying them by income levels.

Once countries are classified by income levels, it is immediately visible that higher levels of GNI per capita do not correspond perfectly to higher average funding for startups in a country. Figure 2 demonstrates some overlap between high income countries and countries with high startup fundings on average, particularly in the case of US, Switzerland and Israel. However, China, along with Russia, Malaysia and Swaziland do not reflect that relationship: while they have some of the highest startup fundings on average, they are not high-income countries - China, Malaysia and Russia are all middle-income countries. Next, Australia, and some Scandinavian countries such as Finland, Iceland and Sweden somewhat contradict the mentioned relationship as well, as startup average fundings seem to be disproportionately low given the country's high GNI per capita. In this case however, more quantitative analysis is required to determine whether the magnitude of this divergence is of any significance. Other than in the mentioned locations, the GNI in most locations appears to be consistent with the average startup funding in that country.

This observation suggests that while we can expect a positive relationship between a country's income level, as measured by GNI per capita, and the average startup funding in that country, this relationship does not apply to middle-income countries. Startups in this subgroup of economies tend to receive disproportionately more funding on average, given the country's income level. This could reflect some consistencies with the earlier findings by Miloud et al, given there is a large overlap between middle-income economies as described by this paper, and East Asian economies, as described by Miloud et al’s research. Therefore, it could be the case that existing structures in middle-income countries have caused for venture capital funding decisions to be more heavily based on rapports, networks and relationships between investors and startups, than the income levels that signal growth potentials of domestic markets. I will be commenting more on the disproportionate funding levels in middle-income economies in the next sections, when I demonstrate the relationship between average funding in different operational categories and country's income level.

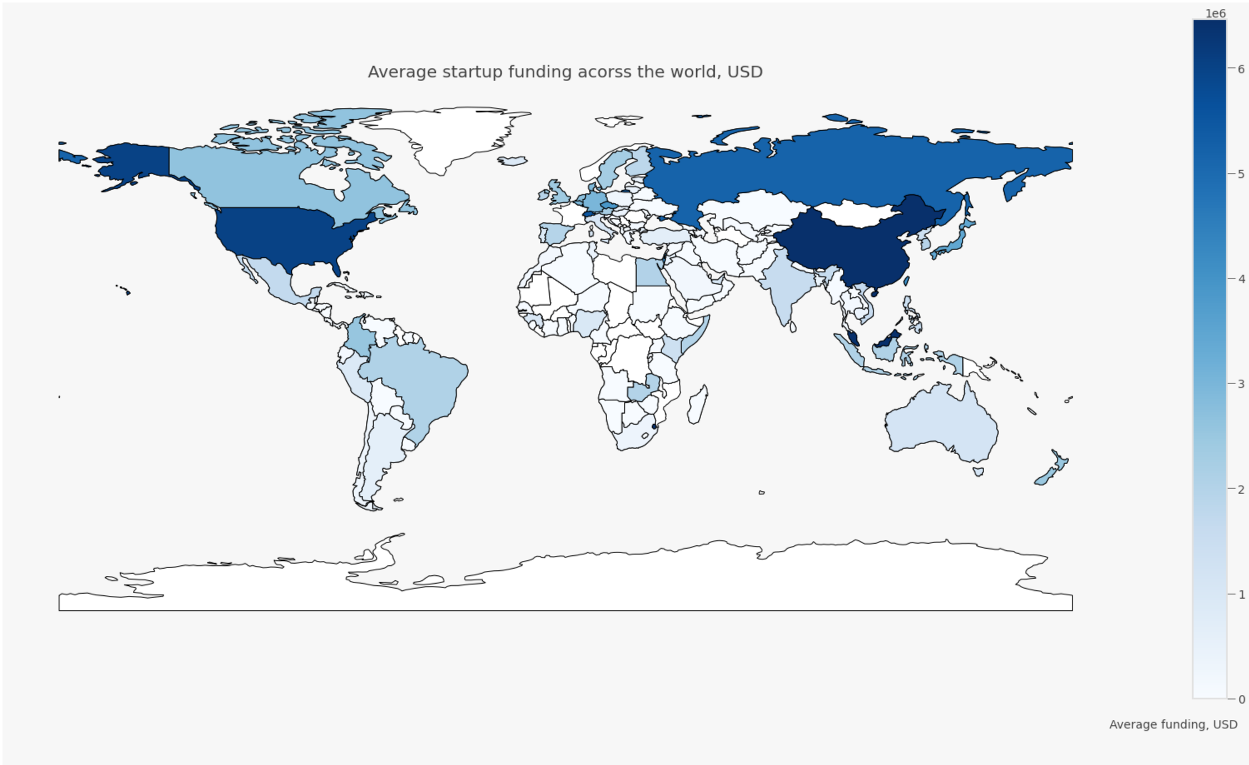


Figure 1, Average startup funding across the world, USD

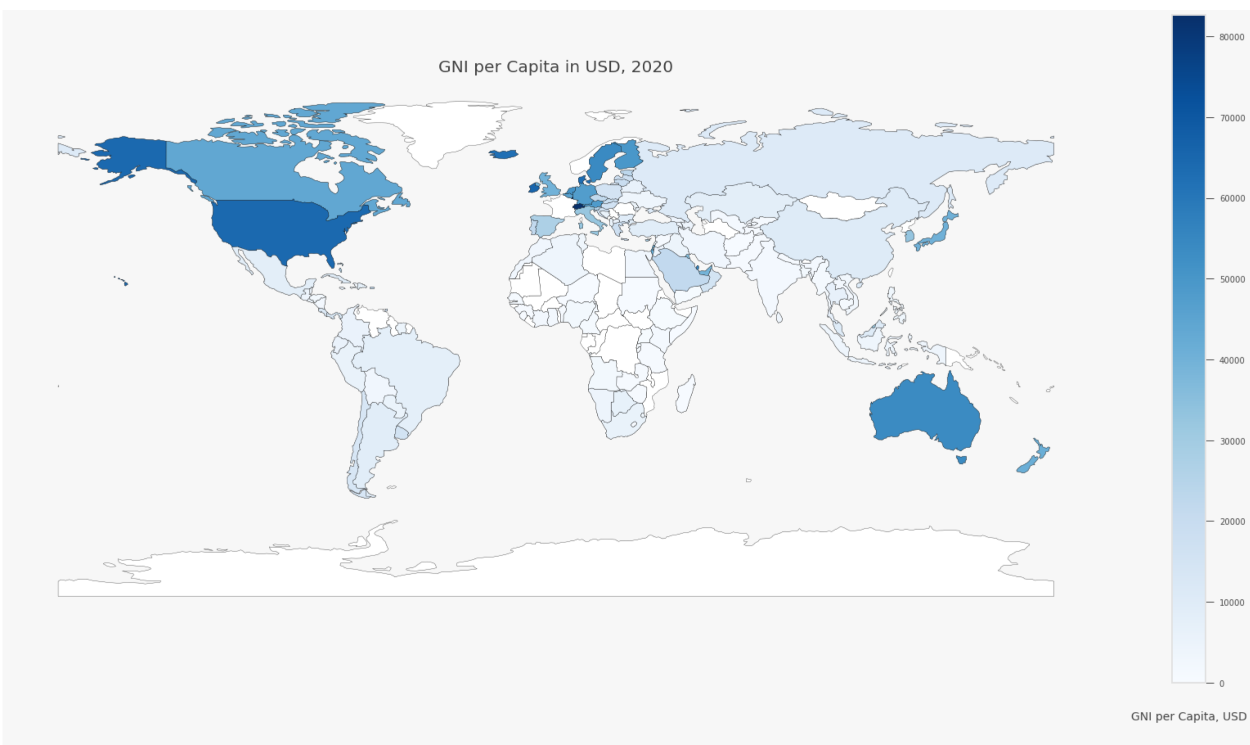


Figure 2, 2020 GNI per capita across the world, USD

Next, out of 15 most popular categories of startups, software was the most popular category of startups, attracting 17,922 out of 91,632 startups in the observation. Furthermore, clean technology, biotechnology and network hosting startups received the highest fundings on average respectively (see figure 3, panel B). Other than being heavily tech-based, what these categories share is being relatively new, as opposed to older tech-based categories that were traditionally often offshore-outsourced. For example, while categories like software and web are significantly more technologically-based, they received significantly less funding on average. This is also evident through the number of startups established in each category. For example, there are much fewer firms both clean technology and biotechnology, than in web and software (see figure 3, panel A). This could reflect the infancy of those industries. It is also likely that high scientific barriers to entry have been keeping the industry small in terms of number of startups, yet fruitful for entrepreneurs.

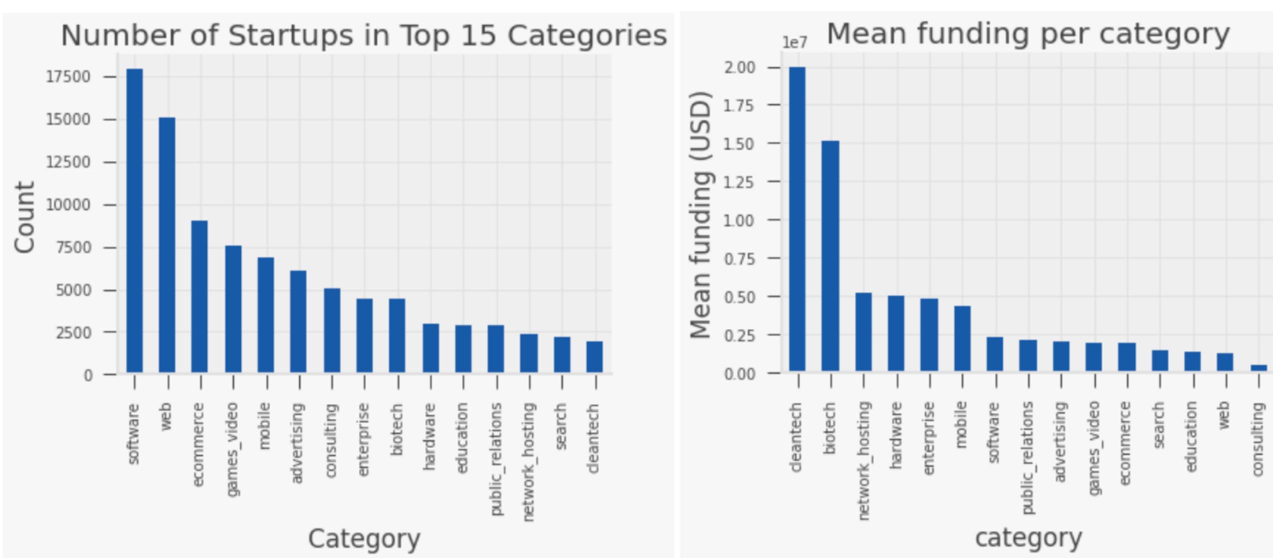


Figure 3. Panel A (left): Number of startups in the top 15 most popular categories. Panel B (right) Mean funding per category, in USD

Next, we visualize how the average venture capital funding for startups varies between countries, particularly based on the country's income level and across these categories. This is the first step towards building a model that does not ignore the intersection of location and sector when explaining factors that systematically affect startup funding. First, we visualize these results simply using the existing categories, and in the next step (regression analysis), we classify them into sectors for an analysis based on the sectoral income levels.

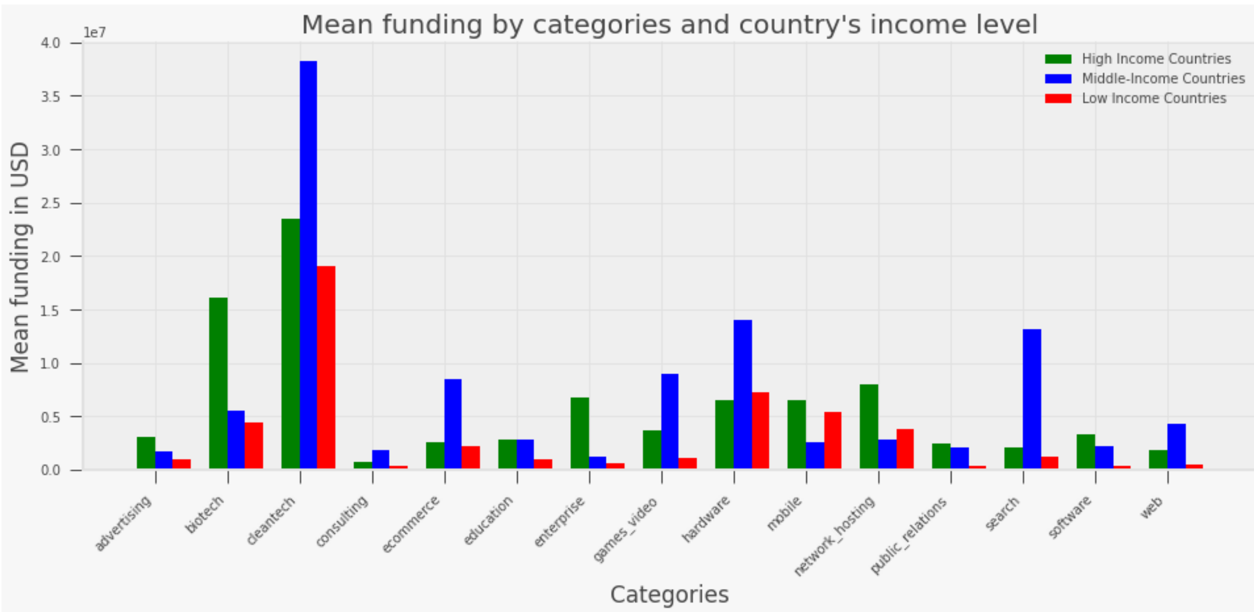


Figure 4, Mean funding (USD) by category and country's income level for top 15 most popular categories

Figure 4, constructed using GNI per capita data and mean funding for startups in 15 most popular categories, is consistent with the previous predictions and hypotheses. Firstly, almost all triads display a difference in fundings for their category, indicating that funding levels vary between high-, middle- and low-income countries. There are few exceptions to this case: ecommerce startups are similarly funded in high- and low-income countries. This similarity is also somewhat observed in hardware, mobile and search industries. This is perhaps because functions under these categories are often very easily offshored and outsourced, hence more competition on a global scale and the insignificant difference in funding. Next, education and, to some extent, public relations and software startups are funded similarly in high income and middle-income countries and biotech startups are almost funded similarly in middle-income and low-income economies. Next, the plot also shows how funding levels vary across startup categories. Again, venture capital funders seem to favor tech-based industries, with clean technology dominating across all income classifications, while consulting maintains the lowest funding in high income and low-income countries, and one of the lowest fundings in middle-income economies.

Finally, the plot suggests that fundings are systematically - for all categories - sub-optimal in low-income economies. However, out of the context of specific industries, the plot shows no systematic relationship between a country's GNI per capita and average funding when comparing high income and middle-income economies. Therefore, it is only wise to compare funding levels across high and middle-income countries within specific startup categories. For example, high income countries, on average, receive more funding in fields of advertising, biotech, enterprise, network hosting and software, while middle income countries on the other hand receive significantly more funding for startups involved with clean technology, ecommerce, video games, hardware, search and web. This could somewhat be explained by the distribution of these categories across high and middle-income countries: middle-income countries, like China, have been dominating in outsourced technological fields, like hardware, web, search and ecommerce, for many years. Therefore, existing networks and rapports between investors and industries are likely strong in these fields, which is proven to be a major factor affecting venture capital funding decisions in East Asia (Miloud, 2012).

Results

Originally, I developed three different hypotheses based on the existing literature and economic theories on the relationship between a startup's funding and the income levels in its domain (geographical and sectoral). More specifically, I anticipated a positive, linear relationship between a startup's funding, and the income level in its country of origin (as represented by GNI per capita) (H1). Similarly, it was also hypothesized that there is a positive, linear relationship between the startup's funding and the income in its specific sector within that country. Given the literature on differences between valuation decisions in East Asia and the west, and my preliminary visual analyses that displayed disproportionate relationships between income and average startup funding in middle-income economies, I predicted that the relationships expected in H1 and H2 would be different for middle-income economies. Specifically, I anticipate a non-linear relationship, potentially parabolic, between GNI per capita in a country and the mean startup funding levels in that country.

To demonstrate these relationships and analyze them formally, I run two main sets of simple regressions. First, I run an ordinary least squares regression where the dependent (Y) variable is the log of mean startup funding (in USD) in a given country, and the independent (X) variable is the log of that country's GNI per capita (measured in USD) in 2020 - a measure of income level in that economy. Within that section, I adjust the independent variable twice to explore how the relationship holds or differs for specific subgroups - particularly, omitting middle-income countries, and omitting both low and middle-income countries.

The second set of simple regressions that I run is also an ordinary least square, aimed particularly at testing whether funding levels in middle-income economies could be better explained using the income of the sectors in which they lie. This is based on the notion that smaller subgroups, like sectors or industries, may contain existing networks and relationships that influence how venture capital firms choose to fund startups. In this section, I run two different sets of regressions to analyze how a startup's funding changes with its sector's income level in that country. Here, the dependent (Y) variable is the log of mean startup funding (in dollars) in a given country, in the given sector. The independent (X) variable is the log of GDP in that country in that sector, as of 2017.

For all regressions, I use a logarithmic scale to be able to respond to skewness towards large values without classifying them as outliers and dropping them. Large values in this study are particularly important as they represent economies with highest funding levels and highest overall or sectoral incomes, which makes them important observations to consider. Furthermore, it is important to note that all countries with a mean startup funding of 0 have been omitted from the regressions in order to avoid errors when converting to a logarithmic scale. Therefore, in this section, I focus on answering questions about the differences in fundings for startups who receive any funding at all, rather than identifying which startups receive funding.

OLS regression set 1: country-wide income and average fundings:

1.1 GNI per Capita on Mean Funding: inclusive

First, I study the relationship between a country's income level, and the average funding for startups in that country. This regression includes countries of low, middle and high income:

$$log\ (average\ start\text{-}up\ funding\ in\ a\ country) = \beta_0 + \beta_1\ log\ (GNI\ per\ capita) + \epsilon$$

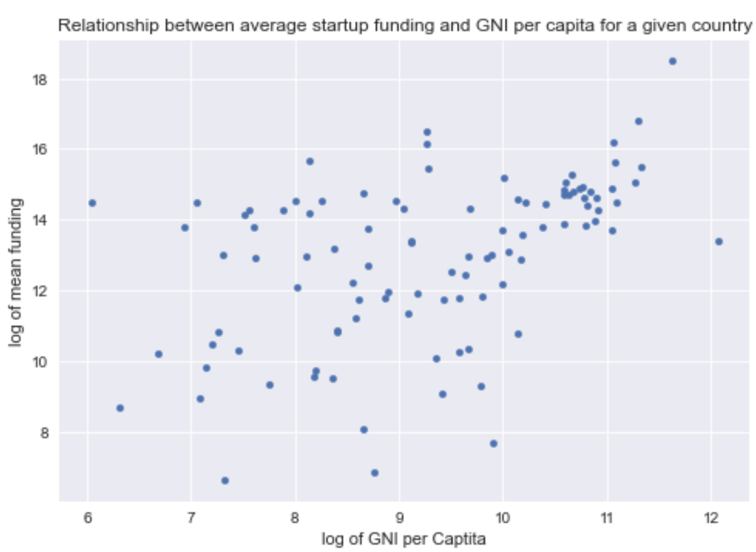


Figure 5, Relationship between average startup funding and GNI per capita in any given country

This regression (Table 1) suggests that a one-percent increase in the GNI per capita of a country would correspond to a 0.8% increase in funding for startups in that country on average, which is economically not very significant. The corresponding P-value rounds up to 0.000, which means that the results are significant at even a 1% level. Furthermore, the large F-statistic and very small suggests that we can reject the null hypothesis that there is no relationship between a country's income level and the average funding for startups in that country. However, according to the r-statistic, a country's income level can only explain 22.4% of the variations in average startup fundings in that country. Visually, the scatterplot shows that the relationship improves (becomes stronger) as income level increases (figure 4). The relationship also fades away most for middle-income economies, which is consistent with our preliminary findings. Therefore, to improve the model, I omit observations for middle income economies next.

OLS Regression Results

Dep. Variable:	log of mean funding	R-squared:	0.267
Model:	OLS	Adj. R-squared:	0.257
Method:	Least Squares	F-statistic:	26.61
Date:	Thu, 07 Apr 2022	Prob (F-statistic):	2.06e-06
Time:	19:28:00	Log-Likelihood:	-154.70
No. Observations:	75	AIC:	313.4
Df Residuals:	73	BIC:	318.0
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.1292	1.388	4.415	0.000	3.362	8.896
log of GNI per Captita	0.7491	0.145	5.159	0.000	0.460	1.039

Omnibus:	5.263	Durbin-Watson:	1.736
Prob(Omnibus):	0.072	Jarque-Bera (JB):	4.430
Skew:	-0.536	Prob(JB):	0.109
Kurtosis:	3.518	Cond. No.	60.2

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 1, OLS regression results for regression 1.1: GNI per capita on mean funding for all countries

1.2 GNI per Capita on Mean Funding: omitting middle-income countries

$$\log(\text{avg funding in a high- or low-income country}) = \beta_0 + \beta_1 \log(\text{GNI per capita}) + \epsilon$$

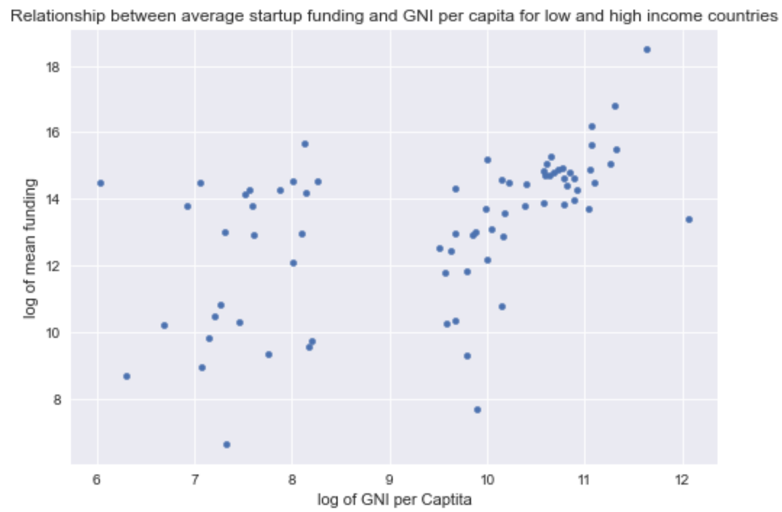


Figure 6, Relationship between average startup funding and GNI per capita in a high- or low-income country

Omitting observations for middle-income economies improves the model. AIC and BIC both decrease, and R-squared improves slightly. The slope of the OLS curve remains mostly similar, and the results are still significant at a 1% level (table 2). It still appears that the model can be further improved, if only applied to high-income economies, as visually, the scatterplot shows a much stronger relationship for high income countries, comparing to low-income ones (figure 6).

OLS Regression Results

Dep. Variable:	log of mean funding	R-squared:	0.267
Model:	OLS	Adj. R-squared:	0.257
Method:	Least Squares	F-statistic:	26.61
Date:	Thu, 07 Apr 2022	Prob (F-statistic):	2.06e-06
Time:	19:28:00	Log-Likelihood:	-154.70
No. Observations:	75	AIC:	313.4
Df Residuals:	73	BIC:	318.0
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.1292	1.388	4.415	0.000	3.362	8.896
log of GNI per Captita	0.7491	0.145	5.159	0.000	0.460	1.039

Omnibus:	5.263	Durbin-Watson:	1.736
Prob(Omnibus):	0.072	Jarque-Bera (JB):	4.430
Skew:	-0.536	Prob(JB):	0.109
Kurtosis:	3.518	Cond. No.	60.2

Table 2, OLS regression results for regression 1.2: GNI per capita on mean funding for high or low-income country

1.3 GNI per Capita on Mean Funding: only high-income countries

$$\log\,(\textit{average startup funding in a high-income country}) = \beta_0 + \beta_1 \log\,(\textit{GNI per capita}) + \epsilon$$

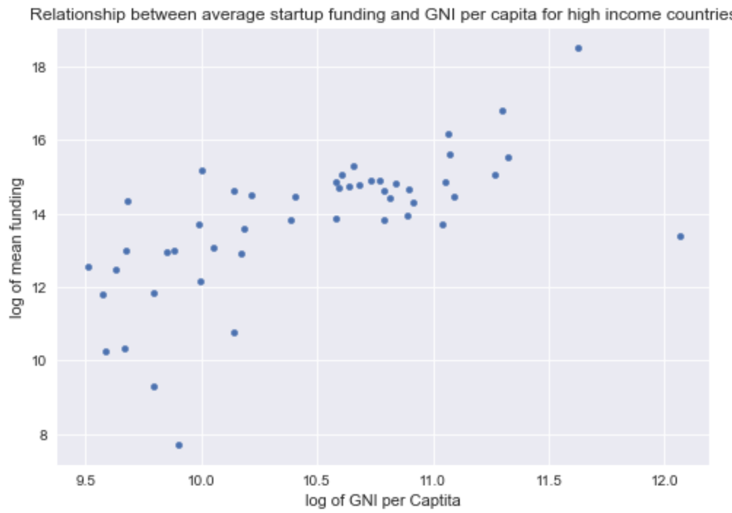


Figure 7, Relationship between average startup funding and GNI per capita in a high-income country

When applied to high income economies only, the model improves significantly. The AIC and BIC decrease further, and according to the R-statistic, a high-income country's income level can explain 45.5% of the variations in average startup fundings in that country (table 3). Furthermore, the slope of the OLS line steepens such that a one-percent increase in the GNI per capita of a high-income country would correspond to a 2.1% increase in funding for startups in that country on average. This finding is much more economically significant, as it highlights the importance of taking a country's income levels into consideration when establishing a startup there. The results are still significant even at a 1% level.

It is important to note that due to problems of endogeneity and confounding variables, this simple regression cannot confirm causality. That is, while there is a relationship between a country's income and its startup funding size, the direction of the effect cannot be determined. Also, higher income levels for a country correspond to many different factors, such as more stability in the economy, higher education levels and less volatility of markets, all of which could theoretically play a role in increasing mean startup funding levels in that country.

OLS Regression Results

Dep. Variable:	log of mean funding	R-squared:	0.455
Model:	OLS	Adj. R-squared:	0.443
Method:	Least Squares	F-statistic:	39.25
Date:	Thu, 07 Apr 2022	Prob (F-statistic):	1.07e-07
Time:	19:28:06	Log-Likelihood:	-85.281
No. Observations:	49	AIC:	174.6
Df Residuals:	47	BIC:	178.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-8.2981	3.533	-2.349	0.023	-15.405	-1.191
log of GNI per Captita	2.1108	0.337	6.265	0.000	1.433	2.789

Omnibus:	20.394	Durbin-Watson:	2.331
Prob(Omnibus):	0.000	Jarque-Bera (JB):	29.941
Skew:	-1.366	Prob(JB):	3.15e-07
Kurtosis:	5.684	Cond. No.	186.

Table 3, OLS regression results for regression 1.3: GNI per capita on mean funding for a high-income country

OLS regression set 2: sector-wide income and average fundings

Next, I study the same relationship on a sectoral level. That is, how does average funding level for startups in a given sector (industrial or service) in a given country change with the GDP of that sector.

2.1.1 Sectoral GDP on Average Funding: Industrial sector

$$\text{log (avg funding for an industrial startup)} = \beta_0 + \beta_1 \text{ log (industrial sector's GDP)} + \epsilon$$

The findings suggest that a one-percent increase in the GDP of the industrial sector of a country would correspond to a 0.158% increase in funding for startups in that country's industrial sector on average. This is a very small change that is of low economic significance (table 4). Furthermore, the results are not significant at any conventional levels, and according to the r-squared, only 2.5% of the variations in average funding for startups in a country's industrial sector can be explained by the GDP of the industrial sector of that country. Finally, the F-statistic is smaller than the conventional levels in economics. Given the limitations of this model, I will break down the independent variable into low, middle and high income to investigate if any of these segments individually display a significant relationship (figure 9).

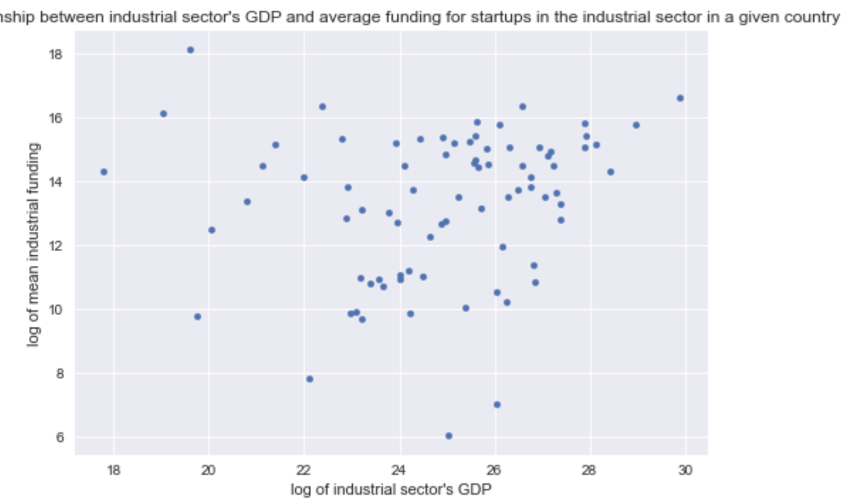


Figure 8, Relationship between industrial sector's GDP and average funding for startups in the industrial sector in any country

```
                                     OLS Regression Results
==========================================================================================
Dep. Variable:     log of mean industrial funding   R-squared:                       0.025
Model:                                        OLS   Adj. R-squared:                  0.013
Method:                             Least Squares   F-statistic:                     2.034
Date:                            Thu, 07 Apr 2022   Prob (F-statistic):              0.158
Time:                                    19:28:12   Log-Likelihood:                -180.03
No. Observations:                              80   AIC:                             364.1
Df Residuals:                                  78   BIC:                             368.8
Df Model:                                       1
Covariance Type:                        nonrobust
==============================================================================================
=====
                                    coef    std err          t      P>|t|      [0.025
0.975]
----------------------------------------------------------------------------------------------
-----
const                             9.3851      2.762      3.398      0.001       3.886      1
4.884
log of industrial sector's GDP    0.1577      0.111      1.426      0.158      -0.062
0.378
==============================================================================
Omnibus:                        8.476   Durbin-Watson:                   2.391
Prob(Omnibus):                  0.014   Jarque-Bera (JB):                8.007
Skew:                          -0.702   Prob(JB):                       0.0182
Kurtosis:                       3.657   Cond. No.                         266.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 4, OLS regression results for regression 2.1.1: Sectoral GDP on Average Funding: Industrial sector

2.1.2 Sectoral GDP on Average Funding: Industrial sector-, low-, middle- and high-income countries individually analyzed

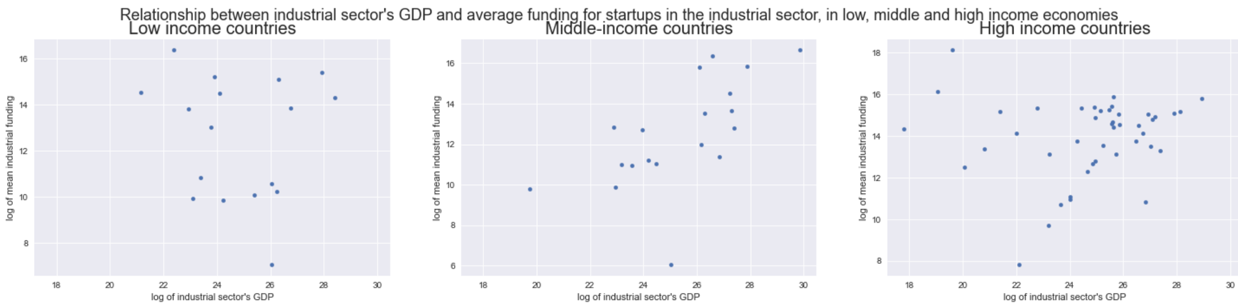


Figure 9, Relationship between industrial sector's GDP and mean funding for startups in the industrial sector. Panel A (left): low-income countries, Panel B (middle): middle-income countries, Panel C (right): high-income countries

The scatterplots in figure 9 display the relationship between the average funding for startups in a country's industrial sector and the industrial sector's GDP in that country, for low-, middle- and high-income economies. Visually, the middle panel shows the strongest relationship between these variables. I ran regressions for each income level separately. The results, as anticipated, showed the strongest and most significant relationship for middle income economies, but the relationship does not hold for low- and high-income economies, and the results are insignificant. I therefore only display the regressions for middle income economies:

$$\log\ (\text{avg funding for industrial startup in a middle-income country}) = \beta_0 + \beta_1 \log\ (\text{industrial sector's GDP}) + \epsilon$$

The findings in table 5 suggest that in middle-income economies, a one-percent increase in the GDP of the industrial sector of the country would correspond to a 0.71% increase in funding for startups in that country's industrial sector on average. While this small change may be of low economic importance, it is statistically significant at a 1% level. Next, over 40% of the variations in average funding for startups in a middle-income country's industrial sector can be explained by the GDP of the industrial sector of that country. Finally, the AIC and BIC decrease significantly when using the model only for middle-income economies, reflecting improvement to the model.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     log of mean industrial funding   R-squared:                       0.404
Model:                                        OLS   Adj. R-squared:                  0.369
Method:                             Least Squares   F-statistic:                     11.53
Date:                            Thu, 07 Apr 2022   Prob (F-statistic):            0.00344
Time:                                    19:28:19   Log-Likelihood:                -40.102
No. Observations:                              19   AIC:                             84.20
Df Residuals:                                  17   BIC:                             86.09
Df Model:                                       1
Covariance Type:                        nonrobust
====================================================================================================
                                       coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
const                               -5.7321      5.396     -1.062      0.303     -17.116       5.652
log of industrial sector's GDP       0.7198      0.212      3.396      0.003       0.273       1.167
==============================================================================
Omnibus:                       12.514   Durbin-Watson:                   1.792
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               10.868
Skew:                          -1.290   Prob(JB):                      0.00436
Kurtosis:                       5.659   Cond. No.                         284.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 5, OLS regression results for regression 2.1.2: Sectoral GDP on Average Funding: Industrial sector in middle-income countries

2.2.1 Sectoral GDP on Average Funding: Service sector

log (avg funding for a service-sector startup) = β0 + β1 log (service sector's GDP) + ϵ

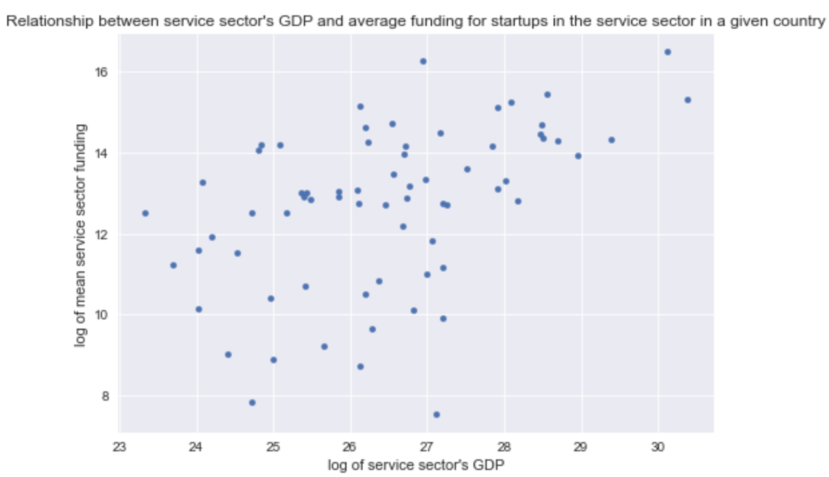


Figure 10, Relationship between service sector’s GDP and average funding for startups in service sector, worldwide

The regression results in table 6 suggest that there is a relationship between the average funding that startups in a country's service sector receive and the GDP generated by the service sector in that country, and that the relationship is significant at a 1% level. More specifically, the model suggests that for every 1% increase in a country's service sector GDP, startups in the service sector in that country receive 0.62% more funding on average. Furthermore, changes in a country's service sector's GDP explain 22.8% of the variances of average funding for service sector startups in that country. Recalling that preliminary evidence suggested different funding patterns for economies of different GDP levels, the model could possibly be improved by running separate regressions for all income groups. This will be the focus of the next regression.

```
                                 OLS Regression Results
=============================================================================================
Dep. Variable:     log of mean service sector funding   R-squared:                    0.228
Model:                                            OLS   Adj. R-squared:               0.216
Method:                                 Least Squares   F-statistic:                  19.47
Date:                                Thu, 07 Apr 2022   Prob (F-statistic):        3.87e-05
Time:                                        19:28:30   Log-Likelihood:             -134.84
No. Observations:                                  68   AIC:                          273.7
Df Residuals:                                      66   BIC:                          278.1
Df Model:                                           1
Covariance Type:                            nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                         -3.7744      3.735     -1.011      0.316     -11.232       3.683
log of service sector's GDP    0.6216      0.141      4.412      0.000       0.340       0.903
==============================================================================
Omnibus:                        8.234   Durbin-Watson:                   1.626
Prob(Omnibus):                  0.016   Jarque-Bera (JB):                7.764
Skew:                          -0.799   Prob(JB):                       0.0206
Kurtosis:                       3.432   Cond. No.                         458.
==============================================================================
```

Table 6, OLS regression results for regression 2.2.1: Sectoral GDP on Average Funding: Service sector

2.2.2 Sectoral GDP on Average Funding: Industrial sector-, low-, middle- and high-income countries individually analyzed

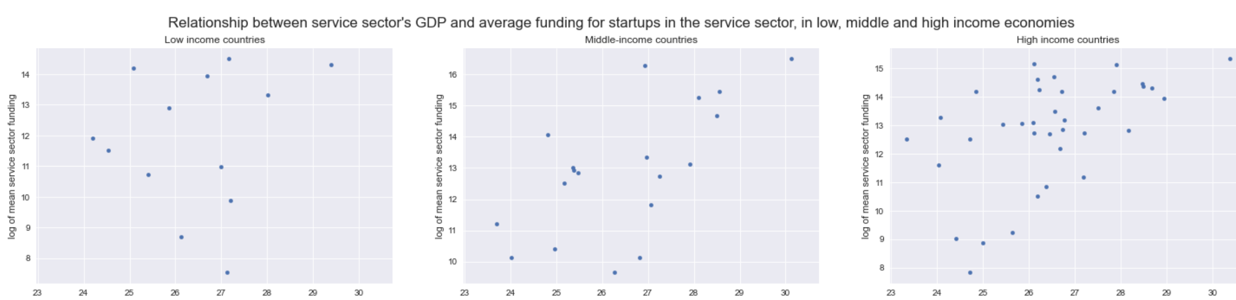


Figure 11, Relationship between service sector's GDP and mean funding for startups in the service sector. Panel A (left): low-income countries, Panel B (middle): middle-income countries, Panel C (right): high-income countries

The panels in figure 11 show how the relationship between the average funding that startups in a country's service sector receive and the GDP generated by the service sector in that country differs in low-, middle- and high-income economies. The scatterplots, as well as individual OLS regressions show once again that the relationship is strongest and most significant for middle income economies, and fades away in low- and high-income economies. I therefore only present the regression output for middle income countries:

$$\textit{log (avg funding for a service-sector startup in a middle-income country)} = \beta_0 + \beta_1 \textit{ log (service sector's GDP)} + \epsilon$$

```
                                OLS Regression Results
==============================================================================================
Dep. Variable:     log of mean service sector funding   R-squared:                       0.409
Model:                                            OLS   Adj. R-squared:                  0.375
Method:                                 Least Squares   F-statistic:                     11.79
Date:                                Thu, 07 Apr 2022   Prob (F-statistic):            0.00317
Time:                                        19:28:51   Log-Likelihood:                -35.289
No. Observations:                                  19   AIC:                             74.58
Df Residuals:                                      17   BIC:                             76.47
Df Model:                                           1
Covariance Type:                            nonrobust
===============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
const                        -7.7641      6.045     -1.284      0.216     -20.518       4.990
log of service sector's GDP   0.7819      0.228      3.433      0.003       0.301       1.262
==============================================================================
Omnibus:                       0.720   Durbin-Watson:                   2.161
Prob(Omnibus):                 0.698   Jarque-Bera (JB):                0.363
Skew:                         -0.331   Prob(JB):                        0.834
Kurtosis:                      2.858   Cond. No.                         427.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 7, OLS regression results for regression 2.2.2: Sectoral GDP on Average Funding: Service sector in middle-income countries

The OLS regression (table 7) suggests that in a middle-income economy, more than 40% of the variations in the mean funding for startups in the service sector can be explained by variations in that country's service sector's GDP. More specifically, a 1% increase in a middle-income country's service sector GDP corresponds to 0.78% more funding for startups in the service sector in that country on average. These results are significant at a 1% level. Finally, AIC and BIC fall comparing to the original model that combined countries of all income groups in one regression, suggesting that service sector's GDP is a better explanator for average service-sector startup funding for middle-income economies.

Discussion of results

From the constructed regression models, the strongest one is model 1.3, which shows how the average funding of startups in a high-income country change with the GNI per capita of that country. This model predicts an economically significant increase of 2.1% in average funding of startups in a high-income country with a 1% increase in the income level in that country, as denoted by GNI per capita. Furthermore, changes in GNI per capita explain 45% of variations in average funding for startups, and the results are statistically significant at a 1% level. This model has one of the lowest AIC and BIC scores (suggesting parsimoniousness in the model) and the highest F-Statistic amongst all other models. This is also a strong model from an economic point of view. Firstly, most startups are clustered in high income countries, and venture capital activity is also highest in those countries. Therefore, a statistically strong model that explains the funding for startups in high income countries specifically is still of high value to many entrepreneurs. Secondly, the model is generalizable to all categories and industries within a high-income country.

For middle income countries specifically, models 2.1.2 and 2.2.2 explain average startup fundings best. It appears that for middle income economies, average funding for startups do not correspond well to the income level in that country, but they change with changes in the income level of the specific sector they are in within that country - namely the service and industrial sectors. This may be explained by the development of specific and distinct networks and corporate relationships in the industrial and service sectors in middle-income economies, as opposed to high income economies. Middle-income economies in East Asia, like China and Malaysia, have historically been hosts to offshore-outsourced functions which were predominantly industrial, such as hardware, web, search and ecommerce. As a result of this historical difference in the development of the industrial sector and its infrastructure in middle income countries, it is therefore likely that the networks in these industries have grown in separate directions, and not integrated much. It is therefore more intuitive to analyze changes in average startup fundings against measures of sectoral income, rather than national income levels.

The mentioned models suggest that in middle income countries, for both the industrial and service sector, changes in that sector's income (as measured by GDP of that sector) explain more than 40% of the variations in the average startup funding levels within that sector, and the results are statistically significant at a 1% level. Given the systematically high average startup fundings in middle income countries, providing models that specifically explain funding patterns in middle income economies may also be of great value to many entrepreneurs.

Finally, no model could explain the funding level for startups in low-income economies. This may be as a result of infrequent startup establishment in those countries, and lack of venture fund activity in low-income countries. In fact, the average venture fund for startups in many low-income countries was 0 - indicating no venture capital activity in that country. Hence, even if a model did explain variations in venture capital funding in low-income economies, it would likely be of small use to entrepreneurs, as low-income countries have been unpopular locations for startup establishments, and venture capital firms operate much less in those regions. Future research could look into developing models and factors that explain of funding in low-income countries.

To summarise these findings in the context of my hypotheses, firstly, the hypothesis that countries with higher income levels have higher startup funding levels on average (H1) only holds in high income countries, and the relationship fades away in middle- and low-income countries. Next, the hypothesis that higher sectoral income in a country corresponds to higher funding for startups in that sector and country (H2) only explains the funding patterns in middle-income economies. This is perhaps due to the networks and relationships that have developed disjointedly in each sector. That is, given in East Asian countries, which largely overlap with middle-income economies, venture capital funding decisions are linked to the relationships and rapports between investors and entrepreneurs, there may be a stronger relationship between startup fundings and income levels when looking at more specific divisions of the economy, rather than a country as a whole. These results, coupled with the fact that none of these variables could explain funding levels in low-income countries, provide support for the last hypothesis (H3), which suggested that the patterns and factors that explain startup funding are not uniform across countries of different income levels. There are fundamentally different funding patterns in middle- and high-income countries, and venture capital activity and startup culture seem to be too weak in low-income countries for a significant relationship to be observed between income levels and startup fundings.

Model comparison

As the final step in evaluating constructed model, I develop a regression tree to predict a startup's funding, given the same independent variables used in the OLS regressions, and then compare the results and errors of the two models. For the OLS regressions, as I progressed through my analysis, to improve the accuracy of my model, I chose both the independent variable that I wanted to split (GNI per capita), and the thresholds on which I wanted to split - same thresholds as I used previously to classify countries into low-, middle- and high-income levels based on their GNI per capita. A regression tree can make this decision more optimally:

$$\min_{j,s}\left[\sum_{i:x_{i,j}\leq s,x_i\in R1}(y_i-\hat{y}_{R1})^2+\sum_{i:x_{i,j}>s,x_i\in R2}(y_i-\hat{y}_{R2})^2\right]$$

The regression tree will iteratively choose the independent variables and thresholds that minimize the mean squared error most. That is, given the independent variables in the dataset, the regression tree provides the most accurate model to predict the dependent variable - a startup's funding. The independent variables used in this regression include the startup country's GNI per capita, industrial sector GDP and service sector GDP, as well as a dummy variable indicating whether the startup's category falls under the industrial sector (1) or service sector (0). Finally, I present the dependent variable - startup's funding - in logarithmic scale as there is a large positive skew that needs to be accounted for. Notice that the dependent variable here, unlike in my OLS regressions, is the funding of an individual startup rather than the average funding for startups in a given country or its sector. Furthermore, the data includes only the startups that have a funding greater than 0 (for scaling purposes), and fall under the categories we classified within the industrial sector and service sector.

The parameter used for regularization in this regression will simply be the depth of the tree: by setting a maximum depth number, we ensure that the tree does not grow to include more parameters than it should, thus avoiding over-fitting. It is important to recognise that there will be random errors and many other variables beyond the scope of this study that impact the funding of a particular startup. Through regularization, we ensure that the model does not reduce the mean squared error term too much to explain the variations that are due to chance, and not the actual change in our independent variables. Due to these reasons, and to avoid complexity, we set the maximum depth to 3 layers. Adding more layers to the regression beyond this point does not produce intuitive results that can be generalizable to the largest number of startups, and the MSE does not decrease noticeably.

The most important difference between the developed regression tree (appendix E) and the OLS regression model, is that the tree chooses the service sector's GDP as the top most indicator of a startup's funding. This is in fact quite intuitive, as venture capital firms themselves fall within the service sector, and there is some proportionality between their earnings (both management fees and carried interest) - that is, to make a revenue, a venture capital firm has to invest. If venture capital revenues contribute to a significant enough proportion of the service sector's GDP and this intuition is correct, this represents an endogeneity problem that future work can address.

The regression tree also consistently chooses very large thresholds for GNI per capita and sectoral GDPs at each given depth, and thus, focuses more closely on high income economies, as we did with OLS regressions. The tree however uses thresholds that could not have been determined using OLS regressions. For example, I split my data using thresholds that classify countries into low, middle or high income, but the regression tree discovered specific thresholds that are statistically stronger turning points for explaining differences in startup fundings. While there may be advantages to having more precise thresholds for sub-sectioning the independent variables, as is the case with the regression tree, it is important to consider the interpretability and application benefits of using the standard thresholds used internationally, such as the GNI per capita thresholds we used for separating countries into low, middle and high income.

Furthermore, the tree successfully identifies and ranks the most important independent variables to split on at each internal node. Recall that with OLS regressions, I used a combination of theoretical hypotheses and some trial and error in order to achieve a narrow enough model to explain startup funding levels - I removed middle-income and low-income countries to improve the model developed in OLS regression set 1, and focused on middle-income economies in OLS regression set 2. This means that the OLS method could only provide insight on the relationship between the independent variables (and sub-sections of them) that I had chosen to analyse, but would not reflect relationships that I had not considered before. For example, the tree identified a country's service sector's GDP as the top-most explanator of startup fundings in that country. As discussed above, there may be an economic explanation for this, linking venture capital firm's income, which contributes to the service sector's income, to how much venture capital firms invest in startups. This is particularly important for reducing endogeneity biases if aiming to establish causality in future studies.

Finally, the mean squared errors of the regression models, and the most important OLS regressions are compared (see table 8). The mean squared error of prediction for the regression tree is larger than that of both model 1.3 and model 2.2.2. This suggests that the algorithm, given the 4 independent variables (startup's country's GNI per capita, industrial sector GDP and service sector GDP and the sector of the startup) and the maximum depth permitted (3), cannot determine a startup's funding as accurately as OLS models 1.3 and 2.2.2. Notice again that the OLS models are different from the regression tree in their dependent variables: they use mean startup fundings for their given country and mean sectoral fundings for a given sector in a given country. This should be considered when evaluating the accuracy of those models against the regression tree.

```
======================== MSE Comparisons ========================

Mean Squared Error for Regression Tree model:    4.402686591239705
Mean Squared Error for OLS model 1.3:            3.5631123553125637
Mean Squared Error for OLS model 2.1.2:          7.0656426521199664
Mean Squared Error for OLS model 2.2.2:          4.295354340699505
```

Table 8, Mean Squared Error (MSE) for regression tree model and OLS models 1.3, 2.1.2 and 2.2.2.

Conclusion and future research

This paper aimed to investigate whether the location and the sector (industrial or service) of a startup, separately or in conjugation with one another, correspond to systematic differences in venture capital fundings of startups, and if so, what are these differences. To answer this question, we described both the country and the sector in which a startup falls in terms of the income generated in that sphere. More precisely, GNI per capita was used for describing income levels in countries – which also allowed for the classification of these locations to low-, middle- and high-income countries – and sectoral GDP was used to define the income generated by the sector in which the startup operates.

We discovered that there indeed are differences between average funding levels for startups, based on the income level of the country that they are located in, although this relationship is not consistent across low-, middle- and high-income countries. For startups in high income countries, there is a positive, linear association between income level of the country they are situated in, and the amount of funding they receive. Simply put, for startups in high-income countries, as the income level of their country increases, so do their fundings. While this aligns with macroeconomic theories that explain investment as a function of income, where a rise in income affects investment levels positively, we cannot establish such causal effect at this stage of analysis. For instance, it could indeed be possible that higher investments are contributing to economic growth, and thus, resulting in larger income levels.

For middle-income economies, this exact relationship does not hold. A better way to model the variations in average startup fundings in such economies is by explaining these differences based on the income level of the sector that they are in, rather than the country as a whole. We discovered that for startups in the industrial sector, higher GDP in the industrial sector corresponds to higher funding levels. Similarly, for firms in the service sector, higher GDP in the service sector corresponds to higher funding levels. This may be explained by a combination of historically different developments of sectors as a result of hosting industrial offshore-outsourced functions and the tendency of venture capital firms to base funding decisions on existing rapports and relationships. It is important to consider the potential endogeneity bias when using service sector's GDP for explaining startup's income, as venture capital firms' earnings, which are arguably proportional to their investments, contribute to the service sector's income. Finally, in low-income countries, no significant relationship exists between income levels in a startups country or sector of operation. This may be due to the lower level of venture capital activity and weak startup culture in low-income countries, which are reflected through the fact that many low-income countries have venture capital funding for their startups.

While past literatures had focused heavier on the importance of capturing venture capital funding for early-stage startups, this paper explains how country-wide and sectoral income levels could hinder or increase a startup's venture capital funding amount, which is more valuable to entrepreneurs. Particularly, this paper develops models that give specific insight into funding levels for startup, based on whether they are situated in high- or low-income countries, which have been more popular locations to startups in the past. This study also adds to past literature on the differences between venture capital firms' valuations and investments in the West and East by providing the implications of those differences (Miloud, 2012): we suggest that in middle-income economies, which largely overlap with East Asian hubs for startups, funding levels for startups tend to be more correlated with the income level of their sector that with the overall income levels in their country.

Future studies or replications of this study could address the mentioned endogeneity problems in an effort to establish a causal effect. Furthermore, they could benefit from categorising startups into sectors more carefully and scientifically to ensure than each startup's funding is in fact being compared against the relevant GDP. Alternatively, researchers could collect income data specific to individual industries (rather than sectors) to ensure this consistency. The latter would perhaps yield more accurate results, as networks of investors and entrepreneurs may be stronger in these narrower categories, and generated income levels would be more accurate determinants of funding decisions. Finally, future research could also work towards developing models that explain funding levels in low-income countries. Given the small frequency of startup establishments in low-income countries, and lack of considerable venture capital activity, this may currently be of low significance. However, with such considerable prospect of growth for startups across the world, such results will eventually be of high economic significance to entrepreneurs.

References

Bruton, G., Ahlstrom, D., & Yeh, K. S. (2004*). Understanding venture capital in East Asia: The impact of institutions on the industry today and tomorrow*. Journal of World Business, 39(1), 72-88.

Davila, A., Foster, G., & Gupta, M. (2003). *Venture capital financing and the growth of startup firms. Journal of business venturing*, 18(6), 689-708.

Gage, D. (2012). *The venture capital secret: 3 out of 4 start-ups fail*. Wall Street Journal, 20.

GNI per capita, Atlas method (current US$) | Data. Data.worldbank.org. (2013). Retrieved 15 April 2022, from https://data.worldbank.org/indicator/NY.GNP.PCAP.CD.

Krishna, A., Agrawal, A., & Choudhary, A. (2016, December). *Predicting the outcome of startups: less failure, more success*. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) (pp. 798-805). IEEE.

List of countries by GDP sector composition - Wikipedia. En.wikipedia.org. (2017). Retrieved 15 April 2022, from https://en.wikipedia.org/wiki/List_of_countries_by_GDP_sector_composition.

List of country codes by alpha-2, alpha-3 code (ISO 3166). Iban.com. (2022). Retrieved 16 April 2022, from https://www.iban.com/country-codes.

Miloud, T., Aspelund, A., & Cabrol, M. (2012). *Startup valuation by venture capitalists: an empirical study*. Venture Capital, 14(2-3), 151-174.

Patel, N. (2022). *90% Of Startups Fail: Here's What You Need to Know About The 10%*. Forbes. Retrieved 15 April 2022, from https://www.forbes.com/sites/neilpatel/2015/01/16/90-of-startups-will-fail-heres-what-you-need-to-know-about-the-10/?sh=21bbb5956679.

Startup Investments. Kaggle.com. (2013). Retrieved 15 April 2022, from https://www.kaggle.com/justinas/startup-investments.

World Bank Country and Lending Groups – World Bank Data Help Desk. Datahelpdesk.worldbank.org. (2020). Retrieved 15 April 2022, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519#High_income.

Appendix A

Summary statistics for startup categories, their frequency and funding within the top 15 most popular categories.

A.1: Overall summary statistics

	category_code
count	91632
unique	15
top	software
freq	17922

A.2: Summary statistics for each category

	funding_total_usd							
	count	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
category_code								
advertising	6,098	2,047,915	10,953,108	0	0	0	0	273,834,120
biotech	4,430	15,094,511	49,783,346	0	0	1,960,175	12,893,180	2,400,000,000
cleantech	1,940	19,933,089	72,709,204	0	0	50,000	9,900,196	1,200,000,000
consulting	5,006	522,743	4,629,420	0	0	0	0	100,000,000
ecommerce	9,065	1,944,639	21,724,468	0	0	0	0	1,100,000,000
education	2,901	1,361,089	16,312,303	0	0	0	0	750,000,000
enterprise	4,441	4,806,655	24,464,747	0	0	0	850,000	1,270,283,000
games_video	7,520	1,958,127	20,097,092	0	0	0	0	860,213,000
hardware	2,951	5,070,205	33,362,268	0	0	0	600,000	1,100,000,000
mobile	6,862	4,357,136	85,790,909	0	0	0	39,753	5,700,000,000
network_hosting	2,350	5,257,320	33,772,869	0	0	0	0	1,050,000,000
public_relations	2,846	2,146,457	26,924,669	0	0	0	0	1,055,750,000
search	2,182	1,426,883	13,475,703	0	0	0	0	448,000,000
software	17,922	2,330,561	13,873,792	0	0	0	0	964,999,998
web	15,118	1,251,379	16,400,664	0	0	0	0	1,147,288,416

Appendix B

Categories in each sector:

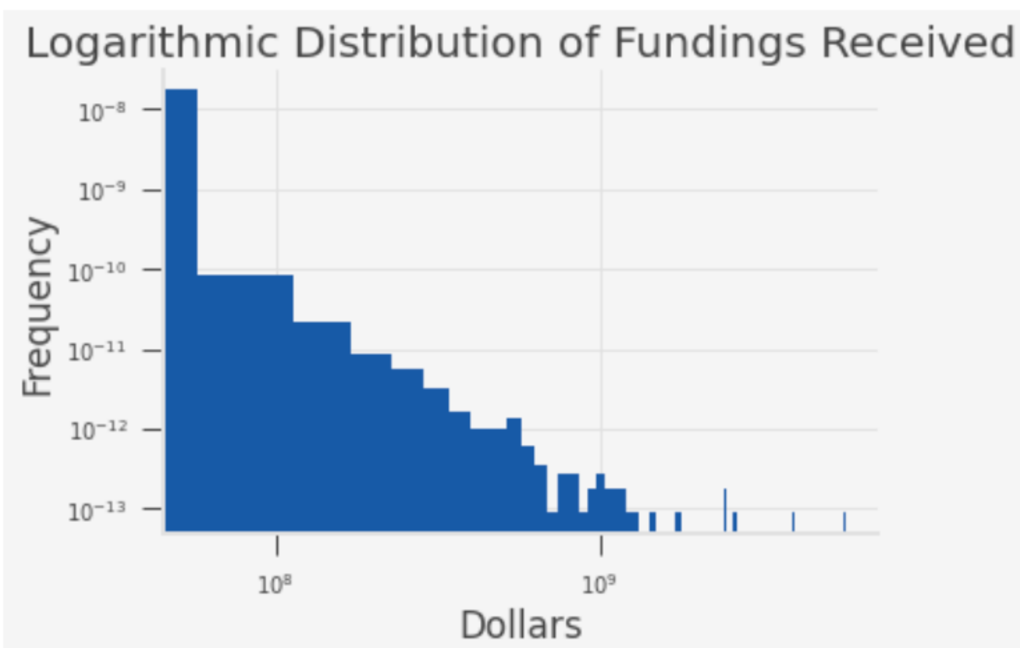
Industrial sector	Service sector
Biotechnology	Advertising
Clean technology	Consulting
Video games	Ecommerce
Hardware	Education
Mobile	Enterprise
Search	Network hosting
Web	Public relations
Software	

Appendix C

C.1: Summary statistics for venture capital funding across all startup categories and countries

	funding_total_usd
count	196,553
mean	2,101,193
std	26,034,848
min	0
25%	0
50%	0
75%	0
max	5,700,000,000

C.2: Logarithmic distribution of fundings received, in USD, by all startups, visualized



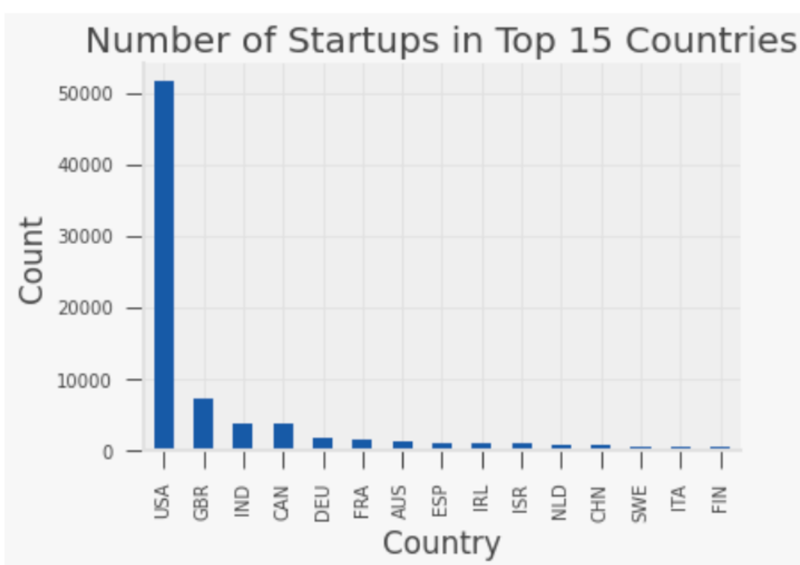
Appendix D

Summary statistics for startup countries, their frequency and funding

D.1. Summary statistics for each of the 15 countries with highest average fundings

	funding_total_usd							
	count	mean	std	min	25%	50%	75%	max
country_code								
BMU	7	107,380,857	200,411,051	0	0	0	112,500,000	526,666,000
LUX	56	19,852,636	92,975,052	0	0	0	131,018	537,779,080
CHN	732	14,603,466	65,146,958	0	0	0	6,305,750	1,100,000,000
CYM	15	10,644,667	21,618,664	0	0	0	8,625,000	65,000,000
MYS	149	10,230,607	62,059,204	0	0	0	0	551,200,000
SWZ	8	6,392,625	14,127,969	0	0	0	2,785,250	40,000,000
USA	51,637	6,014,843	46,964,867	0	0	0	1,000,000	5,700,000,000
CHE	515	5,493,695	30,010,434	0	0	0	125,000	515,000,000
RUS	443	5,178,203	35,318,292	0	0	0	350,000	625,000,000
ISR	1,042	4,299,512	14,447,848	0	0	0	3,000,000	293,000,000
TWN	131	4,181,959	25,220,427	0	0	0	0	238,174,040
CZE	139	3,900,799	24,517,805	0	0	0	0	252,000,000
JPN	470	3,452,392	28,813,110	0	0	0	24,750	540,000,000
NOR	229	3,430,234	19,137,596	0	0	0	500,000	272,120,000
DEU	1,921	3,001,444	17,888,969	0	0	0	0	460,488,000

D.2. Number of startups in 15 countries with highest number of startups, visualized



Appendix E

Regression Tree predicting startup funding based on GNI per capita of its country, GDP of service sector and industrial in its country, and whether the startup is in the industrial sector or service sector