Kickstarter Crowd-funding capstone project

Database building and Visualizations

Abhijeet Mulye

Advisor: Prof. Lee Fleming

Team-mates: Raymond Von Mizener, Kartikeya Mohan Sahai

With additional guidance from Gabe Fierro

ABSTRACT

Our project aimed at creating tools which can help to analyze the exploding phenomenon of crowd-funding, a practice of several small investors pooling their money to fund a venture. We created a MySQL database of venture funding information from Kickstarter (http://www.kickstarter.com), one of the leading crowd-funding platforms. We used Python web scripts to develop the necessary web-scrapers which automatically copied information available in the public domain. As a result of our capstone project, a detailed database has been made available to researchers so that they can query information and study any trends which they desire to investigate. We also created visualizations focusing on the geography of crowd-funding. It was observed that California and New York had outstanding crowd-funding activities in the United States. Also within California, Los Angeles and the San Francisco Bay area dominated the crowd-funding scene in terms of number of ventures, amount of money raised as well as number of donors.

INTRODUCTION

Most of the work on visualizations in crowd-funding space is due to an advisory firm called Massolution (http://www.massolution.com/). The crowd-funding industry report has also indicated that crowd-funding is nearly doubling in volume every year within the United States[1]. Collins, Liam et al. have conducted a similar study in the UK and reported their findings in a paper titled “The rise of future finance: The UK alternative finance benchmarking report”[2].

Furthermore, the JOBS act[3] signed by President Obama in April 2012 contains certain clauses to assist the growth of small businesses typically funded by several small investors. This law has also made investing easier for non-accredited investors. A few clauses of this bill are pending and parliamentary approval is awaited. It is anticipated that such legal changes will further boost the scope and market of crowd-funding in the United States. We focus on geographic distribution of crowd-funding to speculate and answer the question “*Where will the impact of provisions in the JOBS act be most visible?*”

However determining the magnitude of the change is a hard task without the right kind of data. This is one of the motivations behind creating a database of crowdfunding activities. In the paper “How big will be the debt and equity crowd-funding investment market be? Comparisons, assumptions and estimates”[4] Best, Fleming et al. have attempted to provide an estimate. Their estimate has a range from $184M to $3.98B. To arrive at a more specific estimate, more extensive data would be necessary. Academic researchers are not the only people in need of such data.

The White House, World Bank as well as other regulatory bodies are also extremely interested in a database containing detailed information regarding the projects and donors from the viewpoint of understanding this phenomenon and making relevant regulatory changes. The Coleman Fung institute collaborates with several crowd-funding platforms and researchers around the world to make data available for no-profit study. The big commercial platforms such as Kickstarter however are often tight-lipped about their project information.

Kickstarter is one of the prime crowd-funding platforms in the United States. We wanted access to their funding information as soon as possible. We did so by using web-scrapers. By capturing all the public-domain information about every project Kickstarter has ever launched and hosted, we have created an extensive database of projects, both successfully funded and failed.

To back our claim that this database can be used to create visualizations and has the potential to be used for other kinds of more quantitative studies we created maps of crowd-funding activities across the world. Karthikeya Mohan Sahai focused on making interactive global maps in his paper “Kickstarting Crowdfunding”[16] to get a high-level picture of overall crowd-funding activity all over the world. Raymond Von Mizener has applied the techniques of Machine learning in his paper “The crowdfunding database (with respect to machine learning)”[15].

In my paper I have focused on the United States alone and created several graphics with granularity level down to each individual county. It is clearly visible from those graphics that New York, Los Angeles and the San Francisco Bay area dominate the crowd-funding scene in terms of number of ventures, amount of money raised as well as number of donors. We hence speculate that most of the employment opportunities generated by the implementation of the aforementioned JOBS act will be in these regions - if crowdfunding project density is taken to be a proxy for the potential of the region to be a start-up hub.

LITERATURE REVIEW

There are a lot of publications available on crowd-funding on the internet. However few are focused on web-scraping of crowd-funding websites, creating a database of crowd-funding projects or visualizing crowd-funding information particularly by geographical aspects.

Most of the publications (such as the two cited in above section) are concerning growth of crowd-funding in recent years and the impact it will cause on economy.

Some websites/blogs discuss whether crowd-funding will be disruptive for the existing ecosystem of Venture Capitalists and Angel Investors. Most notable of such articles are Ryan Caldbeck's article in Forbes magazine[5] and Lora Kolodny's compilation of interviews in the Wall Street Journal[6]. Ryan downplays the role of crowd-funding with respect to other funding sources concluding that the traditional sources of capital may reduce in scope but will not get replaced by crowdfunding. He writes that crowd-funding should remain complementary to other sources of capital and only when it can provide dividends to investors over the long run it should become a standalone source of capital. On the other hand, many of Lora's interviewees focus only on crowd-funding's potential and strengths.

There also exists some interesting literature regarding how the physical distance between donors and entrepreneurs affects the pledging dynamic.

In a paper related to the spatial proximity of donors[7], Agrawal, Catalini et al conclude that online mechanisms make it easy to invest in early-stage start-ups over long distances. Their conclusion is at odds with established viewpoint that investors heavily favor ventures in their home state or country.

For instance Lin and Viswanathan[8] conclude that investors have been statistically proven to favor projects located physically closest to them in a variety of market situations. According to their paper, the home-state bias was consistent whether the markets were free or regulated in their experimental set-up. The database which we have created currently does not include backer's location. But if extended it can be used to perform analyses similar to the work of above researchers.

METHODS AND MATERIALS

We copied all the information published by Kickstarter in the public domain. We did so by using a technique known as “Web-scraping”. It basically means writing a computer script to imitate a human user surfing the concerned website, to copy all the information from the web host to your local server. Script was written in a programming language known as Python. It was chosen because of availability of open-source libraries dedicated to the purpose of web-scraping such as Beautiful Soup[9].

Fig 1: An Overview of the Scraping process

First step was to obtain a list of individual project URLs.

- Kickstarter's projects page is structured in such a way that more you scroll down, more project descriptions are loaded. This kind of web-page is commonly referred to as an “infinite-scrolling structure”. We were able to use Selenium[10], a package that emulates a web browser, to scrape a list of URLs from this page. Since most web servers lack a physical interface on which browsers such as Mozilla Firefox can render and display websites, we used a virtual display called xvfb[11]. The python code which runs Selenium and generates a list of project URLs to be scraped was written by me.

Second step was to scrape all information about the projects, one URL at a time.

- Aforementioned Beautiful Soup library was helpful in isolating particular section-headers of interest from the project's individual page. Raymond Von Mizener and Karthikeya Mohan Sahai worked on this step. Raymond Von Mizener wrote a python script[15] to convert the JSON output of his web-scraper into a de-serialized format which is easier to use for building a database. He also implemented a parallel method to use the Kickstarter API and get around the use of Selenium with assistance from Gabe Fierro.

Third step was cleaning and parsing the data.

- We used MySQL server hosted at UC Berkeley to achieve this. I also created a database with a certain "schema" to facilitate loading and viewing of the data. [Description of schema is provided on the next page]

Our last step was to carry out visual analysis and create geographical distributions to demonstrate the power of this database. This was done purely for illustrative purposes. The analyses were also carried out in the Python language using libraries such as PANDAS[12]. The SVG file consisting of county boundaries and style description was adopted from Wikimedia Commons[13]. A tutorial[14] written by Nathan Yau on Flowingdata.com was helpful for me to create the graphs. Kartikeya Sahai focused on the gaming industry in his paper "Kickstarting Crowdfunding"[16].

Kickstarter Schema Description v0.1

Access Information

host: rosencrantz.berkeley.edu

user: public

password: <Available on request>

port: 3306

database: kickstarter

Notes:

1. Most text information is in Unicode format (For example, u'Text' instead of Text).

2. All dates such as project creation date, project deadline are Python epoch seconds.

3. Database contains some redundant fields which are Kickstarter's internal variables captured during parsing process.

Following tables contain information about 105,598 Kickstarter projects (hosted online as of Feb'14). Most of the numerical information is in project_data table.

1. Project_data –

#	Column	MySQL Data Type	Commentary
1	disable_communication	varchar(50)	Internal Kickstarter variable – We find it to be associated with "Suspended" projects
2	Currency	varchar(50)	Currency of donation
3	Deadline	Decimal(10,0)	Deadline for funding
4	Currency_trailing_code	varchar(50)	Internal Kickstarter variable
5	project_id	Decimal(10,0)	Unique project identifier

6	state_changed_at	Decimal(10,0)	Date of last status change
7	Goal	Decimal(10,0)	Amount requested by creator
8	Pledged	Decimal(10,0)	Amount pledged by donors
9	State	Varchar(50)	Successful/Failed etc
10	launched_at	Decimal(10,0)	Start date of campaign
11	Blurb	varchar(500)	Detailed project information
12	backers_count	Decimal(10,0)	Number of backers
13	Slug	varchar(200)	Project description
14	Name	varchar(200)	Project name (Similar to above)
15	Country	varchar(50)	
16	created_at	Decimal(10,0)	Creation date of project
17	currency_symbol	varchar(5)	

2. Project_location -

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	Name	varchar(100)	Name of place
3	short_name	varchar(100)	Deadline for funding
4	Country	varchar(100)	
5	Slug	varchar(100)	Place-state combined
6	State	varchar(100)	
7	displayable_name	varchar(200)	Same as Slug
8	project_id3	Decimal(10,0)	Internal Kickstarter Id

3. Project_photo – Links to various images

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	Med	varchar(200)	
3	Little	varchar(200)	
4	Full	varchar(200)	
5	Thumb	varchar(200)	
6	Ed	varchar(200)	
7	Small	varchar(200)	

4. project_url –

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	url	varchar(200)	

5. Category –

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	Position1	Decimal(10,0)	Internal Kickstarter Variable
3	Name	varchar(500)	Music/Technology/Art etc
4	position2	Decimal(10,0)	Internal Kickstarter Variable

6. Category_urls –

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	Discover	varchar(500)	URL to the project category webpage

7. Creator –

#	Column	MySQL Data Type	Commentary
1	project_id	Decimal(10,0)	
2	Name	varchar(100)	Name of project creator
3	Project_id2	Decimal(10,0)	Internal Kickstarter Variable

8. Creator_avatar – Pictures of project creator

#	Column	MySQL Data Type	Commentary
1	creator_id	Decimal(10,0)	Internal – Related to URL of creator's profile/project page
2	small	varchar(500)	
3	medium	varchar(500)	
4	thumb	varchar(500)	

9. creator_profile

#	Column	MySQL Data Type	Commentary
1	creator_id	Decimal(10,0)	Internal – Related to URL of creator's profile/project page
2	Url	varchar(100)	

10. creator_urls

#	Column	MySQL Data Type	Commentary
1	creator_id	Decimal(10,0)	Internal - Related to URL of creator's profile/project page
2	Url	varchar(100)	Related to above

11. location_urls

#	Column	MySQL Data Type	Commentary
1	Id1	Decimal(10,0)	Internal project code
2	nearby_projects	varchar(200)	Kickstarter API URLs - Access is forbidden

DISCUSSIONS

I. Geographical visualizations

We plotted the crowd-funding activity in the United States at county level. All heat-maps have 6 scales of activities – Darkest corresponding to the highest. There are 3 basic types of geographical info-graphics: 1. Number of projects, 2. Amount of money pledged to those projects, and 3. Number of backers.

1. Number of Projects

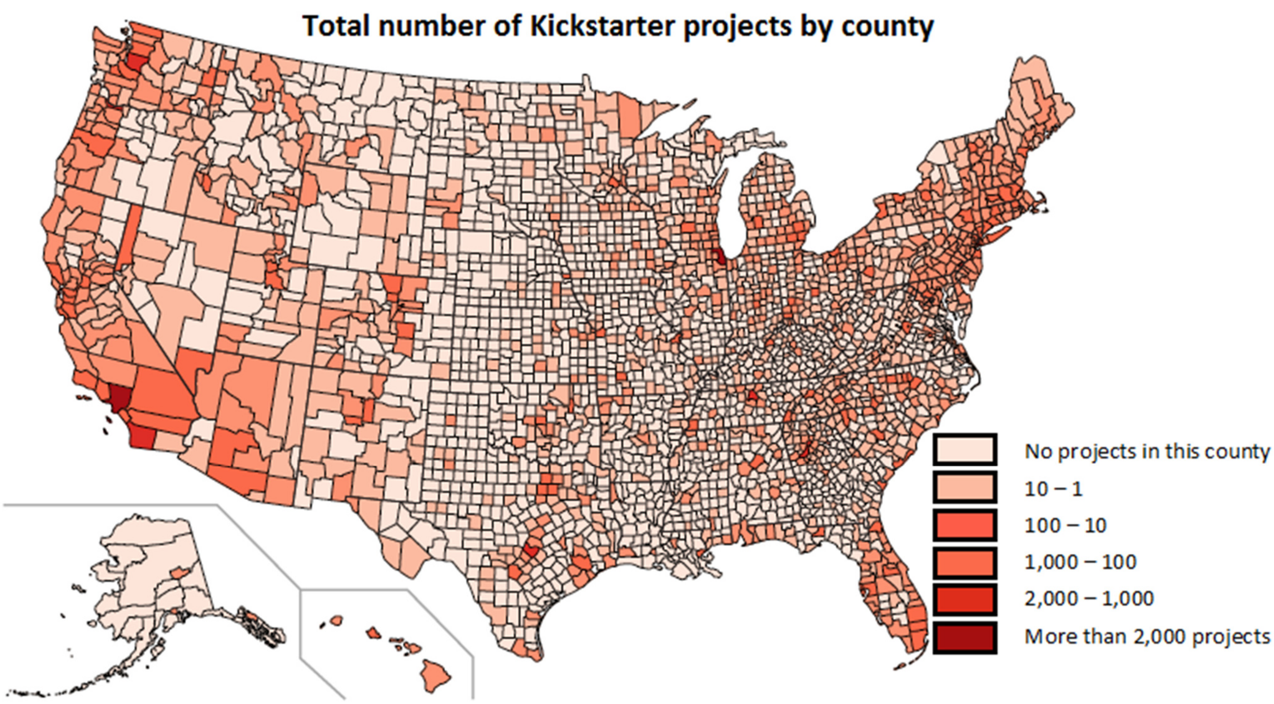

Fig 2. Heat-map showing number of projects in each county

In general the west and east coast seems to have higher crowd-funding activity, in addition to Florida, Washington, and the Boston Massachusetts region.

The results correlate well with the well-known fact that in the United States many of the most reputed institutes of technology/engineering (and universities known for STEM education in general) are concentrated in those areas, leading to a concentration of entrepreneurs and a supporting ecosystem for launching start-ups. Number of projects also correlate well with the population density figures.

2. <u>Total amount pledged</u>

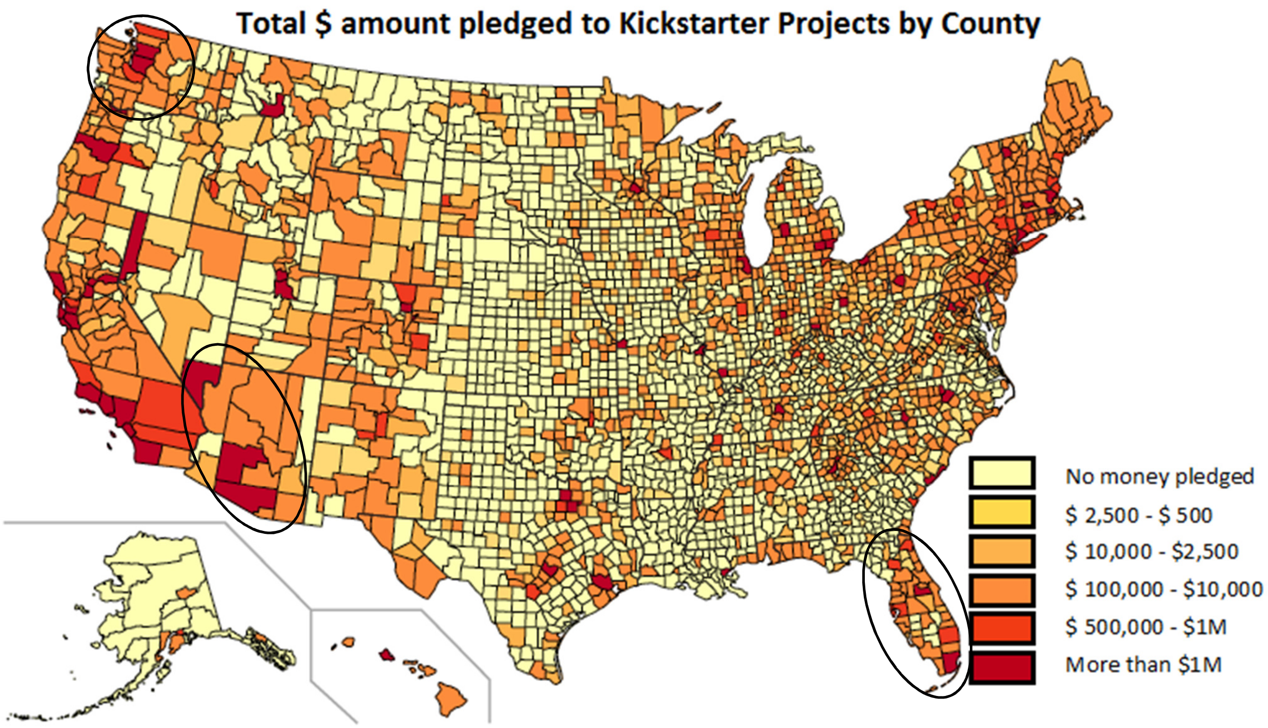

Fig 3. Heat-map showing total funding raised by Kickstarter campaigns

It is interesting to note that other than the East and West Coast, certain regions in Washington and Colorado have also done well to raise money. In terms of total money raised by crowdfunding projects Florida looks almost like California. We speculate that Florida's climate and tourism culture makes it a good destination for launching projects about music or art.

Crowdfunding projects launched in Arizona also seem to do pretty well in terms of raising capital. We speculate that low business and property tax rates makes Arizona an attractive location for starting businesses. Additionally states such as Arizona, Colorado and Washington must also benefit from the fact that workers over there demand lower salaries than their counterparts in California or New York.

3. Total number of backers

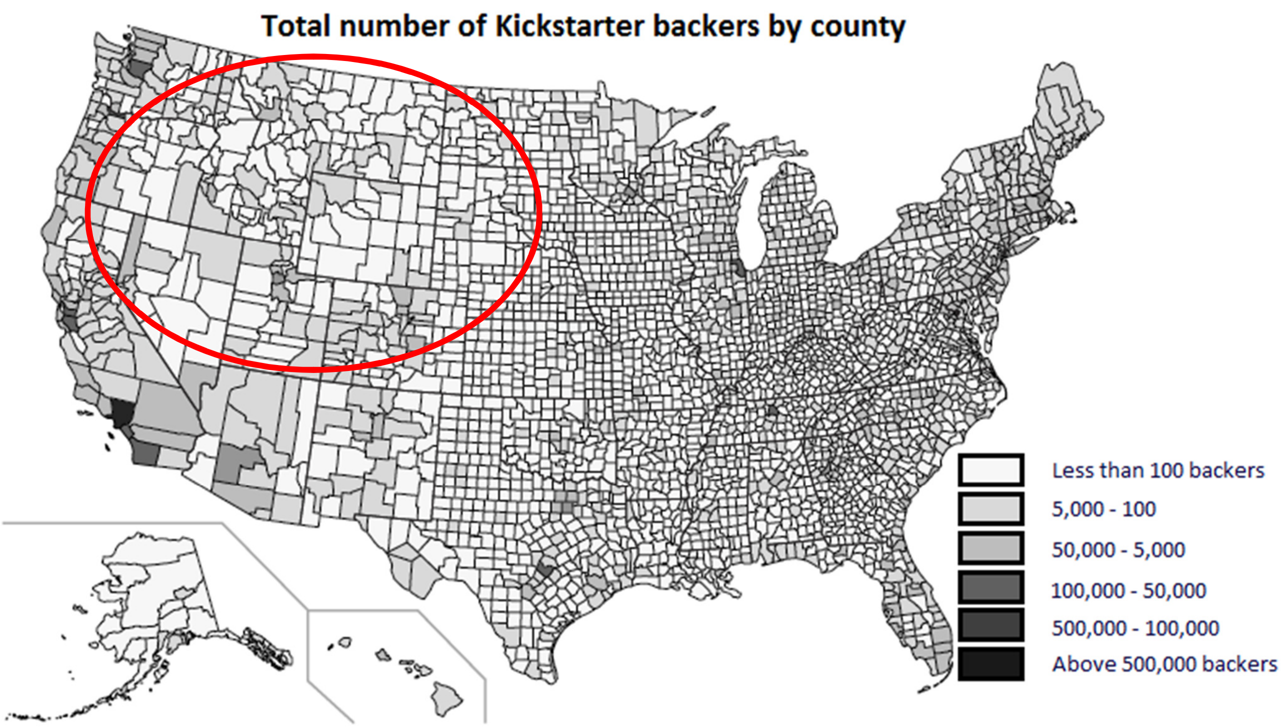

Fig 4. Heat-map showing total number of project backers

From the backer count info-graphic it is visible that fewer people in the Mountain region (specifically Montana/Wyoming) tend to contribute to crowdfunding projects. This can be ascribed to the relative lack of high speed broadband connectivity in those states.[18.] High-speed broadband penetration is a basic necessity for crowdfunding to reach masses.

- As a future extension of this study, it will be interesting to look at cross-county donation paths to verify whether people are more likely to back projects in their home county or whether it is the other way around. To answer the question of whether the home bias works at county level, state level or is it strongest at the country level? (if it works in the online crowdfunding space at all).

- Please see appendix section for the detailed rules used to create the visualizations.

- Such graphics are good to get a sense of the overall picture. However the visual inferences drawn by looking at them need to be backed by a more solid quantitative study comprised of t-tests etc.

- One of the issues with heat-maps is that they tend to correlate with population (i.e. some of the information gets hidden in sparsely populated areas and counts get overstated in densely populated areas). We also adjusted these charts accounting for these effects.

To get a better sense of the “density” of activity we should be looking at number of projects per population of the county to account for how densely populated the county is. To account for the size of county we can look at number of project per unit sq. mile land area of the county. In order to calculate the “per unit population” and “per unit land area” figures we used the United States Census Website[17]

Project count per unit population:

Project count per unit land area:

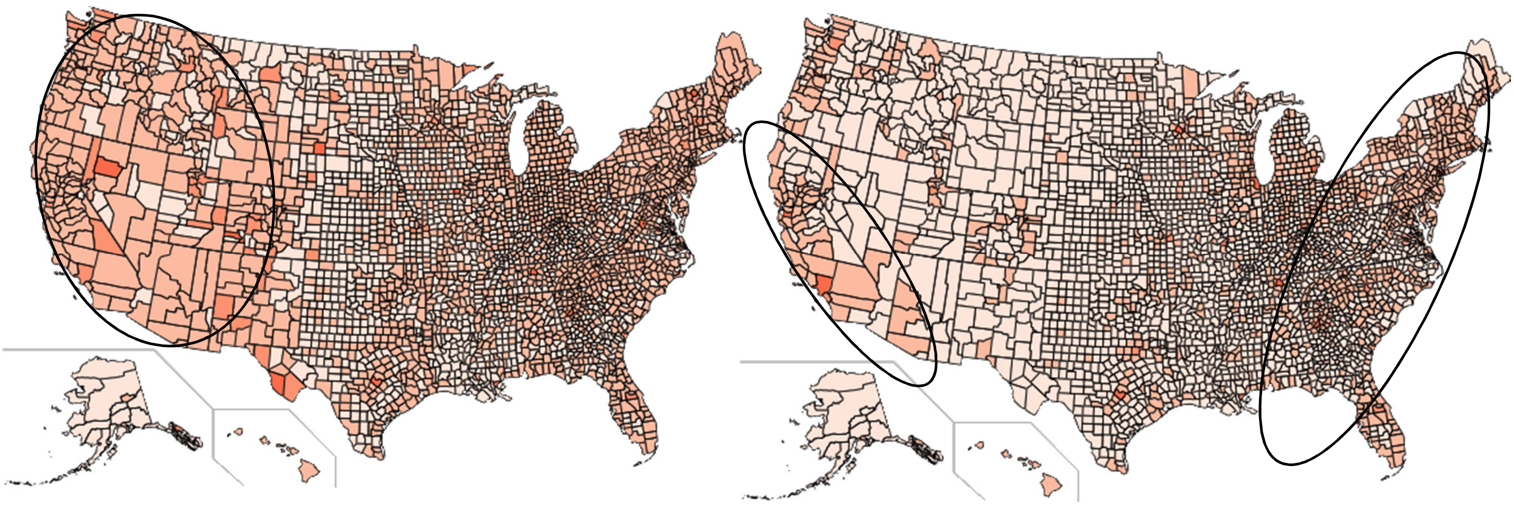

Fig 5. Project count heat-maps adjusted for population/area

Six shades denote 6 roughly equal sextiles. We see that the West and East regions have denser crowdfunding activity in general.

Amount raised per unit population: Amount raised per unit land area:

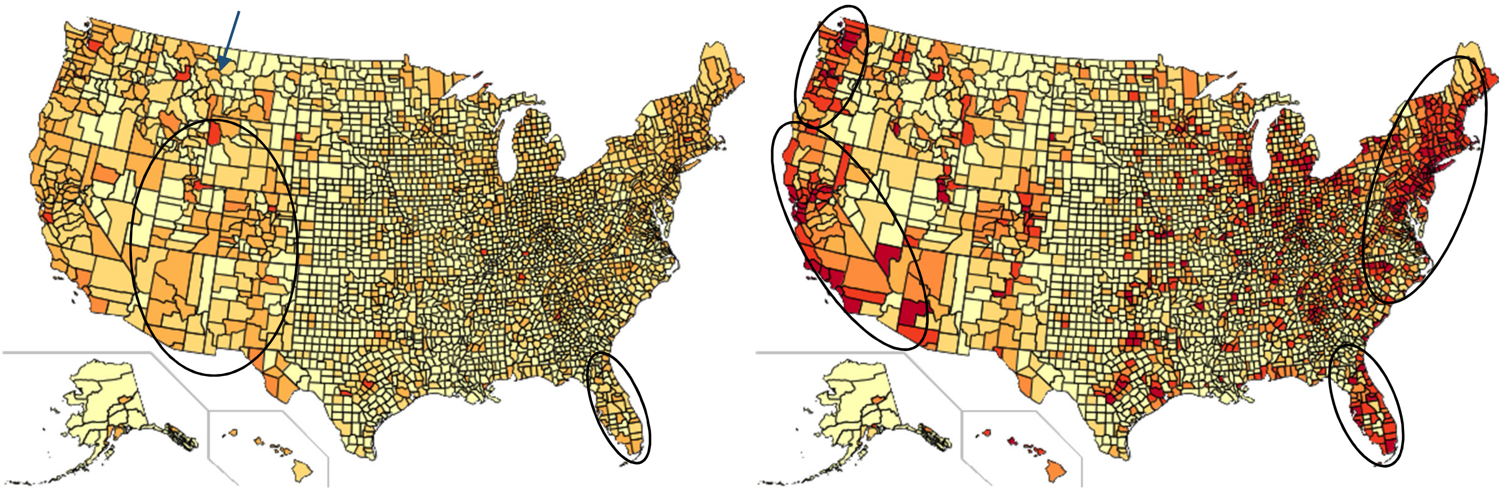

Fig 6. Project funding heat-maps adjusted for population/area

Dividing the counties in six roughly equal sextiles, we notice from the left chart that many states such as Arizona, Florida, Kansas, Texas etc. are (somewhat surprisingly) not far behind the Golden State of California as far as raising money on Kickstarter is concerned. We speculate that while California start-ups get all the media attention, a lot of ventures are being kick-started in those states as well.

Chart on the right perhaps best points out how coastal regions raise more funding per square mile than inner regions.

Backers per unit population: Backers per unit land area:

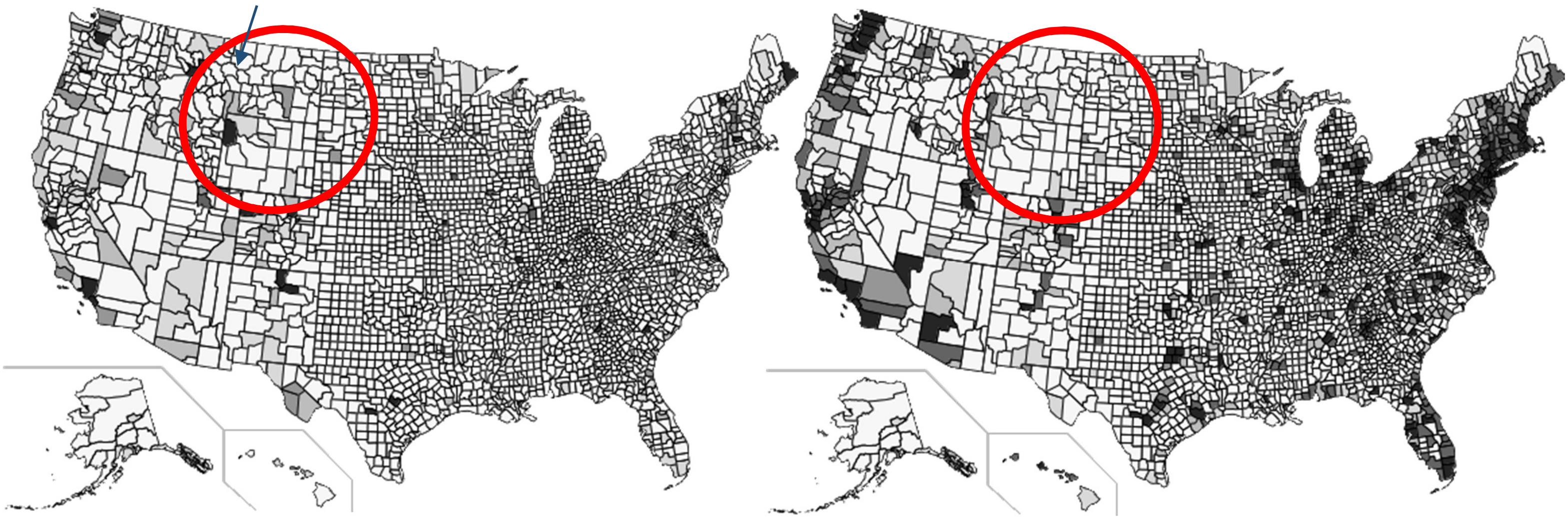

Fig 7. Backer count heat-maps adjusted for population/area

Both of the backer count density charts again divide counties in six roughly equal sextiles and point out the relative lack of crowdfunding backers in the mountain region.

Comparing Fig.6 and Fig.7 we conclude that in certain regions a few backers are donating large amounts to crowdfunding projects (as opposed to a large number of backers donating roughly equitable amounts) i.e. the spread is less.

Only that can explain why certain counties lie in 2nd or 3rd sextile for “total amount raised per unit population”, but lie in the lowest sextile for “number of backers per unit population”. In other words, those counties exhibit a low backer count density in Fig.7, but in Fig.6 the same counties do not fare so poorly. One such example is the Helena -Great Falls region in Montana pointed out with a small blue arrow in both figures.

II. Project success rate and category wise study

We tried to check whether projects with higher goal amounts tend to generate more funding. OLS (ordinary least squares) regression was used to test whether such a linear relationship exists, and whether it is bolstered by factors such as location of project and type of project.

```
In [9]: ols(y=sf_US['pledged'],x=sf_US[['goal']])

Out[9]:
-------------------------Summary of Regression Analysis-------------------------

Formula: Y ~ <goal> + <intercept>

Number of Observations:         82098
Number of Degrees of Freedom:   2

R-squared:         0.0033
Adj R-squared:     0.0032

Rmse:          74532.8327

F-stat (1, 82096):   268.3989, p-value:     0.0000

Degrees of Freedom: model 1, resid 82096

-----------------------Summary of Estimated Coefficients------------------------
      Variable       Coef    Std Err     t-stat    p-value    CI 2.5%   CI 97.5%
--------------------------------------------------------------------------------
          goal     0.0186     0.0011      16.38     0.0000     0.0164     0.0208
     intercept  7665.8969   260.9072      29.38     0.0000  7154.5187  8177.2750
---------------------------------End of Summary---------------------------------
```

Fig 8. OLS regression between pledge amount and goal

Since the R-squared value is very low at 0.33% we conclude that such a simple linear relationship does not exist. Simply asking for more money is not enough to gather more funds.

Is it then possible that projects originating from California, hailed as the Golden state, generate more funding? According to the following study, the answer is again no.

No such simple linear relationship can be established looking at the state of origin alone. Predicting project success based on variables such as location and industry is hard. Also the distribution of funding is highly skewed towards successful projects (i.e. successful projects gather funding several times over their original goal amount, whereas unsuccessful projects fail miserably) as Raymond states in his paper[15.]

It is quite possible that while California is home to more number of projects and start-ups than any other state, its success rate is not significantly higher.

```
In [10]: ols(y=sf_US['pledged'],x=sf_US[['goal','is_CA']])

Out[10]:
-------------------------Summary of Regression Analysis-------------------------

Formula: Y ~ <goal> + <is_CA> + <intercept>

Number of Observations:         82098
Number of Degrees of Freedom:   3

R-squared:         0.0042
Adj R-squared:     0.0042

Rmse:          74498.1346

F-stat (2, 82095):   173.0702, p-value:     0.0000

Degrees of Freedom: model 2, resid 82095

-----------------------Summary of Estimated Coefficients------------------------
      Variable       Coef    Std Err     t-stat    p-value    CI 2.5%   CI 97.5%
--------------------------------------------------------------------------------
          goal     0.0185     0.0011      16.29     0.0000     0.0163     0.0207
         is_CA  5626.4781   639.1604       8.80     0.0000  4373.7236  6879.2326
     intercept  6490.2860   292.9919      22.15     0.0000  5916.0219  7064.5501
---------------------------------End of Summary---------------------------------
```

Fig 9. OLS regression between pledge amount, goal, whether project is located in CA

Finally we added one more Boolean variable that represented whether the project was of the type technology, including software, hardware, electronic media etc. We focused on technology since it has the highest potential to generate employment compared to other Kickstarter project categories such as art, theater, music etc.

```
In [16]: ols(y=sf_US['pledged'],x=sf_US[['goal','is_CA','is_tech']])

Out[16]:
-------------------------Summary of Regression Analysis-------------------------

Formula: Y ~ <goal> + <is_CA> + <is_tech> + <intercept>

Number of Observations:         82098
Number of Degrees of Freedom:   4

R-squared:         0.0052
Adj R-squared:     0.0052

Rmse:          74460.9709

F-stat (3, 82094):   143.1514, p-value:     0.0000

Degrees of Freedom: model 3, resid 82094

-----------------------Summary of Estimated Coefficients------------------------
      Variable       Coef    Std Err     t-stat    p-value    CI 2.5%   CI 97.5%
--------------------------------------------------------------------------------
          goal     0.0184     0.0011      16.21     0.0000     0.0162     0.0206
         is_CA  5578.1236   638.8637       8.73     0.0000  4325.9508  6830.2963
       is_tech 23172.0196  2543.9469       9.11     0.0000 18185.8836 28158.1556
     intercept  6257.7518   293.9563      21.29     0.0000  5681.5974  6833.9063
---------------------------------End of Summary---------------------------------
```

Fig 10. OLS regression between pledge amount, goal, whether project is located in CA, whether it is of type Technology

It can be seen from above snapshot that the R-squared value improved but only marginally.

Hence the conclusion is:

We cannot say for certainty that technology projects from California are able to generate more funding – or at least that these variables do not seem to follow a simple linear relationship.

Below table indicates number of completed projects by state (for states with more than 2,000 completed Kickstarter projects).

State	Total number of projects	Number of Technology projects	Technology Projects as % of total	Number of Music projects	Music Projects as % of total
CA	17031	1611	9.46%	1027	6.03%
NY	11115	647	5.82%	692	6.23%
TX	4343	70	1.61%	356	8.20%
FL	3364	291	8.65%	238	7.07%
IL	3298	206	6.25%	192	5.82%
WA	2840	23	0.81%	192	6.76%
PA	2569	167	6.50%	175	6.81%
MA	2399	171	7.13%	218	9.09%
OR	2285	153	6.70%	165	7.22%
GA	2106	154	7.31%	168	7.98%
TN	2000	6	0.30%	389	19.45%

Fig 11. Fraction of technology projects as a % of all projects – by state

We see that **California does host a higher fraction of technology projects** (9.46%) compared to other top states. However the fraction isn't dramatically higher compared to the nation-wide average of 7.45%.

It is so because **most of the technology projects in CA are located in a handful of counties** such as San Francisco and Los Angeles. Rest of the California counties are virtually indistinguishable from other North American states. This assertion is supported by the heat-map on the next page which is similar to Figure 2, but only plots the number of technology related projects.

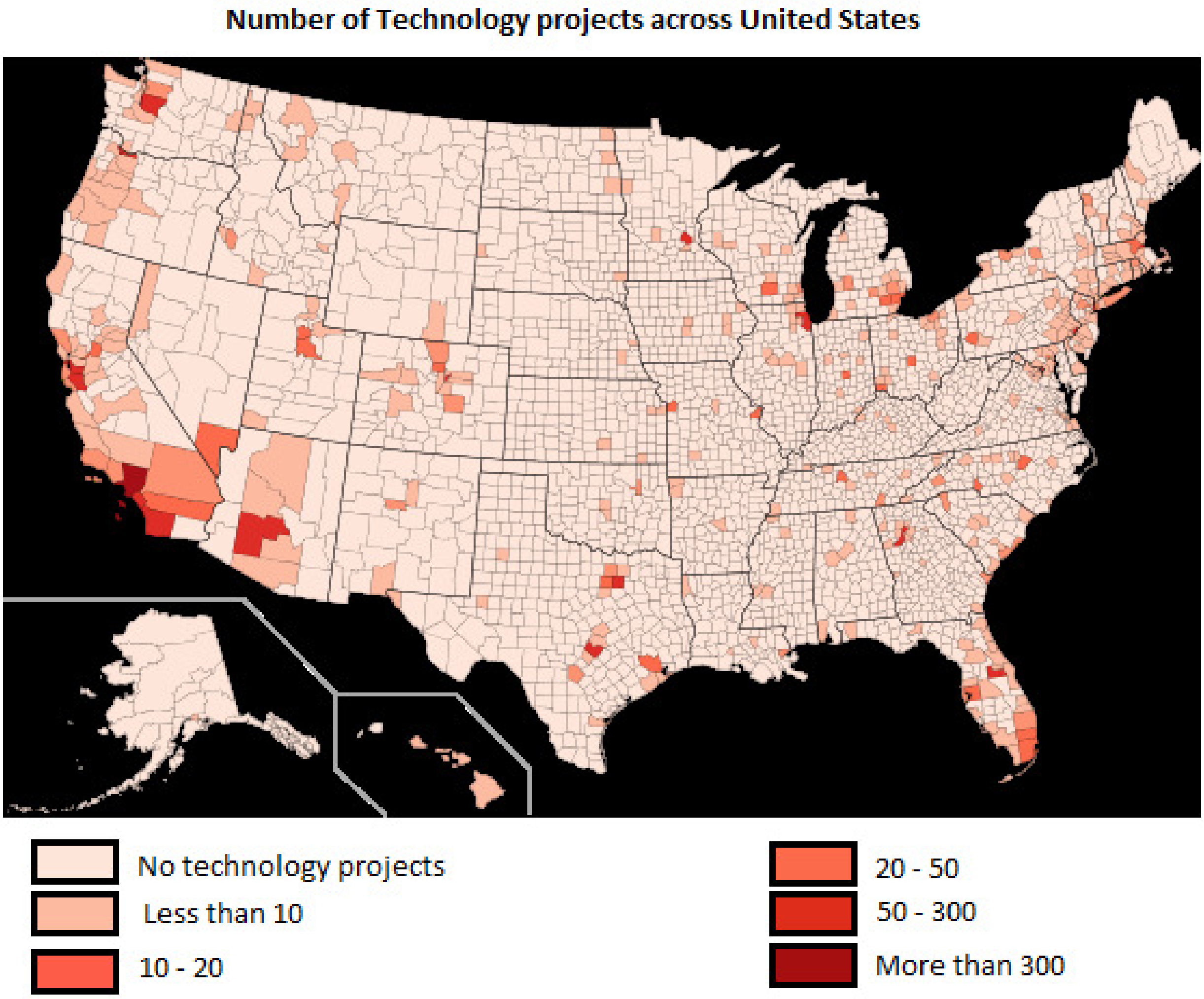

Fig 12. Heat-map showing only the count of technology projects

RESULTS/CONCLUSIONS

There are two parts to my work. First is establishing a database. We created a MySQL database containing information about 105,598 Kickstarter projects (hosted online as of Feb'14). It is accessible at rosencrantz.berkeley.edu. Schema details are available in the section on methodology.

Second, we have visualizations to demonstrate the capabilities of this database. The graphics help to get an overall sense of the distribution of crowdfunding activities in the United States. These also establish beyond doubt that **the crowdfunding activity density across United States is not uniform, also the profiles of projects hosted on Kickstarter (based on their category) vary all over the nation.**

This database can be used to study trends in crowdfunding and verify claims made by crowdfunding platforms regarding their success rate etc. Further category-wise studies can also be carried out to find out which types of projects (Music, art, technology etc.) are more likely to succeed in crowdfunding. One can also study the variables that impact the success of crowdfunding project in general and train a model that determines the probability of success of a project (or expectation of the pledge money) as a function of creator's profile, location, project category, key words in description etc.

One of the shortcomings of this database it that it does not consist of detailed information on backers. If this information is included, home bias effect can be studied. This database should also be combined (and normalized) with other platform's databases (e.g. www.indiegogo.com) in order to increase the number of data points and reduce the impact of biases/irregularities in the data that may be specific to Kickstarter.

REFERENCES:

[1] Massolution. (2014). *Industry Report.* [Online] Available:

http://research.crowdsourcing.org/2013cf-crowdfunding-industry-report

[2] Collins, Swart, Zhang, "The rise of future finance" *in Academic Symposium on Crowdfunding, UC Berkeley Fung Institute*, Berkeley CA, 2013

http://www.funginstitute.berkeley.edu/sites/default/files/Rise_of_Future_Finance_1_1.pdf

[3] Sec.gov Jumpstart Out Business Startups (JOBS) act [Online] Available:

http://www.sec.gov/spotlight/jobs-act.shtml

[4] Best, Neiss, Stralser, Fleming, "How big will the debt and equity crowdfunding investment market be? Comparisons, assumptions and estimates" *in Academic Symposium on Crowdfunding, UC Berkeley Fung Institute*, Berkeley CA, 2013

http://www.funginstitute.berkeley.edu/sites/default/files/Crowdfund_Investment_Paper.pdf

[5] Ryan Caldbeck (2013) *Crowdfunding- Why Angels, Venture Capitalists And Private Equity Investors All May Benefit* [Online] Available:

http://www.forbes.com/sites/ryancaldbeck/2013/08/07/crowdfunding-why-angels-venture-capitalists-and-private-equity-investors-all-may-benefit/

[6] Lora Kolodny (2013) *Angellist and beyond – What VCs really think of crowdfunding* [Online] Available: http://blogs.wsj.com/venturecapital/2013/10/08/angellist-and-beyond-what-vcs-really-think-of-crowdfunding/

[7] Agrawal, Catalini, Goldfarb "The geography of crowdfunding" in *NBER Working paper series* (Feb 2011) [Online] Available: http://www.nber.org/papers/w16820.pdf

[8] Lin and Viswanathan "Home bias in online investments: An empirical study of an online Crowd Funding market" (Feb 2013) [Online] Available:

http://funginstitute.berkeley.edu/sites/default/files/Home_Bias_Online_Investments.pdf

[9] Link to Beautiful Soup official documentation [Online]
http://www.crummy.com/software/BeautifulSoup/bs4/doc/#quick-start

[10] Link to Selenium browsers automation [Online]

http://docs.seleniumhq.org/

[11] Wikipedia page for xvfb, the virtual server display [Online]

http://en.wikipedia.org/wiki/Xvfb

[12] Python data analysis library official page [Online] http://pandas.pydata.org/

[13] Source of SVG graphic [Online] http://en.wikipedia.org/wiki/File:Usa_counties_large.svg

[14] Flowingdata's tutorial on creating python heatmaps [Online]

http://flowingdata.com/2009/11/12/how-to-make-a-us-county-thematic-map-using-free-tools/

[15] Raymond Le Von Mizener "The Crowdfunding Project (with respect to machine learning)" (2014) Master of Engineering thesis, UC Berkeley/Fung Institute

[16] Kartikeya Sahai "Kickstarting Crowdfunding" (2014) Master of Engineering thesis, UC Berkeley/Fung Institute

[17] State and Country quickfacts by United States Census Bureau [Online]

http://quickfacts.census.gov/qfd/download_data.html

[18] Internet in the United States on Wikipedia the free encyclopedia [Online]

http://en.wikipedia.org/wiki/File:US_Wireline_Broadband_31Dec2012.tiff

APPENDIX

Notes about the methodology used to create geographical visuals:

- Only projects which successfully completed campaign or failed to meet goal by their campaign deadline (“Successful” or “Failed”) are included. “Live” and “Suspended” projects are excluded from this work. These graphics show a total of 80,197 projects over 1,827 counties having at least one completed project.
- City-counties (such as San Fransisco) are counted as one. Boroughs are included in the picture, however less information is known regarding projects in such small geographies.
- Some amount of fuzzy-matching and manual inspection was also needed in assigning a project to its county of origin. This is because some founders and backers register their location as, for example, “Lower East Side” or “Manhattan” instead of “New York” or “Hollywood” instead of “Los Angeles”. Wherever possible, such places were manually matched.
- Since the projects are over a period of time, population of each county is the average population according to Q’1 2010 census as well as 2012 and 2013 estimates.