

Process Book

Johnathan Budd and Jared Rosen

Overview and Motivation

While the startup eco system relies on the constant creation of new ideas and companies, many if not all of these companies would never have a chance of succeeding without the support of venture capitalists to invest in their ideas. Some of these venture capital firms have been around for decades while others are startups themselves, but they all play a crucial role in the startup ecosystem. Our goal is to show how these firms work together and make it easy to understand how specific firms invests, geographically, categorically, and over time.

Related Work

//TODO

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

Questions

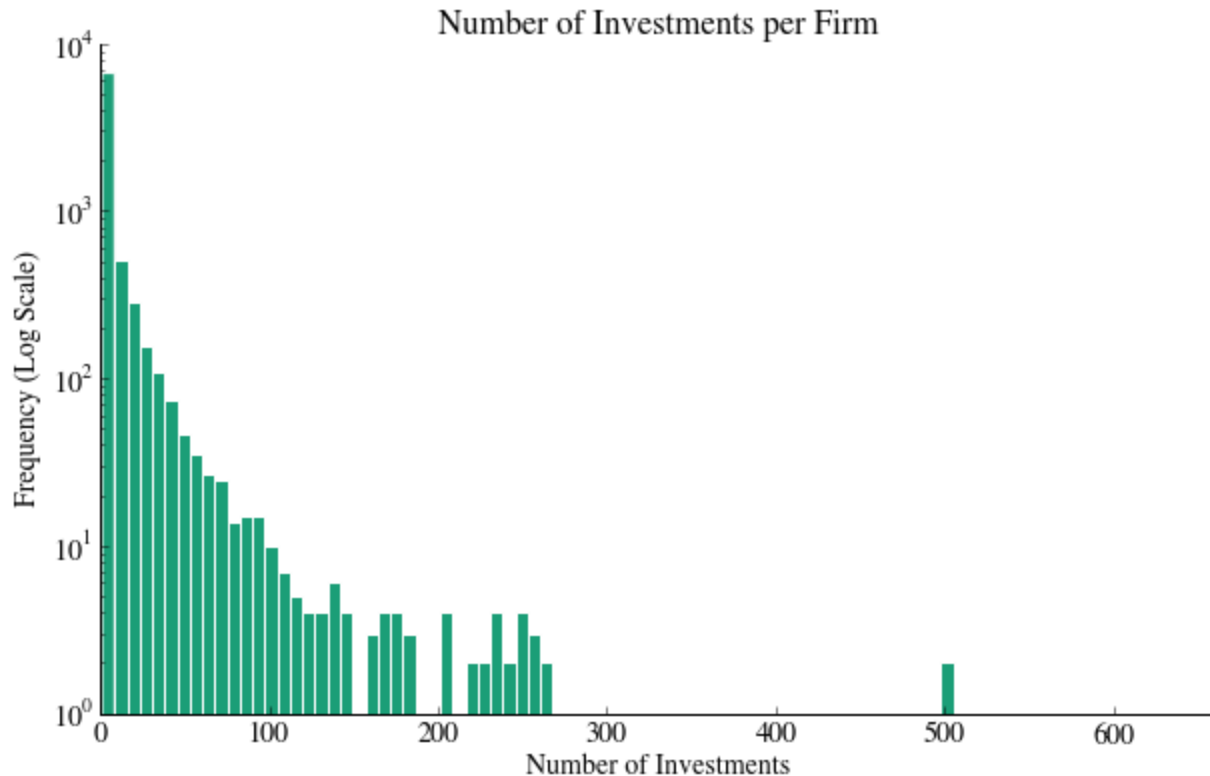
Data / Exploratory Data Analysis

Our data was gathered via the CrunchBase API. The script was written in Python (iPython Notebook) and the data was saved as a json file. To create our final data set we had to make several calls to different api endpoints and restructure the returned information as to make it more suitable for our visualization.

We began by gathering a large amount of raw data on venture firms

- Download list of all financial institutions using the entities endpoint
- For each financial institution query the entity endpoint to get complete information on that firm
- Remove any institutions that have never made an investment or are missing other critical data
- Determine total number of investments for each remaining firm

This left us with a little over 8000 firms. However, many of these entries were for small, orphaned, or outdated firms and, as a result, did not accurately represent the flow of capital through the startup ecosystem. To create a more intelligible visualization we decided to focus on a subset of the the largest and most active firms (as determined by total number of investments). We created a histogram of the number of investments per firm using “square root of the number of firms” bins. We noticed a clear divide in the data in regards to firms with more than versus less than 200 investments.



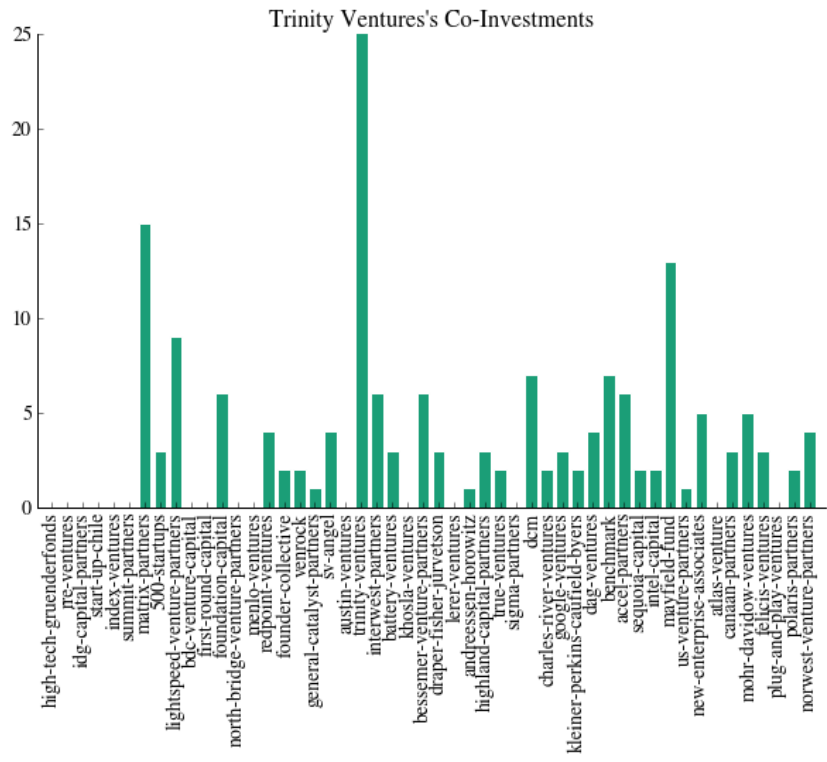
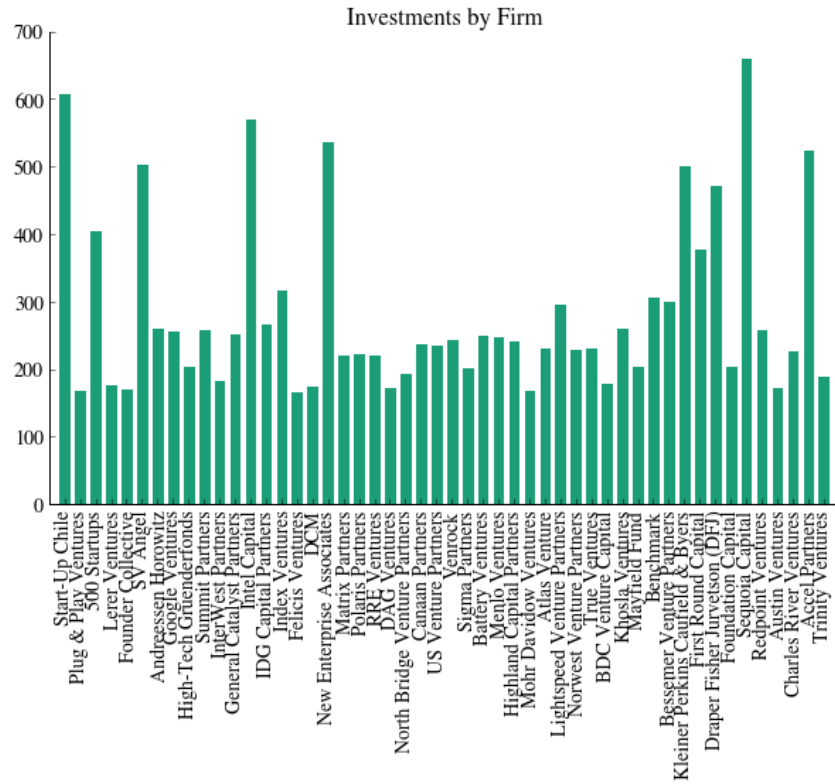
By subsetting the data at this point (and rounding slightly for convenience) we ended with a group of 50 of the most active and influential firms. We then proceed to complete the data gathering process.

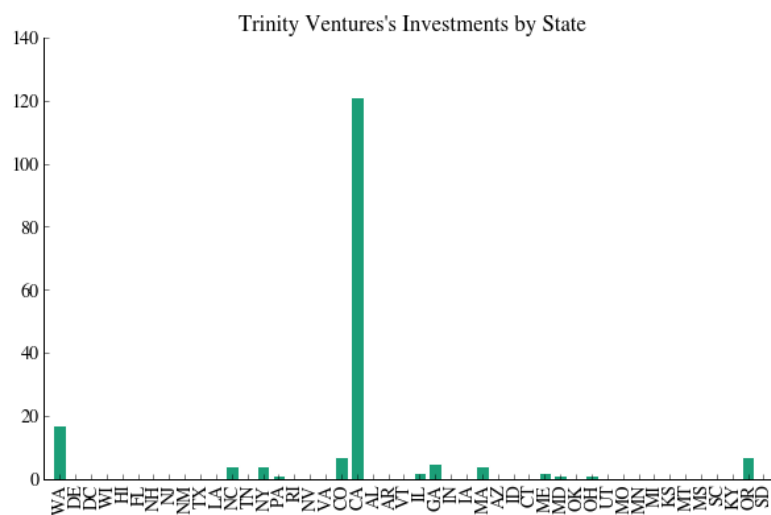
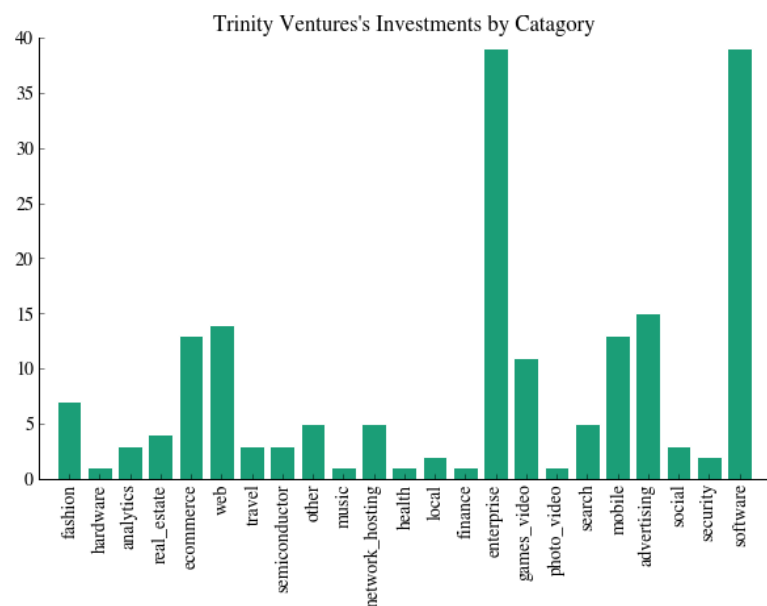
- Determine all companies in the dataset and make a call to the entity API endpoint to determine the company's category and location
- Determine the breakdown of a firm's investments by category, year, stage, and geography
- Calculate the number of co-investments between every firm in our data set

We then did some preliminary data exploration in Python using matplotlib before saving the data as a json file.

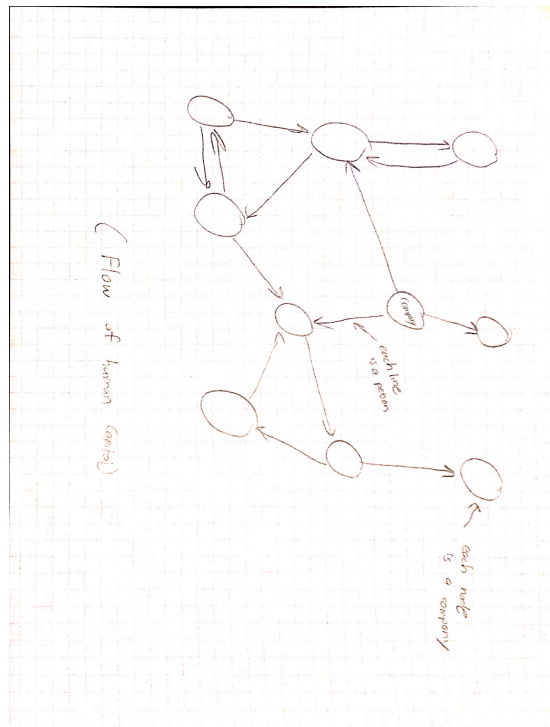
Note: Although the iPython notebook could be run from top to bottom it would take an inordinate amount of time, as such, local files are loaded in and the cells that call the API are commented out. This file (downloadData-Funds.ipynb), as well as the json files (allFirms_2014-04-07-01-06-44_out.json), can be found in the data directory of our repository.

Once we have successfully downloaded our data we made sample visualizations in Python to get proof of concept of our design vision before proceeding to implement everything in d3. Find a sample of these visualizations below.

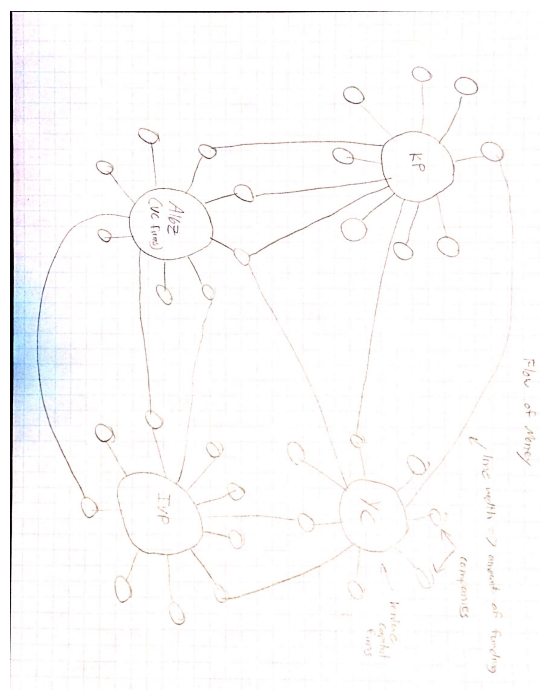




Design Evolution



Use a linked node graph to show how people moves between startups - flow of human capital



Track the flow of money from VC firms to see what companies have similar investors and how money moves throughout the ecosystem

Implementation

- Co Occurrence Matrix
 - This was the first visualization we implemented as it is our main visualization. It shows how often Venture capital firms co-invest in companies and what kind of investments they are making (i.e. seed round, late stage)
- Company Specific Visualizations
 - Bar graph
 - The bar graph shows what categories a firm invests in such as biotechnology or clean energy
 - Line Graph
 - This chart shows the number of investments per year over time. We decided not to include the values from 2014 because having only 4 months of data provided misleading information
 - Heatmap
 - The map shows in which states companies make investments. In order to prevent popular states (California) from distorting the scale and making the rest of the map unreadable we clamped the domain of the color mapping function so it is easier to identify every state that a firm invests in. We are still working on improving this to be more accurate.

Evaluation

//TODO

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?