Process Book

Johnathan Budd | Jared Rosen

Abstract

We have created an interactive visualization to show how venture capital firms work together and how they invest geographically, categorically, and over time. We gathered data on fifty of the top venture capital firms, from which we created four visualizations. A co-occurrence matrix shows how often two firms have invested together and what type of firm they are. A bar graph displays the number of investments each firm makes in categories ranging from biotech to education, a heat map displays the geographic distribution of investments across the country, and a line graph tracks the distribution of investments over time.

- Our project website is: www.vcshowdown.com
- Our screencast can be found at: http://youtu.be/NpDiIC5VKlk

Overview and Motivation

While the startup ecosystem 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

We were inspired by Micke Bostock's co-occurrence matrix. We thought it was a very interesting way to show the relatedness between nodes in a complex data set. We also like the idea of having one primary visualization drive other sub-visualization (like we did in HW3 and HW4 and the Truila example [below]).

- http://bost.ocks.org/mike/miserables/
- http://www.trulia.com/vis/tru247/
- http://visual.ly/vizbox/startup-universe/
- http://pando.com/2013/08/21/introducing-pandomaps-a-new-interactive-tool-for-mapping-startup-ecosystems/

Questions

Original Questions:

- How does money flow between venture capital firms and startups?
- How predictable are the investment patterns of venture capital firms?

While these were the questions we originally set out with in an attempt to understand how money flows in the startup ecosystem. We then refined our goals and decided to focus on the following questions:

- How often do venture capital firms co-invest?
- Do venture capital firms co-invest with firms that have similar profiles?
- Do venture capital firms tend to invest in particular types of companies or over a broad range of categories?
- Do firms concentrate their investments in particular geographic regions?
- How do young firms compare to older more established firms?

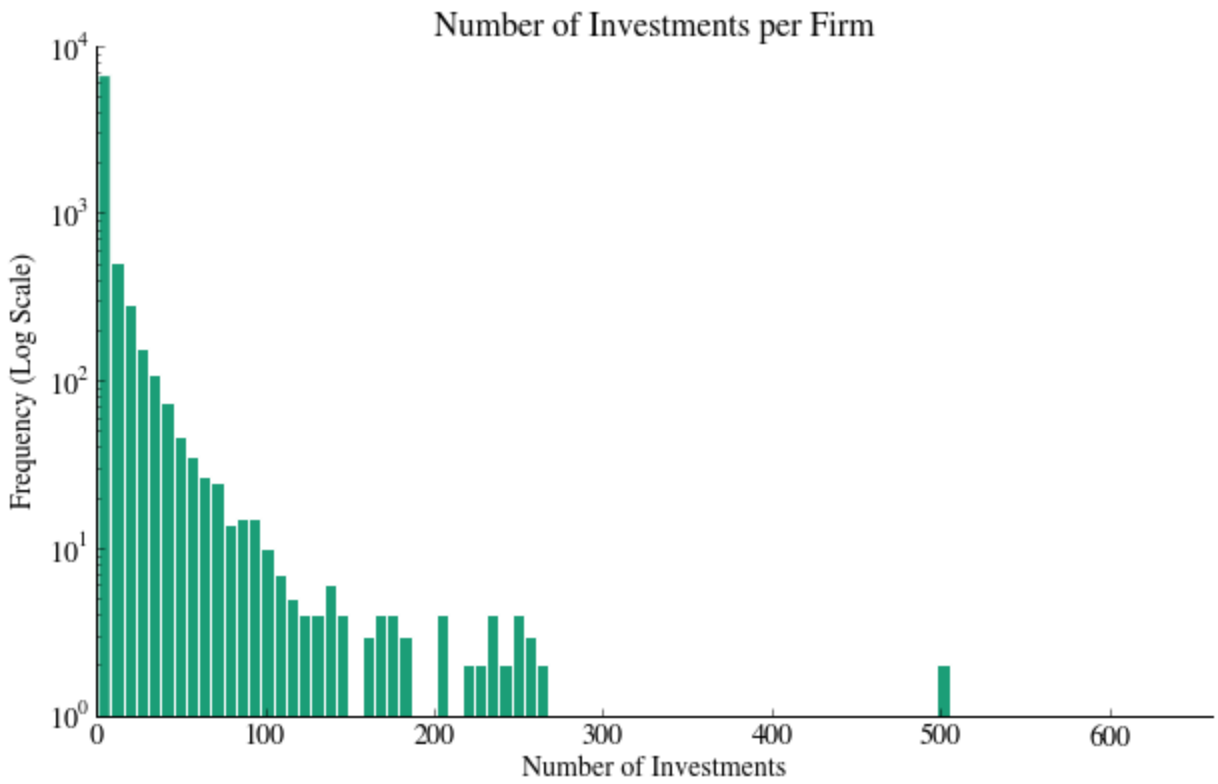
Data / Exploratory Data Analysis

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. All data used in this visualization was downloaded on April 4th, 2014.

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 information
- Determine total number of investments for each of the remaining firms

This left us with a little over 8000 firms. However, many of these entries were for small firms or from orphaned or outdated data and, as a result, did not accurately represent the flow of capital though 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.



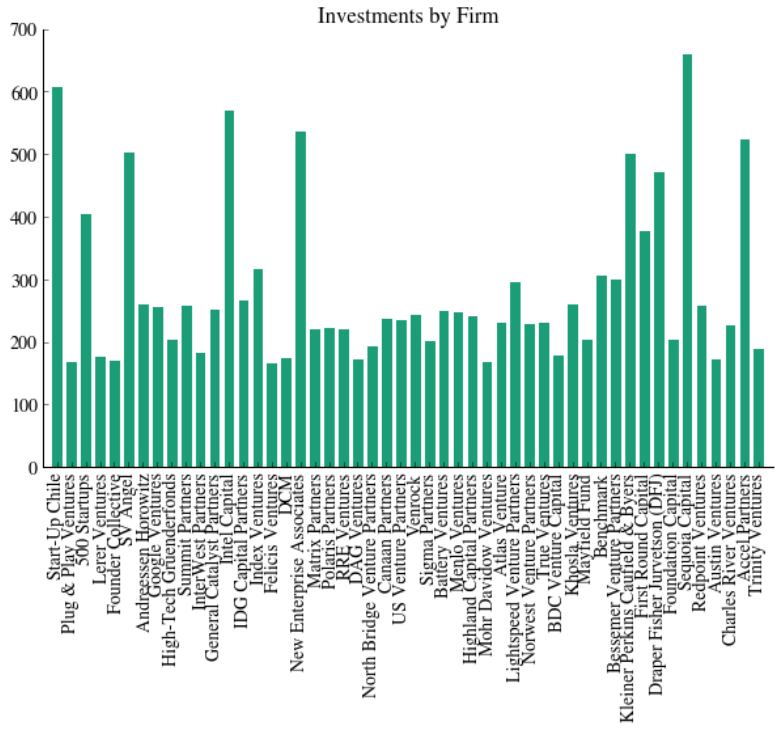
By subsetting the data at this point (and rounding slightly for convenience) we ended with a group of 50 of the most active and influential firms. We then proceed to complete the data gathering process.

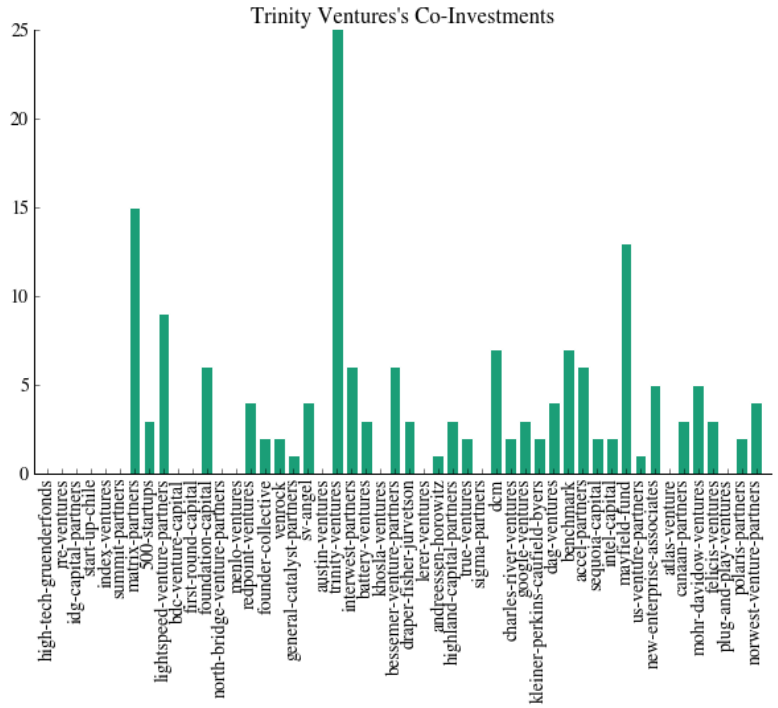
- Determine all companies in the dataset and make a call to the entity API endpoint to determine the companys' category and location
- Determine the breakdown of a firm's' investments by category, year, stage, and geography
- Calculate the number of co-investments between every firm in our data set

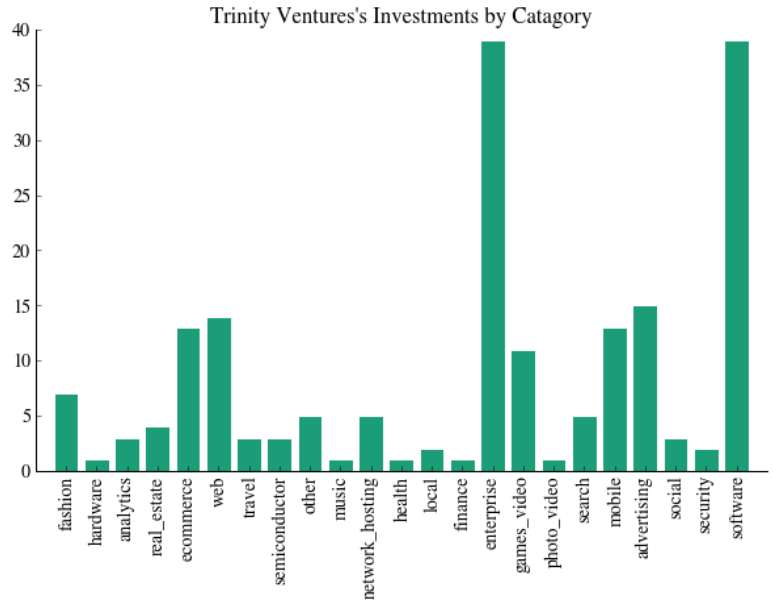
We then did some preliminary data exploration in Python using matplotlib before saving the data as a json file.

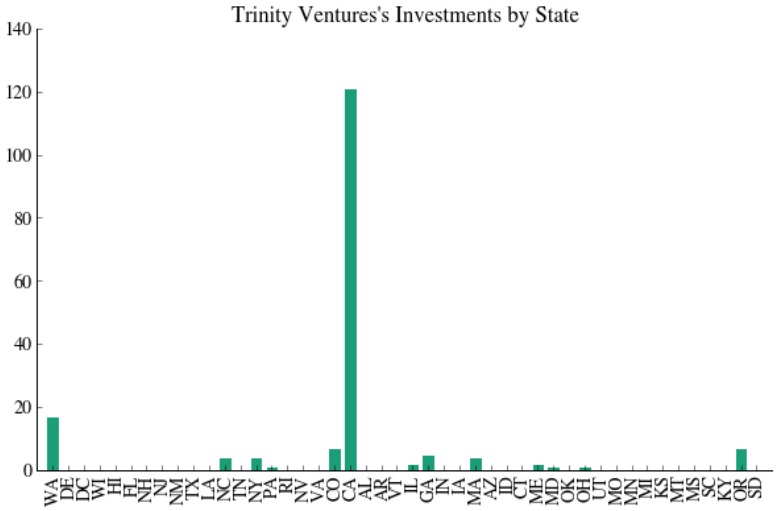
Note: Although the iPython notebook could be run from top to bottom it would take an inordinate amount of time, as such, local files are loaded in and the cells that call the API are commented out. This file (downloadData-Funds.ipynb), as well as the json files (allFirms_2014-04-07-01-06-44_out.json),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

Once we has 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.


Investments by Firm
700
600
500
400
300
200
100
0
Start-Up Chile
Plug & Play Ventures
500 Startups
Lerer Ventures
Founder Collective
SV Angel
Andreessen Horowitz
Google Ventures
High-Tech Gründerfonds
Summit Partners
InterWest Partners
General Catalyst Partners
Intel Capital
IDG Capital Partners
Index Ventures
Felicis Ventures
DCM
New Enterprise Associates
Matrix Partners
Polaris Partners
RRE Ventures
DAG Ventures
North Bridge Venture Partners
Canaan Partners
US Venture Partners
Venrock
Sigma Partners
Battery Ventures
Menlo Ventures
Highland Capital Partners
Mohr Davidow Ventures
Atlas Venture
Lightspeed Venture Partners
Norwest Venture Partners
True Ventures
BDC Venture Capital
Khosla Ventures
Mayfield Fund
Benchmark
Bessemer Venture Partners
Kleiner Perkins Caufield & Byers
First Round Capital
Draper Fisher Jurvetson (DFJ)
Foundation Capital
Sequoia Capital
Redpoint Ventures
Austin Ventures
Charles River Ventures
Accel Partners
Trinity Ventures

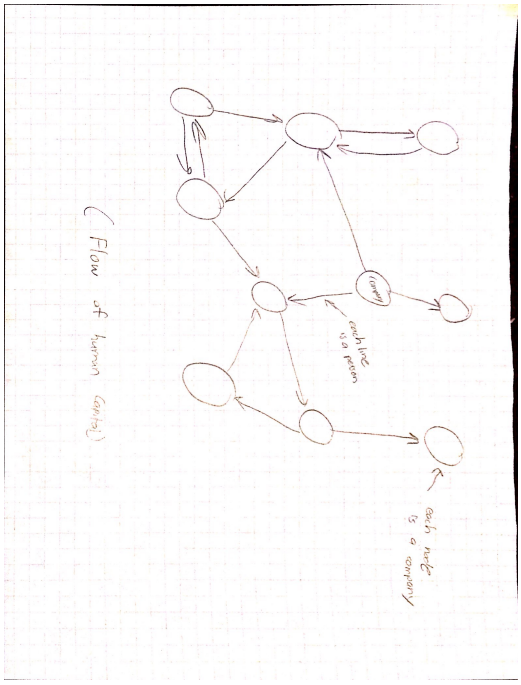

Trinity Ventures's Co-Investments
25
20
15
10
5
0
high-tech-gruenderfonds
pre-ventures
idg-capital-partners
start-up-chile
index-ventures
summit-partners
matrix-partners
500-startups
lightspeed-venture-partners
bdc-venture-capital
first-round-capital
foundation-capital
north-bridge-venture-partners
menlo-ventures
redpoint-ventures
founder-collective
venrock
general-catalyst-partners
sv-angel
austin-ventures
trinity-ventures
interwest-partners
battery-ventures
khosla-ventures
bessemer-venture-partners
draper-fisher-jurvetson
lerer-ventures
andreessen-horowitz
highland-capital-partners
true-ventures
sigma-partners
dcm
charles-river-ventures
google-ventures
kleiner-perkins-caufield-byers
dag-ventures
benchmark
accel-partners
sequoia-capital
intel-capital
mayfield-fund
us-venture-partners
new-enterprise-associates
atlas-venture
canaan-partners
mohr-davidow-ventures
felicis-ventures
plug-and-play-ventures
polaris-partners
norwest-venture-partners


Trinity Ventures's Investments by Catagory
0
5
10
15
20
25
30
35
40
fashion
hardware
analytics
real_estate
ecommerce
web
travel
semiconductor
other
music
network_hosting
health
local
finance
enterprise
games_video
photo_video
search
mobile
advertising
social
security
software

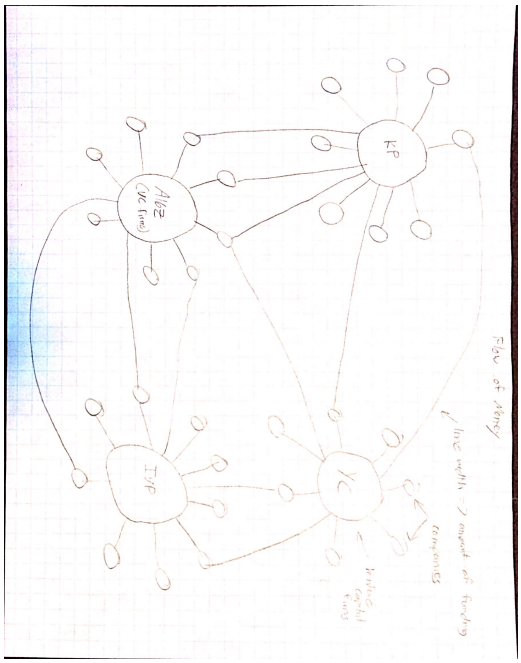

Trinity Ventures's Investments by State
0
20
40
60
80
100
120
140
WA
DE
DC
WI
HI
FL
NH
NJ
NM
TX
LA
NC
TN
NY
PA
RI
NV
VA
CO
CA
AL
AR
VT
IL
GA
IN
IA
MA
AZ
ID
CT
ME
MD
OK
OH
UT
MO
MN
MI
KS
MT
MS
SC
KY
OR
SD


Trinity Ventures's Investments by Catagory
0
5
10
15
20
25
30
35
2002
2000
2001
2006
2007
2004
2005
2014
2008
2009
2011
2010
2013

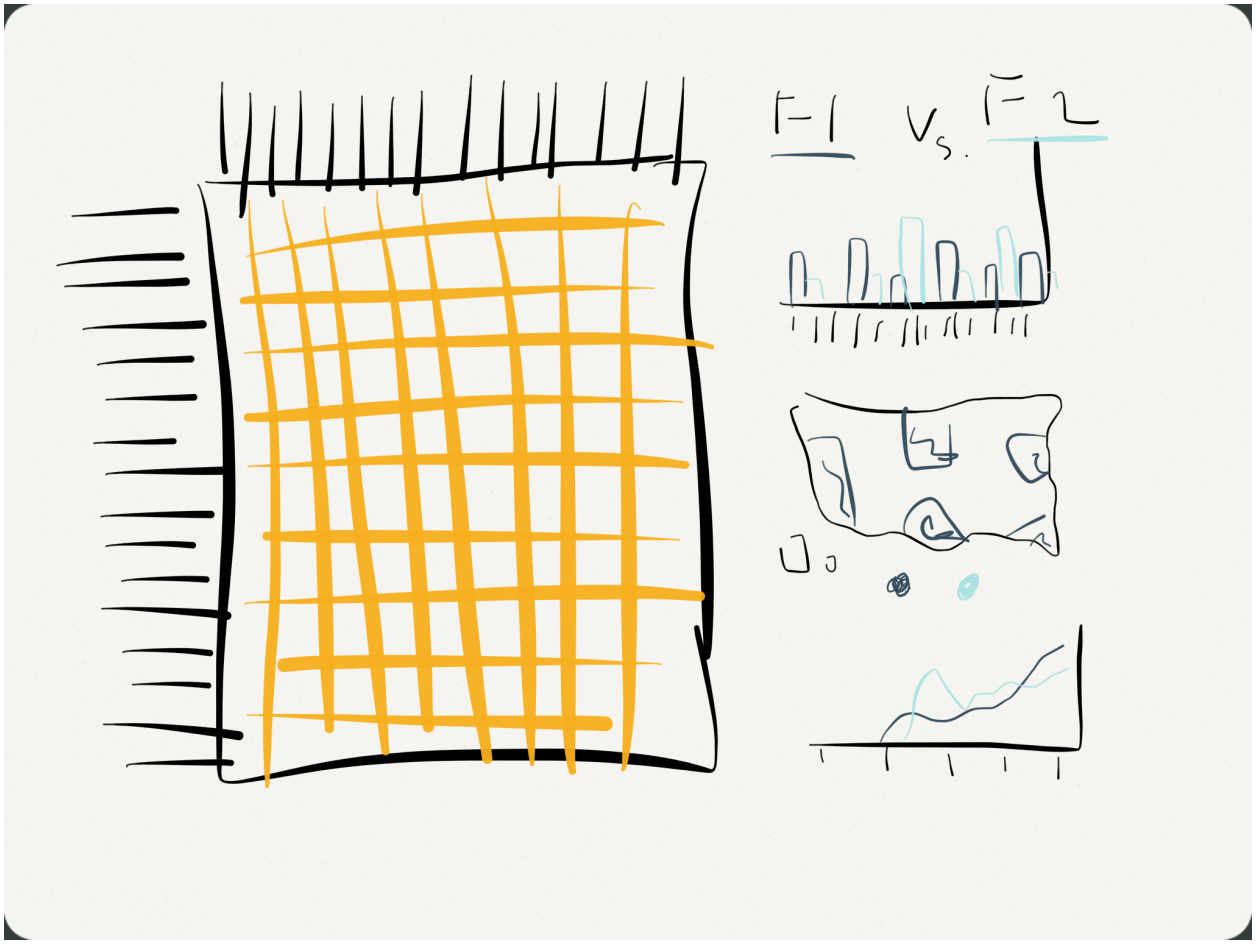
Design Evolution



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

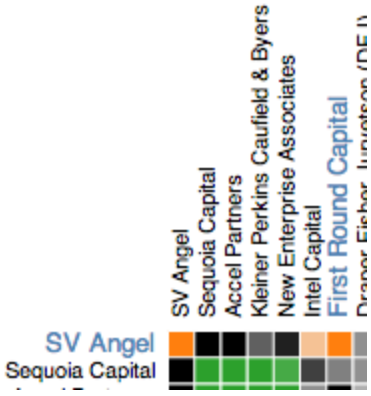
- Co-Occurrence Matrix

The co-occurrence matrix is our main visualization and it is used to control the three sub-visualizations. The x and y axis are identical and both list the fifty venture capital firms we have chosen analyze. The intersection of two firms in the matrix represents the number of times the two firms have co-invested. This value is encoded by the intensity of the cell in the matrix, with darker colors corresponding to more co-investments and blank cells corresponding to zero co-investments.

In addition to representing the number of coinvestments the matrix also uses color to represent the type of investments the firms traditionally make between early, mid and late stage. If both of the firms make the same type of investment the cell is colored according to the below scale, otherwise it is colored black. This is intended to provide insight on how different types of firms work together.

Early → Orange Mid → Green Late → Red

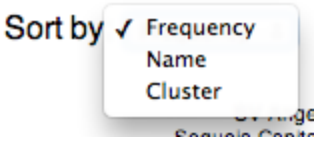
When a user hovers over a cell the two firms corresponding to the cell are highlighted to make it easy to identify and compare specific firms in the larger matrix.



When a cell of the matrix is clicked it is highlighted and the three sub visualizations (bar chart, heatmap, line graph) switch from displaying average values for all fifty firms to specific data for the two firms represented by the cell. In addition the header of the visualization is dynamically updated to display the selected firms. The names of the firms are colored in correspondence with the colors of the three sub visualizations. The text also displays the number of co investments. Clicking on the highlighted cell returns the sub visualizations to displaying the averages for all firms.

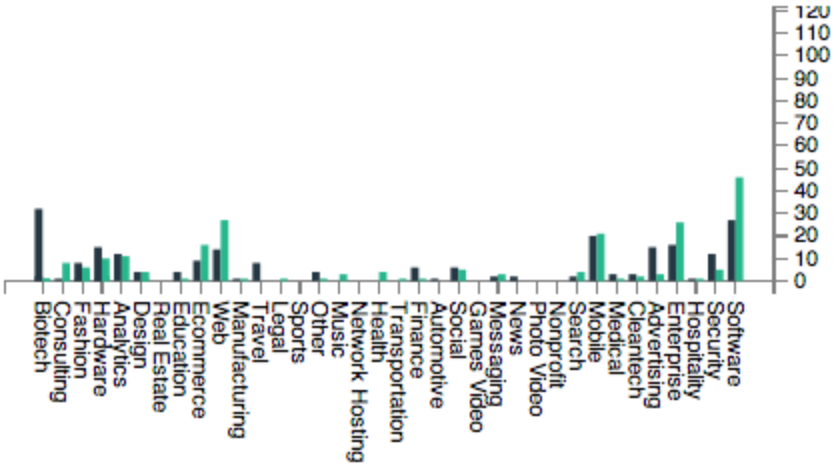
Because the diagonal of the matrix corresponds to all of a firms co-investments with itself, clicking on one of these cells updates the three sub visualizations to display data for only this firm, and the header displays the total number of investments by this firm.

The matrix can be sorted in three ways, alphabetically, by frequency of co-investments, or by clusters of early mid and late stage investments.

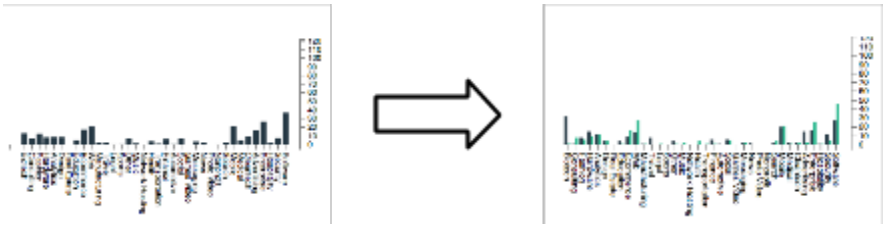


We chose to have the matrix default to sorting by frequency because we felt it was most relevant to the story we are telling with this visualization.

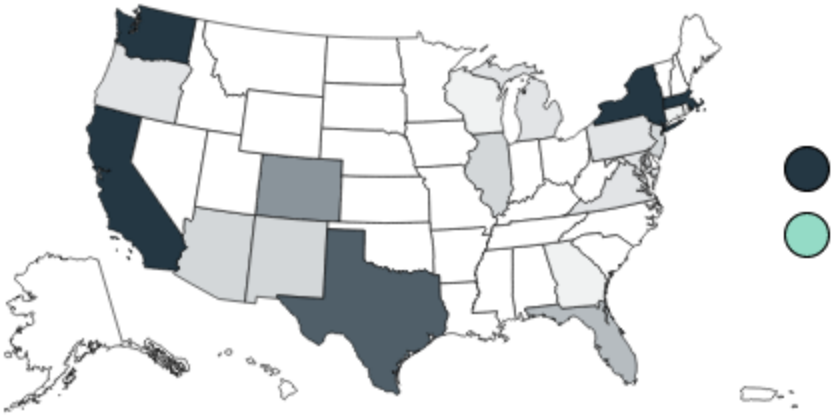
- Bar Chart



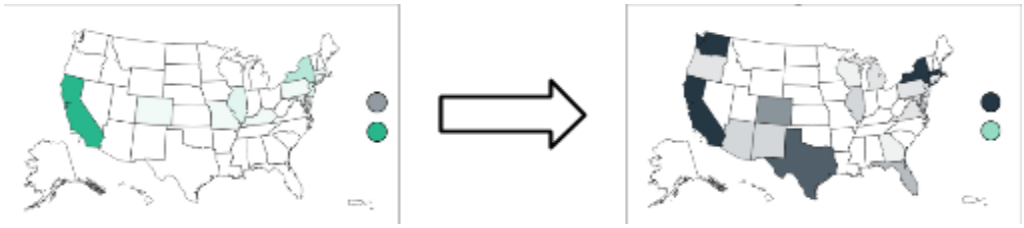
The bar chart displays the number of investments made in different industry categories, ranging from biotech to software. By default the chart displays the averages over all fifty firms. When a cell of the matrix is selected two sets of bars appear representing the investments made by the two corresponding firms.



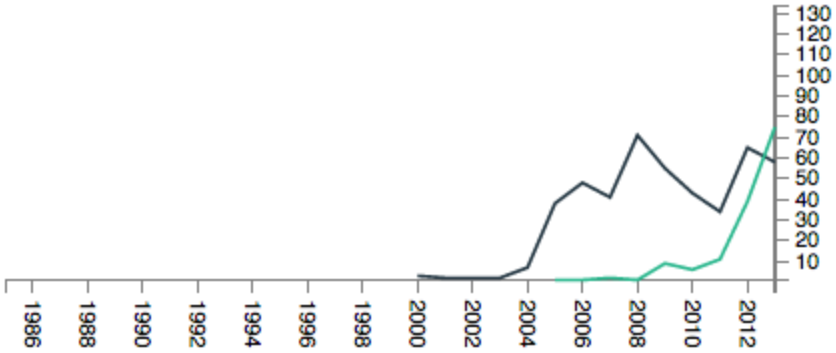
- Heatmap



The heat map displays the number of investments made in each state. This value is encoded by the intensity of the color, with darker colors representing a greater number of investments. Like the bar chart, the map displays average values for all fifty firms by default. When a cell in the matrix is selected a single map is still displayed but two buttons are added next to the map. Clicking on these buttons switches the map between the two firms. It was determined that this was a cleaner solution than generating two separate maps.



- Line Graph



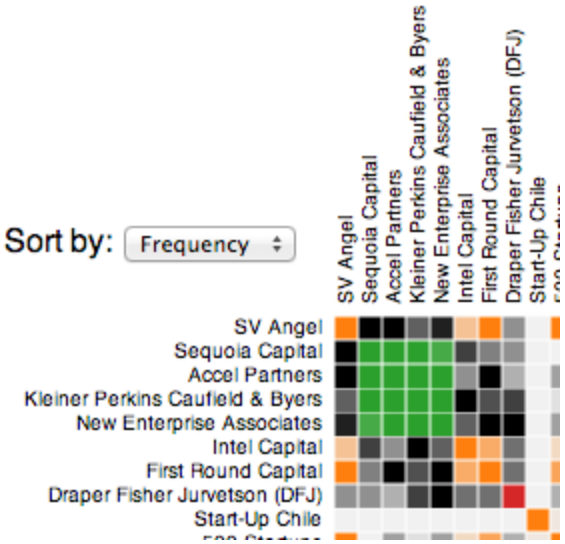
The line graph displays the number of investments made each year by the firms. Like the bar chart and heat map this graph displays the average number of investments per year by default then switches to displaying two separate lines when a cell is selected to show the yearly investments of the two firms.

Evaluation

Our visualization provided a number of key insights into the data surrounding the venture capital ecosystem. While there are thousands of startups there are only a small number of highly active venture capital firms making a few hundred investments, suggesting that venture capital firms provide an aspect of continuity and institutional memory in the constantly changing startup ecosystem. Examining the line graph shows a rapidly accelerating pace of investments over the past few decades with most of the investments being made after 2004 and an almost constant increase year over year since then.

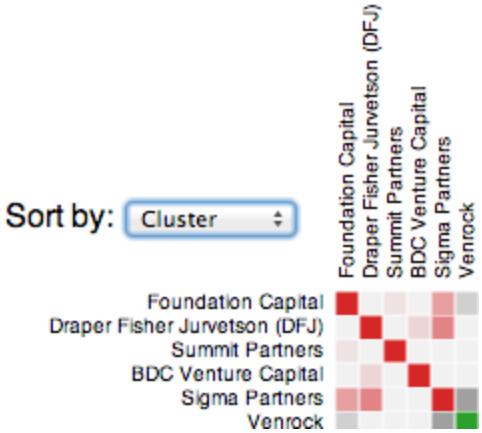
We see a strong bias towards larger coastal population centers in the geographic distribution of investments with the majority concentrated in California, New York, Massachusetts, and Texas. Despite this bias a non trivial number of investments were made in other regions of the country including Utah, Colorado, and Indiana. We also noticed a strong tendency towards and against particular industry categories. The majority of investments were made in companies related to software, web, and enterprise platforms while very few investments were made in the transportation, automotive, and real estate categories. This suggests that firms prefer to invest in companies that require low initial capital to prove their viability.

Many of the insights discussed so far confirm the stereotype of west coast software startups that has become increasingly popular in the recent years. In addition to these insights our visualization provides a more in depth understanding of how venture capital firms work together. When the co-occurrence matrix is sorted by frequency, the firms that have co-invested the most times are located in the upper left corner of the matrix.



We can see from this portion of the matrix that many of the firms that co-invest most often typically make different style of investments, i.e. a firm that typically makes mid stage investments makes a large number of co-investments with early stage firms. This is illustrated by the large number of black and grey cells concentrated near the upper left corner.

While our visualization shows that it is very common for different styles of venture capital firms to co-invest it also shows how rare it is for late stage firms to co-invest.



When the matrix is sorted by cluster and we take into account the fact that the same information is repeated on either side of the diagonal, a necessary evil of a co-occurrence matrix, we see only four instances of two late stage funds co-investing with one another.

Our visualization addresses our primary questions of how often different firms co-invest with one another and how they invest with firms that have similar or differing profiles. It also addresses how investments are distributed geographically and over different industry categories. The one question our visualization does not fully address is how older firms compare to younger firms,

however because the pace of investments and number of firms has increased so rapidly over the past ten years it would be difficult to make many valid conclusions and we felt it was more important to focus on illustrating how firms work together and the overall flow of capital. This visualization could be improved by allowing each of the visualizations to drive the other three visualizations. At the moment the co-occurrence matrix drives the entire visualization because our main goal was to show how firms work together but further insight could be gained by increasing the interactivity of the sub visualizations (for example allow clicking the heatmap to filter by state).