

STARTUPS VIS

Process Book



University of Utah

CS6630

Fall 2017

Joris Gahery, u1178638, u1178638@utah.edu

Lin Jia, u1091732, janicelin1103@gmail.com

Yaodong Zhao, u1054985, yaodong.zhao@utah.edu

GitHub Repository:

<https://github.com/LinJia00/2017-dataviscourse-project>

Website:

<https://linjia00.github.io/2017-dataviscourse-project/>

Table of Contents

1. BackGround Motivation and Overview	Error! Bookmark not defined.
2. Related Work.....	1
3. Questions	3
4. Data	4
5. Exploratory Data Analysis.....	5
6. Design Evolution.....	9
7. Implementation.....	12
8. Evaluation.....	13

1. Background Motivation and Overview

October 28

Angelis has a database of 342 thousand startups located in the United States. However, it doesn't provide us any visualization of it. So, our idea is visualizing these startups' features such as location, focused market, employee numbers, stage and total raised money. We believe that our project can not only extract fascinating insights and exciting visual effects but also it will be very useful for market analysis, job finding and recruiting.

2. Related Works

November 10

Here is the website and pictures to show the startups information but it does not show a very detailed visualization. It cannot show to the users the comparisons between the startups. And it cannot give users a directed visualization in every aspect.

<https://angel.co/information-technology>

Company	Joined	Followers	Signal	
The Hustle	Aug '16	62	...	VIEW JOBS
Patch Homes	Jul '16	50	...	VIEW JOBS
Brave	Jul '16	75	...	VIEW JOBS
Import2	Oct '12	202	...	

Here is a picture gives us an idea to show the stage of startups(last funding type).

<https://www.slideshare.net/ahmerk/business-model-analysis-and-market-research-for-startups-and-entrepreneurs>

STARTUP STAGE WISE APPLICATIONS

Clip slide

STARTUP DEVELOPMENT PHASES



3. Questions

October 28

- How is the trend of these companies growing? (For Users)
- what is the high chance market and location to get a job? (For Users)
- What is the best visualization for this dataset (For Ourselves)

We want to learn and accomplish: By this project, we hope to represent to users the information hidden in the abstract data with appropriate visualization. And, with editable parameters such as filters, we provide user a basic tool to explore the data. And We want to learn more ways to do better designs and how to let data show specifically to the users.

November 10

- How to make a connection with market type, company number and year, and how to control its scale?
- How to aggregate the huge amount companies to show what are the top markets?
- To find a better mark to represent employee number.

4. Data

a. Data source: <https://www.crunchbase.com>

b. Data Processing:

Our original dataset just likes this.

ID	Organization	Organization	Categories	Headquarters Location	Description	CB Rank (Company)	Founded Date	Funding Status	Company Type	Total Funding Amount	Number of Employees
1	/organization/Baremetrics	Business	Intelligence, Data Visualization, Cloud Computing	Birmingham, Alabama, United States	Untee Subscription Analytics and	15,865	14-Oct-13	Seed	For Profit	\$800,000	3
2	/organization/Chrono.gg	PC Games		Alabama, United States, North	Untee Chrono.gg was founded by	16,398	2015	Seed	For Profit	\$1,325,000	2
3	/organization/ComplyMD	Service	Industry, Document Management, Health Care	Birmingham, Alabama, United States	Untee ComplyMD develops and	45,245	2004	Seed	For Profit	\$1,130,000	1
4	/organization/Curse	Video Games, Digital Media, Gaming		Huntsville, Alabama, United States	Untee Curse is a global multimed	3,117	2006	M&A	For Profit	\$58,000,000	4
5	/organization/Fleetio	Automotive, Fleet Management, Software		Birmingham, Alabama, United States	Untee Fleetio is a web-based fle	20,723	1-May-11	Seed	For Profit	\$750,000	1
6	/organization/GLOW	Beauty, Marketplace, Mobile Apps		Birmingham, Alabama, United States	Untee GLOW is a beauty on dem	23,470	Sep-15	Seed	For Profit	\$320,000	2
7	/organization/Graphenics	Nanotechnology, Industrial, Plastics and Rubber Manufacturing		Tuscaloosa, Alabama, United States	Untee Graphenics is a woman-ov	55,697	26-Sep-12	Seed	For Profit	\$20,000	1
8	/organization/IllumiCare	Health Care		Birmingham, Alabama, United States	Untee IllumiCare Smart Ribbons	21,195	2014	Early Stage Venture	For Profit	\$5,760,000	2
9	/organization/KloutCatch	Search Engine		Birmingham, Alabama, United States	Untee Klout Catch is a web reser	71,754	2011	Seed	For Profit	\$40,000	1
10	/organization/Likely!	Social Media Marketing, Photo Sharing		Birmingham, Alabama, United States	Untee We make images successf	25,293	1-Aug-16	Seed	For Profit	\$50,000	1
11	/organization/ModularTec	Document Management, Software		Huntsville, Alabama, United States	Untee Modular Technologies is an information managem	2000	M&A	For Profit	\$7,000,000	3	
12	/organization/MyAcademic	Big Data, Software, Education, Information Technology, Apps		Opelika, Alabama, United States	Untee MyAcademic Big Data, Software, Education, Information Technology, Apps	59,531	2-May-11	Seed	For Profit	\$450,000	2
13	/organization/NitroPCR	Mobile, mHealth, Information Technology, Health Care		Alabama, United States, North	Untee NitroPCR is mobile patient	60,552	2014	Seed	For Profit	\$205,000	2
14	/organization/NotIT	Gamification, Apps		Birmingham, Alabama, United States	Untee Not IT is an online social n	71,724	24-Apr-12	Seed	For Profit	\$55,000	1
15	/organization/Platyp1	Mobile, Social Media Management, Web Design, Web Development		Birmingham, Alabama, United States	Untee Platyp1 is a digital marketi	36,978	1-Dec-13	Seed	For Profit	\$2,000,000	1
16	/organization/PublicVine	Telecommunications, PaaS, Video, Video Streaming, Cloud		Florence, Alabama, United States	Untee PublicVine is a PaaS provider w	28,768	2009	Seed	For Profit	\$3,000,000	1
17	/organization/Sarkitech	Smart Agriculture, Medical		Valhermoso Springs, Alabama, United States	Untee Sarkitech Sensors designs	59,239	2011	Seed	For Profit	\$430,000	1
18	/organization/Sequel	Youth Service Industry, Hospitality, Family		Huntsville, Alabama, United States	Untee Sequel Youth and Family S	21,053	1999	Private Equity	For Profit	\$66,000,000	2
19	/organization/Shift	Delivery, Grocery, E-Commerce		Birmingham, Alabama, United States	Untee Shift is a grocery delivery	2,056	2014	Early Stage Venture	For Profit	\$65,200,000	3
20	/organization/SiliconLabor	Hardware, Software		Huntsville, Alabama, United States	Untee Silicon Laboratories is an i	53,097	1996	IPO	For Profit	\$12,750,039	2
21	/organization/SwellFunds	Charity, Events, Hedge Funds, Non Profit, Financial Services		Birmingham, Alabama, United States	Untee SwellFunds provides	43,195	12-Nov-12	Seed	For Profit	\$500,000	1
22	/organization/TicketBiscuit	Software, Ticketing		Birmingham, Alabama, United States	Untee TicketBiscuit is a Birmingh	5,316	2001	M&A	For Profit	\$5,000,000	2
23	/organization/transOMIC	Biotechnology, Medical, Pharmaceutical		Huntsville, Alabama, United States	Untee transOMIC is a biotechnol	67,507	2012	Seed	For Profit	\$20,000	1
24	/organization/Tutem	Home Renovation, Internet		America, Alabama, United States	Untee Tutem is a leading on-demand platfo	57,208	1-Nov-14	Seed	For Profit	\$100,000	1

The original data has some important attributes includes:

- Organization: company name
- Categories: the company's industry
- Headquarters Location: company located in which state and city
- Founded Date
- Funding Status
- Total Funding Amount
- Number of Employees

November 10

For our project design, we separate and make them more clearly for each chart.

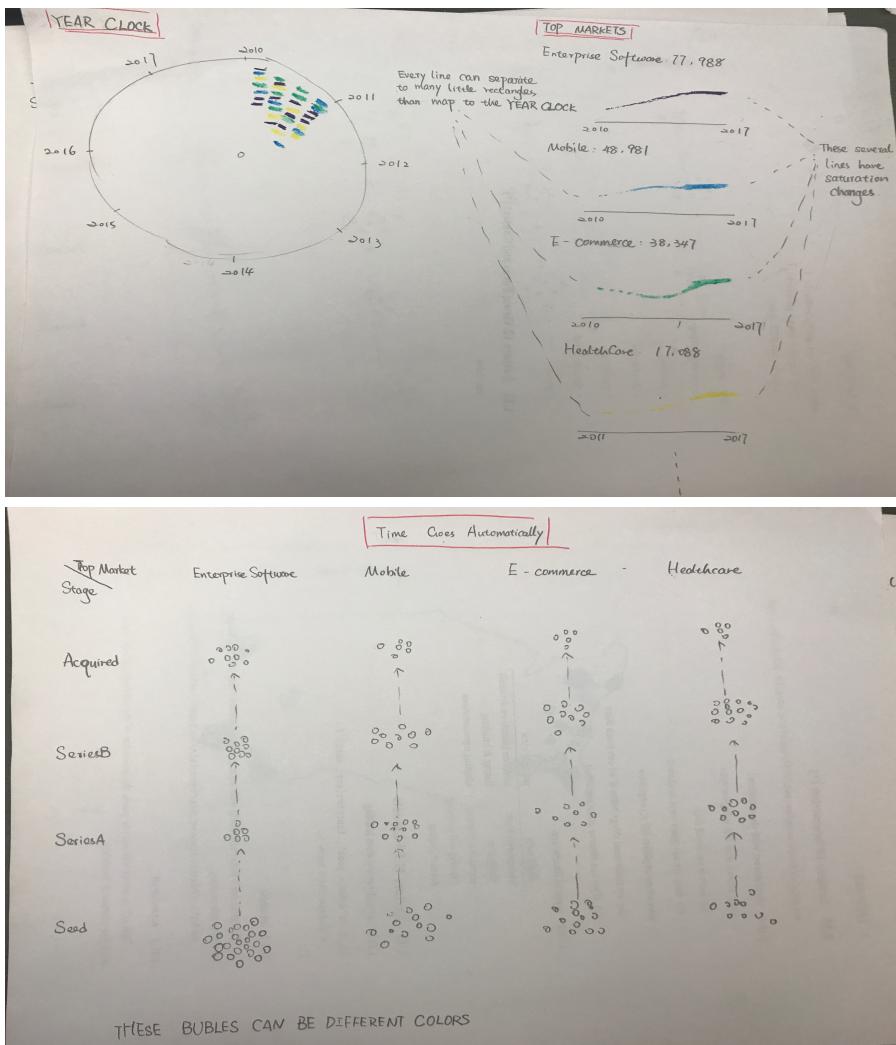
There are six csv file for our whole design:

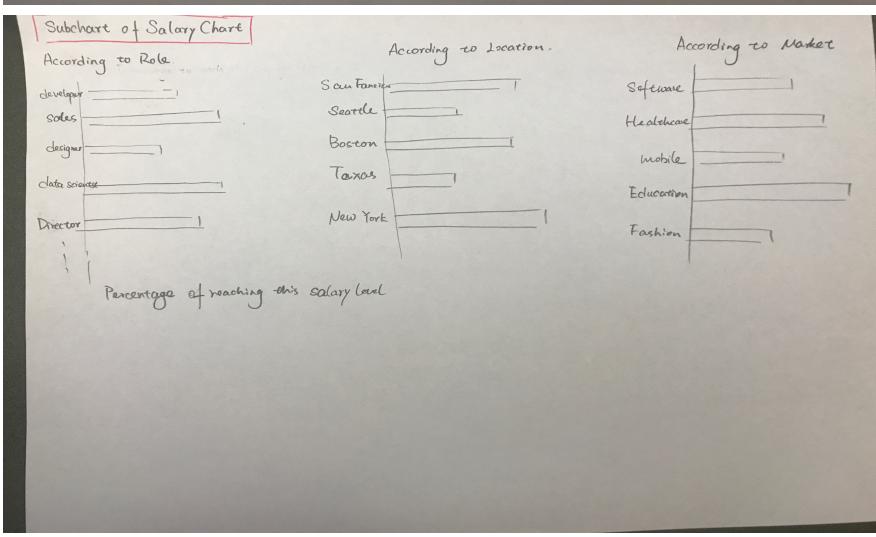
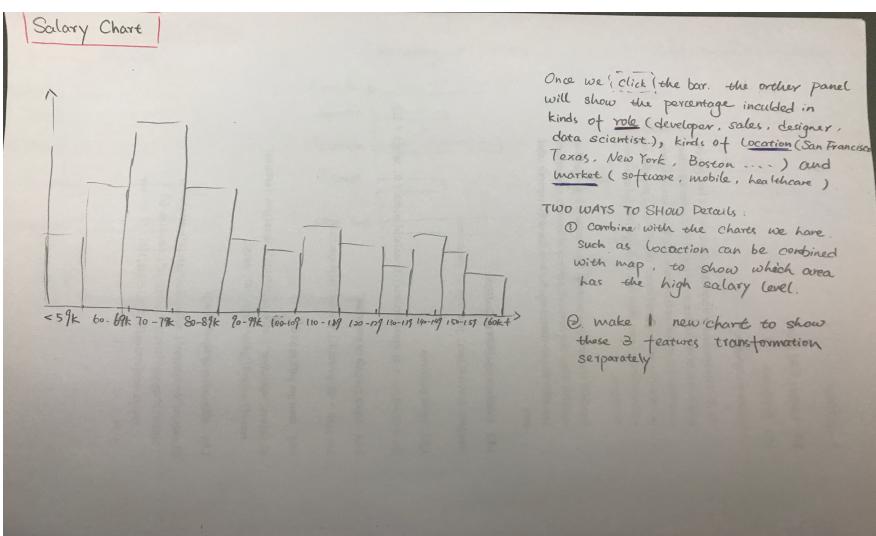
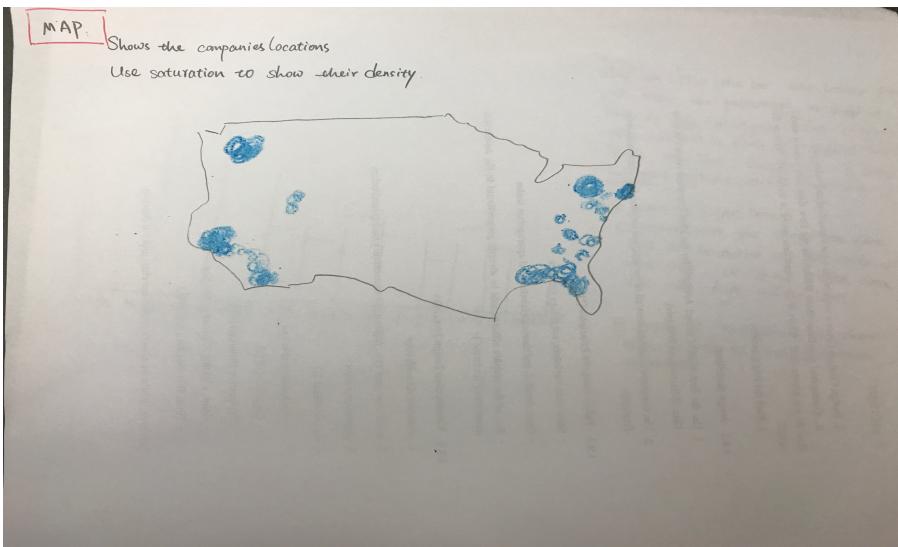
- count-by-city-market.csv
- count-by-city.csv
- count-by-market.csv
- count-by-year-market.csv
- employee-by-market.csv
- employee-by-state.csv

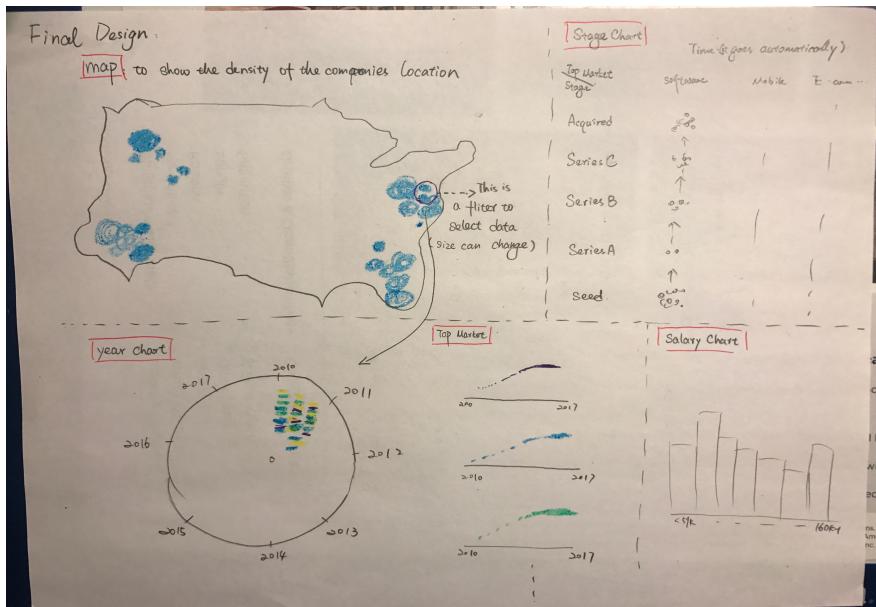
So, it will be easily to handle the data for each specific chart.

5. Exploratory Data Analysis

October 28





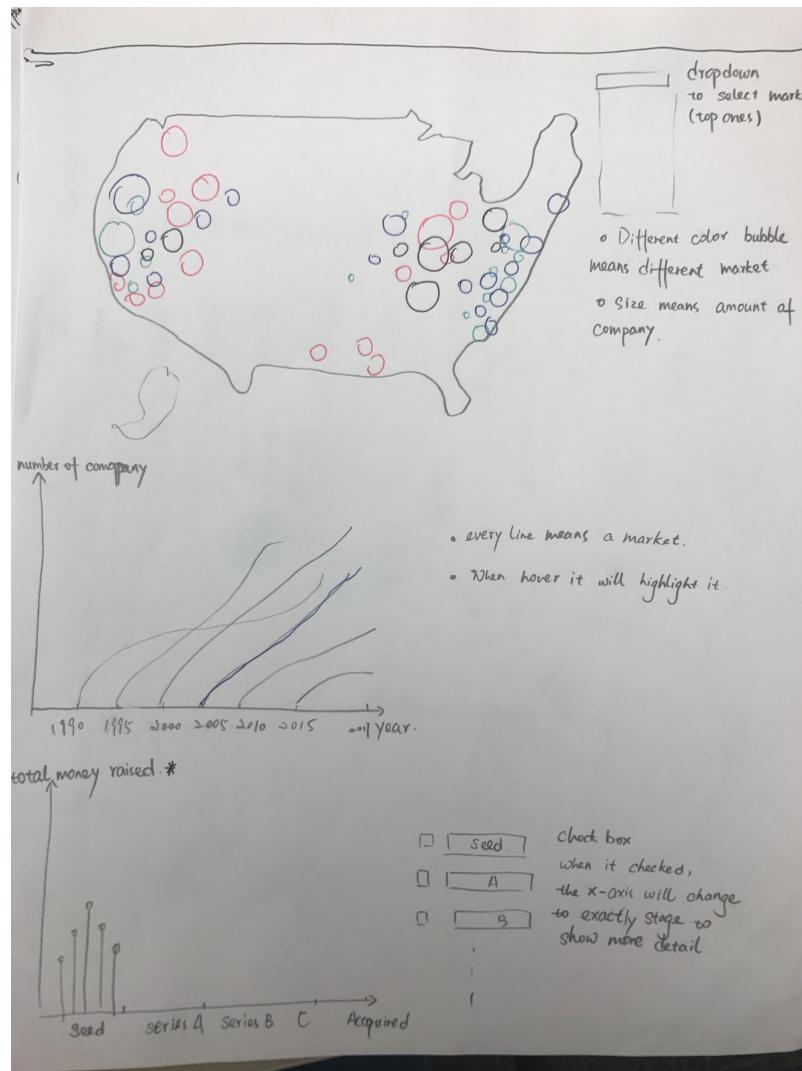


Above are our initial design. Even though our design is attractive and innovative, but at that time we haven't seen the whole dataset. The real dataset we scraped from the website has some difference with our imagine. So, we make some improvements.

6. Design Evolution

November 10

a. Redesign:

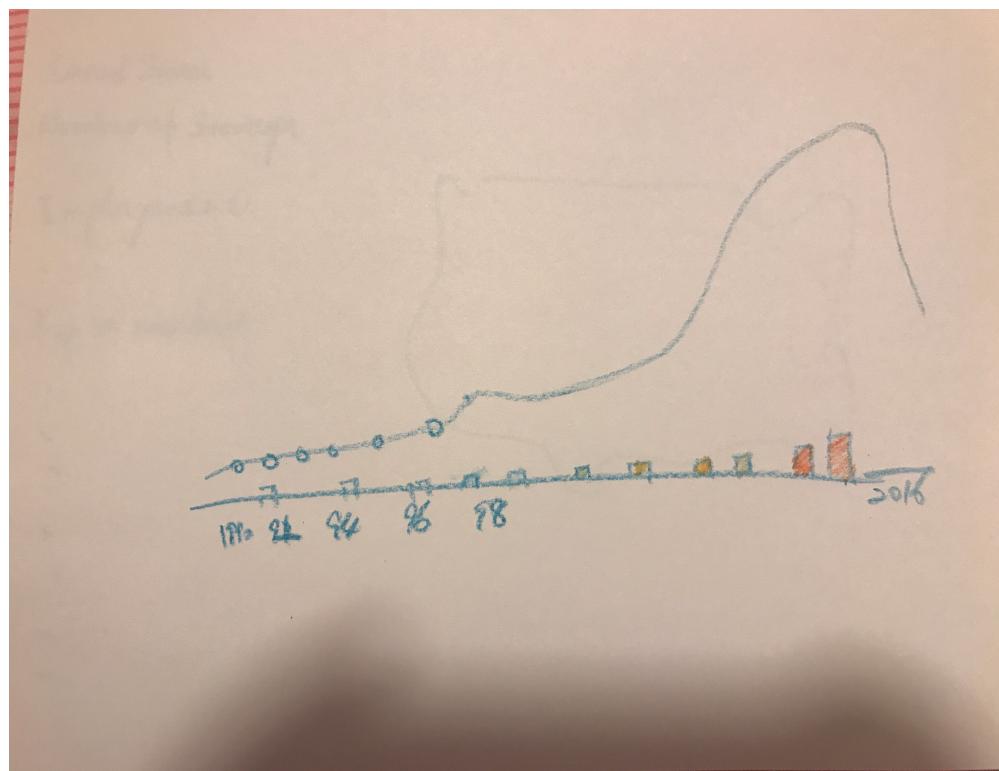
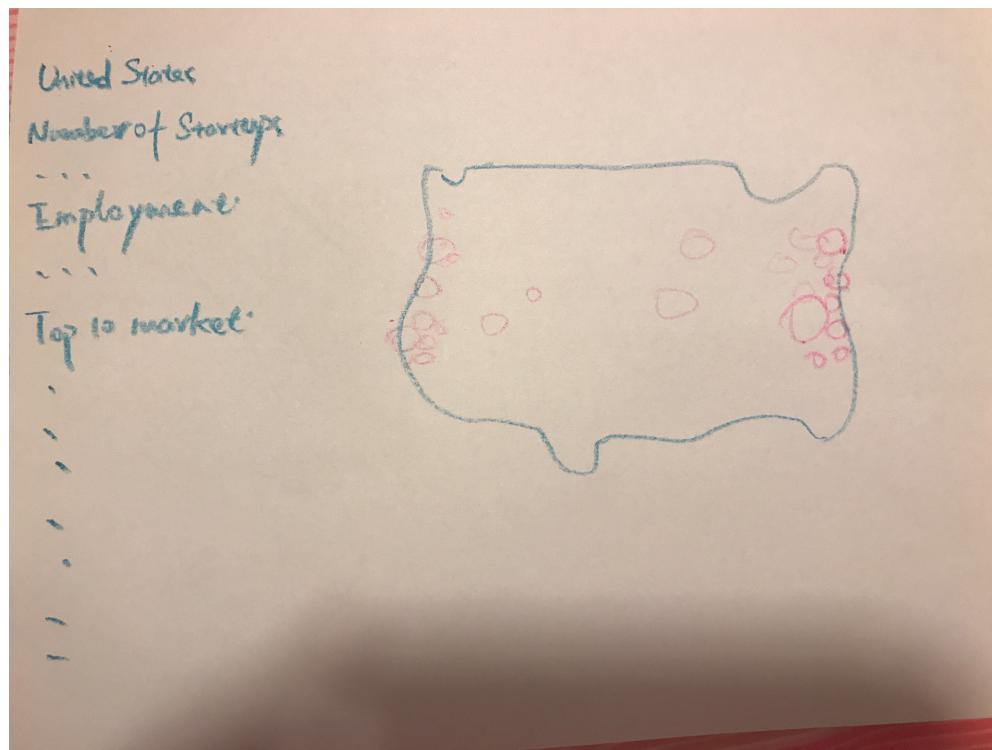


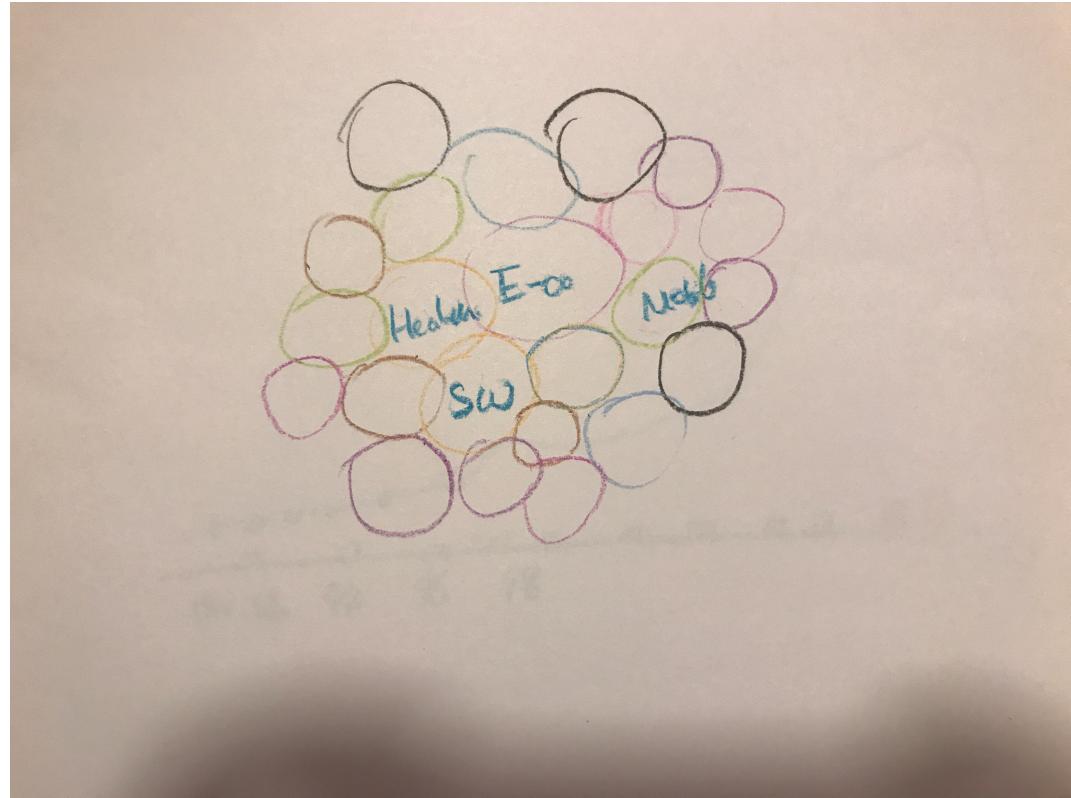
b. Description:

We have encountered some problem when we processing the dataset, so we changed some part of our design. But we will continue to improve the design to show the dataset and reach a better result.

November 25

a. Redesign

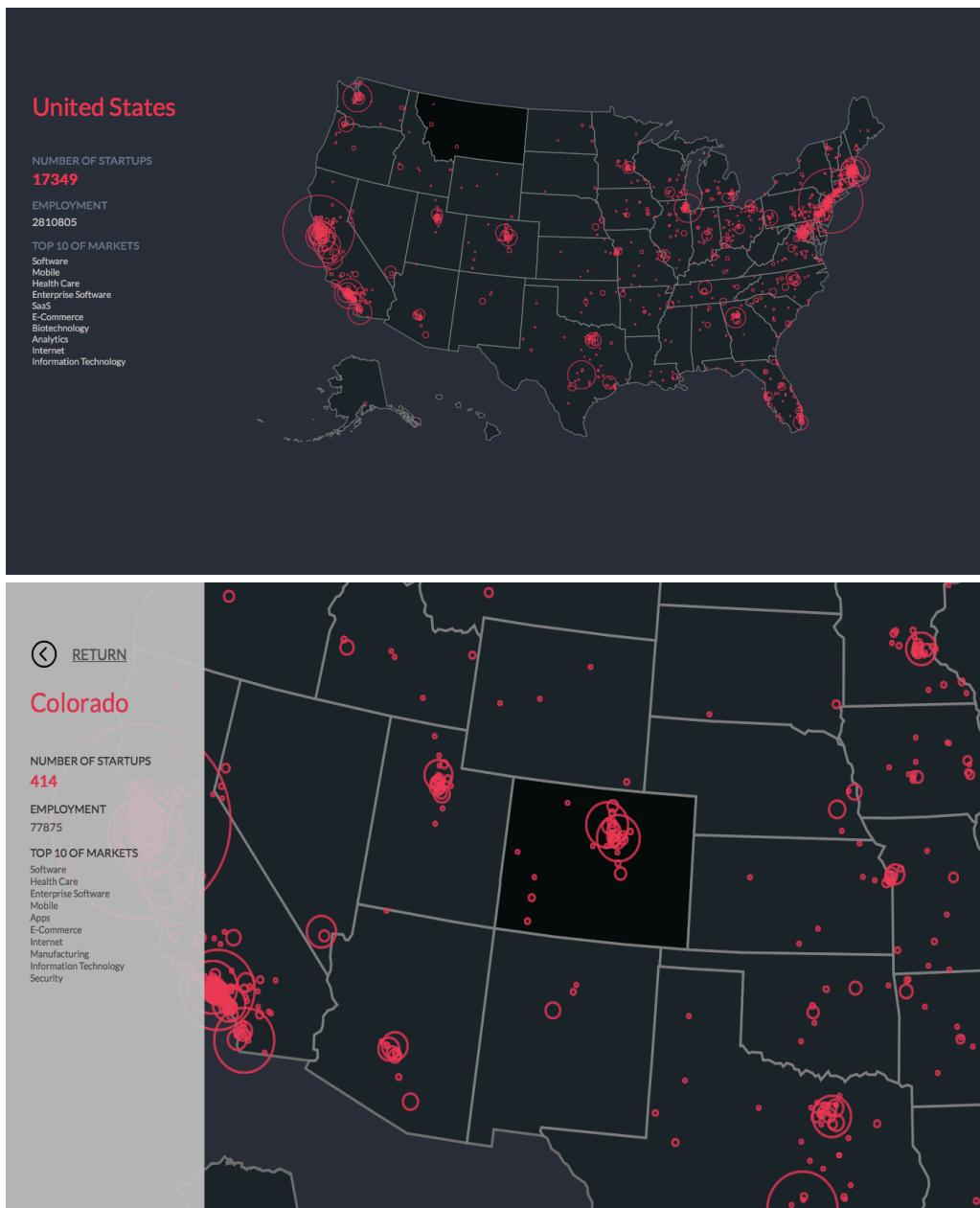


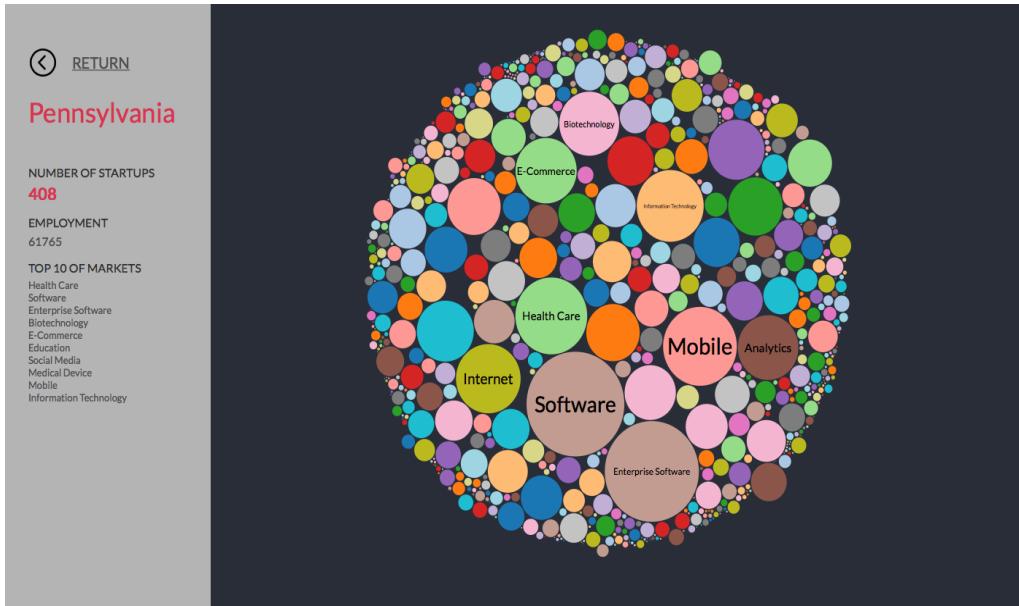
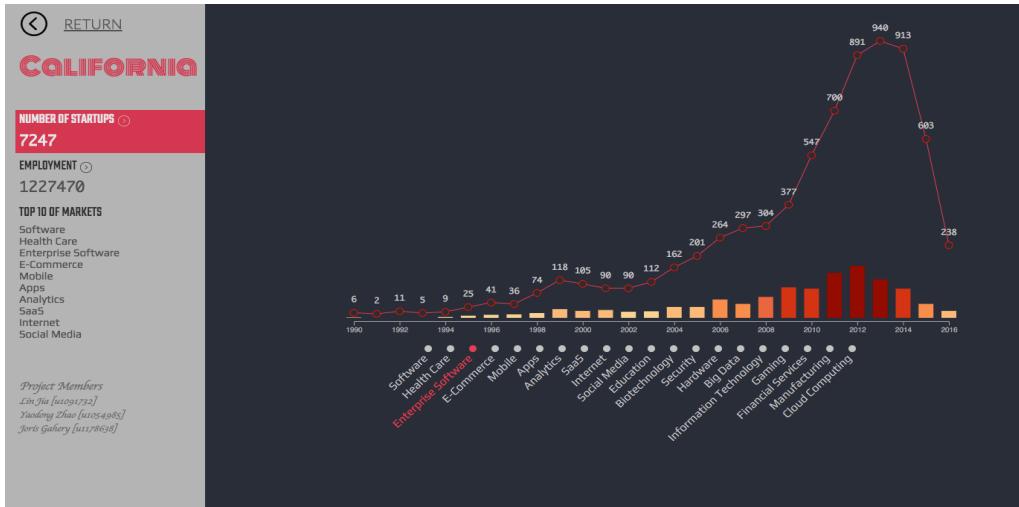


b. Description

We redesign our pages based on our dataset. The main page is the map and the information panel on the left side can not only show the information but also as a role of navigation to the others page. And the employment line chart can show the amount trend of startups in one area from 1990 to 2016. The bubble chart shows which industry has bigger ownership in whole market and which has smaller.

7. Final Implementation





8. Evaluation

- Based on our visualization, from the map, we can directly know the area which has lots of startups that means more opportunity in that area. But the disadvantages are the overlaps of the circles and the small area is hard to select. We choose this as the final design because the effect around the whole pages is better than the other. For the startups-number chart (line chart), we can see the trend both base on the market and the time which can show us the market grows trend. It is easy for user to see which market is boom and its trend in recent years. And the employment chart (bubble chart) show the ownership of each industry in the whole market.
- We answered all the questions except to show the employee numbers.
- It has good effect for the visualization but because the noisy in the dataset we cannot show a better design. Not all the startups in the dataset has the whole feature list as our expected. For example, some companies have the total founded record but some not, some companies have the exactly

start date but some not, etc. At first, our design is innovative but the real data is not like our imagination. So, if we find some more integrated data, the visualization will be better.

Thank You!