

Meetups Data Visualization

Introduction

Meetup is a social networking platform that targets to brings people with similar interest together to participate organized events. Meetup allows members to find and join groups unified by a common interest. As of 2017, there are 32 million users with 280 thousand groups available across 182 countries. The source of the dataset found for this project is from Kaggle (<https://www.kaggle.com/sirpunch/meetups-data-from-meetupcom>) The author collected the data using Meetup API. Python script was used to ping meetup API and collect responses as JSON object. Logical chunks of data were explored and saved as csv files. After the author has collected the data, the data was filtered to include only 3 cities' information (New York, Chicago, San Francisco), which are major cities from East, Central and West of United States.

The reason I chose to use the datasets is because we have attended a lot of meetups in San Francisco ever since we started our MSDS program. It is interesting for me to know from a bigger picture that what the trendy and popular events and topics currently are in these three major cities.

Below listed are the collected dataset in csv format and followed by a Database EER diagram (Figure 1) that shows the relationships among the collected csv files.

- categories.csv
- cities.csv
- events.csv
- group_topics.csv
- groups.csv
- members_topics.csv
- members.csv
- topics.csv
- venues.csv

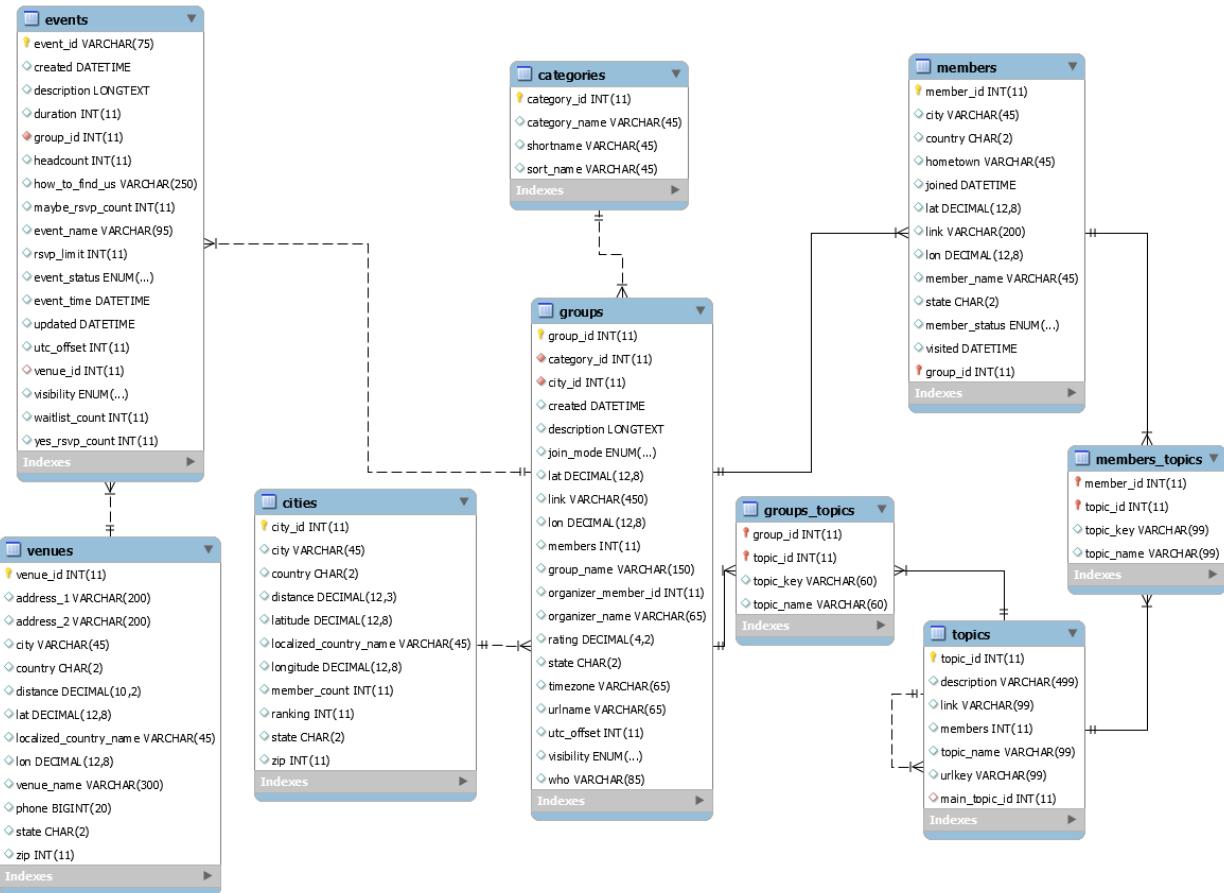


Figure 1 Database EER Diagram

Summary of Data

In this section, I will use visualizations from Tableau to summarize the data with a multitude of different plots ranging from the following types:

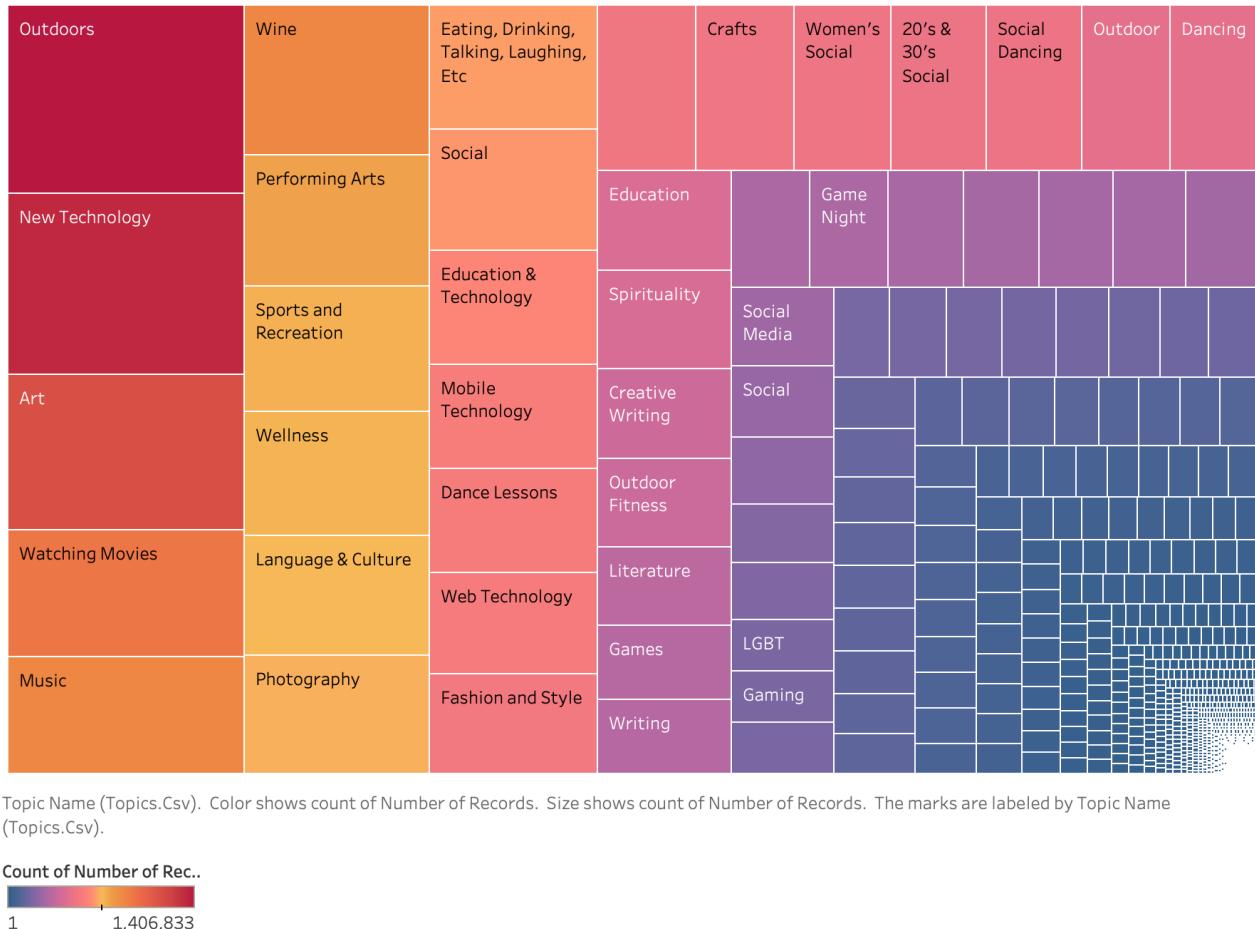
- Histogram
- Bar plot
- Scatterplot
- Bubble Map
- Choropleth Map
- Connection Map
- Heat Map
- Stacked area/Stream graph
- Treemapping
- An interactive plot



Figure 2 Number of records per city level

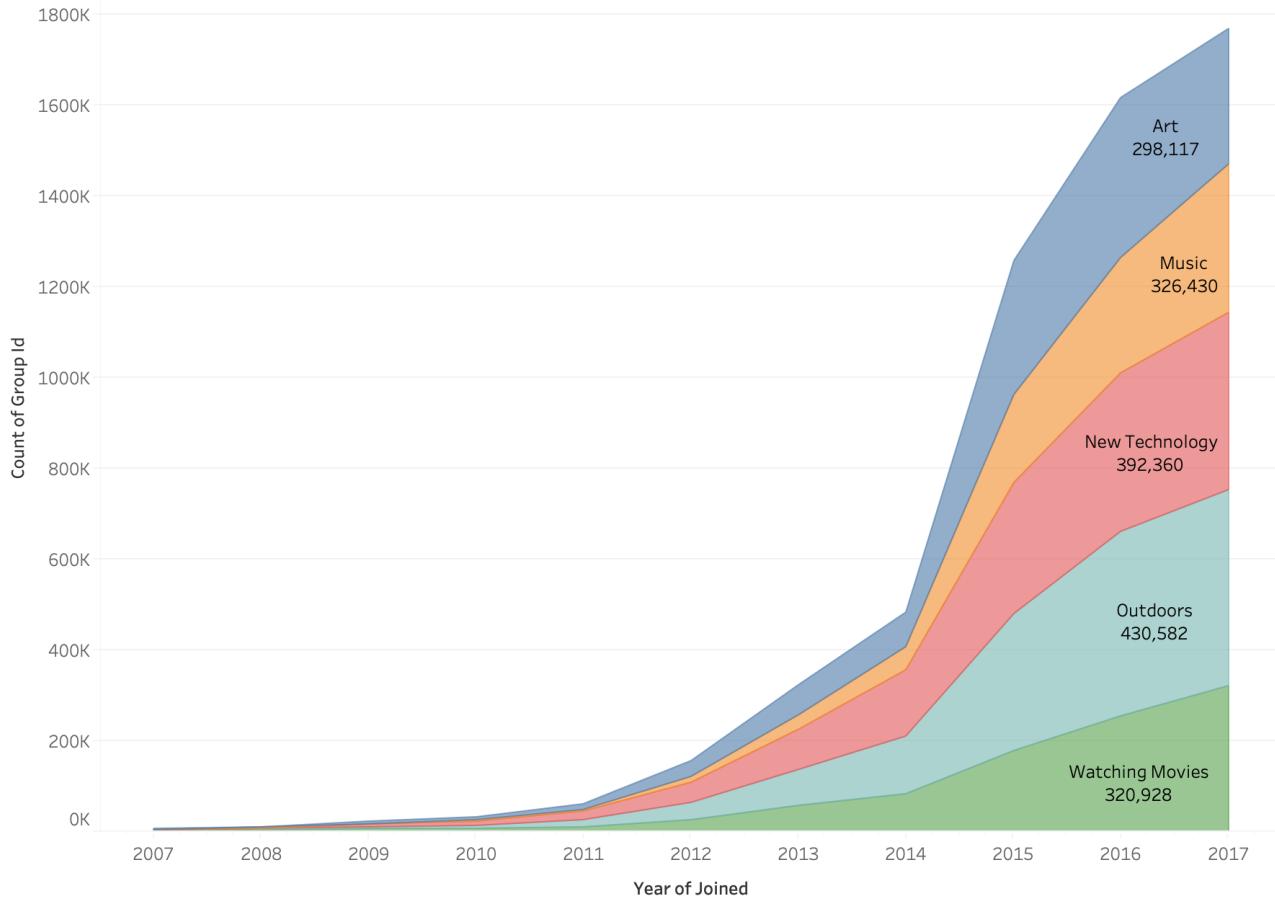
In this dataset, the available cities with its number of records counts are plotted as above in Figure 2. Hence, we will be only looking at the three major cities, which are New York, Chicago and San Francisco in the visualization work in this project.

Popularity of Topics



In Figure 3, I used Treemapping to plot the count of members the topics on meetup. I joined three csv files, members_topics.csv, members.csv and topics.csv together using the first csv file as the bridging file which has two columns one the same as in members.csv and the other one the same as in topics.csv. I counted the number of members to each of the topic level and visualized it as above. We can see that the top 5 most popular topics on meetup is 1) Outdoors, 2) New Technology, 3) Art, 4) Watching Movies and 5) Music.

The top 5 Topic trends



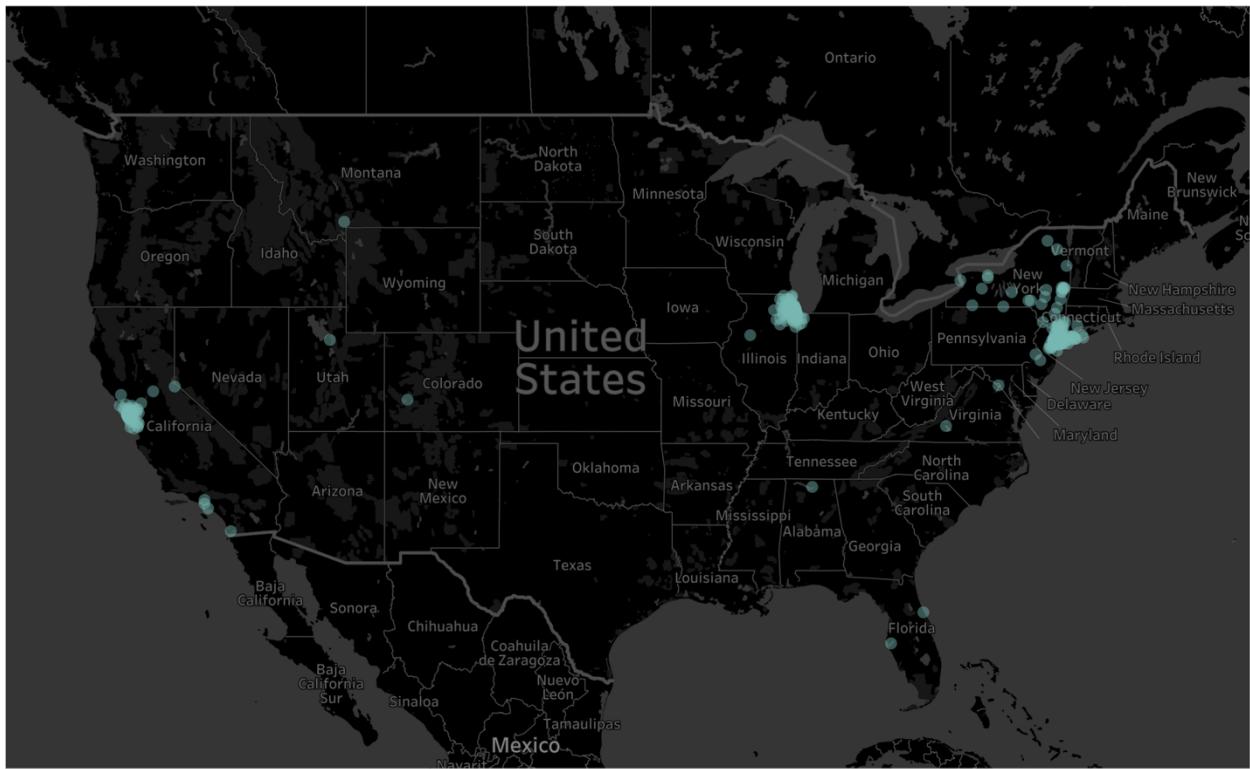
The plot of count of Group Id for Joined Year. Color shows details about Topic Name (Topics.Csv). The marks are labeled by Topic Name (Topics.Csv) and count of Group Id. The data is filtered on Joined Year, which keeps 11 members. The view is filtered on Topic Name (Topics.Csv), which keeps Art, Music, New Technology, Outdoors and Watching Movies.

- Topic Name (Topics.Csv)**
- Art
 - Music
 - New Technology
 - Outdoors
 - Watching Movies

Figure 4 Top Five topic trends

Since I already have the popular topics from Figure 3, I wanted to deep dive into the top 5 most popular topics and plot their trends along the time series data from 2007 to 2017 (latest 10 years that are available in the datasets). As you can see from the above plot, almost all 5 topics began to grow significantly starting from 2014 to 2017. Next, I would like to explore the csv file of venue.

Venue Locations

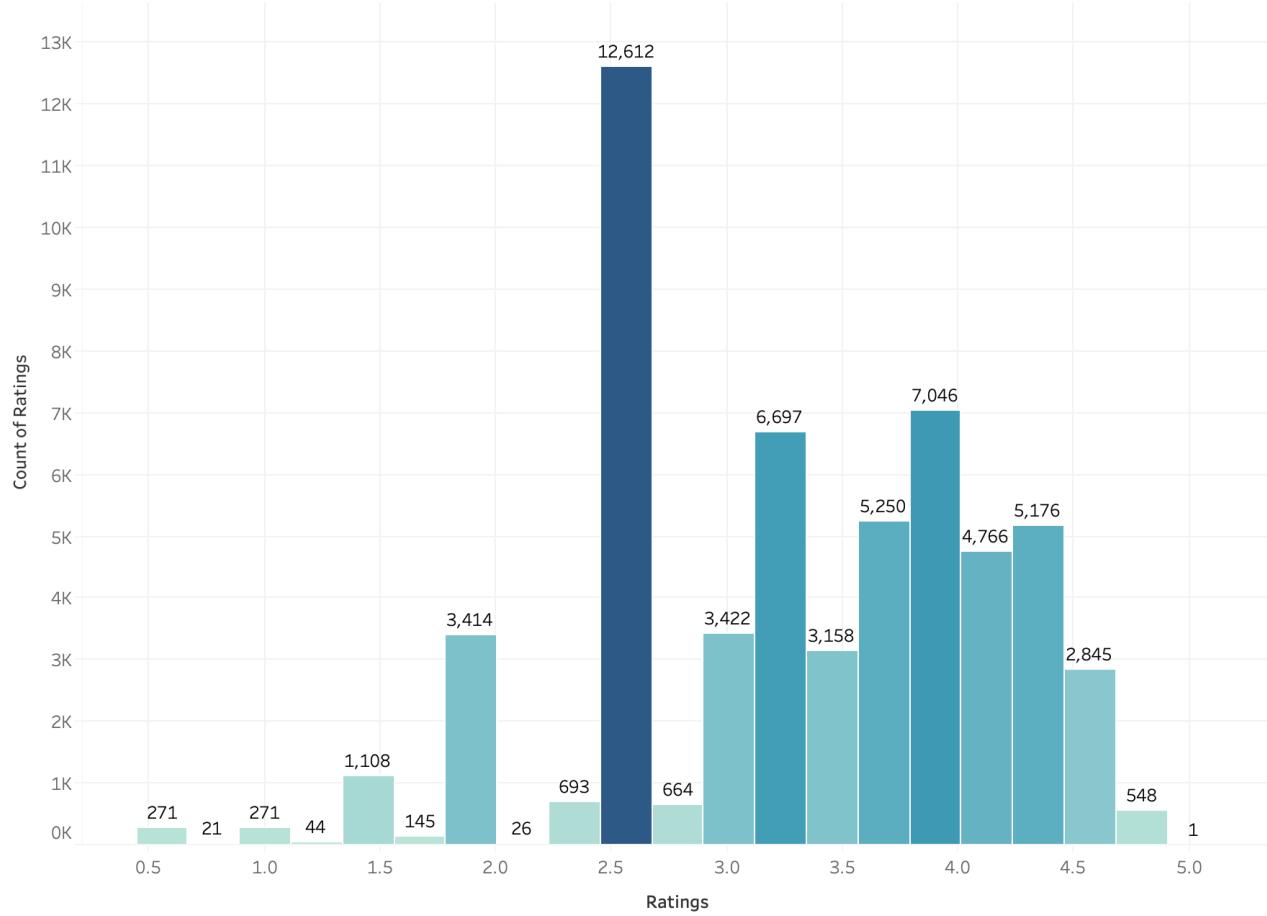


Map based on Longitude (generated) and Latitude (generated). Details are shown for Zip.

Figure 5 Venue Locations

In the venues.csv file, the zip code of each of the events venues are recorded. I used the zip code to map them onto the US map layer. Since this dataset is only about three cities: New York, Chicago, San Francisco. We can see from the above plot that most of the dots are gathering around these three cities. However, there are also minor zip code appearing in the States of Florida, Utah, Virginia, Montana, Alabama and so on. From here I would like to further understand that what are the rating distributions for the venues and how are these ratings distributed in the city areas of each of the major cities (New York, Chicago, San Francisco).

Histogram for Venue Ratings (Normalised)



The trend of count of Normalised Rating for Normalised Rating (bin). Color shows count of Normalised Rating. The marks are labeled by count of Normalised Rating. The data is filtered on Normalised Rating (bin), which excludes 0.0.



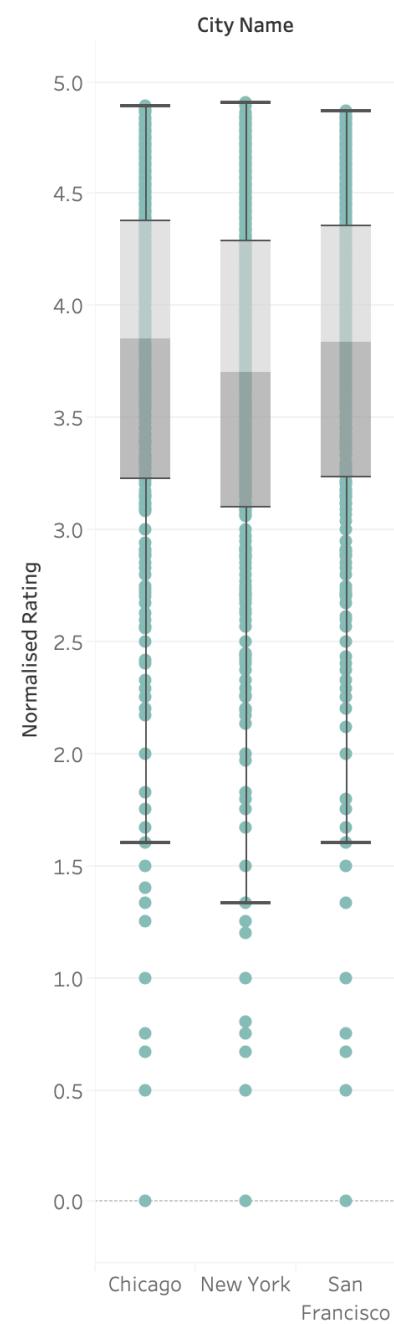
Figure 6 Distribution of Venue Ratings

Figure 6 shows that ratings around 3.0 to 4.5 has a bell curve, which is a normal distribution.

However, rating 2.5 is having the greatest number of counts. The histogram only gives us a high level of rating distributions. Now I want to deep dive into the three cities and see how the city areas are rated with the venues within that area.

Venue Ratings

Distribution



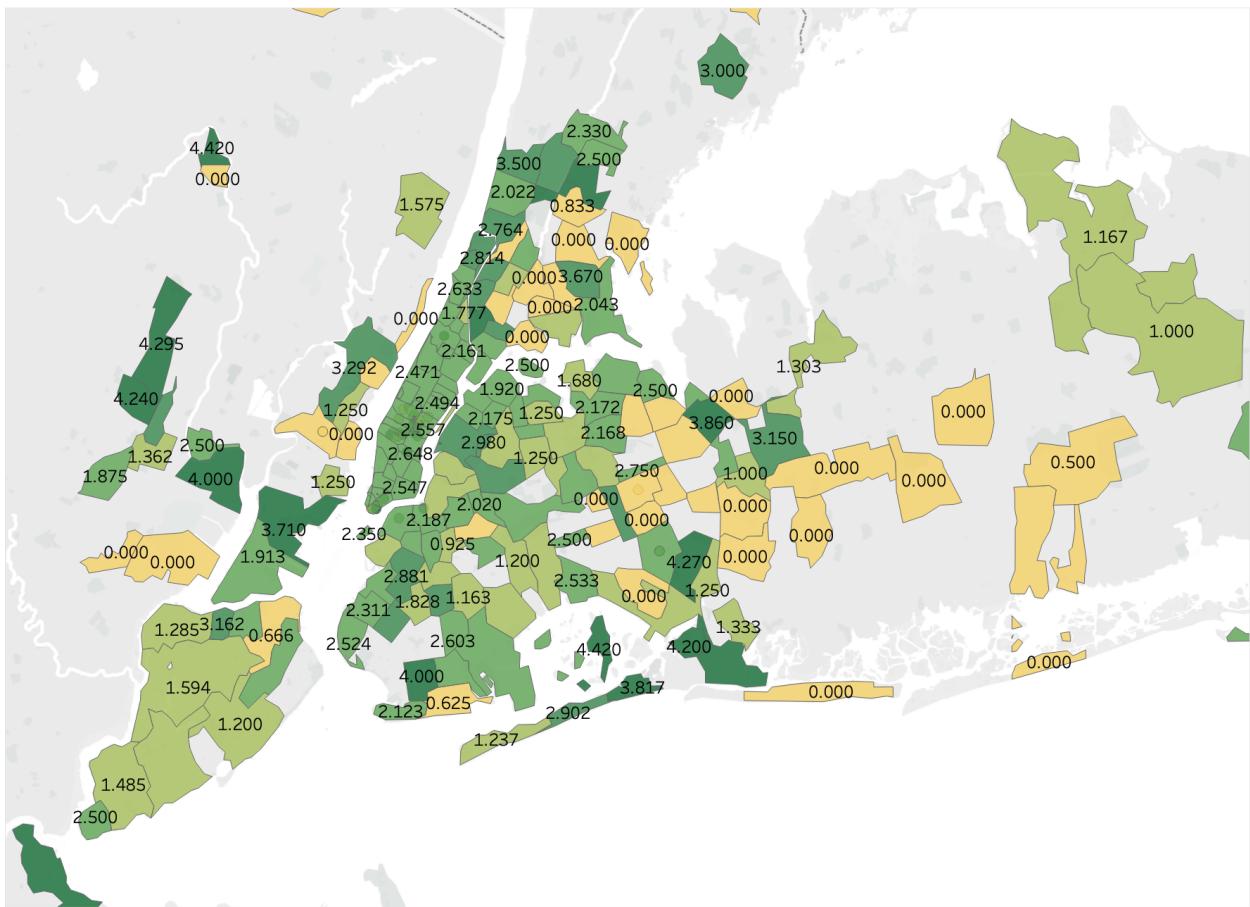
Normalised Rating for each City Name.
The view is filtered on City Name, which keeps San Francisco, Chicago and New York.

Figure 7 Average Venue Ratings

In Figure 7, we can see how the normalized ratings distributions for the three cities of Chicago, New York and San Francisco are. For Chicago and San Francisco, the median ratings are both around 3.8, while for New York City, the median rating is slightly smaller, around 3.6. Although the boxplot also summarizes the 25th and 75th Percentile in the diagram (the dark and light gray boxes), it is still a high-level summary for the distribution of the ratings to the venue. Hence, I would like to further explore the venue ratings mapping to the different zip code areas to each of the three cities.

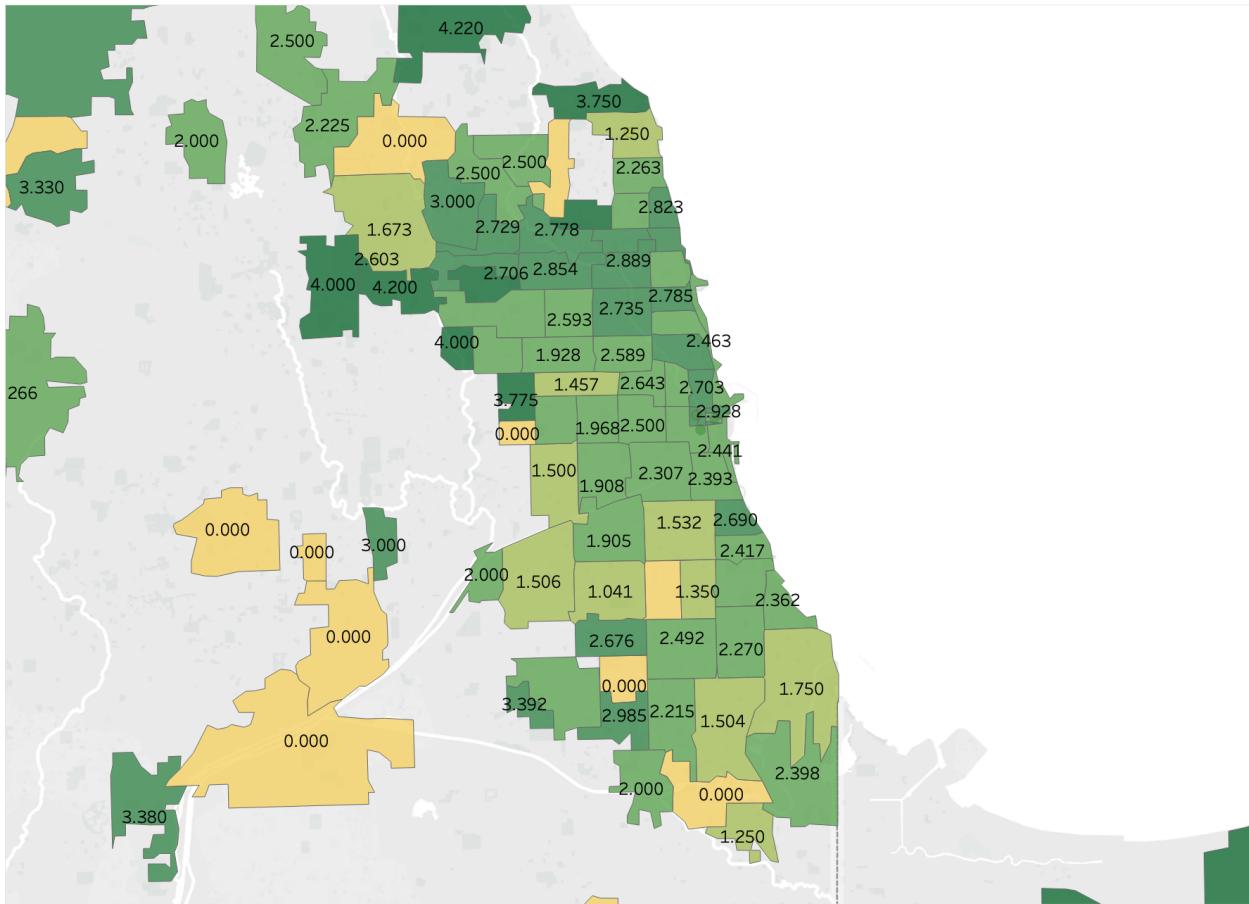
The following three plots, Figure 8, Figure 9, and Figure 10 shows the Average venue rating per Zip code breakdown and mapped it to the city areas.

Average Venue Ratings per Zip code in New York Area

*Figure 8 New York Rating Distributions*

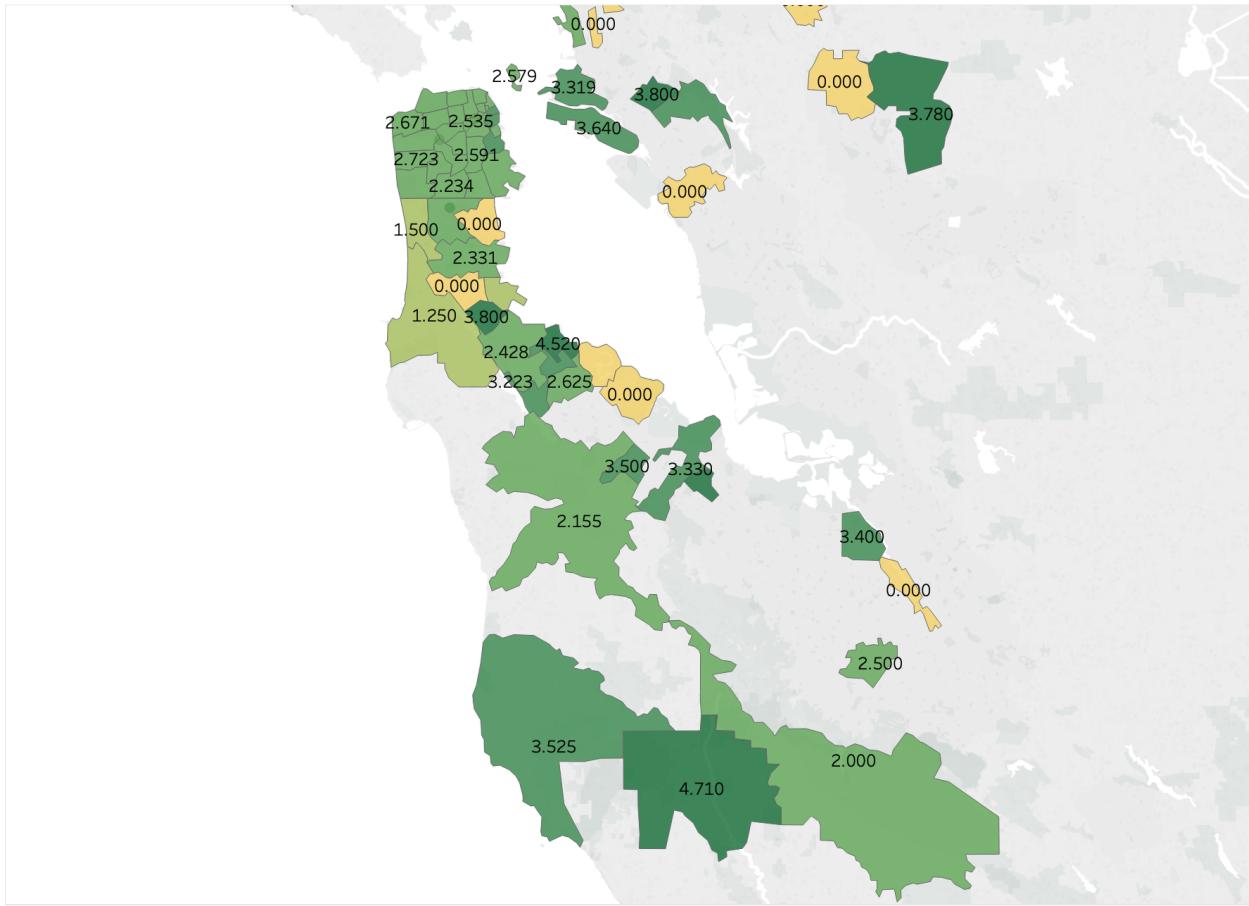
As we can see from Figure 8, most of the city areas in New York city are having average rating around 2.5. Some of the suburban areas have ratings more than 4.0. Other outside areas have even lower ratings (smaller than 2.0 and even 0.0)

Average Venue Ratings per Zip code in Chicago Area

*Figure 9 Chicago Rating Distributions*

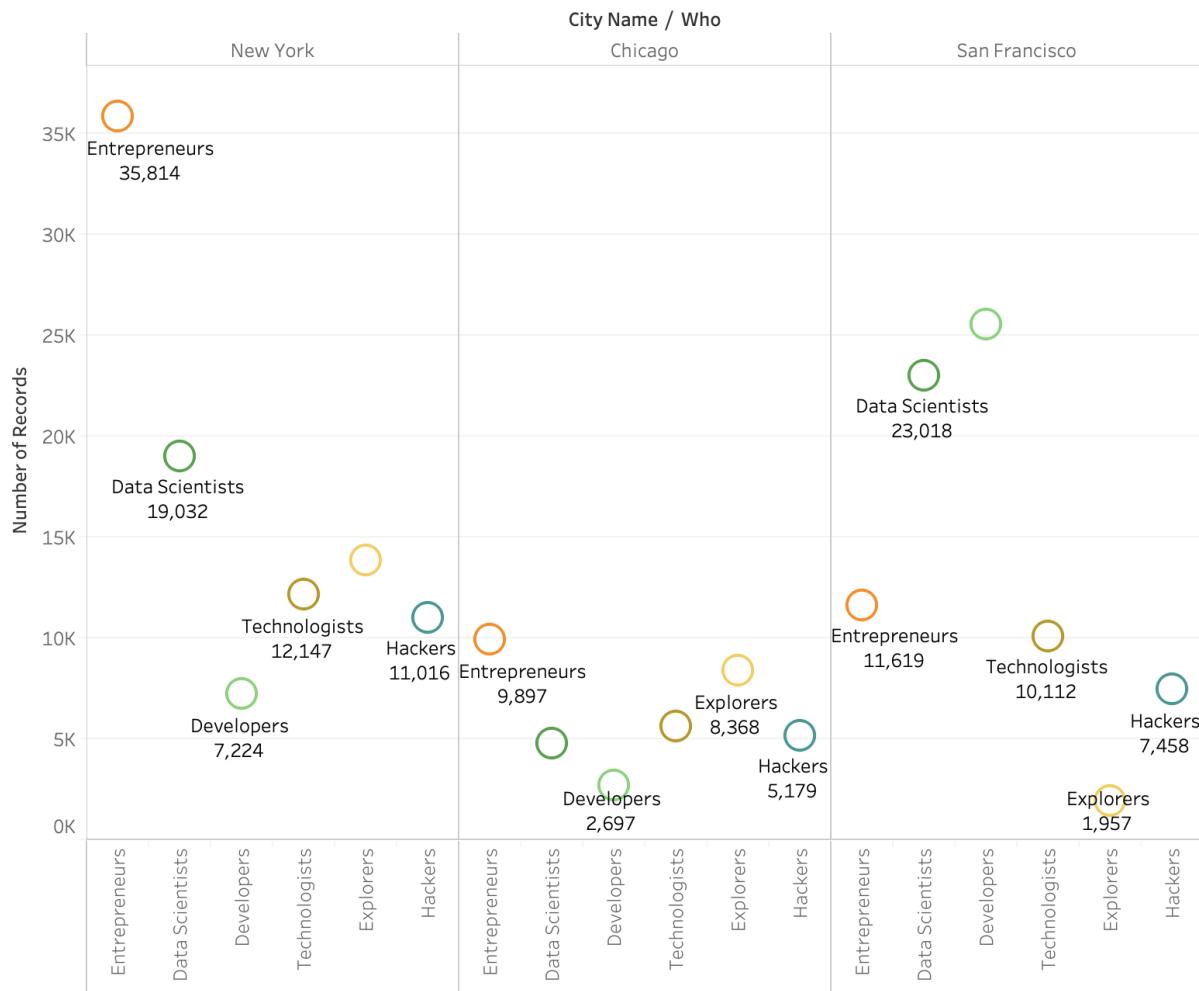
Similarly, in Figure 9, most of the city areas of Chicago are having average rating around 2.7. Some of the suburban areas have ratings more than 4.0. Other outside areas have even lower ratings (smaller than 2.0 and even 0.0)

Average Venue Ratings per Zip code in San Francisco Area

*Figure 10 San Francisco Rating Distributions*

For San Francisco area, as we can see from Figure 10, the same applies in the SF downtown that most of the city areas of San Francisco are having average rating around 2.5. However, Daly City and South San Francisco are having ratings around 1.5 to 1.7. More interestingly, South Bay area such as Palo Alto, Mountain View are having a much higher rating (3.5 and 4.7). The situation is because Palo Alto and Mountain View Areas have many innovative technology companies that provides good venue for these events. Daly City and South San Francisco Areas are mostly residential areas which does not have many venues suitable for events.

Who is on Meetup



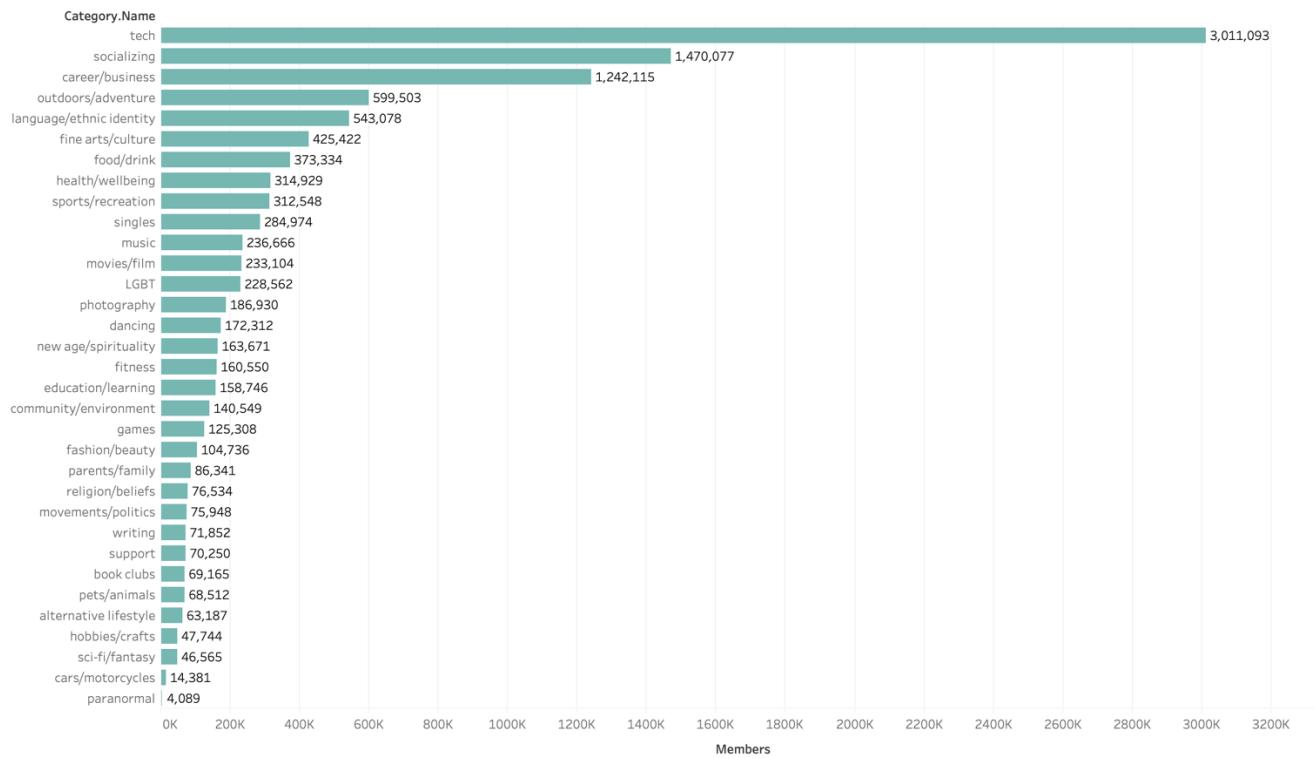
Sum of Number of Records for each Who broken down by City Name. Color shows details about Who. The marks are labeled by Who and sum of Number of Records. The view is filtered on Who, which has multiple members selected.

- Who**
- Orange: Entrepreneurs
 - Green: Data Scientists
 - Light Green: Developers
 - Yellow: Technologists
 - Light Yellow: Explorers
 - Teal: Hackers

Figure 11 Who is on Meetup

Figure 11 shows the identification of people who joined meetup. Interestingly but not surprisingly, New York is having the most of Entrepreneurs while San Francisco is having the most of Developers and Data Scientists. For Chicago, the Entrepreneurs are also its largest number but much smaller comparing to New York's number.

Number of members by categories



Sum of Members for each Category.Name. The marks are labeled by sum of Members.

Figure 12 Number of members by categories

Figure 12 shows the summary of number of members by categories. As we can see tech category is having the largest number of members, almost double the number of the second largest counts category, which is socializing.

The top 10 popular groups in New York

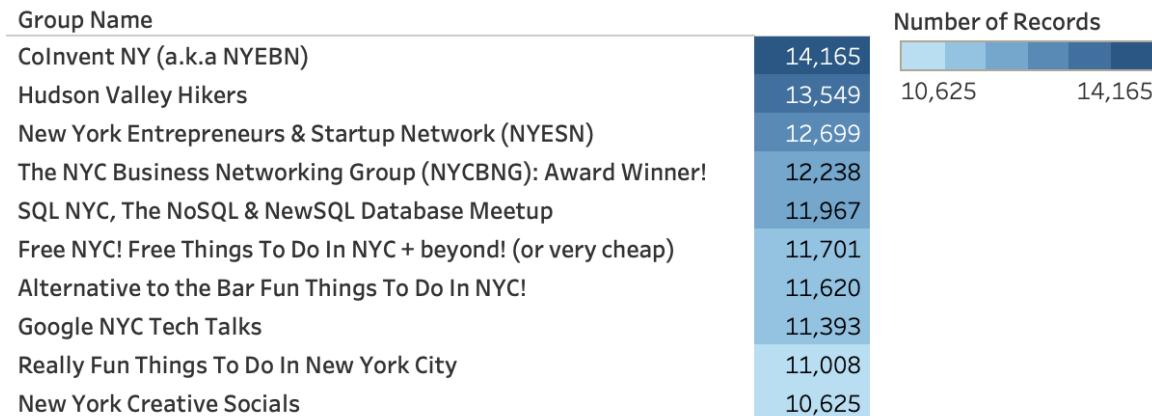


Figure 13 Top 10 popular groups in New York

The top 10 popular groups in Chicago

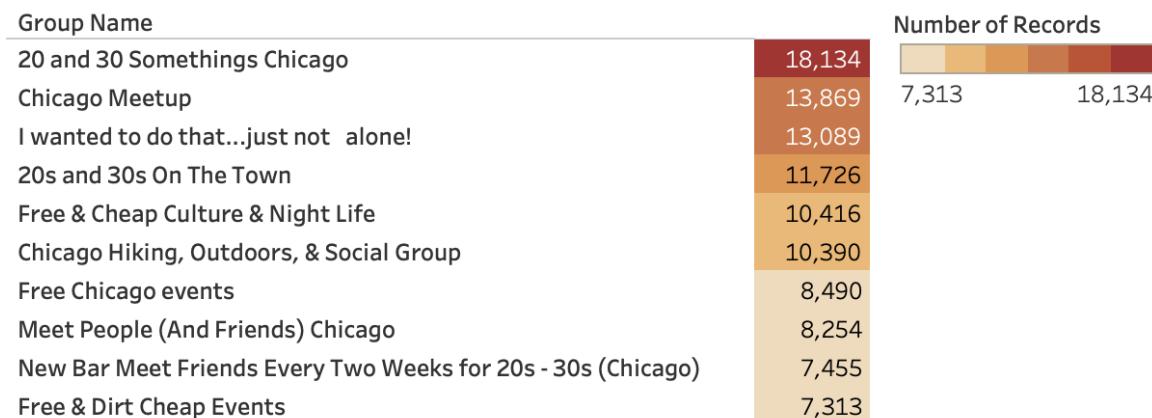


Figure 14 Top 10 popular groups in Chicago

The top 10 popular groups in San Francisco

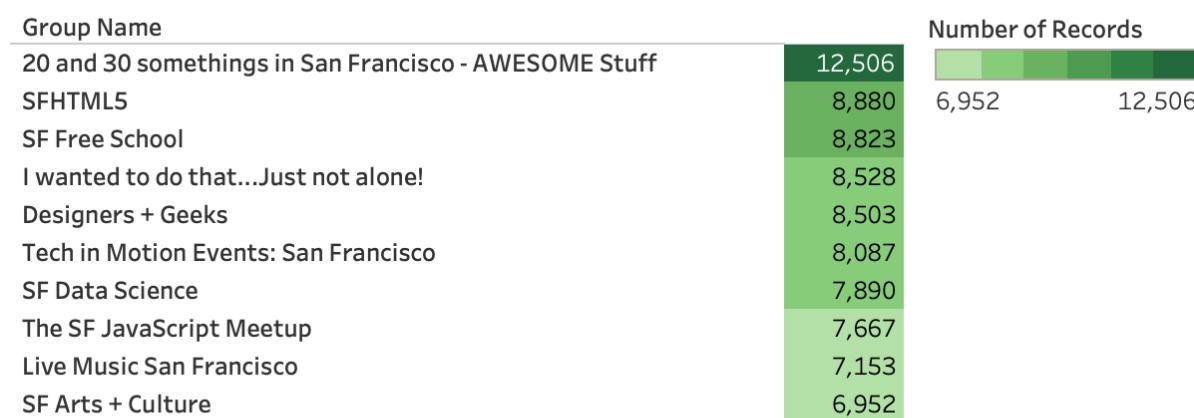


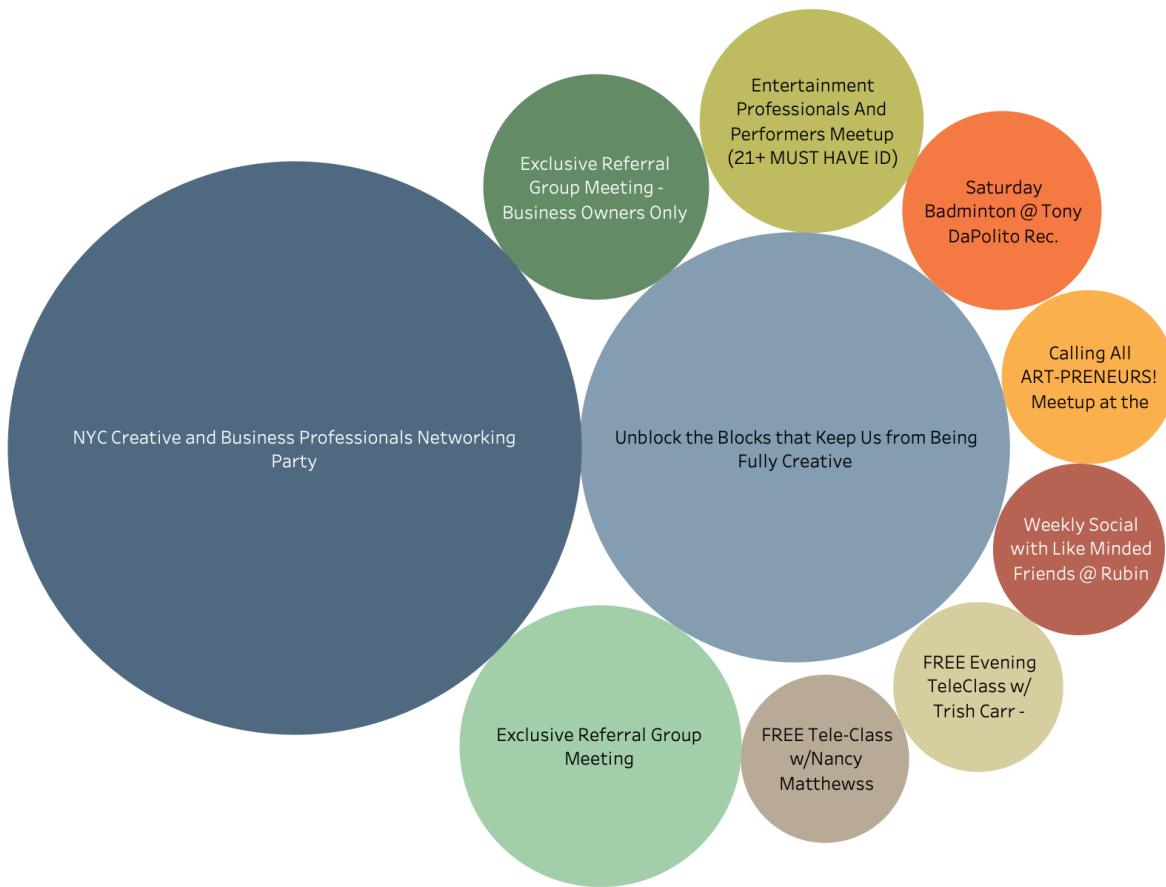
Figure 15 Top 10 popular groups in San Francisco

My Storyline

From the above three plots, Figure 13, Figure 14, and Figure 15, I plotted the top 10 popular groups on meetup at city level for each of the three cities: New York, Chicago, San Francisco. San Francisco is mostly about coding, programming and technical stuff, New York has both entrepreneurs and technical activities while Chicago is more inclusive and have some culture and lifestyle related events.

For the following three plots, Figure 16, Figure 17, and Figure 18, I plotted the Events with the top 10 most members in New York, Chicago and San Francisco respectively. The most popular events in New York City are NYC Creative and Business Professionals Networking Party and Unblock the Blocks that Keep Us from Being Fully Creative. The most popular events in Chicago are Drinks and Dancing at Soundbar Every Friday Night and Drinks and Dancing at Soundbar Every Saturday Night. The most popular events in San Francisco are Coding Drop-In Lab and Yoga for a Cause – Outdoor Yoga Flow. We can see people from the three cities are have quite different life styles, yet all events are very interesting to attend.

Events with the top 10 most members in New York

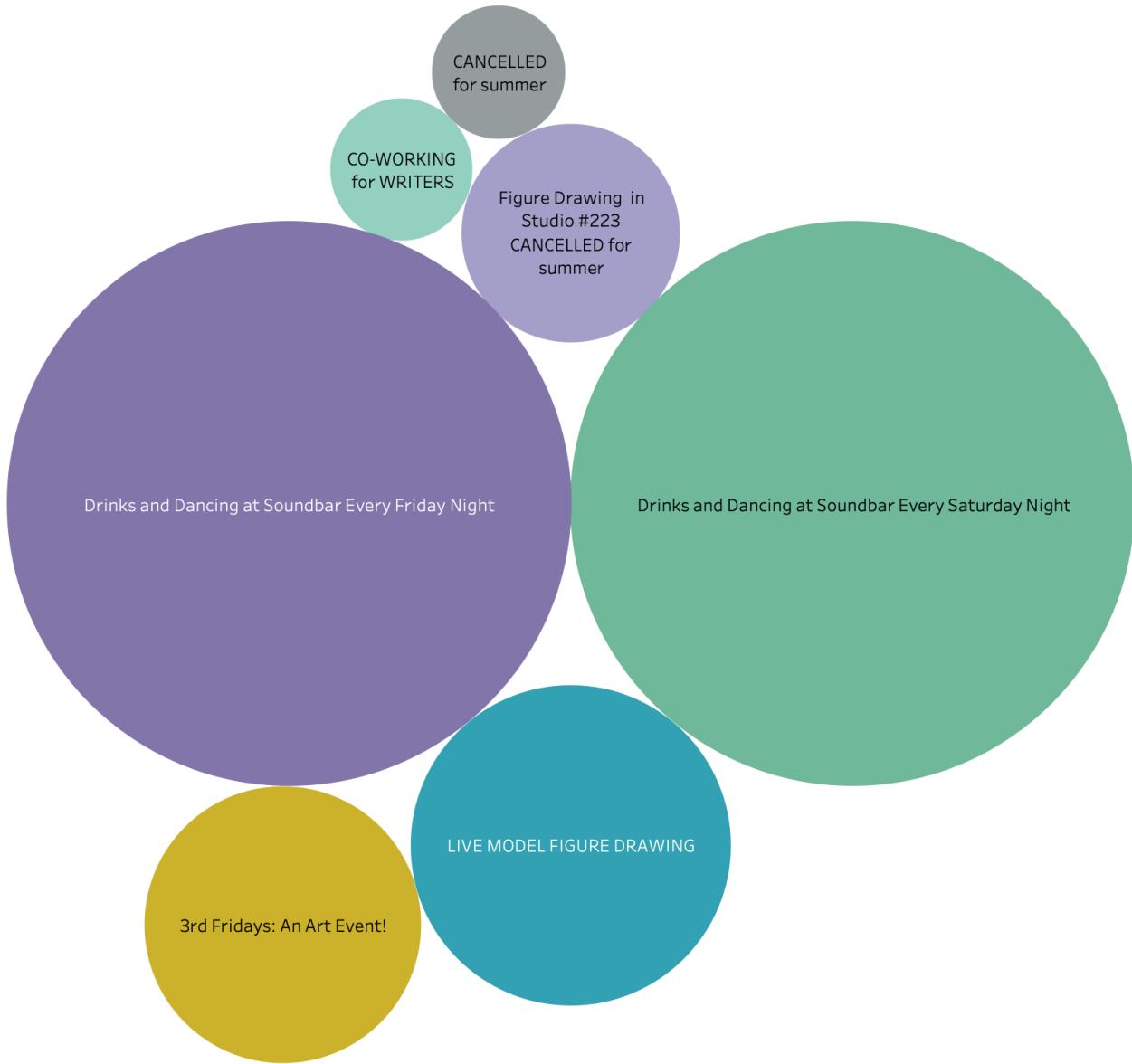


Event Name. Color shows details about Event Name. Size shows sum of Members. The marks are labeled by Event Name. The data is filtered on State, which keeps NY. The view is filtered on Event Name, which keeps 27 members.

- Event Name**
- NYC Creative and Business Professionals Networking Party
 - Unblock the Blocks that Keep Us from Being Fully Creative
 - Exclusive Referral Group Meeting
 - Exclusive Referral Group Meeting - Business Owners Only
 - Entertainment Professionals And Performers Meetup (21+ MUST HAVE ID)
 - Saturday Badminton @ Tony DaPolito Rec. Center 2 PM - 4:30 PM
 - Calling All ART-PRENEURS! Meetup at the Resource & Career Fair on Nov 3rd!
 - Weekly Social with Like Minded Friends @ Rubin Museum Cafe
 - FREE Evening TeleClass w/ Trish Carr - *WOW* Wednesday
 - FREE Tele-Class w/Nancy Matthewss *WOW Wednesday*

Figure 16 Events with the top 10 most members in New York

Events with the top 10 most members in Chicago



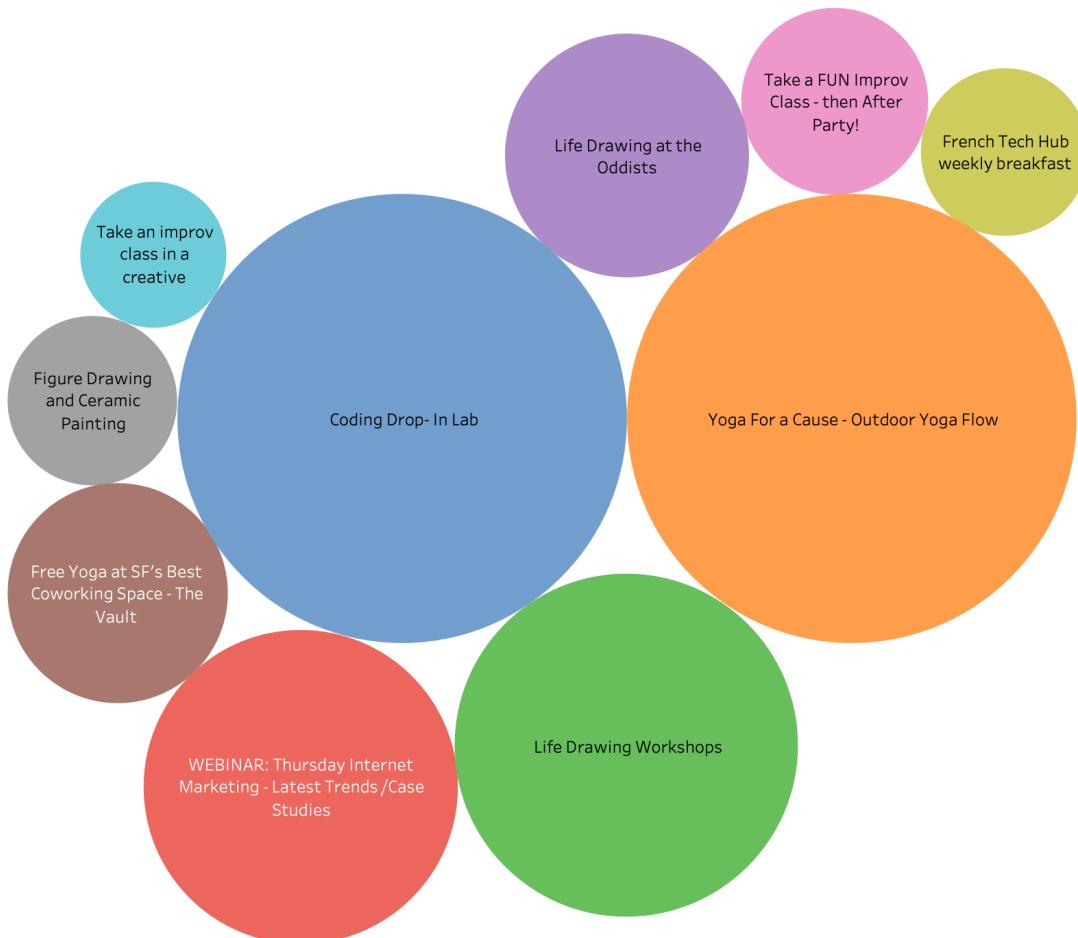
Event Name. Color shows details about Event Name. Size shows sum of Members. The marks are labeled by Event Name. The data is filtered on State, which keeps IL. The view is filtered on Event Name, which keeps 27 members.

Event Name

- Drinks and Dancing at Soundbar Every Friday Night
- Drinks and Dancing at Soundbar Every Saturday Night
- LIVE MODEL FIGURE DRAWING
- 3rd Fridays: An Art Event!
- Figure Drawing in Studio #223 CANCELLED for summer
- CO-WORKING for WRITERS
- CANCELLED for summer Figure Drawing 7-9PM in Studio #223 (BYOB)

Figure 17 Events with the top 10 most members in Chicago

Events with the top 10 most members in San Francisco



Event Name. Color shows details about Event Name. Size shows sum of Members. The marks are labeled by Event Name. The data is filtered on State, which keeps CA. The view is filtered on Event Name, which keeps 27 members.

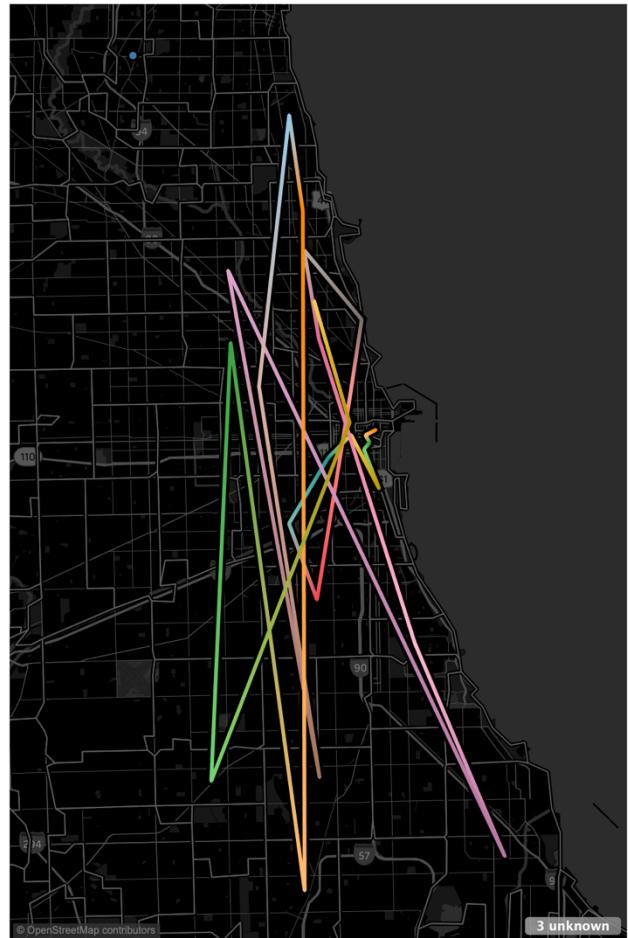
- Event Name**
- Coding Drop-In Lab
 - Yoga For a Cause - Outdoor Yoga Flow
 - Life Drawing Workshops
 - WEBINAR: Thursday Internet Marketing - Latest Trends /Case Studies
 - Life Drawing at the Oddists
 - Free Yoga at SF's Best Coworking Space - The Vault
 - Take a FUN Improv Class - then After Party!
 - Figure Drawing and Ceramic Painting
 - French Tech Hub weekly breakfast
 - Take an improv class in a creative community!

Figure 18 Events with the top 10 most members in San Francisco

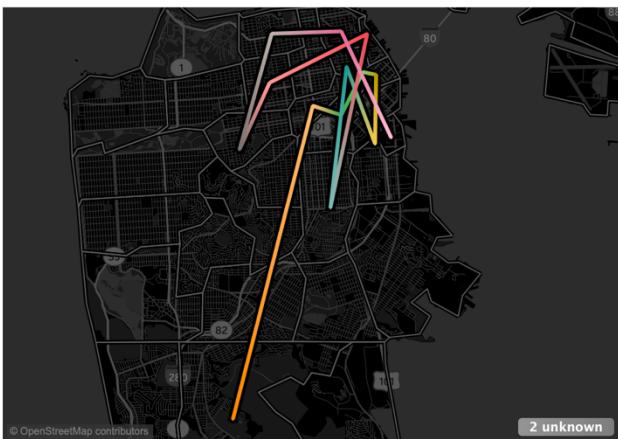
New York City



Chicago



San Francisco



Conclusion

From the above summary of data, we can see the lifestyle in Chicago is casual and relax, people like to have drinks and dance after work (Friday and Saturday nights). People from New York City area generally prefer events for business networking, entrepreneurship and creativity activities while people from San Francisco areas are mostly having coding, data science events and some yoga and drawing events as well. It is very interesting to see how three different cities' events vary from different interests.

Last but not least, I created a radial bar chart to represent the inclusiveness of meetup events in these three cities. We can share different interests while living in the same city. 😊

