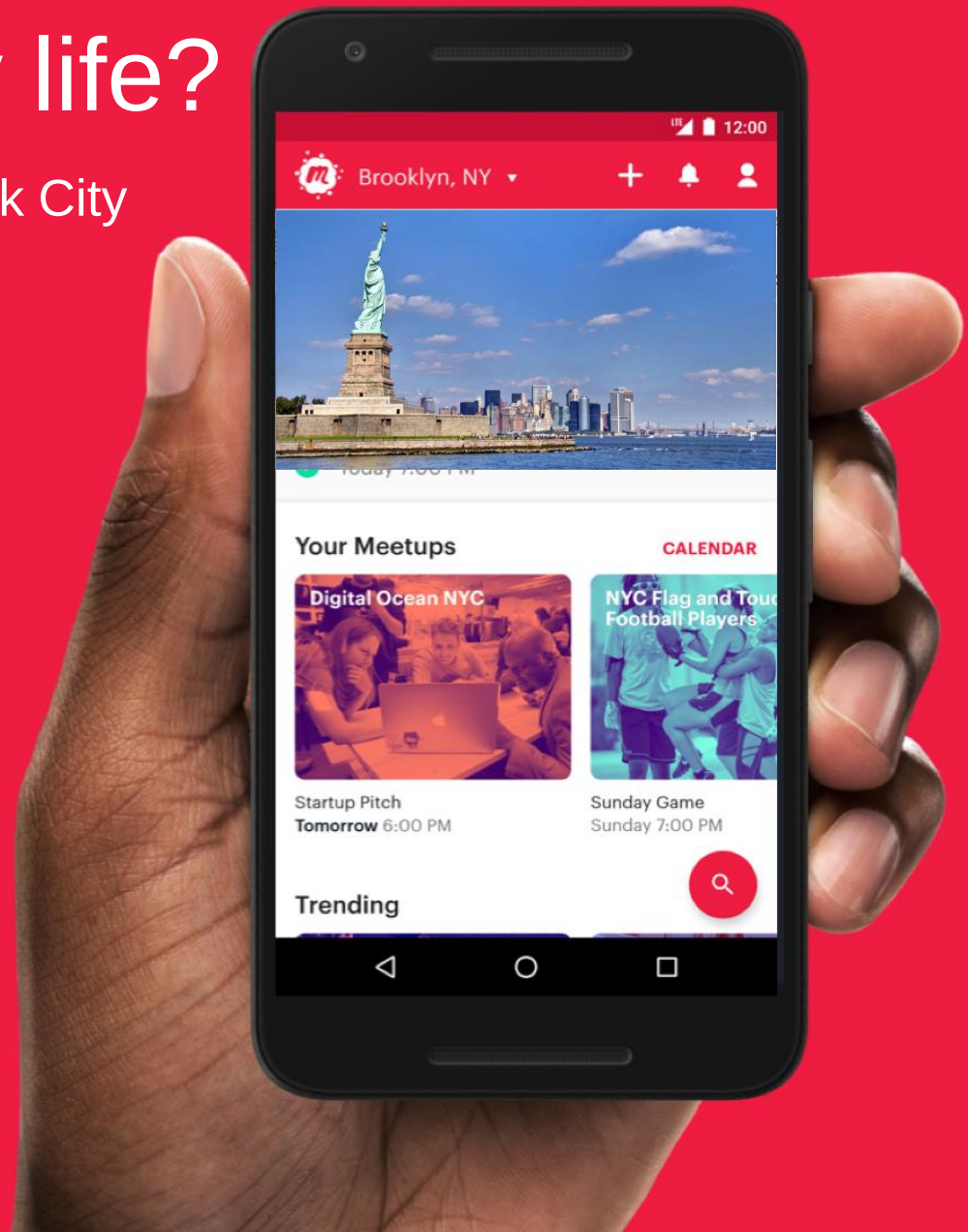
How did Sam live a happy life?

Examine Meetup in New York City

meetup



Group 3

Jake Arendsen, Feifan Gu, Sidi Liu, Yidi Zhang

Content

Sam is a recent college grad who just made the big move to New York City. Because he doesn't know Daniel or Allegra, he lacks friends in the area, so he created an account on Meetup to meet people. He is a data scientist with numerous interests; he is really interested in participating in *deep learning events, sci-fi conventions,* and other social gatherings in the area. He needs a tool to help him use **Meetup** to find his ideal groups, events and friends.

Motivation

- Sam faces a problem not completely unlike something we might face soon.
- In a new city, mostly to completely alone, looking for friends to spend some time with outside of work.
- We'll all likely turn to some form of social networking site.
- These can be cumbersome to navigate at first, so having a tool to find what we want would be very handy.

Overview

- Our goal is to build a dashboard to help a user find one or more groups of interest on Meetup in New York City.

- We hope to accomplish this by first allowing users to explore groups or events on a wide level before eventually having them filter and narrow their scope to learn about the groups and their interconnectivity.

Related Work/Inspiration

- Several of our group members have participated in several meetup events, which were pretty cool, but there's no good tool to figure out which groups are actually good and which events are worth attending.

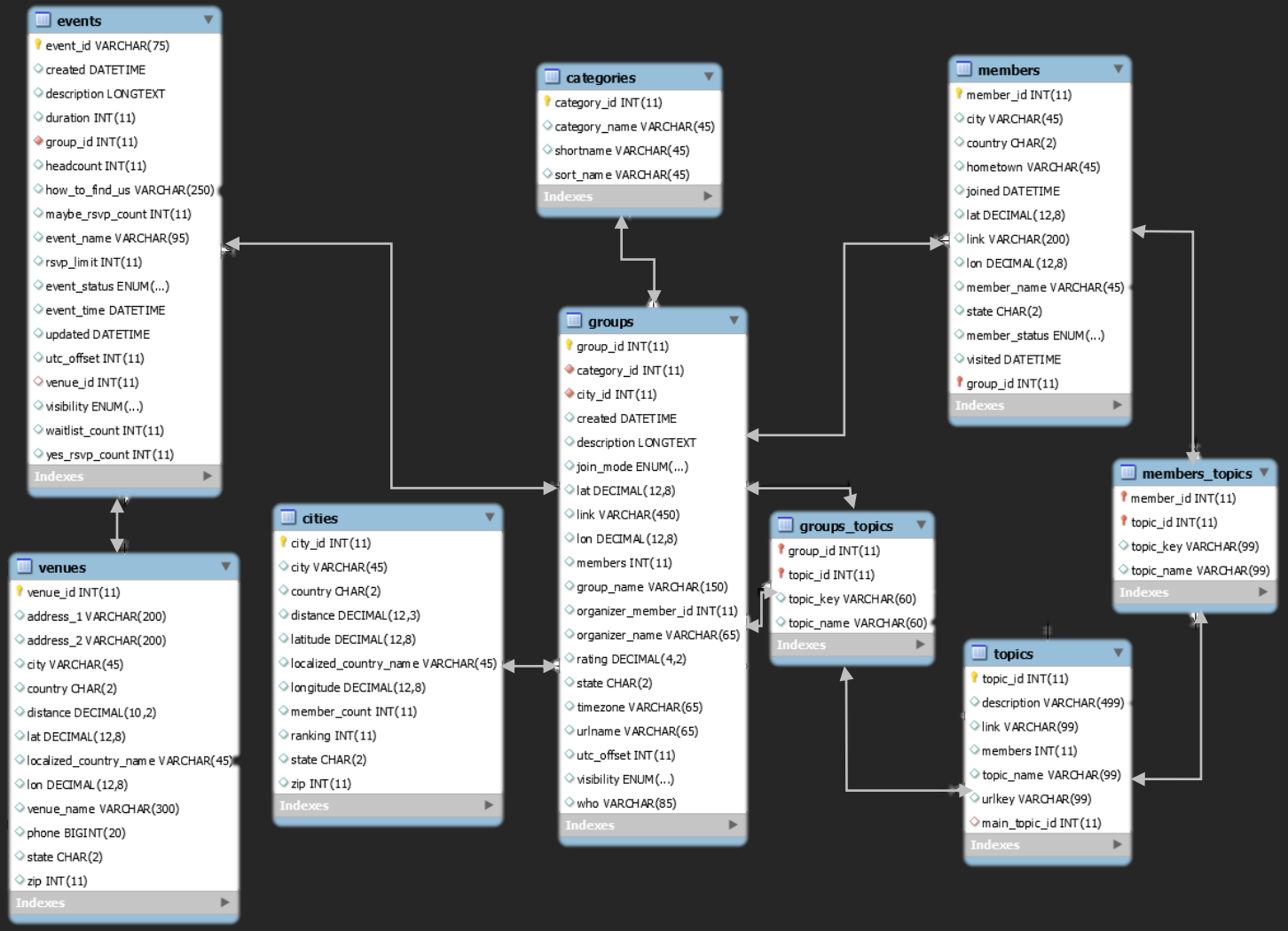
- Additionally, work from our social networks course inspired us to try to create a social network using Meetup data.

Questions

01	What groups would be best for Sam?	• We want to let Sam explore all groups in NYC freely. • Learn things such as rating, number of members, category
02	What kind of events does the group host? What other events are available?	• We want to let Sam explore all events in NYC. • Ideally, he'd be able to filter by both group and category. • We want to present as much information about these events as possible.
03	How can he learn even more about these groups?	• Present an interesting, different set of information about the group. • Potentially give information that makes it easier to compare and contrast the groups.
04	What are the people in this group like?	• We'll try to connect people in this group to other groups of interest. • Hopefully, the other groups will be different categories to help Sam expand his horizons.

Data
Description

Structure



- The data consists of a ton of public records from Meetup found on Kaggle.
- In all, we had access to nine different sheets, all interconnected as shown here.
- We mostly used groups, events, members, and categories in our analysis.

Data
Description

Key variables

	Variables	Counts
Categories	Category name	33
Topic	Topic name	2509
Groups	Group name	16330
	Rating	10282
	Created time	
	# of members	
	Group Link	
Members	Member name	1087923
	Bio	
	Joined Date	
	Member Link	

Variables	Counts
Events	
Event name	5807
Venue name	502
Event Date	
Venue Address	
Yes Rsvp count	
Duration	
Fee.amount	

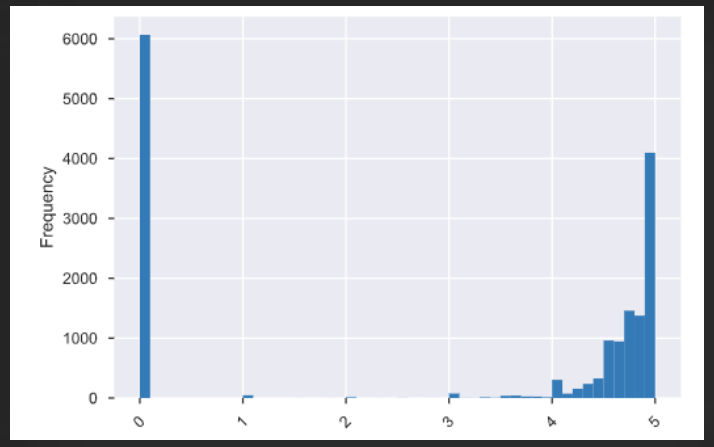
To avoid duplication, we use ids of topics, categories, groups, events, venues and members to represent them in the coding part and show their original name in the visualization.

EDA

Distribution of group location

Value	Count	Frequency (%)
New York	8565	52.4%
San Francisco	4555	27.9%
Chicago	3168	19.4%
South San Francisco	19	0.1%
West New York	11	0.1%
Chicago Ridge	5	< 0.1%
West Chicago	4	< 0.1%
Chicago Heights	2	< 0.1%
North Chicago	1	< 0.1%

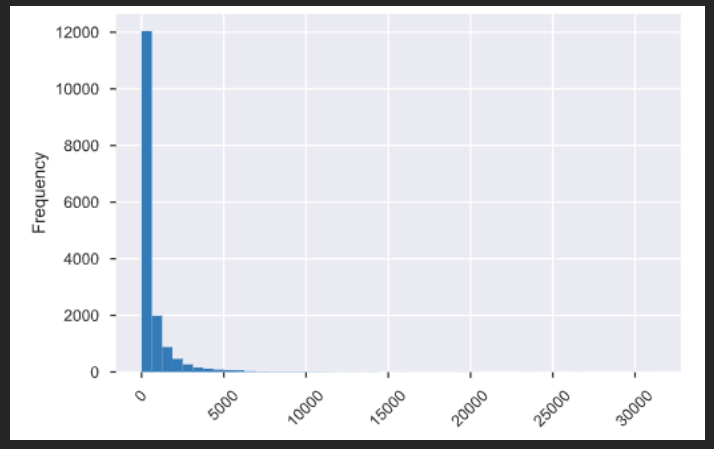
Distribution of group rating



Distribution of group join mode

Count	Frequency (%)
13530	82.9%
2745	16.8%
55	0.3%

Distribution of number of members



Visualize key variables using a package called pandas_profiling in Python

Data
Description

EDA: Event Data

yes_rsvp_count	
Min.	0.000
1st Qu.	1.000
Median	1.000
Mean	3.132
3rd Qu.	2.000
Max	198.000

Duration	
Min.	2,700
1st Qu.	10,800
Median	10,800
Mean	10,769
3rd Qu.	10,800
Max	1,209,600

Through the EDA part, we realized that we are dealing with a giant dataset, actually 1GB in all. So it's necessary for us to narrow the data to show on the viz. Also, the numerical variables inspired us to try to find the elements contributing to the rating.

The main transformation we did is to use the affiliation relationship between groups and members to form edge list to build social networks in R, thus creating new attributes of groups: different centralities.

To Tableau

We go!

Design Evolution - Dashboard 1

We started with the following premise:

> Sam wants to choose a category that he is interested in. Once he made his choice, different topics and groups under that category appear. He is presented with information regarding these groups and he will choose a group that he preferred.

What did we end up doing?

- Sam has the ability to choose a city of his interest. Once he chooses the city, the corresponding categories of Meetup groups will show, with the size corresponding to the popularity of the category (amount of members within groups under that category).
- He can then select any category of his interest. The corresponding topics and groups under those topics would appear on the right.
- He may have a preference to join a newer group or older group. Thus, we provide him with a tool to further sort the groups by the year they were created.
- It is too early for Sam to make his decision, so we decide to provide him with further details to guide him through the process.

Dashboard 1

Design Evolution - Dashboard 1

Topics — Oldest to Newest / Newest to Oldest

Topic Name	Group Name	Year Created	Rating
Adaptive Techno..	Effective Altruism NYC	2013	4.86
	NYC Six Nations/ Tri-Nations /..	2006	4.84
Adult Education	Startupers: Growth hacking f..	2015	4.5
Aging Parents	Coach AGIWAL : Badminton,Te..	2009	4.81
Alternative Spor..	Black U: learn from black expe..	2016	5
	Play! NYC	2012	4.59
Applied Spirituality	Bavaria & Oktoberfest	2017	0
	Christian Women Entrepreneu..	2017	0
	Coach Swissy's School Of Boxi..	2017	0
	Experience Optimization to Ex..	2017	0
	Fitness & Fat Loss for the Offi..	2017	0
	GA Alumni Backend Study Gro..	2017	5
	New York City Christian Caree..	2017	0
	New York Executives in Career..	2017	0
	NYCDEV	2017	0
	Planned Acts Of Kindness - Th..	2017	5
	the cap is released at 21 millio..	2017	0
	Maker Mammas NYC	2016	0
	BizNet New York	2015	0
	New York - Cloud Communicat..	2015	0
	NYC Partner Network	2015	0
	Uptown Wondrous Working W..	2015	0
	BHRP-NYC Young Survival Coa..	2014	0
	Bonsai Meetup	2014	5
	NYC Real-Time Computing (NY..	2014	0
	NYC CompSec	2013	5
	NYC Black Professional Wome..	2012	4.93
	Living Energy~ Global & Local ..	2008	5
Argentine Tango	A Shoulder to Cry On - I Believ..	2017	0
Movies and Video	Acoustic Guitar Players Meetup	2017	5
	Art of Marketing & Sales for P..	2017	0
	Be Your Own TV Audition Guru..	2017	0
	Bhangra Classes & Events in N..	2017	5
	Black & Latino GBTQ Men's Ex..	2017	0
	British Girls in New York City!	2017	0
	Brooklyn Beginner/Intermedi..	2017	5
	Brunch Crew	2017	0
	Bunker Labs New York	2017	0
	Camp Birdy for Grown Ups	2017	0
	CCDE-NYC	2017	0
	Christians Who Drink	2017	0
	Cider and Apple Enthusiasts	2017	4.67

- A Word Cloud can clearly show what are the most popular categories in each city.
- Table is an easy way to show the relevant information (topics, groups etc.) in an organized manner.

Dashboard 1

Design Evolution - Dashboard 2

We started with the following premise:

> Then he checked past events of the groups. His choice is limited to 5 groups here

What did we end up doing?

- We decided it was a bit too early to narrow down to just five groups, so we decided to do an analysis of all events in New York City.
- Due to limitations of the data, we only have event information for three distinct categories of events, so we can't really let Sam choose his categories of interest.
- We gave Sam the ability to filter by both group and category to allow him to gain a maximum understanding of what events look like in NYC.

Design Evolution - Dashboard 2

- Colors represent the category of the group, consistent across all visualizations.
- The top viz contains no axis information as none is necessary; its primary goal is to be a sophisticated filter, as indicated by its tooltip.
- Both the map and the scatter plot contain in depth tooltips, as they are the primary location a user would go to learn about events. Clicking either shows a link to view the event page on Meetup, as well as highlighting it in the other chart.
- The bottom chart reminds a user what groups are rated and gives the user a way to view events for a specific group. It is ordered alphabetically to make finding a group easier.

Dashboard 2

Design Evolution - Dashboard 3

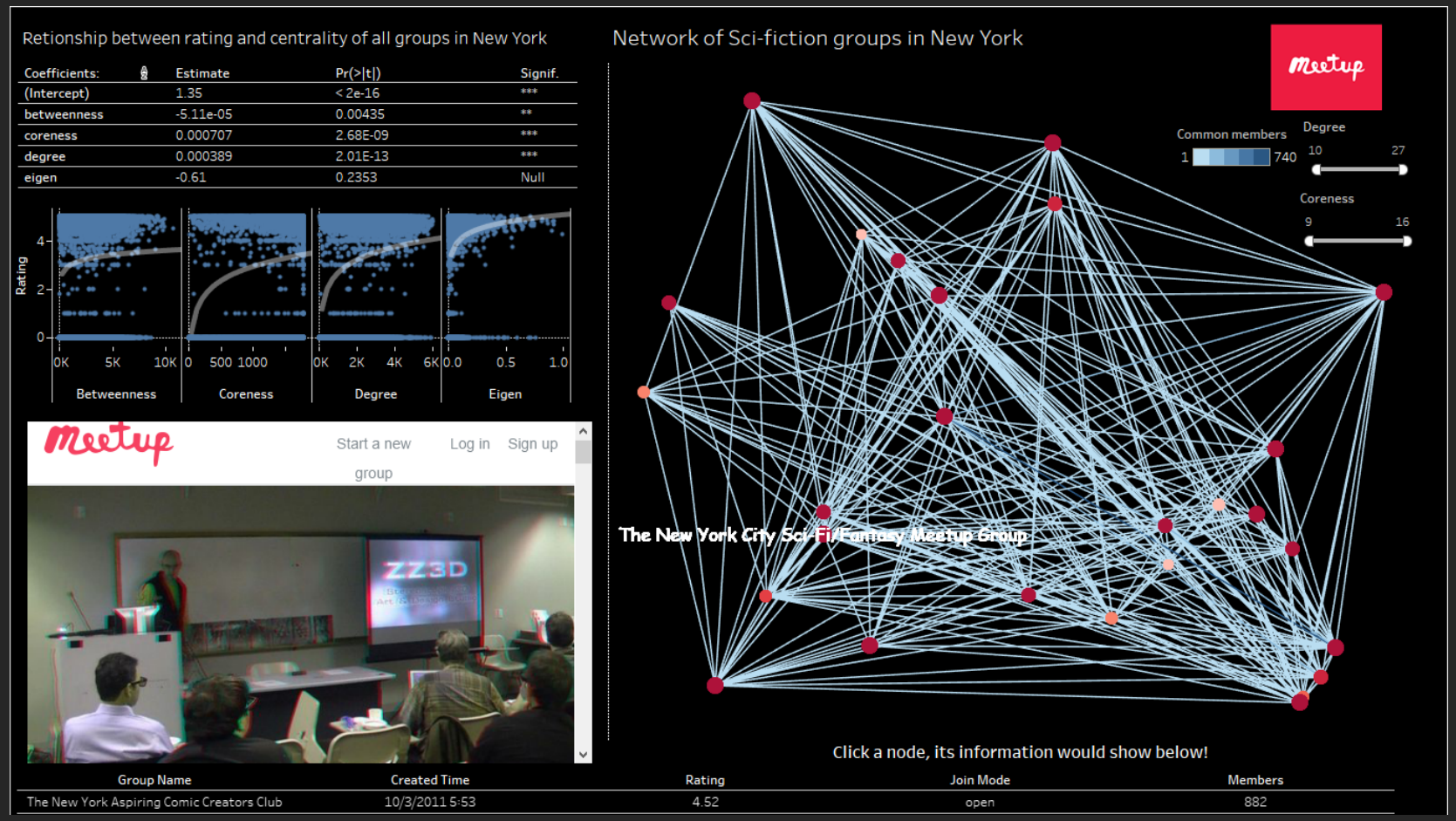
We started with the following premise:

> Finally, he wants to choose the groups in the core, i.e. groups having more relations with other groups (sharing members), so that he can access to more people. He uses the information from the social network to choose several groups to focus on.

What did we end up doing?

- First, we built a social network among groups. We consider two groups to have a connection if they share common members. The weight of the connection is number of shared members.
- Then, we ran a regression to justify that a group's rating does have a positive relationship with its degree centrality and coreness centrality. Degree here means how many groups Sam can get to know through members of one group, and coreness means how large local group network one group is in, helping Sam reach a bigger family of sci-fiction fans.
- Also, we gave Sam the ability to see the detailed information about groups, including its homepage.

Design Evolution - Dashboard 3



- The regression result is in the top-left corner, showing coreness and degree have significantly positive relationship with rating. Below are 4 individual scatter plot with logarithmic trend lines.
- The network between groups can be filtered by groups' degree and coreness. The density of nodes represents degree and size represents coreness, and the density of lines represents the weight of connection.
- By clicking on a node on the network, a link appears that allows the bottom-left webpage to show the group's homepage and the text below would show its name, created time, rating, join mode and number of members.

Dashboard 3

Design Evolution - Dashboard 4

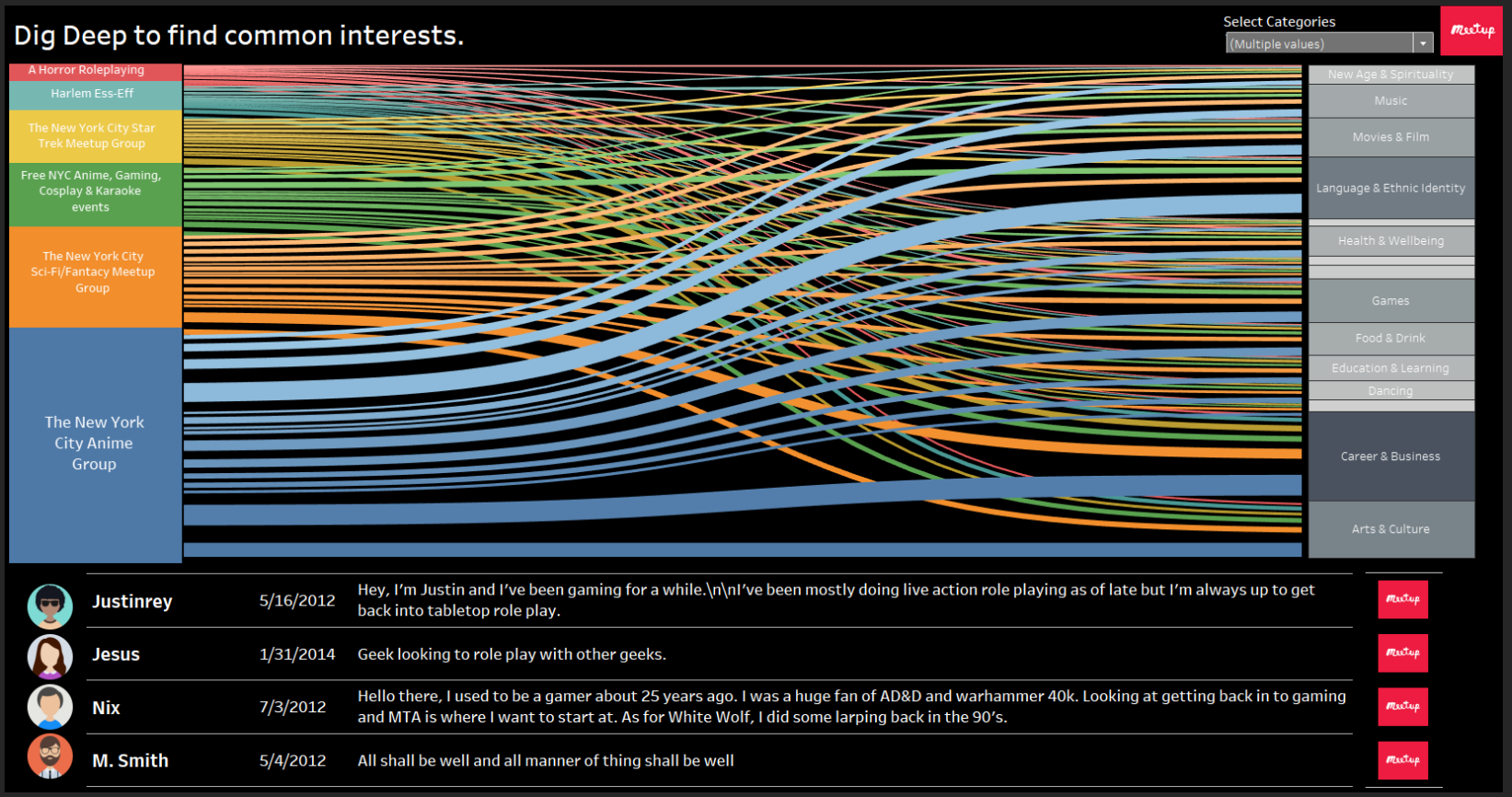
We started with the following premise:

> Sam is now preparing for the upcoming events held by his top 2 groups. After registering, he decided to know more about the people that he will meet. Clicking on a member of a group, he can see the past events, other group memberships, and notable topics of interest for the member. He decided several people to mainly talk to and finally went to the venue with confidence.

What did we end up doing?

- Finally, Sam picks out his top 6 Sci-fiction groups. Then he wants to know more about these groups. Instead of scanning random members' profiles in each group, Sam wants to see the trend of members' interests in a big picture. So we can show him the top 10 categories that those members of his top 6 sci-fiction groups select to join in other groups.
- Besides, we also filter out four most informative member with his or her bio and meetup link for Sam to dig deep and even make friends in advance.
- Also, Sam can check each group's upcoming events to join in through the group link.

Design Evolution - Dashboard 4



- The left side categories of the Sankey chart are top 6 sci-fiction groups that generated from last step. The right-side categories are the top or selected categories that attract members in top 6 sci-fiction groups to join in events.
- The flow size of the Sankey chart represents the number of members into different outputs, a.k.a. categories.
- The bottom table lists four most informative members' profiles along with his/her registered name, joining time, bio and meetup link in each sci-fiction group for Sam's reference.
- By click the link and learn more on his/her personal website, Sam can connect to a few of them who share common interest with him and go to the event together next time.

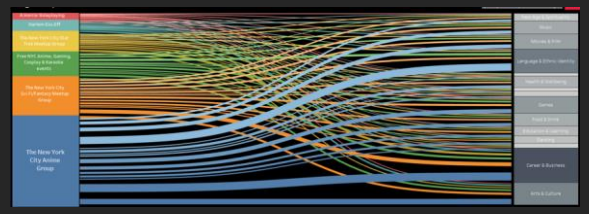
Dashboard 4

Insights & Self-evaluation

Implementation

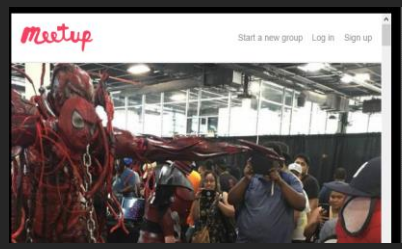
Step 4

Sam can find out what other categories attract the members in his top groups.



Step 3

Sam can check out any group's homepage by pressing the link in the network.

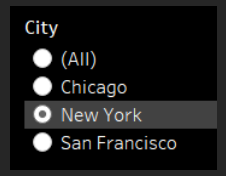


Step 2

Sam can pick a category to further narrow the groups and events displayed to him.

Sam can find out the popular categories in his city.

Step 1



Insights

- Meetup is full of numerous groups, and, as cliché as it sounds, really does seem to have something for everybody.
- While Meetup has numerous events, many of them go either unattended or under-attended. This means if you're using it to make friends, you might have to try a few different events to really meet a variety of people.
- Many of Meetup's groups are very strongly connected through other groups through members, making it very easy to become strongly connected to the network. This also means if you're torn between a few similar groups, you likely shouldn't worry, as you'll meet at least some of the same people at all three.
- Numerous Meetup members are members in vastly different groups. While this means you'll meet exciting people with varied interests, it also means there's no guarantee going to a different style event will introduce you to new people. It might require more than just Meetup to become well ingrained in a social scene in a new city.

Self-Evaluation

We think our visualization is relatively effective at accomplishing our end goal

- Due to limitations in our data, as through of an exploration as we'd like is impossible
- Likewise, the need to filter as aggressively as we initially planned proved unnecessary

The things we are proudest of include

- Incorporating an igraph into Tableau
- Creating our Sankey chart
- The interactivity created across our dashboards, but specifically within our first two

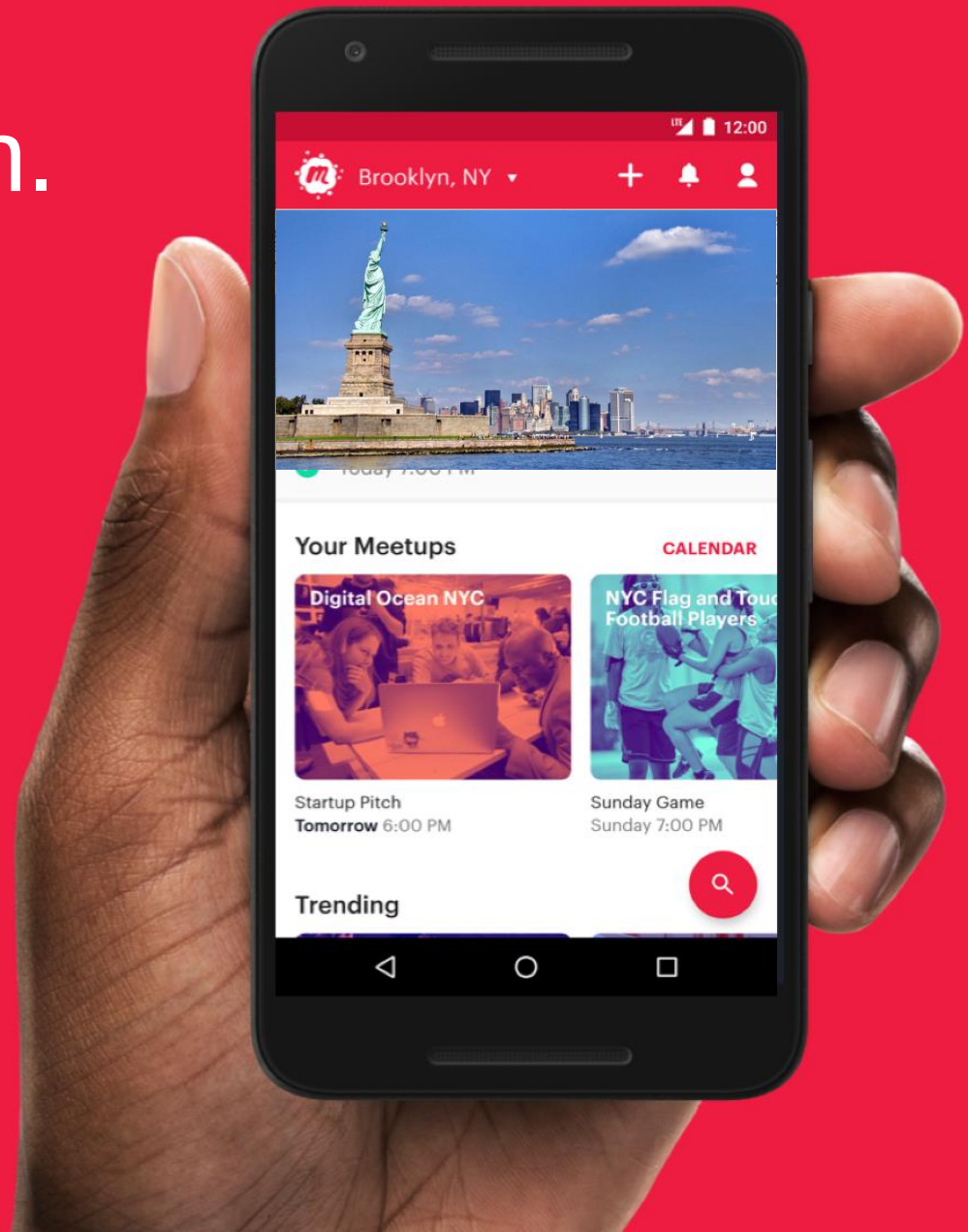
Some things we'd work on more with more time include

- Potentially allowing our last two dashboards to work for a choice of categories
- Making our first two dashboards more informative, likely via tooltips

Sam lives a happy life then.

meetup

THANKS



Appendix

R codes

Internal Edgelist of sci-fiction groups

```
######################## internal edgelist of sci-fiction groups ############################

edgelist <- left_join(affi,affi,by="member_id")
edgelist <- filter(edgelist,group_id.x<group_id.y)
edgelist <- as.data.table(edgelist)
edgelist[,weight:=length(member_id),by=c("group_id.x","group_id.y")]
edge_list <- unique(edgelist[,c(2:4)])

length(unique(union(edge_list$group_id.x,edge_list$group_id.y)))

length(unique(edge_list$group_id.x))

coordinate <- data.frame(group_id=numeric(28))
coordinate$group_id<-unique(union(edge_list$group_id.x,edge_list$group_id.y))
coordinate$LineX<-runif(28, min = 0, max = 100)
coordinate$LineY<-runif(28, min = 0, max = 100)
edge_list1 <- left_join(edge_list,coordinate,by=c("group_id.x"="group_id"))
#fwrite(edge_list1,"edge_list1.csv")
edge_list2 <- left_join(edge_list,coordinate,by=c("group_id.y"="group_id"))
#fwrite(edge_list2,"edge_list2.csv")

edge_list1$group_id.x <- as.character(edge_list1$group_id.x)
edge_list1$group_id.y <- as.character(edge_list1$group_id.y)
g <- graph_from_edgelist(as.matrix(edge_list1[,c(1:2)]))
E(g)$weight <- edge_list1$weight


coordinate$degree <- degree(g)
coordinate$coreness <- coreness(g)

coordinate$group_id <- as.character(coordinate$group_id)

edge_list2$group_id.x <- as.character(edge_list2$group_id.x)
edge_list2$group_id.y <- as.character(edge_list2$group_id.y)

edge_list11 <- left_join(edge_list1,coordinate[,c(1,4,5)],by=c("group_id.x"="group_id"))
fwrite(edge_list11,"edge_list11.csv")
edge_list22 <- left_join(edge_list2,coordinate[,c(1,4,5)],by=c("group_id.y"="group_id"))
fwrite(edge_list22,"edge_list22.csv")


newyork_sci$group_id <- as.character(newyork_sci$group_id)
sci_group_info <- left_join(coordinate,newyork_sci[,c("group_id","group_name","created","description",
sci_group_info <- sci_group_info[,c(1,4:10)]
fwrite(sci_group_info,"sci_group_info.csv")
```

Regression for all groups in NYC

```
######################## regression for all groups in NYC ############################

coordinate <- data.frame(group_id=numeric(6458))
coordinate$group_id<-unique(union(edge_list$group_id.x,edge_list$group_id.y))

edge_list1 <- left_join(edge_list,coordinate,by=c("group_id.x"="group_id"))
#fwrite(edge_list1,"edge_list1.csv")


edge_list1$group_id.x <- as.character(edge_list1$group_id.x)
edge_list1$group_id.y <- as.character(edge_list1$group_id.y)
g <- graph_from_edgelist(as.matrix(edge_list1[,c(1:2)]))
E(g)$weight <- edge_list1$weight
get.data.frame(g)
degree(g)
sort(degree(g))
sort(eigen_centrality(g)$vector)

coordinate$degree <- degree(g)
coordinate$eigen <- eigen_centrality(g)$vector
coordinate$betweenness <- betweenness(g)
coordinate$coreness <- coreness(g)

coordinate$group_id <- as.character(coordinate$group_id)
groups$group_id <- as.character(groups$group_id)
coordinate2 <- left_join(coordinate,groups)

edge_list2$group_id.x <- as.character(edge_list2$group_id.x)
edge_list2$group_id.y <- as.character(edge_list2$group_id.y)

edge_list11 <- left_join(edge_list1,coordinate[,c(1,4,5)],by=c("group_id.x"="group_id"))
fwrite(edge_list11,"edge_list11.csv")
edge_list22 <- left_join(edge_list2,coordinate[,c(1,4,5)],by=c("group_id.y"="group_id"))
fwrite(edge_list22,"edge_list22.csv")


newyork_sci$group_id <- as.character(newyork_sci$group_id)
sci_group_info <- left_join(coordinate,newyork_sci[,c("group_id","group_name","created","description",
sci_group_info <- sci_group_info[,c(1,4:10)]

play <- lm(rating~degree+eigen+betweenness+coreness,coordinate2)
summary(play)
```

R codes

Count other interests than sci-fiction

```
######################### count other interests than sci-fiction ##############################

members2 <- members[group_id%in%newyork_sci$group_id]
member <- members[group_id==8940,member_id]

cat <- data.table(group_id=0,category_id=0)
for (mid in member){
  group <- as.data.frame(members2[member_id==mid,group_id])
  colnames(group)="group_id"
  group2 <- left_join(group,groups[,c(1,2)])
  group3 <- as.data.table(group2)[category_id!=29]
  cat <- rbind(cat,group3)
}
cat <- cat[group_id!=0]

#table(cat$category_id)
catcom <- as.data.frame(table(cat$category_id))
catcom$group_id=8940
names(catcom)=c("other_category","freq","group_id")


for (gid in c(137667, 147811, 3797562, 18524790,22488816)){
  member <- members[group_id==gid,member_id]
  cat <- data.table(group_id=0,category_id=0)
  for (mid in member){
    group <- as.data.frame(members2[member_id==mid,group_id])
    colnames(group)="group_id"
    group2 <- left_join(group,groups[,c(1,2)])
    group3 <- as.data.table(group2)[category_id!=29]
    cat <- rbind(cat,group3)
  }
  cat <- cat[group_id!=0]

  #table(cat$category_id)
  catcomb <- as.data.frame(table(cat$category_id))
  catcomb$group_id=gid
  names(catcomb)=c("other_category","freq","group_id")
  catcom <- rbind(catcom,catcomb)
}


fwrite(catcom,"otherinterest.csv")
```