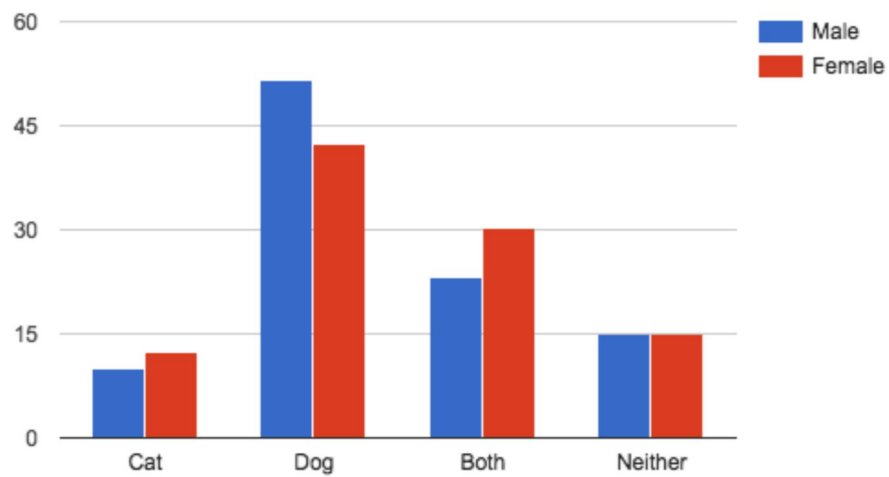# Week 4: Data Viz

Designing ML, Jan 2019
michelle.carney@berkeley.edu
from Info I247/190 with Marti Hearst
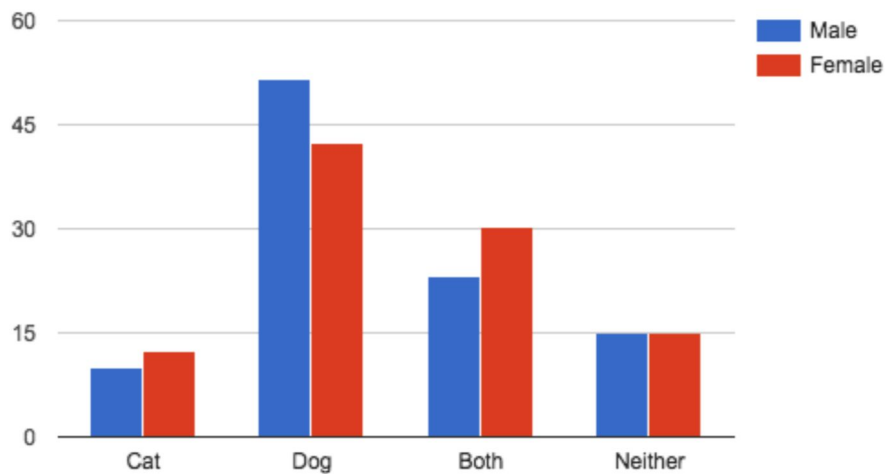
What is Data Viz?
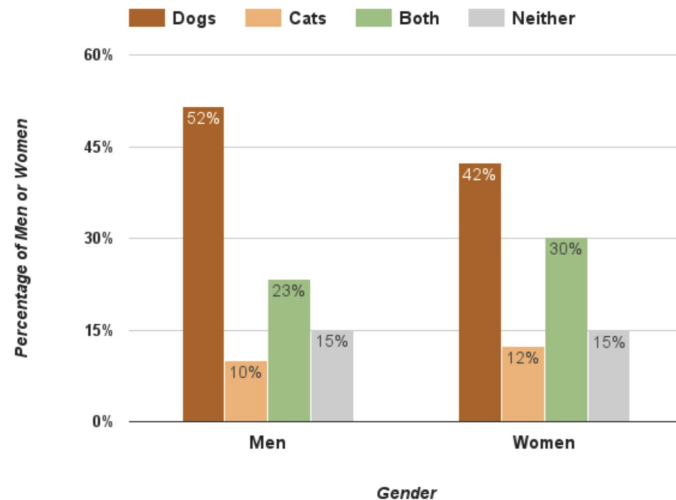
Pet Preference by Gender

Pet Preference by Gender

Legend: Male, Female

	Cat	Dog	Both	Neither

Pet preferences for Men and Women

Legend: Dogs, Cats, Both, Neither

Percentage of Men or Women

	Dogs	Cats	Both	Neither
Men	52%	10%	23%	15%
Women	42%	12%	30%	15%

Gender

Pet preferences of men and women

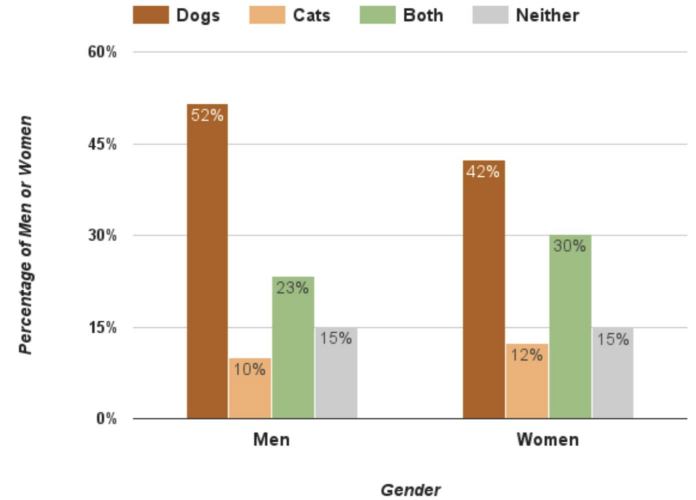|	Cat	Dog	Both	Neither
Men	10.0%	51.6%	23.3%	15.1%
Women	12.4%	42.3%	30.3%	15.0%

Pet preferences for Men and Women

Dogs · Cats · Both · Neither

	Dogs	Cats	Both	Neither
Men	52%	10%	23%	15%
Women	42%	12%	30%	15%

Pet preferences of men and women

Men — Cat 10.0% | Dog 51.6% | Both 23.3% | Neither 15.1%

Women — Cat 12.4% | Dog 42.3% | Both 30.3% | Neither 15.0%

Cat | Dog | Both | Neither
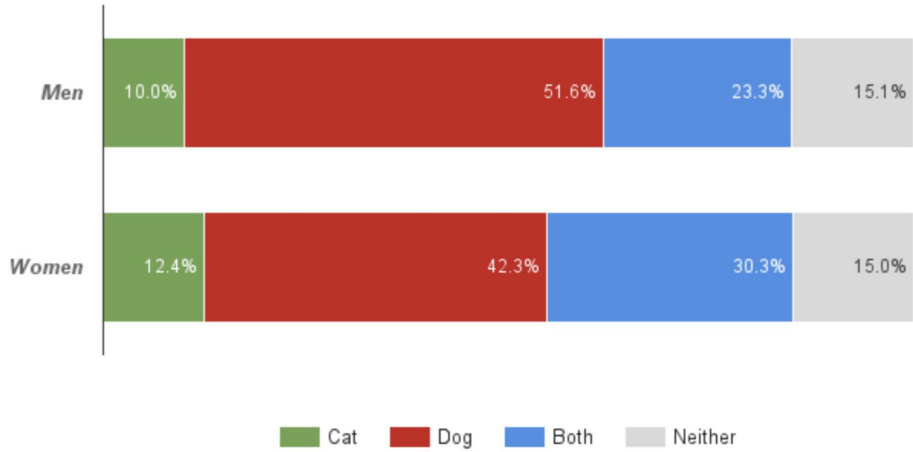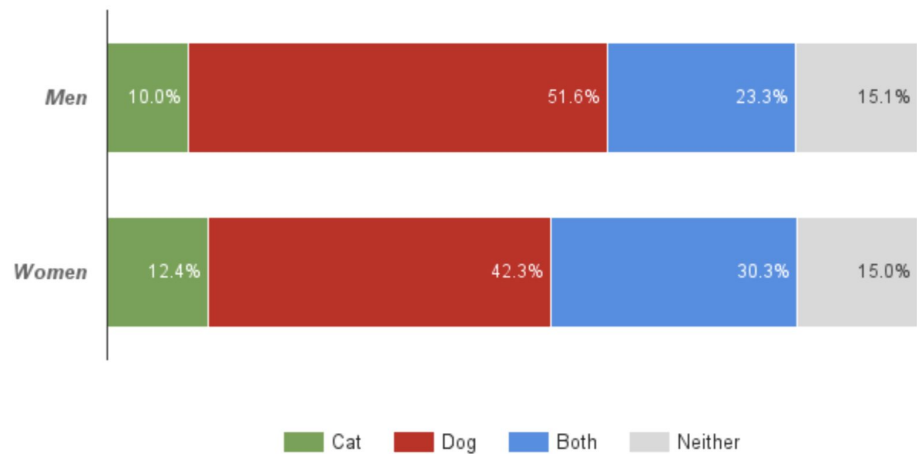
Compared to Women, What Pets Do Men Prefer?

Neither
Cats
Dogs
Both

-10% | -5% | 0% | 5% | 10%

Women Prefer More | Men Prefer More

Pet preferences of men and women

	Cat	Dog	Both	Neither
Men	10.0%	51.6%	23.3%	15.1%
Women	12.4%	42.3%	30.3%	15.0%

Cat Dog Both Neither

Pet Preference by Gender

Cat
Dog
Both
Neither
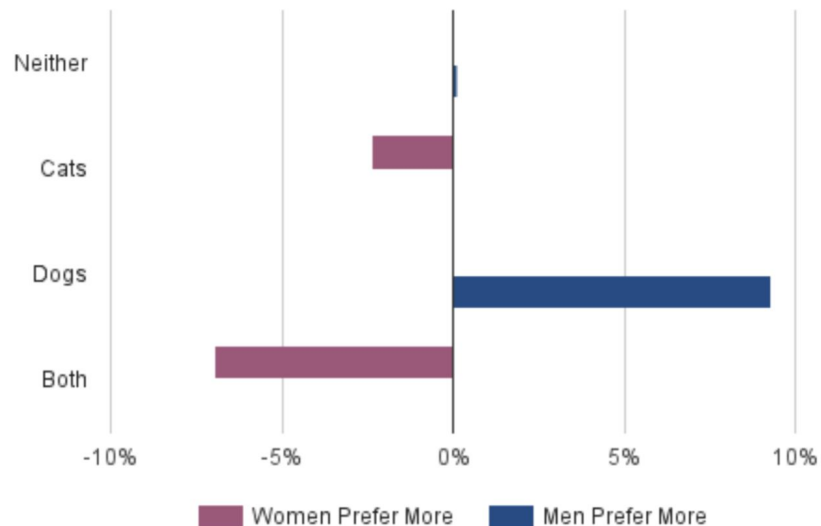
Gender
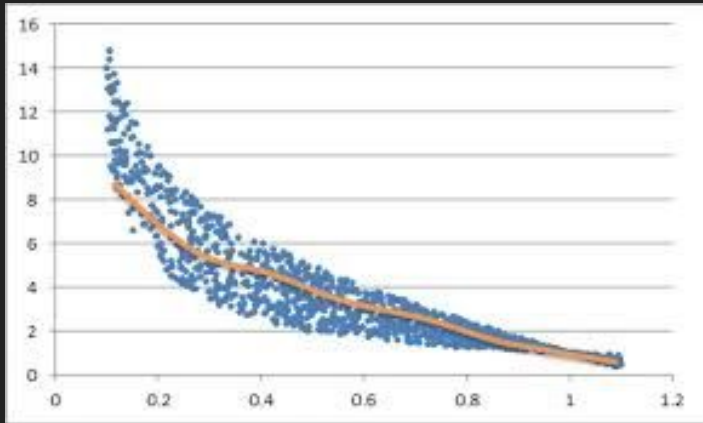
Pre-ML Data Viz (Exploratory Data Analysis)

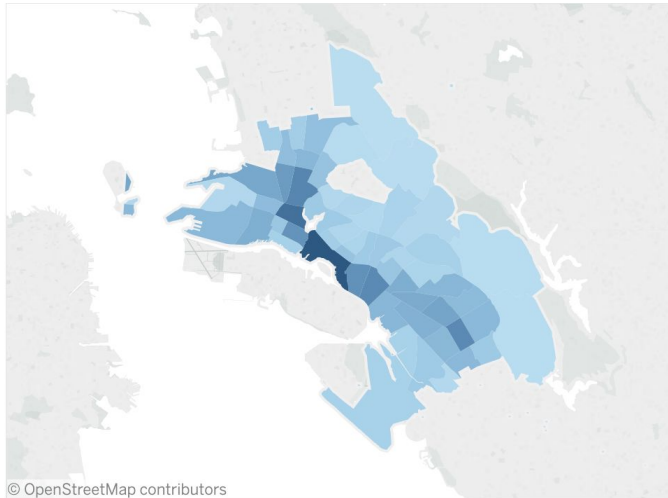Understand what your data actually is!

What is the scope of your data?

What are trends or insights?

What are hypotheses you can test?

Police Beats of Oakland

© OpenStreetMap contributors

Co-occurrence matrix

	Alice (3)	Bob (6)	Carol (5)	Dave (6)	Eve (4)
Alice (3)	1.00	0.33	0.33	0.67	0.33
Bob (6)	0.17	1.00	0.33	0.33	0.33
Carol (5)	0.20	0.40	1.00	0.80	0.40
Dave (6)	0.33	0.33	0.67	1.00	0.50
Eve (4)	0.25	0.50	0.50	0.75	1.00
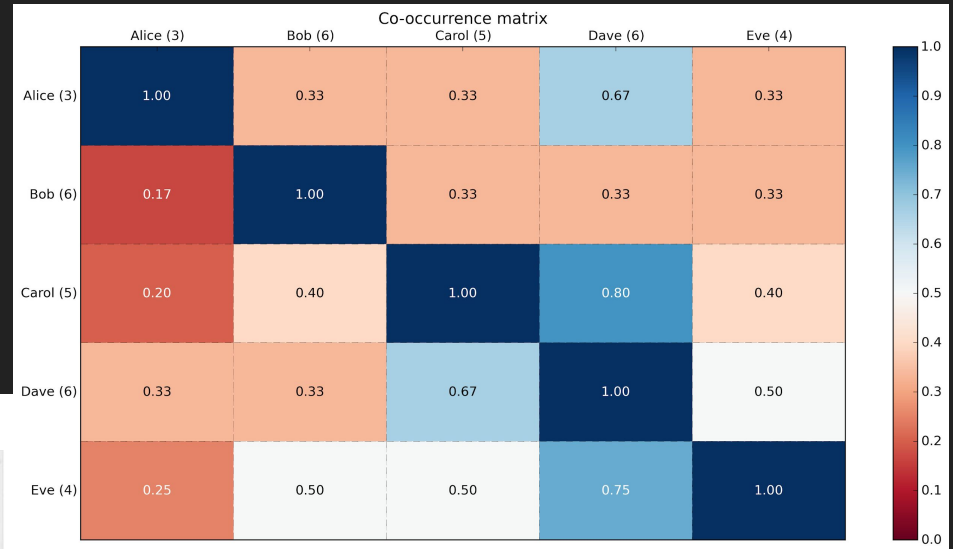
Post-ML Data Viz

Mapping your labels back to your data

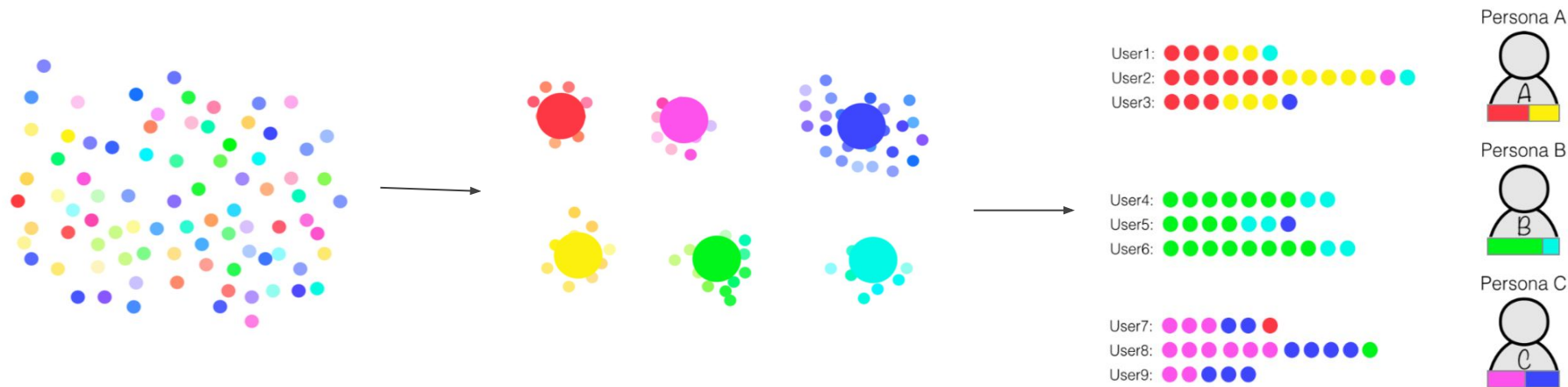What happened with your model predictions?

Does this match what is expected?

Ways for interaction (!!!)

Data Driven Personas: Clustering

Zhang et al (CHI 2016) Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry

Data Viz: Best Practices

Position along a common scale
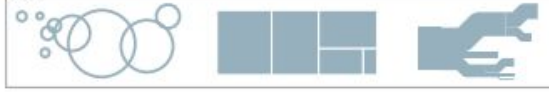
Position along identical, nonaligned scales
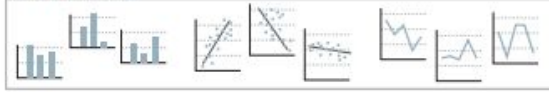
Length

Direction/slope   Angle

Area

Volume

Shading and saturation
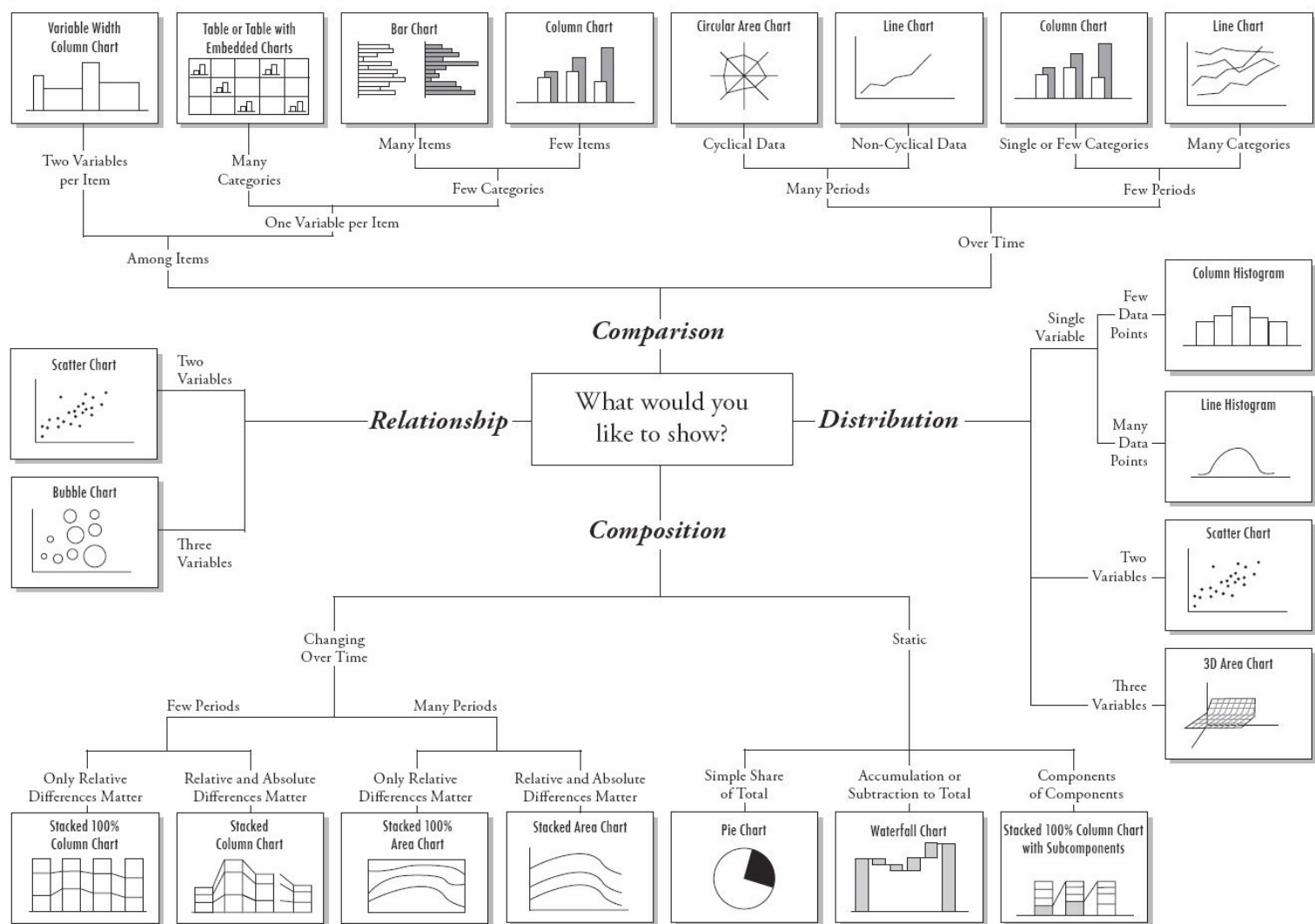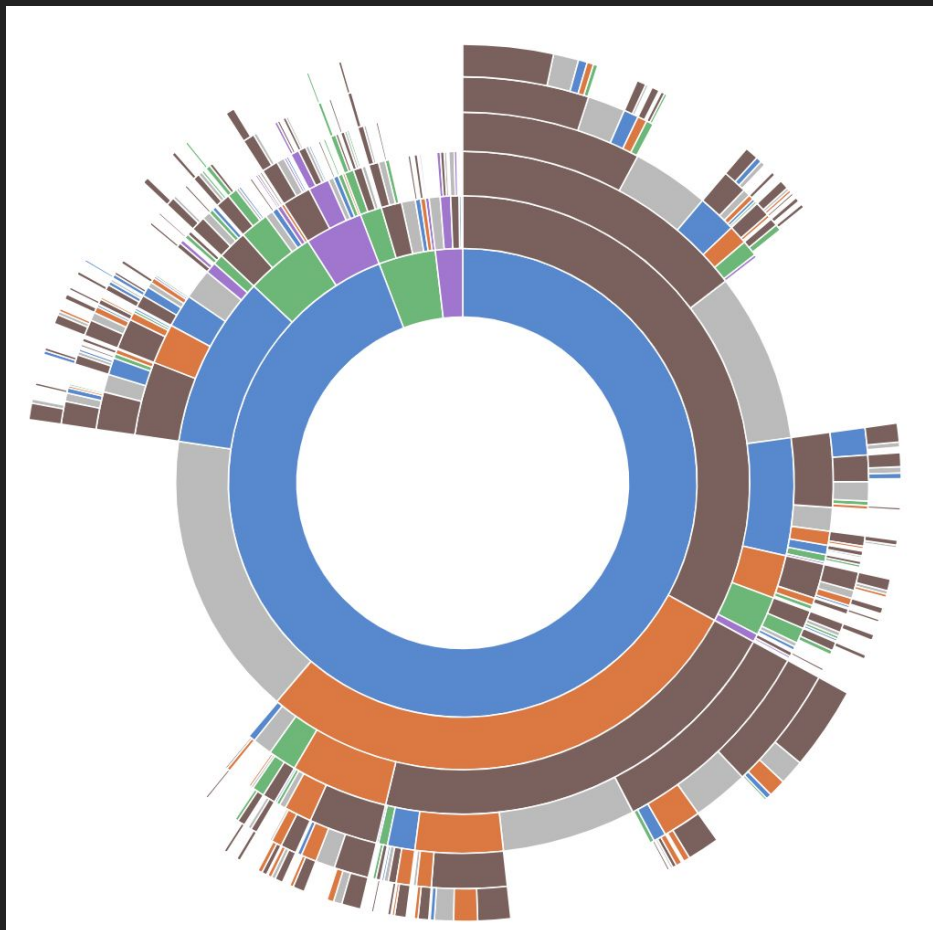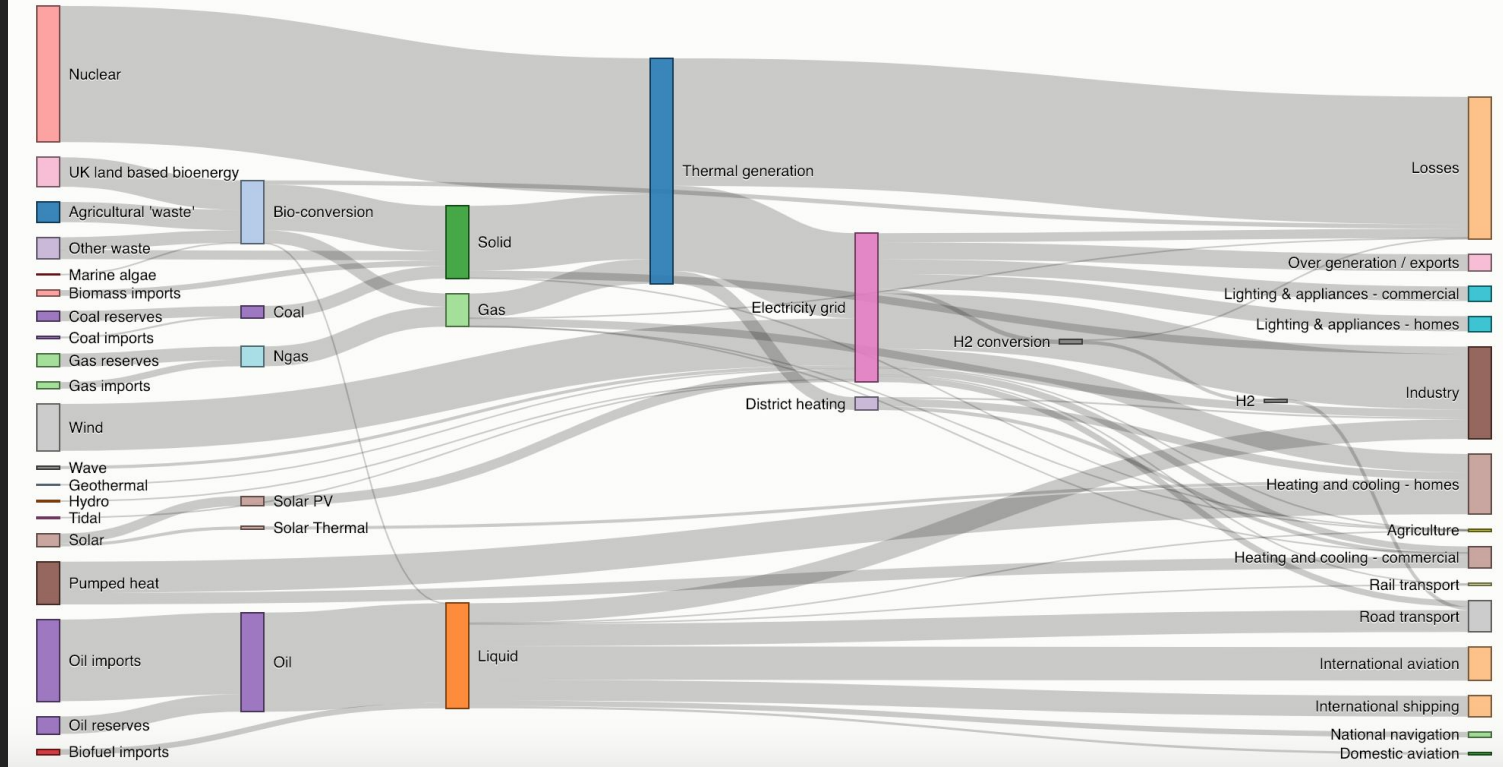
Color hue

Enable accurate estimates

May enable general estimates

Based on Cleveland and Robert McGill (1984)

Comparison

Variable Width Column Chart — Two Variables per Item

Table or Table with Embedded Charts — Many Categories

Bar Chart — Many Items

Column Chart — Few Items

Few Categories

One Variable per Item

Among Items

Circular Area Chart — Cyclical Data

Line Chart — Non-Cyclical Data

Many Periods

Column Chart — Single or Few Categories

Line Chart — Many Categories

Few Periods

Over Time

What would you like to show?

Relationship

Scatter Chart — Two Variables

Bubble Chart — Three Variables

Distribution

Single Variable — Few Data Points — Column Histogram

Many Data Points — Line Histogram

Two Variables — Scatter Chart

Three Variables — 3D Area Chart

Composition

Changing Over Time

Few Periods

Only Relative Differences Matter — Stacked 100% Column Chart

Relative and Absolute Differences Matter — Stacked Column Chart

Many Periods

Only Relative Differences Matter — Stacked 100% Area Chart

Relative and Absolute Differences Matter — Stacked Area Chart

Static

Simple Share of Total — Pie Chart

Accumulation or Subtraction to Total — Waterfall Chart

Components of Components — Stacked 100% Column Chart with Subcomponents

https://bl.ocks.org/kerryrodden/7090426
https://ai.google/research/pubs/pub43261

Sankey Diagrams

https://bost.ocks.org/mike/sankey/
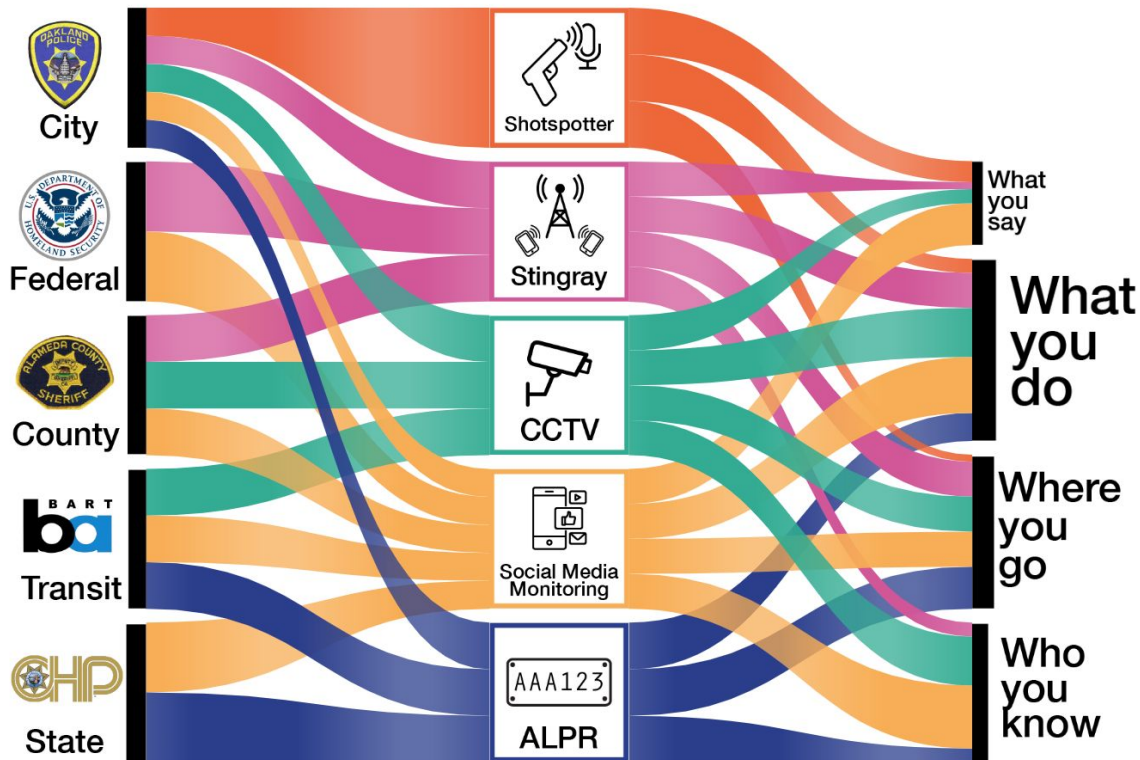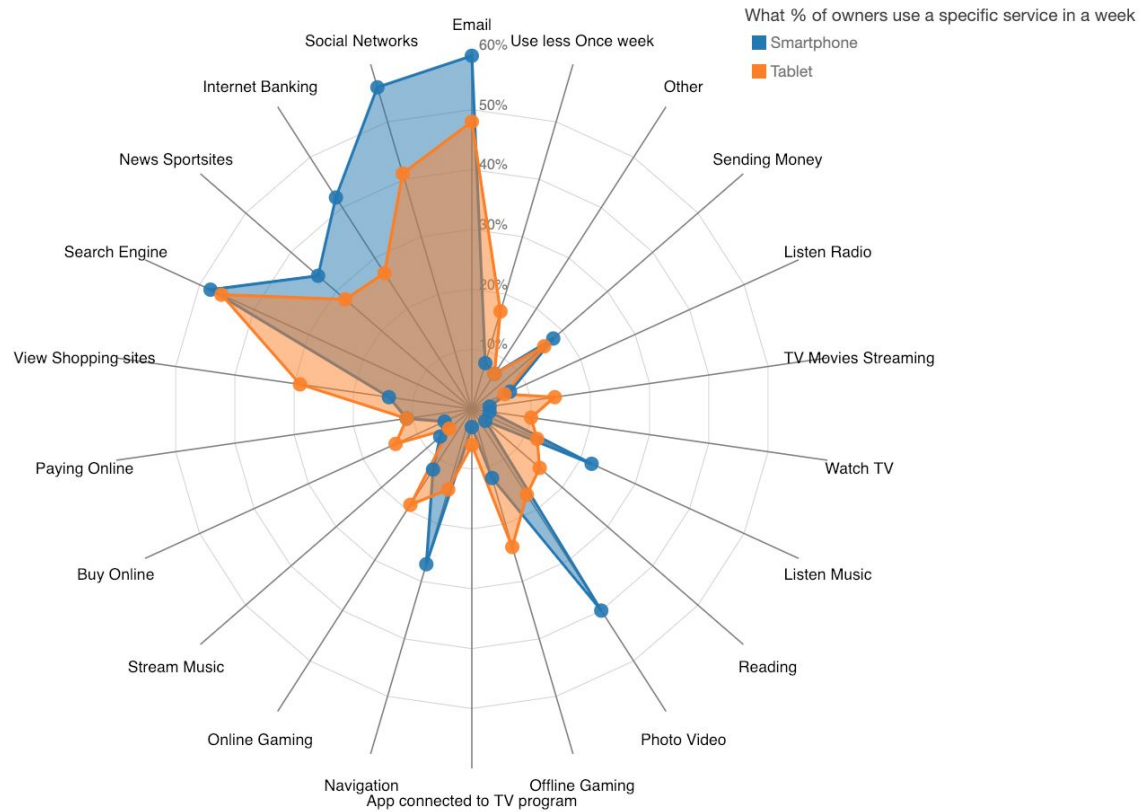
Acoustic Detection Systems (ShotSpotter) & You:

ShotSpotter is a surveillance tool that captures audio from the surrounding environment in order to detect and locate gunshots. The system can be triggered by other loud noises and capture other sounds including human voice.

https://strushucb.github.io/opp/

What % of owners use a specific service in a week
- Smartphone
- Tablet

Email
Social Networks
Internet Banking
News Sportsites
Search Engine
View Shopping sites
Paying Online
Buy Online
Stream Music
Online Gaming
Navigation
App connected to TV program
Offline Gaming
Photo Video
Reading
Listen Music
Watch TV
TV Movies Streaming
Listen Radio
Sending Money
Other
Use less Once week

http://bl.ocks.org/nbremer/6506614

Embedding Projector

DATA

1 tensor found
My tensor

☑ Sphereize data ❓

Load data   Publish

Checkpoint:
Metadata: https://gist.githubusercontent.com/tpipla...
             ...705..401445051...21...6...2...4742.4/

T-SNE   **PCA**   CUSTOM

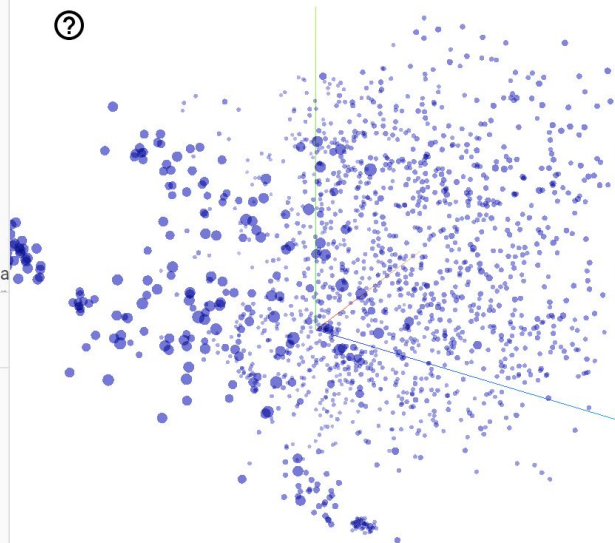X   Y
Component #1   Component #2

Z
Component #3   ☑

PCA is approximate. ❓

Total variance described: 23.2%.

❓ | Points: 1661 | Dimension: 300

Show All Data   Isolate selection   Clear selection

by
Search   .*   label

BOOKMARKS (0) ❓

https://mevey.github.io/EmojiVerse/
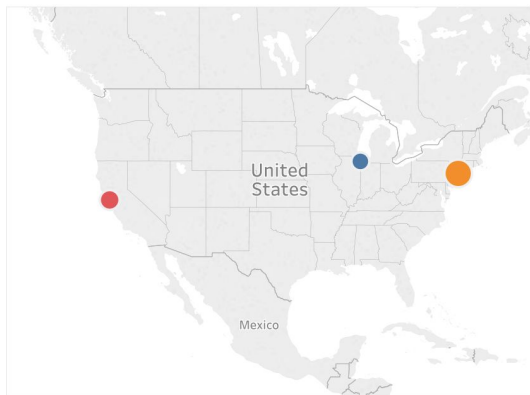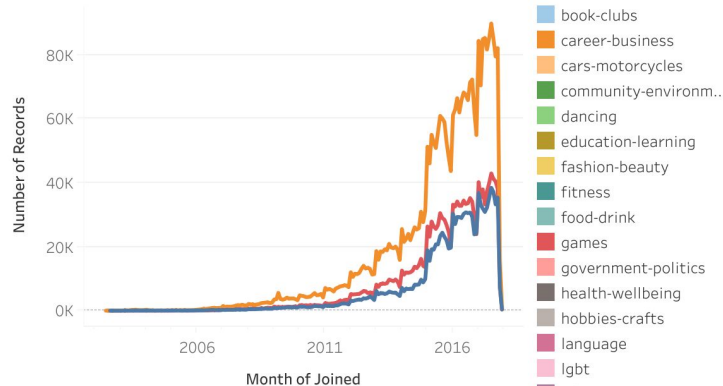
Let's Explore the meetup data! Click on the cities on the map to filter, or the different category subnames. *Questions? michelle.carney@berkeley.edu*
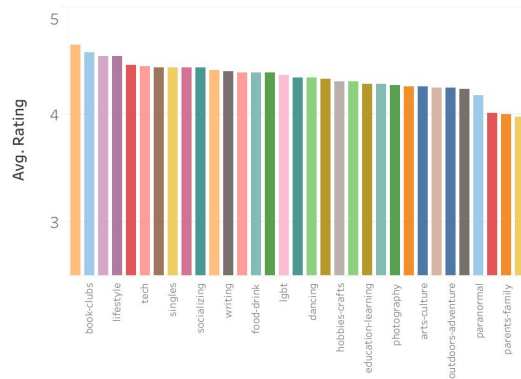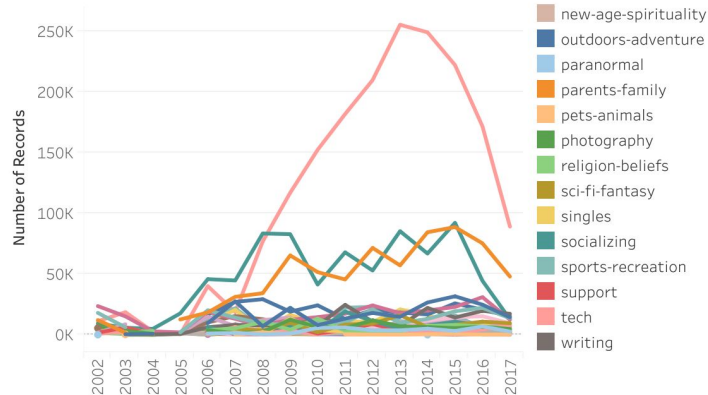
Major Meetup Cities

Membership Growth Overtime by City

Category.Shortname
- arts-culture
- book-clubs
- career-business
- cars-motorcycles
- community-environm..
- dancing
- education-learning
- fashion-beauty
- fitness
- food-drink
- games
- government-politics
- health-wellbeing
- hobbies-crafts
- language
- lgbt
- lifestyle
- movies-film
- music
- new-age-spirituality
- outdoors-adventure
- paranormal
- parents-family
- pets-animals
- photography
- religion-beliefs
- sci-fi-fantasy
- singles
- socializing
- sports-recreation
- support
- tech
- writing

Average Rating by Category

Membership by Category Over Time

Interested in learning more?

Alberto Cairo - Functional Art
http://www.thefunctionalart.com/

Stephen Few - Now You See It
http://www.stephen-few.com/nysi.php

Kerry Rodden
http://www.rodden.org/kerry

Tony Chu
http://www.r2d3.us/visual-intro-to-machine-learning-part-1/