Clustering

Data Science: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM DAVE BLEI AND LAUREN HANNAH
Lecture for Today

- What is clustering?
- K-Means
- Gaussian Mixture Models
Clustering

**Classification:** what is label of new point?

**Clustering:** how should we group these points?

**Clustering:** or is this the right grouping?

**Clustering:** what about this?
Uses:

- genomics
- medical imaging
- social network analysis
- recommender systems
- market segmentation
- voter analysis
Microarray Gene Expression Data

From: “Skin layer-specific transcriptional profiles in normal and recessive yellow (Mc1re/Mc1re) mice” by April and Barsh in *Pigment Cell Research* (2006)
Medical Imaging (MRIs and PET scans)

From: “Fluorodeoxyglucose positron emission tomography of mild cognitive impairment with clinical follow-up at 3 years” by Pardo et al. in Alzheimer's and Dementia (2010)
Social Networks

Twitter Social Network, 20K nodes 250K edges

Aeromental
chrisbrogan

JasonCalacanis
bloggersblog

Forecast

steverubel
ap1978

Webtickle
chrispirillo

jelmpulver
macworld

SteveJobs

grum
prietoj

Michoaca2

TUAW

ambermacarthur

johnedwards

woet

Data Science: Jordan Boyd-Graber

Clustering
Recommender Systems

From: tech.hulu.com/blog/
Market Segmentation

Segment: Milk and Cookies
Life Stage: Family Portrait
Primary Housing: Single Family
Primary Family Type: MC with Children
Average HH Size: 3
Median Age: 32.7
Median Income: 553300

Segment: Aspiring Young Families
Life Stage: High Hopes
Primary Housing: MultiUnit, Single Family
Primary Family Type: Family Mix
Average HH Size: 2.6
Median Age: 29.8
Median Income: 44900

Segment: Prosperous Empty Nester
Life Stage: Senior Styles
Primary Housing: Single Family
Primary Family Type: Married Couple
Average HH Size: 2.4
Median Age: 46.1
Median Income: 64100

Segment: Young and Restless
Life Stage: Solo Acts
Primary Housing: Multi-Units
Primary Family Type: Single Person, Shai
Average HH Size: 2
Median Age: 29.1
Median Income: 38000

Segment: Cozy and Comfortable
Life Stage: Upscale Avenues
Primary Housing: Single Family
Primary Family Type: Married Couples
Average HH Size: 2.6
Median Age: 40.2
Median Income: 58900

Segment: Industrious Urban Fringe
Life Stage: Global Roots
Primary Housing: Single Family
Primary Family Type: Married Couples
Average HH Size: 3.4
Median Age: 28.6

From: mappinganalytics.com/map/segmentation-maps/segmentation-map.html
Voter Analysis

- soccer moms (female, middle aged, married, middle income, white, kids, suburban)
- Nascar dads (male, middle aged, married, middle income, white, kids, Southern, suburban or rural)
- security moms ( ... )
- low information voters
- Ivy League Elites