Interactions II: Recommendations

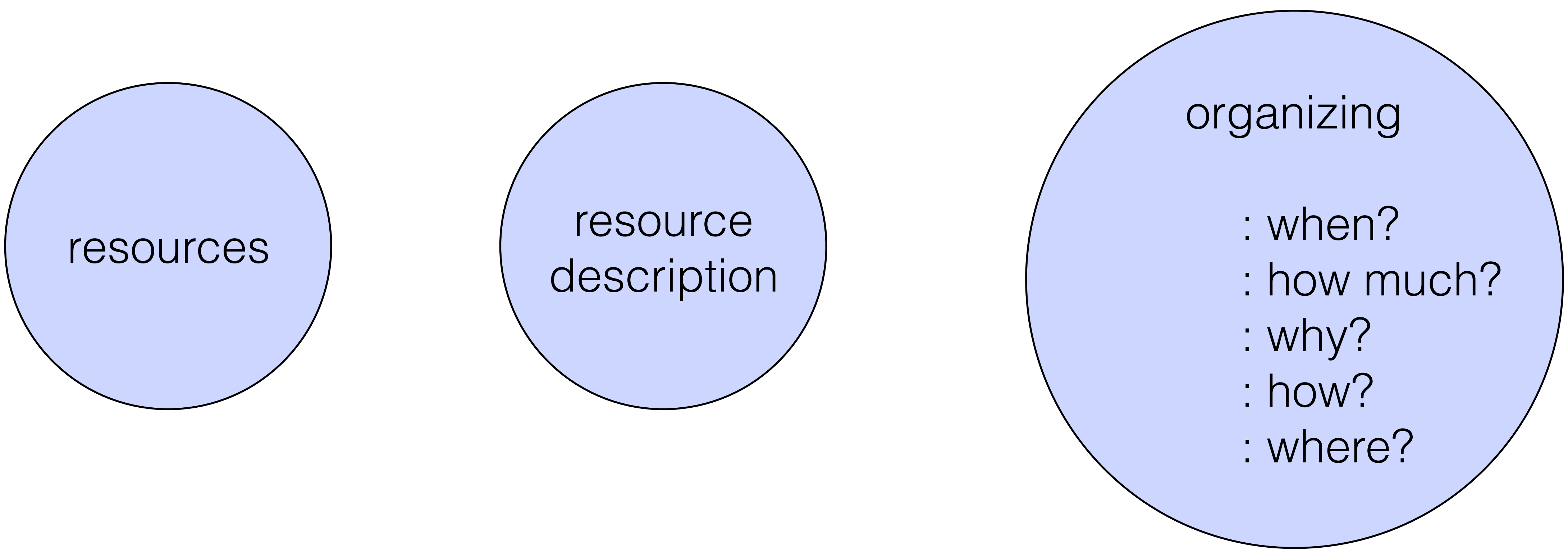
David Bamman
Info 202: Information Organization and Retrieval

Nov. 21, 2016

We design organizing systems because we have some interaction in mind

Recommendation

- Providing recommendations is an interaction that's enabled by organizing systems




smell·bound ('smel-ˌbau̇nd) adj : held as if under a spell by the scent of books
NEW FAVES
EXIT
One Kick
Chelsea Cain
Tom Holt
California
The Magician's Land

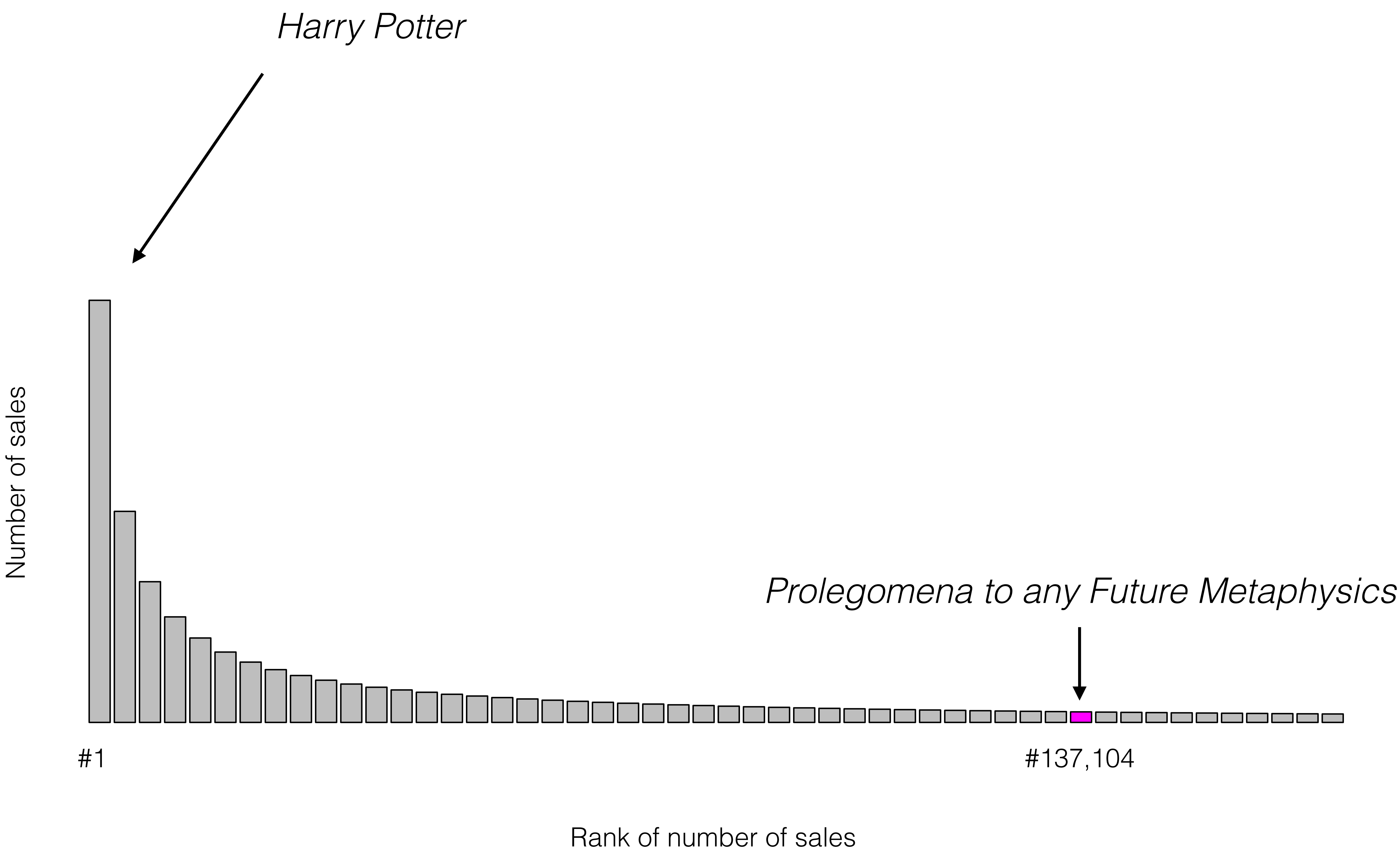

PRODUCE
FRESH
WINE
5.99
EA.

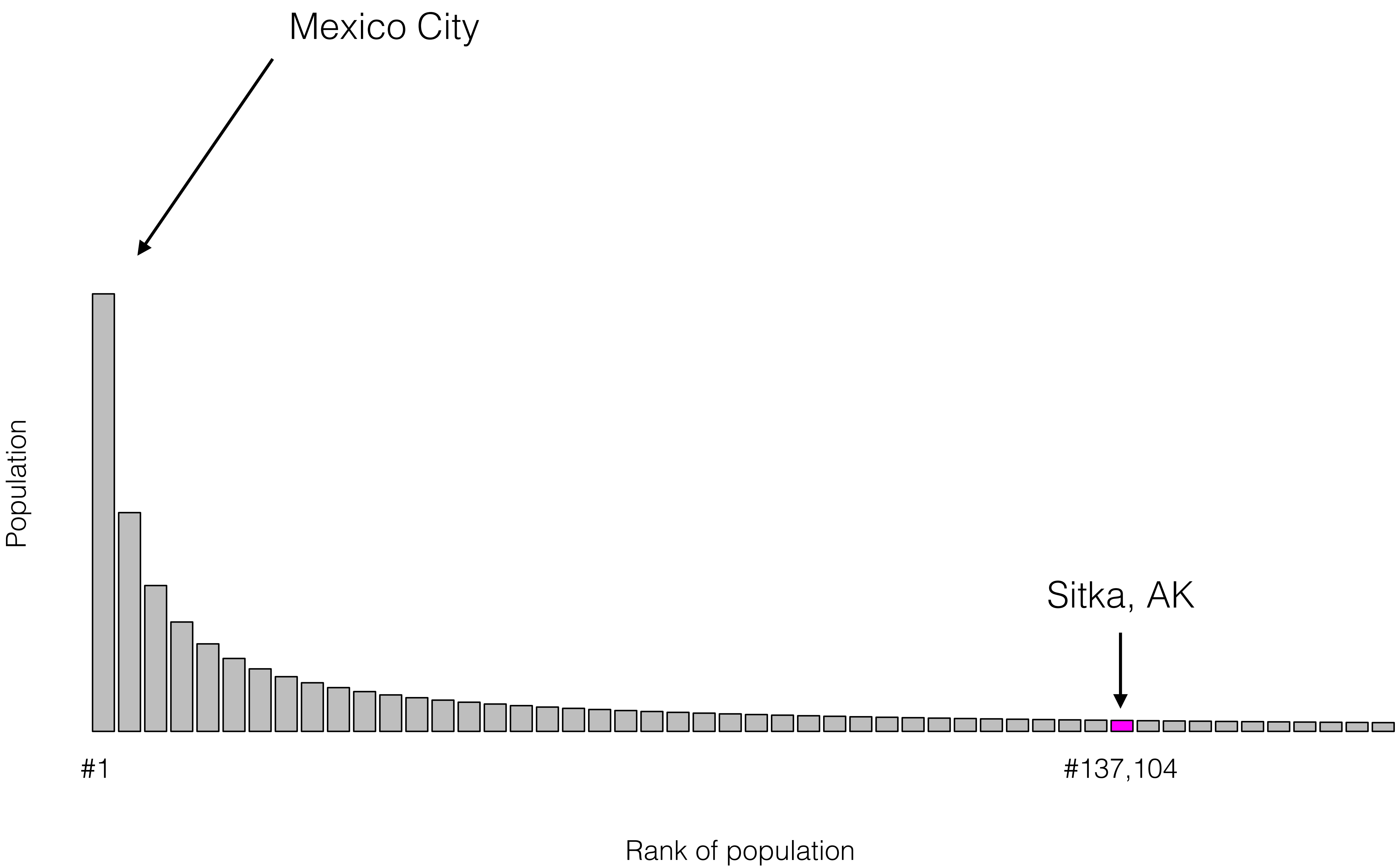
Recommendations

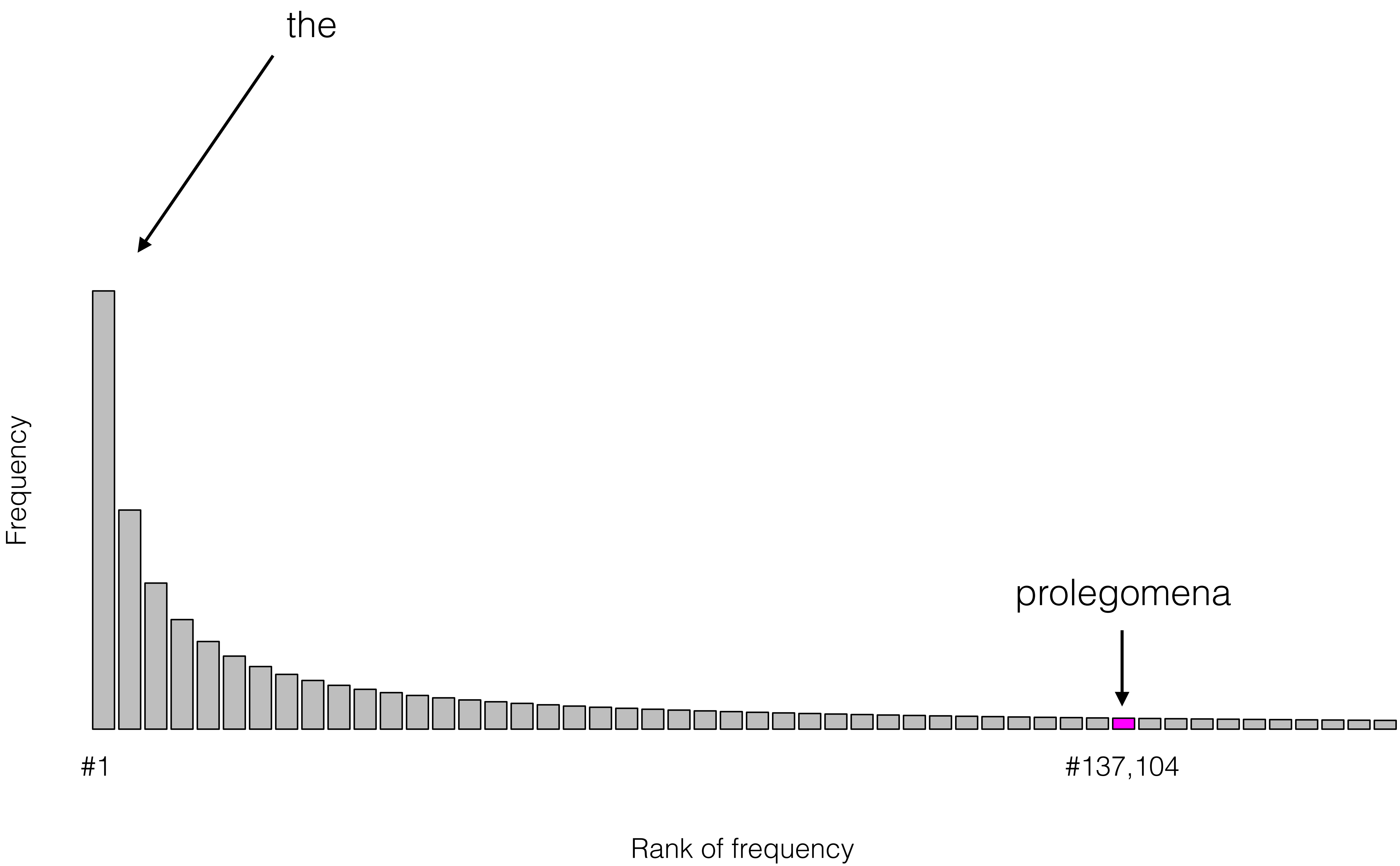
- Physical organizing systems mainly make implicit recommendations at the aggregate level
- Organizing principle #1: promote books that have the highest expected sales among all customers.
- Organizing principle #2: staff recommends books they like.

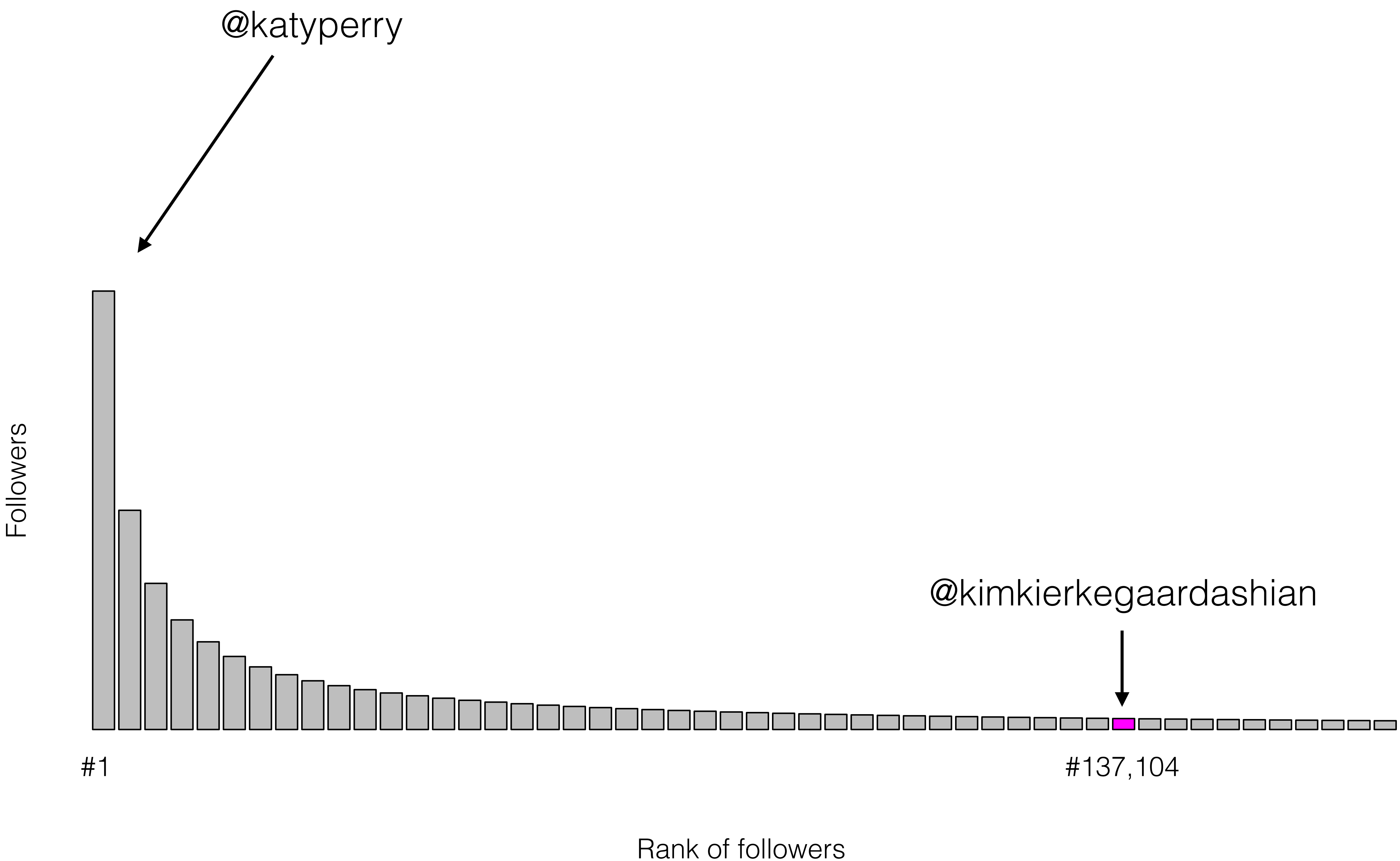
Zipf's law

- For some phenomena, there's a relationship (power law) between the frequency of an event and the rank of that frequency among all events.
  - Social network degree centrality
  - Populations of cities
  - Word frequency
  - Sales


Harry Potter
Prolegomena to any Future Metaphysics
Number of sales
#1
#137,104
Rank of number of sales


Mexico City
Sitka, AK
Population
#1
#137,104
Rank of population


the
Frequency
prolegomena
#1
#137,104
Rank of frequency

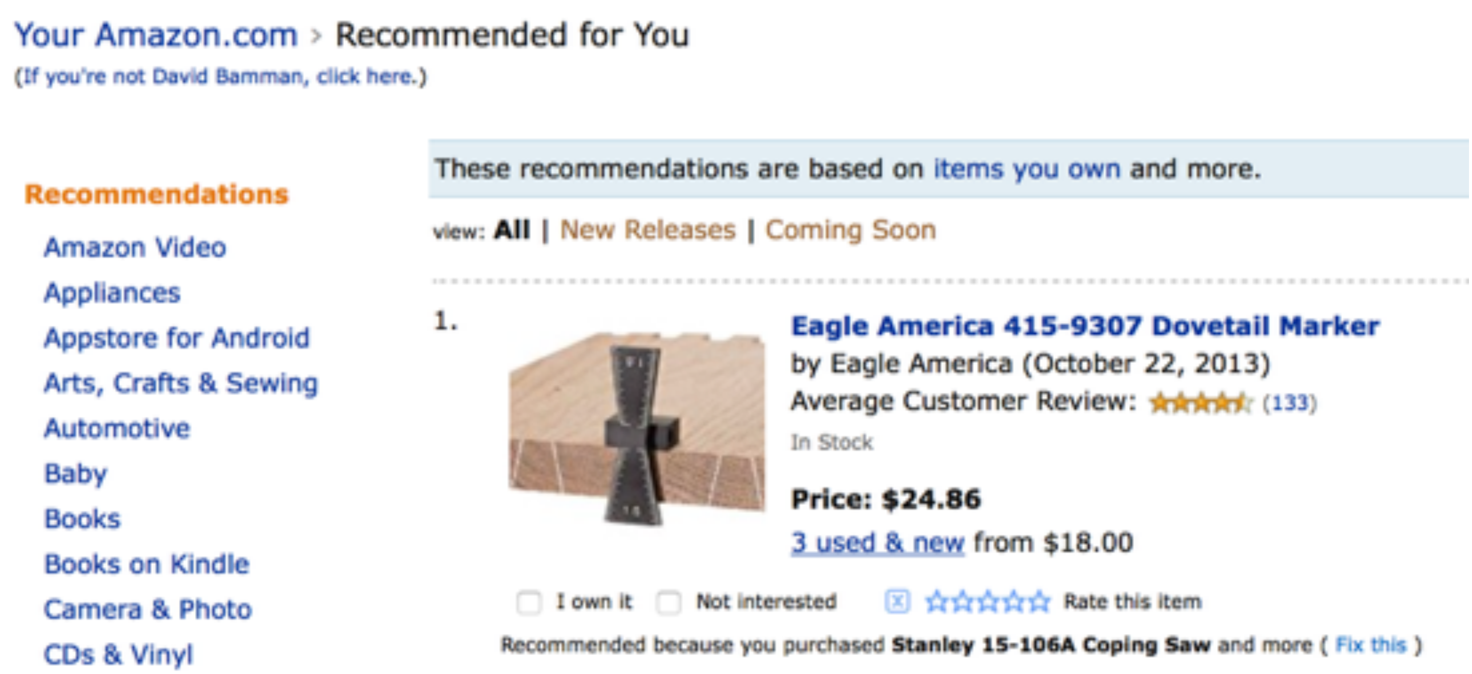

@katyperry
@kimkierkegaardashian
Followers
#1
#137,104
Rank of followers

Long tail

- Aggregate stats (e.g., “bestsellers”) work well for the few items in the frequent end of the tail

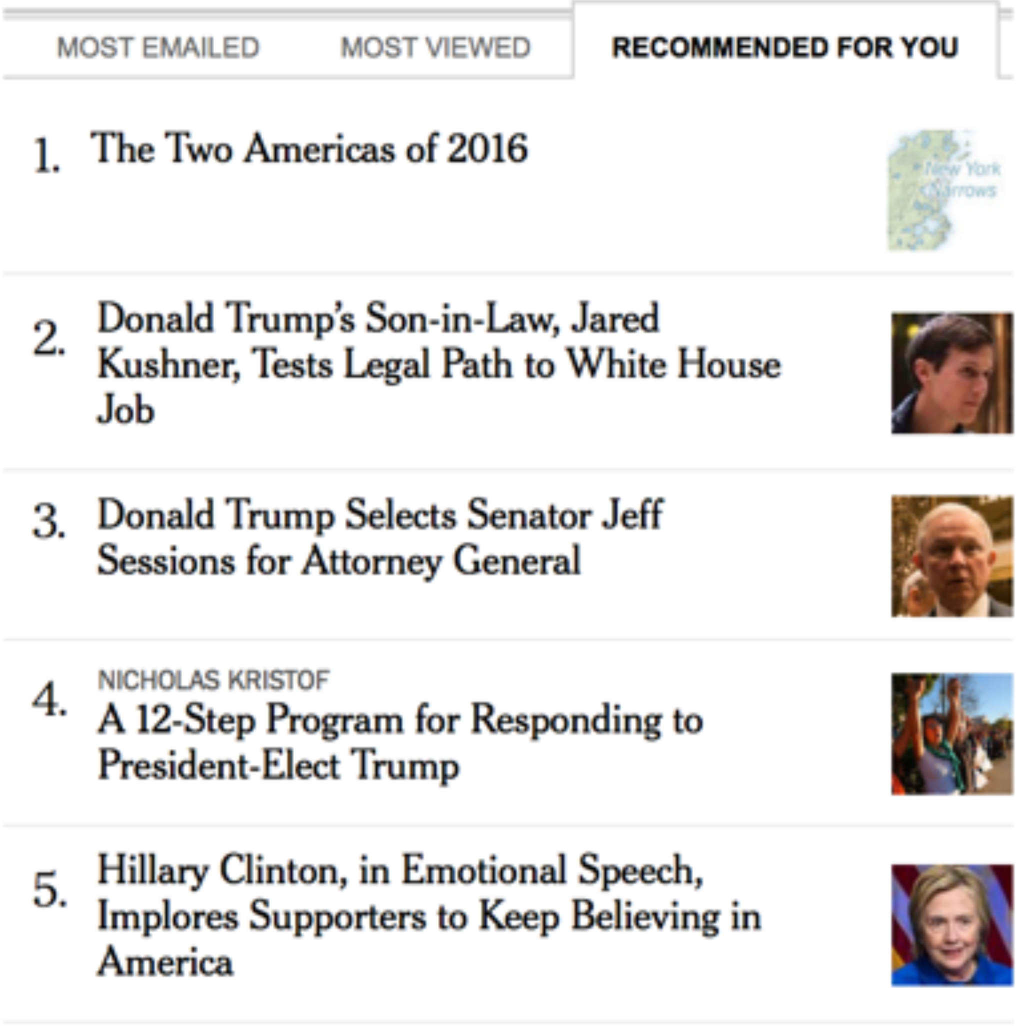
- When there’s a long tail of items with few people who care about them, there’s a lot of be gained by highly customized recommendations

Netflix



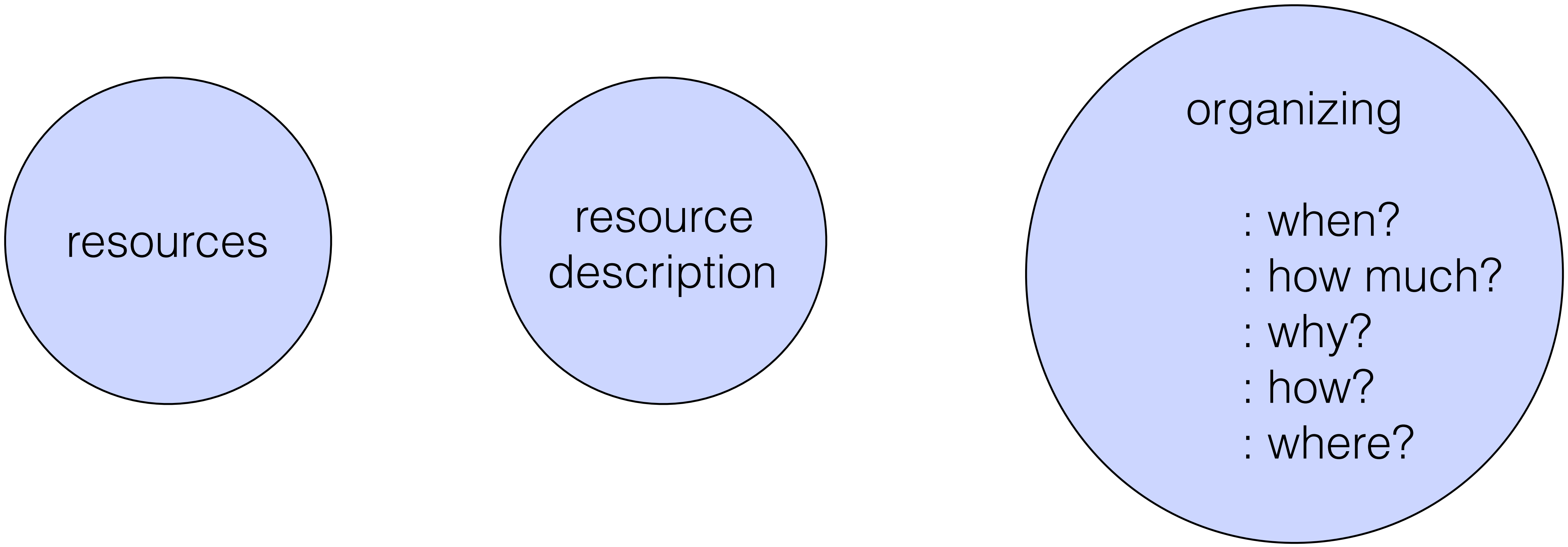
Amazon

Twitter



New York Times

Recommendations via DS



- Automatic recommendations draw on classification, clustering, description, structure

case study: recommendation systems

- Many resources we can marshall to make this prediction.
    - Descriptions of the items themselves
    - Data points given to us by company catalog
    - But considerable flexibility in resource description

case study:
recommendation systems

- Many resources we can marshall to make this prediction.
    - Users who rate movies
    - Recommend movies through the relationships they hold to the people who watch them.

Utility matrix

	Ann	Bob	Chris	David	Erik
Star Wars	5	5	4	5	3
Bridget Jones		4		4	1
Rocky	3		5		
Rambo		?		2	5

How do we get ratings from users?

Methods

- Content based nearest neighbors
- Classification
- Collaborative filtering

Content-based nearest neighbors

- Basic idea: Represent a user's features as the average value of those in the movies they like

- Compare that user representation with each movie to find ones that are most similar

mark hamill	TRUE
harrison ford	TRUE
ben affleck	FALSE
runtime (mins)	121
language=English	TRUE
langauge=Spanish	FALSE
space opera	TRUE
cartoon	FALSE

mark hamill	1
harrison ford	1
ben affleck	0
runtime (mins)	121
language=English	1
langauge=Spanish	0
space opera	1
cartoon	0

	star wars	star wars II	gone girl	Average
mark hamill	1	1	0	0.66
harrison ford	1	1	0	0.66
ben affleck	0	0	1	0.33
runtime (mins)	121	124	149	131.3
language= English	1	1	1	1
language= Spanish	0	0	0	0
space opera	1	1	0	0.66
cartoon	0	0	0	0

Cosine Similarity

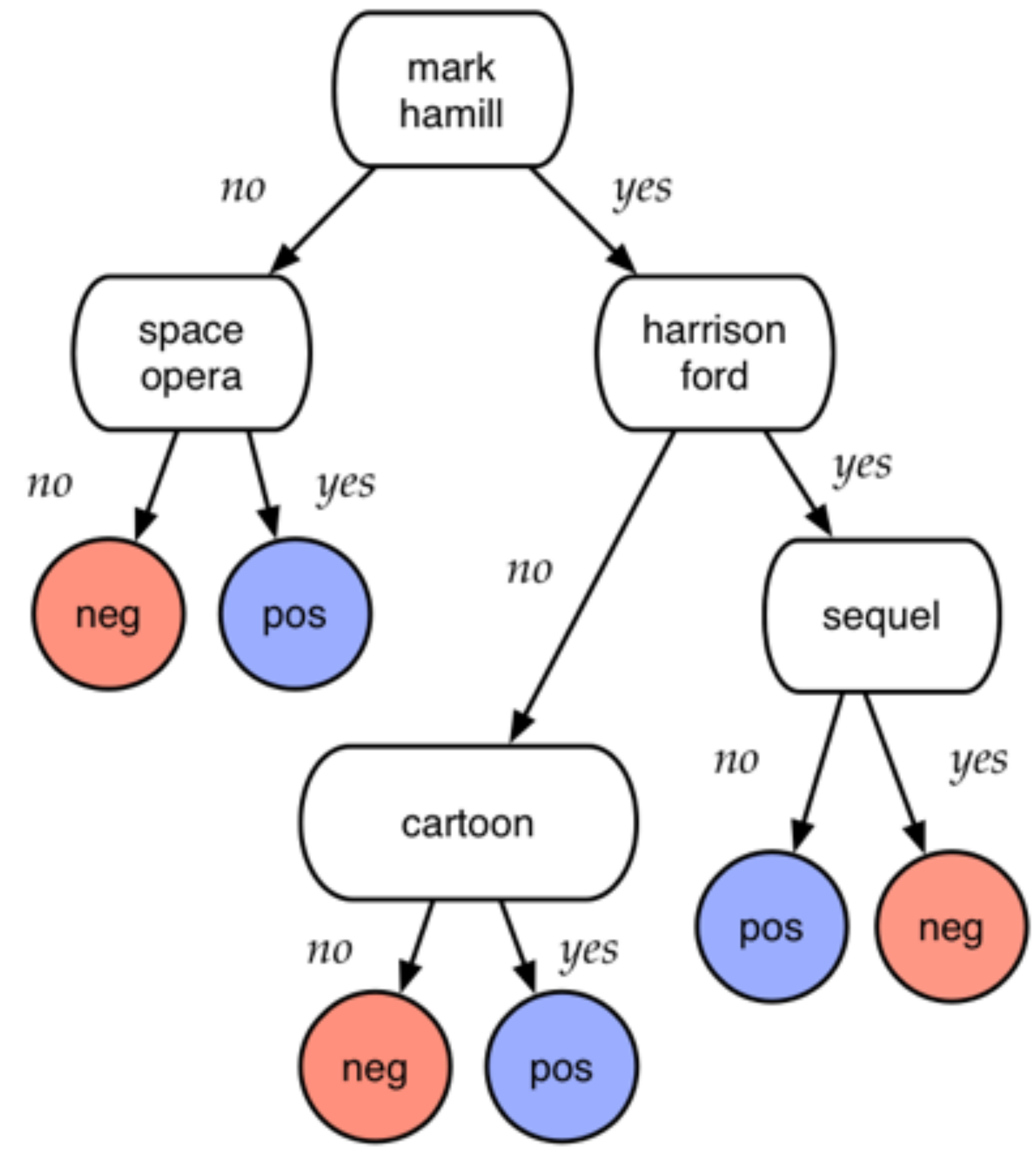
$$cos(x,y) = \frac{\sum_{i=1}^{F} x_i y_i}{\sqrt{\sum_{i=1}^{F} x_i^2}\sqrt{\sum_{i=1}^{F} y_i^2}}$$

- Jaccard similiarty is measure of set overlap.
- Cosine similarity reasons over the value of features (cf. TDO 7.3.6.2)
- Often weighted by TF-IDF to discount the impact of frequent features (cf. 10.4.2.1)

Classification

- Basic idea: train a separate classifier for each user based on their current ratings

- Insight: reassess movies with no rating



Content-based classification

- Content-based recommendation (whether through nearest neighbors or classification) is plagued by data sparsity

- Doesn't consider the way in which other people have rated movies and the structure that exists between them.

	Ann	Bob	Chris	David	Erik
A				5	
B					
C			5		
D					5
E					
F				4	
G					
H					
I					
J					
K					
L			4		
M					
N	3				
O					
P					
Q					
R					
S				2	
T		?			

Collaborative filtering

- Basic idea: rather than recommending based on an item's content (resource description), we'll recommend based on patterns in other user's ratings (and the similarity between users).

- Exploit the assumption that users' tastes have structure

- Learn that if users like A, then they often also like B.

Collaborative filtering

- Two ways we can do this:
  - User-user similarity
  - Item-item similarity

Collaborative filtering

	Ann	Bob	Chris	David	Erik
Star Wars	5	5	4	5	3
Bridget Jones		4		4	1
Rocky	3		5		
Rambo		?		2	5

User-user similarity

1. Represent each user by the movie they’ve rated

2. Identify the K nearest neighbors (e.g., the K users with the highest cosine similarity)

3. Make a predicted rating about an item by averaged those K users’ scores (if they’ve rated it).

	Ann	Bob	Chris	David	Erik
Star Wars	5	5	4	5	3
Bridget Jones		4		4	1
Rocky	3		5		
Rambo				2	5

User-user similarity

	Ann	Bob
Star Wars	5	5
Bridget Jones	0	4
Rocky	3	0
Rambo	0	0

$$cos(x, y) = \frac{\sum_{i=1}^{F} x_i y_i}{\sqrt{\sum_{i=1}^{F} x_i^2}\sqrt{\sum_{i=1}^{F} y_i^2}}$$

Item-item similarity

1. Represent each item by the users who've rated it.

2. Identify the nearest neighbor (e.g., by cosine similarity) to an item that a given user has rated highly

	Ann	Bob	Chris	David	Erik
Star Wars	5	5	4	5	3
Bridget Jones		4		4	1
Rocky	3		5		
Rambo				2	5

Tradeoffs

- Level of granularity
- Users like mixtures of many different kinds of things (multiple movie or music genres, for example) → increase the breadth of recommendations.
- Items often only belong to one genre → increase the precision of recommendations.

Matrix decomposition

- More complex methods explicitly encode the assumption that items and users both contain latent features.
- e.g., “movies with happy endings” — we may not ever see it represented as a feature, but it would explains a lot of the commonalities in how different users rate them.

Matrix decomposition

	Ann	Bob	Chris	David	Erik
SW	5	5	4	5	3
Jones		4		4	1
Rambo	3		5		
Rocky				2	5

=

	F1	F2
SW	0.67	1.3
Jones	-1.4	0.1
Rambo	3.12	0.11
Rocky	-1.3	-0.2

X

	Ann	Bob	Chris	David	Erik
F1	1.7	3.1	-0.7	8.3	-4.5
F2	0.1	-0.2	1.3	7.4	-3.4

Matrix decomposition

- With this (reduced) representation, we can perform the same user-user or item-item queries as before.

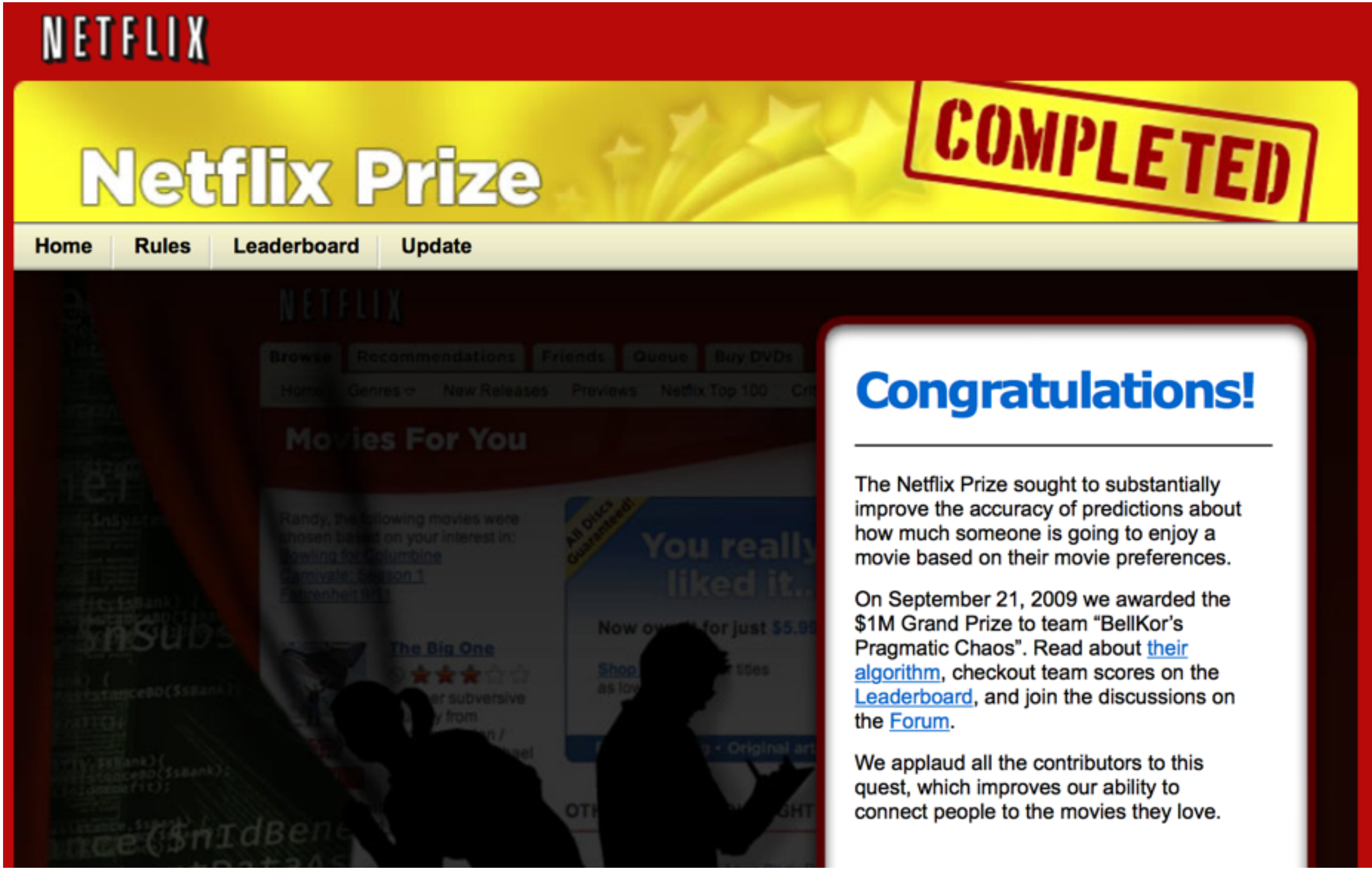
	F1	F2
SW	0.67	1.3
Jones	-1.4	0.1
Rambo	3.12	0.11
Rocky	-1.3	-0.2

X

	Ann	Bob	Chris	David	Erik
F1	1.7	3.1	-0.7	8.3	-4.5
F2	0.1	-0.2	1.3	7.4	-3.4

Latent variables

	observed variables	latent variables
email	text, date, sender	
novels		
social network		
fitbit data		
legislators		
netflix users		


NETFLIX
Netflix Prize
COMPLETED
Home
Rules
Leaderboard
Update
Congratulations!
The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.
On September 21, 2009 we awarded the $1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about their algorithm, checkout team scores on the Leaderboard, and join the discussions on the Forum.
We applaud all the contributors to this quest, which improves our ability to connect people to the movies they love.

Recommendations in an organizing system

- what is being organized?
- why is it being organized?
- how much is it being organized?
- when is it being organized?
- how (or by whom) is it being organized?
- where is it being organized?

- Resources: products (movies, groceries) and the users/customers who interact with them.
- Resource description: deciding what properties of the data we want to use in defining similarity.
- Classification, clustering, latent variable modeling as interactions to support the end goal