Math, Engineering, Sociology

- **Netflix prize in 2009**
  - ➤ **Beat the system, win**
  - ➤ **http://nyti.ms/sPvR**

Collaborative Filtering

- **How does Amazon know what I want?**
  - ➤ **Lots of customers, lots of purchases**

- **How does Pandora know music like Kanye's?**
  - ➤ **This isn't really collaborative filtering, more content-based**

- **How doe Netflix recommend movies?**
  - ➤ **Why did they offer one million $$ to better their method?**

- **Students at Duke who like Compsci also like …**
  - ➤ **Could this system be built?**

From User Rating to Recommendations

21 Jump Street	Lorax	John Carter	Friends with Kids	The Vow
3	-3	5	-2	-3
2	2	3	2	3
4	4	-2	1	-1

- **What should I choose to see?**
  - ➤ **What does this depend on?**
- **Who is most like me?**
  - ➤ **How do we figure this out**

How do we store/use tabular data?

- **How do we access data in FriendScore APT?**
  - ➤ **What does `friends[i][j]` mean?**
    - • **What is `friends[i]`, what is `s[j]` where s is a string?**

- **What is a matrix?**
  - ➤ **Table of numbers? Entries? Two-Dimensional, m X n**
  - ➤ **Many arithmetic operations available**
  - ➤ **Convenient for representing correlated data**

- **In Python what is a list of lists?**
  - ➤ **What is an array? Multi-dimensional?**

If we want to find the best match…

- ● **Assign a number to each possible match**
  - ➢ What's the best match among the numbers?
  - ➢ What about finding the nearest k and using them?

- ● **We need to sort the data to find closest matches**
  - ➢ Differences between lst.sort() and sorted(lst)
  - ➢ How do we change criteria used in sorting?
  - ➢ What sorting algorithm is used, do we need to know this?

- ● **How do assign a number to each possible match**
  - ➢ Rankings, ratings, questions, indirect measures, …

Zephyr Teachout

- ● **Fordham Law Prof**
  - ➢ Duke law school

- ● **Internet Campaign**
  - ➢ Howard Dean 2004
  - ➢ http://cnet.co/GAGR0V

We have so many different parts of our site, right now. We actually build software here. The Internet is really essential for people to achieve some more power with the process.

We have three full-time programmers and our database team has people with (scripting language) PHP and we have a bunch of volunteer coders, as well as a whole community of over 100 people working in open source on Dean-related projects. I am not the person that knows hardware.
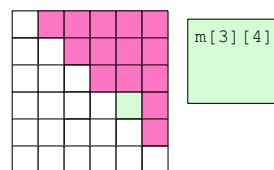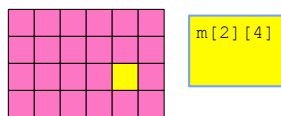

APT Aside

- ● **Common theme to these APTs**
  - ➢ Nested loops
  - ➢ While True: … break

```
for i in range(4):        for i in range(6):
   for j in range(6):        for j in range(i+1,6):
      elt = m[i][j]             elt = m[i][j]
```

m[2][4]

m[3][4]

When to stop iterating?

- ● **Definite loops are "simple"**
  - ➢ for elt in collection: …
  - ➢ for i in range(a,b): …

- ● **What about indefinite loops?**
  - ➢ while attempt < max_allowed: ...
  - ➢ while True: ... break ...

- ● **Example: Thesaurus APT, InternetSecurity APT**
  - ➢ When to stop iterating? When a condition is true
  - ➢ ok = False, … compute … ok = True … if ok: break