Overview of the day

- Announcements
- Final project presentation details
- Final project report details
- Extra credit assignment details
- A4: Crowdwork Ethnography reflection
- Dinner break
- Week 9 reading reflections
- Lecture: Design + HCDS (+ break?)
- Homework review
Please fill out the course evaluation survey! It is open now, and closes December 8: https://uw.iasystem.org/survey/181661
Final project presentation

4 minute oral presentation (10 points)
Slides due: next Thursday, December 7 before class
Submission instructions: submit Canvas link to Google Slides or PDF in Google Drive

Your presentation should demonstrate the following:

- Your ability to present effectively to a professional audience. *Imagine that you are pitching your project to directors/execs at a company you work for.*
- Your ability to communicate the importance of your research to the specified audience.
- Your ability to communicate the implications of your findings accurately and compellingly.
- Your ability to do this in a very short time (*hint: practice beforehand and time yourself!*)

*Important note:* you’ll be interrupted at 4 minutes, and cut off at 4:30. You’ll be graded on how well you completed the assignment requirements before you got cut off.
Final project presentation logistics

- **Presentation order will be random.** You won’t know ahead of time when you’ll be presenting.

- **Time will be tight!** We’re budgeting 2 minutes for transition between speakers. 40 presentations \( \times \) 6 minutes each = 4 hours :/

- **I will be timing each of you.** Using a timer. And I will be annoying about it.

- **You’ll be interrupted at 4 minutes, and cut off at 4:30.** You’ll be graded on how well you completed the assignment requirements before you got cut off.

- **I would like to bring in some coffee/tea and snacks** to help keep everyone awake and happy. I’ll throw in $20. If you donate $10 or more you get to help decide what we order!

---

**In class participation points:** You will be assigned to provide presentation feedback for 2 random colleagues. The feedback can be on presentation style or content. Keep it constructive and polite. Feedback will be posted to Canvas.
Final project report

Written report w/ code and data. At least 1000 words (not including code). 15 points.

**Due date:** Sunday, December 10 at 11:59pm

**Submission instructions:** Submit Canvas link to Github repo w/ notebook and data

The repo should include:

- A Jupyter notebook that contains both your written report and your code
- Your data, or a sample of your data, that can be run in your notebook
- A short written abstract of your study that describes what you did, and what you found.
- An MIT LICENSE file for your code
- A README that contains
  a. Documentation and license information for your data, etc. (same basic reqs as A1 and A2)
  b. Hyperlinks to ALL relevant resources (TOS, API documentation, license deeds, etc)
What the notebook should contain

Your report will be written inside your Jupyter Notebook. You will need to properly document and describe your code at each step, and structure your report in a readable way.

It is up to you whether you want to embed all the analysis code in the body of the report (example), or put the report first and the code after. Either way, every step in your data munging/analysis process needs to be documented in markdown cells, not inline comments.

Clear section headings are required. Here’s one good way to structure your report:

- Introduction
- Background (or Related Work)
- Methods
- Findings
- Discussion (or Implications)
- Conclusion
- References
• **Introduction:** Why is this analysis interesting or important (to people besides you)? Does it solve a real problem or tackle an unresolved research question?

• **Background/Related Work:** What other research has been done in this area? How does this research inform your hypotheses, your analysis, or your system design? What are your hypotheses or research questions?

• **Methods:** Not just your analytical methods: also, why you chose them, and how human-centered considerations such as ethics informed the way you designed your study.

• **Findings:** What did you find? Use words and figures, don’t just point to code.

• **Discussion/Implications:** Limitations of your study (*this is required!*); Why what you found is important; How could future research build on this study?

• **Conclusion:** Restate your research questions/hypotheses and summarize your findings; Explain to the reader how this study informs the *their* understanding of HCDS

• **References:** Any publications (blogs, articles, research papers) you refer to in the text.
Write a ‘meta-reflection’ of four HCDS position papers. Up to 5 points. At least 500 words.
Due date: Next Thursday, December 7 at 4:59pm
Submission instructions: Submit to Canvas discussion “Extra credit reading reflections”

In your meta-reflection cover the following points:

1. why you chose these four papers
2. how these four papers *together* inform your understanding of HCDS
3. how each paper relates to other papers, or assignments from DATA 512

Note: You can’t pick Aragon, C. et al. (2016). *Developing a Research Agenda for Human Centered Data Science*. We’ve already covered that one.

All papers available here: https://cscw2016hcds.wordpress.com/papers/
Review of A4: Crowdwork ethnography
Awesome job!
- Every assignment I read was great
  - A couple missed chunks
  - Assignment text too long?
  - Assignment too long?
However

- We have noticed some people copying other works into their homework
- You are expected to write *everything* originally, with the exception of quotes, which should be identifiable as quotes.
  - Even dataset descriptions
  - Even background sections
  - Everything.
- If we spot any issues like this in your final reports, we are *required* to report you to the Graduate School.
- If you are uncertain:
  - Read [https://integrity.mit.edu/handbook/writing-original-work](https://integrity.mit.edu/handbook/writing-original-work)
  - Send Oliver and Jonathan your draft in advance
Weirdest Hit Award

HIT #9 – "Take Photo of a Person on the Floor 2"

To obtain this HIT I searched "images" and chose from the first ten returned on the page. The Requester was Safely You, the HIT expiration date was Feb 23, 2018, time allotted was 3 hours, and the reward was $0.07. There were no required qualifications and the description was, “Take a photo of a person lying on the floor, DO NOT show person’s face. Conditions for lighting can vary and photo must be taken at eye level or above.”

This was probably the oddest HIT that I came across, especially since there was a sample photo included in Figure 5.

![Figure 5: HIT #9 Sample](image)
Break (15 minutes)
Venmo to:
jonnymorgan.esq@gmail.com
Jonathan-Morgan-24
Week 9 reflections
This reading informs my understanding of human centered data science by illustrating how the more flexible/inclusive designs help people to not compromise on their beliefs. For example: while registering on a social site, a person doesn't have to choose between the binary gender categories. This helps them to retain their self defined identity while using the social media systems. For a data scientist, it’s imperative to be cognizant of the unconscious bias and using custom fields for gender is a step closer in removing the bias.

Question for discussion: Just as race has been eliminated while registering on some of the social sites, will it be effective going gender-less to eliminate gender-based discrimination? If yes, what are the pros and cons of gender-less platforms?
Companies like Linkedin and twitter do not require users to enter their gender but still infer gender using algorithms and send them to advertising companies? I am curious to know the accuracy of these algorithms - if they have poor performance, I suspect it can have widespread negative impact on the larger community?
I am mostly interested in the concept of intersectionality, which is originated from traditional sociology, and now is applied in human-computer interactions (HCI) studies. Intersectionality was first brought up to illuminate how in-equality impacted people based on the intersections of gender, race, and class rather than on any single one of these attributes. Today, in HCI studies, researchers would also like to study users based on the intersections. Gender, race, and class seems to produce abundant groups of people. What are the other possible attributes to bring in? Geo-location, age, education?
If the goal is making data science skills widely accessible, why not create video lectures and detailed, online tutorials? Even if you still run in the in-person classes for 403 students with the apprenticeship model, the online materials can potentially provide skills to hundreds of thousands.
Given the time constraints of the workshops/classes, has there been any considering towards developing follow-up curriculum so those that are interested have some direction in further pursuing the development of their data science/programming skills?
HCDS + design
Part 1: human-centered recommendation systems
(with some week 8 reflections)
Recommender systems

- One of the most familiar types of AI (thanks, Amazon and Facebook)
- Many different sources of signal to build from
  - E.g. user behavior, user demographics, item characteristics
- Many established AI-based approaches
  - E.g. Collaborative filtering, content filtering, naive Bayes, LSI
- ‘dumb’ approaches are also ubiquitous and can be effective
  - E.g. raw item popularity, recency
- Which approach to choose?
Approaches to rec sys design

Engineering-centric

- use whatever sources of data are easiest to access and process
- build whatever models provide the best performance and offline accuracy

Business-centric

- use whatever sources of data we have
- build whatever model provides the best ROI per pre-established metrics

Human-centric

- use sources of data that are ethically appropriate
- build models that are interpretable
- build models that take into account audience, purpose, and context
When listening to my Pandora station for older music, songs that I am familiar with always bring satisfaction whereas novel songs from that era even from the same artist does not.

However when listening to the radio new songs (on my station) give me a greater sense of satisfaction. So depending on the context of what I'm looking for, the effects of novelty could be reversed.
Human-recommender interaction model

Figure 1-1: Aspects of Human-Recommender Interaction. The Aspects are divided into three 'Pillars'

How people evaluate recommendations

Aspects of the recommender dialogue

**Correctness:** user judges the rec to be high quality

**Usefulness:** the rec helps the user with their task

**Transparency:** user understands why they received this rec

**Salience:** the rec stands out, generates an emotional response (pos or neg!)

**Serendipity:** the rec is unexpected, in a good way

**Spread:** the rec list represents the domain well (completeness/recall)
Figure 1: Screen shot of the experiment interface. Clicking on a movie in the list opens a pop-over with additional movie details.
What boggles my mind is that they came up with quantitative (and naïve) measurements for the attributes they used to perform their research on. There were too many assumptions made about how they should quantify attributes such as “novelty” and “understands me.”

When trying to measure human sentiment, such as in the case of a recommender system, how do we go about designing a quantifiable measure with which we can perform analysis and still maintain that the integrity of the sentiment with which we are trying to measure is intact?

More simply, how can we be sure that the method we choose to express an attribute, accurately represents that attribute?
I don’t find the results and analysis process in this paper impressive, partly because it lacks experimental design and analysis from human-centered aspects. For example, I found this paper did not understand participants well.

...active users with at least 15 ratings are invited to participate in the online survey voluntarily. There is no further description of the participants. From a statistics point of view, an analysis of users’ gender, age, location, occupation, etc. may be necessary to show that the participants are representative.
This paper used surveys to figure out user's satisfaction. My question is that how can we handle the bias of a user survey?

Since different people may have different standard to score their satisfaction, the satisfaction degree they give on the survey may not reflect their real satisfaction. It's obviously would cause bias of the survey result.
I can understand why folks would more negatively rate the recommender that suggests more unknown films. But I wonder if we would come to a different conclusion if the metric was not “perception of ratings,” but instead, “satisfaction with the actual movies recommended?” What if users were not shown a list of movies, but instead actually WATCHED the movies recommended, and then were asked to rate these movies?

… the study didn’t take into account the potential disconnect between what users want, and what they think they want. This could open up a whole new can of worms about how to design such recommender systems.
The qualitative data here seems reasonably scalable because it is ranked. How could a similar study be performed with more open-ended questions? Would a like-ness classification be appropriate?
Where can I find more examples of studies like this?
In collaborative and professional settings, I wonder how well UX and Machine Learning teams work together in regards to taking user testing results and feeding them back into the algorithmic designs? Since I haven't yet been exposed to these types of environments myself, I am very curious to about the typical struggles and best working practices.
Case study: Wikipedia reader recommendations

- Evaluating Related Article recommendations (Morgan 2016)
- Evaluating Top Articles recommendations (Morgan 2017)
Meghan Markle
American actress, model, and humanitarian

Matt Lauer
American journalist

Gertrude Jekyll
Garden designer, artist

Slobodan Praljak
Croatian politician and soldier

Bitcoin
Digital cash system and associated currency unit
If users knew more about how algorithms are selecting recommendations, instead of just seeing the recommendations themselves, would they be more likely to discount recommendations due to novelty?

Novelty was shown by the paper to be a negative influence on user opinions of movie recommendation lists. It seems counterproductive to reject recommendations because of this characteristic.
Fundamental question: Why should I, as a user, trust that this algorithm understands who I am, what I like, and what I’m doing?
<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>File size</th>
</tr>
</thead>
<tbody>
<tr>
<td>survey</td>
<td>Oct 12, 2017</td>
<td>–</td>
</tr>
<tr>
<td>Personas / Scenarios</td>
<td>Jul 19, 2017</td>
<td>–</td>
</tr>
<tr>
<td>interview notes</td>
<td>Aug 31, 2017</td>
<td>–</td>
</tr>
<tr>
<td>Consent forms</td>
<td>Jul 18, 2017</td>
<td>–</td>
</tr>
<tr>
<td>Notes for themes</td>
<td>Oct 17, 2017</td>
<td>–</td>
</tr>
</tbody>
</table>
Why don't you want to see this?

- It's not relevant to me
- I've already seen it
- It makes me feel uncomfortable
- It's spam

Why was this item recommended to me?
Recommendations are articles and videos that we think you’ll be interested in, sourced from the millions of items that are being saved to Pocket every day. The more you save and interact with Pocket, the more personalized your Recommendations will be.

Recommendations also come from the people you follow on Pocket. When someone you follow recommends something, it’ll appear in your feed alongside your personalized recommendations from Pocket.”
The categories in this section help advertisers reach people who are most likely to be interested in their products, services, and causes. We've added you to these categories based on information you've provided on Facebook and other activity.

<table>
<thead>
<tr>
<th>About you</th>
<th>Your categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away from family</td>
<td>Away from hometown</td>
</tr>
<tr>
<td>Birthday in November</td>
<td>US politics (very liberal)</td>
</tr>
<tr>
<td>Facebook access (mobile): smartphones and tablets</td>
<td>Frequent Travelers</td>
</tr>
<tr>
<td>African American (US)</td>
<td>Facebook access (mobile): smartphones</td>
</tr>
<tr>
<td>Facebook access (OS): Mac OS X</td>
<td>Gmail users</td>
</tr>
<tr>
<td>Facebook access (network type): WiFi</td>
<td>Facebook access (mobile): all mobile devices</td>
</tr>
</tbody>
</table>
Case study: Citation recommendations

Recommending additional articles to cite in a research paper, based on the articles that are already cited.

“Bayes and PLSI perform well as recommenders in offline simulation experiments... Users, however, were not satisfied with these recommendation lists. These results suggest that the research community’s dependence on offline experiments have created a disconnect between algorithms that score well on accuracy metrics and algorithms that users will find useful.”

“In previous work, we argued that showing one good recommendation in a list of five was enough to satisfy users. It is not that simple: showing one horrible recommendation in five is enough for users to lose confidence in the recommender.

**We call this the Don’t Look Stupid principle:** only show recommendation lists to users when you have some confidence in their usefulness.”

Comparing music recommendations between Echo Nest (Spotify), Google Instant Mix, and iTunes Genius by “WTF score”

“Evaluating playlists is hard. However, there is something that we can do that is fairly easy to give us an idea of how well a playlisting engine works compared to others.

I call it the WTF test. It is really quite simple. You generate a playlist, and just count the number of head-scratchers in the list. If you look at a song in a playlist and say to yourself ‘How the heck did this song get in this playlist’ you bump the counter for the playlist. The higher the WTF count the worse the playlist.”

https://musicmachinery.com/2011/05/14/how-good-is-googles-instant-mix/
<table>
<thead>
<tr>
<th>Track Name</th>
<th>Artist</th>
<th>Album</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BABY I'M A WANT YOU</td>
<td>Aaron Neville</td>
<td>Tell It Like It Is (Aim)</td>
<td>88%</td>
</tr>
<tr>
<td>Tubular Bells I (Full Album)</td>
<td>Mike Oldfield</td>
<td></td>
<td>87%</td>
</tr>
<tr>
<td>Boo Hoo</td>
<td>Bert Kaempfert</td>
<td></td>
<td>86%</td>
</tr>
<tr>
<td>Der Gnom</td>
<td>Mussorgsky, Modest Peter</td>
<td></td>
<td>86%</td>
</tr>
<tr>
<td>A Lovely Way To Spend An........</td>
<td>Engelbert Humperdink</td>
<td>Red Sails in the Sunset</td>
<td>85%</td>
</tr>
</tbody>
</table>
Similar Artists

Paul McCartney
John Lennon
George Harrison
Paul McCartney & Wings
Ringo Starr
Wings
In order to make recommendations that people will actually want to use, **inspire trust**.

Your users should...

- feel the recommendations meet their current needs (audience, purpose, context)
- feel like they understand how the recommendation was made (interpretability)
- not have a ‘WTF’ moment (don’t look stupid)
- not feel like the recommendation is invasive or embarrassing (don’t be creepy)
- feel like they have control over their experience (ask, don’t tell)
Questions to ask yourself

1. How do you know that you’re making human-centered recommendations?
2. How does the presentation of your recommendations affect user trust?
Break (15 minutes)
Human-centered data visualization
Visualizing complex information

- https://fivethirtyeight.com/features/science-isnt-broken/#part1
Visualizing complex information

The Tone Analyzer Service analyzes text at the document level and the sentence level. Use the document level analysis to get a sense of the overall tone of the document, and use the sentence level analysis to identify specific areas of your content where tones are the strongest.

**Document-level**

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Feeling</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>Unlikely</td>
<td>0.91</td>
</tr>
<tr>
<td>Disgust</td>
<td>Likely</td>
<td>0.36</td>
</tr>
<tr>
<td>Fear</td>
<td>Likely</td>
<td>0.23</td>
</tr>
<tr>
<td>Joy</td>
<td>Likely</td>
<td>0.02</td>
</tr>
<tr>
<td>Sadness</td>
<td>Likely</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The Emotion graph represents the likelihood that an emotion tone is present in the text. Learn more.

**Language Style**

<table>
<thead>
<tr>
<th>Style</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical</td>
<td>0.93</td>
</tr>
<tr>
<td>Confident</td>
<td>0.00</td>
</tr>
<tr>
<td>Tentative</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The Language Style and Social Tendencies graphs represent the amount of a language or social tone present in the text. Learn more.

**Social Tendencies**

<table>
<thead>
<tr>
<th>Tendency</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>0.06</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.23</td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.78</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.93</td>
</tr>
<tr>
<td>Emotional Range</td>
<td>0.87</td>
</tr>
</tbody>
</table>

**Sentence-level**

Identify sentences with stronger tones in context or sorted by score. Highlighted sentences indicate the likelihood of a tone present. If more than one tone is present, the stronger one is shown. Click on a sentence to see a breakdown of all tones.
Homework
Homework due next week

Readings (read both, reflect on one)


Final project presentations!

- **Format**: Google slides or PDF, shared with Jonathan and Oliver, linked from Canvas
- **Due before class** (submissions after 5pm marked down)

**Reminder**: Please fill out course evaluation survey! [https://uw.iasystem.org/survey/181661](https://uw.iasystem.org/survey/181661)

See: [https://wiki.communitydata.cc/HCDS_(Fall_2017)#Week_10_November_30](https://wiki.communitydata.cc/HCDS_(Fall_2017)#Week_10_November_30)
Thank you, and an apology
No in-class activity this week!

YOU GET 2 POINTS AND YOU GET 2 POINTS

EVERYONE GETS 2 POINTS
Questions?
Overview of the day

- Announcements
- Final project presentation details
- Final project report details
- Extra credit assignment details
- A4: Crowdwork Ethnography reflection
- Dinner break
- Week 9 reading reflections
- Lecture: Design + HCDS (+ break?)
- Homework review