Human Centered Data Science

DATA 512 — Jonathan T. Morgan & Oliver Keyes

Course overview | Week 1 | September 28, 2017

Introductions

Overview of the day

- Syllabus review
- Pre-course survey results
- *Dinner break (15 min)*
- What do we mean by 'data science'?
- What do we mean by 'human centered'?
- *Coffee break (15 min)*
- How does HCD relate to DS?
- For week 2

Syllabus review

https://wiki.communitydata.cc/HCDS_(Fall_2017)

Class policies

https://wiki.communitydata.cc/HCDS_(Fall_2017)#Policies

Assignments

https://wiki.communitydata.cc/HCDS_(Fall_2017)/Assignments
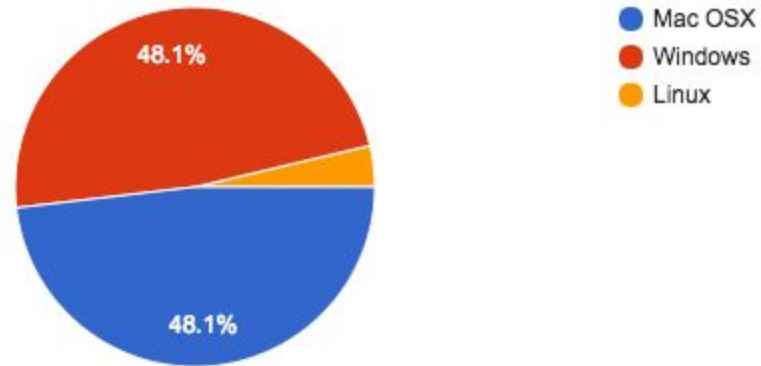
Communication

Questions?

Pre-course survey

What OS do you use?

What operating system is installed on the laptop you plan to use for this course?
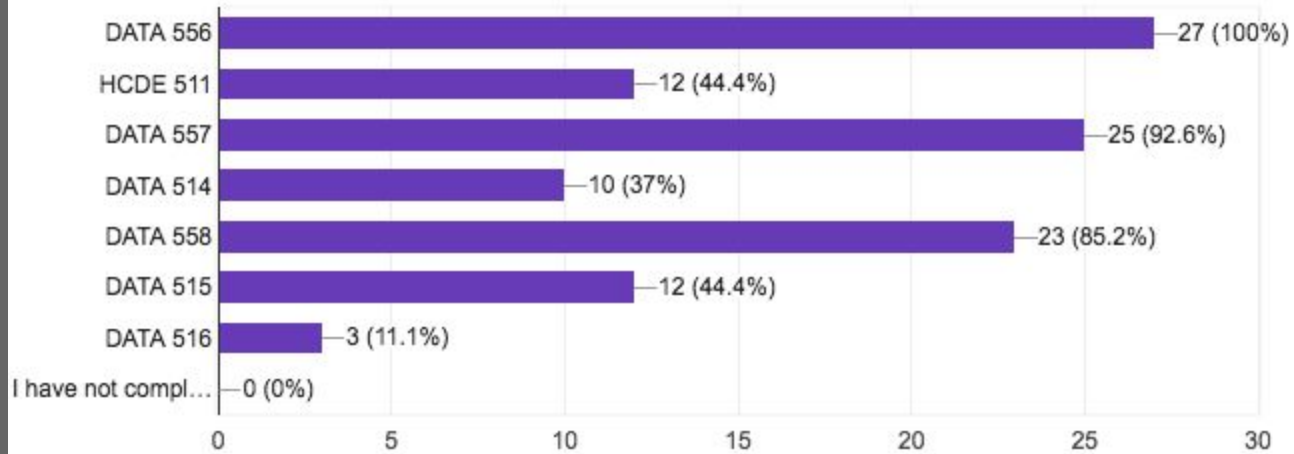
27 responses

- Mac OSX
- Windows
- Linux

48.1%

48.1%

What courses have you taken?

Which UW Data Science Masters program courses (if any) have you taken already?
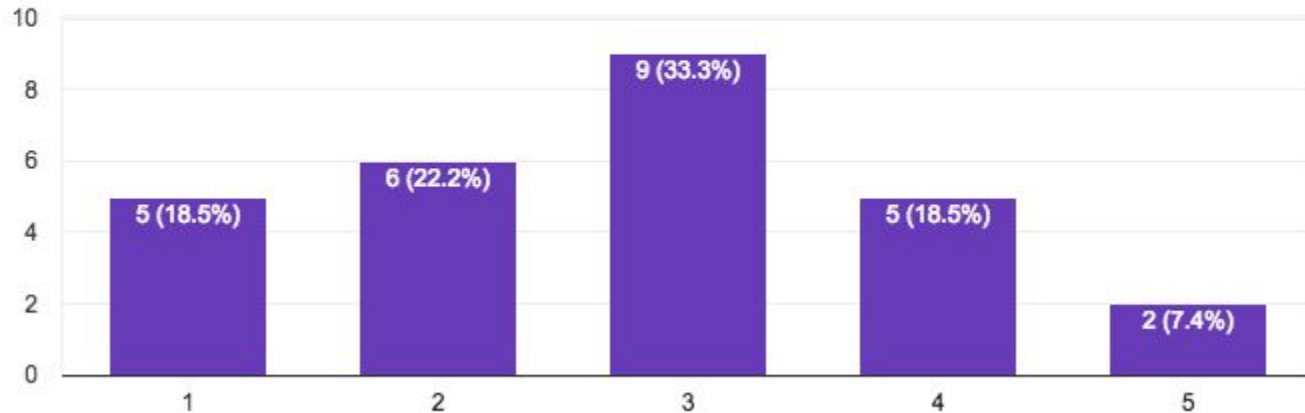
27 responses

How much programming experience do you have?
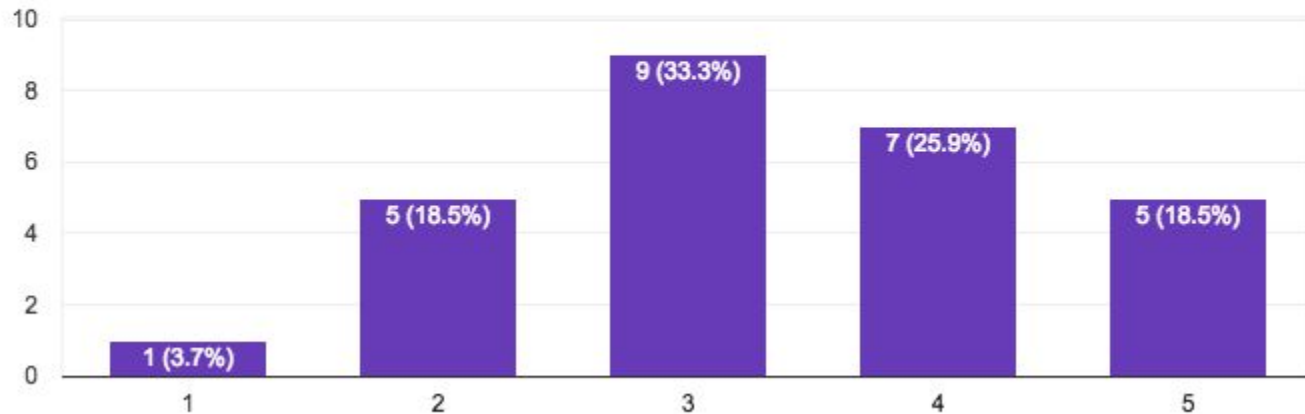
How much quant. research experience do you have?

How much QUANTITATIVE research experience do you have?

27 responses

What kind of quant research?

- My previous employment was for a retirement record keeping company in which identifying anomalies in data was the most common quantitative analysis I performed.
- Created detailed outlier rejection analysis of wind tunnel acoustical data
- visualization with Tableau and Power BI, build ML model for regression, classification and clustering
- data analysis with large amounts of data to answer questions about how Windows is used in the wild
- I have created end to end data pipeline to analyze Windows 10 customer feedback text.
- I do a lot of Product usage analysis
- I have built a supervised machine learning model to predict a fungal blueberry disease
- I have done lots with excel, I have used R and Python, mainly as a MSDS student.
- I've built models to detect fraud
- I have developed and contributed to several packages in R for air quality and satellite data analysis
- Primarily what we have covered thus far in the program.

Have you ever used Jupyter notebooks for research?
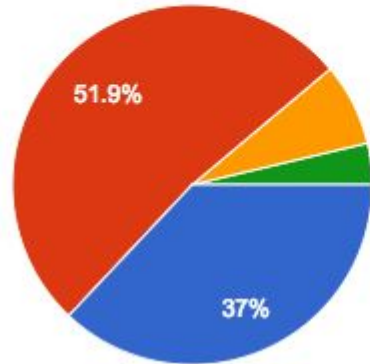
How much qual. research experience do you have?

How much QUALITATIVE research experience do you have?

27 responses
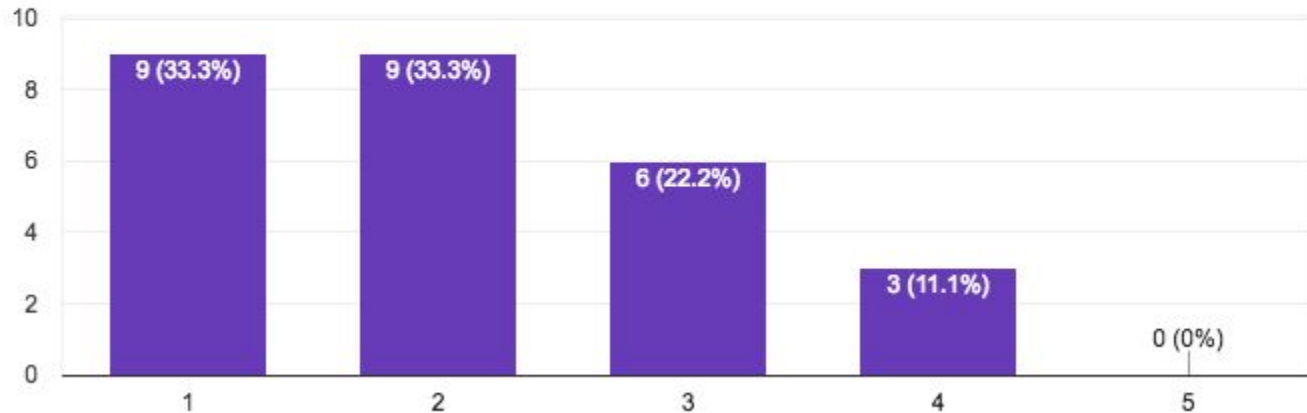
What kind of qual research?

- I have taken courses in survey design
- I took part of usability testing a few times.
- Some experience with product "design" as a Scrum product owner.
- I am the Data Privacy champion for my group.
- I have taken the HCDE course on user centered design
- As a marketing professional I was a consumer of qualitative research, but did not conduct the research.
- Some professional experience in usability
- I took a data visualization / user-centered design course (HCDE 511)
- surveys and progress tracking, informal feedback on visualizations

What kind of data are you most interested in working with?

- Biomedical, finance
- Epidemiological / health data
- Sport data, esp. soccer because I like sports.
- Social media data Government data
- video game data
- industrial predictive monitoring/maintenance
- business data, social media data
- Financial data(stocks, house prices), Networking(latency, packet transmission)
- I would like to get more experience with a broader selection of data. Often, it seemed like our previous data sets were either too small to do in-depth analysis or too unwieldy or fragmented so we spent most of the time trying to clean or standardize the data.
- Data about people's music listening or video game playing habits
- social media and biomedical data
- Text data - challenging to analyze ; Education data - K-12 education for everyone

What do you hope to get from DATA 512?

- Usability testing, interaction design
- Understand the human side of data science as opposed to the quantitative side of data science.
- How data decisions can have human effects
- Being able to communicate and convince the effectiveness of statistical and machine learning models to software engineers would be the biggest takeaway I desire from this course.
- I would love to have a better grasp around the current legal landscape of data collection as it seems more and more that the data companies collect about their users is brought up in the news.
- I am intrigued by algorithmic bias and look forward to learning how it pops up and how to avoid it.
- I would like to better understand the how the interpretation of data might be perceived by individuals or groups. For example, how might data indicating the presence and scope global climate change be interpreted by individuals who may have a political agenda?
- Honestly not sure!

Anything else we should know?

"I believe in ghosts"

What do we mean by *data science*?

Group activity

20 minutes, 4-5 people per group

Sketching the boundaries of data science

- Divide your paper in half: "data science" and "not data science"
- Discuss the question *"What do we mean by data science?"*
- Consider key terms, concepts, activities, examples, data, domains etc…
  - that you associate with data science.
  - that you DO NOT associate with data science.
- Add these to sticky notes, place them on either side of the paper
- Be prepared to discuss your decisions

This activity is not graded. There is no right or wrong answer.

What does 'human centered' mean?

Group activity

20 minutes, 4-5 people per group

Sketching the boundaries of human centered design

- Discuss the question *"What do we mean by human centered-ness?"*
- Come up with AT LEAST three bullet points about what 'human centered' means
- Be prepared to discuss your decisions

This activity is not graded. There is no right or wrong answer.

Lecture: a short history of human-centered design

User-centered design process

Source: https://www.usability.gov/what-and-why/user-centered-design.html

User-centered design principles

- **Audience:** Who is the product for?
- **Purpose:** How will they use the product, what problem will it solve?
- **Context:** What context will they be using the product in, and how could that context impact their experience?

"Human-Centered Design (HCD) is **a process and a set of techniques** used to create new solutions for the world. Solutions include products, services, environments, organizations, and modes of interaction. The reason this process is called "human-centered" is because **it starts with the people we are designing for.**

The HCD process begins by examining the needs, dreams, and behaviors of the people we want to affect with our solutions. We seek to listen to and understand what they want. We view the world through this lens **throughout the design process.**"

ISO 9241-210: HCD for interactive systems

- The design is based upon an explicit understanding of users, tasks and environments.
- Users are involved throughout design and development.
- The design is driven and refined by user-centred evaluation.
- The process is iterative.
- The design addresses the whole user experience.
- The design team includes multidisciplinary skills and perspectives.

Human Centered Systems in the Perspective of Organizational and Social Informatics

Rob Kling & Leigh Star, 1997

Based in an analysis of human tasks

There must be analysis which encompasses the complexity of social organization and the technical state of the art. The analysis cannot be based upon a vague idea of what a generic individual would like, in a stereotypic situation that effectively ignores the varieties of concrete social locations… a human centered analysis must take account of varied social units that structure work and information.

Built to take account of human skills

There are important architectural relationships, such as the question of whether the basic architecture of the system reflect a realistic relationship between people and machines. As with the architecture of buildings, the architecture of machines embody questions of livability, usability and sustainability.

Designed to address human needs

The question of whose purposes are served in the development of a system would be an explicit part of design, evaluation and use. Thus the question of whose ideas get put into the design process is an important one for human centered systems. As well, the question of whose problems are being solved is important.

Monitored for performance in terms of human benefits

Human-centered is not a "one-off" or timeless attribute of a system at a given point in time. Rather, it is a process, one which would take into account how criteria of evaluation are generated and applied, and for whose benefit. It would include the participation of stakeholder groups -- such as involving patient groups in the development of specialist medical technologies, or teachers in the development of instructional technology.

Implications of Kling & Star's HCD

- **Audience:** Who is affected by the system?
- **Purpose:** Whose purposes are served? Whose purposes are not?
- **Context:** How will the system impact people's lived experience? How sustainable and adaptable is the system? What unintended consequences might result from the design and deployment of the system?

Implications of Kling & Star's HCD

- **Ethics**
- **Values**
- **Openness**
- **Communication**
- **Consent**
- **Consequences**
- **Participation**
- **Sustainability**
- **Social impact**
- **Systems thinking**
- **Multidisciplinarity**

Implications of Kling & Star's HCD

Human Centered Design & Engineering (HCDE) faculty and students are advancing the research and design of technologies by using innovative techniques to study human activity and develop meaningful information and sociotechnical systems.

HCDE is designing the future by:
- Considering the role of technology in human activity.
- Prioritizing the needs, desires, and behaviors of people and communities who interact through sociotechnical systems.
- Addressing the specifics of design by working with interdisciplinary communities of researchers to build the technologies of tomorrow.

How does human centered design relate to data science?

Some topics we'll cover

- **Legal and ethical considerations in data collection**
- **Data provenance, preparation, and reproducibility**
- **DS study design best practices**
- **Machine learning:** ethics, best practices, societal impacts
- **Mixed-methods research:** combining quant and qual in DS
- **Human computation:** ethical and effective crowdwork
- **User experience and big data:** UCD & UX for data-driven systems
- **HCDS in the wild:** critical evaluation of data science for social good
- **Scientific communication:** effective and ethical translation for diverse audiences

Group activity

30 minutes, 4-5 people per group

Analyzing HCDS scenarios

- Read through the scenario with your group
- Flesh out the scenario with additional details of your choice (document these)
- Describe at least 3 human-centered design considerations in your scenario
- Follow submission instructions on Canvas (discussion "Week 1 in-class activity: Real-life Human Centered Data Science scenarios")
- Be prepared to discuss your choices

This activity is graded. There are no right or wrong answers; what matters is that you show that you're thinking.

See: https://wiki.communitydata.cc/HCDS_(Fall_2017)/Assignments#Weekly_in-class_activities

For Week 2

Recording lectures

Homework

https://wiki.communitydata.cc/HCDS_(Fall_2017)/Assignments#
Weekly_reading_reflections

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. **Out with every theory of human behavior, from linguistics to sociology.** Forget taxonomy, ontology, and psychology. **Who knows why people do what they do?** The point is they do it, and we can track and measure it with unprecedented fidelity. **With enough data, the numbers speak for themselves.**

Chris Anderson - WIRED

Questions?