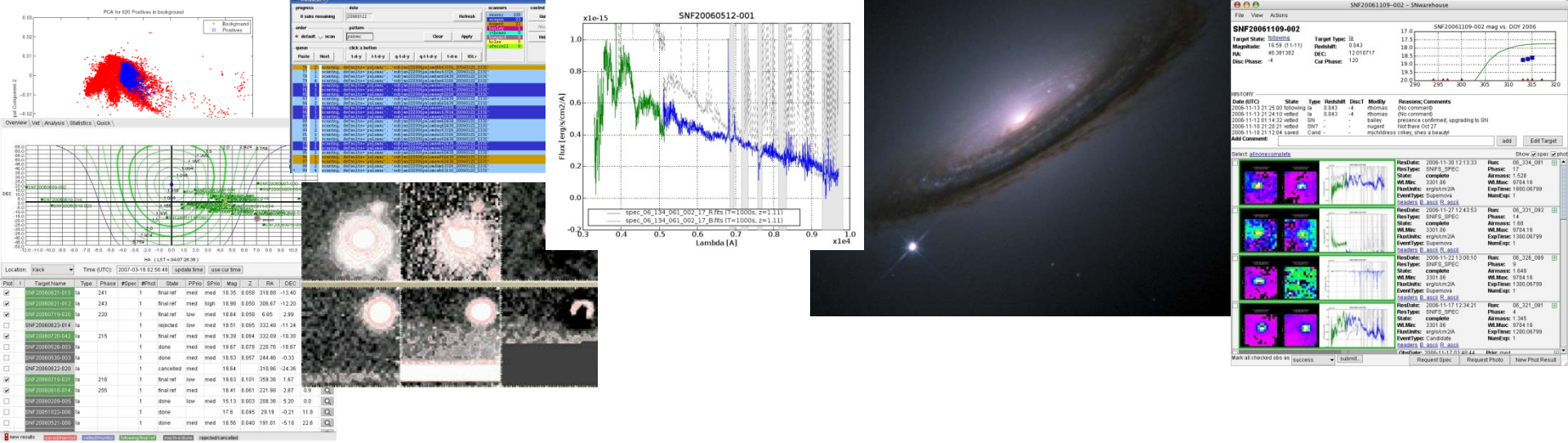


Human-Centered Data Science

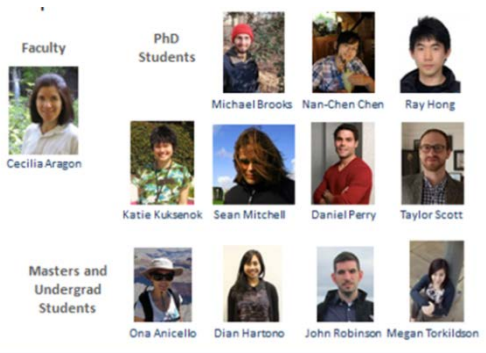
Cecilia R. Aragon

Dept. of Human Centered Design & Engineering

eScience Institute

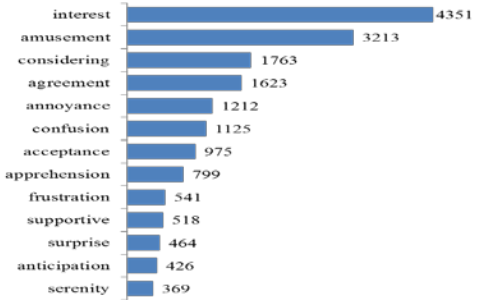
University of Washington

aragon@uw.edu



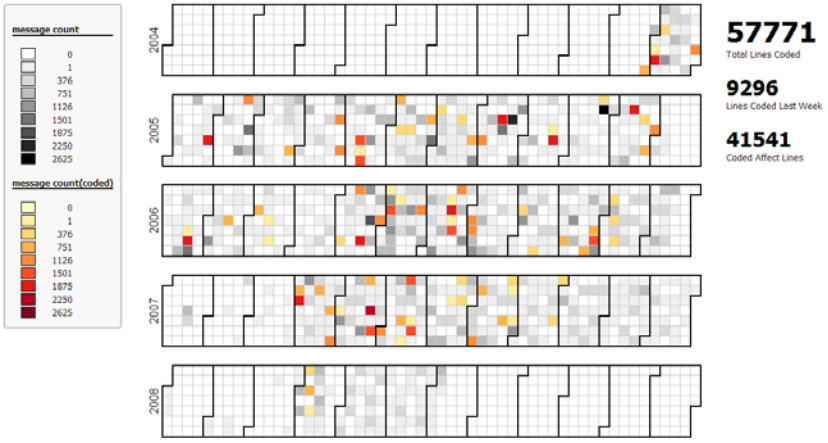
Research with Social Media Text Data

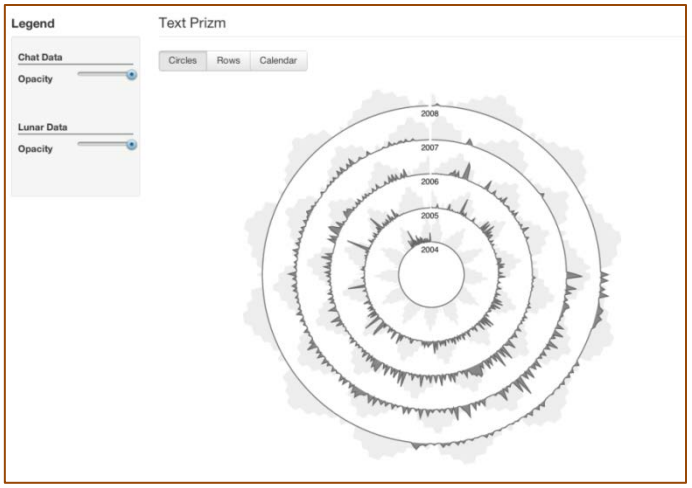
- Ranges along quantitative – qualitative spectrum
- Quantitative
  - Good: Large quantities of data, efficient
  - Bad: Relatively shallow, superficial
- Qualitative
  - Good: deep, explanatory conclusions
  - Bad: Small, focused samples, inefficient
- How to get best of both worlds?
  - Visual analytics, combining machine learning with visualization and qualitative research



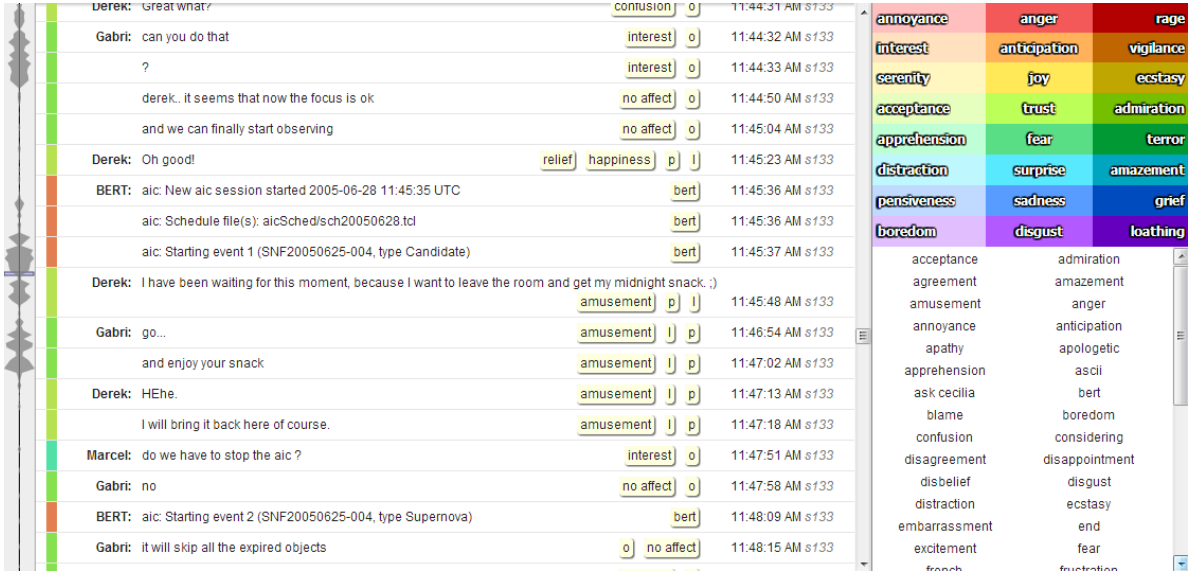
6:07:57	Ray	cool, it worked	*amusement, relief*
6:08:04	Matt	woot	*excitement, joy*
6:08:07	Ray	awesome, I don't think he needs that long of a sleep after turning it off	*acceptance, no affect*
6:08:47		We enhanced eready to detect the sticking	*no affect*
6:08:58	Matt	good job	*supportive, acceptance*
6:09:21		seems it did well there	*happiness, no affect*
6:09:26	Ray	yeah, pretty cool huh?	*interest, agreement, happiness*
6:09:43	Matt	helps keep me from having to stopaic and restart	*no affect*
6:09:55	Ray	indeed, that was the point	*agreement*

"Statistical Affect Detection in Collaborative Chat," Michael Brooks, Katie Kuksenok, Megan Torkildson, Daniel Perry, John Robinson, Paul Harris, Ona Anicello, Taylor Scott, Ariana Zukowski, Cecilia Aragon. *CSCW '13* (2013)





Integrate machine learning and interactive visualization into a qualitative research workflow



- Conclusions based on large data sets
- Maintain context, nuance, depth
- Select tools based on transparency and understandability, not just efficiency

"Statistical Affect Detection in Collaborative Chat," Michael Brooks, Katie Kuksenok, Megan Torkildson, Daniel Perry, John Robinson, Paul Harris, Ona Anicello, Taylor Scott, Ariana Zukowski, Cecilia Aragon. *CSCW '13* (2013)

Human issues in data science

- Ethics: Lilly Irani and Six Silberman, UCSD, Turkopticon and crowdsourcing

  Irani, Lilly C., and M. Silberman. Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. *Proc.of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2013, 611–620.

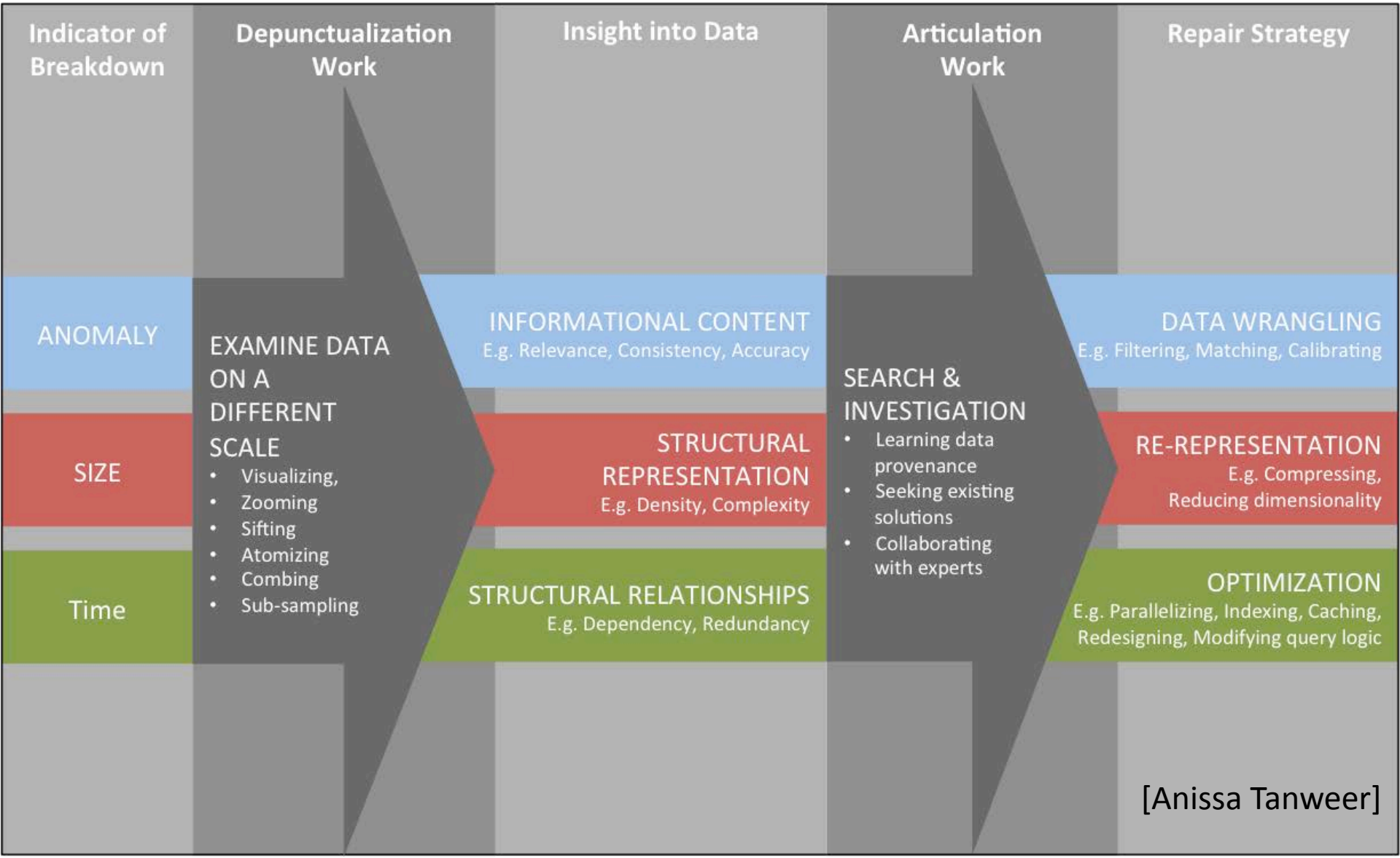
- Design: What happens when you crowd-source design? Daniela Rosner, UW (systems that guarantee banal results?)

  Rosner, Daniela, and Jonathan Bean. "Big data, diminished design." *interactions* 21, no. 3. 2015.

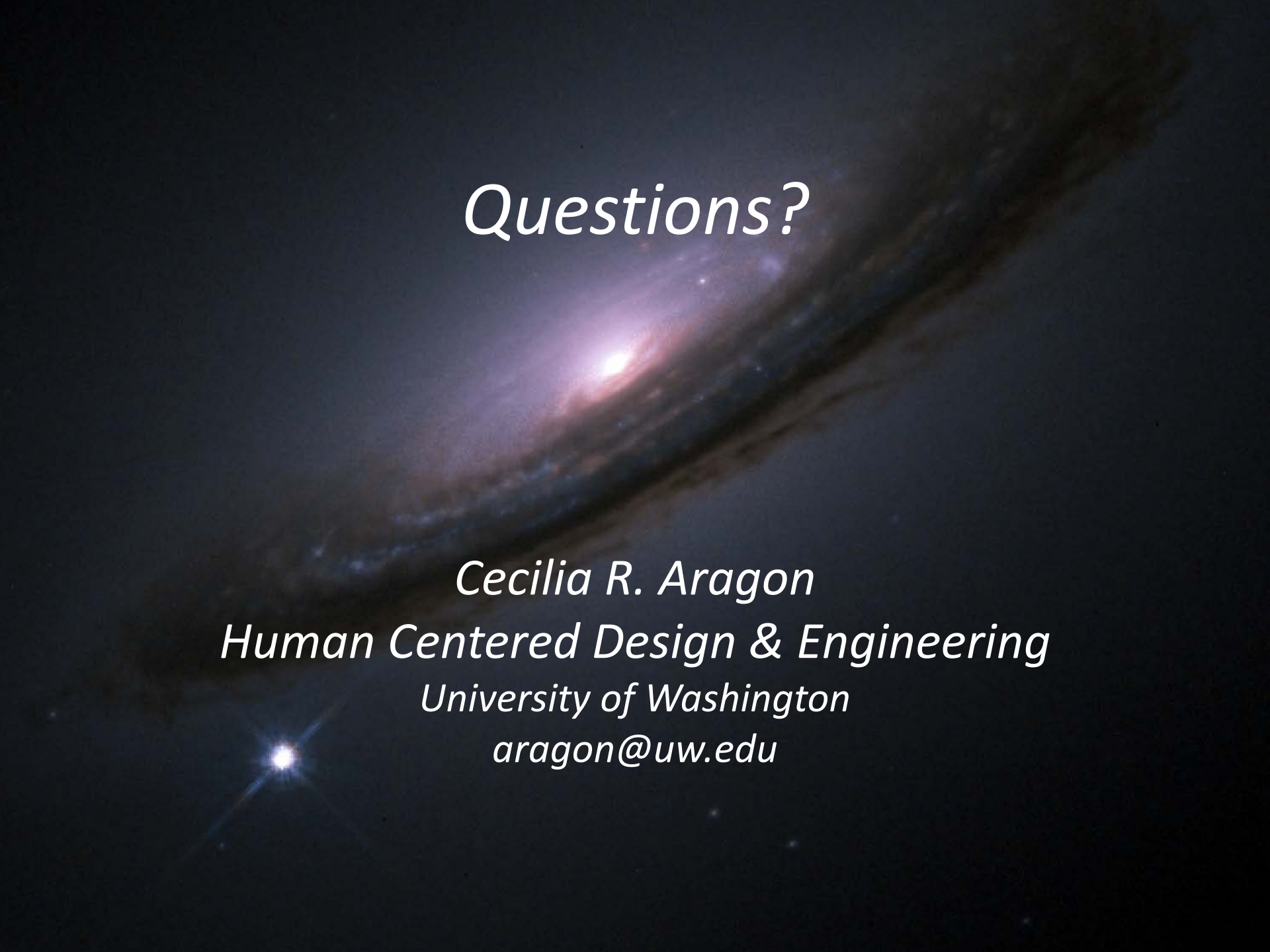
- Data expectations: What does data really mean across different fields? Gina Neff, B. Fiore-Gartland

  Brittany Fiore-Gartland and Gina Neff, UW Dept of Communication, “Communication, Mediation, and the Expectations of Data: Data Valences across Health and Wellness Communities.” 2014.

Typical Breakdown-Repair Process in a Data Science Collaboration



[Anissa Tanweer]


Questions?
Cecilia R. Aragon
Human Centered Design & Engineering
University of Washington
aragon@uw.edu