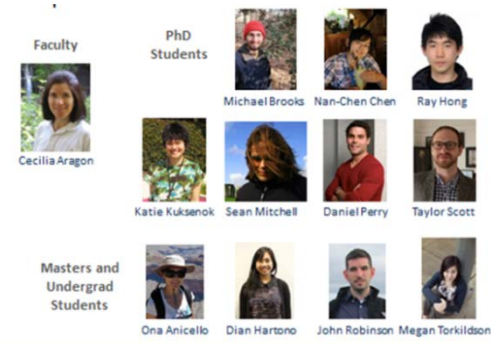


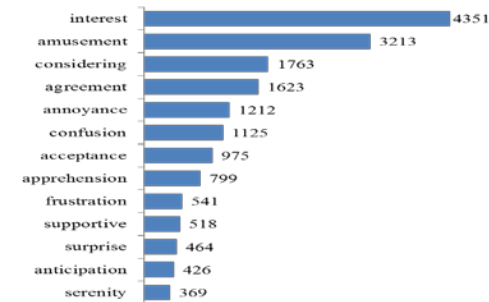
Human-Centered Data Science

Cecilia R. Aragon
Dept. of Human Centered Design & Engineering
eScience Institute
University of Washington
aragon@uw.edu



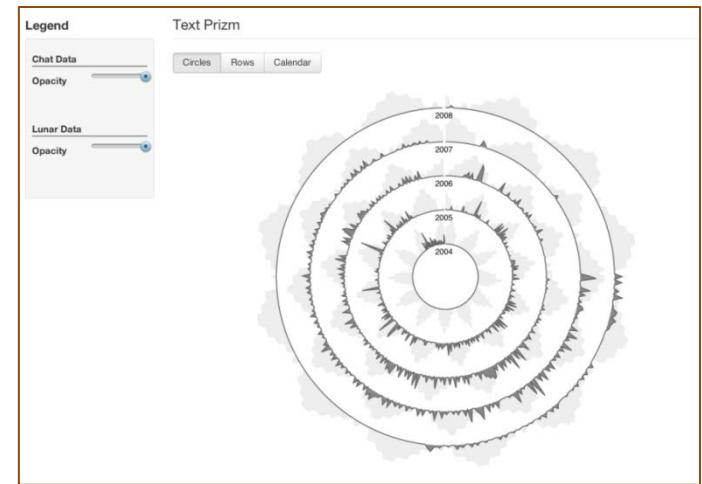
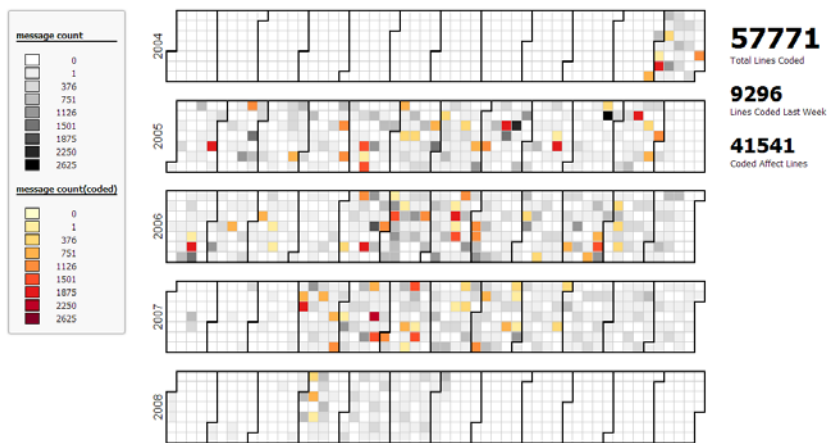
Research with Social Media Text Data

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

["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

Derek:	Great what?	confusion	o	11:44:31 AM s133	
Gabri:	can you do that	interest	o	11:44:32 AM s133	
?		interest	o	11:44:33 AM s133	
derek..	it seems that now the focus is ok	no affect	o	11:44:50 AM s133	
and we can finally start observing		no affect	o	11:45:04 AM s133	
Derek:	Oh good!	relief	happiness	p j	11:45:23 AM s133
BERT:	aic: New aic session started 2005-06-28 11:45:35 UTC	bert		11:45:36 AM s133	
aic:	Schedule file(s): aicSched/sch20050628.tcl	bert		11:45:36 AM s133	
aic:	Starting event 1 (SNF20050625-004, type Candidate)	bert		11:45:37 AM s133	
Derek:	I have been waiting for this moment, because I want to leave the room and get my midnight snack. :)	amusement	p j	11:45:48 AM s133	
Gabri:	go...	amusement	j p	11:46:54 AM s133	
and enjoy your snack		amusement	j p	11:47:02 AM s133	
Derek:	HEhe.	amusement	j p	11:47:13 AM s133	
I will bring it back here of course.		amusement	j p	11:47:18 AM s133	
Marcel:	do we have to stop the aic ?	interest	o	11:47:51 AM s133	
Gabri:	no	no affect	o	11:47:58 AM s133	
BERT:	aic: Starting event 2 (SNF20050625-004, type Supernova)	bert		11:48:09 AM s133	
Gabri:	it will skip all the expired objects	o	no affect	11:48:15 AM s133	

annoyance	anger	rage
interest	anticipation	vigilance
serenity	joy	ecstasy
acceptance	trust	admiration
apprehension	fear	terror
distraction	surprise	amazement
pensiveness	sadness	grief
boredom	disgust	loathing
acceptance	admiration	
agreement	amazement	
amusement	anger	
annoyance	anticipation	
apathy	apologetic	
apprehension	ascii	
ask cecilia	bert	
blame	boredom	
confusion	considering	
disagreement	disappointment	
disbelief	disgust	
distraction	ecstasy	
embarrassment	end	
excitement	fear	
frank	frustration	

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

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.

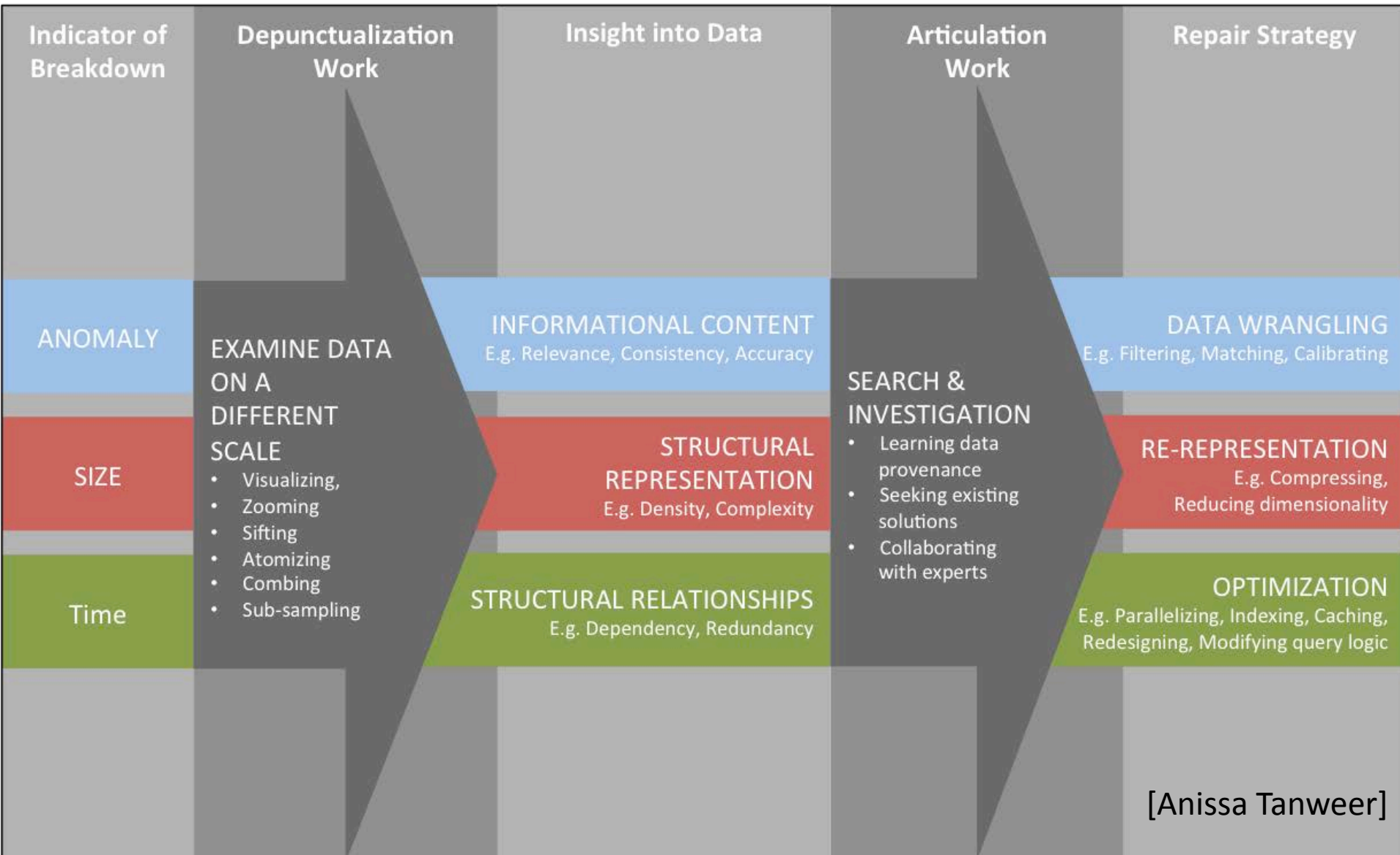
- **Design:** What happens when you crowd-source design? Daniela Rosner, UW (systems that guarantee banal results?)

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





Questions?

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