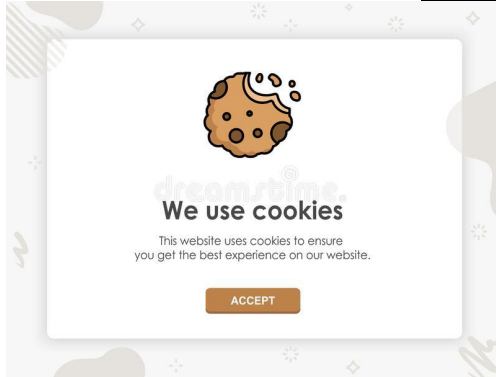


Issues With Archiving and Interpreting Data

By: AJ Herget & Tina Liu

Why should we care about data?

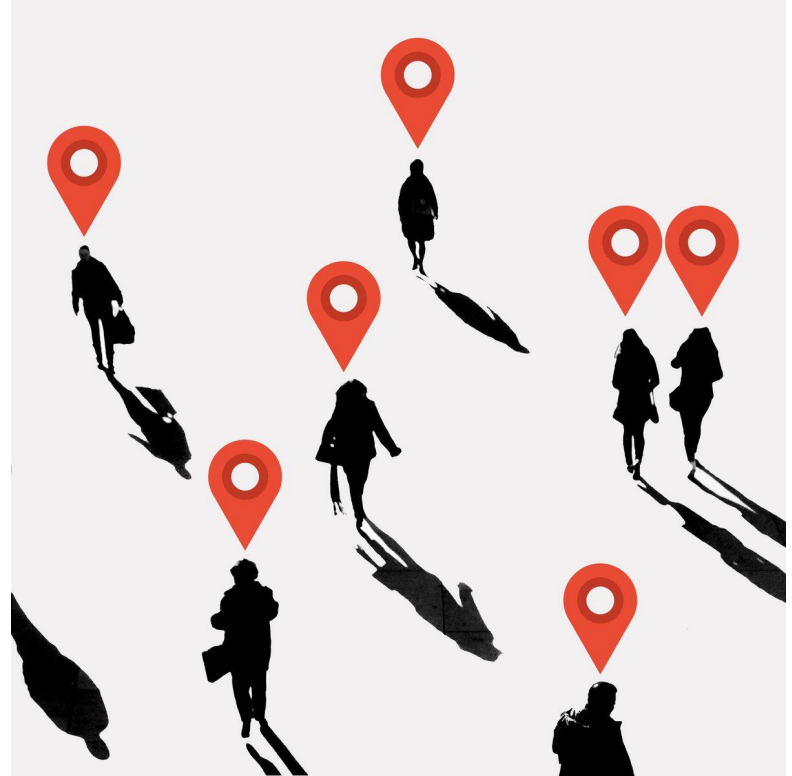
facebook



Online Advertising

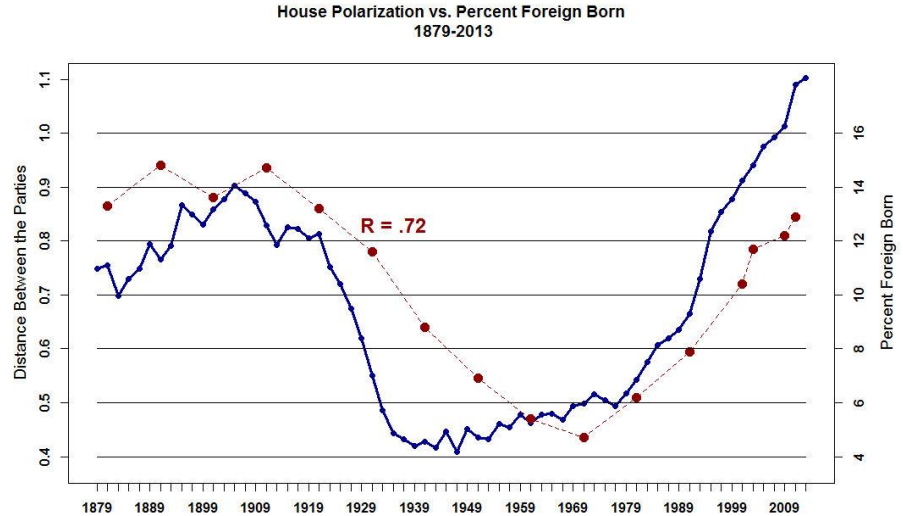
“Critical Questions For Big Data” – Boyd

- What is big data?
- Is data truly objective?
- Does more data = better?
- Context matters
- Is certain data use ethical?
- The digital divide



“We Are All Social Scientists Now” – Grimmer

- How do we make use of this data?
- Need to collaborate with social scientists for their expertise in causal inference



“Data Cleaning” - Chu

	col1	col2	col3	col4	col5
0	2	5.0	3.0	6	NaN
1	9	NaN	9.0	0	7.0
2	19	17.0	NaN	9	NaN

mean()

	col1	col2	col3	col4	col5
0	2.0	5.0	3.0	6.0	7.0
1	9.0	11.0	9.0	0.0	7.0
2	19.0	17.0	6.0	9.0	7.0

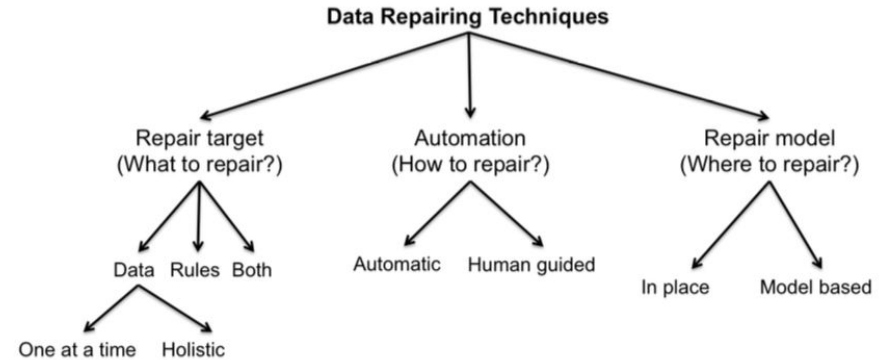
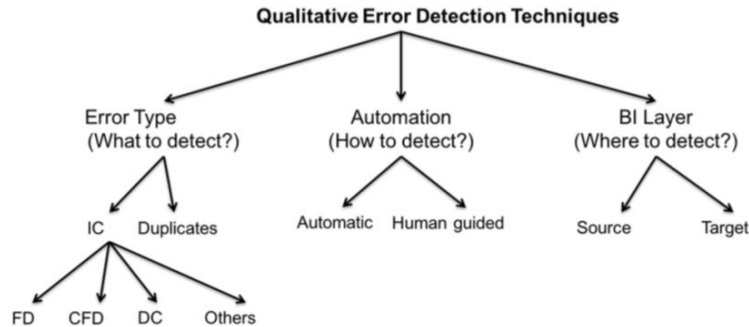


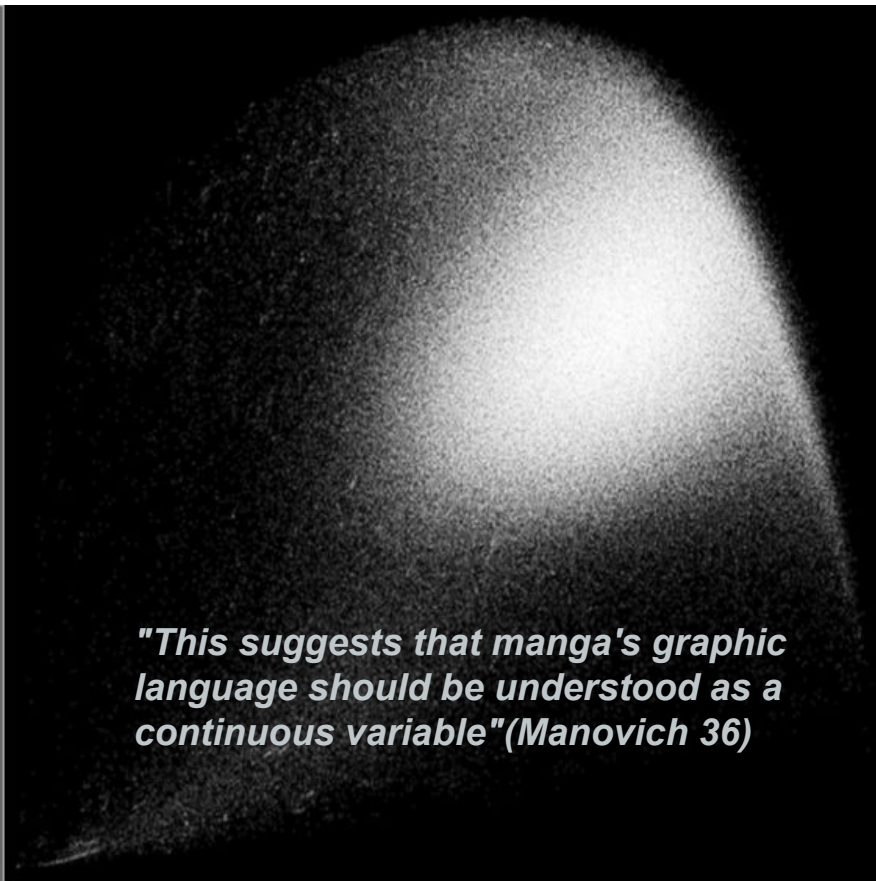
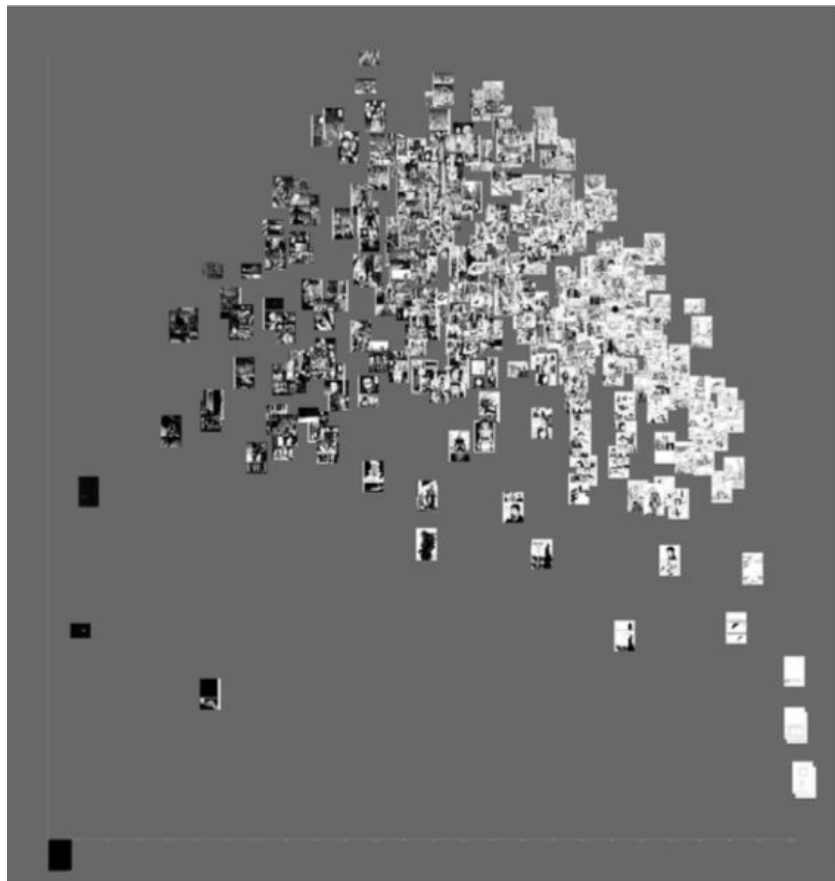
Figure 1: Classification of qualitative error detection techniques.

Figure 2: Classification of data repairing techniques.

"How to Compare One Million Images"

/ Dr. Lev Manovich





"This suggests that manga's graphic language should be understood as a continuous variable"(Manovich 36)

"Archive Fever" / Jacques Derrida

- “The citizens who ... held and signified political power were considered to possess the right to make or represent the law. On account of their publicly recognized authority, it is at their home, in that place which is their house ... that official documents are filed ... It is thus, in this domiciliation, in this house arrest, that archives take place.”
- An archive = a privileged space
- new media technologies structure our capacity for thought and communication

"Defining Data for Humanists" / Trevor Owens

- What is Data *To a Humanist*?
 - Data as Constructed Artifacts
 - Data as Interpretable Texts
 - Data as Processable Information

Conclusions:

- Possibility of overcoming the preexisting bias
 - Manovich; Derrida;
- Make use of big data
 - Grimmer; Owens

“You can measure anything in textual form - with a measurement error. The question is whether that error matters” - Nan Z. Da

Does the error matter?

- Yes: Chu; Boyd
- We agree

Work Cited

- Boyd, Danah, and Kate Crawford. "CRITICAL QUESTIONS FOR BIG DATA." *Information, Communication & Society*, vol. 15, no. 5, 2012, pp. 662–79. *Crossref*, doi:10.1080/1369118x.2012.678878.
- Chu, Xu, et. al. "Data Cleaning: Overview and Emerging Challenges." *Data Cleaning | Proceedings of the 2016 International Conference on Management of Data*, 1 June 2016, dl.acm.org/doi/10.1145/2882903.2912574.
- Da, Nan Z. "The Digital Humanities Debacle." *The Chronicle of Higher Education*, The Chronicle of Higher Education, 23 July 2020, www.chronicle.com/article/the-digital-humanities-debacle/.
- Derrida, Jacques and Eric Prenowitz. "Archive Fever: A Freudian Impression." *Diacritics*, Vol. 25, No. 2, Summer 1995, pp. 9-63.
- Grimmer, Justin. "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together." *Cambridge Core*, Cambridge University Press, 31, Dec. 2014.
- Manovich, Lev. "How to Compare One Million Images." (2012).
- Owen, Trevor. "Defining Data for Humanists: Text, Artifact, Information or Evidence?" *Journal of Digital Humanities*, Vol. 1, No. 1 Winter 2011.