

Life After Linkage

Inference with Many Messy Datasets

Andee Kaplan

Duke University
andrea.kaplan@duke.edu

October 10, 2018
<http://bit.ly/kaplan-ncsu-2018>

Outline

- 1 Record linkage and the downstream task
What is record linkage, and how is it useful?
- 2 Linear regression/binary classification
How to represent a linked dataset via prototyping?
- 3 Population size estimation
How do we count the dead in a civil conflict? When do these estimates fail to capture reality?
- 4 Other work
What else have I been up to?

Record Linkage and the downstream task

Linear regression/binary classification

Population size estimation

Other work

Thank you

- Slides – <http://bit.ly/kaplan-ncsu-2018>
- Contact
 - ▶ Email – andrea.kaplan@duke.edu
 - ▶ Web – <http://andeekaplan.com>
 - ▶ Twitter – <http://twitter.com/andeekaplan>
 - ▶ GitHub – <http://github.com/andeek>

Appendix

Appendix Slide

References I