



Big Data for Cities

Week 6

Curt Savoie
Connor McKay



Agenda

- Recap on last week
- Analysis of multiple data sets
- Let's talk midterm!
- More R Demo
- Q&A



Recap

- Readings / Questions on Comms/Vis?
- Is R getting a bit easier?
- Other Questions on last week?



Multiple Dataset Analysis

- This is a critical problem in Government data work, data is often siloed by agency and is not easily obtained or combined, differences in data definitions prevent easy manipulation
- Don't be afraid to look for connections where none would be apparent
- Be rigorous with your statistics
- Ask yourself if the relationship makes sense in an actionable way
- Data can usually be joined on the following(if they're standardized):
 - Location
 - Time
 - Person
 - Context/Event



Multiple Dataset Analysis Potential Questions

- **311 - Streetlight outages vs property crime**
 - <https://www.citylab.com/equity/2014/02/street-lights-and-crime-seemingly-endless-debate/8359/>
- **Health inspections - property data, food info vs health violations**
 - <https://cseweb.ucsd.edu/~jmcauley/cse255/reports/fa15/036.pdf>
- **Crime vs temperature**
 - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415828/>
- **Education outcomes and housing instability**
 - <http://journals.sagepub.com/doi/abs/10.1177/0885412207299653>



Midterm!

- Pick 1 dataset, no need to combine multiple datasets
- BRAINSTORM!!!
- Turn in your code, visuals, and documentation
- During class next week we will do a review, everyone should take a couple of minutes to discuss their approach, what worked and what didn't, and what they learned by the process.
 - Note: this is **not** a formal presentation! More of a discussion.



For Next Week

- Reading on theory and practice
 - NOTHING!
- In R
 - MIDTERM!