City level data and the Integrated City Sustainability Database:
Modeling building efficiency with the ICSD

Rachel Krause, University of Kansas
RCN Sustainable Human-Building Ecosystems Workshop
May 19, 2016

Presentation overview

- Background on ISCD 1.0
  - Rationale
  - Where the data comes from
- Data Imputation and ICSD 2.0
- Assessing change over time
- ISCD data and the built environment
- Challenges and next steps

Project Rationale

- Local governance provides a valuable setting for the study of political decision-making, policy innovation, and organizational functioning.
- The lack of comprehensive public city-level data has led scholars to:
  - Focus on small number of case studies;
  - Conduct research with qualitative orientation;
  - Rely on surveys with low response rates and limited scope of information.

The ICSD…

- Is the most comprehensive nation-wide dataset of US municipal governments' sustainability programs and policies;
- Adds missing elements to the research infrastructure for the study of local government policy;
- Serves as a baseline to enable the study of policy change over time;
- Will be publically available in the future for use by researchers and practitioners.

Source data for "baseline" 2010-2011 ICSD

Survey Name	Sampling Frame	Respondents	Response Rate
ICMA	8,569 local governments with a population over 10,000	2,176	25.4%
NLC	1,708 mayors in cities over 10,000	442	26.6%
EECBG Grantee Implementation	970 municipal governments receiving EECBG awards, including all cities over 30,000	747	77.0%
Implementation of Energy Efficiency and Sustainability (IBM)	1,180 cities: all with pops over 50,000 and a random sample of 500 cities with pops btwn 20,000 and 50,000	679	57.5%
National Survey of Sustainability Mgmt	601 cities with populations over 50,000	263	44.0%
Municipal Climate Protection	664 cities with populations over 50,000	329	49.5%
Municipal Government Questionnaire	425 cities with populations over 50,000 that have explicit involvement in climate protection	255	60.0%

2010-2011 ICSD Data

- ISCD contains: 2,671 city observations and 1,107 variables

- Approximately 90% of cities 50,000+ responded to at least one of the 7 2010-2011 surveys
  - Each survey a little different and few cities responded to all 7;
  - Problematic in a multi-variate environment;

Common Options for Handling Missing Data

- Listwise or pairwise deletion;
  - Most common option: used in 97% of cases (Peng et al 2006);
  - May cause bias in remaining responses and decreased statistical power;
- Single imputation (e.g. mean replacement; last value replacement)
  - Can underestimate standard deviations and bias coefficients;
- Multiple imputation
  - Generates (20) imputed data sets and runs analysis across the many pooled

Filling in the Gaps in the ICSD

- Utilizes Multiple Imputations by Chained Equations (MICE) implemented in R
- Familiar regression techniques used to predict target variables
  - Predictor variables theoretically selected and empirically verified
  - Average of 95 predictor variables per each target variable

- Imputation conducted on a subset of ICSD data:
  - 683 observations (cities > 50,000)
  - 992 variables (all with < 80% missingness)
- 20 sets of imputed data generated

Filling in the Gaps in the ICSD

- Imputation conducted by the University of Kansas Center for Research Methods and Data Analysis (CRMDA)
- Supported by NSF SoO Grants 1461526/1461506/1461460

(Thanks for the help!)

Benefits of the Imputed ICSD Data

- Can utilize variables from across component surveys without losing observations;
- 100% coverage means no self-selection bias;
- Larger standard errors compensate for increased uncertainty of the data.

Drawbacks of the Imputed ICSD Data

- It’s all just a little more complicated;
- Descriptive statistics are tricky;
- Different disciplines regard imputed data with different degrees of approval/skepticism;
- Depending on the model, the increased number of observations may not be “worth” the larger standard errors.

ICSD at T_2 – 2015 Data

- Three national city sustainability surveys conducted in 2015 (+/-);
- Will enable assessment of change over time.

ICSD and Data on the Built Environment

All ICSD questions are linked to “general concepts”, resulting in the following relevant classifications:

- 47 variables on “Infrastructure”
- 124 variables on “Community-Focused Actions
- 140 variables on “Government-Focused Actions”
- 330 on “Energy”

Examples of Relevant Questions

(NLC, 15)

In terms of energy conservation and efficiency, does the city offer any of the following?

RESIDENTIAL	Available	Not Available	Don't Know
Financing programs or incentives to promote and/or facilitate energy efficiency and conservation upgrades/ retrofits	1	2	3
Energy audit	1	2	3
Financial incentives to promote and/or facilitate renewable energy installation	1	2	3
Education and awareness campaigns to promote energy conservation and efficiency	1	2	3
Green or white (reflective) roofing policies or incentives	1	2	3
Green building code requirements	1	2	3
COMMERCIAL	**Available**	**Not Available**	**Don't Know**
Financing programs or incentives to promote and/or facilitate energy efficiency and conservation upgrades/ retrofits	1	2	3
Financial incentives to promote and/or facilitate renewable energy installation	1	2	3
Green or white (reflective) roofing policies or incentives	1	2	3
Green building code requirements	1	2	3

Examples of Relevant Questions

To practice environmental sustainability, our city has:
[ ] Constructed new buildings based on LEED design or other environmental friendly standards
[ ] Utilized LEED or Commercial Interiors (CI) specifications to renovate existing buildings
[ ] Adopted a green standard as an official minimum criteria for new government buildings

(Hawkins 1c, 1d, 1g)

Examples of Relevant Questions

Does your jurisdiction offer LOANS, GRANTS, REBATES, or TAX ICENTIVES to upgrade or retrofit buildings? *(please select all that apply)*:

[ ] Loans
[ ] Grants
[ ] Rebates
[ ] Tax incentives

(Feiock, 23)

Challenges and Next Steps

- Spreading the word;
- Data distribution;
- Ensuring appropriate/correct use (particularly with imputed data).