Boston Area Research Initiative ● Northeastern & Harvard Universities
E-Mail: bari@northeastern.edu ● Web: www.bostonarearesearchinitiative.net

COVID in Boston: A Database for Research and Teaching

The Boston Area Research Initiative (BARI) is excited to announce a new resource for research and teaching focused on the COVID-19 pandemic: a multisource database that comprehensively captures how the dynamics of Boston shifted before, during, and after the shutdown in response to the pandemic. The effort updates, expands, and enhances the data sets already maintained by BARI's Boston Data Portal, and we are making it available to faculty for Fall courses. Given the diversity of data sets included, we anticipate that it will be useful to a variety of courses—including both substantive and methodological—across a range of disciplines. Important features include:

- A mixture of both administrative records and data scraped from online sources;
- Data on various urban systems and phenomena, covering topic areas such as commerce & economic activity, housing & development, urban planning & infrastructure, crime & disorder, and public health (see Table 1).
- An internal logic for linking data spatially, facilitated by BARI's Geographical Infrastructure for the City of Boston;
- Longitudinal tracking that permits analyses of changes from August 2020 back to March 2020, and, in many cases, months and years earlier;
- Data that support a wide range of methods, including geographical information systems (GIS), dynamic modeling, multilevel modeling, network science, and natural language processing (see Table 2);
- A combination of record- and neighborhood-level data, supporting both methods courses as well as those focused on substantive topics;
- Full variable-by-variable documentation for all datasets.

In addition, the process of construction and release of the data, as well as BARI's infrastructure, provide a unique opportunity for advancing research and teaching using the database. This includes:

- BARI's PhD Data Consultants, who will be available for clarifying questions and to brainstorm incorporation into curricula;
- Multiple courses will work with the data across Colleges and Departments, generating cumulative intelligence on the content and applications on the data that BARI will gather and share.

Online Data Sources

Craigslist: https://doi.org/10.7910/DVN/52WSPT

Yelp: https://doi.org/10.7910/DVN/DMWCBT

Airbnb: https://doi.org/10.7910/DVN/GXLZXF

Places of Interest: https://doi.org/10.7910/DVN/V2UTWH

Administrative Data Sources

Property Assessments: https://doi.org/10.7910/DVN/O2ADLG

Building Permits: https://doi.org/10.7910/DVN/B7DHBK

Code Violations: https://doi.org/10.7910/DVN/TD9YOY

Food Inspections: https://doi.org/10.7910/DVN/6MUQKX

CityScore: https://doi.org/10.7910/DVN/HRPNBQ

Table 1. BARI COVID-19 Datasets and Topics

	Commerce & Economic Activity	Housing & Development	Urban Planning & Infrastructure	Crime & Disorder	Public Health
Online Data					
Craigslist	✔	✔			
Yelp	✔				
Airbnb	✔	✔			
Places of Interest	✔	✔	✔		
Admin. Data					
Property Assessments		✔	✔		
Building Permits		✔	✔		
Code Violations				✔	✔
Food Inspections	✔				✔
CityScore			✔	✔	✔

Table 2. BARI COVID-19 Datasets and Methodological Applications

	Micro-Spatial (GIS point data)	Neighborhood Metrics (GIS polygon data)	Dynamic Modeling	Language Processing	Substantive Courses
Online Data					
Craigslist		✔	✔	✔	✔
Yelp	✔	✔	✔	✔	✔
Airbnb	✔	✔	✔	✔	✔
Places of Interest	✔	✔		✔	✔
Administrative Data					
Property Assessments	✔	✔	✔		✔
Building Permits	✔	✔	✔	✔	✔
Code and Property Violations	✔	✔			
Food Inspections	✔	✔			
CityScore			✔		✔