

# SPPUA 5262: Big Data for Cities Syllabus

Josiah Parry

- **Instructor:** Josiah Parry (j.parry@northeastern.edu)
- **Course Meets:** Thursdays 15:00-16:00 EST (3pm - 4pm)
- **Required Texts:**
  - R for Everyone by Jared P. Lander (available as e-book through NU library)
  - R for Data Science by Hadley Wickham & Garrett Grolemund

## Course Mission

The past decade has seen the rapid proliferation of large-scale digital data sets, from cell phone records to social media posts, from calls for government services to digitized historical records. With their unprecedented volume and detail, these new resources have opened up a window through which we can closely examine the behavioral and social dynamics of everyday life, and nowhere is this truer than in urban areas, which have the greatest density of both people and technology. This has given rise to urban informatics, and a renewed opportunity to understand the city, its people, and its neighborhoods, and to develop new policies and programs that improve urban life. Big Data for Cities takes an experiential approach that enables students to contribute to this young field. Given unprecedented times, we will apply our newfound skills to the urgent question of the shifting urban landscape in the age of COVID-19.

## Course Outcomes

### Course Content

Big Data for Cities introduces students to the burgeoning field of urban informatics with an eye towards five learning outcomes:

1. Describe the Field of Urban Informatics: Students will read articles and websites that expose them to the researchers, policymakers, practitioners, and companies that are incorporating big data into the study and management of the city, from which they will seek to articulate the themes and connections that constitute the field of urban informatics.
2. “Seeing” Neighborhoods through Naturally-Occurring Data: Students will acquire the methods necessary to manage, analyze, and interpret modern digital data (e.g., administrative records) using R statistical software, with a focus on translating data gathered from internet platforms into knowledge that meaningfully describes the dynamics of the city. This experiential aspect of the course is facilitated by the Boston Area Research Initiative (BARI), a center at Northeastern that focuses on urban informatics in the region. The data come from BARI’s efforts to build the Boston Data Portal, which gathers and coordinates data from many sources describing the region; our work will also be informed by the input BARI partners who work in the public, private, and non-profit sectors locally. Data work this semester will focus on data sets that capture the shifting urban landscape in the time of COVID-19.
3. Situating Internet Data in their Real-World Context: Naturally-occurring data are a rich source of information about a city, but it is necessary to verify what real-world events they actually describe and how they are meaningful. Students will undertake City Explorations in order to compare the patterns in their data to independent information about neighborhood contexts. They will also discuss their work with various local

stakeholders, including public officials and community leaders. These experiences will ground our interpretation of the data while also empowering us to utilize them to support the real communities that they describe.

4. Communicating Discoveries: Analysis is only as useful as the insights that can be derived from it. Students will practice reporting and presenting the results of data analyses both orally and in writing in forms that would be used to communicate to a range of audiences.
5. Creating Public Data Products: Students will modify and document improvements they make to the data sets used in the course. These will be stored in BARI's Boston Data Portal, thereby creating new data products that can be used by others, including our local partners and future classes.

## Learning Dimensions: SAIL

Big Data for Cities subscribes to Northeastern University's campus-wide emphasis on experiential learning and does so with a focus on service-learning—that is, applying our newfound technical skills and substantive knowledge to the real-world needs and challenges of local partners. We will do this directly through our analyses of data describing the shifting urban landscape in Boston during COVID-19.

As such, the value of the course will be more than just the five content-specific learning outcomes described above; it will also help students to develop general competencies that are useful across contexts. These learning dimensions are captured by the Center for Advancement of Teaching and Learning through Research's Self-Authored Integrated Learning (SAIL) framework. This course will particularly highlight three of their five dimensions.

SAIL Dimension	Definition	This Course
Intellectual Agility	Learners develop the ability to use knowledge, behaviors, skills, and experiences flexibly in new and unique situations to innovatively contribute to their field.	Leverage quantitative thinking for problem solving. Connect data to real-world phenomena. Adjust project goals with new insights.
Social Consciousness & Commitment	Learners develop the confidence, skills, and values to effectively recognize the needs of individuals, communities, and societies and make a commitment to constructively engage in social action.	Identify civic challenges. Apply data to better understanding and serving the city. Recognize how to use data to better understand and advocate around problems.
Professional & Personal Effectiveness	Learners develop the confidence, skills, behaviors, and values to effectively discern life goals, form relationships, and shape their personal and professional identities to achieve fulfillment.	Coordinate individual and group-level efforts on a collaborative team. Gather and leverage the perspectives of external partners. Select project directions strategically by aligning analytic opportunities with public interest.

## Course Format:

- The course curriculum is broken into three parts, moving through the increasingly sophisticated things we can do with data: Information, Measurement, and Discovery.
- The course will meet weekly remotely through Zoom.
- It is encouraged that all students attend Zoom meetings. Though you will not be graded on attendance.
- We will learn R, a freeware statistical program.
- The typical class meeting will be split into four parts:
  1. a discussion of the week's readings;
  2. workshop on analyzing data in R;
  3. short presentations of student work;
  4. reflections on service-learning and the social impact of our projects.
- The R component of the course takes the form of a flipped classroom, meaning that students are expected to do their technical readings and exercises in advance of class. We will use the class time as a workshop to review these skills.

## Data Collaborations:

- We will learn how to use R to manipulate, analyze, and visualize naturally-occurring data by working with data sets from BARI's COVID in Boston database, which draws from numerous administrative and internet sources. Each student will be assigned to a collaborative team that will analyze one of these data sets. Teams will be constructed using the preferences expressed by students in the pre-semester survey.
- The data collaborations will be the basis of weekly assignments and semester projects. All such assignments will be completed by individuals, but students are encouraged to share documents and modified data with each other. Teams will also work together to combine their advances with the data into data products. Major assignments and data products will be shared back with our municipal partners.
- Data will be made available to students via BARI's Boston Data Portal, hosted by Dataverse. **You will need to create a Dataverse account. See instructions for using Dataverse on Canvas under Course Material - > Technical Support Documents.**

## Assignments

Each week will entail: readings with reading responses; and R-based data explorations. Each unit will also include one city exploration assignment, one midterm (or, for the third unit, the final project), and one service-learning reflection.

## Reading Responses

- Readings are listed in the syllabus and will be posted on Canvas unless otherwise specified.
- Each reading assignment will be accompanied by a short reading prompt meant to guide reading. We will jumpstart class discussion using the Canvas discussion board in advance of class.
  - Students will turn in their response to the reading prompt, which must be at least 250 words, on the Canvas discussion board designated for that week by Thursday at noon.
  - Students will be expected to browse each other's responses as well. It will be required to respond to at least one other student's response by Thursday.
- Posts based on the reading prompt are graded out of 2 points: 2 = complete, 1 = incomplete (under word count / not entirely relevant to the reading / limited original thought), 0 = not turned in. Instructor will award partial points as appropriate.

- Responses to other students will be graded out of 1 point: 1 = complete, 0 = insufficient. There will be no partial credit. These responses do not have a length requirement, but need to comment substantively on the classmate's original post. (For example, "This is a cool idea!" would not be sufficient, but "I really enjoyed the way you framed sensor systems as in terms of their ethical challenges," would be.)

### **R-based Data Exploration**

- R-based data explorations will ask students to practice the skills learned that week in R to analyze their data set. They will be described in detail during each week's in-class workshop and posted on Canvas the week they are assigned.
- R-based data explorations will revolve around the data set that you have elected to focus on for the semester. Though students will work in collaborative groups everyone must turn in their own, independent work. Nonetheless, students are encouraged to collaborate with other group members and classmates in all work. These assignments are intended to build towards the midterm and final projects.
- The write-up should address: Background, or why is this interesting?; Analysis, or what you found out?; and Interpretation, or what does this mean? It should be completed in an RNotebook (.rmd) format that intermingles code and narrative.
- R-based data explorations will be submitted twice, first as a Draft and then as a Final product.

### **City Exploration Assignments**

- There will be one city exploration assignment during each unit of the course in which students will visit a part of the city to which they have never (or rarely) been, either in person or virtually. Each walk will have a different theme, connecting the data curriculum of that unit to the real-world context of the municipality whose data you are analyzing. Guidance on whether to visit places in-person or virtually will be contingent on current circumstances and all city explorations will offer online resources for a virtual option.
- For each city exploration assignment, students will complete a 4-5 pg. walk memo. Each walk memo must also include at least two pictures or screen shots from the exploration (not counted towards length).
- Class on the week of city exploration assignments will conclude with a series of 3 min. student presentations describing their city exploration. Each student will give one such presentation during the semester.
- Rubric for city exploration assignments included on Canvas where assignments are announced.

### **Service-Learning Reflections**

- At three points during the semester (see schedule), each student will write a 2-3 pg. (double-spaced) memo on: how their work in the course has been influenced by the perspectives of external partners and by exposure to real-world contexts; how those experiences have helped them to better understand the material and their data; and the ways in which the partnership has revealed the real-world utility of the skills. Students are also welcome to explore the limitations of these themes, or ways in which they would like to see them continue to grow.
- Service-learning reflections are graded out of 2 points: 2 = complete, 1 = incomplete (short / not fully engaging the question), 0 = not turned in. Instructor will award partial points as appropriate.

### **Midterms and Finals**

Each section of the course will conclude with a major project. These are:

- **Information:** README Documentation will provide a short series of facts derived from the data set that anyone working with the data might find informative. It will include code for replicating the analyses underlying these facts so that the report might be updated regularly or for subsets of the data.

- **Measurement:** Original Measurement will describe a novel set of measures that have been derived from your data set, code that can replicate their construction, and a description of their distribution across the city.
- **Discovery:** Inform, Measure, and Discover will combine the work from the two midterms with original analyses that utilize the measures and insights you have developed across the semester.
- At each of these junctures collaborative teams will work together to organize and document the additions and modifications they have made to their data set, creating a single data product they will upload to the Boston Data Library.
- Rubrics are posted with the fuller assignment descriptions on Canvas.

## Grading

- 10% Participation
- 10% Reading Responses and Other Small Assignments
- 10% Walk Memos
- 15% Week-to-Week R Assignments
- 15% Information Mid-Term Project: README Document 15% Measurement Mid-Term Project: Original Measurement 25% Final Project (Paper and Presentation)

## Academic Honesty:

- Students are expected to abide by Northeastern University's Academic Integrity Policy, which you can read at: <http://www.northeastern.edu/osccr/academicintegrity/>

## Incompletes:

- The grade incomplete (INC) is granted only where a student is approved to make up a single key requirement of the course, such as one of the major assignments.

## Other Expectations:

- Late assignments will be discounted 10% per day After a week, they will no longer be accepted. If an assignment was to be posted online in advance of class to stimulate discussion there will be no opportunity for late credit.
- E-mailed assignments, except in the case of an absence, will be considered 1 day late.
- Late assignments or absences will only be considered excused in the case of a doctor's note or evidence of an academic conflict.