11.s941: Big Data, Visualization, and Society

Data for a Changing Climate in Boston

Eric M. Huntley, MUP
ehuntley@mit.edu

TA: E. Phoebe Holtzman
epkh@mit.edu

DUSPViz RA: Laura Krull
lkrull@mit.edu

MW, 9:30-11:00, 9-217
https://github.com/ericmhuntley/big-data-spring2018

Course Description

Advances in urban science, the rise of big data, the drive to build smarter cities, and the widespread embrace of the open data movement are coalescing into new opportunities for planners to make data actionable through analysis and visualization. Furthermore, many of the most urgent problems facing cities today—perhaps especially those that are the consequences of a changing climate—are problems known, primarily, through their representations in the output of predictive models. These conjoined challenges have made it urgent for planners to develop the ability to collect, analyze, and compellingly visualize urban futures using large and often-messy data sets.
Our work will be oriented around climate change and housing development in Greater Boston. Sea level rise, increasing flood frequency, and intensifying urban heat island effects will impact housing markets, public health, and social vulnerability. Furthermore, the materiality of the built environment has a great deal to do with climate vulnerability; building types are differentially vulnerable to storm surge; open space, pavement types, and vegetation have much to do with the intensity of urban heat islands. Together, these pressures are poised to exacerbate already-acute housing problems in Greater Boston.

Students will develop the technical skills necessary to leverage provided data sets to compellingly visualize the impacts of a changing climate in Greater Boston. Monday class periods will be spent discussing planning and policy issues that will inform our analysis and engaging with current critical debates around the place of data in urban governance and society at large. Wednesdays will be spent learning widely-adopted programming languages and visualization tools, while also allocating time to constructive critiques of student work.

**What Will We Be Learning?**

Students will spend a great deal of their time in this class learning tools essential for working with big data. We’ll learn to script using the Python and JavaScript programming languages. We’ll use common libraries to collect, clean, and analyze data, and develop compelling visualization (e.g., pandas, beautifulsoup, georasters, d3.js, leaflet.js). We will also learn essential data acquisition skills including querying APIs and data scraping.

We’ll also be honing our critical faculties and becoming conversational in the important debates that are emerging around the role of big data in society and the history of data and computation in urban planning. This is especially important as DUSP convenes conversations around urban science—even as we learn to work with data, we will also spending some time thinking seriously about what is at stake both in our professional lives and in the worlds we build.

**Are There Prerequisites?**

No! And the class assumes no prior programming experience. However, *caveat emptor*: this is a coding and data analysis class. As such, while it is not required that you know how to code, it will be required that you approach programming with either some enthusiasm or, failing that, quite a lot of perseverance.
Github

In this class, we'll be making extensive use of Github, a distributed version-management platform. The course Github can be found at [https://github.com/ericmhuntley/big-data-spring2018](https://github.com/ericmhuntley/big-data-spring2018). Students will submit problem sets and work through their projects using Github repositories, allowing instructors to comment on and contribute to their code.

Resources

As we proceed through the class, you may find that you would benefit from supplementing the course tutorials and workshops with additional programming practice. Go for it! I recommend the following resources:

- Codecademy
- Learn Code the Hard Way
- Mike Bostock’s Blocks
- DUSPViz Web Mapping Workshop Series

There are also, of course, conventional books; some that you might find especially useful include:


Course Texts and Library

Students are not required to purchase any books for the course. All readings will be uploaded to the course Stellar site. In addition to the assigned readings, I will be providing a range of books and other materials that will be stored in the 5th floor CRON lab—use these for inspiration and as a way of stepping away from your own work by submerging yourself in the work of others.
Grading

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Sets</td>
<td>45%</td>
</tr>
<tr>
<td>Visualization Precedent</td>
<td>10%</td>
</tr>
<tr>
<td>Midterm Presentation</td>
<td>10%</td>
</tr>
<tr>
<td>Final Process Review</td>
<td>10%</td>
</tr>
<tr>
<td>Final Presentation</td>
<td>25%</td>
</tr>
</tbody>
</table>

Problem Sets

This class will involve completing weekly problem sets. These may seem somewhat unrelenting - this is on purpose! We find the best way to learn coding is practice, practice, practice. These assignments are essential for developing your skills. **All problem sets are due Wednesday at 6pm.**

Late submission will result in an automatic 10% deduction per day late, up to 3 days, after which I will no longer accept work. This is partially so that I can maintain my sanity, and partially so that you can maintain yours—assignments are due so regularly that if you are late it will really set you back.

You can choose to skip one problem set, and yes, I will grant extra credit to students who complete all of them.

Visualization Precedent

Over the course of the semester, each student will give a five-minute presentation on a web visualization that they find compelling. One of our first orders of business will be to schedule these presentations. These should be closely aligned with the goals of student work—they can demonstrate a visualization technique, a type of user interaction, or an interesting model for a final project. Two will be given per class period.

Help! I’m Frustrated!

Technical learning is that it is best accomplished by doing, failing, walking it off, and doing some more. You should expect to spend quite a bit of time fighting through moments of frustration in this class. However, there is a good reason to endure this frustration: mastering a technology always involves learning how to
seek help with problems that exceed your understanding. Use each other as resources! Share resources amongst yourselves!

**Office Hours**

The teaching staff will be available for one three-hour block (TBD) during the week to help students. Of course, it is impossible that a single block of time will align with everyone's schedule, so Eric will also be available for one-on-one consultation appointments on request. It will be very, very helpful if you can submit a summary of your difficulty and a link to the relevant Github repository 12 hours in advance so that teaching staff can spend a bit of time diagnosing the problem - make your and our lives easier by allowing us to spend less time hunting down bugs and more time refining your code!

**Email**

The teaching staff may take up to 24 hours responding to student emails, and we make no guarantees that we will respond as quickly over weekends. While it is truly remarkable that an email can reach our pockets instantaneously, this does not imply that our response must be equally instantaneous!

**A Note on Graduate School and Mental Health**

Academic environments are taxing places; for reasons structural, institutional, and interpersonal, they sometimes do not lend themselves to what most reasonable people would think of as human flourishing. I went to graduate school. In fact, I went to graduate school twice. And without dwelling on the issue, I will say that I am sensitive to the unique strain that the environment we inhabit can, too often, be to the detriment of our mental health and wellbeing.

I have two points here: 1) MIT offers a range of resources for students experiencing mental health challenges and I would really encourage you to be proactive about availing yourself of them; and 2) do not hesitate to let me know if you're struggling. It is, of course, not my intent to know the details of what might be troubling you, but simply to let you know I am sensitive to these matters and that help is available.
Schedule

Week 1

February 7: Course Introduction, Github and Python Setup

Week 2

February 12: Big Data Orientation


February 14: Introduction to Python

- Problem Set 0 Due

Week 3

February 20: Climate Change Adaptation in Boston


February 21: Introduction to Data Wrangling and Queries with Pandas

- Problem Set 1 Due

Week 4

February 26: Critical Questions for Big Data


**February 28: Gathering Data from APIs, Scraping, Exploratory Plots**

- Problem Set 2 Due

**Week 5**

**March 5: Climate Change and Housing Affordability**

- Readings TBA.

**March 7: Data Analysis, Aggregation, and Exploratory Maps**

- Problem Set 3 Due

**Week 6**

**March 12: Climate Modeling**


**March 14: Raster Analysis in Python**

- Problem Set 4 Due
Week 7

March 19: Midterm presentations.
March 21: Midterm presentations.

Week 8

Spring Vacation

Week 9

April 2: Algorithms and Risk


April 4: Web Development (HTML, CSS, JS)

- Problem Set 5 Due

Week 10

April 9: Web Visualizations
April 11: Introduction to D3 - Data Loading, Data Binding

- Problem Set 6 Due

April 12-14: Eric at AAG Conference

Week 11

April 16

No Class
April 18: 2D Plotting, Data Graphs
April 20–21: Climate Changed Symposium

Week 12

April 23: Mapping/Geographic Visualizations 1
April 25: Choropleth Maps
  • Problem Set 7 Due

Week 13

April 30: Mapping/Geographic Visualizations 2
May 2: Mapping Raster Data in D3

Week 14

May 7: Responsive Visualizations
May 9: User Interaction, Event Handlers
  • Problem Set 8 Due

Week 15

May 14: Project Work
May 16: Pin-Up

Week 16

May 23: Final Review