

## CS171 Final Project: Innovative Cities

**Team Members:** Moe Hadhrawi, Yu-Ann Wang

**1. Background and Motivation.** *Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.*

Moe is working in the Changing Places group, MIT Media Lab. He's investigating what makes cities creative/innovative using a tangible interactive decision support system. Moe is interested to connect tangible with graphical data representation to reduce cognitive load in understanding complex systems.

Yu-Ann has lived in multiple cities for work, including NYC, Los Angeles, Hong Kong, and Jakarta. In each place, she was struck by how urban planning has facilitated or disrupted people's ability to get together and work effectively. She is currently taking two courses on digital innovation and sustainable cities. She hopes to tie together these two courses and study why certain cities are so effective in breeding entrepreneurship and innovation.

**2. Project Objectives.** *Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.*

The primary questions that we are trying to answer is that are there certain elements of urban planning or urban composition that is correlated with local innovation? A lot of researchers have investigated this from a sociology and economic standpoint (e.g., Edward Glaeser in the Harvard economics department). but we believe that very few people have investigated this from a data scientists and data visualization perspective. Depending on the quality of data we get, we believe that we could investigate these questions down to the neighborhood level.

While it may be hard to contextualize benefits because we are proving a correlation rather than a correlation, many cities (Denver, NYC, Austin, etc.) are seeking to rebrand their cities and strengthen their support on innovation. Our research may prove to be another helpful data point for these cities.

**3. Data.** *From where and how are you collecting your data? If appropriate, provide a link to your data sources.*

We will be collecting data from nonprofit and city government sources. We hope to be able to draw comparisons between different cities (for now, we're focused on Cambridge/Boston and NYC).

**For NYC:**

Transit data: <http://web.mta.info/developers/index.html>,  
<http://www.nyc.gov/html/dot/html/about/datafeeds.shtml>

NYCEDC: <http://www.nycedc.com>

Small Business Administration: [www.sba.gov/](http://www.sba.gov/)

Other open data (including wifi hotspot locations):

<https://nycopendata.socrata.com/>

**For overall United States:**

Code for America API:

[http://catalog.data.gov/dataset?res\\_format=api&res\\_format\\_limit=0&page=4](http://catalog.data.gov/dataset?res_format=api&res_format_limit=0&page=4)

(although mostly on the healthcare side)

**For Boston:**

We're working with a local civic data meetup group to obtain APIs.

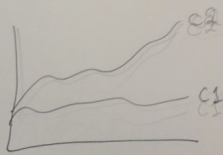
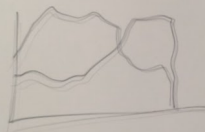
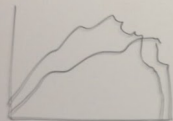
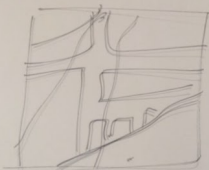
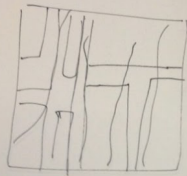
**4. Data Processing.** *Do you expect to do substantial data cleanup? What*

*quantities do you plan to derive from your data? How will data processing be implemented?*

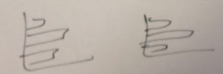
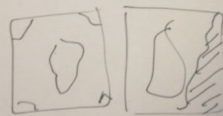
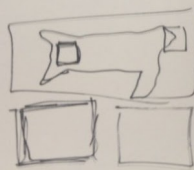
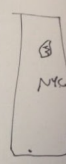
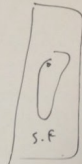
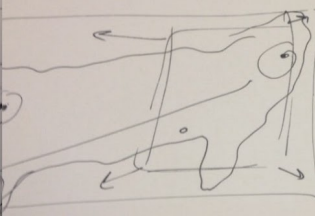
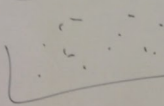
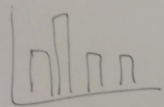
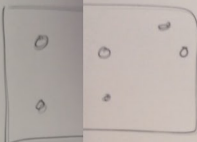
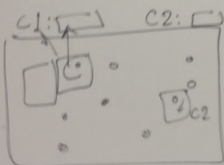
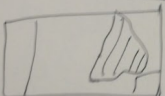
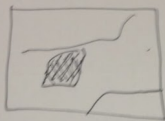
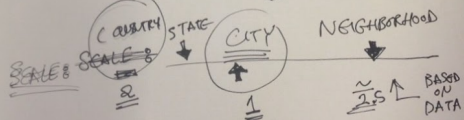
We don't anticipate substantial data cleanup. However, we've been browsing some of these API forums and note that not all APIs are functional. We anticipate that this might be a challenge for us down the road when we set up our web page to pull the information.

**5. Visualization.** *How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.*

We're looking to compare two cities side by side, first on a map basis, but also on a line chart and bar chart basis.



- POPULATION
- EMPLOYMENT
- // CREATIVE INDUSTRIES
- SMALL BUSINESSES



**6. Must-Have Features.** *These are features without which you would consider your project to be a failure.*

We'd like to compare two different cities (right now we think NYC and Cambridge would be good choices). In addition, we'd like to have at least two proxies for innovation (whether it would be patents filed, number of small businesses in an area, or another metric). We'd also like to have at least two dependent variables which do show some relationship (could be anything from population density, population composition, public transit, green buildings, etc.)

**7. Optional Features.** *Those features which you consider would be nice to have, but not critical.*

Ideally, we would be able to compare more than two cities. In addition, we'd like the analysis to be granular enough to target the neighborhood area (i.e., Kendall Square is a lot more different than Brookline). We foresee a challenge being that the data collected from each city is not very consistent.

**8. Project Schedule.** *Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.*

Bulletpoints in bold are designated by CS171 teaching staff.

- **Thursday, March 13: Project proposal due (part of Homework 3)**
- Thursday, March 20: Preliminary hypotheses have been formed. Majority of data should be gathered at this point (reach out to relevant professors and city government officials). All meetings with urban planning/real estate professors should be scheduled at this point.
- Thursday, March 27: Data is fully gathered, initial sketches have already been shared within team and with project TF. Team starts to test out

different government APIs, ensure data is clean and functional.

- **Thursday, April 10: Functional project prototype due - this should implement the majority of our must-haves. This is also the midpoint of the project, team should do a midpoint evaluation of what has gone well so far, what we would like to do differently going forward.**
- **Week of April 14: Project review with the TFs - major objectives is to accept prototype criticisms from TFs and also bounce off ideas for prioritizing nice to have features.**
- Thursday, April 21: About half of nice to have features should be implemented. If any of them cause issues, meet as team and discuss alternatives for data. Meetings with professors from first round of brainstorming should be scheduled so they can provide evaluation on that story we're telling with our prototypes.
- Thursday April 28: Incorporate in feedback from professors. Start assembling deliverables such as process book and screencast.
- **Thursday, May 1: Projects due (including screencast)**
- **Thursday, May 8: Best project presentations and prizes**