**URBAN DATA ANALYTICS: COURSE RESEARCH PROJECT**

For this project you will be using ‘big’ and/or open data sources to answer a research question of your own choice. The project is meant to build on the techniques learned during the course. You will work in groups of up to three students and can choose to work with the same partner from Assignment 1. We advise you to use the CP101 Piazza site to find partners, based on your topic(s) of preference.

For the purposes of this project, we define ***big*** data as data that is not designed for a specific research project but ***found*** online, e.g., data obtained via webscraping or gathered by someone else who made it available to the public (see Appendix). We define ***open*** data as spatial data intentionally made available by publicly funded institutions or local governments through open data portals (see Appendix). Some examples of transactional data include business licenses, crime reports, 311 complaints, traffic patterns, bike sharing ridership, Clipper card transactions, etc.) that can be used without restriction and is published in a machine-readable format; data that has been created and gathered through public investment. You may want to supplement your analyses with ***designed*** data such as data from the American Community Survey or the LEHD program.

**Note**: We will do our best to answer questions posted on Piazza within 24 hours. Questions posted the day assignments are due may not be answered by the time you need to submit your work.

**Deliverables**

The research project consists of three (3) consecutive deliverables. Each deliverable is designed to build on the previous one.

**Deliverable A: Research Proposal**

The first deliverable (Assignment 2) consists of a **one-page** (500 words) **research proposal** describing your main **research** **question**, along with **data** and **methods** you plan to use to answer it. Please identify at least one existing data set for one of your proposed sources (rather than simply data you would like to have) and provide a link to the data set. Please submit your research proposal via bCourses no later than **Saturday, June 13th @11:59 pm**.

***Grading***

* Research question: 40 points
* Data sources: 30
* Methods: 30

**Deliverable B: Story Map**

Deliverable B (Assignment 3) consists of a brief **story map**. A story map is a set of visuals (maps, graphics, images) that tell a story about a place, which in your case is the city, neighborhood, area, or project you are examining to answer your research question. Your story map must also present some preliminary findings. Story maps lead users along a path or timeline with links to related content. Rather than just depicting an event or place, story maps bring perspective and context. You will need to create your story map on **WordPress** (or a similar online content manager) and post the link to your story map on bCourses no later than **Sunday, June 21st @11:59 pm**.

Your story map must include at least **four (4) visuals.** Of these four visuals, present at least **two (2) interactive** **maps** created using **CARTO**. These two maps must be of your own creation. Other visuals might come from secondary sources such as Other visuals might come from secondary sources including Social Explorer, Google Earth, OnTheMap, Google Street View, photographic archives, (or more CARTO!). At a minimum, each visual must include both the visualization and some text providing context. Text should not exceed **1,200 words** (not including citations).

***Sources of Inspiration***

* [John Snow’s Cholera Outbreak Story Map](https://www.arcgis.com/apps/MapSeries/index.html?appid=3b7b69c040e64afe88ea8e577e71829a) [ESRI Story Map Library]
* [Katrina +10: A Decade of Change in New Orleans](https://www.arcgis.com/apps/MapSeries/index.html?appid=597d573e58514bdbbeb53ba2179d2359) [ESRI Story Map Library]
* [Mapping Segregation in DC](https://jmt.maps.arcgis.com/apps/MapJournal/index.html?appid=061d0da22587475fb969483653179091) [ESRI Story Map Library]
* [Transit Deserts in the Bay Area](https://www.ocf.berkeley.edu/~bellanh/) [cp101 story map]
* The [Urban Displacement Project](https://www.urbandisplacement.org/) [see for example case studies]

***Grading***

* Storytelling: 30 points
* Map quality: 25
* Explanatory text: 15
* Other visuals: 25
* Page design: 5

**Deliverable C: Complete Research Project**

Building on Deliverable B, the last deliverable (Assignment 4) consists of an online post containing the complete and final version of your research project. You are expected to significantly improve and extend the work you submitted for Deliverable B and keep the text to a minimum 1,700 words (maximum 5,000 words). You will need to create your final post on **WordPress** (or a similar online content manager) and post the link on bCourses no later than **Tuesday, June 30th @11:59 pm**. Your final project MUST be on a different URL than your story map – that means not overwriting the story map you submitted.

Your final deliverable must include at least **twelve (12) visuals.** Of the twelve visuals, present at least two (**2) new interactive maps** created in **CARTO** (or other mapping tool), in addition to the two maps you presented on the story map. In other words, your final post must have at minimum **four (4) interactive maps**. At least **eight (8)** of the visuals presented must be of your own creation. Other visuals might come from secondary sources including Social Explorer, Google Earth, OnTheMap, Google Street View, photographic archives, etc.

***Sources of Inspiration***

* [How does satisfaction with Bay Area transit systems differ across social media versus official survey data?](https://cp101transitsatisfaction.wordpress.com/) [cp101 project]
* [Urban Impact of Hurricane Sandy in New York City](https://heidid.github.io/cp101-assignment3/index.html) [cp 101 project]
* [You are what you eat: Exploring the connection between food and the city](https://www.ocf.berkeley.edu/~ytse/) [cp101 project]
* [Drawing Discrimination](https://ryanabest.com/ms2-2019/thesis/) [Parson’s School of Design Master Thesis]
* [The Influence of Tech Shuttle Stops on Local Rental Listing](http://www.cp101.org/assets/Philips%20Wolfson%20Batlan%20Asssignment%203.pdf) [cp101 project]

***Grading***

* Research design: 10 points
* Data acquisition, cleaning, and wrangling: 20
* Data analysis process: 25
* Visuals: 25
* Text: 15
* Ethical data considerations: 5

# APPENDIX

# Where to find your big data?

There are many great websites out there to scrape! You may also be able to identify a dataset that somebody else has already scraped for you. For urban data, we often use sites with an explicit spatial component, such as Craigslist (and other rental sites) or Yelp. Other sources for big data include:

* [InsideAirbnb: Adding Data to the Debate](http://insideairbnb.com/behind.html), created by Murray Cox (an independent digital storyteller and activist), provides data scrapped from Airbnb
* [Bay Wheels](https://www.lyft.com/bikes/bay-wheels/system-data) provides geolocated trip data of people using this bike sharing system serving the Bay Area. Other micromobilty companies opening their data to the public include [Capital Bike Share](https://www.capitalbikeshare.com/system-data) (DC area) and [City Bike](https://www.citibikenyc.com/system-data) (NYC)
* [Purple Air](https://www2.purpleair.com/) sells air quality sensors and make the data sensed available to the public (more info [here](https://publiclab.org/notes/jiteovien/08-28-2018/download-analyze-your-purple-air-data))
* Ride hailing services provided by companies such as Uber and Lyft (a.k.a. TNC) provide travel data through the [SFCTA web portal](https://www.sfcta.org/projects/tncs-today). NYC TNC data was curated and published by [FiveThiryEight](https://fivethirtyeight.com/) through their [GitHub repository](https://github.com/fivethirtyeight/uber-tlc-foil-response/)

For additional ideas about where to go to scrape, look at [UCI’s](http://archive.ics.uci.edu/ml/) [Machine Learning data repository](http://archive.ics.uci.edu/ml/)

# Where to find your open data?

Most cities now have an open data portal. Just Google the city name and “open data.” Some municipal open data portals include:

* [NYC Open Data](https://opendata.cityofnewyork.us/)
* [London Data Store](https://data.london.gov.uk/)
* [Data SF](https://datasf.org/opendata/)
* [Open Toronto](https://open.toronto.ca/)

You might also want to explore public datasets via [Amazon’s data website](https://aws.amazon.com/public-datasets/) or [Quora](https://www.quora.com/Data/Where-can-I-find-large-datasets-open-to-the-public)

# Still want more data?

In addition, we have collaborated with a graduate urban analytics class CP255 to build [this list](https://docs.google.com/document/d/102CIekgNeSHhcNSCiAmPgKIZxP6AzM7qCJ-WaIPCAdg/edit) of data sources.