**Exploring, visualizing, and challenging the city open data portal**

Zoë Wilkinson Saldaña

2019-07-22 (note: still revising! will finalize on 07-23)

## Project overview

What would it look like to teach the skills for challenging and critiquing open data alongside the data retrieval itself? How do we help learners understand their data visualization praxis in the context of issues of power, exclusion, and representation?

In this project, I propose creating a short [bookdown](https://bookdown.org/) text. This text will function as a guide to working with open data portals, with a focus on feminist data visualization praxis and approachability for newcomers to data science. I will focus on a few key areas: (1) exclusion and representativeness in dataset selection and coverage; (2) strategies for visualizing bias and ambiguity; (3) community efforts at counter-portals and other city data activism, and (4) visualizing questions of bodies and gender. I will focus on the tidyverse ecosystem of tools in R to fetch, explore, and visualize datasets.

My hope is that this guide succeeds as a learning tool for data science newcomers that subverts the idea of individual, context-less skill acquisition. I also hope to create an opportunity for collaboration with other educators and developers who may be interested to work further with the bookdown text.

## City data portals: history & community activism

****

*Cities rated in the 2017 US Open Data Census. Source:*[*http://us-city.census.okfn.org/*](http://us-city.census.okfn.org/)

In the United States, a growing number of cities (now in the hundreds) curate, store, and make accessible datasets via an open data portal. While the main federal-level aggregator of datasets, data.gov, directly ingests data from a limited number of cities, most remain independent or linked only via a shared vendor platform. Socrata, the largest such vendor, offers a platform and collection of data tools designed to help governments at all levels offer open data services (they claim to work with over 1,000 government organizations).

In some cases, city open data portals are the outcomes of particular policy imperatives and executive orders. Chicago began changing its practices in sharing data online in 2010 to make more easily accessible the results of Freedom of Information Act (FOIA) requests in December 2012. Other cities have noticed the trend towards open data and joined in, leveraging the experiences of their peers and funding from organizations like the Sunlight Foundation. Saint Paul Mayor Chris Coleman announced his city’s data portal in such a context in a 2016 State of the City address.

What determines the structure, mission, and ethos of a city open data portal - and in particular, which datasets are included and which are excluded? I am particularly interested in how we can answer these questions. Some sources of insight include:

1. **Internal knowledge, perspective, and decision-making**: what do we know about the specific decisions made, partnerships, etc. that guide the development of open data portals? What is the role of internal stakeholders in curating the dataset collection?
2. **Technical tools:** E.g. automated detection of bias
3. **Civic tech community of practice perspective:** tools like the US Open Data Census attempt to identify gaps via a “surveying experts” model.
4. **Community activism:** There is a rich and developing history of community groups organizing to protect their interests by achieving representation, privacy, correction of bias, etc. in the portal dataset collection. This is especially true for marginalized and vulnerable populations who are portrayed in datasets.

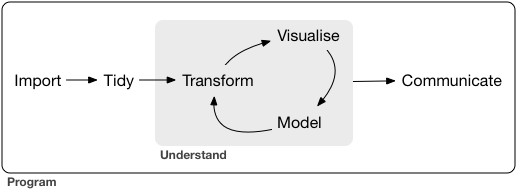
Among these perspectives, community activism is unique – it does not start with the premise of better or more tools, but rather the needs and experiences of community members. In this project, I plan to build upon the [Equitable Open Data Report (2015)](https://datajustice.github.io/report/) and the [Our Data Bodies Playbook (2019)](https://www.odbproject.org/wp-content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf) as frameworks for critical data science learning.



*From* [*Our Data Bodies Playbook (2019)*](https://www.odbproject.org/wp-content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf)

## Critical & feminist data science

In terms of specific practices and methods, the project will focus on exploratory data analysis with city datasets: finding, fetching, parsing, tidying, visualizing, and modeling datasets. I will start with the *R for Data Science* model of exploratory data analysis:



From [*R for Data Science*](https://r4ds.had.co.nz/introduction.html)

I believe that this iterative framework for exploring data can also help us frame what critical, feminist data science praxis might look like. My project will attempt to re-situate these processes in the frameworks of critical practice, and in particular the work of D’Ignazio and Klein (2016) and Oliphant (2017). In particular:

1. The project will introduce exploring data as a process of surfacing **power, omission, labor, and bodies**. In particular, I will focus on datasets that represent individuals but continuously challenge the “individual” represented and the specific context of the creation of that dataset.
2. Similarly, the project will emphasize the importance of **historical, communal, and political context** in questions of open data portals.
3. I will encourage the **learner to situate themselves** and their identity within these very systems. I will discourage the individual scholar/learner model and instead emphasize opportunities for collaboration, community consultation, and sharing analysis as open data.
4. I will place special emphasis on the role of **exploratory data visualization** as a means of provoking critical inquiry. My chosen examples will center on ways to show who is omitted from datasets & strategies for visualizing bias alongside other insights.

## Works cited

D’Ignazio, C., & Klein, L. F. (2016). Feminist data visualization. In Workshop on Visualization for The Digital Humanities (VIS4DH), Baltimore. IEEE.

Oliphant, Tami. (2017). A Case for Critical Data Studies in Library and Information Studies. Journal of Critical Library and Information Studies, 1(1).