

**Visualizing NYC Broadband
White Paper
Will Luckman
Spring 2022**

Visualizing the Political Economy of the Digital Divide

The digital divide is a problem that is technological in nature—it concerns access to computer-enabled digital communications networks—but it isn't a *technical* problem, it is a political one. Technologies are created through social processes, and have social impacts which must be considered. As Stephen Graham pointed out back in his 2002 essay “Bridging Urban Divides? Urban Polarization and Information and Communications Technologies (ICTs),” the effect of the digital divide is a deepening inequality that, due to the scale and reach of these networks, can be documented from the hyper-local to the international level, producing a global subaltern on the wrong side of a planetary borderline of access.¹ But while much good scholarship has focused on the ill effects of the digital divide, if it is a problem we want to *solve*, then we also need to pay close attention to its origins. And we can do that by examining the *politics* of the technology at the heart of the issue: internet infrastructure.

Infrastructure is often physical. Broadband is literally a network of cables and wires laid beneath the ground. But that is only half the story. Infrastructures—like all forms of technology—are “amalgam[s] of technical, administrative, and financial techniques” (Larkin, 2013).² And, as such, infrastructures have specific characteristics that make them useful for exploring certain relationships. There is a cooperative nature to infrastructures due to both their scale, and their function as a substrate on which other forms of social activity occur, and they

¹ Graham, Stephen. “Bridging Urban Divides? Urban Polarization and Information and Communications Technologies (ICTs).” *Urban Studies* Vol.39, No.1 (2002): 33-56

² Larkin, Brian. “The politics and poetics of infrastructure.” *Annual review of anthropology* 42 (2013): 330

require large amounts of capital investment to construct. Therefore, examining the details of their construction and use, beyond the technical and physical dimensions, offers a window into relationships between the powerful actors and institutions building infrastructure, and the types of flow—financial, physical, communicative—they facilitate or obstruct through their design.

Data visualization is an excellent method through which to research and analyze these political aspects of infrastructure. Using visualization techniques we can capture unseen flows of power, embody them, and present them in a format almost as tangible as the physical, built infrastructures they produce. In this way, visualization can also become a tool of struggle. As Jasper Bernes writes in “Logistics, counterlogistics and the communist project,” to develop a counterlogistics, “we might try to graph the flows and linkages around us in ways that comprehend their brittleness as well as the most effective ways they might be blocked as part of the conduct of particular struggles” (Bernes, 2013).³ Data visualization can be deployed in this manner beyond the realm of logistics. In order to understand any complex, invisible challenge and confront it—in this case the capital flows and policy decisions that create an infrastructure that reinforces inequality—we must first convert those complex, shifting, multiscalar oppositional forces into a comprehensible target.

For this project I wanted to examine a hyper-local example of technological infrastructure and visualize the political economy that brought it into being. My goal was twofold: first, to use the case of NYC to demonstrate larger trends in the relationship between the state and private capital, and second, to challenge this relationship and offer an alternative solution to the digital divide as it exists currently in New York City.

³ Bernes, Jasper. “Logistics, counterlogistics and the communist project.” *Endnotes 3* (2013): 194

Finding the Data

I began by looking for any data I could find on New York City broadband provision, with an eye towards economic data. I started my search on NYC OpenData,⁴ a web portal set up in 2013 following the passage of Local Law 11 of 2012 as a repository for newly required municipal data reporting.⁵ I first searched the keyword “broadband” and found a dataset called “Broadband Adoption and Infrastructure by Zip Code”.⁶ That’s also led me to “Broadband Data Dig - Datasets”⁷ where I found franchise coverage maps. I also perused a list of the most popular datasets. From there, I pulled down some demographic data and also found a link out to City budget data hosted on Checkbook NYC, via the City Comptroller’s office.⁸ This last resource ultimately furnished the bulk of my data.

Checkbook NYC hosts aggregate City budget data dating back to 2010, and currently updated through the 2022 budget. The site presents this data with an interactive API and bar chart visualizer that allowed me to drill down into specific budget data by date, agency, and other categories including revenue and spending. It also has advanced search functions that allowed me to set specific parameters and search for specific terms. After acquainting myself with the site and the overall budget, I began targeting contract data on the “Big 3” last-mile Internet Service Providers (ISPs) and their various corporate aliases (Verizon, Time Warner, Charter, Spectrum, Altice, Optimum, Cablevision). First I pulled datasets on contract spending for each of these contractors. Then I went back and pulled data on revenue collected from these same entities. I

⁴ <https://opendata.cityofnewyork.us/>

⁵

<https://www1.nyc.gov/site/doiit/initiatives/open-data.page#:~:text=Facilitating%20greater%20access%20to%20technology.the%20NYC%20Open%20Data%20Portal>

⁶ <https://data.cityofnewyork.us/City-Government/Broadband-Adoption-and-Infrastructure-by-Zip-Code/qz5f-vx82> (last updated 9/17/21, accessed 5/6/22)

⁷ <https://data.cityofnewyork.us/dataset/Broadband-Data-Dig-Datasets/ft4n-yqee> (last updated 5/9/22, accessed 5/6/22)

⁸ <https://www.checkbooknyc.com/>

also began to pull data on spending by various agencies. By looking at the Department of Information Technology and Telecommunications (DoITT), I was able to examine franchise revenue collected, as well as identify a telecom spending category. By isolating that telecom category and returning to the overall City budget, I was able to output data on telecom spending by agency. I was curious about broader trends in government spending with technology contractors, so I also pulled data on NYPD spending with Shotspotter, Inc. and Microsoft.

To supplement this data I turned to a few other sources. To find more information on the ISPs, I looked at public financial reporting required by the SEC, in the form of 10-K statements.⁹
^{10 11} I pulled some additional demographic information from the Census. And I also looked at research put out by the Internet For All coalition. That group, (which, full disclosure, I am involved in organizing), released a white paper in 2021 outlining its plan for a municipal broadband system in NYC which includes a projected budget for network construction, as well as some additional information on the digital divide in NYC and the failures and malfeasance of the private ISPs.¹²

Unfortunately, I also faced some limitations in what I was able to find. Although I was able to uncover franchise territory maps, I wasn't able to find *coverage* maps. To truly tell the story of the digital divide in NYC it would be incredibly useful to have block-by-block maps of where the cable was run and which units were wired. That information can only be supplied by the ISPs themselves. If they have issued detailed reports publicly on the scope of their network, I haven't yet been able to turn them up. I was also unable to find corporate revenue by region. The more I am able to focus narrowly on New York City the more useful my analysis, so I would like

⁹ <https://www.verizon.com/about/sites/default/files/2020-Annual-Report-on-Form-10-K.PDF>

¹⁰ <https://d18rn0p25nwr6d.cloudfront.net/CIK-0001702780/b070b9e3-2e49-4c48-aa8a-281fd8467b31.pdf>

¹¹ <https://ir.charter.com/static-files/a798e04f-1fad-4157-aaf2-3d2866459f51>

¹² <https://internetforall.nyc/research.pdf>

to have data on how much revenue is being generated by these ISPs just within the city. While I was able to generate an approximation of the revenue they're generating from City government contracts, I wasn't able to isolate retail or business-to-business revenue generated only within the NYC market.

There were some limitations within the available data and data structure of the City budget as well. I wanted to know exactly how much money each of the Big 3 providers has paid to the City annually since the start of their franchise agreements. While I was able to locate some revenue data tied to the Big 3, it appears to be projected revenue, and in the case of Verizon, the projection data appears to be inaccurate (as noted in my visualization). I was able to make cuts of reported DoITT franchise revenue, but those revenue numbers are not identified by franchisee. Interestingly, in my research into the franchise agreements, I found that the franchisees are required to submit quarterly reports to DoITT along with accompanying payments. I've written to DoITT requesting access to those reports, but I haven't seen them as of yet. I was also unable to isolate *internet* provision within the City budget. Internet, telephone, and possibly other communications spending was grouped together in the single expense category, "TELEPHONE & OTHER COMMUNICATNS". As a result, some of the comparative spending data in my argument lacks the specificity I desired (although I believe the overall argument, that these sums of money are relatively minor compared to both the overall budget and and the private revenue at play, holds).

Working with the Data

I already described some of the work that went into editing and pulling my datasets from the Checkbook API. But some additional work was required to clean my data before bringing it into Tableau, and once within Tableau, to make it legible. My first task was cleaning the contracting data. Because I used the search function I had exported a number of separate but related datasets that I needed to join, categorize by vendor, and purge of any false positives. I used Python to automate this process, joining the datasets and running a for loop to identify the contracts as belonging to either “Verizon”, “Spectrum”, “Optimum”, or “Other” (for any contracts not associated with one of these providers that may have accidentally slipped through). Once that data was brought into Tableau, I was able to isolate the proper expense category (“TELEPHONE & OTHER COMMUNICATNS”) and graph by the provider categories I had set up.

Additionally, I was able to use Tableau to calculate certain revenue and spending totals I wanted to look at, some of which I then added to a new data set along with other information I collected from the 10-Ks and Internet For All report, so I could do the “big number” comparisons I present in my visualization. The demographic data proved difficult for me to wrangle along with the other data I was preparing, so I opted to use Tableau data layers instead of the datasets I had downloaded. This was easier in some ways, but also had some major limitations, including what data was available, how it could be combined with other data, and an inability to include that data in the tooltips.

Telling the Story

The final form of this project flowed from my research. I knew I wanted to tell a story about broadband, the digital divide, policy, and the relationship between the government and private

contractors, but it wasn't clear at the outset exactly what that story would be. I've described how I found and cleaned my data above, but it wasn't a straightforward process; rather, it was iterative. I had to find data, analyze it, then take those findings into consideration while shaping a narrative. Once I pulled down a few data sets, I began brainstorming the questions I could ask of it: *How much are City agencies currently spending on telecommunications? How much have they spent since 2010? How much is the City earning from franchise agreements? How much are the private providers profiting off of New Yorkers? Who has internet access and who doesn't?* Some of these questions pointed in new directions for research, and as I found new data, I asked new questions of it. When I ran up against the limitations I described above I had to revise my inquiries accordingly.

Eventually I was able to answer many of my questions, and in turn begin to build an argument. I realized that as part of that argument—one ultimately about the failures of private provision of public goods—I also wanted to make the case for an alternative. First I wrote a short outline of the argument and the beats I wanted to hit, and then noted the particular datasets I had (or could try to find) that I would use to make each point. I decided to begin the argument by grounding it in the problem of the digital divide, then present information on how broadband provision currently works in New York, show why that system isn't solving the problem, and finally demonstrate that other solutions to the problem are possible. Next, using that outline, I created a storyboard.

Design Decisions

Because I wasn't working from a single data source or single dataset, and because I wanted to tell a multifaceted story, I knew I would need to use a variety of visualizations. I also

knew that whatever form the project took, I would need to incorporate some text to guide the narrative. My goal for both the text and the visualizations was to keep each as simple and economical as possible. Instead of trying to cram a lot of information into just a few complicated charts, I opted to use many visualizations, using each to make one point clearly, then summarizing that point with a piece of accompanying text. Each pairing would clearly express one small piece of this story, building on the previous, in sequence. To do this effectively, in addition to keeping each item I presented simple, I needed to find the right order, rhythm, and tone with which to present them.

I am trying to illustrate the relationship between money and broadband, the government and private industry, and the scale of the money involved. These goals, particularly scale, led me to incorporate a number bar charts and bubble graphs. Large numbers can be hard to conceive of, and sometimes something that seems large is actually quite small when put in perspective. Bar charts and bubble graphs are good ways to convey this information quickly while also suggesting a relationship between the elements being presented. Presenting them in sequence allowed me to shift scales quickly, while hopefully bringing the viewer along with me.

As I set out to create my visualization I ran into a challenge that steered me away from my original plan: I couldn't figure out how to include visualizations made from unrelated datasets in a single document. This threw off my plans for a sequential story, and I opted to set up some of those sequences where I could, and otherwise try to make the visualizations I intended, and then string them together in a blog post with accompanying text. This was not an optimal solution. My original goal was to create my story visually with minimal text, and then use a blog post to provide additional context on broadband provision in NYC. The plan was to use the blog to summarize my findings, explain some features of the current regime in a bit more

depth, and offer thoughts on the greater trends in politics, technology, and ideology it illustrates. Since I now had to use the blog to sequence and present my visualizations, I tried to achieve some of the rhythm and argument originally intended, while also blending it with the longer textual context and analysis. It quickly became too difficult to parse.

Thankfully I got some great feedback, inspiration, and practical advice through conversations with my peers and professors. The illegibility was obvious, the text had become too central and the relationship between the visualizations—how they informed each other—was lost. My classmate Alex McGlinchy gave me the technical key to solving this problem by showing me how to combine all my visualizations within Tableau. In addition, she is an excellent designer, and I took some inspiration from the cohesive aesthetic of her project and tried to bring that to my own. Another piece of design inspiration came from my professor Michelle McSweeney, who referred to these projects more than once as “posters.” I realized a poster-style vertical format with large, bold text and clean images would be great for combining multiple elements and telling the story I intended. And, once I realized I could restructure my project along the lines I originally envisioned, another professor, Filip Stabrowski, provided excellent advice on how to separate out my visual argument and reframe and expand my accompanying texts.

The final design of the visualization took a form close to my original vision. And by incorporating the notes and inspirations described above, I was able to elevate the entire project beyond that first concept. I wanted my aesthetic and my argument to be in sync: bold, simple, legible, cohesive, persuasive. I decided on big headlines and summaries in a chunky sans-serif font. I presented additional context and information in a more delicate, thin serif that was better for in-depth reading and would be legible at smaller sizes. I worked to reduce the text to the bare

essentials and embraced injecting my own opinion, editorializing openly in the service of my argument. For overall color, I decided on an alert yellow on a dark field. I used transparency to float my visualizations in this field, helping them break their individual frames and strengthening the relationship between visualizations on each page.

Because I used Tableau data layers for the first maps my color options were limited, but I settled on a red-to-green color scheme associated with traffic lights and balance sheets, hopefully resonant with the themes of economic and communications flow on display here. Instead of overlaying the maps, I wanted to show two parallel versions; different data, same map.

To convey the next piece of the story, I combined all my franchise revenue data onto one page and all my telecom spending data on the next page. In this way I was able to work through one important component at a time. I used the size with which I presented the visualizations to help direct the viewer to the most important information (for instance, I reduced the size of the franchise territory maps and telecom spending by agency). I assigned colors to each of the Big 3 providers and used them consistently throughout to avoid having to provide legends every time. I tried to pick colors already loosely associated with each through their branding.

In some ways the maps, the spending, and the revenue data are all there just to provide context and set-up to deliver the final argument in the form of a single, staggered visualization. The bubble chart of “big numbers,” occurred to me a full day after I completed my original storyboard. I knew I needed an idea to cap things off and pull all the research together. Ultimately, I don’t know if it could exist as a visualization without the work of the preceding visualizations to contextualize it, but I think it is the most compelling and effective piece of this project. Again, the goal here was to draw comparisons between different pools of money, and then use scale to offer some additional perspective. If we are able to “zoom out” on this issue,

which is often framed as a problem of limited resources, it should help us realize this is a different sort of problem entirely.

What's Next?

As noted in my introduction, the digital divide is a problem that can be engaged from any number of methodological and theoretical approaches. The research analyzed and presented here is limited regionally to New York City, and temporally to the easily accessible budget data, which begins in 2010. Even within those boundaries there is much more to explore. I opted to look at broadband infrastructure, but the digital divide is expressed across many layers of infrastructure. The websites, platforms, and other types of content hosted via these networks represent their own form of infrastructure, developed by actors with their own set of divergent economic interests. I focused largely on the origin of the digital divide, but its impacts—not just along economic lines, but also along racial and other demographic lines—have much to teach us about how technologies (and the politics behind them) shape new forms of social life and interaction, specifically in relationship to education, incarceration, gentrification, and more. I focused narrowly on at-home broadband provision, but the full story of New York's networks also includes the new rollout of 5-G and WiFi connections, middle-mile networks, administrative and business networks, and the connection points between local networks and national and international broadband infrastructure. And with a longer timescale than I used here, we could examine the role previous forms of infrastructure like phone lines—their regulation and financing—played in influencing the shape of the broadband infrastructure that followed. We

could also trace back municipal spending on such systems to locate trends, or even historical contingencies where conditions may have allowed for a different outcome.

It is a rich vein of study, and one I will continue to build upon. And as we attempt to understand these issues, I think it is important to do so with the conviction that something better is possible and the purpose of achieving that better outcome. Hopefully a strength of the limited scope I deployed here is that it can have a targeted, practical application.