Tools to Source, Store, and Explore Data

Leslie Potts and Usha Yeruva

Data x Power Fellowship Project

Usha Yeruva

Usha (she/her) is based in Brooklyn, New York and is currently working as an Analytics Engineer at the DNC. She's passionate about creating sustainable data infrastructure and long walks in Prospect Park.

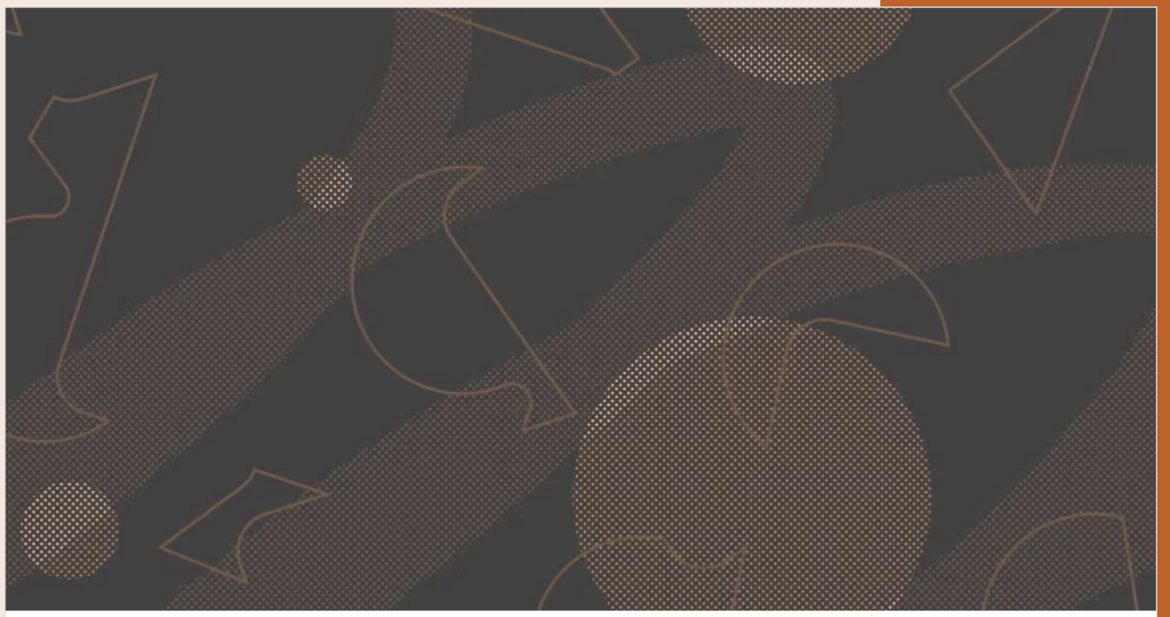
Leslie Potts

Leslie (she/they) is currently based in her hometown of Detroit, Michigan (Anashinabee land). She currently works as the Evaluation & Research Manager at the Youth Engagement Fund. Like many of us, Leslie has recently been focusing a lot on how to stay connected to community and loved ones amid a highly virtual, COVID conscious daily reality.

Introduction

This guide is:

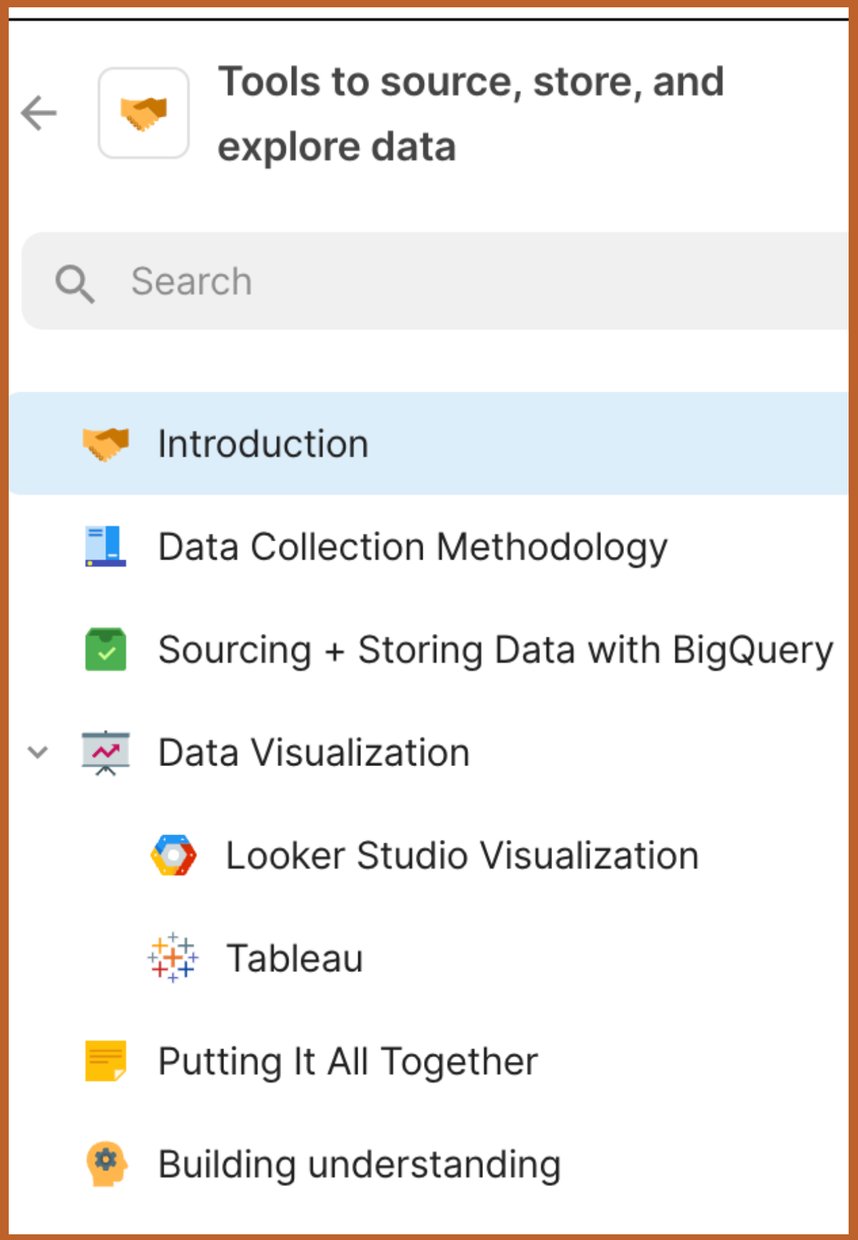
- For folks who have some experience with data analysis tools but are hoping to become more advanced users
- For people hoping to use publicly available data to investigate something they care about
- For people who are hoping to learn how to do more in BigQuery, Looker Studio, and Tableau.



Introduction · Tools to source, store, and explore data

Tools To Source, Store, and Explore Data

Coda / Jan 16



Resource Overview (Coda)

01 **Introduction**

02 **Data Collection Methodology**

03 **Sourcing + Storing Data with BigQuery**

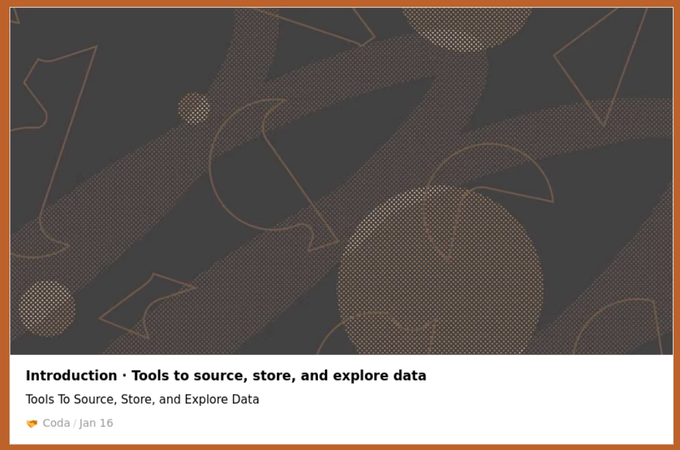
04 **Data Visualization**

4.1 **Looker Studio Visualization**

4.2 **Tableau**

5 **Putting It All Together**

6 **Building Understanding**



Building a Data Collection Plan

Working with ACS data

Other Public Data Resources

Methodology

- Narrowing down the scope of the project
- Being specific allows you to pick truly relevant data sources

Proponents

- Breakdown of data — block groups, census tracts, counties
- Understanding what sort of data is available, how it can be analyzed along with other data

Research

- Other sources — community nonprofits, government sources, commercial data
- Technical understanding of how these pieces of data fit together
- Collecting shapefiles for visualization that match our data collection parameters

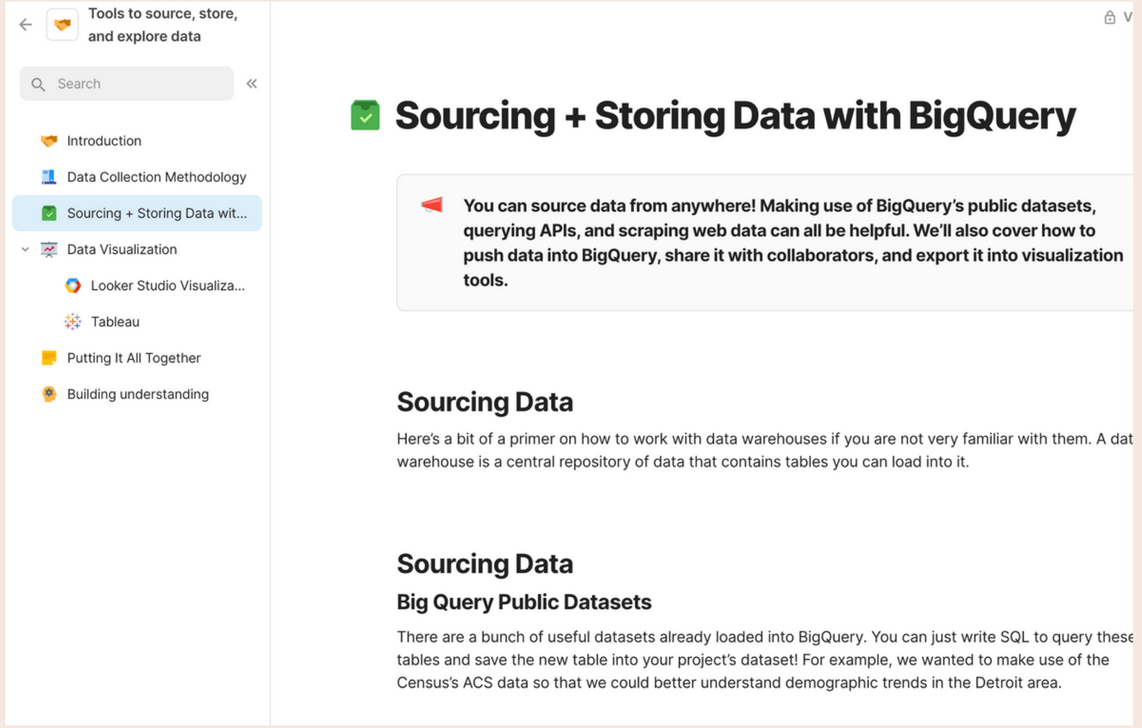
Cleaning and Storing Data

Data Cleaning

- Lots of helpful data lives on the internet in messy, unhelpful spreadsheets.
- Data formulas and organization strategies in Google Sheets allows you to create clean data ready for analysis.

Storing Data

- If you are trying to do a relatively low cost project and you aren't familiar with data warehousing, Google BigQuery is an optimal solution.
- You can upload data from CSVs, write queries, or use the API to add data into BigQuery.



Sourcing Data

Writing SQL to source data

BigQuery Public Datasets

- Public datasets like the ACS are already loaded into BigQuery, so you can just write SQL to bring that data into tables you can begin to analyze.

Querying APIs

Proponents

- Using the Python requests library to pull data out of an API
- Understanding how to use the BigQuery Python API to move data into a table

Scraping Web Data

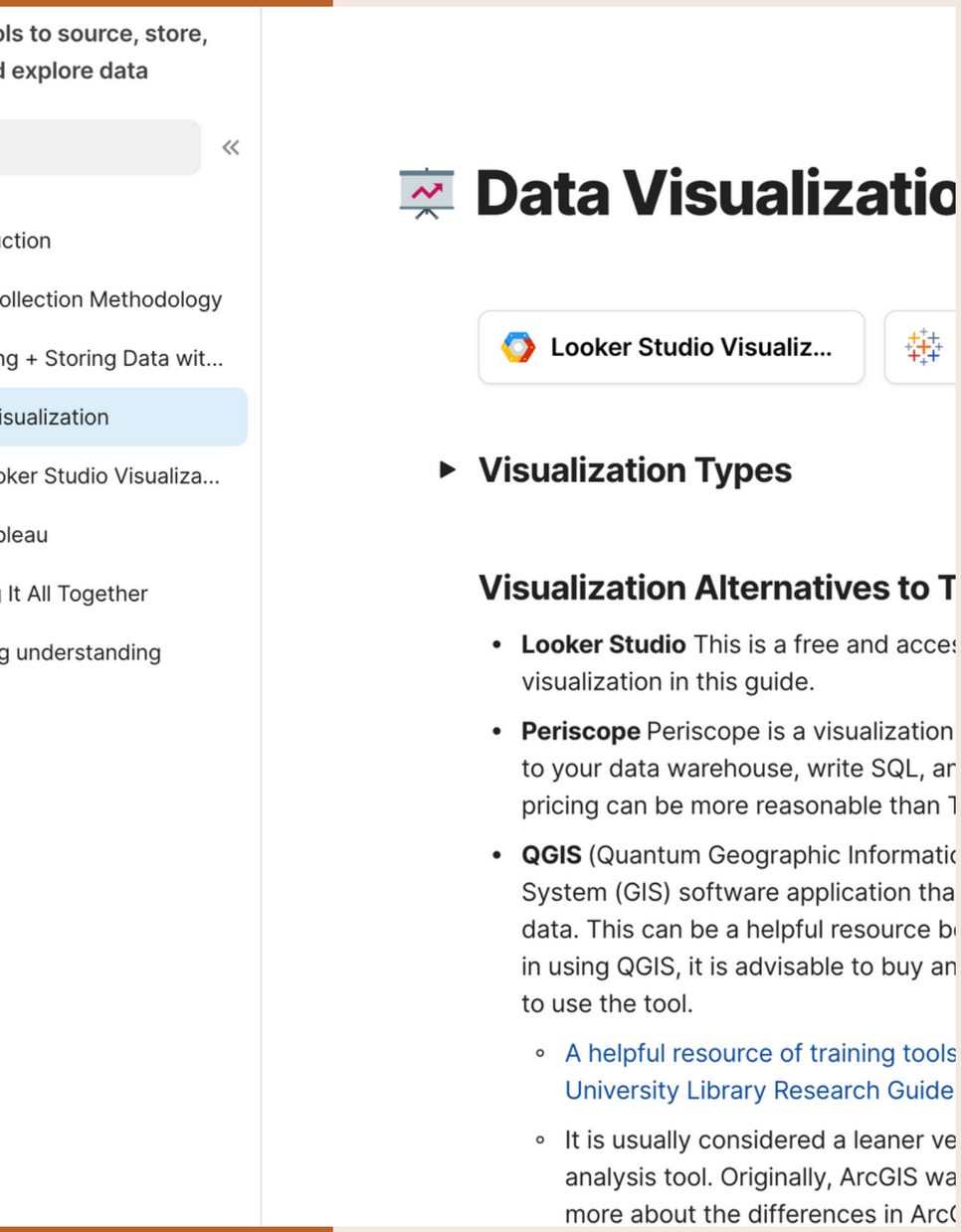
Research

- Using Python's BeautifulSoup library to pull data out of a web page.
- Helpful if the data you need is stored on the web but not in a spreadsheet or API.

Analysis in Looker Studio

Link to our Looker Studio visualization

- How to connect BigQuery and Looker Studio
  - Easy to integrate, simply select your dataset.
  - Write queries in BigQuery and export results into Looker Studio.
- How to add charts and graphs
  - Drag and drop interface, easy to put together simply analyses.



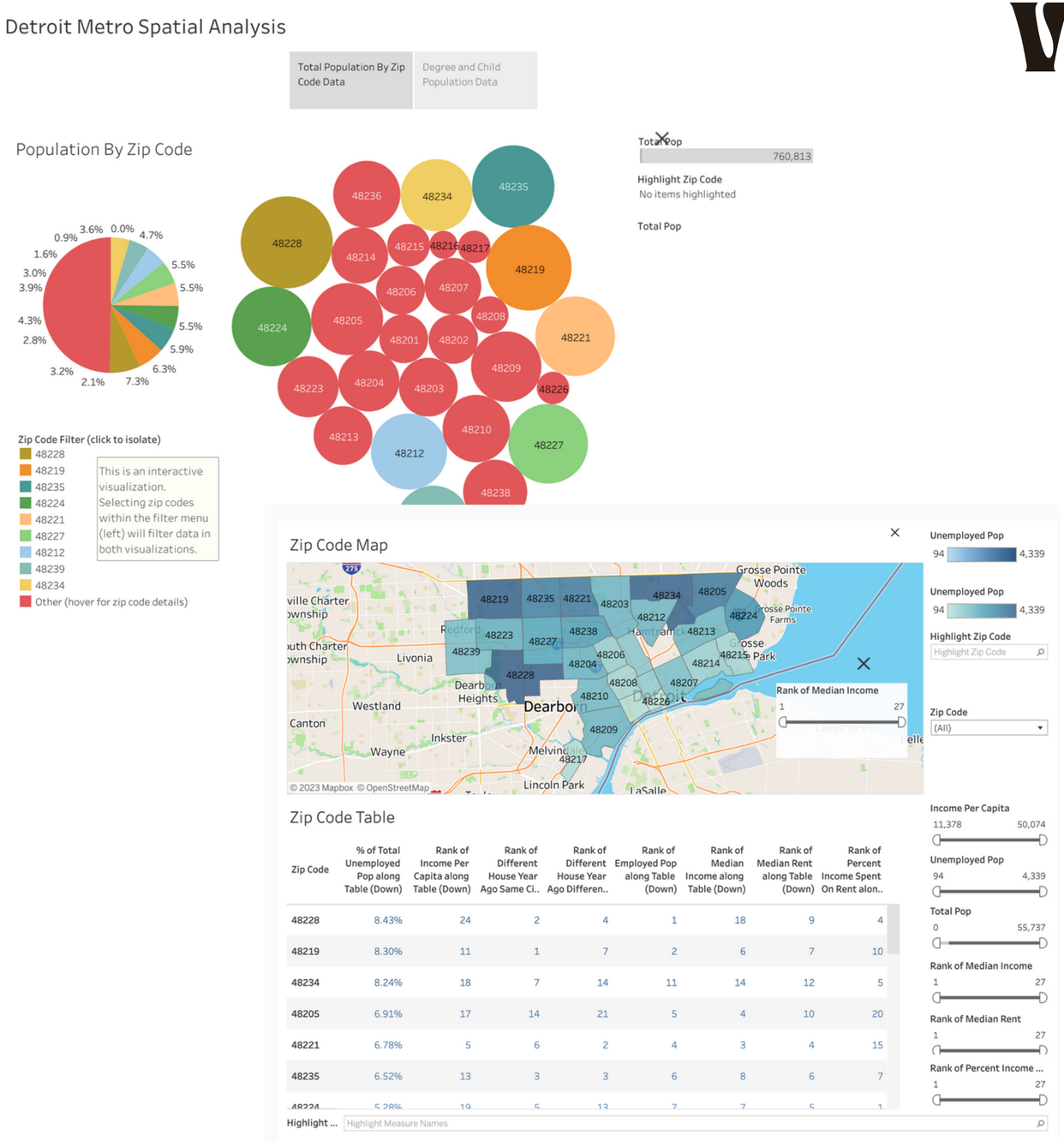
No Coding? No Problem.

While knowing coding languages such as SQL can make it much easier to analyze datasets, it is possible to do analysis and visualization without expertise in this language.

However, you will need to understand data cleaning, data collection, and the functions of the specific tool you are using for your analysis.

As an alternative, we would suggest becoming better acquainted with data cleaning in Google Sheets or Microsoft Excel, as well as a visualization tool like QGIS or Tableau if you are interested in making geospatial visualizations with your data.

Visualization Examples



Looker Studio, Detroit Metro Spatial Analysis:

1- Census Tract Exploration

2- Helpful Stats

Tableau:

- **Detroit Metro Spatial Analysis:** Detroit Community Health Survey Data
- **Population Data By Zipcode** (pictured, top)
- **Detroit Zipcode Metrics Mapping** (pictured, bottom)

Contact

Usha Yeruva

Focus Areas:

Data Warehousing
Big Query
SQL Code
Data Cleaning (SQL code utilization)
Coda.io utilization

email: ushayeruva21@gmail.com

Leslie Potts

Focus Areas:

Data Visualization
ACS Data Cleaning and Analysis
Tableau
QGIS
Data Cleaning (Excel, Google Sheets utilization)

email: contact.lesliepotts@gmail.com