Tools to Source. Store, and Explore Data

Leslie Potts and Usha Yeruva

Data x Power Fellowship Project



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.

Usha Yeruva



Introduction This <u>guide</u> 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.



Tools To Source, Store, and Explore Data 🤝 Coda / Jan 16

Introduction · Tools to source, store, and explore data



Tools to source, store, and explore data

Q Search

Introduction

Data Collection Methodology

Sourcing + Storing Data with BigQuery

Data Visualization



Tableau

- Putting It All Together
- Building understanding

Resource Overview (Coda)

- Introduction 01
- 02 **Data Collection Methodology**
- Sourcing + Storing Data with BigQuery 03
- **Data Visualization** 04
 - 4.1 Looker Studio Visualization
 - 4.2 **Tableau**
- 5 **Putting It All Together**
- **Building Understanding** 6





Introduction \cdot Tools to source, store, and explore data Tools To Source, Store, and Explore Data 😎 Coda / Ian 16

Building a Data Collection Plan

Working with ACS data

Other Public Data Resources

Methodology • Narrowing down the scope of the project

- sources

Proponents

- 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
- our data collection parameters

• Being specific allows you to pick truly relevant data

• Breakdown of data – block groups, census tracts,

• Collecting shapefiles for visualization that match

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.

Fools to source, store, and explore data									
Q Search «									
🤝 Introduction									
📃 Data Collection Methodology									
Sourcing + Storing Data wit									
👻 👳 Data Visualization									
🧿 Looker Studio Visualiza									
💠 Tableau									
Putting It All Together									
Building understanding									

Sourcing + Storing Data with BigQuery

You can source data from anywhere! Making use of BigQuery's public datasets, guerying APIs, and scraping web data can all be helpful. We'll also cover how to push data into BigQuery, share it with collaborators, and export it into visualization tools.

Sourcing Data

Here's a bit of a primer on how to work with data warehouses if you are not very familiar with them. A dat warehouse is a central repository of data that contains tables you can load into it.

Sourcing Data

Big Query Public Datasets

There are a bunch of useful datasets already loaded into BigQuery. You can just write SQL to query these tables and save the new table into your project's dataset! For example, we wanted to make use of the Census's ACS data so that we could better understand demographic trends in the Detroit area

Sourcing Data

Writing SQL to source data

Querying APIs

Scraping Web Data



BigQuery Public Datasets



Proponents

- data out of an API



Research

- - data out of a web page.



• 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.

• Using the Python requests library to pull

• Understanding how to use the BigQuery

Python API to move data into a table

• Using Python's BeautifulSoup library to pull

• 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.



<mark>6</mark> Looker

b Looker Studio

ols to source, store, l explore data

<<

ction

ollection Methodology

ng + Storing Data wit...

isualization

oker Studio Visualiza...

bleau

It All Together

g understanding

💌 Data Visualizatio

🜍 Looker Studio Visualiz...

彝

Visualization Types

Visualization Alternatives to T

- Looker Studio This is a free and access visualization in this guide.
- Periscope Periscope is a visualization to your data warehouse, write SQL, ar pricing can be more reasonable than 1
- QGIS (Quantum Geographic Informatic System (GIS) software application tha data. This can be a helpful resource be in using QGIS, it is advisable to buy an to use the tool.
 - A helpful resource of training tools University Library Research Guide
 - It is usually considered a leaner ve analysis tool. Originally, ArcGIS wa more about the differences in Arc(

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.

Detroit Metro Spatial Analysis



Visualization Examples

- Analysis: **1**- Census Tract Exploration
- 2-Helpful Stats

Tableau:

4.339

•

50,074

4.339

55,737

ank of Median Income

- Detroit Metro Spatial Analysis: Detroit Community Health Survey Data
- top)

0 2023 Map	box © OpenStree	etMap	Linc	oln Park	LaSalle	The second			
Zip Coo	de Table								Income Per 11,378
Zip Code	% of Total Unemployed Pop along Table (Down)	Rank of Income Per Capita along Table (Down)	Rank of Different House Year Ago Same Ci	Rank of Different House Year Ago Differen	Rank of Employed Pop along Table (Down)	Rank of Median Income along Table (Down)	Rank of Median Rent along Table (Down)	Rank of Percent Income Spent On Rent alon	Unemployed P 94
18228	8.43%	24	2	4	1	18	9	4	Total Pop 0
8219	8.30%	11	1	7	2	6	7	10	Rank of Mer
8234	8.24%	18	7	14	11	14	12	5	1
8205	6.91%	17	14	21	5	4	10	20	Rank of Me
8221	6.78%	5	6	2	4	3	4	15	1
8235	6.52%	13	3	3	6	8	6	7	Rank of Pero
8224	5 28%	19	5	13	7	7	5	1	0
ighlight	Highlight Meas	ure Names							

Looker Studio, Detroit Metro Spatial

• **Population Data By Zipcode** (pictured,

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

Data Visualization Tableau QGIS

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

email: ushayeruva21@gmail.com

email: contact.lesliepotts@gmail.com

Leslie Potts

Focus Areas: