Discovering our Community through Data

Provost Academy
Day 1, Morning
August 16, 2021

Introductions!

Joseph Yurko

Teaching Assistant Professor

Dr. Yurko joined the School of Computing and Information (SCI) in 2019, and is currently teaching machine learning and data science related courses. Before joining SCI, he was a data scientist in the manufacturing industry, and prior to that a consultant in the nuclear industry. Dr. Yurko has extensive experience building advanced predictive models to solve challenging problems across engineering disciplines. Those models have been used to optimize performance and interpret system behavior within the presence of uncertainty. Dr. Yurko's industrial experience expanded upon his PhD thesis which used computational Bayesian inference algorithms to quantify the uncertainty in nuclear system simulations. He received his PhD from MIT in Nuclear Science and Engineering in 2014.

Contact

5419 Sennott Square
jyurko@pitt.edu

Eleanor Mattern

Director, Sara Fine Institute
Teaching Assistant Professor

Eleanor "Nora" Mattern is a teaching assistant professor at the University of Pittsburgh's School of Computing and Information. Her teaching and research interests include archives and digital curation, community-centered information work, civic engagement, and information policy and ethics.

Nora previously worked as librarian and researcher at the University of Chicago Library, where she focused on designing and delivering scholarly communications services and supporting digital preservation activities. Before that, she held a joint appointment at the University of Pittsburgh as a visiting faculty member in the University Library System and the School of Computing and Information and a Postdoctoral Researcher in Digital Scholarship Services at the University Library System. She earned her PhD in Library and Information Science from the University of Pittsburgh in 2014.

Contact

Information Sciences Building, Room 610
emm225@pitt.edu
412-624-9331
Personal Website

Alejandro Ciuba

alejandrociuba@pitt.edu

At Pitt, 2021-2022 is ….

Year of
Data and
Society

Let’s begin to get to know one another!

In groups of 5, jot down --

What are words that come to mind when you hear “data”?

Data 101: What is Data?

Today we will talk about:

Data types and formats
Open and Civic Data
Where we find open data
Privacy and ethical issues

Acknowledgement: Western Pennsylvania Regional Data Center and the Carnegie Library of Pittsburgh. *Data 101 Toolkit: What is Data?* [Training Materials].

NEWS ANALYSIS

Census Shows a Nation That Resembles Its Future More Than Its Past

For Democrats, there was much to cheer in the growth of cities and suburbs. But Republicans, imperiled by the falling white population, are still well positioned for redistricting.

The New York Times

A group supporting President Biden's campaign in San Antonio, where Hispanic residents now roughly equal non-Hispanic whites. Texas is on the cusp of becoming a true battleground state. Christopher Lee for The New York Times

News August 12, 2021

Allegheny County gains population for first time in six decades, says census

By Ryan Deto

View of Downtown Pittsburgh

Pittsburgh City Paper

After decades of population loss, Allegheny County bucked its decades long trend and gained population, according to the newly released 2020 Census figures. The county added 27,230 residents compared to 2010, and the population of Pennsylvania's second largest county now stands at 1,250,578.

How does Census data impact communities?

Why We Ask

The 2020 Census is easy. The questions are simple.

The census asks questions that provide a snapshot of the nation. Census results affect your voice in government, how much funding your community receives, and how your community plans for the future.

When you fill out the census, you help:

- Determine how many seats your state gets in Congress.
- Guide how more than $675 billion in federal funding is distributed to states and communities each year.
- Create jobs, provide housing, prepare for emergencies, and build schools, roads and hospitals.

POPULATION COUNT (NUMBER OF PEOPLE LIVING OR STAYING)

We ask this question to collect an accurate count of the number of people at each address on Census Day, April 1, 2020. Each decade, census results determine how many seats your state gets in Congress. State and local officials use census counts to draw boundaries for districts like congressional districts, state legislative districts, and school districts.

United States® Census 2020

AGE AND DATE OF BIRTH

We ask about age and date of birth to understand the size and characteristics of different age groups and to present other data by age. Local, state, tribal, and federal agencies use age data to plan and fund government programs that provide assistance or services for specific age groups, such as children, working-age adults, women of childbearing age, or the older population. These statistics also help enforce laws, regulations, and policies against age discrimination in government programs and in society.

Data || Real Life

“Data represents real life. It is a snapshot of the world, in the same way that a picture catches a small moment in time. **Numbers are always placeholders for something else, a way to capture a point of view** — but sometimes this can get lost. Failing to represent these limitations and nuances and blindly putting numbers in a chart is like reviewing a movie by analyzing the chemical properties of the cellulose on which the images were recorded.”

Giorga Lupi, Data Humanism, The Revolution will be Visualized.

http://giorgialupi.com/data-humanism-my-manifesto-for-a-new-data-wold

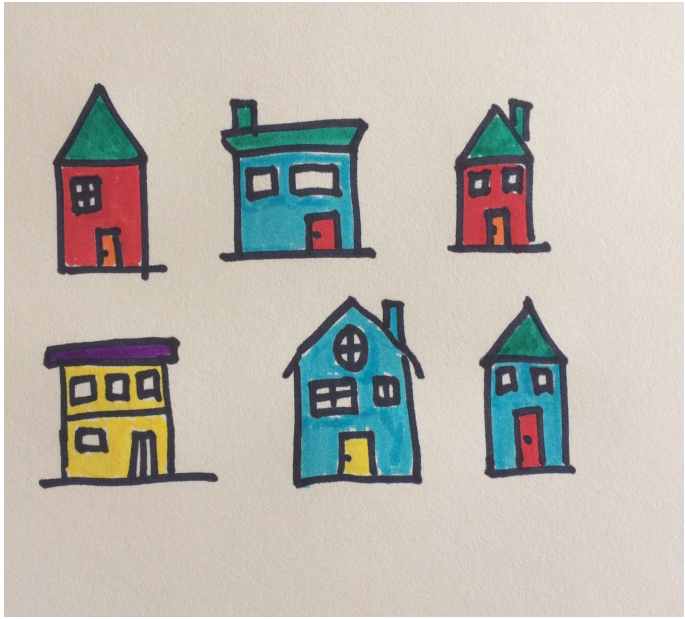
Data

Quantity

Stories of the residents

Style

Neighborhood



Sale Price

Occupancy

Color

Year Built

Condition

From Data 101

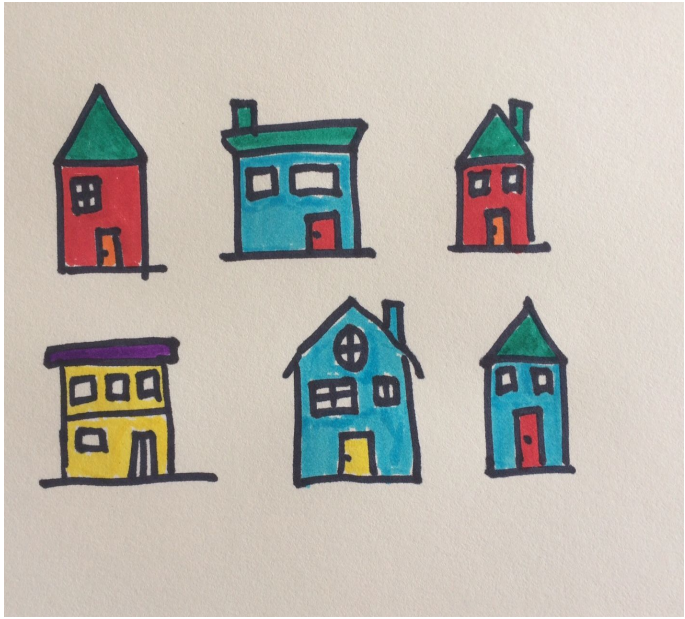
Quantitative Data

Quantity **(6)**

Stories of the residents

Style

Size **(1000 sq ft)**



Average Sale Price **($105,000)**

Color

Condition

From Data 101

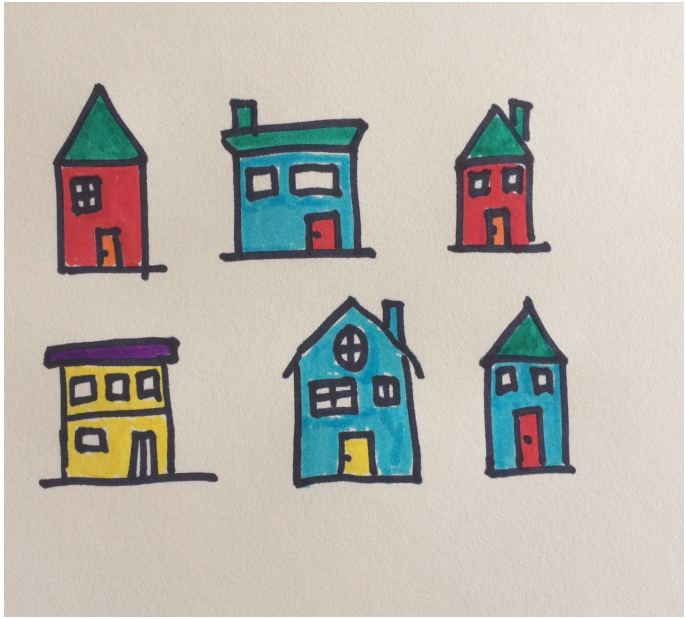
Qualitative Data

Quantity

Stories of the residents **(Phil loves his house)**

Style **(Colonial)**

Size



Average Sale Price

Color **(Red, Blue, Yellow)**

Condition **(Average)**

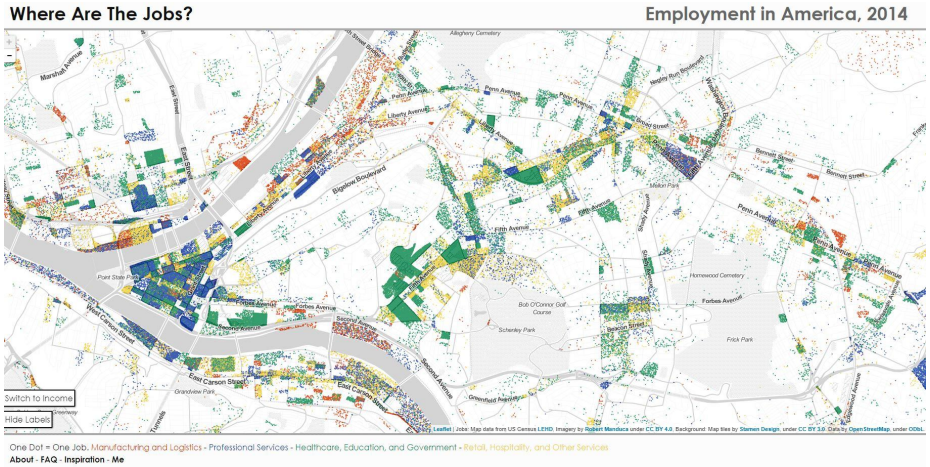
From Data 101

Primary data

Collected first-hand through original research. Ex. Surveys, interviews, or experiments

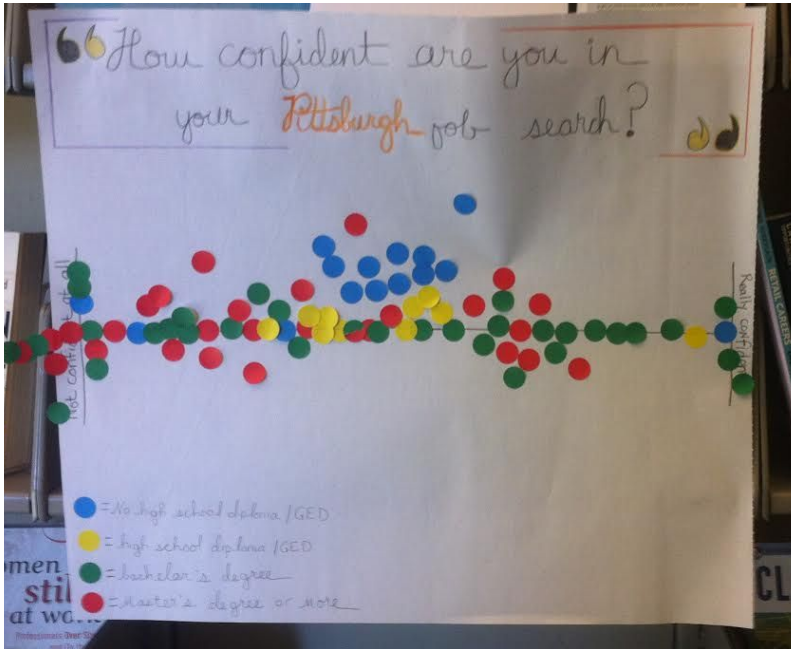
Secondary data

Data that are collected and accessed from outside sources, like the government, companies, or local institutions. Ex. The Census, information collected by government departments.



Data collection methods

existing
records/administrative
surveys
interviews
focus groups
observations
sensors/ambient
experiments
crowdsourced data



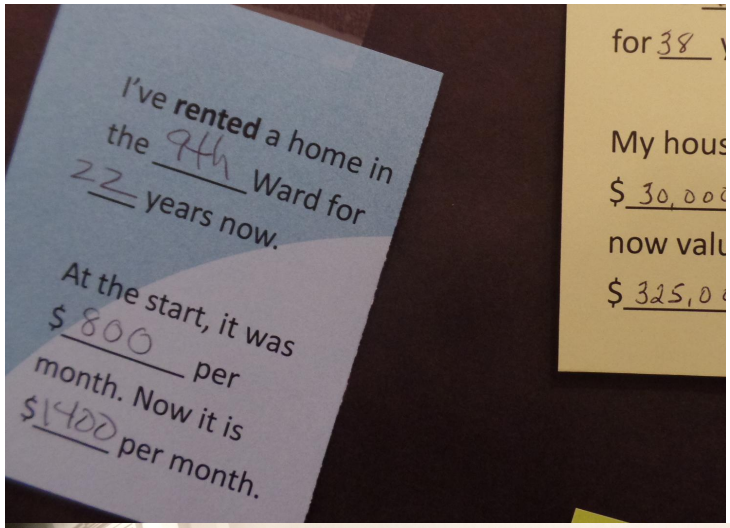
From Data 101

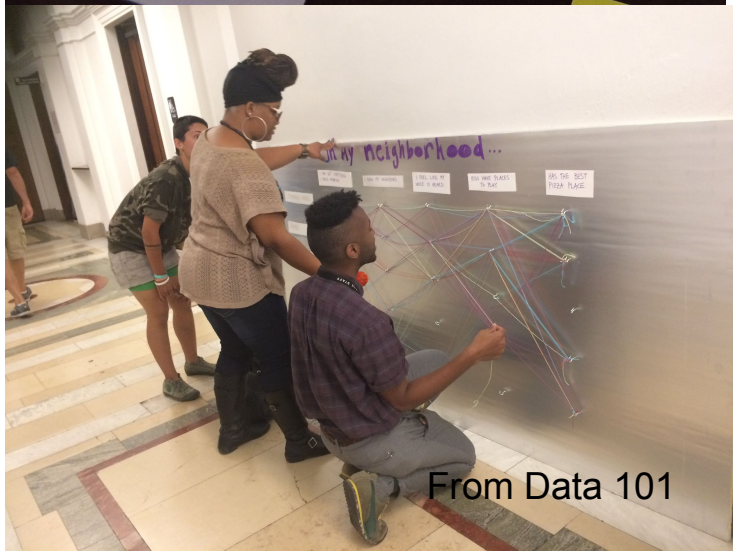
Civic Data

Data related to community life and the choices and policies that affect people and places.

Includes data published by local, state, and federal governments as well as personal data that, if collected & organized, could be used to tell alternative stories.

Data may also become civic data if used for a community purpose.





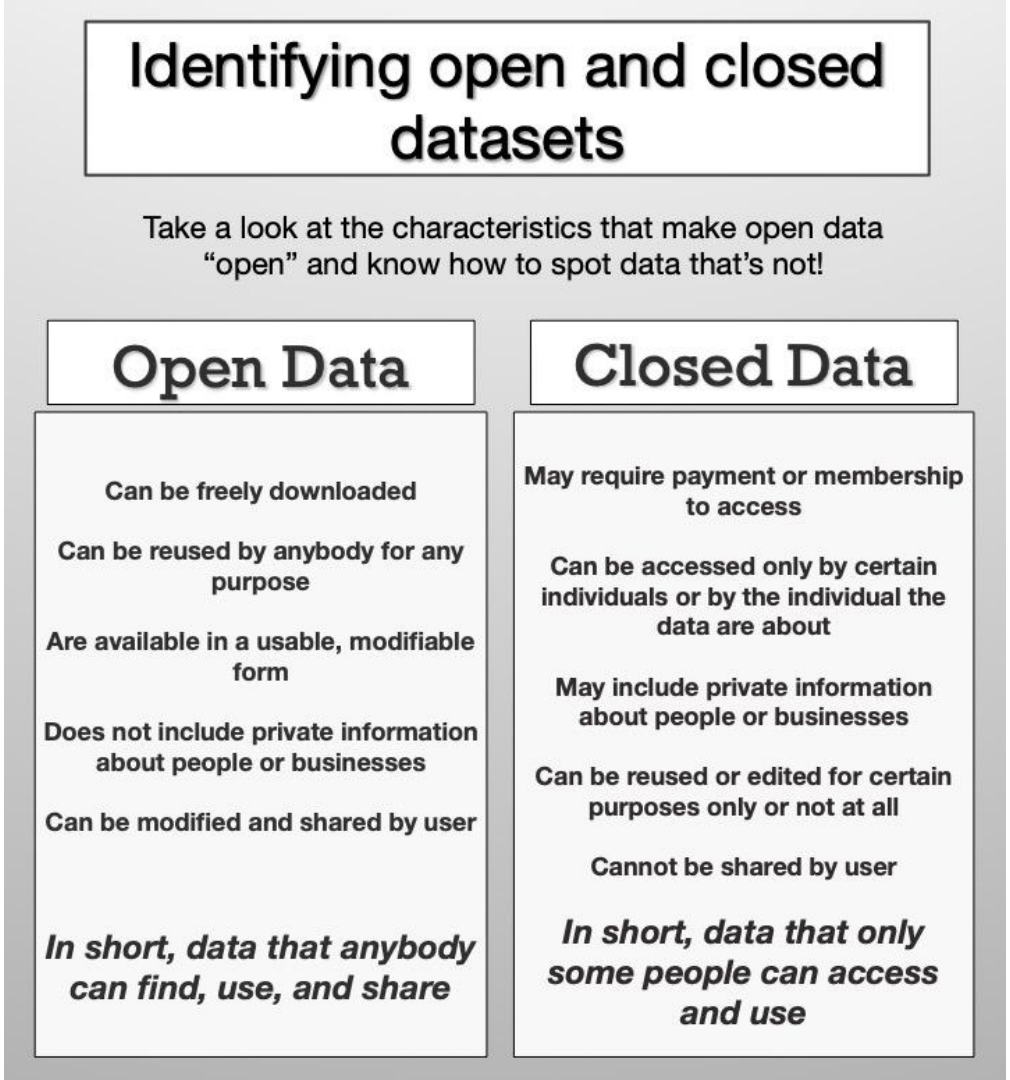
Open data

- Free to access and download
- Reusable by anybody for any purpose
- Available in machine readable, open file formats
- Absent of personally identifiable information
- Timely and as complete as possible
- Modification and resharing permitted

(A few ways) data can be used

- Asking/answering questions
- Telling stories
- Targeting outreach/services
- Spurring discussion
- Checking assumptions
- Learning more about where you live
- Informing policies
- Other?

Ex. Jewish Family and Children's Service of Pittsburgh used local data to look at income and unemployment rates in several South Hills municipalities to help determine where to focus efforts.



Activity

Can you think of any examples of open data? Closed data?