Finding and Acquiring Spatial Data

Class #7 | Geog 215

Introduction to Spatial Data Science

Spring 2020

> *If I cease searching, then, woe is me, I am lost. That is how I look at it - keep going, keep going come what may* - Vincent Van Gogh

What have we covered so far

- R introduction and basic commands
- How to execute R commands and communicate results
  - **Console : Experimentation *(Cooking)***
  - **Scripts : Storing what works *(Recipe for internal use)***
  - **RMarkdown : Communicating to others *(Cookbook)***
- Tidyverse in Action (Bushfires example)
- Representing Location and spatial data

Where is the Spatial Data at ?

How to do I Acquire / get Access to it?

How to I import it in R and work with it?

Data, Data, Data

Foundation data

- “background” spatial data used as reference
- Also referred to as “base map” data
  - For display, a key linkage among features on a map is their relative position
  - **Location relative to a “known” set of features e.g., a map of cities overlaid on a state map**
- May also be used in analysis!

Foundation Data

Vector format

Administrative boundaries/locations

States, cities, counties

Enumeration units

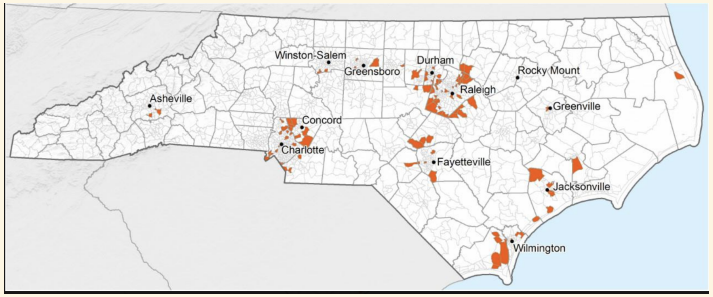
Tracts, Block Groups, Blocks

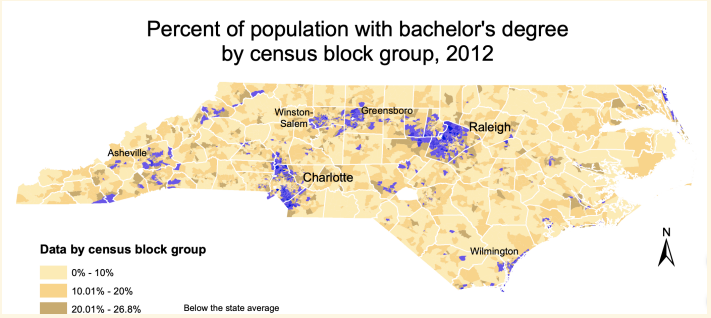
Natural (physical) features

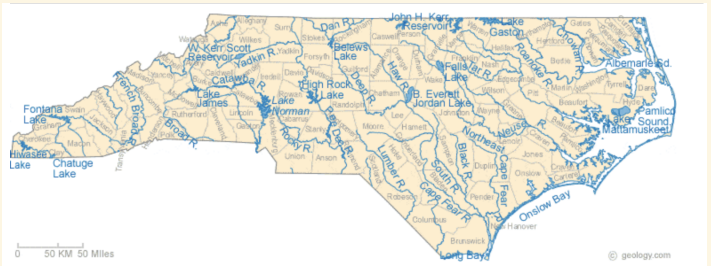
Lakes, rivers

Transportation (or other) infrastructure

Roads, trails, railways







Foundation Data

Raster (usually)

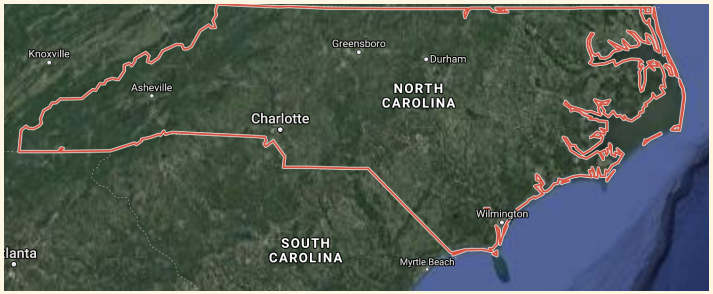
Imagery

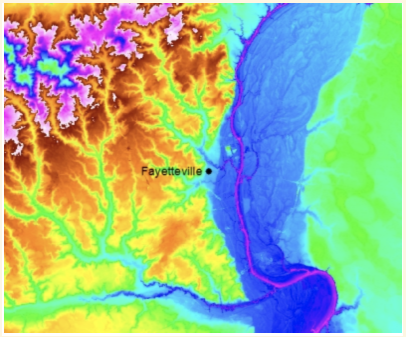
Aerial photos, satellite imagery

Elevation/terrain

Digital Elevation Model

Land cover/use





Foundation Data

Creating and/or developing spatial data requires a large amount of time, effort,and money

- ***Good news***: a large amount of foundation data has already been created and is freely available on the Internet
- ***Bad news***: the availability and quality of this data varies signifcantly from place to place

Demographic Data

Population attributes

- Counts
- Age, gender, race/ethnicity, etc.

Census/survey data

- Data resolution and availability will vary
  - **Generally, higher spatial resolution = less detailed data**

Socioeconomic Data

Population attributes

- Education, Income, Employment, etc
  - **Often, used as a proxy for health or health-related behaviors**
  - **Often, used to assign neighborhood attributes**

Census/survey data

- Data resolution and availability will vary
  - **Generally, higher spatial resolution = less detailed data**

Environ/Neighborhood Data

Neighborhood attributes

- Environmental hazards
  - **Air pollution, water pollution, etc.**
- Environmental conditions
  - **Temperature, precipitation, humidity**
- Disease transmission
  - **Vector distribution (and/or habitat)**
- Amenities
  - **Neighborhood features**
    - **Parks, produce stores, etc.**

Health-related Data

Vital statistics

- Births and deaths

Morbidity Data

- Data about the “health” or “health conditions” in the population

Health care providers

- Locations of hospitals, physicians, emergency services, clinics, etc.

Often, is non-spatial data

Data and information

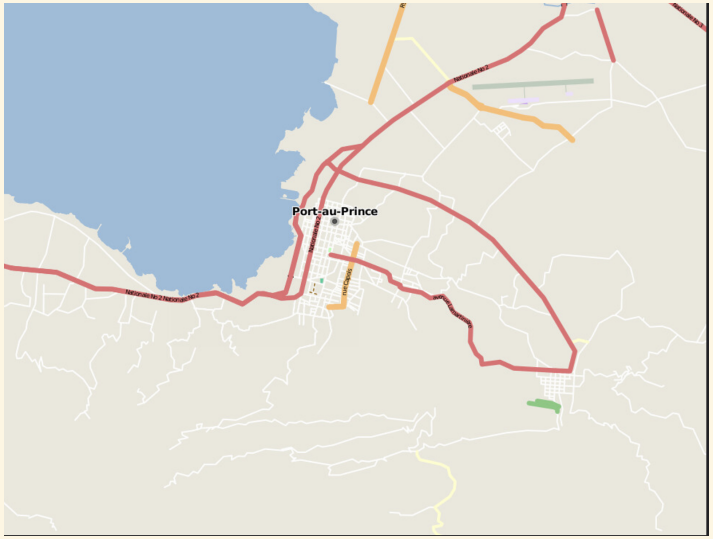
Authoritative information

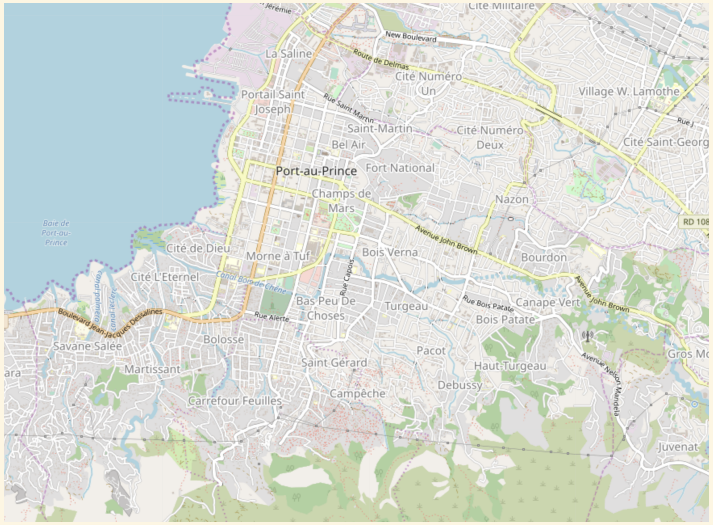
- Accuracy/documentation standards and protocols Metadata provided

Asserted information

- Accuracies are undocumented
- No metadata
- Data about popular places tend to be more accurate
- Accuracy can be higher than legacy authoritative data

For Example


Port-au-Prince


La Saline
Portail Saint Joseph
Cité Numéro Un
Village W. Lamothe
Saint-Martin
Cité Numéro Deux
Bel Air
Port-au-Prince
Fort National
Champs de Mars
Nazon
Cité de Dieu
Morne à Tuf
Bois Verna
Bourdon
Cité L'Eternel
Bas Peu De Choses
Turgeau
Canape Vert
Bois Patate
Bolosse
Savane Salee
Martissant
Saint Gérard
Pacot
Haut-Turgeau
Debussy
Carrefour Feuilles
Campêche
Juvenat

For Example

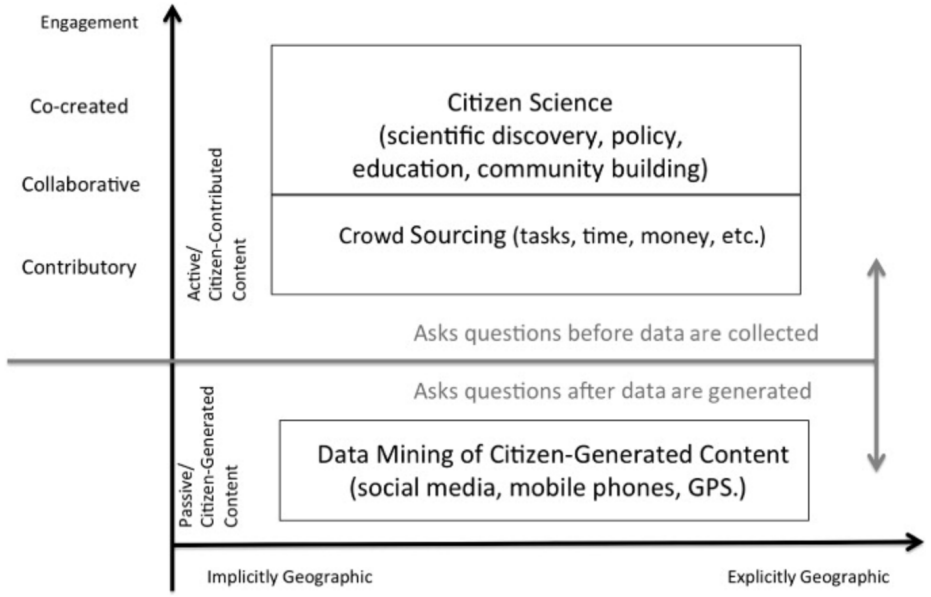

Botongmungdong (Botongmungdong)
Changgwangdong (Changgwangdong)
Ansan Street
Changgwang Street
Tongseongdong (Tongseongdong)
평양직할시 (P'yŏngyang)
해방산거리
Soosong Street
Tonghungdong (Tonghungdong)
Yonggwang
Kyogudong (Kyogudong)
Yokjondong (Yokjondong)
Tongandong (Tongandong)
Google
Cholima Street
Changgwang Street
서문거리
Sosong St
동성동
Seosong Street
Haebangsan St
봉화
Pongwa
해방산거리
봉남동
Pongnam-dong
영광
Yonggwang
동흥동
Tonghung-dong
오탄동
Otan-dong
평양역
Pyongyang
역전동
Yokjon Dong
동안동
Tongan
Dong

Other Examples

Types of Assertive Communication

Figure 1 of 1

Figure 1. A typology of citizen-generated content.



Questions to consider when choosing geospatial data

- Where is the data coming from?
- Does the data cover the geographic extent I need?
- Is the information current?
- Is the scale of the data consistent with my other data?
- How accurate is the data?
- What types of attribute data are available with the data?

Source: https://campusguides.lib.utah.edu/c.php?g=160707&p=1052129

Questions to consider when choosing geospatial data

- How large are the data files?
- Is the data properly formatted to open using geospatial software?
- Is the data consistently projected?
- What metadata accompanies the data?
- The major question: Are there any use restrictions for using or accessing the data?

Source: https://campusguides.lib.utah.edu/c.php?g=160707&p=1052129

Data Acquisition

Data access and sharing

- In the past, data was highly guarded
- Movement towards making sharing data via open availability and unrestrictive use arguments

From secondary sources

- Data format
- Modes of access

Data Acquisition

Data Format

- Consider whether you can work with data

Modes of Access

- Actual steps to acquire the data
- Examples
  - **Web GIS (Must manually point/click)**
  - **Web links (must manually point/click)**
  - **FTP site (can manually point/click or use automation)**
  - **API (request / response approach)**
  - **Web Scraping / Crawling**

Finding Data

Very important to remember...

- Data “existing” in some repository or on the internet is not always analysis-ready
    - **Often, needs substantial formatting and cleaning before it can be used in an analysis**

Finding Data

The data you want may not be available:

- For download via the internet (send an email, make a phone call, file a FIOA) Digitally (paper maps)
- At the extent or resolution you need
- As spatial data (tabular → join to spatial data)
- In a directly useable format (aerial photos / imagery → landcover)

Finding Data

Like many other tasks...

- Explore options
- Be resourceful!

Next Class

Importing Spatial Data in R using different methods

Mining and Mapping Twitter Data!

Complete Data Camp Exercises/ Lab 2

Complete Poll Everywhere Poll