

ENRICHING ANALYTICS WITH GEOGRAPHIC DATA SCIENCE

DANIEL ARRIBAS-BEL
LEVI JOHN WOLF

@darribas
@levijohnwolf
geographicdata.science/book



CONCEPTS

COMPUTATION

CONCEPTS

COMPUTATION

**NOT
EVERYONE
DOES
GEOGRAPHY**

**EVERYBODY
CAN BE A
GEOGRAPHER**

GEOGRAPHY

GEOGRAPHY IS WHAT GEOGRAPHERS DO

Apocryphal, attributed to Wilson (1970),
also earlier by Johnston & Sidaway (2015)

GEOGRAPHY IS WHAT GEOGRAPHERS DO

Apocryphal, attributed to Wilson (1970),
also earlier by Johnston & Sidaway (2015)



GEOGRAPHY

the systematic study
of relationships
between people
and their environments.

GEOGRAPHY

commuting, segregation, climate change, urban planning, campaigns, elections, and voting, housing markets, energy forecasting, operational research, conservation, community dynamics, disease spread, violent conflict, school efficacy, ecology, economic development, ice sheets, rugby digital economy, program evaluation, migration, natural hazards risk, social vulnerability, trade

GEOGRAPHERS

use the relationships
between people and
their environments
to solve problems

**NOT
EVERYONE
DOES
GEOGRAPHY**

**EVERYBODY
CAN BE A
GEOGRAPHER**

NYT: ONE NATION TRACKED

<https://nyti.ms/2uwMIUZ>

Every minute of every day, everywhere on the planet dozens of companies are logging the movements of tens of millions of people.

[...]

Without much effort we spotted visitors to the estates of Johnny Depp, Tiger Woods and Arnold Schwarzenegger, connecting the devices' owners to the residences.

NYT: ONE NATION TRACKED

<https://nyti.ms/2uwMIUZ>

UBIQUITY

Everything has a geographical location & time of occurrence

Every minute of every day, everywhere on the planet dozens of companies are logging the movements of tens of millions of people.

CONTEXT

Location enhances information that is already available

Without much effort we spotted visitors to the estates of Johnny Depp, Tiger Woods and Arnold Schwarzenegger, connecting the devices' owners to the residences.

**Everything has a geographical
location & time of occurrence**

Everything has a geographical location & time of occurrence

Every Monday without fail, a final-year undergrad knocks on my door at 17:55, right as I'm packing up to leave my office, to ask if they can meet to talk about their dissertation.

Everything has a geographical
location & time of occurrence

Event E happens

Everything has a geographical
location & time of occurrence
occurrence

Event E happens
at 17:55
every Monday
as I pack up to leave

Everything has a geographical
location & time of occurrence

occurrence
periodicity

Event E happens
at 17:55
every Monday
as I pack up to leave

Everything has a geographical
location & time of occurrence

occurrence
periodicity
causality

Event E happens
at 17:55
every Monday
as I pack up to leave

Everything has a geographical
location & time of occurrence

occurrence

periodicity

(G-)causality

Event E happens
at 17:55
every Monday
as I pack up to leave

Everything has a geographical
location & time of occurrence

occurrence
periodicity
(G-)causality

Event E happens
at 17:55
every Monday
as I pack up to leave

**TIME: MORE
THAN CLOCK
POSITION**

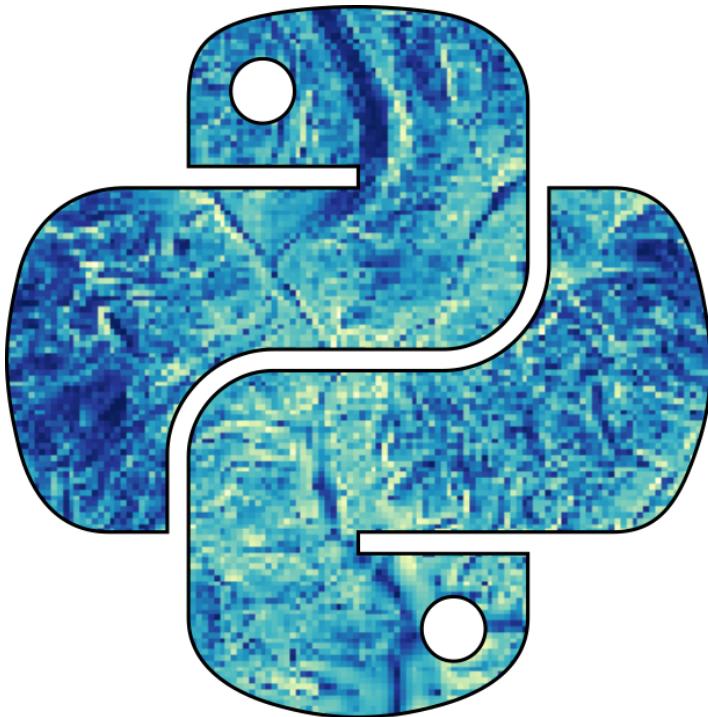
Everything has a geographical
location & time of occurrence

location
relation
contagion

SPACE: MORE THAN EARTH POSITION

Event E happens

at 51°27'N 2°35'W
in Bristol
near me

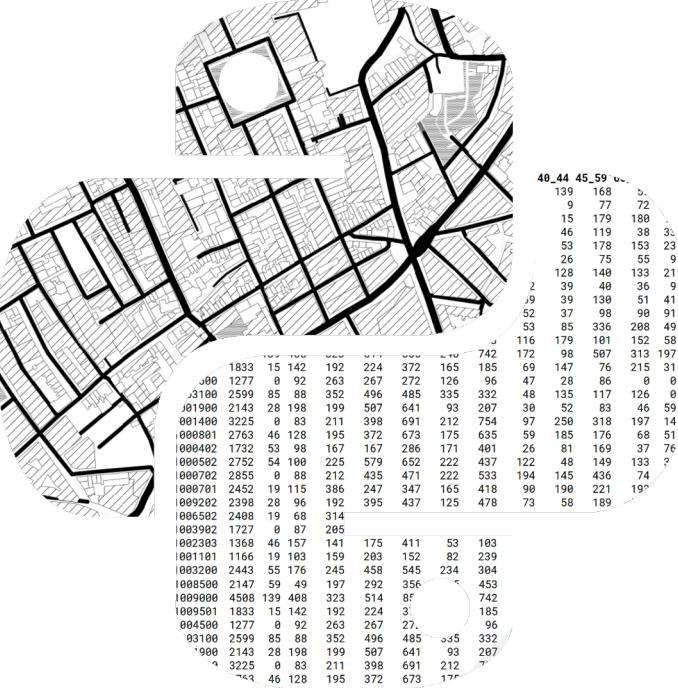


PYTHON FOR GEOGRAPHIC DATA ANALYSIS

Henrikki Tenkanen
Vuokko Heikinheimo
David Whipp

Solving Geographic Information Processing problems
with high-performance Python. CRC Press, 202X.

pythongis.org



GEOGRAPHIC DATA SCIENCE

Sergio Rey
Daniel Arribas-Bel
Levi John Wolf

Methods & models that do (un)supervised learning with
Python using PySAL & Scikit-Learn. CRC Press, 2022.

geographicdata.science/book

GEOGRAPHIC DATA SCIENCE

COMMODITY:

Do standard analysis on your standard data
& chuck the results into Tableau/Alteryx to map

GEOGRAPHIC DATA SCIENCE

COMMODITY:

Do standard analysis on your standard data
& chuck the results into Tableau/Alteryx to map

ENRICHED:

Do standard analysis on data after you have
augmented it using *spatial feature engineering*

EXTENDED:

Use methods that explicitly learn from, analyze,
or summarize the geographical structure in your data

<https://geographicdata.science/book>

COMMODITY:

Do standard analysis on your standard data
& chuck the results into Tableau/Alteryx to map

ENRICHED:

Do standard analysis on data after you have
augmented it using *spatial feature engineering*

EXTENDED:

Use methods that explicitly learn from, analyze,
or summarize the geographical structure in your data

<https://geographicdata.science/book>

COMMODITY:

Do standard analysis on your standard data
& chuck the results into Tableau/Alteryx to map

ENRICHED:

Do standard analysis on data after you have
augmented it using *spatial feature engineering*

EXTENDED:

Use methods that explicitly learn from, analyze,
or summarize the geographical structure in your data

<https://geographicdata.science/book>

ENRICHED:

Do standard analysis on data after you have augmented it using *spatial feature engineering*

<https://geographicdata.science/book>

ENRICHED:

Do standard analysis on data after you have augmented it using *spatial feature engineering*

synthesizing information using spatial relationships within or across data.

<https://geographicdata.science/book>

ENRICHED:

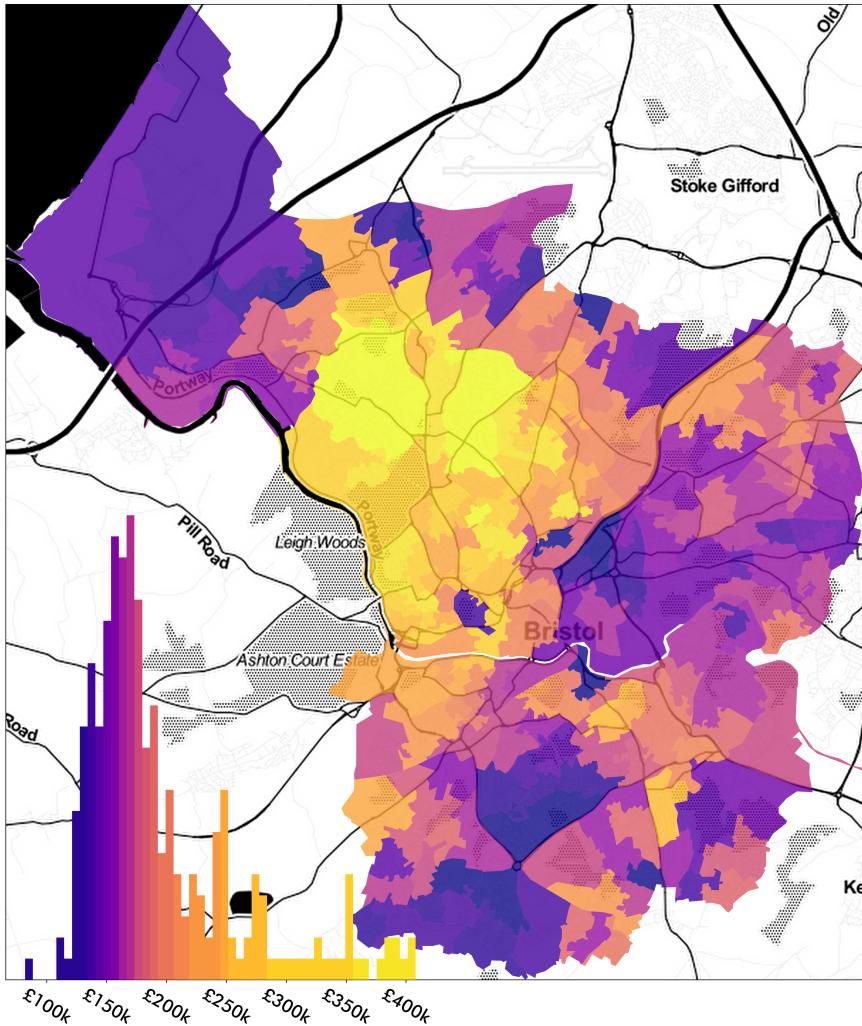
Do standard analysis on data after you have augmented it using *spatial feature engineering*

synthesizing information using spatial relationships within or across data.

connecting datasets together using *map matching*

expressing latent structure using *map synthesis*

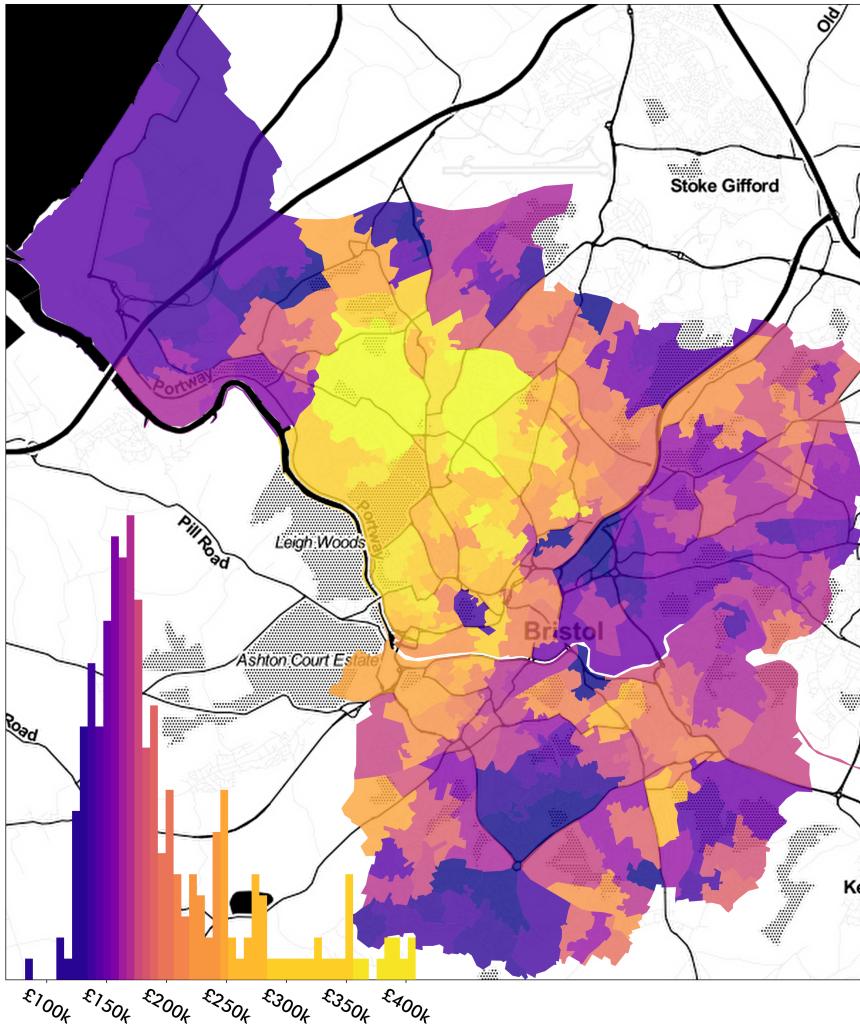
<https://geographicdata.science/book>



Say you're modelling house prices

How good are the schools in the area?
How far is my nearest park?
Is the architecture in the area pretty?
Are there good views?
What's the air quality like?

<https://geographicdata.science/book>



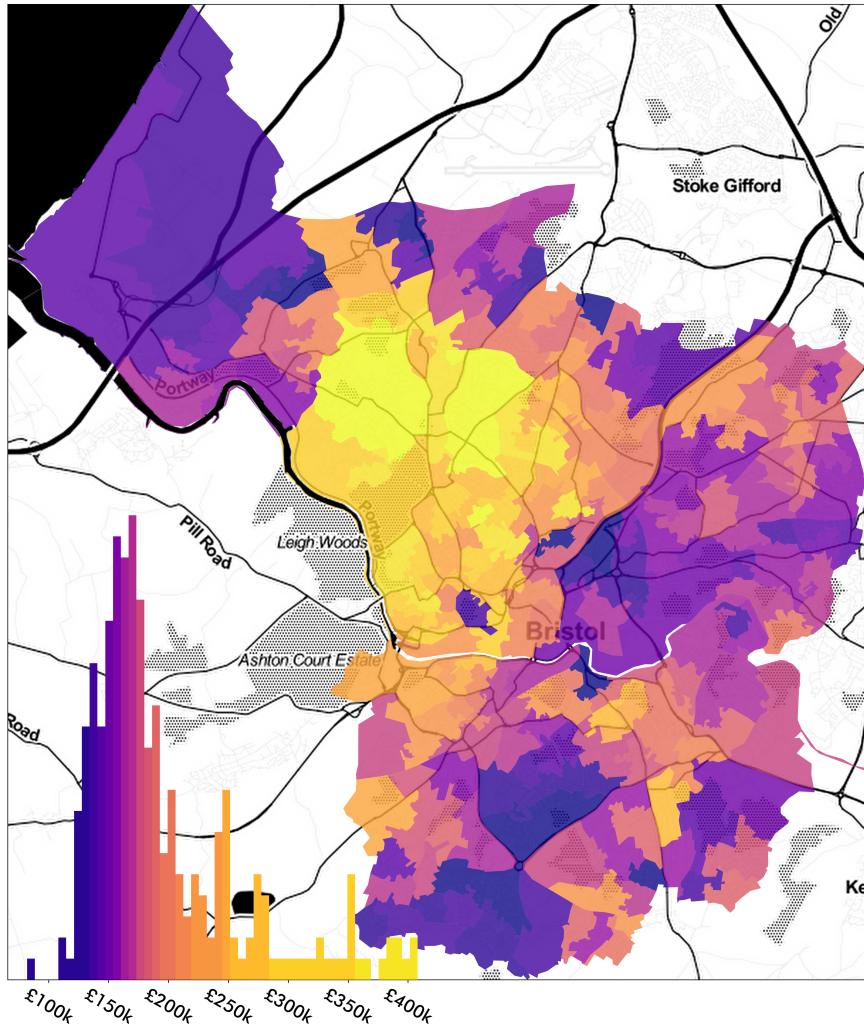
Say you're modelling house prices

How good are the schools in the area?
How far is my nearest park?
Is the architecture in the area pretty?
Are there good views?
What's the air quality like?

MAP MATCHING

Using spatial relationships to transfer information to your target data

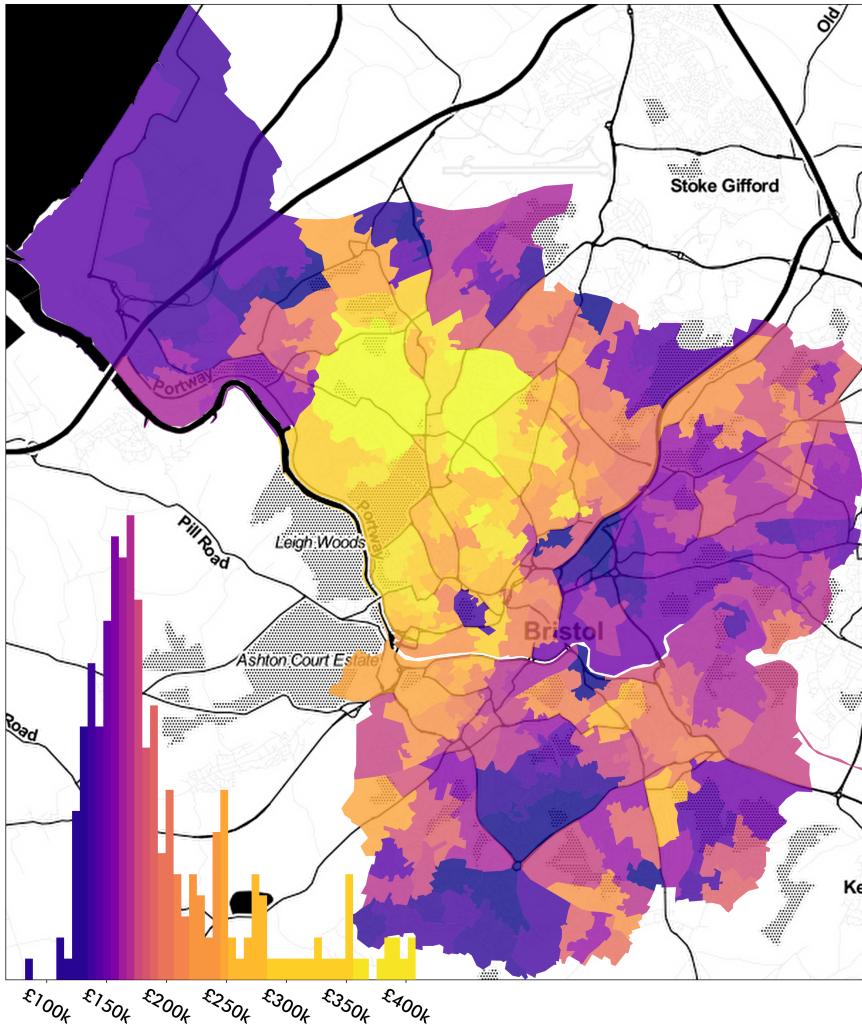
<https://geographicdata.science/book>



Say you're modelling house prices

How much are nearby houses?
How many bedrooms are typical here?
Is it larger than nearby alternatives?
Does any house near have a garden too?

<https://geographicdata.science/book>

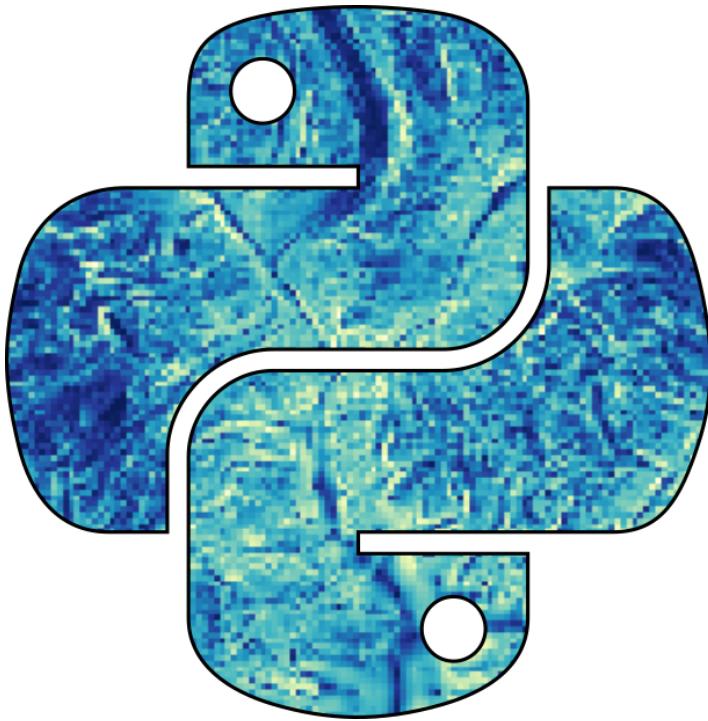


Say you're modelling house prices

How much are nearby houses?
How many bedrooms are typical here?
Is it larger than nearby alternatives?
Does any house near have a garden too?

MAP SYNTHESIS

Using spatial relationships between observations to synthesize new features

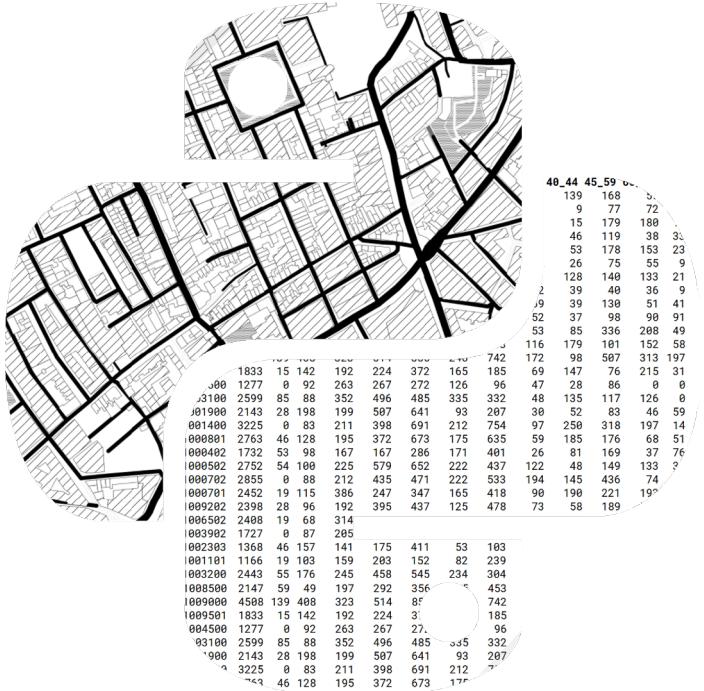


PYTHON FOR GEOGRAPHIC DATA ANALYSIS

Henrikki Tenkanen
Vuokko Heikinheimo
David Whipp

Solving Geographic Information Processing problems
with high-performance Python. CRC Press, 202X.

pythongis.org

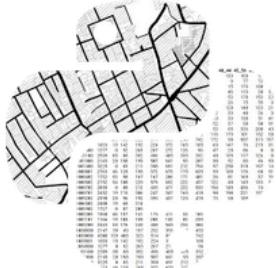


GEOGRAPHIC DATA SCIENCE

Sergio Rey
Daniel Arribas-Bel
Levi John Wolf

Methods & models that do (un)supervised learning with
Python using PySAL & Scikit-Learn. CRC Press, 2022.

geographicdata.science/book



Geographic Data Science with Python

Search this book...



Spatial Feature Engineering

In machine learning and data science, we are often equipped with *tons* of data. Indeed, given the constellation of packages to query data services, free and open source data sets, and the rapid and persistent collection of geographical data, there is simply too much data to even represent coherently in a single, tidy fashion. However, we often need to be able to construct useful *features* from this rich and deep sea of data.

Contents

What is spatial feature engineering?

Feature

Engineering Using Map Matching

Feature

Engineering using Map Synthesis

Conclusion

Questions