



# Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets

**Fernando Chirigati**

Work in collaboration with Harish Doraiswamy,  
Theodoros Damoulas, and Juliana Freire

*SIGMOD 2016*



# Data Exhaust from Cities

*Infrastructure*



*Environment*



*People*

flickr

twitter

**Opportunity:** make cities more efficient and sustainable,  
and improve the lives of citizens

# While understanding NYC...

1. Would a reduction in traffic speed reduce the number of accidents?
2. Why it is so hard to find a taxi when it is raining?

**DAILY Intelligencer**

## Why You Can't Get a Taxi When It's Raining

By Annie Lowrey  Follow @AnnieLowrey



<http://nymag.com/daily/intelligencer/2014/11/why-you-cant-get-a-taxi-when-its-raining.html>

Good luck, lady. Photo: Jacobs Stock Photography/Getty Images

It's pouring rain. You're running late. You desperately want to take a cab to the office. But, of course, there are none to be found. Happens all the time, right? Right, says science — or, to be specific, a new and exhaustive economic analysis of New York City taxi rides and Central Park meteorological data.

# Urban Data Interactions

Uncovering **relationships** between data sets helps us better understand cities!

Uncovering relationships



Uncovering important attributes

*Urban Data Sets are very **Polygamous!***

# Data is available...

... but it's too much work!  
**Big** urban data!



1,200 data sets  
(and counting)

> 300 data sets  
are **spatio-temporal**

8 attributes  
per data set



> 200 attributes

Where to start?  
Which data sets / attributes to analyze?  
Which spatio-temporal slice to analyze?

# Goal: Relationship Queries

*Find all data sets **related** to a given data set D*

Guide users in the data exploration process

Help identify connections amongst disparate data

Identify important variables



Q: Would a reduction in traffic speed reduce the number of accidents?

Find all relationships between Collisions and Traffic Speed data sets

# Goal: Relationship Queries

*Find all data sets **related** to a given data set D*

Guide users in the data exploration

Help identify commonalities between data sets

Identify interesting relationships

**The Data Polygamy Framework**

... efficiently!

**≈2.4 million** possible relationships among NYC Open Data alone for a single spatio-temporal resolution

# (Some) Interesting Relationships

Would a reduction in traffic speed reduce the number of accidents?

*Collisions      Tra*



X

DAILY  
Intelligencer

Things to Know About NYC's New 25-Miles-  
Per-Hour Speed Limit

By Caroline Bankoff [Follow @teamcaroline](#)

<http://nymag.com/daily/intelligencer/2014/11/things-to-know-about-nycs-new-speed-limit.html>



181063216 Photo: Getty Images

Why it is so hard to find a taxi wh

*Come talk to*

Last week, Mayor de Blasio [signed a law](#) lowering New York City's 30-miles-per-hour speed limit to 25. The change is the centerpiece of de Blasio's [Vision Zero](#) plan to drastically reduce New York City traffic deaths,

*"Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets"*, F. Chirigati, H. Doraiswamy, T. Damoulas, and J. Freire. In Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (**SIGMOD**), pp. 1011-1025, 2016

Code, data, and experiments available at:

<https://github.com/ViDA-NYU/data-polygamy>

**Acknowledgments:** National Science Foundation, Moore-Sloan Data Science Environment at NYU, DARPA, and Alan Turing Institute

Thank You!