

Spatial Analytics

Esteban Lopez Ochoa
Master in Business Analytics

30/03/2020

- Substantive
 - From Adam Smith invisible hand to social networks
 - Individual vs. socio-spatial interaction
 - Peer effects, contagion, imitation, trends
 - Spatial externalities
 - Costs/Benefits from activities that impact in a different location
 - Spatial spillovers:
 - Good: neighbor playing the piano
 - Bad: neighbor with COVID-19 and not quarantining
 - Spatial multipliers
 - Investment in a park → housing prices
 - Spatial Mismatch/Disparities
 - Spatial Context

Motivation

- Practical
 - Data: geo-located observations
 - Private sources
 - Public sources
 - Self reported vs. Web scraped
 - Spatial mismatch between data and social processes
 - Labor markets vs. Cities and Counties → Labor Market areas
 - Epidemic spread vs. Data collection in hospitals
 - Neighborhood Effects
 - Spatial interpolation
 - Change of support problem
 - Data at different spatial levels that don't overlap
 - Ej: Parents' income aggregated at the school level used to predict municipalities' income level

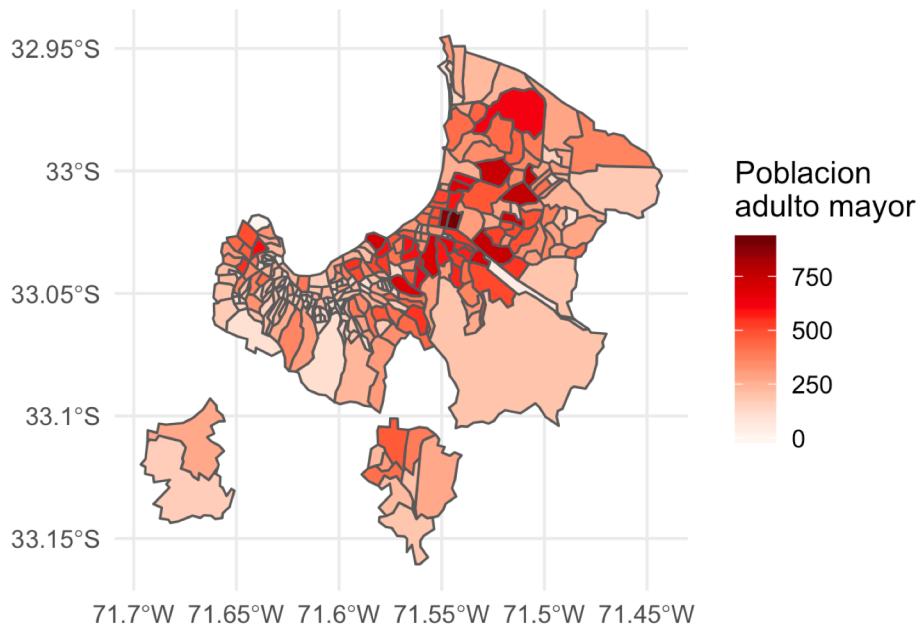
What is Spatial Analytics?

- More than just mapping
 - Added value in the explanation
 - Combination of methods, theory, data manipulation
 - Knowledge discovery
 - “from data, to information, to knowledge, to wisdom”
- Correlation is not causation, and this applies also to spatial analysis.

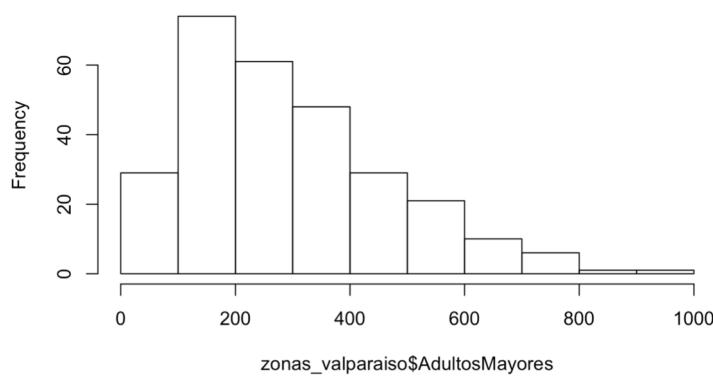
Questions of Spatial Analytics

- Where do things happen
 - Patterns, clusters, hot spots, disparities,..
- Why do things happen
 - Location decisions
- How, things that happen, affect other things (spillovers) and how context affect what happens (interaction)
- Where should things be happening/be located
 - Optimization

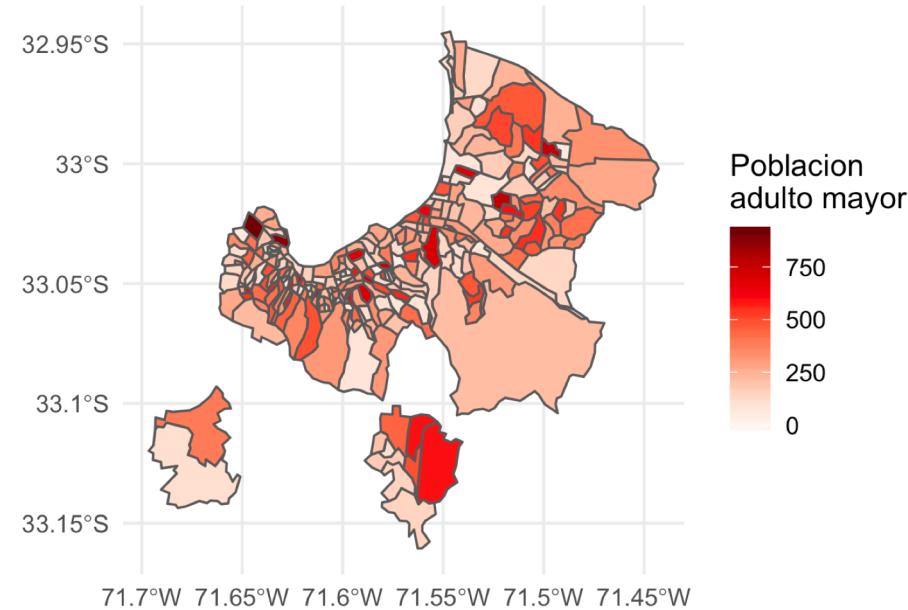
Poblacion de 65 años y más
Valparaíso y Viña del Mar



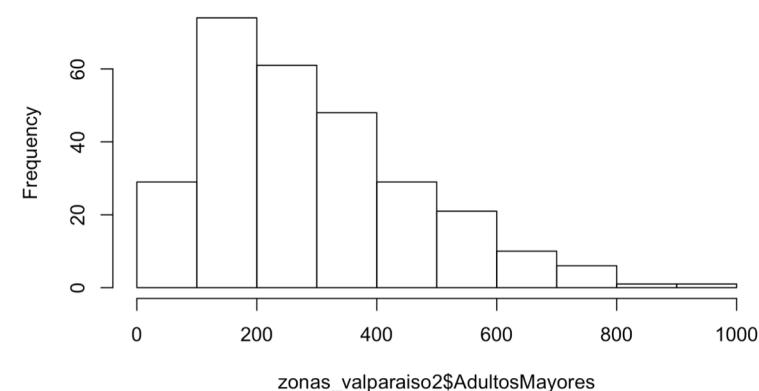
Histograma Adultos Mayores Viña-Valpo



Poblacion de 65 años y más
Valparaíso y Viña del Mar



Histograma Adultos Mayores Viña-Valpo



“Spatial” Big Data

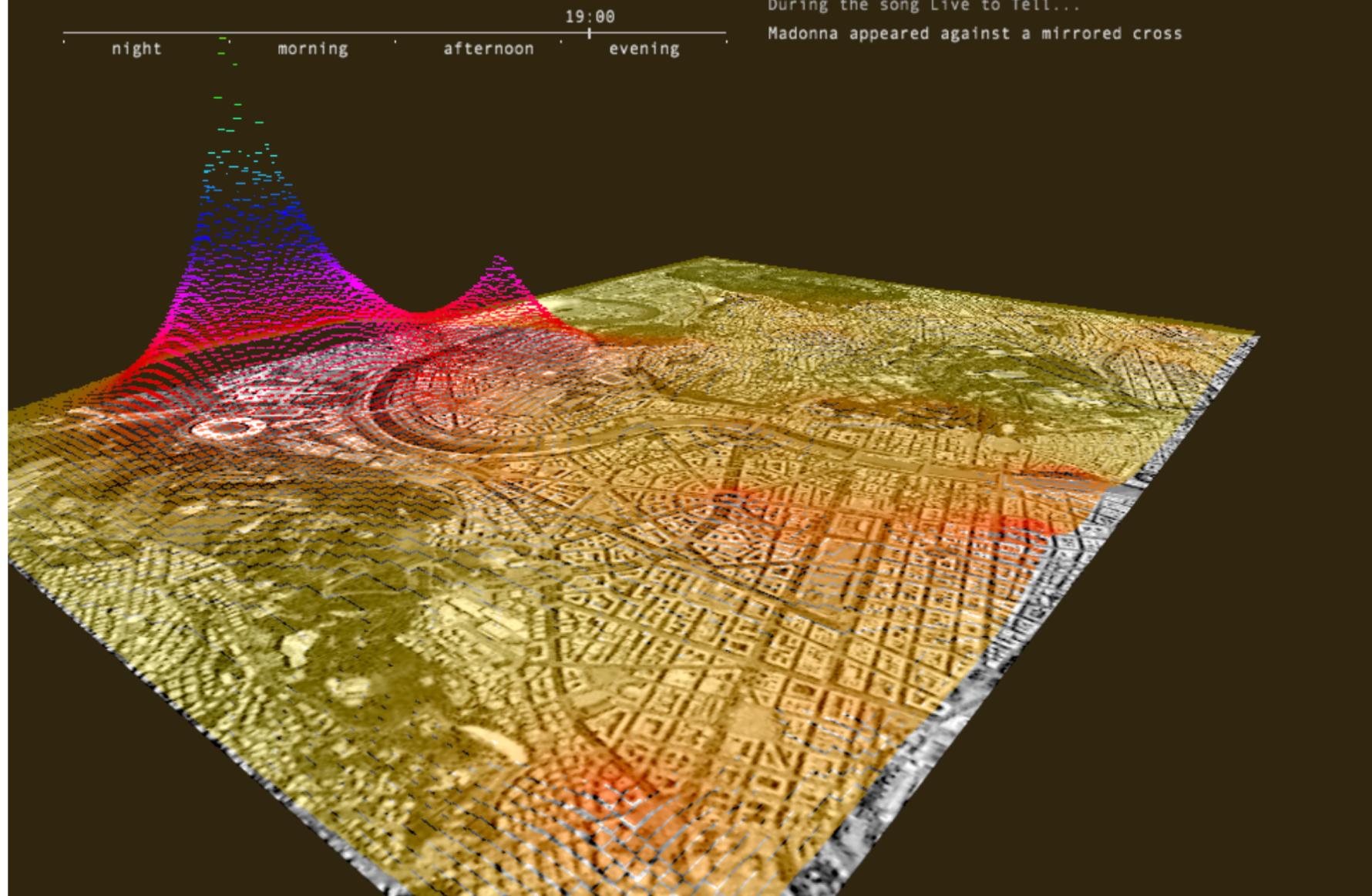
- The five ‘V’
 - Velocity
 - Veracity
 - Volume
 - Variety
 - Value
- Better description
 - Anything that presents a challenge to your normal methods and hardware for data processing and analytics
- Big data for social scientist is toy data for astrophysicist
 - Terabytes vs. Petabytes vs. Hellabytes

Big Data Issues

- Sample sizes
 - Sample vs. Population
 - Does sample size compensate for imprecision?
- Correlation is not causation
- Prediction rather than explanation
- Much newer sources of data with geolocation
 - Client data bases
 - City Sensors
 - Cell phone data

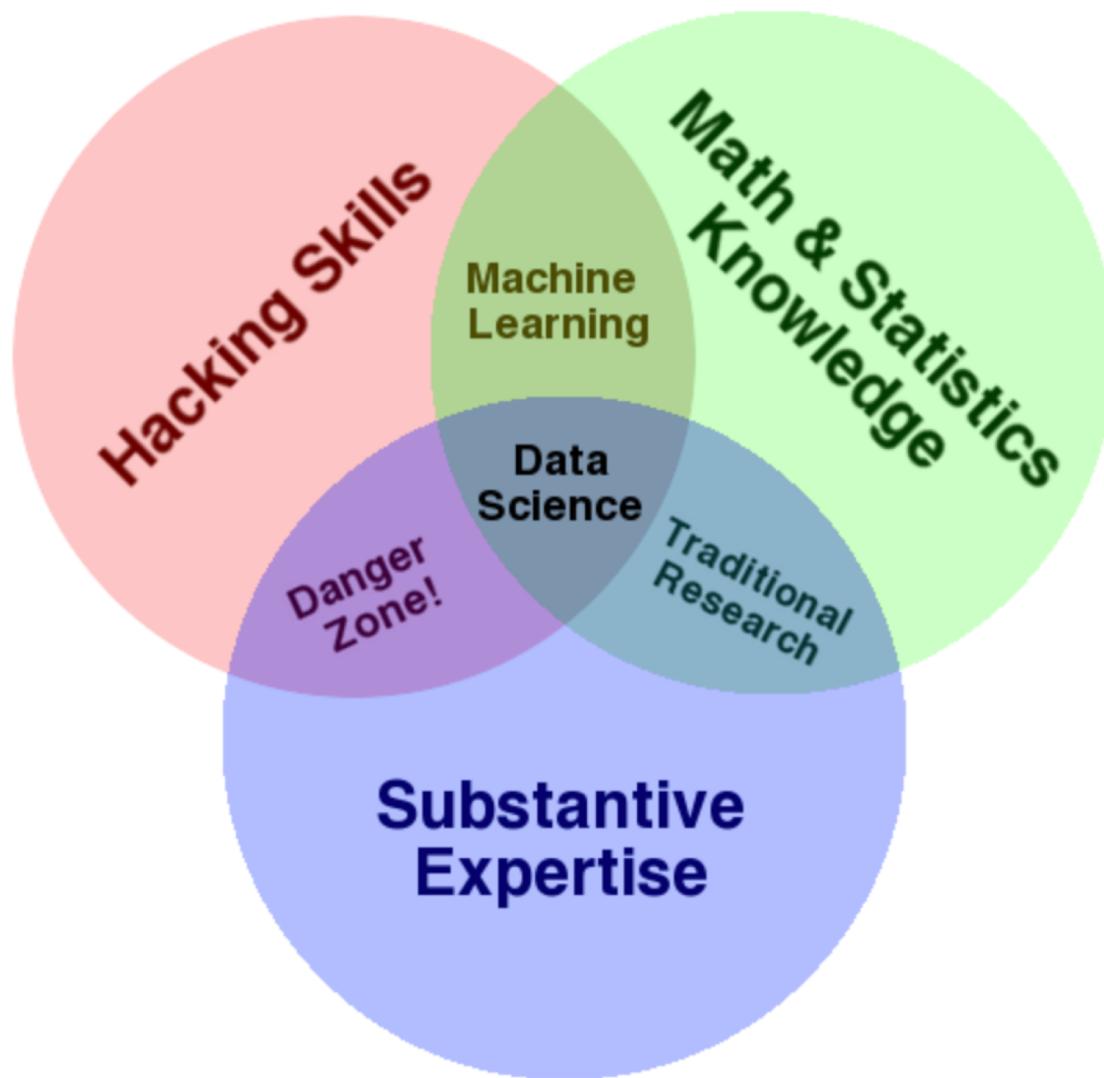
**Madonna Concert
Cellphone activity in Stadio Olimpico Rome
2006-08-06**

But receives rave reviews from 70,000 in attendance
At Rome's Olympic Stadium
Located about three kilometres from the Vatican
During the song Live to Tell...
Madonna appeared against a mirrored cross

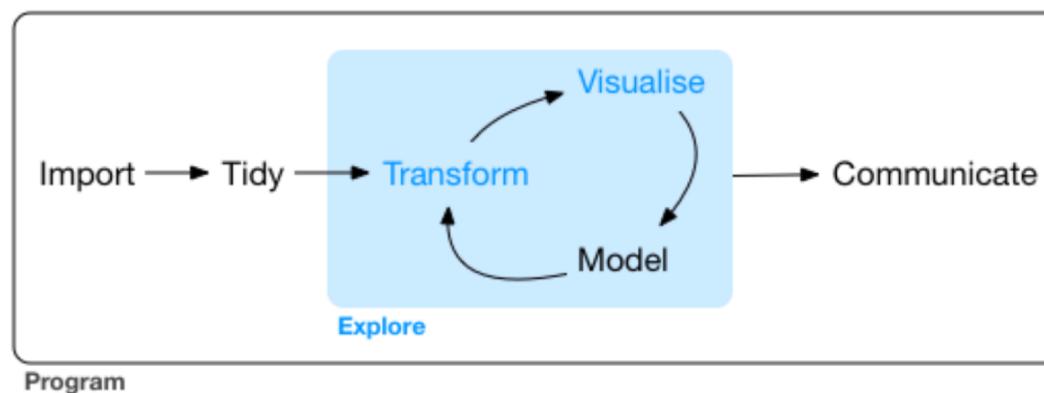


- Computation as the third paradigm to scientific discovery
 - New techniques
 - Simulation
 - Machine learning
 - Data mining
 - Visual data exploration
- Data Driven Science
 - New wave of data challenges the use of theory
 - ‘[theory is dead](#)’ - Noah Smith
 - “Economic Theory Is Dead. Here’s What Will Replace It.”
 - “Why data-driven science is more than just a buzzword”
 - Industry driven research
 - Web marketing, recommending systems, customer profiling...

Data Science



- Explicit treatment of spatial aspects
- Integration of geo-computation, spatial statistics, spatial econometrics, ESDA, spatial optimization, etc.
- 80% is data preparation (Dasu & Jhonson 2003)
 - Algorithms, data structure, workflow



Important questions to ask

- Data source
 - Sampling vs. population?
 - any (spatial) selection bias?
- Spatial units
 - Discrete vs. continuous
 - Is the problem or research question (or variable of interest) measured in the same scale I want to make inference?
 - Locations are given vs. random
 - Crime data – ‘random’ point locations
 - Purchase data – ‘given’ at store locations
- Ecological fallacy
 - Drawing conclusions about individual behavior from aggregate data analysis
 - Multi-level modelling