

Could data from Location-Based Social Networks be used to support urban planning?

Rodrigo Smarzaro, Tiago Lima, Clodoveu Davis Jr.


Federal University of Minas Gerais (Brazil)

7th International Workshop on Location and the Web

26th International World Wide Web Conference - Perth, Australia

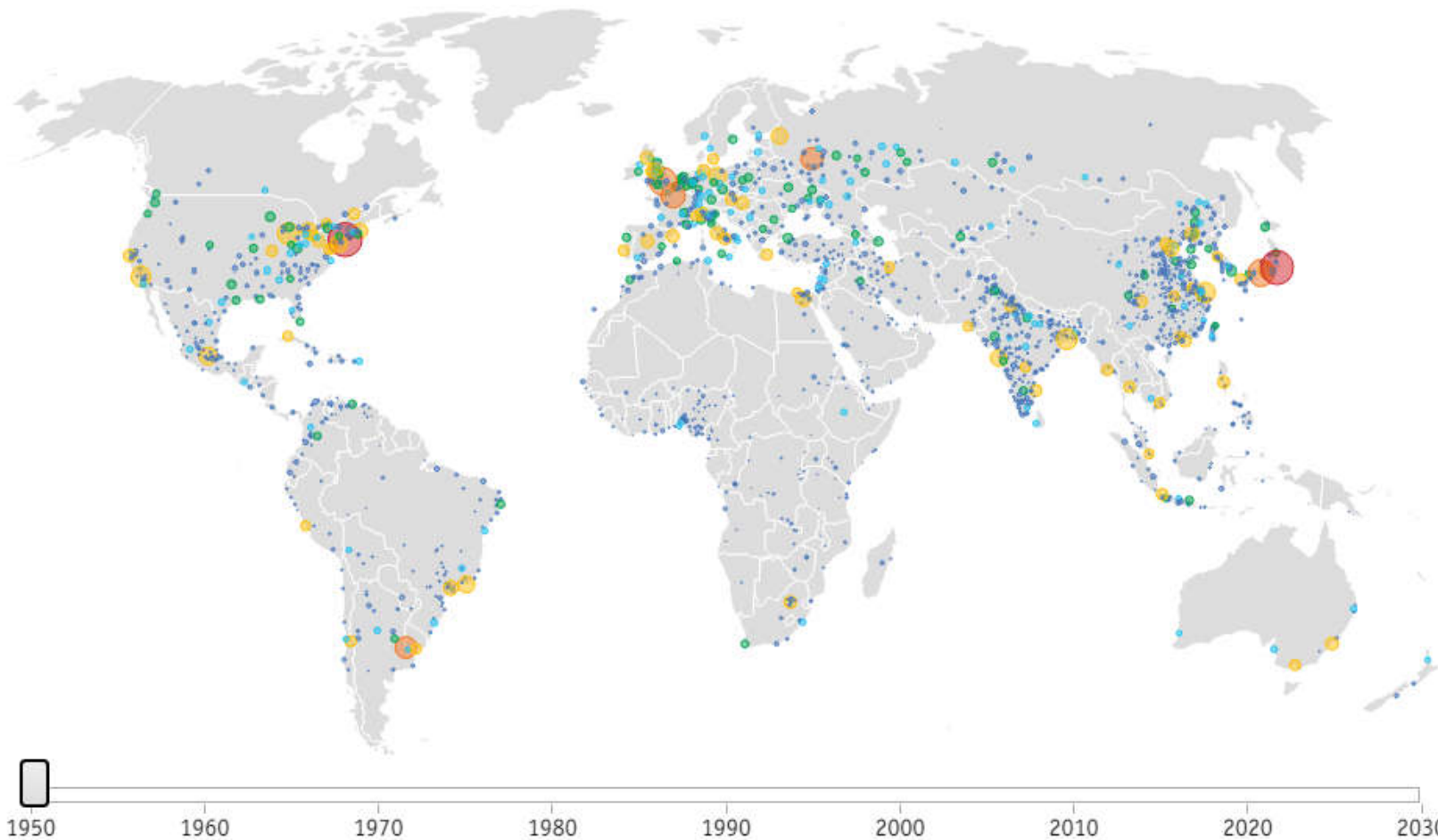


Introduction

Urbanisation, 1950 

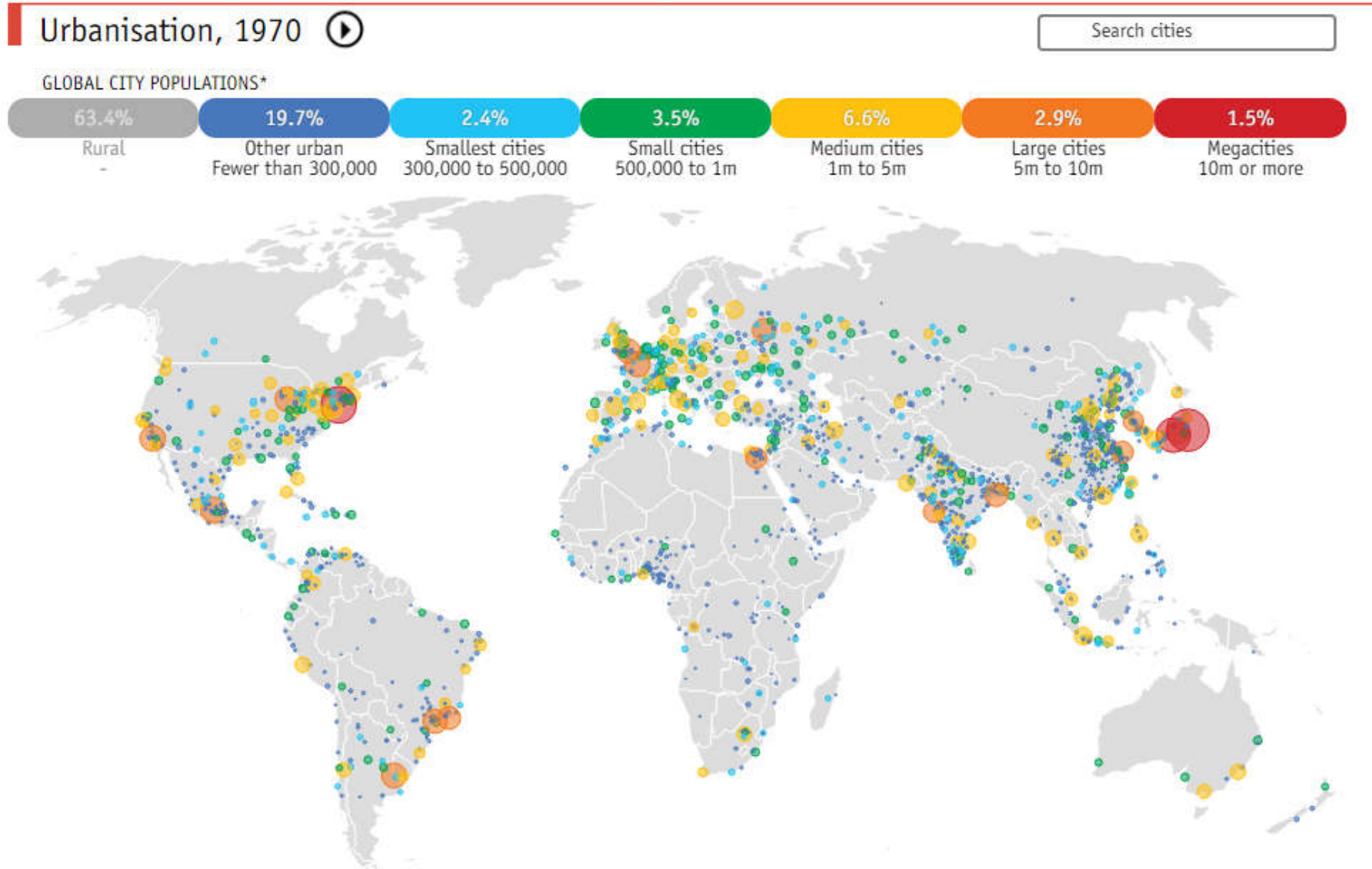
Search cities

GLOBAL CITY POPULATIONS*

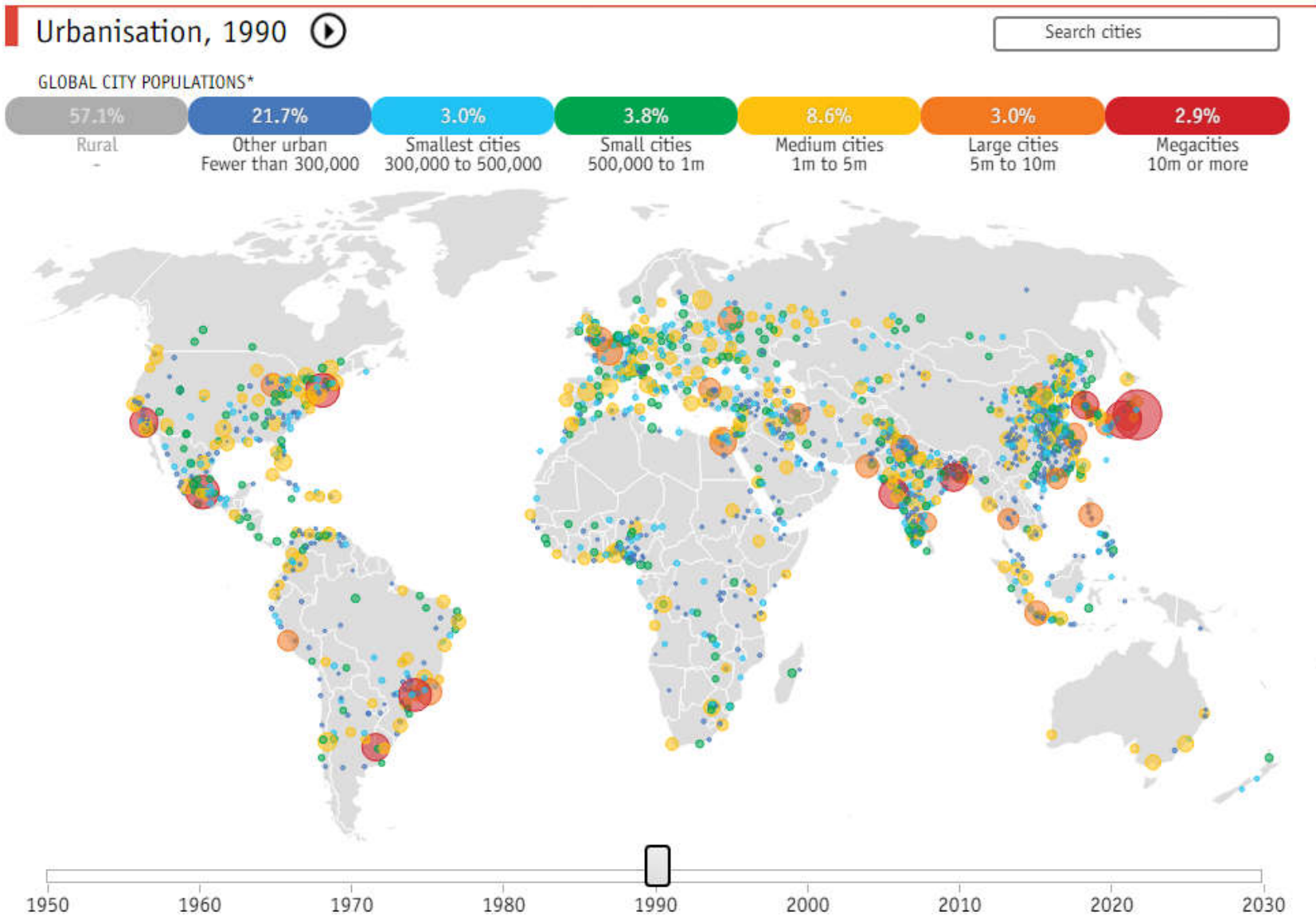


<http://www.economist.com/node/21642053>

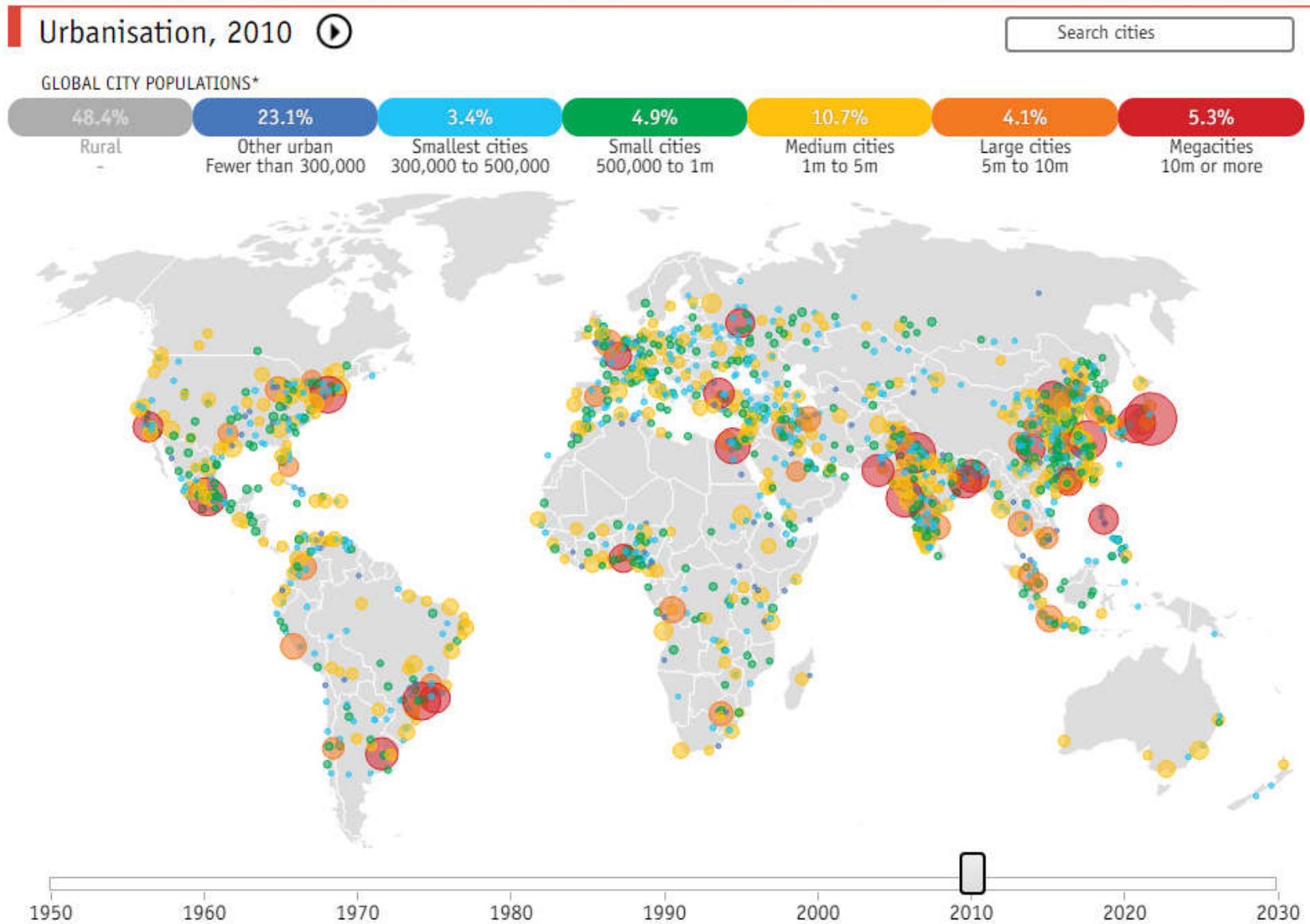
Introduction



Introduction



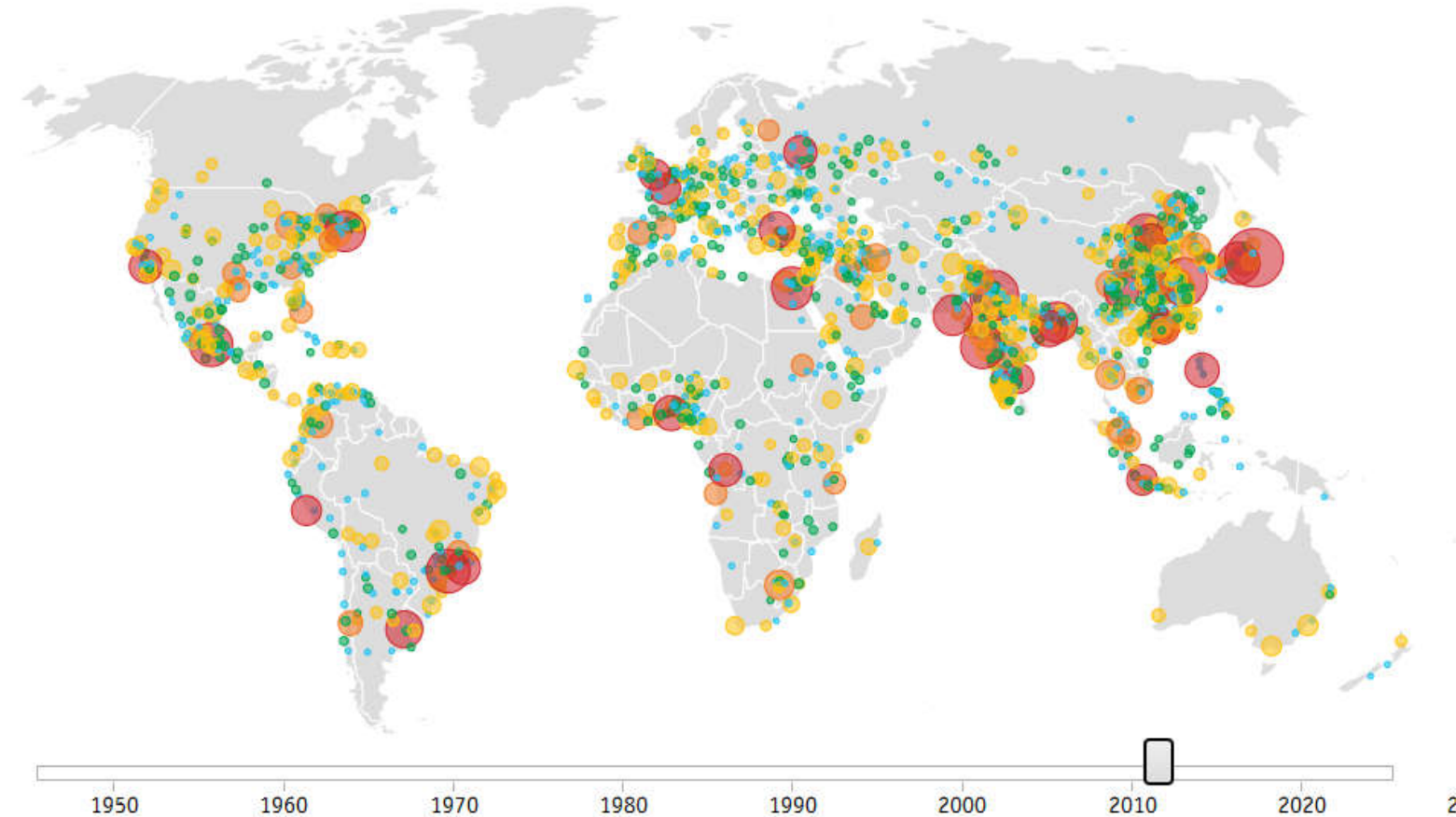
Introduction



Introduction

Urbanisation, 2016

GLOBAL CITY POPULATIONS*



Source: UN

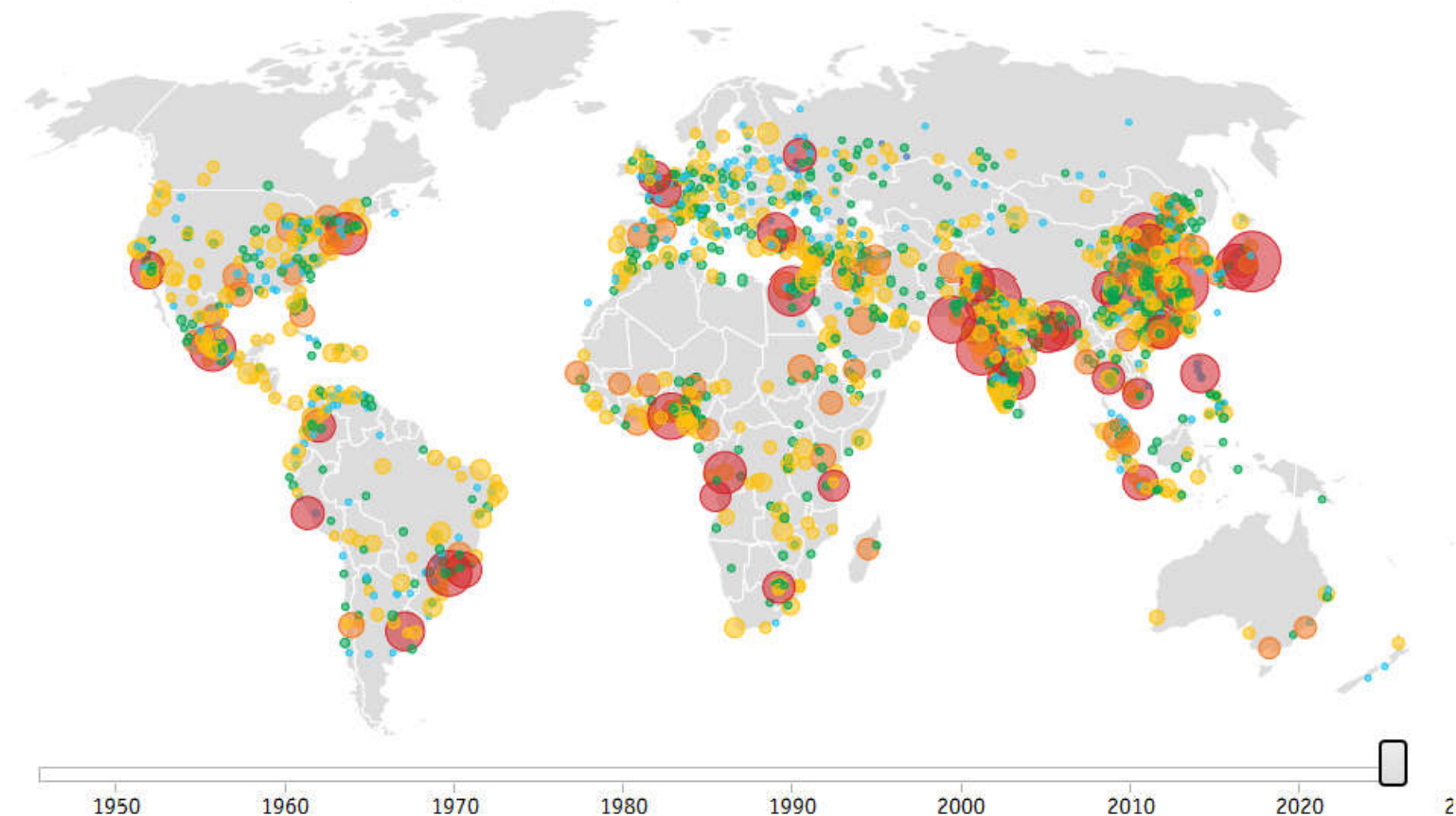
*Dataset comprises urban agglomerations with 300,000 inhabitants or more in 2014.
Data are for countries existing in 2014, mapped on modern borders. Projections from 2014.

<http://www.economist.com/node/21642053>

Introduction

Urbanisation, 2030

GLOBAL CITY POPULATIONS*



Source: UN

*Dataset comprises urban agglomerations with 300,000 inhabitants or more in 2014.
Data are for countries existing in 2014, mapped on modern borders. Projections from 2014.

<http://www.economist.com/node/21642053>

Introduction



- Most of world population lives in cities
 - many challenges arise from this growing urbanization
- Local governments require information to support urban planning
- Indicators helps to measure and monitor performance
 - e..g. availability of services, quality of urban life
- Many efforts aiming to develop indicators to evaluate the performance of countries, regions and cities
 - city as a black box



Introduction

- Quality of Urban Life Index (IQVU)
 - created to support urban planning in Belo Horizonte, a 2.5-million-people city
 - aims to spatially quantify inequalities in the accessibility to services
 - calculated for city subdivisions
- Problem: it lacks of regularity in updating index (1994, 2000, 2006, 2010 and 2012)
 - limits its potential use as a tool to support urban planning
 - possible causes: differences in temporal granularity among the sources of information, difficulties in obtaining data, methodological changes in data generation, political reasons
- We argue that data from location-based social networks (LBSN) can be used to calculate metrics / indexes to supporting urban planning
- Case study: we used data from LBSNs to estimate the Local Availability Index (IOL) – an IQVU component

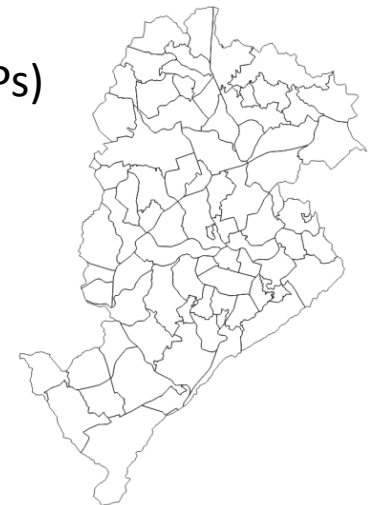
Metrics, Indicators and LBSN Data

- Many studies using alternative data sources to investigate urban problems
- Venerandi et al. (2015) presented a methodology to measure urban deprivation from user-generated content
 - Foursquare and OpenStreetMap
- Quercia and Saez (2014) investigated social media as an alternative data source to verify the relationship between socioeconomic deprivation and the presence of specific economic activities
 - Foursquare
- Shelton et al. (2015) investigated issues like segregation between neighbourhoods, mobility and inequality within the city
 - Twitter (geotagged tweets)
- De Nadai et al. (2016) studied the relationship between urban vitality and diversity
 - Foursquare, OpenStreetMaps

Quality of Urban Life and Local Availability Indexes



- Quality of Urban Life Index (IQVU)
- Aims to quantify spatial inequality of services available and accessible to the population
 - be a tool to support urban planning and resources allocation
 - should be easily and periodically updated (1994, 2000, 2006, 2010, 2012)
- A multidimensional index created by a multidisciplinary team
 - ten dimensions (variables): food supply, culture, education, sports, habitation, urban infrastructure, environment, health, urban services and public security
- Created and applied in Belo Horizonte city (Brazil)
 - uses a political subdivision in regions called Units of Planning (UPs)
- Local Availability Index (IOL) is a component of IQVU
 - measures the availability of services inside a geographic region



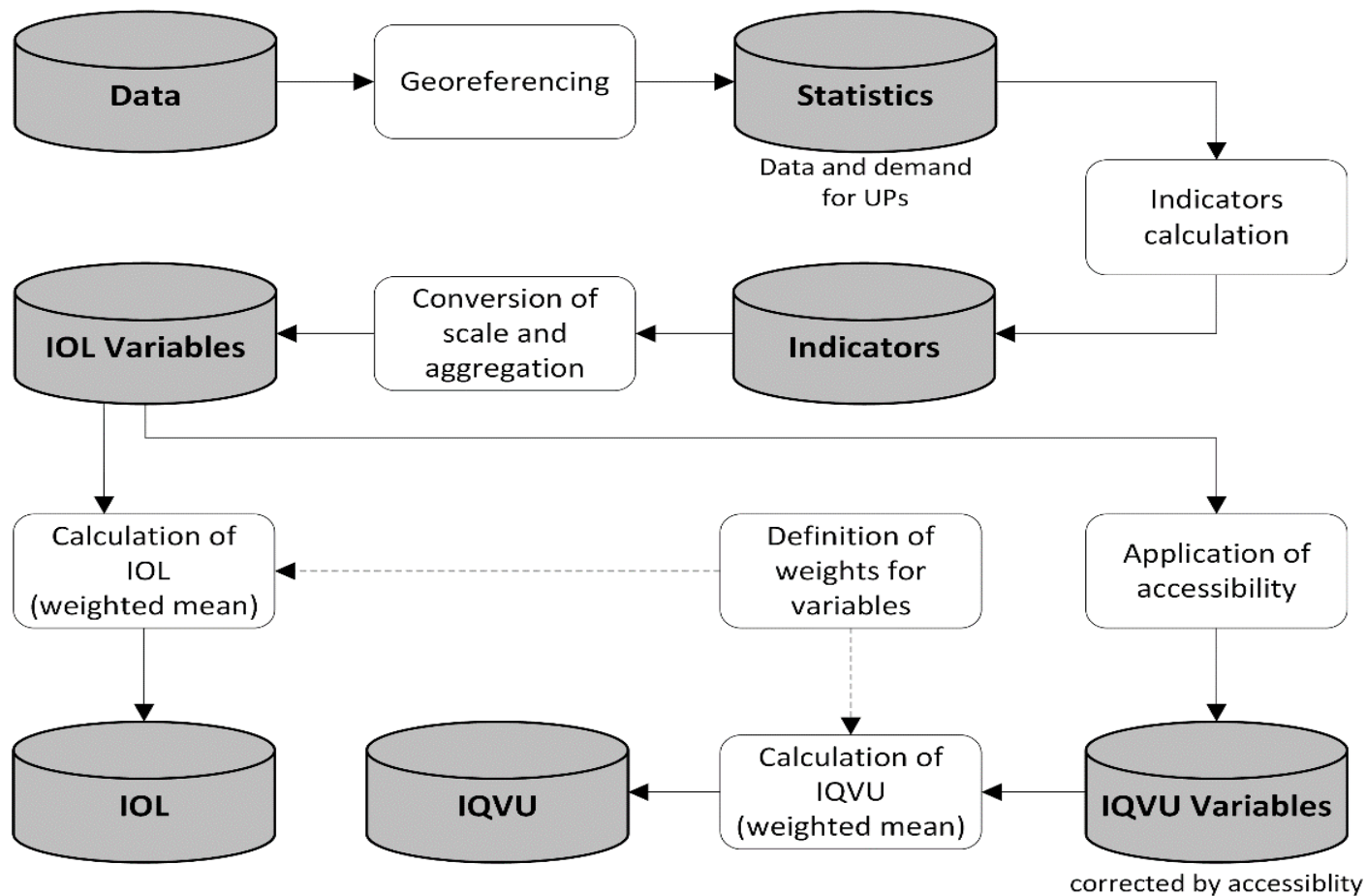
Belo Horizonte

Pampulha Modern Ensemble (World Heritage Site)



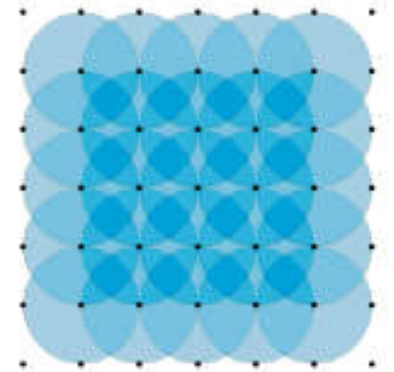
Quality of Urban Life and Local Availability Indexes

| Variable | Component | Indicator |
|----------|------------------|-------------------------------------------------------------------------|
| Supply | Supply equipment | - Hiper and supermarket - Grocery stores and similar |
| Health | Health Care | - Health centers - Other health care equipment - Dental equipment |



Methods: data gathering, characterization and analysis

- Official data sources
- LBSN
 - a grid of ~530k points, with 25m of distance
 - Yelp, Foursquare, Google Places, Facebook
 - data were collected from October 2 to October 25, 2015



Facebook



Foursquare



Google Places



Yelp

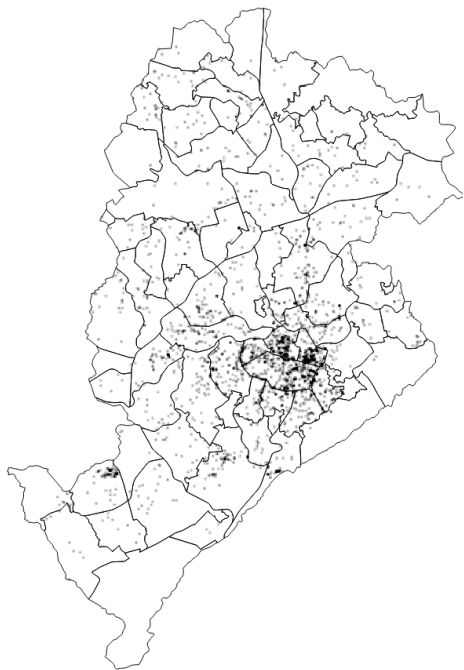
Methods: data gathering, characterization and analysis

- Data processing
 - remove duplicate entries
 - (manual) classification of LBSN categories according to the IQVU indicators
 - LBSN dataset has POI which are not used by IQVU methodology (e.g. laundries and clothing stores)
 - remove unused data

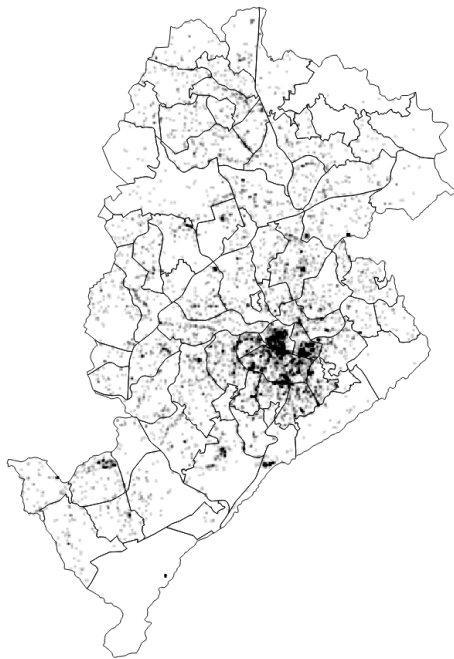
| Operation on data | Facebook | Foursquare | Google | Yelp |
|--------------------------------------------------|----------|------------|-----------|---------|
| Data Gathering (raw data) | 115,137 | 286,227 | 1,389,061 | 354,278 |
| Remove duplicate entries | 42,551 | 91,816 | 169,814 | 57,840 |
| Remove entries without compatible category | 2,214 | 6,827 | 6,456 | 7,669 |

Methods: data gathering, characterization and analysis

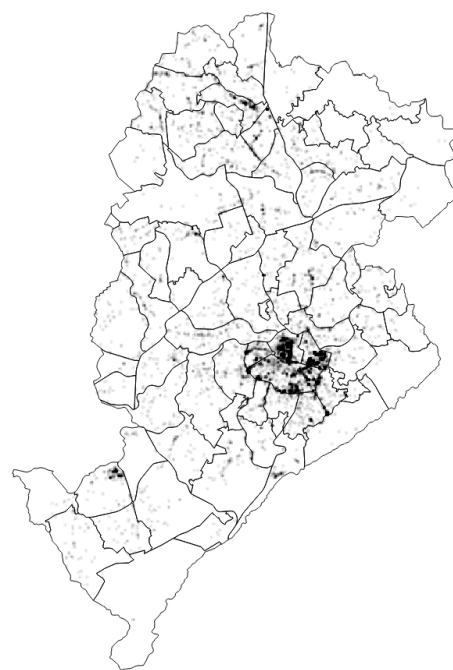
- Differences in the quantity of unique entries and in their spatial distribution
 - downtown concentrates business places
 - deprivation of some regions
- Distribution among categories is also not homogeneous
 - e.g. the number of markets is much higher than the number of hospitals



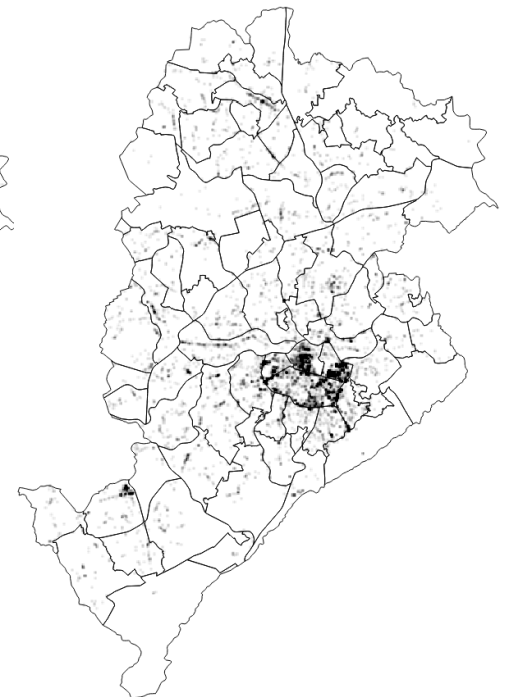
Facebook



Foursquare



Google Places



Yelp

Methods: data gathering, characterization and analysis

| Variable | Component | Indicator | Source |
|----------------------|-------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| Supply | Supply equipment | - Hyper and supermarket - Grocery store and similar | FS/GP FB/FS/GP/Y |
| Culture | Commerce and cultural service | - Cultural equipment (e.g. theater) - Bookstore and stationery - Movie rental store - Magazine stand | FB/FS/GP/Y FB/FS/GP/Y FB/GP/Y FS/Y |
| Education | Childhood education | - % of students enrolled | - |
| | Elementary school | - % of students enrolled - Approval rating | - - |
| | Secondary school | - % of students enrolled | - |
| | | - Approval rating | - |
| Sport | Public space for recreation | - Sport court, field, jogging track | FS/GP |
| Housing | Quality of housing | - Adequate residential area per capita | - |
| | Residential security | - Residential finishing | - |
| | | - Geological risk | - |
| Urban Infrastructure | Environmental healthiness | - Environmental healthiness | - |
| | Electrical power supply | - Electrical power supply | - |
| | Paving | - Possibility of access | - |
| | Collective transport | - Number of vehicles | - |
| | | - Frequency of lines | - |
| Environment | Acoustic comfort | - Tranquility (lack of noise) | - |
| | Air quality | - Absence of collective pollutants | - |
| | Green area | - Green area | FS/GP/Y |
| Health | Health care | - Health centers | FB/FS/GP/Y |
| | | - Other health care services | FB/GP/Y |
| | | - Dental services | FB/GP/Y |
| | Health surveillance | - Absence of life years lost | - |
| Urban Service | Personal service | - Bank agency | FB/FS/GP/Y |
| | | - Gas station | FS |
| | | - Drugstore | FB/FS/GP/Y |
| | Communication and IT service | - Post office | FS/GP/Y |
| | | - Public space for digital inclusion | - |
| Urban Security | Personal security | - Absence of crime against person | - |
| | Patrimonial security | - Absence of crime against patrimony | - |
| | Safety in traffic | - Absence of traffic accident | - |

Methods: calculation of Local Availability Index

- Local Availability Index (IOL)
 - a metric used to calculate IQVU - measure availability of services
 - count, for each indicator, the number of places / UP / population
 - aggregate information from the indicators (e.g. supermarkets) into IQVU variables (e.g. food supply)
 - results are weighted according to the importance of each IQVU variable
- LBSN does not provide data to calculate all indicators of IOL
 - data of IQVU 2012 was used as basis
 - from 36 indicators, it is possible to calculate 15 using data from LBSN
 - 3 of 10 IQVU variables using only LBSN data
 - 6 of 10 IQVU variables using LBSN + Official data

Results

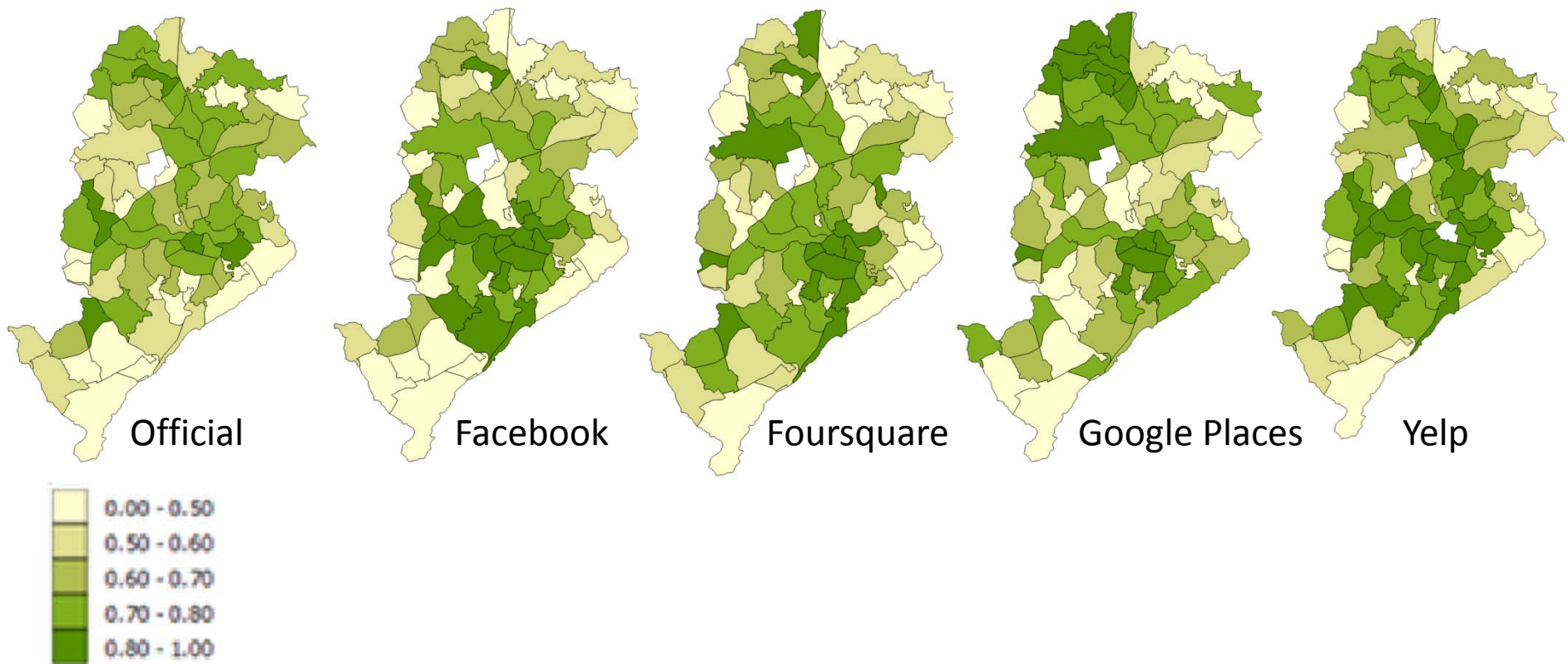
- Evaluation of IOL indicators calculated using data from LBSN
 - values obtained were discretized into the same intervals used by IQVU: [0,0.5), [0.5, 0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 1], labeled 1, 2, 3, 4, 5
 - calculate accuracy, precision and recall using a multi-class approach testing all datasets against official results (2012)
 - good accuracy was achieved, but results for precision and recall were poor

| Indicators | Facebook | | | Foursquare | | | GPlaces | | | Yelp | | |
|----------------------------|--------------|-------|-------|--------------|-------|-------|---------|-------|-------|--------------|-------|-------|
| | acc. | prec. | rec. | acc. | prec. | rec. | acc. | prec. | rec. | acc. | prec. | rec. |
| Hyper and supermarket | 0.392 | | 0.330 | 0.468 | 0.415 | 0.450 | 0.417 | 0.324 | 0.238 | | | |
| Grocery store and similar | 0.367 | 0.231 | 0.206 | 0.126 | 0.112 | 0.156 | 0.341 | 0.146 | 0.178 | 0.303 | 0.199 | 0.204 |
| Cultural equipment | 0.518 | | | 0.746 | | 0.582 | | | | 0.898 | | |
| Bookstore and stationery | 0.405 | 0.127 | 0.152 | 0.531 | | 0.207 | 0.379 | 0.107 | 0.133 | 0.544 | 0.196 | 0.220 |
| Movie rental store | 0.518 | 0.262 | 0.317 | | | | 0.481 | 0.345 | 0.324 | 0.544 | 0.384 | 0.374 |
| Magazine stand | | | | 0.569 | 0.260 | 0.286 | | | | 0.582 | 0.287 | 0.323 |
| Sport court, field | 0.379 | 0.274 | 0.277 | 0.481 | 0.437 | 0.426 | 0.392 | 0.314 | 0.306 | | | |
| Green area | 0.784 | | | 0.835 | | | 0.772 | | | 0.911 | | |
| Health centers | 0.367 | 0.347 | 0.291 | 0.227 | 0.253 | 0.201 | 0.278 | 0.215 | 0.212 | 0.227 | 0.243 | 0.201 |
| Other health care services | 0.708 | 0.310 | 0.368 | | | | 0.746 | 0.352 | 0.428 | 0.848 | 0.487 | 0.403 |
| Dental services | 0.721 | 0.383 | 0.425 | | | | 0.645 | 0.307 | 0.312 | 0.886 | 0.496 | 0.468 |
| Bank agency | 0.481 | 0.174 | 0.195 | 0.721 | 0.295 | 0.352 | 0.658 | 0.405 | 0.497 | 0.506 | 0.158 | 0.172 |
| Gas station | 0.518 | 0.261 | 0.339 | 0.594 | 0.403 | 0.398 | 0.367 | 0.251 | 0.270 | | | |
| Drugstore | 0.189 | 0.079 | 0.106 | 0.443 | 0.231 | 0.299 | 0.291 | | 0.142 | 0.607 | 0.536 | 0.533 |
| Post office | 0.759 | 0.330 | 0.363 | 0.734 | 0.385 | 0.540 | 0.658 | 0.422 | 0.574 | 0.734 | 0.287 | 0.269 |

Results

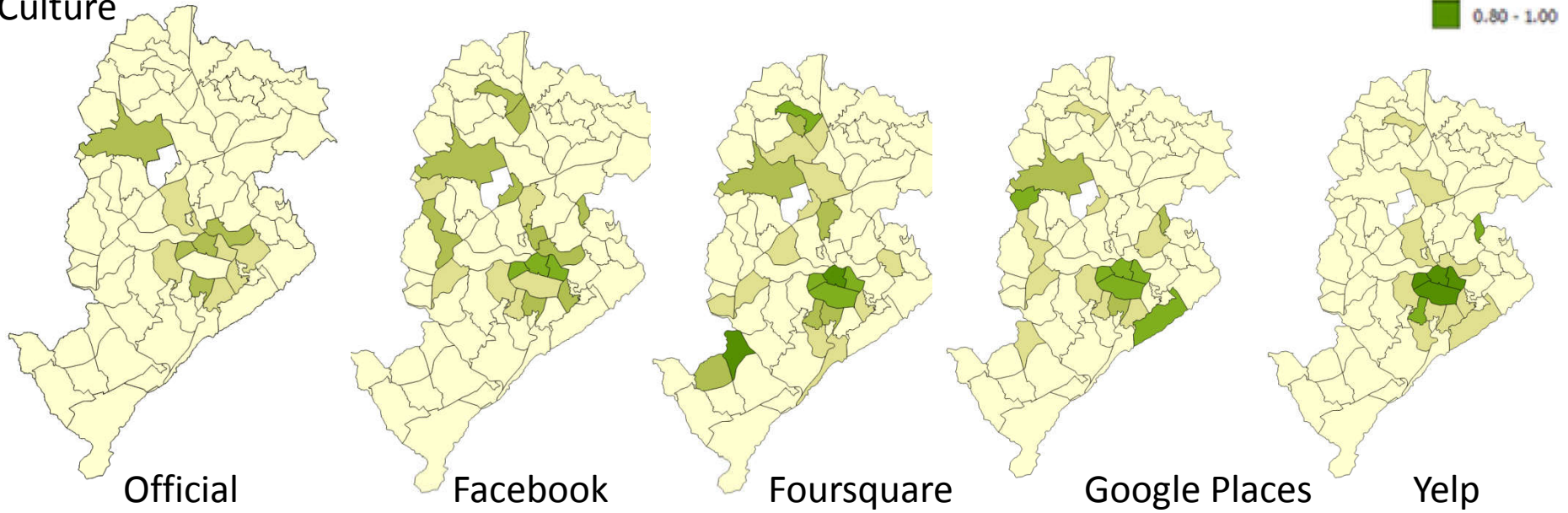
- Spatial patterns of the results of the IOL variables calculated using LBSN

Food Supply

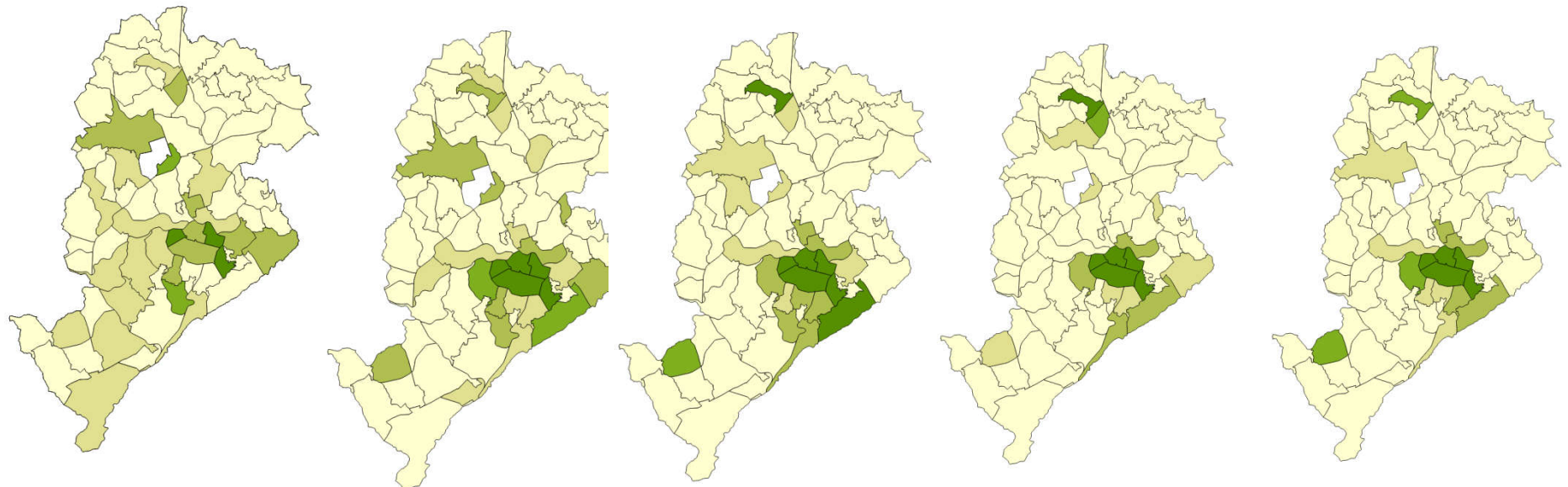


Results

Culture



Health



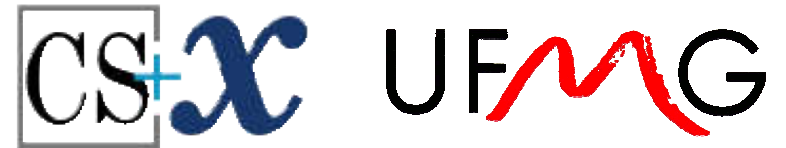
Limitations

- LBSN data can be biased by being crowdsourced
- Lack of some types of information from LBSN
 - it was not possible calculate all indicators
 - this could be solved with government open data initiatives that supports availability of information in machine readable formats
- Lack of mobility data to calculate accessibility (required to calculate IQVU final index)
- Lack of updated results of IQVU for comparison
 - LBSN data are from the third quarter of 2015, the latest IQVU results are from 2012
- With open data initiatives and improvements of this work will be possible to produce frequently and updated results to support urban planning

Final remarks

- There is a potential use of LBSNs data to calculate urban indicators
- Despite the limitations, the achieved results encourage us to perform further investigations
 - How to integrate data from several LBSN? Better results can be achieved with data integration?
 - How to assess quality of urban life? Which indicators should be used?
 - much of data collected were not used
 - some indicators seem irrelevant or less important (e.g. movie rental stores, public phone) and others currently important are not considered (e.g. broadband internet)
 - Challenge: how to evaluate results with an outdated baseline?
- Future work: a VGI framework to dynamically measure individuals perception about the quality of urban life, adjusting the results according to city dynamics and citizen behaviour

Acknowledgments



- LocWeb Organizers and Technical Program Committee
- Brazilian funding agencies

Tiago Lima (tiagofml@gmail.com)
www.labcsx.dcc.ufmg.br

