Big Data , GeoAnalytics and Systems Analysis

Eduardo de Rezende Francisco

eduardo.francisco@fgv.br

FGV EAESP
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

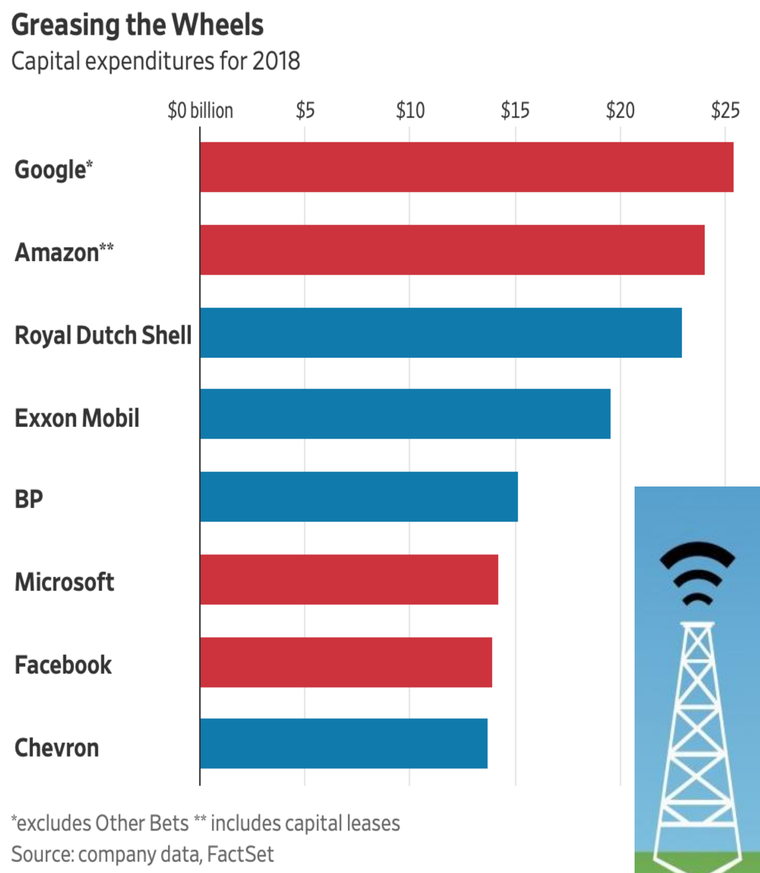
Apresentação

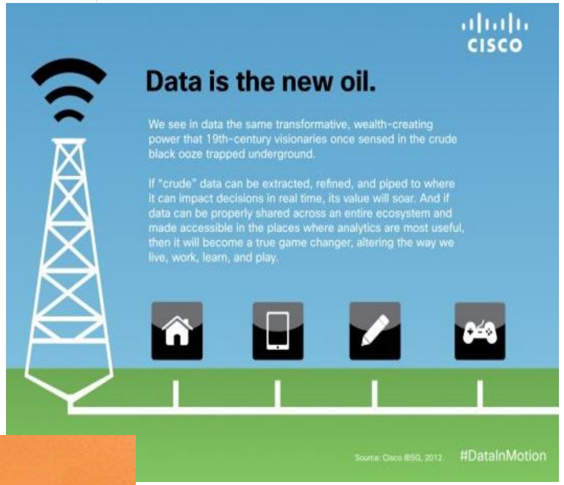
Professor of the TDS (Technology and Data Science) Department at **FGV EAESP**. He teaches disciplines related to GeoAnalytics, Spatial Statistics and Data Science. Bachelor in Computer Science from IME-USP and Master and PhD in Business Administration from FGV EAESP. Deputy Coordinator of the Graduate Course in Business Administration at FGV EAESP. He is a visiting researcher at the Spatial Information Research Center at the University of Otago, New Zealand.

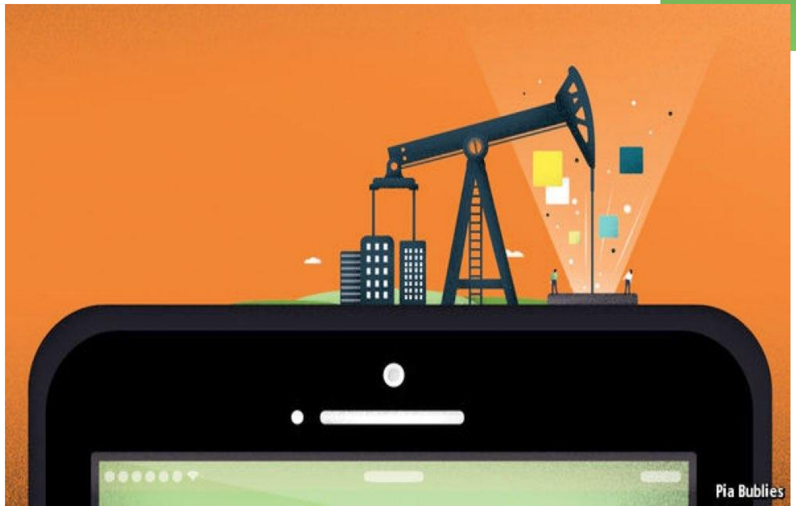
He is CKO (Chief Knowledge Officer) of startup **Meia Bandeirada** and founding partner of **GisBI**, a study and promotion group for Geotechnologies and Business Intelligence.

He has been working in the GIS, Geomarketing and Business Intelligence markets since 1994. He was technology manager of the GIS project and Market Planning project and responsible for the development of Marketing Strategies and Market Research Planning at AES Eletropaulo for 13 years. Participates as a speaker in various Geomarketing, BI, GIS and Utilities events. He is a columnist for GV Executivo and InfoGEO magazines, and a consultant in Spatial Statistics and Predictive Models for Credit and Real Estate. Member of the **Fundação SEADE** Board of Trustees.

eduardo.francisco@fgv.br







The Economist | Topics | Current edition | More

You are now logged in

Regulating the internet giants

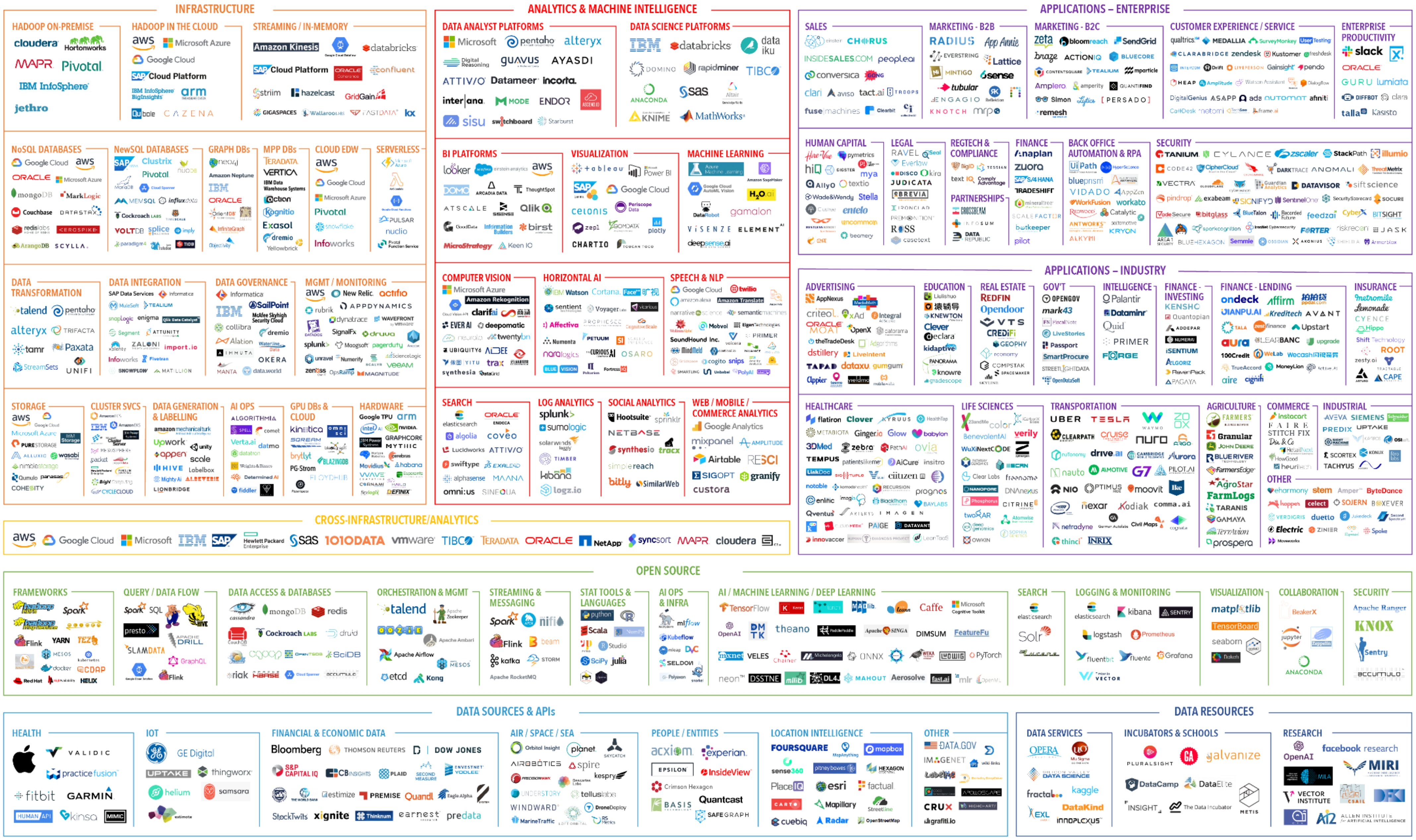
The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

Print edition | Leaders ›
May 6th 2017

A NEW commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era. These

DATA & AI LANDSCAPE 2019



July 16, 2019 - FINAL 2019 VERSION

 mattturck.com/data2019

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

BIG DATA – “Definition”

“Big Data” are <u>data and processes</u> whose scale, distribution, diversity and / or creation speed requires the use of new storage and analysis technologies to allow the capture of the value inserted in them

(FRANCISCO, 2016, adapted from EMC, 2013)

1. Data Volume

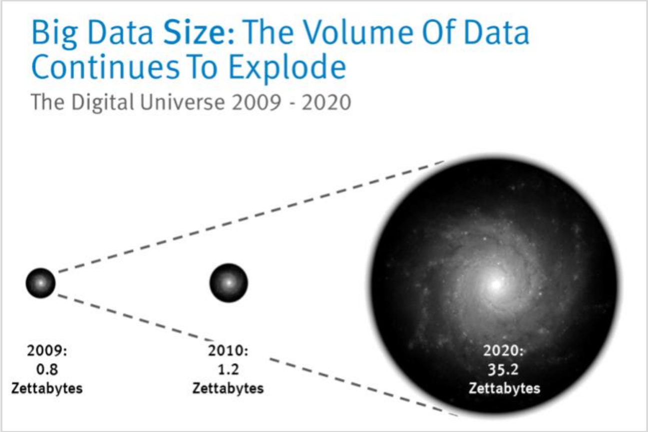
- Billions of lines x billions of columns
- Increase of 44 times from 2009 to 2020 (0,9ZB to 35ZB)

2. Processing Complexity

- Data structures are constantly changing
- Need to analyze such data in real time

3. Data Structure

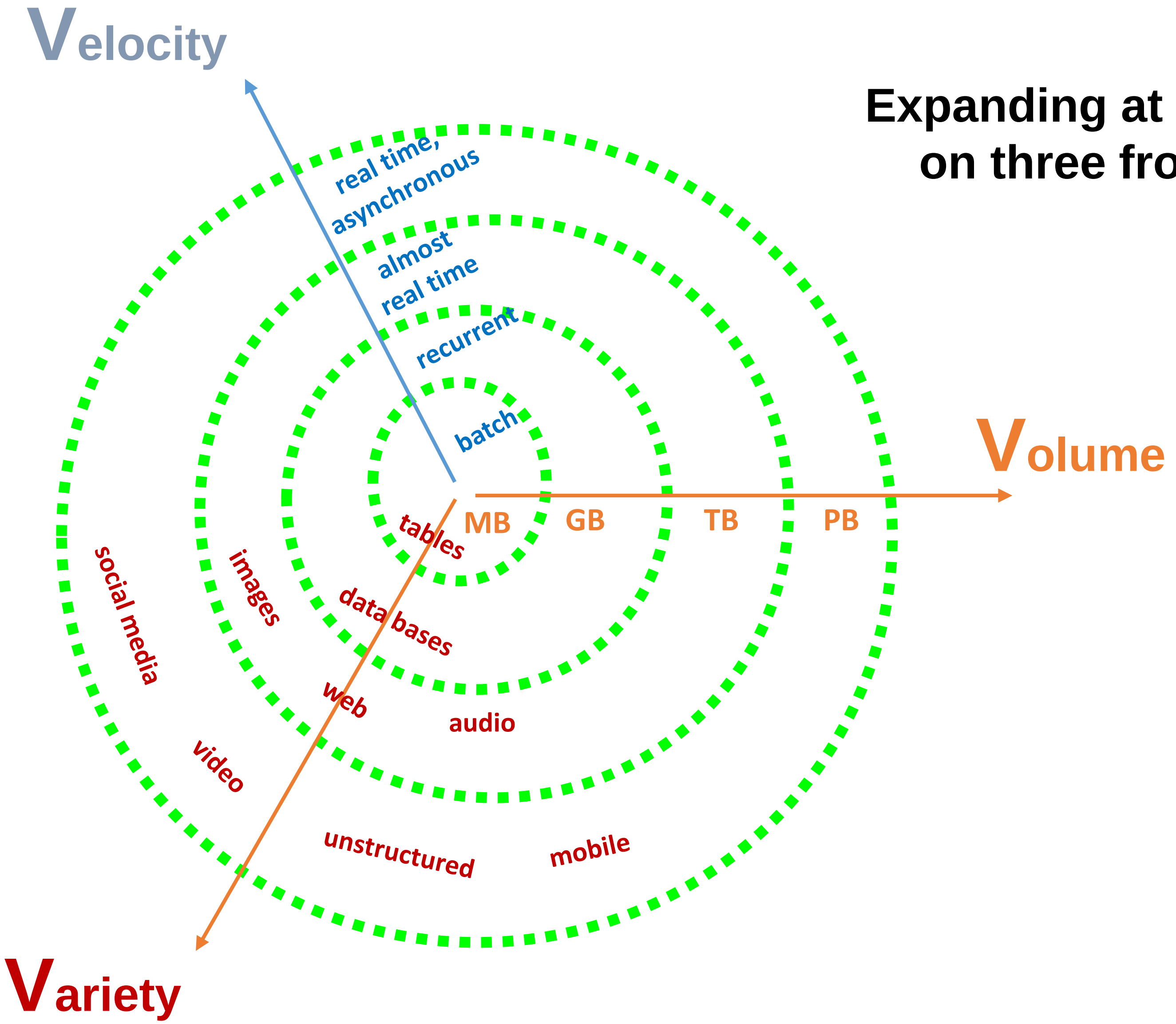
- Large variety (80-90% unstructured) to be analyzed
- These characteristics make it necessary to use parallel and parallel mass computing (MPP) systems



Source: EMC (2013)

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

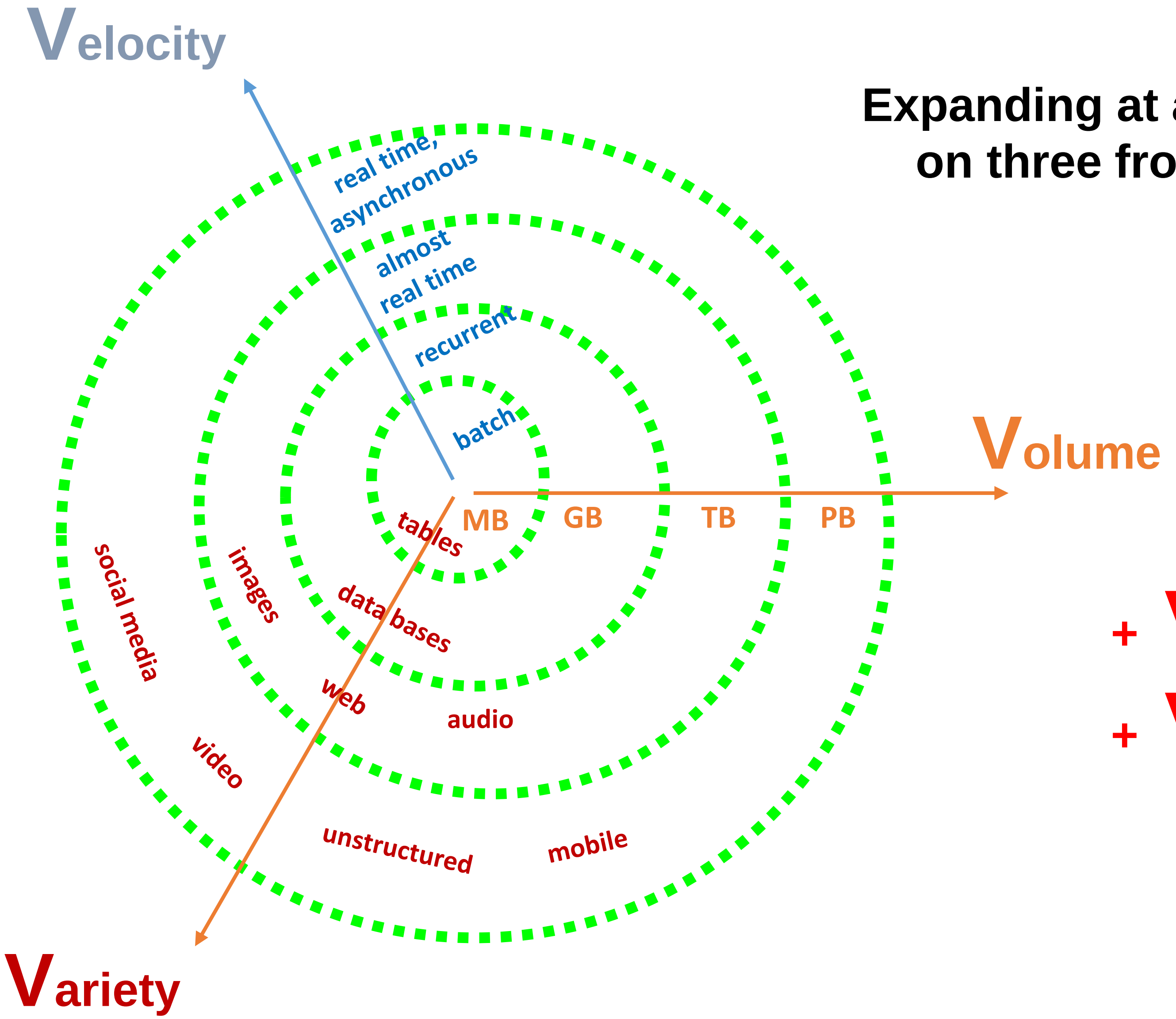
3 Vs



Expanding at an increasing rate on three fronts (the first 3 Vs)

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

5 Vs

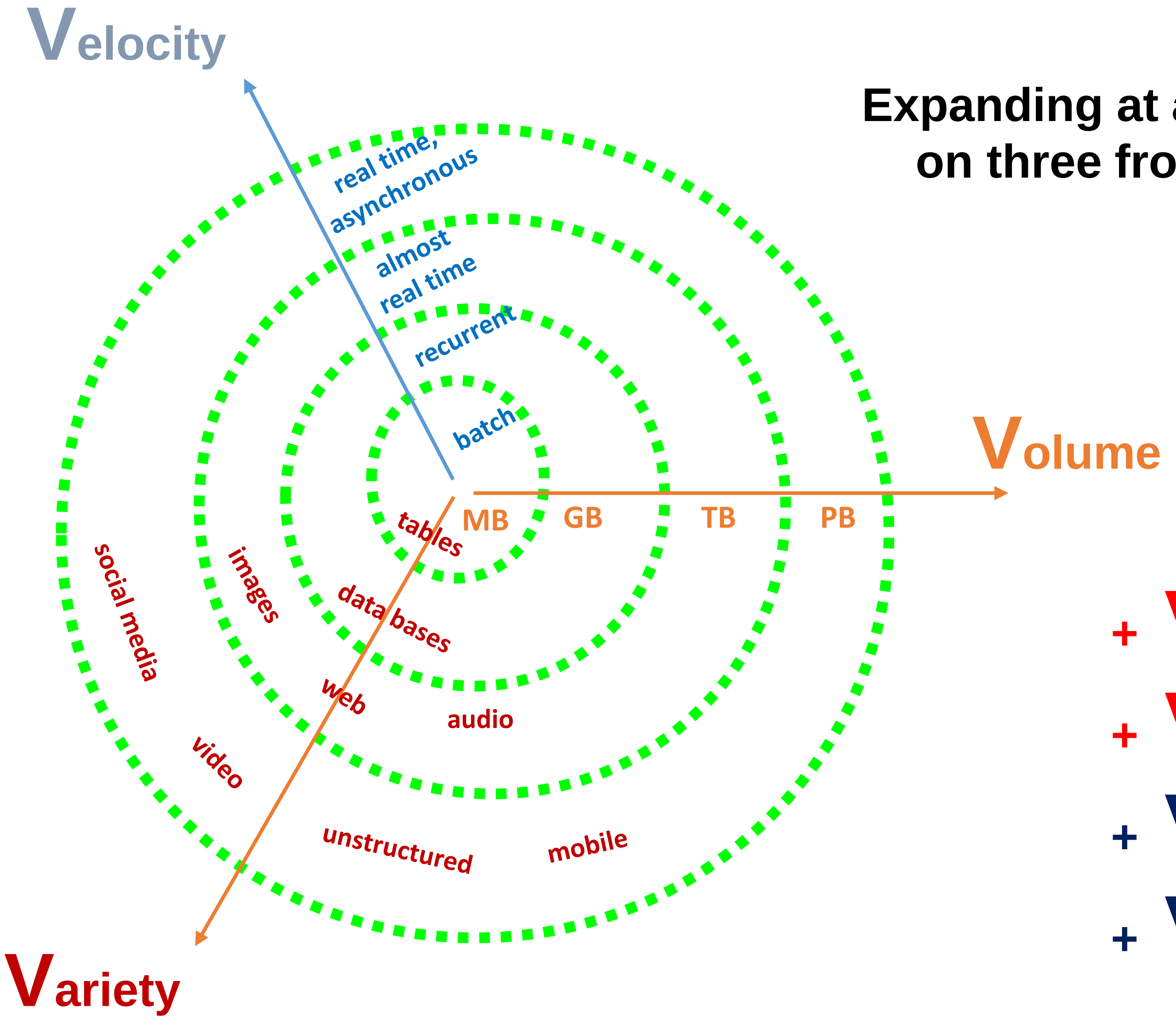


Expanding at an increasing rate on three fronts (the first 3 Vs)

+ **V**eracity
+ **V**alue

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

7 Vs



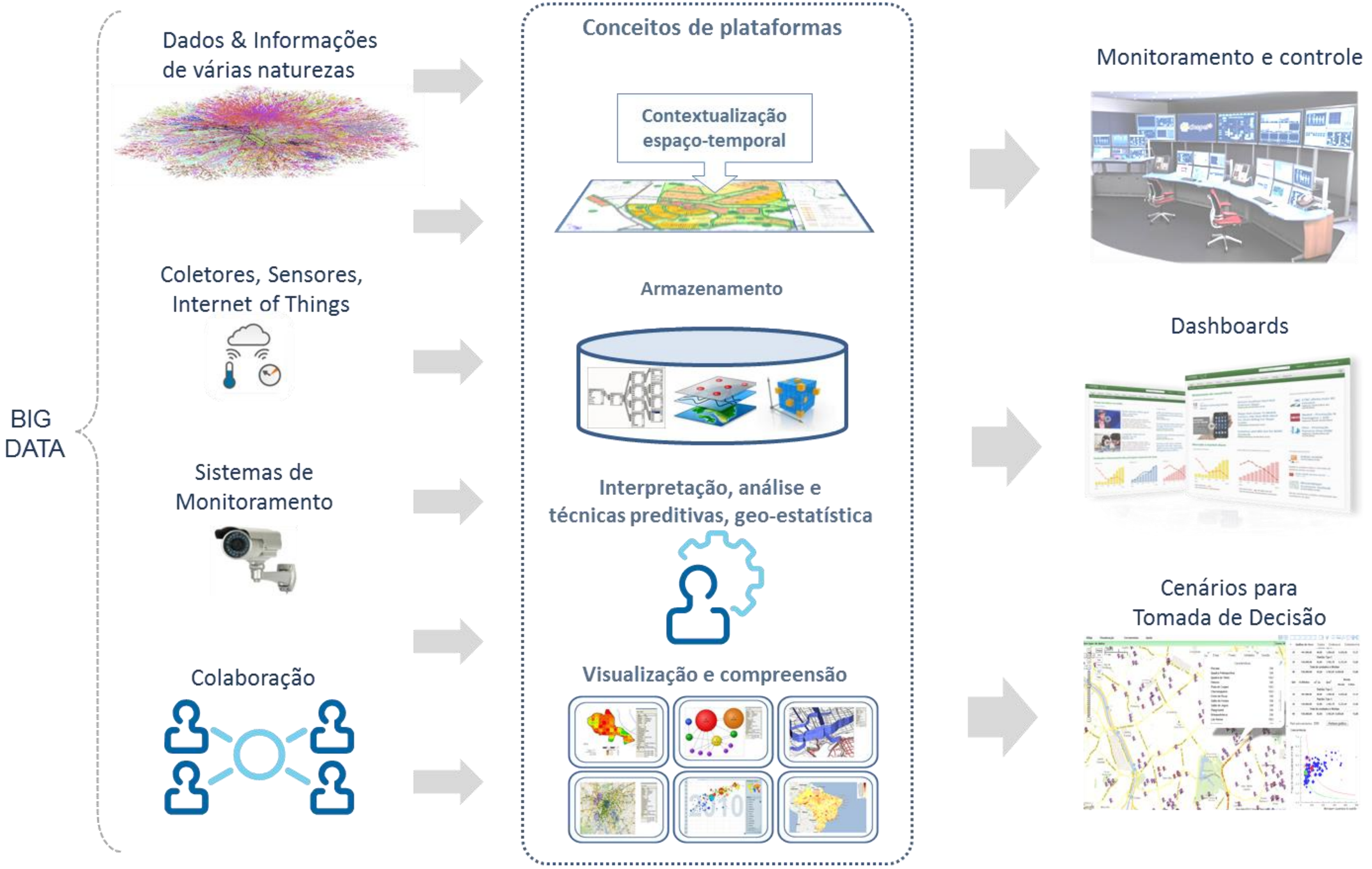
Expanding at an increasing rate on three fronts (the first 3 Vs)

+ **Veracity**
+ **Value**
+ **Variability**
+ **Visualization**

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

FGV

Big Picture


BIG DATA
Dados & Informações de várias naturezas
Coletores, Sensores, Internet of Things
Sistemas de Monitoramento
Colaboração
Conceitos de plataformas
Contextualização espaço-temporal
Armazenamento
Interpretação, análise e técnicas preditivas, geo-estatística
Visualização e compreensão
Monitoramento e controle
Dashboards
Cenários para Tomada de Decisão

Challenges for Systems Analysis in the Big Data Era

- Big Data, Data Science, Analytics ...
- Challenge in the integration of analytical techniques
  - AI, Neuroscience applied to marketing and business, spatial statistics, behavioral models
  - Models to handle stakeholders' relationships
- Computational Challenge
- Challenge in the adoption of AI and Data Science by companies and public organizations
  - Adoption of forceps, apart from core management
  - Analytical Sandbox vs. IT Policy
- **Cultural challenge - new skills of analytical teams and managers**

Use of alternative information (non-structured "Big" data) in the generation of indicators

- Open Data - API or Web Scrapping
- Sectoral Reports, Management Reports, Integrated Reports - Web Scrapping
- Interpretation of Images
- Data Enrichment by Geo-Analysis and Spatial Statistics

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

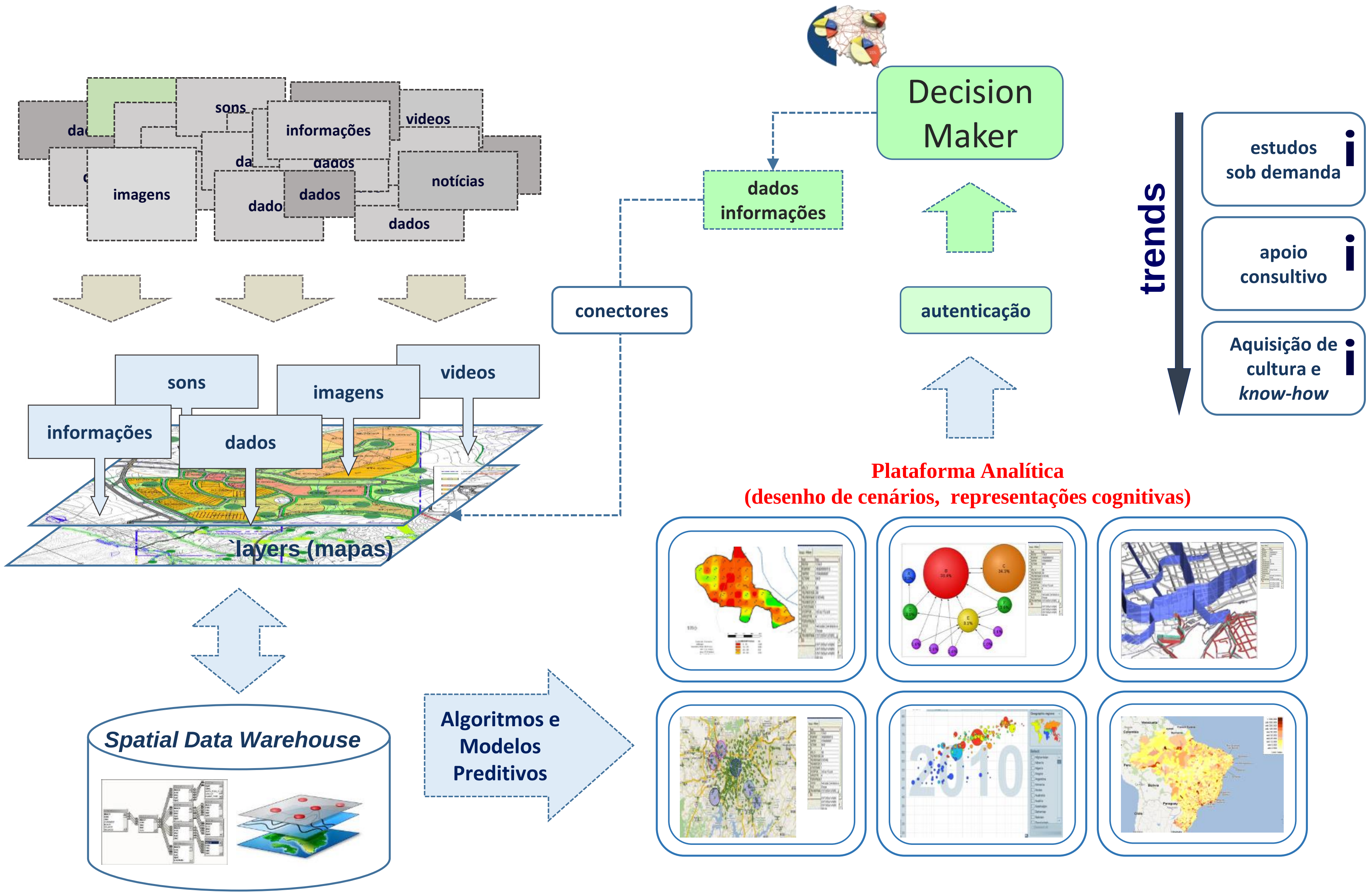
Serious Implications for Scientific Research in Applied Social Sciences

Classic (Inferential) Statistics	Machine Learning
Strong hypotheses	Flexible approach
Theoretical justification	Empirical Efficacy
Exact solution	Approximate solution
Explanation/Interpretation	Prediction

Source: OLLION, 2018

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

Framework: Adaptive Business Intelligence and Big Data



Source: LETOUZÉ, 2017;
MICHALEWICZ et al., 2006 ;
GisBI, 2015

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

Challenges

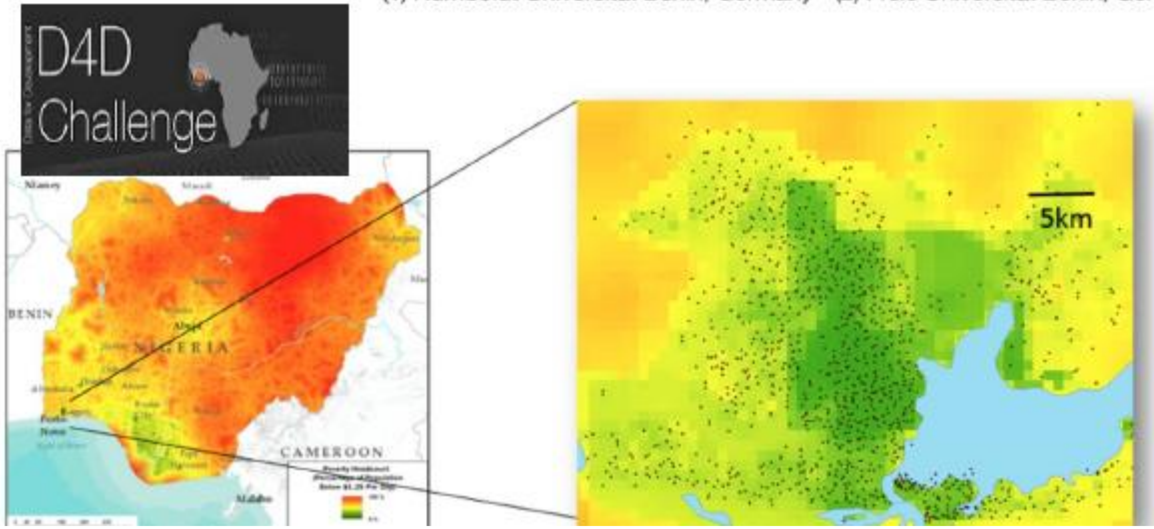
Inferring socio-demographic indicators

Scientific Prize and Ethics Mention: Construction of socio-demographic indicators with digital breadcrumbs

F. Bruckschen [1], T. Schmid [2], T. Zbiranski [1]

We show that socio-demographic indicators such as population, age, literacy, poverty, religion, ethnicity, electricity supply and others can be estimated in unprecedented detail and virtually ad-hoc using antenna-to antenna traffic data only. We offer a uniform approach that can be easily extended to other variables. Results are tested for spatio-temporal robustness and visualized as heat maps.

(1) Humboldt Universität Berlin, Germany - (2) Freie Universität Berlin, Germany



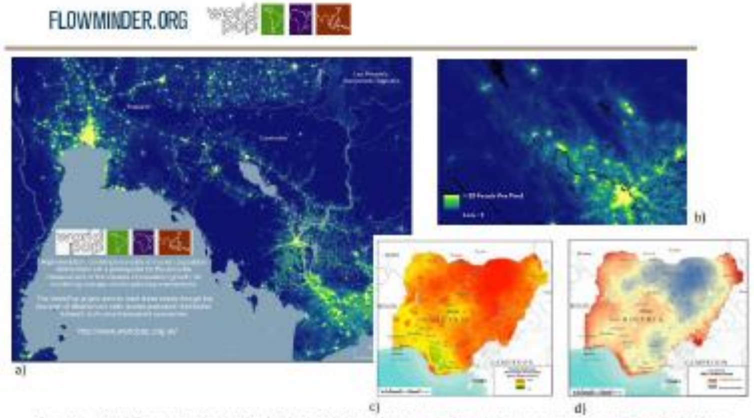


Figure 1: a) WorldPop population density distributions for the the Mekong region; b) close-up picture of population distributions (100x100m) for the Hanoi region; c) Poverty headcounts for Nigeria (<1.25 USD/day) per 1 km²; d) Uncertainty in poverty headcount estimates per 1 km² area.

Source: LETOUZÉ, 2017

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

FGV

OBJ: Analyze the relationship between household income and electricity consumption

Create an income indicator based on electricity

Prediction Model: Electricity Consumption $\xrightarrow{+}$ Household Income

Traditional OLS Regression:

$$\hat{y} = \beta_0 + \beta_1 x$$

$R^2 = 86{,}6\%$

SAR (*Spatial Auto-Regressive*):

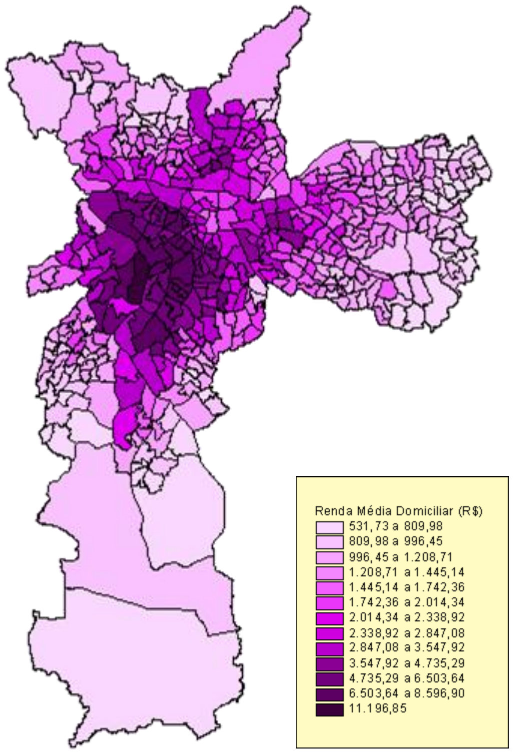
$$\hat{y} = \beta_0 + \beta_1 x + \rho W y$$

$R^2 = 94{,}4\%$

GWR (*Geographically Weighted Regression*):

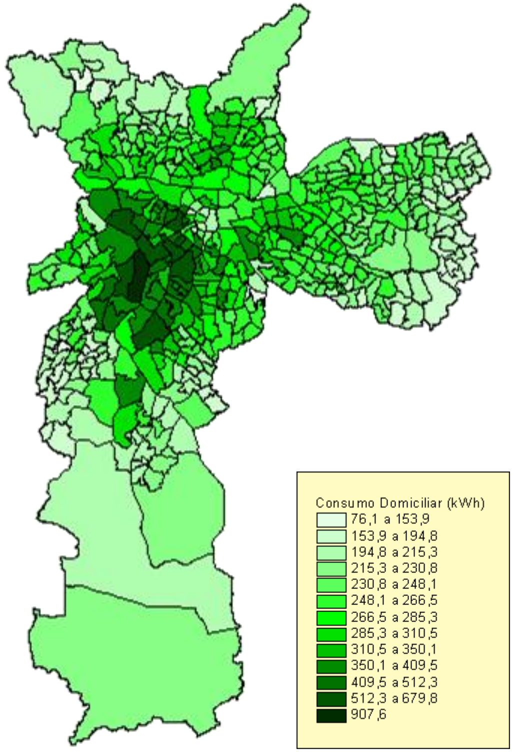
$$\hat{y}_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\ x_i$$

$R^2 = 96{,}8\%$



Household Income

(IBGE)



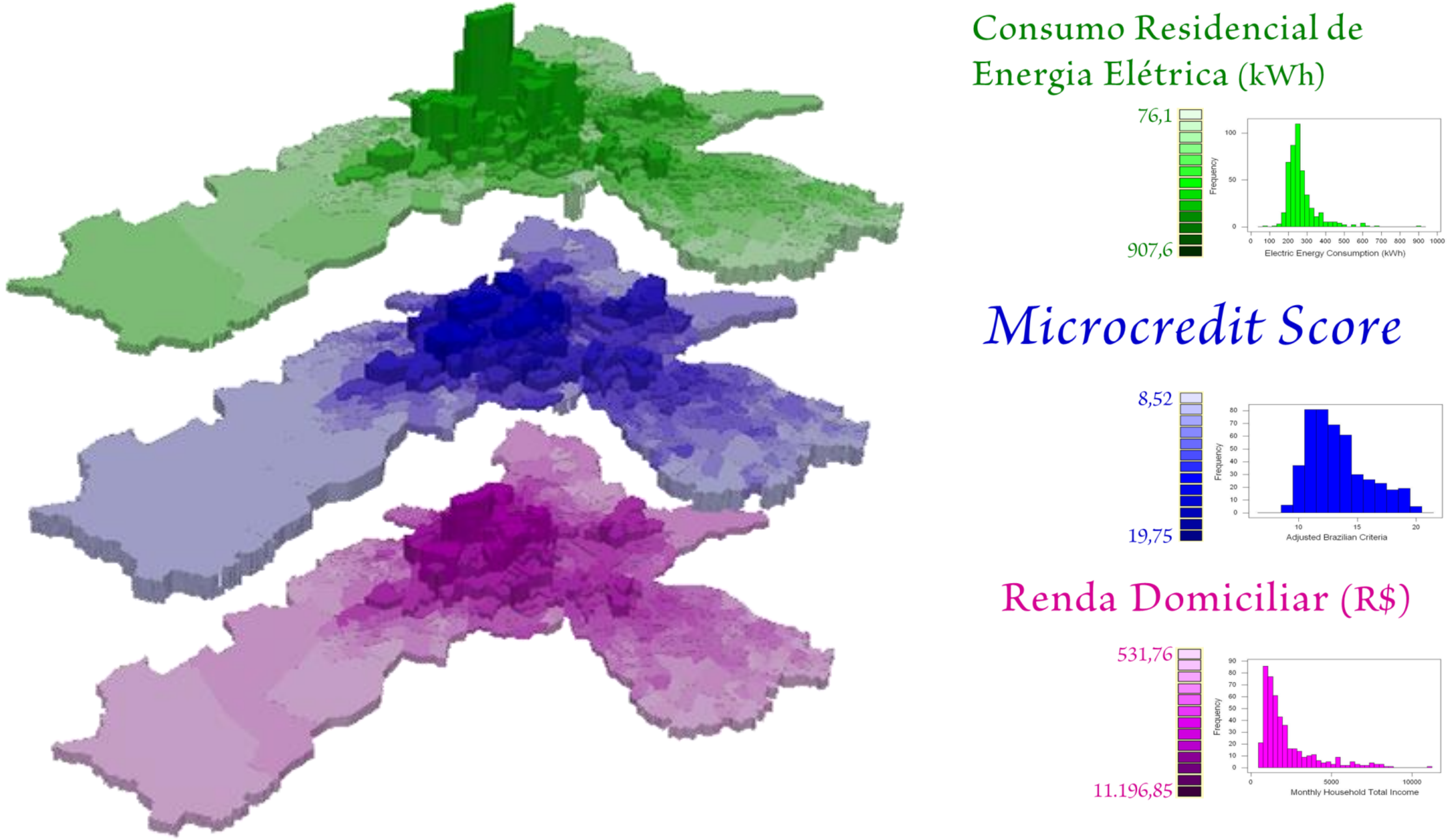
Electricity Consumption

(AES Eletropaulo)

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

FGV

Microcredit Score and Socio-Economic Indicators based on Electric Energy



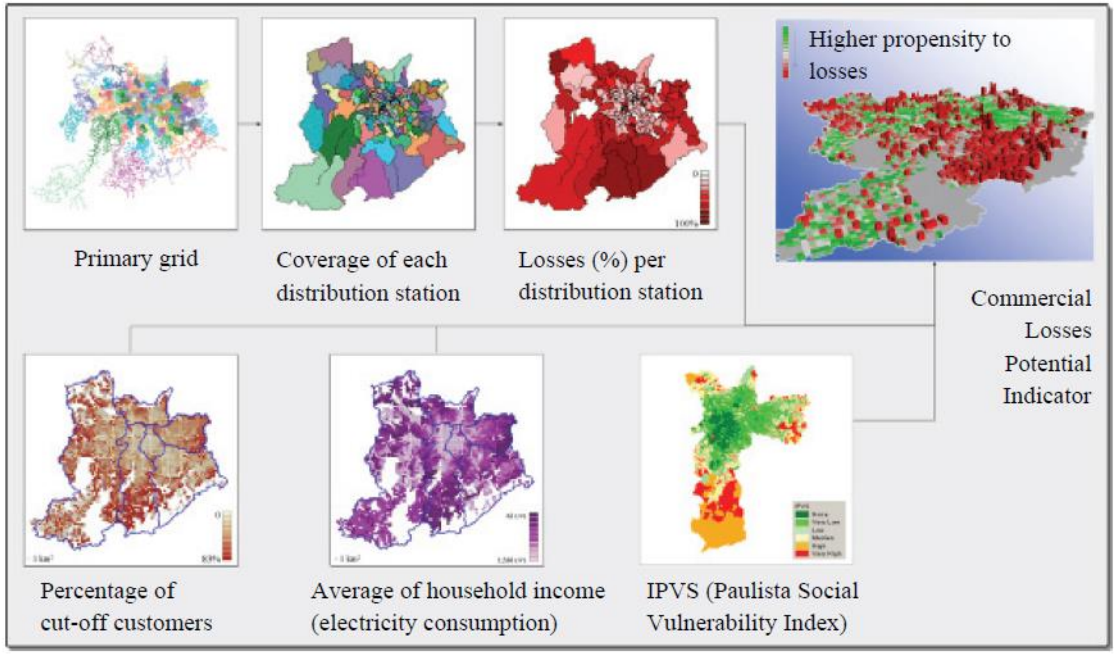
Source: FRANCISCO, 2011

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

- ## Development of an Indicator of Propensity to Energy Commercial Losses using Geospatial Statistical Techniques and Socio-Economic Data: the Case of AES Eletropaulo

FRANCISCO, E., FAGUNDES, E., PONCHIO, M., ZAMBALDI, F. - EnANPAD 2009

  - Use of geospatial techniques for the design and operation of a fraud-prone indicator, and analysis (through GWR) of the association between this indicator and customer satisfaction



Source: FRANCISCO *et al.*, 2009

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

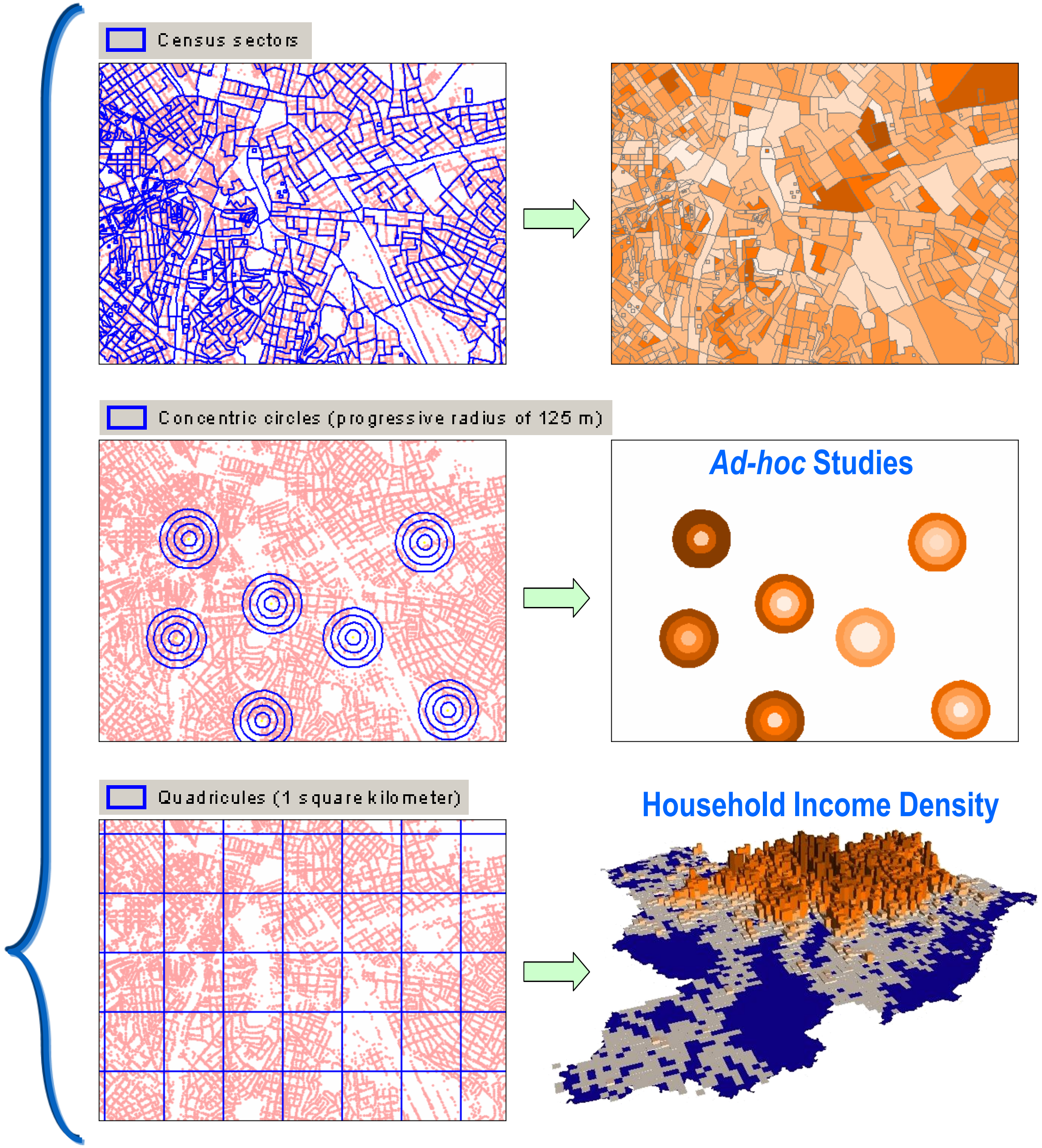
FGV

Managerial Implications

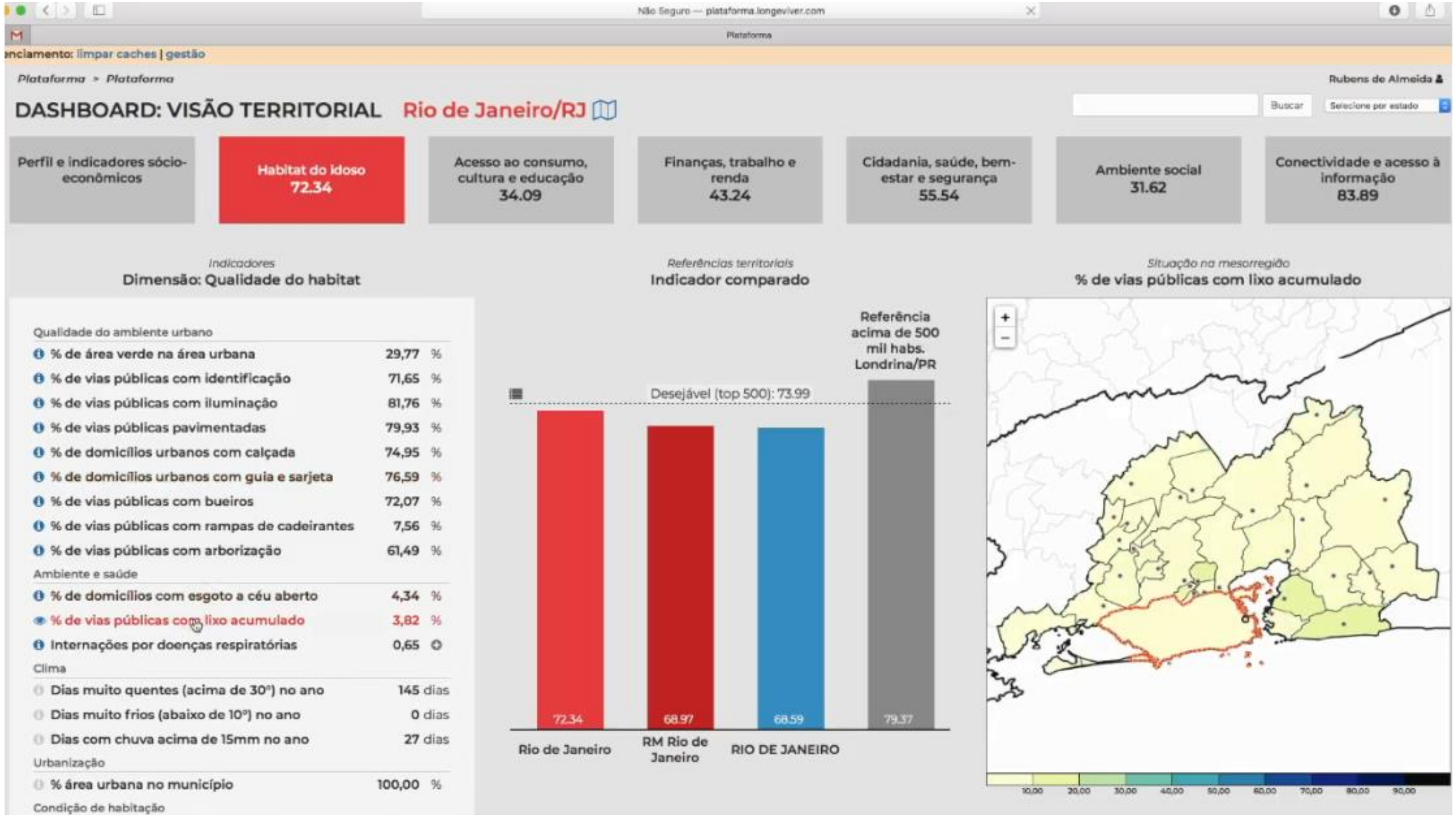
Socio-Economic Indicators based on Electricity Consumption

- **Should be published widely by the electricity distributors**
- **Useful for strategy formulation and decision support**
- ***Support for Customer Relationship Characterization and Management***
- ***Support for Public Policies and Systems Analyses***

Households



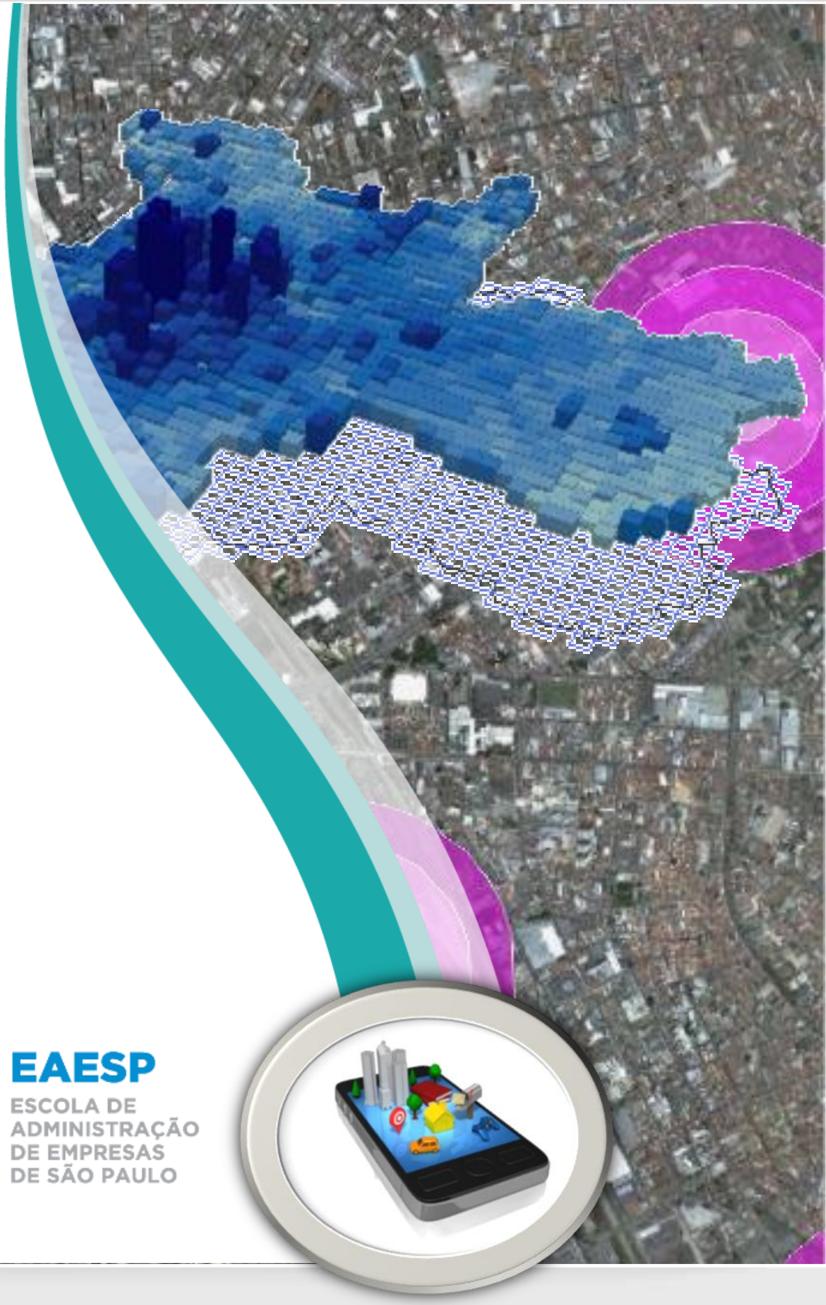
Health and Longevity Research and Study Platform



https://youtu.be/1ET4glwLAe0

Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

Thank you!
¡Gracias!
Obrigado!



Eduardo de Rezende Francisco
eduardo.francisco@fgv.br

FGV EAESP
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO