



This project is co-funded by the  
Seventh Framework Programme for  
Research and Technological  
Development of the European Union



# Exploring government contracting data to hold governments accountable in the Big Data Era

Mihály Fazekas

University of Cambridge, [mf436@cam.ac.uk](mailto:mf436@cam.ac.uk)

Big Data Methods for Social Science and Policy conference,  
Cambridge, UK, 24/9/2015

2015.09.28.



UNIVERSITY OF  
CAMBRIDGE  
Department of Sociology

# The challenge of public procurement

9.3 trillion USD per year\*

20-25% lost to corruption\*\*

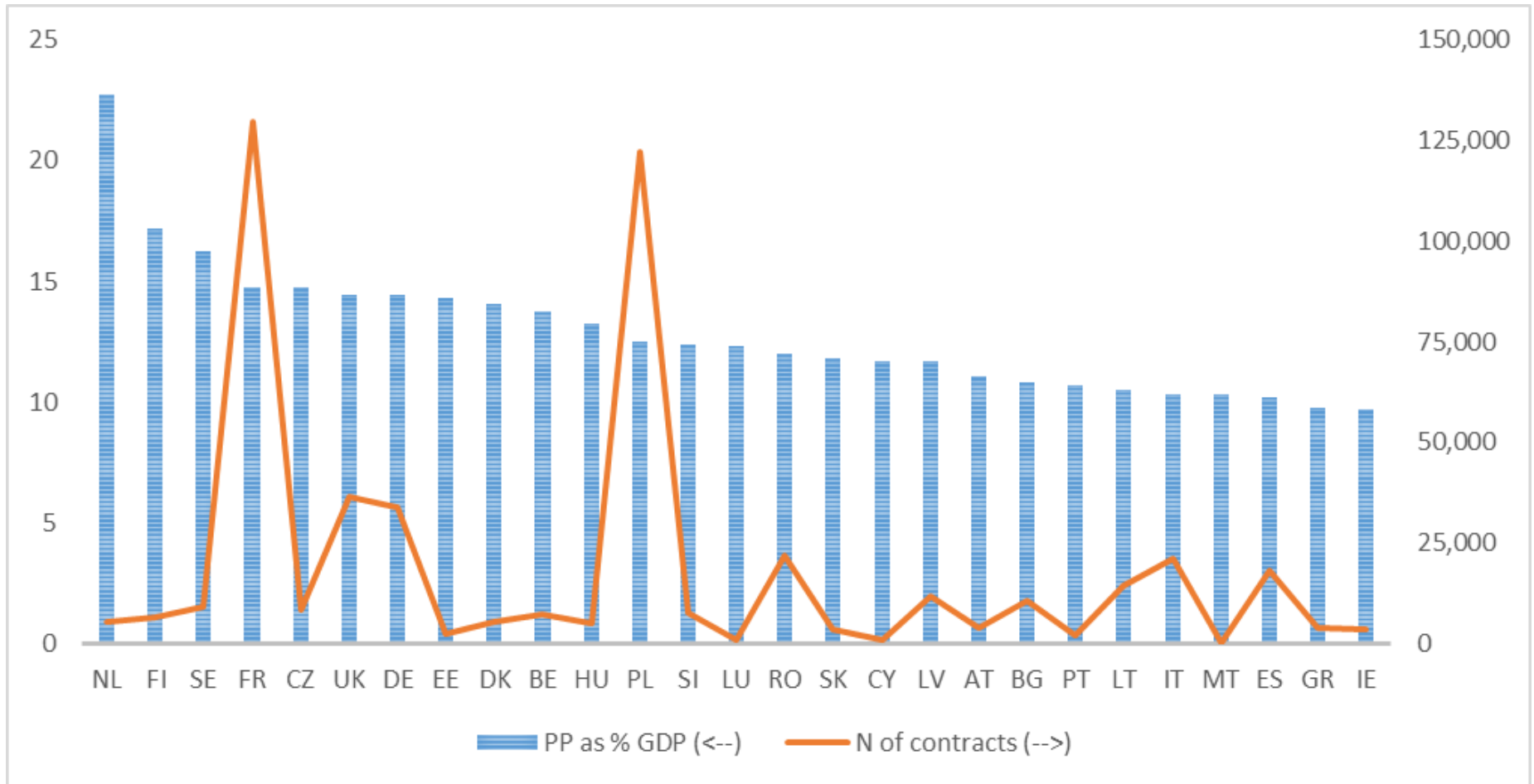
**Should we be worried?**

\* <http://www.open-contracting.org/data-standard-announcement>

\*\* OECD (2013), Implementing the OECD Principles for Integrity in Public Procurement. Progress since 2008. OECD, Paris

2015.09.28.

# Why we need big data in PP?



2012, EU, TED

2015.09.28.

# State of data

- Patchwork of data sources
  - Formats: xml,html,pdf, etc
  - Content
  - Quality
  - Linking
- Administrative data provision is not ready for Big Data
  - Public initiatives: OCDS
  - Civil society/research community: DIGIWHIST

# DIGIWHIST

The Digital Whistleblower. Fiscal Transparency, Risk Assessment and Impact of Good Governance Policies Assessed

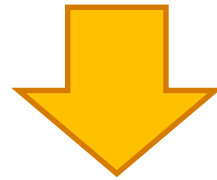
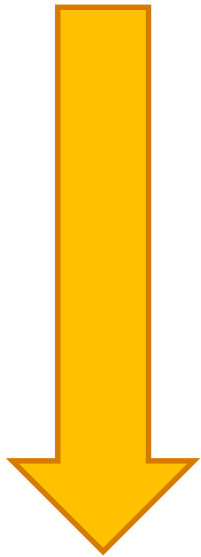
## Goals

1. Advancing anticorruption, transparency, and spending efficiency in public procurement
2. Unlock research potential
3. Enabling losers of corruption to mobilize
4. Helping audit bodies fighting corruption, fraud, and collusion

34 European countries+EC

# DIGIWHIST: Data

Legislation: transparency & procurement



Transactions: contracts & tenders



Actors: Companies, public bodies, individuals

# DIGIWHIST: Indicator building

- built on thorough understanding of context
- ,open-ended'
- actionable+real-time

# DIGIWHIST: Measuring government quality

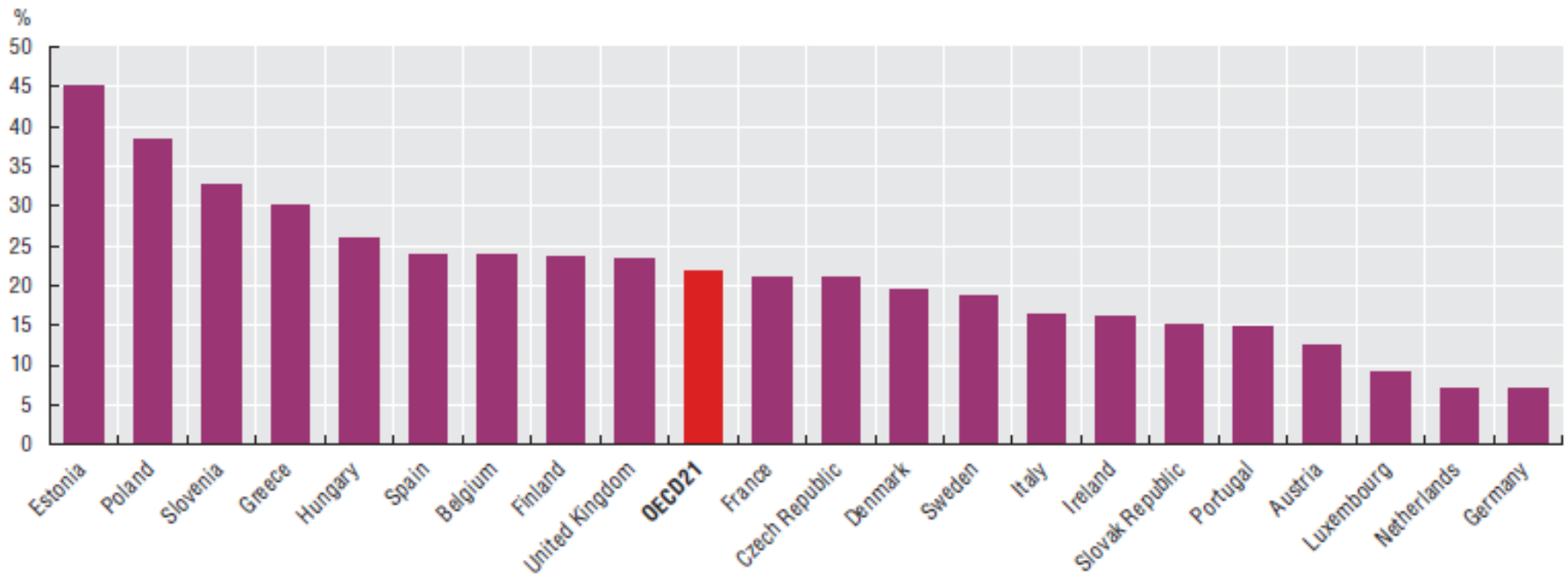
- Transparency
  - Share of spending advertised



# Advertised vs total PP spending

- MOST of the data is missing

40.3 Public procurement tenders advertised in the OJEU as a percentage of total national procurement market (2008)

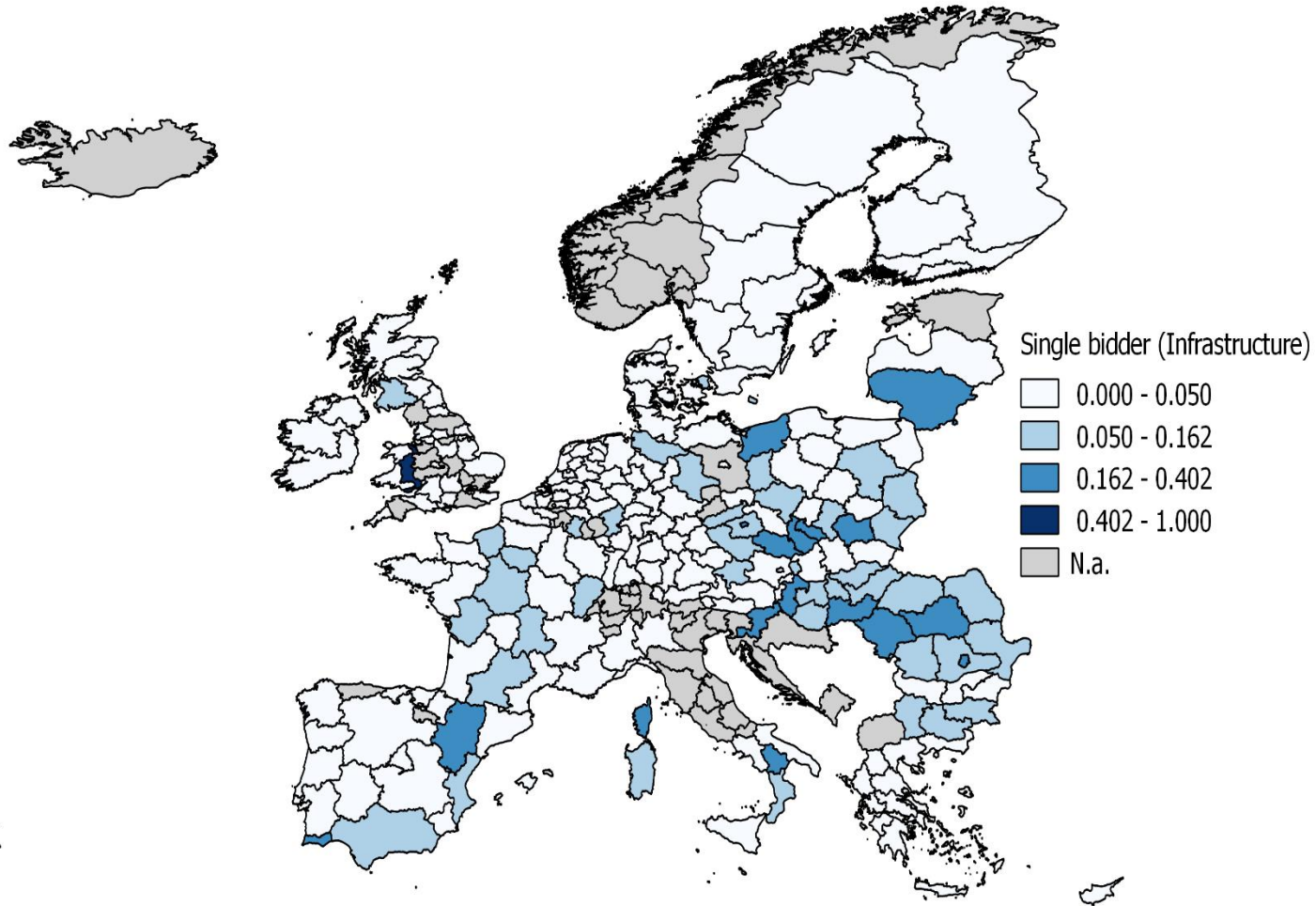


# DIGIWHIST: Measuring government quality

- Transparency
  - Share of spending tendered
- Corruption
  - Tendering red flags: single bidding

# Single bidding

NUTS regions, large infrastructure only



# DIGIWHIST: Measuring government quality

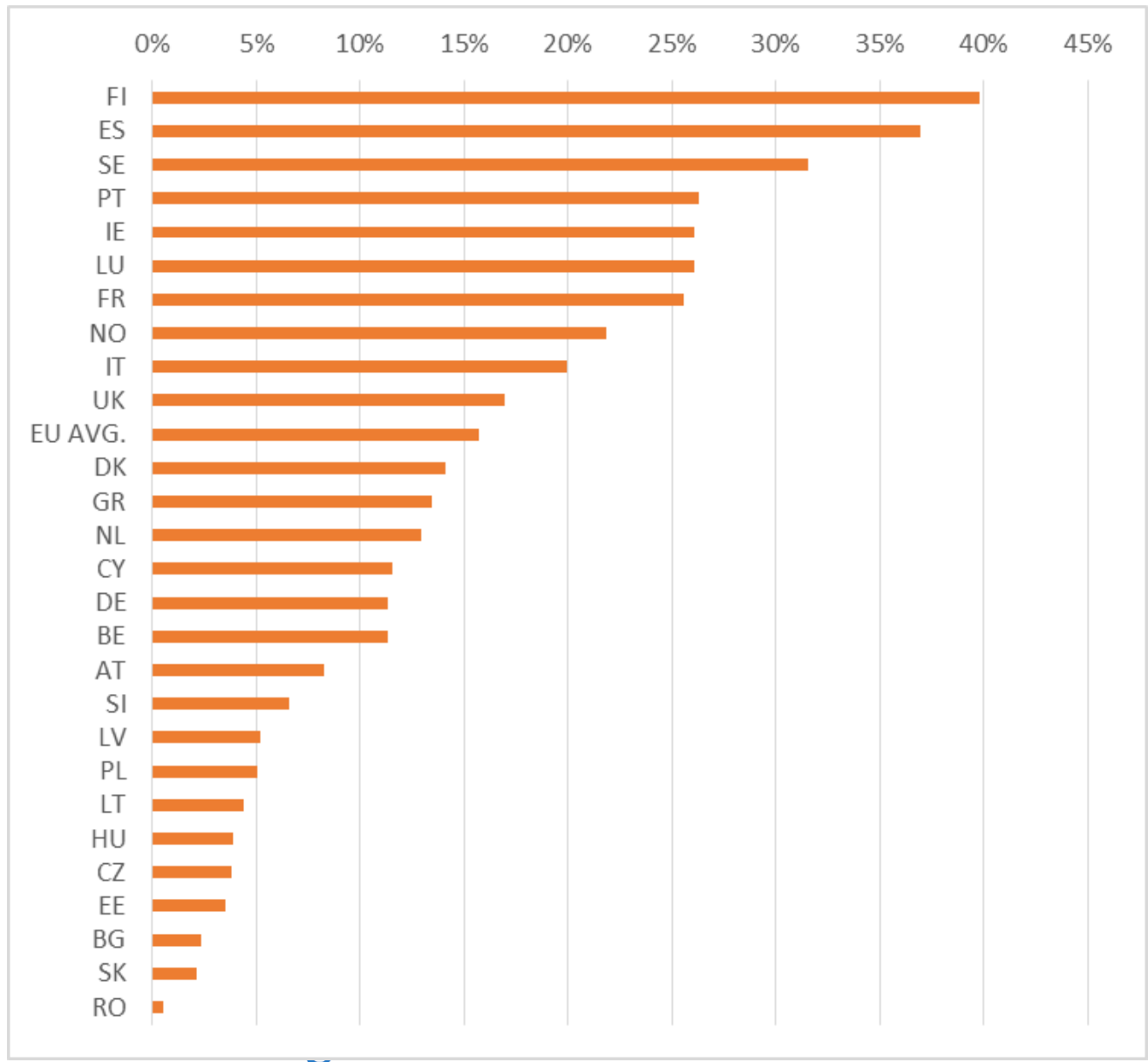
- Transparency
  - Share of spending tendered
- Corruption
  - Tendering red flags: single bidding, exceptions
- Administrative error
  - Missing data

# Average admin error

TED:2009-13

13 mandatory items:

- Organisation name, address
- Contract values
- Subcontract
- Dates



2015.09.28.

# DIGIWHIST: Utilization

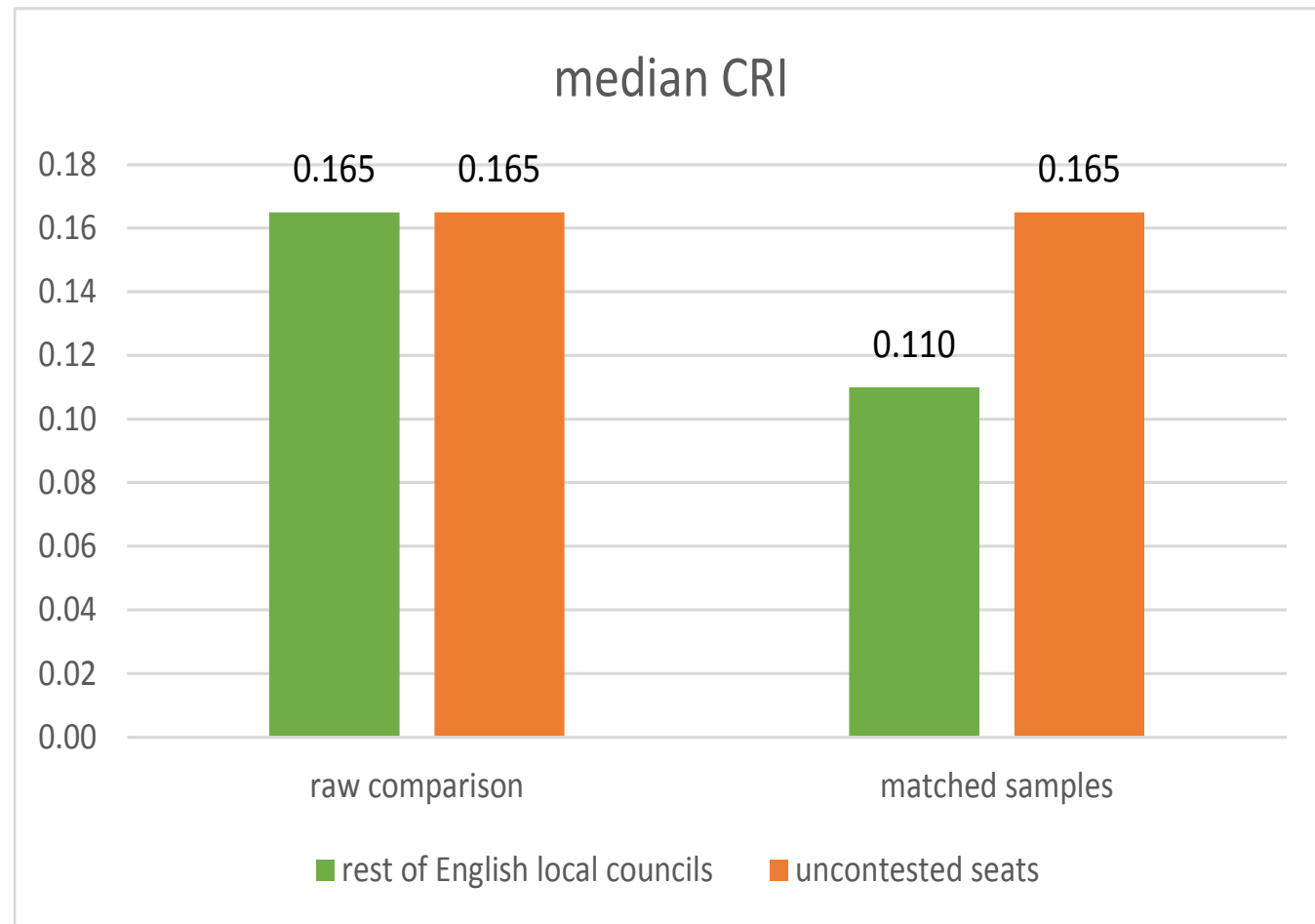
- Data download
- Web portals, mobile apps
- Whistleblower reporting: Big Data+user info
- Risk assessment software for public servants

# 3 brief examples

1. Electoral accountability in the UK
2. Mafia infiltration
3. Court effectiveness

# Electoral accountability

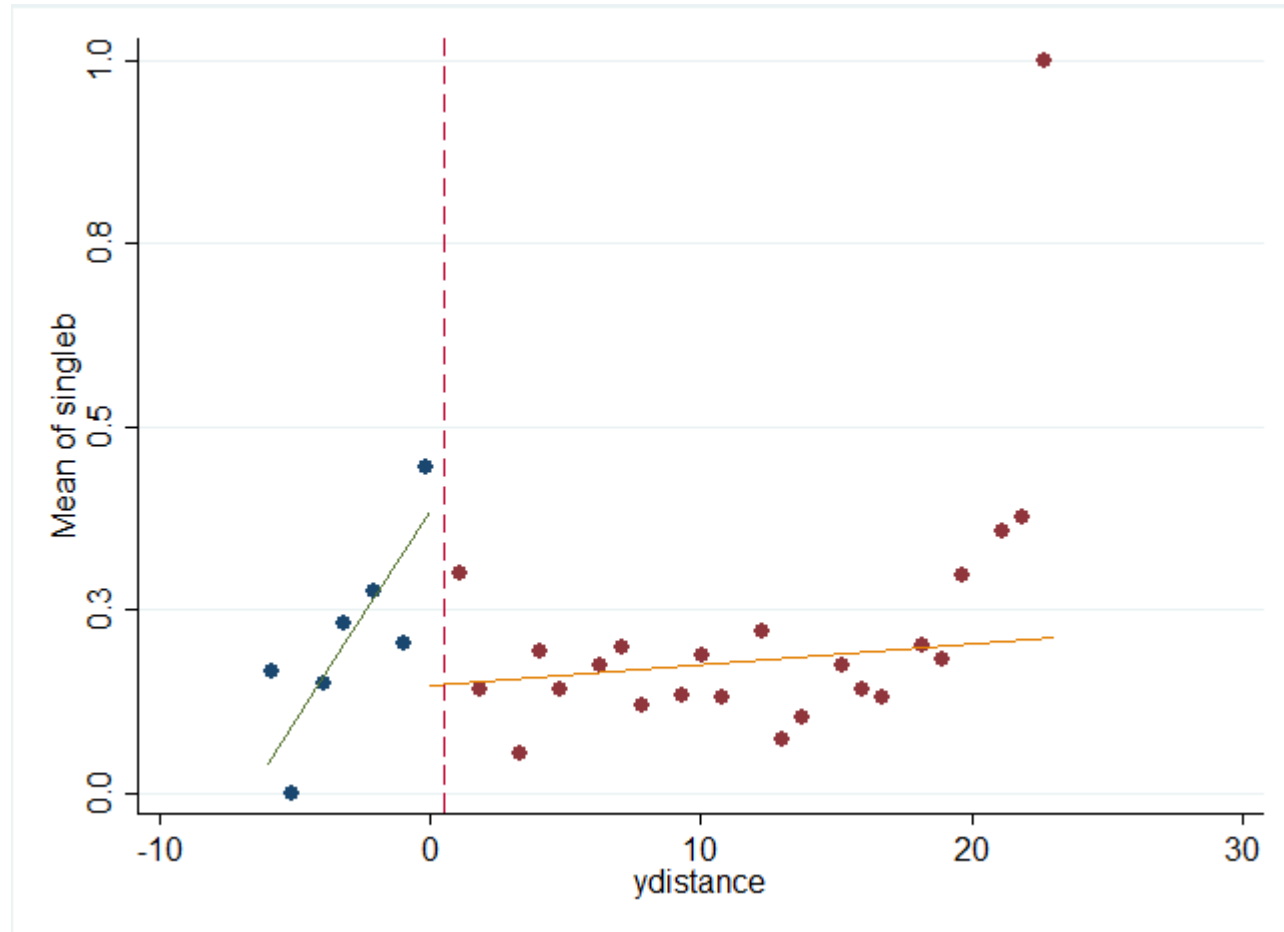
- Lack of electoral competition can lead to corruption even in well governed countries
- UK





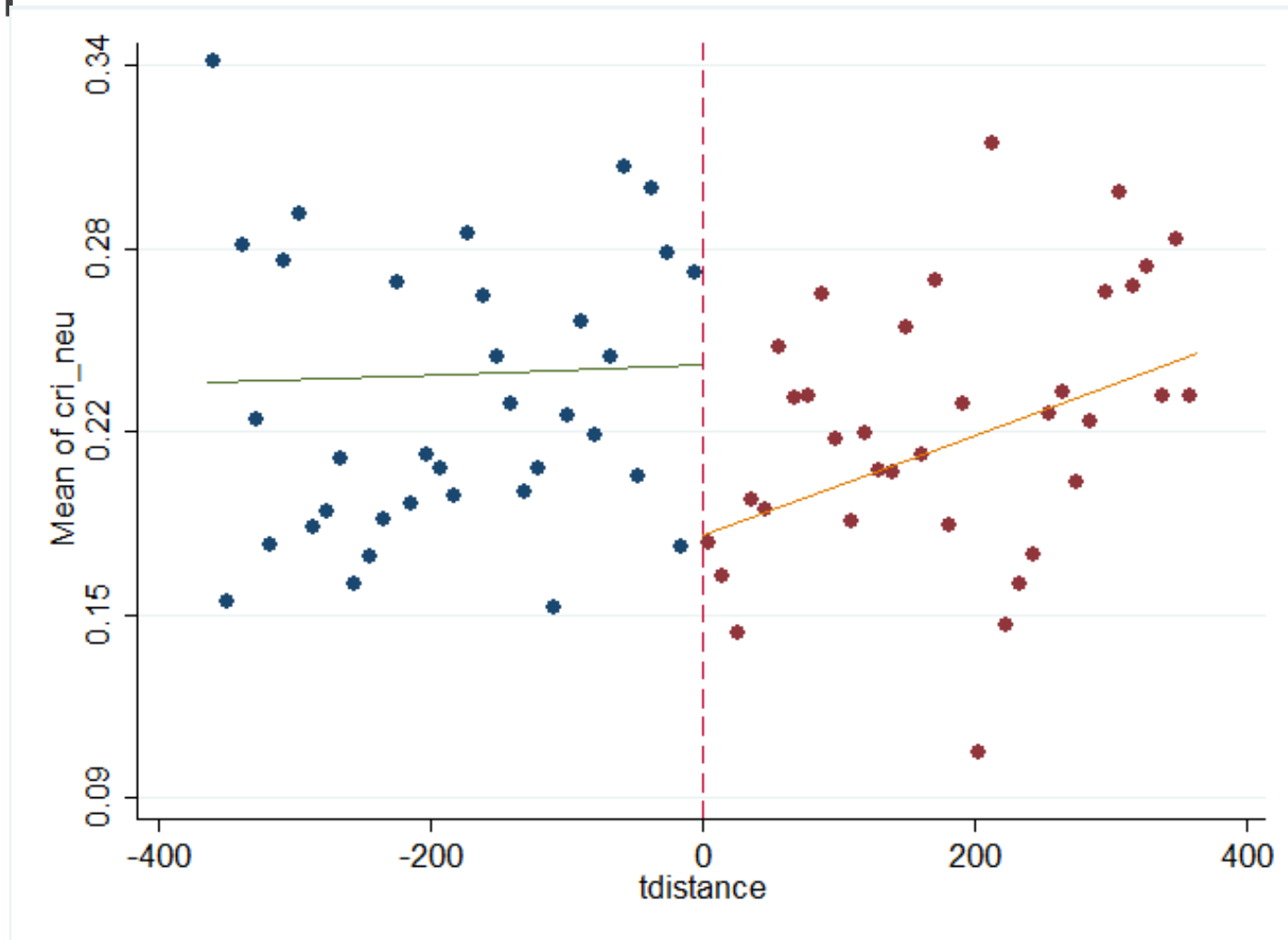
# Mafia infiltration in Italy

- Municipalities dissolved due to mafia infiltration
- Italy



# ECJ as anticorruption actor

Corruption risks in markets judged by ECJ



2015.09.28.

# Areas of collaboration

- Methodological
  - text mining,
  - machine learning
- Sectoral/thematic research
  - Local service provision
  - Education, health, etc.
- Geographical extensions outside Europe
  - Brazil, US, Russia, etc.
- Lobbying governments to better manage admin data

# Furter information

[digiwhist.eu](http://digiwhist.eu)