Everything I knew about protests was wrong

Martin Bekker
Centre for Social Change, UJ

martinjbekker@gmail.com

About

Philosophy, Peace Studies & Conflict Resolution, Development
- not computational

Bafokeng / RBA - not academic

Anthropology, Sociology - did not anticipate coding ante hoc

Theories of protest

Multitude of protests

Understood as signal of social distress, desperation // emancipatory potential

Urban

Conflicting

Desperation (Income poverty, unemployment // Deprivation //

Delivery (Services, Governance) // Democracy // Diplomas //

Demographics (Gender, Age, Foreigners) // Designation //

Data

Case studies

Media data

Administrative data

- Police Data (IRIS)
- Very good and very bad

Methods (1 of 3)

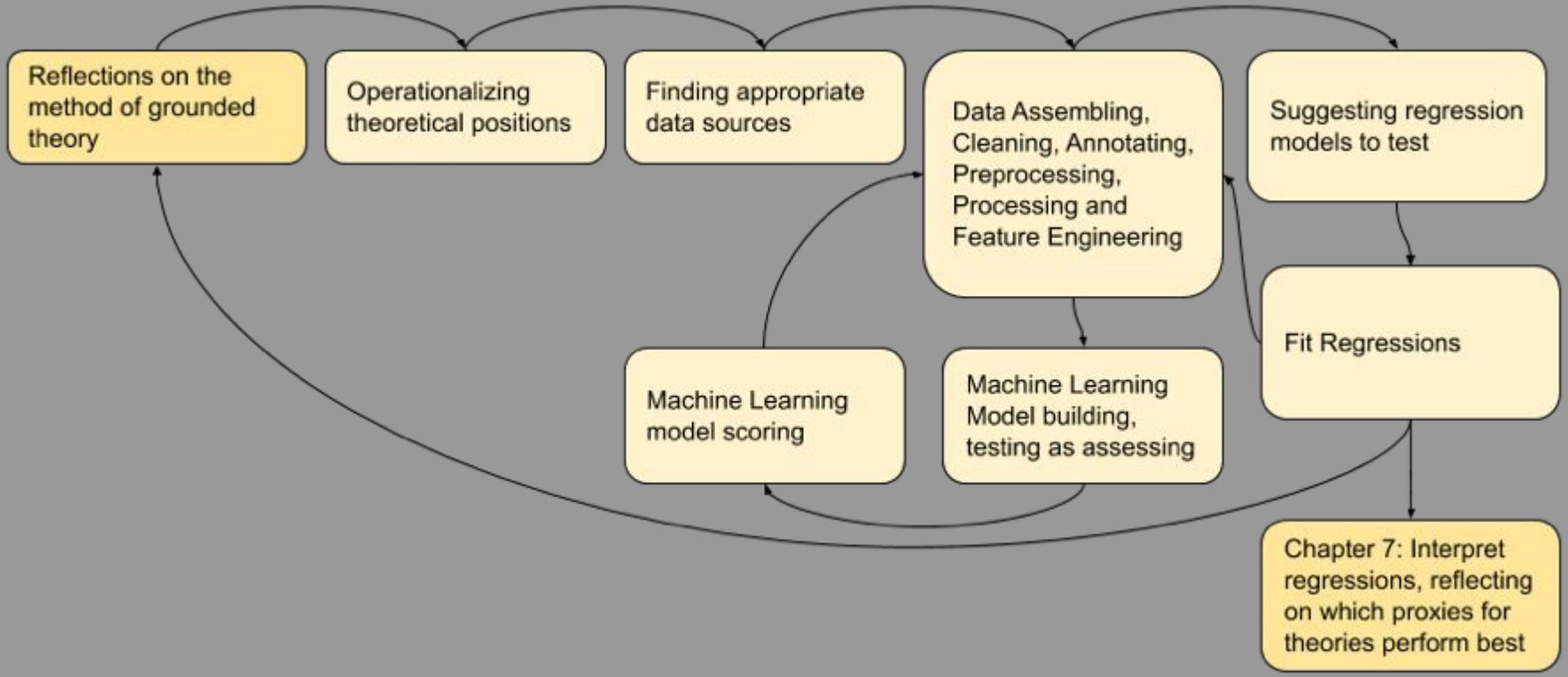
Combining and cleaning

Script to extract crowd sizes

Reading and theming

Coding (manual classification)

Keyword Modelling (ontology)


Reflections on the method of grounded theory
Operationalizing theoretical positions
Finding appropriate data sources
Data Assembling, Cleaning, Annotating, Preprocessing, Processing and Feature Engineering
Suggesting regression models to test
Fit Regressions
Machine Learning Model building, testing as assessing
Machine Learning model scoring
Chapter 7: Interpret regressions, reflecting on which proxies for theories perform best

Methods (2 of 3)

Basics of machine learning

- Labeled data // Classification task // No free lunch
- Train, test (validation)
- Scoring

Allocating to municipalities

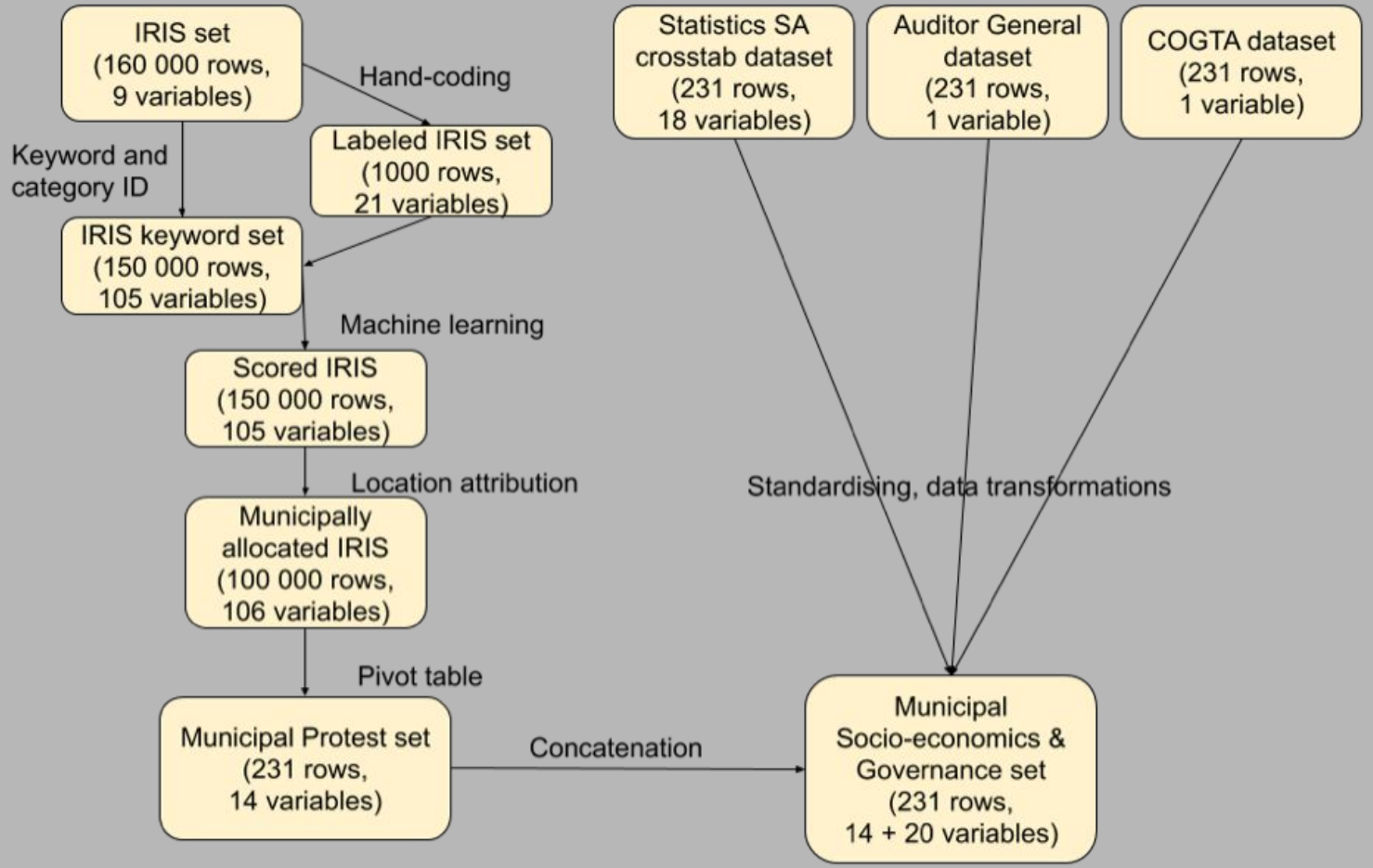
Methods (3 of 3)

Additional data: Statistics South Africa // Cogta // Auditor-General Reports

Municipal frames

Count vs Count per capita vs “General Propensity”

Regressions


IRIS set
(160 000 rows,
9 variables)
Hand-coding
Keyword and
category ID
Labeled IRIS set
(1000 rows,
21 variables)
IRIS keyword set
(150 000 rows,
105 variables)
Machine learning
Scored IRIS
(150 000 rows,
105 variables)
Location attribution
Municipally
allocated IRIS
(100 000 rows,
106 variables)
Pivot table
Municipal Protest set
(231 rows,
14 variables)
Concatenation
Statistics SA
crosstab dataset
(231 rows,
18 variables)
Auditor General
dataset
(231 rows,
1 variable)
COGTA dataset
(231 rows,
1 variable)
Standardising, data transformations
Municipal
Socio-economics &
Governance set
(231 rows,
14 + 20 variables)

Results (Descriptive)

Frequency

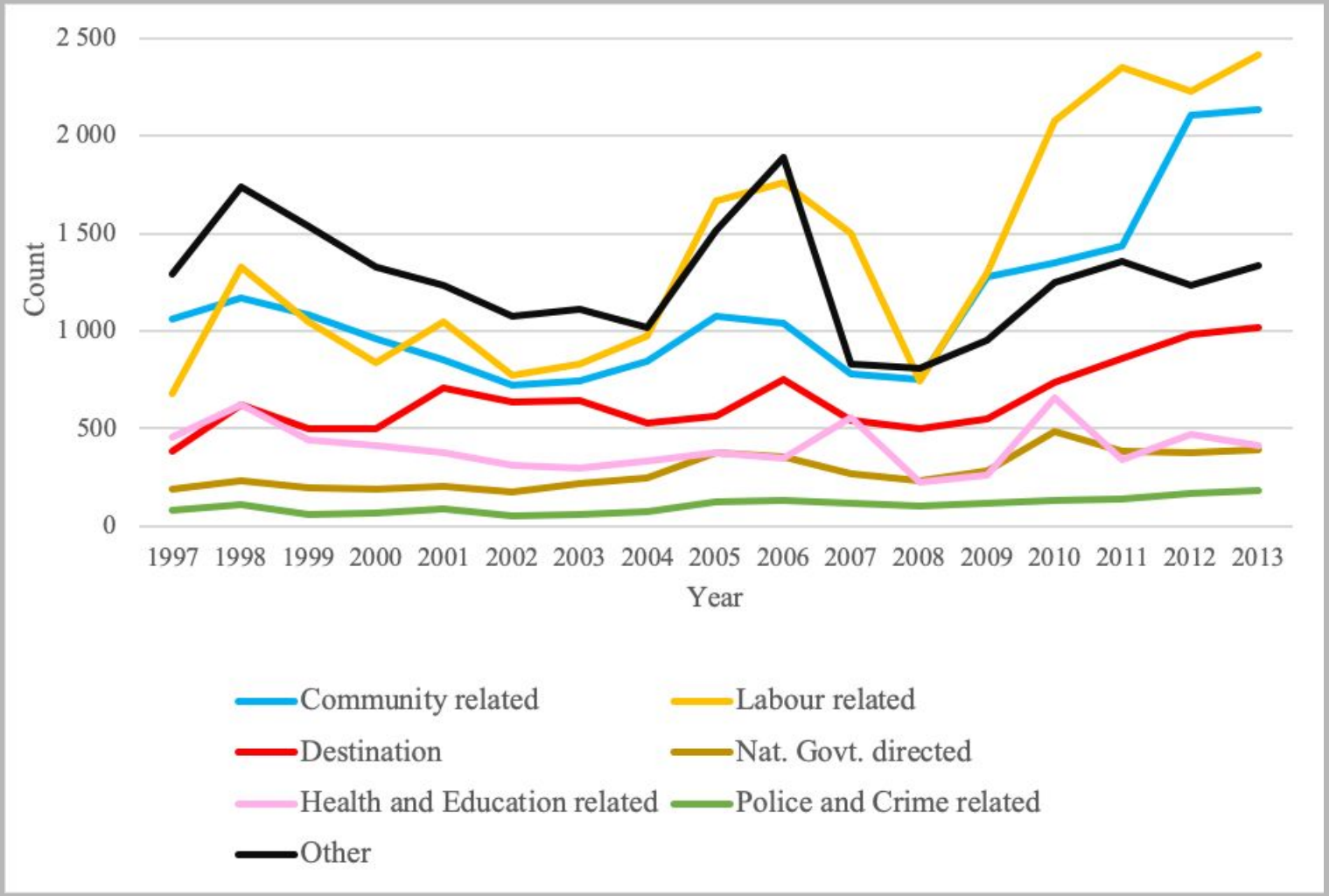
Timeline

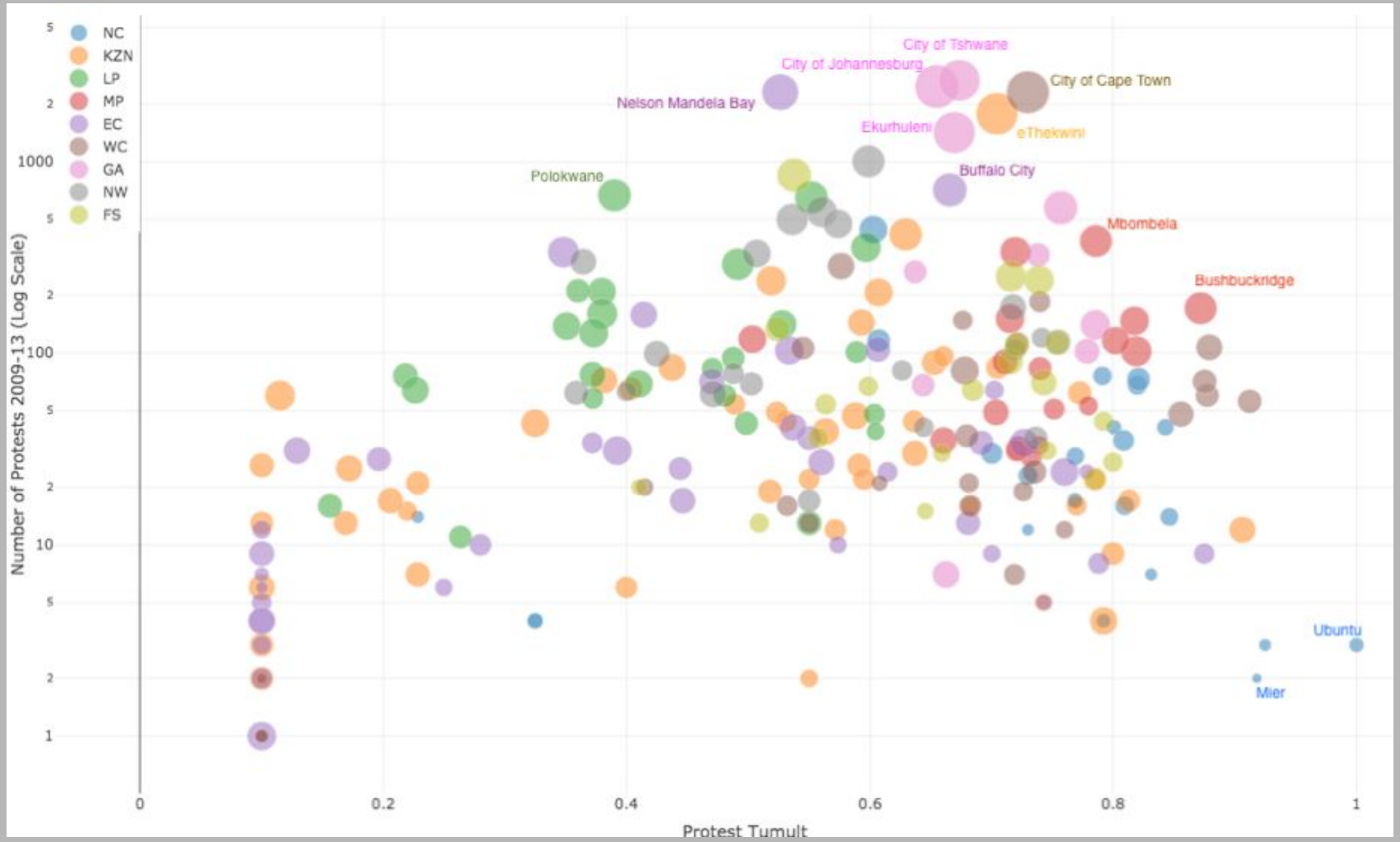
Types of protest

Tumult

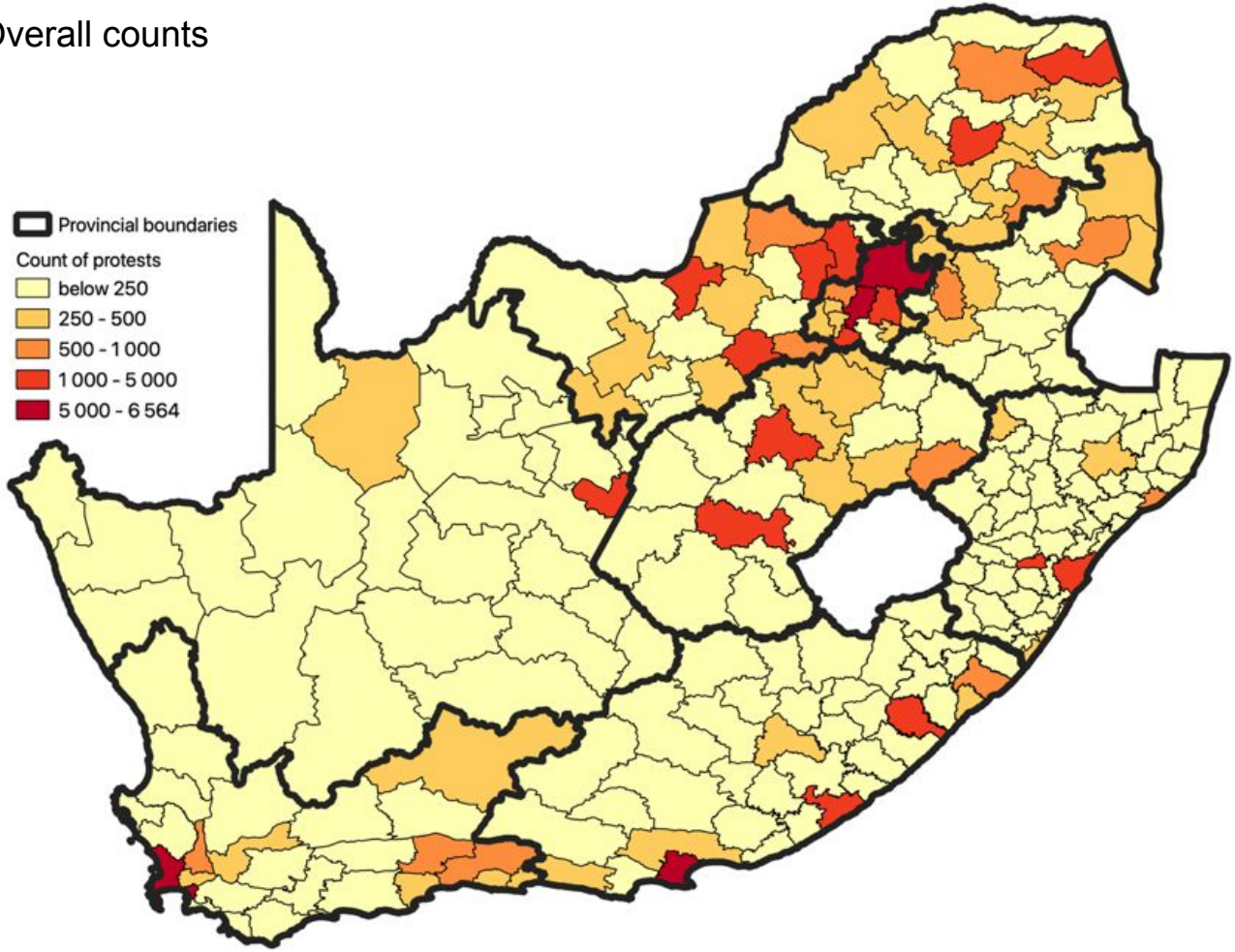
Crowd-size

Rurality

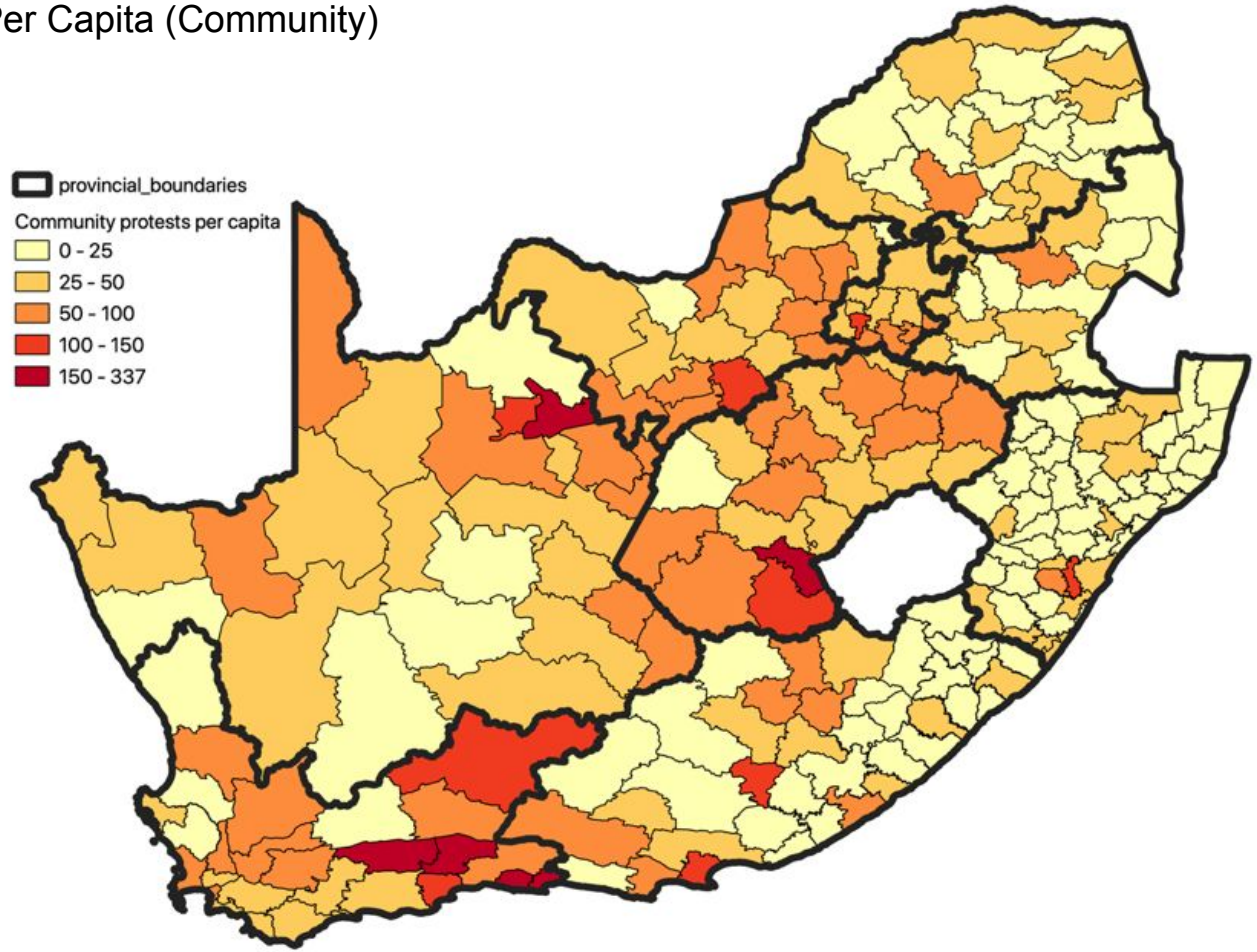

2 500
2 000
1 500
1 000
500
0
Count
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
Year
Community related
Labour related
Destination
Nat. Govt. directed
Health and Education related
Police and Crime related
Other


NC
KZN
LP
MP
EC
WC
GA
NW
FS
City of Tshwane
City of Johannesburg
City of Cape Town
Nelson Mandela Bay
Ekurhuleni
eThekwini
Buffalo City
Polokwane
Mbombela
Bushbuckridge
Ubuntu
Mier
Number of Protests 2009-13 (Log Scale)
Protest Tumult
0
0.2
0.4
0.6
0.8
1
1000
100
10
5
2
1

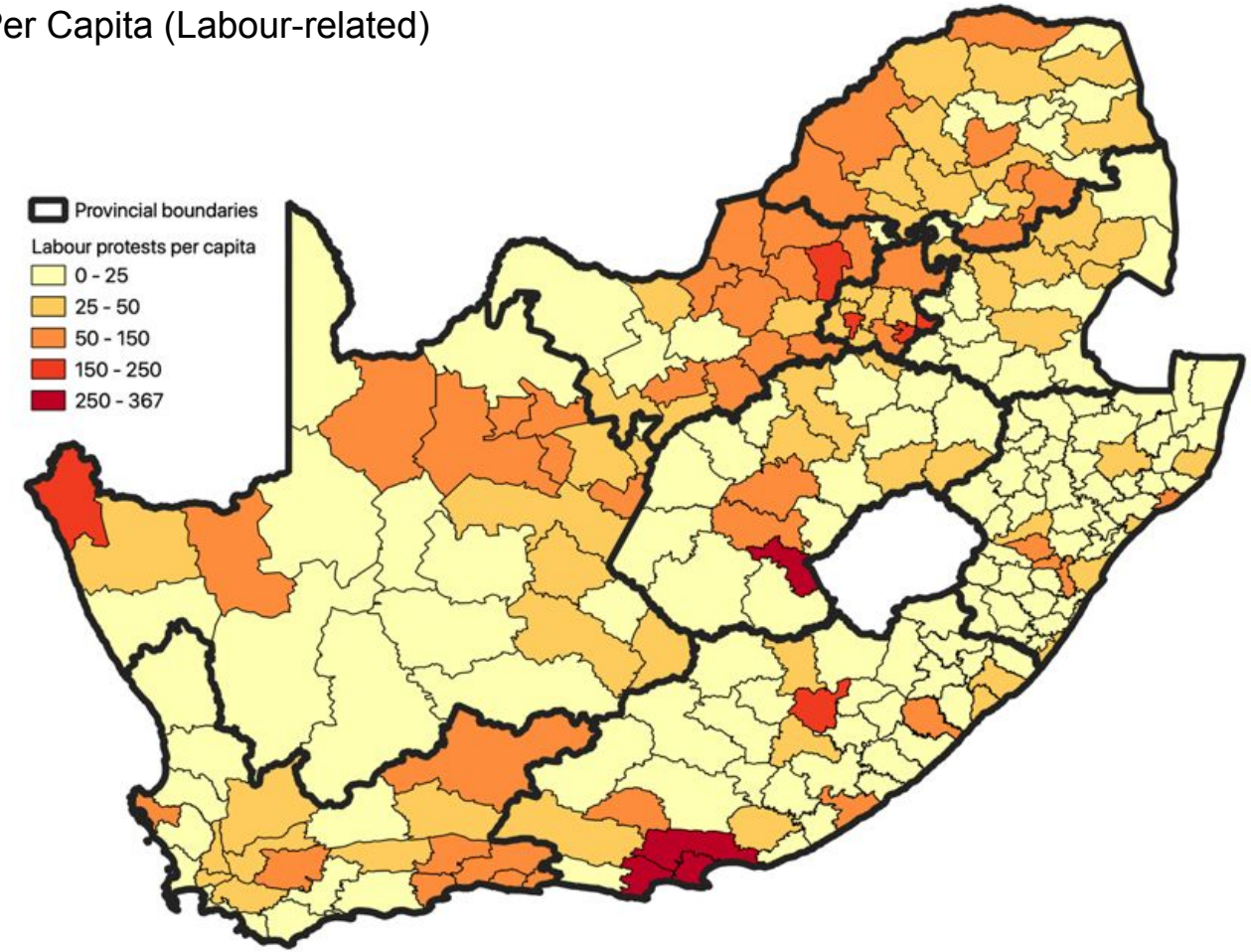
Overall counts



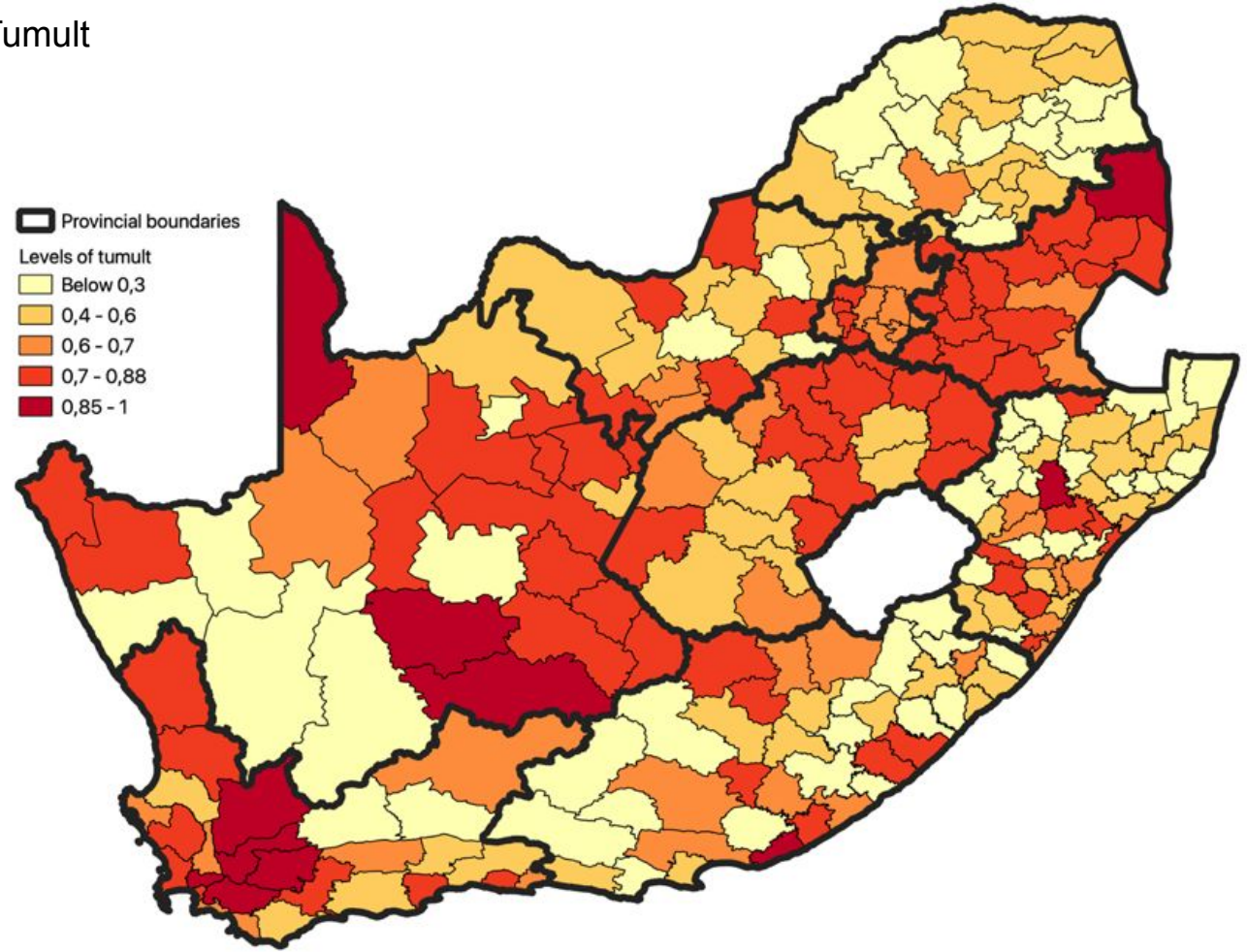
Per Capita (Community)



Per Capita (Labour-related)



Tumult


Provincial boundaries
Levels of tumult
Below 0,3
0,4 - 0,6
0,6 - 0,7
0,7 - 0,88
0,85 - 1

Results (Regressions)

Rejects:

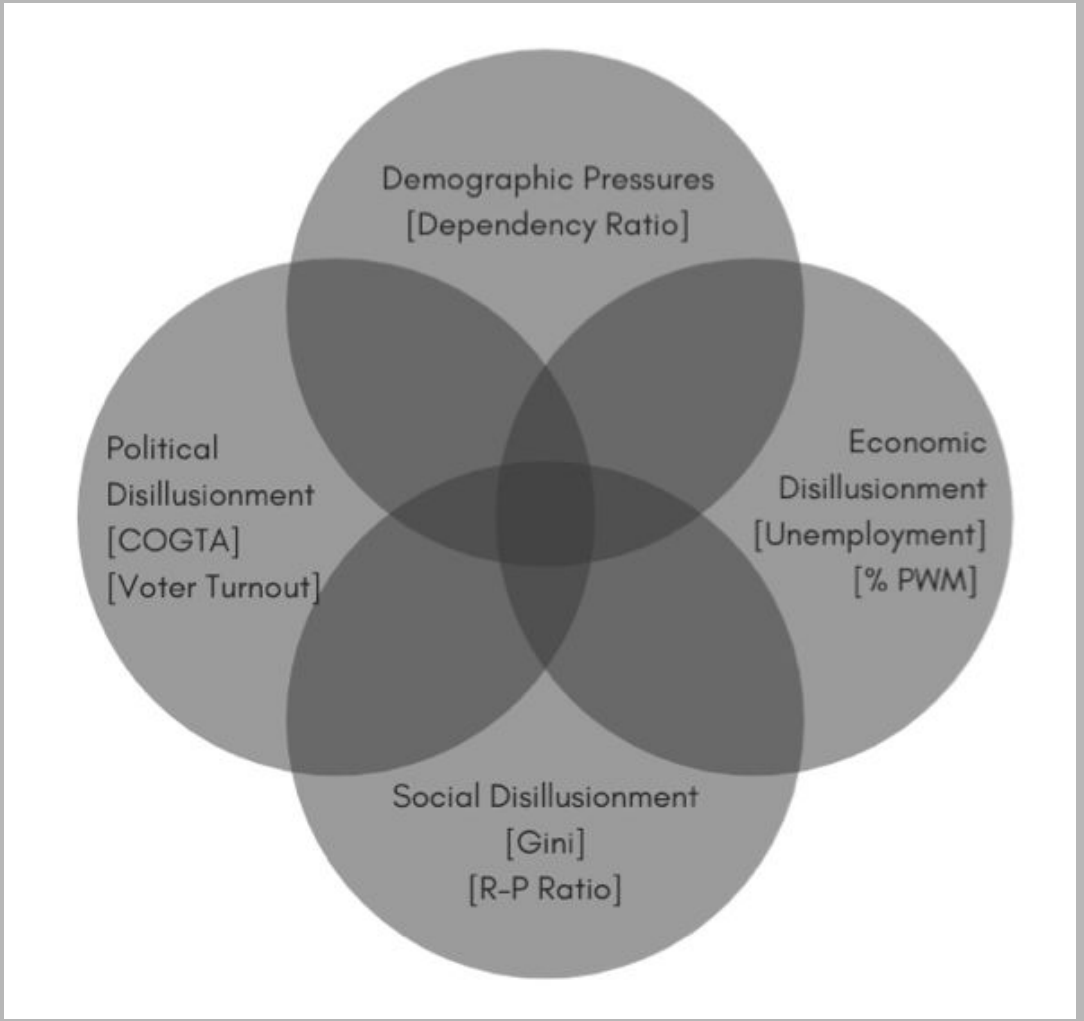
Single-factor theories

Role of foreigners (resource strain, etc) // Services // Informality // Poverty

Corroborates:

Inequality // Unemployment // Political disaffection //

Demographic-education profile


Demographic Pressures
[Dependency Ratio]
Political
Disillusionment
[COGTA]
[Voter Turnout]
Economic
Disillusionment
[Unemployment]
[% PWM]
Social Disillusionment
[Gini]
[R-P Ratio]

Resources

Free courses, online (Edx, Datacamp, Coursera, FastAI, Youtube)

Free software and platforms (Open office, Python IDEs, QGIS, IEEE, Github)

Summary video (of me, sorry): https://www.youtube.com/watch?v=YwwEsajZpRU

Responses

Please answer three questions at https://www.surveymonkey.com/r/LDBSD9S