Estimating Inequality in South Africa with Remote Sensing Data

Sebastian Krantz

Kiel Institute for the World Economy

18th April 2023

Table of Contents

Introduction

- Exploratory work examining the potential of big data sources, in particular remote sensing data, to improve the spatial and temporal resolution of inequality estimates for South Africa.

- Galimberti et al. (2021) estimate nightlights-based Gini-coefficients for 234 countries/territories from 1992 to 2013 $\rightarrow$ significantly correlated with income inequality across countries, but capturing different time dynamics (hypothesis: consumption, informal activities, infrastructure, wealth).

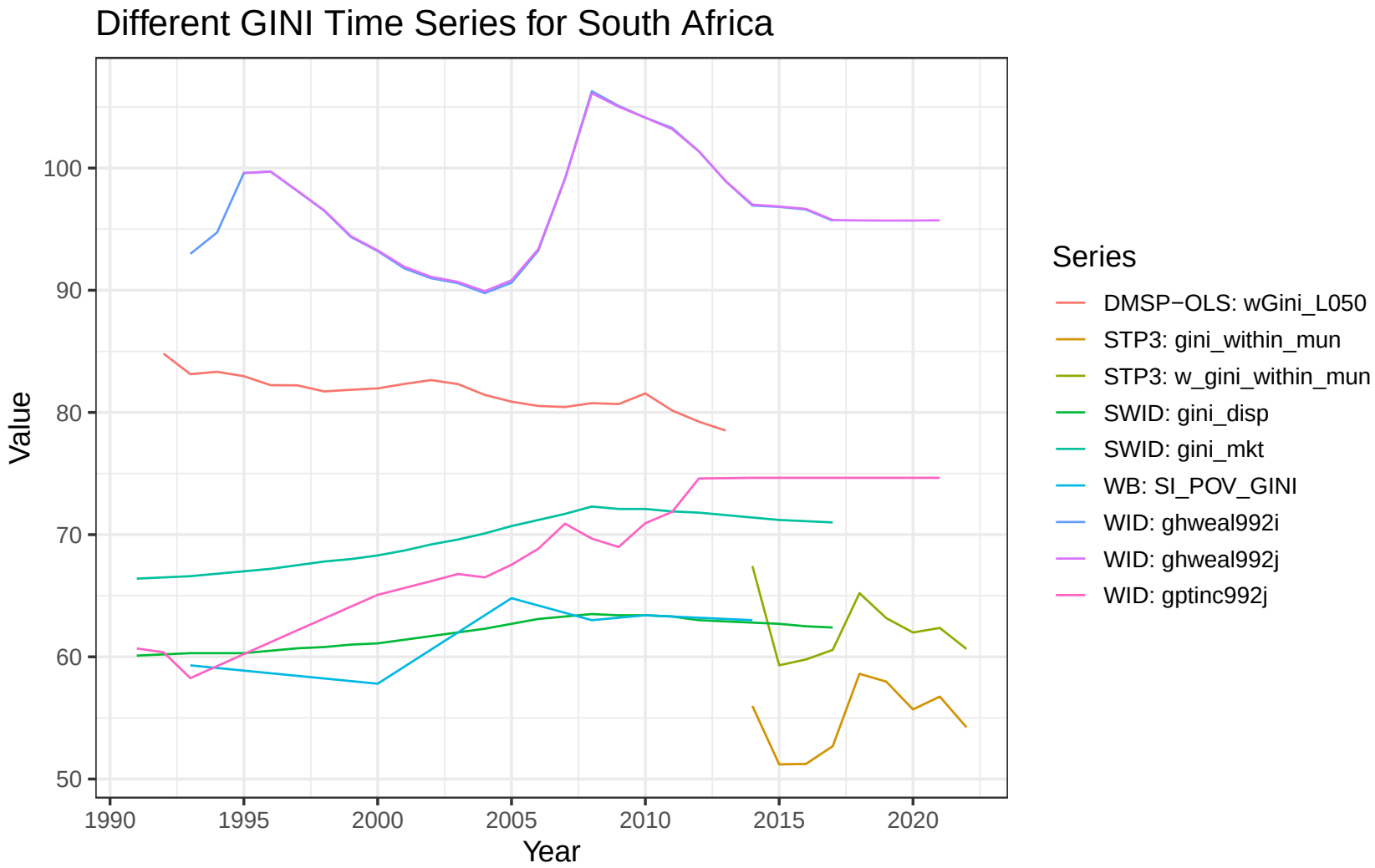
- Lee & Braithwaite (2022) and Chi et al. (2022) use rich data sources (daytime satellite imagery, Open Street Map, Facebook connectivity data) to generate high-resolution wealth estimates (indices) for Sub-Saharan African countries.

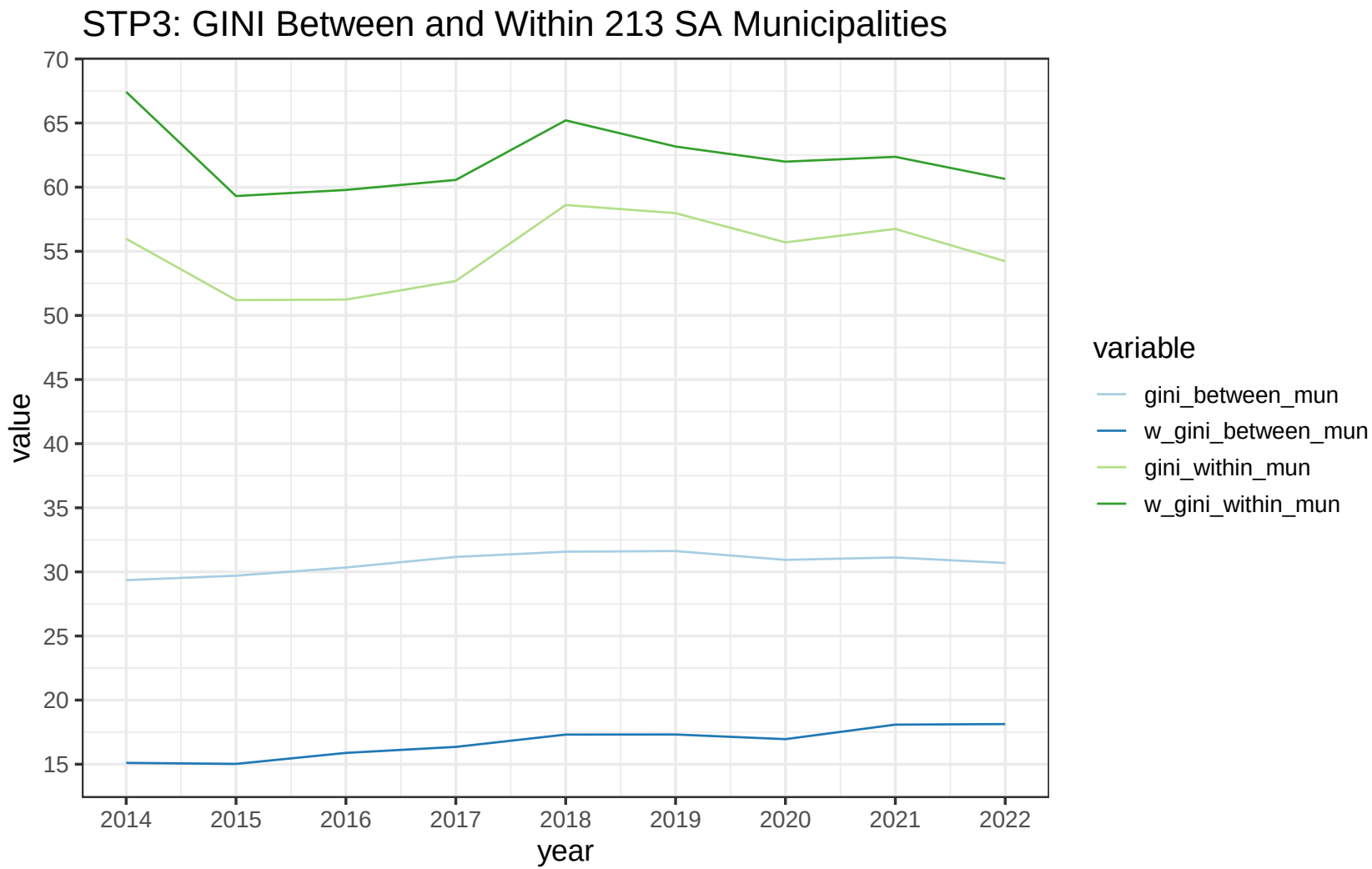
- Spatial Tax Panel V3 (2014-2022) (Nell & Visagie, 2023): spatial inequality in South Africa based on tax records.

Recent Report (2019) *Inequality Trends in South Africa: A multidimensional diagnostic of inequality by Statistics South Africa*

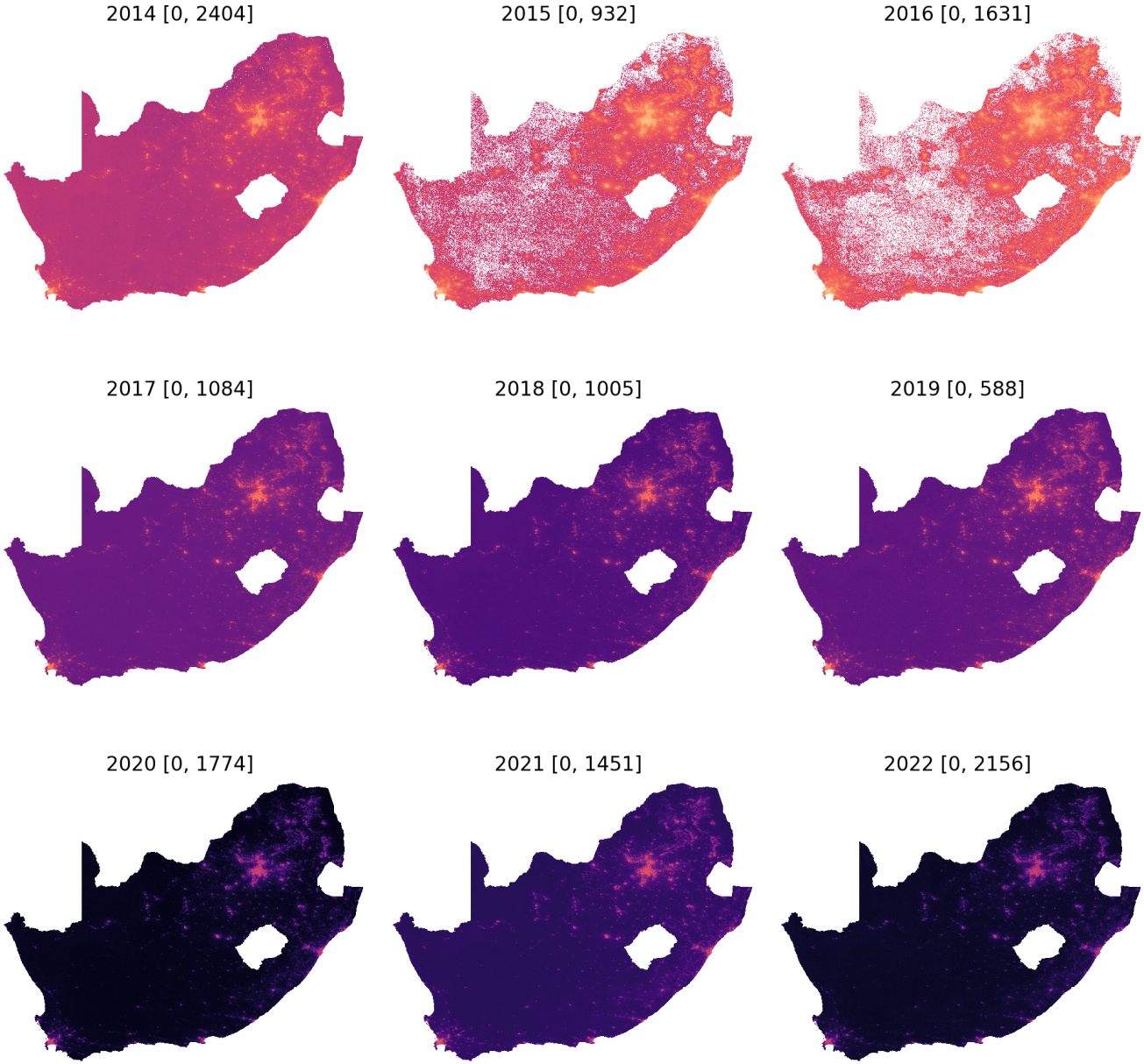
- Most measures suggest decrease in within-group inequality between 2006 and 2015. All provinces, except for Limpopo and Eastern Cape, reported a decrease in Gini coefficients.
- Asset inequality has decreased over time for all provinces, except for Northern Cape and Free State.
- Inequality decreased for Indians/Asians and whites, remained fairly constant for coloureds, increased for black Africans.
- Labour market income is the main driver of income inequality in South Africa, contributing 74.2% towards overall income inequality in the country in 2015.
- School attendance increased between 2002 and 2017, with beneficial effects on inequality. Health inequality remains.

Time Series Estimates: National



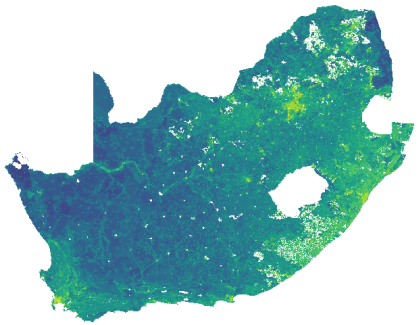

STP3: GINI Between and Within 213 SA Municipalities
value
year
variable
gini_between_mun
w_gini_between_mun
gini_within_mun
w_gini_within_mun
70
65
60
55
50
45
40
35
30
25
20
15
2014
2015
2016
2017
2018
2019
2020
2021
2022

VIIRS DNB Nightlights V1

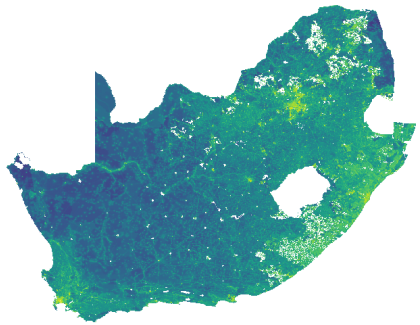


WorldPop Population (UN Adjusted)

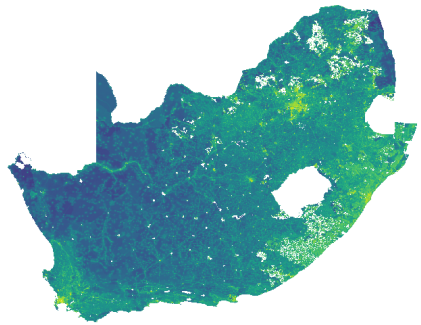
2014 [0, 51382]



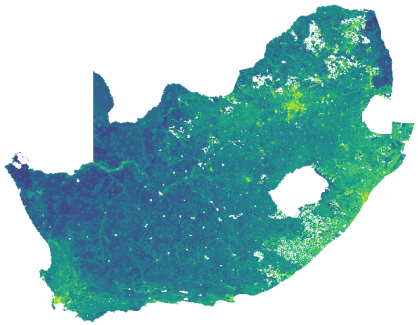
2015 [0, 53015]



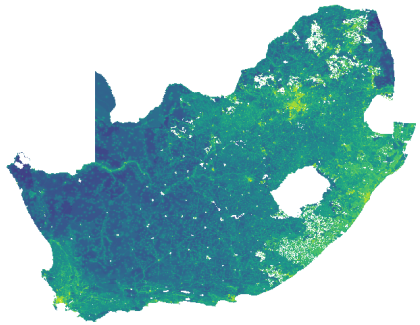
2016 [0, 54531]



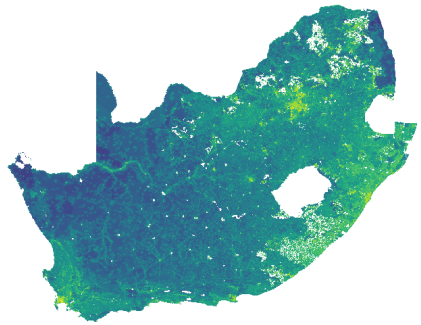
2017 [0, 56181]



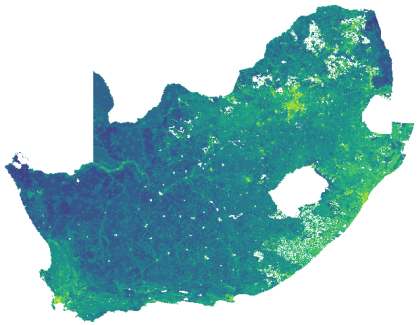
2018 [0, 57838]

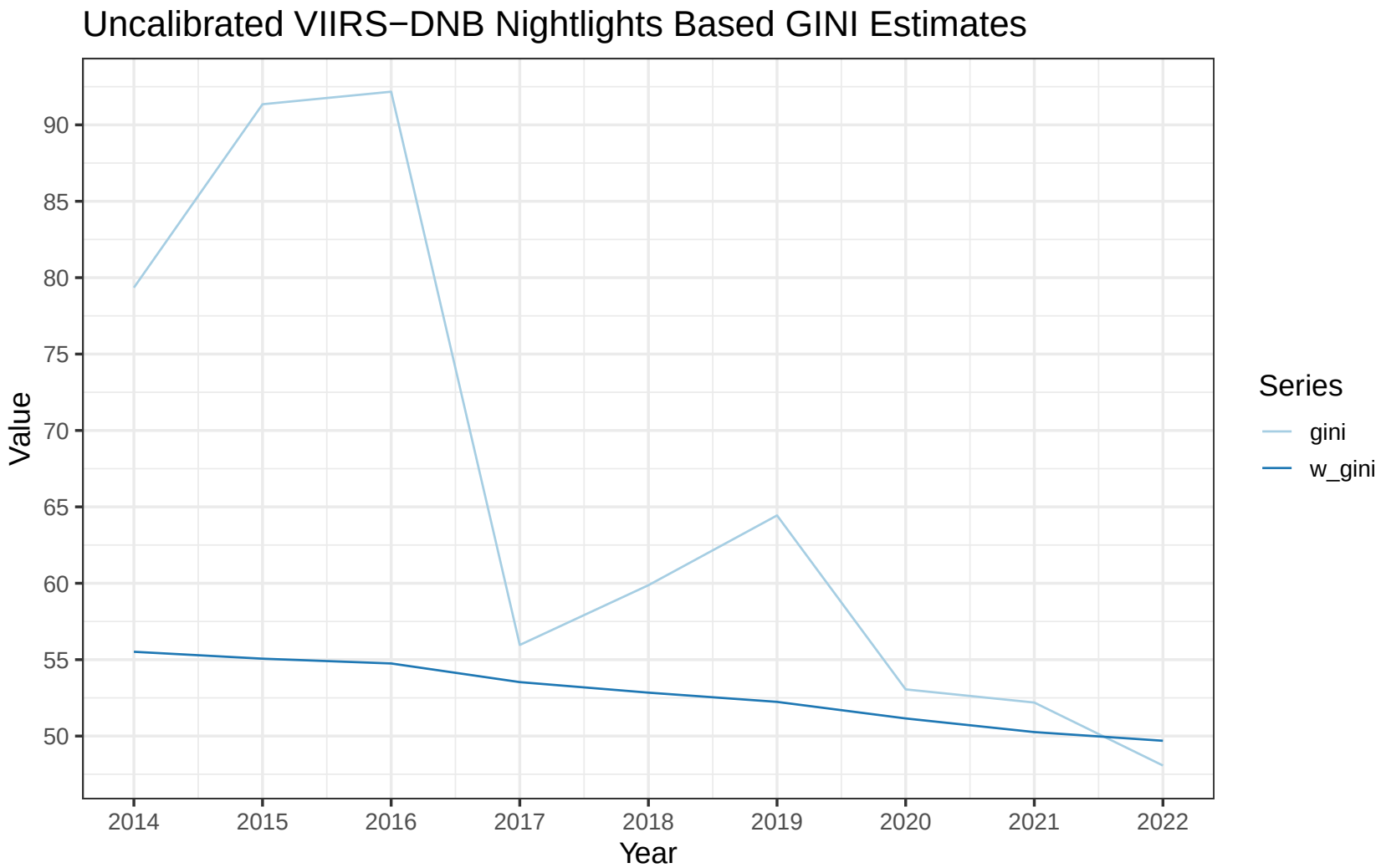


2019 [0, 59471]



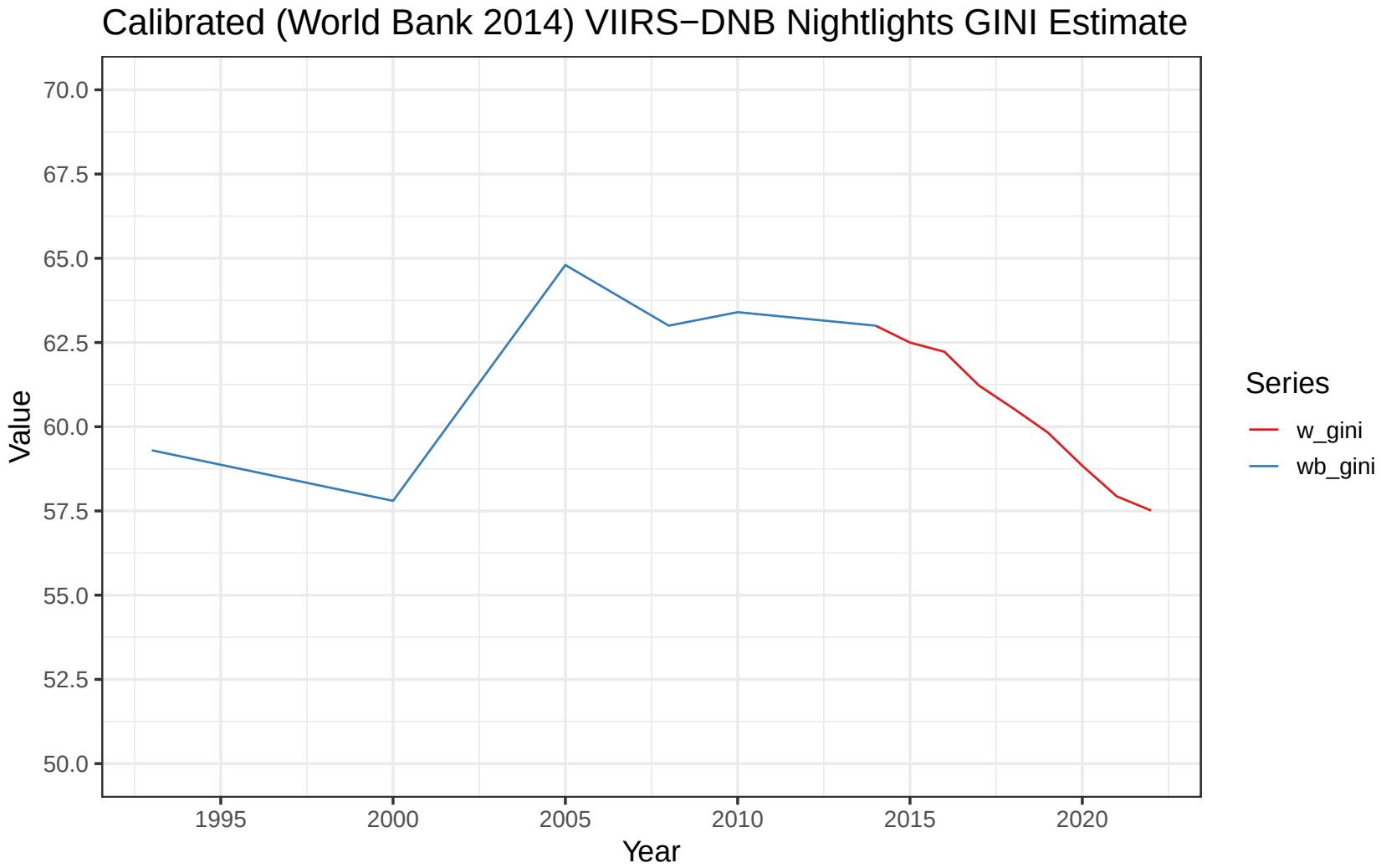
2020 [0, 61069]





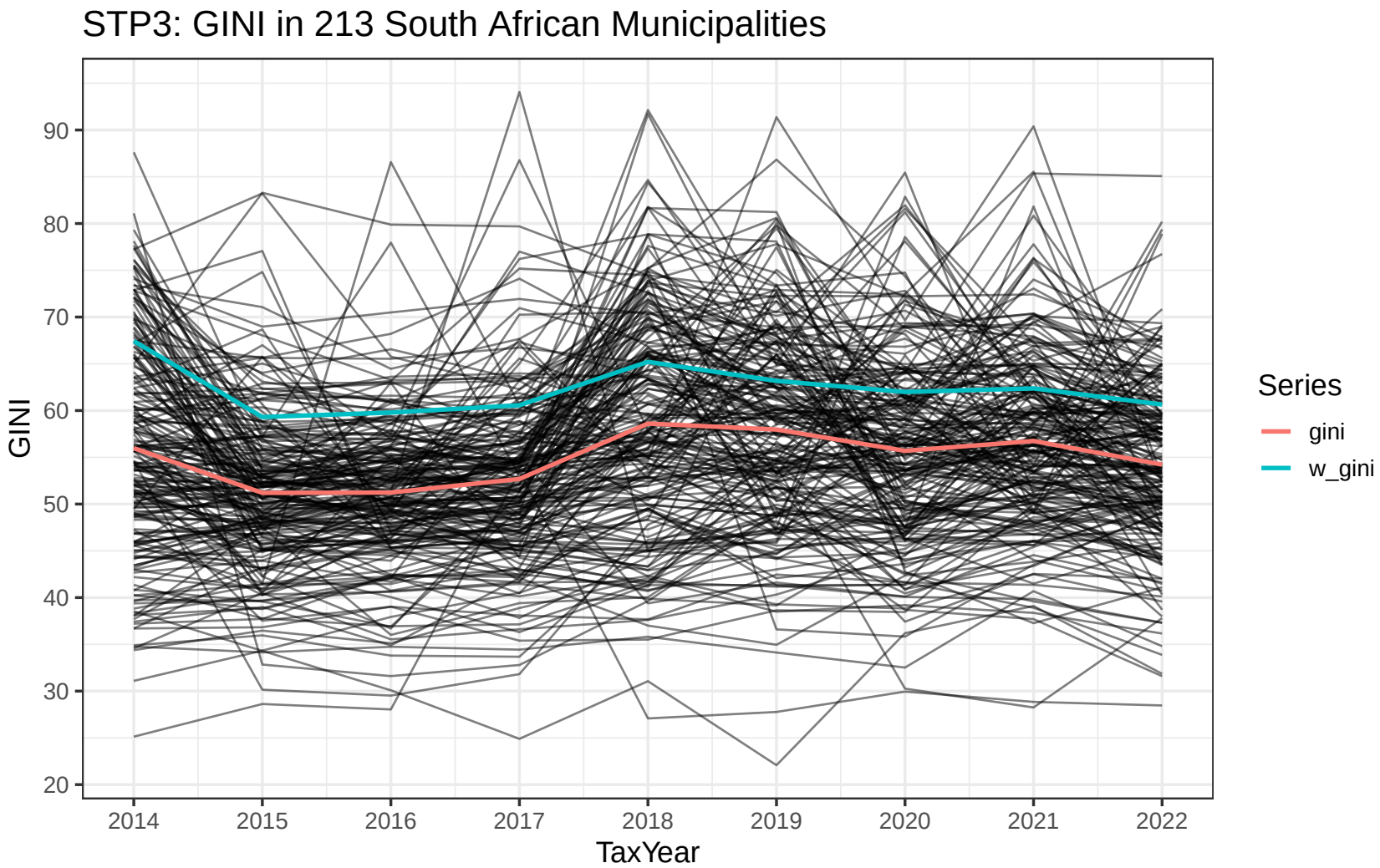
Notes: w_gini is weighted by population[1], gini just uses nightlights. Fluctuations in nightlights are greater in unpopulated areas.

[1] I forecast population in 2021 and 2022 using cell-level linear-time models.

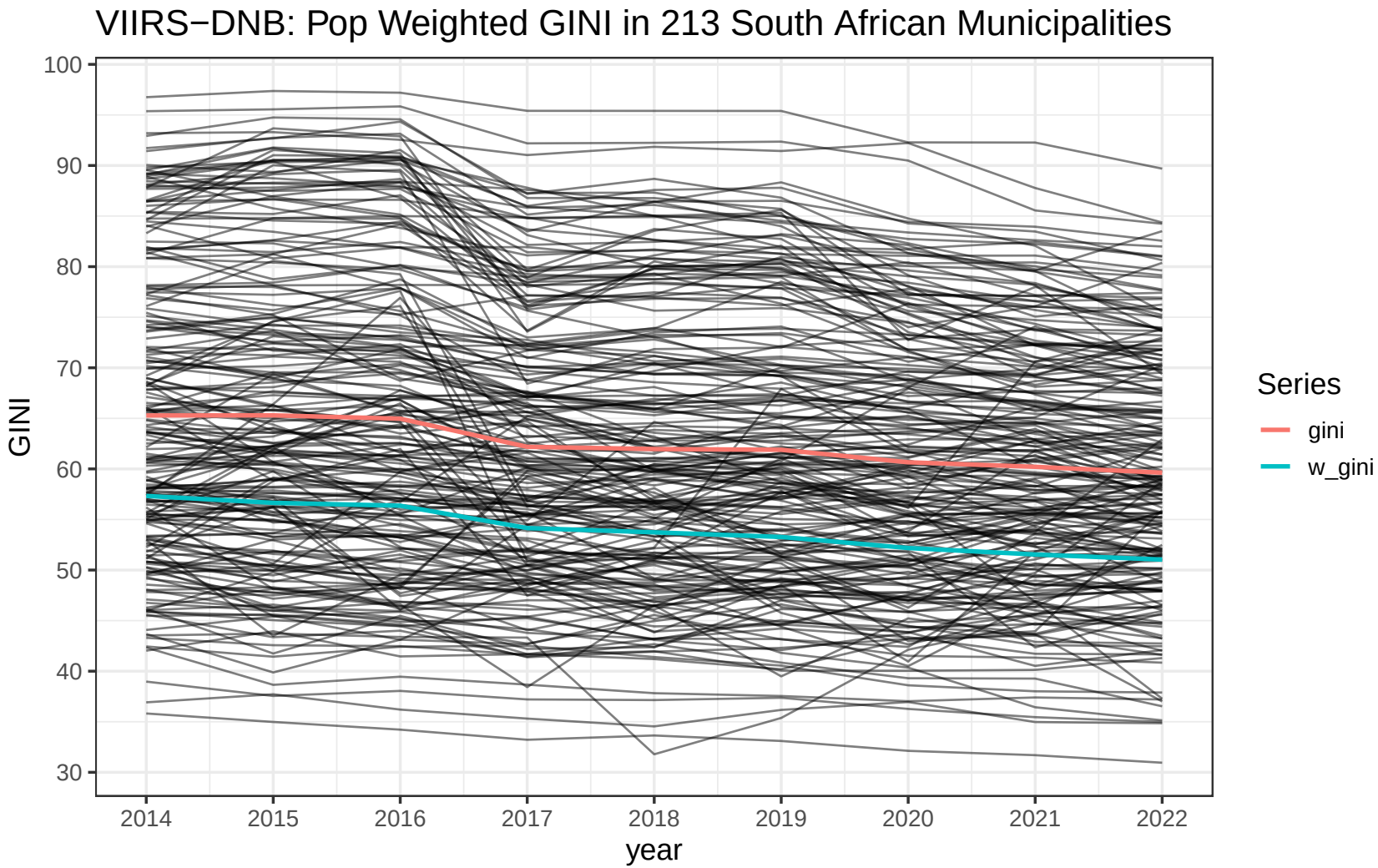


Notes: Calibrated such that `w_gini(NL14`^{k}`, POP14) = 0.63`, the latest World Bank estimate. This is the case for $k = 1.31$.

Time Series Estimates: Municipal



Notes: w_gini here refers to a pop-weighted average of municipal time series.



Notes: w_gini here refers to a pop-weighted average of municipal time series. The time series itself are obtained by a weighted GINI estimation using granular nightlights and population data in each municipality, calibrated using k from the 2020 municipal calibration exercise (see slide 24).

Table: Correlation of STP3 and Nightlights Based Municipal GINIs

Trans	Variable	med_inc	FTE	pop	avg_rad	gini	nl_gini	nl_w_gini
Overall	med_inc	1						
Between	med_inc	1						
Within	med_inc	1						
Overall	FTE	-.006	1					
Between	FTE	-.006	1					
Within	FTE	-.246*	1					
Overall	pop	.038	.964*	1				
Between	pop	.039	.965*	1				
Within	pop	.086*	.204*	1				
Overall	avg_rad	.015	.907*	.910*	1			
Between	avg_rad	.014	.911*	.918*	1			
Within	avg_rad	.157*	.354*	.364*	1			
Overall	gini	-.491*	.171*	.155*	.198*	1		
Between	gini	-.668*	.235*	.214*	.261*	1		
Within	gini	-.078*	.108*	.064*	.222*	1		
Overall	nl_gini	-.002	.022	.071*	.033	.026	1	
Between	nl_gini	.018	.035	.111	.094	.246*	1	
Within	nl_gini	-.142*	-.331*	-.458*	-.839*	-.197*	1	
Overall	nl_w_gini	.199*	-.272*	-.260*	-.325*	-.253*	-.078*	1
Between	nl_w_gini	.215*	-.282*	-.269*	-.331*	-.336*	-.269*	1
Within	nl_w_gini	-.062*	-.210*	-.197*	-.481*	-.106*	.558*	1

Notes: * indicates significance at 5% level. There are 213 municipalities. *Between* data is averaged across years, *Within* data is standardized (scaled and centered) within each municipality. NL GINI's are calibrated using k from the 2020 municipal calibration exercise (see slide 24).

Restricting sample to 61 municipalities with population $>$ 200,000

Table: Correlation of STP3 and Nightlights Based Municipal GINIs

Trans	Variable	med_inc	FTE	pop	avg_rad	gini	nl_gini	nl_w_gini
Overall	med_inc	1						
Between	med_inc	1						
Within	med_inc	1						
Overall	FTE	-.143*	1					
Between	FTE	-.146	1					
Within	FTE	-.326*	1					
Overall	pop	-.145*	.970*	1				
Between	pop	-.149	.972*	1				
Within	pop	.052	.187*	1				
Overall	avg_rad	-.146*	.904*	.899*	1			
Between	avg_rad	-.151	.908*	.908*	1			
Within	avg_rad	.126*	.223*	.340*	1			
Overall	gini	-.704*	.340*	.331*	.373*	1		
Between	gini	-.851*	.410*	.401*	.443*	1		
Within	gini	-.084*	.061	.118*	.237*	1		
Overall	nl_gini	-.217*	-.298*	-.280*	-.335*	.101*	1	
Between	nl_gini	-.277*	-.375*	-.349*	-.412*	.209	1	
Within	nl_gini	-.086*	-.228*	-.598*	-.585*	-.175*	1	
Overall	nl_w_gini	.292*	-.441*	-.464*	-.518*	-.386*	-.180*	1
Between	nl_w_gini	.307*	-.449*	-.472*	-.527*	-.465*	-.327*	1
Within	nl_w_gini	-.028	-.187*	-.449*	-.490*	-.074	.741*	1

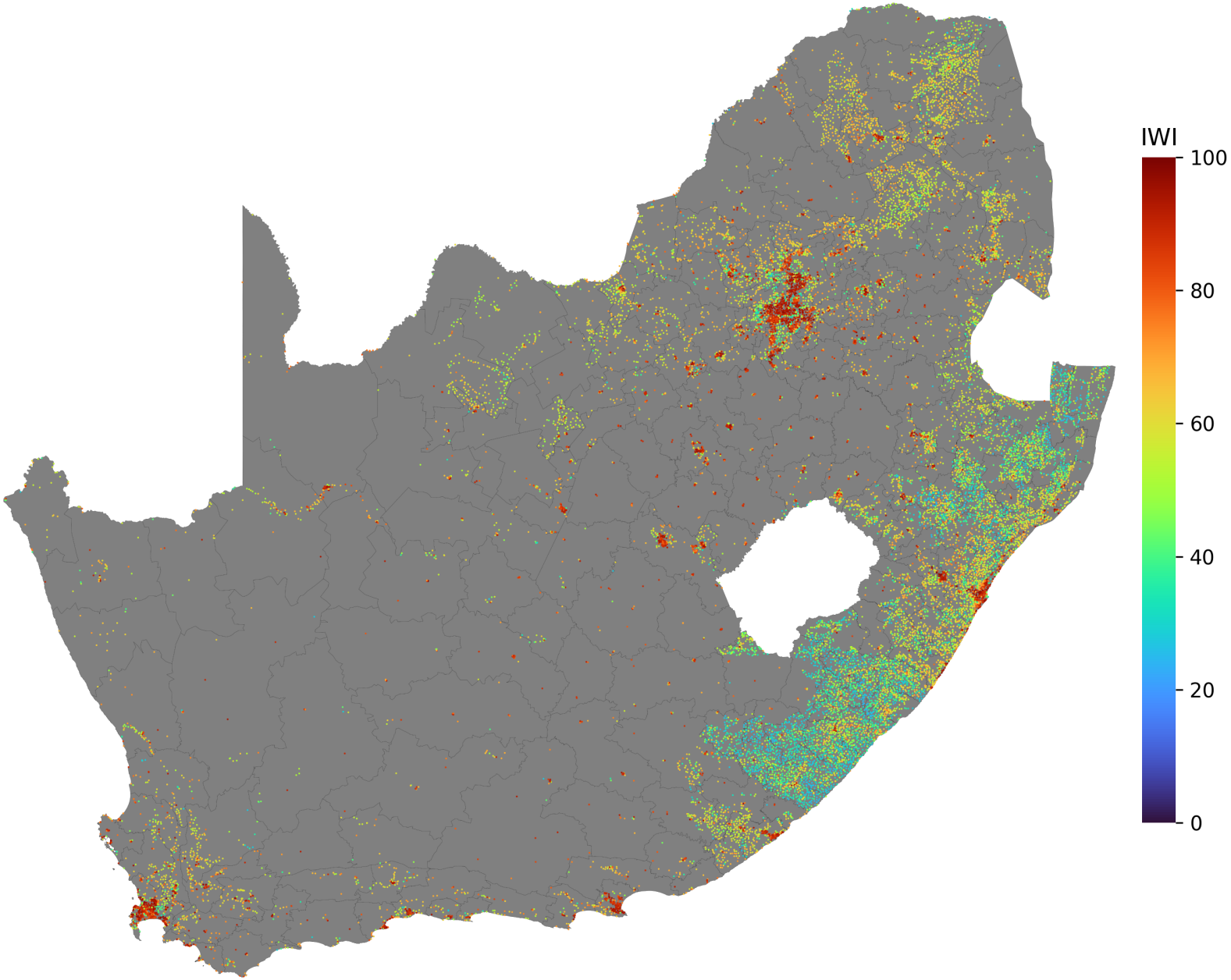
Notes: * indicates significance at 5% level. There are 61 municipalities included. *Between* data is averaged across years, *Within* data is standardized (scaled and centered) within each municipality. NL GINI's are calibrated using k from the 2020 municipal calibration exercise (see slide 24).

Geospatial Estimates

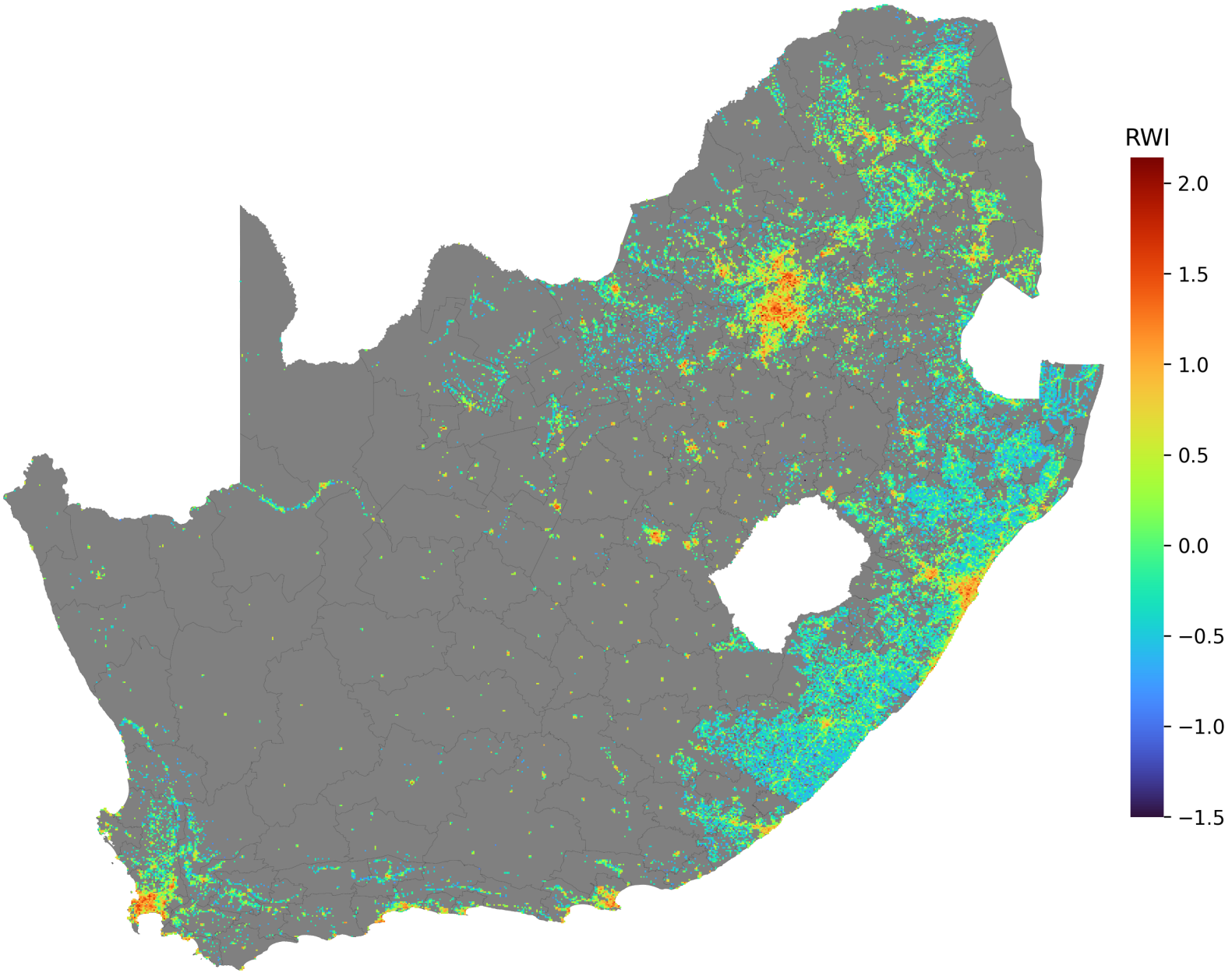
Based on Granular Data

- International Wealth Index by Lee & Braithwaite (2022) (calibrated to DHS data, using OSM and daylight satellite imagery, 1.6km resolution).
- Relative Wealth Index by Chi et al. (2022) (calibrated to LSMS data, daylight satellite imagery, maps, population, and facebook connectivity data, 2.4km resolution).
- VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1, 2020 median composite, 463.83m resolution.
- Spatial Tax Panel V3 (2014-2022) (Nell & Visagie, 2023): GINI estimates for 213 Municipalities + Shapefile, for the years 2014-2022. I use 2020 because last available year of highres (WorldPop) population layer.

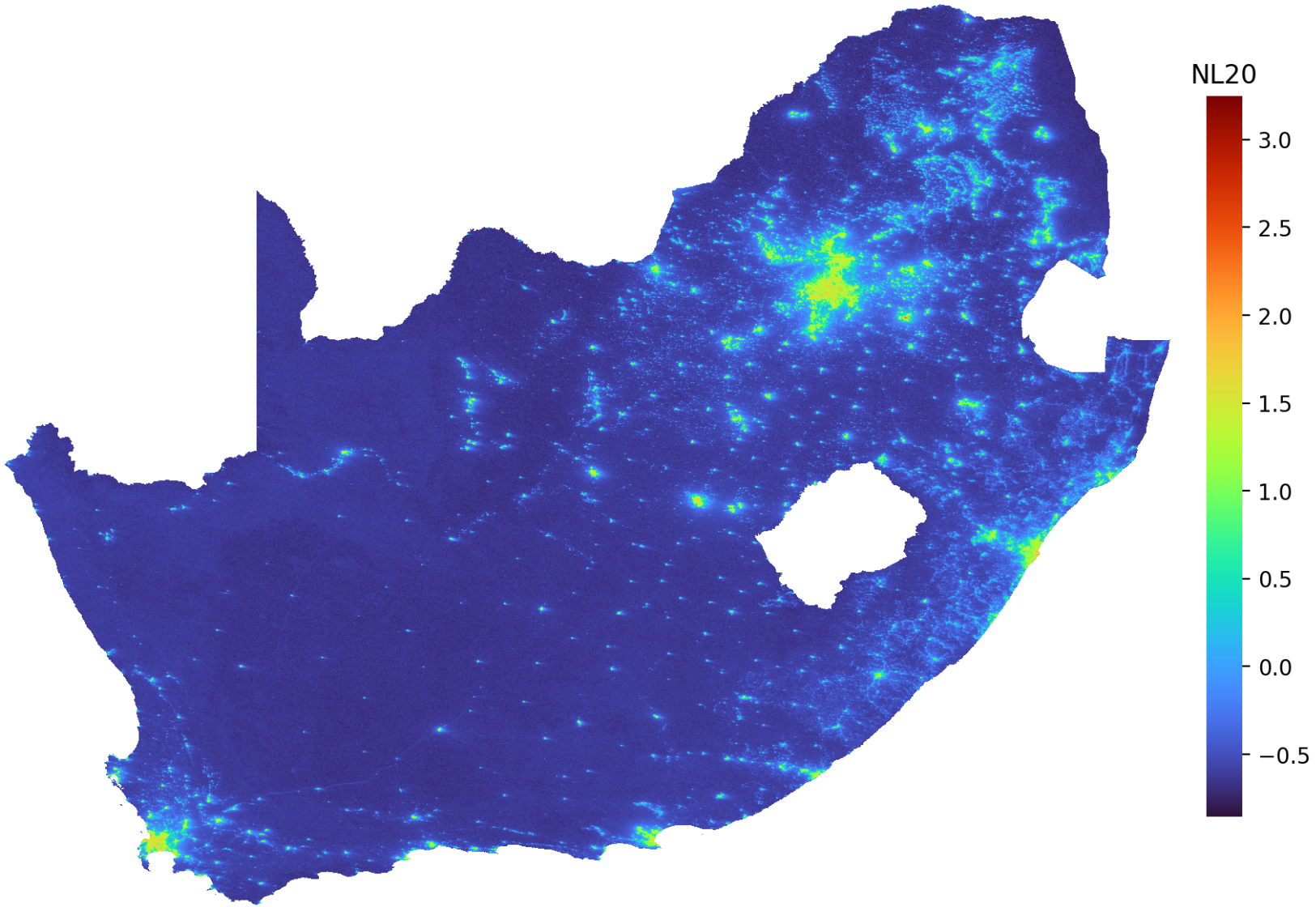
International Wealth Index



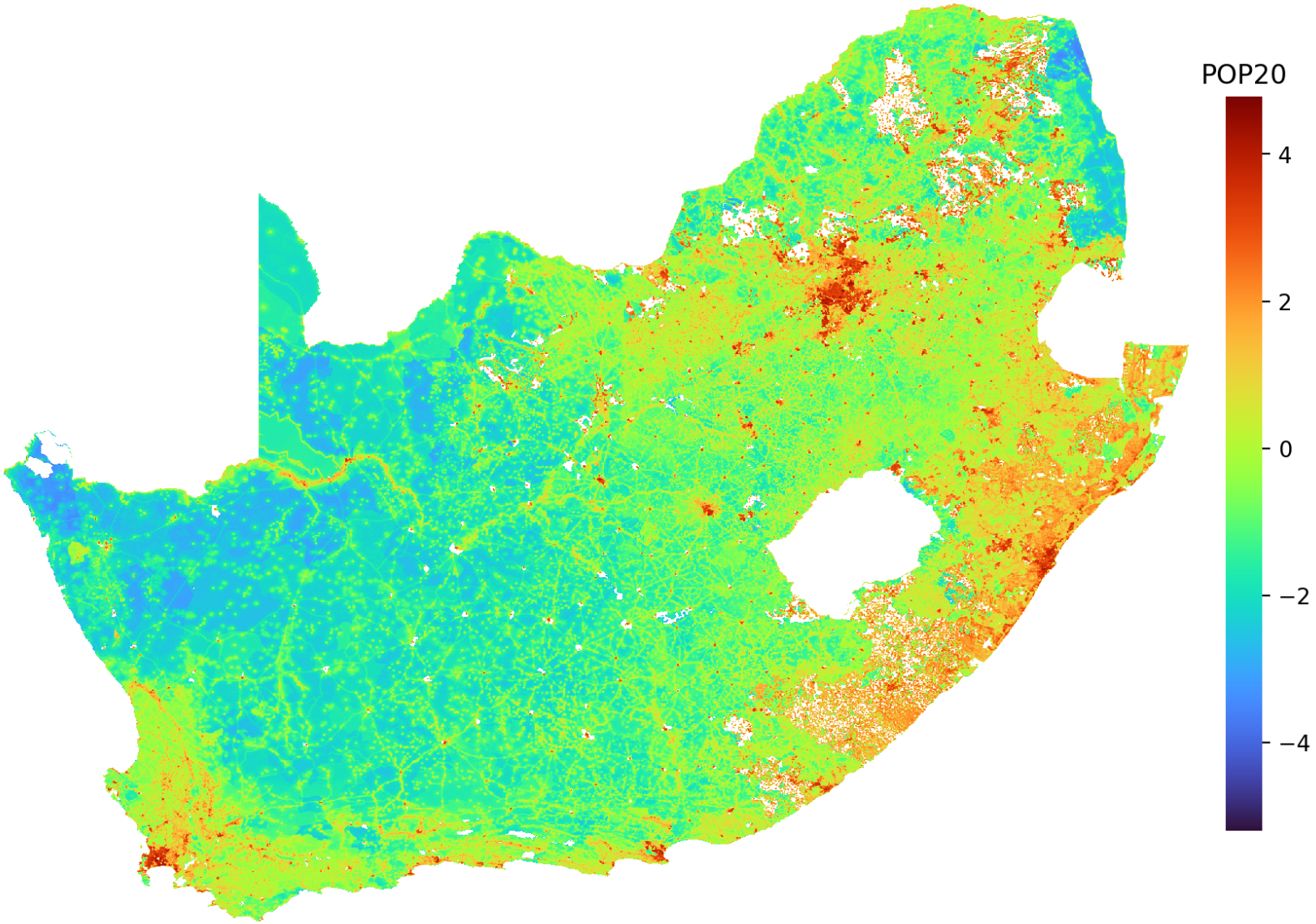
Relative Wealth Index



Log10 VIIRS-DNB 2020

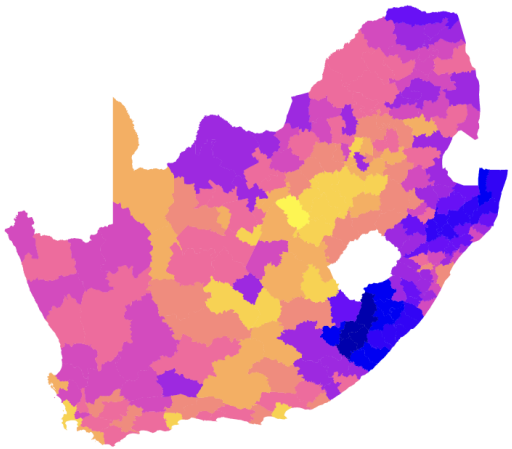


Log10 WorldPop Un-Adjusted Population 2020

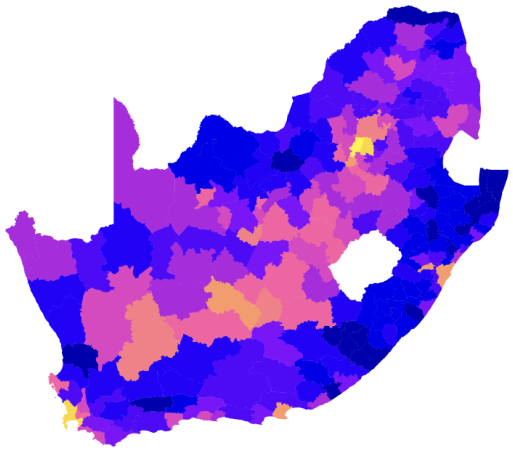


Municipal Estimates

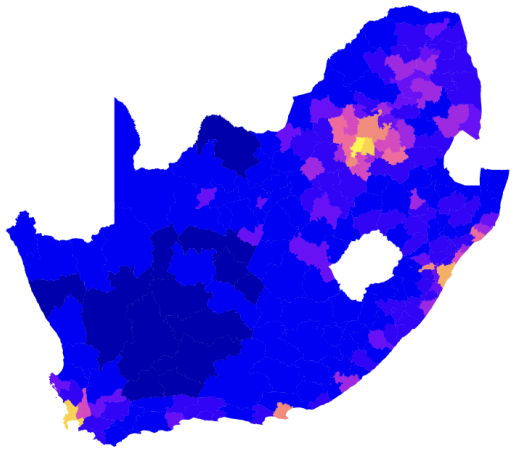
International Wealth Index [36.02, 87]



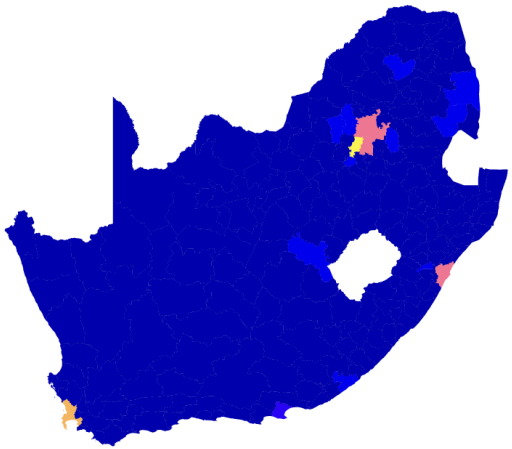
Relative Wealth Index [−0.39, 0.88]



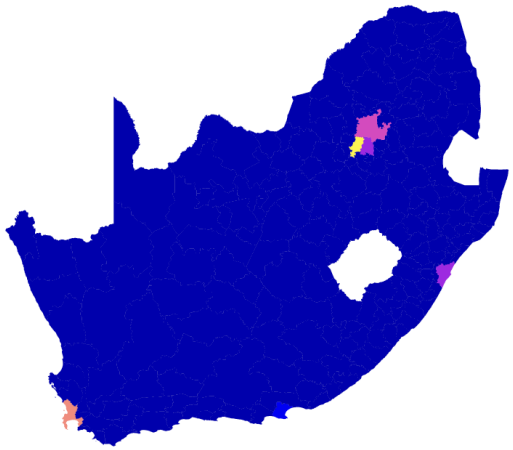
Log10 VIIRS−DNB 2020 [−0.64, 1.27]



WorldPop Population 2020 [9571.6, 5778029.8]



STP3: Full−Time−Eq. Employees 2020 [156.87, 2128914.52]



STP3: Median Income 2020 [3473.16, 34323.55]

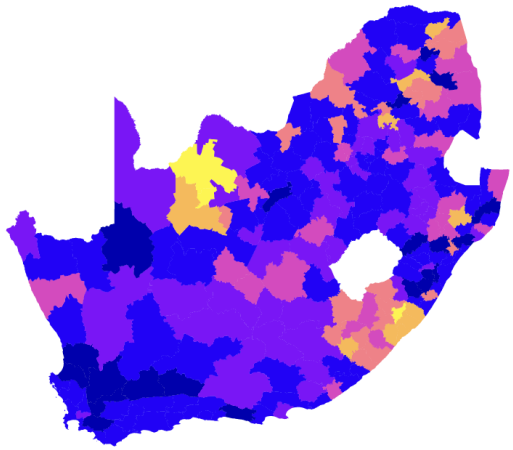


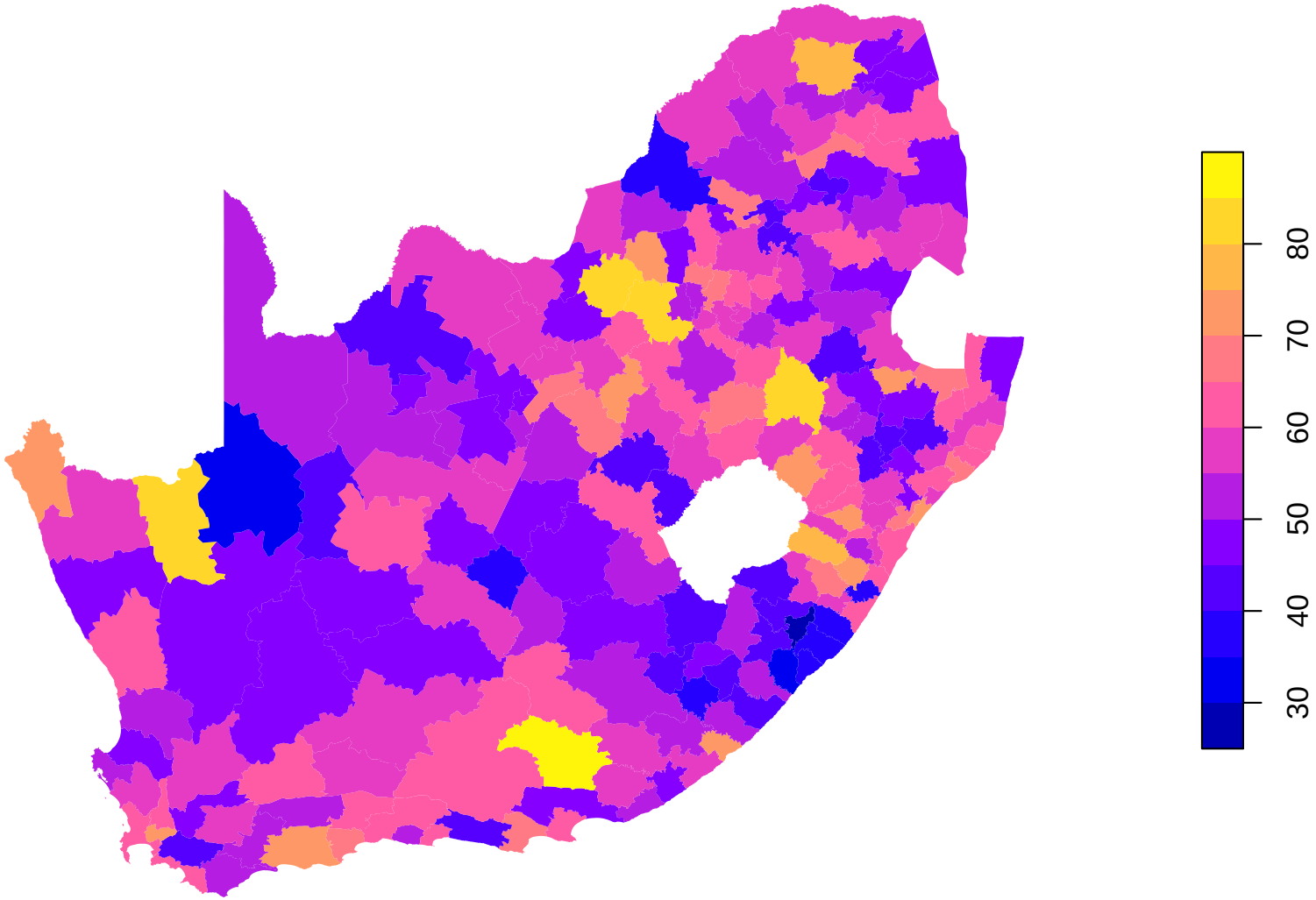
Table: Correlations Across 213 Municipalities

	IWI	RWI	LNL20	POP20	FTE20
IWI	1				
RWI	.753*	1			
LNL20	.236*	.526*	1		
POP20	.204*	.483*	.716*	1	
FTE20	.233*	.475*	.638*	.969*	1
MedInc20	-.304*	-.081	.050	.030	-.012

Notes: * indicates significance at 5% level. There are 213 municipalities.

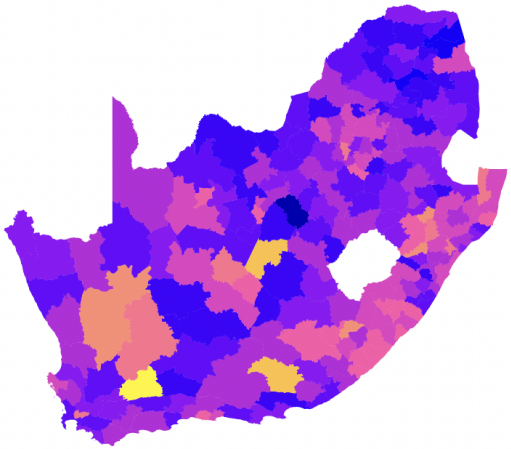
It is notable that median income from STP3 is not correlated with population, RWI or nightlights, and negatively with the IWI!

STP3: GINI Index 2020 [29.95, 85.44]

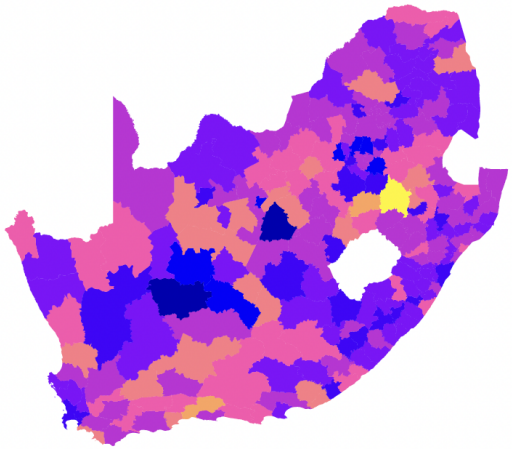


Raw (uncalibrated) estimates, with or without population weights

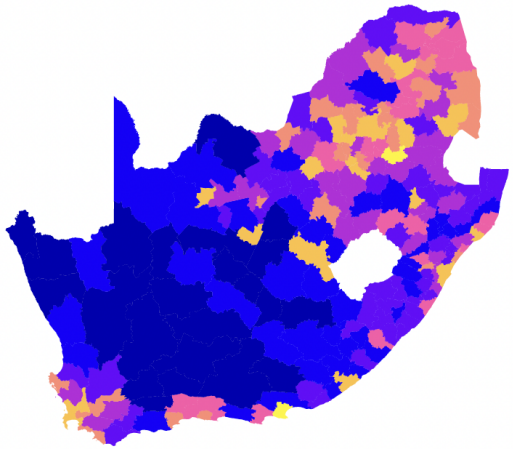
International Wealth Index GINI [1.25, 28.67]



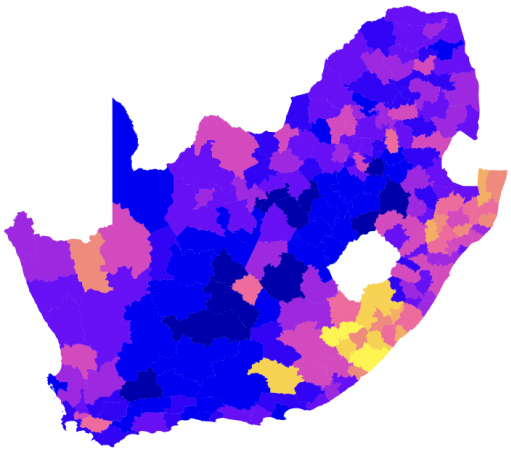
Relative Wealth Index GINI [8.25, 26.26]



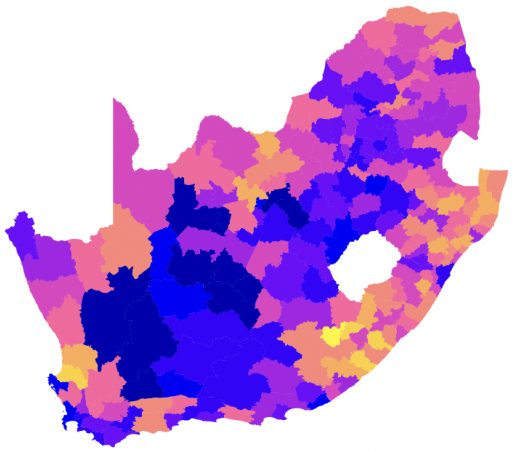
VIIRS–DNB 2020 GINI [2.61, 74.05]



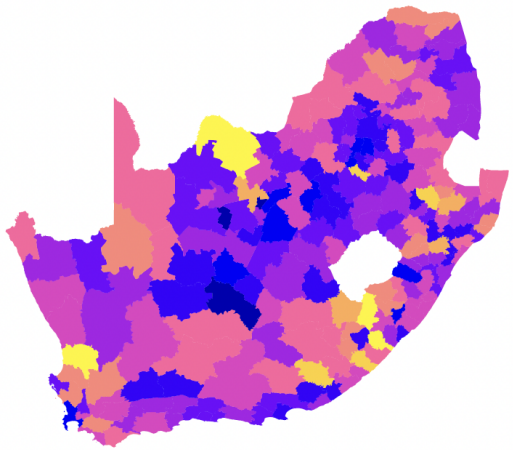
International Wealth Index Weighted GINI [0.01, 21.08]



Relative Wealth Index Weighted GINI [2.03, 23.56]



VIIRS–DNB 2020 Weighted GINI [18.76, 67.33]



Calibration

To increase the realism of estimates, I use a calibration approach, involving finding an optimal power k such that

$$\min_k \quad \frac{1}{N}\sum_{i=1}^{N} |GINI^{\text{RS}} - GINI^{\text{STP3}}|$$

where $GINI^{\text{RS}} = w\_gini(x^k, \text{POP20})$ for $x \in$ IWI, RWI and NL20[2].

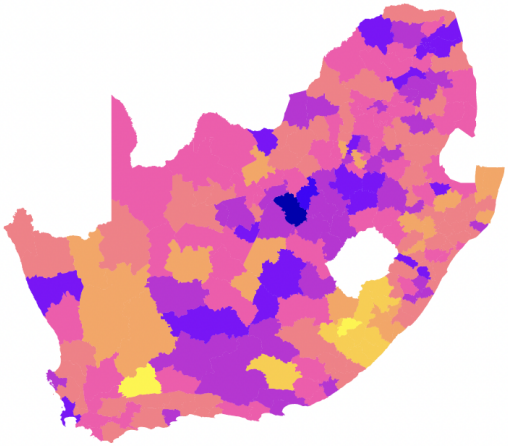
Table: Calibration results, with or without population weights

	Raw		Weighted	
	k	Obj.	k	Obj.
IWI	6.54	12.9	14.72	22.9
RWI	4.07	8.4	7.04	17.9
NL20	1.57	18.7	1.66	14.0

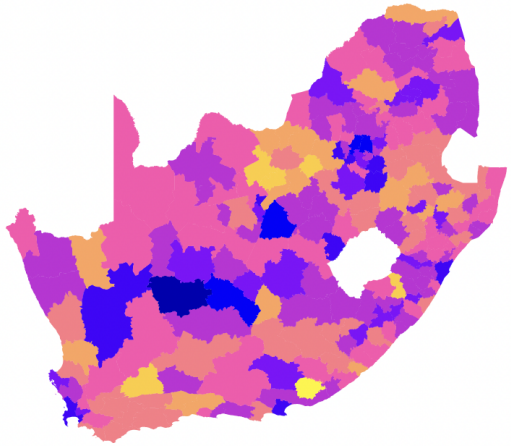
[2]In words: I raise the spatial income/wealth measure to an optimal power k such that weighted municipal GINI estimates most closely resemble the STP3.

Calibrated estimates, with or without population weights

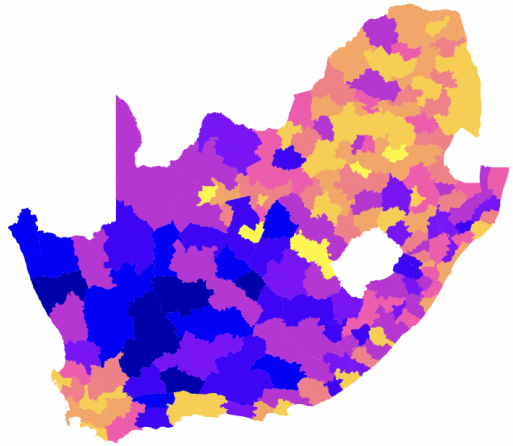
International Wealth Index GINI [8.23, 95.58]



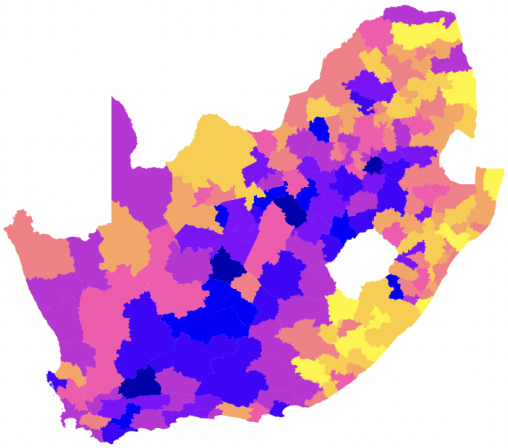
Relative Wealth Index GINI [32.11, 76.28]



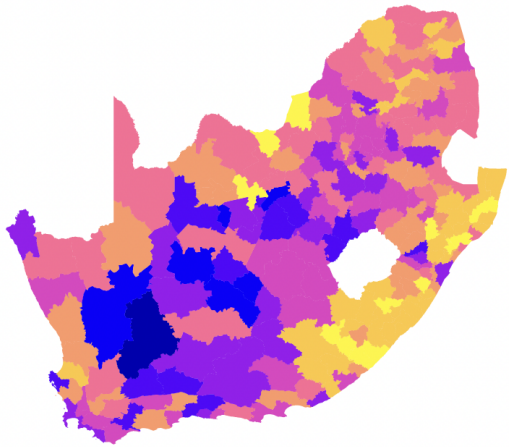
VIIRS–DNB 2020 GINI [6.34, 95.85]



International Wealth Index Weighted GINI [0.01, 97.74]



Relative Wealth Index Weighted GINI [8.66, 86.92]



VIIRS–DNB 2020 Weighted GINI [23.96, 83.37]

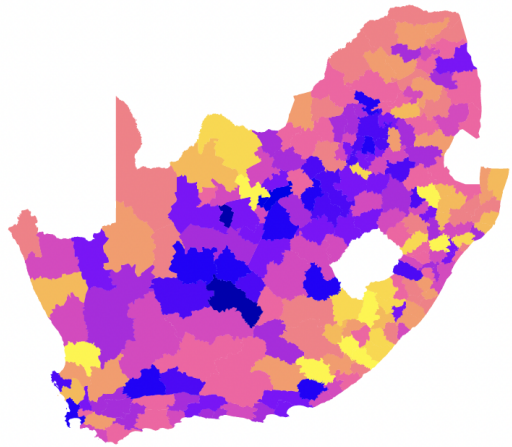


Table: Correlations of Raw (Uncalibrated) Municipal GINI Estimates

	NL20_W_GINI	IWI_W_GINI	RWI_W_GINI	NL20_GINI	IWI_GINI	RWI_GINI
NL20_W_GINI	1					
IWI_W_GINI	.388*	1				
RWI_W_GINI	.568*	.603*	1			
NL20_GINI	-.111	-.041	-.066	1		
IWI_GINI	-.002	.424*	.030	-.106	1	
RWI_GINI	.110	-.194*	.005	-.013	.013	1
STP3_GINI	-.100	-.320*	-.191*	.166*	-.043	.298*

Table: Correlations of Calibrated Municipal GINI Estimates

	NL20_W_GINI	IWI_W_GINI	RWI_W_GINI	NL20_GINI	IWI_GINI	RWI_GINI
NL20_W_GINI	1					
IWI_W_GINI	.569*	1				
RWI_W_GINI	.623*	.678*	1			
NL20_GINI	-.056	.019	-.001	1		
IWI_GINI	.283*	.364*	.242*	-.111	1	
RWI_GINI	.194*	-.080	.194*	.074	.238*	1
STP3_GINI	-.210*	-.341*	-.270*	.185*	-.185*	.160*

Restricting sample to 61 municipalities with population > 200,000

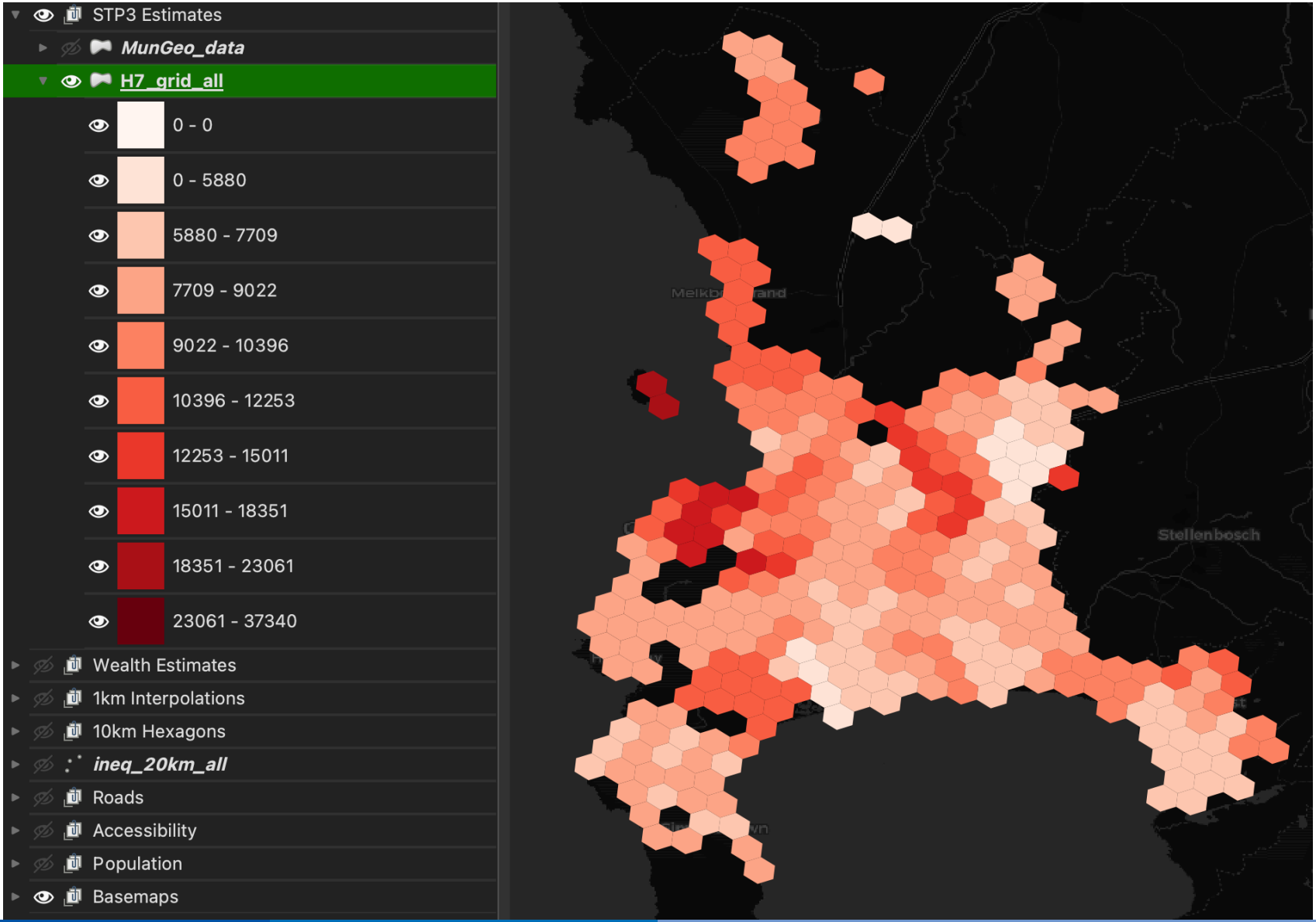
Table: Correlations of Raw (Uncalibrated) Municipal GINI Estimates

	NL20_W_GINI	IWI_W_GINI	RWI_W_GINI	NL20_GINI	IWI_GINI	RWI_GINI
NL20_W_GINI	1					
IWI_W_GINI	.644*	1				
RWI_W_GINI	.778*	.586*	1			
NL20_GINI	-.461*	-.575*	-.631*	1		
IWI_GINI	.166	.529*	.108	-.267*	1	
RWI_GINI	.100	-.123	.128	.190	.086	1
STP3_GINI	-.214	-.331*	-.418*	.366*	.024	.062

Table: Correlations of Calibrated Municipal GINI Estimates

	NL20_W_GINI	IWI_W_GINI	RWI_W_GINI	NL20_GINI	IWI_GINI	RWI_GINI
NL20_W_GINI	1					
IWI_W_GINI	.708*	1				
RWI_W_GINI	.700*	.712*	1			
NL20_GINI	-.387*	-.447*	-.592*	1		
IWI_GINI	.481*	.318*	.352*	-.259*	1	
RWI_GINI	.380*	.102	.301*	.120	.474*	1
STP3_GINI	-.320*	-.375*	-.496*	.347*	-.120	-.119

STP3 also provides median income and GINI for 1.4km hexagons in cities. This shows median income in Cape Town in 2020.



Comparison of Income / Wealth Measures

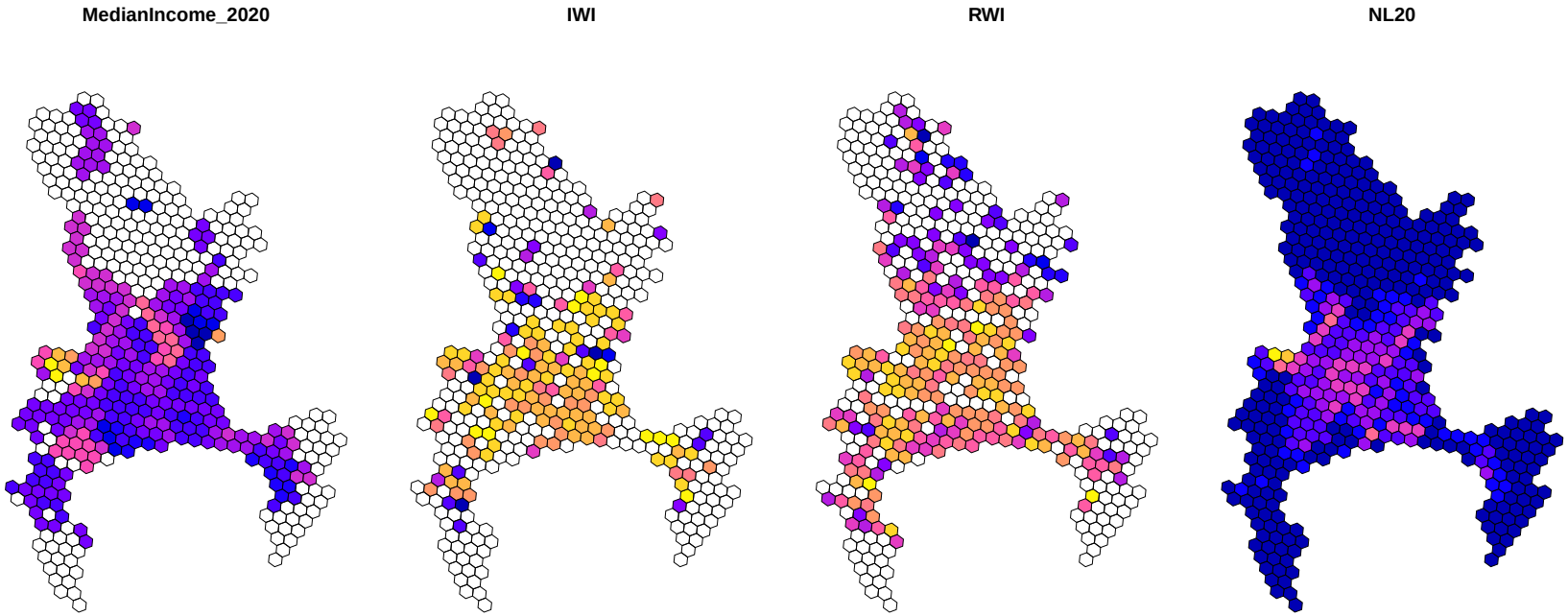
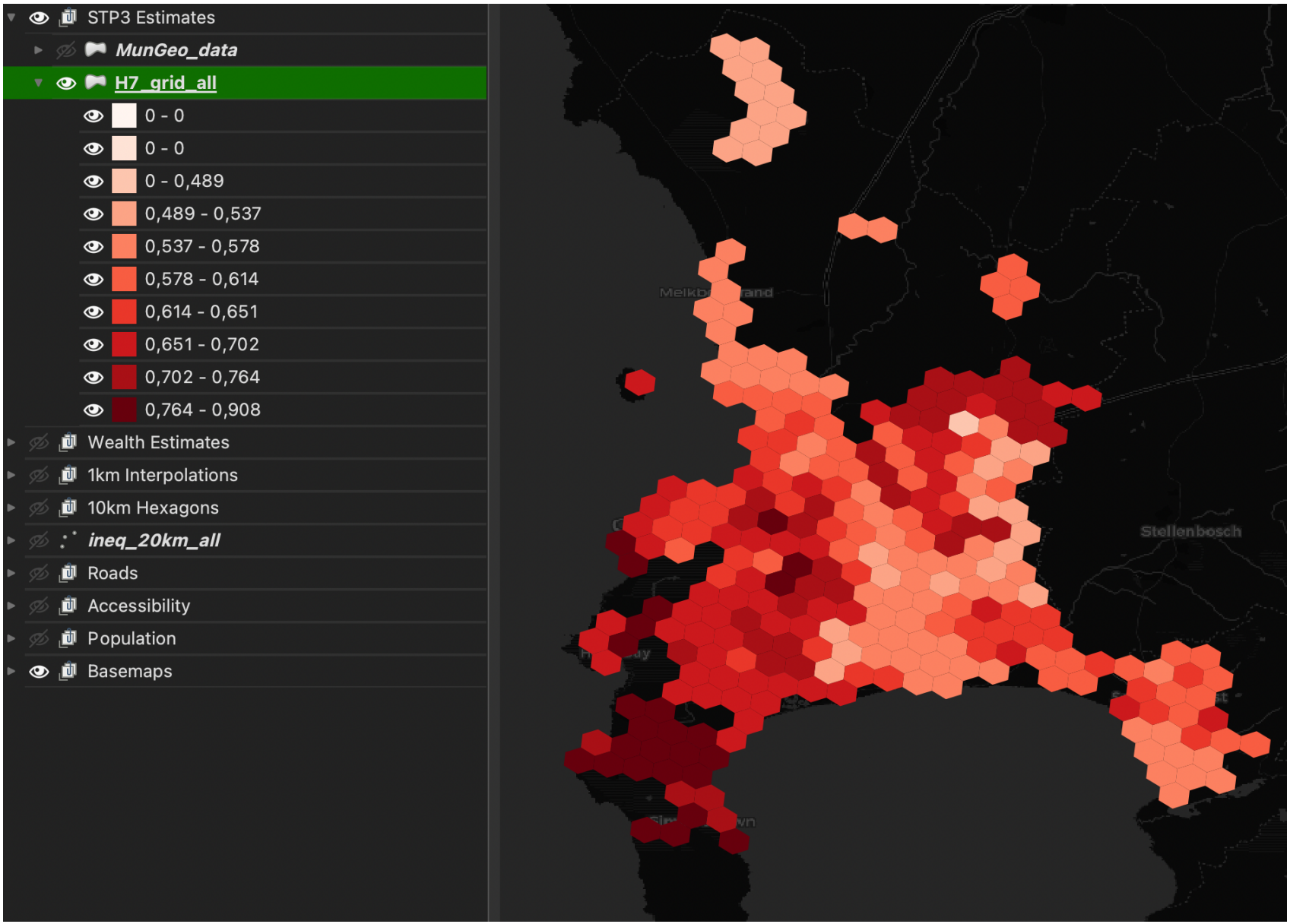


Table: Correltion of Income/Wealth Measures in 1.4km STP3 Hexagons

	Overall (National)				Within Municipalities			
	MedInc20	IWI	RWI	NL20	MedInc20	IWI	RWI	NL20
MedInc20	1 (1904)				1 (1899)			
IWI	-.079* (899)	1 (10623)			-.082* (897)	1 (10623)		
RWI	-.109* (1276)	.645* (6949)	1 (22947)		-.046 (1272)	.427* (6949)	1 (22947)	
NL20	-.024 (1898)	.490* (10611)	.541* (22927)	1 (225311)	.057* (1893)	.368* (10611)	.376* (22927)	1 (225311)

Notes: * indicates significance at 5% level. There are 213 municipalities. Within-data is standardized (scaled and centered) in each municipality.

This shows the GINI in Cape Town in 2020.



IWI (1.6km) and RWI (2.4km) are too low-res to compute a GINI in 1.4km hexagons. **2 Options:**

- Increase the size of Hexagons

→ I use 96km2 (6km diameter) hexagons → 9.53 IWI Obs. (IWI is sparse), 12.4 RWI Obs. and 79.5 NL Obs. per hexagon.

- Interpolate i.e. compute GINI at high resolution using surrounding data points with spatial weights

→ I compute 1km GINIs based on spatial interpolations with 5km, 10km or 30km radius i.e. a weighted GINI is computed with inverse distance × population weights.

Comparison of GINI Estimates on 96km2 Hexagons

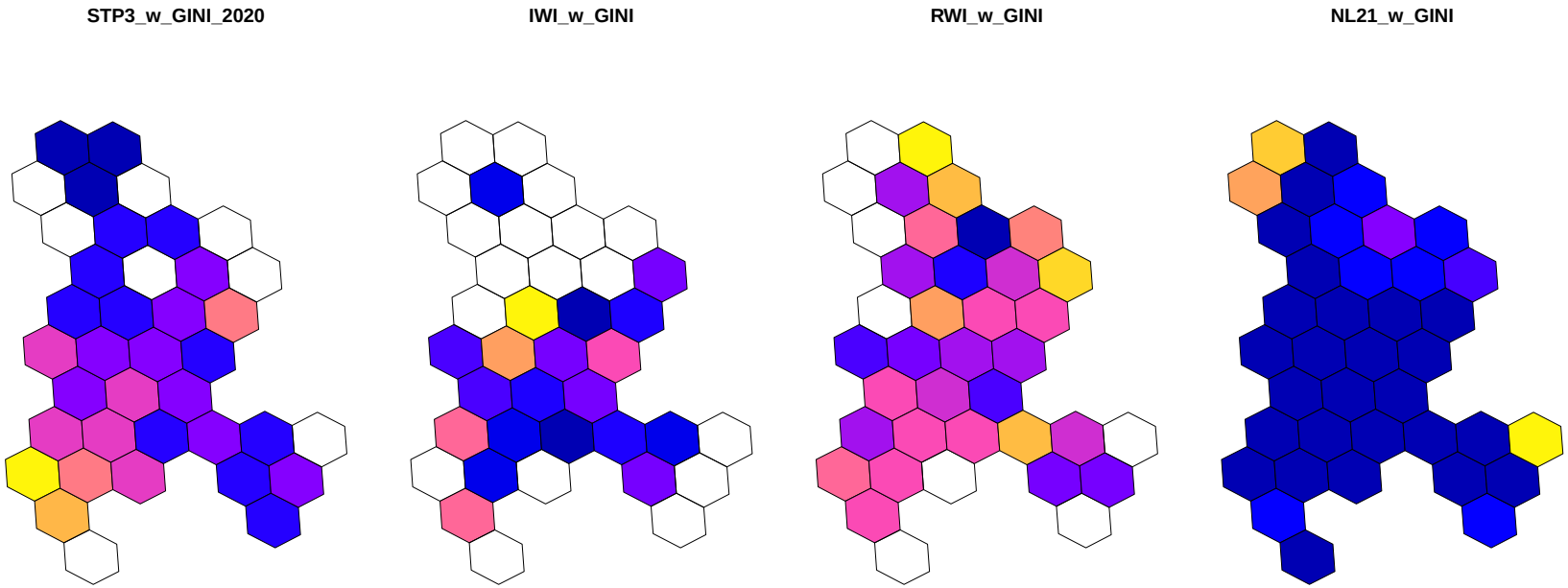
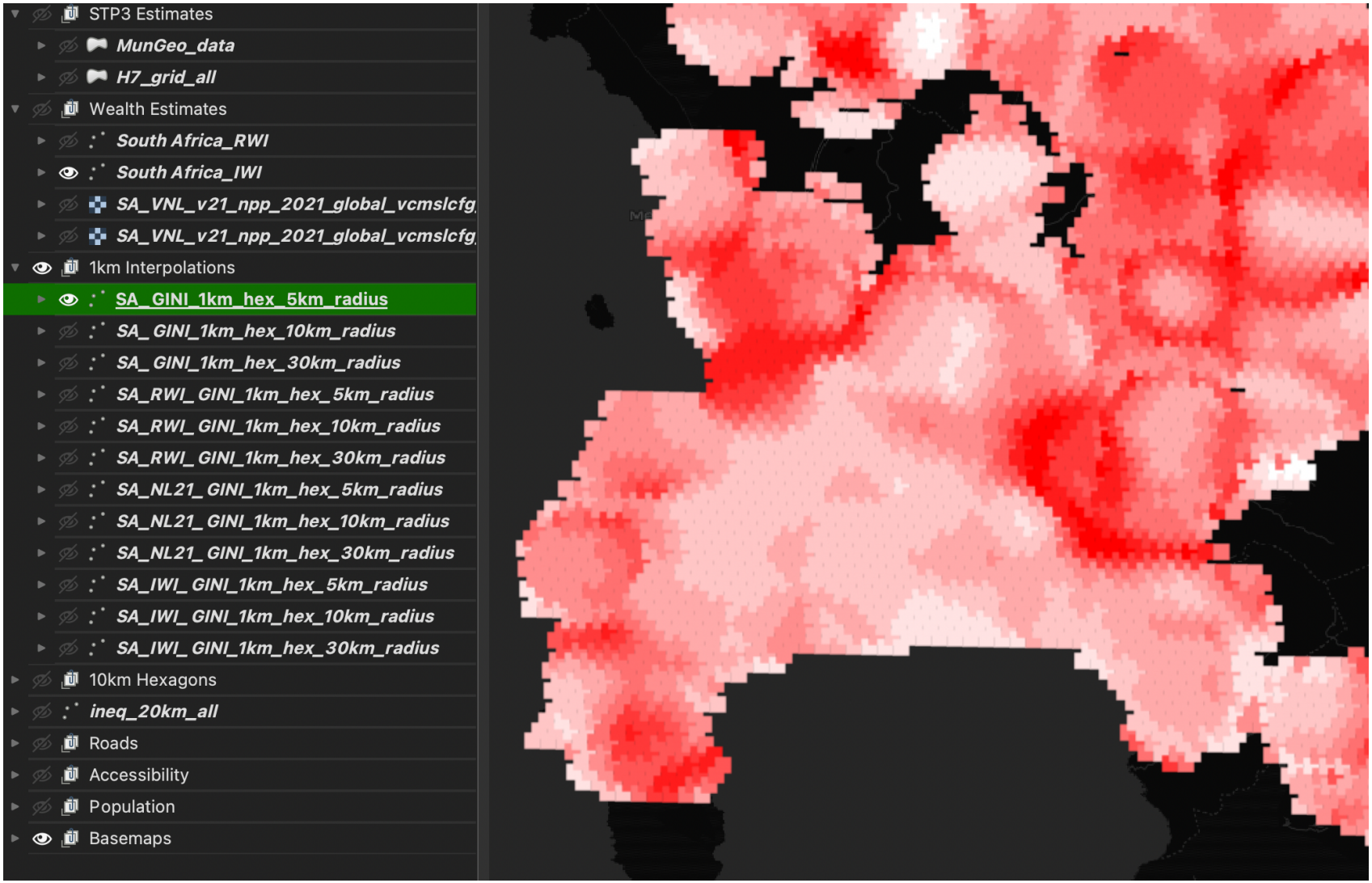


Table: Population Weighted GINI's in 96km2 Hexagons

	IWI_w_GINI	RWI_w_GINI	NL21_w_GINI	WGINI_mean	STP3_w_GINI_2020
IWI_w_GINI	1 (1351)				
RWI_w_GINI	.136* (1333)	1 (2603)			
NL21_w_GINI	.451* (1333)	.075* (2519)	1 (6471)		
WGINI_mean	.827* (1316)	.461* (1316)	.773* (1316)	1 (1316)	
STP3_w_GINI_2020	.084 (143)	-.188* (173)	-.270* (189)	-.093 (142)	1 (189)

Notes: * indicates significance at 5% level

This shows 1km2 weighted GINI with 5km interpolation radius.



Comparison of GINI Estimates 1.4km2 STP3 Hexagons

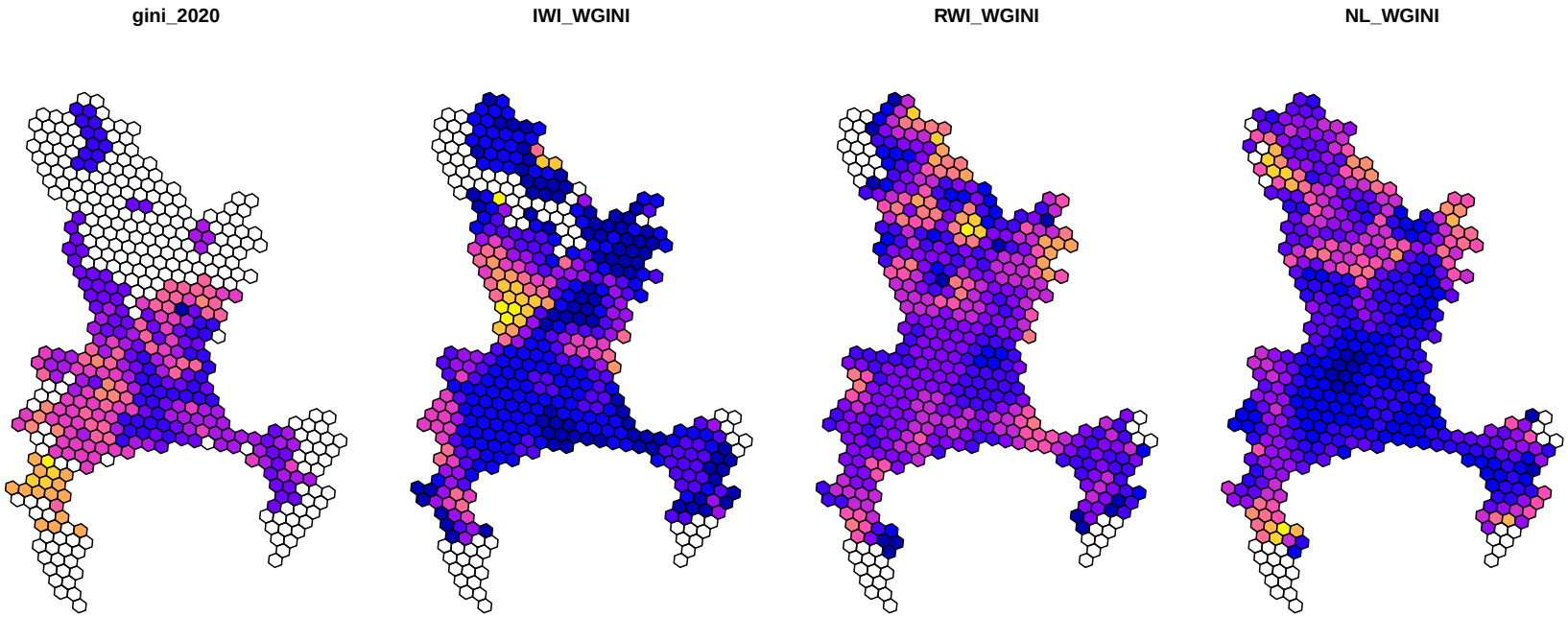


Table: Weighted GINI Interpolations Aggregated to 1.4km STP3

	IWI_WGINI	RWI_WGINI	NL_WGINI	WGINI_mean	STP3_w_GINI_2020
IWI_WGINI	1 (45536)				
RWI_WGINI	.249* (44850)	1 (57638)			
NL_WGINI	.147* (45469)	.169* (57532)	1 (65957)		
WGINI_mean	.842* (44731)	.643* (44731)	.474* (44731)	1 (44731)	
STP3_w_GINI_2020	.138* (1611)	-.124* (1616)	-.107* (1617)	.046 (1611)	1 (1617)

Table: Weighted GINI's in 1km2 Cells with 5km Interpolation Radius

	IWI_WGINI	RWI_WGINI	NL_WGINI	WGINI_mean
IWI_WGINI	1 (174922)			
RWI_WGINI	.213* (172130)	1 (222337)		
NL_WGINI	.167* (174761)	.158* (222083)	1 (255120)	
WGINI_mean	.841* (171969)	.624* (171969)	.492* (171969)	1 (171969)

Notes: * indicates significance at 5% level

Conclusions

Time-Series Estimates

- Nightlights based national GINI estimate shows a declining trend since 2014, in-line with the STP3 estimate from 2018.
- Pretty much useless for municipal GINI time series estimation, which also calls into question national estimate.
- Since Galimberti et al. (2021) show that nightlights are not good at capturing time dynamics in inequality in the average country, these estimates should not be used for research and policy advice.

Geospatial Estimates

- Calibrated municipal GINI estimates based on nightlights or remotely sensed wealth indices are moderately consistent ($\rho = 0.6$, and 0.7 for populated places), but at odds (negatively correlated) with STP3 GINI (negative/zero corr. of indices w. STP3 MedInc).
- High-resolution estimates, based on both 96km2 hexagons and 1km interpolations with various radii (5km preferred) are broadly consistent, and interesting to map structural inequality between communities. But not correlated with tax-based inequality (STP3).

References

Chi, G., Fang, H., Chatterjee, S. & Blumenstock, J. E. (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, *119*(3), e2113658119.

Galimberti, J. K., Pichler, S. & Pleninger, R. (2021). *Measuring inequality using geospatial data* (Tech. Rep.). KOF Working Papers.

Lee, K. & Braithwaite, J. (2022). High-resolution poverty maps in sub-saharan africa. *World Development*, *159*, 106028.

Nell, A. & Visagie, J. (2023). *Spatial tax panel 2014-2022 (version 3)* (Tech. Rep.). National Treasury - Cities Support Programme and Human Sciences Research Council.