Did High Speed Rails Promote Local Economic Activities?: Application of Night Time Satellite Data

Kanghyock Koh [1] Hyunjoo Yang [2]

[1]Korea University

[2]Incheon National University

11/15/2019

Motivation

- Cost of construction of Korean High Speed Rail (KTX) about $ 27 billion USD

  (Yonhap News 2016.9.21.)

- Few studies evaluating the overall impacts of KTX on local economies

- Difficulty of doing empirical analysis at the sub-national level
  - Ideal data are regional level balanced panel data
  - Lack of county- or township-level GDP data
  - Lack of high frequency data
  - Frequent administrative unit changes

Contribution of this paper

Contribution of this paper

- Novel data set of 20-year balanced panel with a local economic activity measure
- Empirical analysis of the effects of KTX on local economic activity

Mixed evidence on the effects of High Speed Rail (HSR)

Positive impacts of HSR

- Better transportation infrastructure improves economic activities
- Housing boom and new industrial parks
- (Sands, 1993; Banister and Berechman, 2003; Kwon 2014)
- Increase in firm productivity (Baek and Park 2015)
- New entry of firms, employment, productivity (Kim and Ahn 2009, 2013; Ahn and Kim 2009)

Effects of High Speed Rail on Local Economic Activity

Negative impacts of HSR

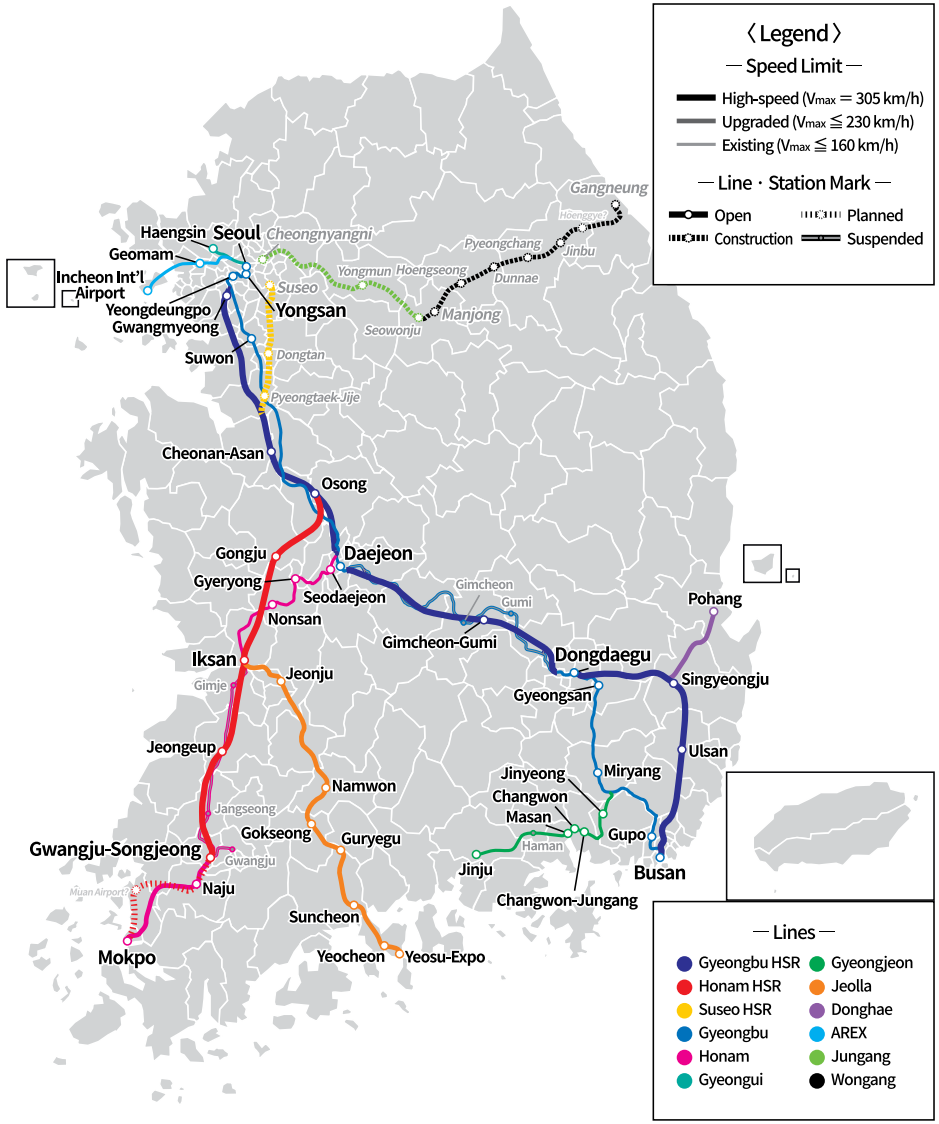
- Economic activities being drained away from larger cities (Whitelegg and Holzapfel, 1993; Thompson, 1995; Vickerman, 1997)
- Cancer patients switch to hospitals in larger cities (Kim et al., 2010)

$\longrightarrow$ Empirical investigation is needed for assessing overall effects

Preview of Findings

- Introduction of the HSR system increases local economic activities of rural townships near HSR stations by 10% in 10 years

The KTX Network



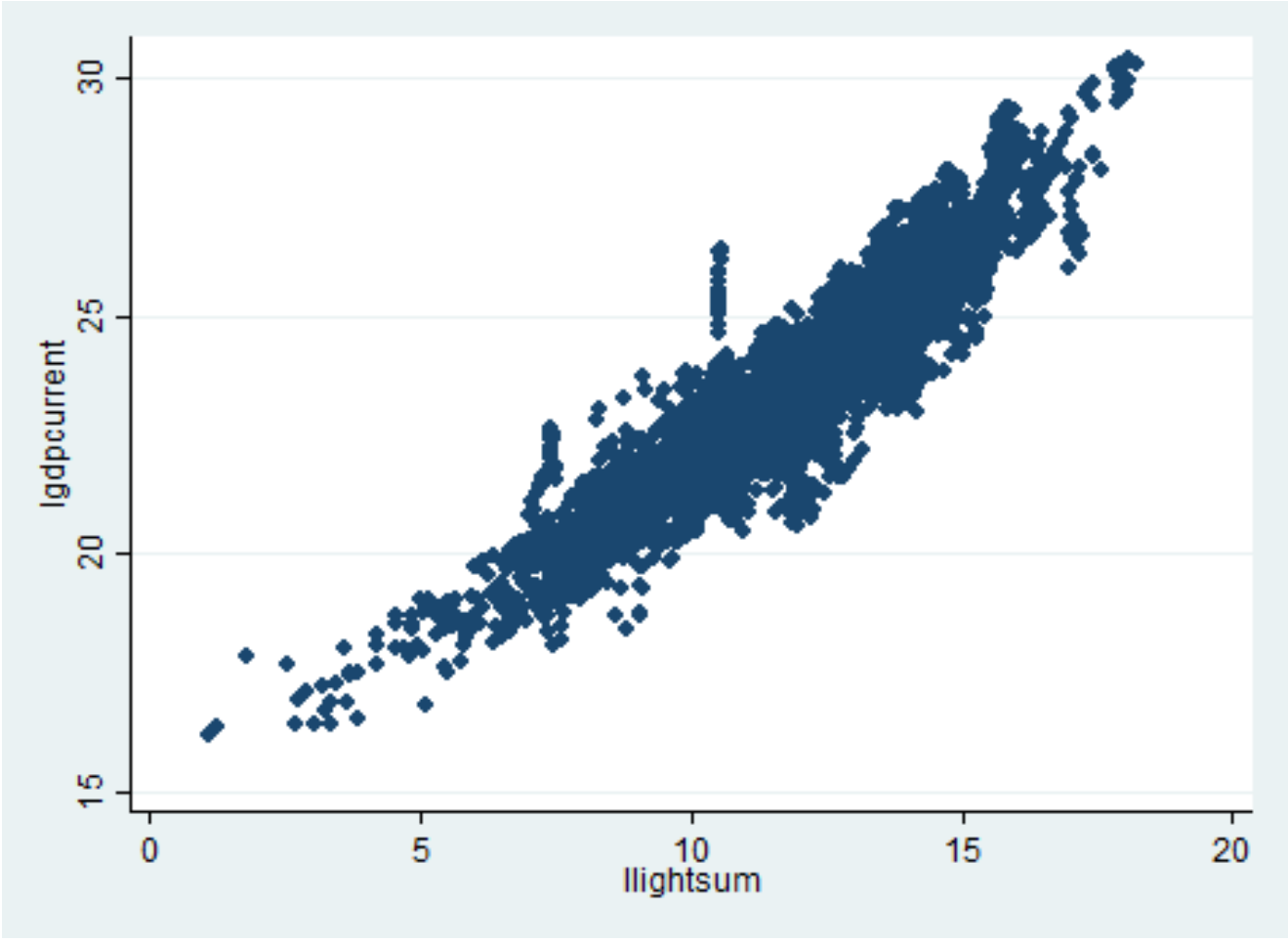
Data

- Township-level measure of economic activity
  - National Oceanic and Atmospheric Administration (NOAA) Night Time Light Data from Defense Meteorological Satellite Program (DMSP), 1994-2013
  - Reliable measure of local economic activities (Henderson et a. 2012; Donaldson and Storeygard 2016)
  - lights-GDP elasticity of 0.3 (Henderson et a. 2012)
  - About 600m $\times$ 600m pixel with light intensity 0-63
- Construction of a 20-year township level balanced panel data set
  - 1,265 townships $\times$ 20 years $=$ 25,300 observations

Data (cont.)

- Construction of a 20-year township level balanced panel data set
  - 1,265 townships $\times$ 20 years $=$ 25,300 observations
- Distance between KTX stations and townships
  - Location of KTX stations are geo-coded using KTX line maps
  - Euclidean distance from the nearest KTX station
- Excluding Gangwon province, islands and urban districts

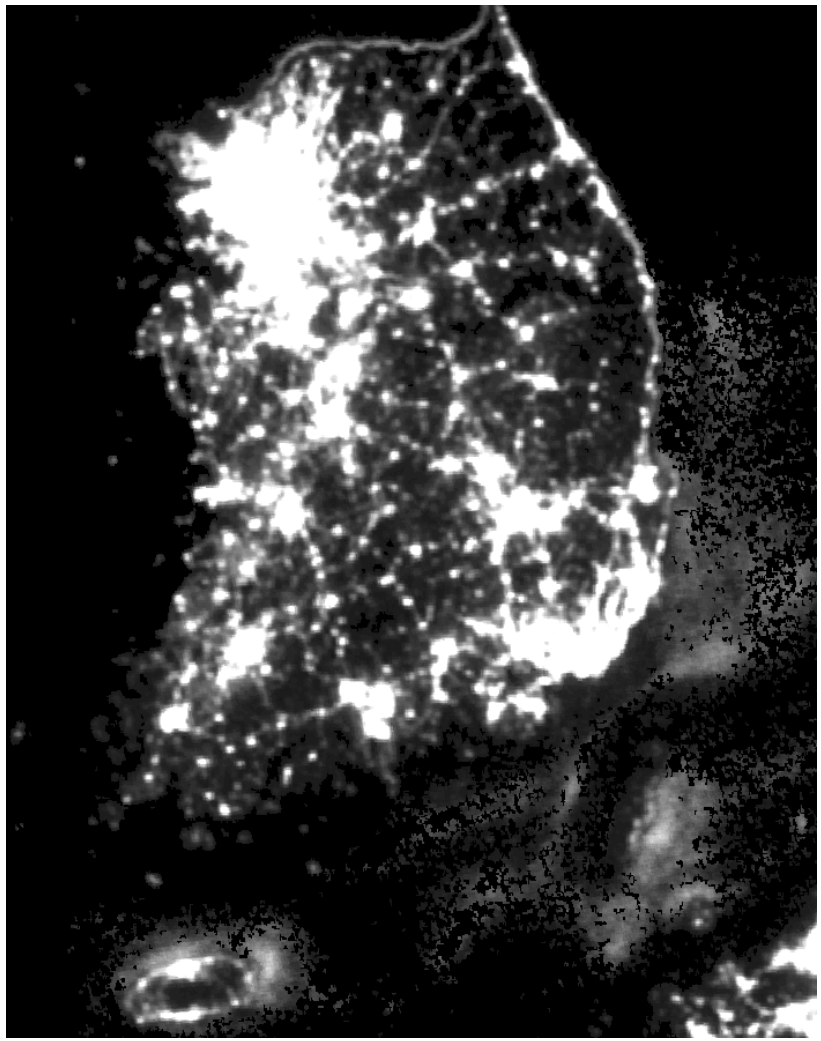
World GDP and Light Intensity 1992-2013



Source: Kyoochul Kim. (2017) North Korean Economy Seen from the Sky, *KDI Policy Research Series Working Paper*

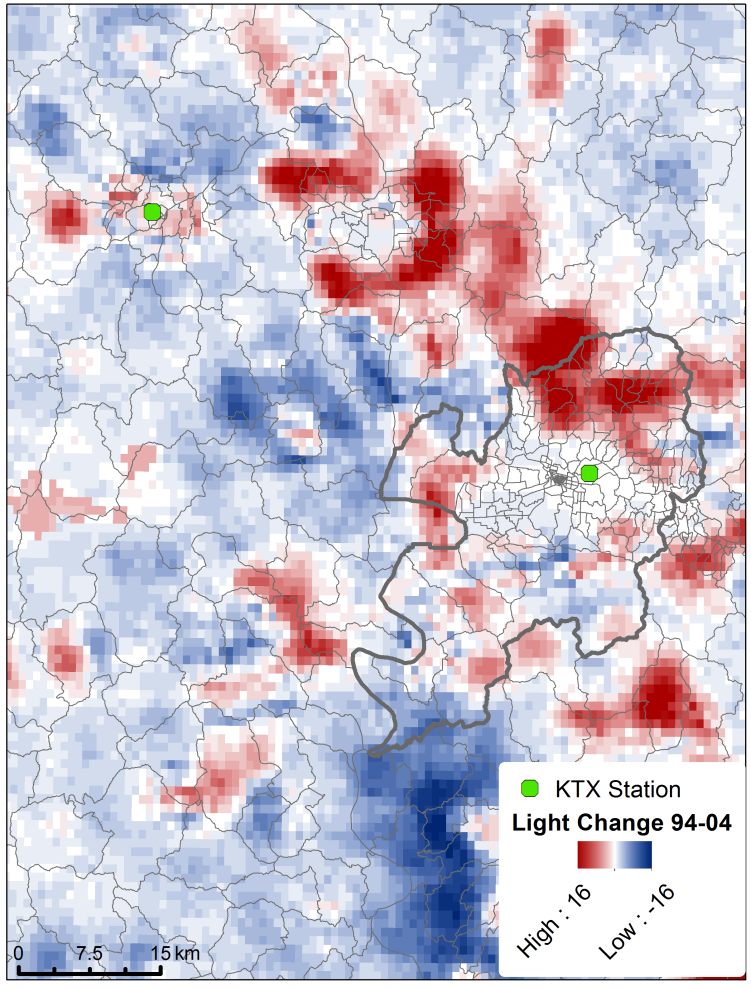
Raw Satellite Images 1994 & 2013

(a) 1994

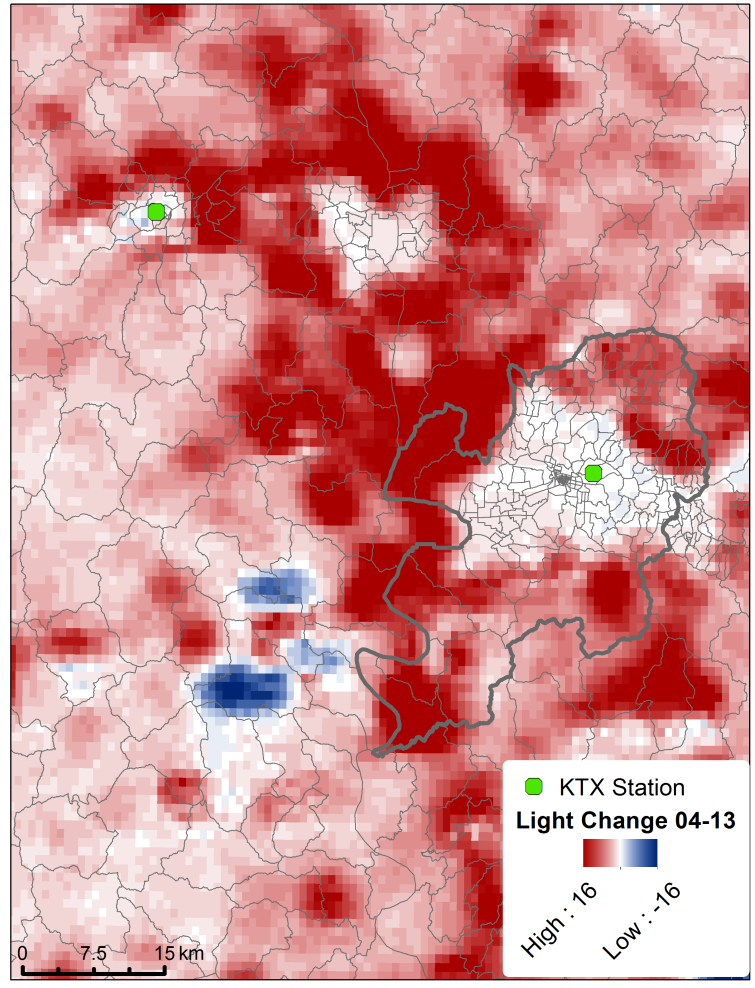


(b) 2013

Growth of Lights near Daegu & and Gimcheon

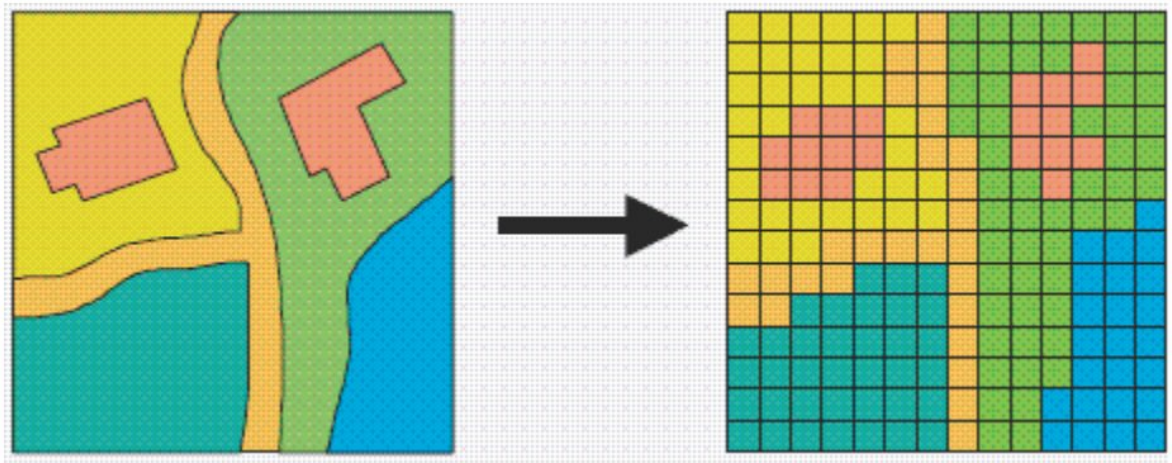


(c) 1994-2004



(d) 2004-2013

GIS data: vector and raster



Vector: points, lines, polygons

Raster: satellite images, altitudes, land cover

Create outcome variable:

Proxy for economic activity using night time satellite images

Data preparation - download data set from NOAA

ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

NOAA > NESDIS > NCEI (formerly NGDC) > EOG

Earth Observation Group

- EOG Home.
- DMSP Archive Description
- Description of DMSP Sensors
- Data Availability
- Data Services and Pricing
- Data Download
- Online Maps and Web Services
- Nighttime Lights Posters
- Presentations
- Publications
- News & Media
- Items of Interest
- Documents About DMSP
- Nighttime Lights Temporal Loops
- Nightsat
- McMurdo Ground Project Data Resources
- FAQ

Version 4 DMSP-OLS Nighttime Lights Time Series

The files are cloud-free composites made using all the available archived DMSP-OLS smooth resolution data for calendar years. In cases where two satellites were collecting data - two c
30 arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude.

More Information

Download Average Visible, Stable Lights, and Cloud Free Coverages || Download Average Lights X Pct

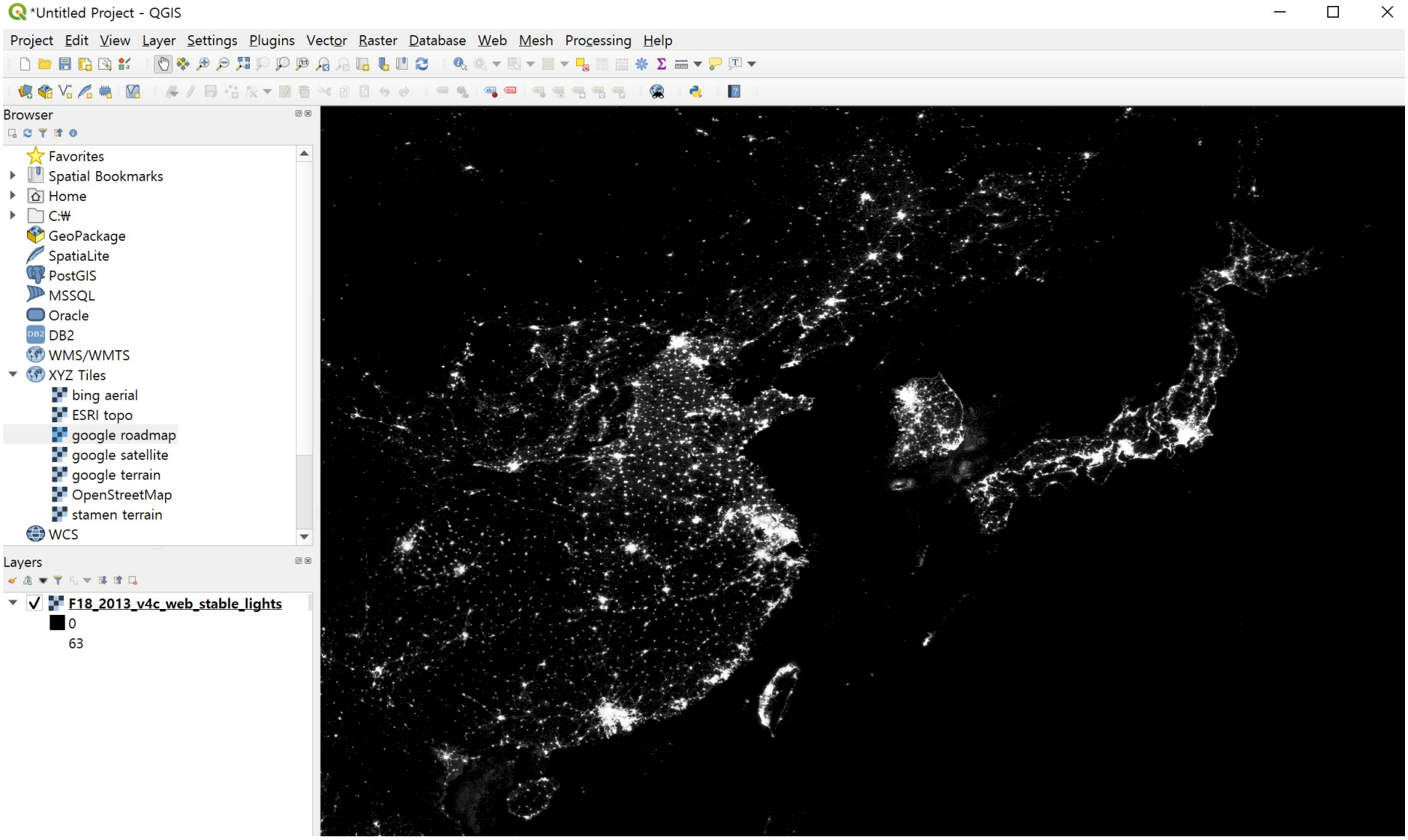
NOTE: Disk space required for compressed data is ~**300 MB** and uncompressed data is ~**3 GB**.
NOTE: These files are gzipped and added to tar files under linux. Winzip and some other windows utilities may convert newlines in the data into a windows linefeed. To avoid this in Win
Miscellaneous and uncheck the 'TAR file smart CR/LF conversion'.

Average Visible, Stable Lights, & Cloud Free Coverages						
Year\Sat.	F10	F12	F14	F15	F16	F18
1992	F101992	-------	-------	-------	-------	-------
1993	F101993	-------	-------	-------	-------	-------
1994	F101994	F121994	-------	-------	-------	-------
1995	-------	F121995	-------	-------	-------	-------
1996	-------	F121996	-------	-------	-------	-------
1997	-------	F121997	F141997	-------	-------	-------
1998	-------	F121998	F141998	-------	-------	-------
1999	-------	F121999	F141999	-------	-------	-------
2000	-------	-------	F142000	F152000	-------	-------
2001	-------	-------	F142001	F152001	-------	-------
2002	-------	-------	F142002	F152002	-------	-------
2003	-------	-------	F142003	F152003	-------	-------
2004	-------	-------	-------	F152004	F162004	-------
2005	-------	-------	-------	F152005	F162005	-------
2006	-------	-------	-------	F152006	F162006	-------
2007	-------	-------	-------	F152007	F162007	-------
2008	-------	-------	-------	-------	F162008	-------
2009	-------	-------	-------	-------	F162009	-------
2010	-------	-------	-------	-------	-------	F182010
2011	-------	-------	-------	-------	-------	F182011
2012	-------	-------	-------	-------	-------	F182012
2013	-------	-------	-------	-------	-------	F182013

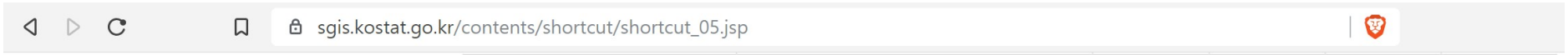
F15 2003 Nighttime Lights Composite

Data preparation - download QGIS software

Data preparation - load satellite images (drag and drop)



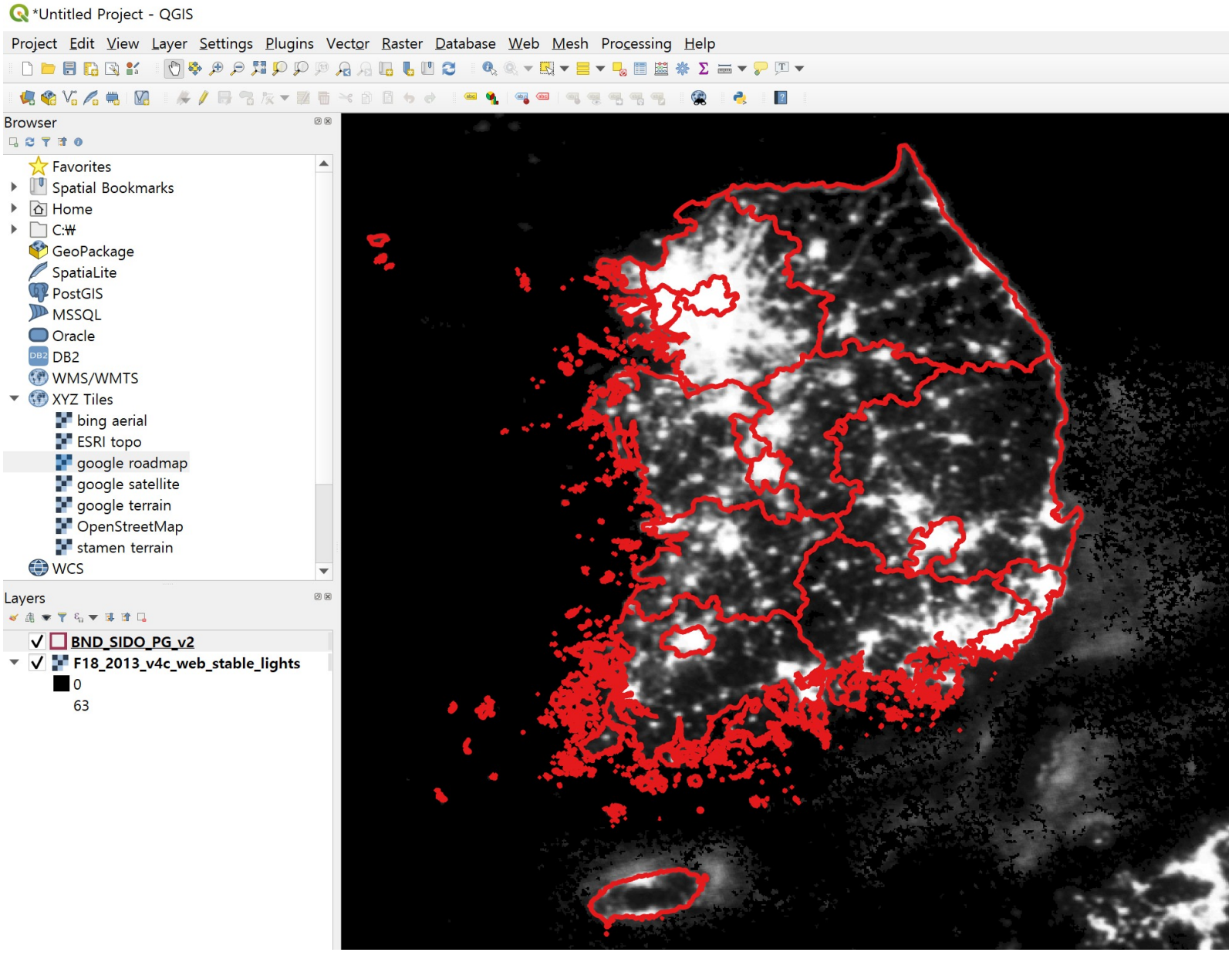
Data preparation - download admin boundaries



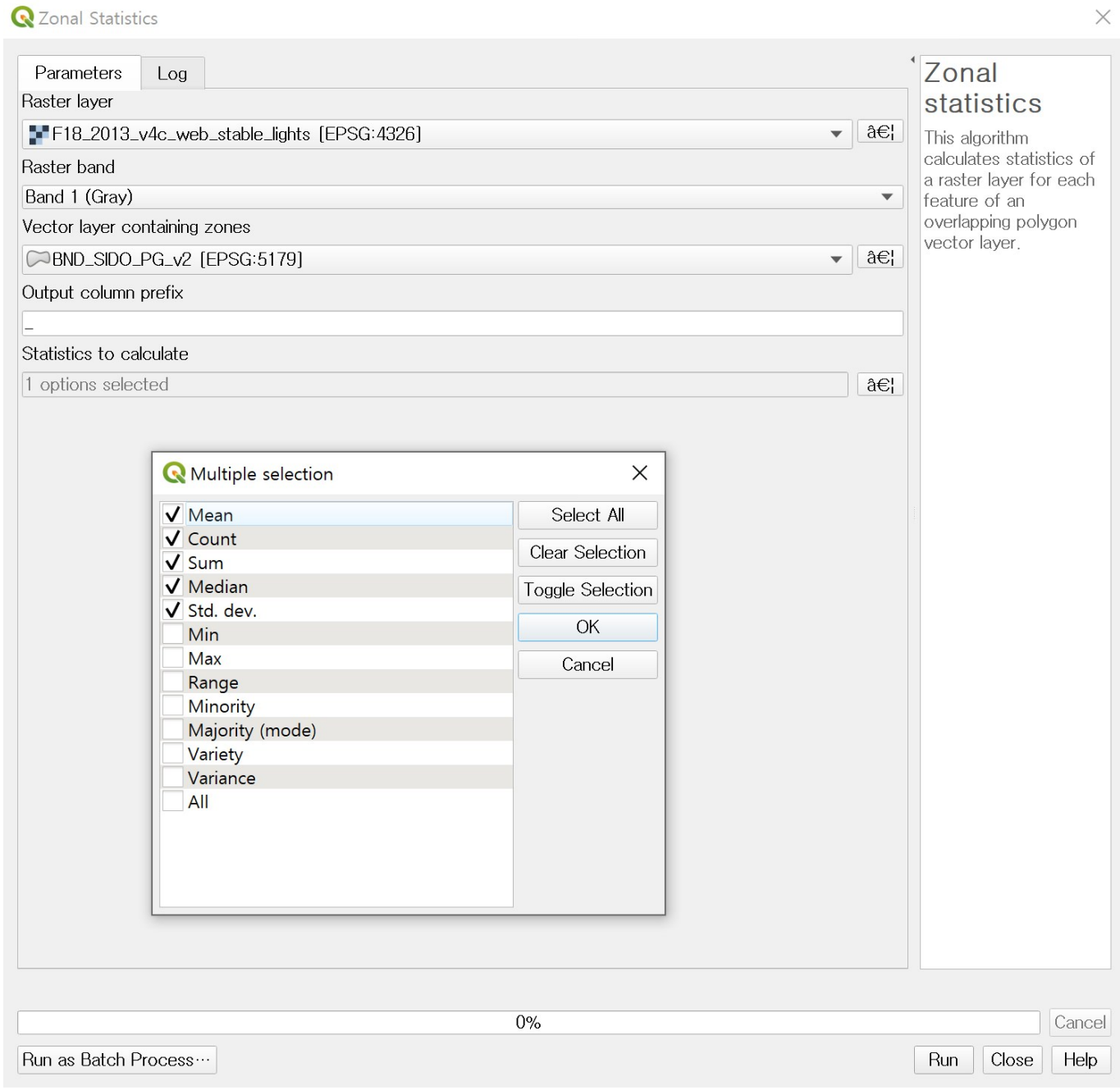
통계지역경계

대상자료명	기준년도	자료형식	공개여부	대상지역	가격
센서스용 행정구역경계(전체)	2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1995, 1990, 1985, 1980, 1975	SHP	공개	전국	무료
센서스용 행정구역경계(시도)	2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1995, 1990, 1985, 1980, 1975	SHP	공개	전국	무료
센서스용 행정구역경계(시군구)	2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1995, 1990, 1985, 1980, 1975	SHP	공개	전국	무료
센서스용 행정구역경계(읍면동)	2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1995, 1990, 1985, 1980, 1975	SHP	공개	전국	무료

Data preparation - load admin. boundaries (drag and drop)



Data preparation - zonal statistics



Data preparation - zonal statistics (results)

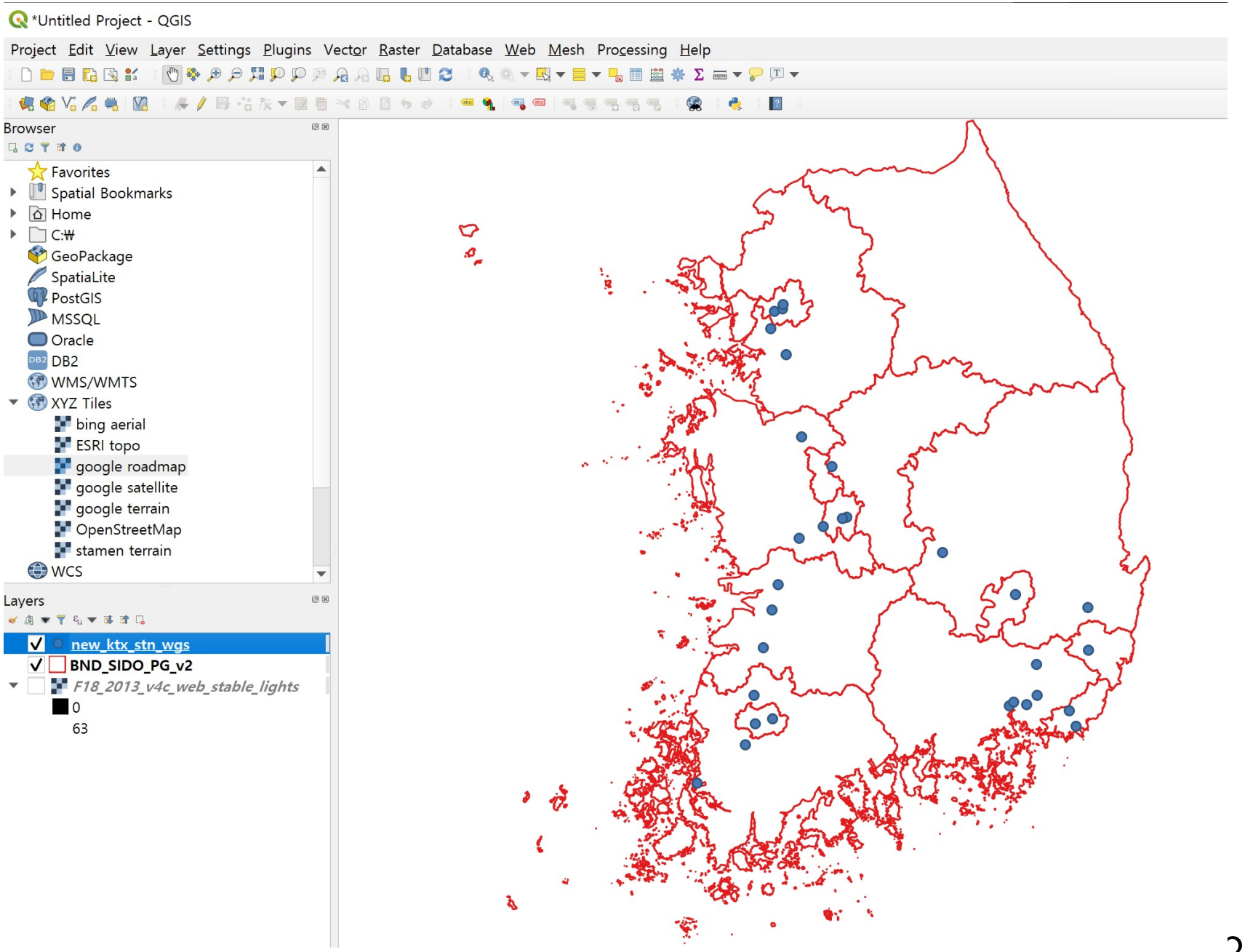
BND_SIDO_PG_v2 :: Features Total: 17, Filtered: 17, Selected: 1

	OBJECTID	BASE_YEAR	SIDO_CD	SIDO_NM	SIDO_NM_EN	_count	_sum	_mean	_median	_stdev
1	1	2015	11	서울특별시	Seoul	894.0000000000...	56146.00000000...	62.80313199105...	63.00000000000...	0.922834383977...
2	2	2015	21	부산광역시	Busan	1135.000000000...	67250.00000000...	59.25110132158...	63.00000000000...	7.227909448695...
3	3	2015	22	대구광역시	Daegu	1268.000000000...	60506.00000000...	47.71766561514...	54.00000000000...	16.49799061204...
4	4	2015	23	인천광역시		1701.000000000...	72724.00000000...	42.75367430922...	48.00000000000...	21.22525241581...
5	5	2015	24	광주광역시	Gwangju	714.0000000000...	34614.00000000...	48.47899159663...	56.00000000000...	16.07363068708...
6	6	2015	25	대전광역시	Daejeon	781.0000000000...	37611.00000000...	48.15749039692...	55.00000000000...	15.72016934716...
7	7	2015	26	울산광역시	Ulsan	1526.000000000...	61828.00000000...	40.51638269986...	41.00000000000...	18.75221420622...
8	8	2015	29	세종특별자치시	Sejong	678.0000000000...	25698.00000000...	37.90265486725...	37.00000000000...	13.69751044393...
9	9	2015	33	충청북도	Chungcheong N.	10764.00000000...	218648.0000000...	20.31289483463...	14.00000000000...	16.03393769907...
10	10	2015	34	충청남도	Chungcheong S.	12068.00000000...	287709.0000000...	23.84065296652...	19.00000000000...	14.95513834931...
11	11	2015	35	전라북도	Jeolla N.	11518.00000000...	233580.0000000...	20.27956242403...	15.00000000000...	14.76021755640...
12	12	2015	36	전라남도	Jeolla S.	18019.00000000...	281761.0000000...	15.63688328986...	10.00000000000...	13.94421573721...
13	13	2015	37	경상북도	Gyeongsang N.	27540.00000000...	462284.0000000...	16.78591140159...	11.00000000000...	14.68771728295...
14	14	2015	38	경상남도	Gyeongsang S.	15147.00000000...	334136.0000000...	22.05954974582...	14.00000000000...	17.92753745139...
15	15	2015	39	제주특별자치도	Jeju	2635.000000000...	51117.00000000...	19.39924098671...	14.00000000000...	14.49779147588...
16	16	2015	31	경기도	Gyeonggi	15284.00000000...	627223.0000000...	41.03788275320...	45.00000000000...	19.67272309940...
17	17	2015	32	강원도	Gangwon	24930.00000000...	296691.0000000...	11.90096269554...	8.000000000000...	12.86967984535...

Create explanatory variable:

Distance from the nearest KTX station

Data preparation - load ktx stations (digitization)



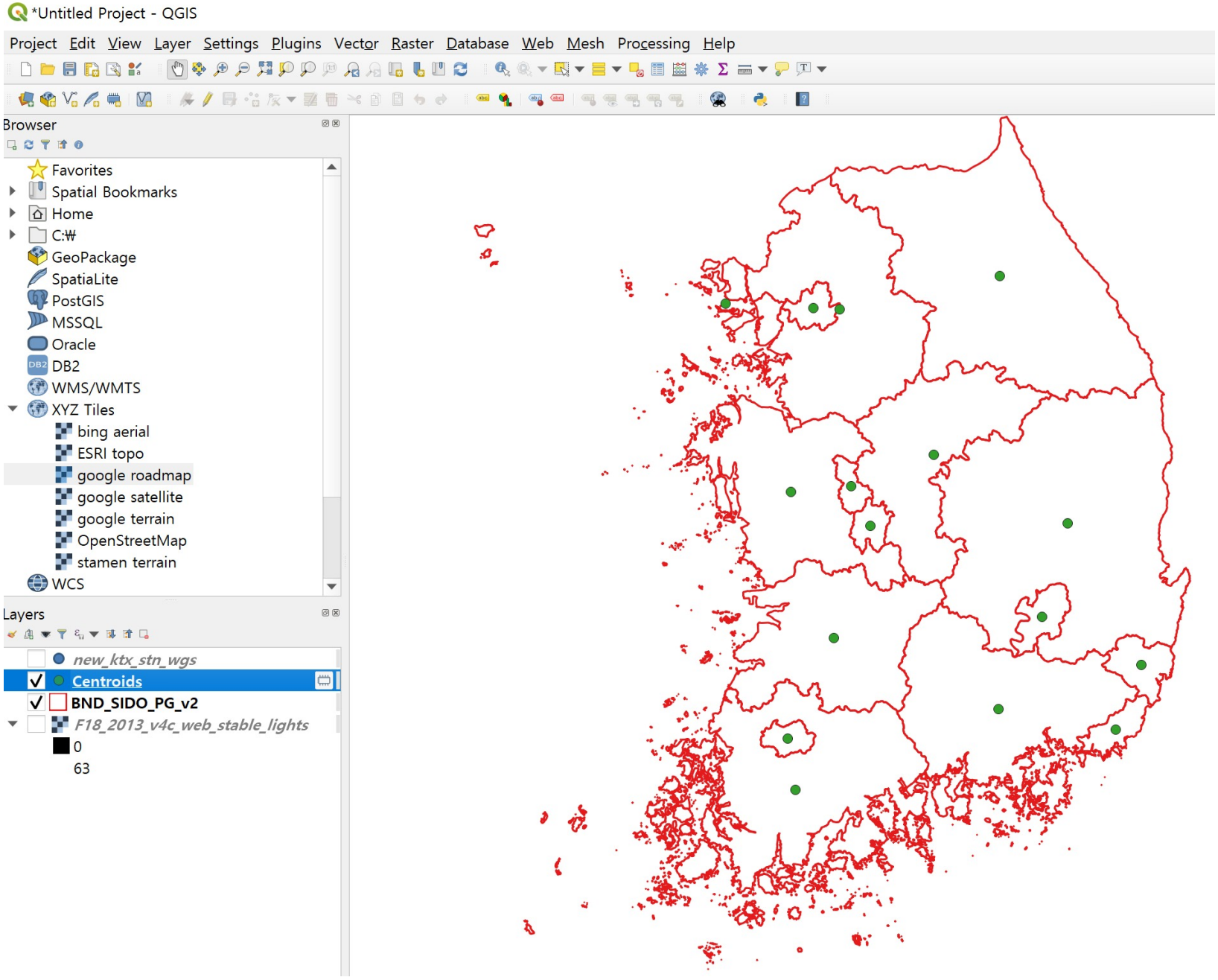
Data preparation - create centroids for admin units

Calculate linear distance (Point A to Point B)

KTX stations are points but admin. units are not

$\longrightarrow$ create points for admin units (centroids)

Data preparation - centroids for admin. units



Data preparation - calculate linear distance (distance matrix)

Data preparation - distance matrix calculation results

distance_matrix :: Features Total: 17, Filtered: 17, Selected: 0

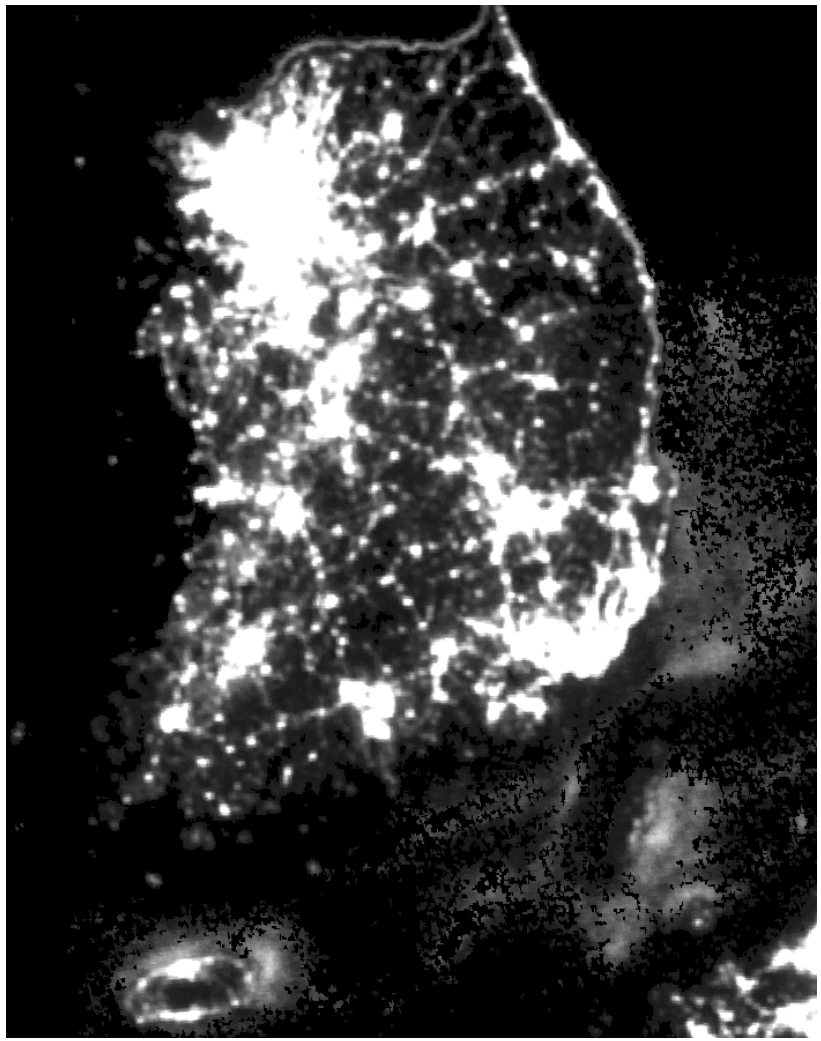
	InputID	TargetID	Distance
1	21	40	5481.527122396...
2	11	49	1883.815571582...
3	29	22	9129.183636982...
4	32	49	118742.0085239...
5	26	29	8930.817285475...
6	25	43	2000.324040809...
7	24	38	4547.781472031...
8	36	24	22035.69575824...
9	35	36	23285.04105391...
10	34	42	37470.36517924...
11	33	22	46800.49058272...
12	31	49	18006.80122811...
13	39	39	156508.5493409...

Final step

- Export created tables to CSV / Excel file
- Merge using ids with Stata
- Repeat for other years

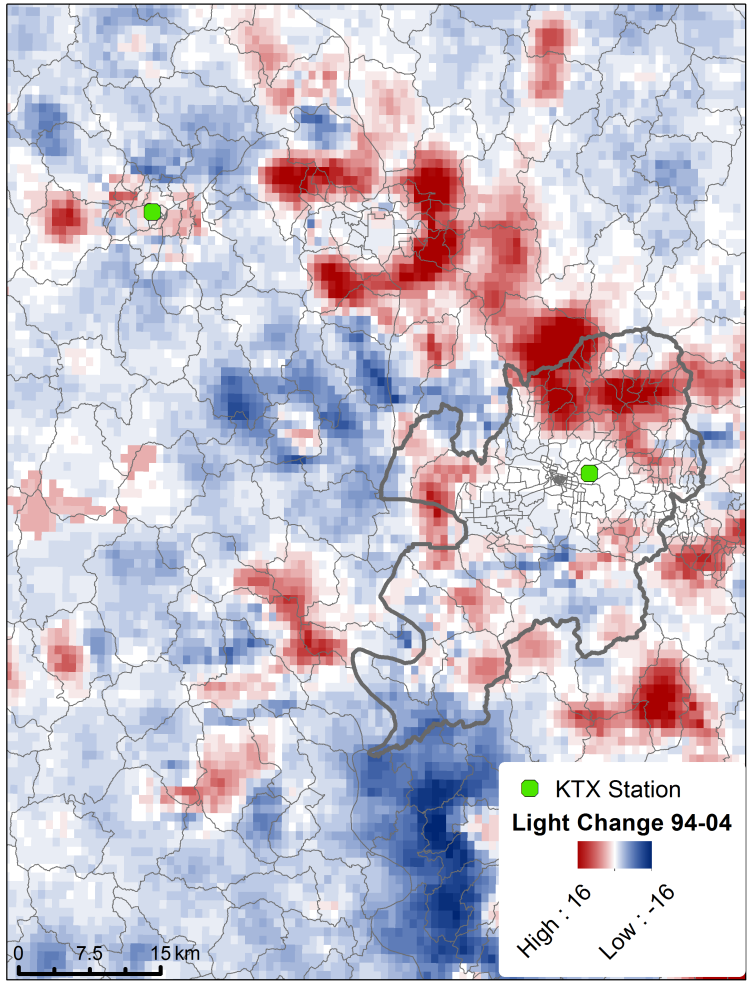
Raw Satellite Images 1994 & 2013

(e) 1994

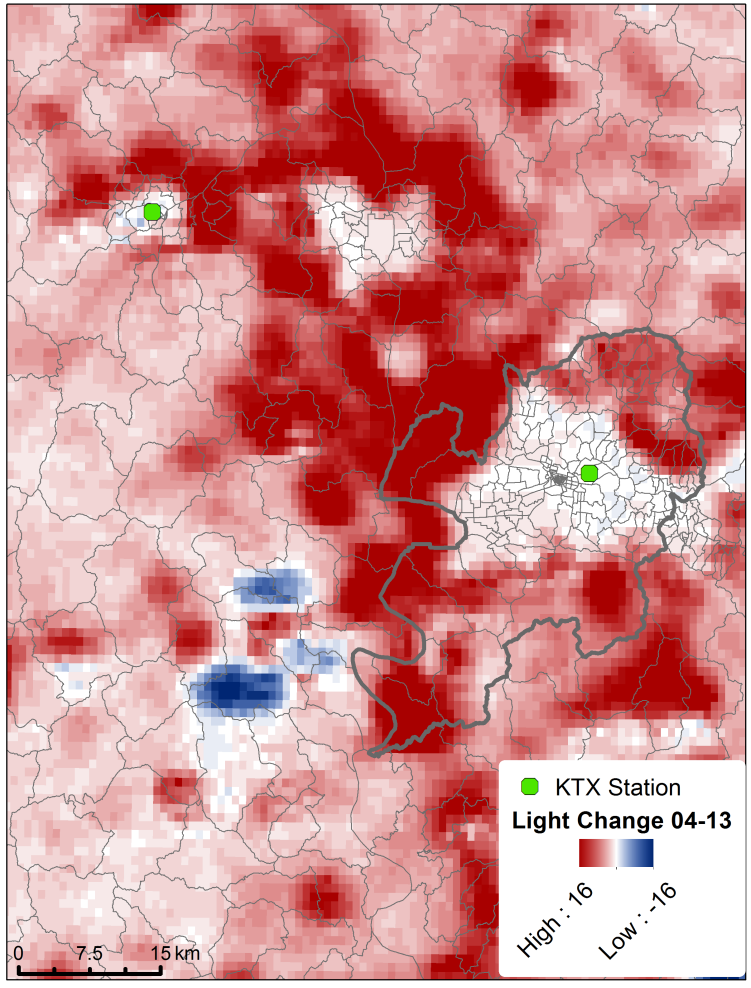


(f) 2013

Growth of Lights near Daegu & and Gimcheon



(g) 1994-2004



(h) 2004-2013

Empirical Strategy

We use difference in differences approach

- Two groups
  - Treatment: townships "closer" to KTX station
  - Control: townships "further away" from KTX station
  - $\longrightarrow$ definition of "closer" will be discussed
- Causal effect of treatment (introduction of KTX)
  - How economic activities of treatment group change compared to control group?

    $\longrightarrow$ Before and after the introduction of KTX
- Important assumption
  - The pre-trend of outcome should be parallel for both groups

Empirical Strategy - Model 1

Difference in differences approach (treatment = 0 or 1)

$$\begin{aligned} LIGHT_{k,m,t} = \alpha_0 &+ \alpha_1 TREAT_{k,m} + \alpha_2 POST_t \\ &+ \alpha_3 TREAT_{k,m} \cdot POST_t \\ &+ X'_{k,m,t}\Psi + \tau_k + \epsilon_{k,m,t} \end{aligned}$$

- $LIGHT_{k,m,t}$ = average light intensity of township m in county k, year t
- $TREAT_{k,m} = 1$ if KTX distance $\leq$ 34 kilometers (median value)
- $POST_t$ = post 2004 dummy
- τ_k = county fixed effects (No township effects yet)
- Distance measured from KTX stations in 2016
- Treatment was assumed to begin in 2004 (underestimated)

Identification

- Key identifying assumption: common trends
  - Trends of economic activity would be the same in both treatment and control groups in the absence of treatment (within the same county)
- Unobserved characteristics of rural units are inconsequential to the choice of routes
  - (Chandra and Thompson, 2000; Michaels, 2008; Redding and Turner, 2015)
- Endogenous selection of KTX stations
  - KTX stations were built where the increased economic activity is predicted
  - Policy makers to choose KTX stations in 1990 where the increase in economic activity in nearby townships would occur exactly 14 years later

Identification (cont.)

1. Did policy makers in 1990 predict the increased economic activity in 2004?
   - Planned operation year delayed multiple times due to financial constraints (Moon and Kim 2012)
   - Actual year of operation was 2004
2. Was the major determinant of the location of original KTX stations increased future economic activity? (Korea Rail Network Authority 2011; Moon and Kim 2012)
   - Not likely (e.g., Daejeon, Daegoo, Kyungjoo)
3. What about newly established KTX stations?
   - Began operation in 2010 (construction began in 2007, 2008)
   - New stations announced in 2003
   - Possible announcement effect (use Kangwon province to check)

Results 1

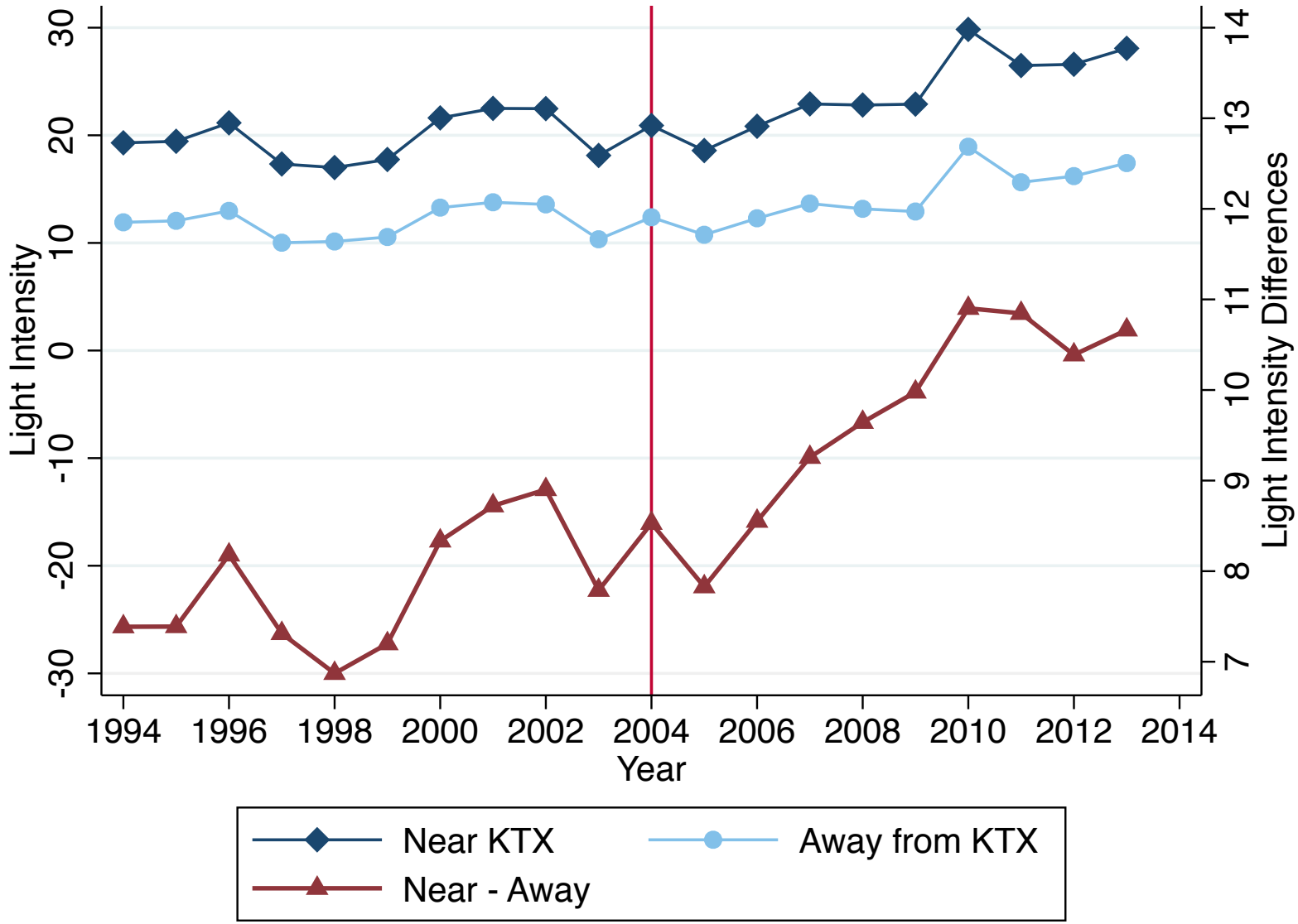

Light Intensity
30
20
10
0
-10
-20
-30
Light Intensity Differences
14
13
12
11
10
9
8
7
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year
Near KTX
Away from KTX
Near - Away

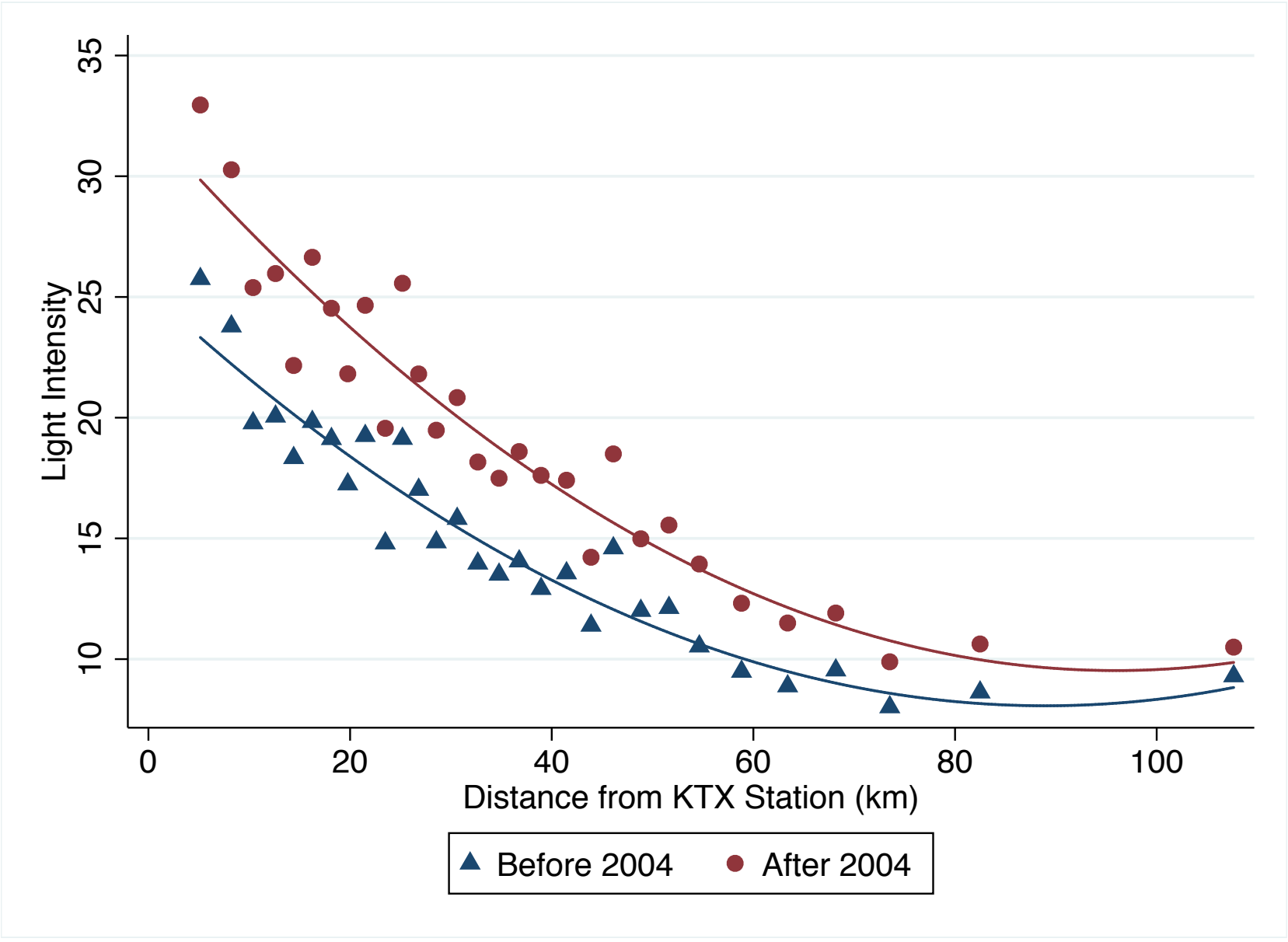
Table 1: Diff-in-Diffs Estimates of the Impact of the KTX on Light

	(1)	(2)	(3)	(4)
treatment × *post*	2.3***	2.3***	2.3***	2.3***
	(0.28)	(0.27)	(0.55)	(0.55)
Observations	25,300	25,300	25,300	25,300
R-squared	0.15	0.21	0.59	0.63
County FE			X	X
Covariate		X		X

Source: U.S. National Oceanic and Atmospheric Administration night time light data, 1994-2013. Standard errors are clusted at the county level.

*** p<0.01, ** p<0.05, * p<0.1.

KTX Distance and Light Intensity



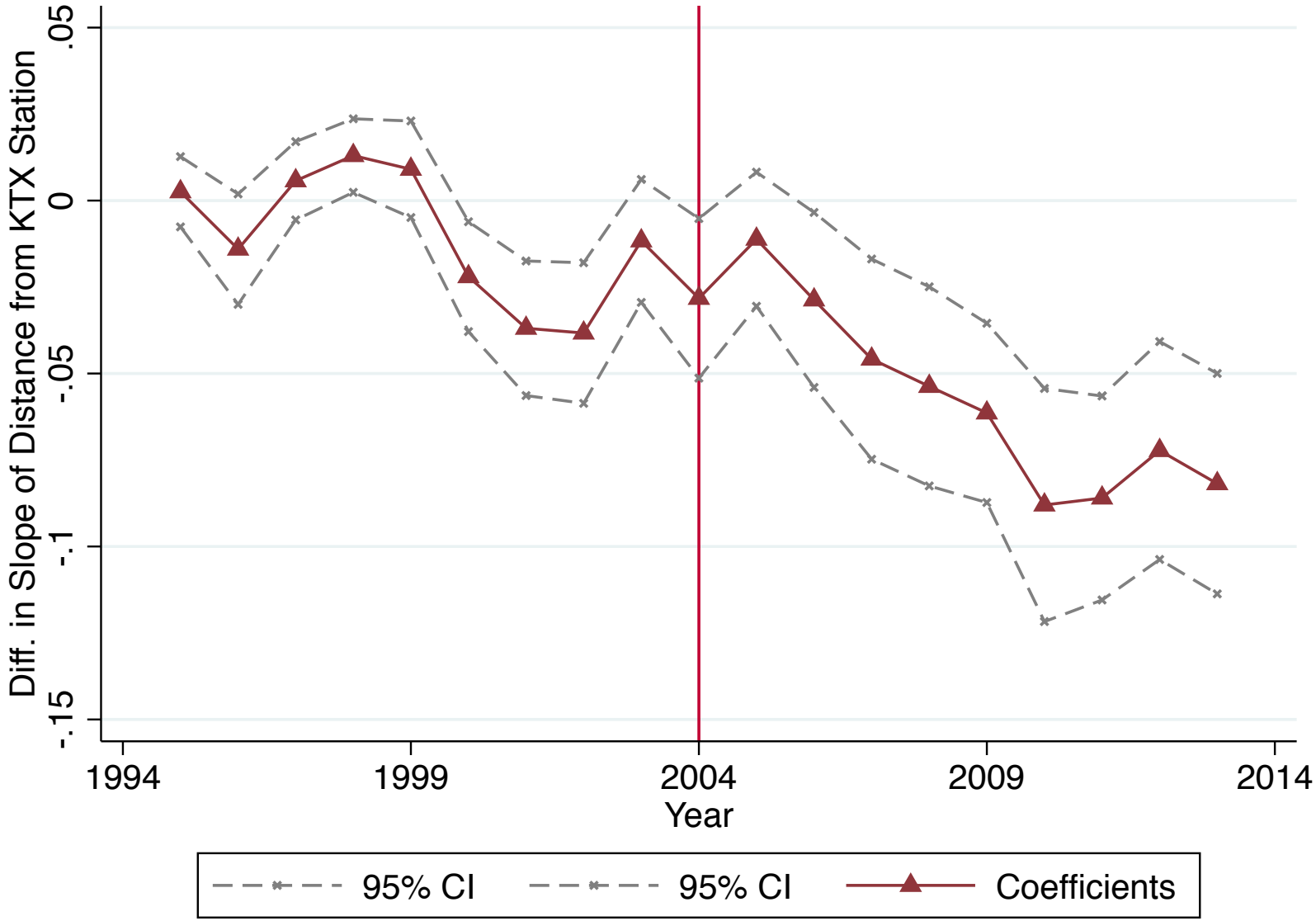
Empirical Strategy - Model 2

Difference in differences approach (continuous treatment)

$$\begin{aligned} LIGHT_{k,m,t} = \delta_0 + \delta_1 DIST_{k,m} + \gamma_t \\ + \textstyle\sum_t \lambda_t DIST_{k,m} \cdot \gamma_t \\ + X'_{k,m,t}\Theta + \sigma_k + \xi_{k,m,t} \end{aligned}$$

- $LIGHT_{k,m,t}$ = average light intensity in county k, township m, year t
- $DIST_{k,m}$ = KTX distance (kilometer)
- γ_t = year fixed effects
- σ_k = county fixed effects

Results 2


Diff. in Slope of Distance from KTX Station
.05
0
-.05
-.1
-.15
1994
1999
2004
2009
2014
Year
95% CI
95% CI
Coefficients

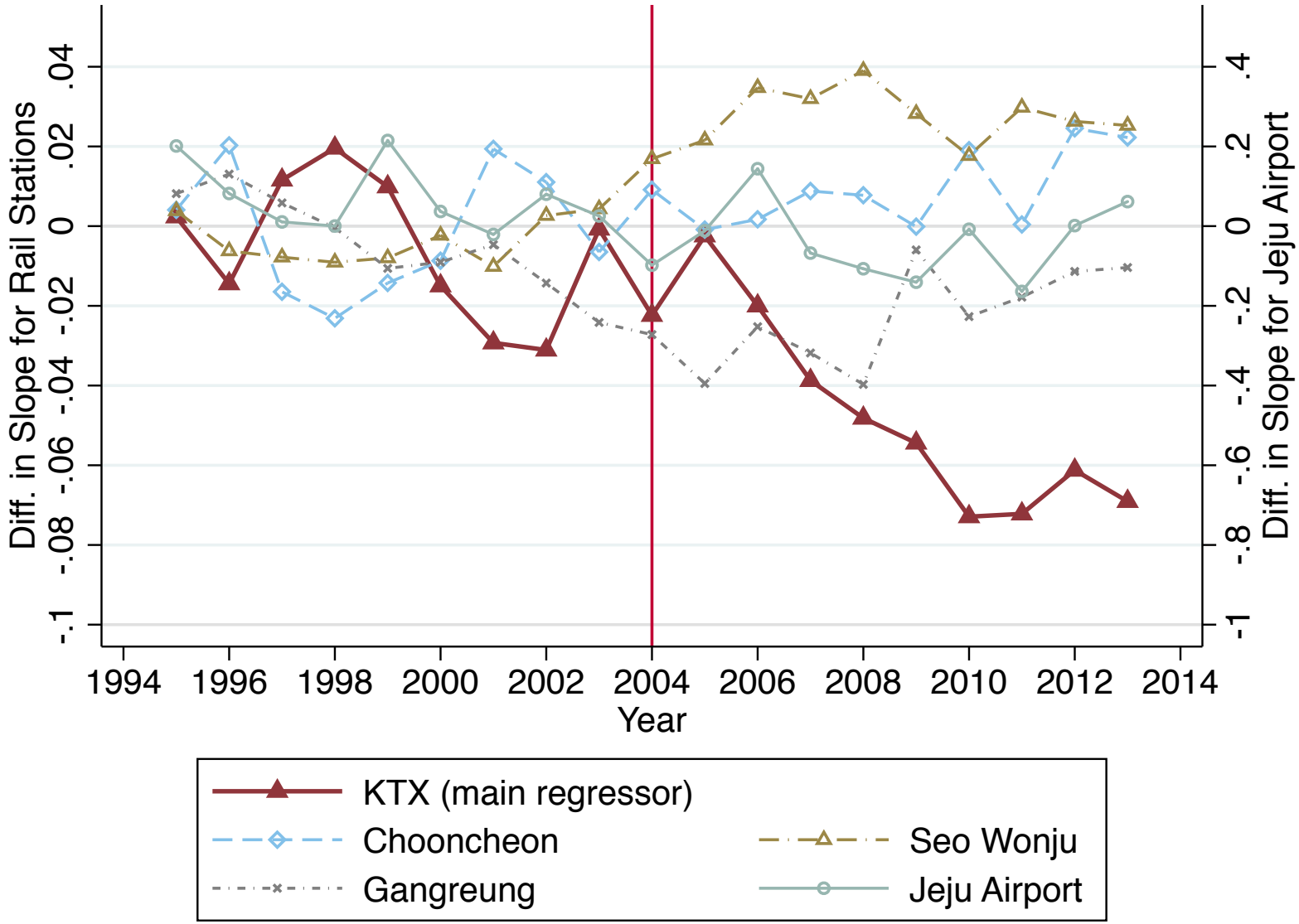
	(1) Coeff.	(1) S.E.	(2) Coeff.	(2) S.E.
1995	0.00	0.02	0.00	0.01
1996	-0.01	0.02	-0.01*	0.01
1997	0.01	0.01	0.01	0.01
1998	0.01	0.01	0.01**	0.01
1999	0.01	0.02	0.01	0.01
2000	-0.02	0.02	-0.02***	0.01
2001	-0.04**	0.02	-0.04***	0.01
2002	-0.04**	0.02	-0.04***	0.01
2003	-0.01	0.01	-0.01	0.01
2004	-0.03*	0.02	-0.03**	0.01
2005	-0.01	0.02	-0.01	0.01
2006	-0.03*	0.02	-0.03**	0.01
2007	-0.05***	0.02	-0.05***	0.01
2008	-0.05***	0.02	-0.05***	0.01
2009	-0.06***	0.02	-0.06***	0.01
2010	-0.09***	0.02	-0.09***	0.02
2011	-0.09***	0.02	-0.09***	0.01
2012	-0.07***	0.02	-0.07***	0.02
2013	-0.08***	0.02	-0.08***	0.02
Observations	25,300		25,300	
R-squared	0.21		0.70	
County FE			X	
Covariate			X	

Source: U.S. National Oceanic and Atmospheric Administration night time light data, 1994-2013. Standard errors are clusted at the county level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Falsification Checks

- Increased economic activities near rail stations after 2004
- Use Gangwon Province without any KTX connections
  - Distance to major rail stations and future KTX stations
- Use Jeju Island
  - Proximity to Jeju International Airport

Results 2


Diff. in Slope for Rail Stations
-.1
-.08
-.06
-.04
-.02
0
.02
.04
Diff. in Slope for Jeju Airport
-1
-.8
-.6
-.4
-.2
0
.2
.4
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year
KTX (main regressor)
Chooncheon
Gangreung
Seo Wonju
Jeju Airport

Conclusions

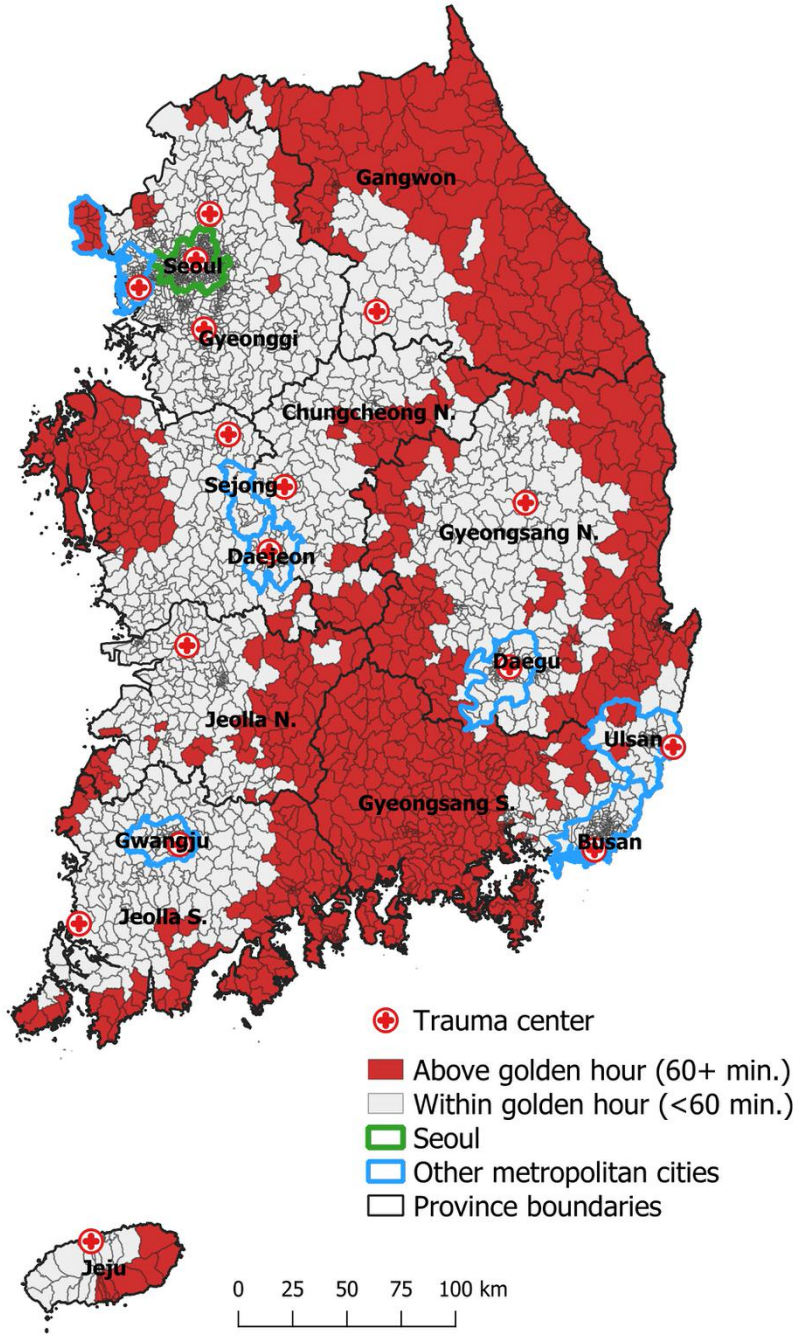
Summary

- Evaluation the impact of HSR system on local economic activity
- Introduction of novel panel data on local economic activity

Future work

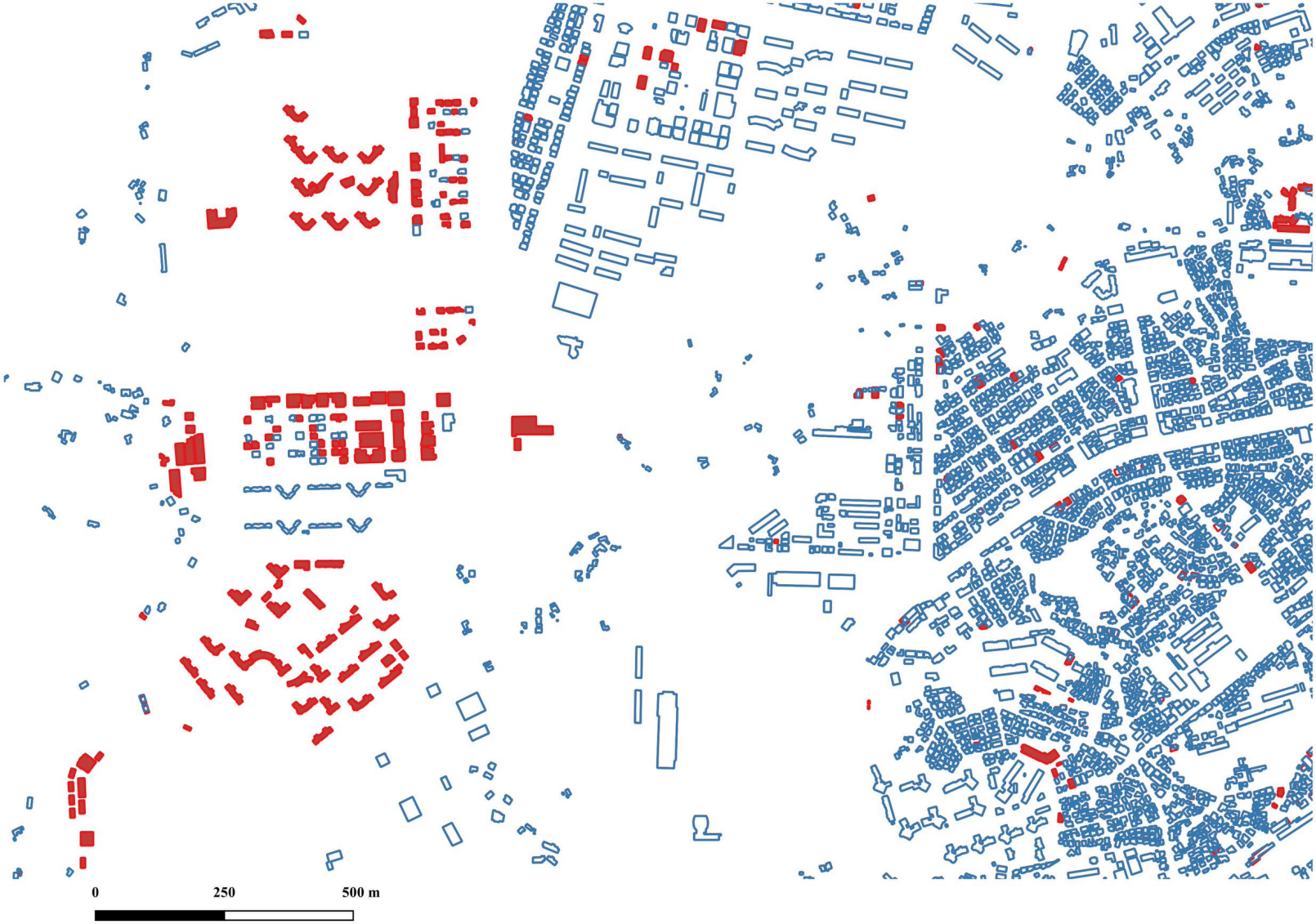
- Economic growth or reorganization?
  - Currently collecting data on population, business, etc.
- Mechanisms?
  - Housing booms, new jobs, region-based policies
  - Entry of firms, employment, productivity
  - Plan to use population and land price data
- From light intensity to $ value

Medical accessibility: trauma centers

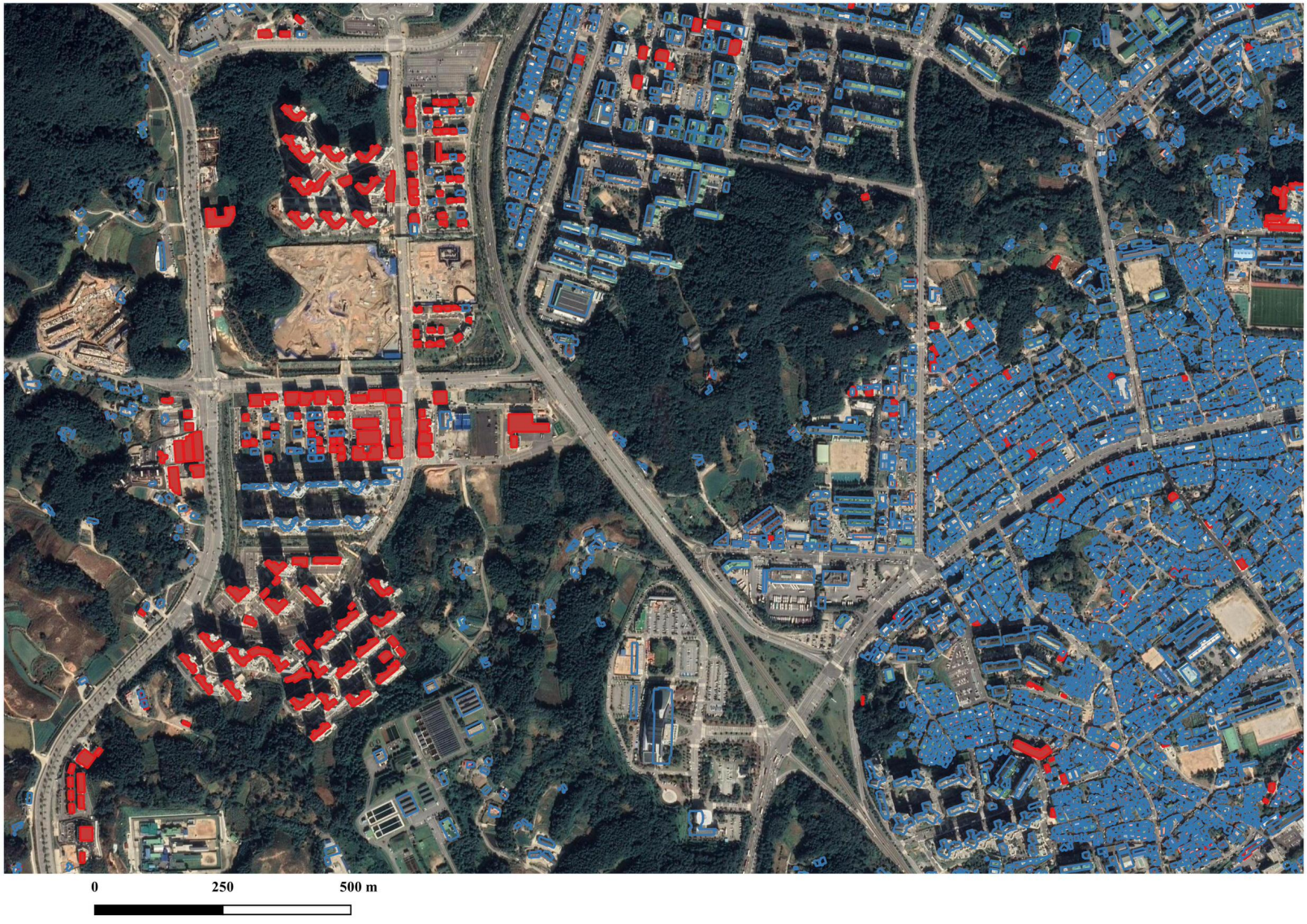


강릉시 유천동 (평창올림픽 이전과 이후)

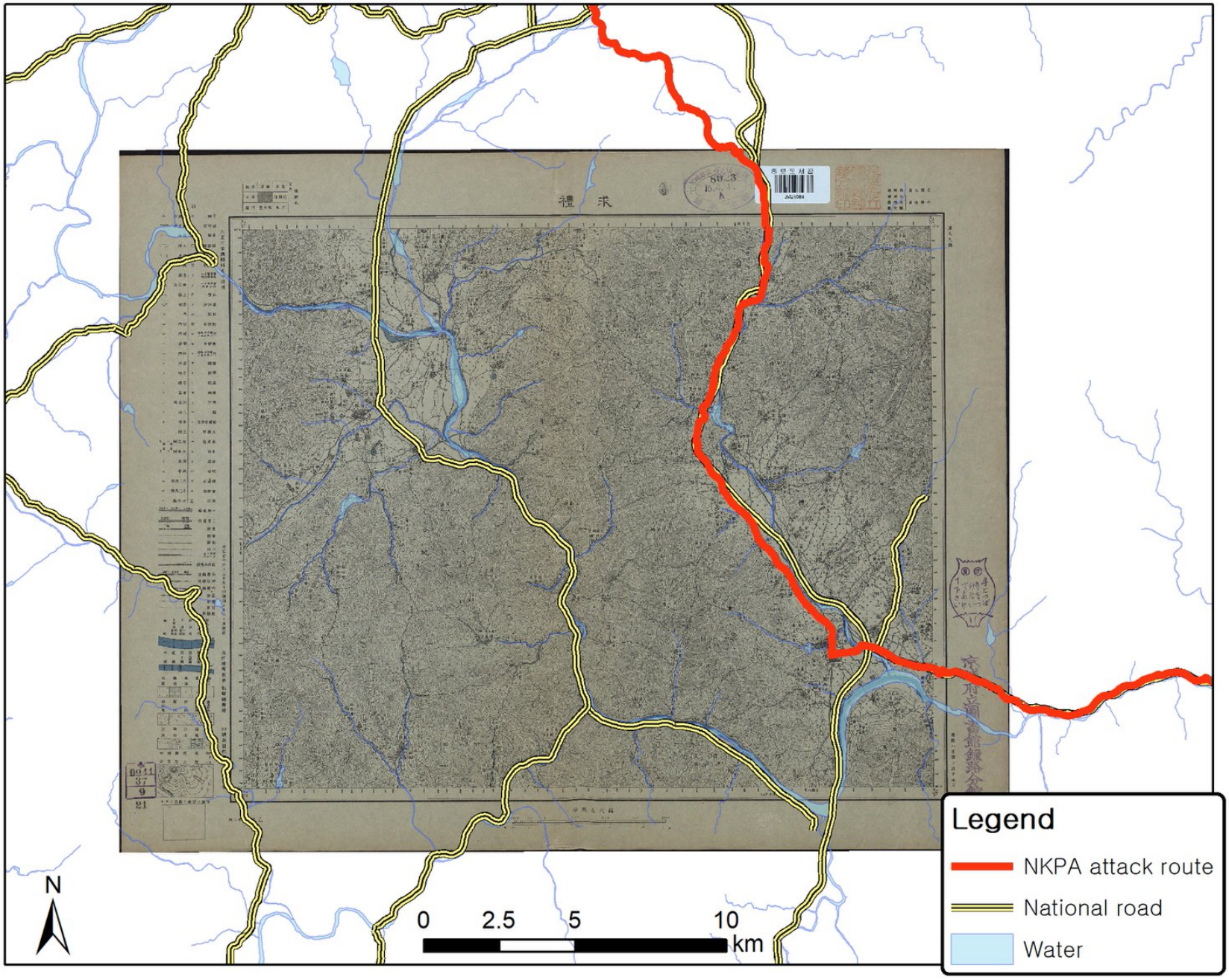
행안부 건물 데이터를 통해 올림픽 이후 새로 생긴 건물 identify



행안부 건물 데이터를 통해 올림픽 이후 새로 생긴 건물 identify

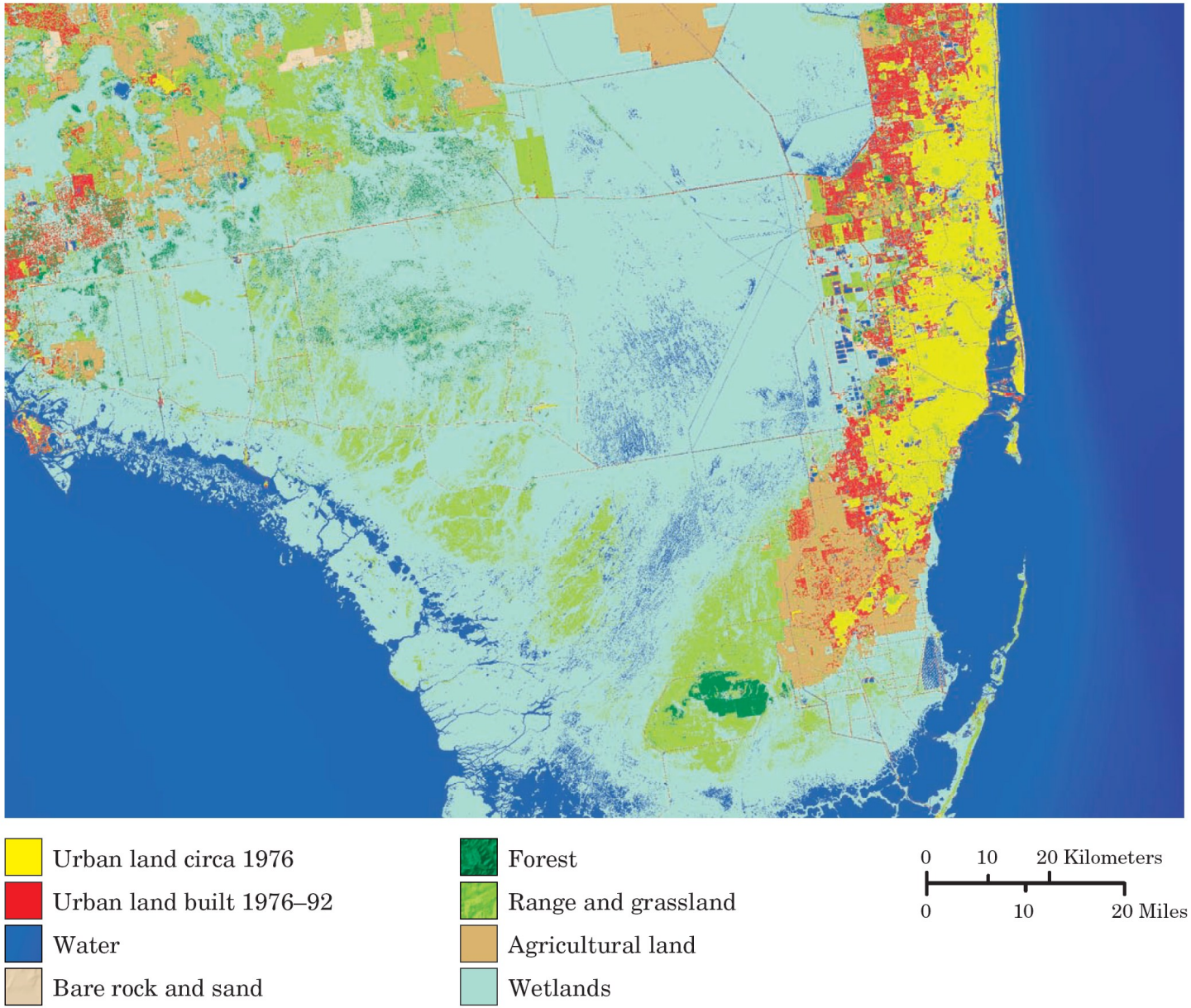


Digitizing 1918 road networks & NKPA attack routes



Burchfield et al. 2006. **QJE**

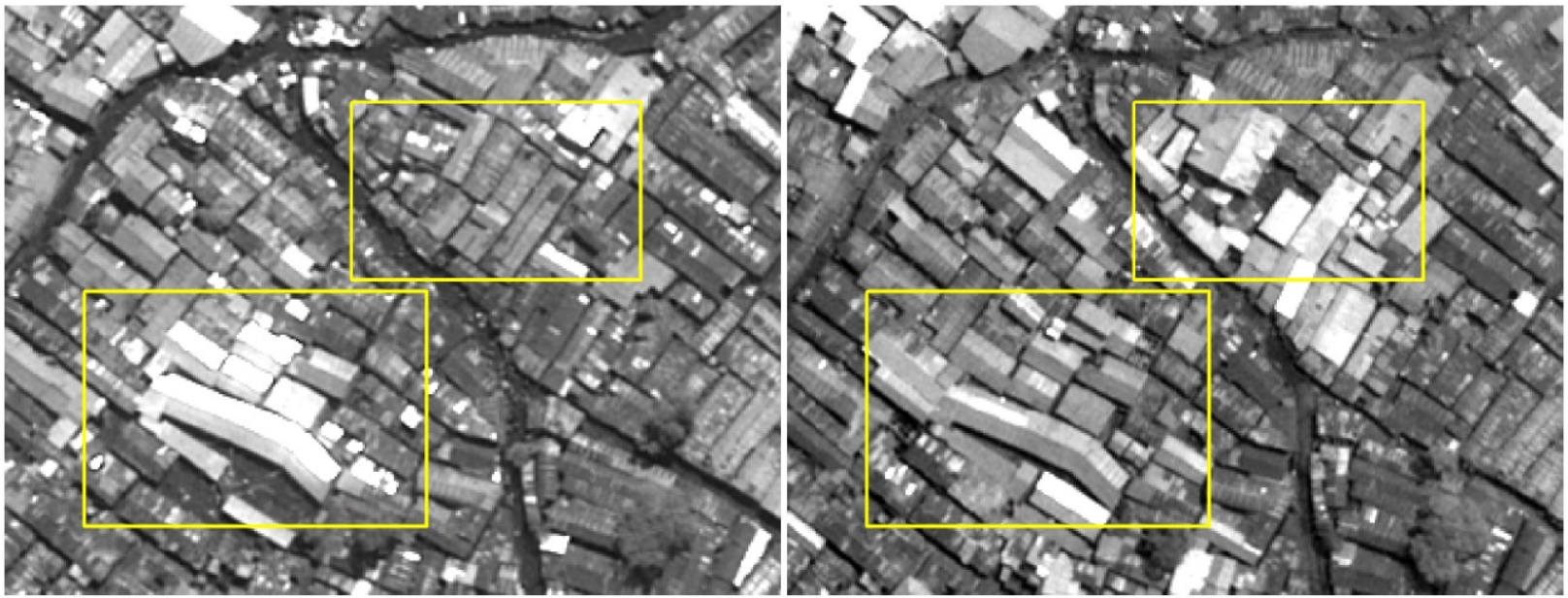
Urban expansion in Miami (1976-1992)



Marx et al. 2019. **AEJ: Applied Economics**

Roof types as a proxy for poverty

Appendix Figure A2: Old and New Roofs in Kibera



Note: Both pictures are taken over the same area of the slum with the same resolution (0.5 meters panchromatic).
The picture in the left panel was taken in July 2009 and that in the right panel in August 2012.
The yellow rectangles highlight clusters of roofs that markedly evolved over the period.
Roofs highlighted in the bottom rectangle degraded while roofs within the top rectangle were upgraded in the same timeframe.
The picture area is approximately 175 meters long and 140 meters wide.

Shiny new roofs v.s. old roofs in villages in Kenya

Varshney et al. 2015. **Big Data**

FIG. 1. Sample satellite image patches of the region with **(a)** thatched roofs in the center and **(b)** a metal roof in the center.

Varshney et al. 2015. **Big Data**

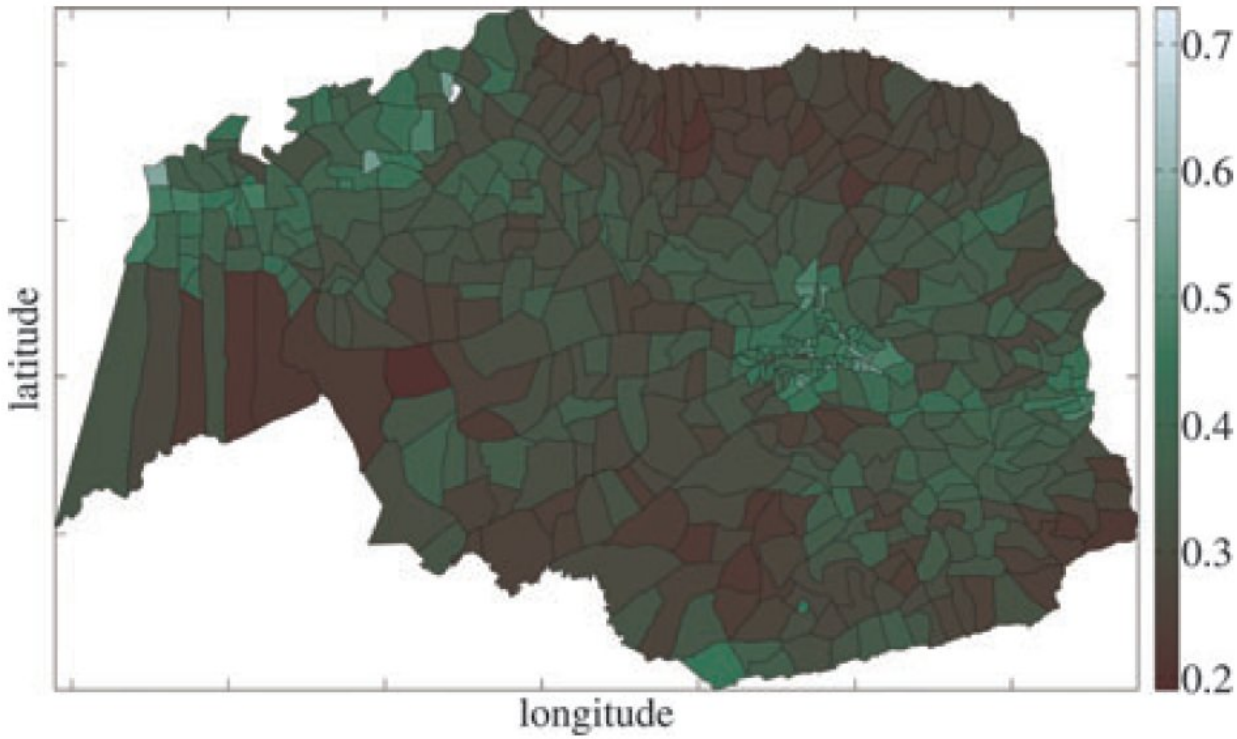


FIG. 3. Estimated proportion of roofs that are metal per village.

Science

RESEARCH

RESEARCH ARTICLES

ECONOMICS

Combining satellite imagery and machine learning to predict poverty

Neal Jean,[1,2*] Marshall Burke,[3,4,5*†] Michael Xie,[1] W. Matthew Davis,[4] David B. Lobell,[3,4] Stefano Ermon[1]

Reliable data on economic livelihoods remain scarce in the developing world, hampering efforts to study these outcomes and to design policies that improve them. Here we demonstrate an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Using survey and satellite data from five African countries—Nigeria, Tanzania, Uganda, Malawi, and Rwanda—we show how a convolutional neural network can be trained to identify image features that can explain up to 75% of the variation in local-level economic outcomes. Our method, which requires only publicly available data, could transform efforts to track and target poverty in developing countries. It also demonstrates how powerful machine learning techniques can be applied in a setting with limited training data, suggesting broad potential application across many scientific domains.

Jean et al. 2016. **Science**

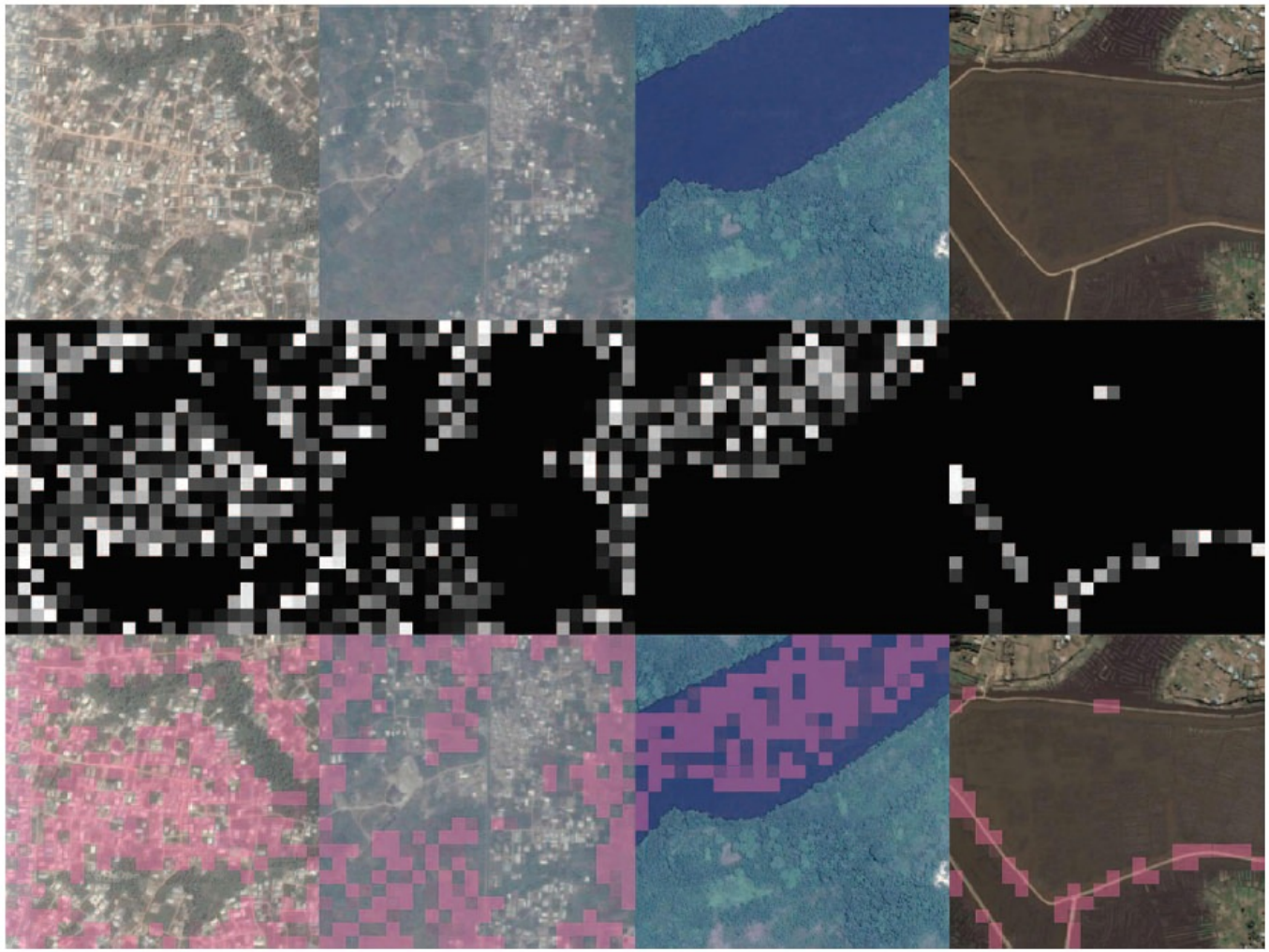


Fig. 2. Visualization of features. By column: Four different convolutional filters (which identify, from left to right, features corresponding to urban areas, nonurban areas, water, and roads) in the convolutional neural network model used for extracting features. Each filter "highlights" the parts of the image that activate it, shown in pink. By row: Original daytime satellite images from Google Static Maps, filter activation maps, and overlay of activation maps onto original images

Jean et al. 2016. **Science**

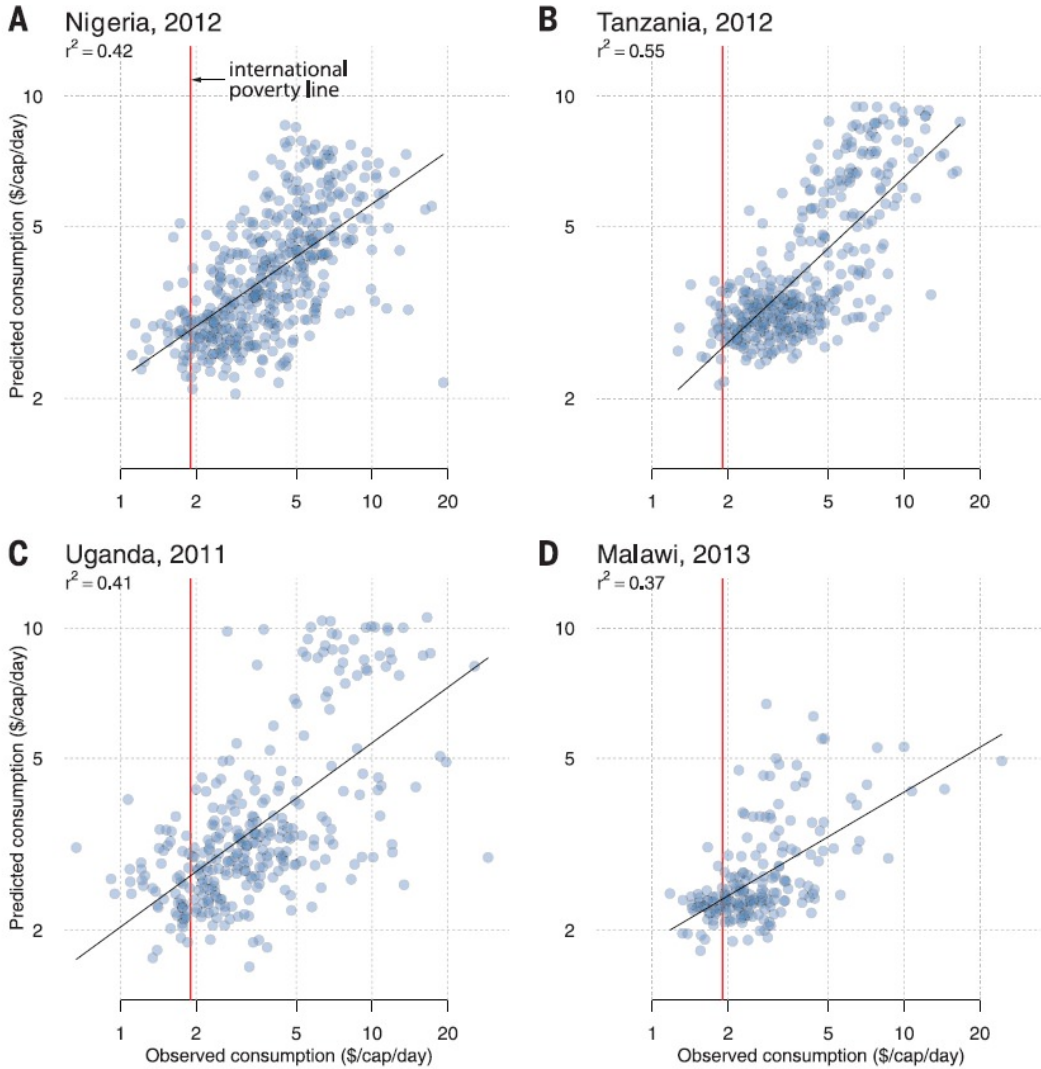
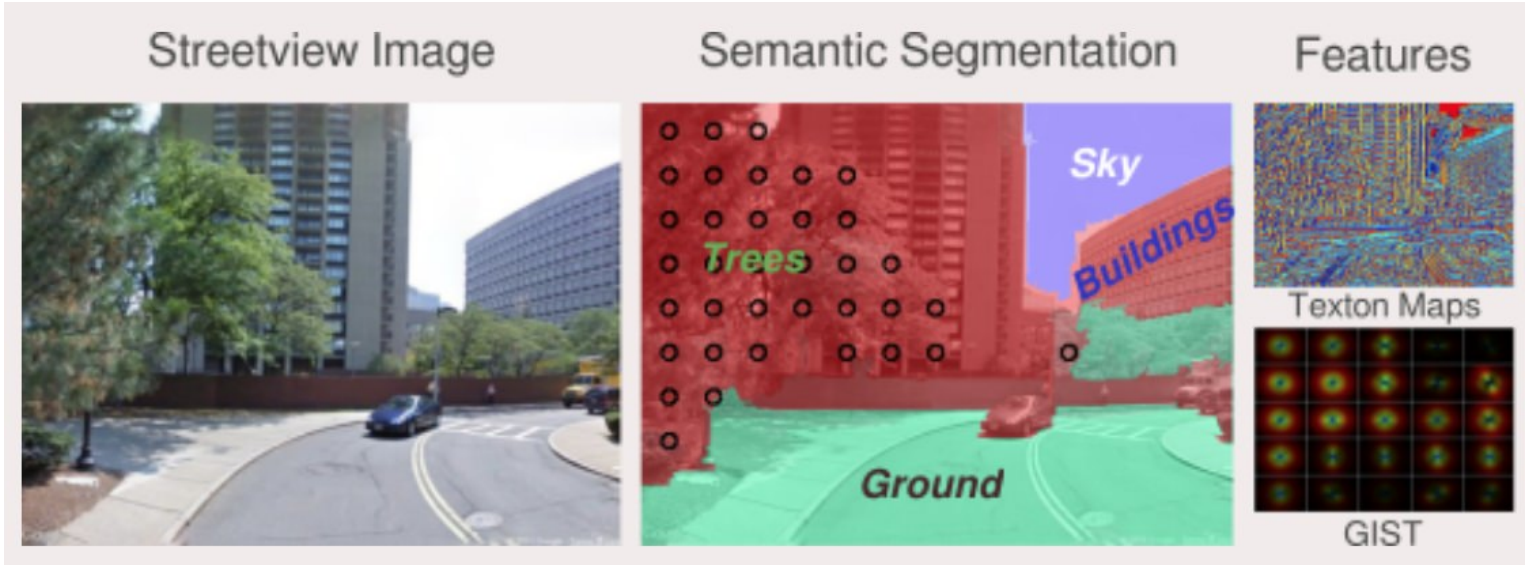


Fig. 3. Predicted cluster-level consumption from transfer learning approach (*y* axis) compared to survey-measured consumption (*x* axis). Results are shown for Nigeria (**A**), Tanzania (**B**), Uganda (**C**), and Malawi (**D**). Predictions and reported r^2 values in each panel are from fivefold cross-validation. Black line is the best fit line, and red line is international poverty line of $1.90 per person per day. Both axes are shown in logarithmic scale. Countries are ordered by population size.

Naik et al. 2017 **PNAS**



Naik et al. 2017 **PNAS**