Measuring inequalities from space

Analysis of satellite raster images with R

Piotr Wójcik
pwojcik@wne.uw.edu.pl
Data Science Lab, dslab.wne.uw.edu.pl

UseR! 2019
Toulouse – France
July 9-12, 2019

UNIVERSITY OF WARSAW
Faculty of Economic Sciences

My bio

- **assistant professor** at the University of Warsaw, Faculty of Economic Sciences (FES UW) since 2008
- **founder and head** of Data Science Lab at FES UW: dslab.wne.uw.edu.pl
- analysis of night-time lights intensity data and day-time satellite images is **one of our projects** in DSLab
- many years of professional experience as a **quantitative analyst** in the financial, telecommunications and marketing research sector
- my LinkedIn profile: www.linkedin.com/in/pwojcik


Night-time lights intensity (NTLI) data

- NTLI data is based on **satellite images** collected and processed by the National Oceanic and Athmosferic Administration (NOAA)
- NOAA provides **two types** of NTLI data:
  - Version 4 **DMSP-OLS** – average annual data for the period 1992–2013
  - Version 1 **VIIRS** – monthly data since April 2012 and averaged annual (only 2015 and 2016)
- NTLI is measured for **pixels** with the size of 30×30 (DMSP-OLS) or 15x15 (VIIRS) arc seconds
- it relates to **less than 1 km2** on the equator (about 0.5 km2 in Europe or USA)
- for each pixel NTLI data is provided in **digital numbers** (DN) on the scale 0–63 (DMSP-OLS) or 0–16384 (VIIRS)


DSLab

Night-time lights intensity (NTLI) data

- NTLI data is based on **satellite images** collected and processed by the National Oceanic and Athmosferic Administration (NOAA)
- NOAA provides **two types** of NTLI data:
  - Version 4 **DMSP-OLS** – average annual data for the period 1992–2013
  - Version 1 **VIIRS** – monthly data since April 2012 and averaged annual (only 2015 and 2016)
- NTLI is measured for **pixels** with the size of 30×30 (DMSP-OLS) or 15x15 (VIIRS) arc seconds
- it relates to **less than 1 km2** on the equator (about 0.5 km2 in Europe or USA)
- for each pixel NTLI data is provided in **digital numbers** (DN) on the scale 0–63 (DMSP-OLS) or 0–16384 (VIIRS)


DSLab

Economic and business applications of NTLI data

- can be easily **agregated for any territorial units**
- **uniformly measured** across the globe
- **independent** of politicians and response rates in surveys
- increasingly used as a **proxy of economic activities** at the regional and local level
- researchers find **strong positive relationship** between NTLI and GDP, and population at the **national level**
- **at the subnational level** the relationship is usually **weaker and unstable**
- it has informational value for **countries with poor quality of national income accounts**
- proxies for **economic well-being** or **market potential** can be calculated for **non-administrative areas** (e.g. **specicic business regions**)

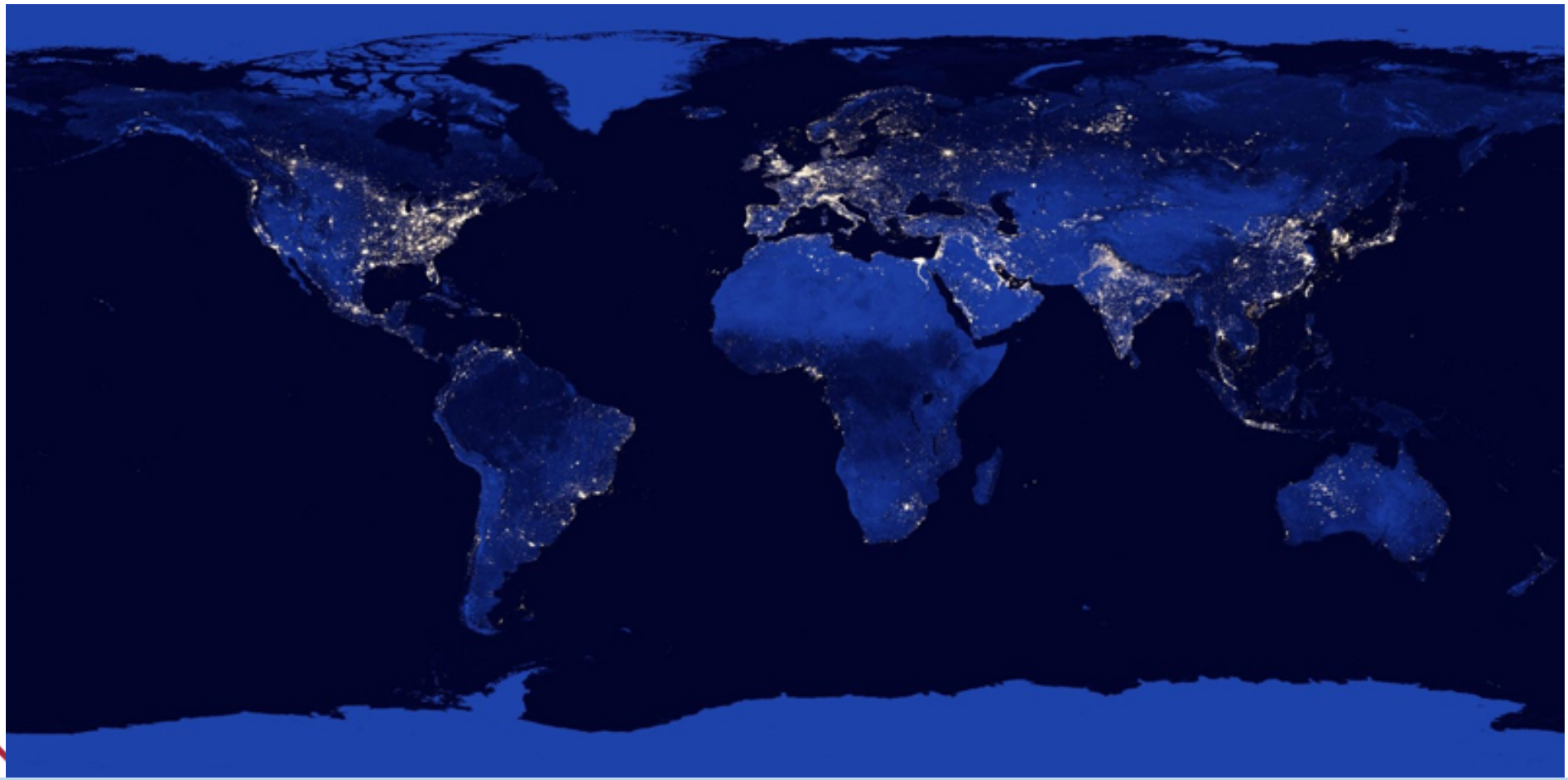
DSLab

Economic and business applications of NTLI data

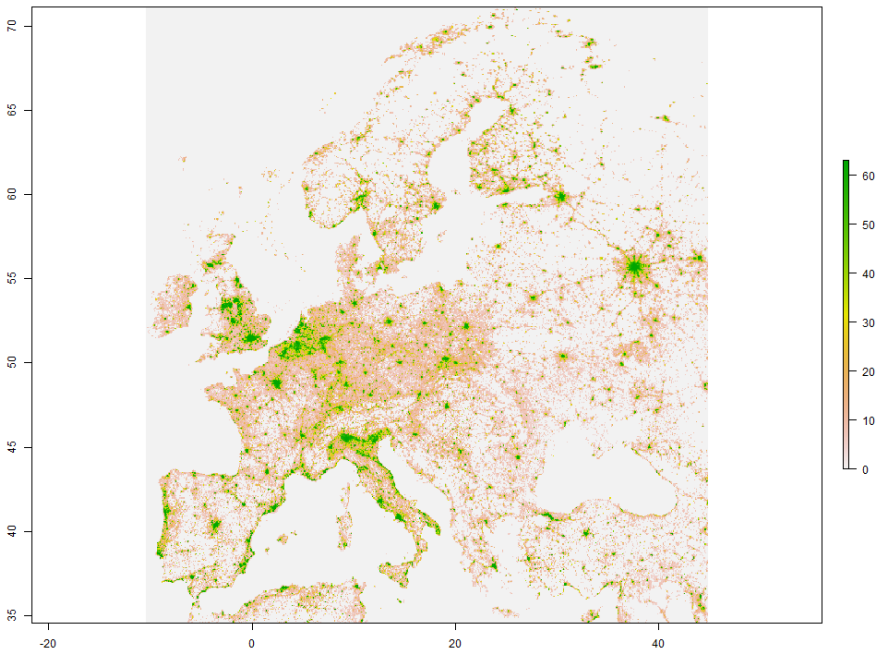
- can be easily **agregated for any territorial units**
- **uniformly measured** across the globe
- **independent** of politicians and response rates in surveys
- increasingly used as a **proxy of economic activities** at the regional and local level
- researchers find **strong positive relationship** between NTLI and GDP, and population at the **national level**
- **at the subnational level** the relationship is usually **weaker and unstable**
- it has informational value for **countries with poor quality of national income accounts**
- proxies for **economic well-being** or **market potential** can be calculated for **non-administrative areas** (e.g. **specicic business regions**)

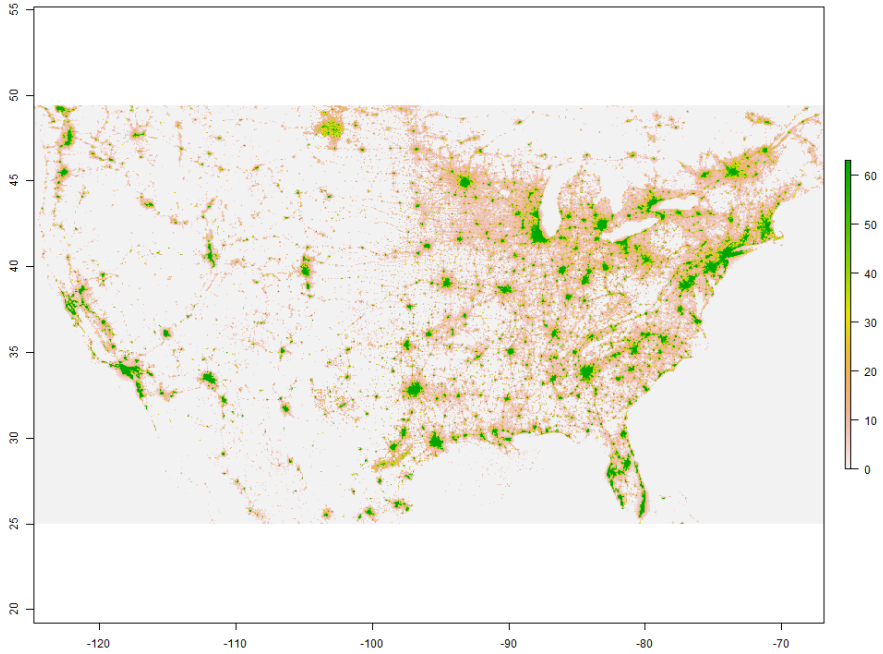
Let the NTLI data speak

- DMSP-OLS data for 2013 used in examples
- **full codes available on**: github.com/ptwojcik/UseR2019



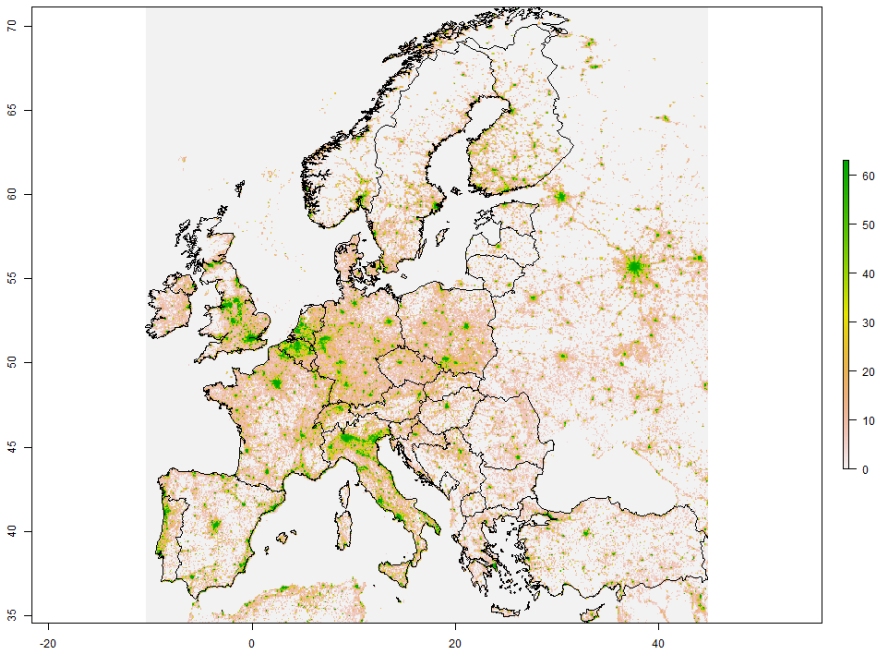
Step 1. importing raster data and map into R

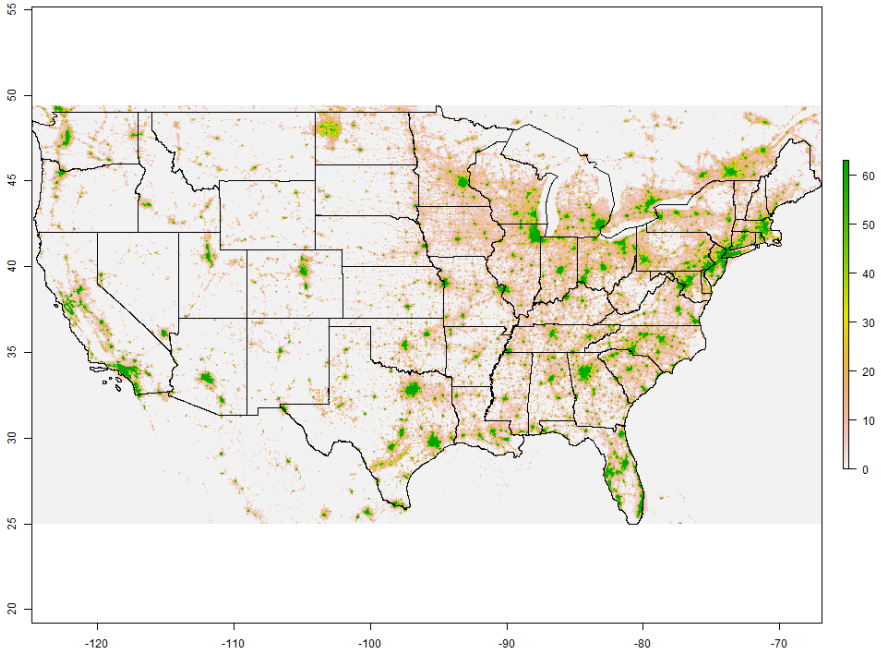




- **importing NTLI data** with the `raster::raster()`
- **importing map data** (e.g. shapefile) with `sf::st_read()`

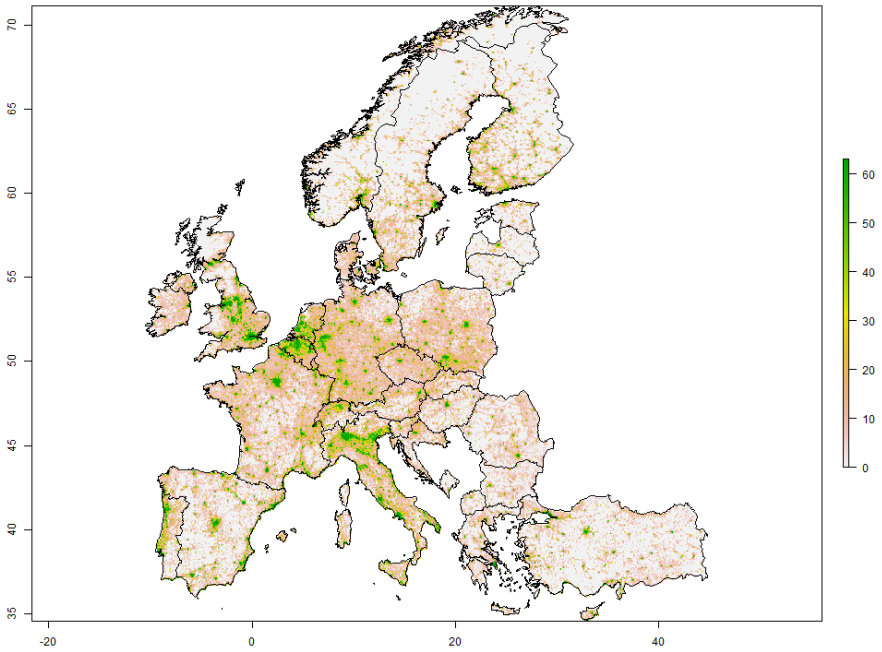
Step 2. checking and adjusting

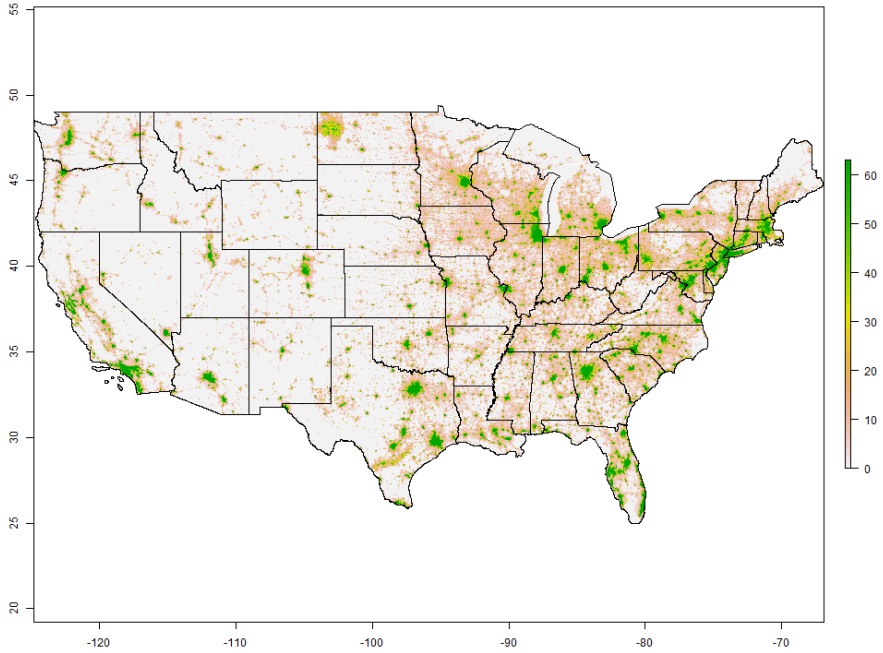




- **checking and adjusting projection** of spatial coordinates in both objects with `raster::crs()` and `sf::st_transform()`
- **checking validity** of `sf` object geometries with `sf::st_is_valid()` and adjusting if needed with `lwgeom::st_make_valid()`

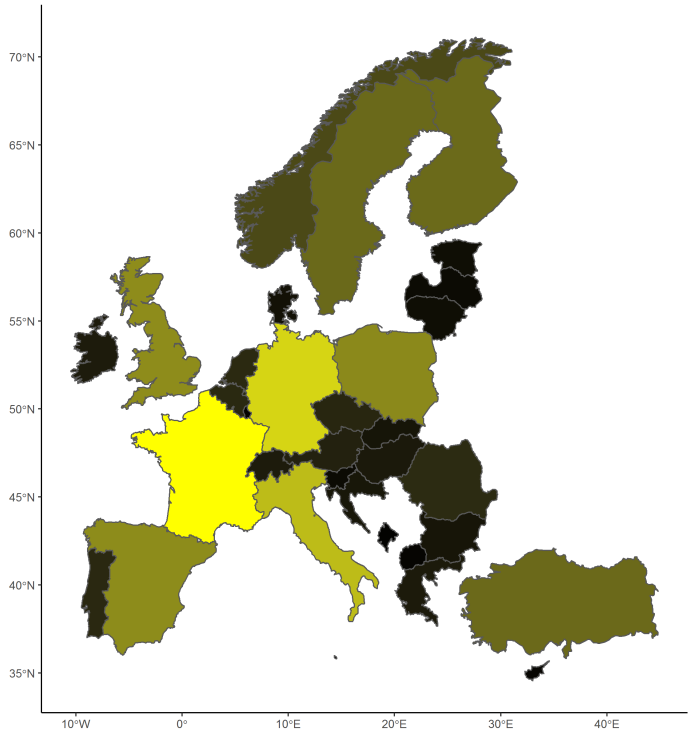
Step 3. limitation of the extent of raster data

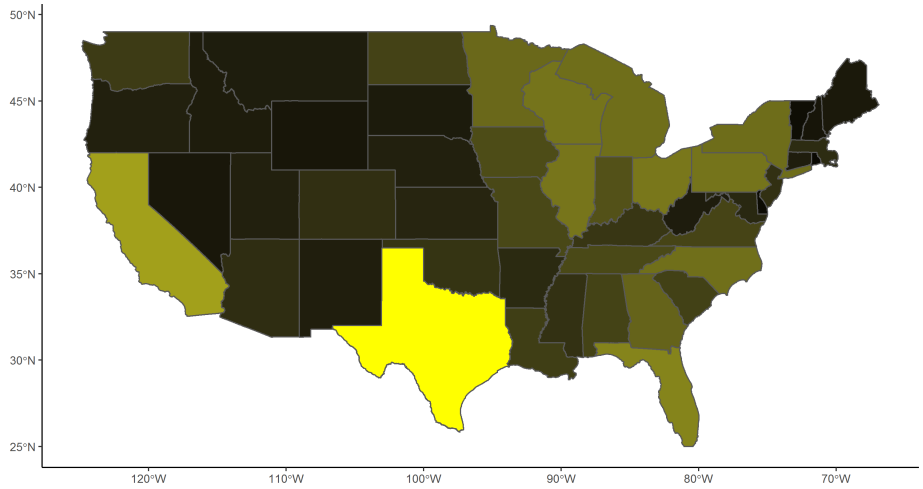




- **limitation of raster data** to the extent of the `sf` object with `raster::extent()`, `raster::crop()` and `raster::mask()`

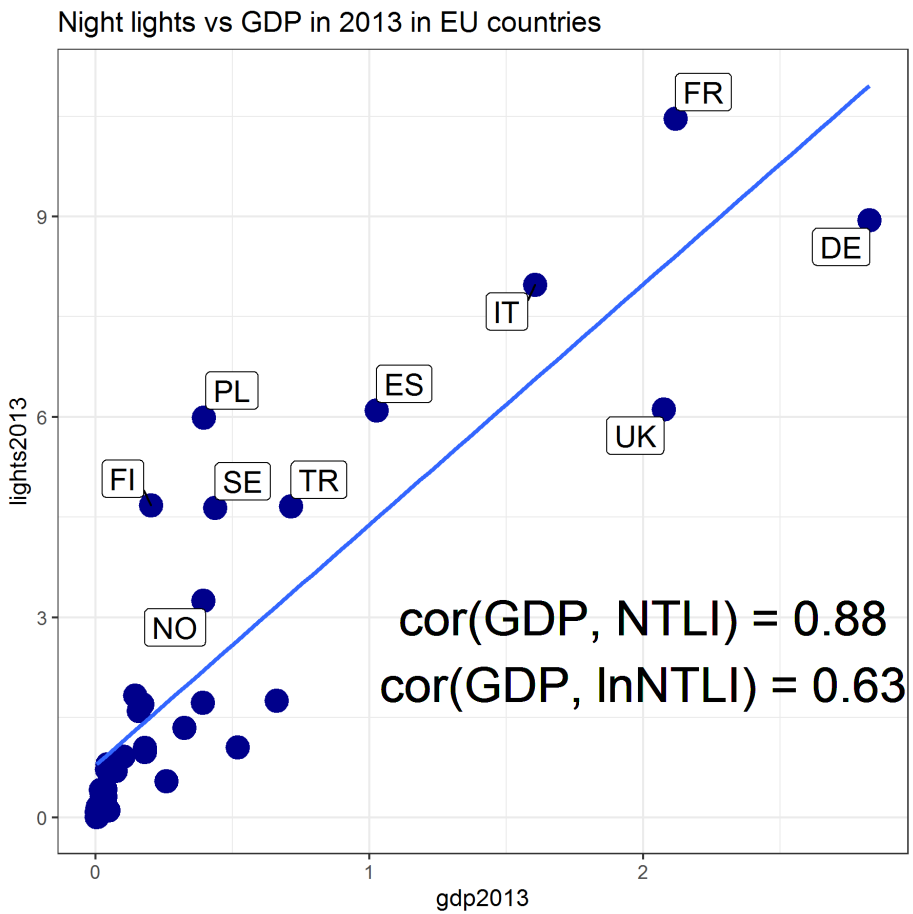
Step 4. aggregation to spatial units

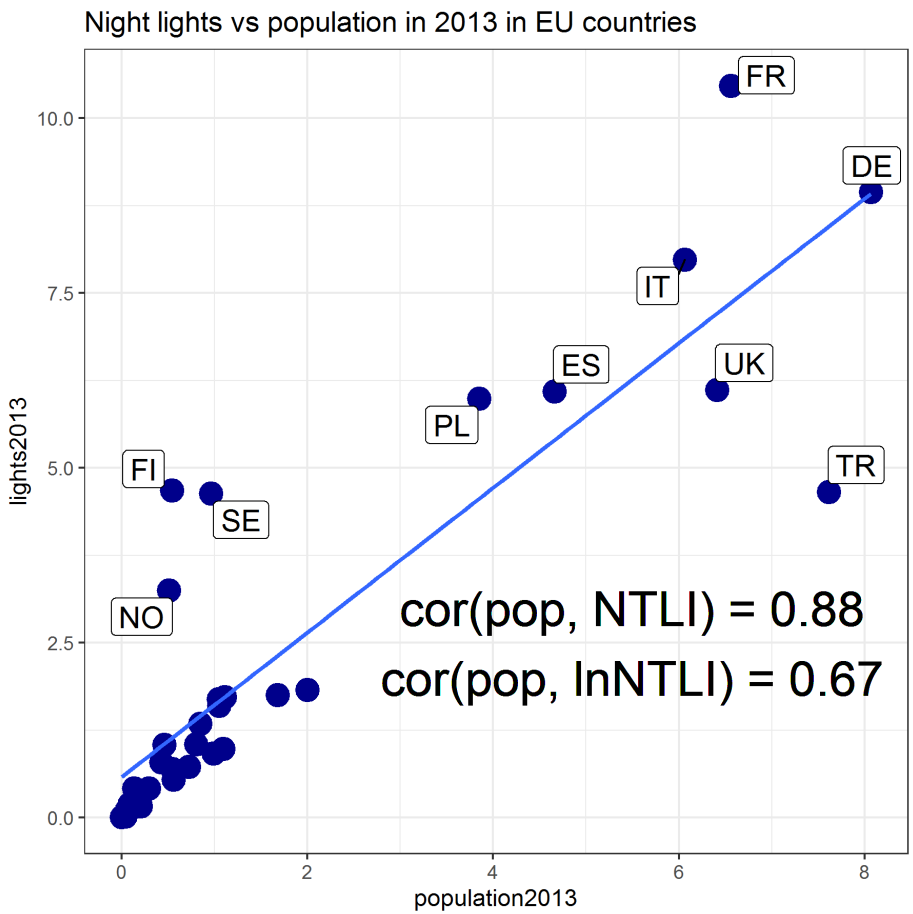




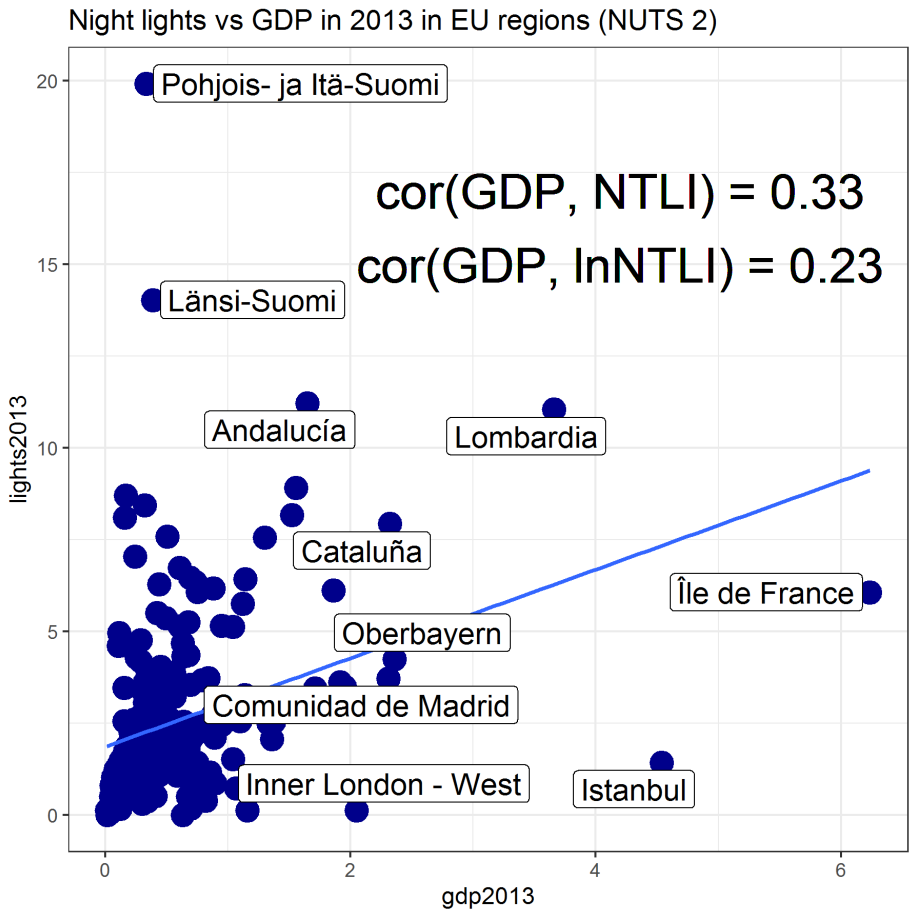
- **aggregation of NTLI** data for spatial units – `raster::extract()`, **much faster if also** `tabularaster::cellnumbers()` used

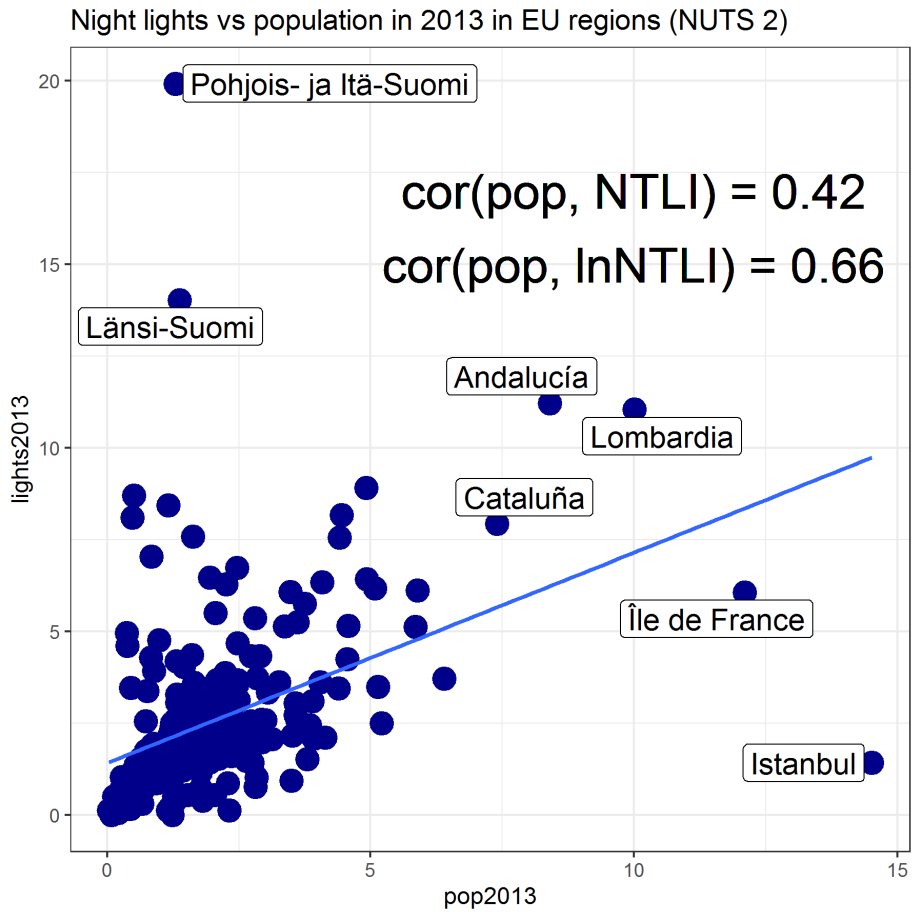
Correlation with GDP and population for EU countries



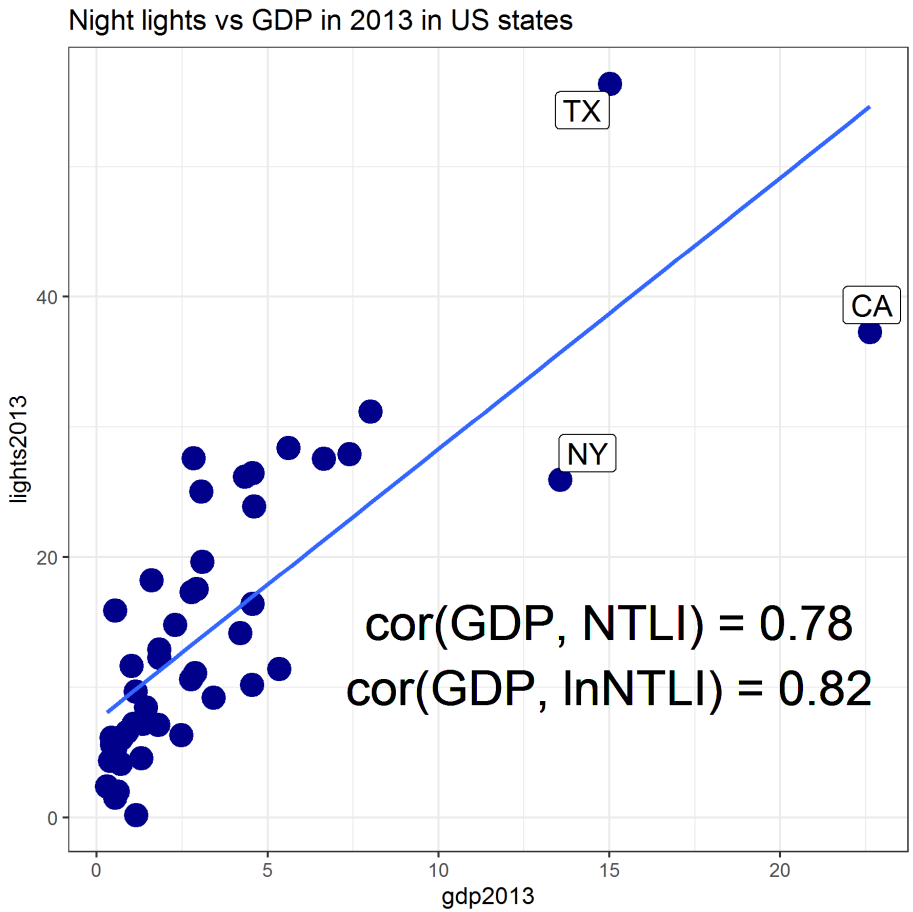


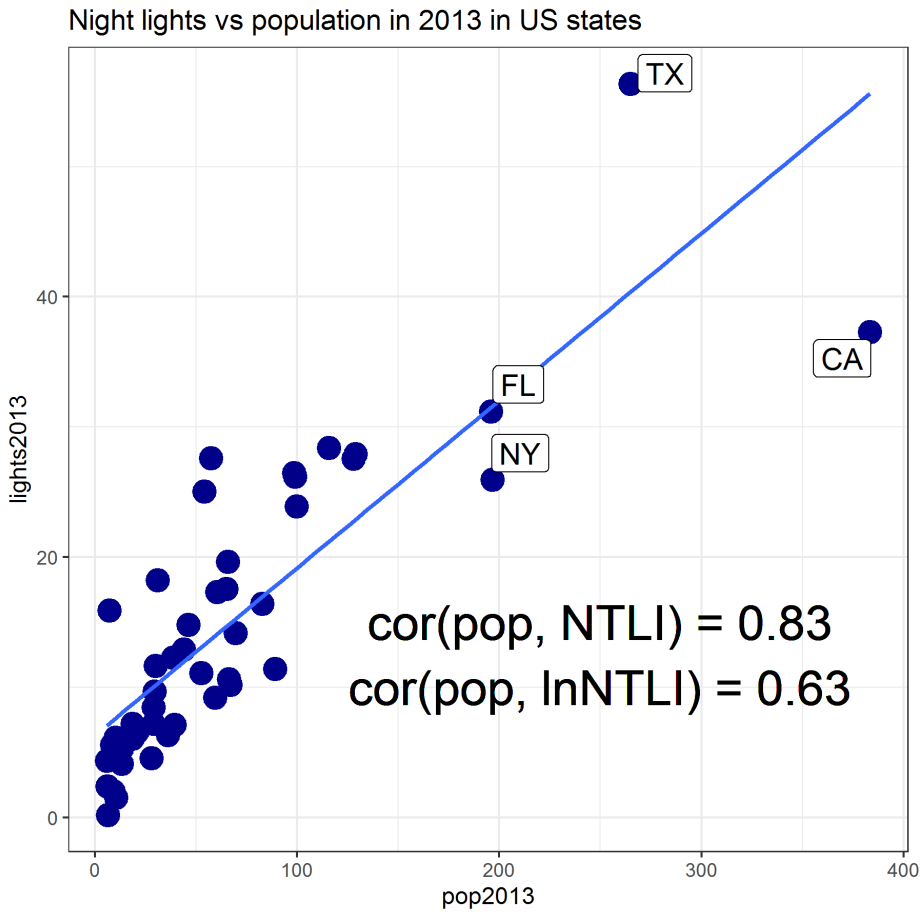
Correlation with GDP and population for EU regions





Correlation with GDP and population for US states





Summary and thank you

- **R has great tools** for raster data and spatial data analysis
- NTLI is a **promising proxy** for market potential or economic well-being
- however, the **relationship is not linear**, urban areas are specific
- **building a model** predicting well-being based on NTLI data **required**
- modern applications use also **high-resolution daytime satellite images** and **image recognition tools**
- **full codes for above examples available on**: github.com/ptwojcik/UseR2019

THANK YOU FOR YOUR ATTENTION!