From Pixels to Geospatial Insights:

IEG's Experience Leveraging Image Data in Evaluations

Virginia Ziulu

Data Scientist
Independent Evaluation Group, World Bank

IEG Independent Evaluation Group | WORLD BANK GROUP
World Bank • IFC • MIGA

1 | Image Data in Evaluations

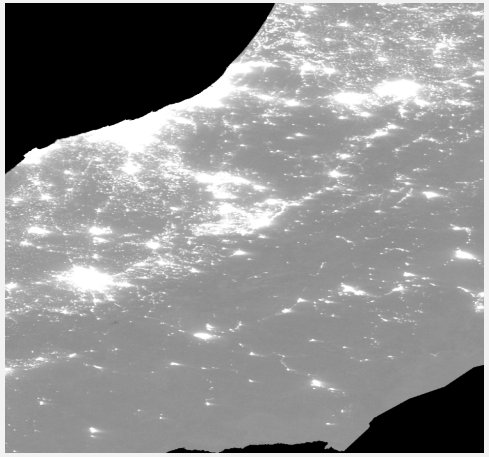
Imagery is a ubiquitous data source

- Images are everywhere.
- 1.81 trillion photos are taken worldwide every year.
- There are 9,000+ satellites actively orbiting the Earth (April 2024).

Image source: Andrej Karpathy, Stanford Computer Science.

There is a wealth of imagery available for Geospatial Analysis

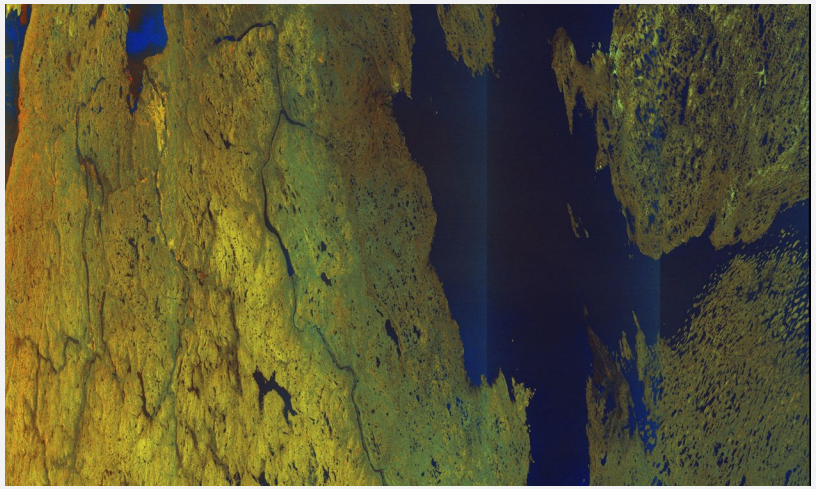
Optical Daylight Satellite Image
(Source: Google Earth)



Optical Nighttime Satellite Image
(Source: VIIRS)

Streetscape Digital Photo
(Source: Mapillary)

Aerial (Drone) Imagery
(Source: Dar Rapid Transit Agency)



Synthetic Aperture Radar (SAR) Image
(Source: ERS-2, ASF DAAC)

Semantic segmentation using convolutional neural networks.
Source: UPSNet-101-M Cityscapes.

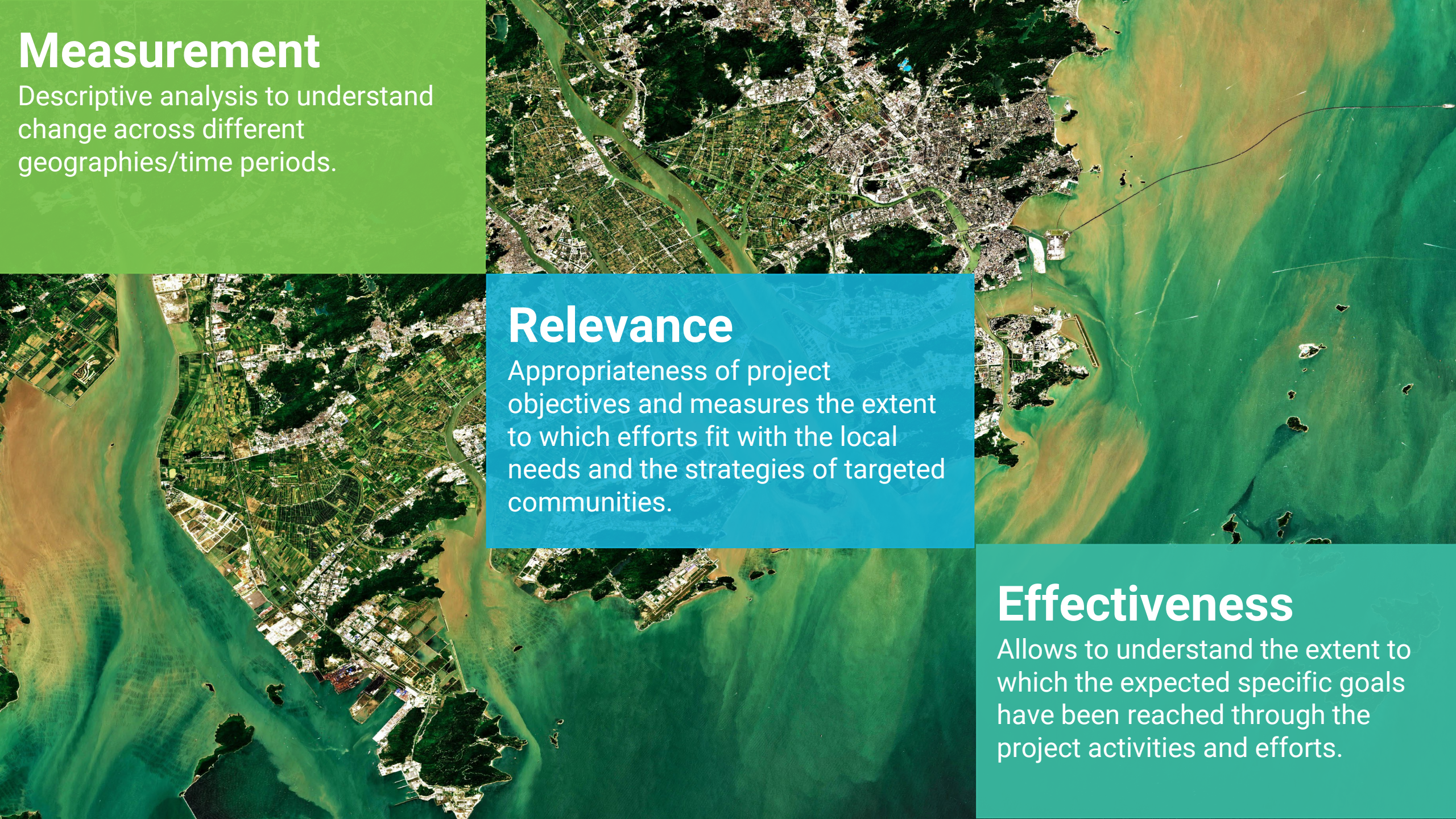
Techniques

- geographic information system (GIS)
- remote sensing
- photogrammetry
- machine learning
- computer vision

2 | Applications in IEG Evaluations

Measurement

Descriptive analysis to understand change across different geographies/time periods.

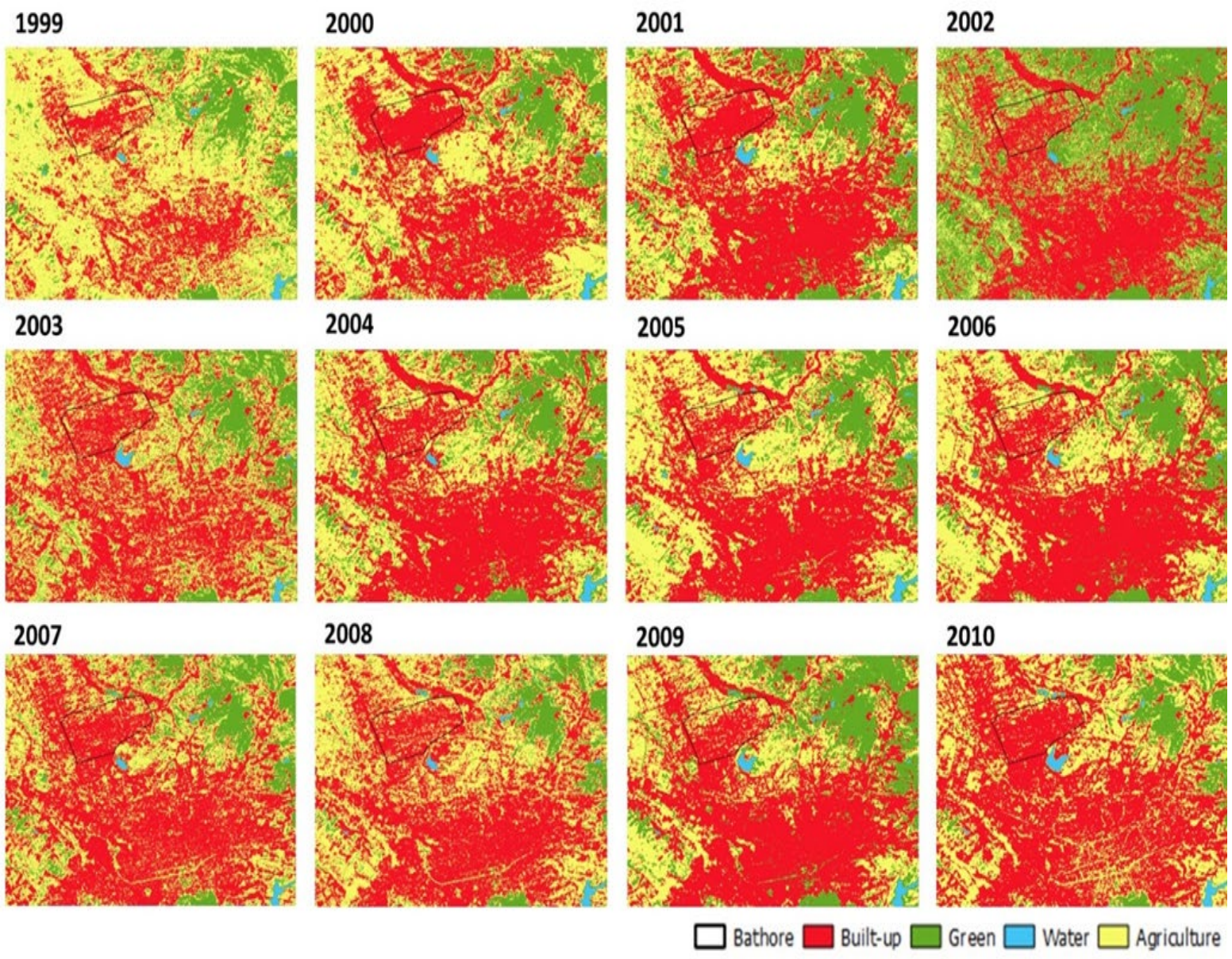


Relevance

Appropriateness of project objectives and measures the extent to which efforts fit with the local needs and the strategies of targeted communities.

Effectiveness

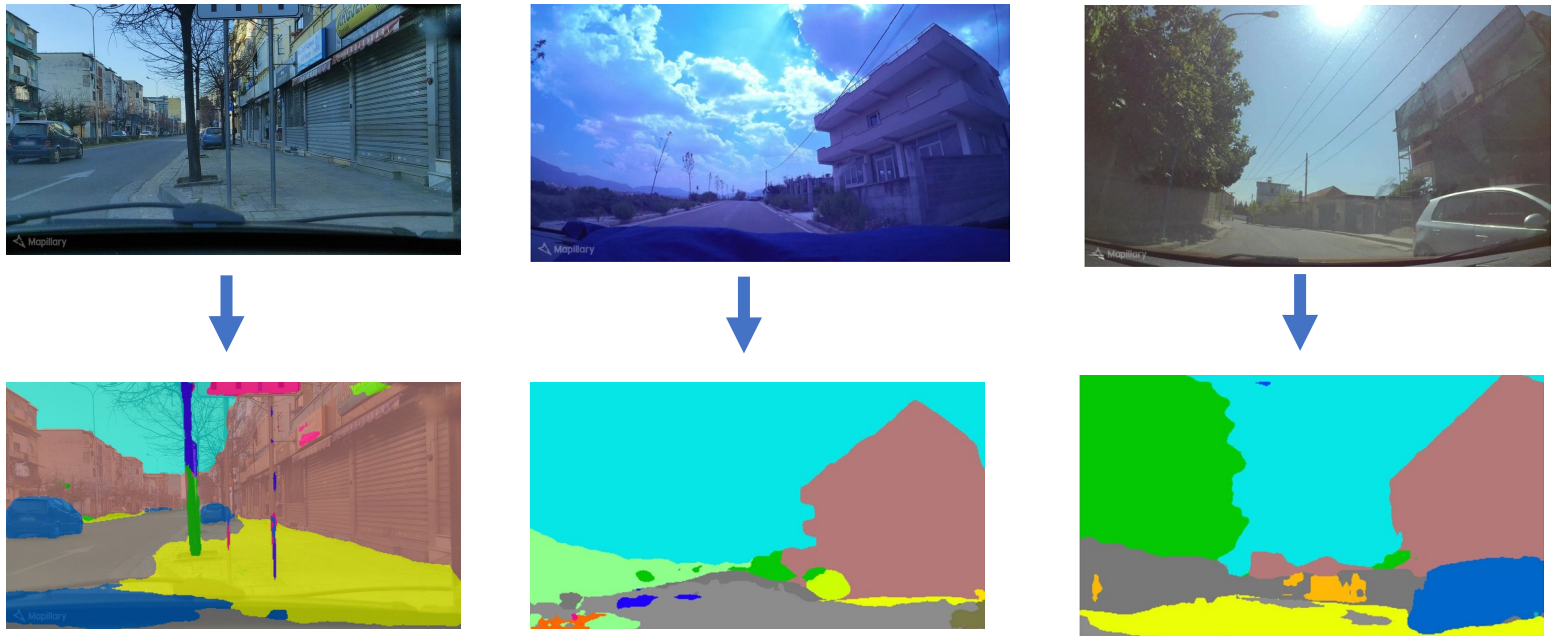
Allows to understand the extent to which the expected specific goals have been reached through the project activities and efforts.

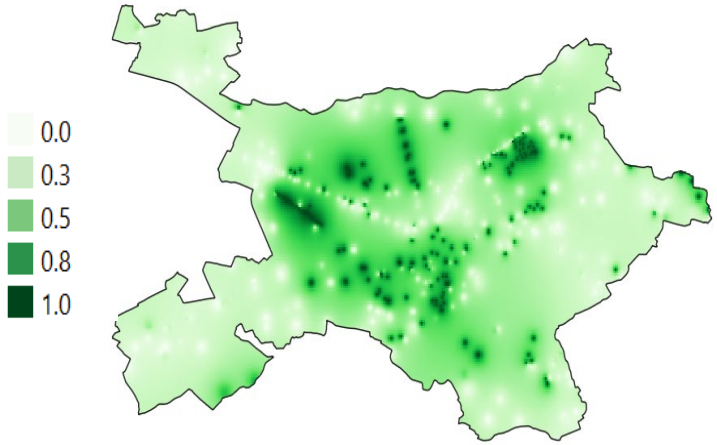


Location: Tirana (Albania). Data source: Landsat.

Measurement

- **Objective:** to understand to what extent horizontal density changed in upgraded neighborhoods In Tirana (Albania).
- **Data source:** optical daylight satellite imagery.
- **Methodology:** supervised classification of satellite imagery to derive land use/land cover model.

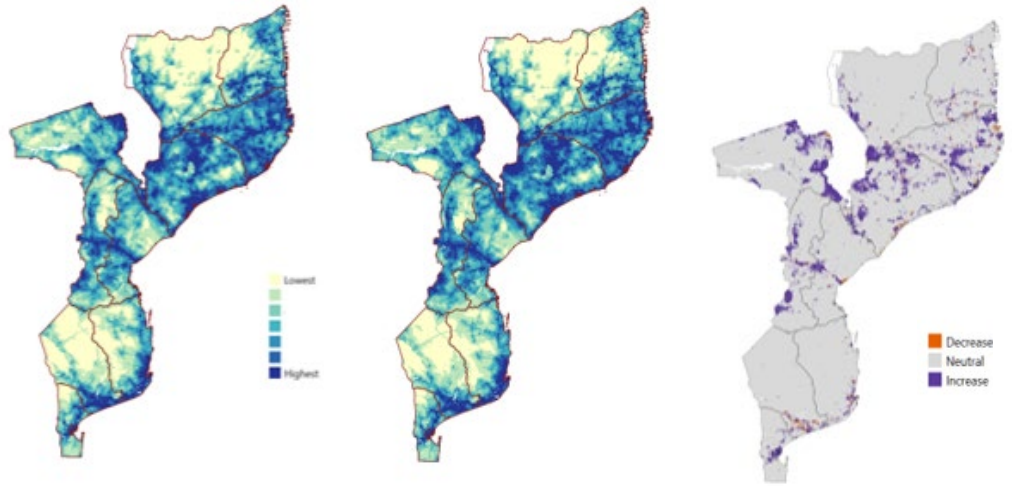




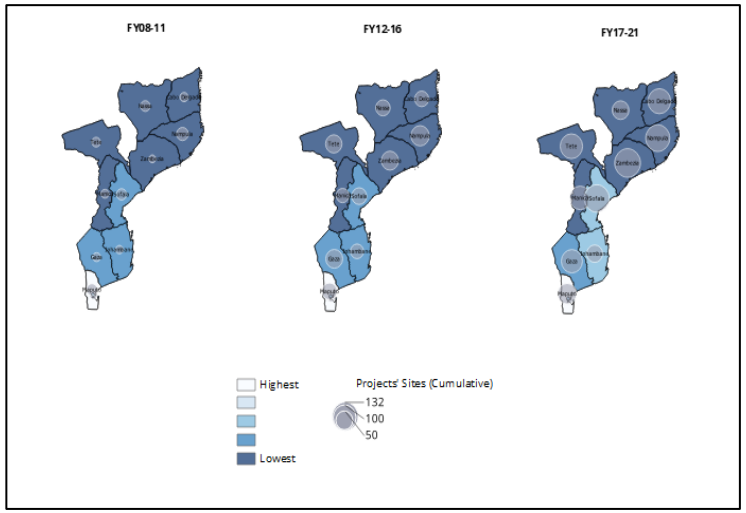
Location: Tirana (Albania). Data source: Mapillary + additional photos taken by the team.

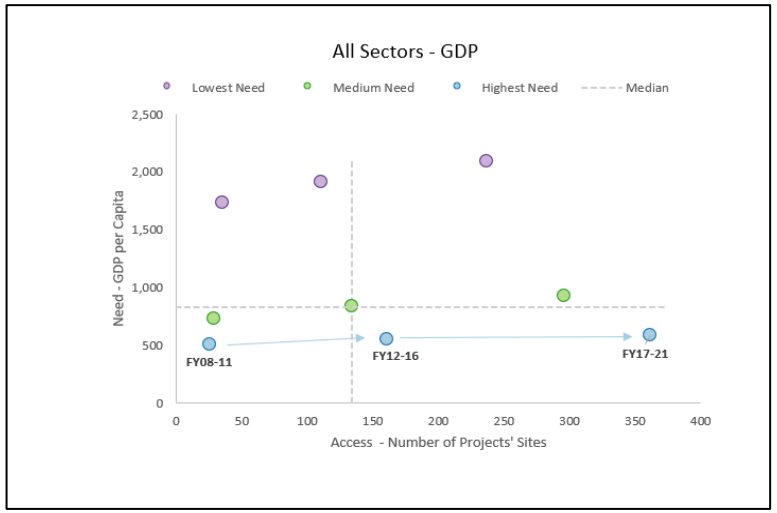
Measurement

- **Objective:** to understand to what extent households in upgraded neighborhoods integrated into the formal economy in Tirana (Albania).
- **Data source:** streetscape digital images.
- **Methodology:** semantic segmentation (deep learning).



Mozambique's geographically disaggregated GDP in 2009 (left) and 2014 (center). Increase/decrease in GDP between 2009 and 2014 (right). Data source: WorldPop.

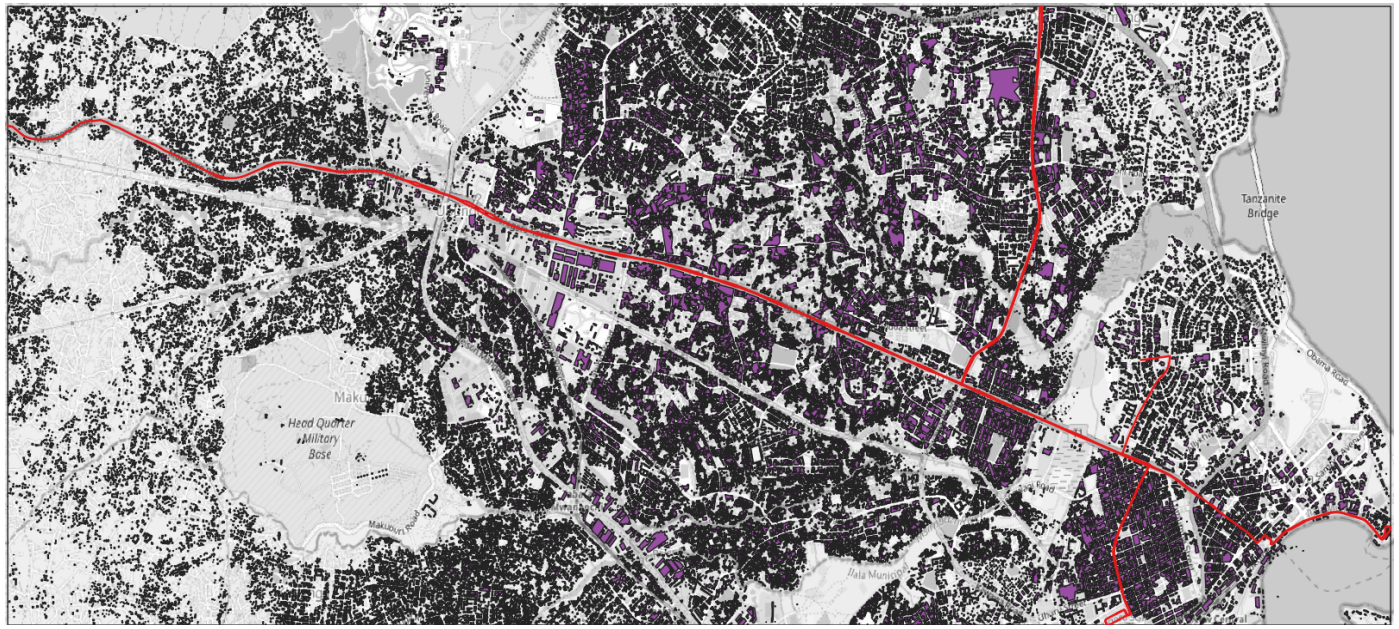


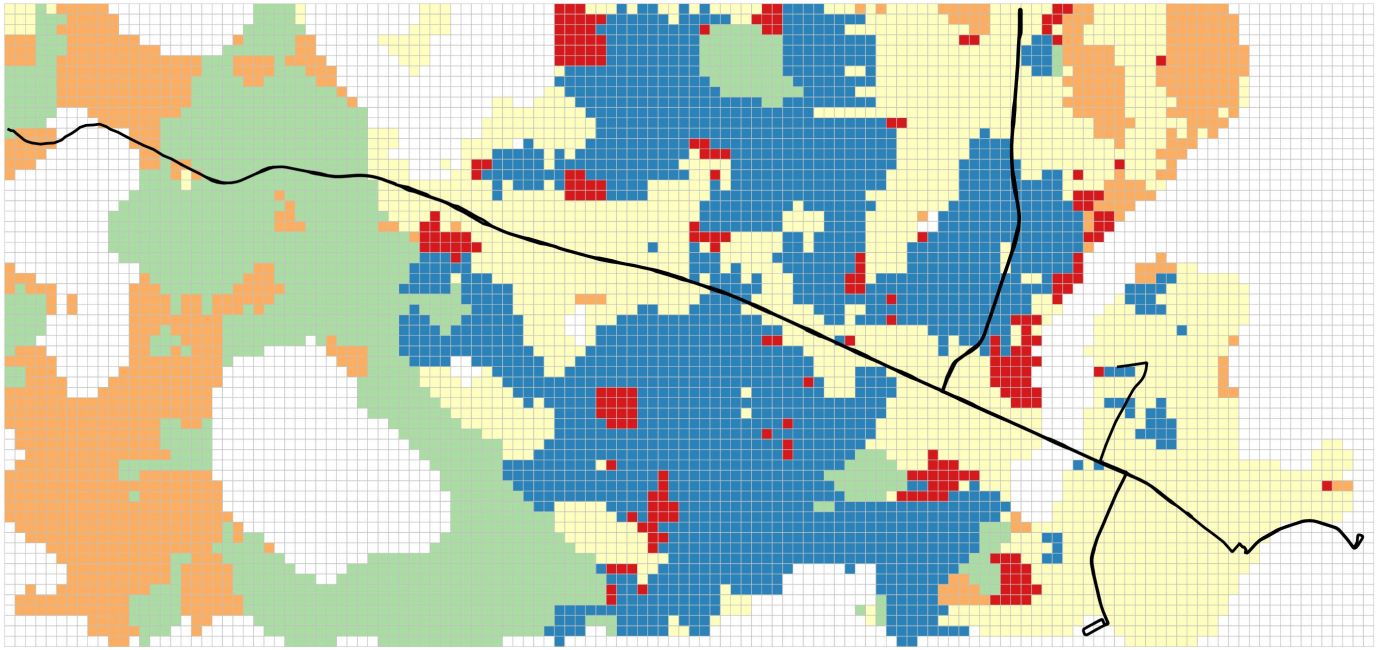


GDP at the province level (need) and number of WB project sites per province (access).

Relevance

- **Objective:** to understand whether the provinces/regions with the highest level of need were targeted by WB.
- **Data source:** gridded raster data (e.g. population, GDP) + geographically disaggregated survey data.
- **Methodology:** gridded geospatial analysis.

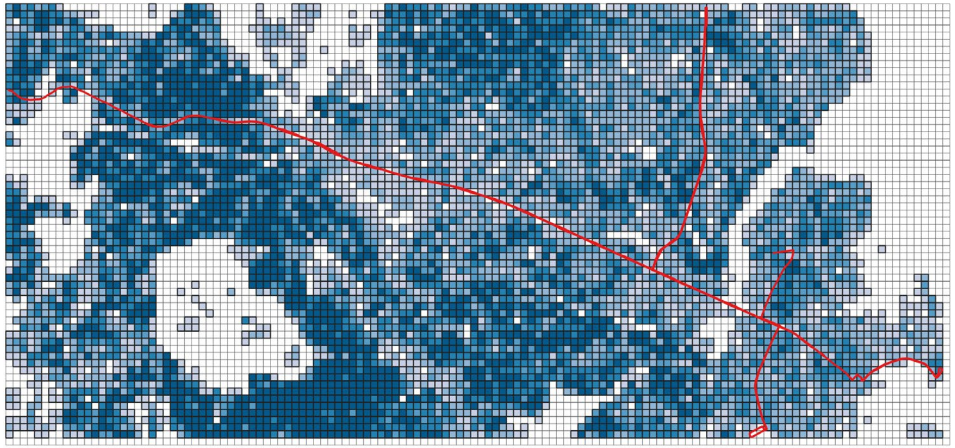




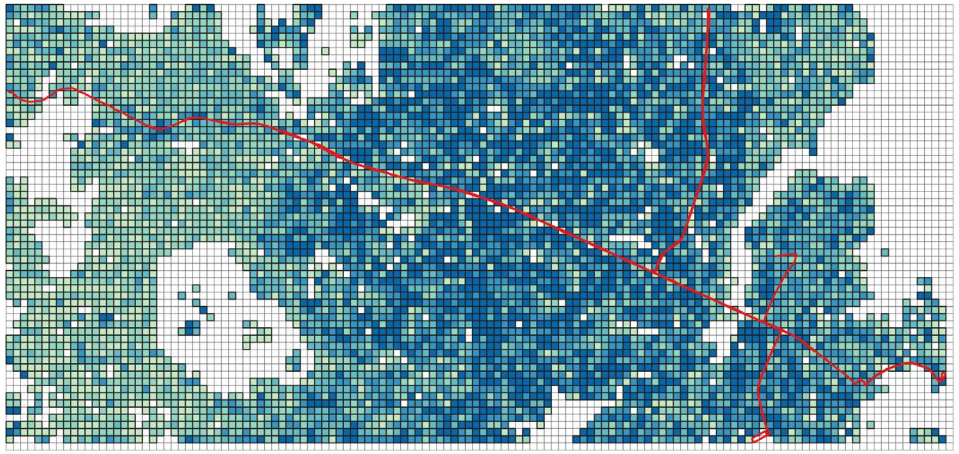
Location: Bus Rapid Transit (BRT), Dar es Salaam (Tanzania). Building parcels source: Microsoft/Bing. Analysis and maps prepared by IEG, World Bank Group.

Effectiveness

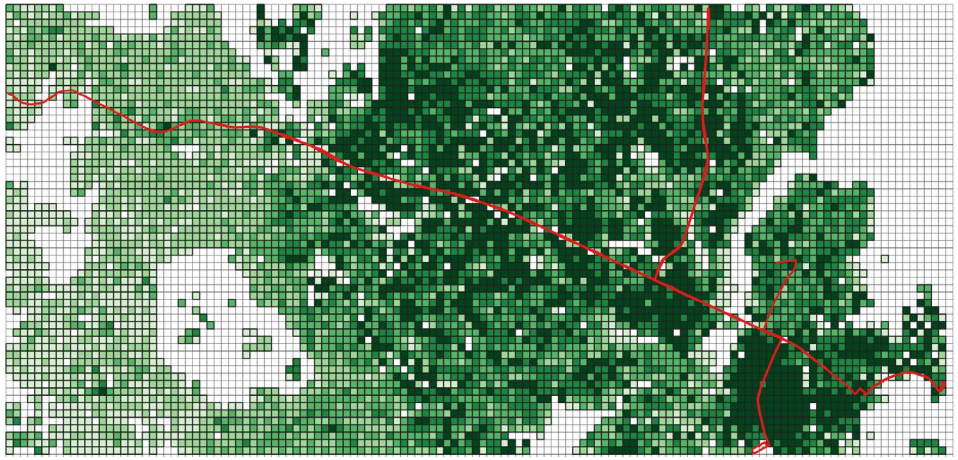
- **Objective:** to assess how different land uses were impacted by a transport project, evaluating its effectiveness in promoting sustainable spatial transformation.
- **Data source:** Microsoft/Bing building parcel data (derived from high-resolution satellite imagery)
- **Methodology:** Gaussian Mixture Model.



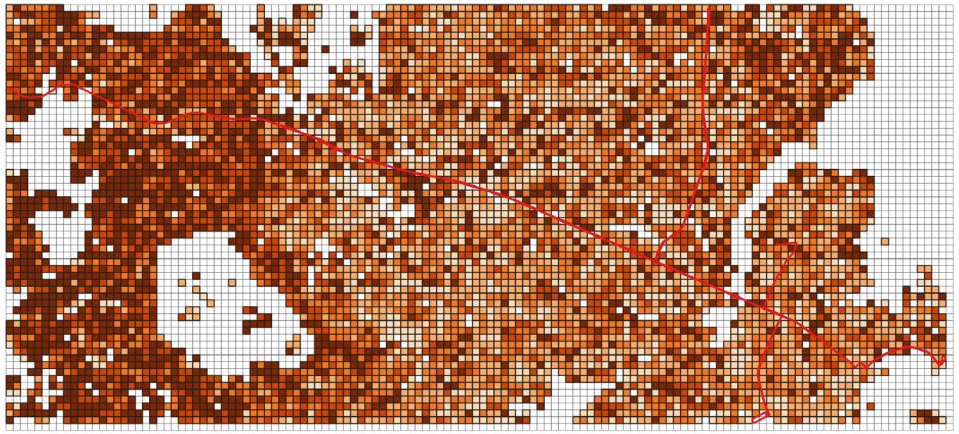
Number of Buildings
- 0 - 0
- 0 - 2
- 2 - 5
- 5 - 7
- 7 - 11
- 11 - 33



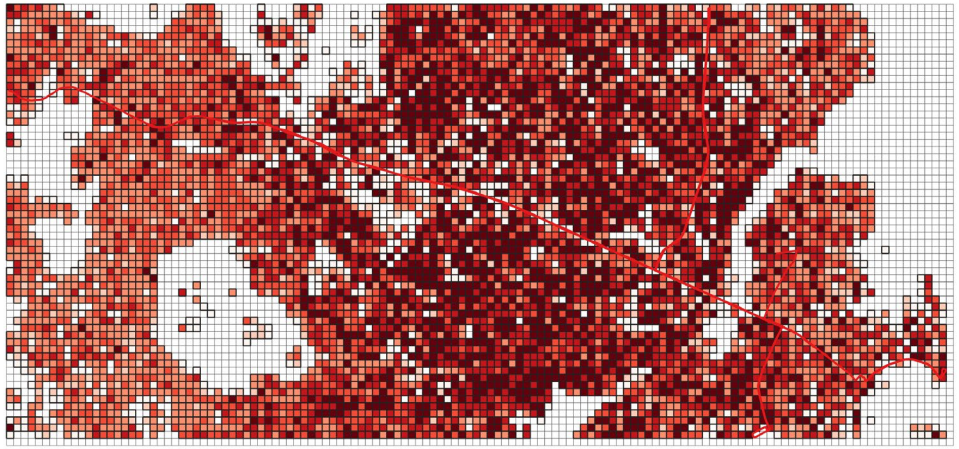
Fractal Dimension Index
- 0 - 0
- 0 - 1.02
- 1.02 - 1.03
- 1.03 - 1.06
- 1.06 - 1.09
- 1.09 - 1.44



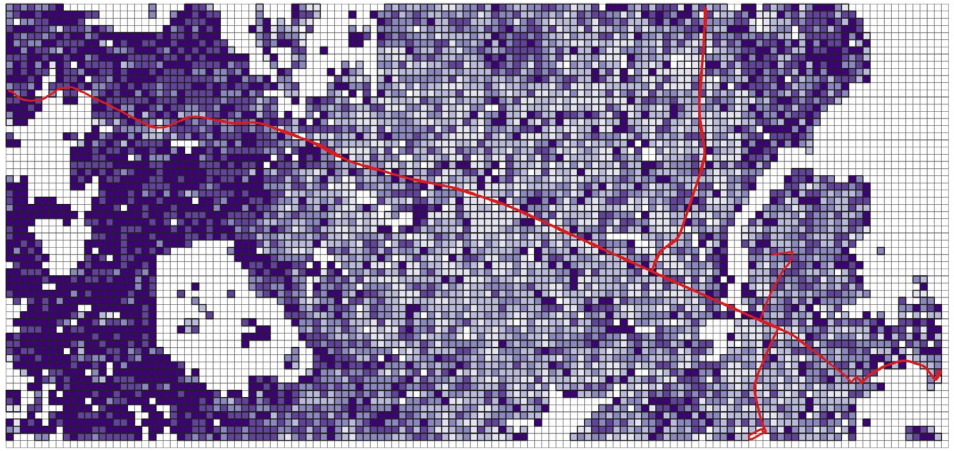
Mean Area (in meters)
- 0 - 0
- 0 - 100
- 100 - 163
- 163 - 266
- 266 - 492
- 492 - 81746



Shape Index
- 0 - 0
- 0 - 0.367
- 0.367 - 0.432
- 0.432 - 0.469
- 0.469 - 0.504
- 0.504 - 0.632



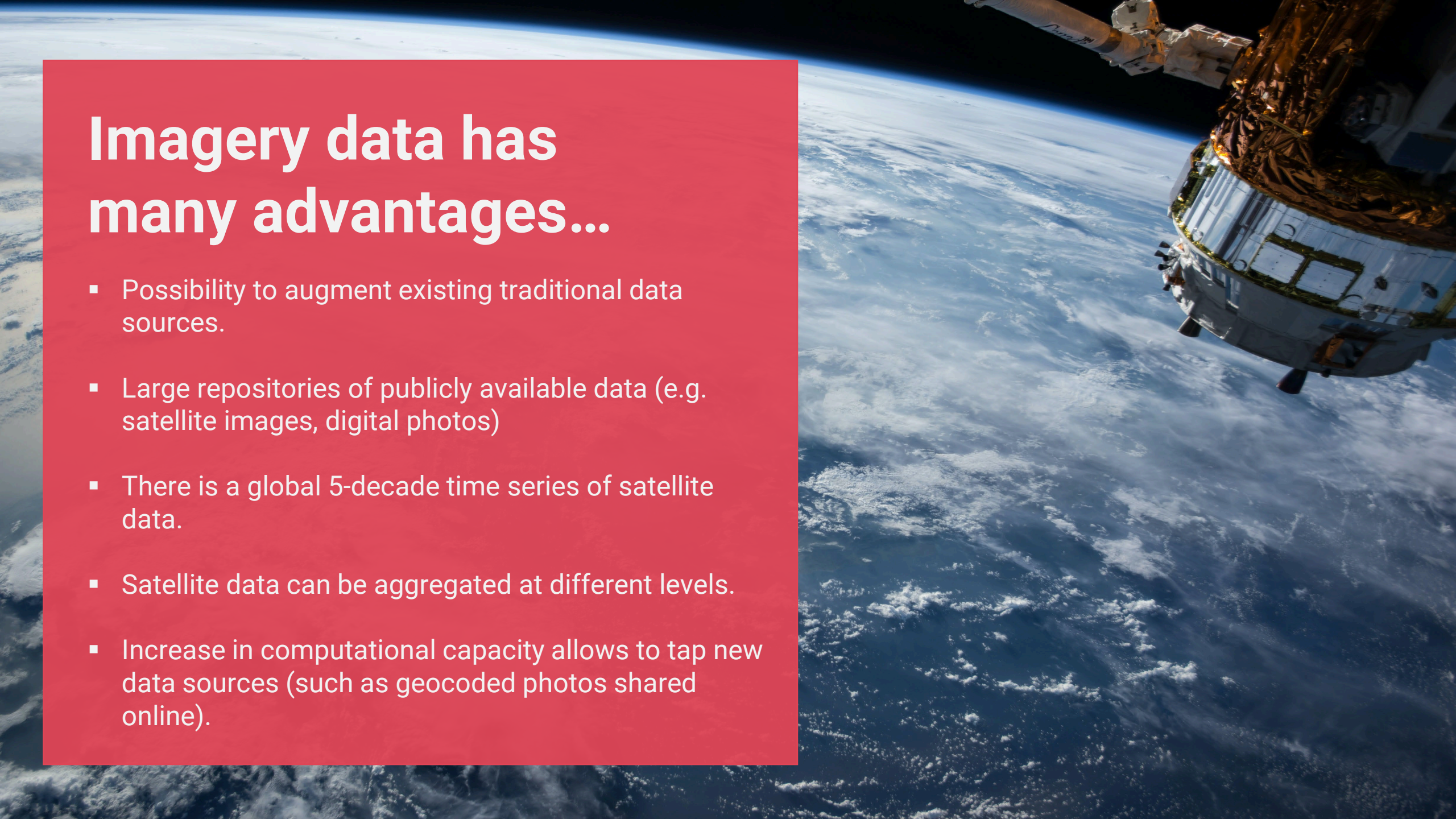
Coefficient of Variation
- 0 - 0
- 0 - 0.29
- 0.29 - 0.57
- 0.57 - 0.76
- 0.76 - 1.03
- 1.03 - 2.57



Compactness Index
- 0 - 0
- 0 - 0.421
- 0.421 - 0.555
- 0.555 - 0.64
- 0.64 - 0.708
- 0.708 - 0.793

Analysis and maps prepared by IEG, World Bank Group.

3 | Practical Considerations



Imagery data has many advantages…

- Possibility to augment existing traditional data sources.
- Large repositories of publicly available data (e.g. satellite images, digital photos)
- There is a global 5-decade time series of satellite data.
- Satellite data can be aggregated at different levels.
- Increase in computational capacity allows to tap new data sources (such as geocoded photos shared online).

… but several challenges remain

- It is essential to construct validity challenges.
- Matching survey data to geocoded image data might be challenging due to coordinates displacement.
- Critical issues when working with crowdsourced data are validity and accuracy.
- Substantial data storage capacity and computation resources might be needed.
- Publicly available satellite data might not have sufficient resolution for some applications.

Conclusion

- Imagery data can be useful to measure dynamics of change across different geographies/time periods.
- Imagery data can be layered with other data sources to construct new variables.
- Imagery data –in combination with other sources of data- can help answer evaluation questions on relevance or effectiveness.
- It is critical to understand the limitations of each type of image data.
- It is important to ‘groundtruth’ imagery data to (1) assess the association between the imagery data proxy and the real phenomenon on the ground; (2) deepen the understanding of the real phenomenon on the ground to enhance the overall validity of findings of the imagery data analysis.

Resources

Leveraging Image Data for Evaluations (Methods Paper)
https://ieg.worldbankgroup.org/evaluations/leveraging-imagery-data-evaluations

Poverty Mapping: Innovative Approaches to Creating Poverty Maps with New Data Sources (Methods Paper)
https://ieg.worldbankgroup.org/evaluations/poverty-mapping-innovative-approaches-creating-poverty-maps-new-data-sources

Why evaluators should embrace the use of geospatial data during Covid-19 (Coronavirus) and beyond (Blog)
https://ieg.worldbankgroup.org/blog/why-evaluators-should-embrace-use-geospatial-data-during-covid-19-coronavirus-and-beyond

When evaluators cannot make it to the field, they can always observe from space (Blog)
https://ieg.worldbankgroup.org/blog/when-evaluators-cannot-make-it-field-they-can-always-observe-space

Impacts of energy efficiency projects in developing countries: Evidence from a spatial difference-in-differences analysis in Malawi (peer reviewed journal)
https://www.sciencedirect.com/science/article/pii/S0973082623000522?casa_token=5LfULYYD9nwAAAAA:0Yx5GJNyhEoDhOnML-I7_6Oxa_2zDPoU7jcigpe2qGzII7dEyc2nLRWmC-u8oO6vnFldEG4on8M

Thank You!