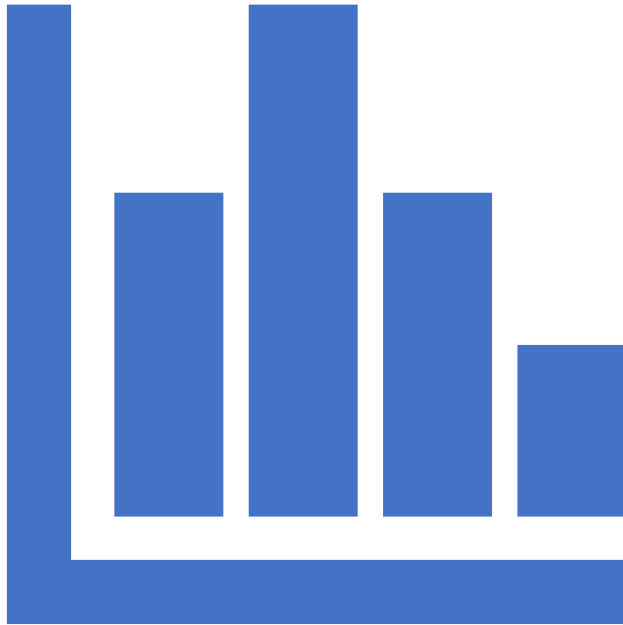




UNITED NATIONS

الاستقيا  
ESCWA

Shared Prosperity Dignified Life



Dr. Ziad Abdallah

# DATA SCIENCE FILLING SDG GAPS PROJECT

# These SDG indicators are ...

## sensitive to the **rural** definition used

- 4.1.1 Children in school & proficiency
- 4.6.1 Adult literacy and numeracy
- 6.1.1 Access to safe drinking water
- 7.1.1 Access to electricity
- 8.10.1 Use of banking services
- 9.c.1 Coverage by mobile network
- 9.1.1 Rural population with access to an all weather road

## sensitive to the **city** definition used

- 11.2.1 Population that has convenient access to public transport
- 11.3.1 Land consumption over population growth
- 11.6.2 Annual mean levels of fine particulate matter (e.g. PM2.5 and PM10) in cities
- 11.7.1 Open public space for public use for all

## Some variations in urban definitions in select Arab Countries

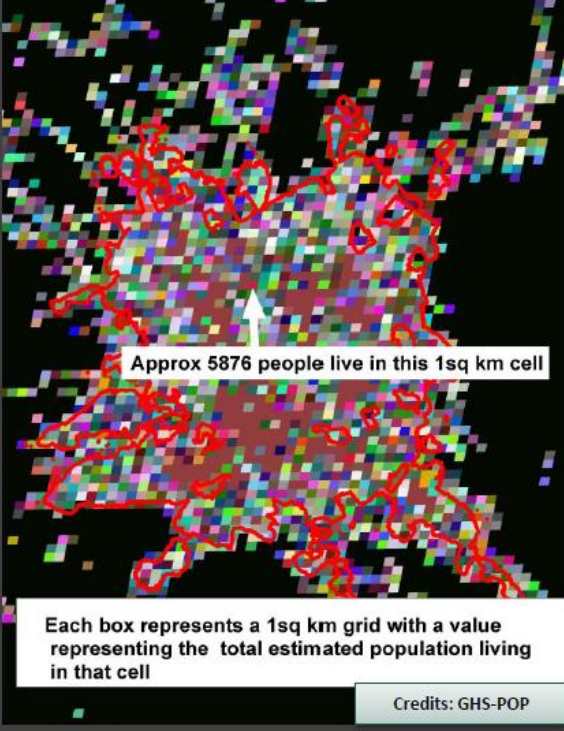
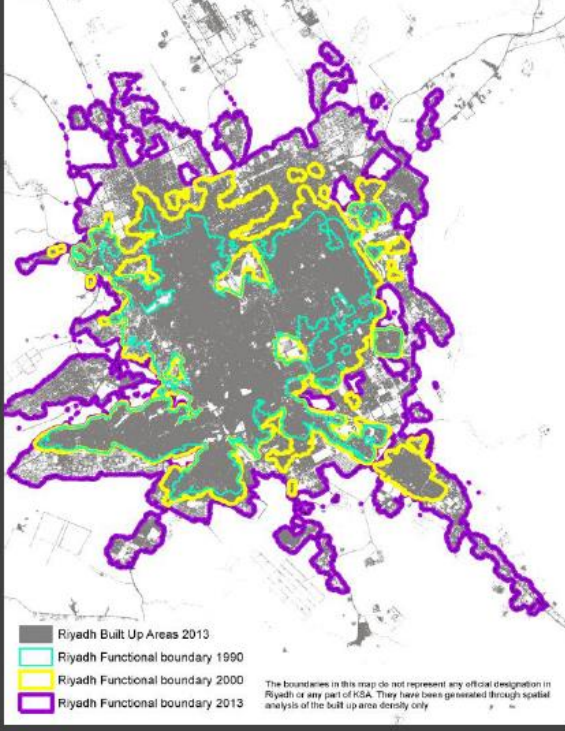
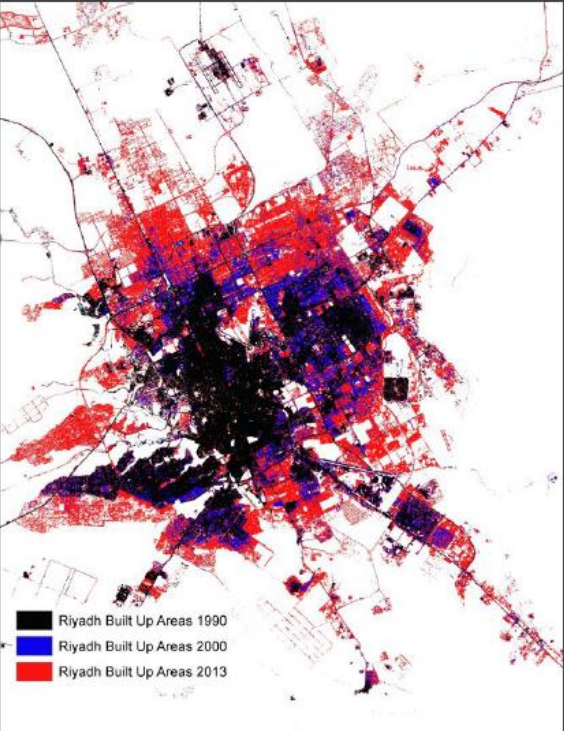
|              | Administrative function | Economic function | Population size | Population density | Urban characteristics | Other criteria | No definition |
|--------------|-------------------------|-------------------|-----------------|--------------------|-----------------------|----------------|---------------|
| Bahrain      | X                       | X                 | X               | X                  | X                     |                |               |
| Jordan       |                         |                   | X               |                    |                       |                |               |
| Morocco      | X                       |                   | X               |                    | X                     |                |               |
| Palestine    |                         |                   |                 |                    |                       | X              |               |
| Saudi Arabia | X                       | X                 | X               |                    |                       | X              |               |
| Sudan        | X                       |                   | X               |                    | X                     |                |               |
| Syria        |                         |                   | X               |                    |                       |                |               |
| Tunisia      |                         |                   |                 |                    |                       |                | X             |
| Yemen        | X                       |                   | X               |                    |                       |                |               |
| Iraq         | X                       | X                 | X               | X                  | X                     |                |               |
| Lebanon      |                         |                   |                 |                    |                       |                | X             |
| Egypt        |                         |                   |                 |                    | X                     | X              |               |
| UAE          |                         |                   | X               |                    |                       | X              |               |

# SDG Indicator 11.3.1 Supporting Sustainable Cities

Ratio of land consumption rate to population growth rate (LCRPGR)

### Concepts

- Land Consumption Rate
- Population Growth Rate



Data needs/  
Indicator inputs

Built up layer

Dynamic & functional city boundaries

Disaggregate population data

- Estimating SDG 11.3.1  
Land Consumption Rate to Population Growth Rate (LCRPGR)

$$LCRPGR = \frac{\text{Land Consumption Rate}}{\text{Population Growth Rate}}$$

$$LCRPGR = \frac{\ln\left(\frac{Urb_{t2}}{Urb_{t1}}\right)}{\ln\left(\frac{Pop_{t2}}{Pop_{t1}}\right)}$$

# Degree of Urbanization

## Definition of DEGURBA

Degree of Urbanisation (DEGURBA).

Method classifies the local units of a country into classes:

**1. Cities**



**1. Cities**

**2. Towns**



**2. Dense Towns**  
**3. Semi-Dense Town**  
**4. Suburb**

**3. Rural areas.**

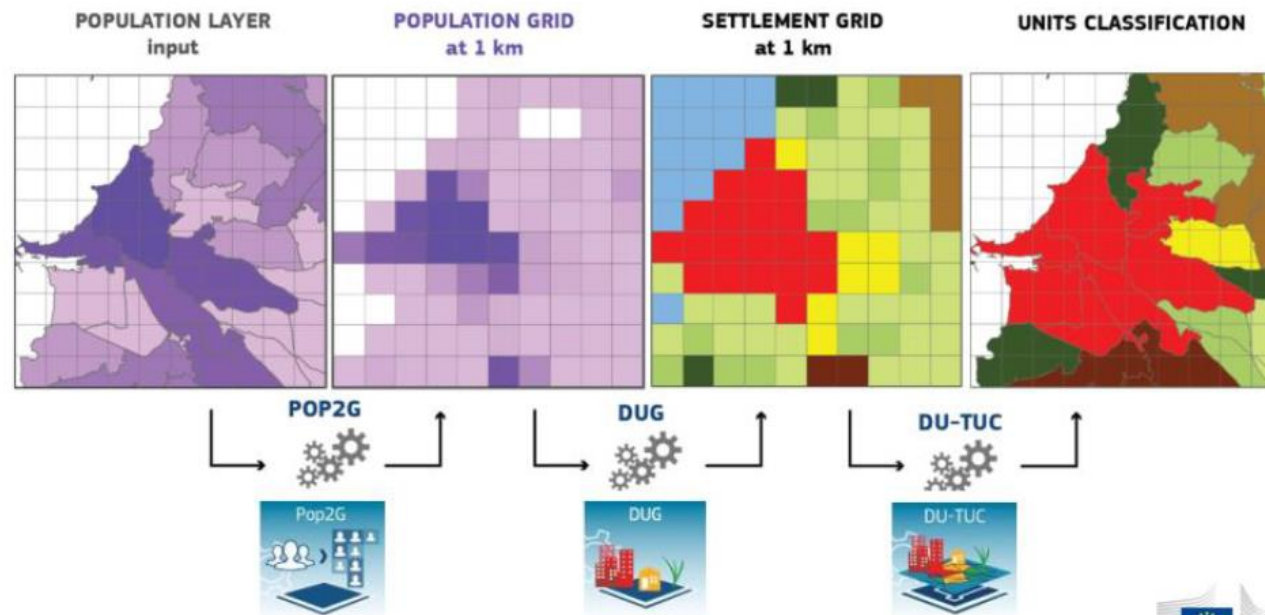


**5. Village**  
**6. Dispersed rural areas.**  
**7. Mostly uninhabited areas**

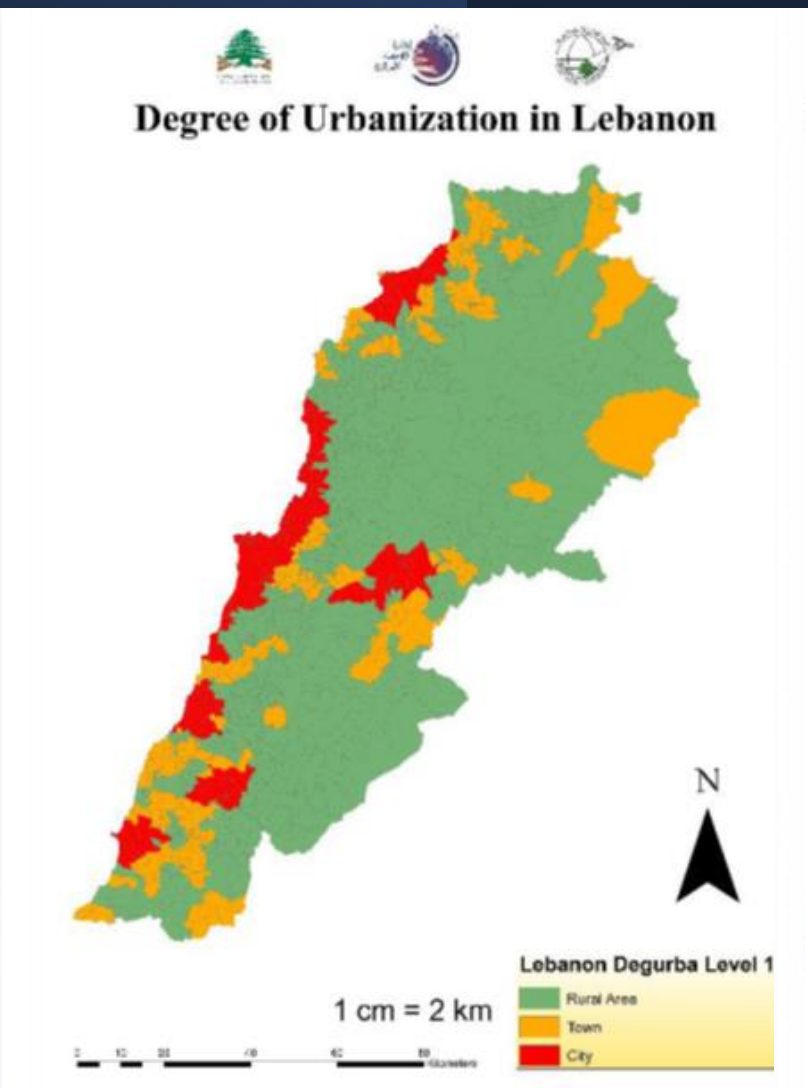
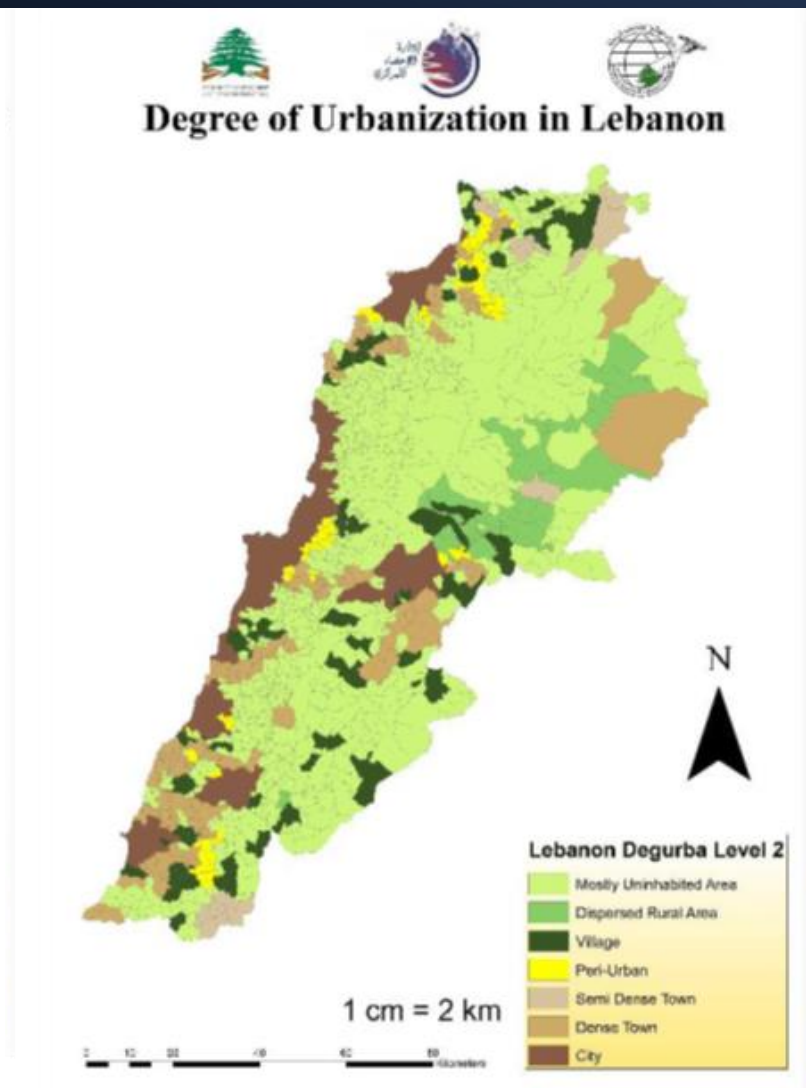
# Degree of Urbanization

## The methodology in three steps:

### Workflow



# Best Practices of CAS in urban data: Degree of Urbanization in Lebanon





# Indicator 11.3.1 compilation

| <b>DEGURBA summary statistics</b> |                                    |           |       |
|-----------------------------------|------------------------------------|-----------|-------|
|                                   | <b>Urban Population</b>            | 1,351,328 | 85.7% |
| <b>DEGURBA L1</b>                 | <b>City</b>                        | 204,453   | 13.0% |
|                                   | <b>Town &amp; Semi-Dense area</b>  | 1,146,875 | 72.7% |
|                                   | <b>Rural Area</b>                  | 225,248   | 14.3% |
| <b>DEGURBA L2</b>                 | <b>City</b>                        | 204,453   | 13.0% |
|                                   | <b>Dense Town</b>                  | 125,756   | 8.0%  |
|                                   | <b>Semi-dense Town</b>             | 89,264    | 5.7%  |
|                                   | <b>Suburban or peri-urban area</b> | 931,855   | 59.1% |
|                                   | <b>Village</b>                     | 100,942   | 6.4%  |
|                                   | <b>Dispersed Rural area</b>        | 124,306   | 7.9%  |
|                                   | <b>Mostly uninhabited area</b>     | -         | 0.0%  |

Develop and implement the data innovation schemes showcasing filled identified indicators

The screenshot displays the Google Earth Engine web interface. The browser address bar shows `code.earthengine.google.com`. The main interface is divided into several panels:

- Scripts Panel:** Shows a list of scripts under the 'Owner' section, including 'SDG 2.4 (copy).2 not co...', 'SDG 2.4.1 not completed', and 'SDG241 with...'. The 'Reader' section shows various example scripts.
- Code Editor:** Contains a JavaScript script titled 'SDG241 with old data'. The script defines a study area (Lebanon), sets start and end dates (2019-01-01 to 2019-12-31), loads Sentinel-2 image data, filters for cloud coverage, and defines a function to calculate NDVI for each image.
- Inspector/Console Panel:** Shows the output of the script: 'Percentage of agricultural land area out of the... JSON' with the value '21.652971466270348'.
- Map Panel:** Displays a map of the region around Beirut and Damascus, with a heatmap overlay showing agricultural land area. The map includes labels for 'Beirut', 'Damascus', 'Al Zabadani', 'Ar Ruhaybah', 'Ad Dumayr', and 'Sabaa Biâr'.

The Windows taskbar at the bottom shows the system tray with the time '1:43 PM' and date '15/03/2023'.

# IMPUTATION

Data preparation Challenge: The Data Transformation

Data processing challenges: Tool Selection for Dealing with Missing Data

SPSS, Stata, Python, and SAS: Strengths & Limitations

Imputation challenges: Multiple Imputation Methods

Validation of the results challenges: Statistical Test Methods, Comparing Data Distributions, etc

But who will accept and how to adopt

Conclusions

# Data preparation Challenge: The Data Transformation

Understanding the Data

Data Clean

Data transformation from long format to wide format

“Similar” data in the same set

```
dcast formula dcast(aql, month + day ~ variable, value.var = "value")
```

| ID variables<br>(left side of formula) | Variable to swing<br>into column names<br>(right side of formula) | Values<br>(value.var) |
|--|---|-----------------------|
|--|---|-----------------------|

Long-format data

| month | day | variable | value |
|-------|-----|----------|-------|
| 5     | 1   | ozone    | 41    |
| 5     | 2   | ozone    | 36    |
| 5     | 3   | ozone    | 12    |
| 5     | 4   | ozone    | 18    |
| 5     | 5   | ozone    | NA    |
| 5     | 6   | ozone    | 28    |

Wide-format data

| month | day | ozone | solar.r | wind | temp |
|-------|-----|-------|---------|------|------|
| 5     | 1   | 41    | 190     | 7.4  | 67   |
| 5     | 2   | 36    | 118     | 8.0  | 72   |
| 5     | 3   | 12    | 149     | 12.6 | 74   |
| 5     | 4   | 18    | 313     | 11.5 | 62   |
| 5     | 5   | NA    | NA      | 14.3 | 56   |
| 5     | 6   | 28    | NA      | 14.9 | 66   |

# Tool Selection for Dealing with Missing Data

| Feature                  | SPSS | Stata |
|--------------------------|------|-------|
| Imputation Methods       | -    | +     |
| Data Management          | +    | -     |
| Estimation and Inference | -    | +     |

# Imputation challenges: Multiple Imputation Methods

Imputation Methods

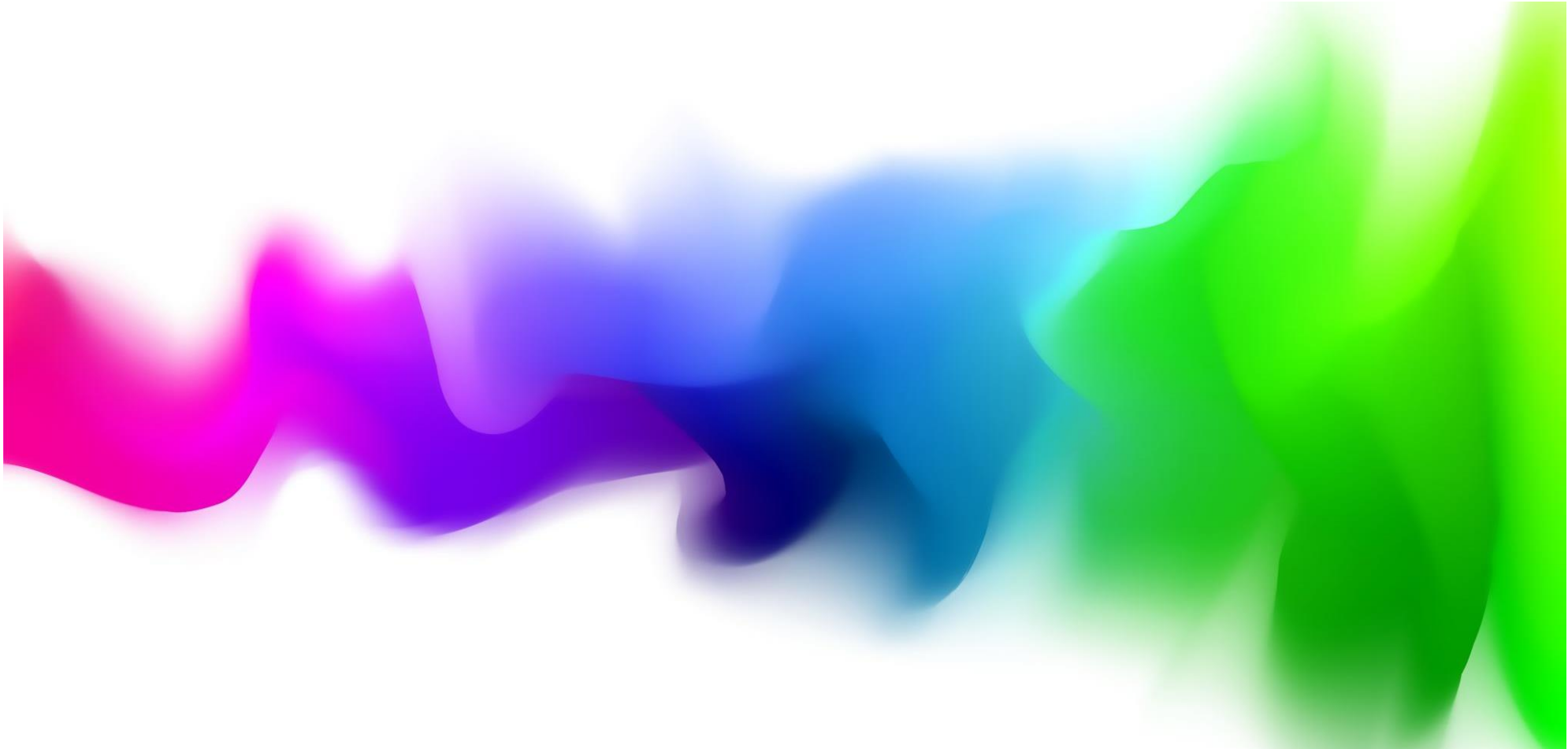
KNN imputation

EM imputation

Multiple imputation by chained equations (MICE)

[Tutorial: Introduction to Missing Data Imputation | by Cambridge Spark | Medium](#)

# Capacity development and regional cooperation



## Basic concept in big data for official statistics and the SDGs

- ❑ Enhance capacity of NSOs in the Arab region
- ❑ Provide solid foundation in fundamental concepts of big data for official statistics
- ❑ Highlight relevance to SDG implementation and monitoring



# Basic concept in big data for official statistics and the SDGs

- ❑ Workshop 1: Basic Concept in Data Science and Big Data for official statistics and the SDGs
- ❑ Workshop 2: Data Science and Big data for official statistics and the SDGs with Python
- ❑ Workshop 3: Data Science and Big data for official statistics and the SDGs: uses cases

# Conclusion Remarks

- Technological advances are rapidly transforming National Statistical Offices (NSOs).
- ICT including Geospatial Information Technology, is a cross-cutting technology to help monitoring SDGs
- Modernization of practices in the Arab region is necessary.
- Focus is on capacity building tools such as specialized training programs, collaborative data partnerships, and adoption of data science and technology.
- Approaches enhance capacity development of individuals and institutions to collect, analyze, and report data.
- Harnessing data science empowers stakeholders to bridge information gaps.
- This ensures accurate and comprehensive progress monitoring for a sustainable future.



Shared Prosperity Dignified Life



Thank you