Socio-Economic Aspects of Urban Development

Unit 5: Urban Data Sources and Methods

Prof. Melanie Krause, Ph.D.

Winter Semester 26/27

Overview of Unit 5

- Where to obtain data for urban analysis
  - Official data from administrative sources
  - Geo-spatial data:
    - ★ Maps of the urban built-up
    - ★ Satellite data of nighttime lights
    - ★ Open Street Map data
- Data analysis methods
  - Statistical data analysis
  - Exercise: Calculating yearly, long-run and average growth rates of rental prices in Leipzig

Administrative Data Sources for Urban Analysis

- When can you obtain data for urban analysis?
- The first port of call will often be the statistical office of the city
  - For example, visit https://statistik.leipzig.de/
- Meta-databases:
  - https://www.govdata.de (Official data from German cities, counties and regions)
  - https://data.europa.eu/en (Official data from across Europe)
- Official websites with data on specific topics for all cities/counties
  - https://www.regionalstatistik.de (Regionaldatenbank Deutschland, official data at the German county level): From sectoral economic data... over the age at marriage... to hazardous waste at the city/county level
  - https://ec.europa.eu/eurostat/web/cities/database (Eurostat data): From educational data ... over tourism... to environmental data at the city/county level

Global Human Settlement Layers

- The official sources will often use the administrative boundaries of cities for their data.
- If you are interested in the actual built-up area, you can see the urban extent from the **Global Human Settlement Layers (GHSL)**, global geo-spatial data sets provided by the European Commission
- They offer a number of high-resolution geo-spatial data sets worldwide gained from satellite data, continuously expanding their offer
- Go to: https://ghsl.jrc.ec.europa.eu/dataToolsOverview.php (Academic Publication: Pesarini et al, 2024)
  - **Built-up surface of the earth**, 100m resolution, from 1975 to 2020 in 5-yearly intervals
  - **Average building height**, 100m resolution, year 2018
  - **European Settlement map**, 2m resolution, year 2015
- You can open the shapefiles in QGIS to zoom into the cities you are interested in

Global Human Settlement Layers



European Settlement Map 2015, R2019

Read the technical details for this product

Current selection:

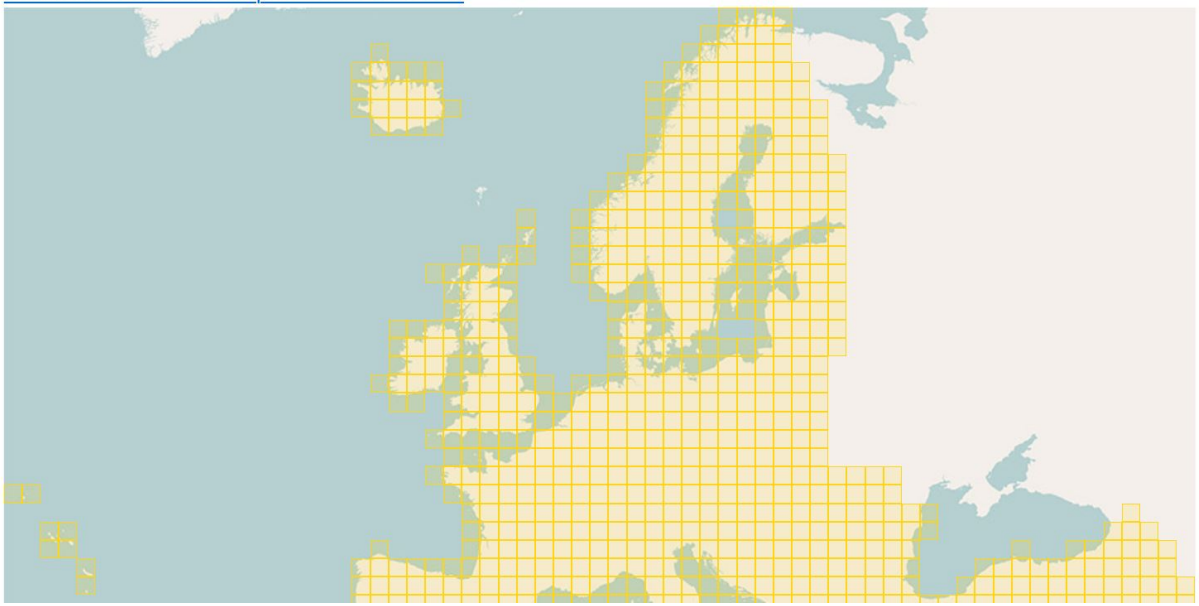
Product: **ESM**, epoch: **2015**, resolution: **2m**, coordinate system: **ETRS89/LAEA Europe (EPSG:3035)**

ⓘ To be noted that some variation might be available only for a certain product (e.g. the 30m resolution is only available for the GHS-BUILT multi-temporal classification)

Download by tiles (click on each box to download a single tile):

Interactive visualisation of the European Settlement Map 2015, R2019

Download the tiles index shapefile for the ESM 2015



GHSL Urban Center Database

While the GHSL data are shapefiles, the authors from the European Commission Florczyk et al (2019) have also released an **excel data base** for selected cities

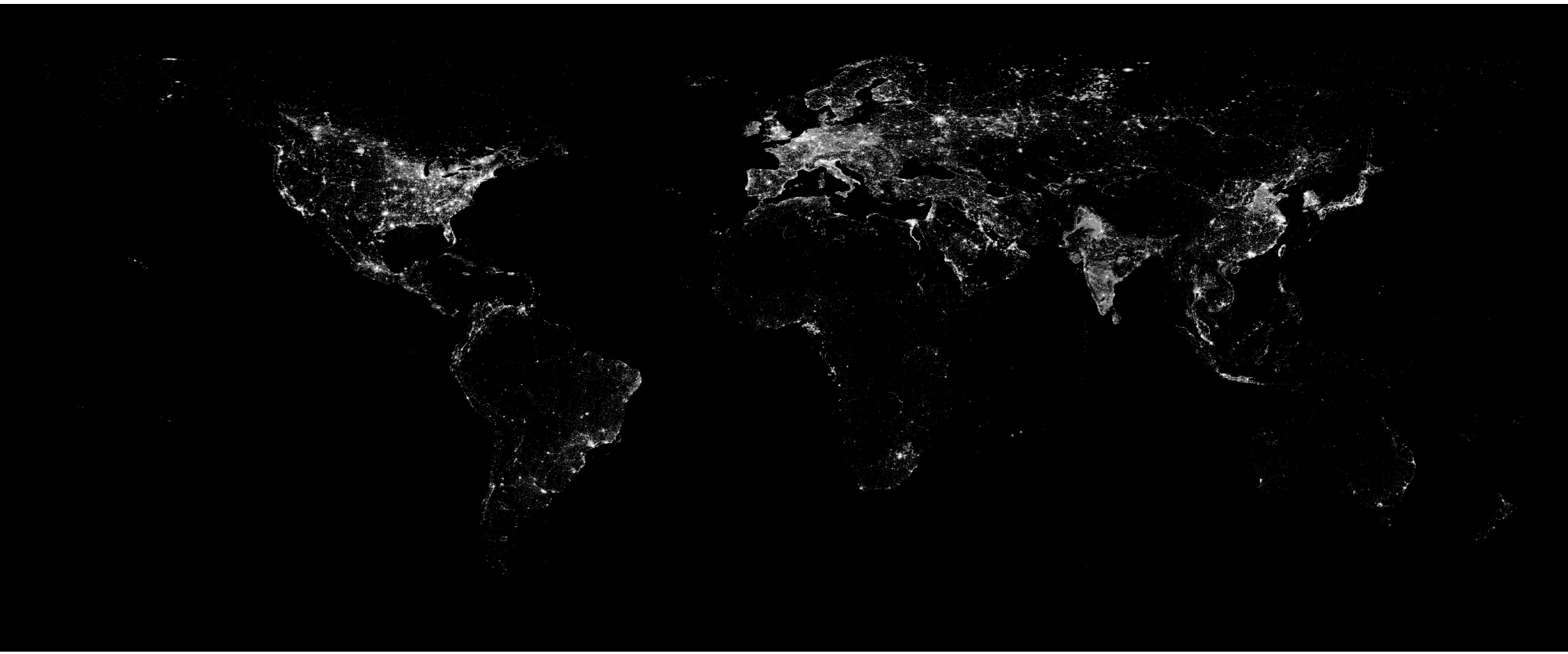
- Go to: https://ghsl.jrc.ec.europa.eu/ucdb2018Overview.php
- Urban extent and population based on GHSL data
- More than 10,000 cities around the world
- Available years like the GSHL data: 1975, 1990, 2000, 2015
- Additional data for these cities have also been included from other sources:
  - Greenness, CO2 emissions, city GDP etc
  - For these variables, refer to the original source. Look at precise measurements and caveats!

Open Street Map and Google Maps

- Information on urban infrastructure, buildings, green spaces, amenities and other places of interest: Open Street Map or Google Maps.
- **Open Street Map (OSM)** is user-generated data: In the Open Street Map project (https://planet.openstreetmap.org/), volunteers around the world collect geo-spatial data
- Detailed maps from around the world with landmarks, shops, schools, cafés, restaurants... $\Longrightarrow$ **Cultural Amenities**
- OSM is an open system, while Google Maps is a closed system and operated by Google. Paper on using OSM in urban economics: Seidel (2023)
- Useful for applications, such as an analysis of cultural amenities in different parts of the city and how they influence rental prices.
- Readily available as a contemporaneous cross-section, no panel: No time changes of the charted characteristics. But recently, some past OSM data have been made available with a wayback machine: https://download.geofabrik.de/

Satellite Data of Nighttime Lights

- Satellite data of nighttime lights have been used increasingly in urban economic research
- Lights as a proxy for local economic activity established by Henderson et al (2012), see also Donaldson and Storeygard (2016) for an overview



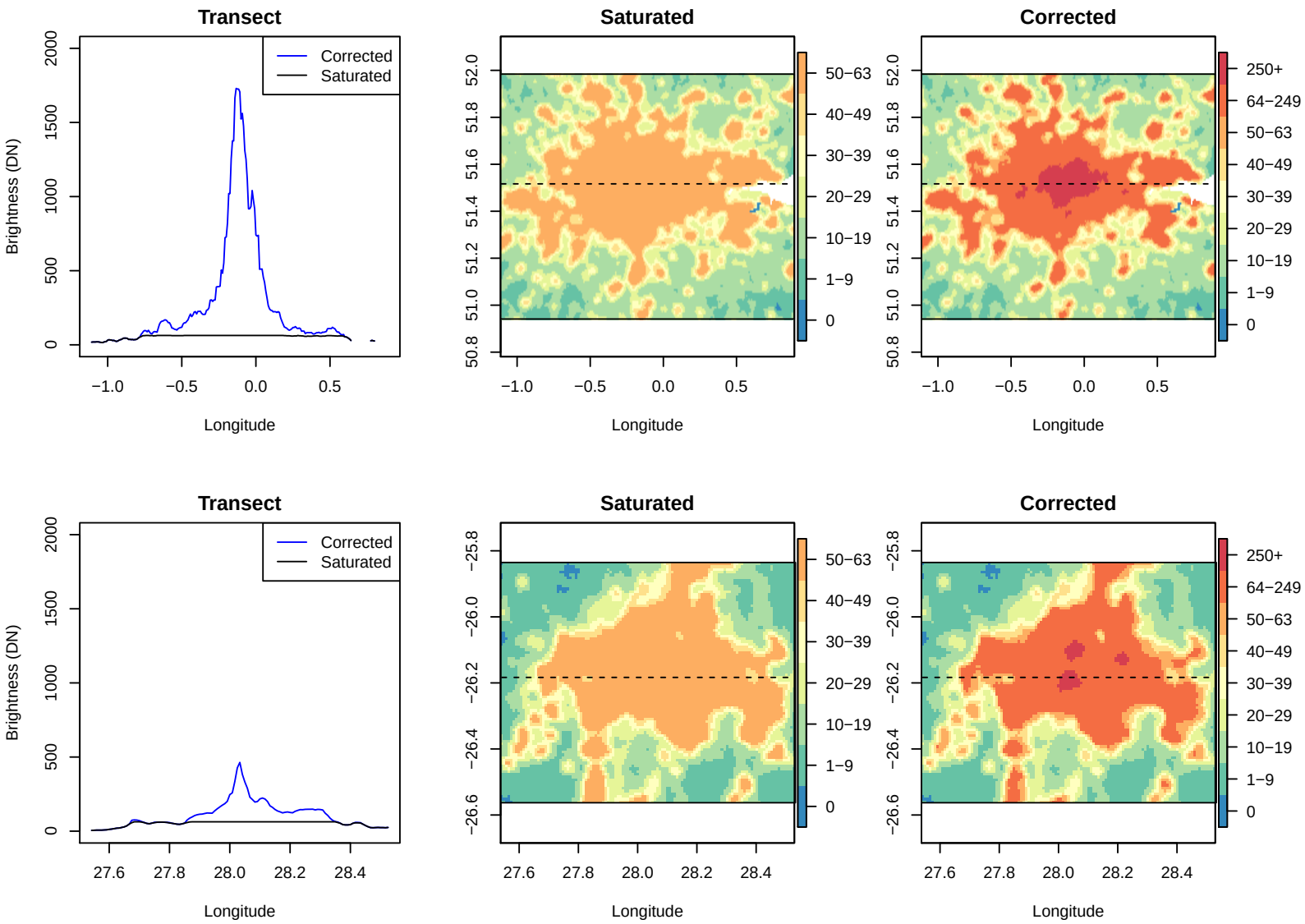
Satellite Data of Nighttime Lights (Since 2012)

- VIIRS data series of nighttime lights (*Visible and Infrared Imager Radiometer Suite*) from the US National Oceanic Administration Agency (NOAA) and National Aeronautics and Space Agency (NASA)
- Satellites measure light intensity around the world:
  - Light intensity at night: Overflight of the satellite around midnight
  - High resolution: pixel size of approx. 20 hectares (450m · 450m, 15 angular seconds)
  - Light intensity is measured in radiance
  - The data is already filtered, cleaned and aggregated by NOAA/NASA over the days (no measurement during cloud cover)
  - Available to download for free for the whole world on a monthly and annual basis 2012-2025 (in particular "Annual VNL V2.2")
  - https://eogdata.mines.edu/products/vnl/

Nighttime Lights in Cities: Top Coding (1992-2013)

- The current VIIRS data series has been available since 2012. For longer-term urban developments, **older satellite data** DMSP-OLS (1992-2013) is used. However, this 1st generation of satellite data suffers from the **top coding** problem:
  - Measurement of brightness on a scale from 0 to 63 Digital Number (DN), saturation of the sensor at 63 DN
  - No differentiation between 'bright' and 'very bright'
  - Many of the world's major cities are affected
  - Differences between urban and rural areas are distorted, no within-city analysis if all districts appear similarly bright
- Solution by Bluhm and Krause (2022): **Top-coding-corrected data set**
  - Pixels affected by top-coding are replaced with values from the Pareto distribution
  - Corrected data series worldwide for 1992-2013: https://www.melanie-krause.de/data.html

London and Johannesburg, Original and Corrected Lights



Uses and Pitfalls of Nighttime Lights in Urban Analysis

- What are we actually measuring?
  - Nighttime light data can be viewed as the product of **population** and **light per capita**.
  - Does it represent investment? Consumption? Public infrastructure?
  - Some sectors of the economy are more light-intensive than others: Energy saving measures and a changing economic structure (services!) lead to a weaker relationship between light growth and economic growth in rich countries
- **Growth of cities** in developing countries: Bluhm and Krause (2022) show that in Sub-Saharan Africa, primary cities grew at least as fast as secondary cities in the period 1992-2013 (no catch-up by secondary cities as previously thought)
- **Identification of city structure**, monocentric vs polycentric urban layout in cities around the world $\Longrightarrow$ Still very useful in all countries, irrespective of light's relation with economic growth

Urban Data Analysis

- Once you have obtained your urban data, you will want to analyze it quantitatively
- **Statistical data analysis** comprises various methodological fields with different goals
  - **Descriptive data analysis**
    - ★ Computing indices, comparing mean values, analyzing the evolution of variables over time $\Longrightarrow$ Tables, histograms, city maps, time lines
  - **Inferential / econometric analysis**
    - ★ Aims to establish a causal relation between two variables, compute the magnitude of the ceteris paribus effect of a unit change of x on y and test its statistical significance $\Longrightarrow$ Very challenging in urban research due to endogeneity, not the goal of the course projects
  - **Predictive analysis**
    - ★ Predicting a y-variable as accurately as possible, based on a number of x-variables that may enter in a nonlinear way $\Longrightarrow$ Using big data and machine learning methods $\Longrightarrow$ 'Big Data and Smart Cities'

Urban Data Analysis

- In this course, we will be analyzing cities quantitatively, but mainly with **descriptive statistics**
  - **Standard statistical techniques**: Documenting mean values, the (graphical) distribution of values such as mean income across different parts of the city, tracing the evolution of variables over time and identifying time trends. Data presentation, interpretation and visualization are important.
  - Critical use of the data and results from the literature: How are the insights presented? Selective use of data to convey a certain message?
  - Be careful with **correlation** and **causality**: Urban life is complex and many variables are highly endogenous.
    - ★ What causes what? Are the streets cleaner in this district because rich people live there? Or does a district with cleaner streets attract wealthier residents?
    - ★ Insights from the literature: When reading a paper that establishes causality (e.g. of policy measure), look at the econometric techniques used.

Growth Rates of Rental Prices

- Analyzing urban evolution over time often involves the calculation of growth rates.
- Exercise: How much have rental prices in Leipzig increased in recent years?
- Let us consider the rental price for existing rental contracts (median price in EUR per square meter without incidental rental costs, source: Citizens' Survey Leipzig)
  - 2022: 6.60
  - 2023: 6.88
  - 2024: 7.14
- Calculate by how much (in %) rental prices increased from 2022 to 2023, and from 2023 to 2024. Which of the two yearly price increases is higher?
- By how much have rents increased on average per year between 2022 and 2024?

Growth Rates of Rental Prices

- Calculation of a growth rate g_0 of a variable x, measured at two points in time 0 and 1 (for example two years)

$$g_0 = \frac{x_1 - x_0}{x_0} \tag{1}$$

- Rental prices increased by 4.24% from 2022 to 2023 and by 3.78 % from 2023 to 2024.
- How to calculate the average of two growth rates g_0 and g_1?
- It's the *geometric* average:

$$g_{av} = \sqrt{(1+g_0)\cdot(1+g_1)} - 1 = \sqrt{1.0424 \cdot 1.0378} - 1 = 0.0400 = 4.00\%$$

- Note that it would be *wrong* to take the arithmetic average of growth rates because growth rates are relative values:
$\frac{1}{2} \cdot (4.24\% + 3.78\%) = 4.01\% \neq 4.00\%$

Growth Rates of Rental Prices

- Is the difference between 4.00% and 4.01% really that much of a deal?
- Let's look at total growth rates and average growth rates over a longer horizon (12 years):
- In 2012, the median rental price in Leipzig was 5.15. In 2024, it was 7.14.
- Compute...
  - ...total price increase (in %) between 2012 and 2024
  - ...the yearly average price increase (in %) between 2012 and 2024

Growth Rates of Rental Prices

- The total rental price increase over the 12 years was

$$g_0 = \frac{x_{12} - x_0}{x_0} = \frac{7.14 - 5.15}{5.15} = 38.64\% \tag{2}$$

- That corresponds to an average yearly increase of

$$g_{av} = \left(\frac{x_{12}}{x_0}\right)^{\frac{1}{12}} - 1 = 2.76\% \tag{3}$$

- ... which is *not* the arithmetic average $\frac{1}{12} \cdot 38.64\% = 3.22\%$.
- The geometric average is always smaller than the arithmetic average. It takes into account that growth rates are relative rates, applied to increasing levels over time (compound interest effect): Only in the first year, the growth rate of 2.76% is applied to the starting value of 5.15. In the second year, it is applied to $5.15 \cdot 1.0276 = 5.29$ and so on.

Growth Rates of Rental Prices

- In general: Starting at a value x_0, the variable reaches a value of x_N after N years. The yearly growth rates are g_0, g_1,... g_{N-1}:

$$x_N = x_0 \cdot (1 + g_0) \cdot (1 + g_1) \cdot (1 + g_2) \cdots (1 + g_{N-1}) \tag{4}$$

- This is equivalent to growing by the average yearly growth rate g_{av}:

$$x_N = x_0 \cdot (1 + g_{av})^N \tag{5}$$

- You therefore obtain the average yearly growth rate as

$$g_{av} = \left(\frac{x_N}{x_0}\right)^{\frac{1}{N}} - 1 \tag{6}$$

or as the geometric mean of the individual growth rates:

$$g_{av} = [(1 + g_0) \cdot (1 + g_1) \cdot (1 + g_2) \cdots (1 + g_{N-1})]^{\frac{1}{N}} - 1 \tag{7}$$

Growth Rates of Rental Prices

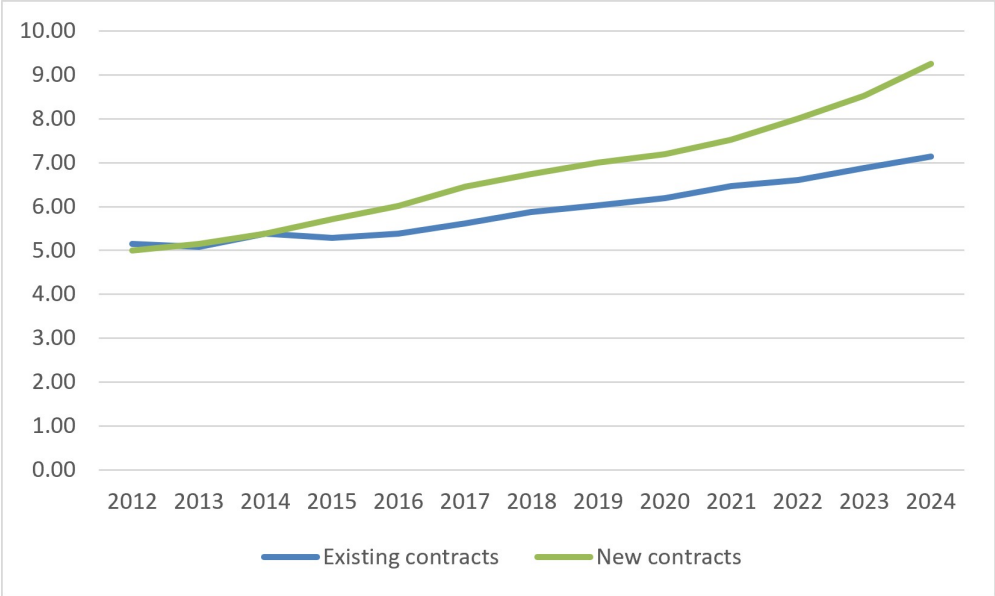
- In the excel sheet *Unit5_RentalPricesLeipzig.xlxs*, you find both the prices for *existing* rental contracts and those for *new* rental contracts (source: Value Marktdatenbank) for the years 2012 to 2024 in Leipzig.
- Compute the yearly growth rates for each of the rental prices. Which differences do you note?
- Compute the average yearly growth rate of each of the rental prices. Which of the two has gone up by more? Why do you think this is the case?
- Plot the two time lines. How does the choice of the minimum and maximum values on the y-axis influence the presentation - and the message that is conveyed?

Growth Rates of Rental Prices

Year	Existing contracts	Yearly growth	New contracts	Yearly growth
2012	5.15		5.00	
2013	5.08	-0.0136	5.15	0.0300
2014	5.38	0.0591	5.38	0.0447
2015	5.29	-0.0167	5.71	0.0613
2016	5.39	0.0189	6.01	0.0525
2017	5.62	0.0427	6.46	0.0749
2018	5.88	0.0463	6.74	0.0433
2019	6.03	0.0255	7.00	0.0386
2020	6.20	0.0282	7.19	0.0271
2021	6.47	0.0435	7.52	0.0459
2022	6.60	0.0201	8.00	0.0638
2023	6.88	0.0424	8.53	0.0662
2024	7.14	0.0378	9.26	0.0856
	Average growth	0.0276	Average growth	0.0527

- The average rental price increase of existing rental contracts was 2.76% per year ($= (B14/B2)^{1/12} - 1$) , while it was 5.27% for new rental contracts on the market. Increasing demand, city growth due to migration from other parts of the country? Inflation feeding into new contracts more quickly?

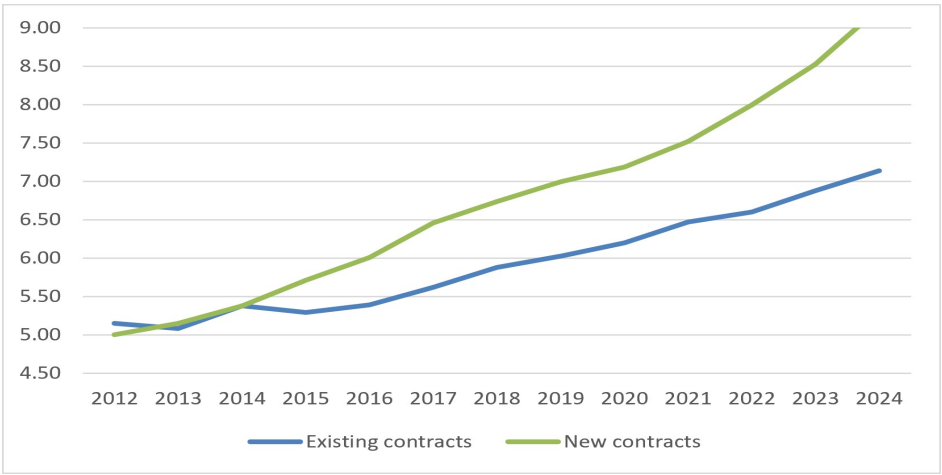
Growth Rates of Rental Prices



- We note a certain divergence of the two time series from 2013/2014 onwards.
- How does the choice of the maximum and minimum values of the axis determine the message that is conveyed about rental prices? (Click on the axis -> *Format axis* to try yourself)

Growth Rates of Rental Prices

- A dramatic increase or staying at a similar level?
- And what about the years on the x-axis? Why start in 2012 and not in 2014 or 2015? What is the yearly average growth rate if you start in 2014?





- *"There are three kinds of lies: Lies, Damned Lies, and Statistics"* (Mark Twain)

Wrap-Up of Unit 5

- The amount of **data** available for urban analysis has increased a lot in recent years, ranging from official statistics coming from administrative sources over the Open Street Map project to satellite data of nighttime lights. The Global Human Settlement Layers provide data on the built-up extent as well as average building height at a fine resolution.
- **Statistical data analysis** can be descriptive, inferential or predictive.
- In **descriptive urban data analysis**, indicators are computed and compared, the distribution of values is analyzed and the evolution of variables over time is examined.
- This often involves the calculation of **growth rates**. The average growth rate over a longer period is the geometric mean of the individual growth rates.

Bibliography of Unit 5 I



Bluhm, R. and M. Krause (2022).

Top Lights - Bright Cities and Their Contribution to Economic Development.

Journal of Development Economics 157, 102880.



Donaldson, D. and A. Storeygard (2016).

The View from Above: Applications of Satellite data in Economics.

Journal of Economic Perspectives 30(4), 171–198.



Florczyk, A., C. Corbane, D. Ehrlich, S. Freire, T. Kemper, L. Maffenini, M. Melchiorri, M. Pesaresi, P. Politis, M. Schiavina, F. Sabo, and L. Zanchetta (2019).

Description of the GHS Urban Centre Database.

Technical report, Publications Office of the European Union, Luxembourg.



Henderson, V., A. Storeygard, and D. N. Weil (2012).

Measuring Economic Growth from Outer Space.

American Economic Review 102(2), 994–1028.

Bibliography of Unit 5 II



Pesaresi, M., M. Schiavina, P. Politis, S. Freire, K. Krasnodębska, J. Uhl, A. Carioli, C. Corbane, L. Dijkstra, P. Florio, H. Friedrich, J. Gao, S. Leyk, L. Lu, L. Maffeninik, I. Mari-Rivero, M. Melchiorri, V. Syrris, J. Van Den Hoek, and T. Kemper (2024).

Advances on the Global Human Settlement Layer by Joint Assessment of Earth Observation and Population Survey Data.

International Journal of Digital Earth 17, 2390454.



Seidel, A. (2023).

A Global Map of Amenities: Public Goods, Ethnic Divisions and Decentralization.

Journal of Development Economics 164, 103113.