



STATISTICS

Workshop on Property Price Indices and Real Estate Statistics Training

International Conference on Real Estate Statistics (ICRES)
Tokyo, Japan, February 18, 2026

Mr. Barra Casey – bcasey@imf.org

Overview of presentation

- Using training material from the **Residential Property Price Index (RPPI) in-person training course** (normally 5 days).
- Primarily aimed at **compiling agencies** – statistical offices and central banks.
- Interspersed with concrete **examples from Kosovo** (with their permission), they have recently published a first-time RPPI.
- Feel free to **ask questions or make a comment**.

IMF capacity development

- IMF **capacity development** (CD) program covers all **real estate price statistics** (and additional indicators) – includes residential, commercial and land transactions.
- CD is **demand-led**, in practice most of the work is on **residential real estate** – we do some work bilaterally with countries on commercial and land prices.
- Elements of CD:
 - RPPIx **online training** course
 - RPPI **in-person regional workshops**.
 - **Bilateral** country engagements.

RPPIx - IMF online training course



 **858**
Registered Participants

 **398**
Active Participants

 **258**
Passed Participants

 **54.26%**
Active/Passed Rate



Location*

● Indonesia	4.02%
● Honduras	3.27%
● Vietnam	3.02%
● Ghana	2.76%
● Kenya	2.76%
● Cambodia	2.51%
● South Africa	2.51%
● United States	2.51%
● Cameroon	2.26%
● Argentina	2.01%

<https://www.imf.org/en/capacity-development/training/icdtc>

- Available in English, Spanish and French

Synthetic House Price Dataset – can be used for training purposes



Contact Us



More information

Please email a request for access to the synthetic dataset to the Division Chief of the Real Sector Division of the Statistics Department (rppi@imf.org) of the IMF.

The request should include information on:

- The reason for access (for example, research, study)
- Organization details (if any)
- Other relevant information

A register of users who receive the Synthetic House Price Dataset will be maintained by the IMF.

Why is the RPPI important for policymaking?

- **Capture trends** in prices and real estate market cycles
- Critical for **financial stability assessments** - assessing risks in banking sector exposure – IMF initiated the Financial Soundness Indicators (FSIs).
- Inform **macroprudential supervision frameworks** - provide early warning signals of systemic risks.
- Supporting **fiscal and monetary policy** analysis
- Households use property price information to **make decisions** on whether to buy or sell.

Aim and scope of the RPPI

- The Residential Property Price Index (RPPI) measures the change in the **average market price of residential property over time**.
- Should be a **monetary transaction** on the **economic territory** of the country.
- **All transactions of residential dwellings** should be covered – all types, all vintages, urban and rural **regardless of the purpose** i.e., let or own occupation.
- Other types of real estate transactions are out of scope e.g. **land, commercial properties and rents**.
- Exclude: **non-market** transactions, **non-household** transactions e.g. corporate, government

General steps for RPPI construction (iterative approach)

1. Identifying a data source(s) – most important aspect
2. Data analysis – getting to know your data.
3. Data preparation, cleaning, filtering
4. Stratification – creating homogenous sub-samples
5. Weights – for aggregation purposes
6. Model selection for hedonic approach
7. Compilation and aggregation
8. Dissemination – statistical release and methods documents

A note on technology

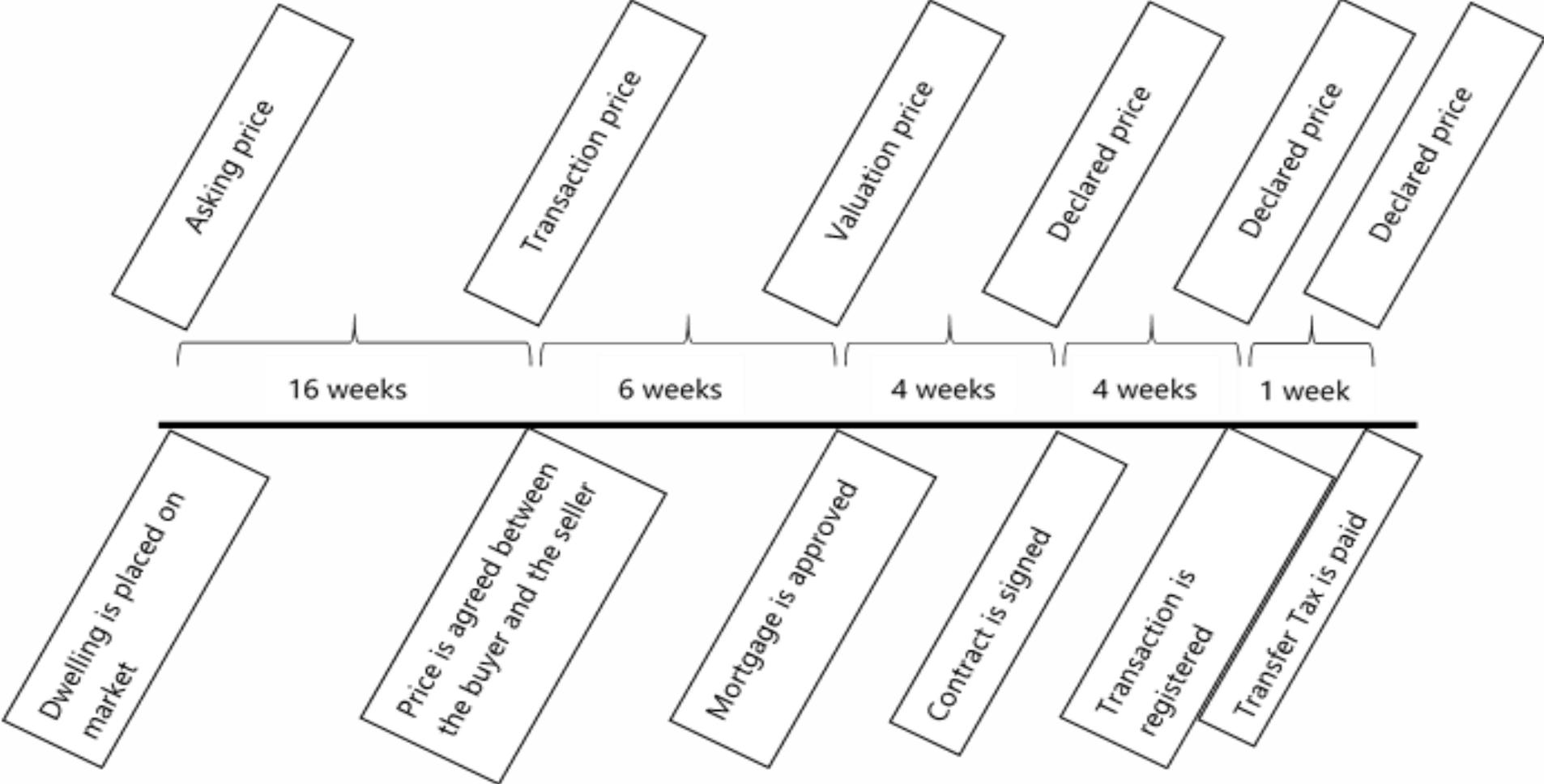
- It is **not possible to compile the RPPI in MS Excel** - even during the development stage of the RPPI.
- Use a **statistical package** such as R, Stata, Python, SAS etc.
- Training might be required – however **AI chatbots** make building the processes in a statistical package very easy.
- IMF assist with this; however, we **encourage countries to build their own production processes**.

Step 1. Identifying a data source(s)

The transaction process

- As a first step in designing an RPPI, compilers should acquire an in-depth understanding of the **real estate transaction process** within their country.
- This will allow the compiler to understand **the type of information generated and where data can be obtained.**

Example of a timeline of the transaction process



Price concepts and their fitness for purpose

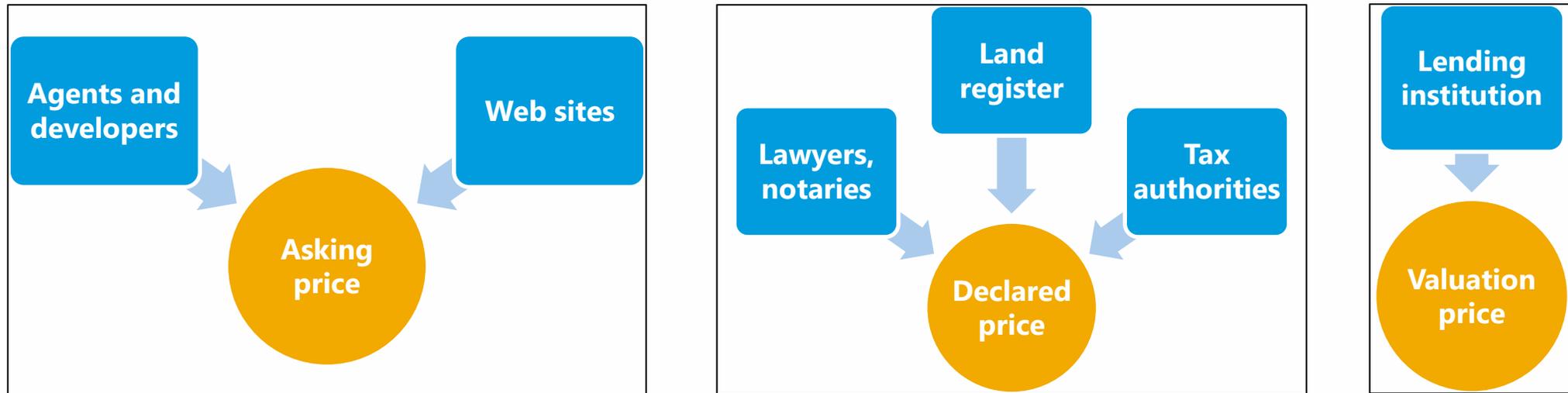
Asking price – Corresponds to the price asked when the dwelling is put on the market. The final transaction price may be higher or lower.

Transaction price – After negotiations the buyer and the seller agree on the price. **This is the target price, e.g., the price actually paid.**

Valuation price – The dwelling is evaluated by the lending institution. The value of the appraisal can be higher, lower or equal to the transaction price.

Declared price – This is the “legal” price of the dwelling, normally used for tax purposes. The registration of the selling price will normally happen later than the transaction date and declared prices may differ from the transaction prices. Depending on the size of these differences it can be or not a suitable proxy.

Data sources and price concepts



Countries have different sources for data available that collect prices from different perspectives. It is important to understand which data source can provide a price closer to the transaction price.

Key criteria

- **Transaction-level** data – not aggregated.
- **Accurate transaction prices** (or best proxy) – and accuracy of the other variables.
- **Key property characteristics** to control for different properties transacting in different periods - will vary by market segment
- **Good coverage** by geography, type of dwelling, vintage
- Can identify the **type of transaction** (market or non-market), **type of seller/buyer**
- **Accessible** (e.g. paper, digital format), **sustainable** collection
- **Timeliness** and **frequency**

Key considerations

No data source is perfect.	+
Getting started is important.	+
Negotiating access to data sources can take time.	+
Consider the pros and cons of data matching.	+

Kosovo example

- Spent **significant amount of time** examining the pros and cons of data from real estate agents, the tax office and the Cadastral Agency – **IMF team facilitate this process.**
- An **inter-governmental group** was established to examine the issue and progress the development of the RPPI
- Project started in 2023 – it was only in 2025 that an **official agreement was reached** between the statistical office and the Cadastral Agency
- There is a lot of back and forth with the data providers **to understand the data collection.**
- Recommendation is to **allocate time for this step** as part of the development process.

Step 2. Data inspection

Data inspection

When investigating economic variables, it is generally useful to start with **graphical representations** to detect those properties of the variable which can be seen by simply looking at the plot of a variable.

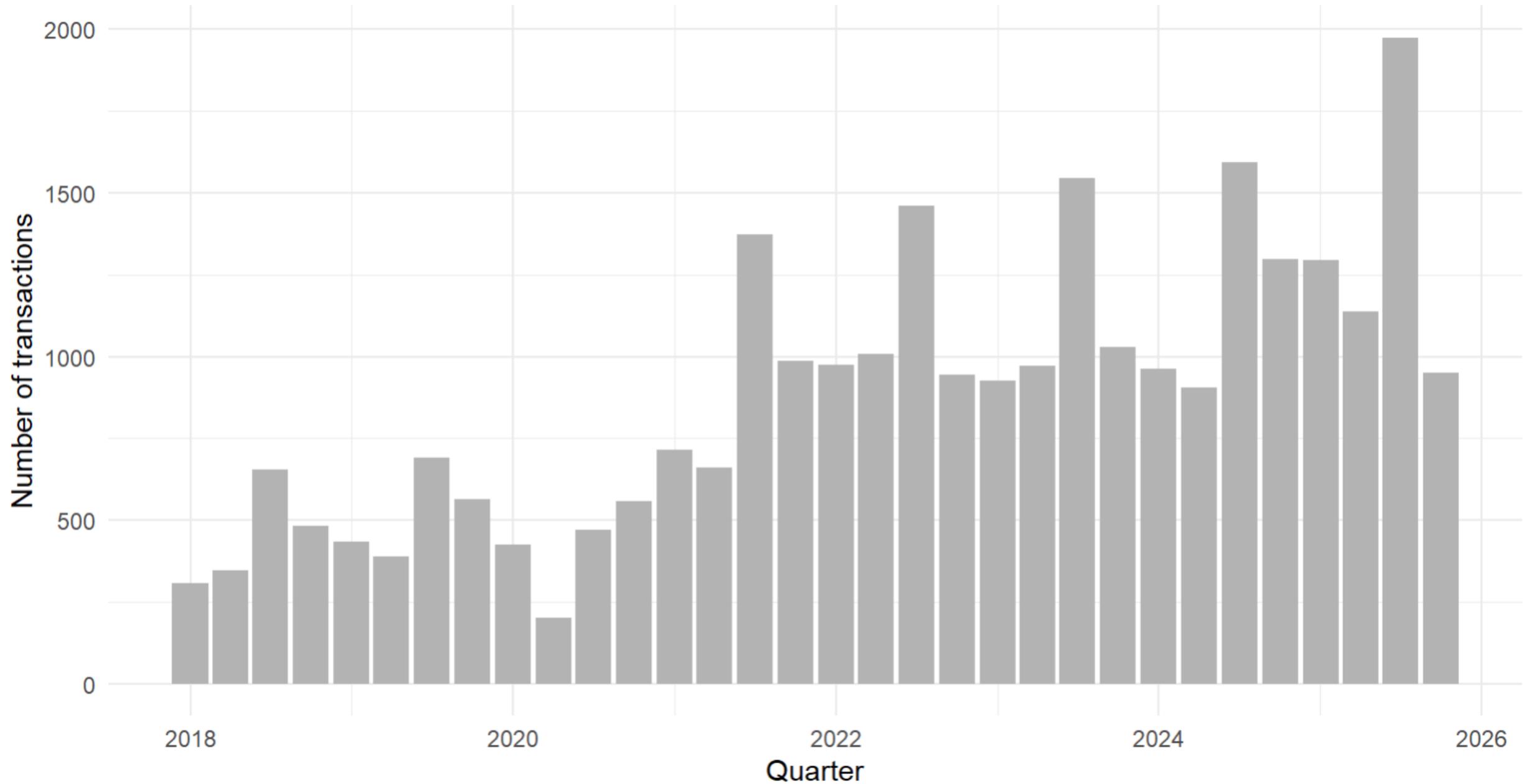
- ▶ Bar charts
- ▶ Scatterplots
- ▶ Boxplots and histograms
- ▶ Time series line plots

Consider **different transformations** of the variable to be analyzed, for example, its levels and its logarithms.

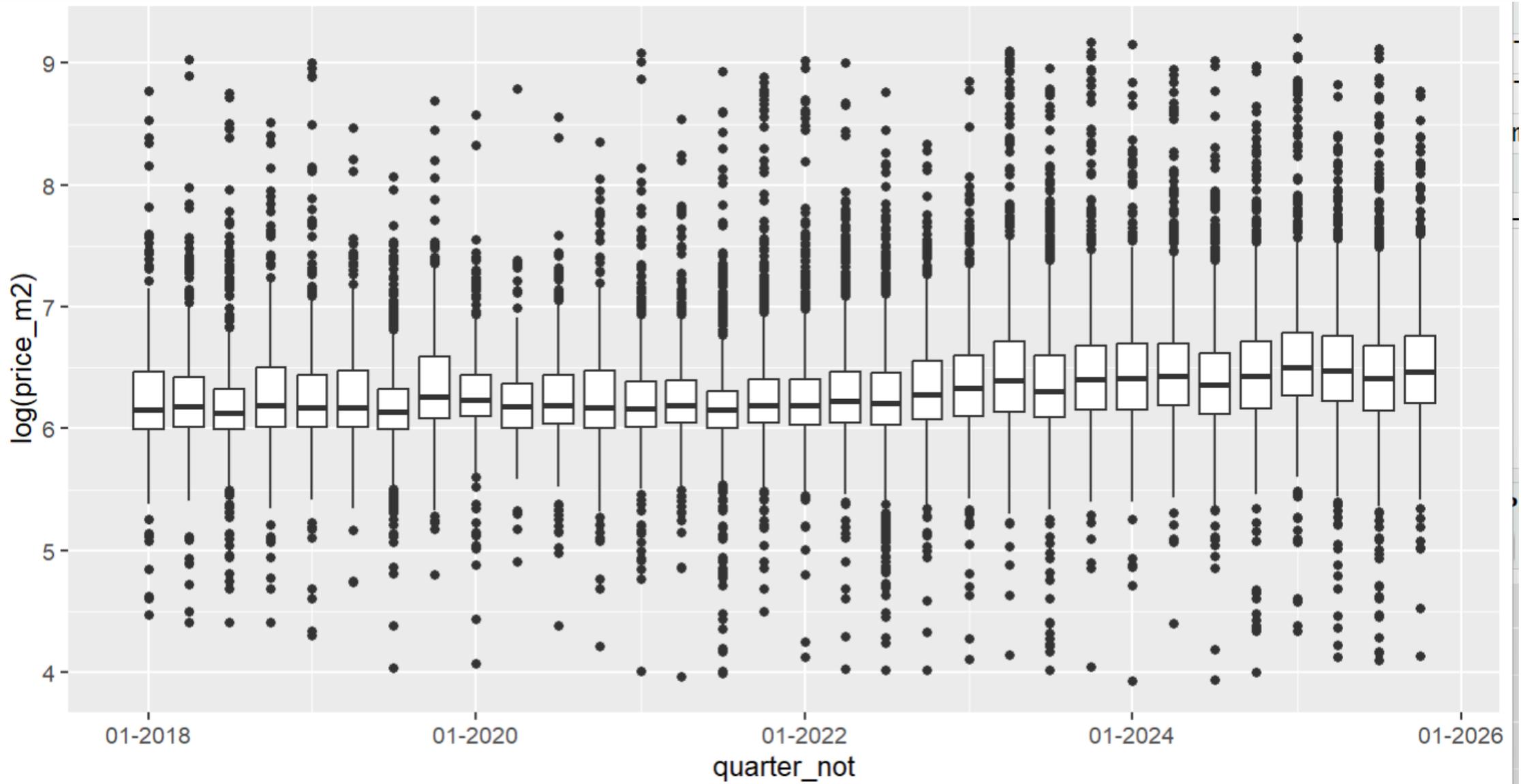
Econometric analysis - helps to **detect regularities in the observations** of a variable and derive 'laws' from them.

Should be completed **before and after the data cleaning** step.

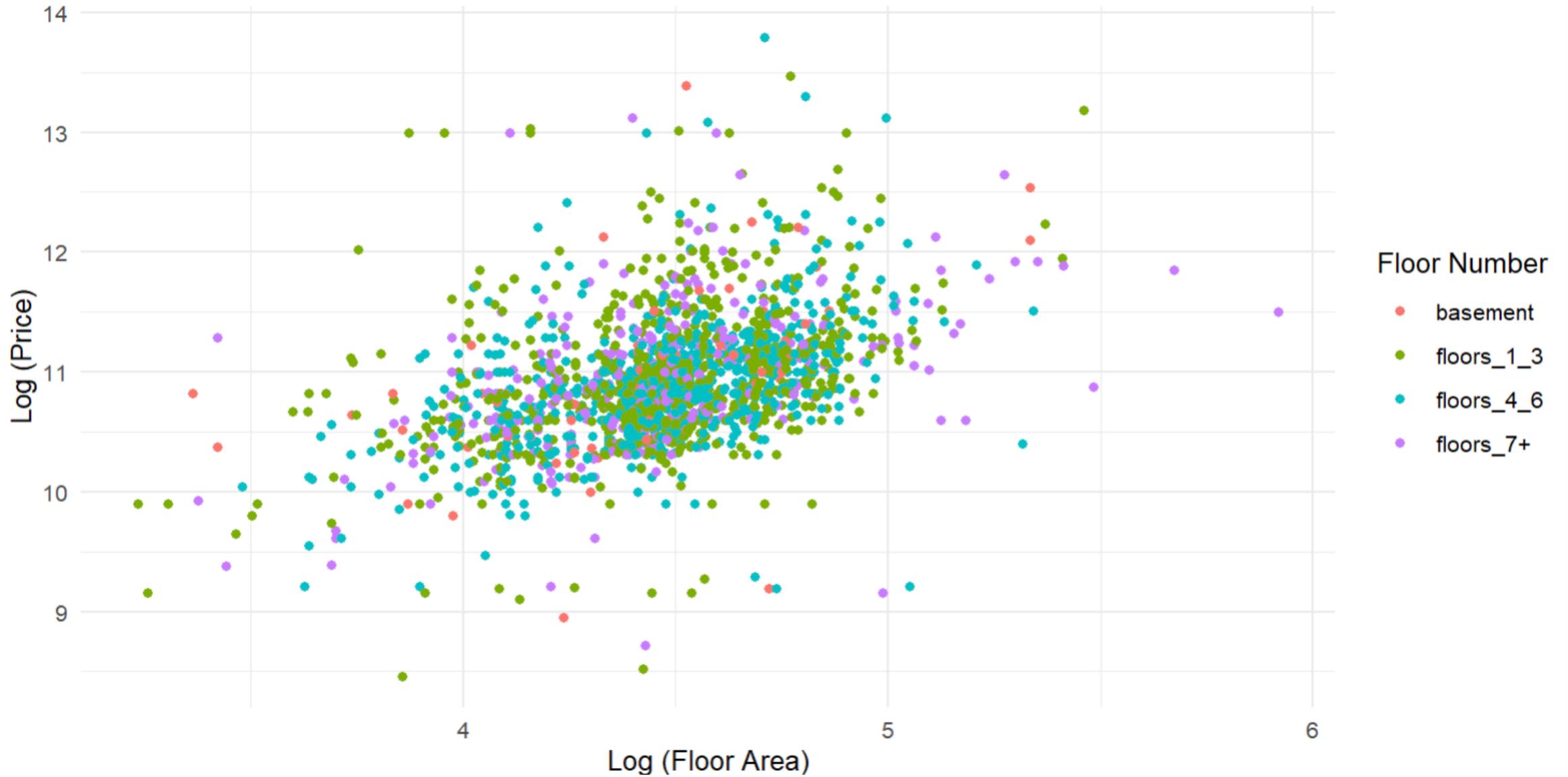
Number of Dwellings by Quarter



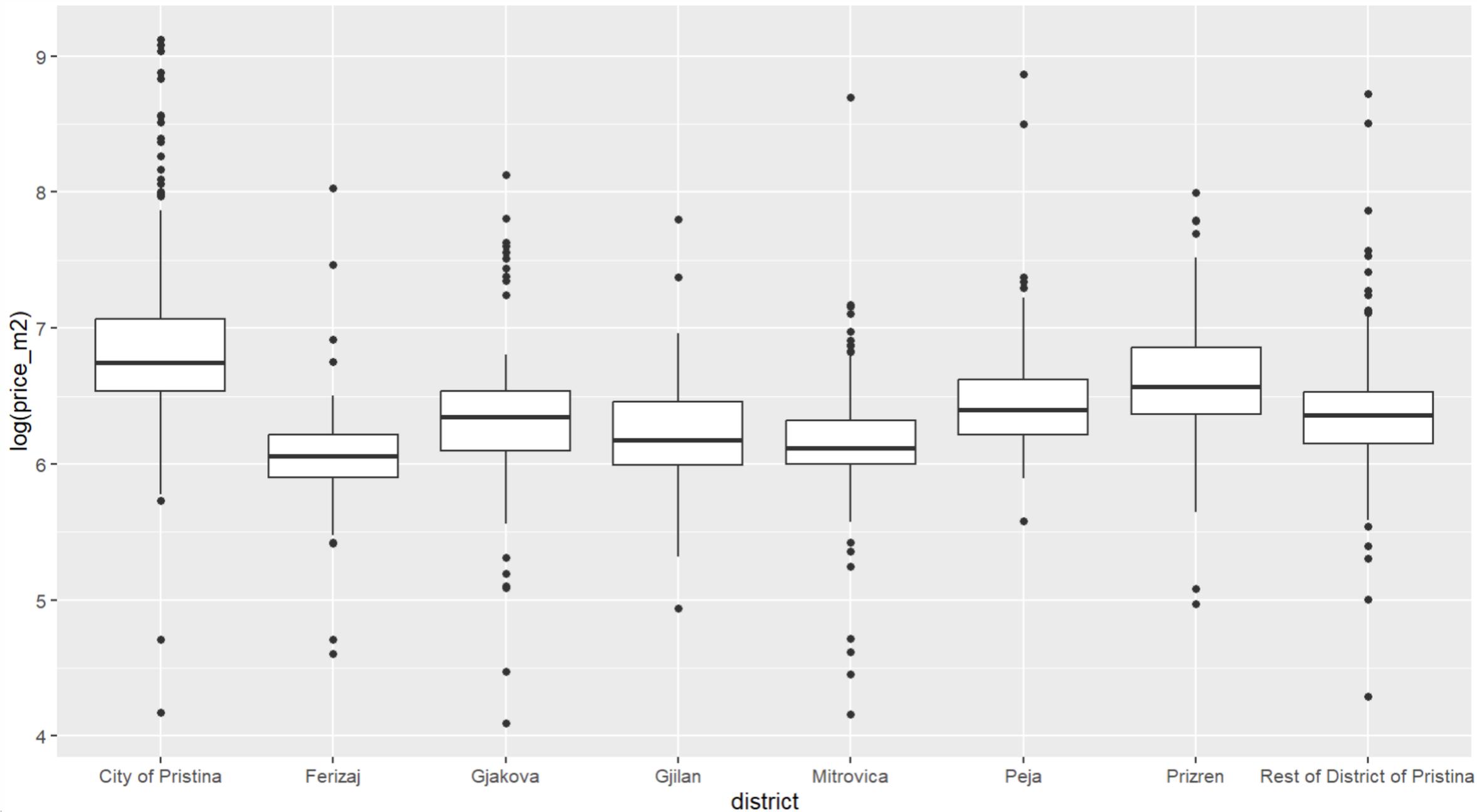
Quarterly boxplots for price per square meter – all residential property



Scatter Plot of Log (Price) vs Log (Floor Area), Latest Quarter



Boxplots for price per square meter by district – latest quarter



Kosovo - timeliness of the Cadastral data

- **Notary registrations are frequently incomplete** at the time of first publication e.g. 60-80% of registrations are received in time for the first publication of the index i.e. 2 months after the reference quarter.
- It **takes time** for the notaries to complete the administrative step of registering the real estate transactions with the Cadastral Agency.
- **Accuracy is improved** by using subsequently updated Cadaster data - compile the RPPI for the latest quarter and publish it as “**preliminary**”, and re-compile the previous quarter, using the updated file, and publish it as “**final**”.
- This revision policy should be **clearly documented** in the RPPI methods paper.

Step 3. Data processing, cleaning, filtering

Data processing

At a minimum, compilers should assess the data in terms of:



- Out of scope are **non-market** transactions (best efforts), **non-household** buyers (if possible) and **non-residential** real estate.
- **Grouping categorical variables** into more manageable categories
- Converting price to a **common currency**, creating **reference period variables** i.e. months and quarters etc.

Back to Kosovo – identifying residential property transactions from the Cadaster file

	% of Total Transactions
Buildings / structures (including land)	33.5
- Residential property	21.5
- Commercial property	9.7
- Mixed-use	1.6
- Infrastructure/public use	0.7
Land only	66.5
- Arable land	35.2
- Grazing land	5.7
- Other agricultural land	12.4
- Forest/natural land	11.8
- Building plot	1.4
Total	100

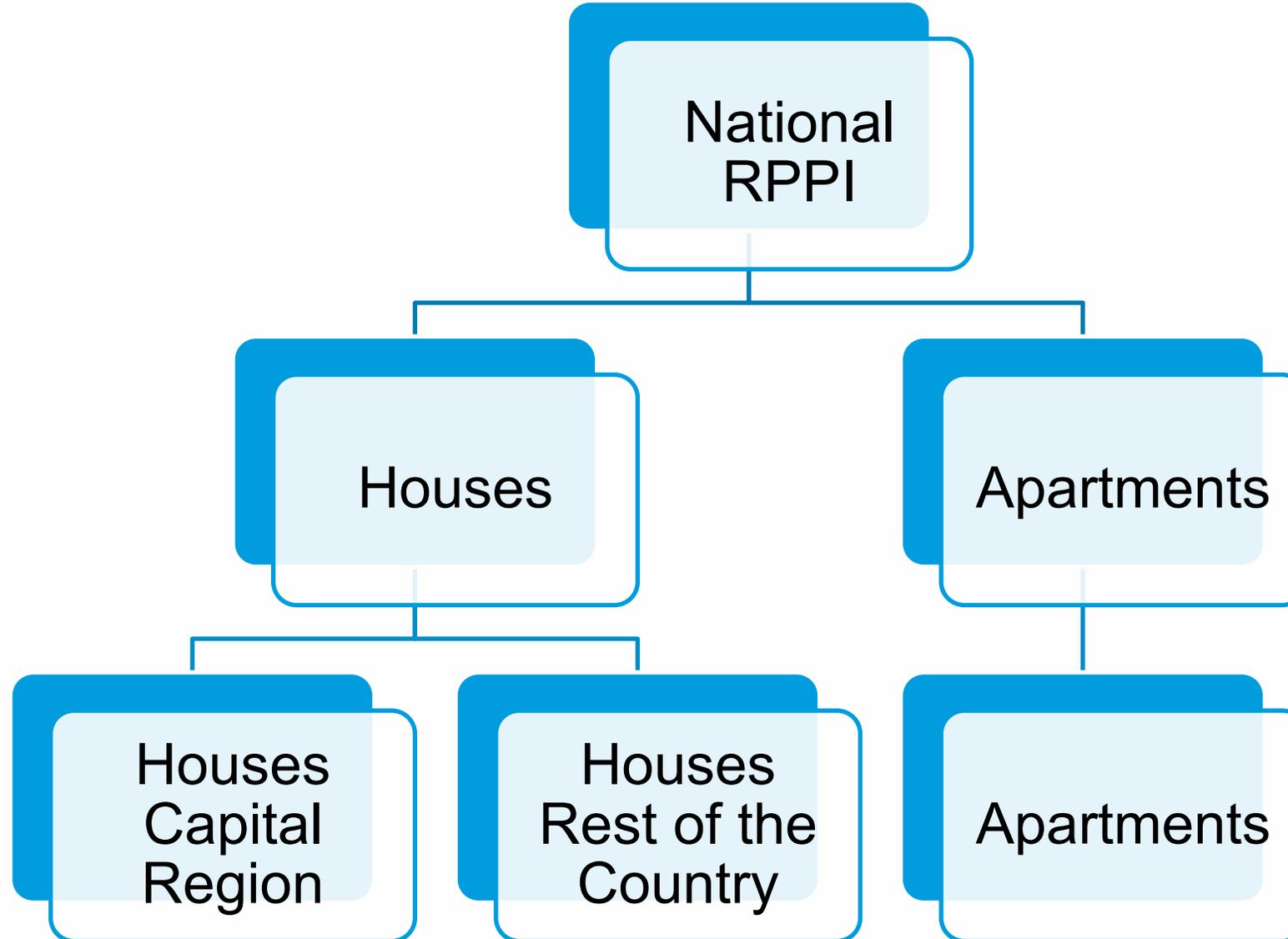
Step 4. Stratification

Stratification

Separating the total sample into **sub-samples** (strata) usually by: **location** (region, district, province) and **type of dwelling** (apartments v houses)

- Goal is to create **homogenous sub-samples** to allow for the calculation of price indices – the same approach is used for the consumer price index.
- This is a very important step as the strata **are used for compilation and dissemination** - compiler needs to consult with users on their needs.
- It is **difficult to amend** the strata after the launch of a new index without revisions to the published figures.
- Stratification is the **first step in the mix adjustment** of the index.
- We use another method to **mix-adjust within the strata** i.e. hedonics.

Stratification



Minimum Number of Observations

What is the minimum number of observations required in a stratum to allow for price index compilation?

* Need sufficient observations – especially during the downturns in the market

Frequency of the Index

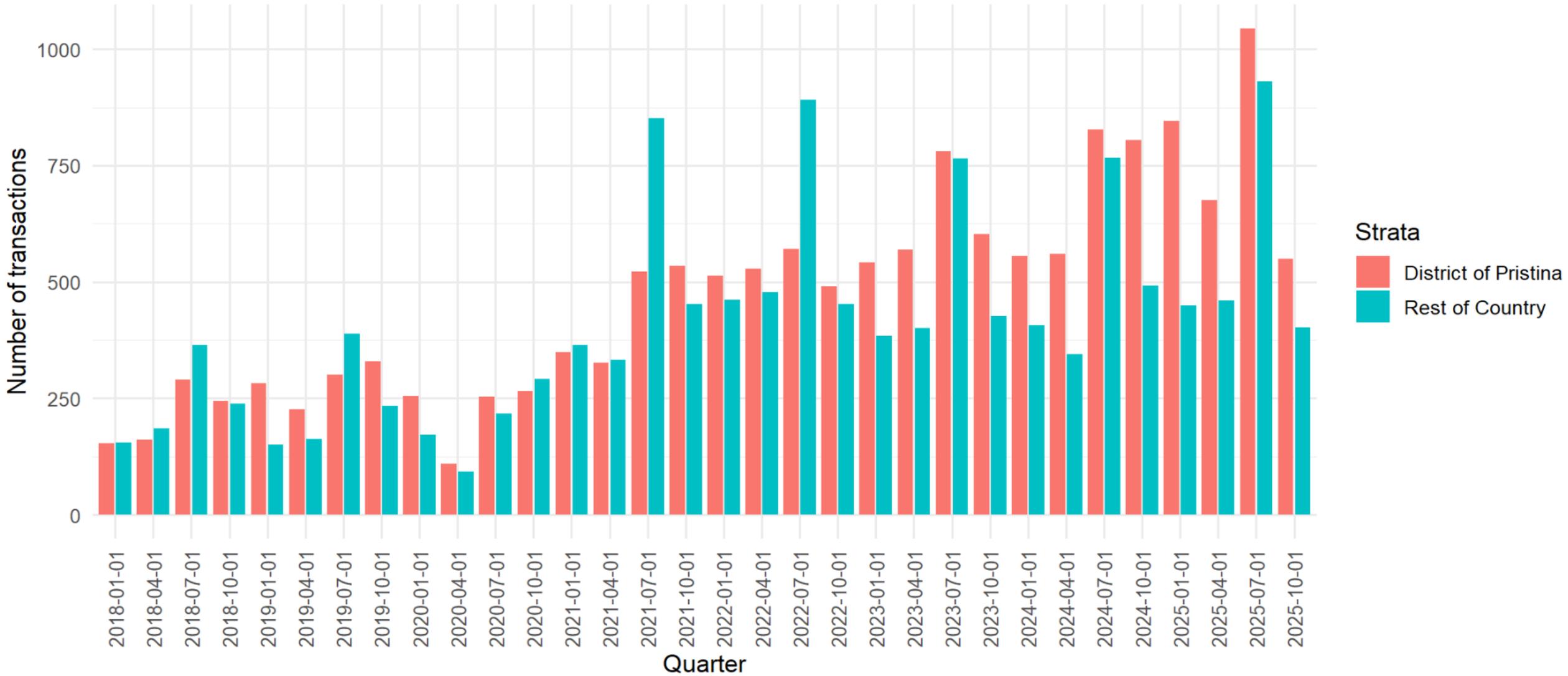


Granularity of the Strata



Kosovo

Number of Dwellings by Quarter



Step 5. Weights

Stock weights vs Flow weights

There are two types of weights that an RPPI compiler can use:

Stock weights:

- Stocks are **positions in time**, i.e., value of all properties in a geographic region on a specific date.
- Can be derived from **census data or valuations for tax purposes**.

Flow (expenditure) weights:

- Flows reflect **changes from one period to the next**, i.e., value of all properties sold during a certain period.
- Can be derived from the **same source used for the prices** (or sometimes from a combination of data sources)

Kosovo RPPI – flows weights updated annually

Year	District of Pristina (%)	Rest of Country (%)
2018	53.8	46.2
2019	51.8	48.2
2020	56.8	43.2
2021	49.8	50.2
2022	52.2	47.8
2023	59.5	40.5
2024	63.7	36.3

Step 6. Model selection for hedonic approach

Measurement challenges

- We generally measure price indices on a **matched model basis** e.g. CPI.
- We want a **constant-quality property price** index but...
 - Properties are **heterogeneous**.
 - Properties are **sold infrequently**
 - **Mix of properties transacted changes** from period to period.
- The **hedonic method** (within the strata) is a **regression-based approach** to adjust average prices to take account of changes in the mix of properties each period.
- Goal: A measure of **pure price change**.

Hedonic approach

- Hedonic regressions recognize that a dwelling is a **bundle of characteristics** (structure attributes and location).
- A set of p characteristics of the dwellings are identified and data over K dwellings are collected.
- A **hedonic regression** of the logarithm of the price of dwelling k , $\ln p_k$, on its set of quality characteristics z_{kj} is given by:

$$\ln p_k = \beta_0 + \sum_{j=1}^p \beta_j \cdot z_{kj} + \varepsilon_k.$$

- [only include **price-determining characteristics** of the properties – **not economic variables** that impact property prices like interest rates, availability of credit etc.]

Kosovo RPPI

- The characteristics of the dwellings (i.e., apartments) included in the models:
 - ▶ **Floor area**
 - ▶ **Location**
 - ▶ **Floor number**
- The **same model was used for both strata**; however, the models can be different.
- In **choosing the explanatory variables** for the model, we need to assess the results and outputs:
 - ▶ Statistical significance of the estimated parameters
 - ▶ Parameters have the expected sign and magnitude – they need to make sense
 - ▶ Explanatory power of the mode (adjusted R-squared)
 - ▶ Graph the residuals

Step 7. Compiling the indices (and aggregating to higher levels)

Options for compiling a hedonic index

The hedonic imputation approach	+
The hedonic characteristics approach	+
The hedonic re-pricing approach	+
The time-dummy hedonic approach	+

The rolling time-dummy (RTD) method

- A hedonic regression is estimated based on pooled data over several consecutive periods, $t = 0, \dots, T$, with dummy variables included for each, but the first, time period ($\delta^0 = 0$) and weights $1/K^t$:

$$\ln p_k = \beta_0 + \sum_{t=1}^T \delta^t \cdot d_k^t + \sum_{j=1}^p \beta_j \cdot z_{kj} + \varepsilon_k,$$

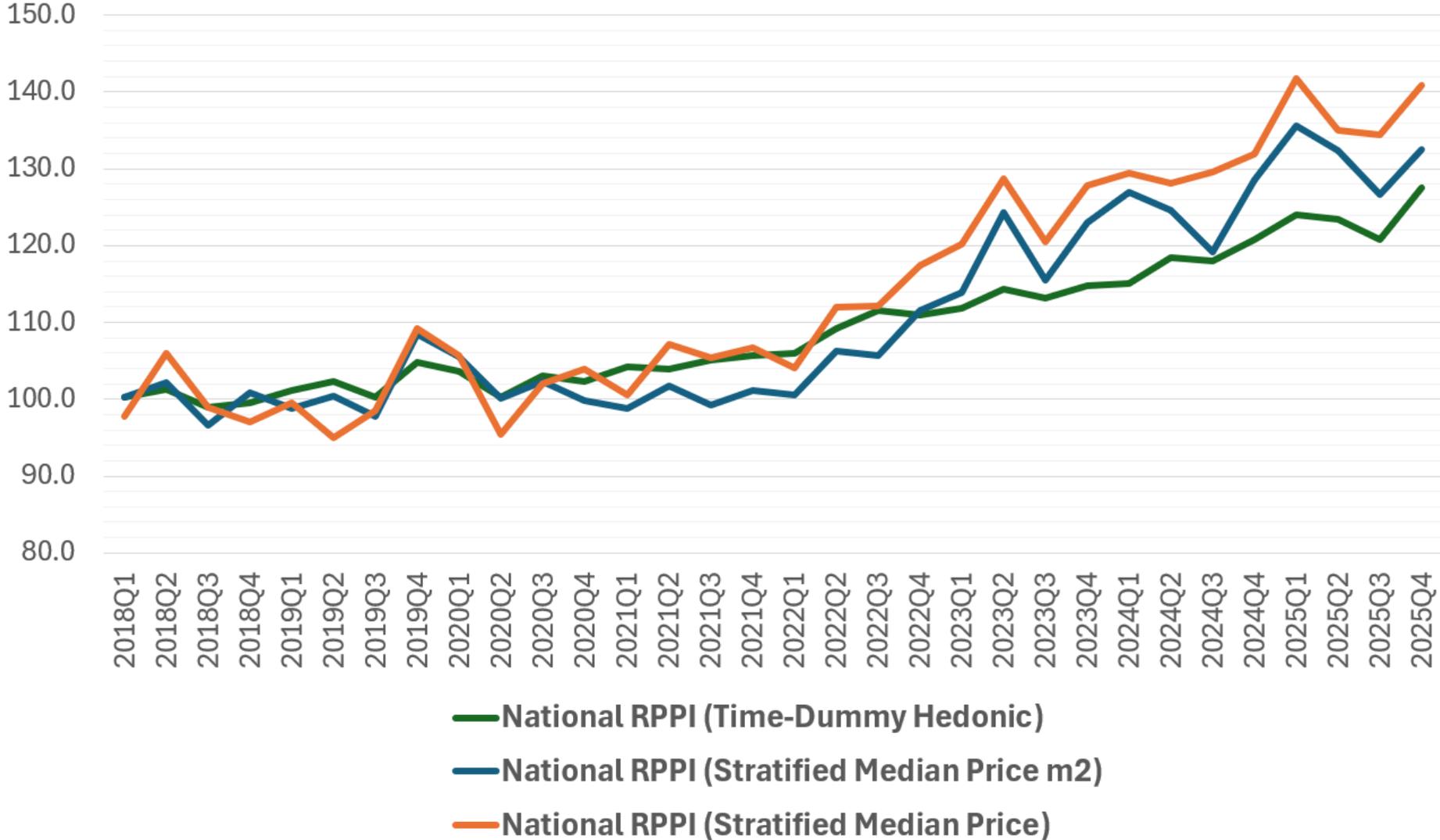
- where the time-dummy variable d_k^t has the value 1 if the observation of dwelling k comes from period t and 0 otherwise.
- The index going from period 0 to period t is given by

$$P_{TDH}^{0,t} = \exp \hat{\delta}^t .$$

Aggregation

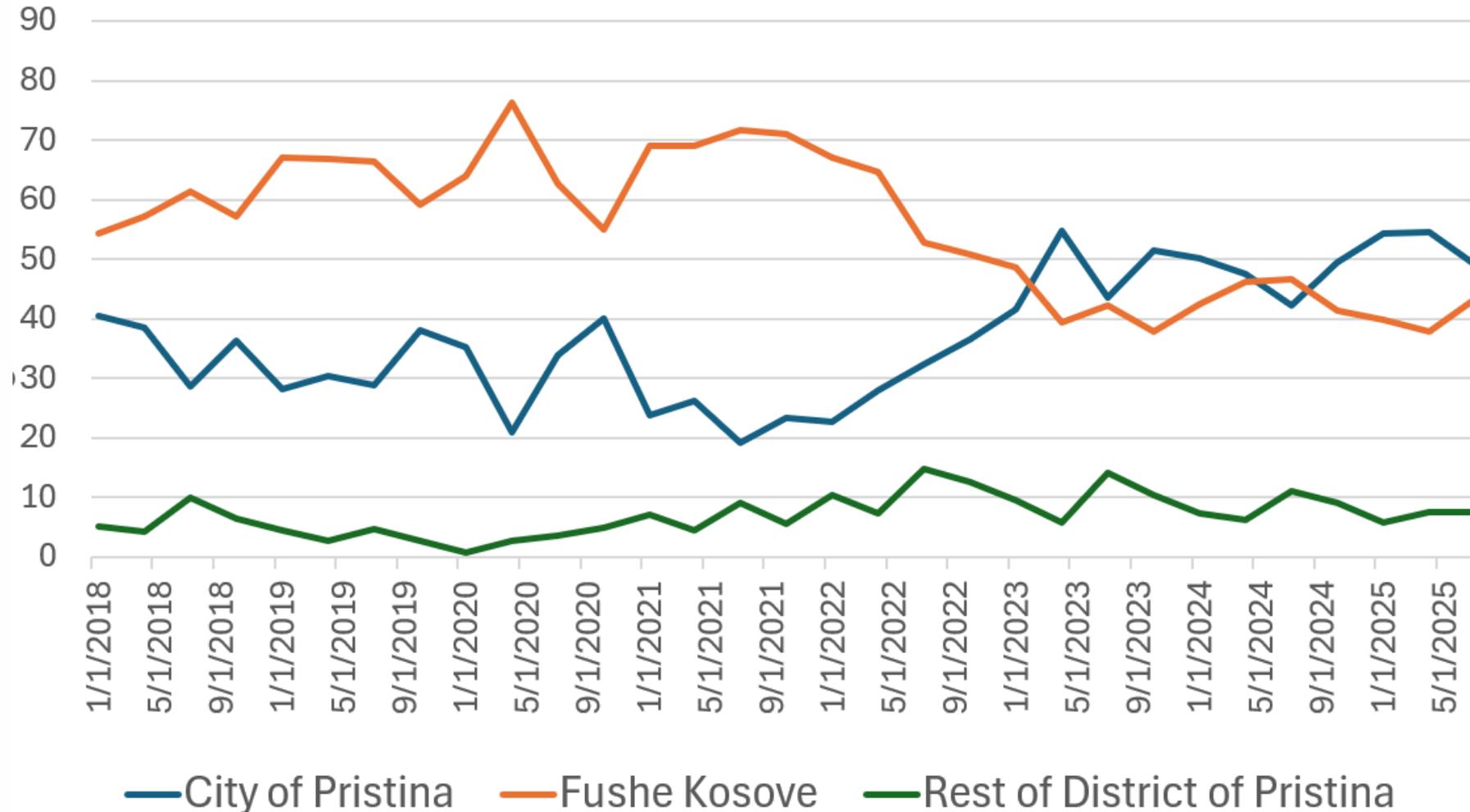
- Options for **fixed weight or chained index**.
- Since the **flow (expenditure) weights** are often available from the same data source as the prices, general recommendation is to **update the weights annually** – this maintains the representativity of the index.
- For **stock weights**, it may not be possible to update annually.
- Weights from the **previous year are applied to the current year strata-level indices** to aggregate to the higher level.

RPPI for Kosovo (2018=100), Hedonic v Stratified Indices, 2018Q1 to 2025Q4



Changing profile of transactions in Kosovo

Location (%) in District of Pristina



Updating the model after initial launch

- The development work can continue after initial launch of the index.
- For example - **adding a new explanatory variable** to the model, change in the specification.
- The new data should be **available for all periods of the rolling window** used by the existing model e.g. 4 quarters or 12 months.
- This will produce a **more accurate estimate of price change**
- Institutional decision required around **revisions** – this will depend on the individual case:
 - ▶ Revise all retrospective periods
 - ▶ Only revise some retrospective periods
 - ▶ Implement the new methods into the future without revisions

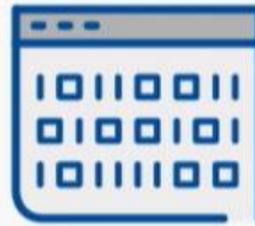
Step 8. Dissemination

Dissemination of the RPPI

Key Aspects to Disseminating the RPPI



**Statistical
Release**



**Accessible
Data**



**Methods
Documentation**

End

Thank you for your participation!