Big data for official statistics: a perspective from price statistics in Brazil

Vladimir Gonçalves Miranda

vladimir.miranda@ibge.gov.br

Brazilian Institute of Gegraphy and Statistics (IBGE)

Prices Indices Coordination/GPLACON

5-th Meeting of the United Nations Network of Economic Statisticians

May 29th, 2024

Outline

- Introductory remarks on CPIs and considerations on potential uses of alternative data sources and data science techniques for CPI production.
- Case studies and discussion on the use of web data and web scraping for CPIs
- Exploration of potential use of administrative records of transacted data.
- Conclusions
- Brief comments on the UN big data Hub for Latin America and Caribe.

Why big data for price statistics?

Technological advancements and commerce practices have provided the appearence of rich alternative data sources such as administrative records, web and scanner data.

Availability of such alternative data sources or/and new techniques open the door for many possibilities to improve the production of price statistics:

- Optimization of collection via use of automatic data acquisition in a more controlled and less prone to human errors manner.
- Incorporation of more elements to the CPI basket: more products and sectors.
- Derivation of more granular and timely indicators: at geographical levels, higher time frequencies.
- Use of more robust index formulations.
- Derivation of more timely and refined CPI weights and consumer patterns.
- Expansion of use to other fields, national and regional PPPs, other price statistics like PPI and construction indicators.

The potential use depends on the source characteristics and structure available.

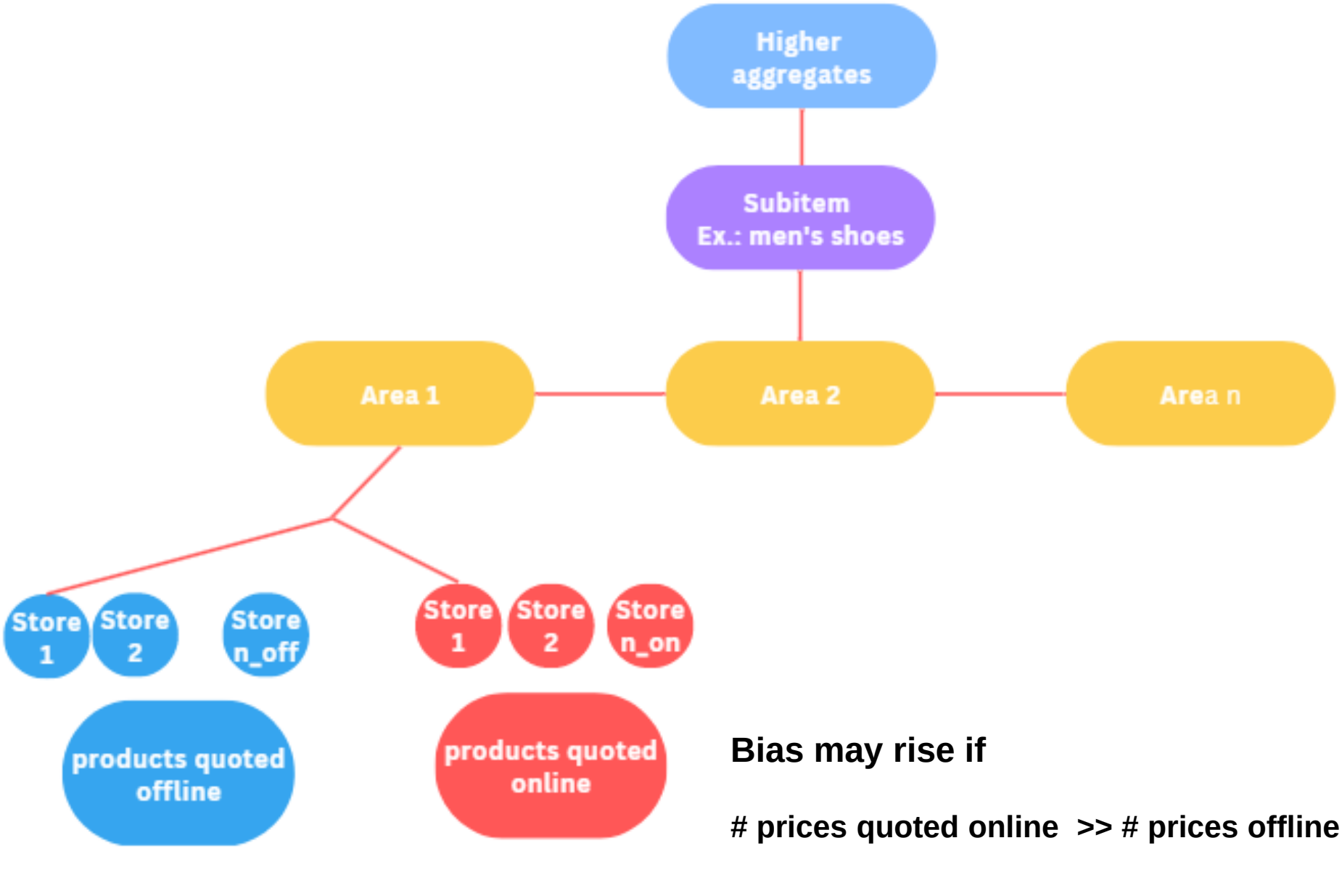
Use may require changes or implementation of many processes in a data pipeline as well as changes in the CPI structure to integrate new data sources.

Core procedures and data integration

- Acquisition
- Preparation
- Classification
- Filtering
- Index calculations

Integration requirements

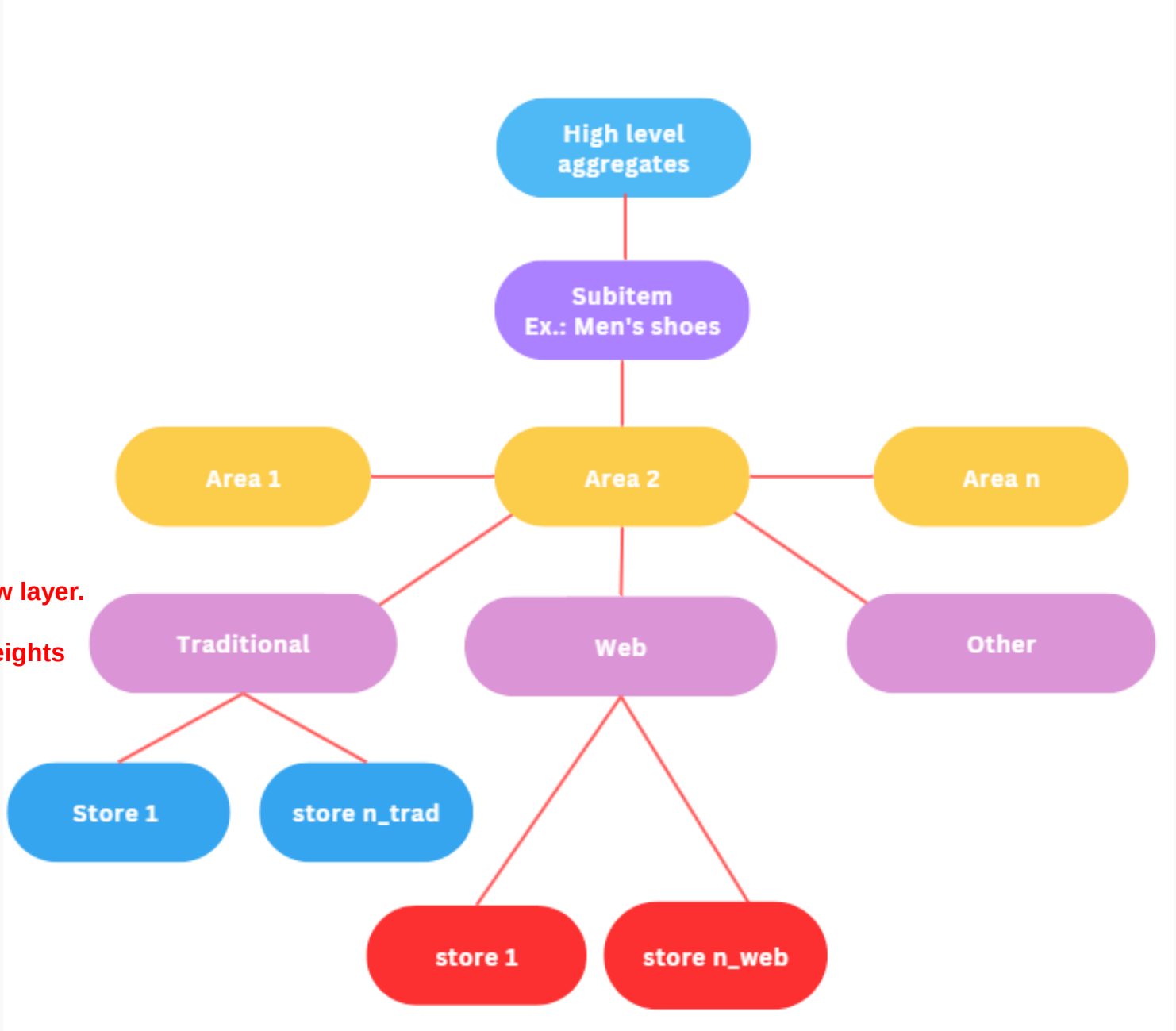
Illustration of integration issues



Bias may rise if

 # prices quoted online >> # prices offline

Illustration of integration issues



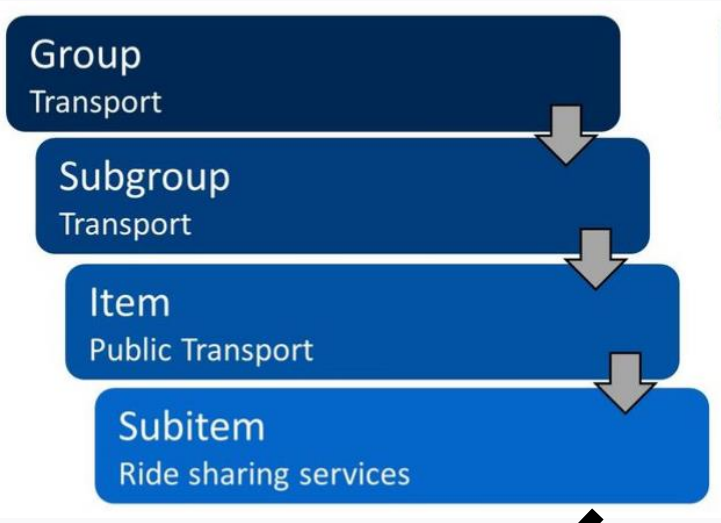
Need to add new layer.

More refined weights data necessary

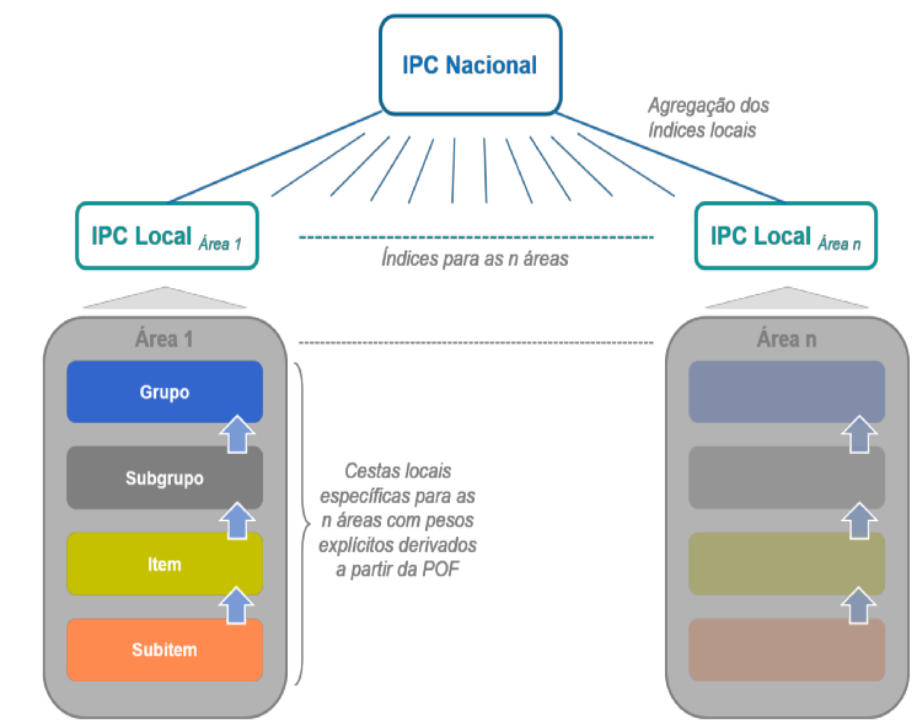
More concretely: CPIs at IBGE

HBS: income and expenditures from households

Classification system



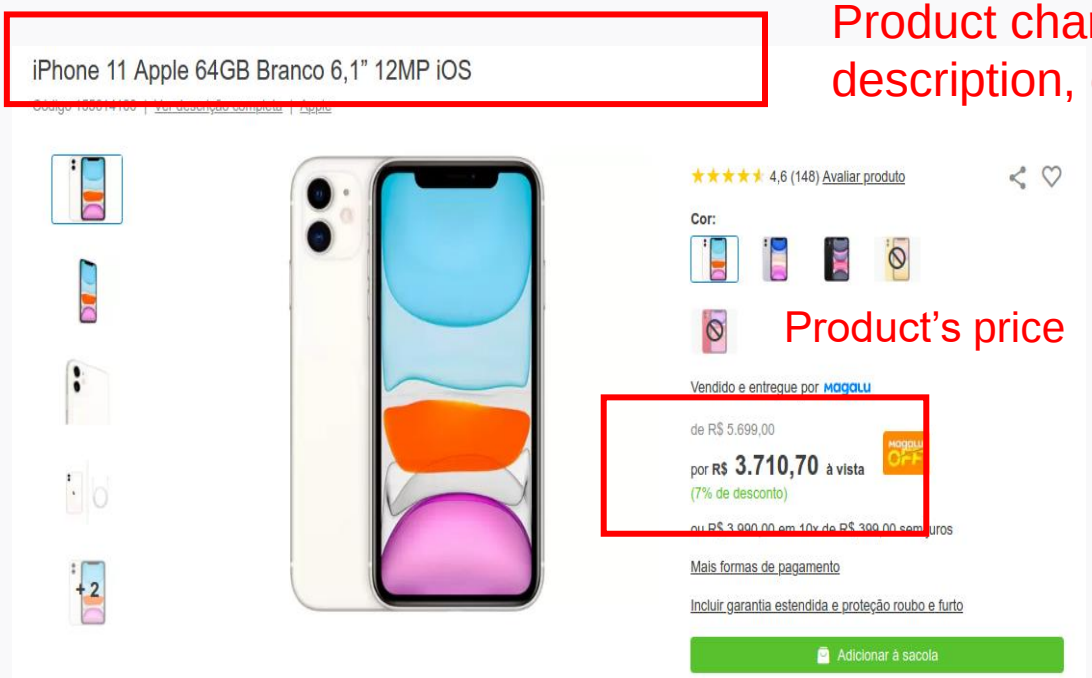
Taxonomy of the system of indicators



Areas covered by different CPIs of the SNIPC.

Web data for prices

Product characteristics: description, code;



Product's price

View from another angle: data within the html page.

```
<div id="anchor-top"></div>
<nav>…</nav>
<div class="header-product js-header-product" data-product="{ "sku": "155614100", "id_pr… "variation_id": "155614100" }">
  <h1 class="header-product__title">iPhone 11 Apple 64GB Branco 6,1" 12MP iOS</h1>
  <small class="header-product__code">…</small>
</div>
<div class="wrapper-product__content wrapper-product__box-prime">
  <div class="showcase-product">
    <ul class="showcase-product__container-thumbs" itemscope="" itemtype="http://schema.org/ImageGallery">…</ul>
    <div class="showcase-product__container-img js-showcase-container js-pop-up js-carousels" data-title="Showcase" data-wrapperid="popup-product" data-content="showcase"> event
```

Other metadata:

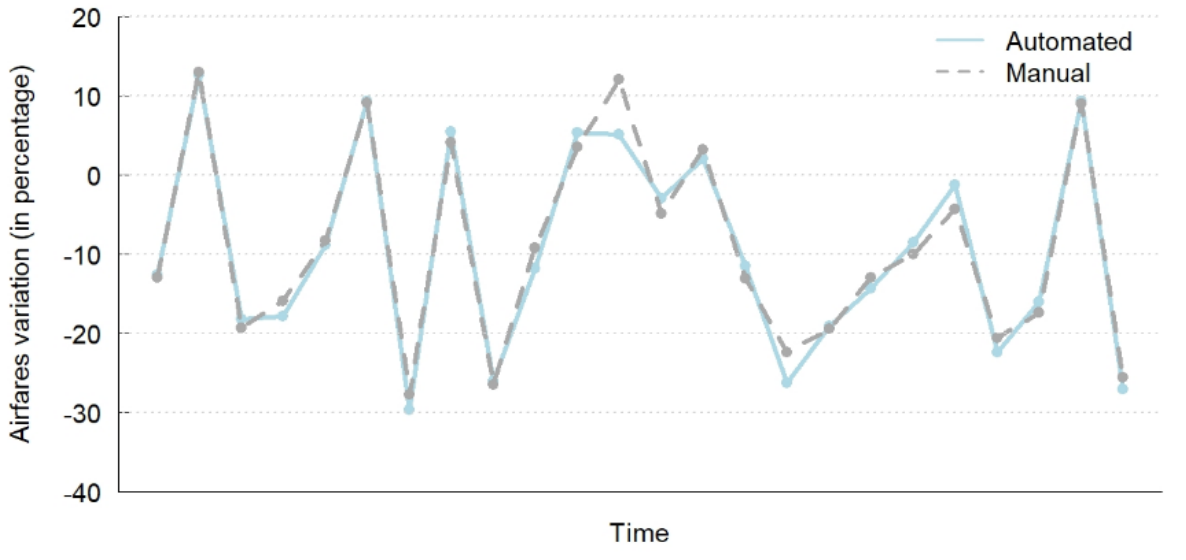
Suporte ao cartão de memória	Não
Tipo de tela	LCD Liquid Retina HD
Tamanho da tela	6,1"
Resolução da tela	1792x828 pixels a 326 ppp
Tecnologia	3G, 4G

```
<div class="information-values__product-page">
  <div class="price-template">
    <div class="price-template__from">de R$ 5.699,00</div>
    <div class="price-template__cash">
      <div class="price-template-price-block">
        por
        <span class="price-template__bold">R$</span>
        whitespace
        <span class="price-template__text">3.710,70</span>
        whitespace
        <span class="price-template__bold">à vista</span>
        <span class="price-template__discount-text--badges">(7% de desconto)</span>
      </div>
```

Example use of web scraping for airfares

Collection automation: Here focus is use of web scraping as a new technique to acquire the data.

Approach aims to emulate the procedures of a human collector inserting the travel parameters in the website, extracting the results and inserting the data in the CPI system.



Key issues for constructing a comparative series:

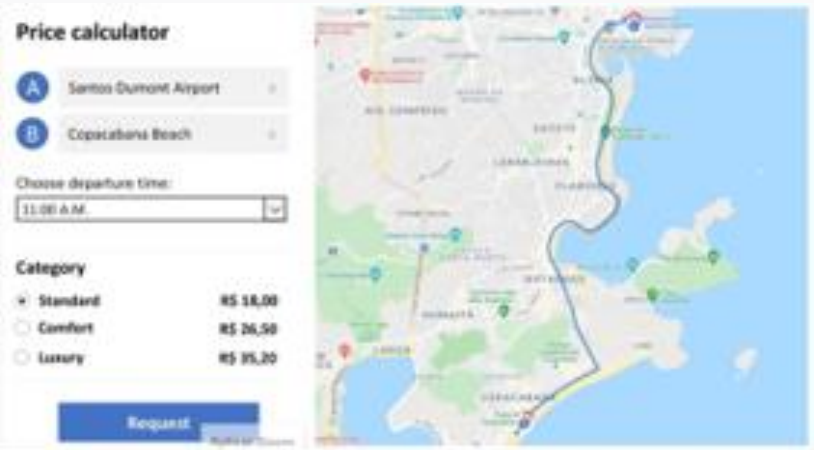
Compare the results with the traditional approach;

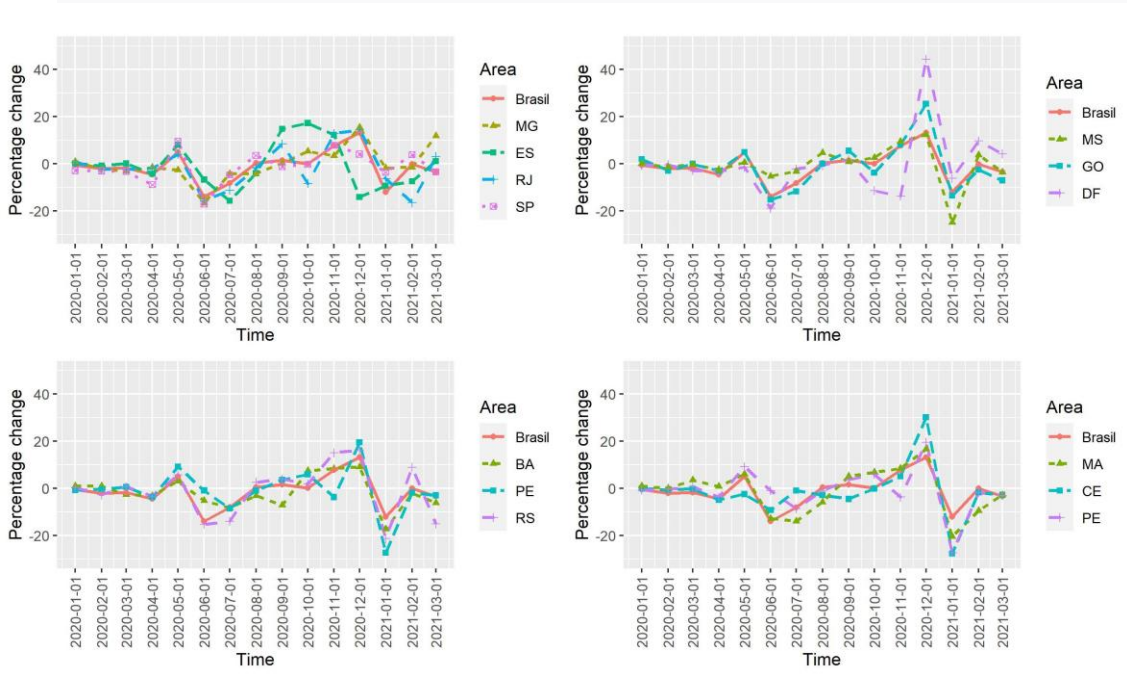
Evaluate the stability of the scrapers.

Technique used for the CPI and later also used for the ICP program.
Can save effort of collection of up to a hundred thousand prices a month.

Example use for ride sharing services

Area	IPCA		INPC	
	Taxi	Ride sharing Services	Taxi	Ride sharing Services
BR	**0,21**	**0,21**	**0,16**	**0,15**
AC	0,54	-	0,55	0,07
PA	0,43	-	0,32	-
MA	0,32	0,11	0,41	0,15
CE	0,18	0,15	0,15	0,16
PE	0,30	0,32	0,15	0,28
SE	0,58	0,11	0,53	0,17
BA	0,38	0,30	0,19	0,21
MG	0,24	0,19	0,17	0,16
ES	0,12	0,10	-	0,09
RJ	0,45	0,31	0,20	0,26
SP	0,16	0,20	0,11	0,12
RS	0,26	0,38	0,20	0,27
MS	0,09	0,23	-	0,28
GO	-	0,26	-	0,09
DF	-	0,25	0,11	0,16

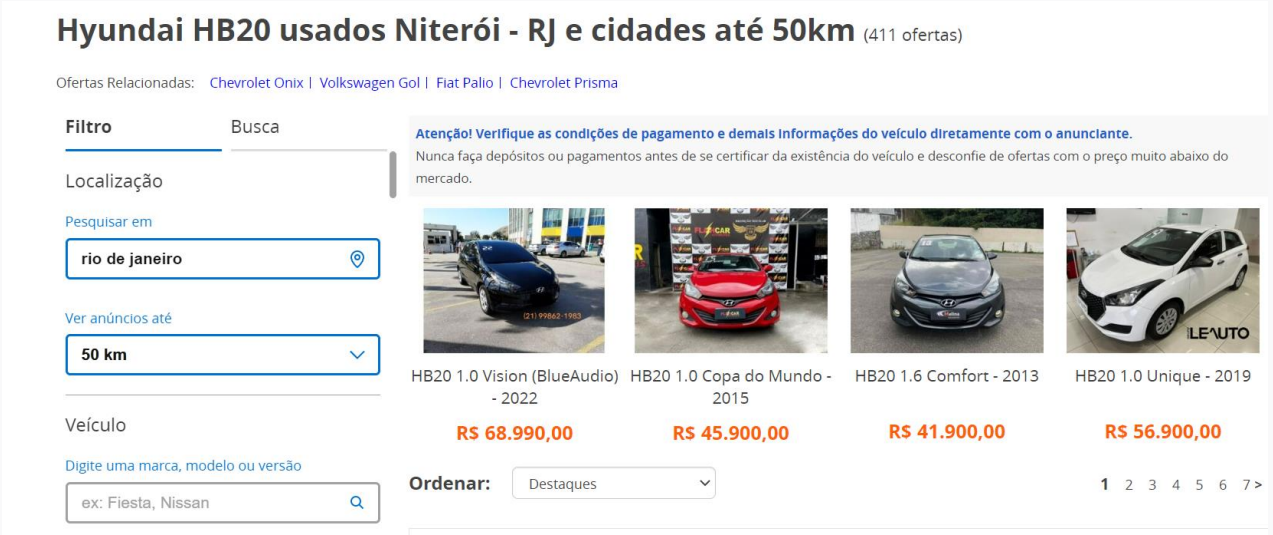


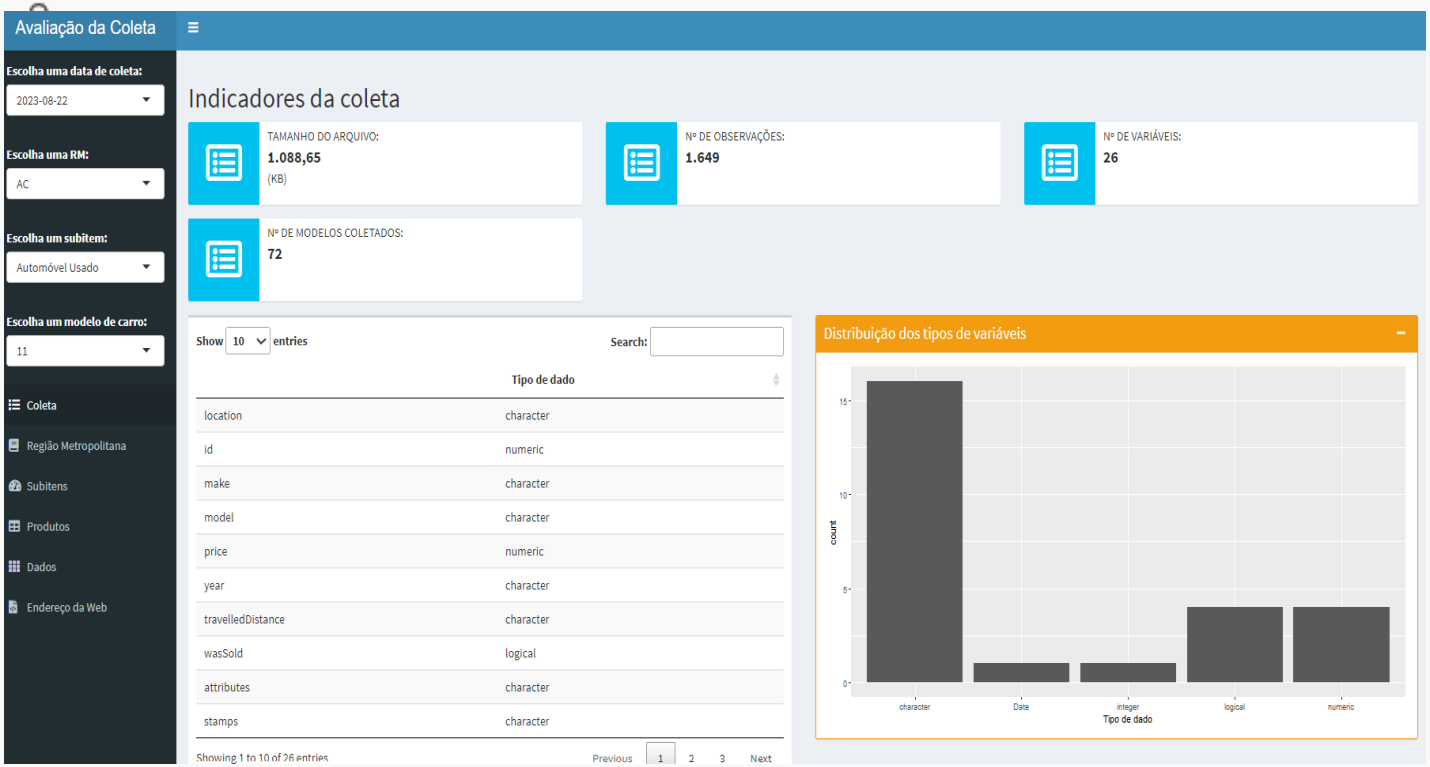


Webscraping + Dashboard: new and used cars

Marketplace platforms for car advertisements can offer a good tool to support the calculation of indices for new and used cars.

Offers according selected car models and geographical locations.





However, important methodological differences might appear from traditional approach: frequency of collection, difference of sources..

Implementation can take longer times.

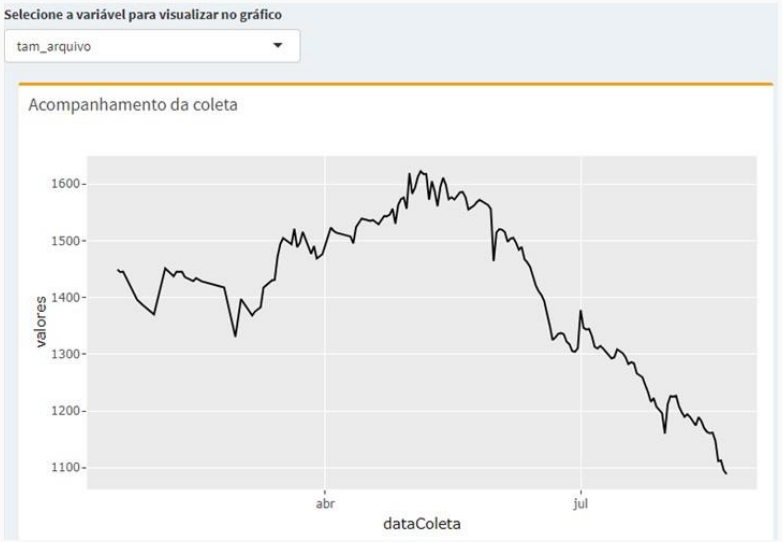
Idea here is while such studies are being made, try to use the data to assist the current processes.

Then the use of a dashboard to assist the control of the data collected and provide further information for the analysts.

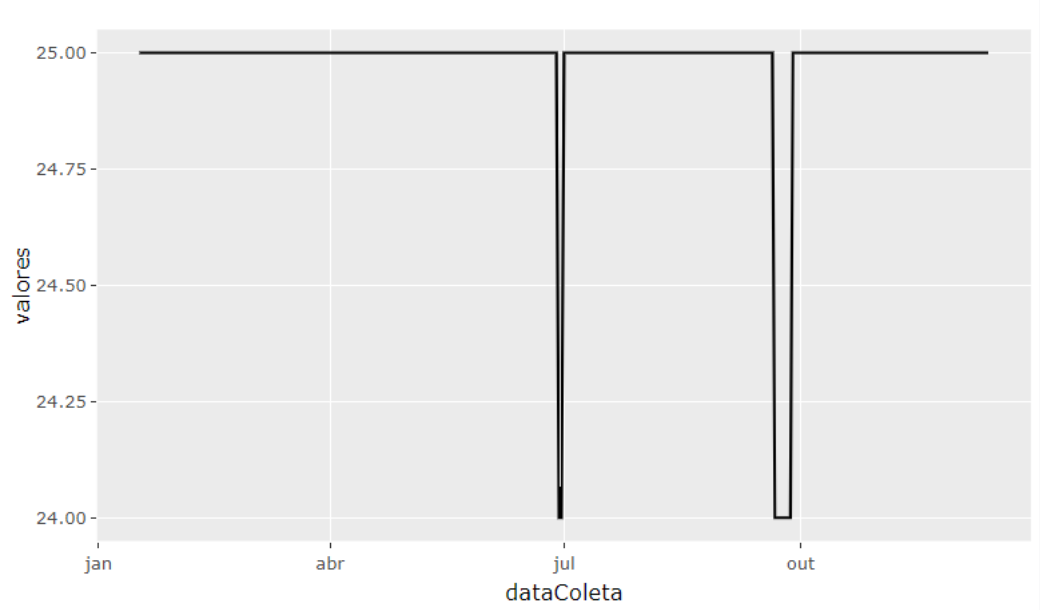
Webscraping + Dashboard: new and used cars

- Metrics useful for quality control of the data collected

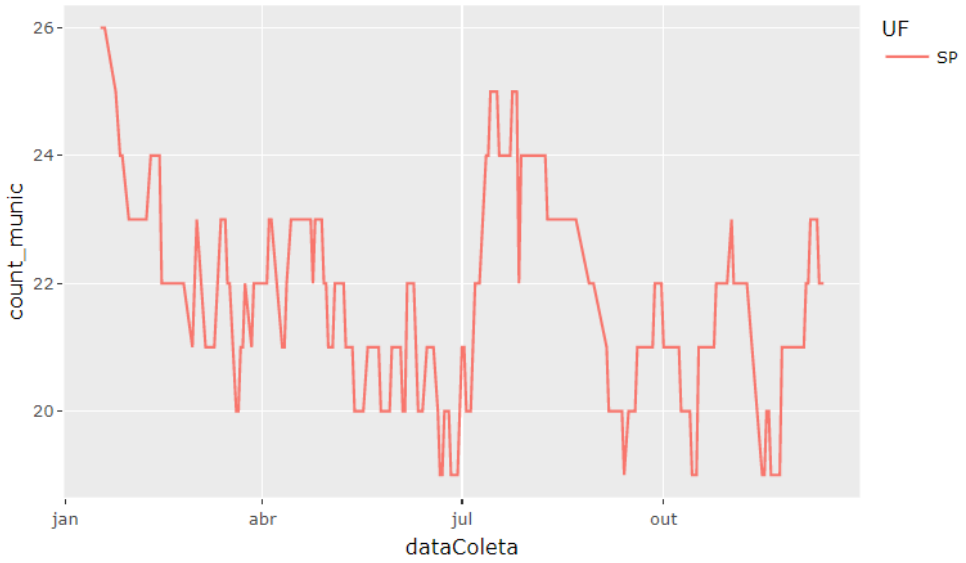
File size variation along time



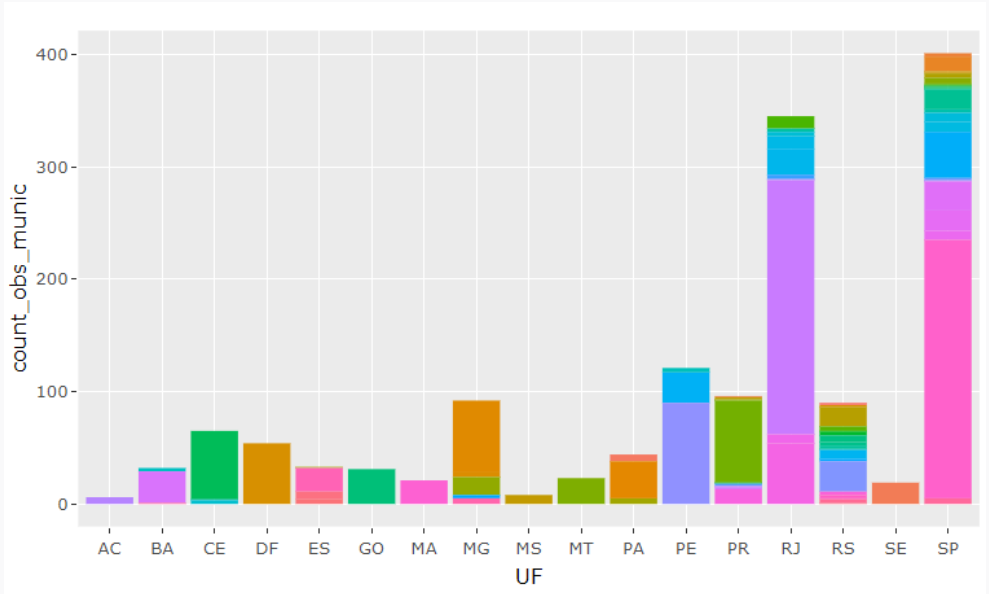
- Number of different areas collected over time



Number of municipalities covered for an area along time.



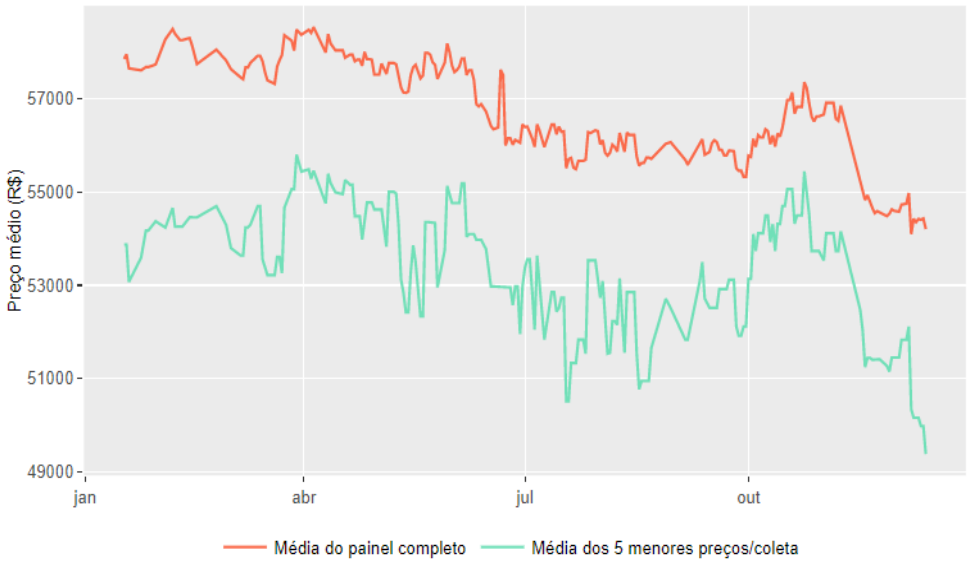
Geographical distribution of observations according each state and municipality for a given day.



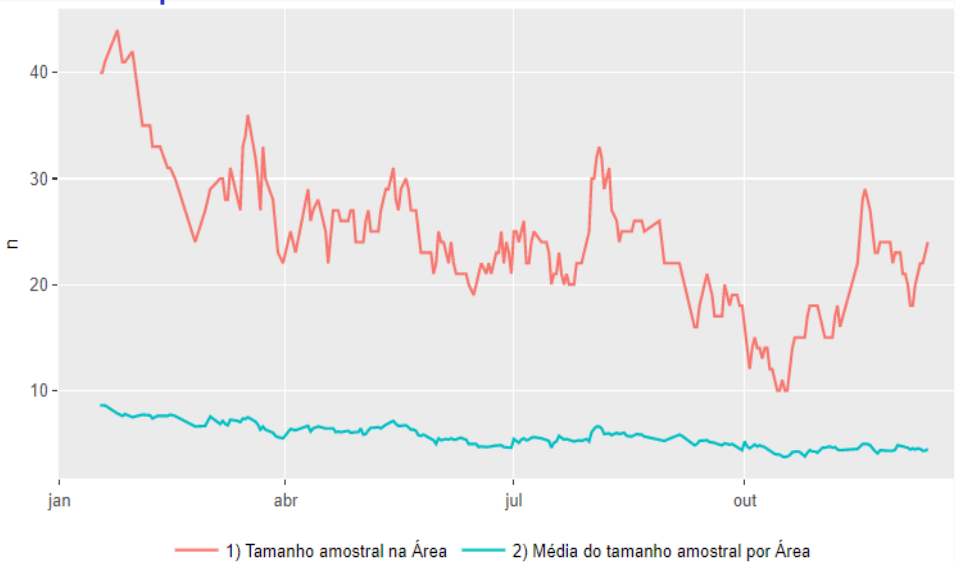
Webscraping + Dashboard: new and used cars

- Results for a given new car model in São Paulo

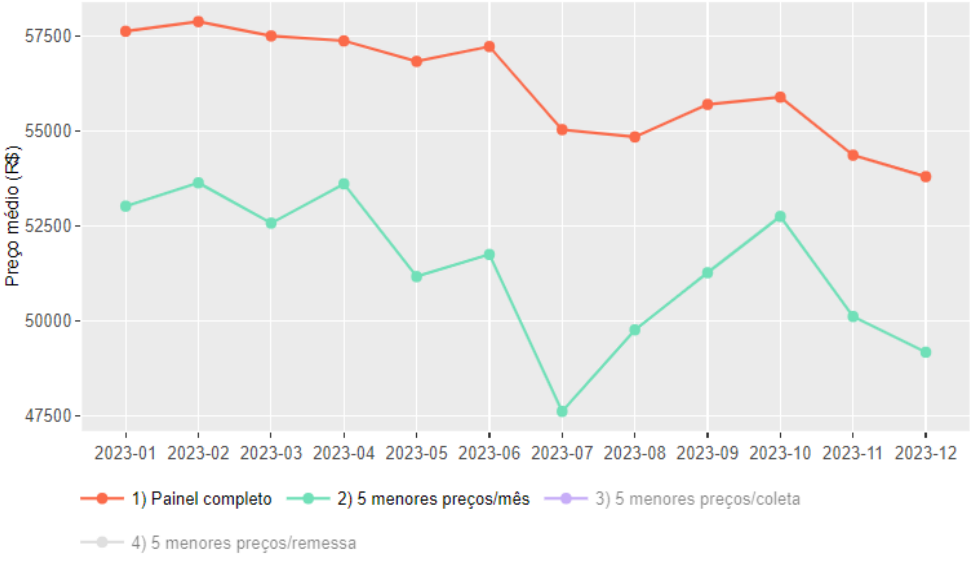
Daily average prices: full panel vs 5 lowest prices



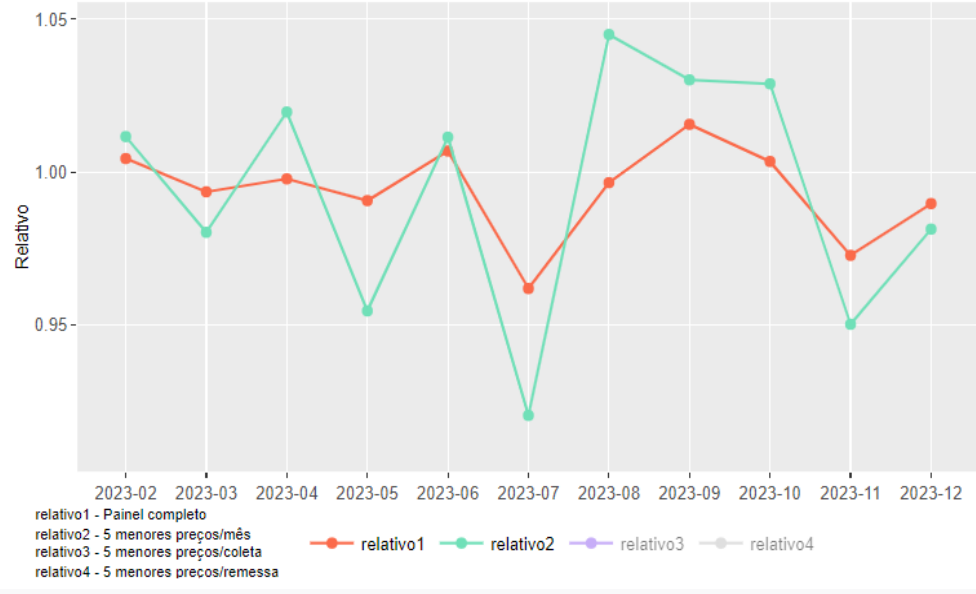
Sample Evolution in the area x average sample across areas



Monthly average prices: full panel vs 5 lowest prices

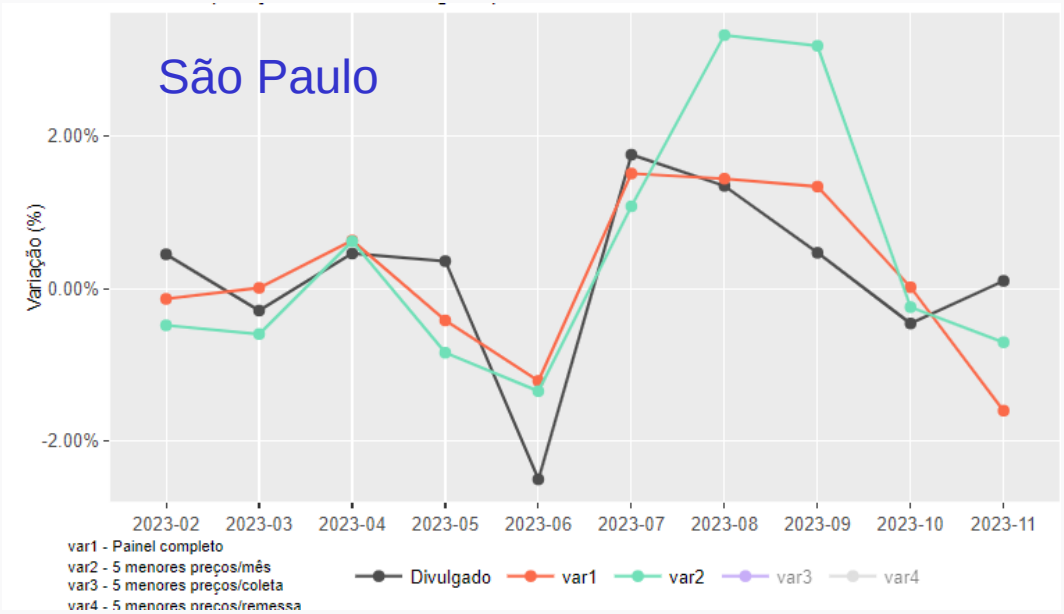


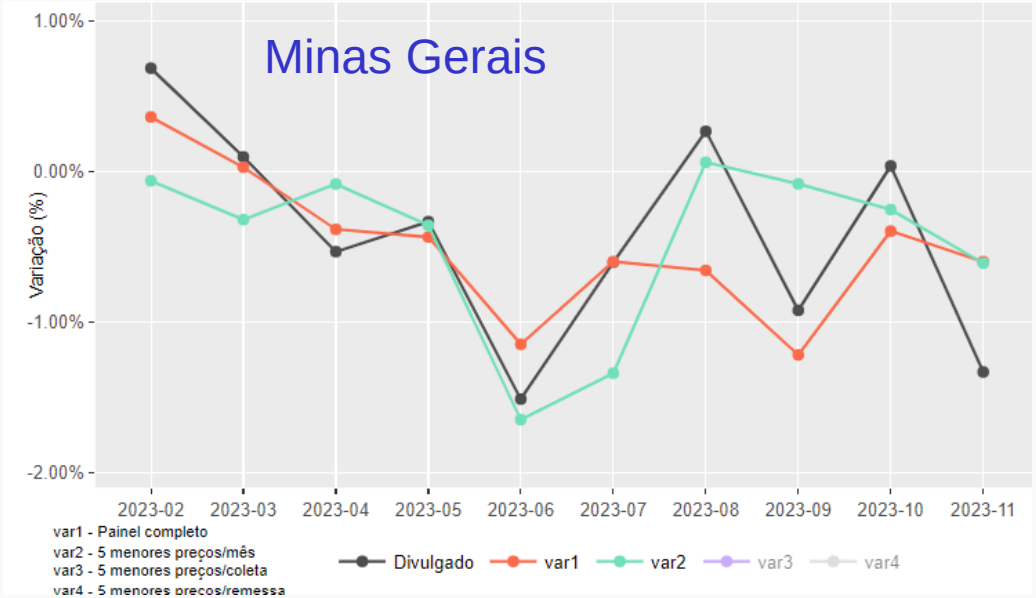
Monthly price relatives: Full panel vs 5 lowest prices



Webscraping + Dashboard: new and used cars

Comparing series for new cars: black (published), red (full month web), green (5 lowest web)





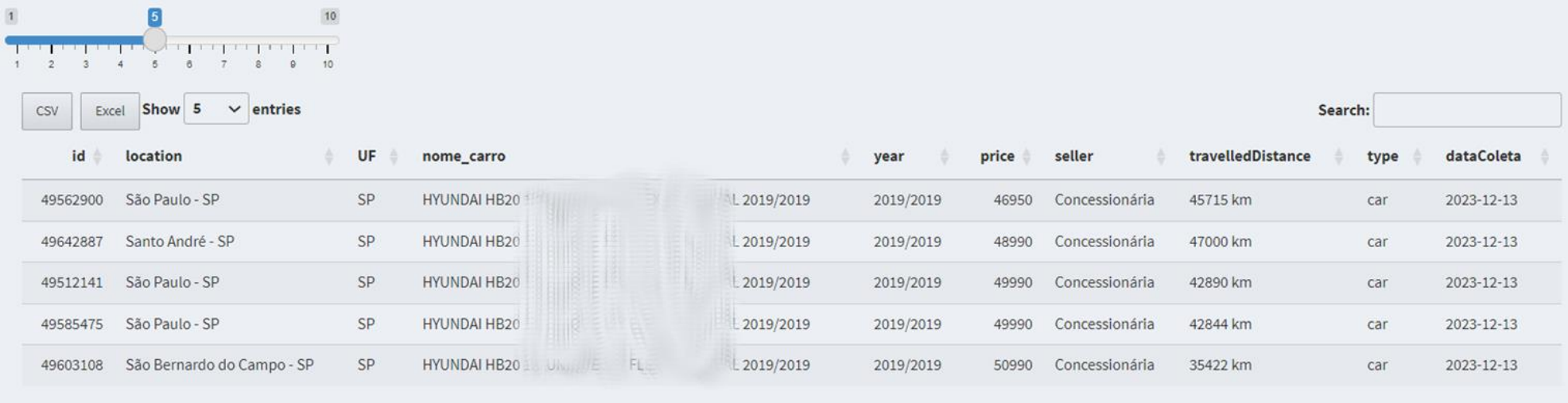
Published and web indices agree reasonably but some more prominent differences can be observed at different periods, or for different areas or methods.

Understanding if the difference is due the sources coverage (new cars prices are collected in brick and mortar stores), methods etc is an important point before putting this into production and require some time to evaluate as a comparative time series need to be constructed.

Meanwhile, the existing results can support the current analysis processes and also the collection for used cars which already make use of web advertisements.

Webscraping + Dashboard: new and used cars

- Microdata can be inspected and exported for support of the manual collection process if necessary.



id	location	UF	nome_carro	year	price	seller	travelledDistance	type	dataColeta
49562900	São Paulo - SP	SP	HYUNDAI HB20	2019/2019	46950	Concessionária	45715 km	car	2023-12-13
49642887	Santo André - SP	SP	HYUNDAI HB20	2019/2019	48990	Concessionária	47000 km	car	2023-12-13
49512141	São Paulo - SP	SP	HYUNDAI HB20	2019/2019	49990	Concessionária	42890 km	car	2023-12-13
49585475	São Paulo - SP	SP	HYUNDAI HB20	2019/2019	49990	Concessionária	42844 km	car	2023-12-13
49603108	São Bernardo do Campo - SP	SP	HYUNDAI HB20	2019/2019	50990	Concessionária	35422 km	car	2023-12-13

Expansion not so straightforward

- We are also studying the use of web scraping for other sectors such as hotels, car rents, electronics and household appliances. Other candidates could include holiday packages, books, bus tickets, rents and condos.

- Different issues may rise for diferent sectors and expansion may not be so straighforward.

- Important points that should be considered for expansion:

i) Stability and maintenance of the scrapers. Challenge for scalability and implementation in production. Fallout plan is mandatory.

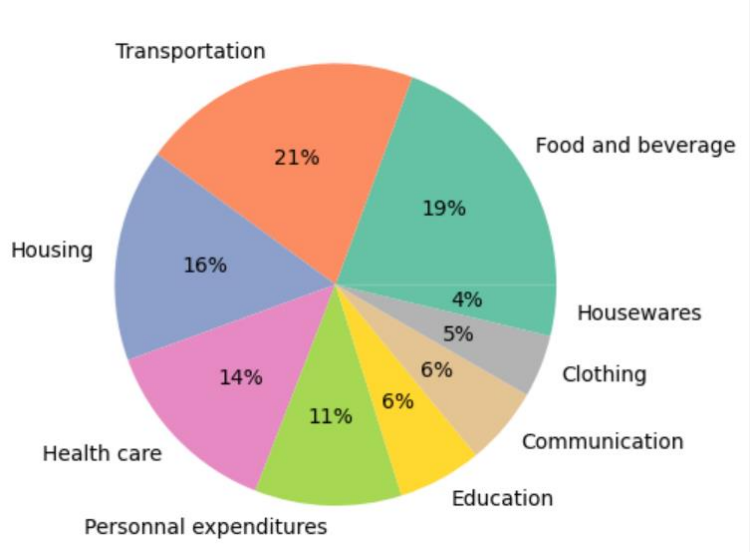
ii) Geographical breakdown might not be possible for sectors such as clothing, food and beverages and electronics. (Via the traditional approach a big retail chain might have several stores that are visited, via online only one price per product might be available for the whole chain).

iii) To make further use of the data a bulk approach might need to be implemented. However, a bulk approch usually requires changes in the weighting structure, and is more demanding in IT resources. Lack of weights for products can also lead to bias.

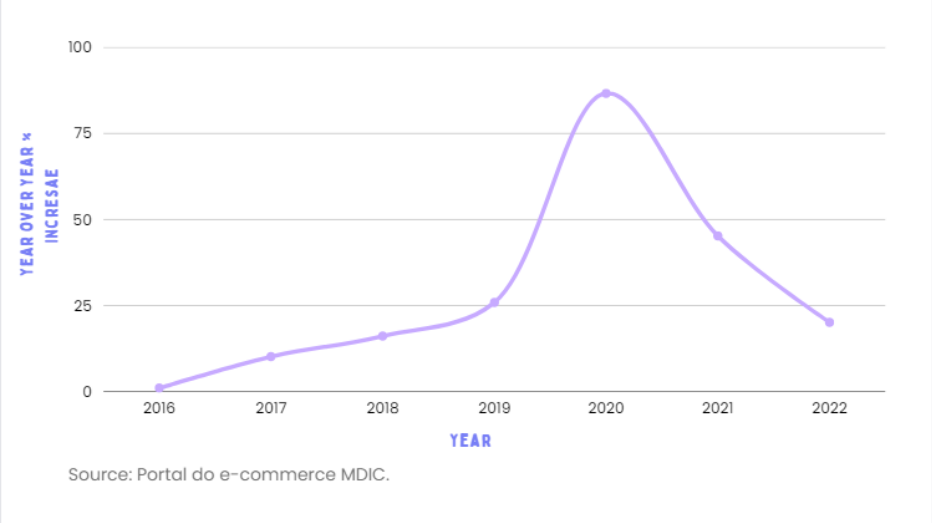
iv) Difference in representativity of ecommerce and actual purchase practices of consumers. Need to check if online indices series are representative of the series derived via traditional methods.

Comparison of market coverage CPI basket and ecommerce

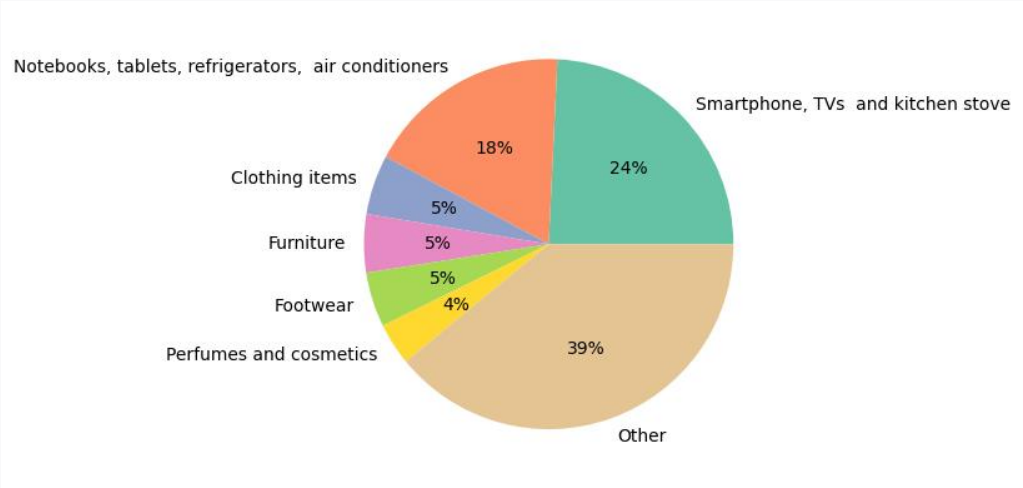
CPI (IPCA) weights (jan 2020): Food, Transportation, housing and health cover almost 70% of the basket.



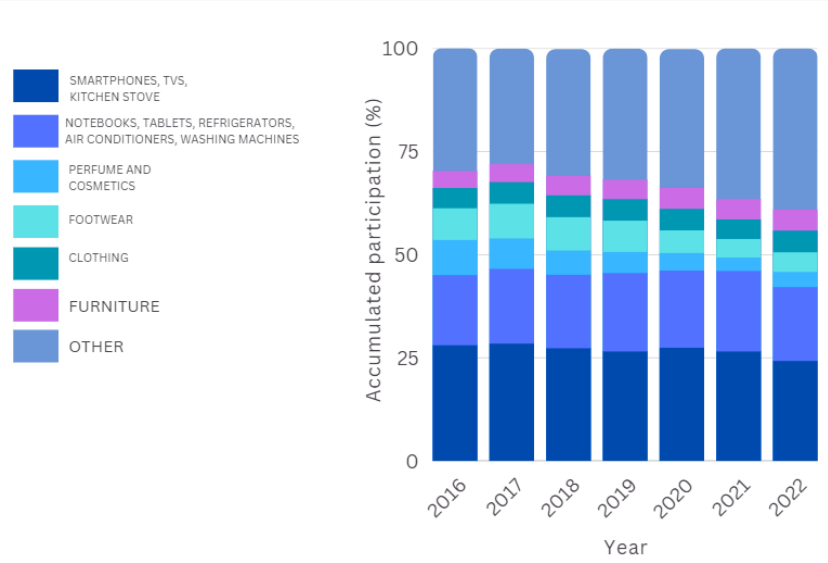
Anual increase of e-commerce sales for products from 2016 – 2022.



Percentage of e-commerce coverage by sector in 2022 (products only).



Distribution of e-commerce sales by categories 2016 – 2022.



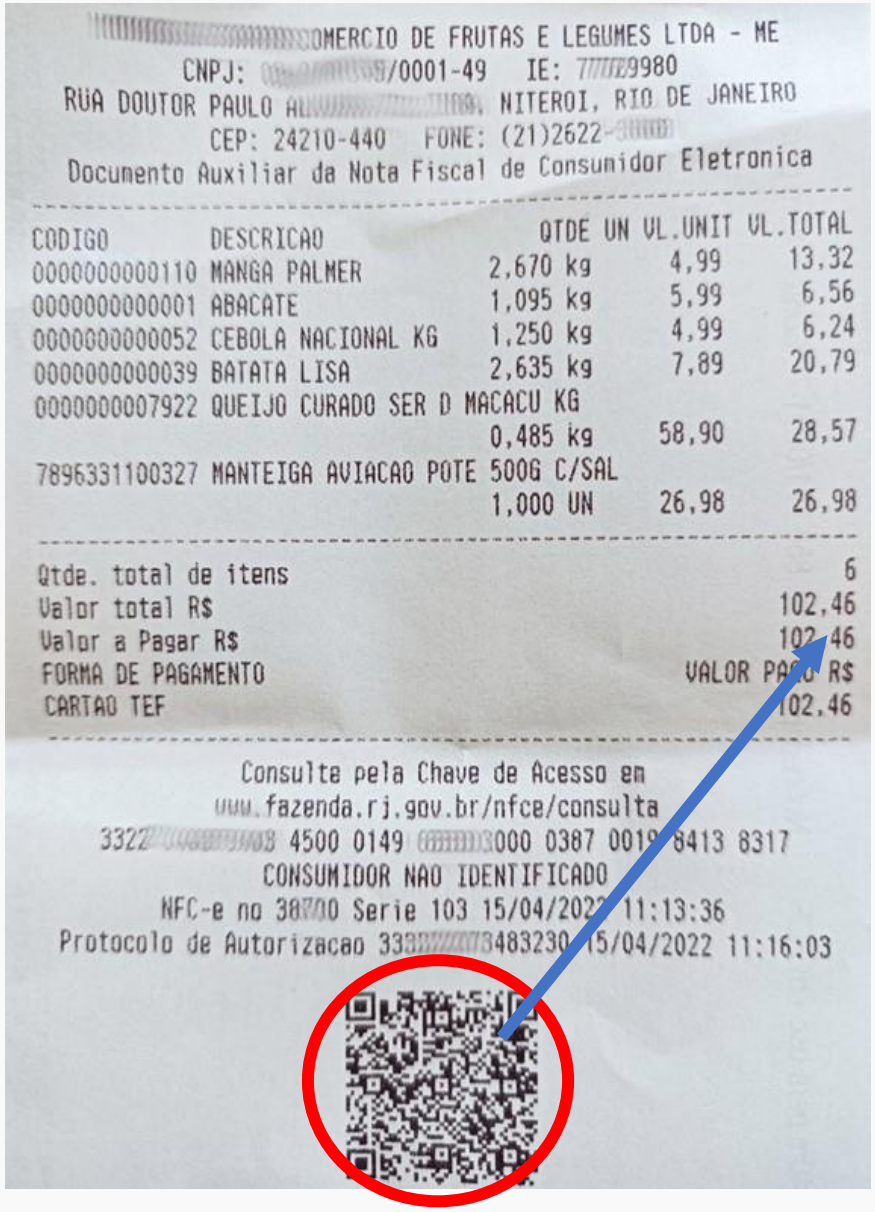
Scanner data

Scanner data

Data that rose with the advent of barcodes for purposes of stock control and analysis by retailers.

It allows a detailed control of products sold and stocks: what was sold, when, amount, turnover etc.

Peculiarity of Brazil: e-receipts



OMERCIO DE FRUTAS E LEGUMES LTDA - ME
CNPJ: [illegible]/0001-49 IE: [illegible]9980
RUA DOUTOR PAULO AL[illegible], NITEROI, RIO DE JANEIRO
CEP: 24210-440 FONE: (21)2622-[illegible]
Documento Auxiliar da Nota Fiscal de Consumidor Eletronica

CODIGO	DESCRICAO	QTDE UN	VL.UNIT	VL.TOTAL
0000000000110	MANGA PALMER	2,670 kg	4,99	13,32
0000000000001	ABACATE	1,095 kg	5,99	6,56
0000000000052	CEBOLA NACIONAL KG	1,250 kg	4,99	6,24
0000000000039	BATATA LISA	2,635 kg	7,89	20,79
0000000007922	QUEIJO CURADO SER D MACACU KG	0,485 kg	58,90	28,57
7896331100327	MANTEIGA AVIACAO POTE 500G C/SAL	1,000 UN	26,98	26,98

Qtde. total de itens 6
Valor total R$ 102,46
Valor a Pagar R$ 102,46
FORMA DE PAGAMENTO VALOR PAGO R$
CARTAO TEF 102,46

Consulte pela Chave de Acesso em
www.fazenda.rj.gov.br/nfce/consulta
3322[illegible]08 4500 0149[illegible]000 0387 0019[illegible]8413 8317
CONSUMIDOR NAO IDENTIFICADO
NFC-e no 38700 Serie 103 15/04/202[illegible]11:13:36
Protocolo de Autorizacao 333[illegible]483230 15/04/2022 11:16:03

[illegible] Comercio de Frutas e Legumes Ltda - Me
CNPJ: [illegible]0001-49
RUA DOUTOR PAULO [illegible] NITEROI , RJ

Filtrar itens...

MANGA PALMER (Código: 100110) Qtde.:2,67 UN: KG Vl. Unit.: 4,99		Vl. Total 13,32
ABACATE (Código: 100001) Qtde.:1,095 UN: KG Vl. Unit.: 5,99		Vl. Total 6,56
CEBOLA NACIONAL KG (Código: 100052) Qtde.:1,25 UN: KG Vl. Unit.: 4,99		Vl. Total 6,24
BATATA LISA (Código: 100039) Qtde.:2,635 UN: KG Vl. Unit.: 7,89		Vl. Total 20,79
QUEIJO CURADO SER D MACACU KG (Código: 107922) Qtde.:0,485 UN: KG Vl. Unit.: 58,9		Vl. Total 28,57
MANTEIGA AVIACAO POTE 500G C SAL (Código: 11055) Qtde.:1 UN: UN Vl. Unit.: 26,98		Vl. Total 26,98
	Qtd. total de itens:	6
	Valor a pagar R$:	102,46

A typical visualization of this same data from the “web app” of the fiscal agencies.

Currently, over 40 billion e-receipts valid in the data base and more than 2.3 million “stores” observed in the last month.

Initial problem: access.

Smells like scanner data and tastes like it … although it is an administrative record.

Proof of concept: inspection of main variables available via web scraping scanned receipts.

Structure of some of the variables found.

Description	Qtd	Unit	Turnover	Code_In	NCM	GTIN	CNPJ (ret_Id)	date	time	purchase_key
maca gala	1.2	KG	11.98	100103	8081000	NA	0345000149	05/05/2022	19:58:06	3,32205E+43
ESC DENT BASIC MACIA	1.0	UN	4.45	715670	96032100	NA	8250054763	01/05/2022	16:22:15	3,32205E+43
DES REXONA ST 48 ME CL	1.0	UN	21.49	277436	33072090	NA	38250054763	01/05/2022	16:22:15	3,32205E+43
BISC AMAN PRINC 400G	2.0	UN	25.28	9878161	19053100	NA	3411243300	24/04/2022	20:28:35	3,32204E+43
SALAMINHO TP HAMB HA	1.0	UN	10.99	1064161	16010000	8421395037341	3411243300	24/04/2022	20:28:35	3,32204E+43
SACOLA VERDE 48X55	1.0	UN	0.01	7636387	39232190	NA	3411243300	24/04/2022	20:28:35	3,32204E+43
AZEITE OLIV TUNI EXT	3.0	UN	17.97	150118	15092000	736532812902	7473009498	22/04/2022	19:45:14	3,32204E+43
ACHOCOLATADO TODDY 4	1.0	UN	9.79	117660	18069000	7894321711263	87473009498	22/04/2022	19:45:14	3,32204E+43
LEITE ZERO LACTOSE N	1.0	UN	7.98	121856	4012010	7898215157410	87473009498	22/04/2022	19:45:14	3,32204E+43
PAO FRANCES JC kg	0.355	kg	6.35	152971	19059090	96881	87473009498	22/04/2022	19:45:14	3,32204E+43
BROINHA MILHO C CANE	0.355	kg	17.71	152925	19059090	NA	87473009498	22/04/2022	19:45:14	3,32204E+43
ARROZ PARBOLIZADO IN	1.0	UN	6.98	112216	10062020	7893500018483	87473009498	22/04/2022	19:45:14	3,32204E+43

Comparison with typical variables found in a scanner dataset.

Date	Store	EAN number	Turn-over	Vol-ume	Unit	Quantity per unit	Product number	Product description
1104	7894	2920080800007	3402,70	211	Gram	300	910076003	Sliced bacon 2x150 G.
1104	7895	2920080800007	2119,65	163	Gram	300	910076003	Sliced bacon 2x150 G.
1104	7896	2920080800007	1516,05	108	Gram	300	910076003	Sliced bacon 2x150 G.
1104	7897	2920080800007	1478,13	105	Gram	300	910076003	Sliced bacon 2x150 G.
1104	7214	2921056000005	302,50	14	Gram	200	911056001	Chicken Fillet
1104	7215	2921056000005	102,50	5	Gram	200	911056001	Chicken Fillet

Potential use for Houlsehold budget surveys

From the very detailed granular data different expenditure information can be obtained.

Example of expenditures obtained at different levels:

By GTIN:

Código_EAN_Comercial	Descrição	
7891000073018	IOG NESTLE NAT DESN 160	171.01
07891010038953	FITA DEN REACH 50M	170.39
SEM GTIN	MASCARA WELLA 500G	169.90
773602315864	MAC MASCARA FALSE LASHES MAXIMIZER PADRAO UN	169.00
07896035301211	GELEIA INBASA 220G	167.70
7896679229575	OVOS CAIPIRA BOM J.12UN	166.80

By store

CNPJ	
[illegible]6517000122	111.72
[illegible]8250054763	105.60
[illegible]4751003736	103.36
[illegible]8250004901	100.67
[illegible]4284012099	96.97
[illegible]6807000191	96.83
[illegible]9311000113	96.60
[illegible]7931000160	96.47
[illegible]2753000196	94.90
[illegible]7473003295	83.30

Combination of GTIN and store

CNPJ	Código_EAN_Comercial	
[illegible]08411243300	02002885010882	176.22
[illegible]5865222995	7896026300216	172.14
[illegible]4133000120	7891000073018	171.01
[illegible]8411243300	07891010038953	170.39
[illegible]77511002182	773602315864	169.00
[illegible]08411243300	07896035301211	167.70

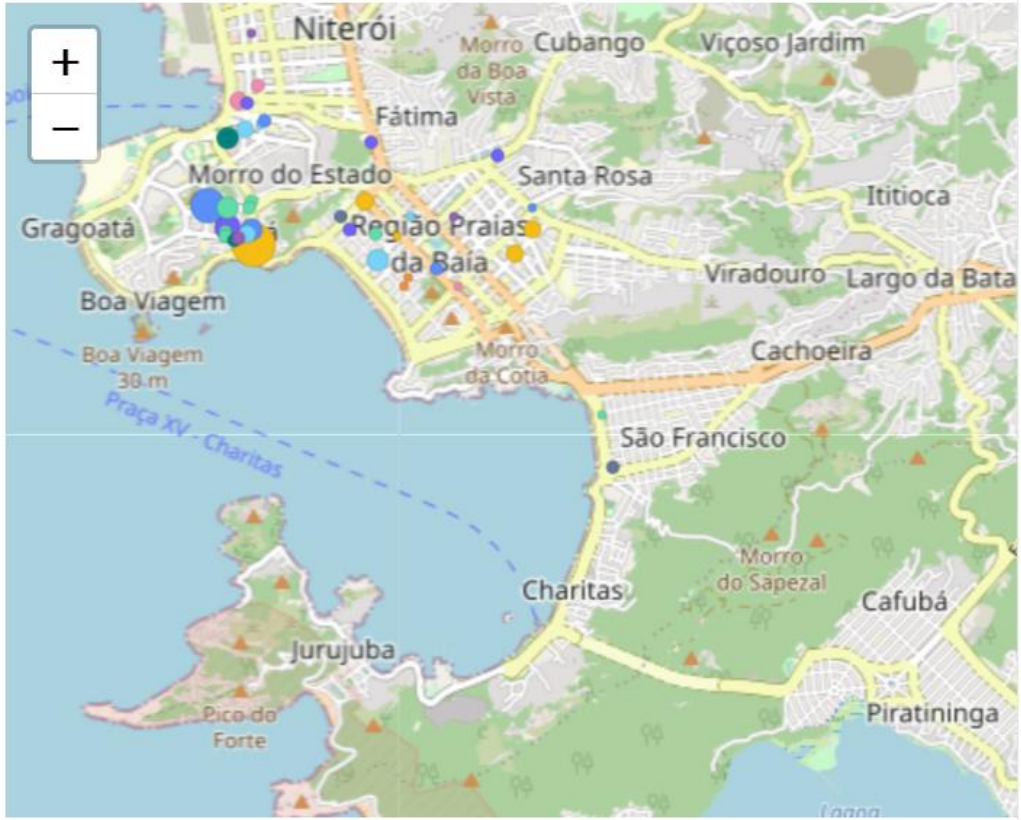
Use of this kind of household scanner data could be very useful for surveys like the household budget survey to derive detailed and precise data on consumer purchases.

Potential use for Houlsehold budget surveys

Retailers addresses also available.

Store name	Store id	Address	District	Zip code	Municipality	Phone	UF	Country
SUPERMERCADO [illegible]A F45	3,33813E+13	RUA MACHADO [illegible]	FLAMENGO	22220060	3304557-Rio de Janeiro		RJ	1058.0
SUPERMERCADO [illegible]A F42	3,33813E+13	RUA SAO [illegible]	BOTAFOGO	22260002	3304557-Rio de Janeiro		RJ	1058.0
SUPER MERCADO [illegible]S/A F37	3,33813E+13	RUA SOA JOAO [illegible]	BOTAFOGO	22270030	3304557-Rio de Janeiro		RJ	1058.0

Adding geographical coordinates, we can map consumption patterns geographical distribution.



Finding the needle in the haystack: classification issues

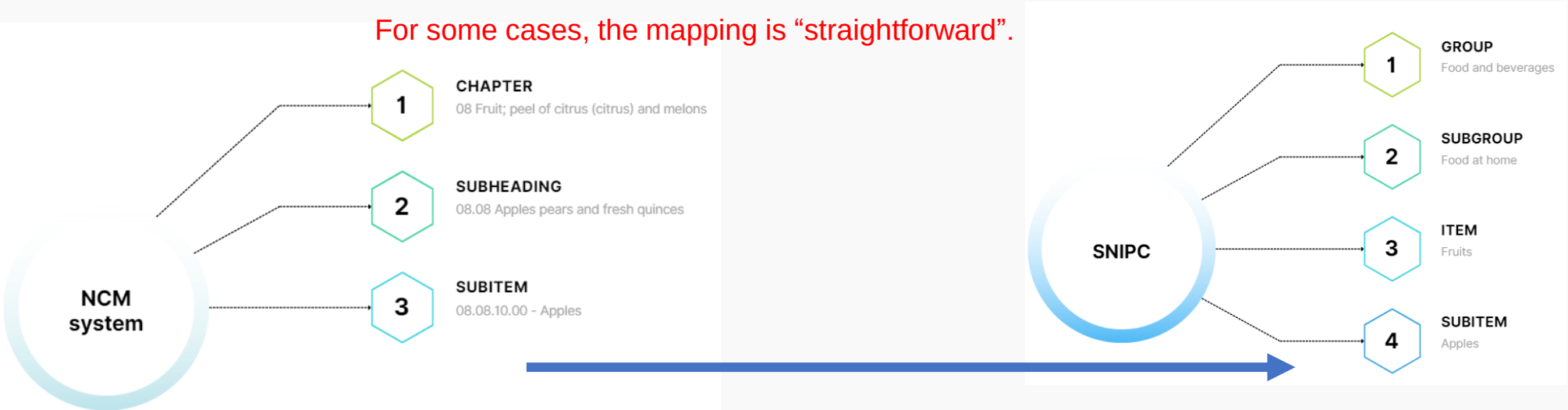
NCM = Mercosul Classification System, MCN in english.

Description	Qtd	Unit	Turnover	Code_In	NCM	GTIN	CNPJ (ret_Id)	date	time	purchase_key
maca gala	1.2	KG	11.98	100103	8081000	NA	5940345000149	05/05/2022	19:58:06	3,32205E+43
ESC DENT BASIC MACIA	1.0	UN	4.45	715679	96032100	NA	33438250054763	01/05/2022	16:22:15	3,32205E+43
DES REXONA ST 48 ME CL	1.0	UN	21.49	277486	33072090	NA	33438250054763	01/05/2022	16:22:15	3,32205E+43
BISC AMAN PRINC 400G	2.0	UN	25.28	9878161	19053100	NA	47508411243300	24/04/2022	20:28:35	3,32204E+43
SALAMINHO TP HAMB HA	1.0	UN	10.99	1054161	16010000	8421395037341	47508411243300	24/04/2022	20:28:35	3,32204E+43
SACOLA VERDE 48X55	1.0	UN	0.11	7636387	39232190	NA	47508411243300	24/04/2022	20:28:35	3,32204E+43
AZEITE OLIV TUNI EXT	3.0	UN	77.97	150118	15092000	736532812902	31487473009498	22/04/2022	19:45:14	3,32204E+43
ACHOCOLATADO TODDY 4	1.0	UN	9.79	117660	18069000	7894321711263	31487473009498	22/04/2022	19:45:14	3,32204E+43
LEITE ZERO LACTOSE N	1.0	UN	7.98	121856	4012010	7898215157410	31487473009498	22/04/2022	19:45:14	3,32204E+43
PAO FRANCES JC kg	0.355	kg	6.35	152971	19059090	96881	31487473009498	22/04/2022	19:45:14	3,32204E+43
BROINHA MILHO C CANE	0.355	kg	17.71	152925	19059090	NA	31487473009498	22/04/2022	19:45:14	3,32204E+43
ARROZ PARBOLIZADO IN	1.0	UN	6.98	112216	10062020	7893500018483	31487473009498	22/04/2022	19:45:14	3,32204E+43

NCM taxonomy:

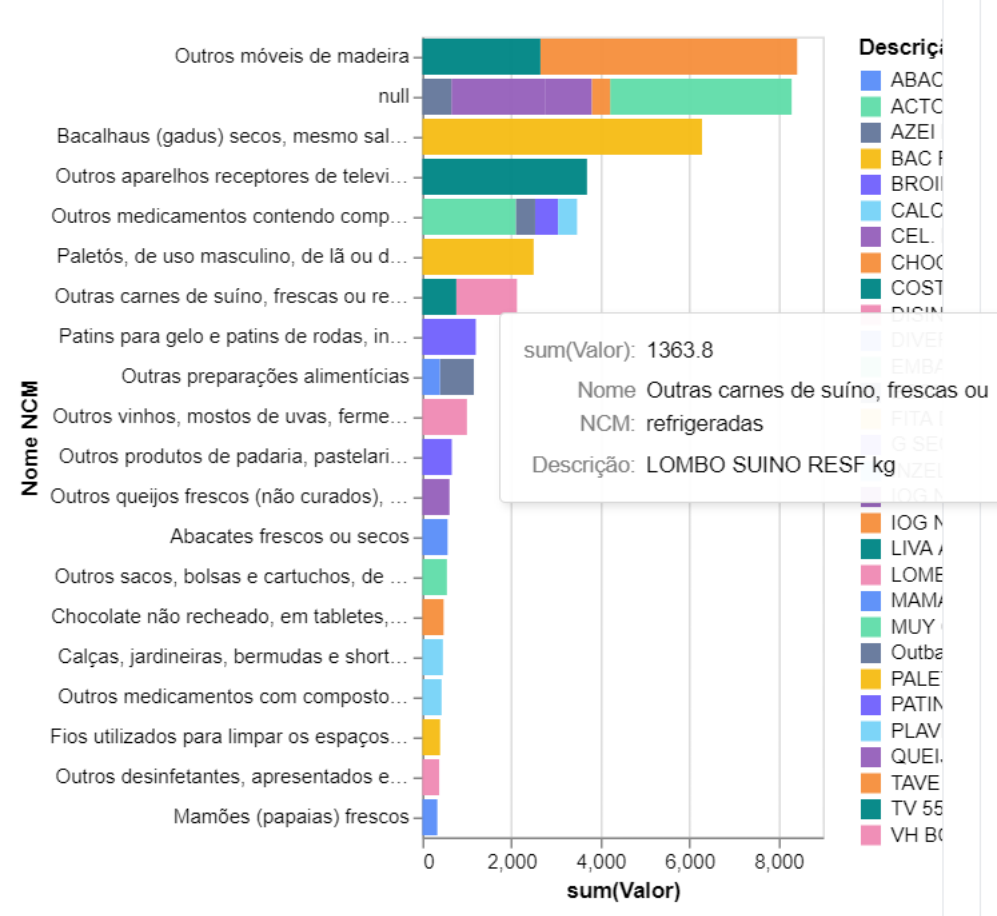
This is extremely valuable to reduce the complexity of the classification task.

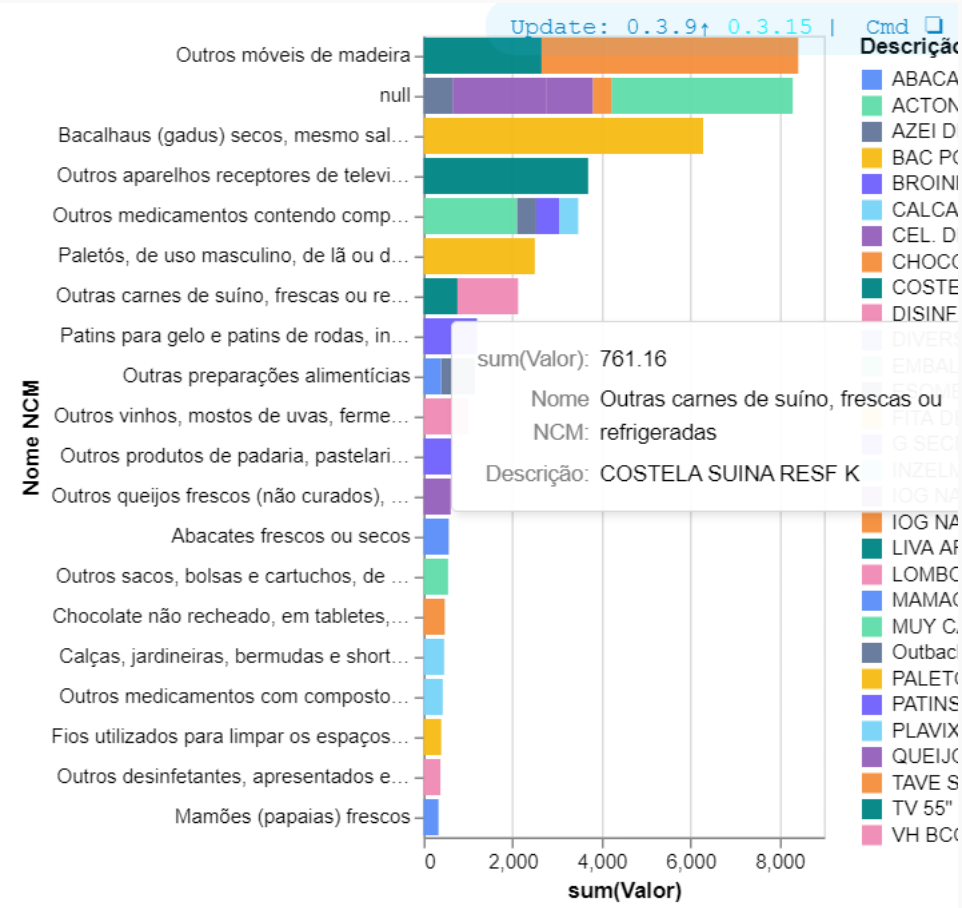
For some cases, the mapping is “straightforward”.



Example of use cases

Distribution of expenditures of products within a given NCM

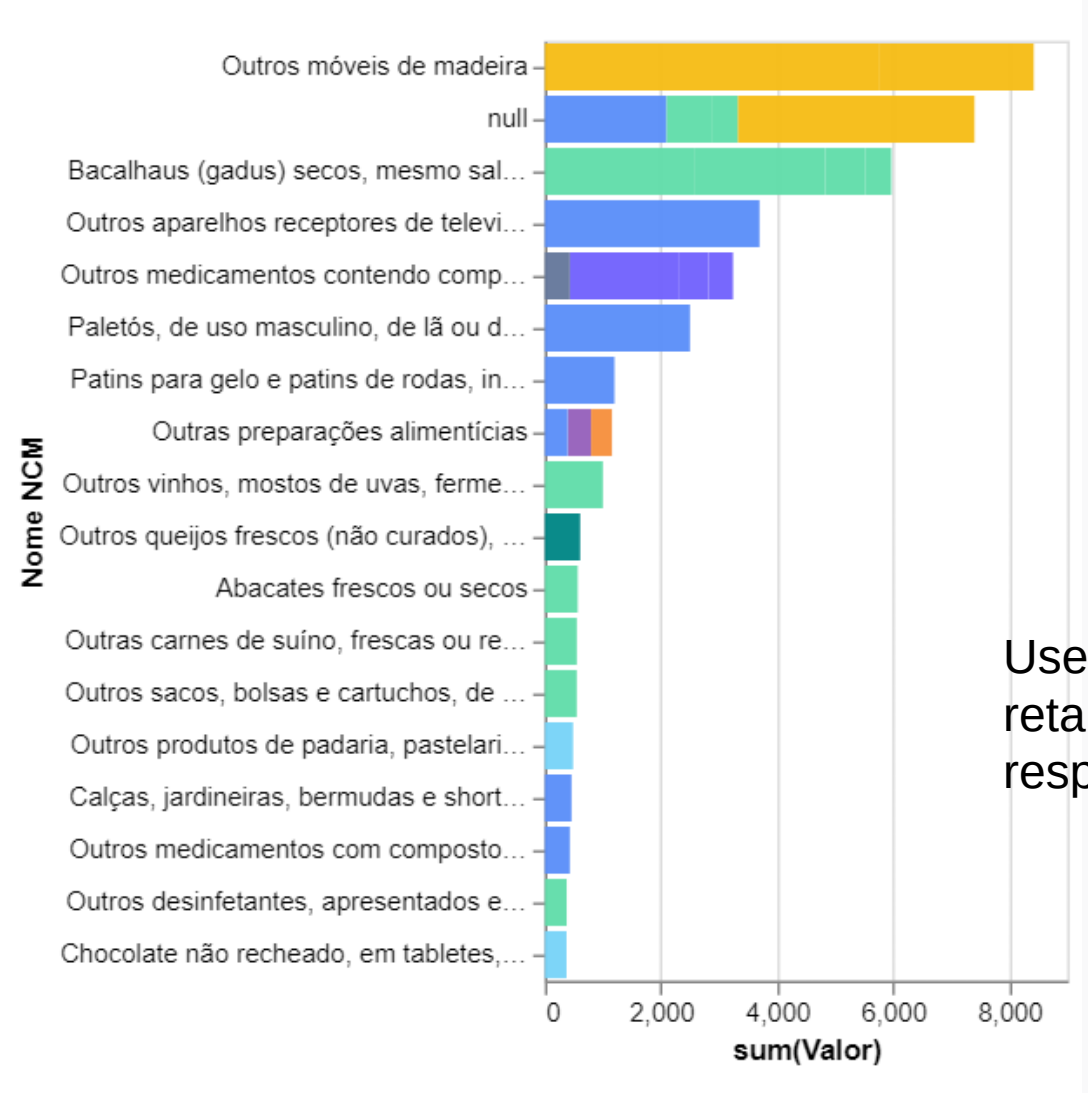




Useful to derive market shares of products.

Example of use cases

Distribution of expenditures according stores for a given NCM



Useful to derive sample of most relevant retailers for a given sector and its respective weight in the market.

Another case study: use for prices indices calculation

Preview of the variables extracted for gasoline from e-receipts for public purchases.

After some filters: 1328 different gas stations have at least one price in the panel from oct/2021-may/2023.

Date	NCM category	Product description	Seller name	Seller Code	Price	Turnover
01/02/2022	Gasoline except for aviation	Gasoline	America Comercio de combustível LTDA	28.549.481/0001-08	6.89	19.98
01/02/2022	Gasoline except for aviation	Gasoline	Posto Arvoredo	27.066.273/0002-76	6.77	54.97

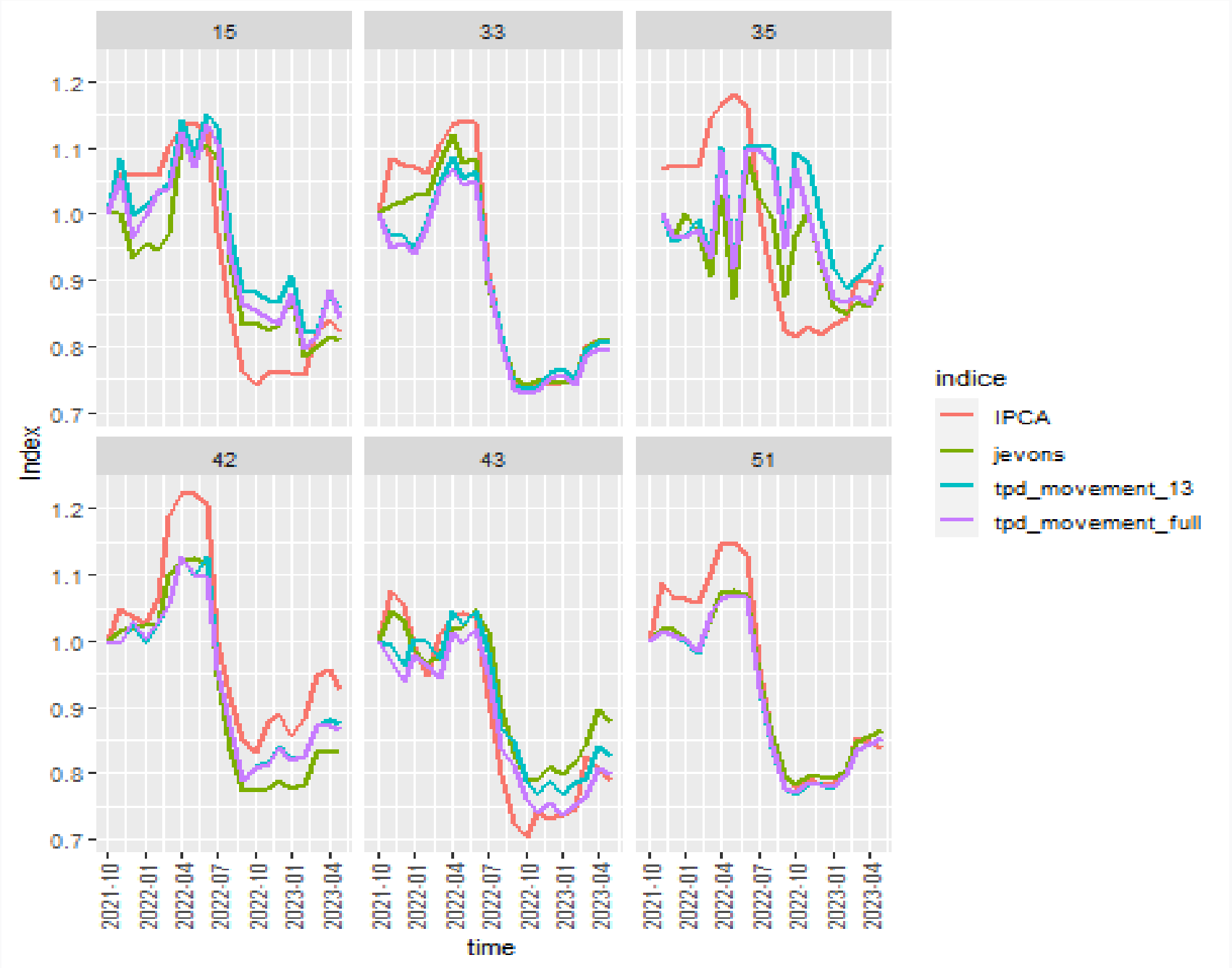
Summary stats	Price	Turnover
Count	12387.0	1.23e+04
Mean	6.24	9.65e+03
Std	0.93	1.26e+05
Min	4.70	6.05e+00
25%	5.45	1.81e+02
50%	6.05	2.91e+02
75%	6.99	8.75e+02
Max	8.58	9.34e+06

Prices range signals that this should be a good proxy for prices paid by families.

However quantities should not be applicable.

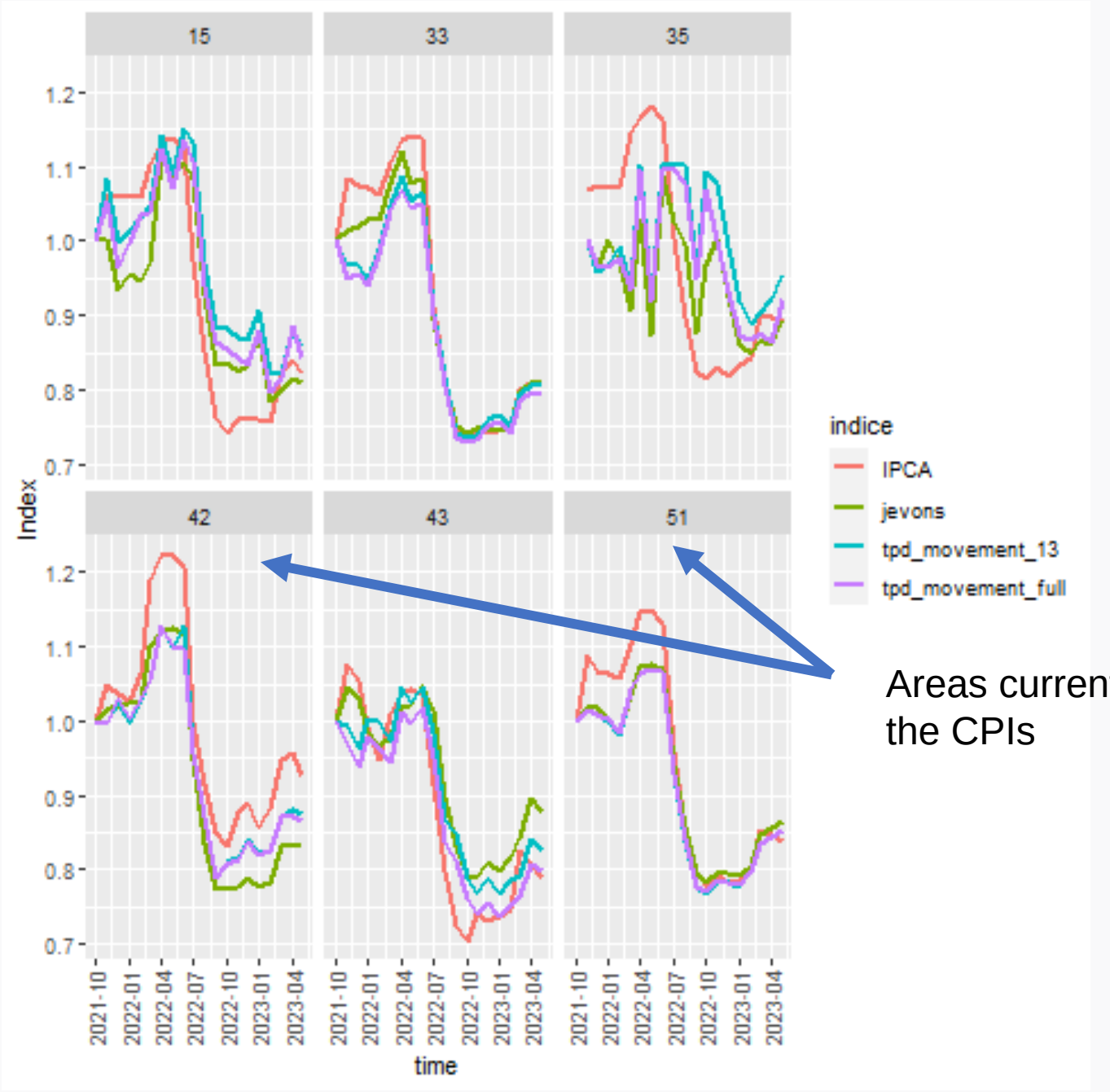
Another case study: use for prices indices calculation

Experimental índices for some areas:



Another case study: use for prices indices calculation

Experimental índices for some areas:



Areas currently not covered in the CPIs

Conclusions

We have inspected the potential and limitations for the use of different alternative data sources for the compilation of price statistics at IBGE. The case studies can reveal some of the potential and challenges associated with the use of such data sources.

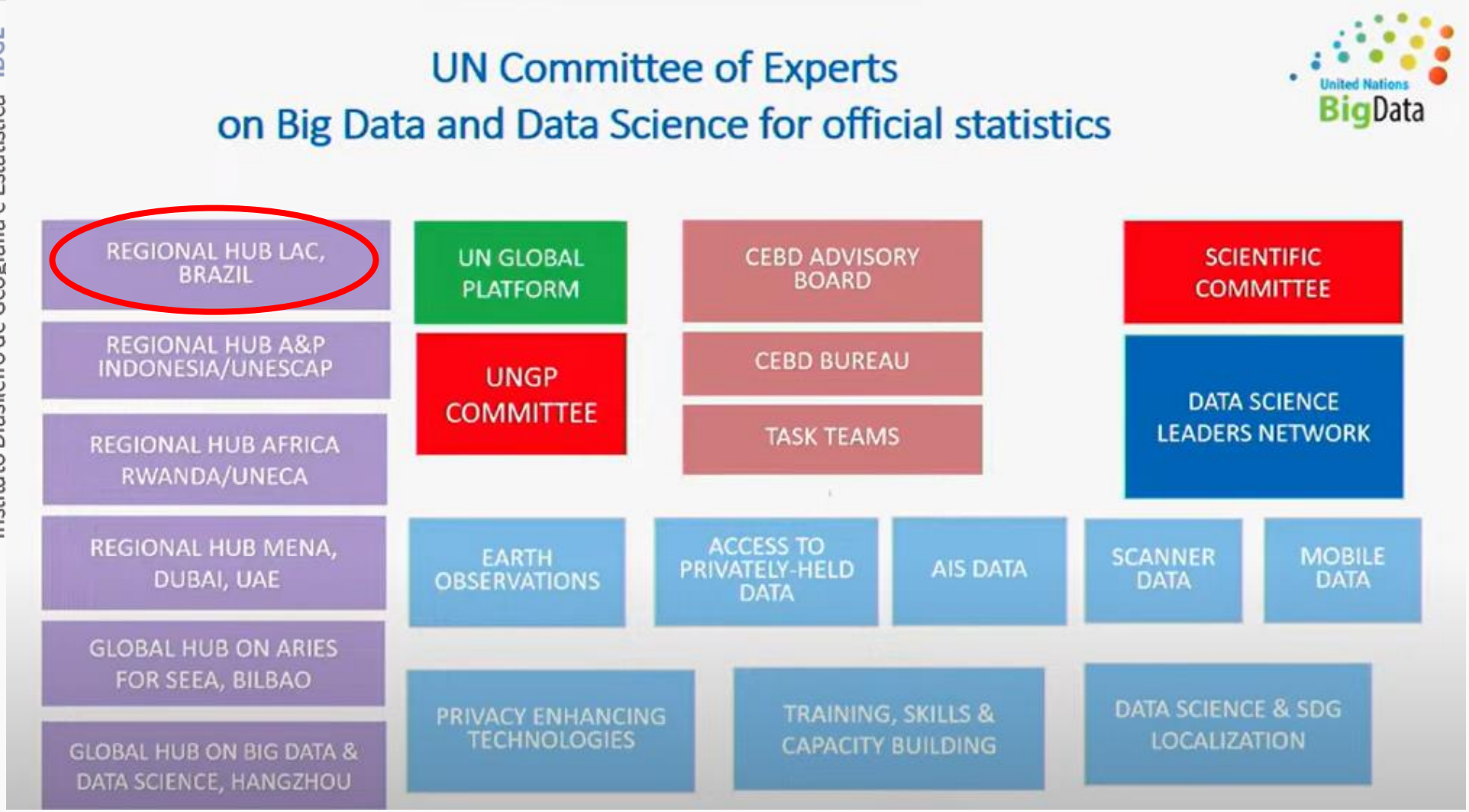
Illustration of use of web scraping can reveal that this is a versatile and powerful tool for different applications. However, caution should be taken for its implementation in production according the sector and application aimed. Requirements for data integration, possible differences in market coverage and sites changes that can block the data collection as well as IT infrastructure should be taken care.

Exploratory studies developed via use of “scanner-like” data contained in the fiscal e-receipts shows that data has potential for use as household scanner-like with potential use for HBS surveys.

Looking at the other side of the coin, stores-like scanner data contained in the data set can be used to derive prices índices with information on prices and quantities transacted.

Access to scanner data sets is still an issue, but exploration of these prelimiary data sets might help to develop a route for such use.

UN big data hub in Brazil



Working program

IBGE

UN Big Data Regional Hub in Brazil

WORKSTREAMS

Strengthening ties and promoting cooperation among producers of official statistics in the Region

Supporting sharing of experiences and knowledge, promoting increased contact and integration among regional producers and users, and increasing use of the knowledge generated.

Training and fostering the interest of young statisticians on the use of Big Data in Official Statistics

Offering online courses and webinars, with theoretical content and hands on activities, on methods, techniques, and tools for the use of Big Data in Official Statistics.

Supporting research on the use of Big Data and Data Science

Broadening the thematic scope of research on the use of Big Data in Official Statistics to gain experience in handling and processing this type of data; improving the accuracy and robustness of the results; developing protocols for incorporation of new data sources into the portfolio of Statistics Institutes in the Region.

Organizing and hosting seminars and conferences

Facilitate the exchange of information and contribute to the discussion on the use of new data sources and technologies, increasing involvement of Latin American and the Caribbean NSIs in developing new methods and algorithms for the global statistical system.

WORKSTREAM 1
Strengthening ties and promoting cooperation between producers of official statistics in the Region

1.1 Enlarge partnership in Latin America and the Caribbean via joint activities
1.2 Use of Big Data in Latin America and the Caribbean – 3rd Consultation

WORKSTREAM 2
Training and fostering the interest of young statisticians on the use of Big Data in Official Statistics

2.1 V Workshop on Use of Big Data for Official Statistics: environmental and climate change indicators
2.2 VI Workshop on Use of Big Data for Official Statistics: environmental and climate change indicators

WORKSTREAM 3
Supporting research on the use of Big Data and Data Science

3.1 Privacy-enhancing technologies (PET) in NSOs in Latin America and the Caribbean
3.2 Informal settlements using satellite imagery
3.3 Big data quality

WORKSTREAM 4
Organizing and hosting seminars and conferences

4.1 Webinar Series: Road to Festival de Datos in Punta del Este, Uruguay
4.2 Webinar Series: New Developments in the Use of Alternative Data
4.3 Special Topic Webinars: developments from the UNCEBD Task Teams
4.4 UN Satellite Datathon 2023 in Rio de Janeiro
4.5 8th International Conference on Big Data and Data Science for Official Statistics: side event

Contribution relating price statistics

First courses given by the Hub:

Instructors Lincoln Silva, Plinio Santos
Assistance of Vladimir Miranda.

Classes videos available online
Day 1 (scraper for static sites via rvest)
https://www.youtube.com/watch?v=J_uk_D8HInk

Day 2 (Use of rSelenium for airfares)
https://www.youtube.com/watch?v=VAZHlufvwEQ

Day 3 (Project discussion and importante remarks on the use of web data)
https://www.youtube.com/watch?v=L1Px9IbMj6w

Instructors: Vladimir Miranda, Lincoln Silva, Jacek Bialek.
Assistance: Plinio Santos.

Topics covered theoretical and practical aspects about the use of alternative data sources for calculation of prices indices.

Applications and exercises via use of two R softwares for data preparation and calculation of índices for alternative data sources.

Ackwoledgements

GPLACON:

Lincoln da Silva, Plinio Marcos dos Santos, Neimar Guimarães

COMEQ

Ingrid Oliveira

COINP

André Almeida, Gustavo Vitti

Thank you for your attention!

vladimir.miranda@ibge.gov.br