Big data in Brazil

Using mobile phone data to estimate SDG indicators

August 2021

Contents

1. Background.. 4
2. Data 5
2.1 Mobile phone data 5
2.1.1 Data access 5
2.1.2 Data characteristics 5
2.2 Reference data 7
2.2.1 Household survey data.. 7
2.2.2 Population estimates 7
2.2.3 Municipal boundaries.. 7
2.2.4 Digital elevation model 8
3. Data quality assurance 8
3.1 Cell tower data 8
3.1.1 Missing values 8
3.1.2 Geographical distribution of cell towers 8
3.1.3 Cell tower occupancy 9
3.2 CDR/IPDR data.. 9
3.2.1 Number of records per day 9
3.2.2 Number of unique subscribers per day.. 10
3.2.3 Subscriber presence in data 10
3.2.4 Diurnal distribution of records.. 12
3.2.5 Weekly distribution of records.. 13
3.2.6 Average number of records per day per subscriber 13
3.2.7 Identification of time zone 14
4. Home anchoring.. 14
5. Calculation of indicators.. 15
6. Results and discussion... 15
6.1 Indicator 17.8.1 15
6.2 Indicator 9.c.1... 22

Figures

FIGURE 1: AREA OF STUDY 6
FIGURE 2: CELL TOWER LOCATION MAP 8
FIGURE 3: NUMBER OF CELL TOWERS PER DAY 9
FIGURE 4: NUMBER OF RECORDS PER DAY 10
FIGURE 5: NUMBER OF SUBSCRIBERS PER DAY..10

FIGURE 6: DISTRIBUTION OF THE NUMBER OF SUBSCRIBERS BY THE TOTAL NUMBER OF DAYS OF MOBILE PHONE VOICE USAGE 11
FIGURE 7: DISTRIBUTION OF THE NUMBER OF SUBSCRIBERS BY THE TOTAL NUMBER OF DAYS OF MOBILE PHONE DATA USAGE 12
FIGURE 8: DISTRIBUTION OF MOBILE PHONE VOICE USAGE BY HOUR 12
FIGURE 9: DISTRIBUTION OF MOBILE PHONE DATA USAGE BY HOUR 13
FIGURE 10: NUMBER OF EVENTS PER WEEK 13
FIGURE 11: PROPORTION OF THE POPULATION USING THE INTERNET BY AREA 17
FIGURE 12: DISTRIBUTION OF THE TECHNOLOGY TO ACCESS THE INTERNET BY AREA 18
FIGURE 13: MAP OF MOBILE NETWORK COVERAGE BY GEOGRAPHICAL AREA 24
FIGURE 14: MAP OF MOBILE NETWORK COVERAGE BY GEOGRAPHICAL AREA AND TECHNOLOGY 25
FIGURE 15: COMPARISON OF MOBILE PHONE COVERAGE BETWEEN ANATEL AND THE MOBILE PHONE STUDY, BY TECHNOLOGY 26

Tables

TABLE 1: ATTRIBUTE OF THE CDR AND IPDR TABLES 5
TABLE 2: ATTRIBUTES OF THE CELL TOWER TABLE 5
TABLE 3: DESCRIPTIVE STATISTICS OF SUBSCRIBER USAGE DURING THE TIME PERIOD 11
TABLE 4: DESCRIPTIVE STATISTICS OF THE AVERAGE NUMBER OF RECORDS PER DAY PER SUBSCRIBER 14
TABLE 5: PROPORTION OF THE POPULATION WITH INTERNET ACCESS, BY TYPE OF TECHNOLOGY 15
TABLE 6: IBGE 2018 AND 2019 POPULATION ESTIMATES PER GEOGRAPHICAL UNIT 19
TABLE 7: INTERNET USE SURVEY ESTIMATES 20
TABLE 8: COMPARISON OF INTERNET USE ESTIMATES FROM MOBILE AND SURVEY DATA 21
TABLE 9: DIFFERENCES BETWEEN MOBILE AND SURVEY ESTIMATES BY GEOGRAPHICAL DISAGGREGATED AREAS 22
TABLE 10: MOBILE NETWORK COVERAGE BY GEOGRAPHICAL AREA 22

1. Background

Over the last decade, national statistical offices have faced challenges in terms of budget constraints and reducing the response burden in national statistical surveys. In addition, the offices are expected to produce geographically disaggregated statistics more frequently. Various ways to meet these demands include the use of new data sources, which also links up with the demand for technological innovation. These new sources include mobile phone data, which is the subject of this case study.

Measuring the information society is a good starting point for exploration of this new data source, given its obvious relationship with data itself and the fact that existing statistics in this field do not satisfactorily meet requirements in terms of temporal frequency and/or geographical granularity. In the case of data currently available on the Brazil SDG Panel, indicator 9.c.1[1] is produced by the Brazilian Telecommunication Agency (ANATEL) annually, with geographical coverage at the country level. Indicator 17.8.1[2] is produced by the Brazilian Institute of Geography and Statistics (IBGE) annually, with disaggregation at the state level.

The aim of this case study is threefold. First, to gain experience in handling and processing this type of data, which is unprecedented at IBGE. Second, to verify the accuracy and robustness of the results through comparisons with similar surveys using traditional data sources. Lastly, to develop a protocol for future incorporation into the Institute's statistical production pipeline.

Partnerships have played a vital role in the development of this study. The Regional Center for Studies on the Development of the Information Society – Cetic.br|NIC.br[3] - has been a longstanding partner of IBGE on research related to the information society and provided essential support during the implementation of this project.

This pilot is one of several undertaken by the International Telecommunication Union (ITU) across various countries to explore the use of mobile phone data in measuring the information society. The study was implemented by IBGE, which carried out data analysis, with support from CETIC. It was led by the ITU ICT Data and Analytics Division, with Positium serving as the project's consultant.

---

[1] Objetivo 9 - Indústria, Inovação e Infraestrutura: Indicador 9.c.1

[2] Objetivo 17 - Parcerias e meios de implementação: Indicador 17.8.1

[3] 15 Years: Cetic.br

2. Data

2.1 Mobile phone data

2.1.1 Data access

In 2019, during the preparations for the 2020 population census, several meetings were held with mobile network operators in Brazil to establish partnerships to assist in the operation. IBGE took this opportunity to request access to a set of mobile phone data to perform exploratory studies on mobility, one of IBGE's longstanding missions. Three of the four main operators in the country were asked to provide IBGE with a dataset restricted to a specific geographical region and a specific time interval for a proof of concept. Despite being well received by the companies, only one responded to the request and provided IBGE with the data. This company is the market leader in the region for which the data was provided, reaching 36 per cent (October 2020) of the market share in the state of Rio de Janeiro, Region I of the General Concession Plan.[4]

IBGE was granted access to the dataset upon signing terms of engagement, which included a commitment not to disclose or share the data with any third parties.

2.1.2 Data characteristics

The set of mobile phone data obtained covers the Rio de Janeiro Metropolitan Area and several neighboring municipalities, totaling 30 municipalities, as shown in the figure below. The Rio de Janeiro Metropolitan Area is the second most populous metropolitan area in the country and the region considered in the study is home to approximately 13.4 million inhabitants (IBGE, 2020).

The period covered by the dataset runs from 1 March to 30 April 2019.

The database is composed of three tables, referring to voice and text data (CDR), Internet data transfers (IPDR), and cell towers.

Table 1: Attribute of the CDR and IPDR tables

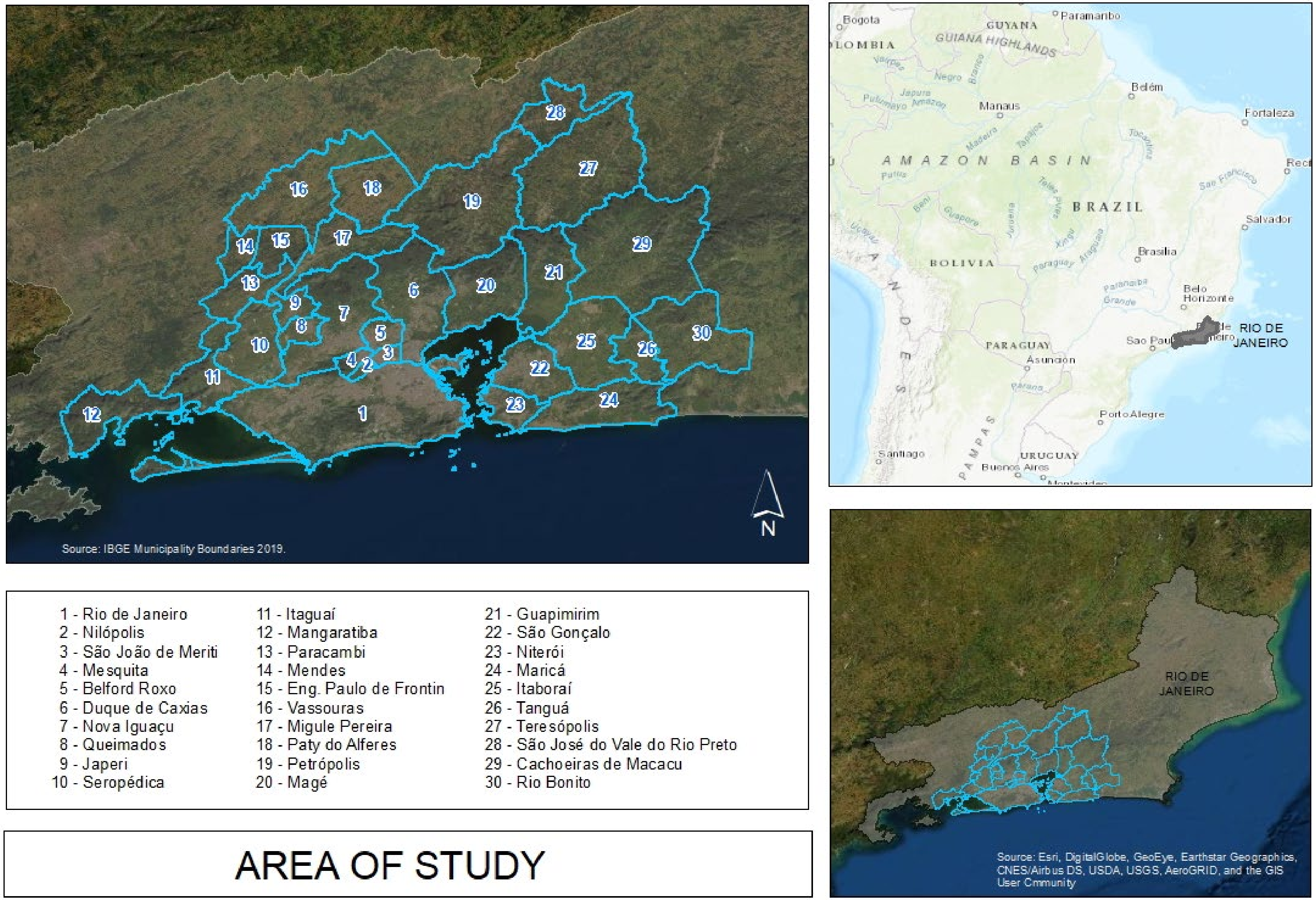
TIME STAMP	COD_SITE	TECHNOLOGY	ID_SUBSCRIBER
DD/MM/YYYY – HH:MM:SS	ID	GSM/3G/4G	Anonymous subscriber ID

Table 2: Attributes of the cell tower table

COD_SITE	LAT	LON	GSM_STATUS	UMTS_STATUS	LTE_STATUS	CITY
ID	Decimal degrees	Decimal degrees	Active/Inactive	Active/Inactive	Active/Inactive	Name

[4] Teleco

Figure 1: Area of study



2.2 Reference data

2.2.1 Household survey data

The Continuous National Household Sample Survey (PNAD Contínua) is a multipurpose rotating short sample survey initiated in 2012, in which households are interviewed five times, once per quarter. The sample covers about 3 000 of 5 570 municipalities in Brazil, and over 800 000 households are interviewed each year, with a focus on labour force statistics. The survey was designed to give quarterly employment estimates for all 27 states, along with metropolitan areas and state capitals. Beyond labour force statistics, a given topic might be investigated in all five interviews or concentrated on in a quarter. Topics include education and ICT. ICT was first investigated in 2016 and has since been investigated in the last quarter (Q4) of each year.

2.2.2 Population estimates

The estimates for the total population of the Brazilian municipalities are calculated based on the mathematical method "AiBi" developed in 1972 by João Lira Madeira and Celso Cardoso da Silva Simões. The basic inputs used in this method are the populations obtained from the most recent national and state-level population projections, along with the population growth of each municipality in the last decade, based on their respective populations listed in the last two population censuses. In addition to these basic inputs, each year updates of Brazil's political-administrative division are incorporated into the municipal population estimates. The updates reflect any changes that occurred to the municipalities' territorial boundaries since the last population census.

Population estimates provided by WorldPop[5] was also utilized, specifically in the calculation of indicator 9.c.1. This is because the municipality represents the smallest geographical unit in IBGE's population estimates, and in the case of indicator 9.c.1, population in small areas is necessary to establish the total population covered by a mobile phone network.

Comparing the two sources of population estimates, there is no significant differences (average of 3.71 per cent and median of 1.85 per cent).

2.2.3 Municipal boundaries

The municipal boundaries constitute reference data complying with the requirements of the official local administrative units: total coverage of the territory, no overlap, hierarchical, geometrical, and containing a unique keycode (geocode). They are presented as vector shapefiles.

[5] WorldPop

2.2.4 Digital elevation model

The elevation model is derived from the Shuttle Radar Topography Mission at a resolution of 1 arc-second (30 metres) with open worldwide distribution. The product used in this study was provided by Embrapa.[6]

3. Data quality assurance

3.1 Cell tower data

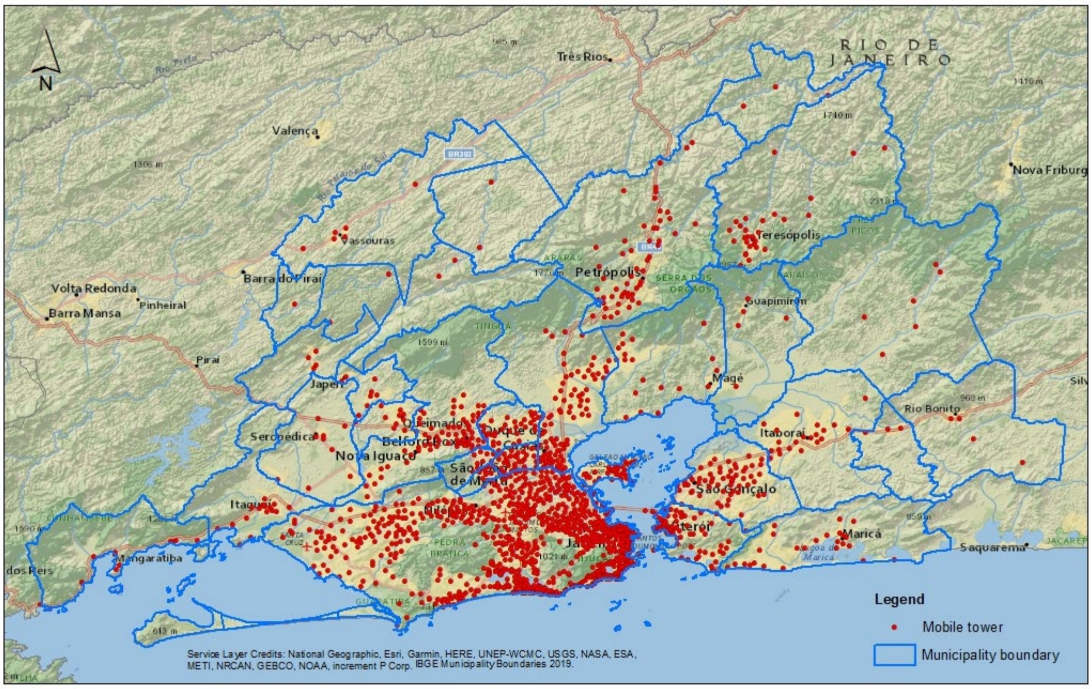
3.1.1 Missing values

The dataset contains no NULL values in any of the mandatory fields, which include CELL_ID, LON, and LAT.

3.1.2 Geographical distribution of cell towers

The dataset contains no incorrect cell tower coordinates, and all regions are covered by at least one cell tower. However, an inconsistency was identified between the coordinates and municipality names: in one municipality, four cell towers matched the coordinates, while in another, they aligned with the attribute data in the table. To prevent overcounting, the records associated with these towers were excluded from the dataset.

Figure 2: Cell tower location map



---

[6] Embrapa: Monitoramento por Satélite

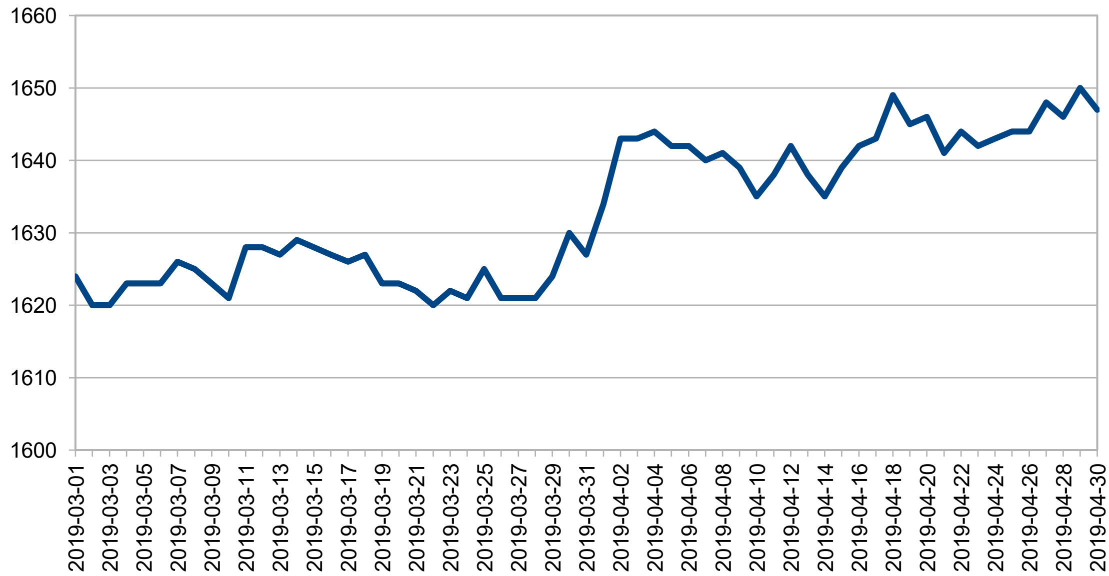
3.1.3 Cell tower occupancy

The raw data includes 1 678 cell towers and, after cleaning, the number of cell towers stands at 1 663 (the difference is 15 cell towers or 0.89 per cent of total). These cell towers have been excluded because they account for one event or less per month and are considered random cell towers.

All cell towers in the CDR/IPDR dataset are in the cell tower dataset; the reverse is also true.

The cell tower distribution for the period of the dataset is shown in the chart below. An incremental rise is noted in the number of cell towers in April (around 1 per cent).

Figure 3: Number of cell towers per day

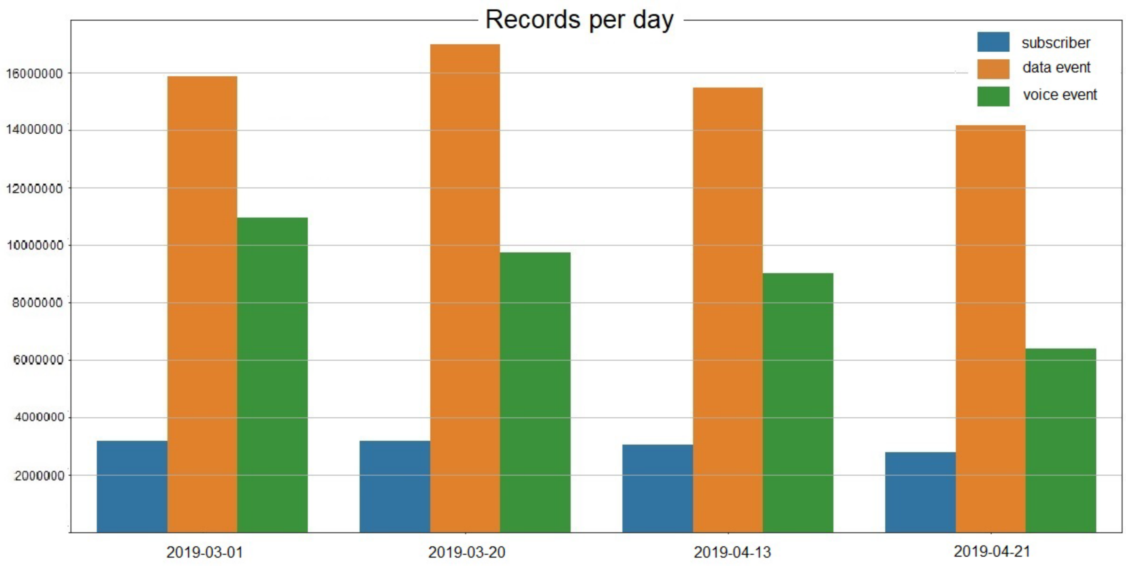


3.2 CDR/IPDR data

3.2.1 Number of records per day

There are no unexplained peaks or lows in records per day, as can be seen in the chart below showing an example of four days during the period of the dataset.

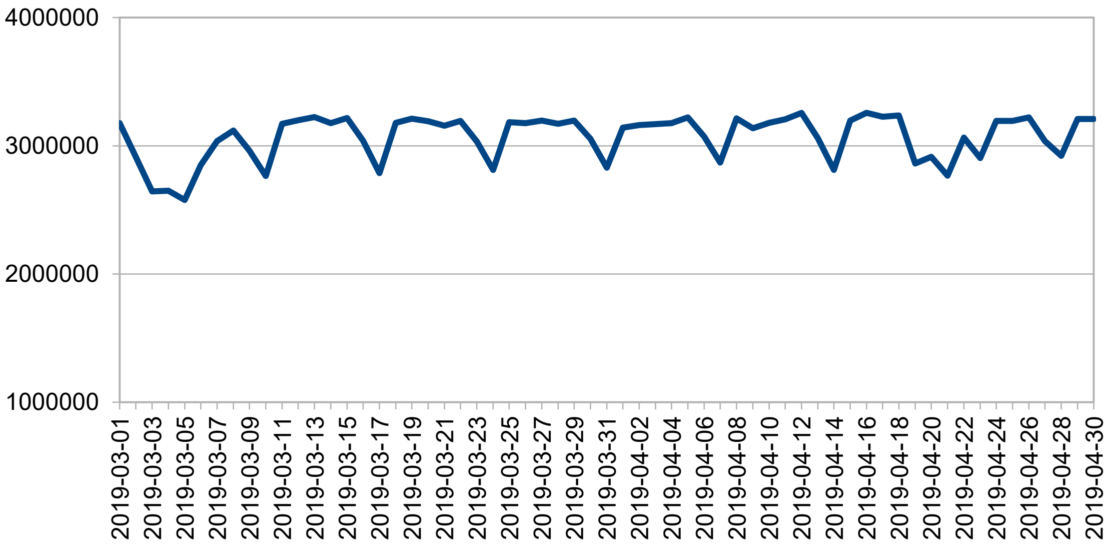
Figure 4: Number of records per day



3.2.2 Number of unique subscribers per day

No unexpected peaks or lows appear in the number of subscribers per day (see chart below). The fluctuations at the beginning and end of the period are due to holidays.

Figure 5: Number of subscribers per day



3.2.3 Subscriber presence in data

Out of the 61 days studied, subscribers used data roughly every second day and voice only every third day on average. However, more than 25% of the subscribers used data nearly every day. The table below shows the statistics relating to subscribers in the 61 days of the dataset, disaggregated by voice, data and total (combined voice and data) records.

Table 3: Descriptive statistics of subscriber usage during the time period

TYPE	MEAN	MIN	MAX	25% PERCENTILE	50% PERCENTILE	75% PERCENTILE	95% PERCENTILE	SD
Total	34.667	1	61	13	37	56	61	21.109
Data	33.212	1	61	11	35	55	61	21.475
Voice	22.065	1	61	7	18	36	55	17.409

The following charts show the distribution of the number of subscribers per number of days in the dataset during the period considered in the study.

It can be noted that, in general, voice subscribers use mobile phones for fewer days than data subscribers.

Figure 6: Distribution of the number of subscribers by the total number of days of mobile phone voice usage

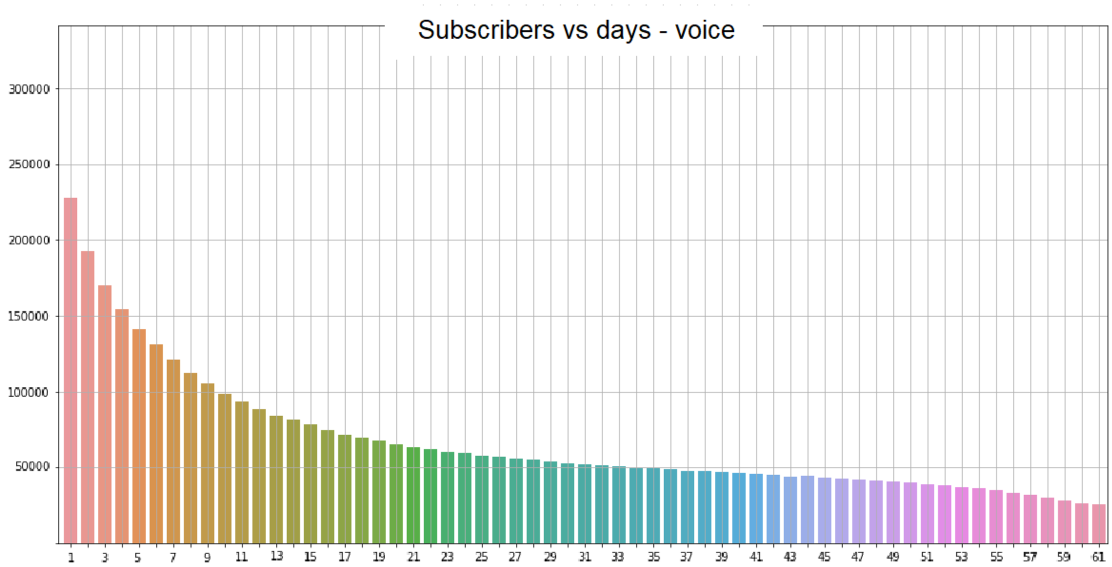
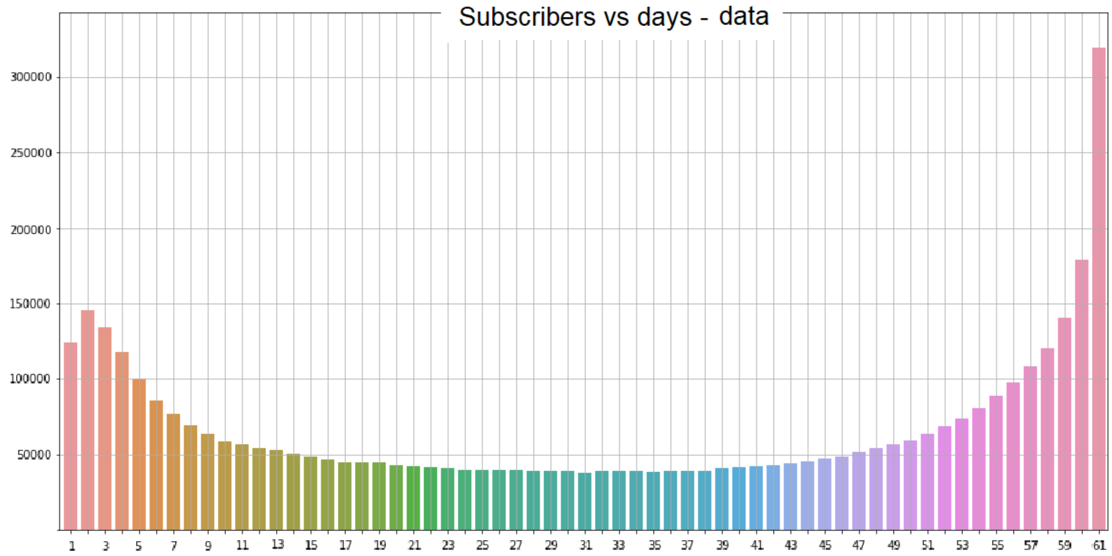


Figure 7: Distribution of the number of subscribers by the total number of days of mobile phone data usage



3.2.4 Diurnal distribution of records

The charts below show the average number of records per hour, between 0000 and 2300 hours, disaggregated by voice and data records. The shape in the chart relating to voice records presents an elephant curve, while the shape in the chart relating to data records presents a somewhat peculiar pattern, with one peak in the early hours (from 0000 to 0200 hours) and another at 0700 hours.

Figure 8: Distribution of mobile phone voice usage by hour

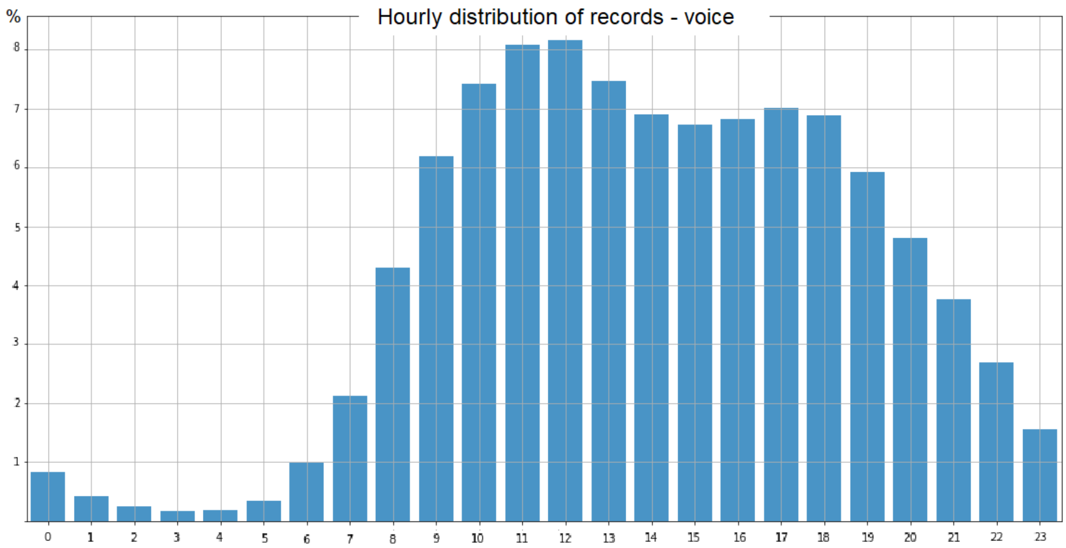
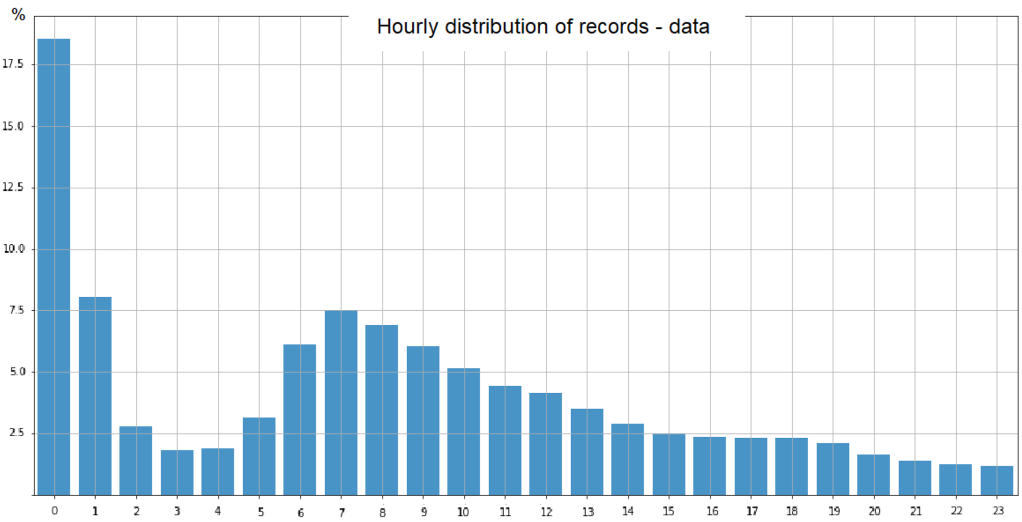


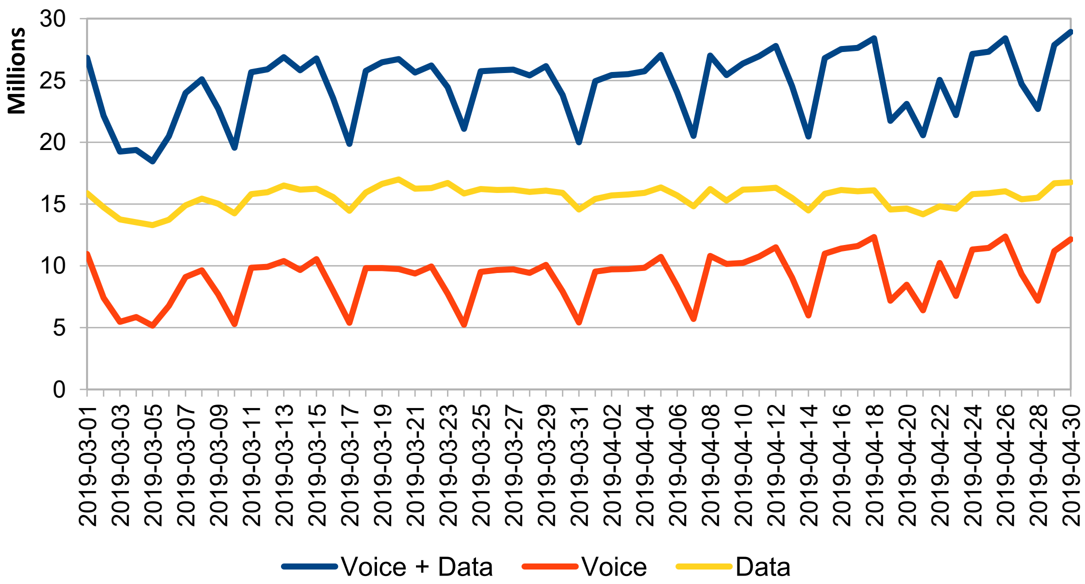
Figure 9: Distribution of mobile phone data usage by hour



3.2.5 Weekly distribution of records

The pattern of the average number of records per day is as expected, showing fewer events at weekends than on weekdays.

Figure 10: Number of events per week



3.2.6 Average number of records per day per subscriber

The table below shows the main descriptive statistics, disaggregated by voice and data records, on four days during the period of the dataset.

Table 4: Descriptive statistics of the average number of records per day per subscriber

Type	Date	Mean	Min	Max	25% percentile	50% percentile	75% percentile	95% percentile	SD
Voice	2019-03-01	5.123	1	3 391	2	3	6	15	523.349
Voice	2019-03-20	4.651	1	1 749	2	3	6	13	477.375
Voice	2019-04-13	4.641	1	2 005	2	3	6	14	478.086
Voice	2019-04-21	4.145	1	1 340	2	3	5	12	435.188
Data	2019-03-01	6.728	1	1 056	5	7	8	11	353.615
Data	2019-03-20	7.016	1	1 380	6	7	8	11	394.653
Data	2019-04-13	6.630	1	929	5	7	8	10	352.211
Data	2019-04-21	6.574	1	1 000	5	7	8	10	381.047

3.2.7 Identification of time zone

Only one time zone is presented in the dataset, identified as GMT-3.

4. Home anchoring

The locations of mobile phone subscribers' residences were inferred based on analysis of the time of the events (voice call and data transfer). The subscriber was considered to be at home if there were events at three specific times (0000 to 0500, 0500 to 0800 and 2100 to 0000 hours) on weekdays. After daily analysis throughout the period covered by the dataset, the cell tower most frequently used at one of these three times, from Monday to Thursday, in the preferential order above, is considered to be the location of the subscriber's residence. Friday was disregarded as people usually start the weekend on this day and might be absent from their usual place of residence.

Using this methodology, it was not possible to identify the place of residence of about 8 per cent (356 299 subscribers) of the 4 053 968 subscribers initially submitted to the anchor processing: 6.5 per cent because there were no events during any of the three defined periods and 1.5 per cent because there were no events on the weekdays defined.

Considering the three periods selected for the home anchoring process, 86 per cent of the subscribers were anchored using the period 0000 to 0500 hours, 1.8 per cent using the period 0500 to 0800 hours and 5 per cent using the period 2100 to 0000 hours.

5. Calculation of indicators

Once the anchor home has been calculated, the indicators proposed by this study can be calculated. The first indicator is the proportion of people using the Internet (SDG indicator 17.8.1). This is calculated by summarizing the number of subscribers in the dataset with only voice events, only data events, and both. This makes it possible to highlight the proportion of people using the Internet (subscribers with only data events, or data and voice events) among those who own a mobile phone (all subscribers).

This indicator can also be enriched by including the Internet access technology. This information is processed by selecting the technology that each subscriber uses most frequently for Internet access. The number of people using the Internet is then disaggregated by technology (2G, 3G and 4G).

The variables are calculated per municipality by adding up all the results from the mobile towers located within the borders of each municipality in the study area, using a point-in-polygon spatial analysis.

Indicator 9.c.1 was calculated using the viewshed approach presented in Section 2.6.2 of this handbook. The following parameters were used:

- Radius: 10 km if technology is 2G; 5 km if 3G; 3 km if 4G; 10 km if general;
- Observer height: 50 m.

6. Results and discussion

6.1 Indicator 17.8.1

The results of indicator 17.8.1 are presented in the table below, providing the proportion of individuals using the Internet, in total and broken down by technology, among individuals using a mobile phone, according to the dataset explored in this study. The results are also presented using maps in Figures 11 and 12.

Table 5: Proportion of the population with Internet access, by type of technology

	INTERNET ACCESS (%)		INTERNET TECHNOLOGY (%)		
MUNICIPALITY	**NO**	**YES**	**2G**	**3G**	**4G**
Belford Roxo	4.78	95.22	2.07	40.47	57.46
Cachoeira de Macacu	9.03	90.97	6.74	54.01	39.25
Duque de Caxias	3.83	96.17	0.98	37.89	61.12
Engenheiro Paulo de Frontin	6.68	93.32	30.87	59.58	9.55
Guapimirim	3.30	96.70	1.34	64.49	34.17
Itaboraí	7.47	92.53	5.08	38.74	56.18
Itaguaí	3.45	96.55	0.64	50.74	48.63
Japeri	4.49	95.51	2.20	52.23	45.57
Magé	3.59	96.41	1.32	51.34	47.34

Mangaratiba	1.81	98.19	0.13	53.89	45.98
Maricá	6.43	93.57	43.61	45.39	11.00
Mendes	5.21	94.79	8.12	87.18	4.71
Mesquita	6.39	93.61	3.99	53.12	42.89
Miguel Pereira	6.22	93.78	4.38	86.67	8.95
Nilópolis	6.18	93.82	3.63	50.82	45.55
Niterói	2.84	97.16	0.45	38.11	61.44
Nova Iguaçu	3.50	96.50	0.72	42.00	57.28
Paracambi	3.58	96.42	0.11	53.67	46.22
Paty do Alferes	8.68	91.32	29.19	41.67	29.13
Petrópolis	2.36	97.64	0.34	50.52	49.14
Queimados	6.68	93.32	4.95	57.56	37.49
Rio Bonito	7.87	92.13	17.90	48.01	34.09
Rio de Janeiro	5.13	94.87	3.70	35.88	60.42
São Gonçalo	5.92	94.08	3.14	42.20	54.66
São João de Meriti	6.71	93.29	5.15	36.94	57.90
São José do Vale do Rio Preto	2.97	97.03	0.79	58.18	41.03
Seropédica	3.00	97.00	0.22	65.61	34.17
Tanguá	9.28	90.72	11.88	49.08	39.03
Teresópolis	7.12	92.88	7.00	48.84	44.15
Vassouras	6.65	93.35	8.07	49.22	42.72

Figure 11: Proportion of the population using the Internet by area

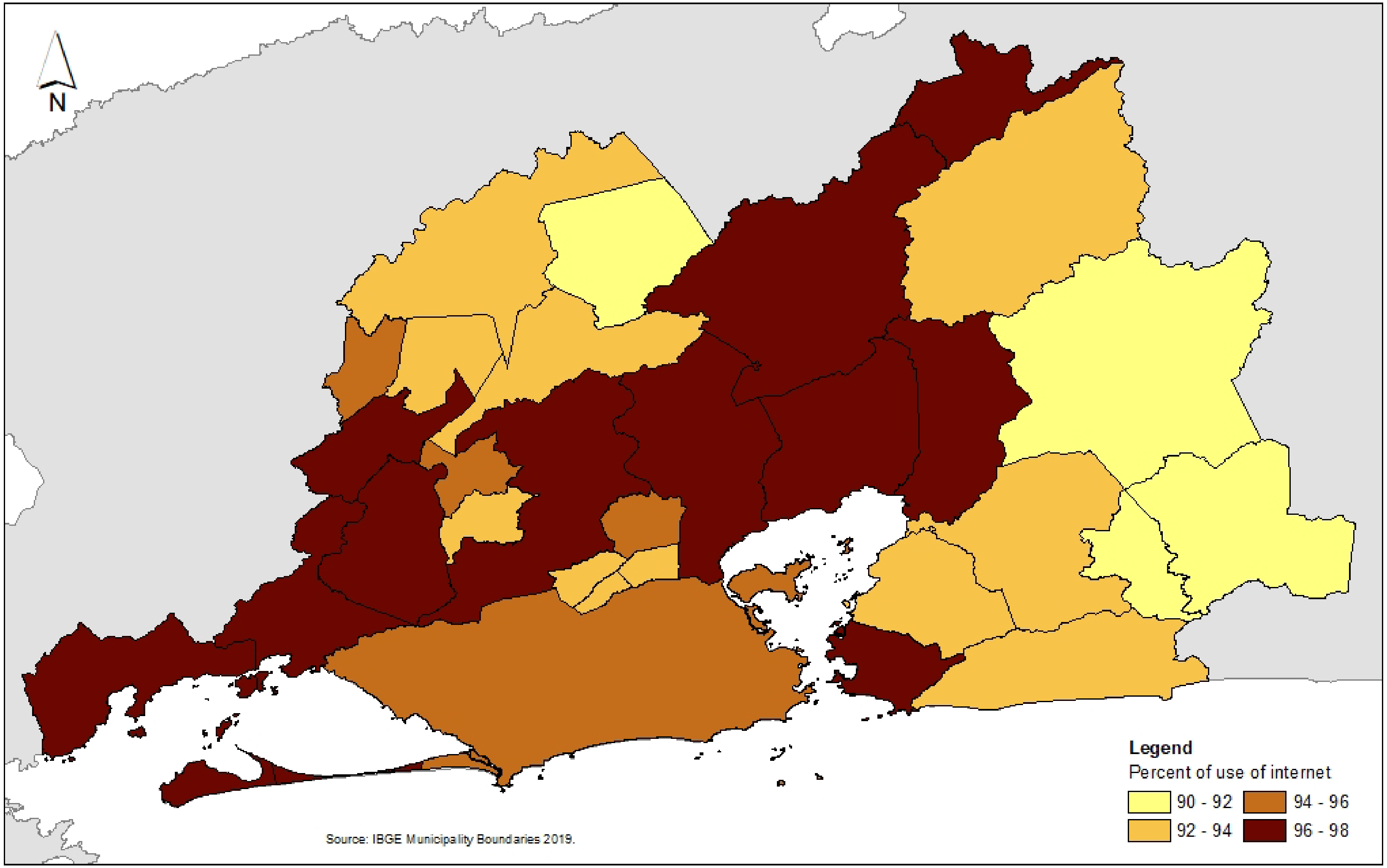
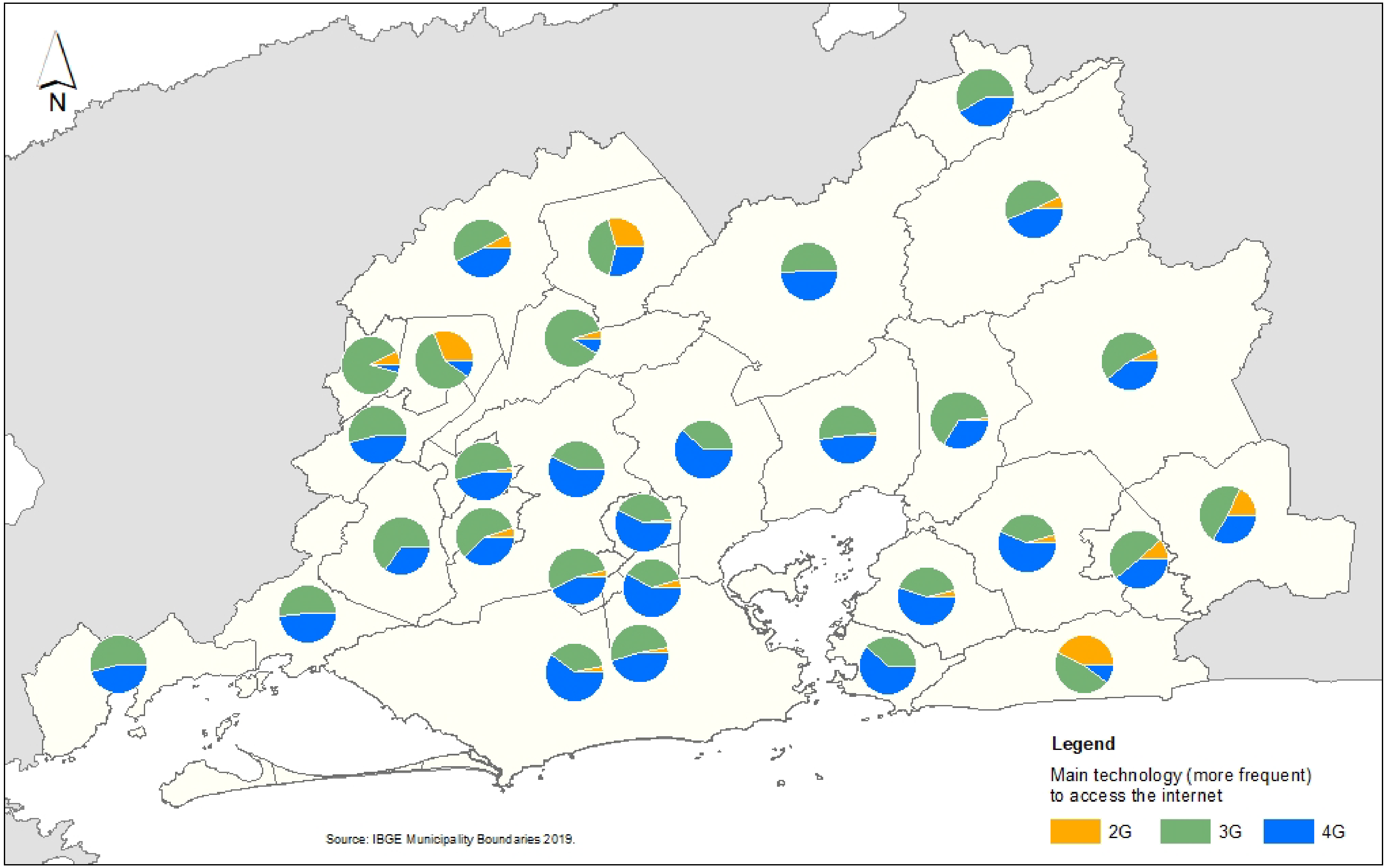


Figure 12: Distribution of the technology to access the Internet by area



For the purpose of the present study, mobile phone data from April to May 2019 will be compared with survey estimates from Q4 2018. After consultation with the Methodological Division in IBGE, it will be possible to disclose estimates: (a) as close as possible to the area of study (i.e. 29 out of 30 municipalities in this area); (b) for the Rio de Janeiro Metropolitan Area; (c) related to the City of Rio de Janeiro; and (d) of differences regarding aggregates of municipalities in the Rio de Janeiro Metropolitan Area. As PNAD Contínua was not designed to produce estimates at these levels, the results need to be interpreted carefully and only as an additional element to contribute to the discussion on the validity of this study.

Table 6 shows 2018 and 2019 population estimates for the 30 municipalities included in the study area, along with Rio de Janeiro Metropolitan Area estimates (19 municipalities).

Table 6: IBGE 2018 and 2019 population estimates per geographical unit

Geographical area	Pop 2018	Pop 2019
Rio de Janeiro*	6 688 927	6 718 903
São Gonçalo*	1 077 687	1 084 839
Duque de Caxias*	914 383	919 596
Nova Iguaçu*	818 875	821 128
Niterói*	511 786	513 584
Belford Roxo*	508 614	510 906
São João de Meriti*	471 888	472 406
Petrópolis	305 687	306 191
Magé*	243 657	245 071
Itaboraí*	238 695	240 592
Teresópolis	180 886	182 594
Mesquita*	175 620	176 103
Nilópolis*	162 269	162 485
Maricá*	157 789	161 207
Queimados*	149 265	150 319
Itaguaí*	125 913	133 019
Japeri*	103 960	104 768
Seropédica*	86 743	82 312
Guapimirim*	59 613	60 517
Rio Bonito	59 814	60 201
Cachoeiras de Macacu	58 560	58 937
Paracambi*	51 815	52 257
Mangaratiba	43 689	44 468
Vassouras	36 702	36 896
Tanguá*	33 870	34 309
Paty do Alferes	27 678	27 769
Miguel Pereira	25 493	25 538

São José do Vale do Rio Preto	21 670	21 795
Mendes	18 578	18 614
Engenheiro Paulo de Frontin	13 929	14 002
Area of study (30 municipalities)	13 374 055	13 441 326
Rio de Janeiro Metropolitan Area (19 municipalities)	12 581 369	12 644 321

* Part of Rio de Janeiro Metropolitan Area

The PNAD Contínua sample for Q4 2018 covered 29 of the 30 municipalities in the table above, the exception being Miguel Pereira. The relevant survey population for this study has been post-stratified to match the population in the capital (City of Rio de Janeiro) and the Metropolitan Area. A difference is therefore to be expected vis-à-vis other areas. For (a), (b) and (c) indicated above, it was less than 1 per cent.

Mobile phone data is only sourced by one operator and therefore covers only about 36 per cent of total mobile phone subscribers in the region (estimated market share). At the same time, the survey data makes it possible to estimate the number of people with a mobile phone and the number with a mobile phone and Internet access on the same device.

Table 7: Internet use survey estimates

Geographical area	Population (above 10 years of age)					
	Total (A)	With mobile phone (B)	With mobile phone and Internet access (C)	(B/A) %	(C/B) %	(C/A) %
Area of study	11 924 615	10 145 078	9 525 418	85.1	93.9	79.9
Rio de Janeiro Metropolitan Area	11 262 615	9 586 898	9 012 308	85.1	94.0	80.0
City of Rio de Janeiro	5 995 720	5 241 579	5 009 556	87.4	95.6	83.6

Source: Q4 2018, PNAD Contínua Survey/IBGE.

To address indicator 17.8.1, the proxy indicator generated locally would be (C/A), which shows almost 80 per cent of the population accessing the Internet in the study area, considering that only access via the owner’s mobile phone is covered. Access with other devices (computers, borrowed phones, etc.) is not covered and no timeframe is considered.

In addition, in the interests of the present study, an evaluation of the relationship between mobile data and survey data is set out most clearly in Table 8. The mobile data and survey data indicators are constructed as follows:

- Mobile data: the proportion of unique users with data records (numerator) in relation to all users identified (denominator).
- Survey data: the proportion of the target population (persons above 10 years of age) with Internet access via their mobile phone (numerator) in relation to the target population (persons above 10 years of age) who indicate they have a mobile phone for personal use (denominator) (see Table 7 (C/B).

Table 8: Comparison of Internet use estimates from mobile and survey data

Geographical disaggregation	Internet access using mobile phone (%)		
	Mobile data	Survey data	Difference
Area of study	93.91	93.89	0.02 p.p.
Rio de Janeiro Metropolitan Area	95.04	94.01	1.04 p.p.
City of Rio de Janeiro	94.87	95.57	-0.70 p.p.

Source: Mobile data and Q4 2018 PNAD Contínua Survey/IBGE.

Results are surprisingly similar for the total area of study, with a difference of up to 1.04 per cent regarding the City of Rio de Janeiro. One would hypothesize that the structure of the mobile data obtained was like that of the mobile phone users in these areas. The operator thus appears to have a relatively broad client base which is representative of the population.

One additional step would be to explore more granular geographical areas to compare survey and mobile data results, with the focus on matching of tower/geographical disaggregation, coverage and the statistical precision of survey data.

Of the 29 municipalities in this study area that are part of the PNAD Contínua sample, four have a variation coefficient less than or equal to 15 (estimates classified as A or B quality[7]) for the estimate of people (above 10 years) with Internet access via their mobile phone. With a laxer criterion, 12 cities have a variation coefficient less than or equal to 30 (estimates ABC). Considering all municipalities, the highest variation coefficient identified was 74.6 per cent.

Bearing in mind the precision of estimates and the fact that the survey was not designed to disclose results at these levels of disaggregation, the following table (Table 9) explores mean and median absolute differences between mobile and

[7] To help communicate the precision level of the estimates, the following classification is used by IBGE regarding variation coefficient intervals (%): A (<=5), B (<=15), C (<=30), D (<=50), E (>50).

survey data, grouping the municipalities by the quality of survey results. Four (more populous) cities with AB estimates have a 1.8 percentage points mean difference, while mean differences reach as high as 5.4 percentage points for 29 municipalities (varying considerably in size).

Table 9 – Mean and median absolute difference between mobile and survey estimates for different geographical disaggregation (percentage points).

Table 9: Differences between mobile and survey estimates by geographical disaggregated areas

Geographical disaggregation	Mean	Median
City of Rio de Janeiro	0.7	0.7
4 municipalities with AB estimates	1.8	1.0
12 municipalities with ABC estimates	2.7	1.6
29 municipalities with ABCDE estimates	5.4	4.8
Municipality with lower difference	0.1	0.1
Municipality with higher difference	18.6	18.6

Source: Mobile data and PNAD Contínua Survey/IBGE.

6.2 Indicator 9.c.1

The following table shows the results for the proportion of the population covered by the mobile network, disaggregated by the technology. The first four columns show the results obtained for this case study and the last three columns show the results presented by ANATEL in its Panorama Data Panel[8] for the year 2020.

The results are shown as maps and charts in Figures 13, 14 and 15.

Table 10: Mobile network coverage by geographical area

	Case study methodology				**ANATEL**		
Geographical area	**2G**	**3G**	**4G**	**General**	**2G**	**3G**	**4G**
1-Belford Roxo	97.44	97.85	97.05	98.10	98.79	99.67	99.81
2-Cachoeira de Macacu	81.68	84.34	79.22	72.69	77.34	85.85	80.23
3-Duque de Caxias	98.37	98.62	97.64	99.14	98.24	99.13	99.54
4-Eng. Paulo de Frontin	46.30	29.00	0.00	26.02	45.24	42.18	41.44
5-Guapimirim	82.51	80.37	71.54	87.08	84.72	90.94	93.35
6-Itaboraí	92.63	90.92	89.29	93.20	91.65	94.84	96.03
7-Itaguaí	93.95	96.98	95.22	97.92	96.37	97.77	97.91
8-Japeri	92.15	91.68	88.83	93.40	96.45	98.74	99.26
9-Magé	84.49	85.82	82.35	90.61	86.56	91.58	94.97

[8] PAINÉIS DE DADOS: Agência Nacional de Telecomunicações

10-Mangaratiba	89.14	88.36	85.74	83.98	82.73	85.80	86.57
11-Maricá	90.88	93.46	89.75	94.87	94.01	97.50	97.96
12-Mendes	52.27	51.52	0.00	59.39	60.26	65.92	71.56
13-Mesquita	99.46	99.46	99.41	100.00	99.98	99.98	99.99
14-Miguel Pereira	60.00	59.70	1.86	69.55	69.77	75.23	34.74
15-Nilópolis	100.00	100.00	100.00	100.00	100.00	100.00	100.00
16-Niterói	96.48	96.79	96.09	98.43	98.59	99.48	99.54
17-Nova Iguaçu	95.94	96.05	95.25	97.63	97.47	98.69	98.83
18-Paracambi	79.61	82.48	81.61	77.72	78.59	91.76	90.60
19-Paty do Alferes	46.30	31.81	28.37	32.85	54.49	46.05	53.41
20-Petrópolis	85.61	86.68	83.66	86.22	88.85	91.53	94.29
21-Queimados	91.14	91.21	88.00	91.23	96.75	99.02	99.61
22-Rio Bonito	71.20	59.87	57.30	76.51	77.61	71.85	75.49
23-Rio de Janeiro	99.30	99.54	99.46	99.92	99.68	99.93	99.95
24-São Gonçalo	95.72	95.28	94.22	96.65	98.52	99.57	99.84
25-São João de Meriti	99.56	99.56	99.52	100.00	100.00	100.00	100.00
26-S. José do Vale do Rio Preto	43.26	42.87	40.54	38.14	50.35	51.37	51.74
27-Seropédica	91.34	89.32	82.94	99.23	86.39	99.27	99.41
28-Tanguá	80.43	91.85	91.94	88.38	79.20	89.83	90.82
29-Teresópolis	87.29	87.90	86.43	86.54	85.81	86.84	88.36
30-Vassouras	78.40	78.23	76.91	71.06	72.01	73.60	75.46
Area of study	96.63	96.79	95.89	97.33	-	-	-
Rio de Janeiro Metro Area	97.54	97.76	97.13	98.53	-	-	-

As we can see, although the figures are similar, discrepancies remain: the differences for 2G technology vary from 10 per cent to 7 per cent, for 3G from 16 per cent to 5 per cent, and for 4G from 72 per cent to 2 per cent. It must nevertheless be borne in mind that we are comparing results from different models that use distinct parameters. The first difference is the methodology. ANATEL uses a propagation model based on existing mobile phone towers in 2020 with specific parameters for distance and antenna height. The case study uses the model described in section 4.2.5, the parameters of which are also presented. The second difference is the population data. ANATEL uses data from the IBGE 2010 population census, whereas the case study uses 2020 population estimates produced by WorldPop. These factors, notably the first, may explain the better coverage rates achieved by ANATEL, mainly for 4G technology.

Figure 13: Map of mobile network coverage by geographical area

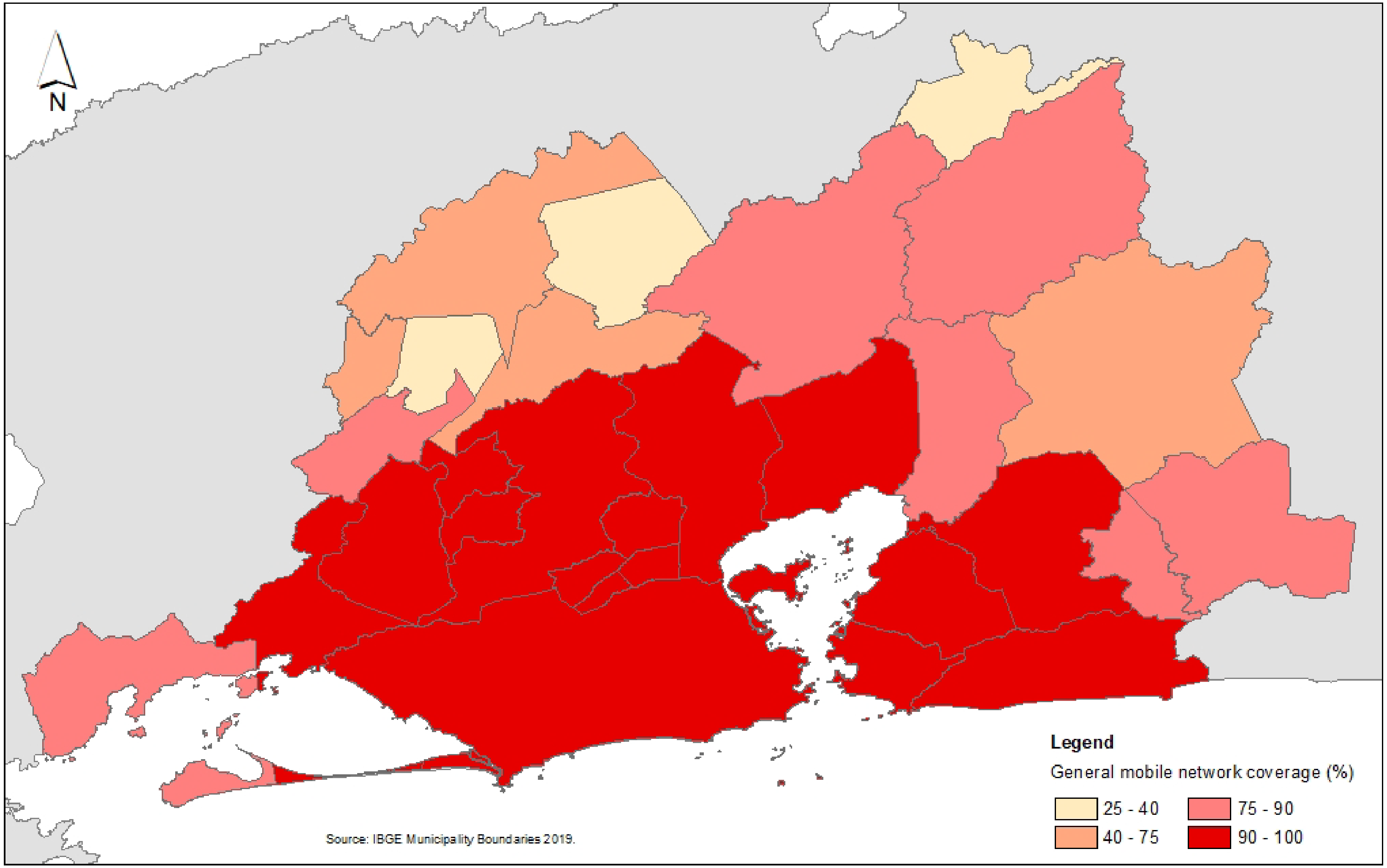


Figure 14: Map of mobile network coverage by geographical area and technology

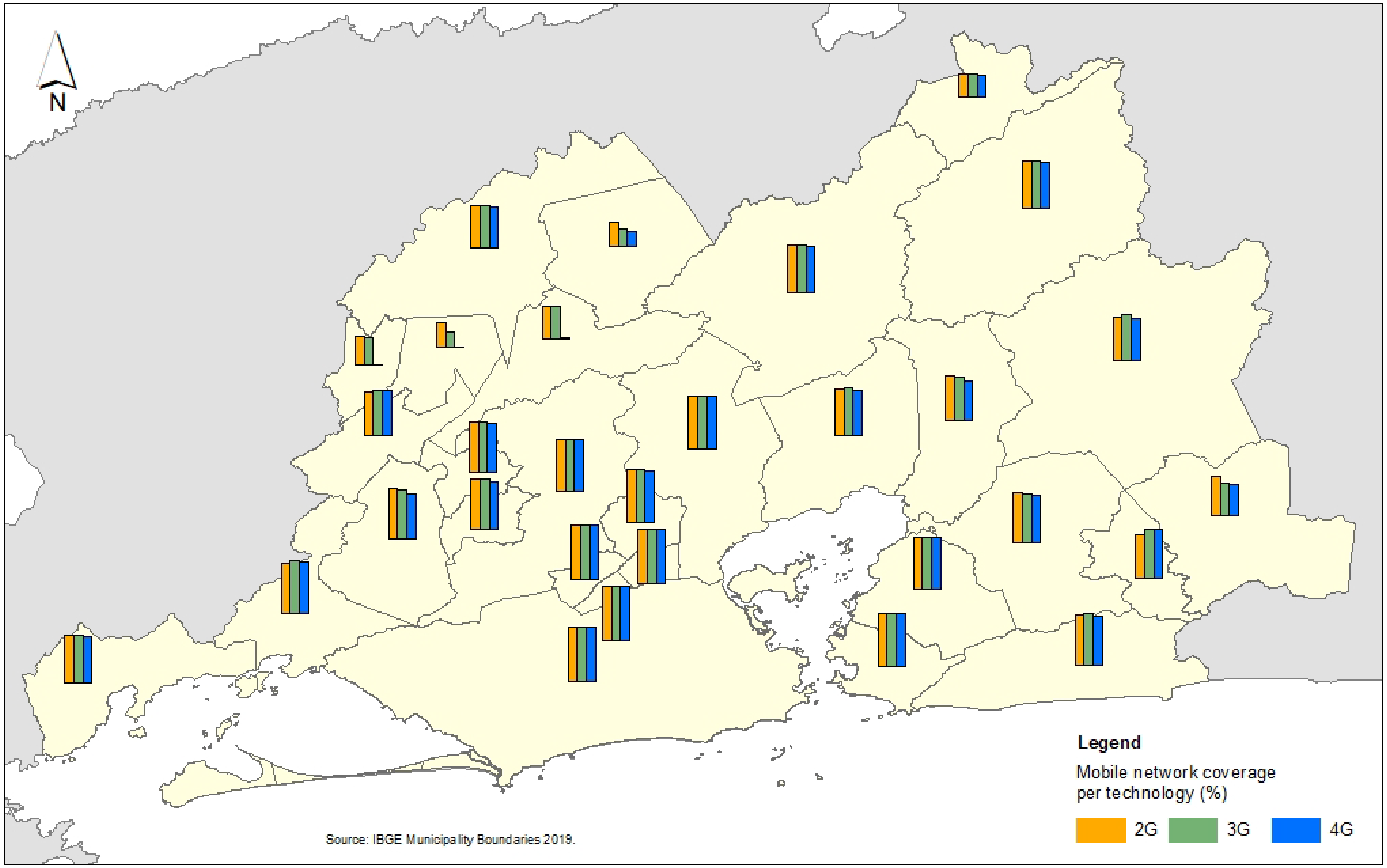


Figure 15: Comparison of mobile phone coverage between Anatel and the mobile phone study, by technology