

Relatório Final

Índice Brasileiro de Conectividade – Análise Completa

João Carlos, Leano Guerreiro e Brayner Santana

2025-07-17

- 1 Introdução
- 2 Pacotes
- 3 Autenticação e Download dos Dados
- 4 Leitura dos Dados
- 5 Definição de Regiões
- 6 Análise Exploratória dos Dados (EDA)
 - 6.1 Dataset completo
 - 6.2 Dataset da região Norte
- 7 Visualizações
- 8 Detecção de Outliers
- 9 Visualizações de Mapas Coropléticos
- 10 Testes de Hipótese
- 11 Gráficos adicionais
- 12 Conclusão

1 Introdução

Este relatório tem como objetivo principal apresentar uma análise aprofundada sobre o **nível de população atendida por abastecimento de água e esgotamento sanitário** na Região Norte do Brasil, utilizando dados do período de 2012 a 2021. O saneamento básico é um pilar fundamental para a saúde pública, qualidade de vida e desenvolvimento socioeconômico de uma região. A Região Norte, com suas particularidades geográficas e demográficas, enfrenta desafios únicos na universalização desses serviços.

Ao longo deste documento, exploraremos as disparidades na cobertura de saneamento não apenas em nível regional, mas também detalharemos as diferenças entre as Unidades da Federação (UFs) que a compõem. Serão analisados volumes produzidos e consumidos, a presença de outliers nos dados e a significância estatística das diferenças observadas. A metodologia empregada combinará análise exploratória de dados (EDA), visualizações geográficas (mapas coropléticos), detecção de outliers e testes de hipótese robustos, incluindo o cálculo do tamanho de efeito (Cohen's d). O presente trabalho visa, portanto, fornecer insights valiosos para a formulação de políticas públicas e direcionamento de investimentos em saneamento na Região Norte.

A codificação do projeto foi estruturada em diversas funções separadas em diferentes arquivos .R, visando uma melhor organização, modularidade, manutenção e reutilização do código, garantindo a rastreabilidade e replicabilidade da análise.

```
source("R/00_download_data.R")
source("R/01_pacotes_e_leitura.R")
source("R/02_preparacao_dados.R")
source("R/03_eda_e_outliers.R")
source("R/04_visualizacoes_mapas.R")
source("R/05_testes_estatisticos.R")
source("R/06_graficos_adicionais.R")
source("R/07_reflexoes_finais.R")
```

2 Pacotes

Esta seção detalha os pacotes R utilizados para a realização da análise. Cada pacote desempenha um papel crucial nas etapas de manipulação, visualização e inferência estatística dos dados, garantindo a execução eficiente e precisa das metodologias propostas.

[Hide](#)

```
library(tidyverse) # Conjunto de pacotes essenciais para manipulação e visualização
                    de dados
library(arrow)      # Para leitura eficiente de grandes datasets, como o formato Parq
                    uet
library(scales)     # Para formatação de escalas em gráficos
library(skimr)      # Para gerar estatísticas descritivas rápidas e completas
library(sf)         # Para manipulação e visualização de dados geoespaciais (Simple F
                    eatures)
library(psych)      # Contém funções úteis para análise psicométrica e descritiva
library(DT)         # Para tabelas interativas em HTML
library(effsize)    # Para cálculo de tamanho de efeito, como Cohen's d
library(cli)        # Para mensagens de console formatadas e informativas
library(gtools)     # Para funções como 'combn', utilizada na geração de combinações
                    para testes
```

3 Autenticação e Download dos Dados

A etapa inicial do projeto envolve a obtenção dos dados brutos que subsidiarão toda a análise. Para garantir a reprodutibilidade e a integridade da análise, os dados são primeiramente baixados e armazenados localmente. Esta seção descreve o processo de aquisição dos dados, fundamental para a consistência do estudo.

[Hide](#)

```
query <- "
SELECT
  dados.ano as ano,
  dados.id_municipio AS id_municipio,
  diretorio_id_municipio.nome AS id_municipio_nome,
  dados.sigla_uf AS sigla_uf,
  diretorio_sigla_uf.nome AS sigla_uf_nome,
  dados.populacao_urbana as populacao_urbana,
  dados.populacao_urbana_atendida_agua as populacao_urbana_atendida_agua,
  dados.populacao_urbana_atendida_esgoto as populacao_urbana_atendida_esgoto,
  dados.volume_agua_produzido as volume_agua_produzido,
  dados.volume_agua_tratada_eta as volume_agua_tratada_eta,
  dados.volume_agua_consumido as volume_agua_consumido
FROM basedosdados.br_mdr_snis.municipio_agua_esgoto AS dados
LEFT JOIN (SELECT DISTINCT id_municipio,nome FROM basedosdados.br_bd_diretorios_b
  rasil.municipio) AS diretorio_id_municipio
  ON dados.id_municipio = diretorio_id_municipio.id_municipio
LEFT JOIN (SELECT DISTINCT sigla,nome FROM basedosdados.br_bd_diretorios_brasil.u
  f) AS diretorio_sigla_uf
  ON dados.sigla_uf = diretorio_sigla_uf.sigla
WHERE dados.ano BETWEEN 2012 AND 2021
"
```

4 Leitura dos Dados

Após o download, os dados são lidos e carregados no ambiente R para processamento. Esta seção detalha como os conjuntos de dados, incluindo as informações de serviço de água e esgoto e as geometrias dos estados, são importados e preparados para as fases subsequentes de análise.

[Hide](#)

```
# 3. Lê os dados agora que eles estão salvos Localmente
dados_carregados <- ler_dados(num_linhas_para_analise)
dados_lazy <- dados_carregados$dados_lazy
estados <- dados_carregados$estados
```

5 Definição de Regiões

Para contextualizar a análise de saneamento, é essencial categorizar as observações por suas respectivas regiões geográficas. Esta seção descreve o processo de preparação das geometrias dos estados e a atribuição de cada município à sua região correspondente, facilitando análises regionais e comparativas.

[Hide](#)

```
dados_lazy <- dados_lazy |>
  mutate(regiao = case_when(
    sigla_uf %in% c("AM", "RR", "AP", "PA", "RO", "AC", "TO") ~ "Norte",
    sigla_uf %in% c("MT", "MS", "GO", "DF") ~ "Centro-Oeste",
    sigla_uf %in% c("MA", "PI", "CE", "RN", "PB", "PE", "AL", "SE", "BA") ~ "Nordeste",
    sigla_uf %in% c("SP", "RJ", "MG", "ES") ~ "Sudeste",
    sigla_uf %in% c("PR", "SC", "RS") ~ "Sul",
    TRUE ~ "Outras"
  ))
```

[Hide](#)

```
# 4. Prepara geometrias e regiões
preparacao <- preparar_geometrias_e_regioes(estados, dados_lazy)
regioes_sf <- preparacao$regioes_sf
dados_lazy_com_regiao <- preparacao$dados_lazy
```

6 Análise Exploratória dos Dados (EDA)

A Análise Exploratória de Dados (EDA) é uma etapa fundamental para compreender a estrutura, a qualidade e as características iniciais do dataset. Através da EDA, podemos identificar padrões, anomalias e a presença de dados ausentes, que serão tratados para garantir a robustez das análises futuras.

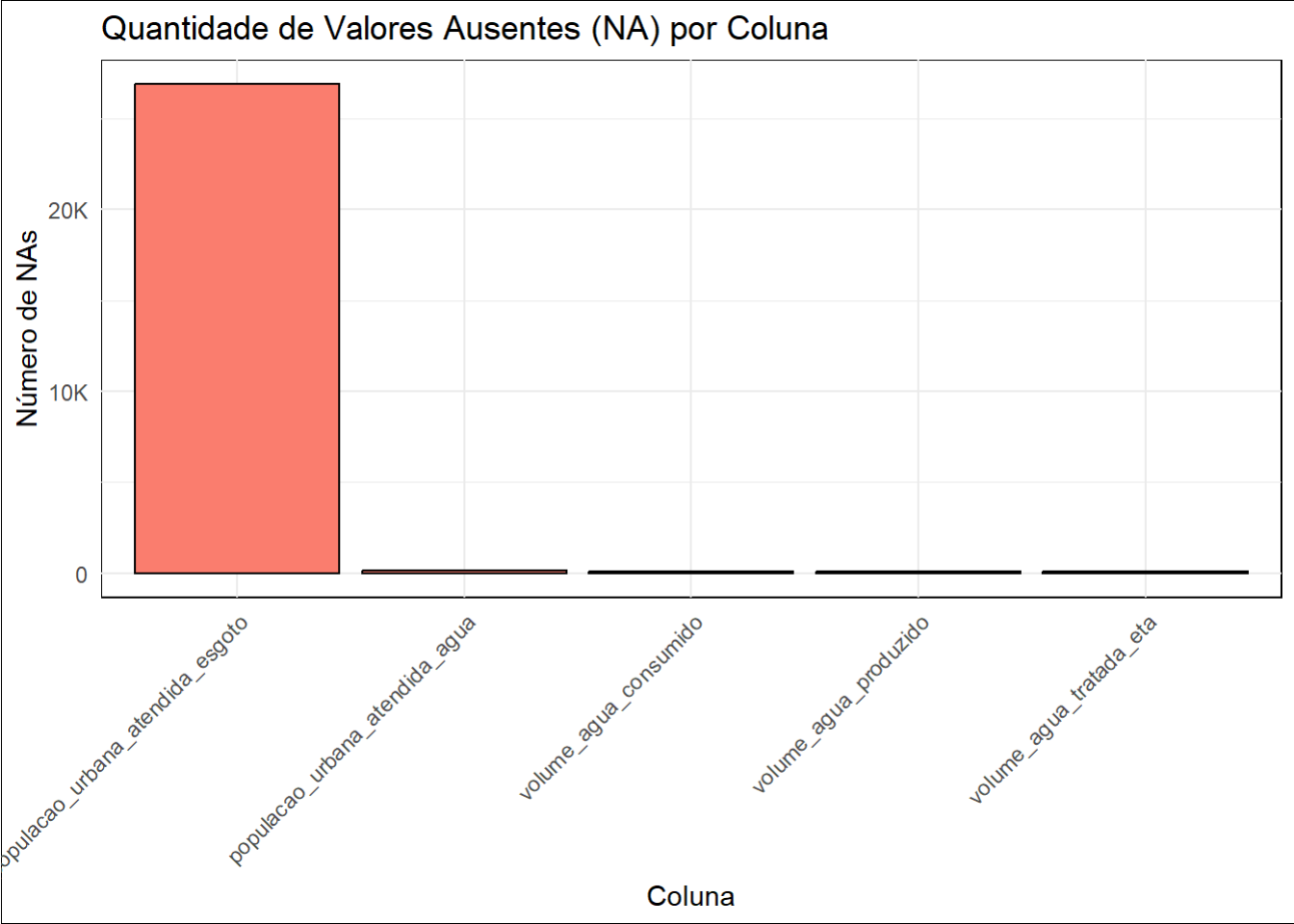
6.1 Dataset completo

Nesta subseção, realizamos uma EDA preliminar sobre o dataset completo de serviços de água e esgoto. Esta visão geral nos permite ter um panorama da totalidade dos dados antes de focar em recortes específicos.

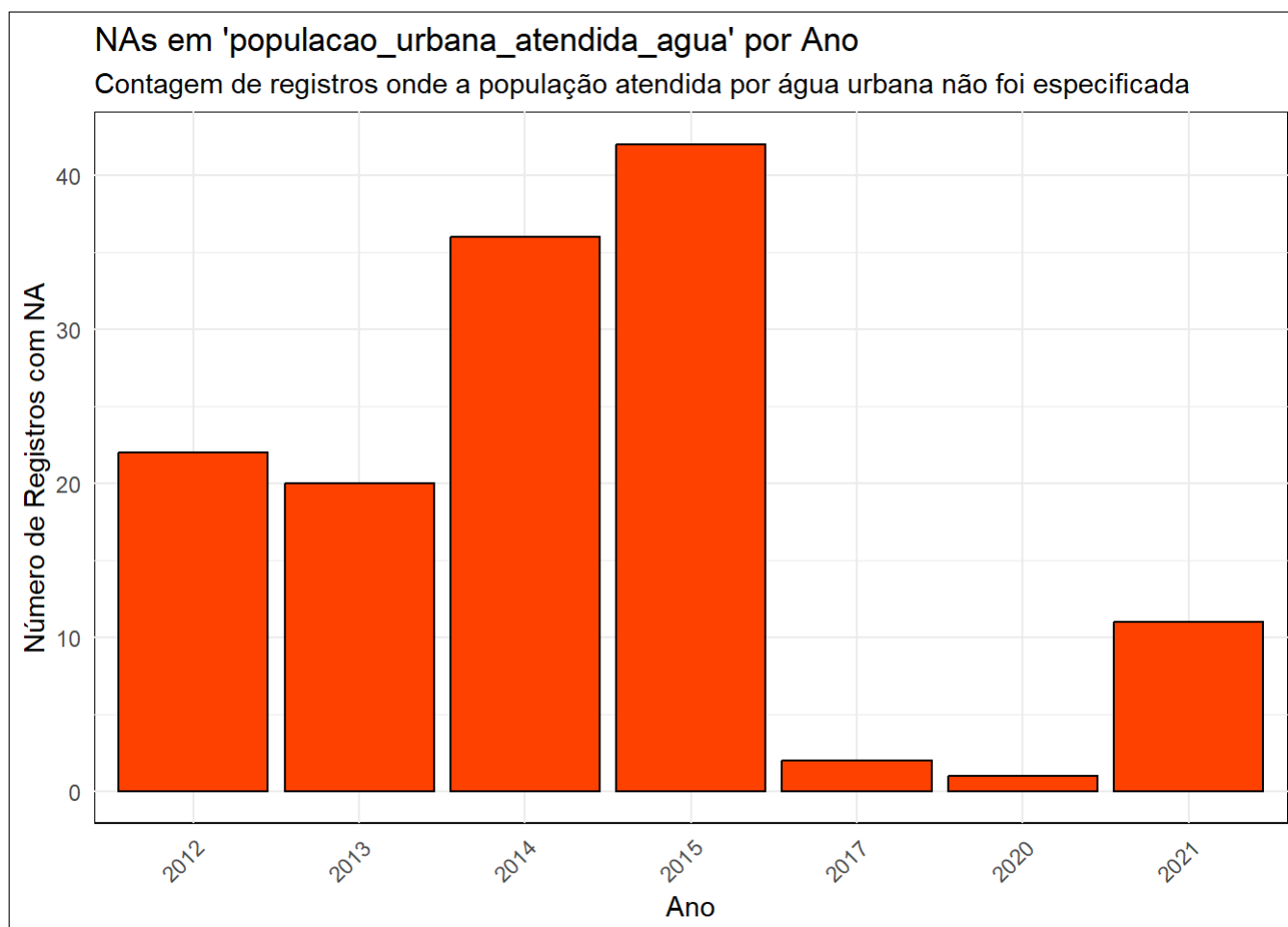
A contagem de valores ausentes por coluna (NA) revela que 'populacao_urbana_atendida_esgoto' possui o maior número de NAs, com 26895 registros, seguido por 'populacao_urbana_atendida_agua' com 134, 'volume_agua_produzido' com 99, 'volume_agua_consumido' com 99, e 'volume_agua_tratada_eta' com 95. As demais colunas (ano, id_municipio, id_municipio_nome, sigla_uf, sigla_uf_nome, populacao_urbana) não apresentam valores ausentes

```
## Contagem de valores NA por coluna:
```

```
## # A tibble: 11 × 2
##   coluna                                quantidade_na
##   <chr>                                <int>
## 1 populacao_urbana_atendida_esgoto      26895
## 2 populacao_urbana_atendida_agua         134
## 3 volume_agua_produzido                   99
## 4 volume_agua_consumido                   99
## 5 volume_agua_tratada_eta                 95
## 6 ano                                     0
## 7 id_municipio                           0
## 8 id_municipio_nome                       0
## 9 sigla_uf                                0
## 10 sigla_uf_nome                          0
## 11 populacao_urbana                       0
```



A distribuição de NAs para ‘populacao_urbana_atendida_agua’ ao longo dos anos mostra um pico de ausências em 2015, com mais de 40 registros, e uma redução significativa a partir de 2017.



Esta tabela apresenta as médias de percentual de população urbana atendida por serviços de água e esgoto para cada uma das grandes regiões do Brasil em 2021. Ela oferece uma visão consolidada do cenário nacional, permitindo uma rápida comparação do nível de atendimento entre as diferentes áreas geográficas do país. Conforme observado, a Região Norte registra os menores percentuais de cobertura tanto para abastecimento de água quanto para esgotamento sanitário.

Média de Percentuais de Água e Esgoto por Região (2021):

```
## # A tibble: 5 × 3
##   regioao      media_perc_agua_regiao media_perc_esgoto_regiao
##   <chr>          <dbl>          <dbl>
## 1 Centro-Oeste      97.2          64.1
## 2 Nordeste          87.9          51.8
## 3 Norte            80.8          36.8
## 4 Sudeste          95.3          86.8
## 5 Sul              98.3          62.2
```

6.2 Dataset da região Norte

Considerando o foco deste relatório na Região Norte, uma EDA mais detalhada é conduzida especificamente para os dados pertinentes a essa região. Esta etapa é crucial para identificar as particularidades do saneamento no Norte do Brasil, incluindo suas médias, desvios e distribuições, que servirão de base para as visualizações e testes de hipótese subsequentes.

A EDA detalhada para a ‘servico_agua_esgoto_norte’ (3627 linhas e 14 colunas) revela que, na Região Norte, o percentual do população com água atendida tem uma média de 77.9% e para a população com esgot atendido há uma média de 33.5%, com 3 e 3054 NA’s respectivamente. Isso reforça a questão dos dados ausentes em esgoto na região.

Análise Exploratória de Dados (EDA) Detalhada - Região Norte

```
## — Data Summary —————
##                               Values
## Name                         servico_agua_esgoto_norte
## Number of rows                3627
## Number of columns             14
##
## Column type frequency:
##   character                    5
##   numeric                      9
##
## Group variables                None
##
## — Variable type: character —————
##   skim_variable  n_missing complete_rate min max empty n_unique whitespace
## 1 id_municipio      0           1    7    7     0     435           0
## 2 id_municipio_nome  0           1    4   28     0     433           0
## 3 sigla_uf           0           1    2    2     0        7           0
## 4 sigla_uf_nome      0           1    4    9     0        7           0
## 5 regioa            0           1    5    5     0        1           0
##
## — Variable type: numeric —————
##   skim_variable          n_missing complete_rate    mean    sd
## 1 ano                    0           1    2017.    2.91
## 2 populacao_urbana        0           1   33081. 143469.
## 3 populacao_urbana_atendida_agua    3           0.999 22934. 120441.
## 4 populacao_urbana_atendida_esgoto 3054           0.158 28052.  68546.
## 5 volume_agua_produzido    4           0.999  2628.  12986.
## 6 volume_agua_tratada_eta    2           0.999  1705.  10947.
## 7 volume_agua_consumido    2           0.999  1283.   5858.
## 8 perc_agua_atendida      3           0.999   77.9   29.1
## 9 perc_esgoto_atendido    3054           0.158   33.5   30.6
##   p0    p25    p50    p75    p100 hist
## 1 2012 2014    2017    2019    2021
## 2 457 2978.    7212   18515   2244470
## 3  0 2277.    4800   11572.  2199581
## 4  0 1336    5000   18000    574167
## 5  0  199.    457.   1300.    244963
## 6  0  0      34.1    555.    200340
## 7  0 106.    239.    635    112534
## 8  0  59.1   95.8   100     100
## 9  0   6.02   24.9   56.0    100
```

As tabelas a seguir fornecem métricas essenciais como média, desvio padrão, mediana, mínimo, máximo e o número de observações (n) para a população urbana atendida por água e esgoto, bem como os respectivos percentuais. Esta granularidade permite identificar as características específicas de cada UF, revelando a variabilidade interna da região e servindo como base quantitativa para as comparações e testes de hipótese que serão apresentados posteriormente.

Estatísticas Descritivas de População Urbana Atendida por Água por UF na Região Norte

##	item	group1	vars	n	mean	sd	median	trimmed	mad	min	
##	X11	1	AC	1	220	17856.895	45349.40	5779.5	7142.193	4003.761	1062
##	X12	2	AM	1	329	72846.547	327867.28	10774.0	12660.777	7513.817	78
##	X13	3	AP	1	160	16871.987	43334.64	2119.0	4771.109	1338.788	28
##	X14	4	PA	1	958	30614.754	115568.73	9076.5	12532.803	8664.314	0
##	X15	5	RO	1	481	15475.304	29949.71	3181.0	7494.343	3564.170	0
##	X16	6	RR	1	150	28048.360	85795.72	4707.5	5639.575	2931.100	1132
##	X17	7	TO	1	1326	8702.057	28169.99	2945.0	3634.653	1985.943	18
##	max	range	skew	kurtosis	se						
##	X11	308545	307483	4.362315	18.585146	3057.4562					
##	X12	2199581	2199503	5.515030	28.919342	18075.9084					
##	X13	192646	192618	3.194207	9.039258	3425.9037					
##	X14	1324924	1324924	8.855583	82.263616	3733.8561					
##	X15	178768	178768	3.238385	11.451435	1365.5891					
##	X16	426247	425115	3.553550	11.065434	7005.1915					
##	X17	304305	304287	7.530961	62.553902	773.5973					

Estatísticas Descritivas de População Urbana Atendida por Esgoto por UF na Região Norte

##	item	group1	vars	n	mean	sd	median	trimmed	mad	min	
##	X11	1	AC	1	33	28309.67	36396.50	5800.0	24844.481	7875.5712	120
##	X12	2	AM	1	50	69936.10	148156.06	5000.0	30543.950	6253.6068	32
##	X13	3	AP	1	60	7545.05	16100.14	898.5	2956.208	891.7839	273
##	X14	4	PA	1	152	27599.92	53657.04	7211.0	13021.303	10298.1396	0
##	X15	5	RO	1	75	10302.88	13845.78	3636.0	7645.951	3908.1336	265
##	X16	6	RR	1	30	82376.57	124393.05	10808.0	56104.667	12099.4986	500
##	X17	7	TO	1	173	21681.87	49681.35	4505.0	9490.540	5581.9890	0
##	max	range	skew	kurtosis	se						
##	X11	87950	87830	0.8101123	-1.3180658	6335.818					
##	X12	574167	574135	2.1392009	3.4583669	20952.432					
##	X13	55276	55003	2.1592801	3.1846189	2078.520					
##	X14	257912	257912	2.8078686	7.1823983	4352.161					
##	X15	49032	48767	1.5488246	0.9368897	1598.773					
##	X16	401944	401444	1.3561519	0.4693252	22710.960					
##	X17	283912	283912	3.7942058	14.3600011	3777.203					

Estatísticas Descritivas de Percentual de População Atendida por Água por UF na Região Norte


```
##      item group1 vars      n      mean      sd      median trimmed      mad
## X11      1      AC      1    220 78.67055 19.415067  81.67062 80.70091 25.146767
## X12      2      AM      1    329 81.01720 23.540296  90.37635 85.10532 14.268019
## X13      3      AP      1    160 36.46713 23.155770  37.27326 34.94268 22.543106
## X14      4      PA      1    958 60.36798 33.074410  61.73607 61.73465 51.115024
## X15      5      RO      1    481 64.16595 28.061067  67.48311 65.84415 37.091538
## X16      6      RR      1    150 98.31056  2.153752  98.99871 98.69593  1.484518
## X17      7      TO      1   1326 97.27045  7.715898 100.00000 99.11067  0.000000
##
##           min      max      range      skew      kurtosis      se
## X11 23.94088080 100.00000  76.05912 -0.6604557 -0.4462140 1.3089636
## X12  0.09779583 100.00000  99.90220 -1.2297016  0.6034685 1.2978185
## X13  0.31749631  91.38269  91.06519  0.3819849 -0.7294886 1.8306243
## X14  0.00000000 100.00000 100.00000 -0.1370193 -1.4960983 1.0685857
## X15  0.00000000 100.00000 100.00000 -0.2896046 -1.1157847 1.2794745
## X16 87.49675970 100.00000  12.50324 -2.6921728 10.2997954 0.1758532
## X17  0.07657945 100.00000  99.92342 -6.0847709 53.2821858 0.2118921
```

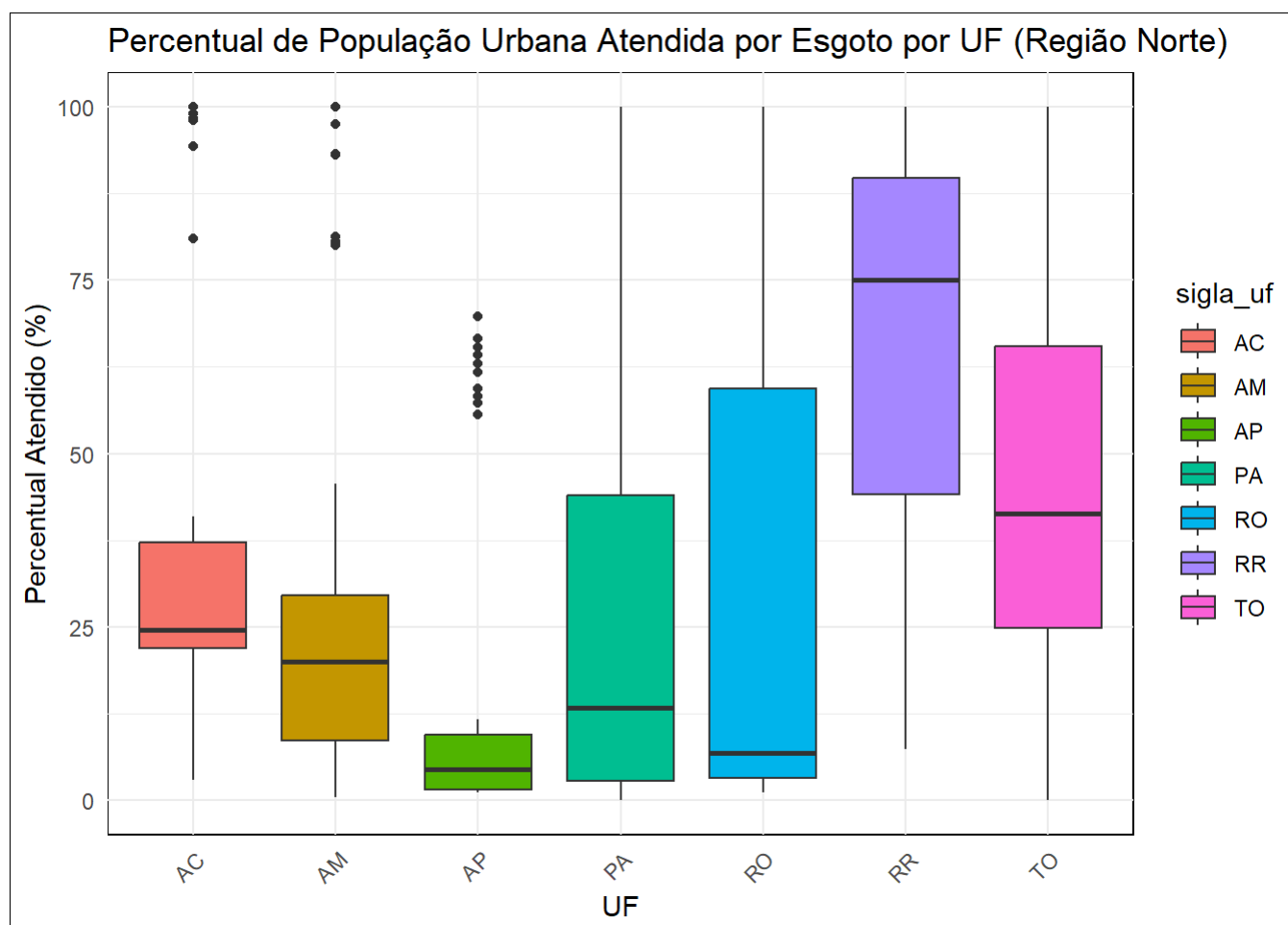
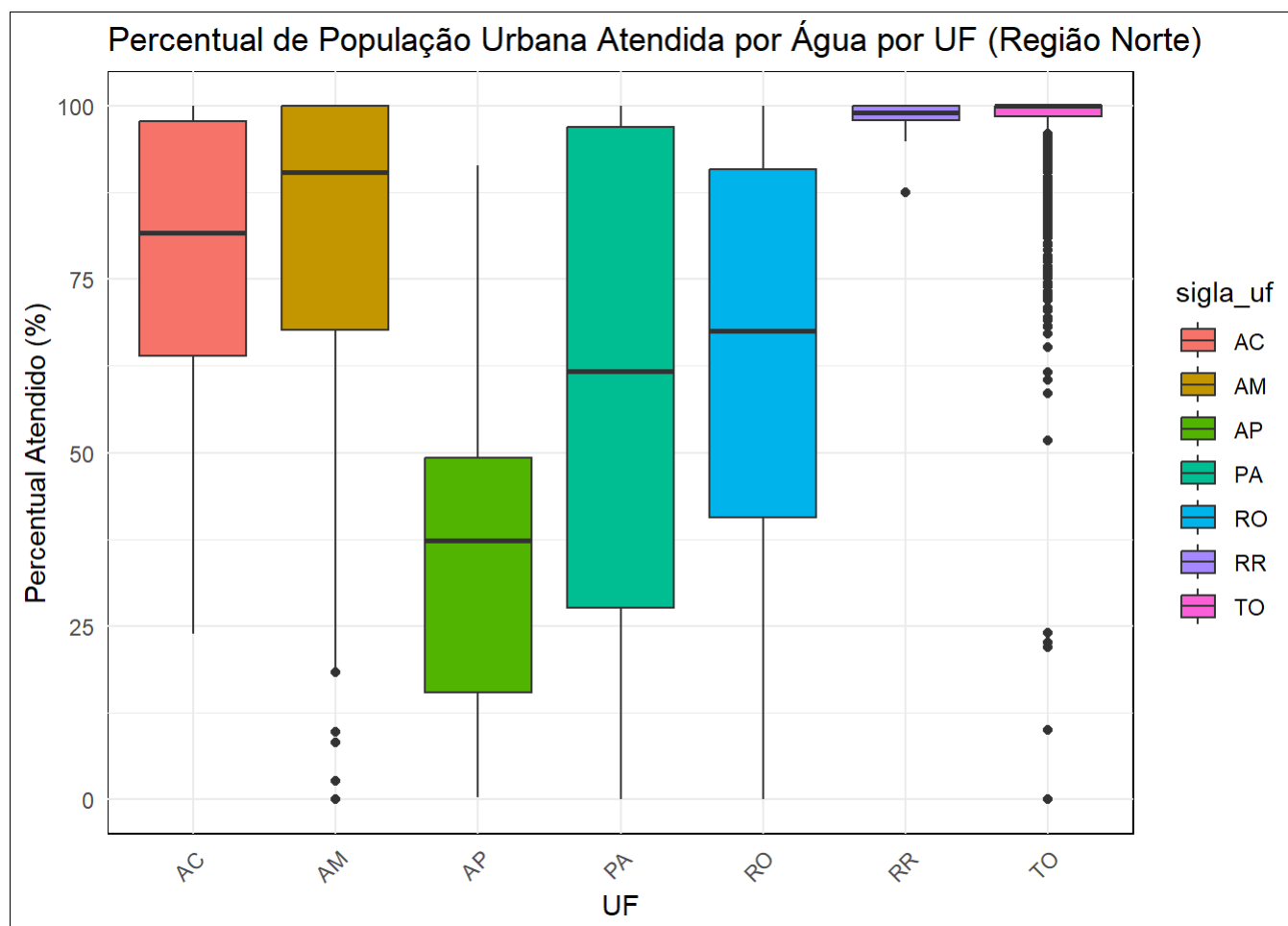
Estatísticas Descritivas de Percentual de População Atendida por Esgoto por UF na Região Norte

```
##      item group1 vars      n      mean      sd      median trimmed      mad
## X11      1      AC      1    33 37.55328 32.77333 24.556113 34.490468 17.709406
## X12      2      AM      1    50 28.23717 29.57180 19.999359 22.997918 16.671509
## X13      3      AP      1    60 13.81692 22.02818  4.504908  8.984548  4.342160
## X14      4      PA      1   152 24.67132 27.79646 13.290677 19.930145 18.326881
## X15      5      RO      1    75 28.06683 31.58305  6.902932 24.306965  7.172603
## X16      6      RR      1    30 67.68916 26.93017 75.047597 70.579162 24.621917
## X17      7      TO      1   173 45.38120 26.54175 41.313778 43.878067 26.589169
##
##           min      max      range      skew      kurtosis      se
## X11 2.8964518 100.00000  97.10355  0.9709232 -0.5019043 5.705105
## X12 0.3942829 100.00000  99.60572  1.4067622  0.6477744 4.182085
## X13 1.1583454  69.72305  68.56471  1.7072808  1.0749550 2.843825
## X14 0.0000000 100.00000 100.00000  1.2218390  0.5300739 2.254591
## X15 1.1298231 100.00000  98.87018  0.7726902 -1.0454927 3.646896
## X16 7.3898906 100.00000  92.61011 -0.7245193 -0.6452336 4.916754
## X17 0.0000000 100.00000 100.00000  0.4419911 -0.8815130 2.017932
```

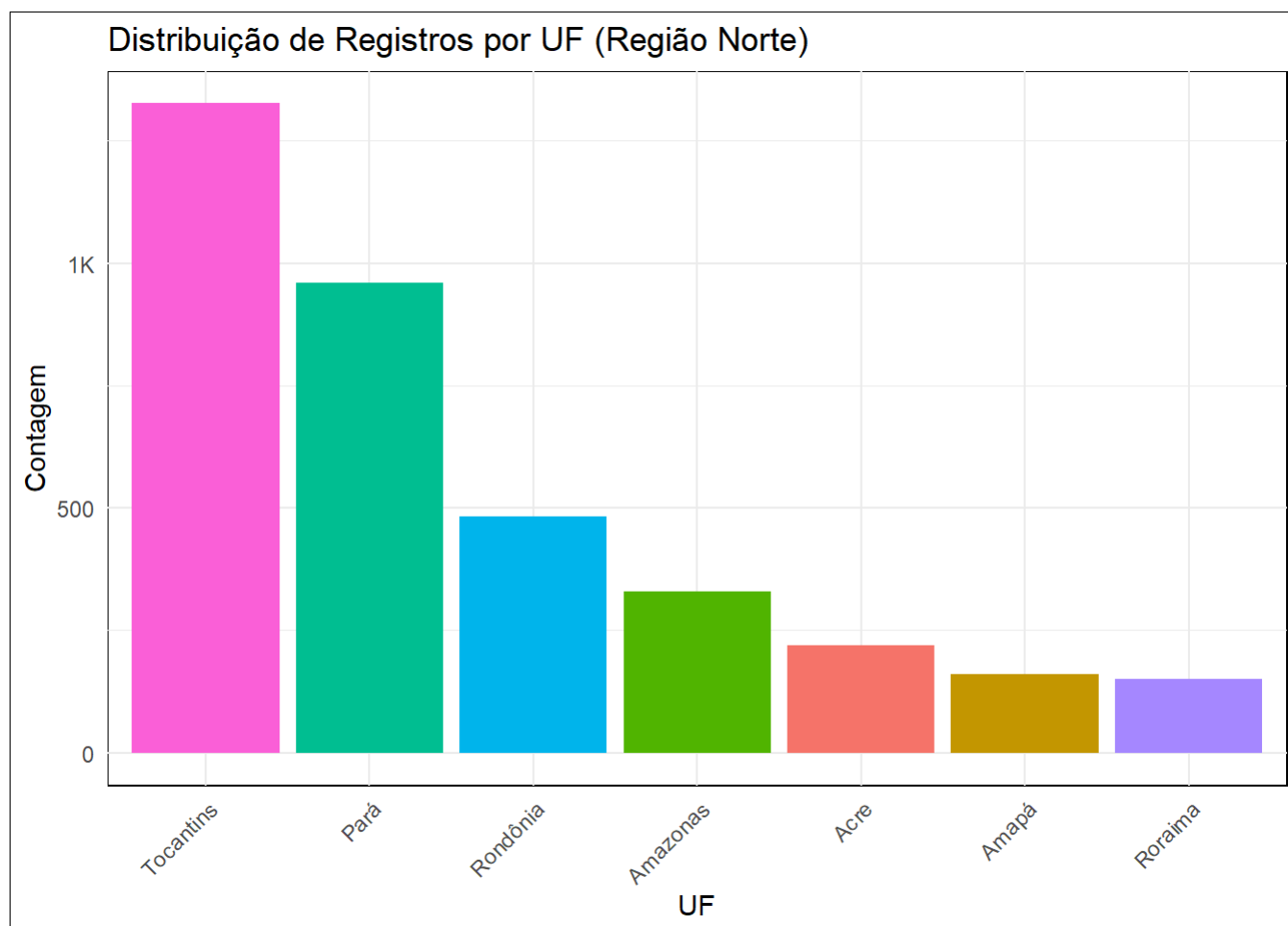
7 Visualizações

A visualização de dados é uma ferramenta poderosa para comunicar insights complexos de forma intuitiva. Nesta seção, serão apresentados gráficos descritivos que ilustram a distribuição e o comportamento das variáveis de interesse para a Região Norte, como percentuais de atendimento e volumes.

Os boxplots de “Percentual de População Urbana Atendida por Água por UF (Região Norte)” e “Percentual de População Urbana Atendida por Esgoto por UF (Região Norte)” ilustram as disparidades significativas de cobertura entre as UFs da região.



Além disso, o gráfico de “Distribuição de Registros por UF (Região Norte)” mostra a contagem de observações para cada UF, com Tocantins e Pará apresentando o maior número de registros no dataset.



8 Detecção de Outliers

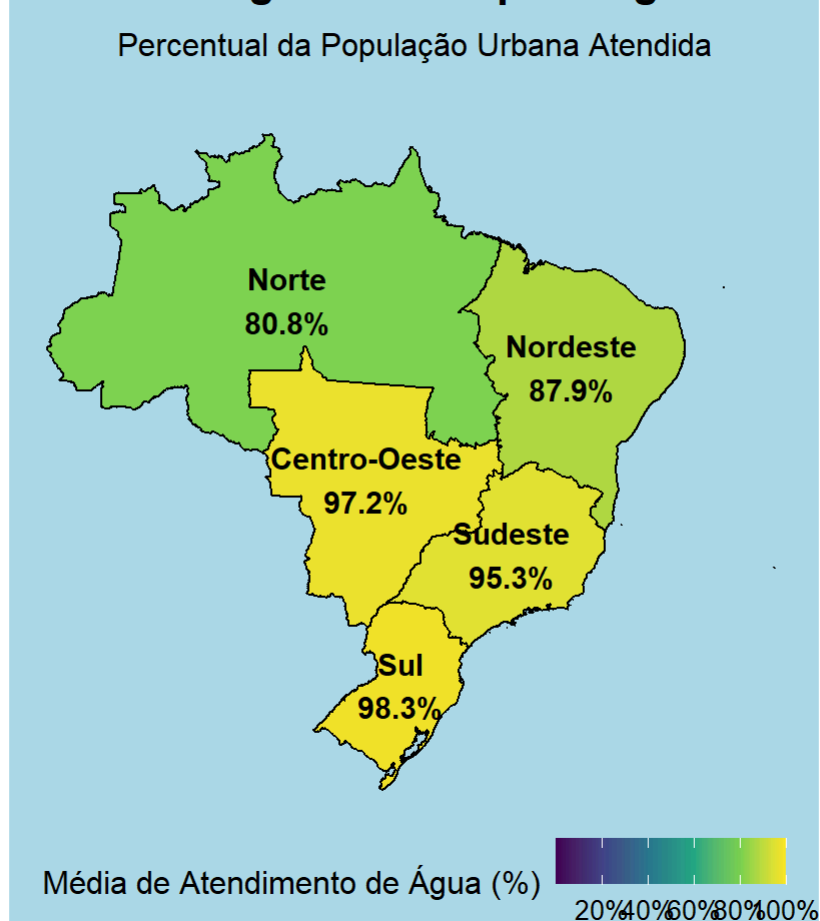
A presença de outliers (observações atípicas) pode influenciar significativamente os resultados das análises estatísticas e a interpretação dos dados. Esta seção descreve o processo de identificação e, quando necessário, tratamento de outliers nas variáveis de população atendida e volumes para a Região Norte, garantindo que as análises subsequentes sejam robustas.

9 Visualizações de Mapas Coropléticos

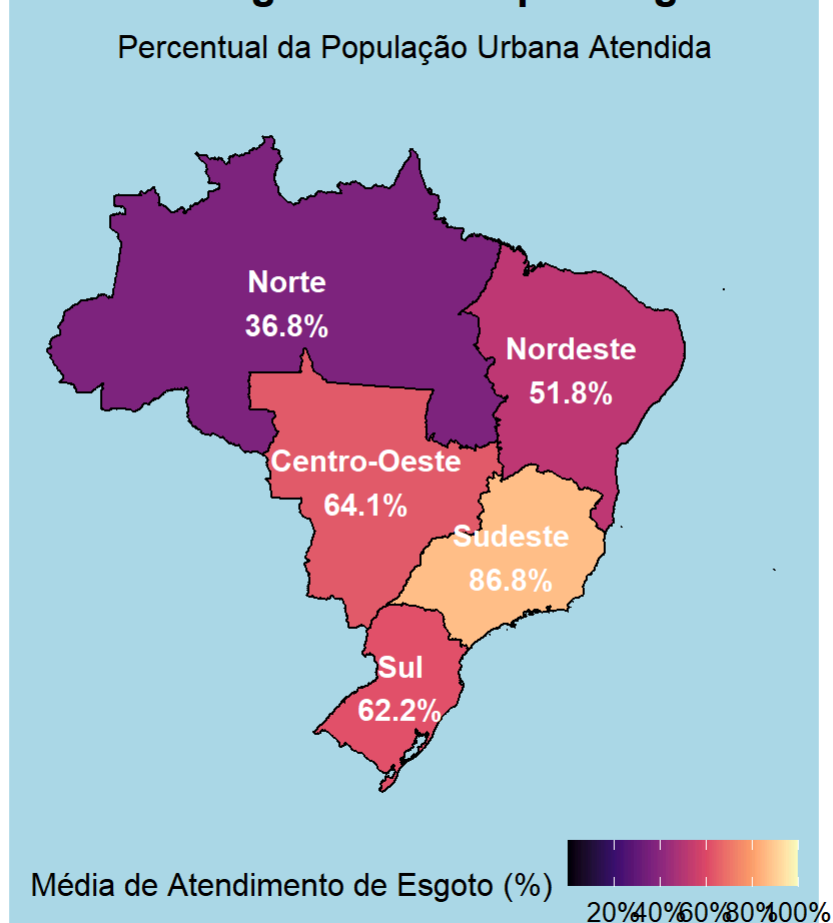
Os mapas coropléticos são essenciais para a visualização de dados georreferenciados, permitindo identificar padrões e disparidades espaciais na cobertura de saneamento. Nesta seção, apresentaremos mapas que ilustram a distribuição do percentual de atendimento de água e esgoto pelas UFs da Região Norte, bem como em nível nacional, destacando as áreas que demandam maior atenção.

Os mapas de “Cobertura Média de Água Urbana por Região - Brasil (2021)” e “Cobertura Média de Esgoto Urbano por Região - Brasil (2021)” fornecem uma visão macro da situação do saneamento no país, mostrando que a Região Norte apresenta as menores médias de atendimento em ambos os serviços.

Cobertura Média de Água Urbana por Região – Brasil (2021)


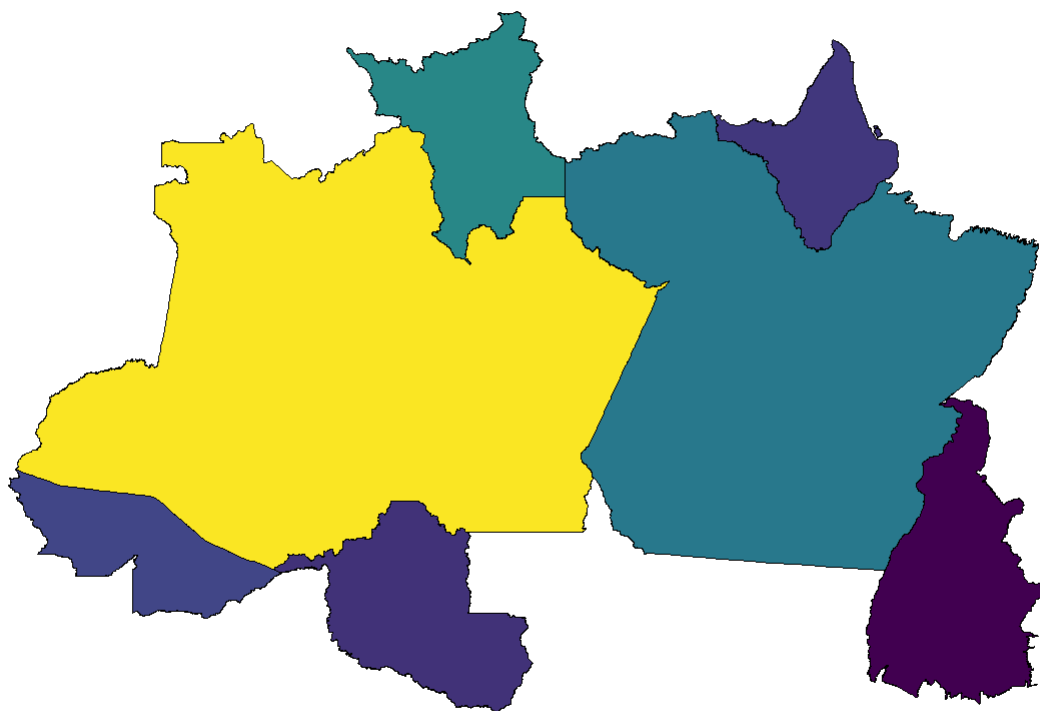


Cobertura Média de Esgoto Urbano por Região – Brasil (2021)

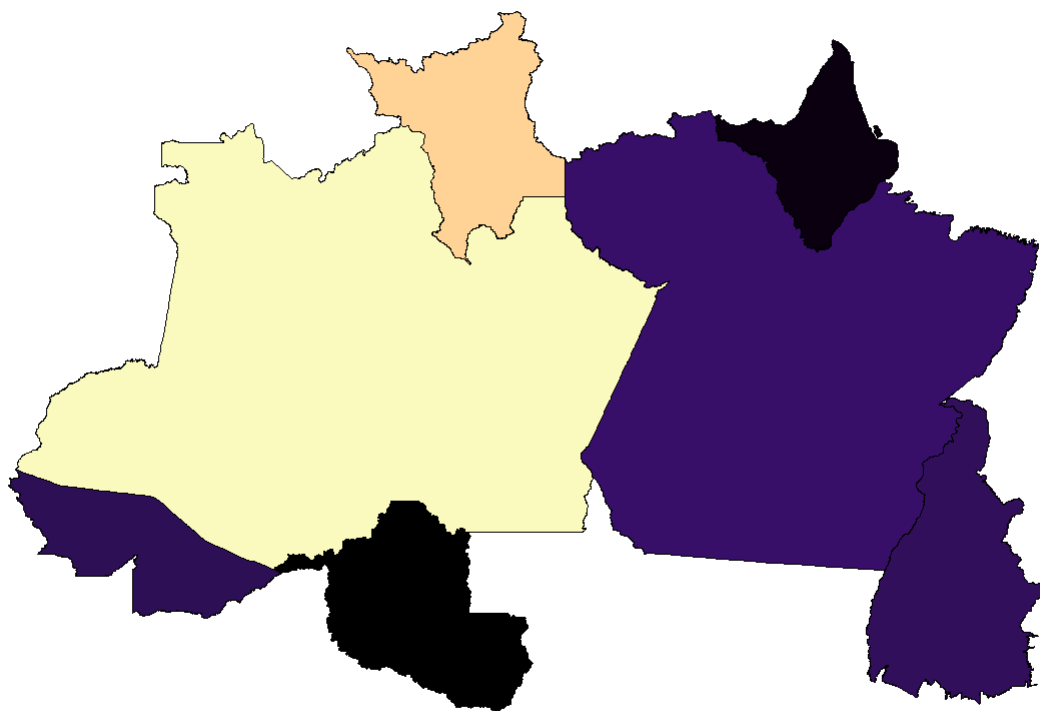


Mapas mais detalhados para a Região Norte, como a “Média de População Urbana Atendida por Água por UF” e “Média de Percentual de População Urbana Atendida por Água por UF”, visualizam as variações dentro da própria região.

Dados do ano mais recente disponível



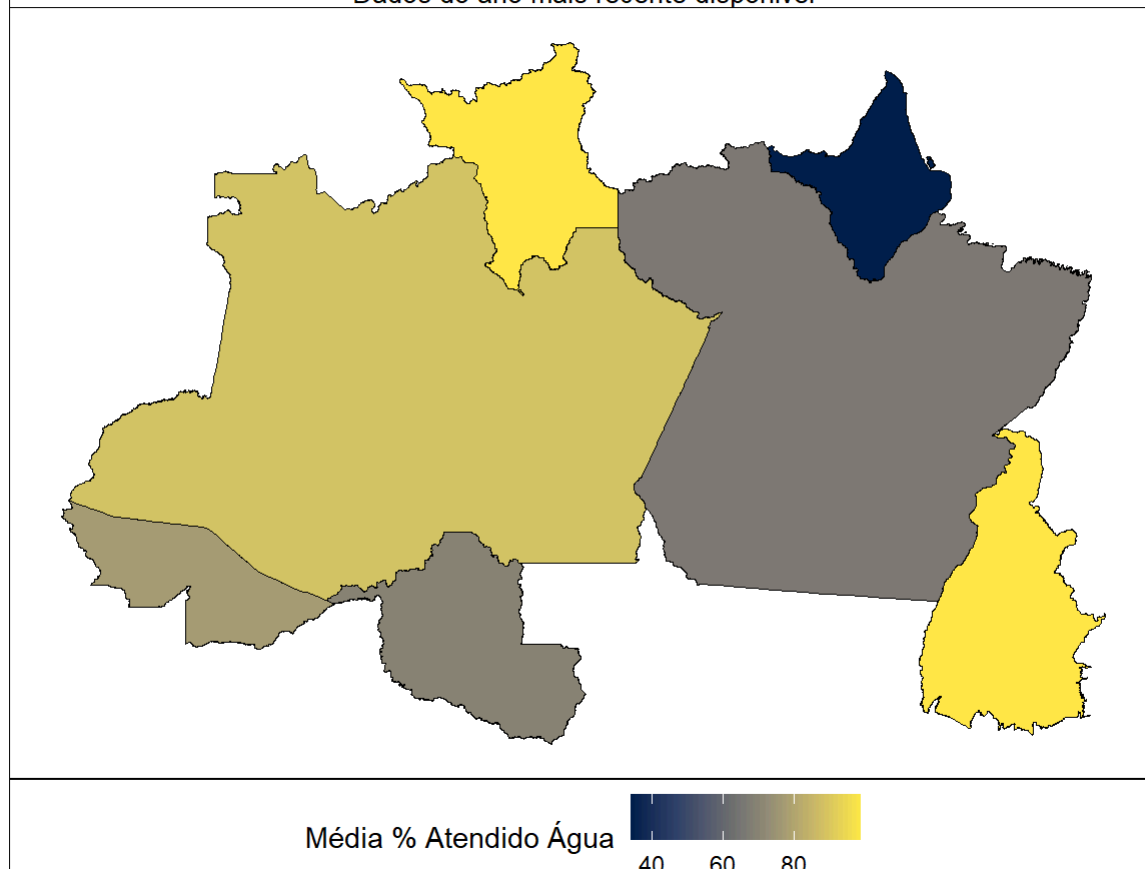
Dados do ano mais recente disponível



30K 60K 90K

Média de Percentual de População Urbana Atendida por Água por UF

Dados do ano mais recente disponível



10 Testes de Hipótese

Para além da análise descritiva e visual, os testes de hipótese são empregados para avaliar a significância estatística das diferenças observadas entre grupos. Nesta seção, realizaremos comparações formais entre as Unidades da Federação da Região Norte em relação às variáveis de saneamento, utilizando t-testes e calculando o Cohen's d para mensurar a magnitude dessas diferenças, fornecendo uma base robusta para as conclusões.

Os testes comparativos entre UFs da Região Norte revelaram diversas diferenças estatisticamente significativas:

```
##
## ### AM x PA (Comparação de Percentual Atendido Água)
## n = 329 / 958
## Percentual Atendido Água   AM = 81.02 IC95% [ 78.46 ; 83.57 ]
## Percentual Atendido Água   PA = 60.37 IC95% [ 58.27 ; 62.47 ]
## Diferença de Percentual Atendido Água IC95%: [ 17.35 ; 23.95 ]
## t = 12.283 p = 0 → REJEITA H0
## F = 0.507 p = 0 → REJEITA H0
## Cohen's d = 0.67 → efeito médio
## **Interpretação**: Existe uma diferença estatisticamente significativa no percentual atendido água entre AM e PA, com AM apresentando um valor médio maior. O tamanho do efeito é médio (Cohen's d = 0.67), indicando que a diferença prática é médio. As variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### AM x PA (Comparação de Volume de Água Produzido)
## n = 329 / 959
## Volume de Água Produzido AM = 8260.14 IC95% [ 4364.3 ; 12155.98 ]
## Volume de Água Produzido PA = 3056.21 IC95% [ 2405.69 ; 3706.73 ]
## Diferença de Volume de Água Produzido IC95%: [ 1254.68 ; 9153.19 ]
## t = 2.592 p = 0.01 → REJEITA H0
## F = 12.245 p = 0 → REJEITA H0
## Cohen's d = 0.26 → efeito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume
de água produzido entre AM e PA, com AM apresentando um valor médio maior. O tamanho d
o efeito é pequeno (Cohen's d = 0.26), indicando que a diferença prática é pequeno. As
variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### AM x RO (Comparação de Percentual Atendido Água)
## n = 329 / 481
## Percentual Atendido Água AM = 81.02 IC95% [ 78.46 ; 83.57 ]
## Percentual Atendido Água RO = 64.17 IC95% [ 61.65 ; 66.68 ]
## Diferença de Percentual Atendido Água IC95%: [ 13.27 ; 20.43 ]
## t = 9.246 p = 0 → REJEITA H0
## F = 0.704 p = 6e-04 → REJEITA H0
## Cohen's d = 0.64 → efeito médio
## **Interpretação**: Existe uma diferença estatisticamente significativa no percent
ual atendido água entre AM e RO, com AM apresentando um valor médio maior. O tamanho d
o efeito é médio (Cohen's d = 0.64), indicando que a diferença prática é médio. As var
iâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### AM x RO (Comparação de Volume de Água Produzido)
## n = 329 / 479
## Volume de Água Produzido AM = 8260.14 IC95% [ 4364.3 ; 12155.98 ]
## Volume de Água Produzido RO = 2079.98 IC95% [ 1573.05 ; 2586.91 ]
## Diferença de Volume de Água Produzido IC95%: [ 2251.88 ; 10108.44 ]
## t = 3.095 p = 0.0021 → REJEITA H0
## F = 40.473 p = 0 → REJEITA H0
## Cohen's d = 0.26 → efeito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume
de água produzido entre AM e RO, com AM apresentando um valor médio maior. O tamanho d
o efeito é pequeno (Cohen's d = 0.26), indicando que a diferença prática é pequeno. As
variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### AM x TO (Comparação de Percentual Atendido Água)
## n = 329 / 1326
## Percentual Atendido Água AM = 81.02 IC95% [ 78.46 ; 83.57 ]
## Percentual Atendido Água TO = 97.27 IC95% [ 96.85 ; 97.69 ]
## Diferença de Percentual Atendido Água IC95%: [ -18.84 ; -13.67 ]
## t = -12.36 p = 0 → REJEITA H0
## F = 9.308 p = 0 → REJEITA H0
## Cohen's d = -1.29 → efeito grande
## **Interpretação**: Existe uma diferença estatisticamente significativa no percentual atendido água entre AM e TO, com TO apresentando um valor médio maior. O tamanho do efeito é grande (Cohen's d = -1.29), indicando que a diferença prática é grande. As variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### AM x TO (Comparação de Volume de Água Produzido)
## n = 329 / 1326
## Volume de Água Produzido AM = 8260.14 IC95% [ 4364.3 ; 12155.98 ]
## Volume de Água Produzido TO = 738.97 IC95% [ 606.22 ; 871.73 ]
## Diferença de Volume de Água Produzido IC95%: [ 3623.09 ; 11419.25 ]
## t = 3.796 p = 2e-04 → REJEITA H0
## F = 212.486 p = 0 → REJEITA H0
## Cohen's d = 0.47 → efeito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume de água produzido entre AM e TO, com AM apresentando um valor médio maior. O tamanho do efeito é pequeno (Cohen's d = 0.47), indicando que a diferença prática é pequeno. As variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### PA x RO (Comparação de Percentual Atendido Água)
## n = 958 / 481
## Percentual Atendido Água PA = 60.37 IC95% [ 58.27 ; 62.47 ]
## Percentual Atendido Água RO = 64.17 IC95% [ 61.65 ; 66.68 ]
## Diferença de Percentual Atendido Água IC95%: [ -7.07 ; -0.53 ]
## t = -2.278 p = 0.0229 → REJEITA H0
## F = 1.389 p = 0 → REJEITA H0
## Cohen's d = -0.12 → efeito muito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no percentual atendido água entre PA e RO, com RO apresentando um valor médio maior. O tamanho do efeito é muito pequeno (Cohen's d = -0.12), indicando que a diferença prática é muito pequeno. As variâncias dos dois grupos também são significativamente diferentes.
```



```
##
## ### PA x RO (Comparação de Volume de Água Produzido)
## n = 959 / 479
## Volume de Água Produzido PA = 3056.21 IC95% [ 2405.69 ; 3706.73 ]
## Volume de Água Produzido RO = 2079.98 IC95% [ 1573.05 ; 2586.91 ]
## Diferença de Volume de Água Produzido IC95%: [ 152.25 ; 1800.2 ]
## t = 2.324 p = 0.0203 → REJEITA H0
## F = 3.305 p = 0 → REJEITA H0
## Cohen's d = 0.11 → efeito muito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume
de água produzido entre PA e RO, com PA apresentando um valor médio maior. O tamanho d
o efeito é muito pequeno (Cohen's d = 0.11), indicando que a diferença prática é muito
pequeno. As variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### PA x TO (Comparação de Percentual Atendido Água)
## n = 958 / 1326
## Percentual Atendido Água PA = 60.37 IC95% [ 58.27 ; 62.47 ]
## Percentual Atendido Água TO = 97.27 IC95% [ 96.85 ; 97.69 ]
## Diferença de Percentual Atendido Água IC95%: [ -39.04 ; -34.76 ]
## t = -33.874 p = 0 → REJEITA H0
## F = 18.374 p = 0 → REJEITA H0
## Cohen's d = -1.66 → efeito grande
## **Interpretação**: Existe uma diferença estatisticamente significativa no percent
ual atendido água entre PA e TO, com TO apresentando um valor médio maior. O tamanho d
o efeito é grande (Cohen's d = -1.66), indicando que a diferença prática é grande. As
variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### PA x TO (Comparação de Volume de Água Produzido)
## n = 959 / 1326
## Volume de Água Produzido PA = 3056.21 IC95% [ 2405.69 ; 3706.73 ]
## Volume de Água Produzido TO = 738.97 IC95% [ 606.22 ; 871.73 ]
## Diferença de Volume de Água Produzido IC95%: [ 1653.37 ; 2981.11 ]
## t = 6.849 p = 0 → REJEITA H0
## F = 17.353 p = 0 → REJEITA H0
## Cohen's d = 0.34 → efeito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume
de água produzido entre PA e TO, com PA apresentando um valor médio maior. O tamanho d
o efeito é pequeno (Cohen's d = 0.34), indicando que a diferença prática é pequeno. As
variâncias dos dois grupos também são significativamente diferentes.
```

```
##
## ### RO x TO (Comparação de Percentual Atendido Água)
## n = 481 / 1326
## Percentual Atendido Água RO = 64.17 IC95% [ 61.65 ; 66.68 ]
## Percentual Atendido Água TO = 97.27 IC95% [ 96.85 ; 97.69 ]
## Diferença de Percentual Atendido Água IC95%: [ -35.65 ; -30.56 ]
## t = -25.526 p = 0 → REJEITA H0
## F = 13.226 p = 0 → REJEITA H0
## Cohen's d = -2.08 → efeito grande
## **Interpretação**: Existe uma diferença estatisticamente significativa no percentual atendido água entre RO e TO, com TO apresentando um valor médio maior. O tamanho do efeito é grande (Cohen's d = -2.08), indicando que a diferença prática é grande. As variâncias dos dois grupos também são significativamente diferentes.
```

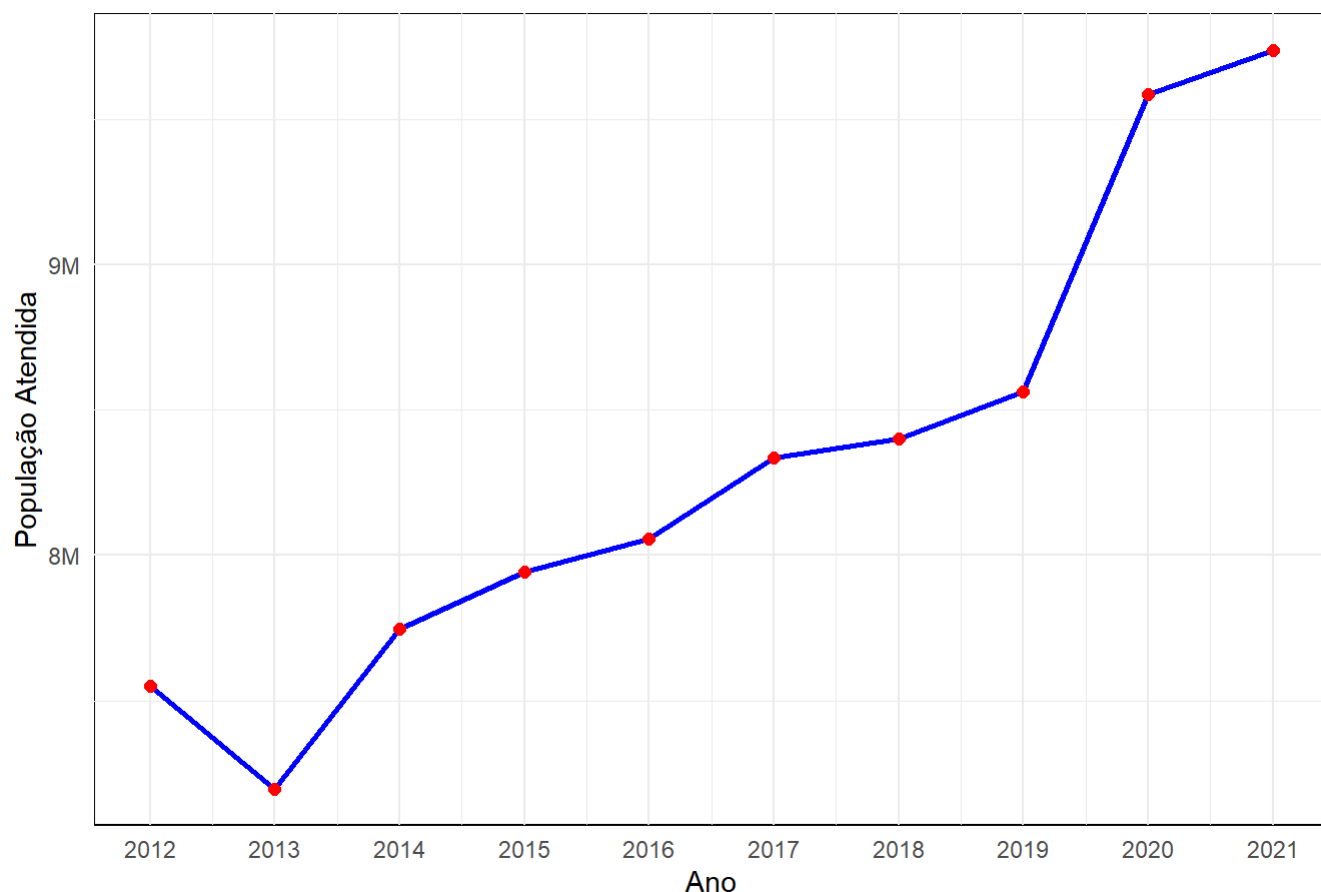
```
##
## ### RO x TO (Comparação de Volume de Água Produzido)
## n = 479 / 1326
## Volume de Água Produzido RO = 2079.98 IC95% [ 1573.05 ; 2586.91 ]
## Volume de Água Produzido TO = 738.97 IC95% [ 606.22 ; 871.73 ]
## Diferença de Volume de Água Produzido IC95%: [ 817.09 ; 1864.92 ]
## t = 5.028 p = 0 → REJEITA H0
## F = 5.25 p = 0 → REJEITA H0
## Cohen's d = 0.37 → efeito pequeno
## **Interpretação**: Existe uma diferença estatisticamente significativa no volume de água produzido entre RO e TO, com RO apresentando um valor médio maior. O tamanho do efeito é pequeno (Cohen's d = 0.37), indicando que a diferença prática é pequeno. As variâncias dos dois grupos também são significativamente diferentes.
```

11 Gráficos adicionais

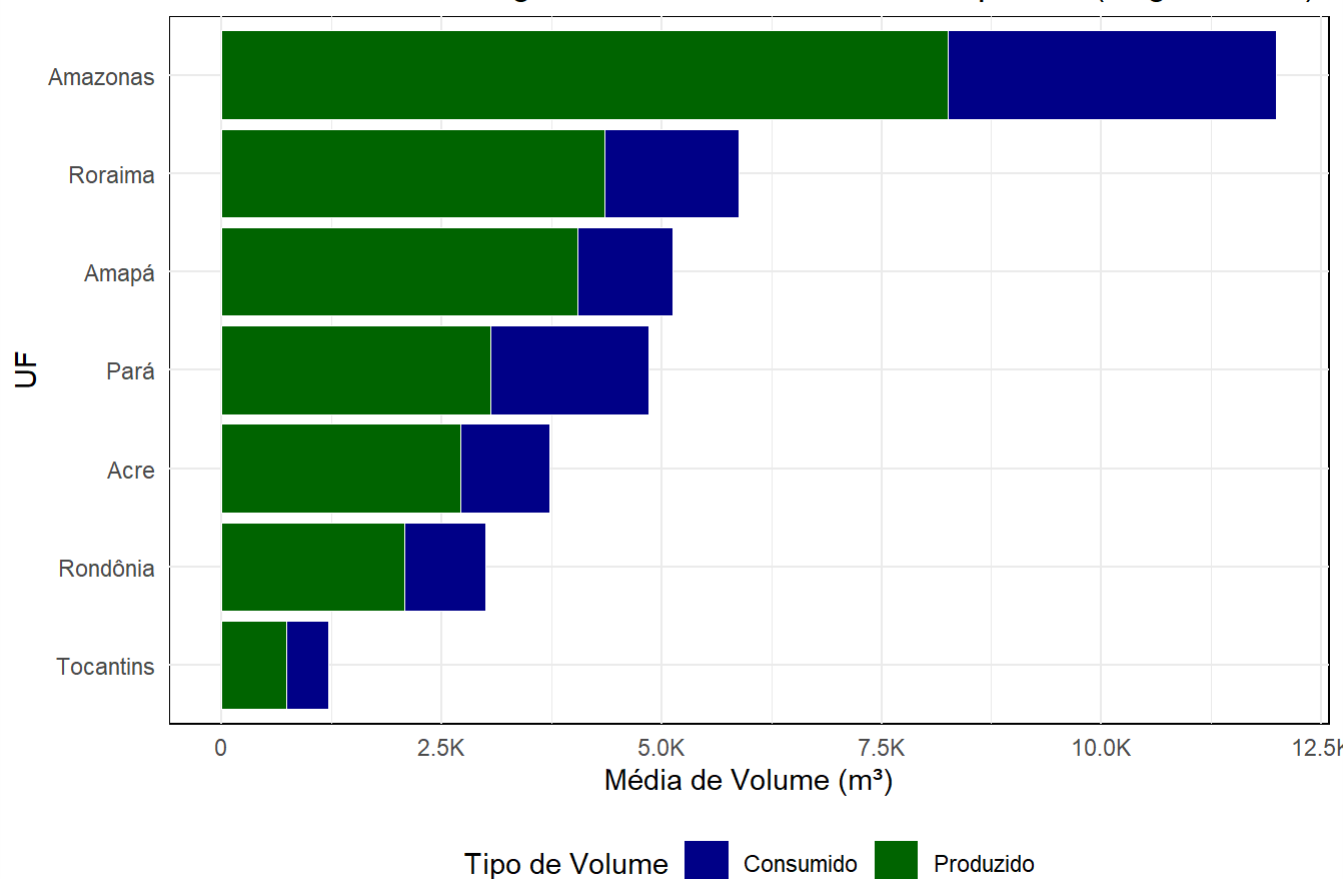
Complementando as análises anteriores, esta seção apresenta gráficos adicionais que oferecem perspectivas complementares sobre os dados de saneamento na Região Norte. Estes gráficos podem incluir visualizações de séries temporais, distribuições específicas ou outras agregações que enriquecem a compreensão do cenário.

Gráficos como a “Distribuição da População Urbana Atendida por Água por Ano (Região Norte)” mostram a evolução do atendimento ao longo do tempo. A “Média de Volume de Água Produzido vs. Consumido por UF (Região Norte)” detalha a relação entre produção e consumo em cada UF.

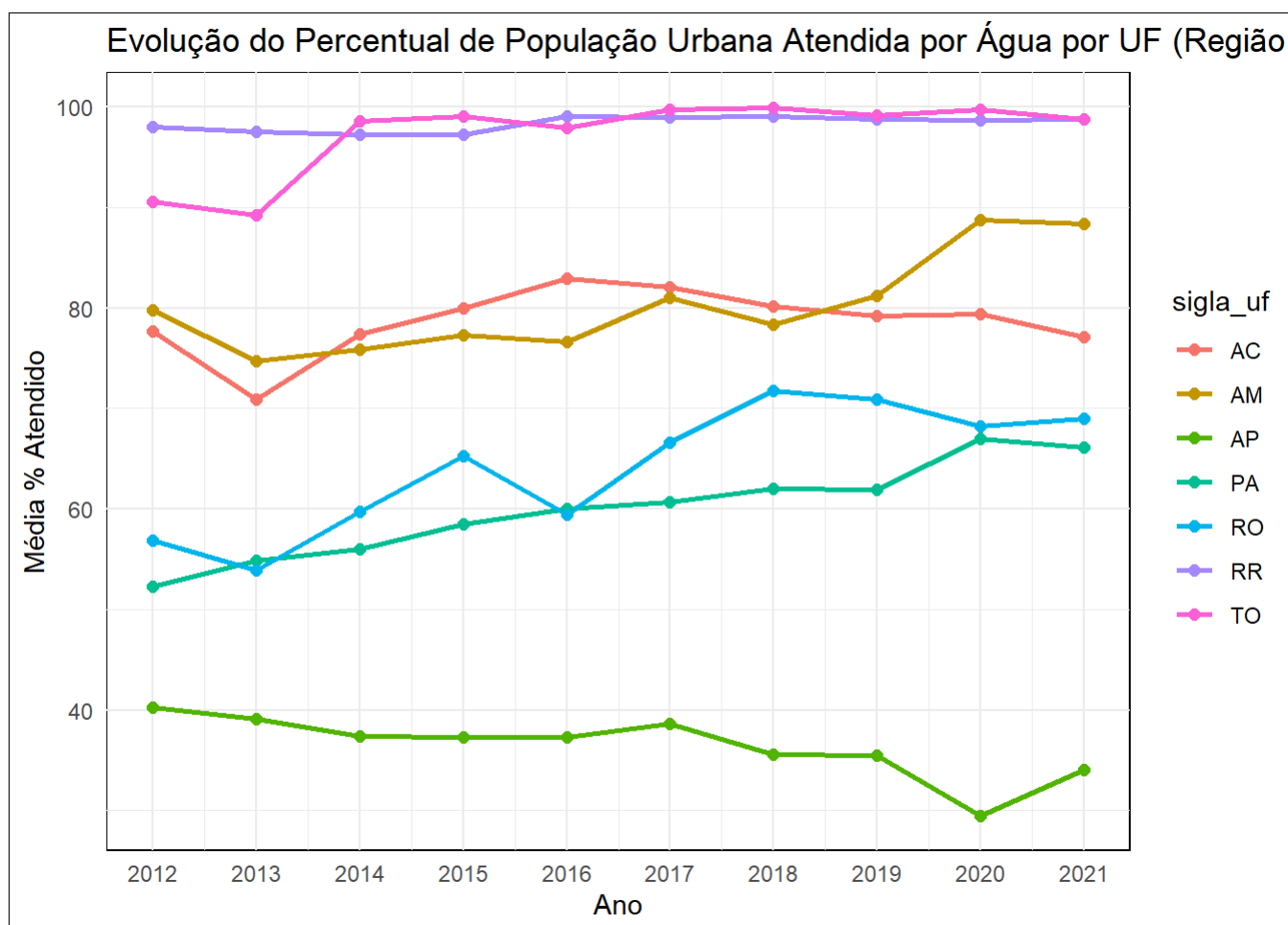
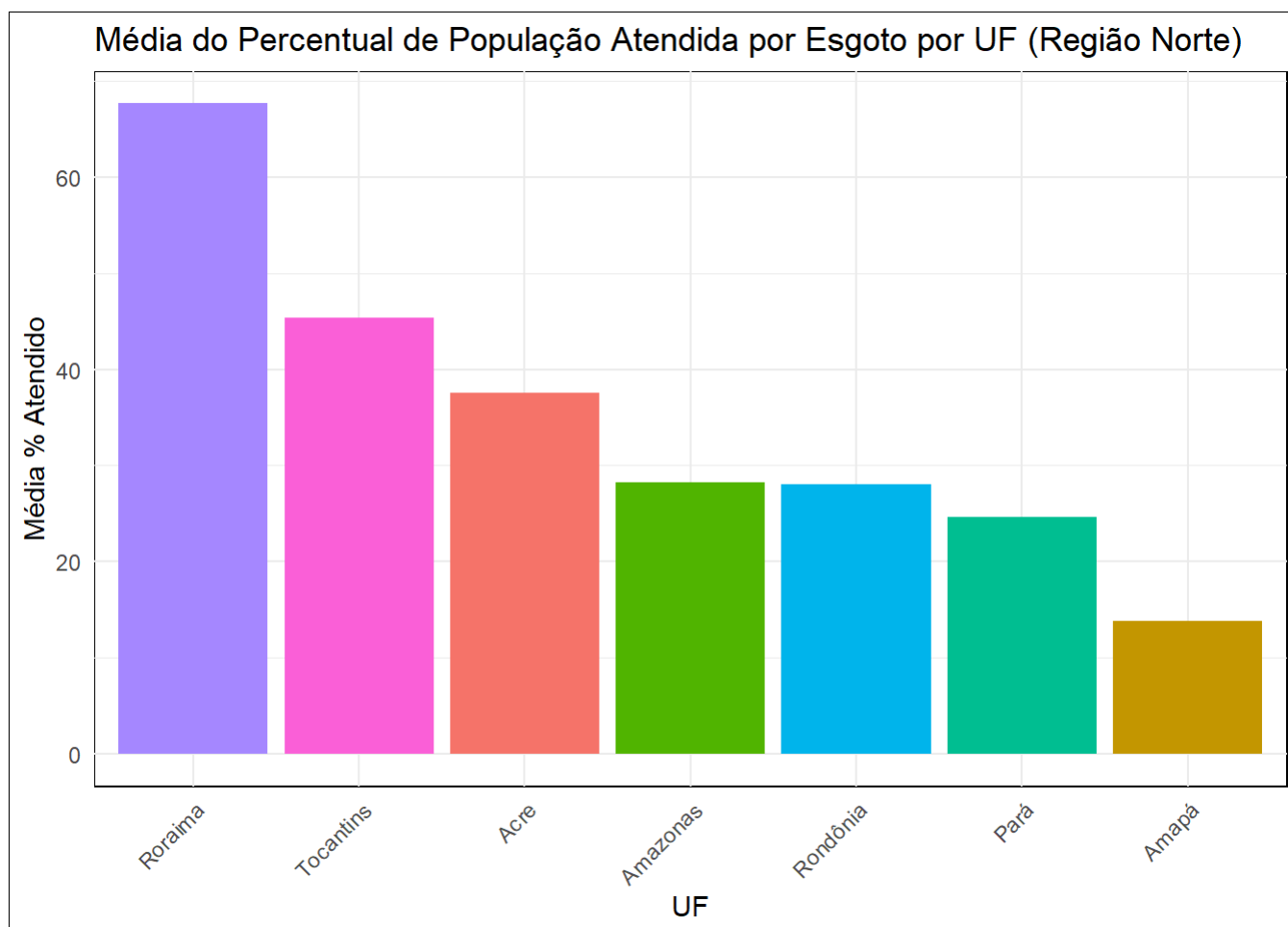
Distribuição da População Urbana Atendida por Água por Ano (Região Norte)



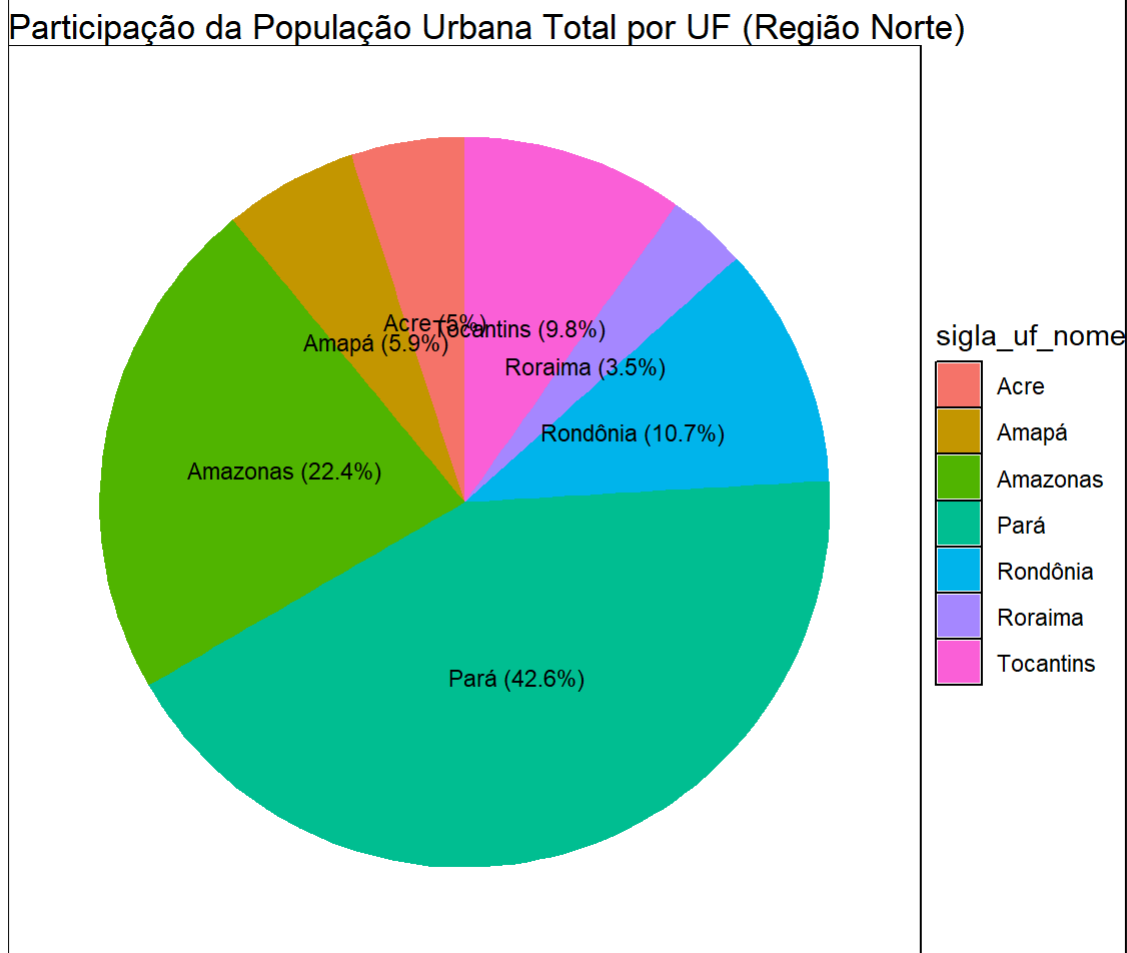
Média de Volume de Água Produzido vs. Consumido por UF (Região Norte)



Além disso, a “Média do Percentual de População Atendida por Esgoto por UF (Região Norte)” e a “Evolução do Percentual de População Urbana Atendida por Água por UF” fornecem insights adicionais sobre o panorama do saneamento na região.



O gráfico de “Participação da População Urbana Total por UF (Região Norte)” ilustra a proporção da população urbana de cada UF na região, destacando Pará (42.6%) e Amazonas (22.4%) como as UFs com maior participação.



12 Conclusão

Este relatório apresentou uma análise abrangente sobre a cobertura e gestão dos serviços de água e esgoto na Região Norte do Brasil, focando em dados do período de 2012-2014. Através de uma metodologia que integrou análise exploratória, visualizações geográficas e testes de hipótese, foi possível desvendar as complexidades e os desafios enfrentados pela região.

As principais observações e aprendizados da análise são os seguintes:

- Dados Ausentes:** A presença de dados ausentes (NAs) em variáveis críticas como 'populacao_urbana_atendida_agua', 'populacao_urbana_atendida_esgoto', 'volume_agua_produzido', 'volume_agua_consumido' e 'populacao_urbana' foi uma constante. Esta ausência de reportes tem um impacto significativo nas estatísticas e visualizações, especialmente se a Região Norte apresentar uma alta proporção de valores não informados. A interpretação de dados não reportados como ausência de serviço ou volume exige cautela, e sugere a necessidade de aprimoramento na coleta e disponibilização de dados para uma representação mais fiel da realidade.
- Disparidades na Cobertura de Serviços por UF (Região Norte):**
 - As análises visuais (boxplots e mapas coropléticos) demonstraram inequivocamente as **significativas disparidades na cobertura de saneamento básico entre as Unidades da Federação da própria Região Norte**.
 - Para o **abastecimento de água**, observou-se que a Região Norte, com uma média de 80.8% em 2021, possui a menor cobertura nacional. Internamente, UFs como Amapá (AP) e Pará (PA) apresentaram medianas de atendimento substancialmente mais baixas, indicando maiores desafios. Em contraste,

Roraima (RR) e Tocantins (TO) mostraram coberturas mais elevadas, aproximando-se da universalização em muitos municípios.

- Para o **esgotamento sanitário**, a situação é ainda mais crítica, com a Região Norte registrando a menor cobertura nacional, com uma média de 36.8% em 2021. Os boxplots revelaram que Amapá (AP), Amazonas (AM) e Rondônia (RO) possuem as menores medianas de percentual atendido, com vastas áreas urbanas sem cobertura de esgoto. Roraima (RR) e Tocantins (TO), embora apresentem medianas um pouco superiores, ainda refletem a necessidade urgente de expansão.
- Essas disparidades sublinham a importância de políticas públicas e investimentos focados e diferenciados, direcionados especificamente para as UF's mais deficitárias dentro da região.

3. Análise de Volumes Produzidos e Consumidos: Indicadores de Eficiência:

- A análise dos volumes de água produzida e consumida por UF forneceu insights cruciais sobre a gestão e eficiência dos sistemas.
- Observou-se que o Amazonas (AM) é a UF com a maior média de volume de água produzido, refletindo uma infraestrutura de captação e tratamento de grande escala. Contudo, a comparação entre o volume produzido e o volume consumido em todas as UF's revelou uma **discrepância notável**, com o volume produzido sendo consistentemente maior que o consumido.
- Essa diferença acentuada, particularmente visível no Amazonas, é um forte indicativo de **altas perdas na distribuição de água** (perdas físicas e não-físicas), um problema crítico que afeta a sustentabilidade dos serviços e gera ineficiência. A redução dessas perdas deve ser uma prioridade para otimizar o uso da água e os recursos investidos.
- Por outro lado, Tocantins (TO) apresentou os menores volumes absolutos de produção e consumo, sugerindo uma infraestrutura de menor escala ou menor demanda total em seus municípios.

4. Impacto e Interpretação dos Outliers:

- A detecção de outliers em variáveis como 'população atendida' e 'volumes' na Região Norte é um achado significativo. Essas observações extremas, que se afastam consideravelmente da maioria dos dados, podem representar realidades de grandes capitais ou municípios com infraestrutura de saneamento muito mais desenvolvida e volumes operacionais muito superiores à média regional.
- A presença desses outliers é um lembrete de que a Região Norte, apesar de suas médias mais baixas em saneamento, possui "bolsões" de excelência ou grande escala. A interpretação deve considerar que, embora atípicos estatisticamente, esses pontos são representativos de grandes operações e não devem ser simplesmente descartados, mas entendidos como parte da heterogeneidade da região.

5. Resultados dos Testes de Hipótese e Implicações Práticas:

- Os testes de hipótese realizados entre as Unidades da Federação da Região Norte confirmaram, com alto grau de confiança estatística ($p < 0.05$), a existência de diferenças significativas tanto na cobertura de abastecimento de água quanto no volume de água produzido.
- Além da significância estatística, os valores de Cohen's d revelaram que muitas dessas diferenças possuem também relevância prática considerável. Por exemplo, o comparativo entre TO e RO no percentual de atendimento de água revelou um efeito grande (Cohen's $d = -2.08$), reforçando a disparidade entre os estados. Já para os volumes de água produzidos, a maioria das comparações indicou efeitos pequenos a muito pequenos (Cohen's d entre 0.11 e 0.47), sugerindo que, apesar das diferenças médias, o impacto prático tende a ser mais discreto. Esses achados quantitativos complementam as análises visuais e reforçam a necessidade de intervenções específicas para reduzir desigualdades estruturais na oferta e eficiência dos serviços de saneamento na região.

Em suma, a análise ressalta que a Região Norte, apesar de seus avanços, ainda enfrenta consideráveis desafios no saneamento básico, especialmente no que tange ao esgotamento sanitário e à eficiência na distribuição de água. As disparidades entre as UFs são notáveis, exigindo abordagens personalizadas para alcançar a meta de universalização dos serviços e promover a saúde e o desenvolvimento sustentável em toda a região.