

Análise de microdados da PNAD COVID19 na capital Salvador - 06/2020

Emanuel Carneiro, Gabriela Lopes, Luíza Beatriz

INTRODUÇÃO

O trabalho tem como objetivo apresentar uma análise exploratória dos dados da PNAD COVID, referentes ao mês junho de 2020. Como base usamos os dados produzidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE)[@pnad2020]., cujo o intuito de captar os impactos da pandemia de COVID-19 sobre diversos aspectos da vida da população brasileira.

O trabalho foi proposto como parte da conclusão da disciplina de Introdução ao R, tendo como objetivos a consolidação dos conhecimentos adquiridos ao longo do semestre, assim como o desenvolvimento de habilidades fundamentais em manipulação, visualização e interpretação de dados utilizando a linguagem R.

Nele será apresentado as principais etapas da análise, incluindo a preparação dos dados, a seleção das variáveis, a elaboração de gráficos e tabelas, bem como uma discussão dos principais achados.

As análises foram direcionadas a temas que despertaram a curiosidade do grupo, com o intuito de promover reflexões relevantes sobre o contexto da pandemia. Importante destacar que o recorte espacial adotado se limita ao município de Salvador (Bahia), permitindo uma investigação mais aprofundada sobre as características locais da população durante o período analisado.

PREPARAÇÃO DOS DADOS

Para o início da preparação dos dados, foram carregados os pacotes essenciais para análise de dados e realizada a importação direta dos microdados da PNAD COVID, referentes ao mês de junho de 2020, por meio da função `get_covid()` do pacote `COVIDIBGE`.

Selecionamos as variáveis relacionadas aos aspectos sociodemográficos (sexo, idade, raça, escolaridade), mercado de trabalho (trabalho remoto, vínculo empregatício, ocupação principal),

saúde (sintomas de COVID-19, busca por atendimento), e domicílio (valor do aluguel, faixa salarial).

```
#selecionando as variaveis

dados_junho <- dados_junho |>
  select(
    UF,CAPITAL,V1032,A002,A003,A004,
    A005,C013,D0051,B0011,B0012,B0013,B0014,B0015,B0016,B0017,B0018,B0019,
    B00110,B00111,B00112,B002,B0031,B0032,B0033,B0034,B0035,B0036,B0037,
    F001, F0022, C01012
  )
```

Afim de tornar o código mais intuitivo, todas as variáveis selecionadas foram renomeadas com termos mais descritivos. Isso facilita o entendimento e a documentação da análise.

```
#renomeando as variaveis selecionadas
dados_junho <- dados_junho |>
  rename(
    estado          = UF,
    capital          = CAPITAL,
    peso            = V1032,
    idade           = A002,
    sexo            = A003,
    raça            = A004,
    escolaridade     = A005,
    trabalho        = C013,
    auxilio_emergencial = D0051,
    febre            = B0011,
    tosse            = B0012,
    dor_garganta     = B0013,
    dificuldade_respiratoria = B0014,
    dor_cabeça       = B0015,
    dor_peito        = B0016,
    nauseas          = B0017,
    nariz_entupido   = B0018,
    fadiga           = B0019,
    dor_olhos        = B00110,
    perda_paladar_ofato = B00111,
    dor_muscular     = B00112,
    procurou_ajuda   = B002,
    ficou_em_casa    = B0031,
```

```

ligou_profissional      = B0032,
automedicação          = B0033,
medicação_orientação_medica = B0034,
visita_sus             = B0035,
visita_particular      = B0036,
outra_providencia      = B0037,
domicilio             = F001,
valor_aluguel          = F0022,
faixa_salarial         = C01012

```

```
)
```

Filtrando somente a capital de Salvador.

```

dados_junho_salvador <- dados_junho |>
  filter(capital == "Município de Salvador (BA)")

```

Também foram realizadas novas recodificações em variáveis categóricas e numéricas, com o objetivo de facilitar a interpretação dos dados e melhorar a visualização nas análises. EX.: Recodificamos a variável `trabalho` para distinguir entre o trabalho “*presencial*” e o “*home office*”.

```

#Recodificação das variaveis
dados_junho_salvador <- dados_junho_salvador |>
  mutate(valor_aluguel = as.factor(valor_aluguel),
         trabalho = fct_recode(trabalho,
                               "Home Office" = "Sim",
                               "Presencial" = "Não"),
         valor_aluguel = fct_recode(valor_aluguel,
                                    "1 - 100" = "0",
                                    "101 - 300" = "1",
                                    "301 - 600" = "2",
                                    "601 - 800" = "3",
                                    "801 - 1.600" = "4",
                                    "1.601 - 3.000" = "5",
                                    "3.001 - 10.000" = "6",
                                    "10.001 - 50.000" = "7",))

```

Tabela 1 – Variáveis demográficas

Variável	Código	Tipo
estado	UF	qualitativa nominal
capital	CAPITAL	qualitativa nominal
peso	V1032	quantitativa contínua
idade	A002	quantitativa contínua
sexo	A003	qualitativa nominal
raça	A004	qualitativa nominal
escolaridade	A005	qualitativa ordinal
trabalho	C013	qualitativa nominal
auxilio_emergencial	D0051	quantitativa continua
valor_aluguel	F0022	quantitativa continua
faixa_salarial	C01012	quantitativa continua

Tabela 2 – Variáveis de Sintomas Relacionados

Variável	Código	Tipo
febre	B0011	qualitativa nominal
tosse	B0012	qualitativa nominal
dor de garganta	B0013	qualitativa nominal
dificuldade respiratória	B0014	qualitativa nominal
dor de cabeça	B0015	qualitativa nominal
dor no peito	B0016	qualitativa nominal
náuseas	B0017	qualitativa nominal
nariz entupido	B0018	qualitativa nominal
fadiga	B0019	qualitativa nominal
dor nos olhos	B00110	qualitativa nominal
perda de paladar ou olfato	B00111	qualitativa nominal
dor muscular	B00112	qualitativa nominal

Tabela 3 – Variáveis sobre Busca por Assistência

Variável	Código	Tipo
procurou ajuda médica	B002	qualitativa nominal
ficou em casa esperando melhora	B0031	qualitativa nominal
ligou para profissional de saúde	B0032	qualitativa nominal
automedicação	B0033	qualitativa nominal
tomou medicação com orientação	B0034	qualitativa nominal
foi a estabelecimento do SUS	B0035	qualitativa nominal
foi a estabelecimento particular	B0036	qualitativa nominal

Variável	Código	Tipo
tomou outra providência	B0037	qualitativa nominal

ANÁLISE EXPLORATÓRIA DOS DADOS

1. Análise dos sintomas por sexo em Salvador.

```
dados_junho_salvador |>
  filter(!is.na(sexo)) |>
  select(sexo, peso, febre:perda_paladar_ofato) |>
  pivot_longer(cols = febre:perda_paladar_ofato,
               names_to = "sintoma",
               values_to = "resposta") |>
  filter(resposta == "Sim") |>
  count(sexo, sintoma, wt = peso) |>
  group_by(sexo) |>
  mutate(freq = n / sum(n)) |>
  ungroup() |>
  ggplot(aes(x = sintoma, y = freq, fill = sexo)) +
  geom_col(position = "dodge") +
  facet_wrap(~sexo) +
  scale_y_continuous(labels = percent_format(decimal.mark = ",")) +
  scale_fill_viridis_d(option = "E") +
  labs(
    title = "Frequência dos sintomas relatados por sexo",
    x = "Sintoma",
    y = "Frequência relativa"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

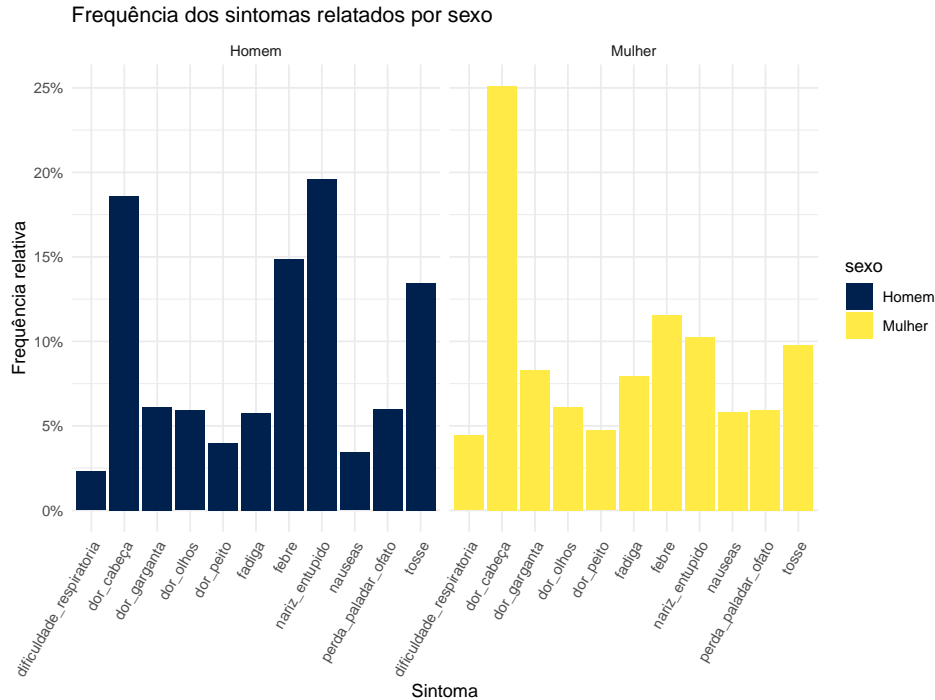


Figure 1: Frequência dos sintomas relatados por sexo

Comentário da análise: Analisando a 1 podemos observar evidências na frequência relativa nos sintomas relatados entre homens e mulheres. Observa-se que os homens relatam com mais frequência febre, dores de cabeça e nariz entupido. Além disso, sintomas como tosse e fadiga também apresentam percentuais expressivos. Já analisando o lado das mulheres no gráfico, conseguimos nota um alto percentual no relato de dores de cabeça, sendo o sintoma mais prevalente. Outros sintomas frequentes incluem febre e tosse. Esses dados indicam que, embora alguns sintomas sejam comuns a ambos os sexos, há diferenças na prevalência, com destaque para a maior proporção de mulheres relatando cefaleia e fadiga.

2. Análise da proporção de beneficiários do auxílio emergencial por faixa de aluguel em Salvador.

```
dados_junho_salvador |>
  filter(!is.na(auxilio_emergencial), !is.na(valor_aluguel)) |>
  group_by(valor_aluguel, auxilio_emergencial) |>
  summarise(total = sum(peso), .groups = "drop") |>
  group_by(valor_aluguel) |>
  mutate(freq = total / sum(total)) |>
  ggplot(aes(x = valor_aluguel, y = freq, fill = auxilio_emergencial)) +
  geom_col(position = "fill", color = "black") +
```

```

scale_y_continuous(labels = scales::percent_format(decimal.mark = ",")
) +
labs(
  title = "Proporção de recebimento do auxílio emergencial\npor
faixa de aluguel em Salvador",
  x = "Faixa de valor do aluguel",
  y = "Proporção (%)",
  fill = "Recebeu auxílio?"
) +
theme_minimal() +
theme(
  axis.text.x = element_text(angle = 45, hjust = 1),
  plot.title = element_text(hjust = 0.5)
) +
scale_fill_viridis_d(option = "E")

```

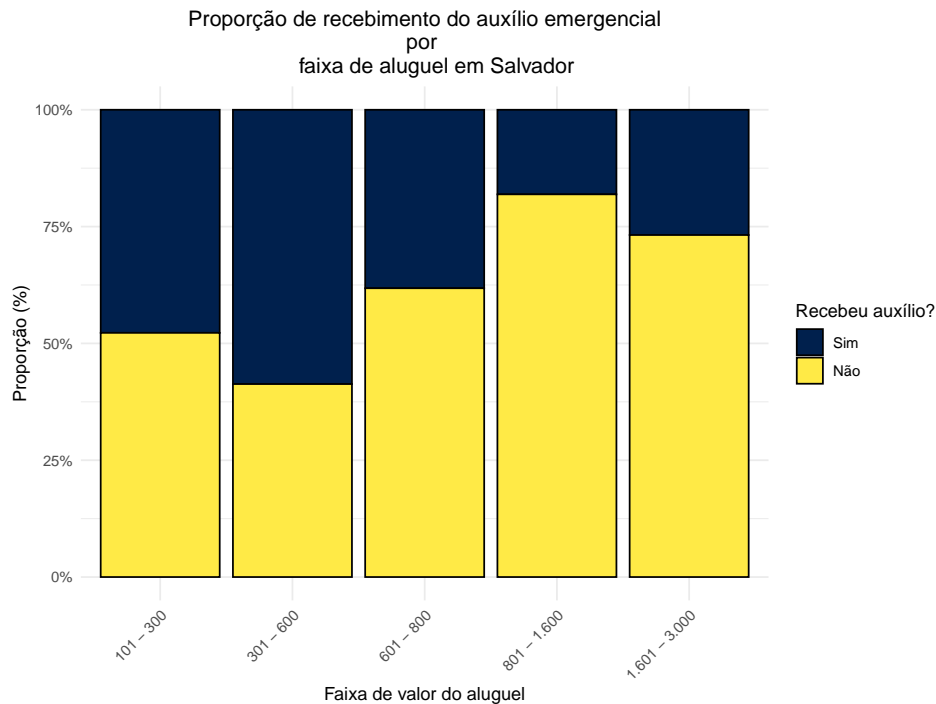


Figure 2: Proporção de recebimento do auxílio emergencial por faixa de aluguel em Salvador

Comentário da análise: Na Figura 2 é evidente que, em Salvador, o recebimento do auxílio emergencial foi proporcionalmente maior entre os domicílios com faixas de aluguel mais baixas. À medida que o valor do aluguel aumenta, a proporção de pessoas

que receberam o benefício diminuí significativamente. Esse padrão sugere que o auxílio emergencial foi direcionado, de forma mais intensa, às famílias em situação de maior vulnerabilidade econômica. O valor do aluguel, nesse contexto, funciona como um indicador indireto da condição socioeconômica dos domicílios, já que aluguéis mais altos tendem a estar associados a famílias com maior renda. Assim, os dados indicam que a política pública cumpriu parcialmente seu objetivo, ao beneficiar, sobretudo, as camadas da população com maiores dificuldades financeiras durante a pandemia.

3. Análise da Distribuição da Renda por Faixa Salarial Segundo Raça e Sexo em Salvador.

```
ggplot(dados_junho_salvador |> filter(!is.na(raça)) |>
  filter(!is.na(faixa_salarial)) |>
  filter(!is.na(sexo)),
  aes( x = raça, y = faixa_salarial, fill = sexo)) +
geom_boxplot() +
facet_wrap(~sexo) +
labs(
  title = "Boxplot da Faixa Salarial por Raça e sexo",
  x = "Raça",
  y = "Faixa Salarial",
  fill = "Sexo"
) +
scale_y_continuous(labels = scales::dollar_format(prefix = "R$ ",
  big.mark = ".",
  decimal.mark = ","))
)+
scale_fill_viridis_d(option = "E")
```

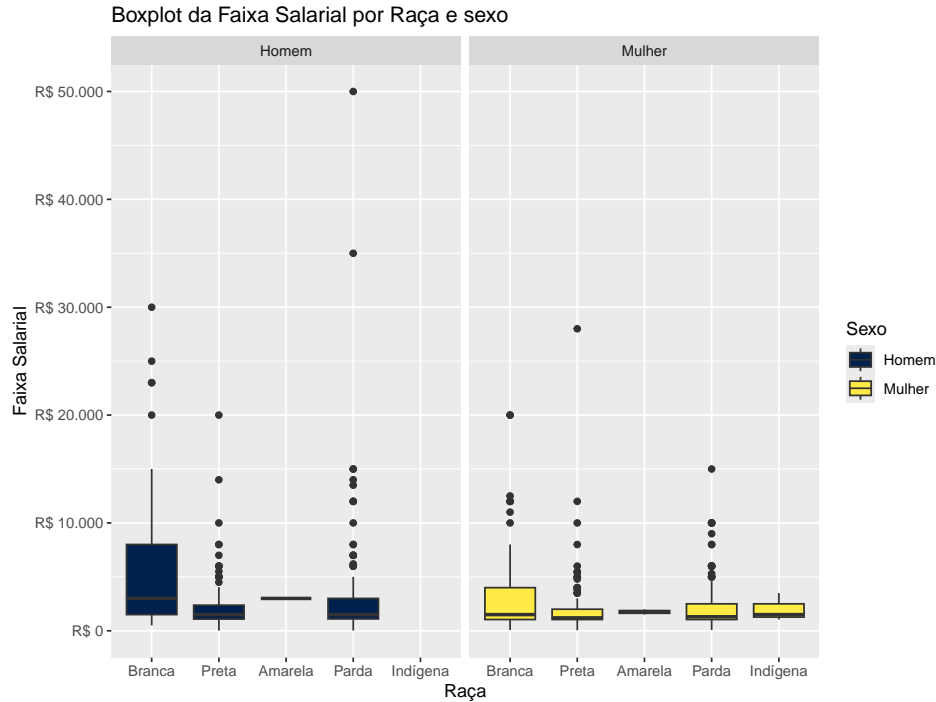



Figure 3: Boxplot da Faixa Salarial por Raça e sexo

Comentário da análise: Na Figura 3 representa boxplots da faixa salarial para diferentes raças, separado por homem e mulher. É possível perceber, que em todas as categorias, a mediana salarial de homens é maior do que a de mulheres. Focando apenas no boxplot de homens, conseguimos perceber que existe uma grande dispersão salarial, e até mesmo outliers chegando a R\$ 30,000. Preta e parda tem medianas menores que a de brancos, com uma concentração em valores pequenos de salários. Já nos das mulheres, é visível que os salários das mulheres brancas tem maior dispersão e mediana relativamente superior as demais raças, ou seja, preta, parda, indígena e amarela tem valores concentrados entre 0 a R\$ 2.000. É perceptível notar que os outliers estão presentes principalmente entre homens brancos e pardos e mulheres brancas.

```
resumo_por_raca <- dados_junho_salvador |>
  group_by(raca) |>
  filter(raca != "Ignorado") |>
  summarise(
    media_ponderada = weighted.mean(faixa_salarial, w = peso, na.rm = TRUE),
  )

resumo_por_raca |>
```

```
kable(caption = "Médias de salários por raça",
      col.names = c("Raça/Cor", "Média Ponderada (R$)"),
      digits = 2,
      align = "lrrr",
      booktabs = T) |>
collapse_rows() |>
kable_styling(latex_options = "striped")
```

Table 4: Médias de salários por raça

Raça/Cor	Média Ponderada (R\$)
Branca	4942.90
Preta	2008.65
Amarela	1991.76
Parda	2343.86
Indígena	1660.36

Comentário da análise: Na Tabela 4, um pequeno resumo das médias dos salários por raça, e conseguimos observar que os brancos teve uma média salarial maior de R\$ 4.924,90, e que os indígenas tiveram uma média salarial menor de R\$ 1.660,36.

4. Análise da distribuição percentual da escolaridade entre trabalhadores em Home Office e Presencial.

```
tabela1 <- dados_junho_salvador |>
  group_by(trabalho) |>
  count(escolaridade, wt = peso) |>
  drop_na() |>
  mutate(freq = n / sum(n))

dados_junho_salvador |>
  group_by(trabalho) |>
  count(escolaridade, wt = peso) |>
  drop_na() |>
  mutate(freq = n / sum(n)) |>
  ggplot(aes(x = escolaridade, y = freq, fill = trabalho)) +
  geom_bar(position = "dodge", stat = "identity",
           colour = "black", width = 0.7) +
  coord_flip() +
  theme_minimal(base_size = 9) +
```

```
labs(
  title = "Distribuição da escolaridade por tipo de trabalho",
  x = "Escolaridade",
  y = "Pocentagem (%)",
  fill = "Tipo de trabalho"
)+
scale_fill_viridis_d(option = "E") +
scale_y_continuous(labels = scales::percent_format(accuracy = 1))
```

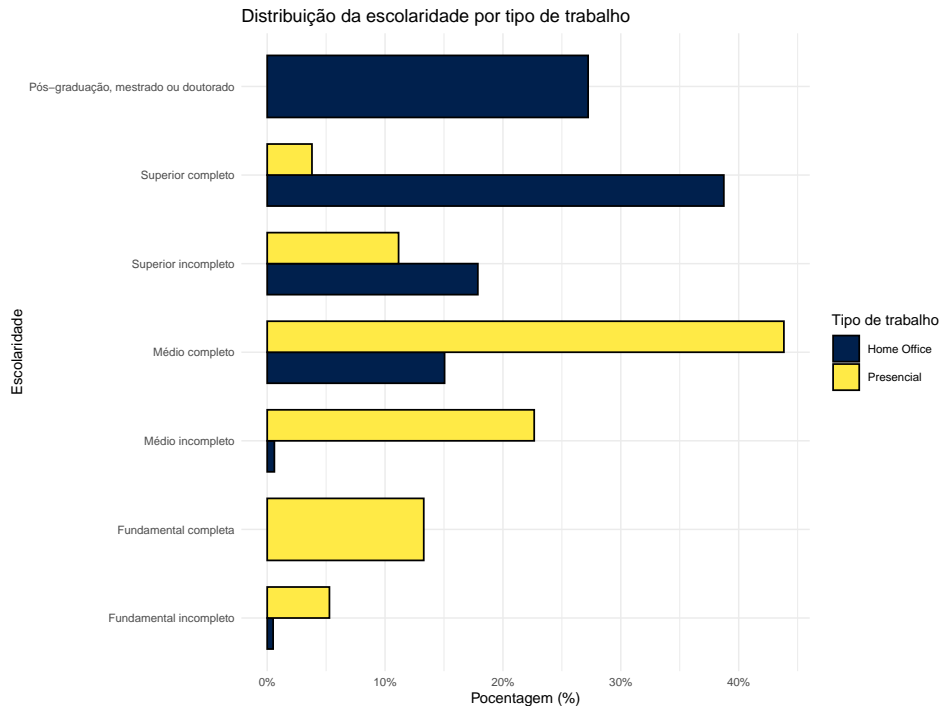


Figure 4: Distribuição da escolaridade por tipo de trabalho

Comentário da análise: Na Figura 4 a distribuição percentual da escolaridade entre os trabalhadores em home office e presencial. Entre os indivíduos com ensino superior completo e pós-graduação, há uma predominância marcante do trabalho remoto. Onde aproximadamente 45% dos trabalhadores com superior completo e 30% dos com pós-graduação atuam em regime home office, refletindo o vínculo entre maior nível de instrução e possibilidade de trabalho remoto. Em contrapartida, trabalhadores com ensino superior médio completo ou fundamental completo concentra-se no trabalho presencial, especialmente os com ensino médio completo quase 50%, no trabalho presencial. Nota-se ainda que os indivíduos com fundamental incompleto quase não estão representados no trabalho remoto. Os dados reforçam a relação entre escolaridade e tipo de trabalho du-

rante o período da pandemia, evidenciando uma desigualdades no acesso ao home office, uma mobilidade mais acessível a pessoas com maior qualificação profissional.

5. Análise das providências tomada pelas pessoas que tiveram os sintomas

```
dados_junho_salvador |>
  select(peso, ficou_em_casa:outra_providencia) |>
  pivot_longer(
    cols = -peso,
    names_to = "providencia",
    values_to = "resposta"
  ) |>
  filter(resposta == "Sim") |>
  count(providencia, wt = peso) |>
  mutate(providencia = case_when(
    providencia == "ficou_em_casa" ~ "Ficou em casa",
    providencia == "automedicação" ~ "Comprou ou
    tomou remédio por conta própria",
    providencia == "medicação_orientação_medica" ~ "Comprou ou
    tomou remédio por orientação médica",
    providencia == "ligou_profissional" ~ "Ligou para profissional
    de saúde",
    providencia == "outra_providencia" ~ "Outra providência",
    providencia == "visita_sus" ~ "Recebeu visita de
    profissional de saúde do SUS",
    providencia == "visita_particular" ~ "Recebeu visita de
    profissional de saúde particular",
    TRUE ~ providencia
  )) |>

ggplot(aes(x = reorder(providencia, -n), y = n)) +
  geom_col(color = "black", show.legend = FALSE) +
  scale_fill_viridis_d(option = "E") +
  labs(
    title = "Providência tomada pelas pessoas que tiveram os sintomas",
    x = NULL,
    y = "Número de Pessoas (absoluto)"
  ) +

  scale_y_continuous(labels = label_number(big.mark = ".",
                                           decimal.mark = ","))

  ) +
```

```
theme_minimal() +

theme(
  axis.text.x = element_text(angle = 45, hjust = 1, size = 10),
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14)
)
```

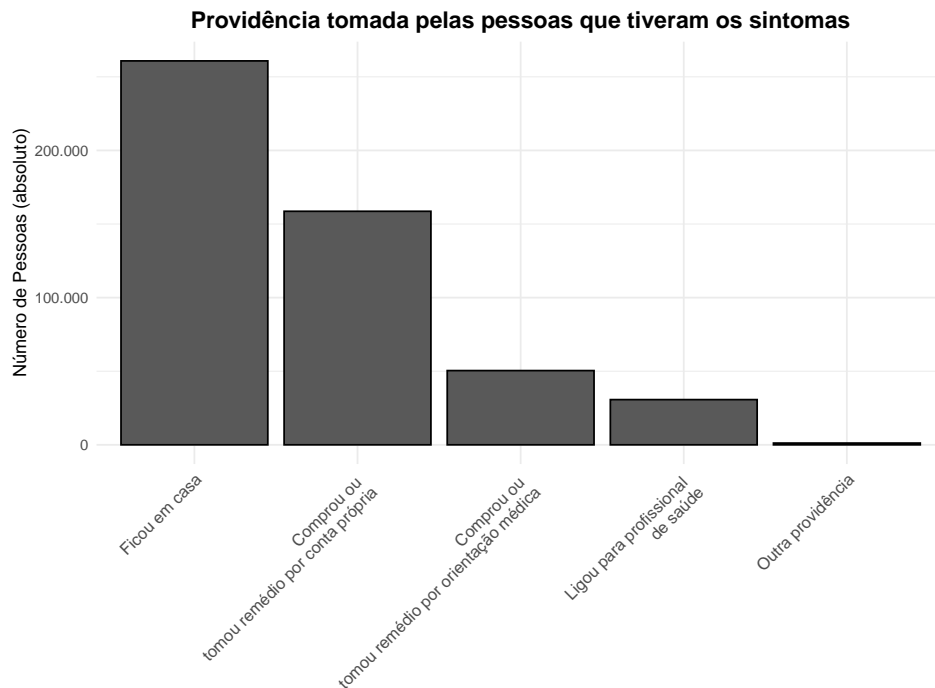


Figure 5: Providência tomada pelas pessoas que tiveram os sintomas

Comentário da análise: Na Figura 5 é visto que o número de pessoas que sentiram os sintomas durante a pandemia em Salvador, elas tomaram atitudes de se precaver, ou seja, fazendo o isolamento social, cerca de mais de 200 mil pessoas, após sentir os sintomas, procuraram em optar em ficar em casa, é notavel também que mais de 100 mil pessoas, decidiram ou comprar um remédio ou tomar um remédio por contra própria, que por um lado é preocupante, pois se automedicar sem ter passado por um exame médico, pode piorar os sintomas, até mesmo naquela época de pandemia, onde não sabíamos como tratar ou precaver o vírus.

CONCLUSÃO E DISCUSSÕES

Portanto, a análise dos dados revelou diversos impactos da pandemia, principalmente as desigualdades socioeconômicas. Notamos distinções na prevalência de sintomas entre homens e mulheres, com homens relatando mais febre e nariz entupido, e mulheres mais dores de cabeça e fadiga, sugerindo variações biológicas e perceptivas. O auxílio emergencial apresentou ser eficaz ao priorizar famílias de menor renda, indicando que a política pública alcançou os mais vulneráveis.

Relacionado a renda e desigualdade, a mediana salarial masculina superou a feminina em todas as categorias raciais, e indivíduos pretos e pardos concentraram-se em faixas salariais mais baixas, evidenciando profundas desigualdade de gênero e raciais no mercado de trabalho. A escolaridade mostrou-se crucial para o acesso ao trabalho remoto: quanto maior o nível de instrução, maior a adesão ao home office, expondo desigualdades no acesso a modalidades de trabalho mais seguras.

Por fim, a resposta da população aos sintomas em Salvador incluiu a adoção do isolamento social por mais de 200 mil pessoas, mas também a automedicação por mais de 100 mil indivíduos. Essa última prática, ressalta os desafios de saúde pública e a importância de orientação médica adequada para evitar agravamento dos sintomas.

Consideramos também que no decorrer da análise enfrentamos algumas dificuldades técnicas, como: Tratamento de variáveis com muitos valores ausentes (NA), necessidade de recodificar e reclassificar variáveis para garantir maior clareza nas análises, interpretação das variáveis disponíveis na base da PNAD COVID e também, encontramos dificuldade no uso da variável “peso” na hora da realização dos gráficos e dos cálculos de médias.

Sobretudo, este trabalho proporcionou uma rica oportunidade de consolidar conhecimentos técnicos em R, incluindo manipulação de dados com o dplyr, visualização com ggplot2, compreender a importância de uma análise crítica e contextualizada dos dados, principalmente quando se tratam de informações sensíveis como desigualdades sociais.

REFERÊNCIAS

IBGE. *PNAD COVID-19*. Instituto Brasileiro de Geografia e Estatística, 2020. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/27946-divulgacao-de-resultados.html>. Acesso em: 22 jul. 2025.

ESTATIDADOS. *Trabalhando com os microdados da PNAD COVID19 no R*. Disponível em: <https://estatidados.com.br/trabalhando-com-os-microdados-da-pnad-covid19-no-r/>. Acesso em: 22 jul. 2025.

ASSUNÇÃO, Gabriel. *COVID*. RPubS. Disponível em: <https://rpubs.com/gabriel-assuncao-ibge/covid>. Acesso em: 22 jul. 2025.

HERMES, ANTONIO. Fluxos de trabalho reproduzíveis, Aula ministrada na disciplina introdução ao R, Estatística, UFRN, Natal/RN, (15/07/2025 - 24/07/2025).

Wickham, H., & Golemund, G. (s.d.). R para Ciência de Dados. (Tradução para o português). Recuperado em 22 de julho de 2025, de <https://pt.r4ds.hadley.nz/>