

# Mapeamento dasimétrico no R Studio

Bruno Vargas Adorno

Curso:  
População, Espaço e Meio Ambiente (**SER-457-3**)

# Objetivos da aula

- Oferecer noções introdutórias do uso de R e Rstudio
- Ilustrar a utilidade da programação em R para desagregar espaciais da população brasileira.

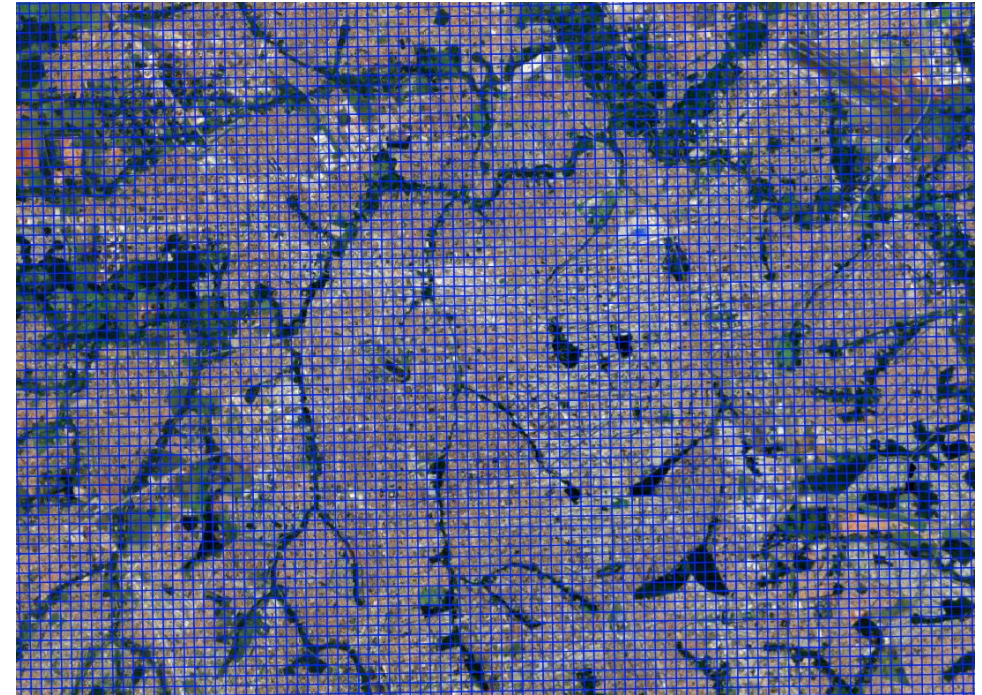
# Desagregar dados espaciais

Setores censitários



- Áreas urbanas ou rurais administrativas + trabalho do recenseador (atualizável a cada censo).
- Cada setor possui um **geocódigo único**
- **Dados coletados são agregados (somados)**

Grades celulares



- Regularidade no tempo e espaço
- Cada grade é identificada por um **geocódigo único**

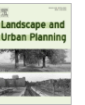
# Por que desagregar dados?

- Perspectiva a partir das minhas pesquisas:
  - Acessibilidade;
    - Importante termos mais exatidão na distância entre origem e destino.



Landscape and Urban Planning

Volume 256, April 2025, 105297



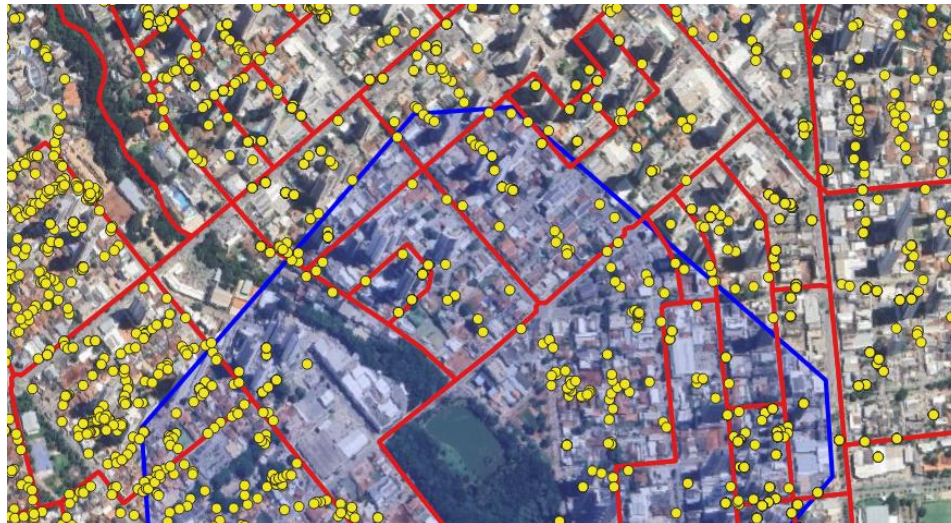
Research Paper

Combining spatial clustering and spatial regression models to understand distributional inequities in access to urban green spaces

Bruno Vargas Adorno <sup>a</sup> ✉, Rafael H.M. Pereira <sup>b</sup> ✉, Silvana Amaral <sup>a</sup> <sup>c</sup> ✉

Parques acessíveis a **600 m – 10 min de caminhada**

Setores censitários

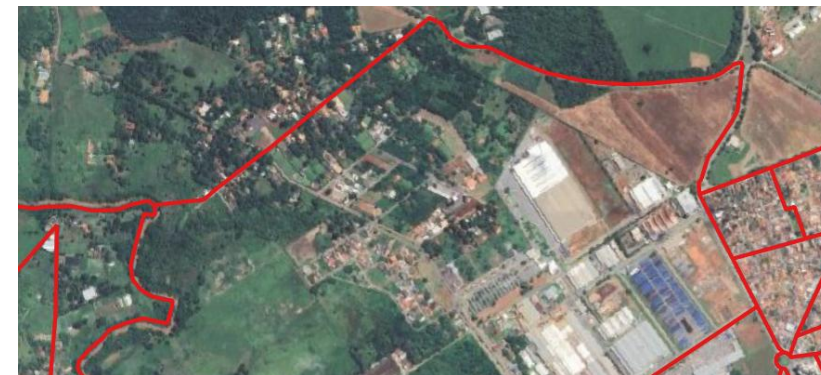
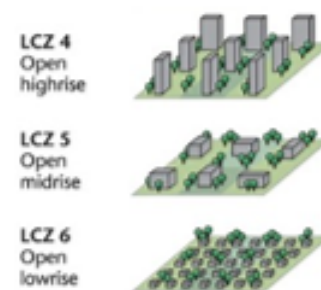
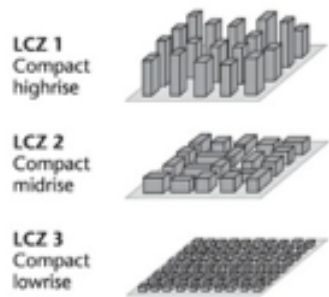


Grades regulares



# Por que desagregar dados?

- Representação de um fenômeno
  - **Zonas Climáticas Locais** são definidas como unidades de paisagens urbanas que abrangem centenas de metros a vários quilômetros;



- Setores com tamanho inferior a área mínima de uma LCZ
- formatos muito alongados, compactos e “complexos”.
- Grids também incluem área de transição



- Setores com tamanho tão grandes que não representam a LCZ dominante do grupo populacional que habita o setor censitário
- Grids também incluem área de transição



# Programação: R studio

- Por que programação?
  - **Documentar, automatizar, e disponibilizar rotina de processamento;**
  - Lidar com grande volume de dados;
  - Pacotes de estatística (espacial);
  - Facilita visualização de dados e produção de figuras de resultados dos estudos
- Por que aprender a utilizar R studio?
  - Gratuito e alimentado por uma grande comunidade;
  - Uma nova ferramenta para currículo;
  - **Disponibilidade de pacotes para análise.**

## Urban Demographics

Structured Procrastination on Cities, Transport, Equity, Spatial Analysis, R



**geobr:** Download Official Spatial Data Sets of Brazil

 Package

CRAN 1.9.1 downloads 139K GitHub code R-CMD-check passing

Citation:

Pereira, R.H.M.; Gonçalves, C.N.; et. al (2019). geobr: Download Official Spatial Data Sets of Brazil. R package. R Found. Stat. Comput., Vienna. <https://CRAN.R-project.org/package=geobr>.



**geobr:** Download Official Spatial Data Sets of Brazil

 Package

pypi package 0.2.2 downloads 226k downloads/month 5k GitHub code

Citation:

Carabetta, J., Pereira, R. H. M., & Gocalves, C. N. (2020). geobr: Download Official Spatial Data Sets of Brazil. Python package version 0.1.5. The Python Package Index (PyPI). Retrieved from <https://pypi.org/project/geobr/0.1.5/>



**censobr:** Download Data from Brazil's Population Census

 Package

CRAN 0.5.0 downloads 15K GitHub code R-CMD-check passing

Citation:

Pereira, R.H.M. and Barbosa, Rogerio J. (2023). censobr: Download Data from Brazil's Population Census. R package. R Found. Stat. Comput., Vienna. <https://CRAN.R-project.org/package=censobr>.

<https://www.urbandemographics.org/software/>

# R studio: vamos conhecer

The image shows a screenshot of the RStudio interface with several annotations pointing to specific features:

- Organization of menus:** An arrow points to the top menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), stating: "Organização amigável dos menus similares a outros softwares conhecidos. +utilizados 'novo projeto', 'abrir arquivos' e 'reiniciar sessão'".
- Files and tables in tabs:** An arrow points to the tab bar at the top, stating: "Arquivos e tabelas abertas no Rstudio organizadas em abas. A depender do tipo de arquivo, por exemplo *Markdown*, podemos formatar texto."
- Code editor and documentation:** An arrow points to the source editor, which displays a document with R code and Markdown text. The text says: "Visualização da documentação dos códigos e de conteúdos das tabelas. Nesse caso do *Markdown* podemos documentar **textos e códigos de forma mais amigável e visualizar o modo de impressão final e código fonte**".
- Console:** An arrow points to the console at the bottom, stating: "Console: apresenta as operações e resultados do que é solicitado nas linhas de código; Apresenta mensagens úteis ao usuário. Ex.: erros: qual foi o erro, em que linha aconteceu o erro?".

The RStudio window shows a project named "POPEA - RStudio". The source editor displays a document titled "(1b) Importing from censobr v.0.5.0" with the following content:

```
{r install censobr}
# install from CRAN
install.packages("censobr")
install.packages("dplyr")

library(censobr)
library(arrows)
library(dplyr)

Pessoas <- read_tracts(year = 2022, dataset = "Pessoas")
ResponsavelRenda <- read_tracts(year = 2022, dataset = "ResponsavelRenda")
```

The console shows the output of the code execution, including messages about the driver 'ESRI Shapefile' and the results of the data loading process.

# R studio: vamos conhecer

## Aba Ambiente

Espaço de trabalho que pode ser salvo, ou importado posteriormente. Onde importamos tabelas e fazemos limpeza de ambiente e memória. No ambiente normalmente são apresentados os dados, valores carregados no Rstudio bem como funções criadas pelo usuário.

Arquivos: mostra o diretório que está salvo o projeto e permite navegar em outros diretórios.

Plots: apresenta figuras que foram criadas a partir de dados e análises do usuário

Help: onde fica armazenada documentação de todas funções dos pacotes carregados no Rstudio.

The screenshot displays the R Studio interface with three panels highlighted by blue arrows:

- Environment Panel:** Shows the current environment with variables like `summed_values` (17102 obs. of 7 variables), `utm_crs` (List of 2), and `Values` (e.g., `epsg_code` as Named num 32722, `hemisphere` as "south", `longitude` as Named num -49.4, `Pessoas` as <object containing active binding>, `ResponsavelRenda` as <object containing active binding>, `state` as "PR", `states` as chr [1:3] "GO" "PA" "PR", `utm_zone` as Named num 22, and `Functions` like `project_UTM` and `tracttoell` as functions).
- Files Panel:** Shows the file explorer with a list of files including `.Rhistory`, `addres`, `Aggreg`, `Aggreg`, `Aggreg`, `censu`, `Dasynt`, `grids`, `POPEA`, and `Result`. The active file is `read_tracts (censobr)`.
- Help Panel:** Shows the documentation for the `read_tracts` function from the `censobr` package. The description states: "Download census tract-level data from Brazil's censuses." The usage is: `read_tracts(year, dataset, as_data_frame = FALSE, showProgress = TRUE, cache = TRUE, verbose = TRUE)`. The arguments are: `year` (Numeric. Year of reference in the format yyyy.) and `dataset` (Character. The dataset to be opened. The following options are available for each edition of the census: 2000 Census: "Basico", "Domicilio", "Responsavel", "Pessoa", "Instrucao", "Morador"; 2010 Census: ...).

# R studio: operações

Abrir arquivo operações\_pacotes\_funções.R no Rstudio.

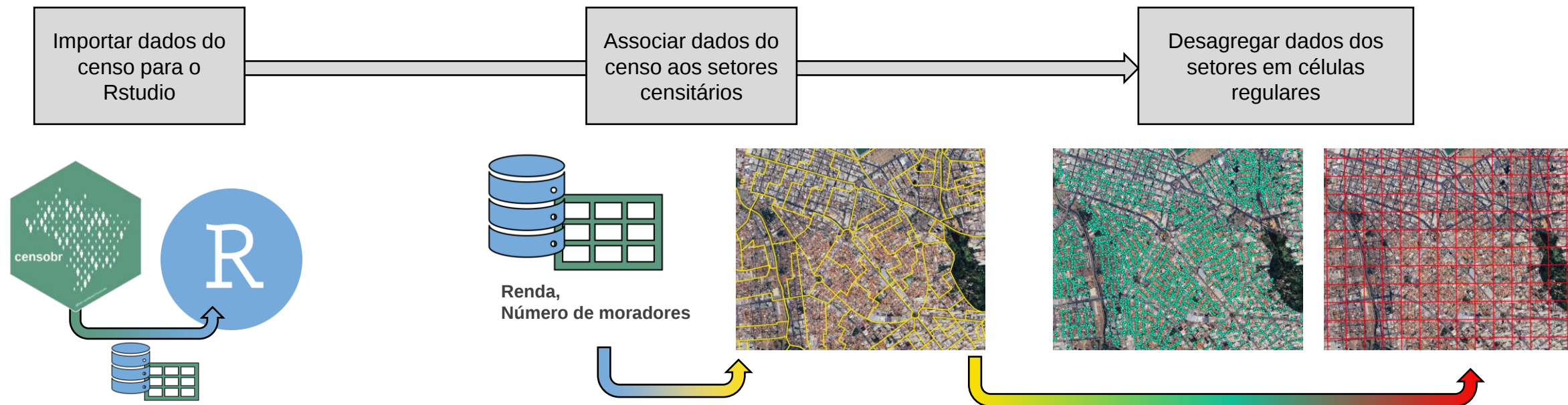
# R studio: pacotes e suas funções úteis

Abrir arquivo `dasymetric_census_2022_tract_to_cell.Rmd` no Rstudio.

# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Um arquivo **Rmd** (R Markdown) é um documento que combina texto formatado em Markdown com código R (e outras linguagens) em um único arquivo. Ele é usado principalmente no ambiente RStudio para criar documentos dinâmicos e reproduíveis.

Nesse arquivo seguiremos o seguinte fluxo de trabalho:



# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Importar dados  
do censo para o  
Rstudio

```
library(censobr)
library(arrow)
library(dplyr)

# investigate how variables are organized to be properly accessed using
"read_tracts"
data_dictionary(year = 2022, dataset = "tracts")

# collect variables from different topics surveyed in census 2022
Pessoas <- read_tracts(year = 2022, dataset = "Pessoas")
ResponsavelRenda <- read_tracts(year = 2022, dataset = "ResponsavelRenda")
head(Pessoas)

# create dataframe with the variables you need. use left join to bind columns
# using "code tract" as common ID

df_pop1 <-

# collect number of residents
Pessoas|> collect() |> select(code_tract, resid = demografia_v01006) %>%

# collect number of black and brown people and sum as "black" variable
left_join(Pessoas|> collect() |> mutate(black = rowSums(across(c(raca_v01318
, raca_v01320)), na.rm = TRUE)) |> select(code_tract, black), by = "code_tract") %>%

# collect number of household's representative who receive income
left_join(ResponsavelRenda|> collect() |> select(code_tract, rep_inc = v06001), by
= "code_tract") %>%

# collect number household's representative and mean income and multiply
left_join(ResponsavelRenda|> collect() |> mutate(income = v06001 * v06004) |>
select(code_tract, income), by = "code_tract")
```

|     | A                 | B                        | D             | E          | F                                                                                                                 | G | H |
|-----|-------------------|--------------------------|---------------|------------|-------------------------------------------------------------------------------------------------------------------|---|---|
| 1   | (censobr) dataset | Tema                     | ável          | Categorias | Descrição                                                                                                         |   |   |
| 2   |                   |                          | Nome Original |            |                                                                                                                   |   |   |
| 3   | Pessoas           | Identificação Geográfica | CD_SETOR      |            | Geocódigo de Setor Censitário                                                                                     |   |   |
| 4   | Pessoas           | Identificação Geográfica | SITUACAO      |            | Situação do Setor Censitário                                                                                      |   |   |
| 5   | Pessoas           | Identificação Geográfica |               | Urbana     | Urbana                                                                                                            |   |   |
| 447 | Pessoas           | Demografia               | V01039        |            | 50 a 59 anos                                                                                                      |   |   |
| 448 | Pessoas           | Demografia               | V01040        |            | 60 a 69 anos                                                                                                      |   |   |
| 449 | Pessoas           | Demografia               | V01041        |            | 70 anos ou mais                                                                                                   |   |   |
| 450 | Pessoas           | Parentesco               | V01042        |            | Pessoa responsável pelo domicílio                                                                                 |   |   |
| 451 | Pessoas           | Parentesco               | V01043        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é cônjuge ou companheiro(a) de    |   |   |
| 452 | Pessoas           | Parentesco               | V01044        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é cônjuge ou companheiro(a) do    |   |   |
| 453 | Pessoas           | Parentesco               | V01045        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é filho(a) do responsável e do cô |   |   |
| 454 | Pessoas           | Parentesco               | V01046        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é filho(a) somente do responsáv   |   |   |
| 455 | Pessoas           | Parentesco               | V01047        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é enteado(a)                      |   |   |
| 456 | Pessoas           | Parentesco               | V01048        |            | Relação de parentesco ou de convivência com a pessoa responsável pelo domicílio é genro ou nora                   |   |   |

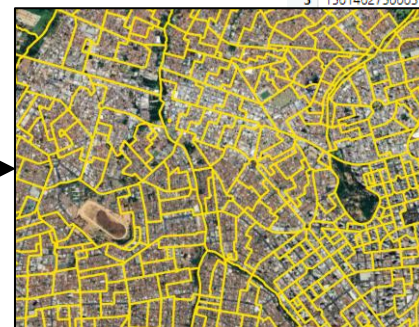
| code_tract      | resid | black | rep_inc | income    |
|-----------------|-------|-------|---------|-----------|
| 110001505000002 | 928   | 601   | 336     | 824218.08 |
| 110001505000003 | 556   | 361   | 208     | 442299.52 |
| 110001505000004 | 222   | 149   | 85      | 141519.90 |
| 110001505000006 | 785   | 514   | 281     | 483994.40 |
| 110001505000007 | 748   | 508   | 291     | 561903.54 |
| 110001505000011 | 631   | 352   | 234     | 744662.88 |

# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Importar dados do censo para o Rstudio

Associar dados do censo aos setores censitários

|   | CD_SETOR        | SITUACAO | CD_SIT | CD_TIPO | AREA_KM2     | CD_REGIAO | NM_REGIAO | CD_UF | NM_UF    |
|---|-----------------|----------|--------|---------|--------------|-----------|-----------|-------|----------|
| 1 | 120040105000708 | Urbana   | 1      | 0       | 3.845045e-02 | 1         | Norte     | 12    | Acre     |
| 2 | 130340305000118 | Urbana   | 1      | 0       | 1.456526e-02 | 1         | Norte     | 13    | Amazonas |
| 3 | 150140275000317 | Urbana   | 1      | 0       | 2.238230e-02 | 1         | Norte     | 15    | Pará     |
|   |                 | Urbana   | 1      | 0       | 1.118959e-02 | 1         | Norte     | 15    | Pará     |
|   |                 | Urbana   | 3      | 0       | 1.867893e-01 | 1         | Norte     | 15    | Pará     |
|   |                 | Urbana   | 1      | 0       | 5.154288e-03 | 2         | Nordeste  | 22    | Piauí    |
|   |                 | Urbana   | 1      | 0       | 2.090193e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 1.668699e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 4.068895e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 3.759162e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 7.710336e-03 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 1.356253e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 3.149096e-02 | 2         | Nordeste  | 23    | Ceará    |
|   |                 | Urbana   | 1      | 0       | 1.143517e-02 | 2         | Nordeste  | 23    | Ceará    |



Já vimos! O que estamos fazendo aqui?

|   | code_tract      | SITUACAO | CD_SIT | CD_TIPO | AREA_KM2     | CD_REGIAO | NM_REGIAO | CD_UF | NM_UF    | resid | black | rep_inc | income   |
|---|-----------------|----------|--------|---------|--------------|-----------|-----------|-------|----------|-------|-------|---------|----------|
| 1 | 120040105000708 | Urbana   | 1      | 0       | 3.845045e-02 | 1         | Norte     | 12    | Acre     | 19    | 17    | 14      | 12472.04 |
| 2 | 130340305000118 | Urbana   | 1      | 0       | 1.456526e-02 | 1         | Norte     | 13    | Amazonas | 61    | 54    | 11      | 14330.03 |
| 3 | 150140275000317 | Urbana   | 1      | 0       | 2.238230e-02 | 1         | Norte     | 15    | Pará     | 44    | 35    | 12      | 22526.40 |
| 4 | 150442205000199 | Urbana   | 1      | 0       | 1.118959e-02 | 1         | Norte     | 15    | Pará     | 50    | 41    | 17      | 18949.39 |
| 5 | 150560110000009 | Urbana   | 3      | 0       | 1.867893e-01 | 1         | Norte     | 15    | Pará     | 23    | 18    | 6       | 25968.00 |

Exclui setores censitários onde tenho valores “não aplicáveis” de residentes

Cria a coluna UF condicionada ao valor das células de outra coluna.

No caso, criamos uma coluna com as siglas dos estados com base no código do IBGE

```
# load library
library(sf)

# All census tracts in Brazil
cntr2022 <- st_read("./census_tracts/BR_setores_CD2022.shp")

#select columns needed
cntr2022 <- cntr2022[,c(1:17,30)]

# Rename one column name
colnames(cntr2022)[1] <- "code_tract"

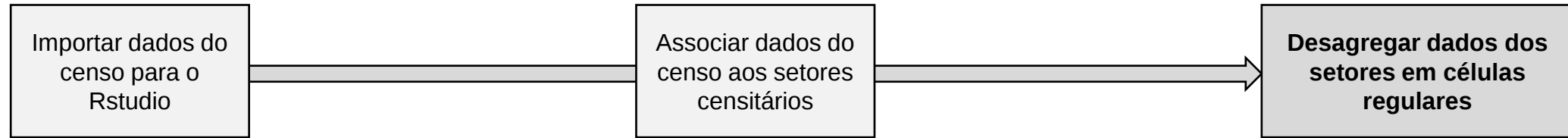
# merge census survey and tract data
cntr_Pop2022 <- left_join(cntr2022, df_pop1, by = "code_tract")

# select important column for the analysis and omit NA values. This NA values are from tract
#centr_Pop2010$women<-centr_Pop2010$head - centr_Pop2010$men
cntr_Pop2022 <- cntr_Pop2022 %>%
  filter(!if_all(19, is.na))

cntr_Pop2022 <- cntr_Pop2022 %>%
  mutate(UF = case_when(
    CD_UF == "11" ~ "RO", CD_UF == "12" ~ "AC", CD_UF == "13" ~ "AM",
    CD_UF == "14" ~ "RR", CD_UF == "15" ~ "PA", CD_UF == "16" ~ "AP",
    CD_UF == "17" ~ "TO", CD_UF == "21" ~ "MA", CD_UF == "22" ~ "PI",
    CD_UF == "23" ~ "CE", CD_UF == "24" ~ "RN", CD_UF == "25" ~ "PB",
    CD_UF == "26" ~ "PE", CD_UF == "27" ~ "AL", CD_UF == "28" ~ "SE",
    CD_UF == "29" ~ "BA", CD_UF == "31" ~ "MG", CD_UF == "32" ~ "ES",
    CD_UF == "33" ~ "RJ", CD_UF == "35" ~ "SP", CD_UF == "41" ~ "PR",
    CD_UF == "42" ~ "SC", CD_UF == "43" ~ "RS", CD_UF == "50" ~ "MS",
    CD_UF == "51" ~ "MT", CD_UF == "52" ~ "GO", CD_UF == "53" ~ "DF"))

#clean unnecessary files
rm(cntr2022, df_pop1)
gc() # clean memory
```

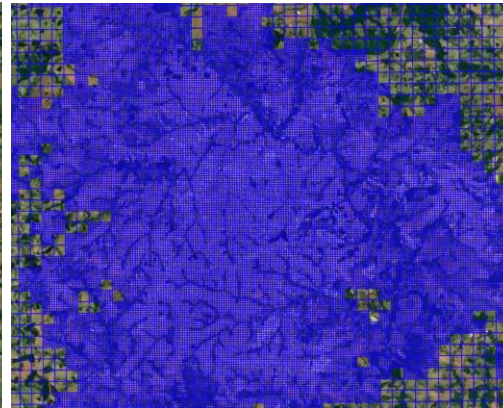
# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd



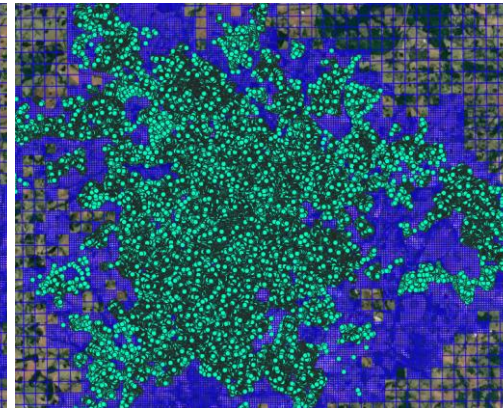
Dados demográficos nos setores



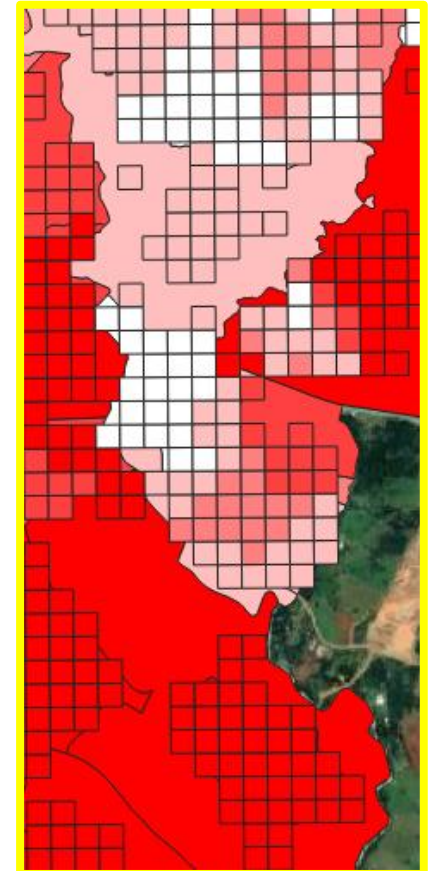
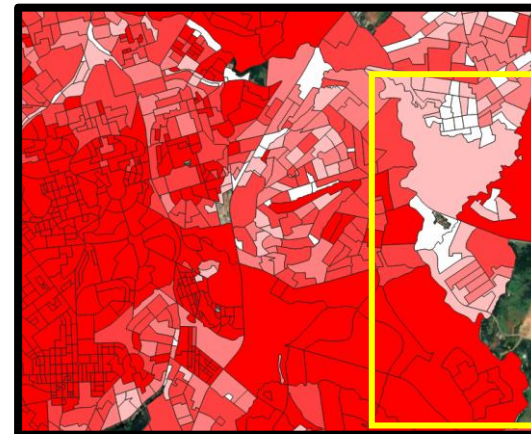
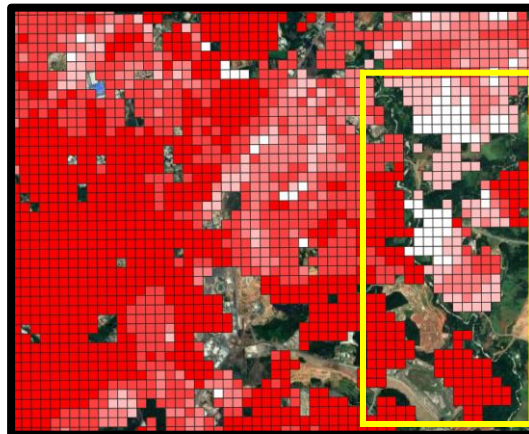
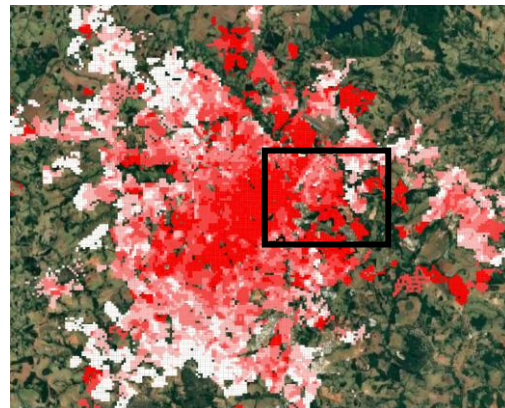
Grade celular



Pontos de endereço



Renda média por célula



# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Importar dados do censo para o Rstudio

Associar dados do censo aos setores censitários

Desagregar dados dos setores em células regulares

```
state = "GO"
cntr_Pop2022_UF = cntr_Pop2022 %>% filter(UF == c("GO"))

tracttozell <- function (state, cntr_Pop2022_UF) {

  #read geolocalized address
  address_2022 <- st_read(paste0("./addresses/",state,"_urbano_conurbado.shp"))

  #read gridcells
  gridcells_2022 <- st_read(paste0("./grids/200m_grid_",state,"_IBGE.shp"))

  address_2022 = st_transform(address_2022,
                             st_crs(cntr_Pop2022_UF))
  gridcells_2022 = st_transform(gridcells_2022,
                               st_crs(cntr_Pop2022_UF))

  # make sure all files have same coordinate system
  st_crs(cntr_Pop2022_UF)
  st_crs(gridcells_2022)
  st_crs(address_2022)

  # Make sure census tract have valid geometry
  cntr_Pop2022_UF <- st_make_valid(cntr_Pop2022_UF)
  print("geometry validated")
  gc()

  # obtaining code_tract id and variables for each address point
  address_2022 <- st_join(address_2022,cntr_Pop2022_UF,left=FALSE)
  print("address_2022 received code_tract")

  # associate gridcells id to address points
  address_2022 <- st_join(address_2022, gridcells_2022[,1], left=FALSE)
  print("address_2022 received grid_id")
}
```

Vamos testar desagregar dados para uma área de estudo situada em Goiás.

Podemos criar uma função para garantir que tudo que está escrito entre chaves { } possa ser replicado para outras áreas de estudo com apenas uma linha de código.

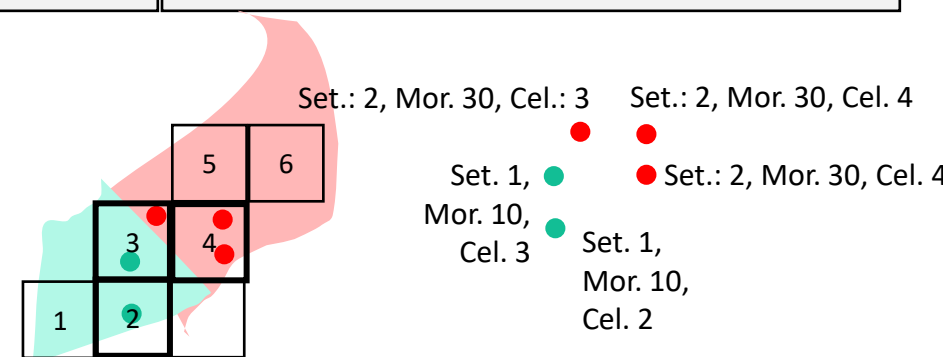
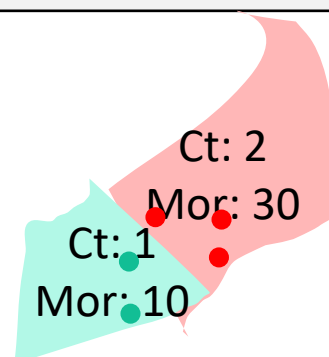
Abrimos os endereços geolocalizados e grades celulares do IBGE

Para estabelecer relações espaciais entre arquivos, é uma boa prática garantir que nossos arquivos compartilham a mesma CRS

Outra boa prática é garantir que setores censitários possuem geometrias válidas

Associamos dados demográficos dos setores censitários aos endereços contidos em cada um.

Associamos identificador único de cada células aos pontos de endereços contidos em cada uma.



# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Importar dados do censo para o Rstudio

Associar dados do censo aos setores censitários

Desagregar dados dos setores em células regulares

```
# remove geometry column turning address_2022 into a data frame, demanding less
# from computer
address_2022 <- st_drop_geometry(address_2022)
gc()

# Count number of address per census tract and add a total column
# As each address represents one household, to disaggregate at the address point
# level we divide the number of people by the total number of household geolocated at
# tract-level.
address_2022 <- address_2022 %>%
  group_by(code_tract) %>%
  mutate(
    total = n(),
    resid_i = resid / total,
    black_i = black / total,
    rep_inc_i = rep_inc / total,
    income_i = income / total) %>%
  ungroup()

print("demographics disaggregated to address")

# # To check if volume was preserved, group by 'code_tract' and calculate the
# sum of 'population'.
# proof <- address_2022 %>%
#   group_by(code_tract) %>%
#   summarise(resid = sum(resid_i))
# # Summarise the values by grid cell

summed_values <- address_2022 %>%
  group_by(ID_UNICO) %>%
  summarise(resid = sum(resid_i),
            black = sum(black_i),
            rep_inc = sum(rep_inc_i),
            income = sum(income_i))

print("demographics aggregated to grid")

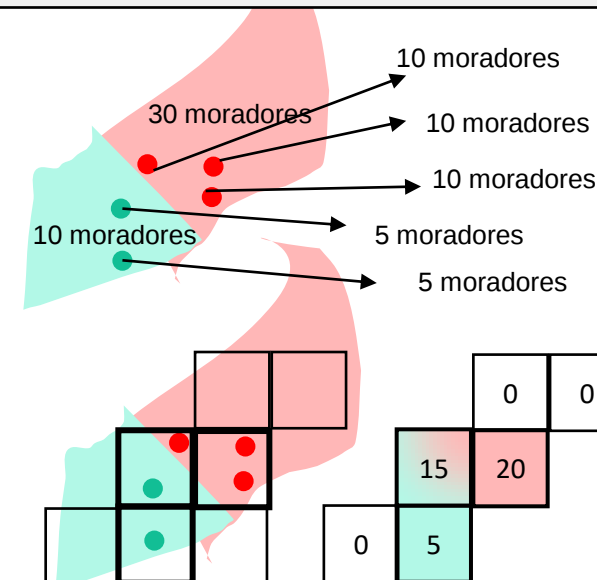
#calculate share value of each variable per resident/representative
summed_values$blackp <- summed_values$black/summed_values$resid
summed_values$incomep <- summed_values$income/summed_values$rep_inc
```

As próximas operações não dependem de info espacial, então, se possível eliminamos a info de geometria e trabalhamos apenas com dado tabular, mais leve.

Agrupamos endereços com mesmo identificador de setor censitário (code\_tract), contamos a quantidade de endereços residenciais registrados e dividimos dados demográficos contabilizados por setor censitário para cada endereço. Assumimos que dentro de cada setor a chance da população se distribuir para cada endereço é a mesma.

Uma vez que definimos a quantidade de moradores por endereço é fácil somar a quantidade por célula

Podemos calcular as taxas de pessoas pretas e a renda média do chefe de família.



# Dasymetric\_census\_2022\_tract\_to\_cell.Rmd

Importar dados do censo para o Rstudio

Associar dados do censo aos setores censitários

Desagregar dados dos setores em células regulares

```
summed_values_grid <- left_join(gridcells_2022[,c(1,11)],summed_values, by =  
'ID_UNICO')  
summed_values_grid <- summed_values_grid %>% filter(!if_all(2, is.na))  
  
print("empty cells deleted")  
  
return(summed_values_grid)  
}
```

Finalmente, podemos juntar nossos **dados agregados às células** ao dado original de **grades celulares**, utilizando o "ID\_ÚNICO" como identificador, e excluir as células que por ventura ficaram sem informação do número de residentes.

**EXTRA:** podemos programar a repetição do processamento por estado utilizando **for** + função criada **tracttocell**

```
states <- c("GO", "PA", "PR")  
  
# Projeção automática para UTM  
summed_values_grid_utm <- st_crs(summed_values_grid)  
  
for (state in states) {  
  cntr_Pop2022_UF = cntr_Pop2022 %>% filter(UF == state)  
  summed_values_grid <- tracttocell(state, cntr_Pop2022_UF)  
  summed_values_grid <- project_UTM(summed_values_grid)  
  st_crs(summed_values_grid)  
  st_write(summed_values_grid, paste0("./Results/demographic_data_",state,".shp"))  
  st_write(cntr_Pop2022_UF, paste0("./Results/demographic_data_ct",state,".shp"))  
  rm(cntr_Pop2022_UF,summed_values_grid)  
  gc()  
}
```