

Taller 2

Ariel Alvarez Marcos Gonzalez Arturo Valenzuela

2025-01-02

```
if(!require(readxl)){install.packages("readxl"); library(readxl)}
```

```
## Loading required package: readxl
```

```
if(!require(dplyr)){install.packages("dplyr"); library(dplyr)}
```

```
## Loading required package: dplyr
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
if(!require(chilemapas)){install.packages("chilemapas"); library(chilemapas)}
```

```
## Loading required package: chilemapas
```

```
## Loading required package: sf
```

```
## Linking to GEOS 3.11.2, GDAL 3.8.2, PROJ 9.3.1; sf_use_s2() is TRUE
```

```
## La documentacion del paquete y ejemplos de uso se encuentran en https://pacha.dev/chilemapas/.
```

```
## Visita https://buymeacoffee.com/pacha/ si deseas donar para contribuir al desarrollo de este software
```

```
if(!require(ggplot2)){install.packages("ggplot2"); library(ggplot2)}
```

```
## Loading required package: ggplot2
```

```
if(!require(stringr)){install.packages("stringr"); library(stringr)}
```

```
## Loading required package: stringr
```

```

if(!require(stringi)){install.packages("stringi"); library(stringi)

## Loading required package: stringi

if(!require(tools)){install.packages("tools"); library(tools)

## Loading required package: tools

if(!require(gridExtra)){install.packages("gridExtra"); library(gridExtra)

## Loading required package: gridExtra

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine

if(!require(cowplot)){install.packages("cowplot"); library(cowplot)

## Loading required package: cowplot

if(!require(rio)){install.packages("rio"); library(rio)

## Loading required package: rio

if(!require(tidyr)){install.packages("tidyr"); library(tidyr)

## Loading required package: tidyr

```

Pull a GITHUB

```

dir.create("data", showWarnings = FALSE) #DATA
dir.create("output", showWarnings = FALSE) # RESULTADOS
dir.create("R", showWarnings = FALSE) # SCRIPT
dir.create("docs", showWarnings = FALSE) # Documentos

readme_content <- "# Análisis CASEN/SERVEL 2021
Análisis de la relación entre participación electoral y condiciones socioeconómicas en Chile.

por motivos de tamaño KB solo fue posible analizar servel y casen 2021
## Estructura del Proyecto
- `data/`: Datos brutos de CASEN y SERVEL
- `output/`: Resultados y visualizaciones
- `R/`: Scripts de análisis

```

```

- `docs/`: Documentación adicional

## Equipo
- Ariel Álvarez
- Marcos González
- Arturo Zamorano"

writeLines(readme_content, "README.md")

# Crear los scripts principales de análisis
script_names <- c(
  "01_load_data.R",      # Para cargar los datos iniciales
  "02_clean_data.R",     # Para limpiar y preparar los datos
  "03_analysis.R",       # Para realizar los análisis principales
  "04_visualizations.R"  # Para crear gráficos y visualizaciones
)

for(script in script_names) {
  file.path("R", script) %>%
    file.create()
}

# Crear el contenido del script principal que configura el proyecto
load_data_content <- '# Configuración inicial del proyecto
if (!require("renv")) install.packages("renv")
renv::init()

# Instalar y cargar paquetes necesarios
packages <- c(
  "tidyverse", # Para manipulación de datos
  "readxl",    # Para leer archivos Excel
  "sf",        # Para datos geográficos
  "here",      # Para manejo de rutas
  "DiagrammeR", # Para diagramas ER
  "usethis"    # Para configuración de Git
)

# Instalar paquetes faltantes
new_packages <- packages[!(packages %in% installed.packages()[,"Package"])]
if(length(new_packages)) install.packages(new_packages)

# Cargar paquetes
lapply(packages, library, character.only = TRUE)

# Configurar Git
usethis::use_git()
'

writeLines(load_data_content, file.path("R", "01_load_data.R"))

# Crear .gitignore - Este archivo evita que se suban archivos innecesarios a GitHub
gitignore_content <- ".Rproj.user
.Rhistory

```

```
.RData
.Ruserdata
*.Rproj"

writeLines(gitignore_content, ".gitignore")
```

Carga y procesamiento de datos electorales

Presidencial 2021

Carga datos de participación Presidencial 2021 primera vuelta

```
metropolitana_2021_1 <- import("https://github.com/Arielo2518/Taller_2_final/raw/main/data/Servel/metropol
datos_comunas1_2021 <- metropolitana_2021_1 %>%
  select(Distrito, Comuna, Local, Mesa, Inscritos, Votación) %>%
  group_by(Distrito, Comuna)
```

Datos agrupados por comuna y porcentaje de participación

```
datos_comunas1_2021_agrupado <- metropolitana_2021_1 %>%
  select(Distrito, Comuna, Local, Mesa, Inscritos, Votación) %>%
  group_by(Distrito, Comuna) %>%
  summarize(
    Total_Inscritos = sum(as.numeric(Inscritos), na.rm = TRUE),
    Total_Votacion = sum(as.numeric(Votación), na.rm = TRUE)
  ) %>%
  mutate(Participacion_Porcentaje = (Total_Votacion / Total_Inscritos) * 100)
```

'summarise()' has grouped output by 'Distrito'. You can override using the
'.groups' argument.

Datos Limpios

```
eleccion_2021_primeravuelta <- datos_comunas1_2021_agrupado %>%
  rename(comuna = Comuna) %>%
  mutate(comuna = str_to_lower(comuna))

eleccion_2021_primeravuelta
```

```
## # A tibble: 52 x 5
## # Groups:   Distrito [7]
##   Distrito   comuna   Total_Inscritos Total_Votacion Participacion_Porcen~1
```

```
##      <chr>      <chr>      <dbl>      <dbl>      <dbl>
## 1 Distrito 10 la granja      102187      45371      44.4
## 2 Distrito 10 macul          100045      54600      54.6
## 3 Distrito 10 providencia    170892      96543      56.5
## 4 Distrito 10 san joaquin     80763      39445      48.8
## 5 Distrito 10 santiago       339067     136398      40.2
## 6 Distrito 10 ñuñoa          198985     123031      61.8
## 7 Distrito 11 la reina        90939      56792      62.5
## 8 Distrito 11 las condes      271386     171671      63.3
## 9 Distrito 11 lo barnech~     84259      55042      65.3
## 10 Distrito 11 peñalolen      190640      95956      50.3
## # i 42 more rows
## # i abbreviated name: 1: Participacion_Porcentaje
```

SEGUNDA VUELTA

Carga datos de participación Presidencial 2021 segunda vuelta

```
metropolitana_2021_2 <- import("https://github.com/Arielo2518/Taller_2_final/raw/main/data/Servel/metropol
datos_comunas2_2021 <- metropolitana_2021_2 %>%
  select(Distrito, Comuna, Local, Mesa, Inscritos, Votación) %>%
  group_by(Distrito, Comuna)
```

Datos agrupados por comuna y porcentaje de participación

```
datos_comunas2_2021_agrupado <- metropolitana_2021_2 %>%
  select(Distrito, Comuna, Local, Mesa, Inscritos, Votación) %>%
  group_by(Distrito, Comuna) %>%
  summarize(
    Total_Inscritos = sum(as.numeric(Inscritos), na.rm = TRUE),
    Total_Votacion = sum(as.numeric(Votación), na.rm = TRUE)
  ) %>%
  mutate(Participacion_Porcentaje = (Total_Votacion / Total_Inscritos) * 100)
```

```
## 'summarise()' has grouped output by 'Distrito'. You can override using the
## '.groups' argument.
```

Datos limpios

```
eleccion_2021_segundavuelta <- datos_comunas2_2021_agrupado %>%
  rename(comuna = Comuna) %>%
  mutate(comuna = str_to_lower(comuna))
eleccion_2021_segundavuelta
```

```
## # A tibble: 52 x 5
## # Groups:   Distrito [7]
##   Distrito   comuna      Total_Inscritos Total_Votacion Participacion_Porcen~1
##   <chr>      <chr>          <dbl>          <dbl>          <dbl>
## 1 Distrito 10 la granja      102187          55913          54.7
## 2 Distrito 10 macul        100045          64016          64.0
## 3 Distrito 10 providencia  170892         104693          61.3
## 4 Distrito 10 san joaquin    80763          48040          59.5
## 5 Distrito 10 santiago      339067         158676          46.8
## 6 Distrito 10 ñuñoa        198985         139003          69.9
## 7 Distrito 11 la reina       90939          63597          69.9
## 8 Distrito 11 las condes     271386         184227          67.9
## 9 Distrito 11 lo barnech~    84259          59439          70.5
## 10 Distrito 11 peñalolen    190640         115888          60.8
## # i 42 more rows
## # i abbreviated name: 1: Participacion_Porcentaje
```

Limpieza

```
remove(datos_comunas1_2021, datos_comunas2_2021, datos_comunas1_2021_agrupado, datos_comunas2_2021_agrupado)
```

```
## Warning in remove(datos_comunas1_2021, datos_comunas2_2021,
## datos_comunas1_2021_agrupado, : objeto
## 'datos_comunas2_2021_agrupadometropolitana_2021_1' no encontrado
```

```
## Warning in remove(datos_comunas1_2021, datos_comunas2_2021,
## datos_comunas1_2021_agrupado, : objeto 'datos_comunas1_2021' no encontrado
```

```
## Warning in remove(datos_comunas1_2021, datos_comunas2_2021,
## datos_comunas1_2021_agrupado, : objeto 'datos_comunas2_2021' no encontrado
```

geolocalización

Primera vuelta

```
eleccion_2021_primeravuelta <- eleccion_2021_primeravuelta %>%
  mutate(Participacion_Porcentaje = as.numeric(Participacion_Porcentaje))

participacion_comuna <- eleccion_2021_primeravuelta %>%
  filter(!is.na(Participacion_Porcentaje) & Participacion_Porcentaje >= 0) %>%
  group_by(comuna) %>%
  summarise(participacion_comuna = sum(Participacion_Porcentaje, na.rm = TRUE))

comunas_santiago <- mapa_comunas %>%
  filter(codigo_region == 13) %>%
  left_join(
    codigos_territoriales %>%
```

```

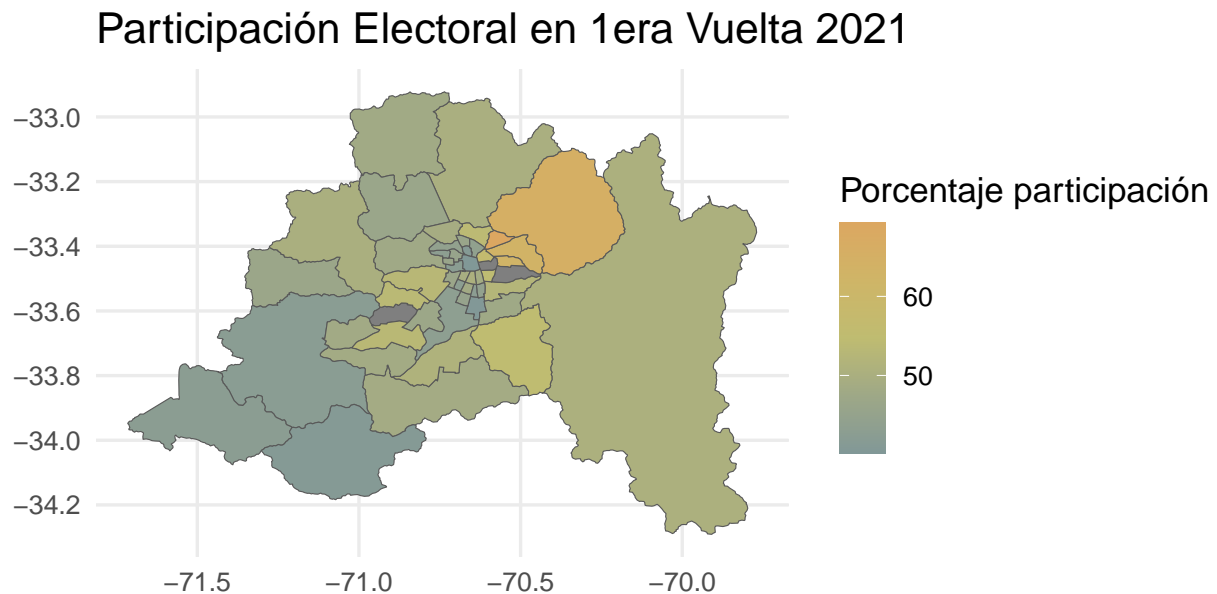
    select(matches("comuna"))
  ) %>%
  mutate(nombre_comuna = str_trim(tolower(nombre_comuna))) %>% # Convertir a minúsculas y quitar espacios
  left_join(participacion_comuna %>%
    mutate(comuna = str_trim(tolower(comuna))), by = c("nombre_comuna" = "comuna"))

## Joining with 'by = join_by(codigo_comuna)'

paleta <- c("#DCA761", "#CFB567", "#BFBC71", "#9EA887", "#819897")

participacion_2021_primeravuelta <- ggplot(comunas_santiago) +
  geom_sf(aes(fill = participacion_comuna, geometry = geometry)) +
  scale_fill_gradientn(colours = rev(paleta), name = "Porcentaje participación") +
  labs(title = "Participación Electoral en 1era Vuelta 2021") +
  theme_minimal(base_size = 13)
participacion_2021_primeravuelta

```



Segunda vuelta

```

eleccion_2021_segundavuelta <- eleccion_2021_segundavuelta %>%
  mutate(Participacion_Porcentaje = as.numeric(Participacion_Porcentaje))

```

```

participacion_comuna <- eleccion_2021_segundavuelta %>%
  filter(!is.na(Participacion_Porcentaje) & Participacion_Porcentaje >= 0) %>%
  group_by(comuna) %>%
  summarise(participacion_comuna = sum(Participacion_Porcentaje, na.rm = TRUE))

comunas_santiago <- mapa_comunas %>%
  filter(codigo_region == 13) %>%
  left_join(
    codigos_territoriales %>%
      select(matches("comuna"))
  ) %>%
  mutate(nombre_comuna = str_trim(tolower(nombre_comuna))) %>% # Convertir a minúsculas y quitar espacios
  left_join(participacion_comuna %>%
    mutate(comuna = str_trim(tolower(comuna))), by = c("nombre_comuna" = "comuna"))

## Joining with 'by = join_by(codigo_comuna)'

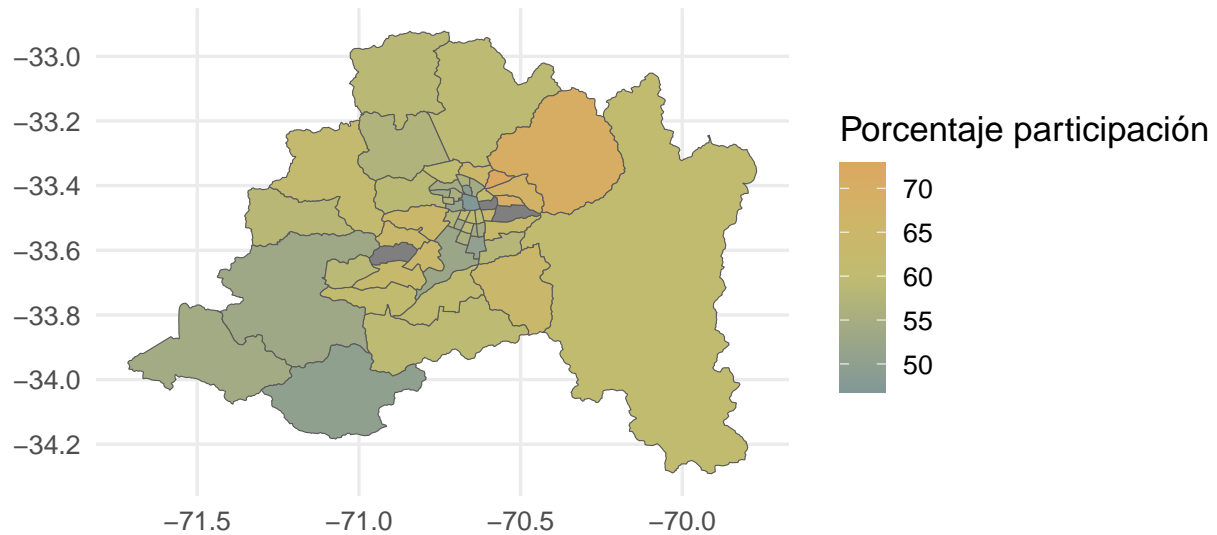
paleta <- c("#DCA761", "#CFB567", "#BFBC71", "#9EA887", "#819897")

participacion_2021_segundavuelta <- ggplot(comunas_santiago) +
  geom_sf(aes(fill = participacion_comuna, geometry = geometry)) +
  scale_fill_gradientn(colours = rev(paleta), name = "Porcentaje participación") +
  labs(title = "Participación Electoral en 2da Vuelta 2021") +
  theme_minimal(base_size = 13)

participacion_2021_segundavuelta

```


Participación Electoral en 2da Vuelta 2021



Gráficos Elecciones

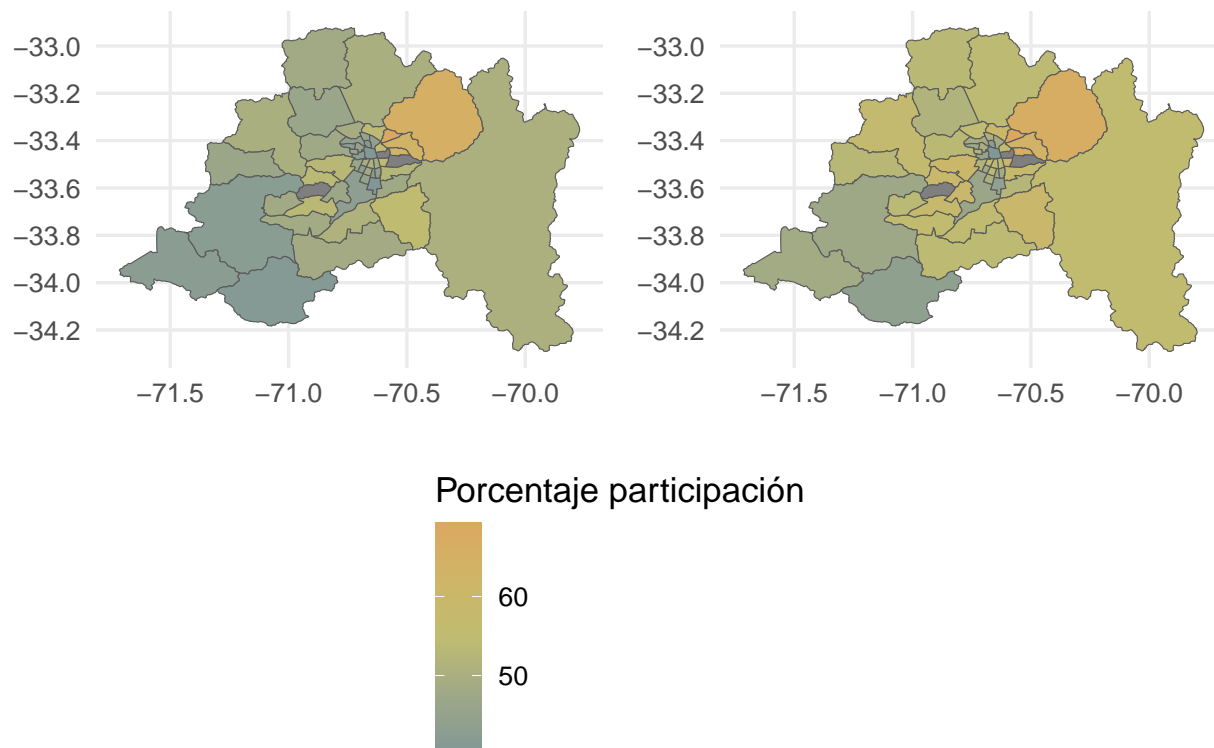
```
# Gráfico 2021
grafico_combinado_2021 <- plot_grid(
  participacion_2021_primeravuelta + theme(legend.position = "none"),
  participacion_2021_segundavuelta + theme(legend.position = "none"),
  align = "v",
  ncol = 2
)

# Extraer la leyenda de uno de los gráficos
leyenda <- get_legend(participacion_2021_primeravuelta)

## Warning in get_plot_component(plot, "guide-box"): Multiple components found;
## returning the first one. To return all, use 'return_all = TRUE'.

# Combinar gráficos y leyenda
plot_grid(grafico_combinado_2021, leyenda, ncol = 1, rel_heights = c(0.81, 0.5))
```

Participación Electoral en 1era Vuelta 2011 Participación Electoral en 2da Vuelta 2011



CASEN

Etiquetar códigos de comunales de Casen a escritos

```
comuna_codes_rm <- c(
  "13101" = "SANTIAGO",
  "13102" = "CERRILLOS",
  "13103" = "CERRO NAVIA",
  "13104" = "CONCHALI",
  "13105" = "EL BOSQUE",
  "13106" = "ESTACION CENTRAL",
  "13107" = "HUECHURABA",
  "13108" = "INDEPENDENCIA",
  "13109" = "LA CISTERNA",
  "13110" = "LA FLORIDA",
  "13111" = "LA GRANJA",
  "13112" = "LA PINTANA",
  "13113" = "LA REINA",
  "13114" = "LAS CONDES",
  "13115" = "LO BARNECHEA",
  "13116" = "LO ESPEJO",
  "13117" = "LO PRADO",
  "13118" = "MACUL",
```

```

"13119" = "MAIPU",
"13120" = "NUNOA",
"13121" = "PEDRO AGUIRRE CERDA",
"13122" = "PENALOLEN",
"13123" = "PROVIDENCIA",
"13124" = "PUDAHUEL",
"13125" = "QUILICURA",
"13126" = "QUINTA NORMAL",
"13127" = "RECOLETA",
"13128" = "RENCA",
"13129" = "SAN JOAQUIN",
"13130" = "SAN MIGUEL",
"13131" = "SAN RAMON",
"13132" = "VITACURA",
"13201" = "PUENTE ALTO",
"13202" = "PIRQUE",
"13203" = "SAN JOSE DE MAIPO",
"13301" = "COLINA",
"13302" = "LAMPA",
"13303" = "TILTIL",
"13401" = "SAN BERNARDO",
"13402" = "BUIN",
"13403" = "CALERA DE TANGO",
"13404" = "PAINE",
"13501" = "MELIPILLA",
"13502" = "ALHUE",
"13503" = "CURACAVI",
"13504" = "MARIA PINTO",
"13505" = "SAN PEDRO",
"13601" = "TALAGANTE",
"13602" = "EL MONTE",
"13603" = "ISLA DE MAIPO",
"13604" = "PADRE HURTADO",
"13605" = "PENAFLOR"
)

comuna_codes_rm <- tibble::tibble(
  comuna = names(comuna_codes_rm), # Extracción de los códigos de CASEN
  nombre_comuna = unname(comuna_codes_rm) # Nombres
)

comuna_codes_rm <- comuna_codes_rm %>%
  mutate(nombre_comuna = str_to_lower(nombre_comuna))

```

Union CASEN 2022 con comunas

#Casen 2022 en sus notas para la investigación comtepla que “no está diseñada para ser representativa a nivel provincial o comunal, por lo que la base principal de Casen ha sido publicada sin incluir las variables de identificación # de estas dos unidades territoriales” sin embargo, estas variables se han puesto a disposición de las y los investigadores en una base de datos complementaria porque unimos ambas bases para identificar valores comunales de CASEN.

```
casen_2021_rm <- import("https://github.com/Arielo2518/Taller_2_final/raw/main/data/CASEN/casen_2021_rm")
casen_comunas_2021 <- import("https://github.com/Arielo2518/Taller_2_final/raw/main/data/CASEN/casen_comunas_2021")
```

```
casen_2021_rm_actualizada <- casen_2021_rm %>%
  left_join(casen_comunas_2021, by = c("folio", "id_persona")) %>% #Las bases se unieron mediante folio
  group_by(folio) %>%
  mutate_all(~ifelse(. %in% c(-88, -99, 97,98,99), NA, .)) # Eliminamos valores faltantes
```

```
## 'mutate_all()' ignored the following grouping variables:
## * Column 'folio'
## i Use 'mutate_at(df, vars(-group_cols()), myoperation)' to silence the message.
```

```
casen_2021_rm_actualizada
```

```
## # A tibble: 38,674 x 17
## # Groups:   folio [13,202]
##   folio id_persona region nse sexo edad e6a ytoth ytotcorh dau dautr
##   <dbl>      <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl>
## 1 1.00e8         1     13     2     2   54    11 2.95e6 2945500     8     7
## 2 1.00e8         2     13     2     2   28    11 2.95e6 2945500     8     7
## 3 1.00e8         3     13     2     2   27    13 2.95e6 2945500     8     7
## 4 1.00e8         4     13     2     2   19    11 2.95e6 2945500     8     7
## 5 1.00e8         5     13     2     1   56    11 2.95e6 2945500     8     7
## 6 1.00e8         1     13     2     1   81     5 6.42e5  963997     4     3
## 7 1.00e8         2     13     2     1   50     9 6.42e5  963997     4     3
## 8 1.00e8         1     13     2     2   59    11 4.15e5  415310     3     3
## 9 1.00e8         2     13     2     1   60     9 4.15e5  415310     3     3
## 10 1.00e8        1     13     2     2   52    11 7.15e5  714750     9     8
## # i 38,664 more rows
## # i 6 more variables: educ <dbl>, expr <dbl>, numper <dbl>, comuna <dbl>,
## #   expp <dbl>, expc <dbl>
```

Etiqueta de comunas por código

```
colnames(casen_2021)
```

```
casen_2021_rm_actualizada <- casen_2021_rm_actualizada %>%
  mutate(comuna = as.character(comuna))

casen_2021 <- casen_2021_rm_actualizada %>%
  left_join(comuna_codes_rm, by = "comuna") %>%
  mutate(comuna = nombre_comuna) %>%
  select(-nombre_comuna)
```

Limpieza CASEN

```
remove(casen_2021_rm, casen_comunas_2021, comuna_codes_rm)
```

Manipulacion CASEN promedio de DECILES por comuna

```
promedio_dau_comuna <- casen_2021 %>%  
  group_by(comuna) %>% # Agrupar por comuna  
  summarise(promedio_dau = mean(dau, na.rm = TRUE)) # Calcular promedio de 'dau' ignorando NA  
  
# Ver el resultado  
promedio_dau_comuna  
  
## # A tibble: 52 x 2  
##   comuna      promedio_dau  
##   <chr>      <dbl>  
## 1 alhue      4.73  
## 2 buin       4.66  
## 3 calera de tango 4.52  
## 4 cerrillos  5.01  
## 5 cerro navia  4.55  
## 6 colina     4.96  
## 7 conchali    4.55  
## 8 curacavi    4.29  
## 9 el bosque   4.27  
## 10 el monte   4.02  
## # i 42 more rows
```

Cruce de participación electoral por promedio de decil CASEN

```
# Reemplazamos Ñ por n para cruce con CASEN  
eleccion_2021_primeravuelta <- eleccion_2021_primeravuelta %>%  
  mutate(across(everything(), ~str_replace_all(., "ñ", "n")))   
  
eleccion_2021_segundavuelta <- eleccion_2021_segundavuelta %>%  
  mutate(across(everything(), ~str_replace_all(., "ñ", "n")))   
  
# Realizar el cruce de bases por 'comuna'  
  
datos_cruzados_primeravuelta <- eleccion_2021_primeravuelta %>%  
  left_join(promedio_dau_comuna, by = "comuna")  
  
datos_cruzados_segundavuelta <- eleccion_2021_segundavuelta %>%  
  left_join(promedio_dau_comuna, by = "comuna")  
  
datos_cruzados_primeravuelta
```

```
## # A tibble: 52 x 6
## # Groups:   Distrito [7]
##   Distrito comuna Total_Inscritos Total_Votacion Participacion_Porcen~1
##   <chr>      <chr>      <chr>          <chr>          <chr>
## 1 Distrito 10 la granja 102187          45371          44.3999725992543
## 2 Distrito 10 macul 100045          54600          54.5754410515268
## 3 Distrito 10 providencia 170892          96543          56.4935748894038
## 4 Distrito 10 san joaquin 80763          39445          48.8404343573171
## 5 Distrito 10 santiago 339067          136398          40.2274476725839
## 6 Distrito 10 nunoa 198985          123031          61.8292836143428
## 7 Distrito 11 la reina 90939          56792          62.4506537349212
## 8 Distrito 11 las condes 271386          171671          63.2571319080572
## 9 Distrito 11 lo barnech~ 84259          55042          65.3247724278712
## 10 Distrito 11 penalolen 190640          95956          50.3336130927403
## # i 42 more rows
## # i abbreviated name: 1: Participacion_Porcentaje
## # i 1 more variable: promedio_dau <dbl>
```

datos_cruzados_segundavuelta

```
## # A tibble: 52 x 6
## # Groups:   Distrito [7]
##   Distrito comuna Total_Inscritos Total_Votacion Participacion_Porcen~1
##   <chr>      <chr>      <chr>          <chr>          <chr>
## 1 Distrito 10 la granja 102187          55913          54.7163533521877
## 2 Distrito 10 macul 100045          64016          63.9872057574092
## 3 Distrito 10 providencia 170892          104693          61.2626688200735
## 4 Distrito 10 san joaquin 80763          48040          59.482683902282
## 5 Distrito 10 santiago 339067          158676          46.7978305172724
## 6 Distrito 10 nunoa 198985          139003          69.856019297937
## 7 Distrito 11 la reina 90939          63597          69.9336918153927
## 8 Distrito 11 las condes 271386          184227          67.8837522937808
## 9 Distrito 11 lo barnech~ 84259          59439          70.5432060670077
## 10 Distrito 11 penalolen 190640          115888          60.7889215274864
## # i 42 more rows
## # i abbreviated name: 1: Participacion_Porcentaje
## # i 1 more variable: promedio_dau <dbl>
```

```
# Creamos una clasificación simplificada de nivel educacional
casen_2021_rm_actualizada <- casen_2021_rm_actualizada %>%
  mutate(nivel_educacional = case_when(
    # Sin educación formal o básica incompleta
    e6a %in% c(1, 2, 3, 4, 5, 6) | is.na(e6a) ~ "Básica o menos",

    # Educación básica completa y media (completa o incompleta)
    e6a %in% c(7, 8, 9, 10) ~ "Media",

    # Educación técnica (en cualquier nivel)
    e6a %in% c(11, 12) ~ "Técnica",

    # Educación universitaria o superior (incluye postgrados)
    e6a %in% c(13, 14, 15) ~ "Universitaria",
```

```

    # Por si hay algún caso no considerado
    TRUE ~ "No especificado"
  ))

# Creamos una visualización que muestra la distribución por comuna
educacion_comuna <- casen_2021_rm_actualizada %>%
  # Agrupamos por comuna y nivel educacional para contar
  group_by(comuna, nivel_educacional) %>%
  summarise(cantidad = n(), .groups = 'drop') %>%
  # Calculamos el porcentaje dentro de cada comuna
  group_by(comuna) %>%
  mutate(porcentaje = (cantidad / sum(cantidad)) * 100) %>%
  ungroup()

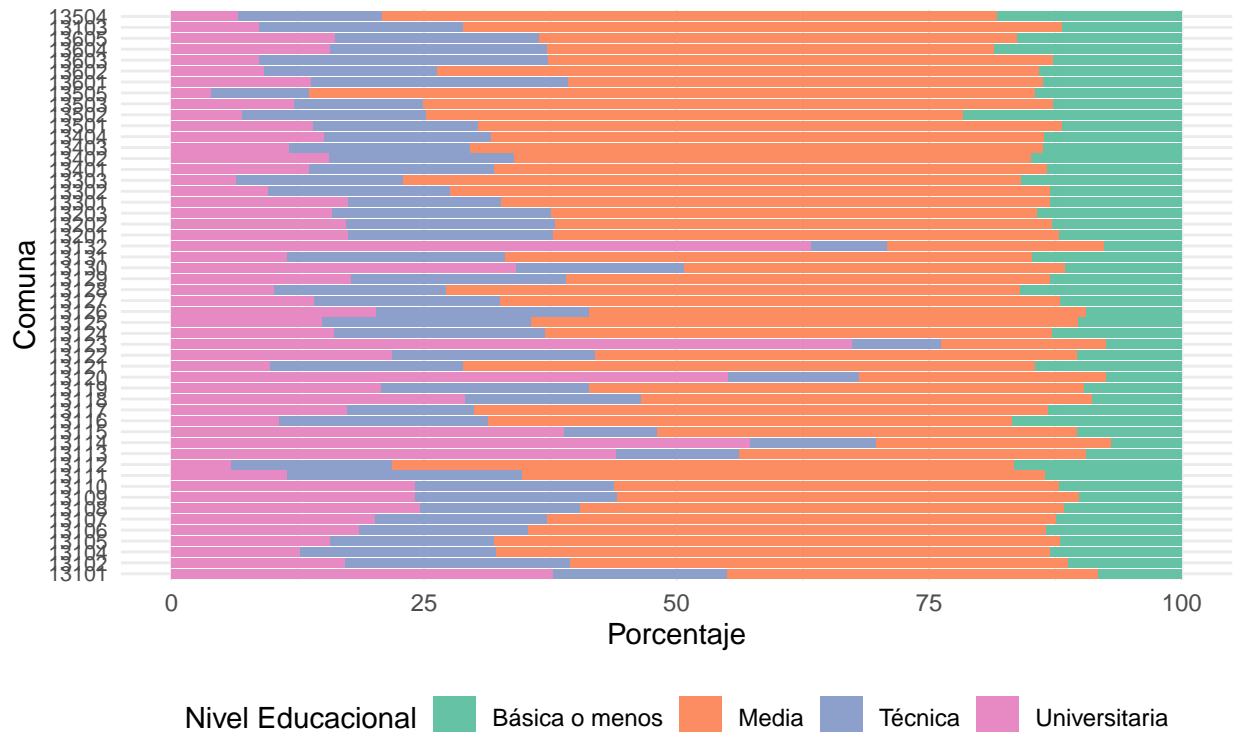
# Creamos un gráfico de barras apiladas por comuna
grafico_educacion <- ggplot(educacion_comuna,
  aes(x = reorder(comuna, porcentaje),
    y = porcentaje,
    fill = nivel_educacional)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  scale_fill_brewer(palette = "Set2") + # Usamos una paleta de colores agradable
  labs(
    title = "Nivel Educativo por Comuna",
    subtitle = "Región Metropolitana - 2021",
    x = "Comuna",
    y = "Porcentaje",
    fill = "Nivel Educativo"
  ) +
  theme_minimal() +
  theme(
    axis.text.y = element_text(size = 8),
    legend.position = "bottom"
  )

# Mostramos el gráfico
grafico_educacion

```

Nivel Educativo por Comuna

Región Metropolitana – 2021



```
# Creamos un resumen general de la distribución
resumen_educacional <- casen_2021_rm_actualizada %>%
  group_by(nivel_educacional) %>%
  summarise(
    cantidad = n(),
    porcentaje = round((cantidad / nrow(casen_2021_rm_actualizada)) * 100, 1)
  ) %>%
  arrange(desc(porcentaje))

# Mostramos el resumen
print(resumen_educacional)
```

```
## # A tibble: 4 x 3
##   nivel_educacional cantidad porcentaje
##   <chr>             <int>      <dbl>
## 1 Media             18159        47
## 2 Universitaria     9070       23.5
## 3 Técnica           6883       17.8
## 4 Básica o menos    4562       11.8
```