

Análisis de Pobreza por Subgrupos Sociales en Estados Unidos (ACS 2016–2023)

2025-09-01

Este informe presenta un análisis detallado de la evolución de la pobreza en Estados Unidos entre 2016 y 2023, con especial atención a las diferencias entre **subgrupos sociales** y **estados**. Utilizando datos del American Community Survey (ACS), se busca identificar patrones persistentes, brechas estructurales y dinámicas territoriales que afectan la distribución de la pobreza.

A través de indicadores clave (KPIs), estimaciones robustas y visualizaciones interactivas, se pretende ofrecer una herramienta útil para investigadores, responsables de políticas públicas y cualquier persona interesada en comprender las desigualdades socioeconómicas en el país.

1. Carga de paquetes necesarios: dplyr, readr, purrr, stringr, ggplot2, tidyr, patchwork, knitr, kableExtra, RColorBrewer, lubridate, scales, gridExtra, cowplot, grid.

2. Lectura condicional del CSV limpio: Este bloque realiza una **lectura condicional del archivo limpio** *df_pobreza_limpio.csv*, optimizando el tiempo de ejecución si el dataset ya ha sido procesado previamente. En caso de que el archivo no exista, se **lee dinámicamente el conjunto completo de archivos CSV** disponibles en la carpeta *data/*, detectando automáticamente el **delimitador correcto** y extrayendo el año desde el nombre del archivo.

Los datos se transforman desde formato ancho a largo, se **limpian los valores numéricos** y se filtran las **categorías sociales relevantes** para el análisis de pobreza. Posteriormente, se reorganizan en formato pivotado, aplicando una función segura para **extraer el primer valor válido** en cada celda, evitando errores por valores faltantes.

Finalmente, se genera un nuevo conjunto de datos con variables clave como **población total**, **población bajo el umbral de pobreza** y **porcentaje de pobreza**, junto con etiquetas temáticas y sociales en español. El resultado se guarda como *df_pobreza_limpio.csv* para **futuras ejecuciones más rápidas y reproducibles**.

```
ruta_limpia <- "../data/df_pobreza_limpio.csv"

if (file.exists(ruta_limpia)) {
  df_final <- read_csv(ruta_limpia,
    col_types = cols(
      Fecha = col_date(),
      Estado = col_character(),
      Grupo_social = col_character(),
      Tematica = col_character(),
      Total_poblacion = col_integer(),
      Bajo_umbral_pobreza = col_integer(),
      Porcentaje_pobreza = col_double()
    ))
} else {
```

```

archivos <- list.files(path = "../data/", pattern = "*.csv", full.names = TRUE)

leer_archivo <- function(ruta) {
  año <- str_extract(ruta, "20\\d{2}")
  print(paste("Leyendo:", ruta))

  primera_linea <- readLines(ruta, n = 1)
  delim_usado <- ifelse(str_detect(primer_linea, ";"), ";", ",")

  read_delim(file = ruta, delim = delim_usado,
             col_types = cols(.default = col_character())) %>%
    mutate(Year = as.integer(año))
}

df_completo <- map_df(archivos, leer_archivo)

df_limpio <- df_completo %>%
  select(Year, everything()) %>%
  select(-`...314`) %>%
  pivot_longer(
    cols = -c(Year, `Label (Grouping)`),
    names_to = c("Estado", "Categoria", "Tipo"),
    names_sep = "!!!",
    values_to = "valor"
  ) %>%
  rename(Grupo = `Label (Grouping)`) %>%
  mutate(
    Grupo = str_trim(Grupo),
    valor = str_remove_all(valor, "[,%±]"),
    valor = str_replace_all(valor, ",", ""),
    valor = na_if(valor, ""),
    valor = as.numeric(valor)
  )

categorias_interes <- c(
  "Under 18 years", "18 to 64 years", "65 years and over",
  "Male", "Female", "White alone", "Black or African American alone",
  "Hispanic or Latino origin (of any race)",
  "Less than high school graduate", "High school graduate (includes equivalency)",
  "Bachelor's degree or higher", "Employed", "Unemployed",
  "Population for whom poverty status is determined" # NUEVA CATEGORÍA
)

df_filtrado <- df_limpio %>%
  filter(!is.na(valor), Tipo == "Estimate", Grupo %in% categorias_interes) %>%
  mutate(Tematica = case_when(
    Grupo %in% c("Under 18 years", "18 to 64 years", "65 years and over") ~ "Edad",
    Grupo %in% c("Male", "Female") ~ "Sexo",
    Grupo %in% c("White alone", "Black or African American alone",
                 "Hispanic or Latino origin (of any race)") ~ "Etnia",
    Grupo %in% c("Less than high school graduate",
                 "High school graduate (includes equivalency)",
                 "Bachelor's degree or higher") ~ "Educación",
  ))

```

```

    Grupo %in% c("Employed", "Unemployed") ~ "Empleo",
    Grupo == "Population for whom poverty status is determined" ~ "Población base"
  ))

extraer_primer_seguro <- function(x) {
  val <- x[!is.na(x)]
  if (length(val) < 1) return(NA_real_) else return(as.numeric(val[1]))
}

df_pivot <- df_filtrado %>%
  pivot_wider(
    names_from = Categoria,
    values_from = valor,
    values_fn = list,
    id_cols = c(Year, Estado, Grupo, Tematica)
  ) %>%
  mutate(across(c(`Total`, `Below poverty level`, `Percent below poverty level`),
    ~ map_dbl(.x, extraer_primer_seguro)))

df_final <- df_pivot %>%
  mutate(
    Grupo_social = case_when(
      Grupo == "18 to 64 years" ~ "Población en edad laboral",
      Grupo == "65 years and over" ~ "Personas mayores",
      Grupo == "Bachelor's degree or higher" ~ "Título universitario",
      Grupo == "Black or African American alone" ~ "Negro o afroamericano",
      Grupo == "Employed" ~ "Empleado",
      Grupo == "Female" ~ "Femenino",
      Grupo == "High school graduate (includes equivalency)" ~ "Educación secundaria",
      Grupo == "Hispanic or Latino origin (of any race)" ~ "Hispano o latino",
      Grupo == "Less than high school graduate" ~ "Educación básica",
      Grupo == "Male" ~ "Masculino",
      Grupo == "Under 18 years" ~ "Menores de edad",
      Grupo == "Unemployed" ~ "Desempleado",
      Grupo == "White alone" ~ "Blanco",
      Grupo == "Population for whom poverty status is determined" ~ "Población evaluable para pobreza"
    ),
    Fecha = as.Date(paste0("01-01-", Year), format = "%d-%m-%Y"),
    Total_poblacion = as.integer(`Total`),
    Bajo_umbral_pobreza = as.integer(`Below poverty level`),
    Porcentaje_pobreza = `Percent below poverty level` / 100
  ) %>%
  select(
    Fecha,
    Estado,
    Grupo_social,
    Tematica,
    Total_poblacion,
    Bajo_umbral_pobreza,
    Porcentaje_pobreza
  )

write_csv(df_final, "../data/df_pobreza_limpio.csv")

```

```
}
```

3. Estimación robusta para el año 2020: Dado que el año 2020 no cuenta con datos oficiales completos debido a la pandemia, se ha implementado una **estimación robusta** basada en la **mediana filtrada con control de dispersión**, aplicada por grupo social y estado. Se excluyen los registros del año 2020 y se seleccionan los **últimos 4 años válidos** (2016–2019), siempre que el **porcentaje de pobreza sea inferior al 80%** y la **desviación estándar no supere 0.5**.

La estimación se calcula a partir de la **mediana del porcentaje de pobreza** y la **mediana de la población total**, obtenidas sobre los años seleccionados. El número estimado de personas bajo el umbral de pobreza se obtiene como el **producto de ambas medianas**, y se genera un nuevo registro etiquetado como “**Estimado**” para el año 2020 en la variable *Origen_dato*.

Este procedimiento permite **preservar la coherencia temporal del análisis** y mantener la **comparabilidad entre estados y subgrupos sociales**, incluso en ausencia de datos observados directamente para ese año.

```
df_modelo <- df_final %>%
  filter(year(Fecha) != 2020)

estimar_2020_mediana_filtrada <- function(df_grupo, n = 4, umbral_max = 0.8, max_sd = 0.5) {
  df_grupo <- df_grupo %>%
    mutate(anio = year(Fecha)) %>%
    filter(Porcentaje_pobreza <= umbral_max)

  años_validos <- sort(unique(df_grupo$anio))
  if (length(años_validos) < n) return(NULL)

  años_usados <- tail(años_validos, n)
  df_filtrado <- df_grupo %>%
    filter(anio %in% años_usados)

  sd_valor <- sd(df_filtrado$Porcentaje_pobreza, na.rm = TRUE)
  if (is.na(sd_valor) || sd_valor > max_sd) return(NULL)

  pred_pobreza <- median(df_filtrado$Porcentaje_pobreza, na.rm = TRUE)
  pred_poblacion <- round(median(df_filtrado$Total_poblacion, na.rm = TRUE))

  if (is.na(pred_pobreza) || is.na(pred_poblacion)) return(NULL)

  pred_bajo_umbral <- round(pred_pobreza * pred_poblacion)

  tibble(
    Fecha = as.Date("2020-01-01"),
    Estado = unique(df_grupo$Estado),
    Grupo_social = unique(df_grupo$Grupo_social),
    Tematica = unique(df_grupo$Tematica),
    Total_poblacion = pred_poblacion,
    Bajo_umbral_pobreza = pred_bajo_umbral,
    Porcentaje_pobreza = round(pred_pobreza, 4),
    Origen_dato = "Estimado"
  )
}
```

```
df_2020_estimado <- df_modelo %>%
  group_by(Estado, Grupo_social, Tematica) %>%
  group_split() %>%
  map_df(estimar_2020_mediana_filtrada)
```

4. Unión de estimaciones y exportación final: Este bloque une los datos originales con las **estimaciones generadas para el año 2020**, asegurando que todos los registros tengan una columna explícita llamada **Origen_dato**, que permite distinguir entre datos observados y estimados. Se filtran los registros originales para excluir el año 2020, y se combinan con los datos sintéticos previamente calculados.

El resultado es un conjunto completo que incluye tanto datos reales como estimaciones, lo que permite **mantener la continuidad temporal del análisis** sin perder trazabilidad. Finalmente, se exporta el dataset consolidado como *df_pobreza_con_estimacion.csv*, listo para ser utilizado en visualizaciones, dashboards o análisis posteriores.

```
# Añadimos la columna de origen a los datos originales (creamos explícitamente la columna)
df_final <- df_final %>%
  mutate(Origen_dato = "Original") %>%
  filter(year(Fecha) != 2020) %>%
  bind_rows(df_2020_estimado)

# Guardamos el conjunto completo con estimaciones incluidas
write_csv(df_final, "../data/df_pobreza_con_estimacion.csv")
```

```
head(df_final)
```

```
## # A tibble: 6 x 8
##   Fecha      Estado  Grupo_social Tematica Total_poblacion Bajo_umbral_pobreza
##   <date>    <chr>    <chr>        <chr>      <dbl>          <dbl>
## 1 2016-01-01 Alabama  Población e~ Poblaci~      4741329      810231
## 2 2016-01-01 Alaska   Población e~ Poblaci~      723968       71566
## 3 2016-01-01 Arizona   Población e~ Poblaci~     6771098     1107464
## 4 2016-01-01 Arkansas  Población e~ Poblaci~     2898630     499225
## 5 2016-01-01 Californ~ Población e~ Poblaci~     38513258    5525524
## 6 2016-01-01 Colorado  Población e~ Poblaci~      5420327     598736
## # i 2 more variables: Porcentaje_pobreza <dbl>, Origen_dato <chr>
```

5. Visualización exploratoria y detección de patrones: La visualización de los datos permite identificar patrones relevantes en la distribución del porcentaje de pobreza, tanto a nivel temporal como regional. Se han utilizado histogramas y boxplots para representar estas variaciones de forma clara y comparativa.

Histogramas y Boxplots por Estado y Año

Los **histogramas** muestran la evolución mensual del porcentaje de pobreza, destacando el año 2020 como estimado mediante un color azul oscuro. Esta diferenciación permite observar posibles sesgos o desplazamientos en la distribución durante ese año, en comparación con los datos observados.

```
df_final <- df_final %>%
  mutate(Estimado = year(Fecha) == 2020)

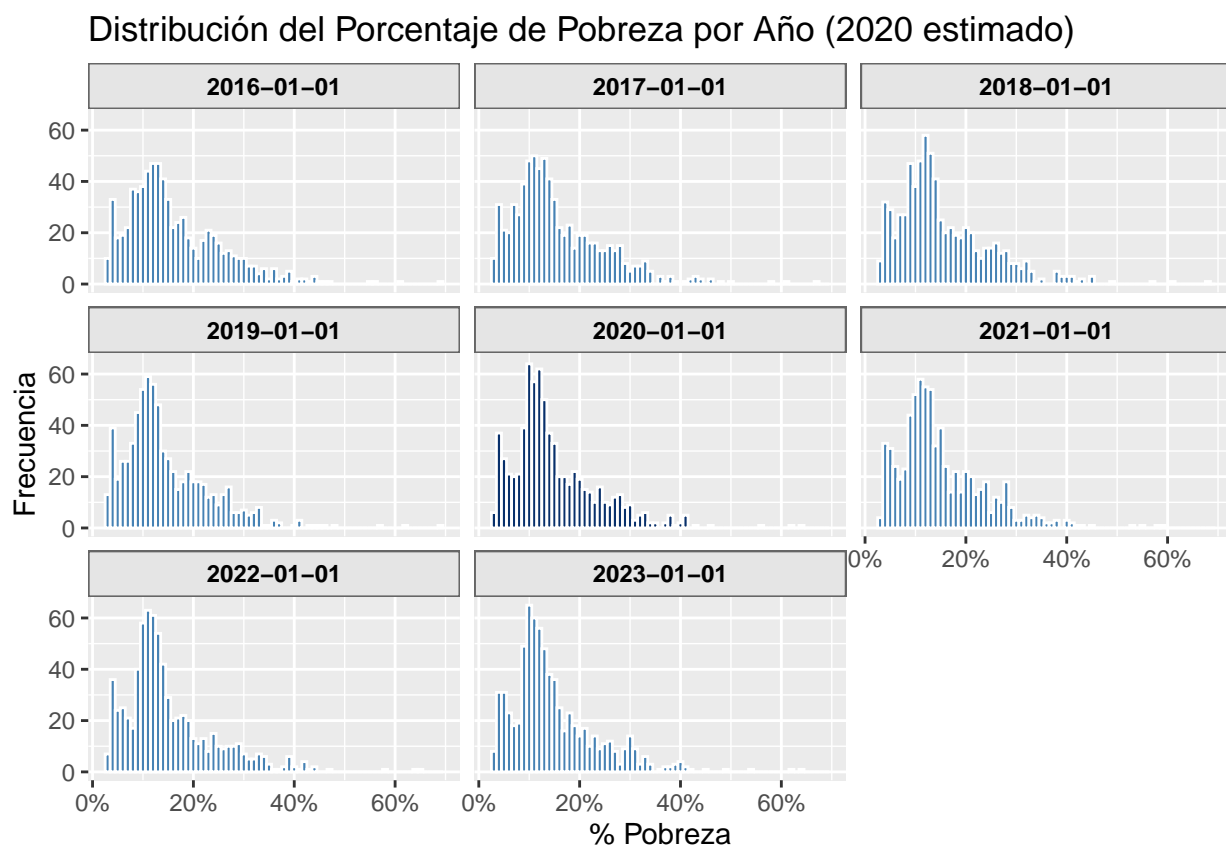
anotacion_2020 <- df_final %>%
  filter(Estimado) %>%
```

```

slice(1)

df_final %>%
  ggplot(aes(x = Porcentaje_pobreza, fill = Estimado)) +
  geom_histogram(binwidth = 0.01, color = "white") +
  facet_wrap(~Fecha) +
  scale_fill_manual(values = c("FALSE" = "steelblue", "TRUE" = "#08306b")) +
  labs(title = "Distribución del Porcentaje de Pobreza por Año (2020 estimado)",
       x = "% Pobreza", y = "Frecuencia") +
  scale_x_continuous(labels = scales::percent_format()) +
  theme(strip.background = element_rect(fill = "gray90", color = "gray40"),
        strip.text = element_text(face = "bold", color = "black"),
        legend.position = "none") #+

```



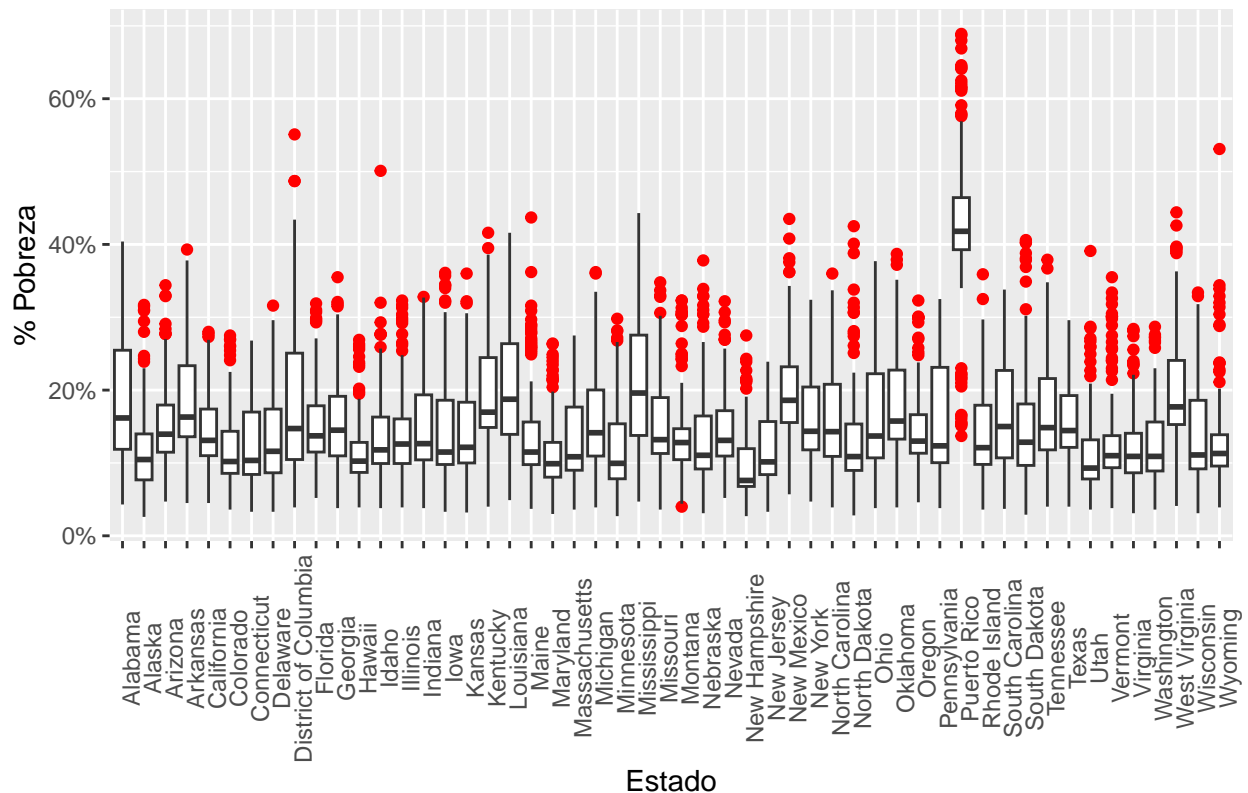
Los **boxplots** por estado revelan una alta variabilidad regional, con diferencias marcadas en la dispersión de los datos. Algunos estados presentan una concentración más estrecha en torno a la mediana, mientras que otros exhiben una mayor amplitud intercuartílica y presencia de outliers extremos.

```

df_final %>%
  ggplot(aes(x = Estado, y = Porcentaje_pobreza)) +
  geom_boxplot(outlier.colour = "red") +
  labs(title = "Variabilidad del Porcentaje de Pobreza por Estado", y = "% Pobreza") +
  theme(axis.text.x = element_text(angle = 90)) +
  scale_y_continuous(labels = scales::percent_format())

```

Variabilidad del Porcentaje de Pobreza por Estado



Se identifican **asimetrías** en la distribución, especialmente en estados con porcentajes de pobreza elevados, lo que sugiere la existencia de subgrupos particularmente vulnerables o dinámicas locales que amplifican la desigualdad.

Justificación metodológica:

El histograma permite visualizar la forma de la distribución, detectar asimetrías, y delimitar los rangos típicos de pobreza por fecha.

El boxplot, por su parte, facilita la detección de valores atípicos (outliers) y la comparación entre estados, destacando aquellos con mayor dispersión o comportamiento extremo.

Esta visualización es clave para orientar análisis posteriores, como la segmentación por subgrupos sociales o la evaluación de políticas focalizadas.

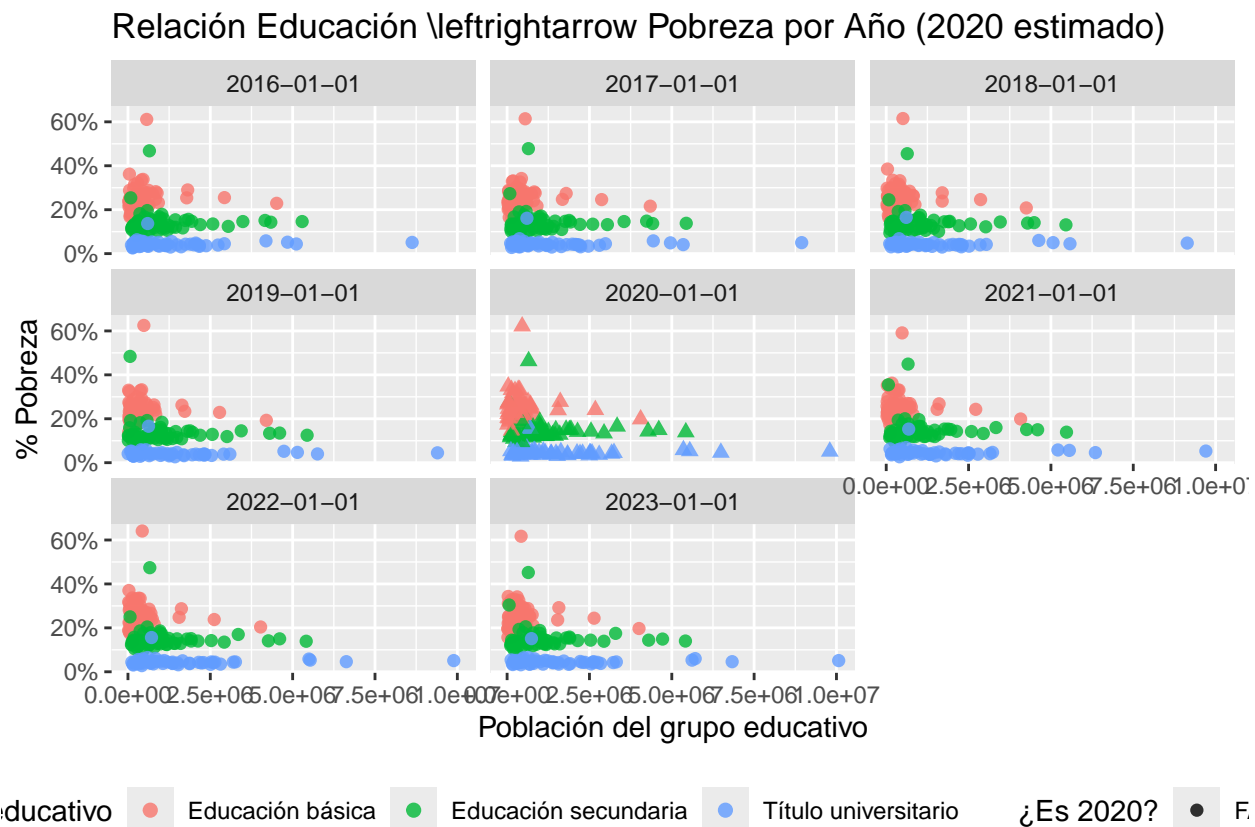
Dispersión: Educación vs Pobreza:

Este conjunto de gráficos de dispersión explora cómo los niveles educativos se relacionan con el porcentaje de pobreza, y si dicha relación varía a lo largo del tiempo. Cada panel representa un año específico, mostrando la relación entre el tamaño poblacional de cada grupo educativo y su nivel de pobreza.

Los puntos están codificados por color según el grupo educativo, y por forma según si el dato corresponde al año 2020 (triángulo) o a años anteriores (círculo). Esta distinción visual permite comparar datos observados con estimaciones o reales, y detectar posibles cambios estructurales.

```
df_final %>%
  filter(Tematica == "Educación") %>%
  mutate(Año = year(Fecha),
         es_2020 = Año == 2020) %>%
  ggplot(aes(x = Total_poblacion, y = Porcentaje_pobreza)) +
```

```
geom_point(aes(color = Grupo_social, shape = es_2020), size = 2, alpha = 0.8) +
facet_wrap(~Fecha) +
labs(
  title = "Relación Educación \\leftarrow Pobreza por Año (2020 estimado)",
  x = "Población del grupo educativo",
  y = "% Pobreza",
  color = "Grupo educativo",
  shape = "¿Es 2020?"
) +
scale_y_continuous(labels = scales::percent_format()) +
scale_shape_manual(values = c("FALSE" = 16, "TRUE" = 17)) +
theme(legend.position = "bottom")
```



Observaciones clave:

- 1- **Relación negativa entre educación y pobreza:** A mayor nivel educativo, menor porcentaje de pobreza. Los grupos con educación básica o sin secundaria presentan consistentemente niveles más altos de pobreza.
- 2- **Persistencia temporal:** Esta relación se mantiene estable en las distintas fechas analizadas, lo que sugiere una correlación estructural entre nivel educativo y vulnerabilidad económica.
- 3- **Variabilidad entre fechas:** Aunque la tendencia general se conserva, algunos paneles muestran ligeras variaciones en la dispersión, posiblemente vinculadas a factores coyunturales o políticas específicas implementadas durante el año 2020.

Justificación analítica:

Este análisis se presenta como un **ejemplo aplicado a la temática de educación**, pero la misma lógica puede extenderse a otras dimensiones disponibles en el dashboard, como **Edad, Sexo, Etnia, Educación, Empleo**. Esto permite construir un enfoque multivariado que revele interacciones complejas entre factores de exclusión y pobreza.

Detectar si los grupos con menor nivel educativo presentan sistemáticamente mayor pobreza ayuda a identificar patrones persistentes de desigualdad, fundamentales para el diseño de políticas públicas focalizadas. Además, si se observa una mejora en la correlación tras un año concreto, podría vincularse a intervenciones estructurales como programas de apoyo educativo, transferencias condicionadas o cambios en el mercado laboral.

Outliers extremos: Detectar estados con valores anómalos por grupo social y año

Este bloque identifica **valores anómalos en el porcentaje de pobreza**, definidos como aquellos que superan el **percentil 95** dentro de cada grupo social y año. El objetivo es detectar **estados con niveles de pobreza excepcionalmente altos**, que podrían requerir atención específica en el análisis o en el diseño de políticas públicas.

Observaciones clave:

Se han filtrado los registros que presentan **porcentajes de pobreza superiores al 95% de su distribución**, agrupados por fecha y grupo social.

Un ejemplo destacado es **Puerto Rico**, que muestra **valores persistentemente elevados** en grupos como **Desempleado** y **Educación básica**, con porcentajes superiores al 64% en varios años consecutivos.

Estos outliers coinciden con lo observado en los **boxplots por estado**, reforzando la idea de que ciertas regiones presentan **niveles estructurales de vulnerabilidad**.

```
df_outliers <-df_final %>%
  group_by(Fecha, Grupo_social) %>%
  filter(Porcentaje_pobreza > quantile(Porcentaje_pobreza, 0.95, na.rm = TRUE)) %>%
  arrange(desc(Porcentaje_pobreza)) %>%
  select(Fecha, Estado, Grupo_social, Porcentaje_pobreza)
```

```
df_outliers %>%
  arrange(Fecha) %>%
  head(5) %>%
  kable(format = "latex", booktabs = TRUE)
```

Fecha	Estado	Grupo_social	Porcentaje_pobreza
2016-01-01	Puerto Rico	Desempleado	0.689
2016-01-01	Puerto Rico	Educación básica	0.611
2016-01-01	Puerto Rico	Menores de edad	0.564
2016-01-01	District of Columbia	Desempleado	0.551
2016-01-01	Puerto Rico	Educación secundaria	0.468

Justificación analítica:

La detección de outliers permite **focalizar el análisis en zonas críticas**, donde la pobreza alcanza niveles extremos.

Este enfoque ayuda a **priorizar intervenciones**, evaluar el impacto de políticas previas y diseñar estrategias más eficaces para grupos sociales específicos.

Además, puede servir como base para **alertas tempranas** o segmentaciones en dashboards interactivos, facilitando la toma de decisiones basada en evidencia.

Justificación conceptual de los gráficos del dashboard (Flex Dashboard): Los gráficos incluidos en el dashboard han sido seleccionados estratégicamente para responder a las preguntas clave del análisis de pobreza. Cada visualización cumple una función narrativa dentro del recorrido del usuario, desde la contextualización inicial hasta la toma de decisiones informadas. No representan un caso específico, sino que ilustran cómo se estructuraría el dashboard para cualquier estado o temática seleccionada.

Vector de colores personalizado

Para garantizar una visualización clara, coherente y perceptivamente efectiva, se ha definido un vector de colores personalizado que codifica cada temática y grupo social del dashboard. Esta paleta cumple varias funciones clave:

Propósito de la codificación

Diferenciación temática: Cada dimensión social (Edad, Sexo, Etnia, Educación, Empleo) tiene un color base que agrupa sus subcategorías.

Consistencia visual: Los colores se mantienen constantes en todos los gráficos, facilitando la lectura transversal.

Agrupación perceptiva: Se aplica el principio de semejanza de la psicología Gestalt, permitiendo al usuario identificar rápidamente los elementos relacionados.

Aplicación en el dashboard

En gráficos de líneas, barras y circulares, los colores permiten distinguir grupos sociales sin necesidad de leer la leyenda.

En los KPIs, los colores refuerzan el significado semántico (por ejemplo, verde para mejora, naranja para empeoramiento).

En interacciones dinámicas, la codificación por color facilita el seguimiento de un grupo específico en múltiples visualizaciones.

Este vector de colores no solo mejora la estética del dashboard, sino que también potencia su capacidad comunicativa, reduciendo la carga cognitiva y aumentando la memorabilidad de los datos.

```
colores_dashboard <- c(
  # Edad
  "Edad" = "#4E79A7",
  "Población en edad laboral" = "#A0CBE8",
  "Personas mayores" = "#86BCB6",
  "Menores de edad" = "#499894",

  # Sexo
  "Sexo" = "#D37295",
  "Masculino" = "#B07AA1",
  "Femenino" = "#D4A6C8",

  # Etnia
  "Etnia" = "#79706E",
  "Negro o afroamericano" = "#BAB0AC",
  "Hispano o latino" = "#9D7660",
  "Blanco" = "#D7B5A6",

  # Educación
  "Educación" = "#59A14F",
  "Educación básica" = "#8CD17D",
  "Educación secundaria" = "#B6992D",
```

```
"Titulo universitario" = "#F1CE63",

# Empleo
"Empleo" = "#E15759",
"Empleado" = "#F28E2B" ,
"Desempleado" = "#EDC948"
)
```

Evolución de la pobreza por grupo social

Este gráfico de líneas representa la evolución temporal de la pobreza según grupos de edad, utilizando como ejemplo el estado de Texas. Cada línea muestra la tasa de pobreza dentro de un grupo específico (menores de edad, población en edad laboral, personas mayores) a lo largo del tiempo. Se incluyen etiquetas en los extremos para facilitar la lectura de los valores iniciales y finales de cada serie.

¿Qué revela esta visualización?

Permite **comparar dinámicas entre grupos sociales**, identificando cuáles han mejorado, empeorado o mantenido niveles estables de pobreza.

Destaca **tendencias persistentes**: por ejemplo, los menores de edad suelen presentar tasas más altas que otros grupos.

Refuerza la lectura con **codificación por color** y anotaciones clave, facilitando la interpretación incluso sin consultar la leyenda.

```
estado_seleccionado <- "Texas"
tematica_seleccionada <- "Edad"

datos <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado, Tematica == tematica_seleccionada) %>%
  dplyr::group_by(Fecha, Grupo_social) %>%
  dplyr::summarise(
    Total = sum(Total_poblacion, na.rm = TRUE),
    Pobre = sum(Bajo_umbral_pobreza, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  dplyr::mutate(Pct = Pobre / Total)

if (nrow(datos) == 0) {
  stop("Sin datos para la combinación seleccionada.")
}

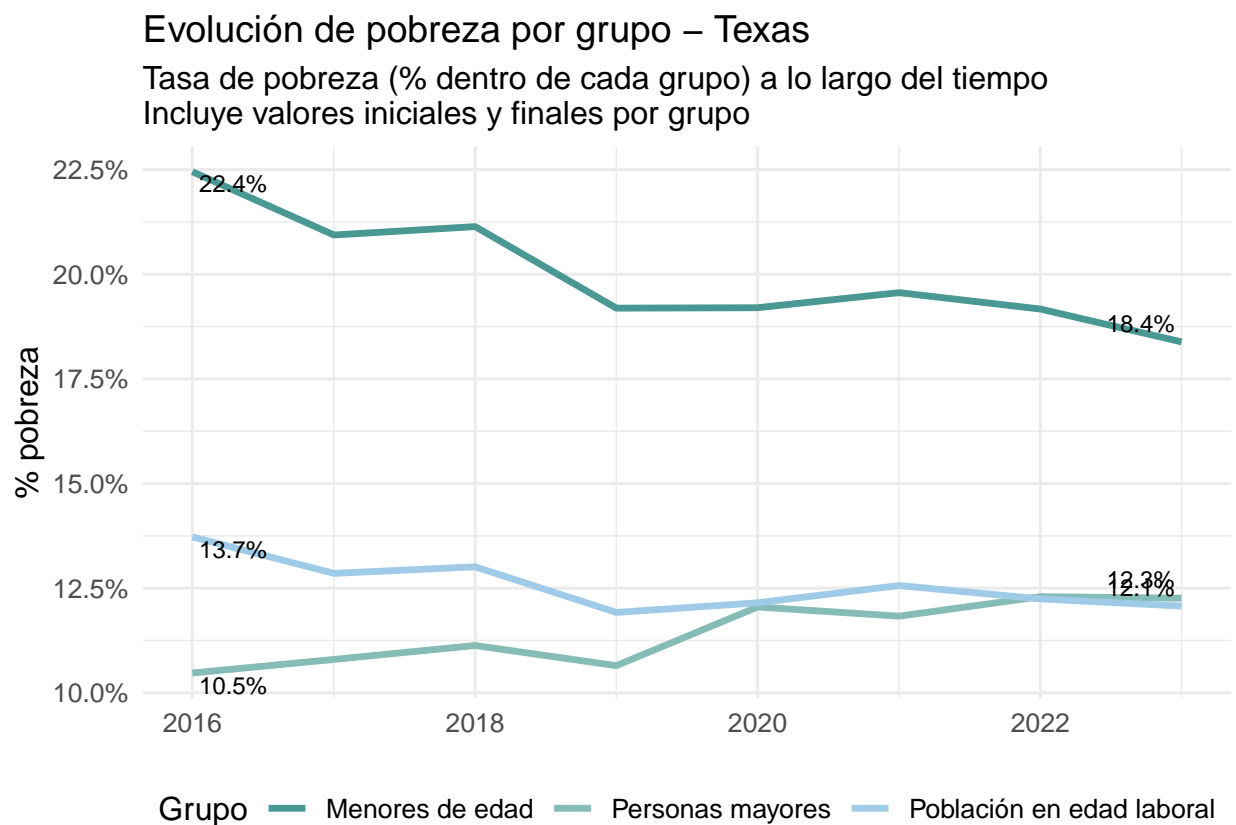
extremos <- datos %>%
  dplyr::group_by(Grupo_social) %>%
  dplyr::summarise(
    Fecha_ini = first(Fecha),
    Pct_ini = first(Pct),
    Fecha_fin = last(Fecha),
    Pct_fin = last(Pct),
    .groups = "drop"
  )

ggplot(datos, aes(x = Fecha, y = Pct, color = Grupo_social)) +
  geom_line(linewidth = 1.2) +
  geom_text(data = extremos,
```

```

aes(x = Fecha_ini, y = Pct_ini, label = scales::percent(Pct_ini, accuracy = 0.1)),
color = "black", hjust = -0.1, vjust = 1.2, size = 3, inherit.aes = FALSE) +
geom_text(data = extremos,
aes(x = Fecha_fin, y = Pct_fin, label = scales::percent(Pct_fin, accuracy = 0.1)),
color = "black", hjust = 1.1, vjust = -0.6, size = 3, inherit.aes = FALSE) +
scale_color_manual(values = colores_dashboard) +
scale_y_continuous(labels = scales::percent_format()) +
labs(
title = paste("Evolución de pobreza por grupo -", estado_seleccionado),
subtitle = "Tasa de pobreza (% dentro de cada grupo) a lo largo del tiempo\nIncluye valores iniciales y finales",
x = NULL, y = "% pobreza", color = "Grupo"
) +
theme_minimal(base_size = 12) +
theme(legend.position = "bottom")

```



Distribución de la pobreza por grupo social – año más reciente

Este gráfico circular muestra la composición total de personas en situación de pobreza, desglosada por grupo social en el año más reciente disponible. A diferencia de los gráficos que analizan tasas internas dentro de cada grupo, esta visualización se enfoca en el volumen absoluto de personas pobres que pertenecen a cada categoría.

¿Qué revela esta visualización?

Permite **dimensionar el peso relativo de cada grupo** dentro del total de pobreza, independientemente de su tamaño poblacional o de su tasa específica.

Destaca que ciertos grupos, como la población en edad laboral o los menores de edad, pueden representar una **proporción significativa del total**, lo que orienta la priorización de políticas públicas.

Refuerza la lectura con **codificación por color** y etiquetas integradas, facilitando la interpretación directa de porcentajes y volúmenes sin necesidad de consultar leyendas externas.

Complementa los gráficos de evolución temporal, ofreciendo una **visión transversal** que conecta tasas relativas con distribución absoluta.

```
estado_seleccionado <- "Texas"
tematica_seleccionada <- "Edad"

ultimo <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado, Tematica == tematica_seleccionada) %>%
  dplyr::summarise(Fecha = max(Fecha, na.rm = TRUE)) %>%
  dplyr::pull(Fecha)

if (length(ultimo) != 1 || !is.finite(ultimo)) {
  stop("No hay fecha válida para esta combinación.")
}

datos <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado,
                Tematica == tematica_seleccionada,
                Fecha == ultimo) %>%
  dplyr::group_by(Grupo_social) %>%
  dplyr::summarise(
    Pobre = sum(Bajo_umbral_pobreza, na.rm = TRUE),
    Total = sum(Total_poblacion, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  dplyr::mutate(Pobre = dplyr::if_else(is.na(Pobre), 0, Pobre)) %>%
  dplyr::filter(Pobre > 0)

if (nrow(datos) == 0) {
  stop(paste0("Sin datos de pobreza para ", format(ultimo, "%Y-%m-%d"), "."))
}

suma_pobre <- sum(datos$Pobre, na.rm = TRUE)
if (suma_pobre == 0) {
  stop("No hay personas bajo pobreza en la fecha seleccionada.")
}

datos <- datos %>%
  dplyr::mutate(
    Pct = Pobre / suma_pobre,
    etiqueta = paste0(scales::percent(Pct, accuracy = 0.1), "\n(n=", scales::comma(Pobre), ")")
  ) %>%
  dplyr::arrange(desc(Pct))

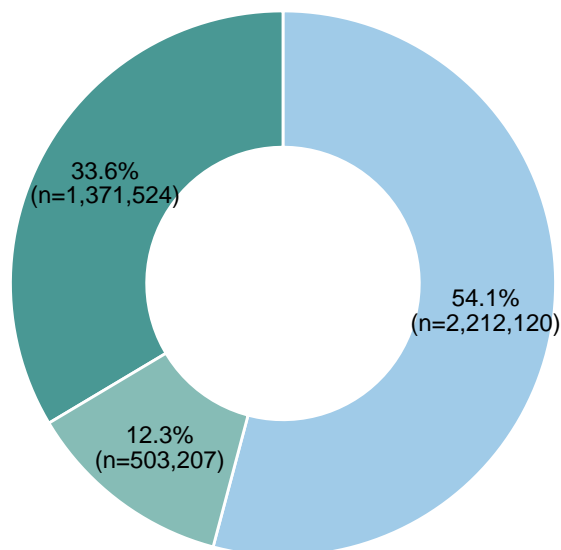
paleta <- colores_dashboard[names(colores_dashboard) %in% datos$Grupo_social]

ggplot(datos, aes(x = 2, y = Pct, fill = Grupo_social)) +
  geom_col(width = 1, color = "white") +
  coord_polar(theta = "y") +
```

```
xlim(0.5, 2.5) +
geom_text(aes(label = etiqueta),
           position = position_stack(vjust = 0.5),
           size = 3, lineheight = 0.9) +
scale_fill_manual(values = paleta) +
labs(
  title = paste("Distribución de personas bajo pobreza -", lubridate::year(ultimo)),
  subtitle = "Participación por grupo dentro del total de personas pobres (volumen abs.)",
  x = NULL, y = NULL, fill = "Grupo",
  caption = "Nota: este gráfico muestra la composición total; la línea anterior refleja tasas de pobreza"
) +
theme_void() +
theme(
  legend.position = "bottom",
  plot.subtitle = element_text(size = 10, face = "italic"),
  plot.caption = element_text(size = 8, color = "grey40")
)
```

Distribución de personas bajo pobreza – 2023

Participación por grupo dentro del total de personas pobres (volumen abs.)



Grupo Menores de edad Personas mayores Población en edad laboral

e gráfico muestra la composición total; la línea anterior refleja tasas de pobreza dentro de cada grupo.

Aportación al total de pobreza por temática

Este gráfico de barras muestra la **distribución de personas pobres según distintas temáticas sociales** en el año más reciente disponible. Cada barra representa el **porcentaje del total de pobreza** que corresponde a una categoría específica, como etnia, edad, sexo o nivel educativo.

¿Qué revela esta visualización?

Permite **comparar el peso relativo de cada temática** en la composición total de la pobreza, destacando cuáles concentran más personas afectadas.

Las etiquetas integradas muestran tanto el **porcentaje como el número absoluto de personas**, lo que facilita una lectura directa sin necesidad de consultar leyendas externas.

Este gráfico es útil para **identificar ejes prioritarios de intervención**: si ciertas temáticas agrupan mayor cantidad de personas pobres, podrían requerir políticas más focalizadas en inclusión, equidad o acceso a oportunidades.

Complementa las visualizaciones anteriores al ofrecer una **perspectiva temática transversal**, que ayuda a entender cómo se distribuye la pobreza más allá de los grupos demográficos clásicos.

```
estado_seleccionado <- "Texas"

tematicas_deseadas <- c("Etnia", "Edad", "Sexo", "Educación")

ultimo <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado) %>%
  dplyr::summarise(Fecha = max(Fecha, na.rm = TRUE)) %>%
  dplyr::pull(Fecha)

if (length(ultimo) != 1 || !is.finite(ultimo)) {
  stop("No hay fecha válida para esta combinación.")
}

datos <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado,
                Fecha == ultimo,
                Tematica %in% tematicas_deseadas) %>%
  dplyr::group_by(Tematica) %>%
  dplyr::summarise(Pobre = sum(Bajo_umbral_pobreza, na.rm = TRUE), .groups = "drop") %>%
  dplyr::mutate(
    Pct = Pobre / sum(Pobre),
    etiqueta = paste0(scales::percent(Pct, accuracy = 0.1), "\n(n=", scales::comma(Pobre), ")")
  ) %>%
  dplyr::arrange(desc(Pct))

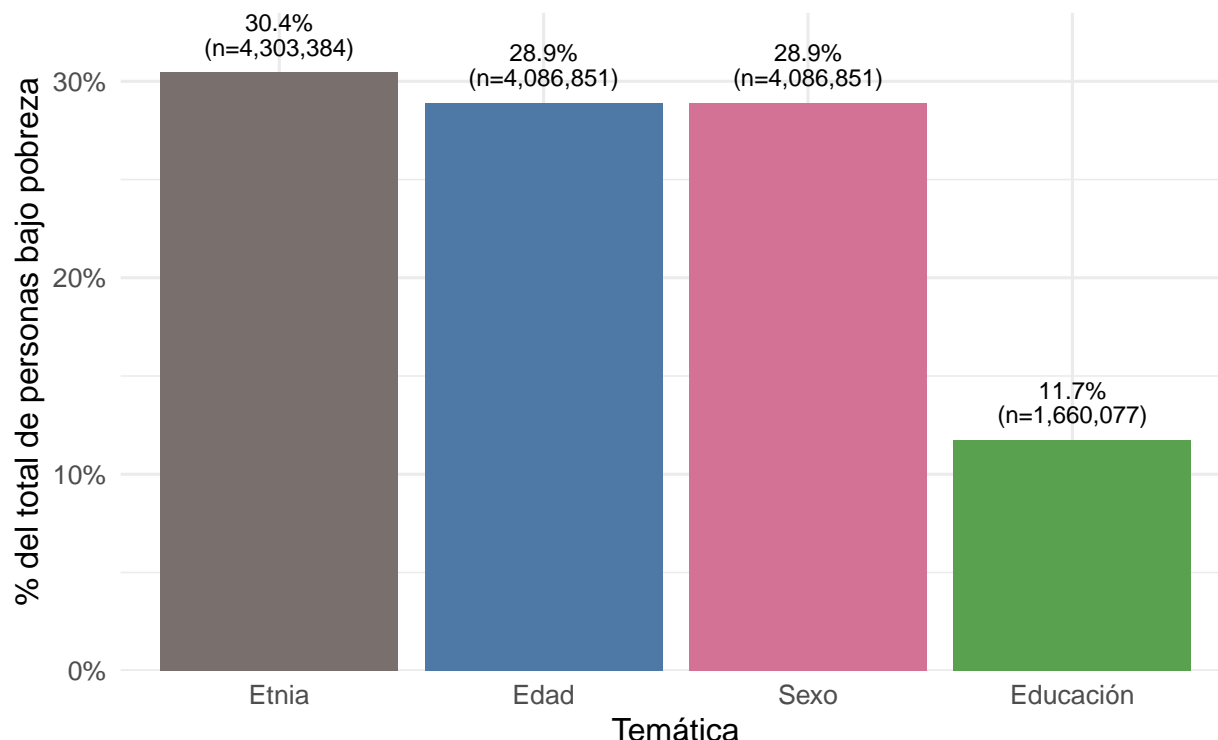
if (nrow(datos) == 0) {
  stop("Sin datos de pobreza para las temáticas seleccionadas.")
}

paleta <- colores_dashboard[names(colores_dashboard) %in% datos$Tematica]

ggplot(datos, aes(x = reorder(Tematica, -Pct), y = Pct, fill = Tematica)) +
  geom_col() +
  geom_text(aes(label = etiqueta), vjust = -0.4, size = 3, color = "black", lineheight = 0.9) +
  scale_fill_manual(values = paleta) +
  scale_y_continuous(labels = scales::percent_format(), expand = expansion(mult = c(0, 0.1))) +
  labs(
    title = paste("Aportación al total de pobreza -", lubridate::year(ultimo)),
    subtitle = "Distribución de personas pobres por temática (suma = 100%)",
    x = "Temática", y = "% del total de personas bajo pobreza"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "none")
```

Aportación al total de pobreza – 2023

Distribución de personas pobres por temática (suma = 100%)



Distribución dentro de la población pobre por situación laboral

Este gráfico de barras muestra la **composición interna de la población pobre** según su situación laboral en el año más reciente disponible. Cada barra representa el porcentaje de personas pobres que se encuentran en cada grupo laboral, como empleados o desempleados.

¿Qué revela esta visualización?

Destaca que una **gran mayoría de personas pobres están empleadas**, lo que sugiere que el empleo no garantiza condiciones suficientes para salir de la pobreza. Esto puede estar vinculado a **bajos salarios, empleos informales o falta de protección social**.

Las etiquetas integradas permiten leer de forma directa tanto el **porcentaje como el número absoluto de personas**, facilitando la interpretación sin necesidad de consultar leyendas externas.

Este gráfico es clave para **replantear políticas laborales y sociales**: si la pobreza afecta mayoritariamente a personas con empleo, se requieren estrategias que aborden la calidad del trabajo, el acceso a derechos laborales y los ingresos mínimos.

Complementa las visualizaciones anteriores al ofrecer una **mirada funcional sobre la pobreza**, centrada en la relación entre empleo y vulnerabilidad económica.

```
estado_seleccionado <- "Texas"
```

```
ultimo <- df_final %>%  
  dplyr::filter(Estado == estado_seleccionado) %>%  
  dplyr::summarise(Fecha = max(Fecha, na.rm = TRUE)) %>%  
  dplyr::pull(Fecha)
```



```

if (length(ultimo) != 1 || !is.finite(ultimo)) {
  stop("No hay fecha válida para esta combinación.")
}

datos <- df_final %>%
  dplyr::filter(Estado == estado_seleccionado,
    Tematica == "Empleo",
    Fecha == ultimo) %>%
  dplyr::group_by(Grupo_social) %>%
  dplyr::summarise(Pobre = sum(Bajo_umbral_pobreza, na.rm = TRUE), .groups = "drop") %>%
  dplyr::mutate(Pct = Pobre / sum(Pobre)) %>%
  dplyr::arrange(desc(Pct))

if (nrow(datos) == 0) {
  stop("Sin datos de 'Empleo' para la fecha actual.")
}

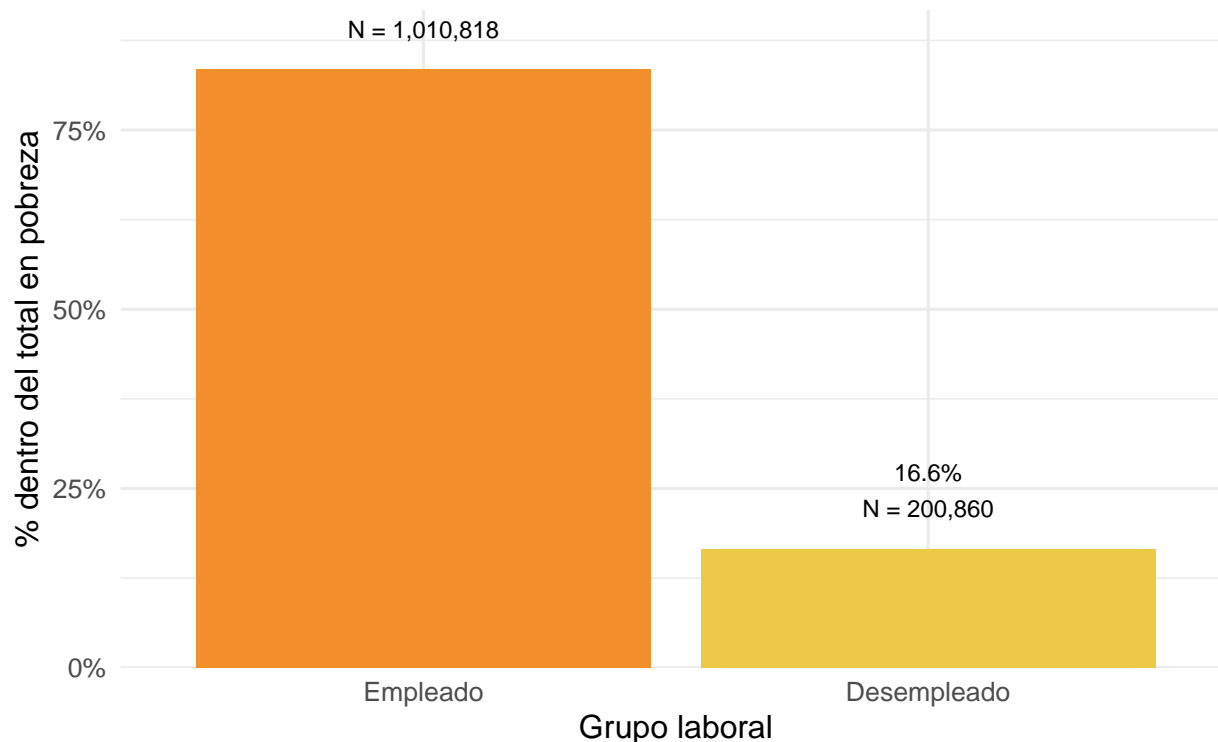
paleta <- colores_dashboard[names(colores_dashboard) %in% datos$Grupo_social]

ggplot(datos, aes(x = reorder(Grupo_social, -Pct), y = Pct, fill = Grupo_social)) +
  geom_col() +
  geom_text(aes(label = paste0(scales::percent(Pct, accuracy = 0.1), "\nN = ", scales::comma(Pobre))),
    vjust = -0.6, size = 3.3) +
  scale_fill_manual(values = paleta) +
  scale_y_continuous(labels = scales::percent_format(), expand = expansion(mult = c(0, 0.1))) +
  labs(
    title = "Distribución dentro de la población pobre",
    subtitle = "Por situación laboral",
    x = "Grupo laboral",
    y = "% dentro del total en pobreza"
  ) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "none")

```

Distribución dentro de la población pobre

Por situación laboral



Indicadores clave de pobreza evolución temporal

Este bloque presenta los **principales indicadores de pobreza** en el estado seleccionado, calculados sobre la población evaluable para pobreza. Se incluyen tres métricas clave: el porcentaje actual, el cambio anual y la variación acumulada en cinco años.

% pobreza actual (2023):

Representa la proporción de personas bajo pobreza en el año más reciente disponible. Este valor resume el estado actual del fenómeno en términos relativos.

Crecimiento anual (desde 2022):

Indica una reducción interanual en la tasa de pobreza, lo que sugiere una mejora o empeoramiento reciente en las condiciones socioeconómicas o en la cobertura de programas sociales.

Crecimiento en 5 años (desde 2018):

Refleja una **tendencia descendente o creciente**, lo que puede atribuirse a cambios estructurales, políticas públicas efectivas o dinámicas demográficas.

Estos KPIs permiten **monitorear el progreso** en la lucha contra la pobreza, identificar puntos de inflexión y evaluar el impacto de intervenciones a lo largo del tiempo. Su lectura integrada con los gráficos anteriores ofrece una **visión completa y estratégica** del fenómeno.

```
df_kpi_texas <- df_final %>%
  dplyr::filter(Estado == "Texas", Grupo_social == "Población evaluable para pobreza") %>%
  dplyr::mutate(Año = lubridate::year(Fecha)) %>%
  dplyr::group_by(Año) %>%
  dplyr::summarise(
```

```

Pob_evaluable = sum(Total_poblacion, na.rm = TRUE),
Pob_pobre = sum(Bajo_umbral_pobreza, na.rm = TRUE),
Porcentaje_pobreza = Pob_pobre / Pob_evaluable,
.groups = "drop"
) %>%
dplyr::arrange(Año)

año_actual <- max(df_kpi_texas$Año)
año_anterior <- año_actual - 1
año_hace_5 <- max(df_kpi_texas$Año[df_kpi_texas$Año <= año_actual - 5])

valor_actual <- df_kpi_texas %>% dplyr::filter(Año == año_actual) %>% dplyr::pull(Porcentaje_pobreza)
valor_anterior <- df_kpi_texas %>% dplyr::filter(Año == año_anterior) %>% dplyr::pull(Porcentaje_pobreza)
valor_hace_5 <- df_kpi_texas %>% dplyr::filter(Año == año_hace_5) %>% dplyr::pull(Porcentaje_pobreza)

crecimiento_anual <- if(length(valor_anterior) > 0) {
  (valor_actual - valor_anterior) / valor_anterior
} else { NA }

crecimiento_5_anos <- if(length(valor_hace_5) > 0) {
  (valor_actual - valor_hace_5) / valor_hace_5
} else { NA }

cat("% pobreza actual (", año_actual, "): ",
    scales::percent(valor_actual, accuracy = 0.1), "\n")

## % pobreza actual ( 2023 ): 13.7%

if (!is.na(crecimiento_anual)) {
  cat("Crecimiento anual (desde ", año_anterior, "): ",
      scales::percent(crecimiento_anual, accuracy = 0.1), "\n")
} else {
  cat("Crecimiento anual: No disponible\n")
}

## Crecimiento anual (desde 2022 ): -2.2%

if (!is.na(crecimiento_5_anos)) {
  cat("Crecimiento en 5 años (desde ", año_hace_5, "): ",
      scales::percent(crecimiento_5_anos, accuracy = 0.1), "\n")
} else {
  cat("Crecimiento en 5 años: No disponible\n")
}

## Crecimiento en 5 años (desde 2018 ): -8.2%

```