vote by mail

Julia Pullano

2025-09-24

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Attaching package: 'janitor'
##
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
##
##
##
## Attaching package: 'maps'
##
##
## The following object is masked from 'package:purrr':
##
##     map
```

Provisional Ballot Usage

```r
# Calculate provisional usage by state
mail_by_state <- eavs %>%
  group_by(state_full) %>%
  summarise(
    mail_ballots = sum(f1d, na.rm = TRUE),  # Provisional ballots cast
    valid_ballots       = sum(f1a, na.rm = TRUE),  # Total voters
    .groups = "drop"
  ) %>%
  mutate(
    mail_rate = (mail_ballots / valid_ballots) * 100
  ) %>%
  arrange(state_full) %>%
  rename(State = state_full,
         `Mail Ballots Cast` = mail_ballots,
```

1

```
        `Valid Ballots Cast` = valid_ballots,
        `Mail Usage (%)` = mail_rate)

# Display the table
knitr::kable(
  mail_by_state,
  caption = "Mail Ballot Usage by State (EAVS 2024)",
  digits = 2
)
```

Table 1: Mail Ballot Usage by State (EAVS 2024)

State	Mail Ballots Cast	Valid Ballots Cast	Mail Usage (%)
ALABAMA	126055	2272911	5.55
ALASKA	48744	340981	14.30
AMERICAN SAMOA	189	10215	1.85
ARIZONA	2597974	3477975	74.70
ARKANSAS	23843	1122278	2.12
CALIFORNIA	0	16164330	0.00
COLORADO	0	3240754	0.00
CONNECTICUT	118362	1820891	6.50
DELAWARE	33659	514367	6.54
DISTRICT OF COLUMBIA	0	328871	0.00
FLORIDA	2945893	10999125	26.78
GEORGIA	268751	5297500	5.07
GUAM	104	30283	0.34
HAWAII	0	522236	0.00
IDAHO	177211	917469	19.32
ILLINOIS	1016208	5717147	17.77
INDIANA	1603815	2986839	53.70
IOWA	0	1674011	0.00
KANSAS	147276	1342102	10.97
KENTUCKY	116324	2086090	5.58
LOUISIANA	119694	2021588	5.92
MAINE	215242	842447	25.55
MARYLAND	744244	3028813	24.57
MASSACHUSETTS	1173112	3512930	33.39
MICHIGAN	2017704	5706503	35.36
MINNESOTA	328830	3271069	10.05
MISSISSIPPI	0	1225176	0.00
MISSOURI	178526	3126837	5.71
MONTANA	0	612423	0.00
NEBRASKA	260899	965145	27.03
NEVADA	0	1486297	0.00
NEW HAMPSHIRE	92945	829090	11.21
NEW JERSEY	828200	4321921	19.16
NEW MEXICO	111167	927923	11.98
NEW YORK	836987	8389626	9.98
NORTH CAROLINA	298269	5756106	5.18
NORTH DAKOTA	89429	371974	24.04
NORTHERN MARIANA ISLANDS	915	12610	7.26
OHIO	1060096	5851625	18.12

| State	Mail Ballots Cast	Valid Ballots Cast	Mail Usage (%)
OKLAHOMA	98548	1573274	6.26
OREGON	0	2269608	0.00
PENNSYLVANIA	1933707	7074875	27.33
PUERTO RICO	132157	1283628	10.30
RHODE ISLAND	51995	522164	9.96
SOUTH CAROLINA	98782	2566404	3.85
SOUTH DAKOTA	159335	435739	36.57
TENNESSEE	86904	3090161	2.81
TEXAS	384221	11488820	3.34
U.S. VIRGIN ISLANDS	613	15952	3.84
UTAH	0	1466896	0.00
VERMONT	234038	361604	64.72
VIRGINIA	474332	4511853	10.51
WASHINGTON	0	3949810	0.00
WEST VIRGINIA	22377	769206	2.91
WISCONSIN	570657	3434185	16.62
WYOMING	38217	271123	14.10

```r
# Get US states map data
us_states <- map_data("state")

# Prepare data for join (convert to lowercase state names)
mail_map_data <- mail_by_state %>%
  mutate(region = tolower(State))

# Join with map data
map_data_joined <- us_states %>%
  left_join(mail_map_data, by = "region")

# Plot
ggplot(map_data_joined, aes(long, lat, group = group, fill = `Mail Usage (%)`)) +
  geom_polygon(color = "white") +
  coord_fixed(1.3) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  labs(
    title = "Mail Ballot Usage by State (EAVS 2024)",
    fill = "% of Valid Ballots Cast"
  ) +
  theme_minimal()
```

Mail Ballot Usage by State (EAVS 2024)
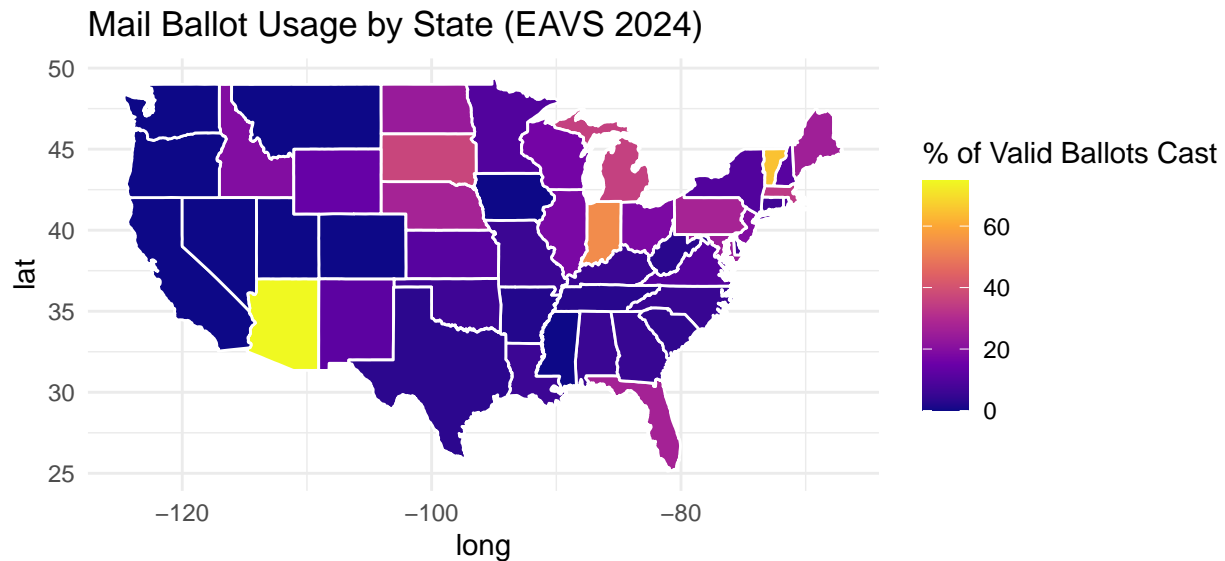
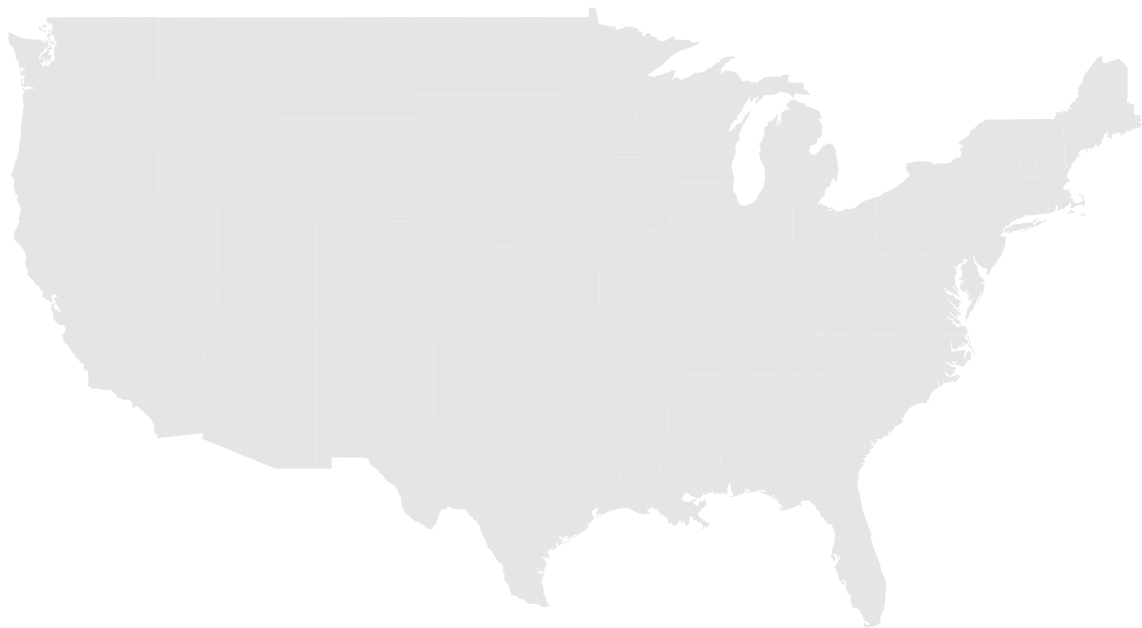## Provisional Ballot Usage Rate – County Map

```r
# Step 1: Summarise provisional usage by county (jurisdiction)
mail_by_state <- eavs %>%
  group_by(state_full, jurisdiction_name) %>%
  summarise(
    mail_ballots = sum(f1d, na.rm = TRUE),
    valid_ballots       = sum(f1a, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    mail_rate = (mail_ballots / valid_ballots) * 100,
    region  = tolower(state_full),
    subregion = tolower(jurisdiction_name)
  )

# Step 2: Get US counties map
us_states <- map_data("state")

# Step 3: Join map with data
map_state_joined <- us_states %>%
  left_join(mail_by_state, by = c("region", "subregion"))

# Step 4: Plot
ggplot(map_state_joined, aes(long, lat, group = group, fill = mail_rate)) +
  geom_polygon(color = NA) +
  coord_fixed(1.3) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  labs(
    title = "Provisional Ballot Usage by State (EAVS 2024)",
    fill = "% of Valid Ballots"
  ) +
  theme_void()
```

Provisional Ballot Usage by State (EAVS 2024)

Provisional Ballot Usage – County (Jurisdiction) Map (FIPS-based)

```r
library(sf)
```

```
## Linking to GEOS 3.11.1, GDAL 3.6.2, PROJ 9.1.1; sf_use_s2() is TRUE
```

```r
library(tigris)
```

```
## To enable caching of data, set `options(tigris_use_cache = TRUE)`
## in your R script or .Rprofile.
```

```r
library(stringr)
options(tigris_use_cache = TRUE)

# 1) Summarize by county FIPS with safe division
provisional_by_fips <- eavs %>%
  mutate(
    # Ensure 5-digit FIPS as characters (keep leading zeros)
    fips5 = case_when(
      is.na(fips_code) ~ NA_character_,
      TRUE ~ str_pad(as.character(fips_code), width = 5, side = "left", pad = "0")
    )
  ) %>%
  group_by(fips5) %>%
  summarise(
    provisional_ballots = sum(f1e, na.rm = TRUE),
    valid_ballots       = sum(f1a, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    provisional_rate = if_else(valid_ballots > 0,
                               100 * provisional_ballots / valid_ballots,
```

```
                                NA_real_)  # avoid Inf/NaN
  )

# 2) County shapes (drop territories)
counties_sf <- tigris::counties(cb = TRUE, class = "sf") %>%
  filter(!STATEFP %in% c("60","66","69","72","78"))  # AS, GU, MP, PR, VI
```

Retrieving data for the year 2024

| |

```
# 3) Join and project
county_map_data <- counties_sf %>%
  left_join(provisional_by_fips, by = c("GEOID" = "fips5"))

county_map_data_proj <- st_transform(county_map_data, 5070)

# 4) Compute a finite max for the legend; fall back if none
finite_max <- suppressWarnings(max(county_map_data_proj$provisional_rate, na.rm = TRUE))
use_limits <- is.finite(finite_max) && finite_max > 0

# 5) Optional visibility: print quick diagnostics to the console
message("Counties matched with data: ",
        sum(!is.na(county_map_data_proj$provisional_rate)), " of ",
        nrow(county_map_data_proj))
```

Counties matched with data: 67 of 3144

```
message("Any zero valid-ballot counties? ",
        any(provisional_by_fips$valid_ballots == 0, na.rm = TRUE))
```

Any zero valid-ballot counties? TRUE

```
# 6) Plot (conditionally apply limits)
p <- ggplot(county_map_data_proj) +
  geom_sf(aes(fill = provisional_rate), color = NA) +
  labs(
    title = "Provisional Ballot Usage by County (EAVS 2024)",
    subtitle = "Rate = (Provisional Ballots Cast ÷ Valid Ballots Cast) × 100",
    fill = "% Provisional",
    caption = "Source: EAVS 2024"
  ) +
  theme_void() +
  theme(legend.position = "right",
        plot.title = element_text(face = "bold"))

if (use_limits) {
  p <- p +
    scale_fill_viridis_c(
      option = "plasma",
      na.value = "grey90",
      limits = c(0, finite_max),
      labels = scales::label_number(accuracy = 0.1)
    )
} else {
  # Let ggplot pick breaks; still set palette and NA color
```
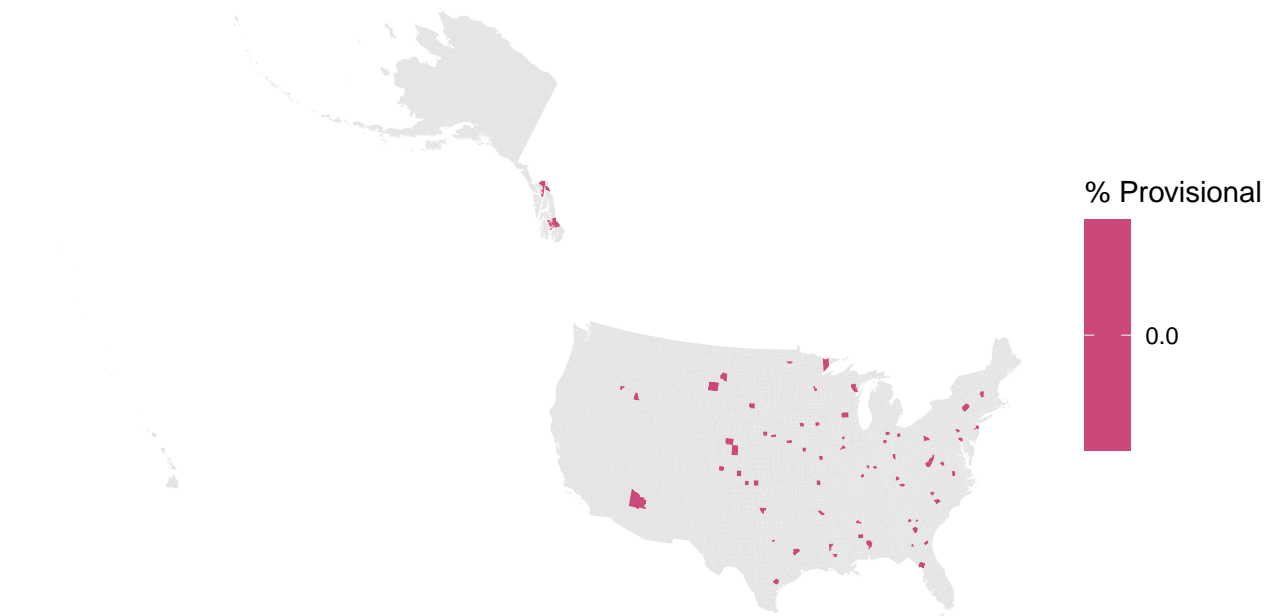
```
  p <- p +
    scale_fill_viridis_c(
      option = "plasma",
      na.value = "grey90",
      labels = scales::label_number(accuracy = 0.1)
    )
}


p
```

Provisional Ballot Usage by County (EAVS 2024)

Rate = (Provisional Ballots Cast ÷ Valid Ballots Cast) × 100

% Provisional

0.0

Source: EAVS 2024