Extra credit

INFO 2950 - Spring 2023

Lina Liu

5/10/23

Setup

Load packages and data:

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- 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 becor
```

```
library(usmap)
library(ggplot2)
library(wesanderson)
library(scales)
```

```
Attaching package: 'scales'

The following object is masked from 'package:purrr':
```

```
    discard

The following object is masked from 'package:readr':

    col_factor
```

```
#| label: load-data
childcare_costs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidy
```

```
Rows: 34567 Columns: 61
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (61): county_fips_code, study_year, unr_16, funr_16, munr_16, unr_20to64...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
counties <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday
```

```
Rows: 3144 Columns: 4
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): county_name, state_name, state_abbreviation
dbl (1): county_fips_code

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Extra credit

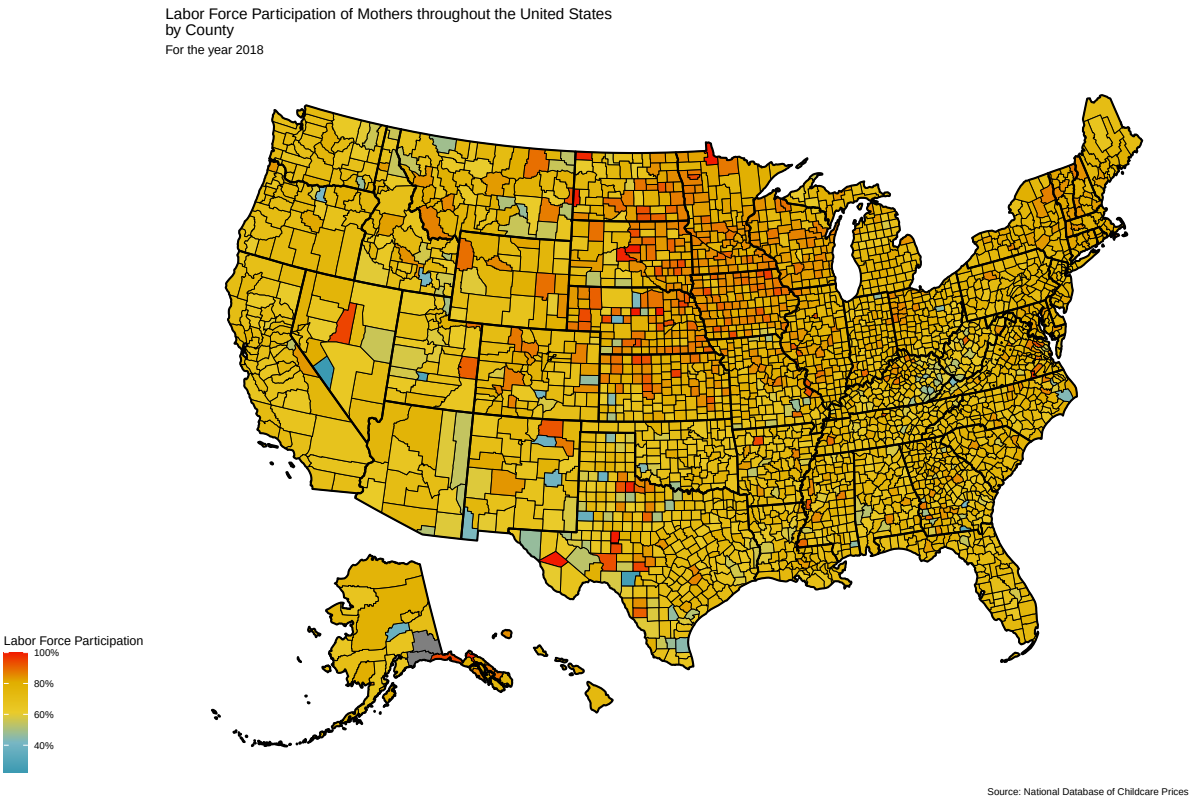
Research Question: What is the labor force participation of mothers who have children throughout the United States in the year 2018?

```
#| label: labor-county
labor_part_2018 <- childcare_costs |>
  # select relevant columns to use for analysis
  select(county_fips_code, study_year, flfpr_20to64_under6, flfpr_20to64_6to17) |>
```

```r
  # filter so we only use the year 2018
  filter(study_year == "2018") |>
  # did this justttt in case if there's n/a values
  drop_na() |>
  # combine the column for labor participation of mothers with children under the age
  # of 6 alongside the labor participation of mothers with children 6 to 17.
  pivot_longer(
    cols = -c("county_fips_code", "study_year"),
    names_transform = parse_number,
    values_to = "pct"
  ) |>
  # get rid of the name column, which is kindaaa unnecessary :p
  select(-name) |>
  # because there's two rows of the same area (one for the labor participation
  # percentage of mothers with children under 6 age group and the other one for
  # the 6-17 age group), I decided to average them together to get the total
  # labor participation
  group_by(county_fips_code) |>
  summarize(pct = mean(pct)) |>
  # aligning the fips code of the data so it matches w/ the fips code and column
  # name that the usmap package gave. I added a leading zero in the beginning &
  # changed the column name to fips
  mutate(
    county_fips_code = as.character(county_fips_code),
    fips = if_else(nchar(county_fips_code) == 4,
                                paste0("0", county_fips_code),
                                county_fips_code),
    pct = pct/100)

#| label: labor-states
state_joined <- inner_join(x = childcare_costs, y = counties,
             by = "county_fips_code")

labor_part_2018_states <- state_joined |>
  select(study_year,
         flfpr_20to64_under6,
         flfpr_20to64_6to17,
         state_abbreviation) |>
  filter(study_year == "2018") |>
  drop_na() |>
  pivot_longer(
```

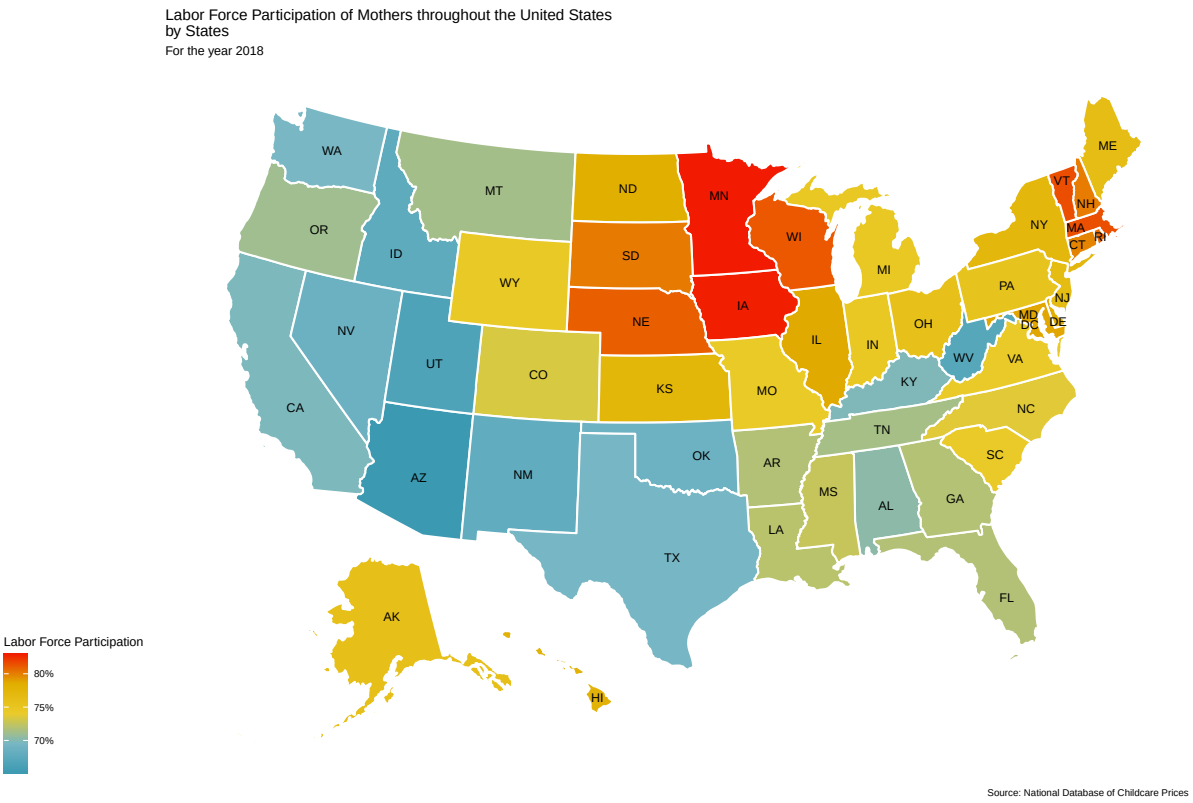
```
    cols = -c("study_year", "state_abbreviation"),
    names_transform = parse_number,
    values_to = "pct"
  ) |>
  group_by(state_abbreviation) |>
  summarize(pct = mean(pct)) |>
  mutate(
    state = state_abbreviation,
    pct = pct/100)
```

```
#| label: visualization-by-county
# I followed this tutorial for the US map:
# https://jtr13.github.io/cc19/different-ways-of-plotting-u-s-map-in-r.html

plot_usmap(data = labor_part_2018,
           values = "pct",
           linewidth = 0.0005)  +
  labs(title = "Labor Force Participation of Mothers throughout the United States
by County",
       subtitle = "For the year 2018",
       caption = "Source: National Database of Childcare Prices",
       fill = "Labor Force Participation") +
  theme(panel.background = element_blank(),
        legend.position = "left") +
  scale_fill_gradientn(colours = wes_palette("Zissou1", 100,
                                             type = "continuous"),
                       labels = percent) +
# I referred to this source to make my state lines thicker:
# https://stackoverflow.com/questions/72543606/how-to-make-census-region-
# border-thicker-on-usmap-packake-r
  geom_polygon(data = usmapdata::us_map(regions = "states"),
               aes(x, y, group = group),
               fill = NA,
               linewidth = 0.7,
               color = "black")
```




```
#| label: visualization-by-state
state_map <- plot_usmap(data = labor_part_2018_states,
           values = "pct",
           linewidth = 0.7,
           labels = TRUE,
           color = "white") +
  labs(title = "Labor Force Participation of Mothers throughout the United States
by States",
       subtitle = "For the year 2018",
       caption = "Source: National Database of Childcare Prices",
       fill = "Labor Force Participation") +
  theme(panel.background = element_blank(),
        legend.position = "left") +
  scale_fill_gradientn(colours = wes_palette("Zissou1", 100,
                                             type = "continuous"),
                       labels = percent)

# Referred to this as source to make state abbr font smaller:
```

```
# https://stackoverflow.com/questions/60806822/how-do-i-change-state-or-counties-
# label-sizes-in-r-with-the-function-usmap
state_map$layers[[2]]$aes_params$size <- 3
print(state_map)
```



Purpose

- The purpose of the visualization is to show how the labor force of the mothers in 2018 compare throughout the United States by different states and counties. Moreover, I thought it was fitting for Mother's day.
- It helps to answer the question of the labor force participation of mothers who have children throughout the United States in the year 2018.
- I chose this type of visualization because it incorporates different areas of the United States, which is perfect for the data. It also kinda looked fun and it's something I haven't done before!

Happy Mother's Day!! :)

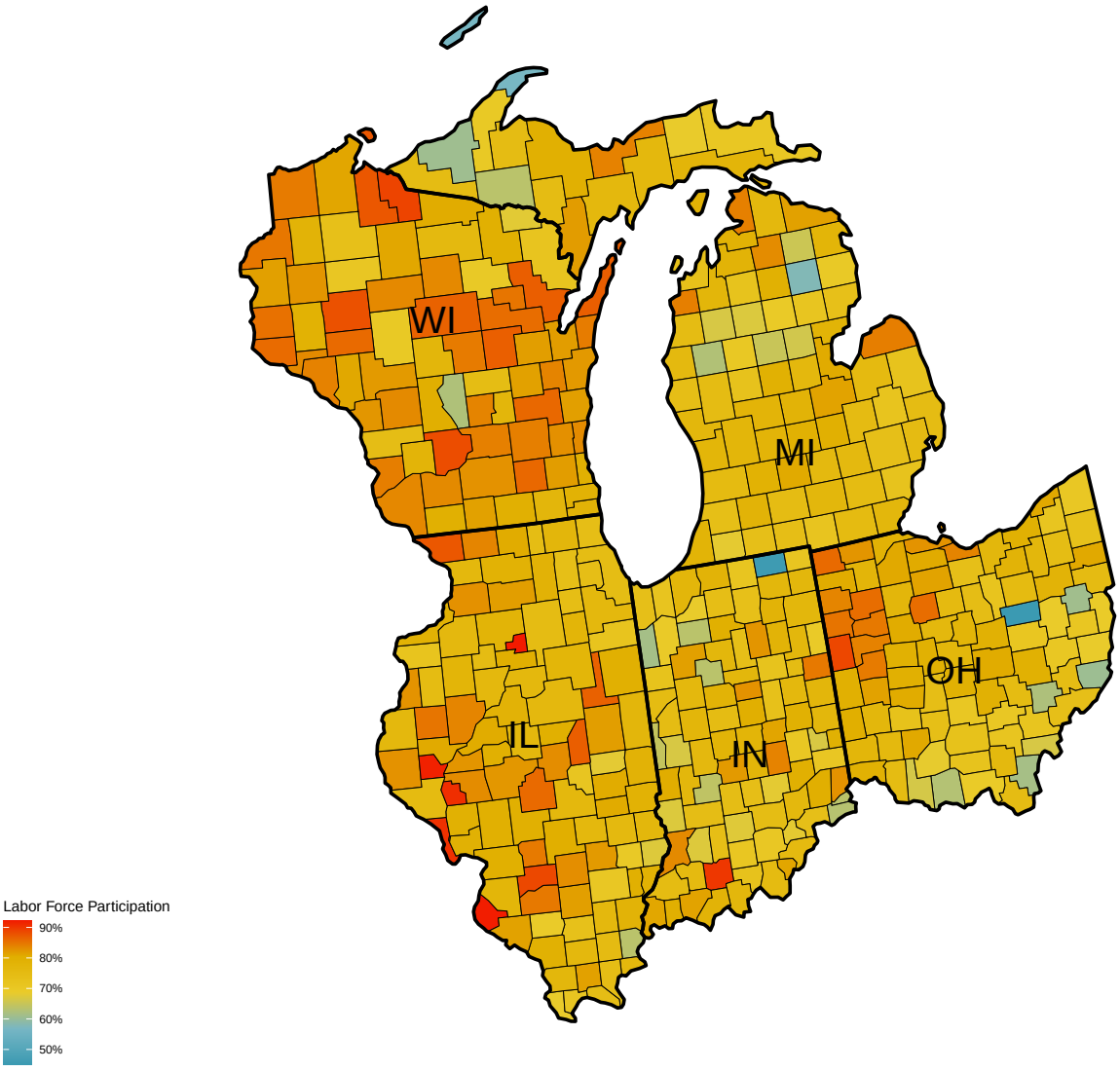
Closer view of the map

The following maps show the labor force participation of mothers by region (for fun!). I only added the code chunk for first map since the code is repetitive.

```
east_north_central <- plot_usmap(data = labor_part_2018,
           values = "pct",
           linewidth = 0.0005,
           include = .east_north_central,
           labels = TRUE)  +
  labs(title = "Labor Force Participation of Mothers in the East North Central region",
       subtitle = "For the year 2018",
       caption = "Source: National Database of Childcare Prices",
       fill = "Labor Force Participation") +
  theme(panel.background = element_blank(),
        legend.position = "left") +
  scale_fill_gradientn(colours = wes_palette("Zissou1", 100,
                                             type = "continuous"),
                       labels = percent) +
    geom_polygon(data = usmapdata::us_map(regions = "states",
                                          include = .east_north_central),
                 aes(x, y, group = group),
                 fill = NA,
                 linewidth = 1,
                 color = "black")

east_north_central$layers[[2]]$aes_params$size <- 8
print(east_north_central)
```

Labor Force Participation of Mothers in the East North Central region

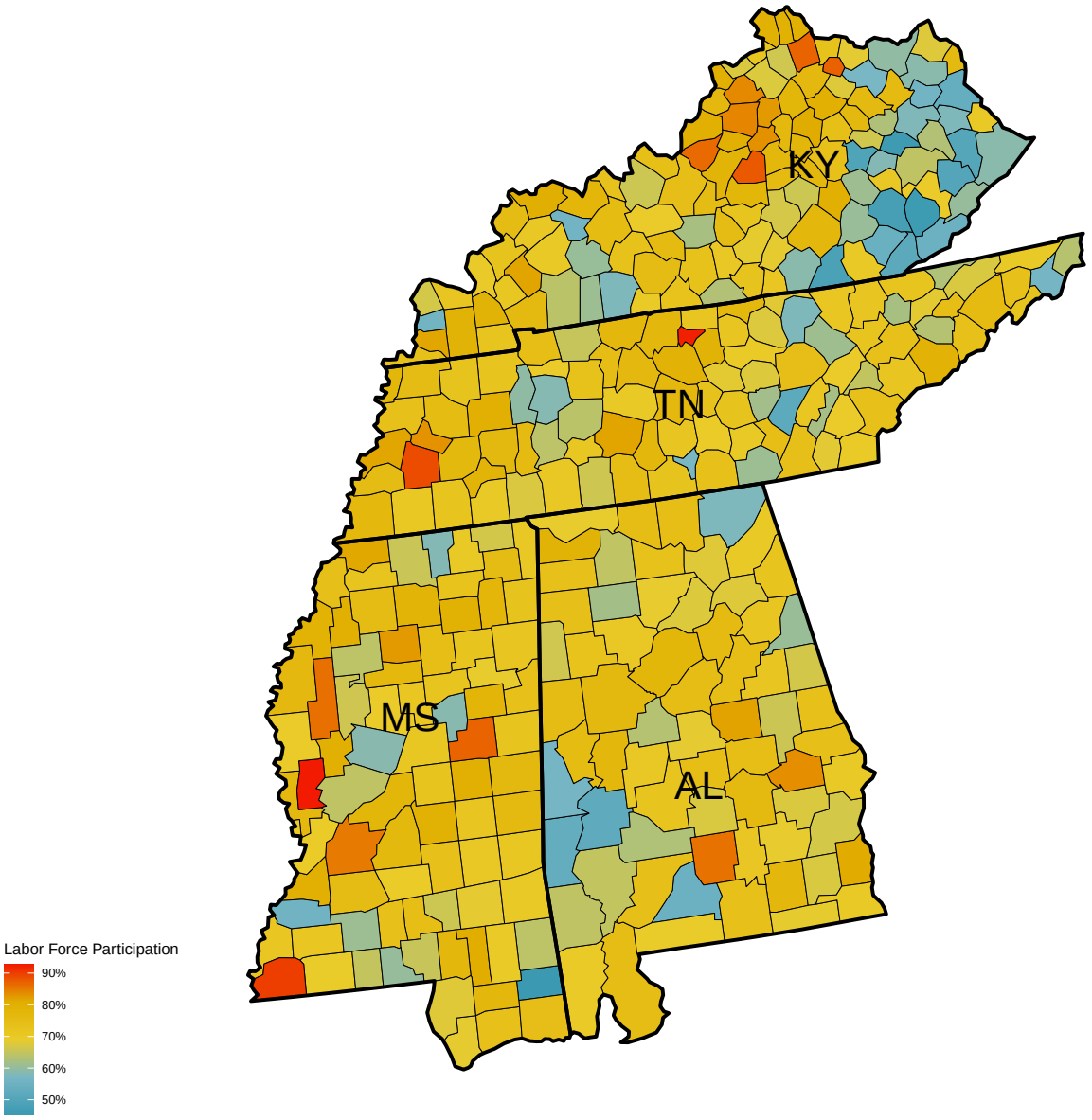
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the East South Central region

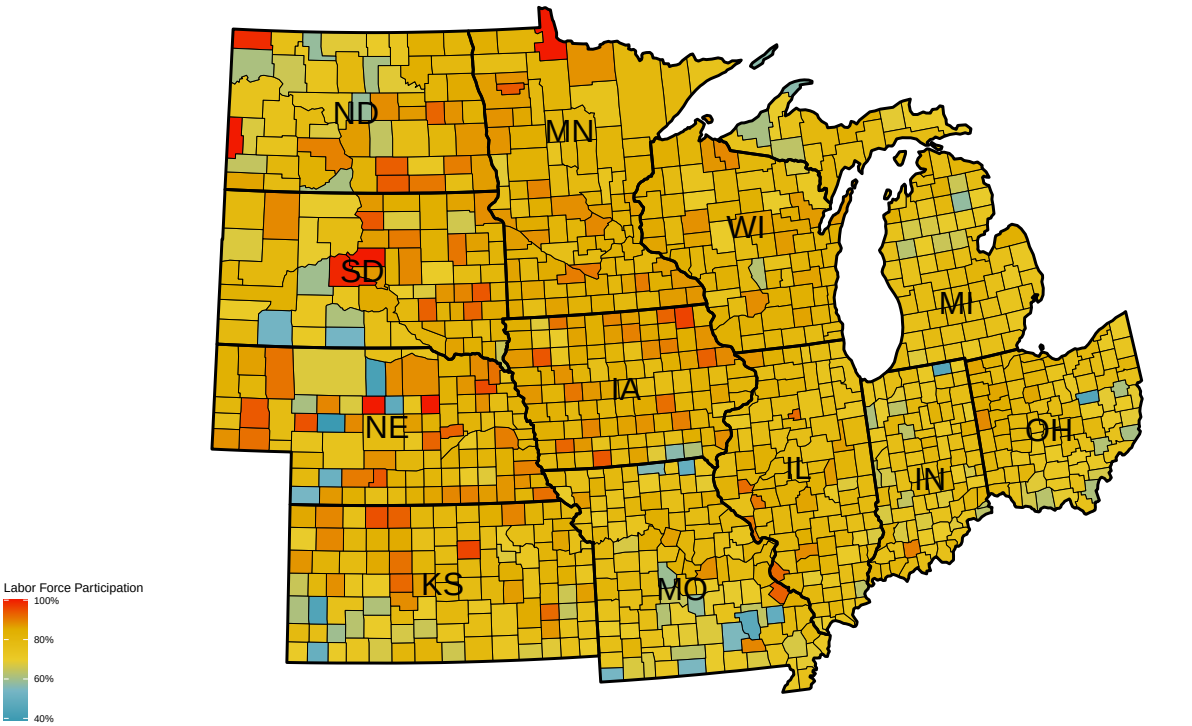
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the Mid−west region

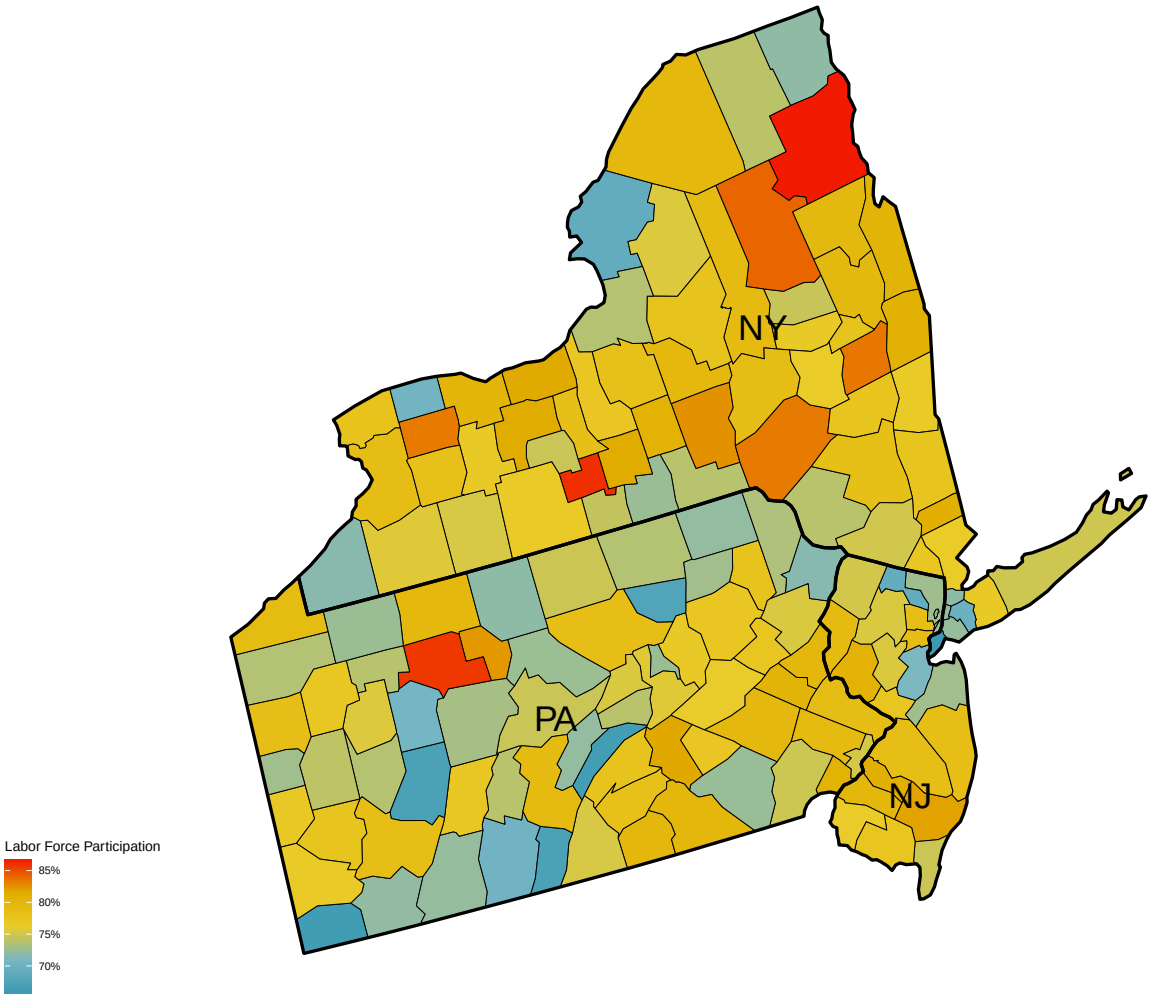
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the Mid−Atlantic region

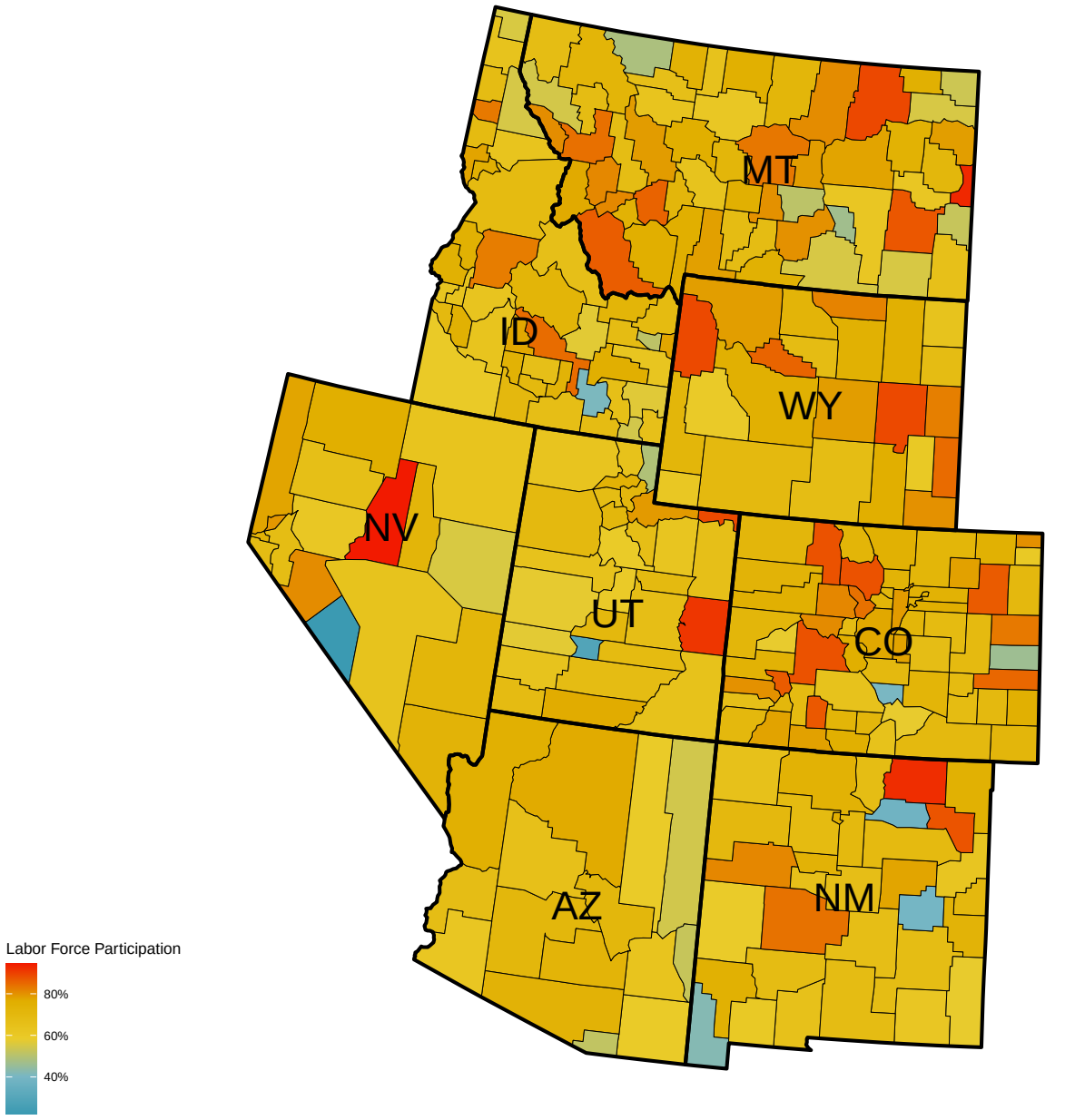
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the Mountain region

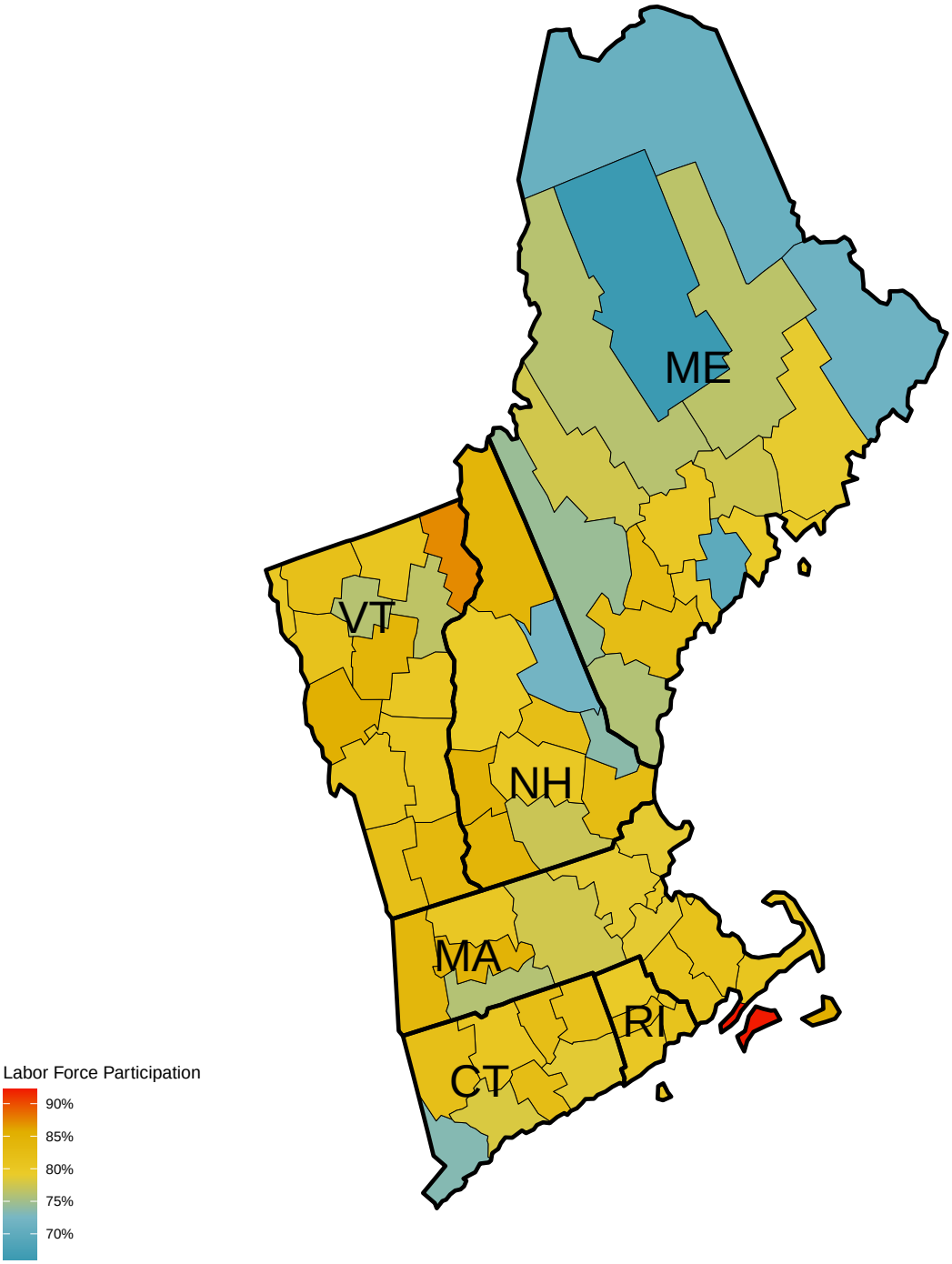
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the New England region

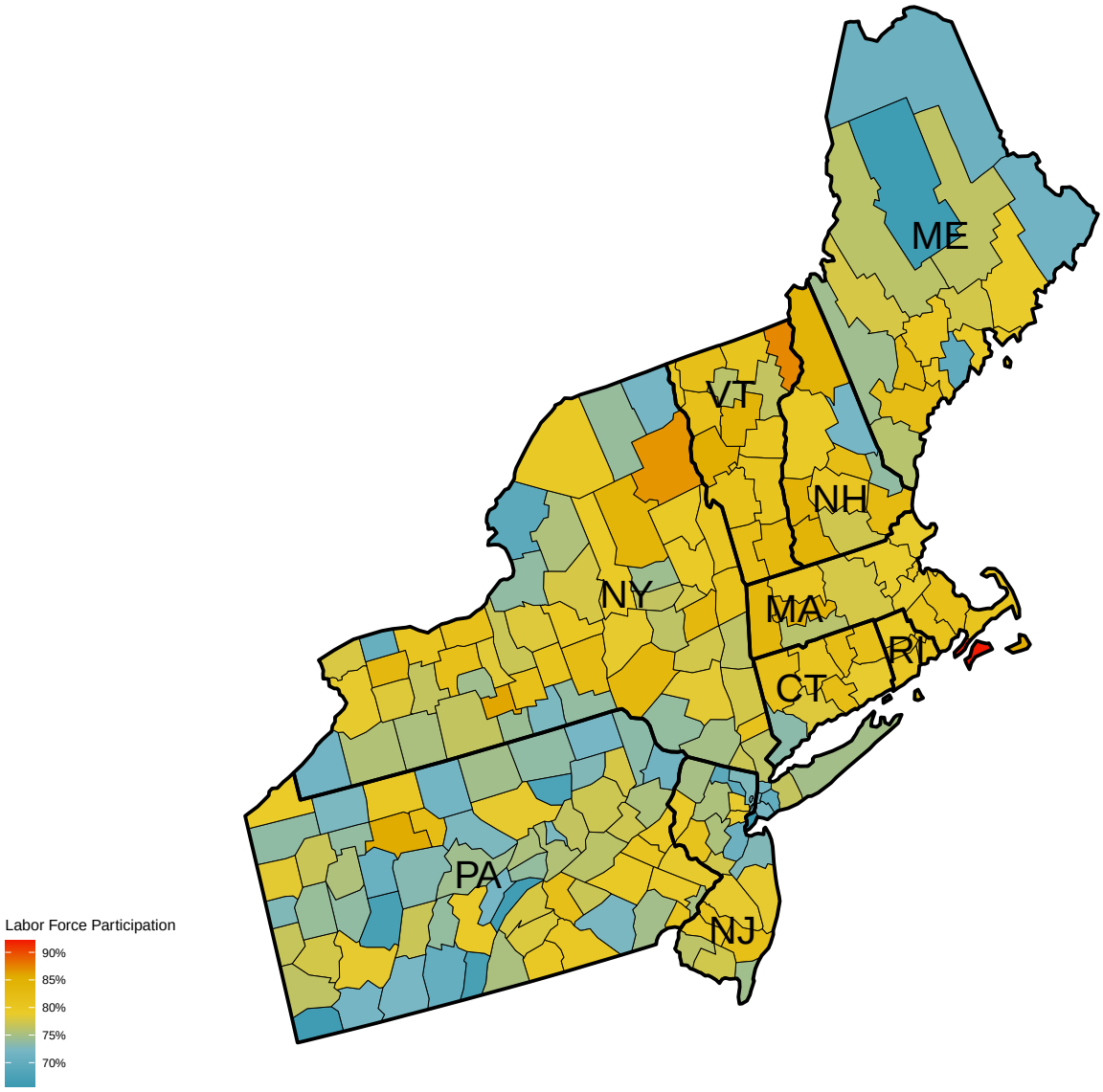
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the Northeast region

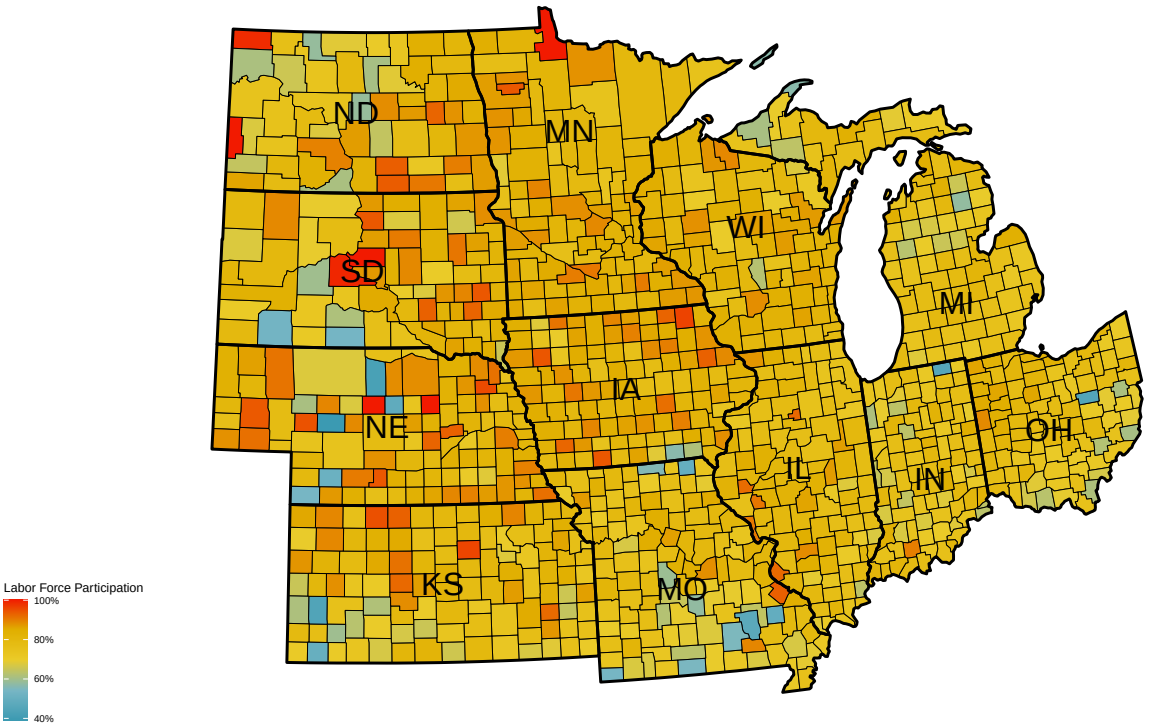
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the North Central region

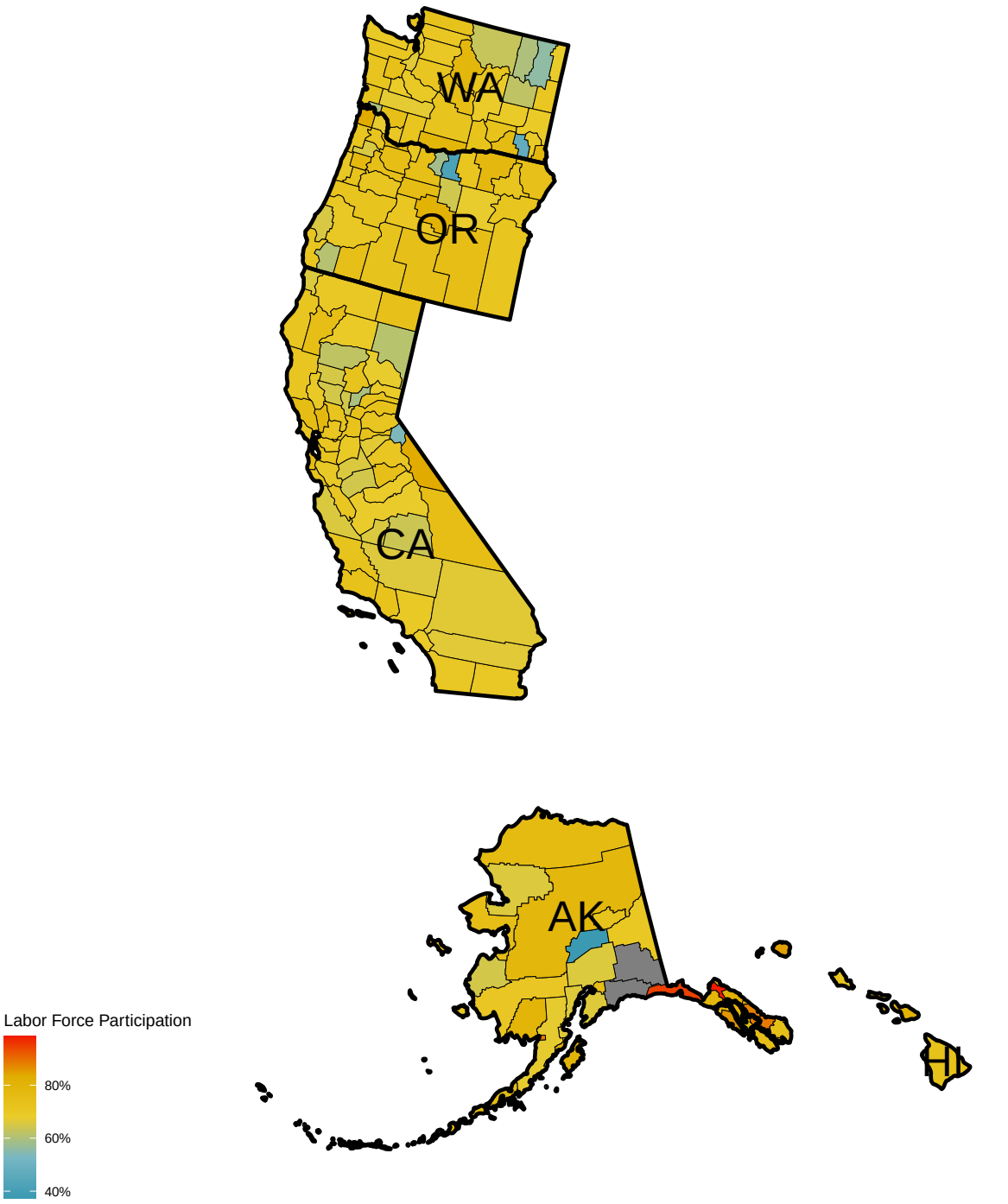
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the Pacific region

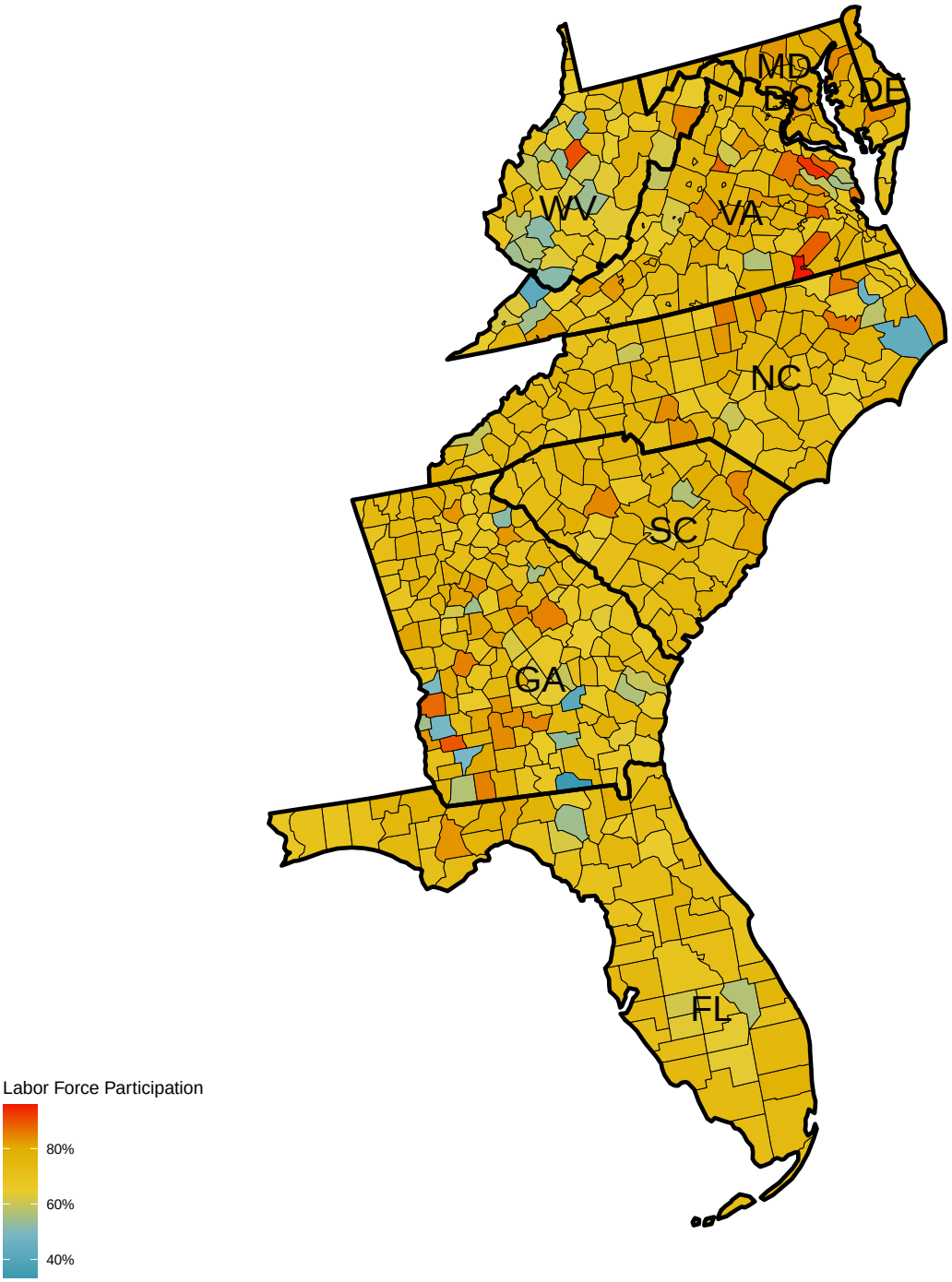
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the South Atlantic region

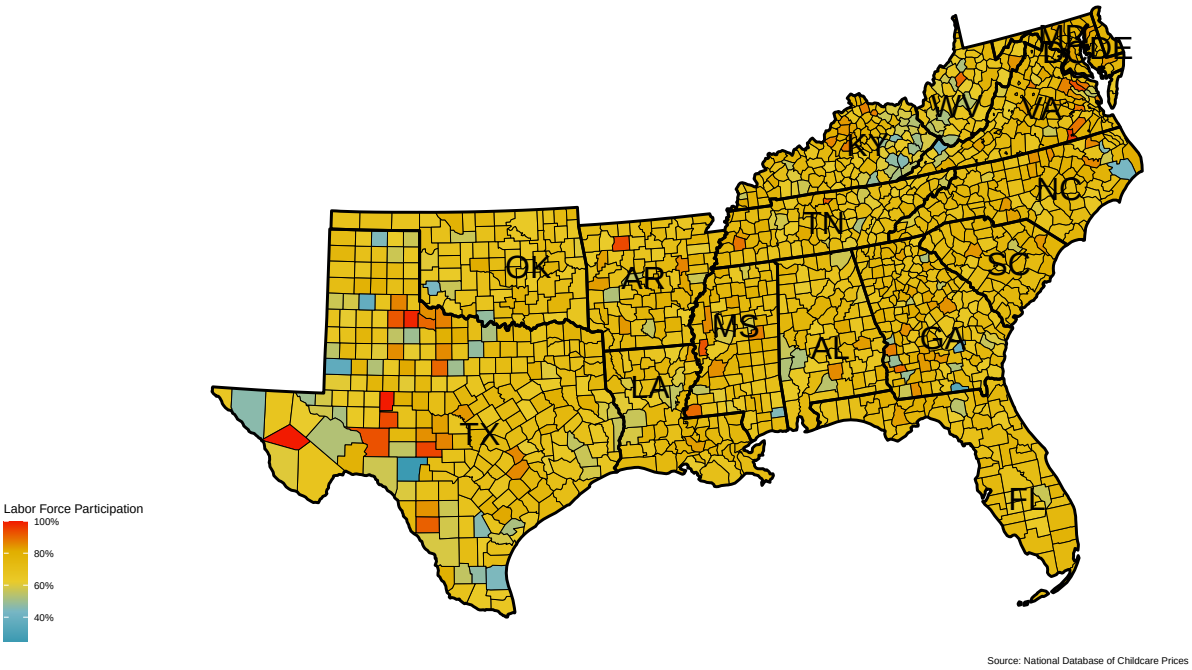
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in South region

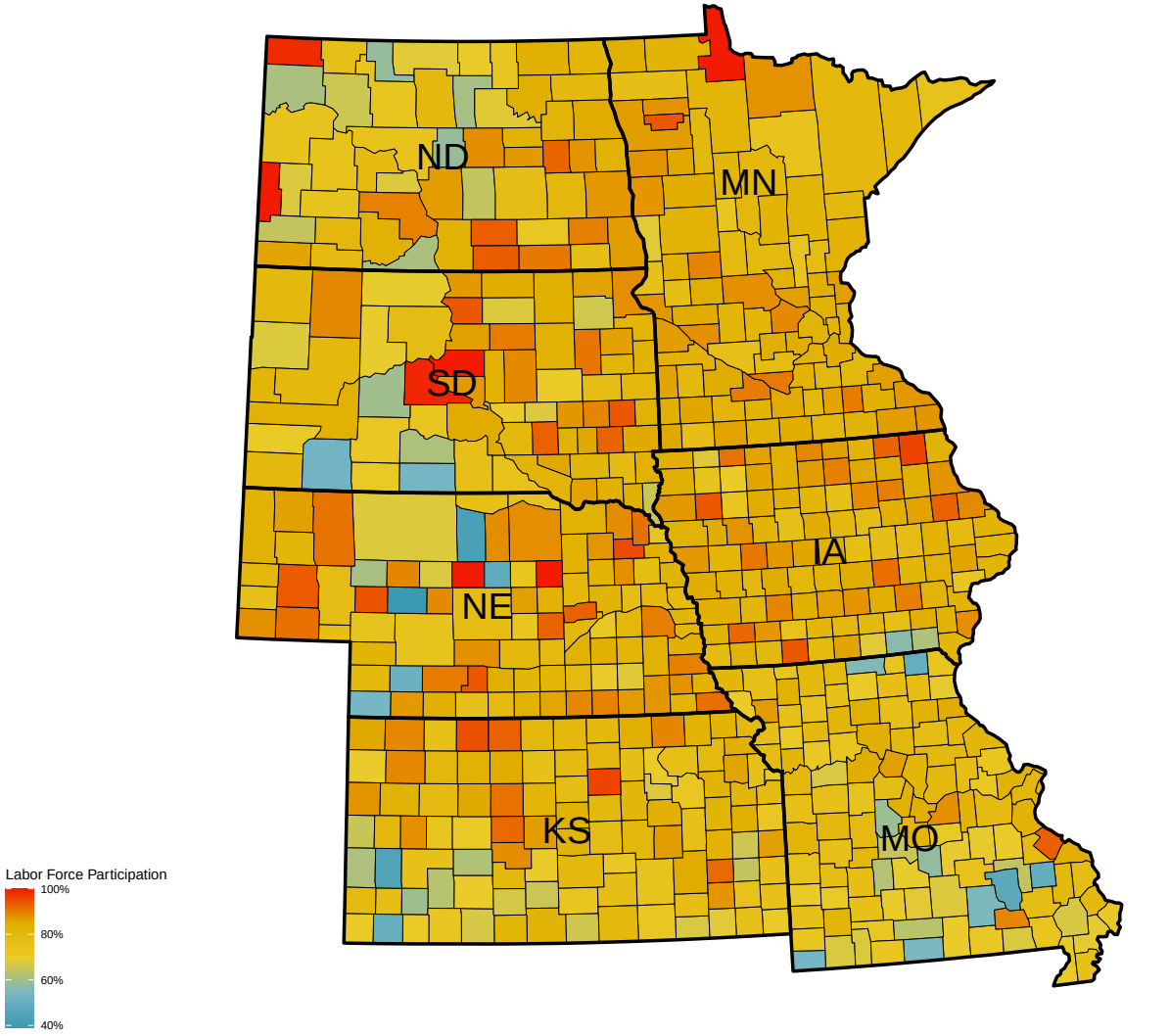
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers West North Central region

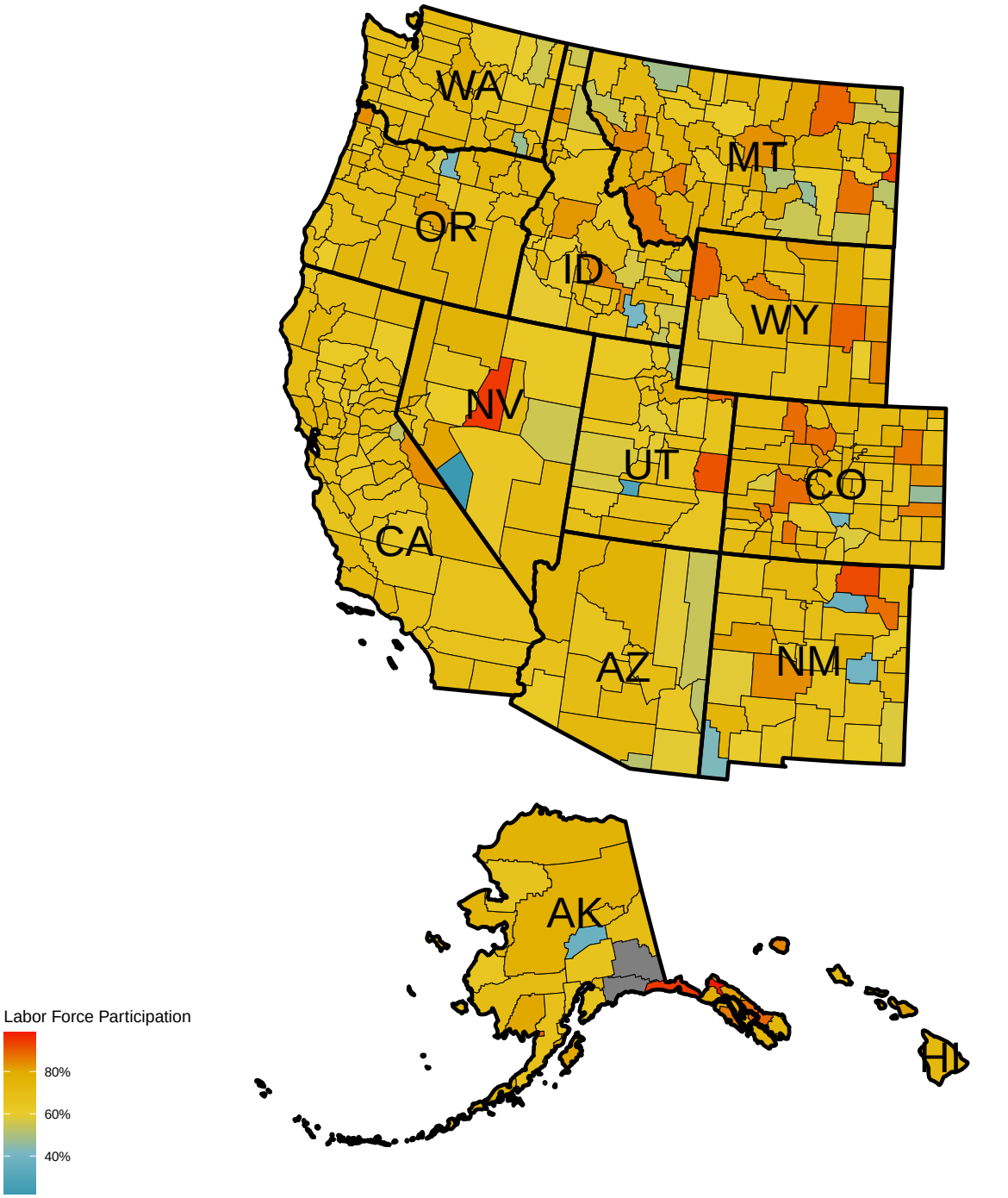
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the West region

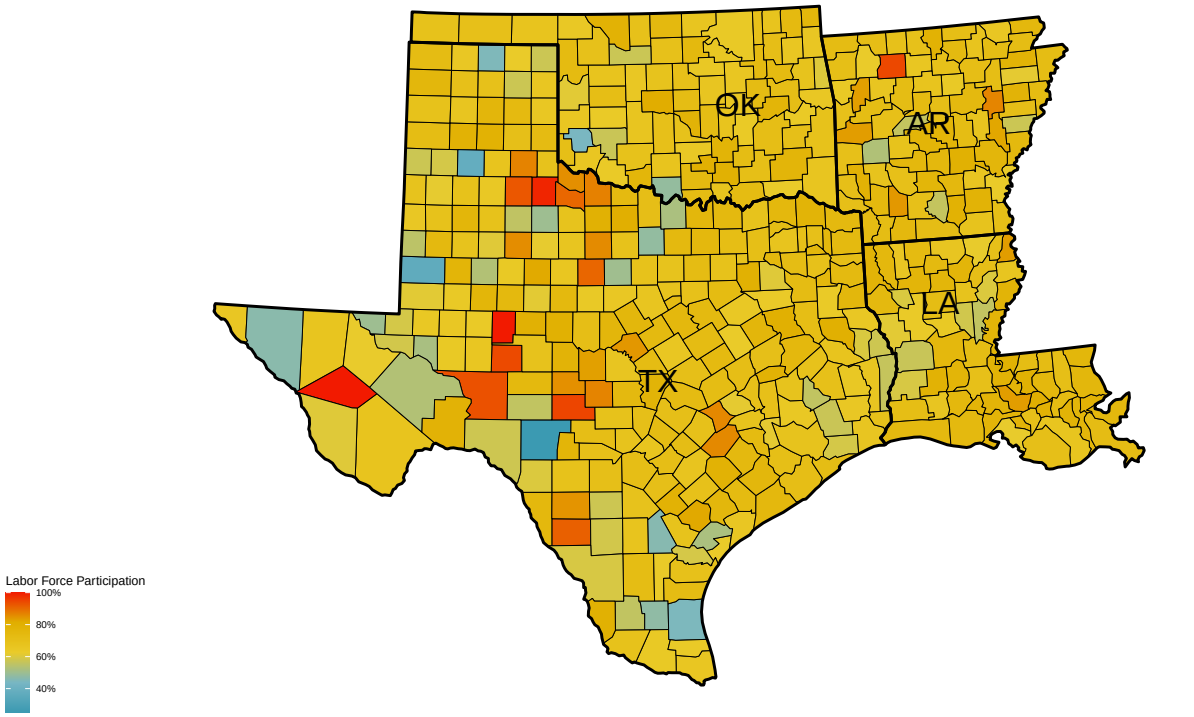
For the year 2018



Source: National Database of Childcare Prices

Labor Force Participation of Mothers in the West South Central region

For the year 2018



Source: National Database of Childcare Prices