Chapter 3

Contents

```
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(terra)
```

```
## terra 1.6.17
```

```
library(dplyr)
```

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

```
## The following objects are masked from 'package:terra':
##
##     intersect, union
```

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

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(spData)
data(us_states)
data(us_states_df)
```

E1. Create a new object called `us_states_name` that contains only the `NAME` column from the `us_states` object using either base R (`[`) or tidyverse (`select()`) syntax. What is the class of the new object and what makes it geographic?

```
us_states_name <- us_states[, "NAME"]
class(us_states_name)
```

```
## [1] "sf"         "data.frame"
```

E2. Select columns from the `us_states` object which contain population data. Obtain the same result using a different command (bonus: try to find three ways of obtaining the same result). Hint: try to use helper functions, such as `contains` or `matches` from **dplyr** (see `?contains`).

```
# E2
us_states_pop1 <- us_states[, c("total_pop_10", "total_pop_15")]

us_states_pop1
```

```
## Simple feature collection with 49 features and 2 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
```

```
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## First 10 features:
##    total_pop_10 total_pop_15                       geometry
## 1       4712651      4830620 MULTIPOLYGON (((-88.20006 3...
## 2       6246816      6641928 MULTIPOLYGON (((-114.7196 3...
## 3       4887061      5278906 MULTIPOLYGON (((-109.0501 4...
## 4       3545837      3593222 MULTIPOLYGON (((-73.48731 4...
## 5      18511620     19645772 MULTIPOLYGON (((-81.81169 2...
## 6       9468815     10006693 MULTIPOLYGON (((-85.60516 3...
## 7       1526797      1616547 MULTIPOLYGON (((-116.916 45...
## 8       6417398      6568645 MULTIPOLYGON (((-87.52404 4...
## 9       2809329      2892987 MULTIPOLYGON (((-102.0517 4...
## 10      4429940      4625253 MULTIPOLYGON (((-92.01783 2...
```

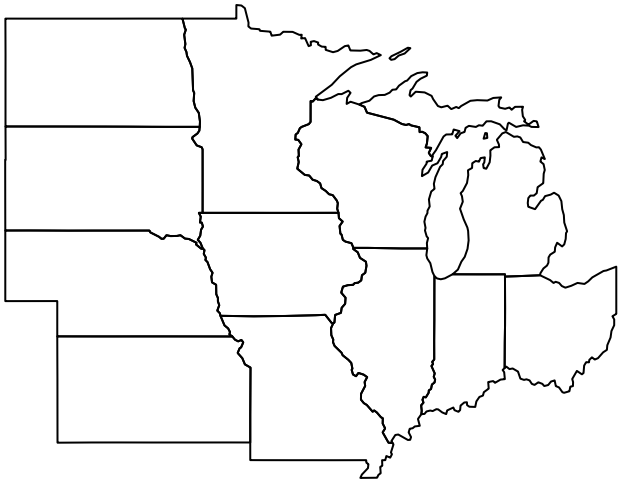
```
us_states |> dplyr::select(total_pop_10, total_pop_15)
```

```
## Simple feature collection with 49 features and 2 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## First 10 features:
##    total_pop_10 total_pop_15                       geometry
## 1       4712651      4830620 MULTIPOLYGON (((-88.20006 3...
## 2       6246816      6641928 MULTIPOLYGON (((-114.7196 3...
## 3       4887061      5278906 MULTIPOLYGON (((-109.0501 4...
## 4       3545837      3593222 MULTIPOLYGON (((-73.48731 4...
## 5      18511620     19645772 MULTIPOLYGON (((-81.81169 2...
## 6       9468815     10006693 MULTIPOLYGON (((-85.60516 3...
## 7       1526797      1616547 MULTIPOLYGON (((-116.916 45...
## 8       6417398      6568645 MULTIPOLYGON (((-87.52404 4...
## 9       2809329      2892987 MULTIPOLYGON (((-102.0517 4...
## 10      4429940      4625253 MULTIPOLYGON (((-92.01783 2...
```

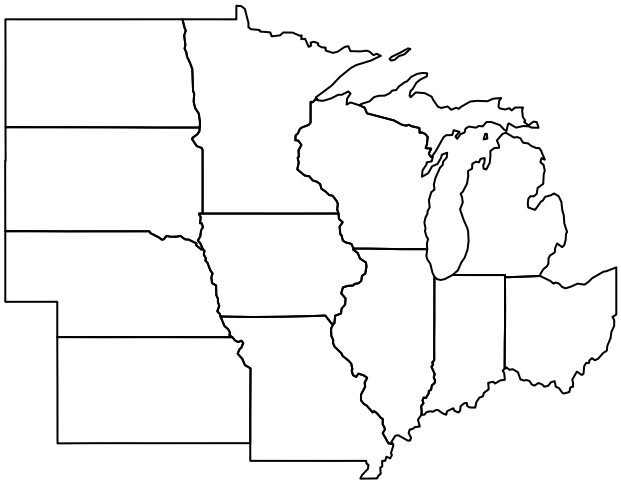
E3. Find all states with the following characteristics (bonus find *and* plot them):

- Belong to the Midwest region.

```
# E3
us_states_midwest <- us_states[us_states$REGION=="Midwest",]
plot(st_geometry(us_states_midwest))
```



```
#E3
plot(st_geometry(us_states |>
  filter(REGION == "Midwest")))
```

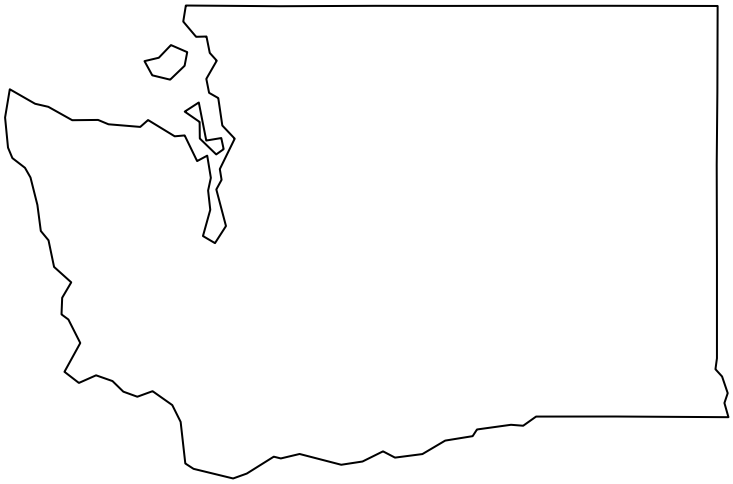


- Belong to the West region, have an area below 250,000 km^2 *and* in 2015 a population greater than 5,000,000 residents (hint: you may need to use the function `units::set_units()` or `as.numeric()`).

```
# E3
us_states_west <- us_states[us_states$REGION=="West",]
us_states_west <- us_states_west[us_states_west$AREA < units::set_units(250000, "km^2"), ]
us_states_west <- us_states_west[us_states_west$total_pop_15 > 5000000, ]
us_states_west
```

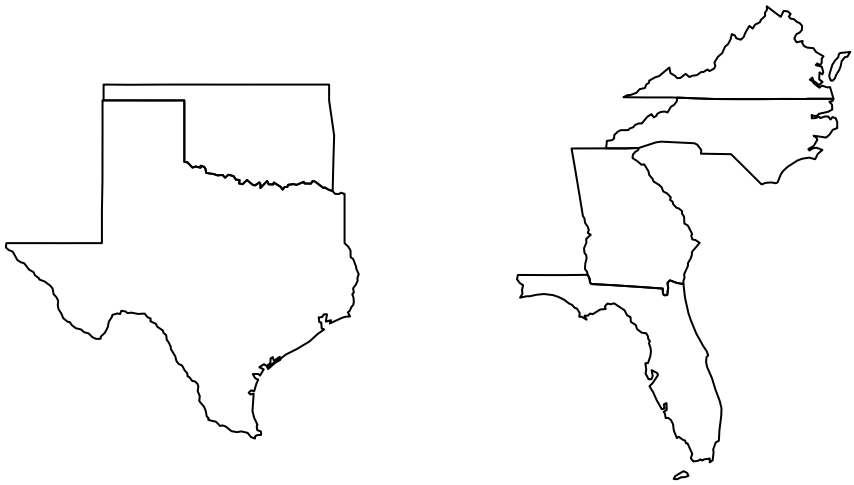
```
## Simple feature collection with 1 feature and 6 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 45.54774 xmax: -116.916 ymax: 49.00236
## Geodetic CRS:  NAD83
##    GEOID       NAME REGION          AREA total_pop_10 total_pop_15
## 47    53 Washington   West 175436 [km^2]      6561297      6985464
##                           geometry
## 47 MULTIPOLYGON (((-122.7699 4...
```

```
plot(st_geometry(us_states_west))
```



- Belong to the South region, had an area larger than 150,000 km^2 or a total population in 2015 larger than 7,000,000 residents.

```
us_states_south <- us_states |> filter(REGION=="South", AREA > units::set_units(150000, "km^2") | total_
plot(st_geometry(us_states_south))
```



E4. What was the total population in 2015 in the `us_states` dataset? What was the minimum and maximum total population in 2015?

```
#E4
us_states |> summarize(total_pop = sum(total_pop_15),
                        min_pop = min(total_pop_15),
                        max_pop = max(total_pop_15))
```

```
## Simple feature collection with 1 feature and 3 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
##   total_pop min_pop  max_pop                       geometry
## 1 314375347  579679 38421464 MULTIPOLYGON (((-122.6096 4...
```

E5. How many states are there in each region?

```
#E5
# n() gives the group size
us_states |>
  group_by(REGION) |>
  summarise(nr_of_states = n())
```

```
## Simple feature collection with 4 features and 2 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## # A tibble: 4 x 3
##   REGION   nr_of_states                                              geometry
##   <fct>           <int>                                  <MULTIPOLYGON [°]>
## 1 Norteast            9 (((-70.8173 42.87229, -70.70382 43.05982, -70.62251 43.~
## 2 Midwest            12 (((-85.48703 45.62121, -85.50127 45.75442, -85.56644 45~
## 3 South              17 (((-81.44412 30.70971, -81.44693 30.81039, -81.40515 30~
## 4 West               11 (((-118.4887 33.41983, -118.4654 33.32606, -118.3102 33~
```

E6. What was the minimum and maximum total population in 2015 in each region? What was the total population in 2015 in each region?

```
#E6
summarize(group_by(us_states, REGION),
          min_pop=min(total_pop_15),
          max_pop=max(total_pop_15),
          tot_pop=sum(total_pop_15))
```

```
## Simple feature collection with 4 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## # A tibble: 4 x 5
##   REGION   min_pop  max_pop   tot_pop                                  geometry
##   <fct>      <dbl>    <dbl>     <dbl>                      <MULTIPOLYGON [°]>
## 1 Norteast  626604 19673174  55989520 (((-70.8173 42.87229, -70.70382 43.05982,~
## 2 Midwest   721640 12873761  67546398 (((-85.48703 45.62121, -85.50127 45.75442~
## 3 South     647484 26538614 118575377 (((-81.44412 30.70971, -81.44693 30.81039~
## 4 West      579679 38421464  72264052 (((-118.4887 33.41983, -118.4654 33.32606~
```

E7. Add variables from `us_states_df` to `us_states`, and create a new object called `us_states_stats`. What function did you use and why? Which variable is the key in both datasets? What is the class of the new object?

```
#E7
us_states_stats <- us_states |>
  left_join(us_states_df, by = c("NAME"="state"))
class(us_states_stats)
```

```
## [1] "sf"         "data.frame"
```

E8. `us_states_df` has two more rows than `us_states`. How can you find them? (hint: try to use the `dplyr::anti_join()` function)

```
#E8
dplyr::anti_join(us_states_df, st_drop_geometry(us_states),
```

```
                by = c("state" = "NAME"))
```

```
## # A tibble: 2 x 5
##   state  median_income_10 median_income_15 poverty_level_10 poverty_level_15
##   <chr>             <dbl>            <dbl>            <dbl>            <dbl>
## 1 Alaska            29509            31455            64245            72957
## 2 Hawaii            29945            31051           124627           153944
```

E9. What was the population density in 2015 in each state? What was the population density in 2010 in each state?

```
#E9
us_pop_density_15 <- us_states$total_pop_15 / us_states$AREA
us_pop_density_15
```

```
## Units: [1/km^2]
##  [1]   36.127786   22.493565   19.582469  276.900364  130.059657   65.520897
##  [7]    7.466293   70.141565   13.579734   37.804767  320.676491   24.794153
## [13]   33.452701    2.664447    9.773012  439.210687  154.660883    3.939560
## [19]   21.272605  109.001118   59.052713    4.220871   38.589593   25.199068
## [25]   29.474967   21.484573   93.768733  178.788485 3633.263718   88.180609
## [31]   21.225799   42.096925   15.541438  220.887166   65.515059   24.147135
## [37]    9.334115   55.114234    6.618638   76.182958  108.134796   15.672536
## [43]  384.119309   59.567250   13.205593   78.332684   39.817723   39.613278
## [49]    2.288421
```

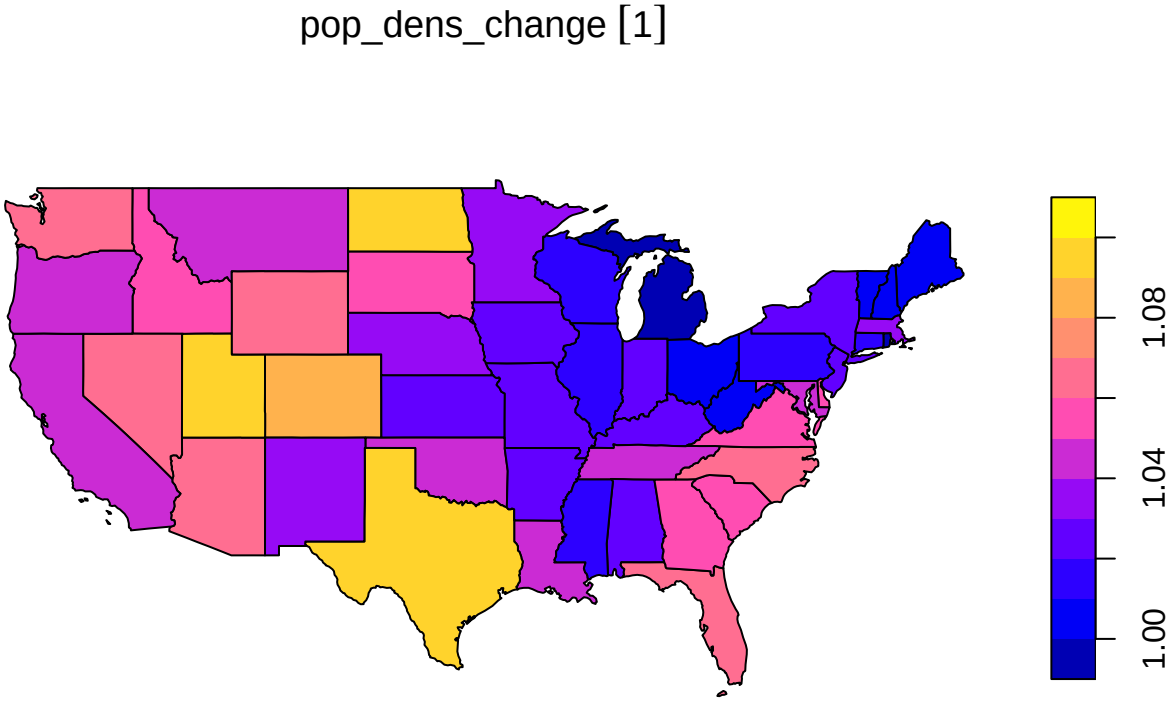
or

```
#E9
us_states2 <- us_states |>
  mutate(pop_dens_15 = total_pop_15/AREA,
         pop_dens_10 = total_pop_10/AREA)

us_states2$pop_dens_15
```

```
## Units: [1/km^2]
##  [1]   36.127786   22.493565   19.582469  276.900364  130.059657   65.520897
##  [7]    7.466293   70.141565   13.579734   37.804767  320.676491   24.794153
## [13]   33.452701    2.664447    9.773012  439.210687  154.660883    3.939560
## [19]   21.272605  109.001118   59.052713    4.220871   38.589593   25.199068
## [25]   29.474967   21.484573   93.768733  178.788485 3633.263718   88.180609
## [31]   21.225799   42.096925   15.541438  220.887166   65.515059   24.147135
## [37]    9.334115   55.114234    6.618638   76.182958  108.134796   15.672536
## [43]  384.119309   59.567250   13.205593   78.332684   39.817723   39.613278
## [49]    2.288421
```

E10. How much has population density changed between 2010 and 2015 in each state? Calculate the change in percentages and map them.

```
#E10
us_states2$pop_dens_change <-
  us_states2$pop_dens_15 / us_states2$pop_dens_10

plot(us_states2[, "pop_dens_change"] )
```



E11. Change the columns' names in `us_states` to lowercase. (Hint: helper functions - `tolower()` and `colnames()` may help.)

```
us_states |>
  setNames(tolower(colnames(us_states)))
```

```
## Simple feature collection with 49 features and 6 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## First 10 features:
##    geoid        name   region             area total_pop_10 total_pop_15
## 1     01     Alabama    South 133709.27 [km^2]      4712651      4830620
## 2     04     Arizona     West 295281.25 [km^2]      6246816      6641928
## 3     08    Colorado     West 269573.06 [km^2]      4887061      5278906
## 4     09 Connecticut Norteast  12976.59 [km^2]      3545837      3593222
## 5     12     Florida    South 151052.01 [km^2]     18511620     19645772
## 6     13     Georgia    South 152725.21 [km^2]      9468815     10006693
## 7     16       Idaho     West 216512.66 [km^2]      1526797      1616547
## 8     18     Indiana  Midwest  93648.40 [km^2]      6417398      6568645
## 9     20      Kansas  Midwest 213037.08 [km^2]      2809329      2892987
## 10    22   Louisiana    South 122345.76 [km^2]      4429940      4625253
##                           geometry
## 1  MULTIPOLYGON (((-88.20006 3...
## 2  MULTIPOLYGON (((-114.7196 3...
## 3  MULTIPOLYGON (((-109.0501 4...
## 4  MULTIPOLYGON (((-73.48731 4...
## 5  MULTIPOLYGON (((-81.81169 2...
## 6  MULTIPOLYGON (((-85.60516 3...
## 7  MULTIPOLYGON (((-116.916 45...
## 8  MULTIPOLYGON (((-87.52404 4...
```

```
## 9  MULTIPOLYGON (((-102.0517 4...
## 10 MULTIPOLYGON (((-92.01783 2...
```

or

```
us_states %>%
  setNames(tolower(colnames(.)))
```

```
## Simple feature collection with 49 features and 6 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## First 10 features:
##    geoid        name   region             area total_pop_10 total_pop_15
## 1     01     Alabama    South 133709.27 [km^2]      4712651      4830620
## 2     04     Arizona     West 295281.25 [km^2]      6246816      6641928
## 3     08    Colorado     West 269573.06 [km^2]      4887061      5278906
## 4     09 Connecticut Norteast  12976.59 [km^2]      3545837      3593222
## 5     12     Florida    South 151052.01 [km^2]     18511620     19645772
## 6     13     Georgia    South 152725.21 [km^2]      9468815     10006693
## 7     16       Idaho     West 216512.66 [km^2]      1526797      1616547
## 8     18     Indiana  Midwest  93648.40 [km^2]      6417398      6568645
## 9     20      Kansas  Midwest 213037.08 [km^2]      2809329      2892987
## 10    22   Louisiana    South 122345.76 [km^2]      4429940      4625253
##                          geometry
## 1  MULTIPOLYGON (((-88.20006 3...
## 2  MULTIPOLYGON (((-114.7196 3...
## 3  MULTIPOLYGON (((-109.0501 4...
## 4  MULTIPOLYGON (((-73.48731 4...
## 5  MULTIPOLYGON (((-81.81169 2...
## 6  MULTIPOLYGON (((-85.60516 3...
## 7  MULTIPOLYGON (((-116.916 45...
## 8  MULTIPOLYGON (((-87.52404 4...
## 9  MULTIPOLYGON (((-102.0517 4...
## 10 MULTIPOLYGON (((-92.01783 2...
```

```
?setNames
```

E12. Using `us_states` and `us_states_df` create a new object called `us_states_sel`. The new object should have only two variables - `median_income_15` and `geometry`. Change the name of the `median_income_15` column to `Income`.

```
us_states_df
```

```
## # A tibble: 51 x 5
##    state                median_income_10 median_income_15 poverty_leve~1 pover~2
##    <chr>                           <dbl>            <dbl>          <dbl>   <dbl>
##  1 Alabama                         21746            22890         786544  887260
##  2 Alaska                          29509            31455          64245   72957
##  3 Arizona                         26412            26156         933113 1180690
##  4 Arkansas                        20881            22205         502684  553644
##  5 California                      27207            27035        4919945 6135142
##  6 Colorado                        29365            30752         584184  653969
##  7 Connecticut                     32258            33226         314306  366351
##  8 Delaware                        29205            30329          93857  108315
##  9 District of Columbia            35264            40884         101767  110365
```

```
## 10 Florida                          24812            24654        2502365 3180109
## # ... with 41 more rows, and abbreviated variable names 1: poverty_level_10,
## #   2: poverty_level_15
```

```
us_states_sel <- left_join(us_states, us_states_df, by = c("NAME" = "state"))
us_states_sel <- dplyr::select(us_states_sel, Income=median_income_15)
us_states_sel
```

```
## Simple feature collection with 49 features and 1 field
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.7042 ymin: 24.55868 xmax: -66.9824 ymax: 49.38436
## Geodetic CRS:  NAD83
## First 10 features:
##    Income                       geometry
## 1   22890 MULTIPOLYGON (((-88.20006 3...
## 2   26156 MULTIPOLYGON (((-114.7196 3...
## 3   30752 MULTIPOLYGON (((-109.0501 4...
## 4   33226 MULTIPOLYGON (((-73.48731 4...
## 5   24654 MULTIPOLYGON (((-81.81169 2...
## 6   25588 MULTIPOLYGON (((-85.60516 3...
## 7   23558 MULTIPOLYGON (((-116.916 45...
## 8   25834 MULTIPOLYGON (((-87.52404 4...
## 9   27315 MULTIPOLYGON (((-102.0517 4...
## 10  24014 MULTIPOLYGON (((-92.01783 2...
```

or

```
us_states_sel <- us_states |>
  left_join(us_states_df, by = c("NAME" = "state")) |>
  dplyr::select(Income = median_income_15)
```

E13. Calculate the change in the number of residents living below the poverty level between 2010 and 2015 for each state. (Hint: See ?us_states_df for documentation on the poverty level columns.) Bonus: Calculate the change in the *percentage* of residents living below the poverty level in each state.