In comparison with other statistical software (e.g., SAS, STATA, and SPSS), R is the best for data visualization. Therefore, in all [posts I have written](https://datascienceplus.com/author/anisa-dhana/) for DataScience+ I take advantage of R and make plots using ggplot2 to visualize all the findings. For example, [previously](https://datascienceplus.com/obesity-trends-in-the-united-states-an-analysis-and-visualization-with-tidyverse-in-r/) I plotted the percentiles of body mass index in the NHANES 2005-2014 and got exactly same results as the paper published in [JAMA](https://jamanetwork.com/journals/jama/fullarticle/2526639).

In this post, I will make a map of the prevalence of Alzheimer disease mortality by the state in the USA. The [Centers for Disease Control and Prevention](https://www.cdc.gov/nchs/pressroom/sosmap/alzheimers_mortality/alzheimers_disease.htm) is providing the data for download, and they have created a beautiful map. I will try to reproduce the same results using several packages in R.

**Libraries and Datasets**

Load the library

library(tidyverse)

library(scales)

library(maps)

library(mapproj)

Download the .CSV file from the Centers for Disease Control and Prevention website (link is above)

dt\_ad <- read.csv("~/Downloads/ALZHEIMERS2016.csv")

head(dt\_ad)

*STATE RATE DEATHS URL*

*1 AL 45.0 2,507 /nchs/pressroom/states/alabama/alabama.htm*

*2 AK 25.8 111 /nchs/pressroom/states/alaska/alaska.htm*

*3 AZ 35.8 3,082 /nchs/pressroom/states/arizona/arizona.htm*

*4 AR 41.3 1,475 /nchs/pressroom/states/arkansas/arkansas.htm*

*5 CA 36.1 15,570 /nchs/pressroom/states/california/california.htm*

*6 CO 34.7 1,835 /nchs/pressroom/states/colorado/colorado.htm*

Load the map data of the U.S. states

dt\_states = map\_data("state")

head(dt\_states)

*long lat group order region subregion*

*1 -87.46201 30.38968 1 1 alabama*

*2 -87.48493 30.37249 1 2 alabama*

*3 -87.52503 30.37249 1 3 alabama*

*4 -87.53076 30.33239 1 4 alabama*

*5 -87.57087 30.32665 1 5 alabama*

*6 -87.58806 30.32665 1 6 alabama*

Now, I have two datasets, one has the rate of mortality from Alzheimer disease and the other have variables with the information to create maps. I need to merge both datasets together but I dont have a similar variable for merge. Therefore, I will create a new region variable form the URL variable in the first dataset and will use to merge with the second dataset. For this purpose, I will use the function separate and gsub. In the end I will merge with states dataset by region.

#get the state name from URL

dt\_ad2 = dt\_ad %>%

separate(URL, c("a","b","c","d", "region"), sep="/") %>%

select(RATE, region)

# removing white space for mergin purposes

dt\_states2 = dt\_states %>%

mutate(region = gsub(" ","", region))

# merge

dt\_final = left\_join(dt\_ad2, dt\_states2)

**Visualization**

The dt\_final dataset have all the variables I need to make the map.

ggplot(dt\_final, aes(x = long, y = lat, group = group, fill = RATE)) +

geom\_polygon(color = "white") +

scale\_fill\_gradient(

name = "Death Rate",

low = "#fbece3",

high = "#6f1873",

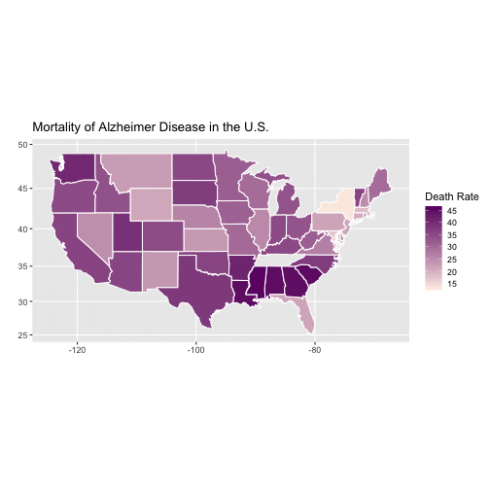
guide = "colorbar",

na.value="#eeeeee",

breaks = pretty\_breaks(n = 5)) +

labs(title="Mortality of Alzheimer Disease in the U.S.", x="", y="") +

coord\_map()

[](https://datascienceplus.com/wp-content/uploads/2018/08/unnamed-chunk-6-1-1.png)