Execution Data in US
R shiny Dashboard

Heejae Roh
Northeastern University
ALY 6070: Communication and Visualization
for Data Analytics
Venkata Duvvuri
March 22, 2023

1. INTRODUCTION

In this module, I will make various visualizations using R shiny and check the the gestalt and design principles of visualization. I used the most recent Execution data from Kaggle. And the original data is provided by deathpenaltyinfo.org. The data covers until the end of 2022, I could say it is the latest statistic. In the process of using R shiny, I referred to many R shiny results until I made this data a conclusion. R shiny was a data format that was more difficult to access than tableau, and data preparation was also necessary. However, I think the biggest advantage is that it can be used as an app.

2. DATA PREPARATION

1. Read Execution data as tibble. Uploaded Map data in rds format.
2. Remove NA values and change column names to match with states.rds data.
3. Obtain population data from Census, read tibble, and left_join with execution data.
4. Calculate relative number of executions based on population data.
5. Organize execution data into state and population using the group_by function and count the number of executions.

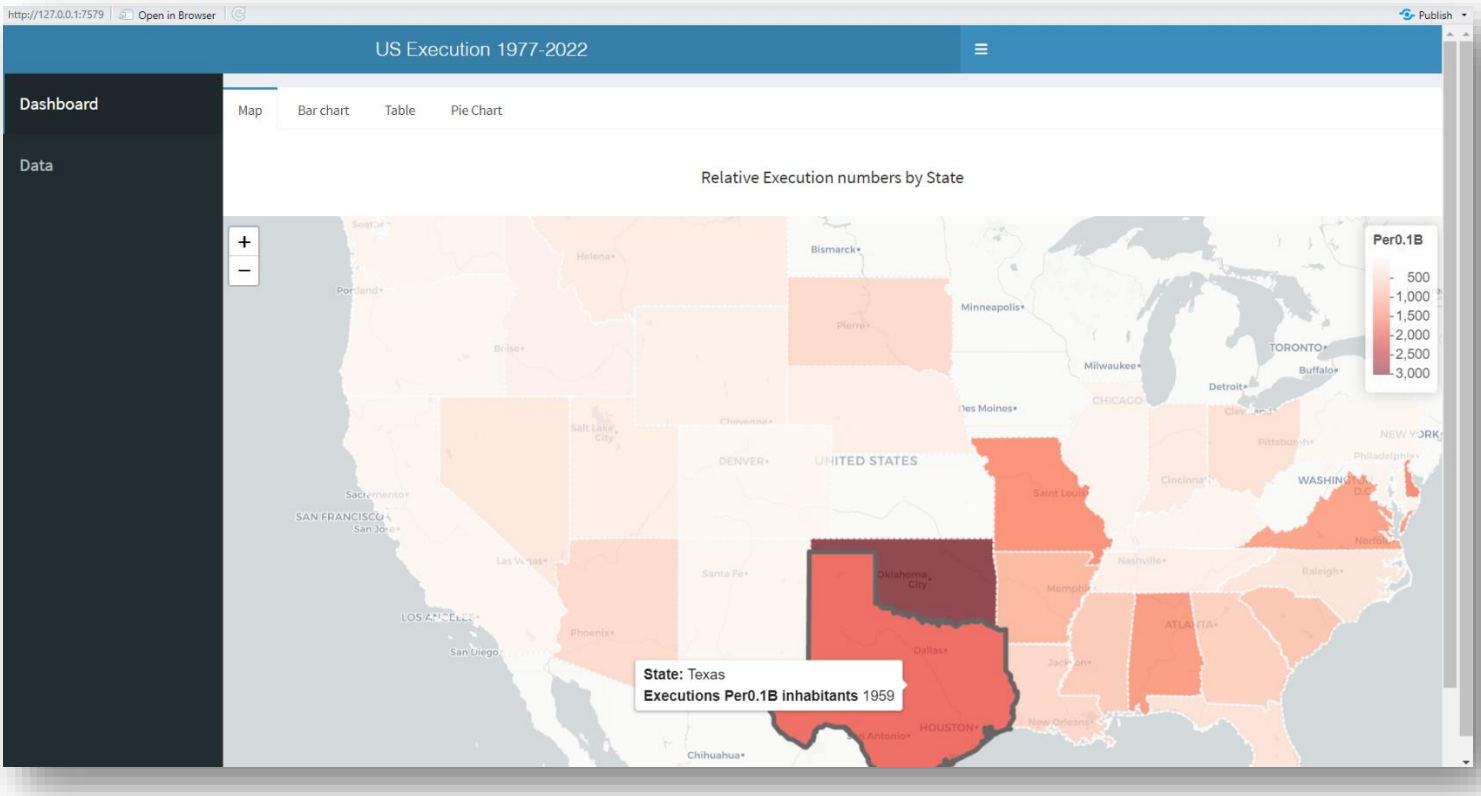

Figure 1 R shiny Dashboard & map tab

3. ANALYSIS

Q. Why did you choose the types of visualizations that you did?

I wanted to utilize US map data which is the 'states.rds' file, and it is compatible with the State name, so it is suitable for joining with data which have 'state'. Therefore, I wanted to find meaningful data among the data with 'state'. So, I searched for data whose number could vary for each state and found ‘Execution’ data. Execution is a very controversial topic. Some countries retain the death penalty, but do not carry out exact executions. With data on controversial subjects, I wanted to know which state had the most executions since 1977 in the United States.

In the picture above, the Top 18 execution states are supposed to appear. If you change the N using the Sidebar, the state of Top N will appear. It could be 3, 5, 10, etc.

Figure 2 Execution number Bar Chart with Choose Top N states

Q. How are the visualizations effective and address the gestalt and design principles discussed in the course?

Proximity and Similarity: In ‘Figure 2, Execution number Bar Chart with Choose Top N states’, I used the color and size of the graph to distinguish each data. I created a gradient using two color ranges: yellow to red and green to red. In this case, if the colors are similar

and the graphs are similar in size, it can be said that they show proximity and similarity. In the picture above, the number of executions by state is predominantly in Texas, and the relative number of executions by state is predominantly in Oklahoma, so only the two appear in different color which is red. In other words, states, other than these two, show proximity and similarity.

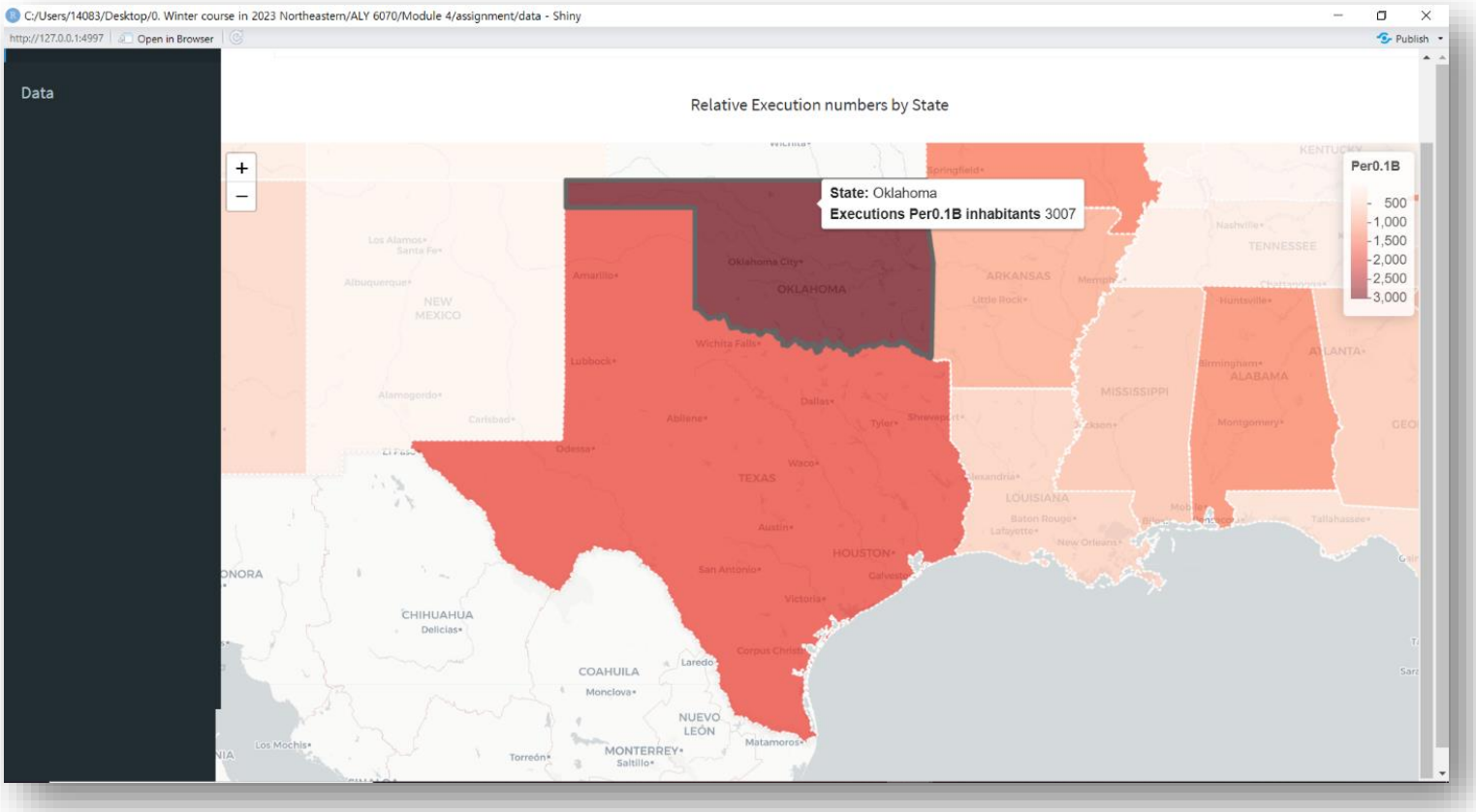

Figure 3 Zoom in Map tab (Oklahoma)

Enclosure: I colored the map in Figure 1 & 3 and then marked the borders. This is what makes an enclosure. And I made the borders clearer when you hover over them. This is an additional element that can further clarify the Enclosure. Users can browse more clearly by state in this Dashboard. In particular, if you hover over it, you can see the relative number of executions according to State name and population.

Closure: And in Figure 3, the groups are different depending on the color. For example, Oklahoma and Texas appear in approximately the same color on the map. Some people may recognize the Enclosure after recognizing the two together. Or, conversely, there are areas that appear white. This means that there was no execution at all. If these states are continuously distributed, they will form a group together and create a closure.

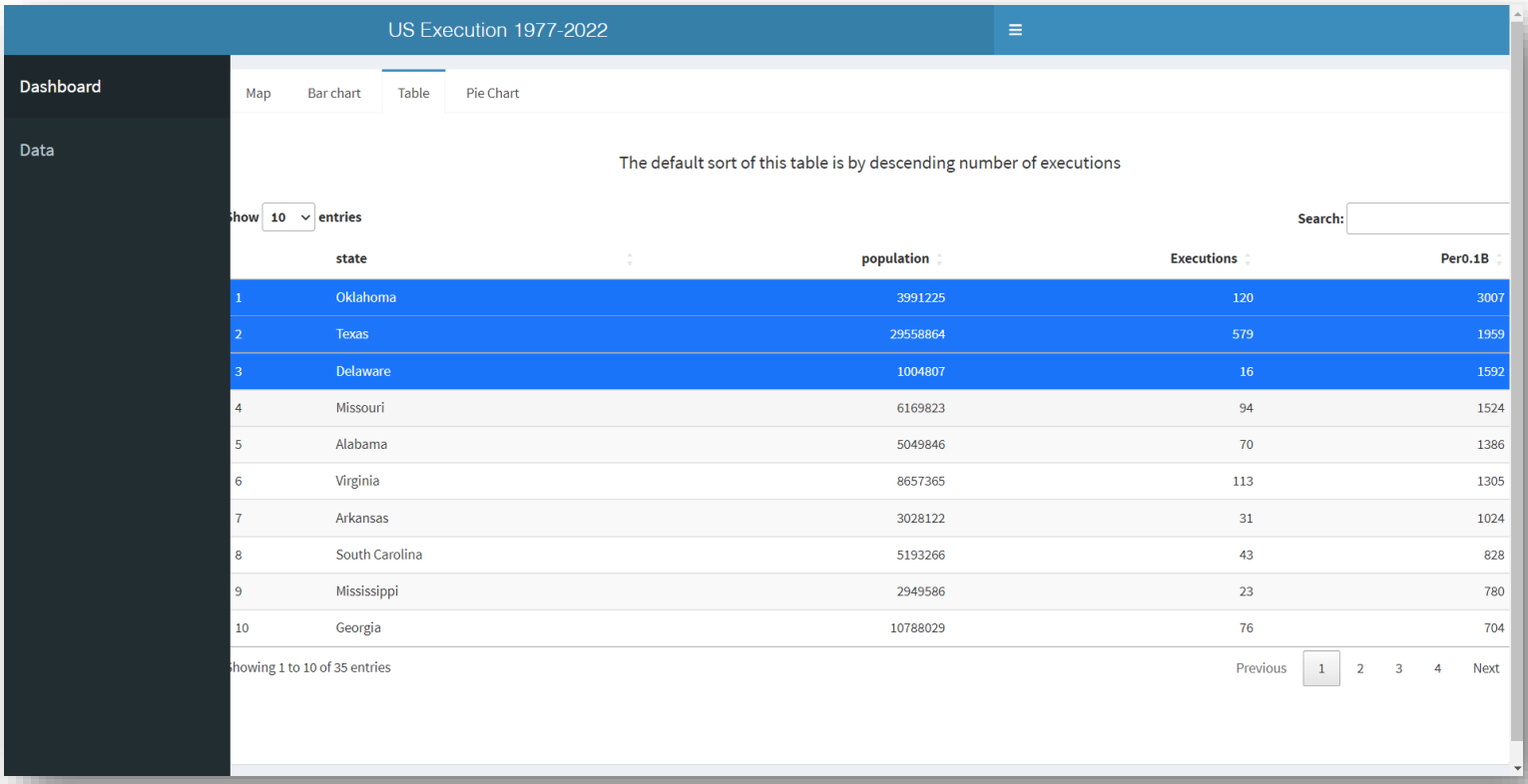

Figure 4 Table of Execution & Population by State

Continuity: In Figure 4 'Table of Execution & Population by State' allows you to adjust Entries. If 10 states come together like the picture above, people can think that these states are states with continuity. You can also feel similarly in the Figure 2 Bar chart. If states have similar numbers, you can think that the numbers are more continuous, and the properties of the state are also continuous.

Connection: Although this dashboard consists of Map, Bar chart, Table, and Pie Chart, but the data has connection according to the name of the state. When the image and data of Texas seen on the map are transferred to the bar chart, the data connected to No. 1 in Total number and No. 2 in Relative number. Also, if you move to the table, the population and other information are linked.

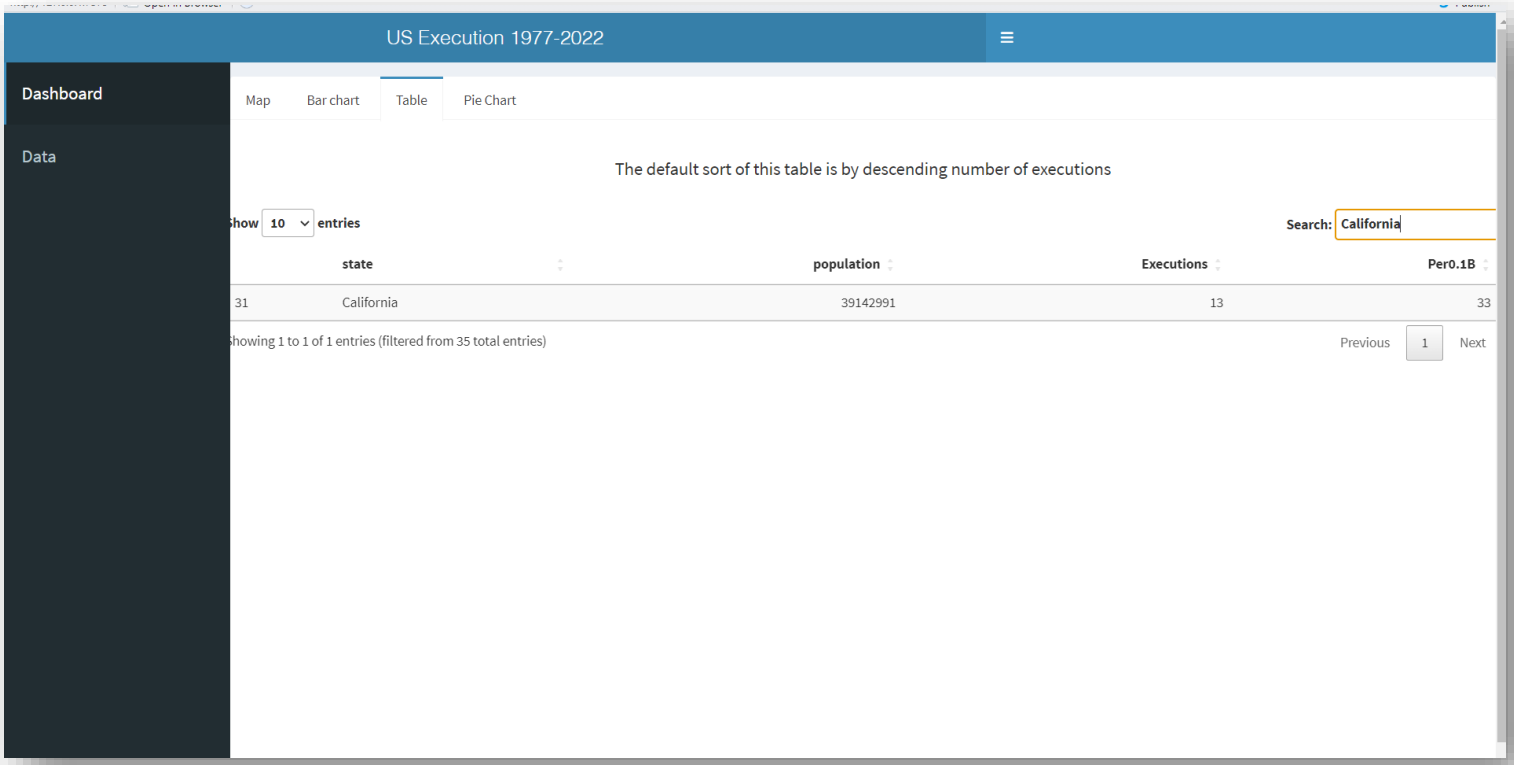

Figure 5 Table. Searching 'California' and Result

Q. How do the visualizations answer the research/business question?

First of all, you can directly find statistics by searching for the name of the state you are interested in in the table as above. Alternatively, you can sort in ascending or descending order according to each criterion by clicking population, execution, and per0.1B (which means the number of executions in 0.1 billion people). in table. Through this, you can see the information by ascending or descending order of each criterion.

Through the Figure 1 Map, we can see which state is most actively executing. Since this includes data up to 2022, you can find a place where relatively high execution takes place by combining all the data until recently. And you can see the entire US data at a glance on the map.

And through the bar chart, you can also check the number of Top N states with the highest total number and the states with the highest number of executions relative to the population.

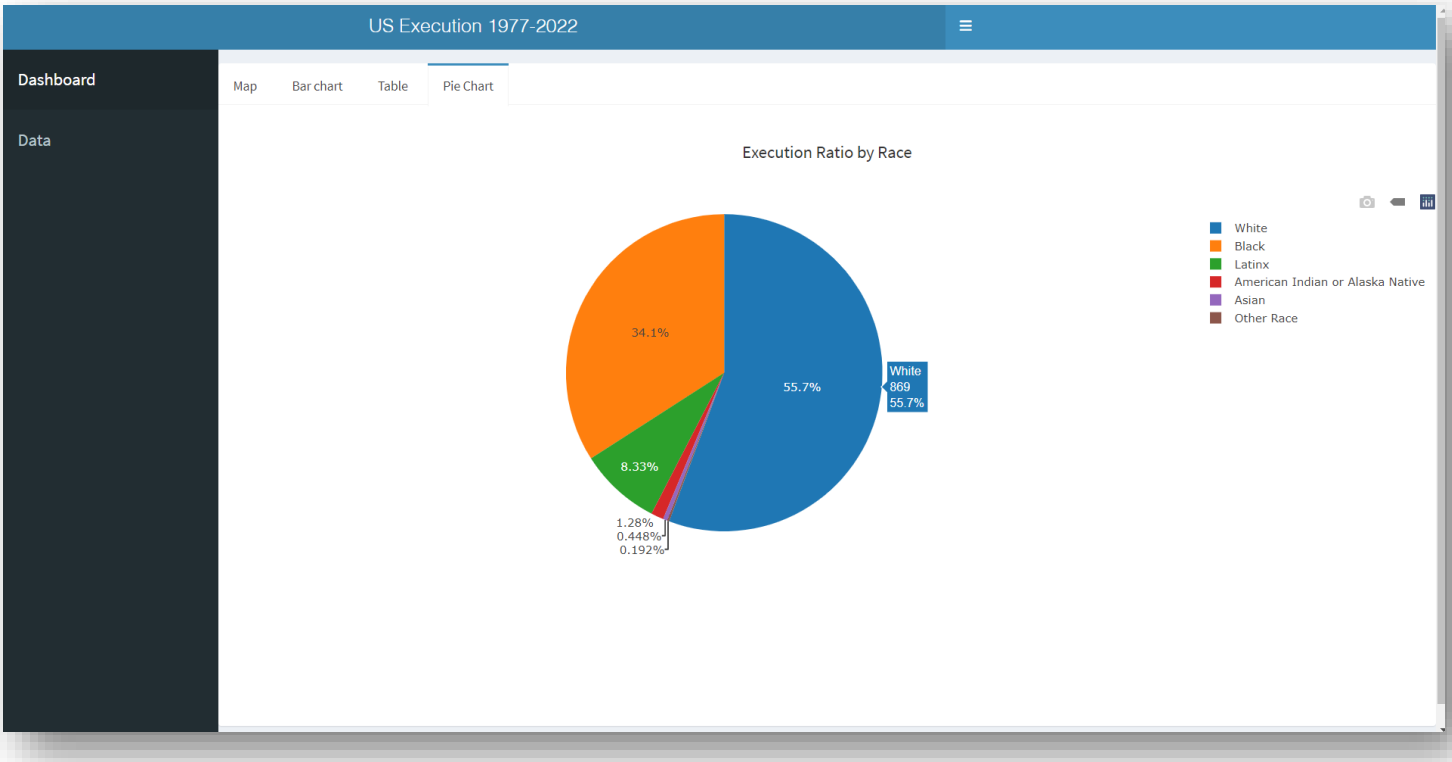

Figure 6 Execution Ratio by Race

Q. What story do the visualizations tell?

Figure 6 shows how the ratio of executions varies depending on the race. Accordingly, we can compare the number of Executions by race. According to the US census in 2021, White is 59.3%, Latino is 18.9%, Black is 13.6%, American Indian and Alaska Native is 1.3%, and Asian appears at 6.1%. Compared to this, Black's execution rate is relatively high, and other races are relatively low. More detailed analysis will be needed, but we can start solving the social problem by finding the cause for this.

Summarizing what this Dashboard says from the beginning, first through the Figure 1 map, you can see how the relative number of executions varies from state to state. Based on this, you can think of execution regulations that vary by state. If there are states where executions are more frequent, you might wonder if there are other factors in that state that facilitate executions. Similarly, in Figure 2, you can think of the absolute number of executions and relative number of executions through the Top N state. You can take a closer look at the data you could see at a glance on the map above.

In Figure 3, you can see the figures briefly through the table. Even if it is not a standout state, you can select and look at the data on table. And you can search for the state name directly through search, or you can look at the data of several states at once by using the alphabet in which the name is entered. As mentioned earlier, this allows you to see how the data in each state changes.

4. CONCLUSION

I learn a lot of new things to use R shiny. R has many useful aspects as a statistics analysis tool. Various tests can be performed, and the possibility of development is high through the new 'package'. However, in terms of visualization or interaction, there are some aspects that are lacking, but R shiny seems to make up for it. I felt I had to put more effort into constructing the code myself through R shiny. I thought that the data containing the state map could now be visualized a little more in R & R shiny.

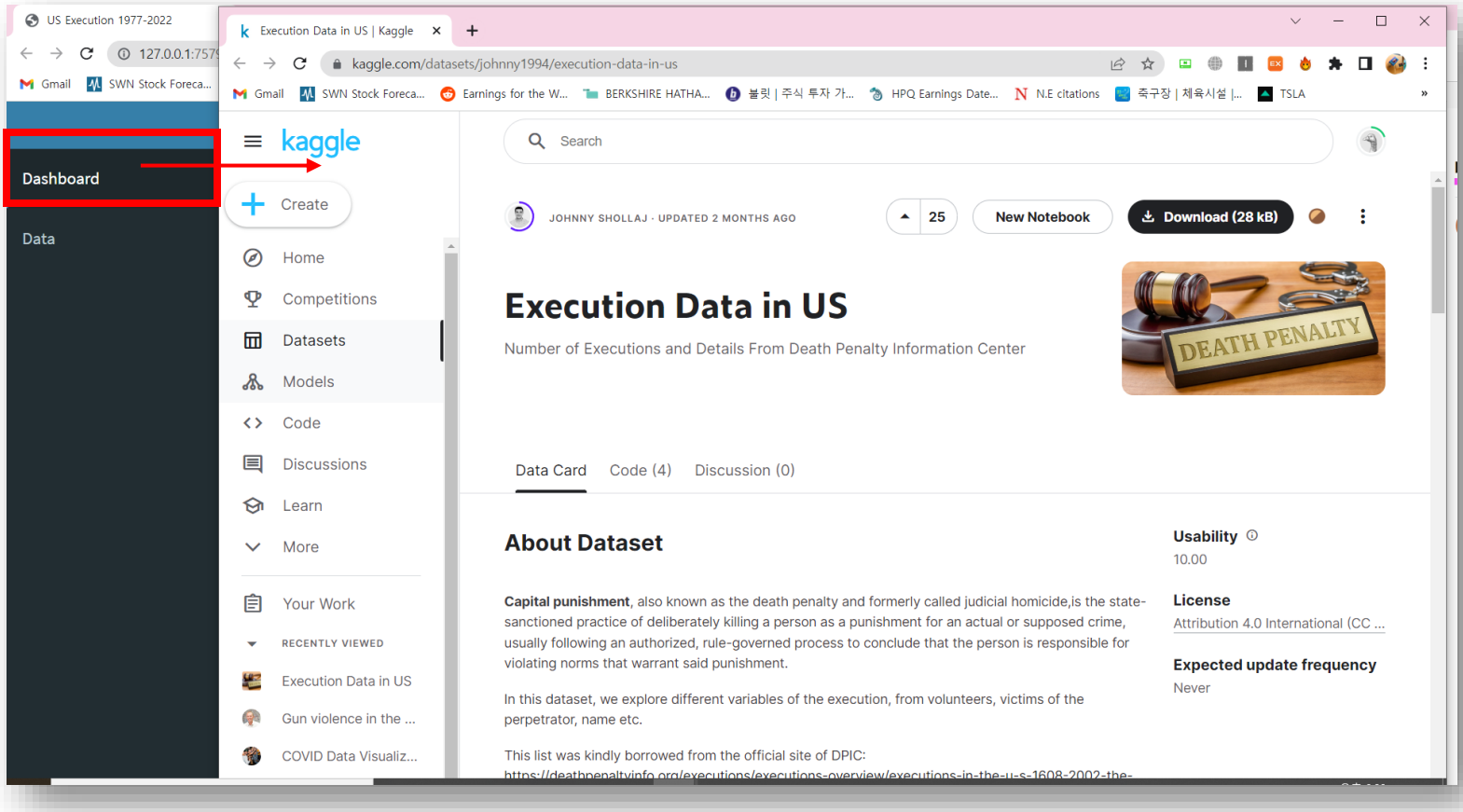

Figure 7 When Clicking 'Data' Dashboard, the Kaggle link pops up

REFERENCE

JOHNNY, SHOLLAJ. (2023). Execution data in US. Kaggle. Retrieved from https://www.kaggle.com/datasets/johnny1994/execution-data-in-us

DPIC. (2023). Executions by state and region since 1976. Death penalty Information center (DPIC). Retrieved from https://deathpenaltyinfo.org/executions/executions-overview/number-of-executions-by-state-and-region-since-1976

Betsy, Reed. (February 9, 2023). Texas executes john balentine, sixth inmate to be put to death in US this year. The Guardian. Retrieved from https://www.theguardian.com/us-news/2023/feb/09/texas-executes-john-balentine-sixth-us-inmate-this-year

plotly. (n.d.). Pie Charts in R. Retrieved from https://plotly.com/r/pie-charts/

Shane. (December 17, 2009). Counting the number of elements with the values of x in a vector. Retrieved from https://stackoverflow.com/questions/1923273/counting-the-number-of-elements-with-the-values-of-x-in-a-vector

yippeee25. (April 28, 2021). How to change row names of dataframe in R ?. Retrieved from https://www.geeksforgeeks.org/how-to-change-row-names-of-dataframe-in-r/

maxime, bordes. (2017). Incorporate a plotly graph into a shiny app. plotly. Retrieved from https://community.plotly.com/t/incorporate-a-plotly-graph-into-a-shiny-app/5329

Erik, Bruin. (2020). Gun violence in the US: EDA and RShiny app. Kaggle. Retrieved from https://www.kaggle.com/code/erikbruin/gun-violence-in-the-us-eda-and-rshiny-app/report

Shiny. (2023). R shiny gallery. Retrieved from https://shiny.rstudio.com/gallery/

APPENDIX

```
library(shiny)
library(dplyr)
library(ggplot2)
library(plotly)
library(leaflet)
library(shinydashboard)
library(tigris)
library(readr)
library(knitr)
library(readr)
library(tibble)
library(stringr)
library(gridExtra)
library(scales)
library(lubridate)
library(ggrepel)
library(rgdal)
library(psych)

setwd("C:\\Users\\14083\\Desktop\\0. Winter course in 2023 Northeastern\\ALY 6070\\Module 4\\assignment\\data")
execut <- as_tibble(data.table::fread("U.S. Executions_cleaned.csv"),header=TRUE,stringsAsFactors = FALSE, na.strings=c("NA", ""))
states <- readRDS("states.rds")

str(execut)
execut <- na.omit(execut)
execut %>% rename(state=State)

#Data preparation
statesPop <- read_csv(str_c("NST-EST2022-POPCHG2020_2022.csv"))
str(statesPop)
statesPop

statesPop
statesPop <- statesPop %>% select(NAME, POPESTIMATE2021)
statesPop <- statesPop %>% filter(!NAME %in% c("United States", "Puerto Rico Commonwealth"))
statesPop <- statesPop %>% rename(state= NAME)
statesPop$state <- as.factor(statesPop$state)
statesPop
colnames(statesPop)[2] <- "population"
```

```
execut_new <- execut %>% rename(state = State)
ExecuteByState <- execut_new %>% group_by(state) %>% summarize(stateExecution=n())
ExecuteByState <-left_join(ExecuteByState, statesPop, by="state")
ExecuteByState$Per100000000 <-
round((ExecuteByState$stateExecution/ExecuteByState$population)*100000000)
ExecuteByState

execute <-  execut_new
execute <- left_join(execute, statesPop, by="state")
str(execute)

# Data Preperation for pie chart
ExecuteRatio <- as.data.frame(table(execute$Race))
ExecuteRatio <-  ExecuteRatio %>% rename(Race = Var1)
ExecuteRatio <-  ExecuteRatio %>% rename(Num = Freq)
ExecuteRatio

rownames(ExecuteRatio) <- ExecuteRatio$Race

ExecuteRatio

header <- dashboardHeader(title = "US Execution 1977-2022", titleWidth=1000)

sidebar <- dashboardSidebar(
  sidebarMenu(
    menuItem(h4("Dashboard"), tabName = "dashboard"),
    menuItem(h4("Data "), href = "https://www.kaggle.com/datasets/johnny1994/execution-data-
in-us"), newtab = FALSE))

body <- dashboardBody(
  tabItems(
    tabItem(tabName= "dashboard",
          fluidRow(
            tabBox(id = "heejae", width=20, height=700,
                   tabPanel("Map",
                     br(),
                     fluidRow(h4("Relative Execution numbers by State", align='center')),
                     br(),
                     fluidRow(leafletOutput("usmap", height=700))),
                tabPanel("Bar chart",
                     br(),
                     h4("Most Executions by State", align="center"),
                     sliderInput(inputId = "num",
                                 label = "Choose Top N states with most executions",
                                 value = 10,
```

```
                    min = 1,
                    max = (n_distinct(execute$state)-1)),
                fluidRow(plotlyOutput("plot1")),
                fluidRow(plotlyOutput("plot2"))),
            tabPanel("Table",
                br(),
                fluidRow(h4("The default sort of this table is by descending number of
executions", align='center')),
                br(),
                fluidRow(DT::dataTableOutput(outputId="table"))),
            tabPanel("Pie Chart",
                br(),
                fluidRow(h4("Execution Ratio by Race", align='center')),
                br(),
                fluidRow(plotlyOutput("plot3"))))

        )    )  ))

ui <- dashboardPage(header, sidebar, body)

server <- function(input, output) {

  execution_select <- reactive({execute %>% group_by(state, population) %>%
summarise(Executions = n()) %>% ungroup() %>% mutate(Per0.1B =
round((Executions/population)*100000000))})

  output$plot1<-renderPlotly({
    ggplotly(execution_select() %>% top_n(input$num, wt=Executions) %>%
           ggplot(aes(x=reorder(state, Executions), y=Executions, fill=Executions, text=state)) +
           geom_bar(stat='identity') + coord_flip() +
           labs(x='', y='Number of Executions', title="Number of Executions by State") +
           scale_fill_gradient(low="yellow", high="red") +
           theme(legend.position="none"),
         tooltip=c("text", "y"))
  })

  output$plot2<-renderPlotly({
    ggplotly(execution_select() %>% top_n(input$num, wt=Per0.1B) %>%
           ggplot(aes(x=reorder(state, Per0.1B), y=Per0.1B, fill= Per0.1B, text=state)) +
           geom_bar(stat='identity') + coord_flip() +
           labs(x='', y='Executions Per0.1B', title="Relative number of Executions") +
           scale_fill_gradient(low="green", high="red") +
           theme(legend.position="none"),
         tooltip=c("text", "y"))
  })
```

```
  output$plot3<-renderPlotly({
    ggplotly(plot_ly(ExecuteRatio, labels = ~Race, values = ~Num, type = 'pie'))
  })

  output$table <- DT::renderDataTable(DT::datatable({execution_select() %>%
arrange(desc(Per0.1B))}))

  states1 <- reactive({tigris::geo_join(states, execution_select(), "NAME", "state", how="inner")})

  state_popup <- reactive({paste0("<strong>State: </strong>",
                                  states1()$NAME,
                                  "<br><strong>Executions Per0.1B inhabitants </strong>",
                                  states1()$Per0.1B) %>%
      lapply(htmltools::HTML)})

  pal <- reactive({colorNumeric("Reds", domain=states1()$Per0.1B)})

  output$usmap <- renderLeaflet({states1() %>%
      leaflet() %>%
      addProviderTiles("CartoDB.Positron") %>%
      setView(-98.483330, 38.712046, zoom = 4) %>%
      addPolygons(data = states1(),
                  fillColor = ~pal()(states1()$Per0.1B),
                  weight = 2,
                  opacity = 1,
                  color = "white",
                  dashArray = "3",
                  fillOpacity = 0.7,
                  highlight = highlightOptions(
                    weight = 5,
                    color = "#666",
                    dashArray = "",
                    fillOpacity = 0.7,
                    bringToFront = TRUE),
                  label = state_popup(),
                  labelOptions = labelOptions(
                    style = list("font-weight" = "normal", padding = "3px 8px"),
                    textsize = "15px",
                    direction = "auto")) %>%
      addLegend(pal = pal(),
                values = states1()$Per0.1B,
                position = "topright",
                title = "Per0.1B")})
  }
#}
shinyApp(ui = ui, server = server)
```