Assignment 3 — R Shiny Application

Sourabh D. Khot (Student ID 002754952)

College of Professional Studies, Northeastern University

ALY 6070: Communication and Visualization for Data Analytics CRN 81182

Professor Venkata Duvvuri

Jun 24, 2022

R Shiny Dashboard

I created the following dashboard using the 'TB_Burden_Country' dataset provided by the instructor. I have used seven variables below, of which six are directly from the dataset and one computed variable. These variables do not have any null values.

- Variables from the dataset: *Country, Year, Est. prevalence of TB (all forms) per 100K, Est. incidence (all forms) per 100K, Est. mortality of TB cases (all forms, excluding HIV) per 100K, and Est. mortality of TB cases who are HIV-positive per 100K*
- Computed variable: *Est. mortaility of TB cases (HIV and non-HIV) per 100K*

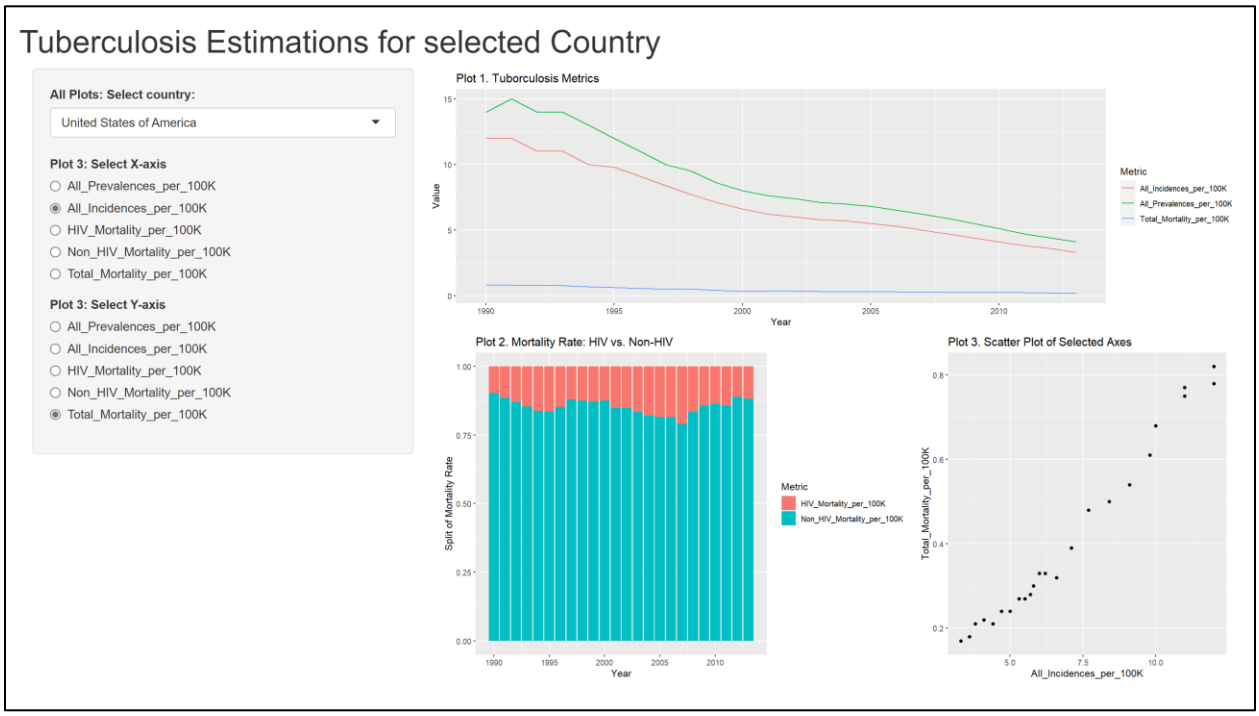

Figure 1. Tuberculosis Dashboard in R Shiny

Include as many different visualizations as needed. These should be of varying types

The dashboard includes three different visualizations: multiple series line chart, 100% stacked vertical bar chart, and scatter plot.

Answer the research/business question and display the key information that the intended audience needs

Each visualization answers the following research questions for the selected country.

- Plot 1: How prevalence, incidences, and total mortality change over the years.
- Plot 2: Within mortality, what is the relative share of HIV vs. Non-HIV cases.
- Plot 3: What is the relationship between selected variables? E.g., mortality rate (dependent variable) vs. incidence rate (independent variable), since incidence rate can be measured and mortality is intuitively correlated with incidence.

Be easy to navigate and visually appealing

The dashboard is implemented using R Shiny and ggplot and is easy and appealing.

Reflect the data accurately and generally communicate the data appropriately

The data is accurately and ethically represented without any modifications.

Tells a story

Overall, the visualizations tell a story about Tuborculosos of the selected country. Referring to figure 1, taking the United States of America as an example, we can see from the first plot that in 1990, per 100K population, there were 14 estimated cases of tuberculosis and 12 estimated incidences and mortality of little less than one. These numbers have steeply reduced to almost half near 2001, then gradually reducing till the latest data of 2013.

Concerning the split of mortality with respect to HIV and non-HIV cases, in 1990, about 10% of tuberculosis mortality was from HIV cases, which increased and peaked till 2007 and then again reduced to about 10% in 2013. Plot 3 shows that mortality has an excellent linear correlation with incidences. Hence, recorded estimated incidences may be used to predict mortality and plan hospital resources accordingly.

References

(n.d.). Retrieved from Mastering-Shiny.org: https://mastering-shiny.org/action-graphics.html

APA Style Table: APA.org. (n.d.). Retrieved from https://apastyle.apa.org/style-grammar-guidelines/tables-figures/tables

Canvas ALY6070: Assignment 3 RShiny Application. (n.d.). Retrieved from https://northeastern.instructure.com/courses/110061/assignments/1351349

Canvas ALY6070: Lesson 5-1 to 5-6 Storytelling with R Shiny: Importance of Context and Application. (n.d.). Retrieved from https://northeastern.instructure.com/courses/110061/pages/module-5-introduction

Knaflic, C. N. (2015). *Storytelling with Data - A Data Visualization Guide for Business Professionals.* New Jersey: Wiley.

R Code

```
library(shiny)
library(ggplot2)
library(tidyr)

df.full <- read.csv("TB_Burden_Country.csv", stringsAsFactors = TRUE, header = TRUE)
colnames(df.full)
summary(df.full)
df <- df.full[,c(1,6,8,28,21,15)]
colnames(df)
colnames(df) =
c("Country","Year","All_Prevalences_per_100K","All_Incidences_per_100K","HIV_Mortality_
per_100K","Non_HIV_Mortality_per_100K")
df["Total_Mortality_per_100K"] <- df["Non_HIV_Mortality_per_100K"] +
df["HIV_Mortality_per_100K"]

ui <- fluidPage(h1("Tuberculosis Estimations for selected Country"),

            sidebarPanel(selectInput(inputId = 'Country',
                                     label = "All Plots: Select country:",
                                     choices = unique(df["Country"]),
                                     selected = "United States of America"),
                         radioButtons(inputId = 'xData', label = "Plot 3: Select X-axis",
                                     choices = names(df[,3:7])),
                         radioButtons(inputId = 'yData', label = "Plot 3: Select Y-axis",
                                     choices = names(df[,3:7]),
                                     selected = names(df[,3:7])[[2]])
                         ),

            mainPanel(

            fluidRow(
              plotOutput(outputId = "Lineplot", width = "100%", height = "300px")),
              splitLayout(cellWidths = c("60%", "40%"), plotOutput("Barplot"),
                          plotOutput("Scatterplot"), height = "50px")
              ),

)
server <- function(input, output) {
```

```
    output$Lineplot <- renderPlot({

    ggplot(data=df[which(df["Country"]==input$Country),c(1:4,7)] %>%
             gather(Metric, Value, -c(Country, Year)),
             aes(x=Year, y=Value, colour=Metric)) +
       geom_line() +
       ggtitle("Plot 1. Tuborculosis Metrics")

     })

    output$Barplot <- renderPlot({

     ggplot(data=df[which(df["Country"]==input$Country),c(1:2,5:6)] %>%
              gather(Metric, Value, -c(Country, Year)),
             aes(x=Year, y=Value, fill=Metric)) +
       geom_bar(position="fill", stat="identity") +
       ggtitle("Plot 2. Mortality Rate: HIV vs. Non-HIV")+
       ylab("Split of Mortality Rate")

    })

    output$Scatterplot <- renderPlot({

     ggplot(data=df[which(df["Country"]==input$Country),],
             aes_string(x=input$xData, y=input$yData)) +
       geom_point() +
       ggtitle("Plot 3. Scatter Plot of Selected Axes")

     })

    }
shinyApp(ui = ui, server = server)
```

<End of Report>