Demographics vs. Income Report

Nushrate Ahmed

2024-10-05

## Assignment Scenario

In this scenario, you play the role of a data scientist hired by a company that provides career services to individuals with lower incomes. The marketing team, working with a small budget, seeks your help to identify which demographics (age, education, gender, etc.) to target for advertising in the US. Using census data, your task is to create an interactive Shiny dashboard to explore trends in various demographics and inform the marketing team’s strategy.

## Objectives

Your task is to create a Shiny dashboard using the UCI Adult dataset that provides insight into the following questions:

* Does education level impact salary?
* What is the relationship between age, sex, and salary?
* Which countries have lower incomes despite high education levels?
* Does the number of working hours contribute to lower income?
* Which work classes face lower incomes?

Afterwards, create a report using [**rmarkdown**](http://rmarkdown.rstudio.com/).

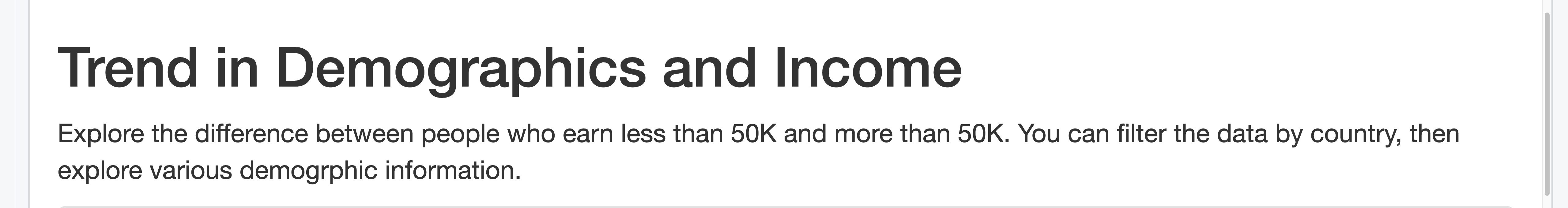
## Dataset Overview

The [**UCI Adult dataset**](https://dataplatform.cloud.ibm.com/exchange/public/entry/view/5fcc01b02d8f0e50af8972dc8963f98e?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDV0151ENSkillsNetwork26095306-2021-01-01&context=cpdaas), derived from census data, contains the following key variables:

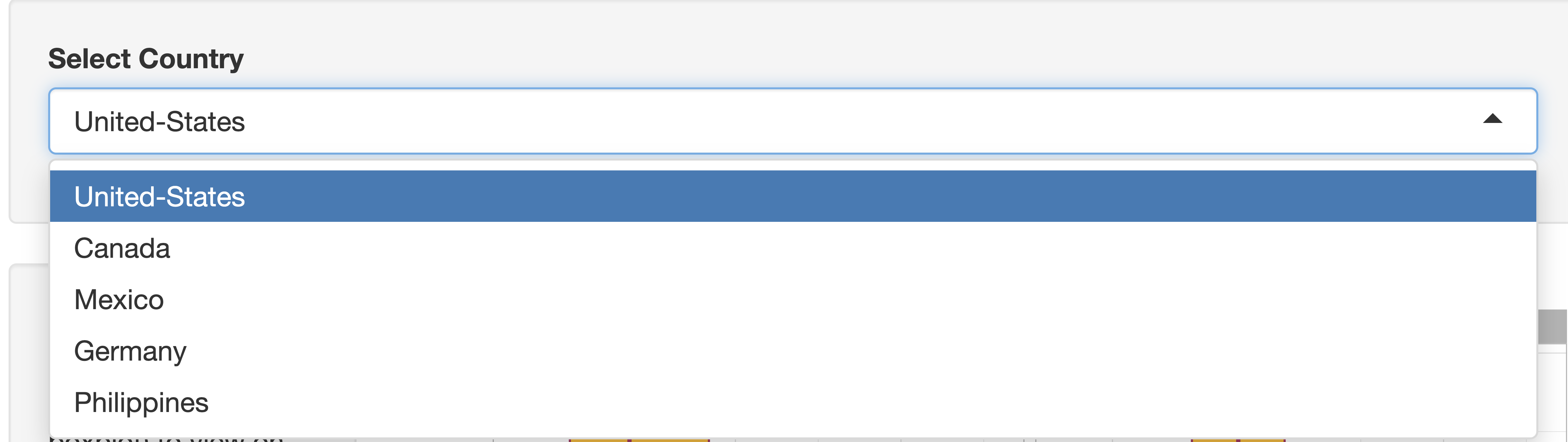
* **native\_country**: The person’s native country
* **age**: The person’s age
* **hours\_per\_week**: The number of hours worked per week
* **education**: The person’s education level
* **workclass**: The class of work (e.g., federal government, private)
* **sex**: Gender of the individual
* **prediction**: Income level, either <=50K or >50K, which is used to facet the graphs

### Tasks:

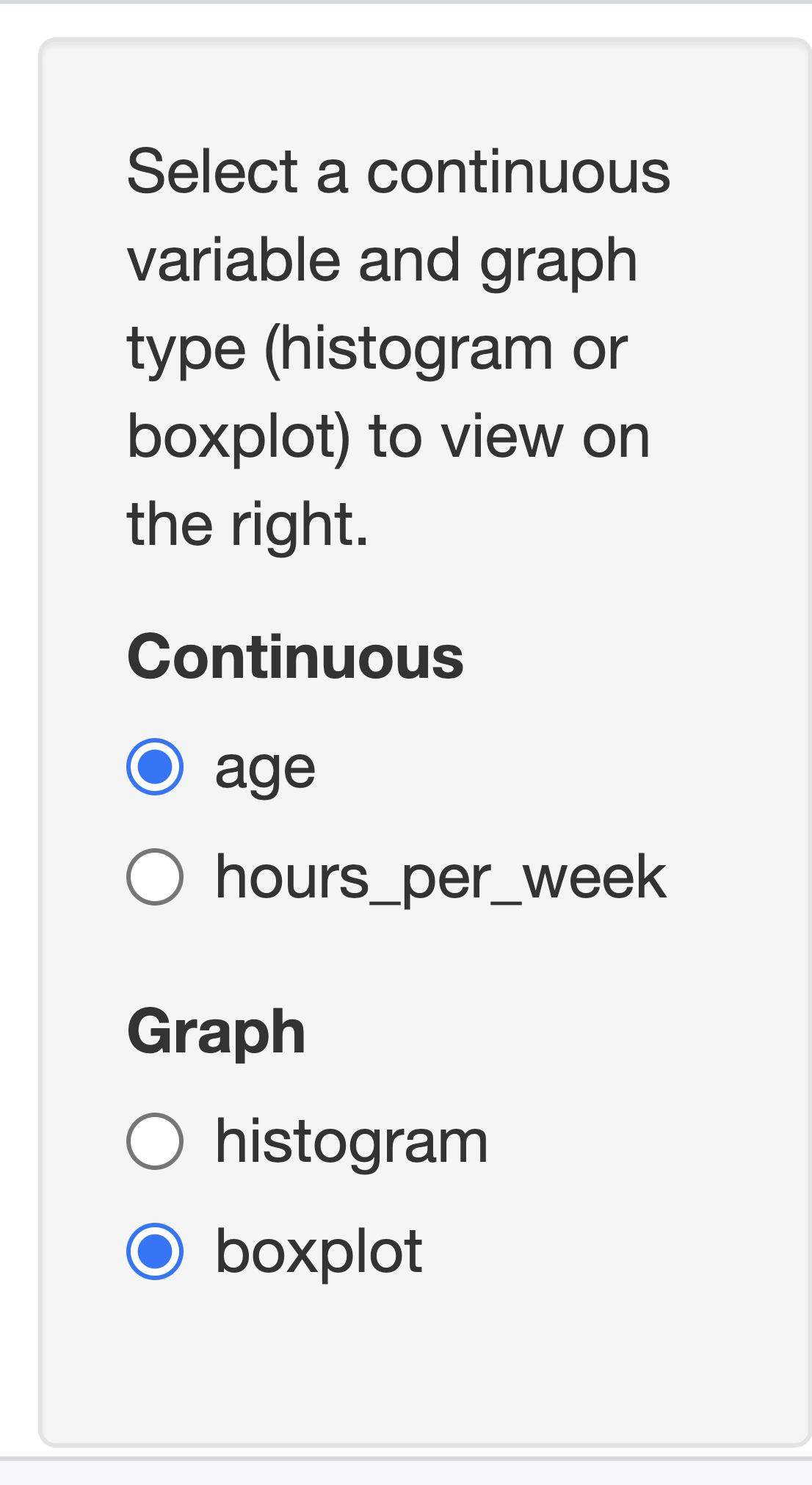
* **Task 1:** Add application title in the UI



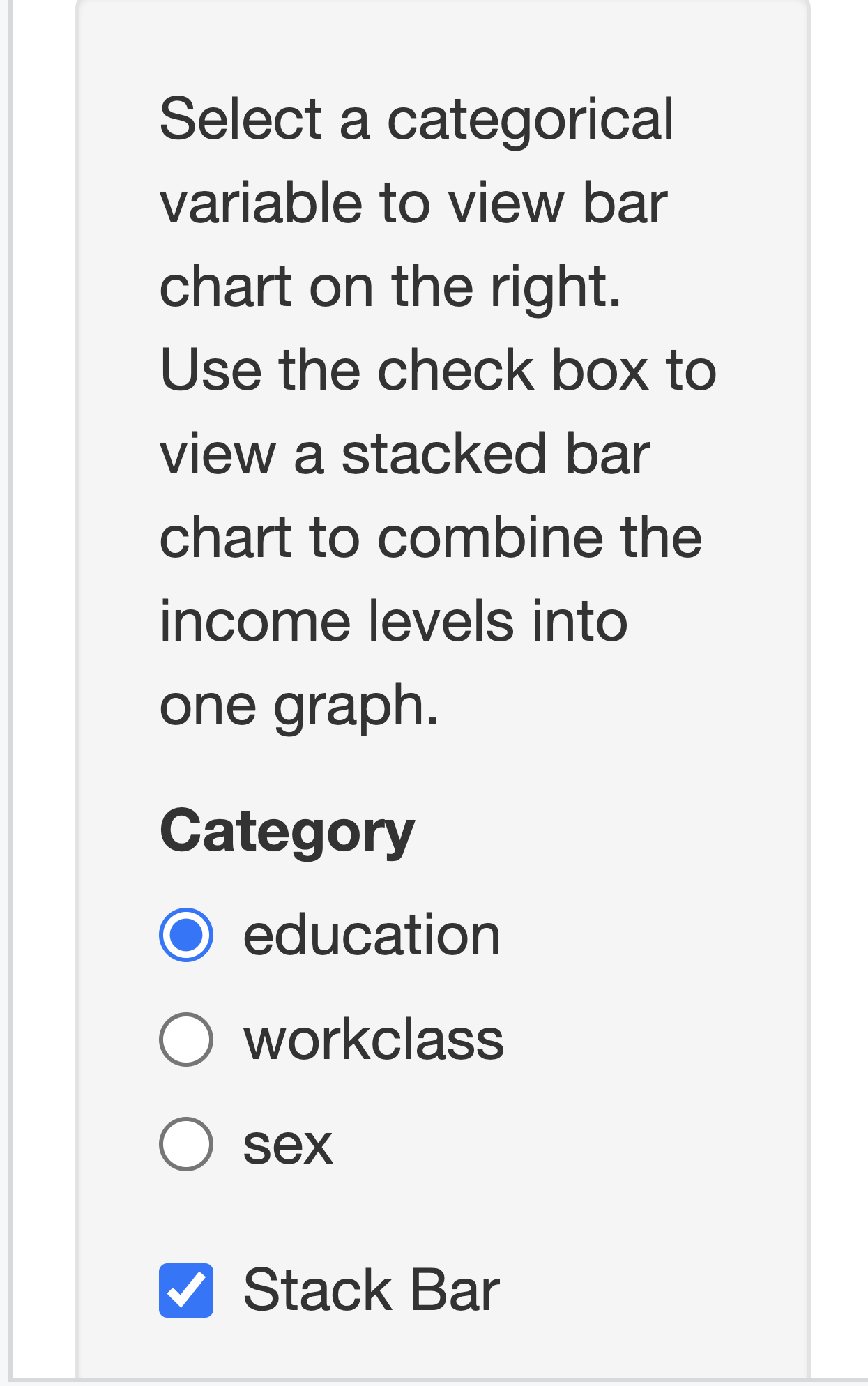
* **Task 2:** Add first fluidRow to select input for country in UI



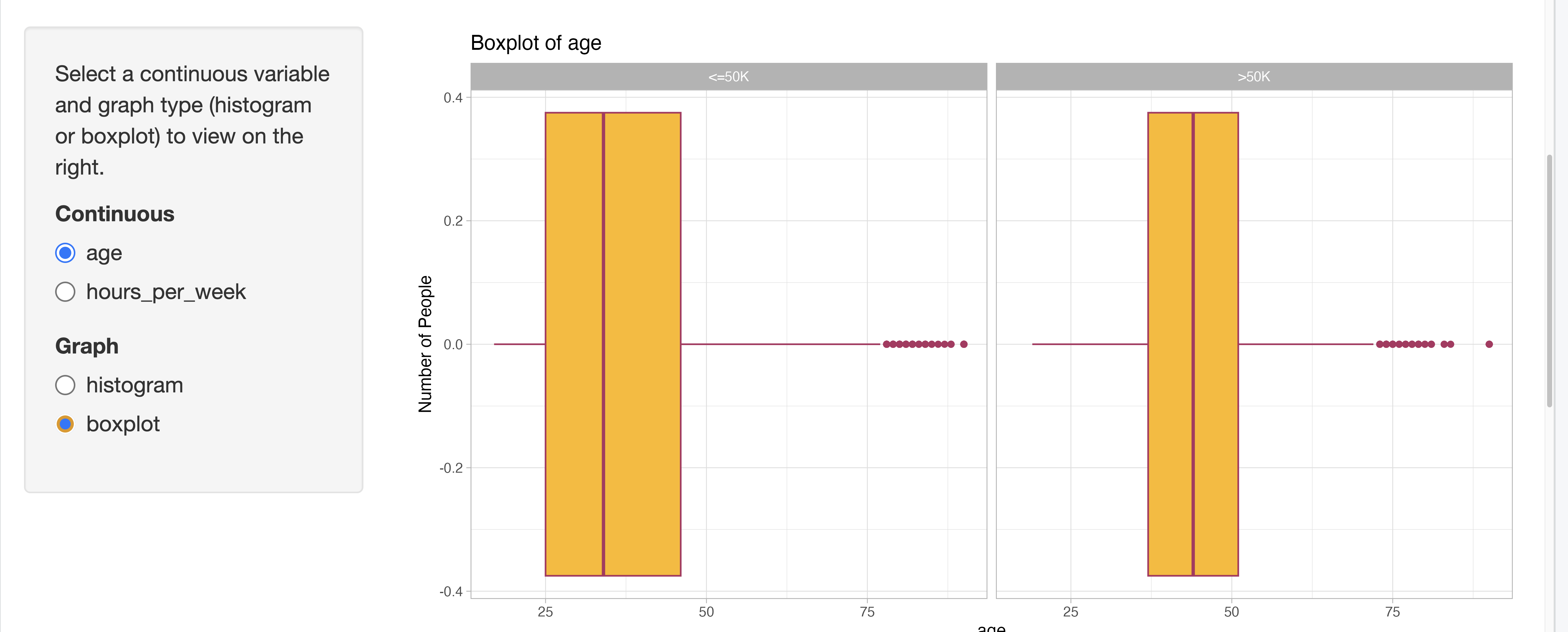
* **Task 3:**. Add second fluidRow to control how to plot the continuous variables in UI



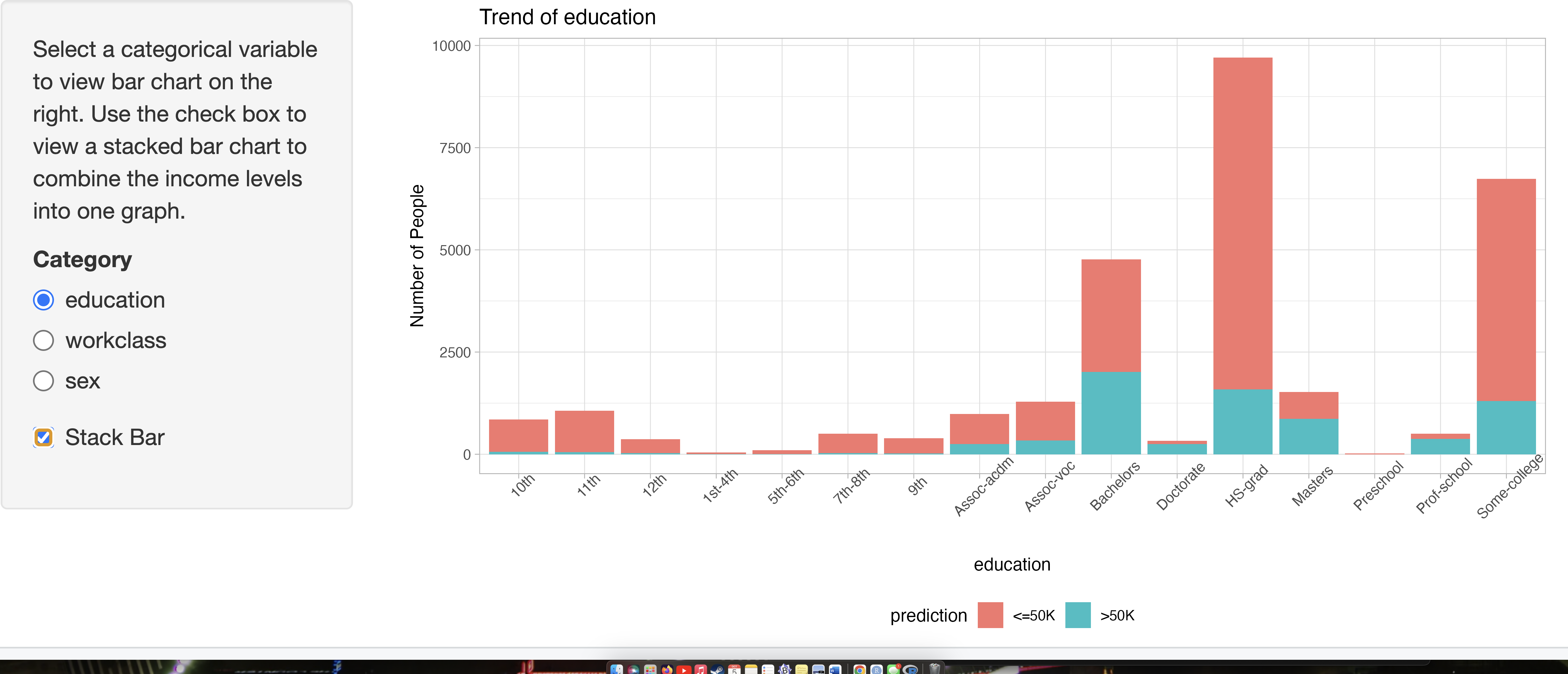
* **Task 4:** Add third fluidRow to control how to plot the categorical variables in UI



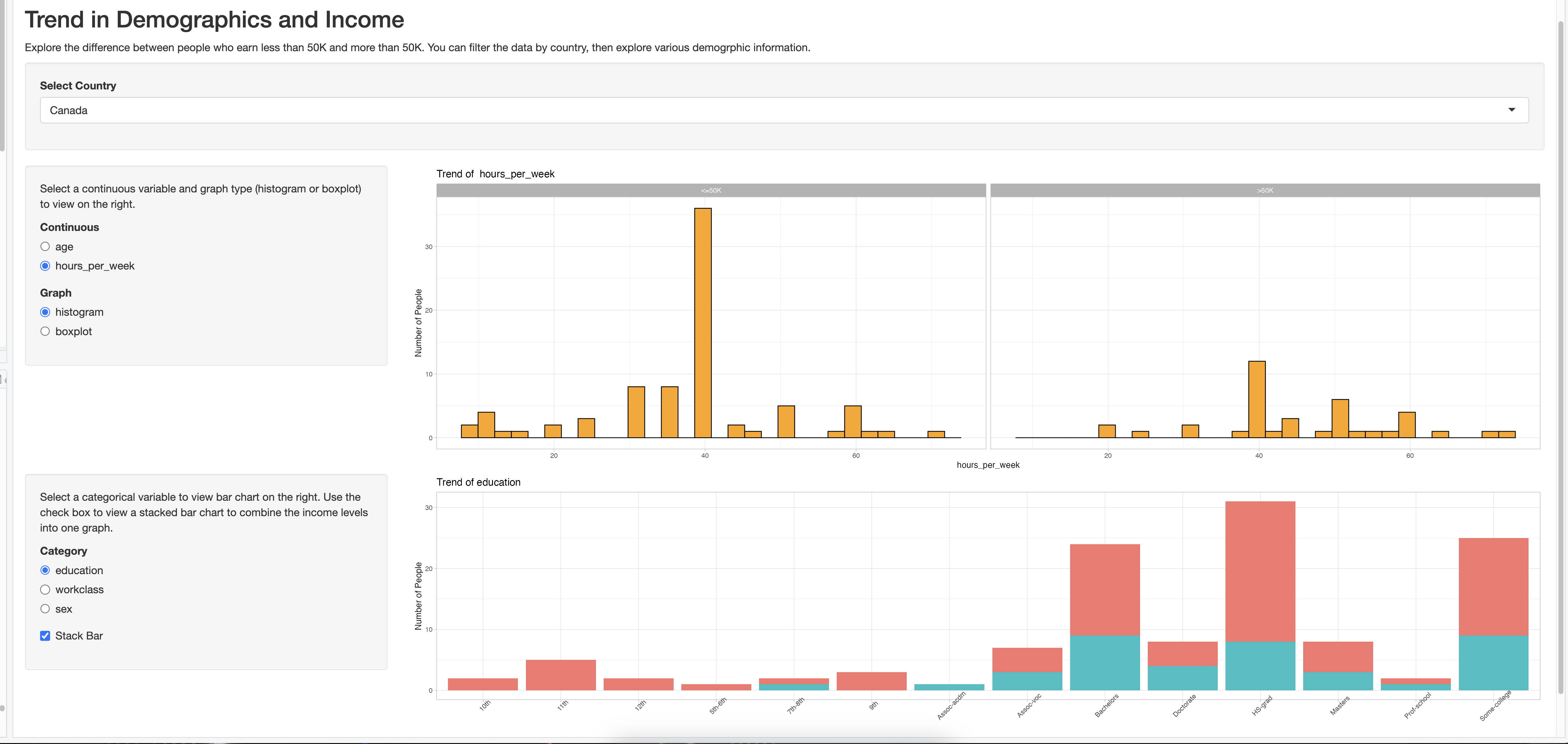
* **Task 5:** Create logic to plot histogram or boxplot in server



* **Task 6:** Create logic to plot faceted bar chart or stacked bar chart in server



* **Task 7:** Optionally change the themes to the graphs



### Final Source Code for ui.R file and server.R file

# Load libraries  
library(shiny)  
library(tidyverse)  
library(rio)  
library(ggplot2)  
library(ggthemes)  
  
# Uncomment ...  
# Load dataset  
# adult=import("/adult.csv")  
#  
  
# Application Layout  
shinyUI(fluidPage(  
 br(),  
 # TASK 1: Application title  
 titlePanel("Trend in Demographics and Income"),  
 p("Explore the difference between people who earn less than 50K and more than 50K. You can filter the data by country, then explore various demogrphic information."),  
   
 # TASK 2: Add first fluidRow to select input for country  
 fluidRow(  
 column(12,   
 wellPanel(selectInput("country","Select Country",choices=c("United-States","Canada","Mexico","Germany","Philippines")))  
 ) # add select input   
 ),  
   
 # TASK 3: Add second fluidRow to control how to plot the continuous variables  
 fluidRow(  
 column(3,   
 wellPanel(  
 p("Select a continuous variable and graph type (histogram or boxplot) to view on the right."),  
 radioButtons("continuous\_variable","Continuous",choices=c("age","hours\_per\_week")),   
 radioButtons("graph\_type","Graph",choices=c("histogram","boxplot"))   
 )  
 ),column(9, plotOutput("p1"))   
 ),  
   
 # TASK 4: Add third fluidRow to control how to plot the categorical variables  
 fluidRow(  
 column(3,   
 wellPanel(  
 p("Select a categorical variable to view bar chart on the right. Use the check box to view a stacked bar chart to combine the income levels into one graph. "),  
 radioButtons("categorical\_variable","Category",choices=c("education","workclass","sex")),   
 checkboxInput("is\_stacked","Stack Bar",value=FALSE)  
 )  
 ),  
 column(9, plotOutput("p2"))  
 )  
)  
)  
  
# Uncomment ...  
# Load libraries  
# library(shiny)  
# library(tidyverse)  
  
# Read in data  
# adult <- read\_csv("adult.csv")  
  
  
# Define server logic  
shinyServer(function(input, output) {  
   
 df\_country <- reactive({  
 adult %>% filter(native\_country == input$country)  
 })  
   
 # TASK 5: Create logic to plot histogram or boxplot  
 output$p1 <- renderPlot({  
 if (input$graph\_type == "histogram") {  
 # Histogram  
 ggplot(df\_country(), aes\_string(x =input$continuous\_variable)) +  
 geom\_histogram(color="black",fill="orange") +   
 labs(y="Number of People", title=paste("Trend of ",input$continuous\_variable)) +   
 facet\_wrap(~prediction)+   
 theme\_light()  
 }  
 else {  
 # Boxplot  
 ggplot(df\_country(), aes\_string(y = input$continuous\_variable)) +  
 geom\_boxplot(color="maroon",fill="darkgoldenrod1") +   
 coord\_flip() +   
 labs(x="Number of People", title=paste("Boxplot of",input$continuous\_variable)) +  
 facet\_wrap(~prediction)+   
 theme\_light()  
 }  
   
 })  
   
 # TASK 6: Create logic to plot faceted bar chart or stacked bar chart  
 output$p2 <- renderPlot({  
 # Bar chart  
 p <- ggplot(df\_country(), aes\_string(x =input$categorical\_variable)) +  
 labs(y="Number of People",title=(paste("Trend of",input$categorical\_variable))) +   
 theme\_light()+  
 theme(axis.text.x=element\_text(angle=45),legend.position="bottom")   
   
   
 if (input$is\_stacked) {  
 p + geom\_bar(aes(fill=prediction))   
 }  
 else{  
 p +   
 geom\_bar(aes\_string(fill=input$categorical\_variable)) +   
 facet\_wrap(~prediction)   
 }  
 })  
   
})

### Dashboard Summary

The interactive dashboard allows users to explore the UCI Adult dataset based on demographics, income levels, and various factors such as education, work class, and country. The findings from this analysis will assist the marketing team in targeting demographics for advertising more effectively.