Title: Web Scraping Assignment Author: Richard Kabiru Date: 2022-07-24

Setup

In this section, the setup for R Markdown is initiated, ensuring that the code chunks are echoed in the final document.

```
knitr::opts_chunk$set(echo = TRUE)
```

Introduction to R Markdown

This part provides a brief explanation of R Markdown, mentioning its capability to format documents for HTML, PDF, and MS Word outputs.

```
This is an R Markdown document. Markdown is a simple formatting syntax for authoring
HTML, PDF, and MS Word documents. For more details on using R Markdown see
<http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both
content as well as the output of any embedded R code chunks within the document.
```

Example: Summary of Cars Dataset

An example demonstrating how to embed R code chunks is shown using the `summary()` function on the `cars` dataset.

```
summary(cars)
```

Including Plots

This section illustrates how to embed plots within the document using the `plot()` function on the `pressure` dataset. The `echo = FALSE` parameter is used to prevent printing of the R code that generated the plot.

```
plot(pressure)
```

Web Scraping Eastern Africa Population Data

In this part, the `read_html` function from the `rvest` package is used to scrape population data of Eastern Africa from a webpage.

```
east_pop <- read_html("https://www.worldometers.info/world-population/eastern-africa-
population/#:~:text=The%20current%20population%20of%20Eastern,among%20subregions%20ranke
```

```
# Get the movie charts from wikipedia url using html_table function
east_africa <- html_table(east_pop)

east_africa[[3]]
```

Data Wrangling and Export

The scraped data is then converted into a data frame, its dimensions are checked, and it is subsequently saved as a CSV file in the specified directory.

```r
# Put the safaricom shares into a data frame
east_africa_df <- data.frame(east_africa[[3]])
east_africa_df

# Numbers of rows and columns
dim(east_africa_df)

# converting data frame into a csv and saving it on the working directory
getwd()
setwd("C:/Users/Rick-Royal/Documents")

write.csv(east_africa_df, "east_africa.csv", row.names = TRUE)
```

This document serves as a comprehensive documentation for the provided R script, explaining each section and the corresponding R code chunks.