SatelliteCountryScraper

This script scrapes information from N2YO.com to return a JSON file for use in a Swift project. There is no publicly available API call to return the owner of a satellite, so this script creates a JSON which allows the user to reference each satellite by its NORAD ID and find its country/owner.

Load Libraries

```
library(tidyverse)
library(jsonlite)
library(httr)
library(rvest)
```

Functions

- get_country_codes:
  - This function scrapes the website in the function's first line. It returns a Tibble with a single column "Code" which contains the list of country codes of countries with satellites currently in orbit
- get_country_link:
  - This function takes a country code as an argument and returns the link to the website containing a table of satellites that the country has currently in orbit.
- get_country_satellites:

  - This function also takes a country code as an argument, however this function calls the get_country_link function for the code and returns a Tibble with 3 columns: "Code", "Country", "NORAD_ID". Each row represents one satellite. The columns are the country code, the full name of the country, and the satellite's NORAD ID, respectively.

- create_satellite_tibble:

  - This function combines the three above functions to create a Tibble containing all satellites from all countries available from the N2YO website. The Tibble is the same format as the get_country_satellites Tibble.

```
get_country_codes <- function() {
  n2yo_page <- read_html("https://www.n2yo.com/satellites/?c=&t=country")

  n2yo_table <- n2yo_page %>%
    html_elements("table") %>%
    .[[3]] %>% # select the second table on the page
    html_elements("td:nth-child(3)") %>% # select the first column of each row
    html_text() %>% # extract the text content
    as_tibble()

  colnames(n2yo_table) <- "Code"

  return(n2yo_table)
}

get_country_link <- function(countryCode) {
  url_prefix <- "https://www.n2yo.com/satellites/?c="
  url_suffix <- "&t=country"
  full_url <- paste(url_prefix, countryCode, url_suffix, sep = "")
}

get_country_satellites <- function(countryCode) {
  countryURL <- get_country_link(countryCode)
  countryPage <- read_html(countryURL)

  country_table <- countryPage %>%
  html_elements("table") %>%
  .[[2]] %>% # select the second table on the page
  html_elements("td:nth-child(2)") %>% # select the first column of each row
  html_text() %>% # extract the text content
```

```
  as_tibble()

  country_name <- country_table %>%
  slice(2) %>%
  pull()

  country_table <- country_table %>%
    mutate(Code = countryCode, Country = country_name) %>%
    slice(-c(1:2)) %>% # convert to tibble and remove the first two rows
    select(Code, Country, value) %>%
    rename(NORAD_ID = value)
}

create_satellite_tibble <- function() {
  countryCodes <- get_country_codes()
  result_tibble <- tibble(Code = character(), Country = character(), NORAD_ID = character(

  for (i in 1:nrow(countryCodes)) {
    code <- countryCodes$Code[i]
    temp_sats <- try(get_country_satellites(code), silent = TRUE)
    result_tibble <- result_tibble %>% add_row(temp_sats)
  }

  return(result_tibble)
}
```

Create the Tibble of Individual Satellites

```
satellites <- create_satellite_tibble()
```

Create a JSON file using the Tibble

The following code converts the satellites Tibble into the output JSON file "satellite_by_country.json". The format of this (prettified) JSON is as follows:

[

{

"code": <character, the country/owner code>,

“name”: <character, the country/owner name,

“satellites”: [

<character, a singular satellite NORAD ID>,

…

]

},

…

]

```r
satellites_grouped <- group_by(satellites, Code)

# create a nested list structure
json_data <- lapply(unique(satellites$Code), function(code) {
  country_name <- trimws(unique(satellites_grouped$Country[satellites_grouped$Code == code
  sat_ids <- satellites_grouped$NORAD_ID[satellites_grouped$Code == code]
  list(code = code, name = country_name, satellites = as.list(sat_ids))
})

# create a JSON file from the list
writeLines(toJSON(json_data, auto_unbox = TRUE), "satellites_by_country.json")
```