Downloading the data

First I identified the webpages from the NRS website that contained the required babynames csv files and then scraped the links to all the csv files with help from the rvest package. I created some helper functions (one to grab the csv links and one to read the csv files into R and tidy them up) to use with the `map()` functions from purrr.

```
library(tidyverse)
library(janitor)
library(rvest)

# Helper functions
get_csv_links <- function(link) {
  read_html(link) %>%
    html_nodes("a") %>%
    html_attr("href") %>%
    str_subset("\\.csv") %>%
    paste0("https://www.nrscotland.gov.uk/", .)
}

read_babynames <- function(link, yr) {
  b <- read_csv(link, skip = 6) %>%
    remove_empty() %>%
    select(-contains("Position")) %>%
    clean_names()

  boy <- b %>%
    select(1:2) %>%
    set_names(c("name", "number_of_babies")) %>%
    mutate(gender = "boy")

  girl <- b %>%
    select(3:4) %>%
    set_names(c("name", "number_of_babies")) %>%
    mutate(gender = "girl")

  bind_rows(boy, girl) %>%
    mutate(year = yr) %>%
    filter(!is.na(number_of_babies))
}

# List of webpages containing the csv files
pages <- c("https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-archive/full-lists-of-babies-first-names-1974-to-1979",
           "https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-archive/full-lists-of-babies-first-names-1980-to-1989",
           "https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-archive/full-lists-of-babies-first-names-1990-to-1999",
           "https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-2000-to-2009",
           "https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/names/babies-first-names/full-lists-of-babies-first-names-2010-to-2014")
```

```
csv_links <- map(pages, get_csv_links) %>%
  unlist()

# Find the years for each csv file
yr <- parse_number(str_extract(csv_links, "[0-9]+\\.csv")) %>%
  if_else(is.na(.), 2018, .) %>%
  if_else(. < 1000, . + 2000, .)

babynames <- map2_df(csv_links, yr, read_babynames)

babynames2 <- babynames %>%
  mutate(decade = paste0(str_sub(year, 1, 3), "0s")) %>%
  group_by(decade, gender, name) %>%
  summarise(number_of_babies = sum(number_of_babies)) %>%
  ungroup()

# Save as rds so it can be quickly read in for the Shiny app
saveRDS(babynames2, "babynames.rds")
```

Shiny App

I created the Shiny app by amending the Shiny template available in RStudio as required – all fairly straightforward stuff and nothing fancy involved at all!

```
# Shiny App: Scotland's most popular babynames by decade

library(shiny)
library(dplyr)
library(ggplot2)
library(scales)
library(stringr)
theme_set(theme_minimal(base_size = 14))

babynames <- readRDS("babynames.rds")


ui <- fluidPage(

    titlePanel("Scotland's Most Popular Babynames"),

    sidebarLayout(
        sidebarPanel(
            selectInput("decade", "Born in Decade:",
                        c("1970s" = "1970s",
                          "1980s" = "1980s",
                          "1990s" = "1990s",
                          "2000s" = "2000s",
                          "2010s" = "2010s")),
            radioButtons("gender", "Gender:",
                         c("Boy" = "boy",
                           "Girl" = "girl")),

            textInput("name_start", "Name starts with", ""),
        ),

        mainPanel(
            plotOutput("barPlot")
```

```
        )
    )
)


server <- function(input, output) {

    output$barPlot <- renderPlot({
        babynames %>%
            filter(decade == input$decade,
                   gender == input$gender,
                   str_detect(name, paste0("^", str_to_title(input$name_start)
))) %>%
            arrange(desc(number_of_babies)) %>%
            mutate(perc = number_of_babies / sum(.$number_of_babies),
                   name = factor(name, levels = rev(.$name))) %>%
            slice(1:20) %>%
            ggplot(aes(x = name, y = perc)) +
            geom_bar(stat = "identity", fill = "orange", width = 0.7) +
            scale_y_continuous(labels = percent, limits = c(0, 1)) +
            labs(x = NULL, y = NULL,
                 caption = "Source: National Records of Scotland\nBabynames Data
1974-2018") +
            coord_flip() +
            theme(panel.grid.major.y = element_blank())
    })
}


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