

TP 4

Yun Shen (Yoshi) Li

2024-03-19

1. Comme je n'avais pas eu d'idée de sites à scraper, j'ai fait quelque chose similaire à ce qui a été présenté dans l'exemple 1 de l'énoncé: un scraping sur la page Wikipedia sur le ranking historique des premiers ministres du Canada parce que la table m'a paru facile à lire et donc facile à extraire des données. Mon idée était de créer une visualisation du ranking des PM durant leurs termes respectifs et voir le changement des opinions publiques à travers le temps. De plus, les rankings sont tirés de Maclean's en 1997, 2011 et 2016, donnant une visualisation plus récentes des rankings.
2. Bon, malheureusement j'ai essayé une tonne de façons pour debugger le problème avec la deuxième colonne contenant le nom des premier ministres qui semble être juste un tas de gunk html que je n'arrive pas à transformer en colonne et je ne sais plus quoi faire.

```
library(tidyverse)
library(lubridate)
library(rvest)

#1. scrapey scrapes

pm_rankings <- read_html(
  "https://en.wikipedia.org/wiki/Historical_rankings_of_prime_ministers_of_Canada") |>
  html_elements("table") |>
  pluck(2) |>
  html_table(fill = TRUE) |>
  rename(pm_name = "Prime Minister",
         party = "Political party",
         mac_97 = "Maclean's 1997[2]",
         mac_11 = "Maclean's 2011[3]",
         mac_16 = "Maclean's 2016[4]") |>
  filter(!is.na(pm_name)) |> #take out where the values are as NA
  select(pm_name, party, mac_97, mac_11, mac_16)

glimpse(pm_rankings)
```

Si il n'y avait eu aucun problème, les variables devrait être comme suivant: la séquence du premier ministre, son nom, son affiliation de parti politique, son ranking dans Maclean's en 1997, 2011 et 2016, son ranking général et finalement son ranking par temps au siège du PM.

3. Je voulais faire un graphique à point avec une ligne reliant les points indiquant le changement de ranking à travers le temps. De plus, j'aurais aimé faire un graphique à barre comparant le ranking des PM par parti (conservateur, libéral, conservateur progressif).
4. Généralement, j'aurais fait la remarque qu'il faut faire attention avec les données provenant d'un web-scraping (même si ce sont des données publique) et l'utilisation de ces données extraits du web, simplement car il est très facile de les utiliser dans une manière qui n'est pas attendue par le collecteur

original de ces données. Mais puisque j'ai scrapé de Wikipedia et que ce sont des données qui sont facilement retrouvables chez Maclean's, je ne crois pas avoir fait quelque chose de mal éthiquement. En plus de ça, Wikipedia est une plateforme basée sur le partage de connaissances et données à tous et toutes, donnant plus ou moins leur consentement pour faire du scraping. Au contraire, si j'avais fait un scraping sur un site pas aussi ouvert que Wikipedia (ex: un journal), il serait mieux de demander l'autorisation au cas où les auteurs sont contre l'idée que l'on utilise leurs informations et données pour des raisons autres que leurs intentions initiales.