

## Lab 04 - Money in U.S. politics

Ivanna Diez de Bonilla (id239)

Aditya Mittal (am3427)

Jemimah Osei (jko35)

2025-03-10

### Setup

Load packages and data:

```
library(tidyverse)
```

```
— Attaching core tidyverse packages — tidyverse 2.0.0
—
✓ dplyr      1.1.4    ✓ readr      2.1.5
✓ forcats    1.0.0    ✓ stringr    1.5.1
✓ ggplot2    3.5.1    ✓ tibble     3.2.1
✓ lubridate  1.9.4    ✓ tidyr      1.3.1
✓ purrr      1.0.4
— Conflicts — tidyverse_conflicts()
—
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
  conflicts to become errors
```

```
library(scales)
```

Attaching package: 'scales'

The following object is masked from 'package:purrr':

discard

The following object is masked from 'package:readr':

col\_factor

```
library(rvest)
```

Attaching package: 'rvest'

The following object is masked from 'package:readr':

guess\_encoding

```
library(robotstxt)
paths_allowed("https://www.opensecrets.org")
```

www.opensecrets.org

```
[1] TRUE
```

## Exercises

### Exercise 1

See the code below stored in lab-04-scrape-pacs.R.

```
# load packages
library(tidyverse)
library(rvest)

# function: scrape_pac
scrape_pac <- function(url) {

  year_extracted <- str_sub(url, -4)
  local_path <- paste0("data/pac/", year_extracted, ".html")

  # read the page
  page <- read_html(local_path)

  # extract the table
  pac <- page |>
    # select node .DataTable (identified using the SelectorGadget)
    html_element(".DataTable-Partial") |>
    # parse table at node td into a data frame
    # table has a header
    html_table(header = TRUE) |>
```

```

# convert to a tibble
as_tibble()

# rename variables
pac <- pac |>
# rename columns
rename(
  name = `PAC Name (Affiliate)`,
  country_parent = `Country of Origin/Parent Company`,
  total = Total,
  dems = Dems,
  repubs = Repubs
)

# fix name
pac <- pac |>
# remove extraneous whitespaces from the name column
mutate(name = str_squish(name))

# add year
pac <- pac |>
# extract last 4 characters of the URL and save as year
mutate(year = year_extracted)

# return data frame
return(pac)
}

# test function
url_2022 <- "https://www.opensecrets.org/political-action-committees-pacs/foreign-connected-pacs/2022"
pac_2022 <- scrape_pac(url_2022)

url_2020 <- "https://www.opensecrets.org/political-action-committees-pacs/foreign-connected-pacs/2020"
pac_2020 <- scrape_pac(url_2020)

url_2000 <- "https://www.opensecrets.org/political-action-committees-pacs/foreign-connected-pacs/2000"
pac_2000 <- scrape_pac(url_2000)

# list of urls

# first part of url
root <- "https://www.opensecrets.org/political-action-committees-pacs/foreign-connected-pacs/"

year <- seq(from = 2000, to = 2022, by = 2)

```

```
# construct urls by pasting first and second parts together
urls <- str_glue("{root}{year}")

# map the scrape_pac function over list of urls
pac_all <- map_dfr(urls, scrape_pac, .progress = TRUE)

# write data
write_csv(pac_all, file = "data/pac-all.csv")
```

## Exercise 2

```
pac_all <- read_csv("data/pac-all.csv")
```

Rows: 2427 Columns: 6

	Column	specification
Delimiter:	" , "	
chr (5):	name, country_parent, total, dems, repubs	
dbl (1):	year	

i Use `spec()` to retrieve the full column specification for this data.  
i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
pac_all_sep <- pac_all |>
  separate_wider_delim(cols = country_parent,
    delim = "/",
    names = c("country", "parent"),
    too_many = "merge")

print(head(pac_all_sep, 10))
```

```
# A tibble: 10 × 7
```

	name	country	parent	total	dems	repubs	year
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>
1	7-Eleven	Japan	Ito-Y...	\$8,5...	\$1,5...	\$7,000	2000
2	ABB Group	Switze...	Asea ...	\$46,...	\$17,...	\$28,5...	2000
3	Accenture	UK	Accen...	\$75,...	\$23,...	\$52,9...	2000
4	ACE INA	UK	ACE G...	\$38,...	\$12,...	\$26,0...	2000
5	Acuson Corp (Siemens AG)	Germany	Sieme...	\$2,0...	\$2,0...	\$0	2000
6	Adtranz (DaimlerChrysler)	Germany	Daiml...	\$10,...	\$10,...	\$500	2000
7	AE Staley Manufacturing (Tate & Lyle)	UK	Tate ...	\$24,...	\$10,...	\$14,0...	2000
8	AEGON USA (AEGON NV)	Nether...	Aegon...	\$58,...	\$10,...	\$47,7...	2000
9	AIM Management Group	UK	AMVES...	\$25,...	\$10,...	\$15,0...	2000
10	Air Liquide America	France	L'Air...	\$0	\$0	\$0	2000

The data set contains 2427 observations and 6 variables.

### Exercise 3

```
pac_all_numeric <- pac_all_sep |>
  mutate(
    total = parse_number(str_remove_all(total, "\\$")),
    dems = parse_number(str_remove_all(dems, "\\$")),
    repubs = parse_number(str_remove_all(repubs, "\\$"))
  )
pac_all_numeric
```

```
# A tibble: 2,427 × 7
  name                country parent total  dems repubs year
  <chr>               <chr>   <chr> <dbl> <dbl> <dbl> <dbl>
1 7-Eleven           Japan  Ito-Y... 8500  1500   7000  2000
2 ABB Group          Switze... Asea ... 46000 17000  28500  2000
3 Accenture          UK      Accen... 75984 23000  52984  2000
4 ACE INA            UK      ACE G... 38500 12500  26000  2000
5 Acuson Corp (Siemens AG) Germany Sieme... 2000  2000    0  2000
6 Adtranz (DaimlerChrysler) Germany Daiml... 10500 10000    500  2000
7 AE Staley Manufacturing (Tate & Lyle) UK      Tate ... 24000 10000  14000  2000
8 AEGON USA (AEGON NV) Nether... Aegon... 58250 10500  47750  2000
9 AIM Management Group UK      AMVES... 25000 10000  15000  2000
10 Air Liquide America France  L'Air...    0    0    0  2000
# i 2,417 more rows
```

### Exercise 4

```
pac_filtered <- pac_all_numeric |>
  filter(country %in% c("Canada", "Mexico"))

# Step 2: Summarize total contributions by year and country
pac_summary <- pac_filtered |>
  group_by(year, country) |>
  summarize(total_contributions = sum(total, na.rm = TRUE), .groups = "drop")

print(pac_summary)
```

```
# A tibble: 23 × 3
  year country total_contributions
  <dbl> <chr>          <dbl>
1  2000 Canada      224815
2  2000 Mexico       33147
3  2002 Canada      214375
4  2002 Mexico         2000
5  2004 Canada      282025
```

```

6 2004 Mexico 0
7 2006 Canada 264512
8 2008 Canada 440050
9 2008 Mexico 178100
10 2010 Canada 699700
# i 13 more rows

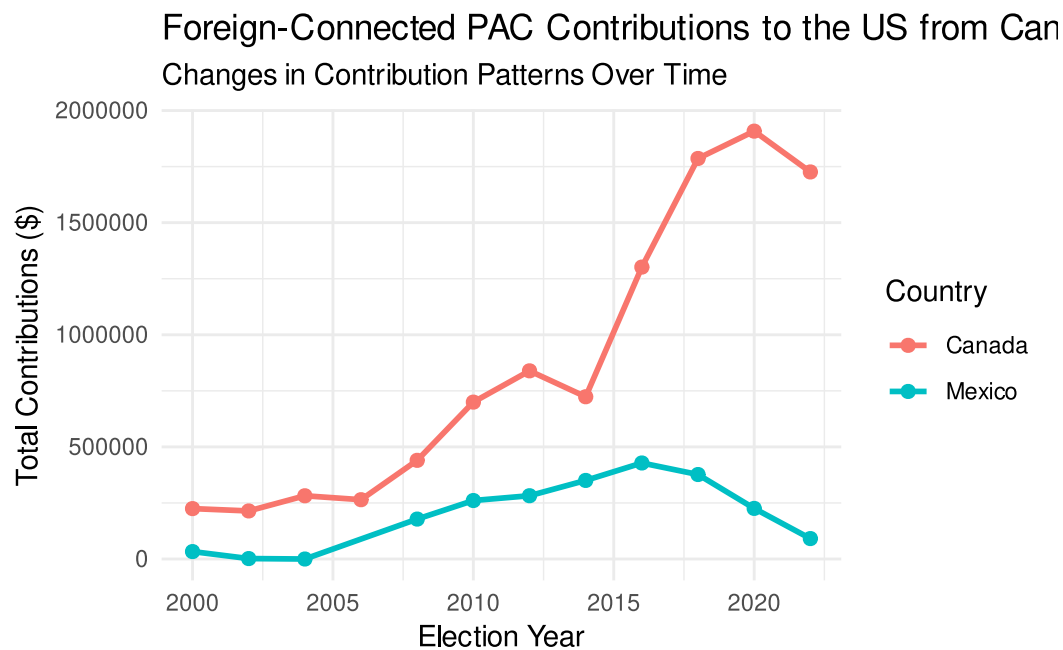
```

```

# Step 3: Create a line plot
ggplot(pac_summary, aes(x = year, y = total_contributions, color = country, group
= country)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  scale_x_continuous(breaks = seq(2000, 2022, by = 5)) +
  labs(
    title = "Foreign-Connected PAC Contributions to the US from Canada & Mexico",
    subtitle = "Changes in Contribution Patterns Over Time",
    x = "Election Year",
    y = "Total Contributions ($)",
    color = "Country"
  ) +
  theme_minimal()

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
 i Please use `linewidth` instead.



The graph shows that Canada's foreign-connected PAC contributions have increased significantly over time, peaking around 2020, while Mexico's contributions saw moderate growth until 2016 but have since declined. Canada's contributions have consistently been higher than Mexico's, with a sharp rise after 2014. This suggests that Canadian entities have become more active in U.S. political funding, while contributions from Mexico have declined, possibly due to economic or regulatory factors.

## Exercise 5

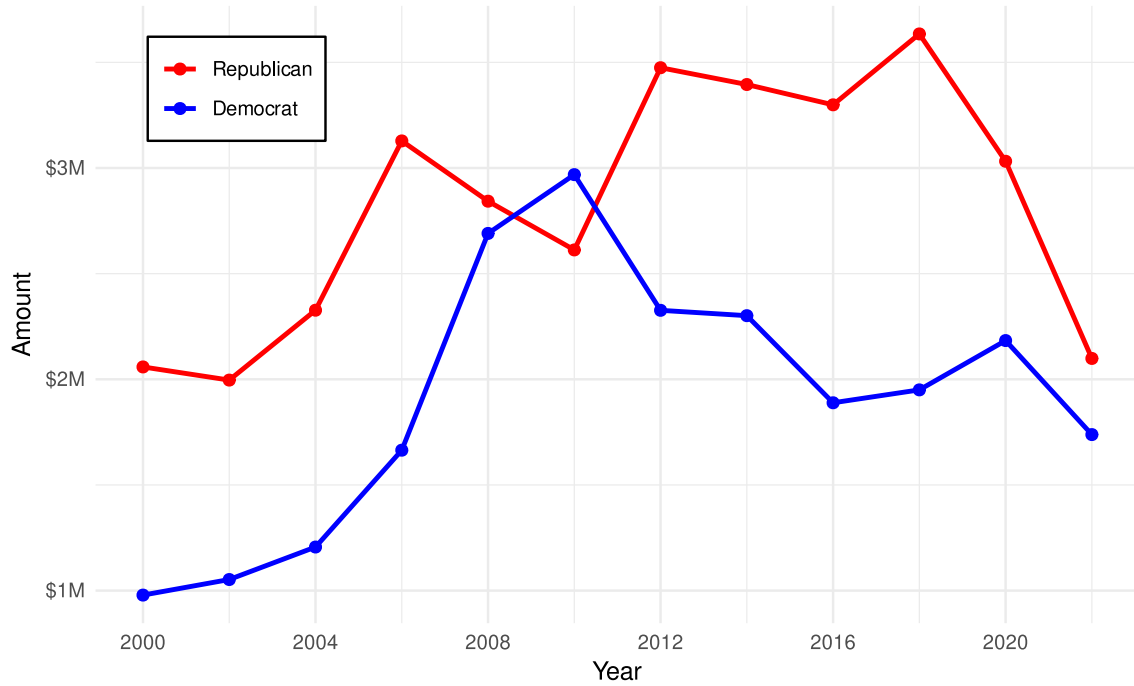
```
UK_pac <- pac_all_numeric |>
  filter(country == "UK")

uk_summary <- UK_pac |>
  pivot_longer(cols = c(dems, repubs), names_to = "party", values_to = "amount")
|>
  mutate(party = recode(party, "dems" = "Democrat", "repubs" = "Republican")) |
  >
  mutate(party = factor(party, levels = c("Republican", "Democrat"))) |>
  group_by(year, party) |>
  summarize(total_contributions = sum(amount, na.rm = TRUE), .groups = "drop")

ggplot(uk_summary, aes(x = year, y = total_contributions, color = party, group
= party)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = c("Republican" = "red", "Democrat" = "blue")) +
  scale_x_continuous(breaks = seq(2000, 2024, by = 4)) +
  scale_y_continuous(labels = label_dollar(scale = 1e-6, suffix = "M")) +
  labs(
    title = "Contributions to US politics from UK-Connected PACs",
    subtitle = "By party, over time",
    x = "Year",
    y = "Amount",
    caption = "Source: Money in Politics",
    color = NULL
  ) +
  theme_minimal() +
  theme(
    legend.position = c(0.05, 0.95),
    legend.justification = c(0, 1),
    legend.background = element_rect(fill = "white")
  )
```

Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2 3.5.0.  
 i Please use the `legend.position.inside` argument of `theme()` instead.

Contributions to US politics from UK-Connected PACs  
By party, over time



Source: Money in Politics

The graph illustrates UK-connected PAC contributions to U.S. political campaigns over time, categorized by party affiliation (Republican and Democrat). Republican contributions have generally been higher, peaking around 2012 and 2018, but experienced a significant decline after 2018, with a sharp drop post-2020. Democratic contributions grew steadily until 2008, briefly surpassing Republican contributions, but then declined and remained consistently lower in recent years. This trend suggests that UK-connected PACs have historically favored Republican candidates, although their overall contributions have fluctuated over time.