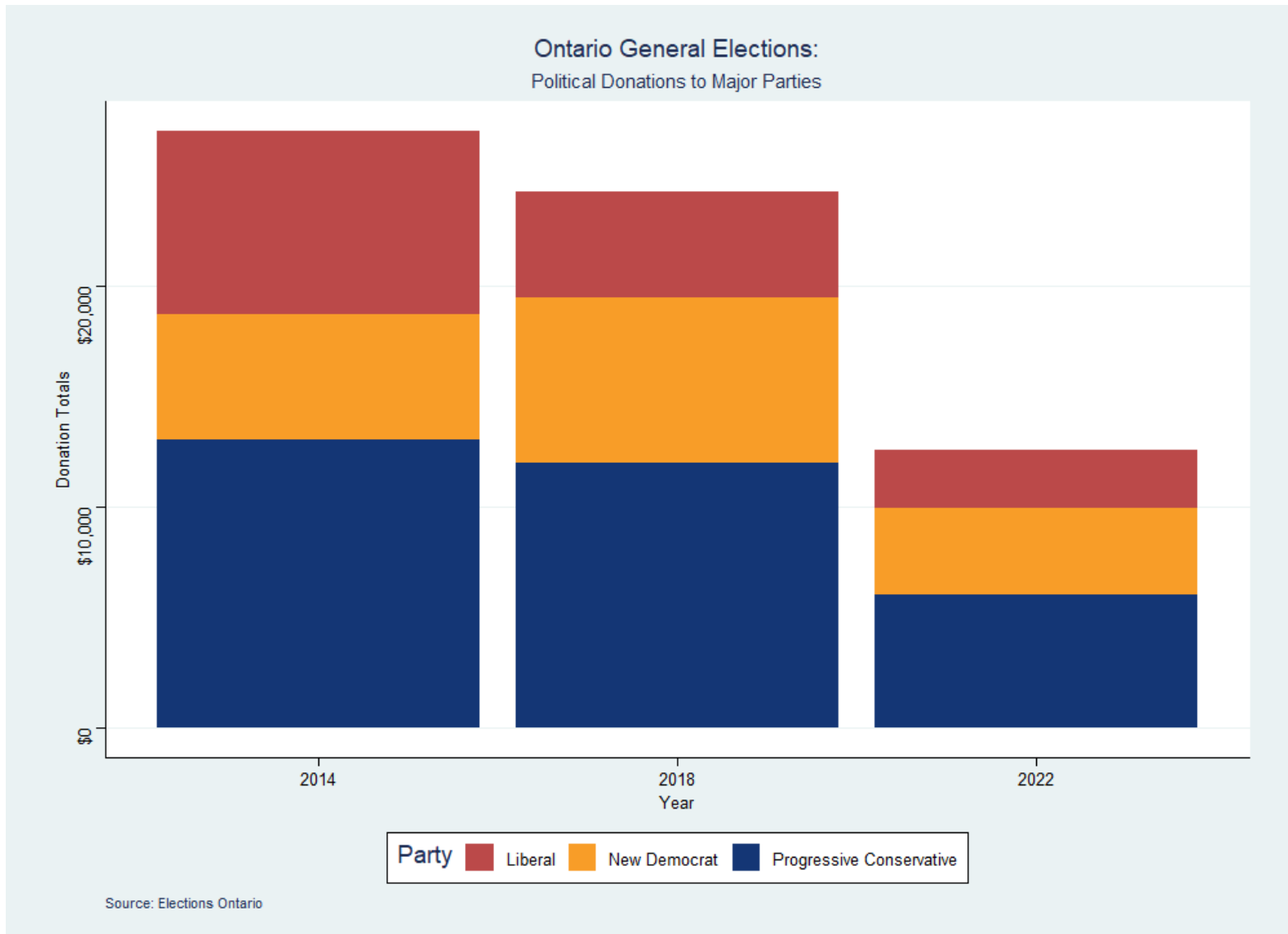


## DSI Dataviz Assignment #4

*Please refer to original image files in GitHub.*

Example One:

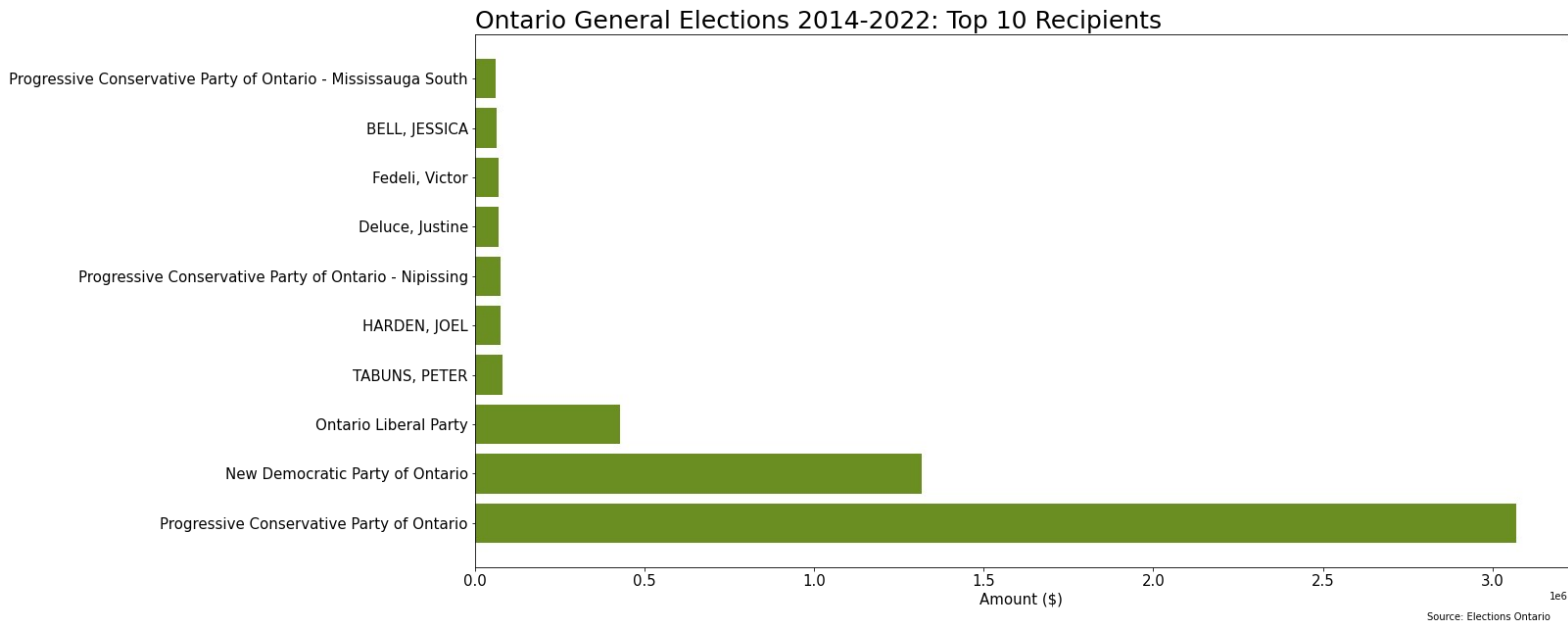
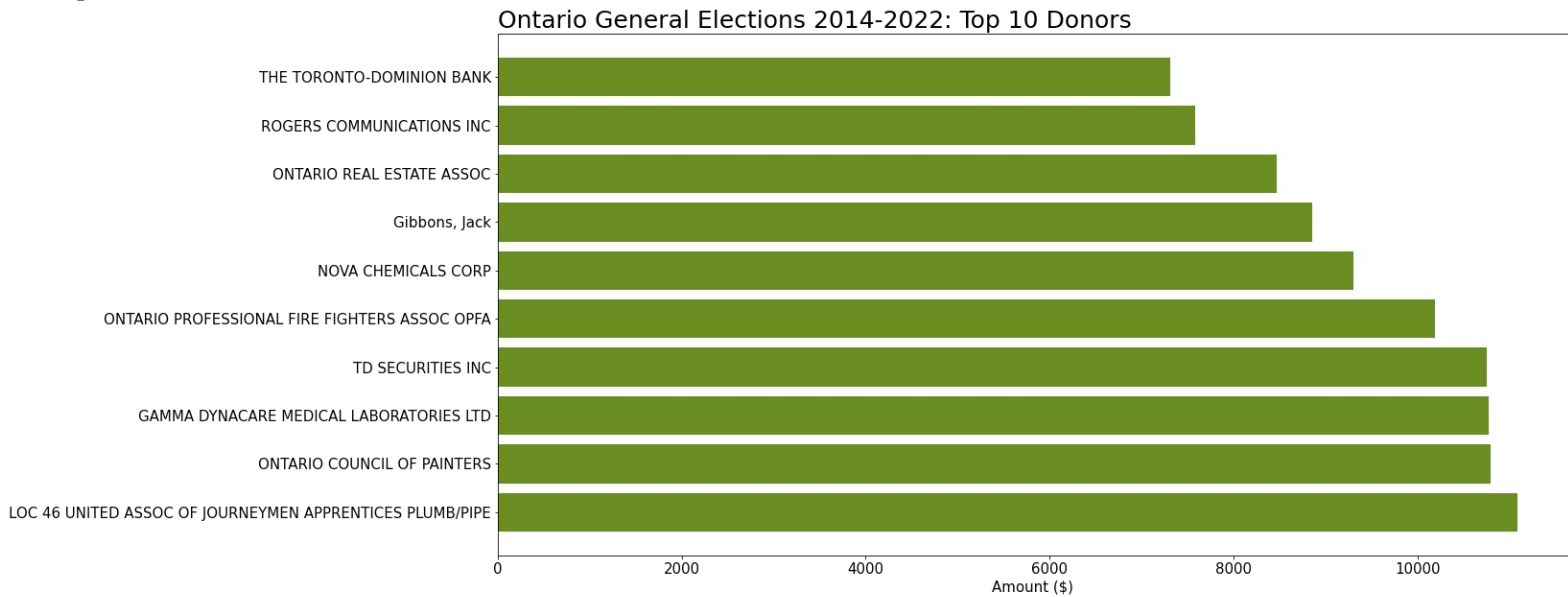


- Software: R Studio
- Intended Audience: Ontario voter, shared online.
- Message: To convey (1) the total amount of political donations made in the last three General Elections in Ontario (2) to compare total donations for each main political party.
- Design Principles:
  - Substantive: Labels are appropriately descriptive. Source is cited. Simple format and relative sizes of colours in stacked bars communicate comparison clearly.
  - Perceptual: Changes in the sizes of the stacked bars clearly communicate three trends:
    - (1) General decline in total donations.
    - (2) Donations to the Progressive Conservative Party (PC) are roughly total the donations to the Liberal and New Democrat (NDP) parties combined every election year.
    - (3) NDP donations outpace the Liberals in 2018 and continue in 2022 which demonstrate the Liberals' defeat in 2018 was not only from voters, but donors as well.
    - Colour choice matches party branding in order to reduce cognitive load; most users should be able to read the graph without referring to the legend (still provided for confirmation).
  - Aesthetic: Stata theme creates a professional look for a rudimentary chart. However, the clashing colours degrade the appearance. There is a tension between taste and perception – both important dataviz elements (Healy, 2019). I chose to prioritize perception and clear understanding.
- Reproducibility: Source link is provided in code. User is given the option to use merged/prepped csv provided in Git repo or to replicate it via instructions in code.
- Accessibility:
  - Visual:
    - Alt-text has been coded with R using a level 2/3 description (Zogheib, 2023).
    - An accessible typeface that will read better on screen is employed (Bringhurst, 2001; Nielsen 2012; Zogheib, 2023).
    - The image would have looked more cohesive if the colours had shared the same values (light-dark); however, I decided to prioritize accessibility by making the colours legible in black and white (see *party\_col\_grayscale.jpg*) (Zogheib, 2023). This information should be accessible to all voters.
  - Language: While Arabic numerals and dollar signs should be understandable to non-English speakers – most adults should know it's a

plot about time x money, the labels are not accessible. I considered using party logos in place of text labels but (1) this would make an already aesthetically-challenged graph unbearably ugly with no guarantee it would increase comprehension. With more resources, I would translate labels for large groups of non-English speakers in Ontario such as Mandarin and Punjabi (Government of Canada, 2022) and languages that use different language characters.

- Tech: png format widely used and easily rendered.
- Potential impact: The graph hopefully prompts viewers to ask questions about our democracy rather than provide answers. Although we cannot draw any causal relationships with funding and election results, voters may have assumed our main parties (or at least the PCs and Liberals) have been operating on a “level playing field” and these graphs illustrate this is not the case. Discrepancies in funding might also prompt us to think more about the influence donations have on our democracy and to research what regulations may exist. As well, we may consider how significant drop in funding in 2022 might be related to Ontario's lowest voter turnout in its history (Rodrigues, 2022).
- Dataset choices: This is a simple dataset with only six variables to work with:
  - Contributor
  - Recipient
  - Recipient.Type (candidate or party)
  - Electoral.District
  - Party.Abbreviation or Party.Name
  - Event or Year
  - Funding dropped significantly in intermediary so I excluded them for simplicity's sake; a more thorough examination should include them. I also excluded Third Party Advertisers due to time restrictions. With more time, these may be included and could change results/trends.
- Underwater labour: Significant public (unionized?) administrative labour is required to process financial donations and records from across an entire province during a busy election year and to ensure the data is clean. Designers and coders are needed to make the data public, usable and accessible. Each party is responsible for reporting its own data and this data must be audited (Elections Ontario, 2023).

Example Two:



- Software: Visual Studio Code (Python/Jupyter)
- Intended Audience: social justice organizers via private screenshare, email, etc.
- Message: To identify the top ten political donors and donees across the last three election years.
- Design Principles:
  - Substantive:
    - Labels are appropriately descriptive.
    - Source is cited.
    - Simple format communicates a comparison clearly. There is a serious issue with how the values of the x-axis is rendered in the bottom graph though; user is unable to determine accurate amounts (I was unable to fix).
  - Perceptual:
    - User requires domain knowledge in order to see two trends: (1) Unions are major political donors; (2) Liberals do not command nearly as many donations as a *party* but through individual candidates/ridings.
  - Aesthetic:
    - I would have preferred to apply the ggtheme “stata” to make all plots similar.
    - The lack of clashing colours make these plots more aesthetically pleasing.
    - With more time, I would have written a function to remove the intense ALL CAPS. There are also random stray characters at the bottom right of the plot; I still need to figure out where they’re coming from.
- Reproducibility: see example one
- Accessibility:
  - Visual:
    - Accessible typeface that will read better on screen is employed (Brighurst, 2001; Nielsen 2012; Zogheib, 2023).
    - High contrast elements increase accessibility (Zogheib, 2023).
    - I was unable to find code to add alt-text in Python :( In real life I would import into R and add it there.
  - Language: Not accessible to non-English speakers. Perhaps with Shiny, one could translate and toggle between languages?
  - Tech: png format widely used and easily rendered.
- Potential impact: These plots are perhaps too detailed for the average voter but as an organizer, they are fascinating. Social justice organizers may have a bias toward focusing on the (negative) influence of corporate donations as we tend to “notice things already primed in memory or repeated often” (Benson & Manoogian, 2008) but the top graph helps to dispel an oversimplification of political donors. The bottom

graph may also help inform organizers' strategic plans. Knowing that the Liberals tend to command donations through candidates and ridings as opposed to the party itself show where the party is weak – as well, this may possibly have been a factor in the Liberal's stunning defeat in 2018 (if candidates become unpopular, there is a less robust donor relationship to the party itself to provide any buffer?).

- Dataset choices: see example one
- Underwater labour: see example one

## Bibliography

- Benson, B., & Manoogian, J. (2018, June 6). *English: Wikipedia's list of 188 cognitive biases*. Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:Cognitive\\_bias\\_codex\\_en.svg](https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg)
- Bringhurst, R. (2001). *The Elements of Typographic Style* (2.4 ed., pp. 193–194). Hartley & Marks.
- Elections Ontario. (2023). *Filing Guidelines for Political Parties*. Elections Ontario. <https://www.elections.on.ca/en/political-entities-in-ontario/political-parties/filing-guidelines-for-political-parties.html>
- Government of Canada, S. C. (2022, August 17). *Mother tongue by geography, 2021 Census*. Statistics Canada. <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/dv-vd/language-langue/index-en.html>
- Kieran Healy. (2019). *Data visualization : a practical introduction* (pp. 5–14). Princeton University Press.
- Nielsen, J. (2012). *Serif vs. Sans-Serif Fonts for HD Screens*. Nielsen Norman Group. <https://www.nngroup.com/articles/serif-vs-sans-serif-fonts-hd-screens>
- Rodrigues, G. (2022, June 3). *Voter turnout in Ontario lowest in history, early data from Elections Ontario shows* | *Globalnews.ca*. Global News. <https://globalnews.ca/news/8891354/ontario-election-2022-low-voter-turnout/>
- Zogheib, C. (2023). *Data Visualization: Visualization with Purpose: Accessible Data Visualization*. [PowerPoint slides]. Data Science Institute, University of Toronto. [https://github.com/UofT-DSI/07-visualization/blob/main/lessons/DSI\\_DataViz\\_5\\_AccessibleDataVisualization.pdf](https://github.com/UofT-DSI/07-visualization/blob/main/lessons/DSI_DataViz_5_AccessibleDataVisualization.pdf)

## Footnotes

1. The following files have been provided in my DataViz repo:
  - Report: DSI-dataviz-04.pdf
  - Code: DSI-DataViz-04.R; DSI-DataViz-04.ipynb
  - Images: DSI-DataViz-04p6.png; DSI-DataViz-04donor10.png; DSI-DataViz-04recip10.png; party\_col\_grayscale.jpg
  - Data: elec.csv
2. While prepping the election data in R, I noticed a lot of Amount values were forced to become null values. This potentially messed up the data but I didn't have time to try cleaning up with Python. Given that I'm not being marked on data cleansing, I've presented the findings here as if they are not problematic but this should not be used as an accurate reflection of the actual data.

## Appendix: Code

*Please refer to original files in GitHub.*

### Example One:

#Visualize Elections ON political donations data from 2014-2022

#Source: <https://finances.elections.on.ca/en/contributions?fromYear=2014&toYear=2023>

#####

#README

#####

#Please feel free to use the elec.csv file in the repo and skip the "read and prepare data" section below. If you wish to download/prepare the data yourself, please keep reading.

#Methodology note: Because file sizes were so large, I downloaded this data in 9 batches, selecting the following restrictions:

#Party (only chose the three major parties) x Event (General Elections for 2014, 2018, 2022),  $3 \times 3 = 9$

#If your internet is better, you may find it easier to just download datasets from the three years here: <https://finances.elections.on.ca/en/downloads>

#If you choose to do the three bulk downloads, you'll have to write your own code to merge the csv files.

#####

#import libraries

#####

library(ggplot2)

library(ggmap)

library(ggthemes)

library(tidyverse)

library(readxl)

#####

#read and prepare data



```
#####
```

```
#open csv files and merge
```

```
#**** make sure to change the filepaths to work with your machine!****
```

```
ndp2022 <- read.csv(file = 'Work/DSI/DataViz/2022NDP.csv')
```

```
lib2022 <- read.csv(file = 'Work/DSI/DataViz/2022LIB.csv')
```

```
pc2022 <- read.csv(file = 'Work/DSI/DataViz/2022PC.csv')
```

```
ndp2018 <- read.csv(file = 'Work/DSI/DataViz/2018NDP.csv')
```

```
lib2018 <- read.csv(file = 'Work/DSI/DataViz/2018LIB.csv')
```

```
pc2018 <- read.csv(file = 'Work/DSI/DataViz/2018PC.csv')
```

```
ndp2014 <- read.csv(file = 'Work/DSI/DataViz/2014NDP.csv')
```

```
lib2014 <- read.csv(file = 'Work/DSI/DataViz/2014LIB.csv')
```

```
pc2014 <- read.csv(file = 'Work/DSI/DataViz/2014PC.csv')
```

```
elec <- rbind(ndp2022, lib2022, pc2022, ndp2018, lib2018, pc2018, ndp2014, lib2014, pc2014)
```

```
#explore data
```

```
head(elec)
```

```
names(elec)
```

```
dim(elec) #63,852 rows, 15 col
```

```
str(elec) #n.b. the Amount column datatype will be chr/character so we'll need to convert
```

```
#convert data-type
```

```
#source: https://stackoverflow.com/questions/31944103/convert-currency-with-commas-into-numeric
```

```
elec$Amount <- as.numeric(gsub('\\$', '', elec$Amount))
```

```
str(elec) #make sure it worked and there are no more $
```

```
#####
```

```
#create objects for plot
```

```
#n.b. colours correspond to political parties/are legible to full colourblindness
```

```
#see party_col_grayscale.jpg
```

```
#####
```

```
party_colours <- c("#BB4949", "#F89D28", "#143675")
```

```
party_names <- c("Liberal", "New Democrat", "Progressive Conservative")
```

```
#####
```

```
#summarize total giving per year/party and plot
```

```
#####
```

```
#plot basic bargraph
```

```
p <- ggplot(data = elec,
```

```
  mapping = aes(x = Year, fill = Party.Abbreviation)) #use abbreviation to avoid inconsistent data
```

```
p1 <- p + geom_bar(position = "stack")
```

```
#correct years on xaxis and change Amount to $ on yaxis
```

```
p2 <- p1 + scale_x_continuous(breaks = c(2014, 2018, 2022)) +
```

```
  scale_y_continuous(labels=scales::dollar_format())
```

```
#change labels and positioning
```

```
p3 <- p2 + labs(title = "Ontario General Elections:",
```

```
  subtitle = "Political Donations to Major Parties",
```

```
  x = "Year",
```

```
  y = "Donation Totals",
```

```
  fill = "Party",
```

```
caption = "Source: Elections Ontario") +  
theme(plot.title = element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5))
```

```
#add custom colours and names for legend
```

```
p4 <- p3 + scale_fill_manual(values = party_colours, labels = party_names)
```

```
p4
```

```
#add theme for a more professional aesthetic
```

```
p5 <- p4 + theme_stata(base_family = "Verdana")
```

```
p5
```

```
#add alt text
```

```
#source: https://ggplot2.tidyverse.org/reference/get\_alt\_text.html
```

```
p6 <- p5 + labs(alt = paste("Barplot graph showing that political donations to the Progressive Conservative Party in the 2014, 2018 and 2022 general elections were about the same amount as total donations to the Liberal and New Democrat Parties combined. We also see NDP donations outpace LIB donations in 2018 and 2022."))
```

```
))
```

```
get_alt_text(p6) #check the alt text works
```

```
p6
```

Example Two:

```
#Visualize Elections ON political donations data from 2014-2022
```

```
#Source: https://finances.elections.on.ca/en/contributions?fromYear=2014&toYear=2023
```

```
#####
```

```
#README
```

```
#####
```

#Please feel free to use the elec.csv file in the repo and skip the "read and prepare data" section below. If you wish to download/prep the data yourself, please keep reading.

#Methodology note:I merged the data into one file in R (see DSI-DataViz-04.R) to create elec.csv so you will need to start from the Methodology note there.

```
#import libraries
```

```
import os
```

```
import sys
```

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
#retrieve data and review
```

```
original_data = 'C://Users/Erica/Documents/Work/DSI/DataViz/elec.csv' #make sure to edit the path to match to your machine!
```

```
df = pd.read_csv(original_data)
```

```
df.head()
```

```
#Identify top 10 donors
```

```
#sort df by Contributor and Amount
```

```
donor10=df.groupby(['Contributor'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False).head(10)
```

```
#plot data
```

```
fig = plt.figure(figsize = (20, 10))
```

```
plt.barh(donor10['Contributor'], donor10['Amount'], color='olivedrab')
```

```
plt.xlabel('Amount ($)', fontsize=15)
```

```
plt.xticks(fontsize=15)
```

```
plt.yticks(fontsize=15)
```

```

plt.title("Ontario General Elections 2014-2022: Top 10 Donors", fontsize=25, loc='left')
plt.gcf().text(0.8,0.05, "Source: Elections Ontario", fontsize=10)

# seeing which party these top donors gave to
# donor10to= df.groupby(['Contributor','Party.Abbreviation'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)
# top10list = ['THE TORONTO-DOMINION BANK','ROGERS COMMUNICATIONS INC','ONTARIO REAL ESTATE
ASSOC','Gibbons, Jack',
#          'NOVA CHEMICALS CORP','ONTARIO PROFESSIONAL FIRE FIGHTERS ASSOC OPFA','TD SECURITIES INC','GAMMA
DYNACARE MEDICAL LABORATORIES LTD','ONTARIO COUNCIL OF PAINTERS','LOC 46 UNITED ASSOC OF JOURNEYMEN
APPRENTICES PLUMB/PIPE']
# df2 = donor10to[donor10to['Contributor'].isin(top10list)]
# df2.groupby(['Contributor','Party.Abbreviation'])[['Amount']].sum().unstack('Party.Abbreviation')

# df2

#Identify top 10 recipients

#sort df by Contributor and Amount
recip10=df.groupby(['Recipient'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False).head(10)

#plot data
fig2 = plt.figure(figsize = (20, 10))
plt.barh(recip10['Recipient'], recip10['Amount'], color='olivedrab')
plt.xlabel('Amount ($)', fontsize=15)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.title("Ontario General Elections 2014-2022: Top 10 Recipients", fontsize=25, loc='left')
plt.gcf().text(0.8,0.05, "Source: Elections Ontario", fontsize=10)

```