My title*

My subtitle if needed

First author Another author

June 27, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

```
##| include: false
##| warning: false
##| message: false
#install.packages("usethis")
#install.packages("gitcreds")
#library(usethis)
#library(gitcreds)
#create_github_token()
# Example dataframe with values and ranges
data_ranges <- data.frame(
  category = c("A", "B", "C", "D"),
  value = c(100, 150, 200, 250),
  range = c(10, 15, 20, 25)  # Example range values
)

# Calculate upper and lower bounds for plotting
data_ranges$lower <- data_ranges$value - data_ranges$range
data_ranges$upper <- data_ranges$value + data_ranges$range

# Print the dataframe
print("Data with Ranges:")
```

```
[1] "Data with Ranges:"
```

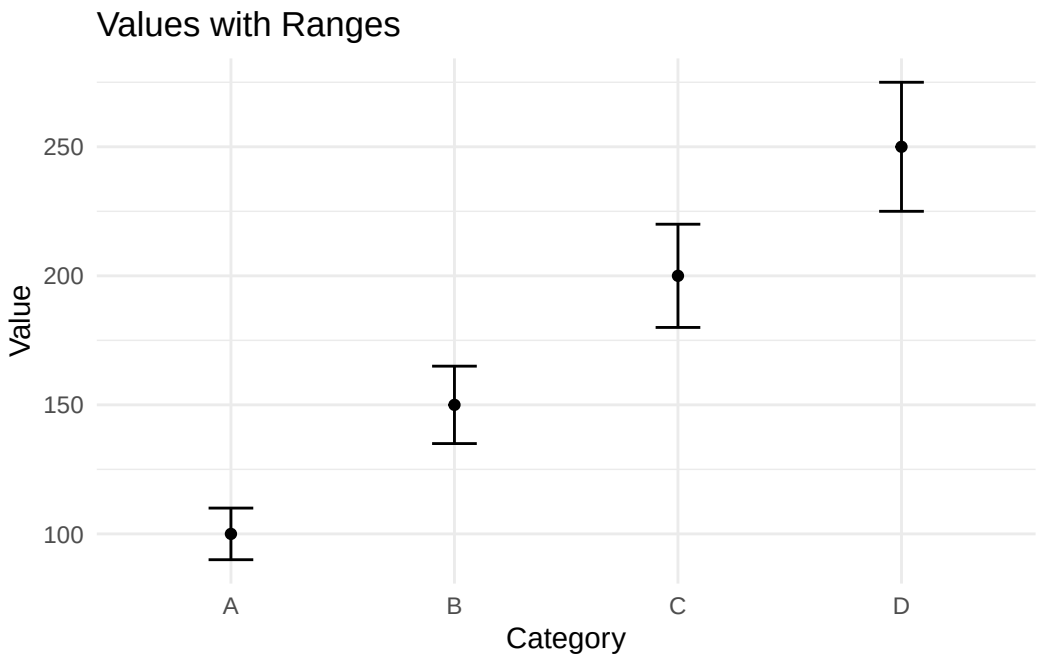
*Code and data are available at: LINK.

```r
print(data_ranges)
```

```
  category value range lower upper
1        A   100    10    90   110
2        B   150    15   135   165
3        C   200    20   180   220
4        D   250    25   225   275
```

```r
# Load necessary library
library(ggplot2)

# Create the plot
ggplot(data_ranges, aes(x = category, y = value)) +
  geom_point() +  # Plot points for values
  geom_errorbar(aes(ymin = lower, ymax = upper), width = 0.2) +  # Error bars for ranges
  labs(title = "Values with Ranges", x = "Category", y = "Value") +
  theme_minimal()
```




```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
```

```
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v lubridate 1.9.3     v tibble    3.2.1
v purrr     1.0.2     v tidyr     1.3.1
-- Conflicts ---------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(palmerpenguins)
```

1 Introduction

In the biggest democratic election the world has ever seen, India held its general election in 2024 over 7 successive phases, beginning on the 19th of April and ending on the 1st of June. The election determined the composition of the Lok Sabha, the lower house of India's Parliament, which comprises 543 seats. A party or coalition needs a minimum of 272 seats to form a government.

During the election period, numerous exit and opinion polls were conducted to gauge public sentiment and predict the potential outcome. Opinion polls were conducted in the months leading up to the election, providing insights into voter preferences and trends. Exit polls were conducted immediately after voters cast their ballots, offering a snapshot of the electorate's choices.

The results of these polls were keenly watched by political analysts, parties, and the public, as they provided an early indication of which party or coalition might secure a majority in the Lok Sabha. Despite their limitations and varying degrees of accuracy, these polls played a crucial role in shaping the narrative and expectations surrounding the 2024 Indian general election.

2 Data

The data used in this paper is obtained by scraping the Wikipedia page for 'general elections in India 2024', and cross-referenced with each article it mentioned. There seems to be no singular database to collect all pre-election and exit poll data, and this is part of the gaping problem examined in this paper. The data points are determined from several National and local newspapers that have published the news online.

While nothing except the results themselves, possible margins, and the sample sizes for the opinion polls are available online, and when one follows links to the actual research/polling

agency hired by the news outlets, there are no publicly available reports available on their websites either.

The three data sets I will be using are Opinion Poll data, Exit Poll data and Mood of the Nation (CVoter).

2.1 Opinion Poll data

In the run-up to the 2024 General Election, various newspapers and agencies conduct pre-election polls to gauge the mindset, mood, and political leaning of the average voter. These polls provide valuable insights into public opinion, allowing political analysts, parties, and the general public to understand potential election outcomes better. They help identify trends, shifts in voter preferences, and the impact of current events and policies on voter behavior.

Containing data from January 2023 up to April 2024, this data consists of the following 7 variables: name of the polling agency, sample size, seats won by NDA, INDIA, and Others. Additionally, the 'Exact' date published is a self-created variable after scouring the references, so a more accurate dataset can be made, and the overlap in a line graph can be avoided. A snapshot of the cleaned data is provided below, which includes the 7 variables and 17 data points in total.

Table 1: A summary atable of cleaned data

Polling agency	Date published	Sample size	NDA	INDIA	Others	Exact
TV9 Bharatvarsh - People's Insight - Polstrat	April 2024	2500000	362	149	32	2024-04-16
ABP News-CVoter	April 2024	57566	373	155	15	2024-04-16
Times Now-ETG	April 2024	271292	384	118	41	2024-04-04
News18	March 2024	118616	411	105	27	2024-03-06
ABP News-CVoter	March 2024	41762	366	156	21	2024-03-16
India TV-CNX	March 2024	162900	378	98	67	2024-03-06
Times Now-ETG	March 2024	323357	378	120	45	2024-03-09
Zee News-Matrize	February 2024	167843	377	93	73	2024-02-28
India Today-CVoter	February 2024	149092	335	166	42	2024-02-08
Times Now-ETG	February 2024	156843	366	104	73	2024-02-08
ABP-CVoter	December 2023	200000	315	185	50	2023-12-25
Times Now-ETG	December 2023	147231	329	158	57	2023-12-13
India TV-CNX	October 2023	54250	315	172	56	2023-10-07
Times Now-ETG	October 2023	135100	307	175	61	2023-10-03
Times Now-ETG	August 2023	110662	311	175	60	2023-08-16
India Today-CVoter	August 2023	160438	306	193	54	2023-08-24
India Today-CVoter	January 2023	140917	298	153	92	2023-02-18

2.2 Mood of the Nation data

Conducted by National Newshouse India Today, the CVoter opinion poll utilizes Computer Assisted Telephone Interviewing among voters aged 18 and above across all 543 Lok Sabha seats. Targeting a sample of 30,000 per quarter with an average response rate of 55%, the tracker is conducted weekly in 11 national languages. According to the official site, CVoter employs a Standard Resilient Distributed dataset to randomly select numbers across telecom lines, ensuring a representative analysis through statistical weighting of the data to reflect the local population as per the latest Census figures. The data is weighted to align with known census profiles, including gender, age, education, income, religion, caste, urban/rural distribution, and past vote recall for the Lok Sabha and Vidhan Sabha elections. For analytics, CVoter uses a proprietary algorithm to calculate provincial and regional vote shares based on the split-voter phenomenon.

The CVoter data offers the most consistent output for opinion polls in India and has been selected due to its standardization and regularity. This consistency allows for the assumption that all other variables are controlled, enabling a focused analysis of trends. The dataset has been independently constructed by aggregating infographics and articles published on the India Today website. The included dataset only has 5 variables, the month, and NDA, INDIA and Others Seats. A snapshot is provided below.

Table 2: A summary table of cleaned data

Month	NDA	INDIA	Others
Jan-20	303	132	108
Aug-20	316	134	93
Jan-21	321	129	93
Aug-21	298	140	105
Jan-22	296	120	127
Aug-22	307	111	125
Jan-23	298	92	153
Aug-23	306	44	193
Jan-24	335	166	42

2.3 Exit poll data

The inception of exit polls in India traces back to 1957, during the second Lok Sabha elections, with the pioneering initiative led by the Indian Institute of Public Opinion.

Exit polls, as the name suggests, are conducted after the culmination of the last phase of voting, in accordance with the guidelines set by the Election Commission of India. Distinguished from opinion polls, which precede elections, exit polls involve surveys administered to voters as they exit polling stations post casting their ballots. These surveys encompass a spectrum of inquiries, delving into voter motivations and party preferences.. As the 2024 Lok Sabha elections concluded, exit poll results for 28 states and eight union territories were slated

for release from 6:30 pm onwards on June 1 2024, in compliance with Election Commission protocols.

In contrast to opinion polls, exit polls are often conducted and released hastily on the same day, rendering the timing irrelevant. Moreover, these exit polls have not disclosed the sample size of the respondents surveyed. As a result, they are presented here merely as indicators, highlighting their potential for inaccuracy and inflation. There are 4 variables, the name of the agency, and the number of seats predicted each for NDA, INDIA and others. The 15 datapoints have been scraped from Wikipedia using rvest.

Table 3: A summary table of cleaned data

agency	NDA	INDIA	Others
ABP News-CVoter	368	167	8
Agni News Services	242	264	37
Dainik Bhaskar	316	173	41
DB Live	221	275	38
India Today-Axis My India	381	148	14
India News-Dynamics	371	125	47
India TV-CNX	386	134	33
NDTV-Jan Ki Baat	377	151	15
News18-CNBC	362	132	47
News 24-Today's Chanakya	400	107	36
News Nation	360	161	22
Republic TV-Matrize	360	126	30
Republic TV-PMarq	359	154	30
Times Now-ETG	358	152	33
TV9 Bharatvarsh - People's Insight - Polstrat	346	162	35

3 References