

# Codebook: parties

This data set is a cleaned version of the data from Clark and Golder's (2006) paper "Rehabilitating Duverger's Theory" [link]. I obtained their replication data from the file `Legislative_new.tab` [link] posted to Matt Golder's Dataverse [link].

The following versions are available:

```
## [1] "parties.csv" "parties.csvy" "parties.dta" "parties.rds"
## [5] "parties.xlsx"
```

The data set is at the **election level**, so that each row of the data set represents one election.

## Load Data

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

# load data
parties_df <- import("data/parties.rds")

# quick look at data
glimpse(parties_df)
```

```
## Observations: 588
## Variables: 7
## $ country      <chr> "Albania", "Albania", "Albania", "Argentina"...
## $ year         <dbl> 1992, 1996, 1997, 1946, 1951, 1954, 1958, 19...
## $ average_magnitude <dbl> 1.00, 1.00, 1.00, 10.53, 10.53, 4.56, 8.13, ...
## $ eneg         <dbl> 1.106929, 1.106929, 1.106929, 1.342102, 1.34...
## $ enep         <dbl> 2.190, 2.785, 2.870, 5.750, 1.970, 1.930, 2...
## $ enlp         <dbl> 1.95, 1.31, 2.07, 3.74, 1.21, 1.26, 1.72, 2...
## $ electoral_system <fctr> Single-Member District, Single-Member Distr...
```

## Variable Descriptions

### country: Country Name

- Coding: The name of the country that held the election.
- Type: character

```
# sort and print country names
sort(unique(parties_df$country))
```

```
## [1] "Albania"          "Argentina"
## [3] "Armenia"          "Australia"
## [5] "Austria"          "Bangladesh"
## [7] "Belgium"          "Benin"
## [9] "Brazil"           "Bulgaria"
## [11] "Canada"           "Central African Republic"
```

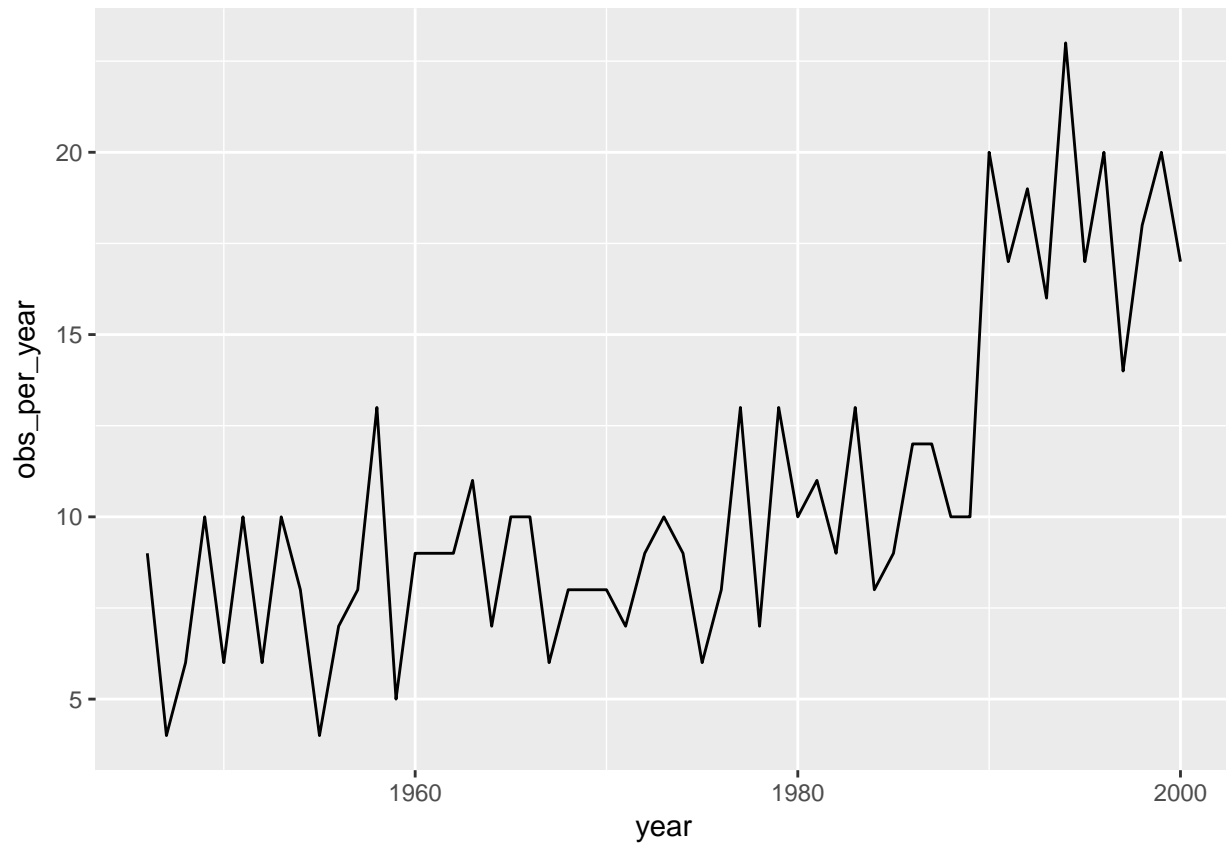
```
## [13] "Chile" "Colombia"
## [15] "Costa Rica" "Croatia"
## [17] "Cyprus" "Czech Republic"
## [19] "Czechoslovakia" "Denmark"
## [21] "Dominica" "Ecuador"
## [23] "El Salvador" "Estonia"
## [25] "Finland" "France"
## [27] "Germany" "Ghana"
## [29] "Greece" "Guatemala"
## [31] "Guyana" "Honduras"
## [33] "Hungary" "India"
## [35] "Indonesia" "Ireland"
## [37] "Israel" "Italy"
## [39] "Jamaica" "Japan"
## [41] "Korea, South (Rep.)" "Latvia"
## [43] "Lithuania" "Macedonia"
## [45] "Malawi" "Mali"
## [47] "Mexico" "Mongolia"
## [49] "Namibia" "Nepal"
## [51] "Netherlands" "New Zealand"
## [53] "Nicaragua" "Niger"
## [55] "Nigeria" "Norway"
## [57] "Pakistan" "Panama"
## [59] "Peru" "Philippines"
## [61] "Poland" "Portugal"
## [63] "Romania" "Sierra Leone"
## [65] "Slovak Republic" "Slovenia"
## [67] "Somalia" "South Africa"
## [69] "Spain" "Sri Lanka"
## [71] "Sweden" "Switzerland"
## [73] "Taiwan" "Thailand"
## [75] "Trinidad and Tobago" "Turkey"
## [77] "Uganda" "United Kingdom"
## [79] "United States" "Venezuela"
## [81] "Zambia"
```

**year:** Year

- Coding: The year of the election.
- Type: integer

```
# calculate observations per year
## group by year
grouped_df <- group_by(parties_df, year)
## calculate obs_per_year
sum_df <- summarize(grouped_df, obs_per_year = n())

# plot observations per year
ggplot(sum_df, aes(x = year, y = obs_per_year)) +
  geom_line()
```

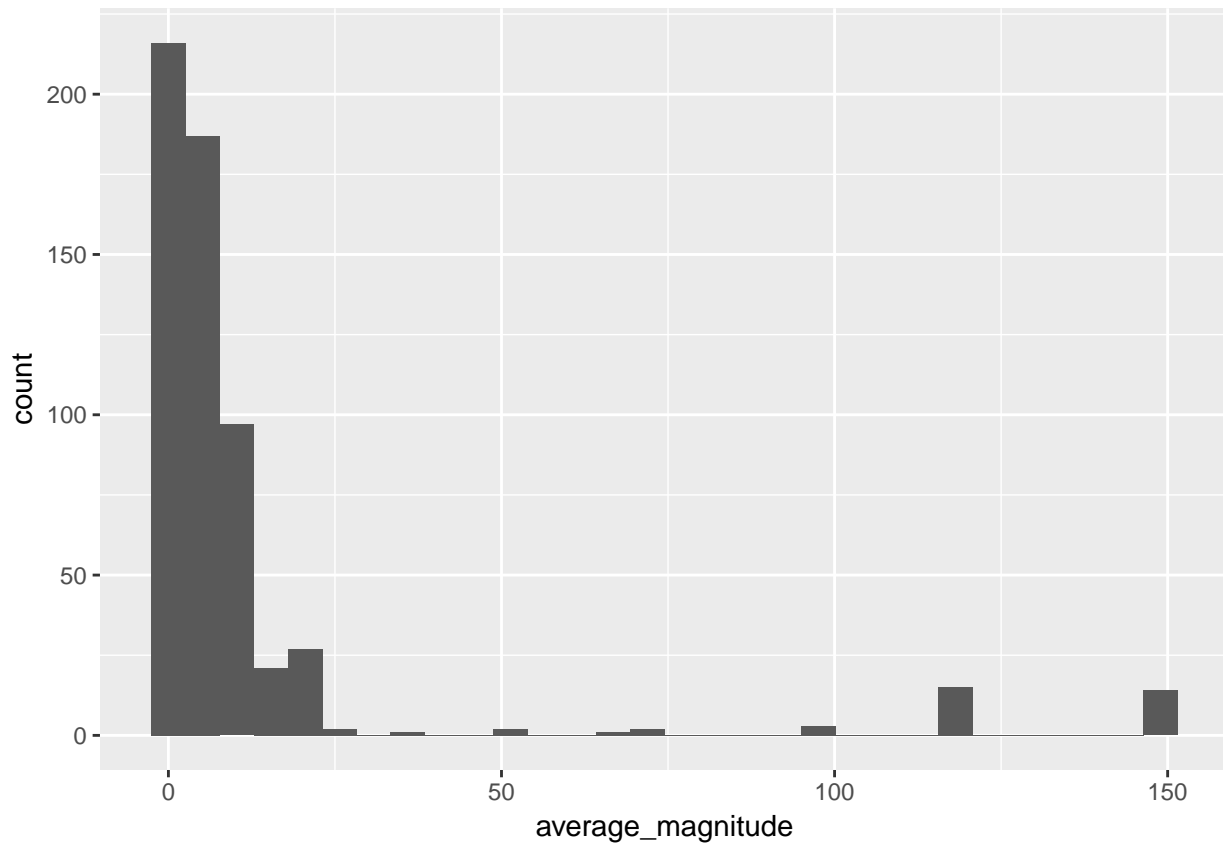


### average\_magnitude: Average District Magnitude

- Coding: The average (or mean) of the district magnitude (the number of seats available in the district) across all the districts in the country. For the U.S. House of Representatives, this would be one, because we have single-member districts (i.e., magnitude of one). In Israel, the average magnitude is 120, because they have a single national district with 120 members.
- Type: double

```
# histogram
ggplot(parties_df, aes(x = average_magnitude)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

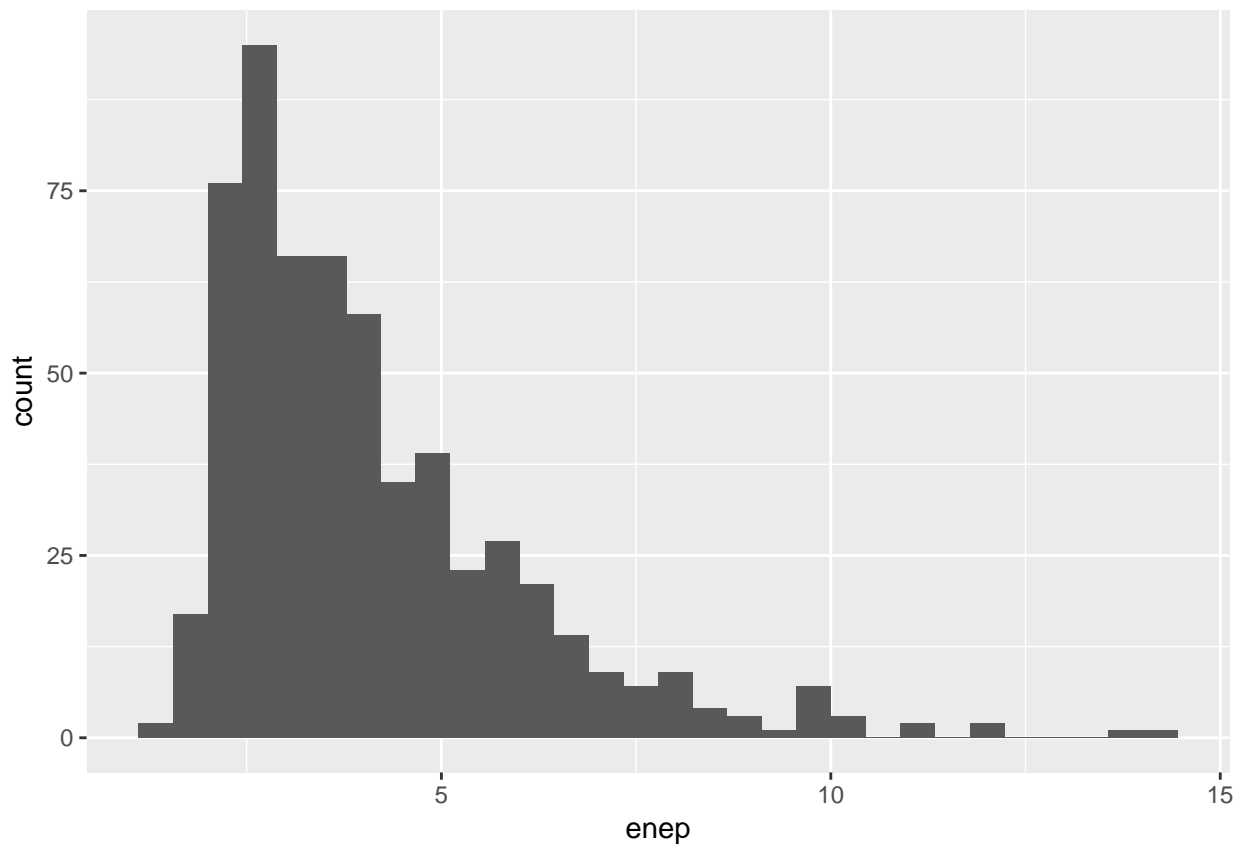


### enep: The Effective Number of Electoral Parties

- Coding: Calculated as  $ENEP_j = \frac{1}{\sum_{i=1}^n v_{ij}^2}$ , where  $ENEP_j$  represents the effective number of electoral parties in election  $j$  and  $v_{ij}$  represents the **vote share** (as a proportion) for party  $i$  in election  $j$ . For the details of this measure, see Clark and Golder (2006) or Clark, Golder, and Golder (ctk), chapter ctk.
- Type: double

```
# histogram
ggplot(parties_df, aes(x = enep)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

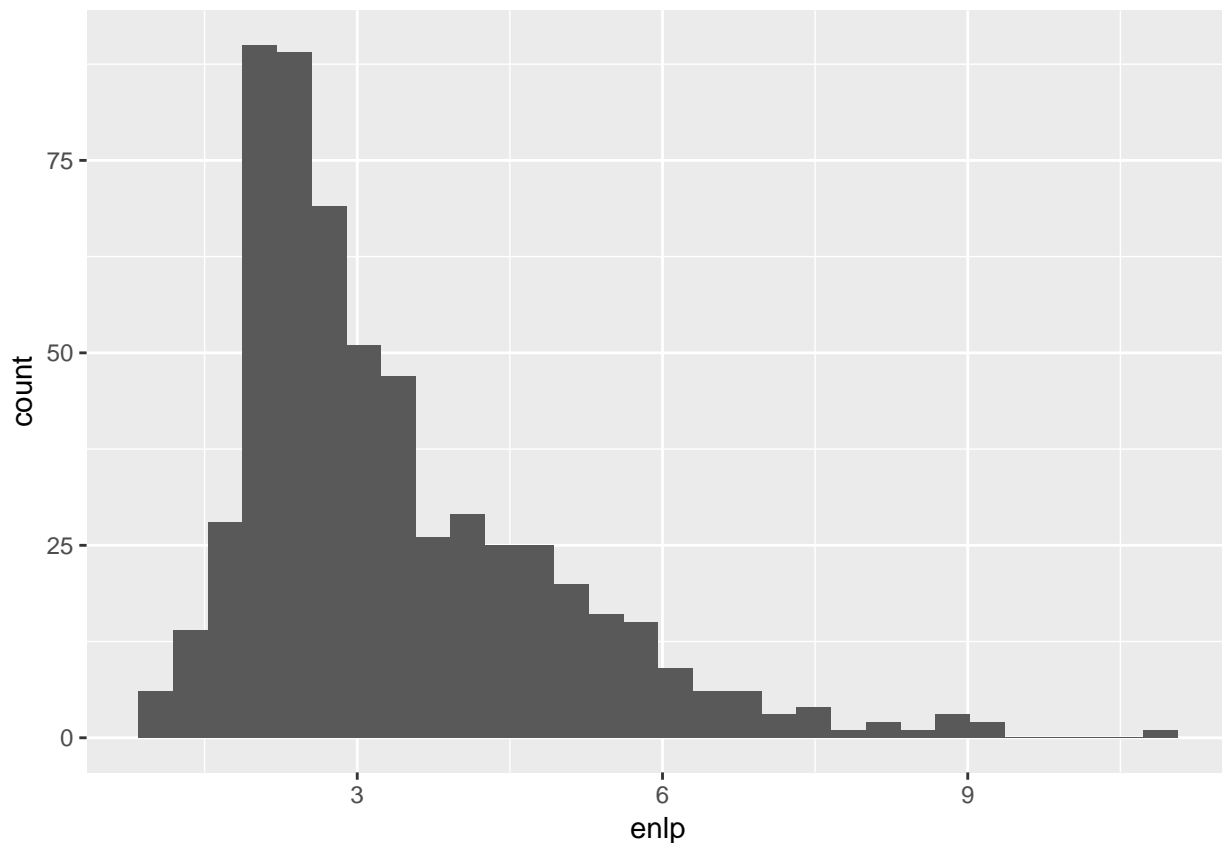


### enlp: The Effective Number of Legislative Parties

- Coding: Calculated as  $ENLP_j = \frac{1}{\sum_{i=1}^n s_{ij}^2}$ , where  $ENLP_j$  represents the effective number of legislative parties in election  $j$  and  $v_{ij}$  represents the **seat share** (as a proportion) for party  $i$  in election  $j$ . For the details of this measure, see Clark and Golder (2006) or Clark, Golder, and Golder (ctk), chapter ctk.
- Type: double

```
# histogram
ggplot(parties_df, aes(x = enlp)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

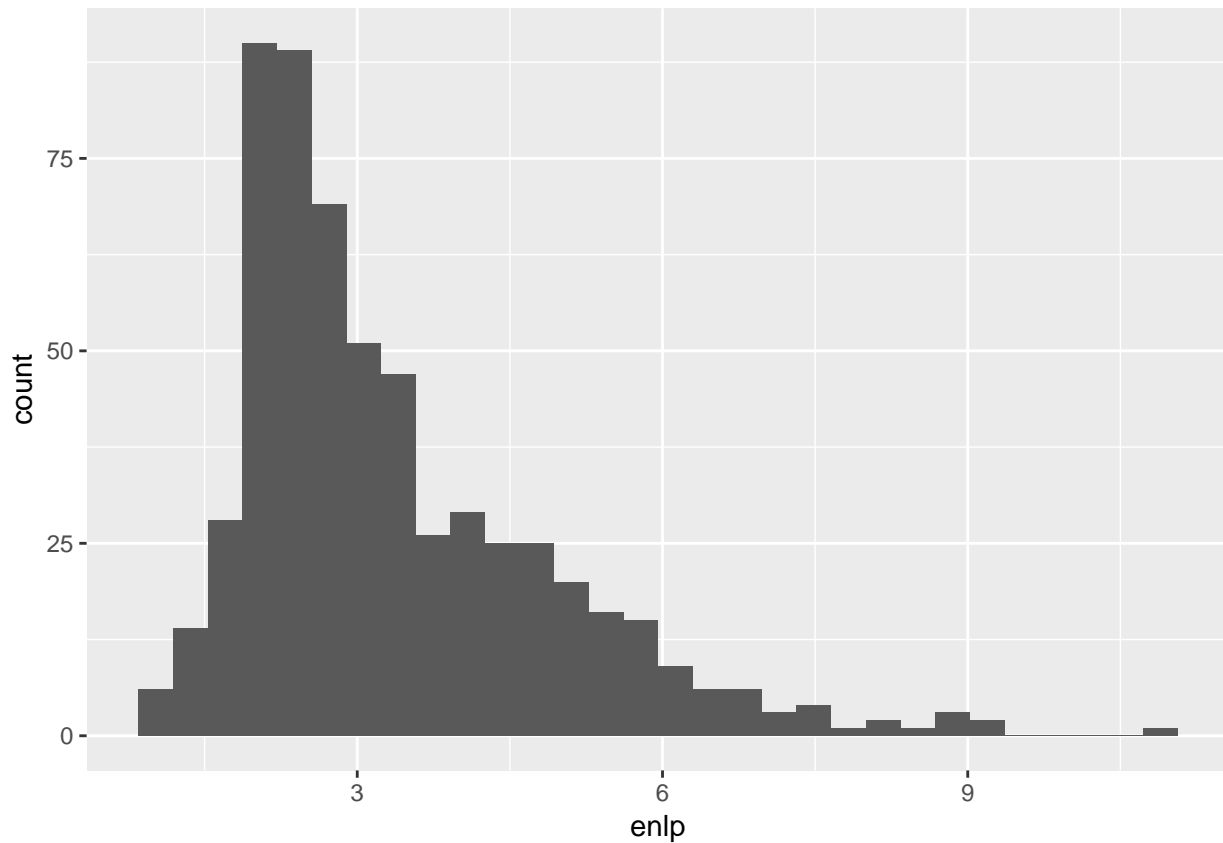


### eneg: The Effective Number of Ethnic Groups

- Coding: Calculated as  $ENEG_j = \frac{1}{\sum_{i=1}^n p_{ij}^2}$ , where  $ENEG_j$  represents the effective number of ethnic groups in the country when election  $j$  occurred and  $p_{ij}$  represents the proportion of the population falling into ethnic group  $i$  when election  $j$  occurred. For the details of this measure, see Clark and Golder (2006) or Clark, Golder, and Golder (ctk), chapter ctk.
- Type: double

```
# histogram
ggplot(parties_df, aes(x = enlp)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



### **electoral\_system: The Type of Electoral System**

- Source: This variable is created from `average_magnitude`.
- Coding:
  - "Single-Member District": when `average_magnitude = 1`.
  - "Small-Magnitude PR": when  $1 < \text{average\_magnitude} \leq 7$ .
  - "Large-Magnitude PR": when `average_magnitude > 7`.
- Type: factor

```
# bar plot  
ggplot(parties_df, aes(x = electoral_system)) +  
  geom_bar()
```

