Codebook: `progresa`

This data set is a cleaned version of the data from De La O's (2013) paper "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico" [link]. I obatined her replication data from the file `DeLaO_AJPS2013_rep_file.dta` posted to her research page [link].

The following versions are available:

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

The data set is at the **region level**, so that each row represents a particular administrative region.

Load Data

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

# load data
progresa_df <- import("data/progresa.rds")

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

```
## Observations: 463
## Variables: 2
## $ condition <chr> "Early Implementation", "Late Implementation", "Earl...
## $ turnout   <dbl> 0.7185629, 0.6583072, 0.6122449, 0.5701358, 0.475609...
```

Variable Descriptions

`condition`: Condition

- Coding: The the exerimental condition of each region.
- Type: factor

```
# sort and print country names
levels(progresa_df$condition)
```

```
## NULL
```

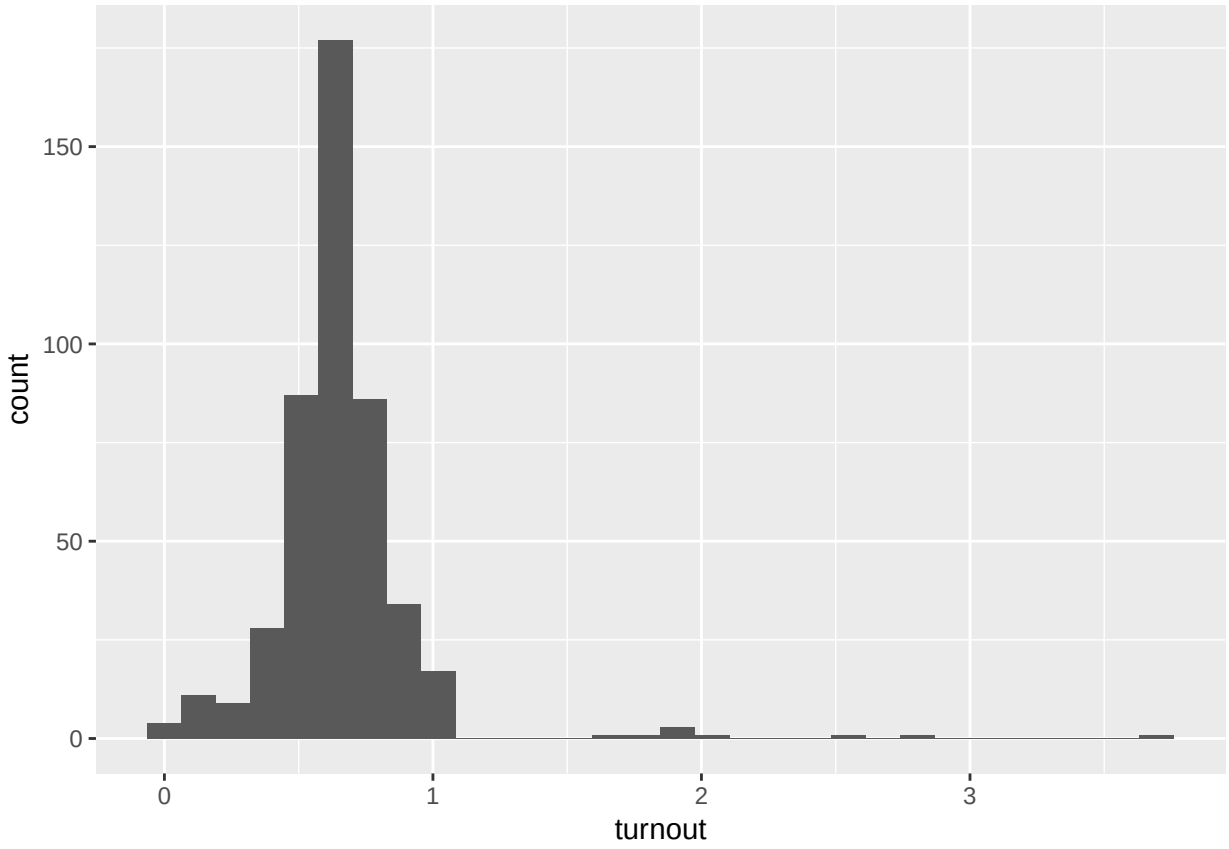
`turnout`: Year

- Coding: The voter turnout in the region, as a proportion.
- Type: double

```
# histogram
ggplot(progresa_df, aes(x = turnout)) +
  geom_histogram()
```

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

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

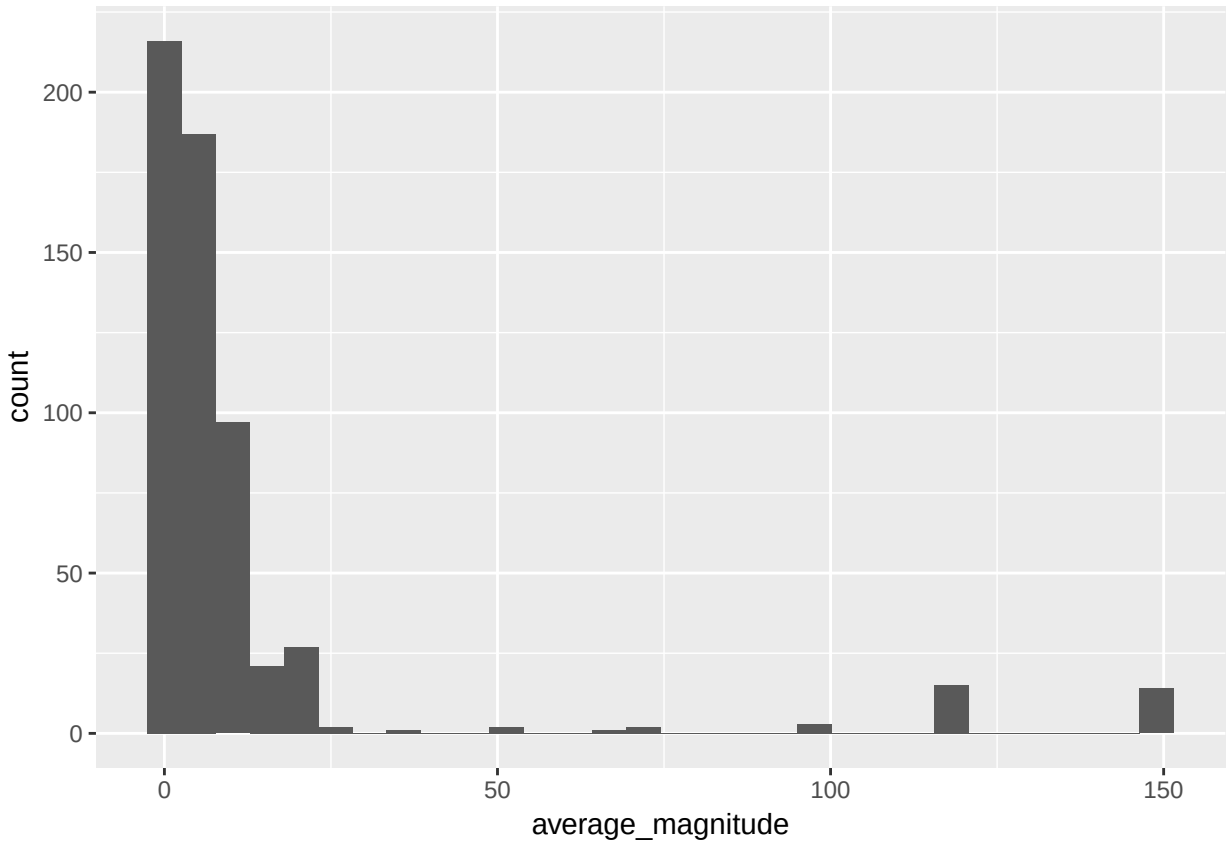


`average_magnitude`: Average District Magnitude

- 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`.
```

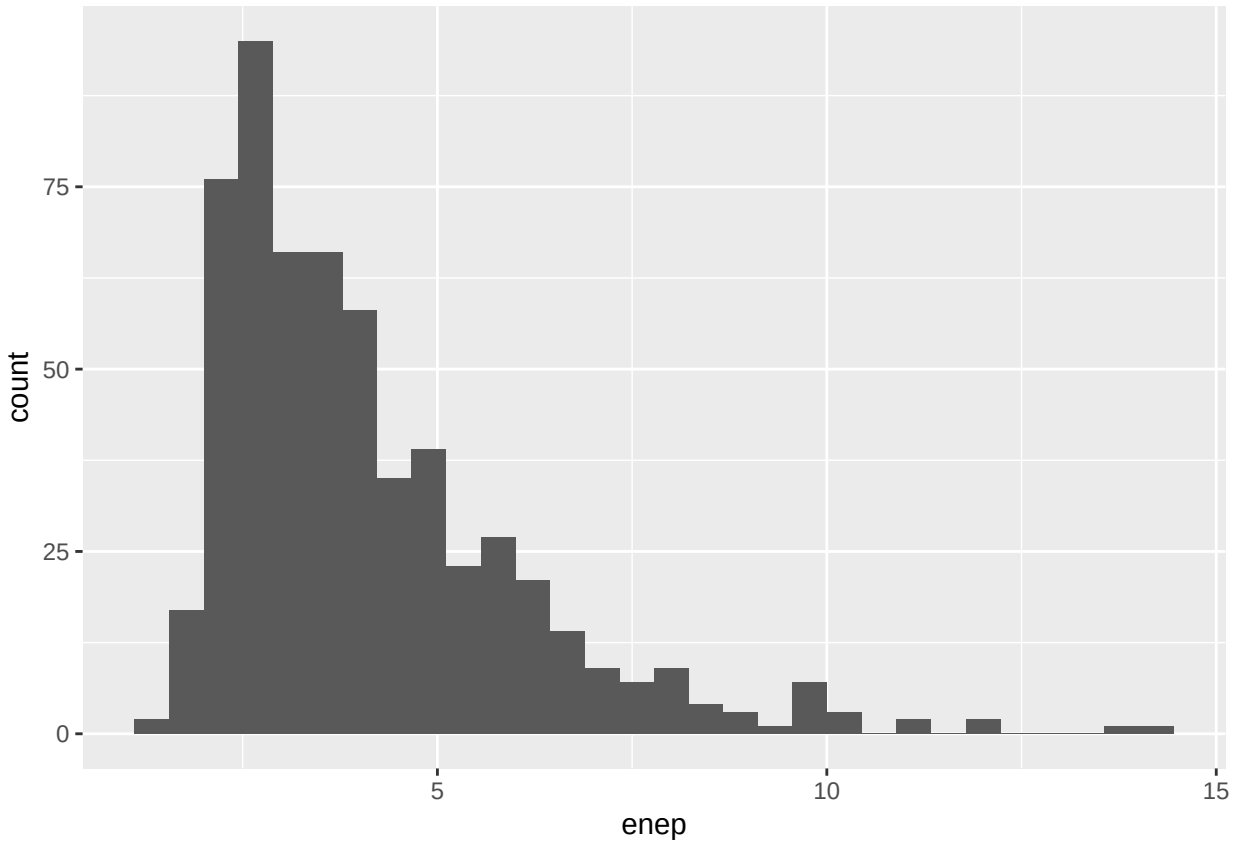


`enep`: The Effective Number of Electoral Parties

- 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`.
```

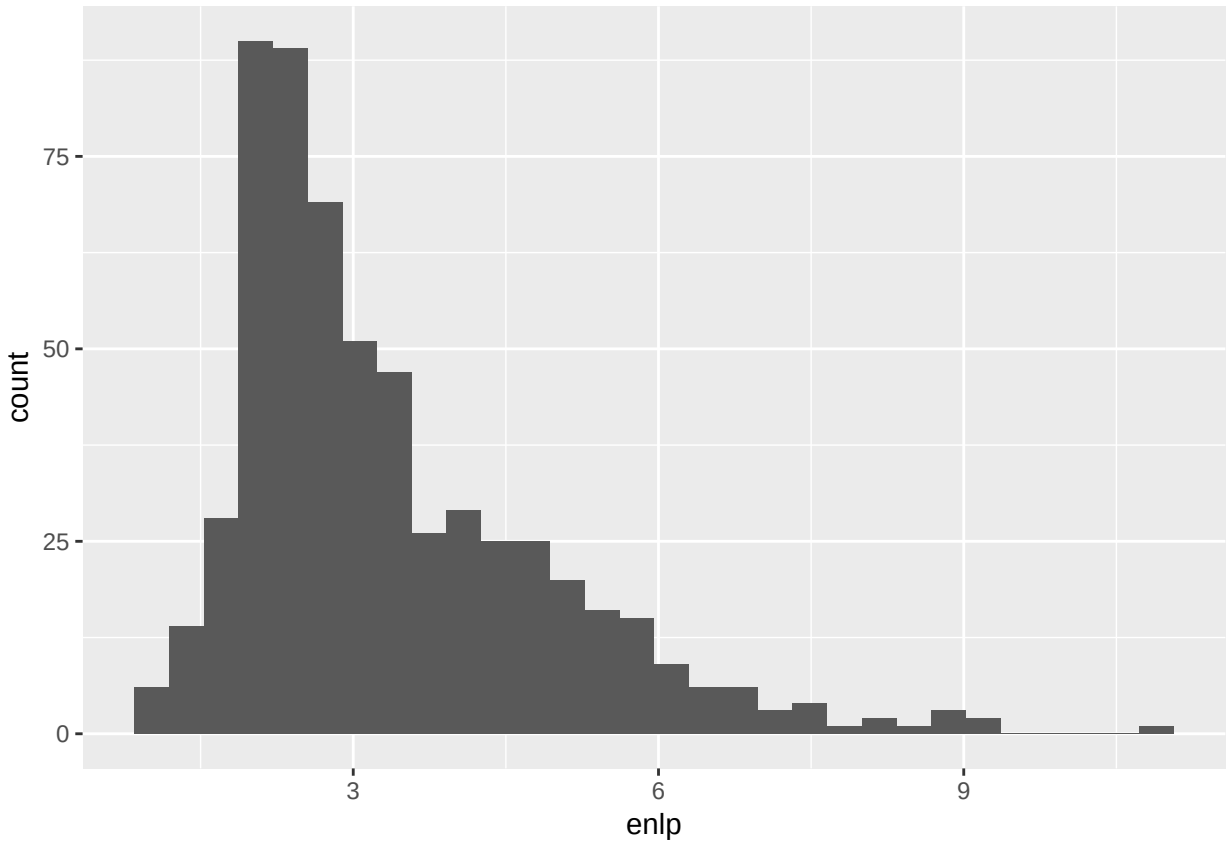


`enlp`: The Effective Number of Legislative Parties

- 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`.
```

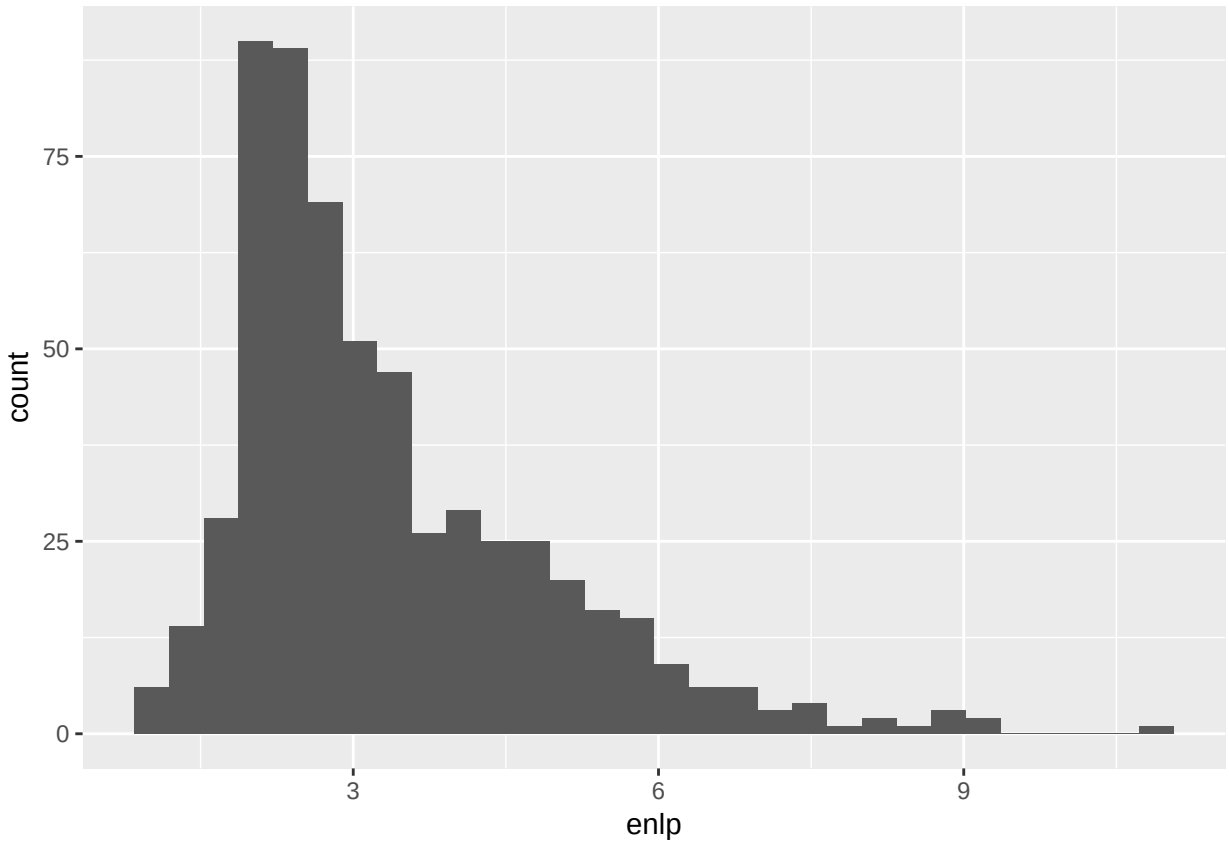


eneg: The Effective Number of Ethnic Groups

- 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()
```

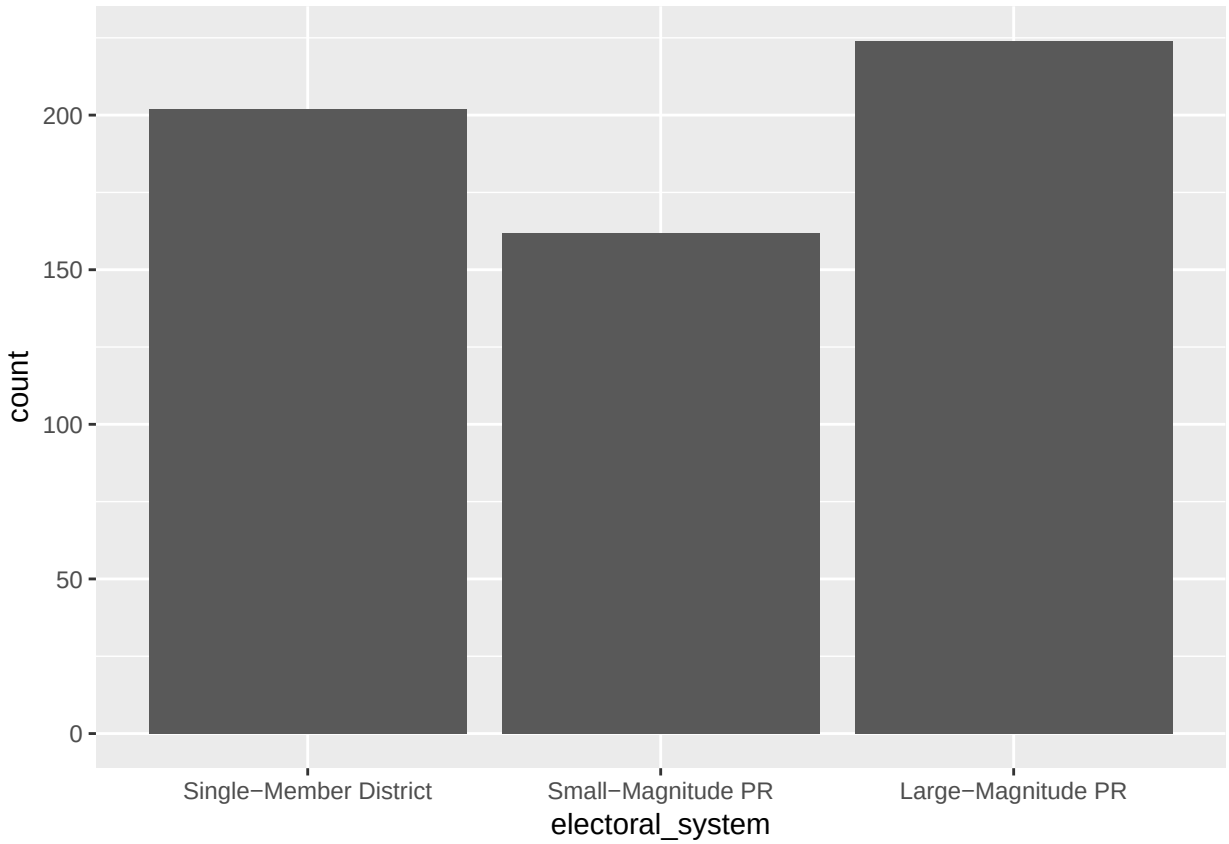
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



`electoral_system`: The Type of Electoral System

- Source: This varaible is created from `average_magnitude`.
- Coding:
  - `"Single-Member Distict"`: when `average_magnitude` $= 1$.
  - `"Small-Magnitude PR"`: when $1 <$ `average_magnitude` ≤ 7.
  - `"Large-Magnitude PR"`: when `average_magnitude` > 7.
- Type: factor

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


count
200
150
100
50
0
Single−Member District
Small−Magnitude PR
Large−Magnitude PR
electoral_system