

# Team proposal

**Anastasiia Saenko. Antara Bhide. Daniela Jiménez Lara**

## **Dataset 1. National Survey of Incarcerated Population in Mexico**

### **General Information**

- Source: INEGI (2021). *Encuesta Nacional de Personas Privadas de a Libertad*. [Data set]. INEGI. <https://www.inegi.org.mx/programas/enpol/2021/>.

The National Survey of Incarcerated Population in Mexico (ENPOL) is a survey implemented by the Mexican National Institute of Statistics and Geography during 2021. Its general objective is to provide information on the experience of the criminal procedure and internment of the incarcerated population aged 18 and over. The general themes covered by the survey are the following:

- Sociodemographic characteristics
- Background
- Judicial procedure
- Penitentiary center.
- Exit expectations.

The geographical breakdown of the survey encompasses 32 federal states and 54 penitentiary centers of interest.

The sample size (observations): **67,584** people aged 18 and over:

- 55,535 men
- 12,049 women

### **Research question 1:**

**How does the type of authority making the arrest and their use of violence influence an inmate's perception for the reason for their incarceration?**

### Predictor variables

We are exploring how the authority that detained the inmate (categorical) and their use of force during arrest and/or violence and threats impact the inmates' perception of why they are in prison. Other control variables that will be used are crime committed, age of inmate, school grade completed by inmate, among others.

### Outcome variables:

Inmate's perception for the reason of incarceration (categorical)

### Research question 2:

**How does the existence of extortion from authorities to inmates during the arrest influence an inmate's sentence?**

### Predictor variables:

We are exploring how the existence of an extortion towards the inmate, from the authorities that committed the arrest, affects the eventual sentence of prison time. Other control variables that will be included are the type of defence (private, public), inmates' income before arrest, age, school grade completed, among others.

### Outcome variables:

Years in prison sentence (continuous)

### Here is a glimpse of the data

Rows: 61,449

Columns: 697

```
$ ...1      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 1~
$ ID_PER    <chr> "2173.8228", "1791.0884", "0188.2951", "1510.0944", "1784.91~
$ CVE_ENT   <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01", "01", ~
$ NOM_ENT   <chr> "Aguascalientes", "Aguascalientes", "Aguascalientes", "Aguas~
$ CEN_INT   <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01", "01", ~
$ NOM_INT   <chr> "CERESO AGUASCALIENTES", "CERESO AGUASCALIENTES", "CERESO AG~
$ COD_RES   <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01", "01", ~
$ SEXO      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ FUERO     <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
$ P1_1      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
```

\$ P1\_1A <dbl> 4, 4, 2, 2, 5, 1, 1, 2, 5, 1, 1, 4, 5, 2, 1, 1, 1, 5, 5, 4, ~  
 \$ P1\_2 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_3 <dbl> 41, 31, 45, 27, 40, 33, 30, 69, 53, 29, 34, 33, 32, 27, 30, ~  
 \$ P1\_4 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_5 <chr> "01", "01", "01", "01", "01", "07", "09", "32", "01", "01", ~  
 \$ P1\_6 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_7 <dbl> 7, 2, 4, 1, 6, 5, 5, 7, 4, 3, 4, 1, 4, 1, 1, 4, 1, 3, 5, 3, ~  
 \$ P1\_8 <dbl> 2, 1, 1, 1, 1, 2, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, ~  
 \$ P1\_9 <chr> NA, "03", "02", "02", "02", NA, "02", NA, NA, "03", "02", "0~  
 \$ P1\_10\_1 <chr> NA, "13", "12", "03", "12", NA, "12", NA, NA, "11", "13", "1~  
 \$ P1\_10\_2 <chr> NA, "10", "10", "10", "10", NA, "10", NA, NA, "10", "09", "1~  
 \$ P1\_10\_3 <chr> NA, "08", NA, NA, NA, NA, NA, NA, NA, "05", NA, "12", "13", ~  
 \$ P1\_10\_4 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "08", "11", NA, ~  
 \$ P1\_10\_5 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "06", "10", NA, ~  
 \$ P1\_10\_6 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "07", NA, NA~  
 \$ P1\_10\_7 <lg1> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_11\_1 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_2 <dbl> NA, 1, 1, 1, 1, NA, 0, NA, NA, 1, 1, 1, 1, 1, 1, 1, 1, NA, N~  
 \$ P1\_11\_3 <dbl> NA, 0, 0, 0, 0, NA, 1, NA, NA, 1, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_4 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_5 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_6 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_7 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_8 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_11\_9 <dbl> NA, 0, 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, NA, N~  
 \$ P1\_12 <dbl> 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ R1\_13X <chr> NA, NA, NA, NA, NA, "0934", NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_14 <dbl> NA, NA, NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
 \$ P1\_15 <dbl> 3, 3, 2, 3, 3, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 2, 3, 3, 3, ~  
 \$ P1\_16 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_17 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_18\_N <chr> "08", "03", "06", "03", "04", "02", "03", "02", "03", "03", ~  
 \$ P1\_18\_G <chr> "05", "03", "02", "02", "03", "06", "03", "03", "03", "03", ~  
 \$ P1\_19 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, ~  
 \$ P1\_20 <chr> "23", "15", "35", "17", "18", "12", "15", "09", "17", "19", ~  
 \$ P1\_21 <chr> NA, "04", "05", "03", "07", "07", "02", "02", "07", "03", "0~  
 \$ P1\_22 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_23 <dbl> 3, 3, 3, 3, 3, 3, 3, 8, 3, 3, 3, 3, 3, 1, 3, 3, 3, 3, 3, ~  
 \$ P1\_24\_1 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_25\_1 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_26\_1 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_27\_1 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_28\_1 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~

[illegible]

\$ P1\_24\_9 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ R1\_24\_9X <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_25\_9 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_26\_9 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_27\_9 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_28\_9 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_29 <dbl> 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 1, 1, 2, 2, 2, 2, 1, 2, ~  
 \$ P1\_30 <dbl> NA, NA, 2, NA, 2, NA, 1, NA, NA, NA, NA, NA, 2, 2, NA, NA, N~  
 \$ P1\_31\_1 <dbl> 2, 1, 2, 2, 1, 2, 2, 2, 2, 1, 1, 2, 2, 2, 2, 1, 1, 2, 2, ~  
 \$ P1\_31\_2 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_31\_3 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, ~  
 \$ P1\_31\_4 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_32 <dbl> 1, 1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, ~  
 \$ P1\_33\_1 <dbl> 2, 1, 2, 1, 1, 2, 1, 2, 1, 2, 2, 1, 1, 2, 2, 2, 1, 1, 1, 2, ~  
 \$ P1\_33\_2 <dbl> 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1, ~  
 \$ P1\_34\_1 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_34\_2 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_35 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_36 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_37 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_38 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_39 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_40 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_1 <dbl> 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_42\_1 <chr> "18", "15", NA, NA, "16", "10", "15", "09", "18", "14", NA, ~  
 \$ P1\_43\_1 <dbl> 1, 1, NA, NA, 1, 1, 1, 2, 1, 2, NA, NA, 2, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_44\_1 <dbl> 1, 1, NA, NA, 1, 1, 1, NA, 1, NA, NA, NA, NA, 1, 1, 1, 1, 1, ~  
 \$ P1\_41\_2 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~  
 \$ P1\_42\_2 <chr> "24", "18", "25", "13", "11", "10", "15", "09", NA, "14", "1~  
 \$ P1\_43\_2 <dbl> 2, 2, 2, 2, 2, 1, 2, 2, NA, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, ~  
 \$ P1\_44\_2 <dbl> NA, NA, NA, NA, NA, 2, NA, NA, NA, NA, 2, NA, NA, 2, NA, NA, ~  
 \$ P1\_41\_3 <dbl> 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 2, 2, 1, 1, 2, 1, 1, 1, 2, ~  
 \$ P1\_42\_3 <chr> NA, "16", NA, "13", "23", "10", "15", "12", NA, "15", "16", ~  
 \$ P1\_43\_3 <dbl> NA, 2, NA, 2, 2, 1, 2, 2, NA, 2, 2, NA, NA, 1, 2, NA, 2, 2, ~  
 \$ P1\_44\_3 <dbl> NA, NA, NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, 2, NA, NA~  
 \$ P1\_41\_4 <dbl> 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2, 2, 1, 1, 2, 1, 2, 2, 2, ~  
 \$ P1\_42\_4 <dbl> NA, 18, NA, 14, NA, 12, NA, NA, NA, 15, NA, NA, NA, 15, 12, ~  
 \$ P1\_43\_4 <dbl> NA, 2, NA, 2, NA, 2, NA, NA, NA, 2, NA, NA, NA, 1, 2, NA, 2, ~  
 \$ P1\_44\_4 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, N~  
 \$ P1\_41\_5 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_42\_5 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_43\_5 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_44\_5 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~

\$ P1\_41\_6 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, ~  
 \$ P1\_42\_6 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 14, ~  
 \$ P1\_43\_6 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, N~  
 \$ P1\_44\_6 <lg1> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_7 <dbl> 2, 1, 2, 2, 1, 2, 1, 2, 2, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, ~  
 \$ P1\_42\_7 <dbl> NA, 18, NA, NA, 20, NA, 15, NA, NA, 15, 18, 17, 18, 15, 12, ~  
 \$ P1\_43\_7 <dbl> NA, 2, NA, NA, 2, NA, 2, NA, NA, 2, 2, 2, 2, 1, 2, NA, 1, NA~  
 \$ P1\_44\_7 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, N~  
 \$ P1\_41\_8 <dbl> 2, 2, 2, 1, 1, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_42\_8 <dbl> NA, NA, NA, 17, 24, NA, 15, NA, NA, NA, NA, NA, NA, 15, NA, ~  
 \$ P1\_43\_8 <dbl> NA, NA, NA, 1, 2, NA, 2, NA, NA, NA, NA, NA, NA, 2, NA, NA, ~  
 \$ P1\_44\_8 <dbl> NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
 \$ P1\_41\_9 <dbl> 2, 1, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, ~  
 \$ P1\_42\_9 <dbl> NA, 16, NA, 26, 24, NA, 15, NA, NA, 17, NA, NA, 22, NA, 16, ~  
 \$ P1\_43\_9 <dbl> NA, 2, NA, 1, 2, NA, 2, NA, NA, 2, NA, NA, 2, NA, 2, NA, 2, ~  
 \$ P1\_44\_9 <dbl> NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
 \$ P1\_41\_10 <dbl> 2, ~  
 \$ P1\_42\_10 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_43\_10 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_44\_10 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_11 <dbl> 2, ~  
 \$ P1\_42\_11 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_43\_11 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_44\_11 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_12 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~  
 \$ P1\_42\_12 <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "15", NA, NA, NA, NA, NA~  
 \$ P1\_43\_12 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, NA, NA, NA, NA, N~  
 \$ P1\_44\_12 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_13 <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 1, 2, 2, 2, 2, 2, ~  
 \$ P1\_42\_13 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 16, NA, NA, 28, NA, 19, ~  
 \$ P1\_43\_13 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, NA, 2, NA, 2, NA,~  
 \$ P1\_44\_13 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_41\_14 <dbl> 2, ~  
 \$ P1\_42\_14 <lg1> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_43\_14 <lg1> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_44\_14 <lg1> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~  
 \$ P1\_45 <dbl> NA, 2, NA, 1, 2, NA, 2, NA, NA, 2, 2, 2, 2, 2, 2, NA, 2, NA,~  
 \$ P1\_46 <dbl> NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
 \$ P1\_47 <dbl> 1, 2, NA, NA, 2, 2, 1, NA, 2, NA, NA, NA, NA, 2, 2, 2, 1, 2,~  
 \$ P1\_48 <dbl> 4, 4, 4, 4, 4, 4, 1, 4, 4, 4, 4, 4, 2, 4, 4, 4, 1, 4, 4, 4, ~  
 \$ EST\_DIS <chr> "3029.1", "3029.1", "3029.1", "3029.1", "3029.1", "3029.1", ~  
 \$ FAC\_PER <dbl> 2.473585, 2.473585, 2.473585, 2.473585, 2.473585, 2.473585, ~  
 \$ FPC <dbl> 0.4380165, 0.4380165, 0.4380165, 0.4380165, 0.4380165, 0.438~

[illegible]

8



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

## **Dataset 2: Life Expectancy data**

### **General Information**

This is a combined data from [World Bank Open Data](#) & [Our World in Data](#). The data explored the projected life expectancy of newborns if the projected mortality patterns at the time of birth were fixed. The dataset is a panel dataset with socio-economic data for years of 2001 to 2019 for 174 countries. It combines several socio-economic variables: basic country indicators such as GPD, unemployment, education and health indicators (disability, health expenditure, sanitation) as well as climate patterns data (emissions) and political indicators (corruption). The data overall has 3306 entries and 16 variables.

### **Research question 1:**

**What's the Impact of Socio-economic factors on life Expectancy?**

#### **Predictor variables**

We are exploring the impact of various socio-economic factors on life expectancy. We will explore relationships between such factors as CO2 emissions, corruption, education and health industries on life expectancy. Some of those variables are also included as control variables.

#### **Outcome variable**

Outcome variable that we are using is life expectancy (in years)

### **Research question 2:**

**How do a country's health expenditure, education expenditure, and income group influence the prevalence of undernourishment among its population ?**

#### **Predictor variables**

Health expenditure, education expenditure and other country-level indicators (gdp, corruptions).

## Outcome variable

Prevalence of Undernourishment (% of the population) ## Here's the glimpse of the data

Rows: 3,306

Columns: 16

```
$ Country.Name      <chr> "Afghanistan", "Angola", "Albania", "An~
$ Country.Code      <chr> "AFG", "AGO", "ALB", "AND", "ARE", "ARG~
$ Region            <chr> "South Asia", "Sub-Saharan Africa", "Eu~
$ IncomeGroup       <chr> "Low income", "Lower middle income", "U~
$ Year              <int> 2001, 2001, 2001, 2001, 2001, 2001, 200~
$ Life.Expectancy.World.Bank <dbl> 56.30800, 47.05900, 74.28800, NA, 74.54~
$ Prevalence.of.Undernourishment <dbl> 47.8, 67.5, 4.9, NA, 2.8, 3.0, 26.1, NA~
$ CO2               <dbl> 730, 15960, 3230, 520, 97200, 125260, 3~
$ Health.Expenditure.. <dbl> NA, 4.483516, 7.139524, 5.865939, 2.484~
$ Education.Expenditure.. <dbl> NA, NA, 3.45870, NA, NA, 4.83374, 2.469~
$ Unemployment       <dbl> 10.809, 4.004, 18.575, NA, 2.493, 17.32~
$ Corruption         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ Sanitation         <dbl> NA, NA, 40.52090, 21.78866, NA, 48.0540~
$ Injuries           <dbl> 2179727.10, 1392080.71, 117081.67, 1697~
$ Communicable       <dbl> 9689193.70, 11190210.53, 140894.78, 695~
$ NonCommunicable    <dbl> 5795426.38, 2663516.34, 532324.75, 1363~
```

## Dataset 3: US School Scores Dataset

Source: National Center for education Statistics - <https://nces.ed.gov/ccd/aboutCCD.asp>

### General Information

This dataset represents the SAT scores of students across the various states of the US from 2005 till 2015 along with their GPAs and score ranges for different subjects. The dataset also provides their financial backgrounds as well as their genders.

This dataset consists of 578 entries and 99 variables.

### Research question 1:

**What is the likeliest SAT score for a student provided a specific financial background.**

### Predictor variables

We are exploring the effect of financial backgrounds, gender, year , and scores in multiple subjects on the outcome variable

### Outcome variable

Outcome variable that we are using is SAT score.

### Research question 2:

What is the likeliest score for math/verbal provided a specific financial background as well as gender

### Predictor variables

GPA, financial background , gender, year , SAT score(and several other subject wise grades)

### Outcome variable

Math/ Verbal score

### Here's the glimpse of the data

Rows: 577

Columns: 99

\$ Year	<int> 2005, 2005, 20~
\$ State.Code	<chr> "AL", "AK", "A~
\$ State.Name	<chr> "Alabama", "Al~
\$ Total.Math	<int> 559, 519, 530,~
\$ Total.Test.takers	<int> 3985, 3996, 18~
\$ Total.Verbal	<int> 567, 523, 526,~
\$ Academic.Subjects.Arts.Music.Average.GPA	<dbl> 3.92, 3.76, 3.~
\$ Academic.Subjects.Arts.Music.Average.Years	<dbl> 2.2, 1.9, 2.1,~
\$ Academic.Subjects.English.Average.GPA	<dbl> 3.53, 3.35, 3.~
\$ Academic.Subjects.English.Average.Years	<dbl> 3.9, 3.9, 3.9,~
\$ Academic.Subjects.Foreign.Languages.Average.GPA	<dbl> 3.54, 3.34, 3.~
\$ Academic.Subjects.Foreign.Languages.Average.Years	<dbl> 2.6, 2.1, 2.6,~
\$ Academic.Subjects.Mathematics.Average.GPA	<dbl> 3.41, 3.06, 3.~
\$ Academic.Subjects.Mathematics.Average.Years	<dbl> 4.0, 3.5, 3.9,~
\$ Academic.Subjects.Natural.Sciences.Average.GPA	<dbl> 3.52, 3.25, 3.~

\$ Academic.Subjects.Natural.Sciences.Average.Years	<dbl> 3.9, 3.2, 3.4,~
\$ Academic.Subjects.Social.Sciences.History.Average.GPA	<dbl> 3.59, 3.39, 3.~
\$ Academic.Subjects.Social.Sciences.History.Average.Years	<dbl> 3.9, 3.4, 3.3,~
\$ Family.Income.Between.20.40k.Math	<int> 513, 492, 498,~
\$ Family.Income.Between.20.40k.Test.takers	<int> 324, 401, 2121~
\$ Family.Income.Between.20.40k.Verbal	<int> 527, 500, 495,~
\$ Family.Income.Between.40.60k.Math	<int> 539, 517, 520,~
\$ Family.Income.Between.40.60k.Test.takers	<int> 442, 539, 2270~
\$ Family.Income.Between.40.60k.Verbal	<int> 551, 522, 518,~
\$ Family.Income.Between.60.80k.Math	<int> 550, 513, 524,~
\$ Family.Income.Between.60.80k.Test.takers	<int> 473, 603, 2372~
\$ Family.Income.Between.60.80k.Verbal	<int> 564, 519, 523,~
\$ Family.Income.Between.80.100k.Math	<int> 566, 528, 534,~
\$ Family.Income.Between.80.100k.Test.takers	<int> 475, 444, 1866~
\$ Family.Income.Between.80.100k.Verbal	<int> 577, 534, 533,~
\$ Family.Income.Less.than.20k.Math	<int> 462, 464, 485,~
\$ Family.Income.Less.than.20k.Test.takers	<int> 175, 191, 891,~
\$ Family.Income.Less.than.20k.Verbal	<int> 474, 467, 474,~
\$ Family.Income.More.than.100k.Math	<int> 588, 541, 554,~
\$ Family.Income.More.than.100k.Test.takers	<int> 980, 540, 3083~
\$ Family.Income.More.than.100k.Verbal	<int> 590, 544, 546,~
\$ GPA.A.minus.Math	<int> 569, 544, 541,~
\$ GPA.A.minus.Test.takers	<int> 724, 673, 3334~
\$ GPA.A.minus.Verbal	<int> 575, 546, 535,~
\$ GPA.A.plus.Math	<int> 622, 600, 605,~
\$ GPA.A.plus.Test.takers	<int> 563, 173, 1684~
\$ GPA.A.plus.Verbal	<int> 623, 604, 593,~
\$ GPA.A.Math	<int> 600, 580, 571,~
\$ GPA.A.Test.takers	<int> 1032, 671, 385~
\$ GPA.A.Verbal	<int> 608, 578, 563,~
\$ GPA.B.Math	<int> 514, 492, 498,~
\$ GPA.B.Test.takers	<int> 1253, 1622, 71~
\$ GPA.B.Verbal	<int> 525, 499, 499,~
\$ GPA.C.Math	<int> 436, 466, 458,~
\$ GPA.C.Test.takers	<int> 188, 418, 1184~
\$ GPA.C.Verbal	<int> 451, 472, 464,~
\$ GPA.D.or.lower.Math	<int> 0, 424, 439, 0~
\$ GPA.D.or.lower.Test.takers	<int> 0, 12, 16, 0, ~
\$ GPA.D.or.lower.Verbal	<int> 0, 466, 435, 0~
\$ GPA.No.response.Math	<int> 0, 0, 0, 0, 0,~
\$ GPA.No.response.Test.takers	<int> 225, 427, 919,~
\$ GPA.No.response.Verbal	<int> 0, 0, 0, 0, 0,~
\$ Gender.Female.Math	<int> 538, 505, 513,~

\$ Gender.Female.Test.takers	<int> 2072, 2161, 98~
\$ Gender.Female.Verbal	<int> 561, 521, 522,~
\$ Gender.Male.Math	<int> 582, 535, 549,~
\$ Gender.Male.Test.takers	<int> 1913, 1835, 83~
\$ Gender.Male.Verbal	<int> 574, 526, 531,~
\$ Score.Ranges.Between.200.to.300.Math.Females	<int> 22, 30, 119, 1~
\$ Score.Ranges.Between.200.to.300.Math.Males	<int> 10, 20, 72, 7,~
\$ Score.Ranges.Between.200.to.300.Math.Total	<int> 32, 50, 191, 1~
\$ Score.Ranges.Between.200.to.300.Verbal.Females	<int> 14, 26, 115, 9~
\$ Score.Ranges.Between.200.to.300.Verbal.Males	<int> 17, 26, 86, 3,~
\$ Score.Ranges.Between.200.to.300.Verbal.Total	<int> 31, 52, 201, 1~
\$ Score.Ranges.Between.300.to.400.Math.Females	<int> 173, 233, 881,~
\$ Score.Ranges.Between.300.to.400.Math.Males	<int> 93, 153, 450, ~
\$ Score.Ranges.Between.300.to.400.Math.Total	<int> 266, 386, 1331~
\$ Score.Ranges.Between.300.to.400.Verbal.Females	<int> 123, 218, 739,~
\$ Score.Ranges.Between.300.to.400.Verbal.Males	<int> 84, 171, 613, ~
\$ Score.Ranges.Between.300.to.400.Verbal.Total	<int> 207, 389, 1352~
\$ Score.Ranges.Between.400.to.500.Math.Females	<int> 514, 696, 3215~
\$ Score.Ranges.Between.400.to.500.Math.Males	<int> 293, 485, 1948~
\$ Score.Ranges.Between.400.to.500.Math.Total	<int> 807, 1181, 516~
\$ Score.Ranges.Between.400.to.500.Verbal.Females	<int> 430, 656, 3048~
\$ Score.Ranges.Between.400.to.500.Verbal.Males	<int> 332, 552, 2398~
\$ Score.Ranges.Between.400.to.500.Verbal.Total	<int> 762, 1208, 544~
\$ Score.Ranges.Between.500.to.600.Math.Females	<int> 722, 813, 3576~
\$ Score.Ranges.Between.500.to.600.Math.Males	<int> 614, 616, 3152~
\$ Score.Ranges.Between.500.to.600.Math.Total	<int> 1336, 1429, 67~
\$ Score.Ranges.Between.500.to.600.Verbal.Females	<int> 690, 729, 3661~
\$ Score.Ranges.Between.500.to.600.Verbal.Males	<int> 617, 596, 3101~
\$ Score.Ranges.Between.500.to.600.Verbal.Total	<int> 1307, 1325, 67~
\$ Score.Ranges.Between.600.to.700.Math.Females	<int> 485, 342, 1688~
\$ Score.Ranges.Between.600.to.700.Math.Males	<int> 611, 445, 2126~
\$ Score.Ranges.Between.600.to.700.Math.Total	<int> 1096, 787, 381~
\$ Score.Ranges.Between.600.to.700.Verbal.Females	<int> 596, 423, 1831~
\$ Score.Ranges.Between.600.to.700.Verbal.Males	<int> 613, 375, 1679~
\$ Score.Ranges.Between.600.to.700.Verbal.Total	<int> 1209, 798, 351~
\$ Score.Ranges.Between.700.to.800.Math.Females	<int> 156, 47, 327, ~
\$ Score.Ranges.Between.700.to.800.Math.Males	<int> 292, 116, 630,~
\$ Score.Ranges.Between.700.to.800.Math.Total	<int> 448, 163, 957,~
\$ Score.Ranges.Between.700.to.800.Verbal.Females	<int> 219, 109, 412,~
\$ Score.Ranges.Between.700.to.800.Verbal.Males	<int> 250, 115, 501,~
\$ Score.Ranges.Between.700.to.800.Verbal.Total	<int> 469, 224, 913,~