Data preparation

Socioeconomic and Gender Disparities: A Multi-Country Study

Andreas Laffert, Research asistant

2025-02-05

1 Presentation

This is the data preparation code for the project "Socioeconomic and Gender Disparities: A Multi-Country Study." The prepared dataset is `SOGEDI_dataset_V1.sav`

In this repository, data processing and cleaning exclude countries with insufficient sample sizes for robust statistical analysis, retaining only observations from Argentina, Chile, Colombia, Spain, and Mexico. However, for anyone wishing to use all cases and countries from the original dataset, it can be accessed at the following link.

2 Libraries

First, we load the necessary libraries. In this case, we use `pacman::p_load` to load and call libraries in one move.

```
if (! require("pacman")) install.packages("pacman")

pacman::p_load(tidyverse,
               sjmisc,
               here,
               sjlabelled,
               haven,
               naniar,
               car,
```

```
               kableExtra)

options(scipen=999)
rm(list = ls())
```

3 Data

We load the database from the the Github repository project.

```
sogedi_db <- haven::read_sav(url("https://github.com/sogedi-project/sogedi-data/raw/re

glimpse(sogedi_db)
```

```
Rows: 4,386
Columns: 283
$ ID                          <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,~
$ StartDate                   <dttm> 2024-04-28 11:11:20, 2024-04-28 11:12:34,~
$ EndDate                     <dttm> 2024-04-28 11:30:12, 2024-04-28 11:31:15,~
$ IPAddress                   <chr> "90.167.243.1", "83.58.124.179", "79.152.1~
$ Duration__in_seconds        <dbl> 1132, 1120, 1192, 1410, 1328, 645, 933, 88~
$ RecordedDate                <dttm> 2024-04-28 11:30:12, 2024-04-28 11:31:16,~
$ ResponseId                  <chr> "R_1eqka09S3bZXYTp", "R_42oDc55cfSucfrX", ~
$ LocationLatitude            <chr> "41.6362", "41.3891", "41.4287", "41.5453"~
$ LocationLongitude           <chr> "-4.7435", "2.1606", "2.2164", "2.4414", "~
$ aten_check_1                <dbl+lbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
$ aten_check_2                <dbl+lbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
$ aten_check_3                <dbl+lbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
$ time_class_1_First_Click    <dbl> 2.695, 12.941, 11.788, 13.477, 228.254, 8.~
$ time_class_1_Last_Click     <dbl> 55.801, 60.736, 38.827, 50.411, 249.534, 2~
$ time_class_1_Page_Submit    <dbl> 57.095, 62.226, 40.360, 51.672, 250.460, 2~
$ time_class_1_Click_Count    <dbl> 20, 6, 6, 6, 12, 9, 6, 6, 6, 7, 11, 16, 6,~
$ eco_in_1                    <dbl+lbl> 6, 6, 7, 6, 6, 4, 4, 3, 6, 3, 7, 7, 5,~
$ eco_in_2                    <dbl+lbl> 6, 6, 7, 6, 6, 4, 5, 4, 3, 4, 1, 6, 5,~
$ eco_in_3                    <dbl+lbl> 7, 6, 7, 6, 6, 4, 2, 3, 5, 3, 5, 4, 6,~
$ jus_ine                     <dbl+lbl> 1, 2, 1, 1, 2, 5, 1, 1, 2, 1, 2, 5, 3,~
$ co_eco                      <dbl+lbl> 7, 7, 6, 4, 5, 3, 6, 6, 3, 2, 1, 4, 5,~
$ time_class_2_First_Click    <dbl> 4.004, 11.257, 7.633, 9.522, 6.466, 9.334,~
$ time_class_2_Last_Click     <dbl> 88.112, 83.882, 68.033, 82.459, 61.386, 32~
```

```
$ time_class_2_Page_Submit    <dbl> 89.238, 85.347, 69.004, 83.534, 62.577, 33~
$ time_class_2_Click_Count    <dbl> 34, 12, 13, 16, 20, 18, 17, 13, 17, 12, 14~
$ pp_pw_1                     <dbl+lbl> 7, 4, 6, 2, 5, 5, 3, 3, 2, 5, 7, 5, 5,~
$ pp_pw_2                     <dbl+lbl> 7, 5, 6, 3, 6, 5, 5, 7, 2, 5, 2, 5, 5,~
$ pp_pw_3                     <dbl+lbl> 7, 6, 7, 2, 5, 3, 3, 5, 2, 4, 2, 4, 5,~
$ pp_pw_4                     <dbl+lbl> 7, 4, 4, 1, 5, 5, 3, 5, 2, 5, 4, 3, 5,~
$ cc_pw_1                     <dbl+lbl> 5, 4, 6, 3, 6, 4, 5, 6, 5, 4, 4, 6, 6,~
$ cc_pw_2                     <dbl+lbl> 4, 2, 4, 2, 5, 4, 4, 4, 2, 4, 2, 4, 4,~
$ cc_pw_3                     <dbl+lbl> 4, 3, 6, 4, 6, 4, 4, 4, 2, 5, 7, 5, 6,~
$ cc_pw_4                     <dbl+lbl> 3, 5, 5, 3, 6, 4, 5, 5, 4, 4, 7, 6, 6,~
$ hc_pw_1                     <dbl+lbl> 1, 1, 1, 1, 1, 4, 2, 2, 1, 3, 1, 2, 4,~
$ hc_pw_2                     <dbl+lbl> 2, 1, 2, 3, 3, 4, 2, 3, 1, 6, 1, 2, 5,~
$ hc_pw_3                     <dbl+lbl> 1, 1, 4, 2, 2, 4, 1, 2, 1, 3, 1, 2, 2,~
$ hc_pw_4                     <dbl+lbl> 2, 2, 2, 1, 2, 3, 4, 2, 1, 4, 1, 2, 5,~
$ time_class_3_First_Click    <dbl> 1.784, 8.404, 5.983, 5.768, 68.795, 4.382,~
$ time_class_3_Last_Click     <dbl> 53.633, 61.879, 60.551, 189.575, 121.331, ~
$ time_class_3_Page_Submit    <dbl> 54.156, 63.590, 61.758, 190.098, 122.122, ~
$ time_class_3_Click_Count    <dbl> 30, 12, 14, 13, 14, 19, 13, 12, 13, 13, 13~
$ pp_pm_1                     <dbl+lbl> 6, 5, 6, 4, 5, 5, 3, 6, 2, 5, 7, 5, 6,~
$ pp_pm_2                     <dbl+lbl> 7, 5, 7, 2, 6, 3, 2, 6, 2, 5, 5, 5, 4,~
$ pp_pm_3                     <dbl+lbl> 7, 6, 6, 3, 5, 3, 3, 6, 2, 5, 7, 3, 5,~
$ pp_pm_4                     <dbl+lbl> 7, 4, 6, 3, 5, 3, 3, 5, 2, 4, 7, 4, 6,~
$ cc_pm_1                     <dbl+lbl> 7, 4, 4, 3, 5, 4, 5, 3, 5, 3, 2, 5, 5,~
$ cc_pm_2                     <dbl+lbl> 4, 2, 1, 1, 4, 4, 3, 4, 2, 2, 2, 4, 4,~
$ cc_pm_3                     <dbl+lbl> 4, 3, 2, 3, 4, 4, 4, 4, 2, 3, 1, 4, 6,~
$ cc_pm_4                     <dbl+lbl> 3, 5, 5, 2, 4, 5, 4, 4, 2, 3, 2, 6, 3,~
$ hc_pm_1                     <dbl+lbl> 3, 1, 4, 3, 3, 3, 2, 4, 4, 4, 7, 3, 5,~
$ hc_pm_2                     <dbl+lbl> 3, 1, 5, 3, 3, 3, 2, 3, 2, 4, 7, 3, 5,~
$ hc_pm_3                     <dbl+lbl> 2, 1, 3, 1, 2, 4, 2, 3, 2, 5, 7, 3, 6,~
$ hc_pm_4                     <dbl+lbl> 3, 2, 4, 2, 5, 4, 2, 4, 1, 5, 7, 3, 5,~
$ time_gender_1_First_Click   <dbl> 1.829, 12.693, 10.808, 18.682, 8.264, 6.60~
$ time_gender_1_Last_Click    <dbl> 126.481, 112.822, 147.886, 102.824, 82.228~
$ time_gender_1_Page_Submit   <dbl> 127.601, 114.335, 149.828, 104.069, 83.877~
$ time_gender_1_Click_Count   <dbl> 49, 20, 28, 23, 20, 31, 25, 24, 20, 23, 27~
$ gen_in_1                    <dbl+lbl> 6, 7, 6, 7, 7, 3, 7, 7, 6, 5, 4, 6, 7,~
$ gen_in_2                    <dbl+lbl> 6, 7, 6, 5, 7, 3, 5, 6, 1, 6, 7, 7, 7,~
$ gen_in_3                    <dbl+lbl> 5, 7, 5, 7, 4, 3, 4, 7, 6, 6, 7, 5, 6,~
$ gen_in_4                    <dbl+lbl> 3, 6, 5, 6, 6, 3, 5, 5, 5, 6, 7, 5, 3,~
$ gen_in_5                    <dbl+lbl> 4, 6, 3, 5, 7, 3, 7, 4, 6, 5, 6, 5, 3,~
$ gen_in_6                    <dbl+lbl> 6, 7, 5, 6, 4, 2, 5, 7, 6, 6, 7, 5, 7,~
$ ps_m_1                      <dbl+lbl> 7, 2, 4, 1, 3, 3, 3, 4, 1, 4, 1, 7, 6,~
```

```
$ ps_m_2                      <dbl+lbl> 6, 1, 2, 5, 1, 4, 1, 4, 1, 1, 1, 5, 4,~
$ ps_m_3                      <dbl+lbl> 6, 2, 4, 3, 4, 2, 4, 4, 1, 4, 7, 3, 6,~
$ hs_m_1                      <dbl+lbl> 1, 1, 2, 1, 2, 3, 2, 2, 1, 3, 1, 2, 4,~
$ hs_m_2                      <dbl+lbl> 1, 1, 5, 1, 3, 3, 1, 2, 1, 2, 1, 2, 5,~
$ hs_m_3                      <dbl+lbl> 1, 2, 1, 1, 2, 4, 1, 2, 1, 3, 1, 3, 5,~
$ shif_1                      <dbl+lbl> 1, 1, 2, 2, 2, 6, 1, 1, 1, 2, 1, 5, 5,~
$ shif_2                      <dbl+lbl> 1, 1, 2, 1, 2, 5, 1, 1, 1, 2, 1, 4, 2,~
$ shif_3                      <dbl+lbl> 1, 1, 1, 4, 2, 3, 1, 1, 1, 2, 3, 5, 3,~
$ femi                        <dbl+lbl> 7, 7, 3, 5, 5, 1, 7, 5, 6, 2, 4, 2, 1,~
$ co_gen                      <dbl+lbl> 7, 7, 3, 4, 5, 3, 6, 5, 2, 2, 1, 4, 4,~
$ jus_gen                     <dbl+lbl> 1, 2, 1, 2, 3, 3, 3, 1, 1, 1, 1, 5, 3,~
$ gen_compe                   <dbl+lbl> 4, 6, 5, 5, 4, 4, 1, 4, 4, 4, 1, 5, 5,~
$ time_contac_1_First_Click   <dbl> 1.842, 12.194, 9.584, 4.779, 10.964, 8.097~
$ time_contac_1_Last_Click    <dbl> 138.959, 125.608, 145.143, 147.327, 288.91~
$ time_contac_1_Page_Submit   <dbl> 139.507, 126.906, 146.760, 148.154, 289.52~
$ time_contac_1_Click_Count   <dbl> 59, 22, 26, 29, 36, 24, 39, 24, 28, 26, 27~
$ ge_ra_wo                    <dbl> 70, 70, 60, 60, 40, 20, 50, 20, 27, 60, 85~
$ ge_ra_me                    <dbl> 30, 30, 40, 40, 60, 80, 50, 80, 73, 40, 15~
$ quan_pw                     <dbl+lbl> 1, 4, 5, 3, 5, 3, 3, 2, 1, 2, 1, 3, 2,~
$ quan_pm                     <dbl+lbl> 1, 4, 5, 3, 5, 4, 3, 3, 1, 2, 1, 3, 2,~
$ quan_rw                     <dbl+lbl> 1, 5, 5, 4, 7, 3, 2, 2, 7, 1, 5, 3, 4,~
$ quan_rm                     <dbl+lbl> 1, 5, 5, 4, 7, 4, 2, 2, 7, 1, 5, 2, 4,~
$ fri_pw                      <dbl+lbl> 1, 1, 2, 3, 3, 4, 2, 1, 1, 3, 1, 2, 1,~
$ fri_pm                      <dbl+lbl> 1, 1, 1, 2, 3, 4, 2, 1, 1, 3, 1, 1, 1,~
$ fri_rw                      <dbl+lbl> 2, 4, 6, 4, 6, 4, 1, 1, 5, 1, 6, 4, 1,~
$ fri_rm                      <dbl+lbl> 2, 5, 6, 3, 6, 4, 1, 1, 5, 1, 7, 4, 1,~
$ qual_pw                     <dbl+lbl> 4, 5, 4, 4, 6, 4, 3, 3, 2, 4, 4, 3, 2,~
$ qual_pm                     <dbl+lbl> 4, 5, 3, 4, 4, 4, 3, 3, 2, 4, 4, 3, 2,~
$ qual_rw                     <dbl+lbl> 2, 5, 6, 3, 5, 4, 3, 4, 4, 4, 7, 4, 3,~
$ qual_rm                     <dbl+lbl> 2, 5, 5, 3, 5, 4, 3, 4, 4, 4, 7, 4, 3,~
$ mobi_up_1                   <dbl+lbl> 4, 3, 3, 5, 2, 3, 1, 3, 1, 4, 5, 5, 6,~
$ mobi_up_2                   <dbl+lbl> 4, 4, 5, 3, 3, 4, 1, 3, 1, 2, 4, 5, 5,~
$ mobi_up_3                   <dbl+lbl> 5, 3, 1, 6, 2, 4, 1, 4, 1, 3, 3, 5, 5,~
$ mobi_down_1                 <dbl+lbl> 5, 6, 6, 6, 5, 4, 5, 5, 6, 4, 5, 3, 2,~
$ mobi_down_2                 <dbl+lbl> 5, 4, 5, 2, 4, 3, 4, 4, 5, 4, 1, 3, 2,~
$ mobi_down_3                 <dbl+lbl> 4, 5, 3, 3, 5, 4, 3, 4, 6, 4, 1, 3, 2,~
$ time_stere_pw_1_First_Click <dbl> 1.813, 14.065, 21.732, 17.635, 11.730, 5.6~
$ time_stere_pw_1_Last_Click  <dbl> 164.968, 172.405, 183.196, 163.738, 148.81~
$ time_stere_pw_1_Page_Submit <dbl> 165.908, 174.555, 185.603, 165.040, 149.59~
$ time_stere_pw_1_Click_Count <dbl> 106, 41, 55, 44, 48, 49, 53, 38, 45, 37, 4~
$ condi_gender                <dbl+lbl> 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1,~
```

```
$ condi_class        <dbl+lbl> 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0,~
$ mor_1              <dbl+lbl> 1, 4, 3, 6, 3, 4, 2, 3, 3, 5, 5, 2, 5,~
$ mor_2              <dbl+lbl> 2, 3, 4, 5, 2, 5, 3, 4, 4, 4, 5, 3, 4,~
$ mor_3              <dbl+lbl> 2, 3, 3, 4, 3, 4, 2, 3, 4, 3, 6, 3, 2,~
$ inm_1              <dbl+lbl> 7, 5, 6, 3, 6, 4, 2, 3, 3, 4, 1, 7, 4,~
$ inm_2              <dbl+lbl> 6, 4, 4, 2, 3, 3, 2, 3, 5, 2, 1, 6, 3,~
$ inm_3              <dbl+lbl> 5, 5, 4, 1, 6, 5, 2, 4, 4, 4, 2, 5, 5,~
$ war_1              <dbl+lbl> 4, 4, 2, 4, 5, 4, 5, 4, 5, 5, 5, 3, 5,~
$ war_2              <dbl+lbl> 2, 3, 4, 5, 4, 3, 3, 4, 4, 4, 5, 3, 4,~
$ war_3              <dbl+lbl> 4, 4, 2, 6, 5, 3, 5, 5, 4, 5, 5, 3, 4,~
$ com_1              <dbl+lbl> 7, 6, 4, 5, 5, 5, 3, 4, 4, 3, 6, 5, 3,~
$ com_2              <dbl+lbl> 6, 6, 5, 5, 5, 5, 3, 5, 5, 4, 6, 5, 2,~
$ com_3              <dbl+lbl> 5, 5, 3, 5, 5, 6, 3, 4, 5, 5, 6, 4, 3,~
$ ph_1               <dbl+lbl> 4, 1, 2, 1, 6, 2, 1, 1, 5, 1, 1, 6, 3,~
$ ph_2               <dbl+lbl> 4, 1, 6, 1, 6, 2, 1, 1, 5, 4, 1, 5, 4,~
$ ah_1               <dbl+lbl> 2, 2, 1, 1, 5, 2, 2, 1, 1, 1, 1, 1, 2,~
$ ah_2               <dbl+lbl> 2, 1, 2, 1, 5, 2, 2, 1, 1, 1, 1, 1, 1,~
$ pf_1               <dbl+lbl> 4, 4, 5, 5, 3, 4, 2, 7, 3, 5, 7, 5, 5,~
$ pf_2               <dbl+lbl> 1, 5, 1, 4, 3, 5, 5, 2, 5, 2, 4, 3, 4,~
$ af_1               <dbl+lbl> 1, 4, 1, 5, 3, 3, 3, 7, 2, 2, 7, 2, 3,~
$ af_2               <dbl+lbl> 1, 3, 2, 7, 4, 4, 2, 7, 4, 5, 4, 4, 5,~
$ ad_1               <dbl+lbl> 1, 4, 2, 5, 3, 5, 1, 5, 2, 3, 2, 3, 4,~
$ ad_2               <dbl+lbl> 4, 4, 5, 6, 2, 5, 3, 7, 2, 6, 7, 4, 6,~
$ co_1               <dbl+lbl> 2, 1, 1, 1, 6, 2, 4, 1, 5, 1, 1, 1, 3,~
$ co_2               <dbl+lbl> 2, 2, 2, 1, 6, 2, 2, 1, 4, 1, 1, 3, 4,~
$ en_1               <dbl+lbl> 1, 1, 1, 2, 2, 2, 4, 1, 3, 1, 1, 1, 1,~
$ en_2               <dbl+lbl> 1, 1, 1, 1, 2, 2, 4, 1, 4, 1, 1, 1, 1,~
$ pi_1               <dbl+lbl> 1, 1, 6, 4, 5, 1, 2, 6, 4, 3, 6, 6, 6,~
$ pi_2               <dbl+lbl> 1, 1, 6, 3, 1, 2, 1, 7, 2, 4, 7, 5, 5,~
$ sk_1               <dbl+lbl> 7, 6, 6, 7, 6, 2, 7, 6, 4, 5, 7, 5, 3,~
$ sk_2               <dbl+lbl> 7, 7, 6, 5, 7, 2, 7, 7, 5, 6, 7, 3, 5,~
$ sk_3               <dbl+lbl> 7, 7, 7, 7, 7, 2, 7, 5, 4, 6, 7, 7, 5,~
$ ex_po_1            <dbl+lbl> NA, NA,  5,  7, NA, NA, NA,  7, NA,  7~
$ ex_po_2            <dbl+lbl> NA, NA,  6,  5, NA, NA, NA,  6, NA,  7~
$ in_po_1            <dbl+lbl> NA, NA,  4,  2, NA, NA, NA,  4, NA,  4~
$ in_po_2            <dbl+lbl> NA, NA,  2,  1, NA, NA, NA,  5, NA,  5~
$ ex_we_1            <dbl+lbl>  7,  7, NA, NA,  7,  4,  7, NA,  6, NA~
$ ex_we_2            <dbl+lbl>  7,  7, NA, NA,  7,  4,  7, NA,  6, NA~
$ in_we_1            <dbl+lbl>  7,  5, NA, NA,  3,  5,  3, NA,  2, NA~
$ in_we_2            <dbl+lbl>  3,  5, NA, NA,  3,  5,  2, NA,  1, NA~
$ carin_control_1    <dbl+lbl> NA, NA,  4,  7, NA, NA, NA,  2, NA,  4~
```

```
$ carin_control_2           <dbl+lbl> NA, NA,  3,  1, NA, NA, NA,  2, NA,  4~
$ carin_attitude_1          <dbl+lbl> NA, NA,  5,  1, NA, NA, NA,  4, NA,  2~
$ carin_attitude_2          <dbl+lbl> NA, NA,  7,  1, NA, NA, NA,  2, NA,  3~
$ carin_reciprocity_1       <dbl+lbl> NA, NA,  3,  4, NA, NA, NA,  3, NA,  3~
$ carin_reciprocity_2       <dbl+lbl> NA, NA,  5,  1, NA, NA, NA,  2, NA,  4~
$ carin_identity_1          <dbl+lbl> NA, NA,  3,  1, NA, NA, NA,  1, NA,  1~
$ carin_identity_2          <dbl+lbl> NA, NA,  1,  2, NA, NA, NA,  5, NA,  1~
$ carin_need_1              <dbl+lbl> NA, NA,  6,  1, NA, NA, NA,  1, NA,  5~
$ carin_need_2              <dbl+lbl> NA, NA,  5,  1, NA, NA, NA,  1, NA,  5~
$ greedy_1                  <dbl+lbl>  7,  6, NA, NA,  7,  2,  3, NA,  7, NA~
$ greedy_2                  <dbl+lbl>  7,  6, NA, NA,  7,  3,  4, NA,  6, NA~
$ greedy_3                  <dbl+lbl>  7,  6, NA, NA,  7,  3,  4, NA,  5, NA~
$ punish_1                  <dbl+lbl>  7,  7, NA, NA,  7,  2,  6, NA,  7, NA~
$ punish_2                  <dbl+lbl>  7,  7, NA, NA,  7,  2,  7, NA,  7, NA~
$ punish_3                  <dbl+lbl>  7,  7, NA, NA,  7,  2,  7, NA,  7, NA~
$ time_dh_1_First_Click     <dbl> 1.898, 22.881, 20.927, 11.899, 12.041, 12.~
$ time_dh_1_Last_Click      <dbl> 48.775, 34.978, 31.391, 25.883, 21.264, 43~
$ time_dh_1_Page_Submit     <dbl> 49.622, 37.663, 32.915, 27.970, 22.766, 44~
$ time_dh_1_Click_Count     <dbl> 26, 4, 5, 6, 4, 9, 5, 5, 5, 4, 4, 8, 4, 4,~
$ asc_pw                    <dbl> 50, 61, 69, 53, 80, 51, 50, 73, 51, 65, 51~
$ asc_pm                    <dbl> 50, 61, 61, 54, 70, 47, 51, 39, 51, 65, 30~
$ asc_rw                    <dbl> 50, 76, 40, 48, 80, 65, 51, 73, 51, 15, 80~
$ asc_rm                    <dbl> 50, 75, 61, 51, 70, 64, 51, 58, 51, 15, 70~
$ time_sexu_First_Click     <dbl> 1.181, 9.759, 7.360, 8.072, 8.461, 2.646, ~
$ time_sexu_Last_Click      <dbl> 71.999, 79.570, 76.440, 46.009, 56.913, 35~
$ time_sexu_Page_Submit     <dbl> 72.541, 81.801, 77.506, 47.515, 57.619, 36~
$ time_sexu_Click_Count     <dbl> 51, 13, 17, 13, 13, 19, 20, 14, 13, 16, 15~
$ wel_abu_1                 <dbl+lbl> 1, 1, 3, 1, 2, 2, 3, 4, 1, 4, 2, 3, 3,~
$ wel_abu_2                 <dbl+lbl> 1, 1, 2, 1, 2, 2, 3, 2, 1, 2, 2, 4, 3,~
$ wel_pa_1                  <dbl+lbl> 7, 2, 7, 1, 6, 2, 3, 6, 5, 7, 7, 5, 6,~
$ wel_pa_2                  <dbl+lbl> 7, 2, 7, 1, 6, 2, 5, 6, 4, 6, 7, 7, 5,~
$ wel_ho_1                  <dbl+lbl> 1, 1, 1, 1, 2, 2, 2, 3, 1, 5, 1, 1, 2,~
$ wel_ho_2                  <dbl+lbl> 1, 1, 1, 1, 2, 2, 4, 4, 1, 6, 1, 4, 2,~
$ pro_pw                    <dbl+lbl> 4, 2, 3, 1, 2, 3, 3, 2, 1, 2, 1, 2, 5,~
$ pro_rw                    <dbl+lbl> 4, 2, 6, 1, 5, 4, 3, 4, 1, 6, 7, 5, 6,~
$ ris_pw                    <dbl+lbl> 6, 2, 6, 1, 6, 4, 3, 3, 4, 4, 7, 6, 6,~
$ ris_rw                    <dbl+lbl> 3, 1, 5, 1, 4, 4, 3, 3, 5, 5, 5, 4, 2,~
$ pre_pw                    <dbl+lbl> 6, 3, 6, 3, 6, 4, 4, 3, 5, 5, 7, 4, 6,~
$ pre_rw                    <dbl+lbl> 3, 1, 4, 3, 2, 4, 2, 3, 3, 2, 2, 5, 1,~
$ time_poli_1_First_Click   <dbl> 1.453, 15.591, 12.917, 12.337, 40.588, 5.5~
$ time_poli_1_Last_Click    <dbl> 106.685, 99.407, 95.733, 90.936, 112.949, ~
```

```
$ time_poli_1_Page_Submit     <dbl> 107.394, 101.436, 96.779, 92.410, 114.003,~
$ time_poli_1_Click_Count     <dbl> 56, 15, 16, 17, 16, 16, 16, 15, 16, 16, 16~
$ redi_1                      <dbl+lbl> 7, 7, 7, 5, 7, 4, 7, 7, 6, 7, 6, 5, 6,~
$ redi_2                      <dbl+lbl> 7, 7, 6, 1, 7, 3, 7, 7, 7, 7, 1, 6, 7,~
$ effec_pw_1                  <dbl+lbl> 1, 1, 5, 1, 3, 3, 2, 2, 2, 3, 2, 4, 2,~
$ effec_pw_2                  <dbl+lbl> 7, 6, 3, 5, 4, 3, 3, 5, 2, 3, 4, 3, 6,~
$ effec_pm_1                  <dbl+lbl> 1, 1, 6, 1, 4, 4, 3, 3, 2, 5, 7, 5, 5,~
$ effec_pm_2                  <dbl+lbl> 7, 6, 3, 4, 3, 4, 3, 4, 2, 4, 7, 5, 3,~
$ poli_progre_1               <dbl+lbl> 7, 7, 5, 6, 7, 2, 7, 6, 6, 6, 7, 6, 5,~
$ poli_progre_2               <dbl+lbl> 7, 7, 5, 6, 7, 3, 5, 7, 6, 6, 7, 6, 6,~
$ poli_restri_1               <dbl+lbl> 7, 4, 6, 1, 6, 3, 4, 4, 4, 6, 6, 3, 4,~
$ poli_restri_2               <dbl+lbl> 3, 6, 5, 1, 4, 3, 2, 6, 3, 4, 7, 5, 5,~
$ aut_pw_1                    <dbl+lbl> 7, 6, 3, 5, 5, 4, 2, 2, 3, 4, 7, 3, 3,~
$ aut_pm_1                    <dbl+lbl> 7, 6, 3, 5, 4, 4, 2, 3, 4, 4, 7, 2, 3,~
$ depe_pw_1                   <dbl+lbl> 6, 2, 5, 1, 6, 4, 5, 4, 4, 4, 7, 5, 5,~
$ depe_pm_1                   <dbl+lbl> 6, 3, 5, 1, 6, 4, 5, 4, 4, 4, 7, 5, 5,~
$ time_violence_First_Click   <dbl> 1.529, 11.600, 32.419, 63.577, 13.948, 6.5~
$ time_violence_Last_Click    <dbl> 85.503, 117.811, 121.627, 225.891, 87.475,~
$ time_violence_Page_Submit   <dbl> 86.202, 120.009, 122.810, 235.969, 93.996,~
$ time_violence_Click_Count   <dbl> 58, 22, 27, 26, 25, 40, 30, 22, 23, 25, 25~
$ condi_viole                 <dbl+lbl> 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1,~
$ hara_pw_1                   <dbl+lbl> 7, 6, 3, 7, 5, 5, 5, 5, 5, 6, 4, 5, 4,~
$ hara_pw_2                   <dbl+lbl> 7, 7, 7, 7, 7, 7, 7, 6, 7, 7, 7, 7, 5,~
$ hara_pw_3                   <dbl+lbl> 7, 6, 2, 7, 6, 7, 7, 5, 6, 7, 7, 7, 6,~
$ abu_pw_1                    <dbl+lbl> 7, 7, 3, 7, 5, 7, 7, 6, 6, 7, 7, 7, 7,~
$ abu_pw_2                    <dbl+lbl> 7, 7, 4, 7, 6, 7, 7, 7, 7, 7, 7, 7, 7,~
$ abu_pw_3                    <dbl+lbl> 7, 7, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,~
$ viole_pw_1                  <dbl+lbl> 7, 5, 7, 2, 3, 3, 3, 6, 4, 7, 6, 5, 2,~
$ viole_pw_2                  <dbl+lbl> 7, 6, 7, 2, 5, 4, 4, 5, 4, 6, 6, 5, 3,~
$ viole_pw_3                  <dbl+lbl> 7, 7, 6, 2, 7, 4, 4, 6, 6, 7, 6, 5, 5,~
$ viole_pw_4                  <dbl+lbl> 7, 5, 6, 2, 5, 4, 4, 6, 4, 6, 6, 4, 2,~
$ viole_pw_5                  <dbl+lbl> 7, 2, 6, 2, 2, 3, 4, 4, 7, 6, 4, 3, 3,~
$ viole_pw_6                  <dbl+lbl> 7, 6, 5, 2, 6, 5, 4, 6, 6, 6, 7, 4, 4,~
$ barri_pw_1                  <dbl+lbl> 6, 5, 7, 2, 2, 3, 6, 6, 6, 7, 7, 7, 5,~
$ barri_pw_2                  <dbl+lbl> 6, 1, 7, 2, 1, 3, 5, 7, 6, 7, 7, 6, 3,~
$ barri_pw_3                  <dbl+lbl> 6, 6, 6, 2, 4, 4, 3, 7, 6, 6, 7, 4, 5,~
$ barri_pw_4                  <dbl+lbl> 6, 3, 6, 2, 3, 4, 6, 7, 6, 6, 7, 4, 2,~
$ barri_pw_5                  <dbl+lbl> 6, 6, 5, 2, 6, 4, 6, 5, 4, 7, 7, 3, 3,~
$ perpe_1                     <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ perpe_2                     <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ perpe_3                     <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
```

```
$ perpe_4                    <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ perpe_5                    <dbl+lbl> 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1,~
$ time_demo_1_First_Click    <dbl> 1.801, 1.418, 2.388, 1.532, 1.449, 0.778, ~
$ time_demo_1_Last_Click     <dbl> 151.620, 113.887, 108.556, 125.513, 61.897~
$ time_demo_1_Page_Submit    <dbl> 152.193, 115.771, 110.955, 126.555, 63.446~
$ time_demo_1_Click_Count    <dbl> 64, 24, 27, 29, 22, 26, 25, 25, 19, 26, 22~
$ age                        <dbl+lbl> 54, 58, 57, 30, 25, 22, 27, 29, 22, 41~
$ sex                        <dbl+lbl> 2, 1, 2, 1, 2, 2, 1, 1, 1, 2, 1, 1, 2,~
$ sex_other                  <chr> "", "", "", "", "", "", "", "", "", "", ""~
$ edu                        <dbl+lbl> 5, 5, 5, 6, 5, 5, 5, 4, 5, 5, 6, 5, 6,~
$ ses                        <dbl+lbl> 6, 6, 6, 7, 7, 7, 6, 5, 5, 4, 6, 8, 6,~
$ hig_ide                    <dbl+lbl> 2, 1, 1, 4, 2, 4, 1, 2, 2, 1, 3, 4, 3,~
$ mid_ide                    <dbl+lbl> 5, 6, 6, 6, 6, 5, 4, 6, 4, 3, 7, 6, 6,~
$ low_ide                    <dbl+lbl> 3, 1, 2, 2, 1, 2, 3, 2, 3, 5, 1, 3, 2,~
$ po                         <dbl+lbl> 1, 2, 2, 3, 2, 5, 1, 2, 2, 1, 5, 6, 6,~
$ country_residence          <dbl+lbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,~
$ country_residence_other    <chr> "", "", "", "", "", "", "", "", "", "", ""~
$ country_residence_recoded  <dbl+lbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,~
$ natio_arge                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_colom                <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_espa                 <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ natio_mex                  <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_chile                <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_peru                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_boli                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_cost                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_cuba                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_ecua                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_elsa                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_guat                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_eqgu                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_hond                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_nica                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_pana                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_para                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_puer                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_domi                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_urug                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_vene                 <dbl+lbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_other                <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ natio_other_text           <chr> "", "", "", "", "", "", "", "", "", "", ""~
```

```
$ lang                  <dbl+lbl> 1, 1, 3, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ lang_other            <chr> "", "", "Catalán", "Catalán", "", "", "", ~
$ lang_recoded          <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ inc                   <dbl> 3200, 1300, 3000, 60000, 3500, 600, 1800, ~
$ currency              <dbl+lbl> 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,~
$ post_code             <chr> "40197", "47001", "08020", "00001", "41005~
$ municipality          <chr> "Segovia", "Valladolid", "sant marti", "-"~
$ n_perso               <dbl+lbl> 3, 1, 4, 2, 3, 3, 3, 2, 1, 3, 1, 3, 4,~
$ ori_sex               <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1,~
$ ori_sex_other         <chr> "", "", "", "", "", "", "", "", "", "", ""~
$ relation              <dbl+lbl> 1, 2, 1, 1, 1, 2, 1, 1, 2, 1, 2, 1, 1,~
$ natio_recoded         <dbl+lbl> 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,~
$ regional_area         <dbl+lbl> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,~
$ PrimarioÚltimo        <dbl+lbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
```

We have 4,386 cases or rows and 283 variables or columns.

4 Processing

4.1 Select

We exclude variables related to attention checks, survey response time, dummy nationalities and the auxiliary variable `PrimarioÚltimo`, which indicates whether there are duplicate cases.

```
db_proc <- sogedi_db %>%
  dplyr::select(-c(matches("^(aten|time)"), 247:269, 283))
```

4.2 Filter

We filter out cases from countries without a sufficiently large sample size for statistical analysis, retaining only those from Argentina, Chile, Colombia, Spain, and Mexico.

```
frq(db_proc$natio_recoded)
```

```
Recodification of nationality based on country of residence, declared nationality, and
# total N=4386 valid N=4386 mean=6.46 sd=5.00

Value |         Label |   N | Raw % | Valid % | Cum. %
------------------------------------------------------
    1 |     Argentine | 857 | 19.54 |   19.54 |  19.54
    2 |      Bolivian |   1 |  0.02 |    0.02 |  19.56
    3 |       Chilean | 860 | 19.61 |   19.61 |  39.17
    4 |     Colombian | 824 | 18.79 |   18.79 |  57.96
    5 |   Costa Rican |   0 |  0.00 |    0.00 |  57.96
    6 |         Cuban |   5 |  0.11 |    0.11 |  58.07
    7 |    Ecuadorian |   3 |  0.07 |    0.07 |  58.14
    8 |    Salvadoran |   2 |  0.05 |    0.05 |  58.19
    9 |       Spanish | 831 | 18.95 |   18.95 |  77.13
   10 |    Guatemalan |   1 |  0.02 |    0.02 |  77.15
   11 | Equatoguinean |   0 |  0.00 |    0.00 |  77.15
   12 |      Honduran |   2 |  0.05 |    0.05 |  77.20
   13 |       Mexican | 837 | 19.08 |   19.08 |  96.28
   14 |    Nicaraguan |   1 |  0.02 |    0.02 |  96.31
   15 |    Panamanian |   1 |  0.02 |    0.02 |  96.33
   16 |    Paraguayan |   2 |  0.05 |    0.05 |  96.37
   17 |      Peruvian |  70 |  1.60 |    1.60 |  97.97
   18 |  Puerto Rican |   0 |  0.00 |    0.00 |  97.97
   19 |     Dominican |   0 |  0.00 |    0.00 |  97.97
   20 |     Uruguayan |   7 |  0.16 |    0.16 |  98.13
   21 |    Venezuelan |  75 |  1.71 |    1.71 |  99.84
   22 |        Rusian |   1 |  0.02 |    0.02 |  99.86
   23 |         Swiss |   1 |  0.02 |    0.02 |  99.89
   24 |          EEUU |   1 |  0.02 |    0.02 |  99.91
   25 |     Brasilian |   4 |  0.09 |    0.09 | 100.00
 <NA> |          <NA> |   0 |  0.00 |    <NA> |   <NA>
```

```r
db_proc <- db_proc %>%
  dplyr::filter(natio_recoded %in% c(1,3,4,9,13))
```

4.3 Recode and transform

Not required.

4.4 Missing values

There is a total of 50.569 missing values in the database, which represents the 5.7% of the total.

```
n_miss(db_proc) # total of NA's
```

```
[1] 50569
```

```
prop_miss(db_proc)*100 # proportion of NA's
```

```
[1] 5.667214
```

Let's see the number and percentage of missing values per variable:

```
db_proc %>%
  select(-c(ex_we_1, ex_we_2, in_we_1, in_we_2,
            ex_po_1, ex_po_2, in_po_1, in_po_2,
            matches("^(greedy|punish|carin)"))) %>%
  miss_var_summary(.) %>%
  filter(pct_miss > 0) %>%
  kable(.,"markdown")
```

variable	n_miss	pct_miss
po	25	0.594
age	21	0.499
ori_sex	11	0.261
inc	10	0.238
relation	8	0.190
country_residence	5	0.119
lang_recoded	2	0.0475
n_perso	1	0.0238

5 Save and export

Finally, we save and export the processed database `db_proc` in `.RData`, `.dta` and `.sav` formats.

```
save(db_proc, file = here("output/data/db_proc.RData"))
haven::write_dta(db_proc, path = here("output/data/db_proc.dta"))
haven::write_sav(db_proc, path = here("output/data/db_proc.sav"))
```