Analysis of the Data from INEGI

INEGI is the Mexican National Institute of Statistics and Geography, which provides a wealth of data on various aspects of Mexican society, including demographics, economics, and geography. The data we will be analyzing comes from the 2020 and 2010 census datasets for Yucatán.

We are particularly interested in how migration patterns have affected the population of the biggest cities in the state, Merida, Valladolid and the coast area, like Progreso and Telchac.

```
file_path_2020 <- "/Users/ecastillo/Documents/coding/analyzing_merida_data/ITER2020 - 31 Yucatán.csv"
data_2020 <- read.csv(file_path_2020, fileEncoding = "UTF-8", sep = ",", header = TRUE)
file_path_2010 <- "/Users/ecastillo/Documents/coding/analyzing_merida_data/RESLOC2010 - 31 Yucatán.csv"
data_2010 <- read.csv(file_path_2010, fileEncoding = "UTF-8", sep = ",", header = TRUE)
# Display the first few rows of the dataset
head(data_2020, 10)
```

```
##    ENTIDAD NOM_ENT MUN                     NOM_MUN  LOC
## 1       31 Yucatán   0 Total de la entidad Yucatán    0
## 2       31 Yucatán   0 Total de la entidad Yucatán 9998
## 3       31 Yucatán   0 Total de la entidad Yucatán 9999
## 4       31 Yucatán   1                       Abalá    0
## 5       31 Yucatán   1                       Abalá    1
## 6       31 Yucatán   1                       Abalá    3
## 7       31 Yucatán   1                       Abalá    5
## 8       31 Yucatán   1                       Abalá    6
## 9       31 Yucatán   1                       Abalá    7
## 10      31 Yucatán   1                       Abalá    8
##                          NOM_LOC  POBTOT PNACENT PNACOE P3YM_HLI P3HLINHE
## 1           Total de la Entidad 2320898 2029698 269765   525092    24640
## 2   Localidades de una vivienda    3950    3295    546     1746      194
## 3  Localidades de dos viviendas     955     856     92      549       98
## 4           Total del Municipio    6550    6518     28     3484       89
## 5                         Abalá    2039    2014     21      548        0
## 6                      Mucuyché     490     490      0      345        7
## 7                     Sihunchén     282     282      0       76        0
## 8                       Temozón     668     666      2      310        1
## 9                      Uayalceh    2608    2603      5     2101       81
## 10                         Peba     239     239      0       68        0
##    P3HLI_HE P5_HLI P5_HLI_NHE P5_HLI_HE PHOG_IND POB_AFRO GRAPROES     PEA
## 1    497656 520580      23929    493856   983257    69599     9.59 1160284
## 2      1549   1742        194      1545     2467      149     6.51    2467
## 3       445    545         97       442      740       46     6.29     504
## 4      3383   3438         88      3338     5197      377     6.77    2809
## 5       548    547          0       547     1139       47     8.15     941
## 6       338    343          7       336      479        1     6.50     212
## 7        76     76          0        76      184        0     6.85     113
## 8       308    309          1       307      592      312     7.60     272
## 9      2009   2059         80      1968     2588       17     5.42    1075
## 10       68     68          0        68      134        0     7.16     109
##    POCUPADA PDESOCUP PE_INAC TOTHOG  POBHOG VIVTOT TVIVHAB TVIVPAR VIVPAR_DES
## 1   1146809    13475  713910 658085 2317081 837334  658351  812181     108752
## 2      2462        5     890   1454    3920   1705    1456    1677        136
## 3       502        2     276    296     955    457     296     454         75
## 4      2774       35    2410   1822    6550   2208    1822    2208        219
```

```
## 5       931       10     719     608     2039     742      608      742          72
## 6       212        0     183     147      490     185      147      185          25
## 7       106        7     129      80      282      96       80       96           5
## 8       269        3     290     173      668     228      173      228          38
## 9      1066        9     911     671     2608     771      671      771          66
## 10      106        3      89      76      239      92       76       92           0
##    VIVPAR_UT OCUPVIVPAR TVIVPARHAB VIVPAR_HAB PROM_OCUP PRO_OCUP_C PSINDER
## 1      70231    2317081     658085     633198      3.52       1.05  505108
## 2        113       3920       1454       1428      2.70       1.08    1143
## 3         86        955        296        293      3.23       1.36     230
## 4        167       6550       1822       1822      3.59       1.46    1193
## 5         62       2039        608        608      3.35       1.29     249
## 6         13        490        147        147      3.33       1.64      59
## 7         11        282         80         80      3.53       1.33      61
## 8         17        668        173        173      3.86       1.42     125
## 9         34       2608        671        671      3.89       1.65     554
## 10        16        239         76         76      3.14       1.31      80
##    PDER_SS
## 1  1810121
## 2     2792
## 3      725
## 4     5356
## 5     1789
## 6      431
## 7      221
## 8      543
## 9     2054
## 10     159
```

```
# Display the first few rows of the dataset
head(data_2010, 10)
```

```
##    ENTIDAD NOM_ENT MUN                    NOM_MUN  LOC
## 1       31 Yucatán   0 Total de la entidad Yucatán    0
## 2       31 Yucatán   0 Total de la entidad Yucatán 9998
## 3       31 Yucatán   0 Total de la entidad Yucatán 9999
## 4       31 Yucatán   1                      Abalá    0
## 5       31 Yucatán   1                      Abalá 9998
## 6       31 Yucatán   1                      Abalá 9999
## 7       31 Yucatán   1                      Abalá    1
## 8       31 Yucatán   1                      Abalá    3
## 9       31 Yucatán   1                      Abalá    5
## 10      31 Yucatán   1                      Abalá    6
##                         NOM_LOC P_TOTAL TOTHOG P3YM_HLI P3HLINHE P3HLI_HE
## 1           Total de la Entidad 1955577 503106   544927    43010   492297
## 2   Localidades de una vivienda    4513   1491     2474      388     2040
## 3  Localidades de dos viviendas    1260    336      844      222      614
## 4           Total del Municipio    6356   1606     3799      206     3545
## 5   Localidades de una vivienda       1      1        1        0        1
## 6  Localidades de dos viviendas       7      2        1        0        1
## 7                         Abalá    1890    523      624        4      614
## 8                      Mucuyché     494    127      391       17      374
## 9                     Sihunchén     334     89      142        0      140
## 10                      Temozón     760    185      419        6      379
```

```
##    P5_HLI P5_HLI_NHE P5_HLI_HE PHOG_IND PNACENT PNACOE VIVTOT T_VIVHAB TVIVPAR
## 1  537516      40273    487751   956352 1772324 156210 638502   507248  634360
## 2    2440        377      2018     3404    4041    380   1987     1503    1975
## 3     824        213       603     1075    1190     58    442      338     440
## 4    3725        200      3477     5505    6322     25   1829     1608    1827
## 5       1          0         1        1       1      0      1        1       1
## 6       1          0         1        6       7      0      3        2       3
## 7     624          4       614     1281    1867     16    628      525     626
## 8     390         17       373      494     494      0    140      127     140
## 9     142          0       140      282     333      1     94       89      94
## 10    419          6       379      727     756      3    203      185     203
##    VIVPAR_HAB TVIVPARHAB
## 1      503106     507145
## 2        1491       1501
## 3         336        338
## 4        1606       1608
## 5           1          1
## 6           2          2
## 7         523        525
## 8         127        127
## 9          89         89
## 10        185        185
```

We also need a mapping between the column names and its meaning:

```
# Create a mapping of column names to their meanings
query_to_readable_en_2020 <- c(
  "ENTIDAD" = "State code",
  "NOM_ENT" = "State name",
  "MUN" = "Municipality or borough code",
  "NOM_MUN" = "Municipality or borough name",
  "LOC" = "Locality code",
  "NOM_LOC" = "Locality name",
  "POBTOT" = "Total population",
  "PNACENT" = "Population born in the state",
  "PNACOE" = "Population born in another state",
  "P3YM_HLI" = "Population aged 3+ speaking an Indigenous language",
  "P3HLINHE" = "Population aged 3+ speaking an Indigenous language and not Spanish",
  "P3HLI_HE" = "Population aged 3+ speaking an Indigenous language and Spanish",
  "P5_HLI" = "Population aged 5+ speaking an Indigenous language",
  "P5_HLI_NHE" = "Population aged 5+ speaking an Indigenous language and not Spanish",
  "P5_HLI_HE" = "Population aged 5+ speaking an Indigenous language and Spanish",
  "PHOG_IND" = "Population in Indigenous census households",
  "POB_AFRO" = "Population identifying as Afro-Mexican or of African descent",
  "GRAPROES" = "Average years of schooling",
  "PEA" = "Economically active population aged 12+",
  "POCUPADA" = "Employed population aged 12+",
  "PDESOCUP" = "Unemployed population aged 12+",
  "PE_INAC" = "Economically inactive population aged 12+",
  "TOTHOG" = "Total census households",
  "POBHOG" = "Population in census households",
  "VIVTOT" = "Total housing units",
  "TVIVHAB" = "Total inhabited housing units",
  "TVIVPAR" = "Total private housing units",
```

```
  "VIVPAR_DES" = "Uninhabited private housing units",
  "VIVPAR_UT" = "Private housing units for seasonal use",
  "OCUPVIVPAR" = "Occupants in inhabited private housing units",
  "TVIVPARHAB" = "Total inhabited private housing units",
  "VIVPAR_HAB" = "Inhabited private housing units",
  "PROM_OCUP" = "Average occupants per inhabited private housing unit",
  "PRO_OCUP_C" = "Average occupants per room in inhabited private housing units",
  "PSINDER" = "Population without health service affiliation",
  "PDER_SS" = "Population with health service affiliation"
)

query_to_readable_en_2010 <- c(
  "ENTIDAD"       = "State code",
  "NOM_ENT"       = "State name",
  "MUN"           = "Municipality or borough code",
  "NOM_MUN"       = "Municipality or borough name",
  "LOC"           = "Locality code",
  "NOM_LOC"       = "Locality name",
  "P_TOTAL"       = "Total population",
  "TOTHOG"        = "Total census households",
  "P3YM_HLI"      = "Population aged 3+ speaking an Indigenous language",
  "P3HLINHE"      = "Population aged 3+ speaking an Indigenous language and not Spanish",
  "P3HLI_HE"      = "Population aged 3+ speaking an Indigenous language and Spanish",
  "P5_HLI"        = "Population aged 5+ speaking an Indigenous language",
  "P5_HLI_NHE"    = "Population aged 5+ speaking an Indigenous language and not Spanish",
  "P5_HLI_HE"     = "Population aged 5+ speaking an Indigenous language and Spanish",
  "PHOG_IND"      = "Population in Indigenous census households",
  "PNACENT"       = "Population born in the state",
  "PNACOE"        = "Population born in another state",
  "VIVTOT"        = "Total housing units",
  "T_VIVHAB"      = "Total inhabited housing units",
  "TVIVPAR"       = "Total private housing units",
  "VIVPAR_HAB"    = "Inhabited private housing units",
  "TVIVPARHAB"    = "Total inhabited private housing units"
)
```

Analysis of the Data from INEGI

We will analyze the data to understand the migration patterns and demographic changes in Yucatán, focusing on the cities of Merida, Valladolid, Progreso, and Telchac. The analysis will include: - Population changes between 2010 and 2020. - How has the migration impacted the density of Indigenous people in these cities. - How has this impacted the occupation of houses.

Lets explore the data

```
# Display the structure of the 2020 dataset
str(data_2020)
```

```
## 'data.frame':    2691 obs. of  36 variables:
##  $ ENTIDAD   : int  31 31 31 31 31 31 31 31 31 31 ...
```

```
##  $ NOM_ENT   : chr  "Yucatán" "Yucatán" "Yucatán" "Yucatán" ...
##  $ MUN       : int  0 0 0 1 1 1 1 1 1 1 ...
##  $ NOM_MUN   : chr  "Total de la entidad Yucatán" "Total de la entidad Yucatán" "Total de la entidad
##  $ LOC       : int  0 9998 9999 0 1 3 5 6 7 8 ...
##  $ NOM_LOC   : chr  "Total de la Entidad" "Localidades de una vivienda" "Localidades de dos viviendas
##  $ POBTOT    : int  2320898 3950 955 6550 2039 490 282 668 2608 239 ...
##  $ PNACENT   : chr  "2029698" "3295" "856" "6518" ...
##  $ PNACOE    : chr  "269765" "546" "92" "28" ...
##  $ P3YM_HLI  : chr  "525092" "1746" "549" "3484" ...
##  $ P3HLINHE  : chr  "24640" "194" "98" "89" ...
##  $ P3HLI_HE  : chr  "497656" "1549" "445" "3383" ...
##  $ P5_HLI    : chr  "520580" "1742" "545" "3438" ...
##  $ P5_HLI_NHE: chr  "23929" "194" "97" "88" ...
##  $ P5_HLI_HE : chr  "493856" "1545" "442" "3338" ...
##  $ PHOG_IND  : chr  "983257" "2467" "740" "5197" ...
##  $ POB_AFRO  : chr  "69599" "149" "46" "377" ...
##  $ GRAPROES  : chr  "9.59" "6.51" "6.29" "6.77" ...
##  $ PEA       : chr  "1160284" "2467" "504" "2809" ...
##  $ POCUPADA  : chr  "1146809" "2462" "502" "2774" ...
##  $ PDESOCUP  : chr  "13475" "5" "2" "35" ...
##  $ PE_INAC   : chr  "713910" "890" "276" "2410" ...
##  $ TOTHOG    : chr  "658085" "1454" "296" "1822" ...
##  $ POBHOG    : chr  "2317081" "3920" "955" "6550" ...
##  $ VIVTOT    : int  837334 1705 457 2208 742 185 96 228 771 92 ...
##  $ TVIVHAB   : int  658351 1456 296 1822 608 147 80 173 671 76 ...
##  $ TVIVPAR   : chr  "812181" "1677" "454" "2208" ...
##  $ VIVPAR_DES: chr  "108752" "136" "75" "219" ...
##  $ VIVPAR_UT : chr  "70231" "113" "86" "167" ...
##  $ OCUPVIVPAR: chr  "2317081" "3920" "955" "6550" ...
##  $ TVIVPARHAB: chr  "658085" "1454" "296" "1822" ...
##  $ VIVPAR_HAB: chr  "633198" "1428" "293" "1822" ...
##  $ PROM_OCUP : chr  "3.52" "2.70" "3.23" "3.59" ...
##  $ PRO_OCUP_C: chr  "1.05" "1.08" "1.36" "1.46" ...
##  $ PSINDER   : chr  "505108" "1143" "230" "1193" ...
##  $ PDER_SS   : chr  "1810121" "2792" "725" "5356" ...
```

```r
# Display the structure of the 2010 dataset
str(data_2010)
```

```
## 'data.frame':    2774 obs. of  22 variables:
##  $ ENTIDAD   : int  31 31 31 31 31 31 31 31 31 31 ...
##  $ NOM_ENT   : chr  "Yucatán" "Yucatán" "Yucatán" "Yucatán" ...
##  $ MUN       : int  0 0 0 1 1 1 1 1 1 1 ...
##  $ NOM_MUN   : chr  "Total de la entidad Yucatán" "Total de la entidad Yucatán" "Total de la entidad
##  $ LOC       : int  0 9998 9999 0 9998 9999 1 3 5 6 ...
##  $ NOM_LOC   : chr  "Total de la Entidad" "Localidades de una vivienda" "Localidades de dos viviendas
##  $ P_TOTAL   : int  1955577 4513 1260 6356 1 7 1890 494 334 760 ...
##  $ TOTHOG    : chr  "503106" "1491" "336" "1606" ...
##  $ P3YM_HLI  : chr  "544927" "2474" "844" "3799" ...
##  $ P3HLINHE  : chr  "43010" "388" "222" "206" ...
##  $ P3HLI_HE  : chr  "492297" "2040" "614" "3545" ...
##  $ P5_HLI    : chr  "537516" "2440" "824" "3725" ...
##  $ P5_HLI_NHE: chr  "40273" "377" "213" "200" ...
##  $ P5_HLI_HE : chr  "487751" "2018" "603" "3477" ...
```

```
##  $ PHOG_IND  : chr  "956352" "3404" "1075" "5505" ...
##  $ PNACENT   : chr  "1772324" "4041" "1190" "6322" ...
##  $ PNACOE    : chr  "156210" "380" "58" "25" ...
##  $ VIVTOT    : int  638502 1987 442 1829 1 3 628 140 94 203 ...
##  $ T_VIVHAB  : int  507248 1503 338 1608 1 2 525 127 89 185 ...
##  $ TVIVPAR   : chr  "634360" "1975" "440" "1827" ...
##  $ VIVPAR_HAB: chr  "503106" "1491" "336" "1606" ...
##  $ TVIVPARHAB: chr  "507145" "1501" "338" "1608" ...
```

We can see that the data types may not be correct, so we will convert them when needed.

Let's see the distribution of the population in the state of Yucatán

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# lets see which are the biggest cities in Yucatán, the data has the total population as MUN == 0, so l
# the LOC = 0 is the total of the municipality
# we only want to keep the
biggest_cities_2020 <- data_2020 %>%
  filter(MUN != 0) %>%
  filter(LOC == 0) %>%
  arrange(desc(POBTOT))

# Lets add the population from 2010
biggest_cities_2010 <- data_2010 %>%
  filter(MUN != 0) %>%
  filter(LOC == 0) %>%
  arrange(desc(P_TOTAL))
```

Now that we have only the totals for each municipality, lets make sure we have the 106 Municipalities that exist in Yucatán.

```r
# Check the number of unique municipalities in the 2020 data
num_municipalities_2020 <- nrow(biggest_cities_2020)
num_municipalities_2010 <- nrow(biggest_cities_2010)
cat("Number of unique municipalities in 2020:", num_municipalities_2020, "\n")
```

```
## Number of unique municipalities in 2020: 106
```

```r
cat("Number of unique municipalities in 2010:", num_municipalities_2010, "\n")
```

```
## Number of unique municipalities in 2010: 106
```

We can see that there are 106 municipalities in both datasets, so we can proceed with the analysis.

What are the biggest cities in Yucatan at the moment (2020)?

```r
# the total population is in the row where MUN is 0 and LOC is 0, so we can filter that out
total_population_2020 <- (
  data_2020 %>%
  filter(MUN == 0, LOC == 0) %>%
  select(POBTOT)
  )$POBTOT



# Display the top 10 biggest cities in Yucatán in 2020, add a column with the percentage of the total p
biggest_cities_2020 %>%
  select(NOM_MUN, POBTOT) %>%
  arrange(desc(POBTOT)) %>%
  mutate(PercentageOfTotal = POBTOT / total_population_2020 * 100) %>%
  head(10)
```

```
##       NOM_MUN POBTOT PercentageOfTotal
## 1      Mérida 995129         42.876895
## 2     Kanasín 141939          6.115693
## 3  Valladolid  85460          3.682195
## 4     Tizimín  80672          3.475896
## 5        Umán  69147          2.979321
## 6    Progreso  66008          2.844072
## 7       Tekax  45062          1.941576
## 8       Ticul  40495          1.744799
## 9      Chemax  38934          1.677540
## 10      Motul  37804          1.628852
```

Lets see the distribution of the population in the biggest cities

Lets analyze the population that was born in the state vs in another state with a pie graph.

```r
pop_born_in_state <- biggest_cities_2020 %>%
  select(NOM_MUN, POBTOT, PNACENT, PNACOE) %>%
  mutate(
    BornInState = PNACENT,
    BornInOtherState = PNACOE
  ) %>%
  select(NOM_MUN, POBTOT, BornInState, BornInOtherState) %>%
  arrange(desc(POBTOT))
```

What were the biggest cities in Yucatan in 2010?

Population Changes

Let's see which cities had the biggest changes in population between 2010 and 2020. For that we need to focus only on total population, population born in the state and population born in another state.

```
population_changes <- biggest_cities_2020 %>%
  select(MUN, NOM_MUN, POBTOT, PNACENT, PNACOE) %>%
  rename(
    Population2020 = POBTOT,
    BornInState2020 = PNACENT,
    BornInOtherState2020 = PNACOE
  ) %>%
  # convert to numeric
  mutate(
    Population2020 = as.numeric(Population2020),
    BornInState2020 = as.numeric(BornInState2020),
    BornInOtherState2020 = as.numeric(BornInOtherState2020)
  ) %>%
  left_join(
    biggest_cities_2010 %>%
      select(MUN, P_TOTAL, PNACENT, PNACOE) %>%
      rename(Population2010 = P_TOTAL,
             BornInState2010 = PNACENT,
             BornInOtherState2010 = PNACOE) %>%
      mutate(
        MUN = as.numeric(MUN),
        Population2010 = as.numeric(Population2010),
        BornInState2010 = as.numeric(BornInState2010),
        BornInOtherState2010 = as.numeric(BornInOtherState2010)
      ),
    by = "MUN"
  ) %>%
  mutate(
    MUN = as.numeric(MUN),
    PopulationChange = Population2020 - Population2010,
    BornInStateChange = BornInState2020 - BornInState2010,
    BornInOtherStateChange = BornInOtherState2020 - BornInOtherState2010,
    PercentageChange = (PopulationChange / Population2010) * 100
  ) %>%
  arrange(desc(PopulationChange))
```

Display the population changes

Total population changes.

```
library(knitr)
kable(population_changes %>%
  select(MUN, NOM_MUN, Population2010, Population2020, PopulationChange, PercentageChange,) %>%
  arrange(desc(PercentageChange)),
  caption = "Population Changes in Yucatán Municipalities (2010-2020)",
  digits = 2)
```

Table 1: Population Changes in Yucatán Municipalities (2010–2020)

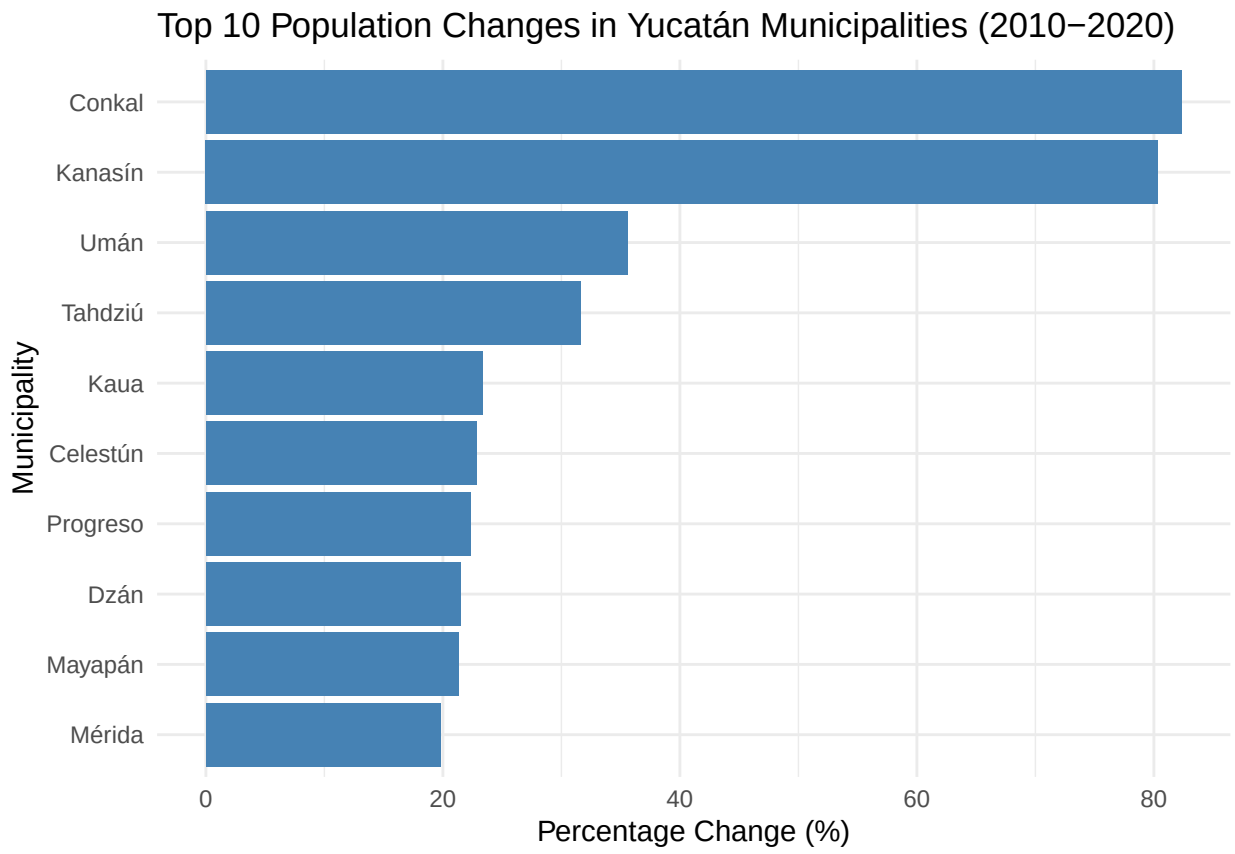
MUN	NOM_MUN	Population2010	Population2020	PopulationChange	PercentageChange
13	Conkal	9143	16671	7528	82.34
41	Kanasín	78709	141939	63230	80.33
101	Umán	50993	69147	18154	35.60
73	Tahdziú	4447	5854	1407	31.64
43	Kaua	2761	3405	644	23.32
11	Celestún	6831	8389	1558	22.81
59	Progreso	53958	66008	12050	22.33
25	Dzán	4941	6003	1062	21.49
49	Mayapán	3269	3965	696	21.29
50	Mérida	830732	995129	164397	19.79
28	Dzilam de Bravo	2463	2936	473	19.20
3	Akil	10362	12285	1923	18.56
92	Tixcacalcupul	6665	7888	1223	18.35
21	Chichimilá	7952	9406	1454	18.28
100	Ucú	3469	4049	580	16.72
19	Chemax	33490	38934	5444	16.26
87	Tetiz	4725	5464	739	15.64
61	Río Lagartos	3438	3974	536	15.59
56	Oxkutzcab	29325	33854	4529	15.44
71	Sudzal	1689	1949	260	15.39
65	San Felipe	1839	2118	279	15.17
102	Valladolid	74217	85460	11243	15.15
105	Yaxkukul	2868	3293	425	14.82
94	Tixmehuac	4746	5444	698	14.71
44	Kinchil	6571	7530	959	14.59
10	Cantamayec	2407	2755	348	14.46
38	Hunucmá	30731	35137	4406	14.34
46	Mama	2888	3296	408	14.13
47	Maní	5250	5968	718	13.68
30	Dzitás	3540	4015	475	13.42
85	Temozón	14801	16680	1879	12.70
55	Opichén	6285	7080	795	12.65
63	Samahil	5008	5631	623	12.44
35	Hoctún	5697	6384	687	12.06
15	Cuzamá	4966	5560	594	11.96
51	Mocochá	3071	3430	359	11.69
75	Teabo	6205	6921	716	11.54
18	Chapab	3035	3385	350	11.53
80	Tekit	9884	11020	1136	11.49
36	Homún	7257	8090	833	11.48
33	Halachó	19072	21255	2183	11.45
52	Motul	33978	37804	3826	11.26
91	Tinum	11421	12700	1279	11.20
79	Tekax	40547	45062	4515	11.14
83	Telchac Puerto	1726	1915	189	10.95
99	Uayma	3782	4191	409	10.81
76	Tecoh	16200	17939	1739	10.73
48	Maxcanú	21704	23991	2287	10.54
104	Yaxcabá	14802	16350	1548	10.46

MUN	NOM_MUN	Population2010	Population2020	PopulationChange	PercentageChange
96	Tizimín	73138	80672	7534	10.30
16	Chacsinkín	2818	3104	286	10.15
66	Santa Elena	3833	4220	387	10.10
39	Ixil	3803	4186	383	10.07
7	Cacalchén	6811	7490	679	9.97
40	Izamal	25980	28555	2575	9.91
90	Timucuy	6833	7503	670	9.81
98	Tzucacab	14011	15346	1335	9.53
53	Muna	12336	13494	1158	9.39
2	Acanceh	15337	16772	1435	9.36
20	Chicxulub Pueblo	4113	4497	384	9.34
45	Kopomá	2449	2677	228	9.31
4	Baca	5701	6195	494	8.67
54	Muxupip	2755	2990	235	8.53
37	Huhí	4841	5250	409	8.45
67	Seyé	9276	10053	777	8.38
81	Tekom	3100	3355	255	8.23
62	Sacalum	4589	4962	373	8.13
32	Espita	15571	16779	1208	7.76
34	Hocabá	6061	6514	453	7.47
14	Cuncunul	1595	1714	119	7.46
89	Ticul	37685	40495	2810	7.46
58	Peto	24159	25954	1795	7.43
23	Chocholá	4530	4863	333	7.35
93	Tixkokob	17176	18420	1244	7.24
103	Xocchel	3236	3451	215	6.64
97	Tunkás	3464	3684	220	6.35
69	Sotuta	8449	8967	518	6.13
6	Buctzotz	8637	9159	522	6.04
29	Dzilam González	5905	6240	335	5.67
95	Tixpéhual	5388	5690	302	5.61
5	Bokobá	2053	2167	114	5.55
64	Sanahcat	1619	1701	82	5.06
17	Chankom	4464	4686	222	4.97
22	Chikindzonot	4162	4363	201	4.83
74	Tahmek	3609	3774	165	4.57
57	Panabá	7461	7766	305	4.09
26	Dzemul	3489	3622	133	3.81
106	Yobaín	2137	2215	78	3.65
60	Quintana Roo	942	976	34	3.61
84	Temax	6817	7037	220	3.23
1	Abalá	6356	6550	194	3.05
24	Chumayel	3148	3244	96	3.05
77	Tekal de Venegas	2606	2683	77	2.95
27	Dzidzantún	8133	8345	212	2.61
68	Sinanché	3126	3206	80	2.56
78	Tekantó	3683	3747	64	1.74
31	Dzoncauich	2772	2818	46	1.66
70	Sucilá	3930	3971	41	1.04
12	Cenotillo	3701	3736	35	0.95
42	Kantunil	5502	5553	51	0.93
72	Suma	1876	1857	-19	-1.01

MUN	NOM_MUN	Population2010	Population2020	PopulationChange	PercentageChange
82	Telchac Pueblo	3557	3512	-45	-1.27
88	Teya	1977	1917	-60	-3.03
8	Calotmul	4095	3949	-146	-3.57
86	Tepakán	2226	2133	-93	-4.18
9	Cansahcab	4696	4466	-230	-4.90

Plot it in a bar graph, only the top 10 biggest changes, do it in percentage with respect to 2010. Increase the resolution of the plot.

```
library(ggplot2)
plot <- ggplot(population_changes %>%
         select(MUN, NOM_MUN, Population2010, Population2020, PopulationChange, PercentageChange) %>%
         arrange(desc(PercentageChange)) %>%
         head(10),
       aes(x = reorder(NOM_MUN, PercentageChange), y = PercentageChange)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(
    title = "Top 10 Population Changes in Yucatán Municipalities (2010-2020)",
    x = "Municipality",
    y = "Percentage Change (%)"
  ) +
  theme_minimal()
ggsave("population_changes_top_10.png", plot, width = 10, height = 6, dpi = 300)
print(plot)
```



Born in state changes. Save the top ten of the table to a file.

```
total_population_2010 <- (
  data_2010 %>%
  filter(MUN == 0, LOC == 0) %>%
  select(P_TOTAL)
  )$P_TOTAL

table <- kable(population_changes %>%
  mutate(
    PercentageChange = round((BornInStateChange / BornInState2010) * 100, 2),
    PercentageChangeToMunicipalityPop = round((BornInStateChange / Population2010) * 100, 2),
    PercentageChangeToTotalPop = round((BornInStateChange / total_population_2010) * 100, 2),
  ) %>%
  select(MUN, NOM_MUN, BornInState2010, BornInState2020, BornInStateChange, PercentageChange, Percentage
  arrange(desc(PercentageChangeToMunicipalityPop))-> population_changes_in_state,
  caption = "Population Changes Born in Yucatán (2010-2020)",
  digits = 2)
# Save the table to a file
write.csv(population_changes_in_state %>%
  select(MUN, NOM_MUN, BornInState2010, BornInState2020, BornInStateChange, PercentageChange, Percentage
  arrange(desc(PercentageChangeToMunicipalityPop)),
  "population_changes_born_in_state.csv", row.names = FALSE, )
print(table)
```

```
##
```

##
Table: Population Changes Born in Yucatán (2010-2020)
##
| MUN|NOM_MUN | BornInState2010| BornInState2020| BornInStateChange| PercentageChange| Perc
|---:|:----------|---------------:|---------------:|-----------------:|----------------:|-----
| 41|Kanasín | 69163| 120754| 51591| 74.59|
| 13|Conkal | 8660| 12431| 3771| 43.55|
| 73|Tahdziú | 4426| 5778| 1352| 30.55|
| 101|Umán | 47441| 61094| 13653| 28.78|
| 43|Kaua | 2713| 3310| 597| 22.01|
| 25|Dzán | 4812| 5835| 1023| 21.26|
| 49|Mayapán | 3252| 3903| 651| 20.02|
| 11|Celestún | 6251| 7555| 1304| 20.86|
| 3|Akil | 10139| 11949| 1810| 17.85|
| 92|Tixcacalcupul | 6575| 7638| 1063| 16.17|
| 59|Progreso | 47133| 55703| 8570| 18.18|
| 56|Oxkutzcab | 27945| 32576| 4631| 16.57|
| 21|Chichimilá | 7768| 8989| 1221| 15.72|
| 100|Ucú | 3370| 3894| 524| 15.55|
| 87|Tetiz | 4685| 5392| 707| 15.09|
| 10|Cantamayec | 2375| 2724| 349| 14.69|
| 19|Chemax | 32916| 37757| 4841| 14.71|
| 94|Tixmehuac | 4659| 5337| 678| 14.55|
| 46|Mama | 2818| 3225| 407| 14.44|
| 105|Yaxkukul | 2761| 3155| 394| 14.27|
| 38|Hunucmá | 29941| 34095| 4154| 13.87|
| 44|Kinchil | 6519| 7394| 875| 13.42|
| 61|Río Lagartos | 3290| 3741| 451| 13.71|
| 47|Maní | 5124| 5782| 658| 12.84|
| 71|Sudzal | 1626| 1835| 209| 12.85|
| 63|Samahil | 4975| 5570| 595| 11.96|
| 35|Hoctún | 5578| 6253| 675| 12.10|
| 55|Opichén | 6234| 6963| 729| 11.69|
| 102|Valladolid | 68496| 77088| 8592| 12.54|
| 15|Cuzamá | 4954| 5525| 571| 11.53|
| 75|Teabo | 6140| 6851| 711| 11.58|
| 85|Temozón | 14606| 16283| 1677| 11.48|
| 50|Mérida | 696800| 790526| 93726| 13.45|
| 80|Tekit | 9813| 10917| 1104| 11.25|
| 18|Chapab | 2995| 3334| 339| 11.32|
| 51|Mocochá | 2980| 3320| 340| 11.41|
| 36|Homún | 7188| 7975| 787| 10.95|
| 91|Tinum | 11042| 12237| 1195| 10.82|
| 79|Tekax | 39302| 43495| 4193| 10.67|
| 76|Tecoh | 16061| 17707| 1646| 10.25|
| 30|Dzitás | 3478| 3836| 358| 10.29|
| 16|Chacsinkín | 2795| 3080| 285| 10.20|
| 39|Ixil | 3740| 4117| 377| 10.08|
| 90|Timucuy | 6761| 7430| 669| 9.89|
| 33|Halachó | 18346| 20187| 1841| 10.03|
| 99|Uayma | 3753| 4114| 361| 9.62|
| 52|Motul | 33053| 36256| 3203| 9.69|
| 104|Yaxcabá | 14523| 15868| 1345| 9.26|
| 48|Maxcanú | 21383| 23335| 1952| 9.13|

```
##  |   7|Cacalchén        |          6721|          7332|           611|          9.09|
##  |   2|Acanceh          |         15118|         16471|          1353|          8.95|
##  |  65|San Felipe       |          1790|          1951|           161|          8.99|
##  |  62|Sacalum          |          4515|          4906|           391|          8.66|
##  |  66|Santa Elena      |          3761|          4083|           322|          8.56|
##  |  45|Kopomá           |          2418|          2616|           198|          8.19|
##  |  40|Izamal           |         25153|         27224|          2071|          8.23|
##  |  53|Muna             |         12041|         13001|           960|          7.97|
##  |  96|Tizimín          |         69665|         75295|          5630|          8.08|
##  |  67|Seyé             |          9185|          9886|           701|          7.63|
##  |  89|Ticul            |         36083|         38915|          2832|          7.85|
##  |  98|Tzucacab         |         13504|         14551|          1047|          7.75|
##  |  54|Muxupip          |          2707|          2906|           199|          7.35|
##  |  37|Huhí             |          4799|          5145|           346|          7.21|
##  |  28|Dzilam de Bravo  |          2273|          2445|           172|          7.57|
##  |  14|Cuncunul         |          1539|          1649|           110|          7.15|
##  |   4|Baca             |          5450|          5822|           372|          6.83|
##  |  34|Hocabá           |          6038|          6434|           396|          6.56|
##  | 103|Xocchel          |          3200|          3408|           208|          6.50|
##  |  32|Espita           |         15213|         16155|           942|          6.19|
##  |   5|Bokobá           |          2020|          2144|           124|          6.14|
##  |  58|Peto             |         23190|         24626|          1436|          6.19|
##  |  83|Telchac Puerto   |          1617|          1718|           101|          6.25|
##  |   6|Buctzotz         |          8417|          8905|           488|          5.80|
##  |  69|Sotuta           |          8389|          8860|           471|          5.61|
##  |  93|Tixkokob         |         16778|         17721|           943|          5.62|
##  |  20|Chicxulub Pueblo |          4016|          4237|           221|          5.50|
##  |  23|Chocholá         |          4416|          4657|           241|          5.46|
##  |  64|Sanahcat         |          1609|          1689|            80|          4.97|
##  |  29|Dzilam González  |          5682|          5973|           291|          5.12|
##  |  95|Tixpéhual        |          5281|          5540|           259|          4.90|
##  |  81|Tekom            |          3046|          3193|           147|          4.83|
##  |  22|Chikindzonot     |          4101|          4280|           179|          4.36|
##  |  17|Chankom          |          4412|          4564|           152|          3.45|
##  |  74|Tahmek           |          3551|          3674|           123|          3.46|
##  |  78|Tekantó          |          3578|          3698|           120|          3.35|
##  |  97|Tunkás           |          3366|          3477|           111|          3.30|
##  |   1|Abalá            |          6322|          6518|           196|          3.10|
##  |  26|Dzemul           |          3383|          3487|           104|          3.07|
##  |  24|Chumayel         |          3110|          3195|            85|          2.73|
##  |  68|Sinanché         |          2974|          3057|            83|          2.79|
##  |  77|Tekal de Venegas |          2519|          2583|            64|          2.54|
##  |  57|Panabá           |          7229|          7370|           141|          1.95|
##  |  84|Temax            |          6770|          6898|           128|          1.89|
##  | 106|Yobaín           |          2047|          2072|            25|          1.22|
##  |  27|Dzidzantún       |          7813|          7854|            41|          0.52|
##  |  70|Sucilá           |          3766|          3778|            12|          0.32|
##  |  42|Kantunil         |          5381|          5380|            -1|         -0.02|
##  |  31|Dzoncauich       |          2713|          2710|            -3|         -0.11|
##  |  60|Quintana Roo     |           925|           921|            -4|         -0.43|
##  |  72|Suma             |          1820|          1799|           -21|         -1.15|
##  |  82|Telchac Pueblo   |          3383|          3330|           -53|         -1.57|
##  |  12|Cenotillo        |          3607|          3551|           -56|         -1.55|
##  |  88|Teya             |          1943|          1884|           -59|         -3.04|
```

```
## |   8|Calotmul         |            3888|            3706|             -182|           -4.68|
## |  86|Tepakán          |            2164|            2047|             -117|           -5.41|
## |   9|Cansahcab        |            4567|            4294|             -273|           -5.98|
```

Born in other state changes.

```
table <- kable(population_changes %>%
  mutate(
    PercentageChange = round(((BornInOtherStateChange / BornInOtherState2010) * 100), 2),
    PercentageChangeToMunicipalityPop = round((BornInOtherStateChange / Population2010) * 100, 2),
    PercentageChangeToTotalPop = round((BornInOtherStateChange / total_population_2010) * 100, 2),
  ) %>%
  select(MUN, NOM_MUN, BornInOtherState2010, BornInOtherState2020, BornInOtherStateChange, PercentageCha
  arrange(desc(PercentageChangeToMunicipalityPop)) -> population_changes_out_state,
  caption = "Population Changes Born in Other States (2010-2020)",
  digits = 2)

  # Save the table to a file
write.csv(population_changes_out_state %>%
  select(MUN, NOM_MUN, BornInOtherState2010, BornInOtherState2020, BornInOtherStateChange, PercentageCha
  arrange(desc(PercentageChangeToMunicipalityPop)),
  "population_changes_born_in_other_state.csv", row.names = FALSE)
print(table)
```

```
##
##
## Table: Population Changes Born in Other States (2010-2020)
##
## | MUN|NOM_MUN          | BornInOtherState2010| BornInOtherState2020| BornInOtherStateChange| Percenta
## |---:|:----------------|--------------------:|--------------------:|----------------------:|--------
## |  13|Conkal           |                382|                3950|                   3568|
## |  41|Kanasín          |               7940|               20552|                  12612|
## |  28|Dzilam de Bravo  |                184|                 484|                    300|
## |  50|Mérida           |             112871|              188353|                  75482|
## | 101|Umán             |               3376|                7881|                   4505|
## |  65|San Felipe       |                 44|                 165|                    121|
## |  59|Progreso         |               6081|                9198|                   3117|
## |  83|Telchac Puerto   |                 89|                 168|                     79|
## |  20|Chicxulub Pueblo |                 74|                 251|                    177|
## |  11|Celestún         |                507|                 783|                    276|
## |  60|Quintana Roo     |                 15|                  52|                     37|
## |  81|Tekom            |                 46|                 159|                    113|
## | 102|Valladolid       |               5392|                8020|                   2628|
## |  30|Dzitás           |                 58|                 172|                    114|
## |  21|Chichimilá       |                159|                 404|                    245|
## |  71|Sudzal           |                 59|                 111|                     52|
## |  96|Tizimín          |               3166|                5144|                   1978|
## |  12|Cenotillo        |                 48|                 147|                     99|
## |  97|Tunkás           |                 86|                 177|                     91|
## | 100|Ucú              |                 68|                 153|                     85|
## |  92|Tixcacalcupul    |                 84|                 238|                    154|
## | 106|Yobaín           |                 87|                 136|                     49|
## |  57|Panabá           |                216|                 386|                    170|
```

```
## |  98|Tzucacab         |  489|  784|  295|
## |   4|Baca             |  246|  366|  120|
## |  23|Chocholá         |  109|  201|   92|
## |  19|Chemax           |  489| 1162|  673|
## |  31|Dzoncauich       |   47|  102|   55|
## |  40|Izamal           |  724| 1234|  510|
## |  27|Dzidzantún       |  294|  440|  146|
## |  33|Halachó          |  707| 1048|  341|
## |  61|Río Lagartos     |  134|  195|   61|
## |  52|Motul            |  775| 1353|  578|
## |  43|Kaua             |   45|   92|   47|
## |  93|Tixkokob         |  344|  630|  286|
## |  48|Maxcanú          |  268|  622|  354|
## |  53|Muna             |  229|  427|  198|
## |  17|Chankom          |   48|  119|   71|
## |  58|Peto             |  869| 1236|  367|
## |  66|Santa Elena      |   61|  119|   58|
## |  85|Temozón          |  165|  385|  220|
## |  99|Uayma            |   20|   76|   56|
## | 104|Yaxcabá          |  261|  477|  216|
## |  32|Espita           |  335|  558|  223|
## |  44|Kinchil          |   42|  130|   88|
## |  29|Dzilam González  |  173|  247|   74|
## |  45|Kopomá           |   29|   59|   30|
## |  74|Tahmek           |   52|   95|   43|
## |   3|Akil             |  189|  305|  116|
## |  72|Suma             |   37|   58|   21|
## |  38|Hunucmá          |  640|  982|  342|
## |  95|Tixpéhual        |   82|  142|   60|
## |  42|Kantunil         |  104|  164|   60|
## |  55|Opichén          |   48|  115|   67|
## |  84|Temax            |   44|  115|   71|
## |  34|Hocabá           |   14|   75|   61|
## |   9|Cansahcab        |  111|  158|   47|
## |   8|Calotmul         |  198|  237|   39|
## |  67|Seyé             |   68|  155|   87|
## |  47|Maní             |   87|  135|   48|
## |  36|Homún            |   43|  108|   65|
## |  49|Mayapán          |    7|   34|   27|
## |  26|Dzemul           |   98|  127|   29|
## |  22|Chikindzonot     |   47|   81|   34|
## |  54|Muxupip          |   38|   60|   22|
## | 103|Xocchel          |   16|   42|   26|
## |  51|Mocochá          |   82|  106|   24|
## |  18|Chapab           |   25|   48|   23|
## |  37|Huhí             |   34|   70|   36|
## |  25|Dzán             |  107|  141|   34|
## |  39|Ixil             |   41|   67|   26|
## |  80|Tekit            |   37|  101|   64|
## |  73|Tahdziú          |   17|   45|   28|
## |  86|Tepakán          |   48|   62|   14|
## |  76|Tecoh            |  118|  218|  100|
## |  70|Sucilá           |  161|  185|   24|
## |   2|Acanceh          |  191|  283|   92|
```

```
## |  91|Tinum            |                347|                414|                  67|
## |   7|Cacalchén        |                 78|                118|                  40|
## |  79|Tekax            |               1166|               1403|                 237|
## |  35|Hoctún           |                 90|                123|                  33|
## |  77|Tekal de Venegas |                 85|                100|                  15|
## |  15|Cuzamá           |                  6|                 34|                  28|
## |   6|Buctzotz         |                172|                219|                  47|
## |  24|Chumayel         |                 28|                 45|                  17|
## |  87|Tetiz            |                 33|                 58|                  25|
## |  69|Sotuta           |                 50|                 94|                  44|
## |  88|Teya             |                 22|                 32|                  10|
## |  94|Tixmehuac        |                 85|                106|                  21|
## |  14|Cuncunul         |                 55|                 62|                   7|
## |  82|Telchac Pueblo   |                144|                159|                  15|
## |  90|Timucuy          |                 44|                 72|                  28|
## | 105|Yaxkukul         |                 97|                107|                  10|
## |  56|Oxkutzcab        |                975|               1052|                  77|
## |  64|Sanahcat         |                  7|                 11|                   4|
## |  46|Mama             |                 58|                 63|                   5|
## |  16|Chacsinkín       |                 20|                 24|                   4|
## |  75|Teabo            |                 52|                 60|                   8|
## |  63|Samahil          |                 27|                 33|                   6|
## |  10|Cantamayec       |                 28|                 30|                   2|
## |   1|Abalá            |                 25|                 28|                   3|
## |  89|Ticul            |               1468|               1419|                 -49|
## |  62|Sacalum          |                 69|                 54|                 -15|
## |   5|Bokobá           |                 28|                 20|                  -8|
## |  68|Sinanché         |                141|                118|                 -23|
## |  78|Tekantó          |                 91|                 47|                 -44|
```

Let's visualize the changes in population

Lets create a plot with the changes in population on top of a map.

```
library(sf)
```

```
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
yucatan_map <- st_read("/Users/ecastillo/Documents/coding/analyzing_merida_data/yucatan.geojson")
```

```
## Reading layer 'Yucatán' from data source
##   '/Users/ecastillo/Documents/coding/analyzing_merida_data/yucatan.geojson'
##   using driver 'GeoJSON'
## Simple feature collection with 106 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -92.3263 ymin: 19.55117 xmax: -87.53315 ymax: 22.58595
## Geodetic CRS:  WGS 84
```

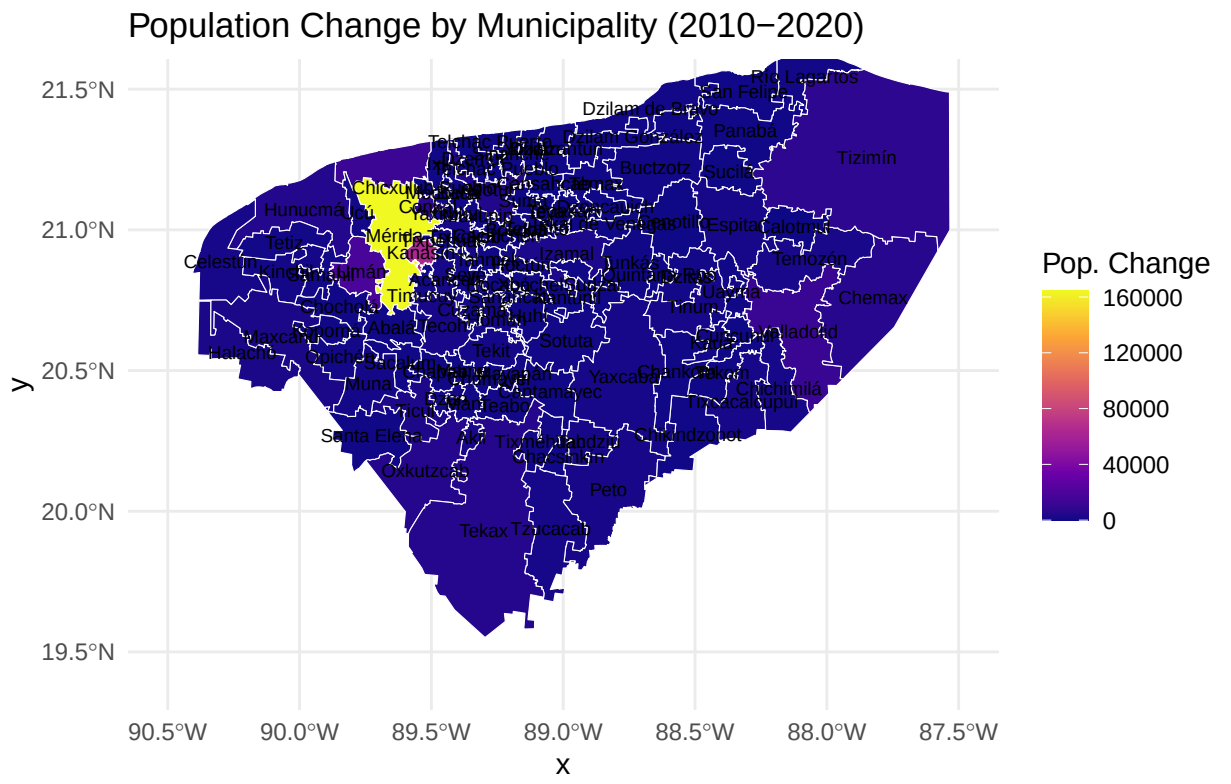
```
yucatan_map <- yucatan_map %>%
  mutate(CVE_MUN = as.numeric(CVE_MUN)) # Ensure CVE_MUN is numeric for joining
```

```r
map_data <- yucatan_map %>%
  left_join(population_changes, by = c("CVE_MUN" = "MUN"))

# Calculate centroids
map_data_centroids <- st_centroid(map_data)
```

```
## Warning: st_centroid assumes attributes are constant over geometries
```

```r
map_plot = ggplot(data = map_data) +
  geom_sf(aes(fill = PopulationChange), color = "white") +
  geom_sf_text(data = map_data_centroids, aes(label = NOMGEO), size = 2.5, color = "black") +
  scale_fill_viridis_c(option = "C") +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 21.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "Population Change by Municipality (2010-2020)",
    fill = "Pop. Change"
  )
print(map_plot)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```



```r
# save to png
ggsave("population_change_map.png", map_plot, width = 10, height = 6, dpi = 300)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

We can see that the biggest increases in population are around the areas of Merida, Valladolid and Tizimin.

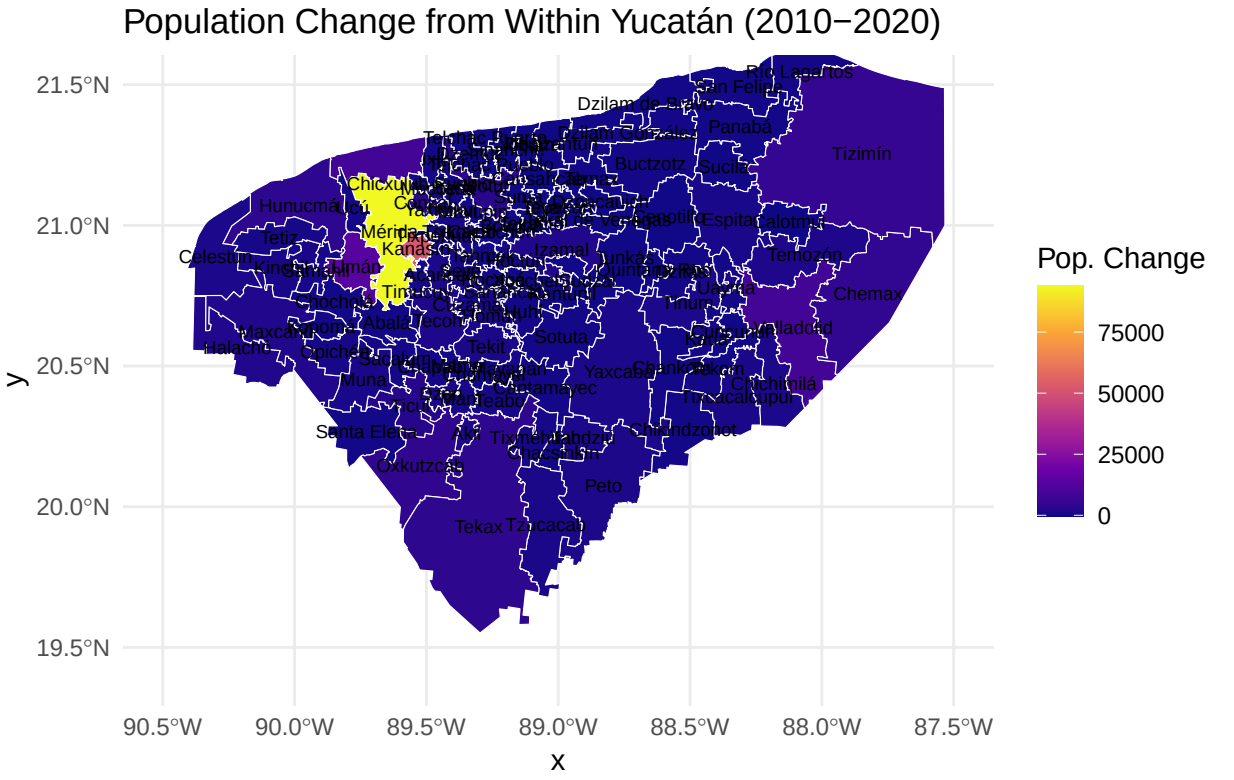
Lets see if we can find the same pattern in the influx of people from outside the state.

Let's visualize the changes in population born in Yucatan state and in other states.

```r
# we need to get the max and min values for the fill scale
global_min <- min(map_data$BornInStateChange, map_data$BornInOtherStateChange, na.rm = TRUE)
global_max <- max(map_data$BornInStateChange, map_data$BornInOtherStateChange, na.rm = TRUE)
```

```r
map_plot = ggplot(data = map_data) +
  geom_sf(aes(fill = BornInStateChange), color = "white") +
  geom_sf_text(data = map_data_centroids, aes(label = NOMGEO), size = 2.5, color = "black") +
  scale_fill_viridis_c(option = "C", limits = c(global_min, global_max)) +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 21.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "Population Change from Within Yucatán (2010-2020)",
    fill = "Pop. Change"
  )
print(map_plot)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

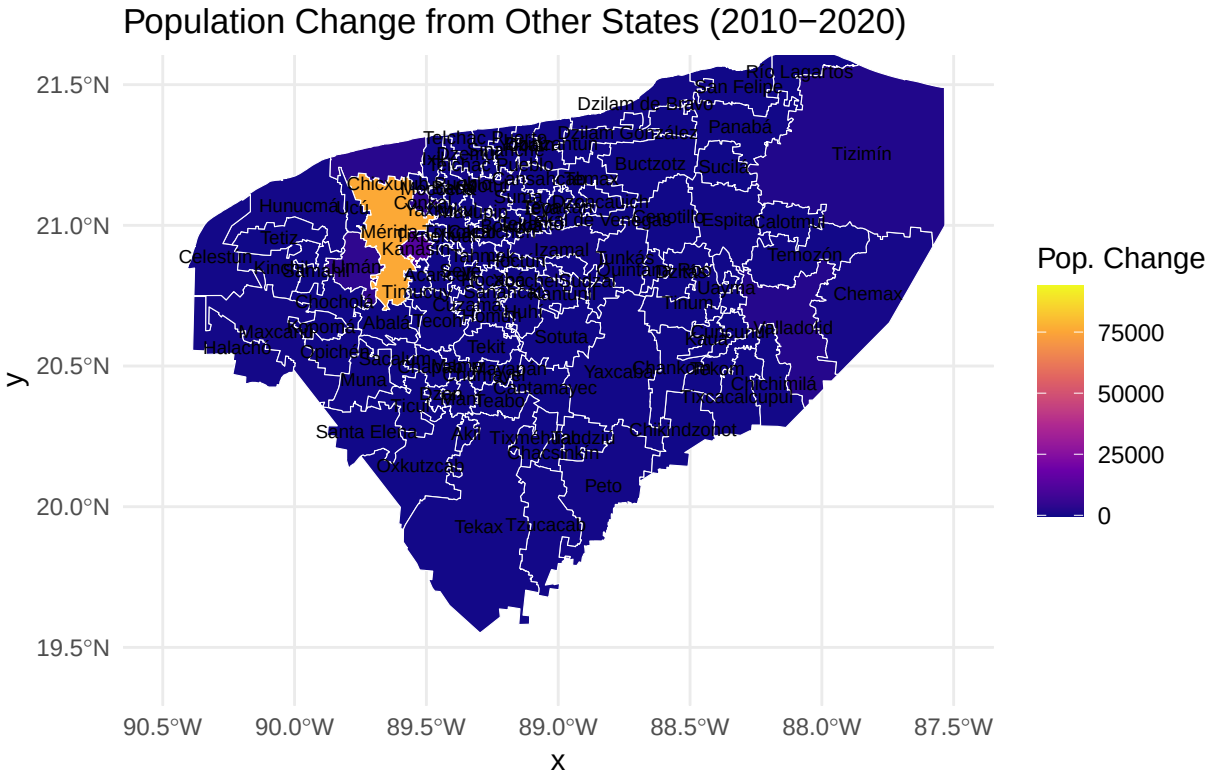



```
# save to png
ggsave("population_change_born_in_state_map.png", map_plot, width = 10, height = 6, dpi = 300)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

```
map_plot = ggplot(data = map_data) +
  geom_sf(aes(fill = BornInOtherStateChange), color = "white") +
  geom_sf_text(data = map_data_centroids, aes(label = NOMGEO), size = 2.5, color = "black") +
  scale_fill_viridis_c(option = "C", limits = c(global_min, global_max)) +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 21.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "Population Change from Other States (2010-2020)",
    fill = "Pop. Change"
  )
print(map_plot)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```




```
# save to png
ggsave("population_change_born_in_other_state_map.png", map_plot, width = 10, height = 6, dpi = 300)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

The trend is similar.

Let's see how the Indigenous population has changed

```
indigenous_population_changes <- biggest_cities_2020 %>%
  select(MUN, NOM_MUN, P3YM_HLI, P3HLINHE, P3HLI_HE, P5_HLI, P5_HLI_NHE, P5_HLI_HE) %>%
  rename(
    IndigenousPopulation2020 = P3YM_HLI,
    IndigenousNotSpanish2020 = P3HLINHE,
    IndigenousAndSpanish2020 = P3HLI_HE,
    IndigenousPopulation5Plus2020 = P5_HLI,
    IndigenousNotSpanish5Plus2020 = P5_HLI_NHE,
    IndigenousAndSpanish5Plus2020 = P5_HLI_HE
  ) %>%
  mutate(
    MUN = as.numeric(MUN),
    IndigenousPopulation2020 = as.numeric(IndigenousPopulation2020),
```

```
    IndigenousNotSpanish2020 = as.numeric(IndigenousNotSpanish2020),
    IndigenousAndSpanish2020 = as.numeric(IndigenousAndSpanish2020),
    IndigenousPopulation5Plus2020 = as.numeric(IndigenousPopulation5Plus2020),
    IndigenousNotSpanish5Plus2020 = as.numeric(IndigenousNotSpanish5Plus2020),
    IndigenousAndSpanish5Plus2020 = as.numeric(IndigenousAndSpanish5Plus2020)
  ) %>%
  left_join(
    biggest_cities_2010 %>%
      select(MUN, P3YM_HLI, P3HLINHE, P3HLI_HE, P5_HLI, P5_HLI_NHE, P5_HLI_HE) %>%
      rename(
        IndigenousPopulation2010 = P3YM_HLI,
        IndigenousNotSpanish2010 = P3HLINHE,
        IndigenousAndSpanish2010 = P3HLI_HE,
        IndigenousPopulation5Plus2010 = P5_HLI,
        IndigenousNotSpanish5Plus2010 = P5_HLI_NHE,
        IndigenousAndSpanish5Plus2010 = P5_HLI_HE
      ) %>%
      mutate(
        MUN = as.numeric(MUN),
        IndigenousPopulation2010 = as.numeric(IndigenousPopulation2010),
        IndigenousNotSpanish2010 = as.numeric(IndigenousNotSpanish2010),
        IndigenousAndSpanish2010 = as.numeric(IndigenousAndSpanish2010),
        IndigenousPopulation5Plus2010 = as.numeric(IndigenousPopulation5Plus2010),
        IndigenousNotSpanish5Plus2010 = as.numeric(IndigenousNotSpanish5Plus2010),
        IndigenousAndSpanish5Plus2010 = as.numeric(IndigenousAndSpanish5Plus2010)
      ),
    by = "MUN"
  ) %>%
  mutate(
    MUN = as.numeric(MUN),
    IndigenousPopulationChange = IndigenousPopulation2020 - IndigenousPopulation2010,
    IndigenousNotSpanishChange = IndigenousNotSpanish2020 - IndigenousNotSpanish2010,
    IndigenousAndSpanishChange = IndigenousAndSpanish2020 - IndigenousAndSpanish2010,
    IndigenousPopulation5PlusChange = IndigenousPopulation5Plus2020 - IndigenousPopulation5Plus2010,
    IndigenousNotSpanish5PlusChange = IndigenousNotSpanish5Plus2020 - IndigenousNotSpanish5Plus2010,
    IndigenousAndSpanish5PlusChange = IndigenousAndSpanish5Plus2020 - IndigenousAndSpanish5Plus2010
  ) %>%
  arrange(desc(IndigenousPopulationChange))
# Display the Indigenous population changes
kable(indigenous_population_changes %>%
  select(MUN, NOM_MUN, IndigenousPopulation2010, IndigenousPopulation2020, IndigenousPopulationChange),
  caption = "Indigenous Population Changes in Yucatán Municipalities (2010-2020)",
  digits = 2)
```

Table 2: Indigenous Population Changes in Yucatán Municipalities (2010–2020)

MUN	NOM_MUN	IndigenousPopulation2010	IndigenousPopulation2020	IndigenousPopulationChange
41	Kanasín	13555	16481	2926
102	Valladolid	38393	40685	2292
19	Chemax	27906	29911	2005
73	Tahdziú	3982	5151	1169

MUN	NOM_MUN	IndigenousPopulation2010	IndigenousPopulation2020	IndigenousPopulationChange
92	Tixcacalcupul	5758	6409	651
21	Chichimilá	6665	7220	555
49	Mayapán	2939	3426	487
85	Temozón	11307	11655	348
94	Tixmehuac	3951	4283	332
25	Dzán	3265	3466	201
43	Kaua	2080	2271	191
30	Dzitás	1707	1778	71
16	Chacsinkín	2519	2581	62
71	Sudzal	911	938	27
10	Cantamayec	1979	2001	22
22	Chikindzonot	3811	3831	20
13	Conkal	1462	1474	12
99	Uayma	3153	3155	2
83	Telchac Puerto	182	183	1
28	Dzilam de Bravo	106	103	-3
61	Río Lagartos	400	391	-9
105	Yaxkukul	307	296	-11
65	San Felipe	143	131	-12
47	Maní	4018	3998	-20
103	Xocchel	1582	1562	-20
66	Santa Elena	2923	2901	-22
3	Akil	5902	5875	-27
35	Hoctún	2424	2394	-30
97	Tunkás	1653	1612	-41
51	Mocochá	385	340	-45
55	Opichén	4398	4350	-48
100	Ucú	809	752	-57
11	Celestún	444	384	-60
106	Yobaín	420	360	-60
104	Yaxcabá	10143	10082	-61
14	Cuncunul	1189	1126	-63
56	Oxkutzcab	17142	17078	-64
46	Mama	2000	1933	-67
5	Bokobá	765	695	-70
64	Sanahcat	793	723	-70
54	Muxupip	1043	965	-78
72	Suma	609	525	-84
26	Dzemul	764	679	-85
60	Quintana Roo	381	291	-90
76	Tecoh	7590	7490	-100
18	Chapab	2067	1958	-109
75	Teabo	4649	4538	-111
82	Telchac Pueblo	462	349	-113
15	Cuzamá	2722	2606	-116
81	Tekom	2625	2505	-120
39	Ixil	532	401	-131
87	Tetiz	1752	1616	-136
45	Kopomá	1084	944	-140
12	Cenotillo	1261	1116	-145
24	Chumayel	2620	2475	-145
62	Sacalum	2562	2408	-154

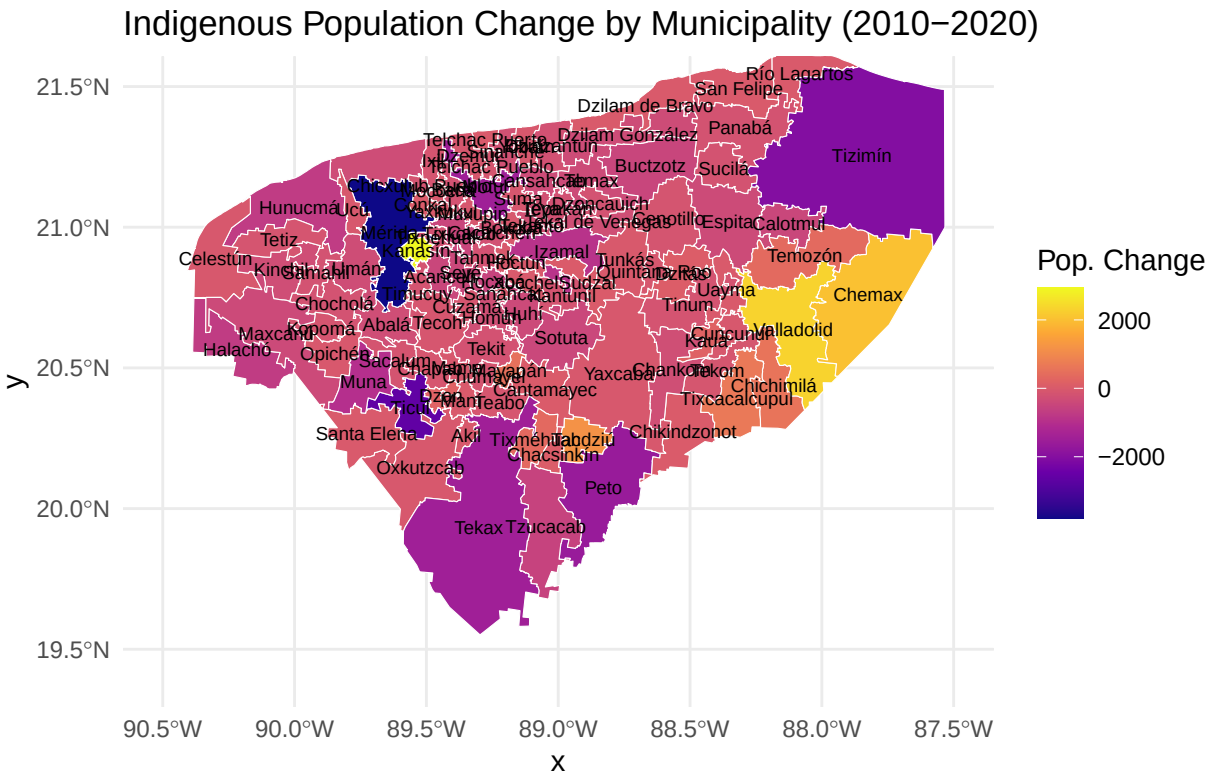
MUN	NOM_MUN	IndigenousPopulation2010	IndigenousPopulation2020	IndigenousPopulationChange
91	Tinum	6978	6816	-162
90	Timucuy	5575	5410	-165
20	Chicxulub Pueblo	471	305	-166
42	Kantunil	2129	1961	-168
70	Sucilá	1727	1554	-173
57	Panabá	1791	1606	-185
31	Dzoncauich	1472	1286	-186
4	Baca	971	782	-189
63	Samahil	1925	1733	-192
23	Chocholá	1313	1121	-192
74	Tahmek	1989	1789	-200
80	Tekit	4675	4474	-201
95	Tixpéhual	1428	1220	-208
27	Dzidzantún	660	439	-221
7	Cacalchén	1612	1391	-221
86	Tepakán	1389	1159	-230
68	Sinanché	829	598	-231
88	Teya	1099	865	-234
78	Tekantó	1301	1044	-257
36	Homún	3763	3473	-290
77	Tekal de Venegas	1346	1056	-290
17	Chankom	3843	3552	-291
44	Kinchil	2134	1821	-313
1	Abalá	3799	3484	-315
101	Umán	10862	10541	-321
59	Progreso	2957	2617	-340
29	Dzilam González	970	619	-351
9	Cansahcab	1101	739	-362
32	Espita	8604	8219	-385
84	Temax	1817	1430	-387
48	Maxcanú	8752	8361	-391
37	Huhí	2148	1749	-399
67	Seyé	2315	1876	-439
69	Sotuta	4301	3843	-458
8	Calotmul	2280	1814	-466
2	Acanceh	4910	4442	-468
93	Tixkokob	2457	1971	-486
6	Buctzotz	2274	1784	-490
34	Hocabá	3582	3051	-531
98	Tzucacab	7305	6750	-555
33	Halachó	11153	10437	-716
38	Hunucmá	6895	6162	-733
40	Izamal	9155	8228	-927
53	Muna	5553	4544	-1009
79	Tekax	23556	22086	-1470
52	Motul	9109	7630	-1479
58	Peto	14152	12526	-1626
96	Tizimín	28213	26185	-2028
89	Ticul	15266	12658	-2608

MUN	NOM_MUN	IndigenousPopulation2010	IndigenousPopulation2020	IndigenousPopulationChange
50	Mérida	74827	71040	-3787

Let's visualize the changes in Indigenous population

```
map_data <- yucatan_map %>%
  left_join(indigenous_population_changes, by = c("CVE_MUN" = "MUN"))
map_plot <- ggplot(data = map_data) +
  geom_sf(aes(fill = IndigenousPopulationChange), color = "white") +
  geom_sf_text(data = map_data_centroids, aes(label = NOMGEO), size = 2.5, color = "black") +
  scale_fill_viridis_c(option = "C") +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 21.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "Indigenous Population Change by Municipality (2010-2020)",
    fill = "Pop. Change"
  )
print(map_plot)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```



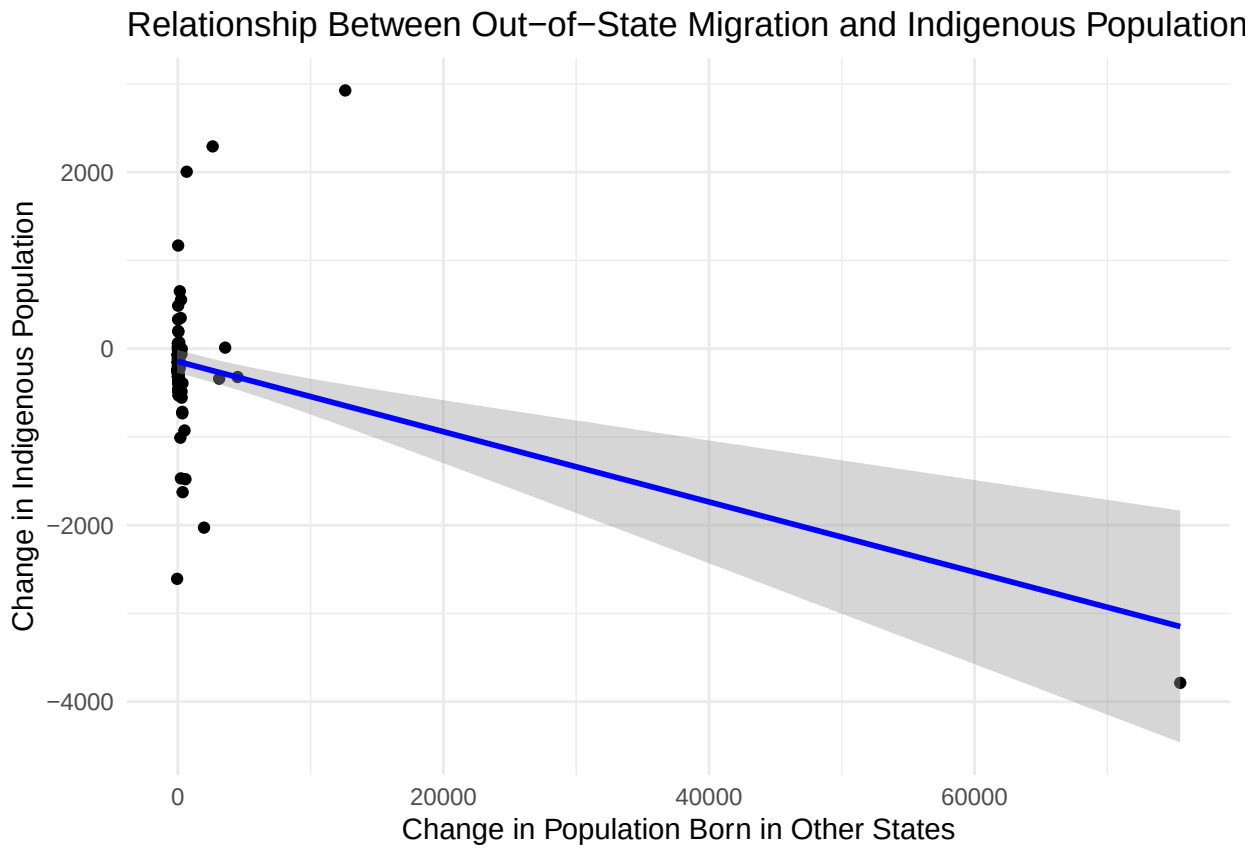
```r
# save to png
ggsave("indigenous_population_change_map.png", map_plot, width = 10, height = 6, dpi = 300)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

We can see that there has been a decrease in the Indigenous population in Merida. A municipality that saw a significant increase is Kanasin, which sits next to Merida. Kanasin is singificantly more affordable than Merida, so it is likely that indigenous people are moving there.

```r
ggplot(population_changes %>%
         left_join(indigenous_population_changes, by = "MUN"),
       aes(x = BornInOtherStateChange, y = IndigenousPopulationChange)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "Relationship Between Out-of-State Migration and Indigenous Population Change",
    x = "Change in Population Born in Other States",
    y = "Change in Indigenous Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



This plot shows the relationship between the change in population born in other states and the change in

Indigenous population. The trend line suggests that as more people from other states move to Yucatán, there is a corresponding decrease in the Indigenous population. But it has one outlier that is skewing the results, so we will first see what it is and then we will remove it and replot it.

```
# Identify the outlier
# join, but keep only one NOM_MUN column
population_changes %>%
  left_join(indigenous_population_changes, by = "MUN") %>%
  mutate(
    ratio = BornInOtherStateChange / IndigenousPopulationChange
  ) %>%
  arrange(desc(ratio)) %>%
  rename(
    NOM_MUN = NOM_MUN.x
  ) %>%
  select(MUN, NOM_MUN, BornInOtherStateChange, IndigenousPopulationChange, ratio) %>%
  tail(10)
```

```
##     MUN         NOM_MUN BornInOtherStateChange IndigenousPopulationChange
## 97   66     Santa Elena                    58                         -22
## 98  104         Yaxcabá                   216                         -61
## 99    3            Akil                   116                         -27
## 100  11        Celestún                   276                         -60
## 101  61    Río Lagartos                    61                          -9
## 102  59        Progreso                  3117                        -340
## 103  65      San Felipe                   121                         -12
## 104 101            Umán                  4505                        -321
## 105  50          Mérida                 75482                       -3787
## 106  28 Dzilam de Bravo                   300                          -3
##           ratio
## 97    -2.636364
## 98    -3.540984
## 99    -4.296296
## 100   -4.600000
## 101   -6.777778
## 102   -9.167647
## 103  -10.083333
## 104  -14.034268
## 105  -19.931872
## 106 -100.000000
```

Conkal is the outlier, it has a very high increase in population born in other states, but a very low increase in Indigenous population. This is likely due to the fact that Conkal is one of the most popular places to move into Merida, it is just next to the city of Merida, a lot of new housing developments both luxurious and affordable.

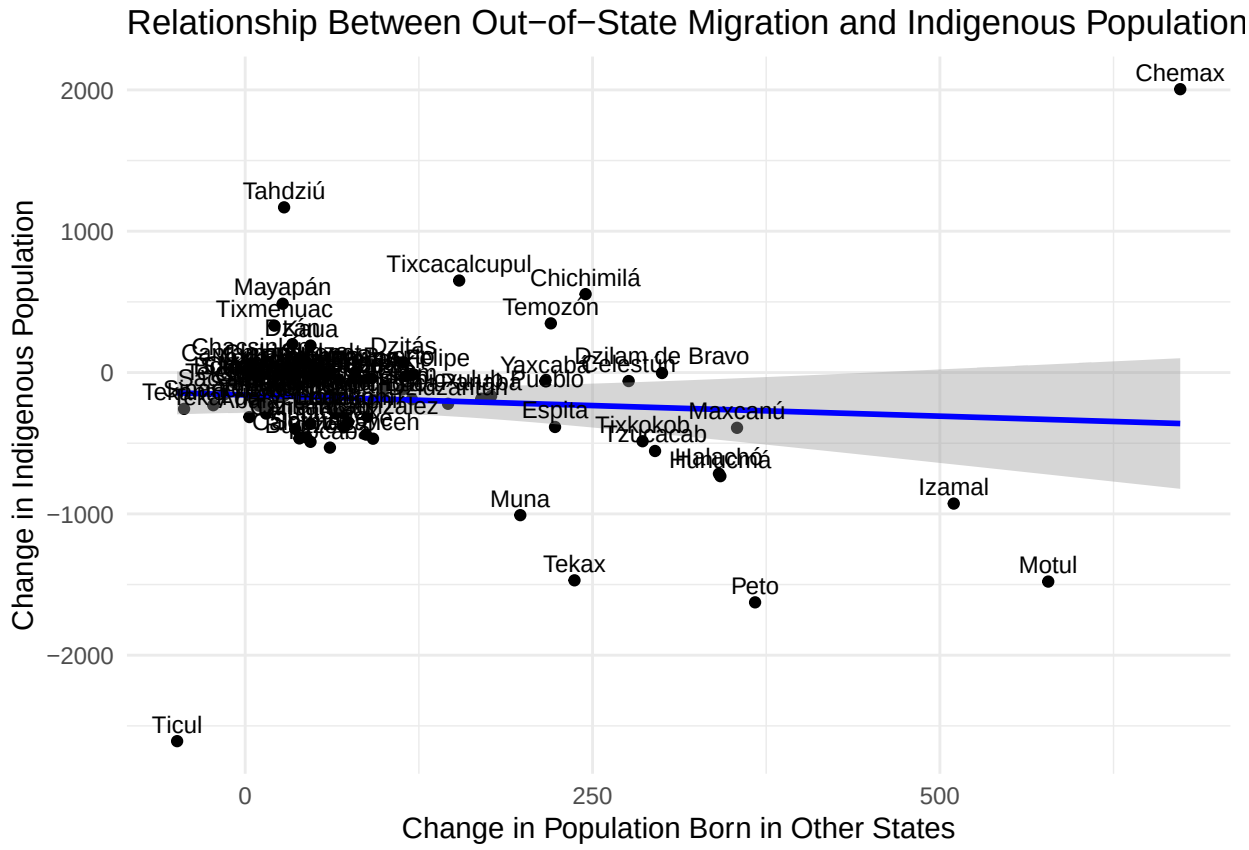
Merida is also skewing the results. Kanasin as well.

Let's remove it to see the plot again more clearly. Add names to the points.

```
population_changes %>%
  left_join(indigenous_population_changes, by = "MUN") %>%
  filter(MUN != 13, MUN != 41, MUN != 50, MUN !=59, MUN != 101, MUN != 102, MUN != 96) %>% # remove Mer
  # remove Conkal, Merida, Dzilam de Bravo
```

```
  ggplot(aes(x = BornInOtherStateChange, y = IndigenousPopulationChange)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  geom_text(aes(label = NOM_MUN.x), vjust = -0.5, size = 3) +
  labs(
    title = "Relationship Between Out-of-State Migration and Indigenous Population Change (Without majo
    x = "Change in Population Born in Other States",
    y = "Change in Indigenous Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

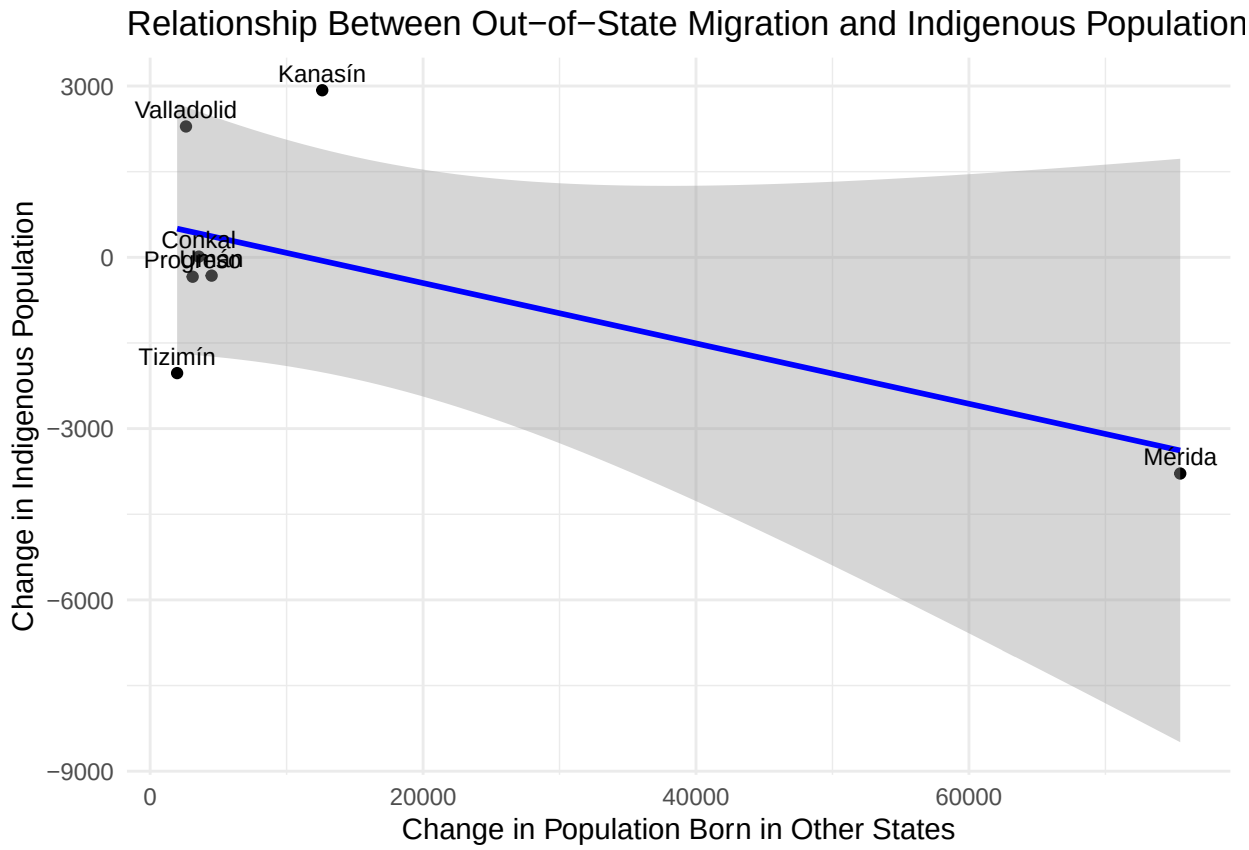


The plot with the removed entries.

```
population_changes %>%
  left_join(indigenous_population_changes, by = "MUN") %>%
  filter(MUN == 13 | MUN == 41 | MUN == 50 | MUN ==59 | MUN == 101 | MUN == 102 | MUN == 96) %>%
  ggplot(aes(x = BornInOtherStateChange, y = IndigenousPopulationChange)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  geom_text(aes(label = NOM_MUN.x), vjust = -0.5, size = 3) +
  labs(
    title = "Relationship Between Out-of-State Migration and Indigenous Population Change (Only Major c
    x = "Change in Population Born in Other States",
```

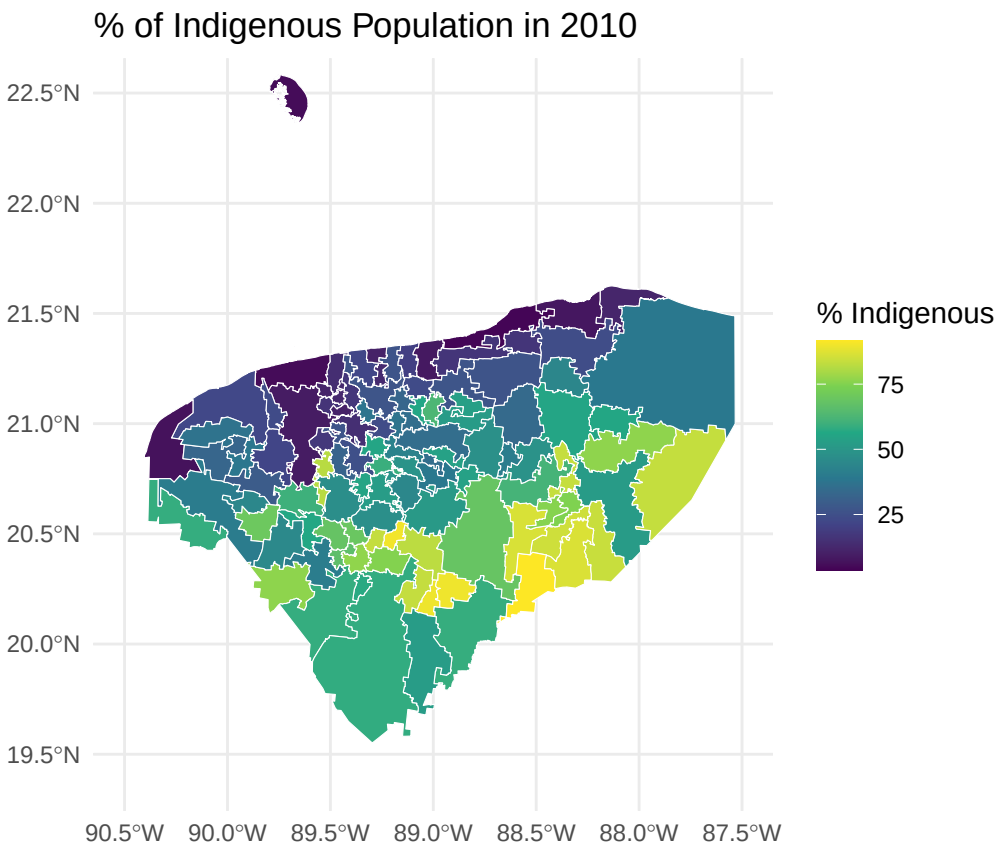
```
    y = "Change in Indigenous Population"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



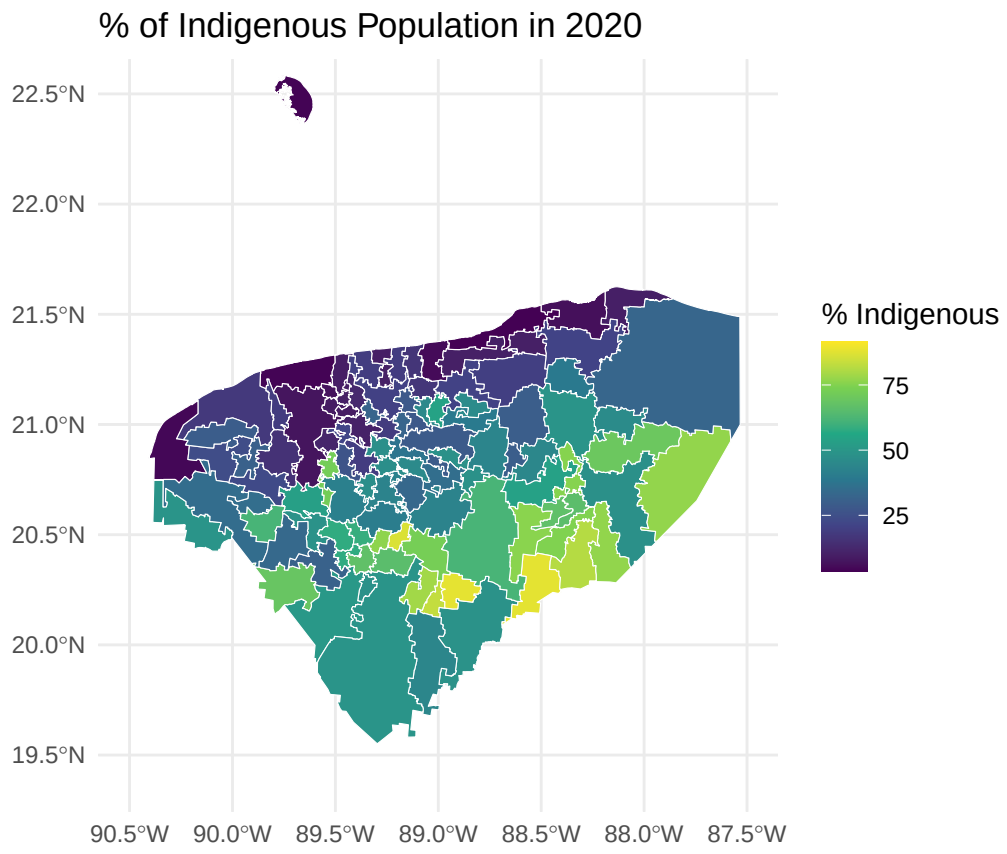

```
map_data <- yucatan_map %>%
  mutate(CVE_MUN = as.numeric(CVE_MUN)) %>%
  left_join(indigenous_population_changes, by = c("CVE_MUN" = "MUN")) %>%
  left_join(population_changes, by = c("CVE_MUN" = "MUN")) %>%
  mutate(
    IndigenousShare2010 = IndigenousPopulation2010 / Population2010 * 100,
    IndigenousShare2020 = IndigenousPopulation2020 / Population2020 * 100
  )

# get the min and max values for the fill scale
global_min <- min(map_data$IndigenousShare2010, map_data$IndigenousShare2020, na.rm = TRUE)
global_max <- max(map_data$IndigenousShare2010, map_data$IndigenousShare2020, na.rm = TRUE)

# Now it has geometry + your data, so this will work
map_plot <- ggplot(map_data) +
  geom_sf(aes(fill = IndigenousShare2010), color = "white") +
  scale_fill_viridis_c(option = "D", limits = c(global_min, global_max)) +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 22.5)) +  # Adjust as needed
```

```
  theme_minimal() +
  labs(
    title = "% of Indigenous Population in 2010",
    fill = "% Indigenous"
  )
print(map_plot)
```




```
ggsave("indigenous_population_share_2010.png", map_plot, width = 10, height = 6, dpi = 300)
```

```
# Now it has geometry + your data, so this will work
map_plot <- ggplot(map_data) +
  geom_sf(aes(fill = IndigenousShare2020), color = "white") +
  scale_fill_viridis_c(option = "D", limits = c(global_min, global_max)) +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 22.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "% of Indigenous Population in 2020",
    fill = "% Indigenous"
  )
print(map_plot)
```



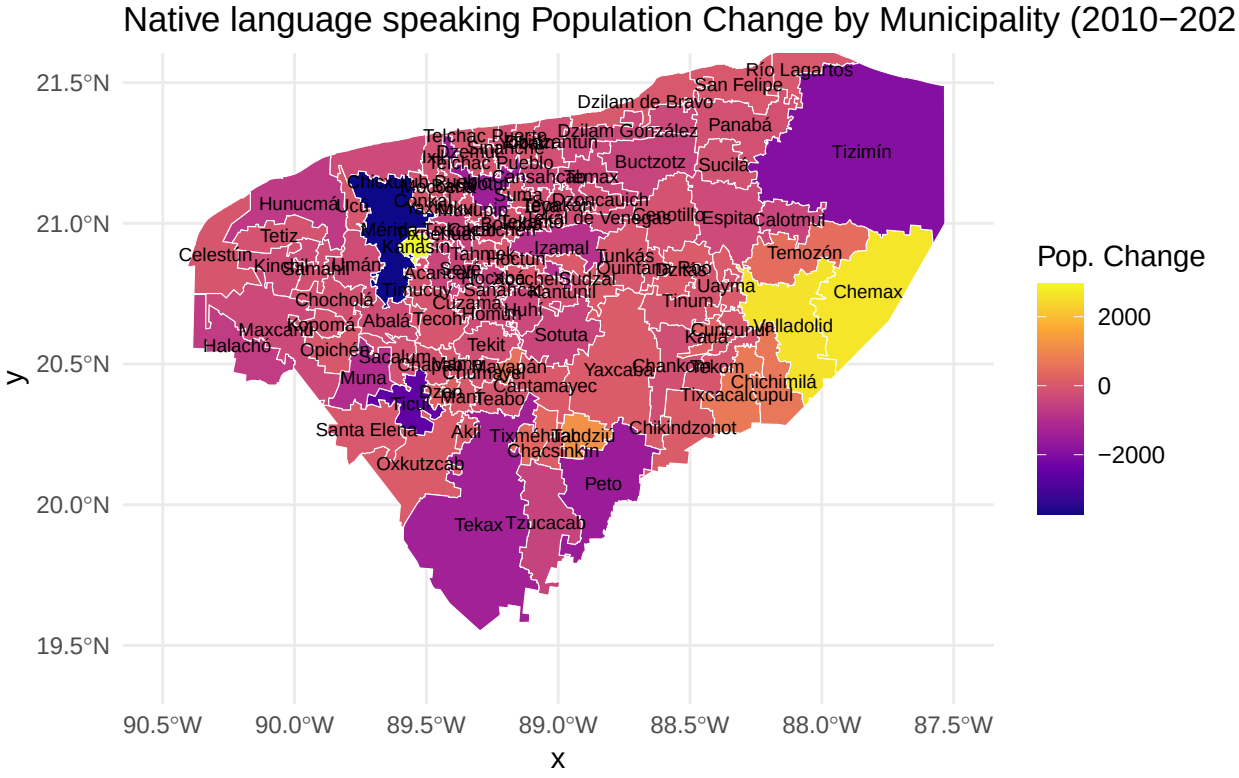

```
ggsave("indigenous_population_share_2020.png", map_plot, width = 10, height = 6, dpi = 300)
```

This shows a trend, the Indigenous population is decreasing in the state of Yucatán. either by being displaced or because their roots are being lost, the metholodgy of INEGI allows people to identify themselves. Another thing that is clear to see is that the indigenous population is concentrated in the south of the state, and from 2010 to 2020 the trend becomer more clear.

Just out of curiosity, lets see how the change in Indigenous population that speak a native language has changed.

```
map_plot <- ggplot(map_data) +
  geom_sf(aes(fill = IndigenousPopulation5PlusChange), color = "white") +
  geom_sf_text(data = map_data_centroids, aes(label = NOMGEO), size = 2.5, color = "black") +
  scale_fill_viridis_c(option = "C") +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 21.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "Native language speaking Population Change by Municipality (2010-2020)",
    fill = "Pop. Change"
  )
print(map_plot)
```

```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```




```
# save to png
ggsave("native_language_population_change_map.png", map_plot, width = 10, height = 6, dpi = 300)
```

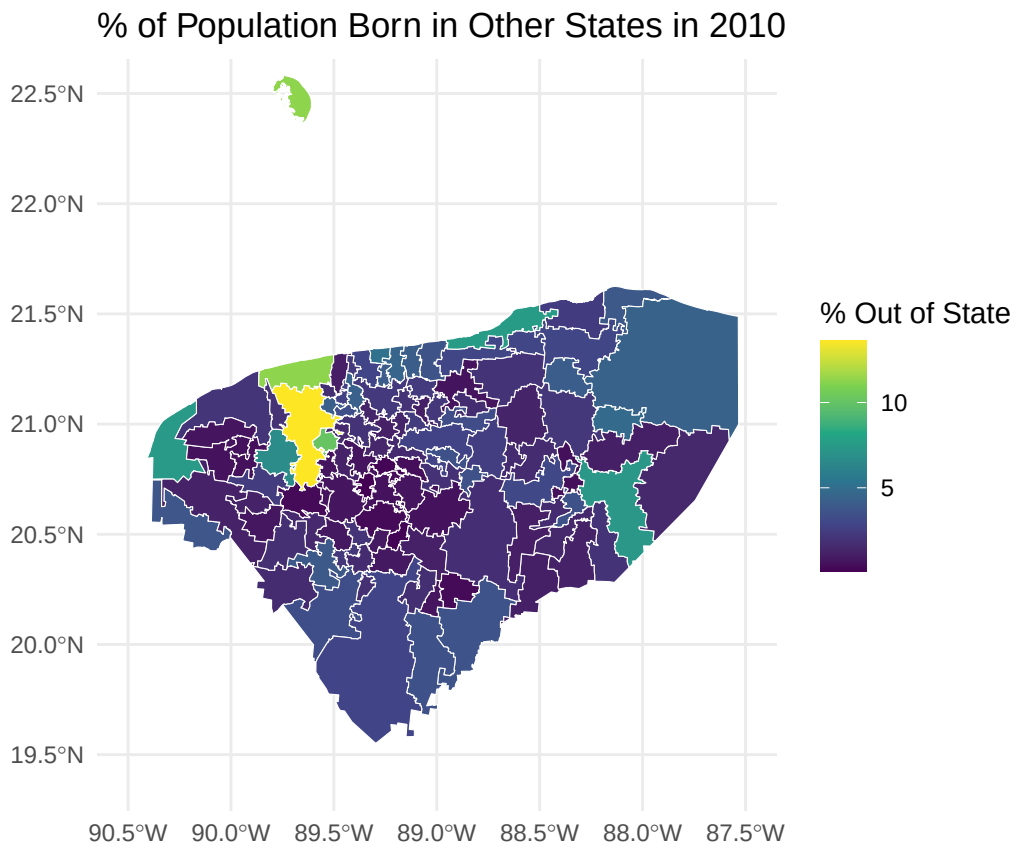
```
## Warning in st_point_on_surface.sfc(sf::st_zm(x)): st_point_on_surface may not
## give correct results for longitude/latitude data
```

Let´s do the same with out of state migration

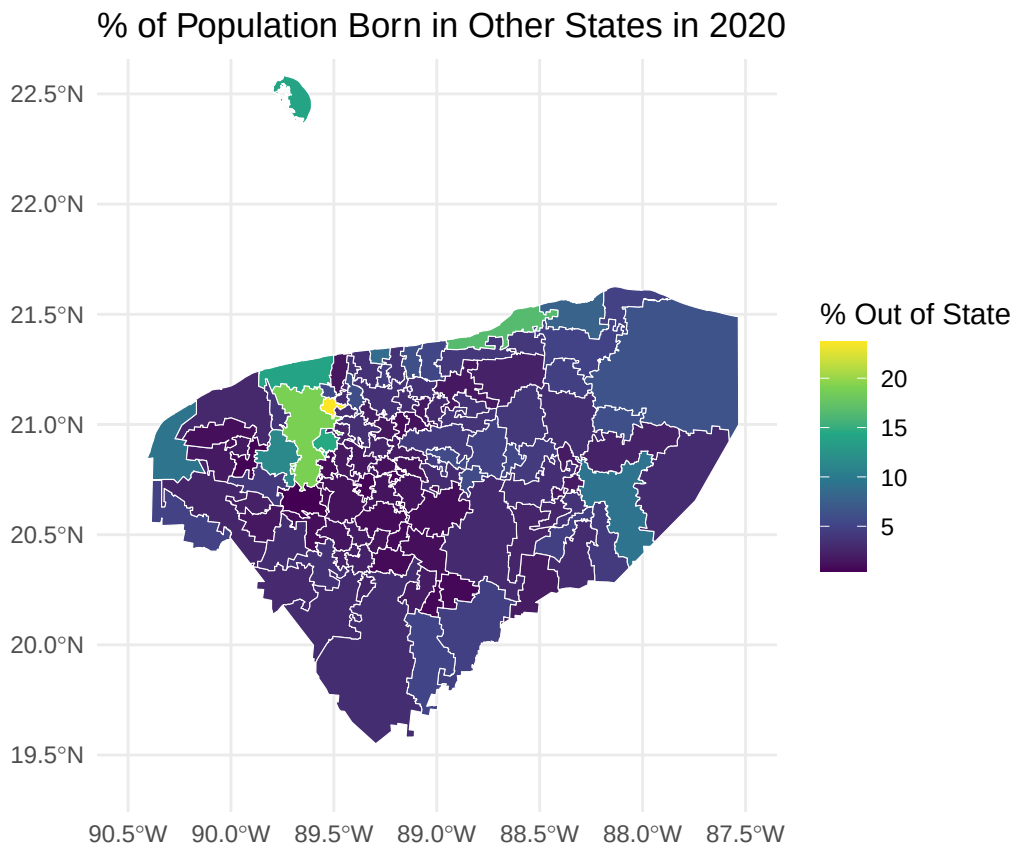
```
# Start from yucatan_map to preserve geometry
map_data <- yucatan_map %>%
  mutate(CVE_MUN = as.numeric(CVE_MUN)) %>%
  left_join(population_changes, by = c("CVE_MUN" = "MUN")) %>%
  mutate(
    OutOfStateShare2020 = BornInOtherState2020 / Population2020 * 100,
    OutOfStateShare2010 = BornInOtherState2010 / Population2010 * 100
  )
```

```
# Now it has geometry + your data, so this will work, lets plot 2010
ggplot(map_data) +
  geom_sf(aes(fill = OutOfStateShare2010), color = "white") +
  scale_fill_viridis_c(option = "D") +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 22.5)) +  # Adjust as needed
```

```
  theme_minimal() +
  labs(
    title = "% of Population Born in Other States in 2010",
    fill = "% Out of State"
  )
```




```
# Now it has geometry + your data, so this will work, lets plot 2020
ggplot(map_data) +
  geom_sf(aes(fill = OutOfStateShare2020), color = "white") +
  scale_fill_viridis_c(option = "D") +
  coord_sf(xlim = c(-90.5, -87.5), ylim = c(19.4, 22.5)) +  # Adjust as needed
  theme_minimal() +
  labs(
    title = "% of Population Born in Other States in 2020",
    fill = "% Out of State"
  )
```

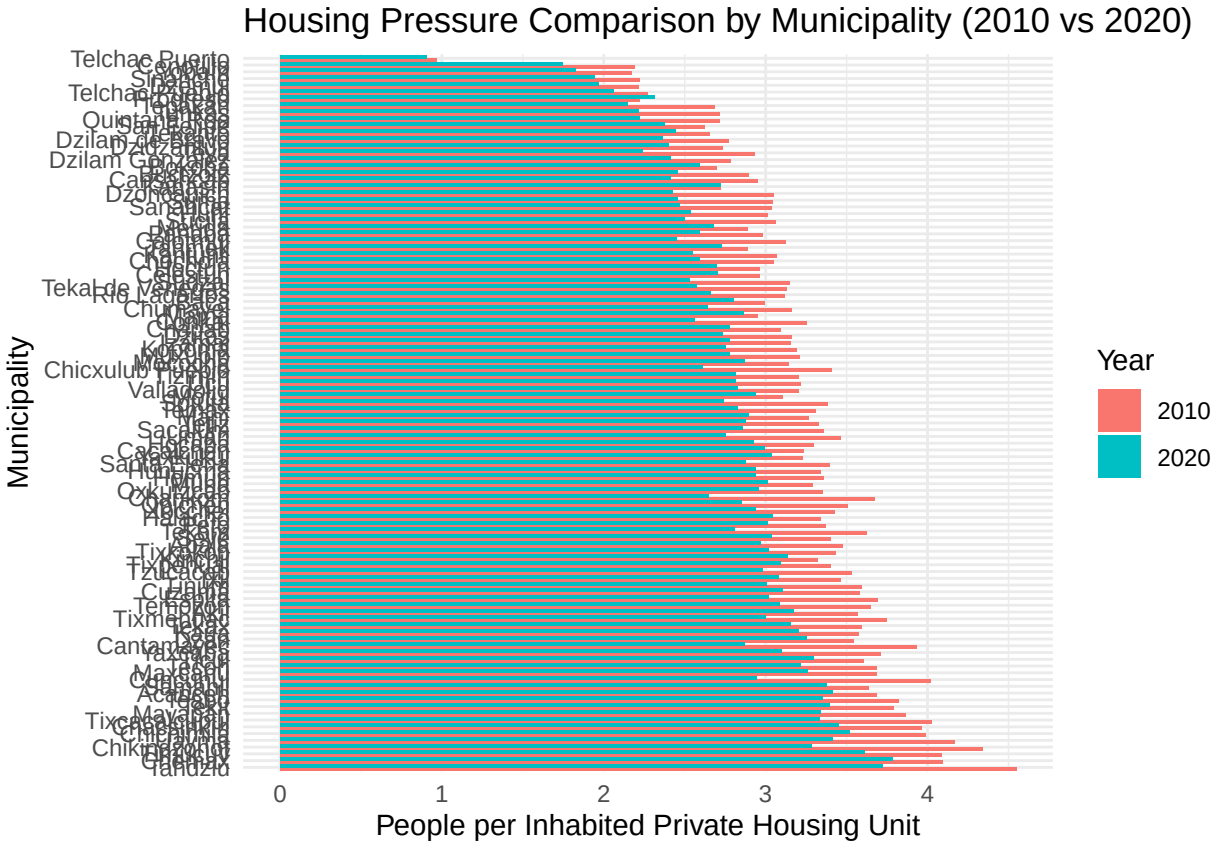


There is an increase in Merida and its surroundings, but apart from that is not as clear as the Indigenous population.

Lets explore the housing pressure

```
housing_pressure_2010 <- biggest_cities_2010 %>%
  mutate(
    POBTOT = as.numeric(P_TOTAL),
    VIVTOT = as.numeric(VIVTOT),
    HousingPressure = POBTOT / VIVTOT
  ) %>%
  select(MUN, NOM_MUN, HousingPressure)

housing_pressure_2020 <- biggest_cities_2020 %>%
  mutate(
    POBTOT = as.numeric(POBTOT),
    VIVTOT = as.numeric(VIVTOT),
    HousingPressure = POBTOT / VIVTOT
  ) %>%
  select(MUN, NOM_MUN, HousingPressure)

# Add a 'Year' column to each and bind them
housing_pressure_2010$Year <- "2010"
housing_pressure_2020$Year <- "2020"
```

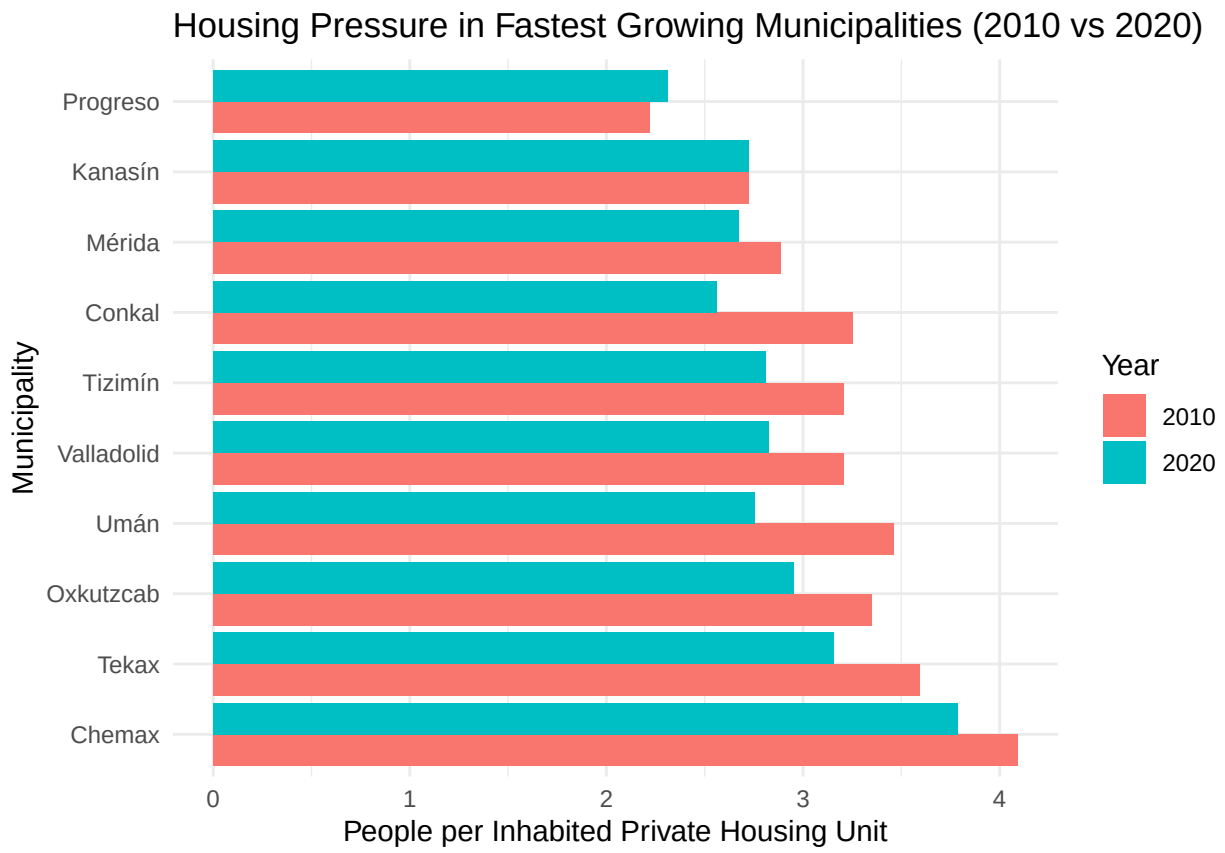
```
housing_pressure_combined <- bind_rows(housing_pressure_2010, housing_pressure_2020)

## plot both the 2010 and 2020 housing pressure, so they are next to each other
ggplot(housing_pressure_combined, aes(x = reorder(NOM_MUN, -HousingPressure), y = HousingPressure, fill
  geom_col(position = "dodge") +
  coord_flip() +
  labs(
    title = "Housing Pressure Comparison by Municipality (2010 vs 2020)",
    x = "Municipality",
    y = "People per Inhabited Private Housing Unit",
    fill = "Year"
  ) +
  theme_minimal()
```



It is not easy to distinguish, lets plot it for the cities that have grown the most.

```
most_grown_cities <- population_changes %>%
  filter(PopulationChange > 0) %>%
  arrange(desc(PopulationChange)) %>%
  head(10) %>%
  select(MUN, NOM_MUN)

plot <- housing_pressure_combined %>%
  filter(MUN %in% most_grown_cities$MUN) %>%
  ggplot(aes(x = reorder(NOM_MUN, -HousingPressure), y = HousingPressure, fill = Year)) +
```

```
  geom_col(position = "dodge") +
  coord_flip() +
  labs(
    title = "Housing Pressure in Fastest Growing Municipalities (2010 vs 2020)",
    x = "Municipality",
    y = "People per Inhabited Private Housing Unit",
    fill = "Year"
  ) +
  theme_minimal()
print(plot)
```



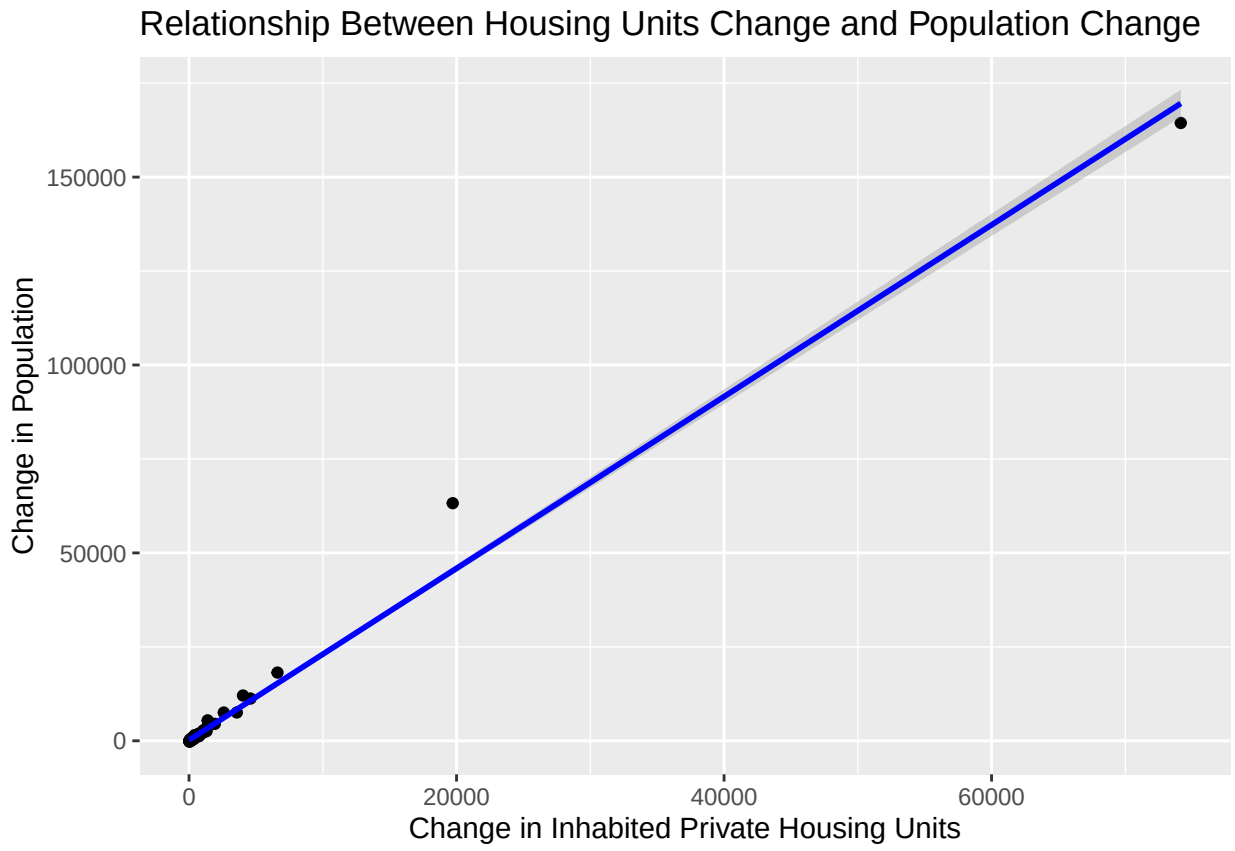

```
# save to png
ggsave("housing_pressure_fastest_growing.png", plot, width = 10, height = 6, dpi = 300)
```

Lets plot the change in housing units vs population change.

```
population_changes %>%
  left_join(biggest_cities_2020 %>%
              select(MUN, TVIVPARHAB, TVIVPARHAB_2020 = TVIVPARHAB) %>%
              mutate(TVIVPARHAB_2020 = as.numeric(TVIVPARHAB_2020)),
            by = "MUN") %>%
  left_join(biggest_cities_2010 %>%
              select(MUN, TVIVPARHAB_2010 = TVIVPARHAB) %>%
              mutate(TVIVPARHAB_2010 = as.numeric(TVIVPARHAB_2010)),
            by = "MUN") %>%
```

```
  mutate(
    HousingUnitsChange = TVIVPARHAB_2020 - TVIVPARHAB_2010,
    PopulationChange = Population2020 - Population2010
  ) %>%
  ggplot(aes(x = HousingUnitsChange, y = PopulationChange)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "Relationship Between Housing Units Change and Population Change",
    x = "Change in Inhabited Private Housing Units",
    y = "Change in Population"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



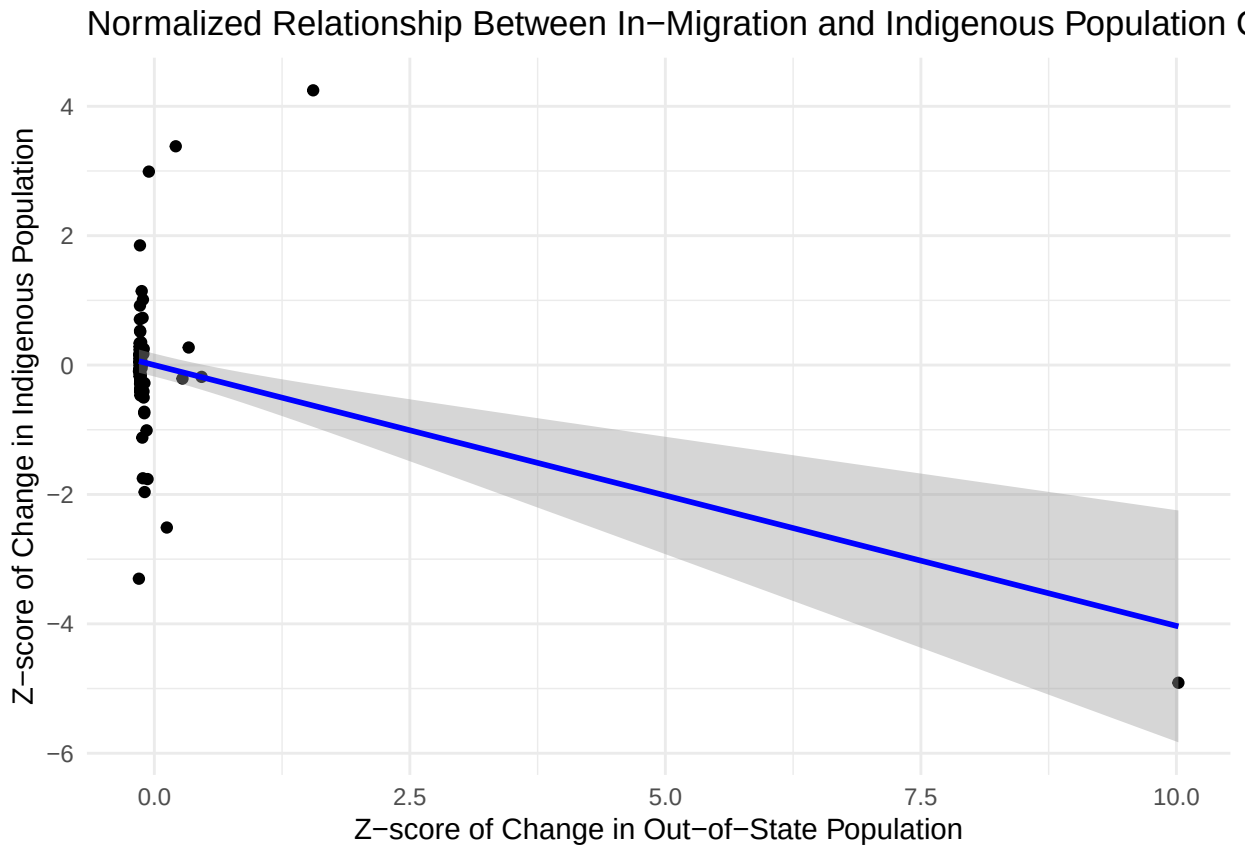
It looks like the number of housing units has increased slightly faster than the population. I checked to see if there's any note in the methodology taken by INEGI, but I did not find anything that suggests the variable does not represent the same anymore. Some conclusions we can take from this and grounded on my personal knowledge, is that the coast and cities like Merida and Valladolid have become hotspots for people to move into and/ or buy to rent out. This could explain the trend, but we would need to do a more in-depth analysis to confirm this.

Lets test the correlation between the change of out of state migration and the change in indigenous population.

Lets normalize it using z-scores, so we can see the relationship more clearly.

```
# Join the two datasets
correlation_data <- population_changes %>%
  select(MUN, NOM_MUN, BornInOtherStateChange) %>%
  left_join(indigenous_population_changes %>%
              select(MUN, IndigenousPopulationChange), by = "MUN") %>%
  mutate(
    BornInOtherStateZ = scale(BornInOtherStateChange)[, 1],
    IndigenousChangeZ = scale(IndigenousPopulationChange)[, 1]
  )


# Plot
ggplot(correlation_data, aes(x = BornInOtherStateZ, y = IndigenousChangeZ)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  labs(
    title = "Normalized Relationship Between In-Migration and Indigenous Population Change",
    x = "Z-score of Change in Out-of-State Population",
    y = "Z-score of Change in Indigenous Population"
  ) +
  theme_minimal()
```

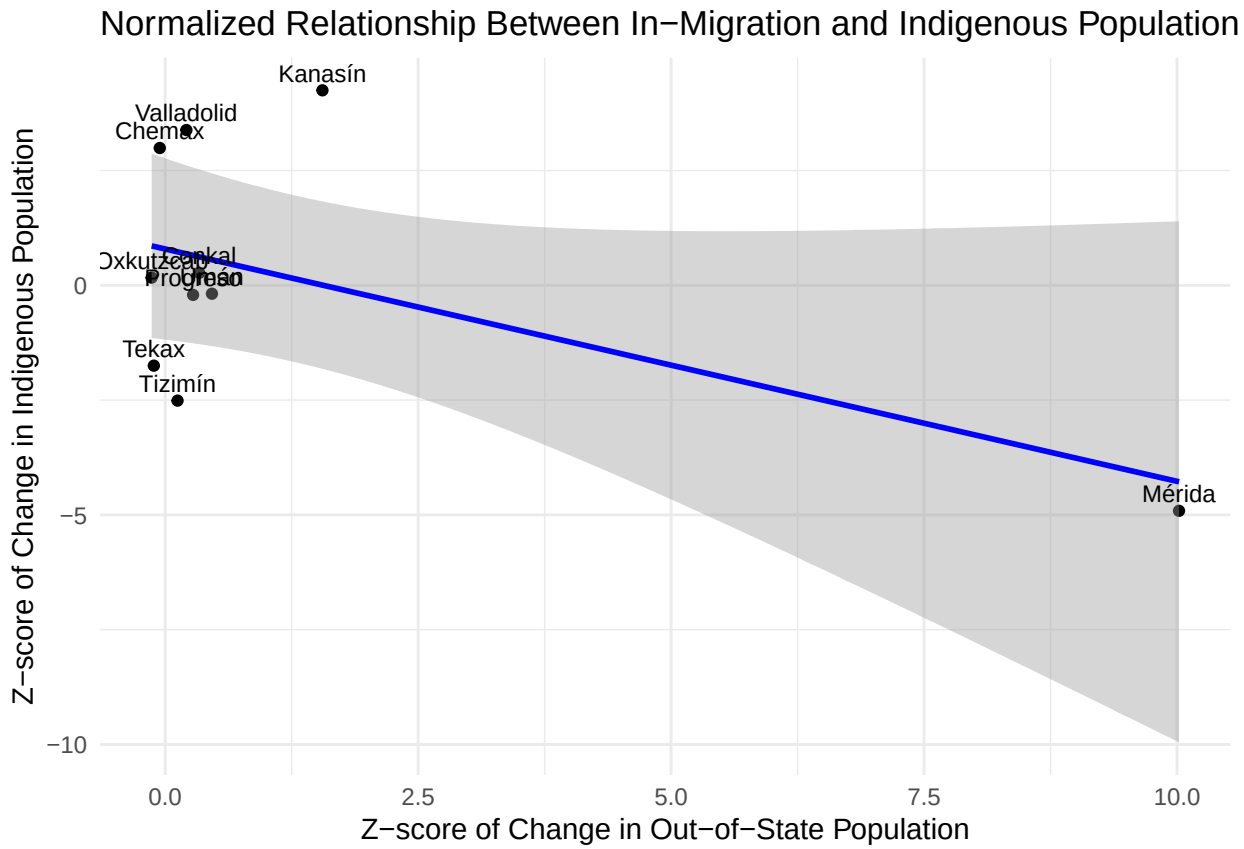
```
## `geom_smooth()` using formula = 'y ~ x'
```



It is not clear because the outlier is skewing the results, lets focus on the most grown cities.

```
# Filter for the most grown cities
plot <- ggplot(correlation_data %>%
         filter(MUN %in% most_grown_cities$MUN),
       aes(x = BornInOtherStateZ, y = IndigenousChangeZ)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = TRUE) +
  geom_text(aes(label = NOM_MUN), vjust = -0.5, size = 3) +
  labs(
    title = "Normalized Relationship Between In-Migration and Indigenous Population Change (Most Grown C
    x = "Z-score of Change in Out-of-State Population",
    y = "Z-score of Change in Indigenous Population"
  ) +
  theme_minimal()
print(plot)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# save to png
ggsave("normalized_relationship_in_migration_indigenous_change.png", plot, width = 10, height = 6, dpi =
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Let's test it.

```
cor.test(correlation_data$BornInOtherStateZ, correlation_data$IndigenousChangeZ)
```

```
##
##  Pearson's product-moment correlation
##
## data:  correlation_data$BornInOtherStateZ and correlation_data$IndigenousChangeZ
## t = -4.4923, df = 104, p-value = 1.832e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5514721 -0.2300595
## sample estimates:
##       cor
## -0.4031238
```