

Bilingual Populations - Percent Point Change

2024-07-23

Loading in Data

```
ddi <- read_ipums_ddi("usa_00009.xml")
all_indicator_data <- read_ipums_micro(ddi)
```

```
## Use of data from IPUMS USA is subject to conditions including that users should cite
the data appropriately. Use command `ipums_conditions()` for more details.
```

```
#2022
poverty_2022 <- read.csv("ACS_DATA/2022/ACSDT5Y2022.B16009-Data.csv")
language_2022 <- read.csv("ACS_DATA/2022/ACSST5Y2022.S1601-Data.csv")
social_2022 <- read.csv("ACS_DATA/2022/ACSCP5Y2022.CP02-Data.csv")
characteristics_2022 <- read.csv("ACS_DATA/2022/ACSST5Y2022.S1603-Data.csv")
limited_eng_2022 <- read.csv("ACS_DATA/2022/ACSST5Y2022.S1602-Data.csv")
household_2022 <- read.csv("ACS_DATA/2022/ACSDT5Y2022.B16002-Data.csv")
education_2022 <- read.csv("ACS_DATA/2022/ACSDT5Y2022.B16010-Data.csv")

#2012
language_2012 <- read.csv("ACS_DATA/2012/ACSST5Y2012.S1601-Data.csv")

#location data
regions <- read.csv("location_data/County_12_Regions.csv")
rural_urban <- read.csv("location_data/rural_urban.csv")
```

CUMULATIVE TRENDS

BILINGUALISM 2012-2022

Overall methodology: Taking people that speak language other than English for all counties, multiplying it by proportion of people that speak given language that speak English 'very well' to compose bilingual population in each county

```

bilingualism_2012 <- language_2012 |>
  select(Geographic.Area.Name, Total..Estimate..Population.5.years.and.over,
    Total..Estimate..Speak.a.language.other.than.English,
    Percent.of.specified.language.speakers..Speak.English.very.well...Estimate..Spe
ak.a.language.other.than.English.,

    #Selecting percents
    Percent.of.specified.language.speakers..Speak.English.very.well...Estimate..Spe
ak.a.language.other.than.English..Spanish.or.Spanish.Creole.,
    Percent.of.specified.language.speakers..Speak.English.very.well...Estimate..Spe
ak.a.language.other.than.English..Asian.and.Pacific.Island.languages.,
    Percent.of.specified.language.speakers..Speak.English.very.well...Estimate..Spe
ak.a.language.other.than.English..Other.Indo.European.languages.,
    Percent.of.specified.language.speakers..Speak.English..less.than.very.well...Es
timate..Speak.a.language.other.than.English..Other.languages.,

    #Selecting bilingual populations
    Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Spanish.or.Spanish.Creol
e,
    Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Asian.and.Pacific.Island.
languages,
    Total..Estimate..Speak.a.language.other.than.English..Other.Indo.European.langu
ages,
    Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Other.languages
  ) |>

  #Converting to decimal for later perposes
  mutate(Total..Estimate..Speak.a.language.other.than.English = Total..Estimate..Speak.
a.language.other.than.English/ 100) |>

  #renaming for clarity cause the other names are so confusing and makes it eas
ier to just code normally later on
  rename(NonEnglish_Language_Estimate = Total..Estimate..Speak.a.language.other.than.Eng
lish,
    Spanish_Estimate = Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Spanis
h.or.Spanish.Creole,
    Asian_Pacific_Estimate = Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..
Asian.and.Pacific.Island.languages,
    Other_Indo_Europe_Estimate = Total..Estimate..Speak.a.language.other.than.Engli
sh..Other.Indo.European.languages,
    Other_Estimate = Total..Estimate..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Other.la
nguages) |>

  #grabbing bilingual proportions from each language category
  rename(Percent_Bilingual = Percent.of.specified.language.speakers..Speak.English.very.
well...Estimate..Speak.a.language.other.than.English.,
    Percent_Spanish_Bilingual = Percent.of.specified.language.speakers..Speak.English.
very.well...Estimate..Speak.a.language.other.than.English..Spanish.or.Spanish.Creole.,
    Percent_Asian_Pacific_Bilingual = Percent.of.specified.language.speakers..Spea
k.English.very.well...Estimate..Speak.a.language.other.than.English..Asian.and.Pacific.I
sland.languages.,
    Percent_IndoEuro_Bilingual = Percent.of.specified.language.speakers..Speak.Englis

```

```

h.very.well...Estimate..Speak.a.language.other.than.English..Other.Indo.European.languages.,
  Percent_Other_Bilingual = Percent.of.specified.language.speakers..Speak.English..less.than.very.well...Estimate..Speak.a.language.other.than.English..Other.languages.) |>

#only category without numbers, so tranforming it based on percentages
mutate(Other_Indo_Europe_Estimate = round(Total..Estimate..Population.5.years.and.over
* (Other_Indo_Europe_Estimate /100), 0)) |>

#making blank values 0
mutate(Percent_Spanish_Bilingual = ifelse(Percent_Spanish_Bilingual == "-", 0, as.numeric(Percent_Spanish_Bilingual)),
  Percent_Asian_Pacific_Bilingual = ifelse(Percent_Asian_Pacific_Bilingual == "-", 0, as.numeric(Percent_Asian_Pacific_Bilingual)),
  Percent_IndoEuro_Bilingual = ifelse(Percent_IndoEuro_Bilingual == "-", 0, as.numeric(Percent_IndoEuro_Bilingual)),
  Percent_Other_Bilingual = ifelse(Percent_Other_Bilingual == "-", 0, as.numeric(Percent_Other_Bilingual)),
  ) |>

#creating total bilingual based on people that speak the language and multiplying by proportion that speak the language and speak english very well
mutate(Spanish_Bilingual = round(Spanish_Estimate * (Percent_Spanish_Bilingual / 100), 0),
  Asian_Pacific_Bilingual = round(Asian_Pacific_Estimate * (Percent_Asian_Pacific_Bilingual / 100), 0),
  IndoEuro_Bilingual = round(Other_Indo_Europe_Estimate * (Percent_IndoEuro_Bilingual / 100), 0),
  Other_Bilingual = round(Other_Estimate * (Percent_Other_Bilingual / 100), 0)
)

```

```

## Warning: There were 4 warnings in `mutate()`.
## The first warning was:
## i In argument: `Percent_Spanish_Bilingual = ifelse(Percent_Spanish_Bilingual ==
##   "-", 0, as.numeric(Percent_Spanish_Bilingual))`.
## Caused by warning in `ifelse()`:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 3 remaining warnings.

```

```

write.csv(file = "FULL_Bilingual_2012.csv", bilingualism_2012)

```

```

bilingualism_2022 <- language_2022 |>
  select(Geographic.Area.Name, Estimate..Total..Population.5.years.and.over,
         Estimate..Total..Population.5.years.and.over..Speak.a.language.other.than.Engli
sh,

         #grabbing people who speak the language proportions from each language category
         Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLI
SH..Spanish,
         Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLI
SH..Asian.and.Pacific.Island.languages,
         Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLI
SH..Other.Indo.European.languages,
         Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLI
SH..Other.languages,

         #grabbing bilingual proportions from each language category

         Estimate..Speak.English.only.or.speak.English.very.well...Percent.of.specified.
language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Spa
nish.,
         Estimate..Speak.English.only.or.speak.English.very.well...Percent.of.specified.
language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Asi
an.and.Pacific.Island.languages.,
         Estimate..Speak.English.only.or.speak.English.very.well...Percent.of.specified.
language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Oth
er.Indo.European.languages.,
         Estimate..Speak.English.only.or.speak.English.very.well...Percent.of.specified.
language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.THAN.ENGLISH..Oth
er.languages.) |>

         #renaming for clarity

         rename(NonEnglish_Language_Estimate = Estimate..Total..Population.5.years.and.over..Sp
eak.a.language.other.than.English,
                Spanish_Estimate = Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUA
GE.OTHER.THAN.ENGLISH..Spanish,
                Asian_Pacific_Estimate = Estimate..Total..Population.5.years.and.over..SPEAK.A.
LANGUAGE.OTHER.THAN.ENGLISH..Asian.and.Pacific.Island.languages,
                Other_Indo_Europe_Estimate = Estimate..Total..Population.5.years.and.over..SPEA
K.A.LANGUAGE.OTHER.THAN.ENGLISH..Other.Indo.European.languages,
                Other_Estimate = Estimate..Total..Population.5.years.and.over..SPEAK.A.LANGUAG
E.OTHER.THAN.ENGLISH..Other.languages,

                Spanish_Bilingual = Estimate..Speak.English.only.or.speak.English.very.well...P
ercent.of.specified.language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTH
ER.THAN.ENGLISH..Spanish.,
                Asian_Pacific_Bilingual = Estimate..Speak.English.only.or.speak.English.very.we
ll...Percent.of.specified.language.speakers..Population.5.years.and.over..SPEAK.A.LANGUA
GE.OTHER.THAN.ENGLISH..Asian.and.Pacific.Island.languages.,
                IndoEuro_Bilingual = Estimate..Speak.English.only.or.speak.English.very.well...Per
cent.of.specified.language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHE

```

```
R.THAN.ENGLISH..Other.Indo.European.languages.,
  Other_Bilingual = Estimate..Speak.English.only.or.speak.English.very.well...Perce
t.of.specified.language.speakers..Population.5.years.and.over..SPEAK.A.LANGUAGE.OTHER.TH
AN.ENGLISH..Other.languages.) |>
```

*#creating total bilingual based on people that speak the language and multiplying by p
roportion that speak the language and speak english very well*

```
mutate(
  Percent_Spanish_Bilingual = ifelse(Spanish_Estimate == 0 | Spanish_Bilingual =
= 0, 0, round(Spanish_Bilingual / Spanish_Estimate, 3)),
  Percent_Asian_Pacific_Bilingual = ifelse(Asian_Pacific_Estimate == 0 | Asian_P
acific_Bilingual == 0, 0, round(Asian_Pacific_Bilingual / Asian_Pacific_Estimate, 3)),
  Percent_IndoEuro_Bilingual = ifelse(IndoEuro_Bilingual == 0 | Other_Indo_Europ
e_Estimate == 0, 0, round(IndoEuro_Bilingual / Other_Indo_Europe_Estimate, 3)),
  Percent_Other_Bilingual = ifelse(Other_Bilingual == 0 | Other_Estimate == 0,
0, round(Other_Bilingual / Other_Estimate, 3))
)
```

```
write.csv(file = "FULL_Bilingual_2022.csv", bilingualism_2022)
```

```
final_bilingual_2012 <- bilingualism_2012 |>
  select(Geographic.Area.Name, Total..Estimate..Population.5.years.and.over, Spanish_Bi
languag, Asian_Pacific_Bilingual, IndoEuro_Bilingual, Other_Bilingual, Percent_Spanish_Bi
languag, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bilingual, Percent_Other_Bila
ngual) |>
  rename(County = Geographic.Area.Name,
    Population_5_Years_Over = Total..Estimate..Population.5.years.and.over) |>
  mutate(County = sub(" County, Texas", "", County))
```

```
final_bilingual_2012 <- left_join(final_bilingual_2012, regions, by = "County")
final_bilingual_2012 <- left_join(final_bilingual_2012, rural_urban, by = "County")
```

```
final_bilingual_2022 <- bilingualism_2022 |>
  select(Geographic.Area.Name, Estimate..Total..Population.5.years.and.over, Spanish_Bi
languag, Asian_Pacific_Bilingual, IndoEuro_Bilingual, Other_Bilingual, Percent_Spanish_Bi
languag, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bilingual, Percent_Other_Bila
ngual) |>
  rename(County = Geographic.Area.Name,
    Population_5_Years_Over = Estimate..Total..Population.5.years.and.over) |>
  mutate(County = sub(" County, Texas", "", County))
```

```
final_bilingual_2022 <- left_join(final_bilingual_2022, regions, by = "County")
final_bilingual_2022 <- left_join(final_bilingual_2022, rural_urban, by = "County")
```

```
write.csv(file = "FINAL_Bilingual_2012.csv", final_bilingual_2012)
write.csv(file = "FINAL_Bilingual_2022.csv", final_bilingual_2022)
```

RURAL/URBAN

Overall methodology: Creating weights for each county based on its proportion of population of people 5 and older compared to the whole state's population of 5 and over. After this weight was multiplied by original number, aggregated by rural and urban status to compose final aggregate.

```
urban_rural_totals <- function(bilangual_data) {
  rural_urban_tot <- bilangual_data |>
  group_by(Rural_Urban_Stat) |>
  summarise(Region_Total = sum(Population_5_Years_Over, na.rm = TRUE))

  aggregated_precursor <- bilangual_data |>
  left_join(rural_urban_tot, by = "Rural_Urban_Stat") |>
  mutate(Weight = Population_5_Years_Over / Region_Total)

  aggregated_data <- aggregated_precursor |>
  group_by(Rural_Urban_Stat) |>
  summarise(
    Total_Spanish_Bilangual = sum(Spanish_Bilangual * Weight, na.rm = TRUE),
    Total_Asian_Pacific_Bilangual = sum(Asian_Pacific_Bilangual * Weight, na.rm = TRUE),
    Total_IndoEuro_Bilangual = sum(IndoEuro_Bilangual * Weight, na.rm = TRUE),
    Total_Other_Bilangual = sum(Other_Bilangual * Weight, na.rm = TRUE),
    Total_Population_5_Years_Over = sum(Population_5_Years_Over * Weight, na.rm = TRUE)
  ) |>
  mutate(
    Percent_Spanish_Bilangual = round(Total_Spanish_Bilangual / Total_Population_5_Years_Over, 3),
    Percent_Asian_Pacific_Bilangual = round(Total_Asian_Pacific_Bilangual / Total_Population_5_Years_Over, 3),
    Percent_IndoEuro_Bilangual = round(Total_IndoEuro_Bilangual / Total_Population_5_Years_Over, 3),
    Percent_Other_Bilangual = round(Total_Other_Bilangual / Total_Population_5_Years_Over, 3)
  )

  return(aggregated_data)
}
```

```
rural_urban_2012 <- urban_rural_totals(final_bilangual_2012)
rural_urban_2022 <- urban_rural_totals(final_bilangual_2022)
```

REGIONS

Overall methodology: Creating weights for each county based on its proportion of population of people 5 and older compared to the whole state's population of 5 and over. After this weight was multiplied by original number, aggregated by region category to compose final aggregate.

```

region_total <- function(bilingual_data) {
  region_totals <- bilingual_data |>
  group_by(Region) |>
  summarise(Region_Total = sum(Population_5_Years_Over, na.rm = TRUE))

  aggregated_precursor <- bilingual_data |>
  left_join(region_totals, by = "Region") |>
  mutate(Weight = Population_5_Years_Over / Region_Total)

  aggregated_data <- aggregated_precursor |>
  group_by(Region) |>
  summarise(
    Total_Spanish_Bilingual = sum(Spanish_Bilingual * Weight, na.rm = TRUE),
    Total_Asian_Pacific_Bilingual = sum(Asian_Pacific_Bilingual * Weight, na.rm = TRUE),
    Total_IndoEuro_Bilingual = sum(IndoEuro_Bilingual * Weight, na.rm = TRUE),
    Total_Other_Bilingual = sum(Other_Bilingual * Weight, na.rm = TRUE),
    Total_Population_5_Years_Over = sum(Population_5_Years_Over * Weight, na.rm = TRUE)
  ) |>
  mutate(
    Percent_Spanish_Bilingual = round(Total_Spanish_Bilingual / Total_Population_5_Years_Over, 3),
    Percent_Asian_Pacific_Bilingual = round(Total_Asian_Pacific_Bilingual / Total_Population_5_Years_Over, 3),
    Percent_IndoEuro_Bilingual = round(Total_IndoEuro_Bilingual / Total_Population_5_Years_Over, 3),
    Percent_Other_Bilingual = round(Total_Other_Bilingual / Total_Population_5_Years_Over, 3)
  )

  return(aggregated_data)
}

```

```

regions_2012 <- region_total(final_bilingual_2012)
regions_2022 <- region_total(final_bilingual_2022)

```

PERCENT POINT CHANGES

```
to_rur_urb_merge_2012 <- rural_urban_2012 |>
  select(Percent_Spanish_Bilingual, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bi
languag, Percent_Other_Bilingual, Rural_Urban_Stat) |>
  filter(Rural_Urban_Stat != "State") |>
  rename(Percent_Spanish_Bilingual_2012 = Percent_Spanish_Bilingual,
          Percent_Asian_Pacific_Bilingual_2012 = Percent_Asian_Pacific_Bilingual,
          Percent_IndoEuro_Bilingual_2012 = Percent_IndoEuro_Bilingual,
          Percent_Other_Bilingual_2012 = Percent_Other_Bilingual
        )

to_rur_urb_merge_2022 <- rural_urban_2022 |>
  select(Percent_Spanish_Bilingual, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bi
languag, Percent_Other_Bilingual, Rural_Urban_Stat) |>
  filter(Rural_Urban_Stat != "State")

differentials_rural_urban <- merge(to_rur_urb_merge_2012, to_rur_urb_merge_2022, by = "R
ural_Urban_Stat") |>
  mutate(Spanish = (Percent_Spanish_Bilingual - Percent_Spanish_Bilingual_2012) * 100,
          Asian_Pacific_Island_Languages = (Percent_Asian_Pacific_Bilingual - Percent_Asi
an_Pacific_Bilingual_2012) * 100,
          Other_Indo_European_Language = (Percent_IndoEuro_Bilingual - Percent_IndoEuro_B
ilingual_2012) * 100,
          Other_Language = (Percent_Other_Bilingual - Percent_Other_Bilingual_2012) * 100
        ) |>
  select(Rural_Urban_Stat, Spanish, Asian_Pacific_Island_Languages, Other_Indo_European_
Language, Other_Language )
```



```
to_merge_regions_2012 <- regions_2012 |>
  select(Percent_Spanish_Bilingual, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bi
languag, Percent_Other_Bilingual, Region) |>
  filter(Region != "State") |>
  rename(Percent_Spanish_Bilingual_2012 = Percent_Spanish_Bilingual,
    Percent_Asian_Pacific_Bilingual_2012 = Percent_Asian_Pacific_Bilingual,
    Percent_IndoEuro_Bilingual_2012 = Percent_IndoEuro_Bilingual,
    Percent_Other_Bilingual_2012 = Percent_Other_Bilingual
  )

to_merge_regions_2022 <- regions_2022 |>
  select(Percent_Spanish_Bilingual, Percent_Asian_Pacific_Bilingual, Percent_IndoEuro_Bi
languag, Percent_Other_Bilingual, Region) |>
  filter(Region != "State")

differentials_regions <- merge(to_merge_regions_2012, to_merge_regions_2022, by = "Regio
n") |>
  mutate(Spanish = (Percent_Spanish_Bilingual - Percent_Spanish_Bilingual_2012) * 100,
    Asian_Pacific_Island_Languages = (Percent_Asian_Pacific_Bilingual - Percent_Asi
an_Pacific_Bilingual_2012) * 100,
    Other_Indo_European_Language = (Percent_IndoEuro_Bilingual - Percent_IndoEuro_B
ilingual_2012) * 100,
    Other_Language = (Percent_Other_Bilingual - Percent_Other_Bilingual_2012) * 100
  ) |>
  select(Region, Spanish, Asian_Pacific_Island_Languages, Other_Indo_European_Language,
Other_Language )

write.csv(file = "Regions_Bilingual_Differential.csv", differentials_regions)
write.csv(file = "Rural_Urban_Bilingual_Differential.csv", differentials_rural_urban)
```