

## covid\_racial\_inequality

```
rm(list=ls())
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.0      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(tidytext)
library(naniar)

# All constant variable
targetRaceEthnicity = c("Non-Hispanic Black", "Non-Hispanic White")

excess_death = read_csv("data/AH_Quarterly_Excess_Deaths_by_State__Sex__Age__and_Race.csv")

##
## -- Column specification -----
## cols(
##   .default = col_character(),
##   Year = col_double(),
##   'Deaths (weighted)' = col_double(),
##   'COVID19 (weighted)' = col_double(),
##   'Deaths (unweighted)' = col_double(),
##   'COVID19 (unweighted)' = col_double(),
##   'Average number of deaths (weighted)' = col_double(),
##   'Average number of deaths (unweighted)' = col_double(),
##   'Number above average (weighted)' = col_double(),
##   'Percent above average (weighted)' = col_double(),
##   'Number above average (unweighted)' = col_double(),
##   'Percent above average (unweighted)' = col_double()
## )
## i Use 'spec()' for the full column specifications.

# Data cleaning
excess_death = excess_death %>%
  dplyr::filter(RaceEthnicity %in% c("Hispanic", "Non-Hispanic Black", "Non-Hispanic White")) %>%
```

```
dplyr::filter(str_detect(AgeGroup, "Years") | AgeGroup %in% c("65+")) %>%
mutate(AgeGroup = case_when(
  AgeGroup == "50-54 Years" ~ "50-64 Years",
  TRUE ~ AgeGroup
))
```

*# Narrow down to specific state and time interval*

```
death_race_age_quarter = excess_death %>%
  filter(Year %in% c(2020, 2021) & State == "WI" & Sex == "All Sexes") %>%
  group_by(RaceEthnicity, AgeGroup, YearQuarter) %>%
  summarise(covidDeath = sum(`COVID19 (weighted)`) %>%
  replace_na(list(covidDeath = 0))
```

## 'summarise()' has grouped output by 'RaceEthnicity', 'AgeGroup'. You can override using the '.groups

```
death_race_age_quarter
```

```
## # A tibble: 75 x 4
## # Groups:   RaceEthnicity, AgeGroup [15]
##   RaceEthnicity AgeGroup   YearQuarter   covidDeath
##   <chr>         <chr>      <chr>         <dbl>
## 1 Hispanic      0-14 Years  2020, Quarter 1         0
## 2 Hispanic      0-14 Years  2020, Quarter 2         0
## 3 Hispanic      0-14 Years  2020, Quarter 3         0
## 4 Hispanic      0-14 Years  2020, Quarter 4         0
## 5 Hispanic      0-14 Years  2021, Quarter 1         0
## 6 Hispanic      15-29 Years  2020, Quarter 1         0
## 7 Hispanic      15-29 Years  2020, Quarter 2         0
## 8 Hispanic      15-29 Years  2020, Quarter 3         0
## 9 Hispanic      15-29 Years  2020, Quarter 4         0
## 10 Hispanic     15-29 Years  2021, Quarter 1         0
## # ... with 65 more rows
```

```
cdc_pop = read.delim("data/Single-Race Population Estimates 2010-2019 by State and Single-Year Age (21)
```

```
cdc_pop = cdc_pop %>%
  # Replace the "" with NA
  replace_with_na(replace = list(Ethnicity = "",
                                Race = "")) %>%
  # Filter the NA value in Ethnicity and Race column
  filter(!is.na(Ethnicity)) %>%
  filter(!is.na(Race))
```

```
cdc_pop = cdc_pop %>%
  dplyr::rename(ageGroup = Five.Year.Age.Groups) %>%
  # Reclassify the Race group
  mutate(Race = case_when(
    Race == "Black or African American" ~ "Black",
```

```

    TRUE ~ Race
  )) %>%
  # Reclassify the Ethnicity group
  mutate(Ethnicity = case_when(
    Ethnicity == "Hispanic or Latino" ~ "Hispanic",
    Ethnicity == "Not Hispanic or Latino" ~ "Non-Hispanic",
    TRUE ~ NA_character_
  )) %>%
  mutate(RaceEthnicity = case_when(
    Ethnicity == "Hispanic" ~ "Hispanic",
    TRUE ~ paste(Ethnicity, Race)
  )) %>%
  # Extract the maxAge in the original ageGroup column
  mutate(maxAge = as.numeric(str_extract(ageGroup, "[0-9]{1,2} ")) %>%
  # reclassify ageGroup based on the maxAge column
  mutate(newAgeGroup = case_when(
    maxAge <= 14 ~ "0-14 Years",
    maxAge <= 29 ~ "15-29 Years",
    maxAge <= 49 ~ "30-49 Years",
    maxAge <= 64 ~ "50-64 Years",
    maxAge >= 65 ~ "65+",
  )) %>%
  dplyr::filter(Ethnicity == "Hispanic" | RaceEthnicity %in% targetRaceEthnicity) %>%
  dplyr::filter(!is.na(newAgeGroup))

pop_race_age = cdc_pop %>%
  filter(States == "Wisconsin") %>%
  group_by(States, newAgeGroup, RaceEthnicity) %>%
  summarise(subPop = sum(Population)) %>%
  ungroup() %>%
  add_count(RaceEthnicity, wt = subPop, name = "popByRace") %>%
  mutate(w_i = subPop/popByRace) %>%
  arrange(RaceEthnicity)

```

## 'summarise()' has grouped output by 'States', 'newAgeGroup'. You can override using the '.groups' arg

```
pop_race_age
```

```
## # A tibble: 15 x 6
```

	States	newAgeGroup	RaceEthnicity	subPop	popByRace	w_i
	<chr>	<chr>	<chr>	<int>	<int>	<dbl>
## 1	Wisconsin	0-14 Years	Hispanic	128402	411208	0.312
## 2	Wisconsin	15-29 Years	Hispanic	107762	411208	0.262
## 3	Wisconsin	30-49 Years	Hispanic	113745	411208	0.277
## 4	Wisconsin	50-64 Years	Hispanic	44266	411208	0.108
## 5	Wisconsin	65+	Hispanic	17033	411208	0.0414
## 6	Wisconsin	0-14 Years	Non-Hispanic Black	93567	369471	0.253
## 7	Wisconsin	15-29 Years	Non-Hispanic Black	95295	369471	0.258
## 8	Wisconsin	30-49 Years	Non-Hispanic Black	95747	369471	0.259
## 9	Wisconsin	50-64 Years	Non-Hispanic Black	55962	369471	0.151
## 10	Wisconsin	65+	Non-Hispanic Black	28900	369471	0.0782
## 11	Wisconsin	0-14 Years	Non-Hispanic White	728477	4587417	0.159

```
## 12 Wisconsin 15-29 Years Non-Hispanic White 852951 4587417 0.186
## 13 Wisconsin 30-49 Years Non-Hispanic White 1123392 4587417 0.245
## 14 Wisconsin 50-64 Years Non-Hispanic White 1058550 4587417 0.231
## 15 Wisconsin 65+ Non-Hispanic White 824047 4587417 0.180
```

```
death_race_age_quarter %>%
  left_join(pop_race_age, by = c("AgeGroup" = "newAgeGroup", "RaceEthnicity" = "RaceEthnicity")) %>%
  arrange(RaceEthnicity, YearQuarter) %>%
  mutate(r_i = covidDeath / subPop,
         w_r = w_i * r_i * 100000) %>%
  group_by(RaceEthnicity, YearQuarter) %>%
  mutate(r_adj = sum(w_r)) %>%
  group_by(RaceEthnicity, YearQuarter) %>%
  summarise(mean(r_adj))
```

## 'summarise()' has grouped output by 'RaceEthnicity'. You can override using the '.groups' argument.

```
## # A tibble: 15 x 3
## # Groups:   RaceEthnicity [3]
##   RaceEthnicity      YearQuarter    'mean(r_adj)'
##   <chr>           <chr>          <dbl>
## 1 Hispanic        2020, Quarter 1      0
## 2 Hispanic        2020, Quarter 2    21.4
## 3 Hispanic        2020, Quarter 3     7.78
## 4 Hispanic        2020, Quarter 4    34.0
## 5 Hispanic        2021, Quarter 1     9.48
## 6 Non-Hispanic Black 2020, Quarter 1     3.52
## 7 Non-Hispanic Black 2020, Quarter 2    34.9
## 8 Non-Hispanic Black 2020, Quarter 3     8.39
## 9 Non-Hispanic Black 2020, Quarter 4    50.6
## 10 Non-Hispanic Black 2021, Quarter 1    16.0
## 11 Non-Hispanic White 2020, Quarter 1     0.349
## 12 Non-Hispanic White 2020, Quarter 2     9.13
## 13 Non-Hispanic White 2020, Quarter 3     9.16
## 14 Non-Hispanic White 2020, Quarter 4    95.3
## 15 Non-Hispanic White 2021, Quarter 1    31.8
```