

# gis project data cleaning

Camille Okonkwo

2024-03-27

```
MMR_edit = read_csv("data/MMR_2019_edit.csv")

## Rows: 281 Columns: 15
## -- Column specification -----
## Delimiter: ","
## chr (7): measure_name, location_id, location_name, race_group, sex_name, age...
## dbl (8): measure_id, sex_id, age_group_id, year_id, metric_id, val, lower, u...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

MMR_2019_edit_Hispanic =
  MMR_edit |>
  filter(year_id == "2019",
         race_group == "Hispanic and any race")

write_csv(MMR_2019_edit_Hispanic, "data/MMR_2019_edit_Hispanic.csv", row.names = FALSE)

MMR_2019_edit_Native =
  MMR_edit |>
  filter(year_id == "2019",
         race_group == "Non-Hispanic American Indian and Alaska Native")

write_csv(MMR_2019_edit_Native, "data/MMR_2019_edit_Native.csv", row.names = FALSE)

MMR_2019_edit_Asian =
  MMR_edit |>
  filter(year_id == "2019",
         race_group == "Non-Hispanic Asian, Native Hawaiian, or Other Pacific Islander")

write_csv(MMR_2019_edit_Asian, "data/MMR_2019_edit_Asian.csv", row.names = FALSE)

MMR_2019_edit_Black =
  MMR_edit |>
  filter(year_id == "2019",
         race_group == "Non-Hispanic Black")

write_csv(MMR_2019_edit_Black, "data/MMR_2019_edit_Black.csv", row.names = FALSE)

MMR_2019_edit_White =
  MMR_edit |>
  filter(year_id == "2019",
         race_group == "Non-Hispanic White")
```

```
write.csv(MMR_2019_edit_White, "data/MMR_2019_edit_White.csv", row.names = FALSE)
```

## Comparative Boxplot

```
str(MMR_edit)
```

```
## spc_tbl_ [281 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ measure_id : num [1:281] 25 25 25 25 25 25 25 25 25 25 ...
## $ measure_name : chr [1:281] "Maternal mortality ratio" "Maternal mortality ratio" "Maternal morta
## $ location_id : chr [1:281] "01" "01" "01" "01" ...
## $ location_name : chr [1:281] "Alabama" "Alabama" "Alabama" "Alabama" ...
## $ race_group : chr [1:281] "Hispanic and any race" "Non-Hispanic American Indian and Alaska Nati
## $ sex_id : num [1:281] 2 2 2 2 2 2 2 2 2 2 ...
## $ sex_name : chr [1:281] "Female" "Female" "Female" "Female" ...
## $ age_group_id : num [1:281] 169 169 169 169 169 169 169 169 169 169 ...
## $ age_group_name: chr [1:281] "10 to 54" "10 to 54" "10 to 54" "10 to 54" ...
## $ year_id : num [1:281] 2019 2019 2019 2019 2019 ...
## $ metric_id : num [1:281] 3 3 3 3 3 3 3 3 3 3 ...
## $ metric_name : chr [1:281] "Rate" "Rate" "Rate" "Rate" ...
## $ val : num [1:281] 16.14 3.92 28.51 78.16 47.1 ...
## $ lower : num [1:281] 4.887 0.122 4.915 55.036 30.41 ...
## $ upper : num [1:281] 48.4 63.9 113.7 114.4 70.5 ...
## - attr(*, "spec")=
## .. cols(
## .. measure_id = col_double(),
## .. measure_name = col_character(),
## .. location_id = col_character(),
## .. location_name = col_character(),
## .. race_group = col_character(),
## .. sex_id = col_double(),
## .. sex_name = col_character(),
## .. age_group_id = col_double(),
## .. age_group_name = col_character(),
## .. year_id = col_double(),
## .. metric_id = col_double(),
## .. metric_name = col_character(),
## .. val = col_double(),
## .. lower = col_double(),
## .. upper = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
MMR_edit = as.data.frame(MMR_edit)
```

```
locations_to_exclude <- c("National", "Midwest Census Region", "Northeast Census Region", "South Census
```

```
boxplot =
  MMR_edit |>
  mutate(
    type =
      ifelse(race_group == "Non-Hispanic Black", "Highlighted", "Normal")) |>
  filter(!location_name %in% locations_to_exclude,
```

```

      race_group != "All racial and ethnic groups") |>
ggplot(aes(
  x = fct_reorder(race_group, val, .fun = median, .desc = TRUE), # Corrected argument name
  y = val,
  color = race_group,
  fill = type)) +
geom_boxplot(alpha = 0.5) +
labs(x = "Race", y = "MMR per 100K",
     title = " U.S. Maternal Mortality Rate, 2019",
     color = "Race Group") +
theme(axis.text.x = element_blank(),
      axis.title.x = element_blank(),
      legend.position = "right") +
guides(fill = FALSE) +
scale_x_discrete("") +
scale_fill_manual(values = c("#67000d", "#fff5f0")) +
scale_color_manual(values = c("Non-Hispanic Black" = "#67000d",
                              "Non-Hispanic American Indian and Alaska Native" = "#d32020",
                              "Non-Hispanic Asian, Native Hawaiian, or Other Pacific Islander" = "#fb9a99",
                              "Non-Hispanic White" = "#f44d38",
                              "Hispanic and any race" = "#fcbea5"))

```

```

## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.

```

```

## This warning is displayed once every 8 hours.

```

```

## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

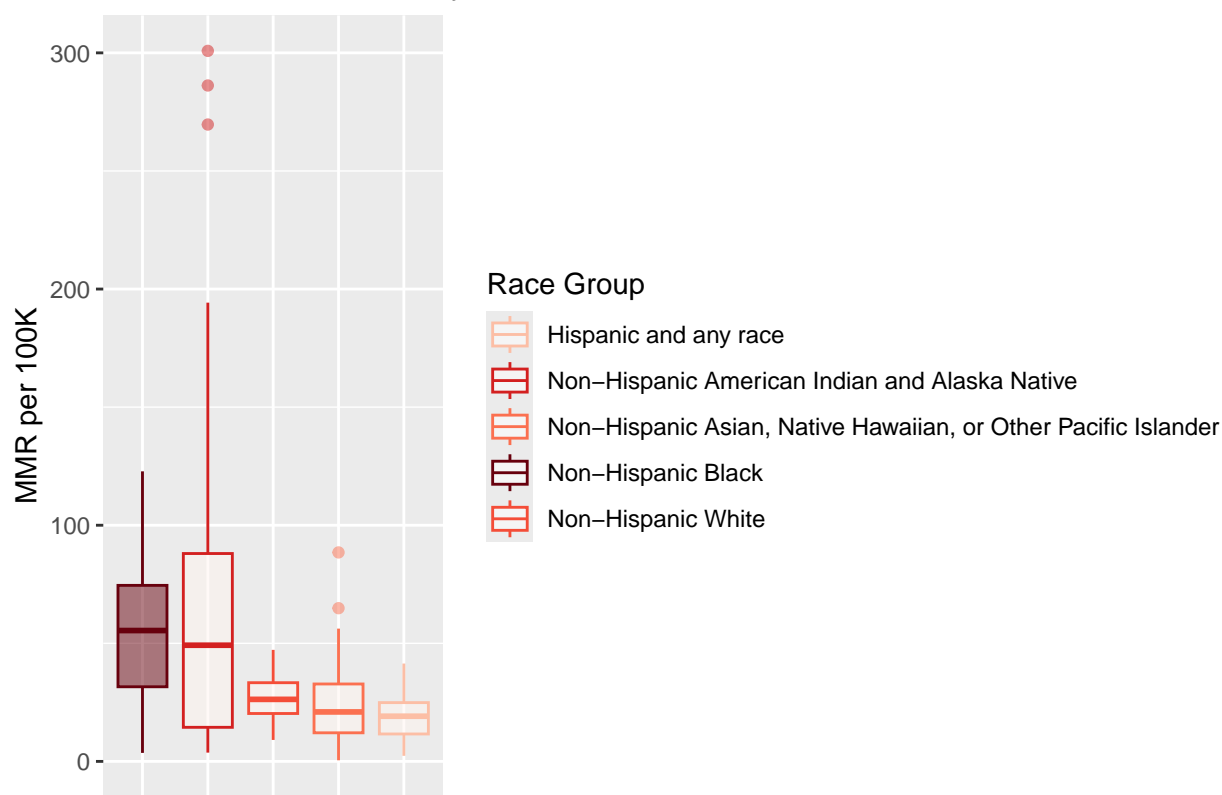
```

```

boxplot # Display the boxplot

```

## U.S. Maternal Mortality Rate, 2019



```
boxplot2 =
  MMR_edit |>
  filter(!location_name %in% locations_to_exclude,
         race_group != "All racial and ethnic groups") |>
  mutate(
    "Non-Hispanic Black" = "Black",
    "Non-Hispanic American Indian and Alaska Native" = "American Indian & Alaska Native",
    "Non-Hispanic Asian, Native Hawaiian, or Other Pacific Islander" = "Asian & Pacific Islander",
    "Non-Hispanic White" = "White",
    "Hispanic and any race" = "Hispanic"
  ) |>
  ggplot(aes(
    x = fct_reorder(race_group, val, .fun = median, .desc = TRUE), # Corrected argument name
    y = val,
    color = race_group,
    fill = race_group)) +
  geom_boxplot(alpha = 0.5) +
  labs(x = "Race", y = "MMR per 100K",
       title = "U.S. Maternal Mortality Rate, 2019",
       color = "Race Group") +
  theme(axis.text.x = element_blank(),
        axis.title.x = element_blank(),
        legend.position = "right") +
  guides(fill = FALSE) +
  scale_x_discrete("") +
  scale_fill_manual(values = c("Black" = "#67000d",
                              "American Indian & Alaska Native" = "#d32020",
```

```

" Asian & Pacific Islander" = "#fb7050",
"White" = "#f44d38",
"Hispanic" = "#fcbea5")) +
scale_color_manual(values = c("Black" = "#67000d",
"American Indian & Alaska Native" = "#d32020",
"Asian & Pacific Islander" = "#fb7050",
"Non-Hispanic White" = "#f44d38",
"Hispanic and any race" = "#fcbea5")) +

theme_classic()

print(boxplot2)

```

## Warning: No shared levels found between `names(values)` of the manual scale and the data's fill values.

### U.S. Maternal Mortality Rate, 2019

