

Excess mortality in Puerto Rico after Hurricane María.

2024-12-16

Prepare

```
library(excessmort)
library(dplyr)
library(lubridate)
library(ggplot2)
library(readxl)
data("puerto_rico_counts")
head(puerto_rico_counts)
```

	agegroup	date	sex	population	outcome
1	0-4	1985-01-01	female	158843.0	2
2	0-4	1985-01-01	male	164476.6	0
3	0-4	1985-01-02	female	158837.8	0
4	0-4	1985-01-02	male	164471.2	0
5	0-4	1985-01-03	female	158832.6	1
6	0-4	1985-01-03	male	164465.9	0

Task 2

```
pre_2017_data <- puerto_rico_counts %>%
  filter(date < as.Date("2017-01-01")) %>%
  mutate(
    year = year(date),
    week_of_year = epiweek(date)
```

```

)

# Aggregate by year, week_of_year, agegroup, and sex, and ensure full weeks (7 distinct days)
weekly_data <- pre_2017_data %>%
  group_by(year, week_of_year, agegroup, sex) %>%
  summarise(
    weekly_outcome = sum(outcome, na.rm = TRUE),
    ndays = n_distinct(date), # Count distinct days in this week-group
    .groups = 'drop'
  ) %>%
  filter(ndays == 7) # Keep only full weeks

# Compute baseline statistics across all pre-2017 years
baseline_stats <- weekly_data %>%
  group_by(agegroup, sex, week_of_year) %>%
  summarise(
    mean_outcome = mean(weekly_outcome, na.rm = TRUE),
    sd_outcome = sd(weekly_outcome, na.rm = TRUE),
    .groups = 'drop'
  )

# Check if data is available
if (nrow(baseline_stats) == 0) {
  stop("No data available for baseline after filtering full weeks.")
}

print("Baseline Weekly Statistics (Pre-2017):")

```

```
[1] "Baseline Weekly Statistics (Pre-2017):"
```

```
print(baseline_stats)
```

```
# A tibble: 1,872 x 5
```

	agegroup	sex	week_of_year	mean_outcome	sd_outcome
	<fct>	<chr>	<dbl>	<dbl>	<dbl>
1	0-4	female	1	6.08	3.38
2	0-4	female	2	6.59	3.40
3	0-4	female	3	5.31	3.64
4	0-4	female	4	5.47	2.90
5	0-4	female	5	5.94	3.17
6	0-4	female	6	5.75	3.57

7	0-4	female	7	5.59	3.64
8	0-4	female	8	5.41	3.04
9	0-4	female	9	6.03	3.11
10	0-4	female	10	5.16	2.92

i 1,862 more rows

```
pre_2017_data <- puerto_rico_counts %>%
  filter(date < as.Date("2017-01-01")) %>%
  mutate(
    year = year(date),
    week_of_year = epiweek(date)
  )

## Combine data into bigger age groups
# - Combine (0-4, 5-9, 10-14) into "0-14"
# - Combine (15-19,20-24,25-29,30-34,35-39) into "15-39"
# - Combine (40-44,45-49,50-54,55-59) into "40-59"
# - Combine (60-64,65-69,70-74) into "60-74"
# - Combine (75-79,80-84,85-Inf) into "75+"
combined_data <- pre_2017_data %>%
  mutate(combined_agegroup = case_when(
    agegroup %in% c("0-4", "5-9", "10-14") ~ "0-14",
    agegroup %in% c("15-19", "20-24", "25-29", "30-34", "35-39") ~ "15-39",
    agegroup %in% c("40-44", "45-49", "50-54", "55-59") ~ "40-59",
    agegroup %in% c("60-64", "65-69", "70-74") ~ "60-74",
    agegroup %in% c("75-79", "80-84", "85-Inf") ~ "75+",
    TRUE ~ agegroup
  ))

# Aggregate by year, week_of_year, combined_agegroup, and sex, ensuring full weeks (7 distinct
weekly_data_combined <- combined_data %>%
  group_by(year, week_of_year, combined_agegroup, sex) %>%
  summarise(
    weekly_outcome = sum(outcome, na.rm = TRUE),
    ndays = n_distinct(date),
    .groups = 'drop'
  ) %>%
  filter(ndays == 7)

# Compute baseline statistics across all pre-2017 years
baseline_stats_combined <- weekly_data_combined %>%
  group_by(combined_agegroup, sex, week_of_year) %>%
```

```

summarise(
  mean_outcome = mean(weekly_outcome, na.rm = TRUE),
  sd_outcome = sd(weekly_outcome, na.rm = TRUE),
  .groups = 'drop'
)

# Check if data is available
if (nrow(baseline_stats_combined) == 0) {
  stop("No data available for baseline after filtering full weeks.")
}

print("Baseline Weekly Statistics With Combined Age Groups:")

```

```
[1] "Baseline Weekly Statistics With Combined Age Groups:"
```

```
print(baseline_stats_combined)
```

```

# A tibble: 520 x 5
  combined_agegroup sex    week_of_year mean_outcome sd_outcome
  <chr>             <chr>      <dbl>         <dbl>      <dbl>
1 0-14             female         1           7.23       3.88
2 0-14             female         2           7.84       4.13
3 0-14             female         3           6.16       3.84
4 0-14             female         4           6.16       3.15
5 0-14             female         5           6.78       3.47
6 0-14             female         6           6.25       3.72
7 0-14             female         7           6.31       3.82
8 0-14             female         8           6.34       3.30
9 0-14             female         9           6.66       3.55
10 0-14            female        10           5.72       3.20
# i 510 more rows

```

```

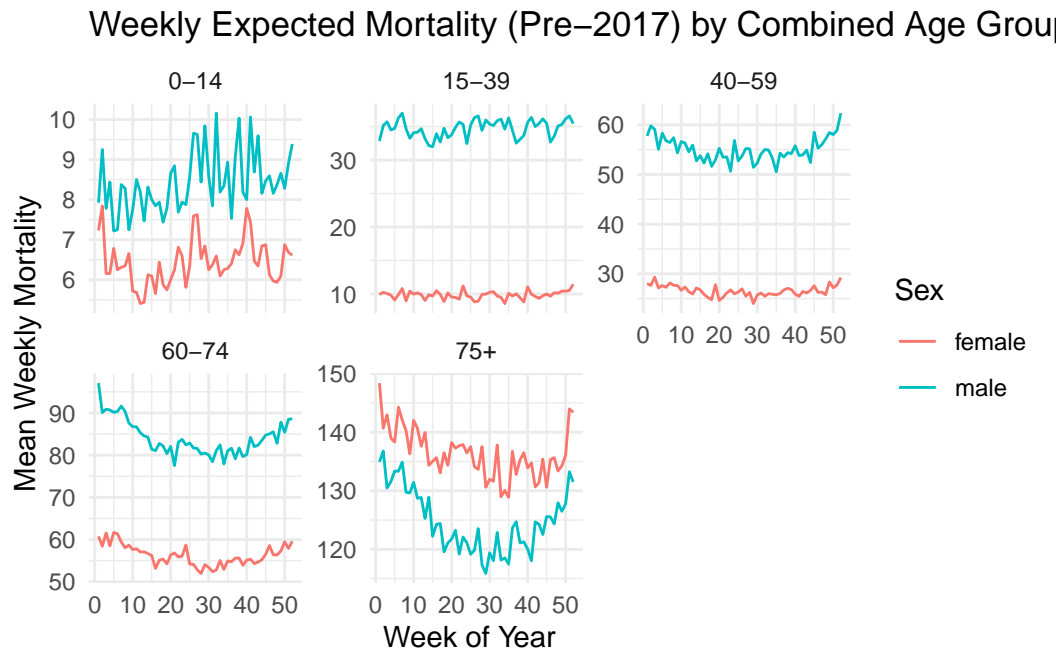
# Plot the combined groups
ggplot(baseline_stats_combined, aes(x = week_of_year, y = mean_outcome, color = sex)) +
  geom_line() +
  facet_wrap(~ combined_agegroup, scales = "free_y") +
  labs(
    title = "Weekly Expected Mortality (Pre-2017) by Combined Age Groups and Sex",
    x = "Week of Year",
    y = "Mean Weekly Mortality",

```

```

  color = "Sex"
) +
theme_minimal()

```



Task 5

```

# Read and prepare NYTimes data
ny_times <- read_excel("../data/ny_times_data.xlsx") %>%
  arrange(Date) %>%
  mutate(Date = as.Date(Date)) # Ensure Date column is in Date format

# Prepare Puerto Rico daily data
daily_data <- puerto_rico_counts %>%
  filter(date >= as.Date("2015-01-01") & date <= as.Date("2017-11-30")) %>% # Filter date range
  group_by(date) %>%
  summarize(Outcome = sum(outcome, na.rm = TRUE), .groups = "drop") %>%
  rename(Date = date) %>% # Rename `date` to `Date` for consistency
  arrange(Date)

# Join datasets and calculate the difference

```

```

compared_data <- ny_times %>%
  left_join(daily_data, by = "Date") %>%
  rename(Outcome_NYTimes = Outcome.x, Outcome_DailyData = Outcome.y) %>%
  mutate(Difference = Outcome_NYTimes - Outcome_DailyData) # Add Difference column

ggplot(compared_data, aes(x = Date)) +
  # Plot the difference
  geom_line(aes(y = abs(Difference), color = "Difference"), size = 1) +
  labs(
    title = "Absolute Difference Between NYTimes and Daily Data Outcomes Over Time",
    x = "Date",
    y = "Absolute_Difference_of_mortality",
    color = "Legend"
  ) +
  scale_x_date(
    breaks = "2 months",          # Add breaks every 2 months
    date_labels = "%Y-%m-%d",    # Format labels as "Year-Month"
    expand = c(0, 0)             # Remove extra space on the x-axis
  ) +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1), # Rotate x-axis labels
    legend.position = "bottom"                         # Move legend to the bottom
  )

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
 i Please use `linewidth` instead.

