Health-Driven Data Visualization

Dr. Sunday Adetunji, MD PhD Epi

2025-04-13

Health-Driven Data Visualization with ggplot2

Rooted in insights from: - *Health Equity Framework* (RWJF, WHO) - *Business Model Generation* (Osterwalder & Pigneur) - *Innovator's DNA* (Dyer et al.) ## 1. Setup: Libraries & Simulated Maternal Health Data

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(scales)
```

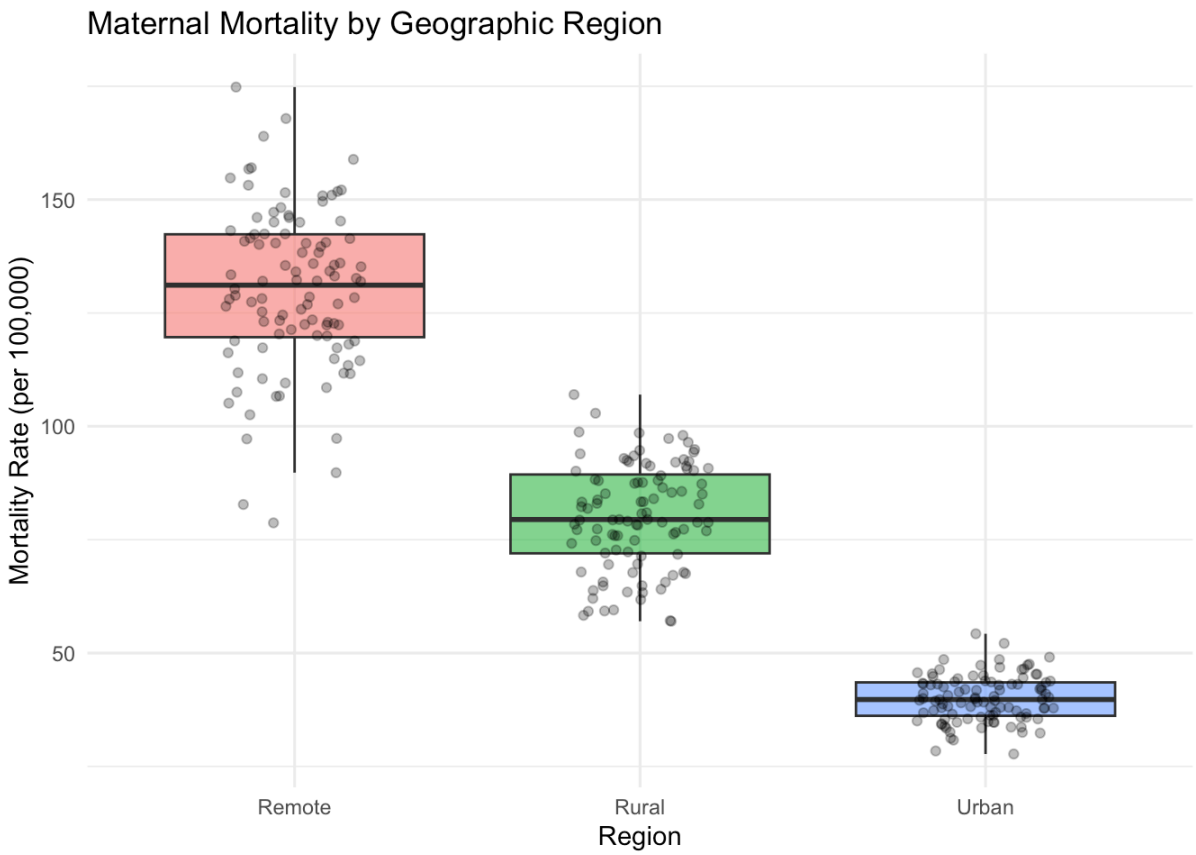
```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
set.seed(2025)
health_data <- tibble(
  region = rep(c("Urban", "Rural", "Remote"), each = 100),
  access = rep(c("High", "Medium", "Low"), each = 100),
  mortality = c(rnorm(100, 40, 5), rnorm(100, 80, 12), rnorm(100, 130, 18))
)
```

2. Boxplot of Maternal Mortality by Region

Insight: Highlights health inequities driven by structural and geographic access

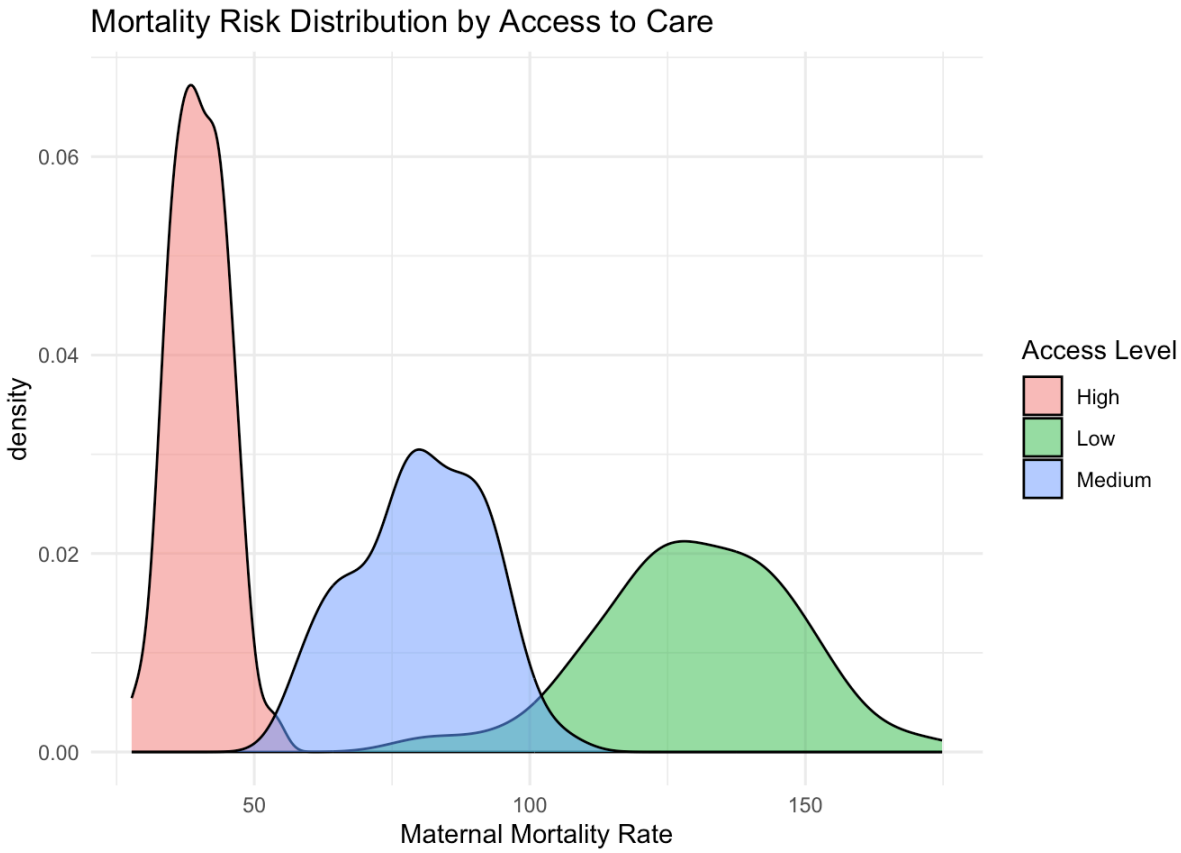
```
ggplot(health_data, aes(x = region, y = mortality, fill = region)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.6) +
  geom_jitter(width = 0.2, alpha = 0.3) +
  labs(
    title = "Maternal Mortality by Geographic Region",
    y = "Mortality Rate (per 100,000)",
    x = "Region"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```



3. Density Plot of Mortality by Access Level

Value: Helps visualize the entire distribution and overlap of risk

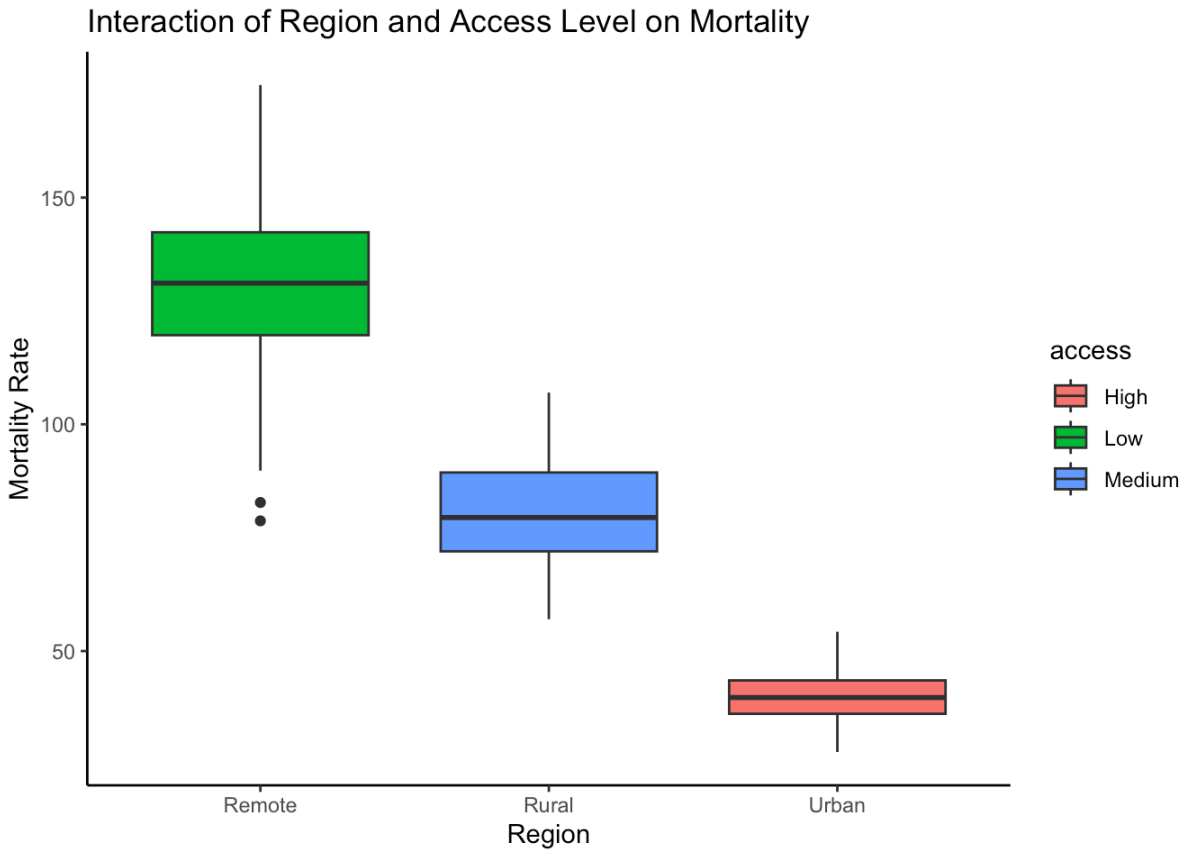
```
ggplot(health_data, aes(x = mortality, fill = access)) +
  geom_density(alpha = 0.5) +
  labs(
    title = "Mortality Risk Distribution by Access to Care",
    x = "Maternal Mortality Rate",
    fill = "Access Level"
  ) +
  theme_minimal()
```



4. Region vs Access Comparison (Interaction)

Purpose: Compare composite effects for intersectional insight

```
ggplot(health_data, aes(x = region, y = mortality, fill = access)) +
  geom_boxplot(position = position_dodge(0.8)) +
  labs(
    title = "Interaction of Region and Access Level on Mortality",
    y = "Mortality Rate", x = "Region"
  ) +
  theme_classic()
```

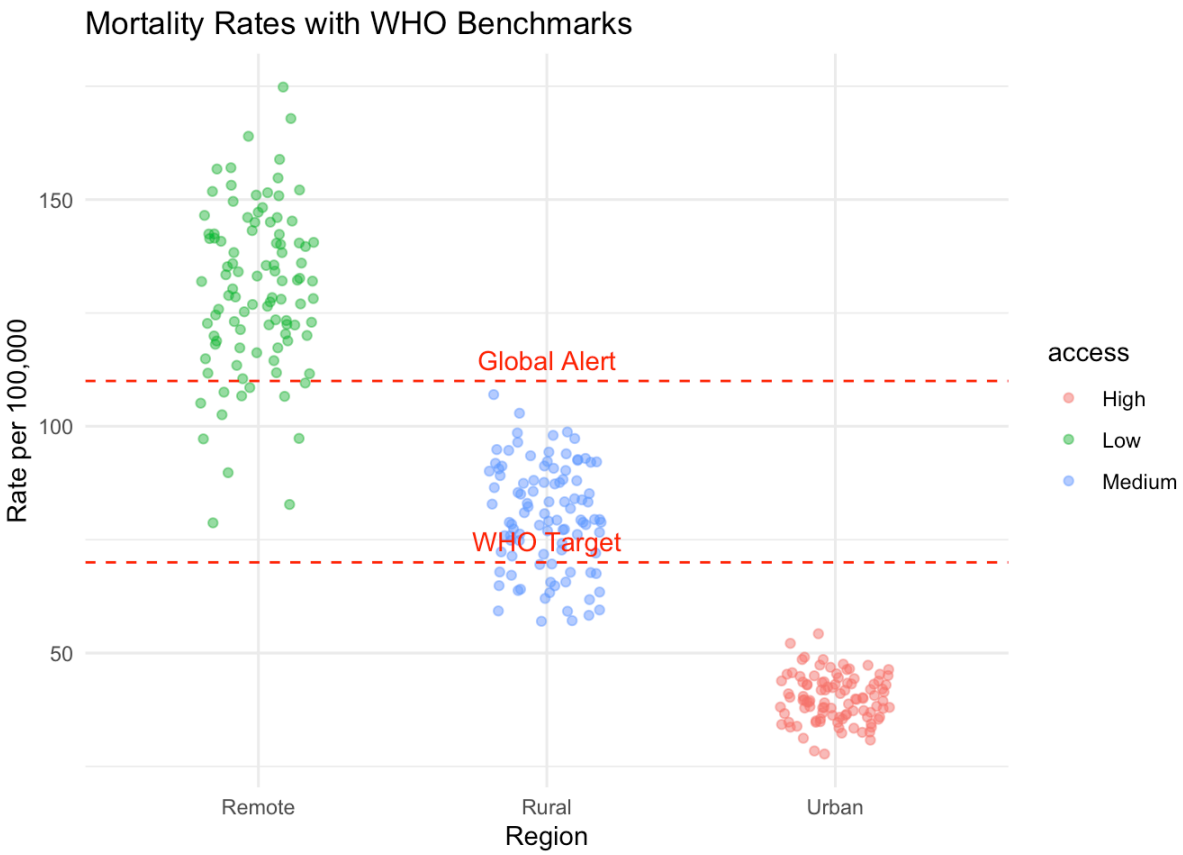


5. Annotated Thresholds and Intervention Targets

Insight: Aligns visual targets with health system policy goals

```
threshold <- tibble(level = c(70, 110), label = c("WHO Target", "Global Alert"))

health_data %>%
  ggplot(aes(x = region, y = mortality)) +
  geom_jitter(aes(color = access), width = 0.2, alpha = 0.5) +
  geom_hline(data = threshold, aes(yintercept = level), linetype = "dashed", color = "red") +
  geom_text(data = threshold, aes(x = 2, y = level, label = label), vjust = -0.6, color = "red") +
  labs(
    title = "Mortality Rates with WHO Benchmarks",
    y = "Rate per 100,000", x = "Region"
  ) +
  theme_minimal()
```



6. Narrative Insight:

- Use these visuals for grant reporting, community-engaged health research, and policymaking.
- Create storylines from region-access interactions to inform targeted interventions.
- Inspire system change by linking data with human impact stories.