```r
library(dplyr)
library(tidyr)
library(ggplot2)
library(gganimate)
library(gifski)
library(transformr)
# 1. BANGLADESH FUTURE IN 2050 (POPULATION PROJECTION)
years_pyr <- seq(1980, 2050, by = 10)
sex <- c("Male","Female")

pyr_df <- expand.grid(
  Year     = years_pyr,
  AgeGroup = paste(seq(0, 75, by = 5),
                   seq(4, 79, by = 5), sep = "-"),
  Sex      = sex
)
pyr_df
# Synthetic rule: ageing + increasing older population
pyr_df <- pyr_df |>
  mutate(
    AgeMid    = as.numeric(sub("-.*","", AgeGroup)) + 2.5,
    base      = 10000 * exp(0.05 * AgeMid),
    trend     = 1 + (Year - 1980)/70,
    SexFactor = ifelse(Sex == "Male", 1.02, 0.98),
    Pop       = round(base * trend * SexFactor),
    PopPlot   = ifelse(Sex == "Male", -Pop, Pop)
  )

p_pyr <- ggplot(pyr_df, aes(x = AgeGroup, y = PopPlot, fill = Sex)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(
    title    = "Population Pyramid {closest_state}",
    subtitle = "Synthetic data (1980–2050)",
    x = "Age Group",
    y = "Population"
  ) +
  transition_states(Year, transition_length = 2, state_length = 1)


animate(p_pyr, nframes = 80, fps = 10)

pyr_2000 <- subset(pyr_df, Year == 2000
                   )
```

```r
anim_save("pyramid_animation.gif",
          animation = animate(p_pyr, nframes = 80, fps = 10))

ggplot(pyr_2000, aes(x = AgeGroup, y = PopPlot, fill = Sex)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Population Pyramid – Year 2000 (Static)")




# 2. AGE MISREPORTING DETECTION
#############################################################

set.seed(222)
age <- sample(10:80, 2000, replace = TRUE)
true_age <- age  # assume underlying "true" ages

# Introduce heaping at multiples of 5 and 10
heap_flag <- rbinom(2000, 1, 0.3)  # 30% may be misreported
reported_age <- ifelse(heap_flag == 1,
                       round(true_age/5)*5,
                       true_age)

age_data <- data.frame(true_age, reported_age)

# Indicator: heaped if last digit 0 or 5
age_data <- age_data |>
  mutate(heaped = ifelse(reported_age %% 5 == 0, 1, 0),
         age_group = cut(reported_age, breaks = seq(10, 80, by = 5),
include.lowest = TRUE))

heaping_summary <- age_data |>
  count(age_group, heaped) |>
  group_by(age_group) |>
  mutate(p = n/sum(n))
heaping_summary

# Simple logistic regression classifier: predict heaped ~ reported_age
heap_model <- glm(heaped ~ reported_age, data = age_data, family = binomial)
summary(heap_model)

# Predicted probability of being heaped by age
```

```r
age_grid <- data.frame(reported_age = 10:80)
age_grid$heap_prob <- predict(heap_model, newdata = age_grid, type = "response")

ggplot(age_grid, aes(x = reported_age, y = heap_prob)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Predicted Probability of Age Heaping",
       x = "Reported Age", y = "Pr(Heaped)")

# 3. PEARL INDEX SIMULATION – DIFFERENT METHODS

set.seed(999)
n_women <- 500

methods <- c("Pill","Condom","Injection")

contraceptive <- data.frame(
  WomanID = 1:n_women,
  Method  = sample(methods, n_women, replace = TRUE, prob = c(0.5,0.3,0.2)),
  MonthsUse = sample(6:24, n_women, replace = TRUE)
)

failure_probs <- c(Pill = 0.04, Condom = 0.08, Injection = 0.02)

contraceptive <- contraceptive |>
  rowwise() |>
  mutate(
    Preg = rbinom(1, size = 1, prob = failure_probs[Method])
  ) |>
  ungroup()

pearl <- contraceptive |>
  group_by(Method) |>
  summarise(
    Pregnancies = sum(Preg),
    WomanMonths = sum(MonthsUse),
    PearlIndex = Pregnancies * 1200 / WomanMonths
  )
pearl

ggplot(pearl, aes(x = Method, y = PearlIndex)) +
  geom_col() +
  theme_minimal() +
  labs(title = "Pearl Index by Contraceptive Method (Synthetic)",
       y = "Pregnancies per 100 woman-years")
```