# Analysis of Countries' Population Growth Across Continents.

Group ID: A 48
Dataset number: 354

Prepared by:
Md Nahedul Bar Chowdhury Ruhul: 24160119,
Abhishek Ashok Kumar: 24140242,
Leslie Lucie Konlack Nanga: 24100607,
Zisan Ahmed: 24162855,
Govardhan Kallayi Kannoth: 24125831

School of Physics, Engineering, and Computer Science (PECS)
University of Hertfordshire
Hatfield, 2025

# Table of Contents

# 1. Introduction

## 1.1. Problem statement and research motivation

When we explored the dataset, we noticed clear contrasts in population change across continents. Many African countries, such as Nigeria with plus 2.08 percent and Ethiopia with plus 2.58 percent, are growing quickly, while several European countries, including Italy with minus 0.33 percent and Poland with minus 1.03 percent, are declining. These patterns led us to question whether the differences between continents are statistically meaningful or simply random. Our motivation is to test whether continents differ in the proportion of countries experiencing growth or decline, addressing a gap in the literature that rarely examines proportional differences rather than growth rates.

## 1.2. The data set

For this analysis, we used the World Population by Country 2025 dataset, which contains 233 countries and includes variables such as yearly population change, fertility rate, net migration, median age and urban population percentage. For our research question, we focused on Yearly Change and Continent. We converted Yearly Change into a Growth Status label, classifying countries like India with plus 0.89 percent as Growing and countries like Japan with minus 0.52 percent as Declining. Each country was assigned to its continent, allowing us to compare Africa, Asia, Europe, the Americas and Oceania and evaluate whether proportions of growing countries differ.

## 1.3. Research question

Exploring the dataset, we saw differences in countries growing or declining across continents. This led us to ask: Is there a difference in the proportion of countries with growing populations among different continents? To investigate this, we compare Growth Status and Continent, applying a Chi square test to assess variation.

## 1.4. Null hypothesis and alternative hypothesis (H0/H1)

To guide our analysis, we defined the hypotheses tested using a Chi square approach.

The Null Hypothesis (H0) states: There is no difference in the proportion of countries with growing populations among different continents. This means Growth Status and Continent are assumed to be independent, with any variation due to random chance.

The Alternative Hypothesis (H1) states: There is a difference in the proportion of countries with growing populations among different continents, as shown in our dataset. This implies a real association between continent and growth status, suggesting demographic factors such as fertility, migration or age structure may influence differences.

# 2. Background research

## 2.1. Research papers

As we kept going with our work, we looked at what researchers said about population. Most talk about big areas or focus on the future. The UN's World Population Prospects 2024 shows Africa is growing fast, while Europe is going down in numbers (United Nations Department of Economic and Social Affairs, Population Division, 2024). Dyson (2011) explains this is because regions are at different stages in how population changes. Bongaarts (2016) says this leads to different problems. For example, Africa has to manage fast growth, while Europe has to plan for older people. Vollset et al. (2020) said that by 2050, most countries might not grow anymore.

These studies give a good picture, but they mostly talk about averages or what might happen later. We saw that few ask if some continents have more growing countries right now. That part is still mostly missing. Our project tries to fill that gap. Instead of using averages, we focus on current data and look at whether a country's growth status is linked to its continent. We use a Chi square test to see if there's a real difference. This gives us a clearer view of how growth and decline are spread globally.

## 2.2. Why RQ is of interest (research gap and future directions according to the literature)

Our research question fills a gap in the current literature. While many studies talk about growth trends across continents (United Nations, 2024; Dyson, 2010; Bongaarts, 2016; Vollset et al., 2020), very few actually test whether some regions have a higher share of growing countries than others at present. That's what makes our question different and worth exploring. Instead of focusing on averages or forecasts, we test if growth status depends on continent using current data and a statistical method. This can support future research on what drives these patterns and help shape targeted population policies and sustainable development strategies worldwide.

# 3. Visualisation

## 3.1. Graph for the RQ output of an R script

To answer our question, we used a stacked bar chart that shows how many countries are growing on each continent (Shehbaz, 2025). We also made a contingency table to display the actual numbers. These graphs were created using R and help show the pattern between growth status and continent clearly and directly.

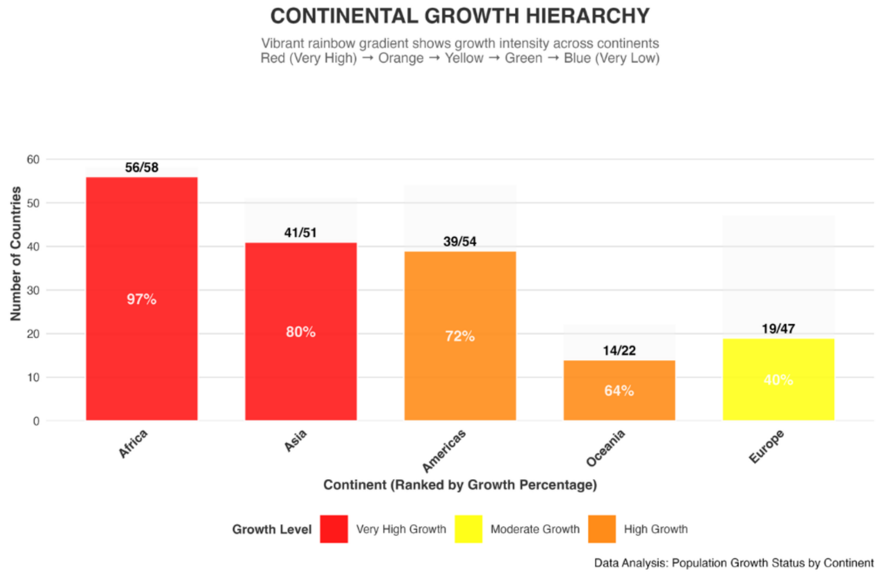


*Figure 1: Gradient Stacked Bar Chart Showing Growth Hierarchy Across Continents*

**Contingency Table (Required for Comparison of Proportions RQ)**

*Table 1: Contingency table showing distribution of growth status across continents (excluding "Other" category)*

| Continent | Total Countries | Growing Countries | Declining Countries | % Growing | Growth Level Category |
|---|---|---|---|---|---|
| Africa | 58 | 56 | 2 | 96.6% | Very High Growth |
| Asia | 51 | 41 | 10 | 80.4% | High Growth |
| Americas | 54 | 39 | 15 | 72.2% | High Growth |
| Oceania | 22 | 14 | 8 | 63.6% | High Growth |
| Europe | 47 | 19 | 28 | 40.4% | Moderate Growth |
| **Global** | **233** | **169** | **63** | **72.8%** | **High Growth** |

## 3.2. Additional information relating to understanding the data

The chart uses a bright color scale that moves from Red to Blue to match how strong the growth is. Africa appears bright Red since almost every country there is growing. The table gives exact numbers so we can back up what we see in the chart with real data (Shehbaz, 2025).

## 3.3. Useful information for the data understanding

Looking at the chart, we saw Africa had the highest growth, with 56 out of 58 countries increasing. Europe was the lowest, with only 19 out of 47 growing. As we move across the continents, the numbers drop (Shehbaz, 2025). This makes it clear that population growth depends a lot on continent.

# 4. Analysis

## 4.1. Statistical test used to test the hypotheses and output

To answer our question, we used a Chi-square test of independence. It fits because both of our variables, continent and growth status, are categories. This test checks if the pattern of growth is random or tied to where countries are. It works well when comparing proportions in groups. Our output showed that the pattern is not random. There's a real difference in how growth is spread across continents, which supports what we expected to find.

**Contingency Table for Analysis:**

*Table 2: Contingency Table showing Declining and Growing Countries across continents. (excluding "Other" category)*

| | **Declining** | **Growing** |
|---|---|---|
| Africa | 2 | 56 |
| America | 15 | 39 |
| Asia | 10 | 41 |
| Europe | 28 | 10 |
| Oceania | 8 | 14 |

## 4.2. The null hypothesis is rejected based on the p-value

**Research Question:** Is there a difference in the proportion of countries with growing populations among continents?

**Hypotheses:** $H_0$: No difference exists; $H_1$: A difference exists.

**Statistical Decision:** Chi-square test, $\chi^2(4) = 43.87$, $p < 0.0001$, $\alpha = 0.05$. Since $p < \alpha$, we **reject $H_0$**.

**Conclusion:** After running the Chi-square test, we obtained a very small p-value ($p < 0.0001$). Which is far below our significance level of 0.05. As a result, we reject the null hypothesis indicating that the differences in population growth across continents are not random. Specifically, Africa had significantly more countries with growing populations than expected. With a residual of 4.69, while Europe had significantly fewer, with a residual of -5.60. These residuals highlight how far each continent's actual counts deviate from what would be expected if growth were evenly distributed, confirming a clear pattern in continental population growth.

# 5. Evaluation – group's experience at 7COM1079

## 5.1. What went well

Our team worked effectively, with each person leading a section but contributing across tasks. A key strength was our communication: meetings kept us aligned, and GitHub prevented version issues. We collaborated well when selecting the statistical test. However, we could improve by setting clearer deadlines earlier, as we became more efficient after finding our rhythm. In future projects, responsibilities and timelines sooner would strengthen consistency. Overall, the teamwork and shared support made the project successful.

## 5.2. Points for improvement

Although teamwork was productive, we sometimes struggled to make early decisions about the direction of the research question. A clearer plan at the start would have saved time during later analysis. Additionally, we could improve consistency in documenting changes and summarising meeting outcomes. Strengthening our understanding of R before beginning the project might also have reduced delays. Better preparation and more structured meetings would help us be more efficient in future projects.

## 5.3. Group's time management

We managed to meet deadlines, but some tasks were completed close to submission due to slow progress early on. As the project advanced, our daily 9 PM meetings helped maintain momentum. Time management improved in the final weeks, though starting the analysis earlier would reduce pressure in future projects.

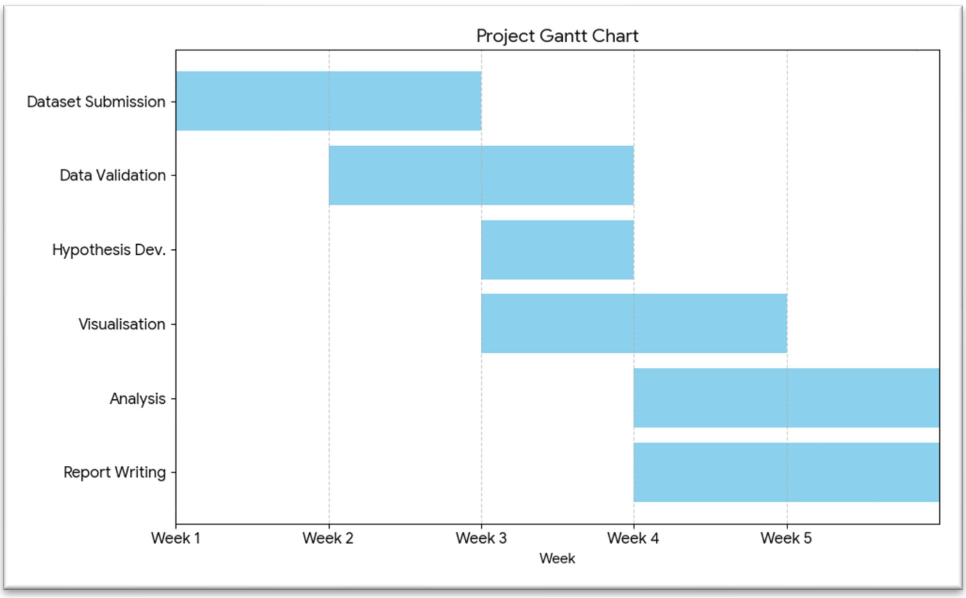


*Figure 2 Project Gantt Chart Based on Weekly updates and Task.*

## 5.4. Project's overall judgement

The project demonstrated strong analytical thinking and effective collaboration. Our research question was clear, and the analysis produced meaningful results. Early progress was slower, but our workflow improved with better organisation. Overall, the project met its objectives and reflected a solid, well-reasoned approach to data analysis in a consistent manner.

## 5.5. Three Most Significant Commits

Commit 1: "Added the dataset" (Nov 25, 2025)
Author: NahedulBarChowdhury <mr25aei@herts.ac.uk>
Impact: It marks the start of the project where the dataset was added to the repository. The team started data exploration cleaning and script development. All analysis and visualisation work is built on this foundation.

Commit 2: "Completed the test and analysis, additionally created R script for test analysis" (Dec 9, 2025)
Author: zisanahmed1227 <za25aeq@herts.ac.uk>
Impact: This is the most important technical commit. It introduced the core statistical testing (including the hypotheses test) and created the key analysis script in R. It supports the research question and determines whether the hypothesis is rejected or accepted.

Commit 3: "Added Visualization (Section 3) and Conclusion (Section 6) on the doc" (Dec 11,2025)
Author: zisanahmed1227 <za25aeq@herts.ac.uk>
Impact: In the final document this commit integrated major report components including key visualisations and the concluding interpretation into the document. It illustrates the transition from the coding to the documentation contributing directly to the final submission.

# 6. Conclusions

## 6.1. Results explained

Our analysis showed clear and significant differences in population growth across continents. Africa stood out with 56 out of 58 countries (96.6 percent) growing. Where Europe had the opposite trend, with only 19 out of 47 (40.4 percent) showing growth. The Chi-square test gave $\chi^2(4) = 43.873$ and a p-value less than 0.0001, meaning we rejected the null hypothesis. So, growth status is clearly connected to continent, and the pattern we found is not random.

## 6.2. Interpretation of the results

What we found shows that population growth isn't random but clearly follows regional patterns. Africa and Asia are seeing strong growth. Where Europe is steadily declining. Since the p-value was so small ($p < 0.0001$), this difference is real and not by chance. It likely comes from structural factors like fertility rates, migration, and policy frameworks. These results matter because they affect global planning, economic planning, and what kind of policies each region might need.

## 6.3. Reasons and implications for future work, limitations of the study

A key limitation was classifying growth as only growing or declining, which oversimplified some cases. The dataset might also be slightly outdated. Future research should explore regional patterns. Use updated data and include other factors like fertility, migration or economy to better explain differences and guide population policy planning. Future work could expand the analysis by including more countries, additional variables or longer time periods to provide a deeper understanding of trends.

# 7. Reference list

Bongaarts, J. (2016, February 24). *Development: Slow down population growth*. Nature News. https://www.nature.com/articles/530409a

Dyson, T. (2011, January 25). *The role of the demographic transition in the process of urbanization.* Wiley Online Library. http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4457.2011.00377.x/abstract

Shehbaz, A. (2025, October 15). 🌍 *world population by country 2025 (latest)*. Kaggle. https://www.kaggle.com/datasets/asadullahcreative/world-population-by-country-2025

Vollset, S. E., Goren, E., Yuan, C.-W., Cao, J., Smith, A. E., Hsiao, T., Bisignano, C., Azhar, G. S., Castro, E., Chalek, J., Dolgert, A. J., Frank, T., Fukutaki, K., Hay, S. I., Lozano, R., Mokdad, A. H., Nandakumar, V., Pierce, M., Pletcher, M., … Murray, C. J. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the global burden of disease study. *The Lancet*, *396*(10258), 1285–1306. https://doi.org/10.1016/s0140-6736(20)30677-2

World population prospects 2024: Summary of results. (n.d.). https://population.un.org/wpp/assets/Files/WPP2024_Summary-of-Results.pdf

-

# 8. Appendices

## A. R code used for analysis and visualisation

Analysis.R code with the appropriate statistics to test the hypotheses.

```
# Here we were running the full analysis
# We installed and loaded the libraries, brought in the dataset, and checked its
# structure and missing values. Then we cleaned the yearly change column by
    fixing
# symbols and converting it into a numeric variable, which allowed us to
    classify
# each country as either growing or declining. After that, we mapped every
    country
# to its continent and checked for any unmapped cases. Once the data was
    ready,
# we created several visual stories to explore how growth patterns varied
    across
# continents, looking at averages, balance, hierarchy, and distribution
    differences.
# Finally, we performed a chi square test to see whether growth status
    depended on
# continent and interpreted the statistical results to support our conclusions.

#instaling Libraries
install.packages("tidyverse")  # For data manipulation
install.packages("countrycode") # For country-to-continent mapping
install.packages("ggplot2")     # For visualizations
install.packages("dplyr")
install.packages("viridis")

# Loading libraries
library(tidyverse)
library(countrycode)
library(ggplot2)
library(dplyr)
library(viridis)

cat("✅ Libraries loaded successfully!\n")

population_data <- read.csv("population_data.csv")

# Basic dataset information
cat("=== DATASET OVERVIEW ===\n")
cat("Total countries:", nrow(population_data), "\n")
cat("Variables available:\n")
print(names(population_data))

# View structure of the dataset
str(population_data)
```

```
head(population_data, 3)

# Check for missing values
missing_values <- sum(is.na(population_data))
cat("\nTotal missing values:", missing_values, "\n")


# Creating New Variable Growth Status
cat("=== CHECKING YEARLY CHANGE VALUES ===\n")
print(head(population_data$Yearly.Change, 10))
# Converting Yearly Change to numeric (handle special minus sign)
# Replacing the special minus sign with regular minus
population_data$Yearly_Change_Numeric <- gsub("−", "-",
        population_data$Yearly.Change)
# Removing the % sign
population_data$Yearly_Change_Numeric <- gsub("%", "",
        population_data$Yearly_Change_Numeric)
# Converting to numeric
population_data$Yearly_Change_Numeric <-
        as.numeric(population_data$Yearly_Change_Numeric)

# Creating Growth Status
population_data$Growth_Status <- ifelse(
  population_data$Yearly_Change_Numeric > 0,
  "Growing",
  "Declining")
table(population_data$Growth_Status)

# Testing with the first few rows
test_rows <- population_data[1:10, ]
test_rows$Growth_Status <- ifelse(
  as.numeric(gsub("%", "", gsub("−", "-", test_rows$Yearly.Change))) > 0,
  "Growing",
  "Declining")


cat("=== TEST RESULTS (First 10 rows) ===\n")
print(test_rows[, c("Country..or.dependency.", "Yearly.Change",
        "Growth_Status")])


# Mapping countries to continent

# First, let's create the continent mapping using countrycode
population_data$Continent <- countrycode(
  sourcevar = population_data$Country..or.dependency.,
  origin = "country.name",
  destination = "continent",
  warn = FALSE
)
```

```
# Handle any countries that couldn't be mapped
na_continent <- sum(is.na(population_data$Continent))
if(na_continent > 0) {
  cat("Countries without continent assignment:", na_continent, "\n")
  # Assign "Other" for unmapped countries
  population_data$Continent[is.na(population_data$Continent)] <- "Other"
}

# Check continent distribution
cat("\n=== CONTINENT DISTRIBUTION ===\n")
continent_table <- table(population_data$Continent)
print(continent_table)

# Visualization

# Story: "Which continents are above/below average in growth?"

# Calculate average growth percentage globally
avg_growing <- mean(population_data$Growth_Status == "Growing") * 100

# Create data for diverging bars
diverging_data <- population_data %>%
  group_by(Continent) %>%
  summarise(
    Total = n(),
    Growing = sum(Growth_Status == "Growing"),
    Percent_Growing = Growing / Total * 100,
    Above_Average = Percent_Growing > avg_growing
  )

diverging_colors <- c("#E74C3C", "#3498DB") # Red-Blue gradient for
        above/below

diverging_bar <- ggplot(diverging_data, aes(x = Continent)) +
  # Bars for percent growing
  geom_bar(
    aes(y = Percent_Growing, fill = Above_Average),
    stat = "identity",
    width = 0.7
  ) +
  # Reference line for average
  geom_hline(
    yintercept = avg_growing,
    color = "#2C3E50",
    linetype = "dashed",
    size = 1,
    alpha = 0.7
  ) +
  # Label for average line
```

```
  annotate(
    "text",
    x = length(unique(diverging_data$Continent)) - 0.5,
    y = avg_growing + 2,
    label = paste0("Global Average: ", round(avg_growing, 1), "%"),
    color = "#2C3E50",
    fontface = "bold",
    size = 3.5
  ) +
  scale_fill_manual(
    values = diverging_colors,
    labels = c("Below Average", "Above Average"),
    name = "Compared to Global Average"
  ) +
  labs(
    title = "CONTINENTS VS GLOBAL AVERAGE",
    subtitle = "Shows which continents have above/below average growth
        percentages\nRed-Blue gradient indicates performance",
    x = "Continent",
    y = "% of Countries Growing",
    caption = "Story: Continental performance compared to global average"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5, color =
        "#2C3E50"),
    plot.subtitle = element_text(size = 11, hjust = 0.5, color = "#7F8C8D",
        margin = margin(b=10)),
    axis.title = element_text(face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom",
    panel.grid.major = element_line(color = "gray95"),
    panel.grid.minor = element_blank()
  ) +
  # Add value labels
  geom_text(
    aes(y = Percent_Growing, label = paste0(round(Percent_Growing, 1),
        "%")),
    vjust = -0.5,
    color = "#2C3E50",
    fontface = "bold",
    size = 4
  )

print(diverging_bar)
ggsave("diverging_bar_gradient.png", diverging_bar, width = 10, height = 7)
```

```
# Shows the balance/imbalance within each continent

# Calculate balance scores
story2_data <- population_data %>%
  group_by(Continent) %>%
  summarise(
    Growing = sum(Growth_Status == "Growing"),
    Declining = sum(Growth_Status == "Declining"),
    Total = n(),
    .groups = 'drop'
  ) %>%
  mutate(
    Balance_Ratio = Growing / Declining,
    Balance_Status = case_when(
      Balance_Ratio > 1.5 ~ "Strongly Growing",
      Balance_Ratio > 1 ~ "Moderately Growing",
      Balance_Ratio == 1 ~ "Perfectly Balanced",
      Balance_Ratio > 0.5 ~ "Moderately Declining",
      TRUE ~ "Strongly Declining"
    )
  )

# Create a diverging color gradient from green (growing) to red (declining)
balance_colors <- c("Strongly Growing" = "#1B5E20",
                    "Moderately Growing" = "#4CAF50",
                    "Perfectly Balanced" = "#FFC107",
                    "Moderately Declining" = "#F44336",
                    "Strongly Declining" = "#B71C1C")

story2_plot <- ggplot(story2_data,
                      aes(x = reorder(Continent, Balance_Ratio))) +

  # Growing side (positive)
  geom_col(aes(y = Growing), fill = "#4CAF50", alpha = 0.8, width = 0.6) +

  # Declining side (negative, shown as mirror)
  geom_col(aes(y = -Declining), fill = "#F44336", alpha = 0.8, width = 0.6) +

  # Zero line
  geom_hline(yintercept = 0, color = "#2C3E50", size = 1) +

  # Balance indicator points
  geom_point(aes(y = (Growing - Declining),
               color = Balance_Status,
               size = abs(Growing - Declining)),
             show.legend = FALSE) +

  scale_color_manual(values = balance_colors) +
  scale_size_continuous(range = c(5, 15)) +
```

```
# Labels for growing side
geom_text(
  aes(y = Growing, label = paste0("↑", Growing)),
  vjust = -0.5,
  color = "#1B5E20",
  fontface = "bold",
  size = 4
) +

# Labels for declining side
geom_text(
  aes(y = -Declining, label = paste0("↓", Declining)),
  vjust = 1.5,
  color = "#B71C1C",
  fontface = "bold",
  size = 4
) +

# Balance ratio labels
geom_text(
  aes(y = (Growing - Declining)/2,
      label = paste0("Ratio: ", round(Balance_Ratio, 2))),
  vjust = -0.5,
  color = "#2C3E50",
  fontface = "bold",
  size = 3.5
) +

# Titles and labels
labs(
  title = "GROWTH-DECLINE CONTINENTAL BALANCE",
  subtitle = "Positive values (green) = Growing countries\nNegative values
         (red) = Declining countries\nSize of dots indicates imbalance
         magnitude",
  x = "Continent (Ordered by Growth:Decline Ratio)",
  y = "Number of Countries\n(Growing ↑ / Declining ↓)",
  color = "Balance Status",
  caption = "Story: Shows growth-decline balance/imbalance within each
         continent"
) +

# Custom y-axis labels
scale_y_continuous(
  breaks = seq(-50, 50, by = 10),
  labels = function(x) ifelse(x < 0, paste0("↓", abs(x)),
                       ifelse(x > 0, paste0("↑", x), "0"))
) +

# Theme
```

```
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5, color =
         "#2C3E50"),
    plot.subtitle = element_text(size = 11, hjust = 0.5, color = "#7F8C8D",
         margin = margin(b=15)),
    axis.title = element_text(face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, face = "bold"),
    legend.position = "bottom",
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    plot.background = element_rect(fill = "white", color = NA)
  ) +

  coord_flip()  # Flip for better readability

print(story2_plot)
ggsave("story2_growth_balance_gradient.png", story2_plot, width = 12,
         height = 8, dpi = 300)
```

```
# Shows the hierarchy of growth dominance across continents


# Calculate growth hierarchy and remove "Other" category
story3_data <- population_data %>%
  filter(Continent != "Other") %>%  # Remove "Other" category
  group_by(Continent) %>%
  summarise(
    Total = n(),
    Growing = sum(Growth_Status == "Growing"),
    Declining = sum(Growth_Status == "Declining"),
    .groups = 'drop'
  ) %>%
  mutate(
    Growth_Ratio = Growing / (Growing + Declining),
    # Create hierarchy levels
    Hierarchy_Level = case_when(
      Growth_Ratio >= 0.8 ~ "Very High Growth",
      Growth_Ratio >= 0.6 ~ "High Growth",
      Growth_Ratio >= 0.4 ~ "Moderate Growth",
      Growth_Ratio >= 0.2 ~ "Low Growth",
      TRUE ~ "Very Low Growth"
    )
  ) %>%
  arrange(desc(Growth_Ratio))
```

```
print("Filtered Data (without 'Other' category):")
print(story3_data)

# Create a vibrant rainbow gradient for the hierarchy levels
hierarchy_colors <- c(
  "Very High Growth" = "#FF0000",    # Red (hottest)
  "High Growth" = "#FF8000",         # Orange
  "Moderate Growth" = "#FFFF00",     # Yellow
  "Low Growth" = "#00FF00",          # Green
  "Very Low Growth" = "#0000FF"      # Blue (coldest)
)


story3_plot <- ggplot(story3_data,
                      aes(x = reorder(Continent, -Growth_Ratio),
                          y = Total)) +

  # Background bar (total countries) - use subtle gradient
  geom_col(fill = "#F0F0F0", alpha = 0.3, width = 0.7) +

  # Growing countries (stacked on top) with vibrant gradient
  geom_col(aes(y = Growing, fill = Hierarchy_Level),
           width = 0.7, alpha = 0.9) +

  # Add subtle gradient overlay for more vibrant effect
  geom_col(aes(y = Growing),
           fill = "white",
           alpha = 0.1,
           width = 0.7) +

  # VIBRANT RAINBOW GRADIENT SCALE
  scale_fill_manual(
    values = hierarchy_colors,
    name = "Growth Level",
    guide = guide_legend(reverse = TRUE)  # Reverse legend to match bar
         order
  ) +

  # Add labels for growing countries - bold black for contrast
  geom_text(
    aes(y = Growing, label = paste0(Growing, "/", Total)),
    vjust = -0.5,
    color = "black",
    fontface = "bold",
    size = 4.5
  ) +

  # Add percentage labels inside bars - white for readability
  geom_text(
```

```
  aes(y = Growing/2,
      label = paste0(round(Growth_Ratio * 100, 0), "%")),
  color = "white",
  fontface = "bold",
  size = 5,
  alpha = 0.9
) +

# Titles and labels
labs(
  title = "CONTINENTAL GROWTH HIERARCHY",
  subtitle = "Vibrant rainbow gradient shows growth intensity across
         continents\nRed (Very High) → Orange → Yellow → Green → Blue
         (Very Low)",
  x = "Continent (Ranked by Growth Percentage)",
  y = "Number of Countries",
  caption = "Data Analysis: Population Growth Status by Continent"
) +

theme_minimal(base_size = 14) +
theme(
  plot.title = element_text(
    face = "bold",
    size = 20,
    hjust = 0.5,
    color = "#333333",
    margin = margin(b = 10)
  ),
  plot.subtitle = element_text(
    size = 13,
    hjust = 0.5,
    color = "#666666",
    margin = margin(b = 20)
  ),
  axis.title = element_text(
    face = "bold",
    size = 13,
    color = "#333333"
  ),
  axis.text.x = element_text(
    angle = 45,
    hjust = 1,
    vjust = 1,
    face = "bold",
    size = 12,
    color = "#333333"
  ),
  axis.text.y = element_text(
    size = 11,
    color = "#333333"
```

```
    ),
    legend.position = "bottom",
    legend.title = element_text(
      face = "bold",
      size = 12,
      color = "#333333"
    ),
    legend.text = element_text(
      size = 11,
      color = "#333333"
    ),
    legend.key.size = unit(1, "cm"),
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_line(color = "#E0E0E0", size = 0.5),
    plot.background = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "white", color = NA),
    plot.margin = margin(20, 20, 20, 20)
  ) +

  # Add subtle glow effect to bars
  geom_col(
    aes(y = Growing),
    fill = NA,
    color = "white",
    size = 1,
    width = 0.71,
    alpha = 0.3
  ) +

  # Ensure y-axis shows proper range
  scale_y_continuous(
    expand = expansion(mult = c(0, 0.15)),
    breaks = seq(0, 60, by = 10),
    limits = c(0, max(story3_data$Total) * 1.15)
  )

print(story3_plot)
ggsave("vibrant_growth_hierarchy_gradient.png", story3_plot, width = 12,
       height = 8, dpi = 300)


# Shows different distribution patterns across continents

# First, let's create population categories (but we won't use population data)
# We'll just use the counts to create interesting patterns

story4_data <- population_data %>%
  group_by(Continent) %>%
  summarise(
```

```
    Growing = sum(Growth_Status == "Growing"),
    Declining = sum(Growth_Status == "Declining"),
    Total = n(),
    .groups = 'drop'
  ) %>%
  mutate(
    # Create pattern categories based on distribution
    Pattern_Type = case_when(
      Growing >= Declining * 2 ~ "Heavily Growing",
      Growing > Declining ~ "Moderately Growing",
      Growing == Declining ~ "Balanced",
      Declining > Growing ~ "Moderately Declining",
      Declining >= Growing * 2 ~ "Heavily Declining"
    )
  ) %>%
  arrange(desc(Growing))

# Create pattern gradient colors
pattern_colors <- c(
  "Heavily Growing" = "#00441B",     # Very dark green
  "Moderately Growing" = "#238B45",  # Medium green
  "Balanced" = "#FFD700",            # Gold
  "Moderately Declining" = "#CB181D", # Medium red
  "Heavily Declining" = "#67000D"    # Very dark red
)

# Create a stacked bar with pattern overlay effect
story4_plot <- ggplot(story4_data) +

  # Base bars (Total)
  geom_col(aes(x = reorder(Continent, -Total), y = Total),
           fill = "#F0F0F0", width = 0.7, alpha = 0.5) +

  # Declining countries (bottom stack)
  geom_col(aes(x = reorder(Continent, -Total), y = Declining),
           fill = "#FF6B6B", width = 0.7, alpha = 0.7) +

  # Growing countries (top stack)
  geom_col(aes(x = reorder(Continent, -Total), y = Growing),
           fill = "#4ECDC4", width = 0.7, alpha = 0.7,
           position = position_nudge(y = story4_data$Declining)) +

  # Pattern overlay based on Pattern_Type
  geom_tile(
    aes(x = as.numeric(reorder(Continent, -Total)),
        y = Total/2,
        width = 0.5,
        height = Total,
        fill = Pattern_Type),
    alpha = 0.2
```

```
) +

scale_fill_manual(values = pattern_colors, name = "Distribution Pattern") +

# Add connecting lines to show the split
geom_segment(
  aes(x = as.numeric(reorder(Continent, -Total)) - 0.35,
      xend = as.numeric(reorder(Continent, -Total)) + 0.35,
      y = Declining,
      yend = Declining),
  color = "gray40",
  linetype = "dashed",
  size = 0.5
) +

# Labels for growing
geom_text(
  aes(x = reorder(Continent, -Total),
      y = Declining + Growing/2,
      label = paste0("G:", Growing)),
  color = "white",
  fontface = "bold",
  size = 4
) +

# Labels for declining
geom_text(
  aes(x = reorder(Continent, -Total),
      y = Declining/2,
      label = paste0("D:", Declining)),
  color = "white",
  fontface = "bold",
  size = 4
) +

# Pattern type labels
geom_text(
  aes(x = reorder(Continent, -Total),
      y = Total + max(story4_data$Total) * 0.05,
      label = Pattern_Type),
  color = "#2C3E50",
  fontface = "bold",
  size = 3.5
) +

# Titles and labels
labs(
  title = "GROWTH DISTRIBUTION PATTERNS",
```

```
  subtitle = "Shows different distribution patterns of growth across
      continents\nBlue = Growing, Red = Declining, Pattern overlay shows
      distribution type",
  x = "Continent (Ordered by Total Countries)",
  y = "Number of Countries",
  caption = "Story: Different patterns of growth distribution across
      continents"
 ) +

 theme_minimal(base_size = 13) +
 theme(
  plot.title = element_text(face = "bold", size = 18, hjust = 0.5, color =
      "#2C3E50"),
  plot.subtitle = element_text(size = 12, hjust = 0.5, color = "#7F8C8D",
      margin = margin(b = 15)),
  axis.title = element_text(face = "bold"),
  axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, face = "bold",
      size = 11),
  legend.position = "bottom",
  panel.grid.major.x = element_blank(),
  panel.grid.minor = element_blank(),
  plot.background = element_rect(fill = "white", color = NA)
 ) +

 scale_y_continuous(expand = expansion(mult = c(0, 0.15)))

print(story4_plot)
ggsave("story4_growth_patterns_gradient.png", story4_plot, width = 11,
      height = 8, dpi = 300)


# Test Analysis

# Prepare data for analysis (remove "Other" category)
population_data_main <- population_data[population_data$Continent !=
      "Other", ]
population_data_main$Continent <-
      droplevels(as.factor(population_data_main$Continent))

# Create final contingency table
final_table <- table(population_data_main$Continent,
      population_data_main$Growth_Status)
cat("Final Contingency Table for Analysis:\n")
print(final_table)

cat("\n--- Statistical Test Selection ---\n")
cat("Test selected: Chi-square Test of Independence\n")
cat("Reason: Both variables are nominal (categorical):\n")
cat("  - Independent variable: Continent (nominal, >2 categories)\n")
cat("  - Dependent variable: Growth Status (nominal, 2 categories)\n")
```

```
cat("  - Purpose: Test if proportions are independent of groups\n")

# Perform Chi-square test
cat("\n--- Chi-square Test Output ---\n")
chi_test <- chisq.test(final_table)

# Display detailed test results
cat("Chi-square Test Results:\n")
cat(sprintf("  χ² statistic: %.3f\n", chi_test$statistic))
cat(sprintf("  Degrees of freedom: %d\n", chi_test$parameter))
cat(sprintf("  P-value: %.6f\n", chi_test$p.value))
cat(sprintf("  Significance level (α): 0.05\n"))

# Calculate and display expected frequencies
cat("\nExpected Frequencies (if no association):\n")
print(round(chi_test$expected, 2))

# Calculate effect size (Cramer's V)
n <- sum(final_table)
k <- min(dim(final_table)) - 1
cramers_v <- sqrt(chi_test$statistic / (n * k))
cat(sprintf("\nEffect Size (Cramer's V): %.3f\n", cramers_v))

# Interpretation of effect size
cat("Effect size interpretation: ")
if(cramers_v >= 0.5) {
  cat("Large effect\n")
} else if(cramers_v >= 0.3) {
  cat("Medium effect\n")
} else if(cramers_v >= 0.1) {
  cat("Small effect\n")
} else {
  cat("Negligible effect\n")
}


# The null hypothesis rejected/not rejected

# Restate the hypotheses
cat("Research Question:",
    "'Is there a difference in the proportion of countries with growing
         populations among different continents?'\n\n")

cat("Null Hypothesis (H₀):",
    "'There is NO difference in the proportion of countries with growing
         populations among different continents.'\n\n")

cat("Alternative Hypothesis (H₁):",
    "'There IS a difference in the proportion of countries with growing
         populations among different continents.'\n\n")
```

```
# Make the statistical decision
alpha <- 0.05
cat("--- Statistical Decision ---\n")
cat(sprintf("P-value: %.6f\n", chi_test$p.value))
cat(sprintf("Alpha level: %.3f\n", alpha))

if(chi_test$p.value < alpha) {
  cat("\n✅ DECISION: REJECT the null hypothesis (H₀)\n")
  cat("   Reason: p-value (", format(chi_test$p.value, scientific = TRUE),
      ") < α (", alpha, ")\n", sep = "")
  cat("   Conclusion: There IS a statistically significant difference in the\n")
  cat("   proportion of growing populations among different continents.\n")

  # Identify which continents contribute most to the difference
  cat("\n--- Post-hoc Analysis: Standardized Residuals ---\n")
  cat("(Values > |2| indicate significant departure from expected)\n")
  std_res <- round(chi_test$stdres, 2)
  print(std_res)

} else {
  cat("\n✅ DECISION: FAIL TO REJECT the null hypothesis (H₀)\n")
  cat("   Reason: p-value (", format(chi_test$p.value, scientific = TRUE),
      ") ≥ α (", alpha, ")\n", sep = "")
  cat("   Conclusion: There is NO statistically significant difference in the\n")
  cat("   proportion of growing populations among different continents.\n")
}

cat(" Statistical test: Chi-square test of independence\n")
cat(sprintf(" Test results: χ²(%d) = %.2f, p = %.4f\n",
            chi_test$parameter, chi_test$statistic, chi_test$p.value))
cat(sprintf(" Effect size: Cramer's V = %.3f\n", cramers_v))

if(chi_test$p.value < 0.05) {
  cat(" Decision: Reject the null hypothesis (H₀)\n")
} else {
  cat(" Decision: Fail to reject the null hypothesis (H₀)\n")
}
```

# B. GitHub log output.

| Commits | Username | Email |
|---|---|---|
| 14 | Coud9Abhishek | aa25anh@herts.ac.uk |
| 3 | Govardhankk04 | gk25aai@herts.ac.uk |
| 13 | LeslieLucieKonlack | lk25aan@herts.ac.uk |
| 22 | NahedulBarChowdhury | mr25aei@herts.ac.uk |
| 19 | zisanahmed1227 | za25aeq@herts.ac.uk |