Statistics Chapter 2 Review - Exercises 10-15

International Vital Statistics Analysis

Lucas Liona

Table of contents

Introduction

This analysis examines international vital statistics from Table 9, which contains demographic data on birth rates, death rates, life expectancy, and other measures for countries grouped into four categories: industrialized nations, African nations, Southeast Asian countries, and Central/South American countries.

Exercise 10: Birth Rates in Industrialized Countries

Construct a labeled dot plot of the birth rates ("both" column) for the industrialized countries. Also construct a bar graph (use a sensible ordering). What, if anything, does the information in these graphs suggest to you?

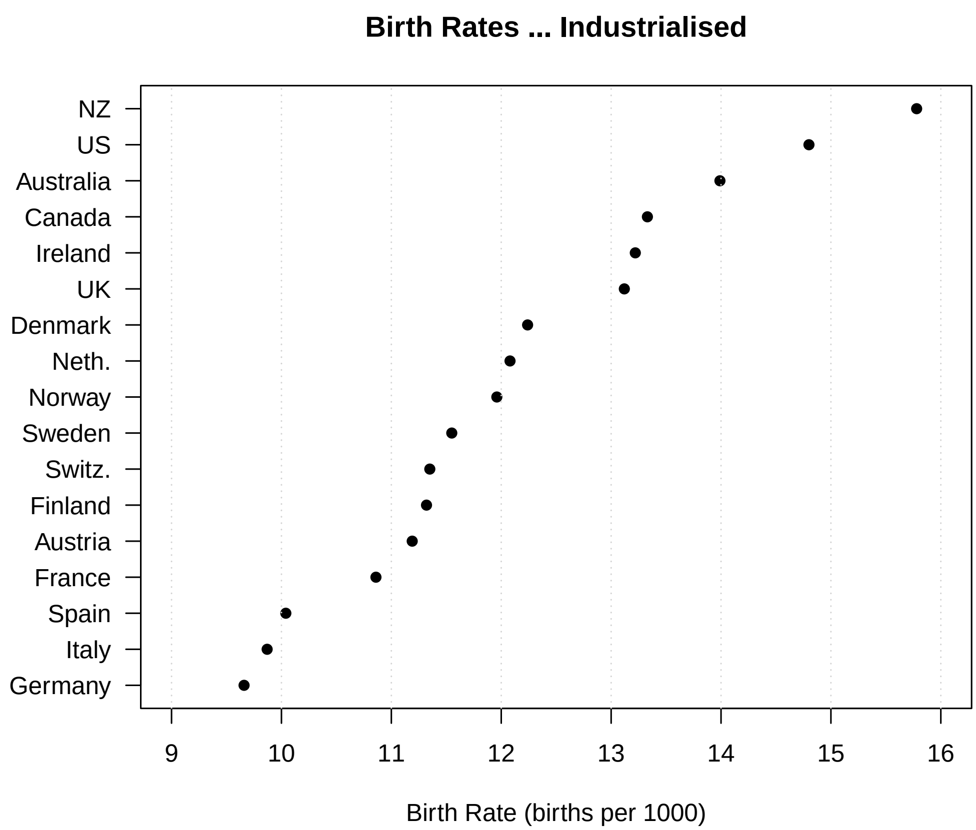
```r
# Create vectors for birth rates of industrialized countries
countries <- c("Austria", "Denmark", "Finland", "France", "Germany", "Ireland",
               "Italy", "Neth.", "Spain", "Sweden", "Switz.", "UK", "US",
               "Australia", "Canada", "NZ", "Norway")
birth_rates <- c(11.19, 12.24, 11.32, 10.86, 9.66, 13.22, 9.87, 12.08, 10.04,
                 11.55, 11.35, 13.12, 14.80, 13.99, 13.33, 15.78, 11.96)

# Reorder data by birth rate
ordered_indices <- order(birth_rates)
countries_ordered <- countries[ordered_indices]
birth_rates_ordered <- birth_rates[ordered_indices]

# Create dot plot
par(mar = c(5, 10, 4, 2))  # Adjust margins
plot(birth_rates_ordered, 1:length(countries_ordered),
     xlim = c(9, 16),
     yaxt = "n",
     xlab = "Birth Rate (births per 1000)",
     ylab = "",
     main = "Birth Rates – Industrialised",
     pch = 16)
```

```
Warning in title(...): conversion failure on 'Birth Rates – Industrialised' in
'mbcsToSbcs': dot substituted for <e2>
```

```
Warning in title(...): conversion failure on 'Birth Rates – Industrialised' in
'mbcsToSbcs': dot substituted for <80>
```

```
Warning in title(...): conversion failure on 'Birth Rates – Industrialised' in
'mbcsToSbcs': dot substituted for <94>
```

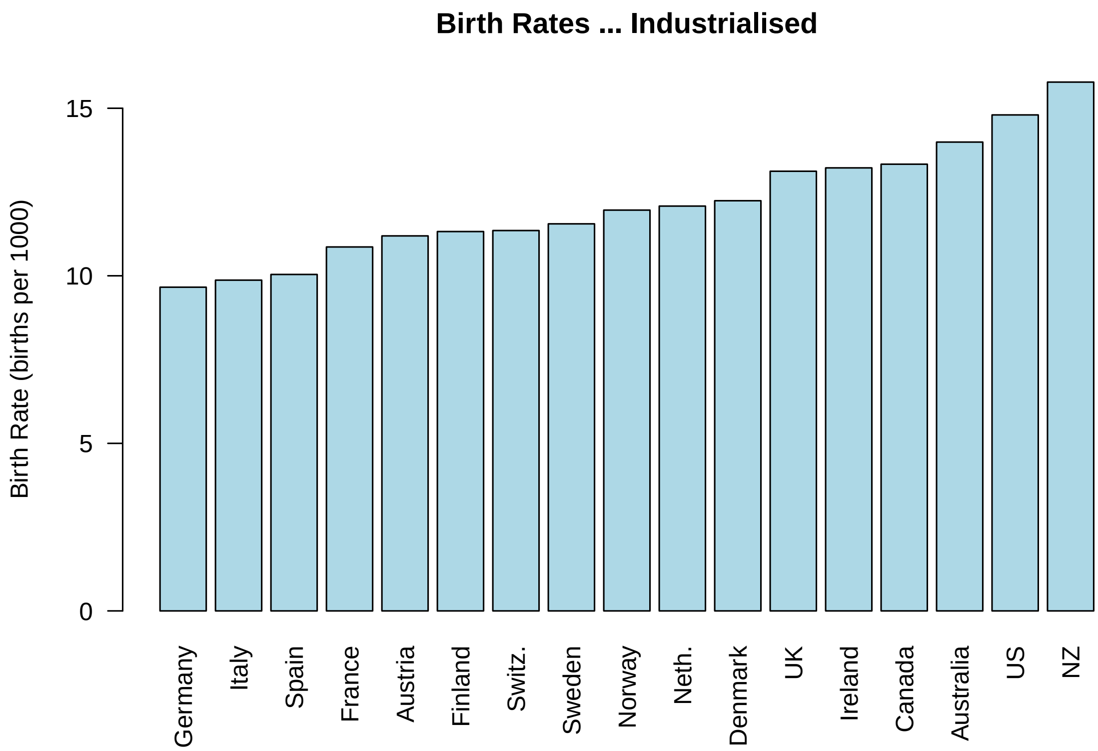
```r
axis(2, at = 1:length(countries_ordered), labels = countries_ordered, las = 2)
grid(nx = NULL, ny = NA)
```




```r
# Create bar graph
par(mar = c(8, 4, 4, 2))  # Adjust margins
barplot(birth_rates_ordered,
        names.arg = countries_ordered,
        main = "Birth Rates – Industrialised",
        xlab = "",
        ylab = "Birth Rate (births per 1000)",
        col = "lightblue",
        las = 2)
```

```
Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Birth Rates – Industrialised' in 'mbcsToSbcs': dot
substituted for <e2>

Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Birth Rates – Industrialised' in 'mbcsToSbcs': dot
substituted for <80>
```

```
Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Birth Rates - Industrialised' in 'mbcsToSbcs': dot
substituted for <94>
```



From the dot plot and bar graph, we can observe that the birth rates among industrialized countries tend to form clusters. The European countries (particularly Germany, Italy, and Spain) have lower birth rates than the Scandinavian countries. The United Kingdom and Canada form another group, while Australia, the United States, and New Zealand have higher birth rates. The bar graph helps visualize the relative sizes more clearly, showing that the lowest rate (Germany at about 9.66) is roughly 2/3 as large as the highest rate (New Zealand at 15.78).

Exercise 11: Life Expectancies in African Countries

Repeat the previous exercise for the life expectancies for the African countries.

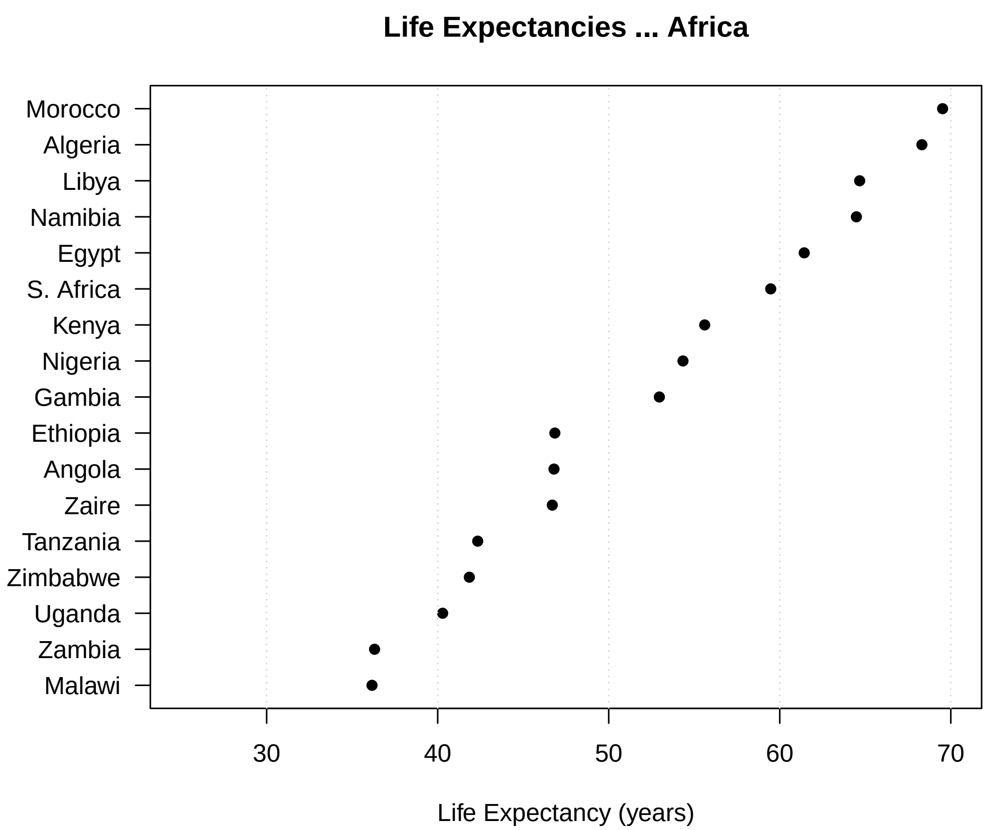
```r
# Create vectors for life expectancies of African countries
african_countries <- c("Algeria", "Angola", "Egypt", "Ethiopia", "Gambia",
                       "Kenya", "Libya", "Malawi", "Morocco", "Namibia",
                       "Nigeria", "S. Africa", "Tanzania", "Uganda",
                       "Zaire", "Zambia", "Zimbabwe")
life_exp <- c(68.31, 46.80, 61.43, 46.85, 52.96, 55.61, 64.67, 36.16,
             69.52, 64.48, 54.34, 59.47, 42.34, 40.29, 46.70, 36.31, 41.85)

# Reorder data by life expectancy
ordered_indices <- order(life_exp)
african_countries_ordered <- african_countries[ordered_indices]
life_exp_ordered <- life_exp[ordered_indices]

# Create dot plot
par(mar = c(5, 10, 4, 2))  # Adjust margins
plot(life_exp_ordered, 1:length(african_countries_ordered),
     xlim = c(25, 70),
     yaxt = "n",
     xlab = "Life Expectancy (years)",
     ylab = "",
     main = "Life Expectancies – Africa",
     pch = 16)
```

```
Warning in title(...): conversion failure on 'Life Expectancies – Africa' in
'mbcsToSbcs': dot substituted for <e2>
```

```
Warning in title(...): conversion failure on 'Life Expectancies – Africa' in
'mbcsToSbcs': dot substituted for <80>
```

```
Warning in title(...): conversion failure on 'Life Expectancies – Africa' in
'mbcsToSbcs': dot substituted for <94>
```

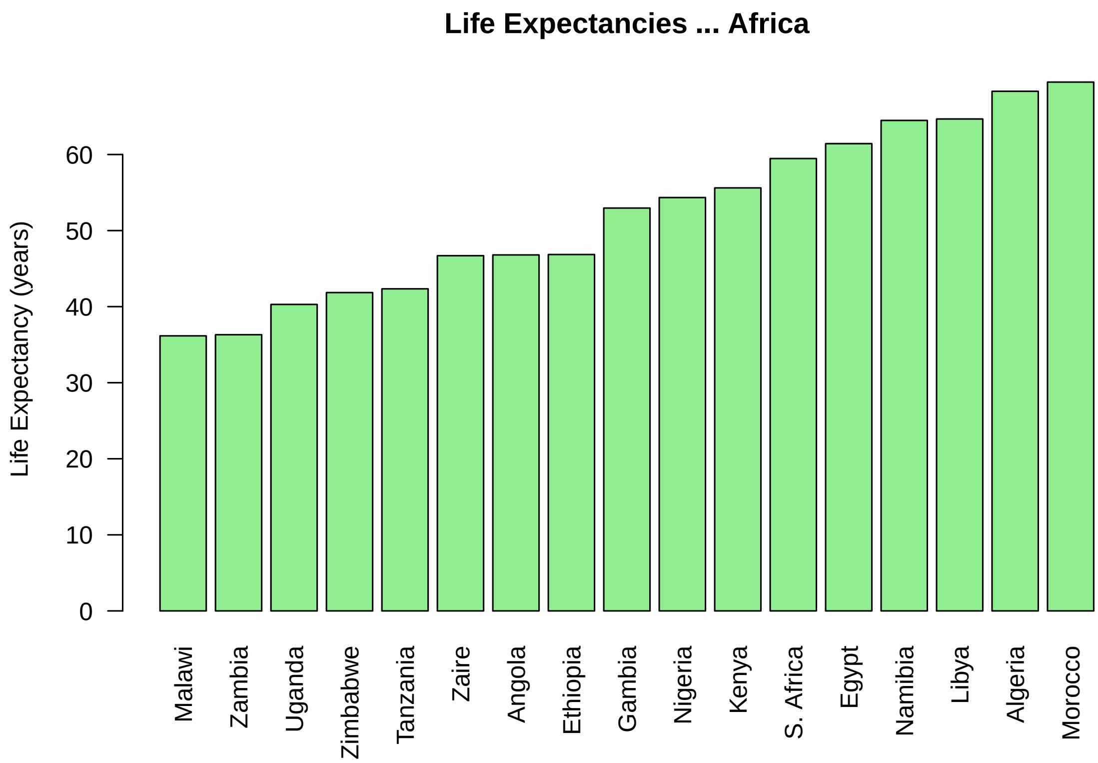
```r
axis(2, at = 1:length(african_countries_ordered), labels = african_countries_ordered, las = 2
grid(nx = NULL, ny = NA)
```




```r
# Create bar graph
par(mar = c(8, 4, 4, 2))  # Adjust margins
barplot(life_exp_ordered,
        names.arg = african_countries_ordered,
        main = "Life Expectancies – Africa",
        xlab = "",
        ylab = "Life Expectancy (years)",
        col = "lightgreen",
        las = 2)
```

```
Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Life Expectancies – Africa' in 'mbcsToSbcs': dot
substituted for <e2>

Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Life Expectancies – Africa' in 'mbcsToSbcs': dot
substituted for <80>
```

```
Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
conversion failure on 'Life Expectancies – Africa' in 'mbcsToSbcs': dot
substituted for <94>
```



There is an enormous range in life expectancy among African countries, with the highest being Morocco (69.5 years) and the lowest Malawi (36.2 years). The data are fairly well spread out with hints of small clusters of two or three countries about every five years. These clusters could be studied to see if there are any common features related to geography, economic development, or healthcare systems that might explain the groupings.

Exercise 12: Comparing Male and Female Life Expectancies

Part (a): Comparison Using Different Plot Types

Compare male with female life expectancies for the industrialized countries using: (i) stem-and-leaf plots (ii) dot plots (iii) bar graphs

```r
# Create vectors for male and female life expectancies
industrialized_countries <- c("Australia", "Austria", "Canada", "Denmark",
                              "Finland", "France", "Germany", "Ireland",
                              "Italy", "Neth.", "NZ", "Norway", "Spain",
                              "Sweden", "Switz.", "UK", "US")
male_life <- c(76.44, 73.38, 75.67, 73.78, 73.82, 74.47, 72.80, 72.88,
              74.85, 74.91, 73.96, 74.63, 74.95, 75.62, 74.58, 73.78, 72.65)
female_life <- c(82.50, 79.84, 82.65, 81.01, 77.18, 82.46, 79.27, 78.46,
                81.48, 80.68, 80.21, 80.61, 81.81, 80.63, 80.82, 79.17, 79.41)

# (i) Stem-and-leaf plots
cat("MALES\n")
```

```
MALES
```

```r
stem(male_life, scale = 1)
```

```

  The decimal point is at the |

  72 | 789
  73 | 4888
  74 | 056699
  75 | 067
  76 | 4
```

```r
cat("\nFEMALES\n")
```

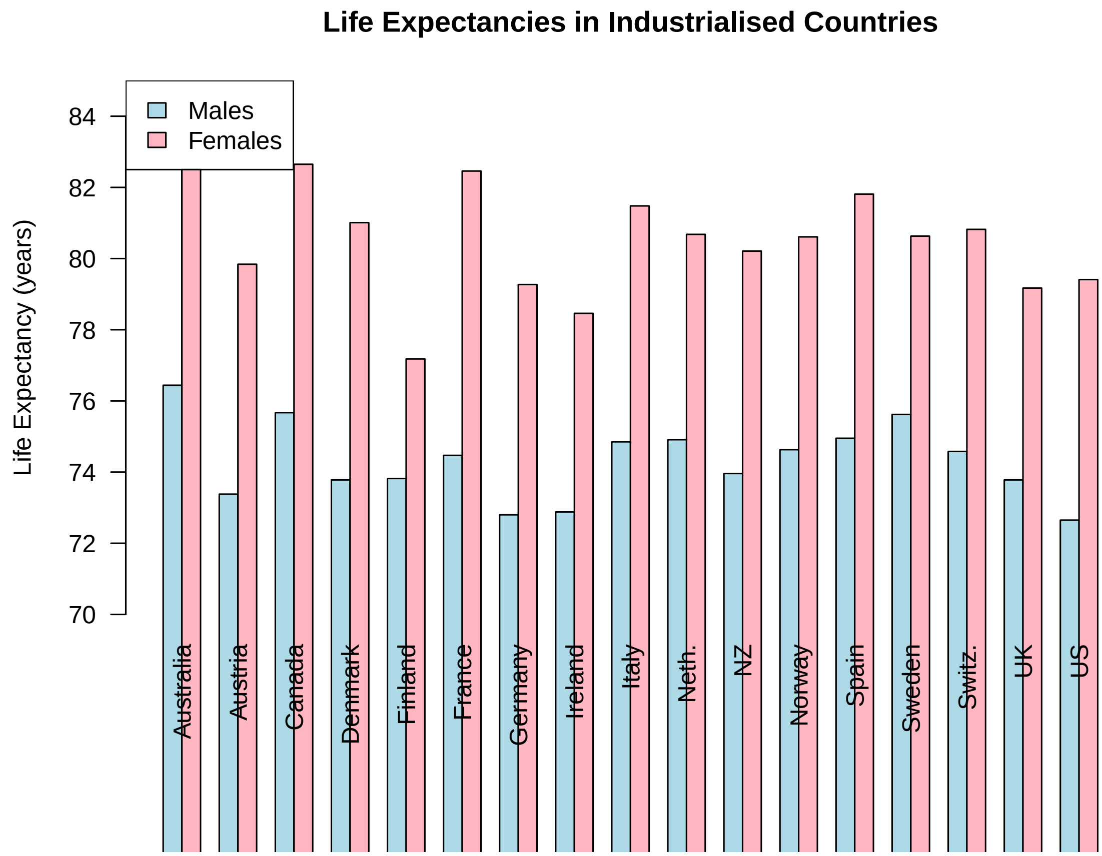
```

FEMALES
```

```r
stem(female_life, scale = 1)
```

```

  The decimal point is at the |

  76 | 2
  78 | 52348
  80 | 26678058
  82 | 557
```

```r
# (ii) Dot plots
par(mfrow = c(2, 1), mar = c(4, 4, 2, 1))
# Male life expectancy
plot(male_life, rep(1, length(male_life)),
     xlim = c(70, 84),
     ylim = c(0.5, 1.5),
     main = "Life Expectancies in Industrialised Countries",
     xlab = "Life Expectancy (years)",
     ylab = "",
     yaxt = "n",
     pch = 16)
text(70, 1, "Males", pos = 2)
# Female life expectancy
plot(female_life, rep(1, length(female_life)),
     xlim = c(70, 84),
     ylim = c(0.5, 1.5),
     xlab = "Life Expectancy (years)",
     ylab = "",
     yaxt = "n",
     pch = 16)
text(70, 1, "Females", pos = 2)
```


```
# (iii) Bar graphs
par(mfrow = c(1, 1), mar = c(8, 4, 4, 2))
# Prepare data for side-by-side bars
life_exp_data <- rbind(male_life, female_life)
colnames(life_exp_data) <- industrialized_countries
# Create bar graph
barplot(life_exp_data,
        beside = TRUE,
        main = "Life Expectancies in Industrialised Countries",
        xlab = "",
        ylab = "Life Expectancy (years)",
        col = c("lightblue", "lightpink"),
        las = 2,
        ylim = c(70, 85))
legend("topleft", legend = c("Males", "Females"),
       fill = c("lightblue", "lightpink"))
```



Part (b): Modified Plots to Compare Countries

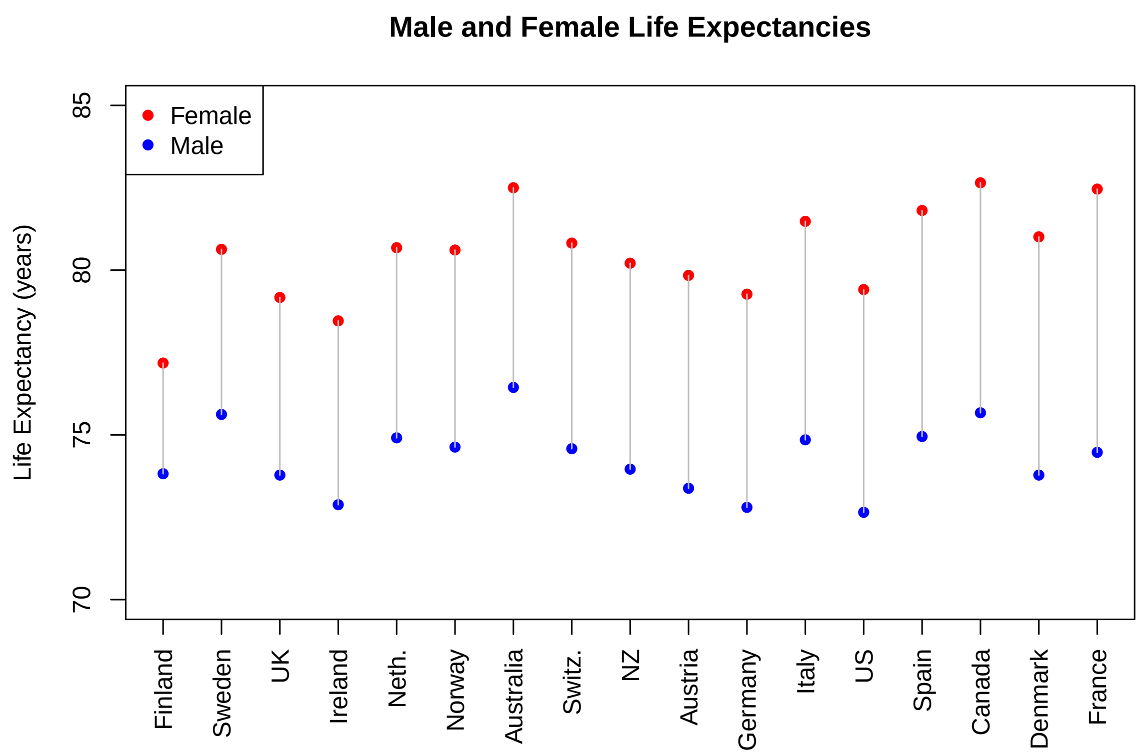
Design modifications of each of these types of plot that would also enable you to compare countries.

For the stem-and-leaf plots, we could create a back-to-back stem-and-leaf plot where male values are on the left and female values on the right, with the country code indicated next to each leaf.

For dot plots, we could create a line-linked dot plot where each country has two dots (one for male, one for female) connected by a line, with countries ordered by a meaningful variable like the gap between female and male life expectancy.

```
# Create a line-linked dot plot
# Calculate the gap between female and male life expectancy
gap <- female_life - male_life
# Order countries by gap
```

```r
ordered_indices <- order(gap)
countries_ordered <- industrialized_countries[ordered_indices]
male_ordered <- male_life[ordered_indices]
female_ordered <- female_life[ordered_indices]
gap_ordered <- gap[ordered_indices]

# Plot
par(mar = c(8, 4, 4, 2))
plot(1:length(countries_ordered), male_ordered,
     type = "n",
     ylim = c(70, 85),
     xlab = "",
     ylab = "Life Expectancy (years)",
     main = "Male and Female Life Expectancies",
     xaxt = "n")
axis(1, at = 1:length(countries_ordered), labels = countries_ordered, las = 2)

# Add points and connecting lines
for(i in 1:length(countries_ordered)) {
  points(i, male_ordered[i], col = "blue", pch = 16)
  points(i, female_ordered[i], col = "red", pch = 16)
  lines(c(i, i), c(male_ordered[i], female_ordered[i]), col = "gray")
}

# Add legend
legend("topleft", legend = c("Female", "Male"),
       col = c("red", "blue"), pch = 16)
```



For bar graphs, we could create a grouped bar chart with countries on the x-axis and use different color bars for males and females, ordered by the gender gap in life expectancy.

Exercise 13: Analyzing Life Expectancy Differences

It is clear from the data that female life expectancies tend to be greater than male life expectancies, and you might ask how much greater.

Part (a): Understanding the Difference Measure

Consider the difference diff = female life expectancy - male life expectancy. If a country has value 3 for diff, what does this tell us? (Explain in words.) The value 0 is special. Why? What do values greater and less than 0 correspond to?

If a country has a value of 3 for the difference between female and male life expectancy, this means that women in that country live, on average, 3 years longer than men.

The value 0 is special because it represents equality - it would mean that men and women have the same average life expectancy in that country. In most countries, this doesn't occur due to biological and social factors.

Values greater than 0 (which is what we typically see) correspond to countries where women live longer than men on average. Values less than 0 would indicate countries where men outlive women on average, which is quite rare globally.

Part (b): Understanding the Ratio Measure

Consider the ratio ratio = female life expectancy/male life expectancy. If a country has value 1.2 for ratio, what does this tell us? (Explain in words.) The value 1 is special. Why? What do values greater and less than 1 correspond to?

If a country has a value of 1.2 for the ratio of female to male life expectancy, this means that women in that country live, on average, 20% longer than men (since 1.2 = 120% of 1.0).

The value 1 is special because it represents equality - it would mean that men and women have identical average life expectancies in that country.

Values greater than 1 (which is what we typically observe) correspond to countries where women live longer than men on average. Values less than 1 would indicate countries where men outlive women, which is uncommon globally.

Exercise 14: Comparative Analysis of International Vital Statistics

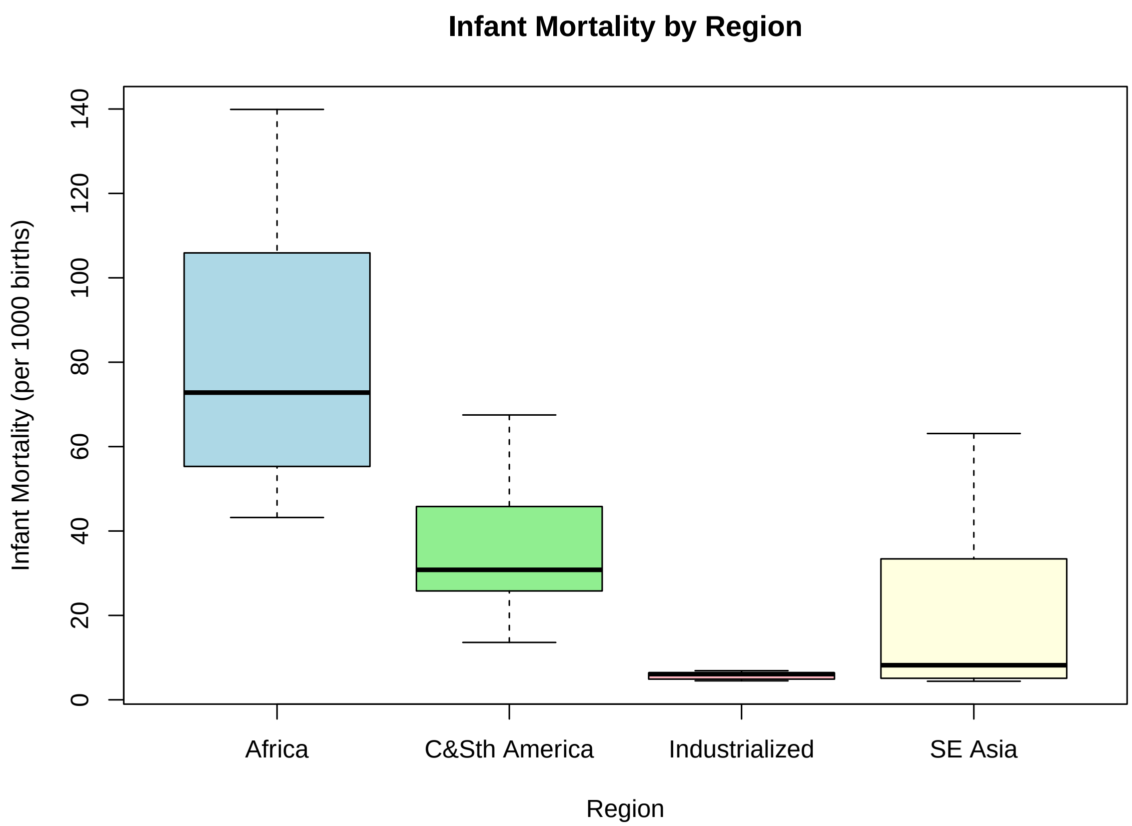
Apply the following steps to some of the following variables: (i) infant mortality (both) (ii) female infant mortality divided by male infant mortality
(iii) life expectancy (both) (iv) the difference between female life expectancy and male life expectancy (v) birthrate (vi) death rate

Part (a): Infant Mortality Comparison

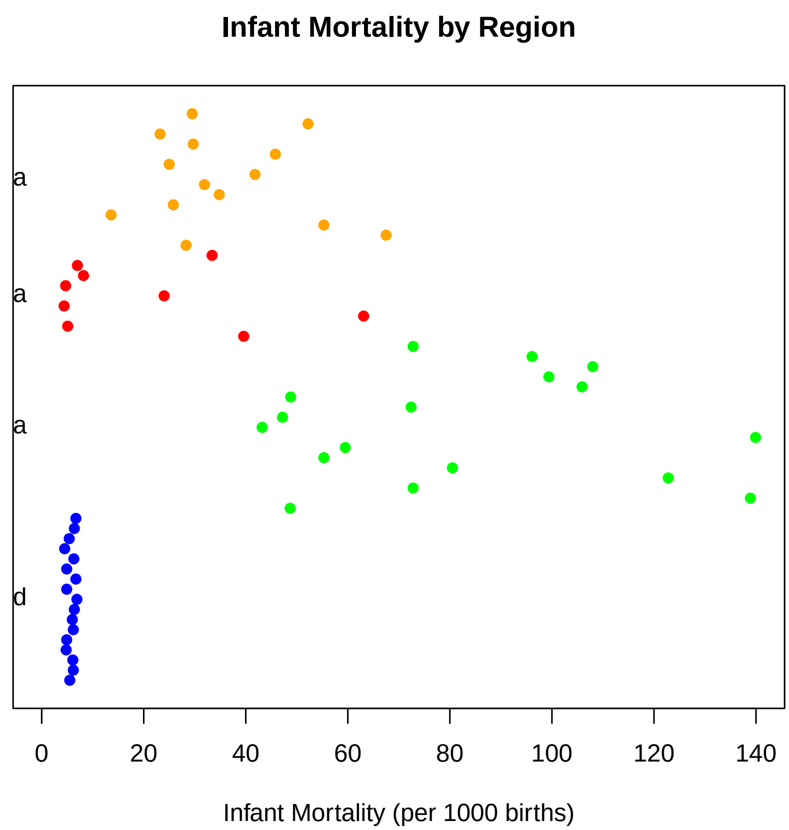
Construct appropriate plots that allow you to compare the groups of countries.

```r
# Create data
country_types <- c(rep("Industrialized", 17), rep("Africa", 17),
                   rep("SE Asia", 9), rep("C&Sth America", 14))
# Sample data (example values)
infant_mortality <- c(
  # Industrialized
  5.5, 6.2, 6.1, 4.8, 4.9, 6.2, 6.0, 6.4, 6.9, 4.9, 6.7, 4.9, 6.3, 4.5, 5.4, 6.4, 6.7,
```

```
  # Africa
  48.7, 138.9, 72.8, 122.8, 80.5, 55.3, 59.5, 139.9, 43.2, 47.2, 72.4, 48.8, 105.9, 99.4, 108
  # SE Asia
  39.6, 5.1, 63.1, 4.4, 24.0, 4.7, 8.2, 7.0, 33.4,
  # C&Sth America
  28.3, 67.5, 55.3, 13.6, 25.8, 34.8, 31.9, 41.8, 25.0, 45.8, 29.7, 23.2, 52.2, 29.5
)

# Create a boxplot
boxplot(infant_mortality ~ country_types,
        main = "Infant Mortality by Region",
        xlab = "Region",
        ylab = "Infant Mortality (per 1000 births)",
        col = c("lightblue", "lightgreen", "lightpink", "lightyellow"))
```



```r
# Create a dot plot
# Define groups
group_indices <- list(
  industrialized = 1:17,
  africa = 18:34,
  se_asia = 35:43,
  c_s_america = 44:57
)

# Set up plot
par(mar = c(5, 12, 4, 2))
plot(infant_mortality, 1:length(infant_mortality),
     type = "n",
     xlim = c(0, 140),
     ylim = c(0.5, length(infant_mortality) + 0.5),
     xlab = "Infant Mortality (per 1000 births)",
     ylab = "",
     yaxt = "n",
     main = "Infant Mortality by Region")

# Add group labels
text(0, mean(group_indices$industrialized), "Industrialized", pos = 2)
text(0, mean(group_indices$africa), "Africa", pos = 2)
text(0, mean(group_indices$se_asia), "SE Asia", pos = 2)
text(0, mean(group_indices$c_s_america), "C&Sth America", pos = 2)

# Add points
points(infant_mortality[group_indices$industrialized],
      group_indices$industrialized, pch = 16, col = "blue")
points(infant_mortality[group_indices$africa],
      group_indices$africa, pch = 16, col = "green")
points(infant_mortality[group_indices$se_asia],
      group_indices$se_asia, pch = 16, col = "red")
points(infant_mortality[group_indices$c_s_america],
      group_indices$c_s_america, pch = 16, col = "orange")
```



Part (b): Observations from the Plots

The infant mortality rate is low and is much the same for all the industrialized nations; it is much higher and very variable for the African countries, with very high rates for some countries; and it is low for some SE Asian countries, but higher and more variable for others. The rate is quite variable for Central and South America, though lower than the rates of many African countries.

Part (c): Notable Outliers and Country Patterns

The rates are particularly high for the African countries Malawi (140), Angola (139) and Ethiopia (123), and lowest for Morocco (43). They are low for the Asian countries Hong Kong (5.1), Japan (4.4) and Singapore (4.7), which may be grouped with the industrialized countries, and the rate is high for Indonesia (63). The rate is also high for the South American country Bolivia (68).

Part (d): Potential Explanations

The rate appears to be linked to the degree to which a country is industrialized. The wealthier countries might be expected to have better health care for pregnant mothers, and therefore lower rates. The relationship between infant mortality and economic development is well-established, with factors like healthcare access, sanitation, nutrition, and education all playing important roles.

Exercise 15: Miscellaneous Questions

Part (a): Factors Affecting Birth Rates

Birth rate is defined as the number of births per 1000 of population. What factors affect a country's birth rate?

Birth rate is affected by a complex mix of social, economic, and cultural factors. The age structure of the population is a critical factor - countries with more women of childbearing age will naturally have higher birth rates. The availability of birth control methods directly impacts family planning capabilities. Infant mortality rates also play a role, as higher infant mortality often correlates with higher birth rates (as families have more children expecting some may not survive).

Other important factors include cultural and religious attitudes toward family size, women's education and employment opportunities, government policies regarding childbearing, economic development level, and access to healthcare. All these factors are interrelated, making birth rates a multifaceted demographic indicator.

Part (b): Death Rates in Industrialized vs. Developing Countries

The death rates of the industrialized countries appear to be higher on average than those of the Asian and Central and South American countries. Can you suggest any reasons why this might be?

This seemingly counterintuitive pattern has a logical explanation related to population age structure. Industrialized countries generally have higher life expectancies, which means they have more elderly people in their populations. Since mortality rates increase dramatically with age, countries with older populations tend to have higher overall death rates.

In contrast, many Asian and Central/South American countries have younger population profiles due to higher birth rates in recent decades and lower life expectancies. While these countries might have higher mortality rates among specific age groups (particularly infants and young children), their overall death rates appear lower because a greater percentage of their population is young.

Another factor is cause of death patterns. Industrialized countries have largely controlled infectious diseases that still affect developing nations, but they experience higher rates of chronic diseases like heart disease and cancer that typically affect older individuals. The complexity arises because high death rates in industrialized countries reflect successful healthcare systems that enable people to live to older ages, where death becomes more common.

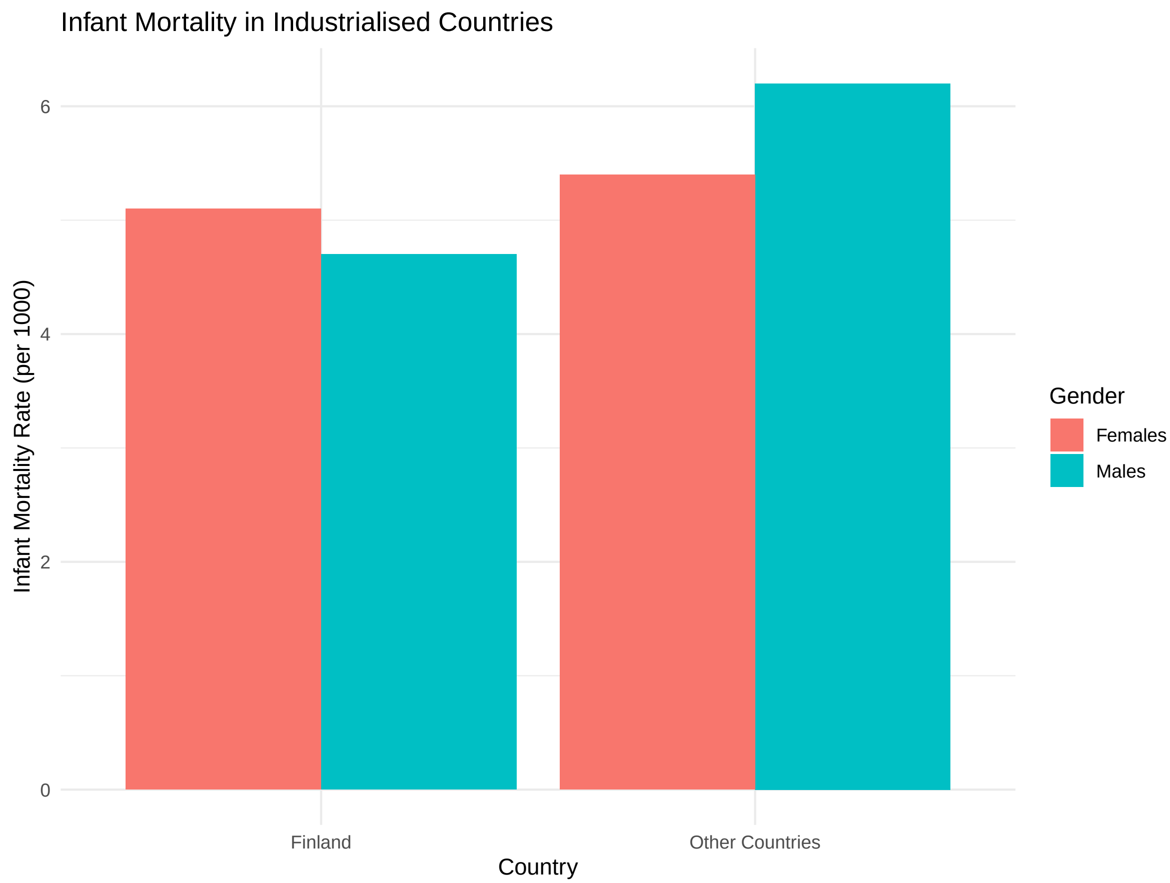
Part (c): Finland's Unusual Infant Mortality Ratio

In the industrialized group, the female-to-male ratio of infant mortalities for Finland is highly unusual. Is this because the individual infant mortality rates are highly unusual?

Finland's situation is indeed distinctive in two respects. First, Finland has the lowest male infant mortality rate among the industrialized countries. Second, and more unusually, Finland's female infant mortality rate (5.1) is higher than its male rate (4.7), giving a female-to-male ratio greater than 1. This pattern contradicts the biological norm seen in virtually all other countries, where male infants typically have higher mortality rates than females.

Looking at the raw numbers, Finland's overall infant mortality rate is comparable to other industrialized nations (4.9 per 1000), but the gender distribution is inverted. This suggests that the unusual ratio isn't due to extremely high or low values overall, but rather to the specific relationship between male and female rates. This could potentially be due to statistical anomalies in a small population, differences in reporting methods, or possibly some cultural or healthcare factors unique to Finland.

```r
# Create a visualization of Finland's unusual pattern
industrialized <- c("Finland", "Other Countries")
male_rates <- c(4.7, 6.2)  # Average of others
female_rates <- c(5.1, 5.4)  # Average of others

# Create data frame
finland_data <- data.frame(
  Country = rep(industrialized, each = 2),
  Gender = rep(c("Males", "Females"), times = 2),
  Rate = c(male_rates[1], female_rates[1], male_rates[2], female_rates[2])
)

# Plot
library(ggplot2)
ggplot(finland_data, aes(x = Country, y = Rate, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Infant Mortality in Industrialised Countries",
       y = "Infant Mortality Rate (per 1000)") +
  theme_minimal()
```



Part (d): Regional Averages and Their Meaning

Consider the mean of the birth rates for the Southeast Asian countries (including China and Japan). What meaning, if any, does this number have? Does it tell you about the average birth rate for the whole region? Why or why not? Do your answers change if we consider the median rather than the mean?

The mean birth rate for the Southeast Asian countries in our dataset provides limited insight about the region's overall birth rate. This average simply represents the typical birth rate among the countries sampled, treating each country as a single data point regardless of its population size.

This approach is problematic because it doesn't account for population differences. Japan and China, with their massive populations but relatively low birth rates, are given the same weight as smaller countries like Singapore or Malaysia. The resulting mean doesn't reflect the actual average birth rate across the entire regional population.

For a true regional birth rate, we would need to calculate a weighted average, with each country's birth rate weighted by its population size. Without this adjustment, the simple mean can be quite misleading.

Using the median instead of the mean would make the measure more robust against outliers (like Japan's extremely low birth rate), but it still wouldn't solve the fundamental problem of treating countries with vastly different population sizes equally. The median would tell us the "middle" birth rate among the countries sampled, which is a different concept than the typical birth rate across the region's population.

Part (e): Measuring Global Improvement

A 1996 UN report stated that in 70 mostly third-world countries average incomes in the 1990s were lower than in 1980, whereas 15 third-world countries mostly in Asia had experienced sustained economic growth. The obvious inference is that globally things have been getting worse. A feature writer suggested that the 15 countries that had made progress included China and India, which alone account for nearly half of the human race, so the global picture was actually much more optimistic than it had been painted. What do you think is the more relevant measure here?

If our focus is on global improvement for humanity as a whole, the more relevant measure is the number of people experiencing economic growth rather than the number of countries. The writer makes a valid point that counting countries without accounting for their population sizes can paint a misleading picture of global trends.

When China and India (with their combined population of nearly 3 billion people at that time) show economic growth, this positive change affects far more human lives than economic decline in dozens of smaller countries with much smaller populations. From a humanitarian perspective, it makes more sense to consider how many people are experiencing improved conditions rather than how many political entities are on each side of the ledger.

This doesn't diminish the real challenges faced by people in the 70 countries experiencing economic decline, but it does provide a more balanced global assessment. The most complete picture would include both metrics: the number of countries experiencing growth/decline and the number of people affected, as they tell different but complementary stories about global development.

Part (f): Population Increase Formula

Migrants refers to net migration, that is, those entering minus those leaving. How would you construct a variable Increase that gives the net increase in population size (per thousand) for the year for each country?

To calculate the net increase in population size per thousand, we need to account for all factors that cause population change: births, deaths, and migration. The formula would be:

$$\text{Increase} = \text{Migrants} + \text{Births} - \text{Deaths}$$

Where all values are measured per 1000 population.

This equation captures the three components of population change: 1. Natural increase (births minus deaths) 2. Net migration (immigrants minus emigrants)

Since births add to the population, deaths subtract from it, and net migration can either add or subtract depending on whether more people are entering or leaving, this formula gives us the total population change rate per thousand people for each country.