Net ID: ras637

RUID: 221003377

Lab1

Raj Shah

2025-02-09

```
# Load required libraries and set global chunk options
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2
```

```
## ── 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(countrycode)
```

```
## Warning: package 'countrycode' was built under R version 4.4.2
```

```
# Load the cleaned dataset
file_path <- "C:/Users/rajsh/OneDrive/Desktop/Inference Data Science
291/cleaned_gapminder_2010.csv"
cleaned_data <- read_csv(file_path)
```

```
## Rows: 175 Columns: 13
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (2): country, continent
## dbl (11): population, lifeexp, income, babies, childmort, co2, gdp,
healthsp...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

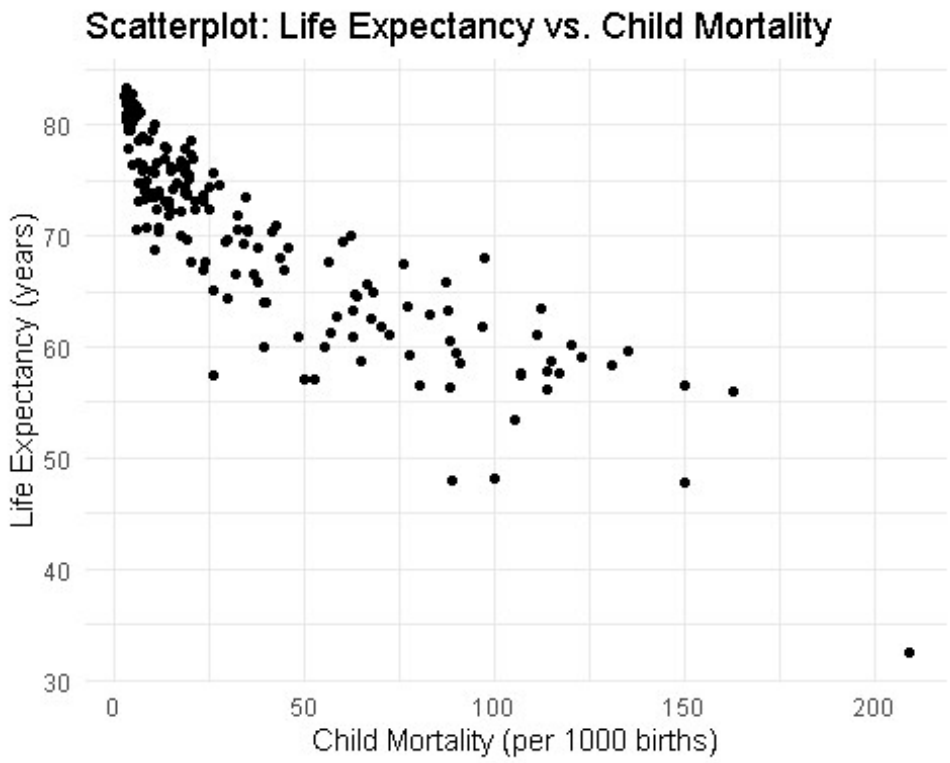
```
# Check column names to ensure they match
colnames(cleaned_data)
```

Net ID: ras637

RUID: 221003377

```
##  [1] "country"     "population"  "lifeexp"     "income"      "babies"
##  [6] "childmort"   "co2"         "gdp"         "healthspend" "water"
## [11] "popdensity"  "murder"      "continent"
```

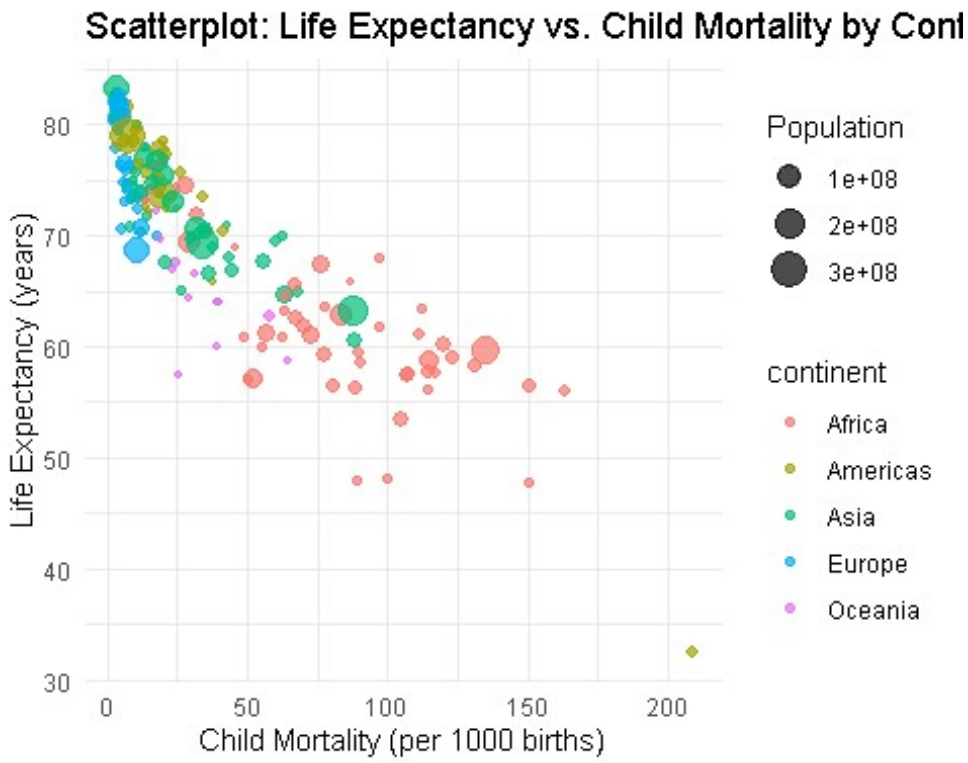
```
# 1. Simple scatterplot of life expectancy vs. child mortality
ggplot(cleaned_data, aes(x = childmort, y = lifeexp)) +
  geom_point() +
  labs(title = "Scatterplot: Life Expectancy vs. Child Mortality",
       x = "Child Mortality (per 1000 births)",
       y = "Life Expectancy (years)") +
  theme_minimal()
```



Net ID: ras637

RUID: 221003377

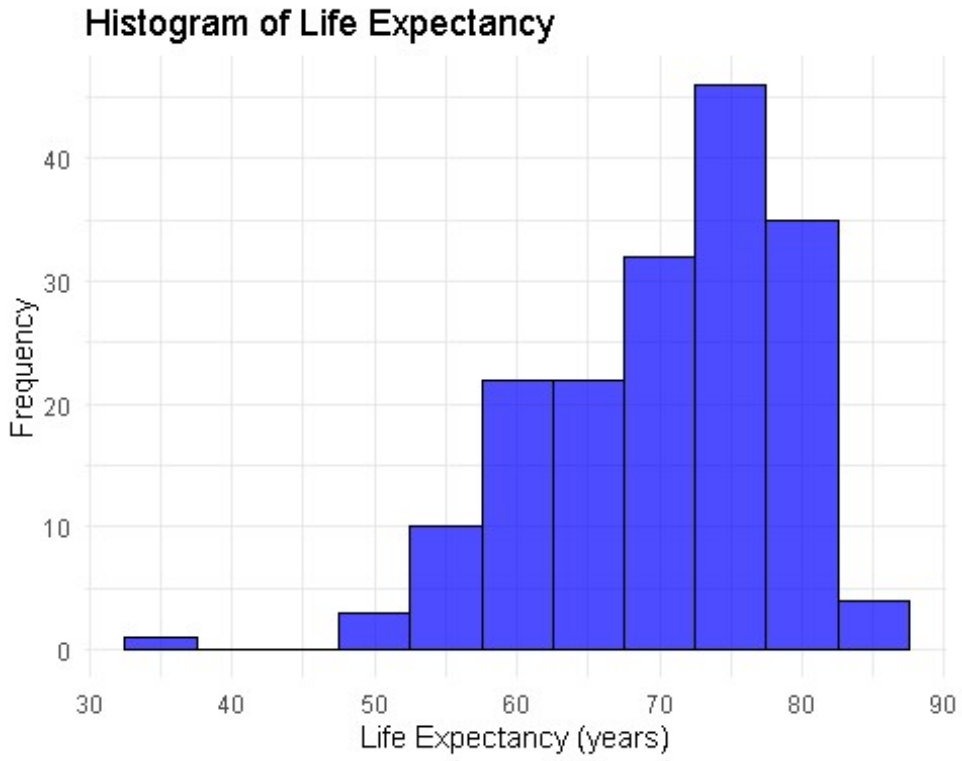
```
# 2. Scatterplot with different colors for continent and size for population
ggplot(cleaned_data, aes(x = childmort, y = lifeexp, color = continent, size
= population)) +
  geom_point(alpha = 0.7) +
  labs(title = "Scatterplot: Life Expectancy vs. Child Mortality by
Continent",
       x = "Child Mortality (per 1000 births)",
       y = "Life Expectancy (years)",
       size = "Population") +
  theme_minimal()
```



Net ID: ras637

RUID: 221003377

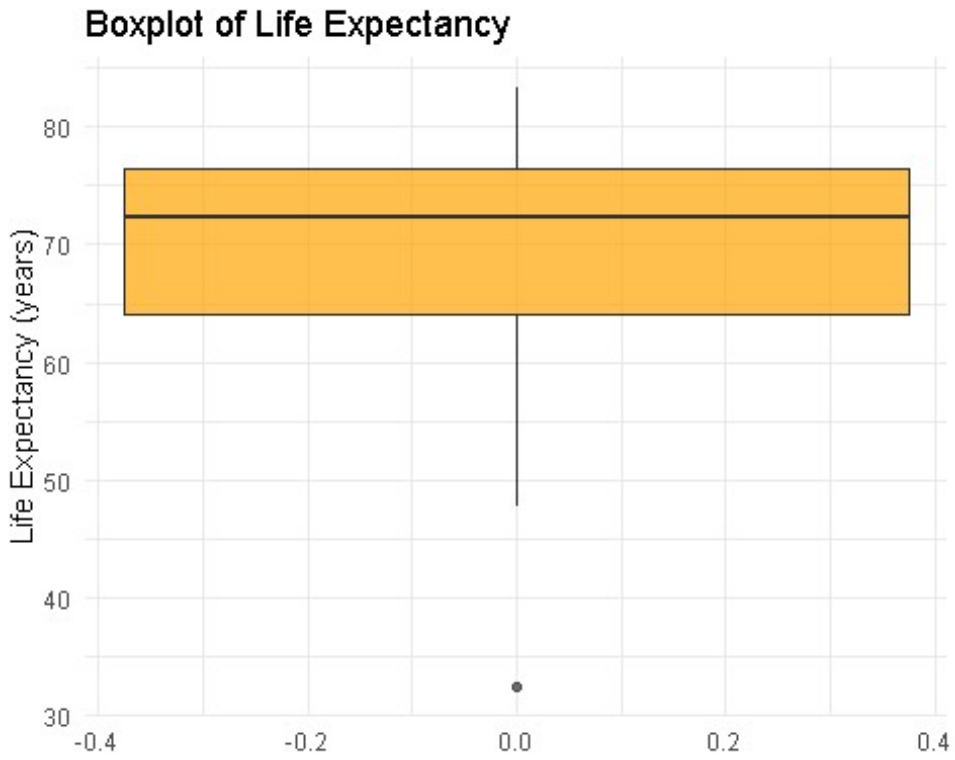
```r
# 3. Histogram of life expectancy
ggplot(cleaned_data, aes(x = lifeexp)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black", alpha = 0.7) +
  labs(title = "Histogram of Life Expectancy",
       x = "Life Expectancy (years)",
       y = "Frequency") +
  theme_minimal()
```



Net ID: ras637

RUID: 221003377

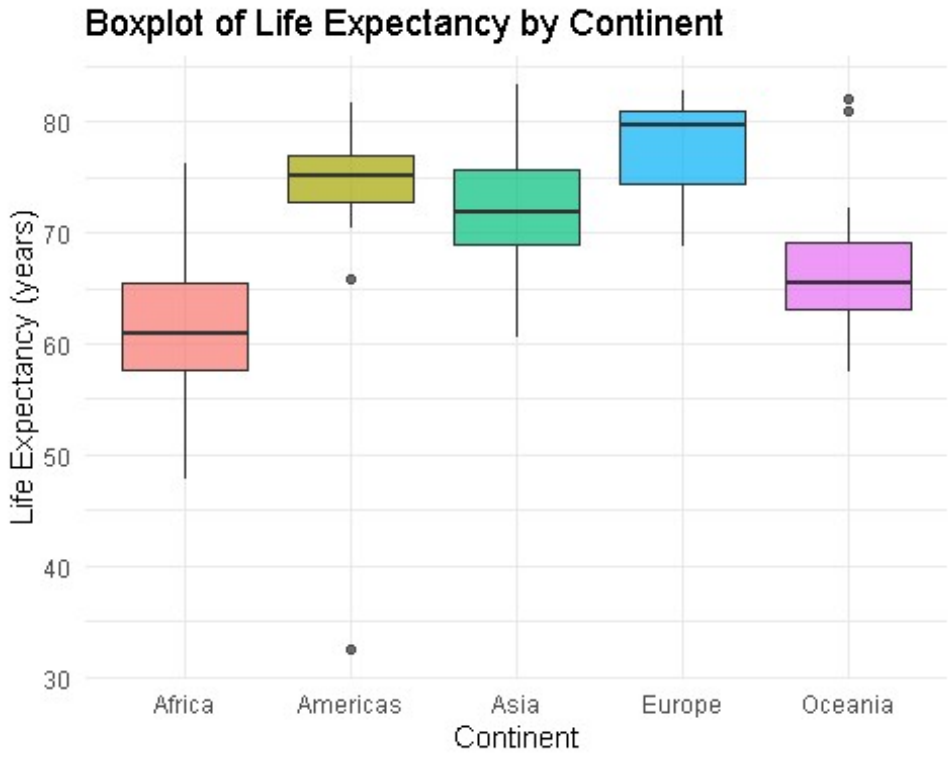
```r
# 4. Boxplot of life expectancy
ggplot(cleaned_data, aes(y = lifeexp)) +
  geom_boxplot(fill = "orange", alpha = 0.7) +
  labs(title = "Boxplot of Life Expectancy",
       y = "Life Expectancy (years)") +
  theme_minimal()
```



Net ID: ras637

RUID: 221003377

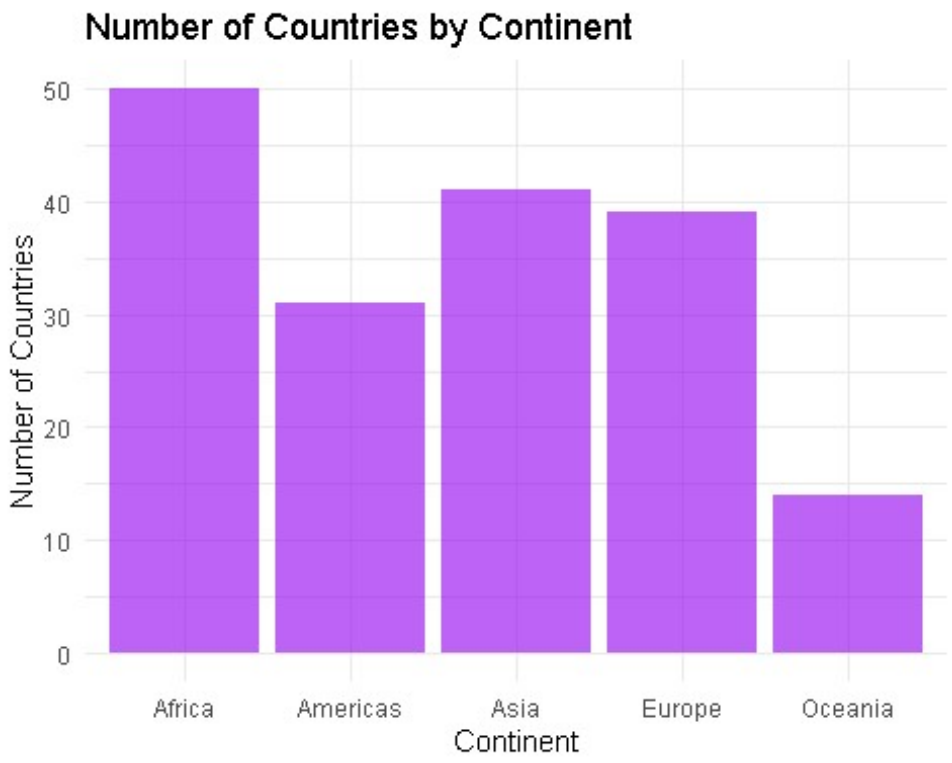
```r
# 5. Side-by-side boxplot of life expectancy by continent
ggplot(cleaned_data, aes(x = continent, y = lifeexp, fill = continent)) +
  geom_boxplot(alpha = 0.7) +
  labs(title = "Boxplot of Life Expectancy by Continent",
       x = "Continent",
       y = "Life Expectancy (years)") +
  theme_minimal() +
  theme(legend.position = "none")
```



Net ID: ras637

RUID: 221003377

```r
# 6. Bar graph of the number of countries by continent
ggplot(cleaned_data, aes(x = continent)) +
  geom_bar(fill = "purple", alpha = 0.7) +
  labs(title = "Number of Countries by Continent",
       x = "Continent",
       y = "Number of Countries") +
  theme_minimal()
```



Net ID: ras637

RUID: 221003377

```r
# 7. Define an indicator for whether the average number of children per woman
is <=2
cleaned_data <- cleaned_data %>%
  mutate(children_indicator = ifelse(babies <= 2, "<=2", ">2"))

# 8. Side-by-side bar graph of continent by children indicator
ggplot(cleaned_data, aes(x = continent, fill = children_indicator)) +
  geom_bar(position = "dodge", alpha = 0.7) +  # Use "dodge" for side-by-side
bars
  labs(title = "Side-by-Side Bar Graph of Continent by Children Indicator",
       x = "Continent",
       y = "Count of Countries",
       fill = "Children Indicator") +
  theme_minimal()
```