

Mini Project #3

Kalide Endale

9/28/2021

Contents

INTRODUCTION: I'll be analyzing the gapminder dataset, a popular rpackage dataset. My main analysis will be to look at life expectancy over different regions.

Developed countries have a higher level of life expectancy as predicted. In these regions, most of the data points are much closer to each other, making a narrow standard distribution (small standard of deviation). However, for developing regions (Africa and Asia), they have a much wider standard of distribution (greater standard of deviation) for life expectancy. It is also interesting to note that the oceanic region has a high life expectancy but we have fewer data points to make any strong conclusions from said data.

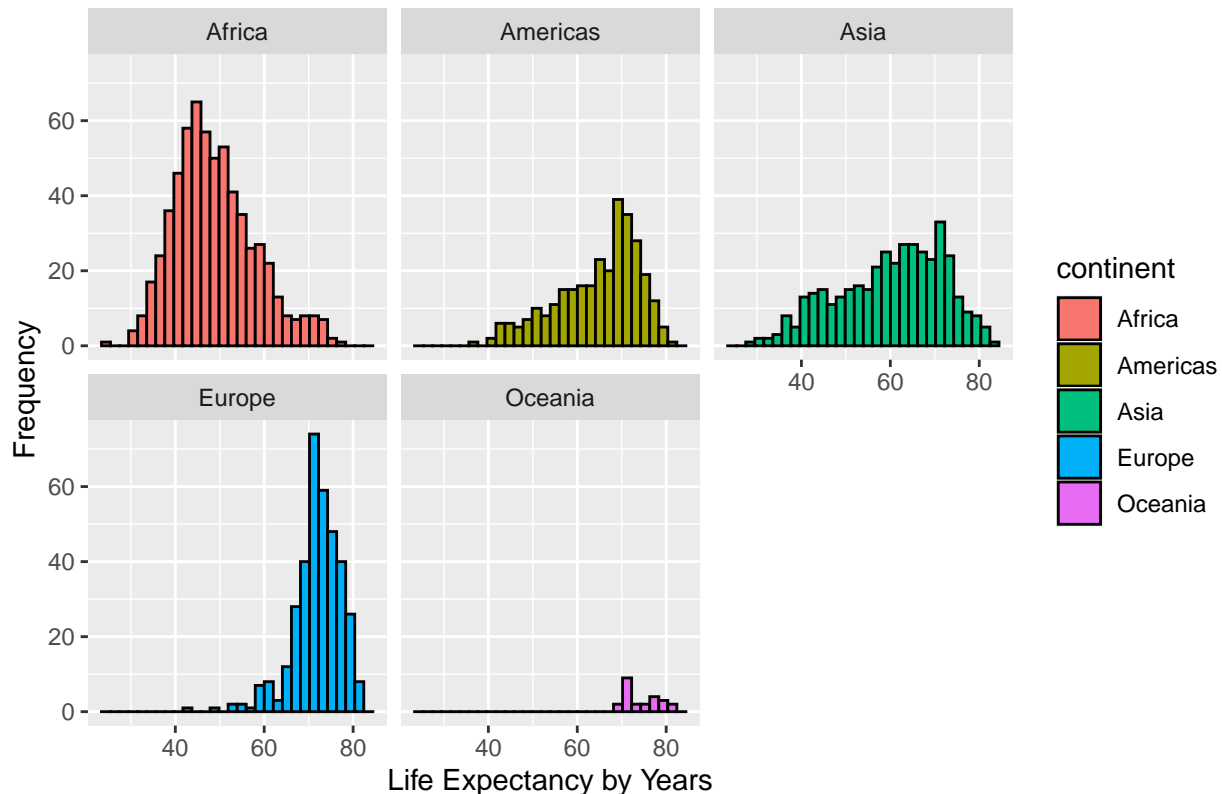
Shapes of Distribution: Africa = Looks Symmetrical ; Americas = Skewed Left ; Asia = Skewed Left ; Europe = Skewed Left ; Oceania = Skewed Right (difficult to say definitively with the small number of data points)

```
library(ggplot2)
library(gapminder)
```

```
ggplot(gapminder, aes(x = lifeExp)) +
  geom_histogram( aes(fill = continent), colour = "Black") +
  labs(title = " The Distribution of Life Expectancy per Continent (bins = 30)", x = "Life Expectancy by Continent") +
  facet_wrap(continent~.)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

The Distribution of Life Expectancy per Continent (bins = 30)



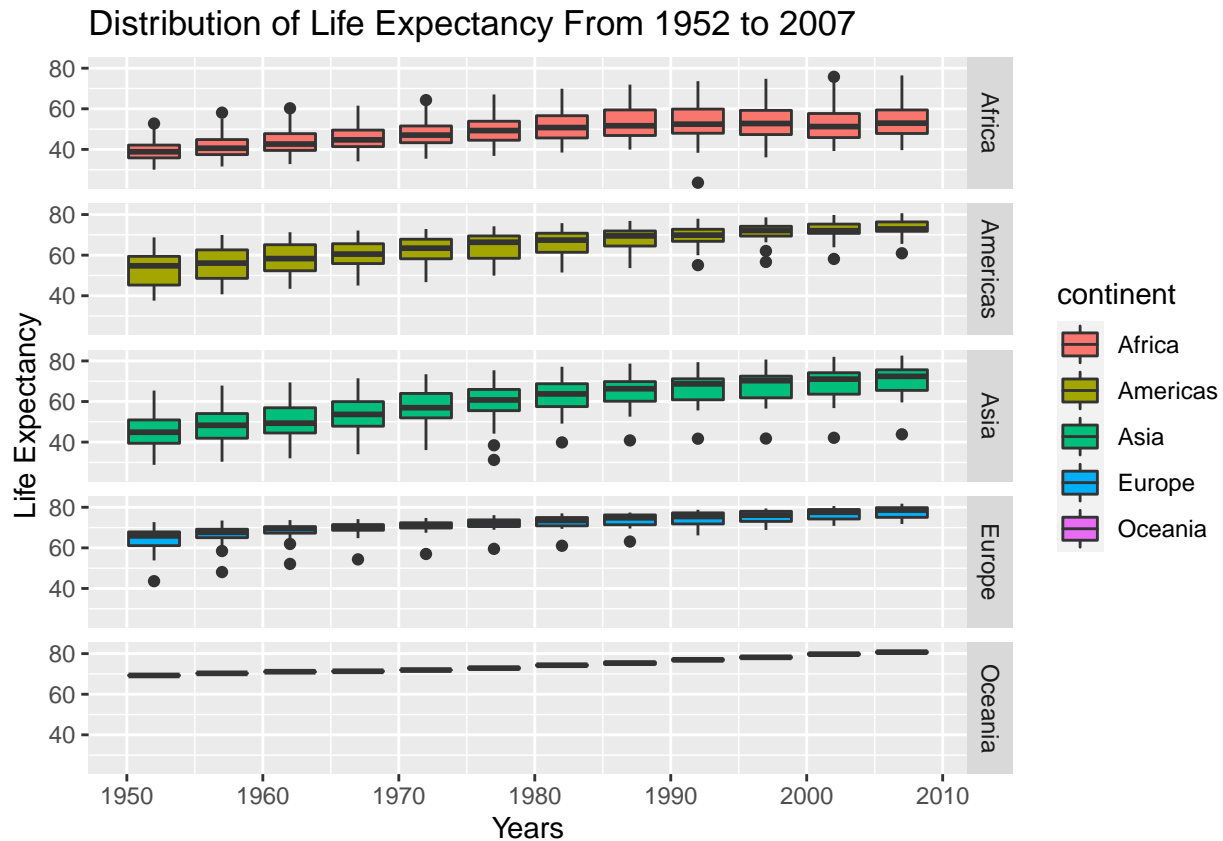
It's interesting to see that there is little variations of Life Expectancy for the Oceania countries over the 55 years of data. Additionally, the US and Europe had a very good life expectancy to GDP per Capita metric to start and this metric only improved over time.

The median life expectancy in Asia in 1952 was about 45 years of age, with a wide variance in the data (displayed by the extended whiskers from the interquartile range). Twenty years later (1972), the life expectancy in Asia has dramatically improved to about 60 years of age, still with a wide variance in the data. This shows that in those 20 years, we could infer there was a lot of economic development occurring in this region. Furthermore, the data also illustrates that the economic development likely varied significantly between Asian countries, leading to a only few big winners, which is probably why we see a noticeable variance in life expectancy in this region by 1972. Thirty-five years later (2007), the data points have consolidated towards a higher median in life expectancy and with fewer outliers.

What is disappointing is the trend and lack of progression we see in the African continent. The data shows that in 1952, the median life expectancy is about 40 years of age for this region and 55 years later(2007), this median has not improved much. The median in 2007 is about 50 years of age. The stagnation in the growth of life expectancy is likely related to the lack of economic growth in this region. Through the 1950s - 1960s, many African countries were in the midst of a battle for independence from their colonial countries. Even post-colonial era, many African nations are facing the problem of neo-colonization. This as a backdrop, it not a surprise that Africa is riddled with proxy wars, aid dependence, instability from artificially drawn borders, corruption of key institution, brain drain, and many other important factors that hampers their growth. However, it's not all bad, the whisker quartile 4 (this is also the max value that is not an outlier) is getting longer. That means that there are some countries in this region that are overcoming the social and economic challenges and improving the condition of life for their citizens.

```
ggplot(gapminder, aes(x = year, y = lifeExp, group = year, fill = continent)) +
  geom_boxplot() +
  labs(title = "Distribution of Life Expectancy From 1952 to 2007", x = "Years", y = "Life Expectancy")
```

```
facet_grid(continent~.)
```



There seems to be a positive curved relationship between GDP per capita and Life Expectancy. From the plots below, you can see that the regions that improved in GDP per capita also improved Life Expectancy. However, It's important to note that there are some outliers in our dataset in Asia, specifically. These data points have a very high GDP per capita but are middle of the pack with regards to Life Expectancy. At first glance it may be difficult to understand why this is the case but further investigation could help us determine why these countries with high GDP per capita are not improving the life expectancy for their citizens (possible factors can be social climate (over-working, stress, etc..), climate change related (the economic growth in these countries may not be sustainable), among other factors.

```
ggplot(gapminder, aes(x= lifeExp, y = gdpPercap)) +
  geom_point() +
  geom_smooth() +
  labs(x = "Life Expectancy", y = "GDP per Capita", title = "Showing Linear Relationship Between Life E
  facet_wrap(continent~.)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Showing Linear Relationship Between Life Expectancy and GDP per Cap

