

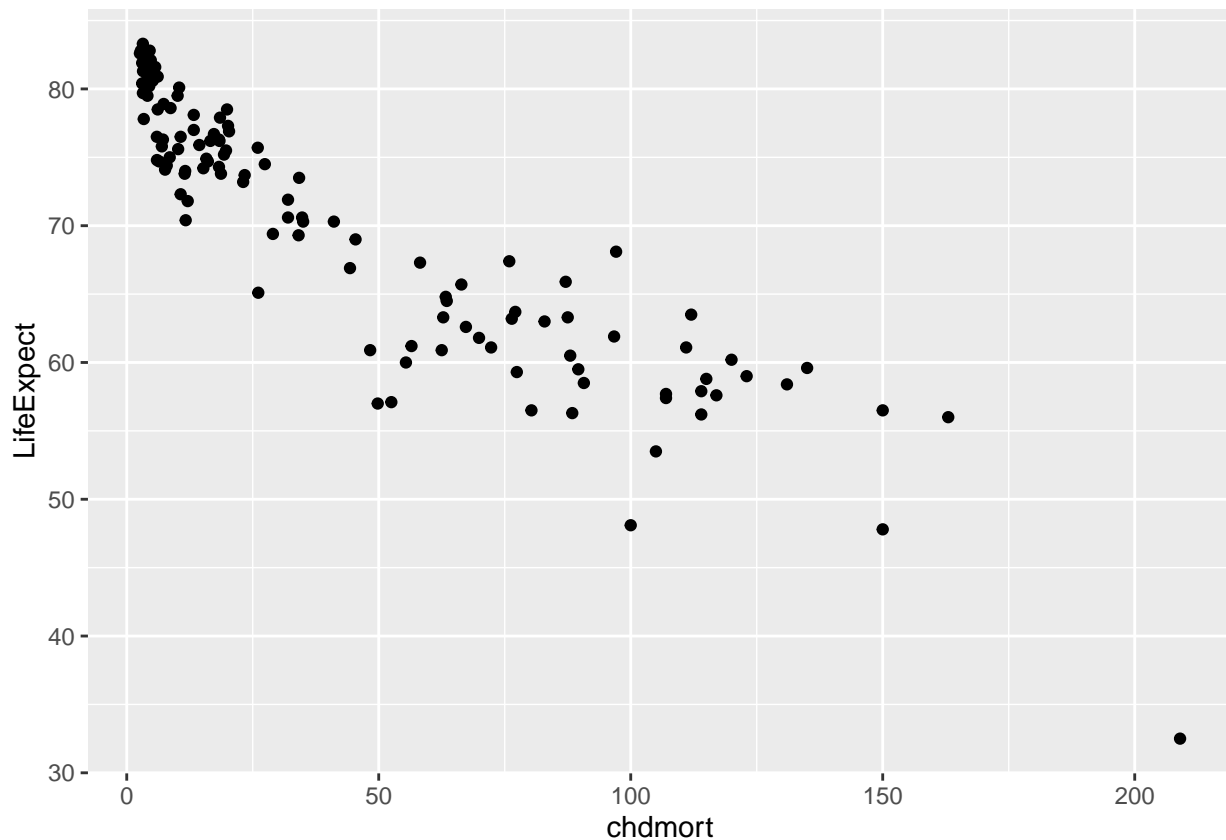
# Assignment1 STAT291

2023-02-02

By Lucas Weston

**Q1:** Have a simple scatterplot of life expectancy vs. child mortality. Is there a positive or negative relationship between the two? Is the relationship approximately linear?

```
d <- read.csv("dataGapminder.csv")
ggplot(data = d, mapping = aes(x = chdmort, y=LifeExpect)) + geom_point()
```



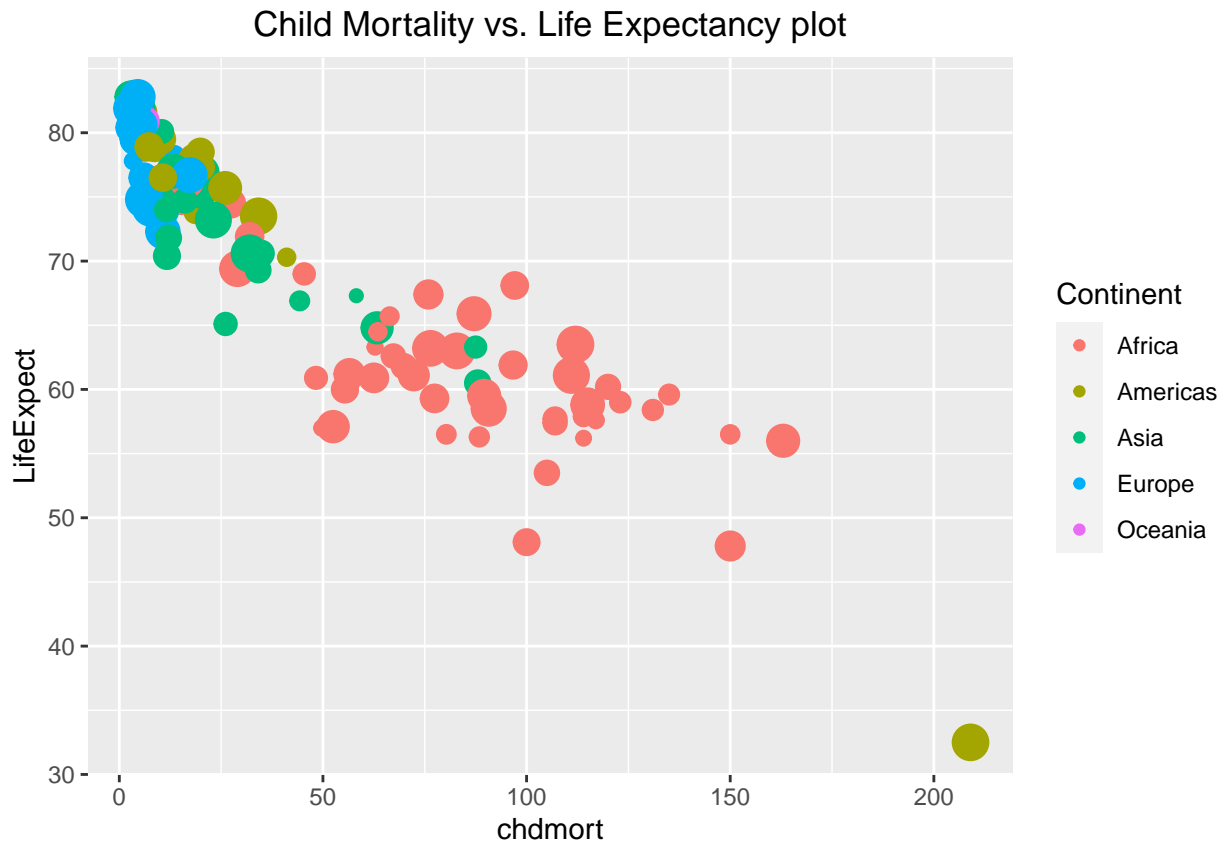
```
cor(d$chdmort, d$LifeExpect)
```

```
## [1] -0.9144401
```

As we can see from the graph, there it is linear. With a lower life expectancy the higher CMR (Child Mortality Rate) is. This also means that the higher life expectancy rate means that there is a lower CMR.

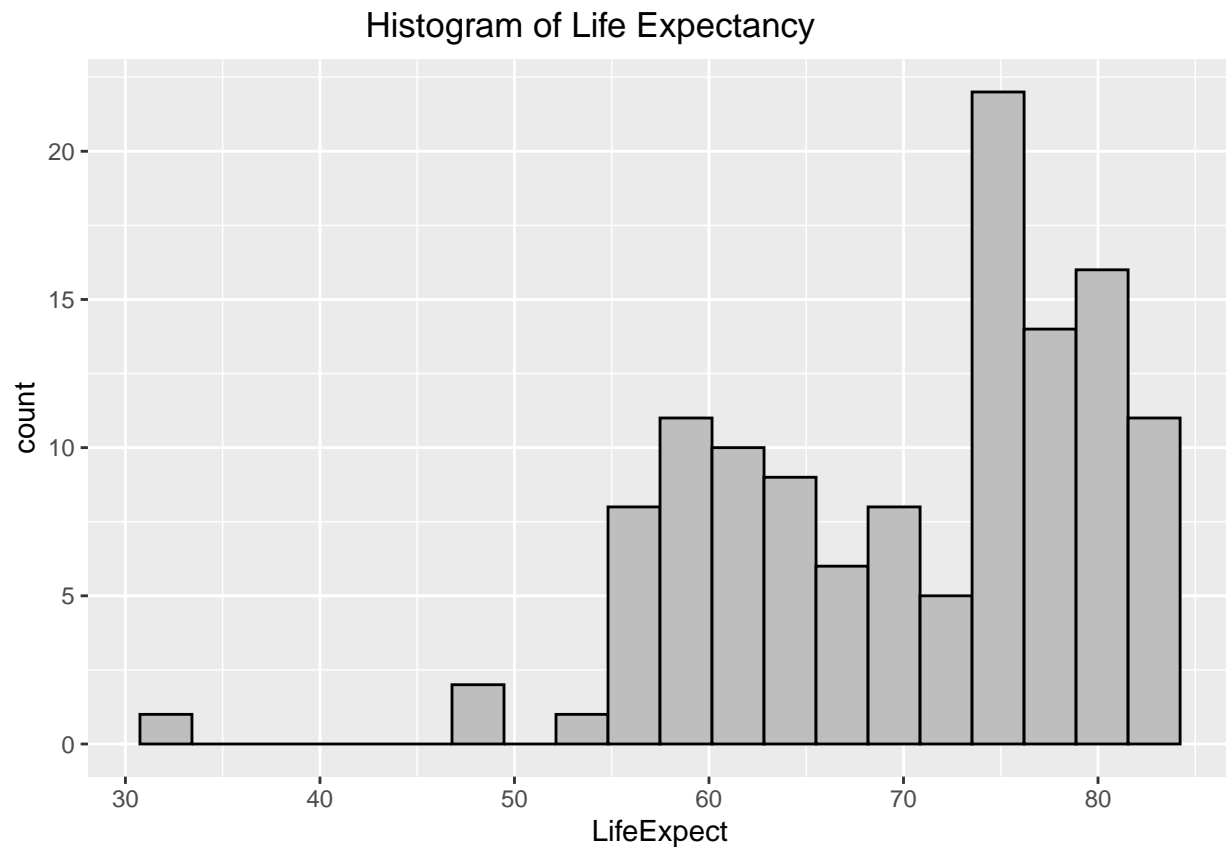
**Q2: Have a scatter plot of life expectancy vs. child mortality, using different color for continent and different size for the population of the country**

```
ggplot(d, aes(x = chdmort, y = LifeExpect, color = Continent)) +  
  geom_point(aes(size = ordered(Population))) + guides(size = "none") +  
  ggtitle("Child Mortality vs. Life Expectancy plot") +  
  theme(plot.title = element_text(hjust = 0.5))
```



**Q3: Have a histogram of life expectancy. Describe the distribution of it**

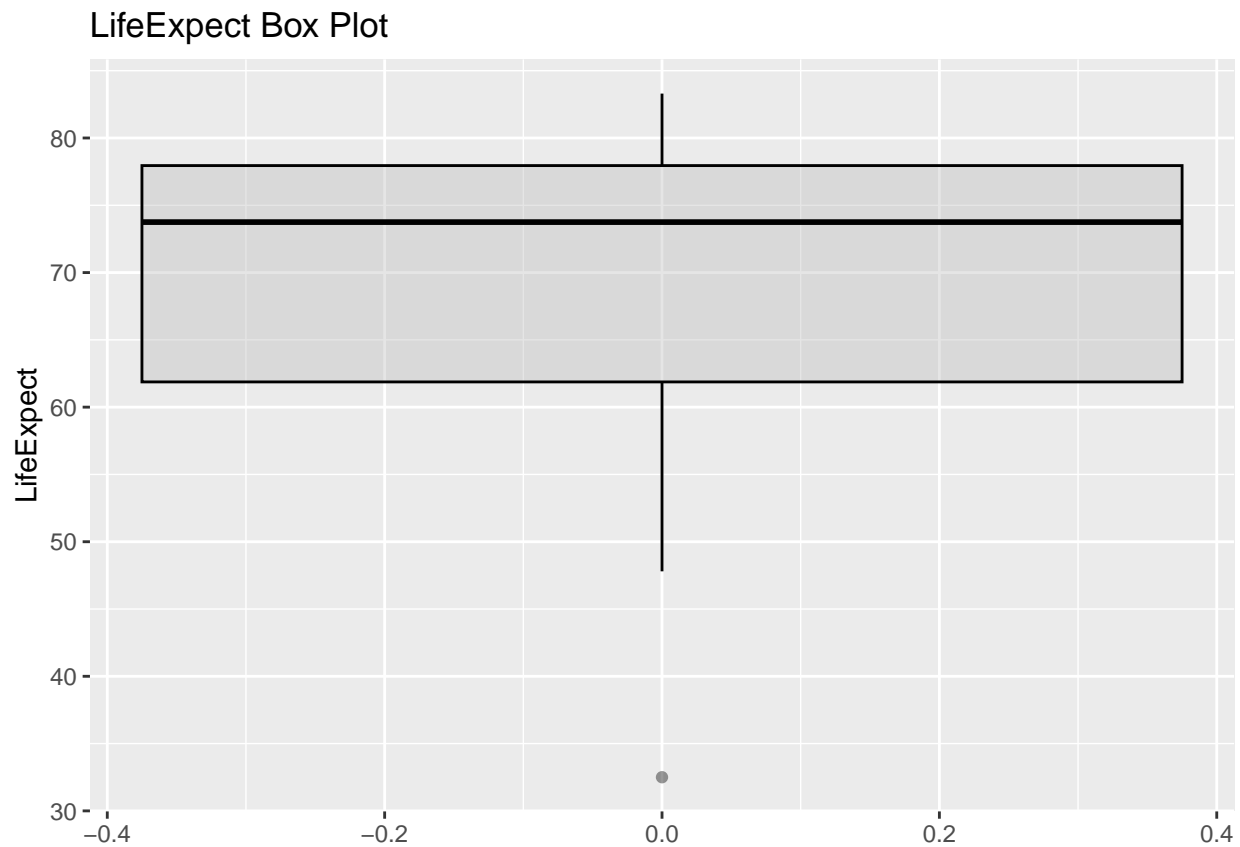
```
ggplot(d, aes(x = LifeExpect)) + geom_histogram(bins = 20, color="black", fill="gray") +  
  ggtitle("Histogram of Life Expectancy") +  
  theme(plot.title=element_text(hjust = 0.4))
```



The histogram seems to be skewed. There is an outlier near the left side but for the most part it is focused on the right with the highest values being around 75-80.

**Q4: Have a boxplot of life expectancy. Are there any standalone points in the boxplot?**

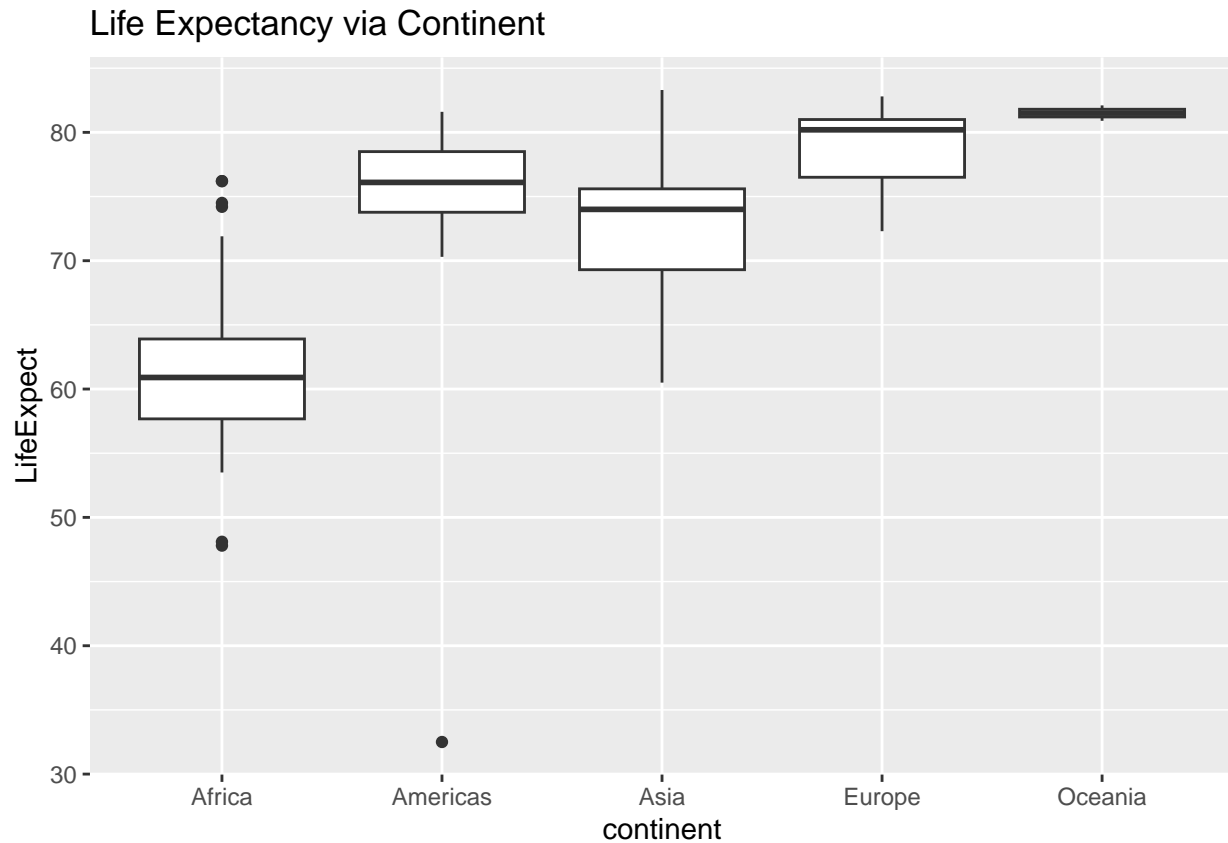
```
ggplot(d,mapping=aes(y= LifeExpect)) +geom_boxplot(alpha= 0.4, color = "black", fill = "gray") +labs(ti
```



This box plot has a single outlier near the far bottom of the graph.

**Q5: Have a side by side boxplot of life expectancy by continent. Are the boxplots similar among 5 continents? Explain some details**

```
ggplot(d, mapping = aes(x = Continent, y = LifeExpect)) + geom_boxplot() + ggtitle("Life Expectancy via
```



**Q6: Have a bar graph of continent. Which continent has the largest number of countries in the dataset? Which continent has the second largest number of countries?**

```
ggplot(d, aes(x = Continent)) +  
  geom_bar(stat = "count", fill="gray") +  
  ggtitle("Continents") + theme(plot.title=element_text(hjust = 0.4))
```

