

# STAT-S 670: Exploratory Data Analysis

## Mini-project: Life Expectancy and COVID

Swarn Gaba | Gayatri Gattani | Gandhar Ravindra Pansare | Rishikesh Kakde

### Executive Summary:

Our project revolves around exploring the dynamics of life expectancy and GDP per capita from 2019 to 2021 across 166 countries spanning five continents. The primary objective is to decipher the intricate relationship between changes in life expectancy and GDP per capita during this critical period.

In the initial phase, we scrutinized the connection between GDP and life expectancy in 2021. The findings suggest a positive correlation—increases in GDP per capita tend to align with higher life expectancy. We further delved into the temporal aspect, analyzing the change in life expectancy from 2019 to 2021. Surprisingly, we observed a declining trend in life expectancy across all continents during this period.

To gain a more nuanced understanding, we examined continent-specific variations in life expectancy changes. The data unveiled distinct patterns, indicating that life expectancy fluctuations differed by region. Our exploration then extended to probe the impact of GDP on life expectancy changes. Intriguingly, our analysis indicated no discernible trend associating GDP per capita with alterations in life expectancy.

This comprehensive analysis, spanning GDP and life expectancy in 2021, changes from 2019 to 2021, and the relationship between GDP and life expectancy changes, provides a multifaceted perspective on the factors influencing global life expectancy dynamics. Our next step involves constructing an interpretable model, possibly incorporating variables like continent, life expectancy in 2019, and GDP per capita in 2019. This modeling endeavor aims to quantify the intricate interplay of these factors and unearth insights that transcend traditional analyses. Through a systematic approach, we aspire to offer a nuanced understanding of the complex relationships within our dataset, shedding light on the dynamics of life expectancy and GDP per capita during this transformative period.

A researcher for a think tank wants to learn how life expectancy changed from 2019 (pre-COVID) and 2021 (post-COVID). They notice the World Bank has data on life expectancy and GDP per capita for those years.

Now that we have the necessary data on life expectancy and GDP per capita for the years 2019 and 2021, the next step is to conduct Exploratory Data Analysis (EDA). EDA allows us to delve into the dataset, explore patterns, and gain insights into how life expectancy changed from the pre-COVID year of 2019 to the post-COVID year of 2021. Through graphical visualizations, summary statistics, and trend analysis, we aim to uncover patterns, variations, and potential factors influencing life expectancy during this period.

**GDP per capita** is given in “PPP (current international dollars)”.

**Life Expectancy** is given in “Years”.

**Reading the data from the file countries.csv -**

```
countries.data <- read.csv("countries.csv", header = TRUE, sep = ",")

new_column_names <- c("Country Name", "Country Code", "Continent")

colnames(countries.data) <- new_column_names

#view(countries.data)
```

Reading the data from the file worldbank-life.csv -

```
worldbank.life.data <- read.csv(file = "worldbank-life.csv", header = TRUE, skip = 4,
                                sep = ",")

columns_to_read <- c("Country.Name", "Country.Code", "X2019", "X2020", "X2021")

worldbank.life.data <- worldbank.life.data |> dplyr::select(all_of(columns_to_read))

new_column_names <- c("Country Name", "Country Code", "Life Expectancy 2019",
                      "Life Expectancy 2020", "Life Expectancy 2021")

colnames(worldbank.life.data) <- new_column_names

#view(worldbank.life.data)
```

Reading the data from the file worldbank-gdp.csv -

```
worldbank.gdp.data <- read.csv(file = "worldbank-gdp.csv", header = TRUE, skip = 4,
                                sep = ",")

columns_to_read <- c("Country.Name", "Country.Code", "X2019", "X2020", "X2021")

worldbank.gdp.data <- worldbank.gdp.data |> dplyr::select(all_of(columns_to_read))

new_column_names <- c("Country Name", "Country Code", "GDP Per Capita 2019",
                      "GDP Per Capita 2020", "GDP Per Capita 2021")

colnames(worldbank.gdp.data) <- new_column_names

#view(worldbank.gdp.data)
```

Merging the data from the three aforementioned files.

```
merge.data.required <- left_join(countries.data, worldbank.life.data,
                                by = c("Country Code", "Country Name"))

merge.data.required <- left_join(merge.data.required, worldbank.gdp.data,
                                by = c("Country Code", "Country Name"))

merge.data.required <- na.omit(merge.data.required)

#view(merge.data.required)
```

After collecting and consolidating data on life expectancy and GDP per capita for the years 2019, 2020, and 2021 across 166 countries, we conducted a preliminary analysis to understand the distribution and central tendencies of the variables. The summary statistics below provide a snapshot of the key measures, shedding light on the variation in life expectancy and economic indicators during this period.

```
summary(merge.data.required)
```

##	Country Name	Country Code	Continent	Life Expectancy 2019
----	--------------	--------------	-----------	----------------------

```
## Length:166          Length:166          Length:166          Min.    :52.91
## Class :character    Class :character    Class :character    1st Qu.:65.96
## Mode  :character    Mode  :character    Mode  :character    Median :73.15
##                                     Mean   :72.24
##                                     3rd Qu.:77.99
##                                     Max.   :84.36
## Life Expectancy 2020 Life Expectancy 2021 GDP Per Capita 2019
## Min.    :52.78      Min.    :52.52      Min.    : 760.6
## 1st Qu.:65.62      1st Qu.:65.27      1st Qu.: 4839.9
## Median :72.15      Median :71.64      Median : 13995.0
## Mean   :71.58      Mean   :71.00      Mean   : 21397.6
## 3rd Qu.:77.17      3rd Qu.:76.42      3rd Qu.: 32290.8
## Max.   :84.56      Max.   :84.45      Max.   :119652.8
## GDP Per Capita 2020 GDP Per Capita 2021
## Min.    : 751.2     Min.    : 787.9
## 1st Qu.: 4868.9     1st Qu.: 5119.7
## Median :13345.0     Median :14508.6
## Mean   :20518.1     Mean   :22489.7
## 3rd Qu.:28312.3     3rd Qu.:33042.2
## Max.   :119534.6     Max.   :131630.8
```

The researcher's major research question is:

### How are changes in life expectancy from 2019 to 2021 related to GDP per capita?

Exploring the relationship between changes in life expectancy from 2019 to 2021 and GDP per capita might pose challenges in immediate comprehension. Hence, we aim to approach this systematically through a series of questions which can be divided into four distinct groups.

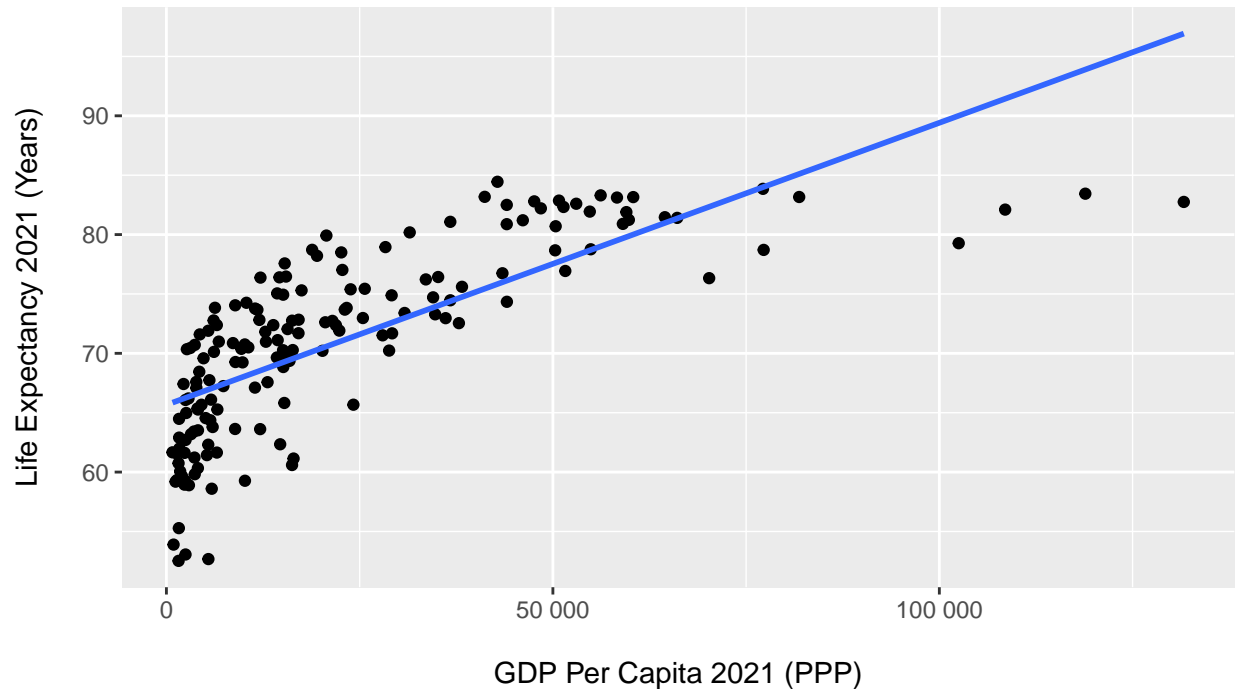
#### 1. GDP and life expectancy in 2021:

How does life expectancy vary with GDP per capita in 2021? Can the trends be well-described by a simple model such as a linear model, or is a more complicated model required?

```
ggplot(merge.data.required, aes(x = `GDP Per Capita 2021`, y = `Life Expectancy 2021`)) +
  geom_point() +
  scale_x_continuous(labels = scales::number_format(scale = 1)) +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE) +
  labs(title = "Figure 1: Life Expectancy VS. GDP Per Capita in 2021",
       subtitle = "Examining the Relationship with a Linear Regression Model,
       Predicting Life Expectancy based on GDP Per Capita.\n",
       x = "\nGDP Per Capita 2021 (PPP)",
       y = "Life Expectancy 2021 (Years)\n")
```

Figure 1: Life Expectancy VS. GDP Per Capita in 2021

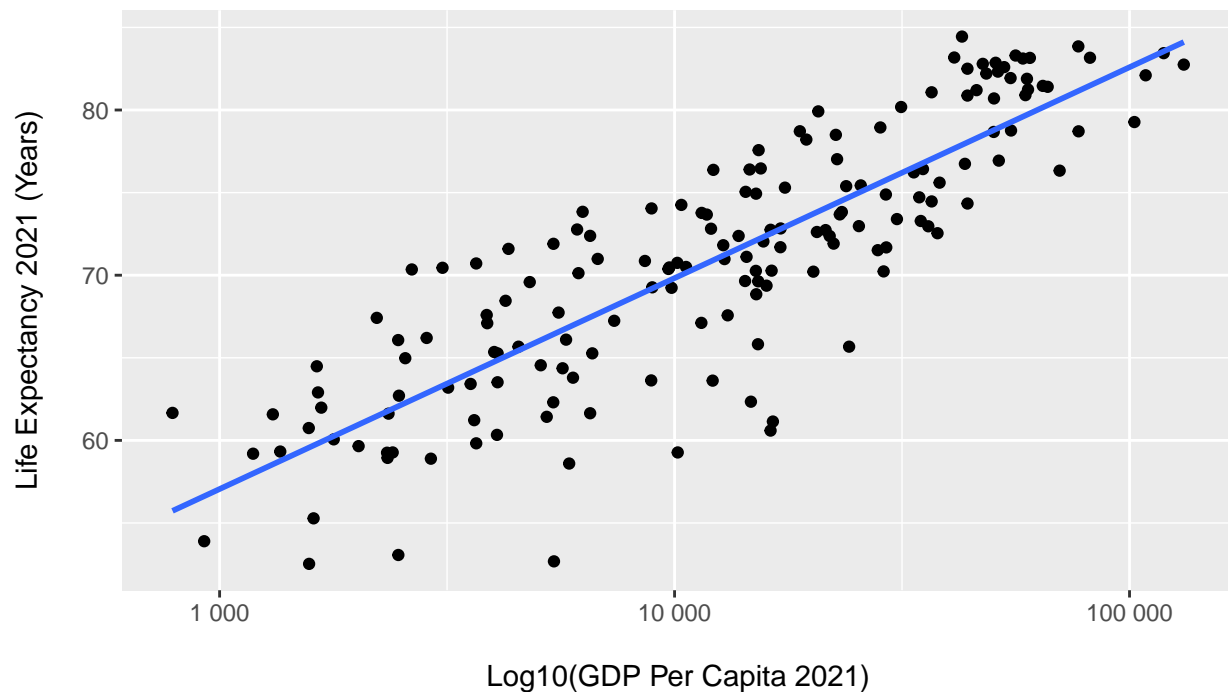
Examining the Relationship with a Linear Regression Model,  
Predicting Life Expectancy based on GDP Per Capita.



In Figure 1, we explore the relationship between GDP per capita and life expectancy in the year 2021. The x-axis represents GDP per capita values, while the y-axis displays corresponding life expectancy values. The scatter plot provides an initial view of the association between these variables. As GDP per capita values span a broad range, we aim to enhance the clarity of patterns by performing a log transformation to the x-axis. This transformation, presented in Figure 2, provides a more discernible representation of the underlying trends, especially across varying scales of GDP per capita.

```
ggplot(merge.data.required, aes(x = `GDP Per Capita 2021`, y = `Life Expectancy 2021`)) +  
  geom_point() +  
  scale_x_log10(labels = scales::number_format(scale = 1)) +  
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE) +  
  labs(title = "Figure 2: Life Expectancy VS. Log-Transformed GDP Per Capita in 2021",  
        subtitle = "Exploring Life Expectancy Patterns with a Linear Regression Model  
        based on Log-Transformed GDP Per Capita.\n",  
        x = "\nLog10(GDP Per Capita 2021)",  
        y = "Life Expectancy 2021 (Years)\n")
```

Figure 2: Life Expectancy VS. Log–Transformed GDP Per Capita in 2021  
Exploring Life Expectancy Patterns with a Linear Regression Model  
based on Log–Transformed GDP Per Capita.



Certainly, to further substantiate the meaningful relationship between GDP per capita and life expectancy in the year 2021, we can calculate the correlation coefficients for both the original data (GDP Per Capita 2021 and Life Expectancy 2021) and the log-transformed data (log(GDP Per Capita 2021) and Life Expectancy 2021).

```
# Correlation Coefficients between GDP Per Capita 2021 and Life Expectancy 2021
```

```
cor(merge.data.required$`GDP Per Capita 2021`, merge.data.required$`Life Expectancy 2021`)
```

```
## [1] 0.7449748
```

```
# Correlation Coefficients between Log10(GDP Per Capita 2021) and Life Expectancy 2021
```

```
cor(log10(merge.data.required$`GDP Per Capita 2021`), merge.data.required$`Life Expectancy 2021`)
```

```
## [1] 0.8505214
```

Upon careful examination of the plots and the calculated correlation coefficients, it is apparent that the correlation between log-transformed variable, log(GDP Per Capita 2021) and Life Expectancy 2021, is slightly higher than that of the original GDP Per Capita 2021 and Life Expectancy 2021.

The relationship between these 2 variables is strongly positive, suggesting that as GDP per capita increases, life expectancy tends to increase as well which is evident by both the visual representation in the plot and the positive correlation coefficient.

The correlation coefficient of 0.85 also tells us that the relationship is strong. Overall, The decision to scale x axis using log transformation resulted in a more effective fit to show a strong and meaningful relationship between these two variables.

Therefore, we observe that the Linear model is good enough to define the relationship between GDP per capita and life expectancy.

To simplify further, we can get an idea of the magnitude of the effect of GDP Per Capita 2021 on predicted Life Expectancy 2021 by fitting a linear model with additive terms for (base 10) log GDP Per Capita 2021 and Continent.

(A model with an interaction provides a better fit, but is not as straightforward to interpret.)

```
model.lm <- lm(`Life Expectancy 2021` ~ log10(`GDP Per Capita 2021`) + Continent,
              data = merge.data.required)

display(model.lm)

## lm(formula = 'Life Expectancy 2021' ~ log10('GDP Per Capita 2021') +
##      Continent, data = merge.data.required)
##               coef.est coef.se
## (Intercept)      26.96    3.00
## log10('GDP Per Capita 2021')  9.91    0.82
## ContinentAmericas      3.88    0.99
## ContinentAsia         4.94    0.95
## ContinentEurope       5.61    1.17
## ContinentOceania      5.11    1.18
## ---
## n = 166, k = 6
## residual sd = 3.72, R-Squared = 0.77
```

### The interpretation for the above linear regression model:

*(Intercept):*

The intercept is 26.96. This is the estimated life expectancy when both log GDP per capita and the country's continent are zero. In practical terms, this represents the estimated life expectancy for a country in Africa (the reference continent) with a log GDP per capita of zero.

*log10(GDP Per Capita 2021):*

For every one-unit increase in the log base 10 of GDP per capita in 2021, there is an estimated increase of 9.91 years in life expectancy. This implies a strong positive relationship between log GDP per capita and life expectancy.

Continent-specific effects indicate that, on average, countries from the Americas, Asia, Europe, and Oceania have higher life expectancies compared to countries in Africa, with the same log GDP per capita.

we created a dataframe named model.lm.df. This dataframe encompasses the original variables from our dataset, alongside two additional columns: .fitted, which contains the fitted values obtained from the linear regression model, and .resid, representing the residuals.

```
model.lm.df <- data.frame(merge.data.required,
                          .fitted = fitted.values(model.lm),
                          .resid = residuals(model.lm))
```

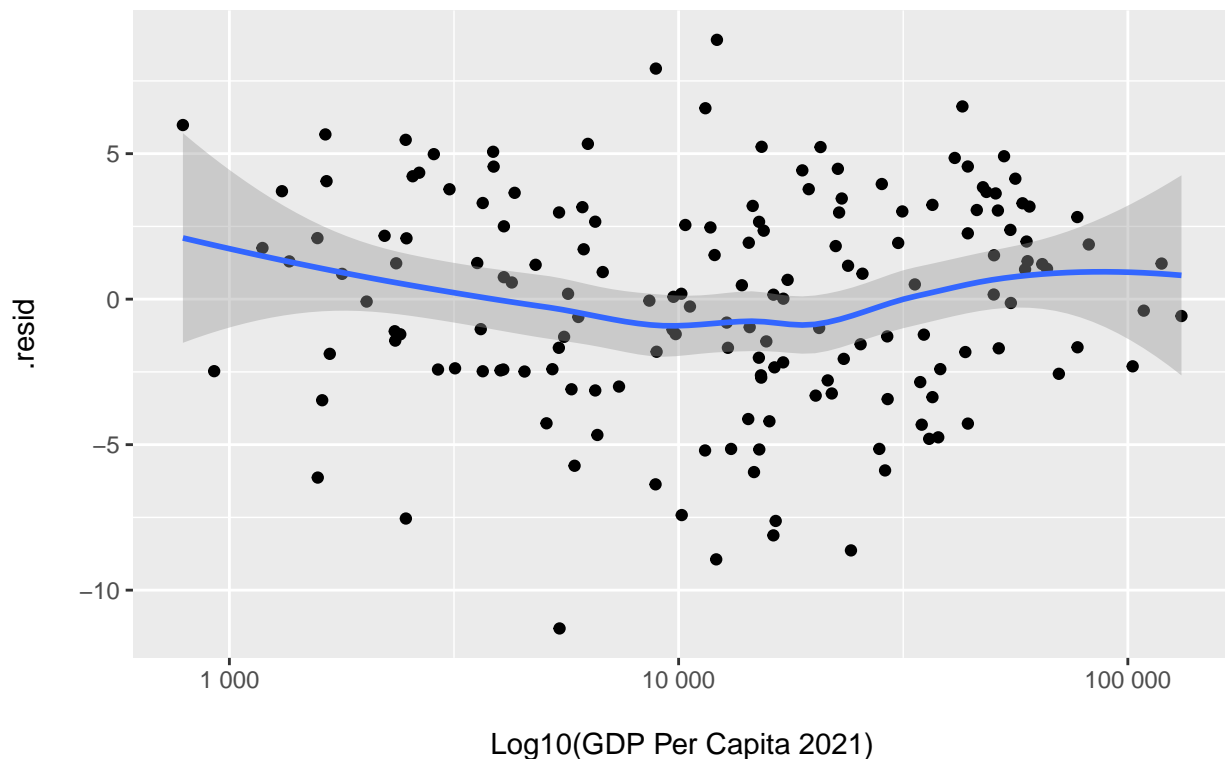
### Residual Plots :

1) **Lack-of-trend check:** Residuals Against Log-Transformed GDP Per Capita 2021

```
ggplot(model.lm.df, aes(x = GDP.Per.Capita.2021, y = .resid)) +
  geom_point() +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  geom_smooth() +
  labs(title = "Figure 3: Residuals Against Log-Transformed GDP Per Capita 2021\n",
       x = "\nLog10(GDP Per Capita 2021)",
       y = ".resid\n")
```

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

Figure 3: Residuals Against Log-Transformed GDP Per Capita 2021



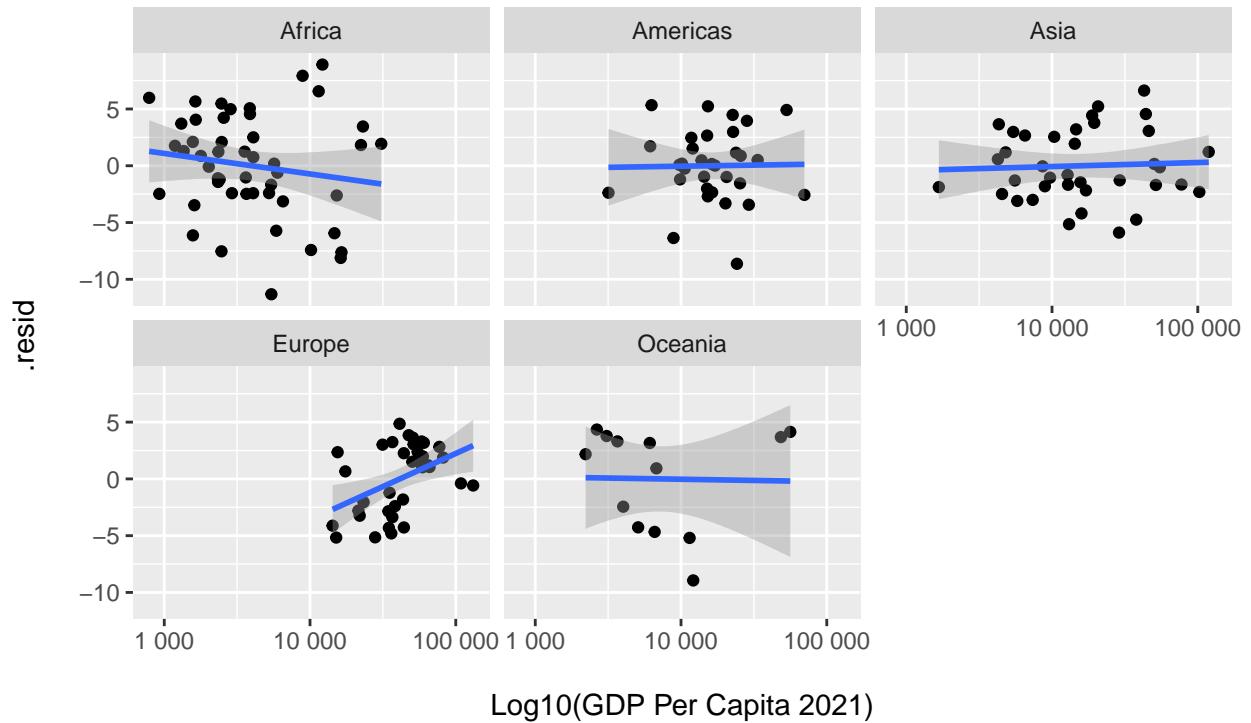
This scatter plot represents the relationship between residuals and log-transformed GDP per capita for the year 2021. The plotted data illustrates that the residuals are distributed randomly about the x-axis. This pattern suggests that the linear model might be the most appropriate fit for explaining the relationship between Life Expectancy and Log-Transformed GDP Per Capita in 2021.

**2) No-interaction check:** Residuals Against Log-Transformed GDP Per Capita 2021, Faceted By Continent

```
ggplot(model.lm.df, aes(x = GDP.Per.Capita.2021, y = .resid)) +
  geom_point() +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  geom_smooth(method = "lm", formula = y ~ x) +
  facet_wrap(~ Continent) +
  labs(title = "Figure 4: Residuals Against Log-Transformed GDP Per Capita 2021",
       subtitle = "Faceted By Continent, Depicted by Linear Regression Model\n",
```

```
x = "\nLog10(GDP Per Capita 2021)",
y = ".resid\n")
```

Figure 4: Residuals Against Log–Transformed GDP Per Capita 2021  
Faceted By Continent, Depicted by Linear Regression Model



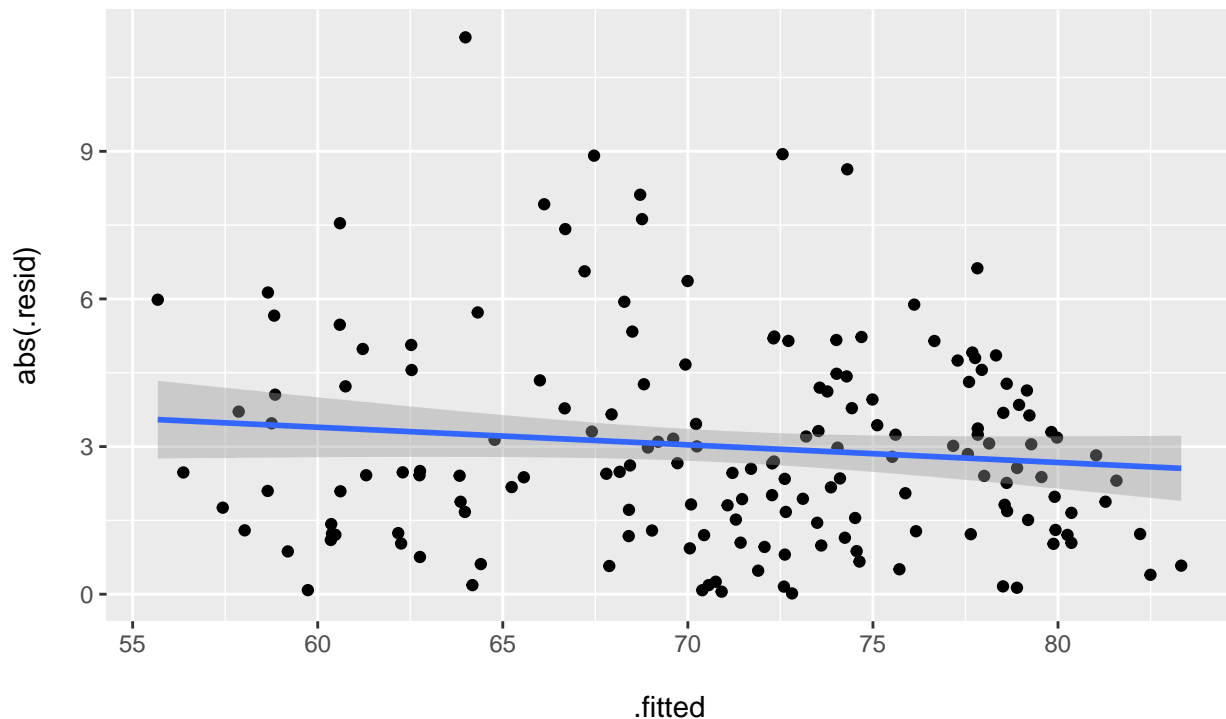
This scatter plot represents the relationship between residuals and log-transformed GDP per capita for the year 2021, faceted by continent. The residual spread suggests the model fits differently across continents, with Europe showing a positive trend as GDP increases. In Africa, Americas, and Asia, residuals are more randomly distributed, indicating a better model fit than in Europe or Oceania. Oceania has fewer data points, leading to greater uncertainty in the model fit as shown by the wider confidence interval.

### 3) Homoskedasticity check: Absolute Residuals Against Fitted Values

```
ggplot(model.lm.df, aes(x = .fitted, y = abs(.resid))) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x) +
  labs(title = "Figure 5: Absolute Residuals Against Fitted Values",
        subtitle = "Depicted by Linear Regression Model\n",
        x = "\n.fitted",
        y = "abs(.resid)\n")
```



Figure 5: Absolute Residuals Against Fitted Values  
Depicted by Linear Regression Model



This scatter plot represents the relationship between absolute residuals and fitted values. The trend line slightly decreases, indicating that higher fitted values have generally smaller residuals, which suggests that the model may be performing better at higher values of the predictor variable. The distribution of points does not show a clear pattern of increasing or decreasing variability of residuals with the increase of fitted values, implying that the variance of residuals is relatively constant—a good sign of homoskedasticity.

**Goodness-of-fit: R-Squared**

```
# Proportion of Variance Explained
var(model.lm.df$.fitted) / var(merge.data.required$`Life Expectancy 2021`)
```

```
## [1] 0.7749469
```

```
# Proportion of Variance Unexplained
var(model.lm.df$.resid) / var(merge.data.required$`Life Expectancy 2021`)
```

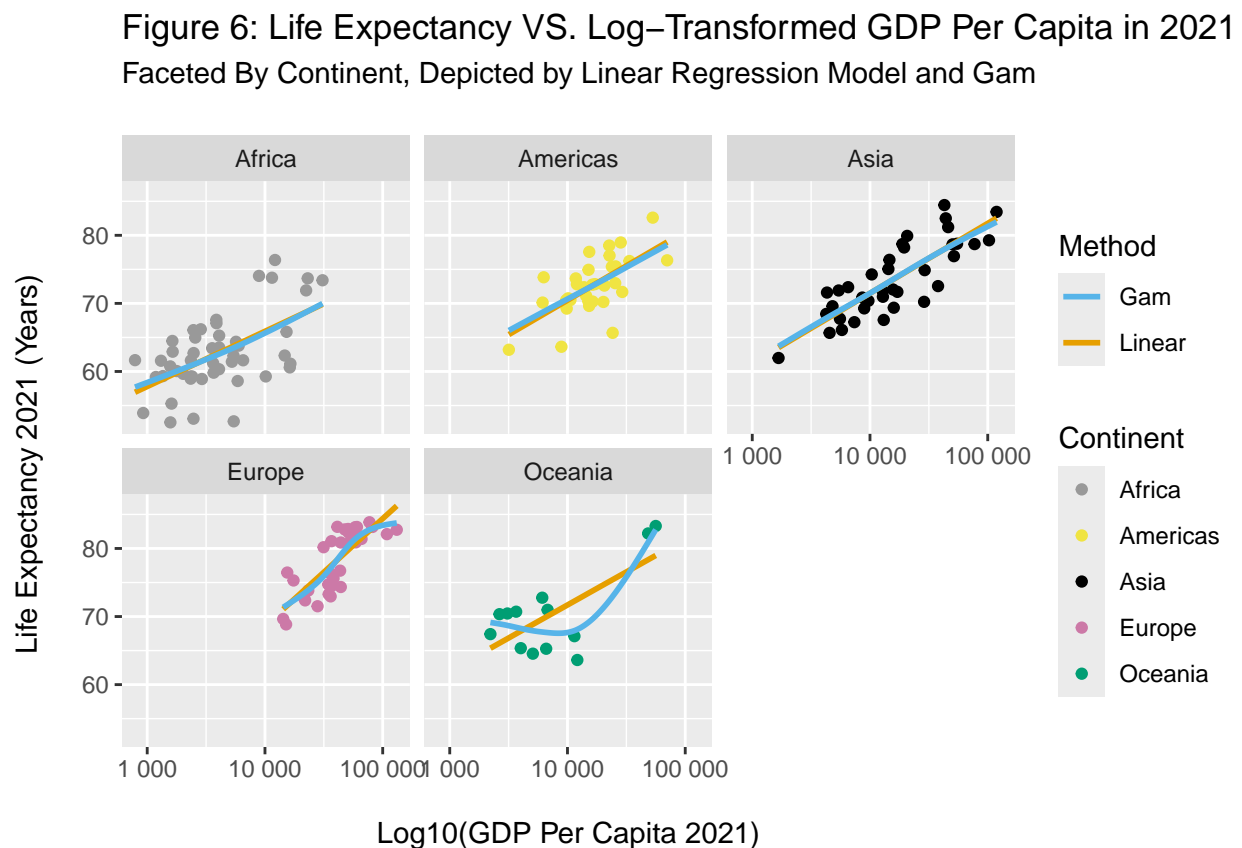
```
## [1] 0.2250531
```

An R-squared value of 0.775 means that approximately 77.5 % of the variation in life expectancy can be explained by the model. The value 0.775 is relatively high, indicating a good fit of the model to the data.

**Is the pattern the same or different for every continent? If some continents are different, which ones, and how is the relationship different in those continents?**

```
ggplot(merge.data.required, aes(x = `GDP Per Capita 2021`, y = `Life Expectancy 2021`)) +
  geom_point(aes(color = Continent)) +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  geom_smooth(aes(linetype = "Linear"), method = "lm", formula = y ~ x,
    se = FALSE, color = "#E69F00") +
  geom_smooth(aes(linetype = "Gam"), method = "gam",
    se = FALSE, color = "#56B4E9") +
  facet_wrap(Continent~.) +
  scale_color_manual(values = c("Africa" = "#999999", "Americas" = "#F0E442",
    "Asia" = "#000000", "Europe" = "#CC79A7",
    "Oceania" = "#009E73")) +
  scale_linetype_manual(name = "Method", values = c("Linear" = "solid", "Gam" = "solid")) +
  labs(title = "Figure 6: Life Expectancy VS. Log-Transformed GDP Per Capita in 2021",
    subtitle = "Faceted By Continent, Depicted by Linear Regression Model and Gam",
    x = "\nLog10(GDP Per Capita 2021)",
    y = "Life Expectancy 2021 (Years)",
    color = "Continent")
```

```
## 'geom_smooth()' using formula = 'y ~ s(x, bs = "cs")'
```



For Americas, Asia, and Africa continents, the trend appears to be linear, as indicated by the fitted linear lines. This suggests that a simple linear model could potentially describe the relationship between 'GDP per capita' and 'life expectancy' within these continents. However, the slopes and intercepts of these linear fits may differ, indicating that the strength of the relationship between GDP and life expectancy and the starting point of life expectancy at the lowest GDP level may vary by continent.

Europe shows a linear trend but with a steeper slope compared to other continents. This might imply that for Europe, changes in GDP per capita have greater effect on life expectancy. The points are more tightly clustered around the line, which suggests less variation in life expectancy at different GDP levels than in other continents.

Oceania presents the most deviation from the linear model, with the line curving sharply upwards. This indicates a nonlinear relationship, where increases in GDP per capita correspond to a greater increase in life expectancy beyond a certain threshold. The observed pattern indicates that to accurately describe the relationship between GDP per capita and life expectancy in Oceania, a more complex model, such as the GAM depicted above, may be necessary, rather than relying solely on a simple linear model.

## 2. Life expectancy changes from 2019 to 2021:

The new variable Life Expectancy Change 2019-21 represents the change in life expectancy from 2019 to 2021 for each country in the dataset.

```
merge.data.required$`Life Expectancy Change 2019-21` <-  
  merge.data.required$`Life Expectancy 2021` - merge.data.required$`Life Expectancy 2019`  
  
#view(merge.data.required)
```

Refining the structures of the dataset to extract key columns and organizing life expectancy data by country, continent, and year for a more insightful analysis and visualization of trends.

```
columns_to_read <- c("Country Name", "Country Code", "Continent", "Year", "Life Expectancy")  
  
merge.data <- merge.data.required |>  
  pivot_longer(cols = starts_with("Life Expectancy 20"), names_to = "Year",  
               values_to = "Life Expectancy") |>  
  mutate(Year = case_when(  
    grepl("Life Expectancy 2019", `Year`, ignore.case = TRUE) ~ "2019",  
    grepl("Life Expectancy 2020", `Year`, ignore.case = TRUE) ~ "2020",  
    grepl("Life Expectancy 2021", `Year`, ignore.case = TRUE) ~ "2021",  
  )) |> dplyr::select(all_of(columns_to_read))  
  
#view(merge.data)
```

We calculated the mean life expectancy for each year and continent, providing a summarized overview of the dataset.

```
merge.data.mean <- merge.data |>  
  group_by(`Year`, `Continent`) |>  
  summarise(Mean = mean(`Life Expectancy`, .groups = 'drop'))
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the  
## '.groups' argument.
```

```
#view(merge.data.mean)
```

How did life expectancy change from 2019 to 2021?

```
summary(merge.data.required$`Life Expectancy Change 2019-21`)
```

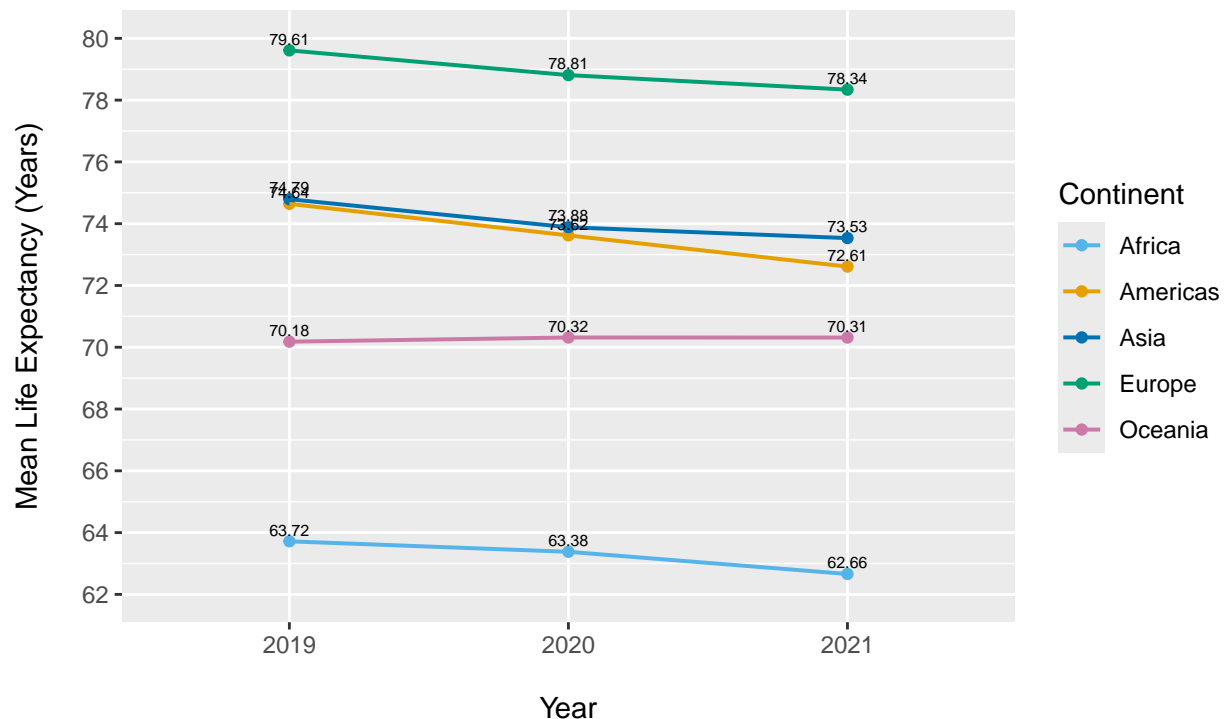
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -5.4610 -1.9432 -0.9040 -1.2432 -0.2665  0.7720
```

### Five-Point Summary Insights:

The mean (-1.2432 years) and the third quartile (-0.2665 years) are less than zero, which confirms that the central tendency of the life expectancy change is towards a decline. The minimum change is quite significant at -5.4610 years, indicating that at least one country experienced a substantial decrease in life expectancy. The maximum change (0.7720 years) shows that the positive changes in life expectancy, where they occurred, were relatively modest.

```
ggplot(merge.data.mean, aes(x = `Year`, y = Mean, color = `Continent`, group = `Continent`)) +
  geom_point() +
  geom_line(linewidth = 0.7) +
  geom_text(aes(label = round(Mean, 2)), vjust = -0.5, size = 2, color = "black") +
  scale_y_continuous(breaks = seq(62, 80, by = 2), limits = c(62, 80)) +
  scale_color_manual(values = cb_palette) +
  labs(title = "Figure 7: Mean Life Expectancy VS. Year",
       subtitle = "Grouped By Continent",
       x = "Year",
       y = "Mean Life Expectancy (Years)")
```

Figure 7: Mean Life Expectancy VS. Year  
Grouped By Continent



To observe life expectancy change from 2019 to 2021, we plotted the mean life expectancy of each continent in each year from 2019 to 2021 considering change in mean life expectancy corresponds to change in life expectancy and observed following :

From above line plot, we can observe that life expectancy shows a *declining trend from 2019 to 2021* across all continents except Oceania. If observed carefully, change in America is highest from 2019 to 2021 that is -2.03, indicating mean life expectancy in America in 2021 is 2.03 lesser than mean life expectancy in 2019. Similarly from 2019 to 2020, and 2020 to 2021 change in mean life expectancy of America is highest being 1.02 and 1.01 respectively showing declining trend. The minimum change is observed in Oceania such that from 2019 to 2020, there is positive trend that is mean life expectancy in 2020 is 0.13 greater than that in 2019 and from 2020 to 2021, the change is -0.01, that is almost negligible thus making Oceania showing the least change from 2019 to 2021 that is -0.13. Overall, it can be seen that Europe among all continents had highest mean of life expectancy in all three years that is 79.61 in 2019, 78.81 in 2020 and 78.34 in 2021 while Africa had that lowest across all three years that is 63.72 in 2019, 63.38 in 2020 and 62.66 in 2021.

### Which countries saw the biggest changes?

```
columns_to_read <- c("Country Name", "Country Code", "Continent",
                     "Life Expectancy Change 2019-21")

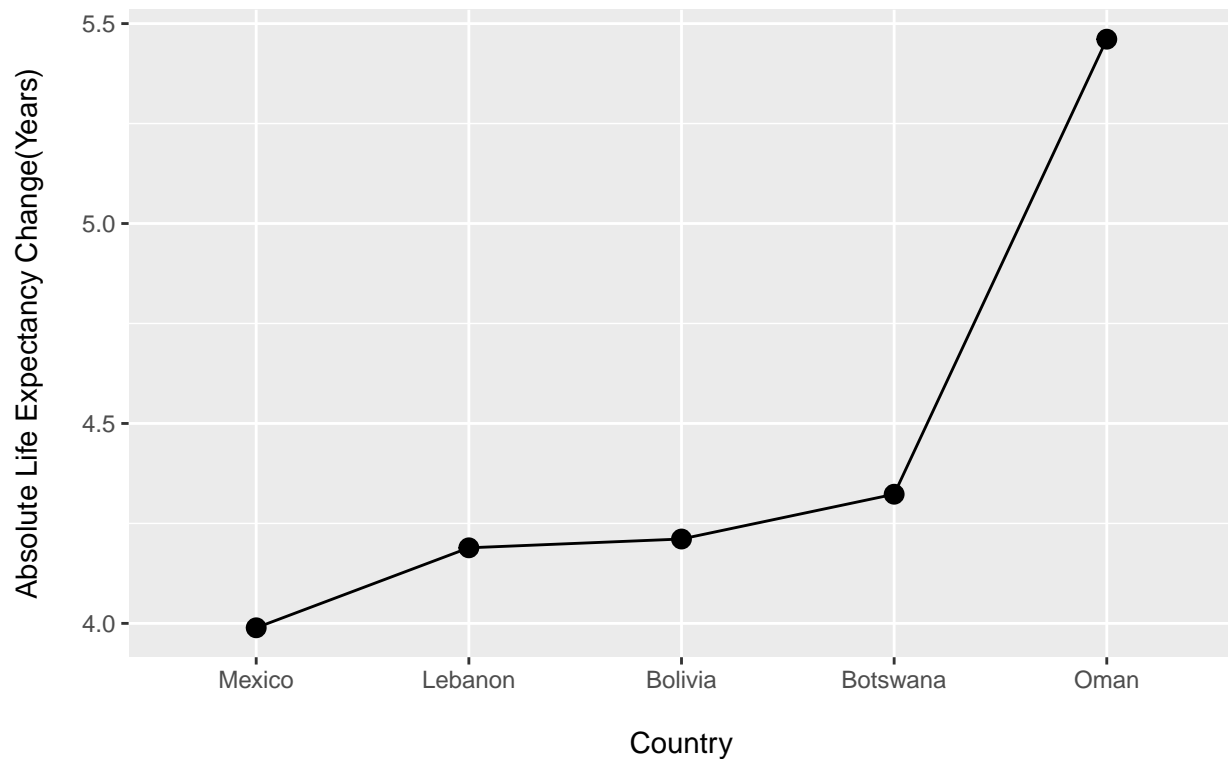
top5.data <- merge.data.required |>
  top_n(5, wt = abs(`Life Expectancy Change 2019-21`)) |>
  arrange(desc(abs(`Life Expectancy Change 2019-21`))) |>
  dplyr::select(all_of(columns_to_read))

knitr::kable(top5.data)
```

Country Name	Country Code	Continent	Life Expectancy Change 2019-21
Oman	OMN	Asia	-5.461
Botswana	BWA	Africa	-4.323
Bolivia	BOL	Americas	-4.211
Lebanon	LBN	Asia	-4.189
Mexico	MEX	Americas	-3.989

```
merge.data.required |>
  top_n(5, wt = abs(`Life Expectancy Change 2019-21`)) |>
  ggplot(aes(x = reorder(`Country Name`, abs(`Life Expectancy Change 2019-21`)),
             y = abs(`Life Expectancy Change 2019-21`))) +
  geom_point(size = 3) +
  geom_line(aes(group=1)) +
  labs(title = "Figure 8: Top 5 Countries by Absolute Life Expectancy Change\n",
       x = "\nCountry",
       y = "Absolute Life Expectancy Change(Years)\n")
```

Figure 8: Top 5 Countries by Absolute Life Expectancy Change



Oman experienced the largest change i.e. decline in life expectancy, with a decrease of approximately 5.46 years.

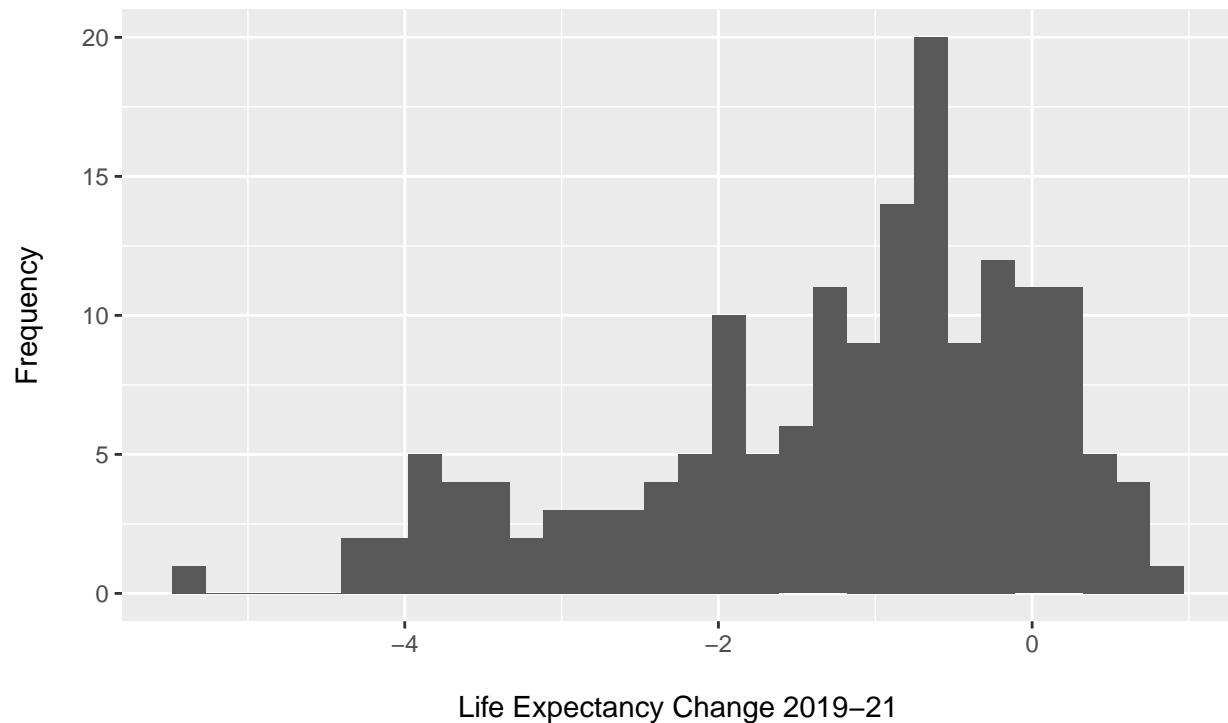
Botswana, Bolivia, Lebanon, and Mexico also saw substantial decreases, ranging from about 3.99 to 4.32 years.

Does the distribution of changes vary by continent, and if so, how?

```
ggplot(merge.data.required, aes(x = `Life Expectancy Change 2019-21`)) +  
  geom_histogram() +  
  labs(title = "Figure 9: Distribution of Life Expectancy Change 2019-21",  
        subtitle = "Depicted By Histogram\n",  
        x = "\nLife Expectancy Change 2019-21",  
        y = "Frequency\n")
```

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

Figure 9: Distribution of Life Expectancy Change 2019–21  
Depicted By Histogram

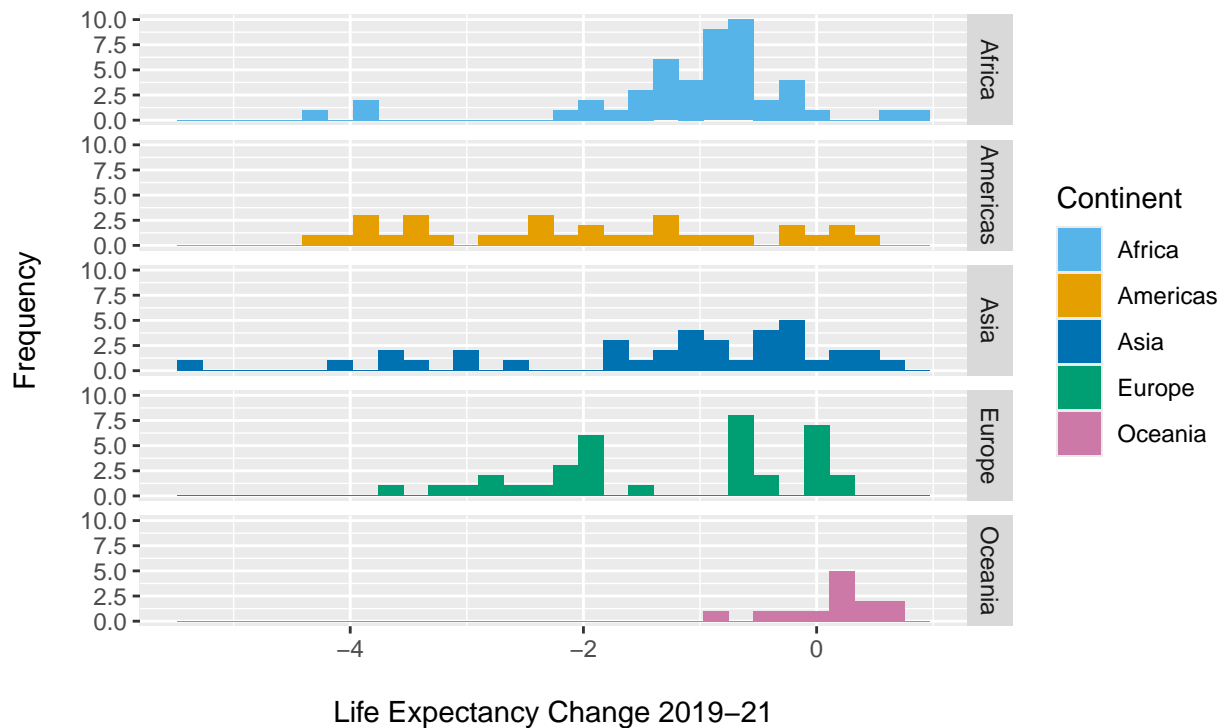


In the histogram, we can see that the most common change in life expectancy was a decrease of around 1 to 2 years, as indicated by the tallest bar. There are also some counts of life expectancy changes between 0 and -1 and -2 and -3, though less frequent. There's a very small count of life expectancy decreases that are more than 4 years, as indicated by the single bar on the far left of the graph.

```
ggplot(merge.data.required, aes(x = `Life Expectancy Change 2019-21`,
                                fill = Continent, group = Continent)) +
  geom_histogram() +
  facet_grid(Continent~.) +
  scale_fill_manual(values = cb_palette) +
  labs(title = "Figure 10: Distribution of Life Expectancy Change 2019-21",
       subtitle = "Faceted By Continent, Depicted By Histogram\n",
       x = "\nLife Expectancy Change 2019-21",
       y = "Frequency\n")
```

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

Figure 10: Distribution of Life Expectancy Change 2019–21  
Faceted By Continent, Depicted By Histogram



**Africa:** Most of the countries saw a decline in life expectancy concentrated around -1.5 to -0.5.

**Americas:** The distribution for Americas show a rather uniformly distributed frequency of the change in life expectancy with small peaks spread out through the distribution.

**Asia:** Asia's histogram doesn't show any discernible pattern, but most of the countries lie between -2 and 0.5, with peak around -0.25.

**Europe:** There are three notable peaks in the distribution at around -2, -0.75, and 0, with highest being at -0.75.

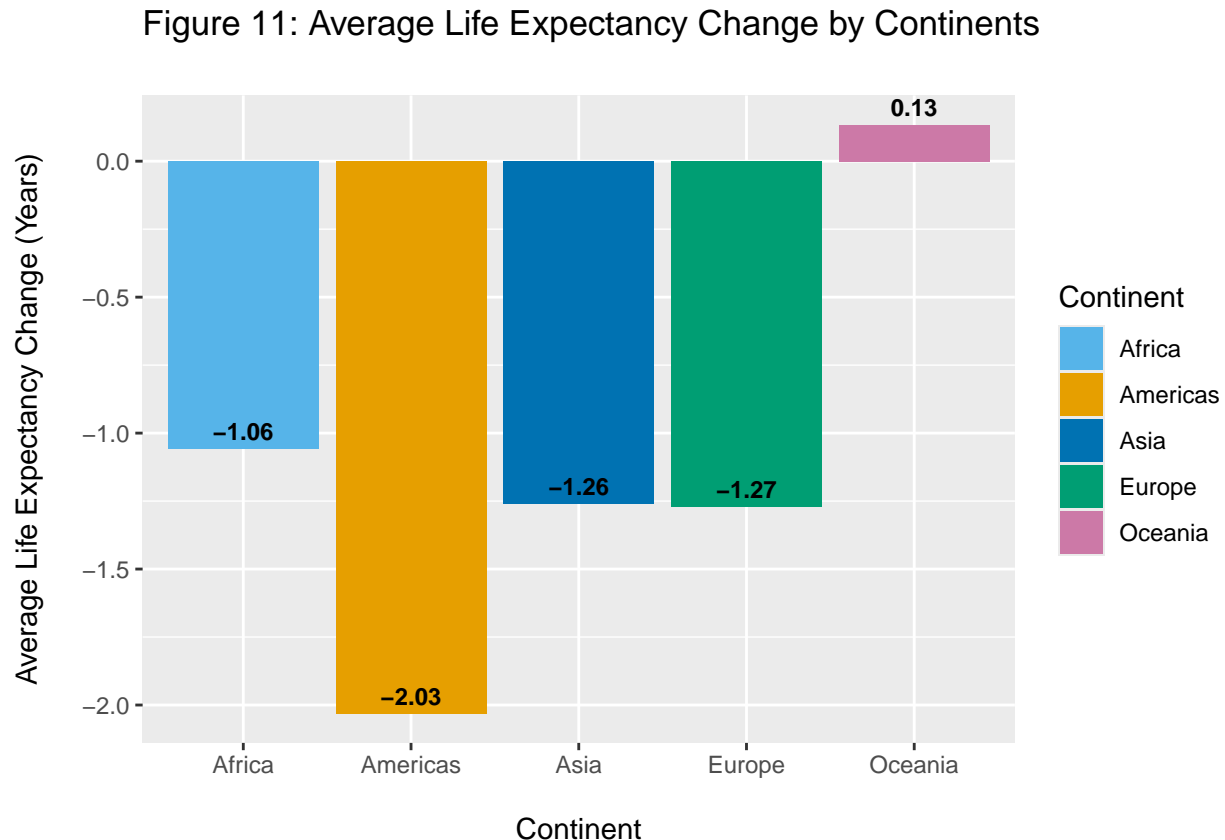
**Oceania:** Due to less number of data points for Oceania, the plot of distribution is not very informative. Most of the countries in Oceania saw a positive change in Life Expectancy. The peak of the distribution is at 0.25. Oceania is the only continent among the three under consideration that has a positive peak value for change in life expectancy in its distribution.

```
continent.data <- merge.data.required |>
  group_by(Continent) |>
  summarise(Avg_Life_Expectancy_Change = mean(`Life Expectancy Change 2019-21`,
                                              na.rm = TRUE))

ggplot(continent.data, aes(x = Continent, y = Avg_Life_Expectancy_Change, fill = Continent)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = cb_palette)+
  geom_text(aes(label = round(Avg_Life_Expectancy_Change, 2)),
            vjust = -0.5,
            size = 3,
            color = "black",
```



```
fontface = "bold") +
labs(title = "Figure 11: Average Life Expectancy Change by Continents\n",
x = "\nContinent",
y = "Average Life Expectancy Change (Years)\n")
```



Life expectancy change is different in all continents.

All continents experienced a decline in average life expectancy, except for Oceania, which saw a slight increase (0.13 years).

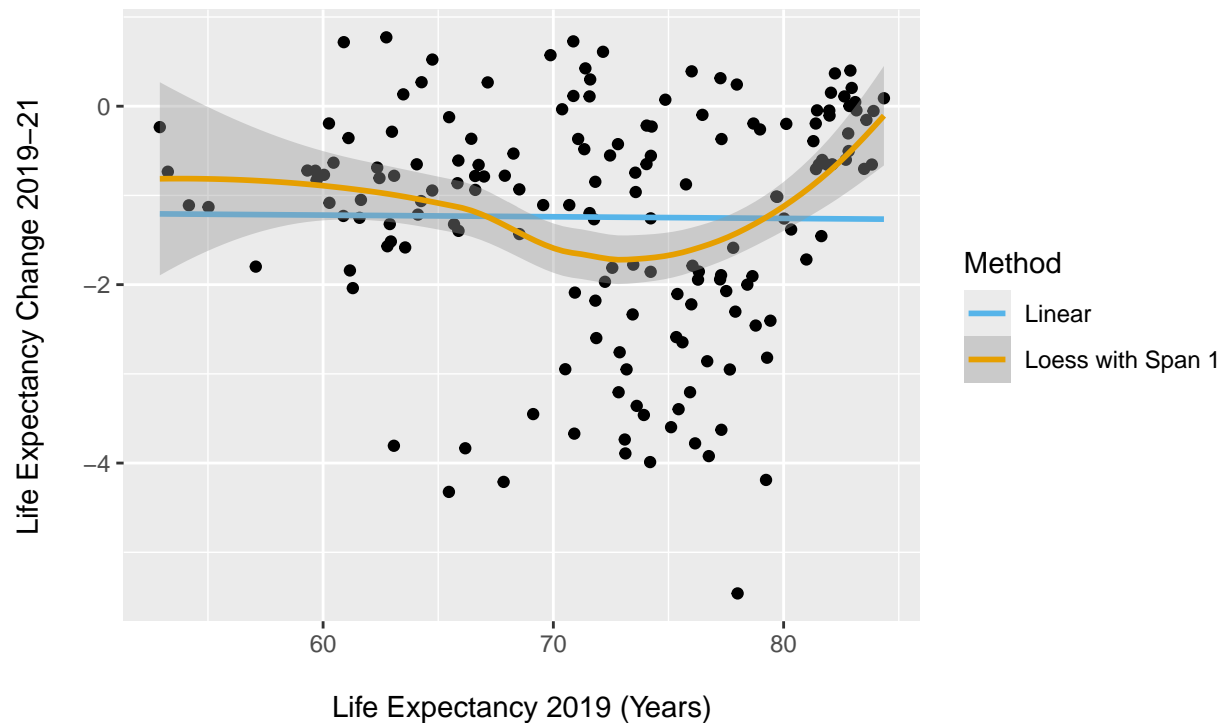
The Americas had the most significant average decline (-2.03 years), followed closely by Europe and Asia. As shown in the plot, the change in Asia and Europe is -1.26 and -1.27 respectively showing life expectancy change distribution is almost same in these two continents.

Africa, while still seeing a decrease, had the smallest average decline among those that decreased.

**Is the change in life expectancy related to 2019 life expectancy?**

```
ggplot(merge.data.required, aes(x = `Life Expectancy 2019`, y = `Life Expectancy Change 2019-21`)) +
  geom_point() +
  geom_smooth(aes(color = "Linear"), method = "lm", formula = y ~ x, se = FALSE) +
  geom_smooth(aes(color = "Loess with Span 1"), formula = y ~ x, method = "loess", span = 1) +
  scale_color_manual(values = cb_palette) +
  labs(title = "Figure 12: Life Expectancy Change 2019-21 VS. Life Expectancy 2019",
  subtitle = "Depicted by Linear Regression Model and Loess with Span 1\n",
  x = "\nLife Expectancy 2019 (Years)",
  y = "Life Expectancy Change 2019-21\n",
  color = "Method")
```

Figure 12: Life Expectancy Change 2019–21 VS. Life Expectancy 2019  
Depicted by Linear Regression Model and Loess with Span 1

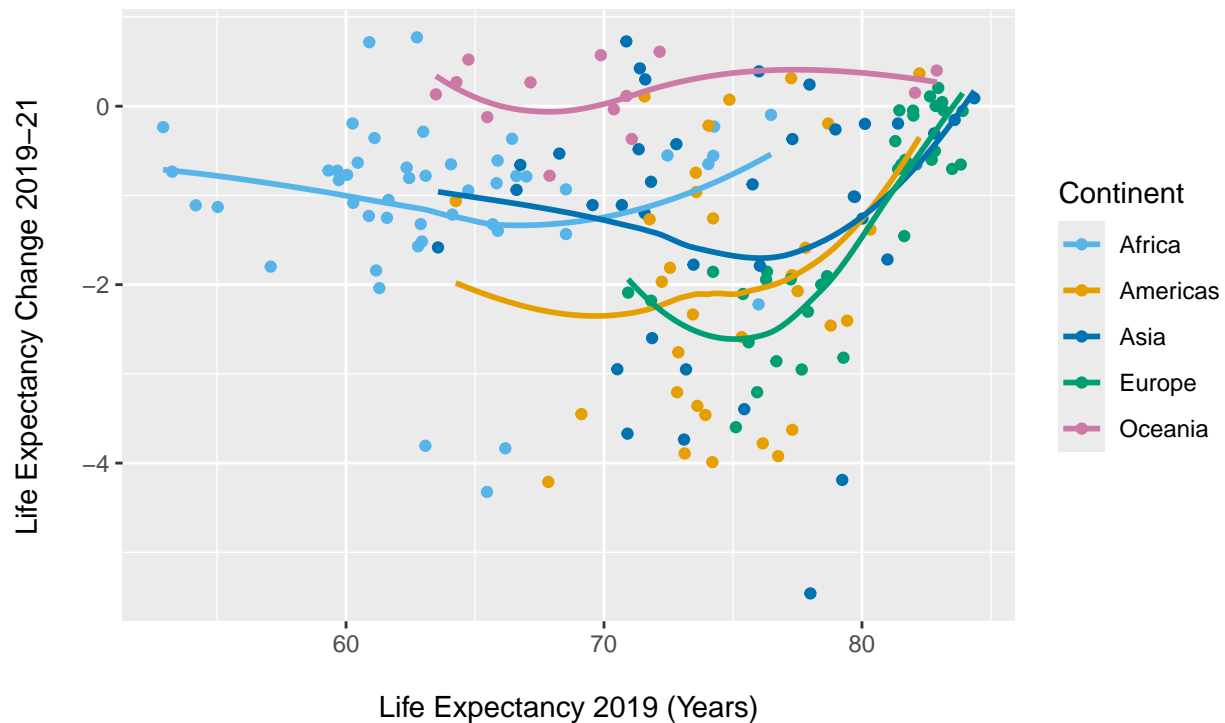


There is no systematic relationship between those two. There doesn't appear to be a strong linear correlation between life expectancy in 2019 and the change by 2021; the cyan linear trend line is quite flat. The orange curve suggests that there may be a non-linear pattern where regions with both lower and higher life expectancies in 2019 experienced smaller changes by 2021, while those with middle-range life expectancies in 2019 saw more variability in the change. For countries with a lower life expectancy in 2019, the change in life expectancy from 2019 to 2021 varies widely. As life expectancy in 2019 increases, the change in life expectancy from 2019 to 2021 initially shows a slight decrease. Then, for countries with the highest life expectancy in 2019, the change in life expectancy from 2019 to 2021 increases.

Does that relationship vary by continent, and if so, how?

```
ggplot(merge.data.required, aes(x = `Life Expectancy 2019`,
  y = `Life Expectancy Change 2019-21`, color = Continent)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE, formula = y ~ x, span = 1) +
  scale_color_manual(values = cb_palette) +
  labs(title = "Figure 13: Life Expectancy Change 2019-21 VS. Life Expectancy 2019",
    subtitle = "Grouped By Continent, Depicted by Loess with Span 1\n",
    x = "\nLife Expectancy 2019 (Years)",
    y = "Life Expectancy Change 2019-21\n")
```

Figure 13: Life Expectancy Change 2019–21 VS. Life Expectancy 2019  
Grouped By Continent, Depicted by Loess with Span 1



The relationship does vary by continent. The shape of the loess curve (span 1 and degree 2) for Africa, Oceania, and Asia looks like the one in the combined plot. Americas is little off but does slightly resemble the combined plot. However, the shape of the loess curve plot for Europe does not resemble the combined plot. Europe presents a distinct pattern with a positive trend, indicating that countries with higher life expectancy in 2019 saw a smaller decrease in life expectancy by 2021.

### 3.Relationship between GDP and life expectancy changes:

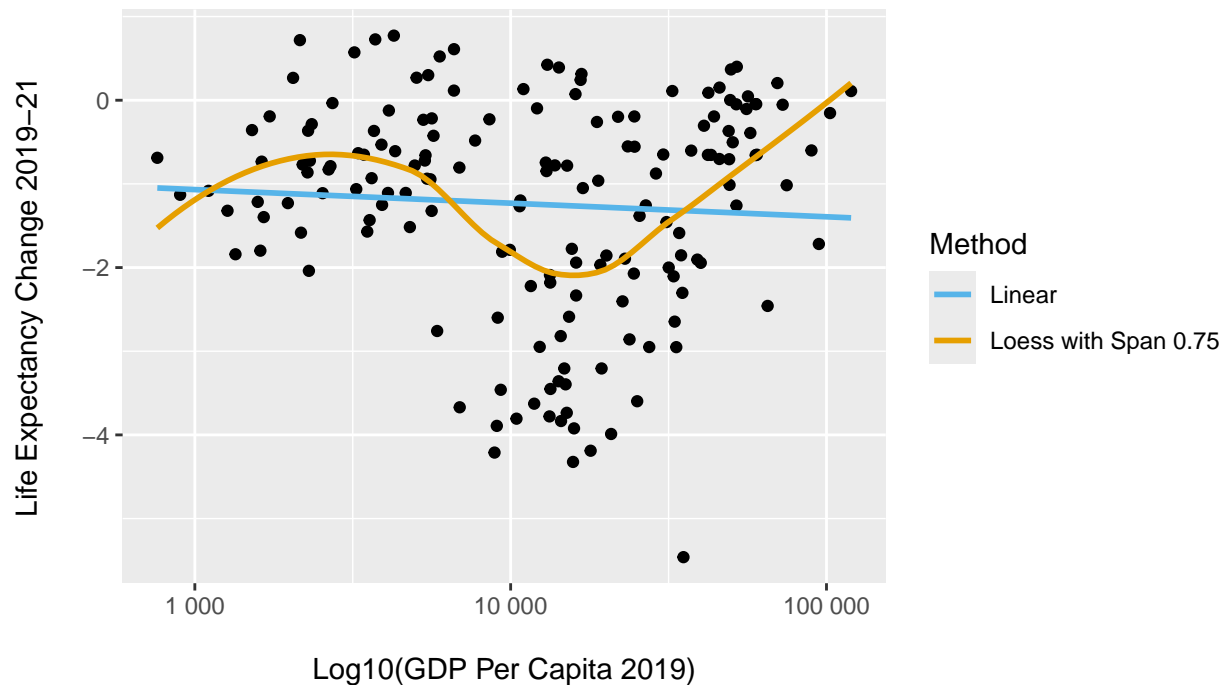
How does the life expectancy change from 2019 to 2021 vary with GDP per capita in 2019? Can the trends be well-described by a simple model such as a linear model, or is a more complicated model required?

```
ggplot(merge.data.required, aes(x = `GDP Per Capita 2019`, y = `Life Expectancy Change 2019-21`)) +
  geom_point() +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  geom_smooth(aes(color = "Linear"), method = "lm", formula = y ~ x,
    se = FALSE) +
  geom_smooth(aes(color = "Loess with Span 0.75"), method = "loess", formula = y ~ x,
    span = 0.75, se = FALSE) +

  scale_color_manual(values = cb_palette) +
  labs(title = "Figure 14: Life Expectancy Change 2019-21 VS. Log-Transformed GDP
    Per Capita 2019",
    subtitle = "Depicted by Linear Regression Model and Loess with Span 0.75\n",
    x = "\nLog10(GDP Per Capita 2019)",
    y = "Life Expectancy Change 2019-21\n",
    color = "Method")
```

Figure 14: Life Expectancy Change 2019–21 VS. Log–Transformed GDP Per Capita 2019

Depicted by Linear Regression Model and Loess with Span 0.75



Initially, for lower values of log-transformed GDP per capita, the LOESS line shows a slight downward trend, suggesting that increases in GDP do not necessarily correspond to increases in life expectancy. As GDP per capita increases further, the LOESS line bottoms out and then begins to ascend, implying that beyond a certain threshold, further increases in GDP per capita are associated with increases in life expectancy. The line peaks and then levels off, indicating that at higher levels of GDP per capita, the life expectancy change becomes less sensitive to GDP growth.

The Loess line indicates a non-linear relationship between the log-transformed GDP per capita and the change in life expectancy, with a noticeable curve in the trend. This suggests that the change in life expectancy does not increase or decrease uniformly with GDP per capita.

The linear model's line is relatively flat and does not capture the non-linear trend shown by the Loess model, indicating that a simple linear model may not be the best fit for this data.

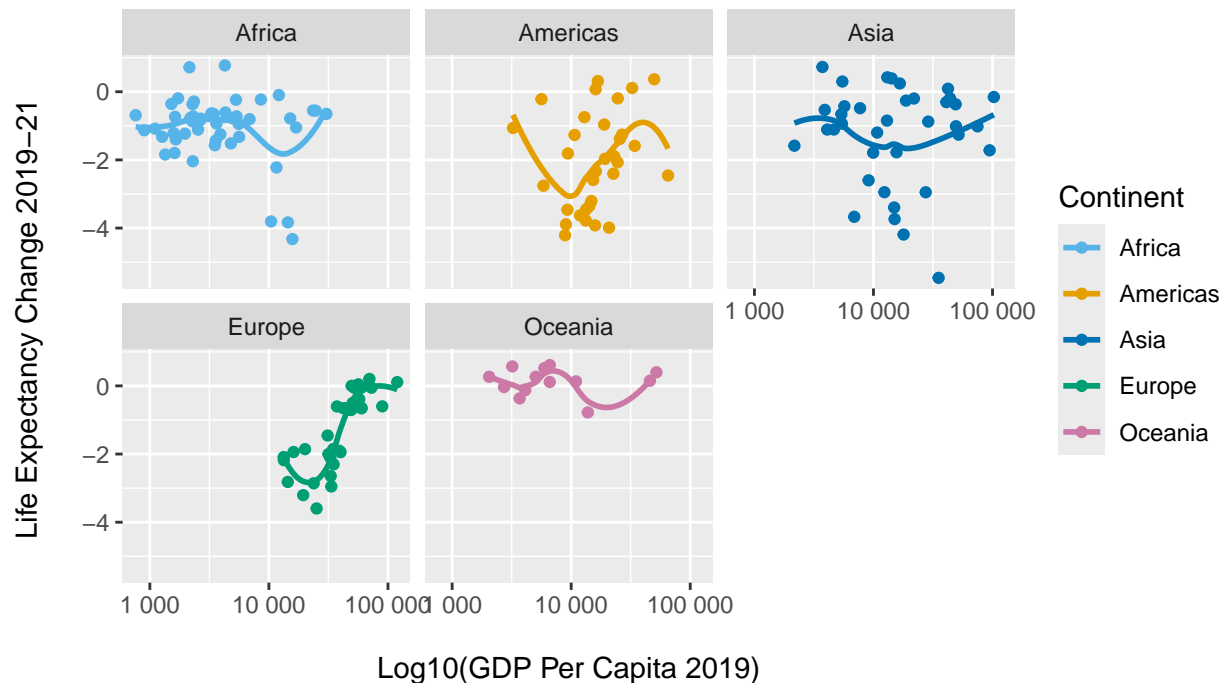
**Is the pattern the same or different for every continent? If some continents are different, which ones, and how is the relationship different in those continents?**

```
ggplot(merge.data.required, aes(x = `GDP Per Capita 2019`,
  y = `Life Expectancy Change 2019-21`, color = Continent)) +
  geom_point() +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  geom_smooth(method = "loess", formula = y ~ x, span = 0.75, se = FALSE) +
  facet_wrap(Continent~.) +
  scale_color_manual(values = cb_palette) +
  labs(title = "Figure 15: Life Expectancy Change 2019-21 VS. Log-Transformed GDP
    Per Capita 2019",
    subtitle = "Faceted By Continent, Depicted by Loess with Span 0.75",
```

```
x = "\nLog10(GDP Per Capita 2019)",
y = "Life Expectancy Change 2019-21\n")
```

Figure 15: Life Expectancy Change 2019–21 VS. Log–Transformed GDP Per Capita 2019

Faceted By Continent, Depicted by Loess with Span 0.75



From the graph, the pattern of life expectancy change in relation to GDP per capita is not uniform across all continents. Here's how the relationship varies:

**Africa:** The trend for Africa in the faceted plot shows an initial slight increase in life expectancy change with a subsequent increase as GDP per capita rises. This is followed by a sharp decrease and then a noticeable increase, which indicates a non-linear relationship. This pattern aligns with the global LOESS curve which also depicts a non-linear relationship, suggesting that the trend in Africa follows the global trend.

**Americas:** The Americas pattern starts with a steep decline, then a significant rise, peaks, and ends with a slight decline. The initial part of this pattern shows some similarity to the global curve. However, towards the higher GDP per capita range, the Americas trend diverges from the global curve.

**Asia:** Asia's pattern follows the global LOESS curve for the available data points, starting from a peak displaying a gradual decline in life expectancy change as GDP per capita increases. This suggests that the global LOESS curve is a good representation of the trend within Asia.

**Europe:** Europe's trend in the faceted plot shows a steep initial decline, followed by a very sharp rise and then leveling off at higher GDP per capita levels. This pattern is more extreme than the global curve.

**Oceania:** The trend for Oceania in the faceted plot exhibits high variability, likely due to fewer data points, resulting in a trend line that may overfit the available data. The pattern shown is not as clearly aligned with the global curve, indicating that a different relationship may exist for Oceania, or that additional data is required for a reliable comparison.

In conclusion, while Africa, the Americas, and Asia display patterns of life expectancy change in relation to GDP per capita that bear resemblance to the global trend, Europe and Oceania show distinct patterns that

differ from the global trend. These differences suggest that the relationship between GDP per capita and life expectancy change is influenced by continent-specific factors, which are not fully reflected in a global analysis.

**4. Modeling:** Can we fit an interpretable model (e.g. a linear model, perhaps with simple transformations) that explains the change in life expectancy from 2019 to 2021 using:

- Continent
- Life expectancy in 2019
- GDP per capita in 2019

#### A) Life Expectancy Change 2019-21 VS. Life Expectancy 2019

In an effort to understand the relationship between the change in life expectancy from 2019 to 2021 and the initial life expectancy in 2019, a loess model was fitted using a second-degree with a span of 1.

```
model.lo <- loess(merge.data.required$`Life Expectancy Change 2019-21`  
                  ~ merge.data.required$`Life Expectancy 2019`, degree = 2, span = 1)  
  
summary(model.lo)
```

```
## Call:  
## loess(formula = merge.data.required$`Life Expectancy Change 2019-21` ~  
##       merge.data.required$`Life Expectancy 2019`, span = 1, degree = 2)  
##  
## Number of Observations: 166  
## Equivalent Number of Parameters: 3.45  
## Residual Standard Error: 1.203  
## Trace of smoother matrix: 3.69 (exact)  
##  
## Control settings:  
##   span      : 1  
##   degree    : 2  
##   family    : gaussian  
##   surface   : interpolate      cell = 0.2  
##   normalize : TRUE  
##   parametric: FALSE  
##   drop.square: FALSE
```

The loess model output indicates a non-linear relationship between the change in life expectancy from 2019 to 2021 and life expectancy in 2019, based on 166 observations. With a span of 1, it suggests a highly responsive model to the data, and the degree of 2 points to the use of quadratic fitting. The residual standard error of 1.203 means that the predictions are, on average, approximately 1.203 units away from the actual values. The effective complexity of the model is moderate, with an equivalent number of parameters at 3.45, and the smoothness of the fit is captured by the trace of the smoother matrix at 3.69.

We created a dataframe named `model.lo.df`. This dataframe encompasses the original variables from our dataset, alongside two additional columns: `.fitted`, which contains the fitted values obtained from the 2 degree loess model with a span of 1, and `.resid`, representing the residuals.

```
model.lo.df <- data.frame(merge.data.required,  
                          .fitted = fitted.values(model.lo),  
                          .resid = residuals(model.lo))
```

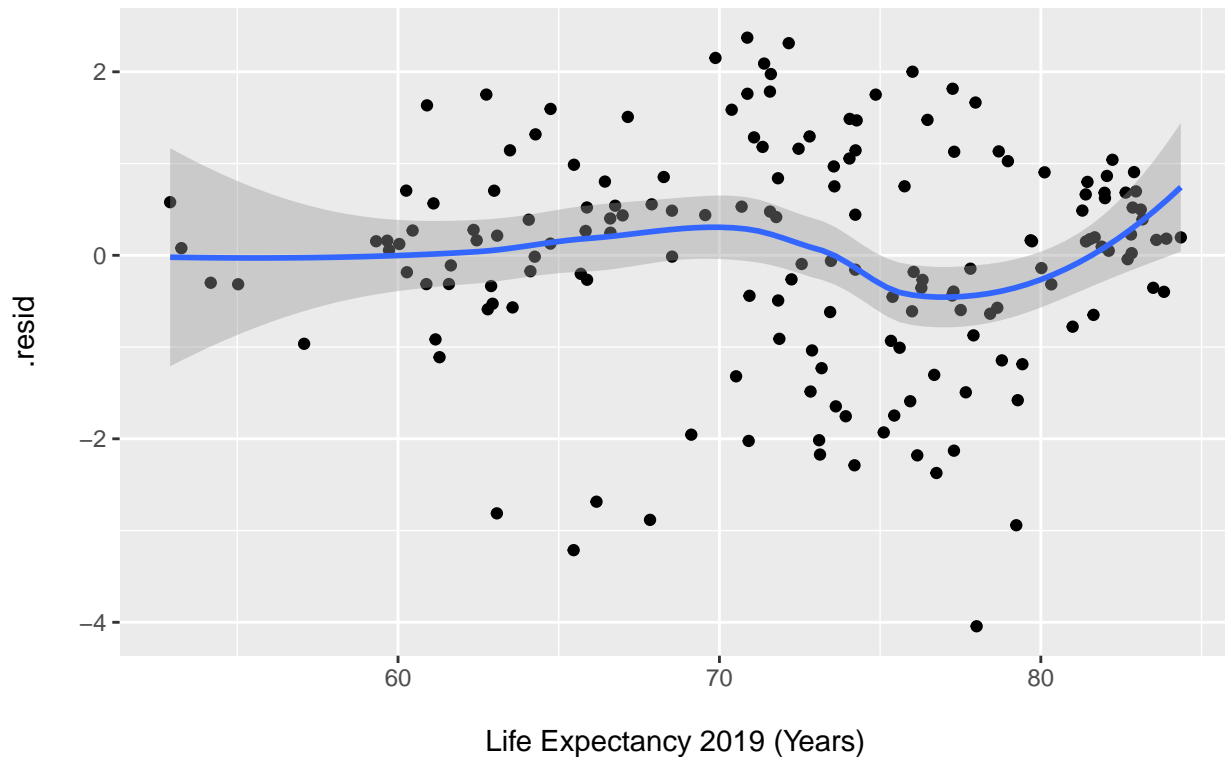
#### Residual Plots :

1) **Lack-of-trend check:** Residuals Against Life Expectancy 2019

```
ggplot(model.lo.df, aes(x = Life.Expectancy.2019, y = .resid)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Figure 16: Residuals Against Life Expectancy 2019\n",
       x = "\nLife Expectancy 2019 (Years)",
       y = ".resid\n")
```

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

Figure 16: Residuals Against Life Expectancy 2019

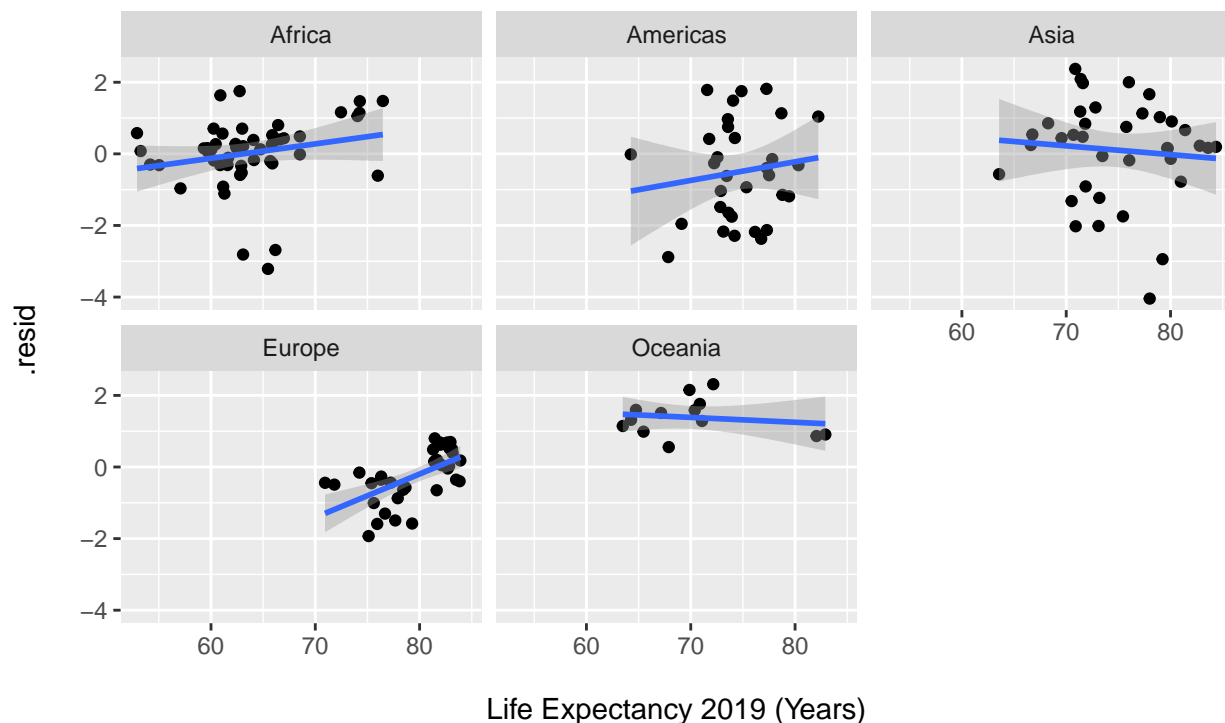


The residual plot against life expectancy in 2019 serves as a diagnostic tool to assess the fit of the loess model (degree 2, span 1). The spread of residuals is largely horizontal and centered around zero, which is an indication of a good fit across the range of life expectancies. The slight upturn at higher life expectancy values suggests that the model may be underpredicting the change in life expectancy for countries with longer lifespans, a nuance that could be addressed by further model adjustments or exploring additional variables that impact these nations differently.

## 2) No-interaction check: Residuals Against Life Expectancy 2019, Faceted By Continent

```
ggplot(model.lo.df, aes(x = Life.Expectancy.2019, y = .resid)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x) +
  facet_wrap(~ Continent) +
  labs(title = "Figure 17: Residuals Against Life Expectancy 2019",
       subtitle = "Faceted By Continent, Depicted by Linear Regression Model\n",
       x = "\nLife Expectancy 2019 (Years)",
       y = ".resid\n")
```

Figure 17: Residuals Against Life Expectancy 2019  
Faceted By Continent, Depicted by Linear Regression Model



The continent-specific residual plots reveal the differences in how the loess model accounts for changes in life expectancy across continents. The plots for Africa, Asia, and Oceania show a random distribution of residuals, suggesting the model's changes are well explained by life expectancy in 2019 alone. However, the plots for the Americas and Europe show patterns that deviate from randomness, with the residuals indicating potential systematic errors in the model's predictions. This suggests that for these continents, additional factors not included in the model might be influencing the changes in life expectancy, or that there are interaction effects between life expectancy and continent that are not being captured.

**Goodness-of-fit: R-Squared**

```
# Proportion of Variance Explained
var(model.lo.df$.fitted) / (var(model.lo.df$.fitted) + var(model.lo.df$.resid))
```

```
## [1] 0.1053124
```

```
# Proportion of Variance Unexplained
var(model.lo.df$.resid) / (var(model.lo.df$.fitted) + var(model.lo.df$.resid))
```

```
## [1] 0.8946876
```

The R-squared value of 0.1053, for the loess model indicates that approximately 10.53% of the variance in the life expectancy change from 2019 to 2021 can be explained by the life expectancy in 2019. This suggests that life expectancy in 2019 is a factor in the change observed over the two years, but there is still a large proportion of the variance that is unexplained by this model. It might be beneficial to consider additional variables like economic, environmental, or healthcare-related variables or different types of models to better understand the factors affecting changes in life expectancy.



## Interpretation of the model:

The analysis conducted using a loess model and residual plots revealed nuanced relationships between life expectancy changes from 2019 to 2021 and the life expectancy in 2019, with variations across continents. The loess model's fit, while generally good, indicated potential underprediction in regions with higher life expectancies and a variance of fit across continents, with Africa, Asia, and Oceania aligning well with the model, and the Americas and Europe exhibiting patterns suggesting a need for additional explanatory variables or model adjustments.

The goodness of fit statistics highlighted that life expectancy in 2019 only modestly explained the changes by 2021, accounting for approximately 10.53% of the variance. This underscores the complexity of predicting life expectancy changes, suggesting that other factors, perhaps economic or health system-related, or even interactions between variables, could provide a more comprehensive understanding of the factors driving changes in life expectancy across different regions.

## B) Life Expectancy Change 2019-21 VS. GDP Per Capita 2019

In an effort to understand the relationship between the change in life expectancy from 2019 to 2021 and the initial GDP Per Capita in 2019, a loess model was fitted using a second-degree with a span of 0.75.

```
model2.lo <- loess(merge.data.required$`Life Expectancy Change 2019-21` ~  
  log10(merge.data.required$`GDP Per Capita 2019`), degree = 2, span = 0.75)  
  
summary(model2.lo)
```

```
## Call:  
## loess(formula = merge.data.required$`Life Expectancy Change 2019-21` ~  
##   log10(merge.data.required$`GDP Per Capita 2019`), span = 0.75,  
##   degree = 2)  
##  
## Number of Observations: 166  
## Equivalent Number of Parameters: 4.74  
## Residual Standard Error: 1.151  
## Trace of smoother matrix: 5.18 (exact)  
##  
## Control settings:  
##   span      : 0.75  
##   degree    : 2  
##   family    : gaussian  
##   surface   : interpolate      cell = 0.2  
##   normalize : TRUE  
##   parametric: FALSE  
##   drop.square: FALSE
```

The loess model summary shows a quadratic relationship between the change in life expectancy from 2019 to 2021 and the logarithm of GDP per capita in 2019 across 166 observations. With a chosen span of 0.75, this model aims for a balance between fitting the data and smoothing out the variations. The residual standard error is relatively low at 1.151, indicating that the model predictions are close to the observed data. The trace of the smoother matrix, a measure of the effective degrees of freedom used in the smoothing process, stands at 5.18, suggesting a moderate level of smoothing.

We created a dataframe named model2.lo.df. This dataframe encompasses the original variables from our dataset, alongside two additional columns: .fitted, which contains the fitted values obtained from the 2 degree loess model with a span of 0.75, and .resid, representing the residuals.

```
model2.lo.df <- data.frame(merge.data.required,
                           .fitted = fitted.values(model2.lo),
                           .resid = residuals(model2.lo))
```

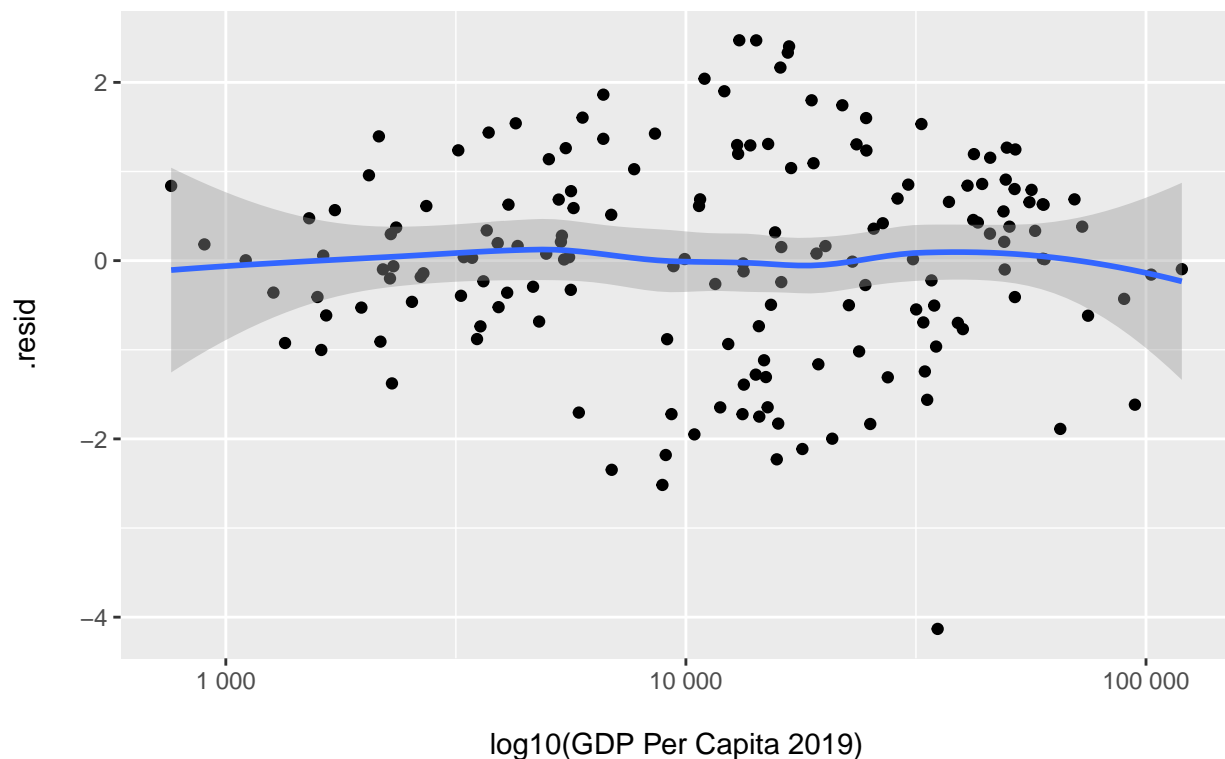
## Residual Plots :

### 1) Lack-of-trend check: Residuals Against GDP Per Capita 2019

```
ggplot(model2.lo.df, aes(x = GDP.Per.Capita.2019, y = .resid)) +
  geom_point() +
  geom_smooth() +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  labs(title = "Figure 18: Residuals Against GDP Per Capita 2019\n",
       x = "\nlog10(GDP Per Capita 2019)",
       y = ".resid\n")
```

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

Figure 18: Residuals Against GDP Per Capita 2019

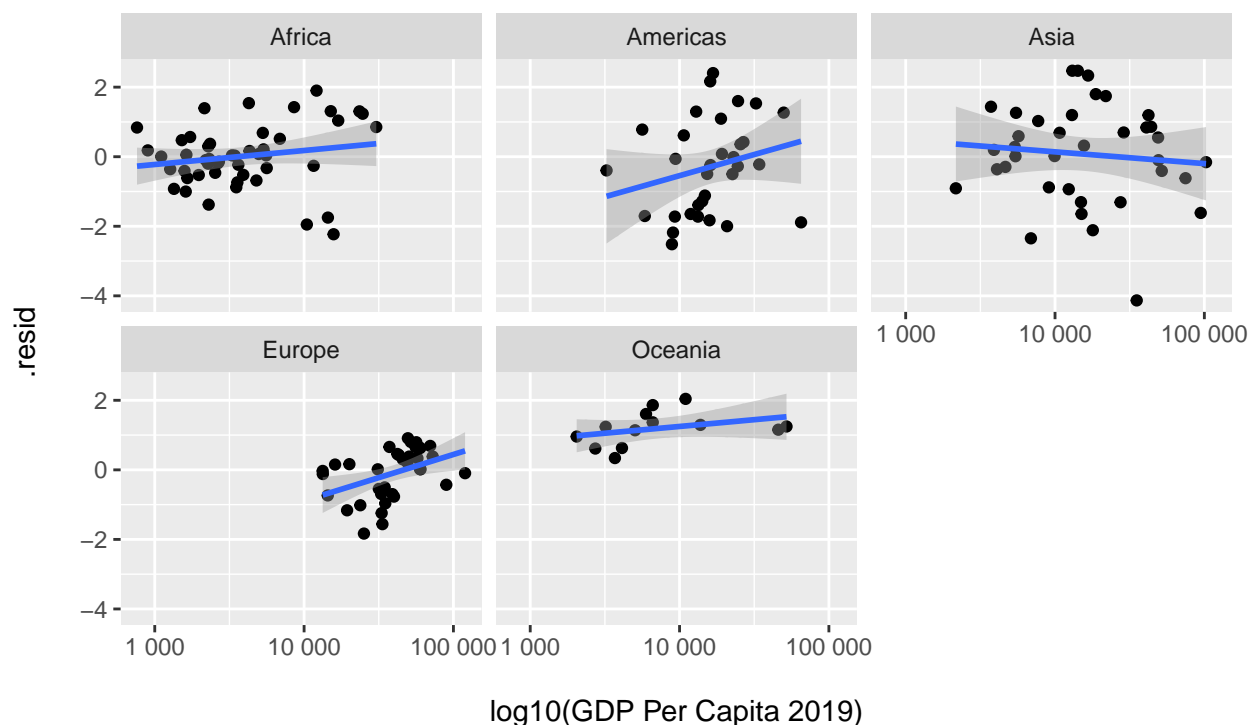


The residual plot against the logarithm of GDP per capita in 2019 shows a spread of data points that do not display any systematic pattern, suggesting that the model's residuals are not biased by GDP levels. This random distribution indicates the model has captured the underlying trend well. However, there is some curvature visible in the loess smoothing line, which implies that there might be a non-linear relationship between GDP and life expectancy change that the model is attempting to capture.

### 2) No-interaction check: Residuals Against GDP Per Capita 2019, Faceted By Continent

```
ggplot(model2.lo.df, aes(x = GDP.Per.Capita.2019, y = .resid)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x) +
  scale_x_log10(labels = scales::number_format(scale = 1)) +
  facet_wrap(~ Continent) +
  labs(title = "Figure 19: Residuals Against GDP Per Capita 2019",
       subtitle = "Faceted By Continent, Depicted by Linear Regression Model",
       x = "\nlog10(GDP Per Capita 2019)",
       y = ".resid")
```

Figure 19: Residuals Against GDP Per Capita 2019  
Faceted By Continent, Depicted by Linear Regression Model



This scatter plot depicts the relationship between residuals and log of GDP Per Capita 2019, faceted by continent. The distribution of residuals suggests variability in how well the model fits the data for different continents. For Africa, Asia, and Oceania, the residuals scatter in a seemingly random pattern along the x-axis, and the corresponding loess smoother is nearly flat, indicating a negligible slope. This pattern implies that a quadratic loess model (degree 2) with a span of 0.75 may effectively represent the relationship in these regions. In contrast, the Americas and Europe exhibit a less random distribution of residuals and fitted smoother with a steeper slope, suggesting that the model may not capture the nuances of the relationship as well in these continents.

**Goodness-of-fit: R-Squared**

```
# Proportion of Variance Explained
var(model2.lo.df$.fitted) / (var(model2.lo.df$.fitted) + var(model2.lo.df$.resid))
```

```
## [1] 0.2088819
```

```
# Proportion of Variance Unexplained
var(model2.lo.df$.resid) / (var(model2.lo.df$.fitted) + var(model2.lo.df$.resid))
```

```
## [1] 0.7911181
```

The R-squared value of 0.208819 for this model indicates that about 20.88% of the variance in the change of life expectancy from 2019 to 2021 can be explained by the log of GDP per capita in 2019. This suggests that the log of GDP per capita is a contributing factor to the changes observed in life expectancy over this period. However, a significant portion, roughly 79%, of the variance remains unaccounted for by this model. To gain a more comprehensive understanding of the determinants of life expectancy changes, it may be useful to explore additional variables or alternative modeling approaches that can capture more of the complexity in the data.

### Interpretation of the model:

The analysis of the change in life expectancy from 2019 to 2021 against GDP per capita in 2019, using a loess regression model, indicates a complex, non-linear relationship. The model, which includes data from various continents, shows a moderate fit with the residuals largely distributed without any clear pattern, suggesting the logarithm of GDP is a significant but not comprehensive predictor of life expectancy changes. The residual standard error is reasonable, hinting at a decent model fit.

However, the goodness-of-fit measure, reflected by the R-squared value, implies that the model only explains around 20.88% of the variation in life expectancy change, leaving a substantial 79% unexplained by GDP per capita alone. The faceted residual plots by continent reveal that the model's effectiveness varies by region, capturing the trend well in Africa, Asia, and Oceania but less so in the Americas and Europe. This suggests that additional factors, which may interact differently across regions, likely influence life expectancy changes alongside GDP.

**Are any interactions necessary? What does your model tell you quantitatively? Is there a good way of visualizing the model? What countries does your model not fit well?**

For Life Expectancy Change 2019-21 VS. Life Expectancy 2019, the analysis indicates that life expectancy in 2019 is not a strong predictor for its change by 2021, with a low proportion of variance explained by the model. No clear indications of interactions were observed in the residual plots. The model's residuals suggest that it fits less effectively for countries with higher initial life expectancies. Visually, the model can be represented with scatter plots and loess curves to identify the trends and deviations. After examining the residuals it is observed that model doesn't fit Oman and Botswana having residuals -4.04296695, -3.21364553 respectively, as these residuals are outliers.

Regarding Life Expectancy Change 2019-21 VS. GDP Per Capita 2019, the log of GDP per capita in 2019 showed a moderate explanatory power for life expectancy changes. The model could be improved by considering interactions, particularly with continent, as the faceted residual plots indicated varied model performance across different regions. Quantitatively, the R-squared value was higher compared to the first model but still left a substantial amount of variance unexplained. A combined visualization using scatter plots with loess curves, faceted by continent, allowed us to discern which countries or regions the model does not fit well, such as those in the Americas and Europe where the residuals deviated from the expected random distribution. After examining the residuals it is observed that model doesn't fit Oman having residuals -4.131762721, as this residual is outlier.

### Conclusions -

**GDP and Life Expectancy in 2021:** Transforming GDP Per Capita with a logarithm linearizes its relationship with Life Expectancy in 2021, revealing a positive correlation where higher GDP per capita is associated with greater life expectancy.

**Life Expectancy Changes from 2019 to 2021:** A general decrease in life expectancy during this period was observed, with Oman experiencing the largest decline, potentially due to the impact of COVID-19. Oceania stands out as the exception, showing an increase in life expectancy, which is not directly correlated

with the region's 2019 life expectancy figures. The variation across continents suggests diverse factors at play beyond just economic indicators.

**Relationship Between GDP and Life Expectancy Changes:** The connection between GDP changes and life expectancy is non-linear, and a loess curve effectively captures this complex relationship on both a global scale and when analyzed by continent.

**Modeling:** The relationship between Life Expectancy Change 2019-21 and Life Expectancy 2019 is best modeled non-linearly, with a loess model fitting most appropriately due to the absence of a discernible pattern. A two-degree loess model with a span of 0.75 is optimal for examining the relationship between Life Expectancy Change and GDP Per Capita 2019, although it fits certain continents better than others, indicating regional disparities in the factors affecting life expectancy changes.

The global trend indicates a reduction in life expectancy in 2021 relative to 2019, with the COVID-19 pandemic likely being a significant contributing factor. The correlation between this change in life expectancy and GDP is not strong, suggesting that economic strength alone is not predictive of these life expectancy trends. Even with a non-linear and flexible approach using a loess model, only 20% of the variation in life expectancy changes could be explained, pointing to the complexity of the underlying factors affecting global health outcomes.