

Data Analysis for Developmental Indicators

Group 3:
Iqra Bismi
Megha Gupta
Saniya Lande
Madhura Pandit

Introduction:

This dataset includes information about Development Indicators in different Continents and countries between years 1960-2014. Following are the key parameters:

- GDP per capita
- Life Expectancy
- Child Mortality per 1000 births
- CO2 Emissions per person

The analysis on this dataset will help us understand better about quality of life in different countries and continents. We have compared various parameters through visualization. These are the following steps of our analysis.

Data wrangling

Using Pandas in Jupyter Notebook we cleaned up the data by editing the column headings, setting the index starting from 1 and by filling the null values with bfill function (i.e backward filling the null values).

```
#editing the column name
df.columns= df.columns.str.capitalize() #making the only the first letter in uppercase
df.columns=["Date", "Continent", "Sub_region", "Country", "GDP_per_Capita", "GDP_grouped", "Life_expectancy", "Life_expectancy_grouped", "Co2_emissions_per_person", "Child_mortality_rate"]
df.head()
```

	Date	Continent	Sub_region	Country	GDP_per_Capita	GDP_grouped	Life_expectancy	Life_expectancy_grouped	Co2_emissions_per_person	Child_mortality_rate
0	1960-11-11 00:00:00	Asia	Southern Asia	Afghanistan	1210	<5000	38.6	0-40		0.0461
1	1960-02-21 00:00:00	Europe	Southern Europe	Albania	2790	<5000	62.7	>60		1.2400
2	1960-09-23 00:00:00	Africa	Northern Africa	Algeria	6520	>= 5000 and <20000	52.0	>=50 and <= 60		0.5540
3	1960-08-22 00:00:00	Europe	Southern Europe	Andorra	15200	>= 5000 and <20000	NaN	NaN		NaN
4	1960-07-24 00:00:00	Africa	Middle Africa	Angola	3860	<5000	42.4	>=40 and < 50		0.0975

```
#counting the number of null values if any
k = df.isna().sum().sum()
#checking the values column-wise
l = df.isna().sum()
print(k)
print(l)
```

```
632
Date                0
Continent            0
Sub_region           0
Country              0
GDP_per_Capita       0
GDP_grouped          0
Life_expectancy      30
Life_expectancy_grouped 30
Co2_emissions_per_person 547
Child_mortality_rate  25
dtype: int64
```

```
# above data tell us there are null values in four categories
#filling the null values using bfill() to backward fill the values
df["Life_expectancy"]=df["Life_expectancy"].bfill()
df["Life_expectancy_grouped"]=df["Life_expectancy_grouped"].bfill()
df["Co2_emissions_per_person"]=df["Co2_emissions_per_person"].bfill()
df["Child_mortality_rate"]=df["Child_mortality_rate"].bfill()
k= df.isna().sum().sum()
l=df.isna().sum()
print(k)
print(l) #null values are removed
```

```
0
Date                0
Continent            0
Sub_region           0
```

```
#setting the index so that index starts from 1
k = [i for i in range(1,len(df)+1)]
df.set_index([pd.Index(k)])
```

	Date	Continent	Sub_region	Country	GDP_per_Capita	GDP_grouped	Life_expectancy	Life_expectancy_grouped	Co2_emissions_per_person	Child_
1	1960-11-11 00:00:00	Asia	Southern Asia	Afghanistan	1210	<5000	38.6	0-40	0.0461	
2	1960-02-21 00:00:00	Europe	Southern Europe	Albania	2790	<5000	62.7	>60	1.2400	
3	1960-09-23 00:00:00	Africa	Northern Africa	Algeria	6520	>= 5000 and <20000	52.0	>=50 and <= 60	0.5540	
4	1960-08-22 00:00:00	Europe	Southern Europe	Andorra	15200	>= 5000 and <20000	42.4	>=40 and < 50	0.0975	
5	1960-07-24 00:00:00	Africa	Middle Africa	Angola	3860	<5000	42.4	>=40 and < 50	0.0975	
...
8906	2000-06-26 00:00:00	Oceania	Micronesia	Kiribati	2010	<5000	59.5	>=50 and <= 60	0.3480	

Data Transformation in Tableau

Describe Field



Years

Role: Discrete Dimension
 Type: Calculated Field
 Contains NULL: Unknown
 Status: Valid

Formula

DATEPART('year', [date])

Domain (20 of 55 members)

1960
 1961
 1962
 1963
 1964
 1965
 1966

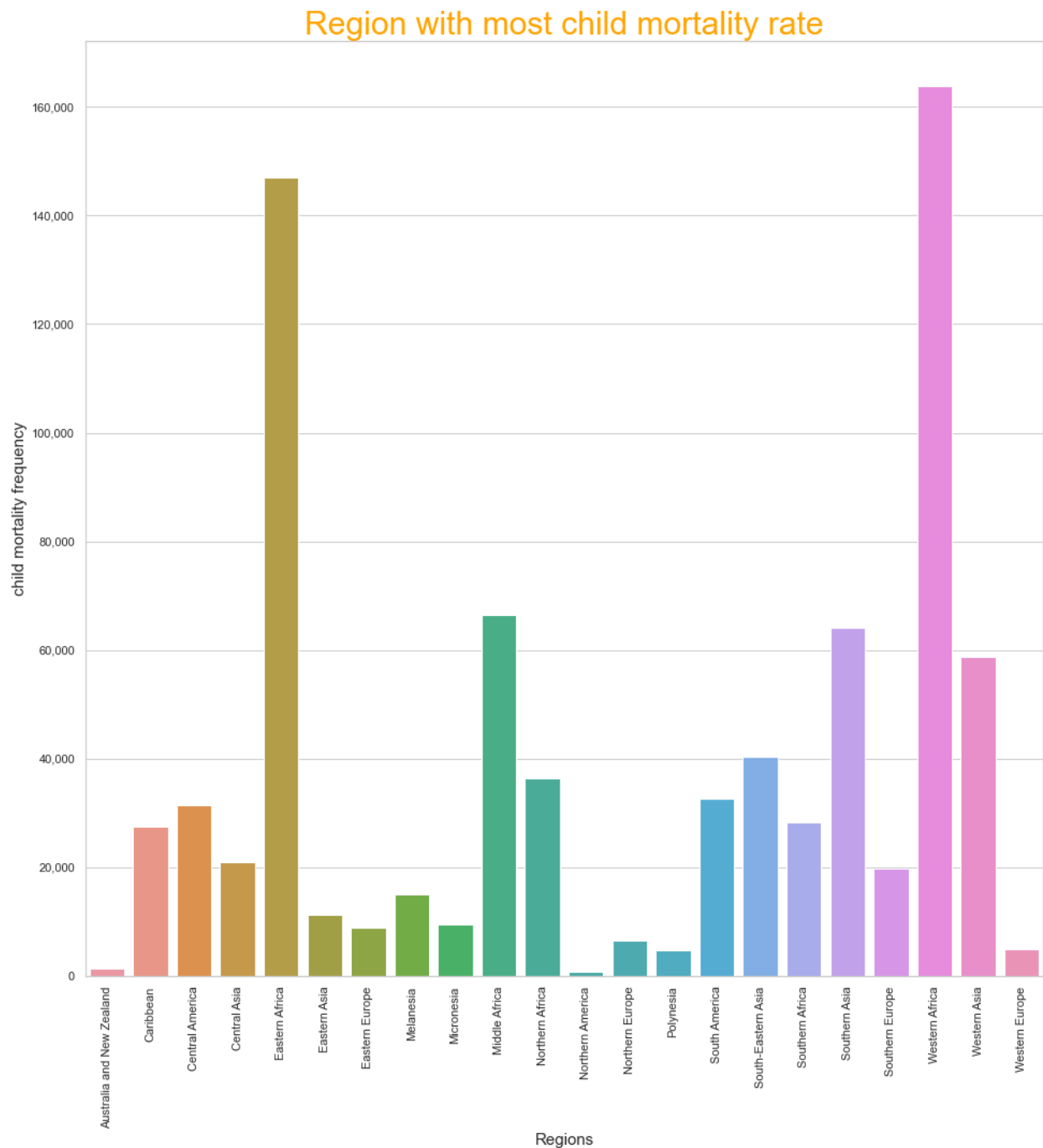
Load

Copy

Data Analysis

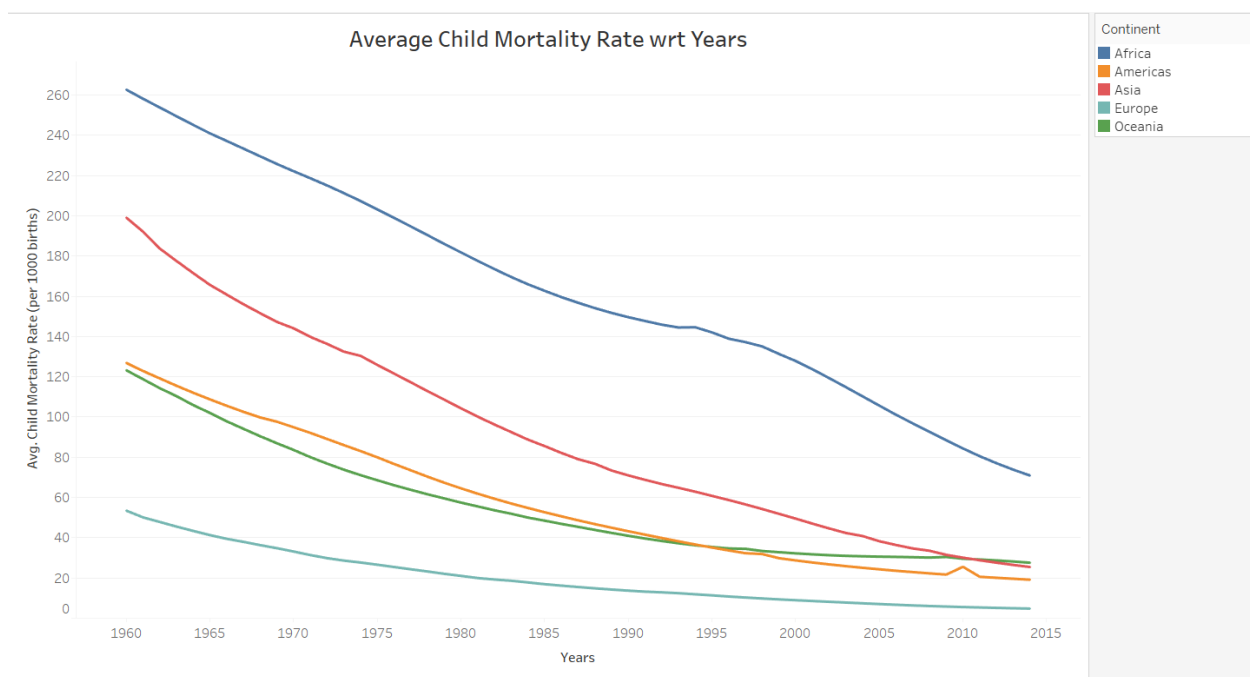
We used Tableau and Seaborn for Data Analysis.

- Which region had the highest Child Mortality rate for the years 1960-2014?



As we can see from the above bar-chart, the Eastern and Western Africa had the highest child mortality rate (around 150k and 165k respectively) whereas Northern America had the lowest rate (merely around 1k).

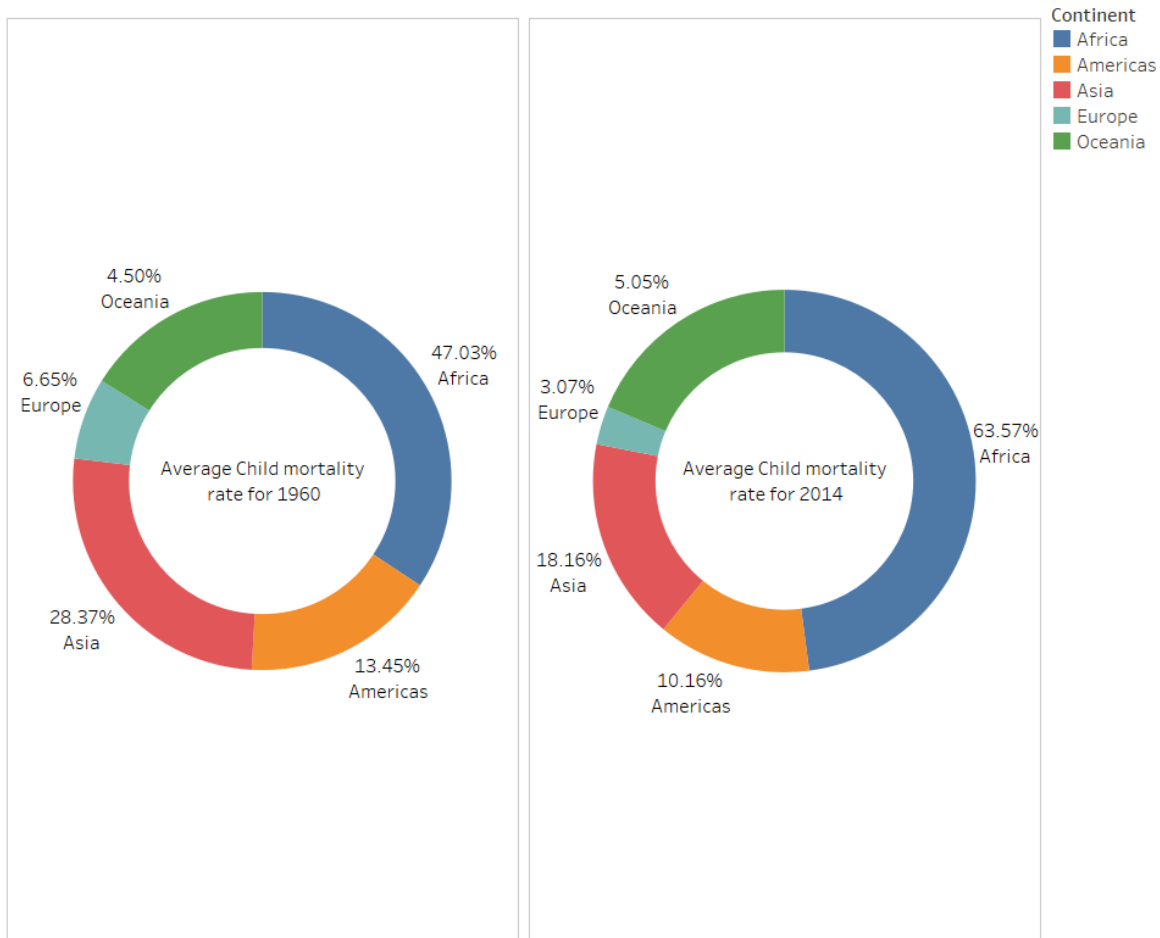
- **To study the trend of Child Mortality Rate wrt years and Continent**



The line chart above clearly shows that there is a decreasing trend for child Mortality rate throughout the years. Africa showed the highest whereas Oceania showed the least child mortality rate.

- **Average Mortality Rate for year 1960 and 2014:**

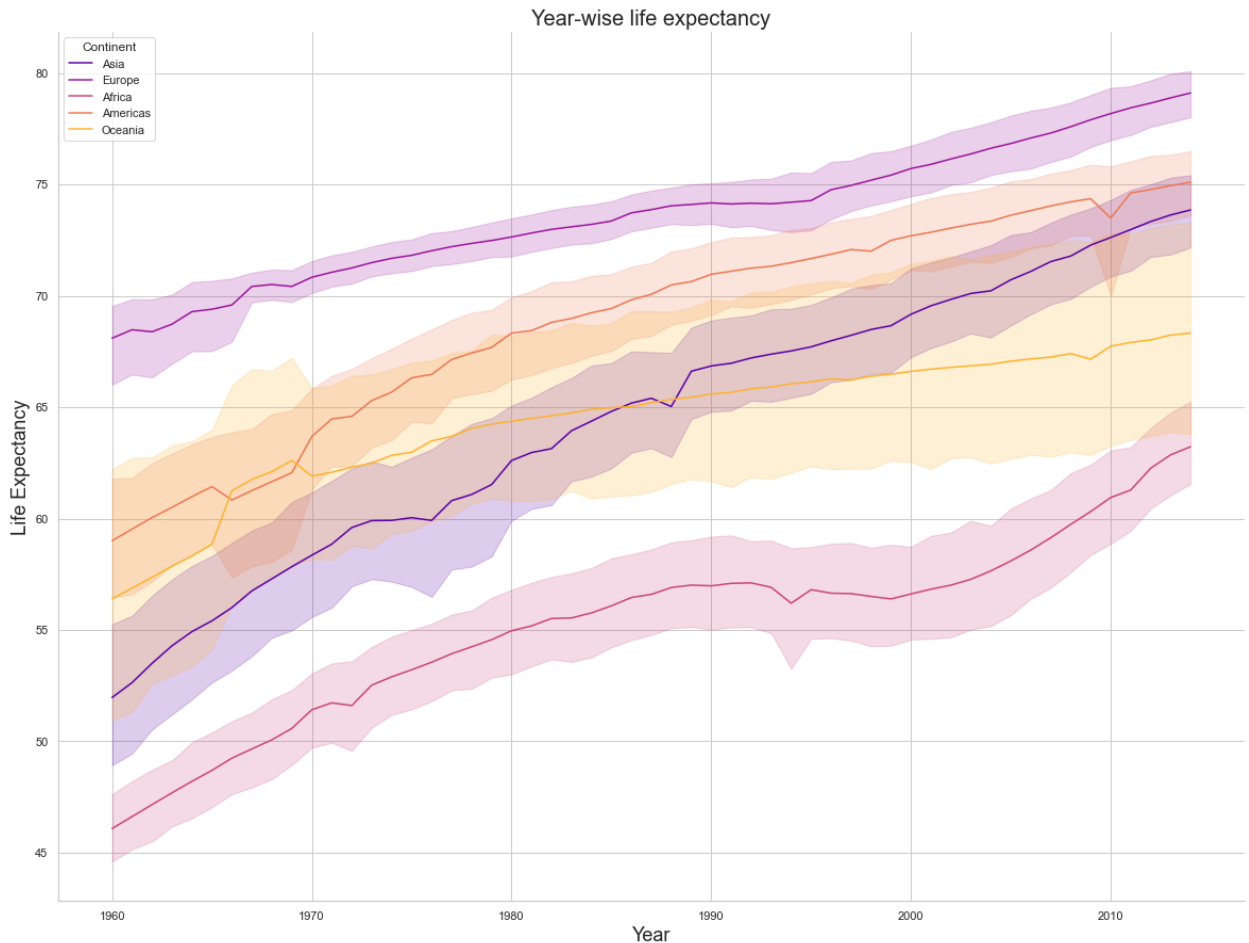
In the below donut chart, we compared the Average child Mortality rate in 1960 and 2014.



As we can see, In Africa, the child Mortality rate has increased by 16.54% whereas, Child Mortality for Asia and Americas collectively has decreased by 13.5%. In Europe the number just decreased to half. However, little fluctuations are observed in Oceania.

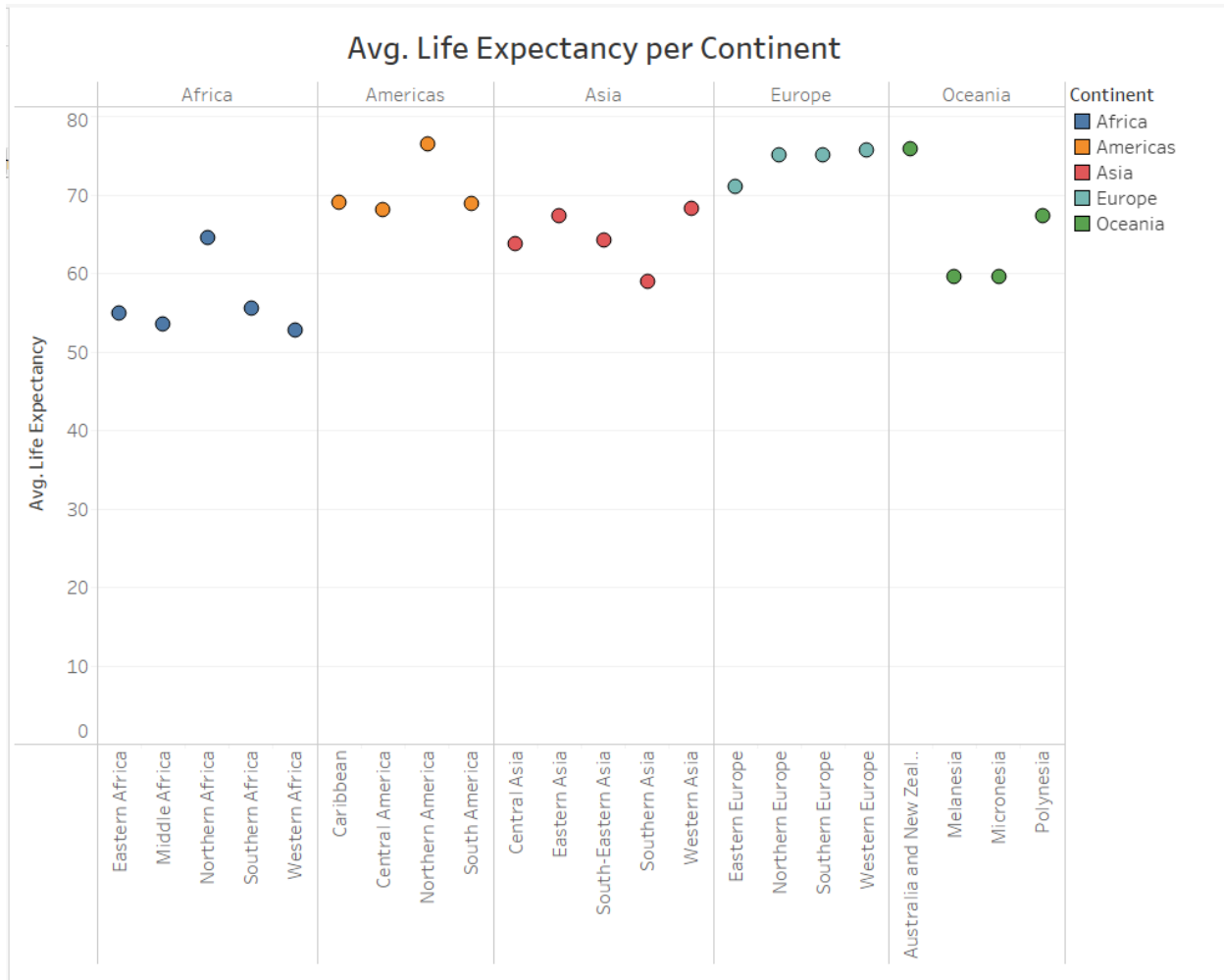
- **Life Expectancy Rate wrt years:**

We would like to see the trend followed by Life Expectancy in the last 50 years.



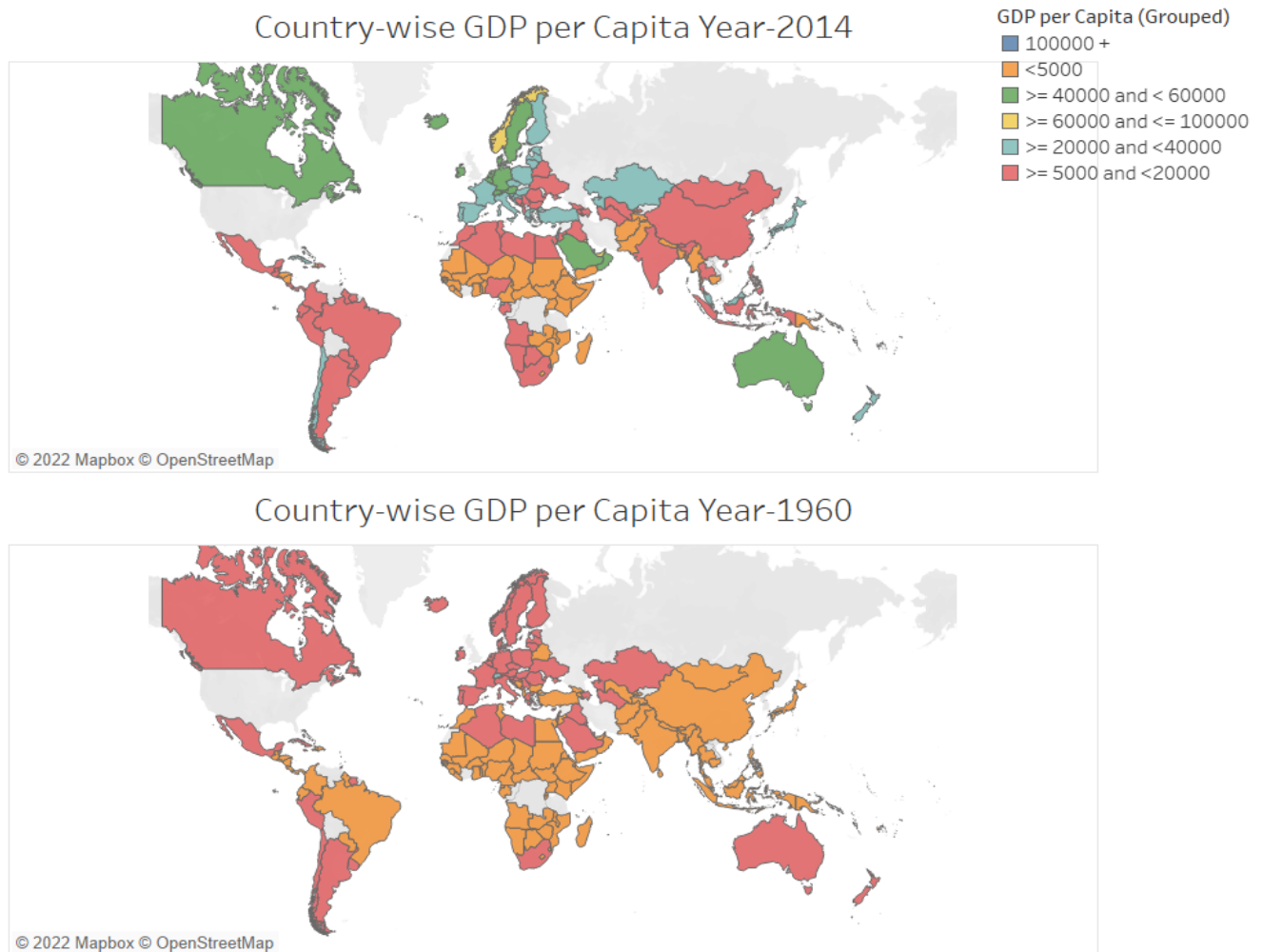
As measured in 5 continents (Asia, Europe, Africa, America, Oceania) Life Expectancy rate has increased significantly assuming the medical care got improved and individuals received easy access to healthcare.

- **Average Life Expectancy per continent:**



As the graph depicts the average life expectancy. We can deduce that people in America and Europe live more as compared to other countries. In Asia and Africa, Life Expectancy is approximately the same, however, we can see fluctuations in Oceania Continent.

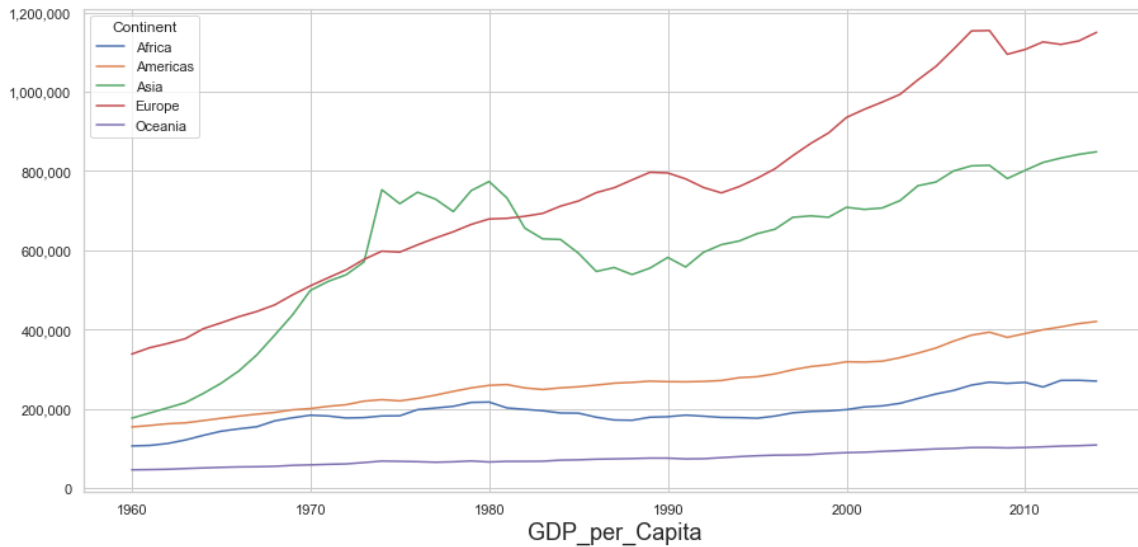
- **Was there any difference in the GDP per capita for various countries for the years 1960 to 2014?**



The Geospatial map above shows that the GDP per capita (grouped) was very low (<5K) for a lot of countries in Africa, Asia and South America, whereas countries like Switzerland and Luxembourg from Europe had the highest GDP per capita (>=20K and < 40K) for the year 1960.

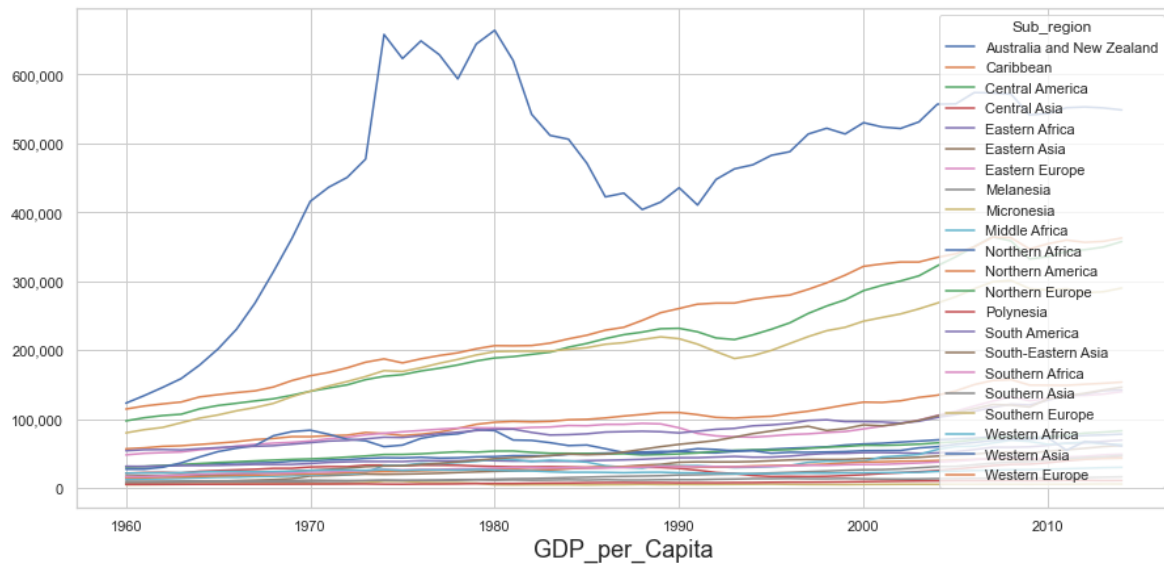
Qatar showed the highest GDP per capita (100K+) for the year 2014. There was a significant increase in the GDP per capita for most of the countries throughout the world in 2014, with exceptions observed in some parts of Africa.

- **Let's study the trend for the GDP per Capita for various continents over the years 1960-2014**



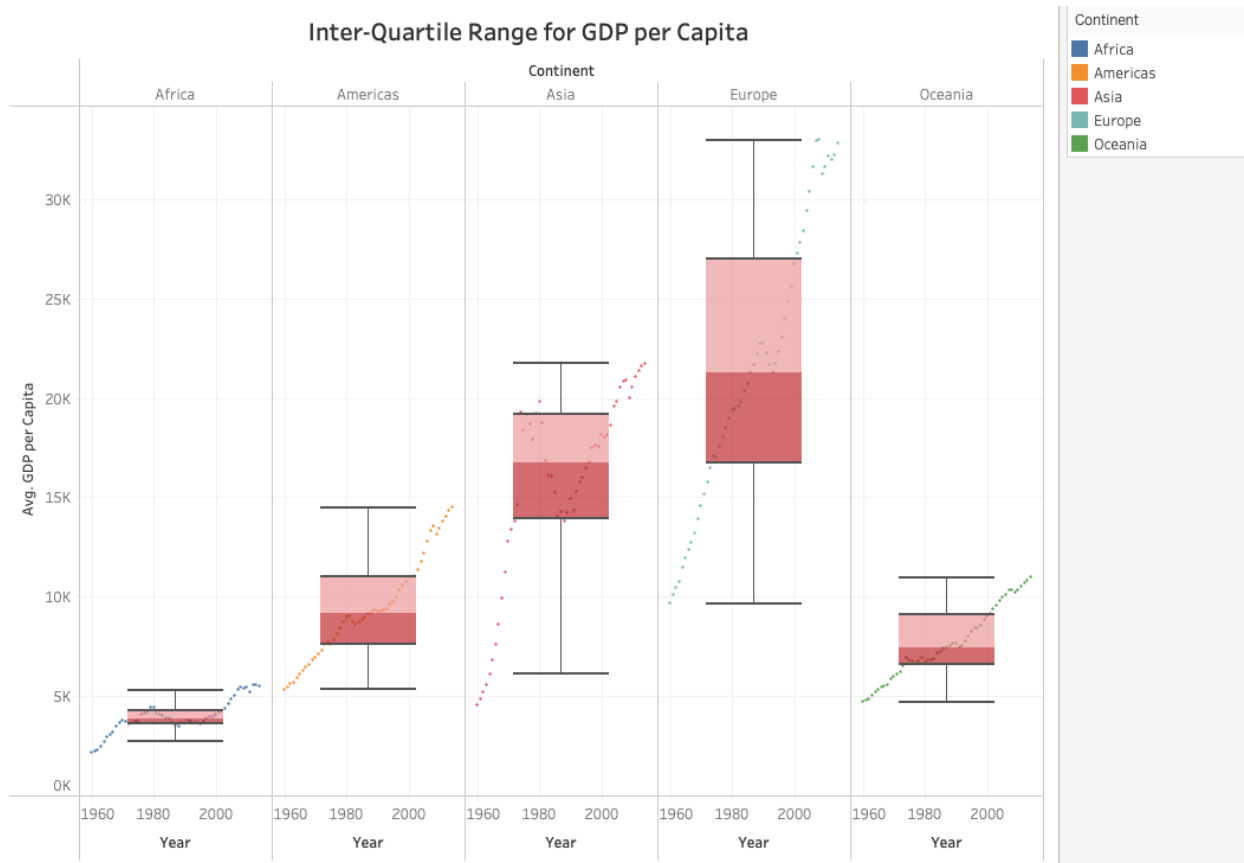
It can be observed from the line chart above that the GDP per capita has increased significantly for Europe over the years 1960 - 2014. The GDP per capita for Asia increased drastically from 1960 to 1980 after which a fall was observed until 1985. GDP started increasing again from 1985 onwards. GDP per capita for Africa was almost constant for the years 1960-2014.

- **Now, let's study the trend for the GDP per Capita for various subregions over the years 1960-2014**



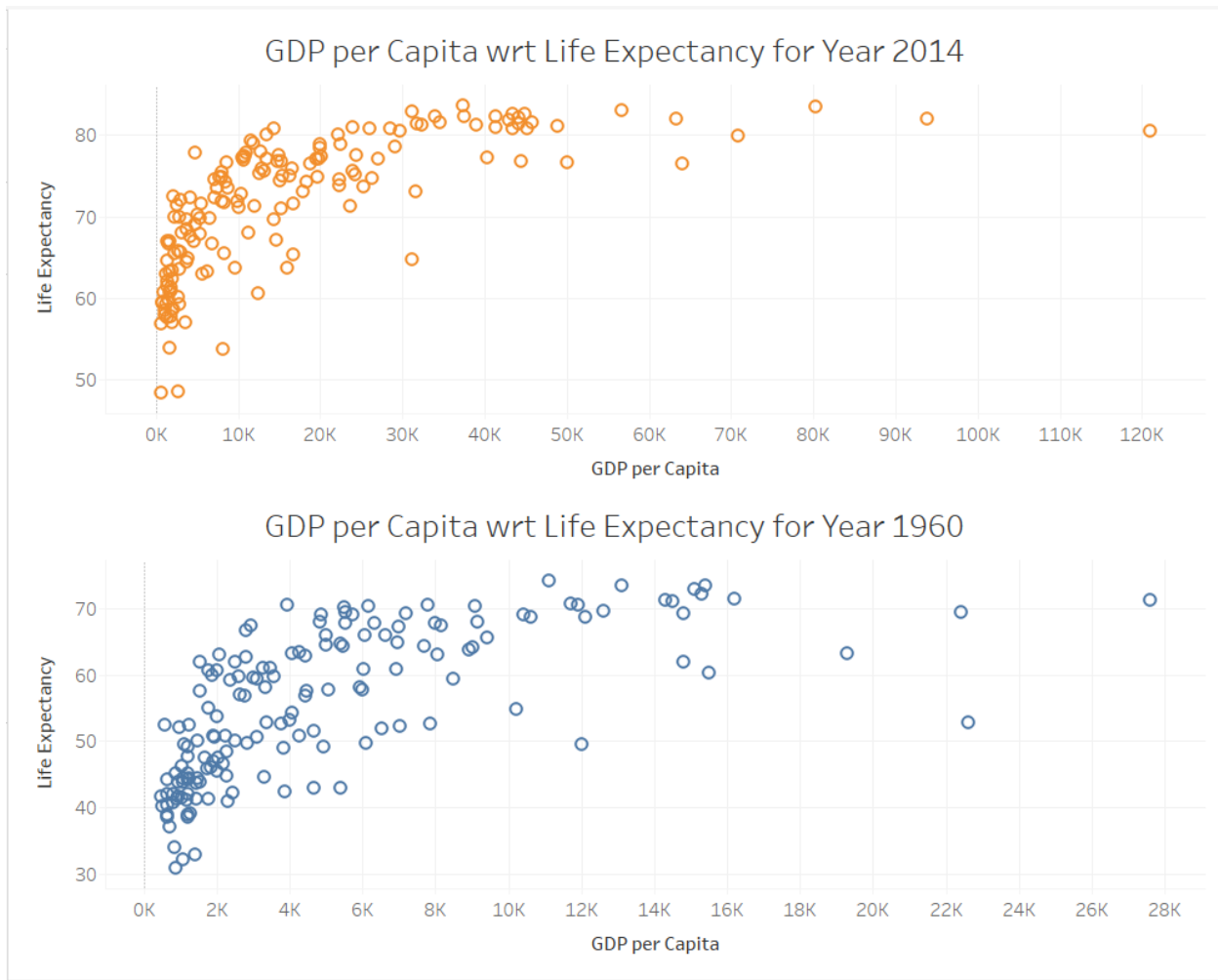
The above line-chart shows that Australia and New Zealand had the highest increase in the GDP per capita compared to other sub-regions across the globe. The GDP per capita decreased from 1980 until 1985 and then it continued to increase until the year 2014.

- **Was there a normal distribution for GDP per capita wrt. five continents?**



Box plot above shows that except for Asia and America, data was approximately symmetric (ie. distributed normally) in all continents. As per the graph, the center of distribution is lowest in Africa as compared to Europe which had the highest mean of around 22k. Also, distribution in Asia is negatively skewed whereas in America data is positively skewed (with Interquartile Range of around 5k).

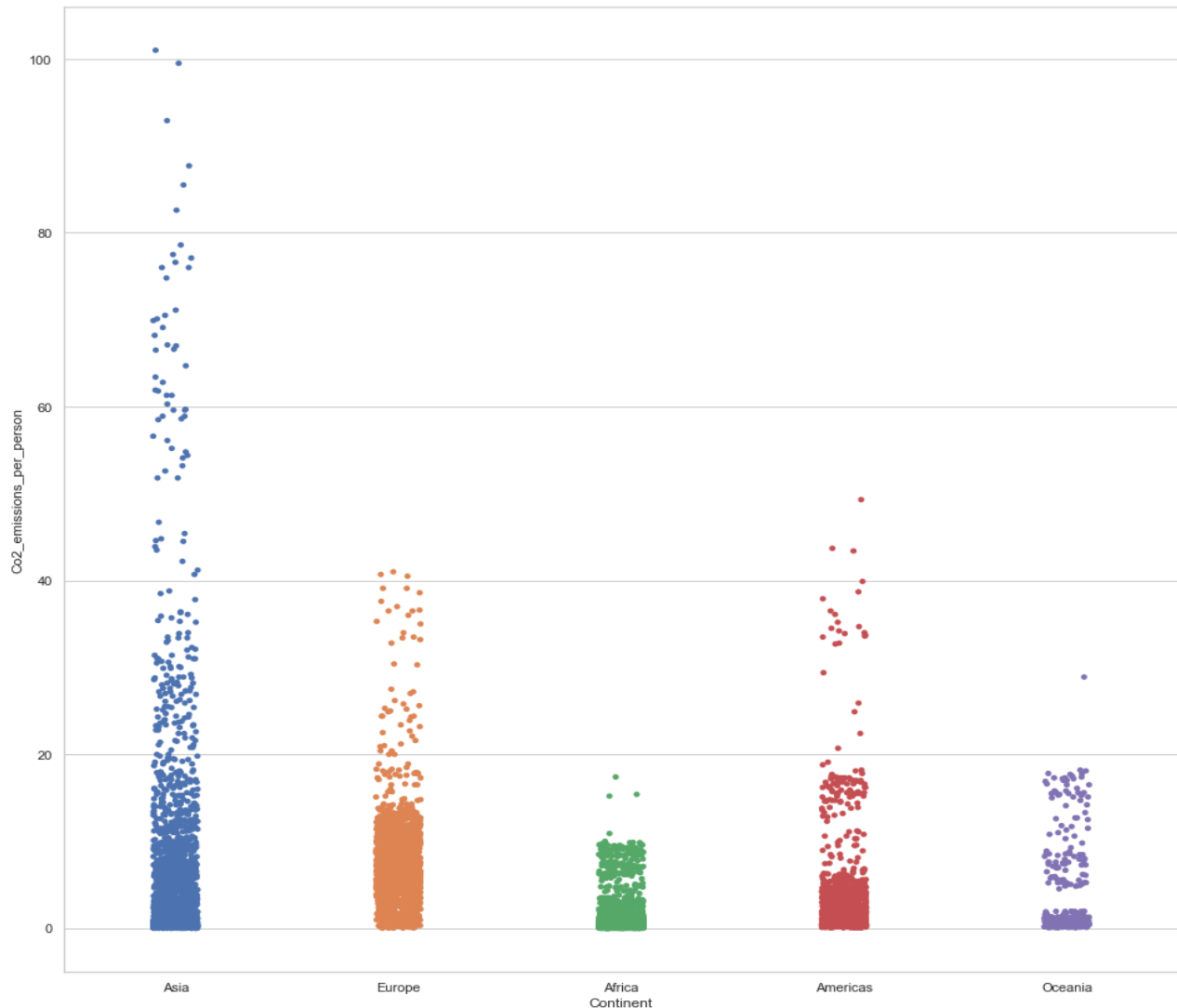
- **Did GDP per capita have any impact on life expectancy?**



With reference to the graph above, it is evident that the GDP per capita and Life expectancy are positively correlated. For the year 1960, GDP per capita and life expectancy were mostly in the range 2k-4k. However, in 2014 the values rose significantly in the range 10k-20k.

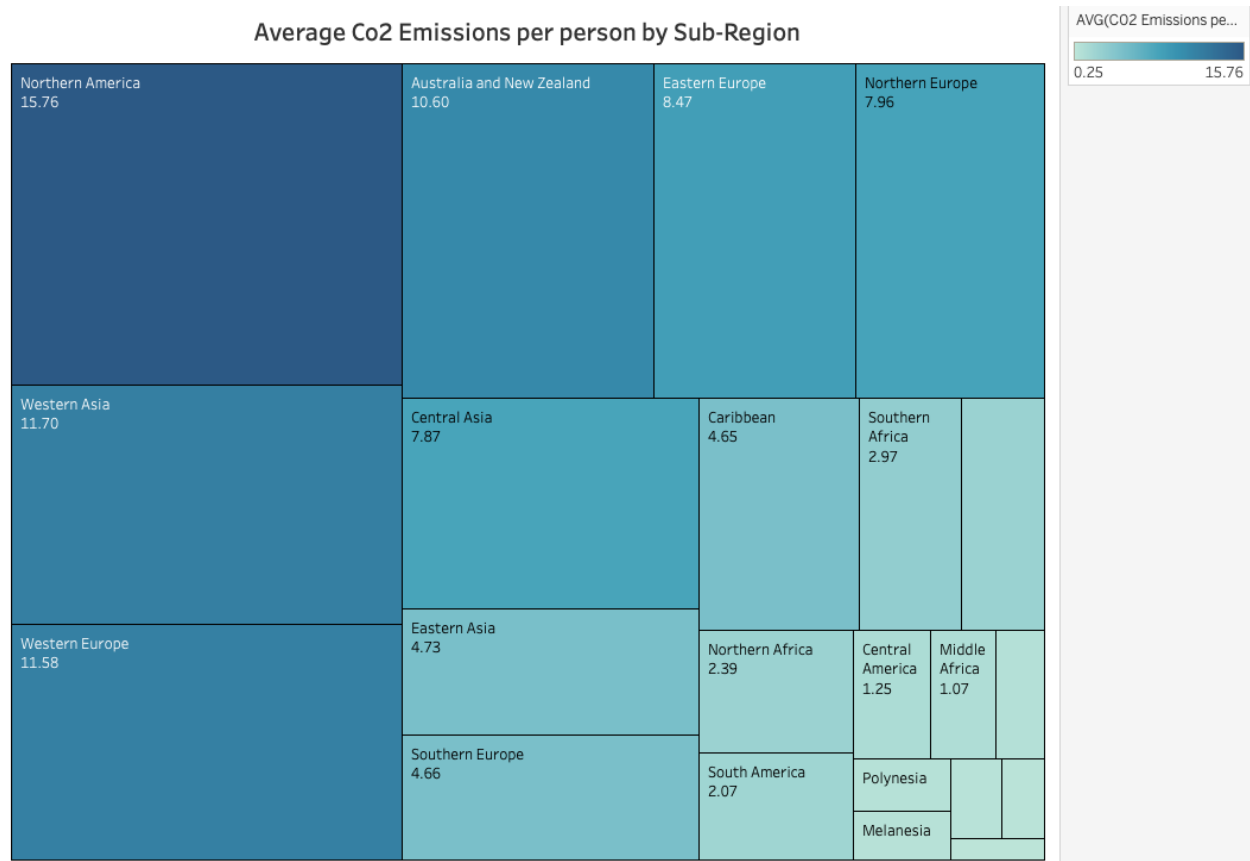
- **How did Development factors like Industrialization affect quality of Life?**

Let us analyze CO2 Emissions per person in different Continents to understand this!



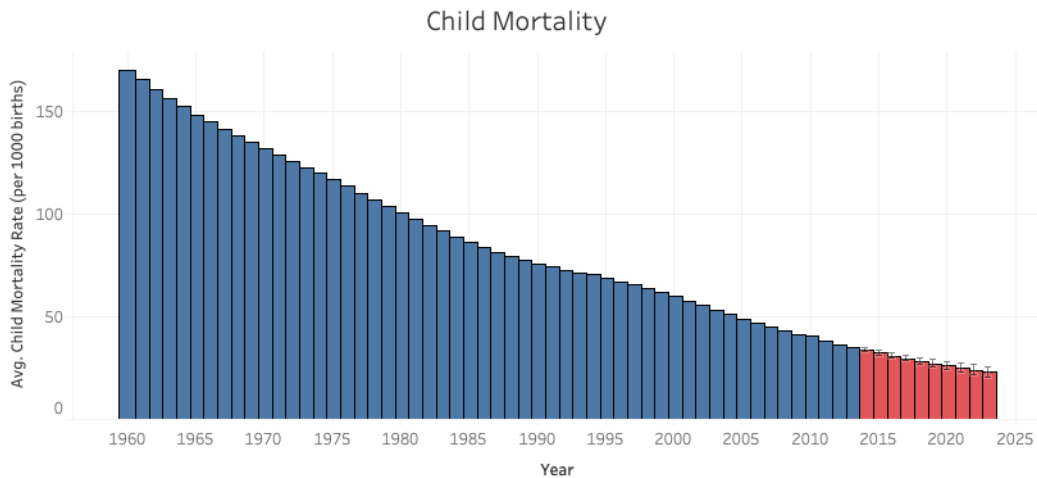
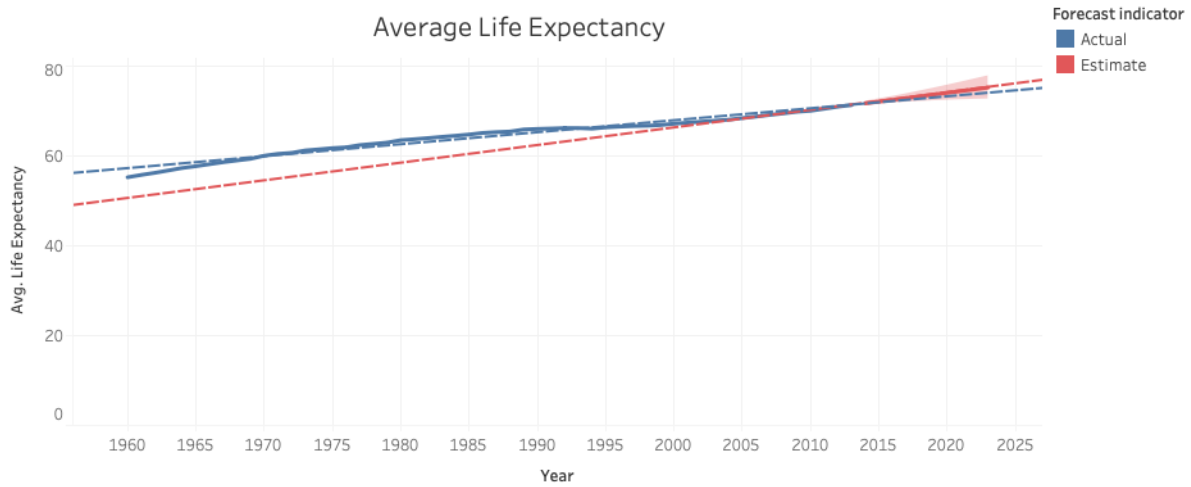
From the Strip plot above, CO2 Emissions per person are most dense and higher in Asian Continent and lowest in Oceania and the African Continent. Population, development & industrial growth plays a major role in this. Developed countries in the European continent had moderate CO2 emission p.p., wherein most populated continent, Asia had maximum CO2 emissions p.p.

- Let us dig deeper for CO2 emissions per person using Heat map



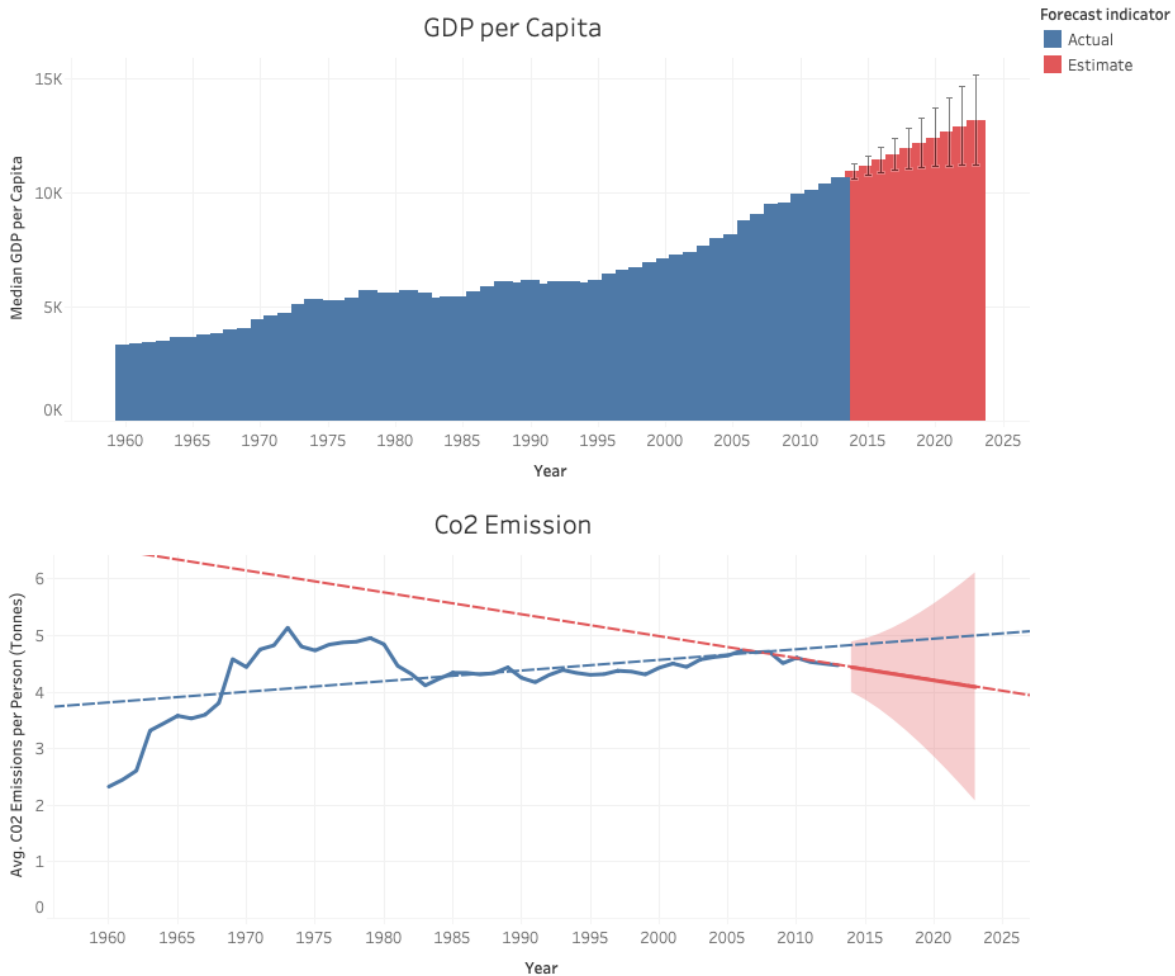
Heatmap is showing CO2 emissions per person per subregion. Northern America has maximum CO2 emission per person compared to all other regions. Regions with more industrial growth over the years had more CO2 emission as a result.

- **How will quality of life be in upcoming years?**



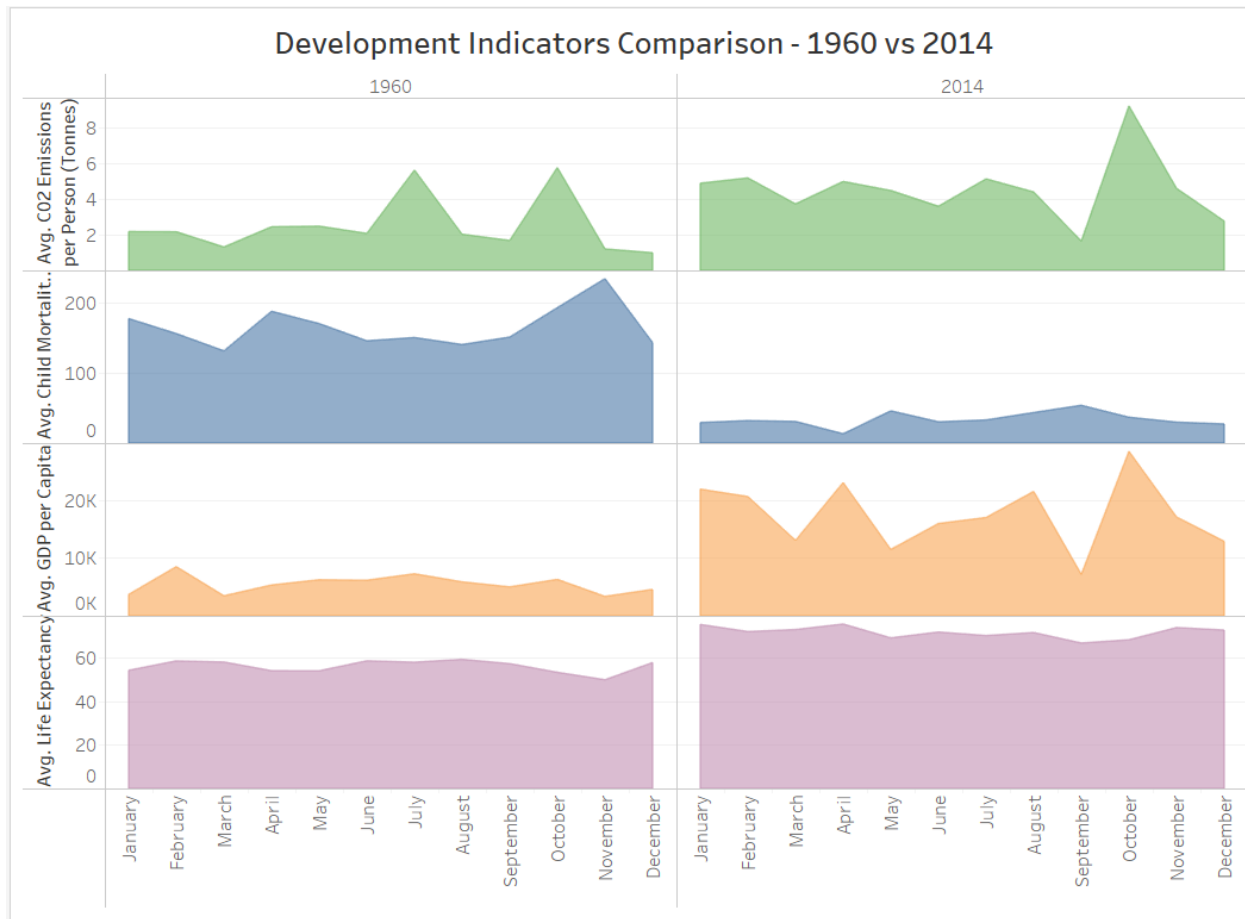
As per the forecast above, life expectancy has an upward trend whereas, child mortality shows downward trend ie. life expectancy will be rising to 70-75 years and child mortality will be falling to 20-25 per 1000 births.

- To study the impact of GDP per capita on CO2 Emission per person in future



As the GDP per capita is rising significantly in all these years, it will adversely affect quality of life due to an increase in the CO2 emission per person. As we can see from the predictive analysis above, till 2025 GDP per capita is likely to increase to 12k. On the other hand, CO2 emission will be in the range 2-6 Tonnes per person. Hence, if we take necessary preventive measures (like more use of renewable sources of energy) the CO2 emission can be as low as 2 Tonnes per person.

Overall comparison of development indicator parameters for the year 1960 versus 2014



- From the comparison above, we can see that the Child mortality rate was higher and Life expectancy was less for the year 1960 which may be due to lack of medical facilities and awareness. However, in 2014, life expectancy has increased and child mortality has significantly reduced which may be because of the development happening over the years.
- As compared to 2014, the GDP per capita was very low in 1960 as the growth in the industrial sector was less. The rapid industrialization over the years led to an increase in the CO2 emissions per person. The average CO2 emissions was around 2 tonnes per person in 1960 which increased to almost 6-7 tonnes per person for the year 2014.

INSIGHTS:

The analysis of development indicators for various regions from 1960 to 2014 gave us some interesting insights which are as follows:

- It can be observed that the developed countries showed higher GDP per capita, therefore lower average child mortality rate and higher life expectancy was observed for them.
- Countries with more industrial growth showed higher amounts of CO2 emissions per person.
- The predictive analysis upto the year 2025 shows a further increase in the CO2 emissions, GDP per capita, life expectancy and a decrease in child mortality rate.