

WORLD HAPPINESS SCORE FROM 2008 AND 2017

14 Nov 2018

MALAK ALMOHAMMADI

Introduction

The World Happiness Report is a landmark survey of the state of global happiness. The World Happiness Report 2018, ranks 156 countries by their happiness levels. It is done by a group of independent experts using the data provided by the yearly Gallup World Poll.

The happiness index is created by many major areas. I Select these majors in my analysis to study the effect of it in the happiness score.

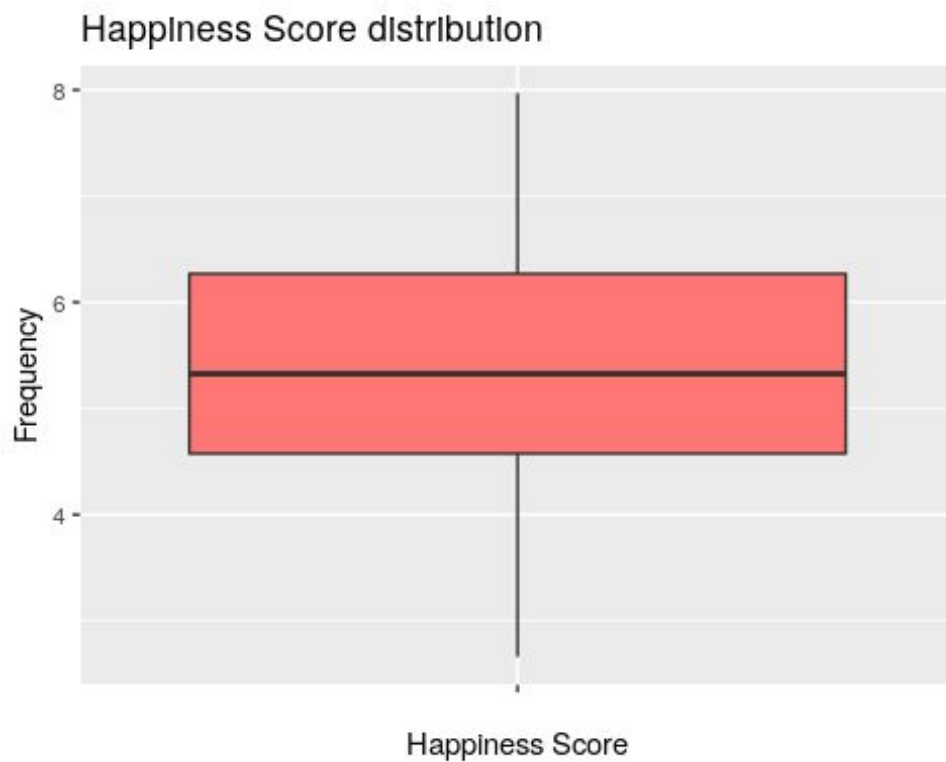
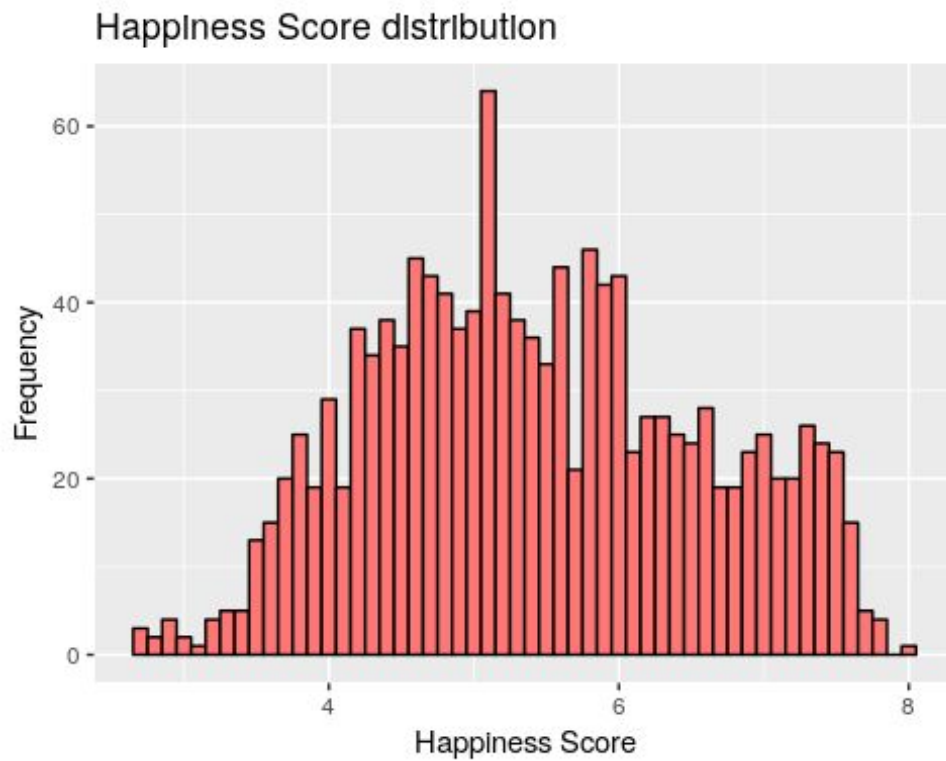
Column Name	Data Type	Description
Country	String	Name of the Country
Score	Float	National AVG response to the questions: "Please imagine a ladder, with the worst possible life as a 0 and the best possible life as a 10. On which step of the ladder do you say you personally feel you stand at this time?". This measure is also known as the Candrill Life Ladder
GDP per Capita	Float	Natural log of GDP per Capita GDP : the total value of all the goods and services produced by a country in a particular year, divided by the number of people living there Source
Social Support	Float	National AVG of the responses to the GWP question: "If you were in trouble, do you have someone to turn to for help? Or friends you can count on to help you out if you need them, or not?"
Healthy Life Expectancy	Float	Life Expectancy AVG based on data by the World Bank and the WDI Life Expectancy : The average number of years that a person can expect to live in "full health" taking into account years lived in less than full health due to disease and/or injury. Source
Positive Affect	Float	AVG of three positive affect measures in the survey: <i>happiness, laugh and enjoyment</i>
Negative Affect	Float	AVG of three negative affect measures in the survey: <i>worry, sadness and anger</i>
Continent	String	Continent of Country

Project Questions :

- Q1) How is the performance of the world happiness score over the years?
- Q2) What are the factors affect the happiness Score?
- Q3) To what extend effect these factors on the happiness Score?
- Q4) How is the performance of happiness score in world continents over the years?
- Q5) How's the relation of the factors and happiness score over the world continents looks like?

Univariate Plots Section

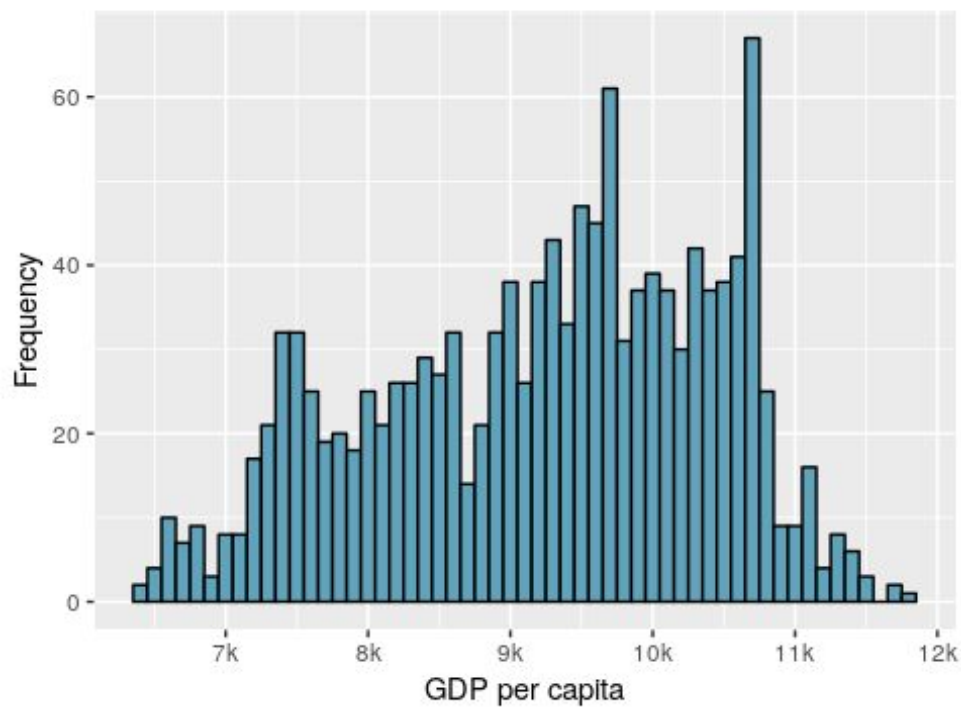
Score	Min.	:2.662	1st Qu.:	4.575	Median :	5.32	Mean:	5.432	3rd Qu.:	6.269	Max.	:7.971
-------	------	--------	----------	-------	----------	------	-------	-------	----------	-------	------	--------



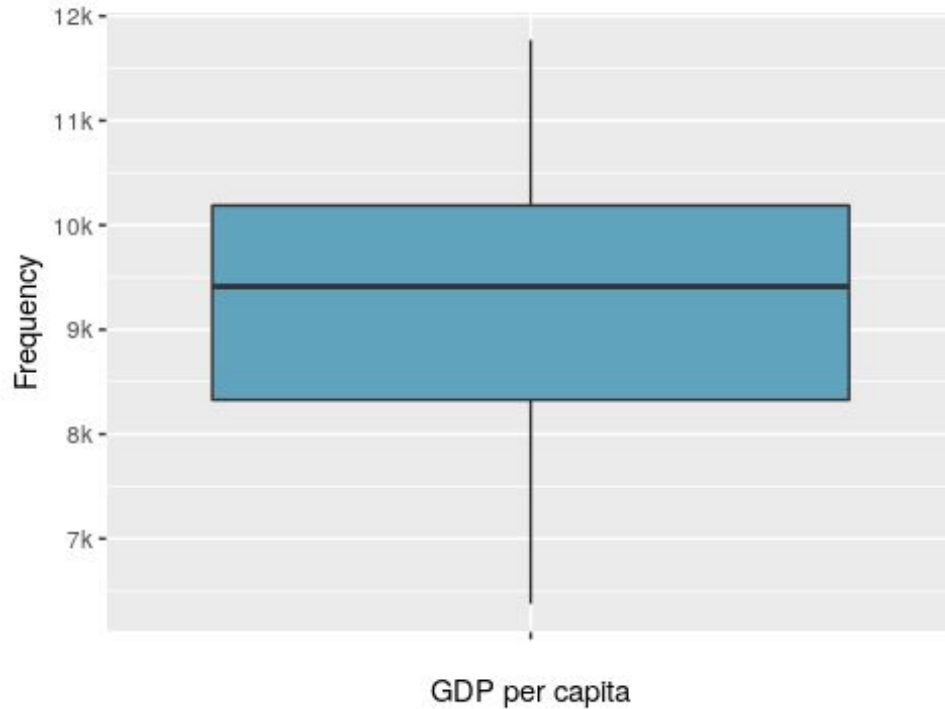
The happiness score is normally distributed, The range of the data is between 2.6 and 7.9, The mean, median and mode are almost equal, Mean = 5.4 and the median = 5.3, Mode = 5. There are no outliers in the Score.

GDP Min.: 6.377 1st Qu.: 8.328 Median : 9.412 Mean: 9.228 3rd Qu.:10.188 Max. :11.770

GDP per capita distribution

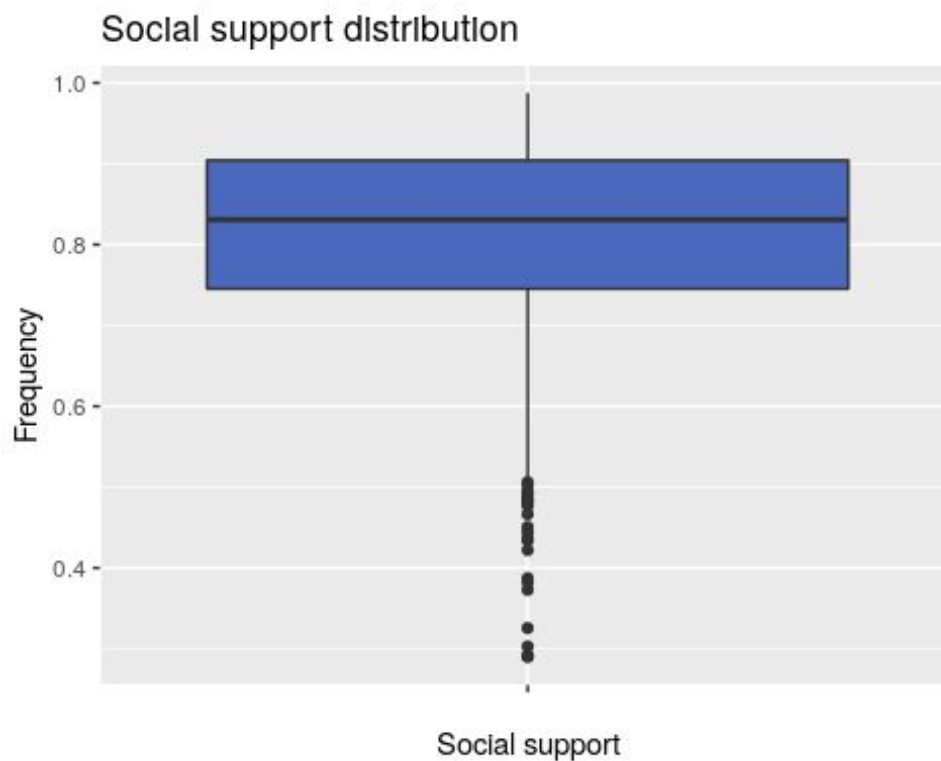
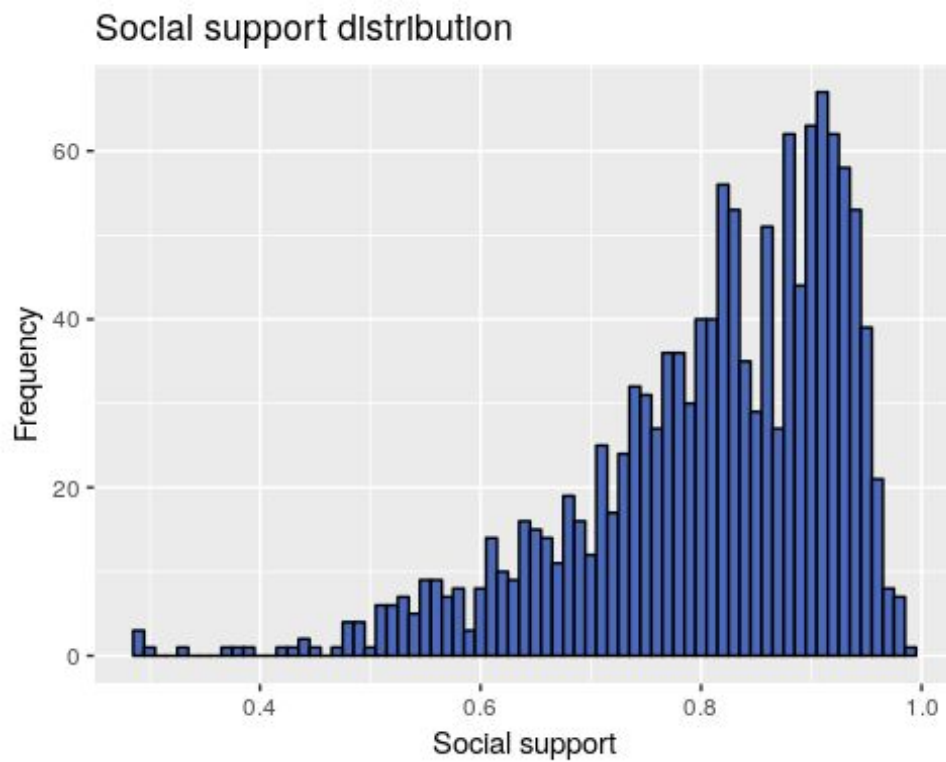


GDP per capita distribution



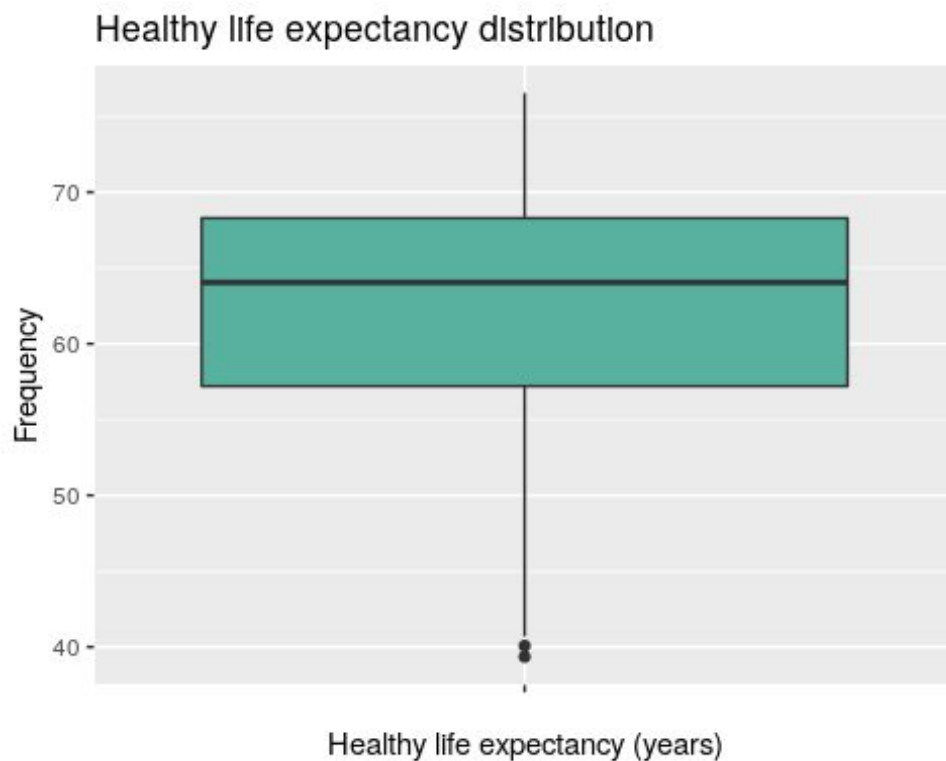
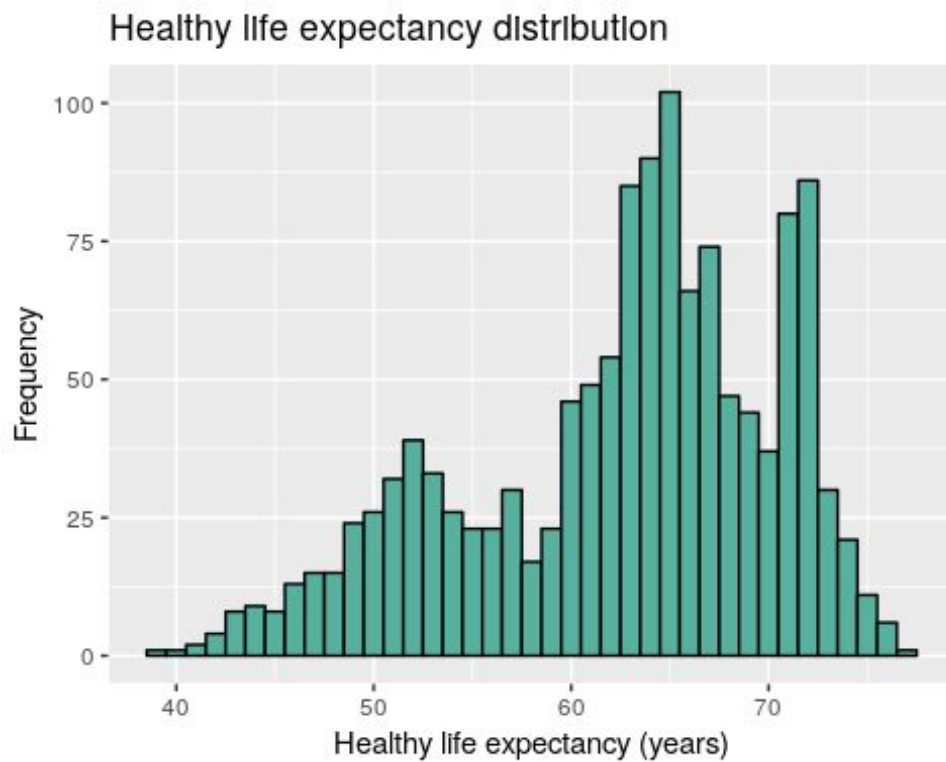
The distribution of GDP per capita is left-skewed, That mean the most of the data is a large number. The range of the data is between 6.3k and 11.8k, the Mean = 9.2k and the median = 9.4k, Mode = 10.8k. There are no outliers in GDP per capita.

Social_s Min.:0.2902 1st Qu.:0.7452 Median :0.8308 Mean:0.8086 3rd Qu.:0.9041 Max.:0.9873



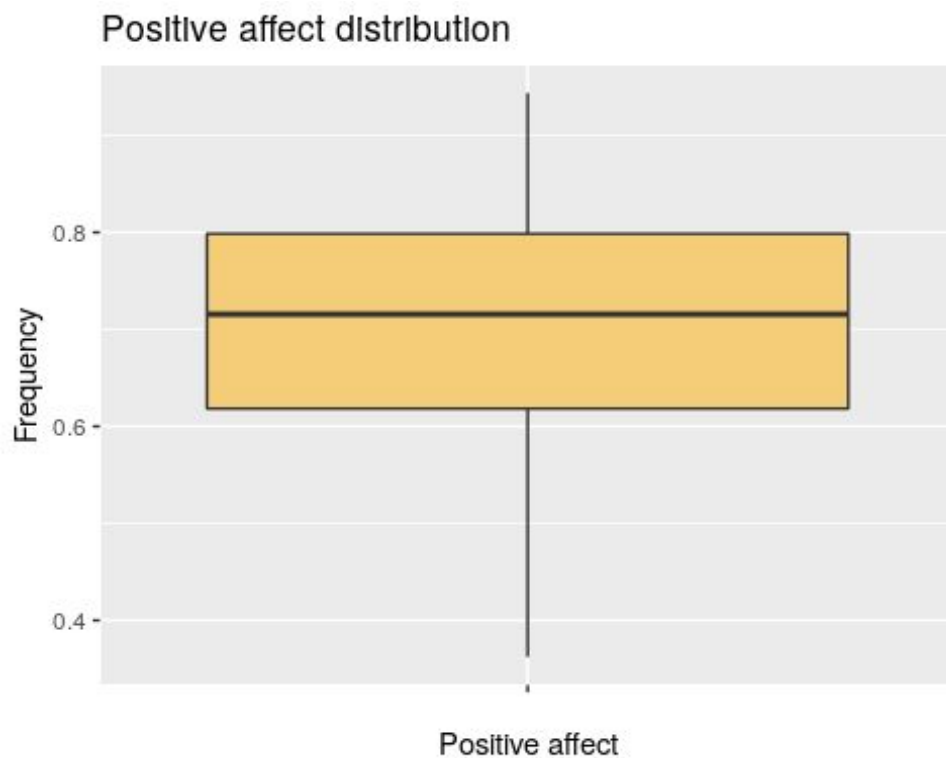
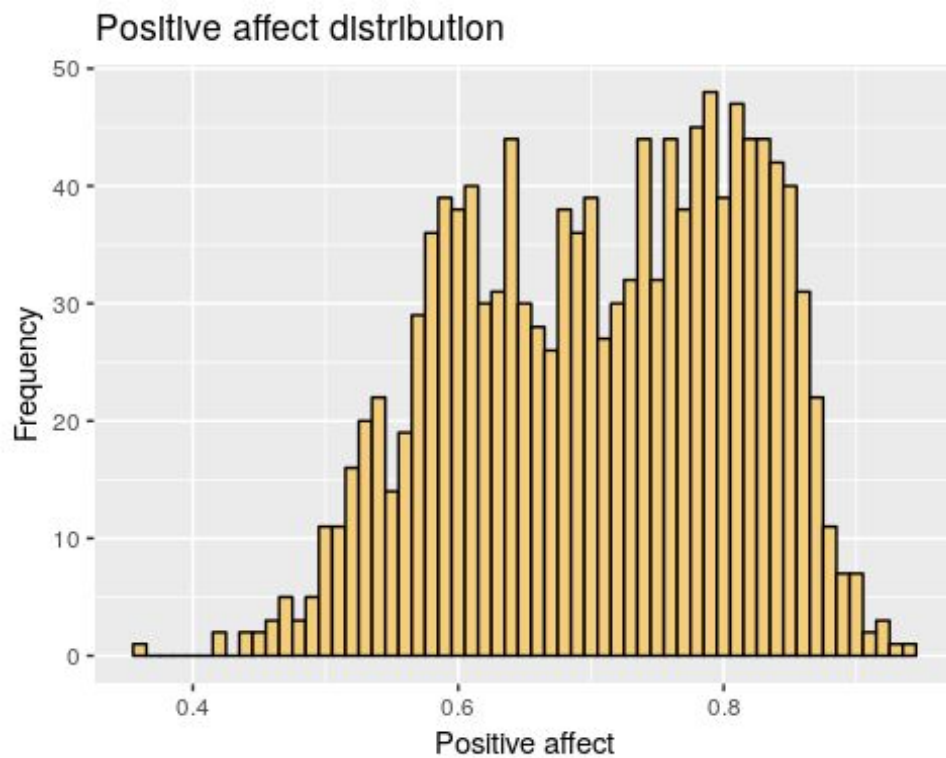
The distribution of Social support is also left-skewed, The range of the data is between 0.2 and 0.9, Most of the data is near 1. The Mean = 0.80 and the median = 0.83, Mode = 10.8. We can see from the box plot there are outliers between 0.3 and 0.5.

Health Min.:39.35 1st Qu.:57.20 Median :64.04 Mean:62.48 3rd Qu.:68.29 Max.:76.54



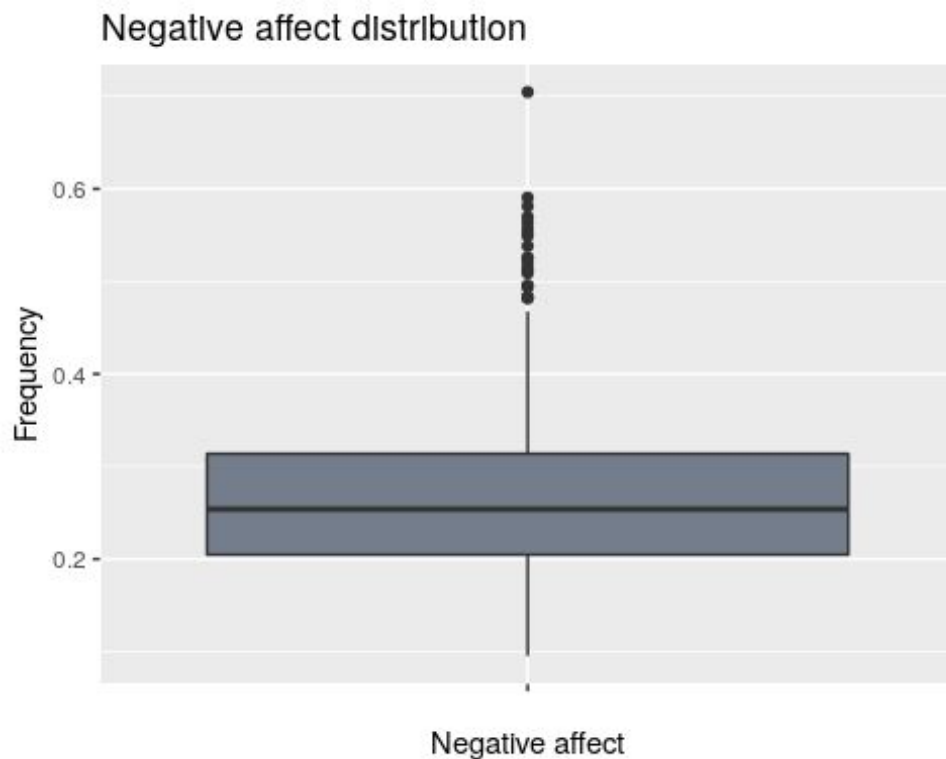
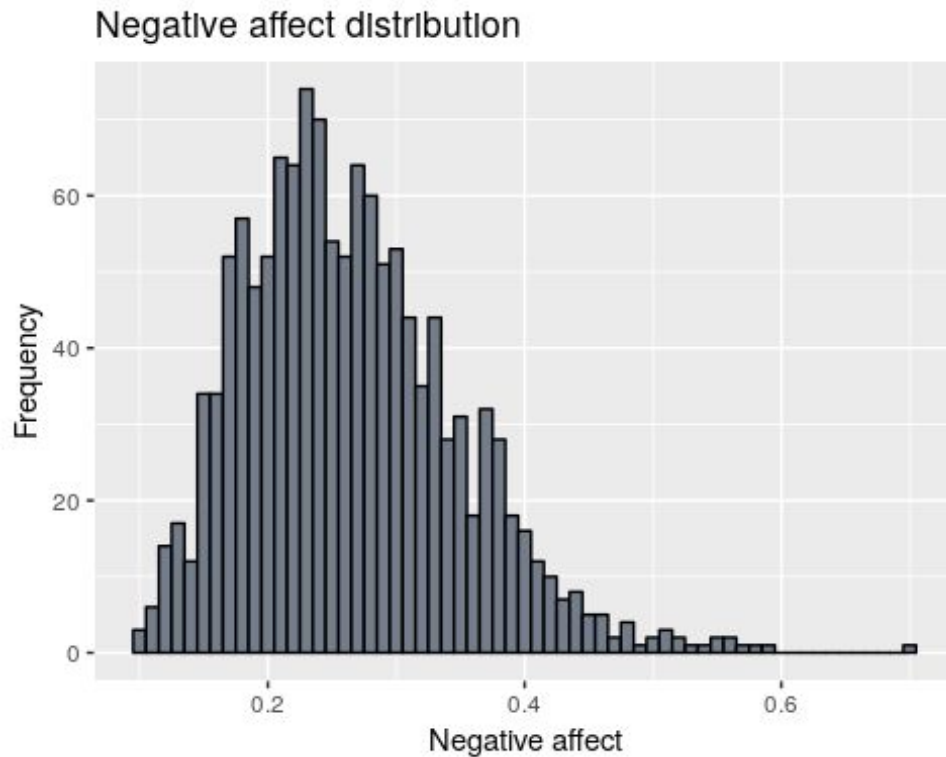
The distribution of Healthy life expectancy is left-skewed, The range of the data is between 39 years and 76 years, Most of the data is between 60 years and 70 years. Which is mean most of the countries citizen live a long time. The Mean = 62 years and the median = 64 years, Mode = 65 years We can see from the box plot there are outliers near 40 years.

Positive_a Min.:0.3625 1st Qu.:0.6183 Median :0.7155 Mean :0.7073 3rd Qu.:0.7986 Max.:0.9436



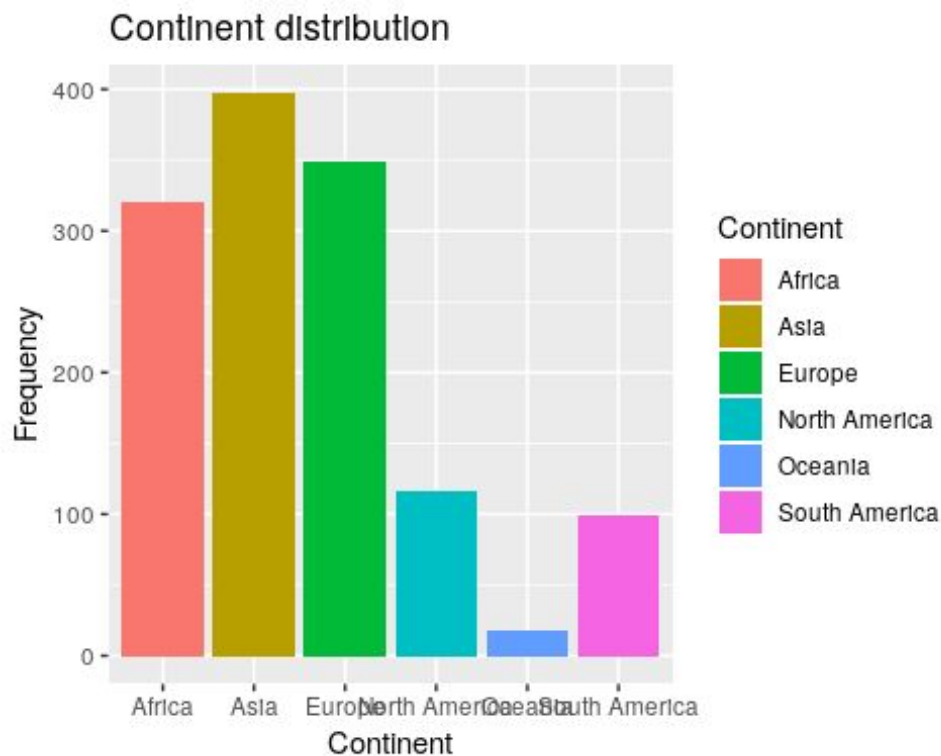
The positive affect distribution is bimodal. This means that there is not a single data value that occurs with the highest frequency, it has 2 modes one at 0.64 and another at 0.85. The mean = 0.70 and the median = 0.71. The range in the data between 0.37 and 0.94. This distribution is surprising me I thought it will be a left skewed. There are no outliers in the positive affect.

Negative_a Min.:0.09549 1st Qu.:0.20466 Median :0.25394 Mean:0.26443 3rd Qu.:0.31382
Max.:0.70459



The negative affect is a right-skewed distribution. This distribution meet my expectation because it makes seance the sadness, worry, and anger, affect in the happiness. The mean is equal to 0.26, the median is equal 0.25, and the mode is equal to 0.23. There are outliers above 0.48.

Continent Africa :320 Asia :398 Europe :349 North America:117 Oceania: 18 South America: 99



From the distribution of the data based on the Continent is highest in Asia by 398 rows this comes from a large number of the country in Asia. Europe comes after by 349 rows, Africa with 320 rows, North America with 117 rows, South America 99 and Oceania with 18 rows only. This rows not depend on the number of the countries in each continent there are multiple rows represent the data of each year.

Univariate Analysis

What is the structure of your dataset?

The dataset contains 1301 observations of 9 features.

What is/are the main feature(s) of interest in your dataset?

The mean feature is happiness score with the value between 0 and 10.

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

I mostly on comparing the factors based on the continents because I want to understand the behavior of happiness around the world.

Did you create any new variables from existing variables in the dataset?

Yes, I created a new column named continent as the categorical variable. I created it from another dataset the match each country with the right continent.

Of the features you investigated, were there any unusual distributions?

Actually, I imagine the distribution before I see the charts. almost all factors except the positive affect meet my expectation. The positive affect has the bimodal distribution.

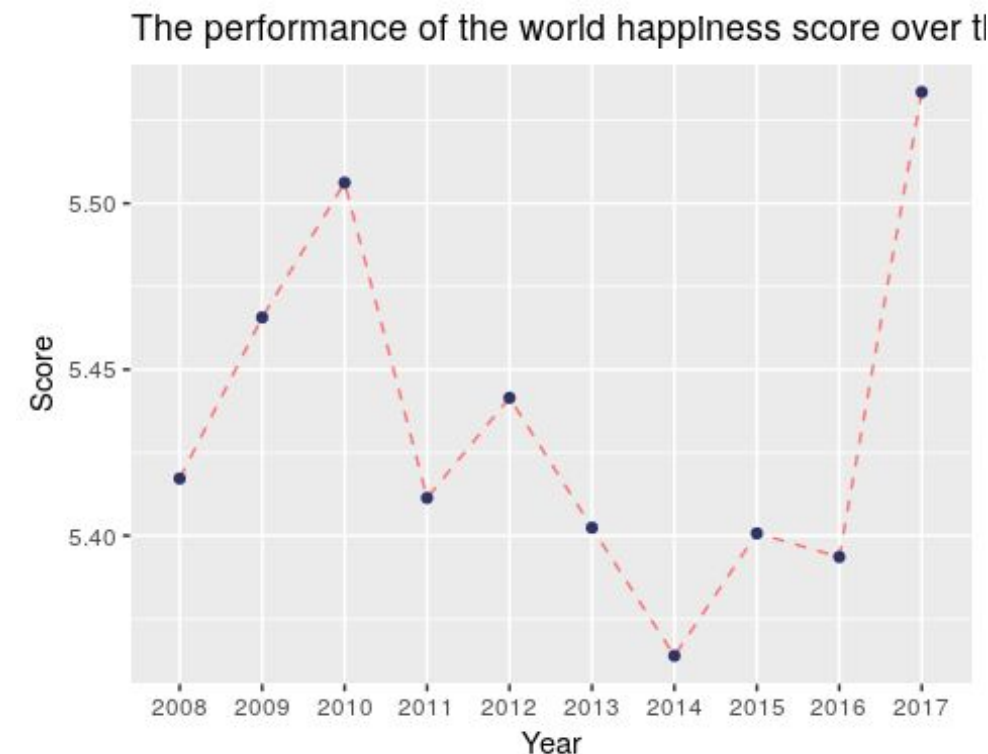
Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

The first step in my project is cleaning the data to be ready to investigate.

Bivariate Plots Section

In this section we will answer this questions : Q1) How is the performance of the world happiness score over the years? Q2) What are the factors affect the happiness Score? Q3) To what extend effect these factors on the happiness Score?

Q1) How is the performance of the world happiness score over the years?



In general the happiness score has highest value at 2017 and lowest value at 2014. From 2008 to 2010 is increased by approximately 0.08 point. From 2010 to 2011 it rapidly dropped by almost 0.05 point. The possible reasons for that is the amount of the data so let us check about it.

```
## # A tibble: 10 x 3
##   Year Score    n
##   <int> <dbl> <int>
## 1  2008  5.42  108
## 2  2009  5.47  112
## 3  2010  5.51  119
## 4  2011  5.41  143
## 5  2012  5.44  139
## 6  2013  5.40  132
## 7  2014  5.36  138
## 8  2015  5.40  138
## 9  2016  5.39  139
## 10 2017  5.53  133
```

From the table, we can see the difference between the number of rows 2010 and 2011 is equal to 24 rows. But this is one possible reason only, we can't judge about this effect till we analyze in more depth.

From 2012 to 2014 it rapidly dropped by almost 0.08 point.

From 2016 to 2017 it rapidly increases by almost 0.14 points.

Q2) What are the factors affect the happiness Score?

Q3) To what extend effect these factors on the happiness Score?

It's interesting when you using helpful packages in R. This packages helps you to reduce the code and give you good results. In this section, I use the correlation matrix to find the correlation coefficient value for all factor. Let's investigate the matrix together.

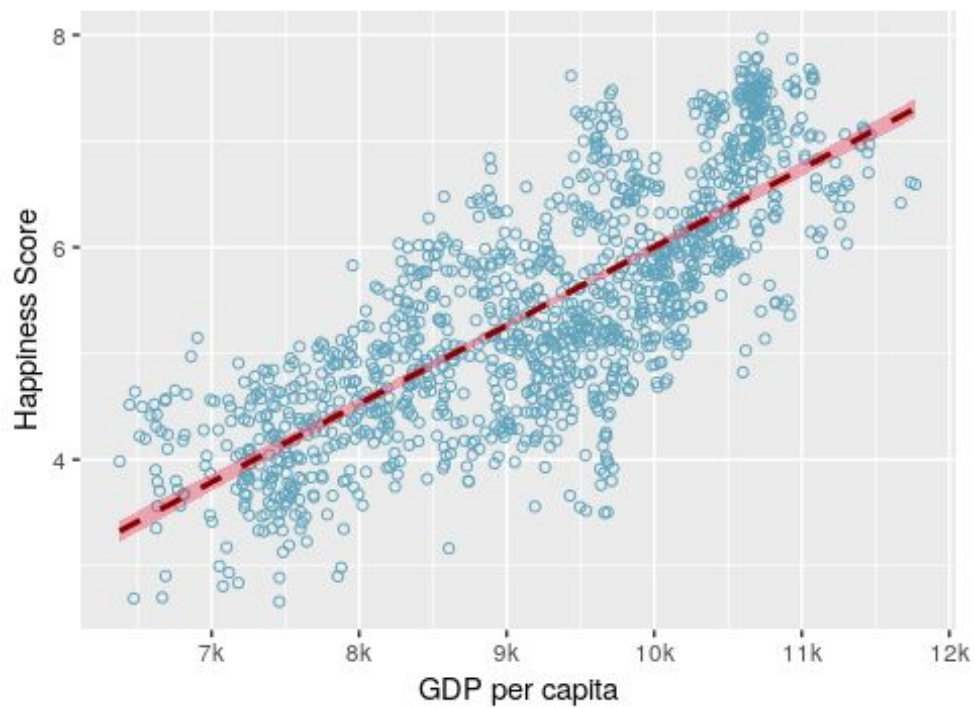


The table below discusses all strength matches based on correlation coefficient values : I used shortcuts to simplify table

-	Score	GDP	Social	Healthy	Positive	Negative
Score	-					
GDP	+Strong	-				
Social_s	+Strong	+Strong	-			
Health	+Strong	+Strong	+Strong	-		
Positive_a	+Strong	+Weak	+Weak	+Weak	-	
Negative_a	-Weak	-Weak	-Weak	-Weak	-Weak	-

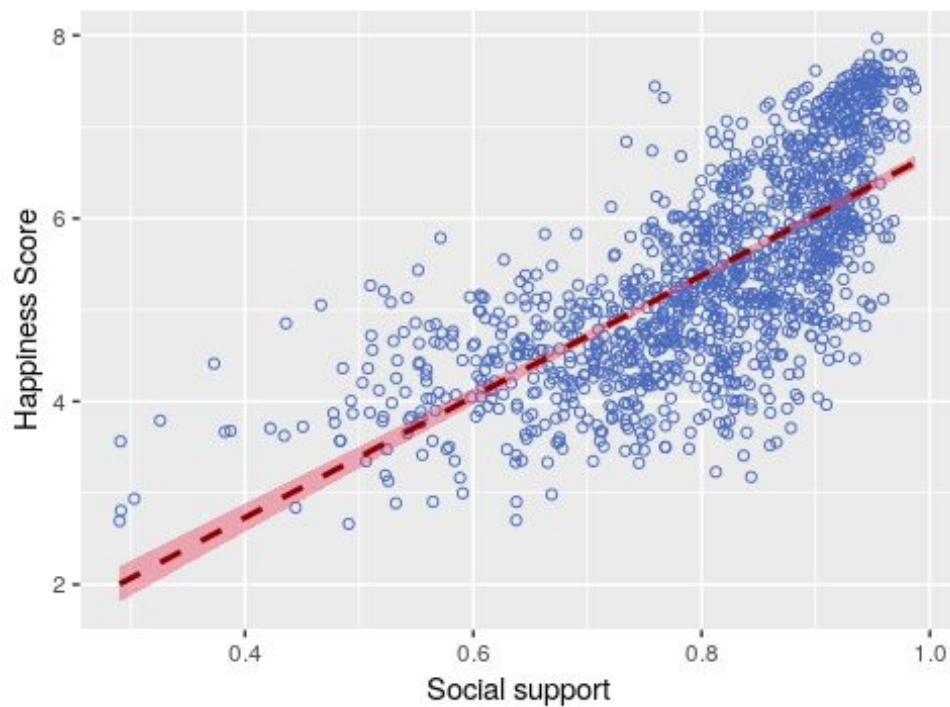
That's great we have a good overview let's see how's the scatter plot of happiness score and other factors look like.

The relationship between GDP per capita and Happiness



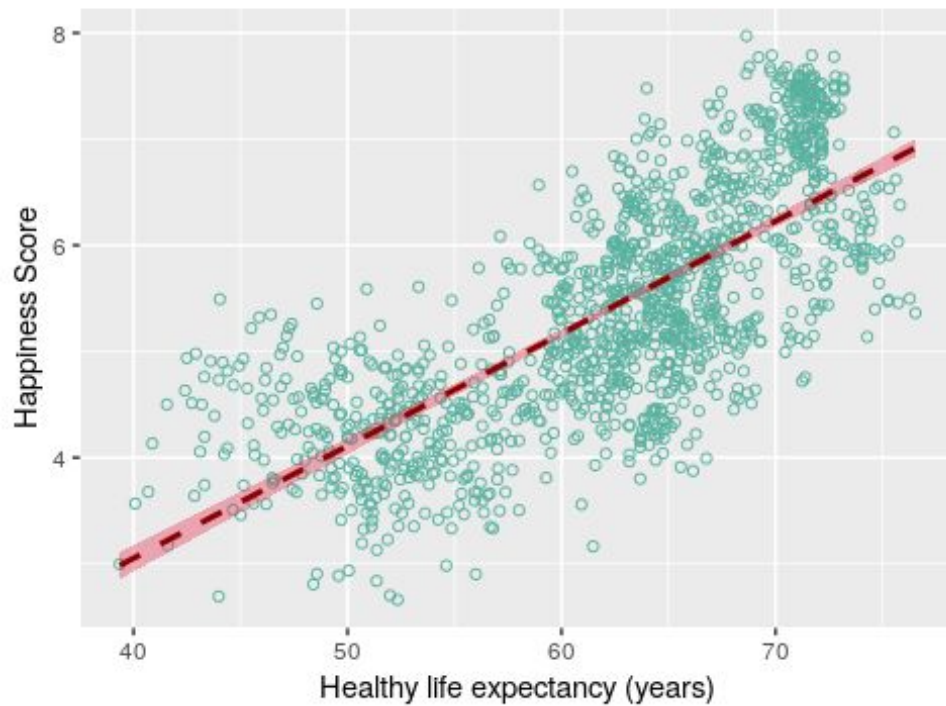
From the chart, we see the relationship between Score and GDP per capita is strong positive.

The relationship between Social support and Happiness



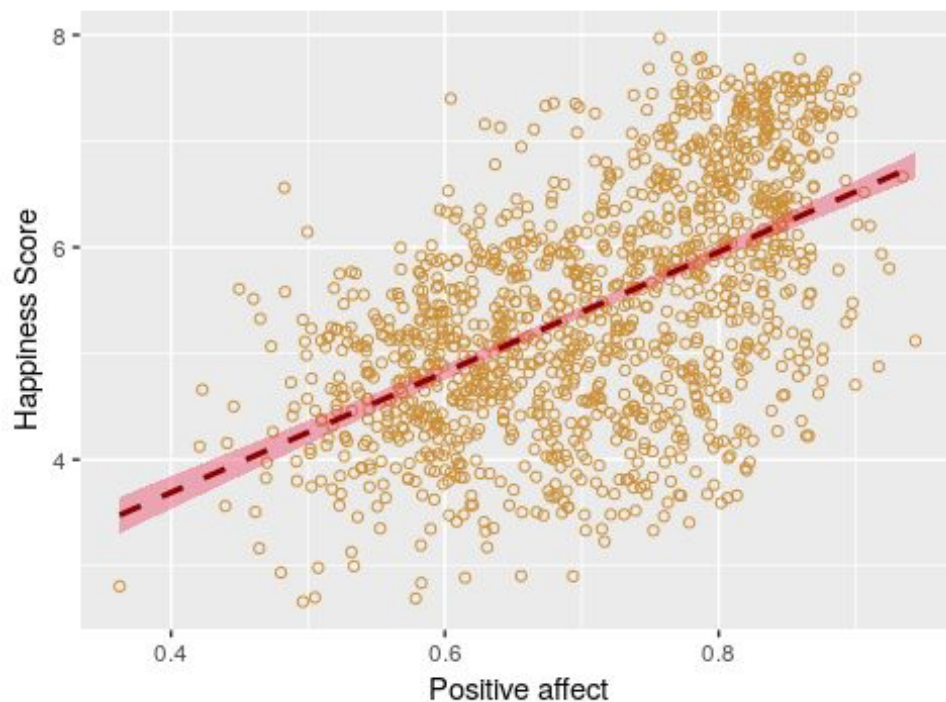
From the chart, we see the relationship between Score and Social support is strong positive.

The relationship between Healthy life expectancy and Happiness Score



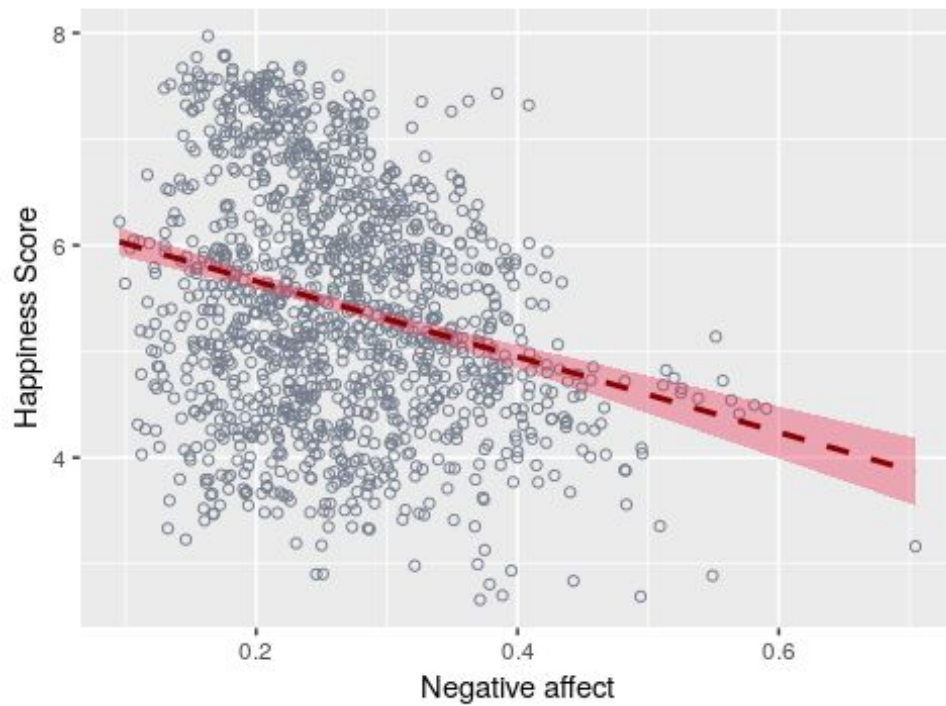
From the chart, we see the relationship between Score and Healthy life expectancy at birth is strong positive.

The relationship between Positive affect and Happiness



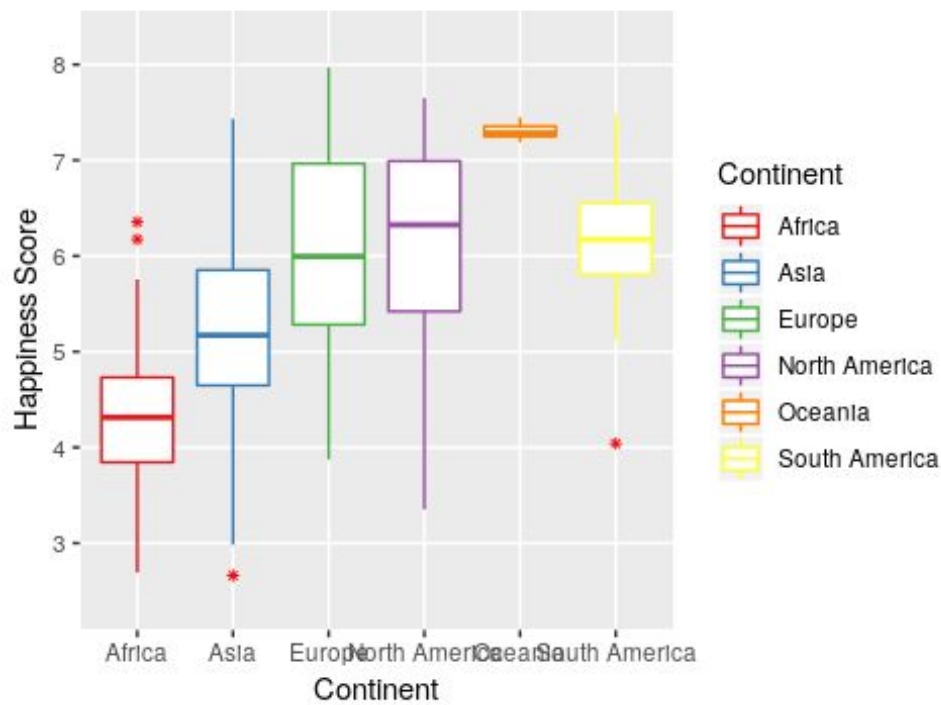
From the chart, we see the relationship between Score and Positive affect is strong positive but less than others.

The relationship between Negative affect and Happiness



From the chart, we see the relationship between Score and Negative affect is weak negative.

The relationship between Continents based on Happiness



I set y range from 0 to 10 to understand hows continents look on the Score ladder.

In general, Europ has the highest score and Africa has the lowest score. Asia has the largest range and Oceania has the lowest range. We should take into consideration the number of countries on each continent.

Africa: The range of score between 2.8 to 5.8, the distribution is almost normal, median equal 4.4, there are outliers above 6.

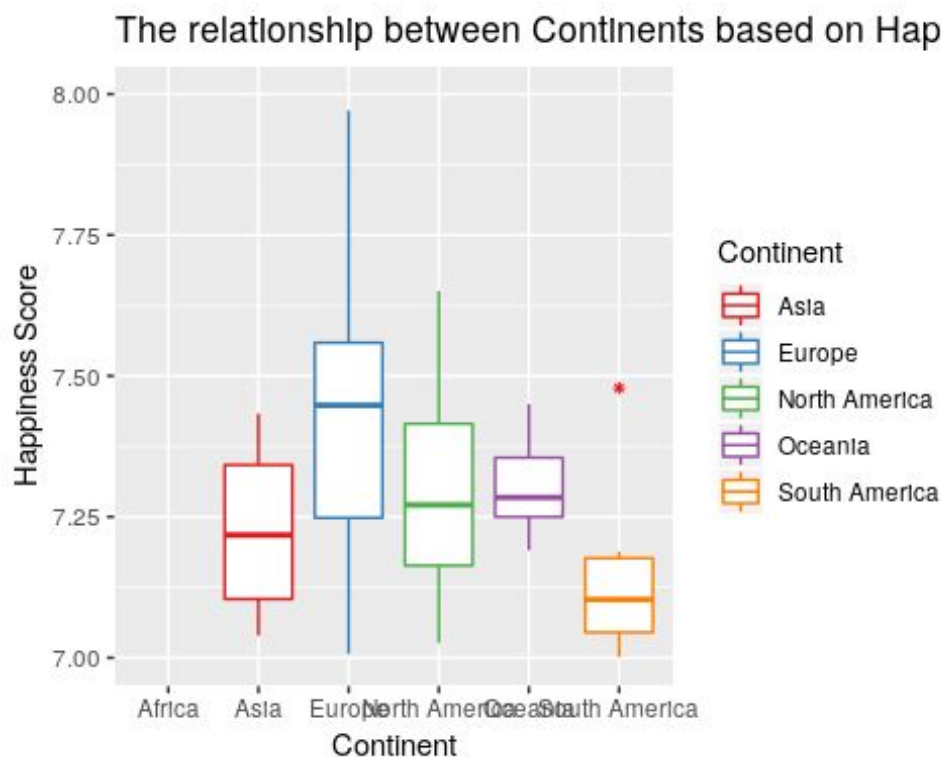
Asia: The score has a wide range between 3 and 7.5, the distribution of the box is right-skewed, median equal 5.1, there are outliers below 3.

Europ: The score has a wide range between 3.9 and 8, the distribution of the box is right-skewed, median equal 6, there are no outliers.

North America: The score has a wide range between 3.5 and 7.8, the distribution of the box is left-skewed, median equal 6.5, there are no outliers.

South America: The score has ranged between 5 and 7.5, the distribution of the box is normal, median equal 6.4, there are outliers around 4.

From this chart, we can't read Oceania so I create a new one zoom in it by decrease y limit



Oceania: The score has ranged between 7.15 and 7.48, the distribution of the box is normal, median equal 7.26, there are no outliers.

Bivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

The happiness score is correlated by other factors shown in the table :

Factor	Correlation coefficient	Relation strength	Relation direction
GDP per capita	0.77	Strong	Positive
Social support	0.7	Strong	Positive
Healthy life expectancy at birth	0.74	Strong	Positive
Positive affect	0.55	Strong	Positive
Negative affect	-0.26	Weak	Negative

Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

The interesting relationships between GDP per capita and Healthy of life expectancy at birth because in the correlation matrix it takes the highest correlation coefficient value with 0.85. GDP per capita is especially useful when comparing one country to another because it shows the relative performance of the countries. A rise in per capita GDP signals growth in the economy and tends to reflect an increase in productivity this increasing reflect on the quality of any service provided by the country to its citizen. The health is one of these important services.

Read more: <https://www.investopedia.com/terms/p/per-capita-gdp.asp>

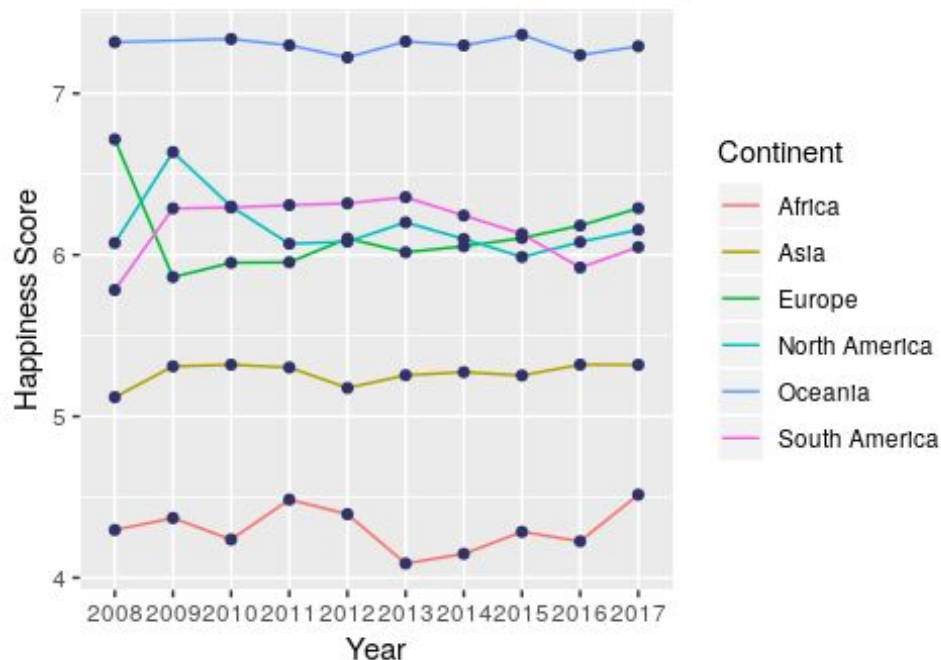
What was the strongest relationship you found?

GDP per capita and Healthy of life expectancy at birth

Multivariate Plots Section

In this section we will answer this questions : Q4) How is the performance of happiness score in world continents over the years? Q5) How's the relation of the factors and happiness score over the world continents looks like? #Q4) How is the performance of happiness score in world continents over the year?

The performance of happiness score
in world continents over the year



Now we separate the general line to many lines represent the Continents. This will help us to understand the years that have drops.

Africa The interesting interval is between 2010 to 2013 because this years have much event based on Arab Spring. The Arab Spring began in late 2010 in response to oppressive regimes and a low standard of living, beginning with protests in Tunisia. The effects of the Tunisian Revolution spread strongly to Libya and Egypt. Sustained street demonstrations took place in Morocco, ,Algeria, and Sudan.

Asia Between 2011 and 2013 Asia also have a problems instead of Arab Spring but the affect here is minor compared with Africa. almost all problems solved in the end of 2011 along with governmental changes in Saudi Arabia ,Bahrain, Jordan, Oman, Kuwait, and Palestinian.

Europe: The interesting interval is between 2008 and 2009 because the happiness score is decreasing by almost 1.2 points. I searched about this during the internet but I didn't find a specific event happened in this year. Maybe the data is not enough or something happened before and the affect has appeared later this depends on 2007 data.

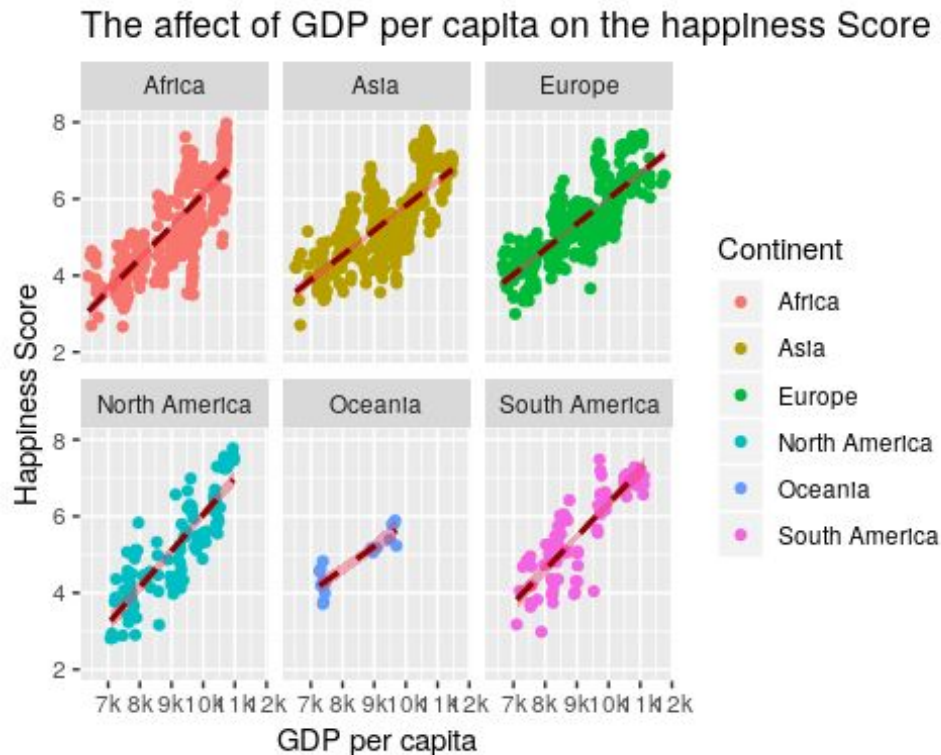
North America: At the blue line, we see the major effect comes from 2009 to 2011. When I searched about something interesting to understand the possible causes occurred there I find something but it's lead us to investigate more about it (did the financial crisis of 2007–2008 affect on the world happiness score as a long-term effect?) maybe a consider it as a question in future work.

South America: The score is increased from 2008 and 2009 by 0.8, then it stays to 2013 then drop again in 2016. I didn't find any interesting information but we should remember based on the missing data in the original dataset that we removed it may be they affect.

Oceania: The line here has tiny changes. This is the smallest continent based on the area and the number of countries. The mean of the score is absolutely affected by this number.

Q5) How's the relation of the factors and happiness score over the world continents looks like?

Here I will study 3 factors : 1) GDP per capita 2) Social support 3) Healthy life expectancy



In General, the GDP per capita has a strong positive relation with happiness score in world continents.

Africa : The GDP per capita range is between 6k to 11k, There is an outlier in 6.2.

Asia : The GDP per capita range is between 6.5k to 11.5k, There is no outliers.

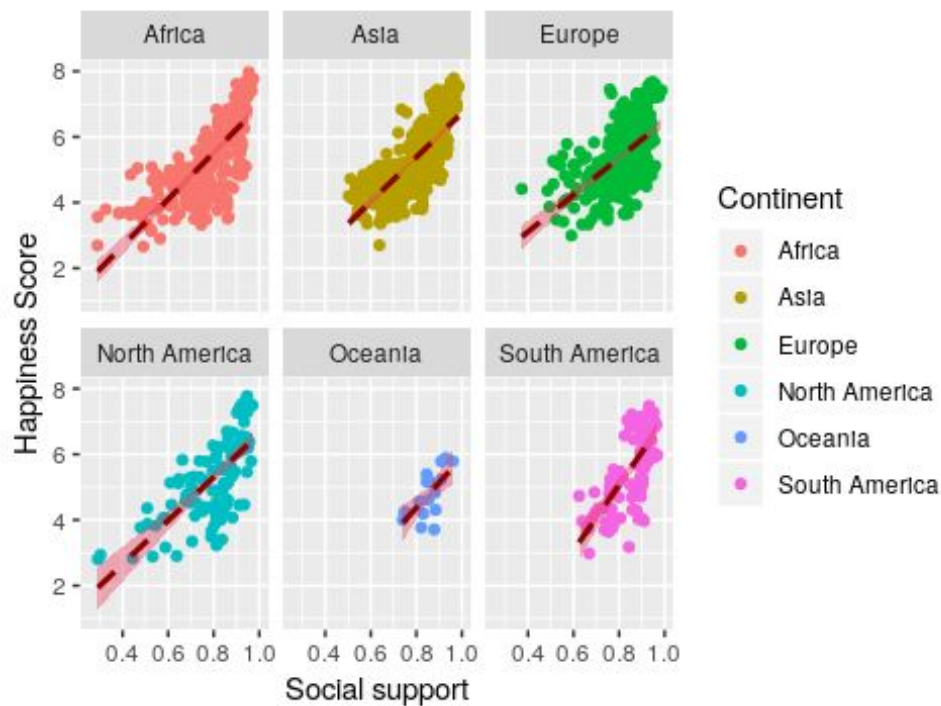
Europe : The GDP per capita range is between 6.5k to 12k, There is no outliers.

North America : The GDP per capita range is between 7k to 11k, There is no outliers

Oceania: I expected this chart. the range is tiny but has a high value between 7k and 9.8k, There is no outliers.

South America: The GDP per capita range is between 7k to 11.2k, There is no outliers.

The affect of Social support on the happiness Score In v



In General, the Social support has a strong positive relation with happiness score in world continents.

Africa : The Social support range is between 0.4 to 1, There is an outlier at 0.3.

Asia : The Social support range is between 0.45 to 1, There is no outliers.

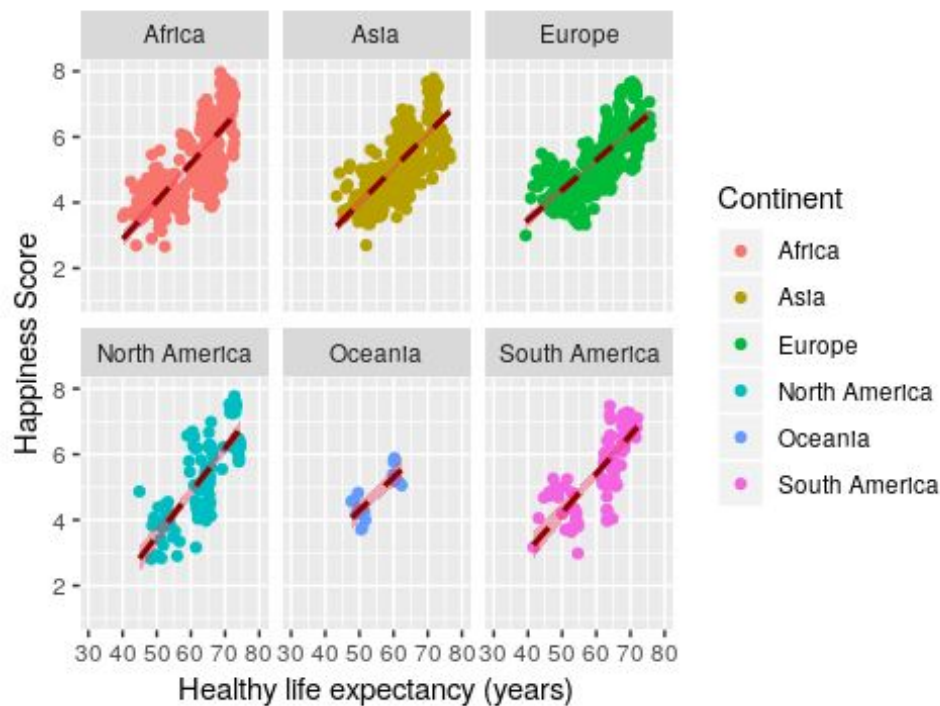
Europe : The Social support range is between 0.5 to 1, There is an outlier at 0.39.

North America : The Social support range is between 0.43 to 1, There is an outlier at 0.3.

Oceania: The range is tiny but has a high value between 0.75 and 1, There is no outliers.

South America: The Social support range is between 0.6 to 1, There is no outliers.

The affect of Healthy life expectancy (years) on the hap



In General, the Healthy life expectancy has a strong positive relation with happiness score in world continents.

Africa : The African people lives about 40 to 75 years , There is an outlier at 0.3.

Asia : The Asian people lives about 40 to 77 years, There is no outliers.

Europe : The European people lives about 40 to 77 years, There is no outliers.

North America : The North America people lives about 59 to 75 years, There is an outlier at 45 years.

Oceania: The Oceania people lives about 45 to 65 years. There is no outliers.

South America: The South America people lives about 40 to 70 years, There is no outliers.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

First, I looked at the trend of the happiness score in the world continents over the years. I found some interesting interval in Africa and Asia between 2010 to 2013. Then, I decided to choose GDP per capita because it has the strongest correlation with happiness score compared with other factors. So I want to understand the distribution of this relation in each continent.

Were there any interesting or surprising interactions between features?

One of the interesting finding is to know how much the people live in each continent.

Final Plots and Summary

Plot One

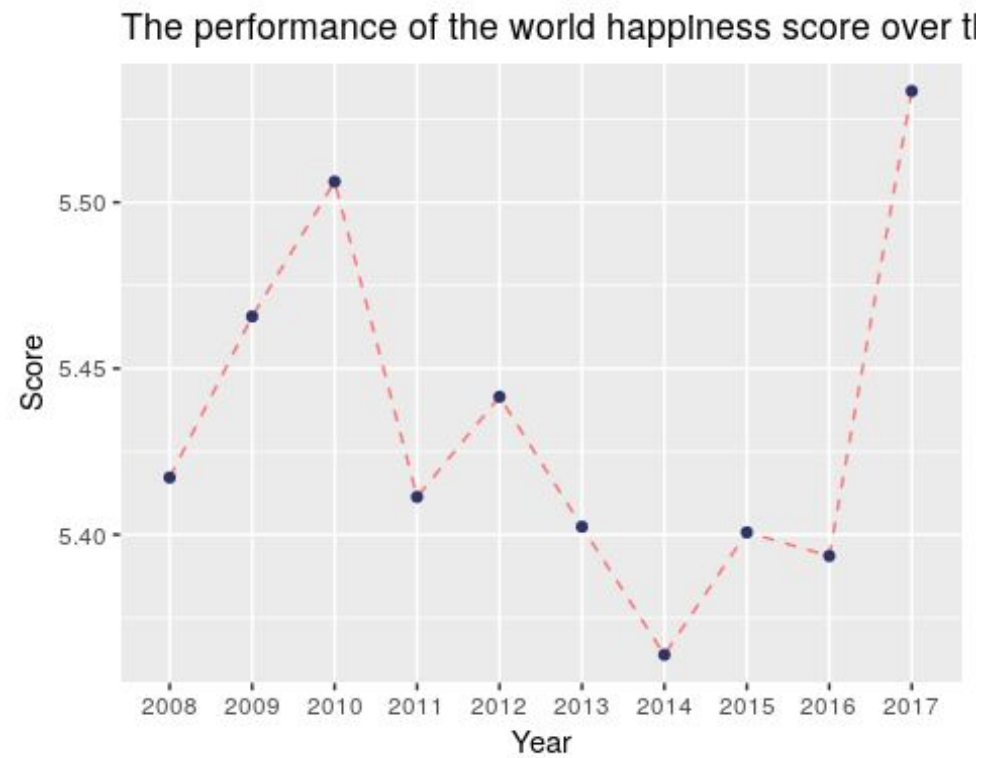


Description One

I chose this chart because it gives us a full coverage of the relations among over data. This function covers the strength and the directions of this relationships.

In this chart, the strength and the direction of all relationships factors in our dataset. The matrix represents the strongest positive relation between GDP per capita and health of Healthy life expectancy at birth. The weakness positive relation between Positive affect with GDP per capita and health of Healthy life expectancy at birth. The strongest negative relationship between Negative affect from one side and Social support and Positive affect from another side. The weakness negative relation between Negative affect and Healthy life expectancy at birth.

Plot Two

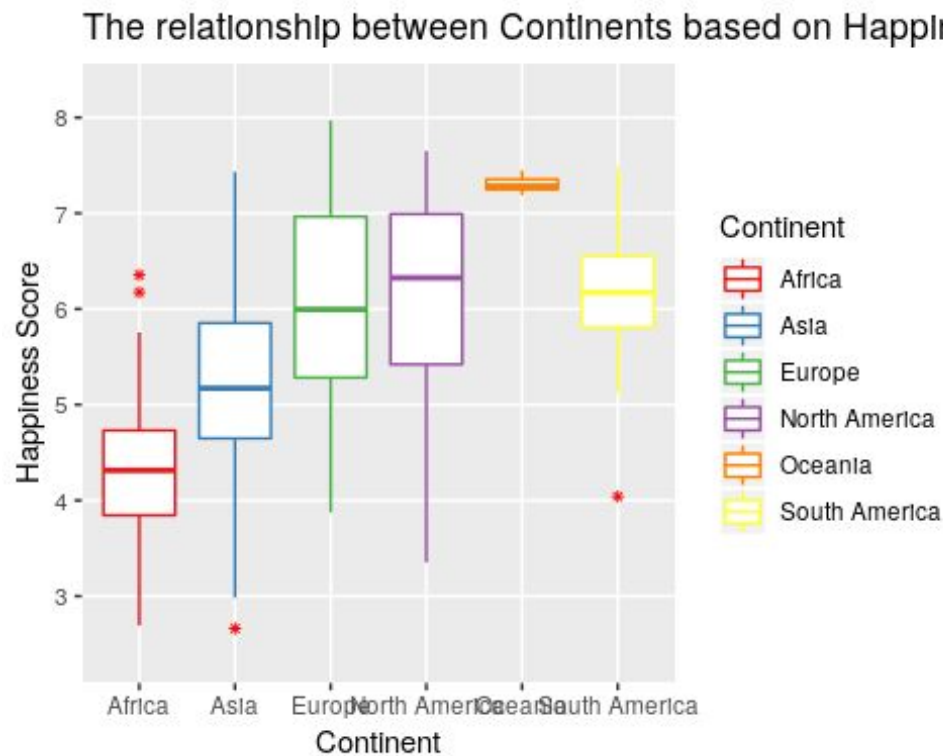


Description Two

I chose this chart because it's a good starting point in your analysis to investigate the main factor over the years.

In general, the happiness score has the highest value at 2017 and lowest value at 2014. From 2008 to 2010 is increased by approximately 0.08 point. From 2010 to 2011 it rapidly dropped by almost 0.05 point. The possible reasons for that are the amount of the data so let us check about it.

Plot Three



Description Three

The continent is the column that I created in my dataset. I created it because I want to study happiness over the world. I thought it's better when I grouping my data instead of study the full list of countries. So I choose this chart because it describes the distribution of the main factor in each group.

In general, Europe has the highest score and Africa has the lowest score. Asia has the largest range and Oceania has the lowest range. We should take into consideration the number of countries on each continent.

Reflection

From this exploratory analysis, we observed the relationships between all factors and our main factor happiness score. There is a strong relation between happiness score and all the factors except Negative affects.

In this project I learned a lot of things which gives me more confidence that I'm in the right way of learning data analysis.

I think the best way in the future work is to think about personal factors. For example, what is the difference between male and female? does the age reflect on happiness? what about the education level? or the amount of reserve money for each person?

Reference

<http://www.sthda.com/english/wiki/ggcorrplot-visualization-of-a-correlation-matrix-using-ggplot2>

<http://www.sthda.com/english/wiki/ggplot2-axis-ticks-a-guide-to-customize-tick-marks-and-labels>

<http://www.sthda.com/english/articles/32-r-graphics-essentials/128-plot-time-series-data-using-ggplot/> <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>

Helpful tool

Online RStudio <https://labs.cognitiveclass.ai/>

Markdown table generator https://www.tablesgenerator.com/markdown_tables#

Cover page design <https://www.canva.com/>