

Computer Project #1

Healthy Life Expectancy



NAME: Tomoe Uchiyama

Professor Alireza Akbari

Course # FOUNDATIONS OF BUSINESS STATISTICS

MSM-503-01/MBA-503-01/IT-502-01 - 24/FA/A

Date: 9/29/2024

CONTENTS

1	Source and Description of Data	3
2	Statement of the Research Objective	3
3	Statistical Analysis	3
3.1	Numerical Methods	3
3.1.1	Numerical Methods-Traditional Method (Excel).....	3
3.1.2	Numerical Methods-Using AI Method.....	6
3.1.3	Numerical Methods-Comparison of Traditional Method (Excel)VS AI Approach..	8
3.2	Graphical Methods	9
3.2.1	The Bar Graph	9
3.2.2	The Pie Chart	9
3.2.3	The Stem and Leaf Display	9
3.2.4	The Box-Whisker Plot	9
3.2.5	The Ogive.....	9
3.2.6	The Area Chart.....	10
3.3	Tabular Methods	10
3.3.1	The Frequency Distribution Table.....	10
3.3.2	The Relative Frequency Distribution Table	10
3.3.3	The Cumulative Frequency Distribution Table	10
3.3.4	The Relative Cumulative Frequency Distribution Table	11
4	Discussion of the Possible Outliers.....	11
5	Conclusion, Summary, and Policy Recommendations.....	11
6	-Appendix-	11

1 SOURCE AND DESCRIPTION OF DATA

This report uses “The World Happiness Report” to study well-being. “The World Happiness Report examines factors that contribute to human happiness, how countries rate their happiness, and the importance of measuring happiness.” The report averages responses from about 1,000 surveys conducted in each country over the past three years. This year’s report is a collaboration between Gallup, the Oxford Wellbeing Research Centre, the UN Sustainable Development Solutions Network, and the WHR’s Editorial Board. The Cantril ladder is used to collect this data. The Cantril ladder is a scale from 0 to 10, with 10 being the best life and 0 being the worst life. Respondents are asked to rate their current life on a scale of 0 to 10.

For the detailed dataset used next, the element is country (137 countries) and the complete list of variables is as follows: Records of happiness score or "ladder score", social support, healthy life expectancy, freedom of life choices, generosity, perception of corruption, and GDP per capita. This time, only healthy life expectancy is used. Also, the healthy life expectancy dataset included the "State of Palestine", but since it was missing data in the healthy life expectancy dataset, it is excluded from this analysis. Therefore, the number of factors in this study is 136.

✧ Report: <https://worldhappiness.report/about/>

✧ Dataset: <https://worldhappiness.report/data/>

2 STATEMENT OF THE RESEARCH OBJECTIVE

The main purpose of this project is to explore, analyze, and describe the healthy life expectancy for 136 countries. The secondary purpose is to explore the potential relationship between the World Happiness Score, and the Healthy life expectancy. The data used are shown in Table 1 in the Appendix.

3 STATISTICAL ANALYSIS

3.1 Numerical Methods

3.1.1 Numerical Methods-Traditional Method (Excel)

Table 2 shows that:

1. Mean: A typical person is expected to have a healthy life expectancy of 64.97 years.
2. Median: The low half of the person had a healthy life expectancy of 65.84 or less.
3. Mode: This data does not show the most prevalent healthy life expectancy.
4. Standard Deviation: The 5.75 years is a measure of the spread of actual healthy Life expectancy around the expected healthy life expectancy of 64.97 years.
5. Kurtosis: The healthy life expectancy distribution is flat.
6. Skewness: The healthy life expectancy distribution is skewed toward the left.
7. Range: Hong Kong S.A.R. of China's healthy life expectancy exceeds Mozambique's healthy life expectancy by 25.75 years.
8. Minimum: Mozambique with a healthy life expectancy of 51.53 years had the lowest healthy life expectancy.
9. Maximum: Hong Kong S.A.R. of China with a Healthy life expectancy of 77.28 years had the highest healthy life expectancy.
10. Sum: The total sum of all healthy life expectancy was 8835.60 years.
11. Count: There are 136 elements in our data set.
12. Q1: The lowest quarter of people had a Healthy life expectancy of at most 60.55 years.
13. Q3: The top quarter of people had a healthy life expectancy of at least 69.54 years.
14. IQR=Q3-Q1: The middle 50% of healthy life expectancy differs by 8.99 years.
15. 20th Percentile: 1 out of 5 respondents of people had a healthy life expectancy of at most 58.60 years.
16. 33rd Percentile: A third of people had a healthy life expectancy of at most 63.43 years.
17. 66th Percentile: Two-thirds of people had a healthy life expectancy of at most 67.68 years.
18. 90th Percentile: 90 percent of people had a healthy life expectancy of at most 71.62 years.
19. Coefficient of Variation: Standard deviation is 8.85% of the mean
20. Z-score for Min: The healthy life expectancy in Mozambique is 2.34 standard deviations below the average.

21. Z-score for Max: The healthy life expectancy in Hong Kong S.A.R. of China is 2.14 Standard Deviations above average.
22. Z-score for Median: The healthy life expectancy in the low half is 0.15 standard deviations above the average.

● Application of Chebyshev's Theorem

Mean (μ)=64.97

Standard Deviation (σ)=5.75

	At least.....%	Low End	High End
Z=1.5	55.56%	56.35	73.60
Z=1.6	60.94%	55.77	74.17
Z=2.2	79.34%	52.32	77.62

At least 55.56% of people's healthy life expectancy is between 56.35 and 73.60 years.

At least 60.94% of people's healthy life expectancy is between 55.77 and 74.17 years.

At least 79.34% of people's healthy life expectancy is between 52.32 and 77.62 years.

● Application of Empirical Rule

Mean (μ)=64.97

Standard Deviation (σ)=5.75

Z-score	Approximate%	Low End	High End
1	68%	59.22	70.72
2	95%	53.47	76.47
3	99.70%	47.72	82.22

Approximately 68% of people had a healthy life expectancy between 59.22 and 70.72 years.

Approximately 95% of people had a healthy life expectancy between 53.47 and 76.47 years.

Approximately 99.70% of people had a healthy life expectancy between 47.72 and 82.22 years.

3.1.2 Numerical Methods-Using AI Method

First, we passed the raw data to Chat GPT and asked the AI to analyze the items analyzed in

3.1.1. The results were as follows:

1. Mean: 64.97 years

This is the average healthy life expectancy, representing the central tendency of the overall data points.

2. Median: 65.84 years

The middle value is when the healthy life expectancy data is arranged in ascending order. This value is less affected by outliers.

3. Mode: 71.15 years

The most frequently occurring value in the healthy life expectancy data. This indicates the concentration of the data.

4. Standard Deviation: 5.75 years

This indicates the spread of healthy life expectancy. A larger value means greater dispersion in the data.

5. Max: 77.28 years

The highest value of healthy life expectancy in the dataset.

6. Min: 51.53 years

The lowest value of healthy life expectancy in the dataset.

7. Q1 (First Quartile): 60.65 years

This shows that 25% of the data falls below this value.

8. Q3 (Third Quartile): 69.41 years

This indicates that 25% of the data is above this value.

9. Range: 25.75 years

The difference between the maximum and minimum values, indicating the overall spread of the data.

10. IQR (Interquartile Range): 8.76 years

The difference between Q3 and Q1, showing the range of the middle 50% of the data.

11. Coefficient of Variation: 8.85%

This measures the relative dispersion of the data, calculated as the standard deviation divided by the mean.

12. Skewness: -0.41

This indicates the asymmetry of the data. A negative value suggests a left-tailed distribution.

13. Kurtosis: -0.78

This measures the sharpness of the peak of the distribution. A negative value indicates a flatter distribution.

14. Z-scores:

These show how far each data point is from the mean, allowing for the identification of outliers.

Percentiles:

15. 20th Percentile: 58.76 years (20% of people are below this value)

16. 33rd Percentile: 63.46 years (33% of people are below this value)

17. 66th Percentile: 67.62 years (66% of people are below this value)

18. 90th Percentile: 71.57 years (90% of people are below this value)

- Chebyshev's Theorem:

This theorem states that for any distribution, a certain percentage of data will fall within k standard deviations from the mean. For example, for $k=2$, at least 75% of the data lies within ± 2 standard deviations from the mean.

- Empirical Rule:

For data following a normal distribution, approximately 68% of the data will fall within ± 1 standard deviation from the mean, about 95% within ± 2 standard deviations, and 99.7% within ± 3 standard deviations. If the distribution holds, you can evaluate the proportion of healthy life expectancy data that falls within specific ranges.

These indicators help deepen the understanding of the healthy life expectancy data and evaluate the differences and trends in health status across countries.

✧ URL: <https://chatgpt.com/>

3.1.3 Numerical Methods-Comparison of Traditional Method (Excel)VS AI Approach

1. Accuracy: I think both have the same accuracy and precision. Chat GPT has proven accuracy in data calculation, but there are some ambiguities in the data explanation. As mentioned in 3.1.2, much of the data does not include actual data in the explanation. For Chebyshev's Theorem and Empirical Rule, I did not specify the Z-score at first, so they provided a non-offensive analysis conclusion. However, when I instructed them to specify the Z-score "you", they set it to 1.5 and performed the analysis. The results are as follows:

A Z-score of 1.5 corresponds to a healthy life expectancy of 73.60 years, indicating that a country with this value has significantly better health outcomes than average. Additionally, we expect a substantial portion of the data (at least 44.4%, or even 86.6% under normality assumptions) to fall within the range of 56.34 to 73.60 years.

Therefore, I think it is necessary to check the accuracy with your own eyes.

2. Ease of Use: I think Chat GPT is more user-friendly because there is no need to remember complex functions, and once you hand over the data, it will calculate and analyze everything for you, so I think Chat GPT is easier for users to use.
3. Efficiency: Chat GPT was much faster than the Traditional Method (Excel). Excel is just as fast at getting the data results, but Chat GPT is faster at interpreting the details.
4. Flexibility: In terms of flexibility, both are available. Chat GPT will perform the analysis you want it to perform. Traditional Method (Excel) also allows you to analyze and create your data, so it is very flexible.
5. Comparison: Using Chat GPT to analyze data is much more efficient than the Traditional Method. However, it doesn't always provide the interpretation you want. Therefore, I think it's necessary to check the detailed data with your own eyes.

3.2 Graphical Methods

3.2.1 The Bar Graph

Graph 1 is a bar graph created based on the Frequency Distribution in Table 3. As shown in Graph 1, the most common number of countries is 65-68, with 34 countries. The approximate numbers are in the center of the graph, and the bars at the ends have the lowest numbers, with one country each at 50-53, 74-77, and 77-80.

3.2.2 The Pie Chart

Graph 2 is a pie chart created based on the Relative Frequency Distribution in Table 4. As shown in graph 2, 56-59, 62-65, and 68-71 account for around 10% of the total. In comparison, 50-53, 74-77, and 77-80 each account for 1% of the total, totaling just 3%. The most common is 65-68, at 25%, but that still only accounts for a quarter of the total.

3.2.3 The Stem and Leaf Display

Graph 3 is a stem and leaf display created based on the raw data in Table 1. Graph 3 makes it easier to see the details of each age, and we can see that 58, 64, 69, and 72 years are the years for healthy life expectancy in eight countries. We can also see that the number of countries that fall into this category decreases as we move toward the minimum and maximum values.

3.2.4 The Box-Whisker Plot

Graph 4 is a box-whisker plot created based on the raw data in Table 1. In Graph 4, there are no data that could be considered outliers. The IQR of 65.838 is not significantly different from the median of 65.84 in Table 2. We can see that the data is relatively balanced.

3.2.5 The Ogive

Graph 5 is an ogive created based on the Relative Cumulative Frequency Distribution in Table 6. As shown in Graph 5, we can see that the healthy life expectancy of at least 80.88% of countries does not exceed 71 years. Therefore, we can see that there are very few countries where the healthy life expectancy exceeds 71 years.

3.2.6 The Area Chart

Graph 6 is an area chart created based on the Relative Frequency Distribution in Table 4. In Graph 6, the number of countries drops sharply from 71-74 to 74-77, from 17.56% to 0.74%. Therefore, there are few countries with a healthy life expectancy exceeding 74-77 years, and at best, it remains just before 71-74 years. Looking at the figure, we can see that the area is relatively wide in the middle, at 65-68 years, and that the average healthy life expectancy of many countries is in that range.

3.3 Tabular Methods

3.3.1 The Frequency Distribution Table

As shown in Table 3, the most common range of healthy life expectancy is 65-68 years old in 34 countries, while the least common ranges are 50-53, 74-77, and 77-80, with only one country each. There are only two countries in the 74-80 range in total, and 134 countries in the 50-74 range, showing that the majority of countries have a healthy life expectancy just before 74 years old.

3.3.2 The Relative Frequency Distribution Table

In Table 4, we can see those 65-68 years old is exactly A quarter of respondents, 56-59, 62-65, 68-71, and 71-74 are in the 10% range. And, 50-65 is 41.92% overall, and 65-80 is 58.08%, so we can see that the healthy life expectancy in many countries is over 65 years old.

3.3.3 The Cumulative Frequency Distribution Table

In Table 5, we can see a sharp increase in the number of countries in the range of 53 to 56. We can see that the numbers are increasing relatively clearly from 50 to 71, but there is almost no increase in the numbers from 71 to 80. The increase is most notable between 62-68 and 65-68, where the number of countries increases by 34 at once.

3.3.4 The Relative Cumulative Frequency Distribution Table

As shown in Table 6, the growth rate from 50 to 74 is relatively high, but from 74 to 80 it hardly changes at all. In other words, the healthy life expectancy in almost all countries is just before 74 years. Also, we can see that the healthy life expectancy of about two out of five countries is up to 65 years.

4 DISCUSSION OF THE POSSIBLE OUTLIERS

First, referring to the Z-score calculated in 3.1.1, Z-score for Min: -2.34, Z-score for Max: 2.14. Outliers are considered when the Z-score is -3 or +3 or higher. The maximum and minimum Z-scores of this data are both within that range, and no outstanding data was observed in Graph 4, so it can be determined that there are no outliers in this data.

5 CONCLUSION, SUMMARY, AND POLICY RECOMMENDATIONS

This time, we investigated healthy life expectancy in 136 countries. The country with the highest healthy life expectancy was Hong Kong S.A.R. of China at 77.28 years, and the country with the lowest healthy life expectancy was Mozambique at 51.53 years. The difference was about 25.75 years. The average life expectancy was 64.97, which is not very different from the median of 65.84. When we investigated the maximum and minimum Z scores, we found no outliers in this dataset. Most countries' healthy life expectancies are between 53 and 74. The most common country with a healthy life expectancy of 65-68 years is 34 countries.

This time, we only investigated healthy life expectancy, so we were unable to compare it with other happiness values, and we were unable to determine how happiness is related to healthy life expectancy. We would like to investigate this relationship next time.

6 -APPENDIX-

Table 1 Healthy Life Expectancy in each Country

Country	Healthy life expectancy	Country 2	Healthy life expectancy3	Country 3	Healthy life expectancy5
Finland	71.15	Hungary	67.5	Senegal	59.999
Denmark	71.25	Argentina	67.2	Mauritania	60.475
Iceland	72.05	Honduras	64.063	Burkina Faso	55.461
Israel	72.697	Uzbekistan	65.301	Namibia	56.851
Netherlands	71.55	Malaysia	65.662	Turkiye	68.663
Sweden	72.15	Portugal	71.25	Ghana	58.763
Norway	71.5	South Korea	73.65	Pakistan	57.313
Switzerland	72.9	Greece	71.15	Niger	56.55
Luxembourg	71.675	Mauritius	63.85	Tunisia	67
New Zealand	70.35	Thailand	68.45	Kenya	58.499
Austria	71.15	Mongolia	60.5	Sri Lanka	67.15
Australia	71.05	Kyrgyzstan	66.852	Uganda	58.913
Canada	71.4	Moldova	65.299	Chad	53.125
Ireland	71.3	China	68.689	Cambodia	61.9
United States	65.85	Vietnam	65.502	Benin	56.095
Germany	71.3	Paraguay	65.9	Myanmar	61.388
Belgium	70.899	Montenegro	67.1	Bangladesh	64.548
Czechia	69.05	Jamaica	66.6	Gambia	57.9
United Kingdom	70.3	Bolivia	63.599	Mali	55.403
Lithuania	67.397	Russia	64.947	Egypt	63.503
France	72.3	Bosnia and Herzegovina	67.275	Togo	57.449
Slovenia	71.052	Colombia	69.35	Jordan	67.6
Costa Rica	70	Dominican Republic	64.399	Ethiopia	60.698
Romania	67.051	Ecuador	69	Liberia	56.7
Singapore	73.8	Peru	69.85	India	60.777
United Arab Emirates	66.243	Philippines	62.038	Madagascar	58.05
Taiwan Province of China	69.6	Bulgaria	66.5	Zambia	55.032
Uruguay	67.5	Nepal	61.847	Tanzania	59.401
Slovakia	68.838	Armenia	67.789	Comoros	59.425
Saudi Arabia	64.399	Tajikistan	62.226	Malawi	58.475
Estonia	69.65	Algeria	66.549	Botswana	54.725
Spain	72.35	Hong Kong S.A.R. of China	77.28	Congo (Kinshasa)	55.375
Italy	72.05	Albania	69.15	Zimbabwe	54.05
Kosovo	65.195	Indonesia	63.048	Sierra Leone	54.9
Chile	70.3	South Africa	56.989	Lebanon	66.149
Mexico	65.8	Congo (Brazzaville)	56.85	Afghanistan	54.712
Malta	71.6	North Macedonia	66.5		
Panama	68.95	Venezuela	64.05		
Poland	69.049	Laos	60.946		
Nicaragua	65.65	Georgia	64.95		
Latvia	66.4	Guinea	54.185		
Bahrain	65.825	Ukraine	64.55		
Guatemala	62.9	Ivory Coast	55.953		
Kazakhstan	65.802	Gabon	58.252		
Serbia	67.088	Nigeria	54.891		
Cyprus	72.802	Cameroon	55.847		
Japan	74.349	Mozambique	51.53		
Croatia	68.95	Iraq	63.415		
Brazil	65.749	Morocco	63.901		
El Salvador	65.597	Iran	66.6		

Table 2 Healthy life expectancy of the numerical values of each item

Healthy life expectancy	
Mean	64.97
Standard Error	0.49
Median	65.84
Mode	#N/A
Standard Deviation	5.75
Sample Variance	33.07
Kurtosis	-0.78
Skewness	-0.41
Range	25.75
Minimum	51.53
Maximum	77.28
Sum	8835.60
Count	136.00
Q1	60.55
Q3	69.54
IQR=Q3-Q1	8.99
20th Percentile	58.60
33rd Percentile	63.43
66th Percentile	67.68
90th Percentile	71.62
Coefficient of Variation	8.85%
Z-score for Min	-2.34
Z-score for Max	2.14
Z-score for Median	0.15

Table 3 Healthy Life Expectancy of the Frequency Distribution

Row Labels	the Frequency Distribution
50-53	1
53-56	13
56-59	15
59-62	11
62-65	17
65-68	34
68-71	19
71-74	24
74-77	1
77-80	1
Grand Total	136

Table 4 Healthy Life Expectancy of the Relative Frequency Distribution

Row Labels	the Relative Frequency Distribution
50-53	0.74%
53-56	9.56%
56-59	11.03%
59-62	8.09%
62-65	12.50%
65-68	25.00%
68-71	13.97%
71-74	17.65%
74-77	0.74%
77-80	0.74%
Grand Total	100.00%

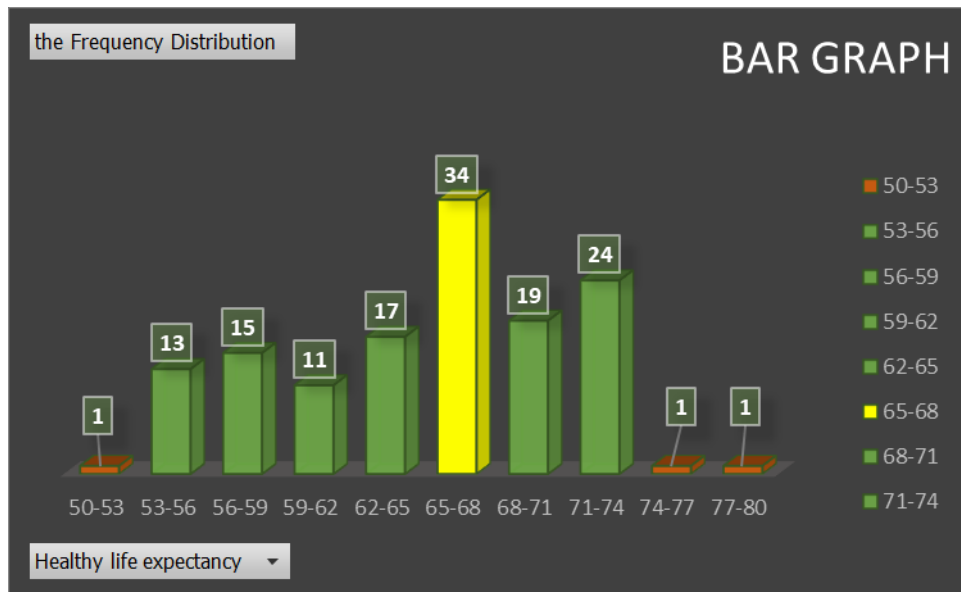
Table 5 Healthy Life Expectancy of the Cumulative Frequency Distribution

Row Labels	the Cumulative Frequency Distribution
50-53	1
53-56	14
56-59	29
59-62	40
62-65	57
65-68	91
68-71	110
71-74	134
74-77	135
77-80	136

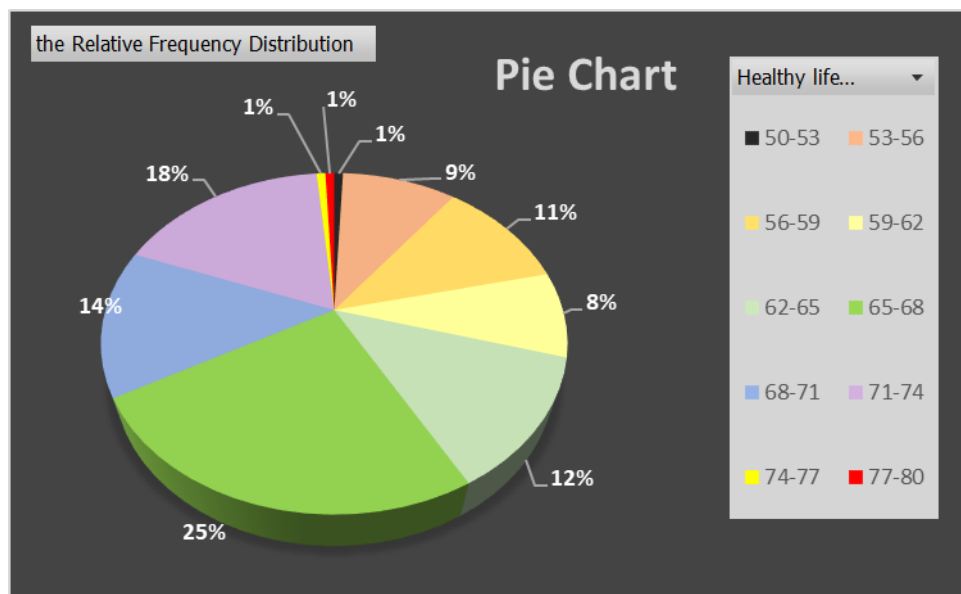
Table 6 Healthy Life Expectancy of the Relative Cumulative Frequency Distribution

Row Labels	the Relative Cumulative Frequency Distribution
50-53	0.74%
53-56	10.29%
56-59	21.32%
59-62	29.41%
62-65	41.91%
65-68	66.91%
68-71	80.88%
71-74	98.53%
74-77	99.26%
77-80	100.00%

Graph 1 Bar Graph



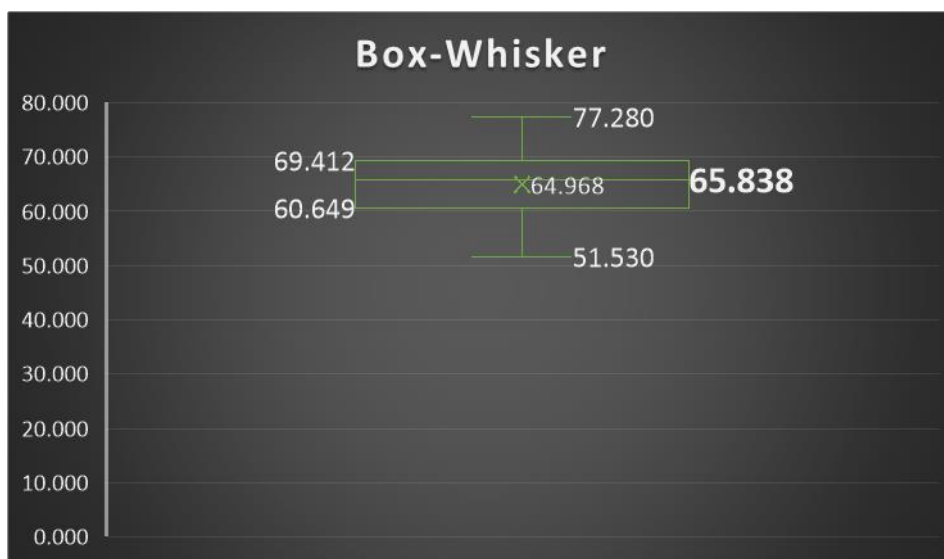
Graph 2 Pie Chart



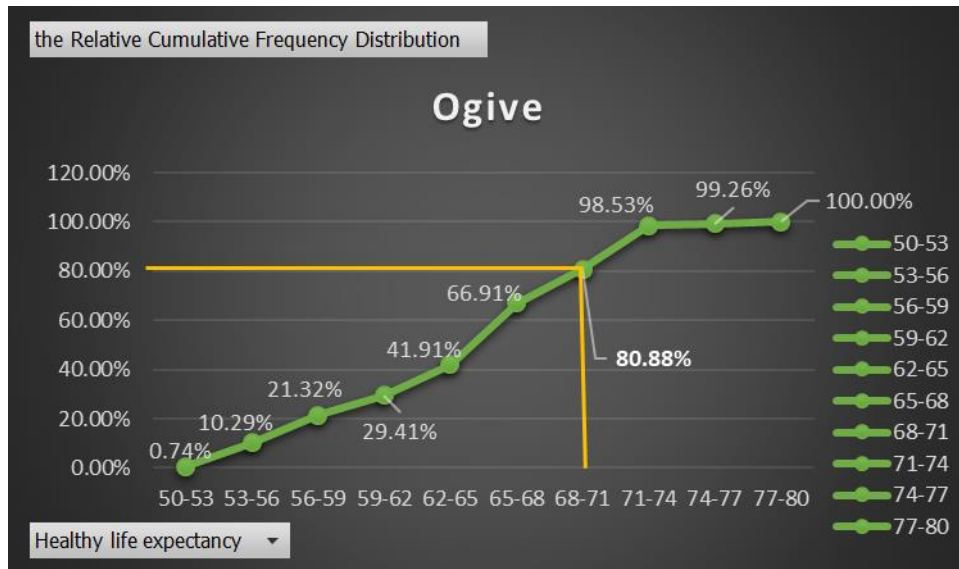
Graph 3 Stem and Leaf Display

Stem	Leaf
51	5
53	1
54	0, 1, 1, 7
55	0, 3, 4, 4, 8, 9
56	0, 5, 7, 8, 8, 8, 9
57	3, 4, 9
58	0, 2, 4, 5, 9, 7, 7, 9
59	4, 9
60	4, 5, 6, 6, 7, 7, 9
61	3, 8, 8, 8, 9
62	0, 2, 2, 9
63	0, 4, 5, 5, 6, 8, 8
64	0, 0, 1, 3, 3, 5, 5, 9
65	1, 1, 2, 2, 5, 5, 7
66	1, 1, 2, 3, 5, 5, 8
67	0, 0, 1, 5, 5, 8, 9
68	4, 6, 6, 8
69	0, 0, 1, 3, 3, 5, 6, 6
70	0, 3, 3, 5
71	0, 0, 0, 2, 5, 5, 5
72	0, 0, 1, 3, 5, 6, 7, 8
73	6, 8
74	3
77	2

Graph 4 Box-Whisker Plot



Graph 5 Ogive



Graph 6 Area Chart

