

# Contents

<b>1 Abstract</b>	<b>4</b>
<b>2 Introduction</b>	<b>4</b>
<b>3 Original Dataset</b>	<b>4</b>
<b>4 Data Manipulation and EDA</b>	<b>5</b>
4.1 Null values . . . . .	5
4.2 Variables and Correlation . . . . .	7
4.3 Exploratory Data Analysis . . . . .	9
4.3.1 Life Satisfaction Distribution Across Countries . . . . .	9
4.3.2 Understanding Relationships Between Metrics (Correlation Analysis) . . . . .	9
4.3.3 A Holistic Comparison of Countries (Radar Chart) . . . . .	11
4.3.4 Visualizing Patterns Through Scatter Plots . . . . .	12
<b>5 Clustering</b>	<b>14</b>
5.1 MClust – Model-Based Clustering** . . . . .	14
5.2 KMeans – K-Means Clustering . . . . .	14
5.3 KMedoids – K-Medoids Clustering . . . . .	14
5.4 Performing the Model Based Clustering. . . . .	14
5.4.1 BIC and Number of Components Used for Clustering . . . . .	15
5.4.2 Classification of Data Points . . . . .	15
5.4.3 Uncertainty . . . . .	16
5.4.4 Density Distribution of Data Within Clusters . . . . .	17
5.5 Analysis of Clusters . . . . .	19
5.6 K-Means . . . . .	22
5.6.1 Comparing K-Means vs MCLUSt . . . . .	23
5.7 K-Medoids Clustering . . . . .	24
5.8 Further Analysis for Clustering . . . . .	26
5.9 Manual clustering vs M-Clust model . . . . .	26
5.10 Manual clustering vs K-Means Cluster . . . . .	27
5.11 Conclusion of Clustering Approach . . . . .	28

<b>6 Fitting the Regression Models</b>	<b>29</b>
6.1 Model 1:Linear Regression . . . . .	29
6.2 Model 2: Linear Regression using Pairwise Interaction effects . . . . .	29
6.3 Model 3: Model with Interaction effects based on relevance of factors. . . . .	30
6.4 Parameters from the Three Models . . . . .	31
6.4.1 Adjusted R-square . . . . .	31
6.4.2 AIC . . . . .	32
6.4.3 BIC . . . . .	32
6.4.4 Residuals . . . . .	33
6.4.5 Normality using Q-Q Plots . . . . .	34
6.4.6 ANOVA . . . . .	37
6.5 Random Intercept Model . . . . .	38
6.6 Fitting the Random Slope and Random Intercept Model . . . . .	40
6.6.1 Random Intercept Model with Random Slope for Education . . . . .	40
6.6.2 Random Intercept Model with Random Slope for Community . . . . .	42
6.7 R Squared . . . . .	44
6.8 Comparison of Models . . . . .	47
<b>7 Conclusion</b>	<b>47</b>
<b>8 Appendix</b>	<b>48</b>
8.1 A: Summary of Clusters with Model Fitted . . . . .	48
8.2 B: Summary of Linear Regression using Pairwise Interaction effects . . . . .	52

# 1 Abstract

This project aims to **estimate the effect of various factors such as health, education, and income among others, on the life satisfaction of people in different regions from different countries**. For this, we carried out a robust approach with MEMs & Robust Estimators. Also, we performed a robust clustering of various factors into different clusters of life satisfaction.

## 2 Introduction

For this project, we utilized the OECD database focusing on **indicators and life-satisfaction scores across various regions within different countries**. This comprehensive dataset captures diverse dimensions of well-being, such as education, employment, health, environment, and social support, which together contribute to a region's overall quality of life. The indicators allow for meaningful comparisons between regions, highlighting disparities and trends. Although the data reflects different years depending on the country, for consistency we adopted the **latest available data**.

It is essential to acknowledge that while some variability exists in the timing of the data, we assumed that any changes within a year or two would be minimal and unlikely to significantly alter policies or life-satisfaction scores. This approach enables us to draw relevant conclusions about the factors influencing regional well-being. By examining key metrics such as income, safety, health, and life satisfaction, this report aims to present a clear picture of well-being across OECD regions and highlight patterns that can inform future policies and initiatives.

## 3 Original Dataset

To execute this project, a dataset containing the different metrics regarding well-being was downloaded and linked to a variable named **df\_wb**.

The variable **df\_wb** contains 447 tuples and 28 columns (25 are the attributes to analyze).

```
## [1] 447 28
```

The variables included are the following:

1. **Country:** Includes the name of all the countries included in the results.
2. **Region:** Includes the name of all the cities from the countries included in the results.
3. **Code:** Code associated to each pair country-city.
4. **Population.with.at.least.secondary.education.(%)**: Percentage of the population completing secondary education.
5. **Employment rate (%)**: Percentage of the working-age population employed.
6. **Unemployment rate (%)**: Percentage of people without jobs actively seeking work.
7. **Household disposable income per capita (USD PPP)**: Average income available per person, in USD.
8. **Homicide rate (per 100k)**: Number of homicides per 100.000 people.
9. **Mortality rate (per 1k)**: Deaths per 1,000 people annually.
10. **Life expectancy**: Average expected lifespan (years).
11. **Air pollution (level of PM2.5,  $\mu\text{g}/\text{m}^3$ )**: Fine particulate air pollution levels.
12. **Voter turnout (%)**: Share of voters participating in elections.
13. **Broadband access (% of household)**: Percentage of households with internet access.
14. **Internet download speed 2021-Q4 (%)**: Internet speed growth/decline in 2021-Q4.
15. **Number of rooms per person**: Average living space per person.
16. **Perceived social network support (%)**: Percentage of people with available social support.
17. **Self assessment of life satisfaction (0-10)**: Subjective rating of overall happiness.
18. **Education (0-10)**: Regional score for education.
19. **Jobs (0-10)**: Score based on employment indicators.
20. **Income (0-10)**: Score for household income.
21. **Safety (0-10)**: Regional score for personal safety.
22. **Health (0-10)**: Score for health indicators.
23. **Environment (0-10)**: Score for environmental quality.
24. **Civic engagement (0-10)**: Score for public participation.
25. **Accessibility to services (0-10)**: Availability of public services.
26. **Housing (0-10)**: Score for housing quality and affordability.
27. **Community (0-10)**: Score for social cohesion.
28. **Life satisfaction (0-10)**: Overall happiness score.

The value type of each attribute appeared to be character, so further data pre-processing and data manipulation will be done.

## 4 Data Manipulation and EDA

### 4.1 Null values

Before analyzing the hypothesis and attributes, a check on the data structure was conducted to prevent potential errors in the future. This involved examining both data types and null values.

First we checked the **null values** to determine their significance and understand which is the best action to take regarding this matter. After checking that there were no null values, but yet we couldn't perform some calculations on the attributes, we did a more detailed analysis to realize that there were null values which were replaced by the **character “.”**.

```

## wb_missing_values
## Country 0.0000000
## Region 0.0000000
## Code 0.0000000
## Population.with.at.least.secondary.education.(%) 5.3691275
## Employment.rate.(%) 3.3557047
## Unemployment.rate.(%) 3.5794183
## Household.disposable.income.per.capita.(USD.PPP) 2.6845638
## Homicide.rate.(per.100k) 0.6711409
## Mortality.rate.(per.1k) 0.0000000
## Life.expectancy 1.7897092
## Air.pollution.(level.of.PM2.5,.µg/m³) 0.0000000
## Voter.turnout.(%) 0.2237136
## Broadband.access.(%.of.household) 1.3422819
## Internet.download.speed.2021-Q4.(%) 0.6711409
## Number.of.rooms.per.person 0.6711409
## Perceived.social.network.support.(%) 2.2371365
## Self.assessment.of.life.satisfaction.(0-10) 2.2371365
## Education.(0-10) 5.3691275
## Jobs.(0-10) 3.3557047
## Income.(0-10) 2.6845638
## Safety.(0-10) 0.6711409
## Health.(0-10) 0.0000000
## Environment.(0-10) 0.0000000
## Civic.engagement.(0-10) 0.2237136
## Accessibility.to.services.(0-10) 0.0000000
## Housing.(0-10) 0.6711409
## Community.(0-10) 2.2371365
## Life.satisfaction.(0-10) 2.2371365

```

Taking this into account, we concluded that there were no significance level of null values (in the form of “..”), so for the moment we decided to keep all the information and look for further actions regarding null values.

```

##      Country n_cities n_missing     p_na
## 1      Canada      13        3 23.07692
## 2      Chile       16        1  6.25000
## 3  Colombia      33        9 27.27273
## 4 Costa Rica       6        6 100.00000
## 5   Finland       5        1 20.00000
## 6    France      18        5 27.77778
## 7   Iceland       2        2 100.00000
## 8    Israel       6        1 16.66667
## 9     Japan      10       10 100.00000
## 10 Lithuania      10        1 10.00000
## 11 New Zealand     14        2 14.28571

```

Also, we can see that, when analyzing the null values by country, there were some cases that have 100% of missing values in some attributes. This is why, we decided to drop **Costa Rica, Iceland and Japan** as they had attributes without values, and this would have not been useful for our project. In addition to them, we also decided to exclude **Türkiye** as we found poor the information contained in the **Life.Satisfaction.(0-10)** attribute.

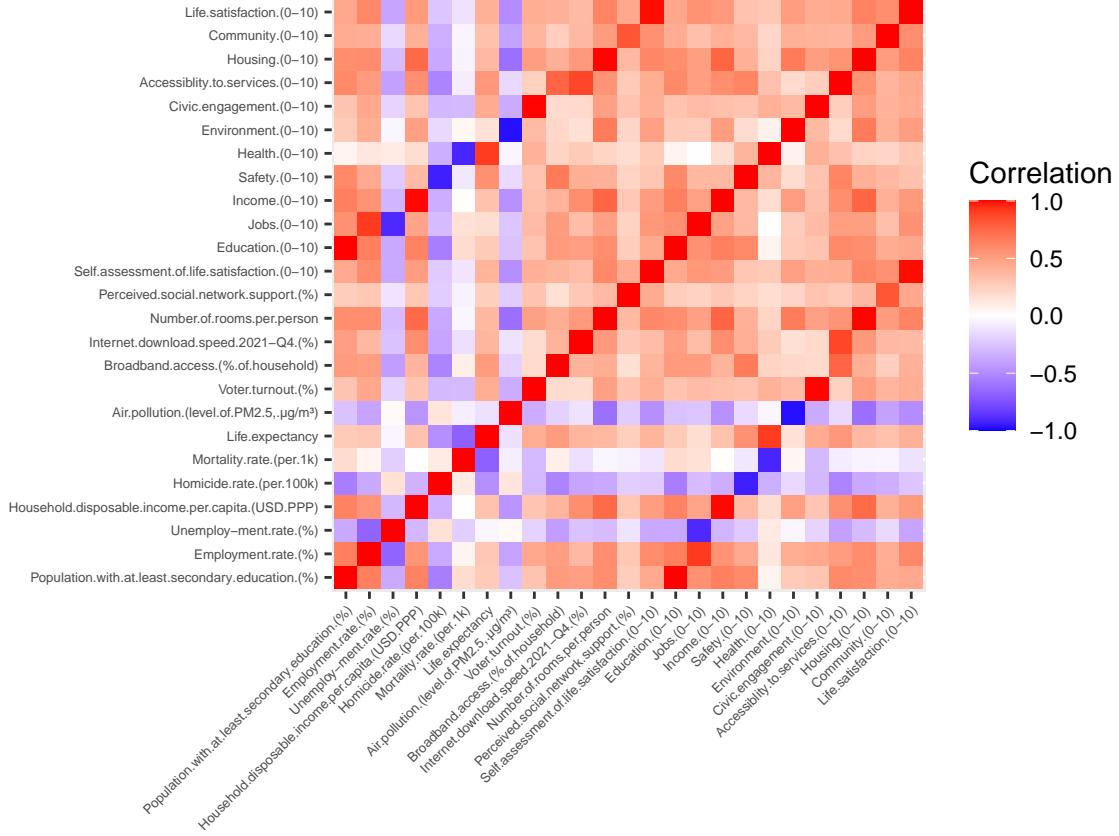
Moreover, for those cases with some null values, we decided to replace them with the **median of the country**. We decided this because the data is divided by city, so we could use the median of the rest cities

of the country for those cities with null values. Furthermore, the median is a better option than the mean as we can avoid the influence of any possible outlier.

Finally, after all this data manipulation, we have the correct data type for each attribute.

## 4.2 Variables and Correlation

Continuing with the data manipulation, we decided to carry out a correlation analysis to determine the need of any further removal of attributes. For this, we decided to plot a correlation heatmap to rapidly see any pair of attributes highly correlated, and with this, determine the removal of one of them.



As a result of this analysis, we decided to drop the following variables, because they can be explained by others (high correlation) and probably have less information than other correlated variables (this was checked manually with the dataset):

1. **Unemployment.rate.(%)**: -0.9 of correlation with Jobs.(0-10).
2. **Life.expectancy**: 0.9 of correlation with Health.(0-10).
3. **Internet.download.speed.2021-Q4.(%)**: 0.87 of correlation with Accessibility.to.services.(0-10).
4. **Perceived.social.network.support.(%)**: 0.83 of correlation with Community.(0-10).
5. **Voter.turnout.(%)**: 0.99 of correlation with Civic.engagement.(0-10).
6. **Air.pollution.(level.of.PM2.5.,µg/m³)**: -0.97 of correlation with Environment.(0-10).
7. **Population.with.at.least.secondary.education.(%)**: 0.99 of correlation with Education.(0-10).
8. **Household.disposable.income.per.capita.(USD.PPP)**: 0.99 of correlation with Income.(0-10).
9. **Employment.rate.(%)**: 0.92 of correlation with Jobs.(0-10).

Additionally, we decided to remove also Homicide.rate.(per.100k) and Mortality.rate.(per.1k) because we assumed that can be represented by Safety (0-10). Also we removed Broadband.access.(%.of.household) and

Number.of.rooms.per.person as we did not consider it useful for the project. Last but not least, we removed Self.assessment.of.life.satisfaction.(0-10) as we directly used Life.satisfaction.(0-10).

Taking into account the resulting dataset, we decided to exclude those countries that have less than 10 cities in the dataset, as we have 10 variables plus the main variable (life satisfaction).

The last step of the data manipulation was to set the upper bound of the scale to 10, as there were some cases with decimals that ended up being a little bit over 10.

So the final dataset used in the models was the following, containing 295 tuples and 14 columns (10 are the attributes to analyze, plus the life satisfaction attribute):

```
## [1] 285 14

##   Country          Region          Code          Education.(0-10)
## Length:285      Length:285      Length:285      Min.    : 0.009965
## Class :character Class :character Class :character 1st Qu.: 4.239527
## Mode  :character Mode  :character Mode  :character Median  : 7.247493
##                                         Mean    : 6.388977
##                                         3rd Qu.: 9.079590
##                                         Max.   :10.000000
##   Jobs.(0-10)       Income.(0-10)      Safety.(0-10)      Health.(0-10)
## Min.    : 0.008579  Min.    : 0.002448  Min.    : 0.00077  Min.    :0.009284
## 1st Qu.: 4.394316  1st Qu.: 0.814674  1st Qu.: 7.63587  1st Qu.:3.422444
## Median  : 6.881235 Median  : 3.393236  Median  : 9.48370  Median :6.089307
## Mean    : 6.183821 Mean    : 3.514659  Mean    : 8.04128  Mean   :5.618996
## 3rd Qu.: 8.286467  3rd Qu.: 4.421304  3rd Qu.: 9.80978  3rd Qu.:7.759107
## Max.   :10.000000  Max.   :10.000000  Max.   :10.00000  Max.   :9.504940
##   Environment.(0-10) Civic.engagement.(0-10) Accessibility.to.services.(0-10)
## Min.    : 0.00532  Min.    :0.009971  Min.    :0.008148
## 1st Qu.: 5.47264  1st Qu.:3.200993  1st Qu.:4.219586
## Median  : 7.26368 Median  :5.113918  Median :6.666018
## Mean    : 6.85327 Mean    :4.955544  Mean   :5.972564
## 3rd Qu.: 8.55721  3rd Qu.:7.103542  3rd Qu.:7.980286
## Max.   :10.00000  Max.   :9.246560  Max.   :9.888185
##   Housing.(0-10)     Community.(0-10)     Life.satisfaction.(0-10)
## Min.    : 0.008387  Min.    : 0.00998  Min.    : 0.009778
## 1st Qu.: 1.516854  1st Qu.: 4.41441  1st Qu.: 4.615385
## Median  : 4.550562 Median  : 7.07207  Median : 6.538462
## Mean    : 4.609031 Mean    : 6.31863  Mean   : 6.059621
## 3rd Qu.: 7.359551  3rd Qu.: 8.37838  3rd Qu.: 8.076923
## Max.   :10.000000  Max.   :10.00000  Max.   :10.000000
```

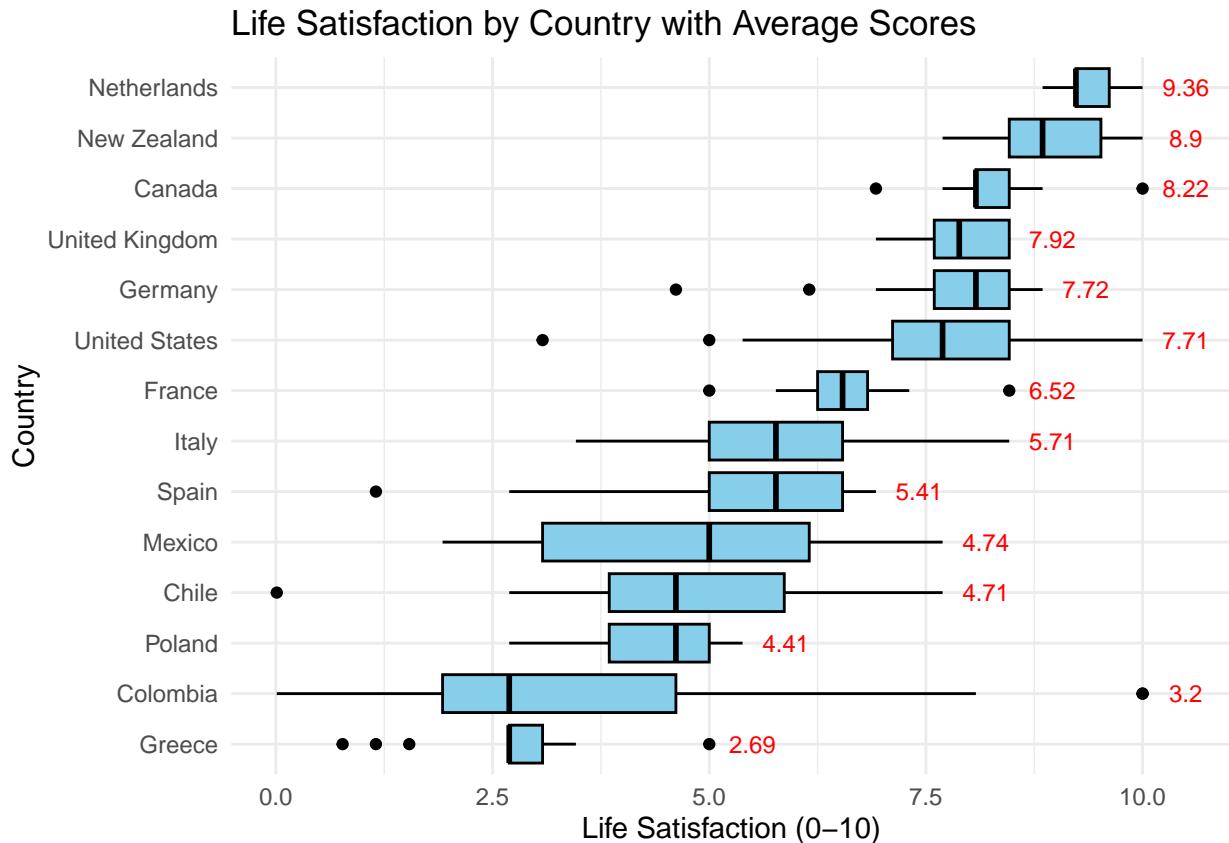
## 4.3 Exploratory Data Analysis

After completing a rigorous data cleaning process to remove redundancies and address multicollinearity, we continued with an Exploratory Data Analysis (EDA) to obtain insights about the factors influencing life satisfaction across 15 countries.

### 4.3.1 Life Satisfaction Distribution Across Countries

We began by analyzing how life satisfaction varies from one country to another. As illustrated in the box plot below, countries like the Netherlands, New Zealand, and Canada stood out with the highest average life satisfaction scores (9.36, 8.9, and 8.22, respectively). On the other hand, countries such as Greece, and Colombia reported the lowest scores.

Interestingly, in countries like Mexico and Colombia, we observed wide variability, suggesting significant regional differences within these nations. This led us to our **first key insight**: life satisfaction is not equally distributed across countries, and economic or social disparities likely play a role.



### 4.3.2 Understanding Relationships Between Metrics (Correlation Analysis)

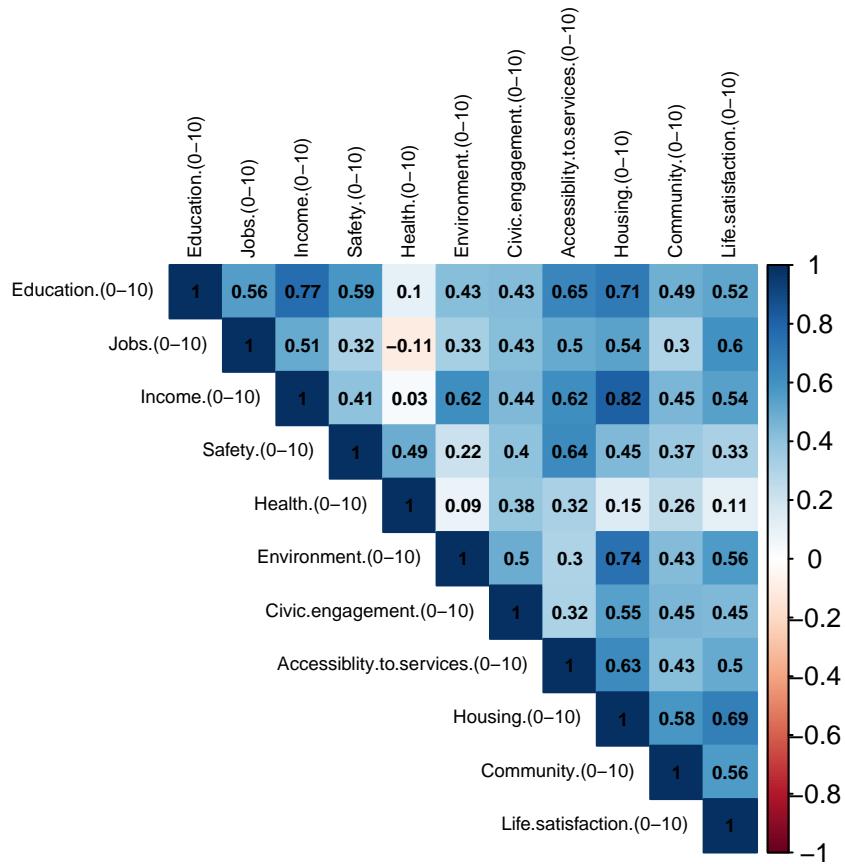
Deeper in our analysis, we explored the correlations between life satisfaction and other well-being metrics using a heatmap:

1. **Housing** had the strongest correlation with life satisfaction (0.69), showing that access to adequate housing and living conditions significantly influences well-being.
2. **Jobs** metric followed closely at 0.6, highlighting the importance of employment opportunities.

3. **Environment** and **Community** stood out with a correlation of 0.56 both, highlighting that a clean and sustainable environment and, connectedness and solidarity among groups in society are also an important part of life.

Surprisingly, **Income** showed a slightly lower correlation at 0.54, suggesting that while important, economic wealth is not the sole driver of life satisfaction.

This analysis revealed that housing conditions, employment opportunities, environment and community are more influential for life satisfaction than income alone.

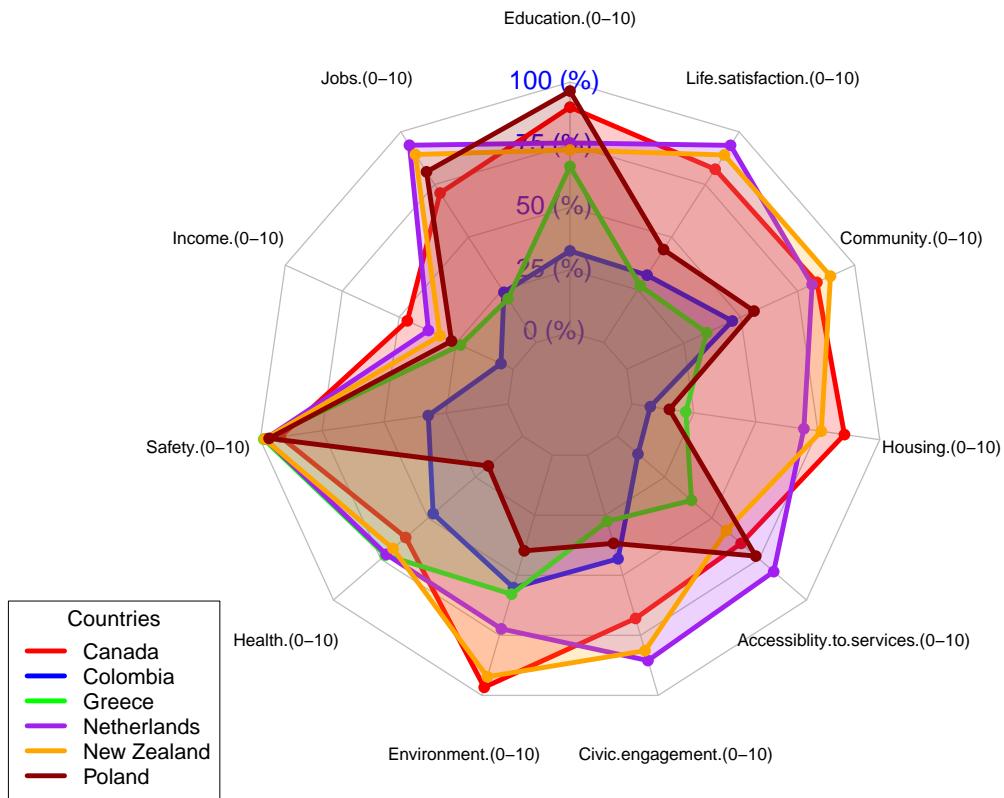


### 4.3.3 A Holistic Comparison of Countries (Radar Chart)

The radar chart highlights the Top 3 countries (Netherlands, Canada, and New Zealand) and the Bottom 3 countries (Poland, Greece, and Colombia) based on key life attributes. Netherlands leads in Jobs, Civic Engagement, and Accessibility to Services, achieving the highest Life Satisfaction. Canada excels in Housing, Environment, and Income, demonstrating a strong economic and living standard. New Zealand stands out with the best Community scores, fostering a strong sense of belonging.

Among the lowest countries, Poland performs well in Education but struggles with Environment and Housing. Colombia, despite having a better Environment than expected, lags in Housing, Accessibility to Services, and especially Safety, making it the least secure. This highlights how gaps in infrastructure, safety, and education can lead to lower life satisfaction across these countries.

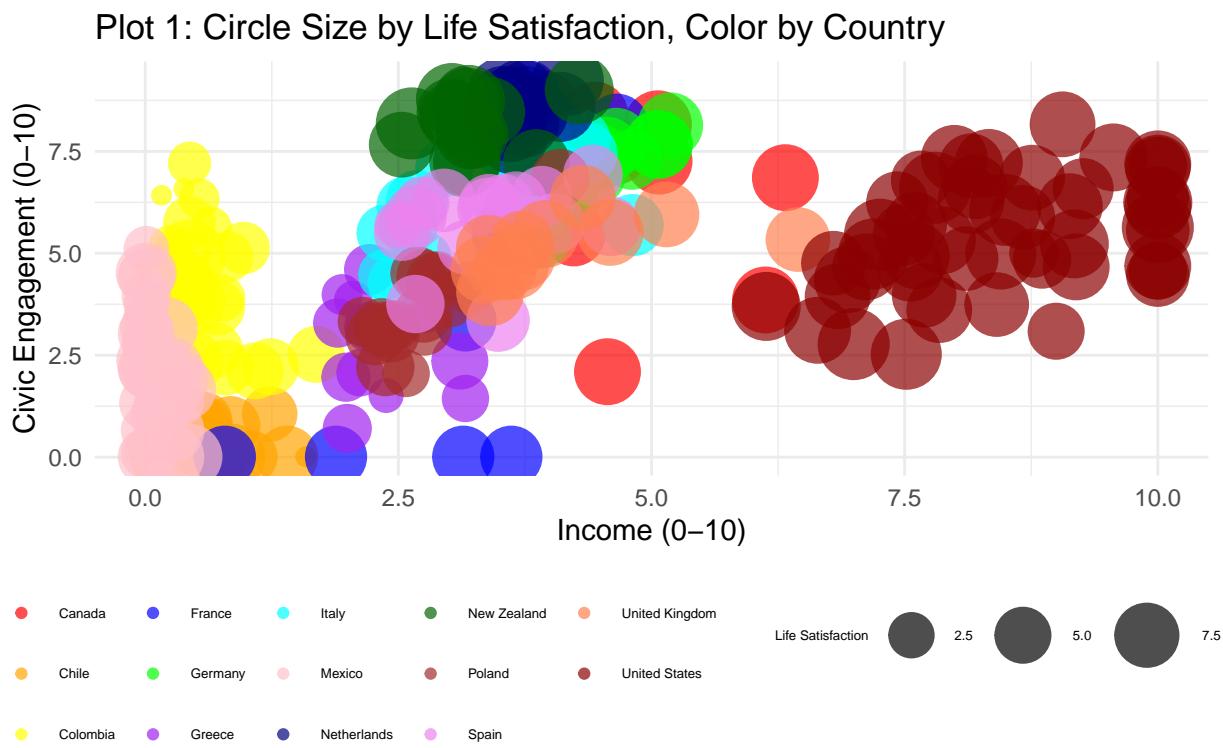
**Top 3 and Bottom 3 Countries – Radar Chart**



#### 4.3.4 Visualizing Patterns Through Scatter Plots

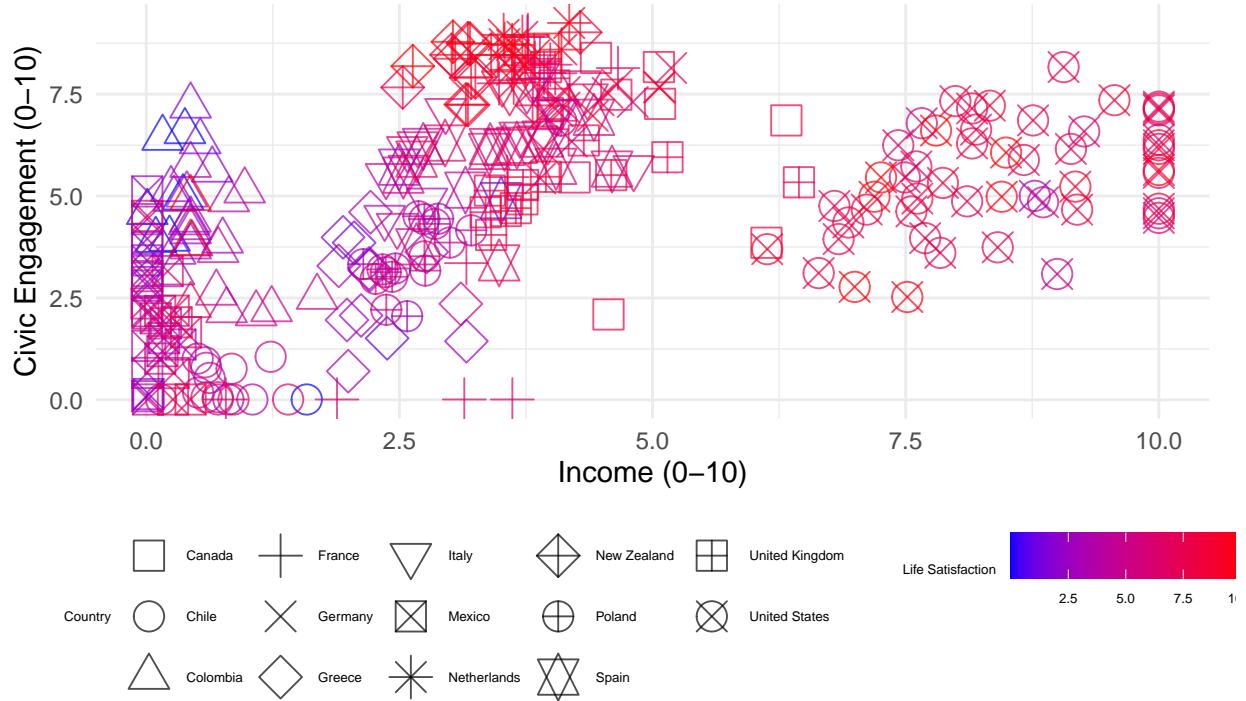
Finally, we created two plots to explore the relationships between Income, Civic Engagement, and Life Satisfaction:

1. **Plot 1: Circle Size by Life Satisfaction, Color by Country:** This plot revealed that countries with larger circles (higher life satisfaction) tend to have higher income and civic engagement levels. For instance, the Netherlands and the United States dominate the upper-right region of the plot. Conversely, smaller circles in the lower-left region highlight countries like Greece, where income and civic engagement remain low.



**2. Plot 2: Different Shapes by Country, Color by Life Satisfaction:** Here, each country is represented with a unique shape, and colors reflect their life satisfaction. Countries with higher satisfaction are visibly clustered in red, whereas lower-satisfaction countries are more dispersed, particularly toward the lower income range. These visualizations allowed us to see the clear relationships between income, civic engagement, and well-being, while also identifying disparities and regional patterns.

Plot 2: Different Shapes by Country, Color by Life Satisfaction



## 5 Clustering

Clustering is a machine learning technique used to group similar data points into clusters. In this project, we used three clustering methods to compare and analyze the results.

Under clustering analysis, we explored various clustering techniques to group countries based on different features, such as life satisfaction, income, and civic engagement. The clustering methods employed include **MClust, K-Means, and K-Medoids**. We use these techniques to analyze the data, create different cluster profiles, and compare the clustering results. We also use dimensionality reduction methods like **PCA and T-SNE** for visualizing the clusters, and various statistical plots to examine the distribution and characteristics of each cluster.

### 5.1 MClust – Model-Based Clustering\*\*

MClust is a probabilistic clustering method that models the data using a mixture of Gaussian distributions. It is based on the assumption that each cluster follows a normal distribution, but each cluster can have different means and variances.

For analysing the clusters, we use BIC criteria and then we plot the classification, uncertainty and density plots.

Classification Plot: The classification step assigns each data point (country) to a specific cluster. The MClust model automatically performs this step and outputs the cluster assignments for each country.

Uncertainty Plot: Uncertainty plots show the probability of a data point belonging to a specific cluster. These plots highlight how certain or uncertain the model is about assigning a particular data point to a cluster.

Density Plot: Density plots visualize the distribution of the data within each cluster. These plots show how the data points are distributed across the different feature values, and help us understand the characteristics of each cluster.

### 5.2 KMeans – K-Means Clustering

K-Means is one of the most widely used clustering algorithms. It works by partitioning the data into K clusters, where each cluster is represented by its centroid (mean of the points within that cluster).

### 5.3 KMedoids – K-Medoids Clustering

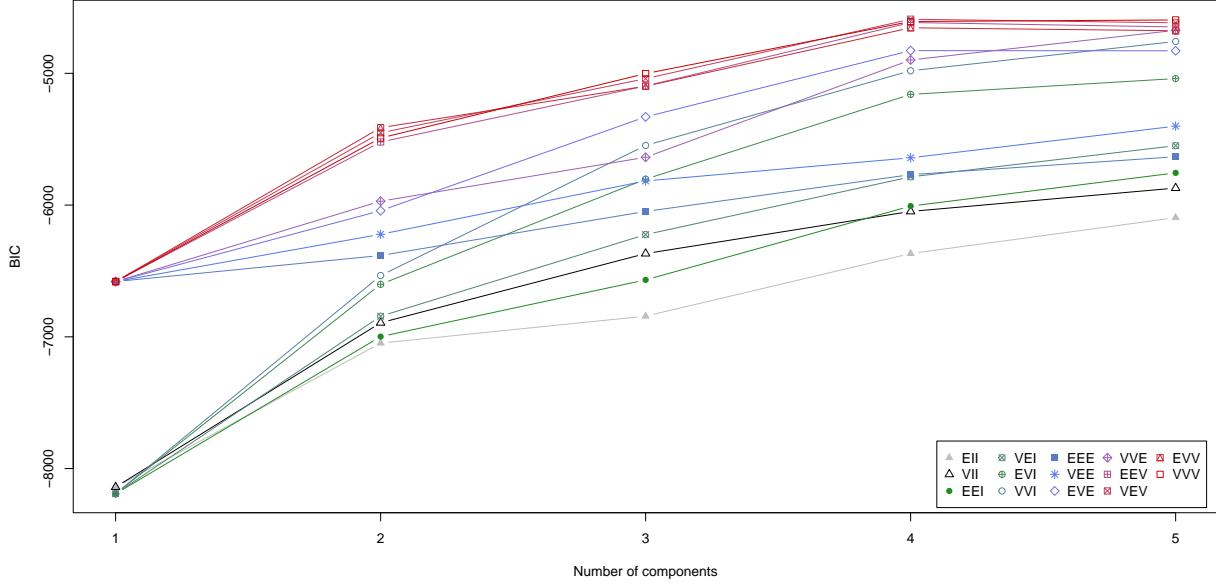
K-Medoids is a clustering algorithm similar to K-Means, but instead of using centroids, it uses actual data points as the center of clusters, making it more robust to noise and outliers.

### 5.4 Performing the Model Based Clustering.

For this model, we used **Mclust (Model Based Clustering)**, which employs **Gaussian Mixture Models (GMM)** to perform the clustering. It automatically selects the optimal number of clusters and model type based on criteria such as the **Bayesian Information Criterion (BIC)**. This approach ensures both accuracy and interpretability in clustering. The summary shows that the ***optimal number of clusters is 4 (Appendix A)***.

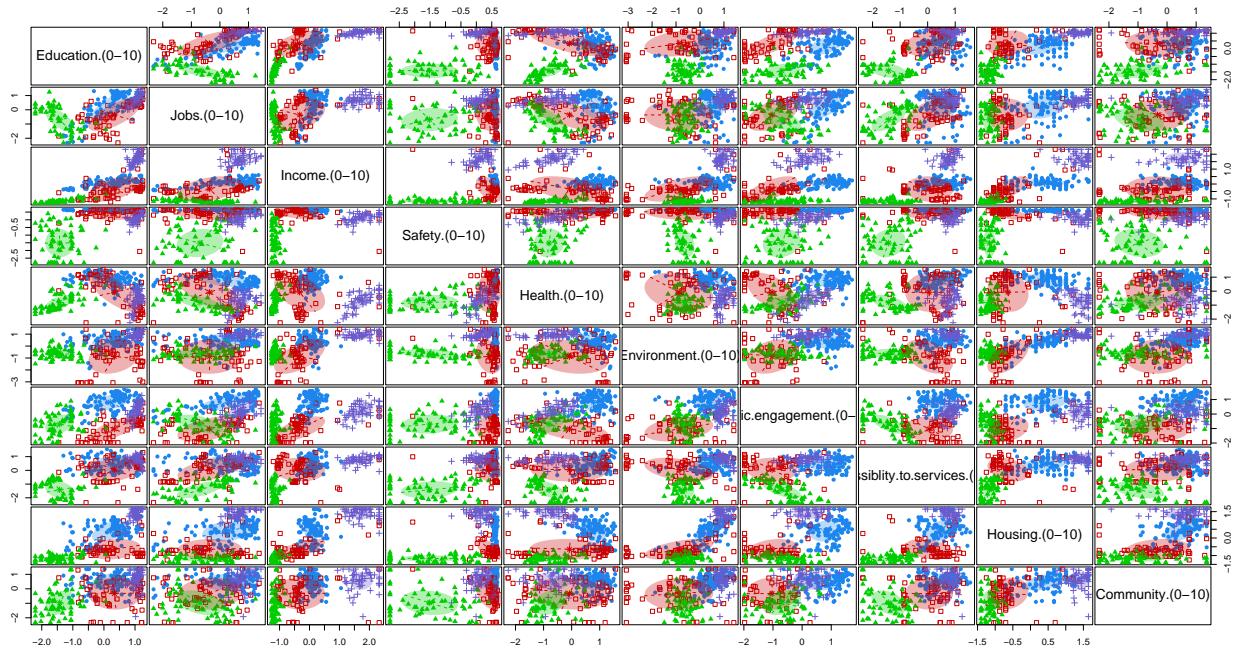
Once we got the 4 clusters, we created a Classification and Uncertainty plots for the different clusters:

#### 5.4.1 BIC and Number of Components Used for Clustering



By evaluating the fit of models with different numbers of clusters and covariance structures, the BIC plot confirmed that the **VEV model with four components** was the most appropriate. This indicated a good balance between model complexity and interpretability. This result reassures us that the identified clusters capture meaningful differences in well-being metrics while avoiding overfitting.

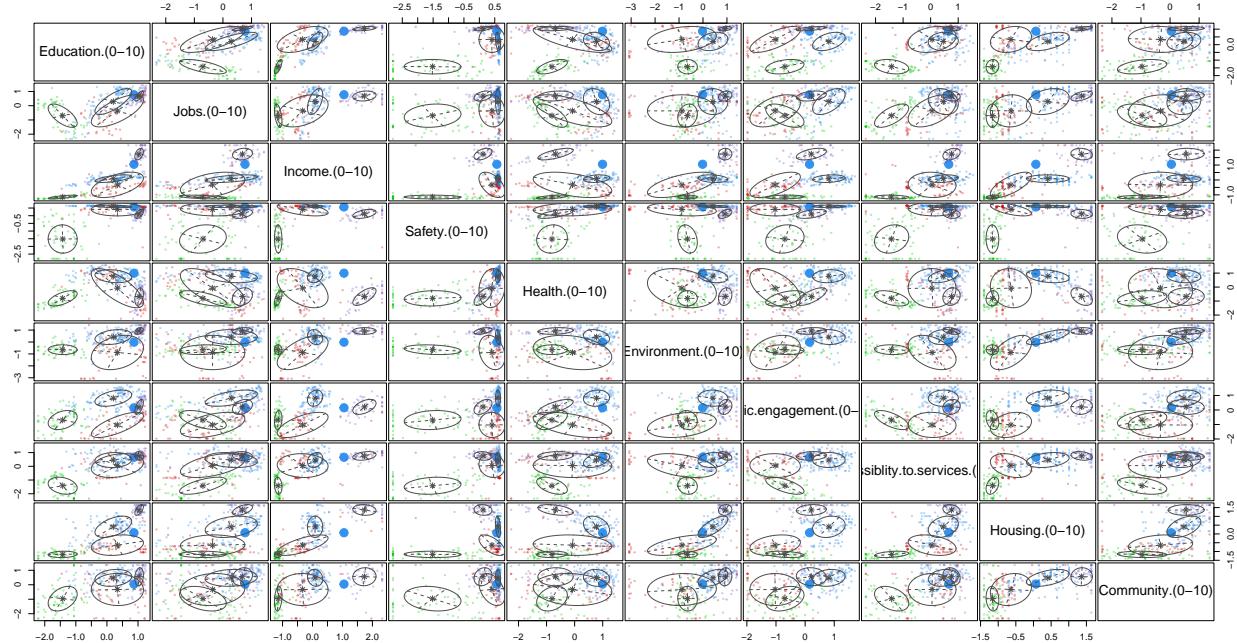
#### 5.4.2 Classification of Data Points



The classification plot of data points across clusters visualized the **assignment of regions to the four clusters based on their scores in the different attributes**. The plot highlighted distinct separations

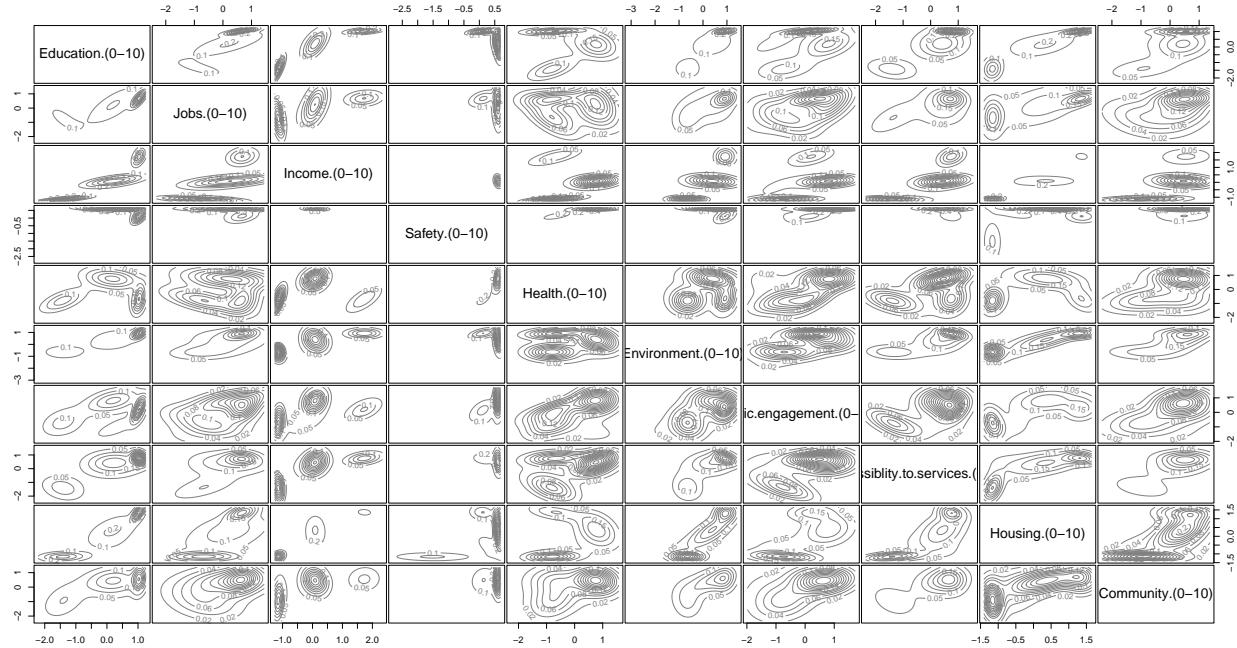
between clusters, emphasizing the significant variations in regional well-being. We can see some clear separation between clusters as for example in Education, Income and Safety, among others. This visualization effectively communicated how regions with similar well-being conditions are grouped together, providing a clear understanding of how specific attributes define life satisfaction levels.

### 5.4.3 Uncertainty



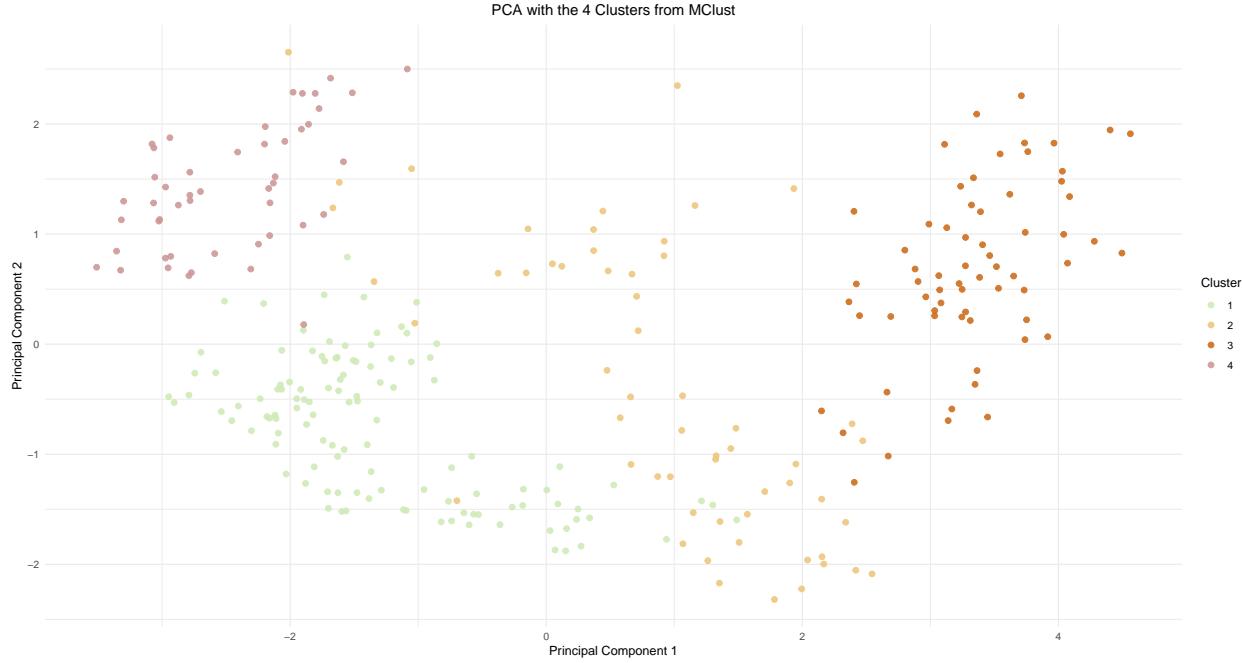
Furthermore, the uncertainty plot offered an additional layer of insight by visualizing the **degree of confidence in assigning data points to their respective clusters**. Most regions exhibited low uncertainty, indicating that the GMM model reliably classified regions into distinct clusters. However, a few regions displayed higher uncertainty, suggesting they share characteristics with multiple clusters.

#### 5.4.4 Density Distribution of Data Within Clusters

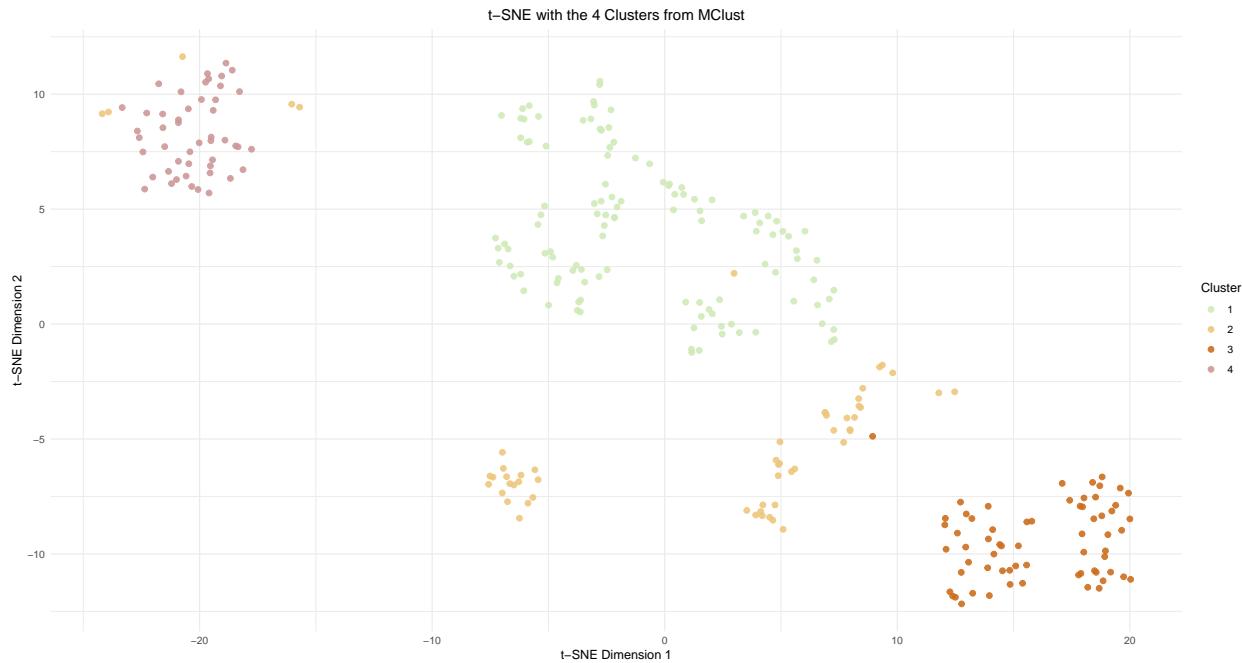


Last but not least, the density distribution plot for data within clusters provide an in-depth look at how key well-being attributes are distributed across the identified clusters. These plots enable a better understanding of the internal variability and the distinctiveness of each cluster. For example, metrics such as **education, income, and safety exhibit unique distributions within clusters**, highlighting the degree of homogeneity or disparity among regions in each grouping.

After this, we applied **Principal Component Analysis (PCA)** to *reduce the dataset's dimensionality*. This allowed us to plot the identified clusters in a two-dimensional space, providing a clear and intuitive representation of their separation and structure.



Furthermore, we employed **t-SNE (t-Distributed Stochastic Neighbor Embedding)** for a better visualization. This technique maps high-dimensional data into a lower-dimensional space while preserving the local structure (cluster), offering an alternative perspective on the clustering results.

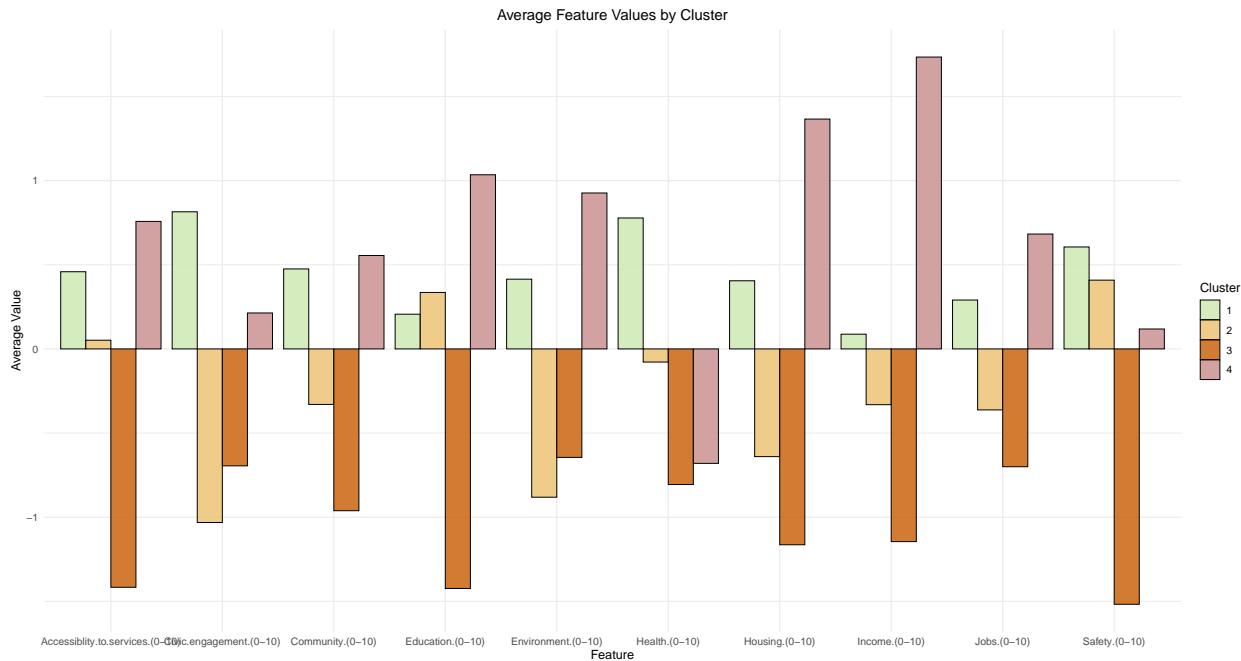


## 5.5 Analysis of Clusters

After obtaining the 4 clusters, we added the assigned clusters to the original dataframe. We then calculated the means for each cluster to better understand the characteristics of the data points within each group. This step helps identify patterns and differences between the clusters.

```
## # A tibble: 4 x 11
##   Cluster `Education.(0-10)` `Jobs.(0-10)` `Income.(0-10)` `Safety.(0-10)`
##   <fct>     <dbl>        <dbl>         <dbl>          <dbl>
## 1 1          0.206        0.290        0.0875        0.606
## 2 2          0.335       -0.362       -0.331        0.409
## 3 3         -1.42        -0.700       -1.14        -1.52 
## 4 4           1.03        0.682        1.73         0.118
## # i 6 more variables: `Health.(0-10)` <dbl>, `Environment.(0-10)` <dbl>,
## # `Civic.engagement.(0-10)` <dbl>, `Accessibility.to.services.(0-10)` <dbl>,
## # `Housing.(0-10)` <dbl>, `Community.(0-10)` <dbl>
```

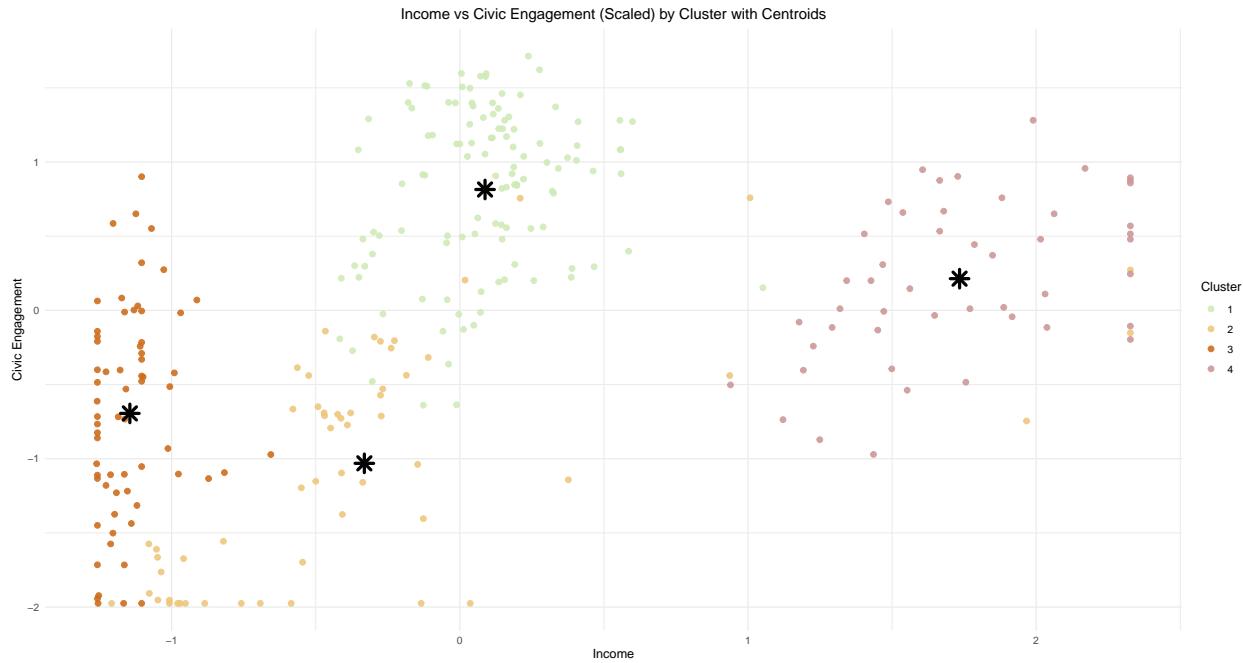
We used a bar plot to show the contribution of each feature in every cluster, with base reference to the average (mean) value of the feature. We can mainly see that for cluster 1, all the features have a positive value, more than the average value for all parameters.



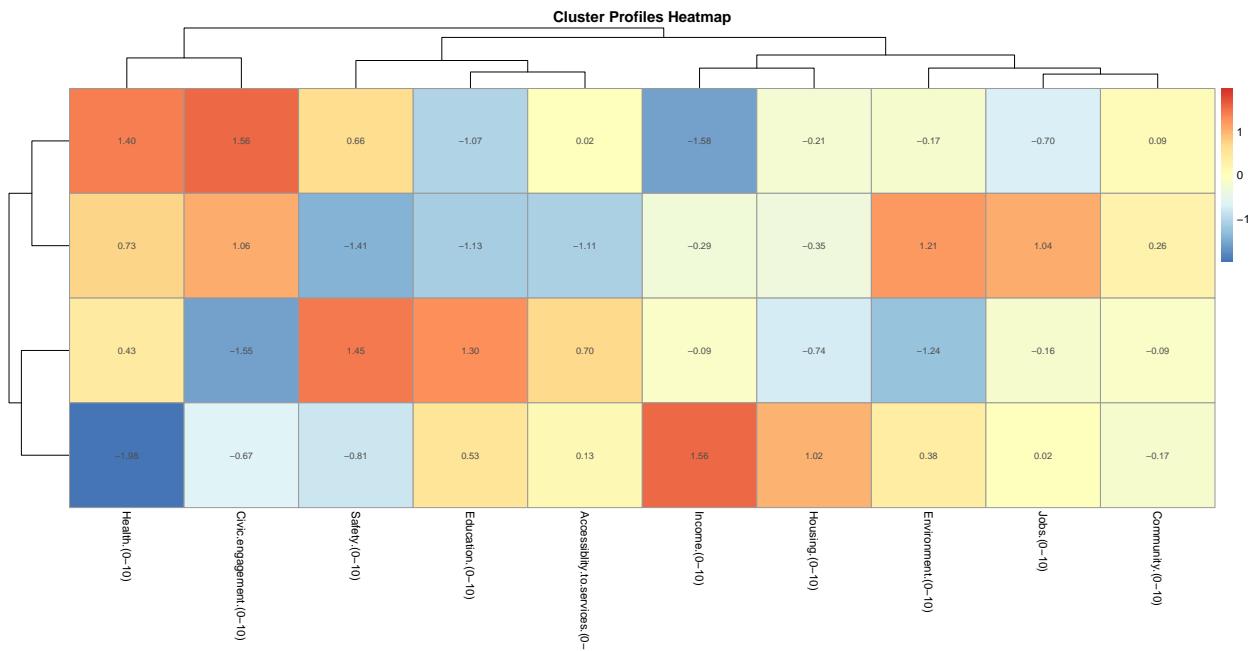
From this table we can clearly conclude that the data highlights a clear disparity in well-being across the clusters.:

- **Cluster 1** performs relatively well in safety (0.606).
- **Cluster 2** shows mixed results, including negative scores in jobs (-0.362) and income (-0.331).
- **Cluster 3** exhibits significantly lower scores across all attributes, with particularly negative values in safety (-1.52) and income (-1.14).
- **Cluster 4** consistently shows the highest scores in most attributes, such as income (1.73) and jobs (0.682).

Moreover, we generated a scatter plot of **Income vs Civic Engagement** for the four clusters, highlighting the distribution of each cluster based on these two variables.

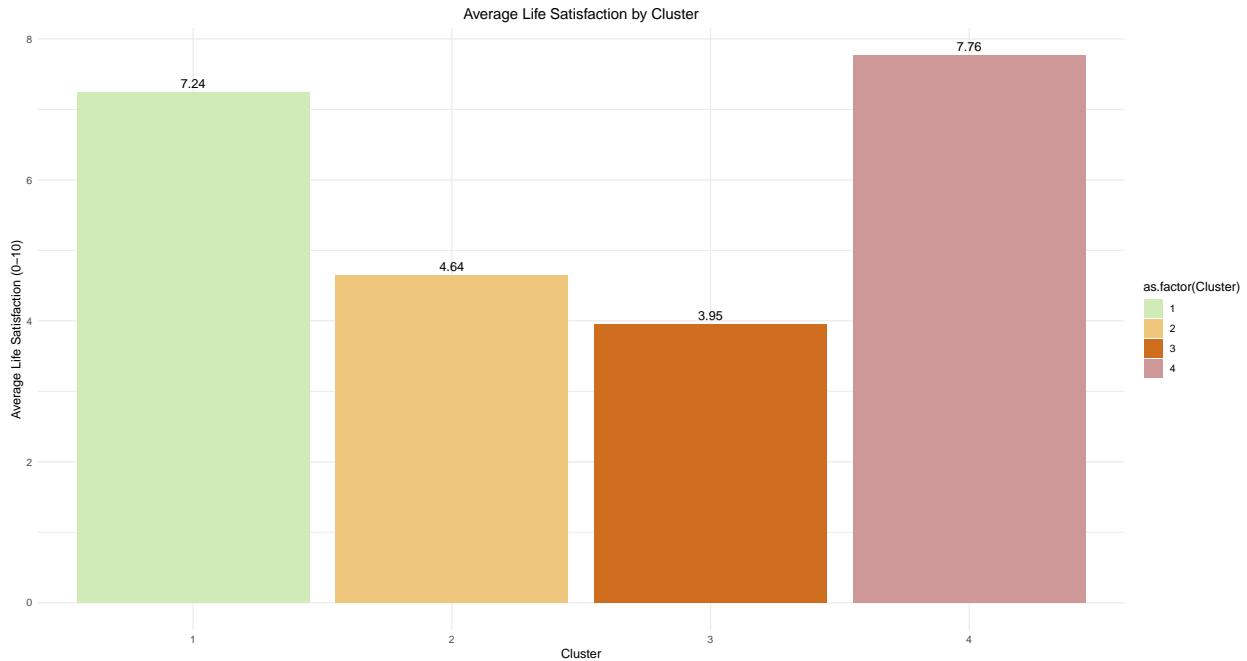


We also created a heatmap displaying the average values of factors in each cluster, with the data scaled to a baseline of 0. This plot is presented alongside the average plot to provide a clear comparison of the cluster characteristics.



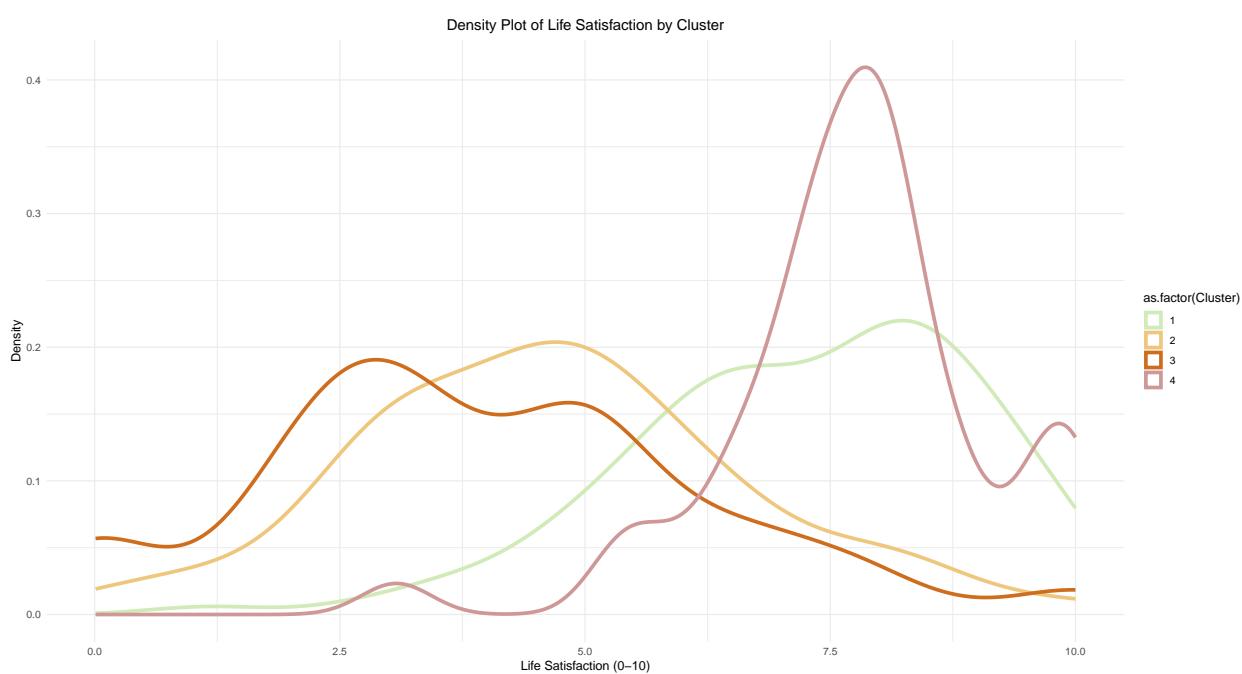
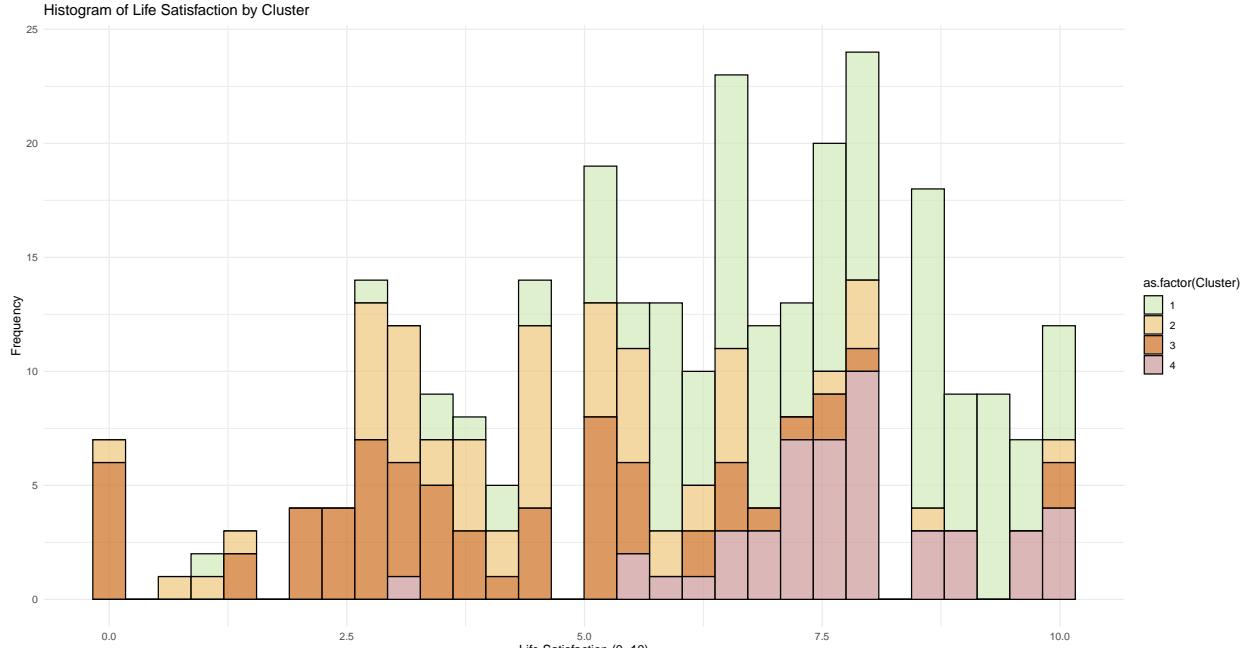
Later on, we went into more detail by plotting the **average Life Satisfaction Score** in each cluster.

```
## # A tibble: 4 x 2
##   Cluster Average_Life_Satisfaction
##   <dbl>                <dbl>
## 1 1                  7.24
## 2 2                  4.64
## 3 3                  3.95
## 4 4                  7.76
```



The **average life satisfaction scores** show that there is not a significant difference between the average values of Cluster 1 and Cluster 4, similarly with Cluster 2 and Cluster 3.

So finally, we created a histogram and a density plot for the four clusters to explore how **Life Satisfaction** behaves and its distribution within each cluster.



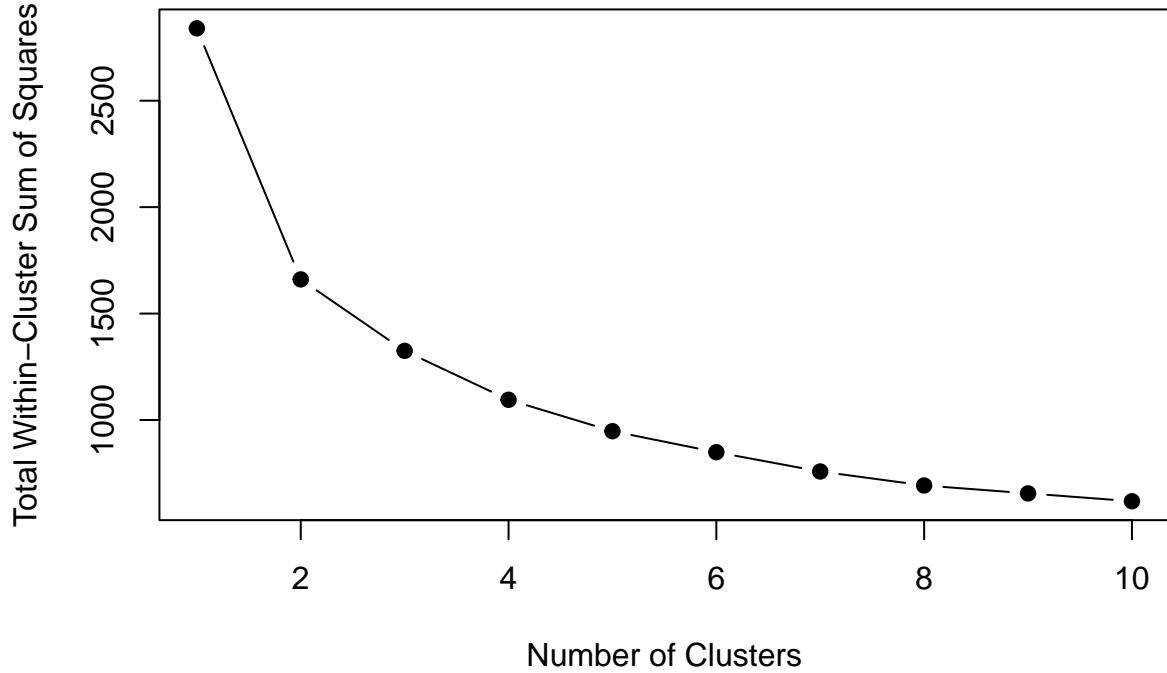
Looking at both the histogram and density plots, we could see that there was not a clear separation among the four clusters.

## 5.6 K-Means

Continuing with our clustering analysis, we decided to do a K-Means clustering to cross check and compare with the Mclust clusters.

When analysing the number of k clusters, we used the **Elbow Method**, and we concluded that there is a kink at  $k=2$ , but we move ahead to choose up to 4 clusters, after which the within-cluster sum of squares starts to fade off.

## Elbow Method for K-means



```
## Count of Cities in Each Cluster from K-Means
```

```
##  
##    1   2   3   4  
##  53  66 104  62
```

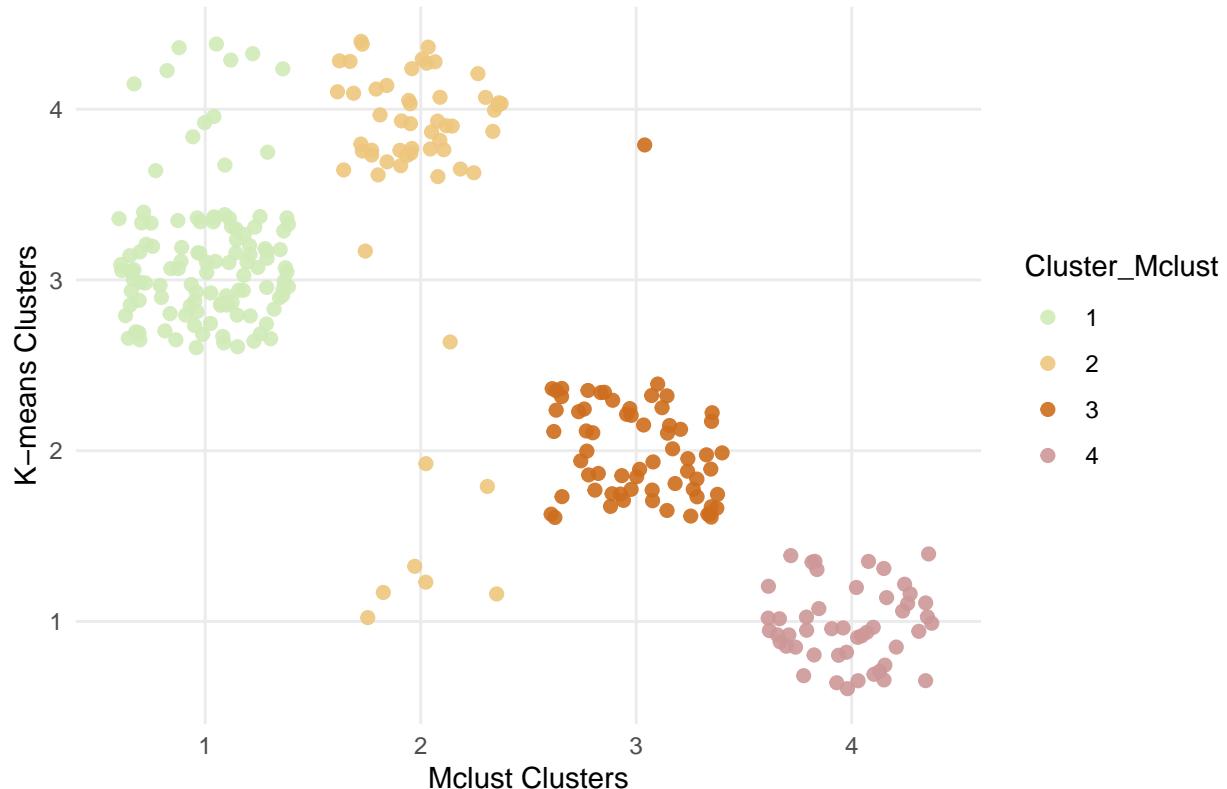
### 5.6.1 Comparing K-Means vs MCLUST

First, we did a **contingency table** to quickly compare both models:

```
##  
## Rows: Mclust Clusters  
## Columns: K-means Clusters  
  
##  
##          1   2   3   4 Sum  
##  1     0   0 102  13 115  
##  2     5   2   2  48  57  
##  3     0   64   0   1  65  
##  4    48   0   0   0  48  
##  Sum  53  66 104  62 285
```

Additionally, we created a scatter plot for comparing the clusters of both models.

## Comparison of Mclust and K–means Clustering



The contingency table and the scatter plot show that the clusters are approximately in line from both the models. So, we go ahead and check the Rand-Index.

The **Rand Index Comparison** is:

```
## [1] 0.807066
```

This also shows that the 4 clusters made by these two models are working in agreement with each other.

## 5.7 K-Medoids Clustering

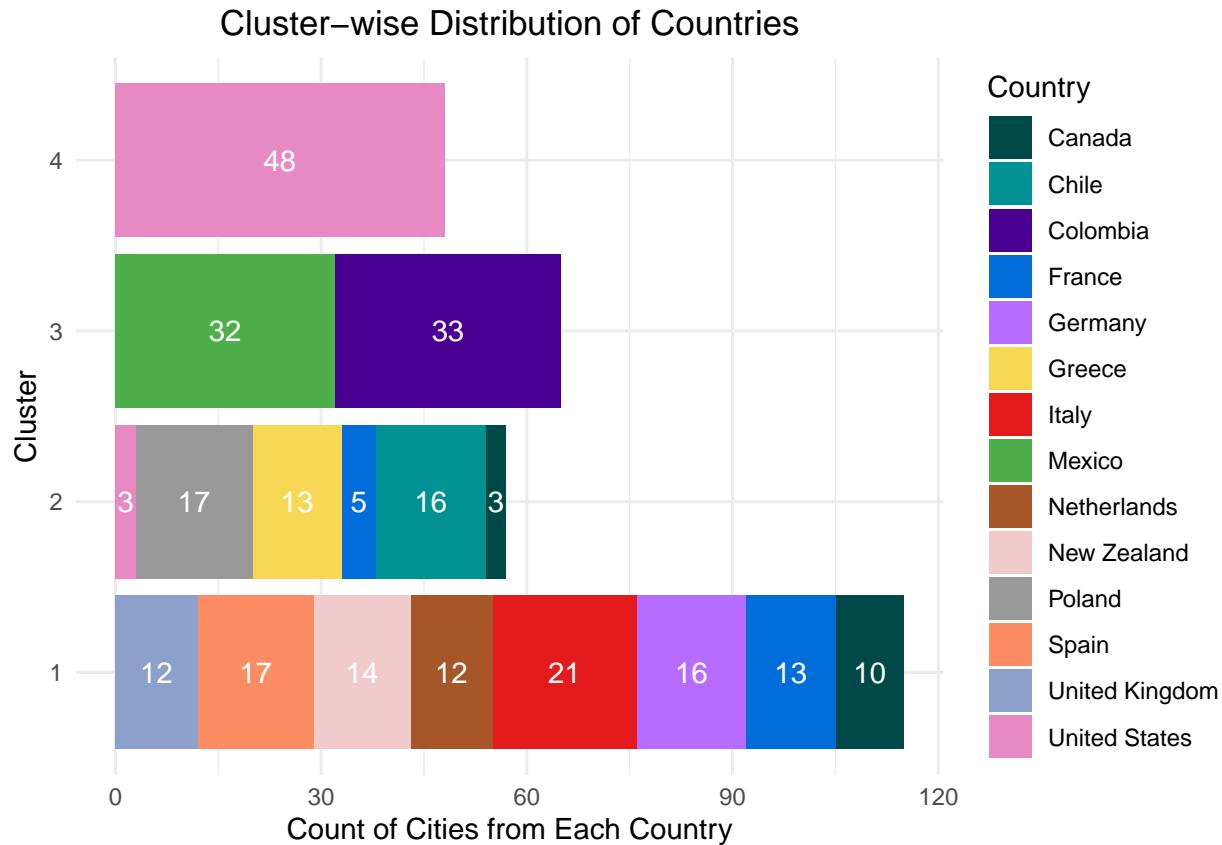
```
##  
##   1   2   3   4  
## 112  52  56  65
```



We see that there is a only a minor agreement between clusters made from K-Means and K-Medoids, which is due to the reason that K-Medoids is more robust, uses representative points and doesn't work on centroids of data.

Furthermore, we see a strong agreement between MClust and K-Means, which means data is probably in a structure suitable for both models. Also, K-Medoids is a distance based approach, while MClust is probabilistic approach.

Moreover, the cluster wise distribution shows that Cluster 4 is majorly composed of regions from USA only (with a higher Life Satisfaction Score), further cluster 3 includes Chile & Colombia.



## 5.8 Further Analysis for Clustering

After all the analysis done, we decided to further analyse **Life Satisfaction**, so we manually divided it in 4 clusters. We divided the range 0-10 into 4 groups:

1. 0-2.5
2. 2.5-5
3. 5-7.5
4. 7.5-10

```
##  
## 0-2.5 2.5-5 5-7.5 7.5-10  
## 21     81    84    99
```

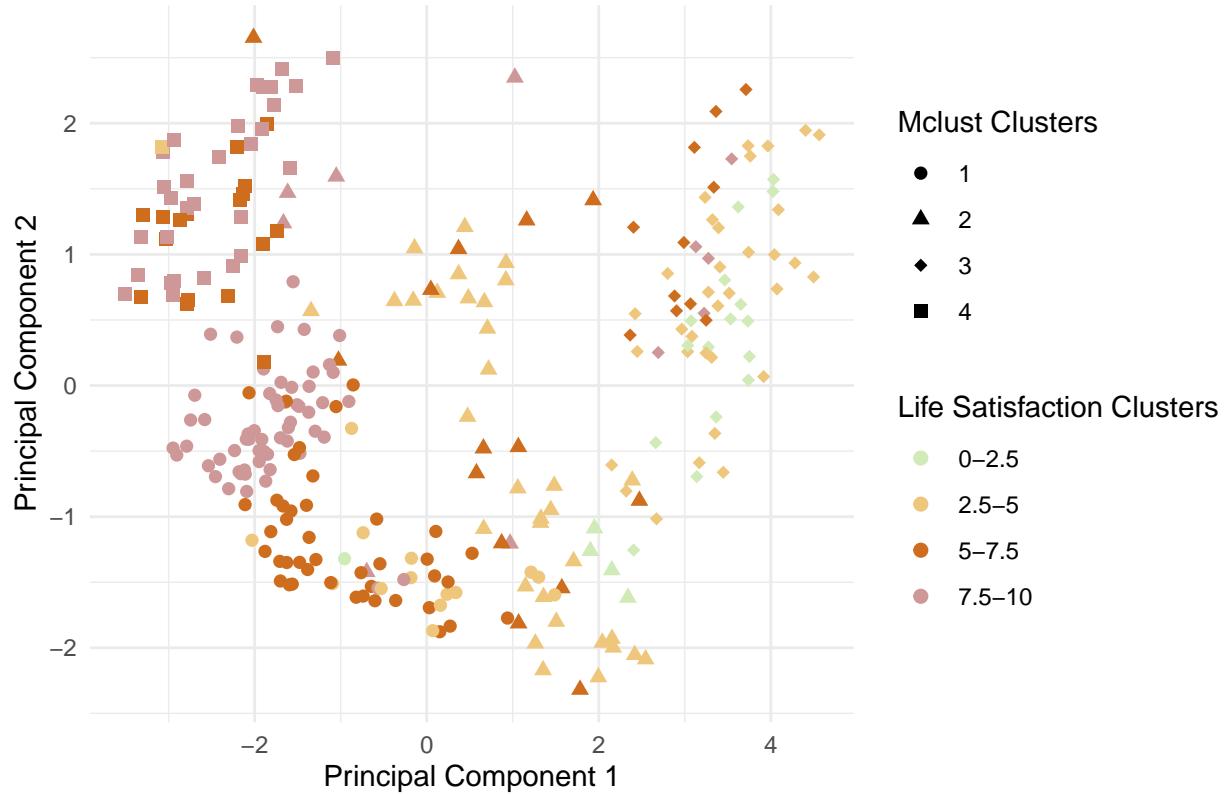
We compared this **manual clustering** vs the **M-Clust model** and the **K-Mean model**.

## 5.9 Manual clustering vs M-Clust model

```
##      mclust_clusters  
##      1 2 3 4  
## 0-2.5 1 4 16 0  
## 2.5-5 14 33 33 1  
## 5-7.5 42 14 11 17  
## 7.5-10 58 6 5 30
```

Additionally, we used a PCA visualization for further comparison between these two clustering methods.

### PCA Comparison of Life Satisfaction and Mclust Clusters

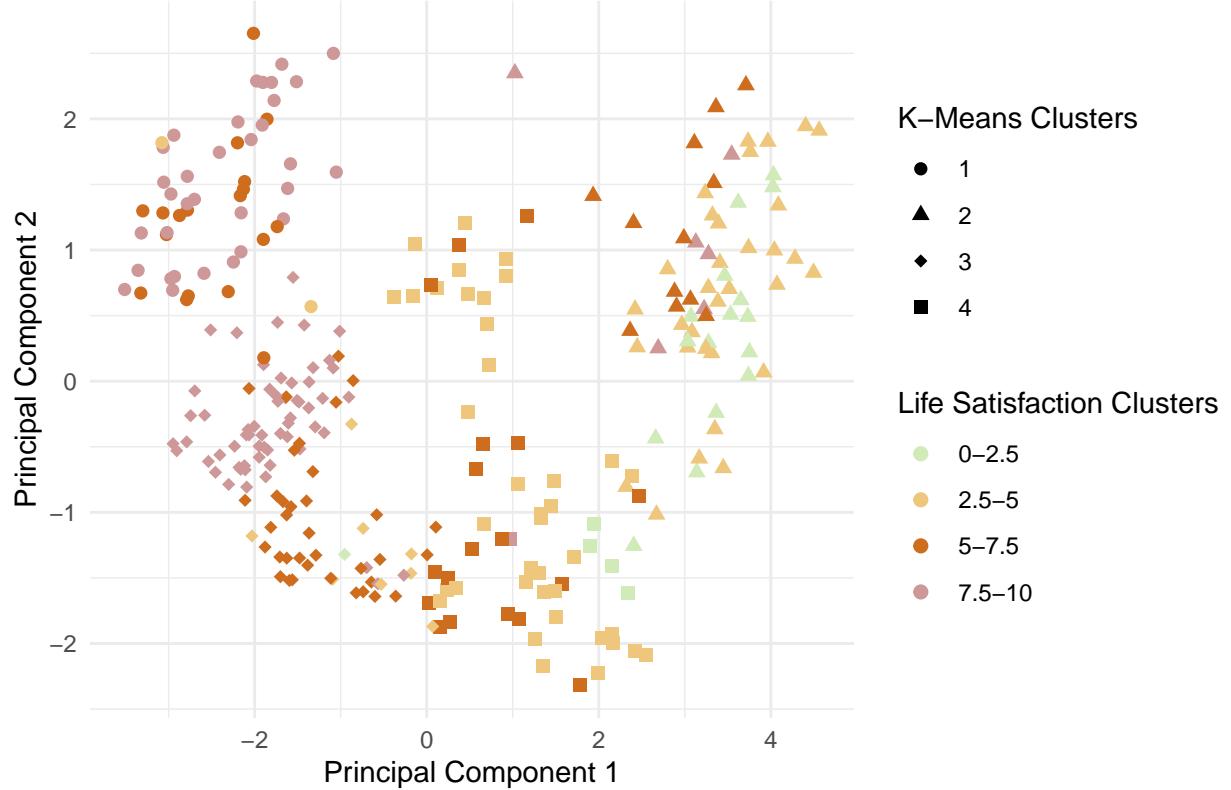


### 5.10 Manual clustering vs K-Means Cluster

```
##          1  2  3  4
## 0-2.5    0 16  1  4
## 2.5-5    2 32  8 39
## 5-7.5   18 12 36 18
## 7.5-10  33  6 59  1
```

Additionally we did used the PCA visualization for further comparison.

## PCA Comparison of Life Satisfaction and K-Means Clusters



The clusters divided manually **don't overlap** the clusters made by K-Means and M-Clust. This is probably because the scores are subjective, and different people measure life satisfaction scores in different perspectives. For some income would be a major factor, and for others community would be a major factor. However, statistically, we see that the **clustering using M-Clust and K-Means show a good overlap for the 4 clusters with a rand index 0.80**.

### 5.11 Conclusion of Clustering Approach

Finally, we compared the metrics:

```
## [1] "ARI between K-Means and K-Medoids: 0.89806810250644"
## [1] "ARI between K-Means and Mclust: 0.807065955471863"
## [1] "ARI between K-Medoids and Mclust: 0.868903212343953"
```

We see that there is only a minor agreement between clusters made from K-Means and K-Medoids, which is due to the reason that K-Medoids is more robust, uses representative points and doesn't work on centroids of data.

Further, we see a strong agreement between MClust and K-Means, which means data is probably in a structure suitable for both models.

Also, K-Medoids is a distance based approach, while MClust is probabilistic approach.

Hence, we conclude that the clusters formed by MClust and K-Means are better representation of the available data for Life Satisfaction Scores. However, for a better robust approach, more data and a detailed

analysis can be performed using further clustering techniques like DBSCAN and spectral based clustering etc.

## 6 Fitting the Regression Models

### 6.1 Model 1:Linear Regression

```
##  
## Call:  
## lm(formula = 'Life.satisfaction.(0-10)' ~ ., data = df_wb %>%  
##   select(-Country, -Region, -Code))  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -6.1996 -0.7953 -0.0019  0.8953  5.5480  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)                 0.360630  0.519000  0.695  0.48773  
## 'Education.(0-10)'          -0.005907  0.060633 -0.097  0.92246  
## 'Jobs.(0-10)'                0.348454  0.049178  7.086 1.17e-11 ***  
## 'Income.(0-10)'             -0.104160  0.068147 -1.528  0.12755  
## 'Safety.(0-10)'              -0.059370  0.049777 -1.193  0.23401  
## 'Health.(0-10)'               0.065423  0.052871  1.237  0.21699  
## 'Environment.(0-10)'         0.214532  0.067339  3.186  0.00161 **  
## 'Civic.engagement.(0-10)'    -0.051276  0.053239 -0.963  0.33633  
## 'Accessibility.to.services.(0-10)' 0.070444  0.063262  1.114  0.26645  
## 'Housing.(0-10)'              0.222854  0.067226  3.315  0.00104 **  
## 'Community.(0-10)'            0.220590  0.043667  5.052 8.01e-07 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 1.542 on 274 degrees of freedom  
## Multiple R-squared:  0.613, Adjusted R-squared:  0.5989  
## F-statistic: 43.41 on 10 and 274 DF,  p-value: < 2.2e-16
```

This regression model explains approximately **61.3%** of the variation in life satisfaction\* (adjusted R-squared = 59.9%). Significant predictors include **Jobs (0-10)**, **Environment (0-10)**, **Housing (0-10)**, and **Community (0-10)**, all positively associated with life satisfaction, as indicated by their significant p-values ( $p < 0.01$ ). Among these, **Jobs (0-10)** has the strongest effect (Estimate = 0.348). Other variables such as **Education**, **Income**, and **Safety** do not show significant associations with life satisfaction at the 5% significance level. The residual standard error of **1.542** suggests moderate variability not explained by the model.

Overall, the model identifies key factors that contribute to life satisfaction but suggests some room for improvement in capturing other influences.

### 6.2 Model 2: Linear Regression using Pairwise Interaction effects

The second model (**Anex B**) includes interaction terms and explains **79.5% of the variation in life satisfaction** (**adjusted R-squared = 74.6%**), showing a substantial improvement in explanatory power

compared to the first model. Significant main effects include **Safety (0-10)**, **Environment (0-10)**, **Civic engagement (0-10)**, and **Community (0-10)**, among others.

Several interaction terms are also significant, such as **Education x Safety**, **Jobs x Income**, and **Safety x Accessibility to Services**, indicating that the relationship between life satisfaction and these predictors depends on other variables. The residual standard error is reduced to 1.227, suggesting better fit and less unexplained variability. However, some predictors and interactions remain insignificant, suggesting that not all modeled interactions are meaningful. This model highlights the complex, interconnected nature of factors influencing life satisfaction.

### 6.3 Model 3: Model with Interaction effects based on relevance of factors.

```
##  
## Call:  
## lm(formula = 'Life.satisfaction.(0-10)' ~ 'Education.(0-10)' +  
##   'Jobs.(0-10)' + 'Income.(0-10)' + 'Safety.(0-10)' + 'Health.(0-10)' +  
##   'Environment.(0-10)' + 'Civic.engagement.(0-10)' + 'Accessibility.to.services.(0-10)' +  
##   'Housing.(0-10)' + 'Community.(0-10)' + 'Education.(0-10)':'Safety.(0-10)' +  
##   'Education.(0-10)':'Accessibility.to.services.(0-10)' + 'Education.(0-10)':'Housing.(0-10)' +  
##   'Jobs.(0-10)':'Income.(0-10)' + 'Jobs.(0-10)':'Safety.(0-10)' +  
##   'Jobs.(0-10)':'Civic.engagement.(0-10)' + 'Jobs.(0-10)':'Accessibility.to.services.(0-10)' +  
##   'Jobs.(0-10)':'Housing.(0-10)' + 'Safety.(0-10)':'Environment.(0-10)' +  
##   'Safety.(0-10)':'Civic.engagement.(0-10)' + 'Safety.(0-10)':'Accessibility.to.services.(0-10)' +  
##   'Safety.(0-10)':'Housing.(0-10)' + 'Accessibility.to.services.(0-10)':'Community.(0-10)',  
##   data = df_wb %>% select(-Country, -Region, -Code))  
##  
## Residuals:  
##       Min     1Q Median     3Q    Max  
## -4.8669 -0.6374  0.0548  0.6978  3.8055  
##  
## Coefficients:  
##                               Estimate Std. Error  
## (Intercept)                   8.6522277  1.8280380  
## 'Education.(0-10)'            0.5587564  0.2485484  
## 'Jobs.(0-10)'                0.0141515  0.1904661  
## 'Income.(0-10)'              0.7883820  0.2409213  
## 'Safety.(0-10)'              -1.0434919  0.1971948  
## 'Health.(0-10)'              -0.0336778  0.0496355  
## 'Environment.(0-10)'          -0.8703014  0.2752127  
## 'Civic.engagement.(0-10)'      -1.2828470  0.2150050  
## 'Accessibility.to.services.(0-10)' -0.0449081  0.3138580  
## 'Housing.(0-10)'              0.5295530  0.3460838  
## 'Community.(0-10)'             0.5235976  0.0745148  
## 'Education.(0-10)':'Safety.(0-10)' -0.0545734  0.0255887  
## 'Education.(0-10)':'Accessibility.to.services.(0-10)' -0.0605636  0.0277395  
## 'Education.(0-10)':'Housing.(0-10)'           0.0458439  0.0213117  
## 'Jobs.(0-10)':'Income.(0-10)'           -0.1198992  0.0317219  
## 'Jobs.(0-10)':'Safety.(0-10)'            0.0004871  0.0220756  
## 'Jobs.(0-10)':'Civic.engagement.(0-10)'      0.0743479  0.0176163  
## 'Jobs.(0-10)':'Accessibility.to.services.(0-10)' 0.0258404  0.0238409  
## 'Jobs.(0-10)':'Housing.(0-10)'           0.0485973  0.0216787  
## 'Safety.(0-10)':'Environment.(0-10)'        0.1220980  0.0299240  
## 'Safety.(0-10)':'Civic.engagement.(0-10)'      0.0946893  0.0232325  
## 'Safety.(0-10)':'Accessibility.to.services.(0-10)' 0.0943920  0.0320156
```

```

## 'Safety.(0-10)':'Housing.(0-10)' -0.1184011 0.0351105
## 'Accessibility.to.services.(0-10)':'Community.(0-10)' -0.0670851 0.0124325
## t value Pr(>|t|)
## (Intercept) 4.733 3.63e-06 ***
## 'Education.(0-10)' 2.248 0.025406 *
## 'Jobs.(0-10)' 0.074 0.940829
## 'Income.(0-10)' 3.272 0.001210 **
## 'Safety.(0-10)' -5.292 2.57e-07 ***
## 'Health.(0-10)' -0.679 0.498054
## 'Environment.(0-10)' -3.162 0.001750 **
## 'Civic.engagement.(0-10)' -5.967 7.86e-09 ***
## 'Accessibility.to.services.(0-10)' -0.143 0.886334
## 'Housing.(0-10)' 1.530 0.127196
## 'Community.(0-10)' 7.027 1.84e-11 ***
## 'Education.(0-10)':'Safety.(0-10)' -2.133 0.033882 *
## 'Education.(0-10)':'Accessibility.to.services.(0-10)' -2.183 0.029905 *
## 'Education.(0-10)':'Housing.(0-10)' 2.151 0.032386 *
## 'Jobs.(0-10)':'Income.(0-10)' -3.780 0.000195 ***
## 'Jobs.(0-10)':'Safety.(0-10)' 0.022 0.982413
## 'Jobs.(0-10)':'Civic.engagement.(0-10)' 4.220 3.37e-05 ***
## 'Jobs.(0-10)':'Accessibility.to.services.(0-10)' 1.084 0.279424
## 'Jobs.(0-10)':'Housing.(0-10)' 2.242 0.025821 *
## 'Safety.(0-10)':'Environment.(0-10)' 4.080 5.98e-05 ***
## 'Safety.(0-10)':'Civic.engagement.(0-10)' 4.076 6.09e-05 ***
## 'Safety.(0-10)':'Accessibility.to.services.(0-10)' 2.948 0.003485 **
## 'Safety.(0-10)':'Housing.(0-10)' -3.372 0.000859 ***
## 'Accessibility.to.services.(0-10)':'Community.(0-10)' -5.396 1.53e-07 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.261 on 261 degrees of freedom
## Multiple R-squared: 0.7533, Adjusted R-squared: 0.7316
## F-statistic: 34.66 on 23 and 261 DF, p-value: < 2.2e-16

```

The regression model incorporates main effects and interaction terms to explain life satisfaction, achieving an adjusted **R-squared of 0.7316**. Key positive predictors include income, community, and jobs interacting with civic engagement, while significant negative effects are observed for safety, civic engagement, and environment. Notable interactions, such as education with accessibility to services and safety with housing, reveal nuanced relationships where factors amplify or mitigate one another's influence. The significant interactions highlight the complexity of factors contributing to life satisfaction, suggesting the need for targeted, context-sensitive policy interventions.

## 6.4 Parameters from the Three Models

### 6.4.1 Adjusted R-square

```

## [1] "Model 1: 0.598909275658412"

## [1] "Model 2: 0.745820074260208"

## [1] "Model 3: 0.731599525077232"

```

Adjusted R<sup>2</sup> measures how well the model explains the variance in the dependent variable while adjusting for the number of predictors. **Model 2** performs best with the highest adjusted R<sup>2</sup> of **0.7458**, showing it explains the most variance. **Model 3** slightly trails behind with **0.7316**, but it likely avoids overfitting by focusing on fewer relevant interaction terms.

#### 6.4.2 AIC

```
## [1] "Model 1: 1068.27959661362"  
  
## [1] "Model 2: 977.147457732896"  
  
## [1] "Model 3: 965.939772745139"
```

AIC penalizes model complexity and evaluates the goodness of fit. Lower values are better. **Model 3** has the lowest AIC (**965.94**), indicating it balances fit and complexity better than the others. **Model 2** has a higher AIC (**977.15**) than Model 3, suggesting it may be slightly more complex or less efficient in terms of fit.

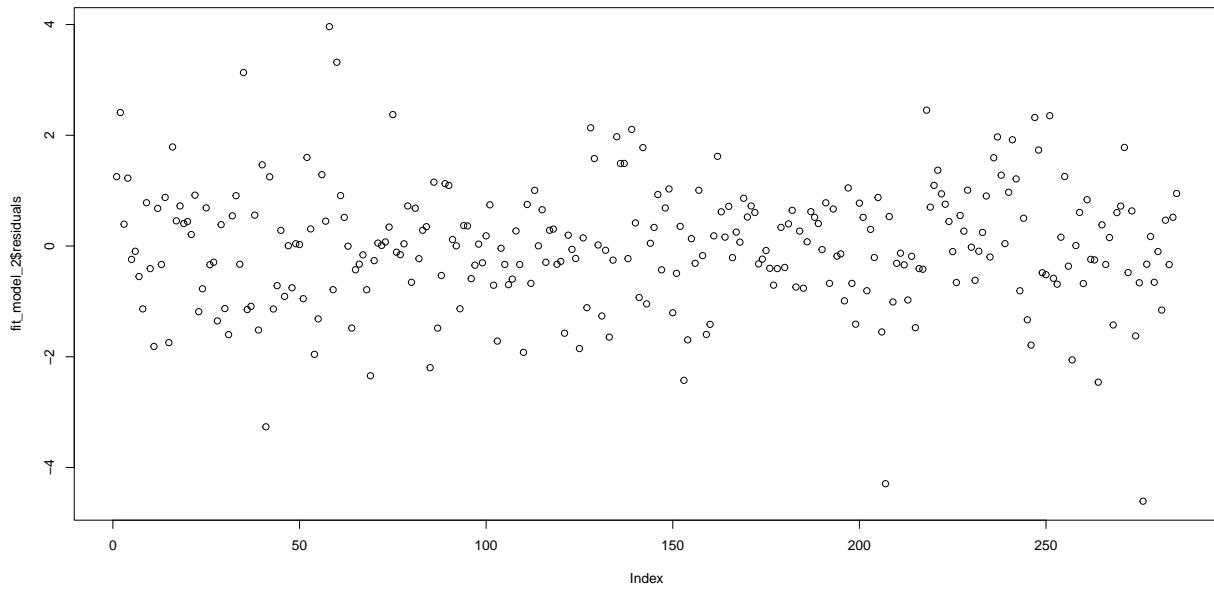
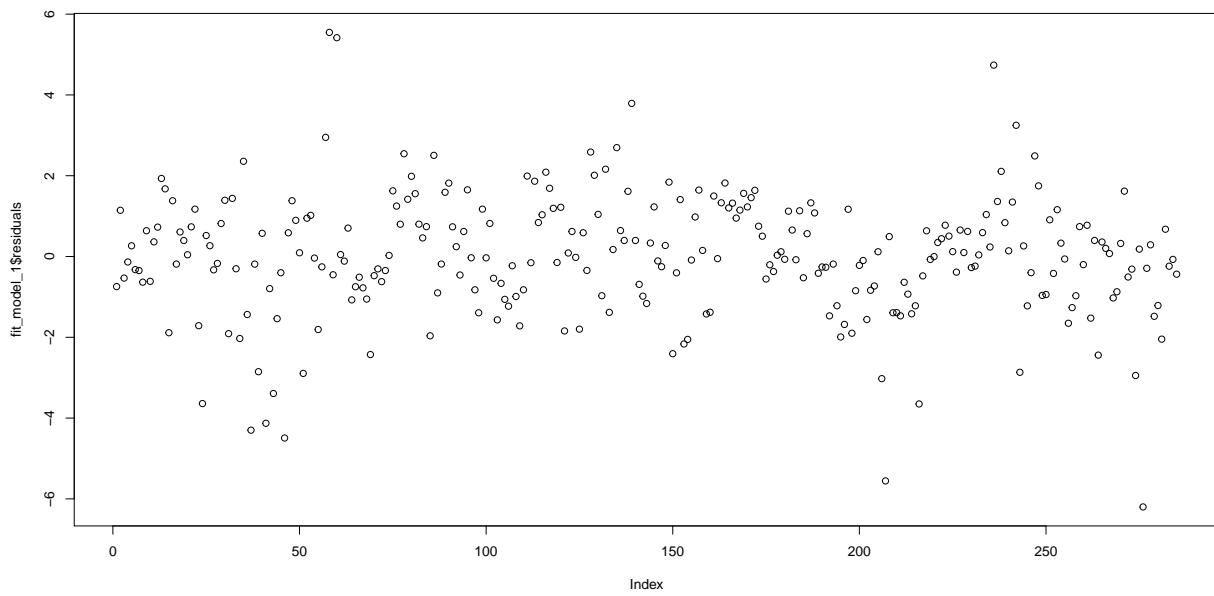
#### 6.4.3 BIC

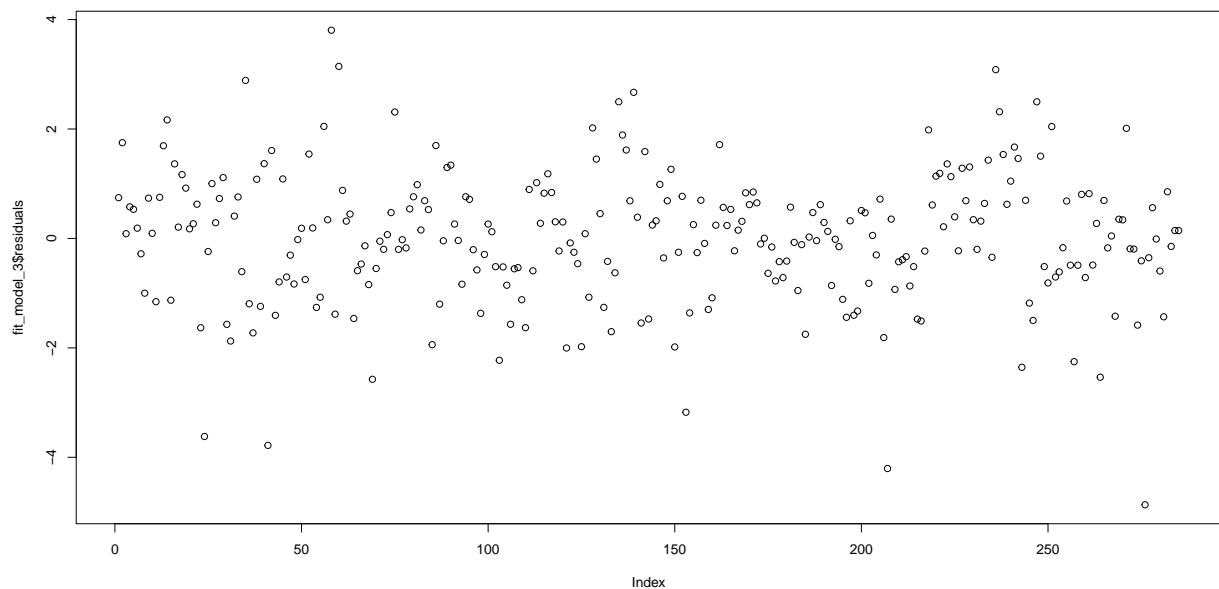
```
## [1] "Model 1: 1112.10946677685"  
  
## [1] "Model 2: 1185.33934100821"  
  
## [1] "Model 3: 1057.25200225186"
```

BIC penalizes complexity more heavily than AIC. Lower values indicate better performance. **Model 3** again has the lowest BIC (**1057.25**), outperforming the others in terms of simplicity and fit. **Model 2** has the highest BIC (**1185.34**), suggesting over-complexity compared to Model 1 and Model 3.

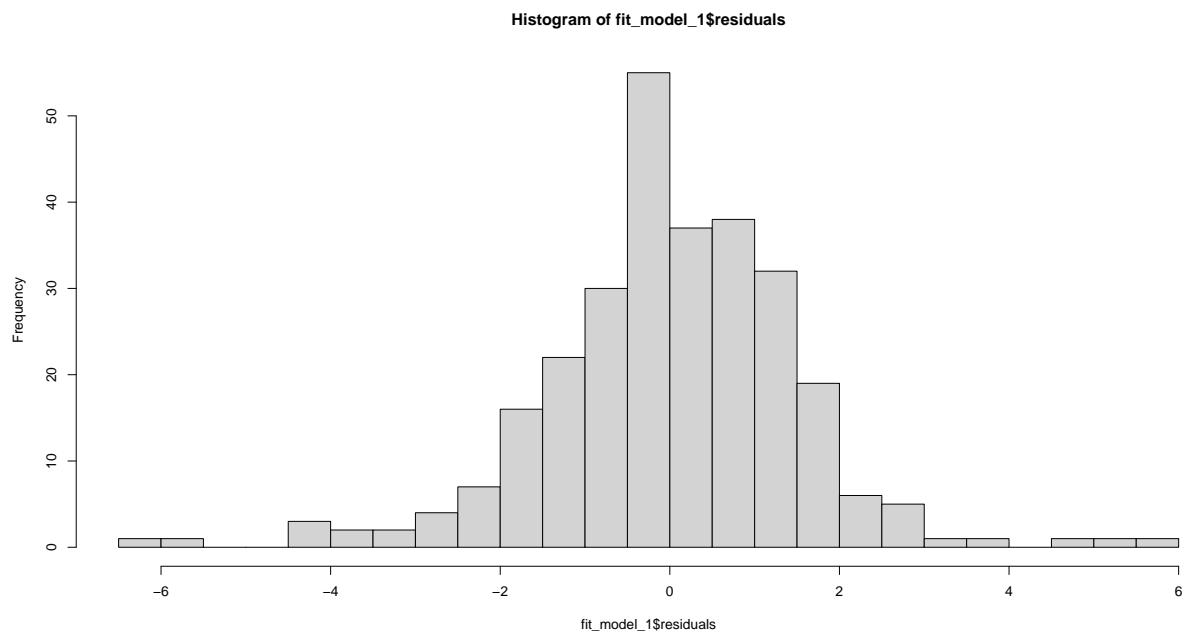
#### 6.4.4 Residuals

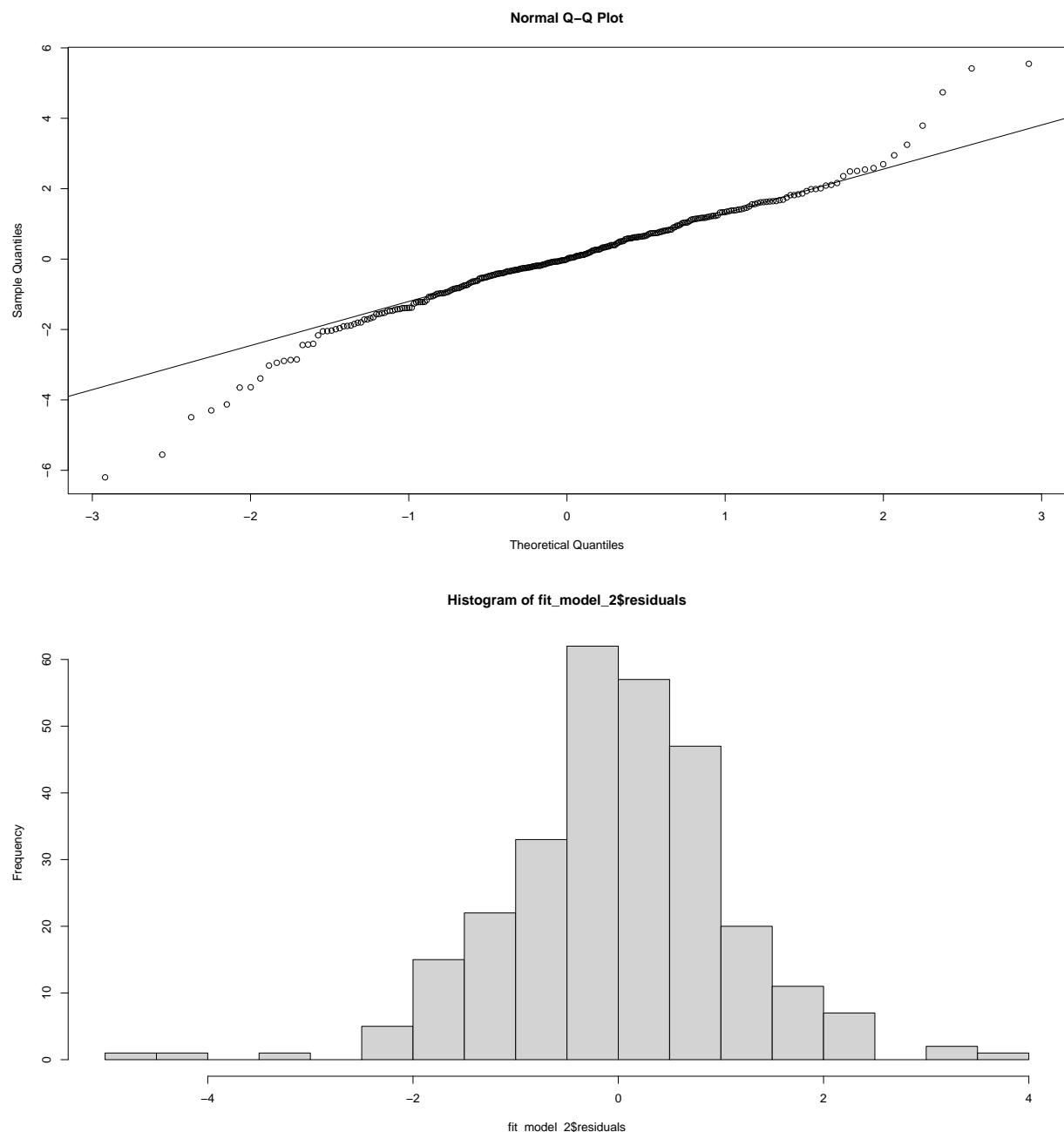
Lower RSE implies better model performance. **Model 3** has the lowest RSE (**1.261**), indicating it predicts life satisfaction with less error compared to simpler models.

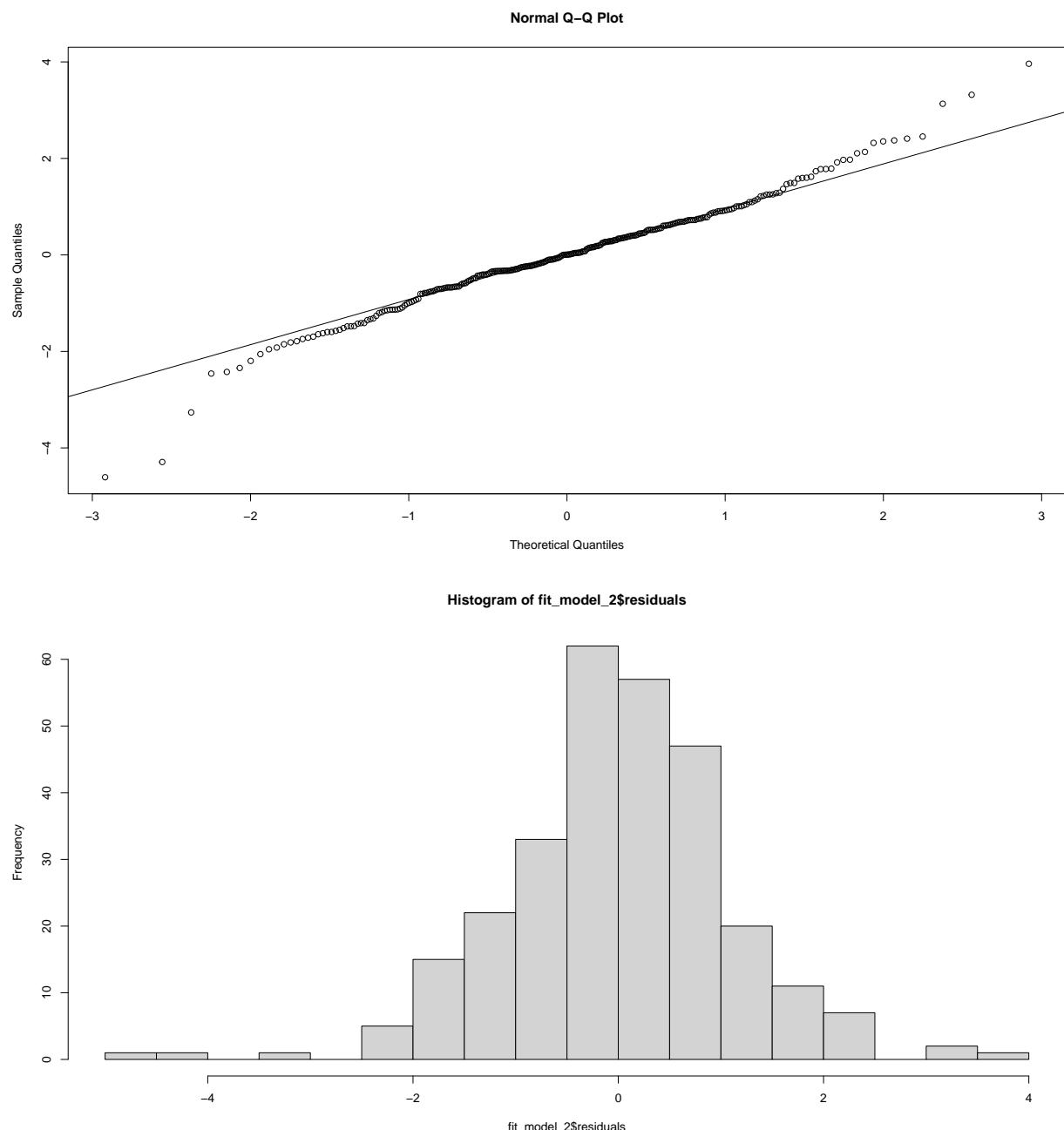


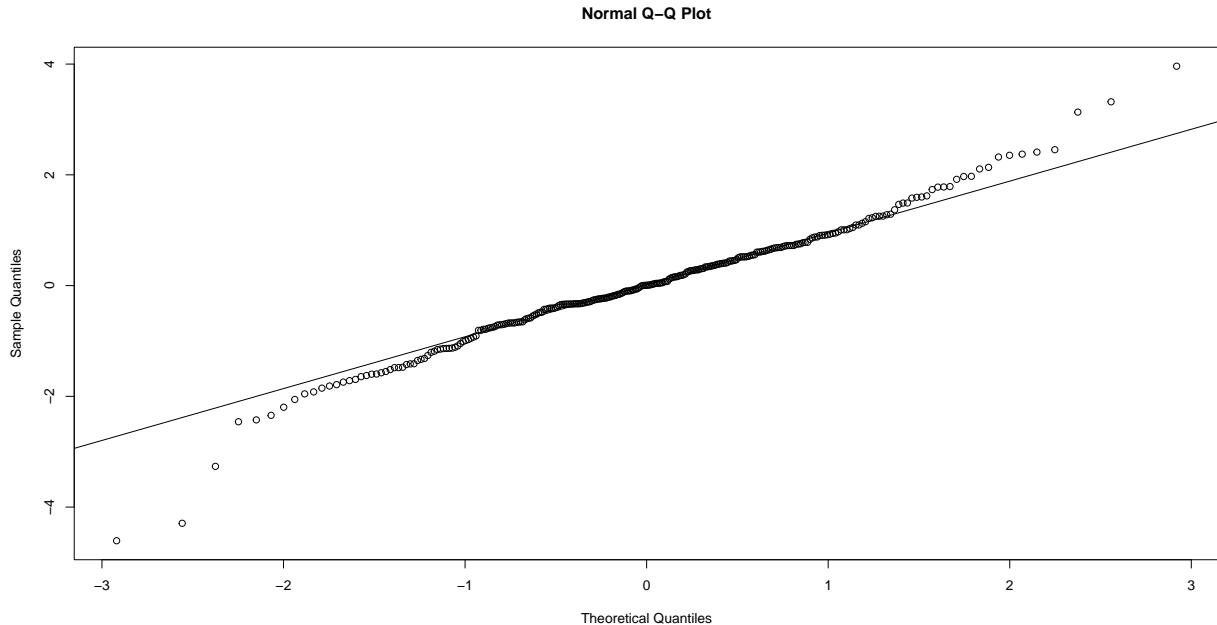


#### 6.4.5 Normality using Q-Q Plots









Conclusion:

- **Model 3** strikes the best balance between goodness of fit (AIC/BIC) and prediction error (RSE).
- **Model 2** has the highest adjusted R<sup>2</sup>, meaning it explains more variance but at the cost of complexity (higher BIC).
- **Model 1** is the simplest but performs worst in all metrics, making it the least suitable.

#### 6.4.6 ANOVA

We used ANOVA to check which model is better from Model 1 and Model 3:

```
## Analysis of Variance Table
##
## Model 1: 'Life.satisfaction.(0-10)' ~ 'Education.(0-10)' + 'Jobs.(0-10)' +
##           'Income.(0-10)' + 'Safety.(0-10)' + 'Health.(0-10)' + 'Environment.(0-10)' +
##           'Civic.engagement.(0-10)' + 'Accessibility.to.services.(0-10)' +
##           'Housing.(0-10)' + 'Community.(0-10)'
## Model 2: 'Life.satisfaction.(0-10)' ~ 'Education.(0-10)' + 'Jobs.(0-10)' +
##           'Income.(0-10)' + 'Safety.(0-10)' + 'Health.(0-10)' + 'Environment.(0-10)' +
##           'Civic.engagement.(0-10)' + 'Accessibility.to.services.(0-10)' +
##           'Housing.(0-10)' + 'Community.(0-10)' + 'Education.(0-10)':'Safety.(0-10)' +
##           'Education.(0-10)':'Accessibility.to.services.(0-10)' + 'Education.(0-10)':'Housing.(0-10)' +
##           'Jobs.(0-10)':'Income.(0-10)' + 'Jobs.(0-10)':'Safety.(0-10)' +
##           'Jobs.(0-10)':'Civic.engagement.(0-10)' + 'Jobs.(0-10)':'Accessibility.to.services.(0-10)' +
##           'Jobs.(0-10)':'Housing.(0-10)' + 'Safety.(0-10)':'Environment.(0-10)' +
##           'Safety.(0-10)':'Civic.engagement.(0-10)' + 'Safety.(0-10)':'Accessibility.to.services.(0-10)' +
##           'Safety.(0-10)':'Housing.(0-10)' + 'Accessibility.to.services.(0-10)':'Community.(0-10)'
##   Res.Df   RSS Df Sum of Sq    F   Pr(>F)
## 1     274  651.16
## 2     261 415.07 13    236.09 11.42 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
```

The Analysis of Variance (ANOVA) test compares Model 1 and Model 3 to assess whether the inclusion of interaction effects in Model 3 significantly improves the model's fit. **Model 3 has fewer residual degrees of freedom** (261 vs. 274) due to the inclusion of 13 additional interaction terms, which increases the number of estimated parameters. **The Residual Sum of Squares (RSS) is considerably lower in Model 3** (415.07 vs. 651.16), indicating that it explains more variability in life satisfaction.

The difference in RSS between the models is **236.09**, demonstrating that the interaction terms in Model 3 significantly enhance the model's performance. The F-statistic (11.42) and the highly significant p-value ( $p < 2.2e-16$ ) confirm that the inclusion of these interaction effects leads to a statistically significant improvement in model fit. While Model 3 captures the relationships in the data better, this improvement comes with increased complexity due to the additional interaction terms.

## 6.5 Random Intercept Model

Since the data has hierarchical structure, we fitted the random intercept model:

```
## Linear mixed model fit by REML [lmerMod]
## Formula: 'Life.satisfaction.(0-10)' ~ 'Education.(0-10)' + 'Jobs.(0-10)' +
##           'Income.(0-10)' + 'Safety.(0-10)' + 'Health.(0-10)' + 'Environment.(0-10)' +
##           'Civic.engagement.(0-10)' + 'Accessibility.to.services.(0-10)' +
##           'Housing.(0-10)' + 'Community.(0-10)' + (1 | Country)
## Data: df_wb %>% select(-Region, -Code)
##
## REML criterion at convergence: 1056.4
##
## Scaled residuals:
##      Min     1Q Median     3Q    Max
## -4.0049 -0.3837  0.0287  0.4401  4.1670
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   Country (Intercept) 1.760     1.327
##   Residual            1.903     1.380
## Number of obs: 285, groups: Country, 14
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                1.99349  0.90626  2.200
## 'Education.(0-10)'         0.24477  0.10371  2.360
## 'Jobs.(0-10)'              0.09457  0.07077  1.336
## 'Income.(0-10)'            -0.01267 0.12532 -0.101
## 'Safety.(0-10)'            -0.08039 0.05615 -1.432
## 'Health.(0-10)'             -0.01965 0.08048 -0.244
## 'Environment.(0-10)'        0.16097  0.07002  2.299
## 'Civic.engagement.(0-10)'   -0.13851 0.07074 -1.958
## 'Accessibility.to.services.(0-10)' 0.11535 0.08865  1.301
## 'Housing.(0-10)'            0.10318  0.08786  1.174
## 'Community.(0-10)'          0.19783  0.04225  4.682
##
## Correlation of Fixed Effects:
##          (Intr) 'Ed.(0-10)' 'J.(0- 'I.(0- 'S.(0- 'Hl.(0-10)' 'En.(0-10)'
## 'Ed.(0-10)' -0.136
## 'Jb.(0-10)' -0.145 -0.241
## 'In.(0-10)'  0.026 -0.290      -0.187
```

```

## 'Sf.(0-10)' -0.385 -0.148      0.085  0.033
## 'Hl.(0-10)' -0.107  0.058      0.004 -0.206 -0.189
## 'En.(0-10)' -0.444 -0.030     -0.058 -0.063  0.039 -0.027
## 'C..(0-10)'  0.078  0.031     -0.333 -0.060 -0.169 -0.282      0.111
## 'A...(0-10)' -0.214 -0.279      0.075 -0.075  0.063 -0.377      0.076
## 'Hs.(0-10)'  0.003 -0.284     -0.022  0.081 -0.014  0.139     -0.227
## 'Cm.(0-10)'  -0.258 -0.080     0.001 -0.003  0.072  0.004      0.005
##           'C..(0 'A...('Hs.(0-10)'
## 'Ed.(0-10)'
## 'Jb.(0-10)'
## 'In.(0-10)'
## 'Sf.(0-10)'
## 'Hl.(0-10)'
## 'En.(0-10)'
## 'C..(0-10)'
## 'A...(0-10) -0.028
## 'Hs.(0-10)' -0.222 -0.139
## 'Cm.(0-10)' -0.034  0.073 -0.129

```

**The Random Intercept Model** was fitted to account for the hierarchical structure in the data, with random intercepts for each country. This model is useful when we expect that the baseline life satisfaction differs across countries, but the relationships between predictors and life satisfaction remain consistent within each country.

The model's random effects show that the variance in life satisfaction between countries (Intercept) is 1.760 with a standard deviation of 1.327. The residual variance (within-country variability) is 1.903 with a standard deviation of 1.380. These results suggest that there is a moderate degree of variability both between countries and within countries.

For the fixed effects, the significant predictors of life satisfaction include **Education** (positive relationship, estimate = 0.24477, p-value < 0.05), **Environment** (positive relationship, estimate = 0.16097, p-value < 0.05), and **Community** (positive relationship, estimate = 0.19783, p-value < 0.001). Other variables, such as **Jobs**, **Income**, **Safety**, **Health**, and **Civic engagement**, did not show significant effects at the conventional 5% level. These results suggest that education, environmental factors, and community engagement have notable impacts on life satisfaction, while income, safety, and health appear to have a more limited influence when accounting for the hierarchical structure.

- **Significant Effects:** Education, Environment, Community, and Civic engagement.
- **Non-significant Effects:** Jobs, Income, Safety, Health, Housing, and Accessibility to services.
- **REML Criterion:** 1131.8. Measure of model fit for restricted maximum likelihood (REML) estimation. Lower values of the REML criterion indicate a better-fitting model.

The **AIC (Akaike Information Criterion)** for this model is **1082.41**, which is a measure of the model's goodness of fit while penalizing for complexity (number of parameters). A lower AIC indicates a better-fitting model when compared to other models. This value suggests that the model balances fit and complexity reasonably well.

The **BIC (Bayesian Information Criterion)** is **1129.892**, which is similar to AIC but applies a stronger penalty for the number of parameters, especially as the number of observations increases. A lower BIC value would generally be preferred for selecting simpler models that fit the data adequately, but in this case, the relatively high BIC indicates that the model may be somewhat complex.

Based on the AIC and BIC, **Model 3** is the most efficient and well-fitting model. **Model 4** (the random intercept model) does not outperform the simpler models in terms of AIC and BIC, though it accounts for the hierarchical structure of the data.

## 6.6 Fitting the Random Slope and Random Intercept Model

### 6.6.1 Random Intercept Model with Random Slope for Education

We extended the random intercept model to allow for both a **random intercept** and a **random slope** for Education across countries. This model accounts for variability in life satisfaction not only at the country level (via the random intercept) but also for differences in the relationship between education and life satisfaction across countries (via the random slope for Education).

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: 'Life.satisfaction.(0-10)' ~ 'Education.(0-10)' + 'Jobs.(0-10)' +
##           'Income.(0-10)' + 'Safety.(0-10)' + 'Health.(0-10)' + 'Environment.(0-10)' +
##           'Civic.engagement.(0-10)' + 'Accessibility.to.services.(0-10)' +
##           'Housing.(0-10)' + 'Community.(0-10)' + (1 + 'Education.(0-10)' |
##           Country)
## Data: df_wb %>% select(-Region, -Code)
##
## REML criterion at convergence: 1048.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0770 -0.4608  0.0149  0.3999  4.1388
##
## Random effects:
##   Groups   Name        Variance Std.Dev. Corr
##   Country (Intercept) 3.9988  1.9997
##             'Education.(0-10)' 0.1283  0.3582  -0.91
##   Residual            1.8175  1.3482
## Number of obs: 285, groups: Country, 14
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                2.71979  1.00237  2.713
## 'Education.(0-10)'         0.18365  0.14257  1.288
## 'Jobs.(0-10)'              0.15170  0.07254  2.091
## 'Income.(0-10)'            -0.01495 0.12780 -0.117
## 'Safety.(0-10)'            -0.11387 0.05482 -2.077
## 'Health.(0-10)'             -0.07032 0.07799 -0.902
## 'Environment.(0-10)'        0.18495  0.07071  2.616
## 'Civic.engagement.(0-10)'  -0.03511 0.07111 -0.494
## 'Accessibility.to.services.(0-10)' 0.06419  0.08633  0.744
## 'Housing.(0-10)'            0.11566  0.09025  1.282
## 'Community.(0-10)'          0.18027  0.04148  4.346
##
## Correlation of Fixed Effects:
##          (Intr) 'Ed.(0-10)' 'J.(0- 'I.(0- 'S.(0- 'Hl.(0-10)' 'En.(0-10)'
## 'Ed.(0-10)' -0.506
## 'Jb.(0-10)'  0.003 -0.213
## 'In.(0-10)'   0.118 -0.261    -0.134
## 'Sf.(0-10)'  -0.304 -0.080    0.036  0.033
## 'Hl.(0-10)'  -0.190  0.055    0.000 -0.166 -0.220
## 'En.(0-10)'  -0.391 -0.002    -0.076 -0.139  0.012  0.026
## 'C..(0-10)'   0.091  0.021    -0.248 -0.152 -0.189 -0.249    0.080
## 'A...(0-10) -0.202 -0.121    -0.038 -0.116  0.067 -0.378    0.101
```

```

## 'Hs.(0-10)' 0.044 -0.202      -0.108 0.096 -0.071 0.213      -0.234
## 'Cm.(0-10)' -0.234 -0.024      -0.018 -0.014 0.086 -0.003      -0.011
## 'C..(0 'A...('Hs.(0-10)'
## 'Ed.(0-10)'
## 'Jb.(0-10)'
## 'In.(0-10)'
## 'Sf.(0-10)'
## 'Hl.(0-10)'
## 'En.(0-10)'
## 'C..(0-10)'
## 'A...(0-10) -0.027
## 'Hs.(0-10)' -0.175 -0.183
## 'Cm.(0-10)' -0.072 0.083 -0.166

```

The **REML criterion** at convergence is 1048.3, indicating the fit of the model to the data. Below is a breakdown of the key findings from the model:

### 1. Random Effects:

- **Country:** The model estimates variability in the intercept and slope across 14 countries.
- **Intercept:** The variance of the random intercept is 3.9982, with a standard deviation of 1.9996.
- **Slope for Education:** The variance for the random slope of Education is 0.1283, with a standard deviation of 0.3582. The negative correlation of **-0.91** between the intercept and slope for Education suggests that countries with higher intercepts (higher life satisfaction overall) tend to have a weaker positive association between Education and life satisfaction.
- **Residual:** The variance of the residual (unexplained by the model) is 1.8176, with a standard deviation of 1.3482.

### 2. Fixed Effects:

- **Intercept:** The estimated intercept is 2.7197, indicating the baseline life satisfaction when all predictors are at their reference levels.
- **Education:** The coefficient for Education is 0.18364, but it is not statistically significant ( $t = 1.288$ ,  $p > 0.05$ ), suggesting that the overall effect of Education on life satisfaction is not significant when accounting for random variation across countries.
- **Jobs:** The coefficient for Jobs is 0.15171, and it is statistically significant ( $t = 2.091$ ), indicating a positive relationship between jobs and life satisfaction.
- **Income:** The coefficient for Income is -0.01495, and it is not statistically significant ( $t = -0.117$ ), suggesting that income has a negligible effect on life satisfaction in this model.
- **Safety:** The coefficient for Safety is -0.11387, and it is statistically significant ( $t = -2.077$ ), suggesting a negative relationship between safety and life satisfaction.
- **Environment:** The coefficient for Environment is 0.18495, and it is statistically significant ( $t = 2.616$ ), suggesting a positive impact of environmental quality on life satisfaction.
- **Community:** The coefficient for Community is 0.18027, and it is statistically significant ( $t = 4.346$ ), indicating that a sense of community is positively associated with life satisfaction.

### 3. Model Diagnostics:

- **The scaled residuals** show that most values are within a reasonable range, with the minimum residual being -4.0770 and the maximum being 4.1388. These values suggest that the model does not exhibit extreme outliers.

- The correlation of fixed effects reveals some relationships between the predictors. For instance, there is a strong negative correlation (-0.506) between the intercept and Education, suggesting that the countries with higher baseline life satisfaction tend to have a weaker association between Education and life satisfaction.

The random intercept and random slope model provides insights into how life satisfaction varies both within countries and across countries. The random slope for Education suggests that while Education may have a positive effect on life satisfaction, the strength of this effect varies considerably across countries. Although the coefficient for Education is not statistically significant overall, the model reveals that the relationship between Education and life satisfaction is not uniform across countries.

Additionally, variables like **Community, Jobs, and Environment** show significant positive effects on life satisfaction, whereas **Safety and Income** are not significantly associated with life satisfaction in this model. This model improves our understanding by accounting for country-level variations, offering a more nuanced perspective on the factors influencing life satisfaction.

This model can serve as an important basis for policy discussions, especially when considering the role of Education and other factors in improving life satisfaction at a national level.

### 6.6.2 Random Intercept Model with Random Slope for Community

The model used here includes random intercepts and random slopes for **Community** in order to account for varying impacts of community engagement on **Life Satisfaction** across different countries. The random slope for community allows for the relationship between community and life satisfaction to differ by country, while the random intercept accounts for the baseline differences in life satisfaction across countries.

```
## Linear mixed model fit by REML [‘lmerMod’]
## Formula: ‘Life.satisfaction.(0-10)’ ~ ‘Education.(0-10)’ + ‘Jobs.(0-10)’ +
##           ‘Income.(0-10)’ + ‘Safety.(0-10)’ + ‘Health.(0-10)’ + ‘Environment.(0-10)’ +
##           ‘Civic.engagement.(0-10)’ + ‘Accessibility.to.services.(0-10)’ +
##           ‘Housing.(0-10)’ + ‘Community.(0-10)’ + (1 + ‘Community.(0-10)’ | 
##           Country)
## Data: df_wb %>% select(-Region, -Code)
##
## REML criterion at convergence: 1045.2
##
## Scaled residuals:
##      Min     1Q   Median     3Q    Max
## -3.8163 -0.3721  0.0712  0.4790  3.6315
##
## Random effects:
## Groups   Name        Variance Std.Dev. Corr
## Country (Intercept) 3.257    1.8047
##           ‘Community.(0-10)’ 0.029    0.1703  -0.75
## Residual            1.782    1.3351
## Number of obs: 285, groups: Country, 14
##
## Fixed effects:
##                               Estimate Std. Error t value
## (Intercept)                2.05779   0.97012  2.121
## ‘Education.(0-10)’          0.20139   0.10045  2.005
## ‘Jobs.(0-10)’               0.13475   0.06885  1.957
## ‘Income.(0-10)’             -0.04688   0.11898 -0.394
## ‘Safety.(0-10)’             -0.06680   0.05419 -1.233
```

```

## 'Health.(0-10)'          -0.01616  0.07737 -0.209
## 'Environment.(0-10)'      0.15195  0.06775  2.243
## 'Civic.engagement.(0-10)' -0.15918  0.06870 -2.317
## 'Accessibility.to.services.(0-10)' 0.10479  0.08517  1.230
## 'Housing.(0-10)'          0.14577  0.08442  1.727
## 'Community.(0-10)'         0.20786  0.06982  2.977
##
## Correlation of Fixed Effects:
##           (Intr) 'Ed.(0-10)' 'J.(0-' 'I.(0-' 'S.(0-' 'Hl.(0-10)' 'En.(0-10)'
## 'Ed.(0-10)'   -0.134
## 'Jb.(0-10)'   -0.129 -0.252
## 'In.(0-10)'    0.032 -0.285     -0.183
## 'Sf.(0-10)'   -0.337 -0.150     0.090  0.027
## 'Hl.(0-10)'   -0.107  0.063     0.011 -0.190 -0.185
## 'En.(0-10)'   -0.384 -0.022     -0.063 -0.071  0.035 -0.026
## 'C..(0-10)'    0.074  0.030     -0.341 -0.060 -0.172 -0.290      0.112
## 'A... (0-10)' -0.197 -0.274     0.062 -0.075  0.053 -0.376      0.083
## 'Hs.(0-10)'    0.020 -0.283     -0.006  0.054 -0.012  0.144     -0.245
## 'Cm.(0-10)'   -0.501 -0.048     0.008  0.000  0.040 -0.008     -0.023
##           'C..(0 'A... ('Hs.(0-10)'
##           'Ed.(0-10)'
##           'Jb.(0-10)'
##           'In.(0-10)'
##           'Sf.(0-10)'
##           'Hl.(0-10)'
##           'En.(0-10)'
##           'C..(0-10)'
##           'A... (0-10)' -0.016
##           'Hs.(0-10)'  -0.231 -0.143
##           'Cm.(0-10)'  -0.021  0.049 -0.077

```

## 1. Random Effects:

- **Country (Intercept):** The variance of the intercept is **3.257**, with a standard deviation of **1.8047**, indicating substantial variability in the baseline life satisfaction across the 14 countries in the study.
- **Country (Community Slope):** The variance for the random slope of **Community** is **0.029**, with a standard deviation of **0.1703**, indicating that the relationship between community engagement and life satisfaction varies somewhat across countries. The negative correlation of **-0.75** suggests that countries with higher baseline life satisfaction tend to have a weaker positive relationship between community engagement and life satisfaction.
- **Residual:** The residual variance is **1.782**, with a standard deviation of **1.3351**, representing unexplained variation in life satisfaction after accounting for both fixed and random effects.

## 2. Fixed Effects:

- **Intercept (2.05779):** The baseline life satisfaction, when all predictors are set to zero, is **2.05779**, and this effect is statistically significant ( $t = 2.121$ ).
- **Education (0.20139):** The coefficient for **Education** is **0.20139**, and it is **statistically significant** ( $t = 2.005$ ). This suggests that higher levels of education are positively associated with life satisfaction.
- **Jobs (0.13475):** The effect of Jobs on life satisfaction is **statistically significant** ( $t = 1.957$ ), indicating that better job satisfaction contributes positively to life satisfaction.
- **Income (-0.04688):** The effect of Income is **not statistically significant** ( $t = -0.394$ ), suggesting that income does not have a clear relationship with life satisfaction in this model.

- **Safety (-0.06680)**: The coefficient for Safety is **not statistically significant** ( $t = -1.233$ ), implying that safety concerns do not significantly impact life satisfaction in this model.
  - **Health (-0.01616)**: The effect of Health is **not statistically significant** ( $t = -0.209$ ), indicating no significant relationship between health and life satisfaction.
  - **Environment (0.15195)**: The coefficient for Environment is **statistically significant** ( $t = 2.243$ ), showing a positive relationship between environmental quality and life satisfaction.
  - **Civic Engagement (-0.15918)**: The coefficient for Civic Engagement is **statistically significant** ( $t = -2.317$ ), suggesting that greater civic engagement is negatively associated with life satisfaction in this model.
  - **Accessibility to Services (0.10479)**: The effect of Accessibility to Services is not statistically significant ( $t = 1.230$ ), indicating that access to services does not significantly affect life satisfaction.
  - **Housing (0.14577)**: The effect of Housing on life satisfaction is **marginally significant** ( $t = 1.727$ ), suggesting a potential positive relationship, though further investigation may be necessary to confirm this.
  - **Community (0.20786)**: The coefficient for Community is **statistically significant** ( $t = 2.977$ ), indicating that greater community engagement is positively associated with life satisfaction. This suggests that individuals who feel more connected to their communities tend to report higher life satisfaction.
3. **Correlation of Fixed Effects**: The correlation of fixed effects shows that **Education and Community** have a slight negative correlation of **-0.048**, indicating a small inverse relationship between these two factors across countries. This suggests that countries with higher education levels might have slightly less emphasis on community engagement, though this correlation is not large enough to be of significant concern.

#### 4. Model Diagnostics:

- **Scaled residuals** range from **-3.8163 to 3.6315**, with most residuals clustered around zero. The absence of extreme values indicates that the model fits the data reasonably well.
- **The negative correlation** between the intercept and slope for **Community** (-0.75) suggests that in countries with higher baseline life satisfaction, the impact of community on life satisfaction is weaker. This could imply that community engagement plays a larger role in improving life satisfaction in countries with lower overall life satisfaction.

The **Random Intercept and Random Slope Model for Community** provides valuable insights into the factors influencing life satisfaction across countries. **Community engagement** is positively related to life satisfaction, with countries that have stronger community ties showing higher life satisfaction. **Education and Jobs** are also significant predictors of life satisfaction, with education showing a positive relationship and jobs contributing to greater satisfaction. **Civic engagement**, however, is negatively associated with life satisfaction, suggesting that greater civic participation may not always enhance life satisfaction in the countries in this study. The model reveals that **environmental quality** is a significant factor in life satisfaction, further emphasizing the importance of quality of life in both personal and community settings.

## 6.7 R Squared

The **R-squared** values provide insights into the proportion of variance in life satisfaction explained by the models, both marginally (accounting for fixed effects) and conditionally (accounting for both fixed and random effects). Here are the results for the three models:

	Model	R_Squared_Marginal
## 1	Random Intercept Model	0.4164078
## 2	Random Intercept with Random Slope for Education	0.4034987
## 3	Random Intercept with Random Slope for Community	0.4276506

```

##   R_Squared_Conditional
## 1      0.6968376
## 2      0.7139324
## 3      0.7065214

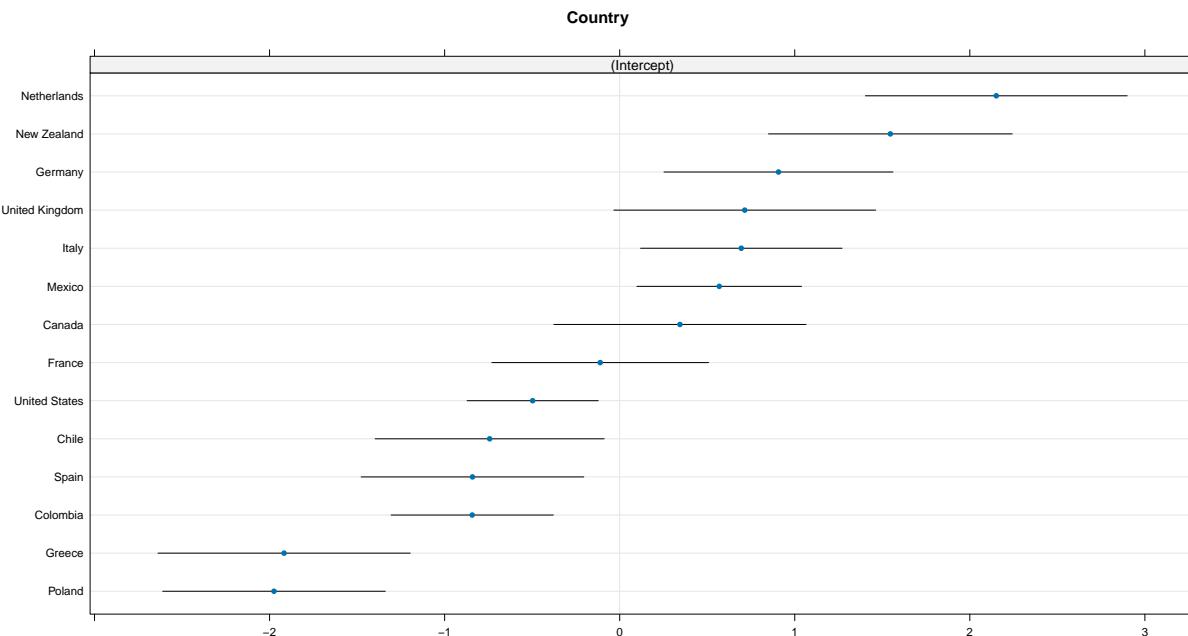
```

- **The Random Intercept Model** explains **41.64%** of the variance in life satisfaction through fixed effects alone (marginal R-squared). After accounting for country-specific variations, the model explains **69.68%** of the total variance in life satisfaction (conditional R-squared).
- \***The Random Intercept with Random Slope for Education** model marginally explains **40.35%** of the variance in life satisfaction. Including the random slopes for education results in a conditional R-squared of **71.39%**, suggesting that the random effects (differences between countries) account for a substantial portion of the variance, with education having a varying impact across countries.
- **The Random Intercept with Random Slope for Community** model has the highest marginal R-squared at **42.77%**, indicating that the model's fixed effects explain slightly more variance compared to the others. The conditional R-squared value of **70.65%** suggests that the inclusion of random slopes for community engagement further explains the variability in life satisfaction between countries.

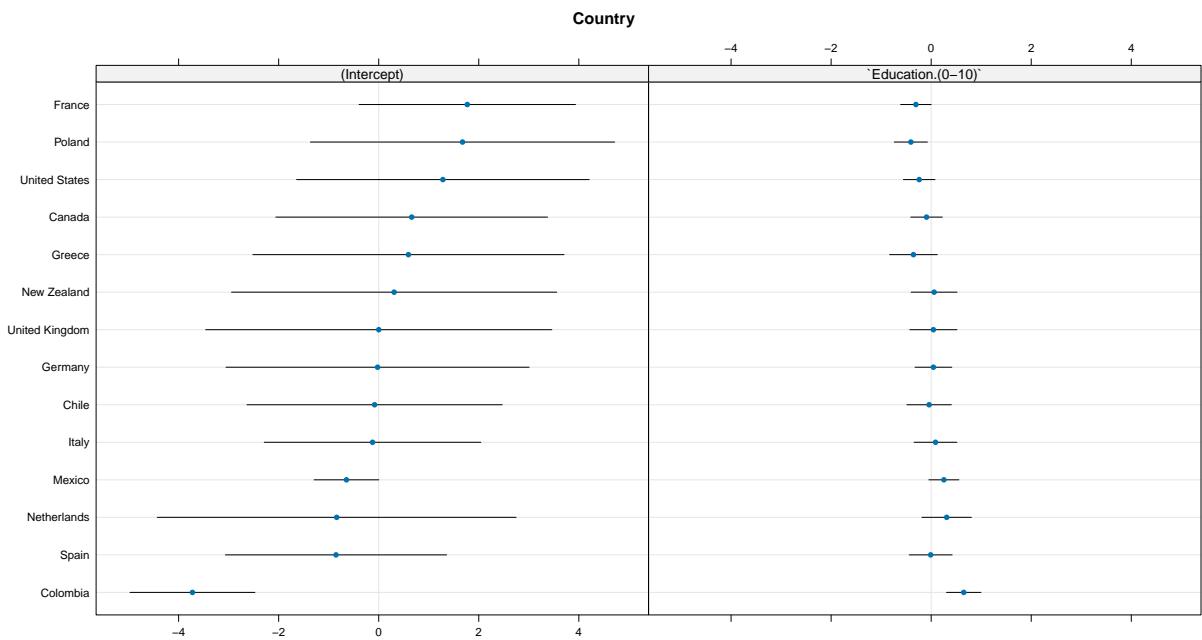
In summary, while all three models show substantial explanatory power for life satisfaction, the **Random Intercept with Random Slope for Education model** provides the highest overall fit, followed closely by the **Community model**. The inclusion of random slopes in both the education and community models improves the explanation of variability in life satisfaction across countries.

Dotplot of Random Effects from the Model with a Random Intercept and Random Slope:

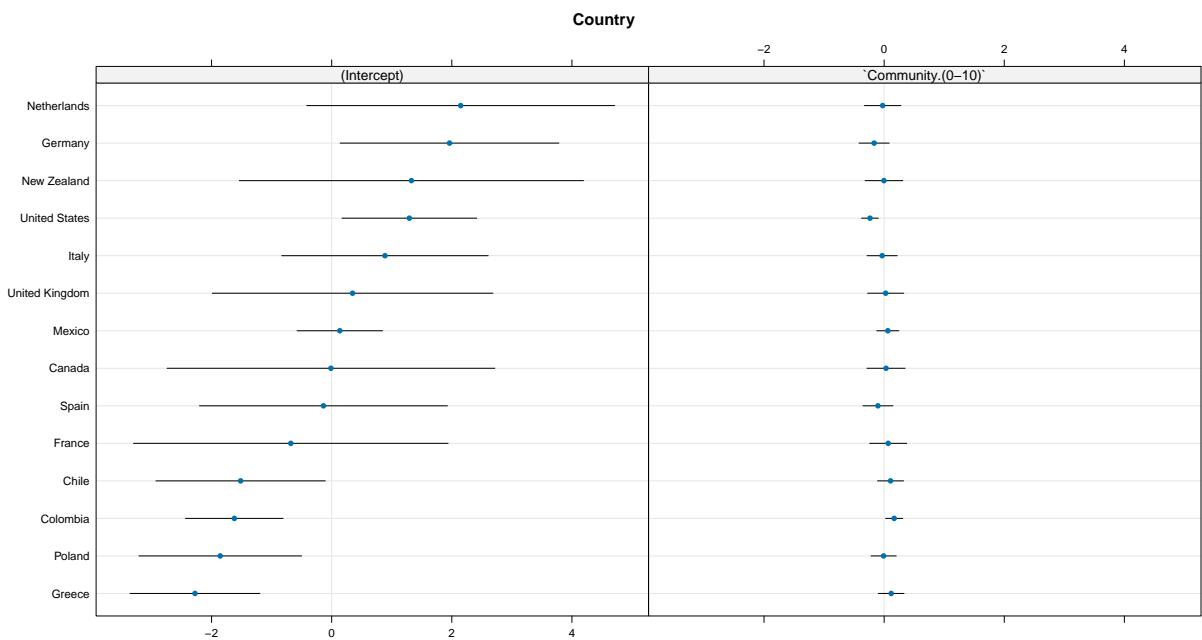
```
## $Country
```



```
## $Country
```



```
## $Country
```



## 6.8 Comparison of Models

```
##               Model      AIC      BIC
## 1             Model 1 1068.2796 1112.109
## 2             Model 2  977.1475 1185.339
## 3             Model 3  965.9398 1057.252
## 4 Model 4 (Random Intercept) 1082.4098 1129.892
## 5           Model 5 (RI+S_Edu 1078.2695 1133.057
## 6           Model 6 (RI+S_Comm) 1075.2022 1129.990
```

**Model 3** appears to provide the best fit in terms of AIC and BIC, suggesting it strikes the best balance between model complexity and accuracy, while the inclusion of random slopes (in Models 5 and 6) does not significantly improve the model's performance as indicated by the AIC/BIC comparison. From this comparison, we see that the Interaction effect no group model gives a better fit for the data, instead of the complex models. This can be due to the fact that we lack enough data points in each category, thus making the complex models work poorly.

## 7 Conclusion

To conclude, the clustering analysis successfully grouped regions into distinct clusters, revealing significant disparities in well-being influenced by factors such as housing, community, and employment. The regression models provided robust insights into the relationships between these factors and **life satisfaction**, emphasizing the importance of **housing and community** as primary determinants, alongside **income and safety**.

While economic factors are crucial, **social and environmental** aspects play an equally, if not more, significant role in shaping **life satisfaction**. However, future research could further refine these models by incorporating additional data and exploring temporal trends to better understand how life satisfaction evolves over time.

## 8 Appendix

### 8.1 A: Summary of Clusters with Model Fitted

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
## 
## Mclust VEV (ellipsoidal, equal shape) model with 4 components:
## 
##   log-likelihood   n   df       BIC       ICL
##             -1627.609 285 236 -4589.206 -4589.333
## 
## Clustering table:
##   1   2   3   4
## 115  57  65  48
## 
## Mixing probabilities:
##   1       2       3       4
## 0.4032620 0.2002456 0.2280711 0.1684213
## 
## Means:
##           [,1]      [,2]      [,3]      [,4]
## Education.(0-10) 0.2059389 0.33610881 -1.4232941 1.0346704
## Jobs.(0-10)      0.2901518 -0.36102661 -0.7000418 0.6824914
## Income.(0-10)     0.0869365 -0.32957114 -1.1448790 1.7340488
## Safety.(0-10)     0.6056543 0.40874312 -1.5171713 0.1183728
## Health.(0-10)     0.7777678 -0.07678917 -0.8056766 -0.6799380
## Environment.(0-10) 0.4146131 -0.88000950 -0.6445377 0.9263714
## Civic.engagement.(0-10) 0.8150896 -1.02976944 -0.6947396 0.2135245
## Accessibility.to.services.(0-10) 0.4584912 0.05241932 -1.4161384 0.7575726
## Housing.(0-10)     0.4050547 -0.63898592 -1.1635679 1.3655462
## Community.(0-10)    0.4754022 -0.32931839 -0.9612754 0.5549891
## 
## Variances:
##   [,1]
## 
##           Education.(0-10)  Jobs.(0-10)  Income.(0-10)
## Education.(0-10) 0.34980136 0.407182085 0.092807170
## Jobs.(0-10)      0.40718209 0.834731225 0.119791045
## Income.(0-10)    0.09280717 0.119791045 0.065108587
## Safety.(0-10)    -0.01655527 -0.006133159 -0.003876554
## Health.(0-10)    -0.05879976 -0.115383455 0.025958817
## Environment.(0-10) 0.10349335 -0.009294753 -0.030499331
## Civic.engagement.(0-10) 0.14509401 0.281653439 0.032168567
## Accessibility.to.services.(0-10) 0.02158839 -0.098242664 0.033742609
## Housing.(0-10)    0.18503138 0.170892287 0.001044623
## Community.(0-10)  0.06030100 0.059233844 -0.022023868
##           Safety.(0-10)  Health.(0-10)  Environment.(0-10)
## Education.(0-10) -0.016555274 -0.05879976 0.103493347
## Jobs.(0-10)      -0.006133159 -0.11538346 -0.009294753
## Income.(0-10)    -0.003876554 0.02595882 -0.030499331
## Safety.(0-10)    0.006718398 0.00822568 -0.015635688
## Health.(0-10)    0.008225680 0.21609910 -0.078631354
## Environment.(0-10) -0.015635688 -0.07863135 0.387884547
```

```

## Civic.engagement.(0-10)      -0.001524007   0.02423316   -0.015511267
## Accessibility.to.services.(0-10) -0.006549369   0.11456670   0.040168215
## Housing.(0-10)              -0.016589154   -0.08974164   0.274621107
## Community.(0-10)             -0.005168471   -0.02580432   0.167395605
##                                         Civic.engagement.(0-10)
## Education.(0-10)                0.145094013
## Jobs.(0-10)                   0.281653439
## Income.(0-10)                 0.032168567
## Safety.(0-10)                  -0.001524007
## Health.(0-10)                  0.024233157
## Environment.(0-10)             -0.015511267
## Civic.engagement.(0-10)          0.312758868
## Accessibility.to.services.(0-10) 0.009385007
## Housing.(0-10)                  0.062255648
## Community.(0-10)                0.056006849
##                                         Accessibility.to.services.(0-10)  Housing.(0-10)
## Education.(0-10)                0.021588389   0.185031378
## Jobs.(0-10)                     -0.098242664   0.170892287
## Income.(0-10)                   0.033742609   0.001044623
## Safety.(0-10)                   -0.006549369   -0.016589154
## Health.(0-10)                   0.114566698   -0.089741638
## Environment.(0-10)              0.040168215   0.274621107
## Civic.engagement.(0-10)          0.009385007   0.062255648
## Accessibility.to.services.(0-10) 0.297895467   0.098310627
## Housing.(0-10)                  0.098310627   0.368370397
## Community.(0-10)                0.056802692   0.169917710
##                                         Community.(0-10)
## Education.(0-10)                0.060300996
## Jobs.(0-10)                     0.059233844
## Income.(0-10)                   -0.022023868
## Safety.(0-10)                   -0.005168471
## Health.(0-10)                   -0.025804318
## Environment.(0-10)              0.167395605
## Civic.engagement.(0-10)          0.056006849
## Accessibility.to.services.(0-10) 0.056802692
## Housing.(0-10)                  0.169917710
## Community.(0-10)                0.340576918
## [,,2]
##                                         Education.(0-10)  Jobs.(0-10)  Income.(0-10)
## Education.(0-10)                 0.704440177   0.645273600   0.37880250
## Jobs.(0-10)                      0.645273600   1.273236111   0.38020307
## Income.(0-10)                     0.378802504   0.380203066   0.72351089
## Safety.(0-10)                     0.001136638   -0.045726193   -0.18315198
## Health.(0-10)                     -0.864678917   -1.025965923   -0.54759686
## Environment.(0-10)                0.339637297   0.001309143   0.73086258
## Civic.engagement.(0-10)            0.529657744   0.472669503   0.46917669
## Accessibility.to.services.(0-10)   0.191802268   0.490562370   -0.04139288
## Housing.(0-10)                   0.110166160   0.033336242   0.34737005
## Community.(0-10)                 -0.024431717   0.114990735   -0.01293922
##                                         Safety.(0-10)  Health.(0-10)  Environment.(0-10)
## Education.(0-10)                 0.001136638   -0.86467892   0.339637297
## Jobs.(0-10)                      -0.045726193   -1.02596592   0.001309143
## Income.(0-10)                     -0.183151982   -0.54759686   0.730862577
## Safety.(0-10)                     0.177700880   0.07287982   -0.155290192

```

## Health.(0-10)	0.072879818	1.96342448	-0.532833668
## Environment.(0-10)	-0.155290192	-0.53283367	2.070902923
## Civic.engagement.(0-10)	-0.010952974	-0.64797029	0.388072287
## Accessibility.to.services.(0-10)	-0.040980803	-0.02348970	-0.513211345
## Housing.(0-10)	-0.150007073	-0.08468139	0.471642398
## Community.(0-10)	-0.184290447	-0.19951018	0.089476918
##	Civic.engagement.(0-10)		
## Education.(0-10)		0.529657744	
## Jobs.(0-10)		0.472669503	
## Income.(0-10)		0.469176690	
## Safety.(0-10)		-0.010952974	
## Health.(0-10)		-0.647970295	
## Environment.(0-10)		0.388072287	
## Civic.engagement.(0-10)		0.753461869	
## Accessibility.to.services.(0-10)		-0.025272599	
## Housing.(0-10)		0.138877483	
## Community.(0-10)		-0.003184435	
##	Accessibility.to.services.(0-10)	Housing.(0-10)	
## Education.(0-10)		0.191802268	0.11016616
## Jobs.(0-10)		0.490562370	0.03333624
## Income.(0-10)		-0.041392877	0.34737005
## Safety.(0-10)		-0.040980803	-0.15000707
## Health.(0-10)		-0.023489700	-0.08468139
## Environment.(0-10)		-0.513211345	0.47164240
## Civic.engagement.(0-10)		-0.025272599	0.13887748
## Accessibility.to.services.(0-10)		0.755407352	-0.16356952
## Housing.(0-10)		-0.163569521	0.34007860
## Community.(0-10)		0.002148064	0.11500946
##	Community.(0-10)		
## Education.(0-10)		-0.024431717	
## Jobs.(0-10)		0.114990735	
## Income.(0-10)		-0.012939225	
## Safety.(0-10)		-0.184290447	
## Health.(0-10)		-0.199510185	
## Environment.(0-10)		0.089476918	
## Civic.engagement.(0-10)		-0.003184435	
## Accessibility.to.services.(0-10)		0.002148064	
## Housing.(0-10)		0.115009461	
## Community.(0-10)		1.225117273	
## [,,3]			
##	Education.(0-10)	Jobs.(0-10)	Income.(0-10)
## Education.(0-10)	0.218292478	-0.249089461	0.042661919
## Jobs.(0-10)	-0.249089461	0.669576935	-0.047674565
## Income.(0-10)	0.042661919	-0.047674565	0.015622258
## Safety.(0-10)	-0.010498767	0.107953617	-0.004919550
## Health.(0-10)	0.155340032	-0.282051158	0.039814031
## Environment.(0-10)	-0.003886977	0.060657513	-0.001295857
## Civic.engagement.(0-10)	0.101859301	-0.296383550	0.015421809
## Accessibility.to.services.(0-10)	-0.078066048	0.329833103	-0.011416315
## Housing.(0-10)	0.007779135	0.001859178	0.005790162
## Community.(0-10)	0.151245621	-0.262665963	0.023542835
##	Safety.(0-10)	Health.(0-10)	Environment.(0-10)
## Education.(0-10)	-0.010498767	0.155340032	-0.0038869767
## Jobs.(0-10)	0.107953617	-0.282051158	0.0606575128

```

## Income.(0-10) -0.004919550 0.039814031 -0.0012958570
## Safety.(0-10) 0.849739159 0.042503745 -0.0925500706
## Health.(0-10) 0.042503745 0.299712175 0.0023038611
## Environment.(0-10) -0.092550071 0.002303861 0.1666364139
## Civic.engagement.(0-10) 0.086947473 0.190595793 -0.0346634692
## Accessibility.to.services.(0-10) 0.102055221 -0.165786860 -0.0247877003
## Housing.(0-10) -0.006502189 0.010872885 -0.0003924336
## Community.(0-10) -0.149741117 0.132660767 -0.0136392994
## Civic.engagement.(0-10)
## Education.(0-10) 0.10185930
## Jobs.(0-10) -0.29638355
## Income.(0-10) 0.01542181
## Safety.(0-10) 0.08694747
## Health.(0-10) 0.19059579
## Environment.(0-10) -0.034663437
## Civic.engagement.(0-10) 0.43308331
## Accessibility.to.services.(0-10) -0.26609935
## Housing.(0-10) -0.03514126
## Community.(0-10) 0.17591681
## Accessibility.to.services.(0-10) Housing.(0-10)
## Education.(0-10) -0.07806605 0.0077791349
## Jobs.(0-10) 0.32983310 0.0018591777
## Income.(0-10) -0.01141631 0.0057901620
## Safety.(0-10) 0.10205522 -0.0065021886
## Health.(0-10) -0.16578686 0.0108728849
## Environment.(0-10) -0.02478770 -0.0003924336
## Civic.engagement.(0-10) -0.26609935 -0.0351412570
## Accessibility.to.services.(0-10) 0.38617419 0.0416321848
## Housing.(0-10) 0.04163218 0.0324726173
## Community.(0-10) -0.20336860 -0.0175575090
## Community.(0-10)
## Education.(0-10) 0.15124562
## Jobs.(0-10) -0.26266596
## Income.(0-10) 0.02354284
## Safety.(0-10) -0.14974112
## Health.(0-10) 0.13266077
## Environment.(0-10) -0.01363930
## Civic.engagement.(0-10) 0.17591681
## Accessibility.to.services.(0-10) -0.20336860
## Housing.(0-10) -0.01755751
## Community.(0-10) 0.72232047
## [,,4]
## Education.(0-10) Jobs.(0-10) Income.(0-10)
## Education.(0-10) 0.019020781 0.03488645 0.022812218
## Jobs.(0-10) 0.034886450 0.12245668 0.024135948
## Income.(0-10) 0.022812218 0.02413595 0.151095287
## Safety.(0-10) 0.021338578 0.03942980 0.050384797
## Health.(0-10) 0.029617392 0.04709618 0.164715557
## Environment.(0-10) 0.014435187 0.02236032 0.002246519
## Civic.engagement.(0-10) 0.042495061 0.05846904 0.099841828
## Accessibility.to.services.(0-10) -0.007249655 -0.01310691 0.040136487
## Housing.(0-10) 0.024885177 0.05468386 -0.014590003
## Community.(0-10) 0.015451488 0.03948197 0.013357092
## Safety.(0-10) Health.(0-10) Environment.(0-10)

```

```

## Education.(0-10)          0.021338578  0.02961739  0.014435187
## Jobs.(0-10)               0.039429803  0.04709618  0.022360316
## Income.(0-10)              0.050384797  0.16471556  0.002246519
## Safety.(0-10)              0.085072399  0.12070134  0.040667466
## Health.(0-10)              0.120701339  0.39759859  0.037280729
## Environment.(0-10)         0.040667466  0.03728073  0.078381753
## Civic.engagement.(0-10)    0.070935696  0.22898453  0.029828024
## Accessibility.to.services.(0-10) -0.013360516  0.02983863  -0.040759637
## Housing.(0-10)             -0.007441072  -0.07947989  0.010579484
## Community.(0-10)            0.033961317  0.00983620  0.014585084
##                                         Civic.engagement.(0-10)
## Education.(0-10)           0.04249506
## Jobs.(0-10)                0.05846904
## Income.(0-10)              0.09984183
## Safety.(0-10)              0.07093570
## Health.(0-10)              0.22898453
## Environment.(0-10)         0.02982802
## Civic.engagement.(0-10)    0.26696445
## Accessibility.to.services.(0-10) 0.02307354
## Housing.(0-10)              0.02013529
## Community.(0-10)            -0.02692342
##                                         Accessibility.to.services.(0-10)  Housing.(0-10)
## Education.(0-10)           -0.007249655  0.024885177
## Jobs.(0-10)                -0.013106907  0.054683862
## Income.(0-10)              0.040136487  -0.014590003
## Safety.(0-10)              -0.013360516  -0.007441072
## Health.(0-10)              0.029838633  -0.079479892
## Environment.(0-10)         -0.040759637  0.010579484
## Civic.engagement.(0-10)    0.023073541  0.020135289
## Accessibility.to.services.(0-10) 0.098684022  -0.024229531
## Housing.(0-10)              -0.024229531  0.092641409
## Community.(0-10)            -0.034832576  0.006234386
##                                         Community.(0-10)
## Education.(0-10)           0.015451488
## Jobs.(0-10)                0.039481971
## Income.(0-10)              0.013357092
## Safety.(0-10)              0.033961317
## Health.(0-10)              0.009836200
## Environment.(0-10)         0.014585084
## Civic.engagement.(0-10)    -0.026923419
## Accessibility.to.services.(0-10) -0.034832576
## Housing.(0-10)              0.006234386
## Community.(0-10)            0.386789965

```

## 8.2 B: Summary of Linear Regression using Pairwise Interaction effects

```

##
## Call:
## lm(formula = 'Life.satisfaction.(0-10)' ~ .^2, data = df_wb %>%
##     select(-Country, -Region, -Code))
##
## Residuals:
##      Min       1Q   Median       3Q      Max

```

```

## -4.6079 -0.6186 0.0057 0.6448 3.9617
##
## Coefficients:
##                                     Estimate
## (Intercept)                      7.289e+00
## 'Education.(0-10)'                 8.599e-01
## 'Jobs.(0-10)'                     -1.504e-01
## 'Income.(0-10)'                   1.892e-01
## 'Safety.(0-10)'                  -7.649e-01
## 'Health.(0-10)'                  -3.434e-01
## 'Environment.(0-10)'                -7.918e-01
## 'Civic.engagement.(0-10)'          -7.091e-01
## 'Accessibility.to.services.(0-10)' 6.830e-02
## 'Housing.(0-10)'                  7.239e-01
## 'Community.(0-10)'                 4.582e-01
## 'Education.(0-10)':'Jobs.(0-10)'   1.210e-02
## 'Education.(0-10)':'Income.(0-10)' 1.364e-02
## 'Education.(0-10)':'Safety.(0-10)' -1.010e-01
## 'Education.(0-10)':'Health.(0-10)' 4.469e-02
## 'Education.(0-10)':'Environment.(0-10)' -3.815e-02
## 'Education.(0-10)':'Civic.engagement.(0-10)' -1.719e-02
## 'Education.(0-10)':'Accessibility.to.services.(0-10)' -9.626e-02
## 'Education.(0-10)':'Housing.(0-10)' 8.845e-02
## 'Education.(0-10)':'Community.(0-10)' 2.396e-02
## 'Jobs.(0-10)':'Income.(0-10)' -1.626e-01
## 'Jobs.(0-10)':'Safety.(0-10)' -5.117e-02
## 'Jobs.(0-10)':'Health.(0-10)' 2.763e-02
## 'Jobs.(0-10)':'Environment.(0-10)' 4.359e-02
## 'Jobs.(0-10)':'Civic.engagement.(0-10)' 5.452e-02
## 'Jobs.(0-10)':'Accessibility.to.services.(0-10)' 4.203e-02
## 'Jobs.(0-10)':'Housing.(0-10)' 9.564e-02
## 'Jobs.(0-10)':'Community.(0-10)' -9.398e-04
## 'Income.(0-10)':'Safety.(0-10)' 7.411e-02
## 'Income.(0-10)':'Health.(0-10)' -6.156e-02
## 'Income.(0-10)':'Environment.(0-10)' 7.238e-02
## 'Income.(0-10)':'Civic.engagement.(0-10)' 9.898e-02
## 'Income.(0-10)':'Accessibility.to.services.(0-10)' 2.363e-02
## 'Income.(0-10)':'Housing.(0-10)' -1.122e-01
## 'Income.(0-10)':'Community.(0-10)' -1.818e-02
## 'Safety.(0-10)':'Health.(0-10)' -3.852e-02
## 'Safety.(0-10)':'Environment.(0-10)' 1.262e-01
## 'Safety.(0-10)':'Civic.engagement.(0-10)' 1.007e-01
## 'Safety.(0-10)':'Accessibility.to.services.(0-10)' 1.385e-01
## 'Safety.(0-10)':'Housing.(0-10)' -1.694e-01
## 'Safety.(0-10)':'Community.(0-10)' 1.874e-02
## 'Health.(0-10)':'Environment.(0-10)' 6.112e-03
## 'Health.(0-10)':'Civic.engagement.(0-10)' 8.490e-03
## 'Health.(0-10)':'Accessibility.to.services.(0-10)' -6.727e-05
## 'Health.(0-10)':'Housing.(0-10)' 3.260e-02
## 'Health.(0-10)':'Community.(0-10)' 1.082e-02
## 'Environment.(0-10)':'Civic.engagement.(0-10)' -3.581e-02
## 'Environment.(0-10)':'Accessibility.to.services.(0-10)' -3.395e-02
## 'Environment.(0-10)':'Housing.(0-10)' -1.116e-02
## 'Environment.(0-10)':'Community.(0-10)' -1.984e-03

```

	Std. Error	t value
## 'Civic.engagement.(0-10)':'Accessibility.to.services.(0-10)'	-2.110e-02	
## 'Civic.engagement.(0-10)':'Housing.(0-10)'	-3.995e-02	
## 'Civic.engagement.(0-10)':'Community.(0-10)'	-3.821e-02	
## 'Accessibility.to.services.(0-10)':'Housing.(0-10)'	1.716e-02	
## 'Accessibility.to.services.(0-10)':'Community.(0-10)'	-8.147e-02	
## 'Housing.(0-10)':'Community.(0-10)'	1.143e-02	
##		
## (Intercept)	2.491e+00	2.926
## 'Education.(0-10)'	4.450e-01	1.932
## 'Jobs.(0-10)'	3.236e-01	-0.465
## 'Income.(0-10)'	7.900e-01	0.239
## 'Safety.(0-10)'	3.345e-01	-2.287
## 'Health.(0-10)'	3.407e-01	-1.008
## 'Environment.(0-10)'	3.635e-01	-2.178
## 'Civic.engagement.(0-10)'	3.286e-01	-2.158
## 'Accessibility.to.services.(0-10)'	4.544e-01	0.150
## 'Housing.(0-10)'	4.937e-01	1.466
## 'Community.(0-10)'	2.330e-01	1.967
## 'Education.(0-10)':'Jobs.(0-10)'	3.497e-02	0.346
## 'Education.(0-10)':'Income.(0-10)'	7.931e-02	0.172
## 'Education.(0-10)':'Safety.(0-10)'	3.764e-02	-2.683
## 'Education.(0-10)':'Health.(0-10)'	4.531e-02	0.986
## 'Education.(0-10)':'Environment.(0-10)'	4.569e-02	-0.835
## 'Education.(0-10)':'Civic.engagement.(0-10)'	3.884e-02	-0.443
## 'Education.(0-10)':'Accessibility.to.services.(0-10)'	4.681e-02	-2.056
## 'Education.(0-10)':'Housing.(0-10)'	4.844e-02	1.826
## 'Education.(0-10)':'Community.(0-10)'	2.573e-02	0.931
## 'Jobs.(0-10)':'Income.(0-10)'	5.602e-02	-2.903
## 'Jobs.(0-10)':'Safety.(0-10)'	3.009e-02	-1.701
## 'Jobs.(0-10)':'Health.(0-10)'	3.620e-02	0.763
## 'Jobs.(0-10)':'Environment.(0-10)'	4.064e-02	1.073
## 'Jobs.(0-10)':'Civic.engagement.(0-10)'	2.525e-02	2.159
## 'Jobs.(0-10)':'Accessibility.to.services.(0-10)'	3.537e-02	1.188
## 'Jobs.(0-10)':'Housing.(0-10)'	3.876e-02	2.468
## 'Jobs.(0-10)':'Community.(0-10)'	2.369e-02	-0.040
## 'Income.(0-10)':'Safety.(0-10)'	6.089e-02	1.217
## 'Income.(0-10)':'Health.(0-10)'	5.559e-02	-1.107
## 'Income.(0-10)':'Environment.(0-10)'	6.138e-02	1.179
## 'Income.(0-10)':'Civic.engagement.(0-10)'	4.879e-02	2.029
## 'Income.(0-10)':'Accessibility.to.services.(0-10)'	5.412e-02	0.437
## 'Income.(0-10)':'Housing.(0-10)'	5.293e-02	-2.119
## 'Income.(0-10)':'Community.(0-10)'	3.522e-02	-0.516
## 'Safety.(0-10)':'Health.(0-10)'	3.543e-02	-1.087
## 'Safety.(0-10)':'Environment.(0-10)'	4.441e-02	2.842
## 'Safety.(0-10)':'Civic.engagement.(0-10)'	3.288e-02	3.063
## 'Safety.(0-10)':'Accessibility.to.services.(0-10)'	4.047e-02	3.422
## 'Safety.(0-10)':'Housing.(0-10)'	5.640e-02	-3.003
## 'Safety.(0-10)':'Community.(0-10)'	1.979e-02	0.947
## 'Health.(0-10)':'Environment.(0-10)'	3.611e-02	0.169
## 'Health.(0-10)':'Civic.engagement.(0-10)'	3.016e-02	0.281
## 'Health.(0-10)':'Accessibility.to.services.(0-10)'	3.268e-02	-0.002
## 'Health.(0-10)':'Housing.(0-10)'	3.871e-02	0.842
## 'Health.(0-10)':'Community.(0-10)'	2.213e-02	0.489
## 'Environment.(0-10)':'Civic.engagement.(0-10)'	3.743e-02	-0.957

## 'Environment.(0-10)':'Accessibility.to.services.(0-10)'	4.990e-02	-0.680
## 'Environment.(0-10)':'Housing.(0-10)'	3.675e-02	-0.304
## 'Environment.(0-10)':'Community.(0-10)'	2.862e-02	-0.069
## 'Civic.engagement.(0-10)':'Accessibility.to.services.(0-10)'	3.673e-02	-0.575
## 'Civic.engagement.(0-10)':'Housing.(0-10)'	4.201e-02	-0.951
## 'Civic.engagement.(0-10)':'Community.(0-10)'	2.501e-02	-1.528
## 'Accessibility.to.services.(0-10)':'Housing.(0-10)'	4.374e-02	0.392
## 'Accessibility.to.services.(0-10)':'Community.(0-10)'	3.238e-02	-2.516
## 'Housing.(0-10)':'Community.(0-10)'	3.288e-02	0.348
##	Pr(> t )	
## (Intercept)	0.003780	**
## 'Education.(0-10)'	0.054544	.
## 'Jobs.(0-10)'	0.642534	
## 'Income.(0-10)'	0.810952	
## 'Safety.(0-10)'	0.023129	*
## 'Health.(0-10)'	0.314536	
## 'Environment.(0-10)'	0.030424	*
## 'Civic.engagement.(0-10)'	0.031991	*
## 'Accessibility.to.services.(0-10)'	0.880641	
## 'Housing.(0-10)'	0.143953	
## 'Community.(0-10)'	0.050444	.
## 'Education.(0-10)':'Jobs.(0-10)'	0.729726	
## 'Education.(0-10)':'Income.(0-10)'	0.863645	
## 'Education.(0-10)':'Safety.(0-10)'	0.007821	**
## 'Education.(0-10)':'Health.(0-10)'	0.324998	
## 'Education.(0-10)':'Environment.(0-10)'	0.404626	
## 'Education.(0-10)':'Civic.engagement.(0-10)'	0.658474	
## 'Education.(0-10)':'Accessibility.to.services.(0-10)'	0.040885	*
## 'Education.(0-10)':'Housing.(0-10)'	0.069173	.
## 'Education.(0-10)':'Community.(0-10)'	0.352756	
## 'Jobs.(0-10)':'Income.(0-10)'	0.004056	**
## 'Jobs.(0-10)':'Safety.(0-10)'	0.090343	.
## 'Jobs.(0-10)':'Health.(0-10)'	0.446111	
## 'Jobs.(0-10)':'Environment.(0-10)'	0.284510	
## 'Jobs.(0-10)':'Civic.engagement.(0-10)'	0.031858	*
## 'Jobs.(0-10)':'Accessibility.to.services.(0-10)'	0.235892	
## 'Jobs.(0-10)':'Housing.(0-10)'	0.014332	*
## 'Jobs.(0-10)':'Community.(0-10)'	0.968388	
## 'Income.(0-10)':'Safety.(0-10)'	0.224831	
## 'Income.(0-10)':'Health.(0-10)'	0.269284	
## 'Income.(0-10)':'Environment.(0-10)'	0.239579	
## 'Income.(0-10)':'Civic.engagement.(0-10)'	0.043651	*
## 'Income.(0-10)':'Accessibility.to.services.(0-10)'	0.662775	
## 'Income.(0-10)':'Housing.(0-10)'	0.035141	*
## 'Income.(0-10)':'Community.(0-10)'	0.606152	
## 'Safety.(0-10)':'Health.(0-10)'	0.278062	
## 'Safety.(0-10)':'Environment.(0-10)'	0.004885	**
## 'Safety.(0-10)':'Civic.engagement.(0-10)'	0.002453	**
## 'Safety.(0-10)':'Accessibility.to.services.(0-10)'	0.000736	***
## 'Safety.(0-10)':'Housing.(0-10)'	0.002968	**
## 'Safety.(0-10)':'Community.(0-10)'	0.344726	
## 'Health.(0-10)':'Environment.(0-10)'	0.865730	
## 'Health.(0-10)':'Civic.engagement.(0-10)'	0.778616	
## 'Health.(0-10)':'Accessibility.to.services.(0-10)'	0.998359	

```

## 'Health.(0-10)':'Housing.(0-10)'          0.400507
## 'Health.(0-10)':'Community.(0-10)'        0.625525
## 'Environment.(0-10)':'Civic.engagement.(0-10)' 0.339767
## 'Environment.(0-10)':'Accessibility.to.services.(0-10)' 0.496966
## 'Environment.(0-10)':'Housing.(0-10)'        0.761678
## 'Environment.(0-10)':'Community.(0-10)'       0.944803
## 'Civic.engagement.(0-10)':'Accessibility.to.services.(0-10)' 0.566176
## 'Civic.engagement.(0-10)':'Housing.(0-10)'      0.342633
## 'Civic.engagement.(0-10)':'Community.(0-10)'    0.127953
## 'Accessibility.to.services.(0-10)':'Housing.(0-10)' 0.695125
## 'Accessibility.to.services.(0-10)':'Community.(0-10)' 0.012563 *
## 'Housing.(0-10)':'Community.(0-10)'           0.728533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 1.227 on 229 degrees of freedom
## Multiple R-squared:  0.795, Adjusted R-squared:  0.7458
## F-statistic: 16.15 on 55 and 229 DF,  p-value: < 2.2e-16

```