Ian Harr

1. An executive summary - This should provide an overview of the problem, analysis, and conclusions. Should not exceed a paragraph.

The problem this project was trying to uncover was to determine the life satisfaction of each individual country during the year 2020. The year 2020 was chosen because it was the year that covid took a full swing globally, so I wanted to see how that affect life satisfaction. The way used to analyze the data pulled from <https://ourworldindata.org/> was hierarchical clustering. The two methods used were AGNED and DIANA clustering methods. The results grouped the highest life satisfaction countries together and the lowest life satisfaction countries together and so on.

2. Problem - Briefly explain and provide context to the problem you are addressing

The problem being uncovered was the discovery of which countries had the highest and lowest life satisfaction based on the year 2020. The year 2020 was chosen because this was the year that covid took a full swing around the globe. I thought it would be interesting to see the affect of this factor on the results. The results of this project could help someone who is interested in moving internationally decide where to re-locate. Coming out of the pandemic many people were looking for a change so maybe the countries with the highest life satisfaction rating during the pandemic would be a good place for them to relocate.

3. Technique - What technique did you use, and why?

The technique that was applied for this project was hierarchical clustering through R programming. Two different methods where used, AGNES which was used for agglomerative clustering, and DIANA which was used for divisive clustering. These methods were applied to my data to group countries together based on life satisfaction in the year 2020. The AGNES method took each object as a single element and combined two elements that were similar into a bigger cluster and so on. This eventually created a dendrogram which grouped the similar countries based on life satisfaction. Then, DIANA was applied, which starts as all objects in a single cluster, then breaks the most heterogeneous cluster into two at each iteration. This was used to create another dendrogram to compare the results of the data.

4. Conclusions - What are your conclusions

The results of both dendrograms grouped the countries with highest life satisfaction together, lowest life satisfaction together, and so on. The two dendrograms created from AGNES and DIANA methods were very similar. They appeared to be flipped with highest life satisfaction countries on the left for the AGNES method but highest life satisfaction countries on the right for the DIANA method. They gave very similar results overall with some of the groups slightly different. Overall, the top three countries with the highest life satisfaction were Finland, Denmark, and Iceland. The bottom three countries with the lowest life satisfaction were Afghanistan, Lebanon, and Zimbabwe.

5. Appendix - Attach all code and references here.

I referenced all of the Module 8 PowerPoint slides and the implementation slides for the programming aspect.

<https://bookdown.org/egarpor/PM-UC3M/app-softw-intro-R.html> was used to refresh how to use some basic R commands.