

— Measures of World Happiness —

ISYE 312:
DATA MANAGEMENT & ANALYSIS FOR INDUSTRIAL ENGINEERS

Spring 2022
Professor Tina Xu

Group 7
Claire Carlson, Emma Muhr, & Karissa Tschida



Table of Contents

Table of Contents	2
Data Exploration	3
Data Set	3
Data Analysis	4
Independent Variables vs. Ladder Score	4
Correlation Plot	5
Outliers	5
Questions	7
Question 1	8
If at all, how does generosity affect ladder score calculations?	8
Question 2	11
How do social support and freedom to make life choices affect the ladder score in Western European countries?	11
Question 3	12
How does the generosity of a country affect their perceptions of corruption?	12
Question 4	13
For Western Europe, Latin America & the Caribbean, and Sub-Saharan Africa, which factor has the highest impact on the ladder score of the region?	13
Question 5	16
Which region of the world has the highest ladder score?	16
Conclusions	17
Resources	19

Data Exploration

Data Set

Across the world, happiness is displayed and perceived in many different ways. For that reason, the Gallup World Poll team seeks to collect a variety of data which is then compiled into the World Happiness Report each year. The “World Happiness Report 2021” data set used for this analysis originally came from Kaggle, which was uploaded to the website by Ajaypal Singh (Singh, 2021). From the “Metadata” section on Kaggle, the World Happiness Report website was accessible. According to the World Happiness Report website, “The World Happiness Report 2021 focuses on the effects of COVID-19 and how people all over the world have fared. [Their] aim was two-fold, first to focus on the effects of COVID-19 on the structure and quality of people’s lives, and second to describe and evaluate how governments all over the world have dealt with the pandemic. In general, [they] try to explain why some countries have done so much better than others” (“World Happiness Report”, 2021). For this analysis, the World Happiness Report, published in 2021, was examined.

In this data set, there are several variables which contribute to the overall happiness ranking of a country. These variables, displayed in the table below, are based on survey information collected between 2018 and 2020.

Variable	Description
Country	Location
Regional Indicator	Region where country is located
Ladder Score (Happiness Score/Life Evaluation Score)	Average life evaluations based on the 2018-2020 surveys
Logged GDP per Capita	Average income of a country
Social Support	Average social support of a country
Healthy Life Expectancy	Average life expectancy of a country
Freedom to Make Life Choices	Average ability of citizens’ freedom to make life choices in a country
Generosity	Average generosity of a country
Perceptions of Corruption	Average perceptions of corruption within a country

Table 1: Variables of Data Set

In the case of this data analysis, we specifically focused on the Country, Regional Indicator, Ladder Score, Logged GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Make Life Choices, Generosity, and Perceptions of Corruption. This data set contains information on 149 countries, meaning there are 149 unique values. In the original data set, there are 20 different variables, but for the case of this analysis, there are only 17 that were essential to the conceptualization of the data, and the 9 (previously mentioned) that were needed for the analysis.

Data Analysis

The overarching goal of this data analysis is to evaluate how different characteristics of a region affect overall happiness. In order to do this, we are posing five questions:

1. If at all, does generosity affect ladder score calculations?
2. How do social support and freedom to make life choices affect the ladder score in Western European countries?
3. How does the generosity of a country affect their perceptions of corruption?
4. For Western Europe, Latin America & the Caribbean, and Sub-Saharan Africa, which factor has the highest impact on the ladder score of the region?
5. Which region of the world has the highest ladder score?

Independent Variables vs. Ladder Score

In order to get a general idea of the independent variables and their relationship with the ladder score, scatter plots are displayed below.

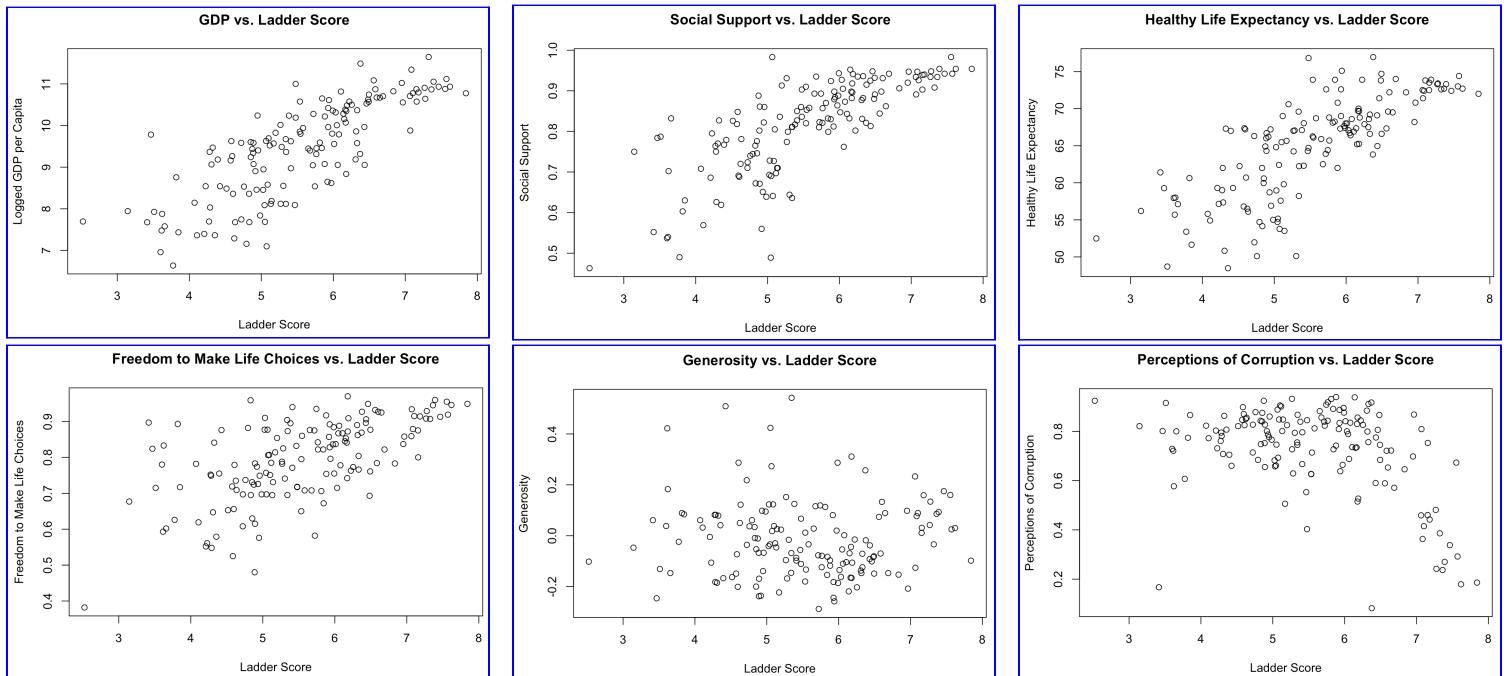


Figure 1: Simple Scatter Plots of Independent Variables vs. Ladder Score

As seen above, it seems there may be some linear or nearly linear relationships between GDP and Ladder Score, Social Support and Ladder Score, as well as Healthy Life Expectancy and Ladder Score. However, this cannot be definitively stated without more data exploration, some of which will be examined in this data analysis.

Correlation Plot

For a closer look at the correlation between different independent variables, a correlation plot was created.

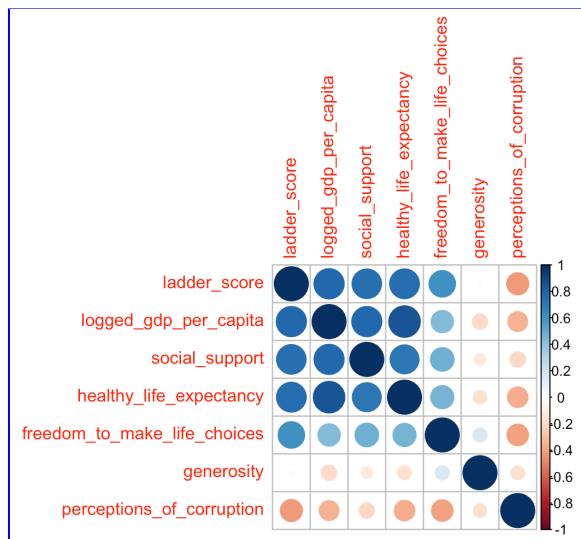


Figure 2: Independent Variables Correlation Plot

As displayed in the scatter plots in Figure 1, the correlation plot similarly displays the idea that there is a relatively strong positive correlation between GDP and Ladder Score, Social Support and Ladder Score, as well as Healthy Life Expectancy and Ladder Score. Interestingly, there seems to be no correlation between Generosity and Ladder score and very little correlation between Generosity and the other independent variables. This observation poses the question: if at all, does generosity affect the ladder score? The answer to this question will be explored later in this analysis.

Outliers

The final thing to consider before beginning this data analysis is if there are any outliers in the data set. First, the original data itself (including its outliers) was plotted to better understand the linearity of the current data (displayed in Figure 3 below).

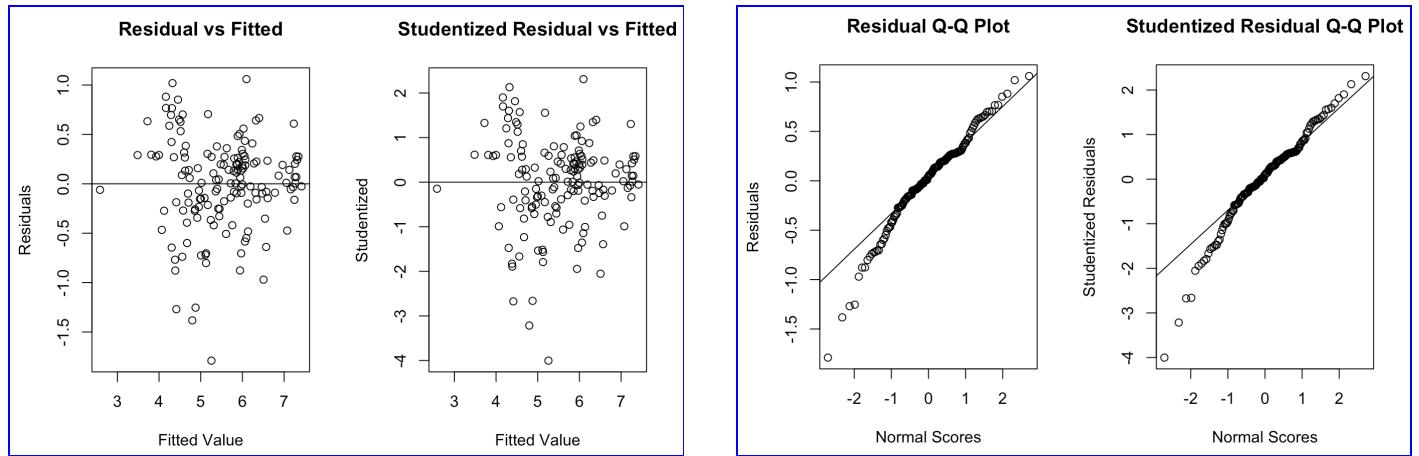


Figure 3: Simple Linear Regression of Data (Including Outliers)

In the Residual and Studentized Residual vs. Fitted plots on the left of Figure 3, the data distribution around the line at 0 follows a slight double bow pattern. This is reflected in the Q-Q Plots on the right, where the data is displaying a tailed distribution. In order to correct this, influential outliers of the data set will be removed using Cook's Distance to identify these points.

We begin by plotting Cook's Distance for each observation. The traditional cut-off point for Cook's Distance is $4/n$, where n is the sample size, which is represented by the blue dashed line in the figure below.

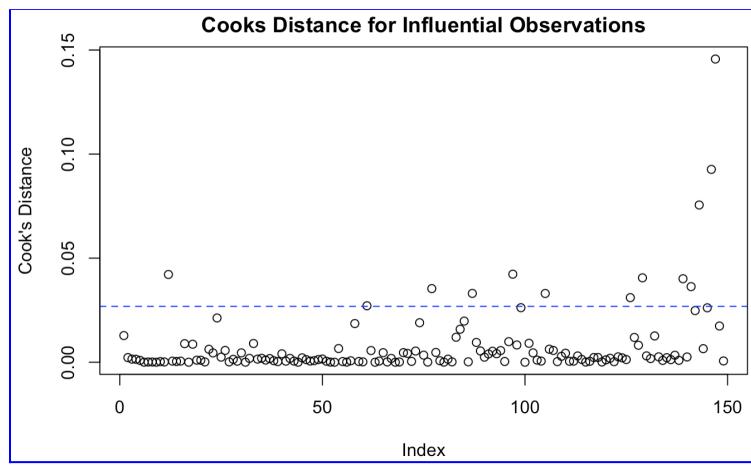


Figure 4: Cook's Distance

To gain a better understanding of where the influential points lie, we also created a bar graph and a chart of Cook's Distance. The plots of which display, in red, the outliers more clearly than Figure 4 does.

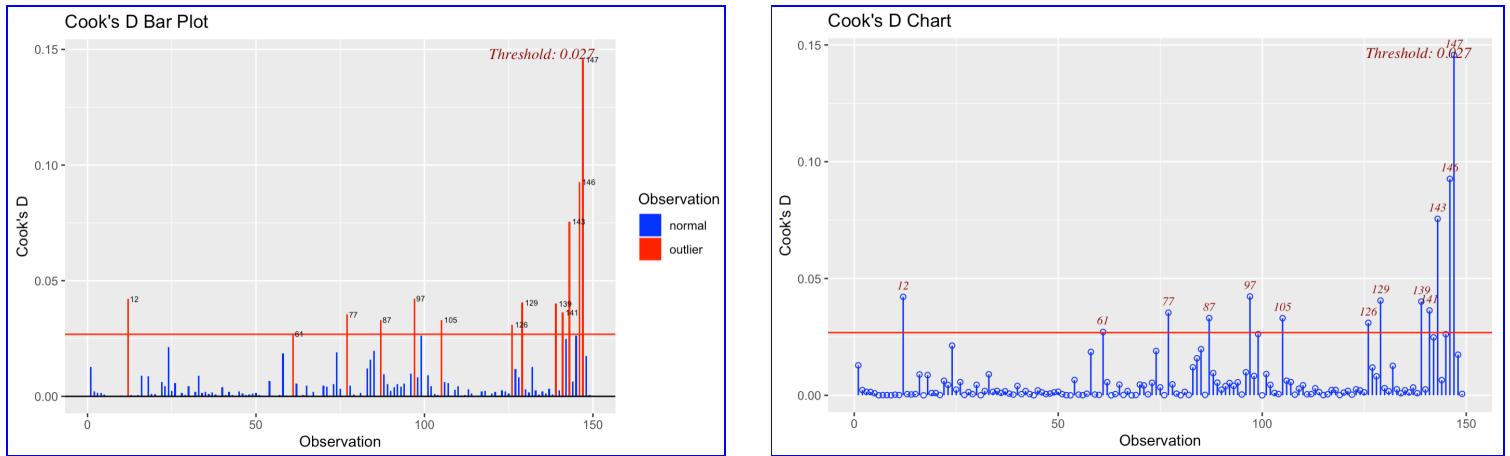


Figure 5: Cook's Distance Bar Plot & Chart

As displayed in both of the plots above, anything above the red line is considered an influential point. Therefore, there are 13 influential points in the data set. Upon identifying these outliers, they were removed from the data set and the Residual vs. Fitted graph, Studentized Residual vs. Fitted graph, Residual Q-Q Plot, and Studentized Residual Q-Q Plot were plotted again, this time without the influential points.

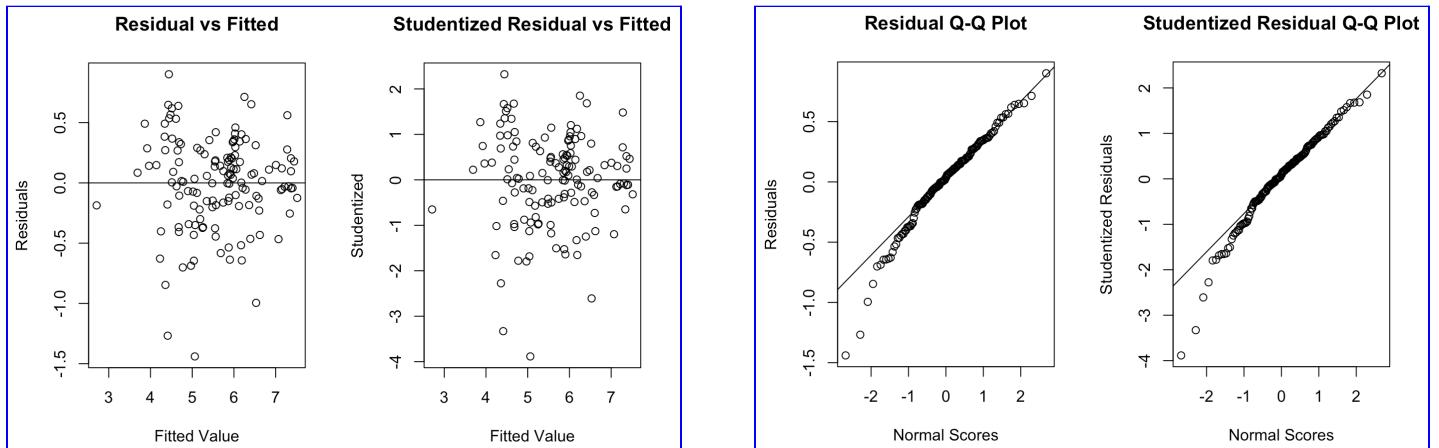


Figure 6: Simple Linear Regression of Data (Excluding Outliers)

Removing the influential points allowed for the data to follow a much better linear pattern on the Q-Q Plots than before. Therefore, no stabilizing transformations will need to be performed on the data in order to complete this analysis.

Questions

As mentioned above in the “Data Analysis” section of the report, the main goal of this data exploration is to evaluate how different characteristics of a region affect overall

happiness. In order to do this, we will explore the answers to five different questions which will allow us to discover how certain attributes pertaining to happiness can impact other attributes.

Question 1

If at all, how does generosity affect ladder score calculations?

After exploring the data and the website's explanation for how the ladder score is calculated, we noticed that it looked like generosity either had very low or little significance in calculating the ladder score, based on how scattered the data was in the scatter plot. To address the question of whether Generosity is a factor when calculating the ladder score, we decided to perform a multi-linear regression analysis. After looking at the graphs for Social Support versus Regional Indicator, the highest ladder scores were in Western Europe. We decided to focus on this region for analysis because zooming in on one gave us a better picture of the data calculations as a whole. We performed the multiple linear regression analysis with social support, freedom to make life choices, and generosity as the independent variables. We created a data frame named 'weurope' that contained the same data points as our original data frame, but the only regional indicator was Western Europe. After plotting a simple scatter plot matrix of the variables used in analysis, Social Support, Freedom to make Life Choices, and Generosity, and Ladder Score, it is still not apparent whether Generosity is linearly related to ladder score.

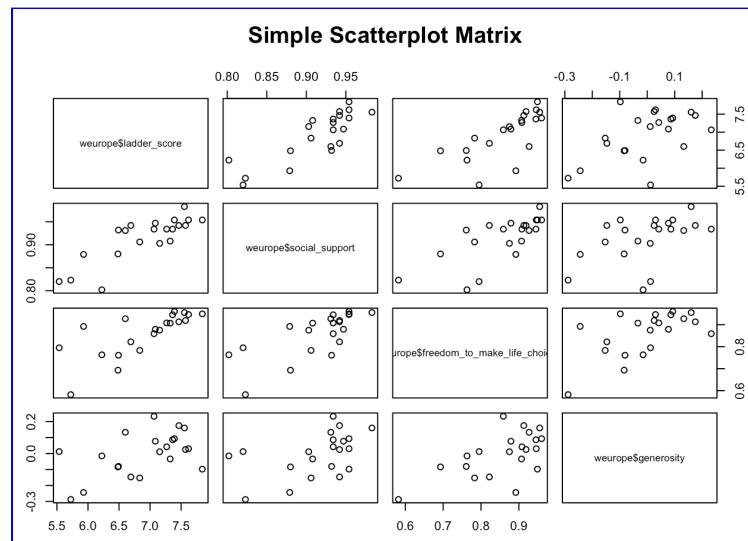


Figure 7: Scatter Plot Matrix of Multiple Linear Regression

Using R, we created a multiple linear regression model to explore the relationship between Ladder Score and Social Support, Freedom to make Life Choices, and Generosity. After looking at the linear models summary statistics we calculated the

regression line shown below where x_1 = social support, x_2 = freedom to make life choices, and x_3 = generosity.

$$y = 8.11x_1 + 1.7743x_2 + 0.3631x_3 - 2.0271$$

To explore whether generosity was an influential factor in calculating ladder score, we calculated R-squared and adjusted R-squared. Since R-squared can be inflated simply by adding more terms to the model, we found adjusted R-squared more important because it penalizes you for added terms to the model that are insignificant. Our adjusted R-squared for this model was 0.6878331, as seen in Figure 8.

Analysis of Variance Table						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
social_support	1	5.9202	5.9202	44.0015	4.229e-06	***
freedom_to_make_life_choices	1	0.3822	0.3822	2.8409	0.1102	
generosity	1	0.0307	0.0307	0.2282	0.6390	
Residuals	17	2.2873	0.1345			

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1						
 R^2 = 1 - (SS(Res) / SS(T))						
adjusted R^2 = 1 - SS(Res)/(n-p) / SS(T)/(n-1)						
 ```{r}						
rsq1 <- 1 - ((2.2873) / (5.92+0.3822+0.0307+2.2873))						
rsq1						
 adj_rsq1 <- 1 - (2.2873 / 17) / ((5.92+0.3822+0.0307+2.2873) / 20)						
adj_rsq1						
 [1] 0.7346581						
[1] 0.6878331						

Figure 8: ANOVA Table with Generosity

To see if the adjusted R-squared would be penalized at all from adding Generosity, our adjusted R-squared when running the same kind of multi-linear regression model, without Generosity, was 0.7012186, as seen in Figure 9.

```

Analysis of Variance Table

Response: ladder_score
 Df Sum Sq Mean Sq F value Pr(>F)
social_support 1 5.9202 5.9202 45.9727 2.373e-06 ***
freedom_to_make_life_choices 1 0.3822 0.3822 2.9681 0.1021
Residuals 18 2.3180 0.1288

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R^2 = 1 - (SS(Res) / SS(T))

adjusted R^2 = 1 - SS(Res)/(n-p) / SS(T)/(n-1)

```{r}
rsq2 <- 1 - ((2.3180) / (5.92+0.3822+2.3180))
rsq2

adj_rsq2 <- 1 - (2.3180 / 18) / ((5.92+0.3822+2.3180) / 20)
adj_rsq2
```

[1] 0.7310967
[1] 0.7012186

```

*Figure 9: ANOVA Table without Generosity*

This shows us that without Generosity in the model, the adjusted R-squared value was higher than when we had it included. To test the hypothesis that Generosity doesn't affect ladder score, our hypothesis was:

$$H_0: \beta_3 = 0$$

$$H_a: \beta_3 \neq 0$$

Test Statistic:

$$t_0: \frac{\widehat{\beta}_3}{\sqrt{\sigma^2 C_{jj}}}$$

```

```{r}
sigmasq <- (sigma(weurope_model_generosity))^2
sigmasq
```

[1] 0.1345448

```{r}
vif <- vif(weurope_model_generosity)
vif
```

social_support freedom_to_make_life_choices generosity
2.120312 2.524734 1.570089

```{r}
t0 <- 0.3631 / (sqrt(0.1345448 * 1.570089))
t0
```

[1] 0.7900065

```

*Figure 10: t-statistic calculation*

For our t-test with 21 observations, and an alpha level of 0.01 on 17 degrees of freedom, the t-table value is 2.898, which is greater than 0.79. This leads us to fail to reject  $H_0$ , meaning for our analysis, Generosity is not statistically significant in calculating Ladder Score.

## Question 2

### How do social support and freedom to make life choices affect the ladder score in Western European countries?

After figuring out that Generosity was not a factor in the Ladder Score, we removed it as an independent variable and refitted the model in Western Europe with only Social Support and Freedom to make Life Choices as the independent variables when calculating Ladder Score, pictured in Figure 11.

To figure out if this was a valid analysis or not, we tested the regression relationship between Social Support and Freedom to make Life Choices.

$$H_0: \beta_1 = \beta_2$$

$$H_a: \beta_j \neq 0 \text{ for at least one } j = [1, 2]$$

```

Call:
lm(formula = ladder_score ~ social_support + freedom_to_make_life_choices,
 data = weurope)

Residuals:
 Min 1Q Median 3Q Max
-0.7618 -0.1386 0.1121 0.2124 0.4276

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.351 1.620 -1.451 0.16389
social_support 8.224 2.398 3.430 0.00299 **
freedom_to_make_life_choices 2.033 1.180 1.723 0.10206

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3589 on 18 degrees of freedom
Multiple R-squared: 0.7311, Adjusted R-squared: 0.7012
F-statistic: 24.47 on 2 and 18 DF, p-value: 7.349e-06

```

Figure 11: Summary of Western Europe Model

Testing the regression relation, using alpha = 0.01, our test shows that the p-value is much less than 0.01, we reject the null hypothesis that  $\beta_1 = \beta_2$ , so there is a significant relationship between the variables in the linear regression model. Our test confirms that there is a linear relationship between  $\beta_1$  and  $\beta_2$ . Which we would expect since these are factors in calculating the ladder score. Our new linear regression equation after redoing the multi-linear regression model is:

$$y = 8.224x_1 + 2.033x_2 - 2.351$$

where  $x_1 = \text{social support}$  and  $x_2 = \text{freedom to make life choices}$

The interpretation of  $\beta_1$  is, for every one unit increase in someone's rating of the social support metric, their ladder score in Western Europe increases by 8.224.

## Question 3

### How does the generosity of a country affect their perceptions of corruption?

Although in Question 1 it was shown that Generosity does not have any effect on the Ladder Score calculations in Western Europe, we wanted to examine the relationship between Perceptions of Corruption and Generosity, across all regions, as, according to the correlation plot (Figure 2), there seems to be a slight negative correlation between the two variables. In order to investigate whether there is a correlation between Generosity and Perceptions of Corruption, we will perform a simple linear regression analysis where Generosity is the independent variable and Perceptions of Corruption is the dependent variable. In order to test the regression relationship between the two variables, we called the linear model to get  $\beta_0$  and  $\beta_1$ .

```

Call:
lm(formula = perceptions_of_corruption ~ generosity, data = world_happiness_2021)

Residuals:
 Min 1Q Median 3Q Max
-0.64601 -0.05258 0.04683 0.11089 0.24822

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.72450 0.01461 49.600 <2e-16 ***
generosity -0.19505 0.09679 -2.015 0.0457 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1774 on 147 degrees of freedom
Multiple R-squared: 0.02688, Adjusted R-squared: 0.02026
F-statistic: 4.061 on 1 and 147 DF, p-value: 0.04571

```

Figure 12: Summary of Perceptions of Corruption & Generosity World Model

The linear regression equation from this model summary is as follows:

$$y = -0.19505x + 0.72450$$

Next, a hypothesis test was conducted to test the linearity of the relationship between Perceptions of Corruption and Generosity at an alpha risk value of 0.1. The hypothesis was as follows:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

In order to calculate the test statistic,  $t_0$ , the standard error of  $\beta_1$  had to be calculated first, which ended up being 0.09679816. Once that value was obtained, we were able to calculate the test statistic as follows:

```
t_0 = (-0.19505 + 0) / (0.09679816)
t_0
...
[1] -2.015018
```

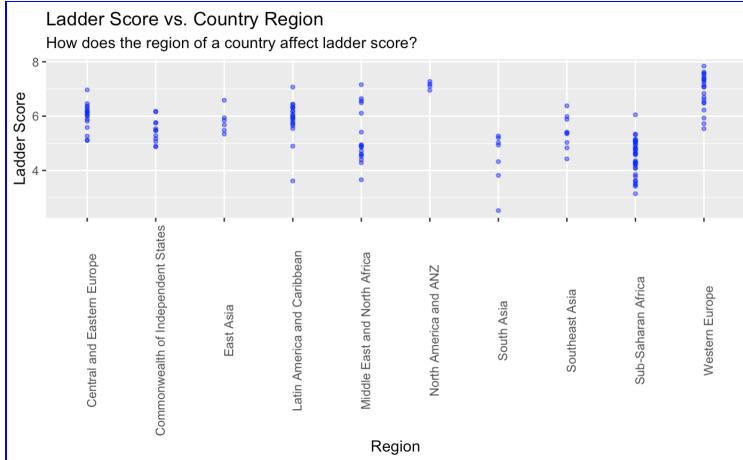
Figure 13: Test Statistic Calculation

From there, we also needed to obtain the value of  $t_{\alpha/2, n-2} = t_{0.1/2, 149-2}$ , which ended up being 1.655285. Finally, we compared the absolute value of  $t_0$  with  $t_{0.1/2, 149-2}$  to decide whether or not the null hypothesis should be rejected. Since  $|-2.015018| > 1.655285$ , we reject the null hypothesis and can therefore say that there is a linear relationship between Generosity and the Perceptions of Corruption. In turn, as displayed in Figure 12, the slope is -0.19505. This means that there is a negative relationship between the two variables and that the Perceptions of Corruption is expected to decrease by 0.19505 as Generosity increases by 1.

## Question 4

**For Western Europe, Latin America & the Caribbean, and Sub-Saharan Africa, which factor has the highest impact on the ladder score of the region?**

To evaluate which factor/independent variable had the highest impact on the Ladder Score of a region, we conducted three simple linear regression analyses on Western European, Latin American & Caribbean, and Sub-Saharan African countries. These three regions were chosen as they are all relatively evenly distributed across the world and because, upon observation of the figure below, Western Europe seemed to have countries with some of the highest ladder scores, Latin America & the Caribbean with relatively average scores, and Sub-Saharan Africa with the lowest ladder scores.



*Figure 14: Ladder Score vs. Region*

First, we evaluated the linear model for Western Europe:

```

Call:
lm(formula = ladder_score ~ logged_gdp_per_capita + social_support +
 healthy_life_expectancy + freedom_to_make_life_choices +
 generosity + perceptions_of_corruption, data = weurope)

Residuals:
 Min 1Q Median 3Q Max
-0.5306 -0.1438 -0.0051 0.1730 0.5004

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.08296 7.64206 -0.011 0.99149
logged_gdp_per_capita 0.37140 0.30832 1.205 0.24833
social_support 6.57583 2.13144 3.085 0.00807 **
healthy_life_expectancy -0.03521 0.09279 -0.379 0.71008
freedom_to_make_life_choices 0.07100 1.24412 0.057 0.95530
generosity 0.17723 0.64510 0.275 0.78754
perceptions_of_corruption -1.00238 0.47709 -2.101 0.05423 .

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3082 on 14 degrees of freedom
Multiple R-squared: 0.8458, Adjusted R-squared: 0.7797
F-statistic: 12.8 on 6 and 14 DF, p-value: 5.594e-05

```

*Figure 15: Western Europe Linear Model Summary*

As displayed under the “Estimate” column in Figure 15, we can see the different beta (slope) values for the relationship between Ladder Score and the other independent variables. Seen above, the highest correlation coefficient for Western European countries and Ladder Score is Social Support; as Social Support increases by 1, the Ladder Score increases by 6.57583.

Second, we took a look at the Latin American & Caribbean countries:

```

Call:
lm(formula = ladder_score ~ logged_gdp_per_capita + social_support +
 healthy_life_expectancy + freedom_to_make_life_choices +
 generosity + perceptions_of_corruption, data = latamerica)

Residuals:
 Min 1Q Median 3Q Max
-0.3983 -0.1434 -0.0241 0.1424 0.5490

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.49714 2.60454 -2.111 0.054749 .
logged_gdp_per_capita 0.02715 0.20799 0.131 0.898157
social_support 0.48564 1.84805 0.263 0.796836
healthy_life_expectancy 0.11484 0.04830 2.378 0.033452 *
freedom_to_make_life_choices 4.13883 0.94984 4.357 0.000776 ***
generosity 0.38819 0.94989 0.409 0.689438
perceptions_of_corruption -0.47465 0.90271 -0.526 0.607874

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3006 on 13 degrees of freedom
Multiple R-squared: 0.8714, Adjusted R-squared: 0.8121
F-statistic: 14.69 on 6 and 13 DF, p-value: 4.078e-05

```

*Figure 16: Latin America & Caribbean Linear Model Summary*

Upon looking at the estimated beta values for this summary, we see that the Freedom to Make Life Choices has the highest value in relation to the Ladder Score; as the Freedom to Make Life Choices increases by 1, the Ladder Score increases by 4.13883.

Lastly, we observed the linear regression model for Sub-Saharan African countries.

```

Call:
lm(formula = ladder_score ~ logged_gdp_per_capita + social_support +
 healthy_life_expectancy + freedom_to_make_life_choices +
 generosity + perceptions_of_corruption, data = ssafrica)

Residuals:
 Min 1Q Median 3Q Max
-1.31892 -0.29280 0.03784 0.43805 0.98049

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.246384 2.074988 0.119 0.906
logged_gdp_per_capita 0.343159 0.211825 1.620 0.116
social_support -1.087070 1.731823 -0.628 0.535
healthy_life_expectancy 0.007795 0.033919 0.230 0.820
freedom_to_make_life_choices 0.757828 1.453040 0.522 0.606
generosity 1.382039 1.006261 1.373 0.180
perceptions_of_corruption 1.608688 1.148309 1.401 0.172

Residual standard error: 0.6356 on 29 degrees of freedom
Multiple R-squared: 0.2196, Adjusted R-squared: 0.05817
F-statistic: 1.36 on 6 and 29 DF, p-value: 0.2636

```

*Figure 17: Sub-Saharan Africa Linear Model Summary*

For the Sub-Saharan African countries, it seems that Perceptions of Corruption has the highest impact on the Ladder Score; when Perceptions of Corruption increase by 1, the Ladder Score increases by 1.608688. However, this seems a bit odd, as one may think

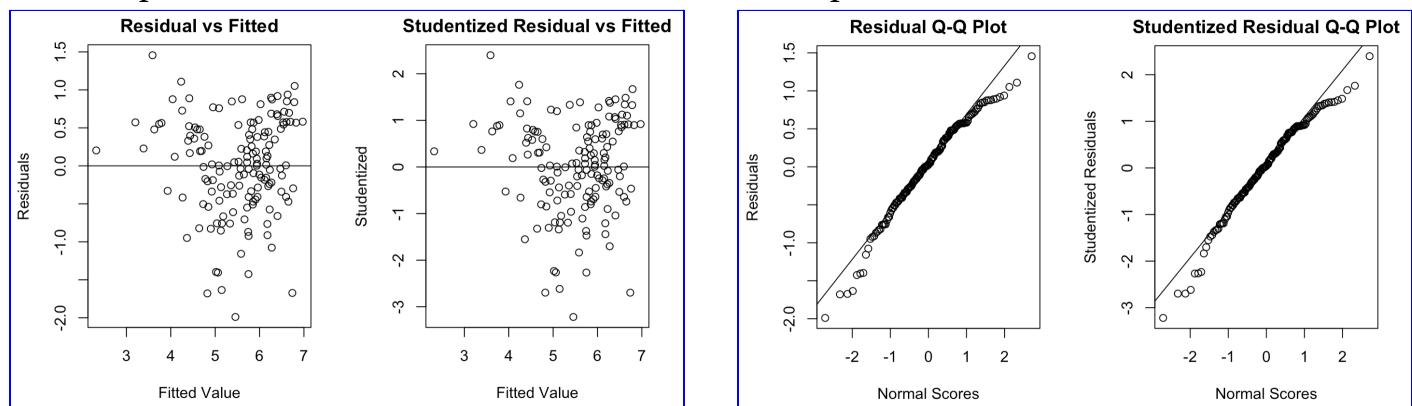
that the more the perception of corruption there is, the less happy people may be. One idea we had that may explain the strange correlation is that, even if there is corruption in the government, it is perceived by the citizens in the country; the corruption is not covered up and citizens feel they understand the on-goings of the government. In turn, citizens may feel they have a better involvement in or understanding of their governmental systems. If they have a higher level of involvement in their government, it may lead citizens to be happier, as it could dispel any feelings of being completely controlled by the government.

In summary, there is no one, certain factor that influences the Ladder Score of regions the most. Around the world, different factors contribute to the happiness of citizens at different levels.

## Question 5

### Which region of the world has the highest ladder score?

To answer this final question, we decided another linear regression would be the best way to go about it. We started by examining the linear model for the countries. As in Question 2, we used Social Support and the Freedom to Make Life Choices as the independent variables and the Ladder Score as the dependent variable.



*Figure 18: Simple Linear Regression for Countries*

Due to the relatively linear relationship between Social Support and the Freedom to Make Life Choices and the Ladder Score, as displayed by the Residual and Studentized Residual Q-Q Plots, we determined that the linearity of the data could be verified and no stabilizing transformations were needed; we could determine the country with the highest Ladder Score without any data modifications. Therefore, the country with the highest Ladder Score/Happiness Score is Finland, with a score of 7.842.

After finding the country with the highest Ladder Score, we needed to determine the region with the highest score. To do this, we sorted our data by region and calculated the average Social Support score, Freedom to Make Life Choices score, and Ladder Score for each region and began this part of our data analysis by determining if a linear

model would be appropriate for the data. Therefore, we have the Fitted and Q-Q Plots displayed below.

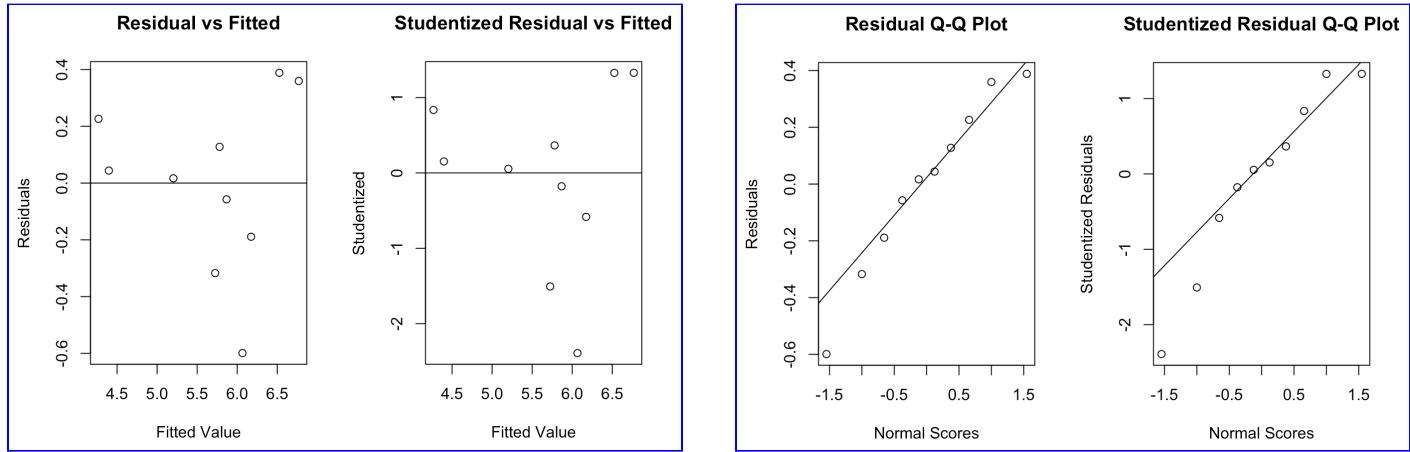


Figure 19: Simple Linear Regression for Regions

As with the simple linear regression plots for countries in Figure 18, the data in the simple linear regression plots for regions also follow a linear pattern on the Q-Q Plots very nicely and the data is randomly distributed on the Fitted Residual plots as well. Therefore, no stabilizing transformations were needed and we were able to determine the region with the highest average ladder score from the plot below.

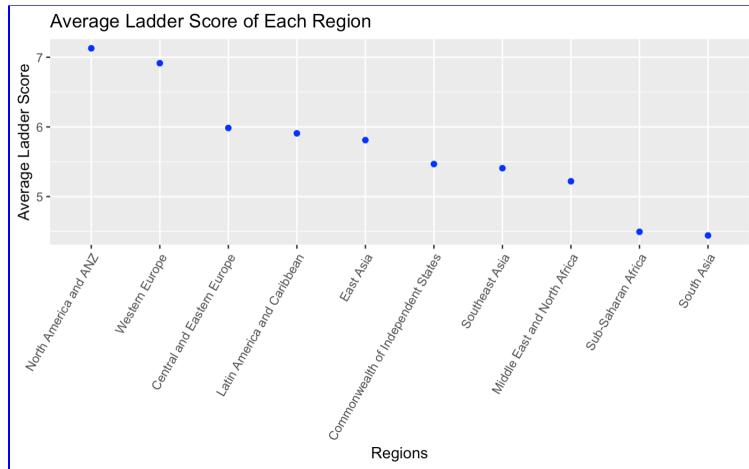


Figure 20: Average Region Ladder Score

As displayed above, the region with the highest average Ladder Score was North America and ANZ with a score of 7.1285 and the region with the lowest average Ladder Score was South Asia with a score of 4.441857.

## Conclusions

There are many factors that go into the calculation of the Ladder Score from the World Happiness 2021 data. One factor that the website reported as a factor influencing

the calculation was the Generosity variable. After further investigation of scatter plots, we tested the coefficient of Generosity's significance in a multiple linear regression in Western Europe, and found that when analyzing Social Support, Freedom to make Life Choices, and Generosity, the coefficient of Generosity was not significant. We removed this metric and dove deeper into how exactly Social Support and Freedom to make Life Choices numerically factors into Western Europe's Ladder Score. Doing a multiple linear regression analysis, we found that for one unit increase in Social Support, the Ladder Score increased by 8.22 units. And for every one unit increase in Freedom to make Life Choices, the Ladder Score increased by 2.033 units. We also explored the most important factors that impacted different regions' Ladder Scores. We investigated the implications that Perceptions of Corruption can have on the Ladder Score, finding the opposite of what we initially believed, which was that happiness would decrease if someone thought their country was corrupt. In conclusion, there was no single metric across all of the regions that contributed most to the overall happiness index.

## **Resources**

Kaggle. (2021). *World Happiness Report 2021*. Retrieved from  
<https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021>

World Happiness Report. (2021). *World Happiness Report 2021*. Retrieved from  
<https://worldhappiness.report/ed/2021/#appendices-and-data>