PSTAT 100 Final Project

Brian Zeng // NetID: brianzeng Jay Udall // NetID: jayudall
Richard Zhao // NetID: richardzhao

Introduction:

In this project, we will be working with data from the World Happiness Report 2023, which includes information related to happiness and well-being by country from 2008-2023.

We want to construct our own "Objective Happiness Index" and ask how well does this MLR model compare to the subjective Life Ladder score across countries in 2022?

Along this report, we will answer the questions:

How much of the variation in Life Ladder scores (2022) can be explained by Log GDP per capita, Social Support, and Healthy Life Expectancy at Birth?

What are the relative contributions of each predictor? Are the objective variables enough to give us a statistically significant answer, or do we need to introduce subjectively gathered data to create a good benchmark?

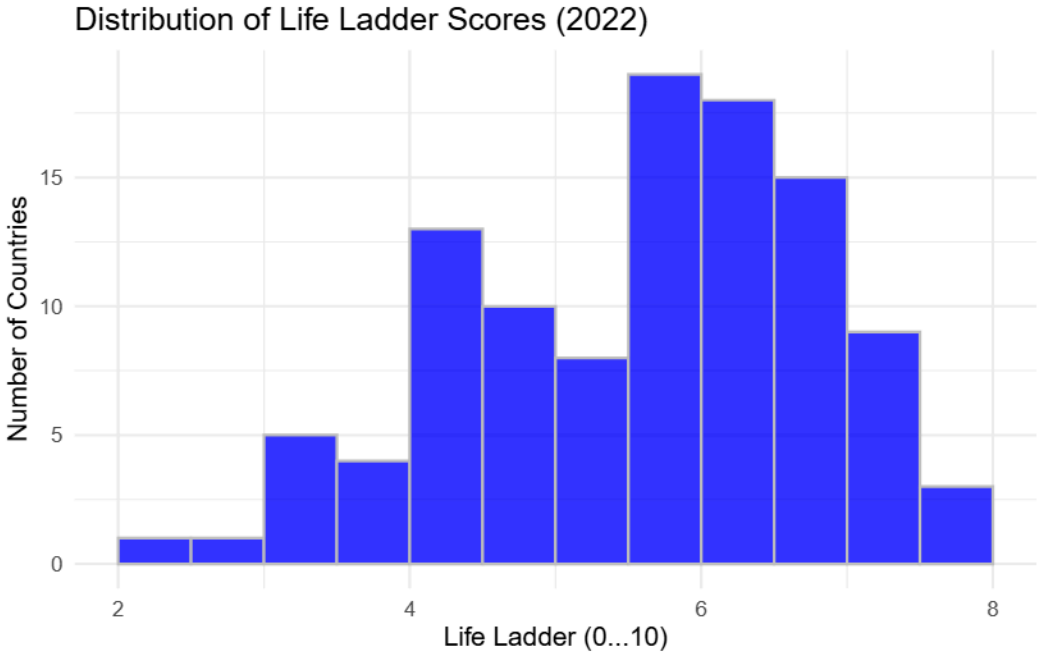
Generally which countries deviate the most from our model expectations??

```
  Country.name year Life.Ladder Log.GDP.per.capita Social.support
1  Afghanistan 2022       1.281                 NA          0.228
2      Albania 2022       5.212              9.626          0.724
3    Argentina 2022       6.261             10.011          0.893
4      Armenia 2022       5.382              9.668          0.811
5    Australia 2022       7.035             10.854          0.942
6      Austria 2022       6.999             10.939          0.876
  Healthy.life.expectancy.at.birth Freedom.to.make.life.choices Generosity
1                           54.875                         0.368         NA
2                           69.175                         0.802     -0.066
3                           67.250                         0.825     -0.128
4                           67.925                         0.790     -0.154
5                           71.125                         0.854      0.153
6                           71.275                         0.856      0.139
```

```
  Perceptions.of.corruption Positive.affect Negative.affect
1                     0.733           0.206           0.576
2                     0.846           0.547           0.255
3                     0.810           0.724           0.284
4                     0.705           0.531           0.549
5                     0.545           0.711           0.244
6                     0.524           0.718           0.226
```

```
[1] 106
```

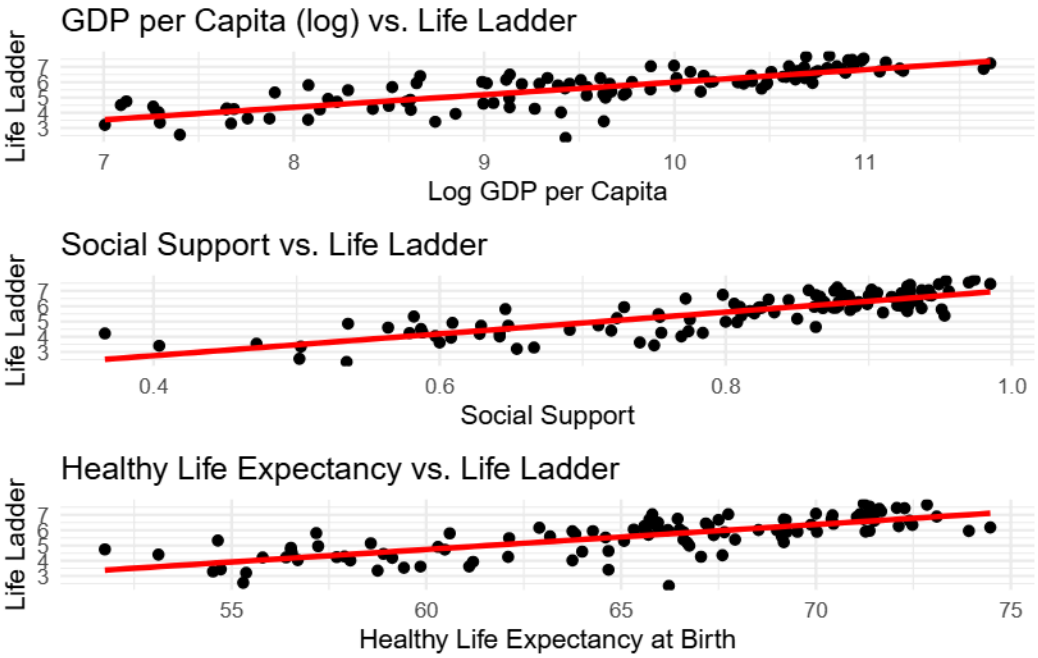
```
  Log.GDP.per.capita_mean Log.GDP.per.capita_sd Social.support_mean
1                9.512217              1.197001           0.7977642
  Social.support_sd Healthy.life.expectancy.at.birth_mean
1         0.1414587                               65.27193
  Healthy.life.expectancy.at.birth_sd Freedom.to.make.life.choices_mean
1                            5.668521                          0.7951415
  Freedom.to.make.life.choices_sd Life.Ladder_mean Life.Ladder_sd
1                       0.1131747         5.596745       1.231311
```



We begin with filtering the WHR2023 dataset to only include countries with complete data for the year 2022. We do this to avoid running into issues with missing data because if we choose to include more years, the amount of missing data only increases, and dropping said missing data would possibly bias our results.

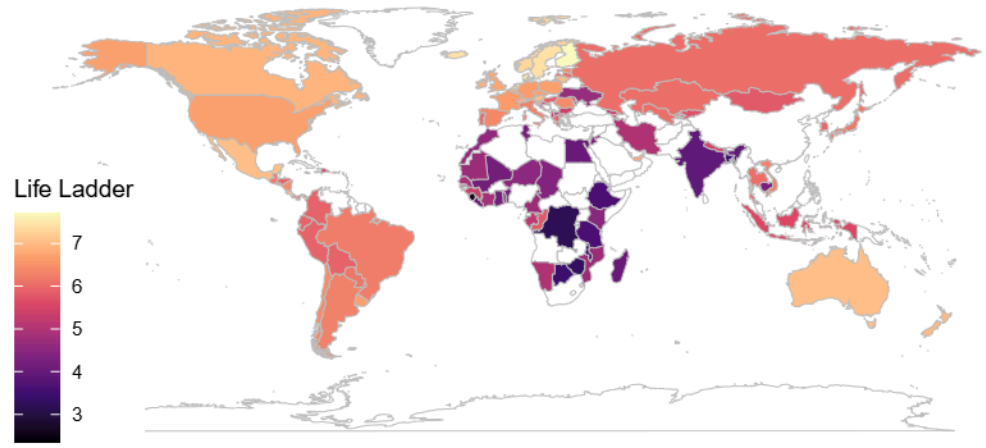
Then, we calculate the necessary summary statistics, being the mean and standard deviation, for each of the variables of interest: Log GDP per capita, Social Support, Healthy Life Expectancy at Birth, and Life Ladder.

A histogram of the Life Ladder scores revealed a slightly left-skewed distribution, with most countries clustering between scores of 4.5 and 7.0. This shows us how the global subjective ratings have moderate to high variability across nations.



To explore the independent relationships between our variables and the Life Ladder, we created scatter plots of each predictor against Life Ladder. All three variables displayed positive linear trends, meaning they show strong signs for predicting happiness. Social Support and GDP per capita in specific showed the most association with Life Ladder.

```
[1] "Congo (Brazzaville)" "Congo (Kinshasa)"    "Czechia"
[4] "United Kingdom"      "United States"
```



Then, we visualize the global distribution of the Life Ladder scores by left joining the WHR2022 dataset with world map data. Country name mismatches such as "Congo (Brazzaville)" had to be amended to make sure everything displayed correctly.

The resulting ggplot depicts Western Europe, North America, and Oceania with the highest happiness scores, while much of Africa and parts of Asia report much lower Life Ladder scores. Depicting the Life Ladder scores helps us visualize the global imbalances in subjective well-being which will come in handy later when we compare our MLR models.

	Log.GDP.per.capita	Social.support
Log.GDP.per.capita	1.0000000	0.7798522
Social.support	0.7798522	1.0000000
Freedom.to.make.life.choices	0.3905048	0.4830239
Healthy.life.expectancy.at.birth	0.8737595	0.7408484

	Freedom.to.make.life.choices
Log.GDP.per.capita	0.3905048
Social.support	0.4830239
Freedom.to.make.life.choices	1.0000000
Healthy.life.expectancy.at.birth	0.3961924

	Healthy.life.expectancy.at.birth
Log.GDP.per.capita	0.8737595
Social.support	0.7408484
Freedom.to.make.life.choices	0.3961924
Healthy.life.expectancy.at.birth	1.0000000

To assess potential multicollinearity between our predictors before fitting the MLR model, we created a cor table with Log GDP per capita, Social Support, Freedom to Make Life Choices, and Healthy Life Expectancy at Birth. Healthy Life Expectancy at Birth in particular showed

major signs of multicollinearity as the correlation between Log GDP per capita and Healthy Life Expectancy was (r = 0.87), followed by the correlation between Social Support and Healthy Life Expectancy (r = 0.74). With correlation ranging from -1 to 1, these values show a very strong relationship and suggest that Healthy Life Expectancy is redundant. On the otherhand, Freedom to Make Life Choices is only somewhat correlated with the other variables (ranging from 0.39 to 0.48), implying that it will provide a distinct contribution to the model.

```
  Country.name year Life.Ladder Log.GDP.per.capita Social.support
1  Afghanistan 2022       1.281                 NA          0.228
2      Albania 2022       5.212              9.626          0.724
3    Argentina 2022       6.261             10.011          0.893
4      Armenia 2022       5.382              9.668          0.811
5    Australia 2022       7.035             10.854          0.942
6      Austria 2022       6.999             10.939          0.876
  Healthy.life.expectancy.at.birth Freedom.to.make.life.choices Generosity
1                           54.875                         0.368         NA
2                           69.175                         0.802     -0.066
3                           67.250                         0.825     -0.128
4                           67.925                         0.790     -0.154
5                           71.125                         0.854      0.153
6                           71.275                         0.856      0.139
  Perceptions.of.corruption Positive.affect Negative.affect
1                     0.733           0.206           0.576
2                     0.846           0.547           0.255
3                     0.810           0.724           0.284
4                     0.705           0.531           0.549
5                     0.545           0.711           0.244
6                     0.524           0.718           0.226
```

```
Call:
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Healthy.life.expectancy.at.birth, data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-2.13205 -0.36082  0.04753  0.41123  1.52047

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -2.51589    0.79320  -3.172  0.00200 **
Log.GDP.per.capita                0.32481    0.11777   2.758  0.00689 **
```

```
Social.support                     4.22241    0.72160   5.851 5.94e-08 ***
Healthy.life.expectancy.at.birth  0.02535    0.02318   1.094  0.27668
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6421 on 102 degrees of freedom
Multiple R-squared:  0.7358,    Adjusted R-squared:  0.728
F-statistic: 94.69 on 3 and 102 DF,  p-value: < 2.2e-16



Call:
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Healthy.life.expectancy.at.birth + Freedom.to.make.life.choices,
    data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-1.69528 -0.30419  0.00809  0.36055  1.60781

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -3.62774    0.75874  -4.781 5.94e-06 ***
Log.GDP.per.capita                 0.33959    0.10715   3.169  0.00202 **
Social.support                     3.26996    0.68649   4.763 6.39e-06 ***
Healthy.life.expectancy.at.birth  0.01868    0.02112   0.884  0.37876
Freedom.to.make.life.choices       2.72475    0.57657   4.726 7.43e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.584 on 101 degrees of freedom
Multiple R-squared:  0.7836,    Adjusted R-squared:  0.7751
F-statistic: 91.45 on 4 and 101 DF,  p-value: < 2.2e-16



Call:
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Freedom.to.make.life.choices, data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-1.89329 -0.31902  0.00765  0.35717  1.57098
```

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -3.1510     0.5331  -5.910 4.56e-08 ***
Log.GDP.per.capita             0.4063     0.0760   5.346 5.52e-07 ***
Social.support                 3.3711     0.6762   4.986 2.54e-06 ***
Freedom.to.make.life.choices   2.7588     0.5747   4.801 5.43e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5833 on 102 degrees of freedom
Multiple R-squared:  0.782, Adjusted R-squared:  0.7756
F-statistic: 121.9 on 3 and 102 DF,  p-value: < 2.2e-16
```

After exploring the Life Ladder variable data extensively, we evaluated three MLR models to identify the most effective combination of predictors for explaining cross-country differences in Life Ladder scores in 2022.

Although the objective variables showed signs of collinearity, we created the first model to include only our objective predictors, including log GDP per capita, Social Support, and Healthy Life Expectancy at Birth. This first MLR produced an adjusted R^2 of 0.728, with GDP and Social Support proving to be statistically significant predictors, as we suspected earlier from the scatter plots ($p < 0.01$). Healthy Life Expectancy on the other hand had a high p-value of 0.277, telling us that it does not contribute much to the model. Using this model as the base line to compare to, we continued on to our other MLR models.

In the second model, we added Freedom to Make Life Choices, a survey based variable that arguably reflects someone's free will, a major factor in happiness. This led to a significant change in adjusted R^2, as it jumped from 0.728 to 0.775. The p-value for Freedom was highly significant with a extremely low p-value ($p < 0.001$). Healthy Life Expectancy still remained statistically insignificant though, which told us that it didn't have much use in our model. While the variable itself is a important factor when it comes to measuring happiness, it's possible that the variable's effects may have already be captured by the other predictors (particularly GDP). Another reason that it ended up being statistically insignificant is that it could've lacked enough variability to make much difference in the model results.

In the third and final model, we kept Log GDP, Social Support, and Freedom to Make Life Choices, while leaving out Healthy Life Expectancy. This MLR model not only maintained the high adjusted R^2 value of 0.775, but also improved residual standard error from Model 1's value of 0.642 on 102 dof, and Model 2's value of 0.584 on 101 dof, to 0.583 on 102 dof.

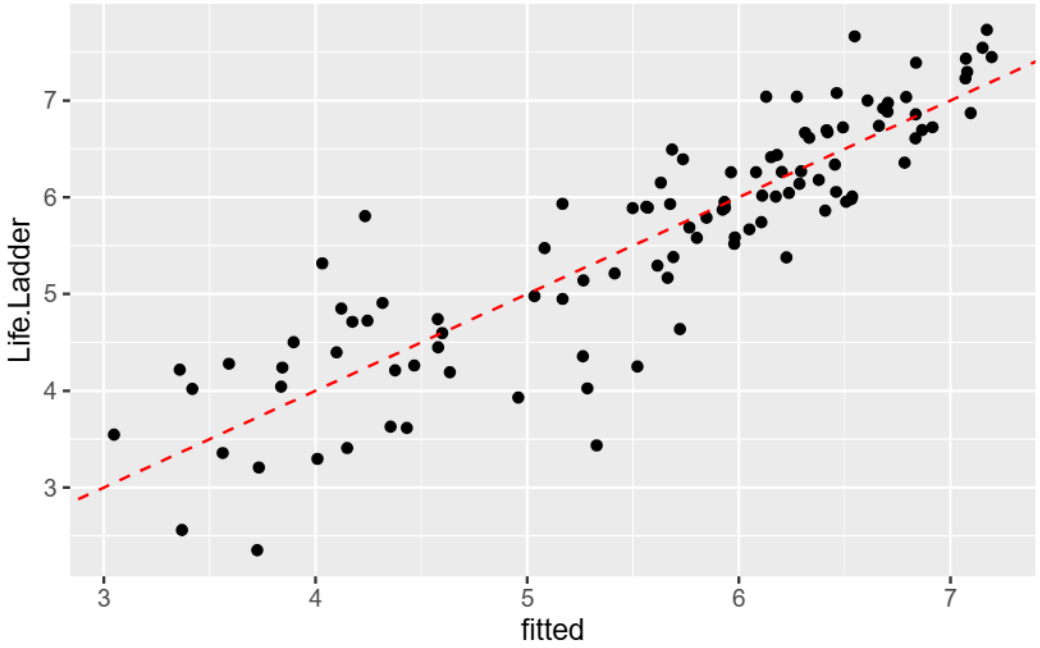
All remaining predictors were statistically significant, with Freedom to Make Life Choices showing a particularly strong relationship with Life Ladder. The results above support the inclusion of Freedom to Make Life Choices as a valuable subjective predictor that complements the objective indicators, and a good replacement to of Healthy Life Expectancy as it resulted

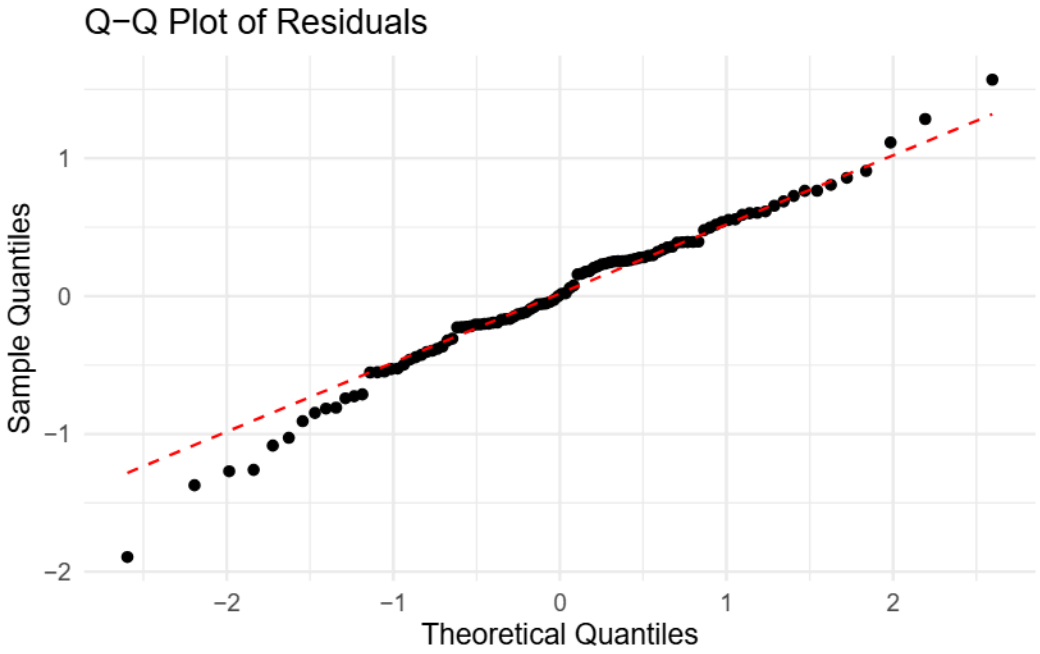
in more statistical contribution. Overall, the third MLR model offers the best balance of objectively gathered data and minimal subjectively gathered data. While not a PCA in the technical sense, this model functions like a conceptual or theoretical reduction of dimensions that influence subjective well-being (Life Ladder). The Life Ladder variable is independently gathered as a survey score, which is not directly calculated from any of the other predictors either objective or subjective, but is strongly influenced by them. All in all, our model provides a strong approximation of happiness levels around the world, based mostly on objective factors than subjective ones.

```
          Log.GDP.per.capita                Social.support
                    2.553747                      2.822942
Freedom.to.make.life.choices
                    1.305141
```

```
   Country.name Log.GDP.per.capita Social.support
1       Albania              9.626          0.724
2     Argentina             10.011          0.893
3       Armenia              9.668          0.811
4     Australia             10.854          0.942
5       Austria             10.939          0.876
6    Bangladesh              8.743          0.404
7       Belgium             10.878          0.923
8         Benin              8.137          0.366
9       Bolivia              9.014          0.824
10     Botswana              9.629          0.750
   Healthy.life.expectancy.at.birth Life.Ladder Freedom.to.make.life.choices
1                            69.175       5.212                         0.802
2                            67.250       6.261                         0.825
3                            67.925       5.382                         0.790
4                            71.125       7.035                         0.854
5                            71.275       6.999                         0.856
6                            64.675       3.408                         0.865
7                            71.050       6.857                         0.890
8                            56.400       4.217                         0.714
9                            63.750       5.929                         0.865
10                           54.725       3.435                         0.739
     fitted
1  5.413231
2  6.202818
3  5.690473
4  6.790510
5  6.608071
```

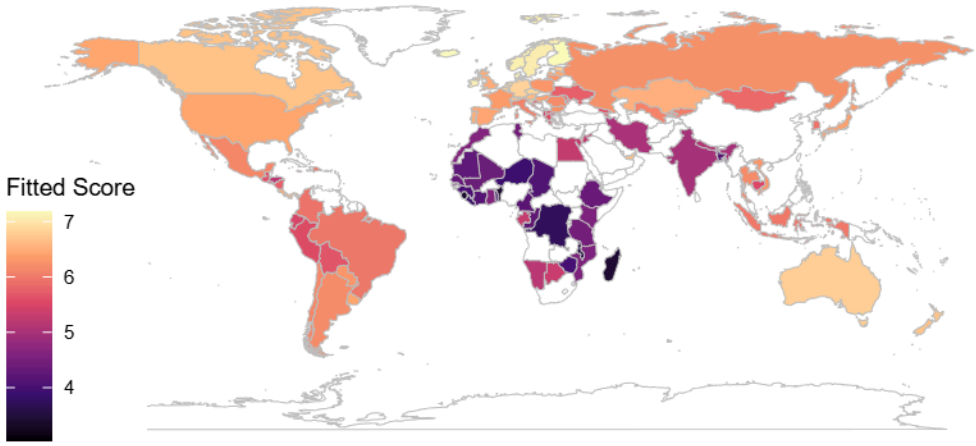
```
6   4.149536
7   6.835528
8   3.358642
9   5.675494
10  5.328293
```





This is a QQPlot, which is used to analyze the distribution of a random variable and its deviation from the normal distribution. It does this by plotting the sample quantities versus the theoretical quantities of if it were to be normal. If our variable followed the normal distribution, the line would be exactly aligned with the dashed red line; however, in our case we can see a deviation at the tails. This tells us that our residuals follow a t-distribution which is has more variation than the normal, and that we violate the Gauss-Markov assumption that the residuals are i.i.d. normal. Our model is doing a good job of explaining the variance, but the estimation of the MLR coefficients could be more precise if the assumptions were valid.

Predicted Life Ladder Scores from MLR Model 3 (2022)



The map using our MLR model shares an extremely similar result to the life ladder map, proving our MLR model with the predictors Log GDP per capita, Social Support, Freedom to Make Life Choice was enough to make a very accurate predictions. Using the Life Ladder map as a baseline, we can see there are slight deviations, India for example, but for the most part the MLR model does a good job predicting happiness levels around the world.

#Appendix

```
library(tidyverse)
library(ggplot2)
library(maps)
library(viridisLite)
library(scales)
library(ggthemes)
library(car)


# DATA READING

whr <- read.csv("whr-2023.csv")
```

```r
years <- sort(unique(whr$year))

whr2022 <- whr %>% filter(year == max(year, na.rm = TRUE))
head(whr2022)
```

```
  Country.name year Life.Ladder Log.GDP.per.capita Social.support
1  Afghanistan 2022       1.281                 NA          0.228
2      Albania 2022       5.212              9.626          0.724
3    Argentina 2022       6.261             10.011          0.893
4      Armenia 2022       5.382              9.668          0.811
5    Australia 2022       7.035             10.854          0.942
6      Austria 2022       6.999             10.939          0.876
  Healthy.life.expectancy.at.birth Freedom.to.make.life.choices Generosity
1                           54.875                         0.368         NA
2                           69.175                         0.802     -0.066
3                           67.250                         0.825     -0.128
4                           67.925                         0.790     -0.154
5                           71.125                         0.854      0.153
6                           71.275                         0.856      0.139
  Perceptions.of.corruption Positive.affect Negative.affect
1                     0.733           0.206           0.576
2                     0.846           0.547           0.255
3                     0.810           0.724           0.284
4                     0.705           0.531           0.549
5                     0.545           0.711           0.244
6                     0.524           0.718           0.226
```

```r
whr2022 <- whr2022 %>%
  select(Country.name, Log.GDP.per.capita, Social.support, Healthy.life.expectancy.at.birth,
         Freedom.to.make.life.choices, Life.Ladder) %>%
  drop_na()

n_countries <- nrow(whr2022)
n_countries
```

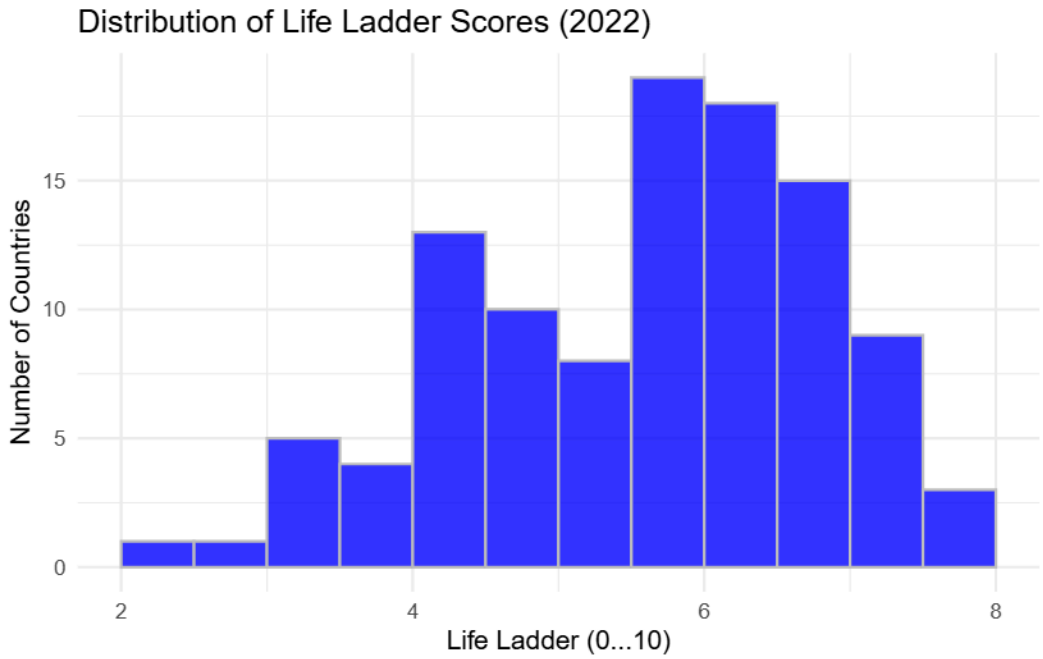
```
[1] 106
```

```r
# HISTOGRAM

whr2022 %>% summarise(across(Log.GDP.per.capita:Life.Ladder, list(mean = mean, sd = sd),
                             na.rm = TRUE))
```

```
  Log.GDP.per.capita_mean Log.GDP.per.capita_sd Social.support_mean
1                9.512217              1.197001           0.7977642
  Social.support_sd Healthy.life.expectancy.at.birth_mean
1         0.1414587                              65.27193
  Healthy.life.expectancy.at.birth_sd Freedom.to.make.life.choices_mean
1                            5.668521                          0.7951415
  Freedom.to.make.life.choices_sd Life.Ladder_mean Life.Ladder_sd
1                       0.1131747         5.596745       1.231311
```

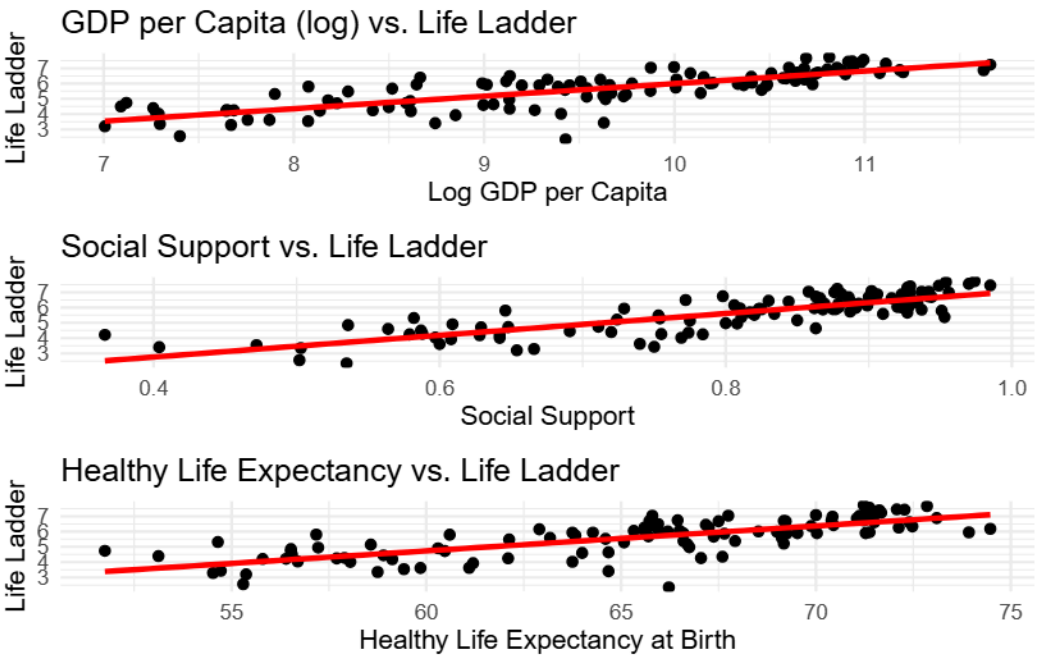
```r
ggplot(whr2022, aes(Life.Ladder)) +
  geom_histogram(binwidth = 0.5, boundary = 0, fill = "blue", color = "gray", alpha = 0.8) +
  labs(title = "Distribution of Life Ladder Scores (2022)",
       x = "Life Ladder (0-10)",
       y = "Number of Countries") +
  theme_minimal(base_size = 10)
```




```r
# OBJECTIVE VARIABLE SCATTER PLOTS

p1 <- ggplot(whr2022, aes(Log.GDP.per.capita, Life.Ladder)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title="GDP per Capita (log) vs. Life Ladder",
       x = "Log GDP per Capita",
```

```
       y = "Life Ladder") +
  theme_minimal(base_size = 10)
p2 <- ggplot(whr2022, aes(Social.support, Life.Ladder)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title="Social Support vs. Life Ladder",
       x = "Social Support",
       y = "Life Ladder") +
  theme_minimal(base_size = 10)
p3 <- ggplot(whr2022, aes(Healthy.life.expectancy.at.birth, Life.Ladder)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title="Healthy Life Expectancy vs. Life Ladder",
       x = "Healthy Life Expectancy at Birth",
       y = "Life Ladder") +
  theme_minimal(base_size = 10)
gridExtra::grid.arrange(p1, p2, p3, ncol = 1)
```




```
# BASELINE LIFE LADDER MAP

world_map <- map_data("world")
map_regions <- unique(world_map$region)
data_regions <- unique(whr2022$Country.name)
```

```r
missing <- setdiff(data_regions, map_regions)
print(missing)
```

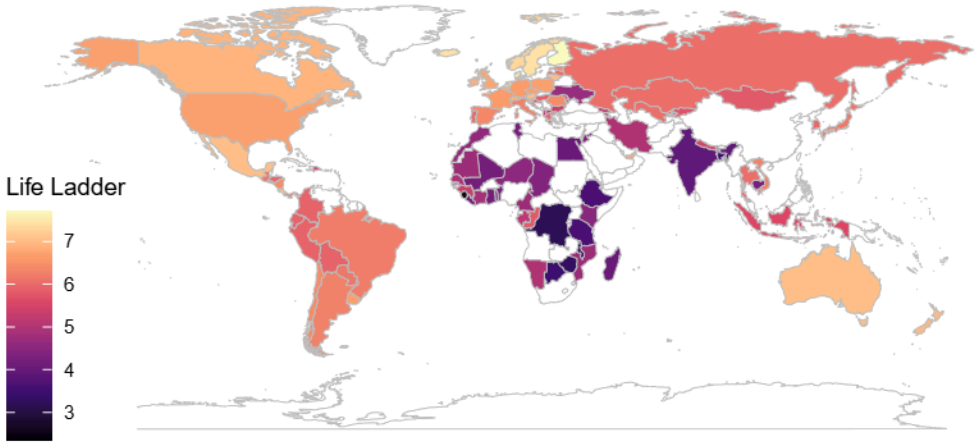
```
[1] "Congo (Brazzaville)" "Congo (Kinshasa)"    "Czechia"
[4] "United Kingdom"      "United States"
```

```r
whr2022_map <- whr2022 %>%
  rename(region = Country.name) %>%
  mutate(
    region = dplyr::recode(region,
      "Congo (Brazzaville)" = "Republic of Congo",
      "Congo (Kinshasa)" = "Democratic Republic of the Congo",
      "Czechia" = "Czech Republic",
      "United Kingdom" = "UK",
      "United States" = "USA"
    )
  )

whr2022_map <- world_map %>%
  left_join(whr2022_map, by = "region")

ggplot(whr2022_map, aes(long, lat, group = group, fill = Life.Ladder)) +
  geom_polygon(color = "gray", size = 0.1) +
  scale_fill_viridis_c(option = "magma", na.value = "white") +
  coord_quickmap() +
  theme_map() +
  labs(
    fill = "Life Ladder",
    title = "Global Distribution of Life Ladder Scores (2022)"
  )
```

Global Distribution of Life Ladder Scores (2022)




```
# COR TABLE

predictors <- whr2022 %>%
  select(Log.GDP.per.capita, Social.support, Freedom.to.make.life.choices,
         Healthy.life.expectancy.at.birth)

cor_matrix <- cor(predictors, use = "pairwise.complete.obs")

print(cor_matrix)
```

```
                                Log.GDP.per.capita Social.support
Log.GDP.per.capita                       1.0000000      0.7798522
Social.support                           0.7798522      1.0000000
Freedom.to.make.life.choices             0.3905048      0.4830239
Healthy.life.expectancy.at.birth         0.8737595      0.7408484
                                Freedom.to.make.life.choices
Log.GDP.per.capita                                 0.3905048
Social.support                                     0.4830239
Freedom.to.make.life.choices                       1.0000000
Healthy.life.expectancy.at.birth                   0.3961924
                                Healthy.life.expectancy.at.birth
Log.GDP.per.capita                                     0.8737595
Social.support                                         0.7408484
Freedom.to.make.life.choices                           0.3961924
Healthy.life.expectancy.at.birth                       1.0000000
```

```r
# MLR MODELS

whr2022 <- whr %>% filter(year == max(year, na.rm = TRUE))
head(whr2022)
```

```
  Country.name year Life.Ladder Log.GDP.per.capita Social.support
1  Afghanistan 2022       1.281                 NA          0.228
2      Albania 2022       5.212              9.626          0.724
3    Argentina 2022       6.261             10.011          0.893
4      Armenia 2022       5.382              9.668          0.811
5    Australia 2022       7.035             10.854          0.942
6      Austria 2022       6.999             10.939          0.876
  Healthy.life.expectancy.at.birth Freedom.to.make.life.choices Generosity
1                           54.875                         0.368         NA
2                           69.175                         0.802     -0.066
3                           67.250                         0.825     -0.128
4                           67.925                         0.790     -0.154
5                           71.125                         0.854      0.153
6                           71.275                         0.856      0.139
  Perceptions.of.corruption Positive.affect Negative.affect
1                     0.733           0.206           0.576
2                     0.846           0.547           0.255
3                     0.810           0.724           0.284
4                     0.705           0.531           0.549
5                     0.545           0.711           0.244
6                     0.524           0.718           0.226
```

```r
#redo whr2022 but this time without dropping Freedom column
#(it may be a significant predictor)
whr2022 <- whr2022 %>%
  select(Country.name, Log.GDP.per.capita, Social.support,
         Healthy.life.expectancy.at.birth, Life.Ladder, Freedom.to.make.life.choices) %>%
  drop_na()

mlr_model1 <- lm(Life.Ladder ~ Log.GDP.per.capita +
                   Social.support +
                   Healthy.life.expectancy.at.birth, data=whr2022)
summary(mlr_model1)
```

```
Call:
```

```
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Healthy.life.expectancy.at.birth, data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-2.13205 -0.36082  0.04753  0.41123  1.52047

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -2.51589    0.79320  -3.172  0.00200 **
Log.GDP.per.capita                 0.32481    0.11777   2.758  0.00689 **
Social.support                     4.22241    0.72160   5.851 5.94e-08 ***
Healthy.life.expectancy.at.birth   0.02535    0.02318   1.094  0.27668
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6421 on 102 degrees of freedom
Multiple R-squared:  0.7358,    Adjusted R-squared:  0.728
F-statistic: 94.69 on 3 and 102 DF,  p-value: < 2.2e-16
```

```r
mlr_model2 <- lm(Life.Ladder ~ Log.GDP.per.capita +
                   Social.support +
                   Healthy.life.expectancy.at.birth +
                   Freedom.to.make.life.choices, data=whr2022)
summary(mlr_model2)
```

```
Call:
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Healthy.life.expectancy.at.birth + Freedom.to.make.life.choices,
    data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-1.69528 -0.30419  0.00809  0.36055  1.60781

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -3.62774    0.75874  -4.781 5.94e-06 ***
Log.GDP.per.capita                 0.33959    0.10715   3.169  0.00202 **
Social.support                     3.26996    0.68649   4.763 6.39e-06 ***
Healthy.life.expectancy.at.birth   0.01868    0.02112   0.884  0.37876
```

```
Freedom.to.make.life.choices      2.72475    0.57657   4.726 7.43e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.584 on 101 degrees of freedom
Multiple R-squared:  0.7836,    Adjusted R-squared:  0.7751
F-statistic: 91.45 on 4 and 101 DF,  p-value: < 2.2e-16
```

```r
#we can see from the regression table that social support, log gdp, and freedom
#to make life choices are statistically significant predictors for life ladder.
#based on healthy life expectancy's p-value of 0.37, we can not confidently assume
#that it adds any value to our model. So can we omit it entirely? To test this,
#we can train the model again and compare the adjusted R^2 values - if it is
#higher with the predictor added, we should keep it in our model.

mlr_model3 <- lm(Life.Ladder ~ Log.GDP.per.capita + Social.support +
                   Freedom.to.make.life.choices, data=whr2022)
summary(mlr_model3)
```

```
Call:
lm(formula = Life.Ladder ~ Log.GDP.per.capita + Social.support +
    Freedom.to.make.life.choices, data = whr2022)

Residuals:
     Min       1Q   Median       3Q      Max
-1.89329 -0.31902  0.00765  0.35717  1.57098

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -3.1510     0.5331  -5.910 4.56e-08 ***
Log.GDP.per.capita             0.4063     0.0760   5.346 5.52e-07 ***
Social.support                 3.3711     0.6762   4.986 2.54e-06 ***
Freedom.to.make.life.choices   2.7588     0.5747   4.801 5.43e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5833 on 102 degrees of freedom
Multiple R-squared:  0.782, Adjusted R-squared:  0.7756
F-statistic: 121.9 on 3 and 102 DF,  p-value: < 2.2e-16
```

```
#the change in the adjusted R^2 value is minimal, so we can assume that healthy life
#expectancy at birth is not significant enough to keep in our model. The value of
#0.7751 tells us that about 78% of the variance in Life Ladder is explained by these
#three predictors.

#the F-statistic value of 0 indicates that we have a statistically significant model
#(there is virtually no chance that all of the true regression coefficients are zero)


# RESIDUALS & QQ PLOT

vif(mlr_model3)
```

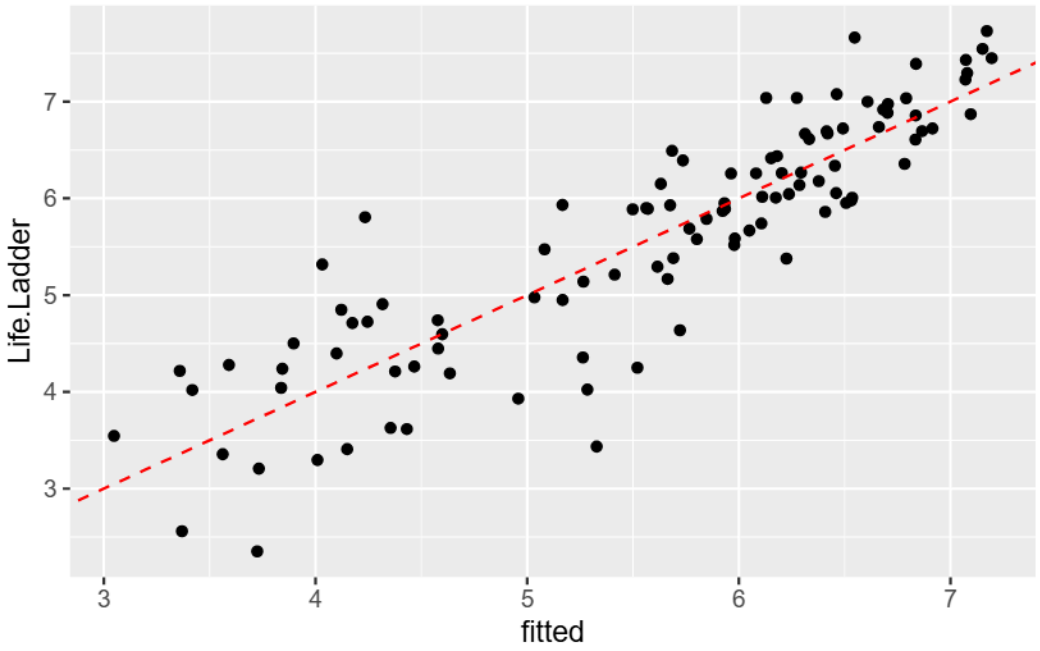
```
          Log.GDP.per.capita                Social.support
                    2.553747                      2.822942
Freedom.to.make.life.choices
                    1.305141
```

```
whr2022$fitted <- predict(mlr_model3)
head(whr2022, 10)
```

```
   Country.name Log.GDP.per.capita Social.support
1       Albania              9.626          0.724
2     Argentina             10.011          0.893
3       Armenia              9.668          0.811
4     Australia             10.854          0.942
5       Austria             10.939          0.876
6    Bangladesh              8.743          0.404
7       Belgium             10.878          0.923
8         Benin              8.137          0.366
9       Bolivia              9.014          0.824
10     Botswana              9.629          0.750
   Healthy.life.expectancy.at.birth Life.Ladder Freedom.to.make.life.choices
1                            69.175       5.212                         0.802
2                            67.250       6.261                         0.825
3                            67.925       5.382                         0.790
4                            71.125       7.035                         0.854
5                            71.275       6.999                         0.856
6                            64.675       3.408                         0.865
7                            71.050       6.857                         0.890
8                            56.400       4.217                         0.714
```

```
9                              63.750          5.929                          0.865
10                             54.725          3.435                          0.739
     fitted
1  5.413231
2  6.202818
3  5.690473
4  6.790510
5  6.608071
6  4.149536
7  6.835528
8  3.358642
9  5.675494
10 5.328293
```

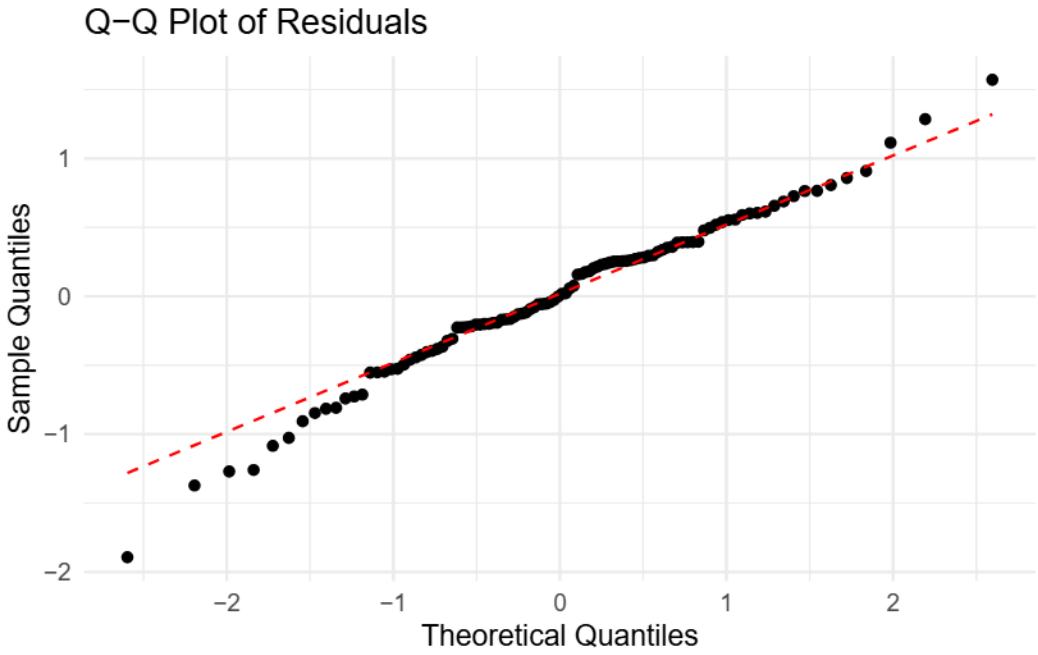
```
ggplot(whr2022,
  aes(x=fitted, y=Life.Ladder)
) +
  geom_point() +
  geom_abline(intercept=0, slope=1, color="red", linetype="dashed")
```




```
resids <- residuals(mlr_model3)

ggplot(data.frame(resids), aes(sample = resids)) +
```

```
  stat_qq() +
  stat_qq_line(color = "red", linetype = "dashed") +
  labs(title = "Q-Q Plot of Residuals",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles") +
  theme_minimal()
```



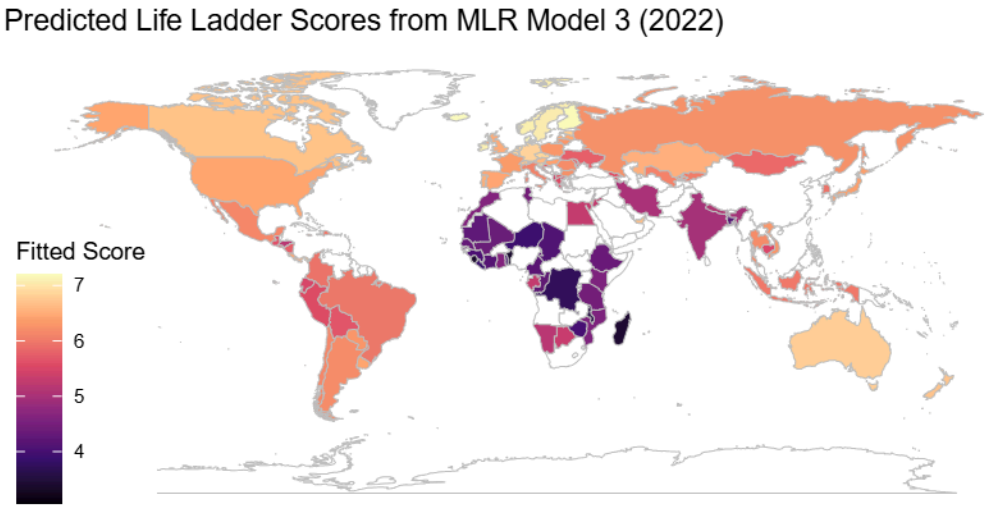

```
# MLR MODEL MAP

whr2022 <- whr2022 %>%
  mutate(fitted = predict(mlr_model3)) %>%
  mutate(region = dplyr::recode(Country.name,
    "Congo (Brazzaville)" = "Republic of Congo",
    "Congo (Kinshasa)" = "Democratic Republic of the Congo",
    "Czechia" = "Czech Republic",
    "United Kingdom" = "UK",
    "United States" = "USA"
  ))

whr2022_map_fitted <- world_map %>%
  left_join(whr2022, by = "region")

ggplot(whr2022_map_fitted, aes(long, lat, group = group, fill = fitted)) +
```

```
  geom_polygon(color = "gray", size = 0.1) +
  scale_fill_viridis_c(option = "magma", na.value = "white") +
  coord_quickmap() +
  theme_map() +
  labs(
    fill = "Fitted Score",
    title = "Predicted Life Ladder Scores from MLR Model 3 (2022)"
  )
```



Sources:

https://stackoverflow.com/questions/7445639/dealing-with-missing-values-for-correlations-calculation

https://sarahpenir.github.io/r/making-maps/

https://ggplot2.tidyverse.org/reference/scale_viridis.html

https://www.statology.org/multiple-r-vs-r-squared/

https://www.rdocumentation.org/packages/car/versions/3.1-3/topics/vif