



WORLD HAPPINESS REPORT

Pre and During COVID-19



Charley Teltschik + Raneem Rahman



Our Dataset: World Happiness Report (2019 & 2020)

- **WHY WE CHOSE IT:**

- To compare happiness levels from pre & during COVID-19 pandemic

2019 → pre-COVID

2020 → during COVID

- **COMPOSITION OF DATASET**

- Countries's levels of happiness based on:
 - Social support
 - Freedom to make life choices
 - Generosity
 - Perception of corruption
 - Healthy life expectancy



Our Dataset: World Happiness Report (2019 & 2020)

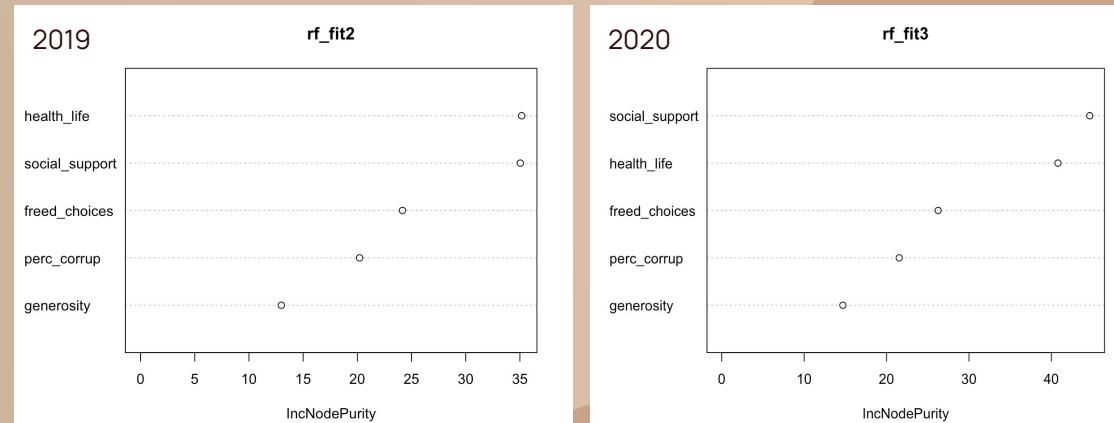
- FACTORS WE LOOKED AT (BASED ON IMPORTANCE):
 - Freedom to make life choices
 - Social support

2019

```
> importance(rf_fit2)
   IncNodePurity
social_support    35.03659
freed_choices    24.16650
generosity       12.98278
perc_corrup      20.21222
health_life      35.16343
```

2020

```
> importance(rf_fit3)
   IncNodePurity
social_support    44.65530
freed_choices    26.25778
generosity       14.69267
perc_corrup      21.53036
health_life      40.79403
> |
```



- IS THERE A SIGNIFICANT DIFFERENCE IN WORLD HAPPINESS BEFORE & DURING COVID-19?
 - Raneem: There is a difference (research to support H_A)
 - Charley: There is not a difference (research to support H_0)
- ●
- ●



Data Cleaning - we both did it :)

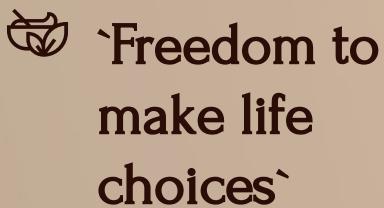
- Made all the chosen column names the same to make the data easier to work with!

```
data2019 <- data2019 %>% rename(`Country` = `Country or region`,  
                                `Happiness score` = Score)  
data2020 <- data2020 %>% rename(`Country` = `Country name`,  
                                `Happiness score` = `Ladder score`)
```

```
# rename columns for 2019 dataset  
data2019 <- data2019 %>%  
  rename(  
    Country = `Country or region`,  
    `Happiness score` = Score,  
    `Healthy life expectancy` = `Healthy life expectancy`  
  )  
  
# rename columns for 2020 dataset  
data2020 <- data2020 %>%  
  rename(  
    Country = `Country name`,  
    `Happiness score` = `Ladder score`,  
    `Healthy life expectancy` = `Healthy life expectancy`,  
    `Freedom to make life choices` = `Freedom to make life choices`  
  )
```



Important Variables



For each country:
National avg. of binary answers where “Yes” = 1 and “No” = 0

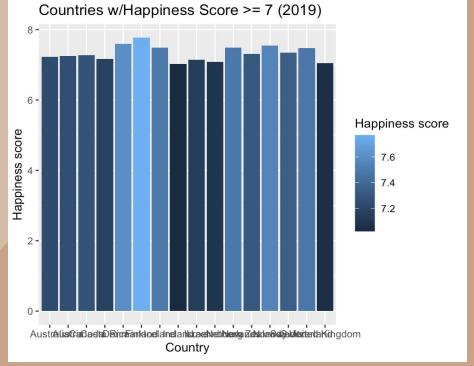
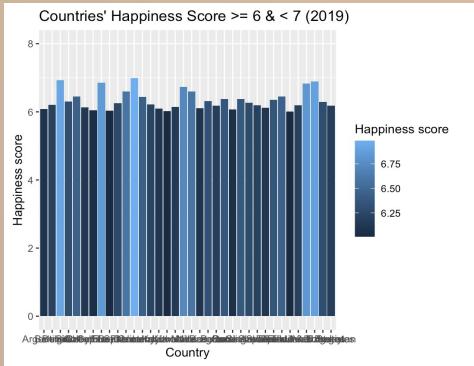
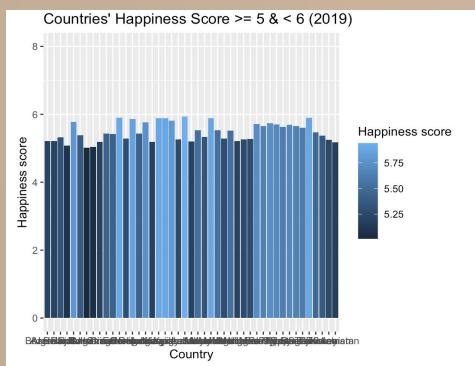
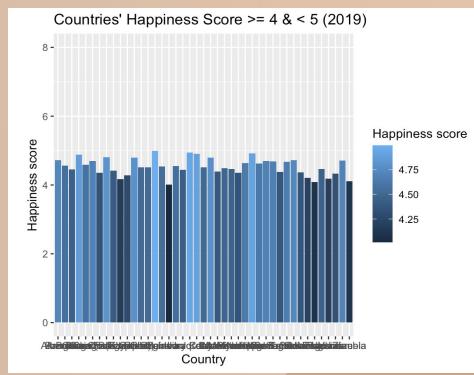
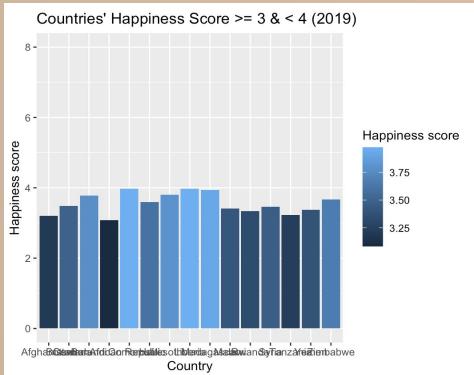
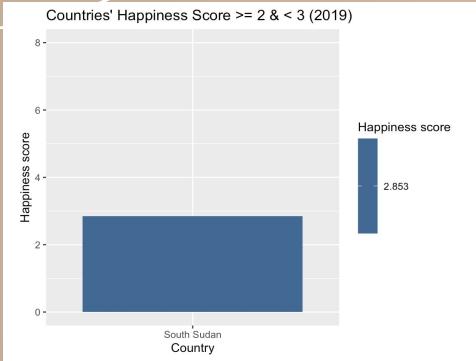


For each country:
National avg. of binary answers where “Yes” = 1 and “No” = 0

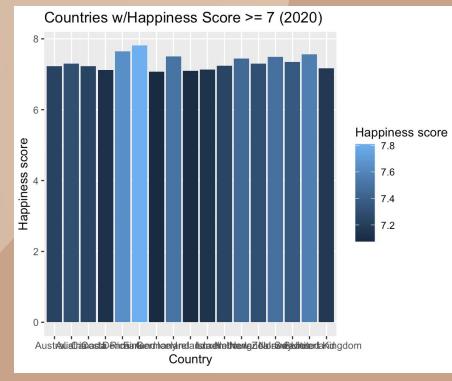
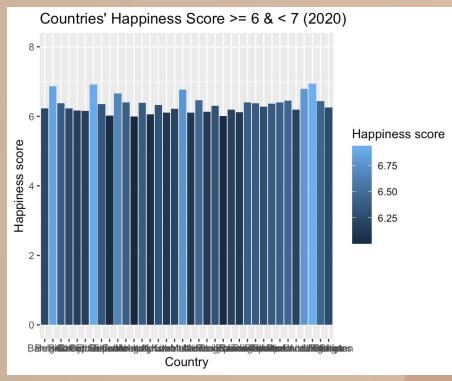
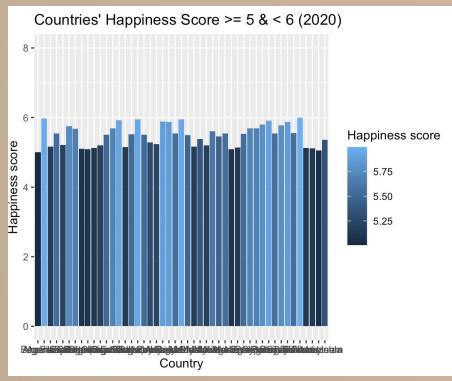
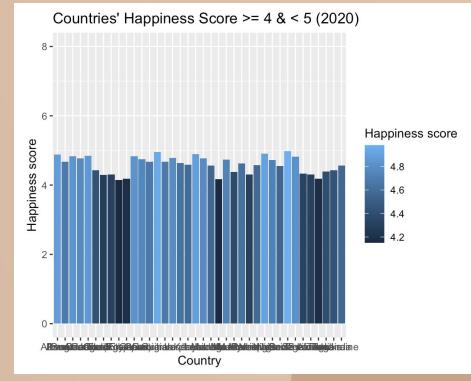
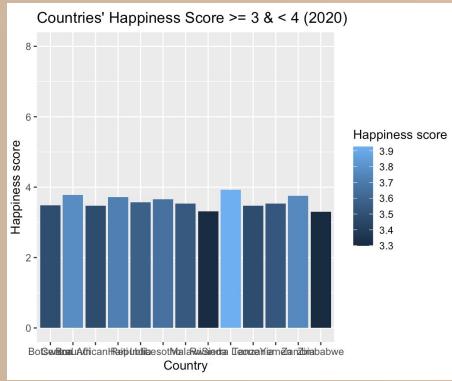
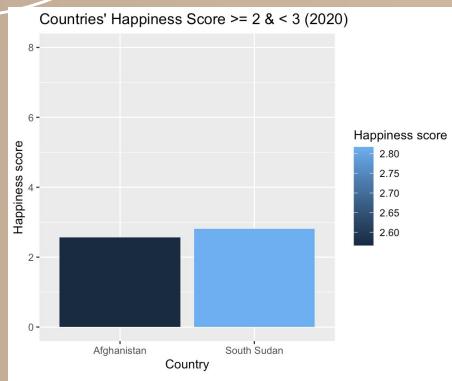


A rating out of a total of 10 to describe the amount of “bliss” the people in each country lived in.

Descriptive Plots (2019)

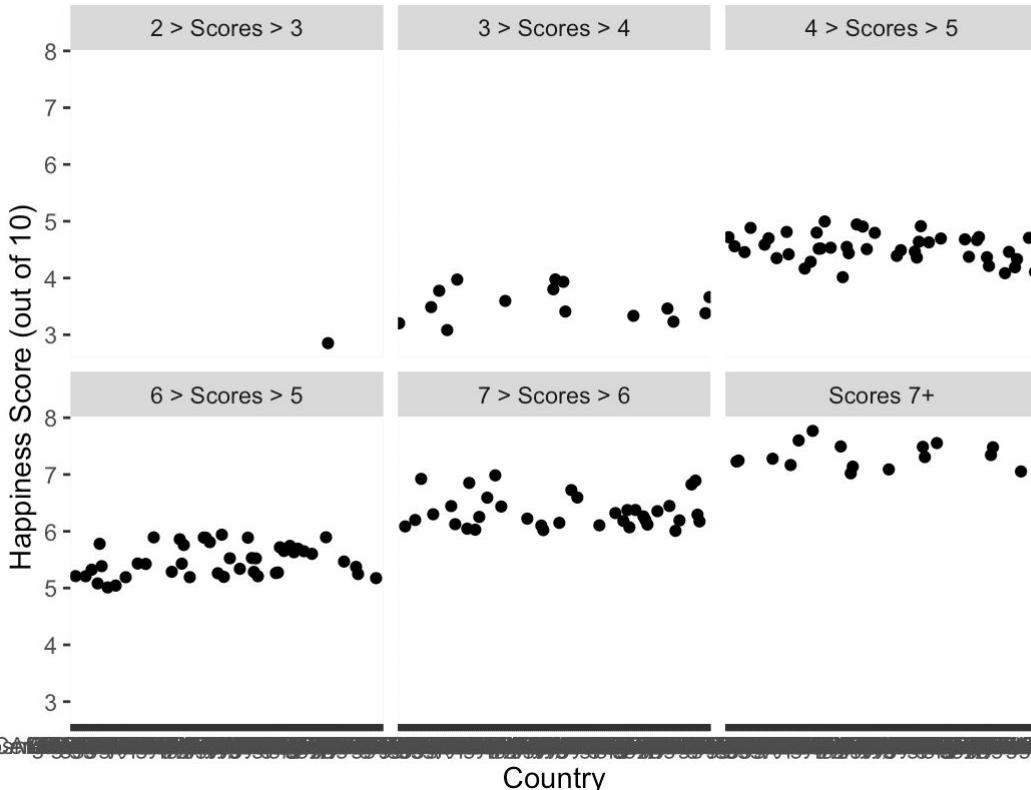


Descriptive Plots (2020)

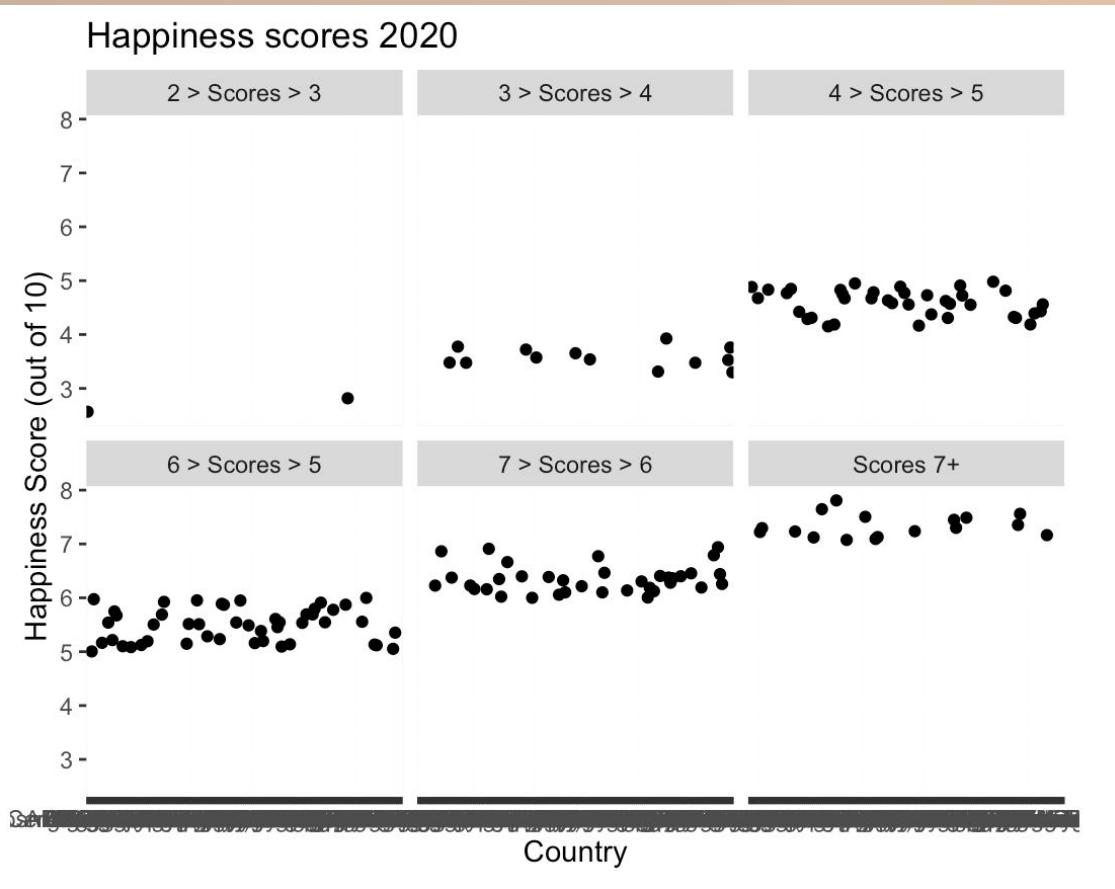


Another View: Descr. Plots (2019)

All Happiness Scores 2019



Another View: Descr. Plots (2020)





H_A : There is a difference in world happiness
between the years 2019 and 2020

MODEL 1 - Linear Regression

2019 Social Support

The 2019 (pre-covid) Model suggests that there is a statistically significant positive relationship between “Happiness score” and “Social Support”

- Our R-squared is 0.6038 which is a good indication of how well the independent variable (social support) explains the dependent variable (happiness score). This R-squared value means that 60.38% of the variability in “Happiness score” can be explained by “Social support” in the model.
- The coefficients reveal that for every one-unit increase in “Social support,” the happiness score is expected to increase by 2.8910 units. The intercept of 1.9124 represents the estimated happiness score when “Social support” is zero.

```
# multiple linear regression model for 2019 dataset
model_2019 <- lm(`Happiness score` ~ `Social support`, data = data2019)
summary(model_2019)

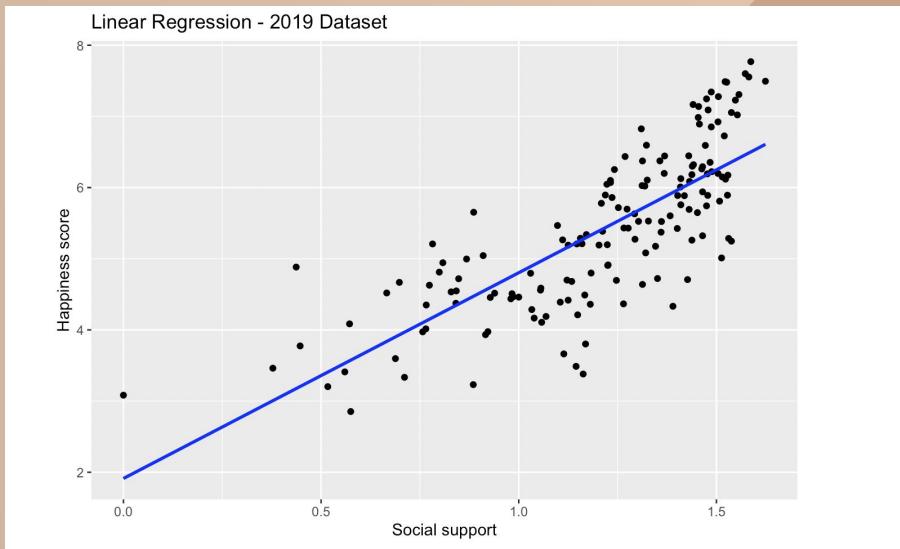
##
## Call:
## lm(formula = `Happiness score` ~ `Social support`, data = data2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.89465 -0.45762 -0.01993  0.54720  1.70721 
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  1.9124    0.2349   8.14 1.25e-13 ***
## `Social support` 2.8910    0.1887  15.32 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7029 on 154 degrees of freedom
## Multiple R-squared:  0.6038, Adjusted R-squared:  0.6012 
## F-statistic: 234.7 on 1 and 154 DF,  p-value: < 2.2e-16
```



MODEL 1 - Analysis

2019 Social Support

- Here we can see the relationship between Social support and Happiness score, and each point represents a country.
- The regression equation:
 - $\text{Happiness Score} = 1.9124 + 2.8910 \times \text{Social Support}$
- The slope of the line is given by the coefficient for Social support (2.8910), indicating the average change in Happiness score for a one-unit increase in Social support.
- As we can see here, the happiness clearly rises as social support increases.



```
library(ggplot2)

# plot for 2019 dataset
plot_2019 <- ggplot(data2019, aes(x = `Social support`, y = `Happiness score`)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Linear Regression - 2019 Dataset",
       x = "Social support",
       y = "Happiness score")
```

MODEL 2 - Linear Regression

2020 Social Support

The 2020 (during covid) Model suggests that there is a statistically significant positive relationship between “Happiness score” and “Social Support”

- Our R-squared is 0.5852. This means that approximately 58.52% of the variability in “Happiness score” can be explained by the “Social support” variable.
- The very low p-value for Social support suggests that the relationship is statistically significant
- An increase in “Social support” is associated with an increase in the “Happiness score.” For every one unit increase in social support, the happiness score is expected to increase by approximately 7.0059 units

```
# multiple linear regression model for 2020 dataset
model_2020 <- lm(`Happiness score` ~ `Social support`, data = data2020)
summary(model_2020)

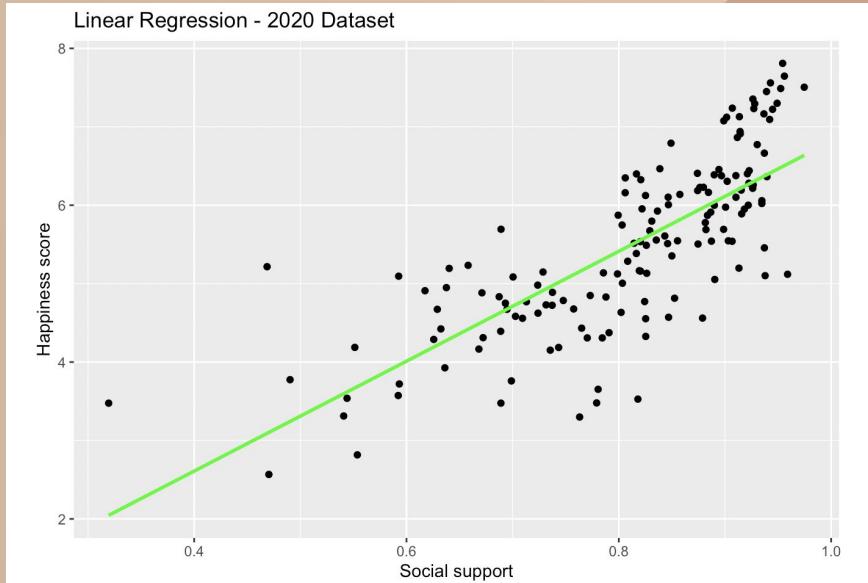
##
## Call:
## lm(formula = `Happiness score` ~ `Social support`, data = data2020)
##
## Residuals:
##      Min        1Q    Median        3Q       Max 
## -2.01071 -0.38261 -0.04146  0.46455  2.12511 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -0.1926     0.3925  -0.491   0.624    
## `Social support`       7.0059     0.4800  14.596  <2e-16 ***
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.7187 on 151 degrees of freedom
## Multiple R-squared:  0.5852, Adjusted R-squared:  0.5825 
## F-statistic: 213.1 on 1 and 151 DF,  p-value: < 2.2e-16
```



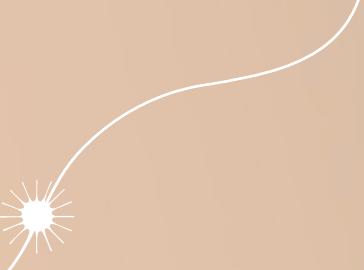
MODEL 2 - Analysis

2020 Social Support

- Here we can see the relationship between Social support and Happiness score
- The regression equation:
 - $\text{Happiness score} = -0.1926 + 7.0059 \times \text{Social support}$
- The slope of this line (7.0059) represents the average change in Happiness score for a one-unit increase in social support
- In this model, compared to the other one, we can see that the social support need has raised, and we still continue to see the same trend that is when social support gets higher, happiness scores do as well

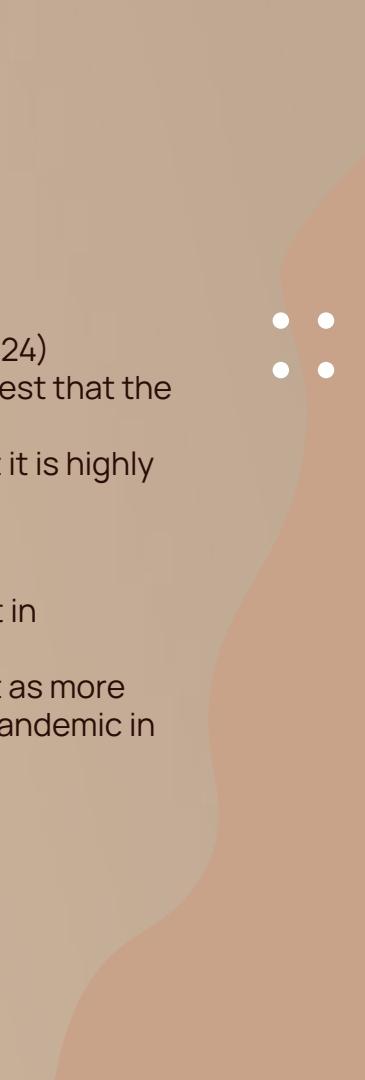


```
# plot for 2020 dataset
plot_2020 <- ggplot(data2020, aes(x = `Social support`, y = `Happiness score`)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "green") +
  labs(title = "Linear Regression - 2020 Dataset",
       x = "Social support",
       y = "Happiness score")
```



Comparison

Social Support, 2019 vs 2020:

- The intercept in the 2020 Model is negative (-0.1926), whereas in 2019, it is positive (1.9124)
 - The social support coefficient in 2020 (7.0059) is higher than in 2019 (2.8910). This suggest that the impact of social support on each country's happiness score is larger in the 2020 data set
 - Both models have extremely low p-values for the 'Social support' variable, indicating that it is highly significant in predicting happiness scores
 - The higher coefficient suggests that on average, the social support had a greater impact in predicting happiness scores during 2020
 - This could be interpreted as an indication that people may have perceived social support as more crucial or beneficial during the challenges and uncertainties associated with COVID-19 pandemic in 2020 compared to the pre-COVID period in 2019
- 

MODEL 3 - Linear Regression

2019 Freedom to make Life Choices

The 2019 (pre covid) Model suggests that there is a significant positive relationship between "Freedom to make life choices" and "Happiness score"

- The r-squared is 0.3212 which means that approximately 32.12% of the the variability in the "Happiness score" can be explained by the variable "Freedom to make life choices"
- The p-value (1.238e-14) is extremely small, which shows evidence against the null hypothesis
- For every one-unit increase in "Freedom to make life choices", its associated with a 4.4026 unit increase in the predicted happiness score

```
library(ggplot2)

# multiple linear regression model for 2019 dataset
model_2019_freedom <- lm(`Happiness score` ~ `Freedom to make life choices`, data = data2019)
summary(model_2019_freedom)

##
## Call:
## lm(formula = `Happiness score` ~ `Freedom to make life choices`,
##     data = data2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7882 -0.5838  0.0149  0.7029  1.8269
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.6788    0.2155 17.075 < 2e-16 ***
## `Freedom to make life choices` 4.4026    0.5158  8.536 1.24e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9201 on 154 degrees of freedom
## Multiple R-squared:  0.3212, Adjusted R-squared:  0.3168
## F-statistic: 72.87 on 1 and 154 DF,  p-value: 1.238e-14
```

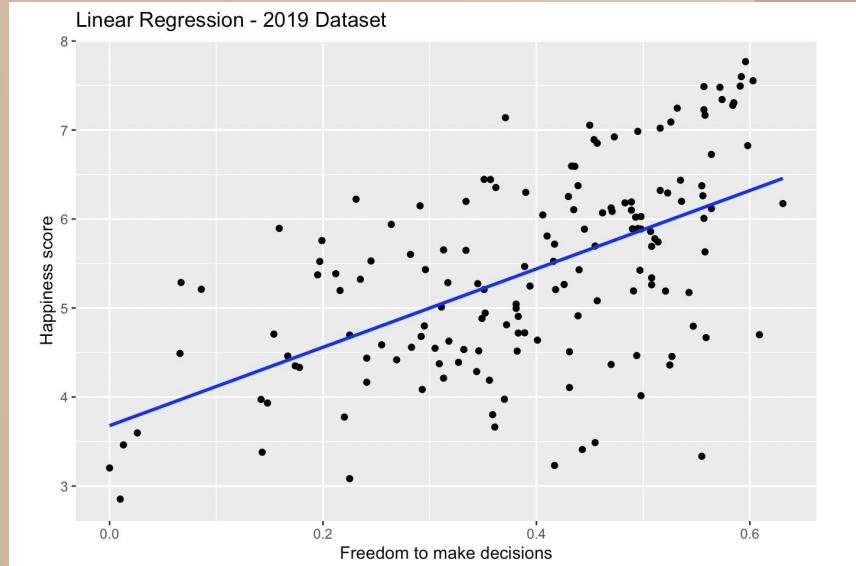


MODEL 3 - Analysis

2019 Freedom to make Life Choices

Here we can see the relationship between Freedom to make life choices and Happiness score, and each point represents a country

- The regression equation:
 - Happiness score = $3.6788 + 4.4026 * \text{Freedom to make life choices}$
- The slope of the line for “Freedom to make life choices” is approximately 4.4026 which means that, on average, for every one-unit increase in the “Freedom to make life choices”, the “Happiness score” is estimated to increase by 4.4026 units
- Freedom to make life choices is associated with a higher happiness score



```
# plot for 2019 dataset
plot_2019_freedom <- ggplot(data2019, aes(x = `Freedom to make life choices`, y = `Happiness score`)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Linear Regression - 2019 Dataset",
       x = "Freedom to make decisions",
       y = "Happiness score")
```

MODEL 4 - Linear Regression

2020 Freedom to make Life Choices

The 2020 (during covid) Model suggests that there is a positive and significant relationship between “Freedom to make choices” and “Happiness score”, which shows that increased freedom is associated with higher happiness scores.

- The R-Squared is 0.3488. This means that 34.88% of the variability in “Happiness score” is explained by the variable “Freedom to make life choices” in the model
- The F-statistic is 80.88 with a very low p-value (9.343e-16), which suggests that the overall model is statistically significant.
- For every one unit increase in Freedom to make life choices, the happiness score is expected to increase by 5.5771 units

```
# multiple linear regression model for 2020 dataset
model_2020_freedom <- lm(`Happiness score` ~ `Freedom to make life choices`, data = data2020)
summary(model_2020_freedom)

##
## Call:
## lm(formula = `Happiness score` ~ `Freedom to make life choices`,
##     data = data2020)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.81474 -0.54132  0.03592  0.62210  1.85491 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                1.1044    0.4912   2.248   0.026 *  
## `Freedom to make life choices` 5.5771    0.6201   8.993 9.34e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.9005 on 151 degrees of freedom
## Multiple R-squared:  0.3488, Adjusted R-squared:  0.3445 
## F-statistic: 80.88 on 1 and 151 DF,  p-value: 9.343e-16
```

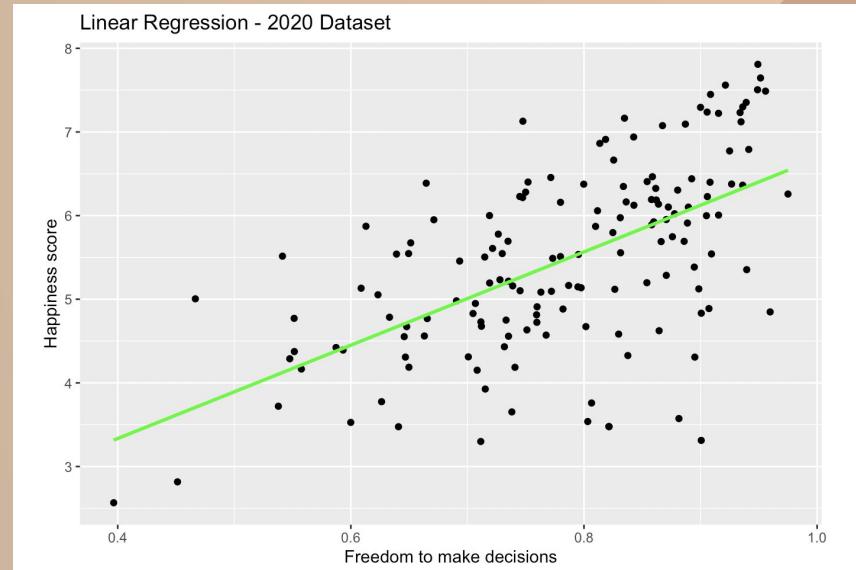


MODEL 4 - Analysis

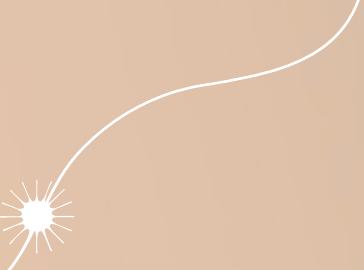
2020 Freedom to make Life Choices

Here we can see the relationship between Freedom to make life choices and Happiness score, and each point represents a country

- The regression equation:
 - Happiness score = $1.1044 + 5.5771 * \text{Freedom to make life choices}$
- The slope is 5.5771, which means that for one unit increase in the freedom to make life choices, the happiness score is expected to increase by 5.5771 units
- This model can explain approximately 34.88% of the variability in the happiness scores
- The coefficient for Freedom is statistically significant with a very low p-value ($9.34e-16$)

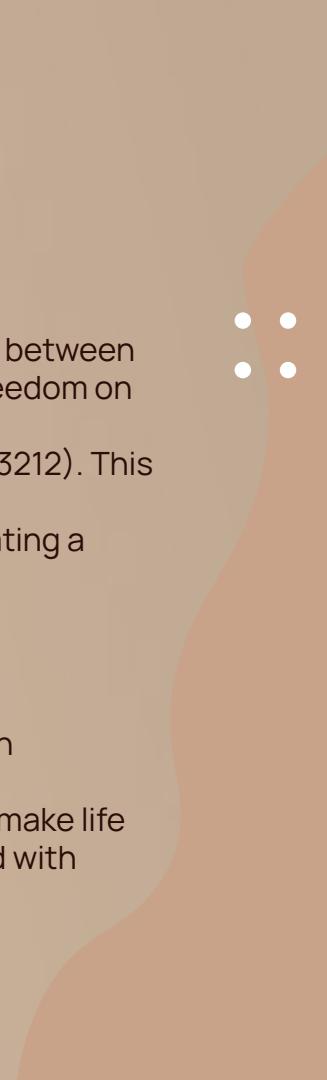


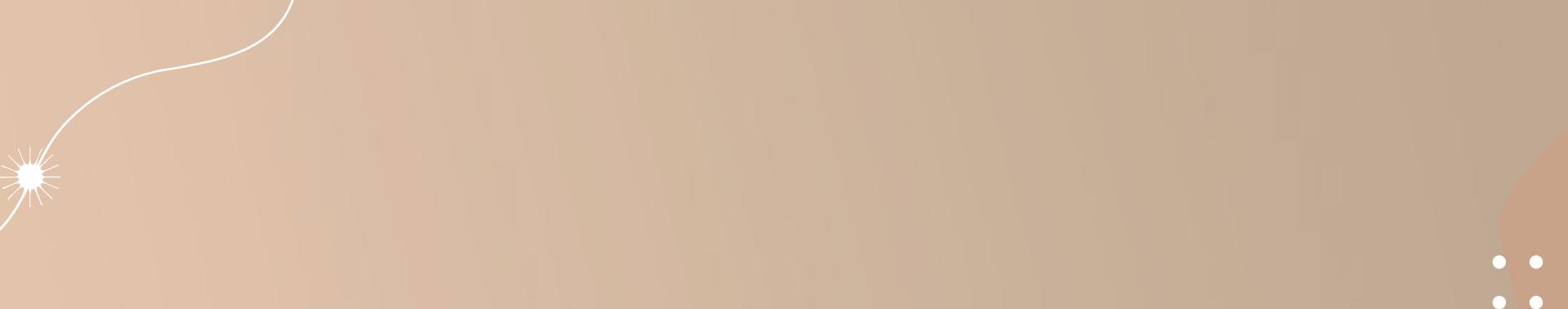
```
# plot for 2020 dataset
plot_2020_freedom <- ggplot(data2020, aes(x = `Freedom to make life choices`, y = `Happiness score`)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "green") +
  labs(title = "Linear Regression - 2020 Dataset",
       x = "Freedom to make decisions",
       y = "Happiness score")
```



Comparison

Freedom to make Life Choices, 2019 vs 2020:

- The intercepts and coefficients for the “Freedom to make life choices” variable are different between the two years. This indicates changes in the baseline happiness score and the impact of freedom on happiness
 - The R-squared value is higher for the 2020 model (0.3488) compared to the 2019 model (0.3212). This indicates that the 2020 model explains more variance in the happiness score
 - The F-Statistic is higher for the 2020 model (80.88) compared to 2019 model (72.87), indicating a better overall fit in 2020
-
- The higher coefficient suggests that on average, the social support had a greater impact in predicting happiness scores during 2020, but not by much
 - This could be interpreted as an indication that people may have perceived the freedom to make life choices a bit more crucial or beneficial during the challenges and uncertainties associated with COVID-19 pandemic in 2020 compared to the pre-COVID period in 2019
- 



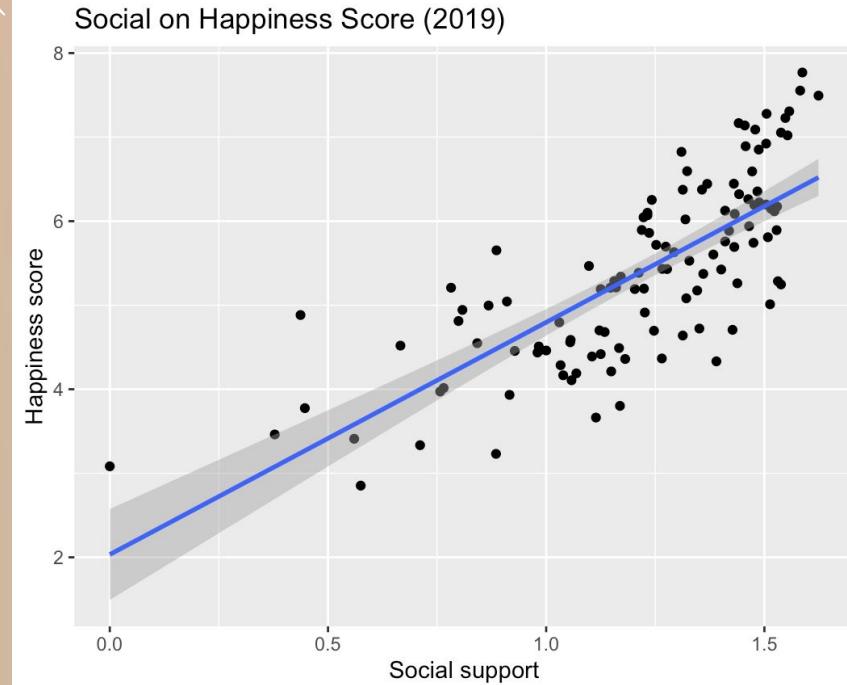
H_0 : There is NOT a difference in world happiness
or a relationship between variables in the years
2019 and 2020

MODEL 1

Social Support

2019

- ROOT MEAN SQUARED ERROR:
 - RMSE values 4.20 & 4.18 > 1 = bad
 - Inaccurate
 - Model could not find a solution
 - OR model is not a better fit w/ `Social support`
- R² VALUE
 - 68% explained variability
 - generally 70% = acceptable



```
get_rmse(data2019_train$`Social support`, preds_train)
## [1] 4.203189
```

```
get_rmse(data2019_test$`Social support`, preds_test)
## [1] 4.181755
```

MODEL 3

Freedom to Make Life Choices 2019

- ROOT MEAN SQUARED ERROR:
 - RMSE values 5.0 & 5.04 > 1 = bad
 - Inaccurate
 - Model could not find a solution
 - OR model is not a better fit w/ `Social support`

- R² VALUE
 - 32.12% explained variability
 - generally 70% = acceptable



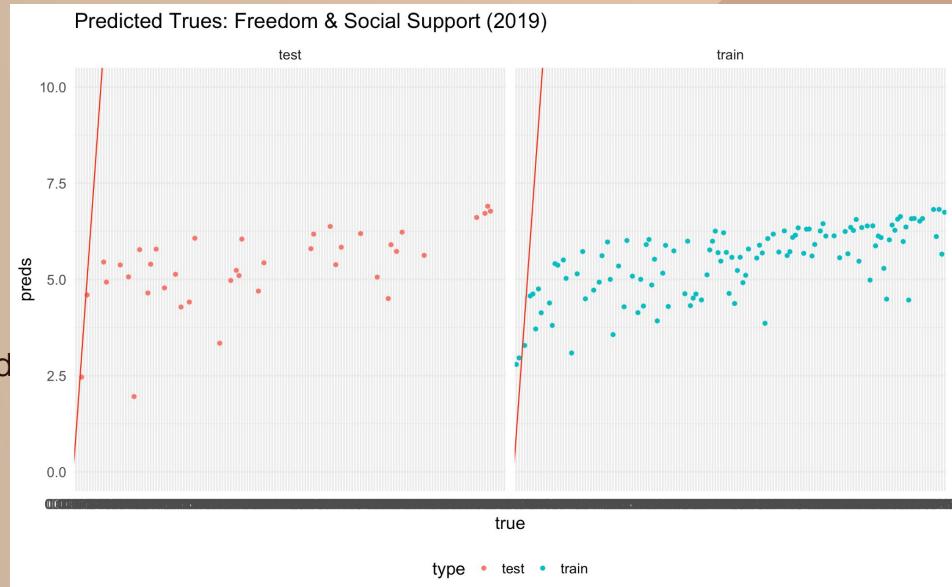
```
get_rmse(data2019_train$`Freedom to make life choices`, preds_train)
## [1] 5.008146
```

```
get_rmse(data2019_test$`Freedom to make life choices`, preds_test)
## [1] 5.047816
```

MODEL 1

Social Support & Freedom of Choice 2019

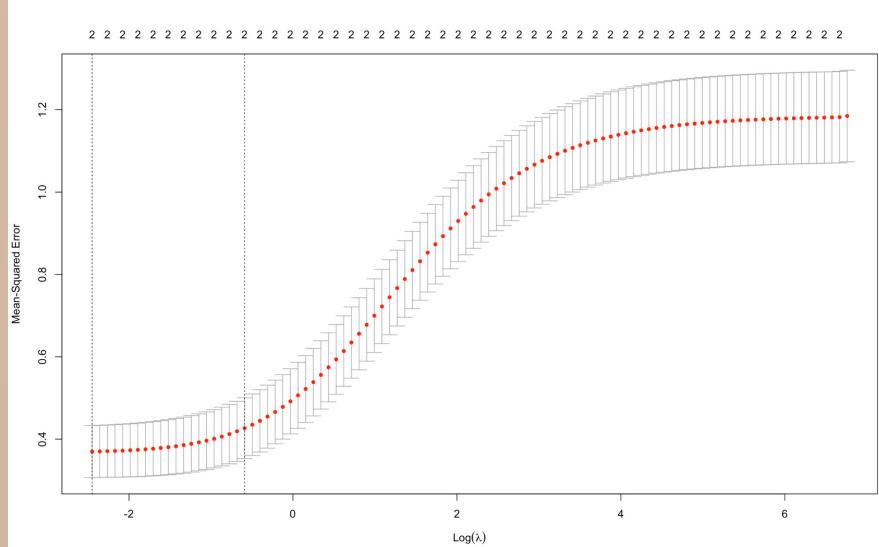
- **PREDICTED TRUE PLOTS:**
 - Train and test data is similar but model may be missing the target and cannot be fit to the data
 - Not totally off so there is some connection, but could definitely be way better



MODEL 1

Social Support & Freedom of Choice 2019

- **LAMBDA = PENALTY TERM**
 - The best lambdas to use are either around -2.3 or -0.8
 - The model uses lambdas 2.3 and 2.7
- **WOULD LIKE TO ADJUST THIS TO LOWER THE ERROR**



(Intercept)	1.698029
social_support	2.367287
freed_choices	2.171513

MODEL 2

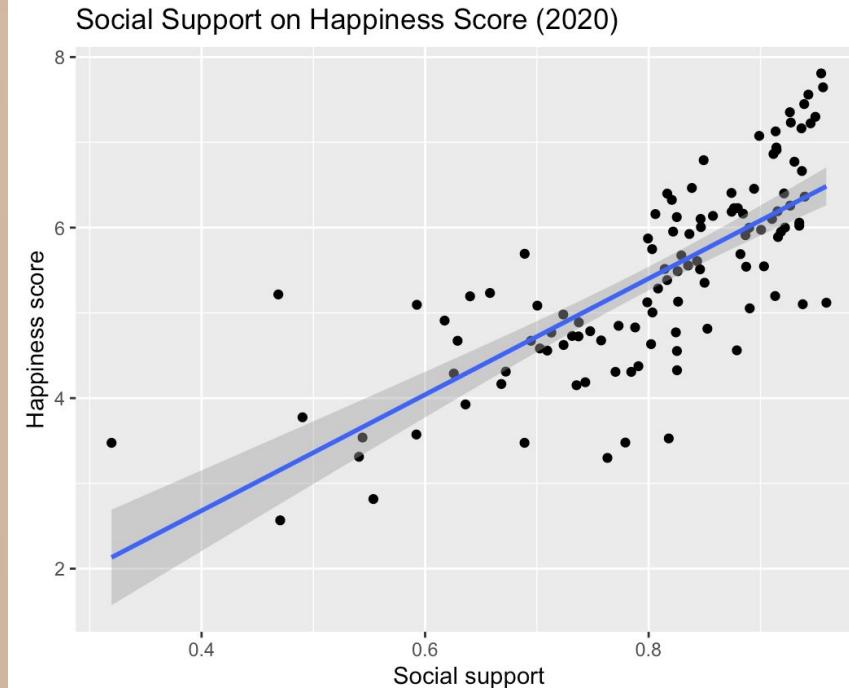
Social Support

2020

- ROOT MEAN SQUARED ERROR:
 - RMSE: 4.68 & 4.77 > 1 = bad
 - Inaccurate
 - Model could not find a solution
 - OR model is not a better fit w/ `Social support`

- R² VALUE
 - 58% explained variability
 - generally 70% = acceptable

• •



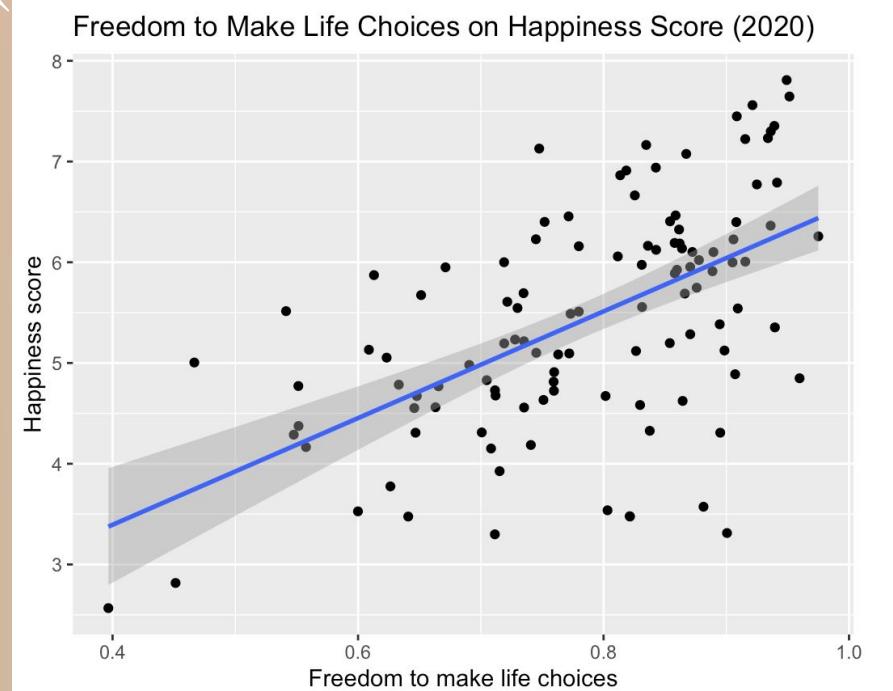
```
get_rmse(data2020_train$`Social support`, preds_train)
## [1] 4.682092
```

```
get_rmse(data2020_test$`Social support`, preds_test)
## [1] 4.77548
```

MODEL 4

Freedom to Make Life Choices 2020

- ROOT MEAN SQUARED ERROR:
 - RMSE values 4.67 & 4.64 > 1 = bad
 - Inaccurate
 - Model could not find a solution
 - OR model is not a better fit w/ `Social support`
 - R² VALUE
 - 34.8% explained variability
 - generally 70% = acceptable
- • •



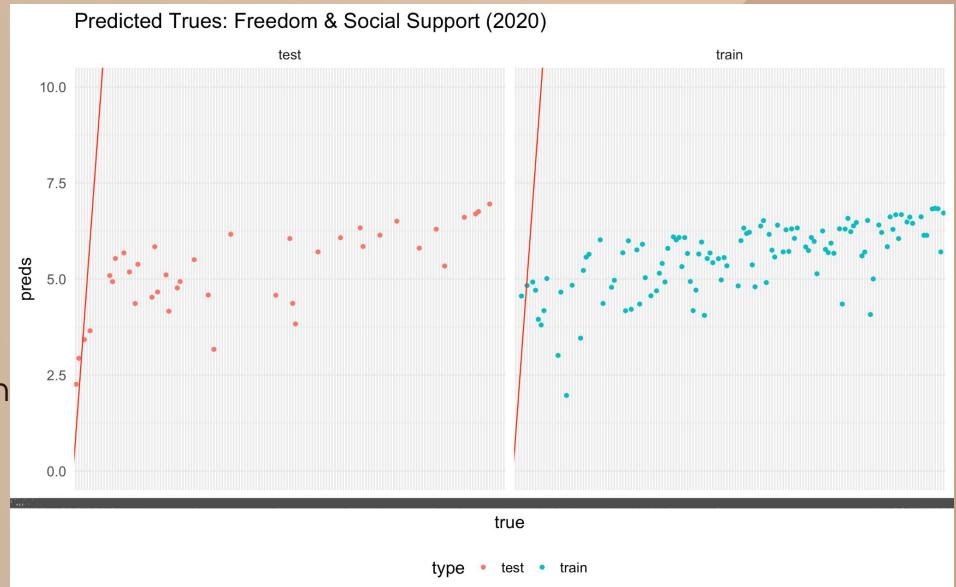
```
get_rmse(data2020_train$`Freedom to make life choices`, preds_train)
## [1] 4.674911
```

```
get_rmse(data2020_test$`Freedom to make life choices`, preds_test)
## [1] 4.648242
```

MODEL 4

Social Support & Freedom of Choice 2020

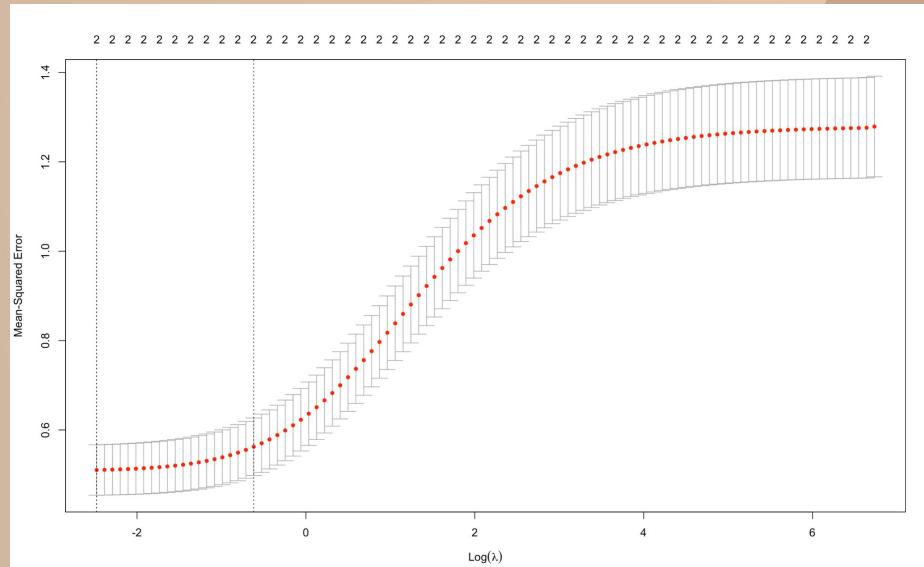
- **PREDICTED TRUE PLOTS:**
 - Train and test data is similar but model may be missing the target and cannot be fit to the data
 - Not totally off so there is some connection, but could definitely be way better



MODEL 4

Social Support & Freedom of Choice 2020

- LAMBDA = PENALTY TERM
 - The best lambdas to use are either around -2.3 or -0.8
 - The model uses lambdas 5.8 and 2.7
- WOULD LIKE TO ADJUST THIS TO LOWER THE ERROR



	s1
(Intercept)	-1.420530
social_support	5.881256
freed_choices	2.706273

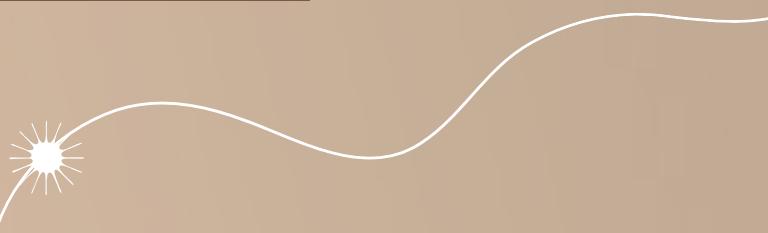


Conclusion

We **FAIL TO REJECT** the null hypothesis due to errors, inaccuracy, underfit models, and untrustworthy results.

We cannot say there is a significant difference in world happiness levels before (2019) and during (2020) the COVID-19 pandemic.

o o o o
o o o o



RESOURCES

- <https://www.kaggle.com/datasets/mathurinache/world-happiness-report?select=2022.csv>
- <https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf?fbclid=IwAR2lh19hPo2qPnQGphe3b5T8f0IX4vjslG5yMgx9FlVYThSJS7QTYbNtkCQ>
- OUR GITHUB LINK:
https://github.com/cteltschik/MGSC_310_FinalProject



Thank You!

