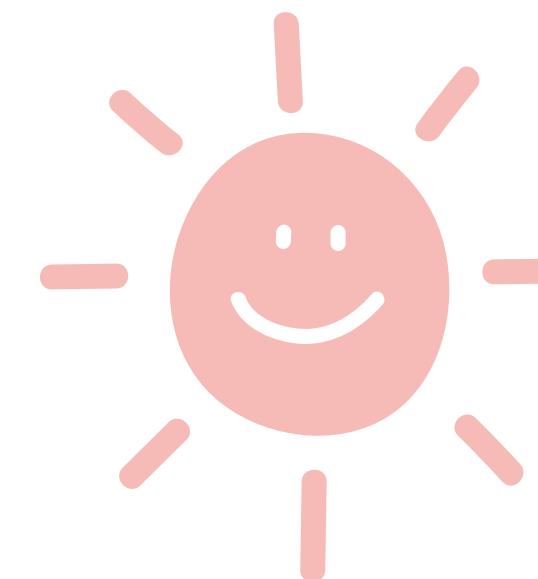


Happy is everything

HAPPINESS-SCORE

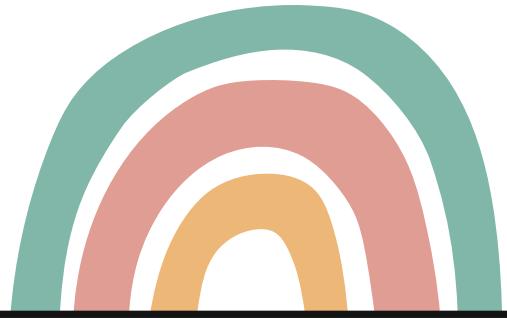
STAT-614

Anna and Qintao



Happiness

ABOUT DATA



Source: Sazidthe1. (2023). World Happiness Report (till 2023) [Data set]. Kaggle.
(<https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors>),
Data Originally from <https://worldhappiness.report/>

The World Happiness Report (2020)

- a. 153 unique observations
- b. 9 columns

i. country

ii. region

iii. happiness_score

iv. gdp_per_capita

v. social_support

vi. healthy_life_expectancy

vii. freedom_to_make_life_choice

viii. generosity

ix. perceptions_of_corruption

The World Happiness Report (2015)

- a. 158 unique observations
- b. 9 columns

Happiness

ABOUT DATA



Variable	Description	Type	Unit
country	Name of the country	Categorical	N/A
region	Region the country belongs to	Categorical	N/A
happiness_score	Avg. life evaluation score (Cantril ladder: 0-10 scale)	Quantitative	Score (0-10)
gdp_per_capita	Log of GDP per capita (PPP, constant 2021 intl. dollars)	Quantitative	Log(USD, PPP-adjusted)
social_support	Prop. answering “yes” to: someone to rely on in times of trouble?	Quantitative (Proportion)	Proportion (0-1)
healthy_life_expectancy	Life expectancy at birth adjusted for health quality	Quantitative	Years
freedom_to_make_life_choices	Prop. satisfied with freedom to choose life activities	Quantitative (Proportion)	Proportion (0-1)
generosity	Residual from donation behavior regressed on GDP per capita	Quantitative	Index (no unit)
perceptions_of_corruption	Prop. believing corruption is widespread	Quantitative (Proportion)	Proportion (0-1)

Happiness

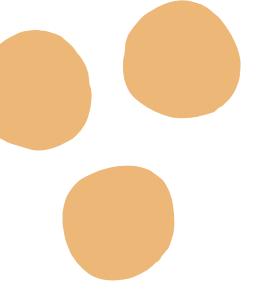


COUNTRY GROUPS

- Central and Eastern Europe (Albania; Armenia; Azerbaijan; Belarus; Bosnia and Herzegovina; Bulgaria; Croatia; Czech Republic; Estonia; Georgia; Hungary; Kosovo; Latvia; Lithuania; Moldova; Montenegro; North Macedonia; Poland; Romania; Russia; Serbia; Slovakia; Slovenia; Ukraine)
- Commonwealth of Independent States (Kazakhstan; Kyrgyzstan; Tajikistan; Turkmenistan; Uzbekistan)
- East Asia (China; Hong Kong S.A.R. of China; Japan; Mongolia; South Korea; Taiwan Province of China)
- Latin America and Caribbean (Argentina; Bolivia; Brazil; Chile; Colombia; Costa Rica; Dominican Republic; Ecuador; El Salvador; Guatemala; Haiti; Honduras; Jamaica; Mexico; Nicaragua; Panama; Paraguay; Peru; Trinidad and Tobago; Uruguay; Venezuela)
- Middle East and North Africa (Algeria; Bahrain; Egypt; Iran; Iraq; Israel; Jordan; Kuwait; Lebanon; Libya; Morocco; Palestinian Territories; Saudi Arabia; Tunisia; Turkey; United Arab Emirates; Yemen)
- North America and ANZ (Australia; Canada; New Zealand; United States)
- South Asia (Afghanistan; Bangladesh; India; Maldives; Nepal; Pakistan; Sri Lanka)
- Southeast Asia (Cambodia; Indonesia; Laos; Malaysia; Myanmar; Philippines; Singapore; Thailand; Vietnam)
- Sub-Saharan Africa (Benin; Botswana; Burkina Faso; Burundi; Cameroon; Central African Republic; Chad; Comoros; Congo (Brazzaville); Congo (Kinshasa); Ethiopia; Gabon; Gambia; Ghana; Guinea; Ivory Coast; Kenya; Lesotho; Liberia; Madagascar; Malawi; Mali; Mauritania; Mauritius; Mozambique; Namibia; Niger; Nigeria; Rwanda; Senegal; Sierra Leone; South Africa; South Sudan; Swaziland; Tanzania; Togo; Uganda; Zambia; Zimbabwe)
- Western Europe (Austria; Belgium; Cyprus; Denmark; Finland; France; Germany; Greece; Iceland; Ireland; Italy; Luxembourg; Malta; Netherlands; North Cyprus; Norway; Portugal; Spain; Sweden; Switzerland; United Kingdom)

Happiness

RESEARCH QUESTIONS

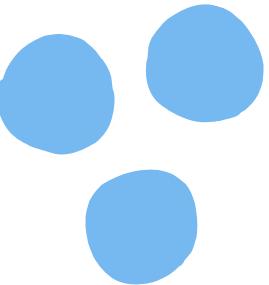


Q1: Do the happiness scores of people on different continents in 2020 differ significantly (**ANOVA Test**)

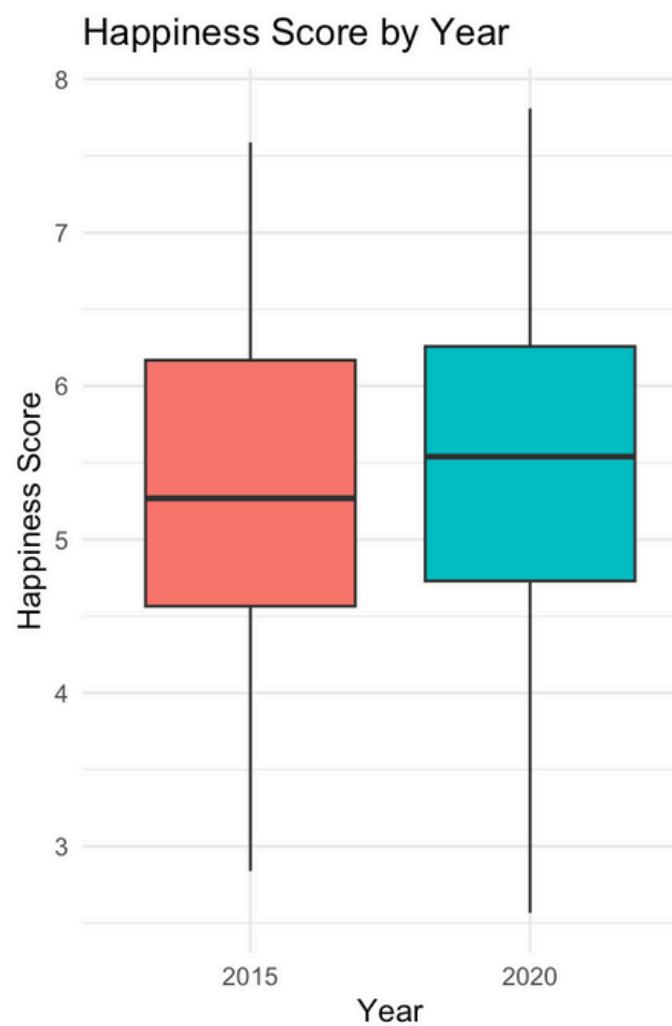
Q2: Have the happiness scores of people in various countries changed between 2015 and 2020 (before and after the COVID-19) (**Pair t-test**)

Q3: How do the different explanatory variables influence the happiness score in 2020 (**Multiple Regression**)

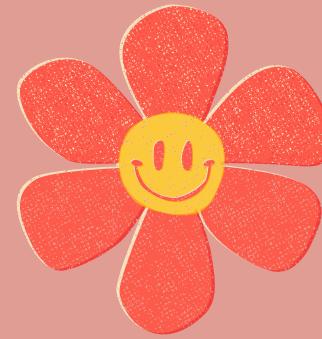
EXPLORATORY DATA ANALYSIS



Summary:



Region	n	Average	Median	SD	IQR	Min	Max	Ratio
Central and Eastern Europe	24	5.66	5.64	0.61	0.98	4.56	6.91	1.52
Commonwealth of Independent States	51	5.71	5.56	0.45	0.52	5.12	6.26	1.22
East Asia	6	5.71	5.69	0.46	0.40	5.12	6.46	1.26
Latin America and Caribbean	21	5.98	6.14	0.66	0.55	3.72	7.12	1.91
Middle East and North Africa	17	5.23	5.01	0.99	1.47	3.53	7.13	2.02
North America and ANZ	41	7.17	7.23	0.16	0.10	6.94	7.30	1.05
South Asia	71	4.48	4.83	1.08	1.22	2.57	5.69	2.22
Southeast Asia	91	5.38	5.35	0.66	1.11	4.31	6.38	1.48
Sub-Saharan Africa	391	4.38	4.43	0.68	0.97	2.82	6.10	2.17
Western Europe	21	6.90	7.09	0.68	1.05	5.51	7.81	1.42

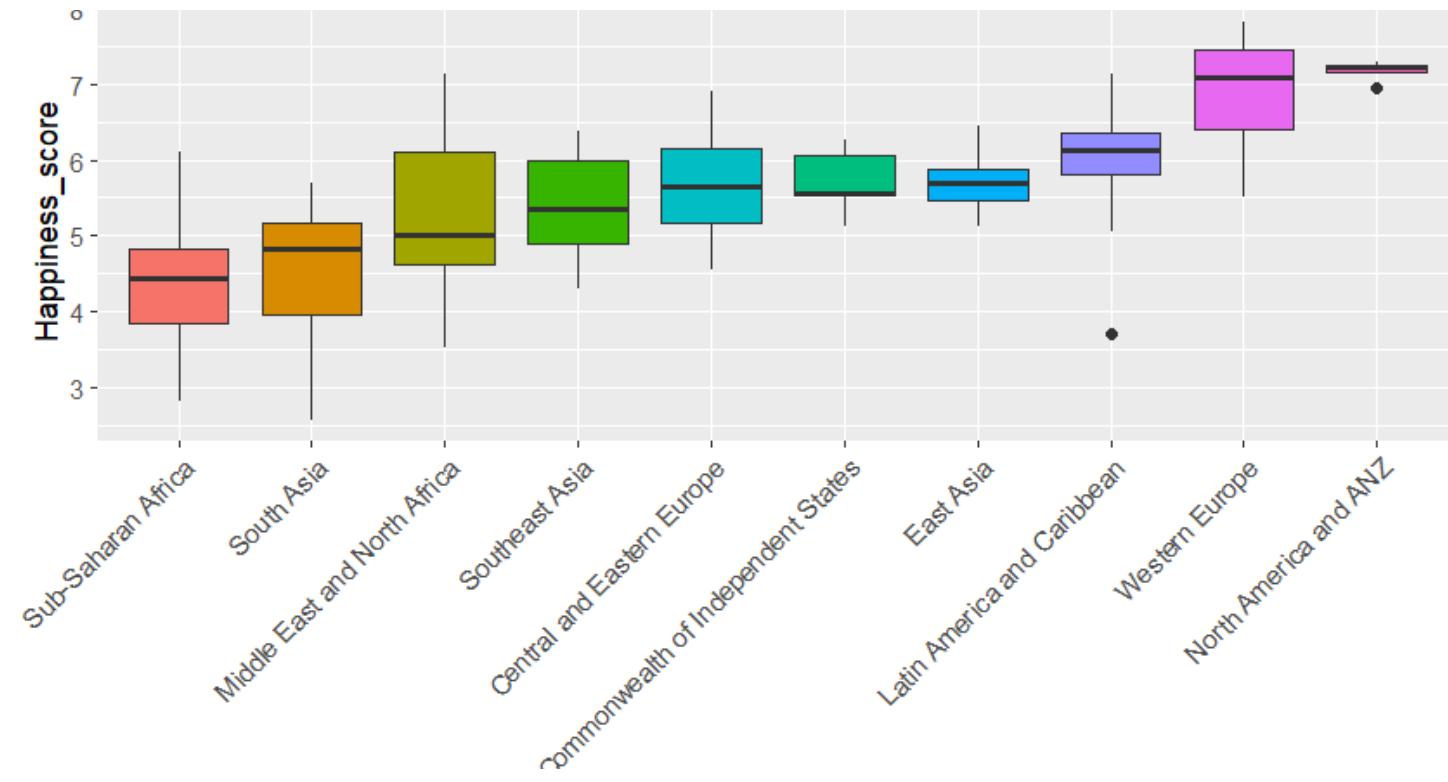


QUESTION 1

Do the happiness scores of people
on different continents
in 2020 differ significantly (ANOVA Test)

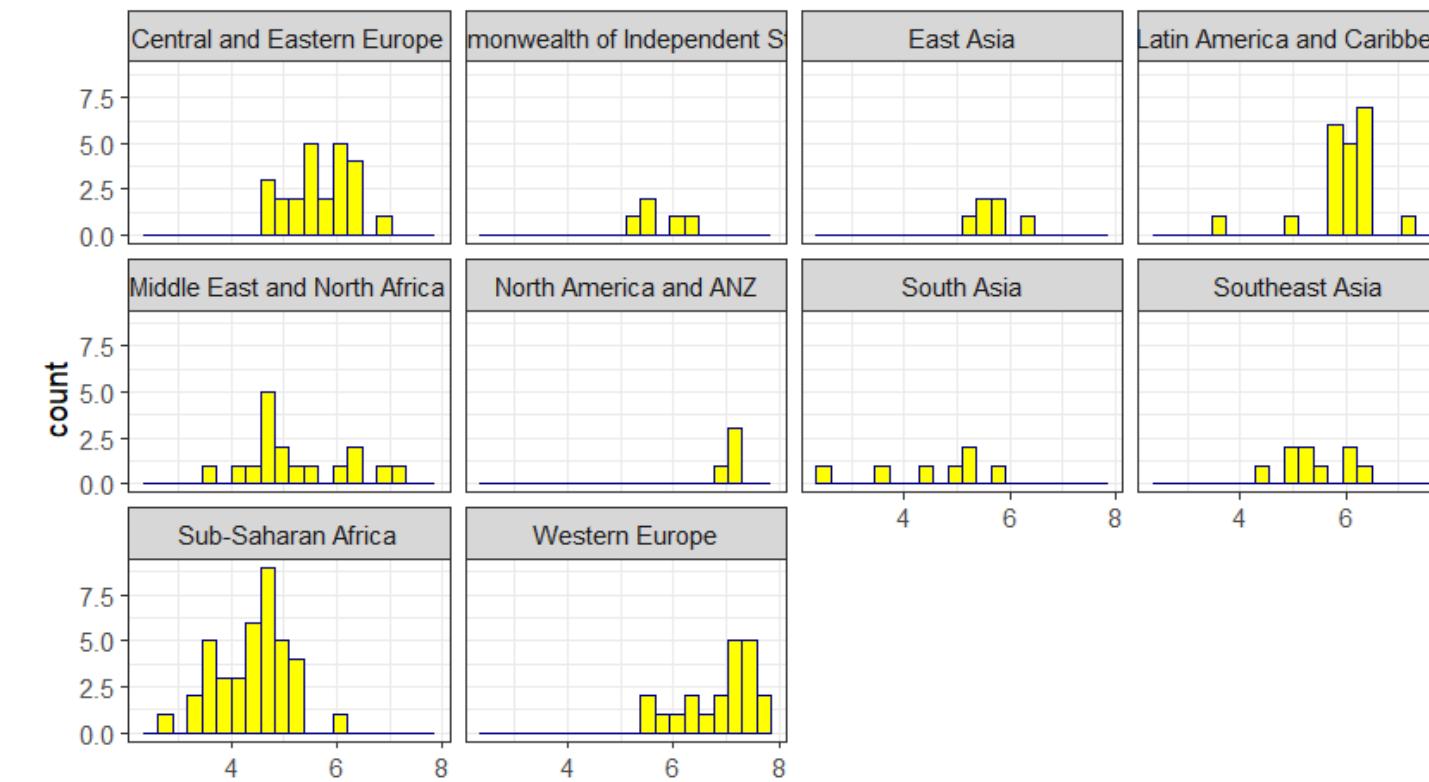


BOXPLOT



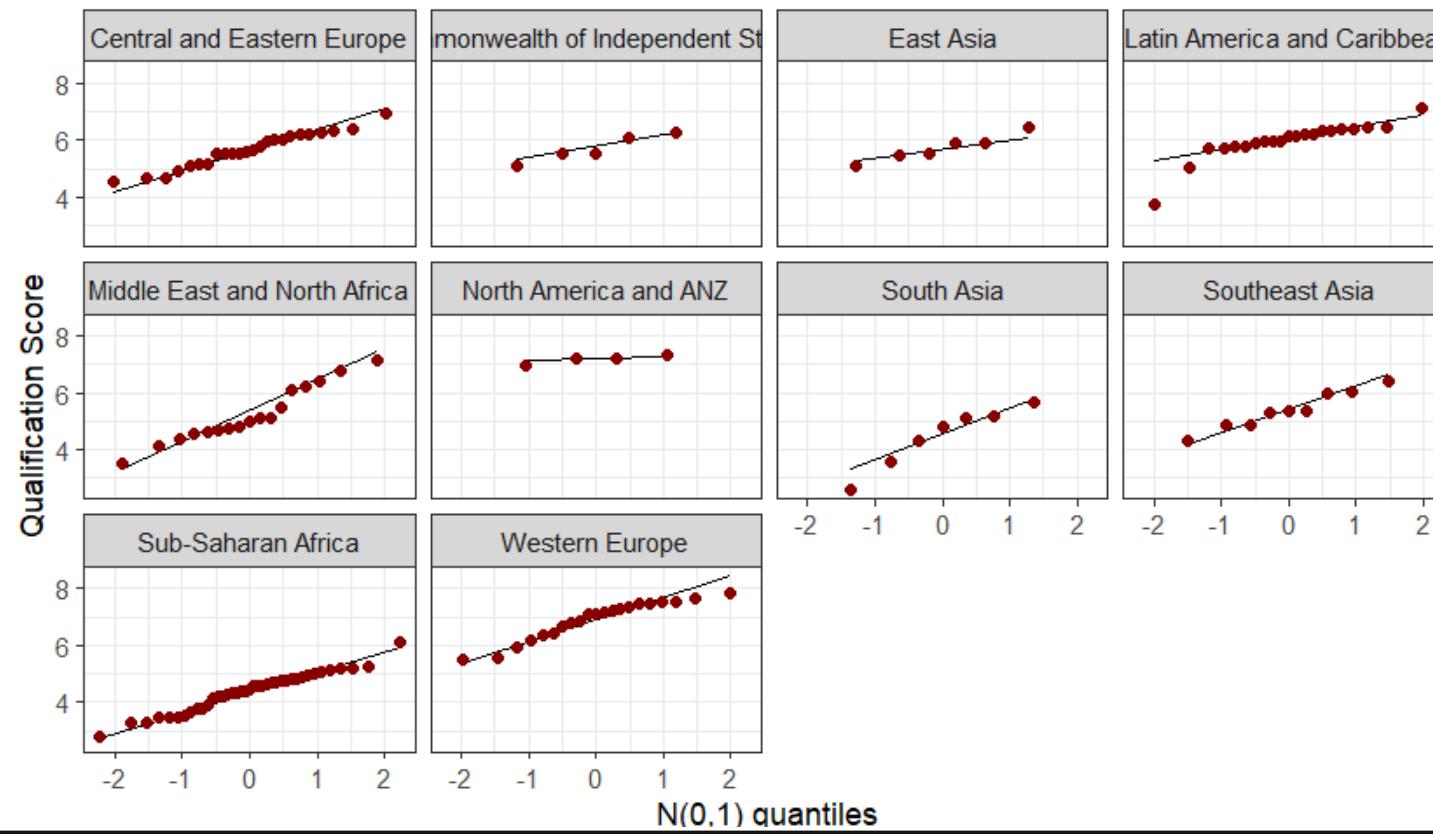
1

HISTOGRAM



2

QQ-PLOT



3

SHAPIRO AND LEVENE TEST

Shapiro-Wilk normality test

```
data: model.fit$residuals  
W = 0.986986, p-value = 0.1635
```

```
Levene's Test for Homogeneity of Variance (center = median)  
Df F value Pr(>F)  
group 9 1.45152 0.17173
```

143

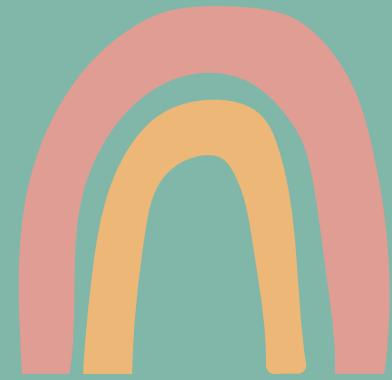
ANOVA

Except Western Europe,
the happiness scores
between North America
and other regions differ !



	pval
Commonwealth of Independent States-Central and Eastern Europe	0.89044
East Asia-Central and Eastern Europe	0.86153
Latin America and Caribbean-Central and Eastern Europe	0.13038
Middle East and North Africa-Central and Eastern Europe	0.05830 .
North America and ANZ-Central and Eastern Europe	0.00013 ***
South Asia-Central and Eastern Europe	0.00017 ***
Southeast Asia-Central and Eastern Europe	0.32540
Sub-Saharan Africa-Central and Eastern Europe	0.00000000156875 ***
Western Europe-Central and Eastern Europe	0.000000034884740 ***
East Asia-Commonwealth of Independent States	0.98432
Latin America and Caribbean-Commonwealth of Independent States	0.43831
Middle East and North Africa-Commonwealth of Independent States	0.18806
North America and ANZ-Commonwealth of Independent States	0.00255 **
South Asia-Commonwealth of Independent States	0.00369 **
Southeast Asia-Commonwealth of Independent States	0.41749
Sub-Saharan Africa-Commonwealth of Independent States	0.00014 ***
Western Europe-Commonwealth of Independent States	0.00098 ***
Latin America and Caribbean-East Asia	0.41944
Middle East and North Africa-East Asia	0.15144
North America and ANZ-East Asia	0.00185 **
South Asia-East Asia	0.00213 **
Southeast Asia-East Asia	0.37863
Sub-Saharan Africa-East Asia	0.000036415915935 ***
Western Europe-East Asia	0.00045 ***
Middle East and North Africa-Latin America and Caribbean	0.00145 **
North America and ANZ-Latin America and Caribbean	0.00258 **
South Asia-Latin America and Caribbean	0.000003232231723 ***
Southeast Asia-Latin America and Caribbean	0.03667 *
Sub-Saharan Africa-Latin America and Caribbean	0.00000000000074 ***
Western Europe-Latin America and Caribbean	0.000051647046219 ***
North America and ANZ-Middle East and North Africa	0.000002368250056 ***
South Asia-Middle East and North Africa	0.02012 *
Southeast Asia-Middle East and North Africa	0.59548
Sub-Saharan Africa-Middle East and North Africa	0.000075601986413 ***
Western Europe-Middle East and North Africa	0.000000000032034 ***
South Asia-North America and ANZ	0.000000012376285 ***
Southeast Asia-North America and ANZ	0.000049921601323 ***
Sub-Saharan Africa-North America and ANZ	0.00000000007587 ***
Western Europe-North America and ANZ	0.48130
Southeast Asia-South Asia	0.01250 *
Sub-Saharan Africa-South Asia	0.75357
Western Europe-South Asia	0.00000000001190 ***
Sub-Saharan Africa-Southeast Asia	0.00022 ***
Western Europe-Southeast Asia	0.000000352904692 ***
Western Europe-Sub-Saharan Africa	< 0.0000000000000002 ***

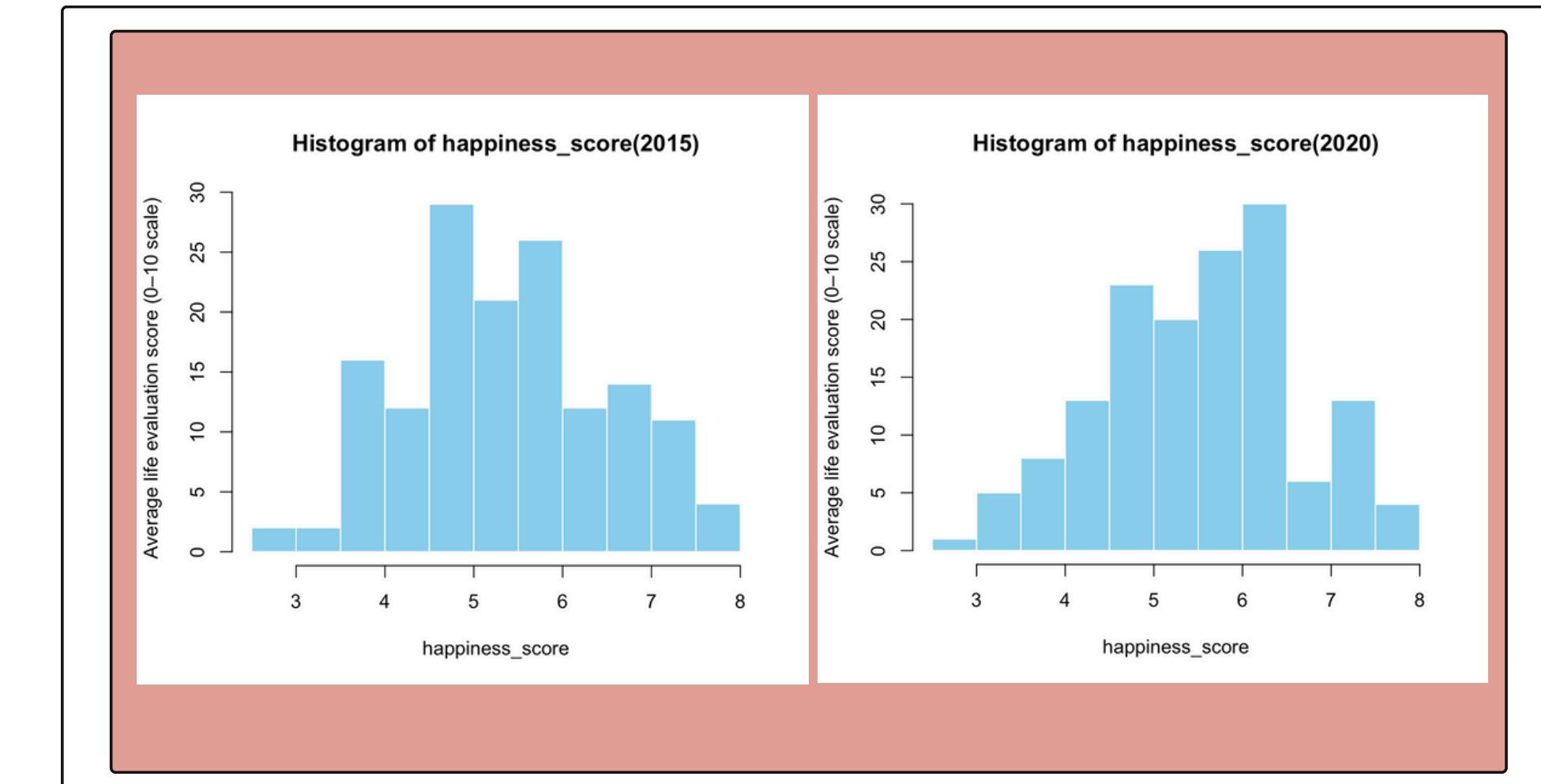
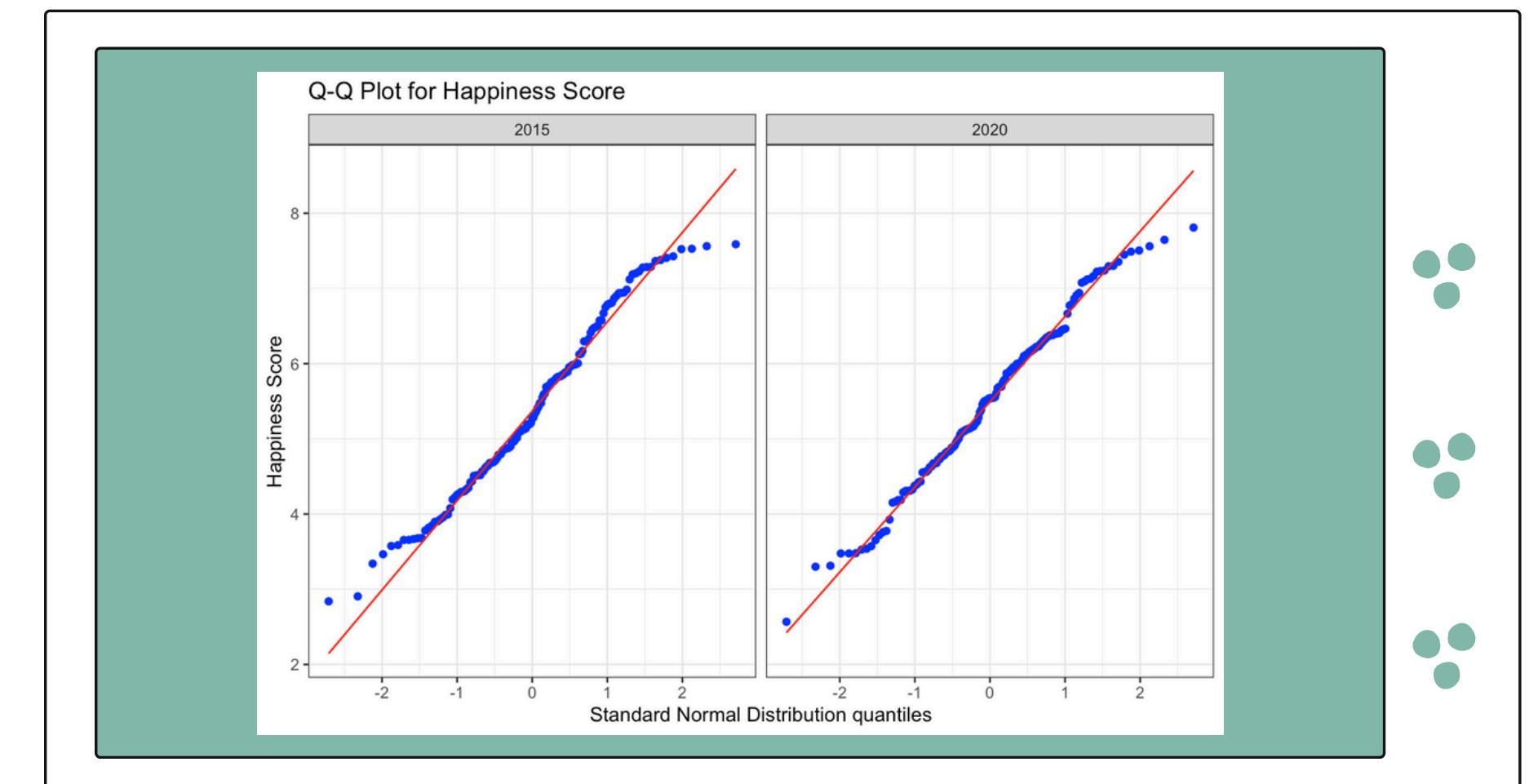
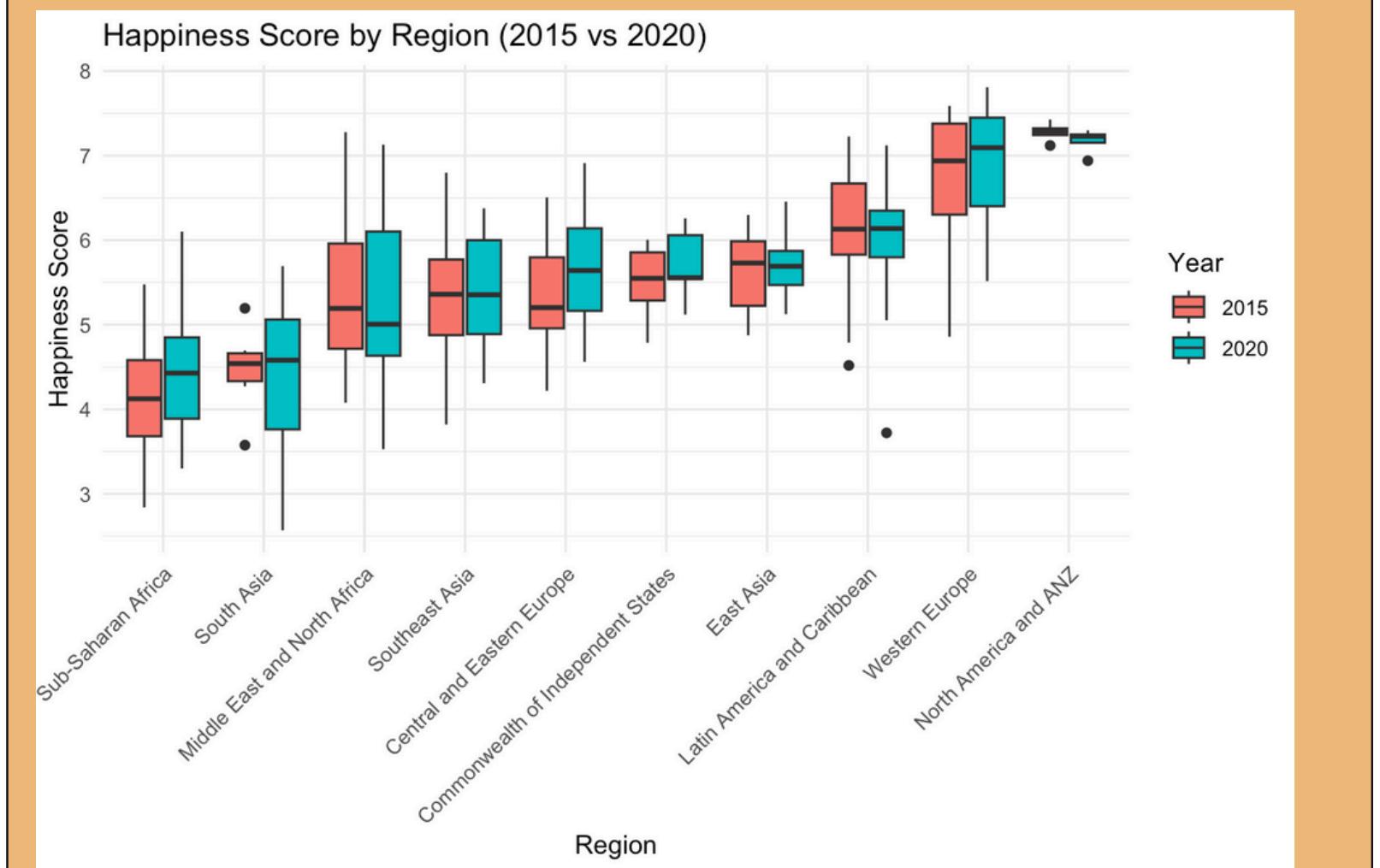
QUESTION 2



HAVE THE HAPPINESS SCORES OF PEOPLE
IN VARIOUS COUNTRIES CHANGED
BETWEEN 2015 AND 2020
(BEFORE AND AFTER THE COVID-19)
(PAIR T-TEST)



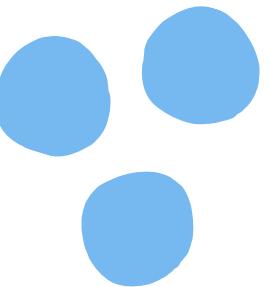
NORMALITY



```
> diff_score <- df_2015$happiness_score - df_2020$happiness_score
> shapiro.test(diff_score)
```

Shapiro-Wilk normality test

```
data: diff_score
W = 0.9852, p-value = 0.112
```



PAIR T-TEST

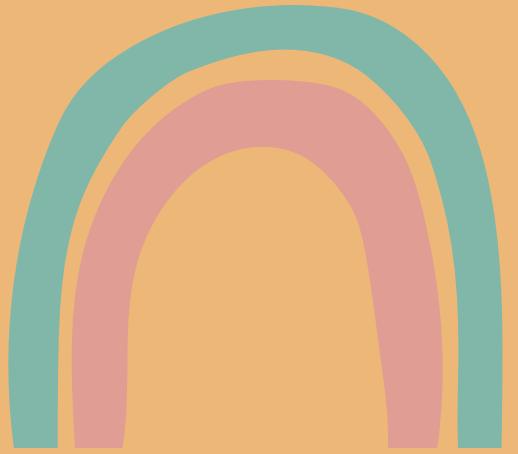
```
> t.test(df_2015$happiness_score, df_2020$happiness_score, paired = TRUE)

Paired t-test

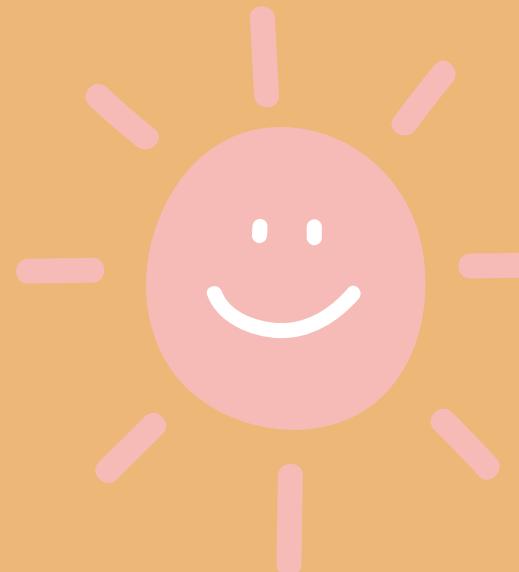
data: df_2015$happiness_score and df_2020$happiness_score
t = -2.34, df = 148, p-value = 0.0206
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
-0.2068116 -0.0174139
sample estimates:
mean difference
-0.112113
```

A paired t-test comparing happiness scores between 2015 and 2020 said a statistically significant decrease ($t = -2.34$, $p = 0.0206$). The average happiness score dropped by 0.11 points

QUESTION 3



HOW DO THE DIFFERENT EXPLANATORY VARIABLES
INFLUENCE THE HAPPINESS SCORE IN 2020
(MULTIPLE REGRESSION)



Happiness

MULTI-REGRESSION



> summary(model51)

Call:
`lm(formula = happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy + freedom_to_make_life_choices + perceptions_of_corruption, data = df_multi)`

Residuals:

Min	1Q	Median	3Q	Max
-1.8508	-0.3453	0.0627	0.3804	1.4712

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.974	0.216	9.14	0.000000000000047 ***
gdp_per_capita	0.690	0.262	2.63	0.00943 **
social_support	1.161	0.280	4.14	0.00005799744202170 ***
healthy_life_expectancy	0.964	0.361	2.67	0.00839 **
freedom_to_make_life_choices	1.604	0.404	3.97	0.00011 ***
perceptions_of_corruption	1.127	0.472	2.39	0.01816 *

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

Residual standard error: 0.57 on 147 degrees of freedom

Multiple R-squared: 0.746, Adjusted R-squared: 0.737

F-statistic: 86.2 on 5 and 147 DF, p-value: <0.000000000000002

> anova(model60, model51)
 Analysis of Variance Table

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model 1: happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy + freedom_to_make_life_choices + generosity + perceptions_of_corruption	1	47.3251				
Model 2: happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy + freedom_to_make_life_choices + perceptions_of_corruption	2	47.8061	-1	-0.481008	1.48393	0.22513

> summary(model60)

Call:

`lm(formula = happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy + freedom_to_make_life_choices + generosity + perceptions_of_corruption, data = df_multi)`

Residuals:

Min	1Q	Median	3Q	Max
-1.7565	-0.3179	0.0665	0.3723	1.4837

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.887	0.227	8.31	0.00000000000061 ***
gdp_per_capita	0.739	0.265	2.79	0.00596 **
social_support	1.153	0.280	4.12	0.000063490577423 ***
healthy_life_expectancy	0.981	0.360	2.72	0.00729 **
freedom_to_make_life_choices	1.482	0.415	3.57	0.00048 ***
generosity	0.621	0.510	1.22	0.22513
perceptions_of_corruption	0.973	0.488	2.00	0.04786 *

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

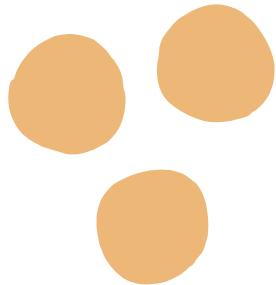
Residual standard error: 0.569 on 146 degrees of freedom

Multiple R-squared: 0.748, Adjusted R-squared: 0.738

F-statistic: 72.4 on 6 and 146 DF, p-value: <0.000000000000002

Happiness

MULTI-REGRESSION



SLR

```
> full_model <- lm(happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy +
+                   freedom_to_make_life_choices + generosity + perceptions_of_corruption,
+                   data = df_multi)
> step_model <- step(full_model, trace = TRUE)
```

Start: AIC=-165.53

```
happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy +
  freedom_to_make_life_choices + generosity + perceptions_of_corruption
```

	Df	Sum of Sq	RSS	AIC
- generosity	1	0.481	47.81	-166.0
<none>			47.33	-165.5
- perceptions_of_corruption	1	1.291	48.62	-163.4
- healthy_life_expectancy	1	2.401	49.73	-160.0
- gdp_per_capita	1	2.525	49.85	-159.6
- freedom_to_make_life_choices	1	4.134	51.46	-154.7
- social_support	1	5.499	52.82	-150.7

Step: AIC=-165.98

```
happiness_score ~ gdp_per_capita + social_support + healthy_life_expectancy +
  freedom_to_make_life_choices + perceptions_of_corruption
```

	Df	Sum of Sq	RSS	AIC
<none>			47.81	-166.0
- perceptions_of_corruption	1	1.856	49.66	-162.2
- gdp_per_capita	1	2.250	50.06	-160.9
- healthy_life_expectancy	1	2.322	50.13	-160.7
- freedom_to_make_life_choices	1	5.132	52.94	-152.4
- social_support	1	5.578	53.38	-151.1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.461	0.145	23.9	<0.0000000000000002 ***
gdp_per_capita	2.316	0.153	15.1	<0.0000000000000002 ***

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

Residual standard error: 0.705 on 151 degrees of freedom
Multiple R-squared: 0.601, Adjusted R-squared: 0.599
F-statistic: 228 on 1 and 151 DF, p-value: <0.0000000000000002

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.137	0.168	18.7	<0.0000000000000002 ***
healthy_life_expectancy	3.372	0.227	14.8	<0.0000000000000002 ***

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

Residual standard error: 0.712 on 151 degrees of freedom
Multiple R-squared: 0.593, Adjusted R-squared: 0.591
F-statistic: 220 on 1 and 151 DF, p-value: <0.0000000000000002

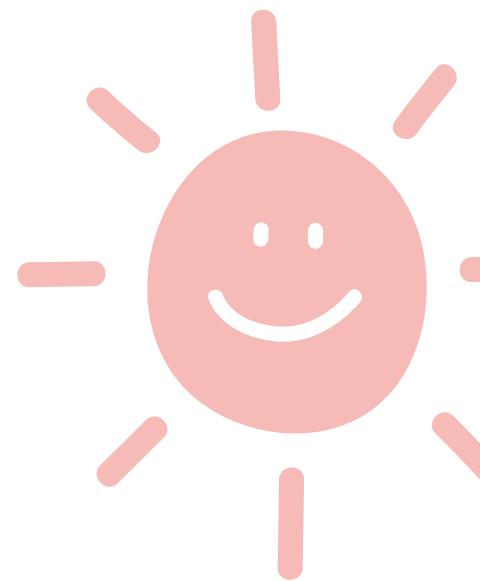
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.328	0.193	27.66	<0.0000000000000002 ***
generosity	0.765	0.899	0.85	0.4

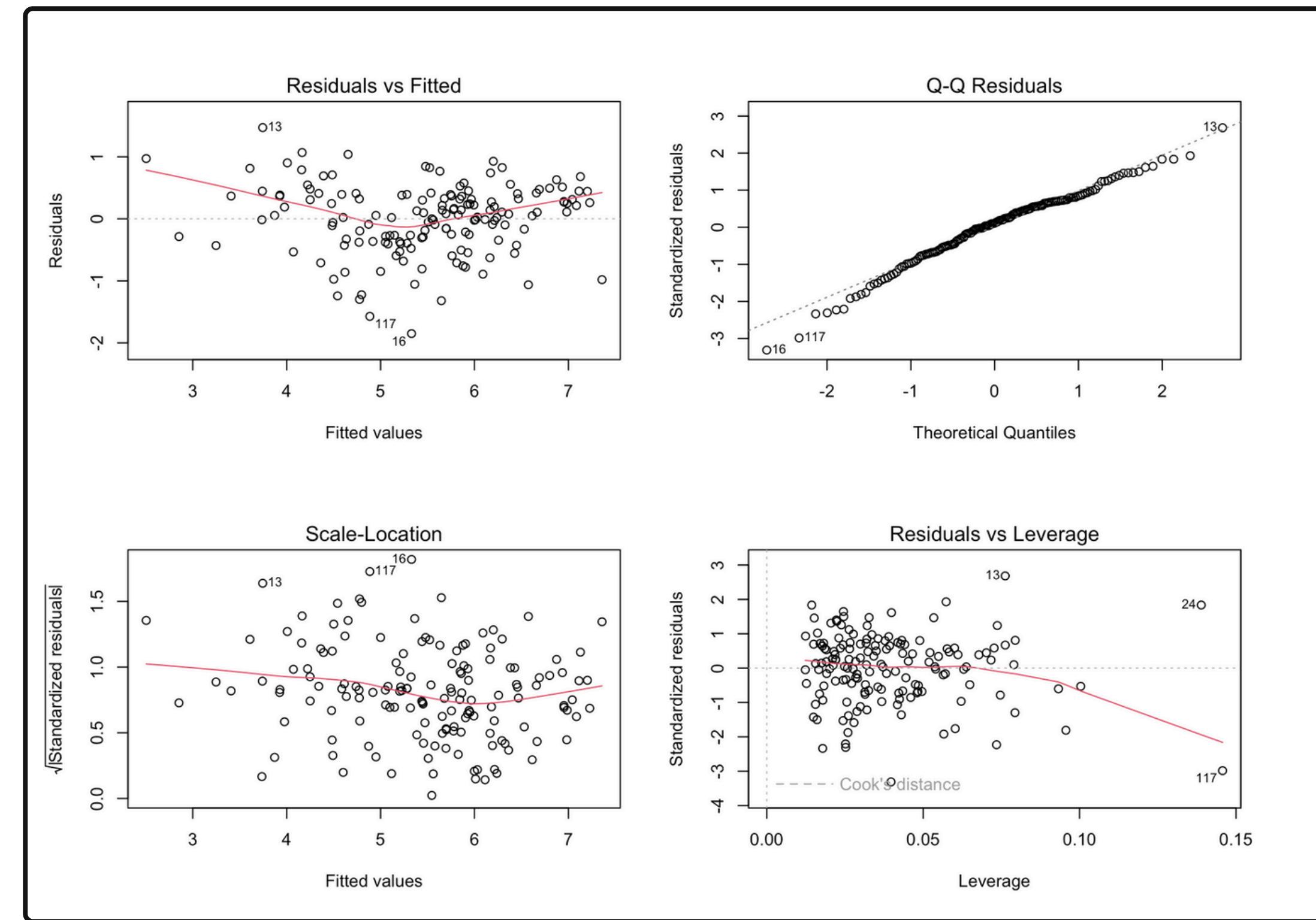
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

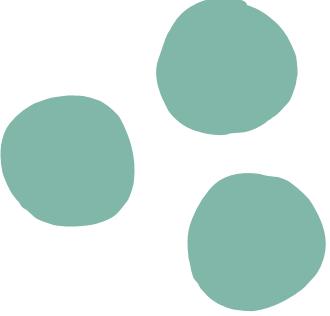
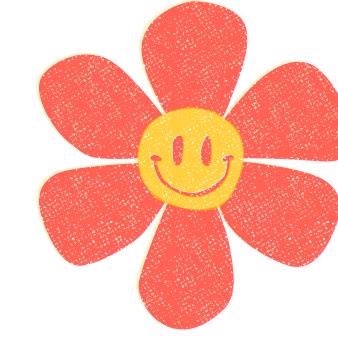
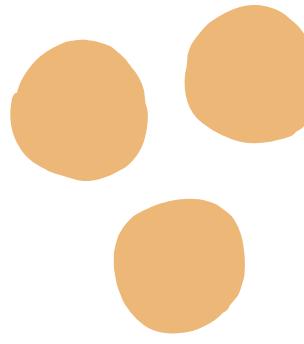
Residual standard error: 1.11 on 151 degrees of freedom
Multiple R-squared: 0.00477, Adjusted R-squared: -0.00182
F-statistic: 0.723 on 1 and 151 DF, p-value: 0.396

Residual standard error: 1.01 on 151 degrees of freedom
Multiple R-squared: 0.175, Adjusted R-squared: 0.17
F-statistic: 32 on 1 and 151 DF, p-value: 0.000000743



CHECKING MODEL ASSUMPTIONS



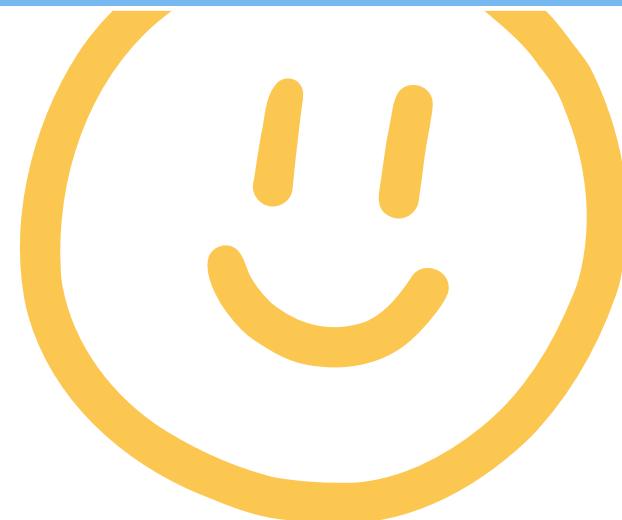
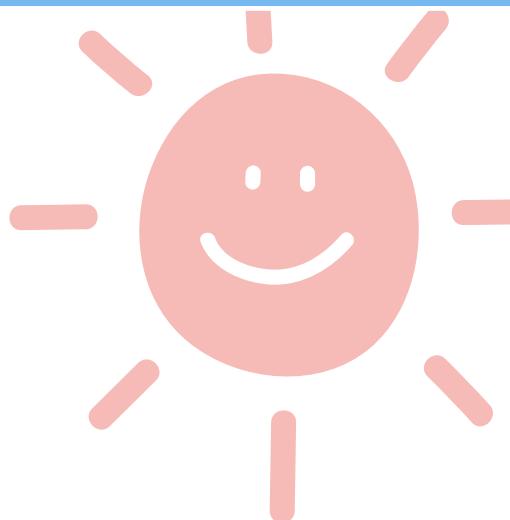


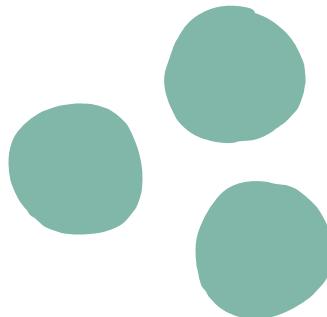
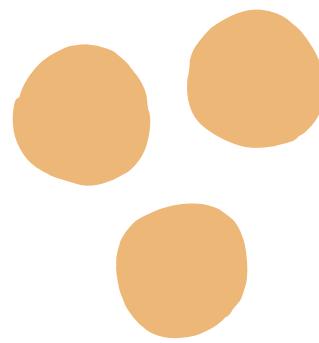
Scope of Inference

THIS DATASET CAN CROSS-NATIONAL COMPARISONS OF SUBJECTIVE HAPPINESS AND RELATED FACTORS.

HOWEVER,

- A. IT REFLECTS CORRELATIONAL RELATIONSHIPS NOT CAUSAL
- B. IT IS NOT GENERALIZABLE BEYOND THE COUNTRIES INCLUDED IN THE DATASET.





Happy is everything

THANKS

STAT-614

Anna and Qintao

