



# **World Happiness Report Data Analysis Final Project**

Monica Collins, Jack Welch, and Zev Rosen



# Why is it Important to Measure Happiness?

- Happiness is a subjective measure of well-being
- Happiness is dependent on a variety of objective and subjective factors making it a good indicator for how a society is functioning
- High happiness scores generally indicate a well functioning society



# What Potential Factors Contribute to Happiness?

- According to Maslow's Hierarchy of Needs, psychological processes that lead to happiness only can occur when more basic human needs are met such as food, bodily, and economic security
- Therefore, potential measurable factors that could contribute to happiness likely relate to the economic security, health, and personal liberty/freedom
- Additional subjective measures that could contribute to happiness include factors increasing interpersonal connectedness such as generosity, social support, and general affect of the population



# Where is this data from/how was it collected

We examined the 2019 World Happiness Report which is an annual report that measures and ranks the happiness of 156 countries. The data used for this report is a composite of data from a Gallup World Poll and The World Health Organization's Global Health Observatory data repository,



# Variables

Dependent Variable: Mean Happiness Score -- continuous numerical measurement of happiness

Positive Affect, Negative Affect, Social support, Freedom, Negative Affect, and Generosity are country ranking based on answers from specific questions in the World Gallup Poll. The value 1 represents the country with the highest average score for the variable and 156 indicates the country with the lowest average score for the value

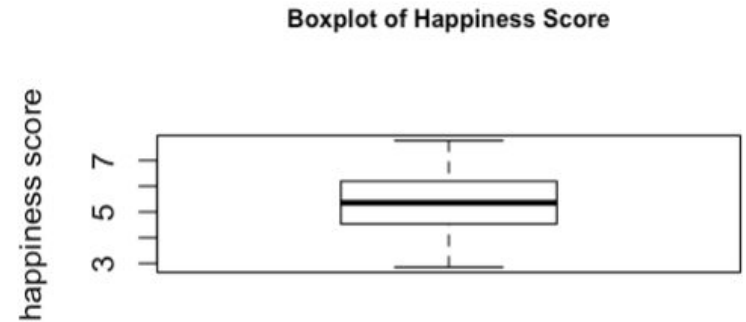
Log of GDP is a ranked list of the log of the GDP per capita where a value of 1 indicates the country with the largest log GDP per capita and a value of 156 indicates the country with the smallest log GDP per capita

Average Healthy Life Expectancy is a ranked list of healthy life expectancy across the countries surveyed with data obtained from the World Health Organization where a value of 1 indicates the country with the highest healthy life expectancy and a value of 156 indicates the country with the lowest healthy life expectancy

# Dependent Variable



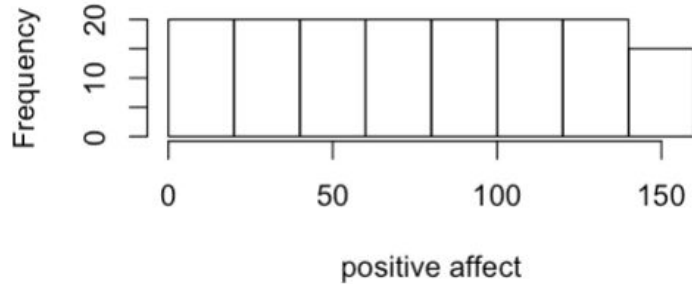
Mean: 5.401  
Median: 5.356  
Standard deviation: 1.12  
Interquartile range: 1.65



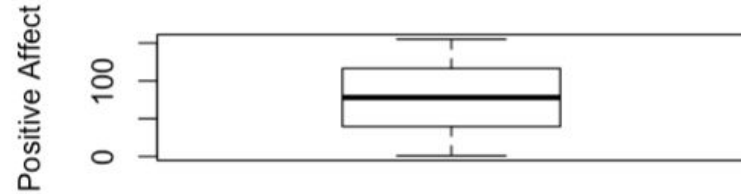
# Explanatory Variables

Mean: 78.0  
Median: 78.0  
Standard deviation: 44.89  
Interquartile range: 77

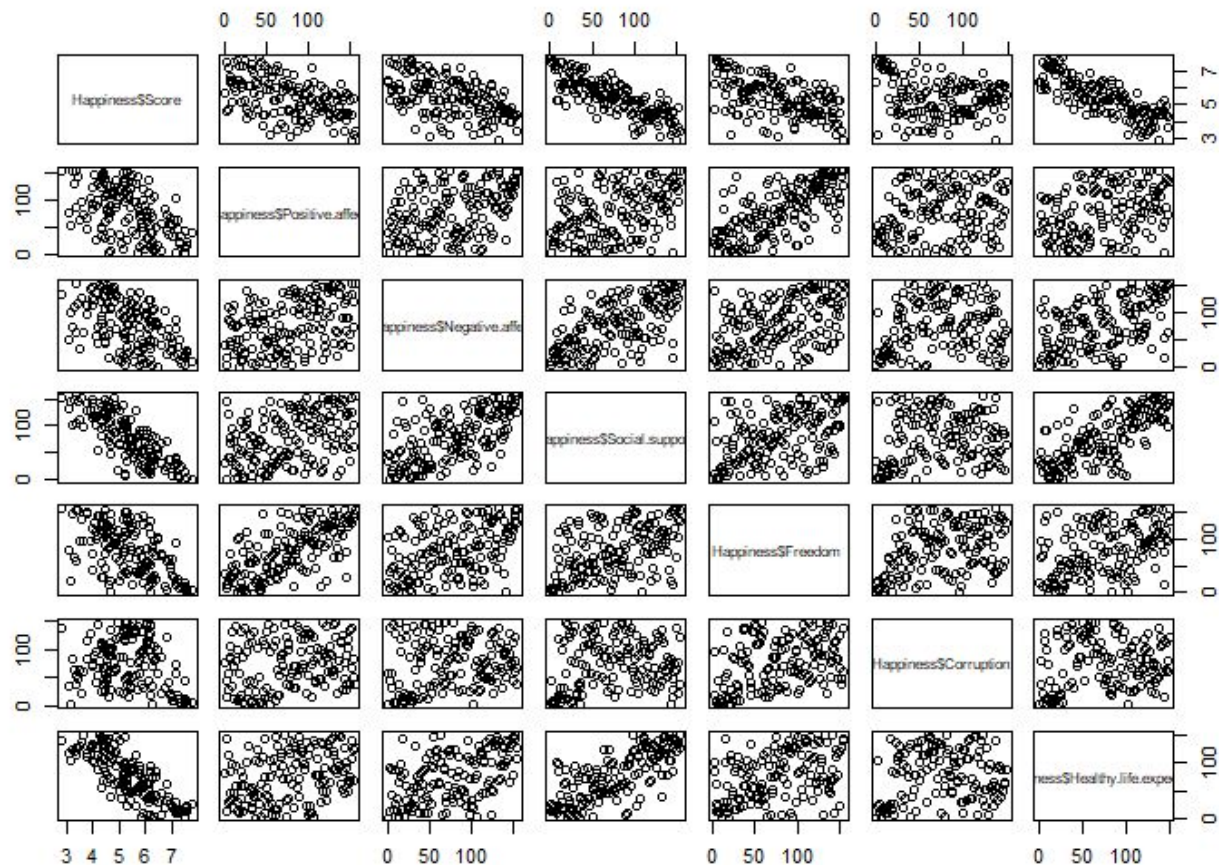
Histogram of Positive Affect



Boxplot of Positive Affect



# Linearity





# Model Analysis

```
> summaryHH(fitset)
```

	model	p	rsq	rss	adjr2	cp	bic	stderr
1	H\$S	2	0.671	58.5	0.669	88.26	-144	0.656
2	H\$S-H\$H	3	0.760	42.7	0.756	30.27	-182	0.562
3	H\$P-H\$S-H\$H	4	0.786	38.0	0.781	14.64	-193	0.533
4	H\$P-H\$S-H\$L-H\$H	5	0.796	36.2	0.790	9.69	-195	0.522
5	H\$P-H\$S-H\$C-H\$L-H\$H	6	0.802	35.1	0.795	7.58	-194	0.516
6	H\$P-H\$N-H\$S-H\$C-H\$L-H\$H	7	0.805	34.6	0.796	7.45	-191	0.514
7	H\$P-H\$N-H\$S-H\$F-H\$C-H\$L-H\$H	8	0.808	34.1	0.798	7.48	-189	0.512
8	H\$P-H\$N-H\$S-H\$F-H\$C-H\$G-H\$L-H\$H	9	0.809	33.9	0.797	9.00	-184	0.513

```
Step: AIC=-177.09
Happiness$Score ~ Happiness$Positive.affect + Happiness$Negative.affect +
  Happiness$Social.support + Happiness$Freedom + Happiness$Corruption +
  Happiness$Log.of.GDP.per.capita + Happiness$Healthy.life.expectancy
```

	Df	Sum of Sq	RSS	AIC
<none>			34.056	-177.09
- Happiness\$Freedom	1	0.5193	34.576	-177.01
- Happiness\$Corruption	1	0.5644	34.621	-176.83
- Happiness\$Negative.affect	1	0.6641	34.720	-176.43
+ Happiness\$Generosity	1	0.1250	33.931	-175.60
- Happiness\$Positive.affect	1	1.5957	35.652	-172.78
- Happiness\$Log.of.GDP.per.capita	1	1.7829	35.839	-172.05
- Happiness\$Healthy.life.expectancy	1	3.2577	37.314	-166.49
- Happiness\$Social.support	1	6.6830	40.739	-154.37

```
> |
```



# Multicollinearity?

vif(fitbest)

Happiness\$Positive.affect  
Happiness\$Freedom

1.983334

Happiness\$Negative.affect

1.818519

Happiness\$Social.support

3.738617

2.427502

Happiness\$Corruption

1.242993

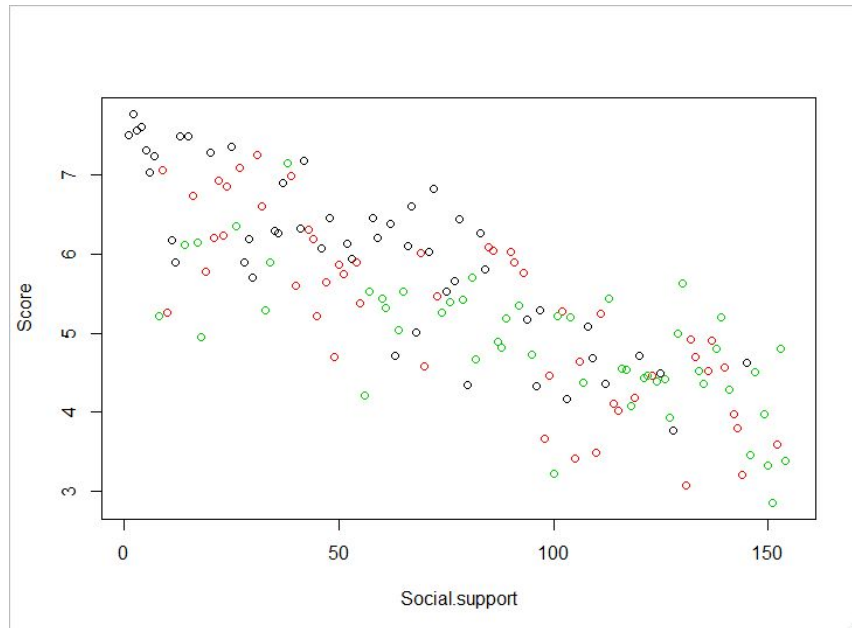
Happiness\$Log.of.GDP.per.capita

5.318331

Happiness\$Healthy.life.expectancy

4.375980

# Interaction Effect



# The Final Model

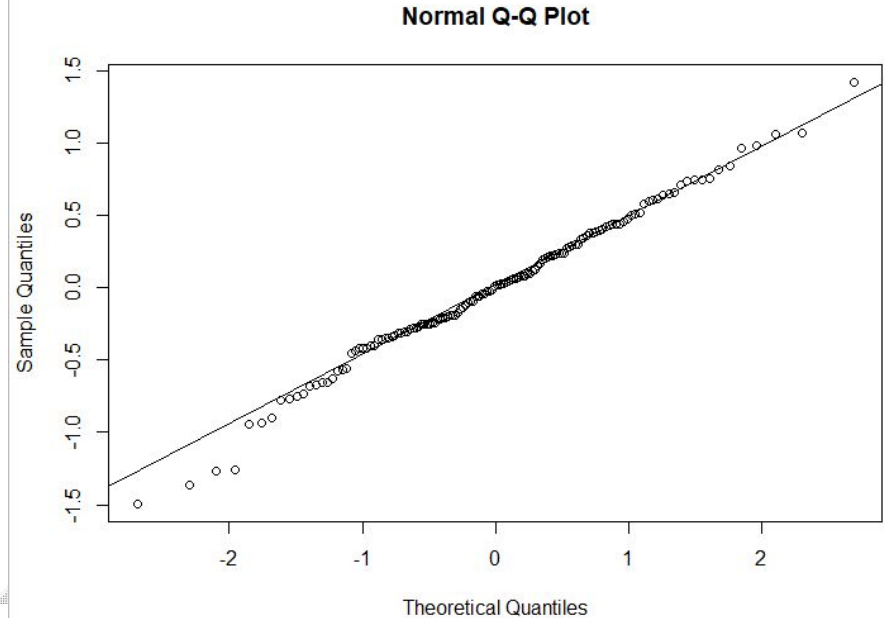
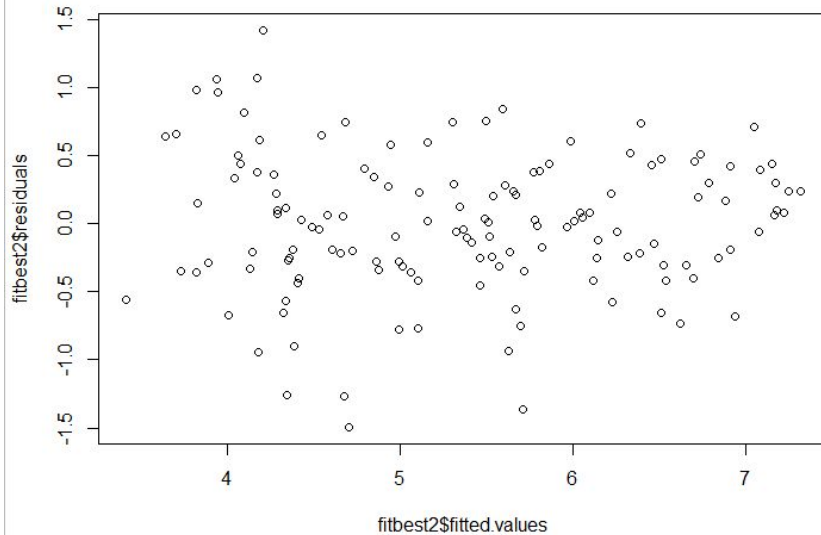
```
Call:
lm(formula = Happiness$Score ~ Happiness$Positive.affect + Happiness$Negative.affect +
    Happiness$Social.support + Happiness$Freedom + Happiness$Corruption +
    Happiness$Healthy.life.expectancy)

Residuals:
    Min       1Q   Median       3Q      Max
-1.4973 -0.3007  0.0155  0.3469  1.4263

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.516787   0.123638  60.797 < 2e-16 ***
Happiness$Positive.affect -0.003393   0.001406  -2.414  0.0171 *
Happiness$Negative.affect  0.001520   0.001324   1.148  0.2530
Happiness$Social.support -0.011093   0.001720  -6.448 1.93e-09 ***
Happiness$Freedom     -0.001890   0.001539  -1.229  0.2214
Happiness$Corruption   -0.002045   0.001136  -1.800  0.0741 .
Happiness$Healthy.life.expectancy -0.011137  0.001546  -7.204 3.91e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5211 on 133 degrees of freedom
(16 observations deleted due to missingness)
Multiple R-squared:  0.8018,    Adjusted R-squared:  0.7929
F-statistic: 89.68 on 6 and 133 DF,  p-value: < 2.2e-16
```

# Regression Diagnostics



## Outliers and Points of High Leverage

```
> summary(influence.measures(fitbest2))
Potentially influential observations of
lm(formula = Happiness$Score ~ Happiness$Positive.affect + Happiness$Negative.affect +
edom + Happiness$Corruption + Happiness$Healthy.life.expectancy) :

      dfb.1_ dfb.H$P. dfb.H$N. dfb.H$S. dfb.Hp$F dfb.Hp$C dfb.H$H. dffit  cov.r  cook.d hat
54    0.03  -0.03   -0.07    0.10    0.08    0.00   -0.12    0.16  1.16_*  0.00  0.10
67   -0.07   0.17   -0.01    0.11    0.01   -0.18    0.01    0.46  0.71_*  0.03  0.03
112  0.09  -0.08   -0.17    0.15   -0.03   -0.05    0.05    0.29  1.18_*  0.01  0.13
125  0.00  -0.10    0.03   -0.05    0.10    0.00    0.01   -0.13  1.18_*  0.00  0.11
130 -0.10   0.22   -0.02   -0.19    0.04   -0.26    0.19   -0.48  0.74_*  0.03  0.03
148 -0.06  -0.13   0.16   -0.08    0.17    0.04   -0.15   -0.40  0.78_*  0.02  0.02
152 -0.24  -0.02   0.02   -0.46    0.32    0.29    0.13   -0.82_*  0.70_*  0.09  0.07
153 -0.15   0.04   0.37   -0.31   -0.04    0.19   -0.03   -0.56  0.79_*  0.04  0.05

> outlierTest(fitbest2)
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferroni p
152 -3.065931      0.0026324      0.36854
> |
```

## Final Model Part II

```
Call:
lm(formula = Score ~ Positive.affect + Negative.affect + Social.support +
    Freedom + Corruption + Healthy.life.expectancy, data = subdat)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.98717	-0.28382	0.00933	0.33957	1.00762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	7.5884024	0.1103173	68.787	< 2e-16	***
Positive.affect	-0.0034330	0.0012934	-2.654	0.00898	**
Negative.affect	0.0008699	0.0012268	0.709	0.47957	
Social.support	-0.0094986	0.0016286	-5.832	4.39e-08	***
Freedom	-0.0030335	0.0014274	-2.125	0.03553	*
Corruption	-0.0021820	0.0010123	-2.156	0.03304	*
Healthy.life.expectancy	-0.0112635	0.0014094	-7.992	7.56e-13	***

---

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

Residual standard error: 0.4564 on 125 degrees of freedom  
(16 observations deleted due to missingness)

Multiple R-squared: 0.8422, Adjusted R-squared: 0.8347

F-statistic: 111.2 on 6 and 125 DF, p-value: < 2.2e-16



# The Happiness Equation:

Happiness(hat) = 7.5884024 - 0.0034330(Positive.affect) + 0.0008699(Negative.affect) -  
0.0094986(Social.support) - 0.0030335(Freedom) - 0.0021820 (Corruption) -  
0.0112635 (Healthy.life.expectancy)



# Interpretation of Regression Coefficients



Controlling for negative affect, social support, freedom, corruption, and healthy life expectancy, a decrease in positive affect rank by 1 is associated with a -0.0034330 change in happiness score, on average.

Controlling for positive affect, social support, freedom, corruption, and healthy life expectancy, a decrease in negative affect rank by 1 is associated with a 0.0008699 change in happiness score, on average.

Controlling for positive affect, negative affect, freedom, corruption, and healthy life expectancy, a decrease in social support rank by 1 is associated with a -0.0094986 change in happiness score, on average.

Controlling for positive affect, negative affect, social support, corruption, and healthy life expectancy, a decrease of 1 in freedom rank is associated with a -0.00300335 change in happiness score, on average.

Controlling for positive affect, negative affect, social support, freedom, and healthy life expectancy, a decrease in corruption rank by 1 is associated with a -0.0021820 change in happiness score, on average.

Controlling for positive affect, negative affect, social support, freedom, corruption, a decrease of 1 in healthy life expectancy rank is associated with a -0.0112635 change in happiness score, on average.

\*A decrease in rank by 1 is your rank number increasing by 1, i.e. Rank 1 to Rank 2



# Conclusion

- Increases in positive affect, social support, freedom, and healthy life expectancy are associated with increase happiness while increased negative affect and corruption are associated with a decrease in happiness
- All data and ranks were obtained from subjective poll questions which is a limitation to the study
- Possible variables for future analysis of happiness could include pollution/quality of environment and population density
- Additional future studies could contrast the happiness equations across models using objective measurements versus subjective poll data