ANALYSIS OF WORLD HAPPINESS REPORT

STA9750
Yu-Lun Tsai
DECEMBER 7, 2023

Introduction

What is World Happiness Report?

The report is a publication of the Sustainable Development Solutions Network, a global initiative of the United Nations. The report primarily uses data from the Gallup World Poll. Each annual report is available to the public to download on the World Happiness Report website.

- GDP Per Capita
- Social Support
- Healthy Life Expectancy
- Freedom of Life Choice
- Generosity
- Perception of Corruption
- Confidence of government

https://www.kaggle.com/datasets/usamabuttar/world-happiness-report-2005-present/data

Introduction

```
> summary(WHR)
 Country.name       Regional.indicator       Year       Life.Ladder    Log.GDP.per.capita
 Length:2199        Length:2199        Min.   :2005   Min.   :1.281   Min.   : 5.527
 Class :character   Class :character   1st Qu.:2010   1st Qu.:4.647   1st Qu.: 8.500
 Mode  :character   Mode  :character   Median :2014   Median :5.432   Median : 9.499
                                       Mean   :2014   Mean   :5.479   Mean   : 9.390
                                       3rd Qu.:2018   3rd Qu.:6.309   3rd Qu.:10.373
                                       Max.   :2022   Max.   :8.019   Max.   :11.664
                                                                      NA's   :20

 Social.support   Healthy.life.expectancy.at.birth Freedom.to.make.life.choices   Generosity
 Min.   :0.2282   Min.   : 6.72                    Min.   :0.2575               Min.   :-0.33753
 1st Qu.:0.7466   1st Qu.:59.12                    1st Qu.:0.6565               1st Qu.:-0.11212
 Median :0.8355   Median :65.05                    Median :0.7698               Median :-0.02267
 Mean   :0.8107   Mean   :63.29                    Mean   :0.7479               Mean   : 0.00010
 3rd Qu.:0.9048   3rd Qu.:68.50                    3rd Qu.:0.8594               3rd Qu.: 0.09207
 Max.   :0.9873   Max.   :74.47                    Max.   :0.9852               Max.   : 0.70271
 NA's   :13       NA's   :54                       NA's   :33                   NA's   :73
 Perceptions.of.corruption Positive.affect  Negative.affect   Confidence.in.national.government
 Min.   :0.0352            Min.   :0.1789   Min.   :0.08274   Min.   :0.0688
 1st Qu.:0.6881            1st Qu.:0.5717   1st Qu.:0.20766   1st Qu.:0.3325
 Median :0.7996            Median :0.6631   Median :0.26067   Median :0.4671
 Mean   :0.7452            Mean   :0.6521   Mean   :0.27150   Mean   :0.4840
 3rd Qu.:0.8688            3rd Qu.:0.7379   3rd Qu.:0.32289   3rd Qu.:0.6188
 Max.   :0.9833            Max.   :0.8836   Max.   :0.70459   Max.   :0.9936
 NA's   :116               NA's   :24       NA's   :16        NA's   :361
```

Introduction – Data Cleaning

```
> summary(WHR)
 Country.Name       Regional.Indicator      Year       Life.Ladder    Log.GDP.Per.Capita
 Length:1982        Length:1982        Min.   :2005   Min.   :2.179   Min.   : 5.527
 Class :character   Class :character   1st Qu.:2011   1st Qu.:4.614   1st Qu.: 8.435
 Mode  :character   Mode  :character   Median :2014   Median :5.404   Median : 9.471
                                       Mean   :2014   Mean   :5.463   Mean   : 9.340
                                       3rd Qu.:2018   3rd Qu.:6.289   3rd Qu.:10.318
                                       Max.   :2022   Max.   :7.971   Max.   :11.664
 Social.Support   Healthy.Life.Expectancy.At.Birth Freedom.To.Make.Life.Choices   Generosity
 Min.   :0.2902   Min.   : 6.72                   Min.   :0.2575               Min.   :-0.337527
 1st Qu.:0.7397   1st Qu.:58.60                   1st Qu.:0.6612               1st Qu.:-0.108381
 Median :0.8348   Median :64.95                   Median :0.7701               Median :-0.022006
 Mean   :0.8087   Mean   :63.17                   Mean   :0.7505               Mean   : 0.001415
 3rd Qu.:0.9061   3rd Qu.:68.72                   3rd Qu.:0.8593               3rd Qu.: 0.091745
 Max.   :0.9873   Max.   :74.47                   Max.   :0.9852               Max.   : 0.702708
 Perceptions.Of.Corruption Positive.Affect  Negative.Affect  Confidence.In.National.Government
 Min.   :0.0352            Min.   :0.1789   Min.   :0.0927   Min.   :0.07879
 1st Qu.:0.6896            1st Qu.:0.5749   1st Qu.:0.2085   1st Qu.:0.33776
 Median :0.8028            Median :0.6654   Median :0.2614   Median :0.47218
 Mean   :0.7470            Mean   :0.6558   Mean   :0.2712   Mean   :0.49460
 3rd Qu.:0.8696            3rd Qu.:0.7413   3rd Qu.:0.3218   3rd Qu.:0.63179
 Max.   :0.9833            Max.   :0.8836   Max.   :0.6067   Max.   :0.99360
```

Introduction – Main Goal

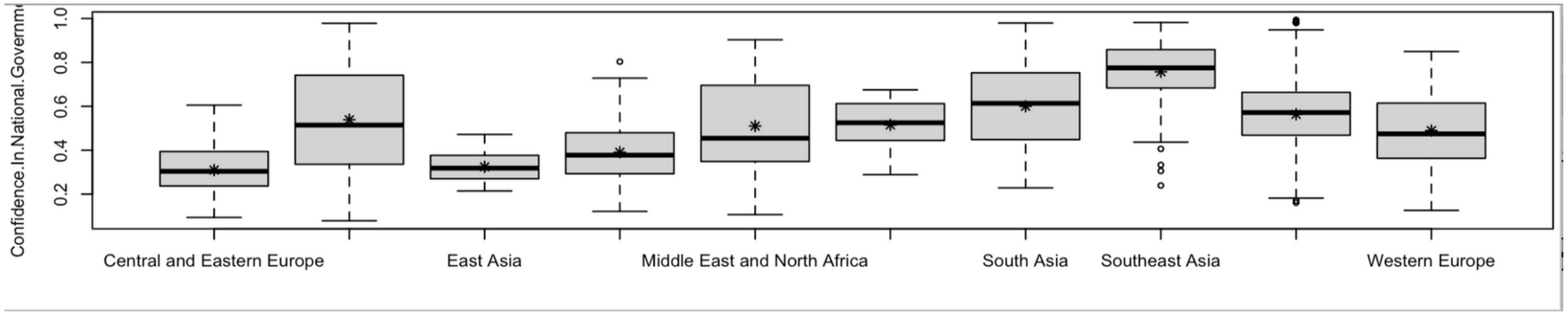
y-variable -> Confidence of national government

x-variable -> Region
Year
Freedom to make choices
Generosity
Healthy Life Expectancy

If there are any relations between Xs and Y?
Which X affects Y the most?

ASSOCIATION ANALYSIS
&
REGRESSION ANALYSIS

VS Region



	Western Europe	Discrepancy	Estimated p-value
Averages (ANOVA)	0.4894	107.2	0
Mean Ranks (Kruskal)	924.8	655.1	0
Medians	0.4744	539.2	0

With 500 permutations, we are 95% confident that
the p-value of ANOVA (means) is between 0 and 0.007
the p-value of Kruskal-Wallis (ranks) is between 0 and 0.007
the p-value of median test is between 0 and 0.007
Note: If 0.05 is in a range, change permutations= to a larger number

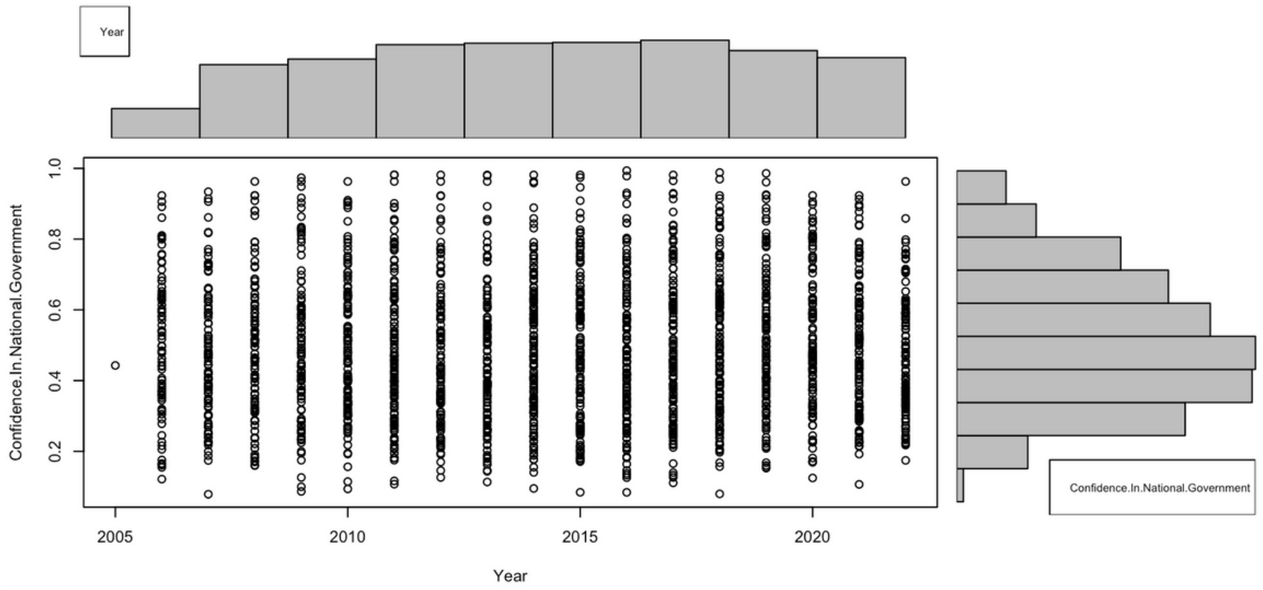
1-pf(107.2, df1=9, df2=1972)
p-value -> 0
Statistically Significant
Conclusive

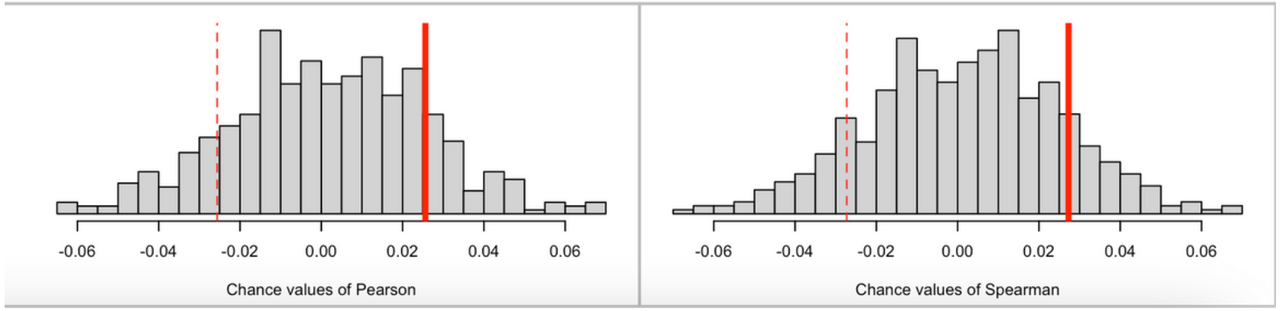
VS Region

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.51802 -0.10356 -0.00521  0.10682  0.43878

Coefficients:
                                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       0.30970    0.01070  28.936  < 2e-16 ***
Regional.IndicatorCommonwealth of Independent States  0.22957    0.01630  14.081  < 2e-16 ***
Regional.IndicatorEast Asia                       0.01460    0.02406   0.607    0.544
Regional.IndicatorLatin America and Caribbean     0.08115    0.01407   5.768 9.28e-09 ***
Regional.IndicatorMiddle East and North Africa    0.20061    0.01756  11.426  < 2e-16 ***
Regional.IndicatorNorth America and ANZ           0.20574    0.02313   8.893  < 2e-16 ***
Regional.IndicatorSouth Asia                      0.29035    0.01995  14.555  < 2e-16 ***
Regional.IndicatorSoutheast Asia                  0.44775    0.01789  25.028  < 2e-16 ***
Regional.IndicatorSub-Saharan Africa              0.25392    0.01311  19.371  < 2e-16 ***
Regional.IndicatorWestern Europe                  0.17969    0.01441  12.468  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1641 on 1972 degrees of freedom
Multiple R-squared:  0.3284,    Adjusted R-squared:  0.3254
F-statistic: 107.2 on 9 and 1972 DF,  p-value: < 2.2e-16
```

East Asia -> P-value 0.544
Positive Correlation
RMSE : 0.1641 Multiple R-squared:32.84%

VS Year

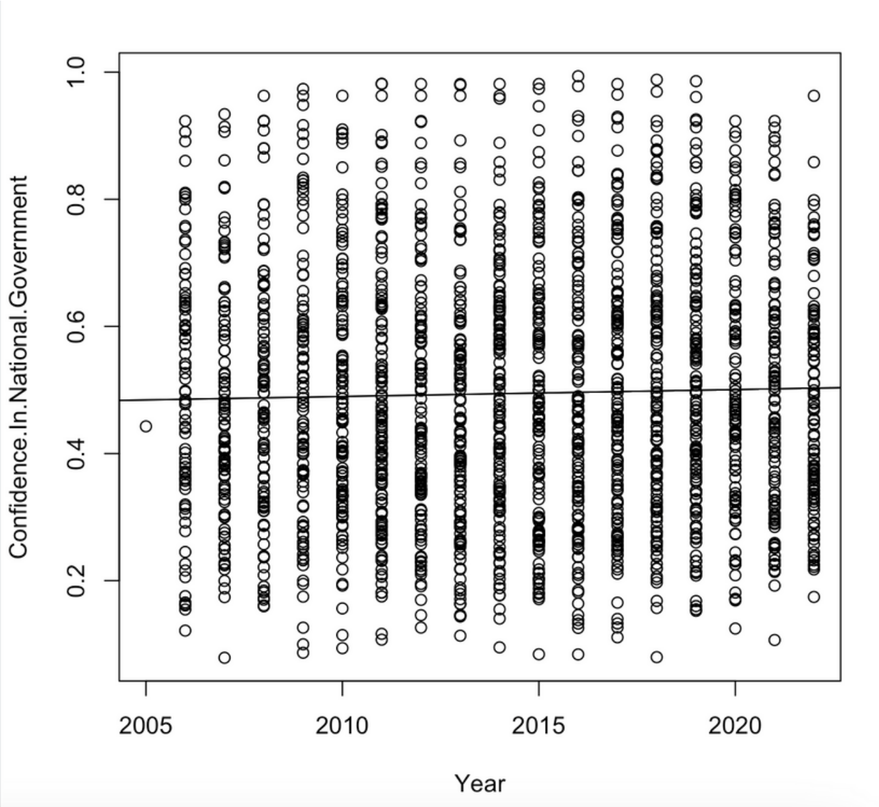





```
Association between Year (numerical) and  Confidence.In.National.Government (numerical)
 using 1982 complete cases
Permutation procedure:
                                    Value Estimated p-value
Pearson's r                  0.02560888             0.276
Spearman's rank correlation 0.02727715             0.252
With 500 permutations, we are 95% confident that:
 the p-value of Pearson's correlation (r) is between 0.237 and 0.317
 the p-value of Spearman's rank correlation is between 0.215 and 0.292
Note:  If 0.05 is in this range, increase the permutations= argument.
```

Non-linear relationship
Correlation -> 0.27
Not Statistically Significant
Conclusive

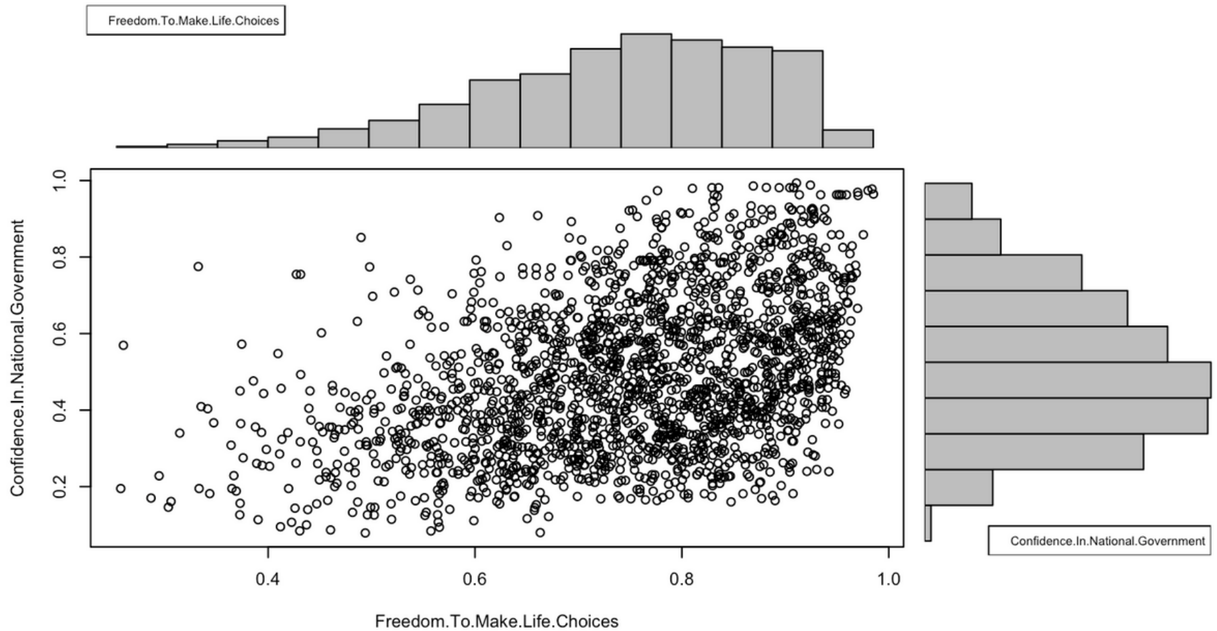
VS Year

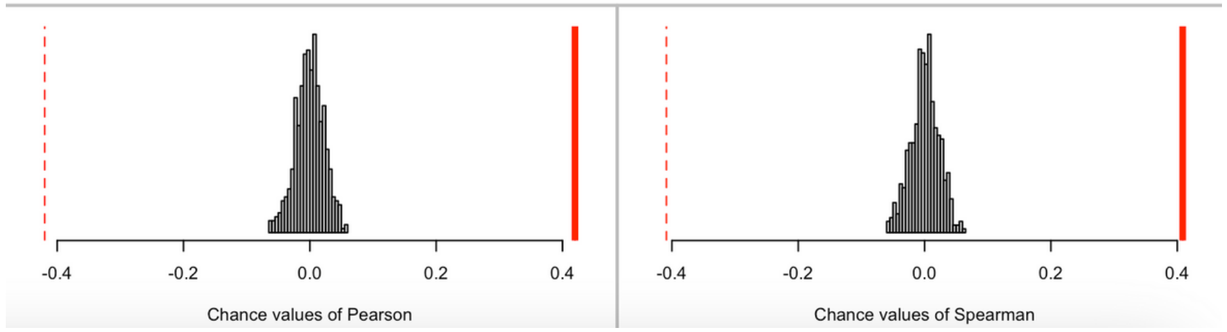



```
Call:
lm(formula = Confidence.In.National.Government ~ Year, data = WHR)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41897 -0.15605 -0.02275  0.13707  0.49712

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.7217526  1.9443537  -0.886    0.376
Year         0.0011003  0.0009653   1.140    0.254

Residual standard error: 0.1997 on 1980 degrees of freedom
Multiple R-squared:  0.0006558, Adjusted R-squared:  0.0001511
F-statistic: 1.299 on 1 and 1980 DF,  p-value: 0.2545
```

Confidence index = -1.7217526 + 0.0011003Year
Each year increase 0.0011003 unit confidence index
RMSE : 0.1997 Multiple R-squared:0.06%

VS Freedom to make choices

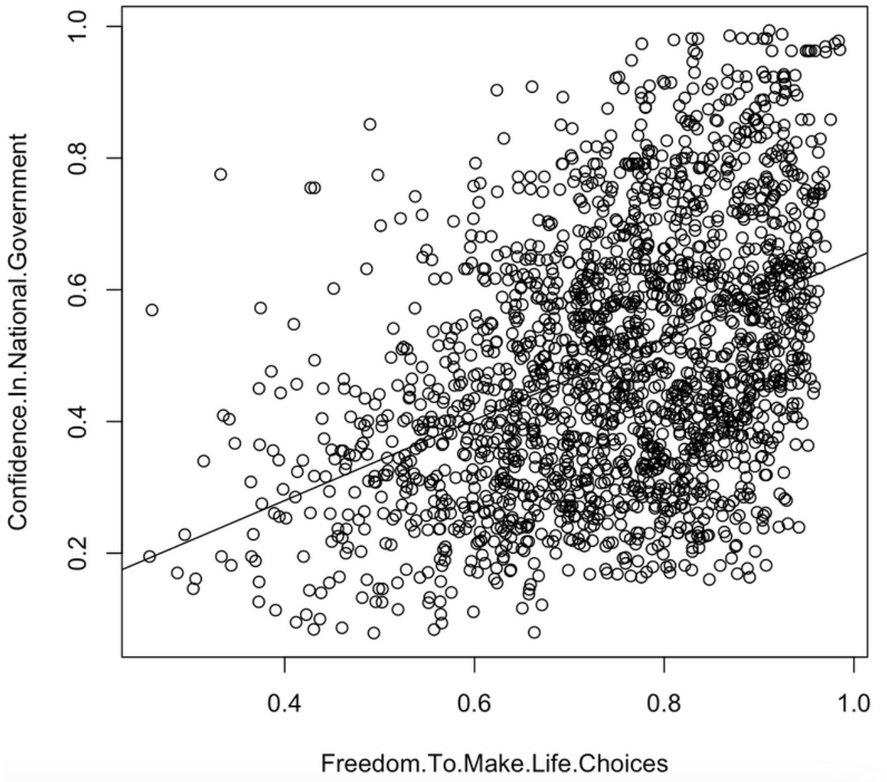





```
> associate(Confidence.In.National.Government~Freedom.To.Make.Life.Choices, data=WHR)
Association between Freedom.To.Make.Life.Choices (numerical) and  Confidence.In.National.G
overnment (numerical)
 using 1982 complete cases
Permutation procedure:
                                  Value Estimated p-value
Pearson's r                 0.4198023                   0
Spearman's rank correlation 0.4086748                   0
With 500 permutations, we are 95% confident that:
 the p-value of Pearson's correlation (r) is between 0 and 0.007
 the p-value of Spearman's rank correlation is between 0 and 0.007
Note:  If 0.05 is in this range, increase the permutations= argument.
```

Non-linear relationship
Monotonic
Correlation -> 0.41
Statistically Significant
Conclusive

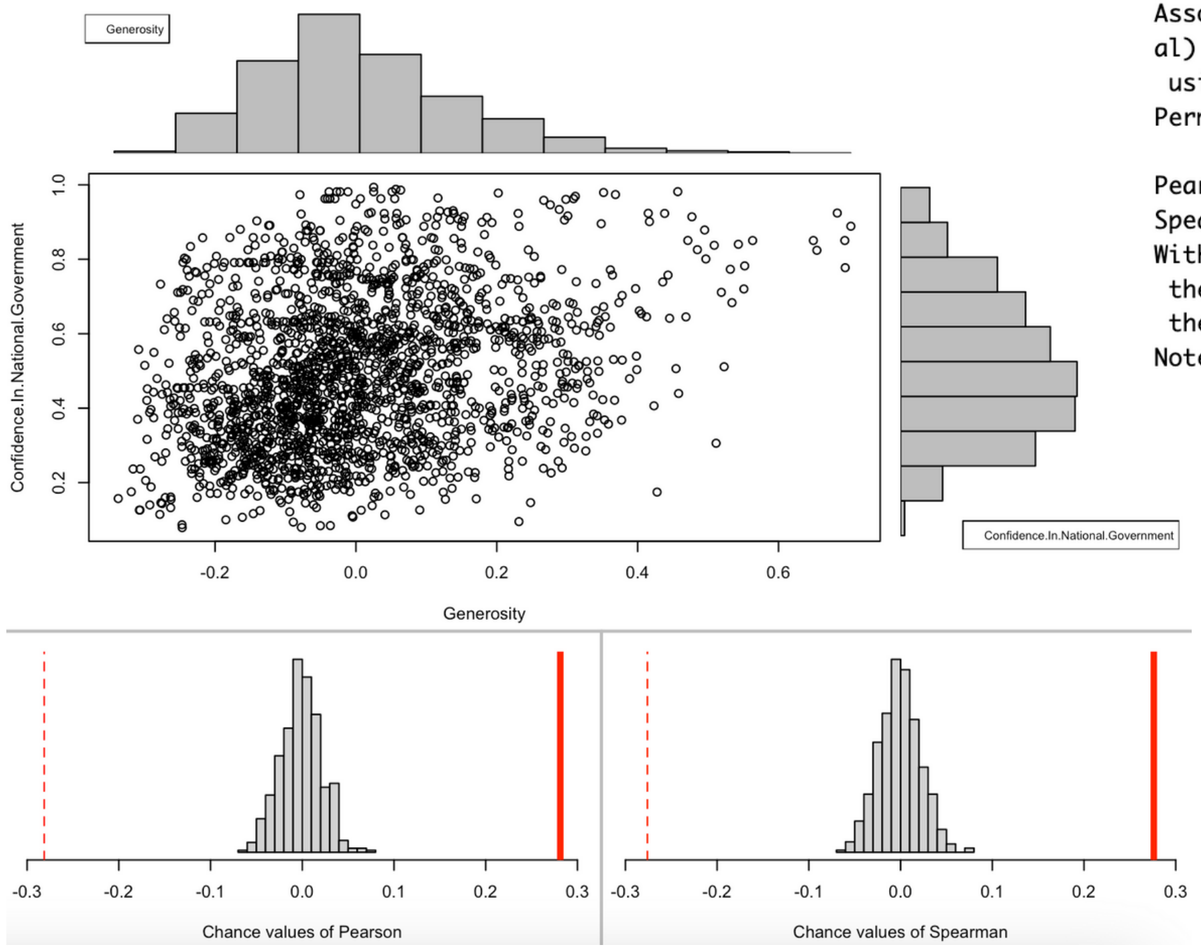
VS Freedom to make choices




```
Call:
lm(formula = Confidence.In.National.Government ~ Freedom.To.Make.Life.Choices,
    data = WHR)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41662 -0.14189 -0.00889  0.12713  0.53652

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.03518    0.02269    1.55    0.121
Freedom.To.Make.Life.Choices  0.61218    0.02974   20.58   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1814 on 1980 degrees of freedom
Multiple R-squared:  0.1762,    Adjusted R-squared:  0.1758
F-statistic: 423.6 on 1 and 1980 DF,  p-value: < 2.2e-16
```

Confidence index = 0.03518 + 0.62328Freedom
Each freedom index increase 0.62328 unit confidence index
RMSE : 0.1814 Multiple R-squared:17.62%

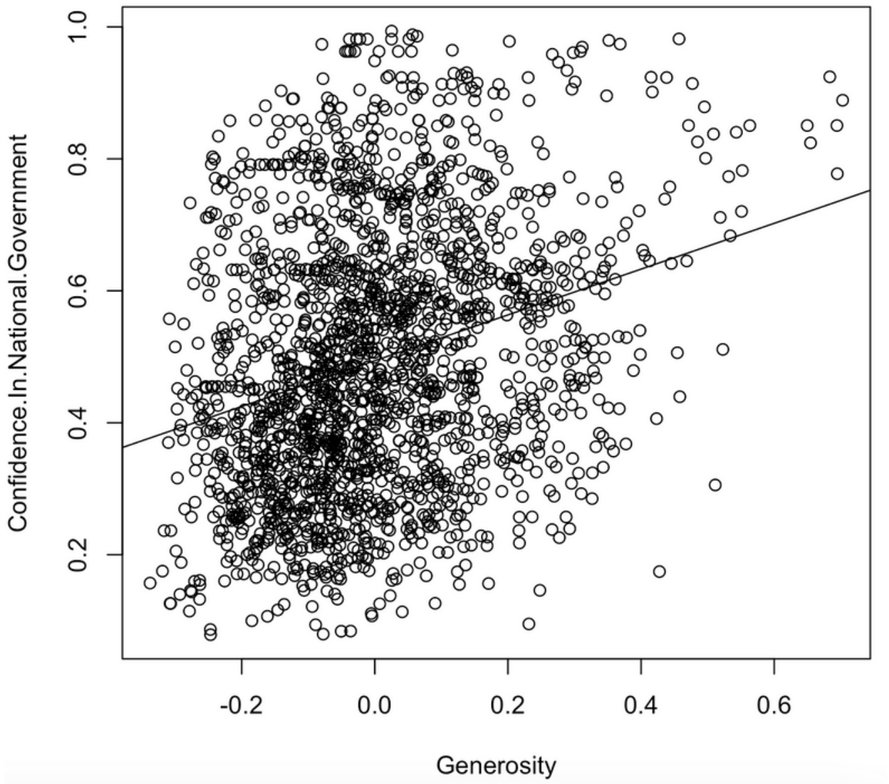
VS Generosity




```
Association between Generosity (numerical) and  Confidence.In.National.Government (numeric
al)
 using 1982 complete cases
Permutation procedure:
                                  Value Estimated p-value
Pearson's r                   0.2813602                 0
Spearman's rank correlation 0.2760717                   0
With 500 permutations, we are 95% confident that:
 the p-value of Pearson's correlation (r) is between 0 and 0.007
 the p-value of Spearman's rank correlation is between 0 and 0.007
Note:  If 0.05 is in this range, increase the permutations= argument.
```

Non-linear relationship
Monotonic
Correlation -> 0.28
Statistically Significant
Conclusive

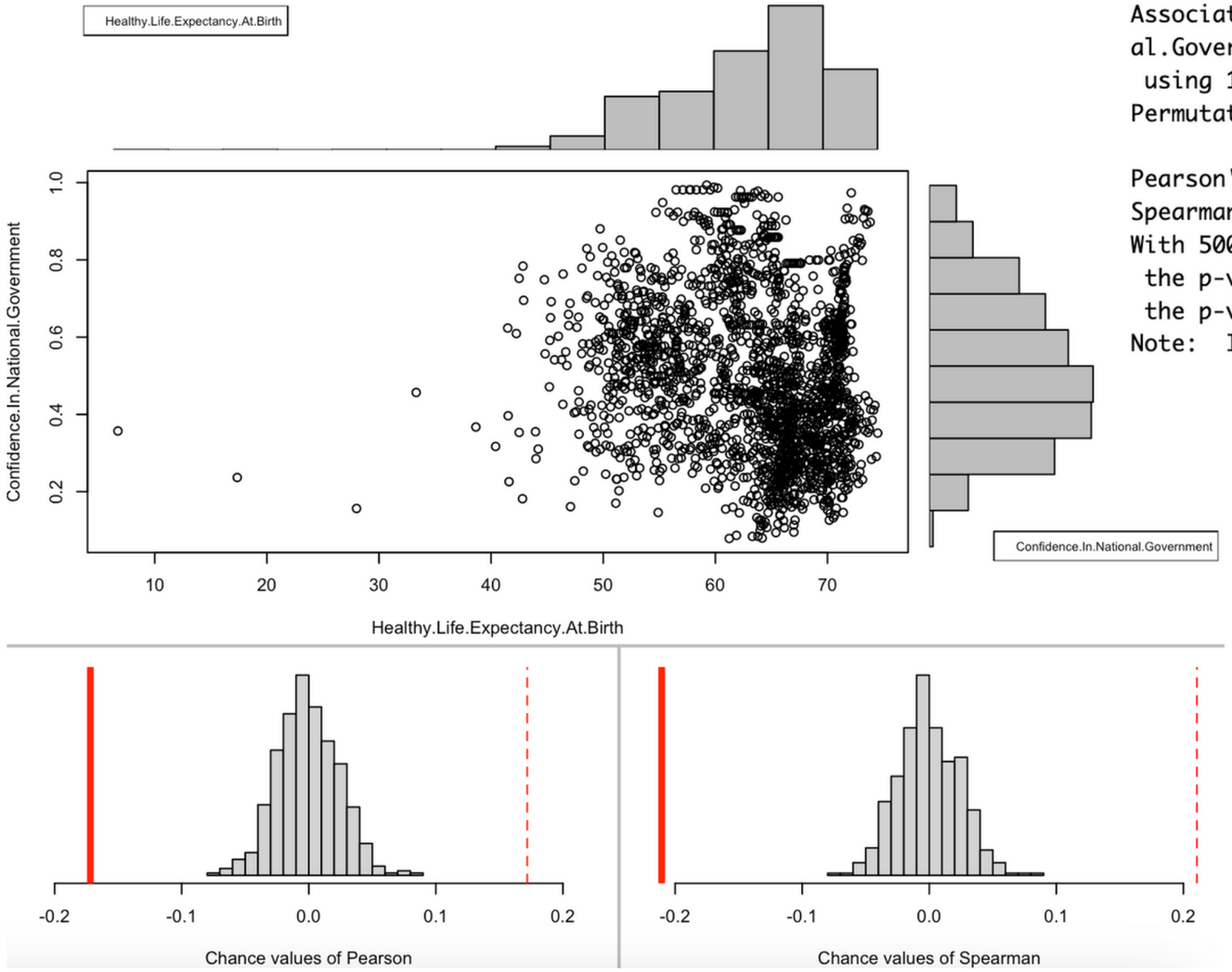
VS Generosity




```
Call:
lm(formula = Confidence.In.National.Government ~ Generosity,
    data = WHR)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4793 -0.1488 -0.0131  0.1274  0.5071

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.494114   0.004307  114.72   <2e-16 ***
Generosity  0.347135   0.026607   13.05   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1917 on 1980 degrees of freedom
Multiple R-squared:  0.07916,   Adjusted R-squared:  0.0787
F-statistic: 170.2 on 1 and 1980 DF,  p-value: < 2.2e-16
```

Confidence index = 0.494114+ 0.347135Generosity
Each Generosity index increase 0.347135 unit confidence index
RMSE : 0.1917 Multiple R-squared:7.916%

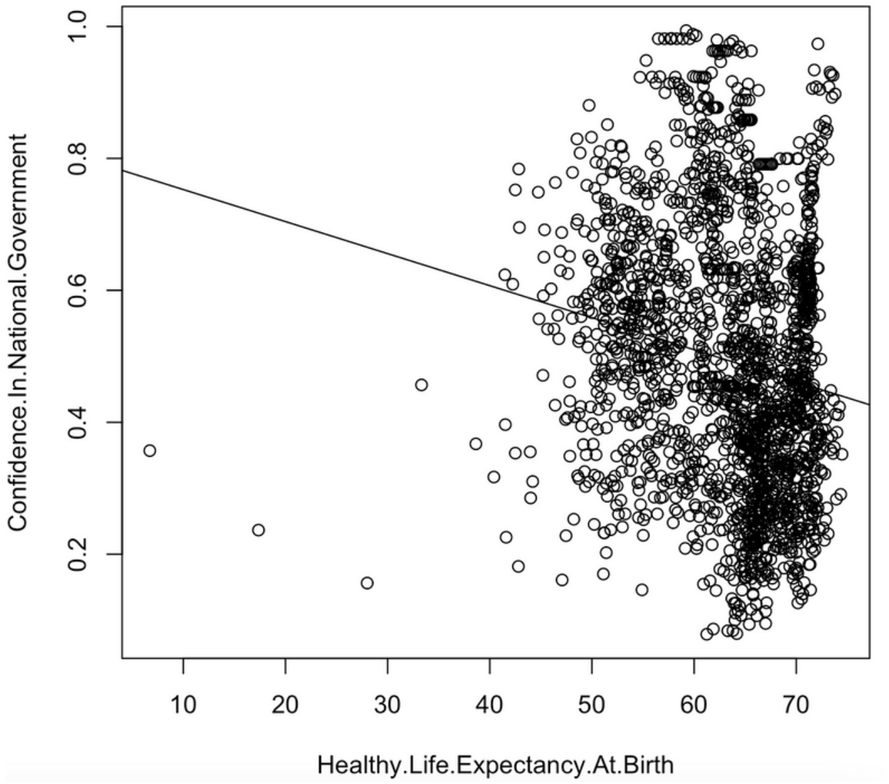
VS Healthy Life Expectancy




```
Association between Healthy.Life.Expectancy.At.Birth (numerical) and  Confidence.In.Nat
al.Government (numerical)
 using 1982 complete cases
Permutation procedure:
                                    Value Estimated p-value
Pearson's r                  -0.1719207                   0
Spearman's rank correlation -0.2106425                    0
With 500 permutations, we are 95% confident that:
 the p-value of Pearson's correlation (r) is between 0 and 0.007
 the p-value of Spearman's rank correlation is between 0 and 0.007
Note:  If 0.05 is in this range, increase the permutations= argument.
```

Non-linear relationship
Monotonic or not
Correlation -> -0.21
Statistically Significant
Conclusive

VS Healthy Life Expectancy




```
Call:
lm(formula = Confidence.In.National.Government ~ Healthy.Life.Expectancy.At.Birth,
    data = WHR)

Residuals:
     Min       1Q   Median       3Q      Max
-0.50901 -0.15123 -0.02237  0.13503  0.52241

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      0.8013271  0.0397442  20.162  < 2e-16 ***
Healthy.Life.Expectancy.At.Birth -0.0048554  0.0006252  -7.766 1.29e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1968 on 1980 degrees of freedom
Multiple R-squared:  0.02956,   Adjusted R-squared:  0.02907
F-statistic:  60.3 on 1 and 1980 DF,  p-value: 1.293e-14
```

Confidence index = 0.8013271- 0.0048554 HLE
Each Healthy life expectancy index decrease 0.0048554 unit confidence index
RMSE : 0.1968 Multiple R-squared:2.956%

Regression Summary

INDEPENDENT VARIABLE	R-SQUARED R^2	RMSE
Region	32.84%	0.1641
Perception of corruption	22.31%	0.1761
Freedom to make life choices	17.62%	0.1814
Generosity	7.916%	0.1917
GDP per capita	3.89%	0.1959
Healthy Life Expectancy	2.956%	0.1968
Social support	2.639%	0.1972
Year	0.06%	0.1997

Additional Packages

dplyr& magrittr

```
data_with_median <- WHR %>%
 group_by(Country.Name) %>%
                    mutate(Confidence_In_National_Government_Median=
median(Confidence.In.National.Government, na.rm = TRUE))
WHR <- data_with_median %>%
                                mutate(Confidence.In.National.Government=
ifelse(is.na(Confidence.In.National.Government),

Confidence_In_National_Government_Median,
                                Confidence.In.National.Government)) %>%
 select(-Confidence_In_National_Government_Median)
```

Conclusion

- In five variables, Region, Freedom to make life choices and Generosity are the three top indicators that influence the index of confidence of national government
  -> Why is it that only East Asia has a higher p-value?
- Healthy life expectancy is the only indicator with a negative correlation
  -> Residents of developed countries have lower confidence in their national government?

Thank You So Much For Listening !