

Assignment 2

World Freedom Data 2016-2017

Sankalp Dangwal S4000592

Introduction

The World Freedom Report examines the levels of Freedom in 155 countries by analyzing survey data collected from nationally representative samples of each country's population between 2013 and 2019. This data helps estimate how economic production, social support, life expectancy, freedom, absence of corruption, and generosity contribute to Freedom evaluations. Our report aims to investigate whether there was a statistically significant change in the Freedom scores of the surveyed countries between 2015 and 2019.

Problem Statement

Measuring Freedom levels is important because it helps identify policies that promote the well-being of populations, making it valuable for politicians, citizens, and policy planners seeking to improve their societies. To address our research question, we will utilize statistical analysis. We will focus on specific data from the World Freedom Report, including the countries' geographical location and their reported Freedom scores for 2015 and 2019. Descriptive statistics will be used to summarize the data, and a boxplot will provide a visual representation of potential differences in the reported Freedom levels between the two years. To determine the statistical significance of any observed changes, we will conduct an independent samples t-test on the data.

Data

The data was obtained from the Kaggle open data website (URL: https://www.kaggle.com/unsdsn/world-happiness?fbclid=IwAR30I6bEoTj8lUUB8et6CpVWFoEd3HAjwNCSHy0G_3UTPGSRNESG1ISjBvo).

The data contains freedom scores for each country, these being a metric measure obtained from the Gallup World Poll which asked respondents to rank the quality of their life on a scale of 0 (the worst) to 10 (the best). The responses were used to measure freedom in the 155 countries surveyed by the poll, and were drawn from nationally representative samples. Gallup weights were used to transform the responses into representative estimates.

Data Cont.

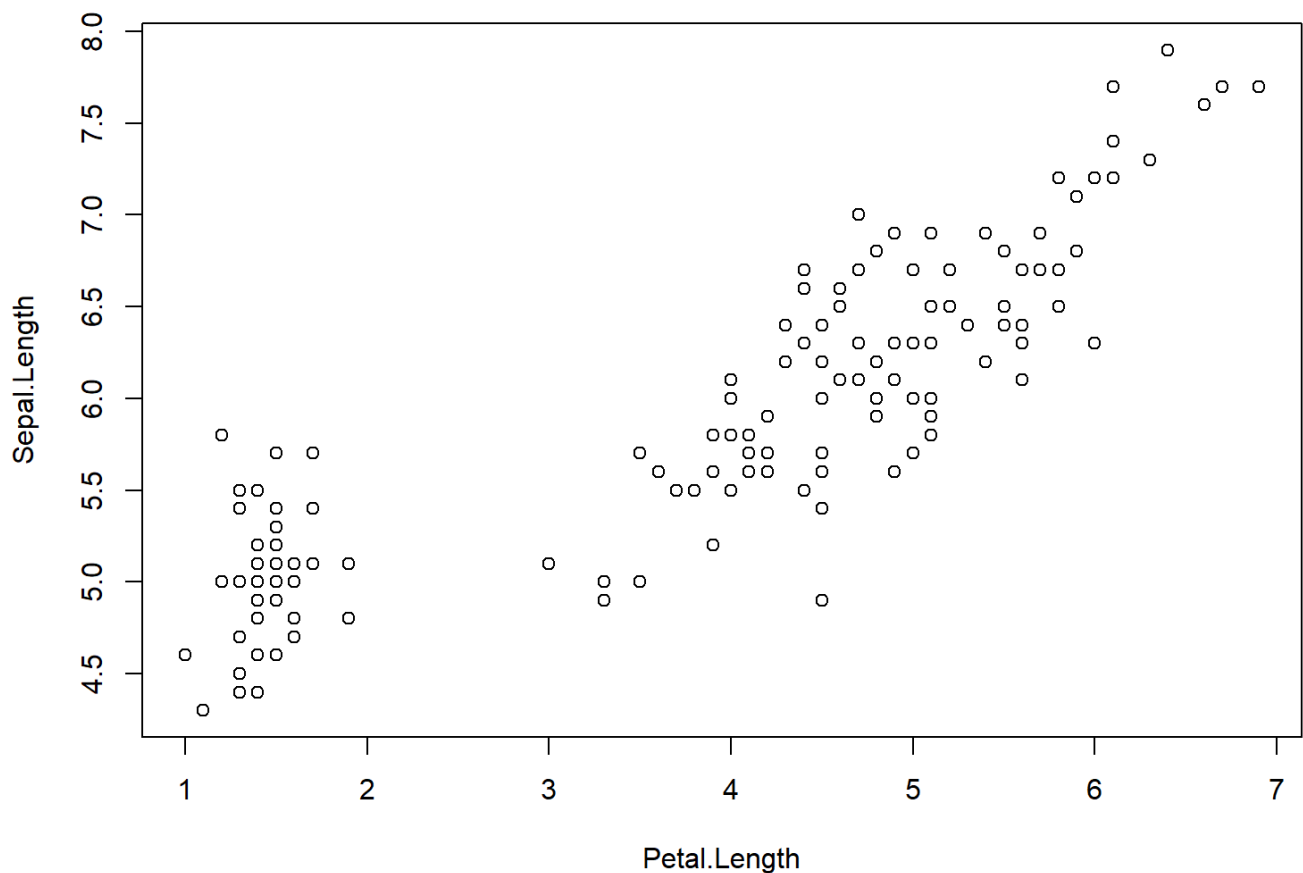
read in data

```
df_2016 = read.csv("C:/Users/sanka/OneDrive/Desktop/Dataset R/2016.csv")
df_2017 = read.csv("C:/Users/sanka/OneDrive/Desktop/Dataset R/2017.csv")
```

Subset and join the relevant data

```
df_2016 %>% select(Country, 'Freedom') -> df_2016_sub
df_2017 %>% select(Country, 'Freedom') -> df_2017_sub
df_2017_sub %>% rename("Country"="Country", "2017"="Freedom") -> df_2017_sub
df_2016_sub %>% rename("2016"="Freedom") -> df_2016_sub
df_2016_sub %>% left_join(df_2017_sub, by = "Country") -> DF
head(DF)
```

```
plot(Sepal.Length ~ Petal.Length, data = iris)
```



Data cont. gather

```
DF <- DF %>% gather("2016", "2017", key = "Year", value = "Freedom")
DF$Year <- as.factor(DF$Year)
str(DF)
```

```
## 'data.frame':   314 obs. of  3 variables:
## $ Country: chr  "Denmark" "Switzerland" "Iceland" "Norway" ...
## $ Year   : Factor w/ 2 levels "2016","2017": 1 1 1 1 1 1 1 1 1 ...
## $ Freedom: num  0.579 0.586 0.566 0.596 0.571 ...
```

Decscriptive Statistics and visualisation

```
table1 <- DF %>%  
  group_by(Year) %>%  
  summarize(Min = min(Freedom, na.rm = TRUE),  
            Q1 = quantile(Freedom, probs = 0.25, na.rm = TRUE),  
            Median = median(Freedom, na.rm = TRUE),  
            Q3 = quantile(Freedom, probs = 0.75, na.rm = TRUE),  
            Max = max(Freedom, na.rm = TRUE),  
            Mean = mean(Freedom, na.rm = TRUE),  
            SD = sd(Freedom, na.rm = TRUE),  
            n = n(),  
            Missing = sum(is.na(Freedom)))
```

Decsriptive Statistics and visualisation cont.

knitr::kable(table1)

Year	Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
2016	0	0.2574800	0.3974700	0.4845300	0.6084800	0.3709939	0.1455068	157	0
2017	0	0.3022089	0.4388801	0.5198853	0.6582487	0.4091140	0.1517532	157	7

Decsriptive Statistics and visualisation cont.

```
sum(sapply(DF, is.nan))
```

```
## [1] 0
```

Decsriptive Statistics and visualisation cont.

```
colSums(is.na(DF))
```

##	Country	Year	Freedom
##	0	0	7

Hypothesis Testing

To examine the hypothesis of a change in reported freedom between 2016 and 2017, we will employ a two-sample t-test. This test is selected due to the expectation that the demographic composition of a country may vary over time, resulting in different samples used to assess happiness levels in 2016 and 2017.

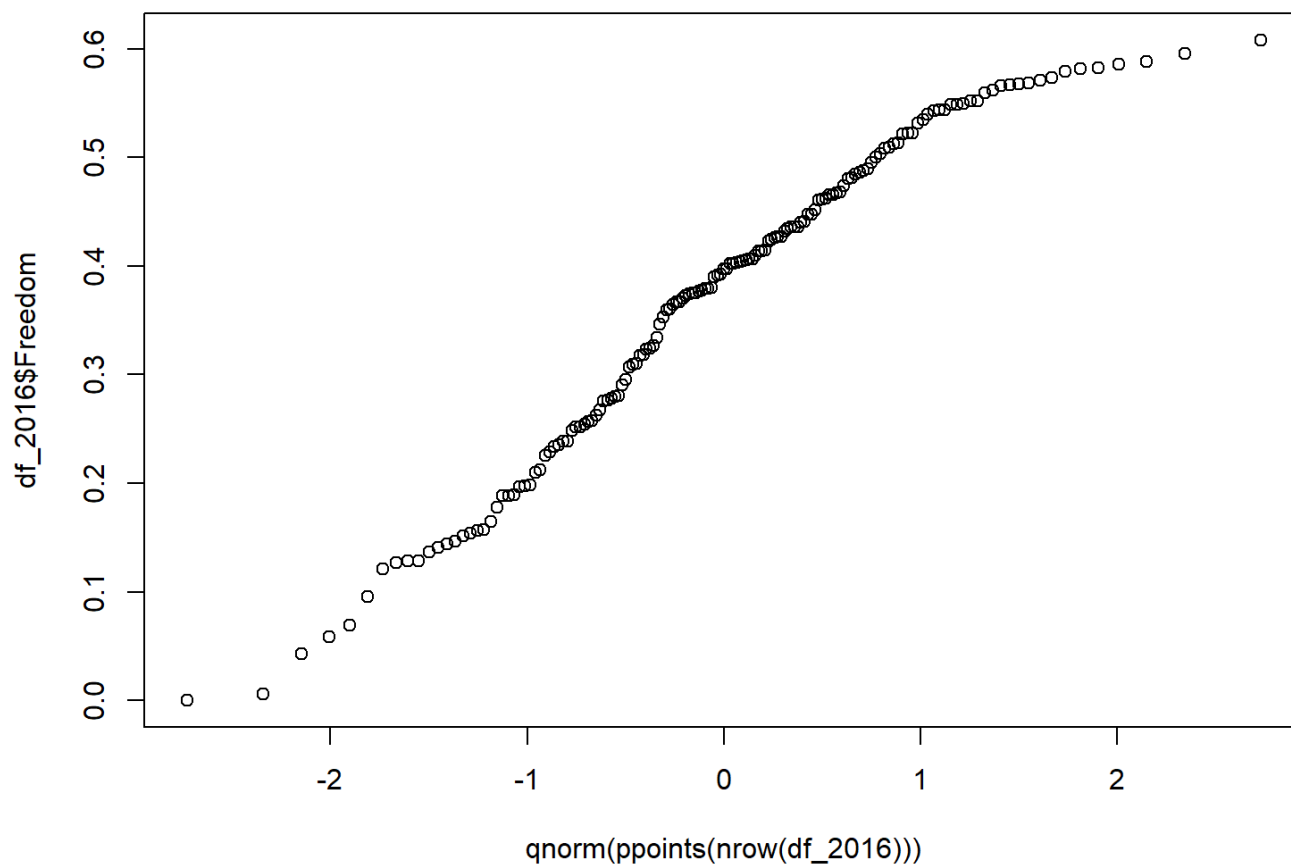
```
model1 <- lm(Sepal.Length ~ Petal.Length, data = iris)
model1 %>% summary()
```

```
##
## Call:
## lm(formula = Sepal.Length ~ Petal.Length, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24675 -0.29657 -0.01515  0.27676  1.00269
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.30660     0.07839   54.94  <2e-16 ***
## Petal.Length   0.40892     0.01889   21.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4071 on 148 degrees of freedom
## Multiple R-squared:  0.76, Adjusted R-squared:  0.7583
## F-statistic: 468.6 on 1 and 148 DF, p-value: < 2.2e-16
```

Hypthesis Testing Cont.

```
df_2016 <- DF %>% filter(Year == 2016)
qqplot(qnorm(ppoints(nrow(df_2016))), df_2016$Freedom, main = "QQPlot for 2016 Freedom")
```

QQPlot for 2016 Freedom



Discussion

Our investigation found that global freedom scores did not significantly change in the period between 2016 and 2017. We reached this conclusion based on the p-value of our test statistic, which was greater than the significance level.

The strengths of our investigation: - The data we used had a large sample size. Most countries on the planet were included in the World Gallup Poll, which strengthens the generalizability of our results. - The data we used was consistent, as all respondents were asked the same question when reporting their freedom levels.

The limitations of our investigation: - The data used to calculate freedom scores was derived from the answers of a survey. Survey responses can sometimes be untrustworthy sources of information as respondents may not answer the questions honestly, may omit answers, and may respond unconsciously. - Freedom is a contestable concept that can take different meanings depending on its cultural context. Comparing freedom levels across different countries using a singular metric does not account for these cultural differences. A similar problem applies to the variations in freedom's meaning from one person to another, making it hard to objectively compare the concept across a population's members.

Directions for future investigations: - Since our investigation did not find a statistically significant change in freedom levels for the period 2016-2017, we advise future research to use alternative data collection methods to explore the question of global freedom levels further. - For example, researchers could use focus groups and interviews to obtain more nuanced categorical data concerning global populations' self-reported freedom levels. Researchers would thus not have to rely on a singular metric to quantify a term as ambiguous as freedom. This would also give researchers more flexibility in accounting for possible cultural differences. - Researchers could also look at the data for a wider timeframe, as the change in freedom levels over time might be more suited to an evaluation in terms of decades rather than years.

One take-home message: - Though our investigation did not find a statistically significant change in freedom levels between 2016 and 2017, this should not discourage further research from being undertaken. Through the application of different statistical methods and the use of different kinds of data, freedom measures may indeed prove to be a useful tool for future policy-making.

References

Data sourced from: https://www.kaggle.com/unsdsn/world-happiness?fbclid=IwAR30I6bEoTj8IUUB8et6CpVWFoEd3HAjwNCSHy0G_3UTPGSRNESGIIISjBvo