

DATA 606 Data Project Proposal

Nicholas Kunze

Data Preparation

```
library(dplyr)
library(ggplot2)
library(psych)

# load data
load("data/WVS_Cross-National_Wave_7_rData_v5_0.rdata")
wvs <- `WVS_Cross-National_Wave_7_v5_0`

# Luckily the data is already tidy, with each row representing one observation,
# a person surveyed. And each column represents a variable.
# As this includes every question from the survey, there are over 600 columns..
# Let's select only those that we need.
wvs <- select(wvs, B_COUNTRY_ALPHA, Q46, Q106, Q107, Q108)

# Rename columns to be more useful - each question in the master pdf maps to a
# column in the df
wvs <- setNames(wvs, c("Country", "Happiness", "Individual_Effort",
                      "Govt_Business_Ownership", "Personal_Responsibility"))

# Exclude respondents who did not respond or did not know any of the relevant questions
wvs <- subset(wvs, Happiness > 0 &
              Individual_Effort > 0 &
              Govt_Business_Ownership > 0 &
              Personal_Responsibility > 0)

# Let's also invert the score of all of the GO scores so that 10 represents the most
# individualistic, like the other two independent variables. I invert the scale by
# using max - VAL + min. Then we swap the name.
wvs$Govt_Business_Ownership <- 10 - wvs$Govt_Business_Ownership + 1
names(wvs)[names(wvs) == "Govt_Business_Ownership"] <- "Private_Business_Ownership"

# Do the same with happiness scores so higher = better
wvs$Happiness <- 4 - wvs$Happiness + 1

wvs[sample(nrow(wvs),5),]
```

```
##      Country Happiness Individual_Effort Private_Business_Ownership
## 33794     IDN         3                 3                         9
## 36323     IRN         3                 1                         3
```

## 8463	BRA	3	1	10
## 84174	UKR	3	7	3
## 1058	ARG	3	6	4
##	Personal_Responsibility			
## 33794		5		
## 36323		1		
## 8463		1		
## 84174		8		
## 1058		6		

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Are countries' populations' economic values, specifically individualist with less government oversight and responsibility, predictive of lower overall happiness?

Cases

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Each case represents a country. There are 64 countries represented in this dataset.

Data collection

Describe the method of data collection.

Data is collected by the World Values Survey. Each participatory nation has a national team responsible for surveying. The minimum sample size is 1200 per nation, with sampling done with different methodology per team. The surveyor will generally perform the data collection with a paper questionnaire at the respondent's place of residence.

A deeper look at the WVS's fieldwork and sampling can be obtained here.

Type of study

What type of study is this (observational/experiment)?

This is an observational study.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

Data is collected by WVS and is available online here: <https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>

For this project, the data archive was downloaded locally as a url to directly access the rdata from was not made available. The local rdat file was loaded then loaded.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E. & Puranen B. (2022): World Values Survey Wave 7 (2017-2022) Cross-National Data-Set. Version: 4.0.0. World Values Survey Association. DOI: doi.org/10.14281/18241.18

Dependent Variable

What is the response variable? Is it quantitative or qualitative?

The response variable is happiness score. It is discrete quantitative. It ranges from 1 (least happy) to 4 (most happy).

Independent Variable(s)

The independent variables are *Individual_Effort*, 'incentivization of individual effort', *Govt_Business_Ownership*, 'belief in government business ownership', and *Personal_Responsibility*, 'personal responsibility over government responsibility for population'. All are discrete quantitative and range from 1 to 10.

Q106	Incomes should be made more equal											There should be greater incentives for individual effort
	1	2	3	4	5	6	7	8	9	10		
Q107	Private ownership of business and industry should be increased											Government ownership of business and industry should be increased
	1	2	3	4	5	6	7	8	9	10		
Q108	Government should take more responsibility to ensure that everyone is provided for											People should take more responsibility to provide for themselves
	1	2	3	4	5	6	7	8	9	10		

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

```
summary(wvs$Happiness)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   3.000   3.143   4.000   4.000
```

```
summary(wvs$Individual_Effort)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   4.000   7.000   6.264   9.000  10.000
```

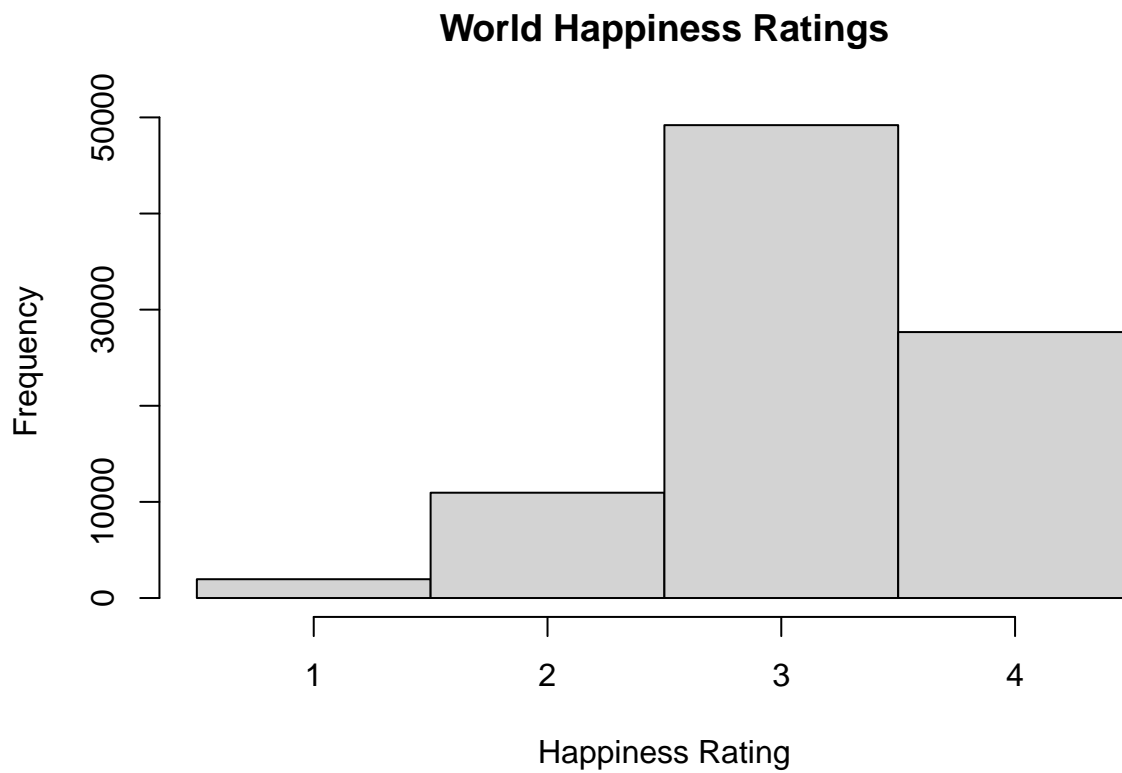
```
summary(wvs$Private_Business_Ownership)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   3.000   6.000   5.365   7.000  10.000
```

```
summary(wvs$Personal_Responsibility)
```

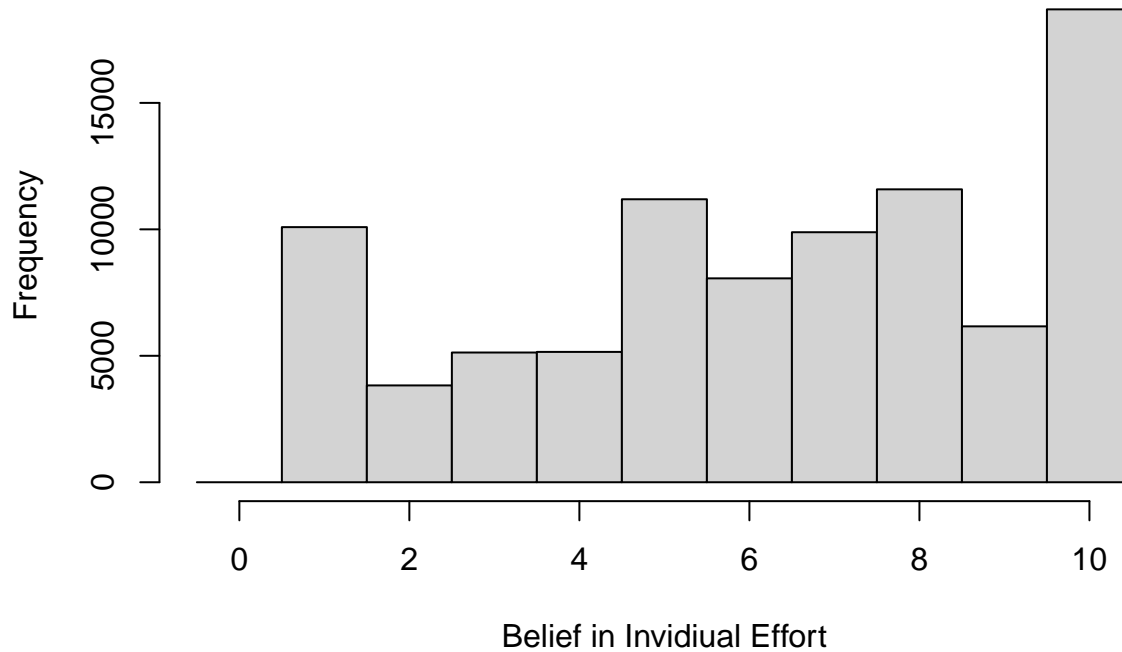
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   2.000   5.000   5.023   7.000  10.000
```

```
hist(wvs$Happiness, main="World Happiness Ratings",xlab="Happiness Rating",breaks=seq(0.5,4.5,by=1))
```



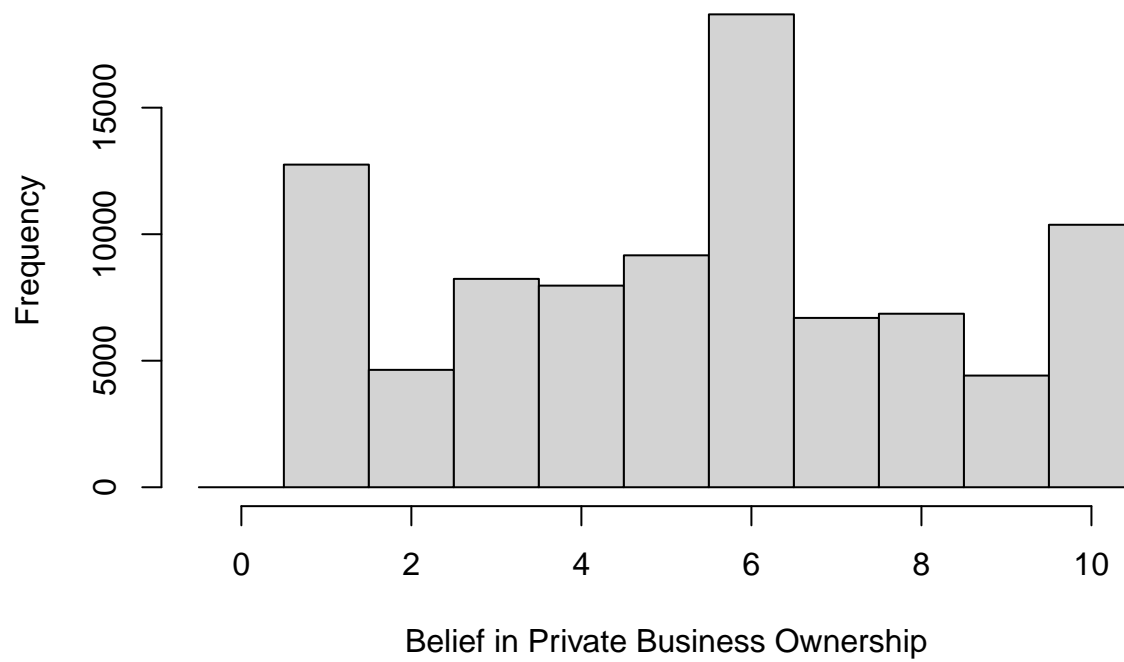
```
hist(wvs$Individual_Effort,
      main="Belief in Individual Effort Incentivization\n over Government-controlled Income Flattening",
      xlab="Belief in Invidiual Effort",
      breaks=seq(-0.5,10.5,by=1))
```

Belief in Individual Effort Incentivization over Government-controlled Income Flattening



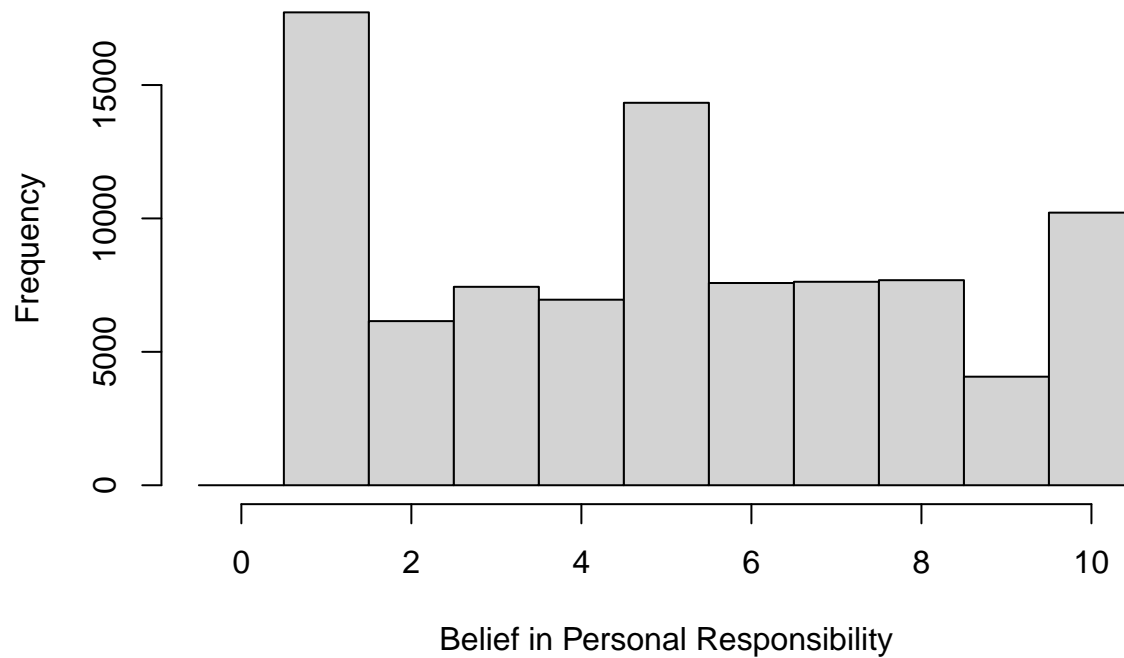
```
hist(wvs$Private_Business_Ownership,  
     main="Belief in Private over Public Business Ownership",  
     xlab="Belief in Private Business Ownership",  
     breaks=seq(-0.5,10.5,by=1))
```

Belief in Private over Public Business Ownership



```
hist(wvs$Personal_Responsibility,  
     main="Belief that People should Provide for Themselves More\n than Government Collective Responsibility",  
     xlab="Belief in Personal Responsibility",  
     breaks=seq(-0.5,10.5,by=1))
```

Belief that People should Provide for Themselves More than Government Collective Responsibility

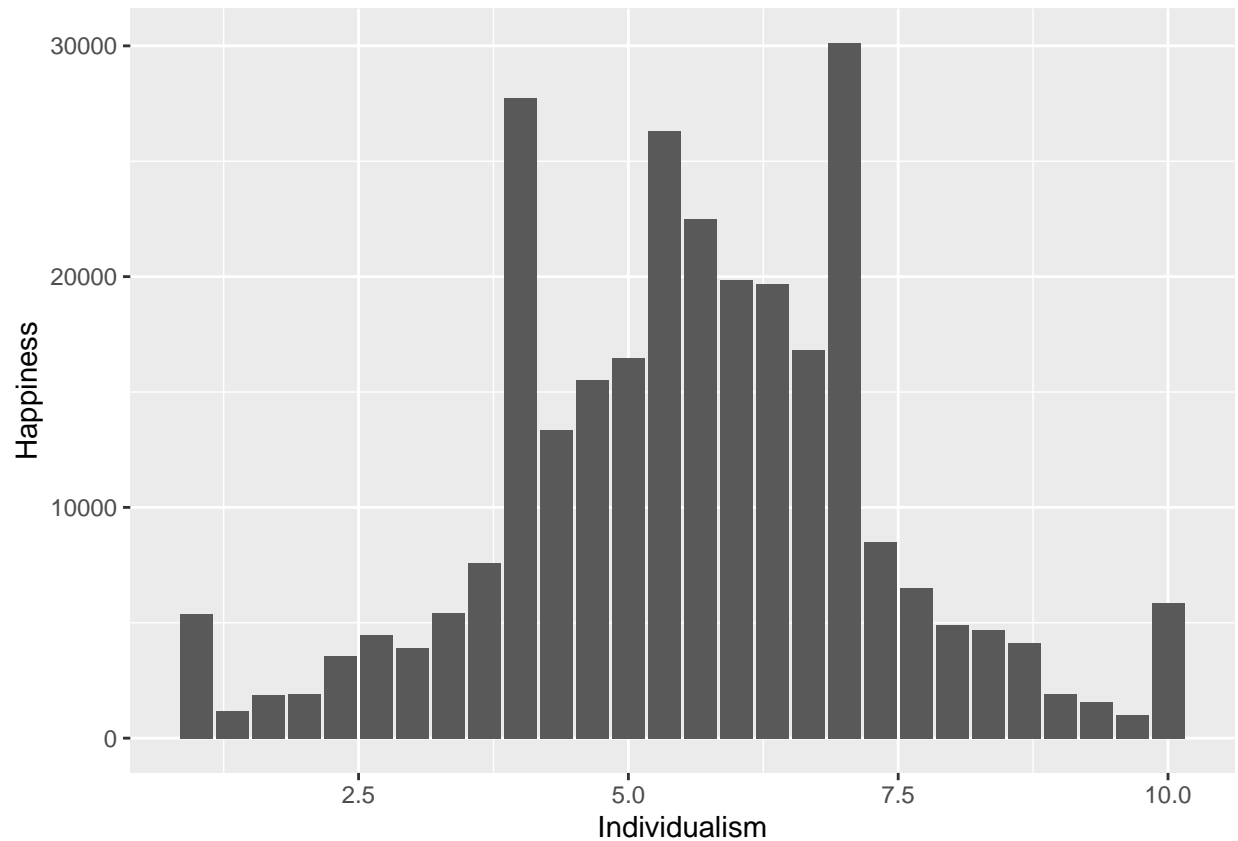


```
# Calculate survey respondent's 'individualism' as mean of the three variables
wvs$Individualism <- rowMeans(wvs[,c("Individual_Effort",
                                     "Private_Business_Ownership",
                                     "Personal_Responsibility")])

cor(wvs$Individualism, wvs$Happiness)
```

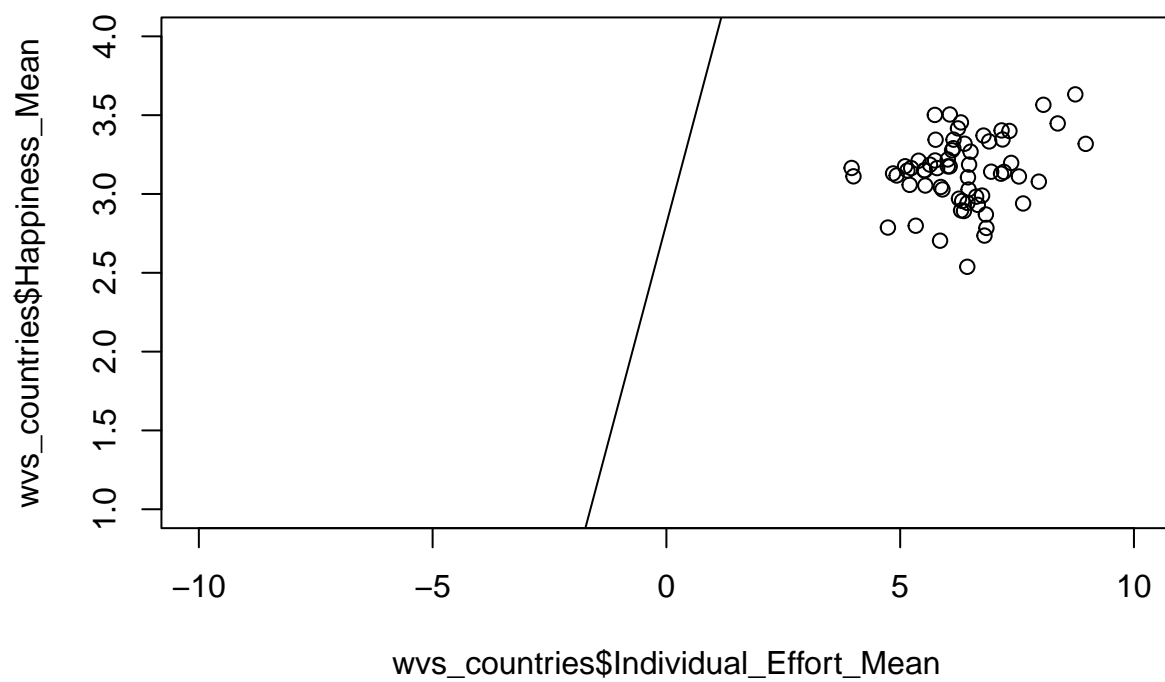
```
## [1] 0.0657328
```

```
ggplot(wvs, aes(x=Individualism, y=Happiness)) +
  geom_bar(stat="identity")
```

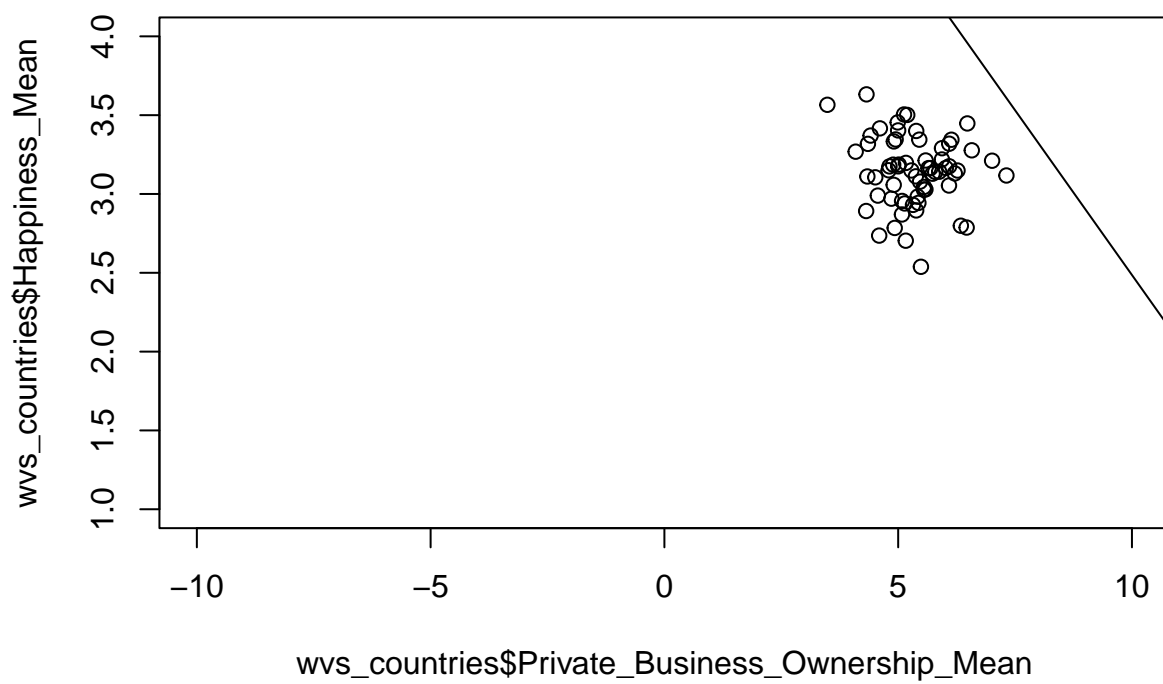


```
wvs_countries <- wvs %>%
  group_by(Country) %>%
  summarise(Country = unique(Country),
            Happiness_Mean = mean(Happiness),
            Individual_Effort_Mean = mean(Individual_Effort),
            Private_Business_Ownership_Mean = mean(Private_Business_Ownership),
            Personal_Responsibility_Mean = mean(Personal_Responsibility))

plot(wvs_countries$Individual_Effort_Mean,
     wvs_countries$Happiness_Mean, xlim=c(-10,10), ylim = c(1,4))
abline(lm(wvs_countries$Individual_Effort_Mean ~ wvs_countries$Happiness_Mean))
```

```
plot(wvs_countries$Private_Business_Ownership_Mean,  
     wvs_countries$Happiness_Mean, xlim=c(-10,10),ylim = c(1,4))  
abline(lm(wvs_countries$Private_Business_Ownership_Mean ~ wvs_countries$Happiness_Mean))
```



```
plot(wvs_countries$Personal_Responsibility_Mean,  
     wvs_countries$Happiness_Mean, xlim=c(-10,10),ylim = c(1,4))  
abline(lm(wvs_countries$Personal_Responsibility_Mean ~ wvs_countries$Happiness_Mean))
```

