THE DEGREES OF FREEDOM

Group C1

Carel Ong Shi Ting (A0157667M)

Swan Htet Lin (A0149926N)

DEFINING FREEDOM

Personal

- Measure by Personal Freedom Index (PFI)

Economic

- Measure by Economic Freedom Index (EFI)

OBJECTIVES

1) Significant predictors that maximise PFI & EFI

2) Does having Economic Freedom ensures Personal Freedom in a country? (vice versa)

OUTLINE

1) Description of dataset
2) Preliminary Plots
3) Model Construction
4) Variable Importance / Relationship
5) Cluster Analysis
6) Conclusion

DESCRIPTION OF DATASET

- 1458 observations with 78 predictor
- 1184 observations with 47 predictors after data cleaning

Response variables:

1) pf_score: Personal Freedom Index. (PFI)

- Ability of an individual to take any plausible actions free from external restraints or entities.

2) ef_score: Economic Freedom Index (EFI)

- Ability of an individual to control his or her own finance and property.

DESCRIPTION OF DATASET

Predictors:

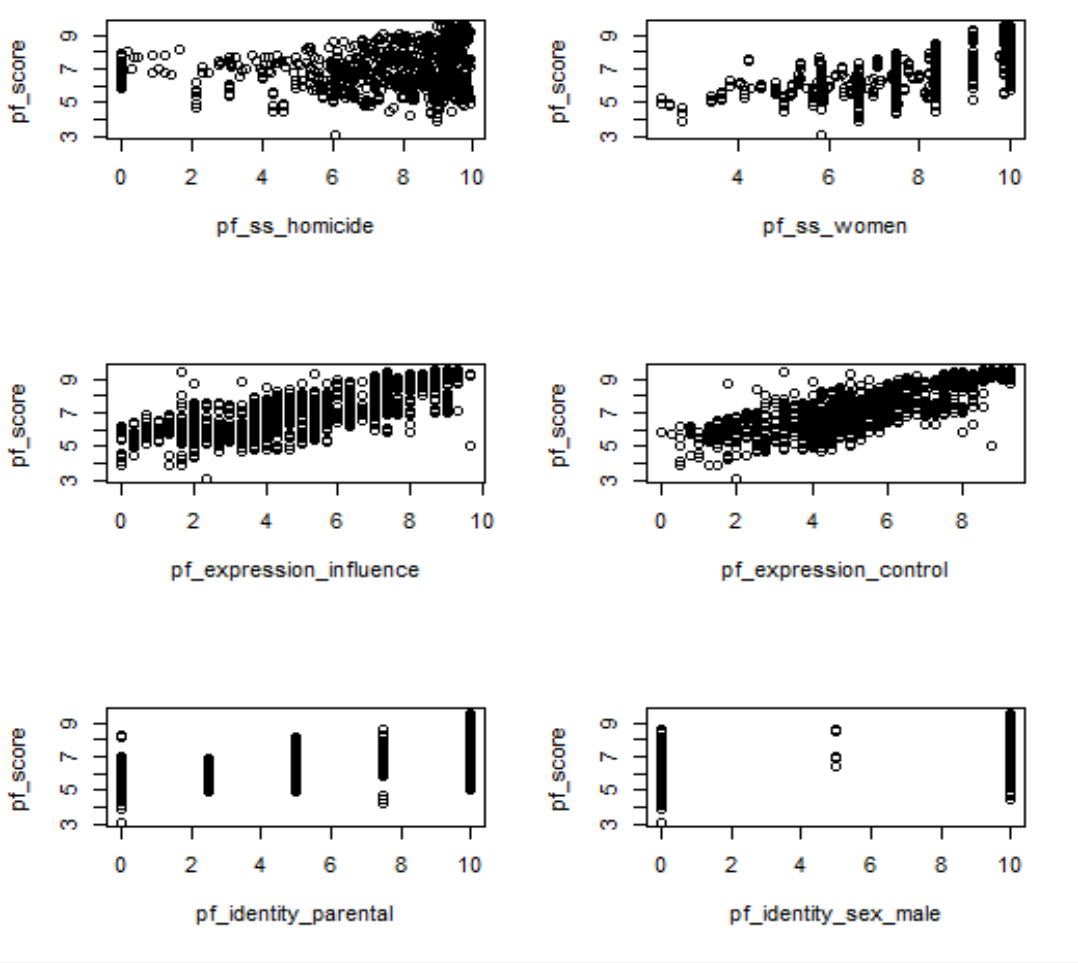
pf_rol (Predictor) – Measure of Rule of Law

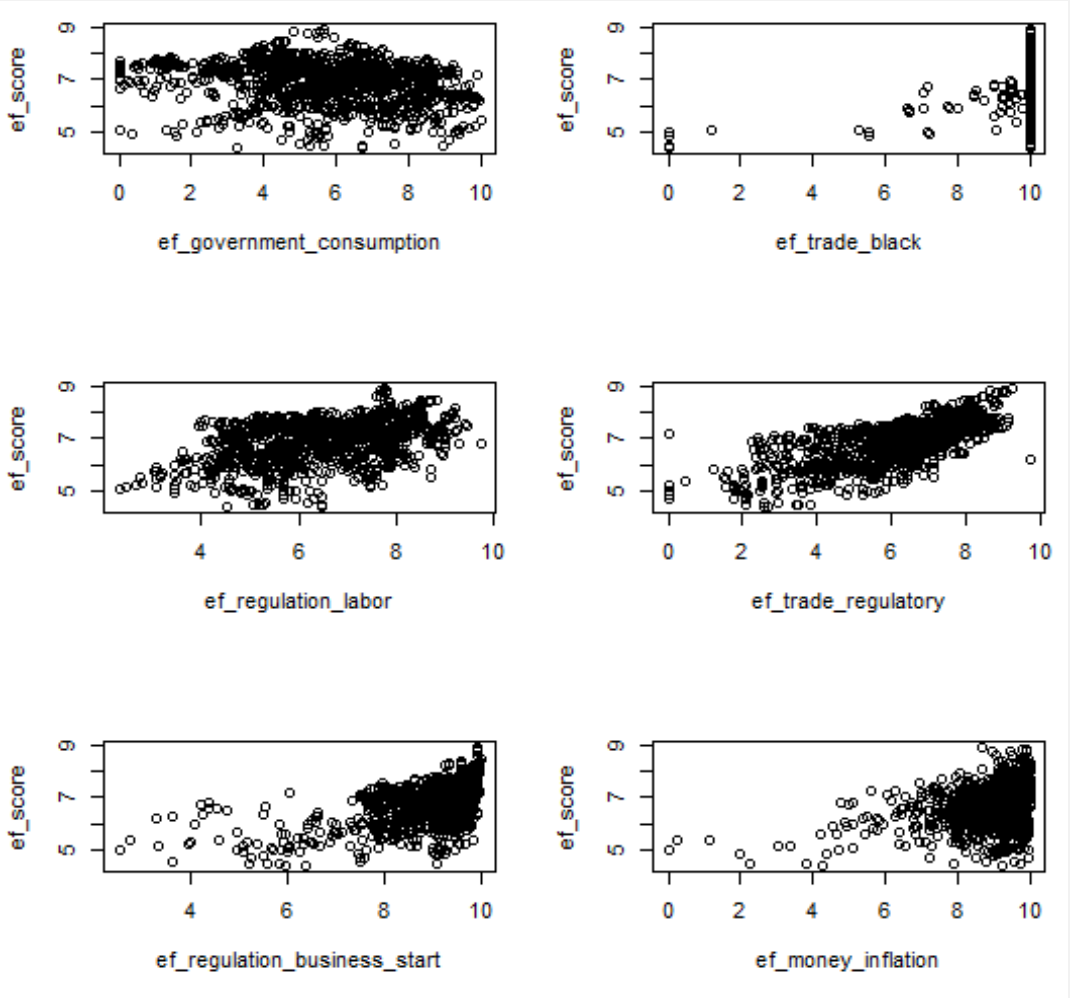
- Rule of Law encompasses mainly government integrity and judicial effectiveness and fairness.
- ⬆ in value suggests that more people are subjected to publicly disclosed legal codes and processes with greater equality.

ef_money (Predictor) – Measure of Sound Money

- Represents the freedom in having money that has a purchasing power determined by the markets and not by governments and political parties.

SCATTER PLOTS

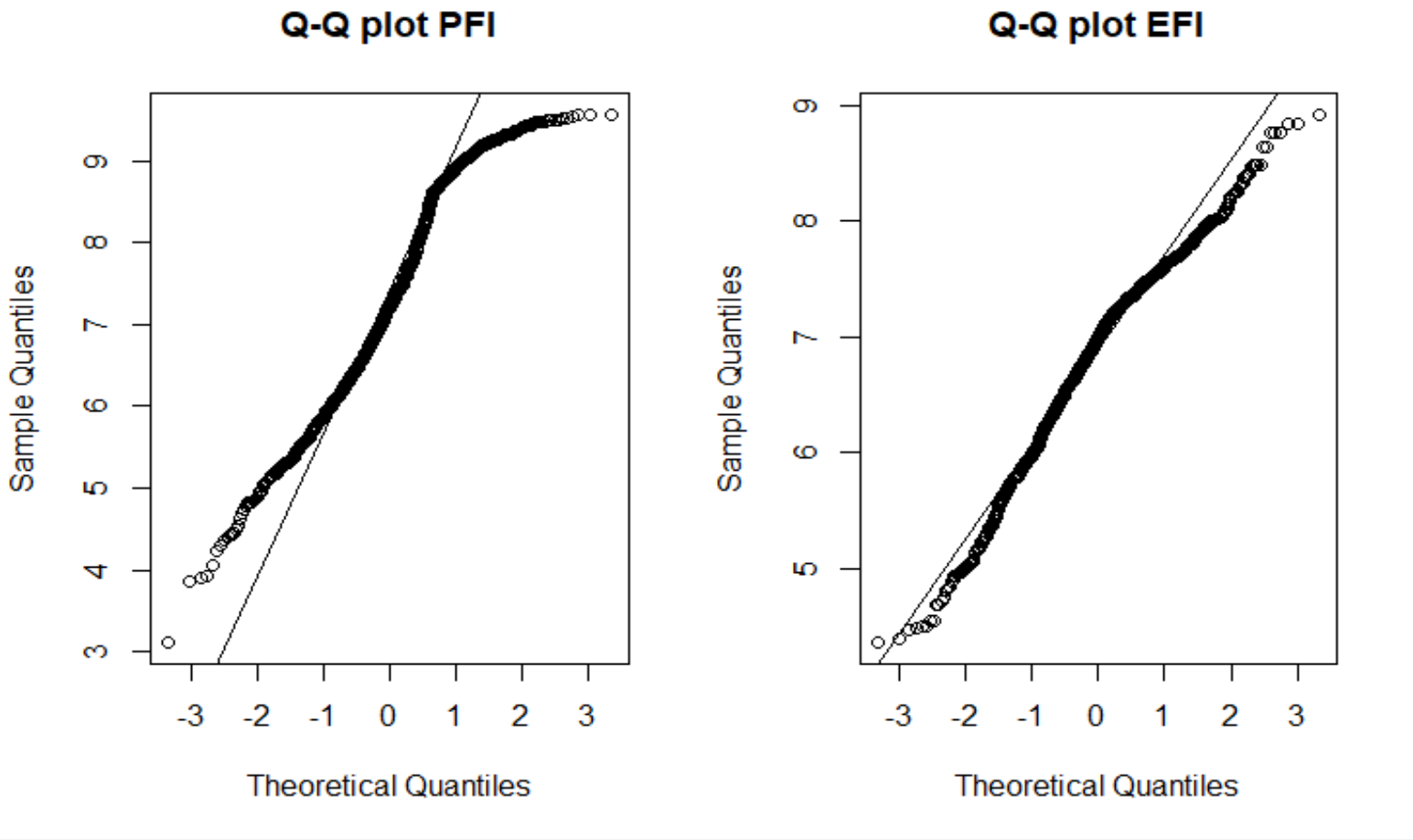




SCATTER PLOTS

Generally linear

NORMALITY PLOTS

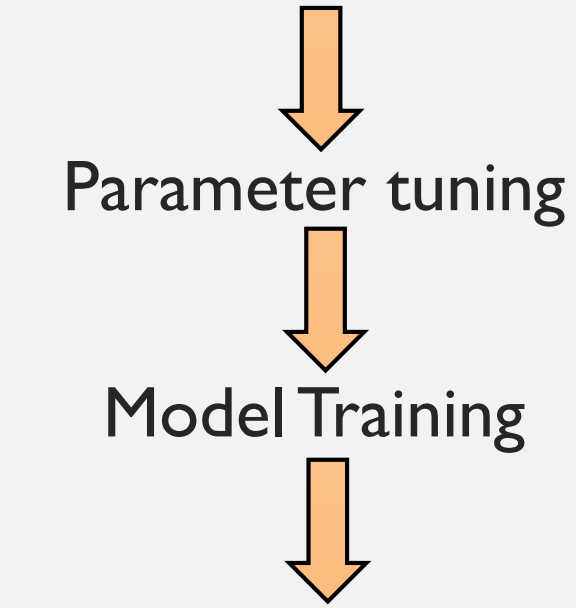


Non-parametric approaches may be suitable to fit the data.

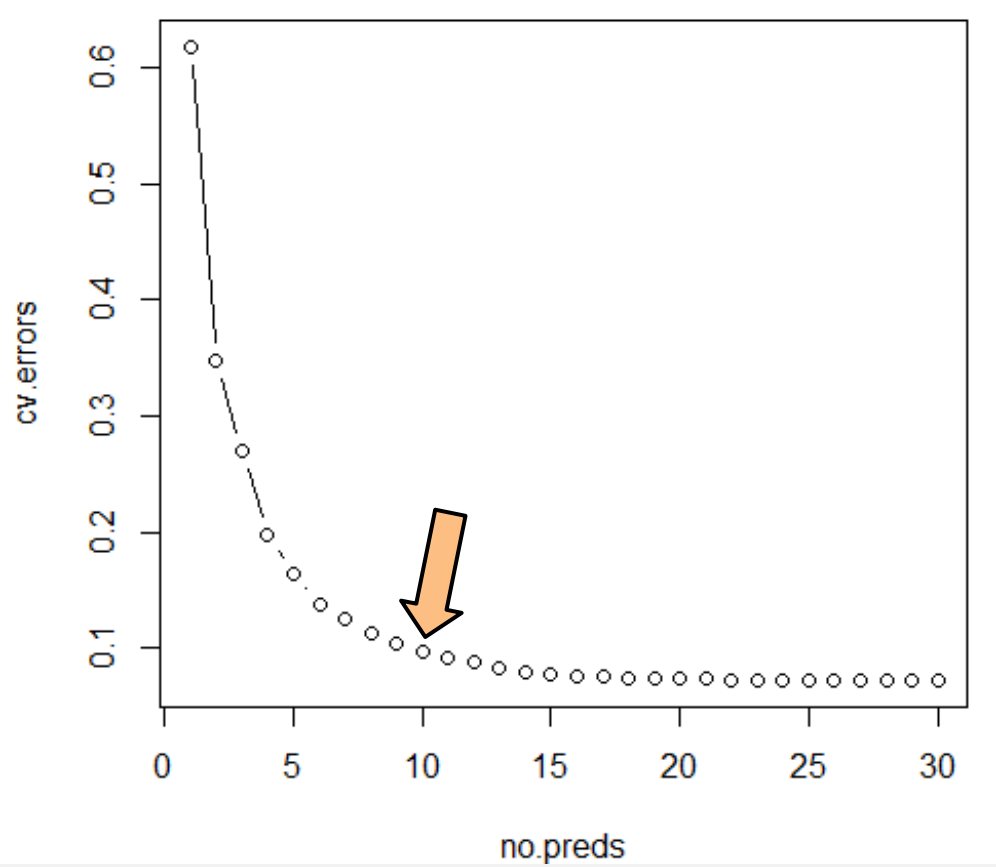
MODEL CONSTRUCTION

On a 80/20 training-test set,

- Multiple Linear Regression model
- Smoothing GAMs model
- Bagging
- Random Forest
- Boosting
- Neural Network



- PFI as response variable
- Selected 10 significant predictors from forward stepwise selection




```
lm(formula = get_model_formula(10, step, "pf_score"), data = hi[train,
    ])

Coefficients:
            (Intercept)          pf_ss_homicide     pf_ss_disappearances
                1.24774                 0.10802                 0.10426
          pf_ss_women     pf_movement_domestic      pf_movement_foreign
                0.16897                 0.05142                 0.04013
pf_expression_influence    pf_expression_control     pf_identity_parental
                0.10829                 0.10397                 0.05917
   pf_identity_sex_male          ef_legal_courts
                0.04326                 0.06334
```

Test MSE = 0.0984

- Introduce smoothing to predictors from linear regression model

```
Anova for Nonparametric Effects
                                Npar Df  Npar F     Pr(F)
(Intercept)
s(pf_ss_homicide, 5)                4   7.1359 1.169e-05 ***
s(pf_ss_women, 5)                   4   7.0988 1.251e-05 ***
s(pf_expression_influence, 5)       4  14.4514 1.896e-11 ***
s(pf_expression_control, 5)         4   8.3692 1.243e-06 ***
pf_identity_parental
pf_identity_sex_male
ef_legal_courts
s(pf_ss_disappearances, 5)          4   3.1737  0.01328 *
pf_movement_domestic
pf_movement_foreign
```

Test MSE = 0.0942

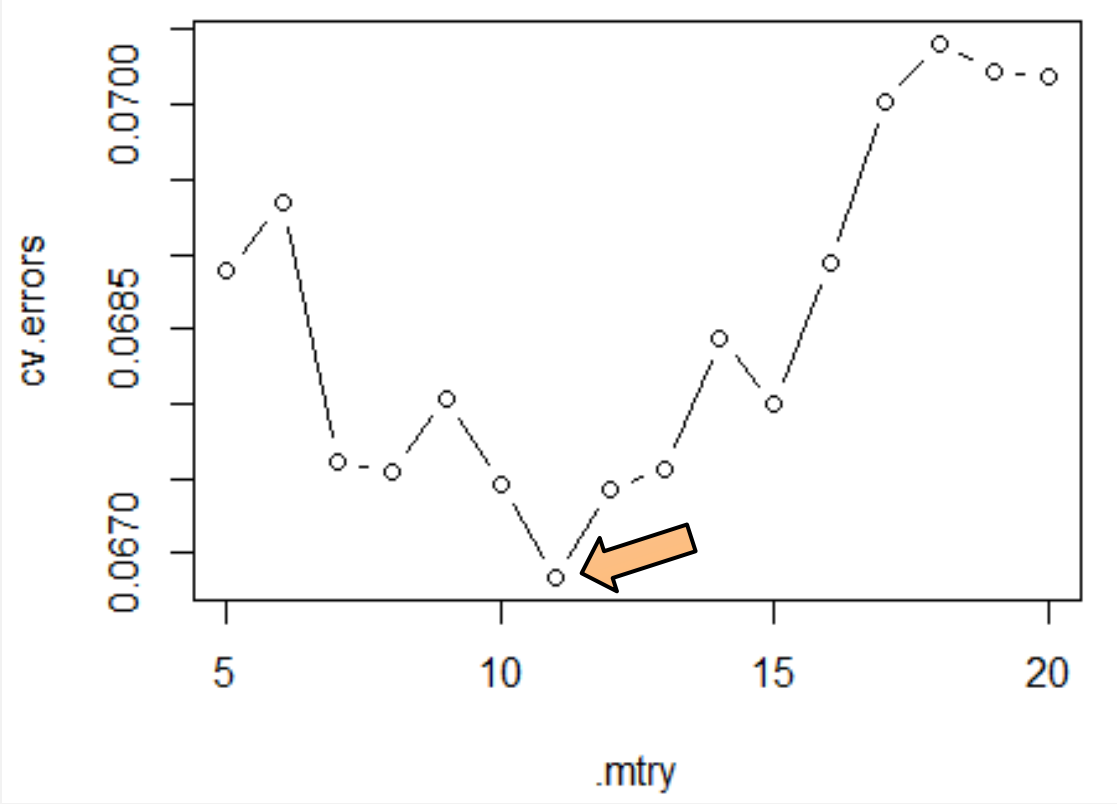
- Tree methods deals with the potential problem of multicollinearity with its built-in feature selection.

- Selected 1000 trees
- All 47 predictors considered for each split

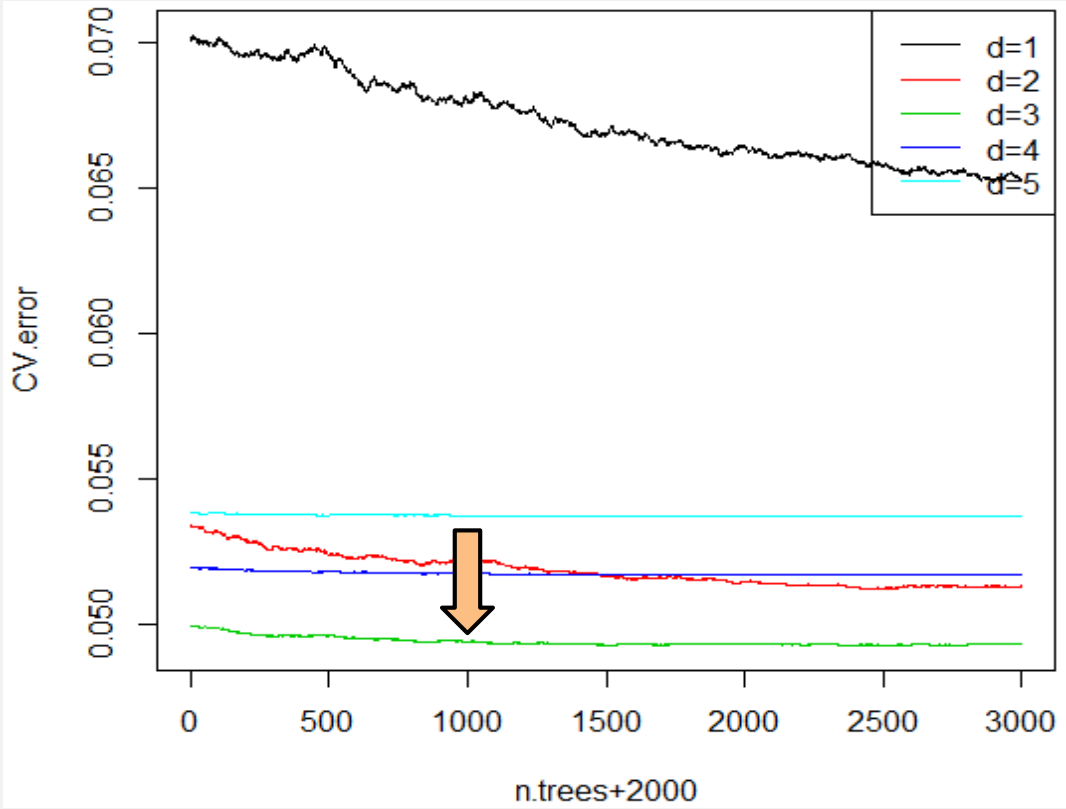
Test MSE = 0.0784

- Selected 1000 trees
- From cross validation, selected 11 predictors to be considered at each split. (mtry = 11)

Test MSE = 0.0642

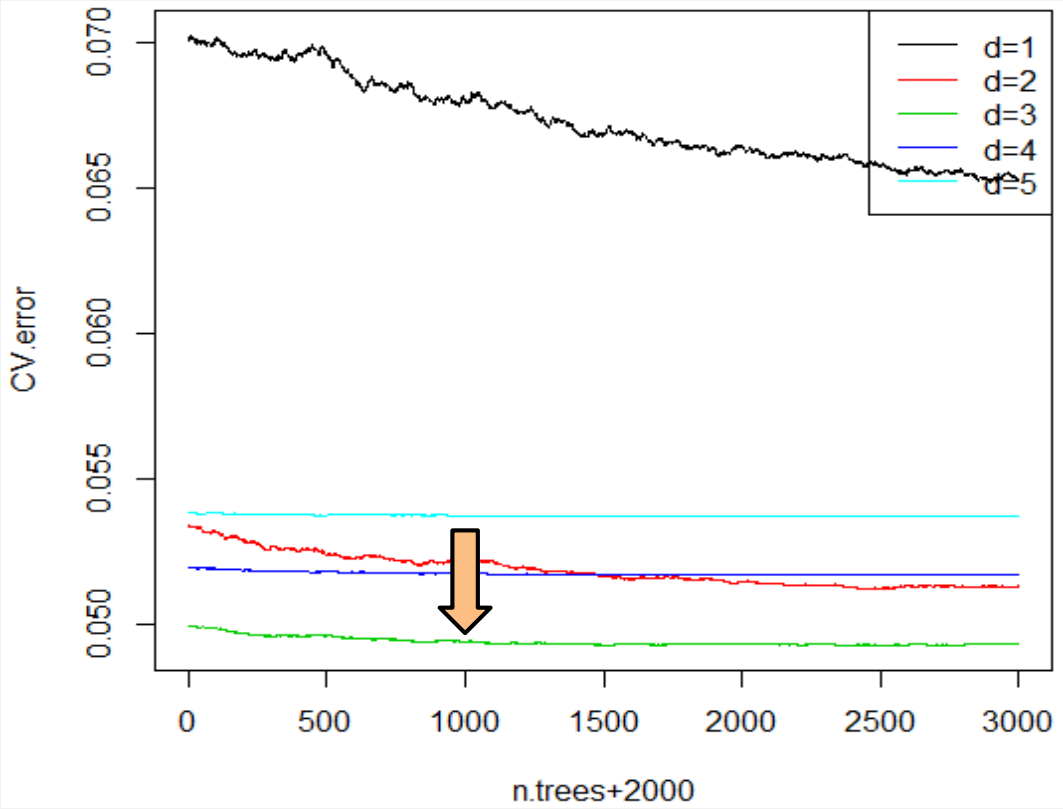


- The larger the learning rate, λ, the smaller number of trees, N is required.
- We tuned Interaction Depth and N, fixed $\lambda = 0.1$,



- The optimal number of trees we chose is 3000 with interaction depth = 3.

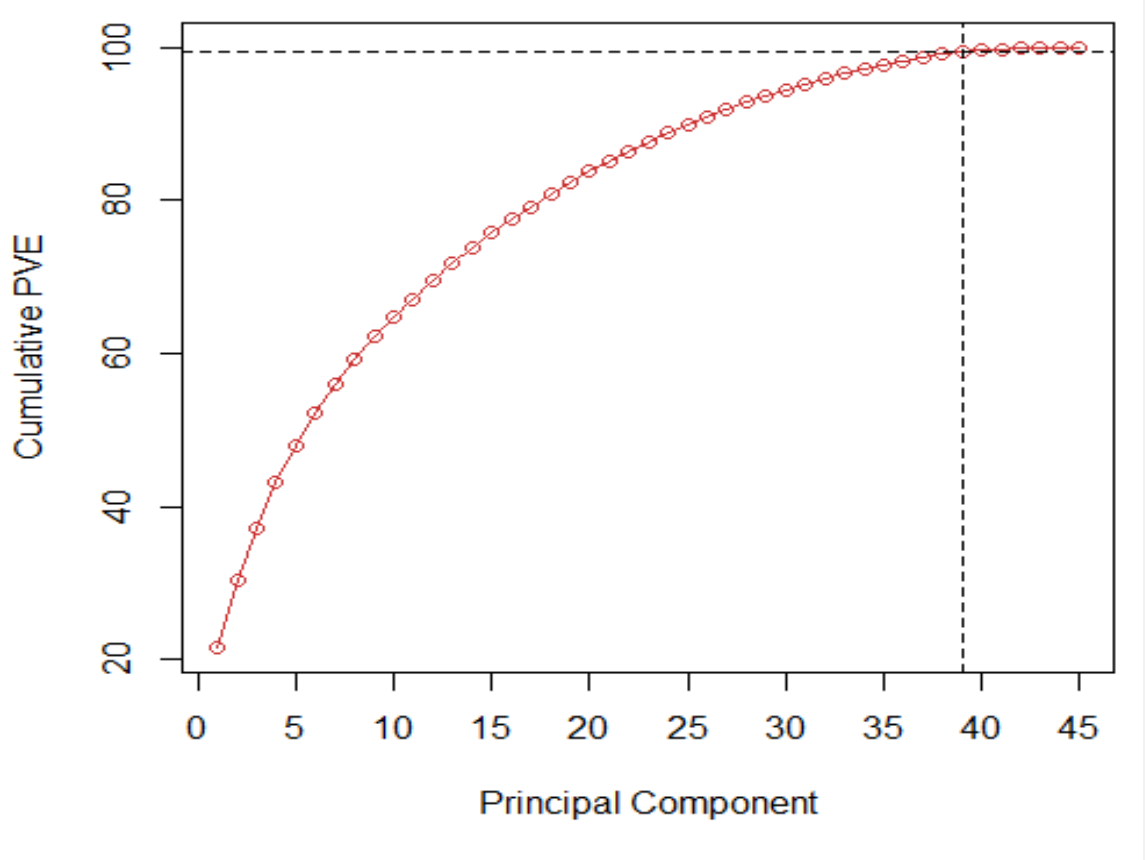
- The larger the learning rate, λ, the smaller number of trees, N is required.
- We tuned Interaction Depth and N, fixed $\lambda = 0.1$,



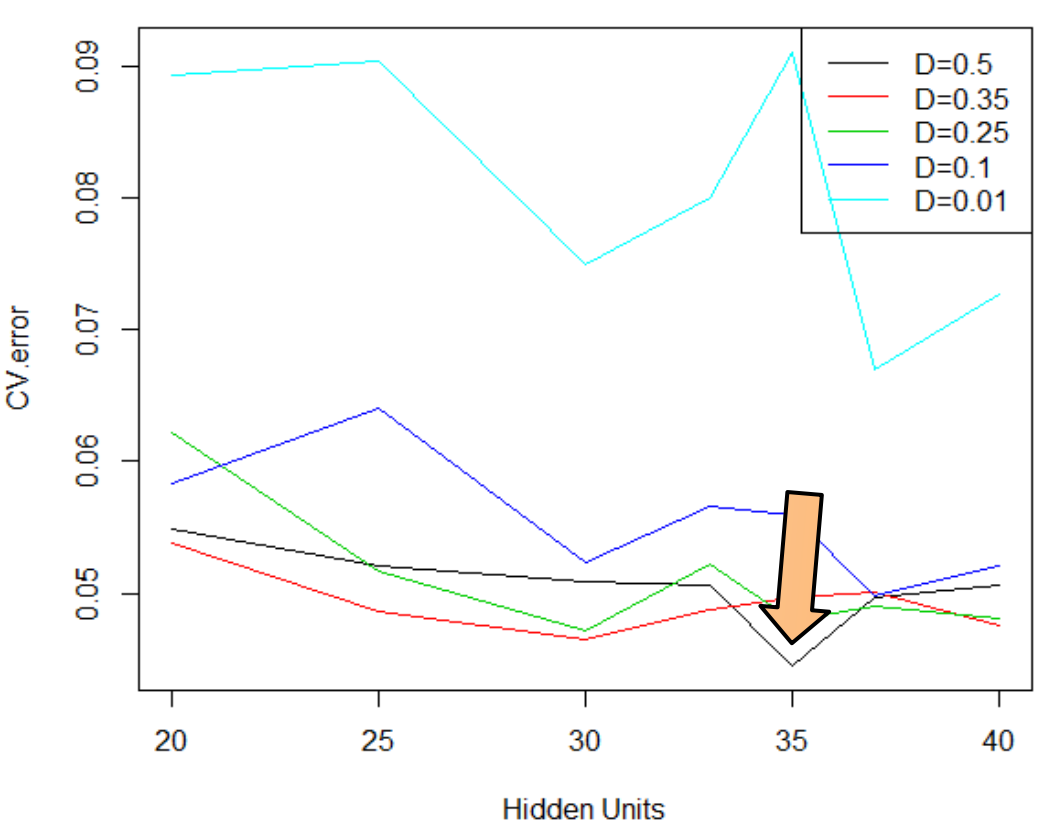
Test MSE = 0.0450

- The optimal number of trees we chose is 3000 with interaction depth = 3.

- PCA selected 39 principal components that explains about 99.5% of the variance
- 39 Inputs to Neural Nets



Neural Nets with 1 hidden layer :
Tuned Number of hidden units, M, and Weight Decay, D
Selected M = 35, and D = 0.5



Test MSE = 0.0411

MODEL PERFORMANCE

For PFI,

Approaches	Test MSE
Multiple Linear Regression (p = 10)	0.0984
Smoothing GAM (p = 10)	0.0942
Bagging (Trees =1000)	0.0784
Random Forest (mtry = 11)	0.0642
Boosting (Trees = 3000, λ = 0.1, d= 3)	0.0450
Neural Network (p=39, M=35, D=0.5)	**0.0411**

MODEL PERFORMANCE

For EFI,

Approaches	Test MSE
Multiple Linear Regression (p = 10)	0.0760
Smoothing GAM (p = 10)	0.0682
Bagging (Trees =1000)	0.0347
Random Forest (mtry = 10)	0.0278
Boosting (trees = 3000, $\lambda = 0.1$, d = 3)	**0.0181**
Neural Network (p=39, M=35, D=0.5)	0.0223

No free lunch theorem

VARIABLE IMPORTANCE

For PFI,

var	rel.inf
pf_expression_control	29.39756043
pf_ss_women	20.31574718
pf_ss_disappearances	8.76300093
pf_expression_influence	6.03968961
pf_ss_homicide	5.59731808
pf_movement_domestic	3.29614093
ef_legal_gender	2.97767054
pf_identity_parental	2.92823482
pf_movement_foreign	2.91018352
pf_identity_sex_male	2.49081807

- pf_expression_control and pf_ss_women are the 2 most important variables in predicting PFI.

For EFI,

var	rel.inf
ef_money_currency	27.978704187
ef_trade_regulatory	12.426804336
ef_regulation_business_start	6.366492989
ef_legal_courts	6.042731840
ef_trade_regulatory_compliance	5.651783473
ef_trade_movement_capital	5.233852402
ef_legal_military	4.002893049
ef_legal_gender	3.728645811
pf_ss_disappearances	2.784232921
ef_regulation_labor	2.487345645

- ef_money_currency and ef_trade_regulatory are the 2 most important variables in predicting EFI.

VARIABLE IMPORTANCE

For PFI,

var	rel.inf
pf_expression_control	29.39756043
pf_ss_women	20.31574718
pf_ss_disappearances	8.76300093
pf_expression_influence	6.03968961
pf_ss_homicide	5.59731808
pf_movement_domestic	3.29614093
ef_legal_gender	2.97767054
pf_identity_parental	2.92823482
pf_movement_foreign	2.91018352
pf_identity_sex_male	2.49081807

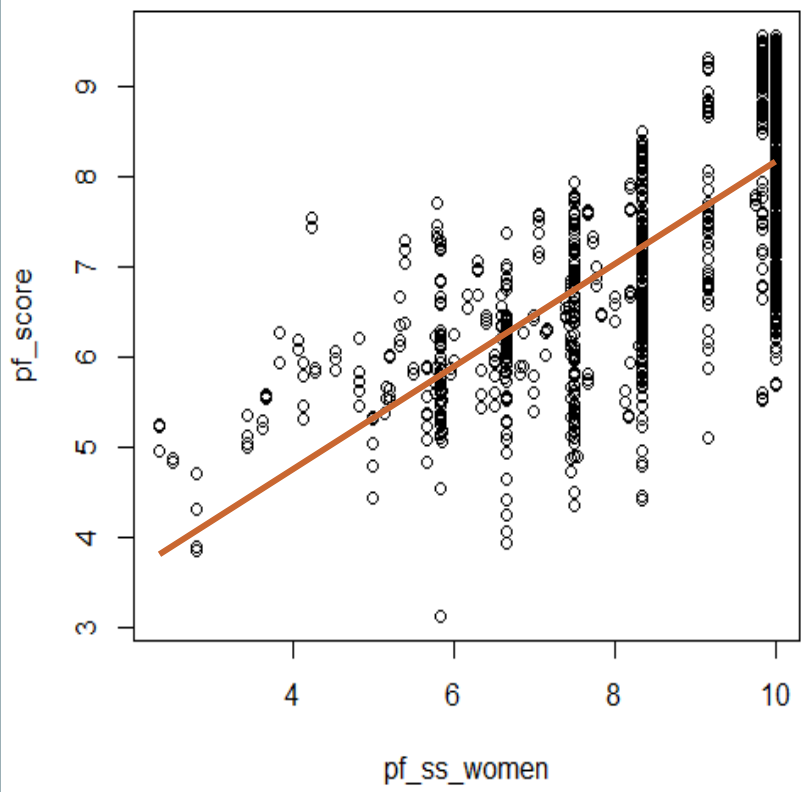
Top 10 Predictors

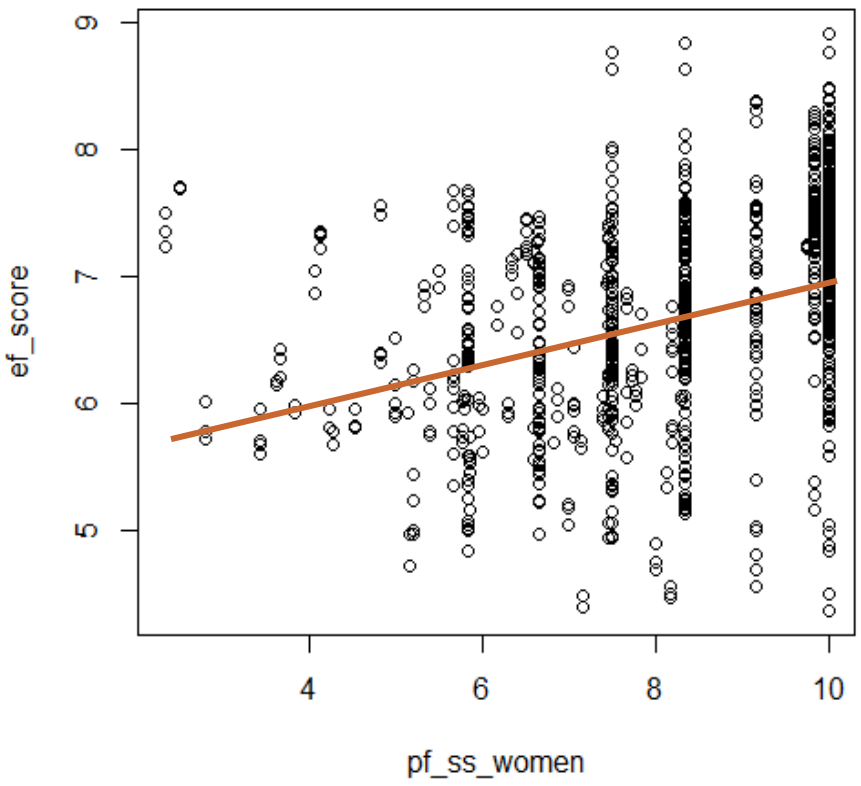
For EFI,

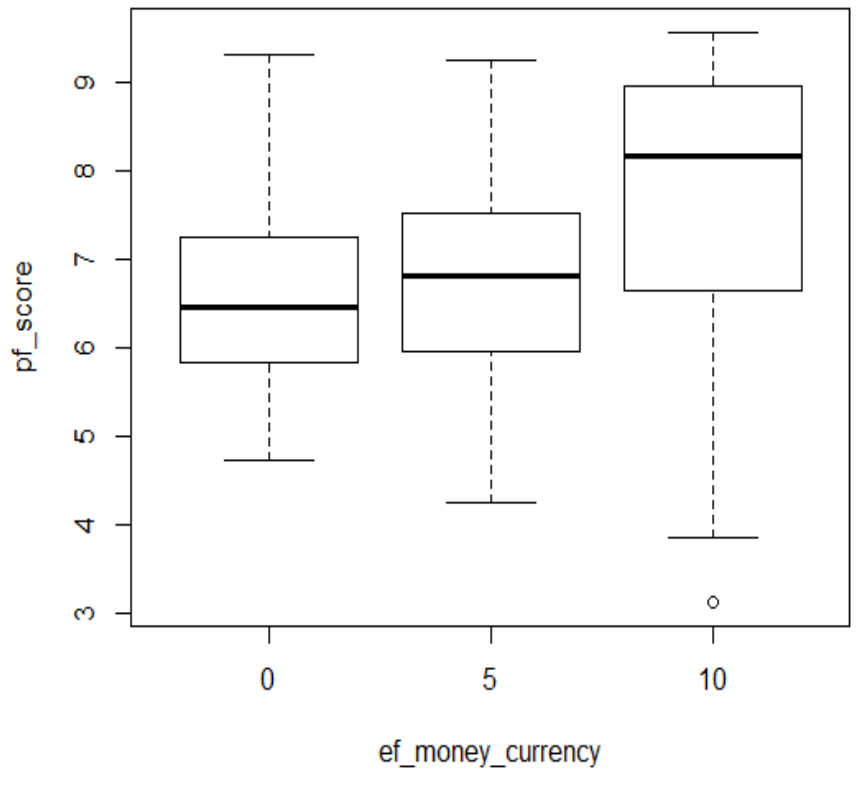
var	rel.inf
ef_money_currency	27.978704187
ef_trade_regulatory	12.426804336
ef_regulation_business_start	6.366492989
ef_legal_courts	6.042731840
ef_trade_regulatory_compliance	5.651783473
ef_trade_movement_capital	5.233852402
ef_legal_military	4.002893049
ef_legal_gender	3.728645811
pf_ss_disappearances	2.784232921
ef_regulation_labor	2.487345645

Top 10 Predictors

- Most of the predictors only greatly affect one of the response variables and not both.

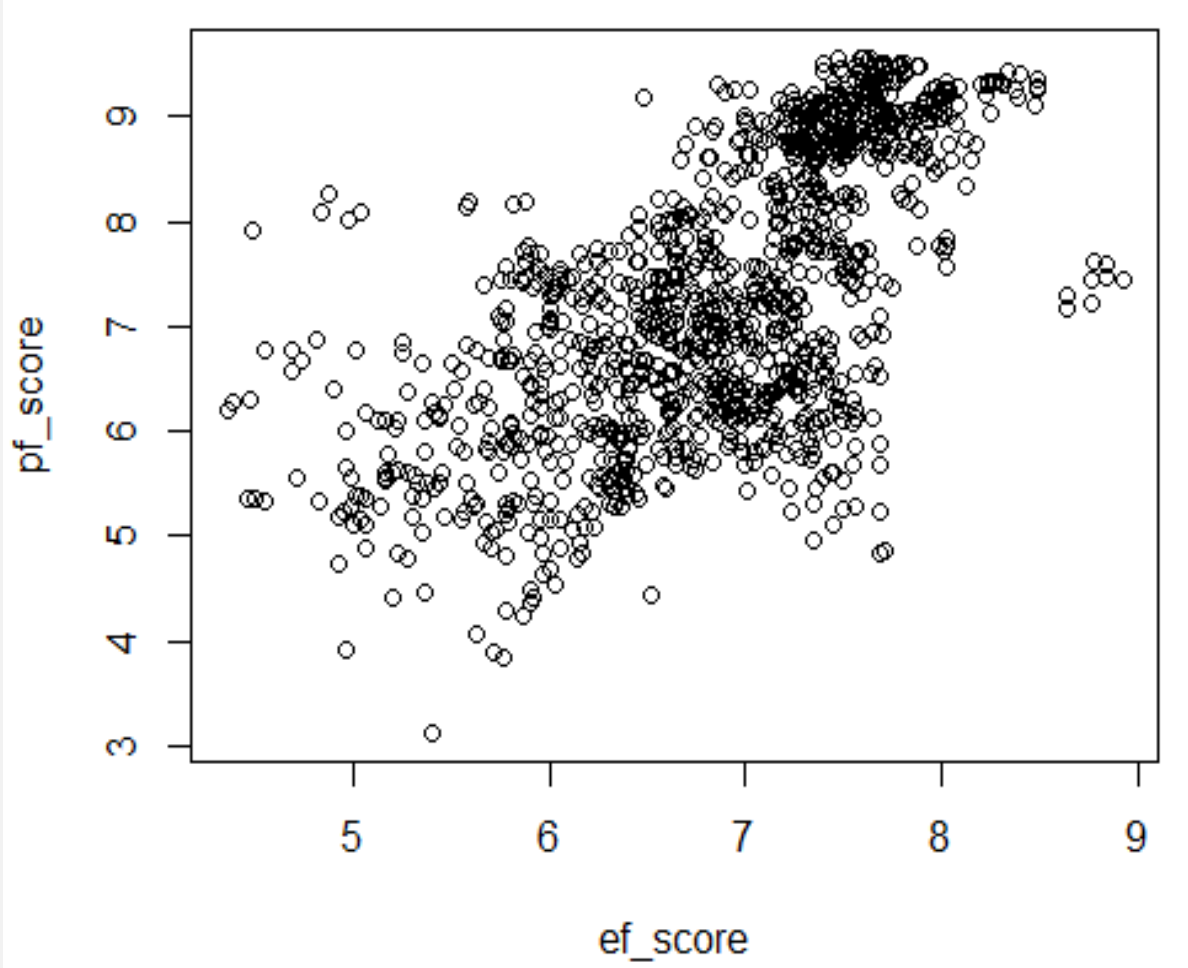




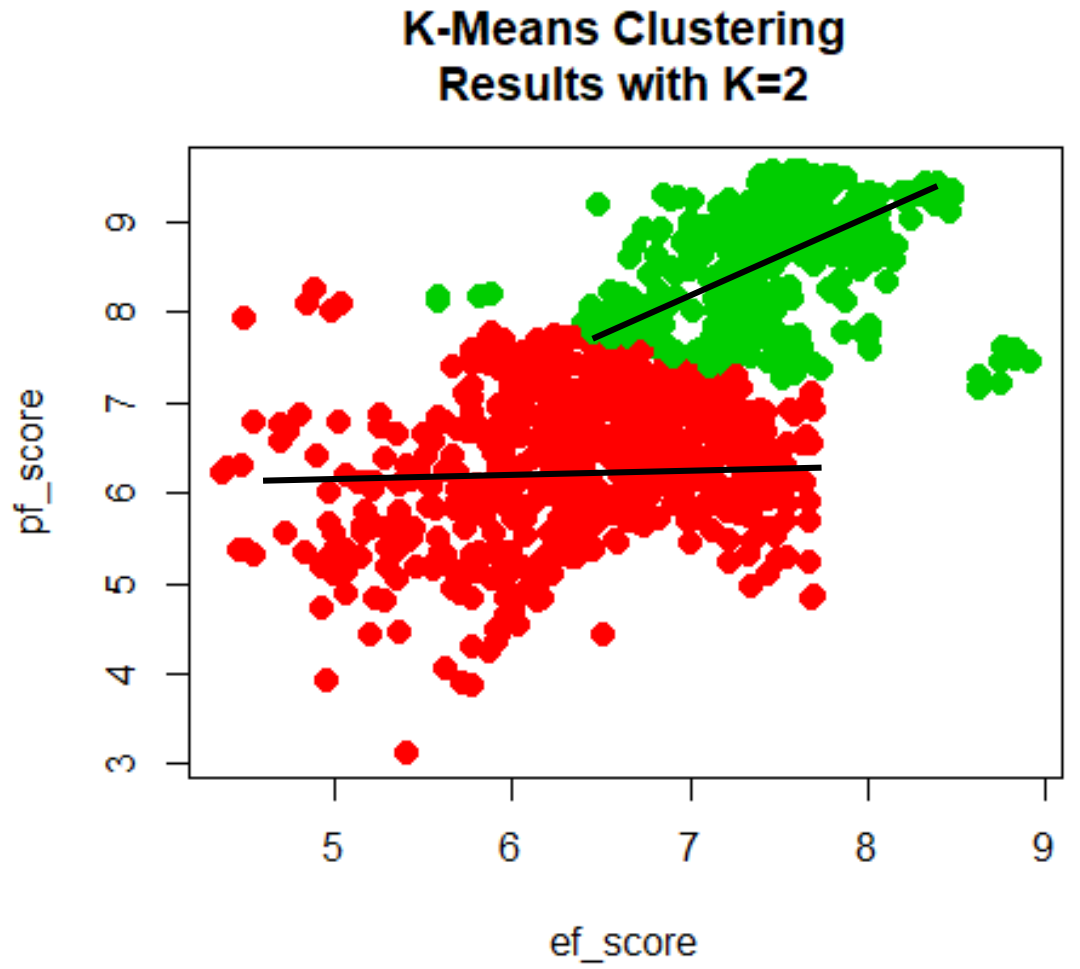


NO SINGLE BEST VARIABLE TO SOLVE THE DIFFERENT ISSUES PERTAINING TO FREEDOM.

RELATIONSHIP BETWEEN PFI AND EFI



RELATIONSHIP BETWEEN PFI AND EFI



	# of observations	Mean (PFI)	Mean (EFI)
Red cluster	689	6.345292	6.426226
Green cluster	495	8.611209	7.465899
Total	1184	7.292614	6.860887

- Relatively positive relationship between the 2 responses in the green cluster, unlike in red cluster.
- Green cluster generally has higher values and mean than those in red clusters.

CONCLUSION

- Correlation does not imply causation.
- PFI and EFI does improve or hurt one another, depending on the state of a country.
- There is no such thing as absolute freedom, especially within large political jurisdictions such as cities and countries.

THANK YOU!