Data_606_Lab_8_Introduction_to_Linear_Regression

Enid Roman

2022-11-13

Getting Started

Load packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(openintro)
```

```
## Loading required package: airports
## Loading required package: cherryblossom
## Loading required package: usdata
```

```
data('hfi', package='openintro')
```

The data

```
glimpse(hfi)
```

```
## Rows: 1,458
## Columns: 123
## $ year                       <dbl> 2016, 2016, 2016, 2016, 2016, 2016,~
## $ ISO_code                   <chr> "ALB", "DZA", "AGO", "ARG", "ARM", ~
## $ countries                  <chr> "Albania", "Algeria", "Angola", "Ar~
## $ region                     <chr> "Eastern Europe", "Middle East & No~
## $ pf_rol_procedural          <dbl> 6.661503, NA, NA, 7.098483, NA, 8.4~
```

```
## $ pf_rol_civil                        <dbl> 4.547244, NA, NA, 5.791960, NA, 7.5~
## $ pf_rol_criminal                     <dbl> 4.666508, NA, NA, 4.343930, NA, 7.3~
## $ pf_rol                              <dbl> 5.291752, 3.819566, 3.451814, 5.744~
## $ pf_ss_homicide                      <dbl> 8.920429, 9.456254, 8.060260, 7.622~
## $ pf_ss_disappearances_disap          <dbl> 10, 10, 5, 10, 10, 10, 10, 10, 10, ~
## $ pf_ss_disappearances_violent        <dbl> 10.000000, 9.294030, 10.000000, 10.~
## $ pf_ss_disappearances_organized      <dbl> 10.0, 5.0, 7.5, 7.5, 7.5, 10.0, 10.~
## $ pf_ss_disappearances_fatalities     <dbl> 10.000000, 9.926119, 10.000000, 10.~
## $ pf_ss_disappearances_injuries       <dbl> 10.000000, 9.990149, 10.000000, 9.9~
## $ pf_ss_disappearances                <dbl> 10.000000, 8.842060, 8.500000, 9.49~
## $ pf_ss_women_fgm                     <dbl> 10.0, 10.0, 10.0, 10.0, 10.0, 10.0,~
## $ pf_ss_women_missing                 <dbl> 7.5, 7.5, 10.0, 10.0, 5.0, 10.0, 10~
## $ pf_ss_women_inheritance_widows      <dbl> 5, 0, 5, 10, 10, 10, 10, 5, NA, 0, ~
## $ pf_ss_women_inheritance_daughters   <dbl> 5, 0, 5, 10, 10, 10, 10, 10, NA, 0,~
## $ pf_ss_women_inheritance             <dbl> 5.0, 0.0, 5.0, 10.0, 10.0, 10.0, 10~
## $ pf_ss_women                         <dbl> 7.500000, 5.833333, 8.333333, 10.00~
## $ pf_ss                               <dbl> 8.806810, 8.043882, 8.297865, 9.040~
## $ pf_movement_domestic                <dbl> 5, 5, 0, 10, 5, 10, 10, 5, 10, 10, ~
## $ pf_movement_foreign                 <dbl> 10, 5, 5, 10, 5, 10, 10, 5, 10, 5, ~
## $ pf_movement_women                   <dbl> 5, 5, 10, 10, 10, 10, 10, 5, NA, 5,~
## $ pf_movement                         <dbl> 6.666667, 5.000000, 5.000000, 10.00~
## $ pf_religion_estop_establish         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_religion_estop_operate           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_religion_estop                   <dbl> 10.0, 5.0, 10.0, 7.5, 5.0, 10.0, 10~
## $ pf_religion_harassment              <dbl> 9.566667, 6.873333, 8.904444, 9.037~
## $ pf_religion_restrictions            <dbl> 8.011111, 2.961111, 7.455556, 6.850~
## $ pf_religion                         <dbl> 9.192593, 4.944815, 8.786667, 7.795~
## $ pf_association_association          <dbl> 10.0, 5.0, 2.5, 7.5, 7.5, 10.0, 10.~
## $ pf_association_assembly             <dbl> 10.0, 5.0, 2.5, 10.0, 7.5, 10.0, 10~
## $ pf_association_political_establish  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_political_operate    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_political            <dbl> 10.0, 5.0, 2.5, 5.0, 5.0, 10.0, 10.~
## $ pf_association_prof_establish       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_prof_operate         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_prof                 <dbl> 10.0, 5.0, 5.0, 7.5, 5.0, 10.0, 10.~
## $ pf_association_sport_establish      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_sport_operate        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ pf_association_sport                <dbl> 10.0, 5.0, 7.5, 7.5, 7.5, 10.0, 10.~
## $ pf_association                      <dbl> 10.0, 5.0, 4.0, 7.5, 6.5, 10.0, 10.~
## $ pf_expression_killed                <dbl> 10.000000, 10.000000, 10.000000, 10~
## $ pf_expression_jailed                <dbl> 10.000000, 10.000000, 10.000000, 10~
## $ pf_expression_influence             <dbl> 5.0000000, 2.6666667, 2.6666667, 5.~
## $ pf_expression_control               <dbl> 5.25, 4.00, 2.50, 5.50, 4.25, 7.75,~
## $ pf_expression_cable                 <dbl> 10.0, 10.0, 7.5, 10.0, 7.5, 10.0, 1~
## $ pf_expression_newspapers            <dbl> 10.0, 7.5, 5.0, 10.0, 7.5, 10.0, 10~
## $ pf_expression_internet              <dbl> 10.0, 7.5, 7.5, 10.0, 7.5, 10.0, 10~
## $ pf_expression                       <dbl> 8.607143, 7.380952, 6.452381, 8.738~
## $ pf_identity_legal                   <dbl> 0, NA, 10, 10, 7, 7, 10, 0, NA, NA,~
## $ pf_identity_parental_marriage       <dbl> 10, 0, 10, 10, 10, 10, 10, 10, 10, ~
## $ pf_identity_parental_divorce        <dbl> 10, 5, 10, 10, 10, 10, 10, 10, 10, ~
## $ pf_identity_parental                <dbl> 10.0, 2.5, 10.0, 10.0, 10.0, 10.0, ~
## $ pf_identity_sex_male                <dbl> 10, 0, 0, 10, 10, 10, 10, 10, 10, 1~
## $ pf_identity_sex_female              <dbl> 10, 0, 0, 10, 10, 10, 10, 10, 10, 1~
## $ pf_identity_sex                     <dbl> 10, 0, 0, 10, 10, 10, 10, 10, 10, 1~
```

```
## $ pf_identity_divorce             <dbl> 5, 0, 10, 10, 5, 10, 10, 5, NA, 0, ~
## $ pf_identity                     <dbl> 6.2500000, 0.8333333, 7.5000000, 10~
## $ pf_score                        <dbl> 7.596281, 5.281772, 6.111324, 8.099~
## $ pf_rank                         <dbl> 57, 147, 117, 42, 84, 11, 8, 131, 6~
## $ ef_government_consumption       <dbl> 8.232353, 2.150000, 7.600000, 5.335~
## $ ef_government_transfers         <dbl> 7.509902, 7.817129, 8.886739, 6.048~
## $ ef_government_enterprises       <dbl> 8, 0, 0, 6, 8, 10, 10, 0, 7, 10, 7,~
## $ ef_government_tax_income        <dbl> 9, 7, 10, 7, 5, 5, 4, 9, 10, 10, 8,~
## $ ef_government_tax_payroll       <dbl> 7, 2, 9, 1, 5, 5, 3, 4, 10, 10, 8, ~
## $ ef_government_tax               <dbl> 8.0, 4.5, 9.5, 4.0, 5.0, 5.0, 3.5, ~
## $ ef_government                   <dbl> 7.935564, 3.616782, 6.496685, 5.346~
## $ ef_legal_judicial               <dbl> 2.6682218, 4.1867042, 1.8431292, 3.~
## $ ef_legal_courts                 <dbl> 3.145462, 4.327113, 1.974566, 2.930~
## $ ef_legal_protection             <dbl> 4.512228, 4.689952, 2.512364, 4.255~
## $ ef_legal_military               <dbl> 8.333333, 4.166667, 3.333333, 7.500~
## $ ef_legal_integrity              <dbl> 4.166667, 5.000000, 4.166667, 3.333~
## $ ef_legal_enforcement            <dbl> 4.3874441, 4.5075380, 2.3022004, 3.~
## $ ef_legal_restrictions           <dbl> 6.485287, 6.626692, 5.455882, 6.857~
## $ ef_legal_police                 <dbl> 6.933500, 6.136845, 3.016104, 3.385~
## $ ef_legal_crime                  <dbl> 6.215401, 6.737383, 4.291197, 4.133~
## $ ef_legal_gender                 <dbl> 0.9487179, 0.8205128, 0.8461538, 0.~
## $ ef_legal                        <dbl> 5.071814, 4.690743, 2.963635, 3.904~
## $ ef_money_growth                 <dbl> 8.986454, 6.955962, 9.385679, 5.233~
## $ ef_money_sd                     <dbl> 9.484575, 8.339152, 4.986742, 5.224~
## $ ef_money_inflation              <dbl> 9.743600, 8.720460, 3.054000, 2.000~
## $ ef_money_currency               <dbl> 10, 5, 5, 10, 10, 10, 10, 5, 0, 10,~
## $ ef_money                        <dbl> 9.553657, 7.253894, 5.606605, 5.614~
## $ ef_trade_tariffs_revenue        <dbl> 9.626667, 8.480000, 8.993333, 6.060~
## $ ef_trade_tariffs_mean           <dbl> 9.24, 6.22, 7.72, 7.26, 8.76, 9.50,~
## $ ef_trade_tariffs_sd             <dbl> 8.0240, 5.9176, 4.2544, 5.9448, 8.0~
## $ ef_trade_tariffs                <dbl> 8.963556, 6.872533, 6.989244, 6.421~
## $ ef_trade_regulatory_nontariff   <dbl> 5.574481, 4.962589, 3.132738, 4.466~
## $ ef_trade_regulatory_compliance  <dbl> 9.4053278, 0.0000000, 0.9171598, 5.~
## $ ef_trade_regulatory             <dbl> 7.489905, 2.481294, 2.024949, 4.811~
## $ ef_trade_black                  <dbl> 10.00000, 5.56391, 10.00000, 0.0000~
## $ ef_trade_movement_foreign       <dbl> 6.306106, 3.664829, 2.946919, 5.358~
## $ ef_trade_movement_capital       <dbl> 4.6153846, 0.0000000, 3.0769231, 0.~
## $ ef_trade_movement_visit         <dbl> 8.2969231, 1.1062564, 0.1106256, 7.~
## $ ef_trade_movement               <dbl> 6.406138, 1.590362, 2.044823, 4.697~
## $ ef_trade                        <dbl> 8.214900, 4.127025, 5.264754, 3.982~
## $ ef_regulation_credit_ownership  <dbl> 5, 0, 8, 5, 10, 10, 8, 5, 10, 10, 5~
## $ ef_regulation_credit_private    <dbl> 7.295687, 5.301526, 9.194715, 4.259~
## $ ef_regulation_credit_interest   <dbl> 9, 10, 4, 7, 10, 10, 10, 9, 10, 10,~
## $ ef_regulation_credit            <dbl> 7.098562, 5.100509, 7.064905, 5.419~
## $ ef_regulation_labor_minwage     <dbl> 5.566667, 5.566667, 8.900000, 2.766~
## $ ef_regulation_labor_firing      <dbl> 5.396399, 3.896912, 2.656198, 2.191~
## $ ef_regulation_labor_bargain     <dbl> 6.234861, 5.958321, 5.172987, 3.432~
## $ ef_regulation_labor_hours       <dbl> 8, 6, 4, 10, 10, 10, 6, 6, 8, 8, 10~
## $ ef_regulation_labor_dismissal   <dbl> 6.299741, 7.755176, 6.632764, 2.517~
## $ ef_regulation_labor_conscription <dbl> 10, 1, 0, 10, 0, 10, 3, 1, 10, 10, ~
## $ ef_regulation_labor             <dbl> 6.916278, 5.029513, 4.560325, 5.151~
## $ ef_regulation_business_adm      <dbl> 6.072172, 3.722341, 2.758428, 2.404~
## $ ef_regulation_business_bureaucracy <dbl> 6.000000, 1.777778, 1.333333, 6.666~
## $ ef_regulation_business_start    <dbl> 9.713864, 9.243070, 8.664627, 9.122~
```

```
## $ ef_regulation_business_bribes      <dbl> 4.050196, 3.765515, 1.945540, 3.260~
## $ ef_regulation_business_licensing   <dbl> 7.324582, 8.523503, 8.096776, 5.253~
## $ ef_regulation_business_compliance  <dbl> 7.074366, 7.029528, 6.782923, 6.508~
## $ ef_regulation_business             <dbl> 6.705863, 5.676956, 4.930271, 5.535~
## $ ef_regulation                      <dbl> 6.906901, 5.268992, 5.518500, 5.369~
## $ ef_score                           <dbl> 7.54, 4.99, 5.17, 4.84, 7.57, 7.98,~
## $ ef_rank                            <dbl> 34, 159, 155, 160, 29, 10, 27, 106,~
## $ hf_score                           <dbl> 7.568140, 5.135886, 5.640662, 6.469~
## $ hf_rank                            <dbl> 48, 155, 142, 107, 57, 4, 16, 130, ~
## $ hf_quartile                        <dbl> 2, 4, 4, 3, 2, 1, 1, 4, 2, 2, 4, 2,~
```

Exercise 1

```
dim(hfi)
```

Dimension of the dataset is 1458 observations and 123 columns.

```
## [1] 1458   123
```

Exercise 2

I would use a scatterplot as its primary use is to observe and show relationships between 2 numeric variables. The relationship seems somewhat linear, but many of the points are scattered randomly.

The relationship seems linear, but some of the points are scattered randomly.

```
ggplot(data = hfi, aes(y = pf_score, x = pf_expression_control)) +
  geom_point()
```

Yes I would be comfortable using a linear model to predict the personal freedom score if I knew a country's pf_expression_control, or its score out of 10, with 0 being the most, of political pressures and controls on media content.

```
## Warning: Removed 80 rows containing missing values (geom_point).
```
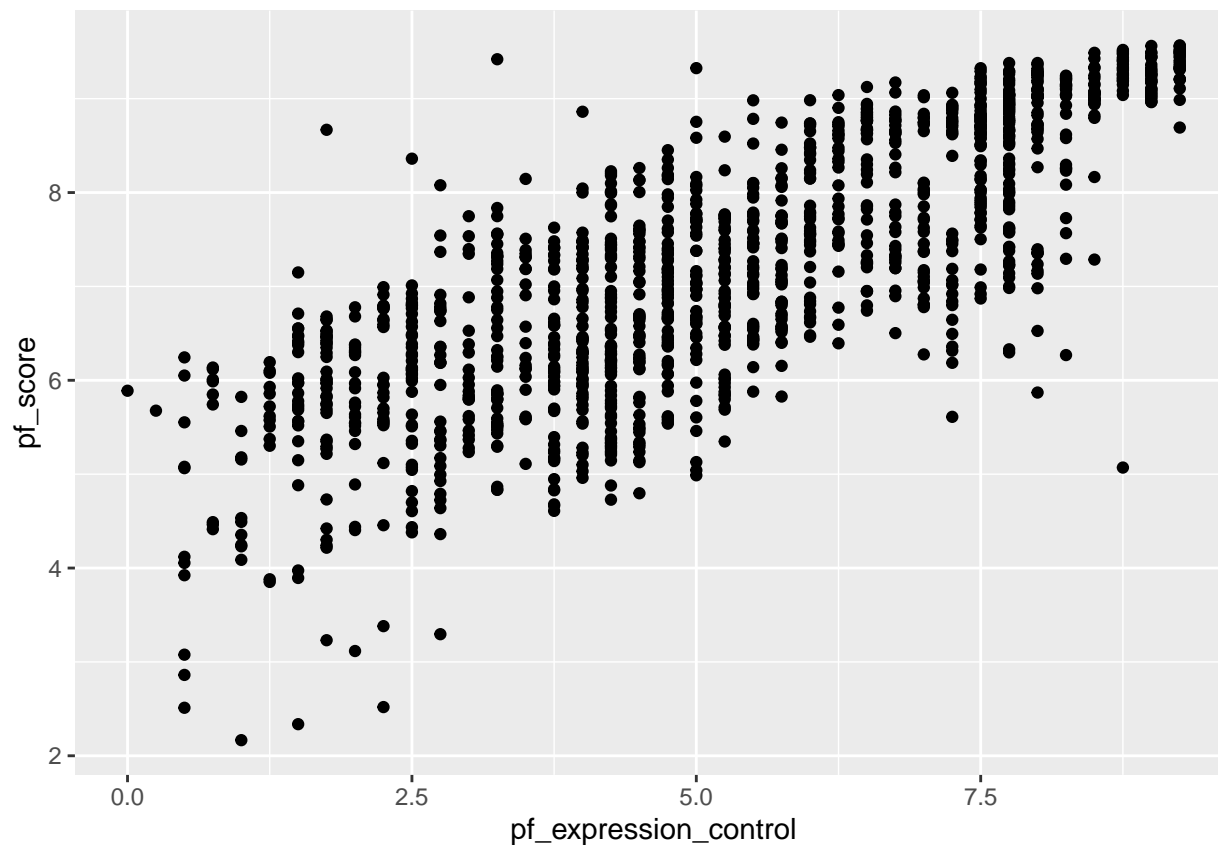
```
hfi %>%
  summarise(cor(pf_expression_control, pf_score, use = "complete.obs"))
```

```
## # A tibble: 1 x 1
##   `cor(pf_expression_control, pf_score, use = "complete.obs")`
##                                                         <dbl>
## 1                                                       0.796
```
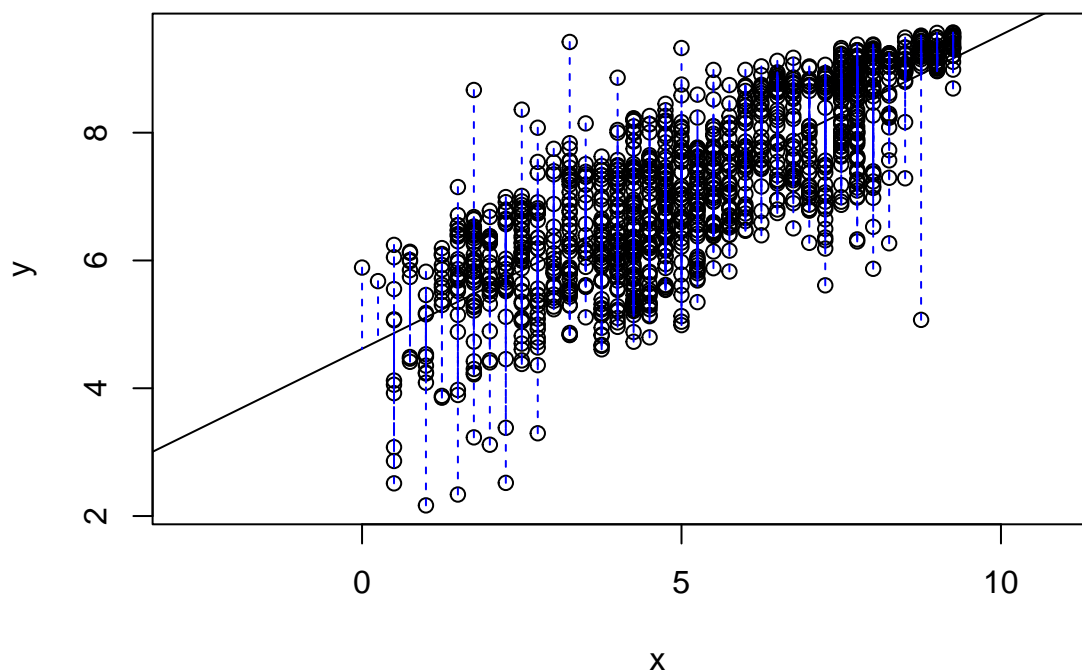
Sum of squared residuals

Exercise 3

```
# subset hfi

hfi1 <- hfi[c("pf_score", "pf_expression_control")]

# Drop NAs

hfi1 <- drop_na(hfi1)
row.names <- NULL

# This will only work interactively (i.e. will not show in the knitted document)
DATA606::plot_ss(x = hfi1$pf_expression_control, y = hfi1$pf_score)
```

The relationship between these two variables is the result of a positive correlation given the upward trend in pf_score as pf_expression_score increases. This positive correlation indicates that it is a strong correlation, since is a positively sloped line. This is confirmed by the ~ 0.8 cor-

relation coefficient value.

```
## Click two points to make a line.
## Call:
## lm(formula = y ~ x, data = pts)
##
## Coefficients:
## (Intercept)            x
##      4.6171       0.4914
##
## Sum of Squares:  952.153
```

DATA606::plot_ss(x = hfi1$pf_expression_control, y = hfi1$pf_score, showSquares = TRUE)
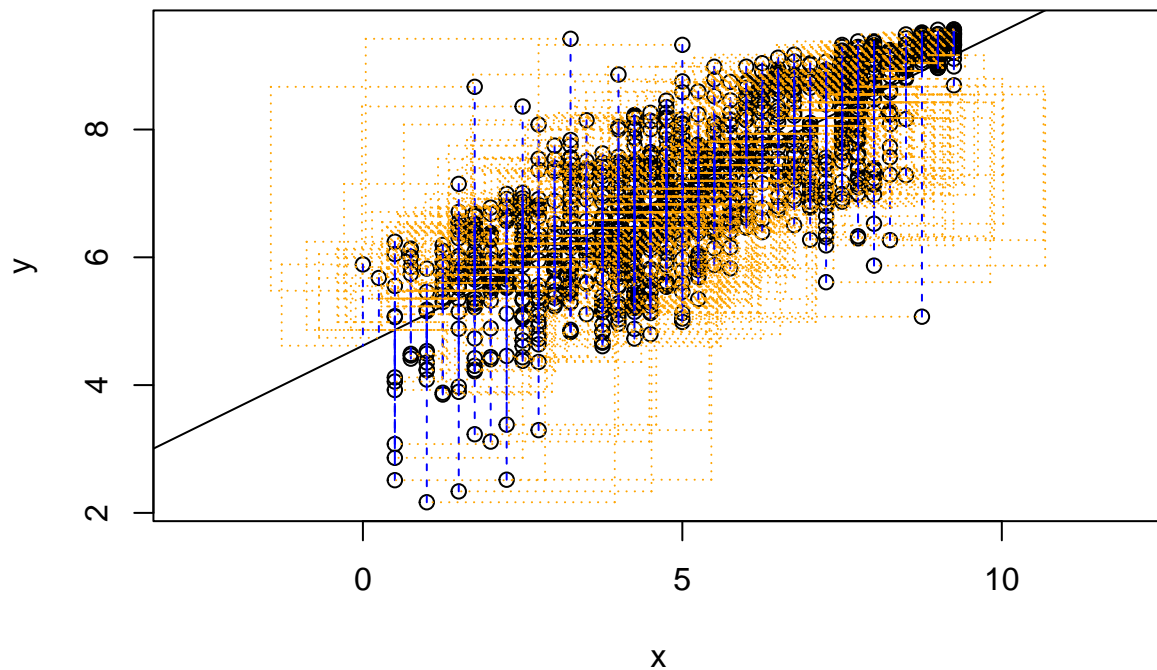
```
## Click two points to make a line.
## Call:
## lm(formula = y ~ x, data = pts)
##
## Coefficients:
## (Intercept)              x
##      4.6171         0.4914
##
## Sum of Squares:  952.153
```

Exercise 4

Not sure if this is working properly. I got 149642. The smallest sum of squared residuals I was able to achieve was **952,153**.

The linear model

```
m1 <- lm(pf_score ~ pf_expression_control, data = hfi)
```

```
summary(m1)
```

```
##
```

```
## Call:
## lm(formula = pf_score ~ pf_expression_control, data = hfi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8467 -0.5704  0.1452  0.6066  3.2060
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.61707    0.05745   80.36   <2e-16 ***
## pf_expression_control  0.49143    0.01006   48.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8318 on 1376 degrees of freedom
##   (80 observations deleted due to missingness)
## Multiple R-squared:  0.6342, Adjusted R-squared:  0.634
## F-statistic:  2386 on 1 and 1376 DF,  p-value: < 2.2e-16
```

Exercise 5

Using the estimates from the R output, the equation of the regression line is

hf_score = 5.153687 - 0.349862 * pf_expression_control

The slope tell us in the context of the relationship between human freedom and the amount of political pressure on media content is that for each additional amount of political pressure on media content score, we would expect the human freedom score to increase by 0.3499.

```
lm5 <- lm(hf_score ~ pf_expression_control, data = hfi)
summary(lm5)
```

The intercept tell us the human freedom score with no amount of political pressure on media is 5.1537.

```
##
## Call:
## lm(formula = hf_score ~ pf_expression_control, data = hfi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6198 -0.4908  0.1031  0.4703  2.2933
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.153687   0.046070  111.87   <2e-16 ***
## pf_expression_control 0.349862   0.008067   43.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.667 on 1376 degrees of freedom
##   (80 observations deleted due to missingness)
## Multiple R-squared:  0.5775, Adjusted R-squared:  0.5772
## F-statistic:  1881 on 1 and 1376 DF,  p-value: < 2.2e-16
```
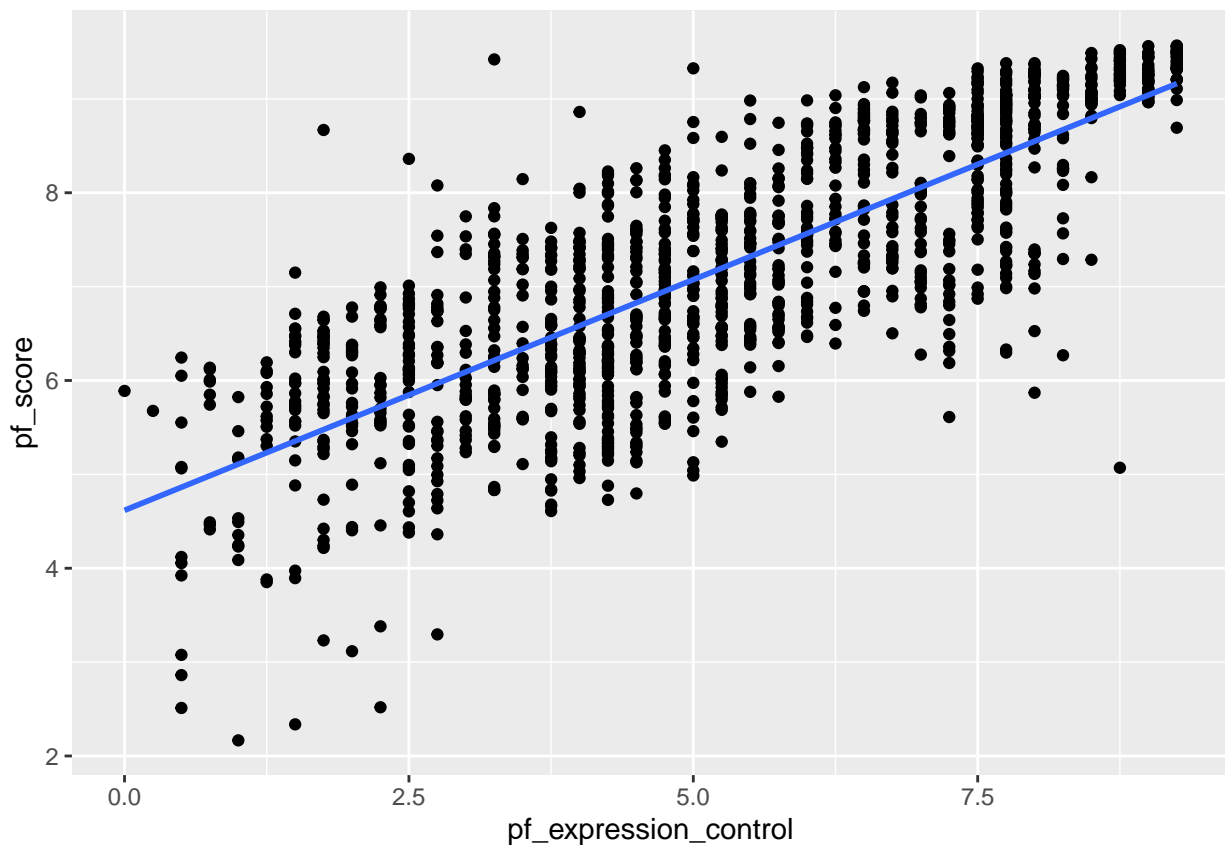
Prediction and prediction errors

```
ggplot(data = hfi, aes(x = pf_expression_control, y = pf_score)) +
  geom_point() +
  stat_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 80 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 80 rows containing missing values (geom_point).
```


Exercise 6

```
pf_exp_control <- 6.7

pf_score6 <- 4.61707 + 0.49143 * pf_exp_control
pf_score6
```

If someone saw the least squares regression line and not the actual data they would predict a country's personal freedom school for one with a 6.7 rating for pf_expression_control is 7.9. This an overestimate by -.48. The residual for this prediction -.48.

```
## [1] 7.909651
```

```
# Check oberved values of pf_score with 6.7 rating
# for `pf_expression_control`

hfi %>%
  group_by(pf_score) %>%
  filter(pf_expression_control == 6.7)
```

```
## # A tibble: 0 x 123
## # Groups:   pf_score [0]
## # ... with 123 variables: year <dbl>, ISO_code <chr>, countries <chr>,
## #   region <chr>, pf_rol_procedural <dbl>, pf_rol_civil <dbl>,
## #   pf_rol_criminal <dbl>, pf_rol <dbl>, pf_ss_homicide <dbl>,
## #   pf_ss_disappearances_disap <dbl>, pf_ss_disappearances_violent <dbl>,
## #   pf_ss_disappearances_organized <dbl>,
## #   pf_ss_disappearances_fatalities <dbl>, pf_ss_disappearances_injuries <dbl>,
## #   pf_ss_disappearances <dbl>, pf_ss_women_fgm <dbl>, ...
```
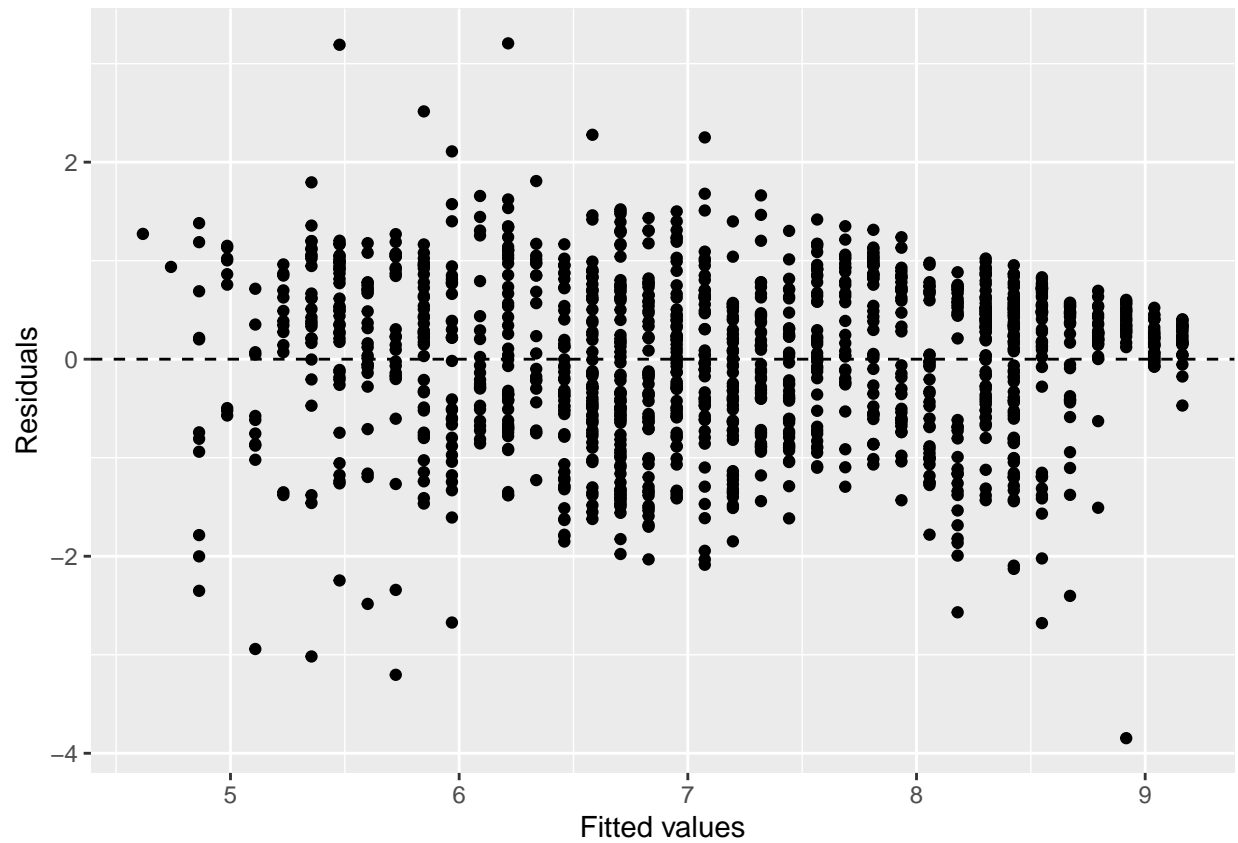
```
# There is none observed value of pf_score with 6.7 rating for pf_expression_score
# I would consider the closest one, case of Belize with
# observed pf_score of 7.43 for 6.75 rating for pf_expression_score

residus <- 7.43 - 7.91
residus
```
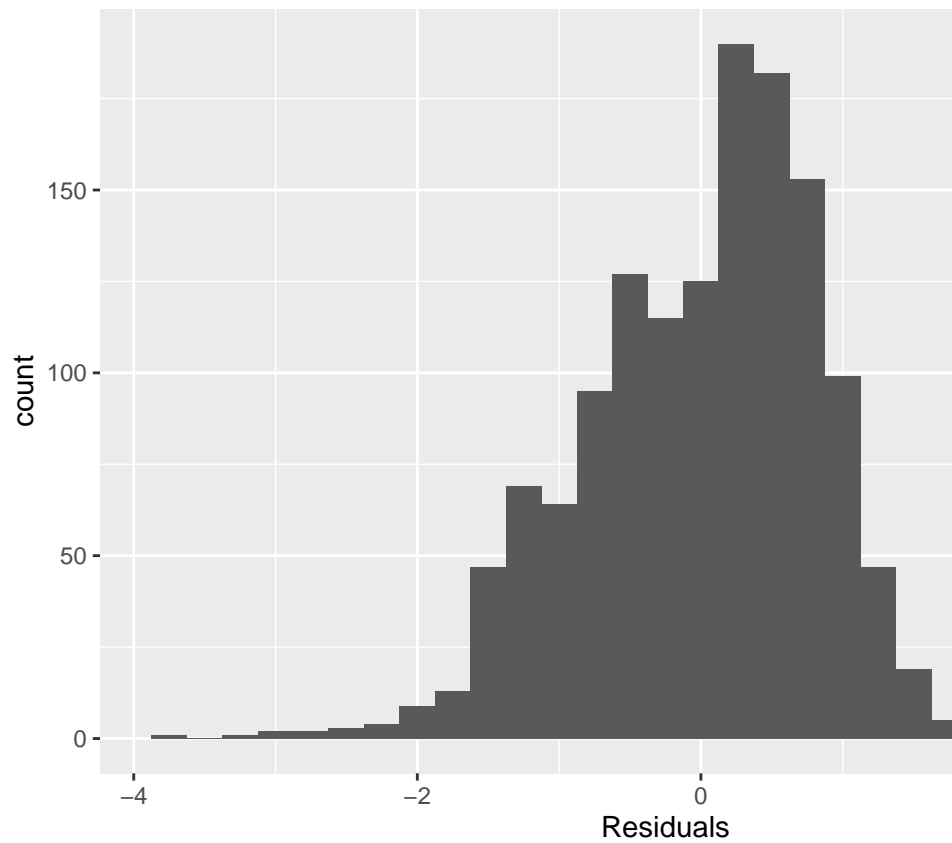
```
## [1] -0.48
```

Model diagnostics

```
ggplot(data = m1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```
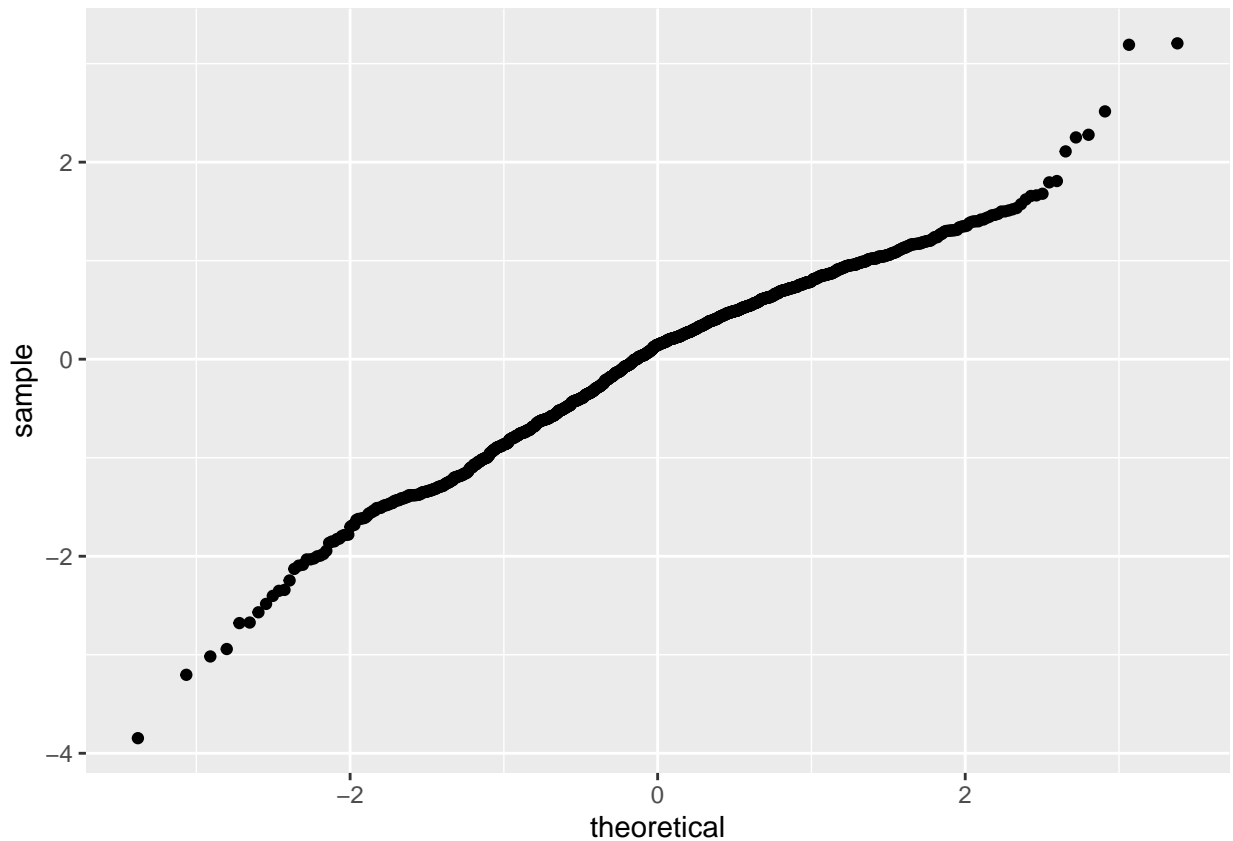
Exercise 7

```
ggplot(data = m1, aes(x = .resid)) +
  geom_histogram(binwidth = .25) +
  xlab("Residuals")
```

There is not an apparent pattern in the residuals plot and this indicates there is a linear rela-

tionship between the two variables.

```
ggplot(data = m1, aes(sample = .resid)) +
  stat_qq()
```

Exercise 8

Both the histogram and the normal probability plot show that the distribution of these data are nearly normal. The nearly normal residuals condition appear to be met.

Exercise 9

The points residuals vs. fitted plot show that points are scattered around 0, there is a constant variability. The constant variability condition appear to be met.

More Practice

```
ggplot(data = hfi, aes(x = pf_rol_procedural, y = pf_ss))+
  geom_point() +
  stat_smooth(method = "lm", se = FALSE)
```

1. I chose to test how the personal sense of security and safety (pf_ss) correlates with the role the procedural justice a given country (pf_rol_procedural). Yes this does seem to be a linear relationship.

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 578 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 578 rows containing missing values (geom_point).
```
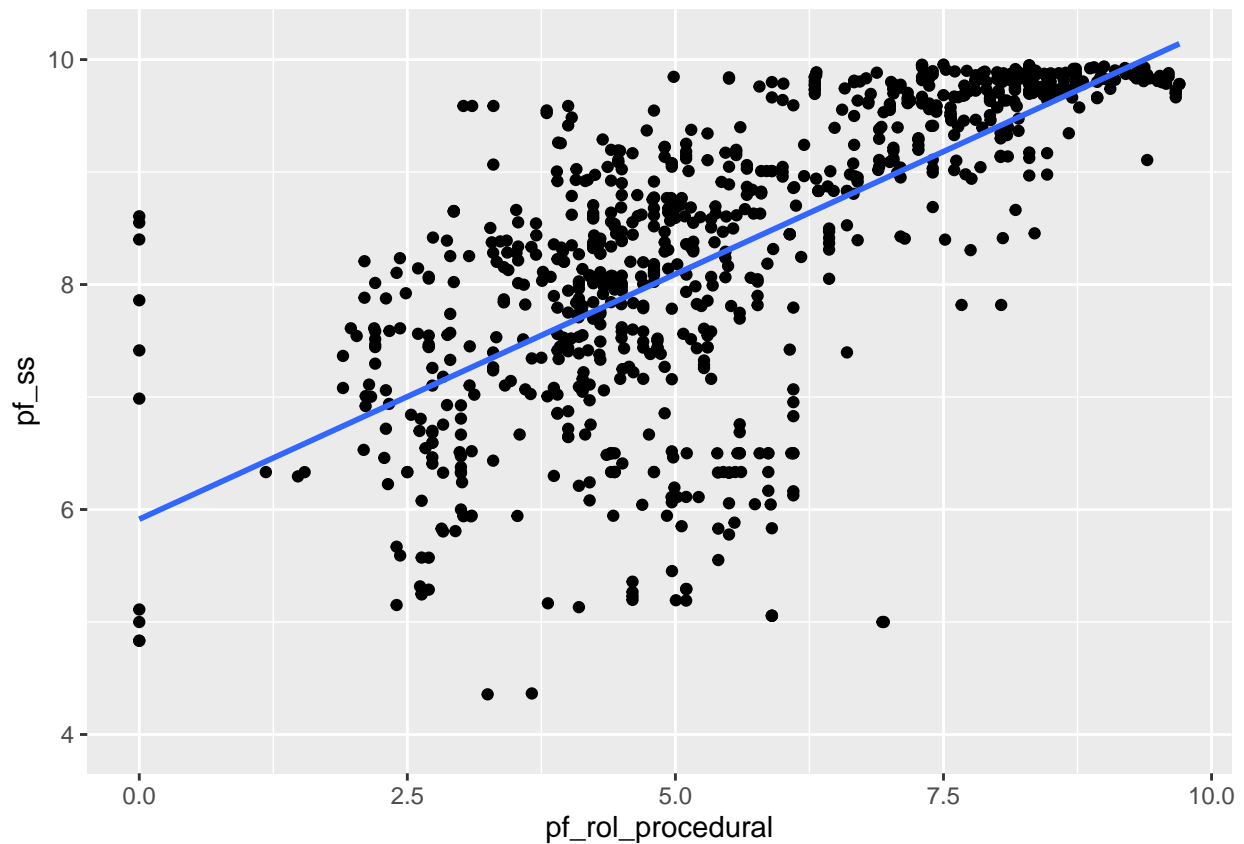


```
lin_model <- lm(data = hfi, formula = pf_ss ~ pf_rol_procedural)
summary(lin_model)
```

2. According to the output, the R2 value in this case is 0.4781, meaning that pf_rol_procedural accounts for approximately 47.81% of the variance seen in pf_ss. This is slightly worse than the model we saw that correlated pf_score and pf_expression_control, but still pretty good!

```
##
## Call:
## lm(formula = pf_ss ~ pf_rol_procedural, data = hfi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9398 -0.3569  0.1601  0.5722  2.6927
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.91391    0.09164   64.53   <2e-16 ***
```

```
## pf_rol_procedural  0.43577    0.01537    28.36    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.948 on 878 degrees of freedom
##    (578 observations deleted due to missingness)
## Multiple R-squared:  0.4781, Adjusted R-squared:  0.4775
## F-statistic: 804.3 on 1 and 878 DF,  p-value: < 2.2e-16
```

3. The relationship between one's feeling of safety and security (pf_ss) as a function of the size of the government they live in (ef_government).

```
ggplot(data = hfi, aes(x = ef_government, y = pf_ss))+
  geom_point() +
  stat_smooth(method = "lm", se = FALSE)
```

There seem to be a negative correlation between these two variables. It seems that countries with larger governments appear to have citizens that feel a lesser sense of safety and security.

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 80 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 80 rows containing missing values (geom_point).
```
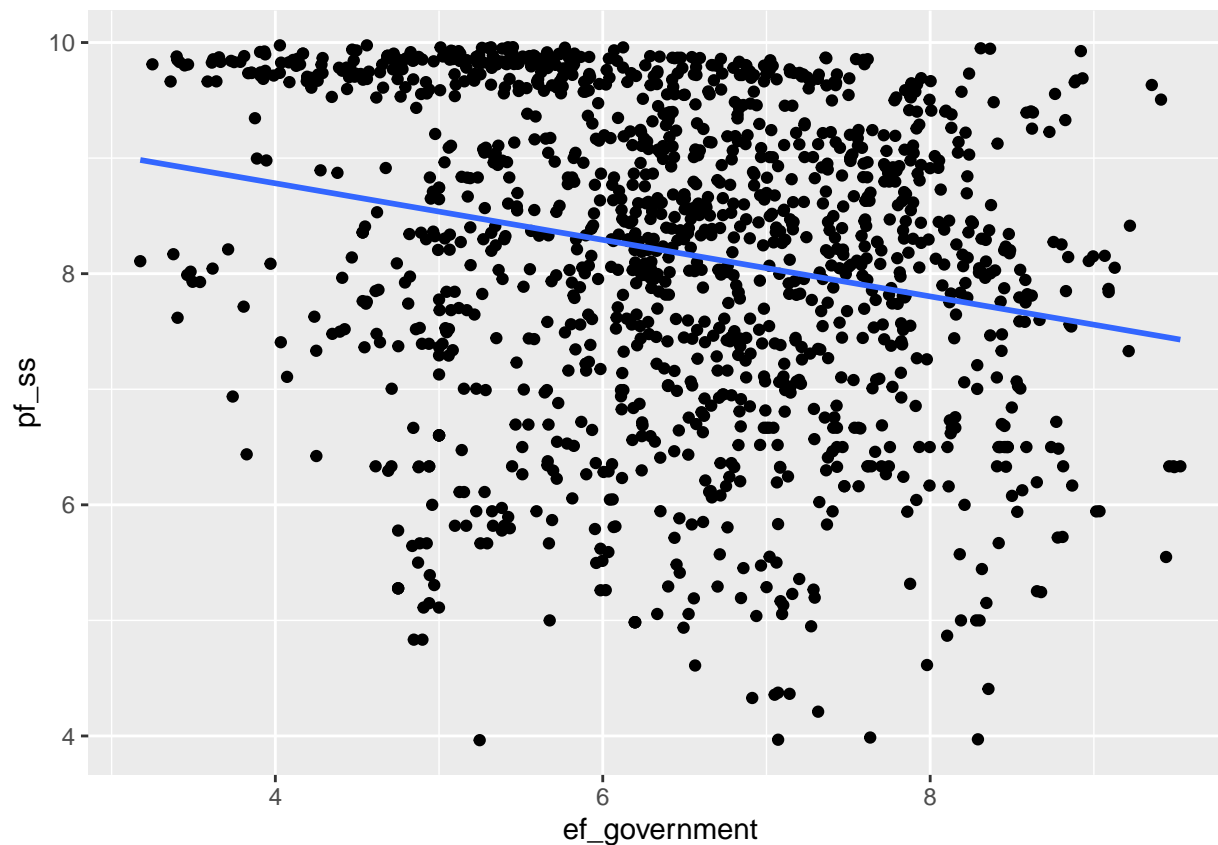
This correlation is confirmed by analyzing the results of the linear fit.

```
lin_model <- lm(data = hfi, formula = pf_ss ~ ef_government)
summary(lin_model)
```

Question to ask is there higher incidence of corruption in larger governments? Do larger governments breed political discord?

```
##
## Call:
## lm(formula = pf_ss ~ ef_government, data = hfi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5127 -0.9362  0.2548  1.1053  2.3476
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     9.76033    0.18660  52.305   <2e-16 ***
## ef_government  -0.24462    0.02836  -8.624   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.334 on 1376 degrees of freedom
##   (80 observations deleted due to missingness)
## Multiple R-squared:  0.05128,    Adjusted R-squared:  0.05059
## F-statistic: 74.38 on 1 and 1376 DF,  p-value: < 2.2e-16
```