

DATA 606 Data Project Proposal

Peter Gatica

Data Preparation

The data I'm working with is in the openintro package and it's called `hfi`, short for Human Freedom Index.

```
### The data
```

```
data("hfi")
```

```
head(hfi,10)
```

```
## # A tibble: 10 x 123
##   year ISO_code countries region pf_rol_procedur~ pf_rol_civil pf_rol_criminal
##   <dbl> <chr>    <chr>    <chr>          <dbl>         <dbl>         <dbl>
## 1  2016 ALB      Albania  Easte~          6.66          4.55          4.67
## 2  2016 DZA      Algeria  Middl~          NA           NA           NA
## 3  2016 AGO      Angola   Sub-S~          NA           NA           NA
## 4  2016 ARG      Argentina Latin~          7.10          5.79          4.34
## 5  2016 ARM      Armenia  Cauca~          NA           NA           NA
## 6  2016 AUS      Australia Ocean~          8.44          7.53          7.36
## 7  2016 AUT      Austria  Weste~          8.97          7.87          7.67
## 8  2016 AZE      Azerbaij~ Cauca~          NA           NA           NA
## 9  2016 BHS      Bahamas Latin~          6.93          6.01          6.26
## 10 2016 BHR      Bahrain  Middl~          NA           NA           NA
## # ... with 116 more variables: 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>,
## #   pf_ss_women_missing <dbl>, pf_ss_women_inheritance_widows <dbl>,
## #   pf_ss_women_inheritance_daughters <dbl>, pf_ss_women_inheritance <dbl>,
## #   pf_ss_women <dbl>, pf_ss <dbl>, pf_movement_domestic <dbl>,
## #   pf_movement_foreign <dbl>, pf_movement_women <dbl>, pf_movement <dbl>,
## #   pf_religion_estop_establish <dbl>, pf_religion_estop_operate <dbl>,
## #   pf_religion_estop <dbl>, pf_religion_harassment <dbl>,
## #   pf_religion_restrictions <dbl>, pf_religion <dbl>,
## #   pf_association_association <dbl>, pf_association_assembly <dbl>,
## #   pf_association_political_establish <dbl>,
## #   pf_association_political_operate <dbl>, pf_association_political <dbl>,
## #   pf_association_prof_establish <dbl>, pf_association_prof_operate <dbl>,
## #   pf_association_prof <dbl>, pf_association_sport_establish <dbl>,
## #   pf_association_sport_operate <dbl>, pf_association_sport <dbl>,
## #   pf_association <dbl>, pf_expression_killed <dbl>,
## #   pf_expression_jailed <dbl>, pf_expression_influence <dbl>,
```

```
## # pf_expression_control <dbl>, pf_expression_cable <dbl>,
## # pf_expression_newspapers <dbl>, pf_expression_internet <dbl>,
## # pf_expression <dbl>, pf_identity_legal <dbl>,
## # pf_identity_parental_marriage <dbl>, pf_identity_parental_divorce <dbl>,
## # pf_identity_parental <dbl>, pf_identity_sex_male <dbl>,
## # pf_identity_sex_female <dbl>, pf_identity_sex <dbl>,
## # pf_identity_divorce <dbl>, pf_identity <dbl>, pf_score <dbl>,
## # pf_rank <dbl>, ef_government_consumption <dbl>,
## # ef_government_transfers <dbl>, ef_government_enterprises <dbl>,
## # ef_government_tax_income <dbl>, ef_government_tax_payroll <dbl>,
## # ef_government_tax <dbl>, ef_government <dbl>, ef_legal_judicial <dbl>,
## # ef_legal_courts <dbl>, ef_legal_protection <dbl>, ef_legal_military <dbl>,
## # ef_legal_integrity <dbl>, ef_legal_enforcement <dbl>,
## # ef_legal_restrictions <dbl>, ef_legal_police <dbl>, ef_legal_crime <dbl>,
## # ef_legal_gender <dbl>, ef_legal <dbl>, ef_money_growth <dbl>,
## # ef_money_sd <dbl>, ef_money_inflation <dbl>, ef_money_currency <dbl>,
## # ef_money <dbl>, ef_trade_tariffs_revenue <dbl>,
## # ef_trade_tariffs_mean <dbl>, ef_trade_tariffs_sd <dbl>,
## # ef_trade_tariffs <dbl>, ef_trade_regulatory_nontariff <dbl>,
## # ef_trade_regulatory_compliance <dbl>, ef_trade_regulatory <dbl>,
## # ef_trade_black <dbl>, ef_trade_movement_foreign <dbl>,
## # ef_trade_movement_capital <dbl>, ef_trade_movement_visit <dbl>,
## # ef_trade_movement <dbl>, ef_trade <dbl>,
## # ef_regulation_credit_ownership <dbl>, ef_regulation_credit_private <dbl>,
## # ef_regulation_credit_interest <dbl>, ef_regulation_credit <dbl>,
## # ef_regulation_labor_minwage <dbl>, ef_regulation_labor_firing <dbl>,
## # ef_regulation_labor_bargain <dbl>, ef_regulation_labor_hours <dbl>, ...
```

There are a lot of columns in this data. I will take a subset the data to include only the variables I am interested in. I will also rename the columns to be more descriptive.

```
(hfi_subset <- hfi[,c("year", "ISO_code", "countries", "region", "pf_rol", "pf_religion_restrictions", "pf_religion")])
```

```
## # A tibble: 1,458 x 19
##   year ISO_code countries region pf_rol pf_religion_restr~ pf_religion
##   <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 2016 ALB Albania Eastern Euro~ 5.29 8.01 9.19
## 2 2016 DZA Algeria Middle East ~ 3.82 2.96 4.94
## 3 2016 AGO Angola Sub-Saharan ~ 3.45 7.46 8.79
## 4 2016 ARG Argentina Latin Americ~ 5.74 6.85 7.80
## 5 2016 ARM Armenia Caucasus & C~ 5.00 5.09 6.22
## 6 2016 AUS Australia Oceania 7.78 8.94 9.08
## 7 2016 AUT Austria Western Euro~ 8.17 5.64 7.93
## 8 2016 AZE Azerbaijan Caucasus & C~ 4.27 5.46 5.46
## 9 2016 BHS Bahamas Latin Americ~ 6.40 6.43 7.94
## 10 2016 BHR Bahrain Middle East ~ 5.90 4.02 7.46
## # ... with 1,448 more rows, and 12 more variables: pf_expression_control <dbl>,
## # pf_expression <dbl>, pf_score <dbl>, pf_rank <dbl>,
## # ef_money_inflation <dbl>, ef_money_currency <dbl>, ef_money <dbl>,
## # ef_score <dbl>, ef_rank <dbl>, hf_score <dbl>, hf_rank <dbl>,
## # hf_quartile <dbl>
```

```
arrange(hfi_west, desc(region))
```

```
## # A tibble: 20 x 19
##   year ISO_code countries region pf_rol pf_religion_restri~ pf_religion
##   <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 2016 AUT Austria Western E~ 8.17 5.64 7.93
## 2 2016 BEL Belgium Western E~ 7.75 6.06 8.13
## 3 2016 DNK Denmark Western E~ 8.69 5.37 7.70
## 4 2016 FIN Finland Western E~ 8.64 5 7.90
## 5 2016 FRA France Western E~ 6.77 7.17 7.25
## 6 2016 DEU Germany Western E~ 8.30 5.83 7.70
## 7 2016 ISL Iceland Western E~ 7.57 4.81 8.12
## 8 2016 IRL Ireland Western E~ 7.61 8.66 9.08
## 9 2016 ITA Italy Western E~ 6.50 6.34 8.05
## 10 2016 LUX Luxembourg Western E~ 7.90 7.87 9.14
## 11 2016 MLT Malta Western E~ 6.88 6.2 8.45
## 12 2016 NLD Netherlands Western E~ 8.52 8.24 8.89
## 13 2016 NOR Norway Western E~ 8.56 6.11 7.41
## 14 2016 PRT Portugal Western E~ 7.17 8.79 9.43
## 15 2016 ESP Spain Western E~ 7.01 6.25 8.00
## 16 2016 SWE Sweden Western E~ 8.54 8.28 8.88
## 17 2016 CHE Switzerland Western E~ 8.25 6.85 8.19
## 18 2016 GBR United King~ Western E~ 7.72 5.18 7.59
## 19 2016 CAN Canada North Ame~ 7.74 8.57 9.00
## 20 2016 USA United Stat~ North Ame~ 6.95 8.66 8.85
## # ... with 12 more variables: pf_expression_control <dbl>, pf_expression <dbl>,
## # pf_score <dbl>, pf_rank <dbl>, ef_money_inflation <dbl>,
## # ef_money_currency <dbl>, ef_money <dbl>, ef_score <dbl>, ef_rank <dbl>,
## # hf_score <dbl>, hf_rank <dbl>, hf_quartile <dbl>
```

```
arrange(hfI_latin_amer, desc(region))
```

```
## # A tibble: 26 x 19
##   year ISO_code countries region pf_rol pf_religion_restr~ pf_religion
##   <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 2016 ARG Argentina Latin Ameri~ 5.74 6.85 7.80
## 2 2016 BHS Bahamas Latin Ameri~ 6.40 6.43 7.94
## 3 2016 BRD Barbados Latin Ameri~ 6.70 7.54 8.49
## 4 2016 BLZ Belize Latin Ameri~ 4.27 8.52 9.06
## 5 2016 BOL Bolivia Latin Ameri~ 3.05 8.93 8.43
## 6 2016 BRA Brazil Latin Ameri~ 4.65 9.08 7.41
## 7 2016 CHL Chile Latin Ameri~ 6.52 8.66 9.27
## 8 2016 COL Colombia Latin Ameri~ 4.42 8.01 8.74
## 9 2016 CRI Costa Rica Latin Ameri~ 6.63 4.81 8.09
## 10 2016 DOM Dominican ~ Latin Ameri~ 4.35 5.46 8.49
## # ... with 16 more rows, and 12 more variables: pf_expression_control <dbl>,
## # pf_expression <dbl>, pf_score <dbl>, pf_rank <dbl>,
## # ef_money_inflation <dbl>, ef_money_currency <dbl>, ef_money <dbl>,
## # ef_score <dbl>, ef_rank <dbl>, hf_score <dbl>, hf_rank <dbl>,
## # hf_quartile <dbl>
```

Research question

Does the happiness factor differ from world region by region based on religious freedom, economic freedom and freedom of expression?

Cases

Each case represents a region of the world such as Western Europe and N. America, Eastern Europe, etc.

Data collection

The report is co-published by the Cato Institute and the Fraser Institute.

Type of study

This is an observational study.

Data Source

The Human Freedom Index presents the state of human freedom in the world based on a broad measure that encompasses personal, civil, and economic freedom. Human freedom is a social concept that recognizes the dignity of individuals and is defined here as negative liberty or the absence of coercive constraint. Because freedom is inherently valuable and plays a role in human progress, it is worth measuring carefully. The Human Freedom Index is a resource that can help to more objectively observe relationships between freedom and other social and economic phenomena, as well as the ways in which the various dimensions of freedom interact with one another.

The report is co-published by the Cato Institute and the Fraser Institute.

Response

The response variable is 2016 happiness factor and is numerical.

Explanatory

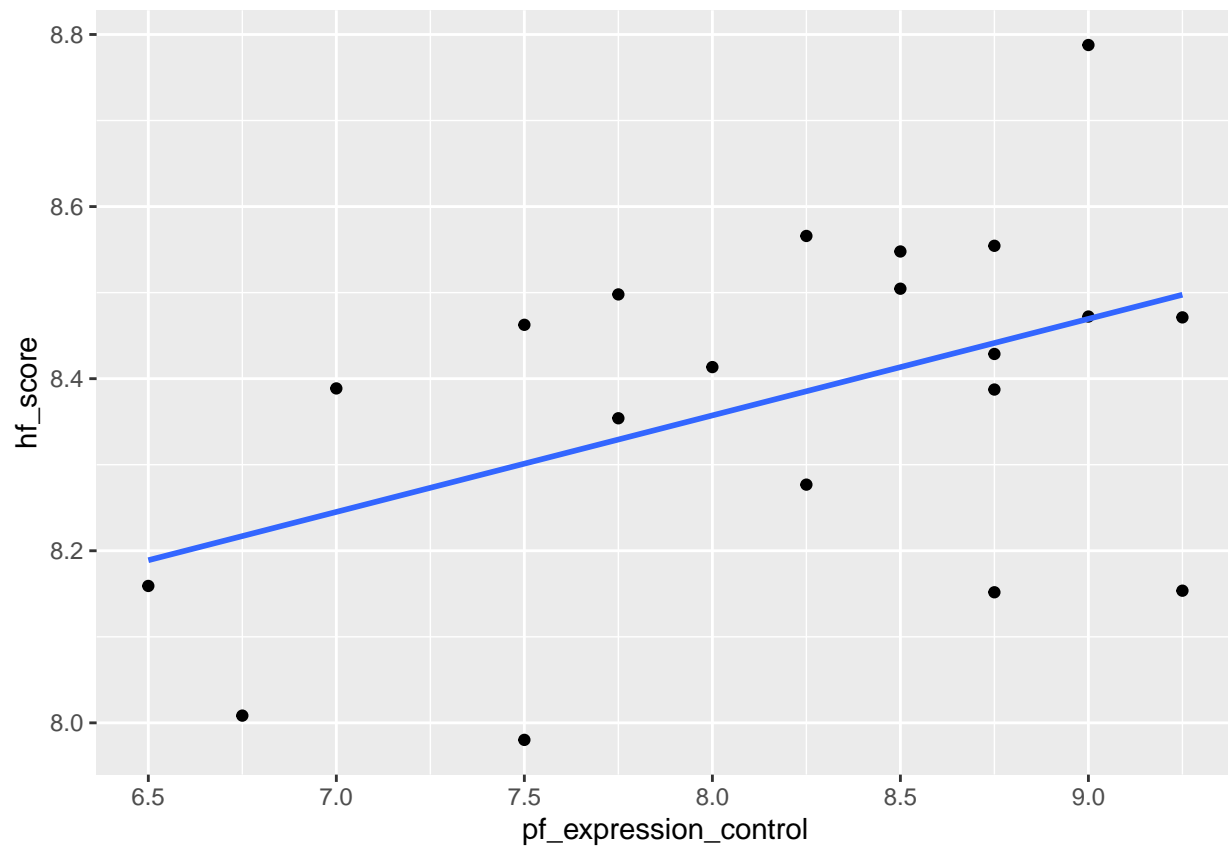
The explanatory variables are ef_score and pf_score and are numerical.

Relevant summary statistics

Provide summary statistics relevant to your research question. For example, if you're comparing means across groups provide means, SDs, sample sizes of each group. This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

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

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



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

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

