Identifying Factors That Affect Political Freedom Within the OECD

Joe Despres, Sabrina Ball, Jacob Haywood, & Joe Sigler

Michigan State University

November 30, 2020

Identifying Factors That Affect Political Freedom Within the OECD

Introduction

2020 is an election year, which has us reflecting on political freedom. Political freedoms are important because they have a dramatic impact on the quality of life for the citizens of a country. Such reflection has us asking: What factors are associated with an increase in political freedom around the world? Political freedom is an abstract concept that does not lend well to quantification. However, Freedomhouse (*Countries and territories*, 2020), a U.S. based think tank and research institute, attempts this by assigning an index of political rights to each country every year based on a fixed criteria. Using the Papaja Package (Aust & Barth, 2020) we take Rmarkdown directly to APA format.

The **Political Rights Index** is composed of three subcategories: electoral process, political participation, and functioning government. Electoral process is a score based on how the current government leaders were selected, mainly how fairly the positions were obtained and kept. As well as the fairness of the current electoral laws, how they are implemented, and the degree to which there is an independent judiciary. Political Participation is evaluated on four criterion, the right to form parties, realistic opposition to current power, political choice free from military, religious powers, economic oligarchies, or any other unaccountable body, and various minority groups having full political rights. Functioning government measures the level of autonomy of the current heads of government when determining policies, safeguards against corruption, and transparency of government operations. Freedom house has measured political freedom in this manner for every country from the years 1978 to present. As we will explain later we are only able to use data from 2000 to 2018.

We have little doubt that this is a meticulously calculated metric, however, this takes discrete values. We are going to be applying a continuous test to these data. Knowing this, we will advise caution is taken in interpretation. One should take these results knowing full well that a core assumption is violated. Despite this, we are reasonably sure that these results can be helpful to confirm suspicion that associations may exist and make policy recommendations. While the exact value of our regression coefficients are not to be interpreted as slope rather an indicator pointing out direction of the association. This will serve as a reasonable measure of political freedom around the world. Additionally, measuring political freedom is done in the same way as measuring physical properties such as weight, velocity, or temperature. Often metrics such as political freedom are made based on scholarly judgment subject to the biases of the people measuring this. Most obvious, these data are collected, assigned, and funded by people in the United States. Naturally, this comes an incalculable cultural bias so we caution our reader to interpret increases or decreases in political freedom to a western notion of what constitutes political freedom.

We began this study by obtaining data from the Gapminder Foundation (Rosling, 2020), a non-profit organization that studies and promotes economic development. After carefully selecting a dependent variable, we browsed data that were also available from Gapminder and selected variables that we felt would be helpful in explaining political freedom in various countries. To explain a country's political rights index we selected: **Corruption Perception Index**, **Education Expenditure**, **Electricity Use**, **Gini Coefficient** a measure of economic inequality, **Internet Users** as a percentage, **Labor Force Participation** Rate, **Military Spending** as percentage of GDP, and **Murders** per million.

Data

These data came in separate sets so we downloaded them individually, cleaned and combined them together using R (R Core Team, 2020) and the Tidyverse Package (Wickham et al., 2019). We included all the code in the appendix. This resulted in a data set with all 196 countries from years 1950 to 2030 (data after 2019 were projections). It was immediately apparent that there was a substantial amount of missing data in an obvious pattern. Less developed countries tended to have more missing data points, and the further back in time the more missing data. Had we conducted this study as it was, it would have been severely biased towards more developed countries as missing data points would be dropped. This forced us to narrow the scope of the study and focus only on countries within the Organization for Economic Co-operation and Development (OECD). This subset of countries has much more complete data. Therefore, we removed data from countries that are not members of the OECD and selected only the years 2000 through 2018. Removing all countries outside of the OECD puts an additional constraint on interpretation. Findings here should also take care to not be interpreted as universally applicable because we are only studying the OECD by necessity.

This remedied most of the missing datapoints. However, even within the OEDC there were a few missing data points. From there, we imputed the median of each country's variable. For instance, Austria's electricity consumption was missing for 2012, so we took the median of Austria's consumption over the 18 year period and inputted it where there was an NA. This afforded us the ability to keep more data with minimal compromise. This is a reasonable procedure because the covariates do not fluctuate wildly year over year. After that, there were a few other cases which we needed to input manually, for example, Iceland's Military spending is 0% of their GDP. This was recorded as NA so we replaced the NA with a 0. This is how we obtained a complete and full data set with no missing data points. The only other adjustment made was to make a few of the variables more intuitive. Corruption perception for example was originally having 0 be the most perception of corruption and 100 being the least. We wanted to have a more intuitive interpretation so we subtracted 100 and multiplied by negative one. This gives us an interpretation of of 0 being the least corrupt and 100 the most corrupt. We did the same for political rights, which originally 1 was the most political freedom and 7 being the least. We want to interpret a higher number to having more political rights.

Descriptive Statistics

Table 1

Descriptive Statistics

Variable	Mean	Median	Std	Min	Max	Range
Political Rights	6.58	7.00	0.80	2.00	7.00	5.00
Year	2009.00	2009.00	5.48	2000.00	2018.00	18.00
Corruption Perception	31.95	29.00	15.87	8.00	71.00	63.00
Education Expenditure	0.20	0.20	0.04	0.10	0.37	0.27
Electricity Use	8598.03	6660.00	7297.72	846.00	54800.00	53954.00
Gini	33.77	32.70	6.41	24.40	57.70	33.30
Internet Users	63.32	69.70	24.25	2.21	99.00	96.79
Life Expectancy	0.60	0.60	0.06	0.45	0.78	0.33
Labor Force Participation	79.18	79.80	2.88	70.20	84.40	14.20
Military Spending	1.69	1.43	1.16	0.00	8.54	8.54
Murders rate	3.17	1.06	6.16	0.15	29.07	28.92

Table 1, made with the KableExtra package (Zhu, 2020) contains basic summary statistics from our dataset. We observed that the mean political rights is 6.6 nearly full with a fairly low standard deviation. This is not ideal. Freedomhouse's classification of political freedom is in discrete values from 1-7. Additionally, countries with a low political freedom have suspect data reporting. We had to select a subset of countries and that subset leans towards having more political freedom. We precede on the understanding that there are moderate problems with the data, first having a discrete response variable and the nature of the distribution being skewed towards the maximum 7/7 political freedom. With these assumption violations in mind, we were not compelled to make any variable transformations and proceed to selecting a model.

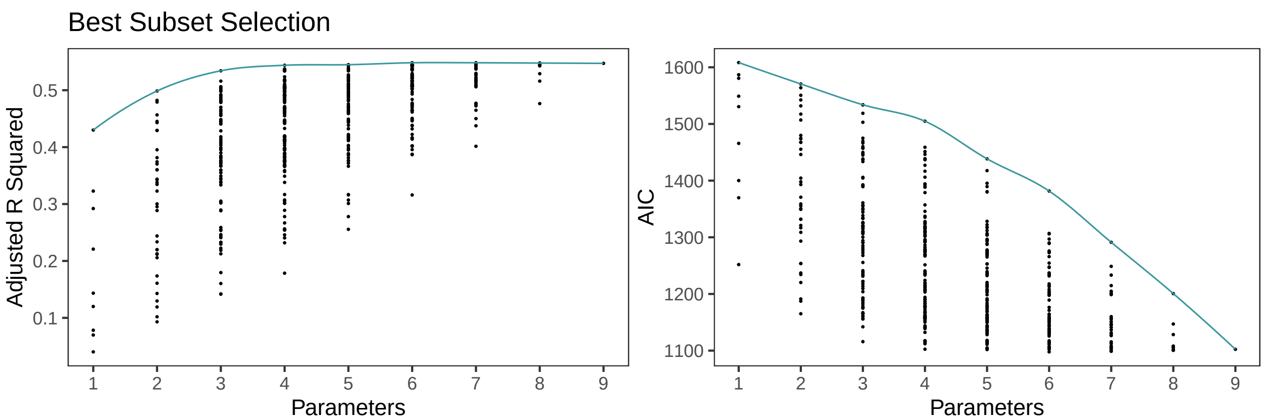


Figure 1

Model Selection

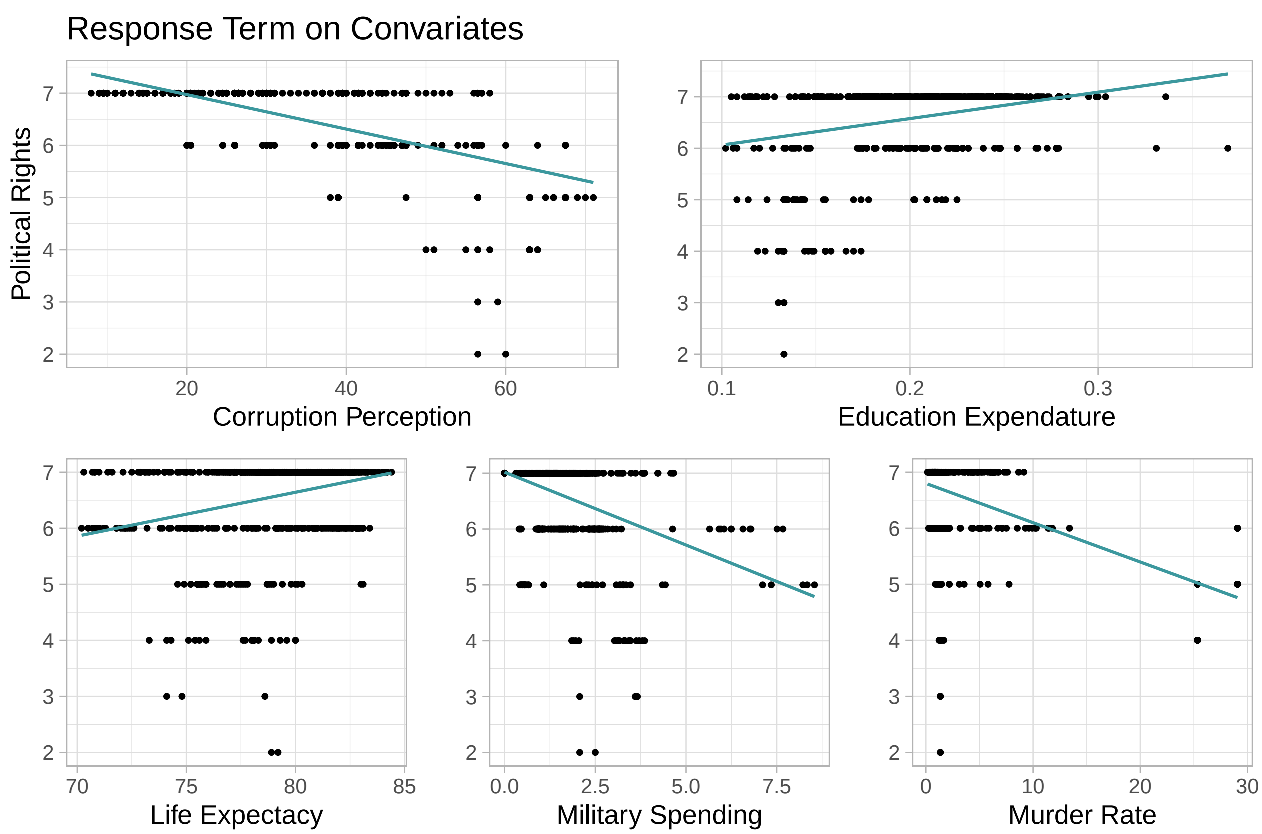
To select a model we performed the best subset algorithm to select several models with the best fit. Graphing the performance of each subset, we can see that the increase in R^2_a and AIC by adding additional parameters diminishes quite rapidly. It was appropriate to narrow our model choices down to three. First, the model with six degrees of freedom, one with five, and one with four. The results from these three models is in Table 2 made using the Stargazer package (Hlavac, 2018). The model with six degrees of freedom contains **Internet Users** with coefficient 0.003, which is statistically significant, however, a one percent increase in the percentage of people using the internet is associated with a 0.003 increase in political freedom is not politically significant. A coefficient that small has us suspecting that the relation that does not seem practically significant. Also, given the violation of assumptions it seems reasonable to exclude it from the model. Looking at the model with five degrees of freedom, all of our coefficients are significant at the 99% level and seem large enough to be indicative of a linear association. The model with four degrees of freedom also has statistically significant coefficients and we would pick it had life expectancy not been significant. This is how we arrived at the model with five degrees of freedom.

Table 2

Regression Results

	Political Rights Index		
	DF = 6	DF = 5	DF = 4
Corruption Perception	-0.022^{***}	-0.023^{***}	-0.023^{***}
	(0.002)	(0.002)	(0.002)
Education Expendature	1.655^{***}	2.076^{***}	2.064^{***}
	(0.550)	(0.525)	(0.526)
Internet Users	0.003^{**}		
	(0.001)		
Life Expectancy	-0.024^{***}	-0.013	
	(0.009)	(0.008)	
Military Spending	-0.149^{***}	-0.160^{***}	-0.162^{***}
	(0.019)	(0.018)	(0.018)
Murder Rate	-0.034^{***}	-0.033^{***}	-0.031^{***}
	(0.004)	(0.004)	(0.004)
Intercept	9.035^{***}	8.306^{***}	7.264^{***}
	(0.731)	(0.672)	(0.120)
Observations	684	684	684
R^2	0.552	0.548	0.546
Adjusted R^2	0.548	0.545	0.544
Residual Std. Error	0.536	0.538	0.539
F Statistic	139.150^{***}	164.500^{***}	204.559^{***}

Note: Standard Errors are in Prentices $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$



The scatter plots look a bit strange at first because of the nature of the discrete response variable, however, this does seem to fit the data reasonably well. There is a fairly linear relationship with a clear sloping line which, except for life expectancy, concur with the regression coefficients. Life expectancy on it's own has a positive association with political freedom, however, when holding the other covariates constant we find that association to be negative.

Now we will examine some regression diagnostics to analyze the performance and identify problems. Let's start with plots of residuals. Figure 2, contains a histogram of the residuals adjacent to a QQ normal plot. It is clear that there are substantial outliers at the lower end of political freedom. Upon further inspection, we found that these outliers are Turkey from 2016 to 2018. These residuals are large and suggest a substantial loss in political freedom. This is confirmed by Freedomhouse (*Countries and territories*, 2020) reporting that a coup attempt in Turkey resulted in a political retaliation against perceived opponents and constitutional changes were made that concentrated political power to the president. After some discussion, we concluded that since the regression results were not severely affected and this is not an entry error removing these data points is inappropriate.

It was clear from the plots that error variance may not be constant. A Brown-Forsythe test confirmed our suspicions informing us that the error variance is not consistent. We calculated a $t^*_{BF} = 9.49$ indicating the error variance is substantially higher when the values of political freedom are lower. This is likely due to the fact that our sample includes predominately values that are on the upper end of the political freedom.

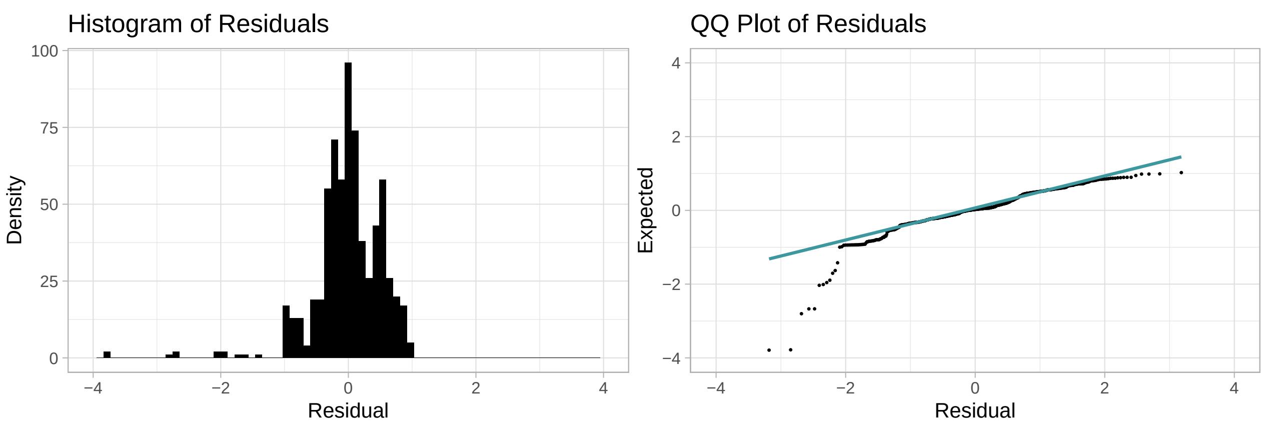


Figure 2

Conclusion

The results are statistically significant but can only be interpreted for countries in the OECD. The model performs reasonably well explaining the political freedom of countries in the 6 and 7 range. However, the model completely falls apart at the one case of a substantial loss in freedom. Again, we don't have an exhaustive sample of countries losing freedom so this is consistent with the dataset we have. However, we can conclude as one of our more robust results is that an increase in education expenditure is associated with an increase in political freedom. The less corrupt a country is perceived to be the more freedom the citizens are afforded. Military spending is associated with a decrease in political freedom. After conducting this study, we are reasonably sure that a relationship exists between political freedom and social, economic, and political factors. We are assured these things matter. Therefore, because we value political freedom, we suggest that when the choice comes, policy makers should choose to spend on education opposed to military when attempting to increase political freedom. Other covariates, policy makers do not have such direct control over.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Countries and territories. (2020). Washington D.C.: Freedomhouse. Retrieved from https://freedomhouse.org

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables.* Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). Retrieved from https://CRAN.R-project.org/package=stargazer

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosling, H. (2020). *Gapminder.* Retrieved from https://www.gapminder.org/data/

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Zhu, H. (2020). *kableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra

Appendix: All code for this report

```r
library(tidyverse)

library(kableExtra)

library(stargazer)

library(patchwork)

library("papaja")

size <- 30

r_refs("r-references.bib")

knitr::opts_chunk$set(echo = FALSE, warning=FALSE, message = FALSE,
                      fig.width = 12, fig.height = 8)

# Read and Clean ----------------------------------------------------------

files <- list.files(path = here::here("raw_data"), pattern = ".csv")

cleaning <- function(df){
  df <- pivot_longer(df, cols = -country, names_to = "year")
}

data <- files %>%
  map(function(x) read_csv(paste0("raw_data/", x))) %>%
  setNames(gsub("\\.csv$", "", files)) %>%
  map(cleaning) %>%
  bind_rows(.id = "id") %>%
  pivot_wider(names_from = id)

# Filtering ---------------------------------------------------------------

countries <- readRDS(here::here("data", "countries.RDS"))

data <- data %>%
  mutate(year = as.numeric(year)) %>%
  filter(year >= 2000, year < 2019, country %in% countries) %>%
```

```
  group_by(country) %>%
  mutate_at(vars(-country),list(~ifelse(is.na(.), median(., na.rm = TRUE), .))) %>%
# Iceland spent 0 so I needed to manually recode that
  mutate(military_spending_pct_of_gdp = replace_na(military_spending_pct_of_gdp, 0),
         murder_per_mil_people = replace(murder_per_mil_people, country == "Mexico", 29.
         murder_per_mil_people = replace(murder_per_mil_people, country == "Chile", 4.4)
         murder_per_mil_people = replace(murder_per_mil_people, country == "Colombia", 2
  relocate(polrights_fh) %>%
  mutate(polrights_fh = (8 - polrights_fh),
 military_spending_pct_of_gdp = military_spending_pct_of_gdp * 100) %>%
  mutate(corruption_perception_index_cpi = (100 - corruption_perception_index_cpi)) %>%
  ungroup()
# Best Subset Selection -------------------------------------------------------------
vars <- data %>%
  select(-country, -year, -ends_with("_fh")) %>%
  names()
models <- list()
for (i in 1:length(vars)) {
  vc <- combn(vars,i)
  for (j in 1:ncol(vc)) {
    model <- as.formula(paste0("polrights_fh", " ~", paste0(vc[,j], collapse = " + ")))
    models <- c(models, model)
    }
  }
subsets <- map(models, function(x) lm(x, data)) %>%
  map(broom::glance) %>%
  setNames(models) %>%
```

```
  bind_rows(.id = "id") %>%
  rename(model = id) %>%
  mutate(Model = str_replace_all(model, "_", " "),
         Model = str_replace(Model, "~", "="),
         Model = str_to_title(Model))
model_df6_formula <- subsets %>% filter(df == 6) %>%
  arrange(desc(adj.r.squared)) %>%
  select(model) %>% head(1) %>% as.character()
fit_df6 <- lm(model_df6_formula, data)
model_df5_formula <- subsets %>% filter(df == 5) %>%
  arrange(desc(adj.r.squared)) %>%
  select(model) %>% head(1) %>% as.character()
fit_df5 <- lm(model_df5_formula, data)
model_df4_formula <- subsets %>% filter(df == 4) %>%
  arrange(desc(adj.r.squared)) %>%
  select(model) %>% head(1) %>% as.character()
fit_df4 <- lm(model_df4_formula, data)
stat <- function(x, df = data, rounding_digits = 2){
    x <- enquo(x)
    df %>%
      summarise(
                       Mean   = mean( !! x),
                     Median   = median( !! x),
                         Std = sd   ( !! x),
                         Min = min  ( !! x),
                         Max = max  ( !! x),
                       Range = max(!! x) - min(!! x) ) %>%
```

```
      mutate_if(is.numeric, round, rounding_digits)
}
data %>%
  select(-country) %>%
map(function(x){stat(x)}) %>%
  bind_rows(.id = "Variable") %>%
  mutate(Variable = replace(Variable,
                            values = c("Political Rights", "Year",
                                       "Corruption Perception",
                                       "Education Expenditure",
                                       "Electricity Use",
                                       "Gini",
                                       "Internet Users",
                                       "Life Expectancy",
                                       "Labor Force Participation",
                                       "Military Spending",
                                       "Murders rate"))) %>%
kable(
  format = "latex",
  booktabs = TRUE,
  escape = FALSE,
  longtable = TRUE,
 caption = "Descriptive Statistics")
r <- subsets %>%
  group_by(df) %>%
  summarise(adj = max(adj.r.squared, na.rm = T))
```

```
subsets %>%
  ggplot(aes(x = df, y = adj.r.squared)) +
  geom_point() +
  geom_smooth(data = r, aes(df, adj),
              se = FALSE, span = 0.5, color = "#3C989E") +
  labs(title = "", x = "Parameters", y = "Adjusted R Squared") +
  scale_x_continuous(breaks = 1:9) +
  theme_test(base_size = size) + labs(title = "Best Subset Selection")

r <- subsets %>%
  group_by(df) %>%
  summarise(AIC = max(AIC, na.rm = T))

subsets %>%
  ggplot(aes(x = df, y = AIC)) +
  geom_point() +
  geom_smooth(data = r, aes(df, AIC),
              se = FALSE, span = 0.5, color = "#3C989E") +
  labs(title = "", x = "Parameters", y = "AIC") +
  scale_x_continuous(breaks = 1:9) +
  theme_test(base_size = size)
stargazer(fit_df6, fit_df5, fit_df4,
                     type = "latex",
                     title = "Regression Results",
                     covariate.labels = c("Corruption Perception",
                                          "Education Expendature",
                                          "Internet Users",
```

```
                                          "Life Expectancy",
                                          "Military Spending",
                                          "Murder Rate",
                                          "Intercept"),
                     dep.var.caption = "Political Rights Index",
                     dep.var.labels = NULL,
                     dep.var.labels.include = FALSE,
                     model.names = FALSE,
                     model.numbers = FALSE,
                     df = FALSE,
                     header = FALSE,
                     column.labels = c("DF = 6", "DF = 5", "DF = 4"),
                     no.space = TRUE,
                     notes.label = "Note: Standard Errors are in Prentices")
size <- 18
a <- data %>%
  ggplot(aes(x = corruption_perception_index_cpi, y = polrights_fh)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "#3C989E") +
  scale_y_continuous(breaks = 1:7) +
  labs(x = "Corruption Perception", y = "Political Rights") +
  theme_light(base_size = size) + labs(title = "Response Term on Convariates")

b <- data %>%
  ggplot(aes(x = edu_exp_gdp_per_person, y = polrights_fh)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE,color = "#3C989E") +
```

```
    scale_y_continuous(breaks = 1:7) +
  labs(x = "Education Expendature", y = "") +
  theme_light(base_size = size)

c <- data %>%
  ggplot(aes(x = life_expectancy_years, y = polrights_fh)) +
  geom_point()+
  geom_smooth(method = "lm", se = FALSE, color = "#3C989E") +
    scale_y_continuous(breaks = 1:7) +
  labs(x = "Life Expectacy", y = "") +
  theme_light(base_size = size)

d <- data %>%
  ggplot(aes(x = military_spending_pct_of_gdp, y = polrights_fh)) +
  geom_point()+
  scale_y_continuous(breaks = 1:7) +
  geom_smooth(method = "lm", se = FALSE, color = "#3C989E") +
  labs(x = "Military Spending", y = "")+ theme_light(base_size = size)

e <- data %>%
  ggplot(aes(x = murder_per_mil_people, y = polrights_fh)) +
  geom_point() +
  scale_y_continuous(breaks = 1:7) +
  geom_smooth(method = "lm", se = FALSE, color = "#3C989E") +
  labs(x = "Murder Rate", y = "") + theme_light(base_size = size)

(a + b) / (c + d + e)
```

```
fit_df5 %>%
  broom::augment() %>%
  ggplot(aes(.resid)) +
  geom_histogram(bins = 75, fill = "black") +
  scale_x_continuous(limits = c(-4,4)) +
  labs(y = "Density", x = "Residual", title = "Histogram of Residuals") +
  theme_light(base_size = size + 10)



fit_df5 %>%
  broom::augment() %>%
  ggplot(aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line(color = "#3C989E", size = 2) +
scale_x_continuous(limits = c(-4,4)) +
scale_y_continuous(limits = c(-4,4)) +
  labs(title = "QQ Plot of Residuals ", x = "Residual", y = "Expected") +
  theme_light(base_size = size + 10)
beepr::beep()
Sys.sleep(1)
```