Tidy Tuesday Freedom Dataset

M. Zhang

2022-03-01 (Update 2022-03-09)

Contents

Introduction

We will be working with the `freedom.csv` dataset via the Tidy Tuesday repository. This analysis will not be styled as a "final report", but rather as a quick walk through of some data analysis and wrangling that took place while exploring the data.

The MSDSO Discord group for the University of Texas Masters in Data Science Online program will be doing weekly explorations of TidyTuesday as an exercise for improving their data science skill sets in a collaborative environment.

For this data set, the analysis was first done independently, then augmented by following along with David Robinson's livestream video (https://youtu.be/VOzUHk3aaBw) for additional learning.

Data Details

This dataset is pulled from the Tidy Tuesday Repository:

Thomas Mock (2022). Tidy Tuesday: A weekly data project aimed at the R ecosystem. https://github.com /rfordatascience/tidytuesday.

The original data is from Freedom House and the United Nationsvia Arthur Cheib.

Freedom House is a nonpartisan organization focused on producing research and reports on themes and trends related to democracy, political rights, and civil liberties. This data set `freedom` contains information in regards to various country's Civil Liberty `CL` and Political Rights `PR` index scores, as well as their Least Developed Country `is_ldc` indicator.

Data Loading

```
#freedom <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/
```

```
freedom <- tt_load("2022-02-22")$freedom %>%
  janitor::clean_names() %>%
  rename(civil_liberties = cl,
         political_rights = pr) %>%
  mutate(country_code = countrycode(country, origin = "country.name", destination = "iso2c"))
```

```
    Downloading file 1 of 1: `freedom.csv`
```

```
freedom
```

```
# A tibble: 4,979 x 9
   country  year civil_liberties political_rights status region_code region_name
   <chr>   <dbl>           <dbl>            <dbl> <chr>        <dbl> <chr>
 1 Afghan~  1995               7                7 NF             142 Asia
 2 Afghan~  1996               7                7 NF             142 Asia
 3 Afghan~  1997               7                7 NF             142 Asia
 4 Afghan~  1998               7                7 NF             142 Asia
 5 Afghan~  1999               7                7 NF             142 Asia
 6 Afghan~  2000               7                7 NF             142 Asia
 7 Afghan~  2001               7                7 NF             142 Asia
 8 Afghan~  2002               6                6 NF             142 Asia
 9 Afghan~  2003               6                6 NF             142 Asia
10 Afghan~  2004               6                5 NF             142 Asia
# ... with 4,969 more rows, and 2 more variables: is_ldc <dbl>,
#   country_code <chr>
```

Data Dictionary

The following tables contain information in regards to the columns available.

`freedom.csv`

variable	class	description
country	character	Country Name
year	double	Year
CL	double	Civil Liberties
PR	double	Political rights
Status	character	Status (Free `F`, Not Free `NF`, Partially Free `PF`)
Region_Code	double	UN Region code
region_name	character	UN Region Name
is_ldc	double	Is a least developed country (binary 0/1)

The definition for "Least Developed Country" is pulled from the United Nations. A country qualifies for LDC if it meets the criteria for Income, Human Assets, and Economic and Environmental Vulnerability. An

important requirement for inclusion is that the country must agree to the classification to be added to the list.

The Civil Liberties and Political Rights score is generated by FreedomHouse, using a methodology inspired by the Universal Declaration of Human Rights which was adopted by the UN General Assembly in 1948, The Civil Liberties score is a combination of 15 separate indicators, and Political Rights score is a combination of 10 separate indicators. Each of these indicators scale from 0 to 4, with 4 representing the greatest amount of freedom. These scores are summarized into indexes for Civil Liberties `civil_liberties` and Political Rights `political_rights` on a scale from 1 to 7, with 1 representing the greatest freedom.

`Status` buckets the combined `CL` and `PR` scores into 3 general categories: Free, Partially Free, and Not Free.

Additional details are available here.

Data Exploration and wrangling

Examining the raw data

We will initially take a precursor inspection of the data, utilizing `summary` for numerical information and `table` for categorical information.

```
glimpse(freedom)
```

```
Rows: 4,979
Columns: 9
$ country         <chr> "Afghanistan", "Afghanistan", "Afghanistan", "Afghani~
$ year            <dbl> 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,~
$ civil_liberties <dbl> 7, 7, 7, 7, 7, 7, 7, 6, 6, 6, 5, 5, 5, 6, 6, 6, 6, 6,~
$ political_rights <dbl> 7, 7, 7, 7, 7, 7, 7, 6, 6, 5, 5, 5, 5, 5, 6, 6, 6, 6,~
$ status          <chr> "NF", "NF", "NF", "NF", "NF", "NF", "NF", "NF", "NF",~
$ region_code     <dbl> 142, 142, 142, 142, 142, 142, 142, 142, 142, 142, 142~
$ region_name     <chr> "Asia", "Asia", "Asia", "Asia", "Asia", "Asia", "Asia~
$ is_ldc          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ country_code    <chr> "AF", "AF", "AF", "AF", "AF", "AF", "AF", "AF", "AF",~
```

```
names(freedom)
```

```
[1] "country"         "year"            "civil_liberties" "political_rights"
[5] "status"          "region_code"     "region_name"     "is_ldc"
[9] "country_code"
```

```
freedom %>%
  select(civil_liberties, political_rights, is_ldc) %>%
  summary(freedom)
```

```
 civil_liberties political_rights     is_ldc
 Min.   :1.000   Min.   :1.000   Min.   :0.0000
 1st Qu.:2.000   1st Qu.:1.000   1st Qu.:0.0000
 Median :3.000   Median :3.000   Median :0.0000
 Mean   :3.369   Mean   :3.411   Mean   :0.2362
 3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:0.0000
 Max.   :7.000   Max.   :7.000   Max.   :1.0000
```

```
freedom %>%
  count(country) %>%
  arrange(n)
```

```
# A tibble: 193 x 2
   country              n
```

```
   <chr>               <int>
 1 South Sudan            10
 2 Montenegro             15
 3 Serbia                 18
 4 Timor-Leste            22
 5 Afghanistan            26
 6 Albania                26
 7 Algeria                26
 8 Andorra                26
 9 Angola                 26
10 Antigua and Barbuda    26
# ... with 183 more rows
```

```r
freedom %>% select(year) %>% table(useNA = "ifany")
```

```
.
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
 189  189  189  189  190  190  190  190  191  191  191  192  192  192  192  192
2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
 193  193  193  193  193  193  193  193  193  193
```

```r
freedom %>% select(status) %>% table(useNA = "ifany")
```

```
.
   F   NF   PF
2219 1257 1503
```

```r
freedom %>% select(region_code) %>% table(useNA = "ifany")
```

```
.
   2    9   19  142  150
1388  364  910 1218 1099
```

```r
freedom %>% select(region_name) %>% table(useNA = "ifany")
```

```
.
  Africa Americas     Asia   Europe  Oceania
    1388      910     1218     1099      364
```

```r
freedom %>% select(is_ldc) %>% table(useNA = "ifany")
```

```
.
   0    1
3803 1176
```

```r
freedom %>%
  distinct(country, region_name) %>% View()
```

We can see that it appears that not all countries have full data for all 26 years. There is data ranging from 1995 through 2020, with a total of 193 countries in the data set.

A precursory glance suggest that not all countries have the full year range of data. There are no unexpected unique values from our initial look at the data. Additionally, the `political_rights` and `civil_liberties` (renamed as `polical_rights` and `civil_liberties`) scores are within the expected ranges given the definitions provided by FreedomHouse.

Checking for NA values

```
freedom %>%
  summarize(across(everything(), ~ sum(is.na(.)))) %>%
  tidyr::pivot_longer(everything()) %>%
  arrange(desc(value)) %>%
  deframe()
```

```
        country            year  civil_liberties political_rights
              0               0                0                0
         status     region_code      region_name           is_ldc
              0               0                0                0
   country_code
              0
```

NA counting methodology was taken from stackexchange

It appears that there are no NA values in this data set. This is reaffirming our findings from when we checked the summary and tables for the dataset earlier.

Recoding

Going forward, we will be using the `region_name` column in lieu of `Region_Code` for simplicity. Additionally, we will recode `is_ldc` into categorical values.

```
freedom = freedom %>%
  mutate(ldc = if_else(is_ldc == 1, "Yes", "No"),
         freedom_index = political_rights + civil_liberties)

freedom %>%
  select(is_ldc, ldc) %>% unique()
```

```
# A tibble: 2 x 2
  is_ldc ldc
   <dbl> <chr>
1      1 Yes
2      0 No
```

Grouping candidates

Let's examine the total number of unique entries per column to see good candidates for faceting or other categorization methods.

```
freedom %>%
  sapply(n_distinct)
```

```
        country            year  civil_liberties political_rights
            193              26                7                7
         status     region_code      region_name           is_ldc
              3               5                5                2
   country_code             ldc    freedom_index
            193               2               13
```
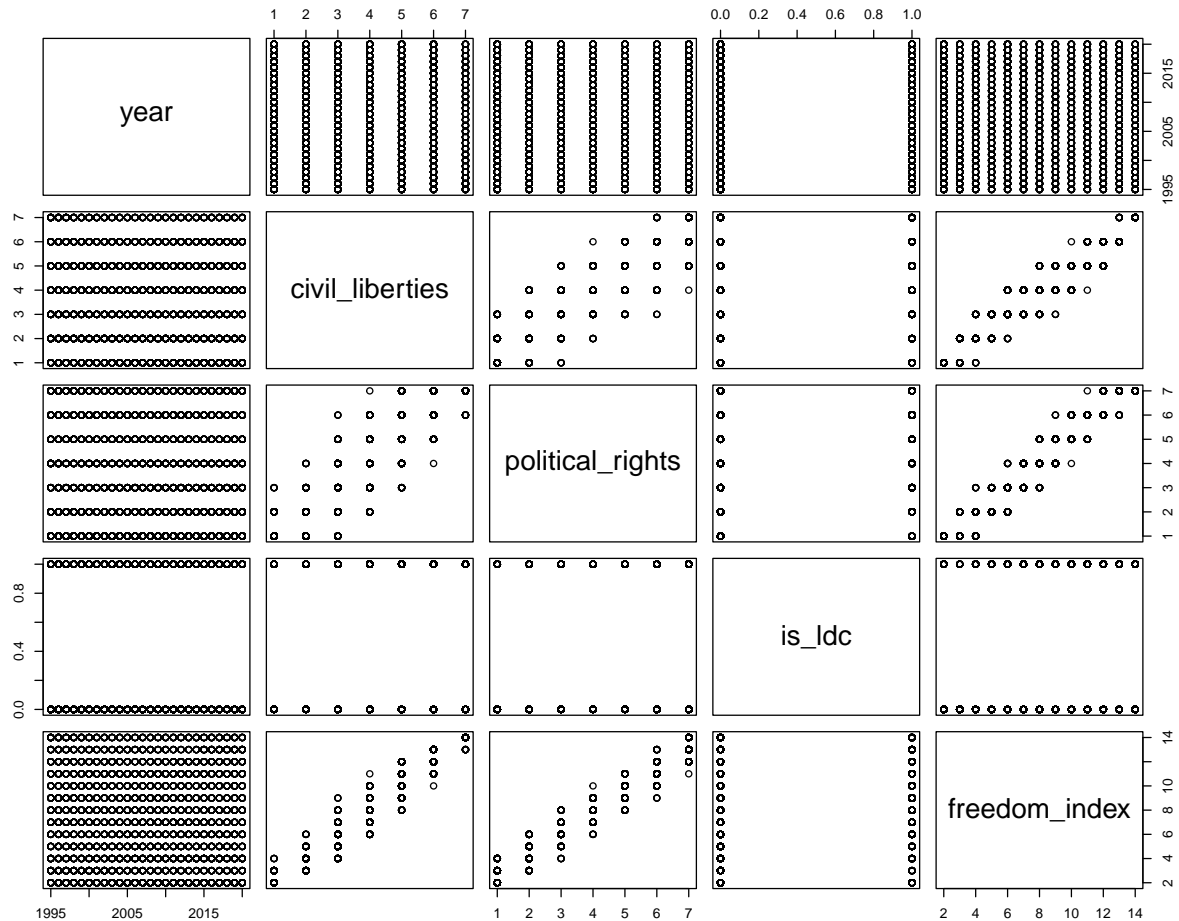
We can see that `LDC`, `Status` and `region_name` are all potential ways to cluster data.
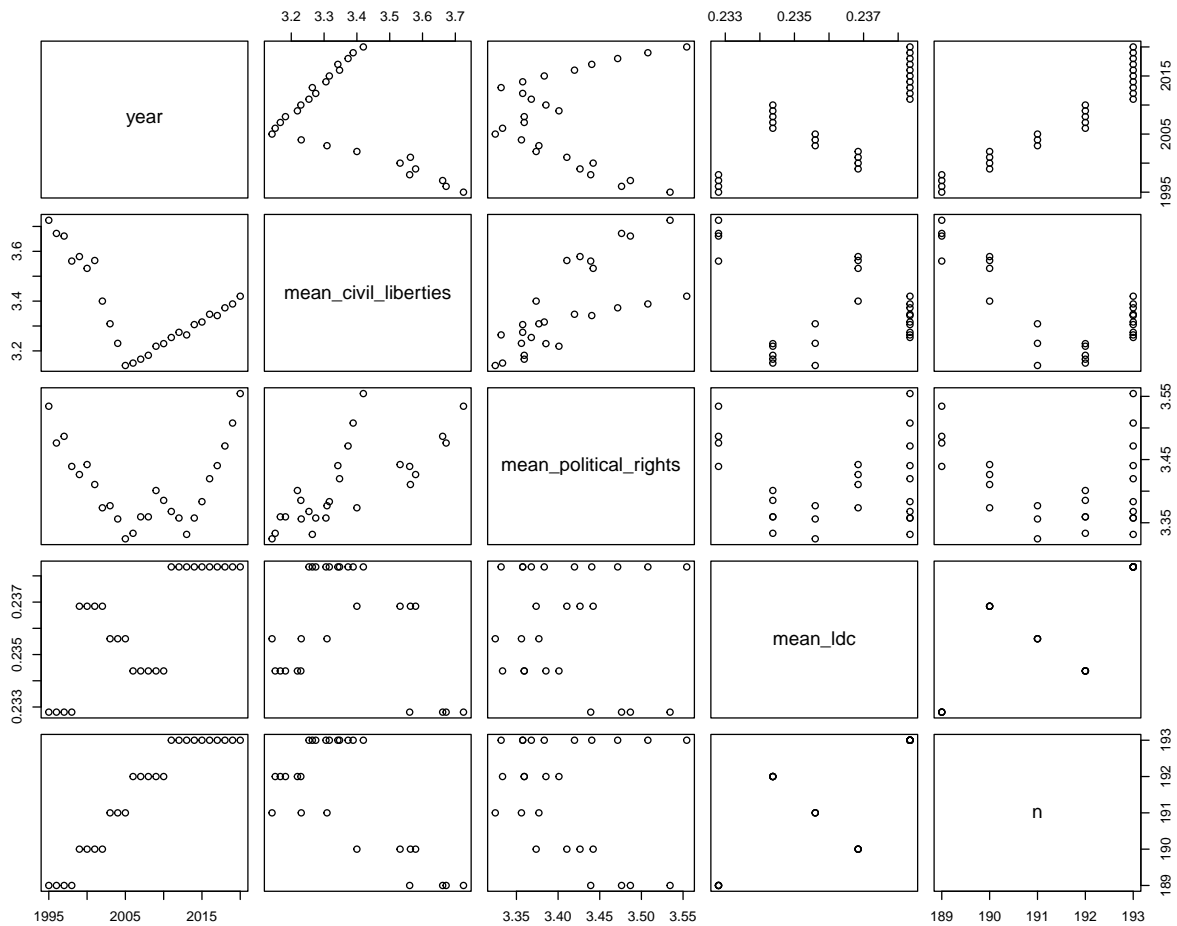
Visual Exploration

Pairwise plot for data - check raw data

```
freedom %>%
  select(year, civil_liberties, political_rights, is_ldc, freedom_index) %>%
  pairs()
```


Check grouped data

```
freedom %>%
  group_by(year) %>%
  summarize(mean_civil_liberties = mean(civil_liberties), mean_political_rights = mean(political_rights)
  select(year, mean_civil_liberties, mean_political_rights, mean_ldc, n) %>%
  pairs()
```

```r
summarize_freedom <- function(tbl) {
  tbl %>%
    summarize(n_countries = n(),
              avg_civil_lib = mean(civil_liberties),
              avg_pol_rights = mean(political_rights),
              pct_free = mean(status == "F"),
              .groups = "drop") %>%
    arrange(desc(n_countries))
}
```

```r
by_region <- freedom %>%
  filter(year == 2020) %>%
  group_by(region_name) %>%
  summarize(n_countries = n(),
            avg_civil_lib = mean(civil_liberties),
            avg_pol_rights = mean(political_rights),
            pct_free = mean(status == "F"))

by_region
```

```
# A tibble: 5 x 5
  region_name n_countries avg_civil_lib avg_pol_rights pct_free
```

```
  <chr>            <int>         <dbl>         <dbl>     <dbl>
1 Africa              54          4.44          4.81     0.167
2 Americas            35          2.57          2.51     0.6
3 Asia                47          4.83          4.98     0.128
4 Europe              43          1.84          1.86     0.767
5 Oceania             14          1.71          1.71     0.857
```

```
by_region %>% ggplot(aes(avg_civil_lib, avg_pol_rights)) +
  geom_point(aes(size = n_countries)) +
  geom_abline(color = "red") +
  geom_text(aes(label = region_name), vjust = 0, hjust = 1.5) +
  expand_limits(x = 0, y = 0, size = 0)
```
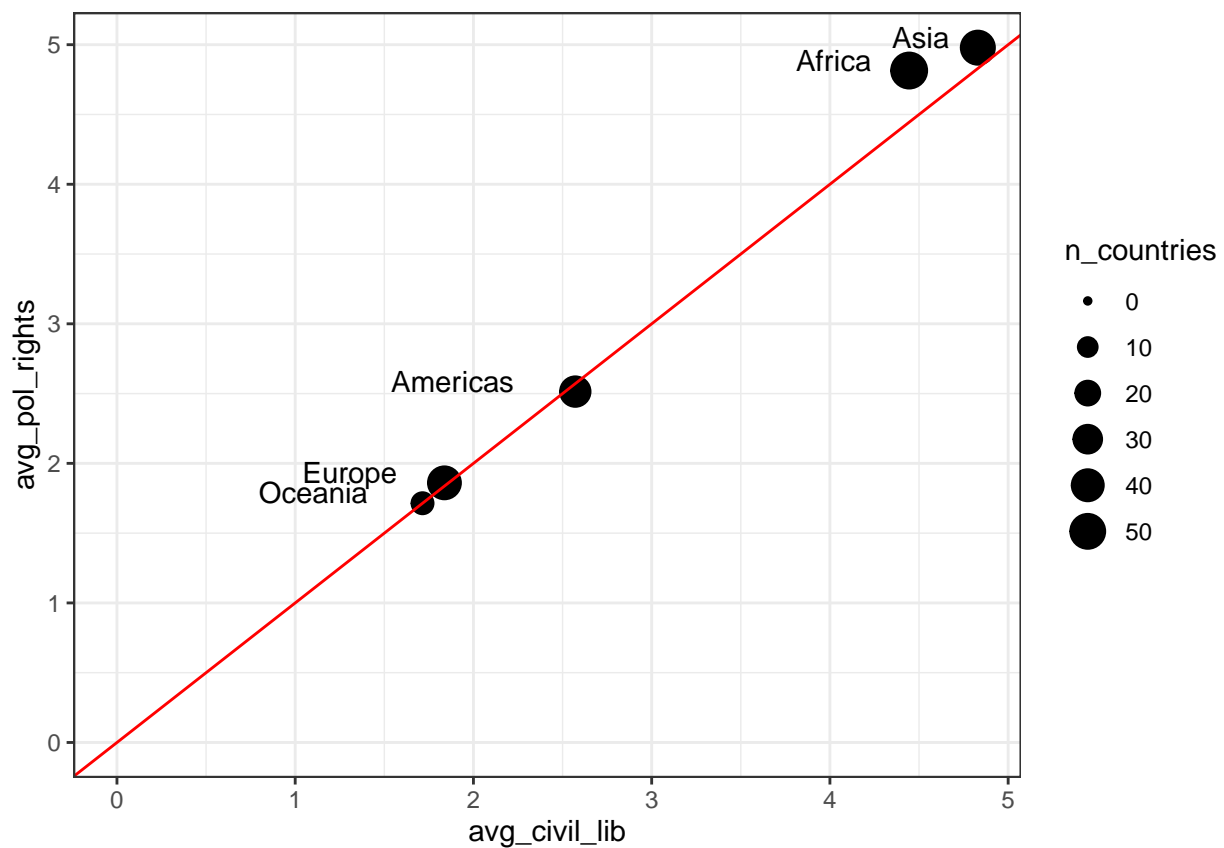


```
freedom %>%
  filter(year == 2020) %>%
  ggplot(aes(civil_liberties, political_rights)) +
  geom_jitter(height = 0.1, width = 0.1) +
  geom_abline(color = "red") +
  expand_limits(x = 0, y = 0, size = 0)
```
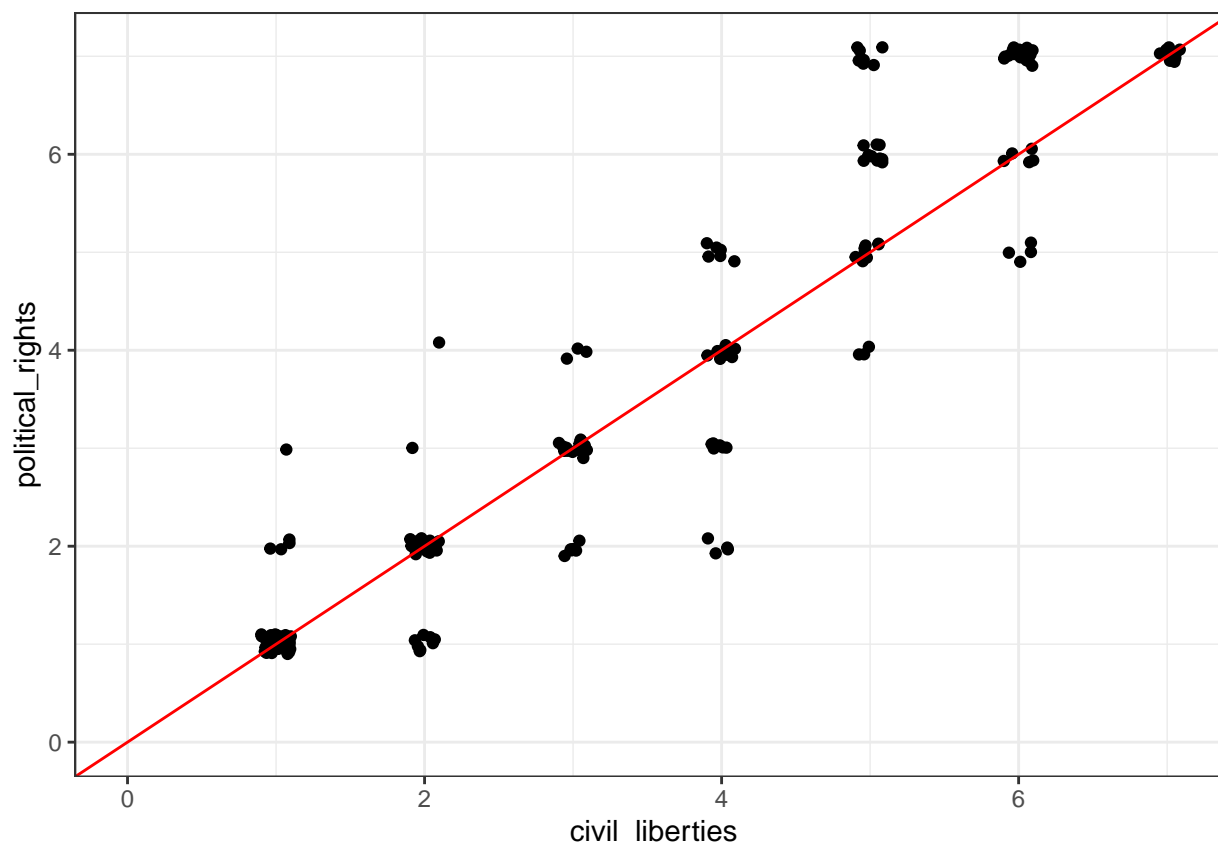
Developing Questions

While examining the data set, the following potential questions arose for investigation:

1. What is the distribution of freedom by Region in a single year?
2. How have the 5 of the best, and 5 of worst non LDC countries shifted in terms of freedom from 1995 to 2020?
3. How are proportions of Free, Partially Free, and Not Free countries shifting over time?
4. How are the distribution of Political Rights, and Civil Liberties fluctuating over time by region?

For the second question, we will need to set criteria for "worst" and "best", and then identify these countries. We will use a combined `political_rights + civil_liberties` score as the gauge of measure. A score of 14 would represent "worst" and 2 would represent "best". We will also exclude countries that do not have data for all years.

```
freedom %>%
  filter(year == 1995, ldc == "No") %>%
  select(country, region_name, ldc, freedom_index) %>%
  arrange(freedom_index) %>% head(20)
```

```
# A tibble: 20 x 4
   country                  region_name ldc    freedom_index
   <chr>                    <chr>       <chr>          <dbl>
 1 Andorra                  Europe      No                 2
 2 Australia                Oceania     No                 2
 3 Austria                  Europe      No                 2
 4 Barbados                 Americas    No                 2
 5 Belgium                  Europe      No                 2
```

```
 6 Belize                               Americas    No                2
 7 Canada                               Americas    No                2
 8 Cyprus                               Asia        No                2
 9 Denmark                              Europe      No                2
10 Dominica                             Americas    No                2
11 Finland                              Europe      No                2
12 Iceland                              Europe      No                2
13 Ireland                              Europe      No                2
14 Liechtenstein                        Europe      No                2
15 Luxembourg                           Europe      No                2
16 Malta                                Europe      No                2
17 Marshall Islands                     Oceania     No                2
18 Micronesia (Federated States of)     Oceania     No                2
19 Netherlands                          Europe      No                2
20 New Zealand                          Oceania     No                2
```

```r
freedom %>%
  filter(year == 1995, ldc == "No") %>%
  select(country, region_name, ldc, freedom_index) %>%
  arrange(desc(freedom_index)) %>% head(20)
```

```
# A tibble: 20 x 4
   country                                region_name ldc   freedom_index
   <chr>                                  <chr>       <chr>         <dbl>
 1 China                                  Asia        No               14
 2 Cuba                                   Americas    No               14
 3 Equatorial Guinea                      Africa      No               14
 4 Iraq                                   Asia        No               14
 5 Libya                                  Africa      No               14
 6 Nigeria                                Africa      No               14
 7 Democratic People's Republic of Korea  Asia        No               14
 8 Saudi Arabia                           Asia        No               14
 9 Syrian Arab Republic                   Asia        No               14
10 Tajikistan                             Asia        No               14
11 Turkmenistan                           Asia        No               14
12 Uzbekistan                             Asia        No               14
13 Viet Nam                               Asia        No               14
14 Indonesia                              Asia        No               13
15 Iran (Islamic Republic of)            Asia        No               13
16 Kenya                                  Africa      No               13
17 Qatar                                  Asia        No               13
18 Algeria                                Africa      No               12
19 Azerbaijan                             Asia        No               12
20 Bahrain                                Asia        No               12
```

```r
freedom %>%
  filter(year == 1995, ldc == "No", freedom_index == 2) %>%
  arrange(freedom_index) %>%
  select(country) %>%
  pull()
```

```
 [1] "Andorra"              "Australia"
 [3] "Austria"              "Barbados"
 [5] "Belgium"              "Belize"
 [7] "Canada"               "Cyprus"
```

```
 [9] "Denmark"                   "Dominica"
[11] "Finland"                   "Iceland"
[13] "Ireland"                   "Liechtenstein"
[15] "Luxembourg"                "Malta"
[17] "Marshall Islands"          "Micronesia (Federated States of)"
[19] "Netherlands"               "New Zealand"
[21] "Norway"                    "Portugal"
[23] "San Marino"                "Sweden"
[25] "Switzerland"               "United States of America"
```

```
freedom %>%
  filter(year == 1995, ldc == "No", freedom_index == 14) %>%
  arrange(freedom_index) %>%
  select(country) %>%
  pull()
```

```
 [1] "China"
 [2] "Cuba"
 [3] "Equatorial Guinea"
 [4] "Iraq"
 [5] "Libya"
 [6] "Nigeria"
 [7] "Democratic People's Republic of Korea"
 [8] "Saudi Arabia"
 [9] "Syrian Arab Republic"
[10] "Tajikistan"
[11] "Turkmenistan"
[12] "Uzbekistan"
[13] "Viet Nam"
```

```
exclude_list <- freedom %>%
  group_by(country) %>%
  summarize(n = n()) %>%
  filter(n != 26) %>%
  select(country) %>% pull()
exclude_list
```

```
[1] "Montenegro"  "Serbia"      "South Sudan" "Timor-Leste"
```

A look at the data showed that there was not enough granularity in the scale to determine soley based off the data, thus we chose the following 5 countries based off interest and recent political events.

```
best_list <- c("Australia", "Canada", "Belgium", "Iceland", "United States of America")
worst_list <- c("China", "Iraq", "Nigeria", "Saudi Arabia", "Viet Nam")
choice_list <-c("Russian Federation", "Ukraine", "Qatar", "Afghanistan", "Rwanda")
best_list
```

```
[1] "Australia"                 "Canada"
[3] "Belgium"                   "Iceland"
[5] "United States of America"
```

```
worst_list
```

```
[1] "China"        "Iraq"         "Nigeria"      "Saudi Arabia" "Viet Nam"
```

Data Visualizations

For the first question, we will look at the most recent year in the data set, 2020, and follow the analysis done by David Robinson:

```
freedom %>%
  filter(year == 2020) %>%
  gather(key = metric, value = value, civil_liberties, political_rights) %>%
  mutate(metric = str_to_title(str_replace_all(metric, "_", " ")),
         region_name = fct_reorder(region_name, value)) %>%
  count(region_name, metric, value) %>%
  ggplot(aes(value, n)) +
  geom_col() +
  facet_grid(region_name ~ metric) +
  scale_x_continuous(name = "World Freedom Index") +
  scale_y_continuous(name = "Count of Countries") +
  ggtitle("Distribution of Freedom Index By Region")
```

Distribution of Freedom Index By Region
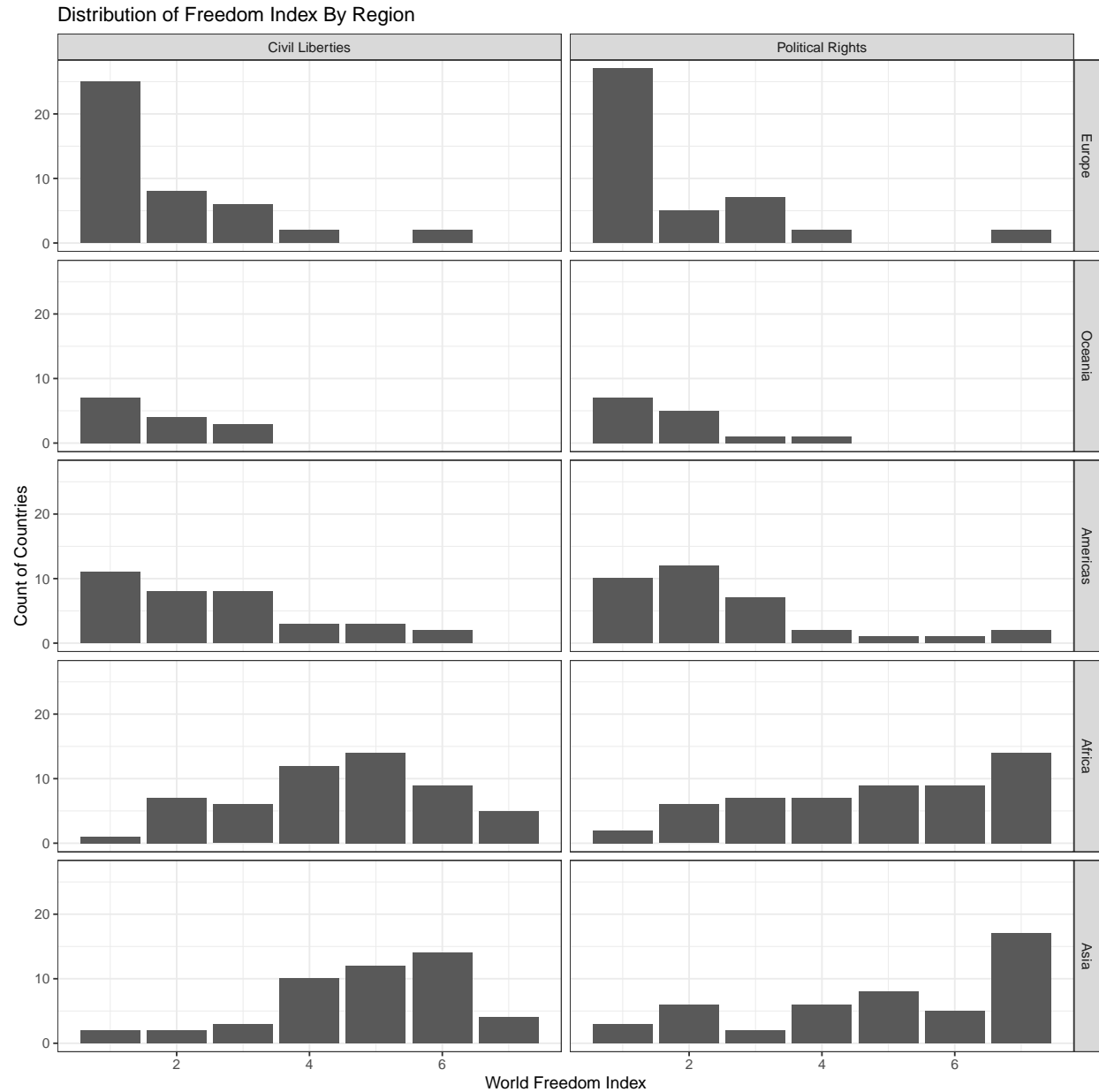
Trends for top and worst 5, and choice

```
freedom %>%
  filter(country %in% best_list) %>%
  pivot_longer(cols = c("political_rights", "civil_liberties"), names_to = "Label", values_to = "Score")
  ggplot() +
  aes(x = year, y = Score, color = Label) +
  facet_wrap(
    vars(country),
    ncol = 1) +
  geom_line(size = 2) +
  scale_y_continuous(
    limit = c(0, 7)) +
```

```
scale_x_continuous(
  name = "Year") +
theme_classic()
```
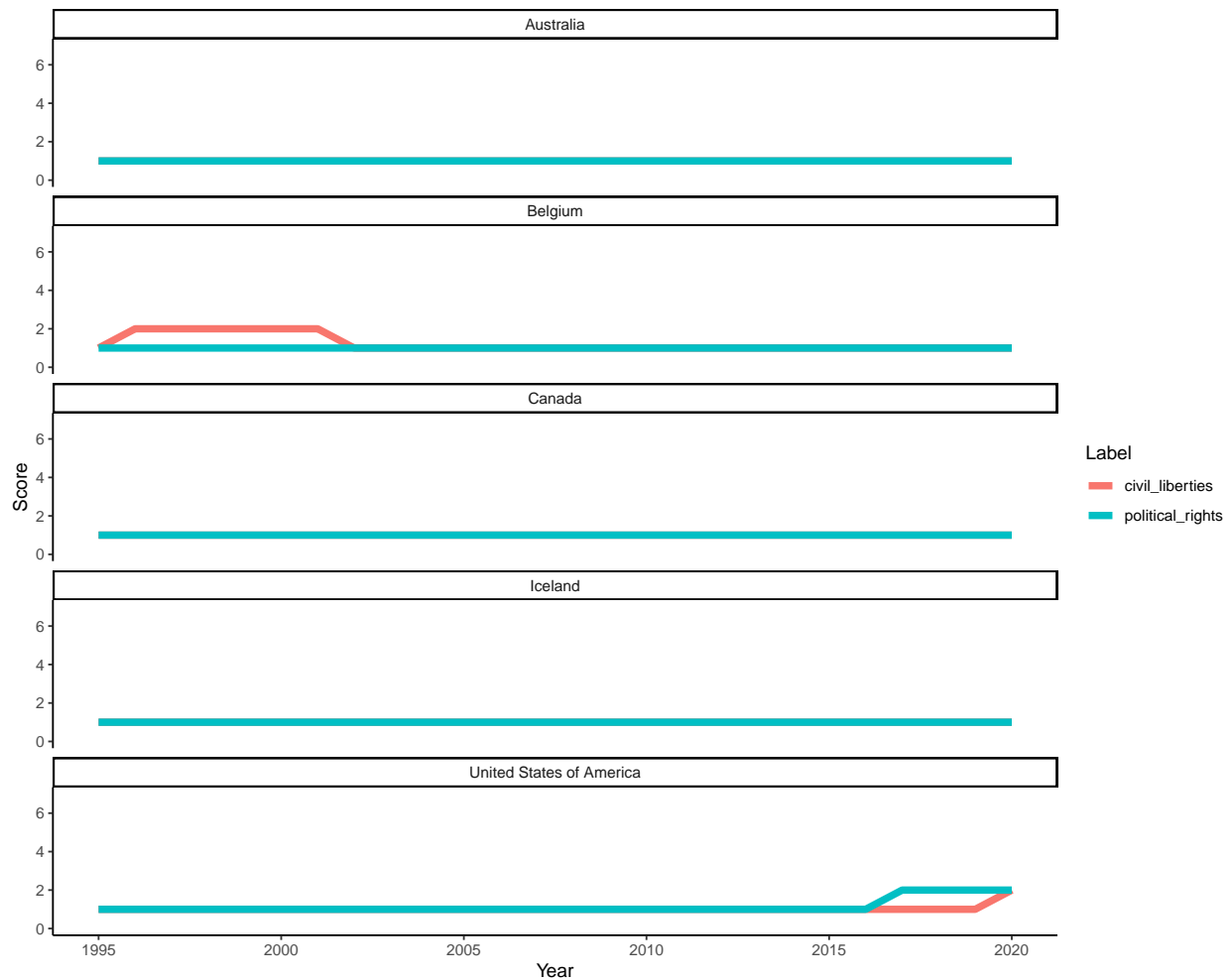


```
freedom %>%
  filter(country %in% worst_list) %>%
  pivot_longer(cols = c("political_rights", "civil_liberties"), names_to = "Label", values_to = "Score")
  ggplot() +
  aes(x = year, y = Score, color = Label) +
  facet_wrap(
    vars(country),
    ncol = 1) +
  geom_line(size = 2) +
  scale_y_continuous(
    limit = c(0, 7)) +
  scale_x_continuous(
    name = "Year") +
  theme_classic()
```
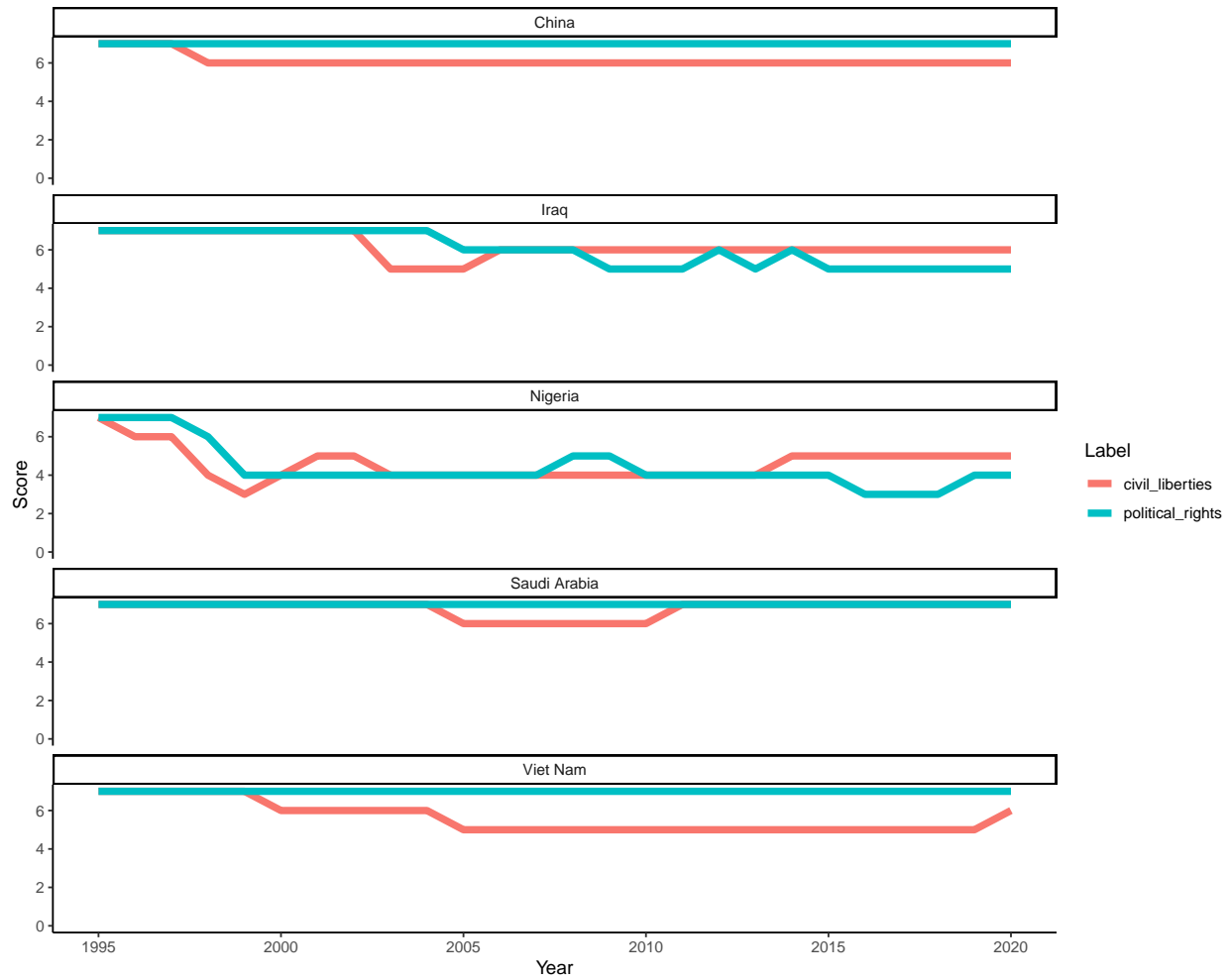
```
freedom %>%
  filter(country %in% choice_list) %>%
  pivot_longer(cols = c("political_rights", "civil_liberties"), names_to = "Label", values_to = "Score")
  ggplot() +
  aes(x = year, y = Score, color = Label) +
  facet_wrap(
    vars(country),
    ncol = 1) +
  geom_line(size = 2) +
  scale_y_continuous(
    limit = c(0, 7)) +
  scale_x_continuous(
    name = "Year") +
  theme_classic()
```
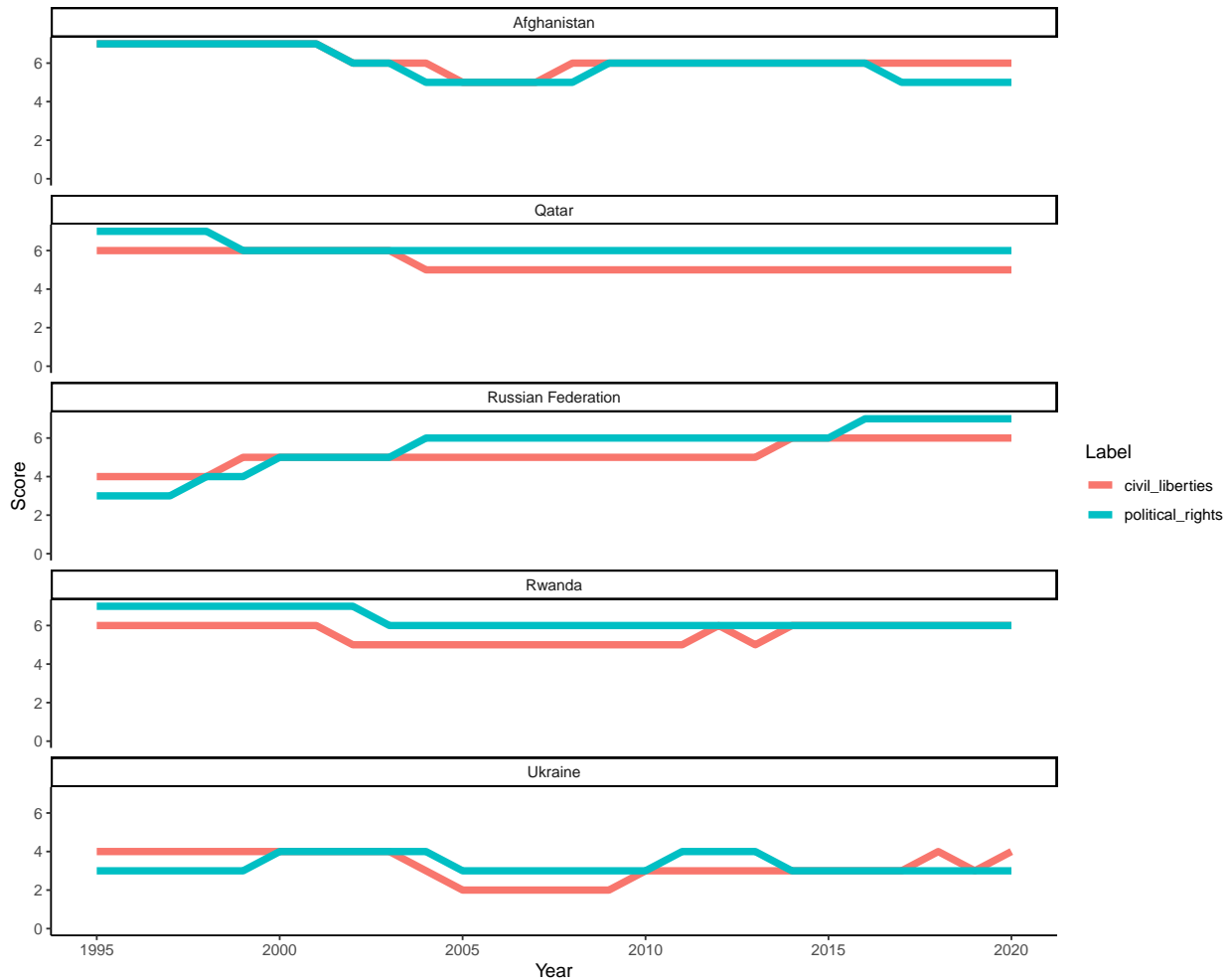
Unfortunately, these plots ended up poor visually. There was also limited movement in the metrics. We could speculate that maybe those with the highest and lowest freedom scores are most "stable" in regards to the operations of their current regime form. A notable exception among those countries scoring "high" in freedom is the U.S., in which this metric shows a decline in freedom starting in 2017.

Overall trends in freedom over time

```
status_colors = c("#56B4E9", "#D55E00", "#000000")

freedom %>%
  filter(!country %in% exclude_list) %>%
  ggplot() +
  aes(x = year, fill = factor(status, levels =c("F", "PF", "NF"))) +
  geom_bar(position = "fill") +
  facet_wrap(
  vars(region_name),
  ncol = 1)  +
  scale_fill_manual(
    name = "",
    values = status_colors,
    breaks = c("F", "PF", "NF"),
```
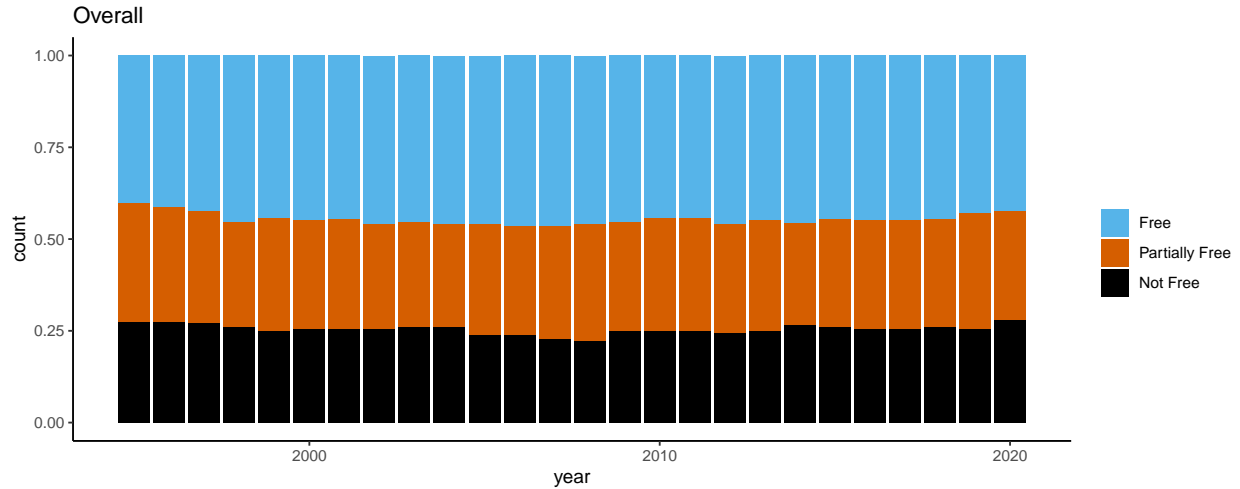
```
        labels = c("Free", "Partially Free", "Not Free")) +
  theme_classic()
```
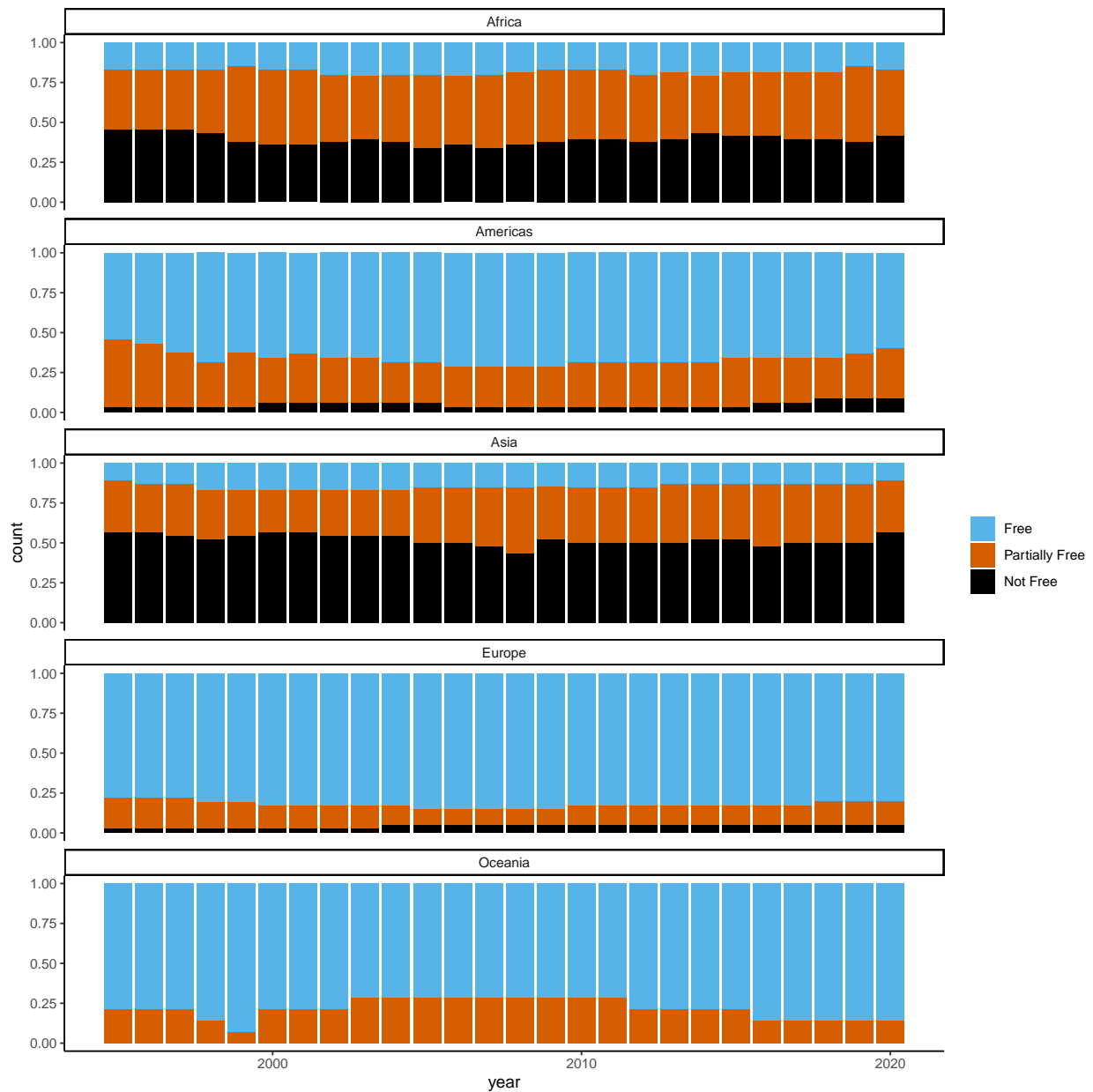


```
freedom %>%
  filter(!country %in% exclude_list) %>%
  ggplot() +
  aes(x = year, fill = factor(status, levels =c("F", "PF", "NF"))) +
  geom_bar(position = "fill") +
  scale_fill_manual(
    name = "",
    values = status_colors,
    breaks = c("F", "PF", "NF"),
    labels = c("Free", "Partially Free", "Not Free")) +
  ggtitle("Overall") +
```

```
theme_classic()
```


Overall

Overall trends in freedom over time by region

```
freedom %>%
  filter(!country %in% exclude_list) %>%
  ggplot() +
  aes(x = factor(year), y = freedom_index) +
  geom_boxplot() +
  facet_wrap(
    vars(region_name),
    ncol = 1) +
  theme_classic()
```
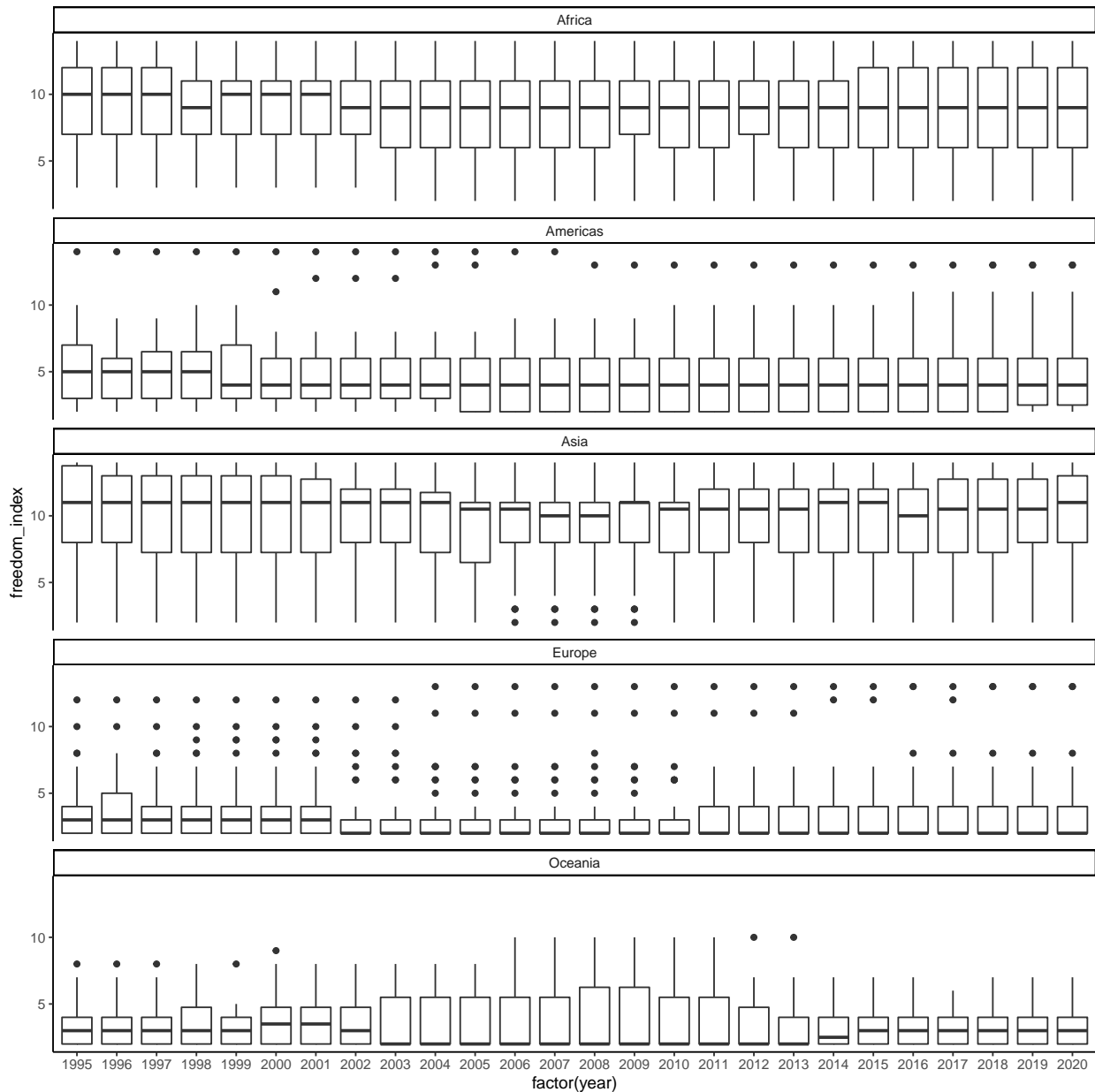
Box plots look quite messy,

We will re-evaluate with line charts as done in David Robinson's example:

```
freedom_gathered <- freedom %>%
  gather(key = metric, value = value, civil_liberties, political_rights) %>%
  mutate(metric = str_to_title(str_replace_all(metric, "_", " ")),
         region_name = fct_reorder(region_name, value))

overall <- freedom_gathered %>%
  group_by(year, metric) %>%
  summarize(avg_rating = mean(value))
```

`summarise()` has grouped output by 'year'. You can override using the `.groups` argument.

```
freedom_gathered %>%
  # mutate(region_name = "Overall") %>%
  # bind_rows(freedom_gathered) %>%
  group_by(year, region_name, metric) %>%
  summarize(avg_rating = mean(value)) %>%
  ggplot(aes(year, avg_rating)) +
  geom_line(size = 1, aes(color = region_name)) +
  geom_line(data = overall, size = 2) +
  facet_wrap(~ metric) +
  expand_limits(y = 1) +
  scale_x_continuous(name = "Year") +
  scale_y_reverse(name = "World Freedom Index", breaks = seq(1,7)) +
  scale_color_discrete(name = "Region", guide = guide_legend(reverse = TRUE)) +
  ggtitle("Change of Freedom Index over time by Region", subtitle = "Black Shows Overall")
```

`summarise()` has grouped output by 'year', 'region_name'. You can override
using the `.groups` argument.

Change of Freedom Index over time by Region
Black Shows Overall

Additional Analysis joined with other data sets

We will perform some additional analysis by pulling in other data sets.

Worldbank Data

More information available at The World Bank.

```
#WDIsearch("gdp") %>% View()

gdp_percap <- WDI(indicator = "NY.GDP.PCAP.CD",
                  extra = TRUE,
                  start = 1995,
                  end = 2020) %>%
```

```r
  as_tibble()

# gdp_percap %>% View()

# freedom %>% distinct(country) %>% anti_join(gdp_percap, by = "country") %>% View()

# gdp_percap %>% filter(str_detect(country, "Egypt"))
```

```r
freedom_join <- freedom %>%
  inner_join(gdp_percap,
             by = c(country_code = "iso2c", "year"),
             suffix = c("", "_wdi")) %>%
  mutate(income = fct_relevel(income, c("Low income", "Lower middle income", "Upper middle income")))

freedom_join %>%
  filter(year == 2020, income != "Not classified") %>%
  group_by(income) %>%
  summarize_freedom() %>%
  arrange(income)
```

```
# A tibble: 4 x 5
  income              n_countries avg_civil_lib avg_pol_rights pct_free
  <fct>                     <int>         <dbl>          <dbl>    <dbl>
1 Low income                   27          5.30           5.56    0
2 Lower middle income          54          4.02           4.15    0.222
3 Upper middle income          53          3.45           3.62    0.377
4 High income                  58          1.91           1.95    0.845
```

```r
freedom_join_gathered <- freedom_join %>%
  filter(income != "Not classified") %>%
  gather(key = metric, value = value, civil_liberties, political_rights) %>%
  mutate(metric = str_to_title(str_replace_all(metric, "_", " ")))

overall_join <- freedom_join_gathered %>%
  group_by(year, metric) %>%
  summarize(avg_rating = mean(value))
```

`summarise()` has grouped output by 'year'. You can override using the `.groups`
argument.

```r
freedom_join_gathered %>%
  # mutate(region_name = "Overall") %>%
  # bind_rows(freedom_gathered) %>%
  group_by(year, income, metric) %>%
  summarize(avg_rating = mean(value)) %>%
  ggplot(aes(year, avg_rating)) +
  geom_line(size = 1, aes(color = income)) +
  geom_line(data = overall_join, size = 2) +
  facet_wrap(~ metric) +
  expand_limits(y = 1) +
  scale_x_continuous(name = "Year") +
  scale_y_reverse(name = "World Freedom Index", breaks = seq(1,7)) +
  scale_color_discrete(name = "World Bank Income", guide = guide_legend(reverse = TRUE)) +
  ggtitle("Change of Freedom Index over time by Income", subtitle = "Black Shows Overall")
```

```
`summarise()` has grouped output by 'year', 'income'. You can override using the
`.groups` argument.
```

Change of Freedom Index over time by Income
Black Shows Overall
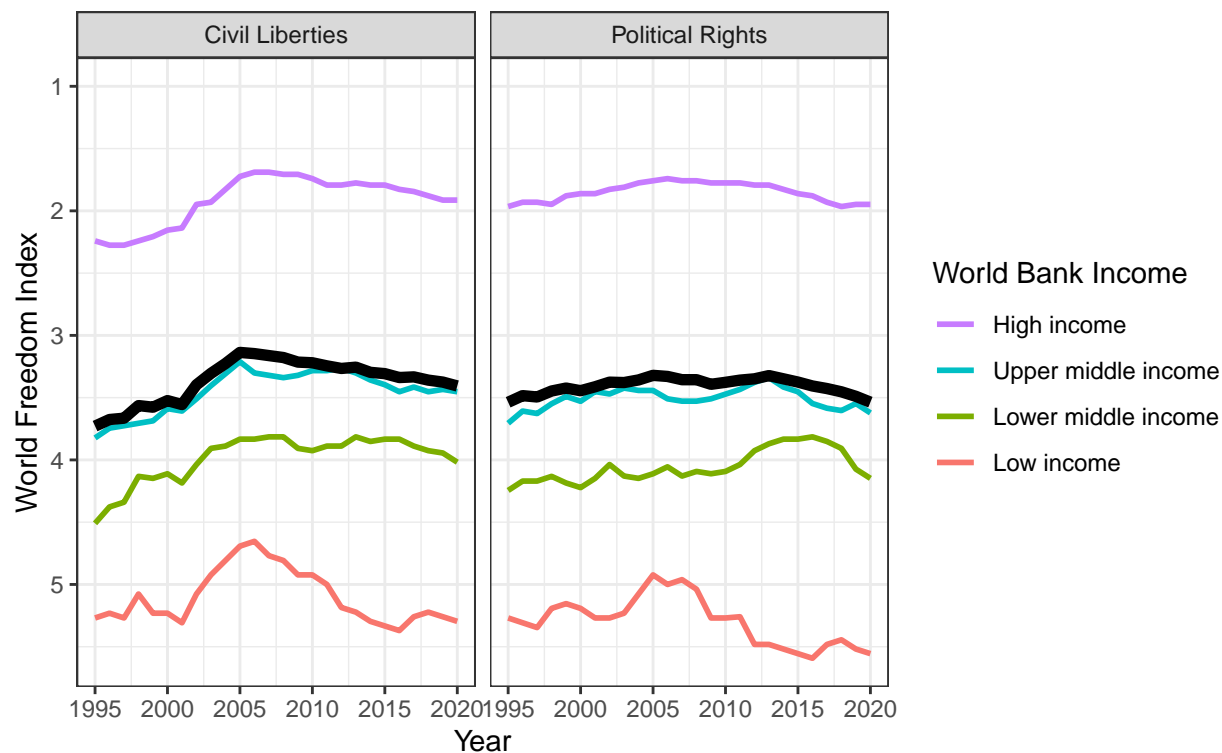
How do we find the biggest outliers?

```
freedom_join_gathered
```

```
# A tibble: 9,906 x 22
   country      year status region_code region_name is_ldc country_code ldc
   <chr>       <dbl> <chr>        <dbl> <chr>        <dbl> <chr>        <chr>
 1 Afghanistan  1995 NF             142 Asia             1 AF           Yes
 2 Afghanistan  1996 NF             142 Asia             1 AF           Yes
 3 Afghanistan  1997 NF             142 Asia             1 AF           Yes
 4 Afghanistan  1998 NF             142 Asia             1 AF           Yes
 5 Afghanistan  1999 NF             142 Asia             1 AF           Yes
 6 Afghanistan  2000 NF             142 Asia             1 AF           Yes
 7 Afghanistan  2001 NF             142 Asia             1 AF           Yes
 8 Afghanistan  2002 NF             142 Asia             1 AF           Yes
 9 Afghanistan  2003 NF             142 Asia             1 AF           Yes
10 Afghanistan  2004 NF             142 Asia             1 AF           Yes
# ... with 9,896 more rows, and 14 more variables: freedom_index <dbl>,
#   country_wdi <chr>, NY.GDP.PCAP.CD <dbl>, status_wdi <chr>,
#   lastupdated <chr>, iso3c <chr>, region <chr>, capital <chr>,
#   longitude <chr>, latitude <chr>, income <fct>, lending <chr>, metric <chr>,
#   value <dbl>
```

```
civil_liberties_2020 <- freedom_join_gathered %>%
    filter(!is.na(NY.GDP.PCAP.CD),
```

```
        metric == "Civil Liberties",
        year == 2020)

lin_mod <- civil_liberties_2020 %>%
  lm(value ~ region_name + log10(NY.GDP.PCAP.CD), data = .)

lin_mod %>%
  augment(data = civil_liberties_2020) %>%
  select(country, region_name, NY.GDP.PCAP.CD, income, value, .fitted, .resid) %>%
  arrange(desc(abs(.resid))) %>%
  head(20) %>%
  ggplot(aes(.fitted, value)) +
  geom_point() +
  geom_abline(color = "red") +
  geom_text_repel(aes(label = country)) +
  scale_x_reverse(name = "Model Civil Liberty") +
  scale_y_reverse(name = "Actual Civil Liberty") +
  expand_limits(x = 1, y = 1) +
  ggtitle("Outliers for Civil Liberties", subtitle = "Lower Number is Better")
```
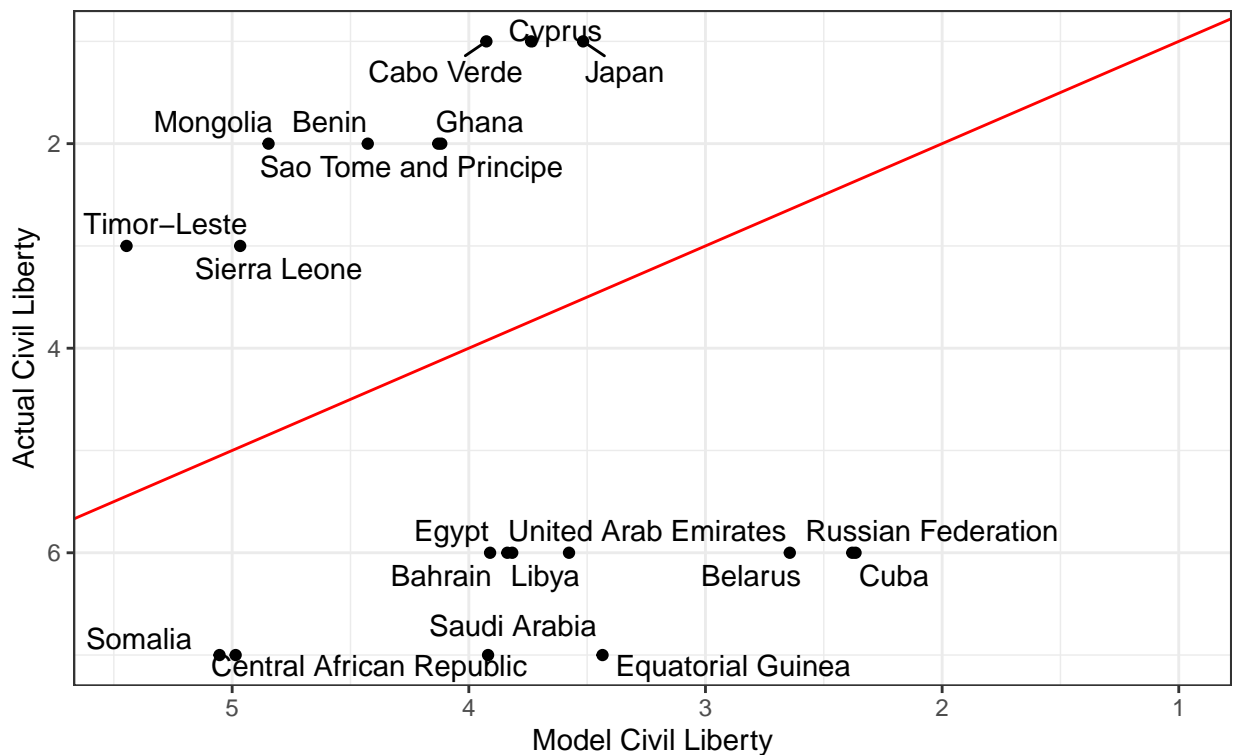
Outliers for Civil Liberties
Lower Number is Better

Map Visual

```
freedom_2020 <- freedom_join_gathered %>%
  filter(year == 2020,
         metric == "Civil Liberties")
```

```
# maps::iso3166

world_map_freedom_2020 <- map_data("world") %>%
  as_tibble() %>%
  regex_left_join(maps::iso3166, c(region = "mapname")) %>%
  left_join(freedom_2020 %>% select(-region), by = c(a2 = "country_code")) %>%
  filter(region != "Antarctica")

world_map_freedom_2020 %>%
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = value)) +
  coord_fixed(1.5) +
  scale_fill_continuous_diverging(palette = "Blue-Red", mid = 3.5, name = "Freedom Index: Civil Libertic
  ggtitle("Civil Liberties Index", subtitle = "In 2020") +
  ggthemes::theme_map()
```

Civil Liberties Index
In 2020
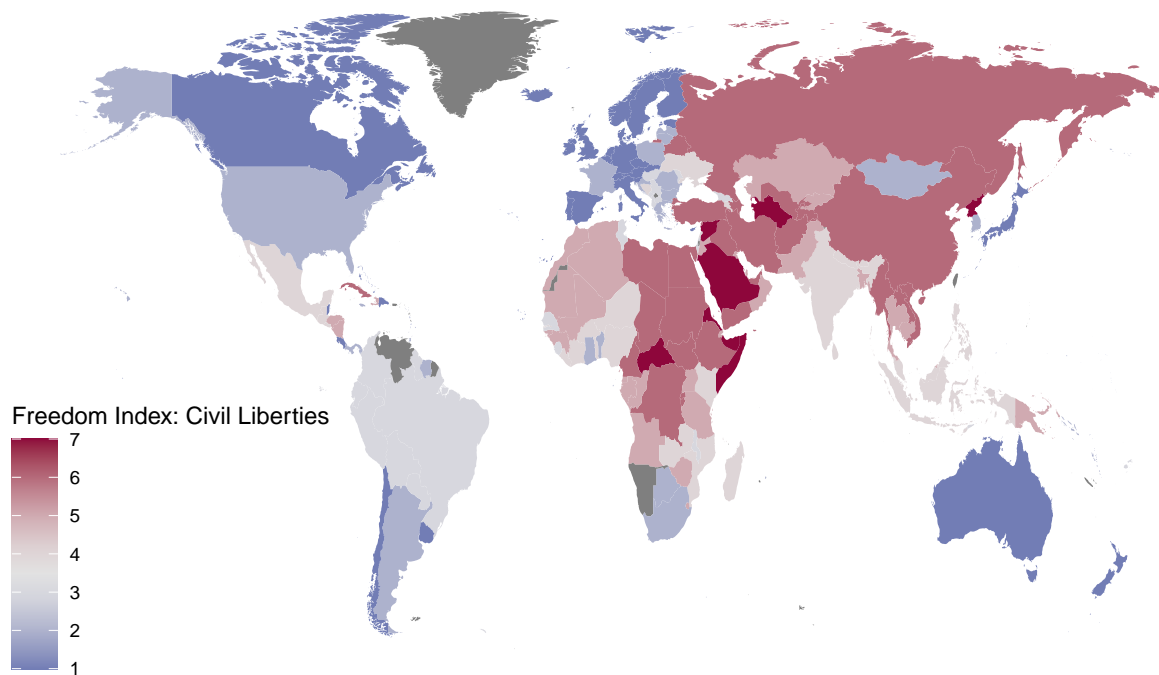
Conclusion

This data set was fairly clean and had a large amount of pre-processing completed on it. My impression is
that it would have been better for Country-level analysis if the data set had supplied the raw indicator scores
which composed the civil_liberties and political_rights metrics. As can be seen in the "Top 5" and "Worst 5"
bar charts, the level of granularity provided gives minimal year by year changes. Region level analysis has
more apparent trends when examining both by region and changes over time. Lastly, by bringing in economic
indicators, we can examine the correlation between freedom and income.