Freedom in Asia

An Analysis in R

Benjamin Absalon

```
freedom <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/maste
r/data/2022/2022-02-22/freedom.csv')

freedom
```

```
## # A tibble: 4,979 × 8
##    country      year    CL    PR Status Region_Code Region_Name is_ldc
##    <chr>       <dbl> <dbl> <dbl> <chr>        <dbl> <chr>        <dbl>
##  1 Afghanistan  1995     7     7 NF             142 Asia             1
##  2 Afghanistan  1996     7     7 NF             142 Asia             1
##  3 Afghanistan  1997     7     7 NF             142 Asia             1
##  4 Afghanistan  1998     7     7 NF             142 Asia             1
##  5 Afghanistan  1999     7     7 NF             142 Asia             1
##  6 Afghanistan  2000     7     7 NF             142 Asia             1
##  7 Afghanistan  2001     7     7 NF             142 Asia             1
##  8 Afghanistan  2002     6     6 NF             142 Asia             1
##  9 Afghanistan  2003     6     6 NF             142 Asia             1
## 10 Afghanistan  2004     6     5 NF             142 Asia             1
## # ℹ 4,969 more rows
```

Data set appeared in Tidy Tuesday on 2022-02-22. See here (https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-02-22/readme.md) for data dictionary and more details.

Question: How has freedom changed over the years in Asia? Can freedom status in Asia be predicted from features in the data?

Introduction: The `freedom` data set originates from Freedom House and United Nations. Countries are scored on civil liberties (`CL`), political rights (`PR`) and labeled as least developed country (`is_ldc`), freedom status (`Status`) for each year between 1995 and 2020.

Both questions request the data be filtered to Asia. `Region_Name` provides the necessary geographical information. `Status` is a label applied to each country indicating whether the country is free, partially free, or not free, according to the model used by the curators. These two attributes are sufficient to create a geospatial plot. Scores for `CL` and `PR` are integers between 1 and 7, and a binary variable, `is_ldc` , provides numeric features to cluster on for the question of prediction.

Approach: Visualizing the change in freedom over the years is done naturally by generating geospatial plots from distinct time points. In this analysis, freedom is viewed at the earliest year (1995) and latest year (2020) present in the data. Countries that show a change in status between the two periods are highlighted.

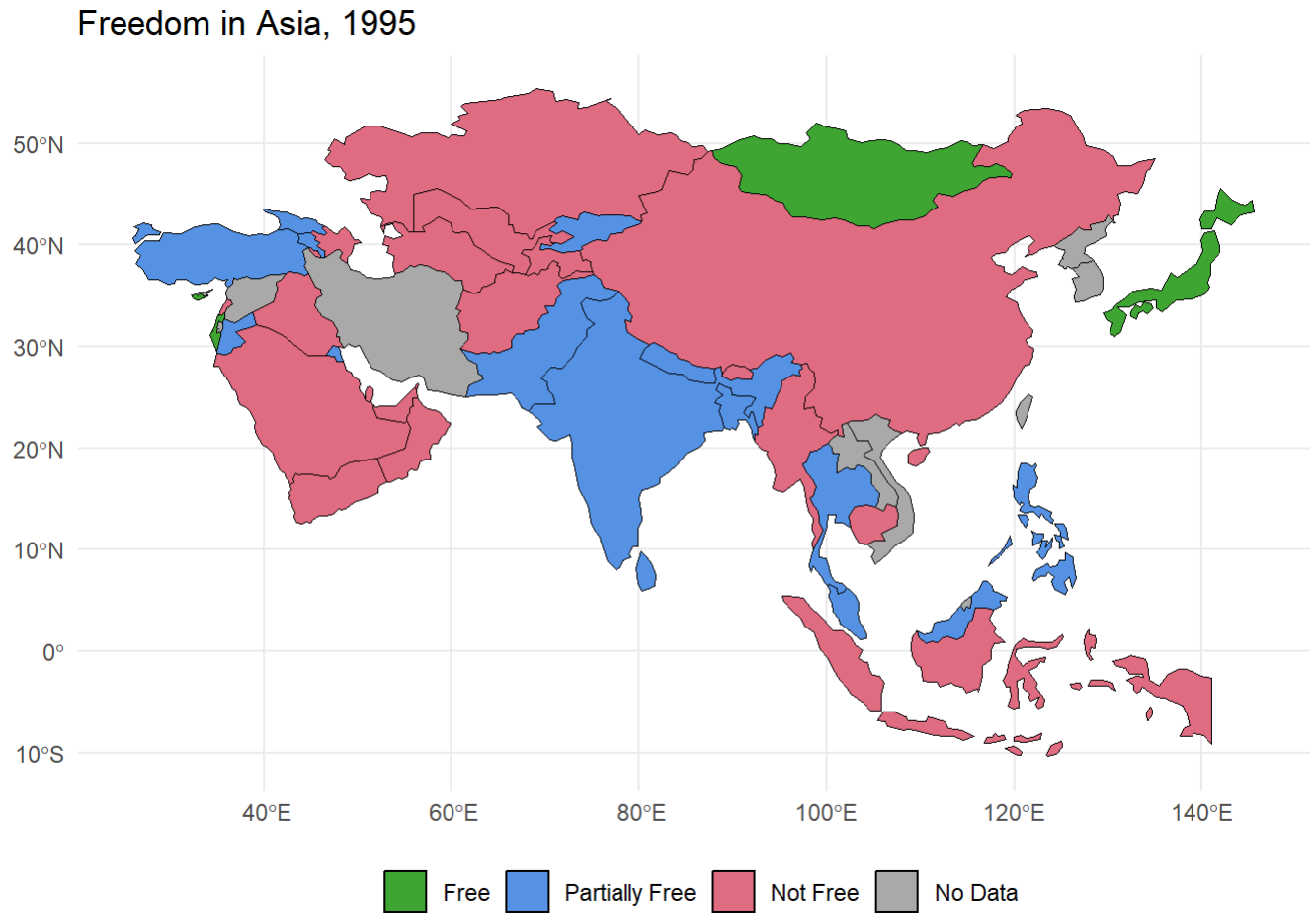
The data set contains two numeric scores and a binary flag. It is assumed the curator's underlying model utilized these variables. Therefore, a clustering analysis is performed in an attempt to recreate statuses. Data has three unique statuses, and consequently, clustering will produce three groups.

Analysis:

```
# general preparation of data set

# merge geometry data and clean up status names
countries <- rnaturalearth::ne_countries(returnclass = "sf")

freedom <- countries %>%
  rename(country = name) %>%
  select(country, region_un, geometry) %>%
  # full join to include countries not in `freedom` with NA years to fill map
  # region_un used for subsetting
  full_join(freedom, by = "country") %>%
  # rename Statuses
  mutate(
    Status = case_when(
    Status == "F" ~ "Free",
    Status == "PF" ~ "Partially Free",
    Status == "NF" ~ "Not Free",
    TRUE ~ NA
    )
  )

# countries with a change in status between 1995 and 2020
status_by_year <- freedom %>%
  data.frame() %>%
  pivot_wider(
    id_cols = country,
    names_from = year,
    values_from = Status
  ) %>%
  mutate(change = `1995` != `2020`)

countries_with_change <- status_by_year %>%
  filter(change == TRUE) %>%
  pull(country) %>%
  unique()

freedom <- freedom %>%
  mutate(change_1995_to_2020 = country %in% countries_with_change)

head(freedom)
```

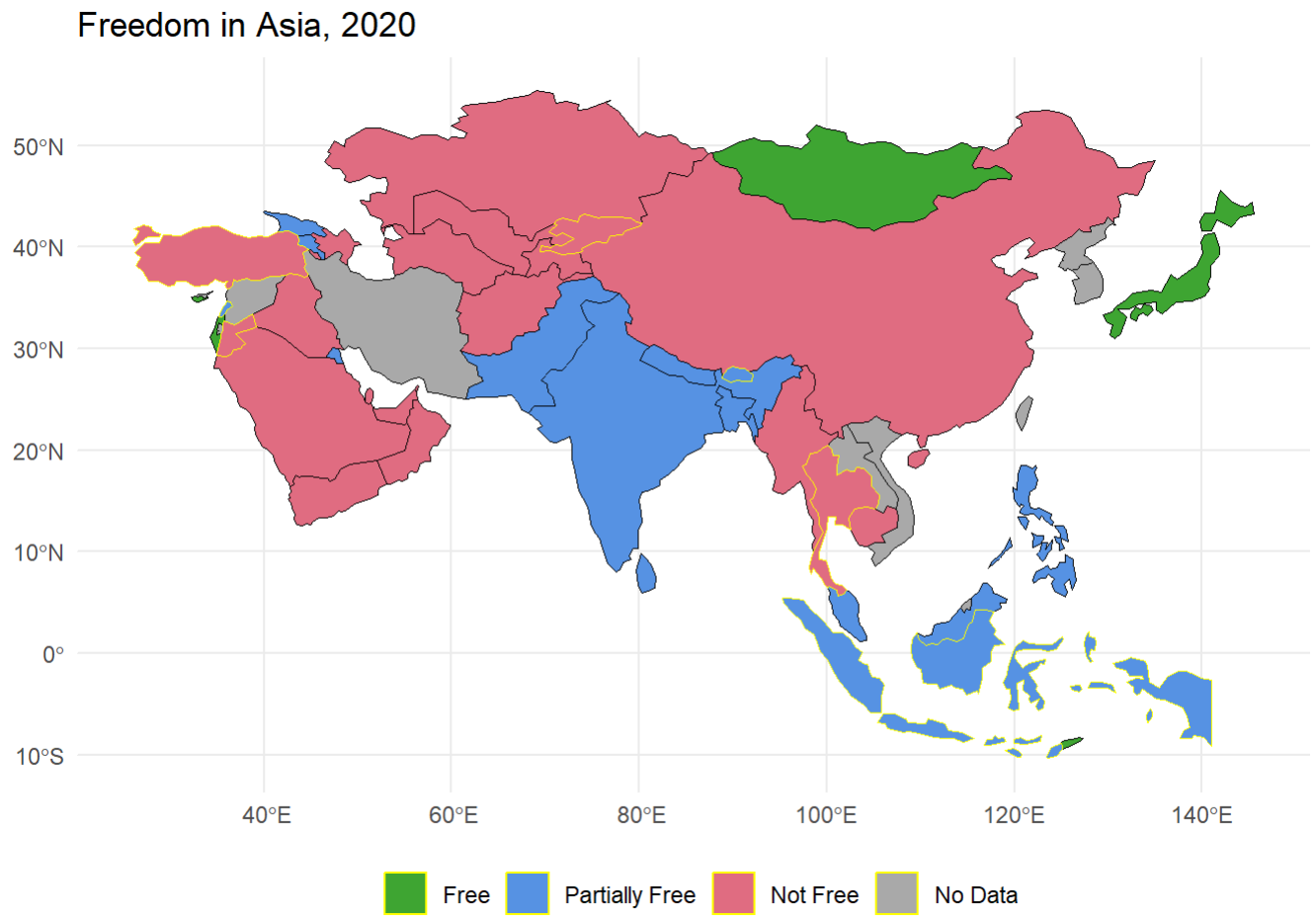
```
## Simple feature collection with 6 features and 10 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -180 ymin: -18.28799 xmax: 180 ymax: -16.02088
## Geodetic CRS:  WGS 84
##   country region_un year CL PR         Status Region_Code Region_Name is_ldc
## 1    Fiji   Oceania 1995  3  4 Partially Free           9     Oceania      0
## 2    Fiji   Oceania 1996  3  4 Partially Free           9     Oceania      0
## 3    Fiji   Oceania 1997  3  4 Partially Free           9     Oceania      0
## 4    Fiji   Oceania 1998  3  4 Partially Free           9     Oceania      0
## 5    Fiji   Oceania 1999  3  2           Free           9     Oceania      0
## 6    Fiji   Oceania 2000  3  6 Partially Free           9     Oceania      0
##                          geometry change_1995_to_2020
## 1 MULTIPOLYGON (((180 -16.067...               FALSE
## 2 MULTIPOLYGON (((180 -16.067...               FALSE
## 3 MULTIPOLYGON (((180 -16.067...               FALSE
## 4 MULTIPOLYGON (((180 -16.067...               FALSE
## 5 MULTIPOLYGON (((180 -16.067...               FALSE
## 6 MULTIPOLYGON (((180 -16.067...               FALSE
```

```
americas_plot <- function(yr, reg) {
  # reusable function for plots
  # yr : year
  # reg : Region_Name
  temp <- freedom %>%
    filter(
      year == yr | is.na(year),
      region_un == reg
    ) %>%
    replace_na(list(Status = "No Data"))

  temp %>%
    ggplot() +
    geom_sf(aes(fill = Status), color = "black") +
    labs(title = glue::glue("Freedom in {reg}, {yr}")) +
    # PSAD56 projection
    coord_sf(crs = 4248) +
    scale_fill_manual(
      breaks = c("Free", "Partially Free", "Not Free", "No Data"),
      values = c("#3EA532", "#5692E3", "#E06C81", "#A9A9A9")
    ) +
    theme_minimal() +
    theme(
      legend.position = "bottom",
      legend.title = element_blank()
    )
}

americas_plot(1995, "Asia")
```



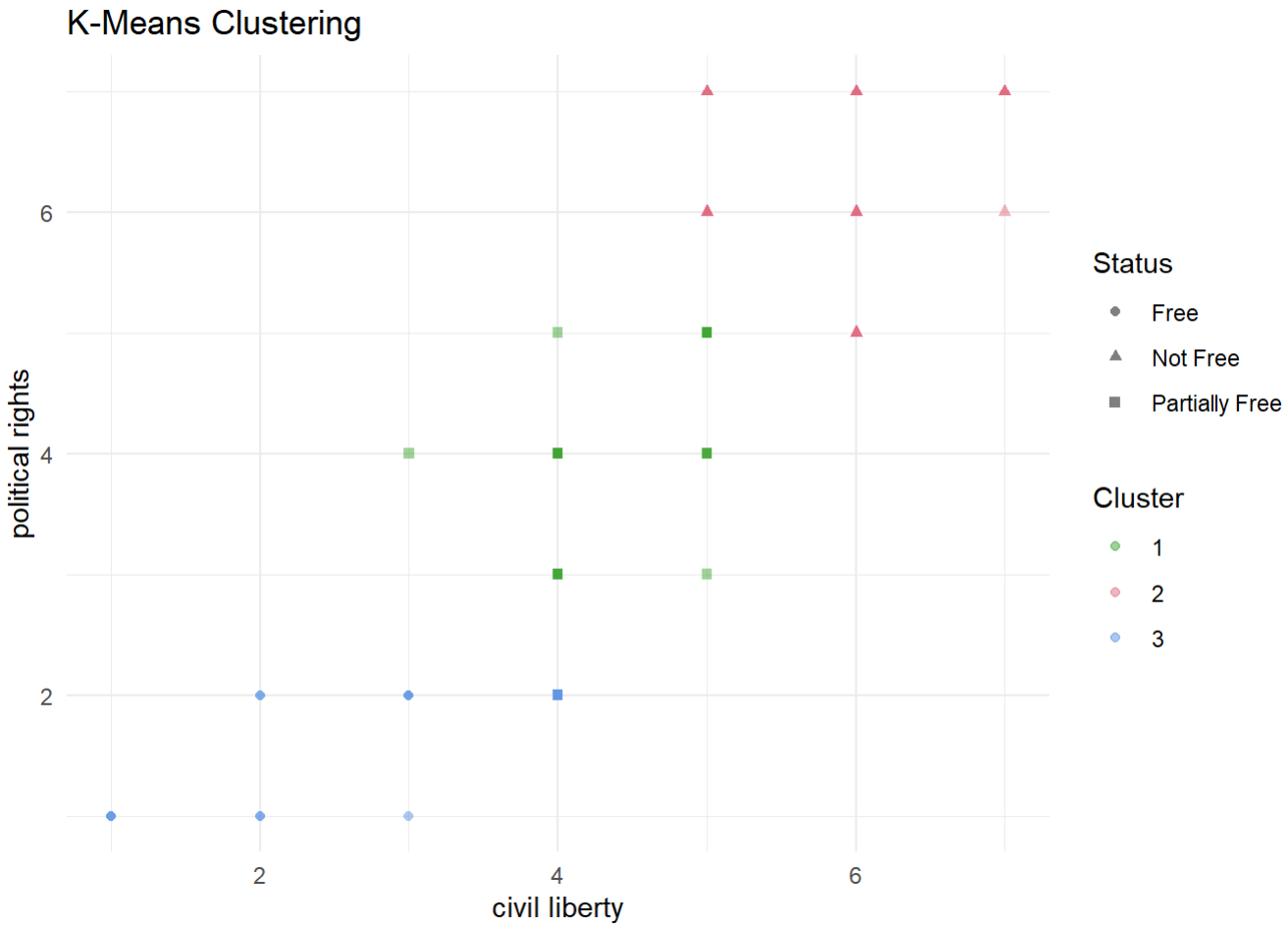

```
americas_plot(2020, "Asia") +
  # highlight borders of countries that changed statuses
  geom_sf(
    data = filter(freedom, Region_Name == "Asia", year == 2020, change_1995_to_2020 == TRUE),
    aes(fill = Status, color = change_1995_to_2020), color = "yellow"
  )
```

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```



Eight countries changed statuses between 1995 and 2020: Turkey, Lebanon, Jordan, Kyrgyzstan, Maldives, Bhutan, Thailand, and Indonesia.

```r
# clustering of original data

# set seed so that output of k-means matches PDF output
set.seed(124242)

asia <- freedom %>%
  data.frame() %>%
  filter(Region_Name == "Asia", year %in% c(1995, 2020))

# fit the model
km_fit <- asia %>%
  select(CL, PR, is_ldc) %>%
  kmeans(centers = 3, nstart = 10)

# augment and plot results
km_fit %>%
  augment(asia) %>%
  ggplot(aes(CL, PR, color = .cluster, shape = Status)) +
  geom_point(alpha = .5) +
  labs(
    title = "K-Means Clustering",
    x = "civil liberty", y = "political rights"
  ) +
  scale_color_manual(
    values = c("#3EA532", "#E06C81", "#5692E3"),
    name = c(".cluster" = "Cluster")
  ) +
  theme_minimal()
```



`CL` and `PR` being integers creates a structured grid of points.

Discussion: Freedom status in Asia has not changed dramatically between 1995 and 2020. Eight countries experienced a change in status, evenly split between changing to Not Free and Partially Free. The periods observed were after the end of the Cold War–a stable time frame compared to historical geopolitics.

Clustering showed a nearly perfect result. One coordinate pair labelled Partially Free was clustered with Free. Not Free and the rest of Partially Free were isolated in their respective clusters. Knowing the number of clusters needed beforehand assisted in the outcome. Additionally, clustering provided strong results likely because the curators' modeled freedom on the same three variables.