Data Building

Stepan Polikanov
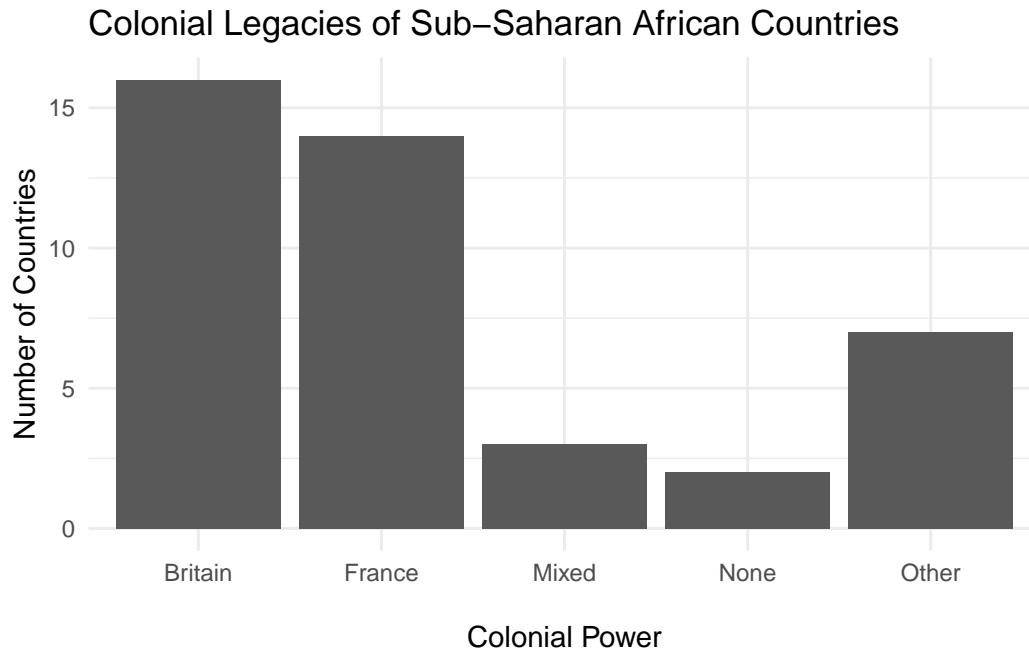
I start with reading in the data.

```
party_class <- read_excel(here("data", "data_raw", "partyreg.xlsx"))
colonial <- read_csv(here("data", "data_raw", "COLDAT_colonies.csv"))
vdem <- readRDS(here("data", "data_raw", "V-Dem-CY-Full+Others-v10.rds"))


cy <- str_c(party_class$ccodecow, party_class$year, sep = "_")
```

I construct dummies for the colonial legacies of sub-Saharan African countries from the Colonial
Dates dataset (Becker, 2019).

```
col_legacy <- colonial |>
  mutate(ccodecow = countryname(country, destination = "cown")) |>
  filter(ccodecow %in% party_class$ccodecow) |>
  reframe(ccodecow, col.britain, col.france,
          col.other = if_else((col.britain == 0 & col.france == 0)
                              & (col.belgium == 1 | col.germany == 1
                                 | col.italy == 1 | col.netherlands == 1
                                 | col.portugal == 1 | col.spain == 1), 1, 0),
          col.none = if_else(col.other == 0 & col.britain == 0
                             & col.france == 0, 1, 0),
          col.type = case_when(col.britain == 1 & col.france == 0
                               & col.other == 0 ~ "Britain",
                               col.france == 1 & col.britain == 0
                               & col.other == 0 ~ "France",
                               col.other == 1 & col.britain == 0
                               & col.france == 0 ~ "Other",
                               col.none == 1 ~ "None",
                               .default = "Mixed"))

ggplot(col_legacy, aes(x = col.type)) +
  geom_bar() +
  labs(title = "Colonial Legacies of Sub-Saharan African Countries",
       x = "\nColonial Power",
       y = "Number of Countries")
```

Colonial Legacies of Sub–Saharan African Countries

The result is that some countries had colonial histories with both Britain and France, either of those a different metropole, or were not colonized at all.

Party system institutionalization data comes from V-Dem v10 (march 2020) (Coppedge et al., 2020).

```
psi <- vdem |>
  select(country_name, ccowcode = COWcode, year, v2xps_party) |>
  mutate(v2xps_party_fill = if_else(is.na(v2xps_party)
                                    & str_c(ccowcode, year, sep = "_") %in% cy,
                                    round(mean(v2xps_party[year >= 2013 & year <= 2018],
                                          na.rm = T), 3), v2xps_party)) |>
  filter(str_c(ccowcode, year, sep = "_") %in% cy)
```

There are two instances when the PSI index is not recorded by V-Dem in the delineated geographical and temporal scope: Mali in 2013 and Guinea in 2013. I fill these with the mean of the PSI index for these countries for the years 2013-2018. This may be suboptimal, so I also provide a sensitivity analysis to this choice.