

# EDA for HIV prevalence survey data

Adam Howes

June 2020

```
source("../src/00_setup.R")
```

```
df <- readRDS("../data/all.rds")
```

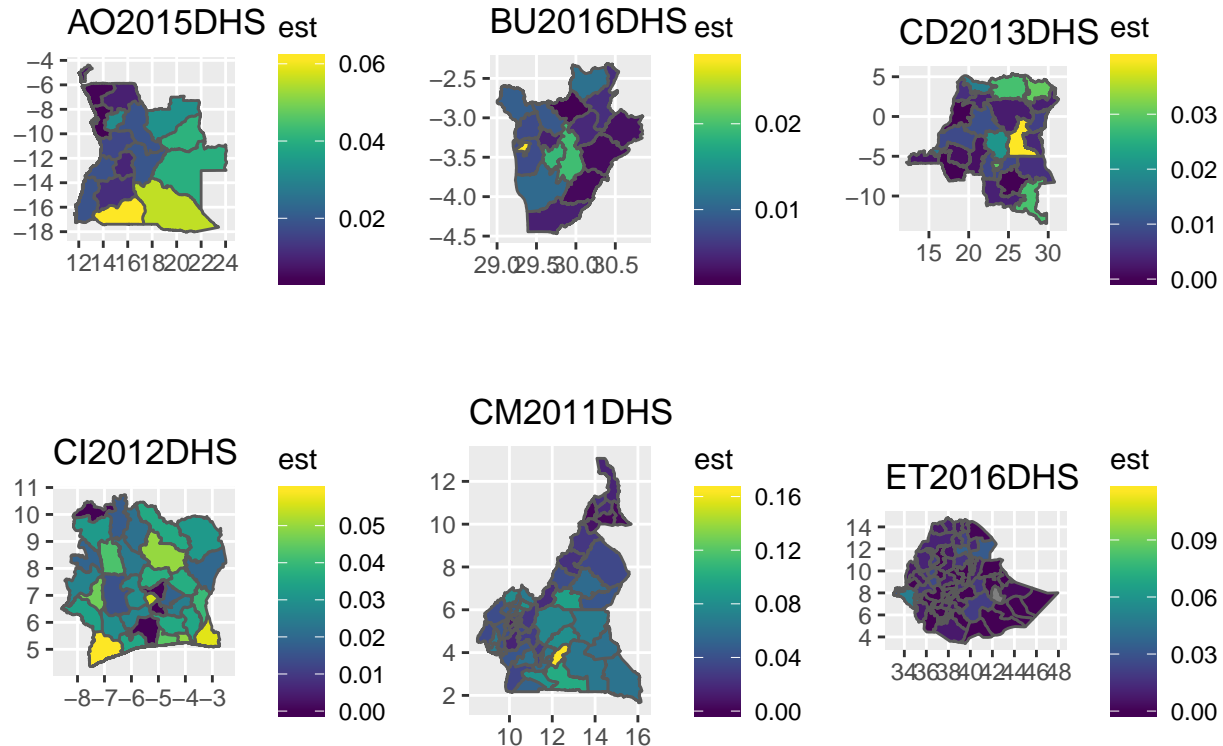
Get the most recent surveys from each country:

```
recent <- c("AO2015DHS", "BU2016DHS", "CD2013DHS", "CI2012DHS", "CM2011DHS",  
            "ET2016DHS", "GH2014DHS", "KE2008DHS", "LS2014DHS", "MW2015DHS",  
            "MZ2015AIS", "NM2013DHS", "RW2015DHS", "SZ2006DHS", "TZ2012AIS",  
            "UG2011AIS", "ZA2016DHS", "ZM2013DHS", "ZW2015DHS")
```

```
est_plot <- function(survey) {  
  df %>%  
    filter(survey_id == survey) %>%  
    ggplot(aes(fill = est)) +  
    geom_sf(aes(geometry = geometry)) +  
    coord_sf() +  
    scale_fill_viridis() +  
    labs(title = paste(survey))  
}
```

For example, plotting the raw estimates of prevalence from the first six surveys:

```
plots <- lapply(recent[1:6], est_plot)  
plot_grid(plotlist = plots)
```



Visualisation of the graph that the adjacency matrix is based on for Malawi:

```
mw <- df %>%  
  filter(survey_id == "MW2015DHS")  
  
mw_sp <- mw %>%  
  as("Spatial")  
  
nb <- poly2nb(mw_sp) # Spatial neighbours object  
  
nb_sf <- nb2lines(nb, coords = coordinates(mw_sp)) %>%  
  as("sf") %>%  
  st_set_crs(st_crs(mw))  
  
ggplot(mw) +  
  geom_sf(fill = "lightgrey", color = "white") +  
  geom_sf(data = nb_sf) +  
  labs(x = "Longitude", y = "Latitude")
```

