Problem set 1: Question 2

Latera Tesfaye Olana

16 January, 2023

Plot:

I spent 46 minutes on this question and below is my attempt - around **15 mins** trying to get best vjust and hjust for better readability - a price for poor programming practice :). I challenged my self to write the code in one sit, without referencing anything, just using what I previously know about ggplot(). I admit, I cheated by reading the documentations just a $_{tiny}$ bit. I couldn't get the coloring of legend texts right. I will keep exercising on this :)

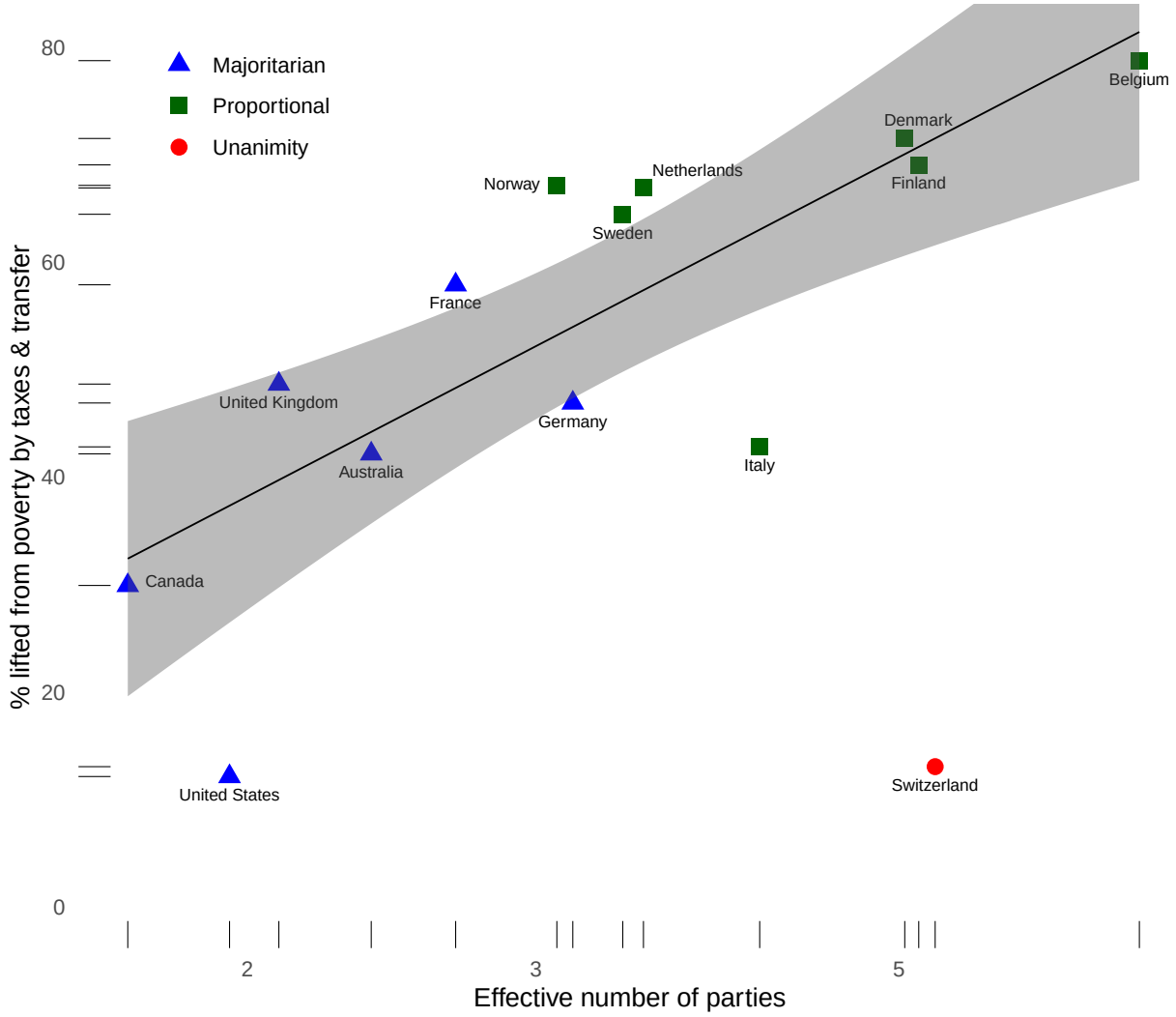

Figure 1: More detailed statistical scatter plot

Supplementary results - Tables

Table 1: Model summary

	Estimate	Naive SE	Robust SE	95%L	95%H	t value	Pr(>|t|)
(Intercept)	1.982	1.0272	1.4096	-1.0892	5.0532	1.4061	0.1851
povertyReduction	0.032	0.0188	0.0254	-0.0233	0.0872	1.2590	0.2320

Code Appendix

```
### Setting up the packages
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
# check if packages are installed; if not, install them
packages <- c("dplyr", "readr", "ggExtra", "plotly",
              "ggplot2","ggstatsplot","ggside","rigr","nlme","lmtest",
              "sandwich","hrbrthemes","MASS")
not_installed <- setdiff(packages, rownames(installed.packages()))
if (length(not_installed)) install.packages(not_installed)

# load packages
library("MASS")
library(dplyr)
library(sandwich)
library(readr)
library(lmtest)
library(nlme)
library(ggstatsplot)
library(ggside)
library(rigr)
library(hrbrthemes)
library(ggExtra)
library(plotly)
library(ggplot2)
# library(tidyverse) # don't load tidyverse package due to conflict with dplyr
### ------------------------------------------------------------
#Loading working directory of the raw data

#Please load your data/directory by changing it with your work directory
#Throughout this code module you will see a tone of places, where
#data is read and written, so please make sure to change them to your
#working directory folder format

working_directory_data <- setwd("C:/Users/latera/Desktop/viz/")

wcgs <- read_csv("data/iverRevised.csv")
### ------------------------------------------------------------
### Q2
# fitting: linear regression

weight_lm <- regress("mean", effectiveParties ~ povertyReduction, data = wcgs)
coef(weight_lm)[,c('Estimate','Naive SE', 'Robust SE','95%L',
                   '95%H','Pr(>|t|)')]
as.data.frame(coef(weight_lm)[,c('Estimate','Naive SE',
                   'Robust SE','95%L','95%H','Pr(>|t|)')])

coef(weight_lm)

h_just <- c(0.5, 0.5, -0.3, 0.3, 0.5, 0.5, 0.5,  0.5,-0.1, 1.3, 0.5, 0.5, 0.5, 0.5)
v_just <- c(2.0, 2.0, 0.1, -1.1, 2.0, 2.0, 2.0, 2.0,-1.0, 0.3, 2.0, 2.0,2.0, 2.0)
#Plotting detailed scatter plot
```

```
p<-ggplot(wcgs, aes(x=effectiveParties, y=povertyReduction)) +
  xlab("Effective number of parties ")+
  ylab("% lifted from poverty by taxes & transfer")+
  geom_point(mapping = aes(shape = partySystem, color = partySystem),size=3)+
  scale_x_continuous(trans='log10')+
  geom_text(aes(label=country), hjust = h_just, vjust = v_just, size=2.5)+

  scale_shape_manual(values = c("Unanimity" = 16, 'Majoritarian' = 17,
                                'Proportional' = 15))+
  scale_color_manual(breaks=c('Majoritarian', 'Proportional', 'Unanimity'),
                     values=c('Majoritarian'='blue', 'Proportional'='darkgreen',
                              'Unanimity'='red'))+
  geom_rug(col="black",linewidth=0.20)+
  coord_cartesian(ylim = c(0,80))+
  theme_bw() +
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())

p <- p + theme(legend.position = c(0.15, 0.9))+
  theme(legend.background = element_rect
        (fill = "transparent"))+
  theme(legend.title = element_blank())

p+stat_smooth(method=function(formula,data,weights=weight) rlm(formula,
                                                               data,
                                                               weights=weight,
                                                                method="MM"),
              fullrange=TRUE,color="black", fill="#555555", level=0.95,linewidth = 0.35)



#Generating tables
kable(coef(weight_lm)%>%round(4), caption = "Model summary")
```