Homework 8

201982188 통계학과 박현주

2020-11-17

Your turn - HW1.

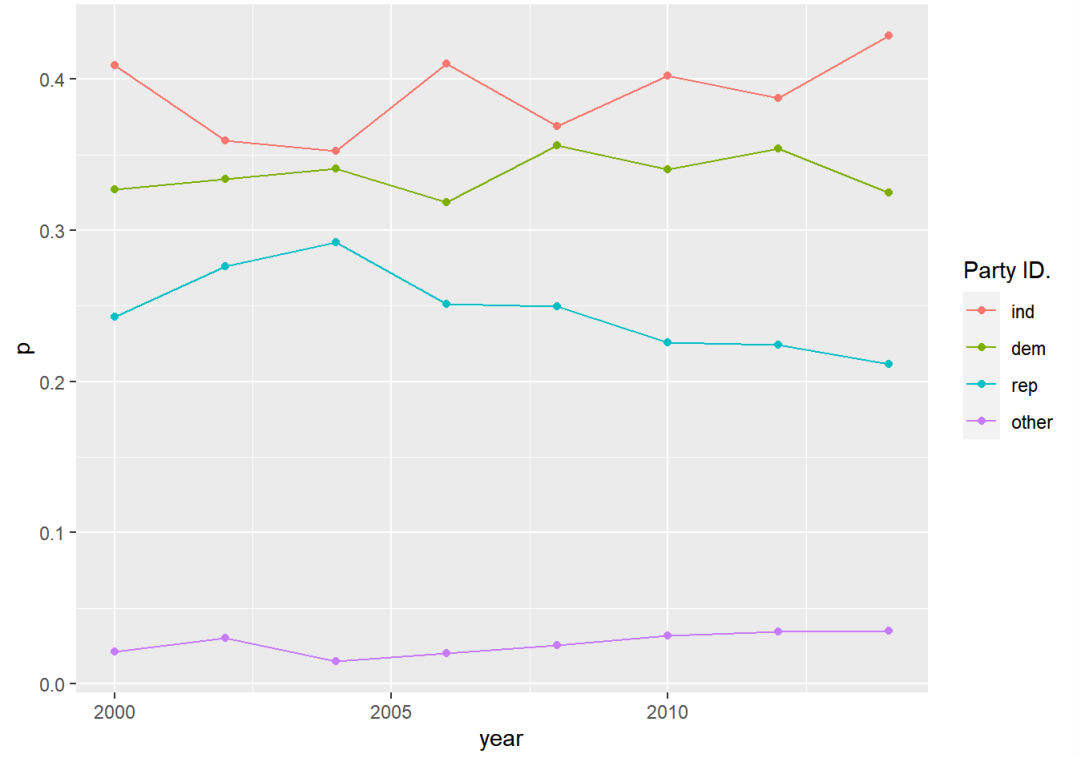
1-1.

How have the proportions of people identifying as Democrat, Republican, and Independent changed over time? Reproduce the following graph.

```
party_id_cat <- function(x){
  if(str_detect(x,'Ind')){
    text <- 'ind'
  }else if(str_detect(x, 'republican')){
    text <- 'rep'
  }else if(str_detect(x,'democrat')){
    text <- 'dem'
  }else{
    text <- 'other'
  }
  return(text)
}
```

```
"%notin%" <- Negate("%in%")
```

```
gss_cat %>%
  mutate(Party_ID = partyid %>% lapply(.,party_id_cat) %>% unlist()) %>%
  group_by(year,Party_ID) %>%
  summarise(n = n()) %>%
  group_by(year) %>%
  mutate(total = sum(n), p = n/total) %>%
  ggplot(aes(x=year, y=p,colour = fct_reorder2(Party_ID, year, p))) +
  geom_point()+geom_line()+
  labs(colour = "Party ID.")
```



1-2.

How could you collapse rincome into a small set of categories? Reproduce the following graph.

```
gss_cat$rincome %>% unique() %>% sort()
```

```
##  [1] No answer      Don't know     Refused        $25000 or more $20000 - 24
##  [6] $15000 - 19999 $10000 - 14999 $8000 to 9999  $7000 to 7999  $6000 to 69
## [11] $5000 to 5999  $4000 to 4999  $3000 to 3999  $1000 to 2999  Lt $1000
## [16] Not applicable
## 16 Levels: No answer Don't know Refused $25000 or more ... Not applicable
```

```r
rincome_cat <- function(x){
  if(x %in% c('$8000 to 9999','$7000 to 7999','$6000 to 6999','$5000 to 5999')){
    text <- "$5000 to 10000"
  }else if(x %in% c('$4000 to 4999','$3000 to 3999','$1000 to 2999')){
    text <- 'Lt $5000'
  }else if(x %in% c('No answer',"Don't know",'Refused','Not applicable')){
    text <- 'Unknown'
  }else if(x == 'Lt $1000'){
    text <- 'Lt $5000'
  }else{
    text <- x
  }

return(text)
}

rincome_levels <- c('Lt $5000','$5000 to 10000','$10000 - 14999',
                    '$15000 - 19999','$20000 - 24999','$25000 or more',
                    'Unknown')[7:1]
gss_cat %>%
  mutate(rincome = as.character(rincome) %>% lapply(.,rincome_cat) %>% unlist()) %>%
  group_by(rincome) %>% summarise(count=n()) %>%
  ggplot(aes(x=factor(rincome,levels = rincome_levels), y=count))+
  geom_bar(stat = 'identity')+
  xlab('rincome')+
  coord_flip()
```