Immigration and

Cheng Ren

a function was written to combine all the data it is too slow to run here

```
#function to paste year
# add_year<-function(name,year){
#   year<-rep(year, 9)
#   ny<-paste(name,year,sep="_")
#   return(ny)
# }
# #extract smaller table
# extr<-function(dataset){
#   colnames(dataset)<-tolower(names(dataset))
#   dataset_st<-dataset[,c('ein','name','address','city','state','zip','totrev','nteefinal','nteefinal1
#                          'cont','secur',"salesexp","invinc","solicit","goods","grprof",
#                          "grrec","exps","fundbal","compens","fundfees","ass_boy","ass_eoy","liab_boy"
#   colnames(dataset_st)<-c('ein','name','address','city','state','zip','totrev','ntee','ntee1',
#                           'tot_pub_sup','secur',"salesexp","invinc","solicit","goods","grprof",
#                           "grrec","exps","fundbal","compens","fundfees","ass_boy","ass_eoy","liab_boy
#   return(dataset_st)
# }
# ## for loop
# filelist<-list.files(pattern = 'csv')
# i=1
# j=1988
# data_com<-data.frame()
# for( i in 1:22){
#   j=j+1
#   data_year<-read.csv(filelist[i])
#   data_year_st<-extr(data_year)
#   data_year_st$year<-j
#   data_com<-rbind(data_year_st,data_com)
#   #assign(paste('data_',j,'_st',sep=''),data_year_st)
# }
# ##name was changed according to IRS
# extr11<-function(dataset){
#   colnames(dataset)<-tolower(names(dataset))
#   dataset_st<-dataset[,c('ein','name','address','city','state','zip','totrev','nteefinal','nteefinal1
#                          'cont','secur',"salesexp","invinc","lessdirfndrsng","goods","grprof",
#                          "grrec","exps","fundbal","compens","fundfees","ass_boy","ass_eoy","liab_boy"
#   colnames(dataset_st)<-c('ein','name','address','city','state','zip','totrev','ntee','ntee1',
#                           'tot_pub_sup','secur',"salesexp","invinc","solicit","goods","grprof",
#                           "grrec","exps","fundbal","compens","fundfees","ass_boy","ass_eoy","liab_boy
#   return(dataset_st)
# }
# ## for loop
# filelist<-list.files(pattern = 'csv')
# i=1
# j=2010
# data_com11<-data.frame()
# for( i in 23:length(filelist)){
```

```
#   j=j+1
#   data_year<-read.csv(filelist[i])
#   data_year_st<-extr11(data_year)
#   data_year_st$year<-j
#   data_com11<-rbind(data_year_st,data_com11)
#   #assign(paste('data_',j,'_st',sep=''),data_year_st)
# }
# #combine
# data_comc <- rbind(data_com11,data_com)
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidycensus)
```

```
## Warning: package 'tidycensus' was built under R version 3.6.2
```

```
library(purrr)
```

```
## Warning: package 'purrr' was built under R version 3.6.1
```

```
library(tidyr)
```

```
setwd('D:/API/bdr/add')
data_comc <- read.csv('data_comc.csv')
```

1.Explore the trends of nonprofit organizations

```
names(data_comc)
```

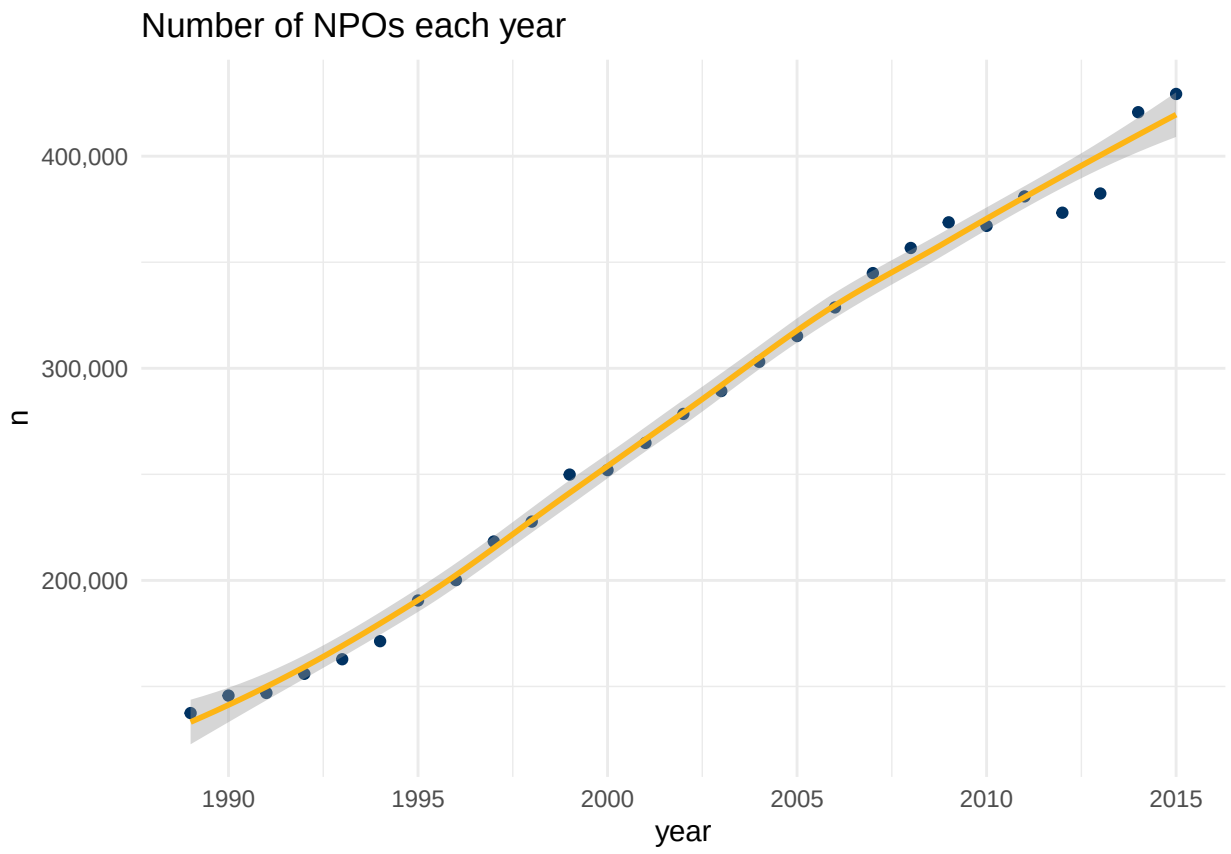
```
##  [1] "ein"           "name"          "address"       "city"
##  [5] "state"         "zip"           "totrev"        "ntee"
##  [9] "ntee1"         "tot_pub_sup"   "secur"         "salesexp"
## [13] "invinc"        "solicit"       "goods"         "grprof"
## [17] "grrec"         "exps"          "fundbal"       "compens"
## [21] "fundfees"      "ass_boy"       "ass_eoy"       "liab_boy"
## [25] "liab_eoy"      "progrev"       "year"          "prg_pre_total"
## [29] "net"           "ast_gain"      "lab_gain"      "gap_end"
```

```
table(data_comc$year)
```

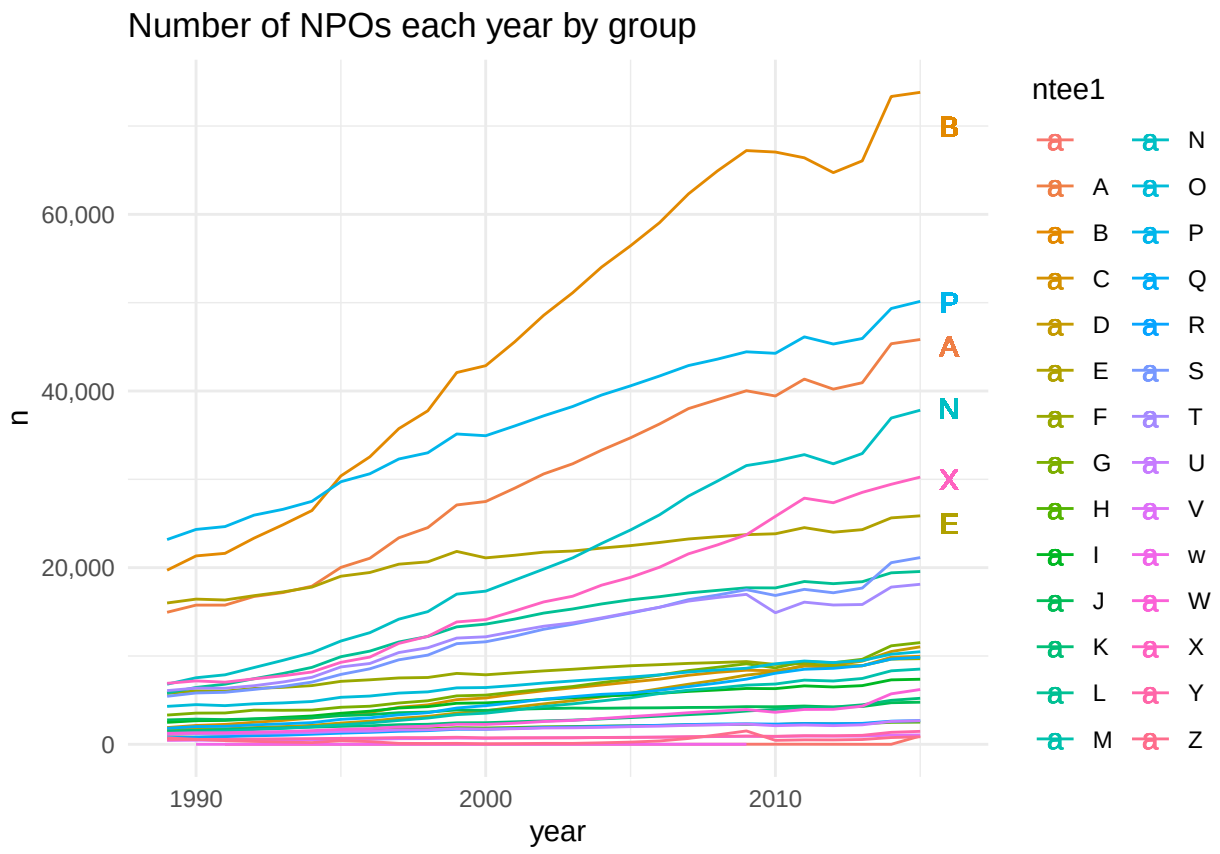
```
##
##   1989   1990   1991   1992   1993   1994   1995   1996   1997   1998
## 137459 145703 146948 155960 162840 171351 190531 200161 218341 227706
##   1999   2000   2001   2002   2003   2004   2005   2006   2007   2008
## 249886 252006 264821 278490 289282 303077 315224 328689 344875 356728
##   2009   2010   2011   2012   2013   2014   2015
## 368816 367146 381035 373358 382401 420735 429338
```

```
#cl<-scale_color_manual(values=c("#FDB515", "#003262"))

increase<-data_comc%>% group_by(year) %>% tally()

p<-ggplot(increase, aes(x=year, y=n))+
  geom_point(color="#003262")+theme_minimal()+
  geom_smooth(color="#FDB515")+
  scale_y_continuous(labels = scales::comma)+
  ggtitle("Number of NPOs each year")
p
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
increase2<-data_comc%>% group_by(year,ntee1) %>% tally()%>%filter(ntee1!="b")
p1<-ggplot(increase2, aes(x=year, y=n,group=ntee1,color=ntee1,label=ntee1))+
  geom_line()+theme_minimal()+
  scale_y_continuous(labels = scales::comma)+
  ggtitle("Number of NPOs each year by group")+
  geom_text(aes(x = 2016, y = 70000, label = "B", color = "B")) +
  geom_text(aes(x = 2016, y = 50000, label = "P", color = "P"))+
  geom_text(aes(x = 2016, y = 45000, label = "A", color = "A"))+
  geom_text(aes(x = 2016, y = 38000, label = "N", color = "N"))+
  geom_text(aes(x = 2016, y = 30000, label = "X", color = "X"))+
  geom_text(aes(x = 2016, y = 25000, label = "E", color = "E"))
p1
```



Overall, the name of organizations was increasing in the past decades. Meanwhile, the top 6 organizations are type B(Education),P(Human Services),A(Arts, Culture & Humanities),N(Recreation & Sports),X(Religion-Related),E(Health Care). However, there are some human services or health service organizations are not in the same category but may serve the same group such as LGBTQ or immigrants. Thus, we need to figure out the algorithm to detect organization in other dimensions such as group of population.

2.Trends of Immigrants

```
years <- list(2009,2010,2011,2012,2013,2014,2015,2016,2017,2018)

multi_year <-
  map(
    years,
    ~ get_acs(
      geography = "state",
      variables = c(fborn = "B06001_049",
                    total_pop = "B01003_001"),
      year = .x,
      geometry = FALSE
    )) %>%
map2(years, ~ mutate(.x, id = .y))
```

```
## Getting data from the 2005-2009 5-year ACS

## Getting data from the 2006-2010 5-year ACS
```

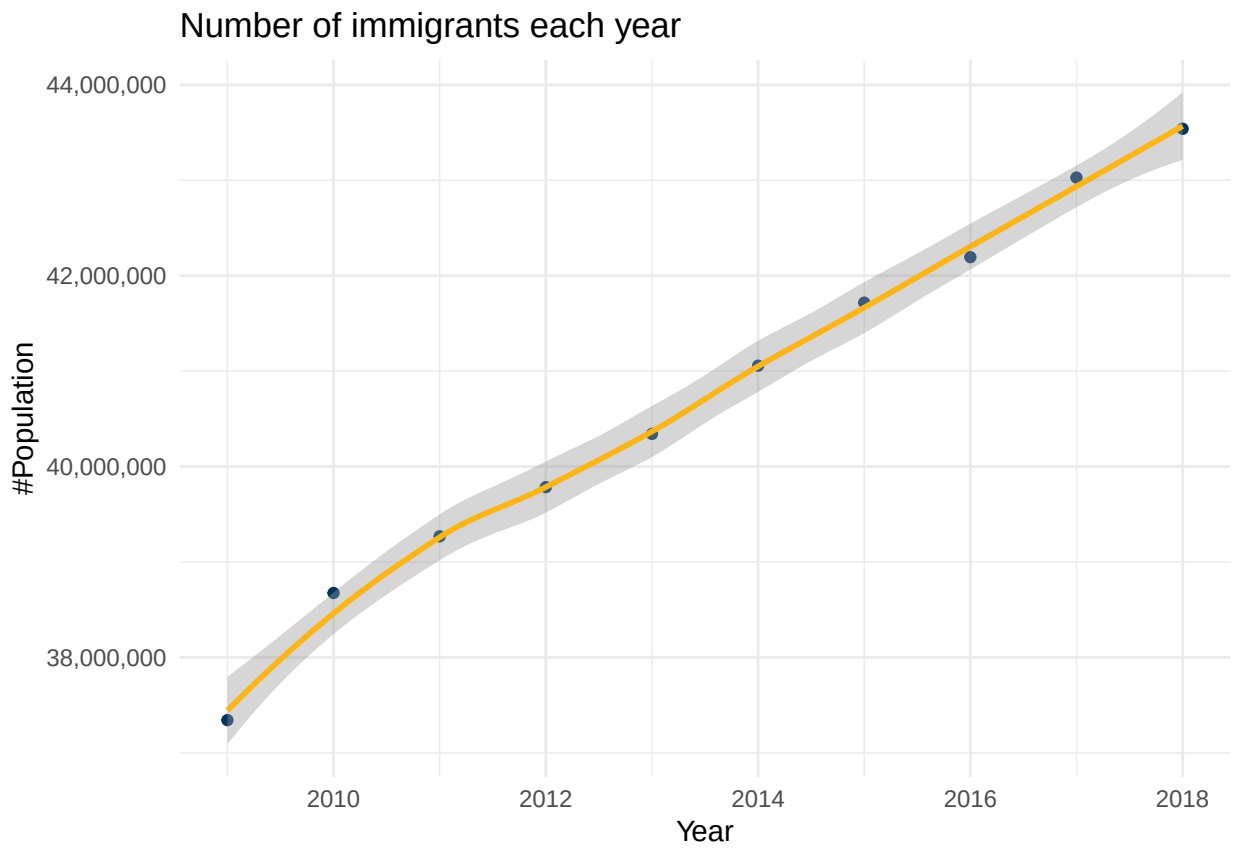
```
## Getting data from the 2007-2011 5-year ACS

## Getting data from the 2008-2012 5-year ACS

## Getting data from the 2009-2013 5-year ACS

## Getting data from the 2010-2014 5-year ACS

## Getting data from the 2011-2015 5-year ACS

## Getting data from the 2012-2016 5-year ACS

## Getting data from the 2013-2017 5-year ACS

## Getting data from the 2014-2018 5-year ACS
```

```
df_c=data.frame( multi_year[1])
for(i in 2:length(multi_year))
{
  df=data.frame(multi_year[i])
  df_c=rbind(df_c,df)
}
```

```
df_fb<-df_c %>%
  filter(variable=='fborn')
df_total<-df_c %>%
  filter(variable=='total_pop')
```

```
df_fb_total <- df_fb%>%group_by(id)%>%summarise(sum_pop=sum(estimate,na.rm = T))
p2<-ggplot(df_fb_total, aes(x=id, y=sum_pop))+
  geom_point(color="#003262")+theme_minimal()+
  geom_smooth(color="#FDB515")+
  scale_y_continuous(labels = scales::comma)+
  ggtitle("Number of immigrants each year")+
  xlab("Year")+ylab("#Population")
p2
```

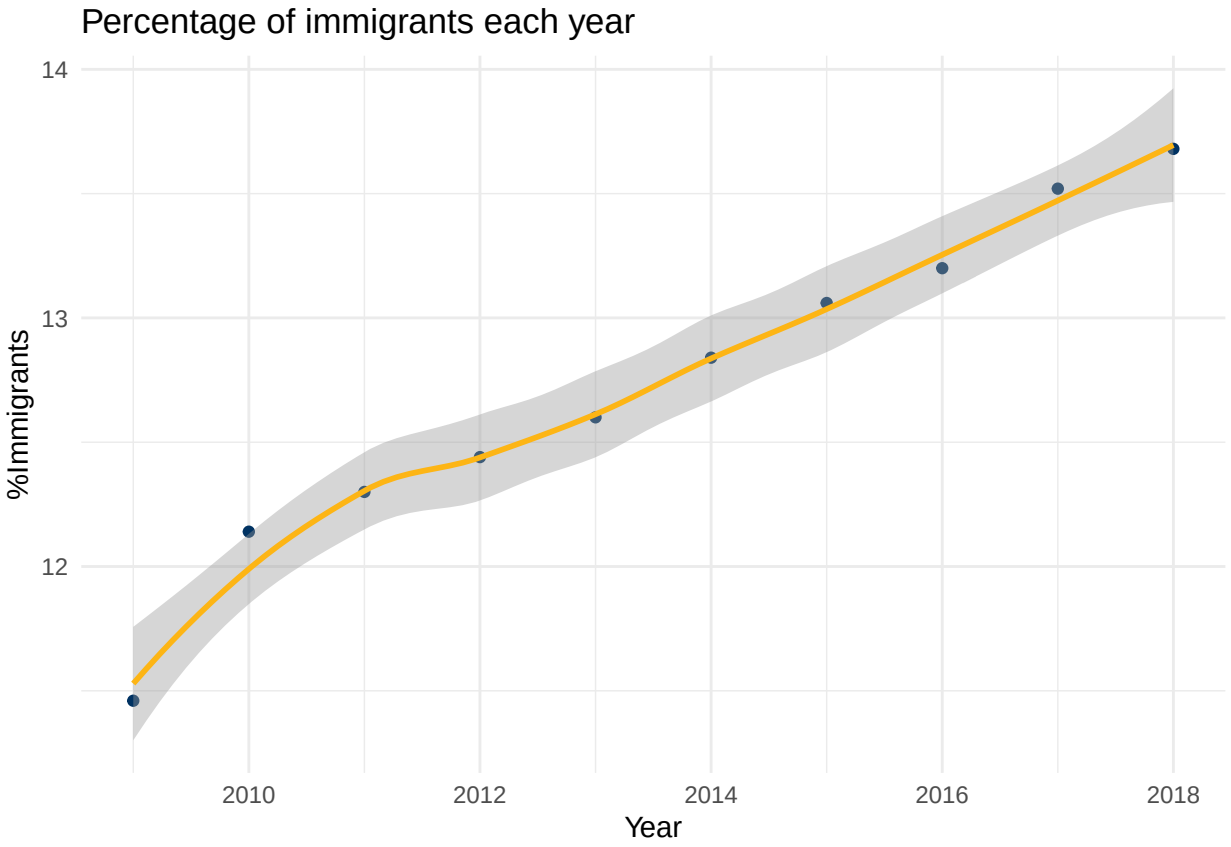
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
df_t_total <- df_total%>%group_by(id)%>%summarise(sum_pop=sum(estimate,na.rm = T))
df_t_total$per <- round(df_fb_total$sum_pop/df_t_total$sum_pop,4)*100
```

```
p3<-ggplot(df_t_total, aes(x=id, y=per))+
  geom_point(color="#003262")+theme_minimal()+
  geom_smooth(color="#FDB515")+
  scale_y_continuous(labels = scales::comma)+
  ggtitle("Percentage of immigrants each year")+
  xlab("Year")+ylab("%Immigrants")
p3
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



In the past ten years, the number of immigrants and percentage of immigrants of total population are both increasing. Here we can see, there will be more demmands from immigrants as the population increases.If we divide the population in to different ethics immigrants, we can provide better personalized service. For example, if we can find all the immigrants realted organizations, we located their position. Meanwhile, there are also some open data related to ethic immigrants concentrated census tract. Then, we can undertsand the trends in each community and compare the mismatch between service providers and demmand. We can identify these organizaitions for immigrants and provide suggestions for some related organizations to expand their market in the right place, where might be a blue sea.