Replication Codes (and Output) for Week 5 Exercise

Chutian Zhou

2/28/2018

0. Read Essential Packages

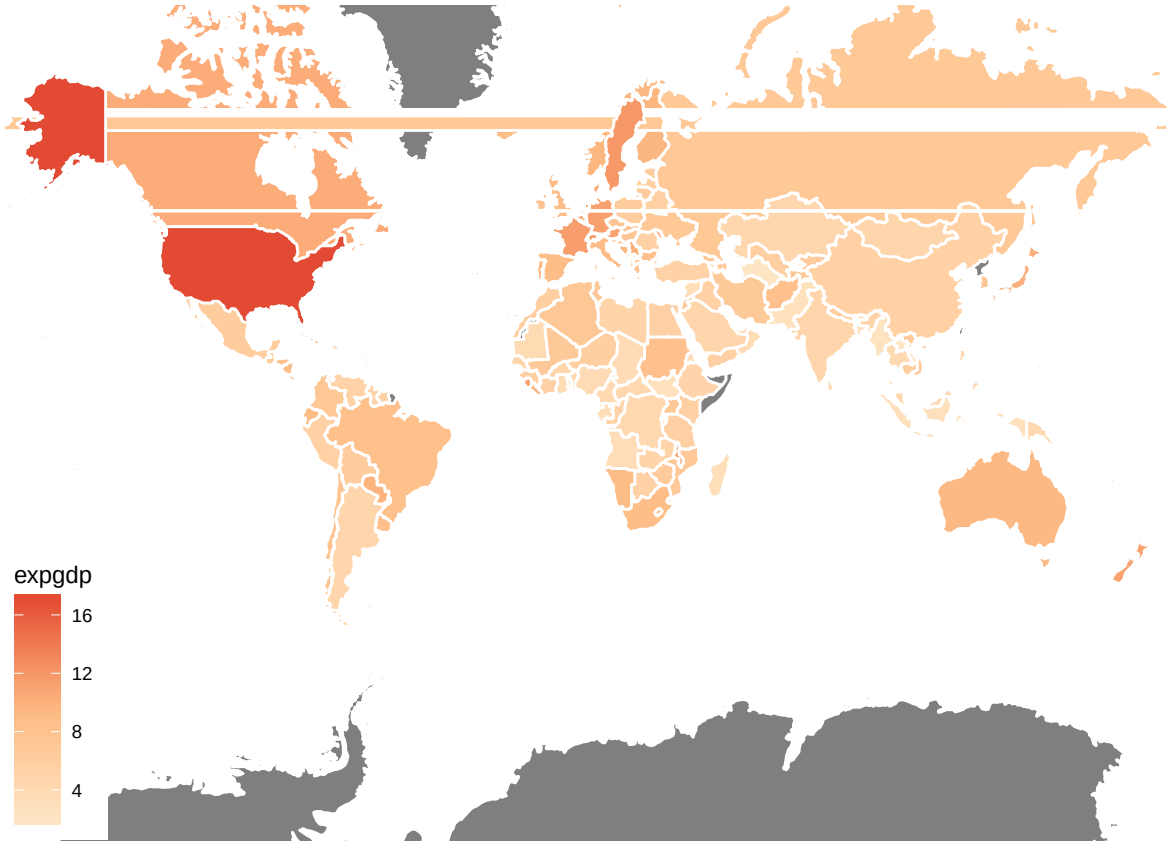
```
library(WDI)
library(dplyr)
library(ggplot2)
library(countrycode)
library(stringr)
library(ggrepel)
library(ggthemes)
library(ggmap)
library(tidyverse)
library(readxl)
```

1. World Development Indicators (WDI)

```
# Import and clean df
df=WDI(indicator="SH.XPD.TOTL.ZS" ,
         start=2014,end=2014,extra =F)
df=df %>% filter(!is.na(SH.XPD.TOTL.ZS))
df<-df[-c(1:46),]
df<-dplyr::rename(df,code=iso2c,expgdp=SH.XPD.TOTL.ZS) # rename "iso2c" as "code"
```

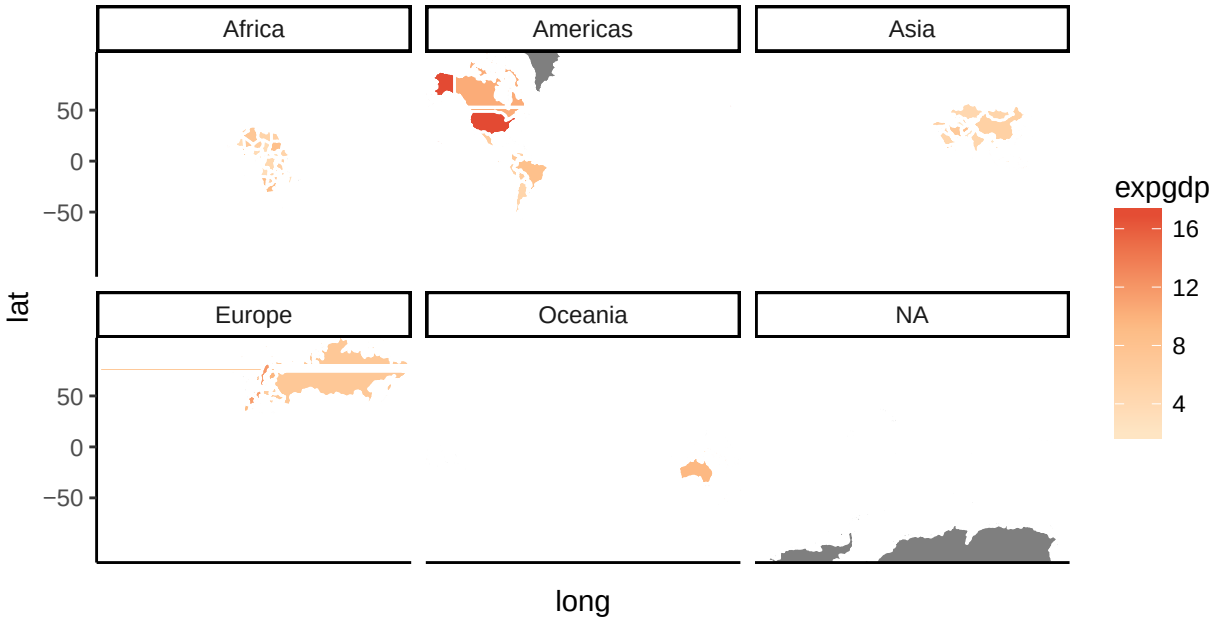
Question 1

```
# Generate data frame "world"
world<-map_data("world")
world<-dplyr::rename(world, country=region)
world$subregion=NULL
world$country<-str_to_title(world$country)
world<-mutate(world,code=
                countrycode(world$country,"country.name","iso2c")) # generate "code"

# Merge world and df together
world.merged<-left_join(world, df, by = "code")

# Draw plot (don't know why there are awkward lines...)
ggplot(world.merged,aes(x=long,y=lat,group=group))+
  geom_polygon(aes(fill = expgdp),color="white")+
  coord_map()+theme_map()+
  scale_fill_gradientn(colours=c("#fee8c8","#fdbb84","#e34a33"))
```



Question 2

```
# Generate "region" for facetting
world.merged$region<-countrycode(world.merged$code,'iso2c','continent')

# Draw plot
ggplot(world.merged,aes(x=long,y=lat,group=group,fill=expgdp))+
  geom_polygon(color="white")+
  coord_map()+theme_classic()+
  facet_wrap(~region)+
  scale_fill_gradientn(colours=c("#fee8c8","#fdbb84","#e34a33"))
```



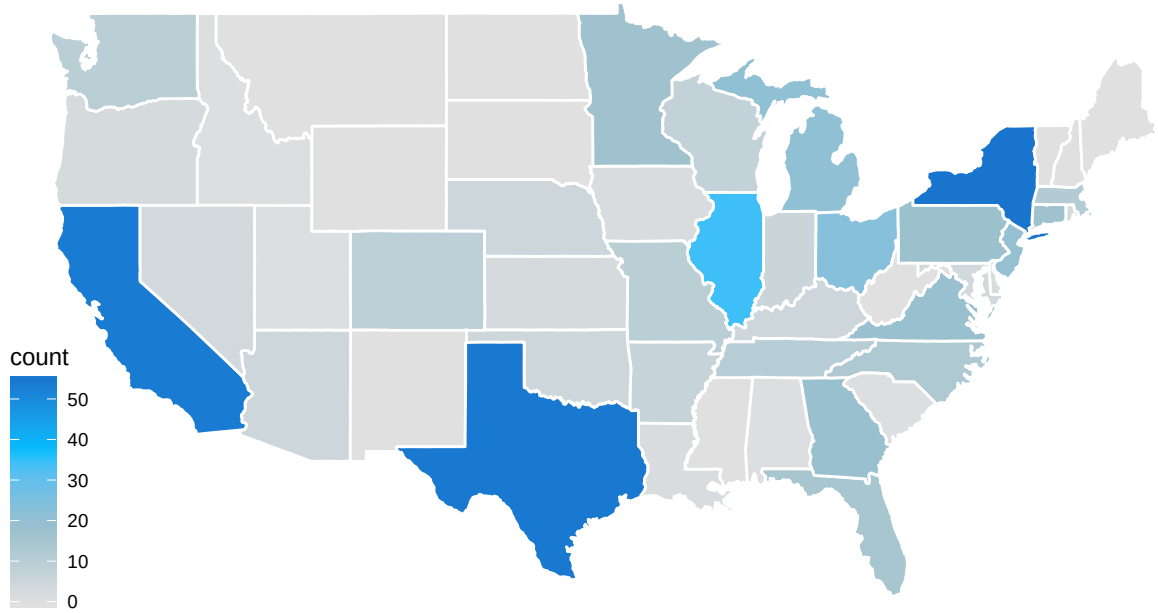
2. Locations of Fortune 500 companies

```
# Get fortune500 data frame ready (codes are provided by Professor)
library(XML)
library(RCurl)
fortune500_url<-getURL("https://www.geolounge.com/fortune-500-list-by-state-for-2015/",
                       .opts=list(ssl.verifypeer=FALSE))
fortune500=readHTMLTable(fortune500_url,header=TRUE,which=1)
colnames(fortune500)<-tolower(colnames(fortune500))
fortune500<-subset(fortune500,select=c("company","streetadd","place","state","zip"))
write.csv(fortune500,"fortune500.csv")
```

Task 1

```
# Generate new data frame bystate (number of companies in a single state)
bystate<-fortune500%>%group_by(state)%>%summarise(count=n())
bystate$state<-as.character(bystate$state)
bystate[8,1]="District Of Columbia"

# Generate data frame us.states and do some cleaning
us.states<-map_data("state")
us.states <-as_data_frame(us.states)
```

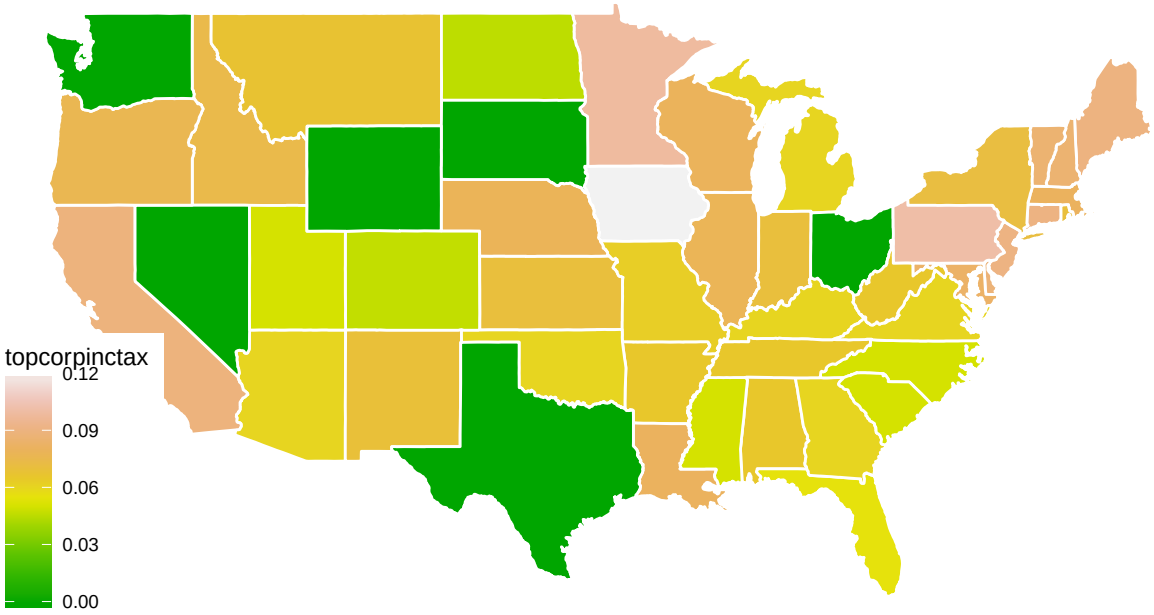
```
us.states<-rename(us.states, state = region)
us.states$state<-str_to_title(us.states$state)
us.states<-left_join(us.states,
                     bystate,by="state") # now us.states has column "count"
us.states<-us.states[,-6]
us.states[is.na(us.states)]<-0 # some states have "NA" in "count". Should be 0

# Draw plot
ggplot(us.states,aes(x =long,y=lat,group=group)) +
  geom_polygon(aes(fill=count),color="white")+theme_map()+
  coord_map(projection="mercator")+
  scale_fill_gradientn(colours=c("gray88","lightblue3","deepskyblue1","dodgerblue3"))
```



Task 2

```
# Read "tax"
tax<-read_excel("State_Corporate_Income_Tax_Rates_2015.xlsx",sheet=2)
tax[51,1]="District Of Columbia"

# Merge us.states and tax
us.states<-left_join(us.states,tax,by="state")

# Draw plot
ggplot(us.states,aes(x=long,y =lat,group=group)) +
```

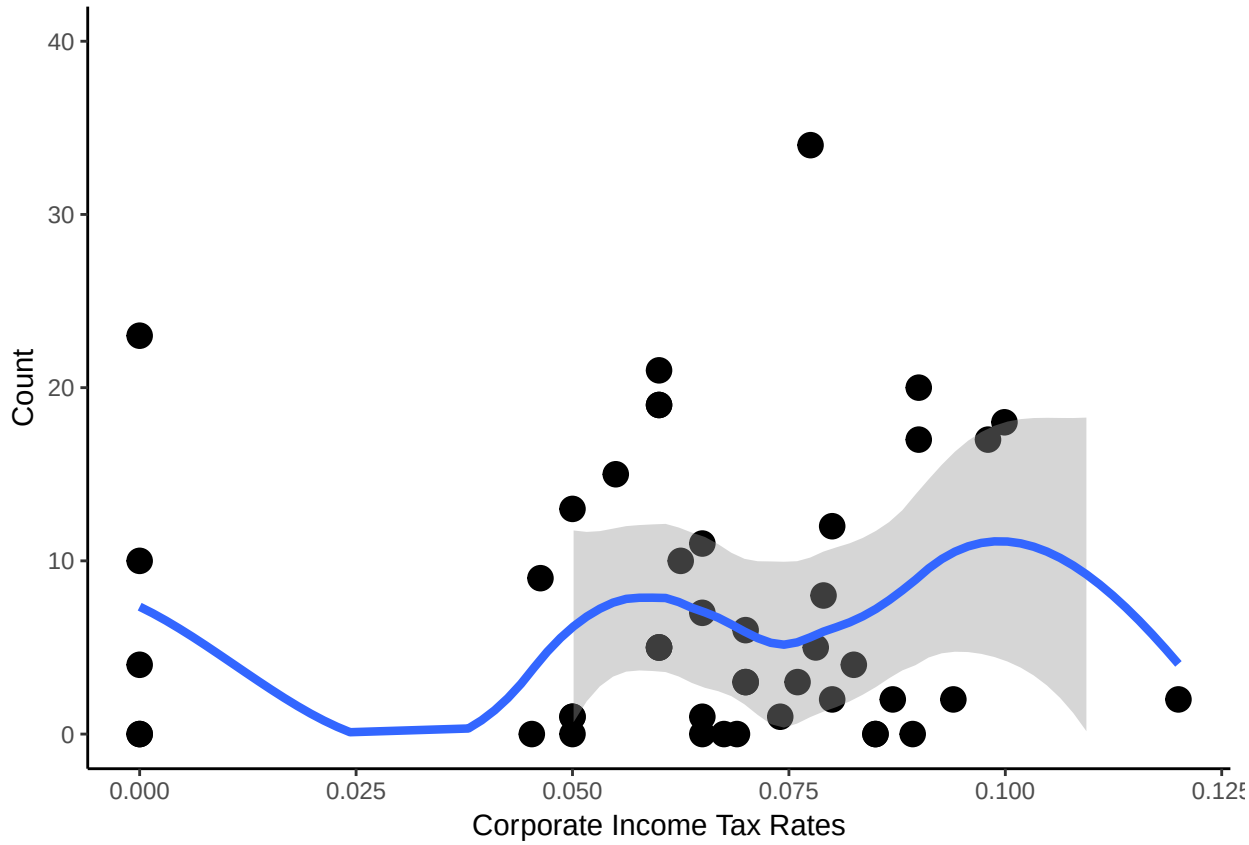
```
  geom_polygon(aes(fill=topcorpinctax), color="white")+theme_map()+
  coord_map(projection="mercator")+
  scale_fill_gradientn(colours = terrain.colors(10))
```



Task 3

(1)

```
# Drop repetitive rows
tax_count=us.states[!duplicated(us.states$state),]

# Draw plot
ggplot(tax_count,aes(topcorpinctax,count))+geom_point(size=4)+
  theme_classic()+geom_smooth(method=loess,size=1.5)+
  scale_y_continuous(name="Count",limits=c(0,40))+
  scale_x_continuous(name="Corporate Income Tax Rates")
```



(2)

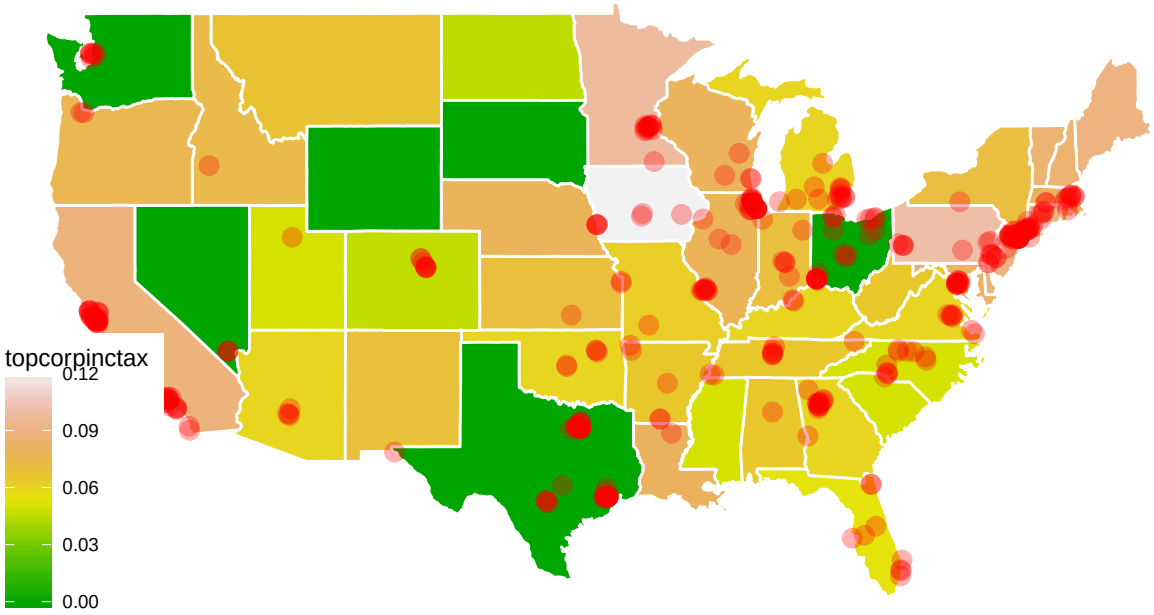
Don't know how to import online table... Should be very easy.

Task 4

(1)

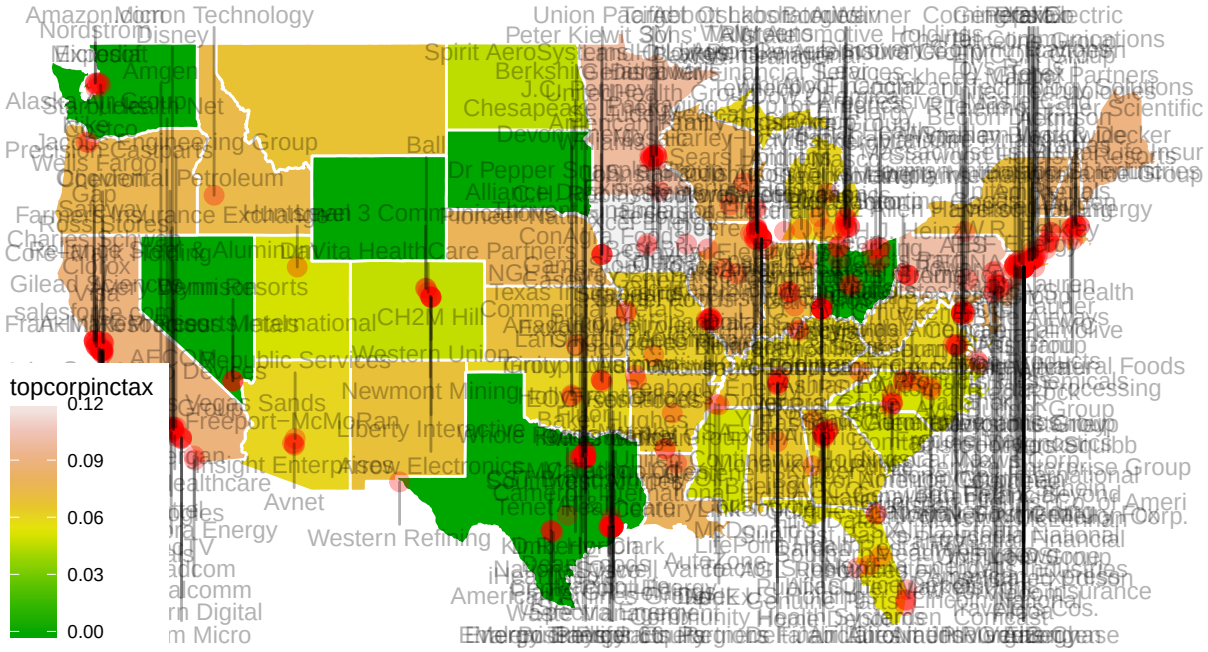
```
# Generate column "address" in which each row has previous columns' info
fortune500$address<-paste(fortune500$streetadd,
                          fortune500$place,
                          fortune500$state,
                          fortune500$zip)

# Run geocode() on fortune500$address
geocodes<-geocode(fortune500$address,output="latlon",source="google")

# Let fortune500 has lon/lat info contained in geocodes data frame
fortune500$lon<-geocodes$lon
fortune500$lat<-geocodes$lat

#Draw plot (notice inherit.aes=FALSE in the aes() in geom_point())
ggplot(us.states,aes(x=long,y=lat,group=group)) +
  geom_polygon(aes(fill=topcorpinctax),color="white")+theme_map()+
```

```
  coord_map(projection="mercator")+
  scale_fill_gradientn(colours=terrain.colors(10))+
  geom_point(aes(x=lon,y=lat),data=fortune500,size=3,alpha=0.3,col="red",inherit.aes=FALSE)
```



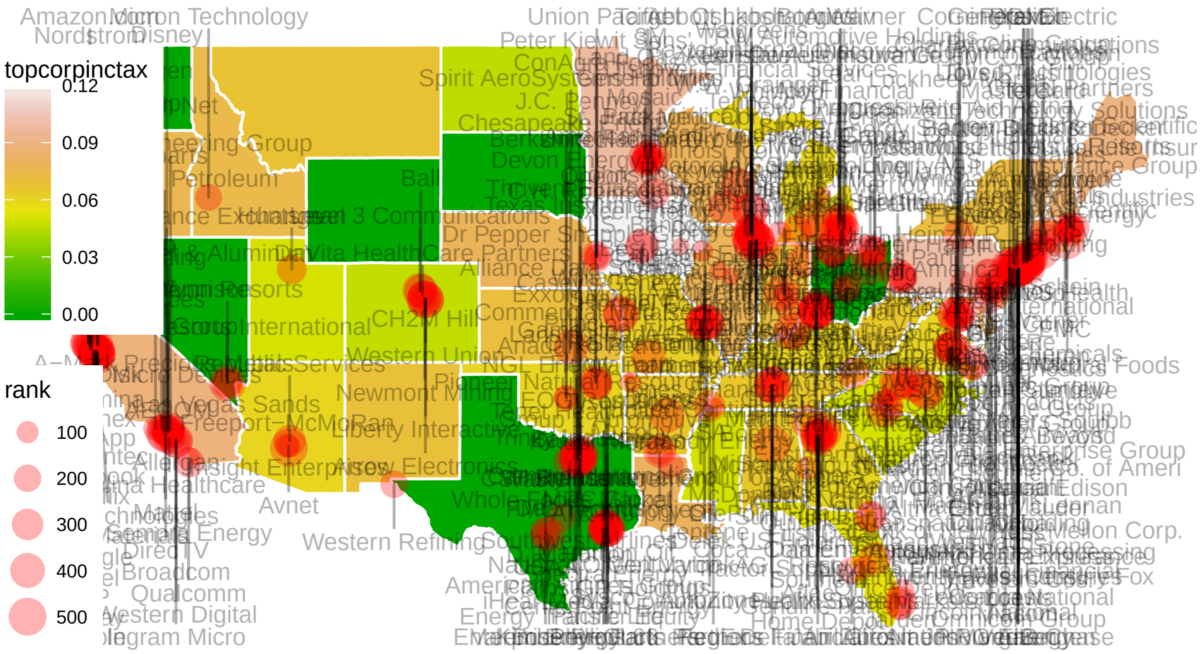
(2)

```
# Notice geom_text_repel() here. It's from ggrepel package
ggplot(us.states,aes(x=long, y =lat, group=group)) +
  geom_polygon(aes(fill=topcorpinctax), color="white")+theme_map()+
  coord_map(projection="mercator")+
  scale_fill_gradientn(colours=terrain.colors(10))+
  geom_point(aes(lon,lat),data=fortune500,size=3,alpha=0.3,col="red",inherit.aes = FALSE)+
  geom_text_repel(aes(label=company,x=lon,y=lat),
                  data=fortune500,size=3,alpha=0.3,col="black",
                  inherit.aes = FALSE,direction = "y")
```



(3)

```
# Generate column "rank", which is exactly the row number
fortune500$rank<-as.numeric(rownames(fortune500))

ggplot(us.states,aes(x=long, y=lat, group=group)) +
  geom_polygon(aes(fill=topcorpinctax),color="white")+theme_map()+
  coord_map(projection="mercator")+
  scale_fill_gradientn(colours=terrain.colors(10))+
  geom_point(aes(lon,lat,size=rank), # size=rank
             data=fortune500,alpha=0.3,col="red",inherit.aes=FALSE)+
  geom_text_repel(aes(label=company,x=lon,y=lat),
                  data=fortune500,size=3,alpha=0.3,col="black",
                  inherit.aes=FALSE,direction="y")
```


topcorpinctax
0.12
0.09
0.06
0.03
0.00
rank
100
200
300
400
500
Micron Technology
Disney
Western Refining
Avnet
Broadcom
Qualcomm
Western Digital
Las Vegas Sands
Newmont Mining
Liberty Interactive
CH2M Hill
Western Union
Spirit AeroSystems
Alliance Data