Exercise
Use R to make a map of the states/provinces/regions showing GDP or income inequality in a country of your choice (except USA).

Step 1: Getting some specific package

```
library(raster) # to get map shape file
library(ggplot2) # for plotting and miscellaneuous things
library(ggmap) # for plotting
library(plyr) # for merging datasets
library(scales) # to get nice looking legends
```

Step 2: Making my own csv file to save data of “GDP of Provinces in China in mid 2012”

I got the data from Wikipedia. URL: https://en.wikipedia.org/wiki/List_of_Chinese_administrative_divisions_by_GDP_per_capita
I use data of 2012 GDP per capita (based on mid-year population) and made a csv file which has 3 variables and 31 objects.
I identify the variable of id in order to match and join values with others.
Another reason that I use the name “id” instead of other names is because there is a parameter named “id” when I get value of shape file from internet. The parameter “id” in shape file corresponds longitude and latitude parameters. So that when I use “join” function, it would be more convenient.
I identify the variable of abbr in order to show the name of each province on map.
The third variable is value of gdp which will be shown on map.

Step 3: Load data from csv file to Rstudio

I use the “read.csv” command to extract information from my csv file.

```
cn.val <- read.csv("F:\\6313 statistic for DS\\project\\1cn_2012_gdp.csv", header = T, sep = ",")
```

After that, I can do operations on my data.

Step 4: Getting a Large Spatial Polygons Data Frame file from internet by using 3 –letter code “CHN”

```
cn.getdata    <- getData("GADM", country = "CHN", level = 1)
```

After execute that, we can find the line of “$NAME_1” shows name of each province which means I got correct data.

Step 5: Extract information from the file we download from the internet

```
cn.usedata <- fortify(cn.getdata)
```

cn.usedata contains several variables that is useful for making map. Such as values of longitude and latitude and id.

Step 6: Combine all the useful data to cn.usedata in order to use later.

```
cn.usedata <- join(cn.usedata, cn.val, by = "id", type = "inner")
```

After this command, we can get values of longitude and latitude. But the value of gdp is wrong because on this step, id equals to 1 without any changes. So the value of gdp is the value of the first province. But on this step I just want to extract data of longitude and latitude.

Step 7 : Set up important information of the map

```
position <- ddply(cn.usedata, .(id),summarize, clat = mean(lat), clong = mean(long))
```

I can get correct order of id from this step and mean value of longitude and latitude which will be used on the place where abbr show on the map.

```
position <- join(cn.val, position, by = "id", type = "inner")
```

After join, I can get correct data of gdp of each province.

Step 8: Find the range of GDP so that color in map can look better.

```
range(cn.usedata$gdp)
[1]     70103  5706792
```

Based on the result, I make the range of brks:
brks <- c(70000, 100000, 2000000,4000000, 6000000)

The value of each parameter can adjust based on result.

Step 9 : Set up important information of the map by using the values of cn.usedata and position.

cn.usedata will provide value of gdp and show in the map with different level of red color. Position will provide abbreviation of each province and show in the map.

```
p <- ggplot() +
  geom_polygon(data = cn.usedata, aes(x = long, y = lat, group = group, fill = gdp),
                    color = "black", size = 0.15) +
  scale_fill_distiller(palette = "Reds", breaks = brks, trans = "reverse") +
  geom_text(data = position, aes(x = clong, y = clat, label = abbr, size = 0.1))+
  theme_nothing(legend = TRUE) +
  labs(title = "GDP of provinces in China in mid 2012", fill = "in millions")
```

Step 10: making the map:

plot(p)