ADS2 week6 Visualizing data - Problem set

Chaochen Wang

2019-10-17

You will need the packages "tidyr","dplyr","ggplot2" for the first three problems and "map" for the optional one.You could use the gather and drop_na in "tidyr" and filter in "dplyr".

Import the GDP dataset and clean the data. (The data are obtained from the world bank https://data.worldbank.org/indicator/NY.GDP.MKTP.CD and trimmed.) 1.Import the dataset and reshape the dataset by putting the GDP values from different years into one variable(column), the key is "Year". 2.Clean the data by removing the imcomplete records. 3.Extract the GDPs of "France", "Germany","Italy" and "Greece" in the years 1960, 1970, 1980, 1990, 2000, 2010 and 2018. Put the data into a new data frame. (Hint: dplyr::filter)

```r
library(ggplot2)
library(tidyr)
setwd("~/Teaching/2019 statistics/week6/data/GDP") # set to your own wo
rking directory
gdp <- read.csv("gdp.csv", na.strings = "",header = T, check.names = F)
#reshape the dataset
gdp <- gather(gdp,key = "Year", value = "GDP", factor_key = T,-CountryN
ame, convert = T)
#remove incomplete records
gdp.noNA <- drop_na(gdp)
#extract subset of data
gdp.sub <- dplyr::filter(gdp.noNA, CountryName %in% c("France","Germany
","Italy","Greece"),Year %in% c( 1960,1970,1980,1990,2000,2010,2018) )
summary(gdp.sub)

##       CountryName        Year           GDP
##   France     :7    Min.   :1960   Min.   :4.447e+09
##   Greece     :7    1st Qu.:1975   1st Qu.:1.215e+11
##   Italy      :7    Median :1990   Median :7.010e+11
##   Germany    :6    Mean   :1991   Mean   :1.082e+12
##   Afghanistan:0    3rd Qu.:2010   3rd Qu.:1.855e+12
##   Albania    :0    Max.   :2018   Max.   :4.000e+12
##   (Other)    :0
```

We want to see the trend of GDP growth in the three countries, please frist use point + line plots to present the data, distinguish countries by colours like this. Please also add the title and the GDP numbers at each point (Hint:use geom_point, geom_line ,geom_text and labs).

```
g <- ggplot(data=gdp.sub, aes(x=Year, y= GDP,color=CountryName))
g1 <- g + geom_point() + geom_line()
g2 <- g1 + geom_text(aes(label =GDP),hjust = 0.1, nudge_x = 0.05 ,size=
2)  + labs(title = "GDP trends in the countries")
g2
```
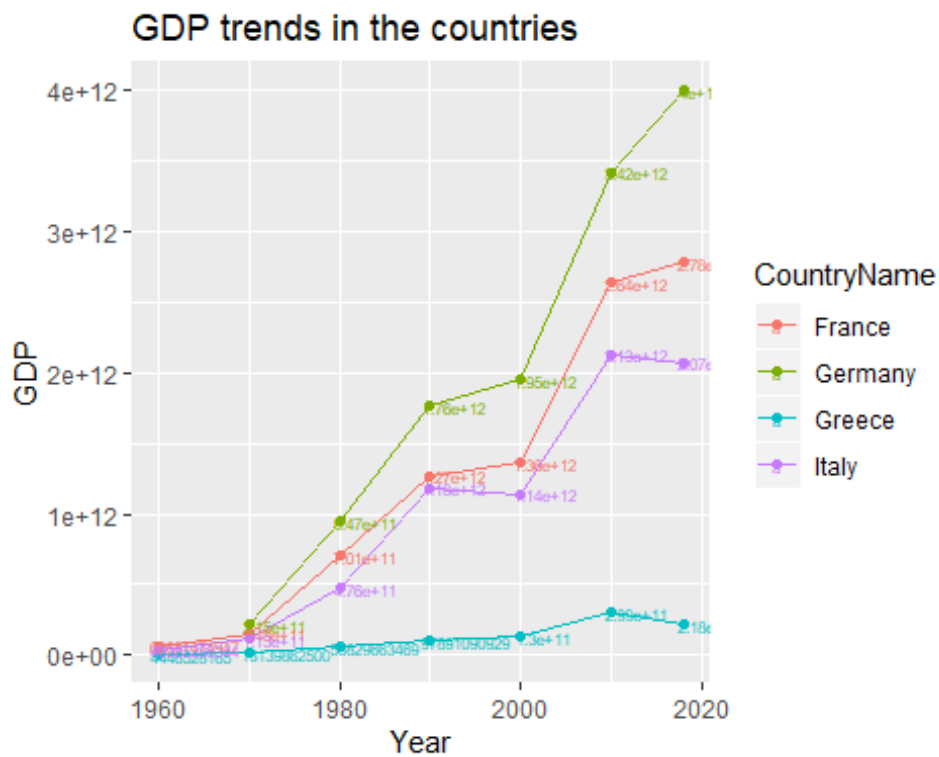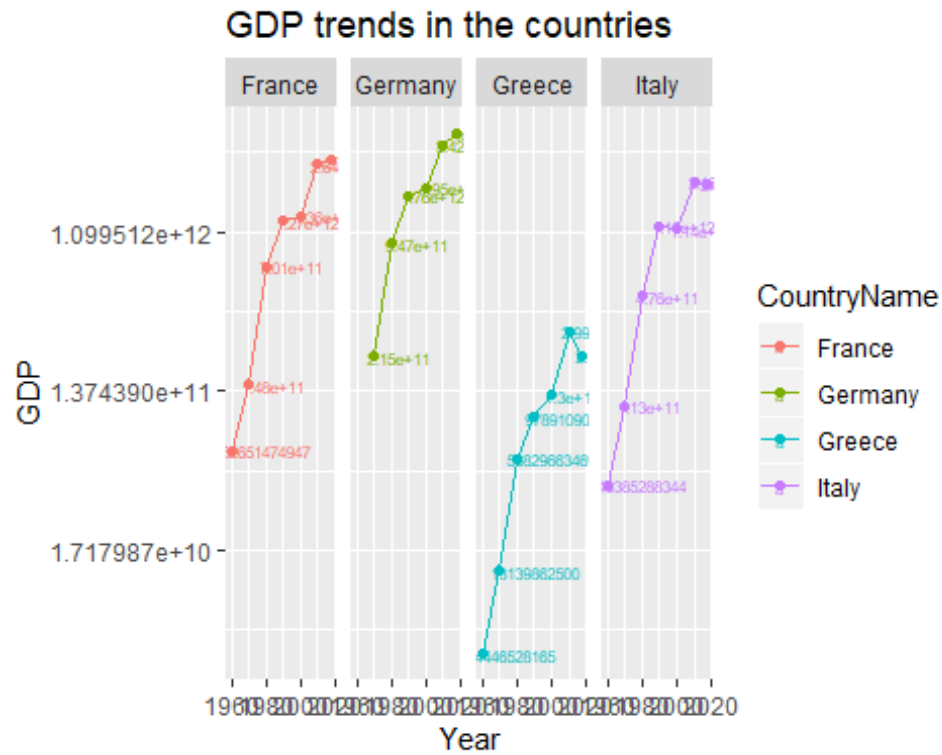

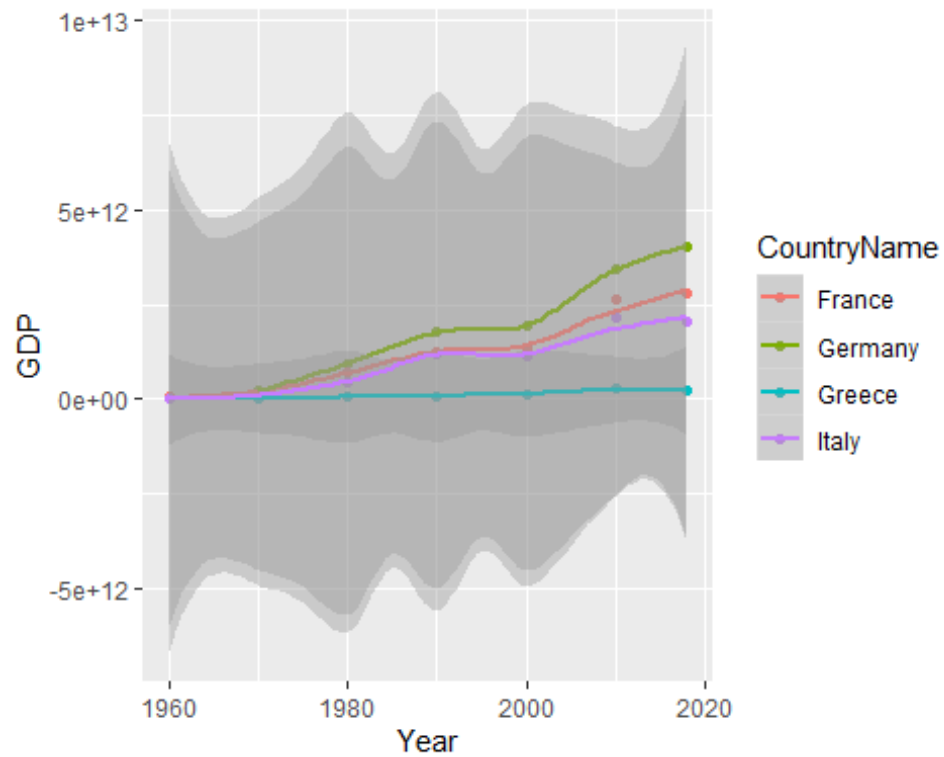Then try to rescale the GDP by log2 and facet the plot by countries.

```
g2 + scale_y_continuous(trans='log2') + facet_wrap(~CountryName,ncol =
4)
```
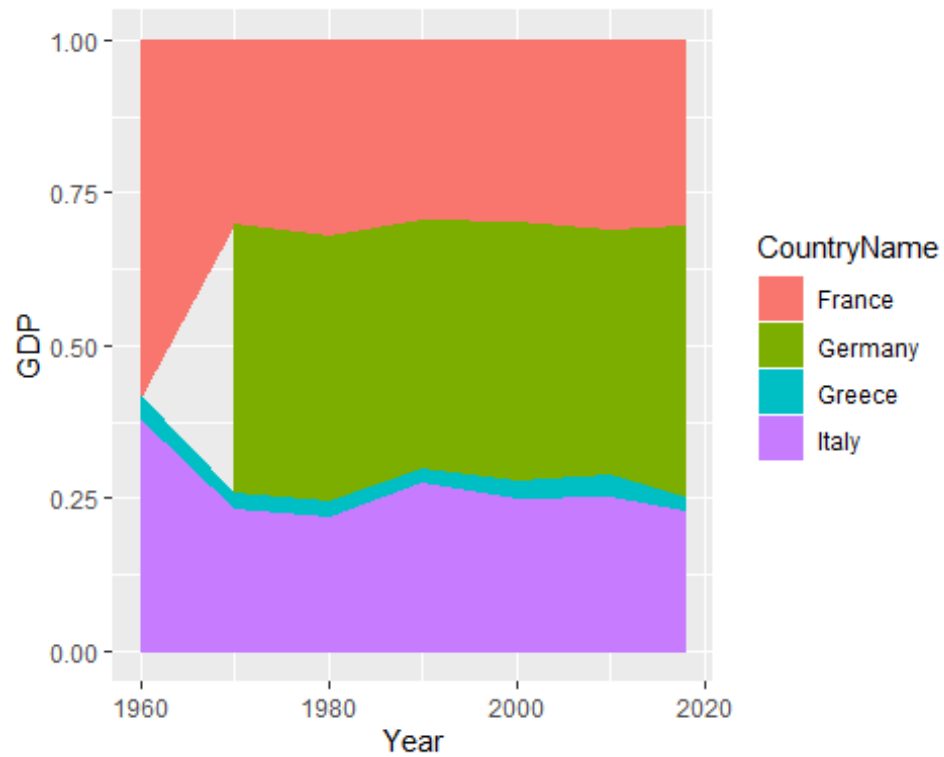
GDP trends in the countries

You can redo the plot using point + fitting with "loess" method.(Hint: geom_point + geom_smooth)

```
g + geom_point() + geom_smooth(method = "loess")
```
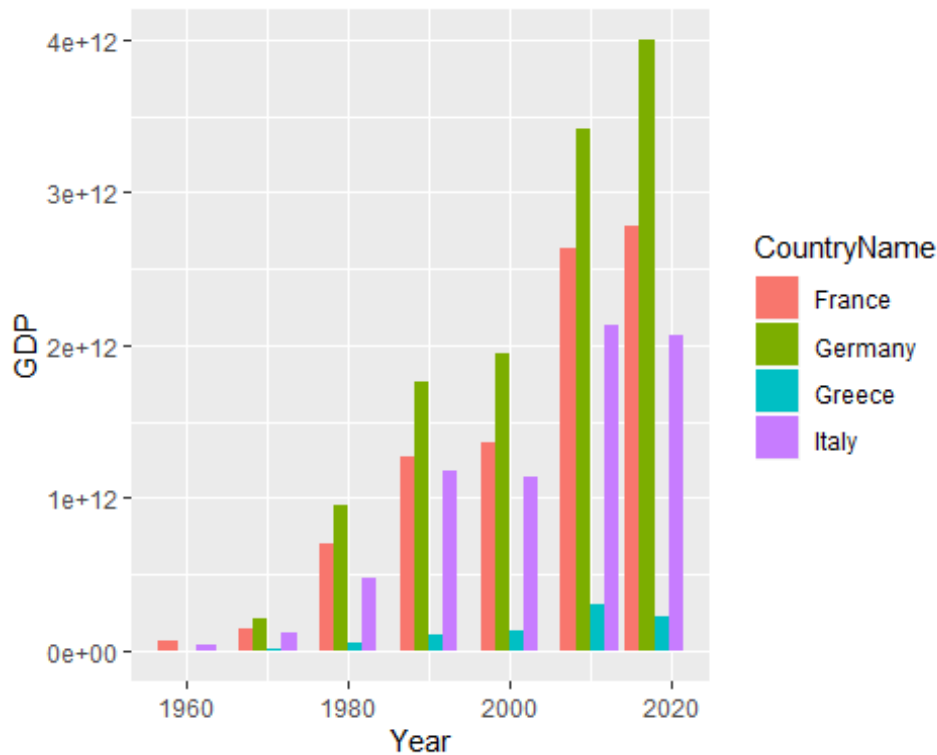
 You can also use area plot and "fill" to show the percentiles of the countries in different years. (Hint:geom_area, use the right position)

```
g + geom_area(aes(fill=CountryName), position = "fill")
```

As another way of comparison, use bar plot to present. (Hint: position)

```
ggplot(data=gdp.sub, aes(x=Year, y= GDP)) +geom_bar(stat = "identity",a
es(group=CountryName,fill=CountryName),position = "dodge")
```

(Optional)We want to show the GDP differences of the three countries in 2018 on the map. First, we need to get the 2018 GDP data out by subsetting the dataset.(Hint: dplyr::filter)
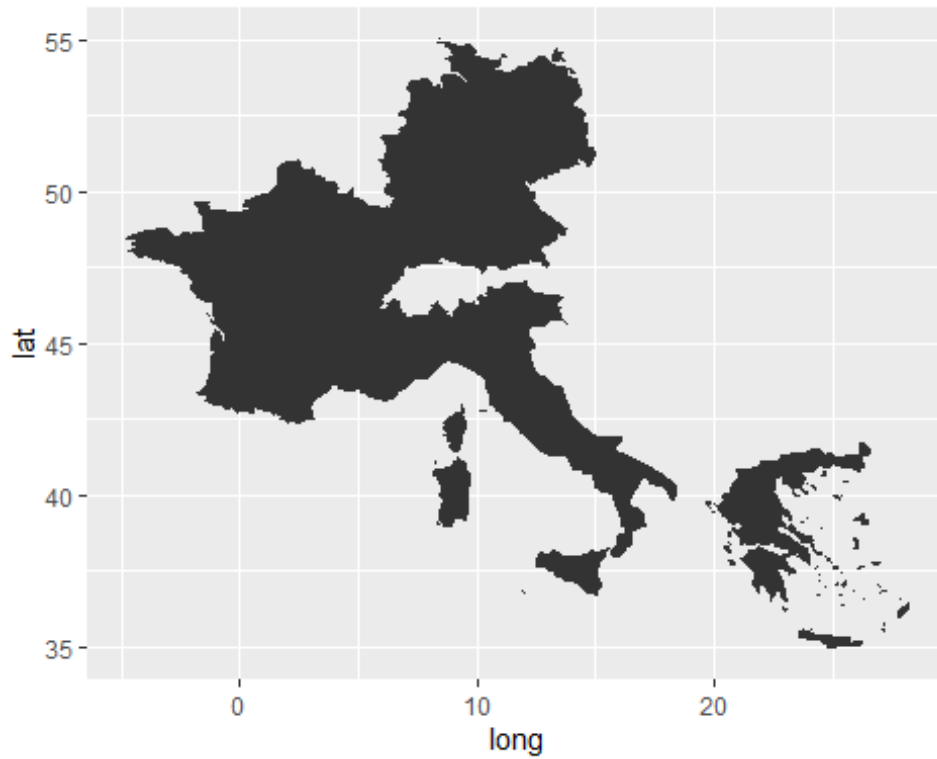
```
gdp.2018 <- dplyr::filter(gdp.sub, Year == "2018")
```

The map data can be obtained from the package "map" using the following code.

```
library(maps)
eu_map <- map_data("world", region = c("France","Germany","Italy","Greece"))
```
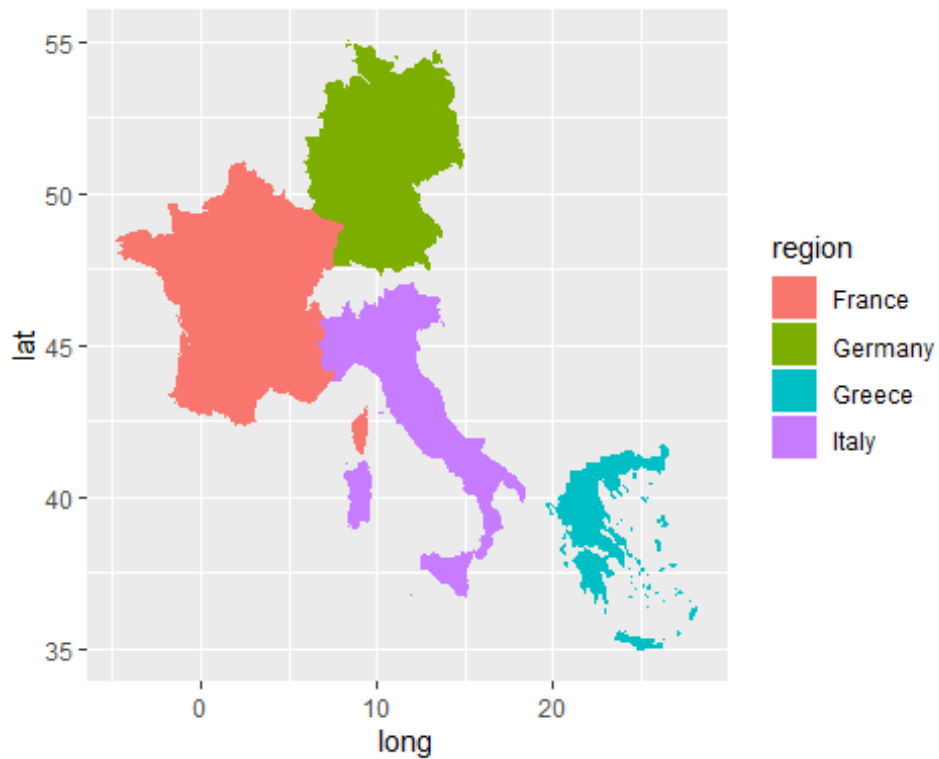
Now you can plot the map using ggplot2 with geom_polygon (Each country is a polygon). Try the code below.

```
ggplot(eu_map, aes(x=long, y=lat,group=group)) + geom_polygon()
```

To better present each country, change the colour of each country in map (Hint: use the fill argument).

```
ggplot(eu_map, aes(x=long, y=lat,group=group)) + geom_polygon(aes(fill=
region))
```
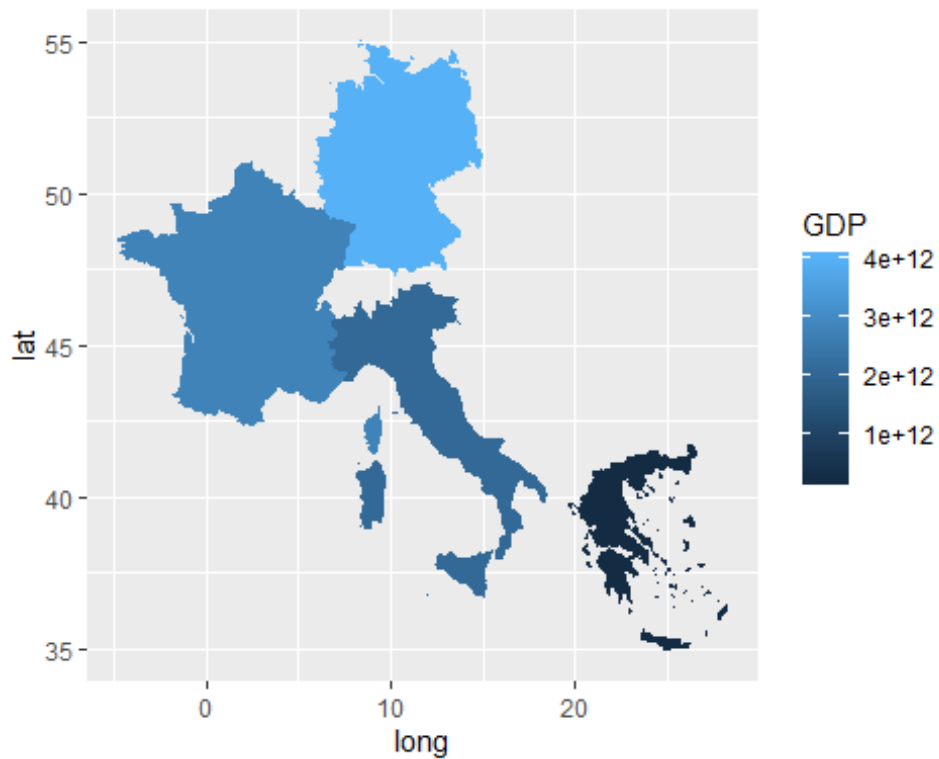
 Now we can try to present the GDP.The first step is to merge the GDP data and map data.(Hint: use dplyr::left_join)

```
gdp.map <- dplyr::left_join(eu_map, gdp.2018, by=c("region"="CountryNam
e"))
```

Then use ggplot to colour the countries based on their GDP values.

```
gmap <- ggplot(gdp.map, aes(x=long, y=lat, group=group)) + geom_polygon
(aes(fill=GDP))
gmap
```

Now, add the title, change the color scale with better labeling and save to a file.

```
gmap <- gmap + labs(title = "2018 GDP in eu countries")+ scale_fill_gra
dient(low="dark blue",high="red",breaks=c(1e+12,2e+12,3e+12),labels=c("
1 Trillion","2 Trillion","3 Trillion"))
gmap
```

2018 GDP in eu countries
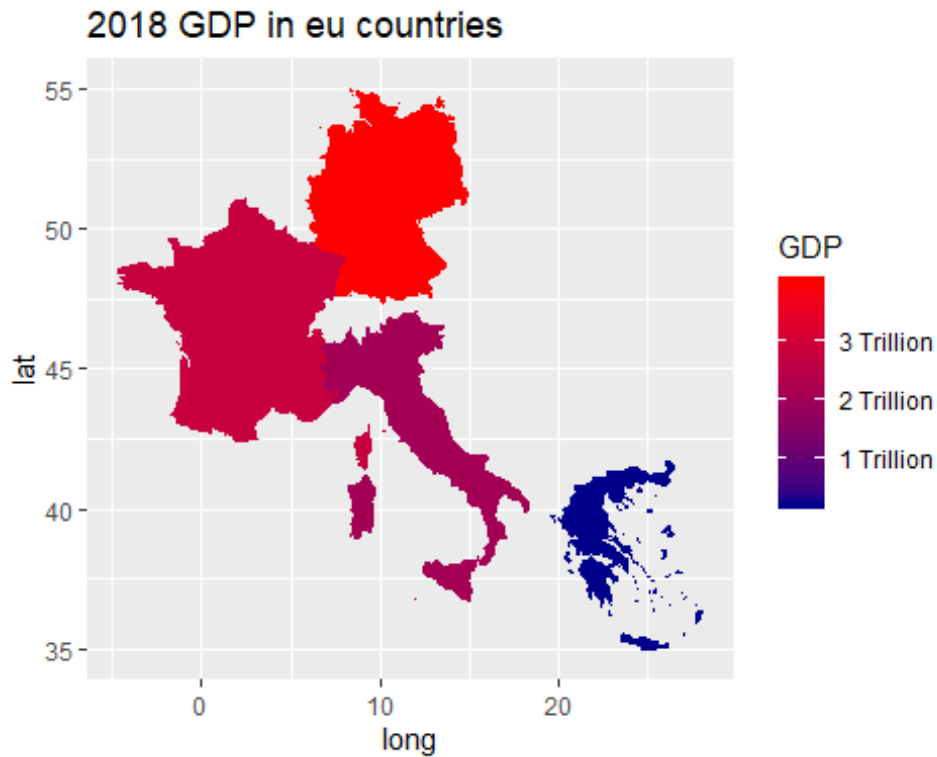
```
png("GDP map.png", width = 600, height = 600)
gmap
dev.off()
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.