Group 9 Global Trade

2022-12-02

[introduction]

Since 2019, COVID-19 has been prevalent, making many changes in our daily lives. Because it is a highly contagious disease, our society has come up with measures to minimize contact with the outside world. However, in the global era, interaction with various countries is very deep-rooted, and in this situation, interruption of interaction would have caused many social problems and confusion. We think trade is the area that has been hit the most by these changes. We would like to analyze how Corona affected the trading and investigate various trade-related figures in this situation.

[question]

While examining the characteristics of countries by level of trade dependence, we found that countries with a high trade dependence generally have low unemployment rates, and countries with a high future population tend to have a low trade dependence because of their sufficient labor. In response, the question arose, "Does the unemployment rate and future population affect determining trade dependence?".

[about trade dependence]

- Trade dependence is the degree of dependence upon foreign trade. In other words, Trade dependence refers to the proportion of exports and imports in gross domestic product(GDP) over a certain period of time (usually one year).
- trade dependence = (Export+Import)/GDP*100
- The fact that the country's trade dependence is high means that the national economy is unstable which leads to an international dependence situation. So, in countries that have high trade dependence, the economic situation will be more affected if internationally damaging issues such as COVID-19 occur.

[library we used]

ggplot2 : visualization library

reshape2 : restructuring the data frame

dplyr : data frame processing

gridExtra : library for putting graphs in one chart

magrittr : use the %>% pipe operator

tidyverse : packages with tools to analyze data

[about dataset]

OECE_Major_Indicator (180 rows, 12 columns) is about major indicators from OECD countries such as GDP, Internet Usage rate and so on.

global_trade_2019_2021 is about export&import amount and export&import share by year and country.

2019_bw(5164 rows, 9 columns)/2020_bw(17314 rows, 9 columns)/2021_bw(3611 rows, 9 columns) is about export&import information by year and country (e.g. amount, share, item) of items with the highest or lowest export growth rate compared to the previous year preprogressing

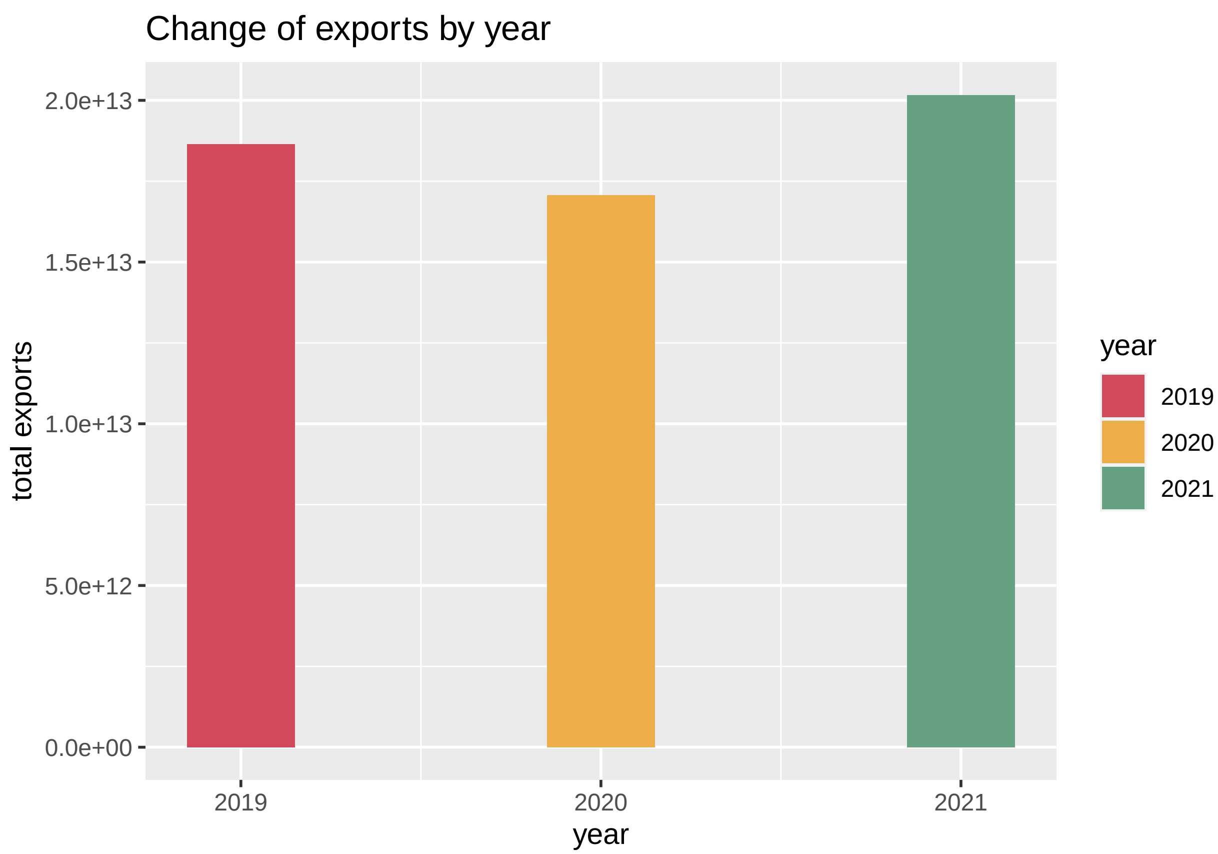
```r
#change column names
names(OECD)<-c("country","year","GDP","GDP_per_capita","GDP_growth","Export","Import"
               ,"Future_population","Unemployment_rate","Consumer_price","Crude_steel_production",
               "Internet_usage_rate")

#remove rows which have NA
OECD<-OECD[!(OECD['country'] == "Aisa"|OECD["country"]=="North America"
             |OECD["country"]=="South America"|OECD["country"]=="Europe"
             |OECD["country"]=="Oceania"|OECD["country"]=="OECD"|OECD["country"]=="World"), ]

#change "-"& ","
OECD$GDP<-gsub("-","0",OECD$GDP)
OECD$GDP<-gsub(",","",OECD$GDP)

#convert character type to numeric type
OECD$GDP<-as.numeric(OECD$GDP)

#multiply 1000 to match the units of exports and imports
OECD$GDP<-sapply(OECD$GDP, function(x) x*1000)

#change "-" or "" into 0
cond1<-OECD['Export']=='-'|OECD['Export']==''
OECD["Export"][cond1]<-0
cond2<-OECD["Import"]=='-'|OECD["Import"]==''
OECD["Import"][cond2]<-0

#remove ","
OECD["Export"]<-gsub(",","",OECD$Export)
OECD["Import"]<-gsub(",","",OECD$Import)

#convert character type to numeric type
OECD["Export"]<-as.numeric(OECD$Export)
OECD["Import"]<-as.numeric(OECD$Import)

#change "," and "-"/change to numeric type
OECD$Future_population<-gsub(",","",OECD$Future_population)
OECD$Future_population<-gsub("-","0",OECD$Future_population)
OECD$Future_population<-as.numeric(OECD$Future_population)

#change "-"/change to numeric type
OECD$Unemployment_rate<-gsub("-","0",OECD$Unemployment_rate)
OECD$Unemployment_rate<-as.numeric(OECD$Unemployment_rate)

#Initialize index
rownames(OECD) <-NULL

#calculate sum of import and export
total=OECD$Import + OECD$Export
OECD$total<-total
```

```r
#convert the value to NA because there is a country in 2020 that did not have data on the GDP
#and calculate trade dependence rate = (export+import)/GDP*100

dependence_rate<-as.numeric(ifelse(OECD$GDP==0,NA,round((OECD$total/OECD$GDP)*100,2)))
OECD$Dependence<-as.numeric(dependence_rate)

#remove un-important variables used
remove(dependence_rate)
remove(cond1)
remove(cond2)
```

- We changed the complex names to simple names to create an easy working environment.
- We removed rows with continental names in country column because they don't have value.
- Using 'gsub' function, we changed all "-" to zero, and all "," to "". This Excel file used"," for numerical separation and marked a value that does not exist as "-". We changed all the columns of a character type which we will use into a numeric type.
- The alignment of index numbers has become in disorder due to the operation of removing certain rows. So, we initialized the index.
- We added a 'total' column to calculate trade dependence. 'total' is the sum of import and export.
- In 2020, Japan didn't have GDP data. so we change 2020's Japan GDP into NA
- In row 115 of our code, We added a new column about trade dependence. Trade dependence is calculated by dividing total(Export+Import) by GDP. Since the unit of 'total' is million and the unit of 'GDP' is billion, we multiplied the calculation result by 1000 to match the units.
- We removed un-important variables used at the end of each chunk.

total Exports, change

```r
#calculate total(import+export) by year
total<-df_global%>%group_by(Inquiry.Criteria)%>%summarize(Export=sum(Export.amount..unit...))

#create bar plot
colorchip <- c("#d1495b", "#edae49", "#66a182")
ggplot(data=total, aes(x=Inquiry.Criteria, y=Export, fill=as.factor(Inquiry.Criteria))) +
  geom_bar(stat="identity", width=0.3)+
  scale_x_continuous(breaks = c(2019,2020,2021))+
  scale_fill_manual(values=colorchip)+
  xlab("year")+
  ylab("total exports")+
  ggtitle("Change of exports by year")+
  labs(fill="year")
```




```
#remove un-important variables used
remove(colorchip)
remove(total)
remove(a)
```

```
## Warning in remove(a):   'a'
```

```
remove(b)
```

```
## Warning in remove(b):   'b'
```

```
remove(ex19)
```

```
## Warning in remove(ex19):   'ex19'
```

```
remove(ex20)
```

```
## Warning in remove(ex20):   'ex20'
```

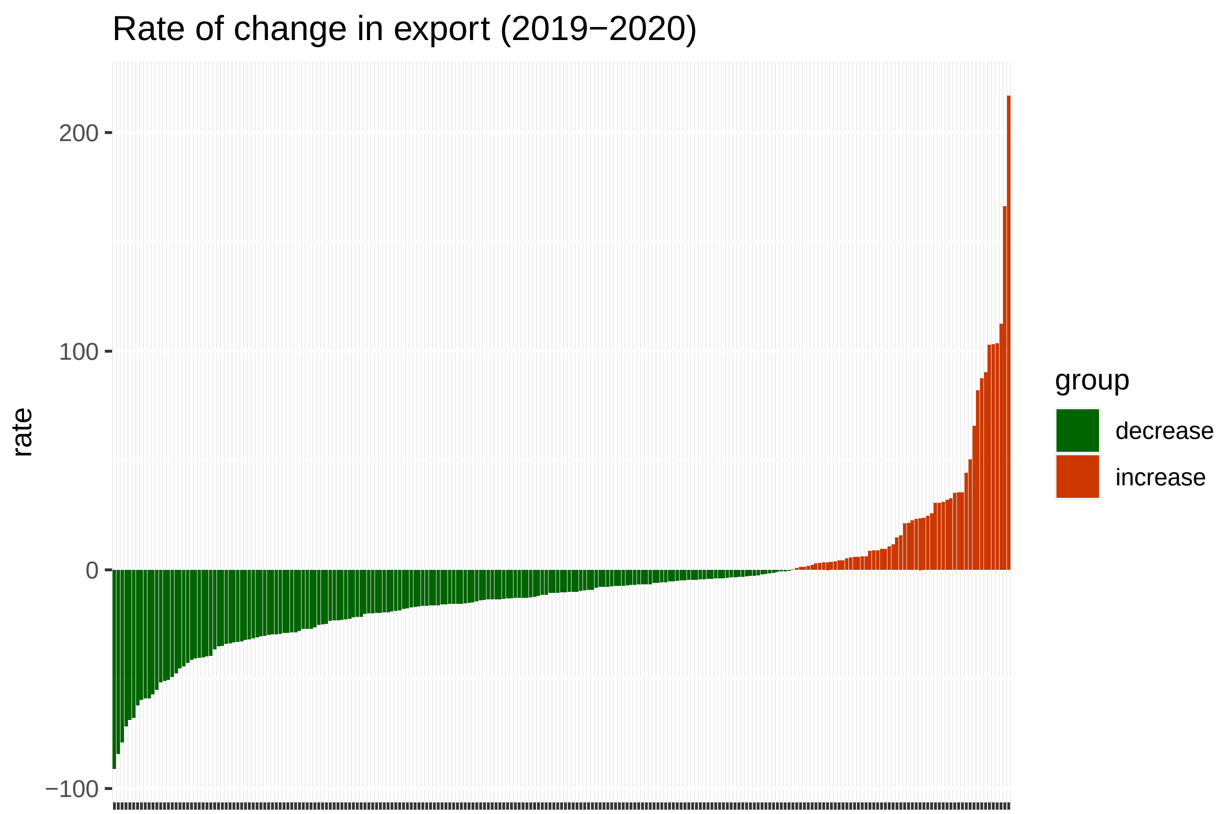
```
remove(ex21)
```

```
## Warning in remove(ex21):   'ex21'
```

This graph shows total exports by year. We can see that total exports decreased in 2020. Although trade field may have been affected by many socioeconomic factors, we can said that COVID-19 is the cause of the decrease in total exports because it has caused a huge change in the socioeconomic field. In addition, exports are higher in 2021 than in 2019. This means that many countries have overcome the COVID-19 situation and developed further.

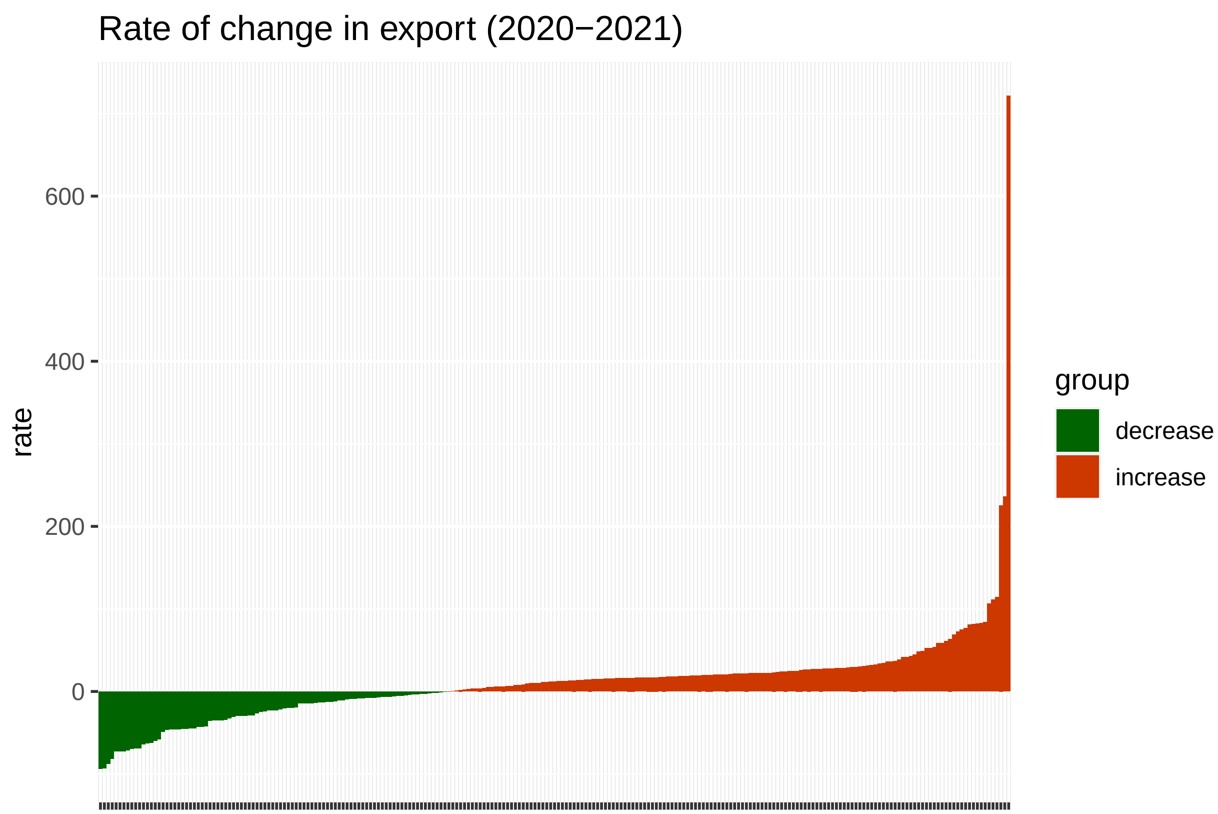
```
#separate dataset by year, calculate sum of export and arrange by country name
ex19<-df_global%>%
```

```r
  filter(Inquiry.Criteria==2019)%>%
  group_by(Export.country)%>%
  summarise(total_ex=sum(Export.amount..unit...))%>%
  arrange(Export.country)
ex20<-df_global%>%
  filter(Inquiry.Criteria==2020)%>%
  group_by(Export.country)%>%
  summarise(total_ex=sum(Export.amount..unit...))%>%
  arrange(Export.country)
ex21<-df_global%>%
  filter(Inquiry.Criteria==2021)%>%
  group_by(Export.country)%>%
  summarise(total_ex=sum(Export.amount..unit...))%>%
  arrange(Export.country)

#calculate the rate of change in total exports compared to the previous year
a<-round((ex20$total_ex-ex19$total_ex)/ex19$total_ex*100,3)
b<-round((ex21$total_ex-ex20$total_ex)/ex20$total_ex*100,3)

exrate1<-data.frame(country=ex20$Export.country, rate=a)
exrate2<-data.frame(country=ex21$Export.country, rate=b)

#Create column about groups based on zero
exrate1["group"]<-ifelse(exrate1$rate<0,"decrease","increase")
exrate2["group"]<-ifelse(exrate2$rate<0,"decrease","increase")

#create diverging bars about rate of change in export from 2019 to 2020
ggplot(data = exrate1,aes(x = reorder(country, rate), y = rate,fill=group))+
  geom_bar(stat = "identity", width=0.9)+
  scale_fill_manual(values = c("darkgreen", "orangered3"))+
  theme(axis.text.x = element_text(size=0.01))+
  ggtitle("Rate of change in export (2019-2020)")+
  xlab("")
```



Using the total amount of exports by year, the change rate was calculated by subtracting the previous year from the latest year and dividing the value by the previous year and multiplying by 100. Countries are divided into increasing and decreasing groups based on zero. And to see the rate of change in total trade exports, the x-axis with the country name is made invisible, and the rate of change is sorted in descending order. As can be seen from the previous graph, the majority of the rate of change in trade exports is decreasing from 2019 to 2020.

```
#create diverging bars about rate of change in export from 2020 to 2021
ggplot(data = exrate2,aes(x = reorder(country, rate), y = rate,fill=group))+
  geom_bar(stat = "identity", width=0.9)+
  scale_fill_manual(values = c("darkgreen", "orangered3"))+
  theme(axis.text.x = element_text(size=0.01))+
  ggtitle("Rate of change in export (2020-2021)")+
  xlab("")
```



From 2020 to 2021, there is more increase than decrease, and the degree of increase is also greater.

export

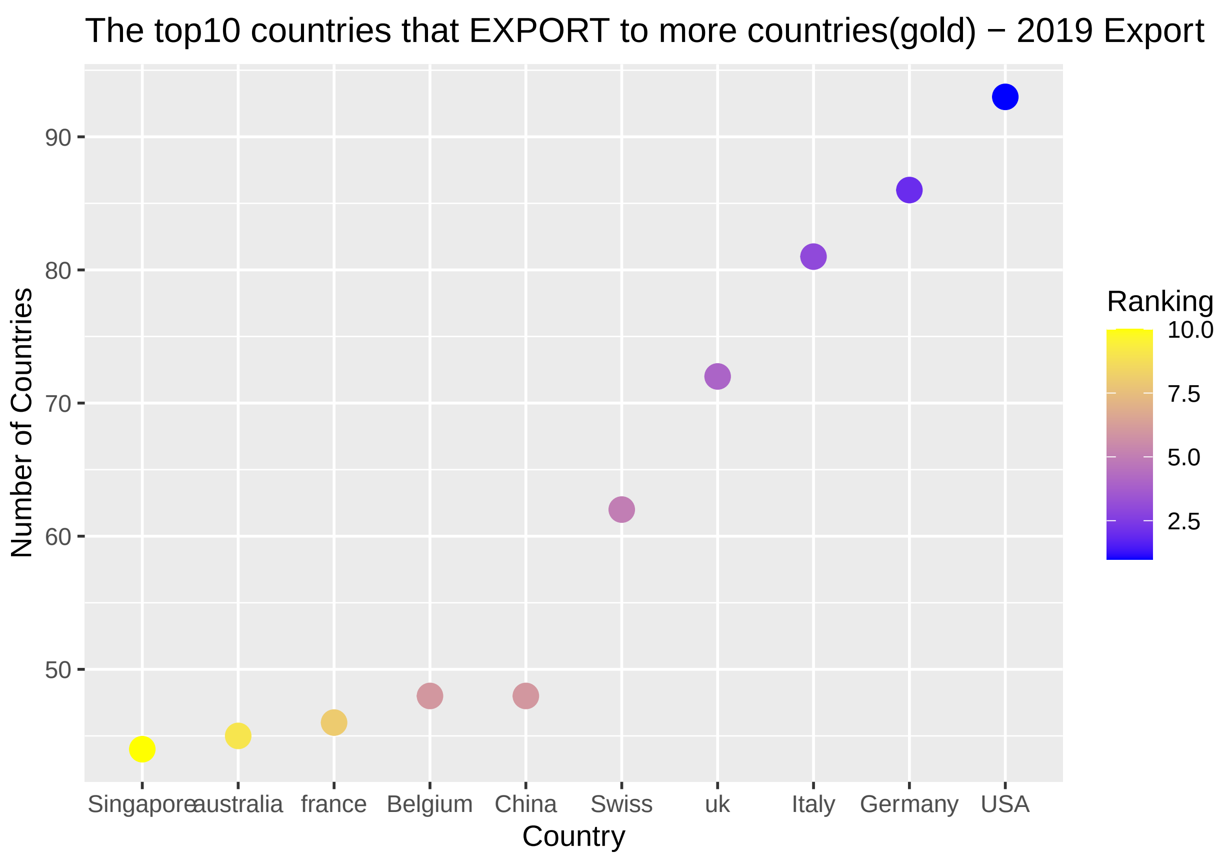
2019

```
#Ranking means : ranking = 1 means the country that exported for more different countries that year
#filter the dataframe in 3 columns (country name | number of countries it exports to | ranking)

df_19_rank_b <- df_19%>%filter(best_worst=="best")%>%group_by(Export.country, Import.country) %>%
  summarise(n_distinct(Import.country))  %>% tally() %>% mutate(rank=min_rank(desc(n)))
```

```
## `summarise()` has grouped output by 'Export.country'. You can override using
## the `.groups` argument.
```

```
#order the dataframe by ranking
df_19_rank_filtered_b<- df_19_rank_b%>%filter(rank<=10)
df_19_rank_filtered_b<-na.omit(df_19_rank_filtered_b[order(df_19_rank_filtered_b$n,decreasing=FALSE),])

#Change the country name column to factor
colnames(df_19_rank_filtered_b)[2] <- "Countries"
factor <- c(df_19_rank_filtered_b$Countries)
df_19_rank_filtered_b$Export.country <- factor(df_19_rank_filtered_b$Export.country

Ranking<-abs(desc(df_19_rank_filtered_b$rank))

#plotting the graph (countries | number of different countries that that country exports to)
```

```
ggplot(df_19_rank_filtered_b,aes(x=Export.country, y=Countries,color=Ranking))+
  geom_point(size=4)+
  xlab("Country")+
  ylab("Number of Countries")+
  ggtitle("The top10 countries that EXPORT to more countries(gold) - 2019 Export")+
  labs(color="Ranking")+
  scale_color_gradient(low="blue", high="yellow")
```




```
#Removed variables not important
remove(factor)
remove(Ranking)
remove(df_19_rank_b)
remove(df_19_rank_filtered_b)
```

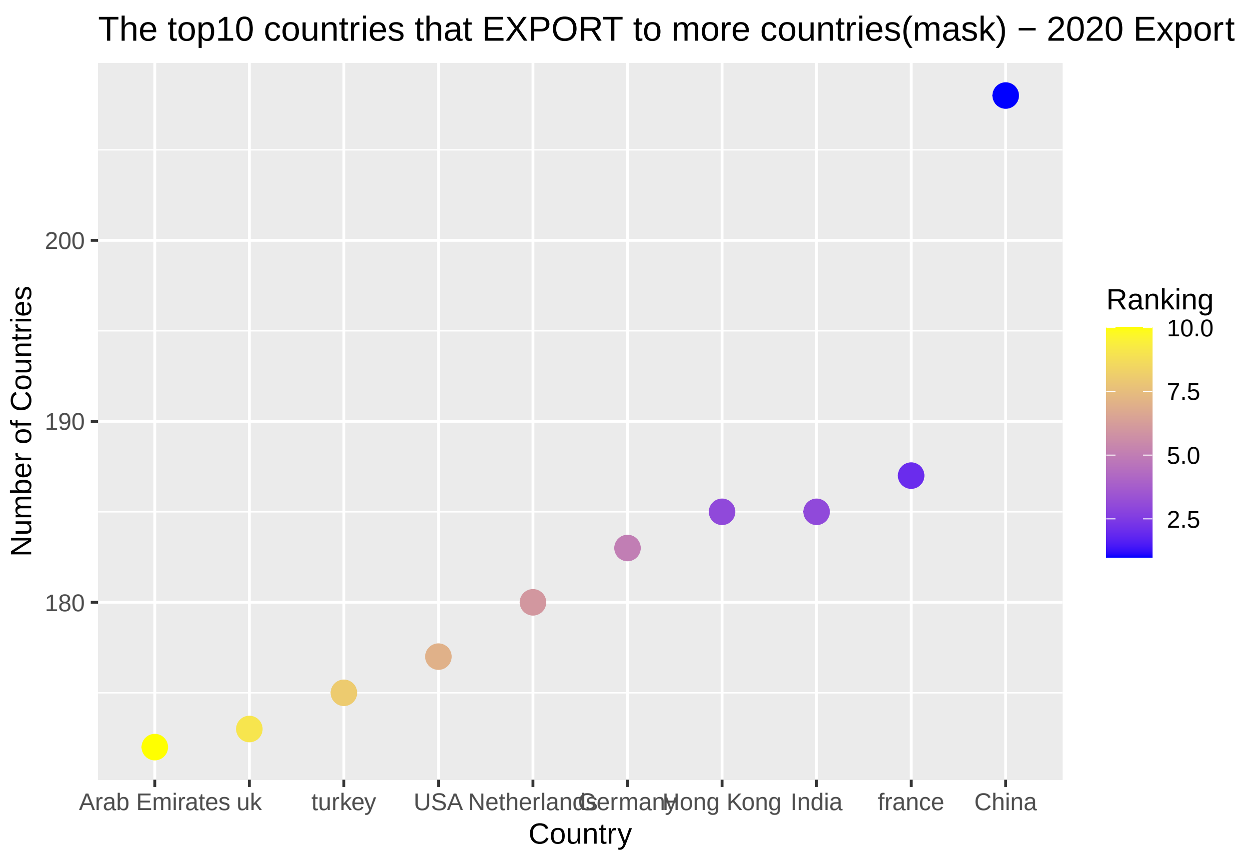
we extracted datasets of items with the highest export growth rate compared to the previous year from bw_datasets. And we select and visualized the top 10 percent of countries with a large number of countries exported. This graph show the top 10 countries that export to more countries. The item with the highest export growth rate compared to the previous year was gold, and the United States sold it to the largest number of countries.

2020

```
#Ranking means : ranking = 1 means the country that exported for more different countries that year
#filter the dataframe in 3 columns (country name | number of countries it exports to | ranking)

df_20_rank_b <- df_20%>%filter(best_worst=="best")%>%group_by(Export.country, Import.country) %>%
  summarise(n_distinct(Import.country))  %>% tally() %>% mutate(rank=min_rank(desc(n)))
```

```
## `summarise()` has grouped output by 'Export.country'. You can override using
## the `.groups` argument.
```

```r
#order the dataframe by ranking
df_20_rank_filtered_b<- df_20_rank_b%>%filter(rank<=10)
df_20_rank_filtered_b<-na.omit(df_20_rank_filtered_b[order(df_20_rank_filtered_b$n,decreasing=FALSE),])

#Change the country name column to factor
colnames(df_20_rank_filtered_b)[2] <- "Countries"
factor <- c(df_20_rank_filtered_b$Countries)
df_20_rank_filtered_b$Export.country <- factor(df_20_rank_filtered_b$Export.country

Ranking<-abs(desc(df_20_rank_filtered_b$rank))

#plotting the graph (countries | number of different countries that that country exports to)
ggplot(df_20_rank_filtered_b,aes(x=Export.country, y=Countries, color=Ranking))+
  geom_point(size=4)+
  xlab("Country")+
  ylab("Number of Countries")+
  ggtitle("The top10 countries that EXPORT to more countries(mask) - 2020 Export")+
  labs(color="Ranking")+
  scale_color_gradient(low="blue", high="yellow")
```




```r
#Removed variables not important
remove(factor)
remove(Ranking)
remove(df_20_rank_b)
```

```
remove(df_20_rank_filtered_b)
```

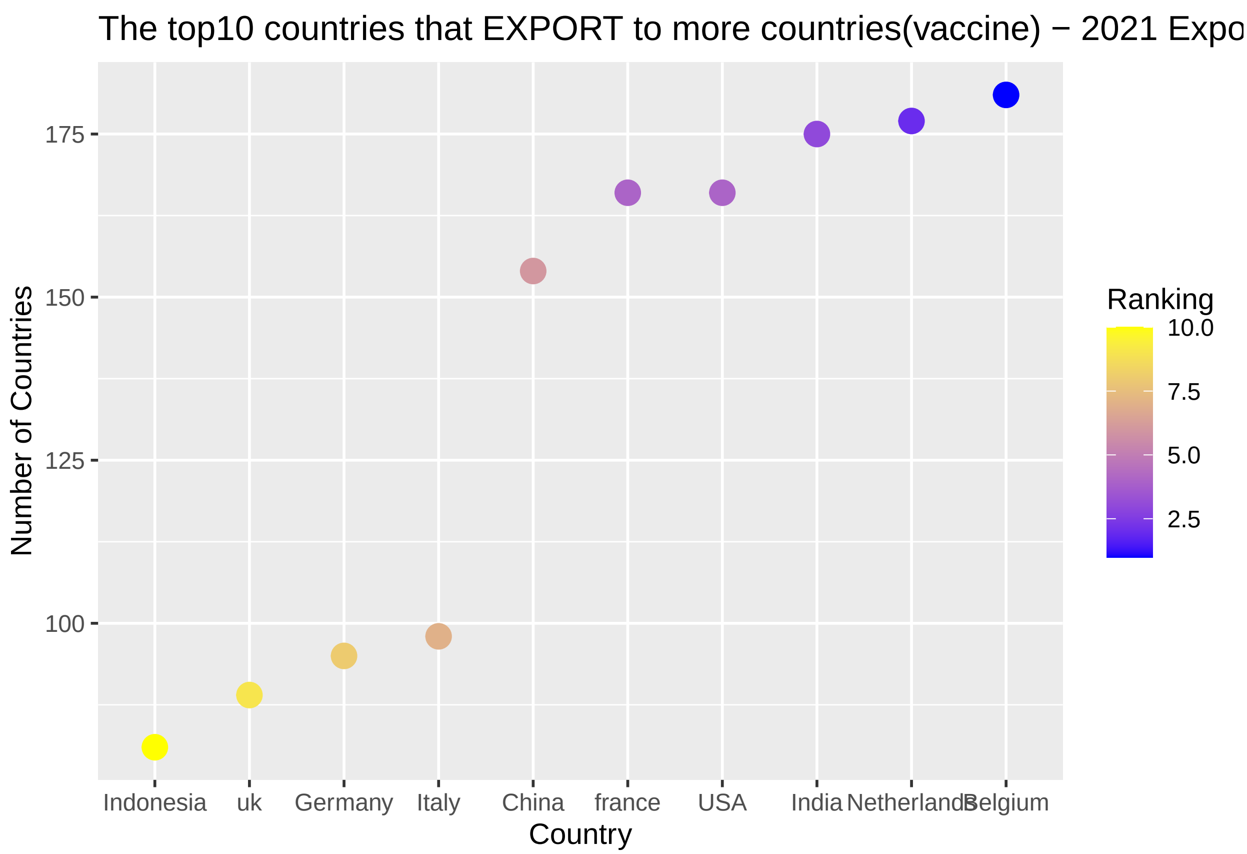
In 2020, as displayed in the graph, China exported to more than 200 different countries, while France explored to around 185. With this, we can also easily check for ourselves which item was also exported the most from this country, which also made sense in 2020 that the masks were the most exported items.

2021

```
#Ranking means : ranking = 1 means the country that exported for more different countries that year
#filter the dataframe in 3 columns (country name | number of countries it exports to | ranking)

df_21_rank_b <- df_21%>%filter(best_worst=="best")%>%group_by(Export.country, Import.country) %>%
  summarise(n_distinct(Import.country))  %>% tally() %>% mutate(rank=min_rank(desc(n)))
```

```
## `summarise()` has grouped output by 'Export.country'. You can override using
## the `.groups` argument.
```

```
#order the dataframe by ranking
df_21_rank_filtered_b<- df_21_rank_b%>%filter(rank<=10)
df_21_rank_filtered_b<-na.omit(df_21_rank_filtered_b[order(df_21_rank_filtered_b$n,decreasing=FALSE),])

#Change the country name column to factor
colnames(df_21_rank_filtered_b)[2] <- "Countries"
factor <- c(df_21_rank_filtered_b$Countries)
df_21_rank_filtered_b$Export.country <- factor(df_21_rank_filtered_b$Export.country

#plotting the graph (countries | number of different countries that that country exports to)
Ranking<-abs(desc(df_21_rank_filtered_b$rank))
ggplot(df_21_rank_filtered_b,aes(x=Export.country, y=Countries, color=Ranking))+
  geom_point(size=4)+
  xlab("Country")+
  ylab("Number of Countries")+
  ggtitle("The top10 countries that EXPORT to more countries(vaccine) - 2021 Export")+
  labs(color="Ranking")+
  scale_color_gradient(low="blue", high="yellow")
```




```
#Removed variables not important
remove(factor)
remove(Ranking)
remove(df_21_rank_b)
remove(df_21_rank_filtered_b)
```

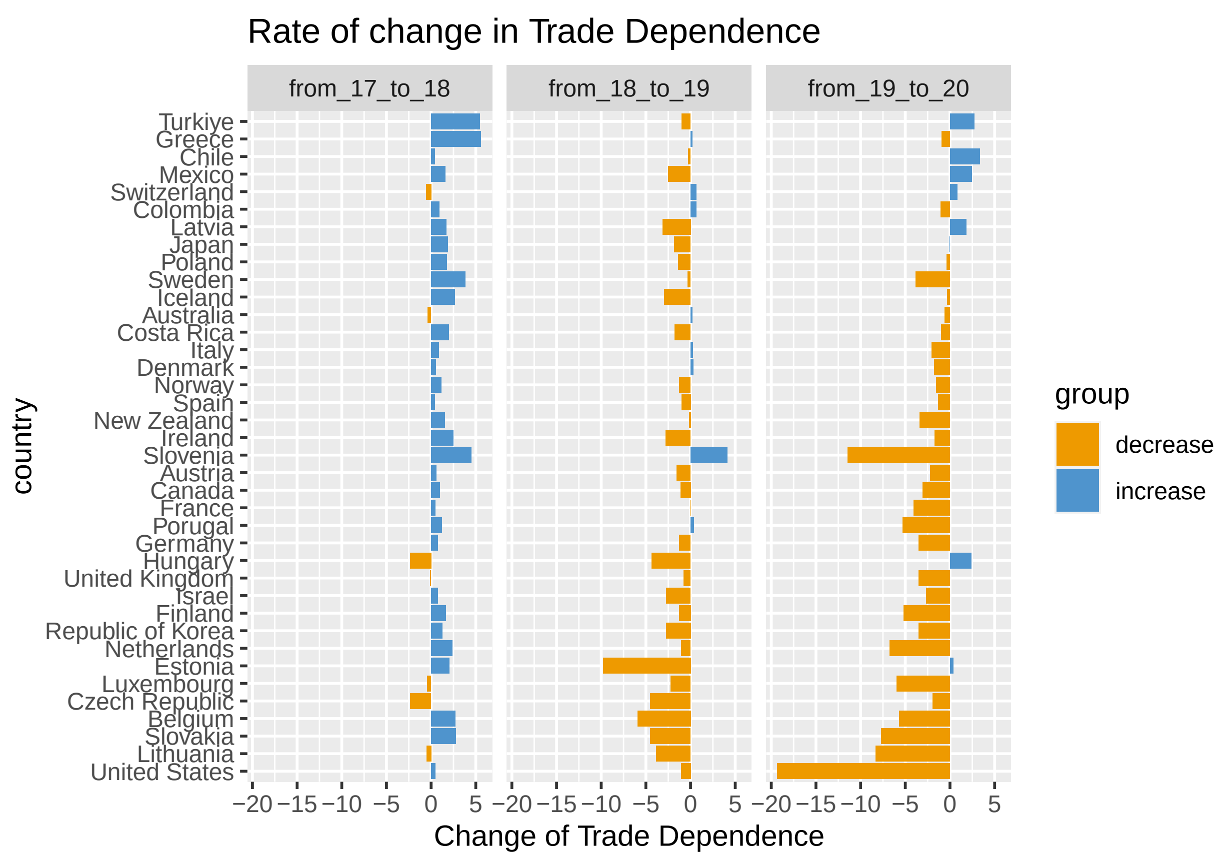
Vaccines are the items with the highest export growth rate in 2021. Among them, Belgium has exported the most to countries, as the biggest major areas of manufacturing in Belgium are pharmaceutical, and Pfizer and BNTX's factories are in Belgium. It can be seen that countries with high pharmaceutical-related capabilities are at the top.

Overall, Corona has made Changes in trade exports.

Change in Trade dependence

```
#separate dataset by year and arrange
rate17<-OECD%>%filter(year==2017)%>%select(year,country, Dependence)%>%arrange(country)
rate18<-OECD%>%filter(year==2018)%>%select(year,country, Dependence)%>%arrange(country)
rate19<-OECD%>%filter(year==2019)%>%select(year,country, Dependence)%>%arrange(country)
rate20<-OECD%>%filter(year==2020)%>%select(year,country, Dependence)%>%arrange(country)

#calculate rate of change in rate dependence compared to the previous year
from_17_to_18<-ifelse(rate17$Dependence> rate18$Dependence
                      ,-abs(rate17$Dependence - rate18$Dependence)
                      ,abs(rate17$Dependence-rate18$Dependence))
country<-rate18$country
```

```
from_18_to_19<-ifelse(rate18$Dependence> rate19$Dependence
                      , -abs(rate18$Dependence - rate19$Dependence)
                      ,abs(rate18$Dependence- rate19$Dependence))
from_19_to_20<-ifelse(rate19$Dependence > rate20$Dependence
                      , -abs(rate19$Dependence - rate20$Dependence)
                      , abs(rate19$Dependence - rate20$Dependence))
#set to zero change rate to country that do not have a GDP value
from_19_to_20[20]<-0
rate<-data.frame(country,from_17_to_18,from_18_to_19,from_19_to_20)

#change columns into value
rates<-melt(rate, id=c("country"),measure=c("from_17_to_18","from_18_to_19","from_19_to_20"))

#Create column about groups based on zero
rates["group"]<-ifelse(rates$value<0,"decrease","increase")

#create diverging bars about rate of change in trade dependence
ggplot(data=rates ,aes(x = reorder(country, value), y = value,fill=group))+
  geom_bar(stat = "identity", width=0.9)+
  coord_flip()+
  facet_wrap(~variable)+
  scale_fill_manual(values = c("orange2", "steelblue3"))+
  ggtitle("Rate of change in Trade Dependence")+
  xlab("country")+
  ylab("Change of Trade Dependence")
```



```r
#Removed variables not important
remove(rate)
remove(rate17)
remove(rate18)
remove(rate19)
remove(rate20)
remove(rates)
remove(from_17_to_18)
remove(from_18_to_19)
remove(from_19_to_20)
```

This graph is about changes in trade dependence by year. The change in trade dependence is calculated by determining the sign (+) and (-) if the trade dependence in the latest year is greater than that of the previous year or not and adding the sign to the difference in trade dependence in the two years. Since Japan's GDP value in 2020 did not exist, the difference in Japan's trade dependence in 2020 is set to 0. In order to divide the graph by year, we transformed a column with values for the year into categorical type data and add these values as new columns. In addition, the increased group and decreased group is divided based on 0. Blue is the group with increased trade dependence, and orange is the decreased group. Looking at this graph, the trade dependence of most countries decrease from 2018 to 2019 due to atmosphere of protectionism. Even The trade dependence decreases significantly from 2019 to 2020 because of COVID-19 which has deactivated global trade. Since High trade dependence means low economic independence of the country, a decrease in trade dependence means that Corona has had a positive impact on the trade.

correlation/ linear regression test(unemployment rate & trade dependence)

```r
#Put column values in variables
dependence=OECD$Dependence
unemployment_rate=OECD$Unemployment_rate

#Remove rows that could not calculate trade dependence due to lack of GDP value
dependence<-dependence[-117]
unemployment_rate<-unemployment_rate[-117]

#Pearsonr analysis between trade dependence and unemployment rate
cor.test(dependence,unemployment_rate, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  dependence and unemployment_rate
## t = -2.6272, df = 149, p-value = 0.00951
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.35810053 -0.05244107
## sample estimates:
##        cor
## -0.2104071
```

```r
#linear regression analysis
summary(lm(dependence~unemployment_rate))
```

```
##
## Call:
## lm(formula = dependence ~ unemployment_rate)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -66.89 -29.62 -10.43  12.31 104.85
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        86.2195     6.7925  12.693  < 2e-16 ***
## unemployment_rate  -2.3866     0.9084  -2.627  0.00951 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40.65 on 149 degrees of freedom
## Multiple R-squared:  0.04427,    Adjusted R-squared:  0.03786
## F-statistic: 6.902 on 1 and 149 DF,  p-value: 0.00951
```

The trade dependence column and the unemployment rate column are extracted from the OECD data set. Since Japan's GDP value in 2020 did not exist, the trade dependence can't be not calculated. so the value is deleted. The unemployment rate is also deleted to maintain the same length of the vector. And as a result of performing the correlation analysis, the p-value is 0.00, which is lower than 0.05, and it means that the two variables have a statistically significant correlation. It is found that the correlation coefficient is -0.21 and two variables have a weak negative correlation. Since trade dependence and unemployment rate have a statistically significant correlation, we do a linear regression analysis between trade dependence (dependent variable) and unemployment rate(independent variable). In the results, the p-value is lower than 0.05, so this model is statistically significant. In addition, when the unemployment rate rises by 1, the trade dependence decreases by about -23.86. Considering that the R-squared value is 0.03 to 0.04, the unemployment rate explain only 4% of the trade dependence. In other words, it can not be a variable that directly affects trade dependence.

correlation/ linear regression test(future population & trade dependence)

```r
#Put column values in variables
dependence=OECD$Dependence
future_population=OECD$Future_population

#Remove rows that could not calculate trade dependence due to lack of GDP value
dependence<-dependence[-117]
future_population<-future_population[-117]

#Pearsonr analysis between trade dependence and Future population
cor.test(dependence,future_population, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  dependence and future_population
## t = -5.1683, df = 149, p-value = 7.477e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5174003 -0.2454516
## sample estimates:
##        cor
## -0.3898941
```

```r
#linear regression analysis
summary(lm(dependence~future_population))
```

```
##
## Call:
## lm(formula = dependence ~ future_population)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -50.41 -27.98 -13.18  18.99  95.65
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        8.037e+01  3.641e+00  22.074  < 2e-16 ***
## future_population -2.767e-04  5.353e-05  -5.168 7.48e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.29 on 149 degrees of freedom
## Multiple R-squared:  0.152,  Adjusted R-squared:  0.1463
## F-statistic: 26.71 on 1 and 149 DF,  p-value: 7.477e-07
```

Next, the trade dependence column and future population column are extracted from the OECD data set, and as mentioned above, the value of the location is deleted because Japan's GDP value in 2020 did not exist. And as a result of the correlation analysis, it is found that because the p-value was 0.00 lower than 0.05 and the correlation coefficient was -0.38, the two variables have a statistically significant correlation and a general negative correlation. Based on this information, we do a linear regression analysis between trade dependence (dependent variable) and future population (independent variable). In the results, the p-value is lower than 0.05, so the model is statistically significant. In addition, when the future population rises by this 1, the trade dependence decreases by about -0.002. Given the R-squared value of 14-15, the future population only explain 15% of the trade dependence. That is, it can't be a variable that directly affects trade dependence.

Based on the linear regression model of the unemployment rate and future population for trade dependence, we found out that there will be not only two variables but also other variables that affect trade dependence. However, since the linear regression model was statistically significant and the p-value was lower than 0.05 in the correlation analysis between the two variables(unemployment rate & future population) and trade dependence, the two variables and trade dependence would have had a significant relationship.

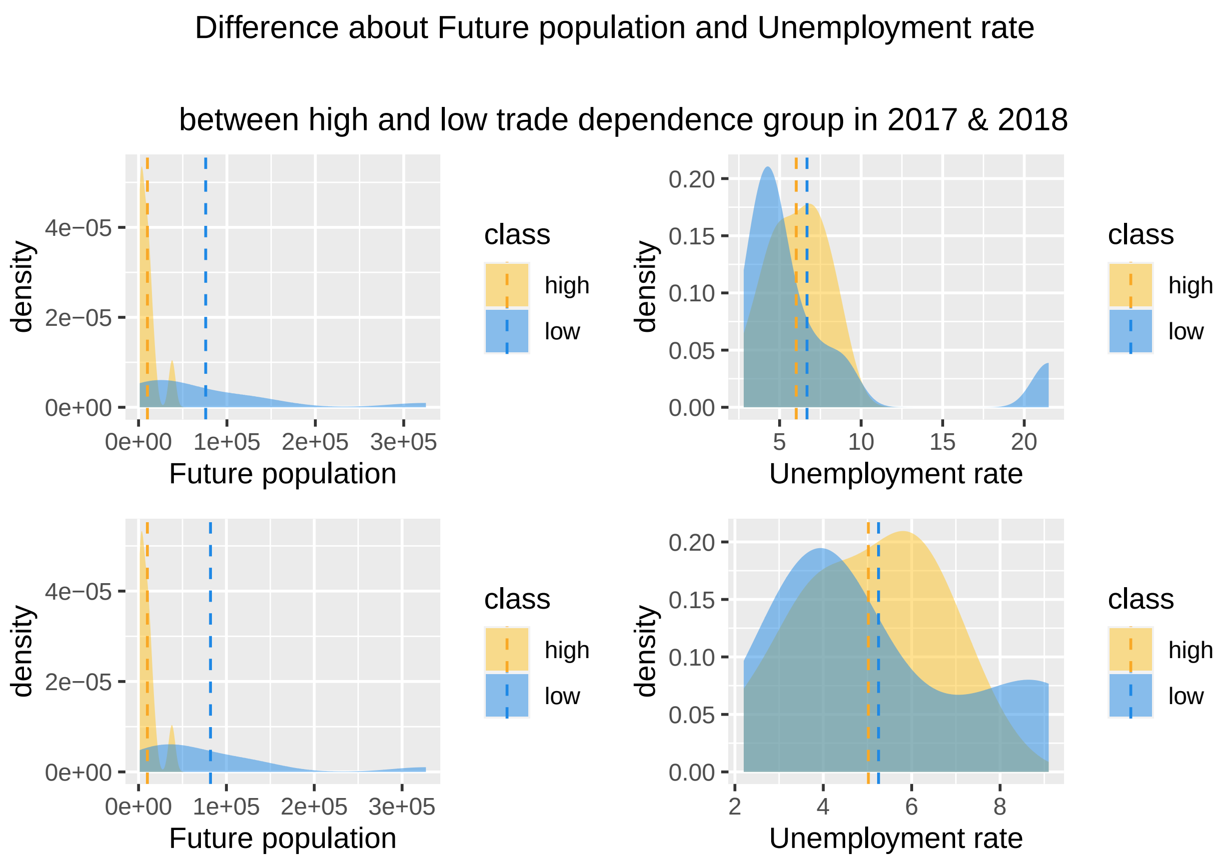
Difference about Future population and Unemployment rate between high and low trade dependence group

---

```r
#extract dataset whose year is 2017
OECD17<-subset(OECD, year==2017)

#Extract top 10 countries with higher trade dependence than average
high1<-OECD17%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate)
            ,total_dependence=sum(Dependence))%>%
  arrange(desc(total_dependence))%>%head(10)
high1['class']<-rep('high',10)

#Extract top 10 countries with less than trade dependence than average
```

```r
low1<-OECD17%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate)
            ,total_dependence=sum(Dependence))%>%
  arrange(total_dependence)%>%head(10)
low1['class']<-rep('low',10)

#extract dataset whose year is 2018
OECD18<-subset(OECD, year==2018)

#Extract top 10 countries with higher trade dependence than average
high2<-OECD18%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(desc(total_dependence))%>%head(10)
high2['class']<-rep('high',10)

#Extract top 10 countries with less than trade dependence than average
low2<-OECD18%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(total_dependence)%>%head(10)
low2['class']<-rep('low',10)

#Combine datasets for high-dependent and low-dependent countries
total1<-rbind(high1,low1)
total2<-rbind(high2,low2)

#calculate the mean about future population and unemployment rate by group
means1<-total1%>%group_by(class)%>%summarise(FP1.mean=mean(total_Future_population)
                                             ,UR1.mean=mean(total_unemployment_rate))

means2<-total2%>%group_by(class)%>%summarise(FP2.mean=mean(total_Future_population)
                                             ,UR2.mean=mean(total_unemployment_rate))

#Extract mean values and classes
FP1<-means1%>%select(class, FP1.mean)
UR1<-means1%>%select(class, UR1.mean)

FP2<-means2%>%select(class, FP2.mean)
UR2<-means2%>%select(class, UR2.mean)

#create density plot about future population and unemployment by group in 2017&2018

a = ggplot(total1, aes(x = total_Future_population, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=FP1, aes(xintercept=FP1.mean, color=class),linetype="dashed")+
```

```
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Future population")

b = ggplot(total1, aes(x = total_unemployment_rate, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=UR1, aes(xintercept=UR1.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Unemployment rate")

c = ggplot(total2, aes(x = total_Future_population, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=FP2, aes(xintercept=FP2.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Future population")

d = ggplot(total2, aes(x = total_unemployment_rate, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=UR2, aes(xintercept=UR2.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Unemployment rate")


grid.arrange(a,b,c,d, nrow=2, ncol=2, top="Difference about Future population and Unemployment rate
             \n between high and low trade dependence group in 2017 & 2018")
```

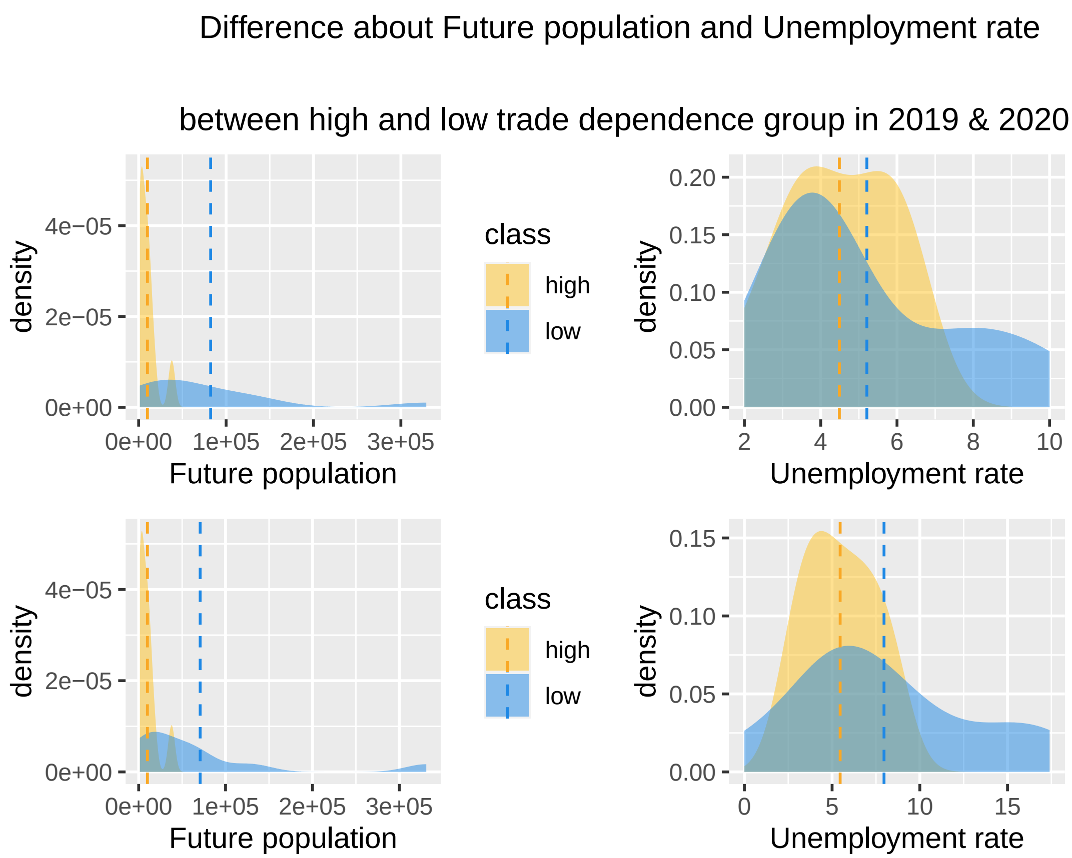



```
remove(a)
remove(b)
remove(UR1)
remove(FP1)
remove(UR2)
remove(FP2)
remove(high1)
remove(low1)
remove(high2)
remove(low2)
remove(means1)
remove(means2)
remove(OECD17)
remove(OECD18)
remove(total1)
remove(total2)
```

To further clarify the above view, we extract the top 10 countries with high trade dependence and the bottom 10 countries with low trade dependence to visualize the difference between the unemployment rates of the two groups and the future population. This graph is from 2017 and 2018. Blue is the group with low trade dependence, yellow is the group with high trade dependence, and the dotted line represents the average of the group's unemployment rate and future population values. Judging from the average line, it can be seen that countries with low trade dependence have relatively high unemployment rates and future populations. However, it is difficult to say that the unemployment rate has a remarkable difference between countries with low trade dependence and countries with high trade dependence. Rather, at the midpoint, countries with high trade dependence tend to have higher unemployment rates.

```r
#extract dataset whose year is 2019
OECD17<-subset(OECD, year==2019)

#Extract top 10 countries with higher trade dependence than average
high1<-OECD17%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population)
            ,total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(desc(total_dependence))%>%head(10)
high1['class']<-rep('high',10)

#Extract top 10 countries with less than trade dependence than average
low1<-OECD17%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(total_dependence)%>%head(10)
low1['class']<-rep('low',10)

#extract dataset whose year is 2020
OECD18<-subset(OECD, year==2020)

#Extract top 10 countries with higher trade dependence than average
high2<-OECD18%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(desc(total_dependence))%>%head(10)
high2['class']<-rep('high',10)

#Extract top 10 countries with less than trade dependence than average
low2<-OECD18%>%select(year, country, Future_population, Unemployment_rate,Dependence)%>%
  group_by(country)%>%
  summarise(total_Future_population=sum(Future_population),
            total_unemployment_rate=sum(Unemployment_rate),
            total_dependence=sum(Dependence))%>%
  arrange(total_dependence)%>%head(10)
low2['class']<-rep('low',10)

#Combine datasets for high-dependent and low-dependent countries
total1<-rbind(high1,low1)
total2<-rbind(high2,low2)

#calculate the mean about future population and unemployment rate by group
means1<-total1%>%group_by(class)%>%summarise(FP1.mean=mean(total_Future_population)
                                            ,UR1.mean=mean(total_unemployment_rate))

means2<-total2%>%group_by(class)%>%summarise(FP2.mean=mean(total_Future_population)
                                            ,UR2.mean=mean(total_unemployment_rate))
```

```r
#Extract mean values and classes
FP1<-means1%>%select(class, FP1.mean)
UR1<-means1%>%select(class, UR1.mean)

FP2<-means2%>%select(class, FP2.mean)
UR2<-means2%>%select(class, UR2.mean)

#create density plot about future population and unemployment by group in 2019&2020

a = ggplot(total1, aes(x = total_Future_population, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=FP1, aes(xintercept=FP1.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Future population")

b = ggplot(total1, aes(x = total_unemployment_rate, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=UR1, aes(xintercept=UR1.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Unemployment rate")

c = ggplot(total2, aes(x = total_Future_population, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=FP2, aes(xintercept=FP2.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Future population")

d = ggplot(total2, aes(x = total_unemployment_rate, fill=class)) +
  geom_density(color=NA,alpha=0.5)+
  scale_fill_manual(values=c("#ffc425", "#1E88E5"))+
  geom_vline(data=UR2, aes(xintercept=UR2.mean, color=class),linetype="dashed")+
  scale_color_manual(values=c("#F9A825", "#1E88E5"))+
  xlab("Unemployment rate")


grid.arrange(a,b,c,d, nrow=2, ncol=2, top="Difference about Future population and Unemployment rate
             \n between high and low trade dependence group in 2019 & 2020")
```




```
remove(a)
remove(b)
remove(UR1)
remove(FP1)
remove(UR2)
remove(FP2)
remove(high1)
remove(low1)
remove(high2)
remove(low2)
remove(means1)
remove(means2)
remove(OECD19)
```

```
## Warning in remove(OECD19):   'OECD19'
```

```
remove(OECD20)
```

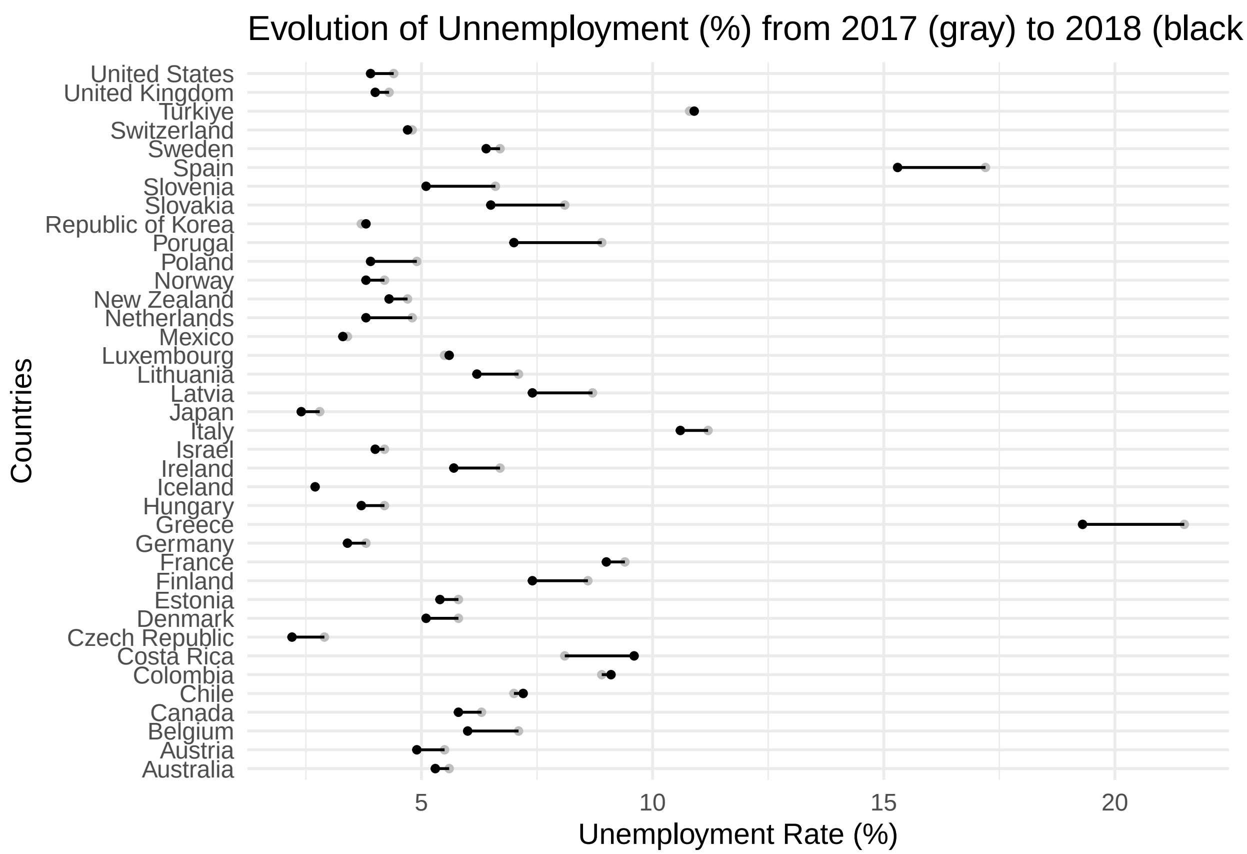
```
## Warning in remove(OECD20):   'OECD20'
```

```
remove(total1)
remove(total2)
```

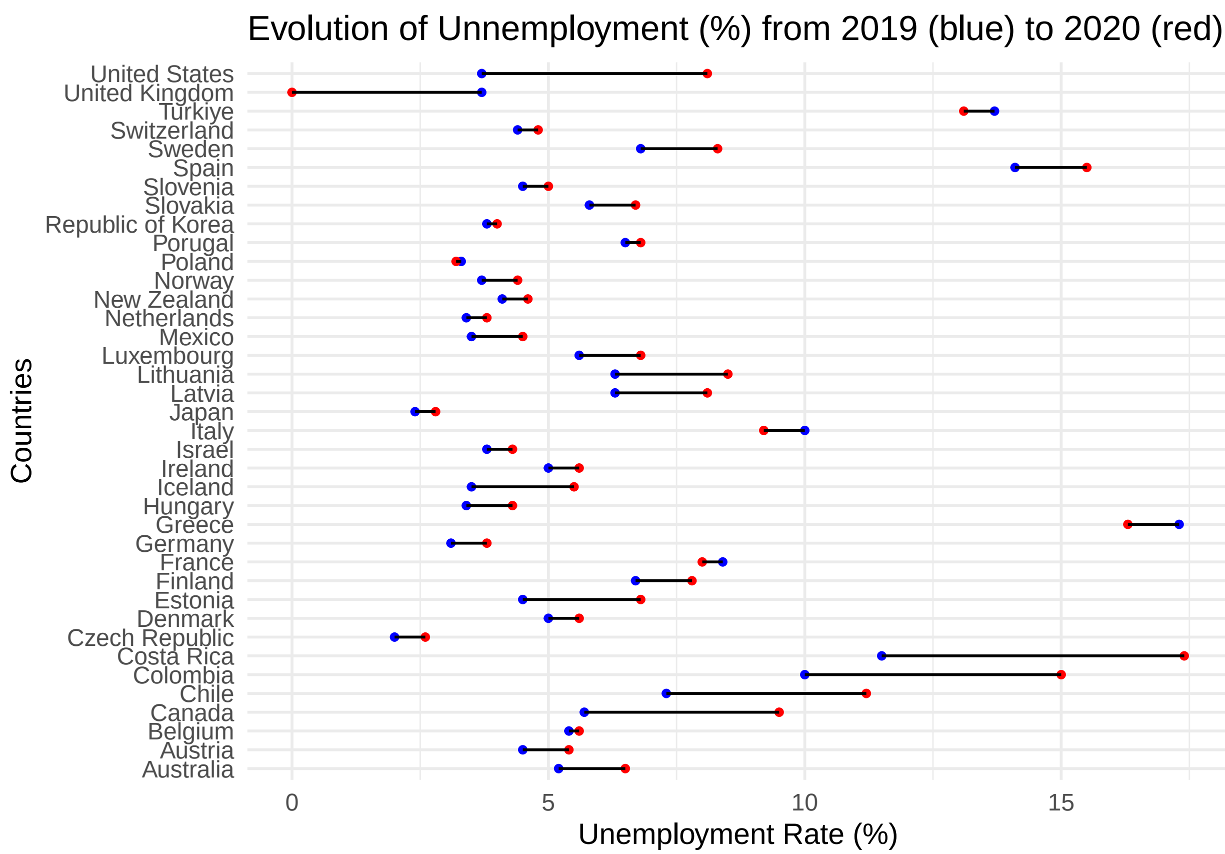
This graph is for 2019 and 2020. As time goes by, it can be seen that the gap between the average unemployment rate in countries with high dependence and countries with low dependence is getting bigger. In conclusion, although the unemployment rate and future population cannot be seen as variables that completely explain trade dependence, but they have some meaningful relationship with trade dependence and can be figures that represent the characteristics of trade dependence.

unemployment rate

```r
#clearly visible that the unnemployment have been increased a lot

#Separating dataframes by years
OECD_17<-OECD%>%filter(year==2017)
OECD_18<-OECD%>%filter(year==2018)
OECD_19<-OECD%>%filter(year==2019)
OECD_20<-OECD%>%filter(year==2020)

#Merging the datasets in order for them all contain the column with the unemployment rate of each year
df<-OECD_17
df['Unemployment_rate2018'] <- OECD_18$Unemployment_rate
df['Unemployment_rate2019'] <- OECD_19$Unemployment_rate
df['Unemployment_rate2020'] <- OECD_20$Unemployment_rate

#chose colors to be ploted grey for 2017 unemployment black for 2018
grey = c(df$Unemployment_rate)
black = c(df$Unemployment_rate2018)

#create a small dataframe to be more efficient to plot
countries = c(df$country)
graph = data.frame(grey, black, countries)

#plotting
graph %>%
  pivot_longer(c("grey", "black")) %>%
  ggplot() +
  geom_point(aes(x = value, y = countries, color = name), size = 1) +
  scale_color_identity() +
  xlab("Unemployment Rate (%)") +
  geom_line(aes(x = value, y = countries)) +
  theme_minimal() +
  ylab("Countries") +
  ggtitle("Evolution of Unnemployment (%) from 2017 (gray) to 2018 (black)")
```




```
#chose colors to be ploted blue for 2019 unemployment red for 2020
blue = c(df$Unemployment_rate2019)
red = c(df$Unemployment_rate2020)

countries = c(df$country)
graph = data.frame(blue, red, countries)

#plotting 2nd graph
graph %>%
  pivot_longer(c("blue", "red")) %>%
  ggplot() +
  geom_point(aes(x = value, y = countries, color = name), size = 1) +
  scale_color_identity() +
  xlab("Unemployment Rate (%)") +
  geom_line(aes(x = value, y = countries)) +
  theme_minimal() +
  ylab("Countries") +
  ggtitle("Evolution of Unemployment (%) from 2019 (blue) to 2020 (red)")
```




```
#remove un-important variables used
remove(black)
remove(blue)
remove(countries)
remove(grey)
remove(red)
remove(graph)
remove(OECD_17)
remove(OECD_18)
remove(OECD_19)
remove(OECD_20)
```

After dividing the unemployment rate data by year, we visualized the changes in the unemployment rate by specifying different colors for each year. This graph shows Evolution of Unemployment (%) from 2017 (blue | grey) to 2018 and from 2019 to 2020(red | black). From 2017 to 2018, the unemployment rate in most countries is decreasing. On the contrary, it is increasing from 2019 to 2020.

On the second graph, since the unemployment rate increases, we can see that dozens of countries got a huge impact on their unemployment rate during that year when COIVD-19 spread. It is also noticeable that South Korea on both graphs represents a stabilized unemployment rate, which can be related to a "strong auto-sustainable economy" based on our analysis.

[Difficulties and solutions]

• With the main dataset, we can only know the flow of trade, and can't know information on various variables or figures related to trade. -> solution : We tried to understand the relationship between trade dependence and other variables through a sub-dataset about OECD countries.

- We have difficult to understanding the linear regression results. -> solution : Through googling, we studied the meaning of the numbers from linear regression results.

[limit]

- The number of rows in the OECD dataset is too small to generalize the results of our analysis.

[Conclusion]

- There was a negative impact of a decrease in trade exports due to COVID-19. But, Corona had a positive impact by helping the countries' economic independence by reducing their trade dependence.

- Because R-squared values of results is very small and the correlation coefficient is not large enough, the two variables(future population, unemployment rate) can't explain trade dependence, and it is highly likely that trade dependence will be determined due to the influence of not only these two variables but also other variables. But on the basis of p-value which is lower than 0.05 and our analysis with graph, these two variables can be said that they have some significant relationship with trade dependence and they are indicators representing the feature of trade dependence.

- Trade dependence is not just a trade-related field, but a figure formed based on various fields such as population and society. Therefore, in order to increase the country's economic independence, the country must come up with policies to reduce trade dependence from various perspectives.