Data science 1 page

Global Trade

JOÃO VASCO
Dayun Kang /강다윤
Suhwa Kwon /권수화
Hongcheon Liu /유홍천

content

Data science 2 page

Introduction

Data science 3 page

		Size of datasets
global_trade_2019_2021.csv	Main dataset	83903 rows, 8 columns
2020_bw.csv	Sub dataset	17314 rows, 9 columns
2019_bw.csv	Sub dataset	5164 rows, 9 columns
2021_bw.csv	Sub dataset	3611 rows, 9 columns
OECD_Major_Indicator.csv	Sub dataset	180 rows, 12 columns

Source : Kotra trade investment big data, global trade status statistics data,
Korean statistical information service(KOSIS)

Preliminary Findings

Data science 4 page

- cleaning

Changing OECD dataset columns

Gross.domestic.product..GDP...current.year.s.price...US..1.billion.GDP. -> GDP
GDP.Growth.... -> GDP_growth
Export...1.million. -> Export
Earnings...1.million. -> Import
Future.population..1.000.people. -> Future_population
Etc…

Change “-” to “0” / omit Nas / delete “ , ”

```
country                : chr
year                   : int
17 ...
GDP                    : chr
GDP_per_capita         : chr
GDP_growth             : chr
Export                 : chr
Import                 : chr
Future_population      : chr
Unemployment_rate      : chr
Consumer_price         : chr
Crude_steel_production: chr
Internet_usage_rate    : chr
```

← Convert to numeric type

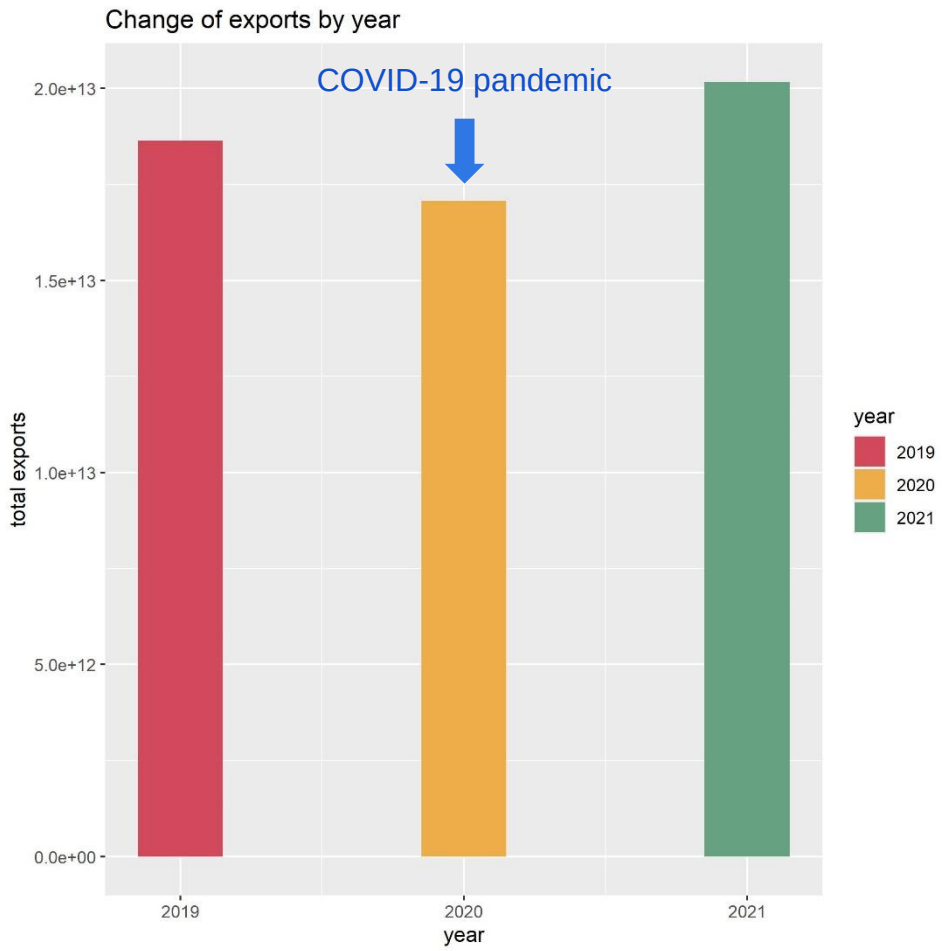
Library we used

- Ggplot2
- Reshape2
- Dplyr
- GridExtra
- Tidyverse
- Magrittr

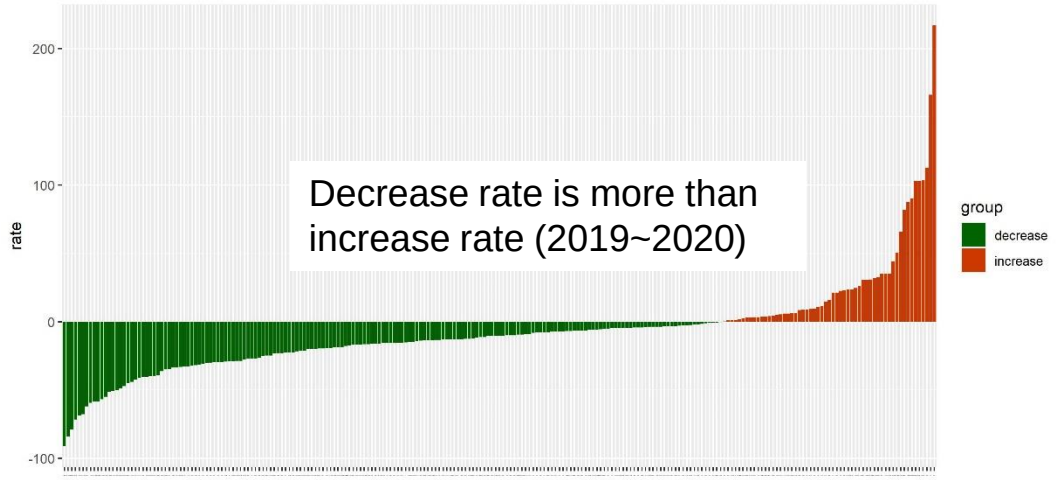
Preliminary Findings

Data science 5 page

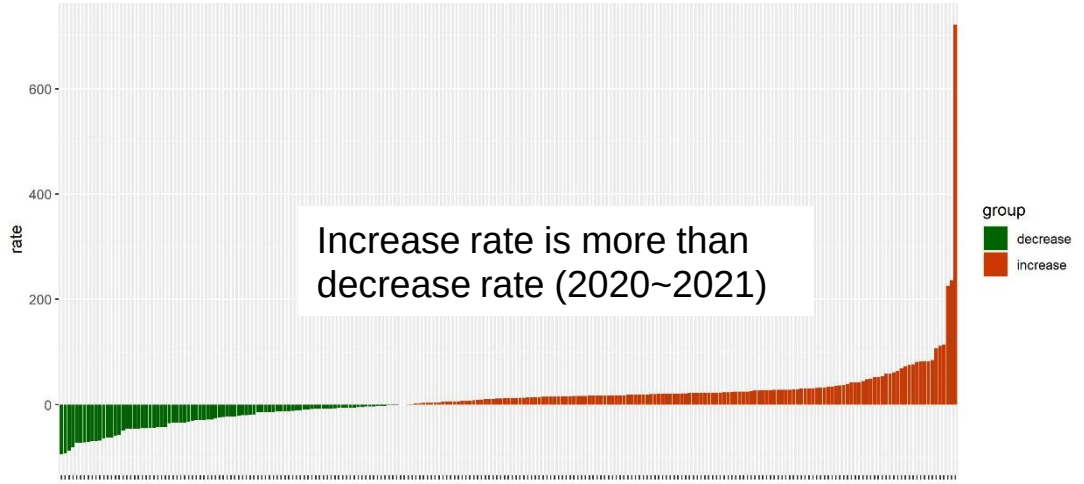
- Basic Stats



Change rate of Export 2019-2020



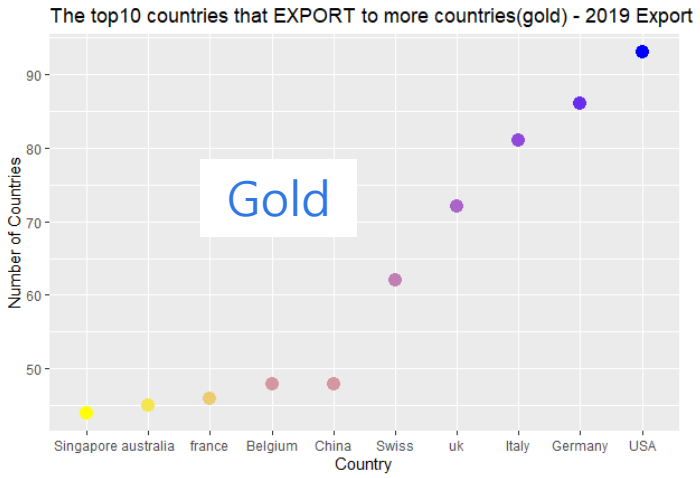
Change rate of Export 2020-2021

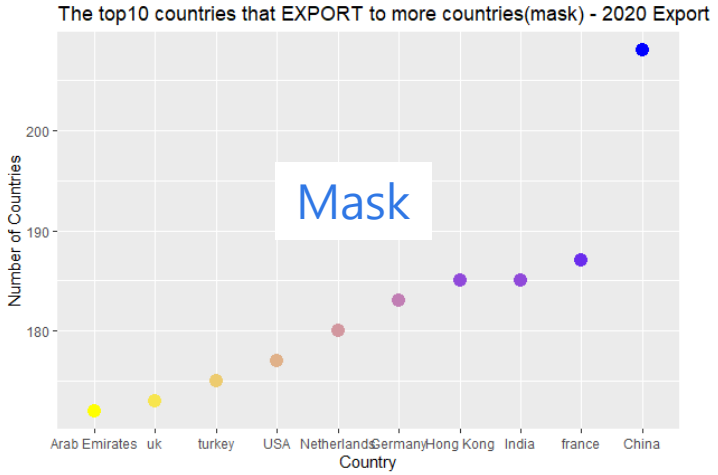


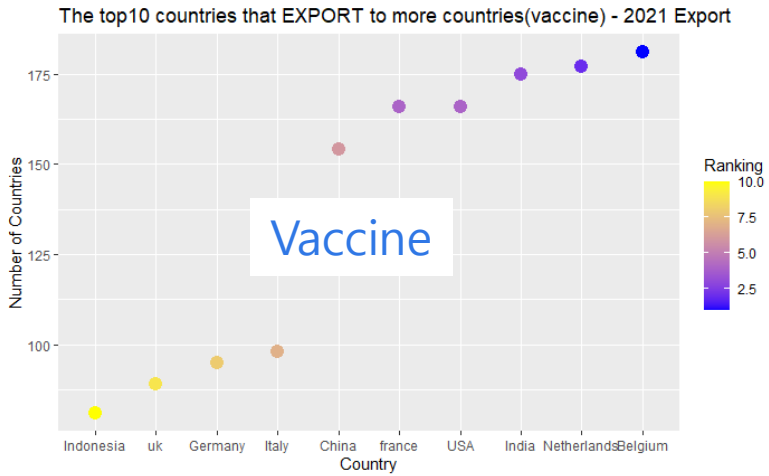
Preliminary Findings

Data science 6 page

- Basic Stats The top 10 countries that EXPORT to more countries (2019-2021)







Hypothesis/ Question

Hypothesis :

"Corona would have had a negative impact on world trade."

Question :

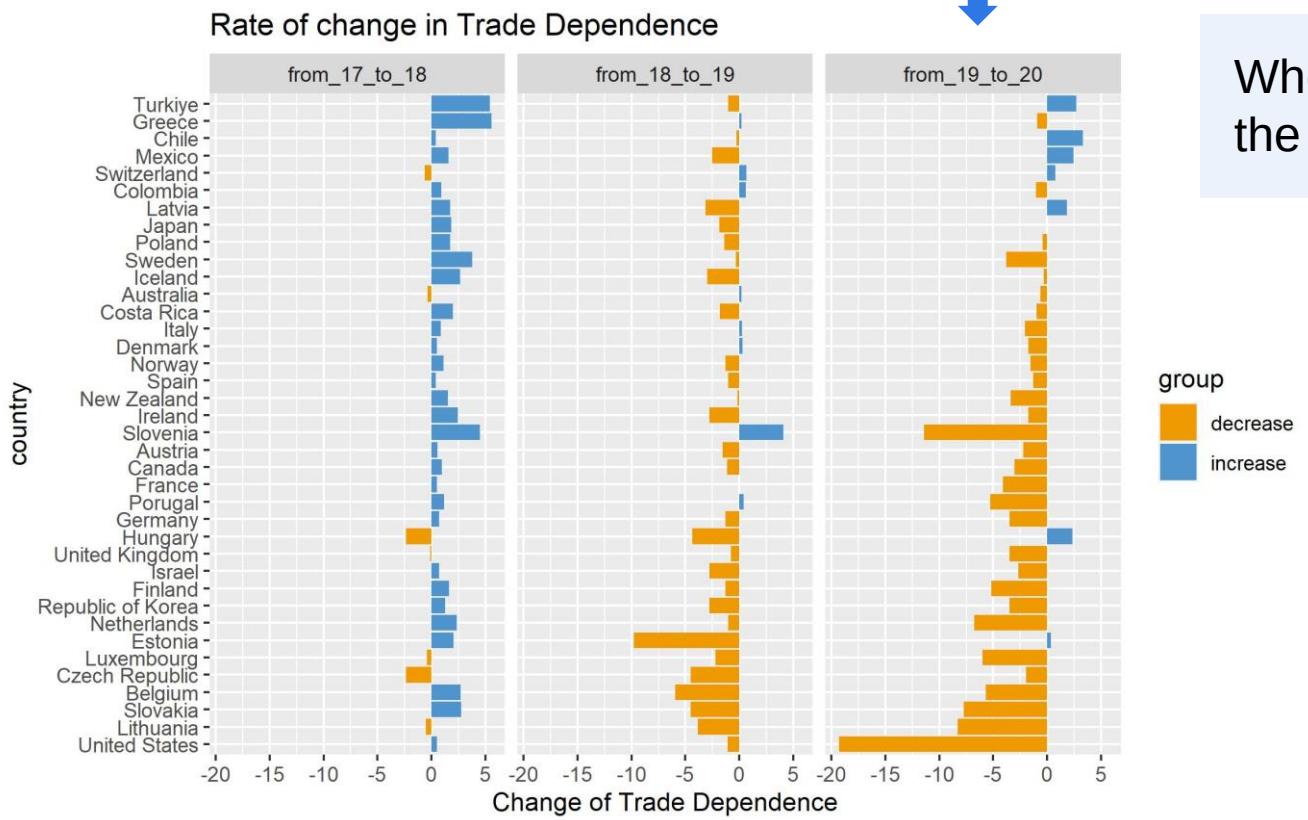
"Do unemployment rate and future population affect determining trade dependence?"

$$\text{Trade Dependence} = \frac{\text{Export + Import}}{\text{GDP}}$$

→ Indicator of the degree of dependence on trade in a country's economy

Change of Trade Dependence by year

COVID-19 pandemic



When the COVID-19 pandemic flourished,
the rate of decrease in trade dependence increased

Correlation analysis

-Unemployment and Trade dependence/ Future population and Trade dependence

Unemployment rate

```
        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
```

Future population

```
        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
```

The two variables have a statistically significant correlation with trade dependence

Linear regression analysis

-Unemployment and Trade dependence/ Future population and Trade dependence

Unemployment rate

```
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
```

Future population

```
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
```

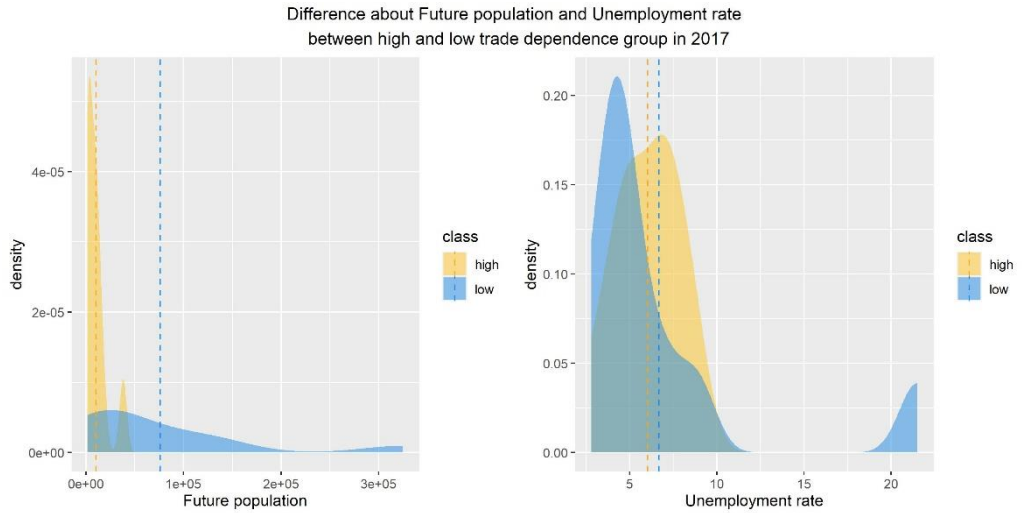
The linear model of the two variables is statistically significant.
However, it is difficult to say that the two variables explain trade dependence due to the small value of the Adjusted R-squared

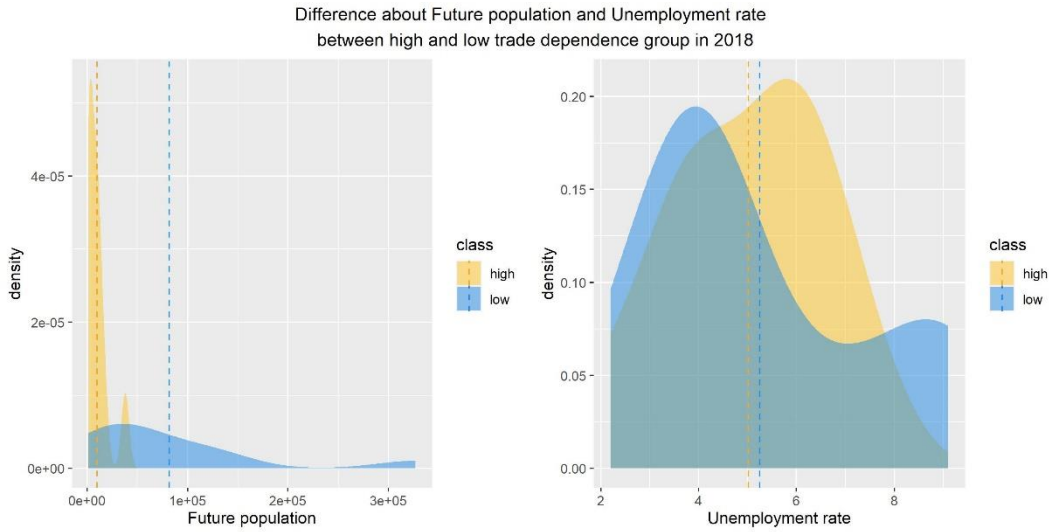
EDA

Data science 11 page

Visualization

-Between Top 10 high trade dependence group(yellow) and Top 10 low trade dependence group(blue)



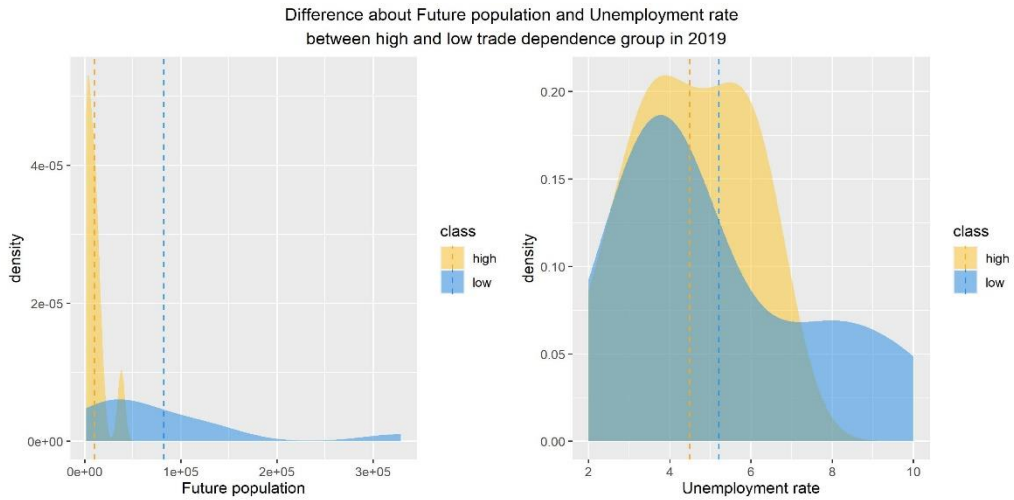


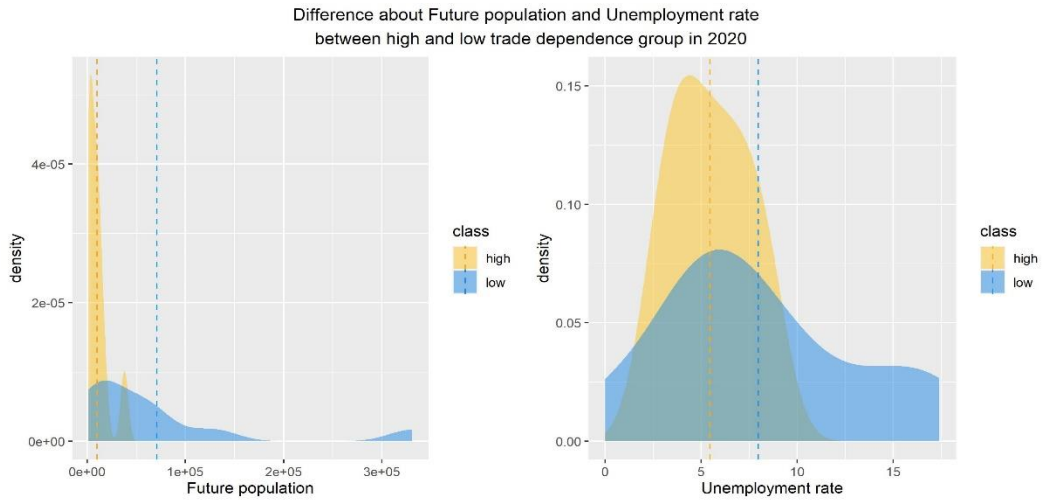
EDA

Data science 11 page

Visualization (cond.)

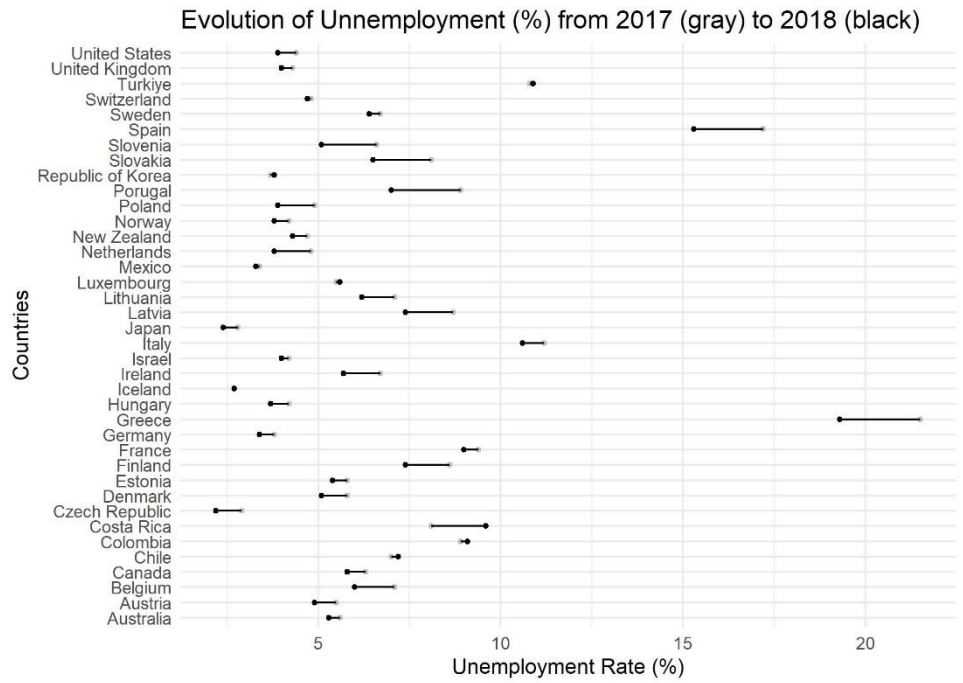
-Between Top 10 high trade dependence group(yellow) and Top 10 low trade dependence group(blue)



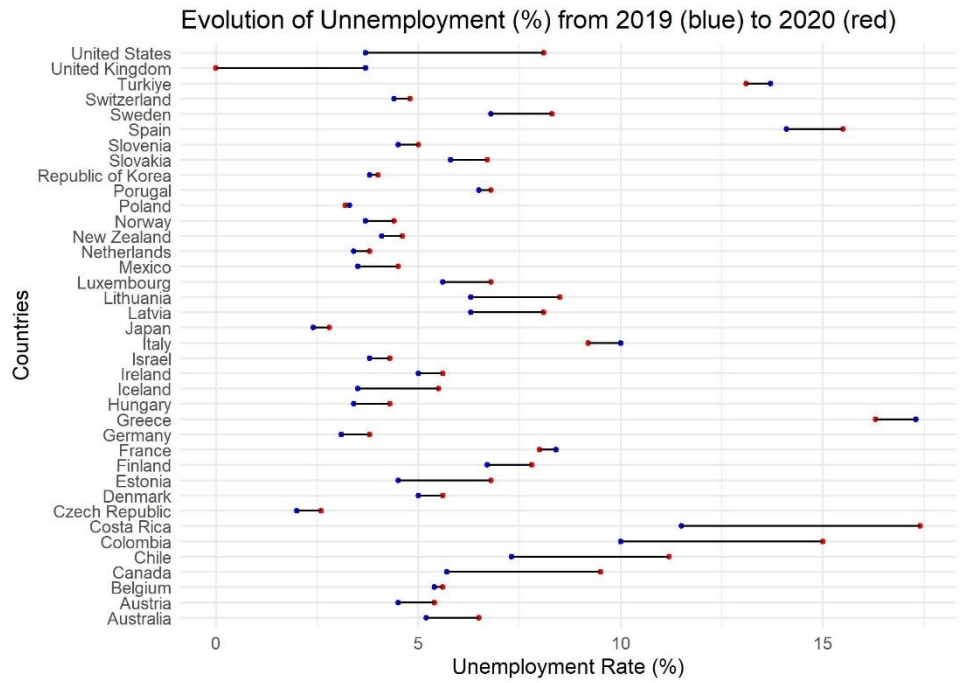


Rate of change in unemployment rate by year (OECD)

Unemployment rate decrease (from 2017 to 2018)



Unemployment rate Increase (from 2019 to 2020)



Challenges

•What was difficult? 

- 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.

- We have difficult to understanding the linear regression results.

• What did you do about it?

-Through the searching, we found trade-related figures, trade dependence, and analyzed them using them.

-Through googling, we studied the meaning of the numbers from linear regression results.

Conclusions

- Corona had a positive impact by lowering trade dependence , a figure that indicates how much the country's economy is affected by global issues.
- The unemployment rate and future population is characteristics that appear in groups with high and low trade dependence.