CS492(B) final project: A study on COVID-19 vaccination dataset

Team 9 20180526 Lee Hyejin 20190436 Lee Kyungmin

I. Introduction

I-1 Motivation

Today, COVID-19 vaccination is a very popular topic since most of the countries are doing their best on it. Some of the countries are about to form the social immunity (herd immunity), but Korea is in the beginning. Since there are various datasets related to COVID-19 vaccination, we chose this topic for our project.

We thought that by analyzing the real data, we have to get an insight about the current worldwide vaccination progress and the details of COVID-19 vaccination.

I-2 Previous Works

There are some previous works exist related to COVID-19 vaccination.

In Kaggle, there are various COVID-19 vaccination-related tasks such as tracking the progress of COVID-19 vaccination. Those tasks, however, are only efficient to know about the vaccination progress itself. Since we wanted to know deeply about the effect of vaccination and the details of it, these tasks were not enough to fulfill our needs.

Also, there is a paper published, with a title "COVID-19 dynamics after a national immunization program in Israel." This paper shows the effect of vaccination in COVID-19 seriousness, but gives only a qualitative measure.

I-3. Objectives

Here, our final goal is to analyze COVID-19 vaccination datasets in various ways and propose the **guidelines for vaccination**. To achieve the goal, we started with four big questions.
1. Does vaccination really alleviates COVID-19 seriousness?
2. What is more important between vaccination and several social distancing policies?
3. Which adverse effects are common in the vaccines?
4. Can we have less stringent social distancing after vaccination?

II. Datasets

II-1. Our World In Data COVID-19 dataset

(link : https://ourworldindata.org/coronavirus)

In this project, we mainly used "Our World In Data COVID-19 dataset". This dataset inicludes 12 csv files in total.

The first csv file, 'owid-covid-data.csv' includes daily measures by country for COVID-19 seriousness, vaccination progress, and some additional information of the countries.

```
##  [1] "iso_code"
##  [2] "continent"
##  [3] "location"
##  [4] "date"
##  [5] "total_cases"
##  [6] "new_cases"
##  [7] "new_cases_smoothed"
##  [8] "total_deaths"
##  [9] "new_deaths"
## [10] "new_deaths_smoothed"
## [11] "total_cases_per_million"
## [12] "new_cases_per_million"
## [13] "new_cases_smoothed_per_million"
## [14] "total_deaths_per_million"
## [15] "new_deaths_per_million"
## [16] "new_deaths_smoothed_per_million"
## [17] "reproduction_rate"
## [18] "icu_patients"
## [19] "icu_patients_per_million"
## [20] "hosp_patients"
## [21] "hosp_patients_per_million"
## [22] "weekly_icu_admissions"
## [23] "weekly_icu_admissions_per_million"
## [24] "weekly_hosp_admissions"
## [25] "weekly_hosp_admissions_per_million"
## [26] "new_tests"
## [27] "total_tests"
## [28] "total_tests_per_thousand"
## [29] "new_tests_per_thousand"
## [30] "new_tests_smoothed"
## [31] "new_tests_smoothed_per_thousand"
## [32] "positive_rate"
## [33] "tests_per_case"
## [34] "tests_units"
## [35] "total_vaccinations"
## [36] "people_vaccinated"
## [37] "people_fully_vaccinated"
## [38] "new_vaccinations"
## [39] "new_vaccinations_smoothed"
## [40] "total_vaccinations_per_hundred"
## [41] "people_vaccinated_per_hundred"
## [42] "people_fully_vaccinated_per_hundred"
## [43] "new_vaccinations_smoothed_per_million"
## [44] "stringency_index"
## [45] "population"
## [46] "population_density"
## [47] "median_age"
## [48] "aged_65_older"
## [49] "aged_70_older"
## [50] "gdp_per_capita"
## [51] "extreme_poverty"
## [52] "cardiovasc_death_rate"
## [53] "diabetes_prevalence"
## [54] "female_smokers"
```

```
## [55] "male_smokers"
## [56] "handwashing_facilities"
## [57] "hospital_beds_per_thousand"
## [58] "life_expectancy"
## [59] "human_development_index"
```

There are three kinds of variables in the dataset. First is the observations of COVID-19 seriousness. This includes new_cases, new_deaths, icu_patients and hosp_patients. THe second one is the observations of COVID-19 vaccination and this includes people_vaccinated and people_fully_vaccinated. The last one is the extra information about the countries such as population, female_smokers or life_expentancy.

```
##    new_cases        new_deaths        icu_patients   hosp_patients
##  Min.   :-74347   Min.   :-1918.0   Min.   :    0   Min.   :     0
##  1st Qu.:      2   1st Qu.:    0.0   1st Qu.:   32   1st Qu.:   118
##  Median :     70   Median :    2.0   Median :  186   Median :   702
##  Mean   :  6021   Mean   :  141.5   Mean   : 1096   Mean   :  4832
##  3rd Qu.:    780   3rd Qu.:   18.0   3rd Qu.:  726   3rd Qu.:  2857
##  Max.   :905992   Max.   :17906.0   Max.   :30022   Max.   :133214
##  NA's   :2584     NA's   :12256     NA's   :81158   NA's   :78919
```

Between various COVID-19 seriousness measures, we found that icu_patients and hosp_patients have lots of NA values inside, so we decided not to use those measures. For the all analysis, we will use only cases and deaths for the analysis.

II-2. VAERS (Vaccine Adverse Effect Report System) dataset

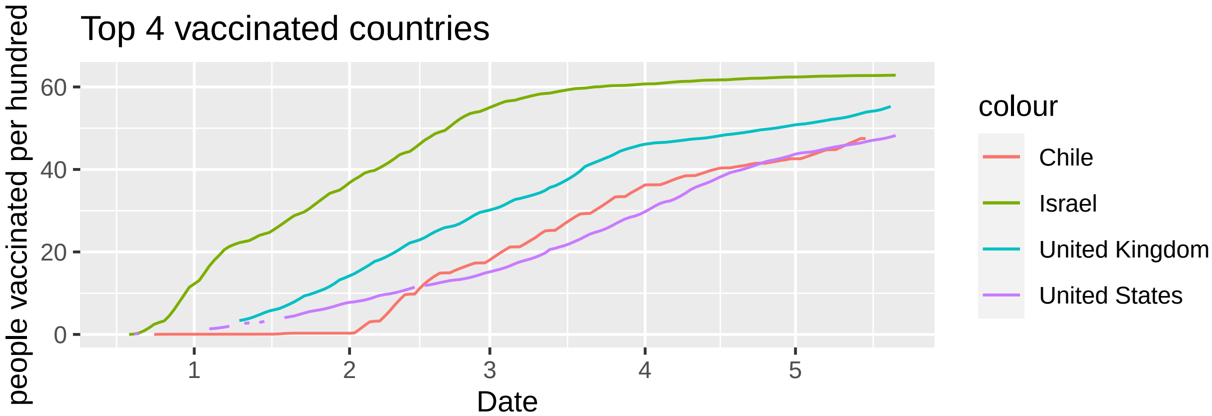
(link : https://vaers.hhs.gov/data.html)

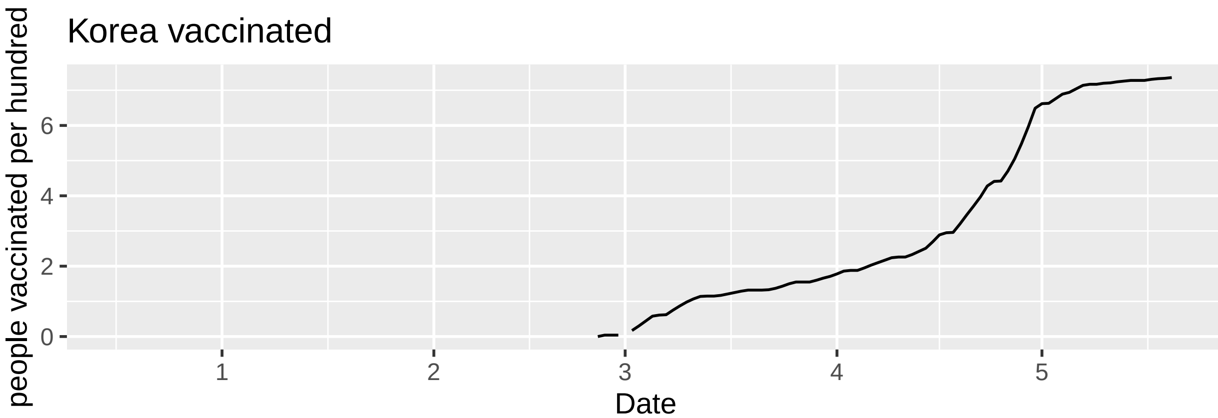
For vaccination adverse effect analysis, we used VAERS (Vaccination Adverse Effect Report System) dataset. This dataset is not only for COVID-19 and includes several other vaccines, but in 2021 data, it mainly contains data about COVID-19 vaccines. Three datasets were used in total, VAERSdata, VAERSVAX and VAERSSYMPTOMS.

III. Data Anaylsis and Interpretation

III-1. Data overview

First, we got a brief look about how vaccination is going on all around the world.





These two are the representative figures. Some top vaccinated countries are reaching about 60%. Especially, Israel is showing the fastest vaccination all around the world. However, Korea's vaccination is still about 10%, which is far less compared to those top vaccinated countries.

III-2. Does vaccination really alleviates COVID-19 seriousness?

In this section, we tried to find out whether there is an effect of vaccination on COVID-19 seriousness. For this purpose, we first preprocessed the data. Here, we thought that the new cases or cases per million itself cannot represent the "alleviation" effect. Thus we made a new variable, "case_change". Case change is defined as the new cases after vaccination devided by the total cases before vaccination.

```
full_dataset %>%
  filter(!is.na(people_vaccinated)) %>%
  select(date) %>%
  summary()
```

```
##       date
##  Min.   :2020-12-02
##  1st Qu.:2021-02-23
##  Median :2021-03-27
##  Mean   :2021-03-21
##  3rd Qu.:2021-04-22
##  Max.   :2021-05-21
```

In the above summary, we could find that the first vaccination case was reported in "2020-12-02".

Thus, case_change is defined by mean(new_cases_per_milion between 2021-05-03 and 2021-05-16) / total_cases_per_million before 2020- 12-01.

So, the data processing is like the code below. "after_vaccination", and "before_vaccination" was extracted from the dataset, joined by country, and the new variable "case_change" was added. With the same principle, "death_change" was also defined.

```
start_date <- as.Date("2021-05-3"); end_date <- as.Date("2021-05-16")
after_vaccination <- owid_covid %>%
  filter(date >= start_date & date <= end_date) %>%
  filter(population >= 5000000) %>%
  group_by(iso_code) %>%
  summarise(new_cases = mean(new_cases_per_million, na.rm = TRUE),
            new_deaths = mean(new_deaths_per_million, na.rm = TRUE),
            people_vaccinated = max(people_vaccinated_per_hundred),
            people_fully_vaccinated = max(people_fully_vaccinated_per_hundred)
  ) %>%
  na.omit()

start_date_2 <- as.Date("2021-1-3"); end_date_2 <- as.Date("2021-12-01")
before_vaccination <- owid_covid %>%
  filter(date >= start_date_2 & date <= end_date_2) %>%
  group_by(iso_code) %>%
  summarise(new_cases = max(total_cases_per_million),
            new_deaths = max(total_deaths_per_million),
  ) %>%
  na.omit()

#join two tibbles
compare <- inner_join(before_vaccination, after_vaccination, by="iso_code")


#make new variables
new_compare <- compare %>%
  mutate(case_change = new_cases.y / new_cases.x * 100,
         death_change = new_deaths.y / new_deaths.x * 100)
new_compare <- new_compare %>%
  filter(!is.na(death_change))
```

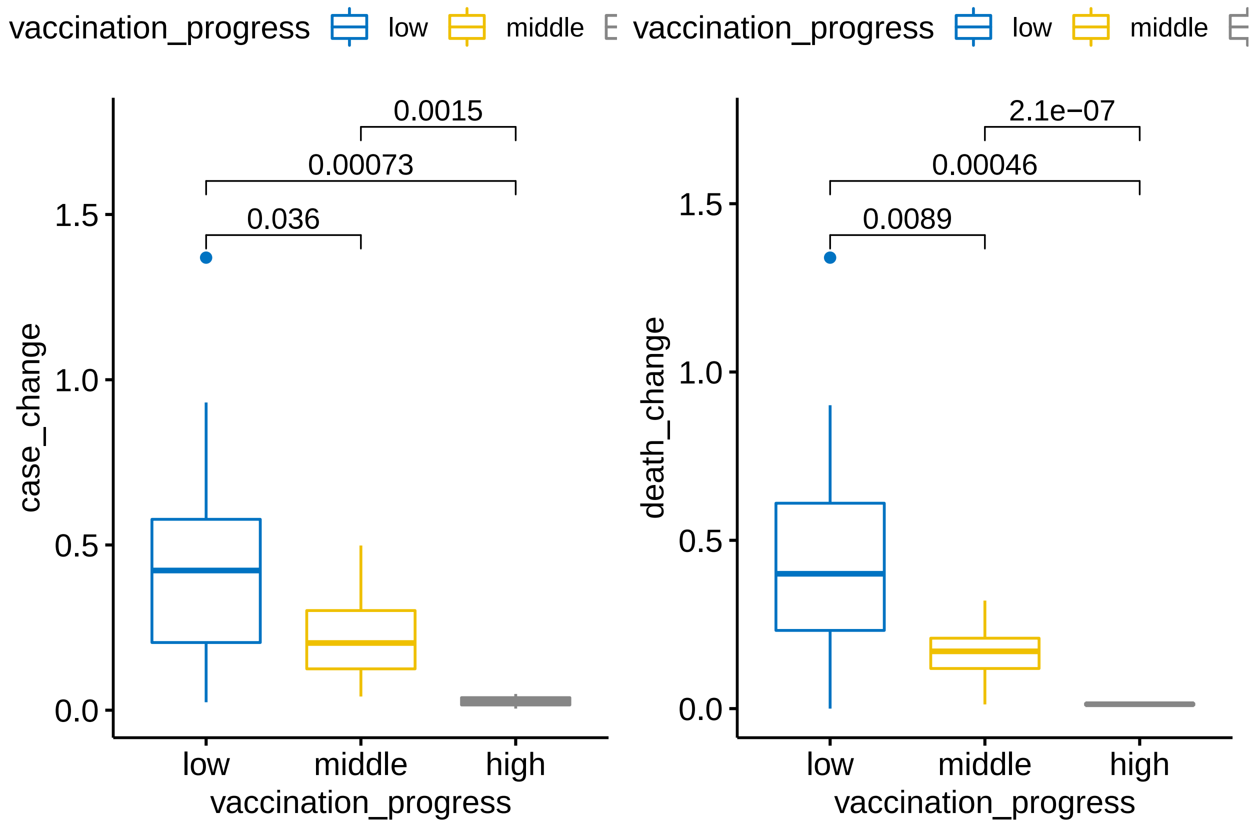
Then, we wanted to compare the effect of vaccination categorically. Countires were assigned a new variable "vaccination_progress". If the variable "people_vaccinated" is under 20, then it was assigned low. Between 20 and 50 were assigned middle, and over 50 was assigned as high.

This is the result of categorization.

```
table(new.df$vaccination_progress)
```

```
##
##    low middle   high
##     14     19      2
```

Lastly, we made boxplots, and annotated the results of t.test with the boxplots.



Boxplot clearly shows that vaccination progress significantly affects both the cases and deaths of COVID-19. Although most of the vaccines used gives the full effect after administration twice, this plot shows that even one dose of vaccination can affect COVID-19 seriousness.

III-3. What is more important between vaccination and several social distancing policies?

Then, we moved onto our next question, "What is more important between vaccination and several social distancing policies?". For analysis, we tried to construct a model which can explain the COVID-19 seriousness by vaccination and several social distancing policies.

Preprocessing First, we pre-processed the datasets. All the datasets from our world in data was inner-joined by country and date. Since the daily case or deaths has a fluctuation which can prevent the model fitting, we averaged those values by week. Then, we had to decide how to see the "seriouness change". Here, we used "3 week interval" since the real effect of vaccination and social distancing arises between 2-3weeks. Thus, we made a new variable "case_change" which is defined as averaged case devided by averaged case after 3 weeks.

Model 1. linear regression for the whole world data: First, we used linear regression.

```
model1.analysis <- lm(case_change~people_vaccinated + contact_tracing+facial_coverings+restrictions_inte
                        cancel_public_events+restriction_gatherings+
                        close_public_transport+school_closures+
                        stay_home_requirements+workplace_closures,
               data = mut_week.data)
summary(model1.analysis)
```

```
##
## Call:
## lm(formula = case_change ~ people_vaccinated + contact_tracing +
##     facial_coverings + restrictions_internal_movements + cancel_public_events +
##     restriction_gatherings + close_public_transport + school_closures +
##     stay_home_requirements + workplace_closures, data = mut_week.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##   -75.5   -20.2    -8.6     1.1 12094.9
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       63.2654    11.0996   5.700 1.25e-08 ***
## people_vaccinated                 -0.3309     0.4453  -0.743  0.45745
## contact_tracing                    8.0579     3.9112   2.060  0.03942 *
## facial_coverings                  -6.8394     2.2686  -3.015  0.00258 **
## restrictions_internal_movements  -1.8516     3.6278  -0.510  0.60979
## cancel_public_events              -9.4334     5.6774  -1.662  0.09665 .
## restriction_gatherings            -7.3609     2.7765  -2.651  0.00804 **
## close_public_transport            10.0806     4.2664   2.363  0.01817 *
## school_closures                   -1.7606     3.2872  -0.536  0.59226
## stay_home_requirements             0.6002     3.6770   0.163  0.87034
## workplace_closures                -2.1379     4.0043  -0.534  0.59343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 197.1 on 6385 degrees of freedom
## Multiple R-squared:  0.007444,   Adjusted R-squared:  0.00589
## F-statistic: 4.789 on 10 and 6385 DF,  p-value: 6.929e-07
```

Here, we got the multiple R-squared value of 0.007444 and adjusted R-squared value of 0.00589. We thought a better model was needed.

Model 2. linear regression for Isreal data: One thing was that even if some countires show same policy response, it does not really mean that they are making same policies with same stringency. Thus, we aimed to model country by country, not with the whole world data.

For modeling, we chose Israel as the model country since its vaccination rate is really high, and are already being used as a model country for COVID-19 immunization by several researchers. We again tried a second model, linear regression model for Israel data.

```
model2.df <- mut_week.data %>%
  filter(iso == "ISR")

model2.analysis <- lm(case_change~people_vaccinated + contact_tracing+facial_coverings+restrictions_inte
                        cancel_public_events+restriction_gatherings+
                        close_public_transport+school_closures+
                        stay_home_requirements+workplace_closures,
               data = model2.df)

summary(model2.analysis)
```

```
##
## Call:
## lm(formula = case_change ~ people_vaccinated + contact_tracing +
##     facial_coverings + restrictions_internal_movements + cancel_public_events +
##     restriction_gatherings + close_public_transport + school_closures +
##     stay_home_requirements + workplace_closures, data = model2.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49618 -0.05619 -0.01519  0.02927  0.36161
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -0.299206   0.334935  -0.893   0.3765
## people_vaccinated               -0.004103   0.001928  -2.128   0.0389 *
## contact_tracing                  0.194972   0.077855   2.504   0.0160 *
## facial_coverings                 0.175756   0.091252   1.926   0.0606 .
## restrictions_internal_movements  0.189379   0.239887   0.789   0.4341
## cancel_public_events            -0.010082   0.054308  -0.186   0.8536
## restriction_gatherings          -0.024238   0.050219  -0.483   0.6317
## close_public_transport          -0.420690   0.098114  -4.288  9.7e-05 ***
## school_closures                  0.091532   0.063708   1.437   0.1579
## stay_home_requirements          -0.271524   0.242440  -1.120   0.2688
## workplace_closures               0.014349   0.073144   0.196   0.8454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1559 on 44 degrees of freedom
## Multiple R-squared:  0.764,  Adjusted R-squared:  0.7104
## F-statistic: 14.25 on 10 and 44 DF,  p-value: 8.58e-11
```

Here, it showed Multiple R-squared value of 0.764 and adjusted R-squared value of 0.7104. We could find that most of the variables where shown not to be significant in ths model. However, vaccination rate was shown to be significant. Also, some other variables, contact_tracing and close_public_transport were also significant.

```
model2_1.analysis <- lm(case_change~ people_vaccinated +
                          contact_tracing+
                          close_public_transport,
               data = model2.df)

summary(model2_1.analysis)
```

```
##
## Call:
## lm(formula = case_change ~ people_vaccinated + contact_tracing +
##     close_public_transport, data = model2.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.49519 -0.09168 -0.00744  0.05227  0.56937
##
## Coefficients:
```

```
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.365661   0.137241   2.664  0.01030 *
## people_vaccinated     -0.003644   0.001124  -3.242  0.00210 **
## contact_tracing        0.190850   0.062030   3.077  0.00336 **
## close_public_transport -0.545816   0.074946  -7.283 1.95e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1689 on 51 degrees of freedom
## Multiple R-squared:  0.6788, Adjusted R-squared:  0.6599
## F-statistic: 35.93 on 3 and 51 DF,  p-value: 1.273e-12
```

When eliminating the insignificant variables and re-modeled, it showed multiple R-squared value of 0.6788 and adjusted R-squared value of 0.6599. Also, all variables were shown to be significant. Although some of the social distancing policies were more significant than vaccination in this model, we concluded that vaccination significantly affects COVID-19 seriousness.

Model 3. Random Forest for Israel data: Next, we tested if another model would work better than the simple linear regression. Here, we used random forest model.

We used tidymodels for the job, and after extracting the recipe and applied the correlation filter, stay_home_requirements was removed due to high correlation.

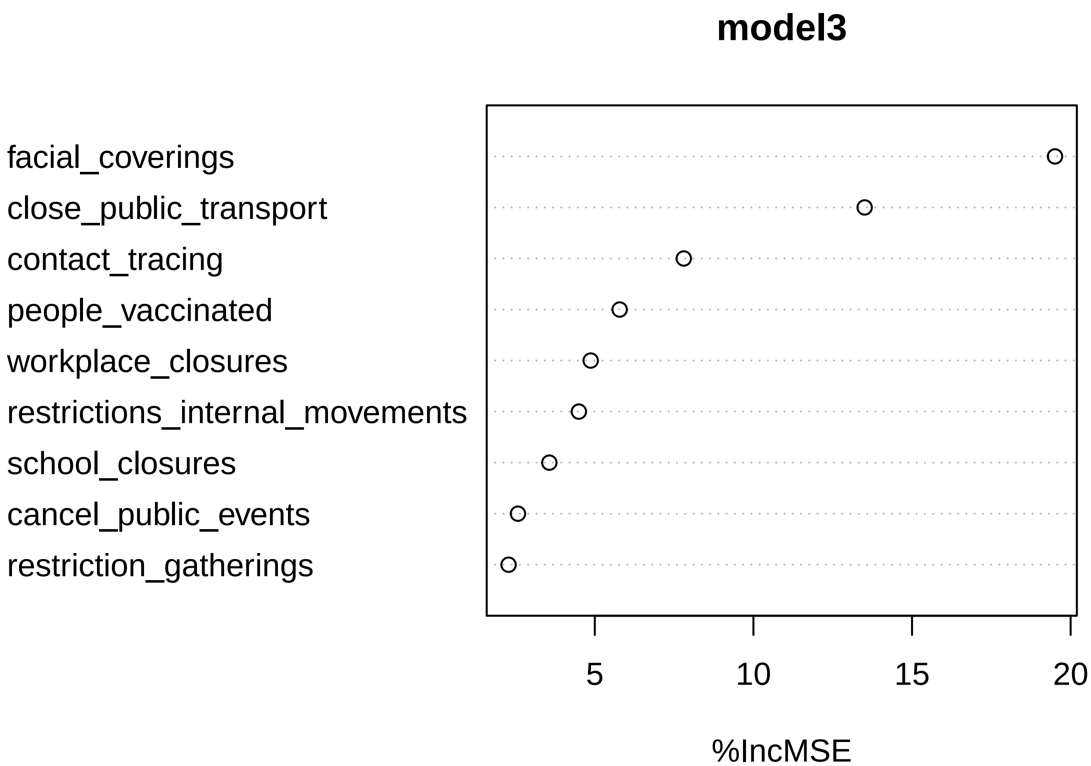
```
set.seed(1234)
model3 <- randomForest(case_change~people_vaccinated +
                         contact_tracing+facial_coverings+
                         restrictions_internal_movements+
                         cancel_public_events+restriction_gatherings+
                         close_public_transport+school_closures+
                         workplace_closures,
                         data=model3.train, mtry = 3, importance = TRUE)

model3 %>%
  predict(model3.test) %>%
  bind_cols(model3.test) %>%
  metrics(truth=case_change, estimate=...1)
```

```
## New names:
## * NA -> ...1
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard      0.0499
## 2 rsq     standard      0.974
## 3 mae     standard      0.0298
```

When tested on the testing data, we could got the r squared value of 0.974. Also, the other metrics were good, too.



Back to our previous goal, finding out the importance of vaccination or the social distancing policies, we plotted the importance values by MSE increase. Here, we could see that vaccination is in the fourth place. Its importance was a bit lower than some of the social distancing policies. However, vaccination was still important in the model.

One more thing interesting in two models was that school_closure was insignificant or not important in those models. Actually, there are so many disadvantages reported about the school closure policy, Korea government decided to fully open the schools next semester. Our observation supports this policy.

III-4. Which adverse effects are common in the vaccines?

Then, we tried to analyze the adverse effects of COVID-19 vaccines. In the datasets, there were only FDA-approved vaccines, we only used the Moderna and Pfizer vaccines.

```
pfizer %>%
  group_by(symptom) %>%
  summarise(freq = n()/ nrow(pfizer) * 100) %>%
  arrange(desc(freq)) %>%
  head(20)
```

```
## # A tibble: 20 x 2
##    symptom              freq
##    <chr>               <dbl>
##  1 Headache             4.91
##  2 Fatigue              3.85
##  3 Chills               3.66
```

```
##   4 Pyrexia            3.61
##   5 Pain               3.42
##   6 Nausea             2.79
##   7 Dizziness          2.75
##   8 Pain in extremity  2.03
##   9 Myalgia            1.61
##  10 Arthralgia         1.57
##  11 Dyspnoea           1.38
##  12 Injection site pain 1.21
##  13 Rash               1.17
##  14 Pruritus           1.15
##  15 Asthenia           1.00
##  16 Vomiting           0.976
##  17 Paraesthesia       0.935
##  18 Lymphadenopathy    0.910
##  19 Malaise            0.893
##  20 Diarrhoea          0.892
```

```
moderna %>%
  group_by(symptom) %>%
  summarise(freq = n()/ nrow(moderna) * 100) %>%
  arrange(desc(freq)) %>%
  head(20)
```

```
## # A tibble: 20 x 2
##    symptom                  freq
##    <chr>                   <dbl>
##  1 Headache                 4.52
##  2 Pyrexia                  3.90
##  3 Chills                   3.84
##  4 Fatigue                  3.66
##  5 Pain                     3.56
##  6 Injection site erythema  2.92
##  7 Nausea                   2.60
##  8 Injection site pain      2.58
##  9 Pain in extremity        2.41
## 10 Injection site pruritus  2.20
## 11 Injection site swelling  2.18
## 12 Dizziness                2.16
## 13 Pruritus                 1.67
## 14 Injection site warmth    1.54
## 15 Myalgia                  1.44
## 16 Erythema                 1.42
## 17 Rash                     1.42
## 18 Arthralgia               1.40
## 19 Injection site rash      1.13
## 20 Dyspnoea                 1.01
```

When arranging the most-occuring side effects of the vaccines, Those top 5 side effects were headache, fatigue, chills, pyrexia and pain. These are the most common side effects for every vaccines, which can be seen as 'normal' effect of the vaccines, by an immune response.

Then, we manually found the number of people administered those vaccines, and the percentage of age groups of vaccine administration.

(link: https://www.statista.com/statistics/1198516/covid-19-vaccinations-administered-us-by-company/)
(link: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html)

```
moderna_by_age_group <- moderna %>%
  mutate(age_group = ifelse(CAGE_YR>=65, "over_65", "under_65"))
moderna_by_age_group %>%
  group_by(age_group) %>%
  summarise(number = n()) %>%
  na.omit() %>%
  cbind(total_vac = c(moderna_over_65, moderna_under_65)) %>%
  mutate(symp_freq = number / total_vac * 100)
```

```
##   age_group number total_vac symp_freq
## 1   over_65 136438  31753603 0.4296772
## 2  under_65 315865  93754710 0.3369057
```

```
pfizer_by_age_group <- pfizer %>%
  mutate(age_group = ifelse(CAGE_YR>=65, "over_65", "under_65"))
pfizer_by_age_group %>%
  group_by(age_group) %>%
  summarise(number = n()) %>%
  na.omit() %>%
  cbind(total_vac = c(pfizer_over_65, pfizer_under_65)) %>%
  mutate(symp_freq = number / total_vac * 100)
```

```
##   age_group number total_vac symp_freq
## 1   over_65  70262  40262628 0.1745092
## 2  under_65 199114 120787885 0.1648460
```

Both in Moderna and Pfizer, side effect reporting rate did not showed significant difference between two age groups. However, we could find that Moderna's reporting rate was a lot bigger than Pfizer's reporting rate.

```
pfizer_over_65 <-
  pfizer_by_age_group %>%
  filter(age_group == "over_65")
pfizer_over_65 %>%
  group_by(symptom) %>%
  summarize(freq = n()/ nrow(pfizer_over_65)* 100) %>%
  arrange(desc(freq)) %>%
  head(10)
```

```
## # A tibble: 10 x 2
##    symptom            freq
##    <chr>             <dbl>
##  1 Headache           3.54
##  2 Fatigue            3.12
##  3 Chills             2.84
##  4 Pyrexia            2.63
##  5 Dizziness          2.44
##  6 Pain               2.36
##  7 Nausea             2.27
##  8 Dyspnoea           1.67
##  9 Pain in extremity  1.64
## 10 Death              1.63
```

```r
pfizer_under_65 <- pfizer_by_age_group %>%
  filter(age_group == "under_65")
pfizer_under_65 %>%
  group_by(symptom) %>%
  summarize(freq = n() / nrow(pfizer_under_65) * 100) %>%
  arrange(desc(freq)) %>%
  head(10)
```

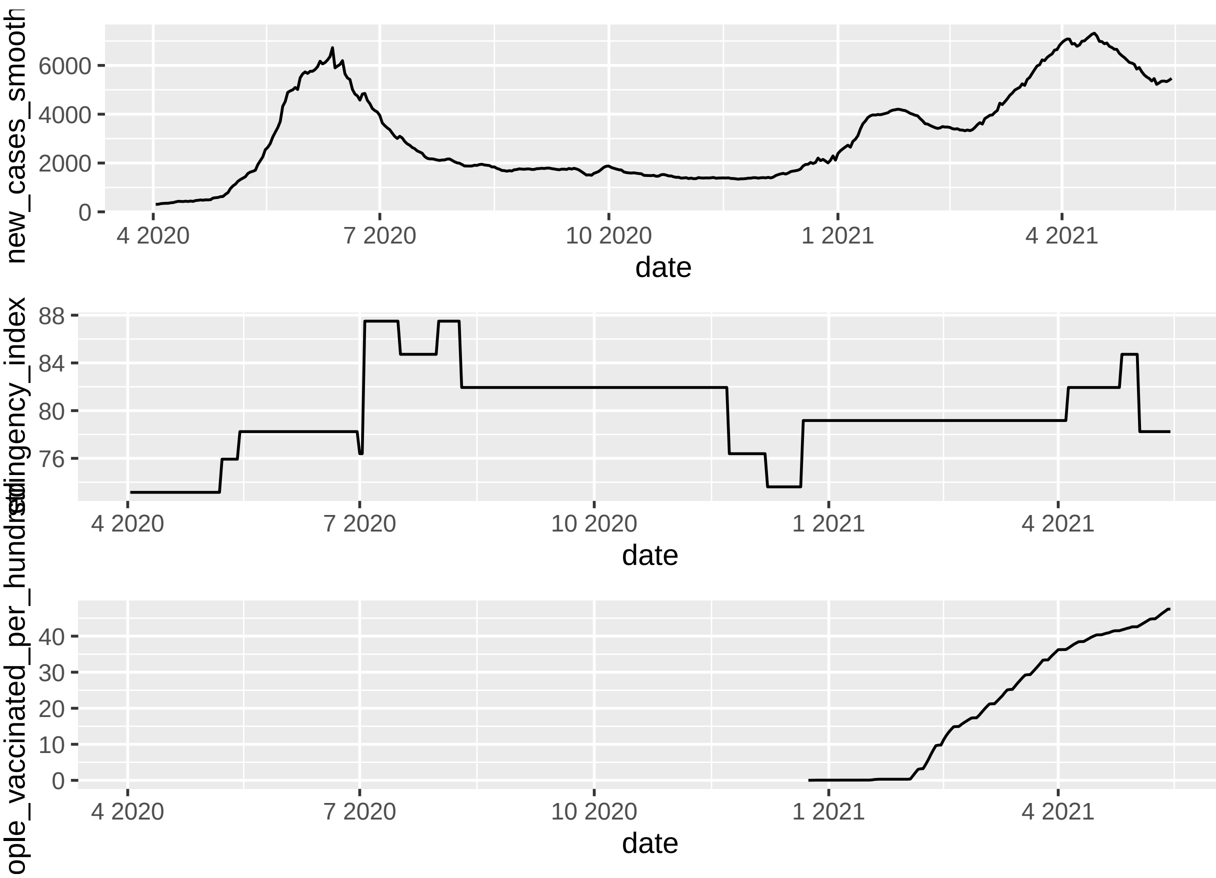
```
## # A tibble: 10 x 2
##    symptom             freq
##    <chr>              <dbl>
##  1 Headache            5.50
##  2 Fatigue             4.19
##  3 Chills              4.06
##  4 Pyrexia             4.03
##  5 Pain                4.01
##  6 Dizziness           3.18
##  7 Nausea              3.15
##  8 Pain in extremity   1.92
##  9 Myalgia             1.78
## 10 Injection site pain  1.69
```

We then analyzed the common adverse effects of Pfizer, which is now used in Korea, by age groups. Most of them are same, but we could find that noticable proportion of people over 65s died.

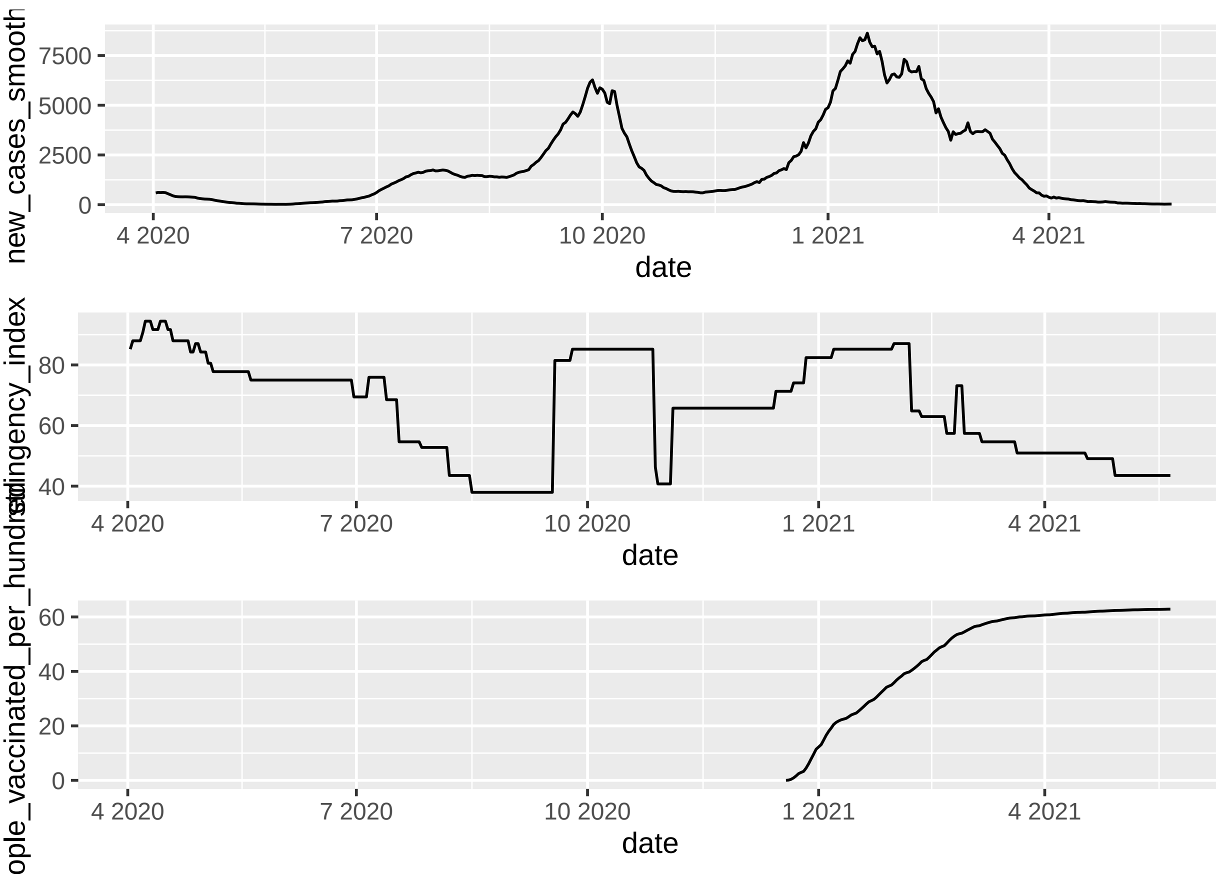
Overall, there was difference in side effect reporting between different vaccines, and no difference between the age groups. However, death is sometimes reported in over 65s, and we thought that this is due to the physical condition or the underlying diseases.

III-5. Can we have less stringent social distancing after vaccination?

Our last question was whether we can have less stringent social distancing after vaccination or not. Here, we tried to find some examples from top vaccinated countries, and we found two great examples.



The first example is Chile. In Chile,vaccination rate is about 50%, but they did less stringent social distancing during vaccination. When the vaccination rate was about 40%, their daily new cases were in the peak.



Compared to it, Isreal did strong social distancing before the vaccination rate reaches about 50%. After that, even if they relieved social distancing policy, their daily new cases decreased over time.

In this sector, we concluded that it is better to keep strong social distancing before 40 ~ 50% of people are at least once vaccinated.

IV. Conclusion

Like this, we made four analysis points, and did a necessary analysis for each of the points. First, we checked that COVID-19 become less serious when more people are vaccinated. Also, by modeling, we foud that vaccination rate significantly affects the COVID-19 case change. However, some social distancing policies were shown to be more important than vaccination rate. Latly,we concluded that it is better to keep social distancing stringent during vaccination time.

While working on the project, we thought that if there are more datasets, we could do more analysis about various points. For example, vaccine adverse effect report system only contains vaccines that are approved by FDA. if there is Astrazeneca vaccine adverse effect dataset, it will be good to compare two vaccines, Astrazeneca and Moderna which are the most widely used vaccines in Korea. Also, vaccination progress by age can enable an analysis about the effect of vaccination by age.

V. R&R

Kyungmin made data overview plots, tried linear modeling with different observations and countries. Also searched for VAERS dataset and did preprocessing.
Hyejin did OWID-COVID data preprocessing, made III-2 boxplot, made three models and analyzed VAERS dataset.