

# Covid-19 Nursing home Mortality rate

Sina Mokhtar

2022-06-21

## load libraries

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(tidyr)
library(usdata)
```

## load nursery home datasets

```
load( "Data/nurse_df.csv")
```

## Top 5 state with cumulative mortality rate

```
# summerize data to state level and filter top 5 state with cumulative mortality rate
# (after "2021-06-30")
# Note: mortality rate = death / occupied bed
nurse_state_cum_top5_list = nurse_df %>%
  drop_na(state) %>%
  group_by(date, state) %>%
  summarize(deaths = sum(deaths, na.rm = TRUE),
            occupied_bed = sum(occupied_bed, na.rm = TRUE)) %>%
  mutate(mortalityRate = (deaths / occupied_bed) ) %>%
  mutate(state_ab = state2abbr(state)) %>%
```

```
dplyr::filter(date >= "2021-06-30") %>%
group_by(state) %>%
mutate(mortalityRate_cum = cumsum(mortalityRate)) %>%
group_by(state) %>%
dplyr::arrange(desc(mortalityRate_cum)) %>%
dplyr::select(state) %>%
distinct() %>%
head(5)
```

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

## Compute county level mortality rate and vaccine rate

```
# summarize data to county level and filter by date and top 5 state
# Note: mortality rate = death / occupied bed
# Note: occupancyRate = occupied_bed / total_bed
nurse_county = nurse_df %>%
  drop_na(state, county) %>%
  filter(date >= "2021-06-30") %>%
  filter(state %in% nurse_state_cum_top5_list$state) %>%
  group_by(date, state, county) %>%
  mutate(nurse_shortage = dplyr::case_when(nurse_shortage == "Y" ~ 1,
                                           nurse_shortage == "N" ~ 0)) %>%
  summarize(deaths = sum(deaths, na.rm = TRUE),
            occupied_bed = sum(occupied_bed, na.rm = TRUE),
            nurse_shortage = mean(nurse_shortage, na.rm = TRUE),
            total_bed = sum(total_bed, na.rm = TRUE),
            resident_vaccineRate = mean(resident_vaccineRate, na.rm = TRUE),
            personnel_vaccineRate = mean(personnel_vaccineRate, na.rm = TRUE),
            resident_vaccineRate_boost = mean(resident_vaccineRate_boost, na.rm = TRUE),
            personnel_vaccineRate_boost = mean(personnel_vaccineRate_boost, na.rm = TRUE)) %>%
  ungroup() %>%
  mutate(mortalityRate = (deaths / occupied_bed) * 100,
         occupancyRate = (occupied_bed / total_bed) * 100,
         state_abb = state2abbr(state),
         occupancyRate = if_else(occupancyRate < 100, occupancyRate, NULL)) %>%
  drop_na(mortalityRate) %>%
  mutate(mortalityRate = ifelse(mortalityRate == Inf, NA, mortalityRate))
```

## 'summarise()' has grouped output by 'date', 'state'. You can override using the  
## '.groups' argument.

## Linear models for mortalityRate VS nurse\_shortage

```
fit = lm(mortalityRate ~ nurse_shortage,  
         data = nurse_county)
```

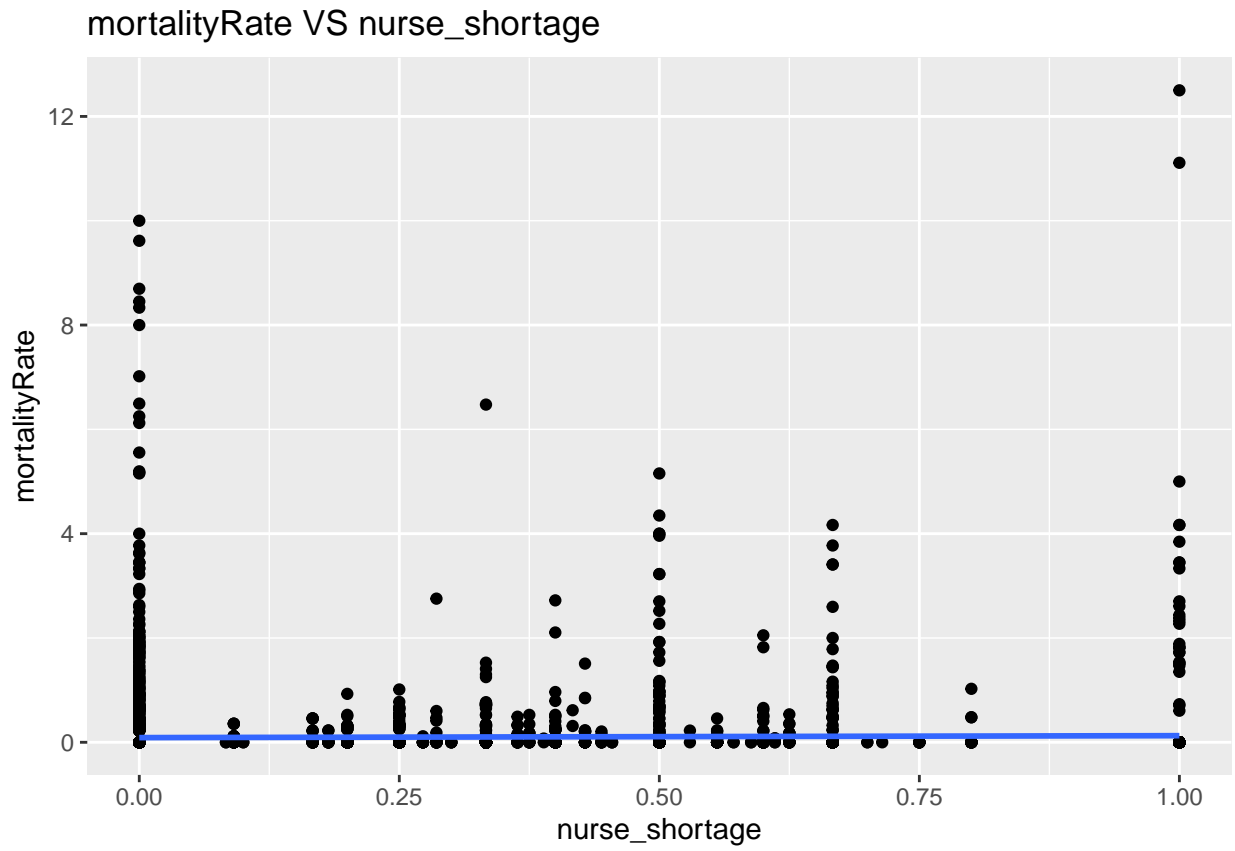
```
summary(fit)
```

```
##  
## Call:  
## lm(formula = mortalityRate ~ nurse_shortage, data = nurse_county)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.1272 -0.1048 -0.0898 -0.0898 12.3728   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  0.089842   0.008686   10.34  <2e-16 ***  
## nurse_shortage 0.037311   0.020393    1.83   0.0674 .    
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.5576 on 6132 degrees of freedom  
## (5 observations deleted due to missingness)  
## Multiple R-squared:  0.0005456, Adjusted R-squared:  0.0003826   
## F-statistic: 3.347 on 1 and 6132 DF, p-value: 0.06736
```

```
#plot the model  
fig_mortalityRate_nurse_shortage = ggplot(data = nurse_county,  
                                           aes(y = mortalityRate,  
                                               x = nurse_shortage)) +  
  labs(title = "mortalityRate VS nurse_shortage") +  
  geom_point() +  
  geom_smooth(method='lm',  
              formula= y~x,  
              se = FALSE)  
fig_mortalityRate_nurse_shortage
```

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



## Linear models for mortalityRate VS resident\_vaccineRate

```
fit = lm(mortalityRate ~ resident_vaccineRate,
        data = nurse_county)

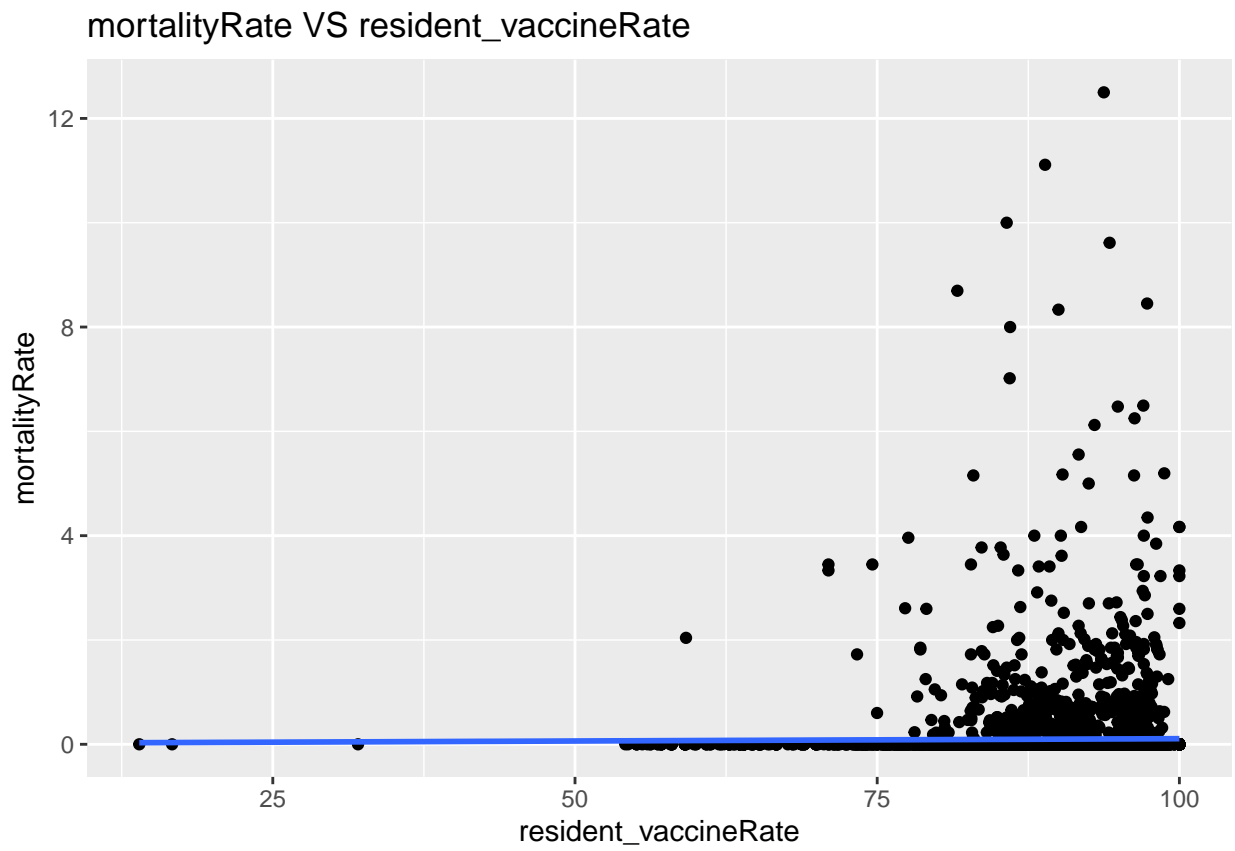
summary(fit)

##
## Call:
## lm(formula = mortalityRate ~ resident_vaccineRate, data = nurse_county)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1076 -0.1038 -0.0996 -0.0942 12.3979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.0186428   0.0891138   0.209   0.834
## resident_vaccineRate 0.0008900   0.0009813   0.907   0.364
##
## Residual standard error: 0.5584 on 6116 degrees of freedom
## (21 observations deleted due to missingness)
## Multiple R-squared:  0.0001345, Adjusted R-squared:  -2.901e-05
## F-statistic: 0.8226 on 1 and 6116 DF,  p-value: 0.3645

#plot the model
fig_mortalityRate_resident_vaccineRate = ggplot(data = nurse_county,
        aes(y = mortalityRate,
             x = resident_vaccineRate)) +
  labs(title = "mortalityRate VS resident_vaccineRate") +
  geom_point() +
  geom_smooth(method='lm',
             formula= y~x,
             se = FALSE)
fig_mortalityRate_resident_vaccineRate

## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```



## Linear models for mortalityRate VS personnel\_vaccineRate

```
fit = lm(mortalityRate ~ personnel_vaccineRate,
        data = nurse_county)

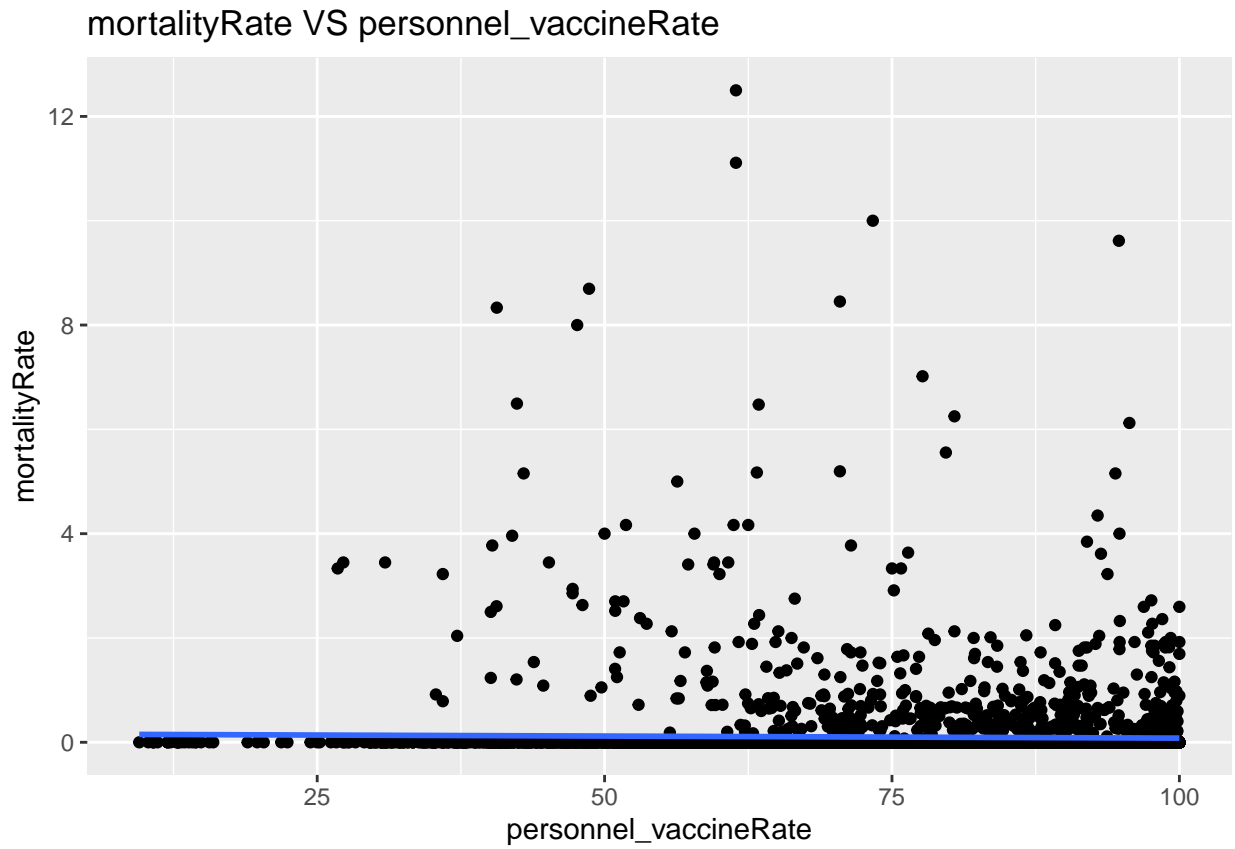
summary(fit)

##
## Call:
## lm(formula = mortalityRate ~ personnel_vaccineRate, data = nurse_county)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1511 -0.1078 -0.0962 -0.0853  12.3895
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1585970   0.0316733     5.007 5.68e-07 ***
## personnel_vaccineRate -0.0007824   0.0004065    -1.925  0.0543 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5583 on 6116 degrees of freedom
## (21 observations deleted due to missingness)
## Multiple R-squared:  0.0006053, Adjusted R-squared:  0.0004419
## F-statistic: 3.704 on 1 and 6116 DF, p-value: 0.05432

#plot the model
fig_mortalityRate_personnel_vaccineRate = ggplot(data = nurse_county,
        aes(y = mortalityRate,
            x = personnel_vaccineRate)) +
  labs(title = "mortalityRate VS personnel_vaccineRate") +
  geom_point() +
  geom_smooth(method='lm',
            formula= y~x,
            se = FALSE)
fig_mortalityRate_personnel_vaccineRate

## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```





## Linear models for mortalityRate VS resident\_vaccineRate\_boost

```
fit = lm(mortalityRate ~ resident_vaccineRate_boost,
        data = nurse_county)

summary(fit)

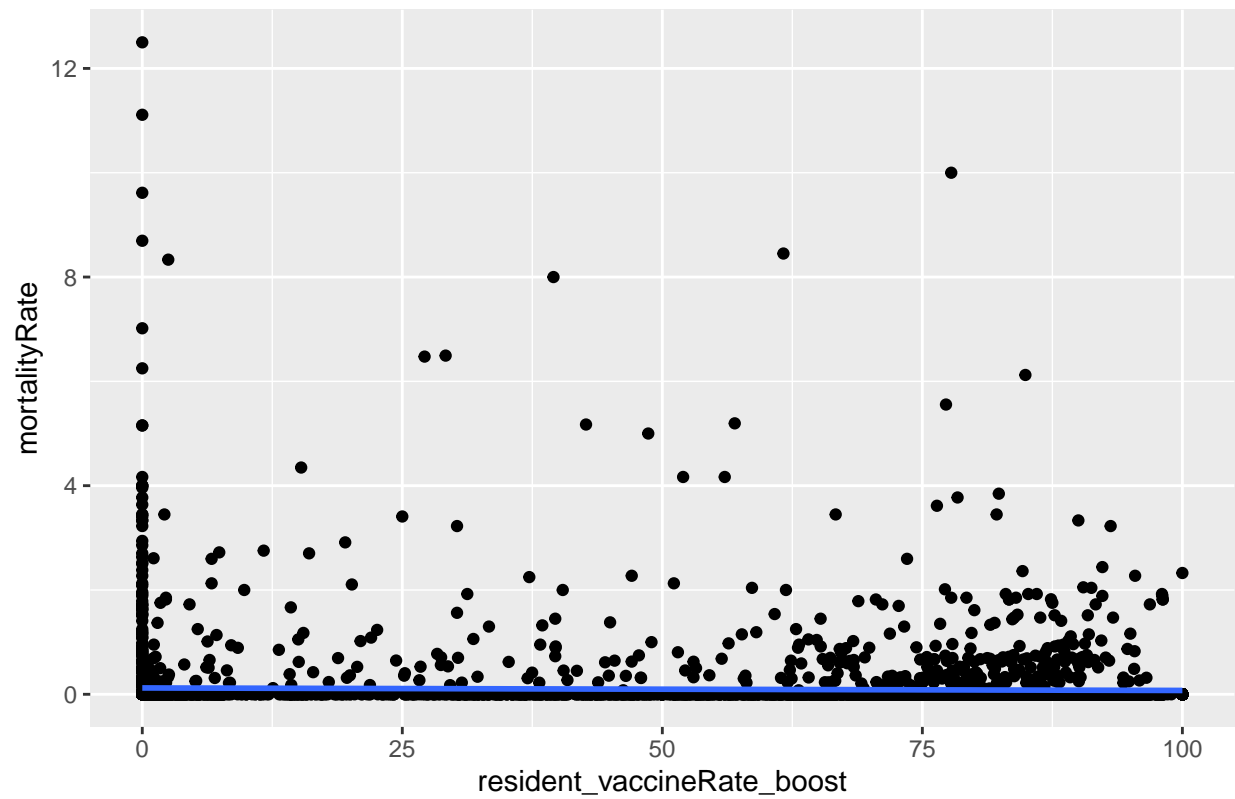
##
## Call:
## lm(formula = mortalityRate ~ resident_vaccineRate_boost, data = nurse_county)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1219 -0.1219 -0.0881 -0.0794  12.3781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1218670   0.0110731   11.006 < 2e-16 ***
## resident_vaccineRate_boost -0.0004821   0.0001801   -2.676  0.00746 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5581 on 6116 degrees of freedom
## (21 observations deleted due to missingness)
## Multiple R-squared:  0.00117,    Adjusted R-squared:  0.001007
## F-statistic: 7.164 on 1 and 6116 DF,  p-value: 0.00746

#plot the model
fig_mortalityRate_resident_vaccineRate_boost = ggplot(data = nurse_county,
              aes(y = mortalityRate,
                  x = resident_vaccineRate_boost)) +
  labs(title = "mortalityRate VS resident_vaccineRate_boost") +
  geom_point() +
  geom_smooth(method='lm',
              formula= y~x,
              se = FALSE)
fig_mortalityRate_resident_vaccineRate_boost

## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```

mortalityRate VS resident\_vaccineRate\_boost



## Linear models for mortalityRate VS personnel\_vaccineRate\_boost

```
fit = lm(mortalityRate ~ personnel_vaccineRate_boost,
        data = nurse_county)

summary(fit)

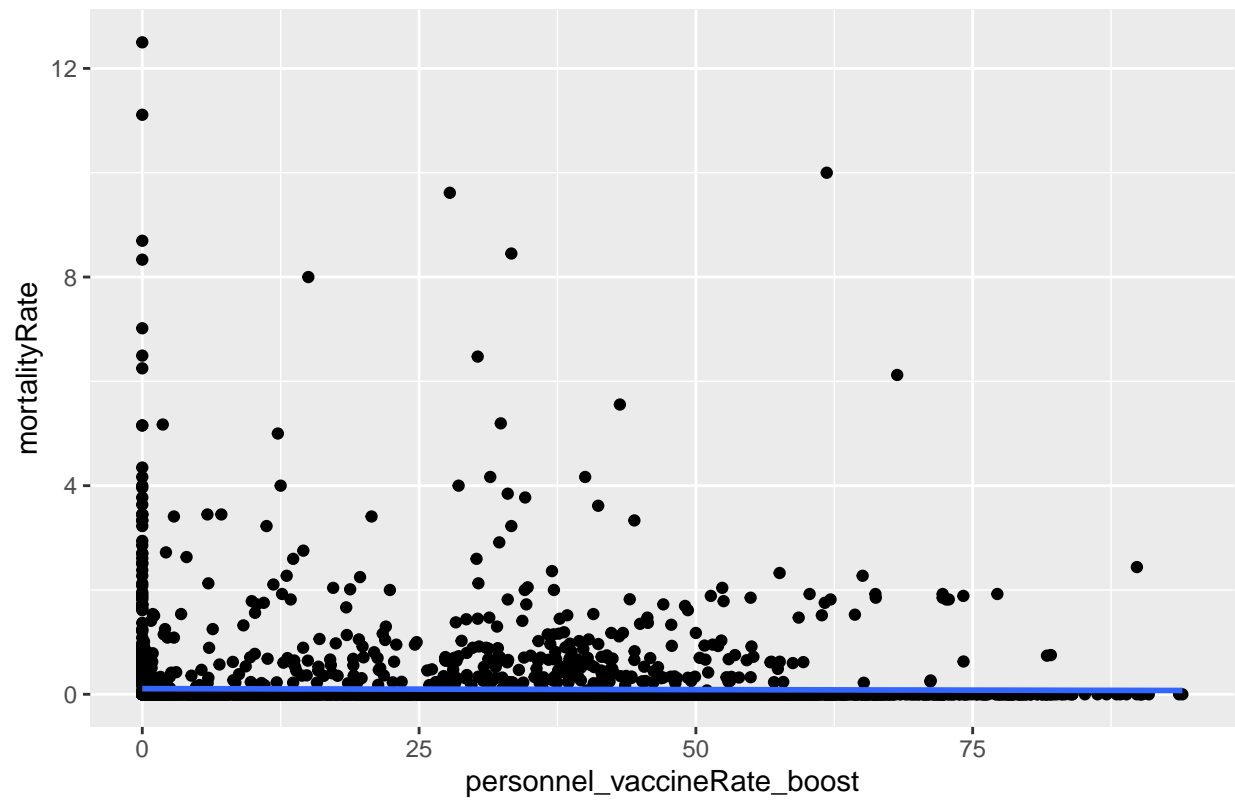
##
## Call:
## lm(formula = mortalityRate ~ personnel_vaccineRate_boost, data = nurse_county)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1081 -0.1081 -0.0977 -0.0904  12.3919
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1080727   0.0103291   10.463  <2e-16 ***
## personnel_vaccineRate_boost -0.0003729   0.0003139   -1.188    0.235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5584 on 6116 degrees of freedom
## (21 observations deleted due to missingness)
## Multiple R-squared:  0.0002307, Adjusted R-squared:  6.723e-05
## F-statistic: 1.411 on 1 and 6116 DF, p-value: 0.2349

#plot the model
fig_mortalityRate_personnel_vaccineRate_boost = ggplot(data = nurse_county,
              aes(y = mortalityRate,
                  x = personnel_vaccineRate_boost)) +
  labs(title = "mortalityRate VS personnel_vaccineRate_boost") +
  geom_point() +
  geom_smooth(method='lm',
              formula= y~x,
              se = FALSE)
fig_mortalityRate_personnel_vaccineRate_boost

## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```

mortalityRate VS personnel\_vaccineRate\_boost



## Linear models for mortalityRate VS occupancyRate

```
fit = lm(mortalityRate ~ occupancyRate,
        data = nurse_county)

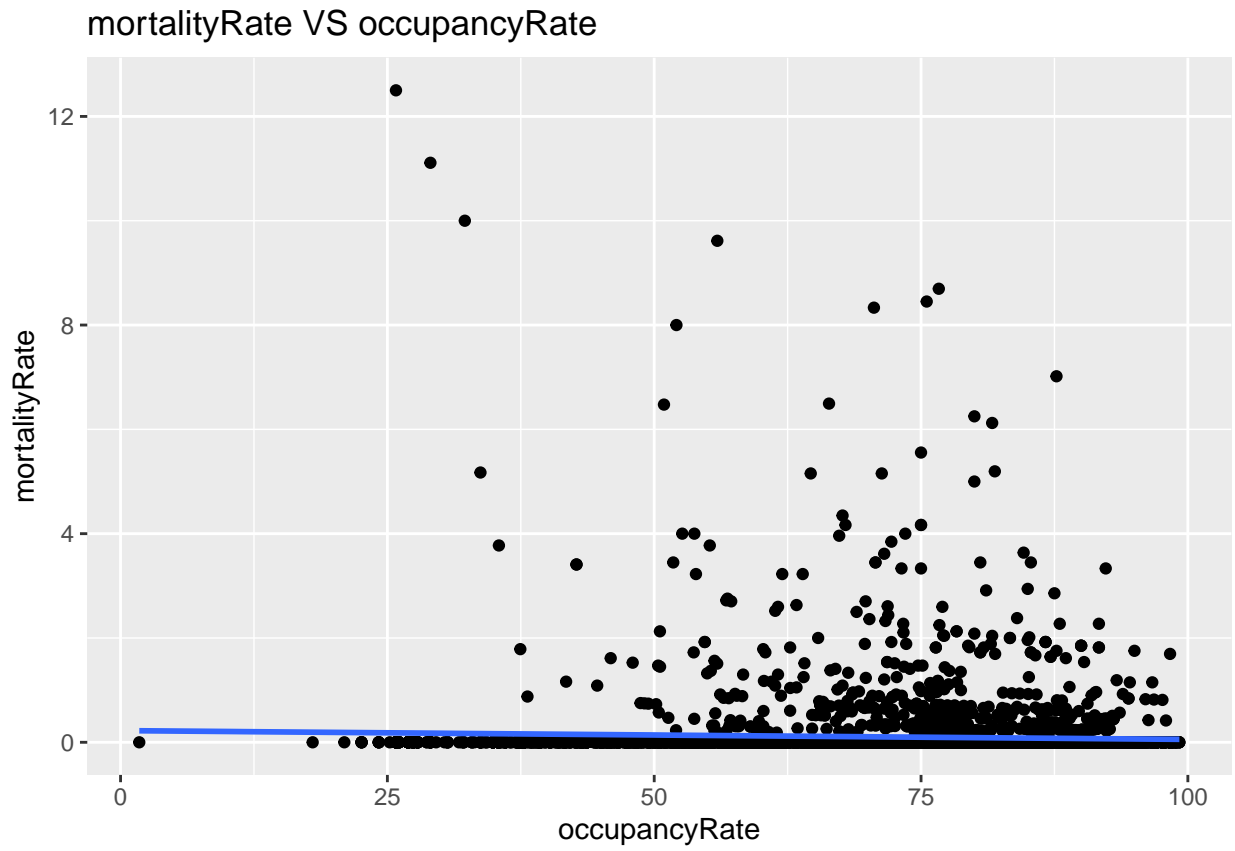
summary(fit)

##
## Call:
## lm(formula = mortalityRate ~ occupancyRate, data = nurse_county)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2211 -0.1126 -0.0939 -0.0755  12.3194
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2240793  0.0353856   6.333 2.58e-10 ***
## occupancyRate -0.0016841  0.0004699  -3.584 0.000341 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.56 on 6069 degrees of freedom
## (68 observations deleted due to missingness)
## Multiple R-squared:  0.002112, Adjusted R-squared:  0.001947
## F-statistic: 12.84 on 1 and 6069 DF, p-value: 0.0003412
```

```
#plot the model
fig_mortalityRate_occupancyRate = ggplot(data = nurse_county,
        aes(y = mortalityRate,
             x = occupancyRate)) +
  labs(title = "mortalityRate VS occupancyRate") +
  geom_point() +
  geom_smooth(method='lm',
             formula= y~x,
             se = FALSE)
fig_mortalityRate_occupancyRate
```

```
## Warning: Removed 68 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 68 rows containing missing values (geom_point).
```



## Time series of the nurse\_shortage

```
fig_time_series_county_nurse_shortage = ggplot(data = nurse_county,
        aes(x = date,
            y = nurse_shortage)) +
  labs(title = "nurse_shortage in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
            formula= y~x,
            se = FALSE)
ggsave("Result/fig_time_series_county_nurse_shortage.jpg",
        fig_time_series_county_nurse_shortage,
        height=4,
        width=8,
        scale=1.65)
```

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
```

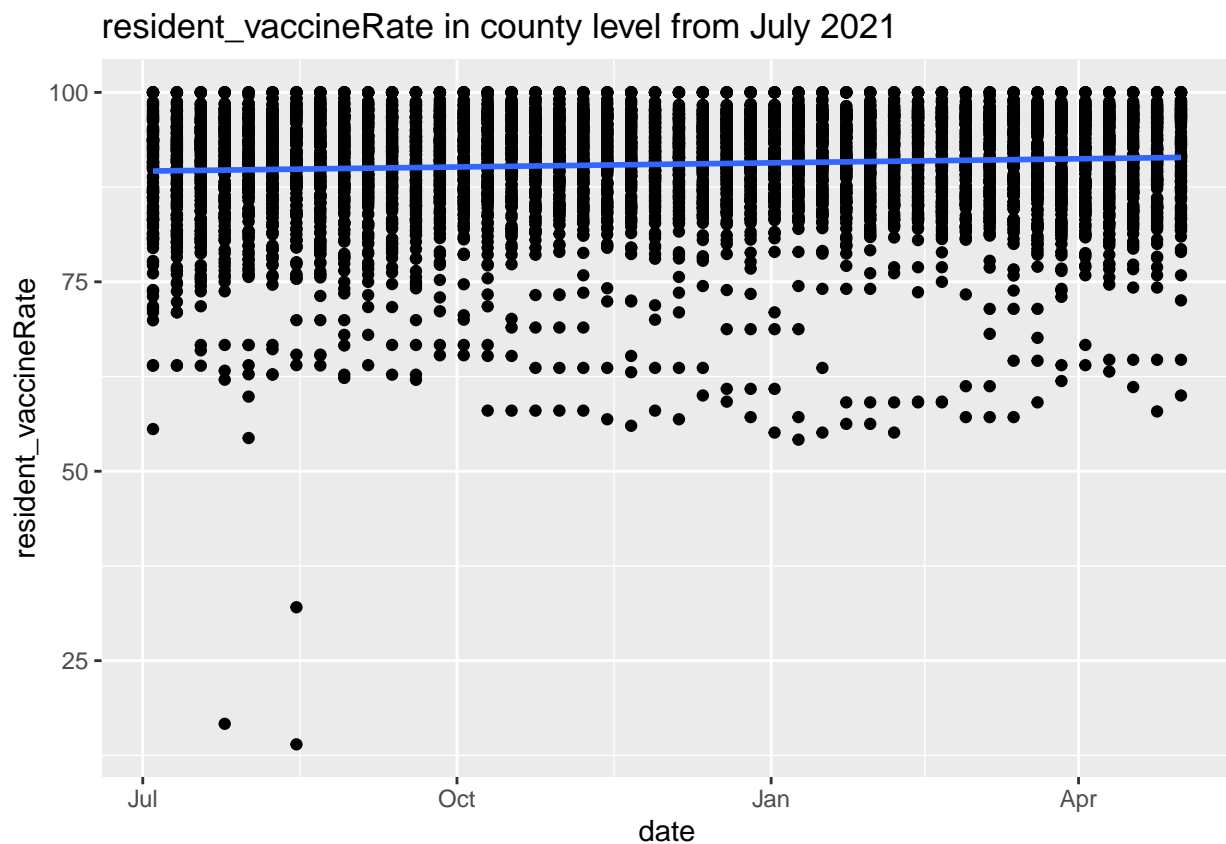
```
## Warning: Removed 5 rows containing missing values (geom_point).
```

## Time series of the resident\_vaccineRate

```
fig_time_series_county_resident_vaccineRate = ggplot(data = nurse_county,
                                                    aes(x = date,
                                                        y = resident_vaccineRate)) +
  labs(title = "resident_vaccineRate in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
              formula= y~x,
              se = FALSE)
fig_time_series_county_resident_vaccineRate
```

## Warning: Removed 21 rows containing non-finite values (stat\_smooth).

## Warning: Removed 21 rows containing missing values (geom\_point).



```
ggsave("Result/fig_time_series_county_resident_vaccineRate.jpg",
        fig_time_series_county_resident_vaccineRate,
        height=4,
        width=8,
        scale=1.65)
```

## Warning: Removed 21 rows containing non-finite values (stat\_smooth).

## Removed 21 rows containing missing values (geom\_point).

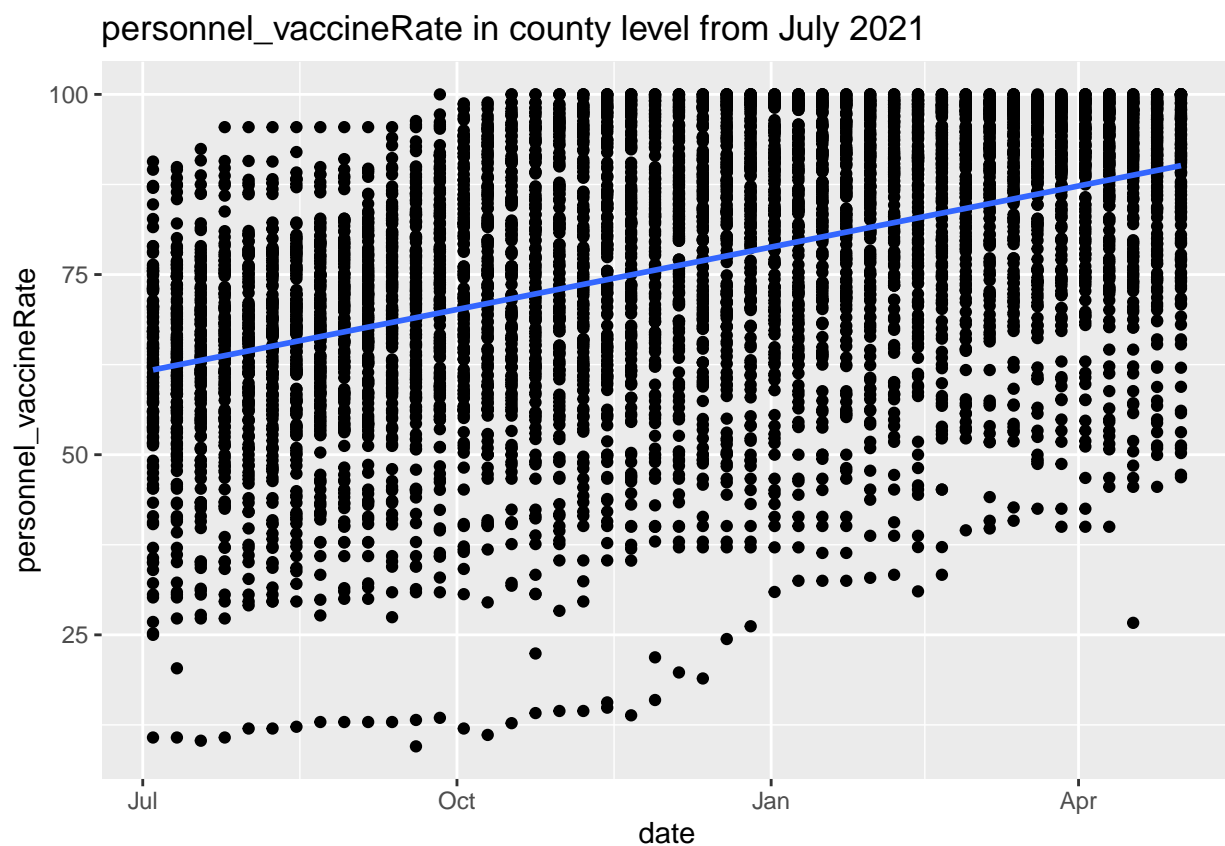


## Time series of the personnel\_vaccineRate

```
fig_time_series_county_personnel_vaccineRate = ggplot(data = nurse_county,
              aes(x = date,
                  y = personnel_vaccineRate)) +
  labs(title = "personnel_vaccineRate in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
              formula= y~x,
              se = FALSE)
fig_time_series_county_personnel_vaccineRate
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



```
ggsave("Result/fig_time_series_county_personnel_vaccineRate.jpg",
        fig_time_series_county_personnel_vaccineRate,
        height=4,
        width=8,
        scale=1.65)
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

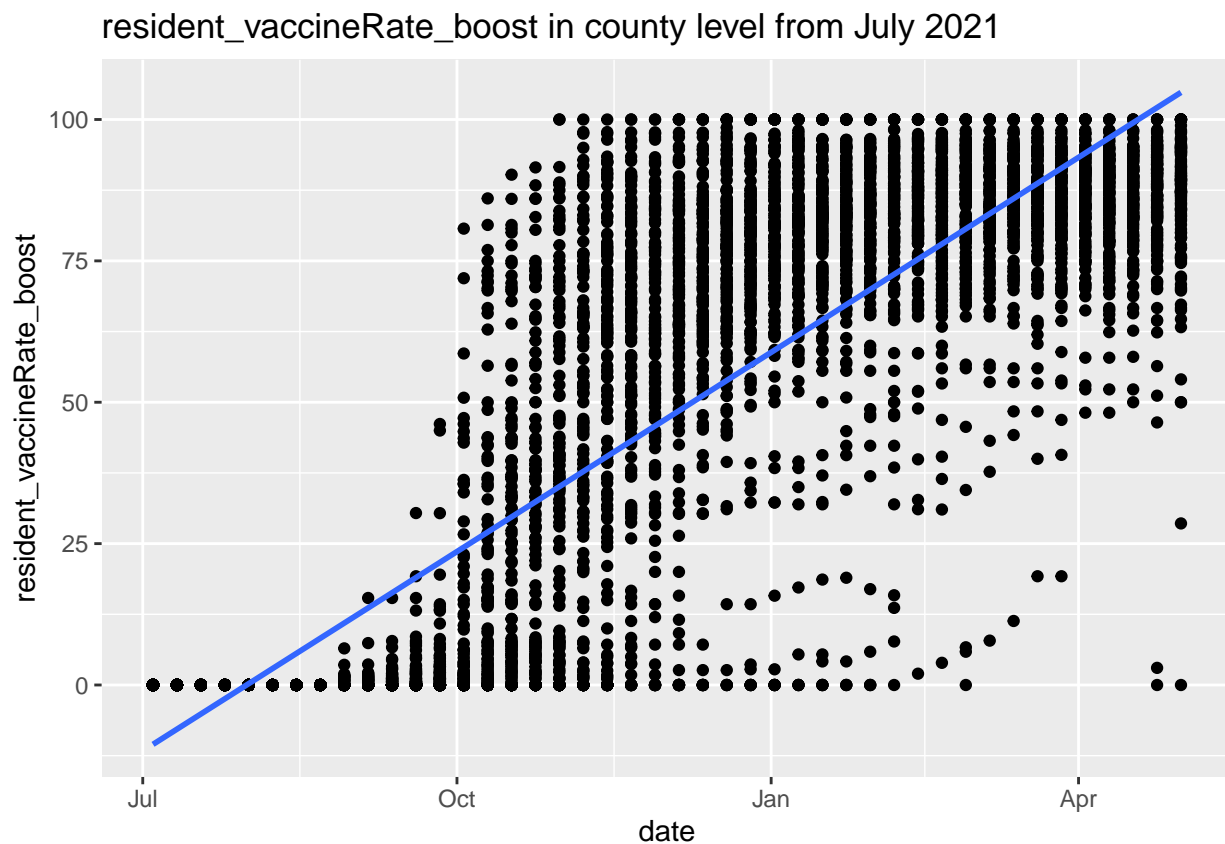
```
## Removed 21 rows containing missing values (geom_point).
```

## Time series of the resident\_vaccineRate\_boost

```
fig_time_series_county_resident_vaccineRate_boost = ggplot(data = nurse_county,
  aes(x = date,
      y = resident_vaccineRate_boost)) +
  labs(title = "resident_vaccineRate_boost in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
    formula= y~x,
    se = FALSE)
fig_time_series_county_resident_vaccineRate_boost
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



```
ggsave("Result/fig_time_series_county_resident_vaccineRate_boost.jpg",
  fig_time_series_county_resident_vaccineRate_boost,
  height=4,
  width=8,
  scale=1.65)
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

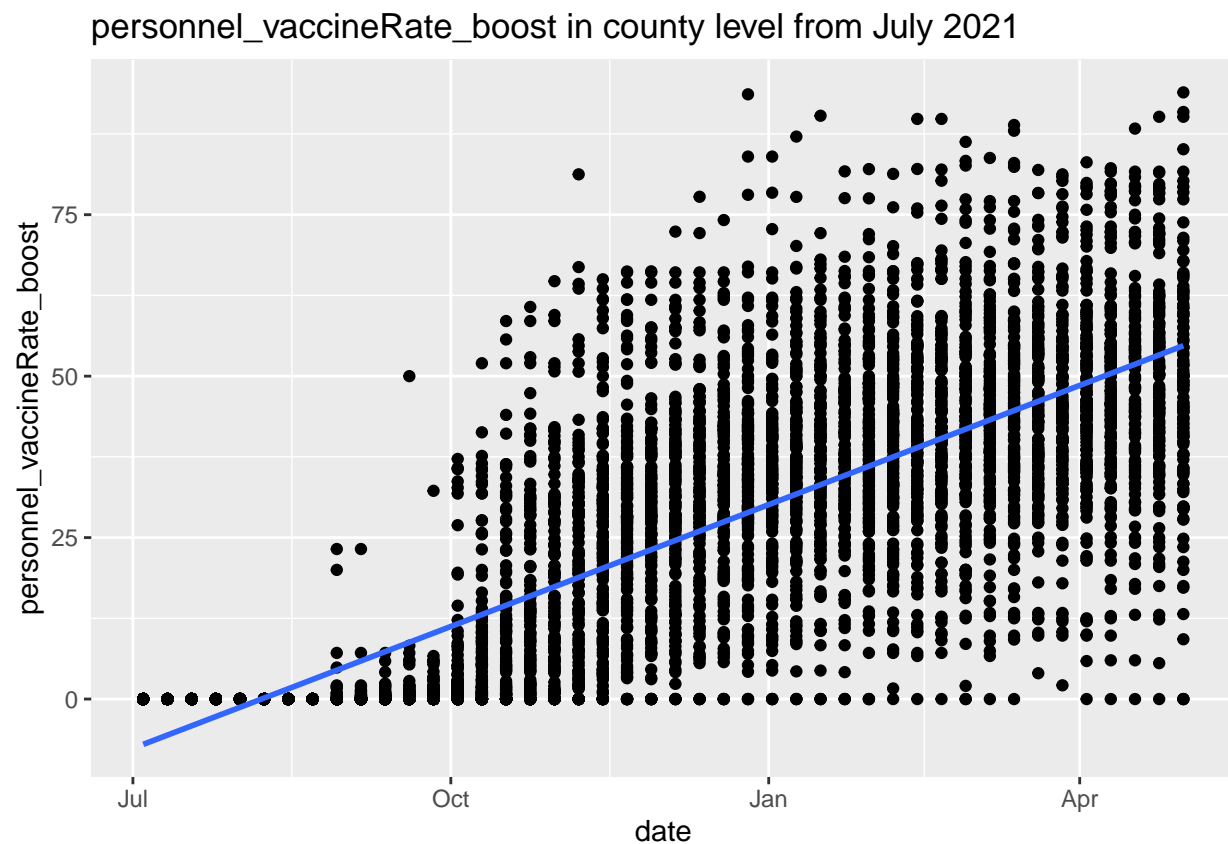
```
## Removed 21 rows containing missing values (geom_point).
```

## Time series of the personnel\_vaccineRate\_boost

```
fig_time_series_county_personnel_vaccineRate_boost = ggplot(data = nurse_county,
  aes(x = date,
      y = personnel_vaccineRate_boost)) +
  labs(title = "personnel_vaccineRate_boost in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
      formula= y~x,
      se = FALSE)
fig_time_series_county_personnel_vaccineRate_boost
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```



```
ggsave("Result/fig_time_series_county_personnel_vaccineRate_boost.jpg",
  fig_time_series_county_personnel_vaccineRate_boost,
  height=4,
  width=8,
  scale=1.65)
```

```
## Warning: Removed 21 rows containing non-finite values (stat_smooth).
```

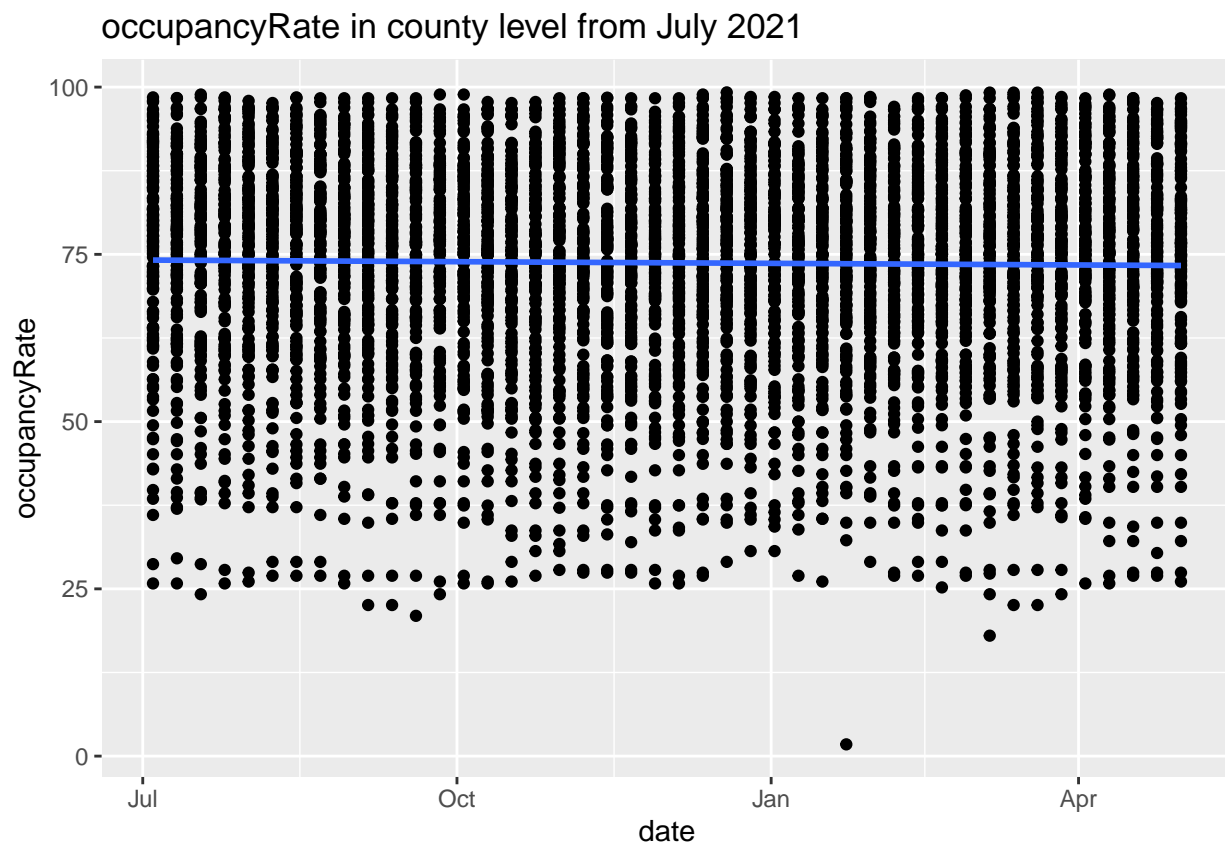
```
## Removed 21 rows containing missing values (geom_point).
```

## Time series of the occupancyRate

```
fig_time_series_county_occupancyRate = ggplot(data = nurse_county,
      aes(x = date,
          y = occupancyRate)) +
  labs(title = "occupancyRate in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
              formula= y~x,
              se = FALSE)
fig_time_series_county_occupancyRate
```

```
## Warning: Removed 68 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 68 rows containing missing values (geom_point).
```



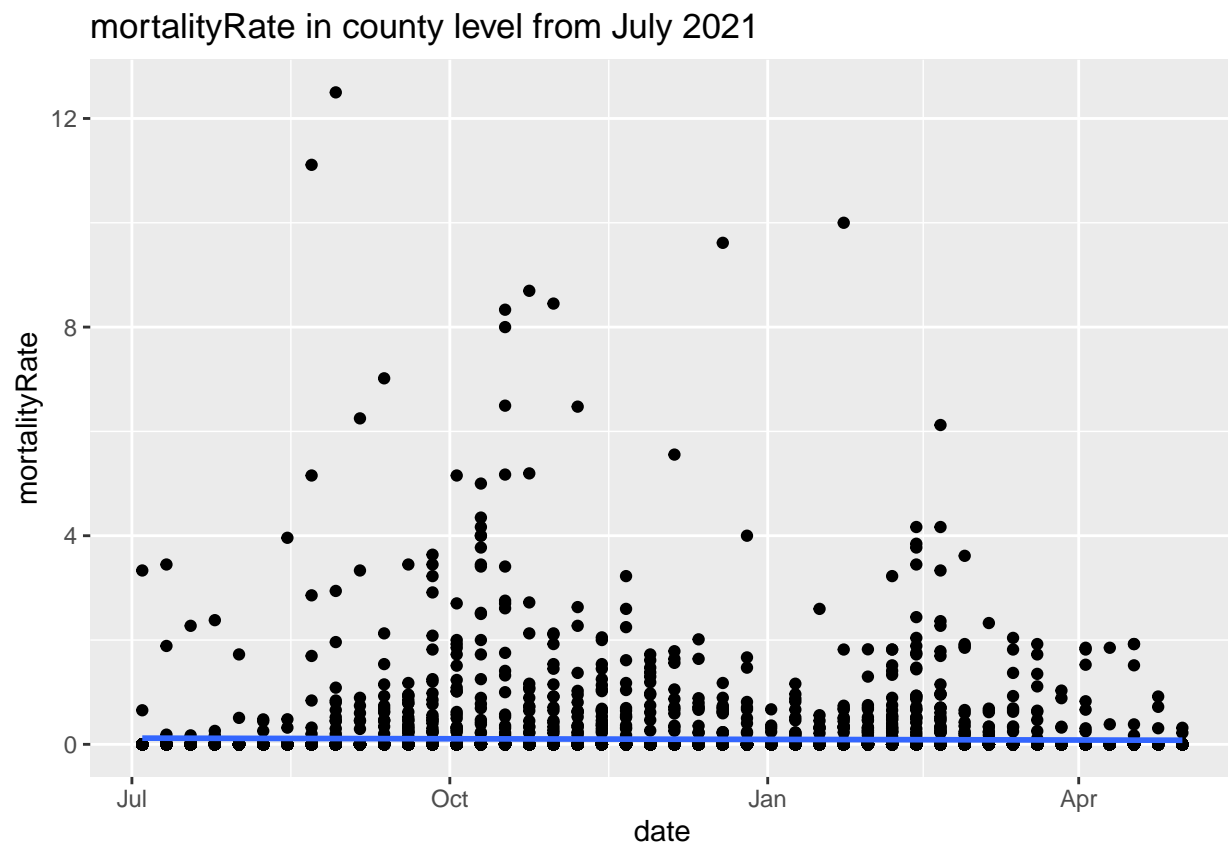
```
ggsave("Result/fig_time_series_county_occupancyRate.jpg",
  fig_time_series_county_occupancyRate,
  height=4,
  width=8,
  scale=1.65)
```

```
## Warning: Removed 68 rows containing non-finite values (stat_smooth).
```

```
## Removed 68 rows containing missing values (geom_point).
```

## Time series of the mortalityRate

```
fig_time_series_county_mortalityRate = ggplot(data = nurse_county,
      aes(x = date,
          y = mortalityRate)) +
  labs(title = "mortalityRate in county level from July 2021 ") +
  geom_point() +
  geom_smooth(method='lm',
      formula= y~x,
      se = FALSE)
fig_time_series_county_mortalityRate
```



```
ggsave("Result/fig_time_series_county_mortalityRate.jpg",
  fig_time_series_county_mortalityRate,
  height=4,
  width=8,
  scale=1.65)
```