

# GY7702 2021/22 Assignment 1

219047470

7/30/2022

## Libraries

```
library(car)
library(dplyr)
library(tidyverse)
library(magrittr)
library(GGally)
library(fmsb)
library(ggplot2)
library(plotly)
```

## Part 1: Loading Data into R

```
link <- 'C:/Users/khalsz/Downloads/data5/data6/'
setwd(link)
data1 <- read.csv('covid19_vaccinations_2020-12-06_2021-10-02_long.csv')
data2 <- read.csv('ukpopestimatesmid2020on2021geography.csv')
laddata1 <- data1 %>%
  dplyr::filter(area_name == c('Bolton', 'Waltham Forest', 'Halton'))
```

```
## Warning in area_name == c("Bolton", "Waltham Forest", "Halton"): longer object
## length is not a multiple of shorter object length
```

```
#Checking NA
laddata1[is.na(laddata1)]
```

```
## character(0)
```

Data contains zero NA

```
#Converting dataframe to wide format
widelad <- laddata1 %>% reshape(timevar = 'metric', v.names = 'value',
                               idvar = c('date_reported', 'area_name'), direction = 'wide')

#Replacing all NA values with Zero
```

```
widelad[is.na(widelad)] <- 0

#Joining data
joinedlad <- widelad %>% dplyr::left_join(
  data2, by = c('area_code' = 'i..Code')
)

# Removing prefix 'value' from columns
new_colnames <- gsub("^value.", "", colnames(joinedlad))
names(joinedlad) <- new_colnames

#Joining data
joinedlad <- joinedlad %>%
  dplyr::select(-Name) %>%
  dplyr::rename( Population = All.ages )

#Replacing coma and converting variables to numeric
joinedlad$Population <- as.numeric( gsub(',', '', joinedlad$Population))

#Converting variable 'date_reported' to data format
joinedlad$date_reported <- as.Date(joinedlad$date_reported)

#Creating new variable 'Perc_New_Cases'
joinedlad_w_p <- joinedlad %>%
  dplyr::mutate(Perc_New_cases = round( (new_cases/Population) *100, 3) )

joinedlad_w_p$Perc_New_cases <- as.character(paste0(joinedlad_w_p$Perc_New_cases, '%'))
```

## Part 2: 2.1

```
joinedlad_w_p[order(joinedlad_w_p$date_reported),] %>%
  dplyr::select(date_reported, area_name, new_cases, Perc_New_cases) %>%
  dplyr::group_by(area_name) %>%
  head(3) %>%
  knitr::kable()
```

date_reported	area_name	new_cases	Perc_New_cases
2020-12-06	Halton	0	0%
2020-12-06	Bolton	52	0.018%
2020-12-06	Waltham Forest	106	0.038%

## Part 2: 2.2

```
widen_joinlad <- joinedlad_w_p %>%
  dplyr::select(date_reported, area_name, new_cases) %>%
```

```

dplyr::arrange( date_reported) %>%
dplyr::group_by(area_name) %>%
dplyr::mutate(pre = dplyr::lag(new_cases, 7, default = NA)) %>%
dplyr::mutate(pre_pec = ((new_cases - pre)/pre)*100) %>%
dplyr::ungroup() %>%
dplyr::select(date_reported, area_name, pre_pec) %>%
tidyr::pivot_wider(names_from = area_name, values_from = pre_pec) %>%
dplyr::filter(dplyr::between(date_reported, as.Date('2021-07-01'), as.Date('2021-07-31')))

widen_joinlad <- data.frame(widen_joinlad)

widen_joinlad[widen_joinlad == Inf] <- 0
widen_joinlad[is.na(widen_joinlad)] <- 0

widen_joinlad <- data.frame( lapply(widen_joinlad, function(x) round(x, 2)))

for (i in 2:length(names(widen_joinlad))){
  widen_joinlad[,i] = as.character(paste0(widen_joinlad[,i], '%'))
}
widen_joinlad %>%
  knitr::kable()

```

date_reported	Halton	Bolton	Waltham.Forest
2021-07-01	40.54%	0%	0%
2021-07-02	-100%	0%	0%
2021-07-03	0%	0%	30.77%
2021-07-04	-100%	0%	0%
2021-07-05	-100%	0%	0%
2021-07-06	0%	-100%	0%
2021-07-07	0%	0%	-100%
2021-07-08	-100%	0%	0%
2021-07-09	0%	-100%	0%
2021-07-10	0%	0%	2.35%
2021-07-11	0%	19.64%	0%
2021-07-12	0%	-100%	0%
2021-07-13	10.29%	0%	0%
2021-07-14	0%	-100%	0%
2021-07-15	0%	0%	-100%
2021-07-16	-100%	0%	0%
2021-07-17	0%	0%	-100%
2021-07-18	0%	-100%	0%
2021-07-19	45.31%	0%	0%
2021-07-20	-100%	0%	0%
2021-07-21	0%	0%	1.19%
2021-07-22	-100%	0%	0%
2021-07-23	0%	0%	-38.12%
2021-07-24	0%	-53.22%	0%
2021-07-25	0%	0%	-22.22%
2021-07-26	-100%	0%	0%
2021-07-27	0%	-100%	0%
2021-07-28	0%	0%	-21.18%

date_reported	Halton	Bolton	Waltham.Forest
2021-07-29	0%	-27.36%	0%
2021-07-30	0%	0%	-12.8%
2021-07-31	0%	-100%	0%

## Part 2: 2.3

### Building function for Max of each month

```
ave_month <- function(LAD){
high_M <- joinedlad %>%
  dplyr::filter(area_name == LAD) %>%
  dplyr::select(date_reported, new_first_vaccine_dose) %>%
  dplyr::slice_max(new_first_vaccine_dose) %>%
  dplyr::select(date_reported) %>%
  dplyr::pull(1,1) %>%
  lubridate::month()

joinedlad %>%
  dplyr::filter(area_name == LAD) %>%
  dplyr::filter(lubridate::month(date_reported) == high_M) %>%
  dplyr::mutate(week_no =lubridate::week(date_reported) ) %>%
  dplyr::group_by(week_no) %>%
  dplyr::select(new_first_vaccine_dose, date_reported) %>%
  dplyr::mutate(Weeks = as.character(paste(first(date_reported), 'to', last(date_reported)))) %>%
  dplyr::ungroup() %>%
  dplyr::group_by(Weeks) %>%
  dplyr::summarise( Average_number_of_first_doses = mean(new_first_vaccine_dose)) %>%
  dplyr::select(Weeks, Average_number_of_first_doses) %>%
  knitr::kable()

}

ave_month('Halton')
```

## Adding missing grouping variables: 'week\_no'

Weeks	Average_number_of_first_doses
2021-03-01 to 2021-03-04	449.2500
2021-03-05 to 2021-03-11	101.0000
2021-03-12 to 2021-03-18	549.2857
2021-03-19 to 2021-03-25	722.0000
2021-03-26 to 2021-03-31	222.3333

```
ave_month('Waltham Forest')
```

## Adding missing grouping variables: 'week\_no'

Weeks	Average_number_of_first_doses
2021-05-01 to 2021-05-06	150.66667
2021-05-07 to 2021-05-13	224.00000
2021-05-14 to 2021-05-20	221.71429
2021-05-21 to 2021-05-27	96.28571
2021-05-28 to 2021-05-31	901.00000

```
ave_month('Bolton')
```

```
## Adding missing grouping variables: 'week_no'
```

Weeks	Average_number_of_first_doses
2021-03-01 to 2021-03-04	855.0000
2021-03-05 to 2021-03-11	588.5714
2021-03-12 to 2021-03-18	334.2857
2021-03-19 to 2021-03-25	575.4286
2021-03-26 to 2021-03-31	527.3333

```
#Part 2: 2.3 average daily new first vaccination per week
```

```
high_M2 <- joinedlad %>%
  dplyr::select(date_reported, new_first_vaccine_dose) %>%
  dplyr::slice_max(new_first_vaccine_dose) %>%
  dplyr::select(date_reported) %>%
  dplyr::pull(1,1) %>%
  lubridate::month()

joinedlad %>%
  dplyr::filter(lubridate::month(date_reported) == high_M2) %>%
  dplyr::mutate(week_no =lubridate::week(date_reported) ) %>%
  dplyr::group_by(week_no) %>%
  dplyr::select(new_first_vaccine_dose, date_reported) %>%
  dplyr::mutate(Weeks = as.character(paste(first(date_reported), 'to', last(date_reported)))) %>%
  dplyr::ungroup() %>%
  dplyr::group_by(Weeks) %>%
  dplyr::summarise( Average_number_of_first_doses = mean(new_first_vaccine_dose)) %>%
  dplyr::select(Weeks, Average_number_of_first_doses) %>%
  knitr::kable()
```

```
## Adding missing grouping variables: 'week_no'
```

Weeks	Average_number_of_first_doses
2021-03-01 to 2021-03-04	543.7500
2021-03-05 to 2021-03-11	258.0476
2021-03-12 to 2021-03-18	308.2381
2021-03-19 to 2021-03-25	517.7143

Weeks	Average_number_of_first_doses
2021-03-26 to 2021-03-31	249.8889

## Part 3

```
boltcomp <- function(x){
  data1 %>%
    filter(area_name == 'Bolton') %>%
    filter(metric == x) %>%
    dplyr::mutate(Month = lubridate::month(date_reported)) %>%
    dplyr::group_by(Month) %>%
    dplyr::summarise(mon_max = max(value))
}

englcomp <- function(x){
  data1 %>%
    filter(metric == x) %>%
    dplyr::mutate(Month = lubridate::month(date_reported)) %>%
    dplyr::group_by(Month) %>%
    dplyr::summarise(mon_max = max(value))
}

boltcomp('cumulative_cases')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1    21778
## 2     2    24370
## 3     3    25655
## 4     4    26321
## 5     5    30407
## 6     6    33922
## 7     7    38316
## 8     8    41202
## 9     9    44337
## 10    10    44529
## 11    12    17525
```

```
englcomp('cumulative_cases')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1    87592
## 2     2    96411
## 3     3   100120
## 4     4   101349
## 5     5   102676
```

```
## 6      6 110264
## 7      7 131320
## 8      8 147119
## 9      9 160047
## 10     10 160725
## 11     12  57057
```

```
boltcomp('new_cases')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1     208
## 2     2     150
## 3     3      64
## 4     4      37
## 5     5     285
## 6     6     163
## 7     7     278
## 8     8     145
## 9     9     154
## 10    10     109
## 11    12     166
```

```
englcomp('new_cases')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1    1701
## 2     2     595
## 3     3     179
## 4     4     110
## 5     5     285
## 6     6     630
## 7     7    1219
## 8     8    1101
## 9     9    1220
## 10    10     412
## 11    12    1423
```

```
boltcomp('new_first_vaccine_dose')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1    3764
## 2     2    3022
## 3     3    3016
## 4     4    1246
## 5     5    4234
## 6     6     929
```

```
## 7      7      381
## 8      8      236
## 9      9      186
## 10     10       44
## 11     12     577
```

```
englcomp('new_first_vaccine_dose')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1     10649
## 2     2     11390
## 3     3     11832
## 4     4      3531
## 5     5     4453
## 6     6     4746
## 7     7     2691
## 8     8     1350
## 9     9      964
## 10    10     1913
## 11    12     3078
```

```
boltcomp('new_second_vaccine_dose')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1       470
## 2     2        70
## 3     3     1700
## 4     4     4343
## 5     5     3825
## 6     6     2876
## 7     7     1581
## 8     8      868
## 9     9      400
## 10    10      143
## 11    12        12
```

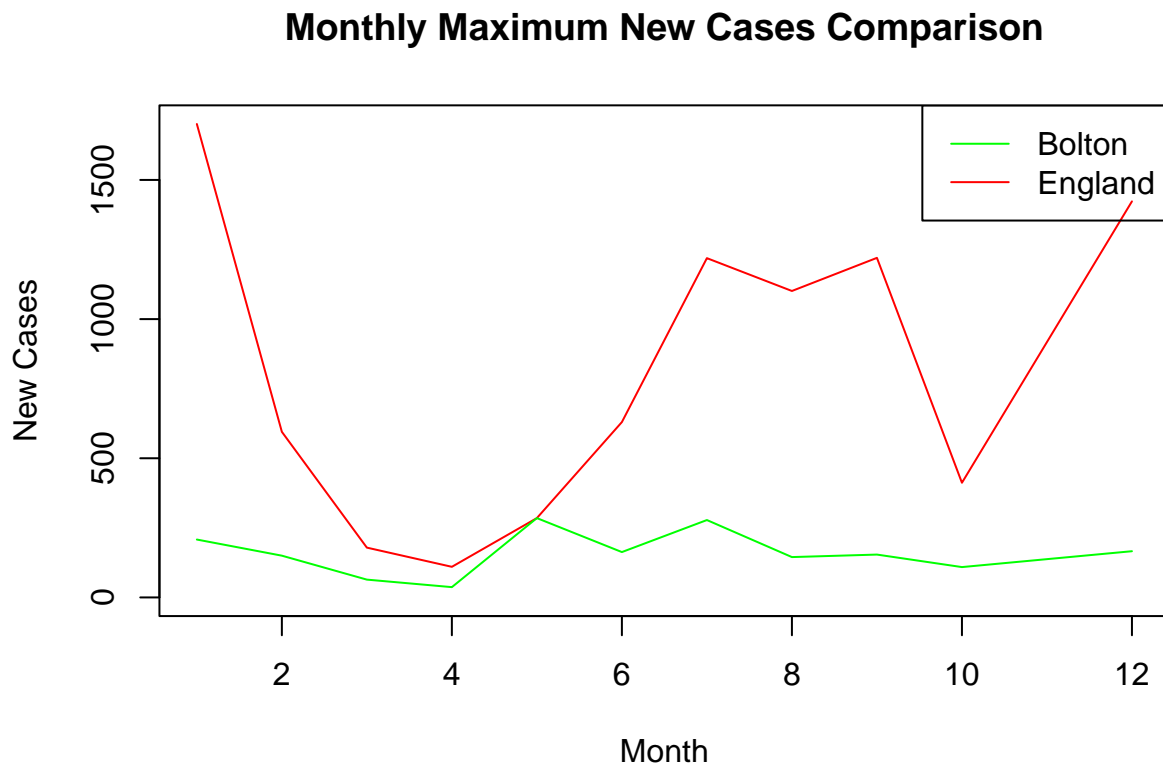
```
englcomp('new_second_vaccine_dose')
```

```
## # A tibble: 11 x 2
##   Month mon_max
##   <dbl>   <int>
## 1     1     2544
## 2     2     1113
## 3     3     8704
## 4     4     9789
## 5     5    11079
## 6     6     9389
## 7     7     3658
```



```
## 8      8      4189
## 9      9      2310
## 10     10      754
## 11     12      248
```

```
#New Cases Comparison Chart
plot(englcomp('new_cases'), type = 'l', col = 'red',
     main = 'Monthly Maximum New Cases Comparison', xlab= 'Month',
     ylab = 'New Cases', ylim=c(1, 1700))
lines(boltcomp('new_cases'), type = "l", col = "green")
legend("topright",
     legend = c("Bolton", "England"),
     col = c("green", "red"),
     lty = 1)
```



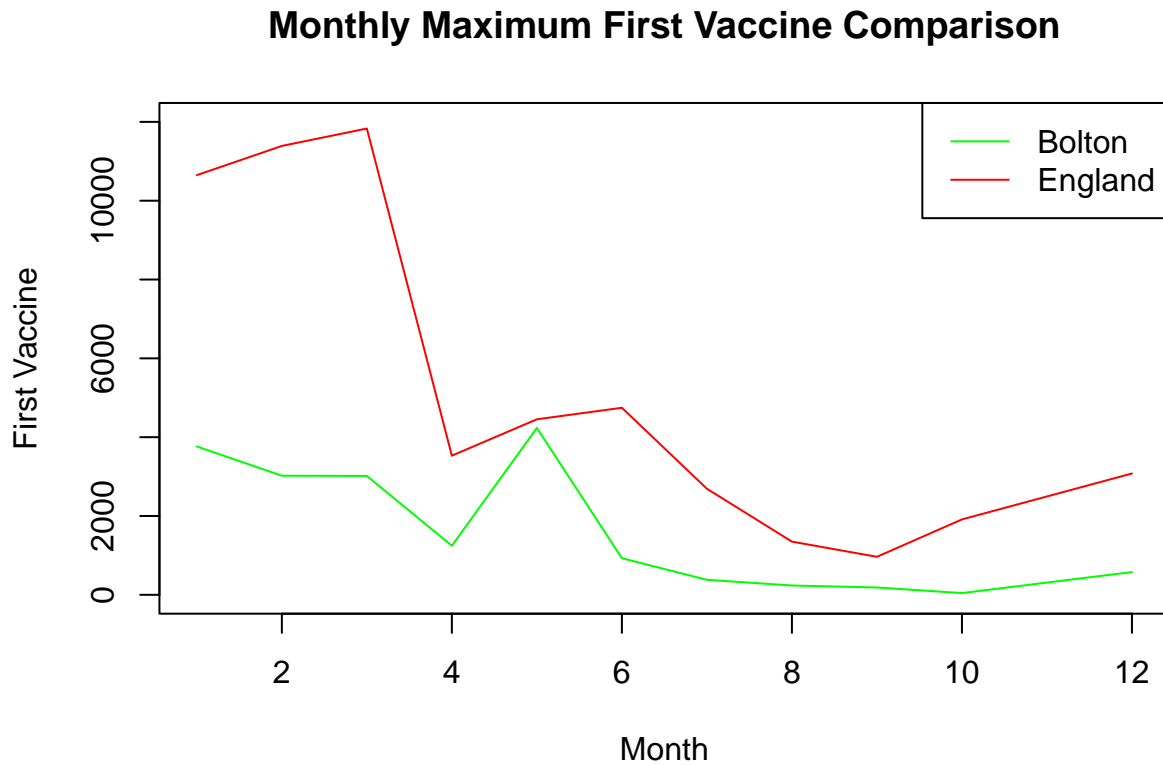
The chart shows that England had its highest maximum new cases in January while Bolton had its highest maximum new cases in May. However, both England and Bolton had their lowest maximum new cases on April. This shows that there is a high chance that new cases were low across the country in April. Also, England and Bolton had the same new cases trend from January to April.

```
#First Vaccine dose Comparison
plot(englcomp('new_first_vaccine_dose'), type = 'l', col = 'red',
     main = 'Monthly Maximum First Vaccine Comparison',
     xlab= 'Month', ylab = 'First Vaccine', ylim = c(1, 12000))
lines(boltcomp('new_first_vaccine_dose'), type = "l", col = "green")
legend("topright",
```

```

legend = c("Bolton", "England"),
col = c("green", "red"),
lty = 1)

```



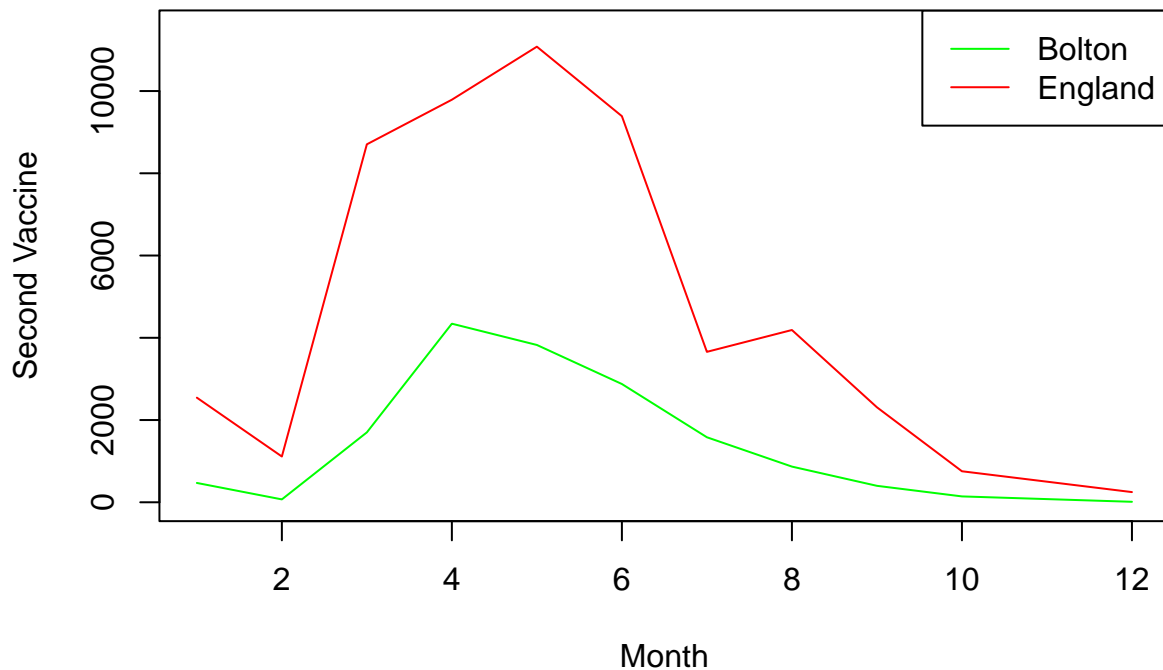
Unlike England which had its highest maximum first vaccine in March, Bolton had its highest maximum first vaccine in May. While England had its lowest maximum first vaccine in September, Bolton had its own in October. Both England and Bolton had similar first vaccine trend from October to December.

```

plot(englcomp('new_second_vaccine_dose'), type = 'l', col = 'red',
     main = 'Monthly Maximum Second Vaccine Comparison',
     xlab = 'Month', ylab = 'Second Vaccine', ylim = c(1,11500))
lines(boltcomp('new_second_vaccine_dose'), type = "l", col = "green")
legend("topright",
     legend = c("Bolton", "England"),
     col = c("green", "red"),
     lty = 1)

```

## Monthly Maximum Second Vaccine Comparison



Both England and Bolton displayed similar Maximum second vaccine trend from January to April. However, England reached its peak in May while Bolton peaked in April. A fall in trend was also noticed in both from May to December and both reach their lowest maximum second vaccine in December.

### Part 4: 4.0

Based on the information offered, it is evident that the data is a UK covid-19 data from 2020-12-06 to 2021-10-02, containing information about number of new and cumulative covid cases and number of first and second vaccines collected by people in different Local Authority District (LAD) of England. For the purpose of analysis, three LAD: Bolton, Waltham Forest and Halton were assigned to me. To efficiently perform this analysis, I used libraries like dplyr, tidyverse and pastec, among others.