class19 Pertussis and the CMI-PB project

Ruofan Kang(A17236920)

Pertussis is a a sever lung infection also known a whooping cough.

We will begin by investigating the number of Pertussis cases per year in the US.

This data is availabe on the CDC website here

```
#1 echo=FALSE

cdc <- data.frame( year = c(1922L,1923L,1924L,1925L,1926L,
                                          1927L,1928L,1929L,1930L,1931L,
                                          1932L,1933L,1934L,1935L,1936L,1937L,
                                          1938L,1939L,1940L,1941L,1942L,
                                          1943L,1944L,1945L,1946L,1947L,1948L,
                                          1949L,1950L,1951L,1952L,1953L,1954L,
                                          1955L,1956L,1957L,1958L,1959L,
                                          1960L,1961L,1962L,1963L,1964L,1965L,
                                          1966L,1967L,1968L,1969L,1970L,
                                          1971L,1972L,1973L,1974L,1975L,1976L,
                                          1977L,1978L,1979L,1980L,1981L,
                                          1982L,1983L,1984L,1985L,1986L,1987L,
                                          1988L,1989L,1990L,1991L,1992L,1993L,
                                          1994L,1995L,1996L,1997L,1998L,
                                          1999L,2000L,2001L,2002L,2003L,2004L,
                                          2005L,2006L,2007L,2008L,2009L,
                                          2010L,2011L,2012L,2013L,2014L,2015L,
                                          2016L,2017L,2018L,2019L,2020L,
                                          2021L), cases = c(107473,164191,165418,152003,
                                          202210,181411,161799,197371,166914,
                                          172559,215343,179135,265269,180518,
                                          147237,214652,227319,103188,183866,
                                          222202,191383,191890,109873,133792,
                                          109860,156517,74715,69479,120718,
```

```
                                        68687,45030,37129,60886,62786,
                                        31732,28295,32148,40005,14809,11468,
                                        17749,17135,13005,6799,7717,9718,
                                        4810,3285,4249,3036,3287,1759,
                                        2402,1738,1010,2177,2063,1623,1730,
                                        1248,1895,2463,2276,3589,4195,
                                        2823,3450,4157,4570,2719,4083,6586,
                                        4617,5137,7796,6564,7405,7298,
                                        7867,7580,9771,11647,25827,25616,
                                        15632,10454,13278,16858,27550,18719,
                                        48277,28639,32971,20762,17972,
                                        18975,15609,18617,6124,2116)
       )
```

Lets have a wee look at this table

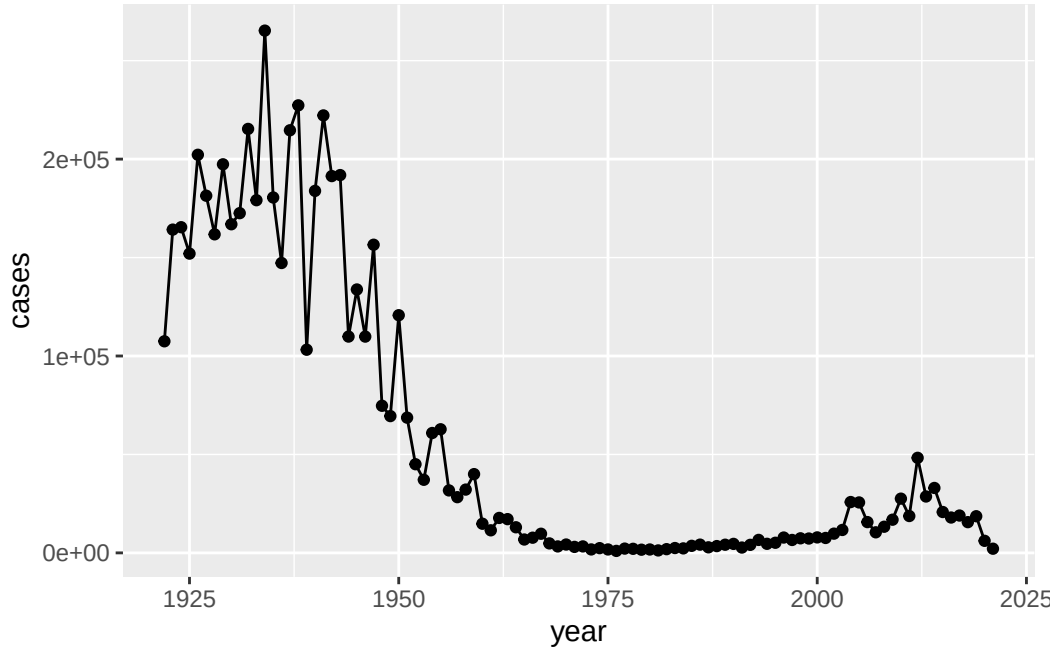
```
head(cdc)
```

```
  year  cases
1 1922 107473
2 1923 164191
3 1924 165418
4 1925 152003
5 1926 202210
6 1927 181411
```

I want a nice plot of the number of cases per year.

> Q1. With the help of the R "addin" package datapasta assign the CDC pertussis case number data to a data frame called cdc and use ggplot to make a plot of cases numbers over time.

```
library(ggplot2)

ggplot(cdc) +
  aes(x = year, y = cases) +
  geom_point() +
  geom_line()
```



Q2. Using the ggplot geom_vline() function add lines to your previous plot for the 1946 introduction of the wP vaccine and the 1996 switch to aP vaccine (see example in the hint below). What do you notice?

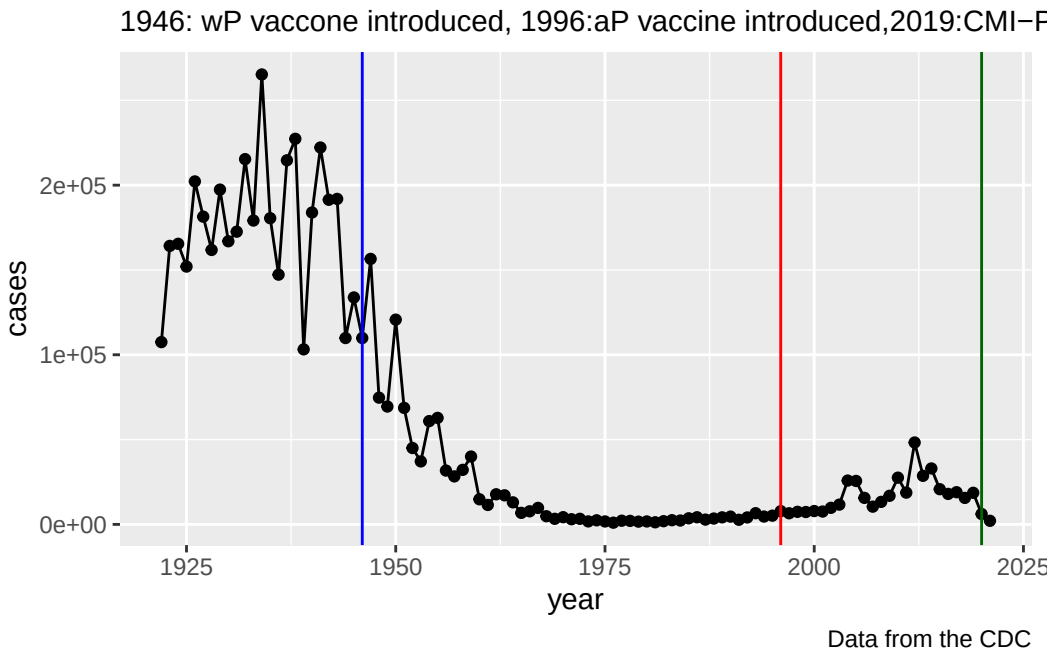
```
ggplot(cdc) +
  aes(x = year, y = cases) +
  geom_point() +
  geom_line() +
  geom_vline(xintercept = 1946, Linetype="dashed", col="blue") +
  geom_vline(xintercept = 1996, Linetype="dashed", col="red") +
  geom_vline(xintercept = 2020, Linetype="dashed", col="darkgreen") +
  labs(titl="Pertussis cases per year in the US",
      subtitle="1946: wP vaccone introduced, 1996:aP vaccine introduced,2019:CMI-PB projec
      caption="Data from the CDC")
```

```
Warning in geom_vline(xintercept = 1946, Linetype = "dashed", col = "blue"):
Ignoring unknown parameters: `Linetype`
```

```
Warning in geom_vline(xintercept = 1996, Linetype = "dashed", col = "red"):
Ignoring unknown parameters: `Linetype`
```

```
Warning in geom_vline(xintercept = 2020, Linetype = "dashed", col =
```

```
"darkgreen"): Ignoring unknown parameters: `Linetype`
```



Q3. Describe what happened after the introduction of the aP vaccine? Do you have a possible explanation for the observed trend?

There is a lag and then cases rise with a ~3 year cycle perhaps similar to that observed pior to the first wP vaccine introduction.

3.Exploring CMI-PB data

Why is this vaccine-preventable disease on the upswing? To answer this question we need to investigate the mechanisms underlying waning protection against pertussis. This requires evaluation of pertussis-specific immune responses over time in wP and aP vaccinated individuals.

This is the goals of the CMI-PB project: https://www.cmi-pb.org/

The CMI-PB project makes its data available via "API-endpoint" that return JSON format.

We will use the 'jasonlite package to access this data. The main function in this package is called 'read_json()'

```
library(jsonlite)

# Subject table
subject <- read_json("https://www.cmi-pb.org/api/subject", simplifyVector = TRUE)
specimen <- read_json("http://cmi-pb.org/api/specimen", simplifyVector = TRUE)
titer <- read_json("http://cmi-pb.org/api/v4/plasma_ab_titer", simplifyVector = TRUE )
```

Have a wee peak at these new object

```
head(subject)
```

```
  subject_id infancy_vac biological_sex              ethnicity  race
1          1          wP         Female Not Hispanic or Latino White
2          2          wP         Female Not Hispanic or Latino White
3          3          wP         Female                 Unknown White
4          4          wP           Male Not Hispanic or Latino Asian
5          5          wP           Male Not Hispanic or Latino Asian
6          6          wP         Female Not Hispanic or Latino White
  year_of_birth date_of_boost      dataset
1    1986-01-01    2016-09-12 2020_dataset
2    1968-01-01    2019-01-28 2020_dataset
3    1983-01-01    2016-10-10 2020_dataset
4    1988-01-01    2016-08-29 2020_dataset
5    1991-01-01    2016-08-29 2020_dataset
6    1988-01-01    2016-10-10 2020_dataset
```

```
head(titer)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1     IgE               FALSE   Total 1110.21154       2.493425
2           1     IgE               FALSE   Total 2708.91616       2.493425
3           1     IgG                TRUE      PT   68.56614       3.736992
4           1     IgG                TRUE     PRN  332.12718       2.602350
5           1     IgG                TRUE     FHA 1887.12263      34.050956
6           1     IgE                TRUE     ACT    0.10000       1.000000
   unit lower_limit_of_detection
1 UG/ML                 2.096133
2 IU/ML                29.170000
3 IU/ML                 0.530000
4 IU/ML                 6.205949
```

```
5 IU/ML                 4.679535
6 IU/ML                 2.816431
```

Q4. How many aP and wP infancy vaccinated subjects are in the dataset?

```
table(subject$infancy_vac)
```

```
aP wP
60 58
```

Q5. How many Male and Female subjects/patients are in the dataset?

```
table(subject$biological_sex)
```

```
Female   Male
    79     39
```

Q6. What is the breakdown of race and biological sex (e.g. number of Asian females, White males etc…)?

```
table(subject$race, subject$biological_sex)
```

```
                                        Female Male
  American Indian/Alaska Native              0    1
  Asian                                     21   11
  Black or African American                  2    0
  More Than One Race                         9    2
  Native Hawaiian or Other Pacific Islander  1    1
  Unknown or Not Reported                   11    4
  White                                     35   20
```

```
#1 echo=FALSE, message=FALSE,warning=FALSE
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v lubridate 1.9.3     v tibble    3.2.1
v purrr     1.0.2     v tidyr     1.3.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()  masks stats::filter()
x purrr::flatten() masks jsonlite::flatten()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

working with dates

dates can really suck to work with and do math. The lubridate package makes this much easir.
It is parts of the **tidyverse** that includes dplyr,ggplots,etc

```
today()
```

```
[1] "2023-12-05"
```

```
today() - mdy("11-23-2001")
```

```
Time difference of 8047 days
```

```
time_length( today() - ymd("2001-11-23"),  "years")
```

```
[1] 22.03149
```

Now add the age of each subject to the subject table.

```
subject$age <- ymd(subject$date_of_boost) - ymd(subject$year_of_birth)

subject$age_years <- time_length(subject$age,"years")

head(subject)
```

```
  subject_id infancy_vac biological_sex              ethnicity  race
1          1          wP         Female Not Hispanic or Latino White
2          2          wP         Female Not Hispanic or Latino White
3          3          wP         Female                 Unknown White
4          4          wP           Male Not Hispanic or Latino Asian
5          5          wP           Male Not Hispanic or Latino Asian
6          6          wP         Female Not Hispanic or Latino White
  year_of_birth date_of_boost      dataset        age age_years
1    1986-01-01    2016-09-12 2020_dataset 11212 days  30.69678
2    1968-01-01    2019-01-28 2020_dataset 18655 days  51.07461
3    1983-01-01    2016-10-10 2020_dataset 12336 days  33.77413
4    1988-01-01    2016-08-29 2020_dataset 10468 days  28.65982
5    1991-01-01    2016-08-29 2020_dataset  9372 days  25.65914
6    1988-01-01    2016-10-10 2020_dataset 10510 days  28.77481
```

Q7. Using this approach determine (i) the average age of wP individuals, (ii) the average age of aP individuals; and (iii) are they significantly different?

The mean age for the aP group was 21, while for the wP group, it was 31, indicating a difference of 10 years. Both the median and mean for the wP group exceeded those for the aP group.

```
library(dplyr)

ap <- subject %>% filter(infancy_vac == "aP")

round( summary( time_length( ap$age, "years" ) ) )
```

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   19      20      20      21      21      28
```

```
wp <- subject %>% filter(infancy_vac == "wP")

wp_age_summary <- round(summary(time_length(wp$age, "years")))
print(wp_age_summary)
```

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   23      26      29      31      34      51
```

```
t.test(ap$age, wp$age)
```

```
    Welch Two Sample t-test

data:  ap$age and wp$age
t = -12.088 days, df = 69.647, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4186.413 days -3000.486 days
sample estimates:
Time differences in days
mean of x mean of y
 7608.533 11201.983
```

Q8. Determine the age of all individuals at time of boost?

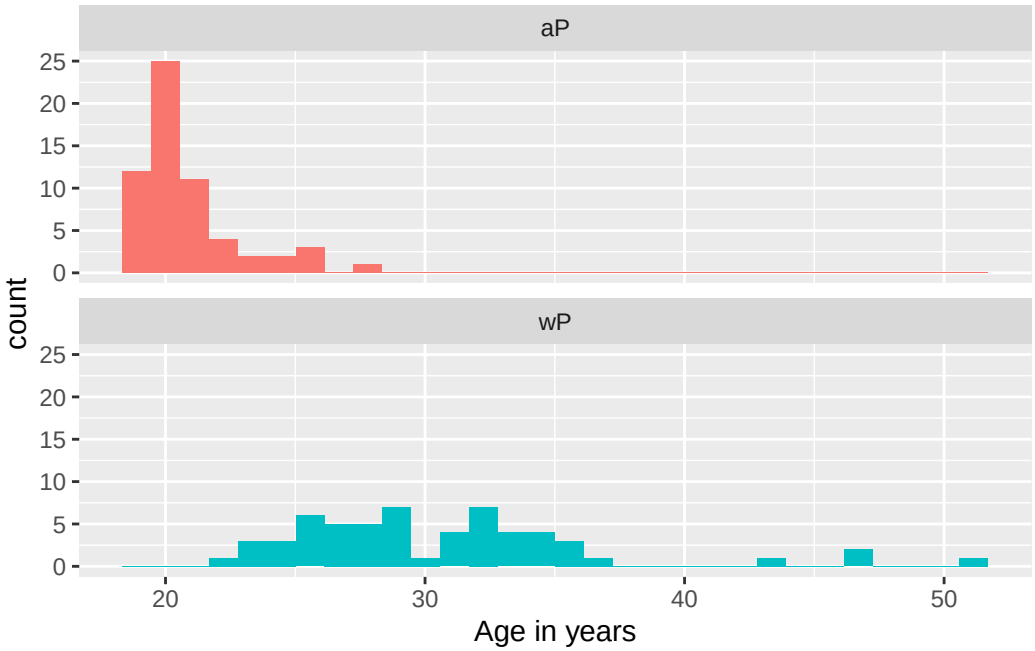
```
int <- ymd(subject$date_of_boost) - ymd(subject$year_of_birth)
age_at_boost <- time_length(int, "year")
head(age_at_boost)
```

```
[1] 30.69678 51.07461 33.77413 28.65982 25.65914 28.77481
```

Q9. With the help of a faceted boxplot or histogram (see below), do you think these two groups are significantly different?

```
ggplot(subject) +
  aes(time_length(age, "year"),
      fill=as.factor(infancy_vac)) +
  geom_histogram(show.legend=FALSE) +
  facet_wrap(vars(infancy_vac), nrow=2) +
  xlab("Age in years")
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# Or use wilcox.test()
x <- t.test(time_length( wp$age, "years" ),
       time_length( ap$age, "years" ))

x$p.value
```

```
[1] 9.121472e-19
```

Merge or join tables

Q9. Complete the code to join specimen and subject tables to make a new merged data frame containing all specimen records along with their associated subject details:

```
meta <- inner_join(specimen,subject)
```

```
Joining with `by = join_by(subject_id)`
```

```
head(meta)
```

```
  specimen_id subject_id actual_day_relative_to_boost
1           1          1                           -3
2           2          1                            1
3           3          1                            3
4           4          1                            7
5           5          1                           11
6           6          1                           32
  planned_day_relative_to_boost specimen_type visit infancy_vac biological_sex
1                             0         Blood     1          wP         Female
2                             1         Blood     2          wP         Female
3                             3         Blood     3          wP         Female
4                             7         Blood     4          wP         Female
5                            14         Blood     5          wP         Female
6                            30         Blood     6          wP         Female
               ethnicity  race year_of_birth date_of_boost      dataset
1 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
2 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
3 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
4 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
5 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
6 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
         age age_years
1 11212 days  30.69678
2 11212 days  30.69678
3 11212 days  30.69678
4 11212 days  30.69678
5 11212 days  30.69678
6 11212 days  30.69678
```

Antiboday measurments in the blood

Q10. Now using the same procedure join meta with titer data so we can further analyze this data in terms of time of

```
abdata <- inner_join(titer, meta)
```

```
Joining with `by = join_by(specimen_id)`
```

```
dim(abdata)
```

```
[1] 41810    22
```

Q11. How many specimens (i.e. entries in abdata) do we have for each isotype?

```
table(abdata$isotype)
```

```
 IgE  IgG IgG1 IgG2 IgG3 IgG4
6698 3240 7968 7968 7968 7968
```

Q12. What are the different $dataset values in abdata and what do you notice about the number of rows for the most "recent" dataset?

```
dataset_counts <- table(abdata$dataset)

print(dataset_counts)
```

```
2020_dataset 2021_dataset 2022_dataset
       31520         8085         2205
```

```
abdata <- inner_join(titer, meta)
```

```
Joining with `by = join_by(specimen_id)`
```

```
head(abdata)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1     IgE               FALSE   Total 1110.21154       2.493425
2           1     IgE               FALSE   Total 2708.91616       2.493425
3           1     IgG                TRUE      PT   68.56614       3.736992
4           1     IgG                TRUE     PRN  332.12718       2.602350
5           1     IgG                TRUE     FHA 1887.12263      34.050956
6           1     IgE                TRUE     ACT    0.10000       1.000000
   unit lower_limit_of_detection subject_id actual_day_relative_to_boost
1 UG/ML                 2.096133          1                           -3
2 IU/ML                29.170000          1                           -3
3 IU/ML                 0.530000          1                           -3
4 IU/ML                 6.205949          1                           -3
5 IU/ML                 4.679535          1                           -3
```

```
6 IU/ML                  2.816431          1                          -3
  planned_day_relative_to_boost specimen_type visit infancy_vac biological_sex
1                             0         Blood     1          wP         Female
2                             0         Blood     1          wP         Female
3                             0         Blood     1          wP         Female
4                             0         Blood     1          wP         Female
5                             0         Blood     1          wP         Female
6                             0         Blood     1          wP         Female
               ethnicity  race year_of_birth date_of_boost      dataset
1 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
2 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
3 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
4 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
5 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
6 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
         age age_years
1 11212 days  30.69678
2 11212 days  30.69678
3 11212 days  30.69678
4 11212 days  30.69678
5 11212 days  30.69678
6 11212 days  30.69678
```

Q.How many isotypes are we measuring for all these individuals?

```
table(abdata$isotype)
```

```
 IgE  IgG IgG1 IgG2 IgG3 IgG4
6698 3240 7968 7968 7968 7968
```

Lets focus on one of these IgG

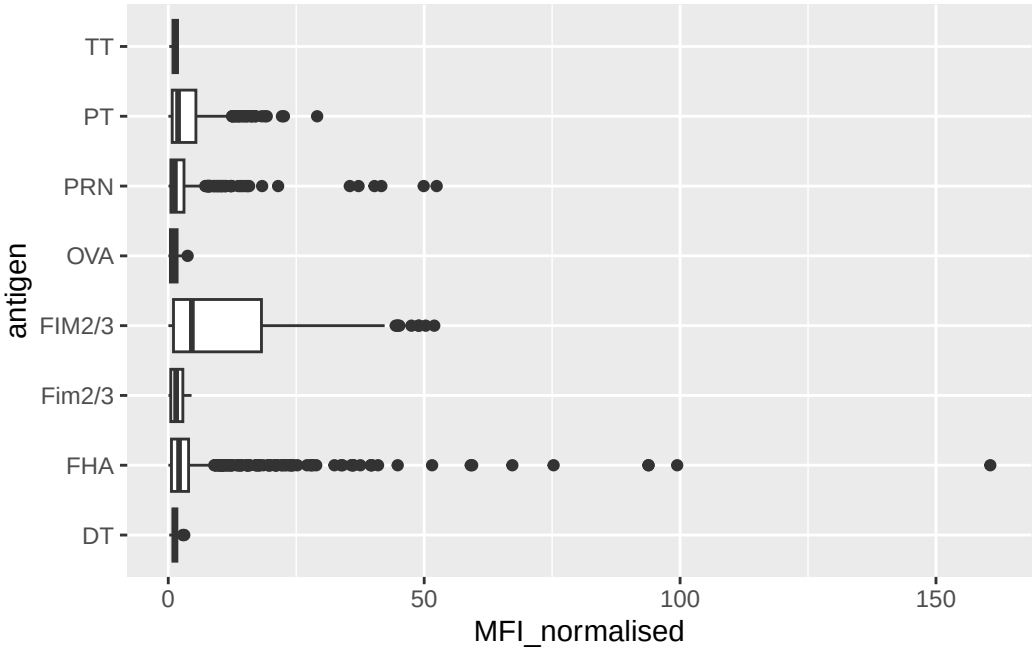
```
igg <- abdata %>% filter(isotype =="IgG")
head(igg)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1     IgG                TRUE      PT   68.56614       3.736992
2           1     IgG                TRUE     PRN  332.12718       2.602350
3           1     IgG                TRUE     FHA 1887.12263      34.050956
4          19     IgG                TRUE      PT   20.11607       1.096366
```

```
5          19     IgG                TRUE     PRN  976.67419       7.652635
6          19     IgG                TRUE     FHA   60.76626       1.096457
   unit lower_limit_of_detection subject_id actual_day_relative_to_boost
1 IU/ML                 0.530000          1                           -3
2 IU/ML                 6.205949          1                           -3
3 IU/ML                 4.679535          1                           -3
4 IU/ML                 0.530000          3                           -3
5 IU/ML                 6.205949          3                           -3
6 IU/ML                 4.679535          3                           -3
  planned_day_relative_to_boost specimen_type visit infancy_vac biological_sex
1                             0         Blood     1          wP         Female
2                             0         Blood     1          wP         Female
3                             0         Blood     1          wP         Female
4                             0         Blood     1          wP         Female
5                             0         Blood     1          wP         Female
6                             0         Blood     1          wP         Female
               ethnicity  race year_of_birth date_of_boost      dataset
1 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
2 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
3 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
4                Unknown White    1983-01-01    2016-10-10 2020_dataset
5                Unknown White    1983-01-01    2016-10-10 2020_dataset
6                Unknown White    1983-01-01    2016-10-10 2020_dataset
         age age_years
1 11212 days  30.69678
2 11212 days  30.69678
3 11212 days  30.69678
4 12336 days  33.77413
5 12336 days  33.77413
6 12336 days  33.77413
```

Box plot of MFI_normalised vs antigen

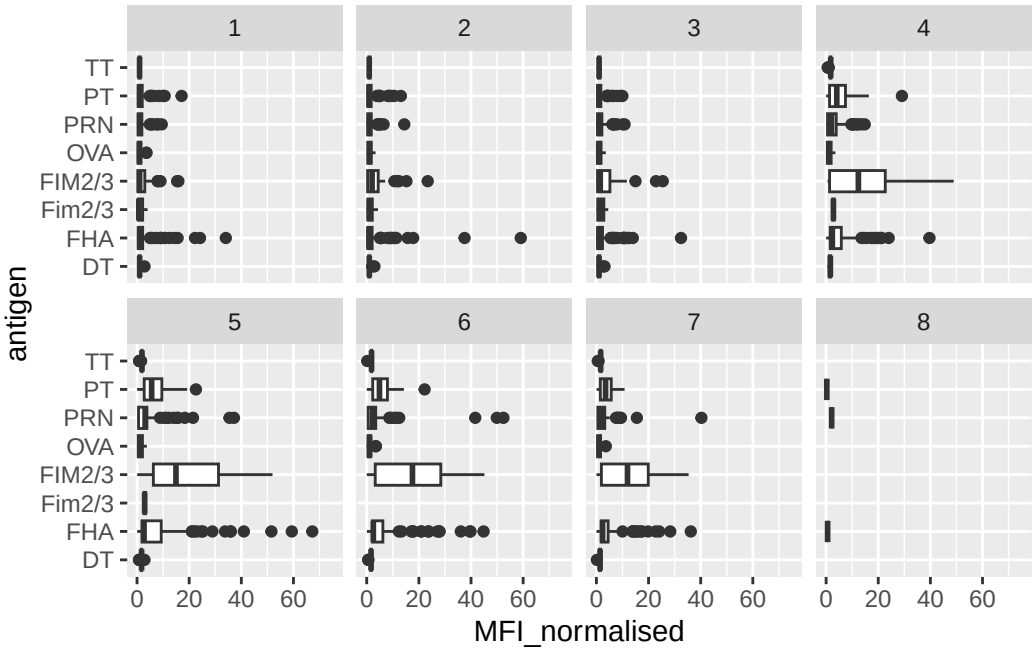
```
ggplot(igg)+
  aes(MFI_normalised ,
      antigen) +
  geom_boxplot()
```



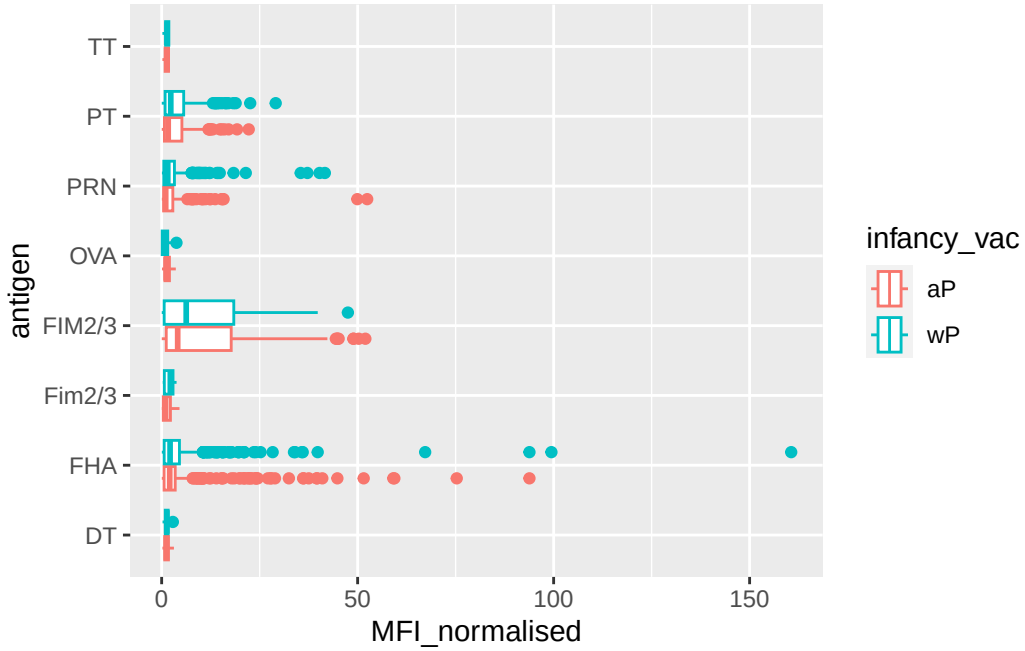
Q13. Complete the following code to make a summary boxplot of Ab titer levels (MFI) for all antigens:

```
ggplot(igg)+
  aes(MFI_normalised ,
      antigen) +
  geom_boxplot() +
  xlim(0,75) +
 facet_wrap(vars(visit),nrow = 2)
```

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

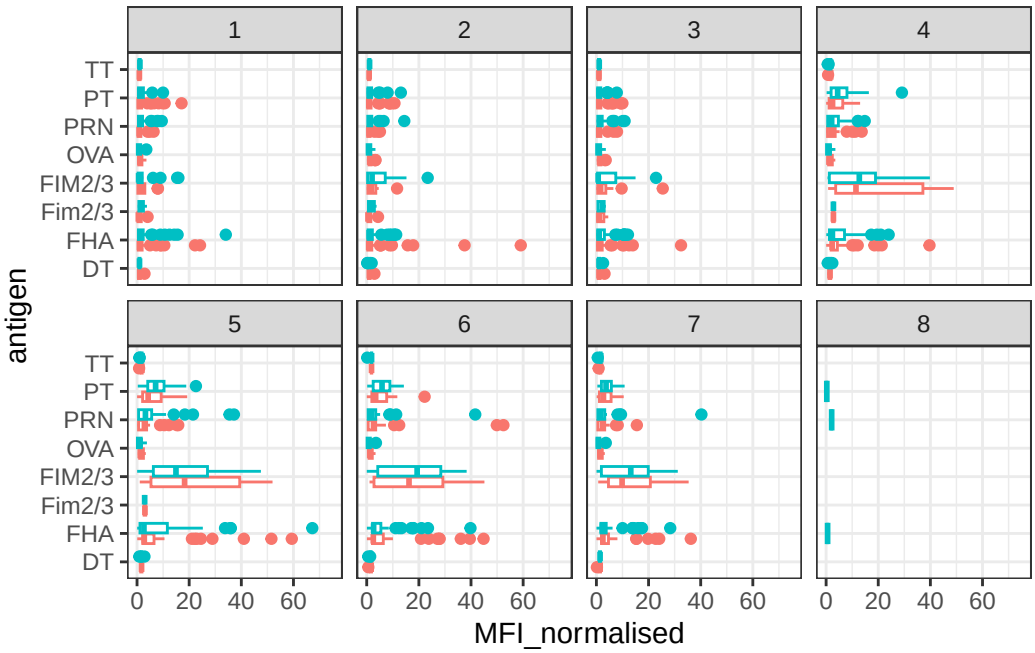



```
ggplot(igg)+
  aes(MFI_normalised ,
      antigen,
      col=infancy_vac) +
  geom_boxplot()
```



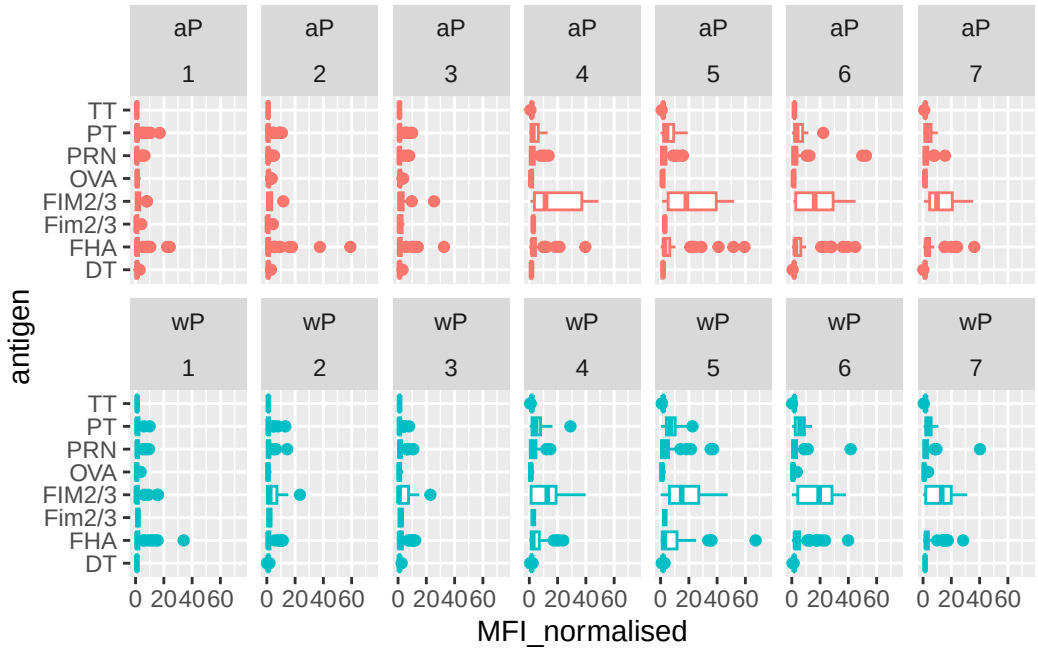

```
ggplot(igg)+
  aes(MFI_normalised ,
      antigen,
      col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit), nrow=2) +
  xlim(0,75) +
  theme_bw()
```

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



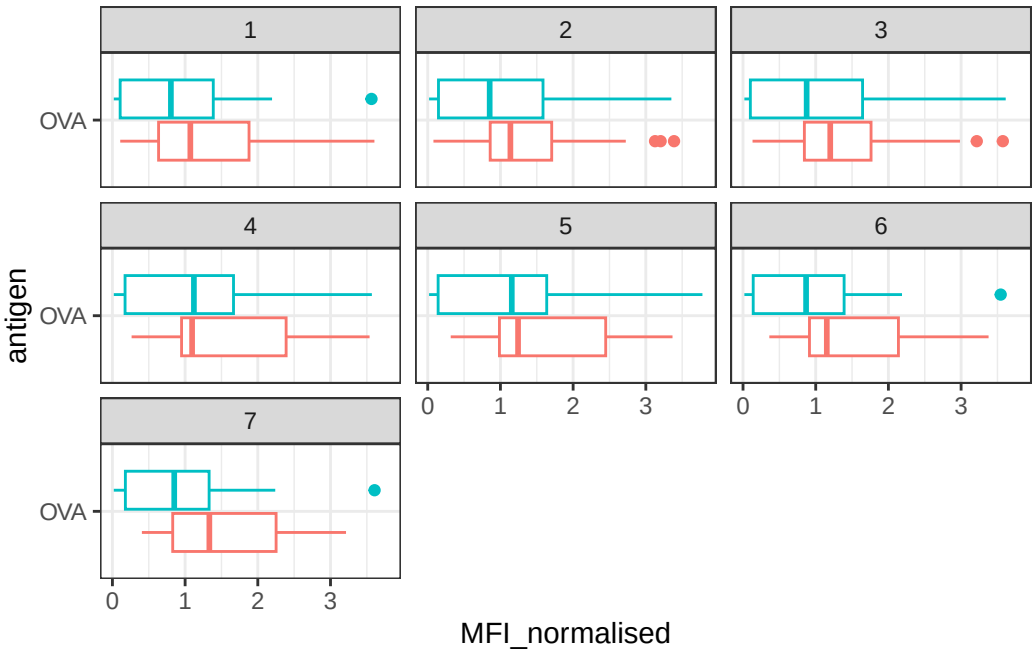

```
igg %>% filter(visit != 8) %>%
ggplot() +
  aes(MFI_normalised, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  xlim(0,75) +
  facet_wrap(vars(infancy_vac, visit), nrow=2)
```

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

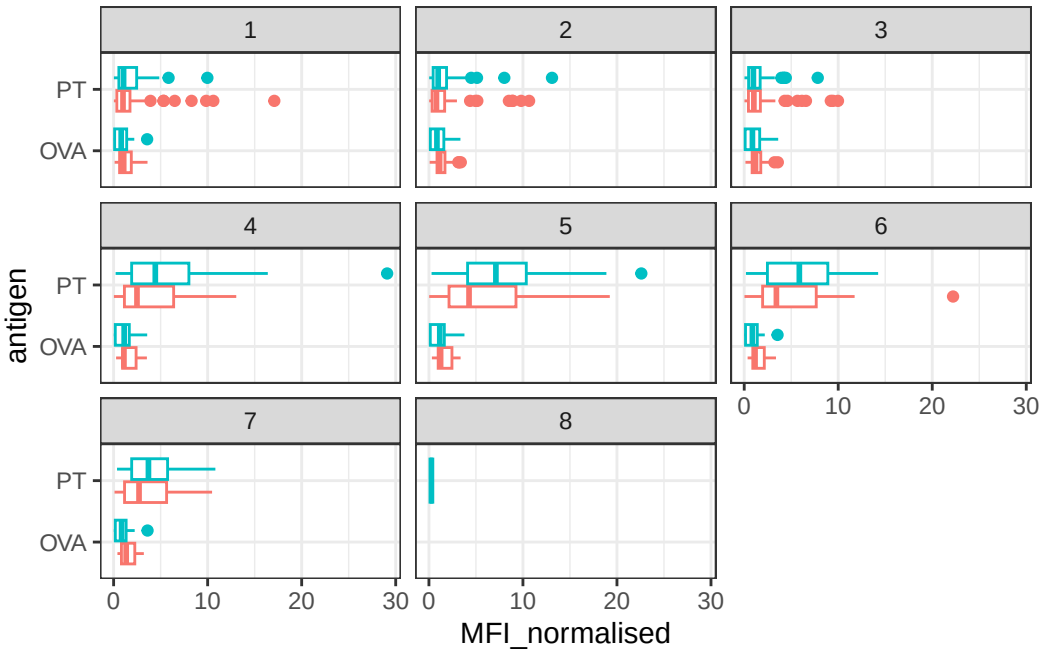


Q15. Filter to pull out only two specific antigens for analysis and create a boxplot for each. You can chose any you like. Below I picked a "control" antigen ("OVA", that is not in our vaccines) and a clear antigen of interest ("PT", Pertussis Toxin, one of the key virulence factors produced by the bacterium B. pertussis).

```
filter(igg, antigen=="OVA") %>%

  ggplot() +
  aes(MFI_normalised, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```




```
filtered_igg <- filter(igg, antigen %in% c("OVA", "PT"))


  ggplot(filtered_igg) +
  aes(MFI_normalised, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



Q16. What do you notice about these two antigens time courses and the PT data in particular?

Over time, both aP and wP vaccines elicited immune responses to PT antigens, with aP vaccines (red) generally showing a higher median response than the wP vaccine (teal), potentially indicating a stronger immune reaction to the acellular pertussis vaccine. Additionally, the OVA antigen levels remained consistent over time for both vaccine types, confirming its suitability as a control.

Q17. Do you see any clear difference in aP vs. wP responses?

The aP vaccine response (red) shows a generally higher median level of MFI compared to the wP vaccine (teal) . The response to PT increased over time for both vaccines, peaking at around 5, but the peak for aP seems to be higher, perhaps the aP vaccine has a stronger response. aP and wP vaccines show very similar trends of increase and decrease. Also, both vaccine types showed outliers, indicating individual differences in immune response.

```
head(igg)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1     IgG                TRUE      PT   68.56614       3.736992
2           1     IgG                TRUE     PRN  332.12718       2.602350
3           1     IgG                TRUE     FHA 1887.12263      34.050956
```

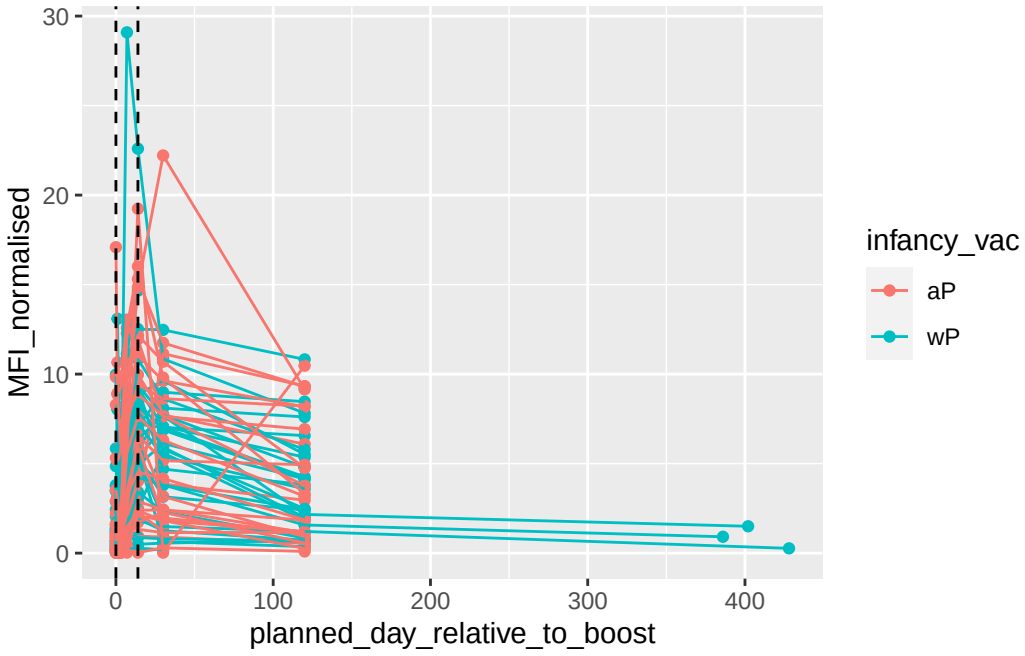
```
4          19     IgG                TRUE      PT   20.11607       1.096366
5          19     IgG                TRUE     PRN  976.67419       7.652635
6          19     IgG                TRUE     FHA   60.76626       1.096457
   unit lower_limit_of_detection subject_id actual_day_relative_to_boost
1 IU/ML                 0.530000          1                           -3
2 IU/ML                 6.205949          1                           -3
3 IU/ML                 4.679535          1                           -3
4 IU/ML                 0.530000          3                           -3
5 IU/ML                 6.205949          3                           -3
6 IU/ML                 4.679535          3                           -3
  planned_day_relative_to_boost specimen_type visit infancy_vac biological_sex
1                             0         Blood     1          wP         Female
2                             0         Blood     1          wP         Female
3                             0         Blood     1          wP         Female
4                             0         Blood     1          wP         Female
5                             0         Blood     1          wP         Female
6                             0         Blood     1          wP         Female
               ethnicity  race year_of_birth date_of_boost      dataset
1 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
2 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
3 Not Hispanic or Latino White    1986-01-01    2016-09-12 2020_dataset
4                Unknown White    1983-01-01    2016-10-10 2020_dataset
5                Unknown White    1983-01-01    2016-10-10 2020_dataset
6                Unknown White    1983-01-01    2016-10-10 2020_dataset
         age age_years
1 11212 days  30.69678
2 11212 days  30.69678
3 11212 days  30.69678
4 12336 days  33.77413
5 12336 days  33.77413
6 12336 days  33.77413
```

Focus in on IgG to the pertussis Toxin(PT) antigen in the 2021 dataset

```
igg.pt <- igg %>% filter(antigen == "PT", dataset=="2020_dataset")
```

```
ggplot(igg.pt) +
  aes(planned_day_relative_to_boost,
      MFI_normalised,
      col=infancy_vac,
      group=subject_id) +
```

```
  geom_point() +
  geom_line() +
  geom_vline(xintercept = 0, linetype = "dashed", col = "black") +
  geom_vline(xintercept = 14, linetype = "dashed", col = "black")
```

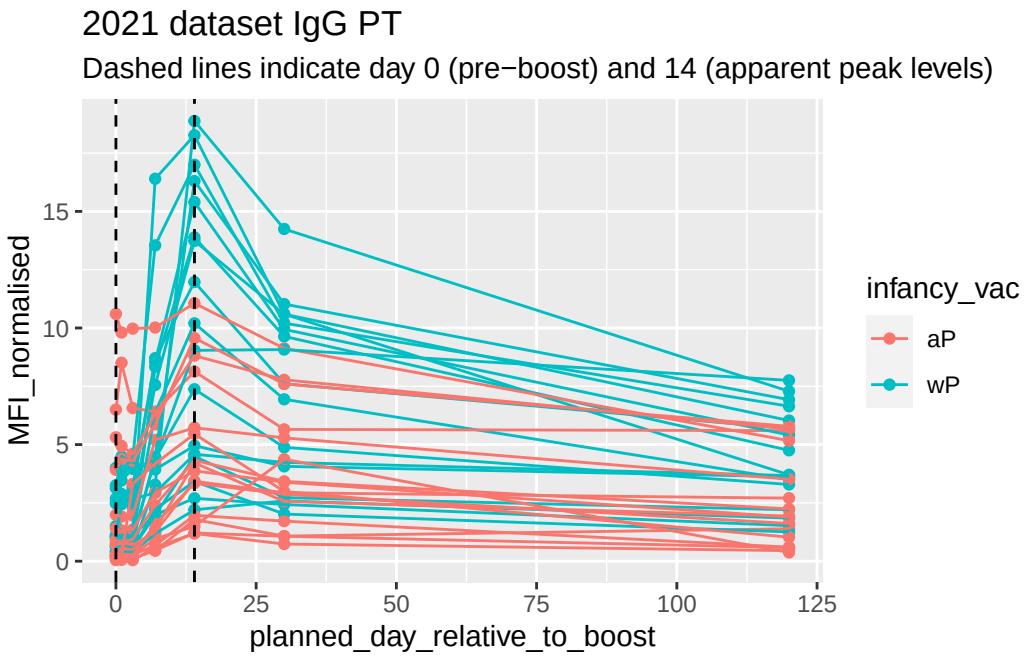



```
  labs(title="2020 dataset IgG PT",
         subtitle = "Dashed lines indicate day 0 (pre-boost) and 14 (apparent peak levels)")
```

```
$title
[1] "2020 dataset IgG PT"

$subtitle
[1] "Dashed lines indicate day 0 (pre-boost) and 14 (apparent peak levels)"

attr(,"class")
[1] "labels"
```

```
  abdata.21 <- abdata %>% filter(dataset == "2021_dataset")

  abdata.21 %>%
```

```
  filter(isotype == "IgG",  antigen == "PT") %>%
  ggplot() +
    aes(x=planned_day_relative_to_boost,
        y=MFI_normalised,
        col=infancy_vac,
        group=subject_id) +
    geom_point() +
    geom_line() +
    geom_vline(xintercept=0, linetype="dashed") +
    geom_vline(xintercept=14, linetype="dashed") +
  labs(title="2021 dataset IgG PT",
       subtitle = "Dashed lines indicate day 0 (pre-boost) and 14 (apparent peak levels)")
```



Q18. Does this trend look similar for the 2020 dataset?

No, they looks not similar.