Class 19

Yu-Chia Huang (PID: A59026739)

Import the data:

Pertussis is a bacterial infection that causes a severe cough. Often named "whooping cough".

Let's have a look at case numbers of Pertussis in the US.

> 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.

##1. Investigating pertussis cases by year

```
#install.packages("datapasta")
```

```
head(cdc)
```

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

```
library(ggplot2)
#install.packages("tidyverse")
library(tidyverse)
```

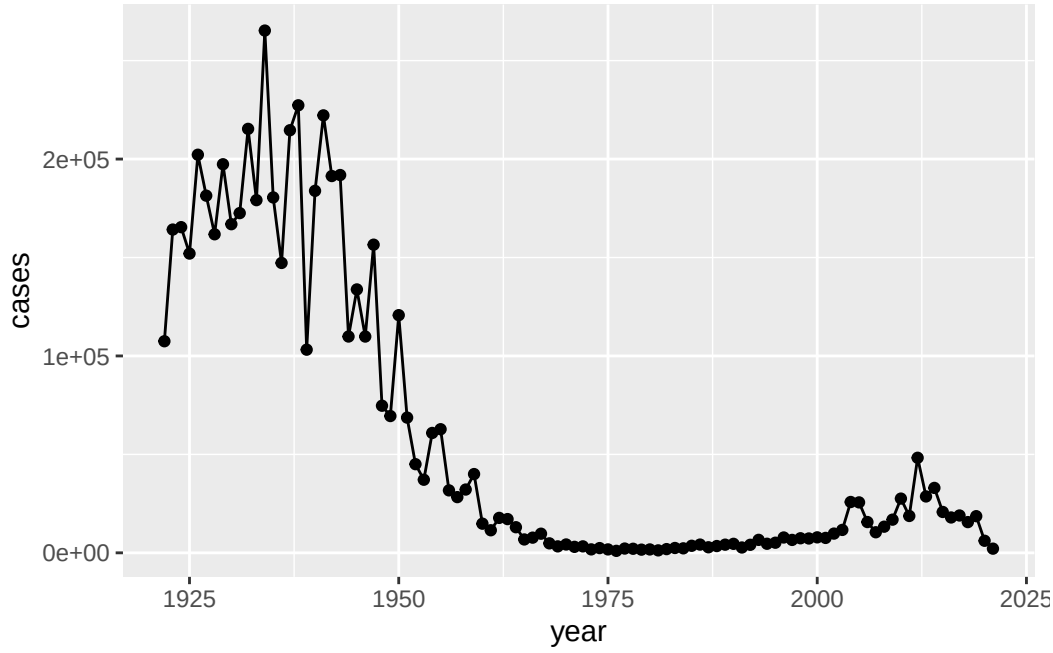
```
Warning: package 'tidyverse' was built under R version 4.3.2
```

```
Warning: package 'forcats' was built under R version 4.3.2
```

```
Warning: package 'lubridate' was built under R version 4.3.2
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
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 dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
ggplot(cdc) +
  aes(year, cases) +
  geom_point() +
  geom_line() +
  labs()
```

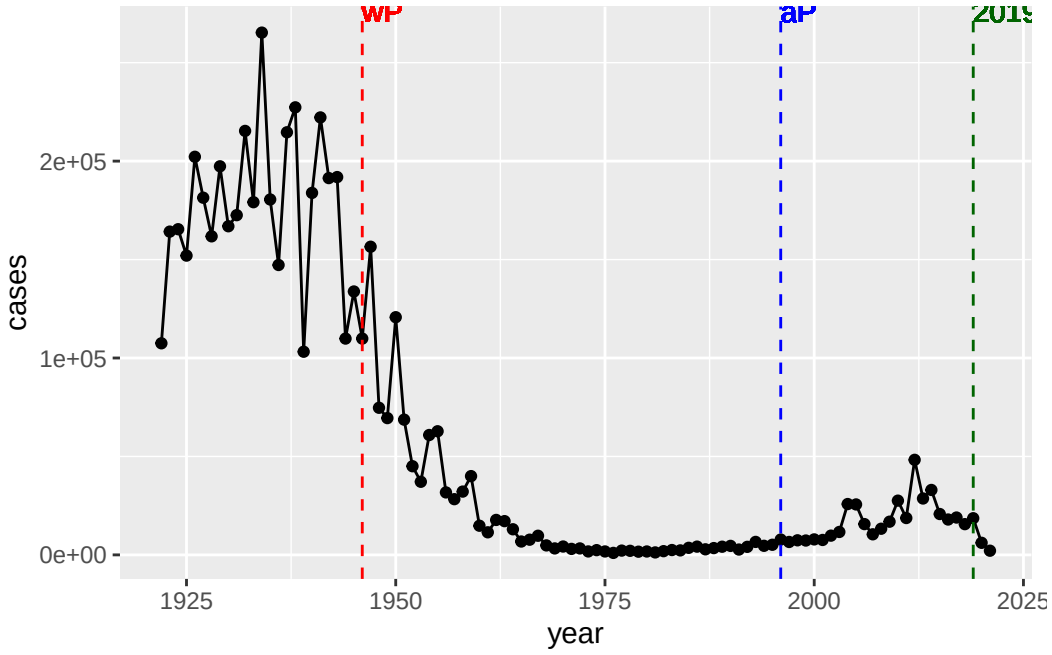


##2. A tale of two vaccines (wP & aP)

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(year, cases) +
  geom_point() +
  geom_line() +
  labs() +
  #v = vertical
  geom_vline(xintercept = 1946, color = "red", linetype = "dashed") +
  geom_vline(xintercept = 1996, color = "blue", linetype = "dashed") +
  geom_vline(xintercept = 2019, color = "darkgreen", linetype = "dashed") +
  geom_text(aes(x = 1946, y = max(cdc$cases), label = "wP"), vjust = -0.5, hjust = 0, colo
  geom_text(aes(x = 1996, y = max(cdc$cases), label = "aP"), vjust = -0.5, hjust = 0, colo
  geom_text(aes(x = 2019, y = max(cdc$cases), label = "2019"), vjust = -0.5, hjust = 0, co
```

```
Warning: Use of `cdc$cases` is discouraged.
i Use `cases` instead.
Use of `cdc$cases` is discouraged.
i Use `cases` instead.
Use of `cdc$cases` is discouraged.
i Use `cases` instead.
```



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

The case number dropped in 1946 but increased in 1947. Then the case number dropped a lot in 1948 and 1949 but increased a little bit in 1950. The overall trend of the cases after aP (1946) dropped, yet it fluctuated a lot. It may be due to several reasons such as the coverage of the vaccination and the improved sensitivity of the detection.

```
library(dplyr)
```

```
cdc %>% filter(year >= 1945 & year <= 1950)
```

```
  year  cases
1 1945 133792
2 1946 109860
3 1947 156517
4 1948  74715
5 1949  69479
6 1950 120718
```

##CMI-PB project

The CMI-PB project collects and makes available data on the immune response to Pertussis booster vaccination.

We will access the data via the API (application programming interface). We will use the **jsonite** package to access the data using the `read_json()` function.

```
library(jsonlite)
```

```
Warning: package 'jsonlite' was built under R version 4.3.2
```

```
Attaching package: 'jsonlite'

The following object is masked from 'package:purrr':

    flatten
```

```
subject <- read_json("http://cmi-pb.org/api/subject", simplifyVector = TRUE)
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
```

```
nrow(subject)
```

```
[1] 118
```

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
```

Q. Make a histogram of the subject age distribution and facet by infancy_vac.

##Side-Note: Working with dates

```
library(lubridate)
today() - mdy("02-11-1996")
```

```
Time difference of 10164 days
```

```
today() - ymd("1996-02-11")
```

```
Time difference of 10164 days
```

```
time_length(today() - ymd("1996-02-11"), "year")
```

```
[1] 27.82752
```

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?

```
subject$age <- today() - ymd(subject$year_of_birth)
library(dplyr)
wP <- subject %>% filter(infancy_vac == "wP")
aP <- subject %>% filter(infancy_vac == "aP")
```

```
mean(time_length(wP$age, "years"))
```

```
[1] 36.33525
```

```
mean(time_length(aP$age, "years"))
```

```
[1] 26.03851
```

```
aP_age <- time_length(aP$age, "years")
wP_age <- time_length(wP$age, "years")
t.test(aP_age, wP_age)
```

```
    Welch Two Sample t-test

data:  aP_age and wP_age
t = -12.436, df = 65.411, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.950080  -8.643385
```

```
sample estimates:
mean of x mean of y
 26.03851  36.33525
```

(i) 36.33525
(ii) 26.03851
(iii) Yes. p value < 2.2e-16

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

```
subject$age <- ymd(subject$date_of_boost) - ymd(subject$year_of_birth)
subject$age_year <- time_length(subject$age, "years")
head(subject$age_year)
```

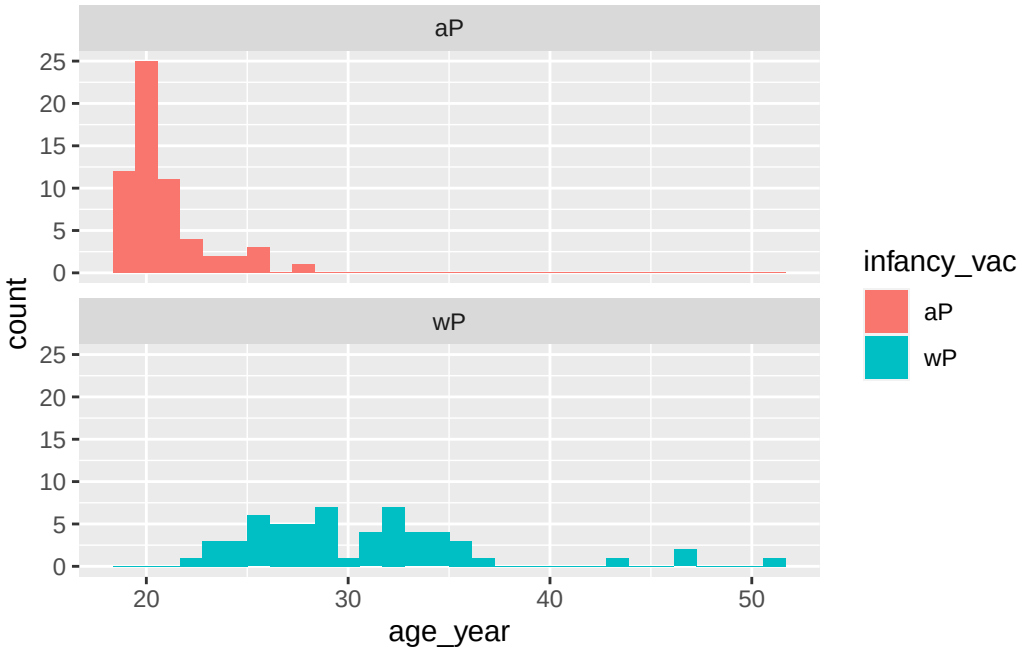
```
[1] 30.69678 51.07461 33.77413 28.65982 25.65914 28.77481
```

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

Yes.

```
ggplot(subject) +
  aes(age_year,
      fill = infancy_vac) +
  facet_wrap(vars(infancy_vac), ncol = 1) +
  geom_histogram()
```

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




```
table(subject$dataset)
```

```
2020_dataset 2021_dataset 2022_dataset
          60           36           22
```

##Joining multiple tables

```
specimen <- read_json("http://cmi-pb.org/api/specimen", simplifyVector = TRUE)
head(specimen)
```

```
  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
1                             0         Blood     1
```

```
2                            1         Blood     2
3                            3         Blood     3
4                            7         Blood     4
5                           14         Blood     5
6                           30         Blood     6
```

```
titer <- read_json("http://cmi-pb.org/api/v4/plasma_ab_titer", simplifyVector = TRUE)
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
```

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:

I want to merge (join) the specimen and subject table together.

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

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

```
dim(meta)
```

```
[1] 939  15
```

```
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_year
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
```

Q10. Now using the same procedure join meta with titer data so we can further analyze this data in terms of time of visit aP/wP, male/female etc.

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

```
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?

The most recent dataset is 2022_dataset. It only contains 2205 rows, which is the fewest.

```
table(abdata$dataset)
```

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

##4. Examine IgG Ab titer levels

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

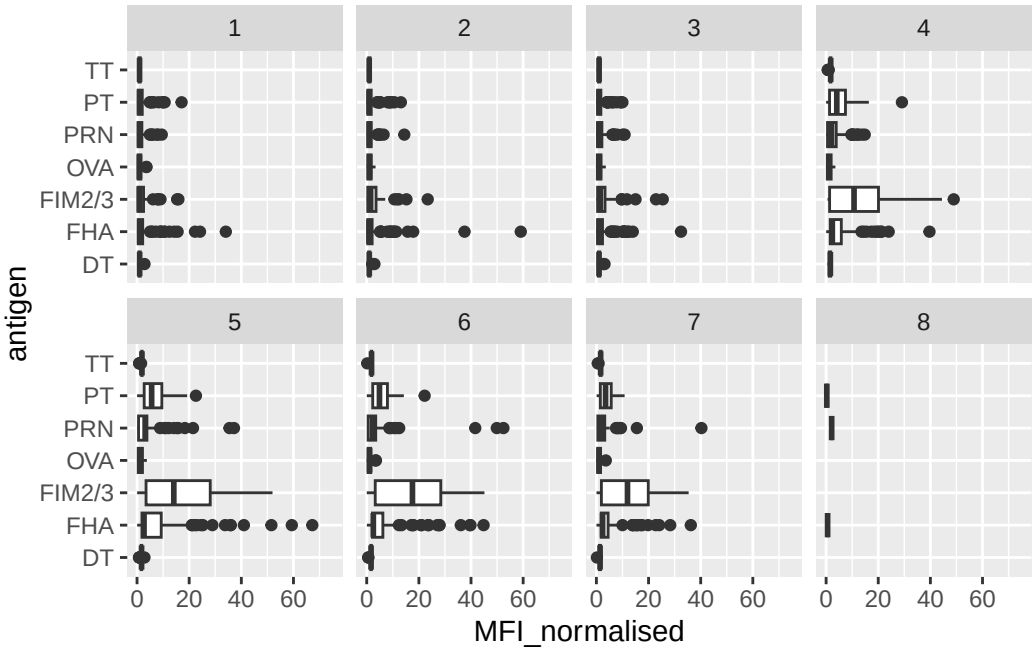
```
  specimen_id subject_id actual_day_relative_to_boost
1           1          1                           -3
2           1          1                           -3
3           1          1                           -3
4           2          1                            1
5           2          1                            1
6           2          1                            1
  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                             1         Blood     2          wP         Female
5                             1         Blood     2          wP         Female
6                             1         Blood     2          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_year isotype is_antigen_specific antigen        MFI
1 11212 days 30.69678     IgG                TRUE      PT   68.56614
2 11212 days 30.69678     IgG                TRUE     PRN  332.12718
3 11212 days 30.69678     IgG                TRUE     FHA 1887.12263
4 11212 days 30.69678     IgG                TRUE      PT   41.38442
5 11212 days 30.69678     IgG                TRUE     PRN  174.89761
6 11212 days 30.69678     IgG                TRUE     FHA  246.00957
  MFI_normalised  unit lower_limit_of_detection
1       3.736992 IU/ML                0.530000
2       2.602350 IU/ML                6.205949
3      34.050956 IU/ML                4.679535
4       2.255534 IU/ML                0.530000
5       1.370393 IU/ML                6.205949
6       4.438960 IU/ML                4.679535
```

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()`).
```



Q14. What antigens show differences in the level of IgG antibody titers recognizing them over time? Why these and not others?

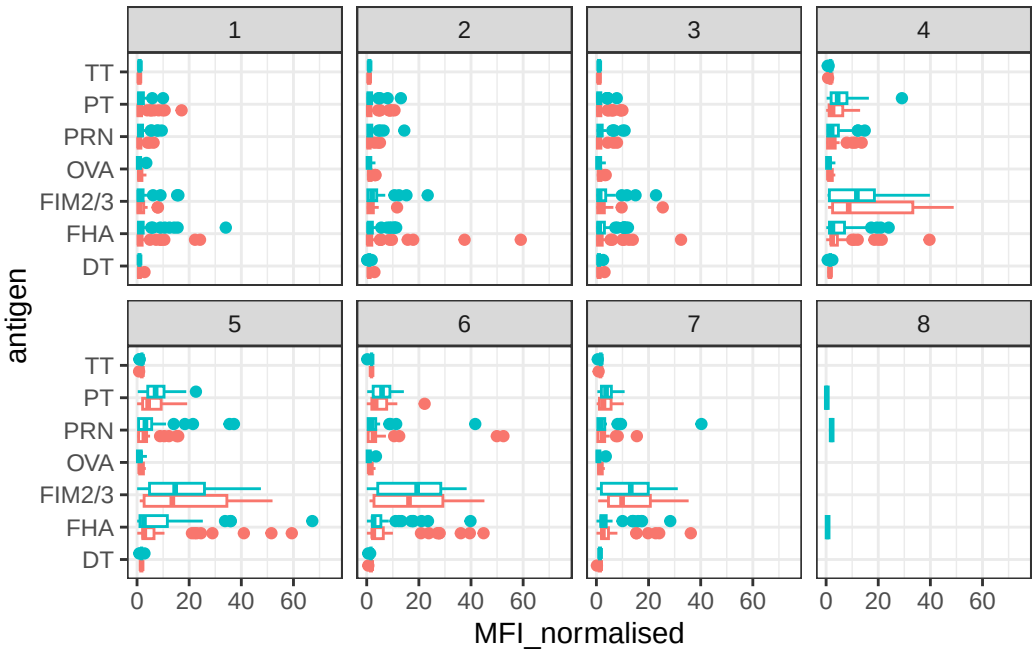
PT (Pertussis toxin complex), PRN (Pertactin autotransporter), FHA (Filamentous hemagglutinin), and FIM2/3 (Mixture of Fim2 and Fim3). Fimbrial protein is the pilus of Bordetella pertussis. All the antigens listed above are contained in the wP and aP vaccine.

The remaining antigens listed below are proteins from other species as controls: TT: Tetanus toxin OVA: Ovalbumin DT: Diphtheria toxin

We can attempt to examine differences between wP and aP here by setting color and/or facet values of the plot to include infancy_vac status (see below). However these plots tend to be rather busy and thus hard to interpret easily.

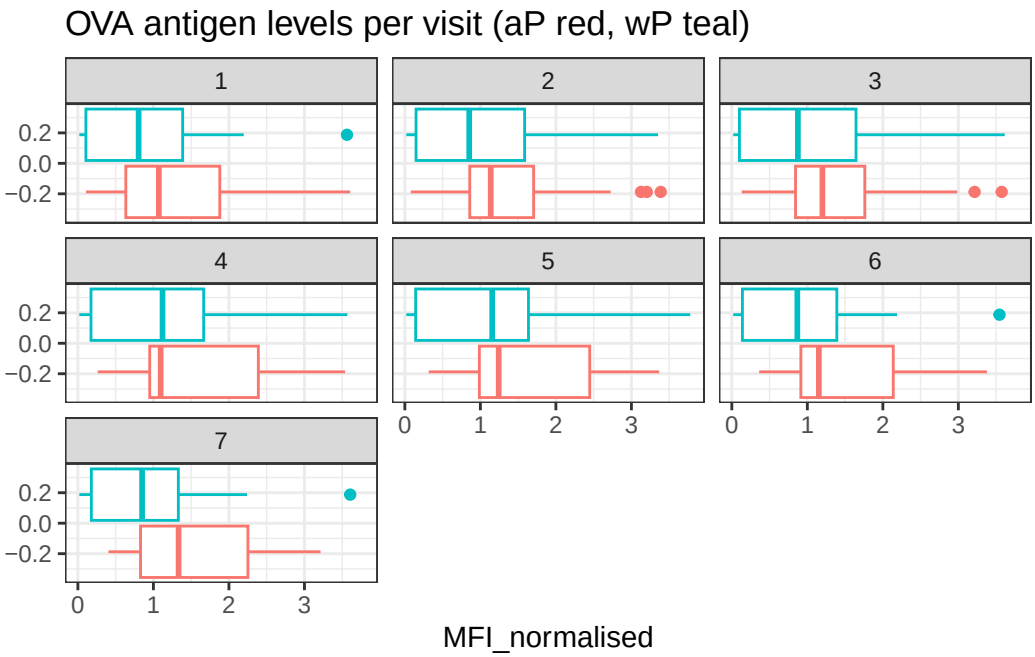
```
ggplot(igg) +
  aes(MFI_normalised, antigen, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit), nrow=2) +
  xlim(0,75) +
  theme_bw()
```

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



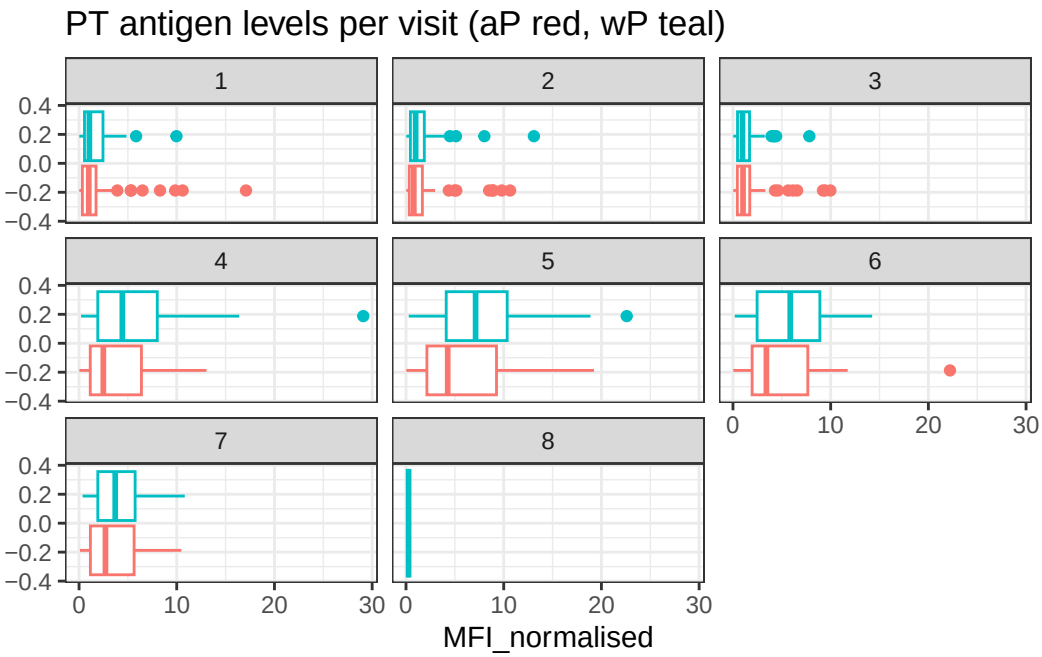
Q15. Filter to pull out only two specific antigens for analysis and create a boxplot for each. You can chose any you like. Below I picked a "control" antigen ("OVA", that is not in our vaccines) and a clear antigen of interest ("PT", Pertussis Toxin, one of the key virulence factors produced by the bacterium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

```
filter(igg, antigen=="OVA") %>%
  ggplot() +
  aes(MFI_normalised, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw() +
  ggtitle("OVA antigen levels per visit (aP red, wP teal)")
```



For PT antigen:

```
filter(igg, antigen=="PT") %>%
  ggplot() +
  aes(MFI_normalised, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw() +
  ggtitle("PT antigen levels per visit (aP red, wP teal)")
```

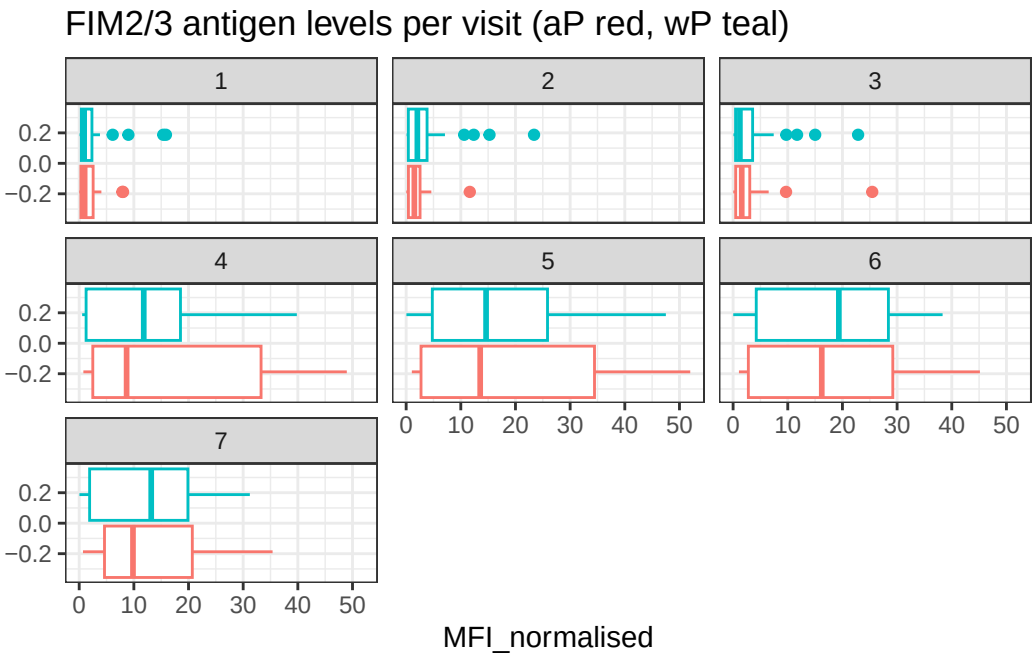


Q16. What do you notice about these two antigens time courses and the PT data in particular?

The levels of anti-PT IgG increased a lot and reached the peak at around visit 5. The levels of anti-OVA IgG didn't change at different time point and the levels were quite low.

For FIM2/3:

```
filter(igg, antigen=="FIM2/3") %>%
  ggplot() +
  aes(MFI_normalised, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw() +
  ggtitle("FIM2/3 antigen levels per visit (aP red, wP teal)")
```



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

I didn't see obvious difference in aP vs, wP responses regarding the concentrations and the time to reach the peak of the IgG levels. (Though the anti-PT IgG levels wP is a little bit higher at visit 5.)

It's a mistake that there are FIM2/3 and Fim2/3. Let's filter out which dataset is it of Fim2/3.

```
oops <- abdata %>% filter(antigen == "Fim2/3")
table(oops$dataset)
```

```
< table of extent 0 >
```

```
table(abdata$dataset)
```

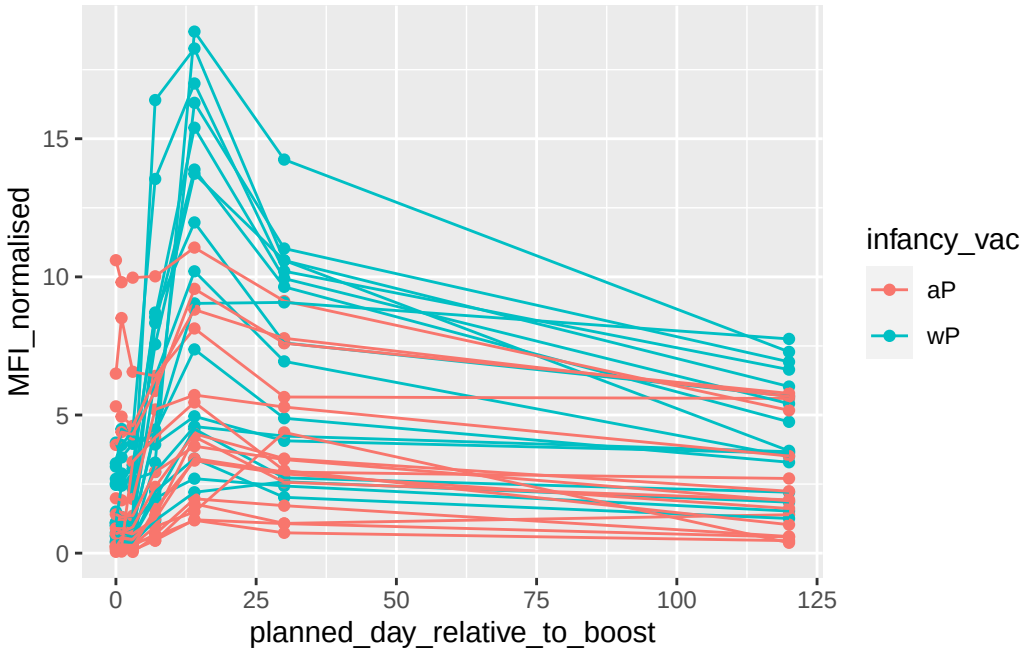
```
2020_dataset 2021_dataset 2022_dataset
       31520         8085         2205
```

Select (or filter) for the 2021 dataset and isotype IgG I want a time course (`planned_day_relative_to_boost`) of IgG levels (`MFI_normalized`) for "PT" antigen.

```
#abdata$planned_day_relative_to_boost
```

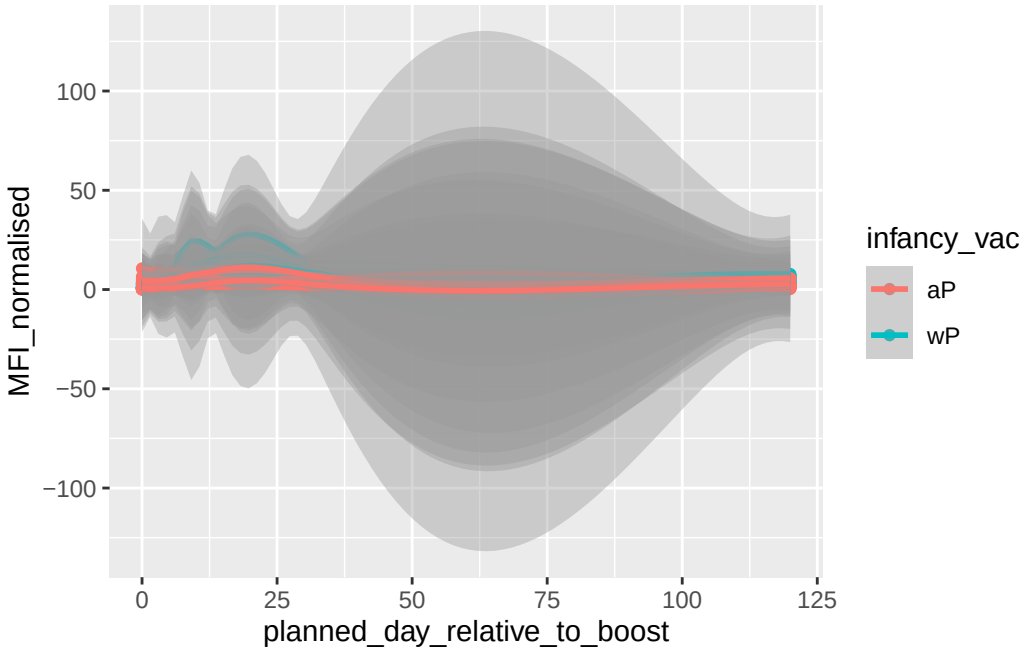
```
abdata_21_IgG_PT <- abdata %>% filter(dataset == "2021_dataset",
                                      isotype == "IgG",
                                      antigen == "PT")
```

```
ggplot(abdata_21_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac,
      group = subject_id) + #Group by subject_id!
  geom_point() +
  geom_line()
```



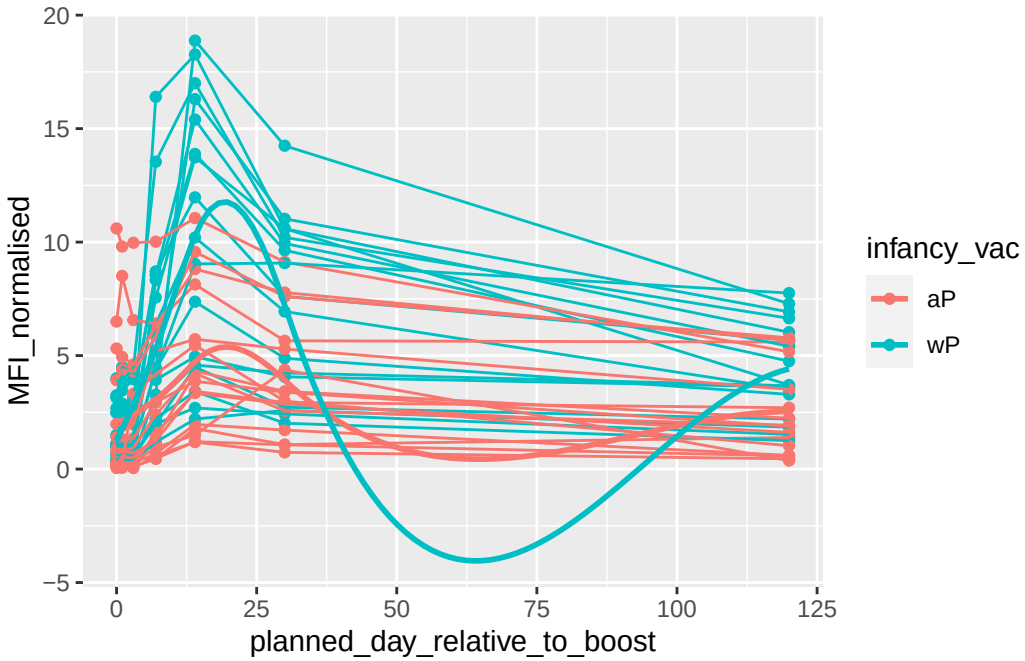

```
ggplot(abdata_21_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac,
      group = subject_id) + #Group by subject_id!
  geom_point() +
  geom_line() +
  geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(abdata_21_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac) +
  geom_point() +
  geom_line(aes(group = subject_id)) +
  geom_smooth(se=FALSE)
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(abdata_21_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac) +
  geom_point() +
  geom_line(aes(group = subject_id), linewidth=0.5, alpha=0.5) +
  geom_smooth(se=FALSE, span=0.4, linewidth=3) +
  geom_vline(xintercept = 0, linetype="dashed") +
  geom_vline(xintercept = 14, linetype="dashed")
```

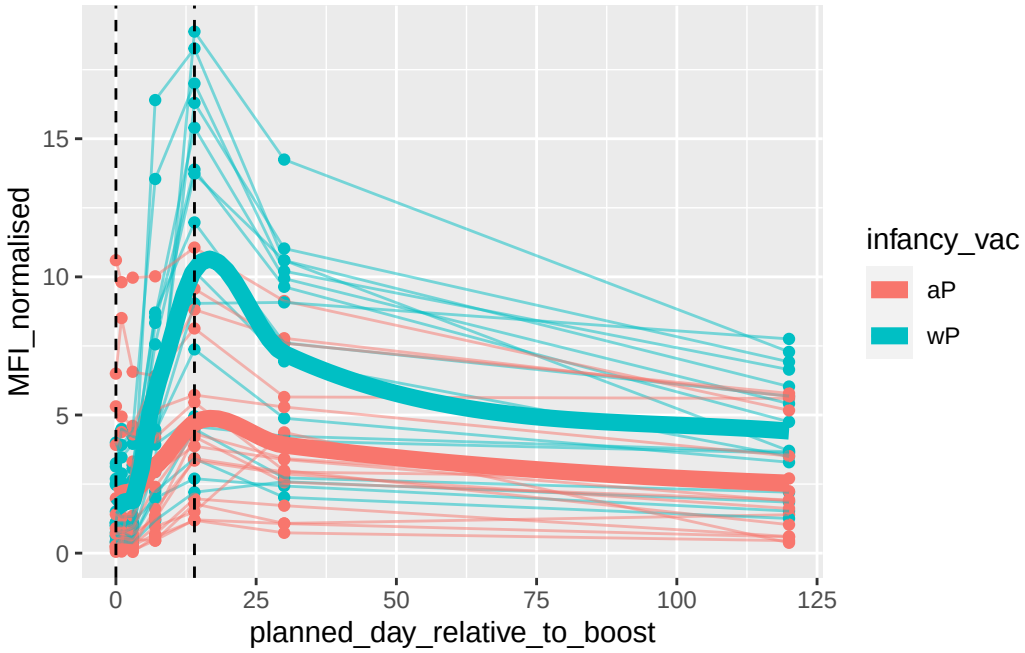
```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -0.6
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 3.6
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1.8382e-16
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 11364

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -0.6

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 3.6

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1.4316e-16

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 11364
```



Let's do the 2022 dataset.

```
abdata_22_IgG_PT <- abdata %>% filter(dataset == "2022_dataset",
                                      isotype == "IgG",
                                      antigen == "PT")
```

```
ggplot(abdata_22_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac) +
  geom_point() +
  geom_line(aes(group = subject_id), linewidth=0.5, alpha=0.5) +
  geom_smooth(se=FALSE, span=0.4, linewidth=3) +
  geom_vline(xintercept = 0, linetype="dashed") +
  geom_vline(xintercept = 14, linetype="dashed")
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -30.15
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 15.15
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0
```

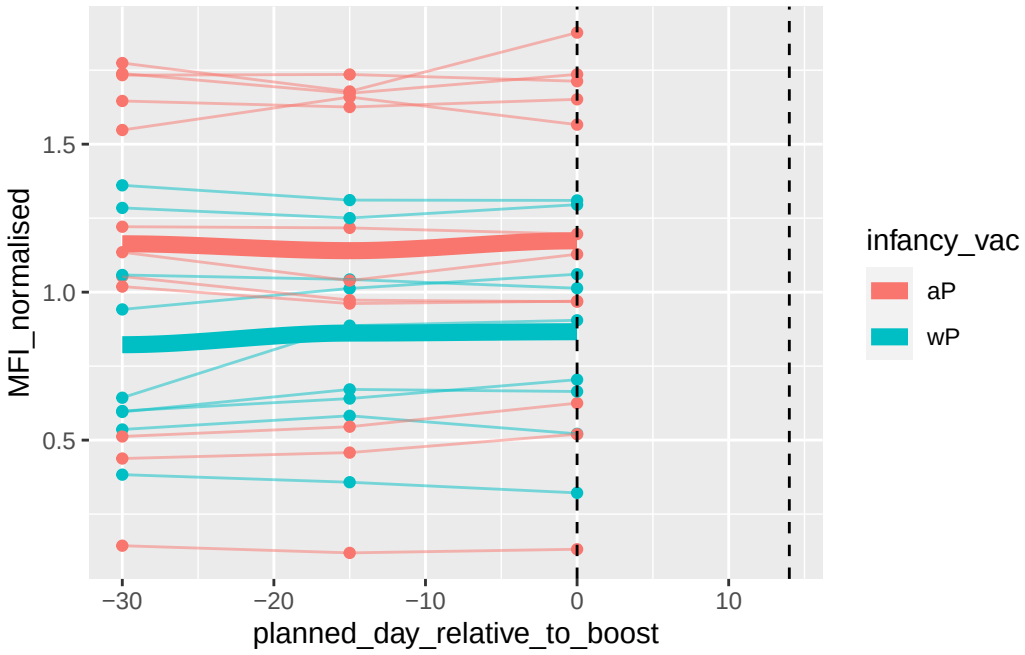
```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 229.52
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -30.15
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 15.15
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 229.52
```



2020-dataset

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

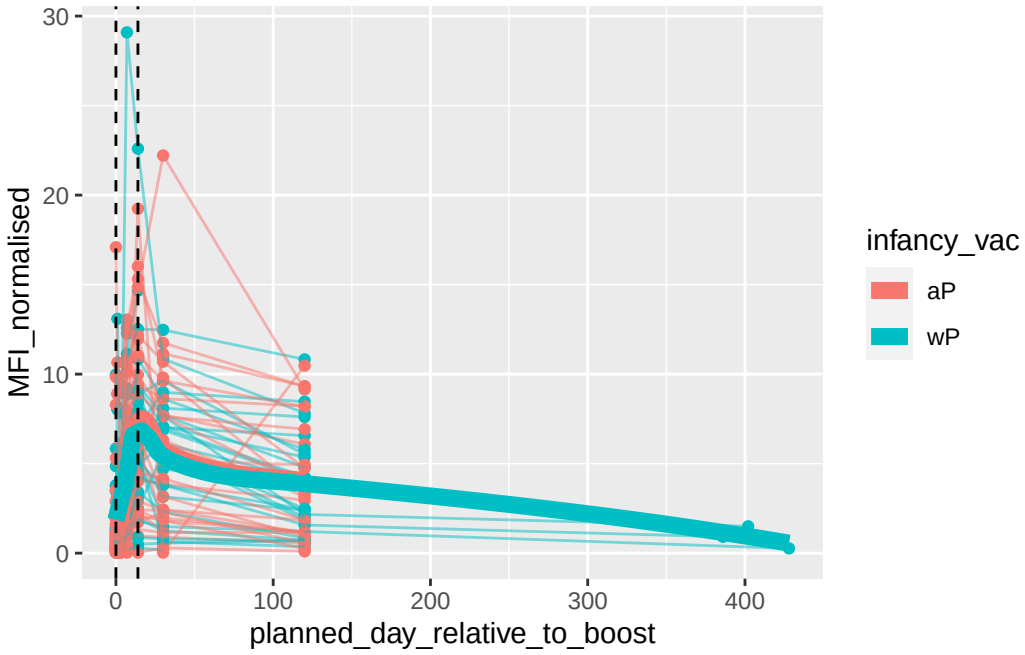
Yes. The peak of the anti-PT IgG MFI_normalised are both around day 14. But the avergage of wP seems to be higher than aP in 2021 but not in 2020.

```
abdata_20_IgG_PT <- abdata %>% filter(dataset == "2020_dataset",
                                      isotype == "IgG",
                                      antigen == "PT")
```

```
ggplot(abdata_20_IgG_PT) +
  aes(planned_day_relative_to_boost, MFI_normalised,
      col = infancy_vac) +
  geom_point() +
  geom_line(aes(group = subject_id), linewidth=0.5, alpha=0.5) +
  geom_smooth(se=FALSE, span=0.4, linewidth=3) +
  geom_vline(xintercept = 0, linetype="dashed") +
  geom_vline(xintercept = 14, linetype="dashed")
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -0.6

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 3.6

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 2.3736e-16

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 11364

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at -2.14

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 5.14

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 4.7012e-16

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 9
```



##5. Obtaining CMI-PB RNASeq data

```
url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_gene_id=eq.ENSG00000211896.
```

```
rna <- read_json(url, simplifyVector = TRUE)
```

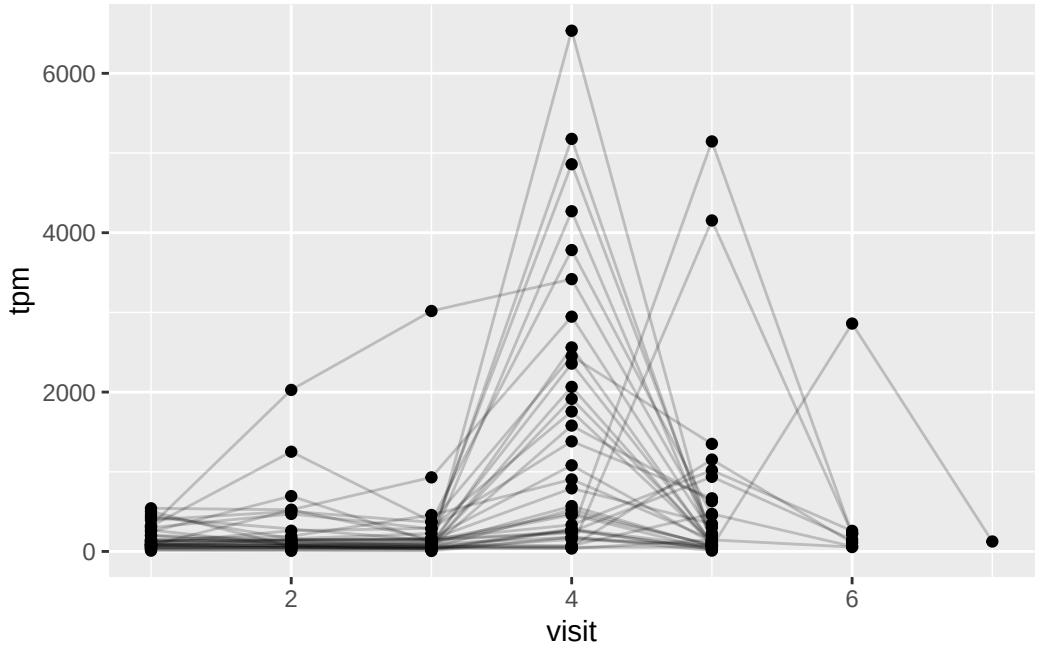
To join the `rna` expresion data with the metadata `meta`:

```
ssrna <- inner_join(rna, meta)
```

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

Q19. Make a plot of the time course of gene expression for IGHG1 gene (i.e. a plot of visit vs. tpm).

```
ggplot(ssrna) +
  aes(visit, tpm, group=subject_id) +
  geom_point() +
  geom_line(alpha=0.2)
```



Q20.: What do you notice about the expression of this gene (i.e. when is it at it's maximum level)?

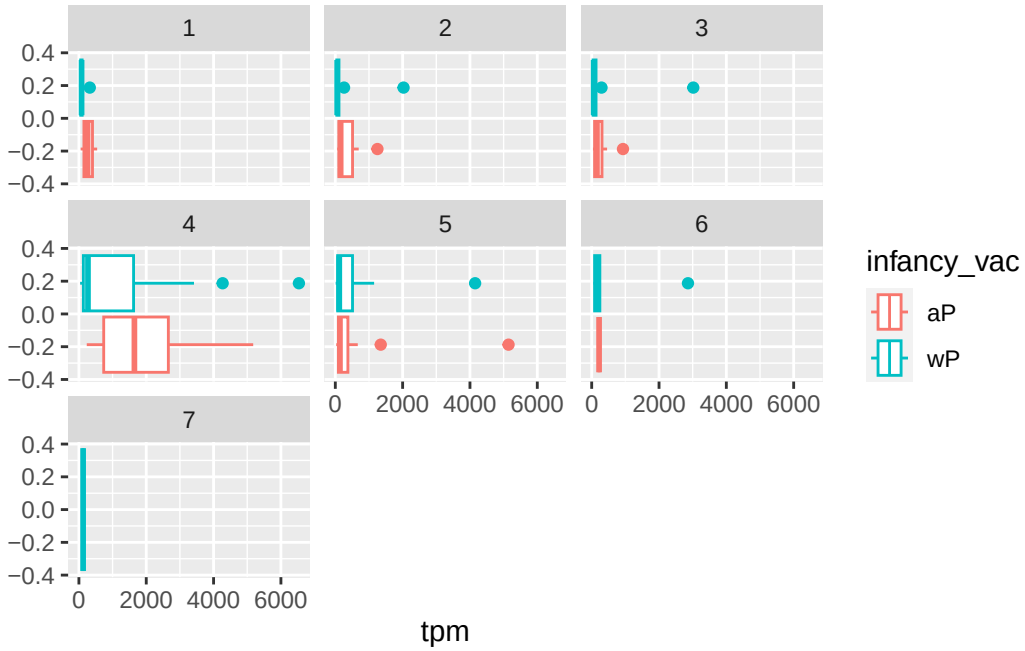
IGHG1 is the IgG heavy chains. The time to reach the maximum level of IGHG1 is at visit 4.

> Q21. Does this pattern in time match the trend of antibody titer data? If not, why not?

The expression of IGHG1 reached the peak at visit 4, which is earlier than the trend of antibody titer data (around visit 5 to 6). It is possibly because the gene expression is earlier, and the antigen-specific plasma cells would later significantly expand. Also, the memory of B cell can remain the serum antibody levels for days to several weeks and even lifelong.

To compare aP and wP on IGHG1:

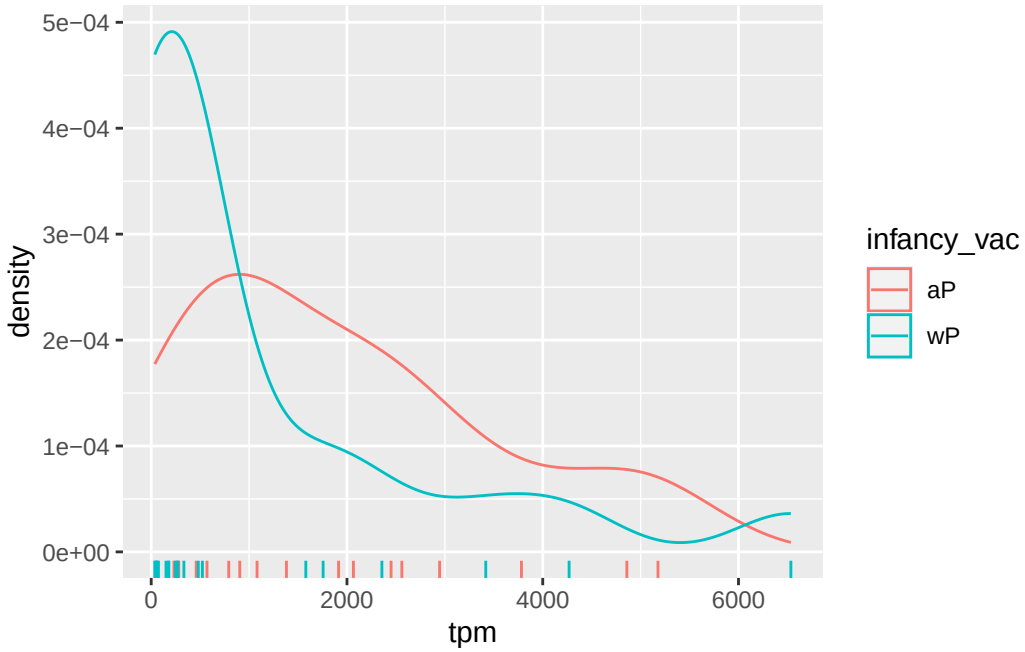
```
ggplot(ssrna) +
  aes(tpm, col=infancy_vac) +
  geom_boxplot() +
  facet_wrap(vars(visit))
```



To filter visit 4:

```
ssrna %>%
  filter(visit==4) %>%
```

```
ggplot() +
  aes(tpm, col=infancy_vac) + geom_density() +
  geom_rug()
```



There is no obvious difference between aP and wP.