Class 19

Tomi Lane Timmins

```
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),
  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)
)
```

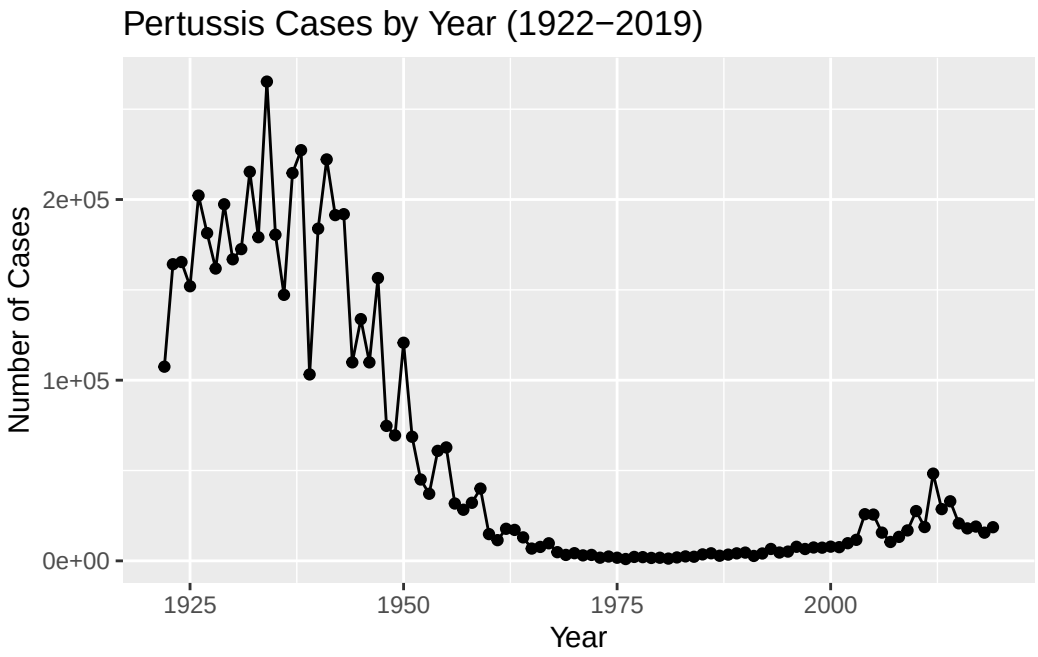
```
head(cdc)
```

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

1 Investigating pertussis cases by year

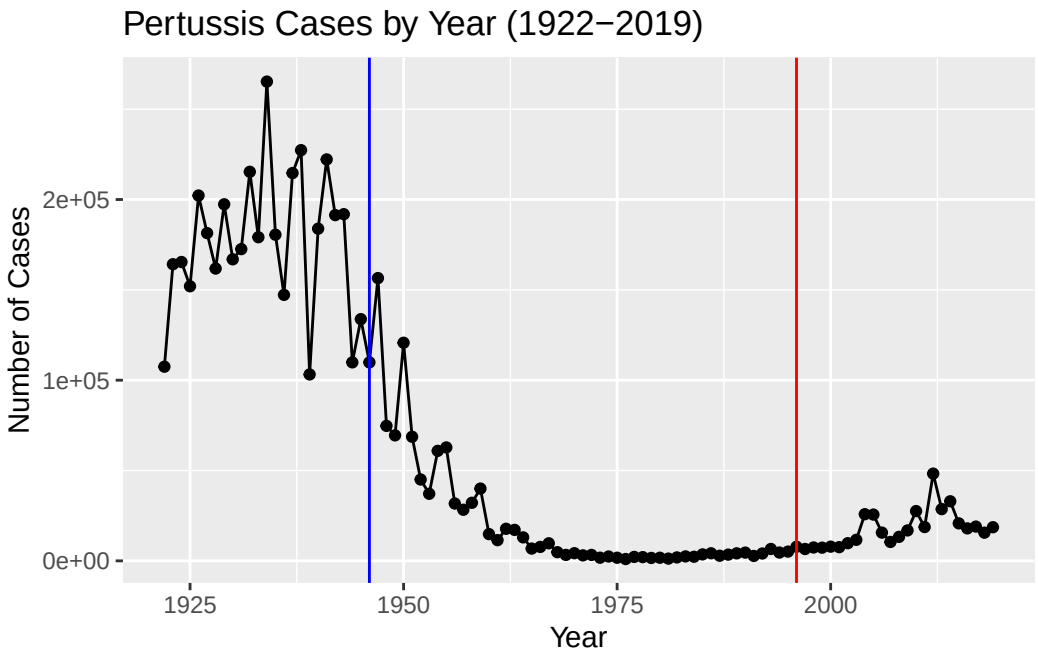
Q1. With the help of the R "addin" package datapasta assign the CDC pertussis case number data to a data frame called cdc and use ggplot to make a plot of cases numbers over time.

```
library(ggplot2)

baseplot <- ggplot(cdc) +
  aes(Year, Cases) +
  geom_point() +
  geom_line() +
  labs(title = "Pertussis Cases by Year (1922-2019)", x = "Year", y = "Number of Cases")

baseplot
```



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?

```
baseplot + geom_vline(xintercept = 1946, col = "blue") +
  geom_vline(xintercept = 1996, col = "red")
```



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

The number of observed cases increased after the introduction of the new vaccine. It could be due to bacterial resistance, decreased effectiveness, or less people obtaining vaccines.

Additional points for discussion: How are vaccines currently approved?

Typically we examine 'Correlates of protection' and need to conclude a study in finite time. For the aP vaccine there is an induction of pertussis toxin (PT) antibody titers in infants at equivalent levels to those induced by the wP vaccine. The aP vaccines also had less side effects (reduction of sore arms, fever and pain).

It is impossible to discover a effect 10 years post vaccination in the current trial system.

Some things make a difference such as time of day one is vaccinated - morning gives more immunity than afternoon for some reason.

It is unclear what differentiates people that have been primed with aP vs. wP long term.

CMI-PB project is an attempt to make data on this question open and examinable by all.

Exploring CMI-PB data

The CMI-PB project collects data on aP and wP individuals and their immune response to infection and/or booster shots.

CMI-PB project provides scientific community with this info. It tracks and makes available the long-term humoral and cellular immune response data for a large number of individuals who received these vaccinations (DTwP or DTaP followed by Tdap boosters)

CMI-PB data is in JSON format. To read these, we will use 'read.json()' function from the 'jsonlite' package.

```
#Allows us to read, write and process JSON data

library(jsonlite)
```

We pasted the url from the "subject" table on CMI-PB

```
subject <- read_json("https://www.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
```

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

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

```
aP wP
47 49
```

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

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

```
Female   Male
    66     30
```

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                                     18    9
  Black or African American                  2    0
  More Than One Race                         8    2
  Native Hawaiian or Other Pacific Islander  1    1
  Unknown or Not Reported                   10    4
  White                                     27   13
```

Side Note: Working with Dates

```
library(lubridate)
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

What is today's date?

```
today()
```

```
[1] "2023-03-14"
```

How many days have passed since the year 2000?

```
today() - ymd("2000-01-01")
```

```
Time difference of 8473 days
```

What is this in years?

```
time_length( today() - ymd("2000-01-01"), "years")
```

```
[1] 23.19781
```

```
age_days <- today() - ymd(subject$year_of_birth)

age_years <- time_length(age_days, "years")

age_years
```

```
 [1] 37.19644 55.19781 40.19713 35.19781 32.19713 35.19781 42.19576 38.19576
 [9] 27.19781 41.19644 37.19644 41.19644 26.19576 30.19576 34.19576 36.19713
[17] 43.19781 26.19576 29.19644 36.19713 30.19576 28.19713 30.19576 33.19644
[25] 47.19781 51.19781 51.19781 33.19644 25.19644 25.19644 32.19713 28.19713
[33] 28.19713 25.19644 25.19644 35.19781 30.19576 36.19713 31.19781 30.19576
[41] 25.19644 24.19713 26.19576 23.19781 25.19644 23.19781 23.19781 26.19576
[49] 24.19713 25.19644 23.19781 27.19781 24.19713 25.19644 23.19781 42.19576
[57] 40.19713 38.19576 32.19713 31.19781 35.19781 40.19713 26.19576 41.19644
[65] 26.19576 35.19781 34.19576 26.19576 33.19644 40.19713 32.19713 26.19576
[73] 25.19644 26.19576 38.19576 29.19644 38.19576 26.19576 25.19644 25.19644
[81] 26.19576 25.19644 27.19781 25.19644 26.19576 26.19576 26.19576 25.19644
[89] 25.19644 26.19576 26.19576 26.19576 27.19781 26.19576 26.19576 26.19576
```

```
subject$age <- 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
1    1986-01-01    2016-09-12 2020_dataset 37.19644
2    1968-01-01    2019-01-28 2020_dataset 55.19781
3    1983-01-01    2016-10-10 2020_dataset 40.19713
4    1988-01-01    2016-08-29 2020_dataset 35.19781
5    1991-01-01    2016-08-29 2020_dataset 32.19713
6    1988-01-01    2016-10-10 2020_dataset 35.19781
```

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?

```
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
```

```
mean(filter(subject, infancy_vac == "aP")$age)
```

```
[1] 25.5156
```

```
  mean(filter(subject, infancy_vac == "wP")$age)
```

```
[1] 36.36006
```

T- test

```
  ap.age <- filter(subject, infancy_vac == "aP")$age
  wp.age <- filter(subject, infancy_vac == "wP")$age

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

```
    Welch Two Sample t-test

data:  ap.age and wp.age
t = -12.092, df = 51.082, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.644857  -9.044045
sample estimates:
mean of x mean of y
 25.51560  36.36006
```

Based on T-test, these are significantly different populations in terms of age.

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

```
  age_at_boost <- time_length( ymd(subject$date_of_boost) - ymd(subject$year_of_birth), "yea


  age_at_boost
```

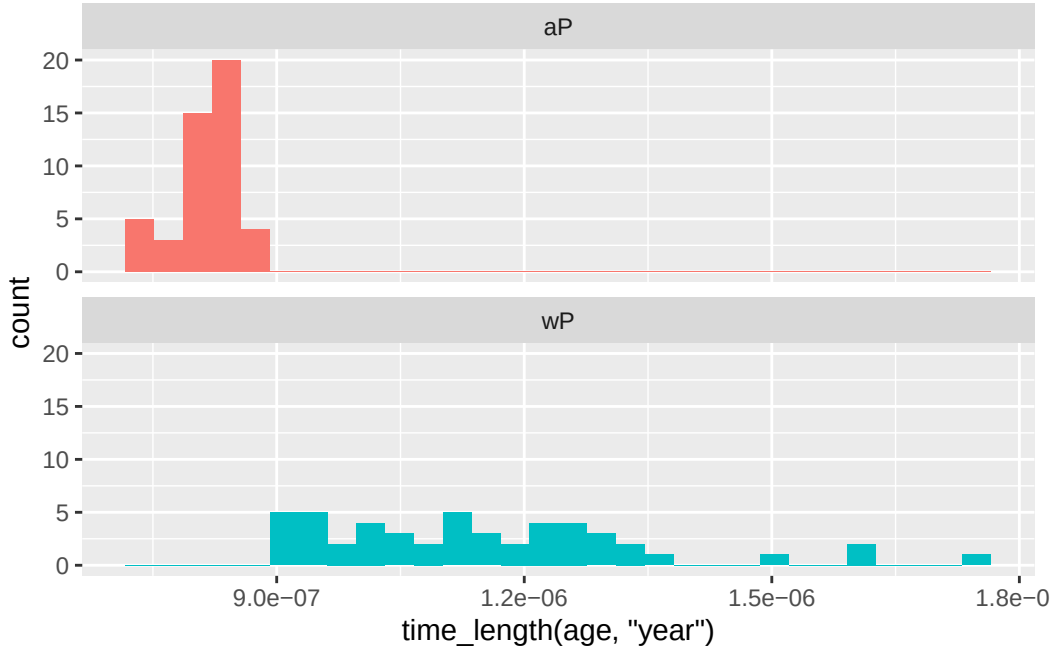
```
 [1] 30.69678 51.07461 33.77413 28.65982 25.65914 28.77481 35.84942 34.14921
 [9] 20.56400 34.56263 30.65845 34.56263 19.56194 23.61944 27.61944 29.56331
[17] 36.69815 19.65777 22.73511 32.26557 25.90007 23.90144 25.90007 28.91992
[25] 42.92129 47.07461 47.07461 29.07324 21.07324 21.07324 28.15058 24.15058
[33] 24.15058 21.14990 21.14990 31.20876 26.20671 32.20808 27.20876 26.20671
[41] 21.20739 20.26557 22.26420 19.32375 21.32238 19.32375 19.32375 22.41752
[49] 20.41889 21.41821 19.47707 23.47707 20.47639 21.47570 19.47707 35.65777
```

```
[57] 33.65914 31.65777 25.73580 24.70089 28.70089 33.73580 19.73443 34.73511
[65] 19.73443 28.73648 27.73443 19.81109 26.77344 33.81246 25.77413 19.81109
[73] 18.85010 19.81109 31.81109 22.81177 31.84942 19.84942 18.85010 18.85010
[81] 19.90691 18.85010 20.90897 19.04449 20.04381 19.90691 19.90691 19.00616
[89] 19.00616 20.04381 20.04381 20.07940 21.08145 20.07940 20.07940 20.07940
```

Q9. With the help of a faceted boxplot (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)
```

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



Joining Multiple Tables

Read the specimen and ab_titertables into R and store the data as ‘specimen’and ’titer’:

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

titer <- read_json("https://www.cmi-pb.org/api/ab_titer", simplifyVector = TRUE)
```

```
head(specimen)
```

```
  specimen_id subject_id actual_day_relative_to_boost
1           1          1                           -3
2           2          1                          736
3           3          1                            1
4           4          1                            3
5           5          1                            7
6           6          1                           11
  planned_day_relative_to_boost specimen_type visit
1                             0         Blood     1
2                           736         Blood    10
3                             1         Blood     2
4                             3         Blood     3
5                             7         Blood     4
6                            14         Blood     5
```

The subject_id column corresponds to the subject table information.

```
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:

Note: 'inner_join()' merges two data sets, but only keeps observations in x that have a matching key in y (if a row is missing in either, it will be dropped). You may lose data with this method. 'full_join()' merges the datasets without dropping data.

```
dim(specimen)
```

```
[1] 729   6
```

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

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

```
dim(meta)
```

```
[1] 729  14
```

```
head(meta)
```

```
  specimen_id subject_id actual_day_relative_to_boost
1           1          1                           -3
2           2          1                          736
3           3          1                            1
4           4          1                            3
5           5          1                            7
6           6          1                           11
  planned_day_relative_to_boost specimen_type visit infancy_vac biological_sex
1                             0         Blood     1          wP         Female
2                           736         Blood    10          wP         Female
3                             1         Blood     2          wP         Female
4                             3         Blood     3          wP         Female
5                             7         Blood     4          wP         Female
6                            14         Blood     5          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
1 37.19644
2 37.19644
3 37.19644
4 37.19644
5 37.19644
6 37.19644
```

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(titer, meta)
```

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

```
dim(abdata)
```

```
[1] 32675    21
```

```
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
1 37.19644
2 37.19644
3 37.19644
4 37.19644
5 37.19644
6 37.19644
```

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 1413 6141 6141 6141 6141
```

Q12. What do you notice about the number of visit 8 specimens compared to other visits?

```
table(abdata$visit)
```

```
   1    2    3    4    5    6    7    8
5795 4640 4640 4640 4640 4320 3920   80
```

There is a much lower sample size for visit 8 specimens in comparison to the other visits. Data is missing for many of the individuals and so it would be best to exclude visit 8.

Examine IgG1 Ab titer levels

We want to examine abdata for IgG1 isotype. We use 'filter()' to isolate the IgG1 isotype and exlude the visit 8 entries.

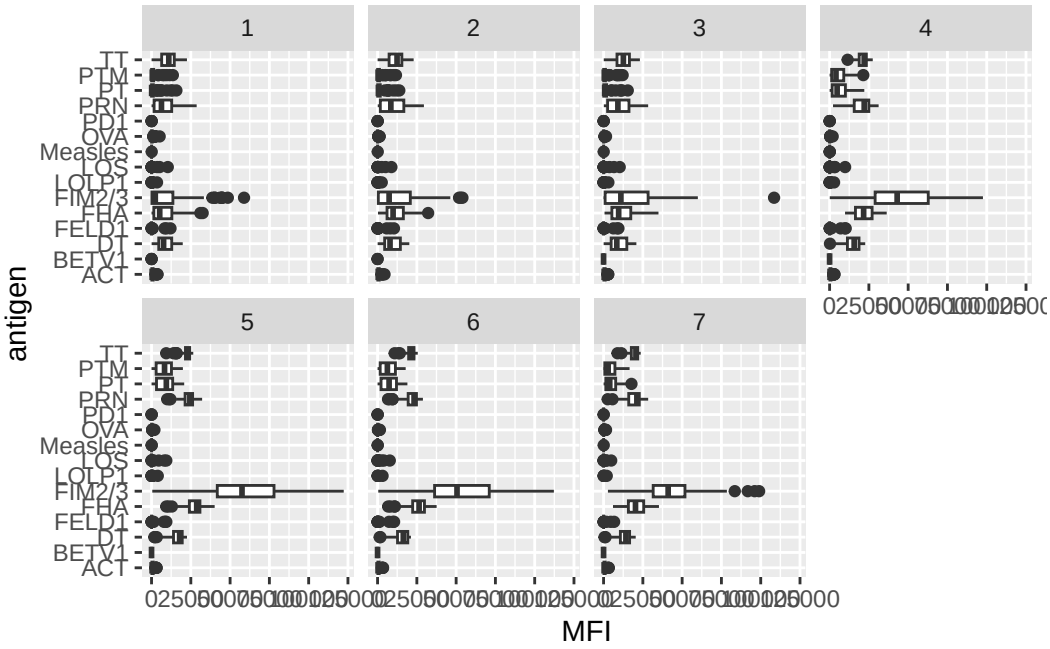
```
ig1 <- abdata %>%
  filter(isotype == "IgG1", visit!=8)

head(ig1)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1    IgG1                TRUE     ACT 274.355068      0.6928058
2           1    IgG1                TRUE     LOS  10.974026      2.1645083
3           1    IgG1                TRUE   FELD1   1.448796      0.8080941
4           1    IgG1                TRUE   BETV1   0.100000      1.0000000
5           1    IgG1                TRUE   LOLP1   0.100000      1.0000000
6           1    IgG1                TRUE Measles  36.277417      1.6638332
   unit lower_limit_of_detection subject_id actual_day_relative_to_boost
1 IU/ML                3.848750          1                           -3
2 IU/ML                4.357917          1                           -3
3 IU/ML                2.699944          1                           -3
4 IU/ML                1.734784          1                           -3
5 IU/ML                2.550606          1                           -3
6 IU/ML                4.438966          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
1 37.19644
2 37.19644
3 37.19644
```

```
4 37.19644
5 37.19644
6 37.19644
```

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

```
ggplot(ig1) +
  aes(MFI, antigen) +
  geom_boxplot() +
  facet_wrap(vars(visit), nrow = 2)
```



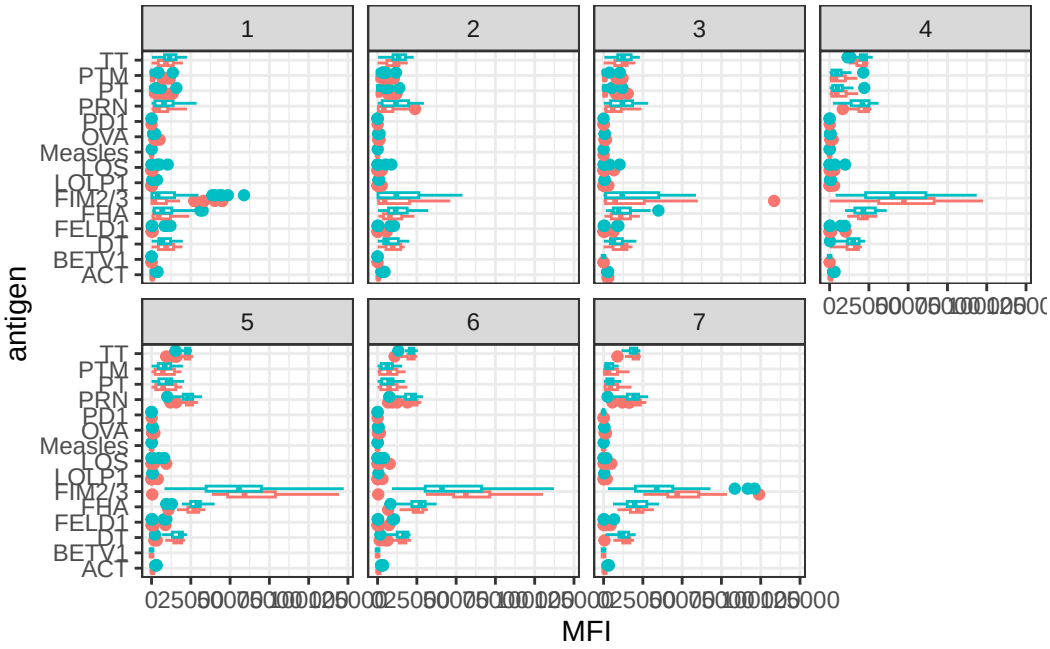
Q14. What antigens show differences in the level of IgG1 antibody titers recognizing them over time? Why these and not others?

FIM2/3, FHA and PT show differences.

FIM 2/3 show a much higher difference in the level of IgG1 antibody titers recognizing them over time. Looking at Uniprot, we can see that fimbrial proteins are pili on the surface of pertussis. They are involved in cell adhesion. The vaccine is likely targeting these proteins (FIM2/3) on the cell and that is why we see increased levels of antibody titers recognizing them.

Now we can examine differences between wP and aP.

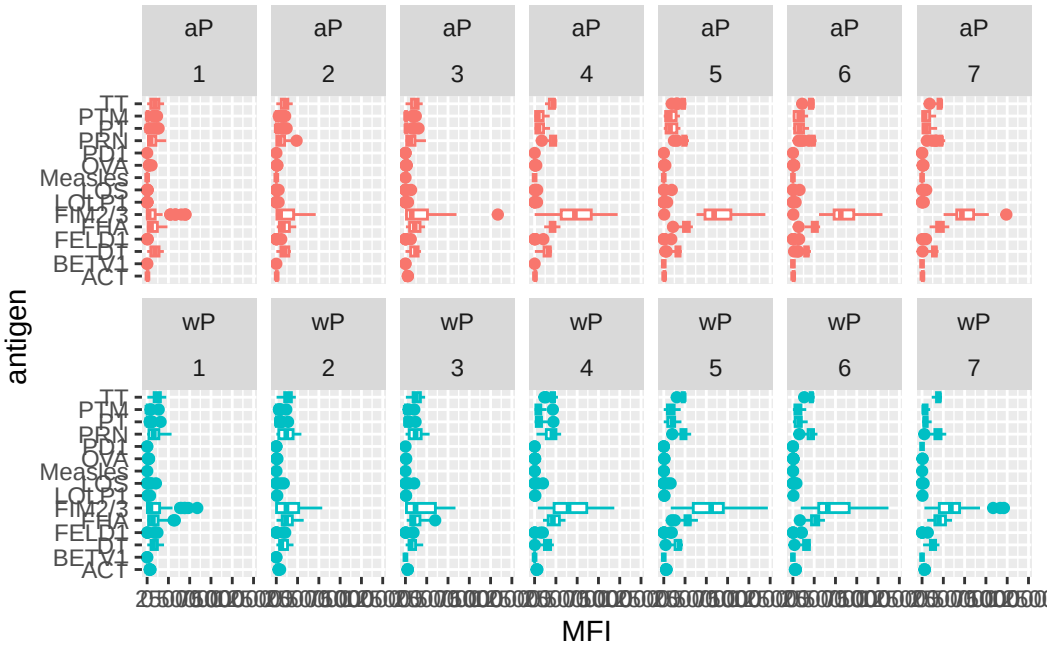
```
ggplot(ig1) +
  aes(MFI, antigen, col = infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit), nrow = 2) +
  theme_bw()
```



We see an increase in DT, (diptheria toxin) - the vaccine is targeting this - as well as FHA and FIM 2/3.

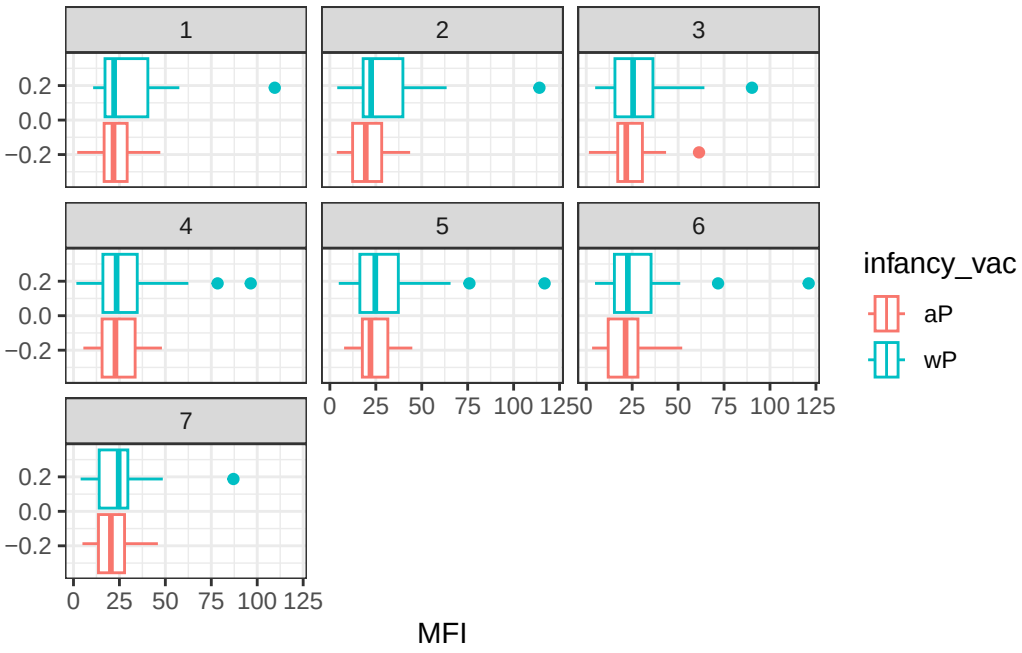
Again by faceting with infancy_vac:

```
ggplot(ig1) +
  aes(MFI, antigen, col = infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(infancy_vac, visit), nrow = 2)
```



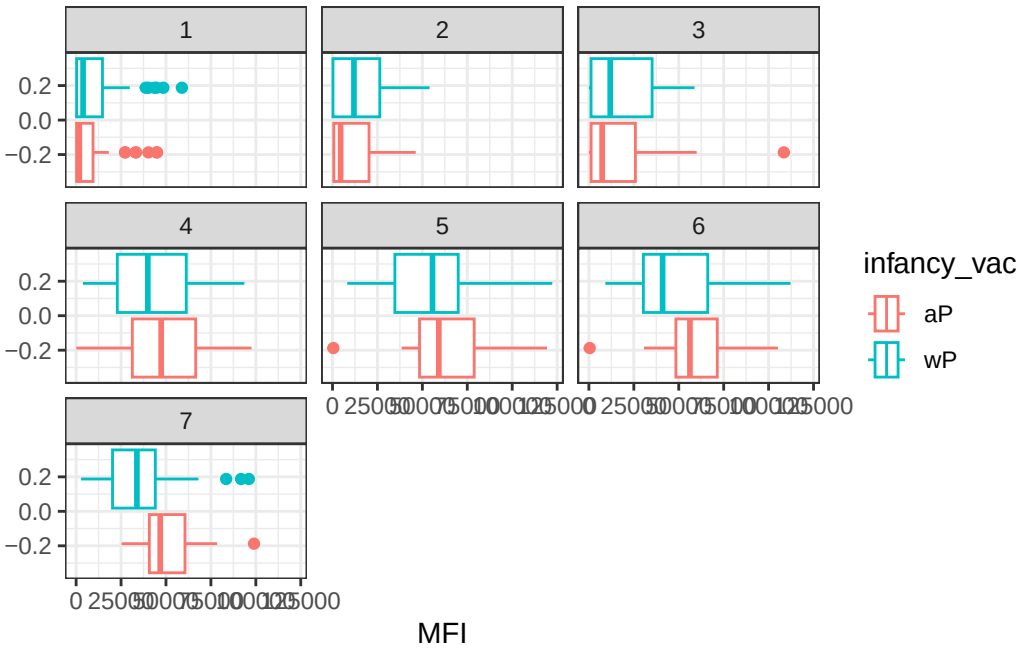
Q15. Filter to pull out only two specific antigens for analysis and create a boxplot for each. You can chose any you like. Below I picked a "control" antigen ("Measles", that is not in our vaccines) and a clear antigen of interest ("FIM2/3", extra-cellular fimbriae proteins from B. pertussis that participate in substrate attachment).

```
filter(ig1, antigen == "Measles") %>%
  ggplot() +
  aes(MFI, col = infancy_vac) +
  geom_boxplot(show.legend = TRUE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



For FIM 2/3:

```
filter(ig1, antigen == "FIM2/3") %>%
  ggplot() +
  aes(MFI, col = infancy_vac) +
  geom_boxplot(show.legend = TRUE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



Q16. What do you notice about these two antigens time course and the FIM2/3 data in particular?

Measles was the control, so there are consistent antibody levels after each visit. FIM 2/3 has an increase in antibody levels up to visit 6, and then decreases at visit 7.

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

While aP and wP responses stay around the same levels, wP has higher average levels of antibodies for the first 3 visits and then aP shows higher average levels of antibodies from visits 4 to 7.

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)
```

This link is for a key gene involved in expressing any IgG1 gene, in particular the IGHG1 gene.

Use '_join()' for RNA and meta (which is specimen and subject)

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

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

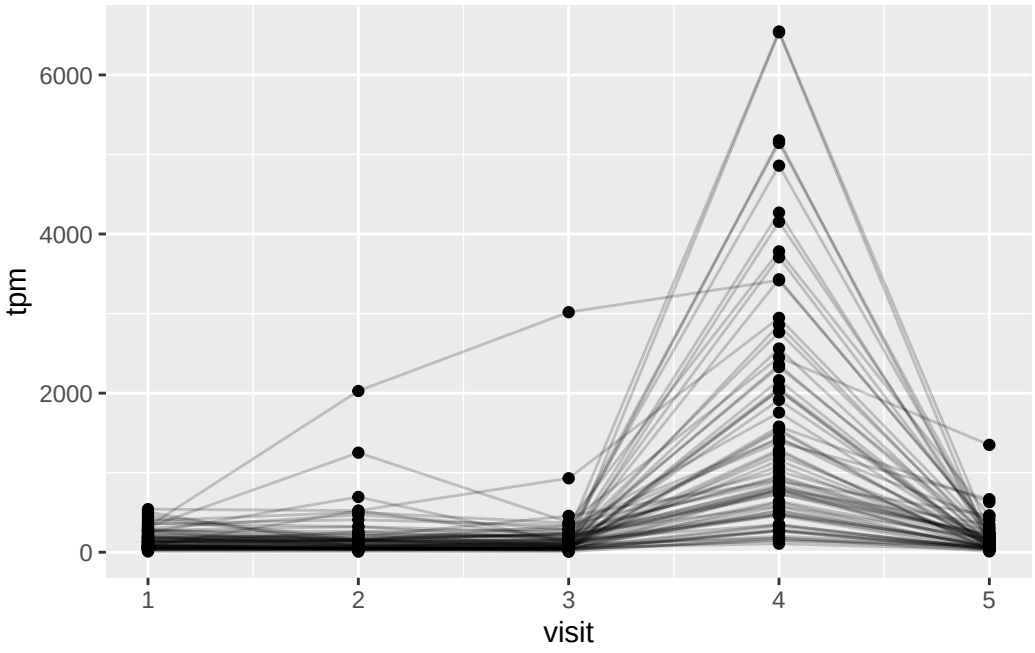
```
head(ssrna)
```

```
  versioned_ensembl_gene_id specimen_id raw_count      tpm subject_id
1         ENSG00000211896.7         344     18613  929.640         44
2         ENSG00000211896.7         243      2011  112.584         31
3         ENSG00000211896.7         261      2161  124.759         33
4         ENSG00000211896.7         282      2428  138.292         36
5         ENSG00000211896.7         345     51963 2946.136         44
6         ENSG00000211896.7         244     49652 2356.749         31
  actual_day_relative_to_boost planned_day_relative_to_boost specimen_type
1                            3                             3         Blood
2                            3                             3         Blood
3                           15                            14         Blood
4                            1                             1         Blood
5                            7                             7         Blood
6                            7                             7         Blood
  visit infancy_vac biological_sex              ethnicity                race
1     3          aP         Female     Hispanic or Latino More Than One Race
2     3          wP         Female Not Hispanic or Latino               Asian
3     5          wP           Male     Hispanic or Latino More Than One Race
4     2          aP         Female     Hispanic or Latino               White
5     4          aP         Female     Hispanic or Latino More Than One Race
6     4          wP         Female Not Hispanic or Latino               Asian
  year_of_birth date_of_boost      dataset      age
1    1998-01-01    2016-11-07 2020_dataset 25.19644
2    1989-01-01    2016-09-26 2020_dataset 34.19576
3    1990-01-01    2016-10-10 2020_dataset 33.19644
4    1997-01-01    2016-10-24 2020_dataset 26.19576
5    1998-01-01    2016-11-07 2020_dataset 25.19644
6    1989-01-01    2016-09-26 2020_dataset 34.19576
```

Q18. 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)
```



Q19.: What do you notice about the expression of this gene (i.e. when is it at it's maximum level)?

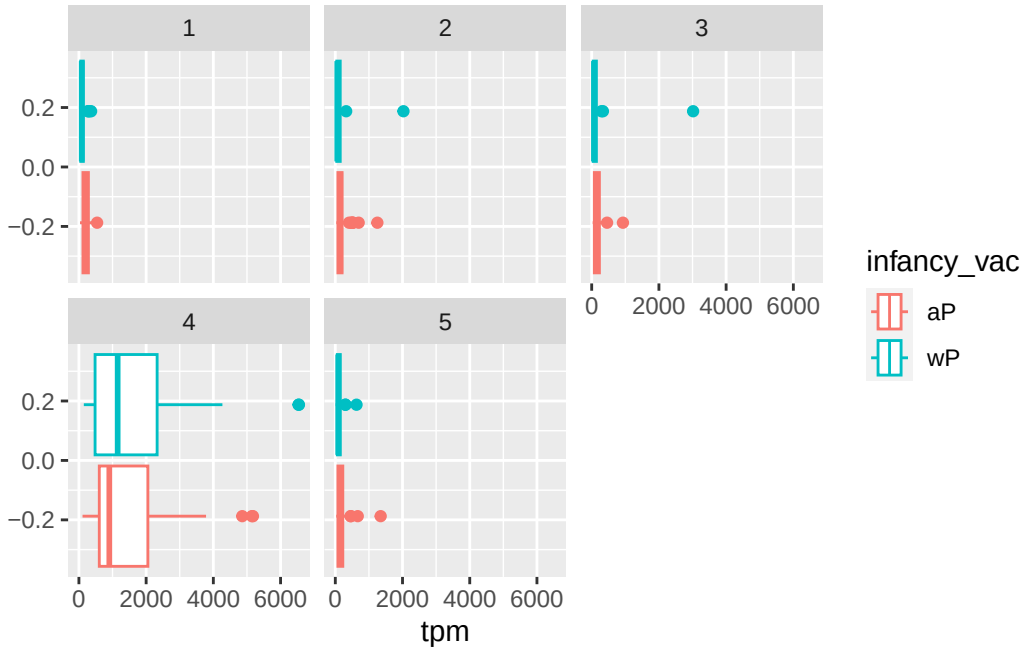
The maximum level of the expression of the IGHG1 gene is at visit 4. Between visit 3 and 4, there is a rapid peak and between visit 4 and 5 there is a rapid decline.

Q20. Does this pattern in time match the trend of antibody titer data? If not, why not?

No, the antibody titer data peaks around visit 6 and declines much more slowly. (peaks earler and declines slower)

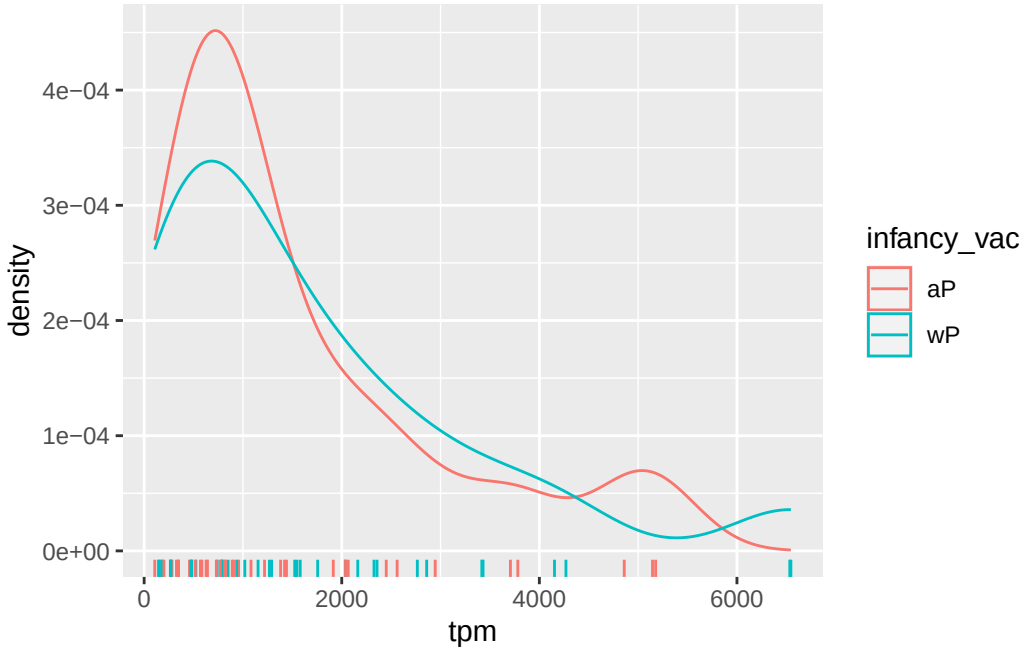
We can learn more by color/facet with infancy_vac status (the aP or wP they received as an infant)

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



We can filter for visit 4:

```
ssrna %>%
  filter(visit==4) %>%
  ggplot() +
    aes(tpm, col=infancy_vac) + geom_density() +
    geom_rug()
```



Is RNA-Seq expression levels predictive of Ab titers?

Yes, it appears so

What differentiates aP vs. wP primed individuals?

That is what we are trying to distinguish. Aside from the physical responses (the reason people switched from wP to aP), they appear to be different in the quickness, peak and decline of the immune response.

What about decades after their first immunization? Do you know? Contact Bjoern and Barry for your trip to Sweden :-)

We cannot do clinical trials for that long (lack of funding), so it is unclear. That is why we are collecting this data.

```
sessionInfo()
```

```
R version 4.2.2 (2022-10-31)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Big Sur ... 10.16

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
```

```
LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] dplyr_1.1.0     lubridate_1.9.2 jsonlite_1.8.4  ggplot2_3.4.1

loaded via a namespace (and not attached):
 [1] rstudioapi_0.14  knitr_1.42       magrittr_2.0.3   tidyselect_1.2.0
 [5] munsell_0.5.0    timechange_0.2.0 colorspace_2.1-0 R6_2.5.1
 [9] rlang_1.0.6      fastmap_1.1.1    fansi_1.0.4      tools_4.2.2
[13] grid_4.2.2       gtable_0.3.1     xfun_0.37        utf8_1.2.3
[17] cli_3.6.0        withr_2.5.0      htmltools_0.5.4  yaml_2.3.7
[21] digest_0.6.31    tibble_3.2.0     lifecycle_1.0.3  farver_2.1.1
[25] vctrs_0.5.2      glue_1.6.2       evaluate_0.20    rmarkdown_2.20
[29] labeling_0.4.2   compiler_4.2.2   pillar_1.8.1     generics_0.1.3
[33] scales_1.2.1     pkgconfig_2.0.3
```