

# Lab15: Pertussis Mini Project

Sam Altshuler (PID: A59010373)

3/9/2022

Libraries being used:

```
library(ggplot2)
library(jsonlite)
library(lubridate)
library(dplyr)
```

## Class 15: Investigating Pertussis Resurgence

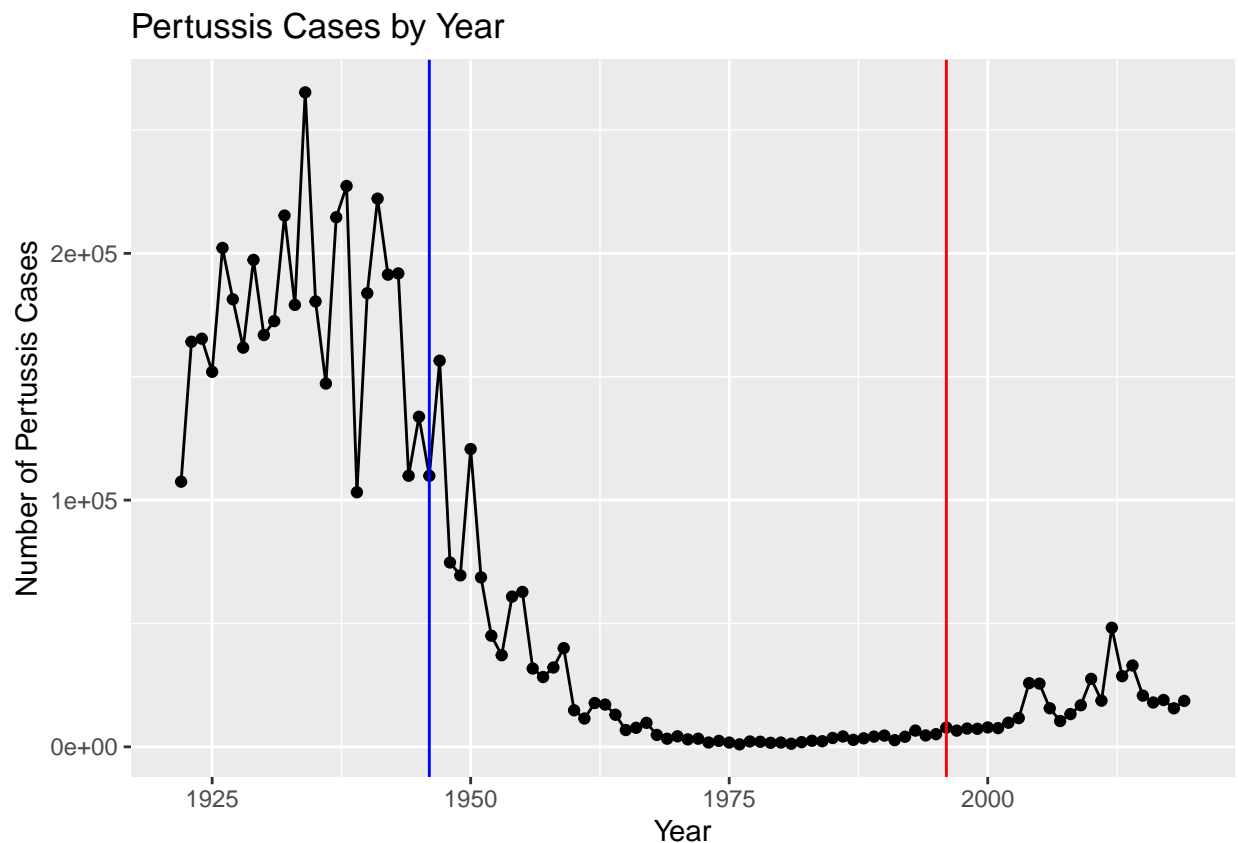
Data obtained from <https://www.cdc.gov/pertussis/surv-reporting/cases-by-year.html>.

```
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),
  No..Reported.Pertussis.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)
```

Plot cases of Pertussis over time.

```
ggplot(cdc) +
  aes(x= Year, y = No..Reported.Pertussis.Cases)+
  geom_point()+
  geom_line()+
  labs(title = "Pertussis Cases by Year", x = "Year", y = "Number of Pertussis Cases")+
  geom_vline( xintercept = 1946, color = "blue")+
  geom_vline( xintercept = 1996, color = "red")
```



The vaccine is first introduced in 1946 as a whole-cell Pertussis vaccine (blue line). In 1996, the vaccine was switched to acellular Pertussis (red line) in order to lessen the side effects from the wP vaccine. Pertussis cases seems to increase again after the implementation of the aP vaccine. This could be due to an increase in vaccine hesitancy, more people travelling to/from areas of high pertussis case rates, or even the aP vaccine doesn't have as sustained of immune response (waning immunity) as the wP vaccine.

## Exploring CMI-PB Data

We will access the JSON formatted data via the CMI-PB API and the `jsonlite` package.

```
subject <- read_json("https://www.cmi-pb.org/api/subject", simplifyVector = TRUE)
head(subject, 3)
```

```
##   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
##   year_of_birth date_of_boost  study_name
## 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
```

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

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

```
##
## aP wP
## 47 49
```

There are 47 individuals who got the aP vaccine and 49 individuals that got the wP vaccine.

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

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

```
##
## Female   Male
##    66    30
```

There are 66 females and 30 males in the dataset.

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
```

## Joining multiple tables

```
# Complete the API URLs...
specimen <- read_json("https://www.cmi-pb.org/api/specimen", simplifyVector = TRUE)
titer <- read_json("http://cmi-pb.org/api/ab_titer", simplifyVector = TRUE)
```

Now to make sense of the data and ask questions about aP vs wP of the ab titer data for example. I need to merge (“join”) the subject data with these new tables.

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

```
## Joining, by = "subject_id"
```

Can do either an inner\_join() or full\_join() but we want to do inner\_join() because we only want to analyze subjects that are in both datasets.

Now we want to add in the ab Titer data to further analyze the data.

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

```
## Joining, by = "specimen_id"
```

```
dim(abdata)
```

```
## [1] 32675    19
```

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
```

The number of visit 8 specimens are much much lower than visits 1-7.

## Examine IgG1 Ab titer levels

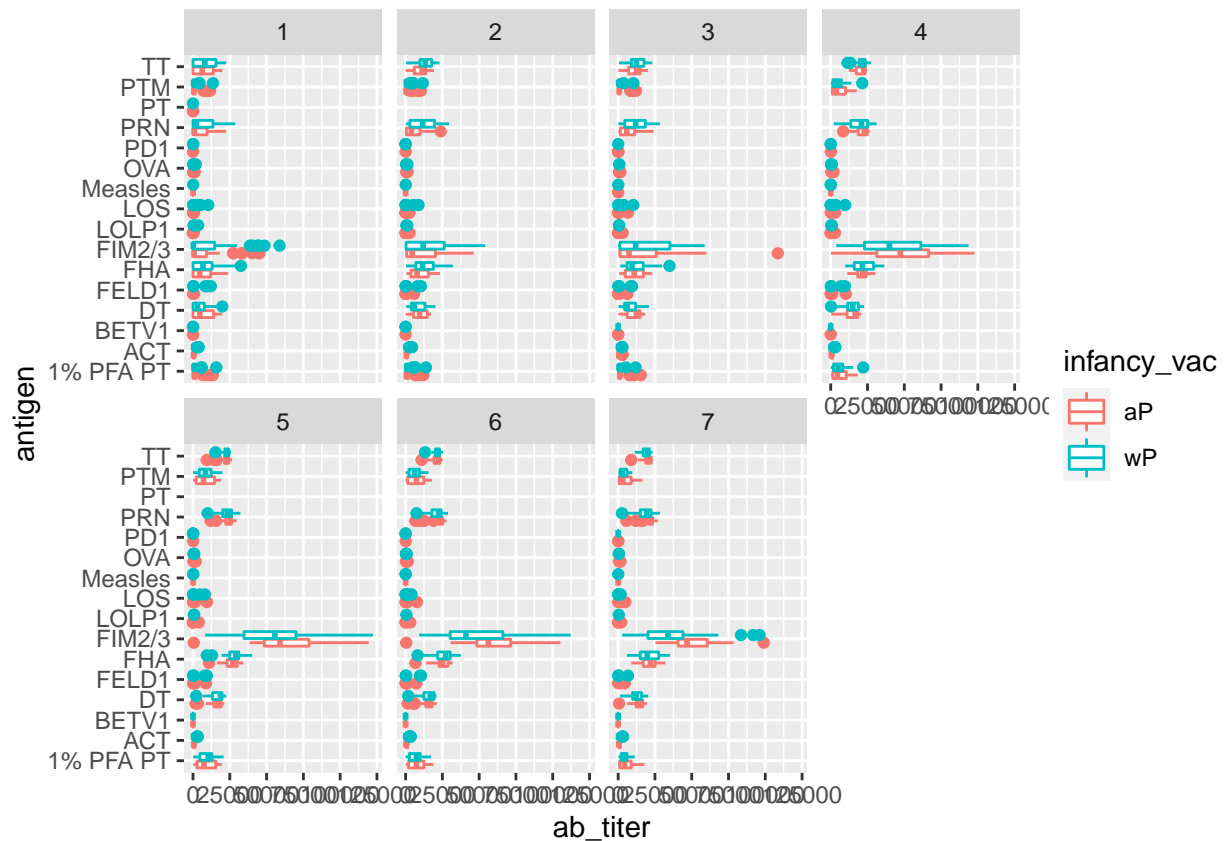
Now we are going to use the dataset built/joined above to focus in on the IgG1 isotype and exclude visit 8 (which is a low number).

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

```
##   specimen_id isotype is_antigen_specific antigen  ab_titer  unit
## 1           1   IgG1                TRUE     ACT 274.355068 IU/ML
## 2           1   IgG1                TRUE     LOS 10.974026 IU/ML
## 3           1   IgG1                TRUE    FELD1 1.448796 IU/ML
## 4           1   IgG1                TRUE    BETV1 0.100000 IU/ML
## 5           1   IgG1                TRUE    LOLP1 0.100000 IU/ML
## 6           1   IgG1                TRUE Measles 36.277417 IU/ML
##   lower_limit_of_detection subject_id actual_day_relative_to_boost
## 1                   3.848750           1                    -3
## 2                   4.357917           1                    -3
## 3                   2.699944           1                    -3
## 4                   1.734784           1                    -3
## 5                   2.550606           1                    -3
## 6                   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 study_name
## 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
```

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

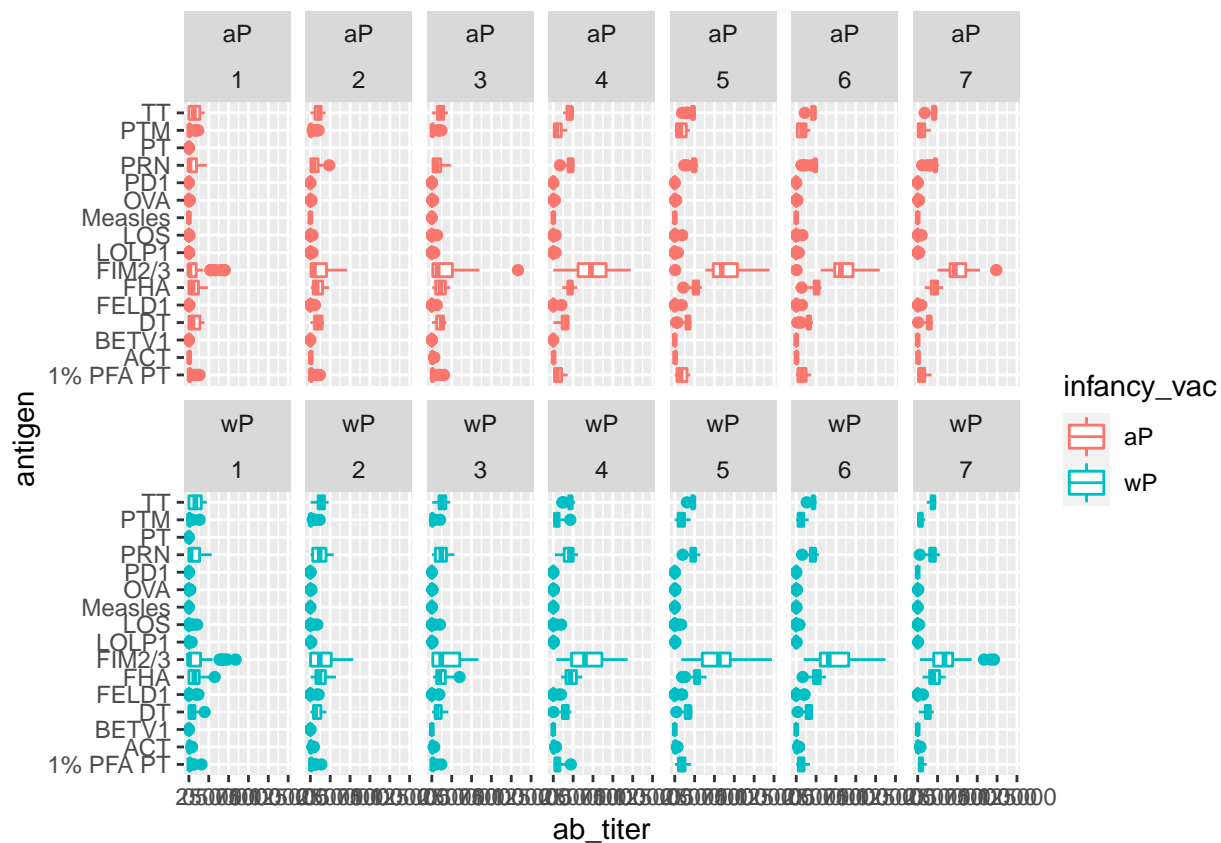
```
ggplot(ig1) +
  aes(ab_titer, antigen, color = infancy_vac) +
  geom_boxplot() +
  facet_wrap(vars(visit), nrow=2)
```



Q14. What antigens show differences in the level of IgG1 antibody titers recognizing them over time? Why these and not others? - The antigens that increase as the number of visits increases are the components of the vaccine. The other antibodies for the antigens aren't increasing because the individuals aren't being exposed to these antigens.

Now I can add `infancy_vac` to the facetting.

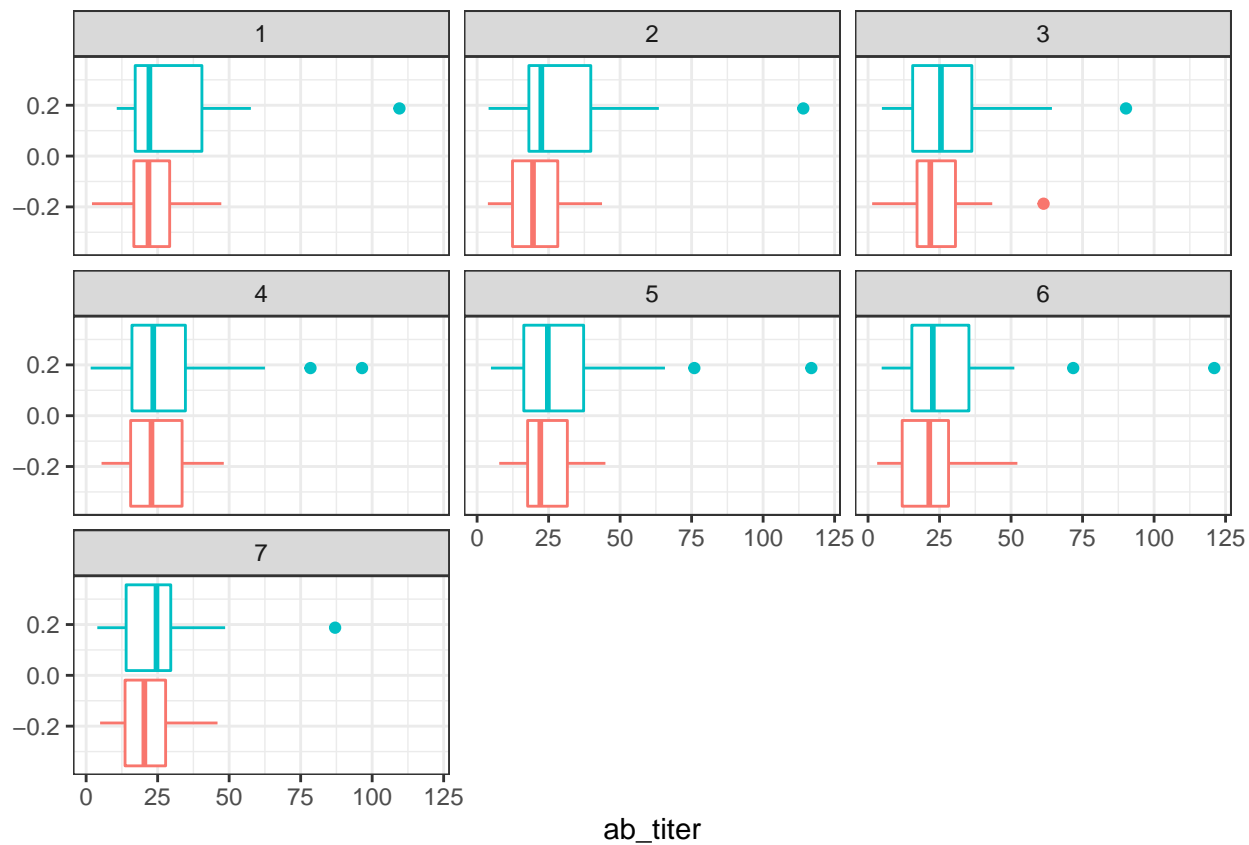
```
ggplot(ig1) +
  aes(ab_titer, antigen, color = infancy_vac) +
  geom_boxplot() +
  facet_wrap(vars(infancy_vac, visit), nrow=2)
```



```

ig1_measles <- ig1 %>% filter(antigen == "Measles")
ggplot(ig1_measles) +
  aes(ab_titer, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()

```

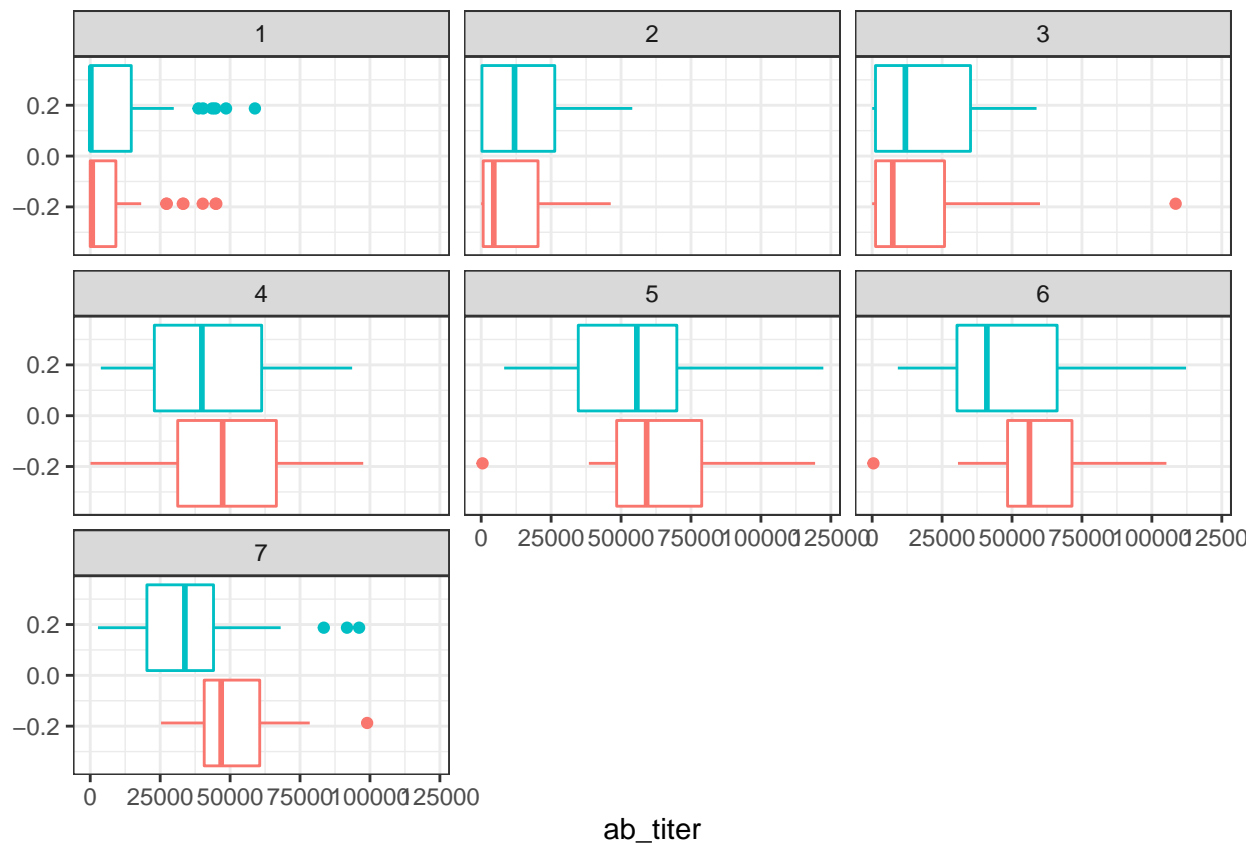


```

ig1_fim <- ig1 %>% filter(antigen == "FIM2/3")
ggplot(ig1_fim) +
  aes(ab_titer, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()

```





The FIM2/3 data seems to increase much faster and to a higher level than any other antibody response to an antigen.

HOWEVER, it doesn't appear that the aP data is any different than the wP data.

## CMI-PB RNASeq data

```
url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_gene_id=eq.ENS00000211896.7"
rna <- read_json(url, simplifyVector = TRUE)
```

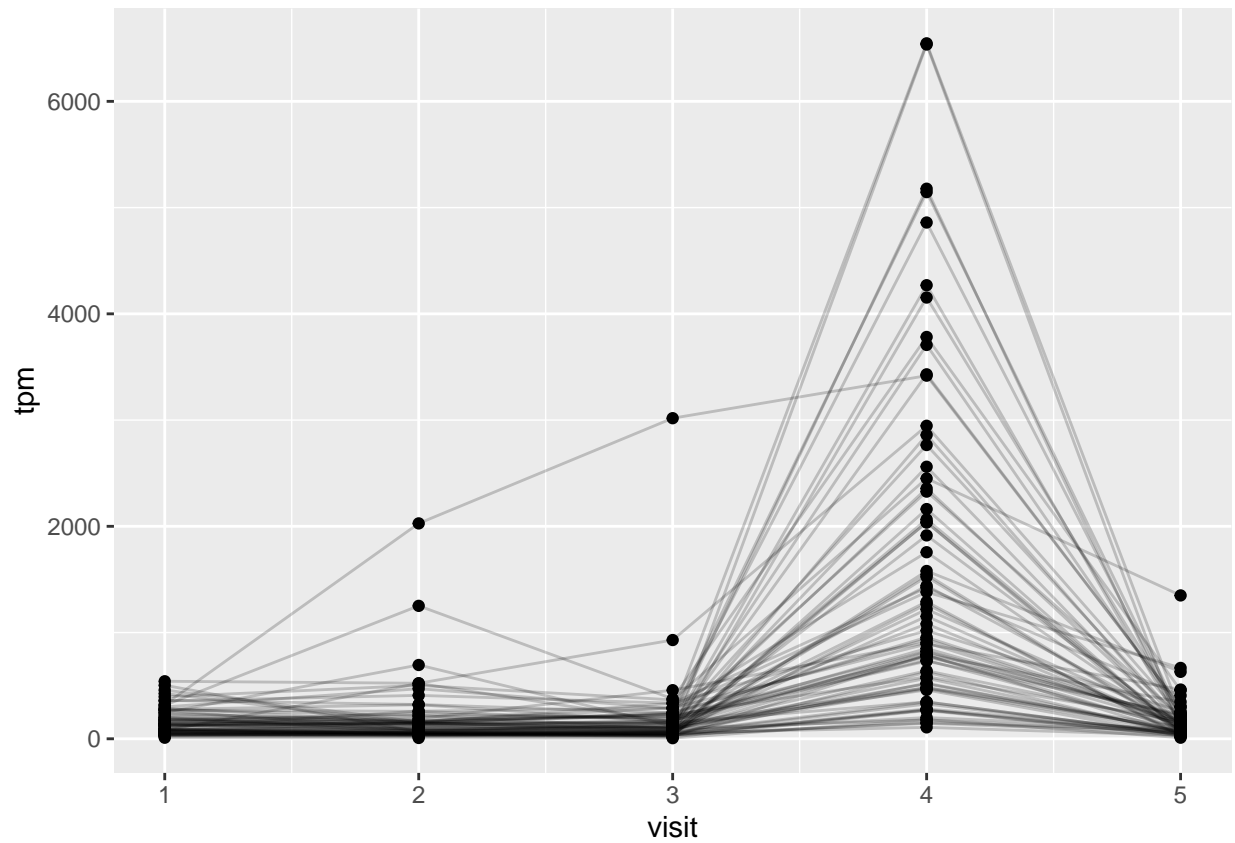
Join this data to the meta dataset.

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

```
## Joining, by = "specimen_id"
```

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(x = visit, y = 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)? - This gene is at it's maximum expression around the fourth visit Q20. Does this pattern in time match the trend of antibody titer data? If not, why not? - This pattern seems to match the first initial increase in FIM2/3 antibodies.