

# Lab 19 - Pertussis

AUTHOR

Catherine Diep

## Lab 19 - Pertussis Resurgence Mini Project

---

### Q1.

---

```
cdc <- data.frame(  
  Year = c(1922L,  
           1923L, 1924L, 1925L, 1926L, 1927L,  
           1929L, 1930L, 1931L, 1932L, 1933L,  
           1936L, 1937L, 1938L, 1939L, 1940L,  
           1942L, 1943L, 1944L, 1945L, 1946L,  
           1949L, 1950L, 1951L, 1952L, 1953L,  
           1955L, 1956L, 1957L, 1958L, 1959L,  
           1961L, 1962L, 1963L, 1964L, 1965L,  
           1968L, 1969L, 1970L, 1971L, 1972L,  
           1974L, 1975L, 1976L, 1977L, 1978L,  
           1981L, 1982L, 1983L, 1984L, 1985L,  
           1987L, 1988L, 1989L, 1990L, 1991L,  
           1994L, 1995L, 1996L, 1997L, 1998L,  
           2000L, 2001L, 2002L, 2003L, 2004L,  
           2006L, 2007L, 2008L, 2009L, 2010L)
```

```

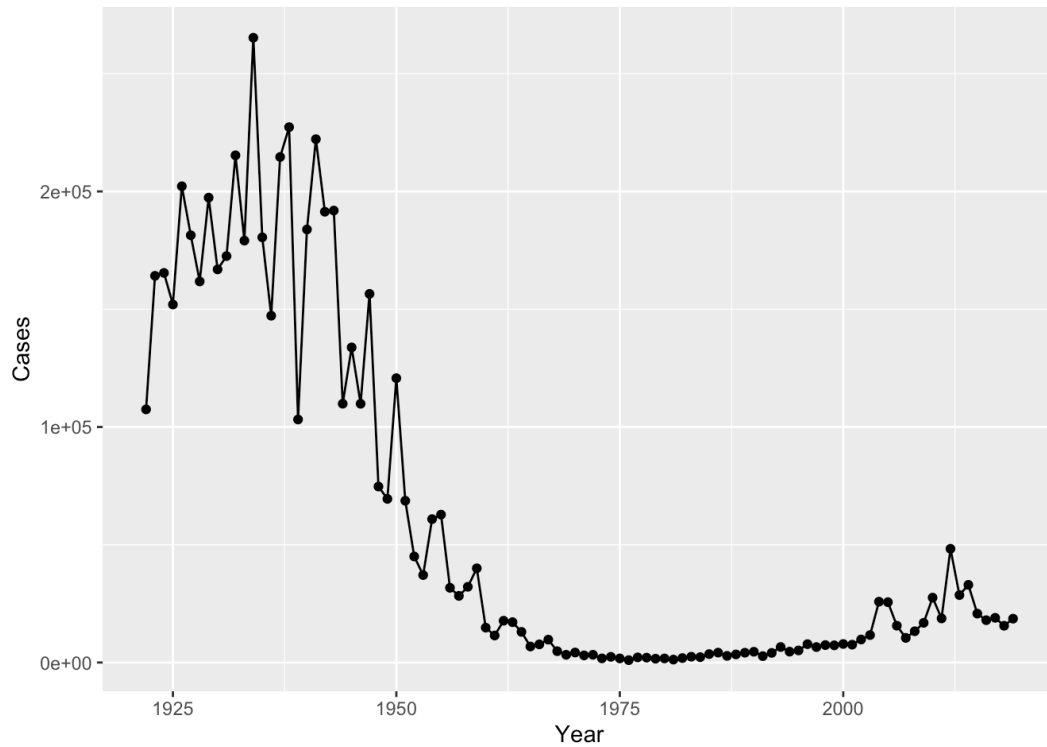
2013L,2014L,2015L,2016L,2017L,
2019L),
Cases = c(107473,
164191,165418,152003,202210,
161799,197371,166914,172559,
265269,180518,147237,214652,
183866,222202,191383,191890,
133792,109860,156517,74715,6
68687,45030,37129,60886,6278
32148,40005,14809,11468,1774
13005,6799,7717,9718,4810,32
3036,3287,1759,2402,1738,101
1623,1730,1248,1895,2463,227
4195,2823,3450,4157,4570,271
4617,5137,7796,6564,7405,729
7580,9771,11647,25827,25616,
13278,16858,27550,18719,4827
32971,20762,17972,18975,1560
)

```

```

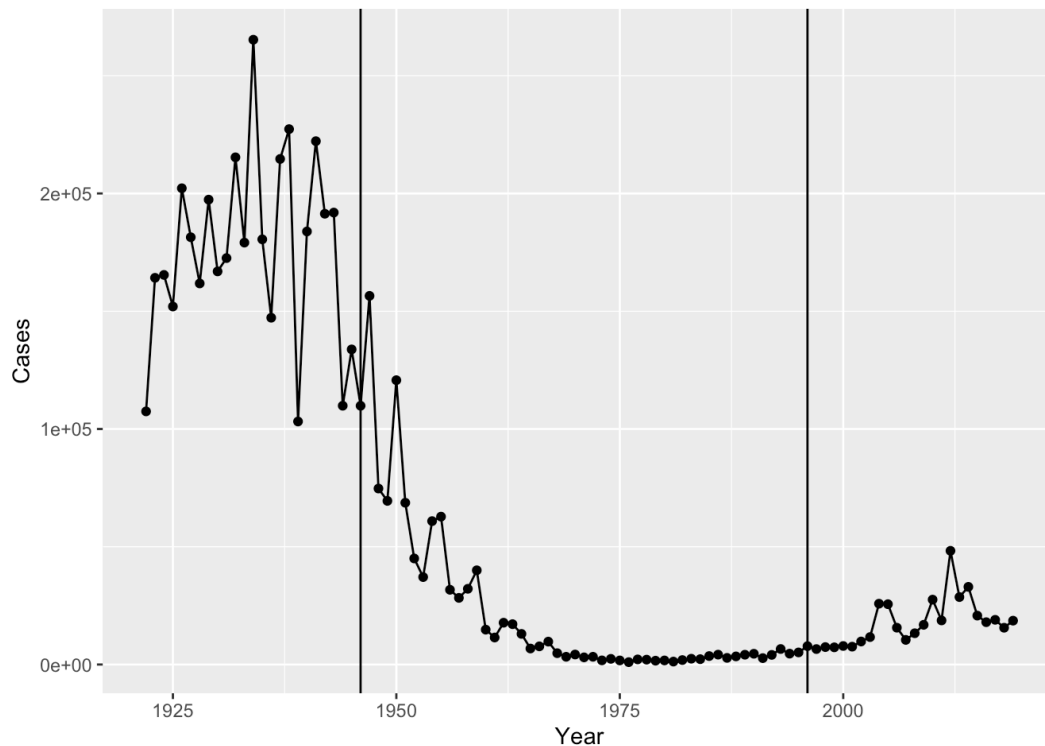
library(ggplot2)
ggplot(cdc) +
  aes(Year, Cases) +
  geom_point() +
  geom_line() +
  labs("Year", "Number of cases")

```



## Q2.

```
ggplot(cdc) +  
  aes(Year, Cases) +  
  geom_point() +  
  geom_line() +  
  labs("Year", "Number of cases") +  
  geom_vline(xintercept=1946) +  
  geom_vline(xintercept=1996)
```



After the introduction of the wP vaccine, the number of pertussis cases appears to exponentially decrease. After the introduction of the aP vaccine, the number of cases increases.

### Q3.

After the introduction of the aP vaccine, there were more pertussis cases per year. This may be because the aP vaccine is not as effective as the wP vaccine, vaccine hesitancy so fewer children were getting the vaccine, and also wearing immunity from the old wP vaccine.

```
library(jsonlite)
subject <- read_json("https://www.cmi-pb.org/api/subject", simplify = FALSE)

head(subject, 3)
```

	subject_id	infancy_vac	biological_sex	ethnicity
1	1	wP	Female	Not Hispanic or Latino
2	2	wP	Female	Not Hispanic or Latino

	White				
3	3	wP	Female		Unknown

	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	

## Q4.

```
nrow(subject[subject$infancy_vac == 'wP',])
```

```
[1] 49
```

```
nrow(subject[subject$infancy_vac == 'aP',])
```

```
[1] 47
```

47 aP, 49 wP

## Q5.

```
nrow(subject[subject$biological_sex == 'Female',])
```

```
[1] 66
```

```
nrow(subject[subject$biological_sex == 'Male',])
```

```
[1] 30
```

66 female, 30 male

## Q6.

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

	American Indian/Alaska Native	Asian	Black or African American
Female	0	18	
Male	1	9	
More Than One Race			
Native Hawaiian or Other Pacific Islander			
Female	8		
Male	2		
Unknown or Not Reported			
White			
Female	10	27	
Male	4	13	

The breakdown of race and biological sex are in the table above.

## Q7.

```
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

```
library(lubridate)
```

Attaching package: 'lubridate'

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

date, intersect, setdiff, union

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

```
ap <- subject %>% filter(infancy_vac == "aP")
round( summary( time_length( ap$age, "years" ) ) )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
23	25	26	26	26	27

```
wp <- subject %>% filter(infancy_vac == "wP")
round( summary( time_length( wp$age, "years" ) ) )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
28	32	35	36	40	55

aP average age: 26, wP average age: 36, they are significantly different

## Q8.

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

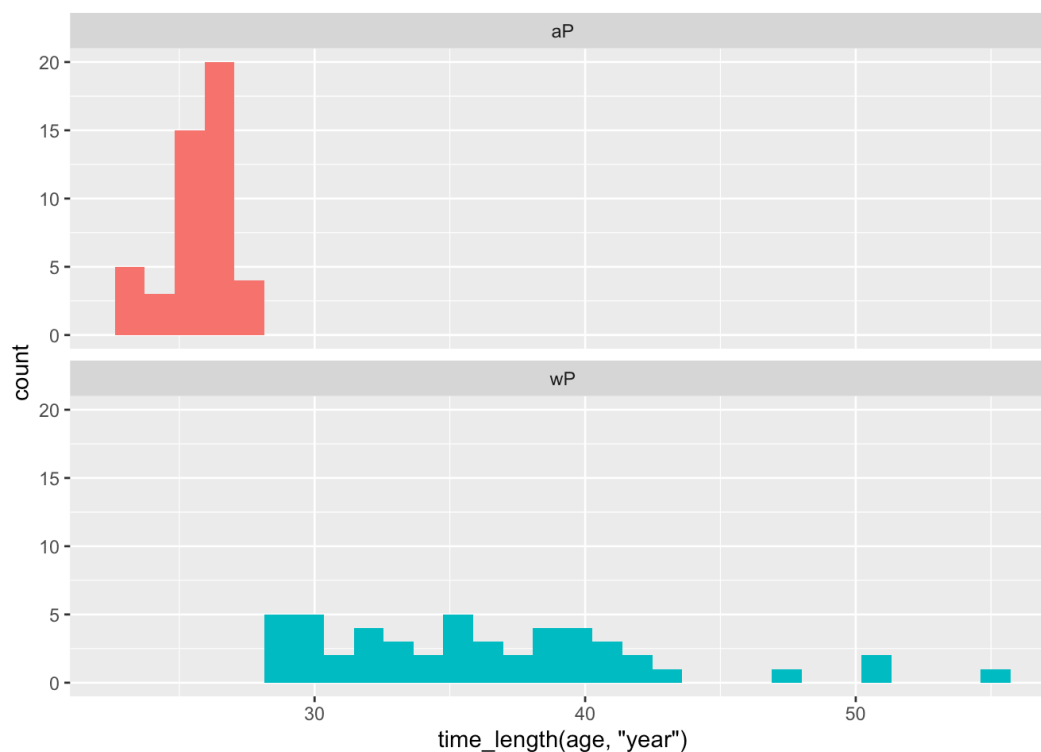
```
[1] 30.69678 51.07461 33.77413 28.65982 25.65914 28.77481
```

Above are the ages of individuals at time of their boost.

## Q9.

```
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`.



Yes, these groups are significantly different.

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

## Q9.

```
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	planned_day_relative_to_boost	specimen_type	visit	infancy_vac	biological_sex	ethnicity	race	year_of_birth	date_of_boost	dataset	age
1	1	1			-3								
2	2	1			736								
3	3	1			1								
4	4	1			3								
5	5	1			7								
6	6	1			11								
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												
1													
2													
3													
4													
5													
6													
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													
1													
2													
3													
4													
5													
6													

## Q10.

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

Joining with `by = join\_by(specimen\_id)`

```
dim(abdata)
```

```
[1] 32675    21
```

## Q11.

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

```

IgE  IgG IgG1 IgG2 IgG3 IgG4
6698 1413 6141 6141 6141 6141

```

Above are the number of specimens for each isotype.

## Q12.

```
table(meta$visit)
```

```

1  2  3  4  5  6  7  8  9 10 11
96 96 96 96 96 90 86 42 19 11  1

```

There are fewer visit 8 specimens than visit 1-7 specimens

```
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

```

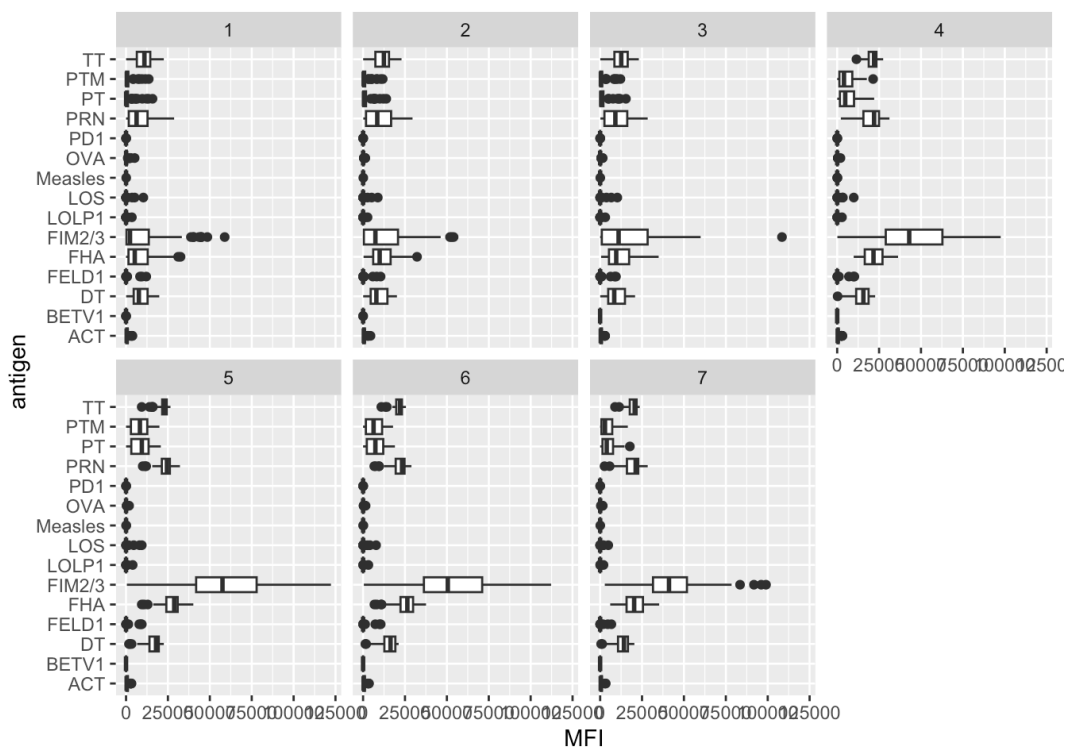
4 Not Hispanic or Latino White	1986-01-01	2016-09-12
2020_dataset		
5 Not Hispanic or Latino White	1986-01-01	2016-09-12
2020_dataset		
6 Not Hispanic or Latino White	1986-01-01	2016-09-12
2020_dataset		

age

1 13592 days
2 13592 days
3 13592 days
4 13592 days
5 13592 days
6 13592 days

## Q13.

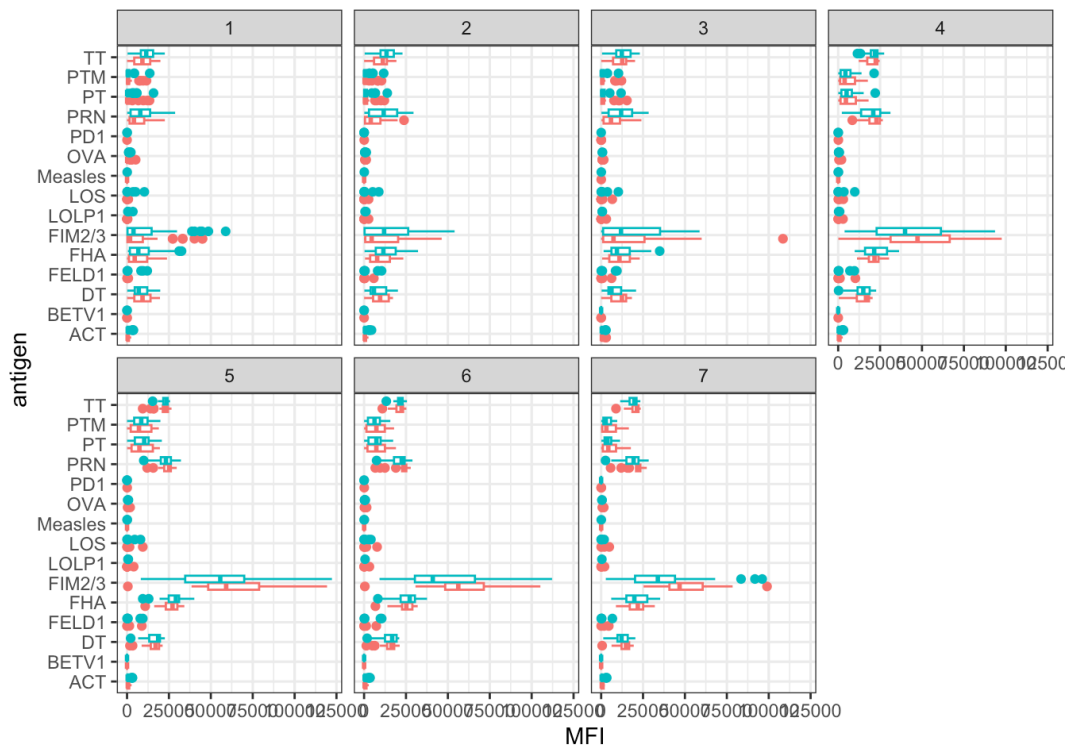
```
ggplot(ig1) +
  aes(MFI, antigen) +
  geom_boxplot() +
  facet_wrap(vars(visit), nrow=2)
```



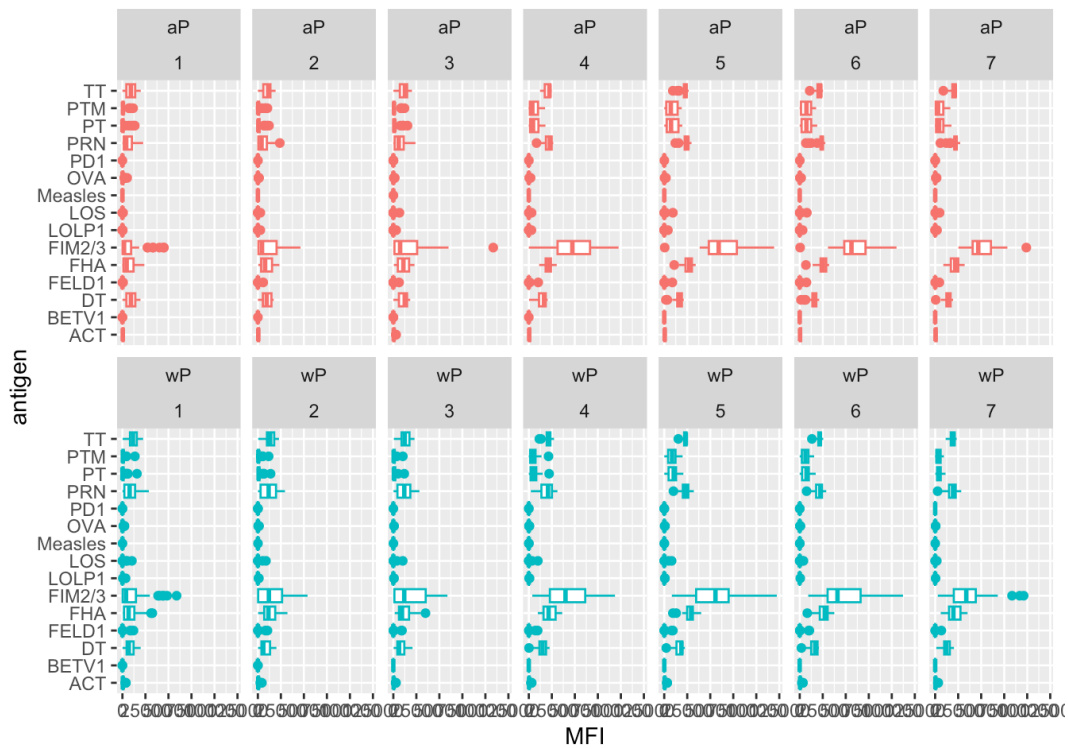
## Q14.

TT, PTM, PRN, FIM2/3, FHA, and DT show differences in IgG levels over time. This may be because these antibodies are produced via vaccination, so after vaccination antibodies increase and then decrease over the long term.

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

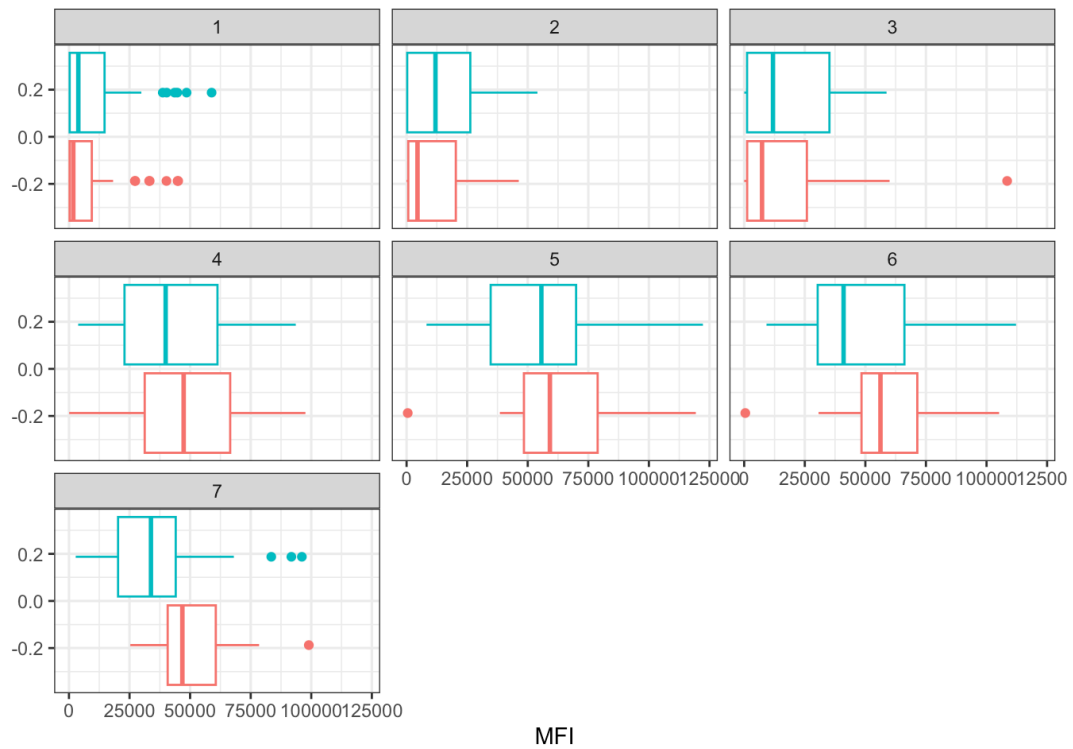


```
ggplot(ig1) +
  aes(MFI, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(infancy_vac, visit), nrow=2)
```

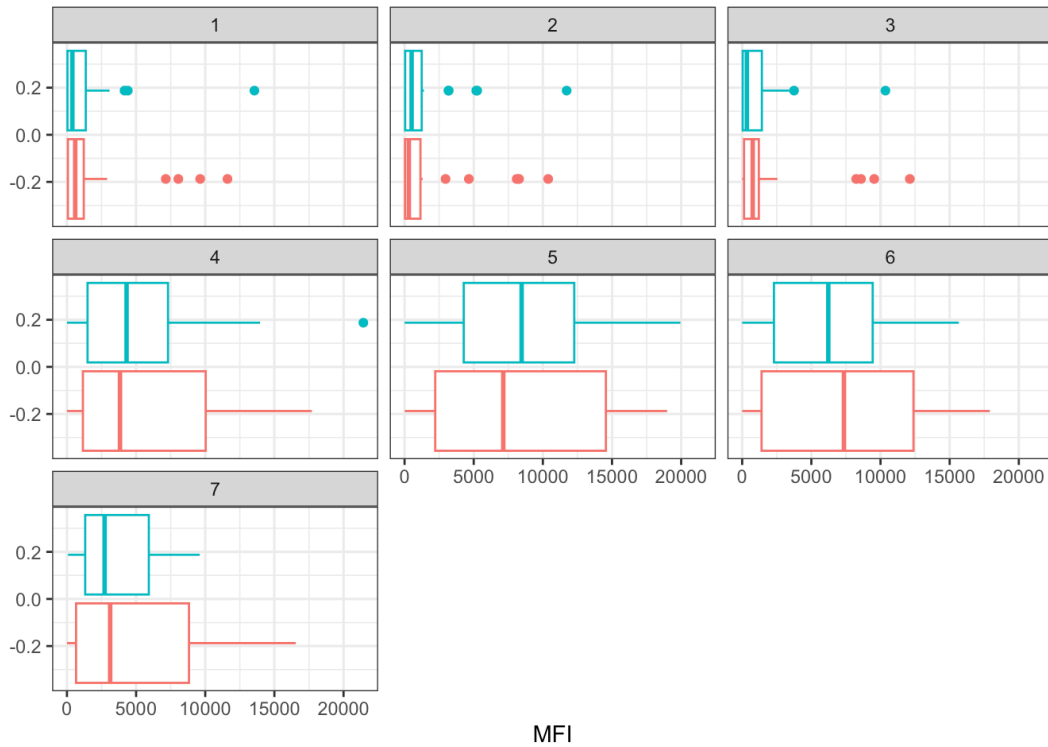


## Q15.

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

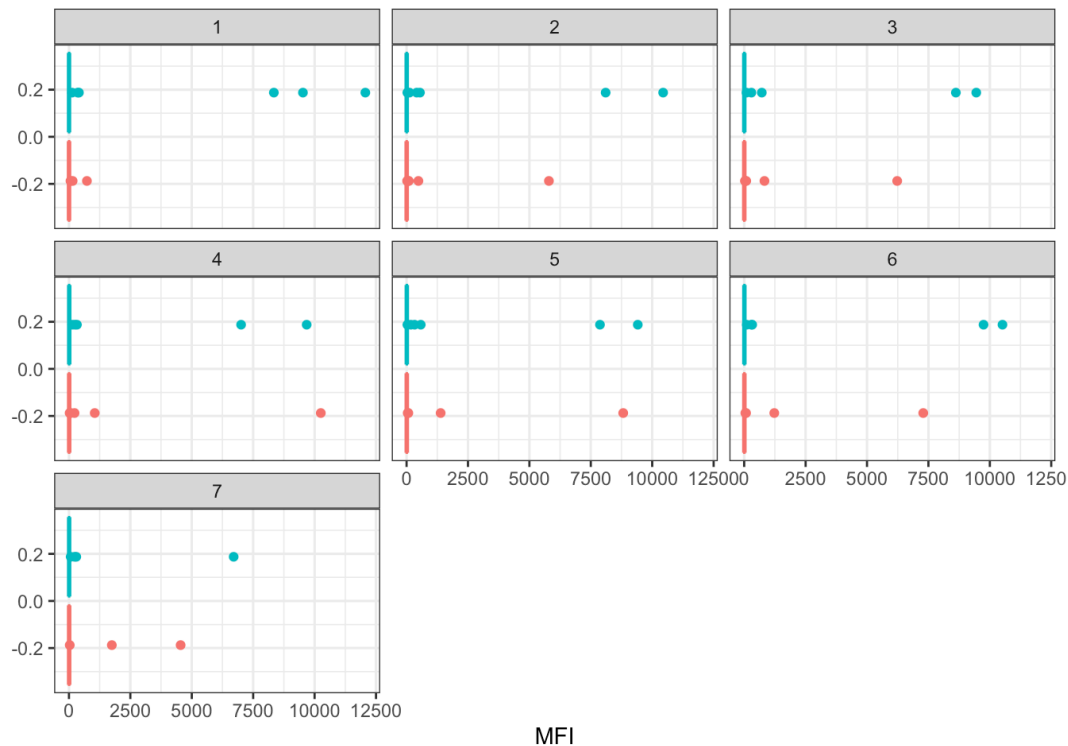


```
filter(ig1, antigen=="PTM") %>%
  ggplot() +
    aes(MFI, col=infancy_vac) +
    geom_boxplot(show.legend = FALSE) +
    facet_wrap(vars(visit)) +
    theme_bw()
```



```
filter(ig1, antigen=="FELD1") %>%
  ggplot() +
    aes(MFI, col=infancy_vac) +
    geom_boxplot(show.legend = FALSE) +
    facet_wrap(vars(visit)) +
    theme_bw()
```





## Q16.

Both FIM2/3 and PTM peak at around visits 5-6, with more PTM than FIM2/3 antigens detected; however, at visit 7 there appears to be more FIM2/3 than PTM. Meanwhile, FELD1 is stagnant at low levels of 0.

## Q17.

There is not too much of a difference between wP and aP responses for any of the antigens.

```
url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_11896.7"
```

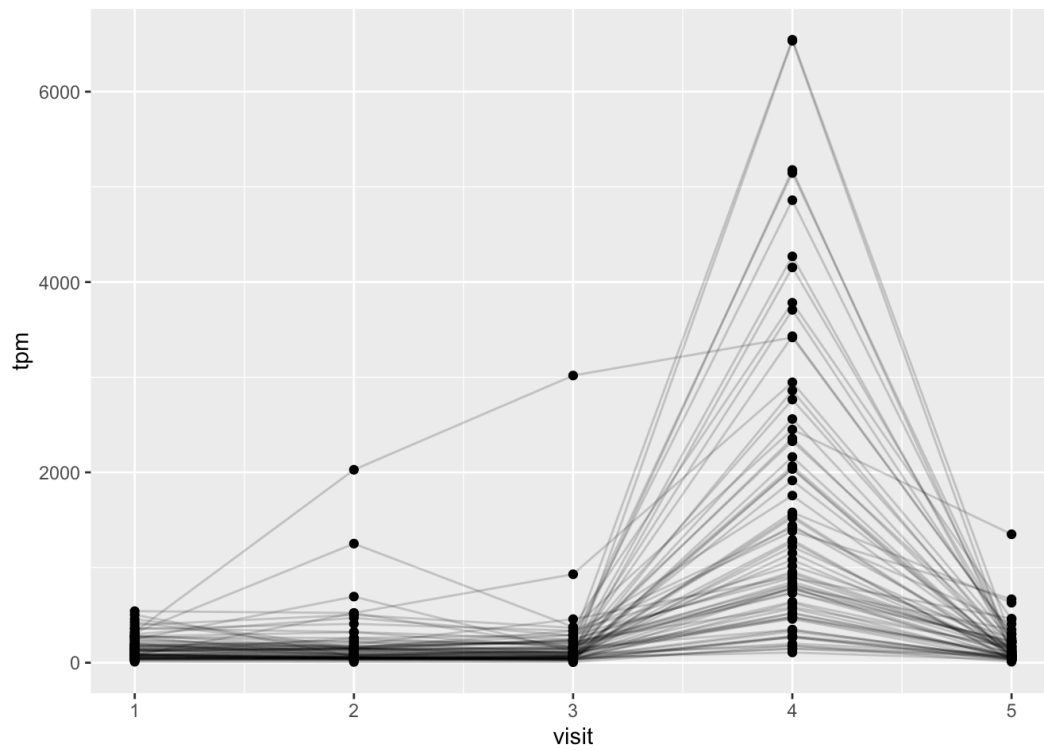
```
rna <- read_json(url, simplifyVector = TRUE)
```

```
#meta <- inner_join(specimen, subject)
ssrna <- inner_join(rna, meta)
```

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

## Q18.

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



## Q19.

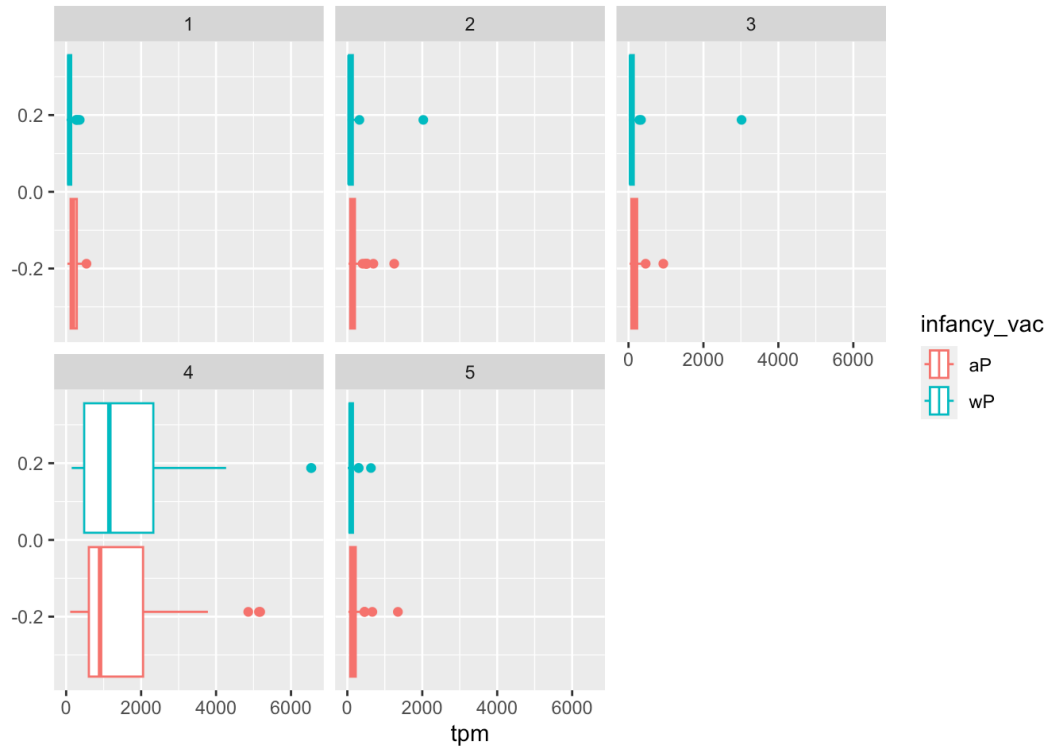
The gene increases expression after visit 3, hitting its peak at visit 4, and then decreasing expression to lower levels at visit 5.

## Q20.

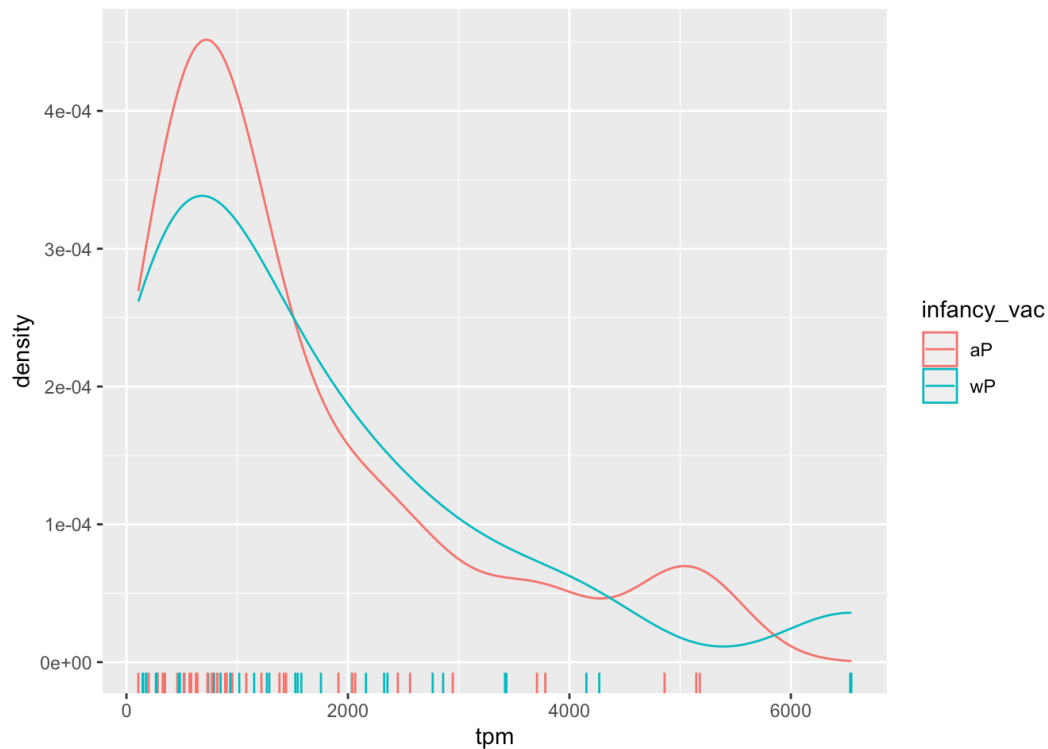
This does match the trend of antibody titer data which shows that there are the maximum number of antibodies around visit 5-6; expression would

result in these antibodies which are long-lived and persist after they are made.

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



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



```
sessionInfo()
```

R version 4.2.1 (2022-06-23)

Platform: x86\_64-apple-darwin17.0 (64-bit)

Running under: macOS Catalina 10.15.7

Matrix products: default

BLAS:

/Library/Frameworks/R.framework/Versions/4.2/Resources/lib/lib  
Rblas.0.dylib

LAPACK:

/Library/Frameworks/R.framework/Versions/4.2/Resources/lib/lib  
Rlapack.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] lubridate\_1.9.2 dplyr\_1.1.0 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_4.2.1  
[13] grid_4.2.1         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_3.1.8        lifecycle_1.0.3  
farver_2.1.1  
[25] htmlwidgets_1.6.1  vctrs_0.5.2         glue_1.6.2  
evaluate_0.20  
[29] rmarkdown_2.20     labeling_0.4.2       compiler_4.2.1  
pillar_1.8.1  
[33] generics_0.1.3     scales_1.2.1        pkgconfig_2.0.3
```

...