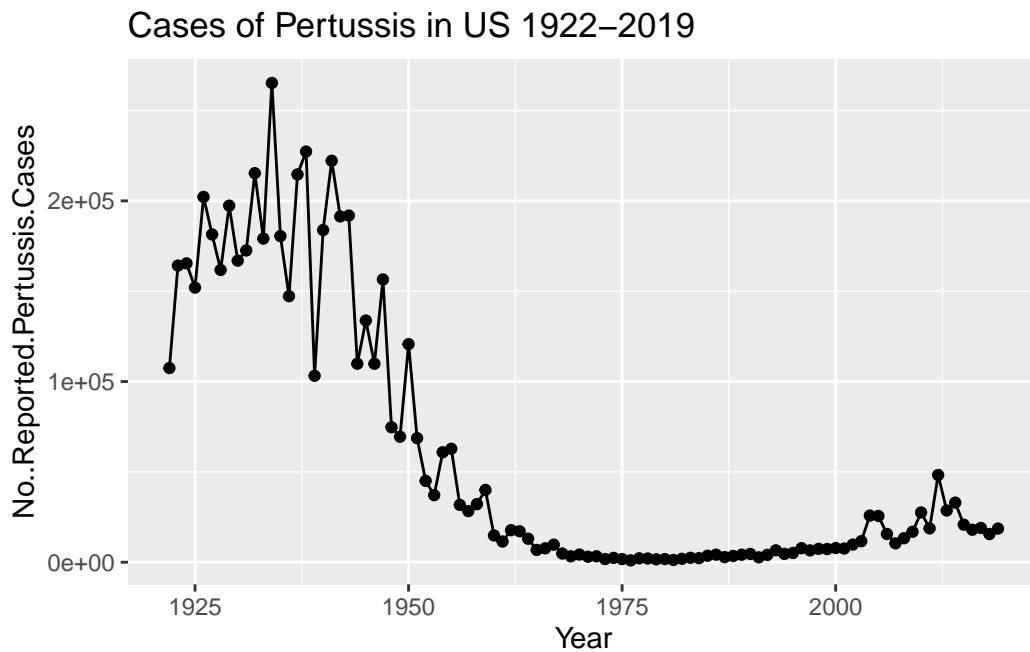


# class19

Colin Mach

Q1.) Plot of case numbers of persussis over time.

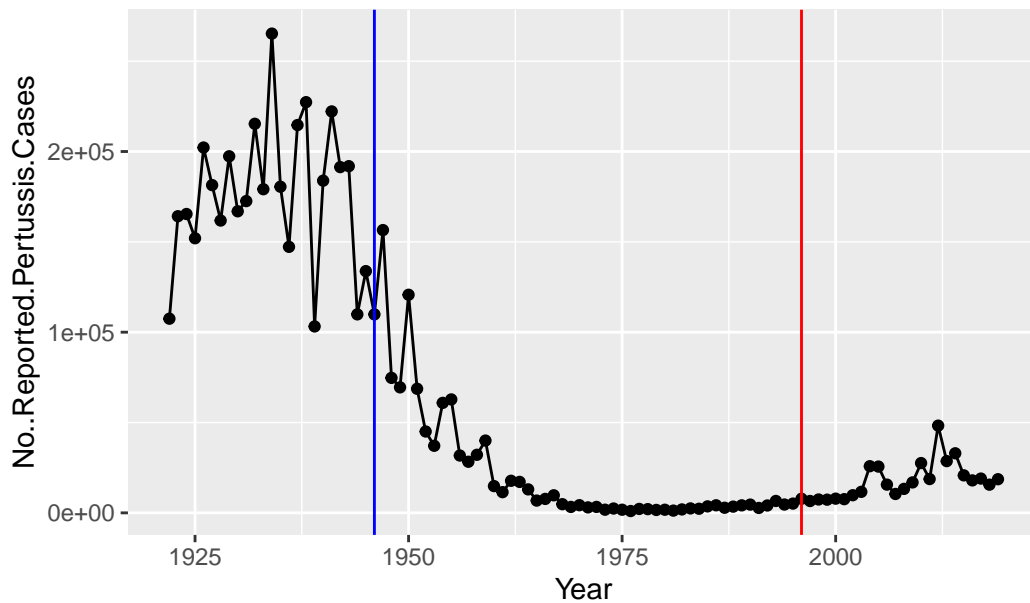
```
library(ggplot2)
baseplot <- ggplot(cdc) + aes(x = Year, y = No..Reported.Pertussis.Cases) + geom_line() +
baseplot
```



Q2.)

```
baseplot + geom_vline(xintercept = 1946, color = "blue") + geom_vline(xintercept = 1996, c
```

Cases of Pertussis in US 1922–2019



Q3.) What happened during these time periods?

There was a resurgence after the switch to the aP vaccine since the vaccine may be less effective than the wP vaccine and may not provide as much immunity since the pertussis may have evolved and adapted from the aP vaccine. The spike was also correlated with vaccine hesitancy since that was when people were misinformed that vaccines cause autism which has had a lasting impact since the 2000s.

## CMI-PB

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

Q4.) Number of infancy vaccinated subjects in the dataset?

```
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	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	

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

```
aP wP
47 49
```

There are 96 vaccinated subjects

Q5.) How many Male and Female subjects are in the dataset

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

```
Female  Male
66      30
```

There are 66 female and 30 male subjects

Q6.) What is the breakdown of race and biological sex

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

	American Indian/Alaska Native	Asian	Black or African American
Female	0	18	2
Male	1	9	0

	More Than One Race	Native Hawaiian or Other Pacific Islander
Female	8	1
Male	2	1

	Unknown or Not Reported	White
Female	10	27
Male	4	13

Q7.) average age of wP individuals and average age of aP individuals and are they significantly different

```
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)
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

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

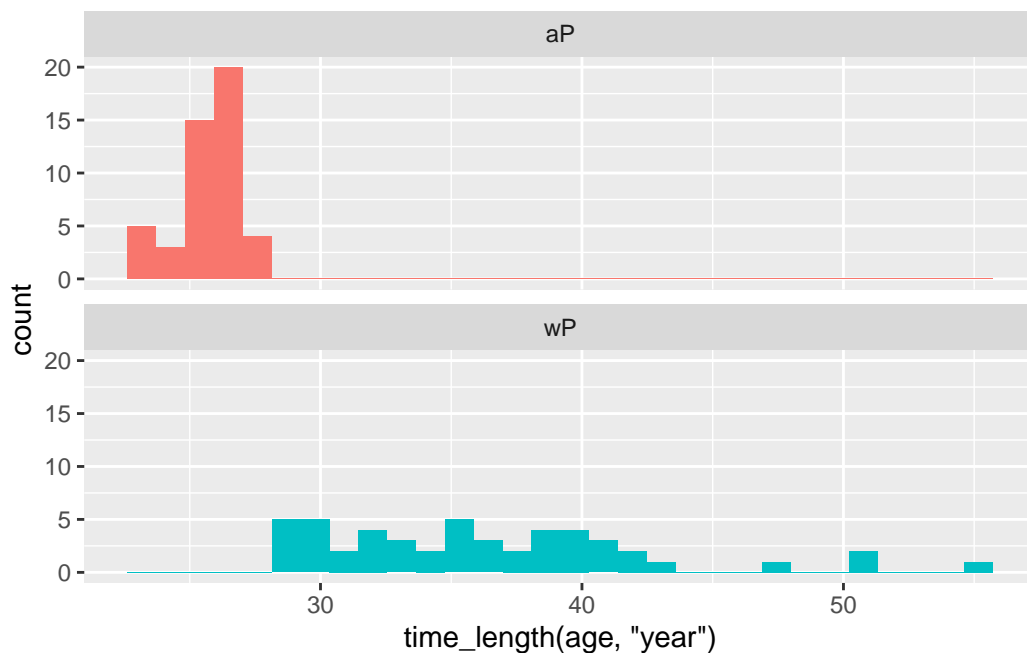
```
int <- ymd(subject$date_of_boost) - ymd(subject$year_of_birth)
age_at_boost <- time_length(int, "year")
head(age_at_boost)
```

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

Q9.) Are these two groups statistically different in terms of average age

```
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`.



Graph shows that they are very significantly different since one is skewed to the left while the wP is more spread out

```
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)
```

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
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 13586 days
```

```
2 13586 days
3 13586 days
4 13586 days
5 13586 days
6 13586 days
```

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
```

Q12.)

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

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

```
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

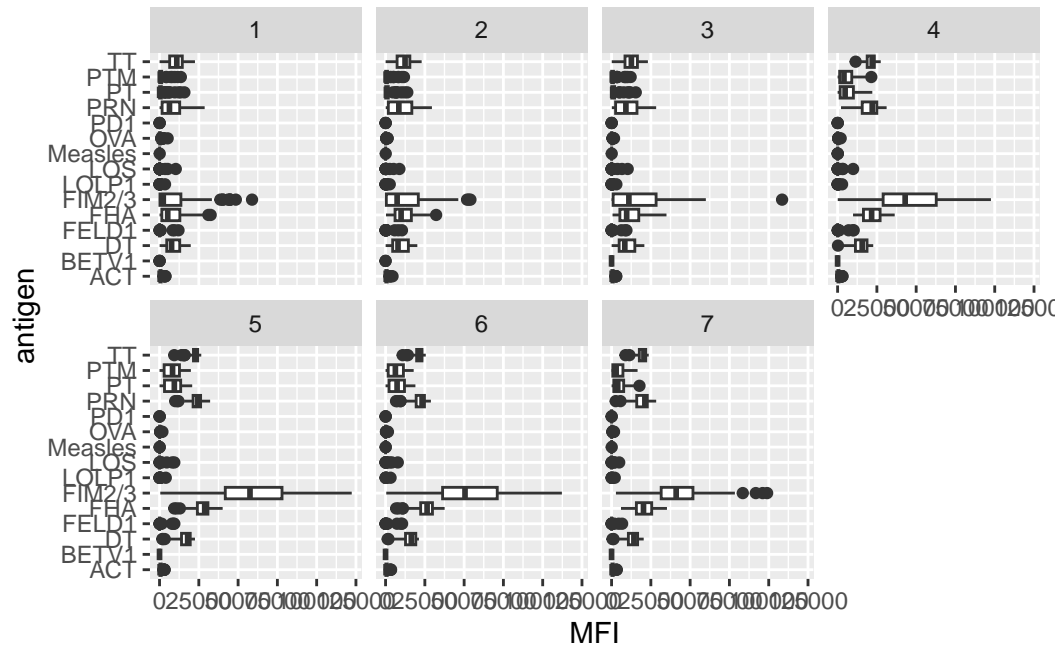
  

	age
1	13586 days
2	13586 days
3	13586 days
4	13586 days
5	13586 days
6	13586 days

Q13.)

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

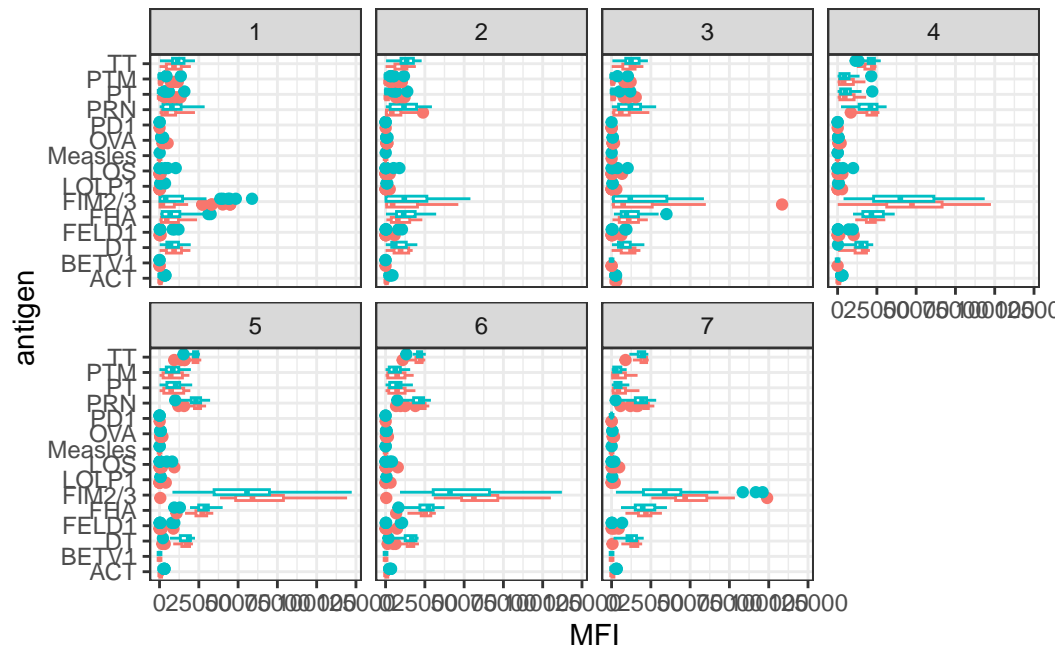




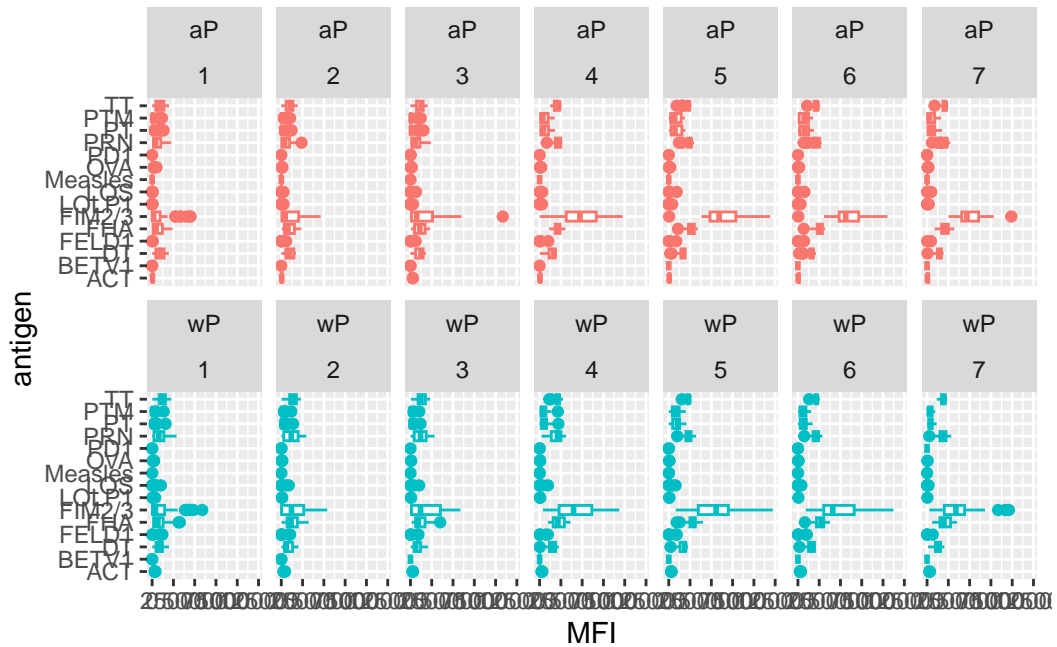
Q14.)

FIM 2/3 show differences in the level of igG1 antibody titers recognizing them over time and FHA average increases over time.

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

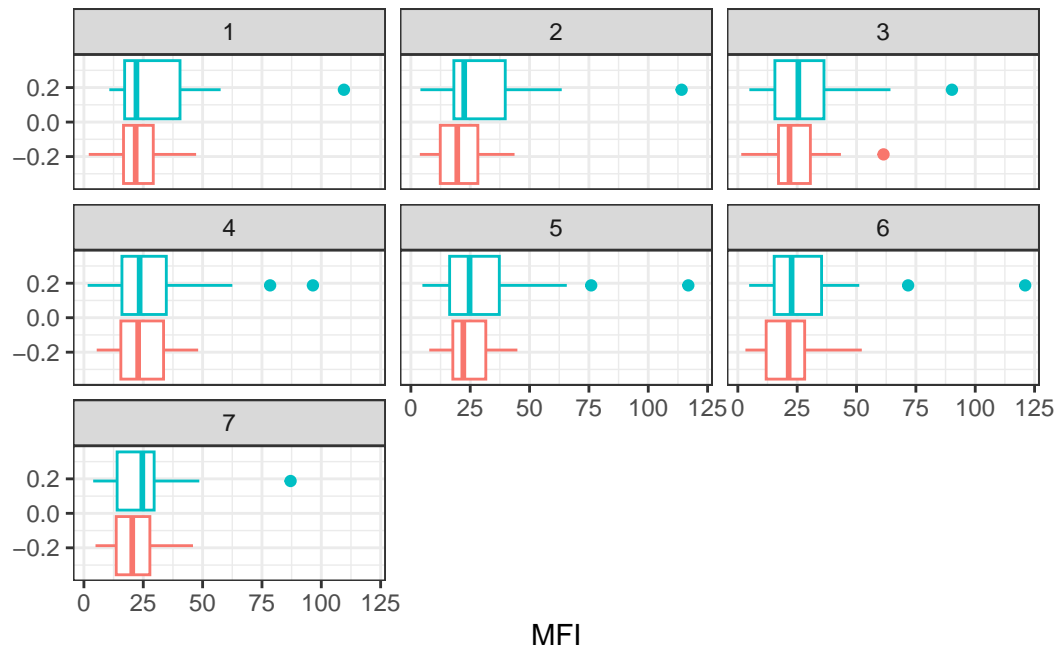


```
ggplot(ig1) +
  aes(MFI, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(infancy_vac, visit), nrow=2)
```

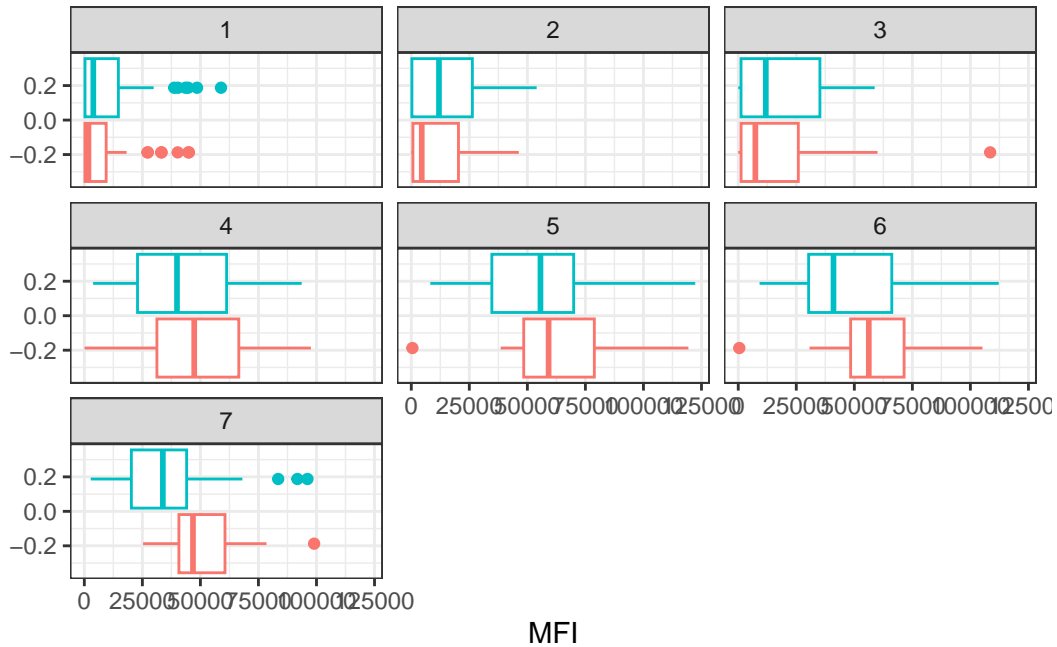


Q15.)

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



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



Q16.)

FIM 2/3 increases over time and then peaks at visit 5 and then declines after.

Q17.)

aP seems to have a higher detection antigen levels after visit 4 than wP which is more pronounced as visits increase.

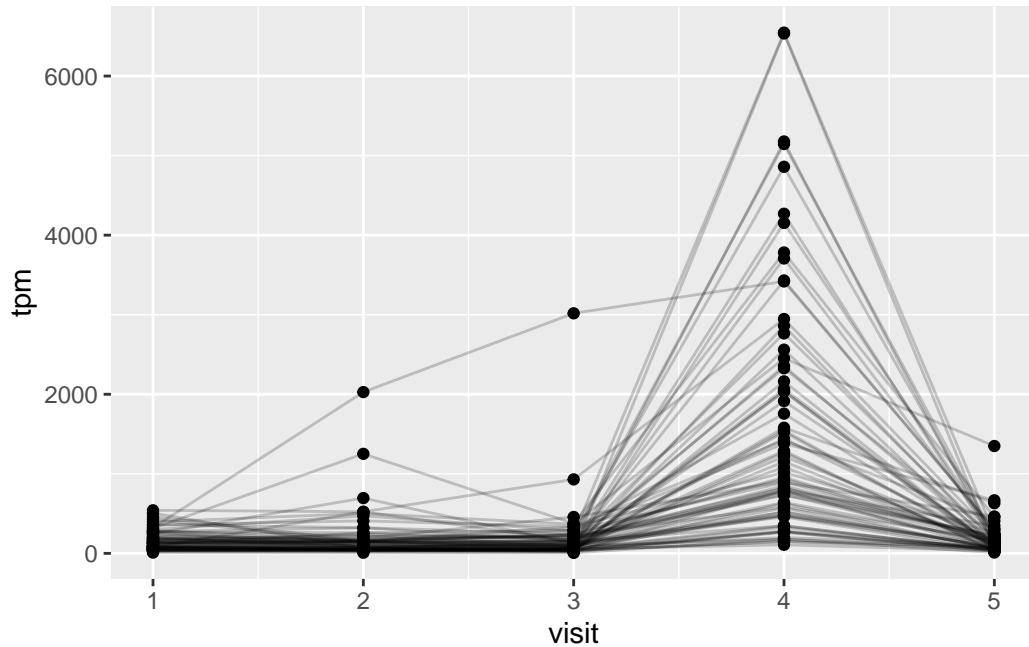
```
url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_gene_id=eq.ENSOG00000211896."

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



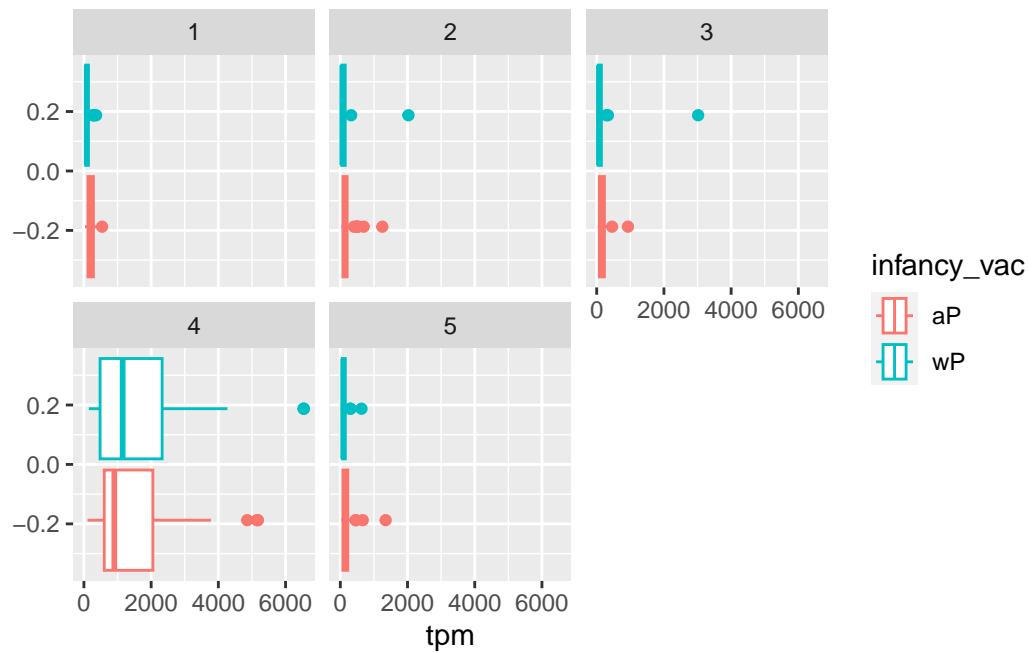
Q19.)

Expression peaks around Visit 4.

Q20.)

No this does not match the titer data since antigen has a long shelf life meaning that continuous gene expression is not needed in order to provide the necessary amount of antigen levels as they also peak at different times and that gene expression leads to the antigen so it makes sense that expression would peak before antigen.

```
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()
```

