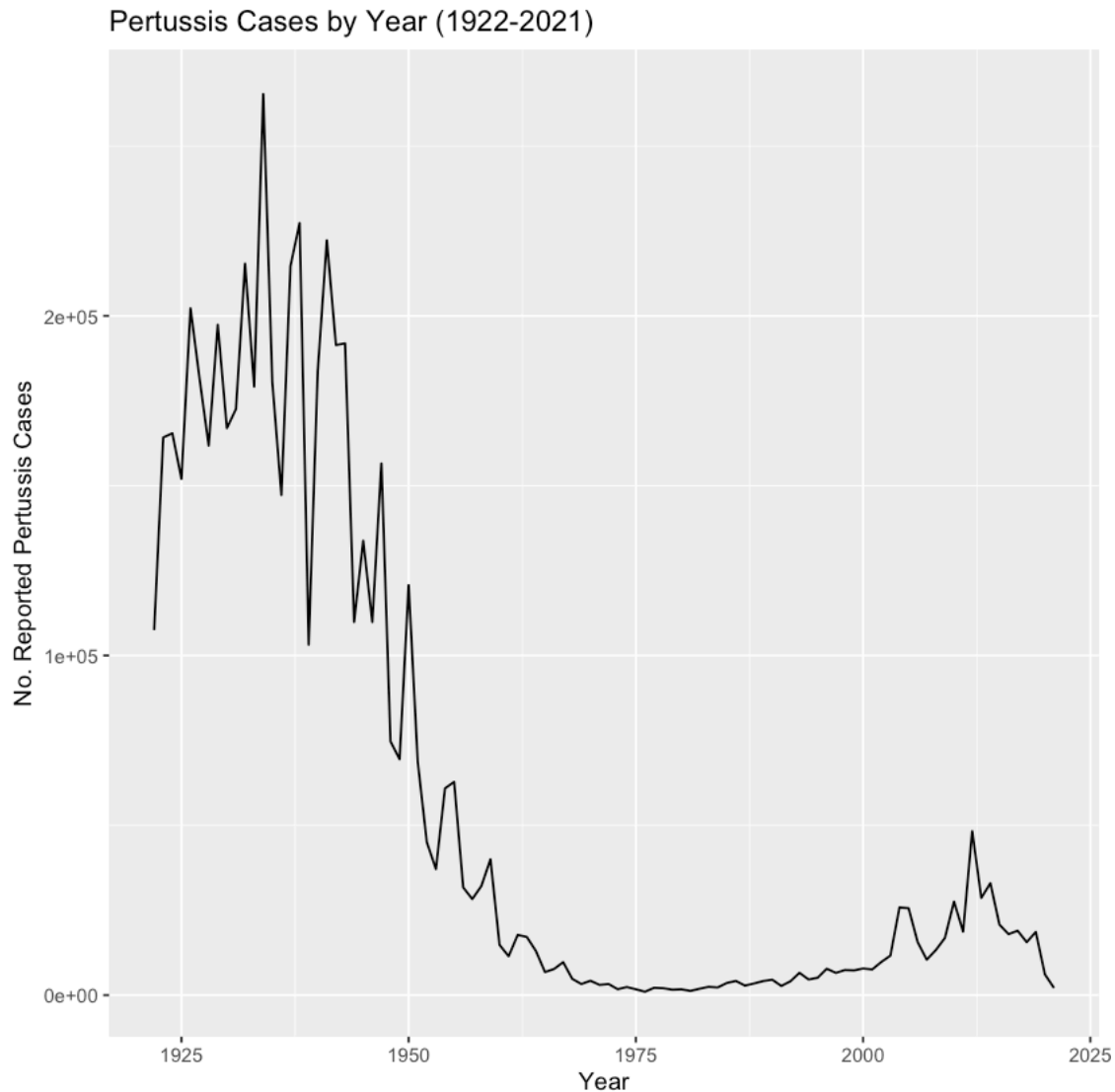


# 18\_miniproject

Rosa Chavez (PID: A16487039)

```
[1]: library(ggplot2)
cdc <- read.csv("Pertussis_Cases_by_Year_1922-2021.csv")
names(cdc) <- c("Year", "Cases")

ggplot(cdc, aes(x = Year, y = Cases)) +
  geom_line() +
  labs(title = "Pertussis Cases by Year (1922-2021)",
       x = "Year",
       y = "No. Reported Pertussis Cases")
```



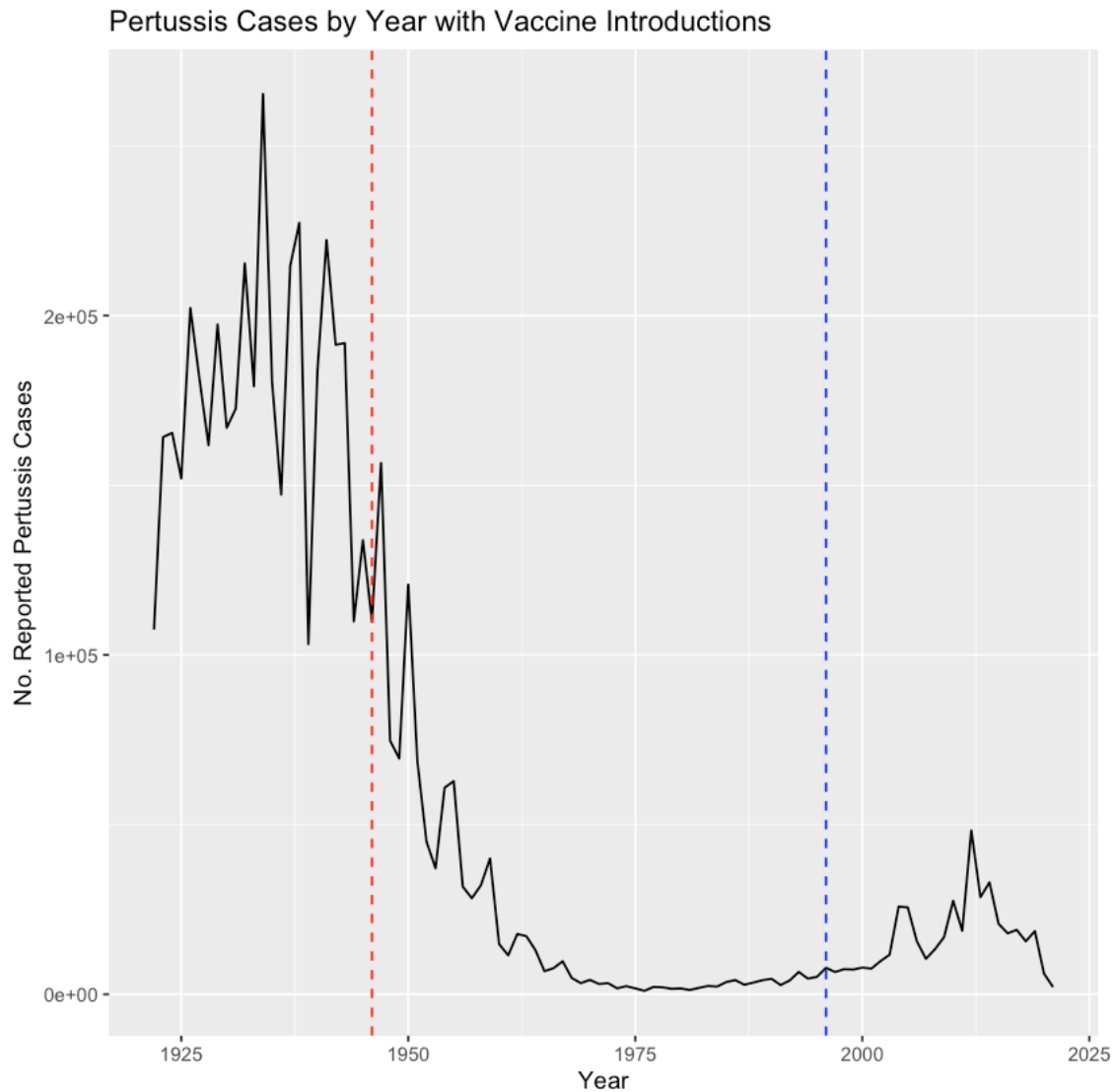
Q2.

I noticed that cases dropped significantly after the first vaccine in 1946 but there was a resurgence after the new type of vaccine in 1996. This is possible evidence that the newer vaccine is not effective or as effective.

```
[2]: library(ggplot2)

ggplot(cdc, aes(x = Year, y = Cases)) +
  geom_line() +
  geom_vline(xintercept = 1946, linetype = "dashed", color = "red") +
  geom_vline(xintercept = 1996, linetype = "dashed", color = "blue") +
  labs(title = "Pertussis Cases by Year with Vaccine Introductions",
        x = "Year",
```

```
y = "No. Reported Pertussis Cases")
```



Q3. I noticed that cases dropped significantly after the first vaccine in 1946 but there was a resurgence after the new type of vaccine in 1996. This is possible evidence that the newer vaccine is not effective or as effective.

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

	subject_id	infancy_vac	biological_sex	ethnicity	race	year_of_
	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
A data.frame: 3 × 8	1	wP	Female	Not Hispanic or Latino	White	1986-01-
	2	wP	Female	Not Hispanic or Latino	White	1968-01-
	3	wP	Female	Unknown	White	1983-01-

```
[4]: ap_count <- sum(subject$infancy_vac == "aP")
wp_count <- sum(subject$infancy_vac == "wP")
```

```
[5]: ap_count
```

60

```
[6]: wp_count
```

58

```
[7]: male_count <- sum(subject$biological_sex == "Male")
female_count <- sum(subject$biological_sex == "Female")
```

```
[8]: male_count
```

39

```
[9]: female_count
```

79

```
[10]: 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

Q7. Yes it is significant with a P-value of 6.813505e-19

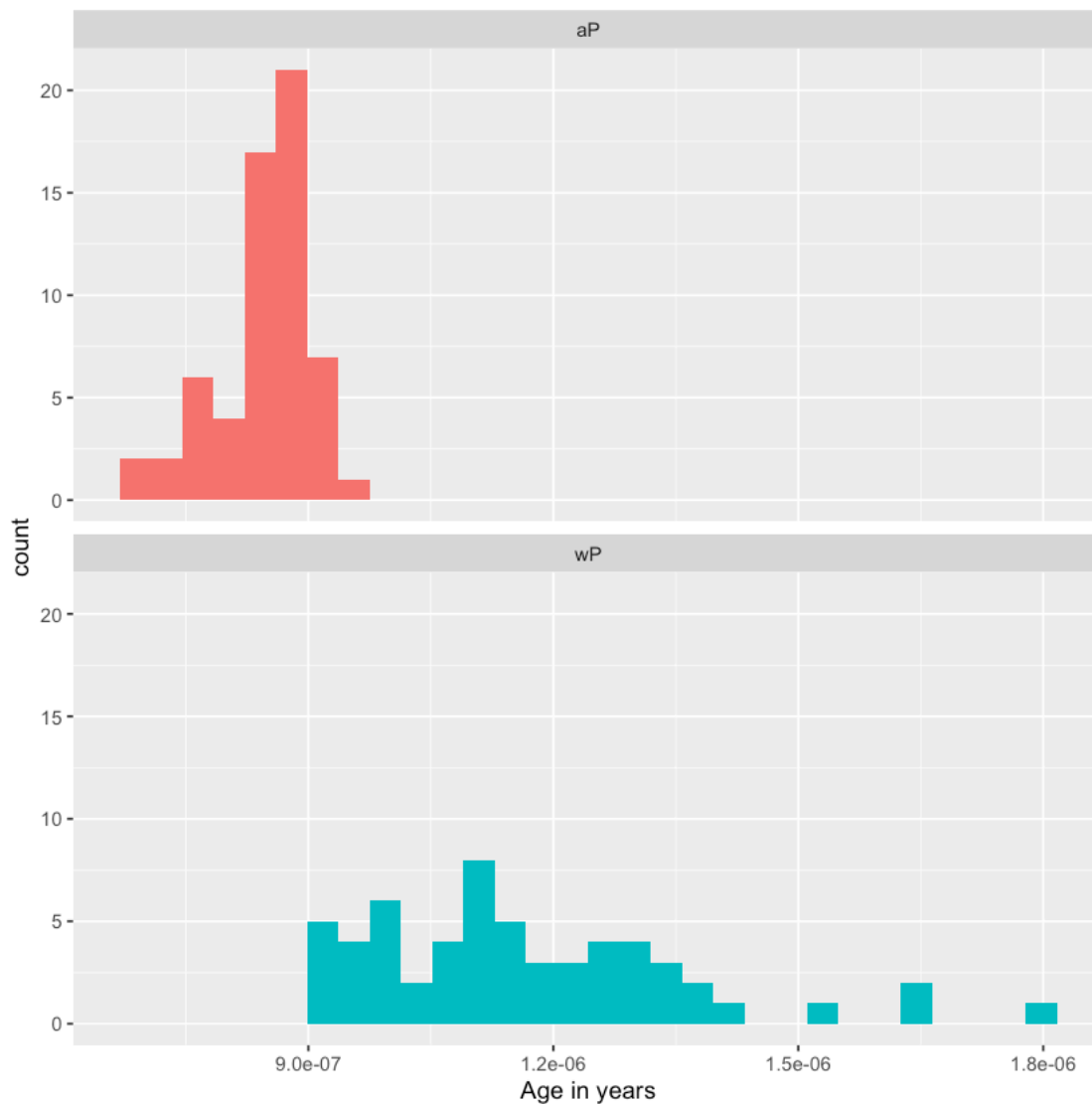
```
[15]: subject$date_of_boost <- ymd(subject$date_of_boost)
subject$age_at_boost <- as.numeric(difftime(subject$date_of_boost,
↪subject$year_of_birth, units = "days")) / 365.25
print(head(subject$age_at_boost))
```

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

Q9. Yes this shows significant difference because the graph shows two different patterns for the ages of people who got the old (wP) versus the new (aP)

```
[16]: ggplot(subject) +
  aes(time_length(age, "year"),
  fill=as.factor(infancy_vac)) +
  geom_histogram(show.legend=FALSE) +
  facet_wrap(vars(infancy_vac), nrow=2) +
  xlab("Age in years")
```

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



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

```
[18]: titer <- read_json("https://www.cmi-pb.org/api/plasma_ab_titer", simplifyVector_
      ↪= TRUE)
```

```
[19]: library(dplyr)
      meta <- left_join(specimen, subject)
      dim(meta)
      head(meta)
```

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

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

1. 939 2. 15

A data.frame: 6 × 15

	specimen_id <int>	subject_id <int>	actual_day_relative_to_boost <int>	planned_day_relative_t <int>
1	1	1	-3	0
2	2	1	1	1
3	3	1	3	3
4	4	1	7	7
5	5	1	11	14
6	6	1	32	30

```
[20]: abdata <- inner_join(titer, meta)
      dim(abdata)
```

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

1. 46906 2. 22

```
[21]: table(abdata$isotype)
```

```

  IgE  IgG IgG1 IgG2 IgG3 IgG4
6698 4255 8983 8990 8990 8990
```

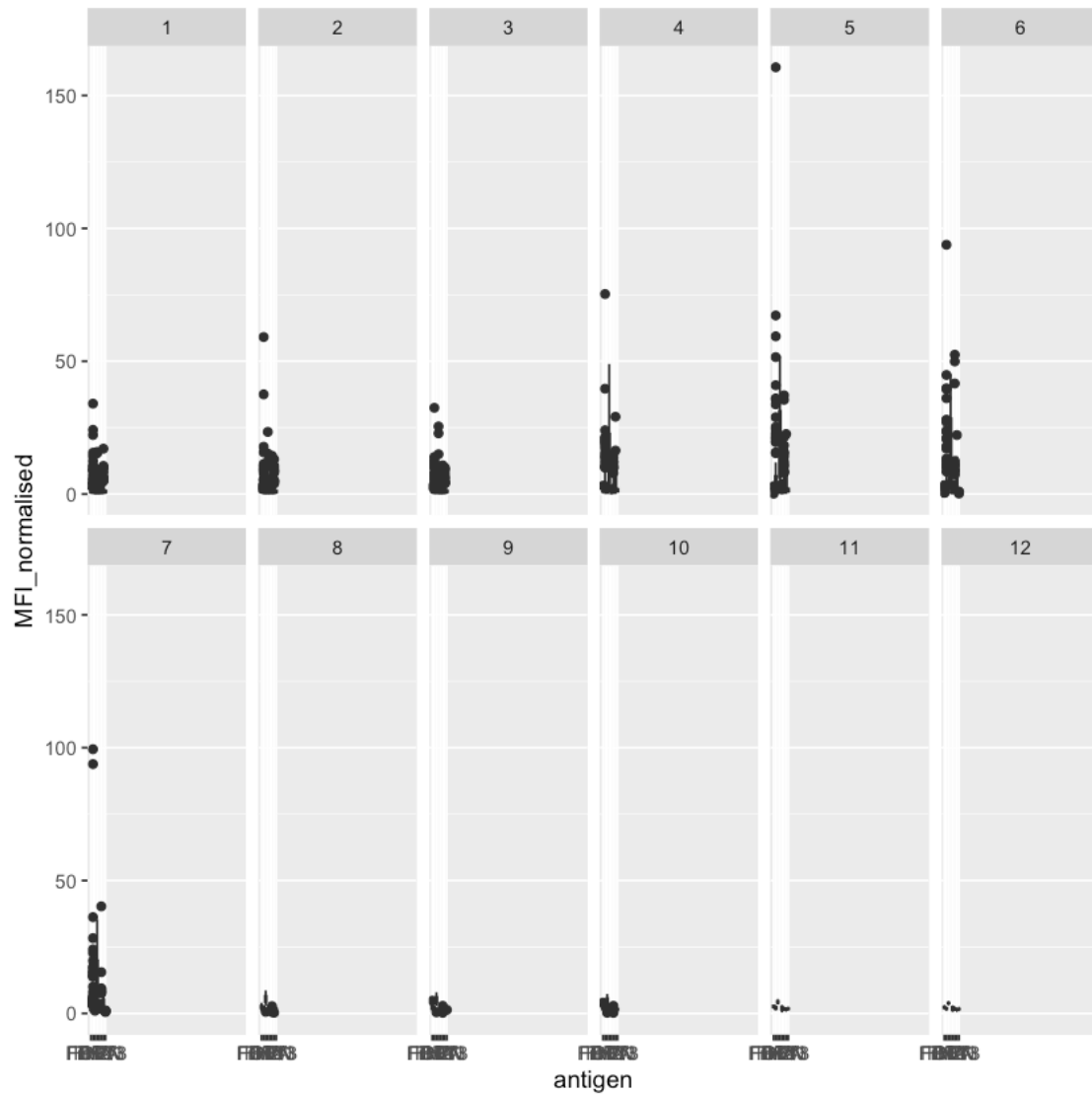
```
[22]: dataset_values <- table(abdata$dataset)
      dataset_values
```

2020_dataset	2021_dataset	2022_dataset
31520	8085	7301

```
[23]: igg <- abdata %>% filter(isotype == "IgG")
      head(igg)
```

		specimen_id <int>	isotype <chr>	is_antigen_specific <lgl>	antigen <chr>	MFI <dbl>	MFI_normalised <dbl>
	1	1	IgG	TRUE	PT	68.56614	3.736992
	2	1	IgG	TRUE	PRN	332.12718	2.602350
A data.frame: 6 × 22	3	1	IgG	TRUE	FHA	1887.12263	34.050956
	4	19	IgG	TRUE	PT	20.11607	1.096366
	5	19	IgG	TRUE	PRN	976.67419	7.652635
	6	19	IgG	TRUE	FHA	60.76626	1.096457

```
[24]: ggplot(igg) +
      aes(x=antigen, y=MFI_normalised) +
      geom_boxplot() +
      coord_cartesian(xlim=c(0, 75)) +
      facet_wrap(~visit, nrow=2)
```



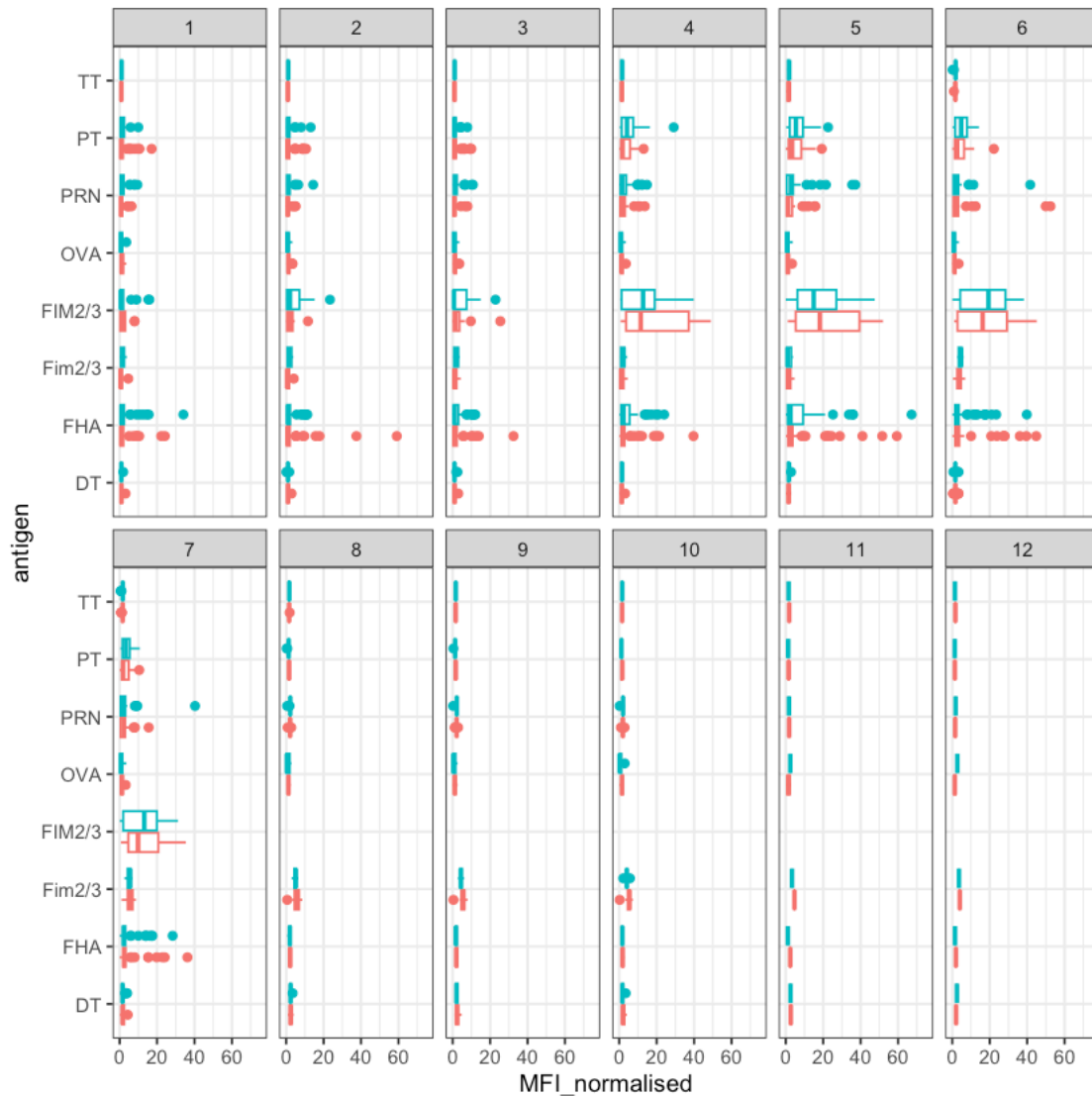
Q14. The antigens that show differences are FIM2/3, FHA, Fim2/3, PT, PRN, DT, TT and OVA

```
[25]: 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 message:

"Removed 5 rows containing non-finite outside the scale range  
(`stat\_boxplot()`)."

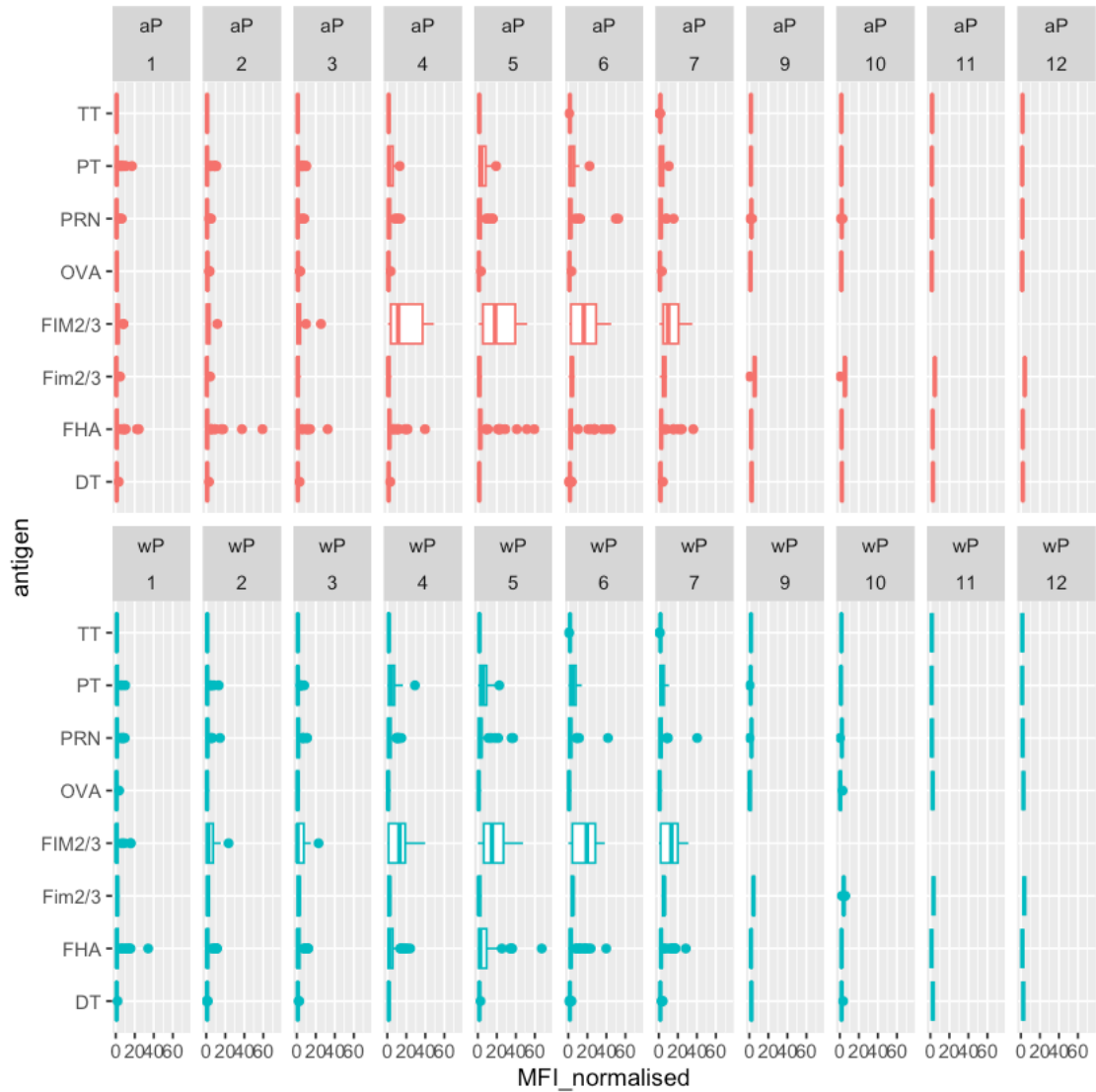




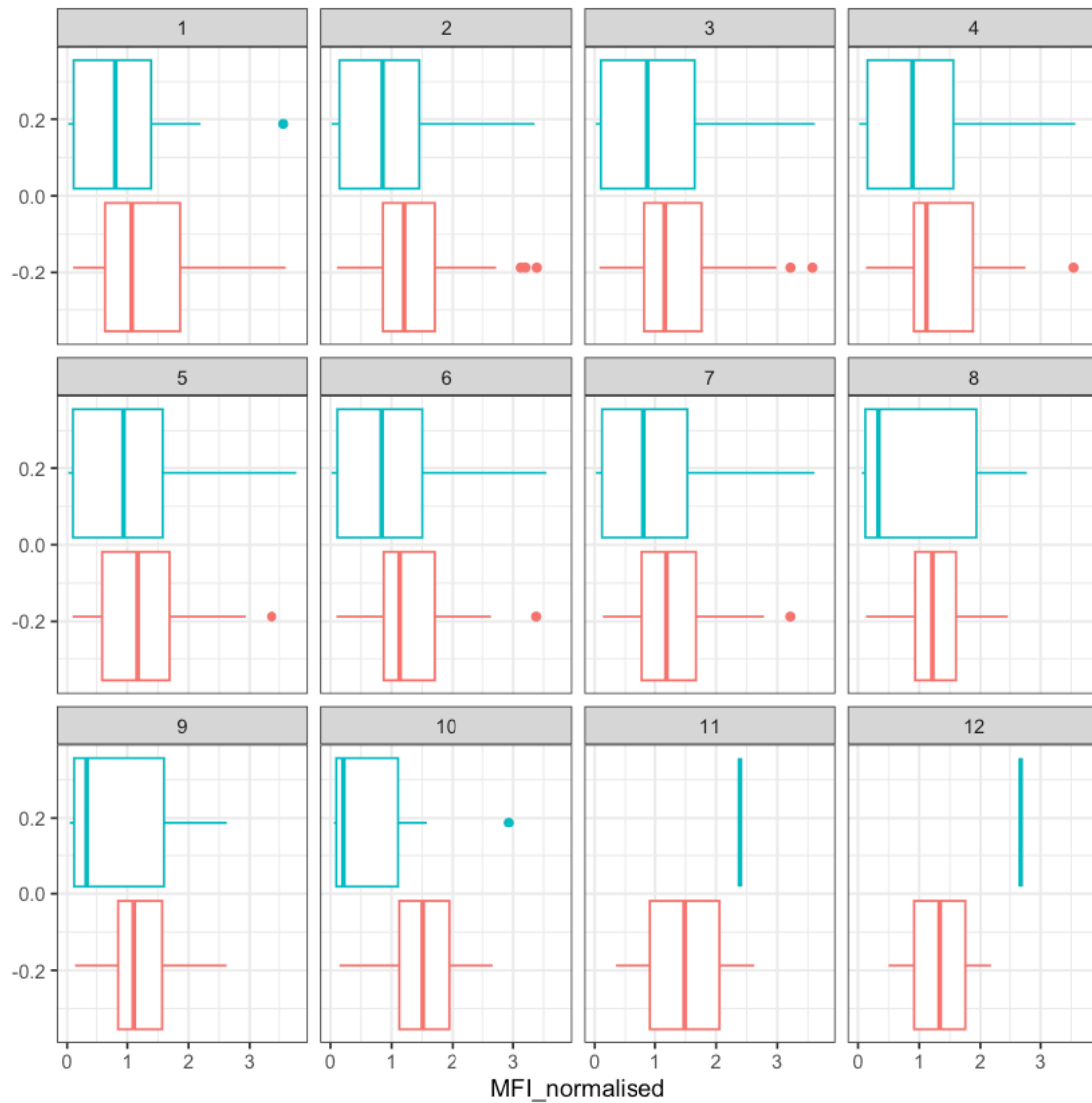
```
[26]: igg %>% filter(visit != 8) %>%
  ggplot() +
  aes(MFI_normalised, antigen, col=infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  xlim(0,75) +
  facet_wrap(vars(infancy_vac, visit), nrow=2)
```

Warning message:

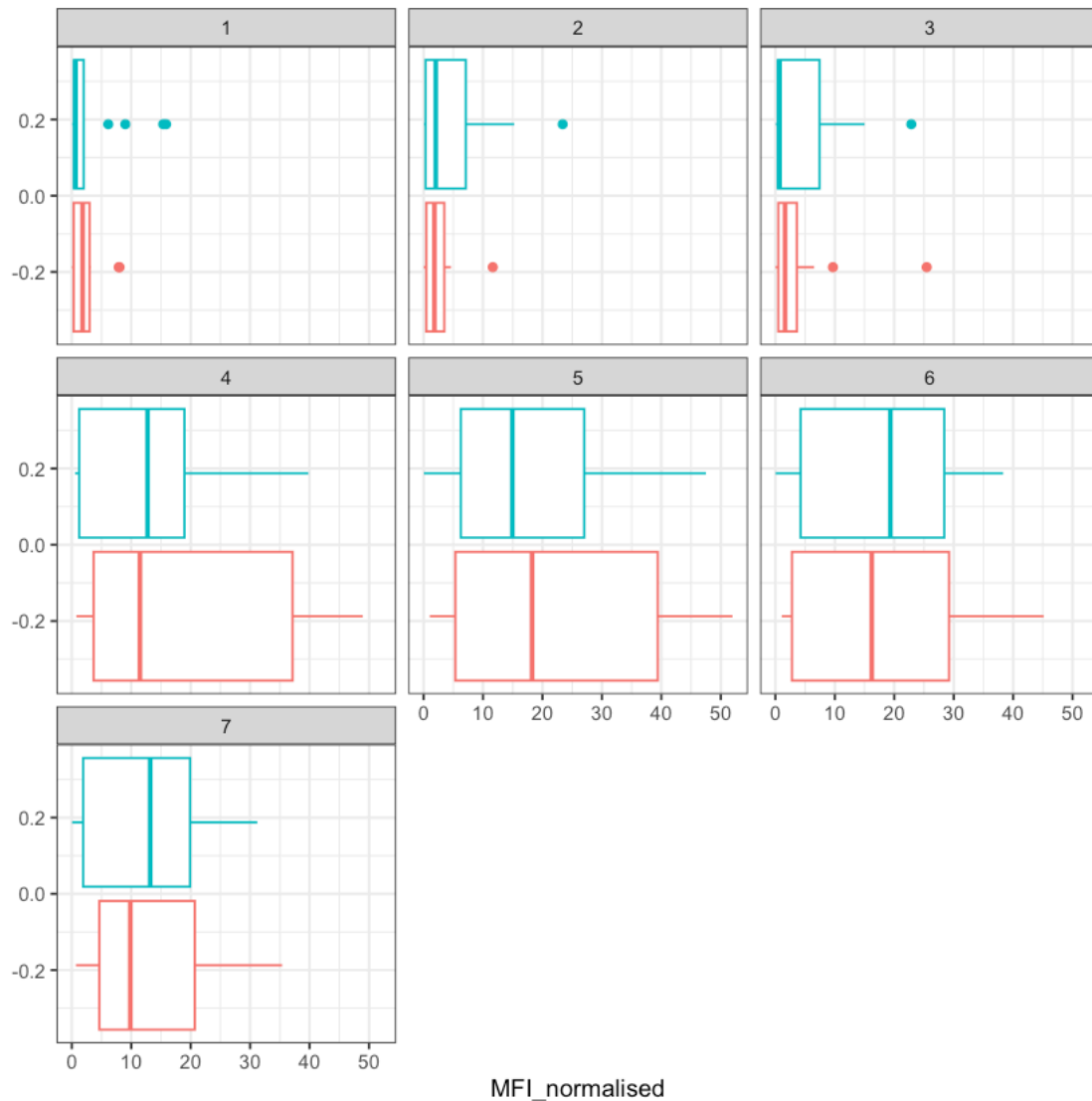
"Removed 5 rows containing non-finite outside the scale range  
(`stat\_boxplot()`)."



```
[27]: filter(igg, antigen=="OVA") %>%
  ggplot() +
  aes(x=MFI_normalised, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



```
[28]: filter(igg, antigen=="FIM2/3") %>%
  ggplot() +
  aes(x=MFI_normalised, col=infancy_vac) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



Q16. PT & FIM2/3 antigens show a clear trend and antibody levels rise after vaccination in 5, which is showing how the vaccine does a good job a boosting the body's immune response.

Q17. The immune responses to the PT & FIM2/3 antigens in people who got either the wP or aP vaccine show differences in how strong the response is at different times, but they do not favor one vaccine over the other which suggest that both vaccines are similarly effective at triggering the immune system.

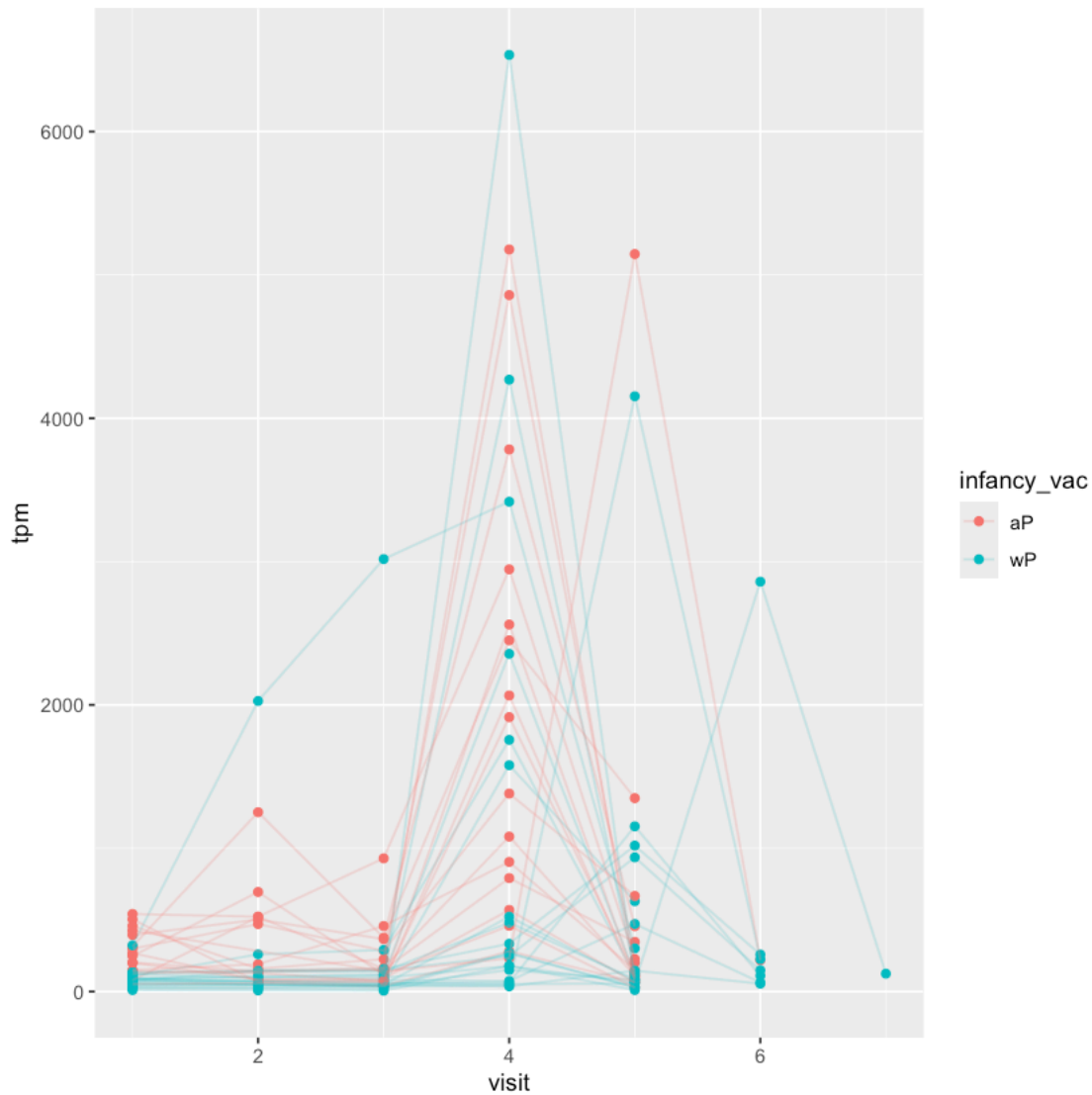
Q18. Yes the trends look similar for both datasets

```
[29]: url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_gene_id=eq.
      ↪ENSG00000211896.7"
      rna <- read_json(url, simplifyVector = TRUE)
```

```
[30]: #meta <- inner_join(specimen, subject)
      ssrna <- inner_join(rna, meta)
```

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

```
[31]: ggplot(ssrna) +
      aes(x=visit, y=tpm, color=infancy_vac, group=subject_id) +
      geom_point() +
      geom_line(alpha=0.2)
```



Q20. The peak mentioned earlier shows the body's reaction to the vaccine is strongest at that point and it rapidly decreases after the peak which is typical as the immediate response to the vaccination wears off

Q21. Yes the pattern in time match the trend of antibody titer data