Class 18: Pertussis mini project

Jihyun In

Background

Pertussis (aka whoppping cough) is a common lung infection cuased by the bacteria *B. pertussis.* The CDC track s cases of pertussis in the US.

https://tinyurl.com/pertussiscdc

Examining cases of pertussis by year

We can use the `datapasta` package to scrape case numbers from the CDC website.

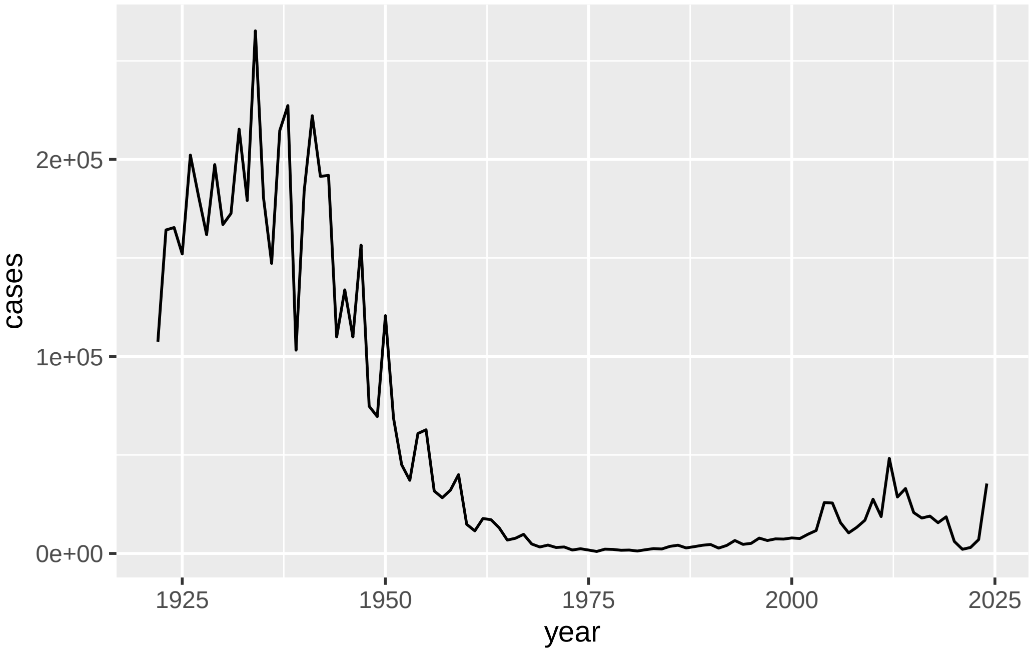
```r
head(cdc)
```

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

> Q. Make a plot of pertussis cases per year using ggplot

```r
library(ggplot2)
```

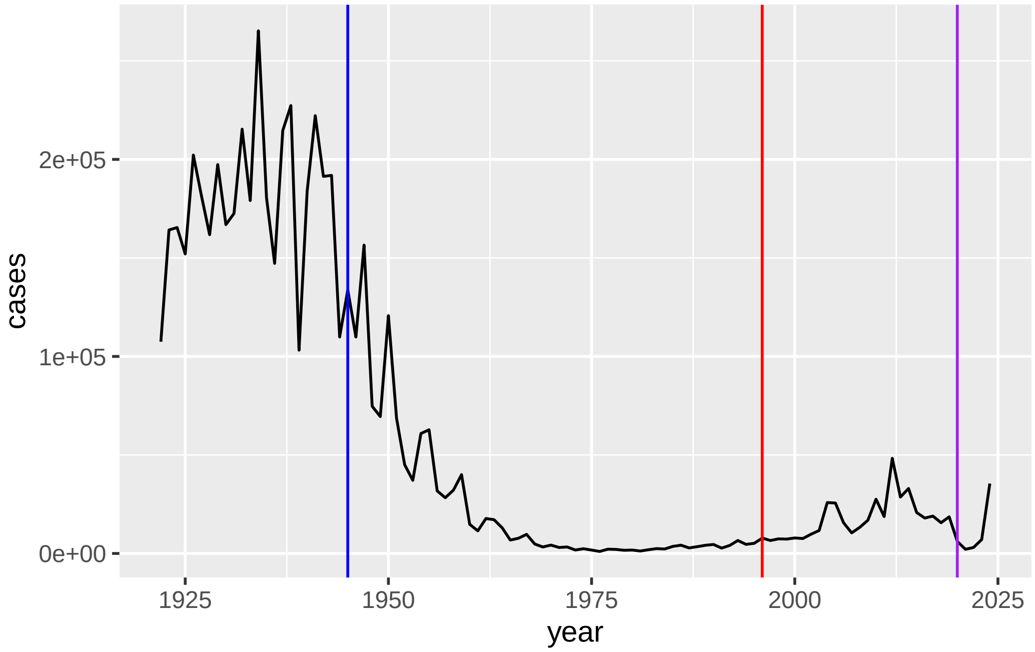
```r
cases <- ggplot(cdc) +
  aes(year, cases) +
  geom_line()
cases
```



Q2. Add some key time points in our history of interaction with pertussis.These include wP roll-out (the first vaccine) in 1945 and the switch to aP in 1996.

We can use `geom_vline()`

```
cases <- cases +  geom_vline(xintercept=1945, col="blue")  + geom_vline(xintercept=1996, col=
cases
```



Mounting evidence suggests tha the newer **aP** vaccine is less effective over the long term than the older **wP** vaccine that it replaced. In other words, the vaccine efficacy wane smore rapidly with aP than with wP.

Enter the CMI-PB project

CMI-PB (computational models of immunity - pertussis boost)'s major goal is to investigate how the immune system responds differently to aP vs. wP vaccinated individuals and be able to predict this

CMI-PB makes all their collected data freely avialbe and they store it in a databased composed different tables. Here we will access a few of these.

```r
library(jsonlite)

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
```

How many subjects(i.e. enrolled people are there)

```r
nrow(subject)
```

```
[1] 172
```

how many ap and wp subjects are there?

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

```

aP wP 
87 85 
```

Q. How many male/female are in the dataset

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

```

Female   Male 
   112     60 
```

Q. how about gender and race

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

```
                                           Female Male
  American Indian/Alaska Native                 0    1
  Asian                                        32   12
  Black or African American                     2    3
  More Than One Race                           15    4
  Native Hawaiian or Other Pacific Islander     1    1
  Unknown or Not Reported                      14    7
  White                                        48   32
```

> Q. Is this representative of the US population?

No. It’s more representative of UCSD students.

Let’s read soe other database table from CMI

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

We want to join these tables to get all our information

```
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
```

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

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

```r
dim(meta)
```

```
[1] 1503   13
```

one more join

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

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

```r
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 infancy_vac biological_sex
1 UG/ML                 2.096133          1          wP         Female
2 IU/ML                29.170000          1          wP         Female
3 IU/ML                 0.530000          1          wP         Female
4 IU/ML                 6.205949          1          wP         Female
5 IU/ML                 4.679535          1          wP         Female
6 IU/ML                 2.816431          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
  actual_day_relative_to_boost planned_day_relative_to_boost specimen_type
1                           -3                             0         Blood
2                           -3                             0         Blood
3                           -3                             0         Blood
4                           -3                             0         Blood
5                           -3                             0         Blood
6                           -3                             0         Blood
```

```
  visit
1     1
2     1
3     1
4     1
5     1
6     1
```

```
dim(abdata)
```

```
[1] 52576    20
```

Q. How many Ab isotypes are there in the dataset?

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

```

  IgE   IgG  IgG1  IgG2  IgG3  IgG4 
 6698  5389 10117 10124 10124 10124 
```

How many differen tantigens are measured in the dataset?

```
table(abdata$antigen)
```

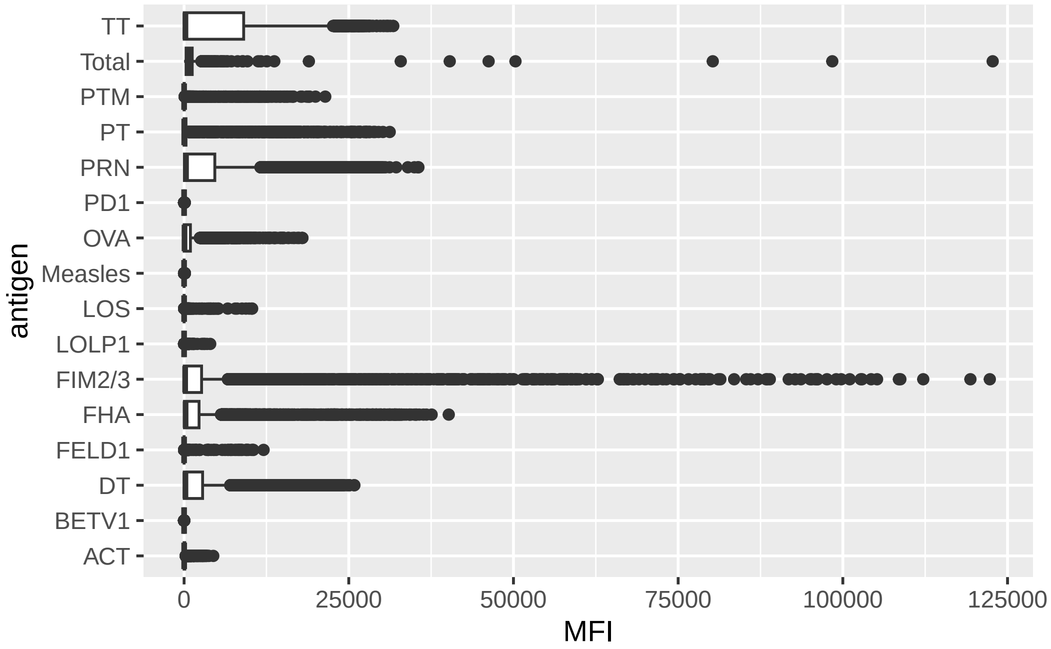
```

    ACT   BETV1      DT   FELD1     FHA  FIM2/3   LOLP1     LOS Measles     OVA 
   1970    1970    4978    1970    5372    4978    1970    1970    1970    4978 
    PD1     PRN      PT     PTM   Total      TT 
   1970    5372    5372    1970     788    4978 
```

boxplot

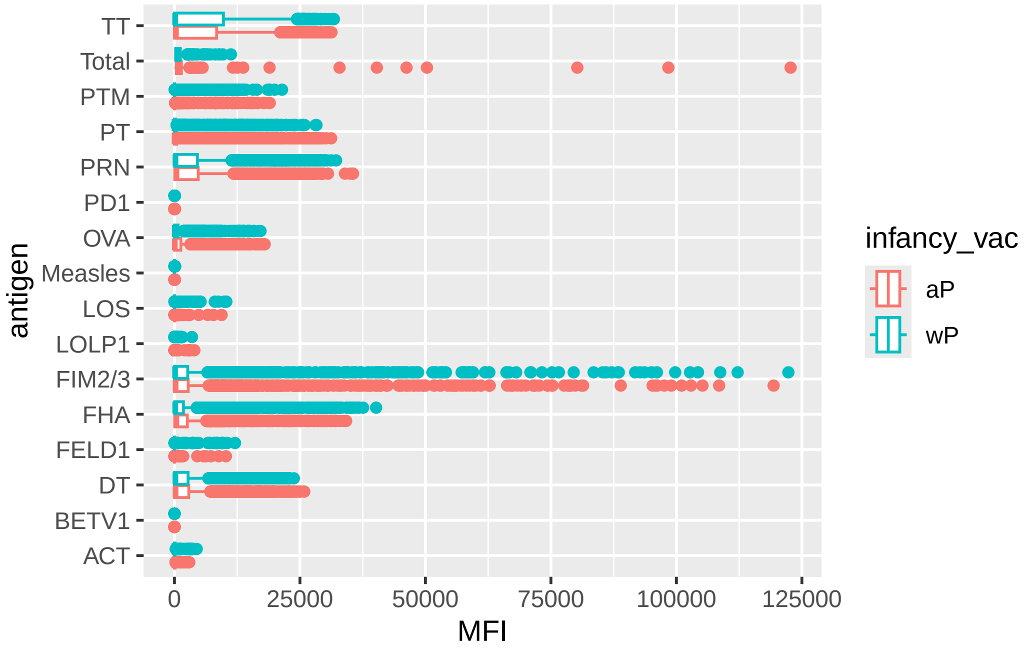
```
ggplot(abdata) + aes(MFI, antigen ) + geom_boxplot()
```

```
Warning: Removed 1 row containing non-finite outside the scale range
(`stat_boxplot()`).
```



```r
ggplot(abdata) + aes(MFI, antigen, col = infancy_vac) + geom_boxplot()
```

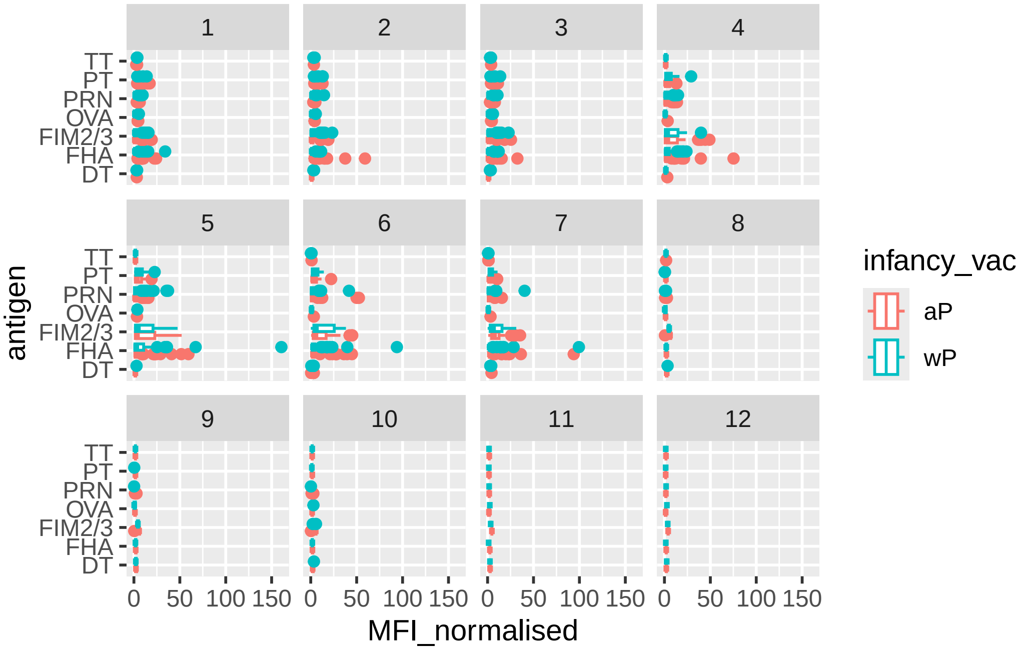
```
Warning: Removed 1 row containing non-finite outside the scale range
(`stat_boxplot()`).
```



```r
igg <- abdata |> filter(isotype == "IgG")
head(igg)
```

```
  specimen_id isotype is_antigen_specific antigen        MFI MFI_normalised
1           1     IgG                TRUE      PT   68.56614       3.736992
2           1     IgG                TRUE     PRN  332.12718       2.602350
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
   unit lower_limit_of_detection subject_id infancy_vac biological_sex
1 IU/ML                 0.530000          1          wP         Female
2 IU/ML                 6.205949          1          wP         Female
3 IU/ML                 4.679535          1          wP         Female
4 IU/ML                 0.530000          3          wP         Female
5 IU/ML                 6.205949          3          wP         Female
6 IU/ML                 4.679535          3          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                Unknown White    1983-01-01    2016-10-10 2020_dataset
5                Unknown White    1983-01-01    2016-10-10 2020_dataset
6                Unknown White    1983-01-01    2016-10-10 2020_dataset
  actual_day_relative_to_boost planned_day_relative_to_boost specimen_type
1                           -3                             0         Blood
2                           -3                             0         Blood
3                           -3                             0         Blood
4                           -3                             0         Blood
5                           -3                             0         Blood
6                           -3                             0         Blood
  visit
1     1
2     1
3     1
4     1
5     1
6     1
```

```r
ggplot(igg) + aes(MFI_normalised, antigen, col = infancy_vac) + geom_boxplot() + facet_wrap(
```



Focus in further just one of these - let’s pick PT (pertussis toxin)

```r
table(igg$dataset)
```

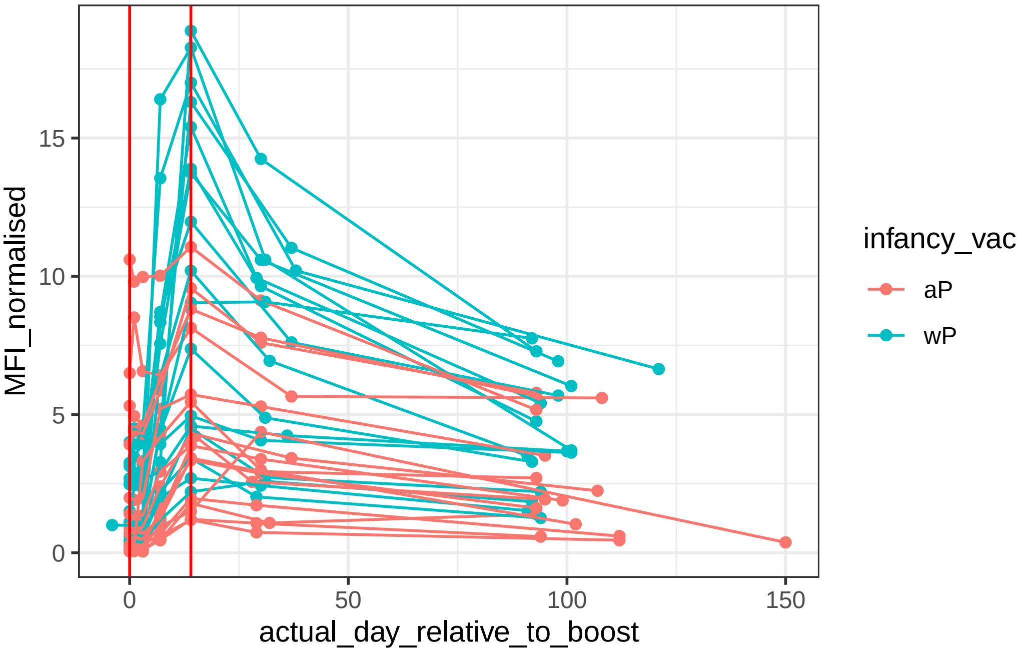
```
2020_dataset 2021_dataset 2022_dataset 2023_dataset 
        1182         1617         1456         1134 
```

```r
pt_igg <- abdata |>
  filter(isotype == "IgG",
         antigen == "PT",
         dataset == "2021_dataset")
```

```r
dim(pt_igg)
```

```
[1] 231  20
```

```r
ggplot(pt_igg) + aes(actual_day_relative_to_boost, MFI_normalised, col = infancy_vac, group=
```


MFI_normalised
0
5
10
15
actual_day_relative_to_boost
0
50
100
150
infancy_vac
aP
wP

p