

class19

Steven Gan

2022-12-03

Table of contents

| | |
|---|-----------|
| 1. Investigating pertussis cases by year | 1 |
| 2. A tale of two vaccines (wP & aP) | 3 |
| 3. Exploring CMI-PB data | 4 |
| The CMI-PB API returns JSON data | 4 |
| Side-Note: Working with dates | 6 |
| Joining multiple tables | 8 |
| 4. Examine IgG1 Ab titer levels | 9 |
| 5. Obtaining CMI-PB RNASeq data | 18 |
| 6. Working with larger datasets | 20 |
| Working with long format data | 21 |
| Convert to “wide” format | 21 |
| Filtering | 21 |
| PCA | 22 |

1. Investigating pertussis cases by year

Q1:

```
# install.packages("datapasta")
library(datapasta)
```

```

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)
)

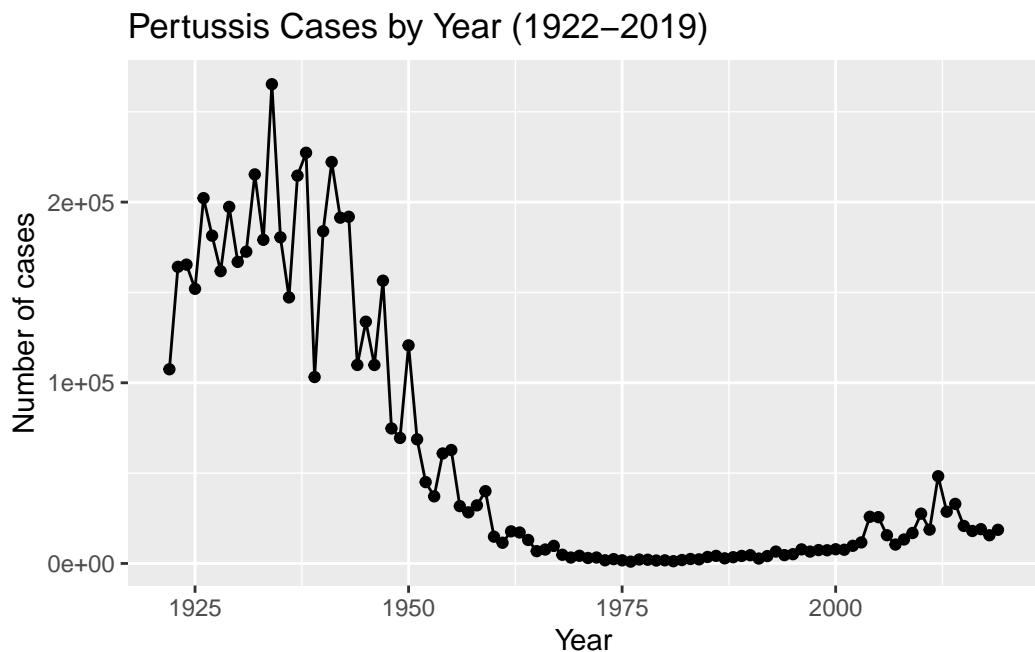
```

```

library(ggplot2)

p <- ggplot(cdc) +
  aes(Year, No..Reported.Pertussis.Cases) +
  geom_point() +
  geom_line() +
  labs(x = "Year", y = "Number of cases",
       title = "Pertussis Cases by Year (1922–2019)")
p

```



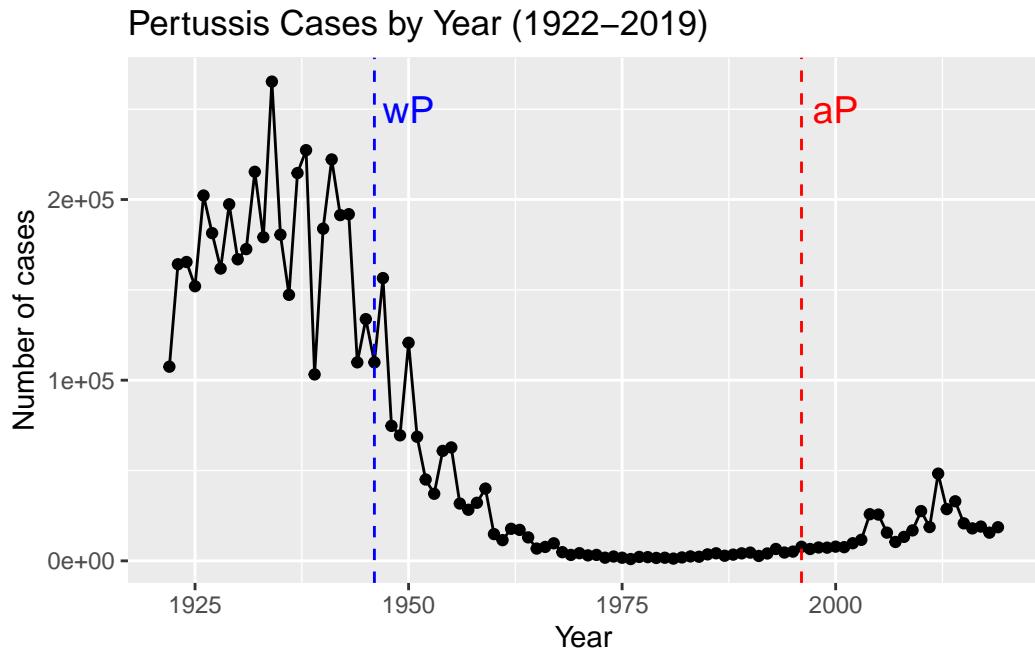
2. A tale of two vaccines (wP & aP)

Q2:

```

p + geom_vline(xintercept = 1946, linetype = "dashed", col = "blue") +
  geom_vline(xintercept = 1996, linetype = "dashed", col = "red") +
  annotate(geom = "text", x = 1950, y = 250000, label = "wP",
           col = "blue", size = 5) +
  annotate(geom = "text", x = 2000, y = 250000, label = "aP",
           col = "red", size = 5)

```



After the application of wP, number of cases dramatically decreased to approaches 0, indicating the robust effect of vaccination. After the change of wP to aP, number of cases slightly increased, but still significantly lower than without vaccination.

Q3:

Number of cases increased slightly after introducing aP. A possible explanation is the effect of aP is weaker than wP, since aP is made from acellular antigen with weakened toxin.

3. Exploring CMI-PB data

The CMI-PB API returns JSON data

```
# install.packages("jsonlite")
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

```

Q4:

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

```
aP wP
47 49
```

Q5:

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

```
Female   Male
66       30
```

Q6:

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

| | American Indian/Alaska Native | Asian | Black or African American | Other |
|--------|-------------------------------|---|---------------------------|-------|
| 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 |

Side-Note: Working with dates

```

library(lubridate)

today()

[1] "2022-12-04"

today() - ymd("2001-02-11")

Time difference of 7966 days

time_length(today() - ymd("2005-02-27"), "years")

[1] 17.76591

Q7:
(i)

library(dplyr)

subject$age <- time_length(today() - ymd(subject$year_of_birth), "years")

subject %>%
  filter(infancy_vac == "wP") %>%
  summarise(Mean = mean(age)) %>%
  round()

Mean
1    36

(ii)

```

```

subject %>%
  filter(infancy_vac == "aP") %>%
  summarise(Mean = mean(age)) %>%
  round()

```

Mean
1 25

(iii)

```

t.test(filter(subject, infancy_vac == "wP")$age,
       filter(subject, infancy_vac == "aP")$age)

```

Welch Two Sample t-test

```

data: filter(subject, infancy_vac == "wP")$age and filter(subject, infancy_vac == "aP")$age
t = 12.092, df = 51.082, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 9.044045 12.644857
sample estimates:
mean of x mean of y
 36.08627 25.24182

```

Q8:

```

time_length(ymd(subject$date_of_boost) - ymd(subject$year_of_birth), "years")

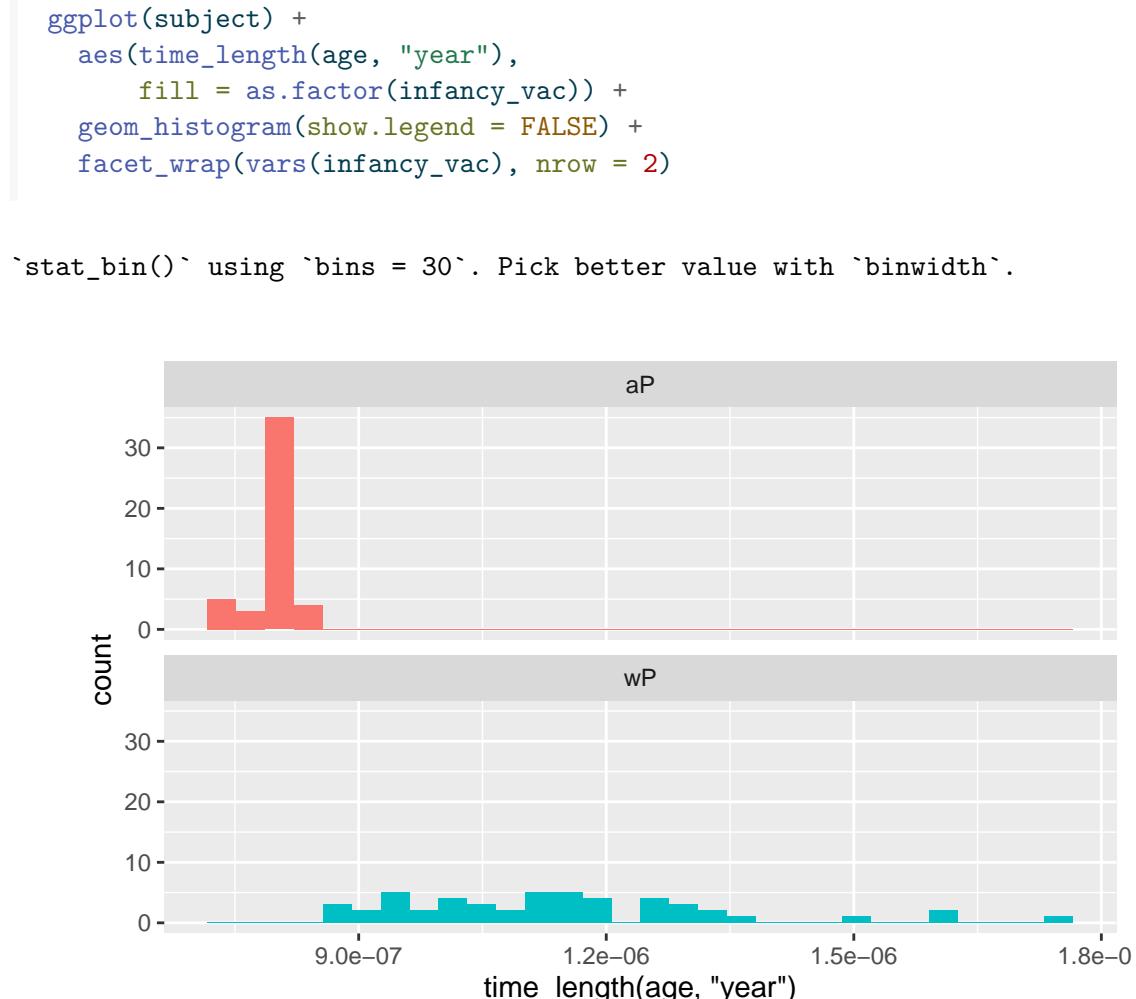
```

```

[1] 30.69678 51.07461 33.77413 28.65982 25.65914 28.77481 35.84942 34.14921
[9] 20.56400 34.56263 30.65845 34.56263 19.56194 23.61944 27.61944 29.56331
[17] 36.69815 19.65777 22.73511 32.26557 25.90007 23.90144 25.90007 28.91992
[25] 42.92129 47.07461 47.07461 29.07324 21.07324 21.07324 28.15058 24.15058
[33] 24.15058 21.14990 21.14990 31.20876 26.20671 32.20808 27.20876 26.20671
[41] 21.20739 20.26557 22.26420 19.32375 21.32238 19.32375 19.32375 22.41752
[49] 20.41889 21.41821 19.47707 23.47707 20.47639 21.47570 19.47707 35.65777
[57] 33.65914 31.65777 25.73580 24.70089 28.70089 33.73580 19.73443 34.73511
[65] 19.73443 28.73648 27.73443 19.81109 26.77344 33.81246 25.77413 19.81109
[73] 18.85010 19.81109 31.81109 22.81177 31.84942 19.84942 18.85010 18.85010
[81] 19.90691 18.85010 20.90897 19.04449 20.04381 19.90691 19.90691 19.00616
[89] 19.00616 20.04381 20.04381 20.07940 21.08145 20.07940 20.07940 20.07940

```

Q9:



They are significantly different

Joining multiple tables

```
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, by = "subject_id"

dim(meta)

[1] 729 14
```

Q10:

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

Joining, 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
```

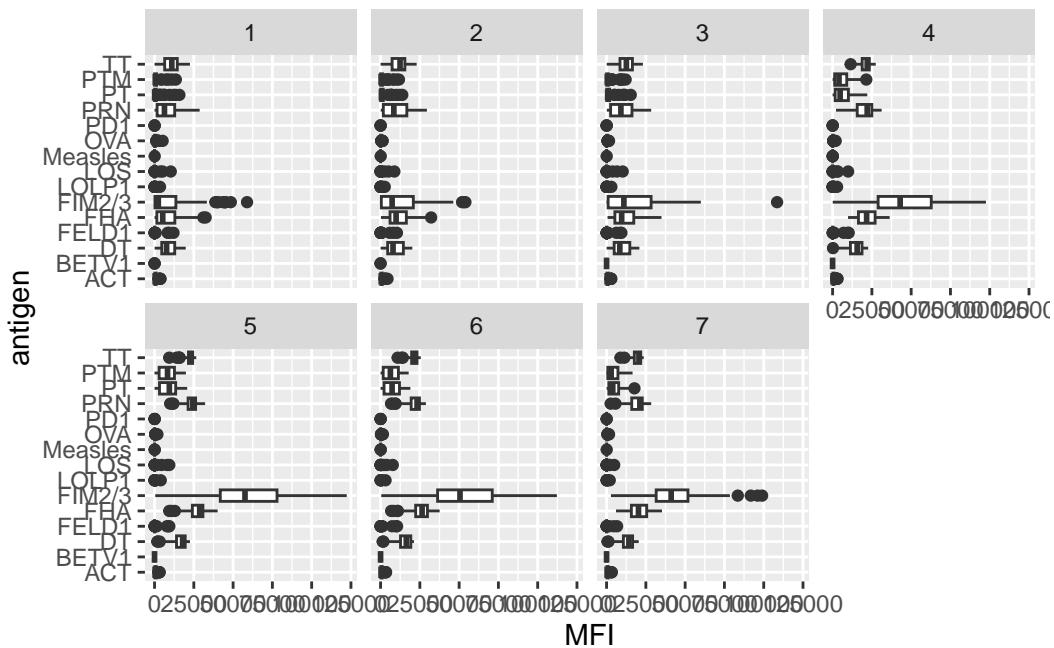
The number of visit 8 specimens are significantly lower than others.

4. Examine IgG1 Ab titer levels

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

Q13:

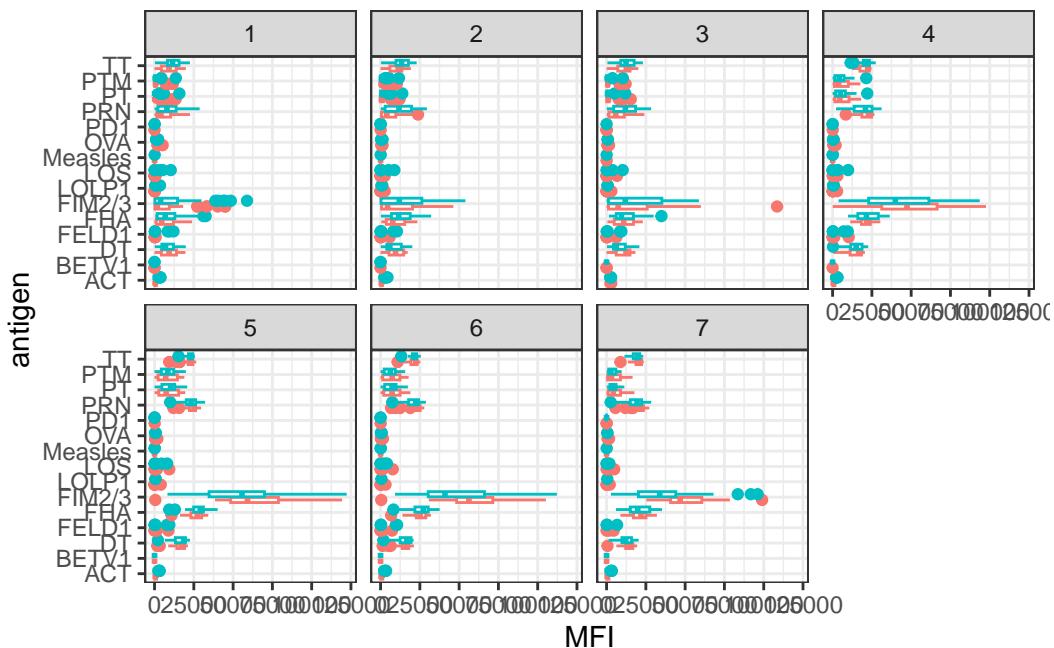
```
ggplot(ig1) +
  aes(MFI, antigen) +
  geom_boxplot() +
  facet_wrap(vars(visit), nrow = 2)
```



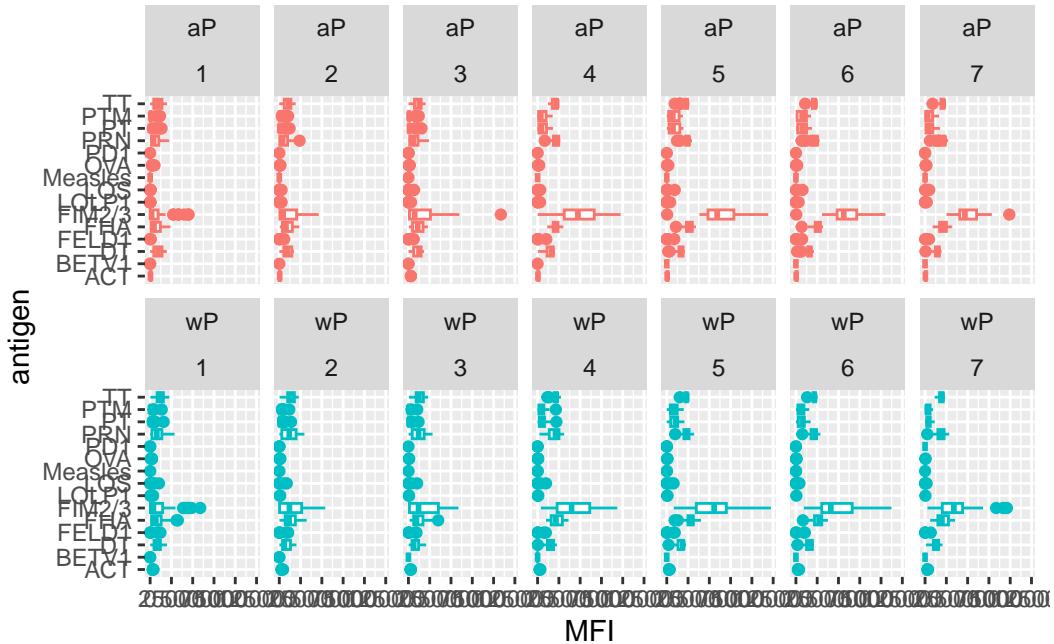
Q14:

FIM2/3 exhibited most significant differences in the level of IgG1 antibody titers overtime.

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

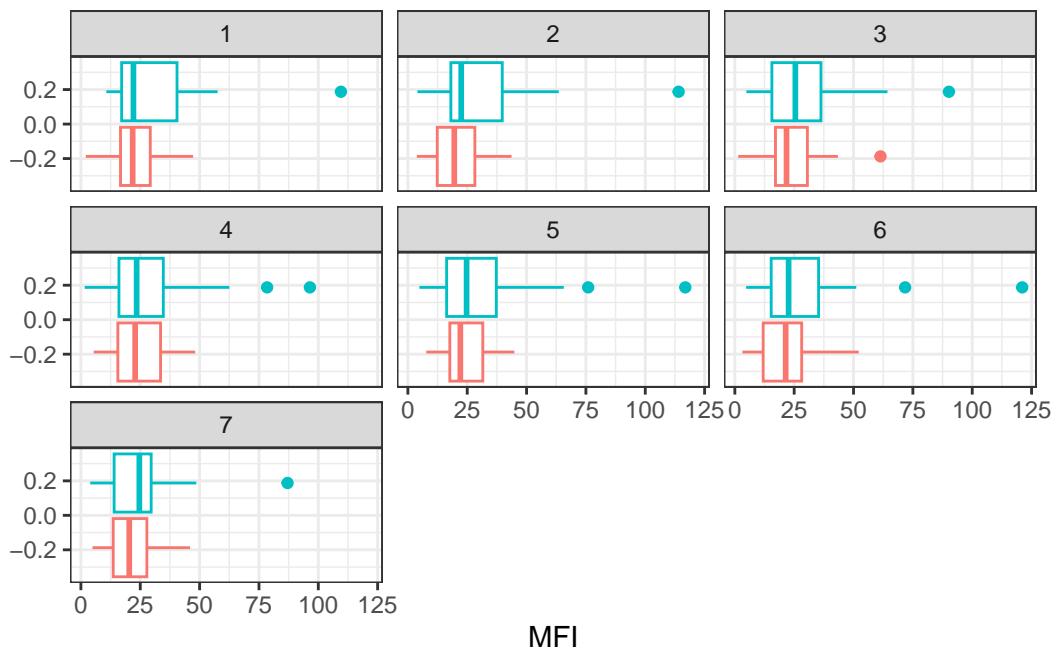


```
ggplot(ig1) +
  aes(MFI, col = infancy_vac ) +
  geom_boxplot(show.legend = FALSE) +
  facet_wrap(vars(infancy_vac, visit), nrow = 2)
```

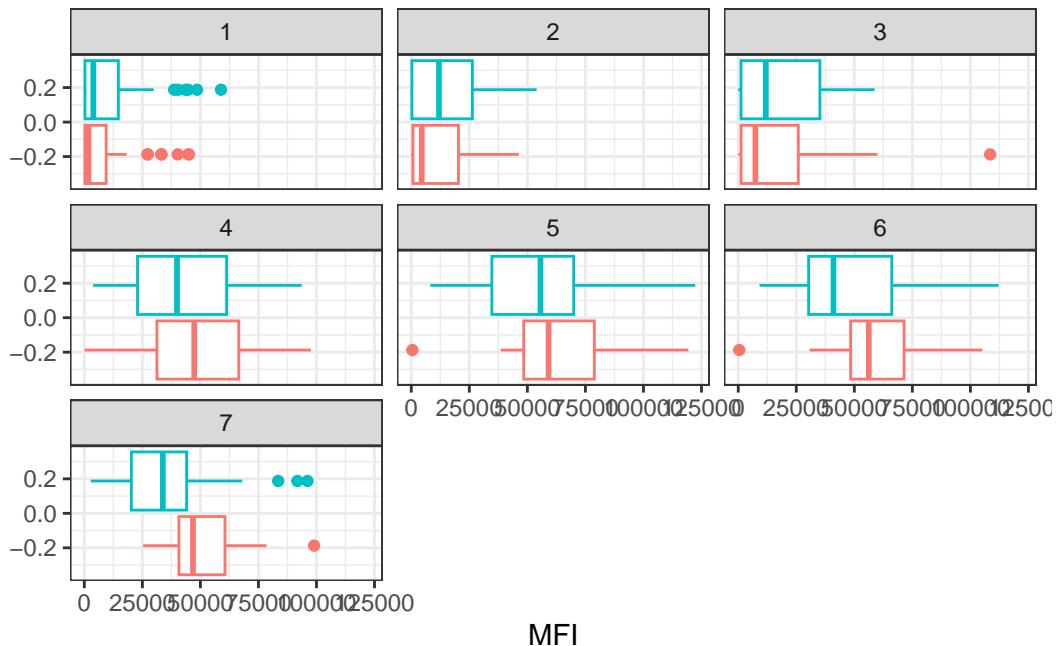


Q15:

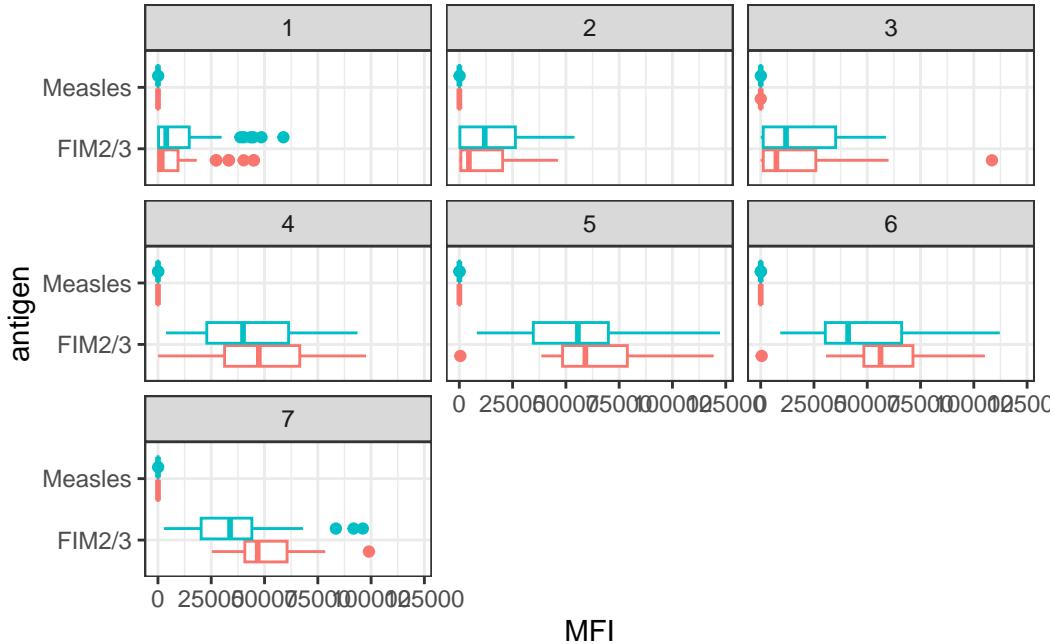
```
filter(ig1, antigen == "Measles") %>%
  ggplot() +
  aes(MFI, col = infancy_vac) +
  geom_boxplot(show.legend = F) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



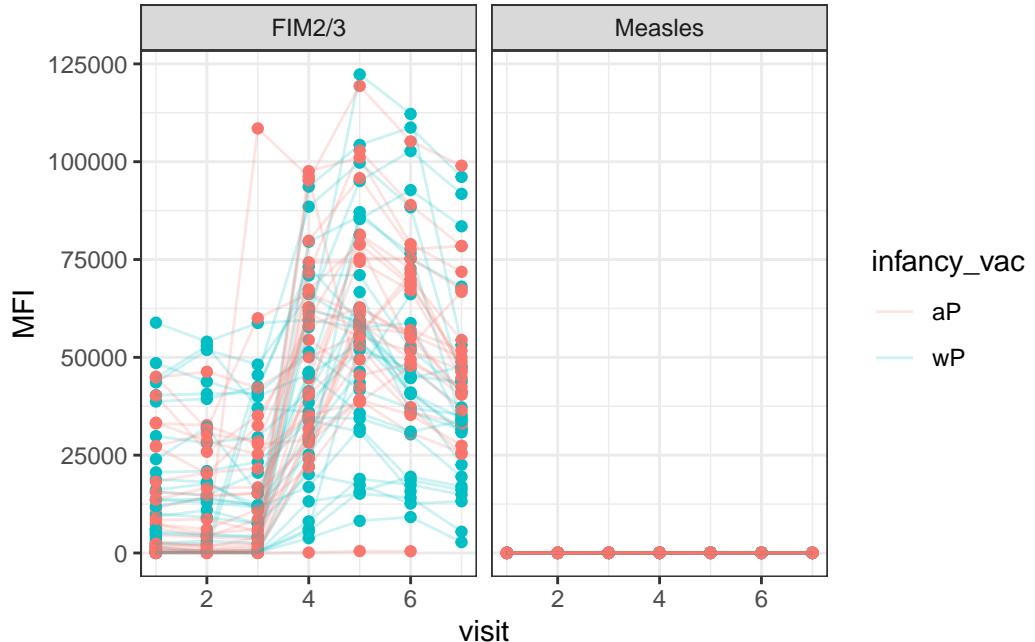
```
filter(ig1, antigen == "FIM2/3") %>%
  ggplot() +
  aes(MFI, col = infancy_vac) +
  geom_boxplot(show.legend = F) +
  facet_wrap(vars(visit)) +
  theme_bw()
```



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

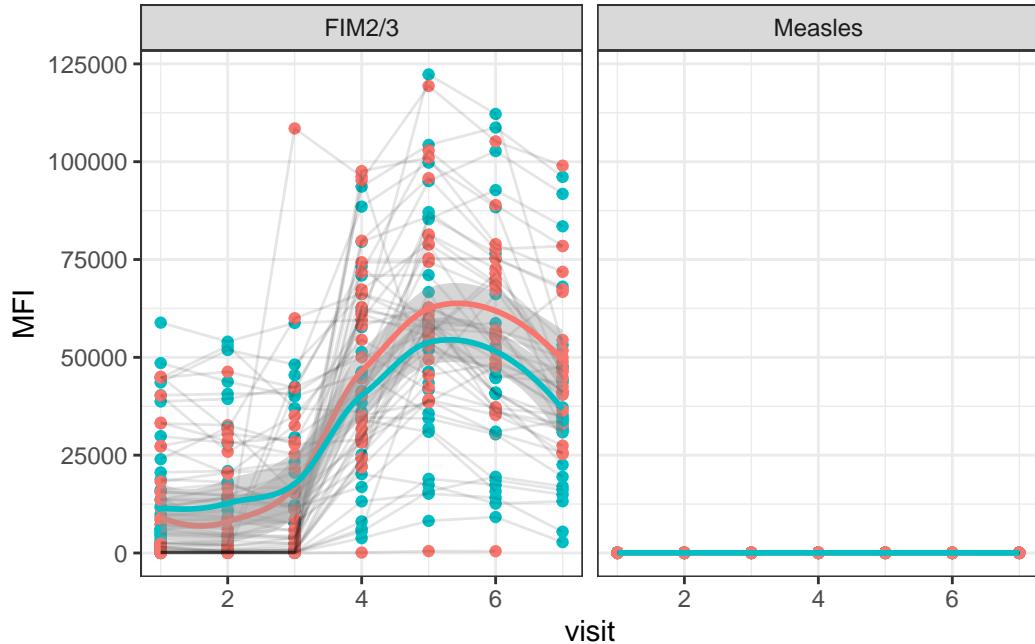


```
filter(ig1, antigen == "FIM2/3" | antigen == "Measles") %>%
  ggplot() +
  aes(visit, MFI, col = infancy_vac, antigen, group = subject_id) +
  geom_point(show.legend = F) +
  geom_line(alpha = 0.2) +
  facet_wrap(vars(antigen)) +
  theme_bw()
```



```
filter(ig1, antigen == "FIM2/3" | antigen == "Measles") %>%
  ggplot() +
  geom_point(aes(visit, MFI, col = infancy_vac), show.legend = F) +
  geom_line(aes(visit, MFI, group = subject_id), alpha = 0.1) +
  geom_smooth(aes(visit, MFI, col = infancy_vac), show.legend = F) +
  facet_wrap(vars(antigen)) +
  theme_bw()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Q16:

FIM2/3 rises and peaks at visit 5, whereas Measles do not rise at all.

Q17:

```
t.test(filter(ig1, antigen == "FIM2/3" & infancy_vac == "wP" & visit == 5)$MFI,
       filter(ig1, antigen == "FIM2/3" & infancy_vac == "aP" & visit == 5)$MFI)
```

Welch Two Sample t-test

```
data: filter(ig1, antigen == "FIM2/3" & infancy_vac == "wP" & visit == 5)$MFI and filter(ig1, antigen == "FIM2/3" & infancy_vac == "aP" & visit == 5)$MFI
t = -1.0479, df = 55.49, p-value = 0.2992
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-21445.415  6716.825
sample estimates:
mean of x mean of y
55662.16  63026.45
```

There are no significant difference between wP and aP responses.

5. Obtaining CMI-PB RNASeq data

```
url <- "https://www.cmi-pb.org/api/v2/rnaseq?versioned_ensembl_gene_id=eq.ESG00000211896"

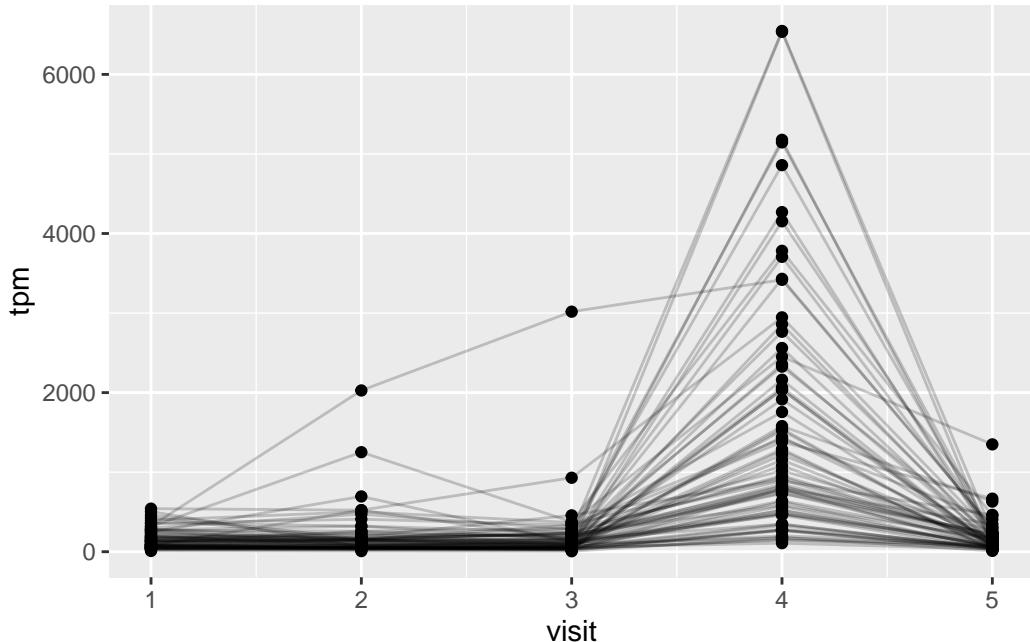
rna <- read_json(url, simplifyVector = TRUE)

ssrna <- inner_join(rna, meta)

Joining, by = "specimen_id"
```

Q18:

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



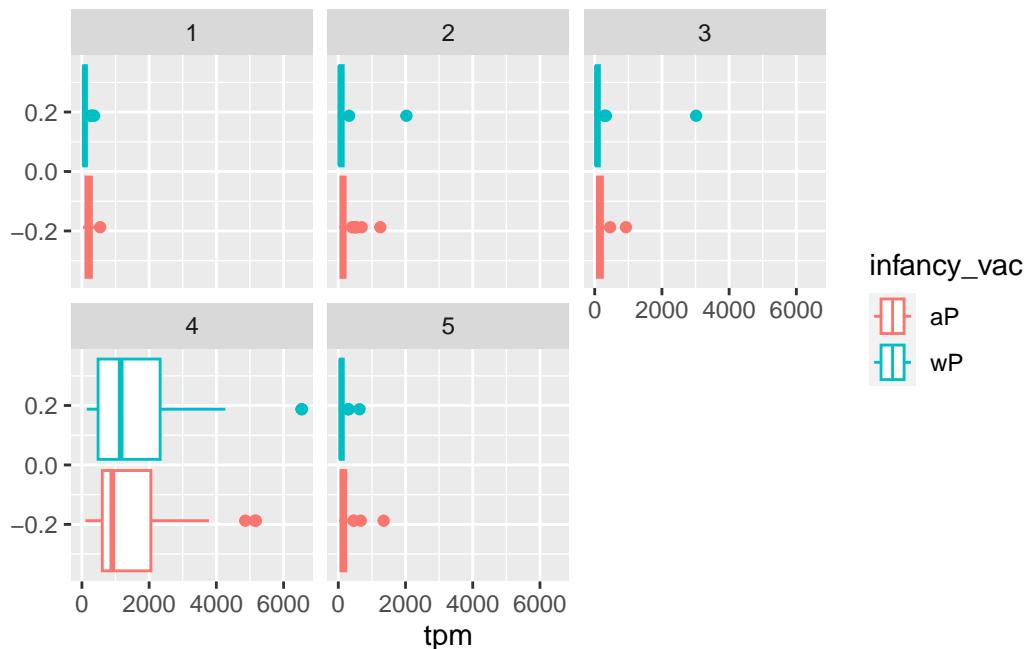
Q19:

The expression of IGHG1 gene peaks at visit 4, and drops back to basal line quickly after it.

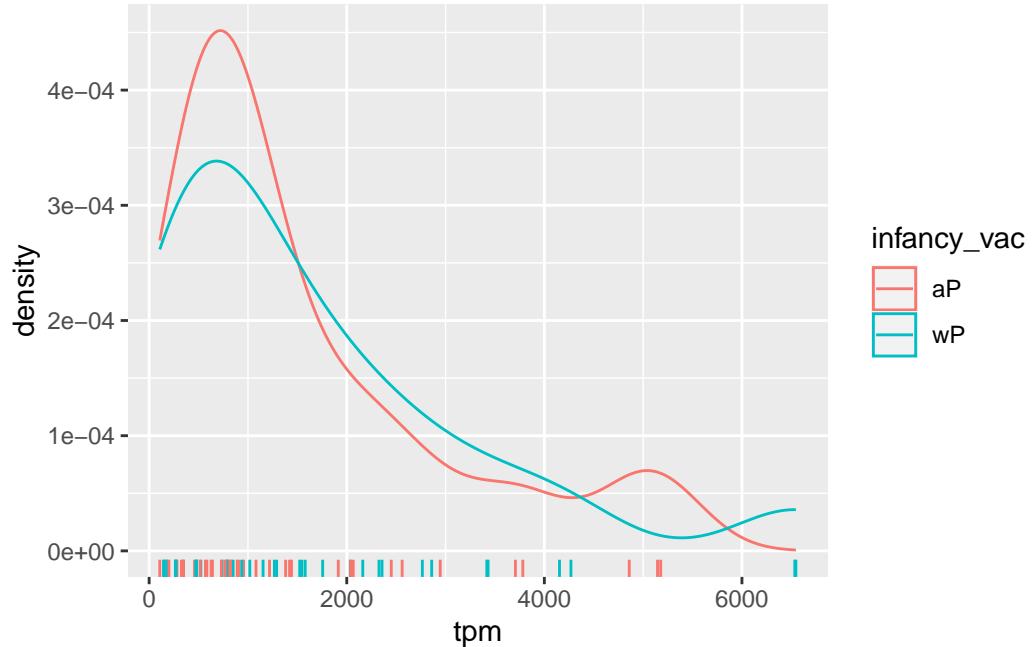
Q20:

The peaking of transcription level is a visit early than the peaking of antibody titer data. This could be explained by the fact that the RNA expression changes is prior to protein level changes.

```
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()
```



6. Working with larger datasets

```
rnaseq <- read.csv("2020LD_rnaseq.csv")
```

```
head(rnaseq, 3)
```

| | versioned_ensembl_gene_id | specimen_id | raw_count | tpm |
|---|---------------------------|-------------|-----------|-----|
| 1 | ENSG00000229704.1 | 209 | 0 | 0 |
| 2 | ENSG00000229707.1 | 209 | 0 | 0 |
| 3 | ENSG00000229708.1 | 209 | 0 | 0 |

```
dim(rnaseq)
```

```
[1] 10502460        4
```

Working with long format data

```
n_genes <- table(rnaseq$specimen_id)
head(n_genes , 10)

  1      3      4      5      6     19     20     21     22     23
58347 58347 58347 58347 58347 58347 58347 58347 58347 58347

all(n_genes[1] == n_genes)

[1] TRUE
```

Convert to “wide” format

```
library(tidyr)

rna_wide <- rnaseq %>%
  select(versioned_ensembl_gene_id, specimen_id, tpm) %>%
  pivot_wider(names_from = specimen_id, values_from = tpm)

dim(rna_wide)

[1] 58347   181
```

Filtering

```
rna_wide <- as.data.frame(rna_wide)
rownames(rna_wide) <- rna_wide[,1]
rna_wide <- rna_wide[-1]

rna <- rna_wide[rowSums(rna_wide == 0) == 0,]
```

PCA

```
rna.pr <- prcomp(t(rna), scale = T)
summary(rna.pr)
```

Importance of components:

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | |
|------------------------|----------|----------|----------|----------|----------|----------|---------|
| Standard deviation | 60.6583 | 49.9705 | 35.91043 | 31.0968 | 27.09718 | 21.50696 | |
| Proportion of Variance | 0.2241 | 0.1521 | 0.07855 | 0.0589 | 0.04473 | 0.02818 | |
| Cumulative Proportion | 0.2241 | 0.3762 | 0.45478 | 0.5137 | 0.55840 | 0.58658 | |
| | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 | |
| Standard deviation | 19.17286 | 17.13482 | 16.67289 | 16.27738 | 15.03324 | 14.32096 | |
| Proportion of Variance | 0.02239 | 0.01788 | 0.01693 | 0.01614 | 0.01377 | 0.01249 | |
| Cumulative Proportion | 0.60897 | 0.62685 | 0.64379 | 0.65993 | 0.67369 | 0.68618 | |
| | PC13 | PC14 | PC15 | PC16 | PC17 | PC18 | |
| Standard deviation | 13.53315 | 12.67410 | 12.25801 | 11.70855 | 11.47470 | 11.01182 | |
| Proportion of Variance | 0.01116 | 0.00978 | 0.00915 | 0.00835 | 0.00802 | 0.00739 | |
| Cumulative Proportion | 0.69734 | 0.70712 | 0.71628 | 0.72463 | 0.73265 | 0.74003 | |
| | PC19 | PC20 | PC21 | PC22 | PC23 | PC24 | |
| Standard deviation | 10.76559 | 10.29998 | 10.0082 | 9.63628 | 9.44306 | 9.42803 | |
| Proportion of Variance | 0.00706 | 0.00646 | 0.0061 | 0.00566 | 0.00543 | 0.00541 | |
| Cumulative Proportion | 0.74709 | 0.75356 | 0.7597 | 0.76531 | 0.77075 | 0.77616 | |
| | PC25 | PC26 | PC27 | PC28 | PC29 | PC30 | PC31 |
| Standard deviation | 9.27496 | 9.10433 | 8.96355 | 8.91515 | 8.7878 | 8.60664 | 8.44559 |
| Proportion of Variance | 0.00524 | 0.00505 | 0.00489 | 0.00484 | 0.0047 | 0.00451 | 0.00434 |
| Cumulative Proportion | 0.78140 | 0.78645 | 0.79134 | 0.79618 | 0.8009 | 0.80540 | 0.80974 |
| | PC32 | PC33 | PC34 | PC35 | PC36 | PC37 | PC38 |
| Standard deviation | 8.4007 | 8.26138 | 8.18063 | 8.12140 | 8.0032 | 7.85126 | 7.6919 |
| Proportion of Variance | 0.0043 | 0.00416 | 0.00408 | 0.00402 | 0.0039 | 0.00375 | 0.0036 |
| Cumulative Proportion | 0.8140 | 0.81820 | 0.82228 | 0.82629 | 0.8302 | 0.83395 | 0.8376 |
| | PC39 | PC40 | PC41 | PC42 | PC43 | PC44 | PC45 |
| Standard deviation | 7.63912 | 7.60838 | 7.52189 | 7.37348 | 7.15183 | 7.00860 | 6.9008 |
| Proportion of Variance | 0.00355 | 0.00353 | 0.00345 | 0.00331 | 0.00312 | 0.00299 | 0.0029 |
| Cumulative Proportion | 0.84111 | 0.84464 | 0.84808 | 0.85139 | 0.85451 | 0.85750 | 0.8604 |
| | PC46 | PC47 | PC48 | PC49 | PC50 | PC51 | PC52 |
| Standard deviation | 6.6551 | 6.54580 | 6.45333 | 6.31855 | 6.26560 | 6.08552 | 5.90039 |
| Proportion of Variance | 0.0027 | 0.00261 | 0.00254 | 0.00243 | 0.00239 | 0.00226 | 0.00212 |
| Cumulative Proportion | 0.8631 | 0.86571 | 0.86825 | 0.87068 | 0.87307 | 0.87533 | 0.87745 |
| | PC53 | PC54 | PC55 | PC56 | PC57 | PC58 | PC59 |
| Standard deviation | 5.83567 | 5.79800 | 5.7288 | 5.65386 | 5.5802 | 5.45969 | 5.36913 |
| Proportion of Variance | 0.00207 | 0.00205 | 0.0020 | 0.00195 | 0.0019 | 0.00182 | 0.00176 |
| Cumulative Proportion | 0.87952 | 0.88157 | 0.8836 | 0.88551 | 0.8874 | 0.88923 | 0.89098 |

| | PC60 | PC61 | PC62 | PC63 | PC64 | PC65 | PC66 |
|------------------------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 5.33129 | 5.26305 | 5.21117 | 5.07668 | 5.00510 | 4.9627 | 4.92934 |
| Proportion of Variance | 0.00173 | 0.00169 | 0.00165 | 0.00157 | 0.00153 | 0.0015 | 0.00148 |
| Cumulative Proportion | 0.89271 | 0.89440 | 0.89606 | 0.89763 | 0.89915 | 0.9006 | 0.90213 |
| | PC67 | PC68 | PC69 | PC70 | PC71 | PC72 | PC73 |
| Standard deviation | 4.86683 | 4.81317 | 4.74124 | 4.71886 | 4.66506 | 4.6271 | 4.57028 |
| Proportion of Variance | 0.00144 | 0.00141 | 0.00137 | 0.00136 | 0.00133 | 0.0013 | 0.00127 |
| Cumulative Proportion | 0.90357 | 0.90499 | 0.90635 | 0.90771 | 0.90904 | 0.9103 | 0.91161 |
| | PC74 | PC75 | PC76 | PC77 | PC78 | PC79 | PC80 |
| Standard deviation | 4.55773 | 4.54837 | 4.47374 | 4.4403 | 4.41226 | 4.37648 | 4.36788 |
| Proportion of Variance | 0.00127 | 0.00126 | 0.00122 | 0.0012 | 0.00119 | 0.00117 | 0.00116 |
| Cumulative Proportion | 0.91288 | 0.91414 | 0.91536 | 0.9166 | 0.91774 | 0.91891 | 0.92007 |
| | PC81 | PC82 | PC83 | PC84 | PC85 | PC86 | PC87 |
| Standard deviation | 4.32596 | 4.30933 | 4.28836 | 4.27554 | 4.2513 | 4.23927 | 4.21706 |
| Proportion of Variance | 0.00114 | 0.00113 | 0.00112 | 0.00111 | 0.0011 | 0.00109 | 0.00108 |
| Cumulative Proportion | 0.92121 | 0.92234 | 0.92346 | 0.92458 | 0.9257 | 0.92677 | 0.92786 |
| | PC88 | PC89 | PC90 | PC91 | PC92 | PC93 | PC94 |
| Standard deviation | 4.19902 | 4.15401 | 4.14085 | 4.12231 | 4.12170 | 4.10010 | 4.08789 |
| Proportion of Variance | 0.00107 | 0.00105 | 0.00104 | 0.00104 | 0.00103 | 0.00102 | 0.00102 |
| Cumulative Proportion | 0.92893 | 0.92998 | 0.93103 | 0.93206 | 0.93310 | 0.93412 | 0.93514 |
| | PC95 | PC96 | PC97 | PC98 | PC99 | PC100 | PC101 |
| Standard deviation | 4.07552 | 4.03781 | 4.02476 | 4.00741 | 4.00552 | 3.99172 | 3.97242 |
| Proportion of Variance | 0.00101 | 0.00099 | 0.00099 | 0.00098 | 0.00098 | 0.00097 | 0.00096 |
| Cumulative Proportion | 0.93615 | 0.93714 | 0.93813 | 0.93911 | 0.94009 | 0.94106 | 0.94202 |
| | PC102 | PC103 | PC104 | PC105 | PC106 | PC107 | PC108 |
| Standard deviation | 3.95087 | 3.94057 | 3.91823 | 3.90615 | 3.89913 | 3.89378 | 3.87923 |
| Proportion of Variance | 0.00095 | 0.00095 | 0.00094 | 0.00093 | 0.00093 | 0.00092 | 0.00092 |
| Cumulative Proportion | 0.94297 | 0.94391 | 0.94485 | 0.94578 | 0.94670 | 0.94763 | 0.94854 |
| | PC109 | PC110 | PC111 | PC112 | PC113 | PC114 | PC115 |
| Standard deviation | 3.86776 | 3.85663 | 3.8336 | 3.81309 | 3.78876 | 3.78379 | 3.77524 |
| Proportion of Variance | 0.00091 | 0.00091 | 0.0009 | 0.00089 | 0.00087 | 0.00087 | 0.00087 |
| Cumulative Proportion | 0.94946 | 0.95036 | 0.9513 | 0.95214 | 0.95302 | 0.95389 | 0.95476 |
| | PC116 | PC117 | PC118 | PC119 | PC120 | PC121 | PC122 |
| Standard deviation | 3.76252 | 3.75080 | 3.74101 | 3.72862 | 3.70700 | 3.69654 | 3.69508 |
| Proportion of Variance | 0.00086 | 0.00086 | 0.00085 | 0.00085 | 0.00084 | 0.00083 | 0.00083 |
| Cumulative Proportion | 0.95562 | 0.95648 | 0.95733 | 0.95818 | 0.95901 | 0.95984 | 0.96068 |
| | PC123 | PC124 | PC125 | PC126 | PC127 | PC128 | PC129 |
| Standard deviation | 3.68374 | 3.66810 | 3.65651 | 3.64075 | 3.63739 | 3.6163 | 3.60312 |
| Proportion of Variance | 0.00083 | 0.00082 | 0.00081 | 0.00081 | 0.00081 | 0.0008 | 0.00079 |
| Cumulative Proportion | 0.96150 | 0.96232 | 0.96314 | 0.96394 | 0.96475 | 0.9656 | 0.96634 |
| | PC130 | PC131 | PC132 | PC133 | PC134 | PC135 | PC136 |
| Standard deviation | 3.59875 | 3.58476 | 3.57679 | 3.56872 | 3.55644 | 3.55038 | 3.53451 |
| Proportion of Variance | 0.00079 | 0.00078 | 0.00078 | 0.00078 | 0.00077 | 0.00077 | 0.00076 |

| | | | | | | | |
|------------------------|---------|-----------|---------|---------|---------|---------|---------|
| Cumulative Proportion | 0.96713 | 0.96791 | 0.96869 | 0.96946 | 0.97024 | 0.97100 | 0.97176 |
| | PC137 | PC138 | PC139 | PC140 | PC141 | PC142 | PC143 |
| Standard deviation | 3.51807 | 3.51427 | 3.50386 | 3.49349 | 3.47856 | 3.47470 | 3.45298 |
| Proportion of Variance | 0.00075 | 0.00075 | 0.00075 | 0.00074 | 0.00074 | 0.00074 | 0.00073 |
| Cumulative Proportion | 0.97252 | 0.97327 | 0.97402 | 0.97476 | 0.97550 | 0.97623 | 0.97696 |
| | PC144 | PC145 | PC146 | PC147 | PC148 | PC149 | PC150 |
| Standard deviation | 3.44836 | 3.44306 | 3.43366 | 3.41538 | 3.3965 | 3.3900 | 3.3844 |
| Proportion of Variance | 0.00072 | 0.00072 | 0.00072 | 0.00071 | 0.0007 | 0.0007 | 0.0007 |
| Cumulative Proportion | 0.97768 | 0.97841 | 0.97912 | 0.97984 | 0.9805 | 0.9812 | 0.9819 |
| | PC151 | PC152 | PC153 | PC154 | PC155 | PC156 | PC157 |
| Standard deviation | 3.36809 | 3.35319 | 3.33843 | 3.33554 | 3.32960 | 3.32031 | 3.30105 |
| Proportion of Variance | 0.00069 | 0.00068 | 0.00068 | 0.00068 | 0.00068 | 0.00067 | 0.00066 |
| Cumulative Proportion | 0.98263 | 0.98331 | 0.98399 | 0.98467 | 0.98534 | 0.98601 | 0.98668 |
| | PC158 | PC159 | PC160 | PC161 | PC162 | PC163 | PC164 |
| Standard deviation | 3.29069 | 3.28356 | 3.27465 | 3.26242 | 3.24435 | 3.23475 | 3.22352 |
| Proportion of Variance | 0.00066 | 0.00066 | 0.00065 | 0.00065 | 0.00064 | 0.00064 | 0.00063 |
| Cumulative Proportion | 0.98734 | 0.98799 | 0.98865 | 0.98930 | 0.98994 | 0.99057 | 0.99121 |
| | PC165 | PC166 | PC167 | PC168 | PC169 | PC170 | PC171 |
| Standard deviation | 3.20725 | 3.19890 | 3.18227 | 3.16515 | 3.15970 | 3.1470 | 3.1320 |
| Proportion of Variance | 0.00063 | 0.00062 | 0.00062 | 0.00061 | 0.00061 | 0.0006 | 0.0006 |
| Cumulative Proportion | 0.99183 | 0.99246 | 0.99307 | 0.99368 | 0.99429 | 0.9949 | 0.9955 |
| | PC172 | PC173 | PC174 | PC175 | PC176 | PC177 | PC178 |
| Standard deviation | 3.12120 | 3.09159 | 3.06421 | 3.04720 | 3.04144 | 3.01235 | 2.97749 |
| Proportion of Variance | 0.00059 | 0.00058 | 0.00057 | 0.00057 | 0.00056 | 0.00055 | 0.00054 |
| Cumulative Proportion | 0.99609 | 0.99667 | 0.99724 | 0.99781 | 0.99837 | 0.99892 | 0.99946 |
| | PC179 | PC180 | | | | | |
| Standard deviation | 2.96893 | 3.677e-14 | | | | | |
| Proportion of Variance | 0.00054 | 0.000e+00 | | | | | |
| Cumulative Proportion | 1.00000 | 1.000e+00 | | | | | |

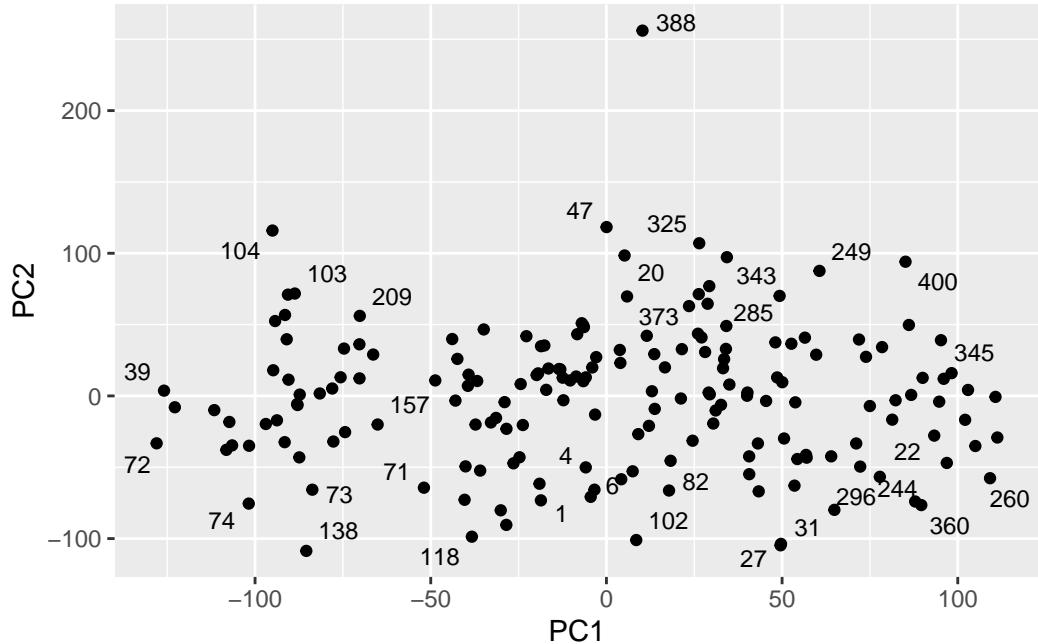
```

library(ggrepel)

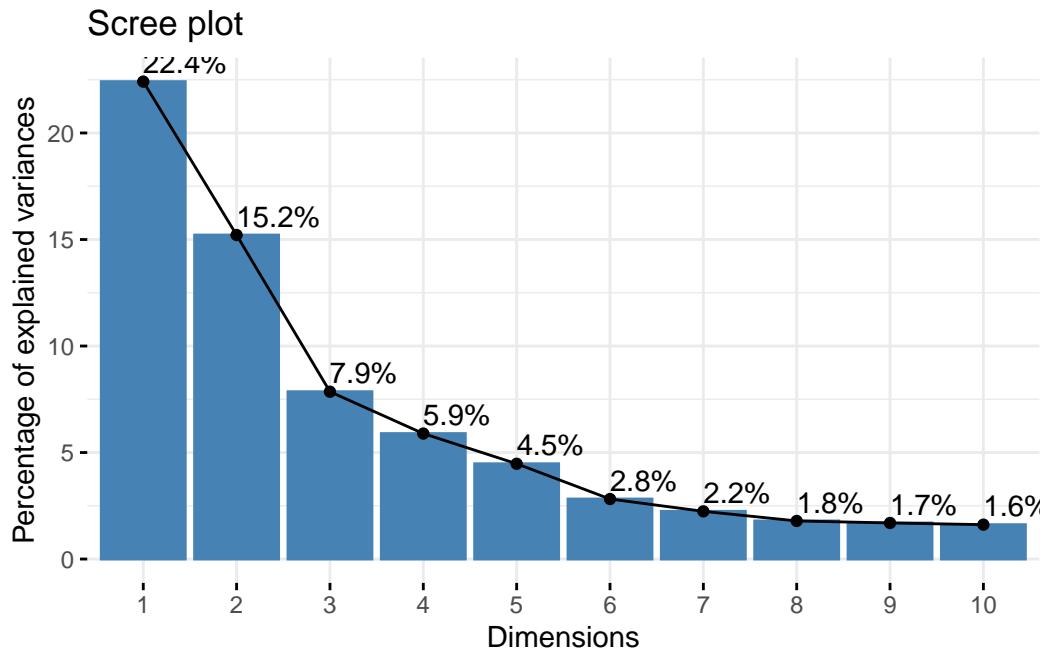
rna.pr$x %>%
  as.data.frame() %>%
  ggplot() +
  aes(PC1, PC2, label = rownames(rna.pr$x)) +
  geom_point() +
  geom_text_repel(size = 3.3, max.overlaps = 5)

```

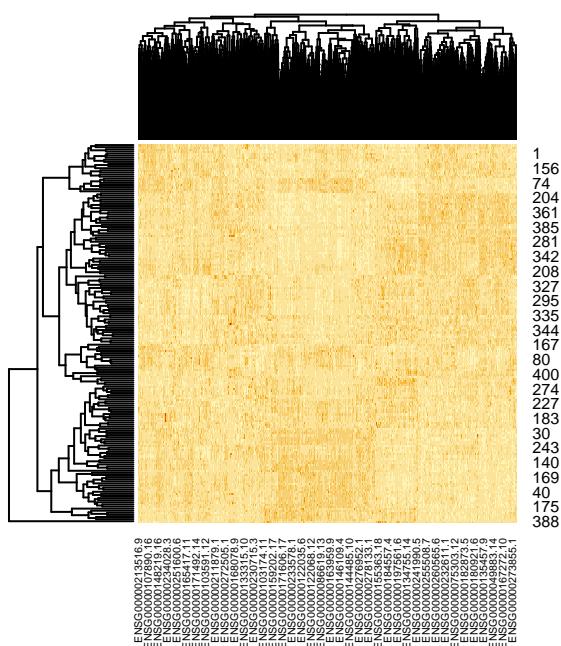
Warning: ggrepel: 147 unlabeled data points (too many overlaps). Consider increasing max.overlaps



```
library(factoextra)  
fviz_eig(rna.pr, addlabels = TRUE)
```



```
rna.norm <- t(scale(t(rna)))
heatmap(t(rna.norm))
```



```
heatmap(rna.pr$x)
```

