

Mini Project: Investigating Pertussis Resurgence

Jennifer

1. Investigating pertussis cases by year

Let's assign the data to a data frame and plot it.

Q1. With the help of the R “addin” package datapasta assign the CDC pertussis case number data to a data frame called cdc and use ggplot to make a plot of cases numbers over time.

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

Q2. Using the ggplot geom_vline() function add lines to your previous plot for the 1946 introduction of the wP vaccine and the 1996 switch to aP vaccine (see example in the hint below). What do you notice?

Q3. Describe what happened after the introduction of the aP vaccine? Do you have a possible explanation for the observed trend?

After the introduction of the aP vaccine, there is a significant resurgence in the number of pertussis cases. This could be due to bacterial evolution.

3. Exploring CMI-PB data

The CMI-PB API (like most APIs) sends responses in JSON format. To read these types of files into R we will use the read_json() function from the jsonlite package.

```
# Allows us to read, write and process JSON data
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
1          1          wP      Female Not Hispanic or Latino White
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. How many aP and wP infancy vaccinated subjects are in the dataset?

96

Q5. How many Male and Female subjects/patients are in the dataset?

66 female subjects and 30 male subjects

Q6. What is the breakdown of race and biological sex (e.g. number of Asian females, White males etc...)?

```
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

```

Two of the columns of subject contain dates in the Year-Month-Day format.

```
library(lubridate)
```

Loading required package: timechange

Attaching package: 'lubridate'

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

date, intersect, setdiff, union

What is today's date (at the time I am writing this obviously)

```
today()
```

```
[1] "2022-11-29"
```

How many days have passed since new year 2000?

```
today() - ymd("2000-01-01")
```

Time difference of 8368 days

What is this in years?

```
time_length( today() - ymd("2000-01-01"), "years")
```

```
[1] 22.91034
```

,

The average age for ap individuals is 24 and for wp individuals is 35.

>Q8. Determine the age of all individuals at time of boost?

```
::: {.cell}
```

```
```{r .cell-code}  
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. With the help of a faceted boxplot (see below), do you think these two groups are significantly different?

Q9. Complete the code to join specimen and subject tables to make a new merged data frame containing all specimen records along with their associated subject details: