

# HW2

## Jiayou's Graph

Aim: the graph is aiming to show us that the vaccine indeed reduced by a significant amount of cases for various diseases from history.

I fell like the graph personally when I hover on the bubbles, it tells me exact how many cases in that bubble. However, there are two things that really makes trouble when viewing. It is not that bad for me, especially I think I have no idea how to redo it for now. Technically, it is very interesting to me.

Advantages:

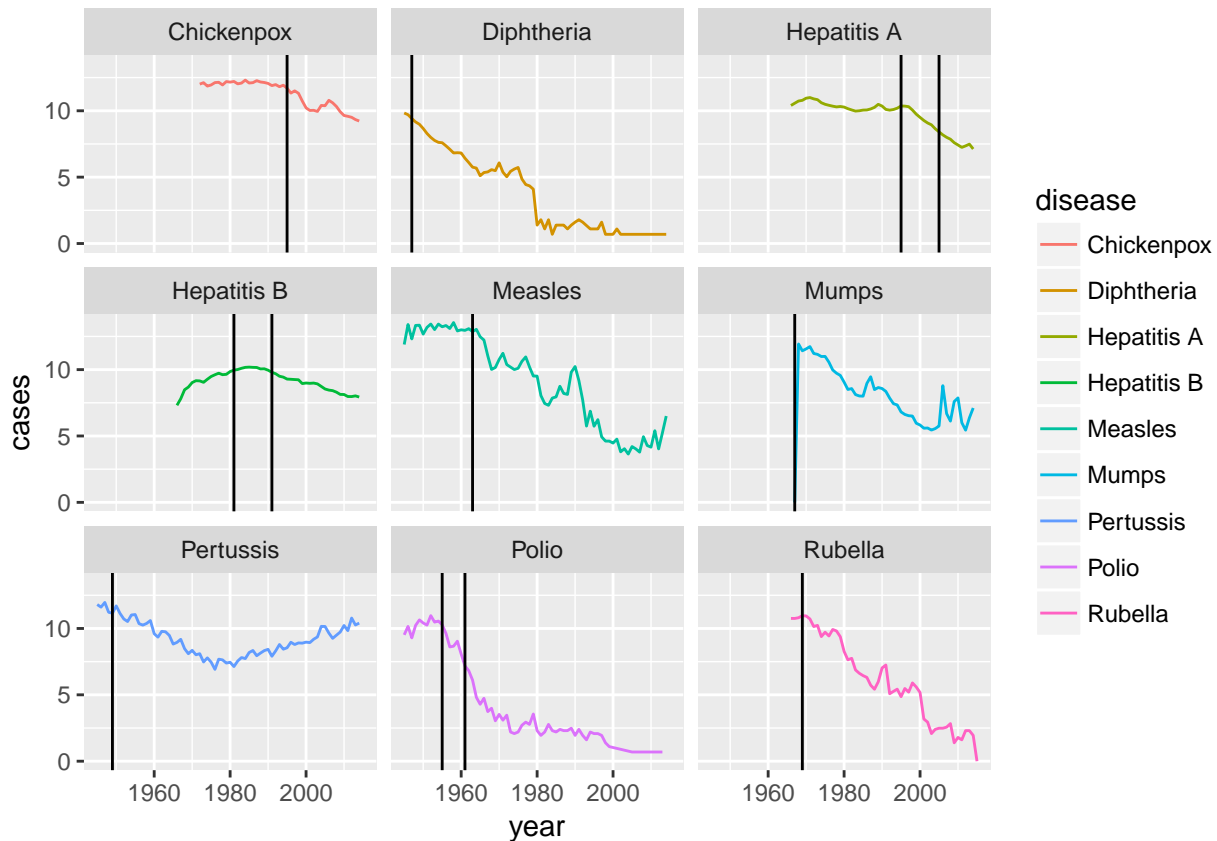
1. No information is lost, we see the graph and just hover on the bubbles, we can get all the information we need, which is cool to be a graph with full information.
2. The decent usage of hue, identifying the year of vaccine clearly, as well as the shape(scale) of the bubbles, indicating the cases' size.

Disadvantages:

1. The overlapping of bubbles are severe problems preventing me to see the patterns clearly, I may group the problem as bad usage of luminance as the luminance in the picture actually means nothing, but it is a bit misleading to me.
2. The size of the bubbles give us the information, if it could be something like lines, or violin plot, it could say more spaces for each year, which may make the graph no overlapping.
3. The size of the graph is too large, if we could shrink the length, it may be better for viewing.
4. If the graph is horizontally rather than vertically, it may be better for me.

## Graph 1

```
library(readr)
library(ggplot2)
library(tidyverse)
aa<-read.csv('vaccine_data_online.csv')
bb<-(aa %>% mutate(cases =log(cases+1))
      %>% mutate(vaccine=if_else(vaccine==FALSE,FALSE,TRUE)) )
cc<- (bb %>% filter(vaccine==TRUE))
gg0<-ggplot(bb,aes(x=year,y=cases,colour=disease))
gg1<-gg0+geom_line()+facet_wrap(~disease)+geom_vline(data=cc,aes(xintercept = year))
gg1
```



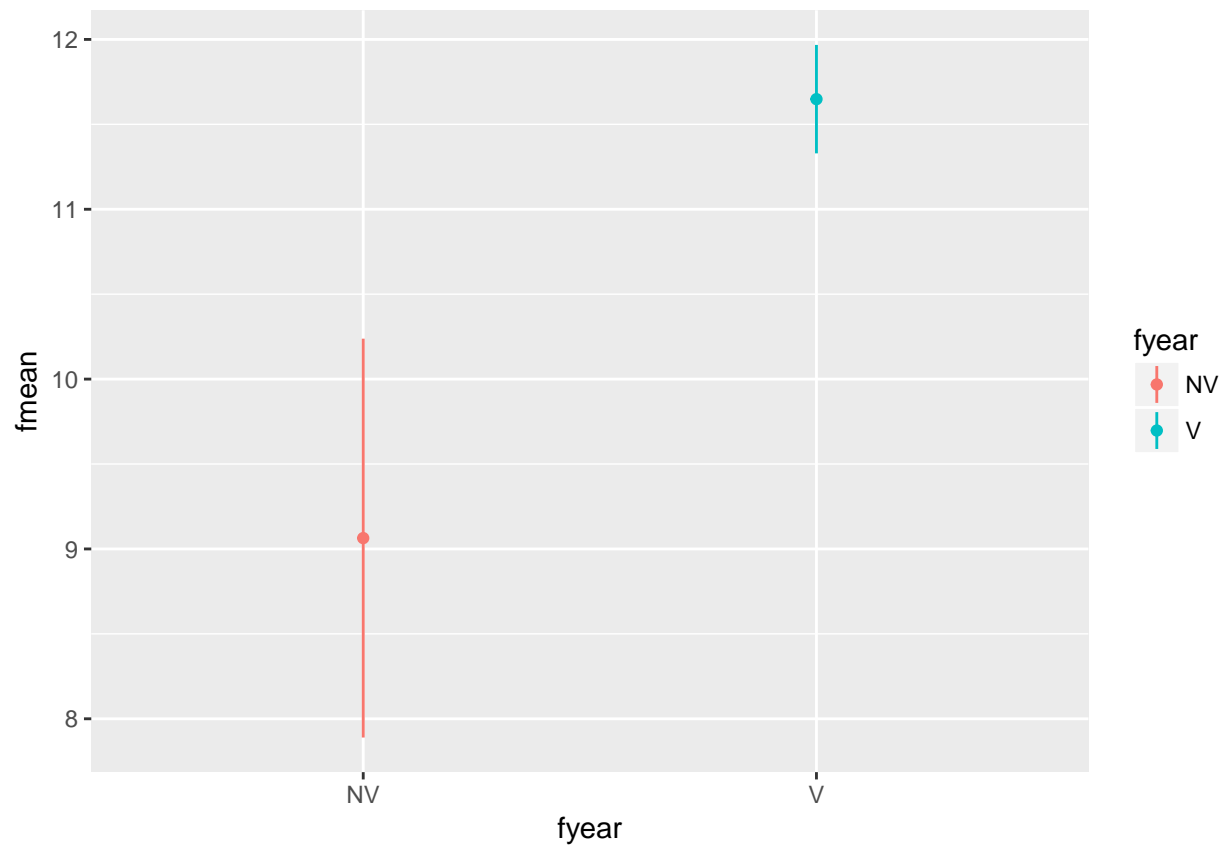
The graph has balanced view over different kinds of diseases. It shows clearly, directly and a bit ugly using abline to indicate when the vaccine appeared. I think it is more efficient comparing to the bubble plot.

Alignment is another thing we made, but I consider if we can put all these lines in the same plot with direct label, it may work better when the scale is good enough.

## Graph 2

Sorry, I have not come up with any good idea of the second graph. So I repeat again this graph, just using one of the disease 'Pertussis'. I tried to group the data by cut at the vaccine time, but I did not make it.

```
dd<-bb %>% filter(disease=='Pertussis') %>% mutate(fyear=if_else(year<1949,'V','NV'))
#dd
#dd$fyear
###
(dd
 %>% group_by(fyear)
 %>% summarise(fmean=mean(cases),
               se=sd(cases)))->ee
#ee
(ggplot(ee,aes(fyear,fmean,colour=fyear))+geom_point() +geom_linerange(aes(ymin=fmean-se,ymax=fmean+se)))
```



From the graph, we can tell that vaccine makes the mean of cases graphically significantly smaller.

This graph compares the scale by summarizing the data, using length (CI) instead of area. We may also apply a boxplot or violin plot. However, I have tried the whole afternoon and night, but just don't know how to make the cases to the y axis. So I end up doing this again.