

# HFG case data

Marlin

2023-06-15

## Data placeholder

This vignette covers the traditionally collected case data during the high frequency period. This case data has four different metrics of cases over a period of 55 weeks.

```
library(DSIWastewater)
library(dplyr)
library(ggplot2)

data(HFGCase_data, package = "DSIWastewater")
```

The data is broken down by site and date. Site is where it was collected from a range of 10 locations. For each date and site there is four metrics of case data. These generally agree but have some slight variation. In this data a case number of less than 5 is represented with -999.

```
HFGCase_data%>%
  filter(!is.na(ReportedCases))%>%
  ggplot(aes(x = date), size = .5)+
  geom_point(aes(y = ReportedCases, color = "Reported Cases"))+
  geom_point(aes(y = EpisodeCases, color = "Episode Cases"))+
  geom_point(aes(y = CollectedCases, color = "Collected Cases"))+
  geom_point(aes(y = ConfirmedCases, color = "Confirmed Cases"))+
  scale_y_log10()+
  facet_wrap(~site)
```



Otherwise the signals measured were N1 and N2, different genes of Covid-19, PMMOV, BCoV, HF183, and CrP, which are suppose to functions as normalizes. These signals can have a couple of different companion columns. The Ct version is what the machine measured which was the number of spins to detect the genes. This is an exponential decay transformation to the Base column. The other extra column is the LOD which says if the measurement was bellow the Labs level of detection. This might mean the exact number is less reliable.

```
head(HFGCase_data)
```

##	date	site	ReportedCases	EpisodeCases	CollectedCases	ConfirmedCases
## 1	2020-10-30	Hudson	NA	NA	NA	NA
## 2	2021-01-04	Hudson	NA	NA	NA	NA
## 3	2021-01-07	Hudson	NA	NA	NA	NA
## 4	2021-01-10	Hudson	NA	NA	NA	NA
## 5	2021-01-12	Hudson	NA	NA	NA	NA
## 6	2021-01-14	Hudson	NA	-999	NA	NA

We used this data set in some analysis here