

# HFG waste data

Marlin

2023-06-15

## HFG data source

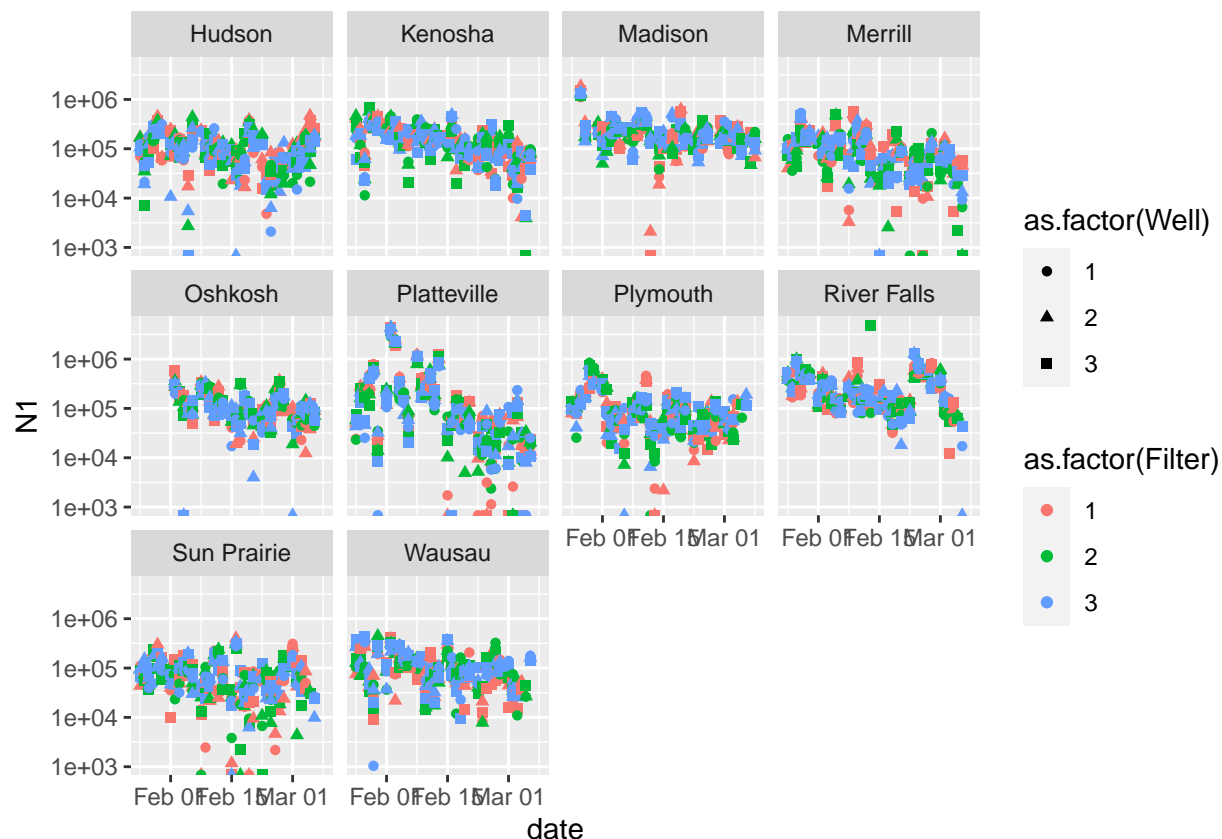
This vignette covers the high frequency data collected by the Wisconsin State Lab of Hygiene. This Wastewater data covers 6 weeks of 9 measurements a day of 6 day a week. This has 17 columns where 4 categorical variables and 13 measurement columns which analyse 6 signals.

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

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

The data is broken down by site, date, filter and well. Site is where it was collected from a range of 10 locations. Filter and well are what level of replicates it is. Filter is the high level category representing the different collections. Well contains the info about what technique batch it was. The numbering of the categories is arbitrary from day to day.

```
#fix titles
#fix overplotting points and axis
#drop well category in plot
HFGWaste_data%>%
  ggplot(aes(x = date, y = N1, color = as.factor(Filter), shape = as.factor(Well)), size = .5)+
  geom_point()+
  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.

```
#change to table form
#over explain is better then under explain
#have explanation as an appendix
head(HFGWaste_data)
```

##	site	date	Filter	Well	N1Ct	N1	N1LOD	N2CT	N2	N2LOD	PMMOVCT
## 1	Hudson	2021-01-25	1	1	33.25	70268	FALSE	33.70	40121	TRUE	24.96
## 2	Hudson	2021-01-25	1	2	31.86	171547	FALSE	32.87	67953	FALSE	24.89
## 3	Hudson	2021-01-25	1	3	32.88	88856	FALSE	33.21	54756	TRUE	25.15
## 4	Hudson	2021-01-25	2	1	32.20	138137	FALSE	32.11	110372	FALSE	25.21
## 5	Hudson	2021-01-25	2	2	31.89	168132	FALSE	32.16	107241	FALSE	25.22
## 6	Hudson	2021-01-25	2	3	32.72	98634	FALSE	32.88	67862	FALSE	25.18
##	PMMOV	BCoV	HF183CT	HF183	CrPCT	CrP	Notes				
## 1	8335368	1.92	28.05	236887243	29.66	119023953	<NA>				
## 2	8756036	2.06	28.22	210693707	29.92	99677418	<NA>				
## 3	7268880	1.95	28.60	162485283	30.06	90503607	<NA>				
## 4	6945237	4.15	27.88	265156508	29.94	98203103	<NA>				
## 5	6917604	4.76	28.27	203828374	30.05	91220128	<NA>				
## 6	7103485	6.10	27.71	296482397	29.95	97912594	<NA>				

We used this data set in some analysis here