noisevspopulation

Kyllan Wunder

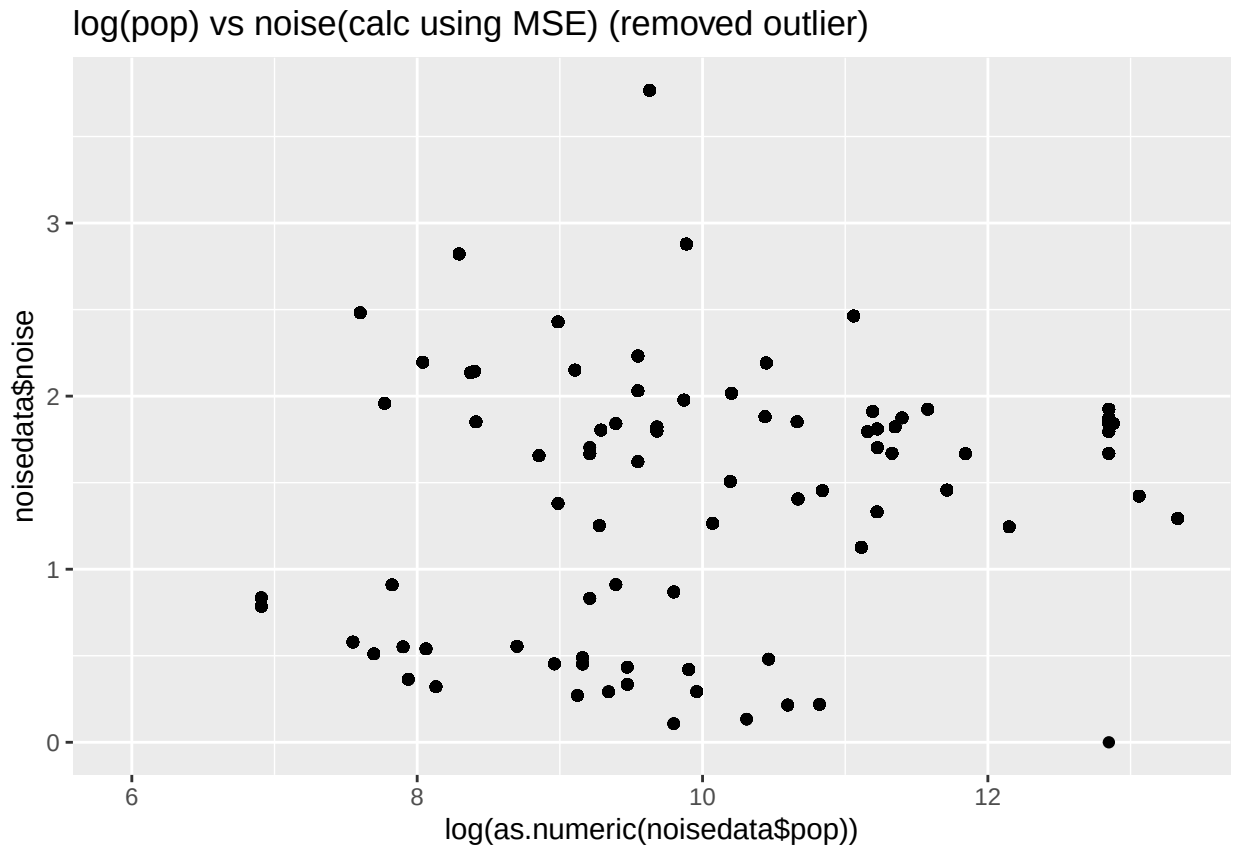
2023-05-12

```
MSE <- function(baseVal, meanVal){
  ret <- mean((baseVal - meanVal)^2)
  return(ret)
}


data(WasteWater_data, package = "DSIWastewater")


noisedata <- WasteWater_data %>%
  drop_na(site) %>%
  drop_na(PMMoV)

#sqrt(sum((mean-ppmov)^2)/n)
noisedata <- noisedata %>%
  group_by(site) %>%
  mutate(noise = MSE(log(as.numeric(PMMoV)),
                     mean(log(as.numeric(PMMoV)),na.rm = TRUE)))
```

```
## Warning: There were 4 warnings in `mutate()`.
## The first warning was:
## i In argument: `noise = MSE(log(as.numeric(PMMoV)),
##   mean(log(as.numeric(PMMoV)), na.rm = TRUE))`.
## i In group 59: `site = "Rib Lake"`.
## Caused by warning in `log()`:
## ! NaNs produced
## i Run `dplyr::last_dplyr_warnings()` to see the 3 remaining warnings.
```

```
#largest noise value that skews graph
noisedata <- noisedata %>% filter(site != "Lake Geneva WWTP")

ggplot(noisedata, aes(log(as.numeric(noisedata$pop)), noisedata$noise)) +
  geom_point() +
  ggtitle("log(pop) vs noise(calc using MSE) (removed outlier)")
```

```
## Warning: Removed 942 rows containing missing values (`geom_point()`).
```

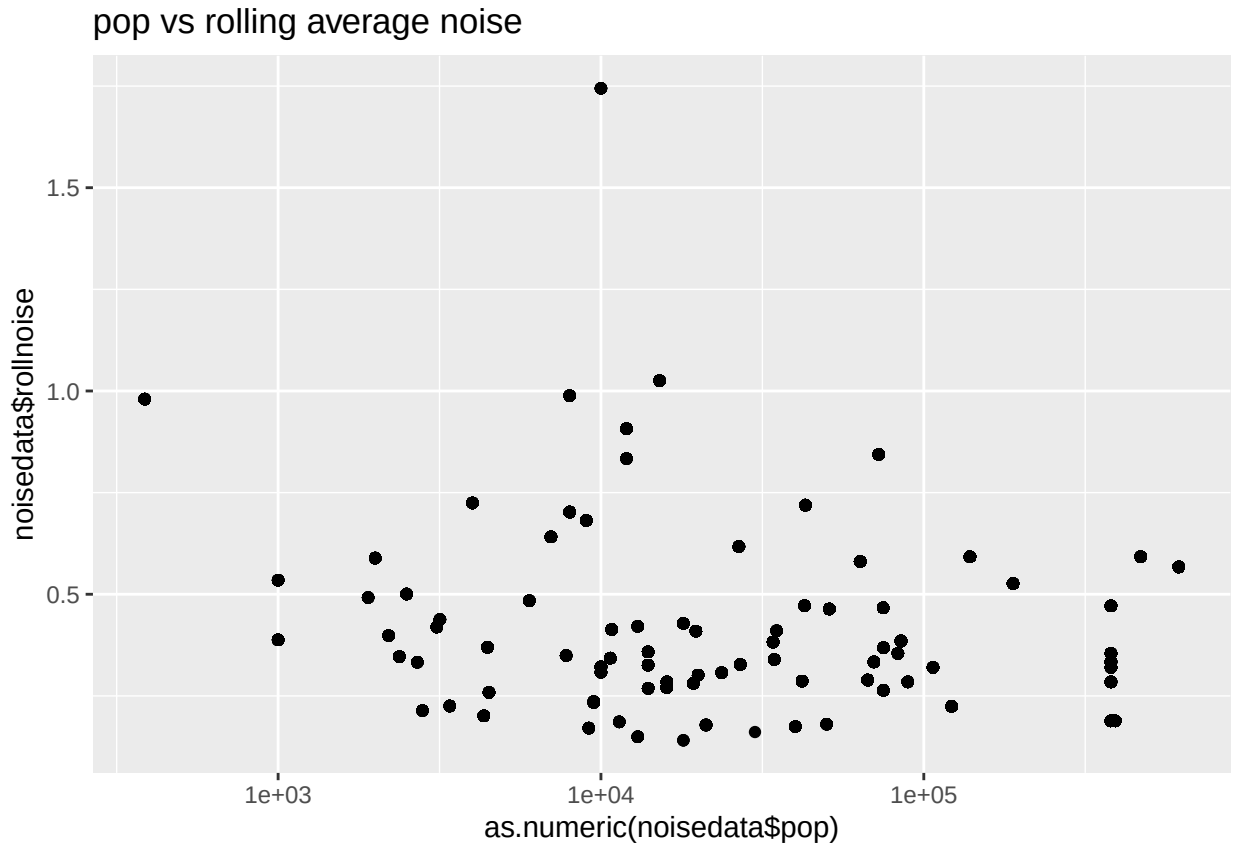



```
noisedata <- noisedata %>%
  group_by(site) %>%
  mutate(rollaverage = rollmean(log(as.numeric(PMMoV)), k=7, fill = NA)) %>%
  drop_na(rollaverage)
```

```
## Warning: There were 2 warnings in `mutate()`.
## The first warning was:
## i In argument: `rollaverage = rollmean(log(as.numeric(PMMoV)), k = 7, fill =
##   NA)`.
## i In group 59: `site = "Rib Lake"`.
## Caused by warning in `log()`:
## ! NaNs produced
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```
noisedata <- noisedata %>%
  group_by(site) %>%
  mutate(rollnoise = MSE(log(as.numeric(PMMoV)), rollaverage))

ggplot(noisedata) +
  aes(as.numeric(noisedata$pop), noisedata$rollnoise)+
  geom_point()+
  scale_x_log10() +
  ggtitle("pop vs rolling average noise")
```

```
## Warning: Removed 410 rows containing missing values (`geom_point()`).
```

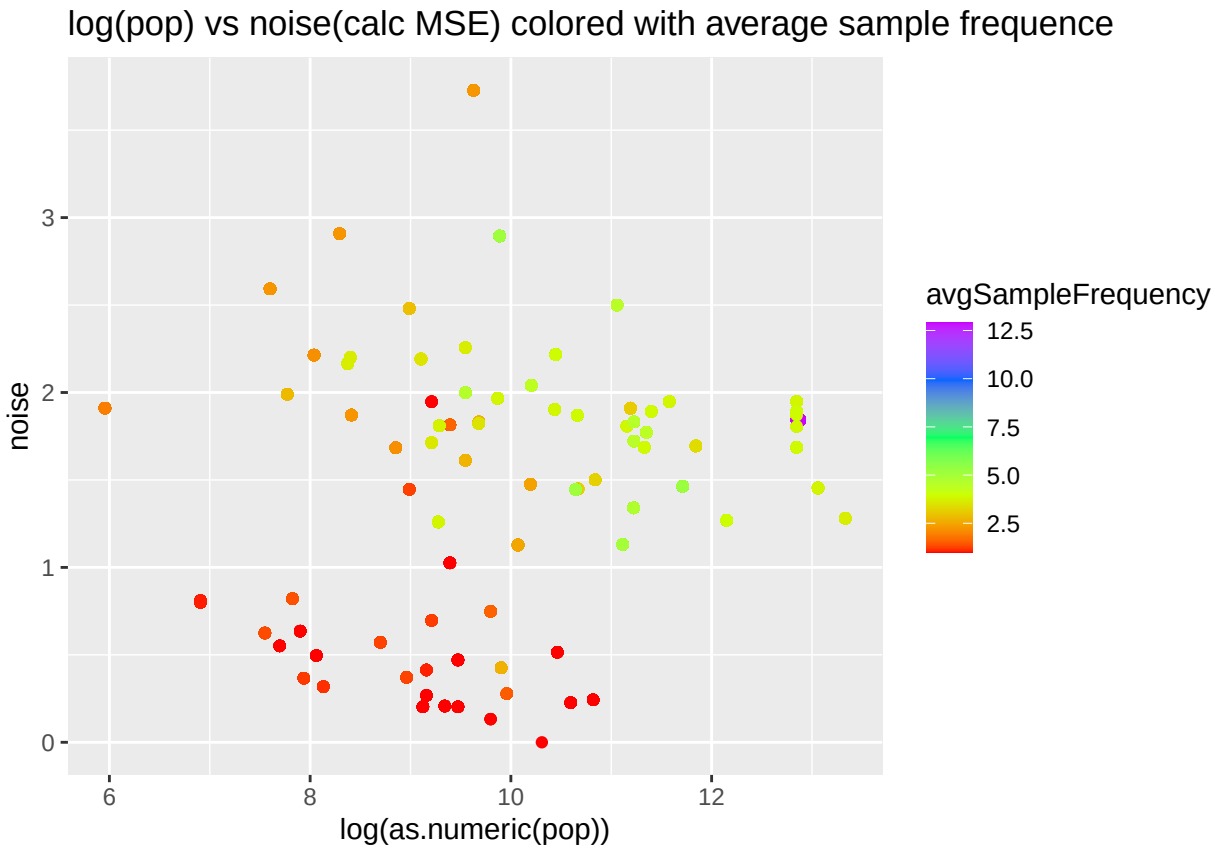



```
#not much improvement using the rolling average over 7 points
```

```
noisedatasample <- noisedata

noisedatasample <- noisedatasample %>%
  group_by(site, week = week(anydate(date))) %>% mutate(sampleFrequency = n()) %>%
  ungroup()


noisedatasample <- noisedatasample %>% group_by(site) %>%
  mutate(avgSampleFrequency = mean(sampleFrequency))
```

```
noisedatasample <- noisedatasample %>%
  group_by(site) %>%
  mutate(noise = MSE(log(as.numeric(PMMoV)),
                     mean(log(as.numeric(PMMoV)),na.rm = TRUE)))

#largest noise value that skews graph
#noisedatasample <- noisedatasample %>% filter(site != "Lake Geneva WWTP")

ggplot(noisedatasample, aes(log(as.numeric(pop)), noise)) + #/avgSampleFrequency
  geom_point(aes(color=avgSampleFrequency)) +
  scale_colour_gradientn(colours=rainbow(5)) +
  ggtitle("log(pop) vs noise(calc MSE) colored with average sample frequence")
```

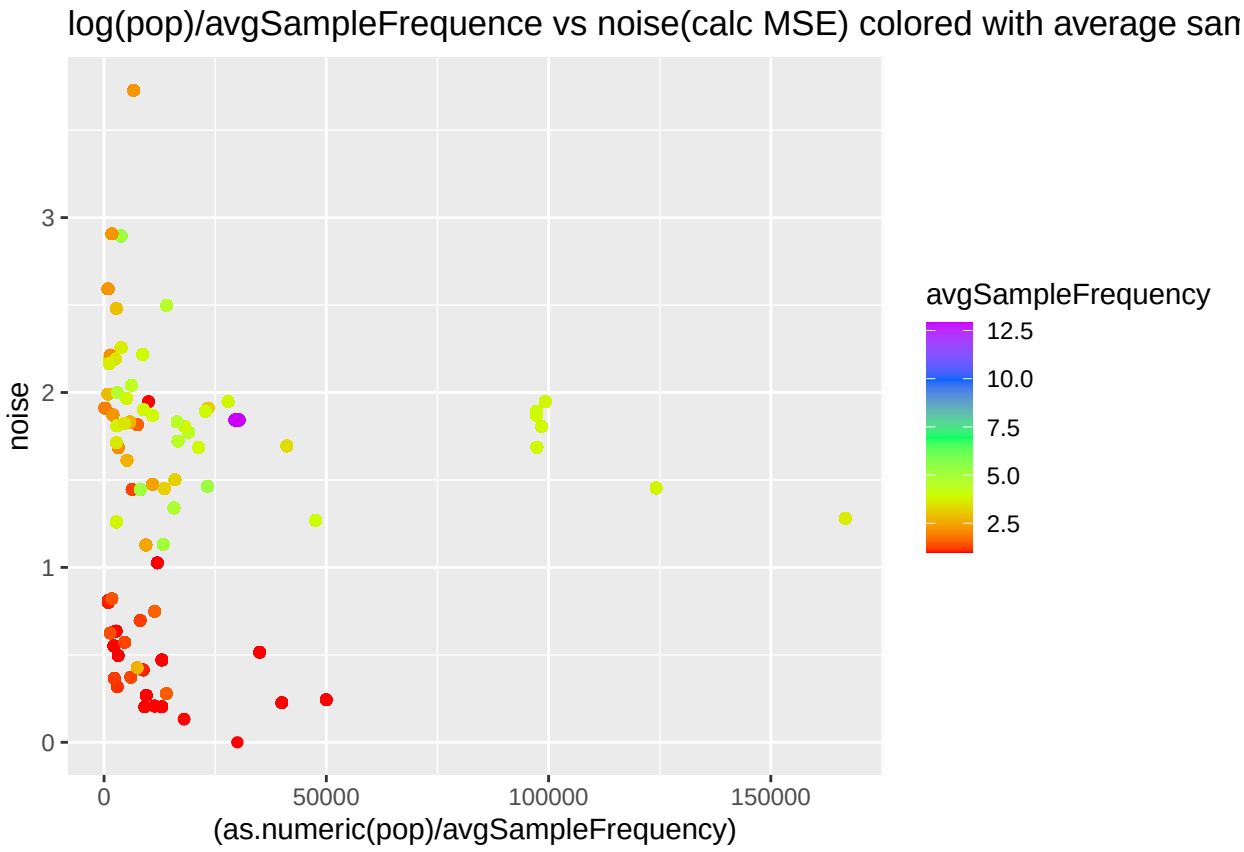
```
## Warning: Removed 410 rows containing missing values (‘geom_point()‘).
```




```
#Madison MSD WWTF has 12
```

```
ggplot(noisedatasample, aes((as.numeric(pop)/avgSampleFrequency), noise)) +
  geom_point(aes(color=avgSampleFrequency)) +
  scale_colour_gradientn(colours=rainbow(5)) +
  ggtitle("log(pop)/avgSampleFrequence vs noise(calc MSE) colored with average sample frequence")
```

```
## Warning: Removed 410 rows containing missing values (‘geom_point()‘).
```




```
noisedatasample %>% #filter(grepl("Madison", site)) %>%
ggplot(aes(week, avgSampleFrequency)) +
  geom_point() +
  facet_wrap(~site) +
  ggtitle("average sample frequencey by week by site ")


noisedatasample %>%
  group_by(site) %>%
ggplot(aes(log(as.numeric(population_served)), avgSampleFrequency)) +
  geom_point() +
  ggtitle("log(pop) vs average sample frequency")
```