

Preprocessing:

```
library(ShortRead)
fqsample <- readFastq(dirPath = "data",
                      pattern = "SRR1971253.fastq")
abc <- alphabetByCycle(sread(fqsample))
# transpose nucleotides A, C, G, T per column
nucByCycle <- t(abc[1:4,])
# tidy dataset
nucByCycle <- nucByCycle %>%
  as.tibble() %>% # convert to tibble
  mutate(cycle = 1:50) # add cycle numbers
```

```
> # glimpse nucByCycle
> glimpse(nucByCycle)
Observations: 50
Variables: 5
$ A    <int> 16986, 13152, 13763, 14843, 16460, 15655, 15679, 15829, 14912...
$ C    <int> 16457, 13441, 15756, 15559, 14197, 15316, 14320, 15151, 15618...
$ G    <int> 16880, 12168, 13299, 14340, 14566, 15266, 16315, 14841, 14984...
$ T    <int> 10968, 22575, 18518, 16594, 16110, 15099, 15022, 15515, 15822...
$ cycle <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18...
>
> # make an awesome plot!
> nucByCycle %>%
  # gather the nucleotide letters in alphabet and get a new count column
  gather(key = alphabet, value = count, -cycle) %>%
  ggplot(aes(x = cycle, y = count, color = alphabet)) +
  geom_line(size = 0.5) +
  labs(y = "Frequency") +
  theme_bw() +
  theme(panel.grid.major.x = element_blank())
```

