

Adapter lengths for all datasets

2025-03-19

Load all data

```
a60_unmod <- read.table(file = "a60_unmod_polyA_position.tsv", sep = "\t", header = TRUE)
a60_30 <- read.table(file = "a60_30_polyA_position.tsv", sep = "\t", header = TRUE)
a60_60 <- read.table(file = "a60_60_polyA_position.tsv", sep = "\t", header = TRUE)

a120_unmod <- read.table(file = "a120_unmod_polyA_position.tsv", sep = "\t", header = TRUE)
a120_1mod <- read.table(file = "a120_1mod_polyA_position.tsv", sep = "\t", header = TRUE)
a120_2mod <- read.table(file = "a120_2mod_polyA_position.tsv", sep = "\t", header = TRUE)
a120_4mod <- read.table(file = "a120_4mod_polyA_position.tsv", sep = "\t", header = TRUE)
```

Define lengths of adapter

```
a60_unmod["adapter_length"] <- a60_unmod$start - 1
a60_30["adapter_length"] <- a60_30$start - 1
a60_60["adapter_length"] <- a60_60$start - 1

a120_unmod["adapter_length"] <- a120_unmod$start - 1
a120_1mod["adapter_length"] <- a120_1mod$start - 1
a120_2mod["adapter_length"] <- a120_2mod$start - 1
a120_4mod["adapter_length"] <- a120_4mod$start - 1
```

Histograms of all the data in each set

```
datasets <- list(
  a60_30 = a60_30,
  a60_60 = a60_60,
  a60_unmod = a60_unmod,
  a120_1mod = a120_1mod,
  a120_2mod = a120_2mod,
  a120_4mod = a120_4mod,
  a120_unmod = a120_unmod
)

n_bins_visible <- 100
breaks_common <- seq(from = 0, to = 8000, length.out = n_bins_visible + 1)

for (name in names(datasets)) {
  adapter_length <- datasets[[name]]$adapter_length
```

```

adapter_length <- na.omit(adapter_length)
adapter_length <- adapter_length[adapter_length >= 0 & adapter_length <= 8000]

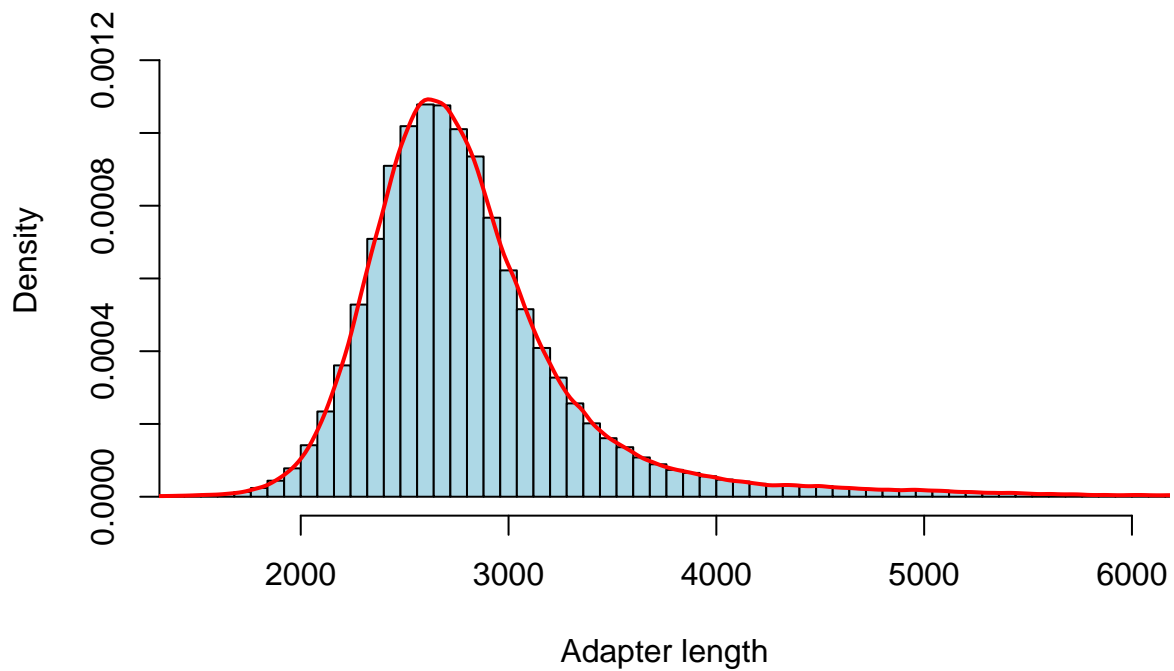
if (length(adapter_length) == 0) next

hist(adapter_length,
      probability = TRUE,
      main = paste("Density Plot of adapter lengths -", name),
      xlab = "Adapter length",
      ylab = "Density",
      xlim = c(1500, 6000),
      ylim = c(0, 0.0013),
      col = "lightblue",
      border = "black",
      breaks = breaks_common)

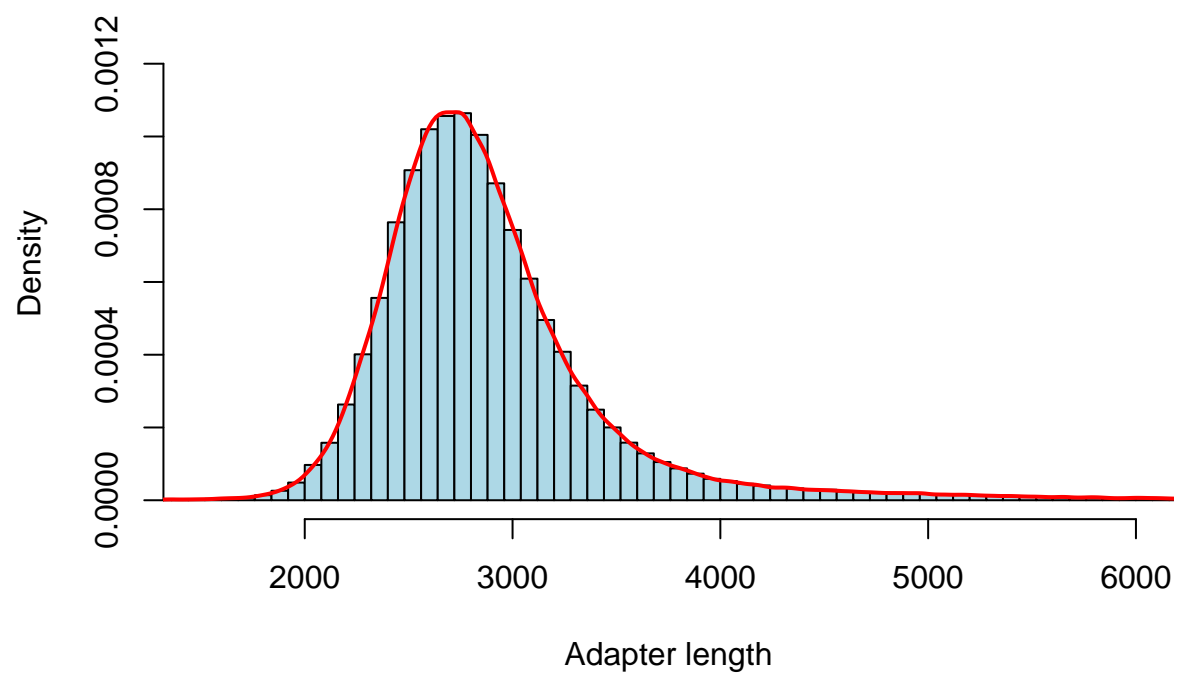
lines(density(adapter_length), col = "red", lwd = 2)
}

```

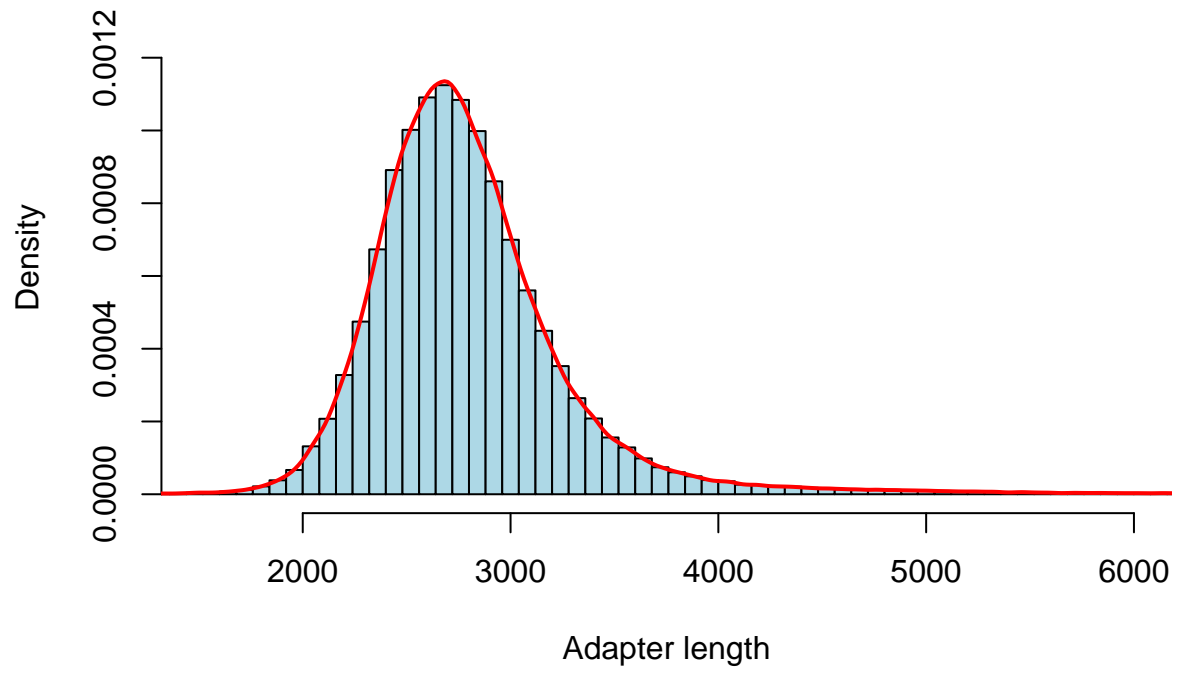
Density Plot of adapter lengths – a60_30



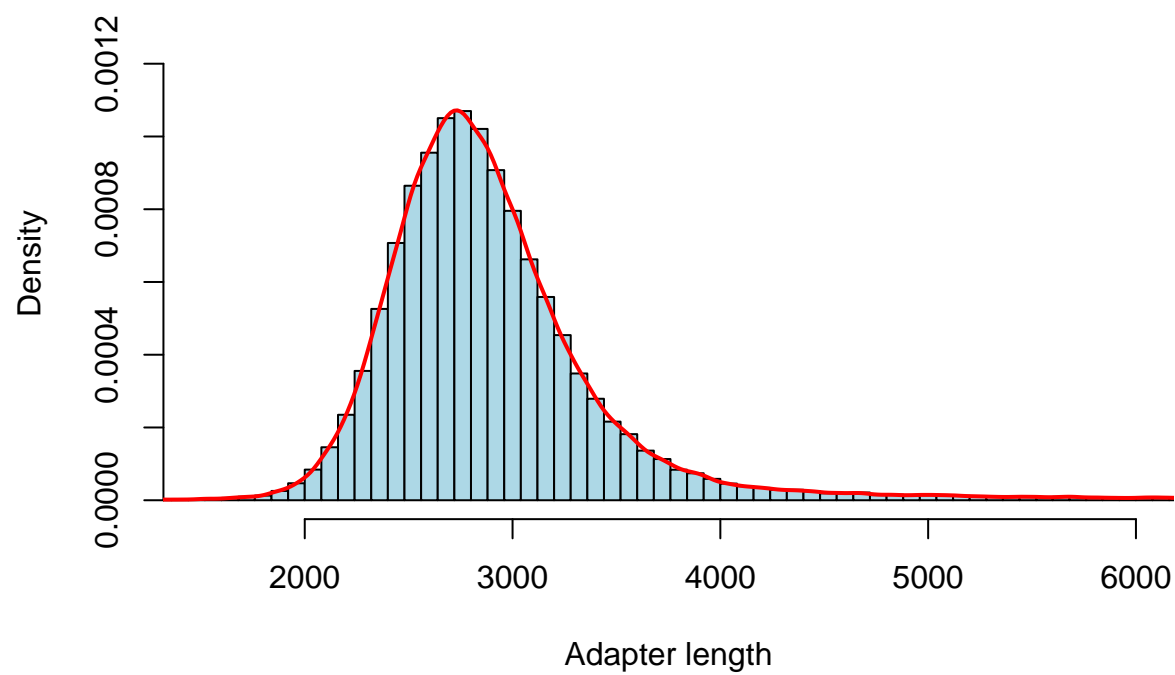
Density Plot of adapter lengths – a60_60



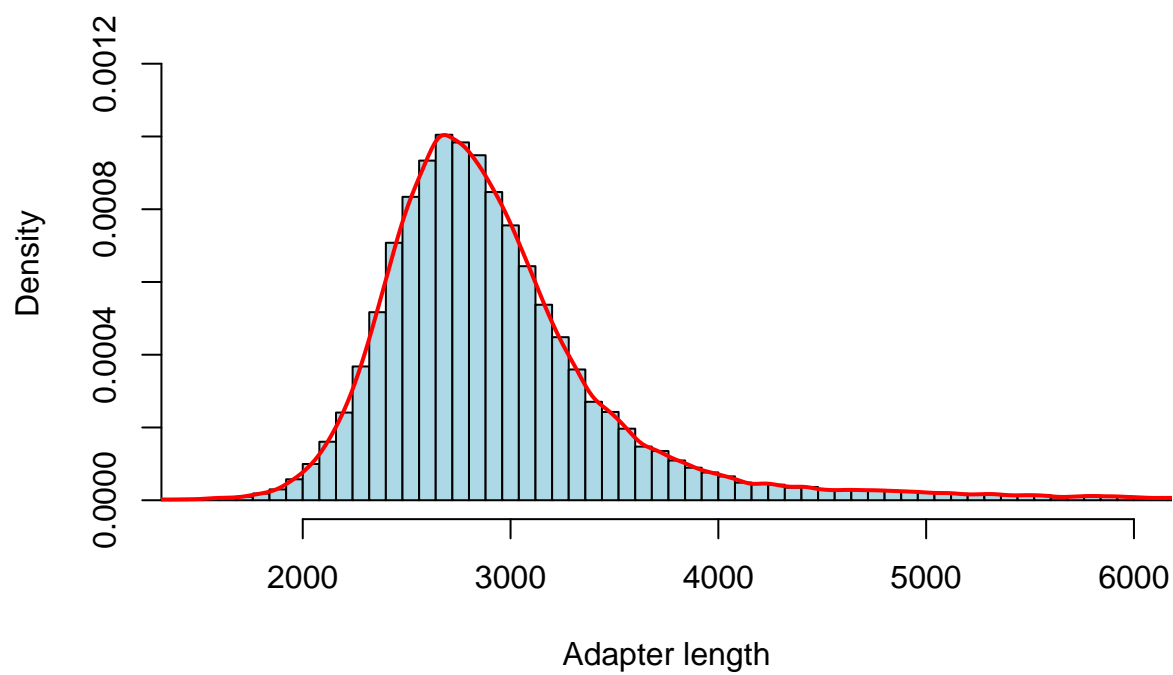
Density Plot of adapter lengths – a60_unmod



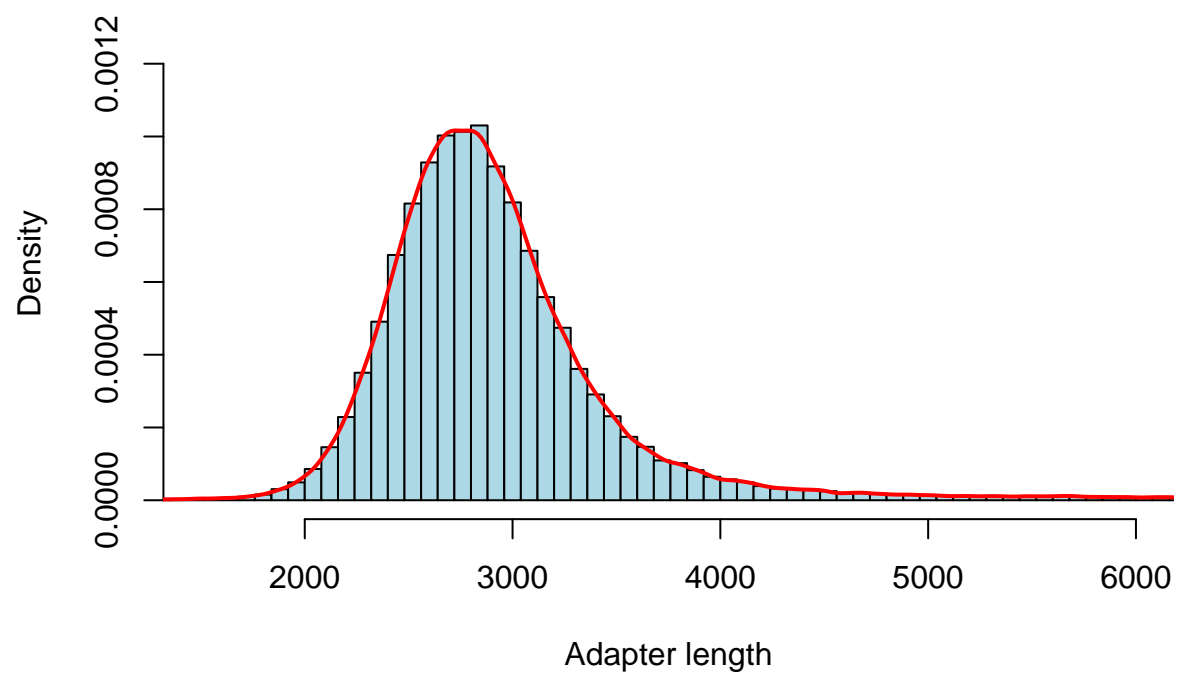
Density Plot of adapter lengths – a120_1mod



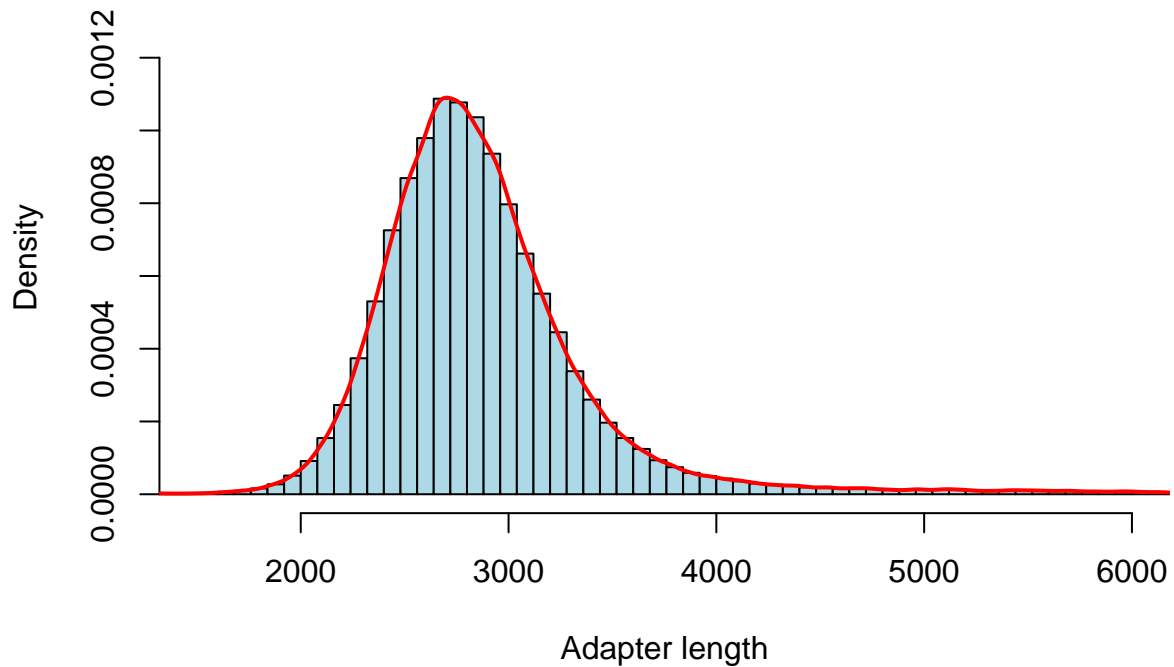
Density Plot of adapter lengths – a120_2mod



Density Plot of adapter lengths – a120_4mod



Density Plot of adapter lengths – a120_unmod



Histograms of all the data with log transformation

```
n_bins_visible <- 100
breaks_common <- seq(log(1), log(8000), length.out = n_bins_visible + 1)

for (name in names(datasets)) {
  adapter_length <- datasets[[name]]$adapter_length

  adapter_length <- na.omit(adapter_length)
  adapter_length <- adapter_length[adapter_length > 0 & adapter_length <= 8000]

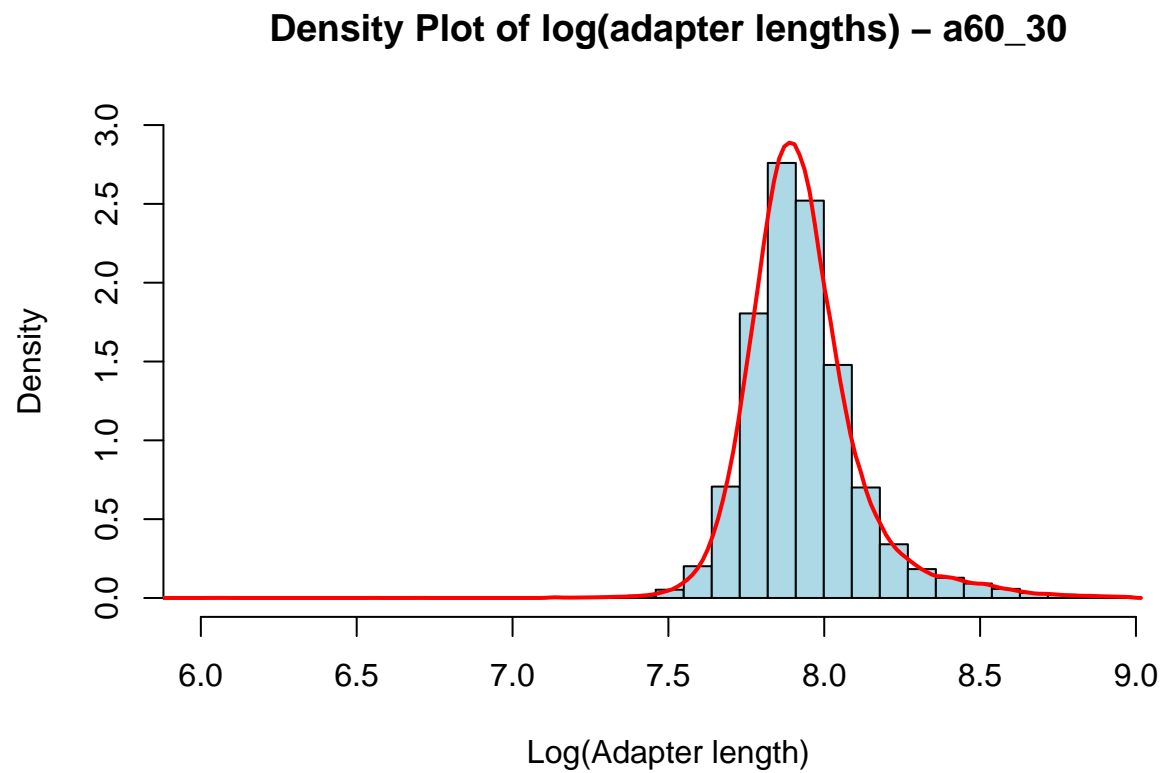
  if (length(adapter_length) == 0) next

  adapter_length <- log(adapter_length)

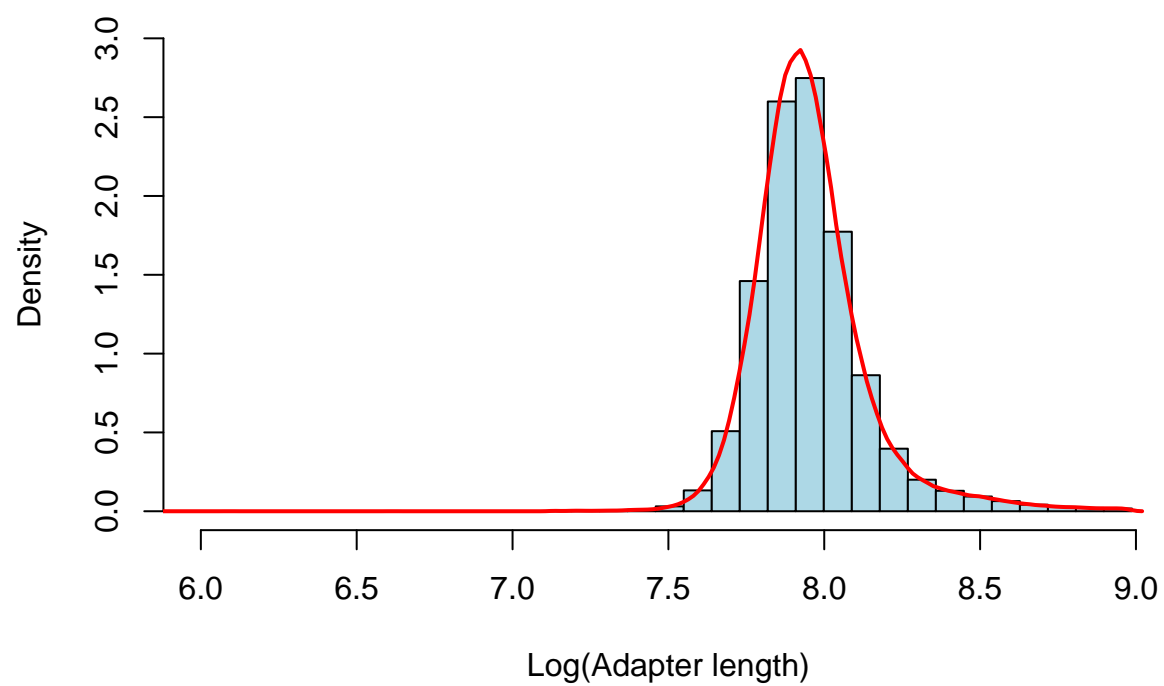
  hist(adapter_length,
        probability = TRUE,
        main = paste("Density Plot of log(adapter lengths) -", name),
        xlab = "Log(Adapter length)",
        ylab = "Density",
        xlim = c(6,9),
        ylim = c(0,3),
        col = "lightblue",
        border = "black",
```



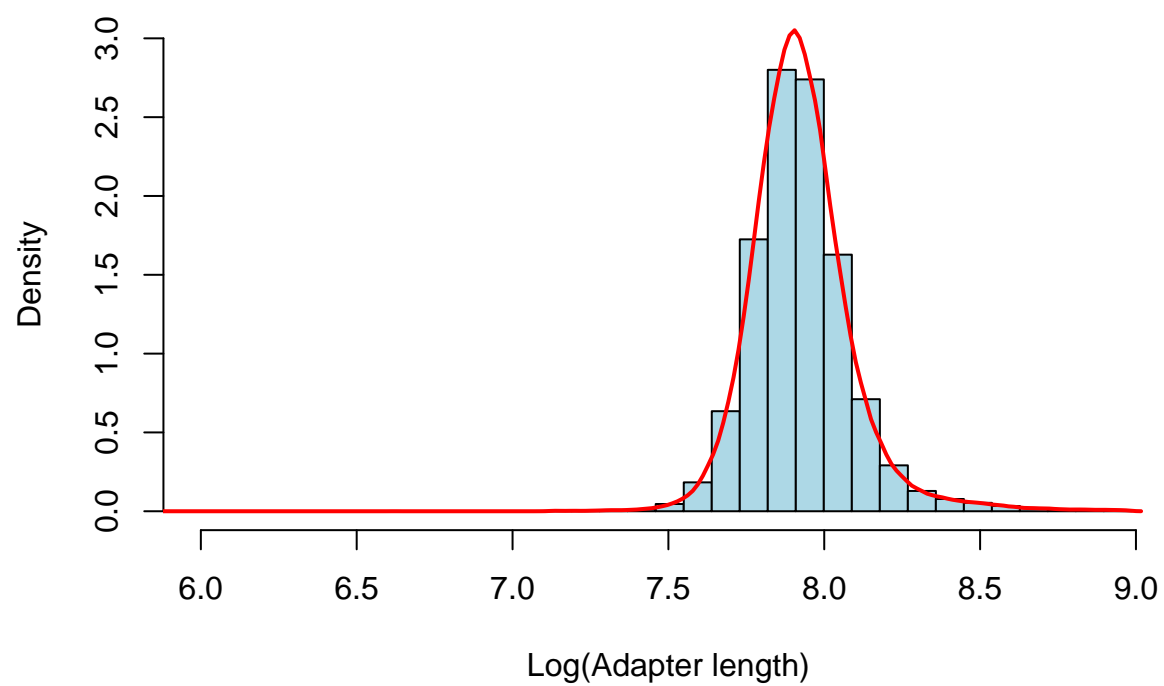
```
breaks = breaks_common)\n\nlines(density(adapter_length), col = "red", lwd = 2)\n}
```



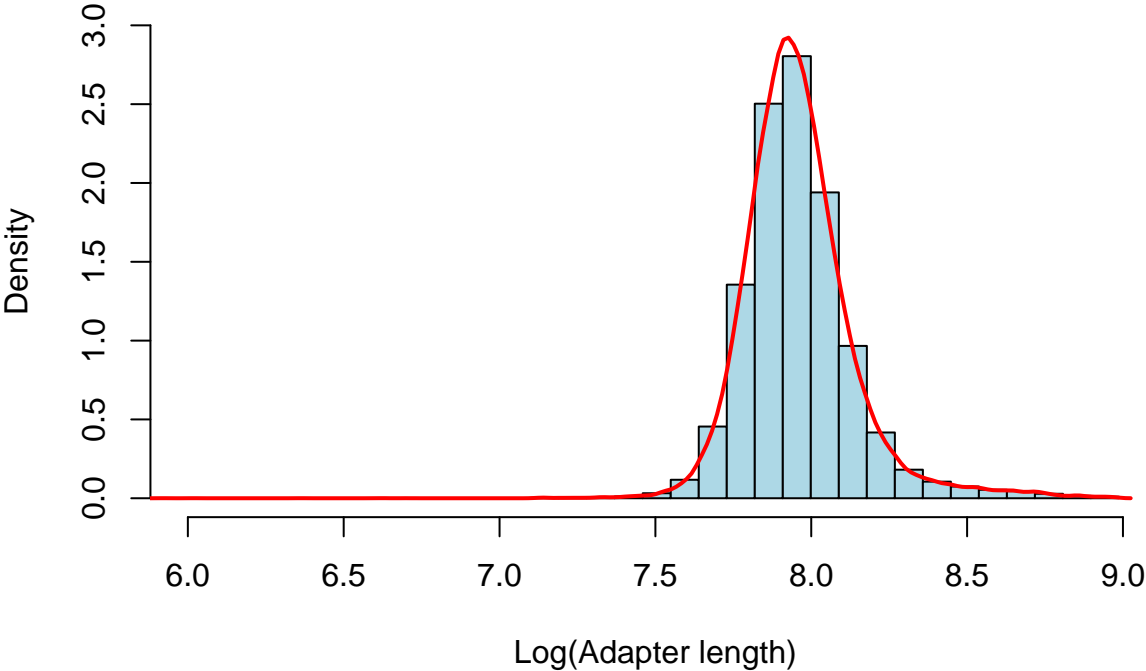
Density Plot of log(adapter lengths) – a60_60



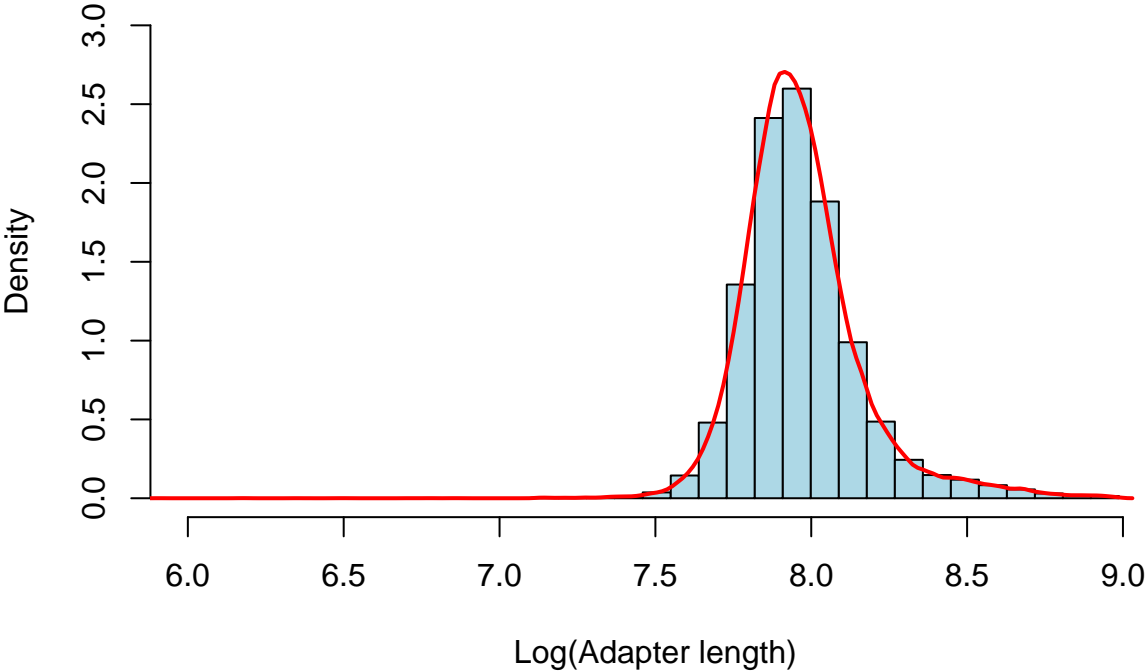
Density Plot of log(adapter lengths) – a60_unmod



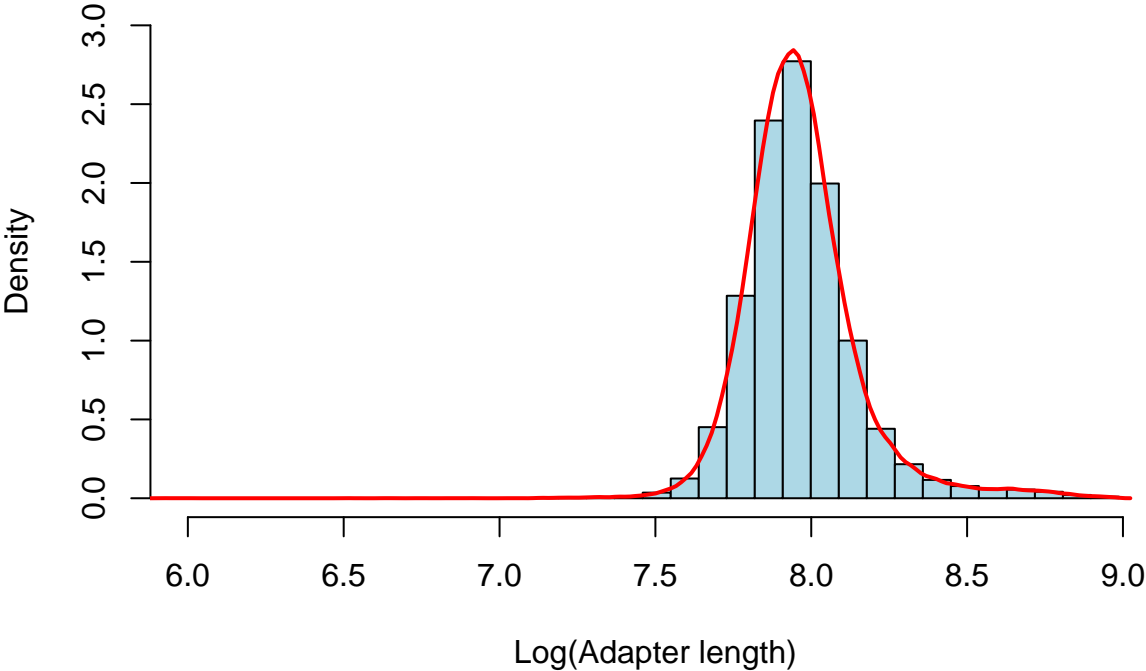
Density Plot of log(adapter lengths) – a120_1mod



Density Plot of log(adapter lengths) – a120_2mod



Density Plot of log(adapter lengths) – a120_4mod



Density Plot of log(adapter lengths) – a120_unmod

