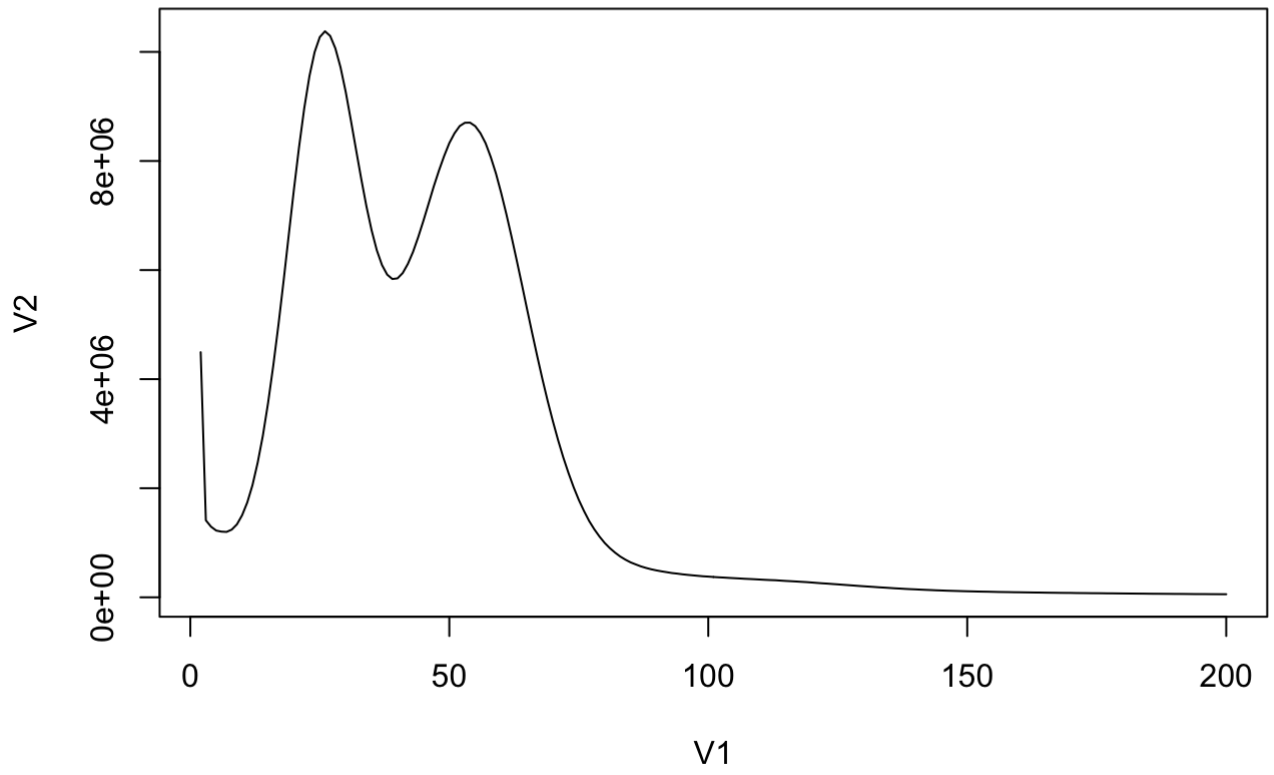
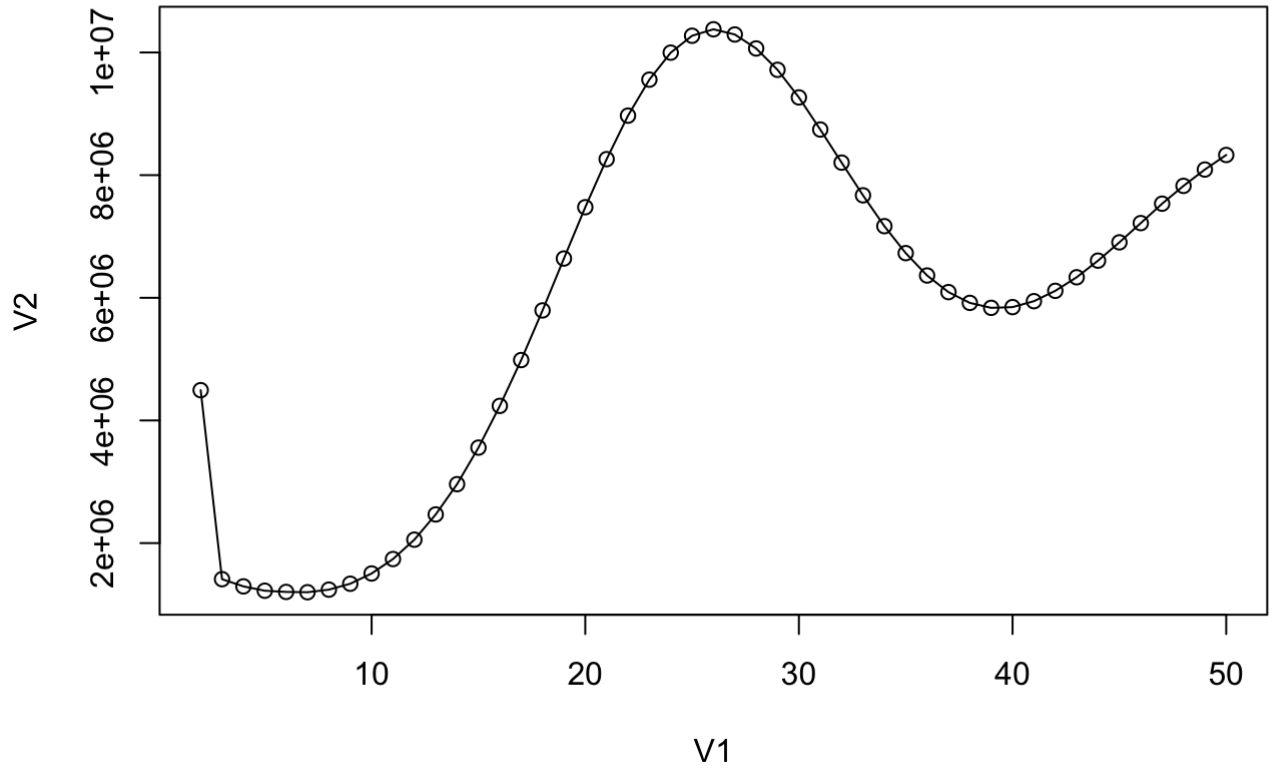
```
setwd("~/Desktop/JellyFish2")
```

Genome size estimation using kmer size 19

```
dataframe19 <- read.table("19mer_out.histo")
plot(dataframe19[2:200,], type="l")
```




```
plot(dataframe19[2:50,], type="l")
points(dataframe19[2:50,])
```



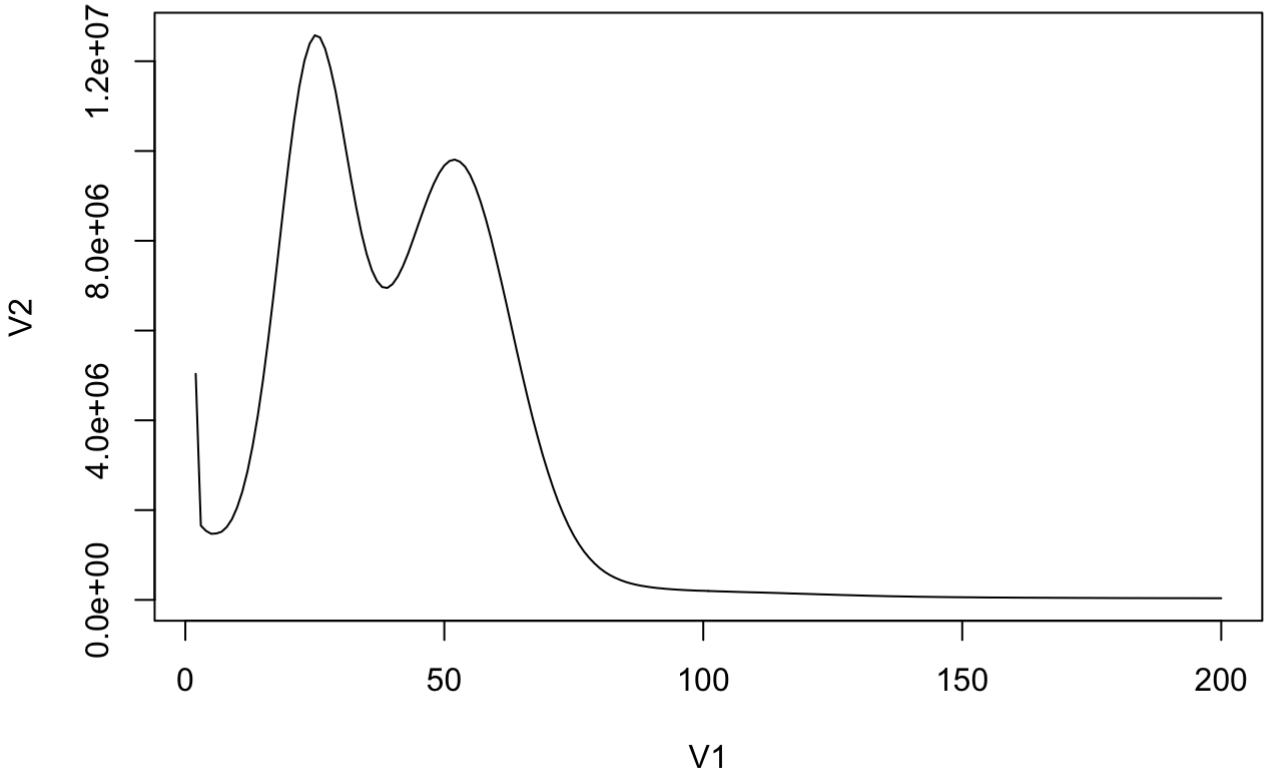

```
dataframe19[22:29,]
```

```
##    V1       V2
## 22 22  8969948
## 23 23  9556977
## 24 24  9997232
## 25 25 10272627
## 26 26 10378051
## 27 27 10293187
## 28 28 10065227
## 29 29  9717226
```

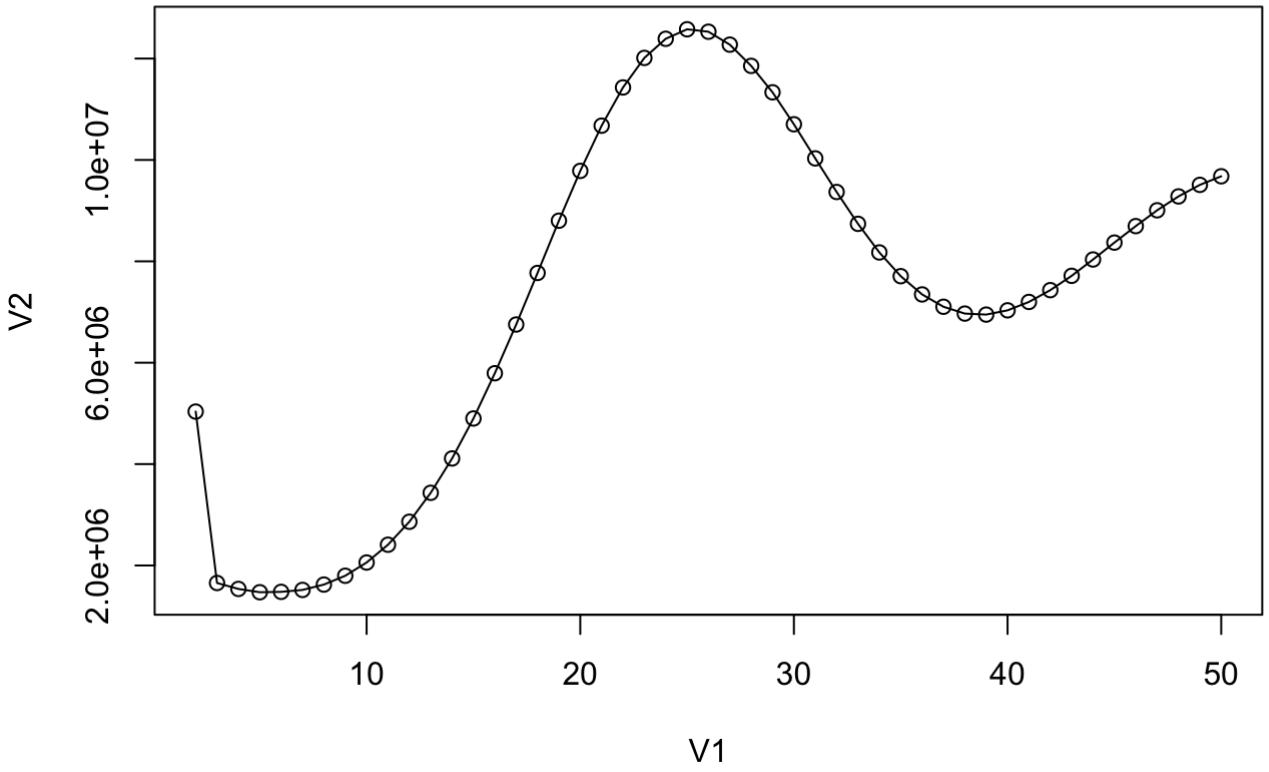
```
a <- sum(as.numeric(dataframe19[2:10000,1]*dataframe19[2:10000,2]))/26
```

Genome size estimation using kmer size 21

```
dataframe21 <- read.table("21mer_out.histo")
plot(dataframe21[2:200,], type="l")
```




```
plot(dataframe21[2:50,], type="l")
points(dataframe21[2:50,])
```



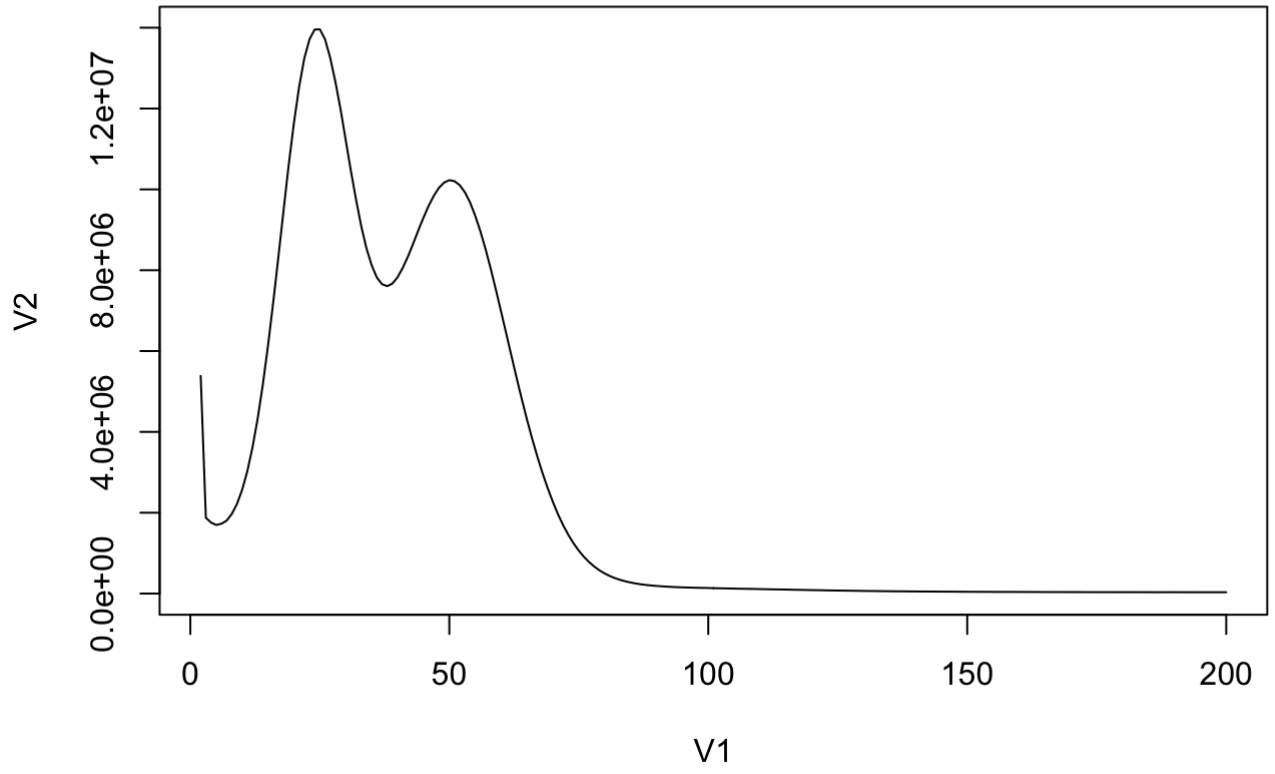

```
dataframe21[22:29,]
```

```
##    V1       V2
## 22 22 11430202
## 23 23 12012893
## 24 24 12390128
## 25 25 12577573
## 26 26 12529493
## 27 27 12274426
## 28 28 11855744
## 29 29 11334373
```

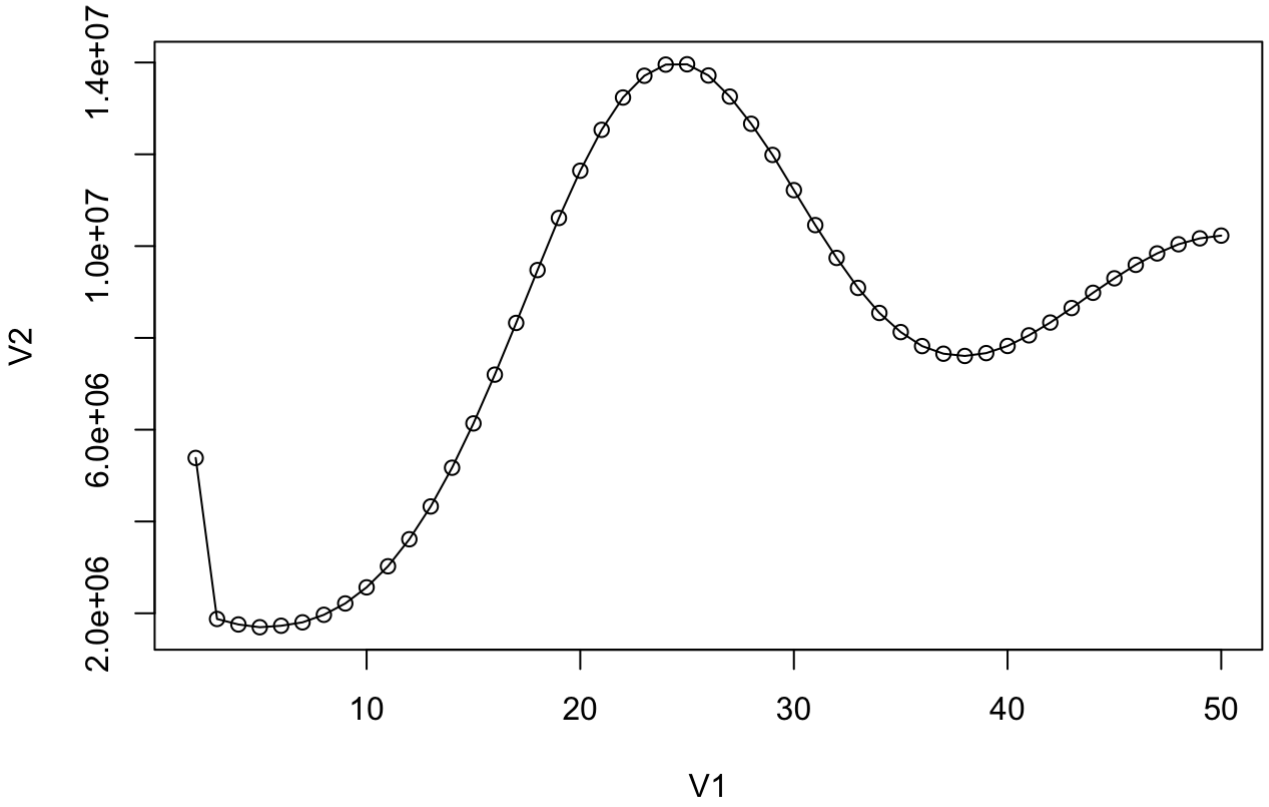
```
b <- sum(as.numeric(dataframe21[2:10000,1]*dataframe21[2:10000,2]))/25
```

Genome size estimation using kmer size 23

```
dataframe23 <- read.table("23mer_out.histo")
plot(dataframe23[2:200,], type="l")
```




```
plot(dataframe23[2:50,], type="l")
points(dataframe23[2:50,])
```



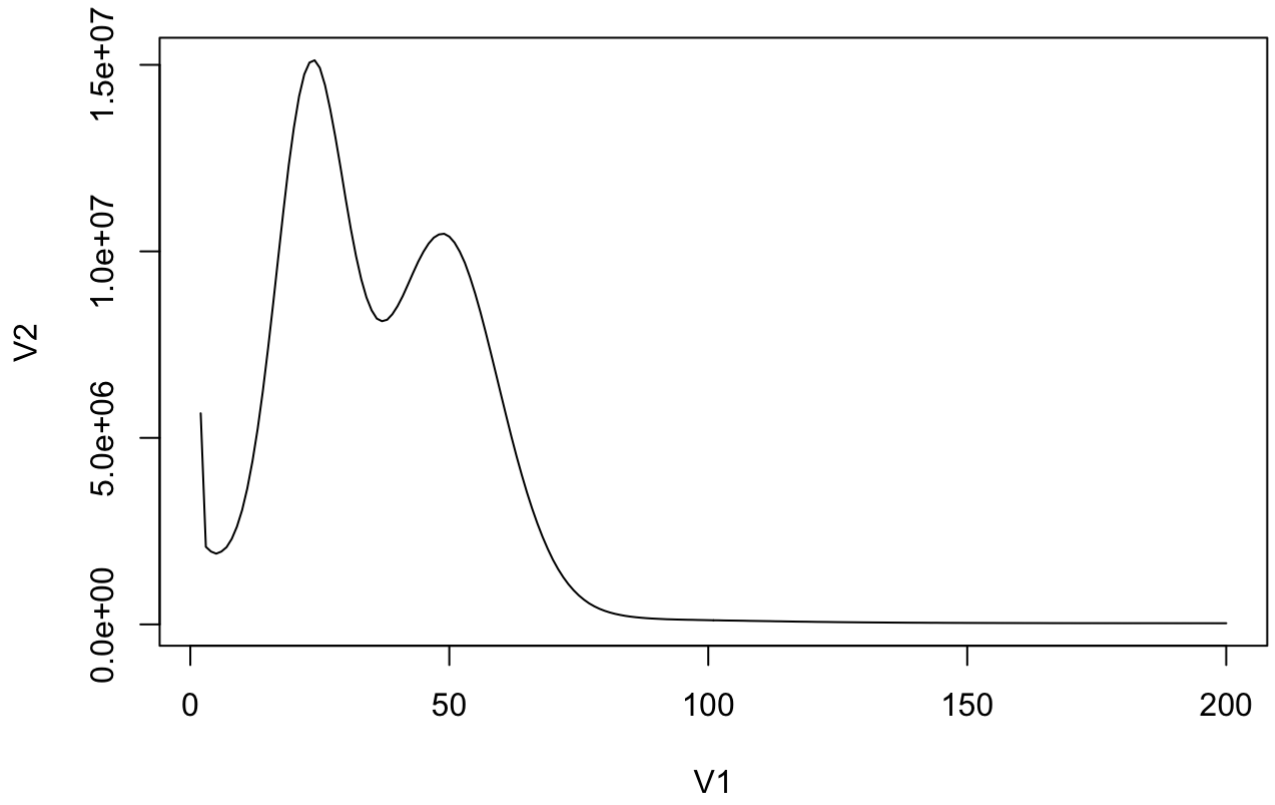

```
dataframe23[22:29,]
```

```
##    V1       V2
## 22 22 13237047
## 23 23 13711104
## 24 24 13954308
## 25 25 13961126
## 26 26 13717190
## 27 27 13257090
## 28 28 12664852
## 29 29 11986311
```

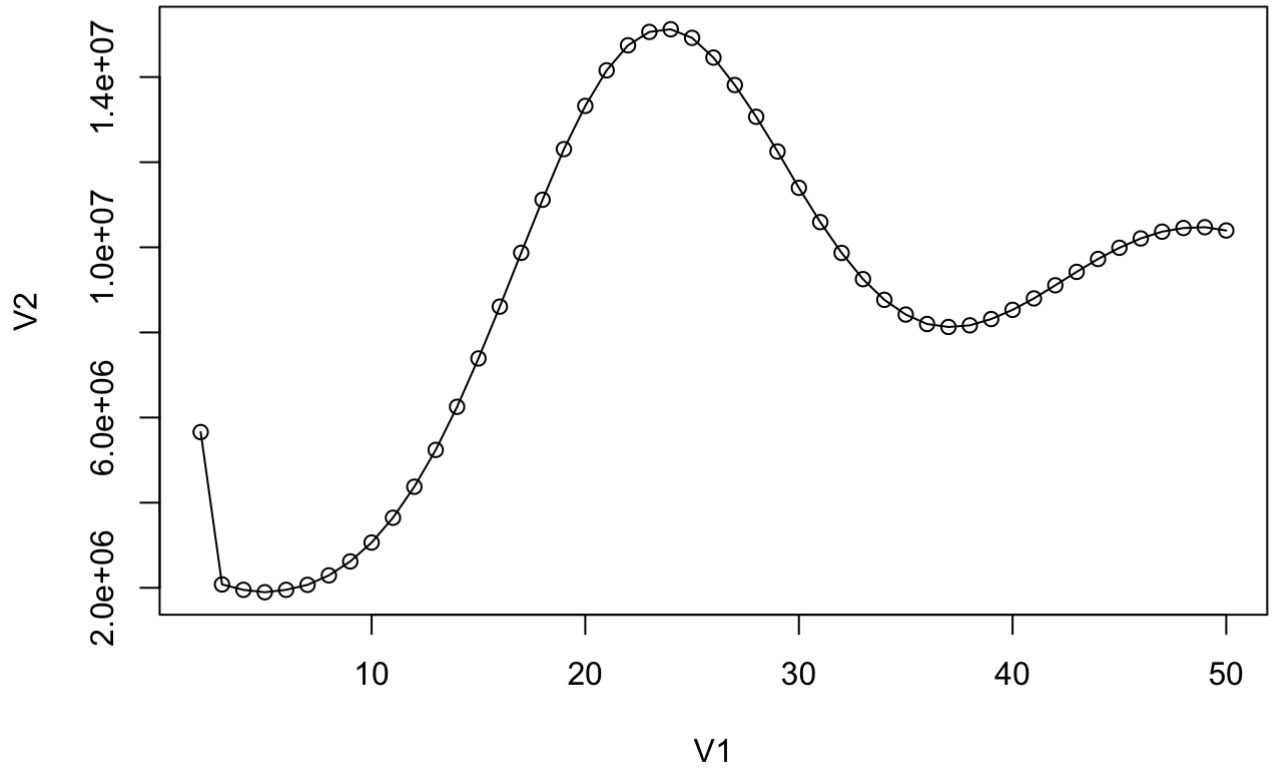
```
c <- sum(as.numeric(dataframe23[2:10000,1]*dataframe23[2:10000,2]))/25
```

Genome size estimation using kmer size 25

```
dataframe25 <- read.table("25mer_out.histo")
plot(dataframe25[2:200,], type="l")
```




```
plot(dataframe25[2:50,], type="l")
points(dataframe25[2:50,])
```



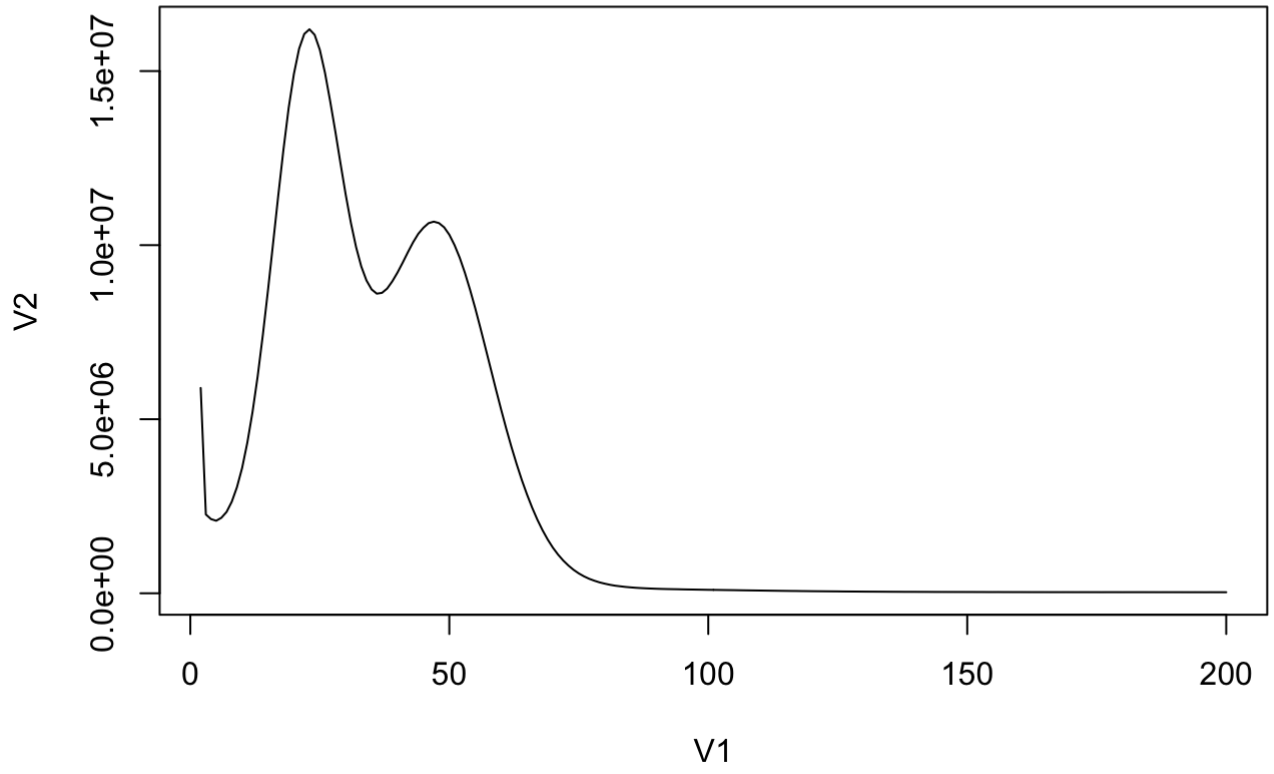

```
dataframe25[22:29,]
```

```
##    V1       V2
## 22 22 14746160
## 23 23 15060548
## 24 24 15122493
## 25 25 14920963
## 26 26 14457627
## 27 27 13812782
## 28 28 13066500
## 29 29 12251534
```

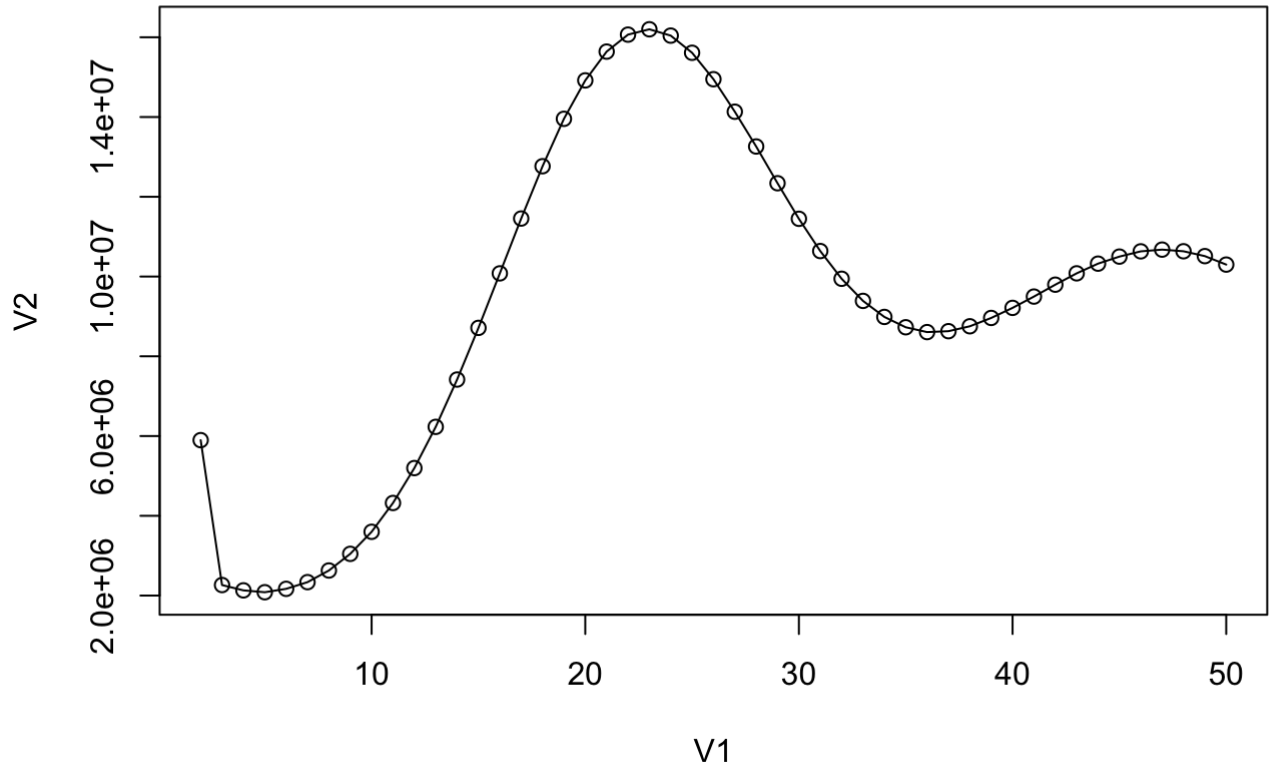
```
d <- sum(as.numeric(dataframe25[2:10000,1]*dataframe25[2:10000,2]))/24
```

Genome size estimation using kmer size 27

```
dataframe27 <- read.table("27mer_out.histo")
plot(dataframe27[2:200,], type="l")
```




```
plot(dataframe27[2:50,], type="l")
points(dataframe27[2:50,])
```



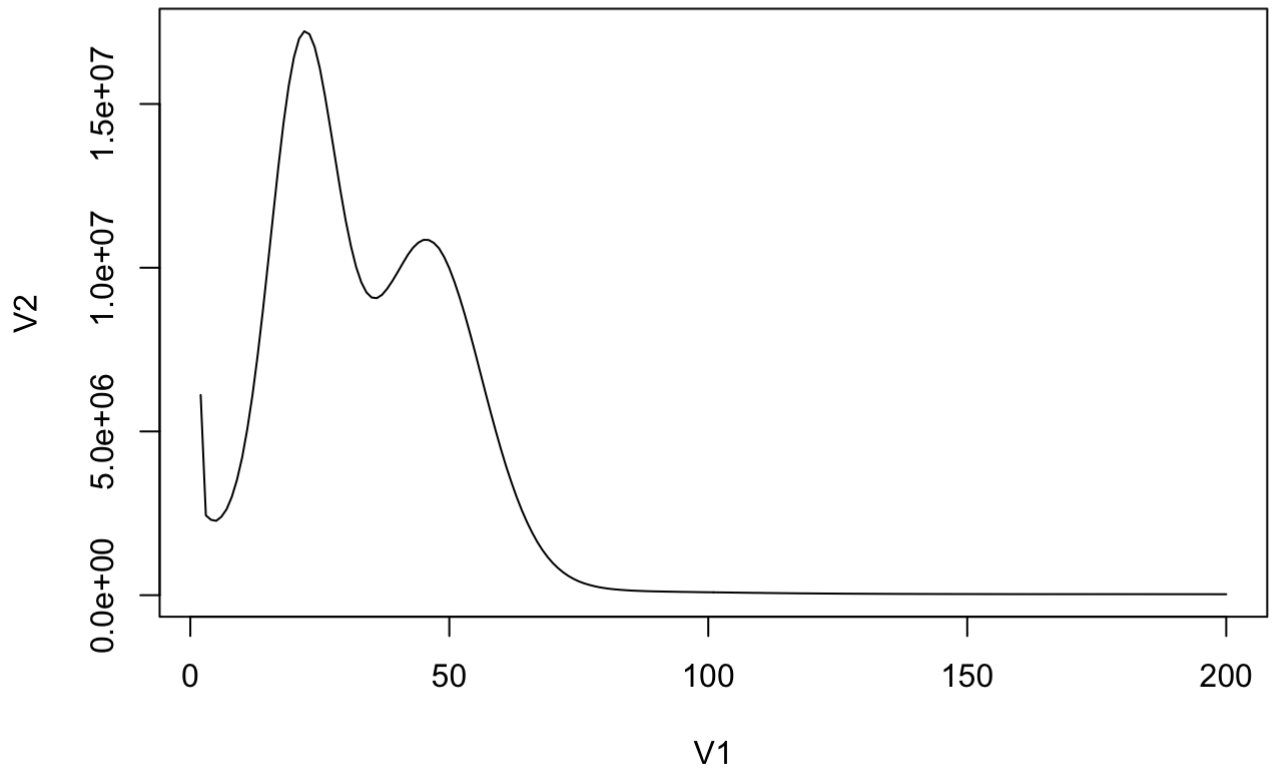

```
dataframe27[22:29,]
```

```
##    V1       V2
## 22 22 16066268
## 23 23 16200257
## 24 24 16041407
## 25 25 15611227
## 26 26 14948401
## 27 27 14133390
## 28 28 13261229
## 29 29 12339859
```

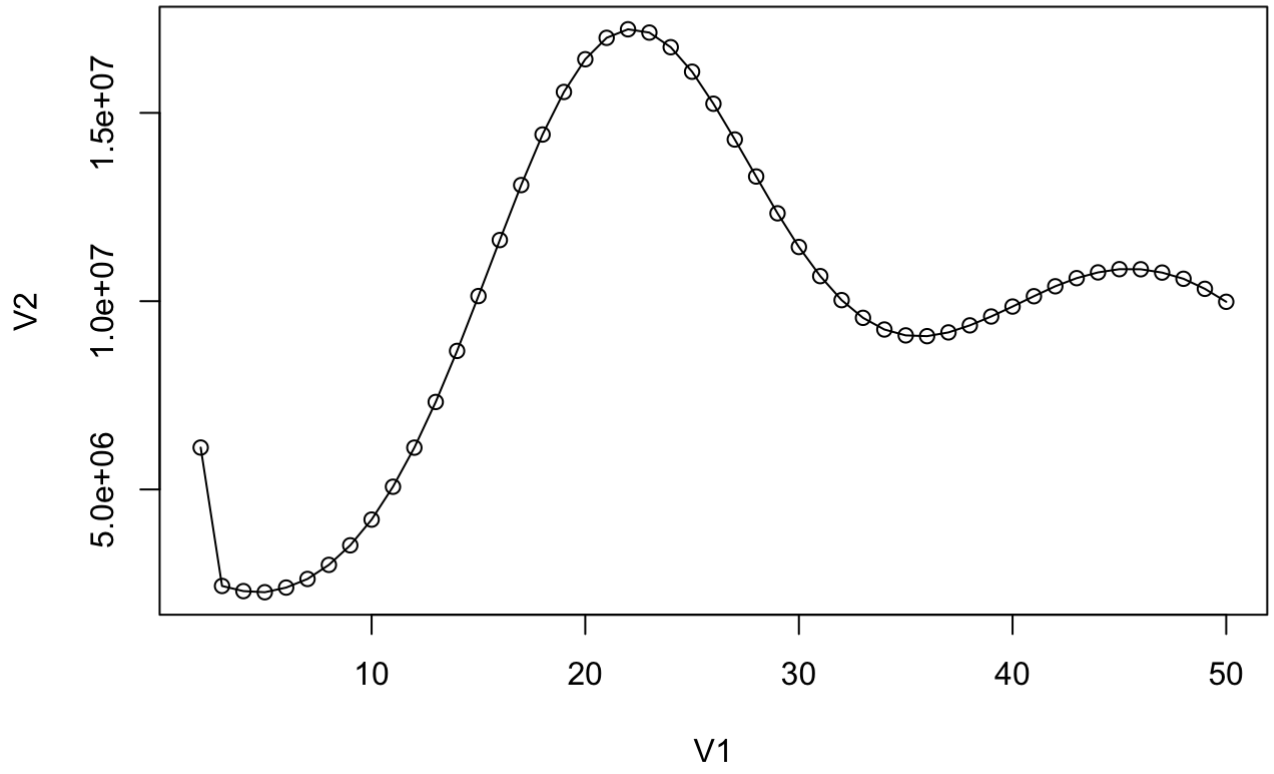
```
e <-sum(as.numeric(dataframe27[2:10000,1]*dataframe27[2:10000,2]))/23
```

Genome size estimation using kmer size 29

```
dataframe29 <- read.table("29mer_out.histo")
plot(dataframe29[2:200,], type="l")
```




```
plot(dataframe29[2:50,], type="l")
points(dataframe29[2:50,])
```



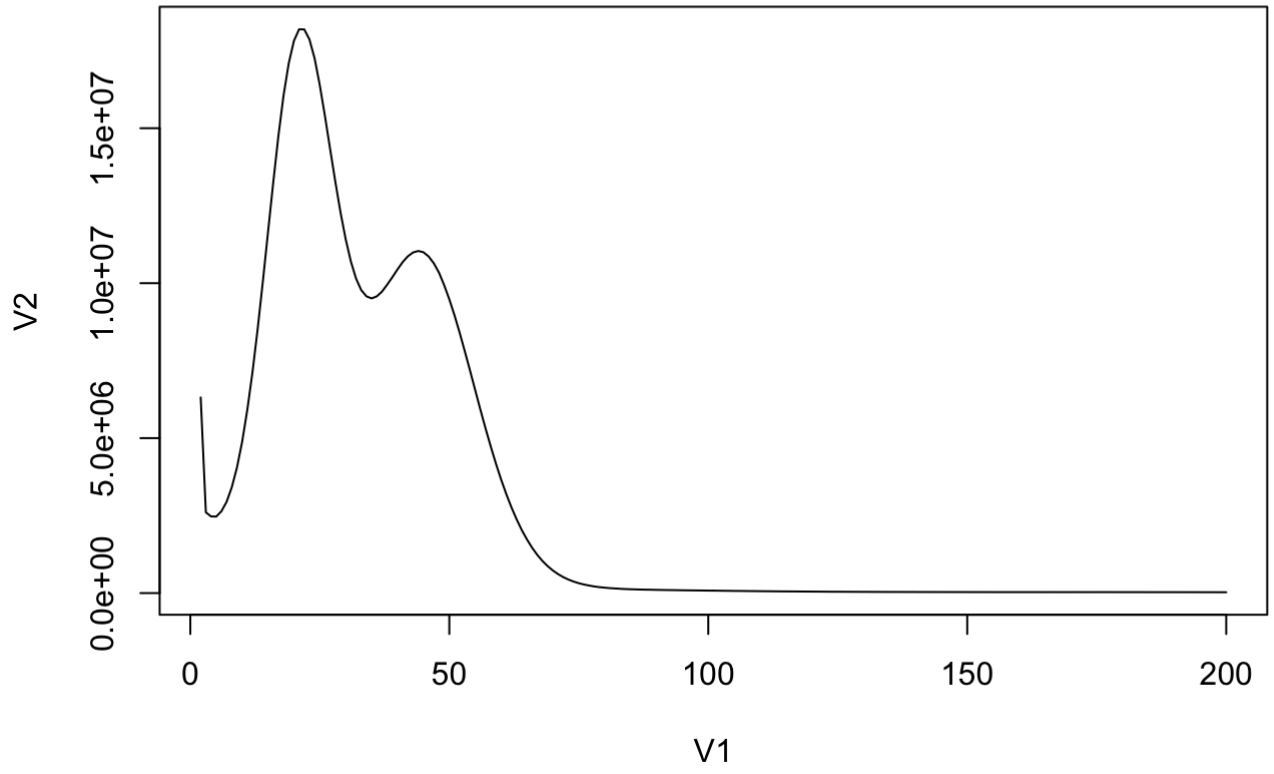

```
dataframe29[20:27,]
```

```
##    V1       V2
## 20 20 16424329
## 21 21 16992575
## 22 22 17219622
## 23 23 17131897
## 24 24 16743861
## 25 25 16093097
## 26 26 15242897
## 27 27 14293131
```

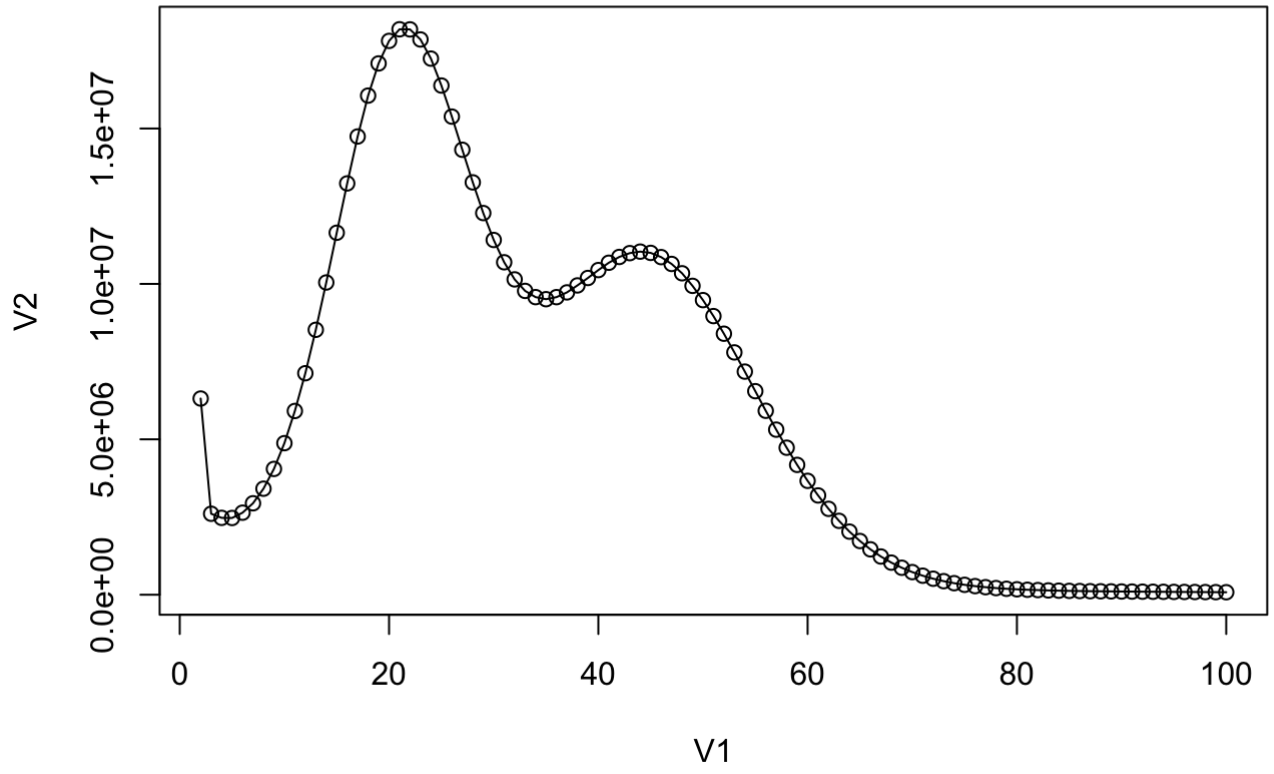
```
f <- sum(as.numeric(dataframe29[2:10000,1]*dataframe29[2:10000,2]))/22
```

Genome size estimation using kmer size 31

```
dataframe31 <- read.table("31mer_out.histo")
plot(dataframe31[2:200,], type="l")
```




```
plot(dataframe31[2:100,], type="l")
points(dataframe31[2:100,])
```

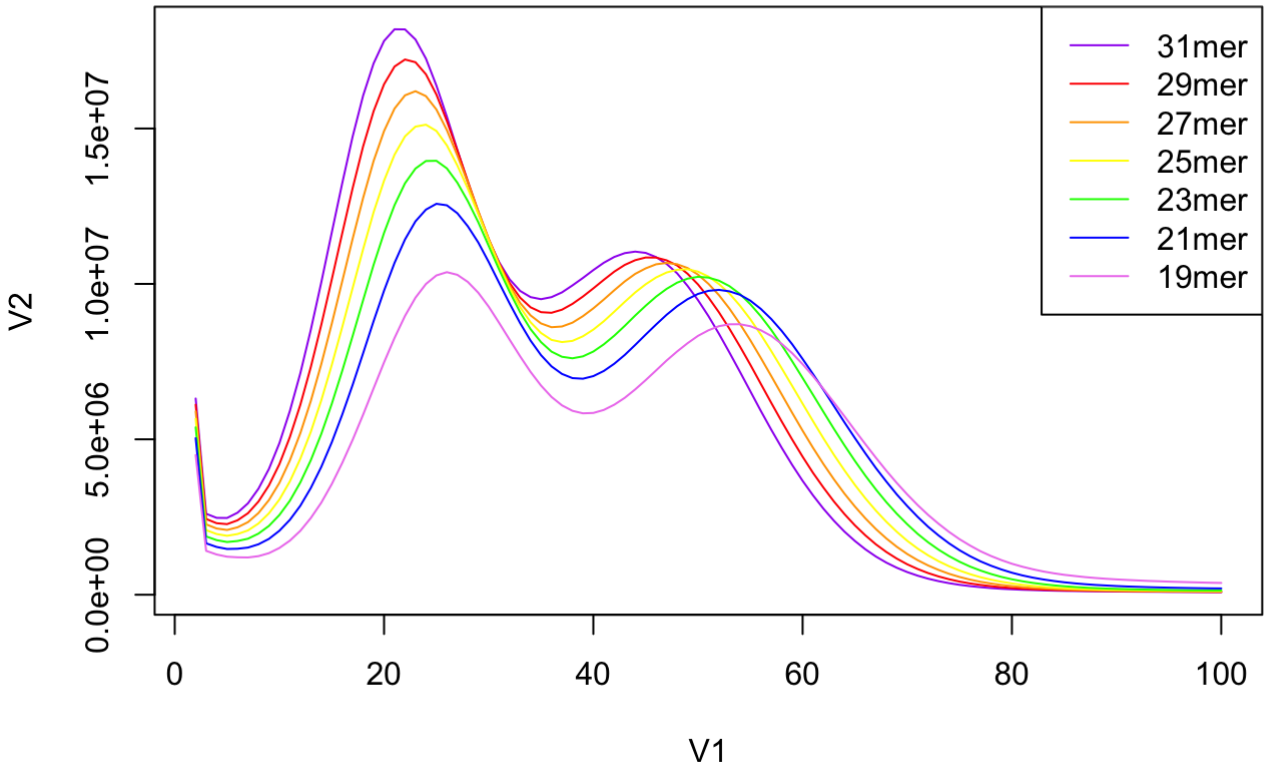



```
dataframe31[20:37,]
```

```
##    V1       V2
## 20 20 17817944
## 21 21 18191868
## 22 22 18188347
## 23 23 17863722
## 24 24 17248253
## 25 25 16381947
## 26 26 15382315
## 27 27 14315735
## 28 28 13267281
## 29 29 12279867
## 30 30 11413846
## 31 31 10696314
## 32 32 10145720
## 33 33  9775426
## 34 34  9575348
## 35 35  9508708
## 36 36  9574711
## 37 37  9727387
```

```
g <- sum(as.numeric(dataframe31[2:10000,1]*dataframe31[2:10000,2]))/22
```

```
plot(dataframe31[2:100,], type="l", col="purple")
lines(dataframe29[2:100,], col="red")
lines(dataframe27[2:100,], col="orange")
lines(dataframe25[2:100,], col = "yellow")
lines(dataframe23[2:100,], col = "green")
lines(dataframe21[2:100,], col = "blue")
lines(dataframe19[2:100,], col = "violet")
legend("topright",legend=c("31mer", "29mer","27mer", "25mer", "23mer", "21mer",
"19mer"), lty=c(1,1), col=c( "purple", "red", "orange", "yellow", "green", "blue", "viol
et"))
```



Make a table of results

```
require(data.table)
```

```
## Loading required package: data.table
```

```
DT = data.table("kmer" = c("19mer", "21mer", "23mer", "25mer", "27mer", "29mer",
"31mer"), "size estimate" = c(a, b, c, d, e, f, g))

DT
```

```
##     kmer size estimate
## 1: 19mer    1072526962
## 2: 21mer    1092265476
## 3: 23mer    1068252581
## 4: 25mer    1087284540
## 5: 27mer    1107731692
## 6: 29mer    1130064418
## 7: 31mer    1102271949
```