test1

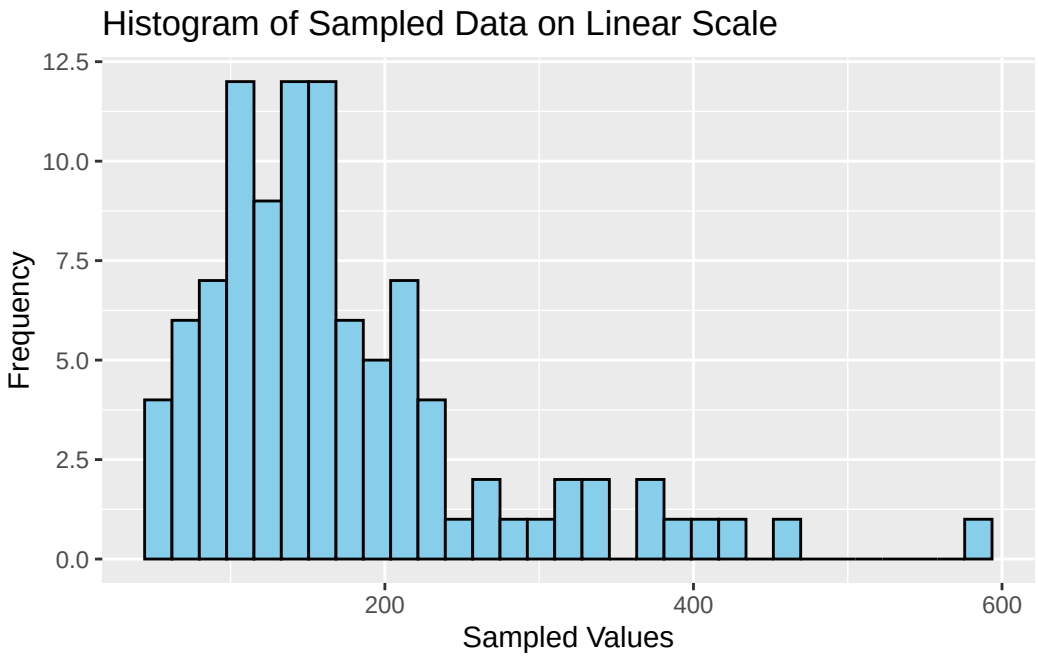
Week 1 Exercise Part B

To generate 100 values from a log-normal distribution

```
library(ggplot2)
set.seed(33)
sampled_values <- rlnorm(100,5,0.5)
```

Linear scale histogram

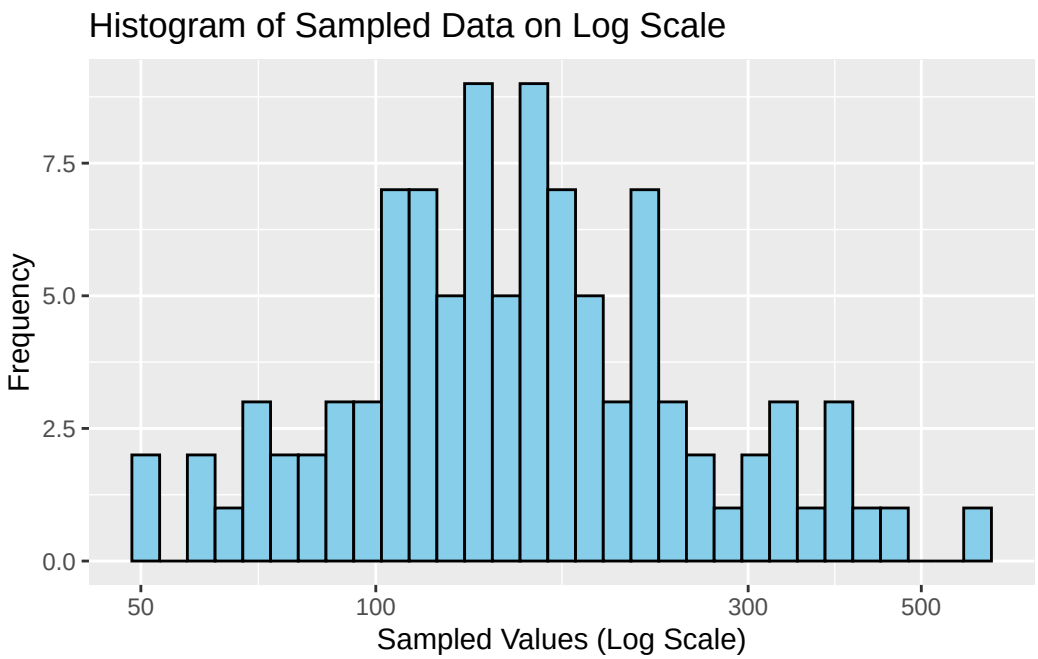
```
# Create a histogram of the sampled data on a linear scale
linear_plot <- ggplot(data.frame(sampled_values), aes(x = sampled_values)) +
  geom_histogram(binwidth = diff(range(sampled_values)) / 30,
                 fill = "skyblue",
                 color = "black") +
  labs(title = "Histogram of Sampled Data on Linear Scale",
       x = "Sampled Values",
       y = "Frequency")
print(linear_plot)
```



Form the plot, we can find that most data under 200 and very few data more than 220.

Log scale histgram

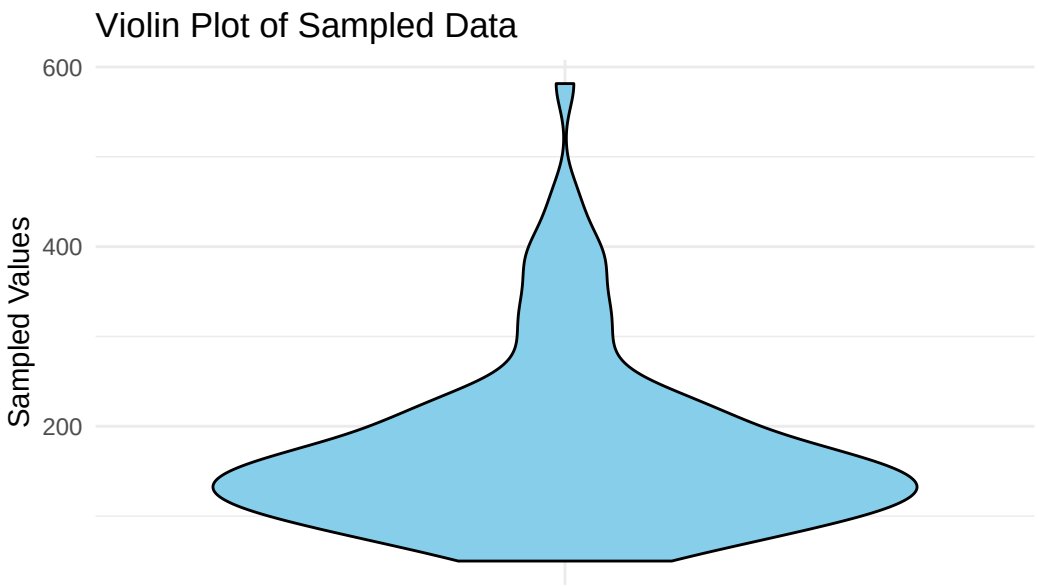
```r
# Create a histogram of the sampled data on a log scale
log_plot <- ggplot(data.frame(sampled_values), aes(x = sampled_values)) +
  geom_histogram(binwidth = diff(log10(range(sampled_values))) / 30,
                 fill = "skyblue",
                 color = "black") +
  scale_x_log10() +
  labs(title = "Histogram of Sampled Data on Log Scale",
       x = "Sampled Values (Log Scale)",
       y = "Frequency")
print(log_plot)
```



Because we changed linear to log scale, we have a very different shape. Most data is around 100-300. Also values smaller than 100 and values bigger than 300 don't have a big difference.

Violin plot

```
# Create a violin plot of the sampled data
violin_plot <- ggplot(data.frame(sampled_values), aes(x = "", y = sampled_values)) +
  geom_violin(fill = "skyblue", color = "black") +
  labs(title = "Violin Plot of Sampled Data",
       x = "",
       y = "Sampled Values") +
  theme_minimal()
print(violin_plot)
```



From the violin plot, we can know the data shape directly.

To get the mean and standard deviation

```
summary(sampled_values)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  49.97  110.93  149.82  174.03  211.40  581.48 
```

```
sd(sampled_values)
```

```
[1] 96.67352
```

The mean of the 100 values is 174.03 and the standard deviation is 96.67352.