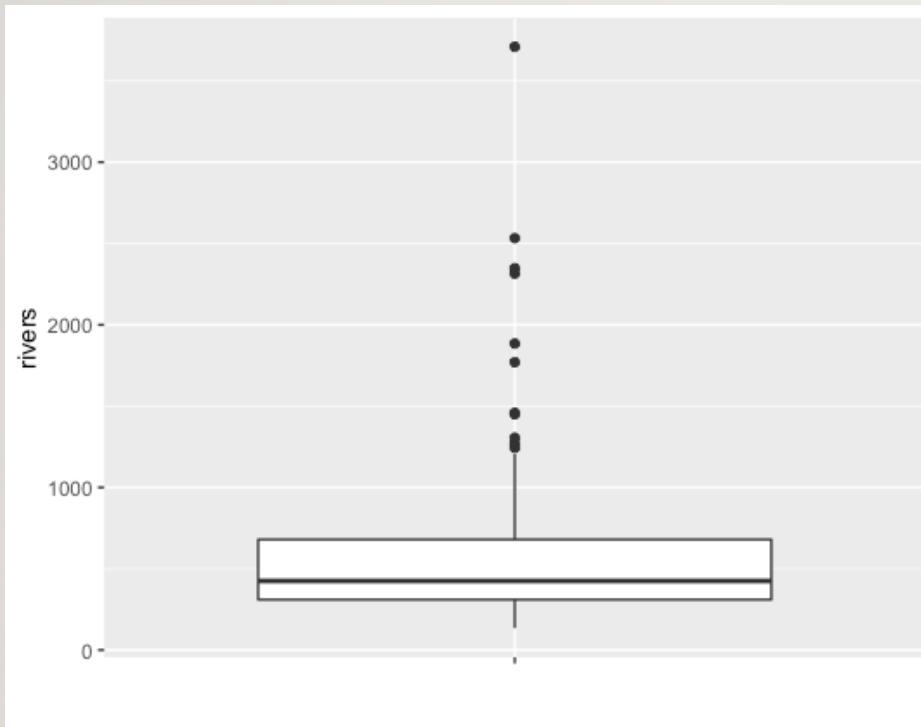
A wide-angle aerial photograph showing a river flowing through a dense, green forest. The river curves and splits into smaller streams, creating a complex network of waterways. The surrounding land is covered in lush vegetation, with the river providing a clear blue contrast.

MAJOR RIVERS IN NORTH AMERICA

STATISTICAL PROGRAMMING IN R: LESSON 5 HANDS-ON

AUTUMN HEYMAN

MAJOR RIVERS IN NORTH AMERICA BOX PLOT



Summary

Min	I st Quar	Median	Mean	3 rd Quar	Max	IQR
135.0	310.0	425.0	591.2	680.0	3710.0	370

Outliers

Upper Range	Lower Range
1235	-60

Several data points are clearly above the upper range limit of 1235.

A few data points are near the upper range limit.

No data points are below the lower range limit.

MAJOR RIVERS IN NORTH AMERICA

BOX PLOT: R CODE

```
rr = data.frame(rivers)

head(rr)

l <- ggplot(rr, aes(x = "", y = rivers))

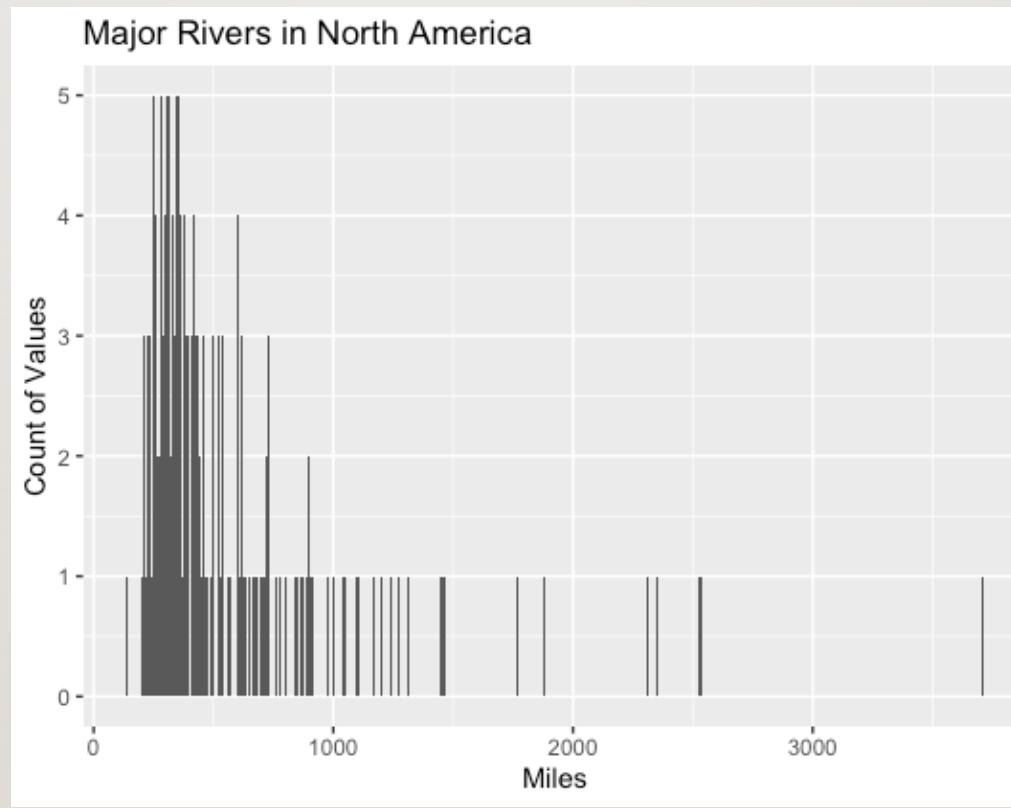
l + geom_boxplot() + xlab("")

summary(rr)

IQR <- 680-310

Outlier.Range <- IQR * 1.5
```

MAJOR RIVERS IN NORTH AMERICA HISTOGRAM



MAJOR RIVERS IN NORTH AMERICA

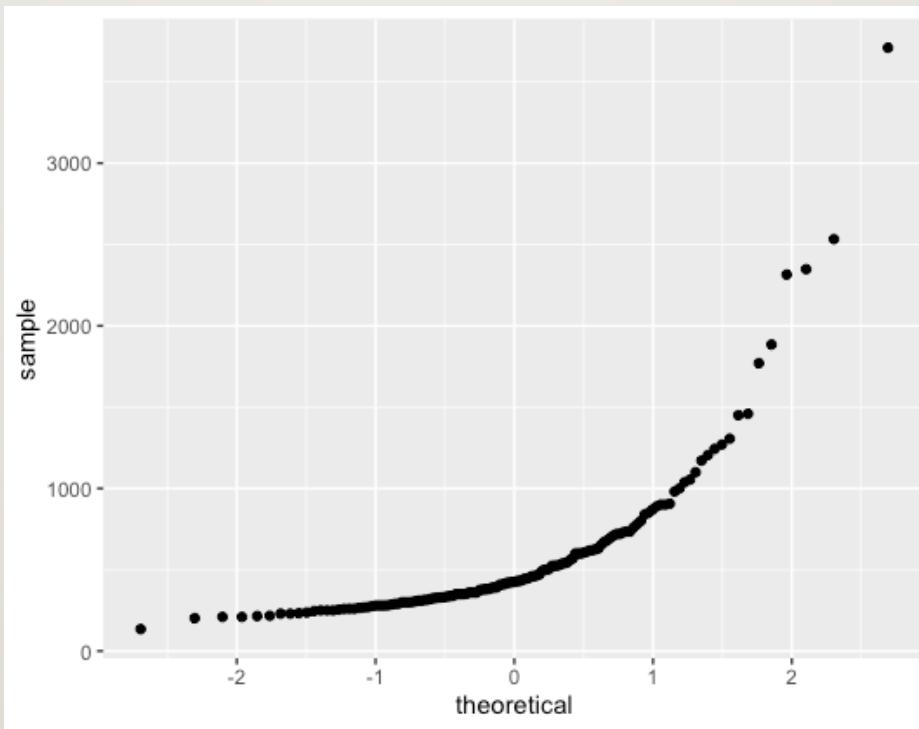
HISTOGRAM: R CODE

```
rr = data.frame(rivers)

m <- ggplot(rr, aes(x = rivers))

m + geom_histogram(binwidth=10) + ggtitle("Major Rivers in North America") +
  xlab("Miles") + ylab ("Count of Values")
```

MAJOR RIVERS IN NORTH AMERICA NORMAL PROBABILITY PLOT



MAJOR RIVERS IN NORTH AMERICA

NORMAL PROBABILITY PLOT: R CODE

```
rr = data.frame(rivers)  
ggplot(rr, aes(sample = rivers)) + geom_qq()
```

MAJOR RIVERS IN NORTH AMERICA: SUMMARY OF FINDINGS

Outlier Data:

Above Upper Range Limit: 11 data points, ranging from 1243-3710

Below Lower Range Limit: No data points fall below the lower limit.

Normal Distribution: NO

Evidence

Histogram: The histogram does not have the shape of a normal distribution. It does not depict the bell-shaped curve indicative of a normal distribution.

Normal Probability Plot: The data in the normal probability plot does not fall on a straight line. The data does not come from a normal distribution.