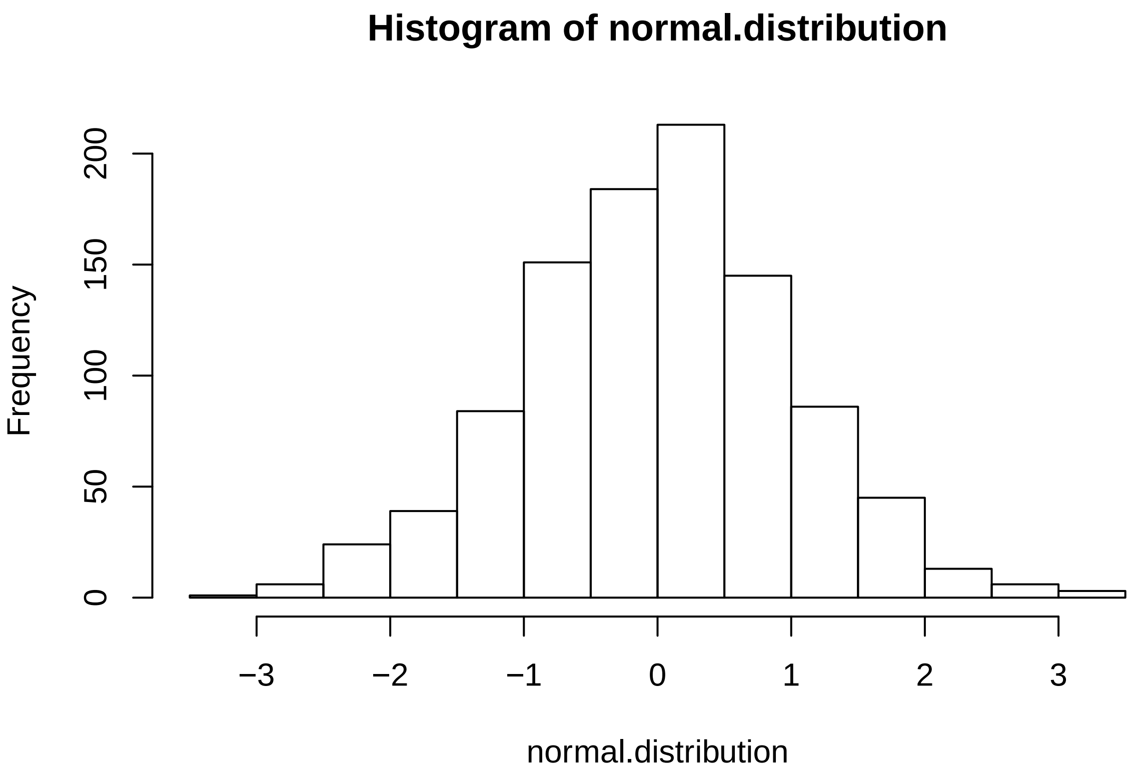
Homework 4 - bootstrap

Bootstrap Normal Distribution

First we create the normal distribution called `normal.distribution`

```
normal.distribution = rnorm(1000)
```

We would expect the histogram to already show a normal distribution



Bootstrap function

Now we will introduce a bootstrap function that will take the `distribution` as its only argument and will return a new distribution based on creating 1000 means from 50 random samples out of the original distribution.

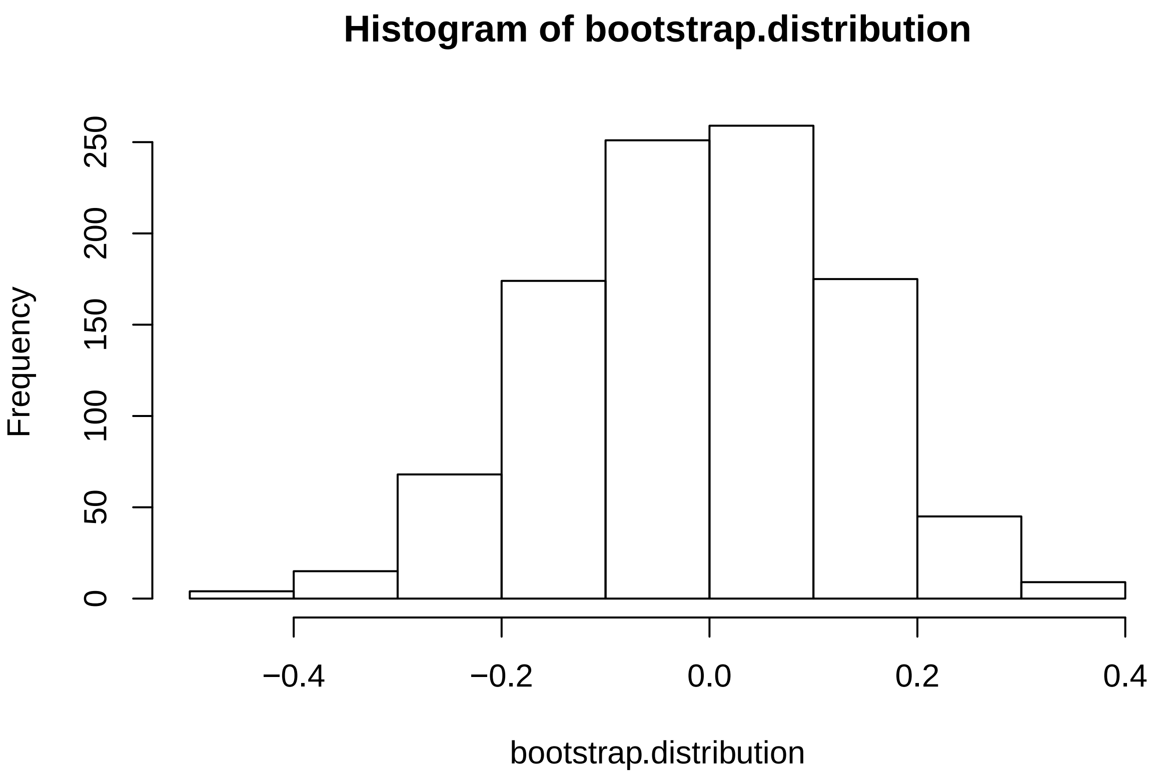
```
bootstrap = function(distribution) {
  # sample size
  n = 50
  #number of simulations
  nsim = 1000
  lotsa.means = numeric(nsim)
  for (i in 1:nsim) {
    x = sample(distribution, 50)
    lotsa.means[i] = mean(x)
  }
```

```
  lotsa.means
}
```

Now, let's see what happens when we apply the `bootstrap` method against our `normal.distribution` variable and graph the results:

```
bootstrap.distribution = bootstrap(normal.distribution)

hist(bootstrap.distribution)
```



As we can see, the bootstrap distribution is even more normal than the original and the amount of variance is much smaller in the bootstrap distribution

```
sd(normal.distribution)
```

```
## [1] 1.00616
```

```
sd(bootstrap.distribution)
```

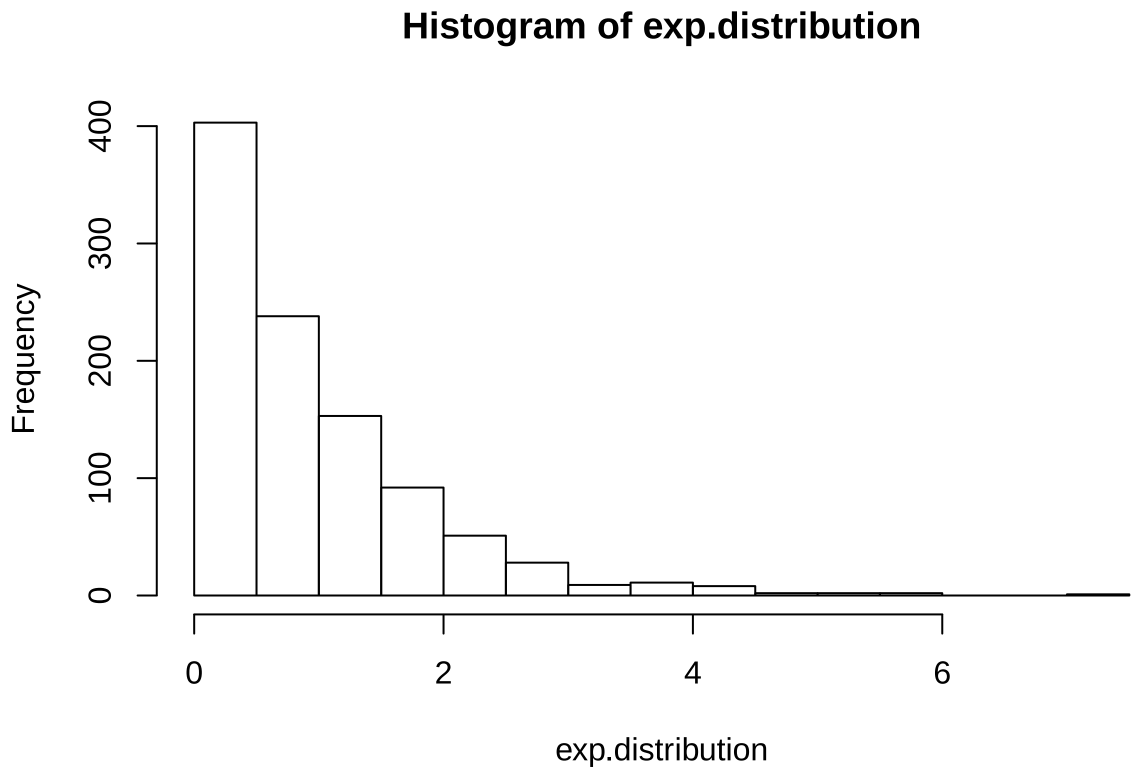
```
## [1] 0.138954
```

Bootrstrap Exponential Distribution

Now let's create an exponential distribution and apply the same steps that we did with the normal distribution above.

```r
exp.distribution = rexp(1000)
```

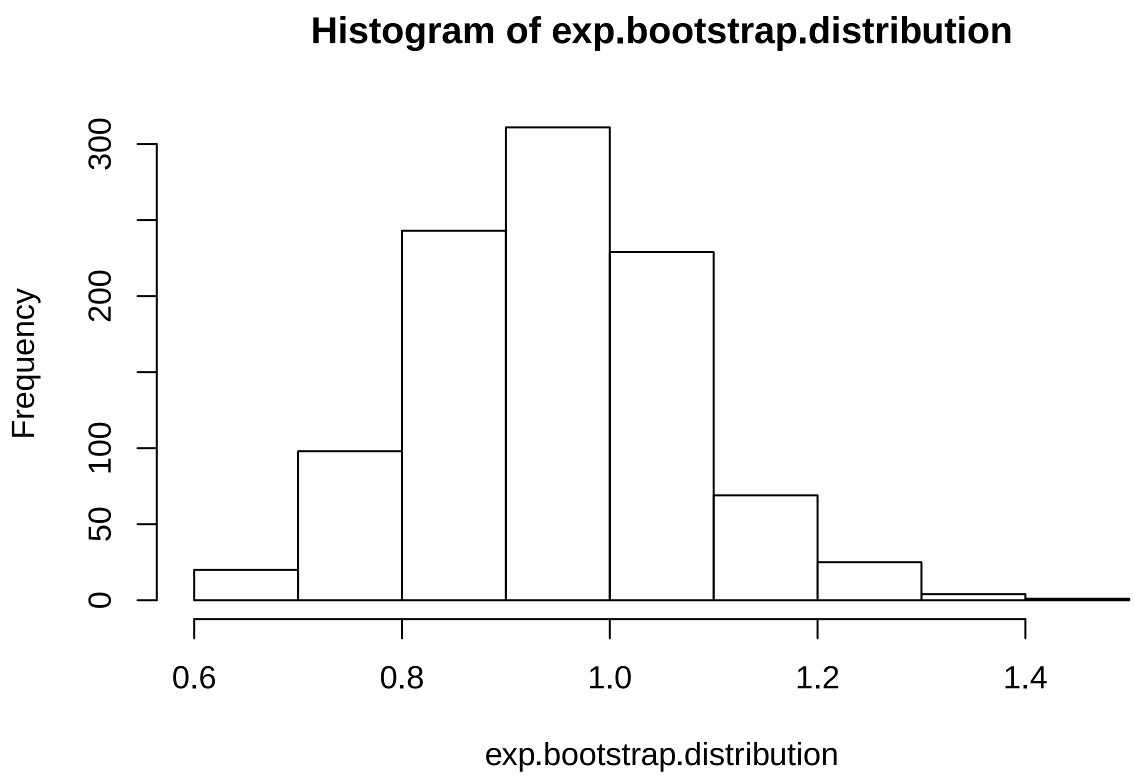
As we can see, the data is heavily skewed with a strong, right-tailed distribution.



Applying the `bootstrap` method against our `exp.distribution` variable, we can see that the new distribution is almost completely normal…an amazing transformation from the original, right-tailed distribution.

```r
exp.bootstrap.distribution = bootstrap(exp.distribution)

hist(exp.bootstrap.distribution)
```



Notice that the bootstrap distribution is centered around 1

```
mean(exp.bootstrap.distribution)
```

```
## [1] 0.9468496
```

Finally, notice that the variance is much less within the bootstrap distribution

```
sd(exp.distribution)
```

```
## [1] 0.9196133
```

```
sd(exp.bootstrap.distribution)
```

```
## [1] 0.1260915
```

Thus, the true power of the Central Limit Theorem has been revealed and demonstrated.

> The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough. stattrek.com