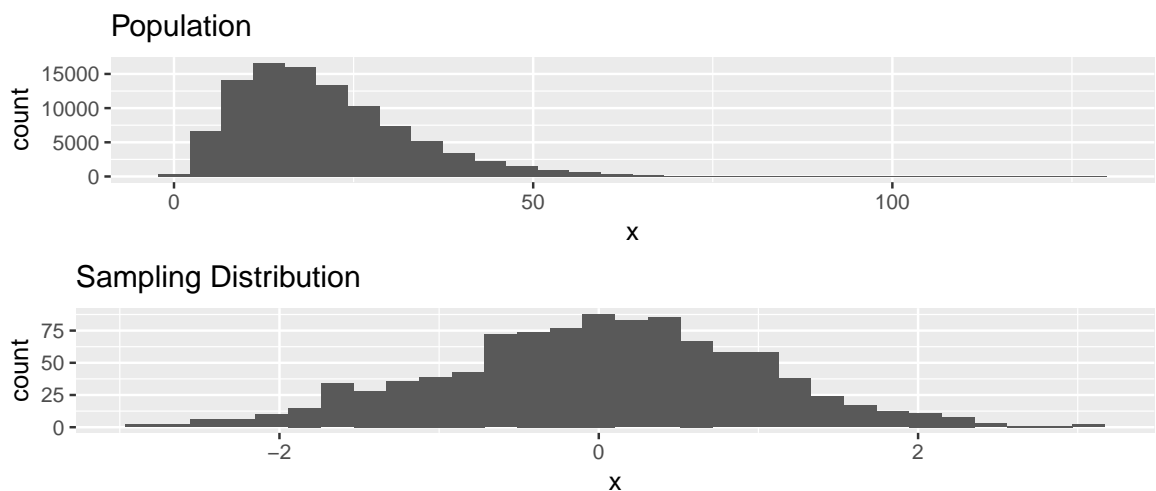


1. Write an R function that replicates our sampling distribution examples we played with in class. You will use the Gamma distribution $X \sim \Gamma(\text{shape} = 3, \text{scale} = 7)$.

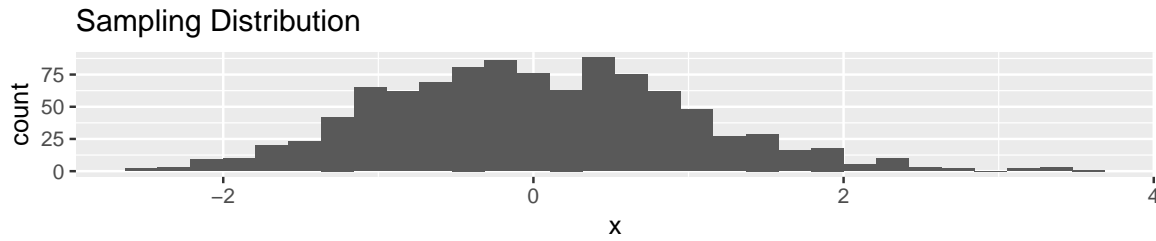
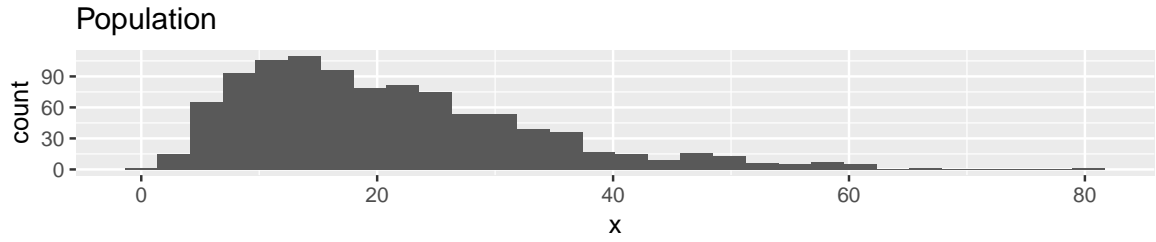
```
library(ggplot2)
library(gridExtra)
f <- function(N, R, shape=3, scale=7) {
  xbars <- rep(NA, R)
  for (r in 1:R) {
    xbars[r] <- mean(rgamma(N, shape=shape, scale=scale))
  }
  z <- (xbars - 3*7)/(sqrt(3)*7/sqrt(N))
  X <- rgamma(N*1e2, shape=shape, scale=scale)
  pop_plot <- ggplot(data.frame(x=X), aes(x=x)) +
    geom_histogram(bins=30) +
    labs(title="Population")
  sampdist_plot <- ggplot(data.frame(x=z), aes(x=x)) +
    geom_histogram(bins=30) +
    labs(title="Sampling Distribution")
  grid.arrange(pop_plot, sampdist_plot, nrow=2)
}
```

```
f(1001, 1e3) # a
```

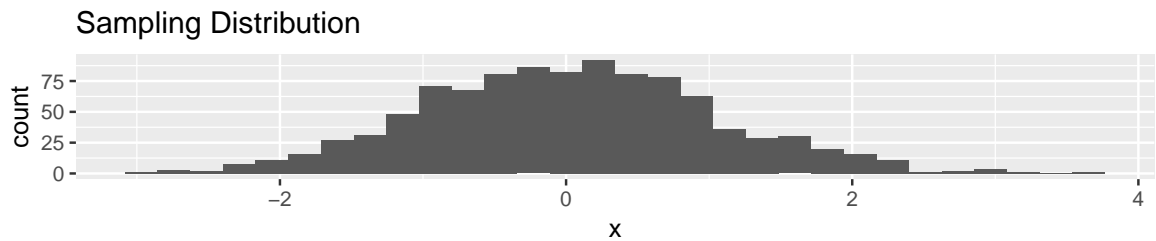
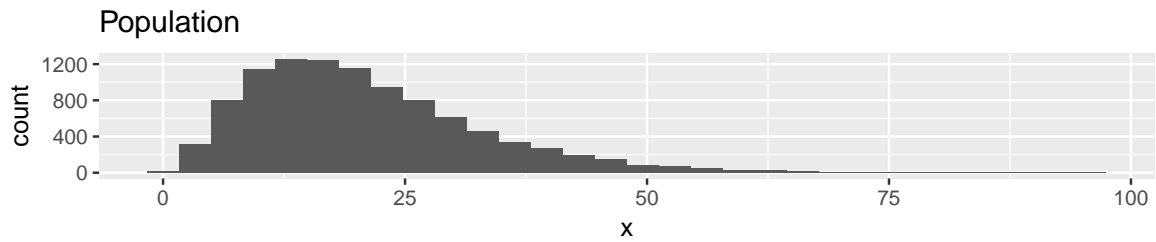


The gamma distribution is right skewed, but the sampling distribution is nearly symmetric nonetheless. This is the effect of the Central Limit Theorem.

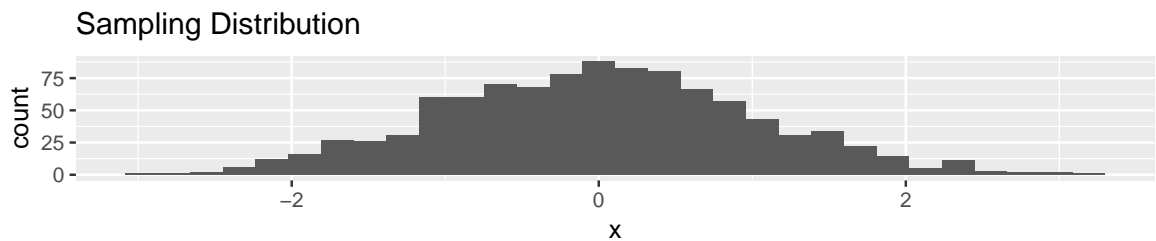
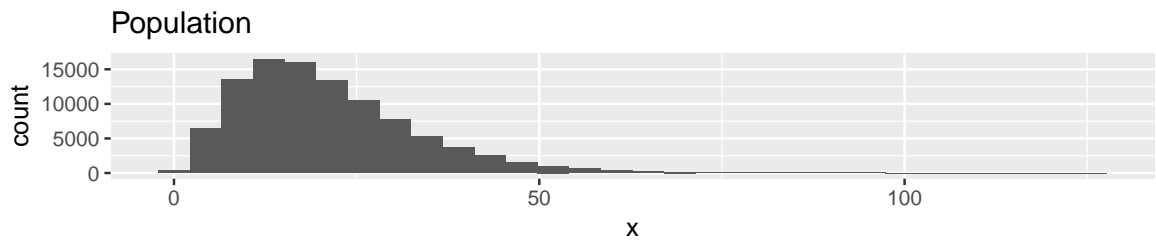
```
f(10, 1e3) # b and c
```



f(100, 1e3)

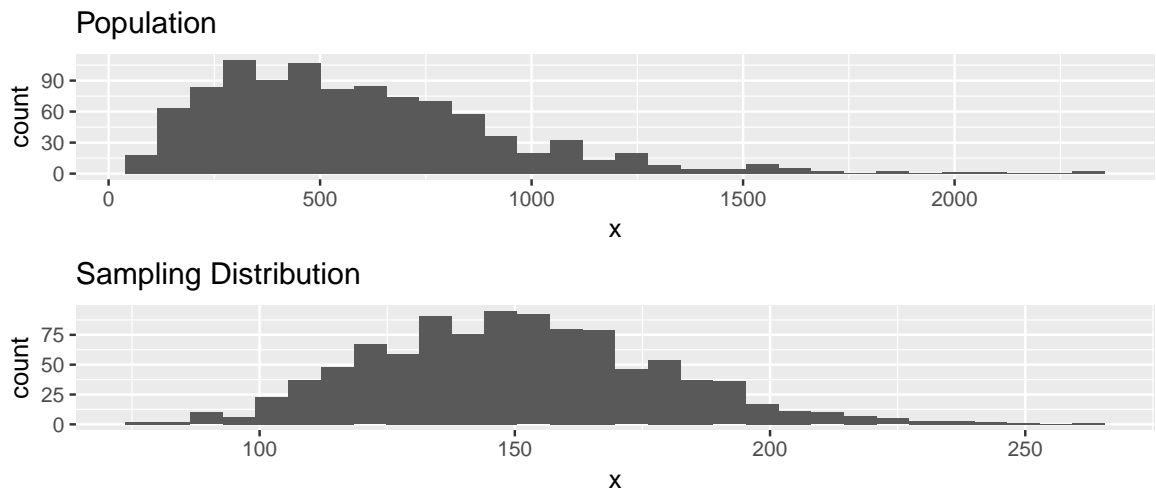


f(1000, 1e3)

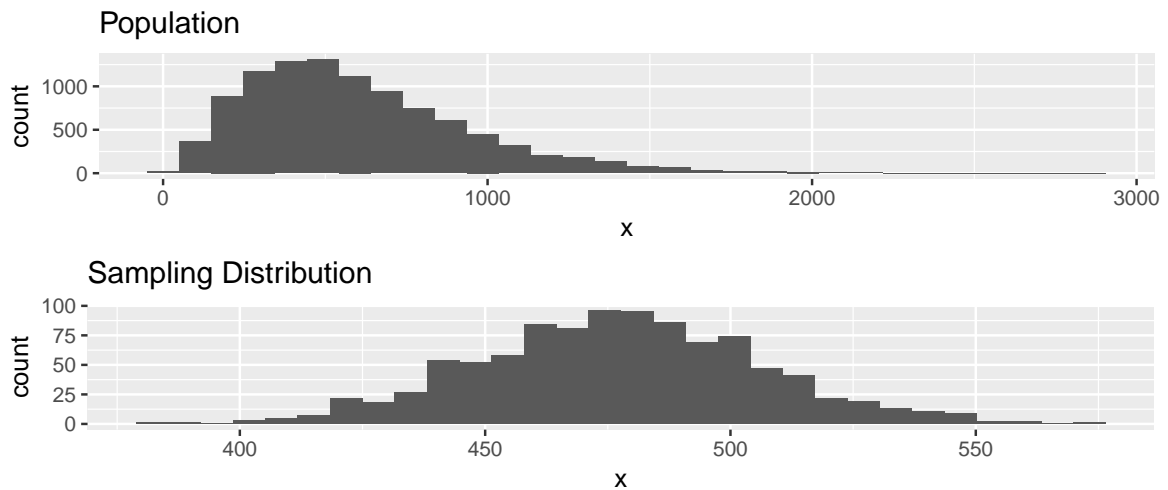


We see that as the sample size increases from 10 to 100 to 1000, the sampling distribution becomes increasingly symmetric and smoother – hence the normal approximation improves. Notice though, that because we standardized the sample means, that is calculated the z-score for each sample mean, the standard normal distribution applies throughout. If we did not calculate the z-scores, the standard error would decrease as the sample size increases, and the sampling distribution would become narrower. This is an important distinction.

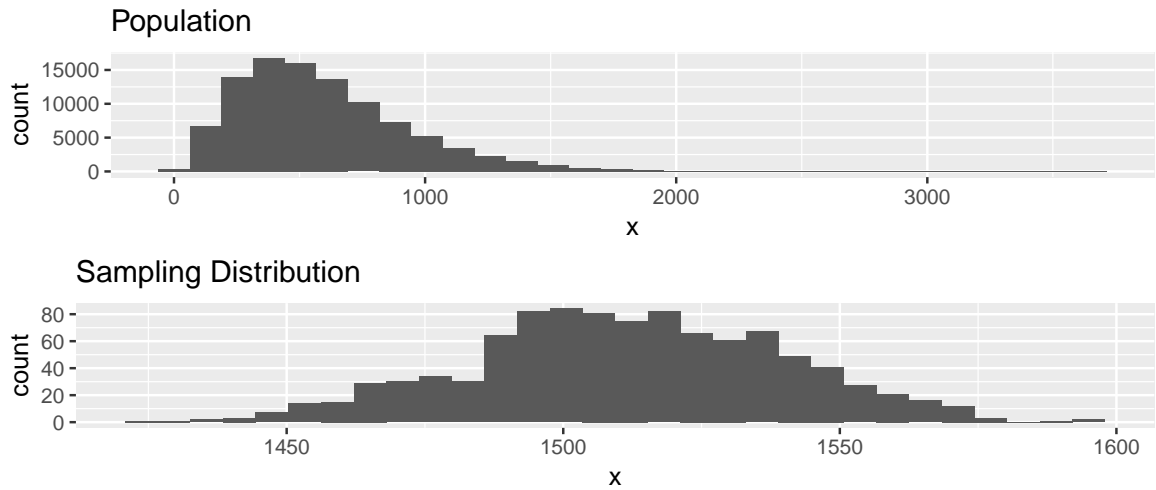
```
f(10, 1e3, scale=200) # d
```



```
f(100, 1e3, scale=200)
```



```
f(1000, 1e3, scale=200)
```



Notice, that despite the increased skew of the gamma distribution, the sampling distribution still looks symmetric.