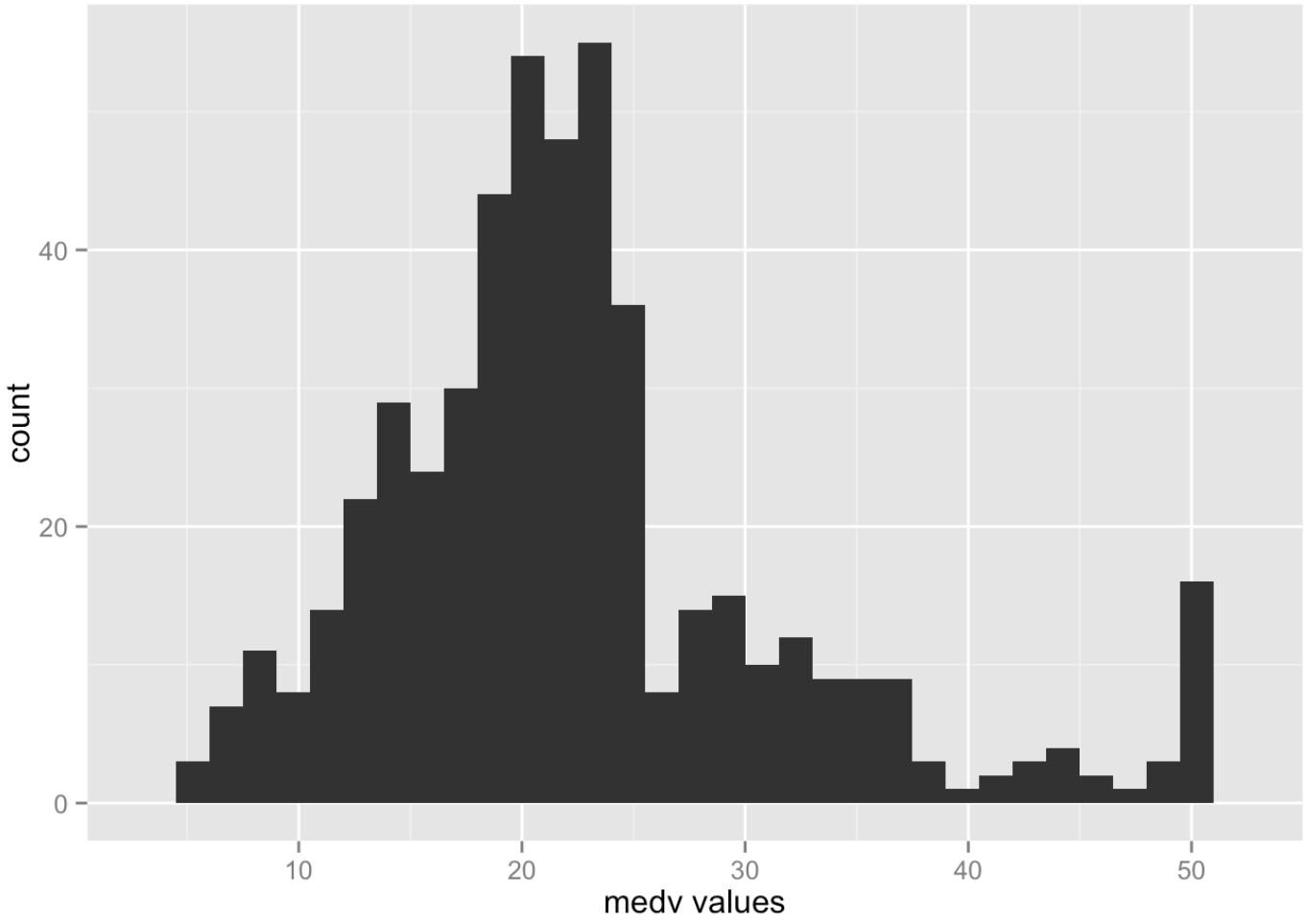
Problem 4

Chapter 5, Exercise 9 (Sec. 5.4, p. 201)

```
m = Boston$medv
qplot(m, xlab = 'medv values')
```



Part A

```
m.mean = mean(m)
m.mean
```

```
## [1] 22.53281
```

Part B

```
m.std_err = sd(m)/sqrt(nrow(Boston))
m.std_err
```

```
## [1] 0.4088611
```

This standard error is fairly low, though it could be better.

Part C

```
# Inspiration for the following code comes from:
#   stats.stackexchange.com/questions/22472/use-of-standard-error-
#   of-bootstrap-distribution

m.bs = boot(m, function(d, i) mean(d[i]), 1000)

# The standard error from bootstrap is simply the standard deviation
# of the bootstrap distribution:
sd(m.bs$t)
```

```
## [1] 0.4033504
```

The bootstrap method's standard error is very similar to that of part (b).

Part D

```
boot.ci(m.bs, type='bca')
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = m.bs, type = "bca")
##
## Intervals :
## Level       BCa
## 95%   (21.79, 23.37 )
## Calculations and Intervals on Original Scale
```

```
t.test(Boston$medv)
```

```
##
##  One Sample t-test
##
## data:  Boston$medv
## t = 55.111, df = 505, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  21.72953 23.33608
## sample estimates:
## mean of x
##  22.53281
```

These result in nearly identical confidence intervals, with a difference of `< 0.1` on each end.

Part E

```
m.median = median(m)
m.median
```

```
## [1] 21.2
```

Part F

```
m.bs = boot(m, function(d, i) median(d[i]), 1000)
sd(m.bs$t)
```

```
## [1] 0.3724187
```

The standard error of the median using the bootstrap (`0.385`) is somewhat lower than that of the mean (`0.419`).

Part G

```
u10 = quantile(m, c(.10))
```

Part H

```
m.bs = boot(m, function(d, i) quantile(d[i], c(.10)), 1000)
sd(m.bs$t)
```

```
## [1] 0.5010284
```

Our standard error increases significantly when we try to predict the 10th percentile. This is understandable; since our data basically takes the form of a normal distribution, the bulk of the data is in the middle, while there are far fewer examples to draw upon on the edges.