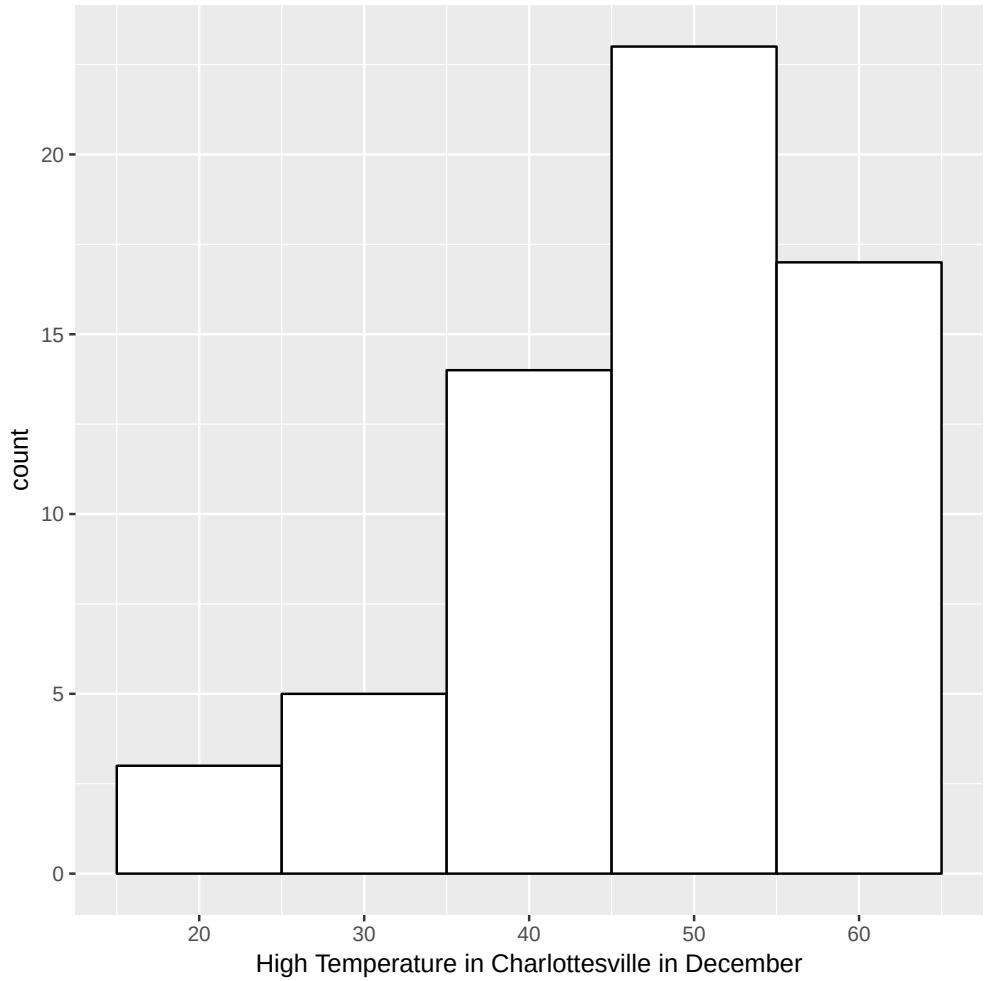
Statistics 3080
Homework 8
David Smith

Problem 1a

```
> set.seed(7311986)
> library(ggplot2)
> weather <- read.table("weather_full.txt")
> ggplot(weather, aes(x=V1)) +
+   geom_histogram(binwidth=10, fill="white", colour="black") +
+   labs(x="High Temperature in Charlottesville in December")
> print("The population distribution is bell-shaped and left-skewed,")

[1] "The population distribution is bell-shaped and left-skewed,"

> print("similar to a chi-square distribution but flipped.")

[1] "similar to a chi-square distribution but flipped."
```

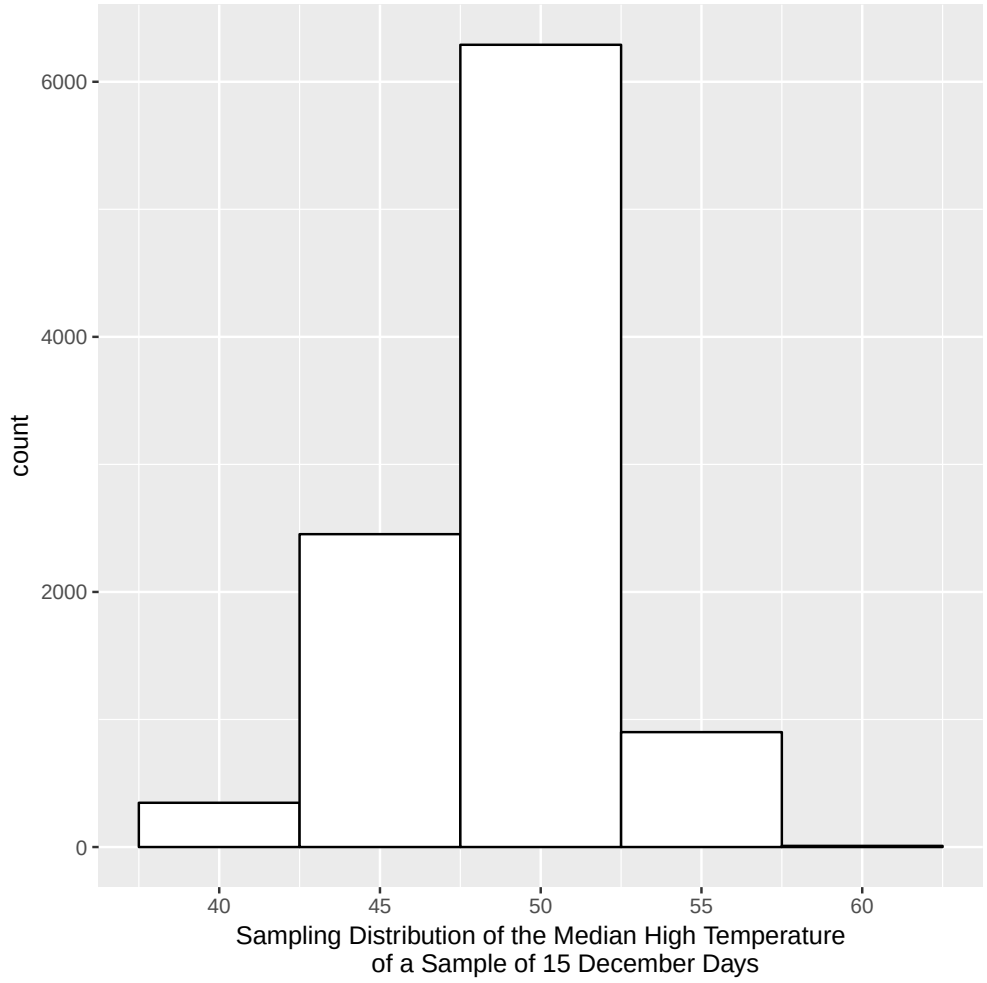


Problem 1b

```
> K <- 10000
> weather_samps <- replicate(K, sample(weather$V1, size=15))
> weather_meds <- apply(weather_samps, 2, median)
> mean(weather_meds)

[1] 49.1588

> sd(weather_meds)

[1] 2.92141
```

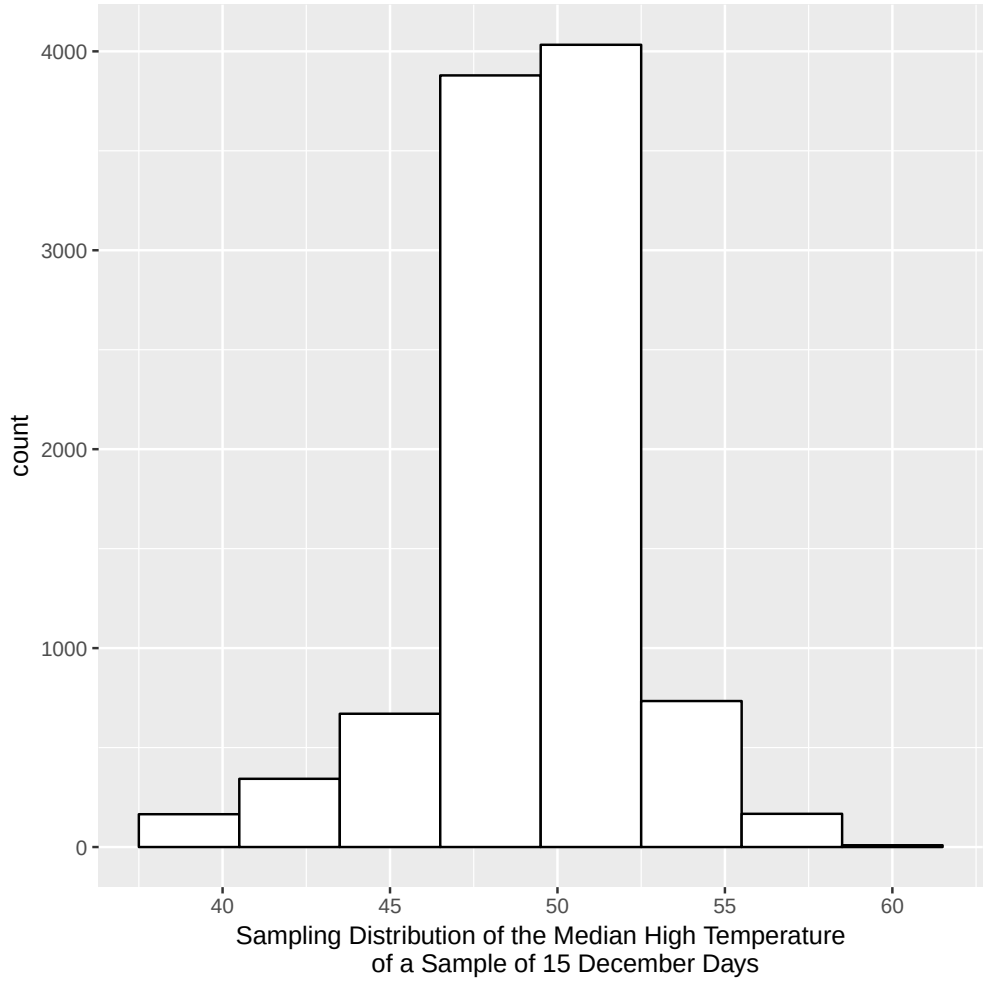
Problem 1c

```
> ggplot() + aes(weather_meds) +
+   geom_histogram(binwidth=5, fill="white", colour="black") +
+   labs(x="Sampling Distribution of the Median High Temperature
+        of a Sample of 15 December Days")
> print("The sampling distribution appears normal, with a small")

[1] "The sampling distribution appears normal, with a small"

> print("variance, since the graph is tall in the middle and")

[1] "variance, since the graph is tall in the middle and"

> print("quickly thins near the tails. The distribution does appear")

[1] "quickly thins near the tails. The distribution does appear"

> print("to be slightly left-skewed.")

[1] "to be slightly left-skewed."
```



Problem 1d

```
> ggplot() + aes(weather_meds) +
+   geom_histogram(binwidth=3, fill="white", colour="black") +
+   labs(x="Sampling Distribution of the Median High Temperature
+        of a Sample of 15 December Days")
> print("Again, the sampling distribution appears normal, with a")

[1] "Again, the sampling distribution appears normal, with a"

> print("small variance. The normality is more clear now, and")

[1] "small variance. The normality is more clear now, and"

> print("there is only an extremely slight left skew to the graph.")

[1] "there is only an extremely slight left skew to the graph."
```



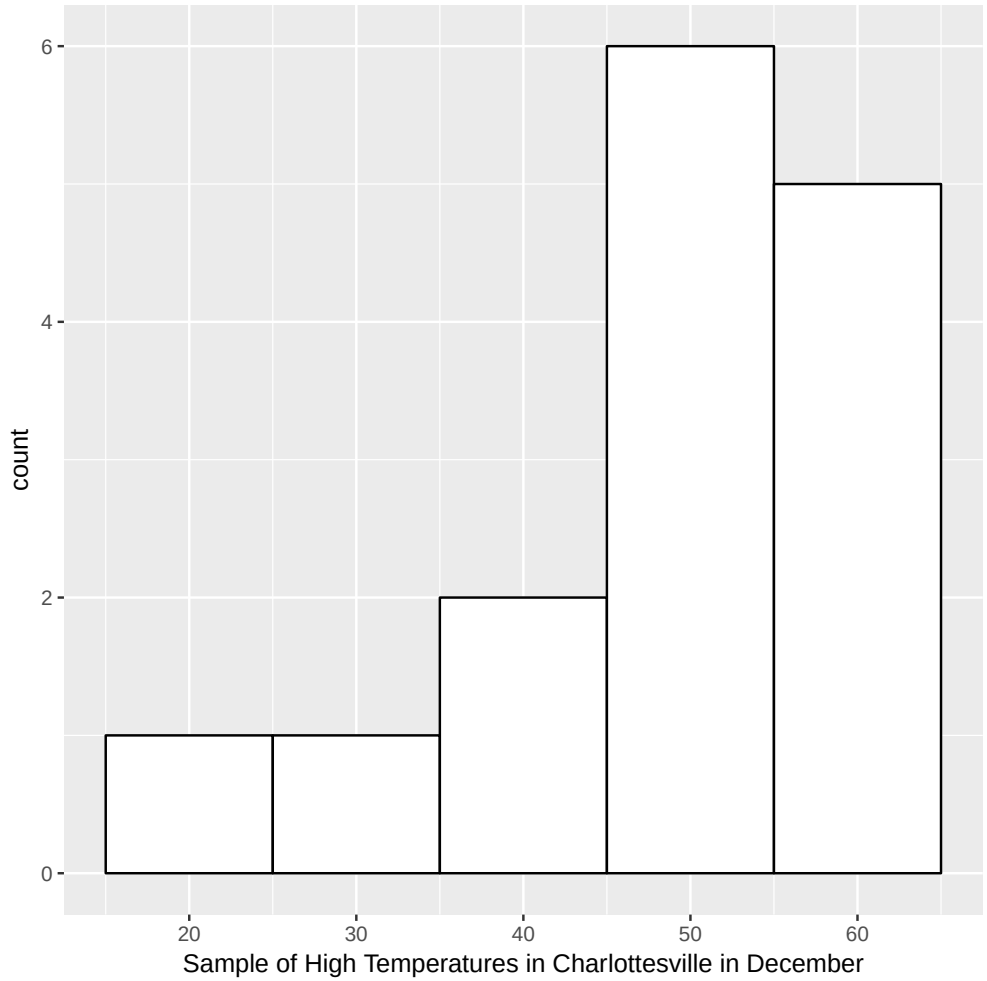
Problem 1e

```
> print("Since the sampling distribution appears normal, and")
[1] "Since the sampling distribution appears normal, and"
> print("has low variance as well, using a methodology based on the")
[1] "has low variance as well, using a methodology based on the"
> print("Central Limit Theorem to evaluate the median is appropriate.")
[1] "Central Limit Theorem to evaluate the median is appropriate."
```

Problem 2a

```
> w_samp <- read.table("weather_samp.txt")
> ggplot(w_samp, aes(x=V1)) +
+   geom_histogram(binwidth=10, fill="white", colour="black") +
+   labs(x="Sample of High Temperatures in Charlottesville in December")
> print("The sample distribution appears to be slighlty bell-shaped,")
```

```
[1] "The sample distribution appears to be slighlty bell-shaped,"
> print("and is left-skewed. The assumptions for bootstrapping hold, since")
[1] "and is left-skewed. The assumptions for bootstrapping hold, since"
> print("this sample is independent and representative (it is presumably")
[1] "this sample is independent and representative (it is presumably"
> print("an SRS, and comparison to the population distribution in problem")
[1] "an SRS, and comparison to the population distribution in problem"
> print("1 confirms this), and the underlying population does not have heavy")
[1] "1 confirms this), and the underlying population does not have heavy"
> print("tails, as was seen in the previous problem.")
[1] "tails, as was seen in the previous problem."
```



Problem 2b

```
> B <- 10000
> boot_samp <- replicate(B, sample(w_samp$V1, replace=TRUE))
> samp_med <- median(w_samp$V1)
> boot_meds <- apply(boot_samp, 2, median)
> boot_err <- sort(boot_meds - samp_med)
> p2.5 <- 250
> p97.5 <- 9750
> boot_ci <- samp_med - boot_err[c(p97.5, p2.5)]
> boot_ci

[1] 44 56
```

Problem 2c

```
> boot_sd <- sd(boot_meds)
> boot_sd

[1] 3.07159

> lower_ci <- samp_med - 1.96*boot_sd/sqrt(15)
> upper_ci <- samp_med + 1.96*boot_sd/sqrt(15)
> c(lower_ci, upper_ci)

[1] 48.44556 51.55444
```

Problem 2d

```
> print("The normal bootstrap confidence interval is wider, while the")

[1] "The normal bootstrap confidence interval is wider, while the"

> print("z confidence interval is significantly shorter. This is")

[1] "z confidence interval is significantly shorter. This is"

> print("consistent with my answer to problem 1e, since we concluded")

[1] "consistent with my answer to problem 1e, since we concluded"

> print("that the sampling distribution of medians was normal and had")

[1] "that the sampling distribution of medians was normal and had"

> print("small variance, and thus using the Central Limit Theorem to")

[1] "small variance, and thus using the Central Limit Theorem to"
```

```
> print("evaluate the median (such as calculating a z confidence interval")

[1] "evaluate the median (such as calculating a z confidence interval"

> print("as we have done here) was appropriate.")

[1] "as we have done here) was appropriate."
```

Problem 3a

```
> flights <- read.table("flights.txt")
> boot_samp <- replicate(B, sample(flights$V1, replace=TRUE))
> samp_perc <- quantile(flights$V1, 0.95)
> boot_perc <- apply(boot_samp, 2, quantile, probs=0.95)
> boot_err <- sort(boot_perc - samp_perc)
> p0.5 <- 50
> p99.5 <- 9950
> boot_ci <- samp_perc - boot_err[c(p99.5, p0.5)]
> boot_ci

[1] 49.90 55.45
```

Problem 3b

```
> print("The conservative endpoint of this confidence interval is")

[1] "The conservative endpoint of this confidence interval is"

> print("the upper endpoint of 55.45, since this is the most cautious")

[1] "the upper endpoint of 55.45, since this is the most cautious"

> print("value to check against the number of available seats: if this")

[1] "value to check against the number of available seats: if this"

> print("endpoint is less than or equal to the number of available seats,")

[1] "endpoint is less than or equal to the number of available seats,"

> print("the entire interval is, and thus we can be very sure that the")

[1] "the entire interval is, and thus we can be very sure that the"

> print("claim is true.")

[1] "claim is true."
```

Problem 3c

```
> print("The algorithm has not worked according to this sample, since")
[1] "The algorithm has not worked according to this sample, since"
> print("the upper bound of the confidence interval for the 95th percentile")
[1] "the upper bound of the confidence interval for the 95th percentile"
> print("is greater than 55, the number of available seats, indicating that")
[1] "is greater than 55, the number of available seats, indicating that"
> print("more passengers arrive than there are seats on the plane more than")
[1] "more passengers arrive than there are seats on the plane more than"
> print("5 percent of the time.")
[1] "5 percent of the time."
```

References:

- None