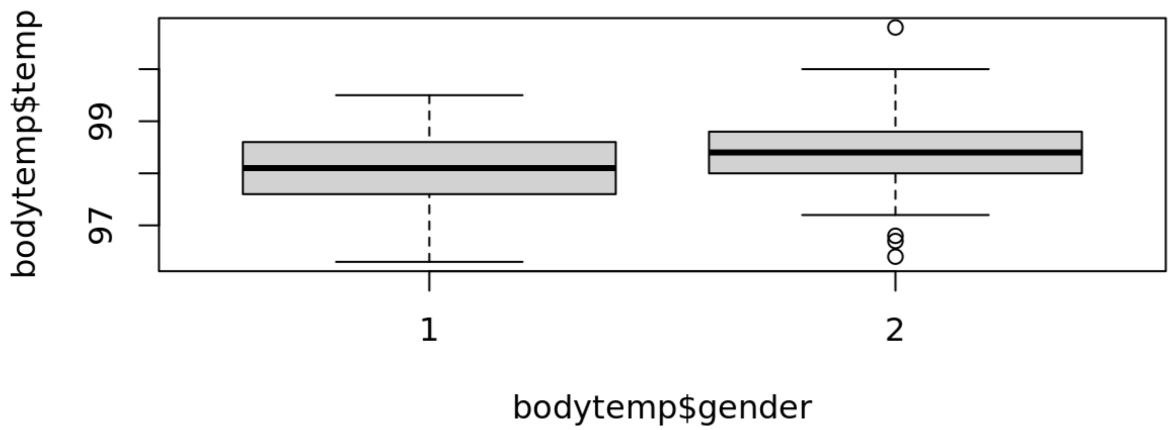
Question 1: Produce side-by-side boxplots that compare the body temperature of females against the body temperature of males. Based on your plots, which gender has a higher average body temperature?

```
> boxplot(bodytemp$temp ~ bodytemp$gender)
```



Based on the plot it seems like females (denoted #2) have a higher average body temperature.

Question 2: State the null and alternative hypotheses, both in words and in symbols.
Null hypothesis: There is no association between gender and average body temperature
mu female=mu male
Alternative Hypothesis: There is an association between gender and average body temperature.
Mu female≠ mu male

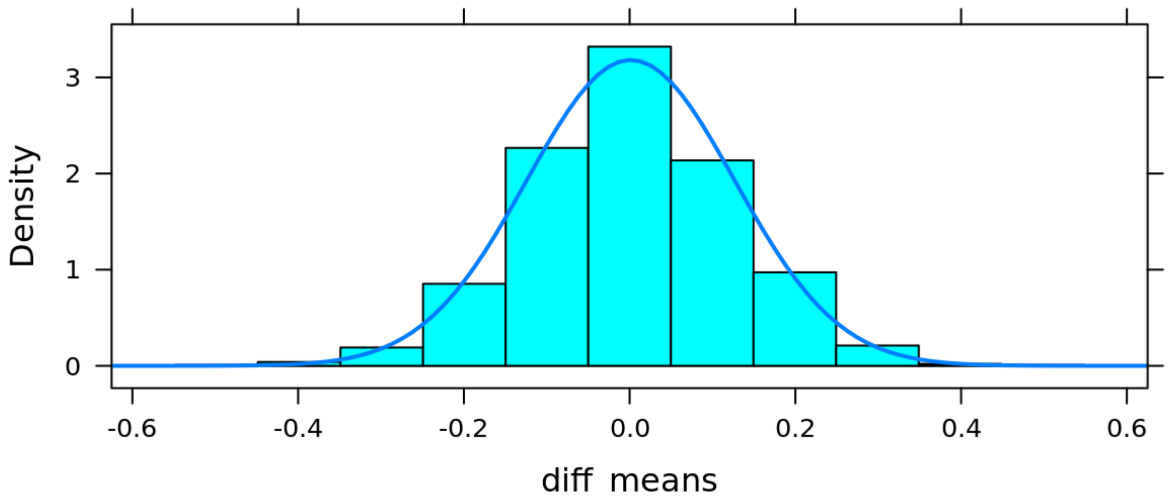
```
obs_means <- mean(temp ~ gender, data = bodytemp)
obs_mean_diff <- obs_means[1] - obs_means[2]
obs_mean_diff
```

```
## 1
## -0.2892308
```

Question 3: Interpret what the value -0.2892308 of obs_mean_diff means in the context of the data
Obs_mean_diff is the difference in average body temperature of men and the average body temperature of women

Question 4: Produce a histogram of the null distribution of the difference in sample means with a normal curve superimposed. Do you think the null distribution is approximately normal? Comment on the histogram and check the validity conditions

```
> histogram(diff_means, fit="normal")
```



The null distribution appears to be approximately normal because it is unimodal and approximately symmetrical. The body temperature is approximately normal because there are at least 20 observations in each group of the quantitative variable.

Question 5: What proportion of simulated statistics are at least as extreme as the observed statistic obs_mean_diff? In other words, what is the (simulation-based) p-value?

```
> mean(diff_means >= abs(obs_mean_diff) | diff_means <= obs_mean_diff)
[1] 0.016
```

The simulation based p-value is 0.016. Therefore the proportion of statistics that are atleast as extreme as the observed statistic is 0.016 or 1.6%.

Question 6: Using a significance level of α = 0.05, decide whether or not you would reject the null hypothesis in favor of the alternative. Interpret this decision in context.
Based on the significance level of 0.05, I would reject the null hypothesis in favor of the alternative hypothesis because the p-value of 0.016 is less than 0.05. In the context of the study this means that we reject that there is no association between gender and average body temperature and deem the association between gender and average body temperature plausible.

Question 7: Use the t.test() function to conduct a two-sample t-test to decide if the true mean body temperature is different between males and females. Use a significance level of α = 0.05. Is your theory-based p-value similar to the simulation-based p-value from Question 5?

```
t.test(temp ~ gender, data = bodytemp)

        Welch Two Sample t-test
```

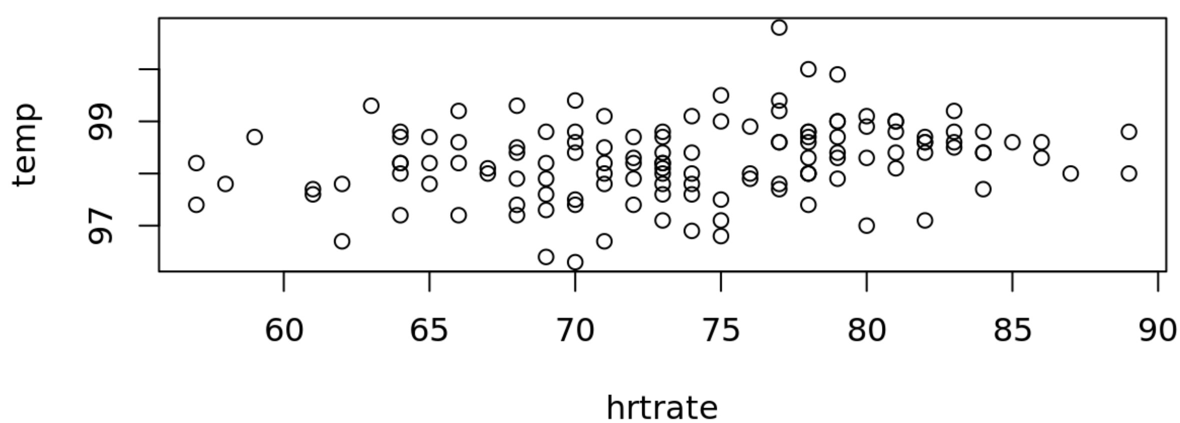
```
data:  temp by gender
t = -2.2854, df = 127.51, p-value = 0.02394
alternative hypothesis: true difference in means between group 1 and group 2 is not
equal to 0
95 percent confidence interval:
 -0.53964856 -0.03881298
sample estimates:
mean in group 1 mean in group 2
       98.10462        98.39385
```

The theory-based p-value is 0.02394. Although this is slightly higher than the observed p-value it still indicates strong evidence against the null hypothesis and using the significance level of 0.05 we can reach the same conclusion to reject the null hypothesis in favor of the alternative hypothesis. Since we reach the same conclusion and both p-values indicate similar strength of evidence, the theory-based and simulation-based p-value are similar.

Question 8: Interpret the 95% confidence interval from the t.test() output in the context of the data.
The 95% confidence interval is between -0.5396 and -0.0388. This means that we are 95% confident that the long run difference in average body temperature between males and females is between -0.5396 and -0.0388. This means that the average body temperature is lower for males than females.

Question 9: Produce and describe a scatterplot that considers the relationship between heart rate and temperature. In deciding which variable to use as the explanatory and which as the response, think about the research question we are asking.

```
> plot(temp ~ hrtrate, data = bodytemp)
```



The scatter plot has high variability between 97-100 and seems to have a slight positive correlation. The area between 65-80 is the most dense area on the scatter plot. There appears to be few outliers one and only one appears to jump out from the data around 77bpm and 100 degrees. This is because heart-rate and body-temperature are regulated to stay as constant as possible

__Question 10:__ Compute the sample correlation between heart rate and body temperature.

```
> cor(temp ~ hrtrate, data = bodytemp)
[1] 0.2536564
```

The correlation coefficient between heart rate and body temperature is 0.2537. As expected this is a small positive correlation.

__Question 11:__ Compute the (simulation-based) p-value for the observed slope. Do we have strong evidence against the null hypothesis and in favor of the alternative?

```
> obs_fit <- lm(temp ~ hrtrate, data = bodytemp)
> slopevalues <- obs_fit$coefficients[2]
>
> mean(slopes >= slopevalues | slopes <= -slopevalues)
[1] 0.004
```

The simulation-based p-value for the observed slope is 0.004. Based on the p-value, we have strong evidence against the null hypothesis in favor of the alternative hypothesis.

__Question 12:__ Use the summary() function to conduct a test of significance to decide if there is a statistically significant linear association between heart rate and body temperature. Use a significance level of α = 0.05. Review: What would the slope coefficient be if you switch the roles of x and y? Would it stay the same? What about the p-value?

```
> bestfit <- lm(temp ~ hrtrate, data = bodytemp)
> summary(bestfit)

Call:
lm(formula = temp ~ hrtrate, data = bodytemp)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.85017 -0.39999  0.01033  0.43915  2.46549 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 96.306754   0.657703 146.429  < 2e-16 ***
hrtrate      0.026335   0.008876   2.967  0.00359 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.712 on 128 degrees of freedom
Multiple R-squared:  0.06434,	Adjusted R-squared:  0.05703 
F-statistic: 8.802 on 1 and 128 DF,  p-value: 0.003591
```

>

The p-value is 0.003591 which means there is strong evidence against the null hypothesis, which means there is strong evidence against no linear association between heart rate and body temperature. This means that it is likely there is a linear association between heart rate and body temperature. If you switched the values of x and y the coefficient would change signs and go from positive correlation to negative correlation. The p-value would be the same magnitude and the sign would change.