Week 4 Assignment

Jessalyn Chuang

#(1) Data Tidying and Visualization

##a. Tidying

```
urchins_data <- read.csv(file = here("./Assignments/week4/urchins.csv"),
                         stringsAsFactors = TRUE)

urchins_data <- urchins_data[,c("YEAR", "MONTH", "SITE", "TREATMENT", "SIZE",
                                "COUNT", "COMMON_NAME")]

urchins_data <- filter(urchins_data, SITE == "MOHK" & COMMON_NAME == "Red Urchin" & (YEAR == 2015 | YEA

urchins_data <- expand.dft(urchins_data, freq = "COUNT")
```

##b. Histogram of size distributions of red urchins in 2015 and 2016

```
n_2015 <- sum(urchins_data$YEAR == 2015)
n_2016 <- sum(urchins_data$YEAR == 2016)
range_2015 <- urchins_data %>%
  filter(YEAR == 2015) %>%
  summarize(range = max(SIZE) - min(SIZE)) %>%
  pull(range)
range_2016 <- urchins_data %>%
  filter(YEAR == 2016) %>%
  summarize(range = max(SIZE) - min(SIZE)) %>%
  pull(range)

bin_number_2015 <- round(2*n_2015^(1/3))
bin_number_2016 <- round(2*n_2016^(1/3))

bin_width_2015 <- range_2015/(2*n_2015^(1/3))
bin_width_2016 <- range_2016/(2*n_2016^(1/3))

ggplot(urchins_data, aes(x = SIZE, fill = TREATMENT)) +
  geom_histogram(data = filter(urchins_data, YEAR == 2015),
                 bins = bin_number_2015, binwidth = bin_width_2015) +
  geom_histogram(data = filter(urchins_data, YEAR == 2016),
                 bins = bin_number_2016, binwidth = bin_width_2016) +
  facet_wrap(~ YEAR) +
  labs(title = "Size Distribution of Red Urchins (2015-2016)",
       x = "Urchin Size (cm)",
       y = "Count",
       fill = "Treatment",
       caption = "Figure 1: Histogram of red urchin size distributions at Mohawk Reef in 2015 and 2016,
```

```
       categorized by treatment type. The x-axis represents urchin size (cm), and the y-axis indicates
       the count of individuals in each size bin.2015 data has 14 bins with bin width 0.49,
       2016 data has 12 bins with bin width 0.60 based on the Rice Rule.") +
  theme(plot.caption = element_text(hjust = 0.5))
```

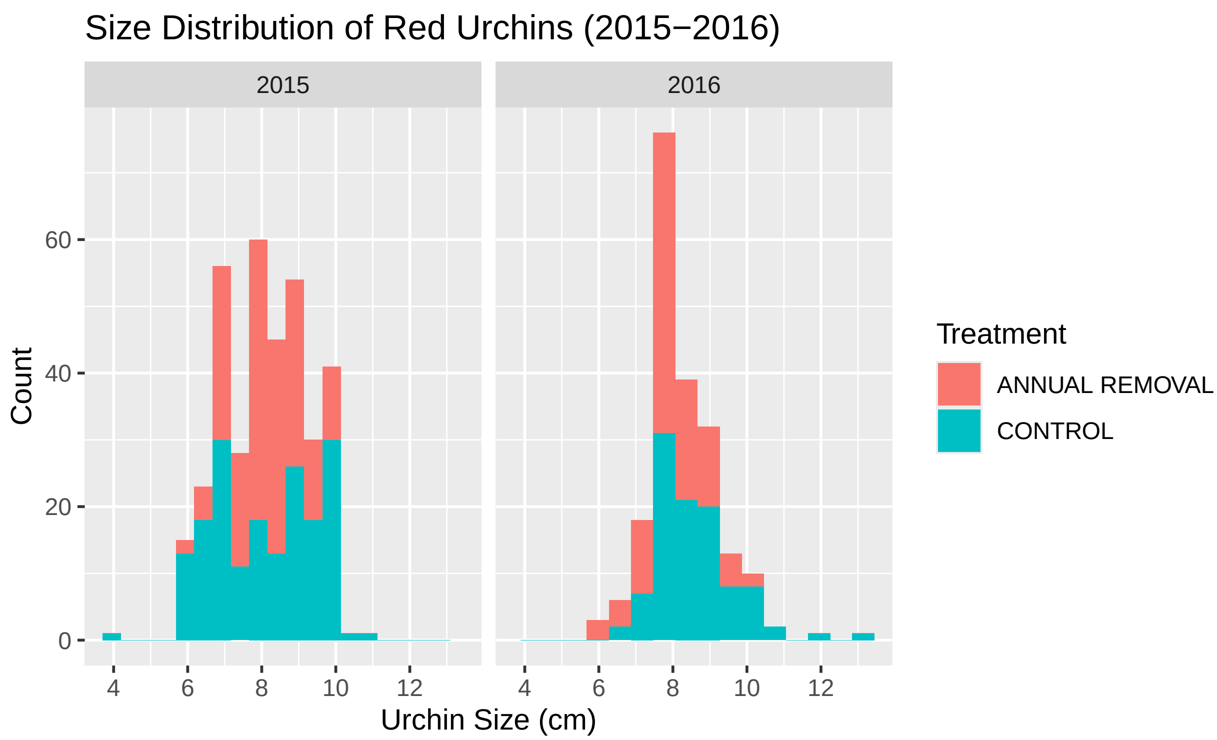

Figure 1: Histogram of red urchin size distributions at Mohawk Reef in 2015 and 2016, categorized by treatment type. The x−axis represents urchin size (cm), and the y−axis indicates the count of individuals in each size bin.2015 data has 14 bins with bin width 0.49, 2016 data has 12 bins with bin width 0.60 based on the Rice Rule.

#(2) Confidence Intervals

##a. 95% two-sided confidence interval for red urchin sizes in 2015 across treatment types

```
#2015
urchins_2015 <- urchins_data %>%
  filter(YEAR == 2015)
mean_2015 <- mean(urchins_2015$SIZE)
urchins_2015_CI <- t.test(urchins_2015$SIZE, mu = mean_2015, conf.level = 0.95,
                          alternative = "two.sided")
print(paste("The 95% confidence interval for red urchin sizes in 2015 is [",
            round(urchins_2015_CI$conf.int[1], 2), ",",
            round(urchins_2015_CI$conf.int[2], 2), "] cm"))
```

```
## [1] "The 95% confidence interval for red urchin sizes in 2015 is [ 8.08 , 8.32 ] cm"
```

##b. 95% two-sided confidence interval for red urchin sizes in 2016 across treatment types

```
#2016
urchins_2016 <- urchins_data %>%
```

```
  filter(YEAR == 2016)
mean_2016 <- mean(urchins_2016$SIZE)
urchins_2016_CI <- t.test(urchins_2016$SIZE, mu = mean_2016, conf.level = 0.95,
                          alternative = "two.sided")
print(paste("The 95% confidence interval for purple urchin sizes in 2016 is [",
            round(urchins_2016_CI$conf.int[1], 2), ",",
            round(urchins_2016_CI$conf.int[2], 2), "] cm"))
```

```
## [1] "The 95% confidence interval for purple urchin sizes in 2016 is [ 8.14 , 8.42 ] cm"
```

#(3) One-sample Hypothesis Tests

##(ai) One-sided one-sample t-test to see if red urchins in 2015 from control group are significantly less than the mean red urchin size of 9 cm reported by researchers at another location

Null hypothesis: Red urchins in the 2015 from the control group do not have lower sizes that the population mean red urchin size of 9 cm Alternative hypothesis: Red urchins in the 2015 from the control group have lower sizes that the population mean red urchin size of 9 cm

```
urchins_2015_control <- urchins_data %>%
  filter(YEAR == 2015 & TREATMENT == "CONTROL")
urchins_2015_control_test <- t.test(urchins_2015_control$SIZE, mu = 9,
                                    conf.level = 0.95, alternative = "less")
print(paste("Resulting p-value = ", urchins_2015_control_test$p.value))
```

```
## [1] "Resulting p-value =  8.84853252845733e-14"
```

With a 95% confidence level, since p-value is less than 0.05, we reject the null hypothesis in favor of the alternative hypothesis which is that red urchins in the 2015 from the control group have significantly lower sizes that the mean red urchin size of 9 cm reported by researchers at another location.

##(3aii) One-sided one-sample t-test to see if red urchins in 2016 from control group are significantly less than the mean red urchin size of 9 cm reported by researchers at another location

Null hypothesis: Red urchins in the 2016 from the control group do not have lower sizes that the population mean red urchin size of 9 cm Alternative hypothesis: Red urchins in the 2016 from the control group have lower sizes that the population mean red urchin size of 9 cm

```
urchins_2016_control <- urchins_data %>%
  filter(YEAR == 2016 & TREATMENT == "CONTROL")
urchins_2016_control_test <- t.test(urchins_2016_control$SIZE, mu = 9,
                                    conf.level = 0.95, alternative = "less")
print(paste("Resulting p-value = ", urchins_2016_control_test$p.value))
```

```
## [1] "Resulting p-value =  2.66601172211481e-05"
```

With a 95% confidence level, since p-value is less than 0.05, we reject the null hypothesis in favor of the alternative hypothesis which is that red urchins in the 2016 from the control group have significantly lower sizes that the mean red urchin size of 9 cm reported by researchers at another location.

#(4) Two-sample Hypothesis Tests

##(ai) Two-sample two-sided t-test for mean red urchin sizes between the two treatments in 2015

```r
red_2015_control <- urchins_2015_control
red_2015_annual <- urchins_data %>%
  filter(YEAR == 2015 & TREATMENT == "ANNUAL REMOVAL")

#testing for equal variances first
red_2015_vartest <- var.test(x = red_2015_control$SIZE, y = red_2015_annual$SIZE,
                             alternative = "two.sided",
                             conf.level = 0.95)

red_2015_vartest
```

```
##
##  F test to compare two variances
##
## data:  red_2015_control$SIZE and red_2015_annual$SIZE
## F = 2.2319, num df = 179, denom df = 174, p-value = 1.556e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.659458 2.999889
## sample estimates:
## ratio of variances
##           2.231887
```

```r
#The p-value <0.05, so we reject the null hypothesis and we retain the alternative
#hypothesis that the variances are not equal

red_2015_ttest <- t.test(red_2015_control$SIZE, red_2015_annual$SIZE)

red_2015_ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  red_2015_control$SIZE and red_2015_annual$SIZE
## t = -0.29451, df = 313.51, p-value = 0.7686
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2810157  0.2078411
## sample estimates:
## mean of x mean of y
##  8.180556  8.217143
```

The mean red urchin size in 2015 was not significantly different between the control group (mean ± standard deviation = 8.18 ± 1.38, n = 180) and the annual removal group (mean ± standard deviation = 8.22 ± 0.92, n = 175) as determined by a two-sample, two-sided t-test (t = -0.29, df = 313.51, p = 0.77, $\alpha = 0.05$)

##(aii) Two-sample two-sided t-test for mean red urchin sizes between the two treatments in 2016

```r
red_2016_control <- urchins_data %>%
  filter(YEAR == 2016 & TREATMENT == "CONTROL")
red_2016_annual <- urchins_data %>%
  filter(YEAR == 2016 & TREATMENT == "ANNUAL REMOVAL")
```

```r
#testing for equal variances first
red_2016_vartest <- var.test(x = red_2016_control$SIZE, y = red_2016_annual$SIZE,
                             alternative = "two.sided",
                             conf.level = 0.95)

red_2016_vartest
```

```
##
##  F test to compare two variances
##
## data:  red_2016_control$SIZE and red_2016_annual$SIZE
## F = 1.5464, num df = 100, denom df = 99, p-value = 0.03084
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.041326 2.295565
## sample estimates:
## ratio of variances
##           1.546409
```

```r
#The p-value <0.05, so we reject the null hypothesis and we retain the alternative
#hypothesis that the variances are not equal

red_2016_ttest <- t.test(red_2016_control$SIZE, red_2016_annual$SIZE)

red_2016_ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  red_2016_control$SIZE and red_2016_annual$SIZE
## t = 4.0132, df = 190.99, p-value = 8.597e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2768581 0.8120528
## sample estimates:
## mean of x mean of y
##  8.554455  8.010000
```

The mean red urchin size in 2016 was significantly different between the control group (mean $\pm$ standard deviation = 8.55 $\pm$ 1.06, n = 101) and the annual removal group (mean $\pm$ standard deviation = 8.01 $\pm$ 0.85, n = 100) as determined by a two-sample, two-sided t-test (t = 4.0132, df = 190.99, p = 8.597e-05, $\alpha = 0.05$)

##(b) Two-sample two-sided t-test for mean red urchin sizes in the control group between 2015 and 2016

```r
#testing for equal variances first
red_control_vartest <- var.test(x = red_2015_control$SIZE, y = red_2016_control$SIZE,
                             alternative = "two.sided",
                             conf.level = 0.95)

red_control_vartest
```

```
##
```

```
##  F test to compare two variances
##
## data:  red_2015_control$SIZE and red_2016_control$SIZE
## F = 1.691, num df = 179, denom df = 100, p-value = 0.004139
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.184118 2.373493
## sample estimates:
## ratio of variances
##           1.691033
```

```
#The p-value <0.05, so we reject the null hypothesis and we retain the alternative
#hypothesis that the variances are not equal

red_control_ttest <- t.test(red_2015_control$SIZE, red_2016_control$SIZE)

red_control_ttest
```

```
##
##  Welch Two Sample t-test
##
## data:  red_2015_control$SIZE and red_2016_control$SIZE
## t = -2.5383, df = 252.7, p-value = 0.01174
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.66399679 -0.08380299
## sample estimates:
## mean of x mean of y
##  8.180556  8.554455
```

The mean red urchin size was significantly different between the 2015 control group data (mean $\pm$ standard deviation = 8.18 $\pm$ 1.38, n = 180) and the 2016 control group data (mean $\pm$ standard deviation = 8.55 $\pm$ 1.06, n = 101) as determined by a two-sample, two-sided t-test (t = -2.54, df = 252.7, p = 0.01, α = 0.05)