ESM 206 Assignment 4 Problem 2

Robert Heim

11/14/2018

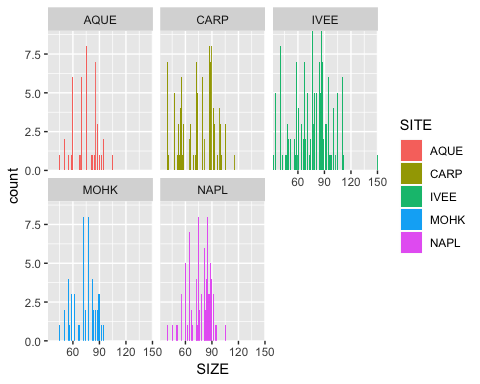
# Problem 2: Compare Mean Lobster Size By Site in 2017

# Get data into case and tidy format  
  
count <- as.data.frame(lobster\_size) #Coerce to data.frame   
  
lobster\_size\_tidy <- expand.dft(count, freq = "COUNT") #Expand data to tidy format  
  
lobster\_size\_tidy\_2017 <- lobster\_size\_tidy %>%   
 filter(YEAR == 2017) #Use only observations only from 2017

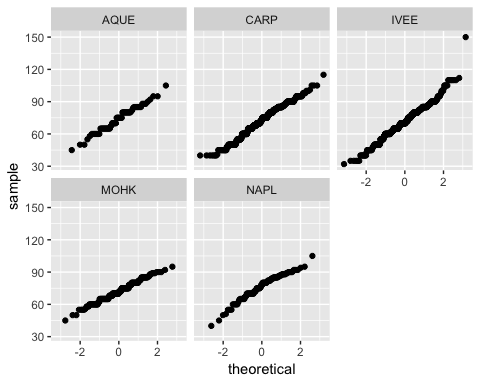
# Create a summary table for mean lobster carapace length in 2017 at each site  
  
size\_mean\_2017 <- lobster\_size\_tidy\_2017 %>%  
 group\_by(SITE) %>%  
 summarize(Mean\_size= mean(SIZE),Standard\_deviation= sd(SIZE), Sample\_size= length(SITE), Standard\_error= Standard\_deviation/sqrt(Sample\_size))

# Single factor: Site   
# Number of levels in the factor: 5 (AQUE, CARP, IVEE, MOHK, and NAPL)  
# Random variable: carapace length (mm)  
  
#Question: Is there a significant difference in carapace length (mm) for AQUE, CARP, IVEE, MOHK, and NAPL locations in 2017?  
  
# H0: Mean lobster carapace length across all sites are equal.   
# HA: At least two mean lobster caparace lengths differ significantly between sites.   
  
# Do some data exploring to find whether the assumption of normality is met for the carapace length data. Create a histogram and qq-plot.  
  
lobster\_2017\_hist <- lobster\_size\_tidy\_2017 %>%  
 ggplot(aes(x= SIZE)) +  
 geom\_histogram(binwidth = 1, aes(fill= SITE))+  
 facet\_wrap((~SITE)) +  
 scale\_y\_continuous(expand = c(0,0), limits = c(0,9))+   
 scale\_x\_continuous(expand = c(0,0))  
  
lobster\_2017\_hist

## Warning: Removed 49 rows containing missing values (geom\_bar).



lobster\_2017\_qq <- lobster\_size\_tidy\_2017 %>%  
 ggplot(aes(sample= SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE)  
  
lobster\_2017\_qq



# Lobster size data for 2017 appear normally distributed at each site  
  
# Use a one-way ANOVA to test for a difference between any two of the sites  
  
lobster\_aov <- aov(SIZE ~ SITE, data = lobster\_size\_tidy\_2017)  
  
summary(lobster\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## SITE 4 2355 588.6 3.424 0.0085 \*\*  
## Residuals 1663 285871 171.9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

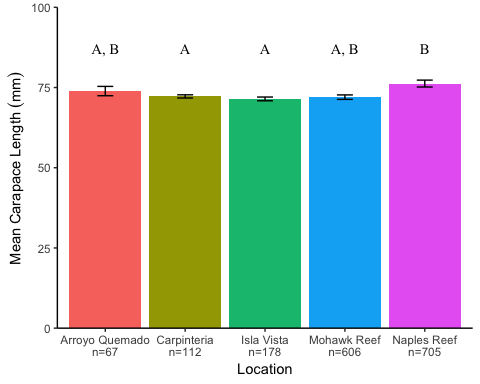
# P-value= 0.009, meaning at least two samples of lobster carapance length were taken from sites with different means. Which ones are different? Are all five different from each other? Or something else?   
  
# Need to run a Post-hoc Tukey's test to find where the differences are.   
  
lobster\_ph <- TukeyHSD(lobster\_aov)   
  
lobster\_ph

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = SIZE ~ SITE, data = lobster\_size\_tidy\_2017)  
##   
## $SITE  
## diff lwr upr p adj  
## CARP-AQUE -1.6657352 -6.24294710 2.911477 0.8582355  
## IVEE-AQUE -2.4433772 -7.05292315 2.166169 0.5968998  
## MOHK-AQUE -1.8955224 -7.02720717 3.236162 0.8514711  
## NAPL-AQUE 2.3366205 -3.19311600 7.866357 0.7775633  
## IVEE-CARP -0.7776420 -2.76097123 1.205687 0.8216104  
## MOHK-CARP -0.2297872 -3.23309697 2.773523 0.9995765  
## NAPL-CARP 4.0023556 0.36042398 7.644287 0.0228728  
## MOHK-IVEE 0.5478548 -2.50450730 3.600217 0.9882889  
## NAPL-IVEE 4.7799976 1.09751057 8.462485 0.0037001  
## NAPL-MOHK 4.2321429 -0.08607271 8.550358 0.0579286

Mean lobster carapace length in 2017 differed significantly at the five Long-Term Ecological Resesearch Sites (one-way ANOVA, F(4,1663)= 3.42, p= 0.009; Figure 1). Post hoc analysis by Tukey’s HSD revealed that carapace length differed significantly between Naples Reef and Carpinteria (pairwise p= 0.023). Carapace length also differed significantly between Naples Reef and Isla Vista (0.004).

“Sediment bulk densities (g/cm3 ) differed significantly in the four river reaches studied (one-way ANOVA, F(3,92) = 85.0, p < 0.001, n = 24 for each; Table 1). Post-hoc analysis by Tukey’s HSD revealed that sediment density in Reach A differed significantly from all other reaches (pairwise p < 0.001) and had the lowest mean bulk density (0.38 g/cm3 ). Sediment bulk densities in the downstream reaches were more than double that measured in Reach A (Figure 1), indicating that Reach A may currently offer the most suitable spawning habitat.”

# Use a bar graph to show mean lobster carapace length at each site. Include error bars and letters.   
  
   
lobster\_size\_col <-ggplot(size\_mean\_2017, aes(x = SITE, y = Mean\_size)) +  
 geom\_col(aes(fill = SITE)) +  
 geom\_errorbar(aes(ymin= Mean\_size-Standard\_error, ymax= Mean\_size+Standard\_error), width=.2) +   
 theme\_classic() +  
 theme(legend.position= "none") +  
 labs(x= "Location", y= expression(Mean~Carapace~Length~(mm))) +  
 scale\_y\_continuous(limit= c(0,100), expand=c(0,0)) +  
 scale\_x\_discrete(labels=c("AQUE" = "Arroyo Quemado\nn=67",   
 "NAPL" = "Naples Reef\nn=705",  
 "MOHK" = "Mohawk Reef\nn=606",  
 "IVEE" = "Isla Vista\nn=178",  
 "CARP" = "Carpinteria\nn=112")) +  
 annotate("text", x = 1, y = 87, label = "A, B", family = "Times New Roman") +  
 annotate("text", x = 2, y = 87, label = "A", family = "Times New Roman") +  
 annotate("text", x = 3, y = 87, label = "A", family = "Times New Roman") +  
 annotate("text", x = 4, y = 87, label = "A, B", family = "Times New Roman") +  
 annotate("text", x = 5, y = 87, label = "B", family = "Times New Roman")  
  
lobster\_size\_col



**Figure 1. Mean Lobster Carapace Length at five sites in the Santa Barbara Channel in 2017.** Error bars represent ± 1 standard error. Like letters indicate values that do not differ significantly. Data retreived from Santa Barbara Coastal Long-Term Ecological Research [1].

# 3 Changes in lobster size at each MPA and non-MPA site (comparing only 2012 and 2017 sizes).

# Use a one-sided, two-sample t-test to compare lobster size between 2012-2017 at each MPA site  
  
#HO: Lobsters at IVEE are not significantly larger in 2017 than in 2012  
#HA: Lobsters at IVEE are significantly larger in 2017 than in 2012  
  
# Create a new df with lobster size for IVEE at 2012 and 2017   
  
lobster\_size\_IVEE\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE") %>%   
 filter(YEAR == "2012")   
   
lobster\_size\_IVEE\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE") %>%   
 filter(YEAR == "2017")   
  
# Find the mean and sd for IVEE for 2012 and 2017  
  
IVEE\_2017\_mean <- mean(lobster\_size\_IVEE\_2017$SIZE)  
  
IVEE\_2017\_sd <-sd(lobster\_size\_IVEE\_2017$SIZE)  
   
IVEE\_2012\_mean <- mean(lobster\_size\_IVEE\_2012$SIZE)  
  
IVEE\_2012\_sd <-sd(lobster\_size\_IVEE\_2012$SIZE)  
  
IVEE\_2017\_mean - IVEE\_2012\_mean

## [1] 5.375222

# Run a one-sided t-test for each MPA, comparing 2012 to 2017   
  
lobster\_IVEE\_ttest<- t.test(lobster\_size\_IVEE\_2017$SIZE, lobster\_size\_IVEE\_2012$SIZE, var.equal= TRUE, alternative= "greater")  
  
lobster\_IVEE\_ttest

##   
## Two Sample t-test  
##   
## data: lobster\_size\_IVEE\_2017$SIZE and lobster\_size\_IVEE\_2012$SIZE  
## t = 1.885, df = 630, p-value = 0.02995  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.6777847 Inf  
## sample estimates:  
## mean of x mean of y   
## 71.45215 66.07692

# Need to do an effect size test using Cohen's D  
# First, make vectors  
  
IVEE\_2017\_vector <- lobster\_size\_IVEE\_2017 %>%  
 pull(SIZE)  
  
IVEE\_2012\_vector <- lobster\_size\_IVEE\_2012 %>%  
 pull(SIZE)  
  
effect\_size <- cohen.d(IVEE\_2017\_vector,IVEE\_2012\_vector)  
effect\_size

##   
## Cohen's d  
##   
## d estimate: 0.3775177 (small)  
## 95 percent confidence interval:  
## inf sup   
## -0.01633002 0.77136540

Mean lobsters carapace length in 2017 [71.45 ± 14.32 ( ± ), n=26] was not signficantly different than in 2012 [66.08 ± 12.09 ( ± ), n=606] at Isla Vista [t(630) = 1.88, p < 0.03, = 0.05]. The effect size is moderate (Cohen’s d= 0.38). From the samples, lobster carapace length in 2017 was approximately 5.38 mm greater than lobster caparace length in 2012.

# Run a one-sided, two-sample t-test to see if size in 2017 is greater than in 2012  
  
lobster\_size\_NAPL\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "NAPL") %>%   
 filter(YEAR == "2012")   
   
lobster\_size\_NAPL\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "NAPL") %>%   
 filter(YEAR == "2017")   
  
# Find the mean and sd for NAPL for 2012 and 2017  
  
NAPL\_2017\_mean <- mean(lobster\_size\_NAPL\_2017$SIZE)  
  
NAPL\_2017\_sd <-sd(lobster\_size\_NAPL\_2017$SIZE)  
   
NAPL\_2012\_mean <- mean(lobster\_size\_NAPL\_2012$SIZE)  
  
NAPL\_2012\_sd <-sd(lobster\_size\_NAPL\_2012$SIZE)  
  
NAPL\_2017\_mean - NAPL\_2012\_mean

## [1] 3.232143

# Run a one-sided t-test for to see if size in 2017 is greater than in 2012  
  
lobster\_NAPL\_ttest<- t.test(lobster\_size\_NAPL\_2017$SIZE, lobster\_size\_NAPL\_2012$SIZE, var.equal= TRUE)  
  
lobster\_NAPL\_ttest

##   
## Two Sample t-test  
##   
## data: lobster\_size\_NAPL\_2017$SIZE and lobster\_size\_NAPL\_2012$SIZE  
## t = 0.67636, df = 116, p-value = 0.5002  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -6.232765 12.697051  
## sample estimates:  
## mean of x mean of y   
## 76.23214 73.00000

# Lobster sizes at NAPL in 2017 are not significantly different than in 2012  
  
NAPL\_2017\_vector <- lobster\_size\_NAPL\_2017 %>%  
 pull(SIZE)  
  
NAPL\_2012\_vector <- lobster\_size\_NAPL\_2012 %>%  
 pull(SIZE)  
  
effect\_size <- cohen.d(NAPL\_2017\_vector,NAPL\_2012\_vector)  
effect\_size

##   
## Cohen's d  
##   
## d estimate: 0.2834216 (small)  
## 95 percent confidence interval:  
## inf sup   
## -0.5473456 1.1141889

Mean lobsters carapace length in 2017 [76.23 ± 11.39 ( ± ), n=6] was not signficantly different than in 2012 [73 ± 11.75 ( ± ), n=112] at Naples Reef [t(630) = 0.68, p < 0.5, = 0.05]. The effect size is small (Cohen’s d= 0.28). From the samples, lobster carapace length in 2017 was approximately 3.23 mm greater than lobster caparace length in 2012.

# One-sided t-test for AQUE nonMPA   
  
lobster\_size\_AQUE\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE") %>%   
 filter(YEAR == "2012")   
  
lobster\_size\_AQUE\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE") %>%   
 filter(YEAR == "2017")   
  
lobster\_AQUE\_ttest<- t.test(lobster\_size\_AQUE\_2017$SIZE, lobster\_size\_AQUE\_2012$SIZE, var.equal= TRUE)  
  
lobster\_AQUE\_ttest

##   
## Two Sample t-test  
##   
## data: lobster\_size\_AQUE\_2017$SIZE and lobster\_size\_AQUE\_2012$SIZE  
## t = 1.2622, df = 103, p-value = 0.2097  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.654312 7.445357  
## sample estimates:  
## mean of x mean of y   
## 73.89552 71.00000

# Find the mean and sd for AQUE for 2012 and 2017  
  
AQUE\_2017\_mean <- mean(lobster\_size\_AQUE\_2017$SIZE)  
  
AQUE\_2017\_sd <-sd(lobster\_size\_AQUE\_2017$SIZE)  
   
AQUE\_2012\_mean <- mean(lobster\_size\_AQUE\_2012$SIZE)  
  
AQUE\_2012\_sd <-sd(lobster\_size\_AQUE\_2012$SIZE)  
  
AQUE\_2017\_mean - AQUE\_2012\_mean

## [1] 2.895522

# Lobster sizes at AQUE in 2017 are not significantly different than in 2012  
  
AQUE\_2017\_vector <- lobster\_size\_AQUE\_2017 %>%  
 pull(SIZE)  
  
AQUE\_2012\_vector <- lobster\_size\_AQUE\_2012 %>%  
 pull(SIZE)  
  
effect\_size <- cohen.d(AQUE\_2017\_vector,AQUE\_2012\_vector)  
effect\_size

##   
## Cohen's d  
##   
## d estimate: 0.2563169 (small)  
## 95 percent confidence interval:  
## inf sup   
## -0.1479675 0.6606014

Mean lobsters carapace length in 2017 [73.9 ± 11.89 ( ± ), n=38] was not signficantly different than in 2012 [71 ± 10.15 ( ± ), n=67] at Arroyo Quemado [t(103) = 1.26, p < 0.21, = 0.05]. The effect size is small (Cohen’s d= 0.26). From the samples, lobster carapace length in 2017 was approximately 2.90 mm greater than lobster caparace length in 2012.

# One-sided t-test for CARP nonMPA   
  
lobster\_size\_CARP\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "CARP") %>%   
 filter(YEAR == "2012")   
  
lobster\_size\_CARP\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "CARP") %>%   
 filter(YEAR == "2017")   
  
lobster\_CARP\_ttest<- t.test(lobster\_size\_CARP\_2017$SIZE, lobster\_size\_CARP\_2012$SIZE,var.equal= TRUE)  
  
lobster\_CARP\_ttest

##   
## Two Sample t-test  
##   
## data: lobster\_size\_CARP\_2017$SIZE and lobster\_size\_CARP\_2012$SIZE  
## t = -1.3361, df = 781, p-value = 0.1819  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.257332 0.998958  
## sample estimates:  
## mean of x mean of y   
## 72.22979 74.35897

# Find the mean and sd for CARP for 2012 and 2017  
  
CARP\_2017\_mean <- mean(lobster\_size\_CARP\_2017$SIZE)  
  
CARP\_2017\_sd <-sd(lobster\_size\_CARP\_2017$SIZE)  
   
CARP\_2012\_mean <- mean(lobster\_size\_CARP\_2012$SIZE)  
  
CARP\_2012\_sd <-sd(lobster\_size\_CARP\_2012$SIZE)  
  
CARP\_2017\_mean - CARP\_2012\_mean

## [1] -2.129187

# Lobster sizes at CARP in 2017 are not significantly different than in 2012  
  
CARP\_2017\_vector <- lobster\_size\_CARP\_2017 %>%  
 pull(SIZE)  
  
CARP\_2012\_vector <- lobster\_size\_CARP\_2012 %>%  
 pull(SIZE)  
  
effect\_size <- cohen.d(CARP\_2017\_vector,CARP\_2012\_vector)  
effect\_size

##   
## Cohen's d  
##   
## d estimate: -0.1594364 (negligible)  
## 95 percent confidence interval:  
## inf sup   
## -0.39380971 0.07493682

# Lobster sizes at CARP in 2017 are not significantly greater than in 2012

Mean lobsters carapace length in 2017 [72.23 ± 13.21 ( ± ), n=78] was not signficantly different than in 2012 [74.36 ± 14.62 ( ± ), n=705] at Carpinteria [t(103) = -1.34, p < 0.182, = 0.05]. The effect size is small (Cohen’s d= -0.16). From the samples, lobster carapace length in 2017 was approximately 2.13 mm less than lobster caparace length in 2012.

# One-sided t-test for MOHK nonMPA   
  
lobster\_size\_MOHK\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "MOHK") %>%   
 filter(YEAR == "2012")   
  
lobster\_size\_MOHK\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "MOHK") %>%   
 filter(YEAR == "2017")   
  
lobster\_MOHK\_ttest<- t.test(lobster\_size\_MOHK\_2017$SIZE, lobster\_size\_MOHK\_2012$SIZE,var.equal= TRUE, alternative = "less")  
  
lobster\_MOHK\_ttest

##   
## Two Sample t-test  
##   
## data: lobster\_size\_MOHK\_2017$SIZE and lobster\_size\_MOHK\_2012$SIZE  
## t = -4.0689, df = 259, p-value = 3.138e-05  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -3.121847  
## sample estimates:  
## mean of x mean of y   
## 72.00000 77.25301

# Lobster sizes are signficantly different at MOHK in 2012 and 2017  
  
# Find the mean and sd for MOHK for 2012 and 2017  
  
MOHK\_2017\_mean <- mean(lobster\_size\_MOHK\_2017$SIZE)  
  
MOHK\_2017\_sd <-sd(lobster\_size\_MOHK\_2017$SIZE)  
   
MOHK\_2012\_mean <- mean(lobster\_size\_MOHK\_2012$SIZE)  
  
MOHK\_2012\_sd <-sd(lobster\_size\_MOHK\_2012$SIZE)  
  
MOHK\_2017\_mean - MOHK\_2012\_mean

## [1] -5.253012

# Lobster sizes at MOHK in 2017 are not significantly different than in 2012  
  
MOHK\_2017\_vector <- lobster\_size\_MOHK\_2017 %>%  
 pull(SIZE)  
  
MOHK\_2012\_vector <- lobster\_size\_MOHK\_2012 %>%  
 pull(SIZE)  
  
effect\_size <- cohen.d(MOHK\_2017\_vector,MOHK\_2012\_vector)  
effect\_size

##   
## Cohen's d  
##   
## d estimate: -0.5408116 (medium)  
## 95 percent confidence interval:  
## inf sup   
## -0.8066597 -0.2749635

Mean lobsters carapace length in 2017 [72 ± 9.28 ( ± ), n=83] was signficantly less than in 2012 [77.25 ± 10.59 ( ± ), n=178] at Mohawk Reef [t(259) = -4.07, p < 0, = 0.05]. The effect size is moderate (Cohen’s d= -0.54). From the samples, lobster carapace length in 2017 was approximately 5.25 mm less than lobster caparace length in 2012.

# 4 Proportion of legal lobsters at the five sites

# Make a new data frame with numbers of lobster that are and are not above the legal minimum of 82.6 mm.   
  
legal\_2017 <- lobster\_size\_tidy\_2017 %>%  
 mutate(above\_legal= case\_when(SIZE < 82.6 ~ 'Below', SIZE >= 82.6 ~ 'Above')) %>%  
 count(SITE, above\_legal) %>%  
 spread(above\_legal, n) %>%  
 select(-SITE)  
  
legal\_2017

## # A tibble: 5 x 2  
## Above Below  
## <int> <int>  
## 1 16 51  
## 2 179 526  
## 3 130 476  
## 4 24 154  
## 5 37 75

# Rename rows in legal\_2017  
  
rownames(legal\_2017) <- c("Arroyo Quemado", "Carpinteria", "Isla Vista", "Mohawk Reef", "Naples Reef")

## Warning: Setting row names on a tibble is deprecated.

# Question: Does the proportion of lobsters of legal size differ significantly across the five sites?  
  
# Run a chi-square test  
  
legal\_2017\_prop <- prop.table(as.matrix(legal\_2017), 1)  
size\_x2 <- chisq.test(legal\_2017)  
size\_x2

##   
## Pearson's Chi-squared test  
##   
## data: legal\_2017  
## X-squared = 18.497, df = 4, p-value = 0.0009864

legal\_2017\_prop

## Above Below  
## Arroyo Quemado 0.2388060 0.7611940  
## Carpinteria 0.2539007 0.7460993  
## Isla Vista 0.2145215 0.7854785  
## Mohawk Reef 0.1348315 0.8651685  
## Naples Reef 0.3303571 0.6696429

# Make a table with proportions of legal lobsters  
  
legal\_2017\_totals <- legal\_2017 %>%  
 mutate(total= Above + Below) %>%  
 mutate(prop\_legal= Above/(Above+Below))   
   
legal\_2017\_totals

## # A tibble: 5 x 4  
## Above Below total prop\_legal  
## <int> <int> <int> <dbl>  
## 1 16 51 67 0.239  
## 2 179 526 705 0.254  
## 3 130 476 606 0.215  
## 4 24 154 178 0.135  
## 5 37 75 112 0.330

legal\_2017\_totals\_2 <- legal\_2017\_totals %>%  
 select(Above, total, prop\_legal) %>%  
 round(2)  
  
rownames(legal\_2017\_totals\_2) <- c("Arroyo Quemado", "Carpinteria", "Isla Vista", "Mohawk Reef", "Naples Reef")

## Warning: Setting row names on a tibble is deprecated.

colnames(legal\_2017\_totals\_2) <- c("Legal Lobsters", "Total Lobsters", "Proportion Legal Lobsters")

Based on the samples from each Long-Term Ecological Research Site in 2017: Arroyo Quemado (n=67), Carpinteria (n=705), Isla Vista (n=606), Mohawk Reef (n=178), and Naples Reef (n=112), the proportion of lobsters that are above the legal minimum carapace length of 82.6 mm differs significantly between sites (X-squared(4) = 18.497, p< 0.001).