ESM206-assignment4

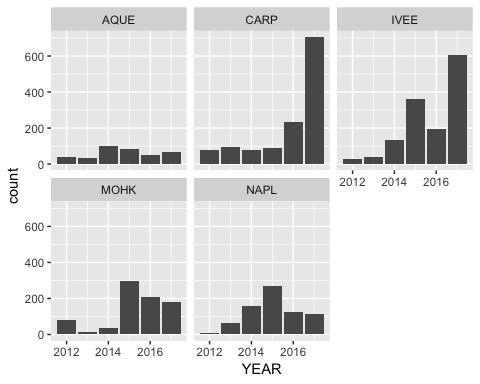
Claire Madden, Bridget Gibbons, Andrew Paterson

11/12/2018

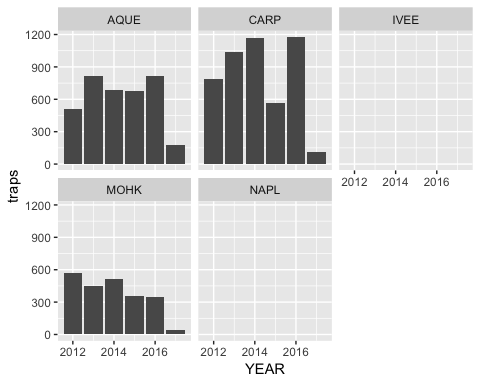
lobster\_size2 <- as.data.frame(lobster\_size)  
lobster\_size\_tidy <- expand.dft(lobster\_size2, freq = "COUNT")  
  
#If you guys don't want to change the site name here, no problem!  
size\_tidy <- lobster\_size\_tidy %>%   
 mutate(Site=  
 case\_when(  
 SITE == "NAPL" ~ "Naples Reef",  
 SITE == "AQUE" ~ "Arroyo Quemado",  
 SITE == "MOHK" ~ "Mohawk Reef",  
 SITE == "IVEE" ~ "Isla Vista",  
 SITE == "CARP" ~ "Carpinteria"  
 )  
 ) %>%   
 select(-SITE)  
  
traps\_tidy <- lobster\_traps %>%   
 mutate(Site=  
 case\_when(  
 SITE == "NAPL" ~ "Naples Reef",  
 SITE == "AQUE" ~ "Arroyo Quemado",  
 SITE == "MOHK" ~ "Mohawk Reef",  
 SITE == "IVEE" ~ "Isla Vista",  
 SITE == "CARP" ~ "Carpinteria"  
 )  
 ) %>%   
 select(-SITE)

# Part 1 (Claire)

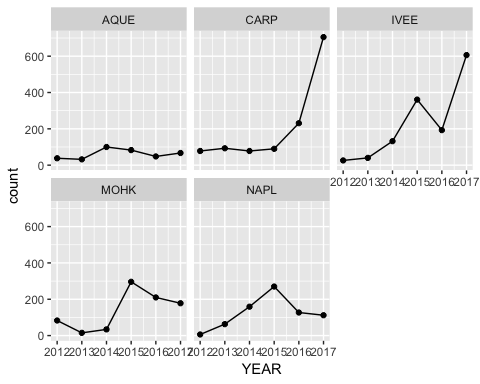
abundance\_summary <- lobster\_size %>%  
 group\_by(SITE, YEAR) %>%   
 summarize(  
 count = sum(COUNT))  
  
abundance\_column <- ggplot(abundance\_summary, aes( x = YEAR, y = count))+  
 geom\_col()+  
 facet\_wrap(~SITE)  
  
abundance\_column



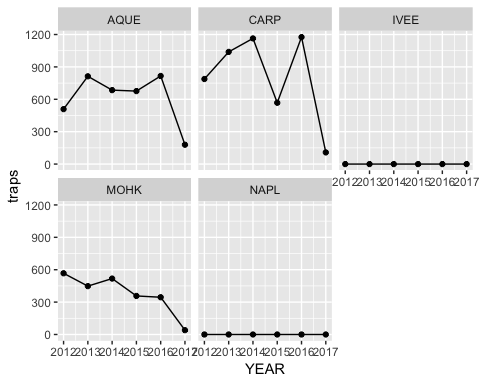
traps\_simple <- lobster\_traps %>%   
 filter(SITE == "AQUE" | SITE == "NAPL" | SITE == "MOHK" | SITE == "IVEE" | SITE == "CARP")  
  
traps\_summary <- traps\_simple %>%   
 group\_by(SITE, YEAR) %>%   
 summarize(  
 traps = sum(TRAPS))  
  
traps\_column <- ggplot(traps\_summary, aes( x = YEAR, y = traps))+  
 geom\_col()+  
 facet\_wrap(~SITE)  
  
traps\_column



abundance\_scatter <- ggplot(abundance\_summary, aes(x = YEAR, y = count))+  
 geom\_point()+  
 geom\_line()+  
 facet\_wrap(~SITE)  
  
abundance\_scatter



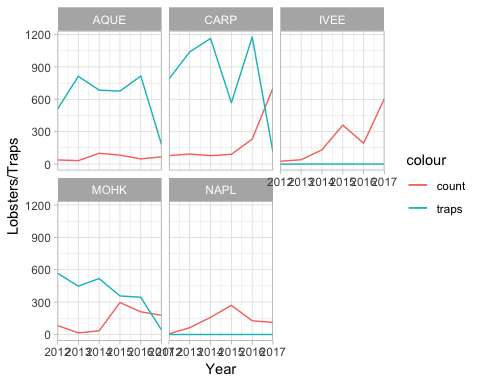
traps\_scatter <- ggplot(traps\_summary, aes(x = YEAR, y = traps))+  
 geom\_point()+  
 geom\_line()+  
 facet\_wrap(~SITE)  
  
traps\_scatter



# my attempt at joining the data before graphing and or getting both graphs for one location on the same plane  
  
abundance\_traps <- full\_join(abundance\_summary, traps\_summary)

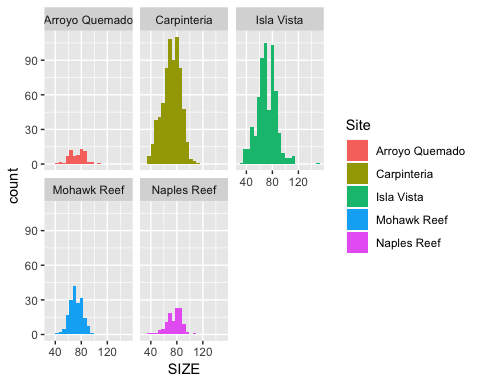
## Joining, by = c("SITE", "YEAR")

both\_scatter <- ggplot(abundance\_traps, aes(x = YEAR))+  
 geom\_line(aes(y=count, color = "count"))+  
 geom\_line(aes(y=traps, color = "traps"))+  
 facet\_wrap(~SITE)+  
 theme\_light()+  
 scale\_x\_continuous(expand = c(0,0))+  
 labs( x = "Year",  
 y = "Lobsters/Traps")  
  
both\_scatter

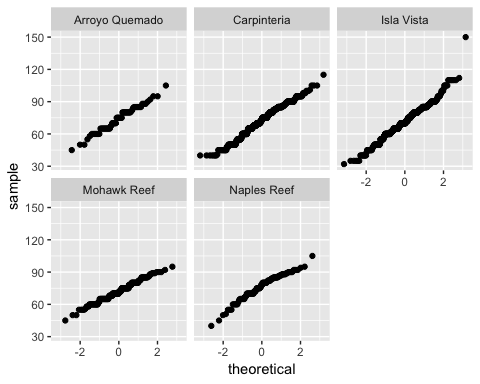


# Part 2 (Andrew)

#Filter for just 2017 results for Part 2, and Exploratory Histogram and QQ Plots  
  
size\_2017 <- size\_tidy %>%   
 filter(YEAR == "2017") %>%  
 select(SIZE, Site)  
  
carapace\_hist <- ggplot(size\_2017, aes(x = SIZE))+  
 geom\_histogram(bins = 23, aes(fill = Site))+  
 facet\_wrap(~Site)  
  
carapace\_hist



carapace\_qq <- ggplot(size\_2017, aes(sample = SIZE))+  
 geom\_qq()+  
 facet\_wrap(~Site)  
  
carapace\_qq



#The data looks roughly normally distributed, and all sites have a sample size of greater than 30.

#Levene's Test for Equal Variances, since we have more than two groups  
#H0: Variances are equal  
#HA: Variances are not equal  
  
carapace\_levene <- leveneTest(SIZE ~ Site, data = size\_2017)

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to  
## factor.

carapace\_levene

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 4 8.3893 1.065e-06 \*\*\*  
## 1663   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

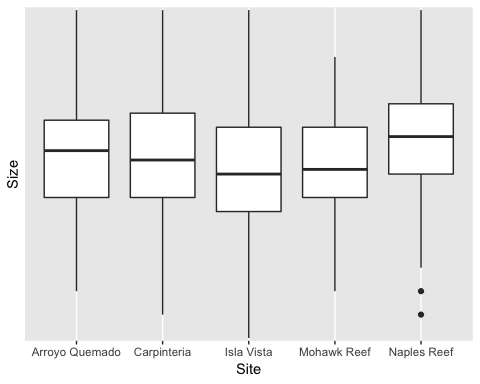
#We reject the null hypothesis of equal variances (p < .001) based on Levene's Test. However, we can examine the variances of each site to see if the largest is less than four times the smallest variance.   
  
carapace\_variances <- size\_2017 %>%   
 group\_by(Site) %>%   
 summarize(  
 Variance = var(SIZE)  
 )  
  
carapace\_variances

## # A tibble: 5 x 2  
## Site Variance  
## <chr> <dbl>  
## 1 Arroyo Quemado 141.   
## 2 Carpinteria 174.   
## 3 Isla Vista 205.   
## 4 Mohawk Reef 86.1  
## 5 Naples Reef 130.

#The largest variance is well less than four times the smallest variance, so we can assume accept the null hypothesis of equal variances and run an ANOVA.  
  
#H0: Mean carapace sizes are equal at all sites  
#HA: Mean carapace sizes vary for lobsters from at least two sites.  
  
carapace\_aov <- aov(SIZE ~ Site, data = size\_2017)  
summary(carapace\_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

#boxplot for ANOVA in case we want to include this  
lobster\_anova <- ggplot(size\_2017, aes(x = Site, y = SIZE))+  
 geom\_boxplot()+  
 scale\_x\_discrete()+  
 scale\_y\_discrete()+  
 ylab("Size")+  
 xlab("Site")  
  
lobster\_anova



#With p = .0085, we reject the null, and assume mean carapace length varies between lobsters from at least two sites. At least two populations were taken from populations with different means.  
  
#Post-hoc testing using Tukey's HSD  
  
carapace\_ph <- TukeyHSD(carapace\_aov)  
carapace\_ph

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = SIZE ~ Site, data = size\_2017)  
##   
## $Site  
## diff lwr upr p adj  
## Carpinteria-Arroyo Quemado -1.6657352 -6.24294710 2.911477 0.8582355  
## Isla Vista-Arroyo Quemado -2.4433772 -7.05292315 2.166169 0.5968998  
## Mohawk Reef-Arroyo Quemado -1.8955224 -7.02720717 3.236162 0.8514711  
## Naples Reef-Arroyo Quemado 2.3366205 -3.19311600 7.866357 0.7775633  
## Isla Vista-Carpinteria -0.7776420 -2.76097123 1.205687 0.8216104  
## Mohawk Reef-Carpinteria -0.2297872 -3.23309697 2.773523 0.9995765  
## Naples Reef-Carpinteria 4.0023556 0.36042398 7.644287 0.0228728  
## Mohawk Reef-Isla Vista 0.5478548 -2.50450730 3.600217 0.9882889  
## Naples Reef-Isla Vista 4.7799976 1.09751057 8.462485 0.0037001  
## Naples Reef-Mohawk Reef 4.2321429 -0.08607271 8.550358 0.0579286

#Lobster sizes differed significantly only between Naples Reef and Carpenteria, and Naples Reef and Isla Vista. Do we want to run Cohen's d here? It might get very clunky to convey all of the effect sizes, or maybe we could make a table showing all of the differences and effect sizes?

**Figure 1: Mean Lobster Carapace Length at Five LTER Sites in the Santa Barbara Channel**

#Comparing the lobster sizes across the sites, and putting this information in a table.  
  
size\_summary <- size\_2017 %>%   
 group\_by(Site) %>%   
 summarize("Mean Size" = round(mean(SIZE), 2),  
 "Sample Size" = length(Site))  
  
size\_summary

## # A tibble: 5 x 3  
## Site `Mean Size` `Sample Size`  
## <chr> <dbl> <int>  
## 1 Arroyo Quemado 73.9 67  
## 2 Carpinteria 72.2 705  
## 3 Isla Vista 71.4 606  
## 4 Mohawk Reef 72 178  
## 5 Naples Reef 76.2 112

size\_final <- kable(size\_summary, col.names = c("Site", "Mean Carapace Size (mm)", "Sample Size"), align = "c") %>%   
 kable\_styling(bootstrap\_options = c("striped"))  
  
size\_final

Site

Mean Carapace Size (mm)

Sample Size

Arroyo Quemado

73.90

67

Carpinteria

72.23

705

Isla Vista

71.45

606

Mohawk Reef

72.00

178

Naples Reef

76.23

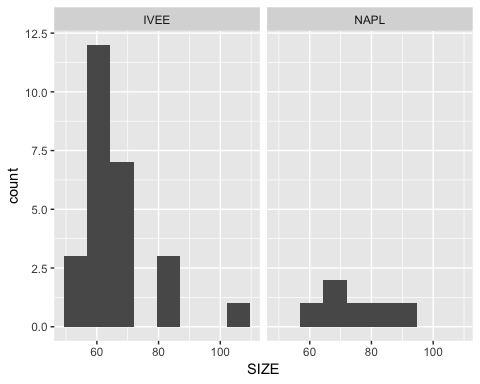
112

# Part 3 (Bridget)

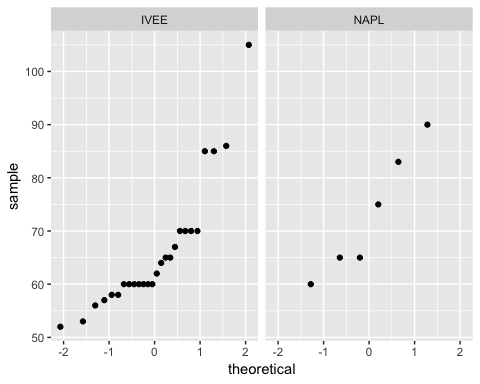
# The following chunks are comparing 2012 and 2017 for each site individually

#MPA Sites data frames  
  
mpa\_size\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE" | SITE == "NAPL", YEAR == "2012") %>%   
 select(YEAR, SITE, SIZE)  
  
mpa\_size\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE" | SITE == "NAPL", YEAR == "2017") %>%   
 select(YEAR, SITE, SIZE)

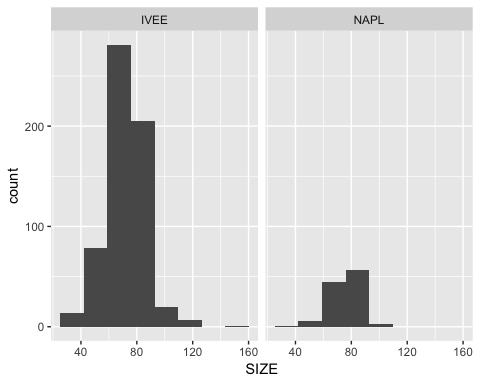
#MPA Sites exploratory histograms and qq plots  
  
mpa\_2012\_hist <- ggplot(mpa\_size\_2012, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~ SITE)  
  
mpa\_2012\_hist



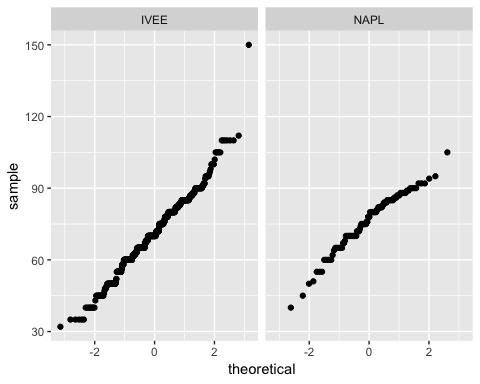
mpa\_2012\_qq <- ggplot(mpa\_size\_2012, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE)  
  
mpa\_2012\_qq



mpa\_2017\_hist <- ggplot(mpa\_size\_2017, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~ SITE)  
  
mpa\_2017\_hist

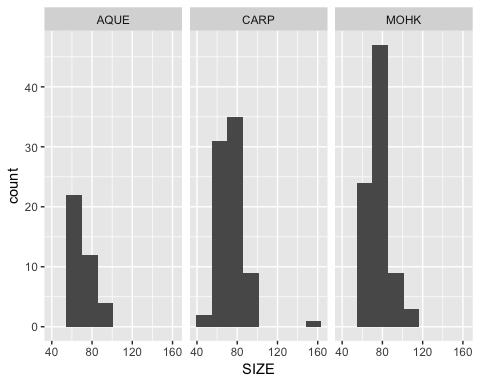


mpa\_2017\_qq <- ggplot(mpa\_size\_2017, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE)  
  
mpa\_2017\_qq

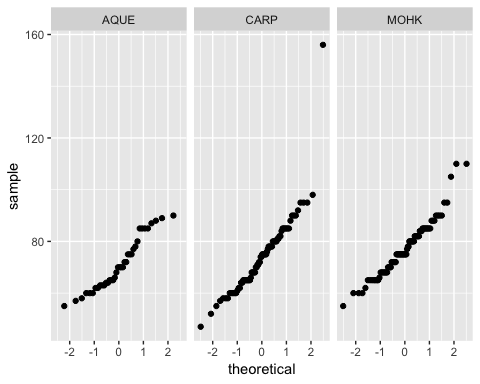


#NonMPA sites data frames  
  
nonmpa\_size\_2012 <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE" | SITE == "MOHK" | SITE == "CARP", YEAR == "2012") %>%   
 select(YEAR, SITE, SIZE)  
  
nonmpa\_size\_2017 <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE" | SITE == "MOHK" | SITE == "CARP", YEAR == "2017") %>%   
 select(YEAR, SITE, SIZE)

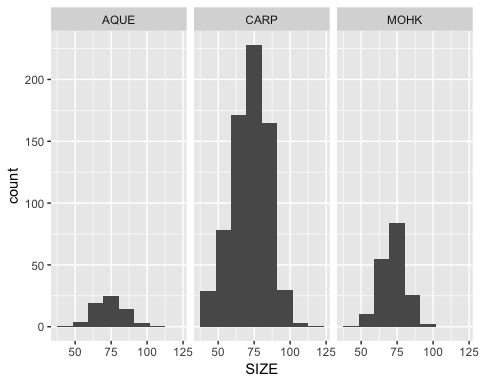
#NonMPA Sites exploratory histograms and qq plots  
  
nonmpa\_2012\_hist <- ggplot(nonmpa\_size\_2012, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~ SITE)  
  
nonmpa\_2012\_hist



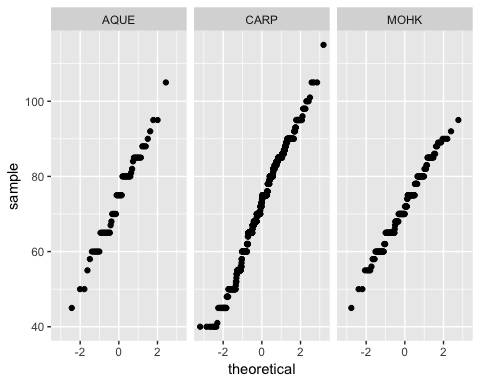
nonmpa\_2012\_qq <- ggplot(nonmpa\_size\_2012, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE)  
  
nonmpa\_2012\_qq



nonmpa\_2017\_hist <- ggplot(nonmpa\_size\_2017, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~ SITE)  
  
nonmpa\_2017\_hist



nonmpa\_2017\_qq <- ggplot(nonmpa\_size\_2017, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE)  
  
nonmpa\_2017\_qq



#Site 1: Arroyo Quemado - nonMPA  
  
#Make vectors for the 2012 and 2017 values  
  
aq\_2012 <- nonmpa\_size\_2012 %>%   
 filter(SITE == "AQUE") %>%   
 pull(SIZE)  
  
aq\_2017 <- nonmpa\_size\_2017 %>%   
 filter(SITE == "AQUE") %>%   
 pull(SIZE)  
  
#Run an F-test for equal variance  
#H0: ratio of variances = 1  
#HA: ratio of variances NOT = 1  
  
aq\_ftest <- var.test(aq\_2012, aq\_2017)  
aq\_ftest

##   
## F test to compare two variances  
##   
## data: aq\_2012 and aq\_2017  
## F = 0.72863, num df = 37, denom df = 66, p-value = 0.2986  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.419142 1.327868  
## sample estimates:  
## ratio of variances   
## 0.7286314

#Retain null, variances are equal. Run student's t-test  
  
aq\_ttest <- t.test(aq\_2012, aq\_2017, var.equal = TRUE)  
aq\_ttest

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

#p-valye of 0.2097, no significant difference in lobster sizes  
  
#calculate effect size  
  
aq\_effsize <- cohen.d(aq\_2012, aq\_2017)  
aq\_effsize

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

#effect size is small (-0.256)

#Site 2: Carpinteria - nonMPA  
  
#Make vectors for the 2012 and 2017 values  
  
carp\_2012 <- nonmpa\_size\_2012 %>%   
 filter(SITE == "CARP") %>%   
 pull(SIZE)  
  
carp\_2017 <- nonmpa\_size\_2017 %>%   
 filter(SITE == "CARP") %>%   
 pull(SIZE)  
  
#Run an F-test for equal variance  
#H0: ratio of variances = 1  
#HA: ratio of variances NOT = 1  
  
carp\_ftest <- var.test(carp\_2012, carp\_2017)  
carp\_ftest

##   
## F test to compare two variances  
##   
## data: carp\_2012 and carp\_2017  
## F = 1.2244, num df = 77, denom df = 704, p-value = 0.2043  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.896208 1.750406  
## sample estimates:  
## ratio of variances   
## 1.224405

#Retain null, variances are equal. Run student's t-test  
  
carp\_ttest <- t.test(carp\_2012, carp\_2017, var.equal = TRUE)  
carp\_ttest

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

#p-valye of 0.1819, no significant difference in lobster sizes  
  
#calculate effect size  
  
carp\_effsize <- cohen.d(carp\_2012, carp\_2017)  
carp\_effsize

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

#effect size is negligible (-0.159)

#Site 3: Mohawk Reef - nonMPA  
  
#Make vectors for the 2012 and 2017 values  
  
mohk\_2012 <- nonmpa\_size\_2012 %>%   
 filter(SITE == "MOHK") %>%   
 pull(SIZE)  
  
mohk\_2017 <- nonmpa\_size\_2017 %>%   
 filter(SITE == "MOHK") %>%   
 pull(SIZE)  
  
#Run an F-test for equal variance  
#H0: ratio of variances = 1  
#HA: ratio of variances NOT = 1  
  
mohk\_ftest <- var.test(mohk\_2012, mohk\_2017)  
mohk\_ftest

##   
## F test to compare two variances  
##   
## data: mohk\_2012 and mohk\_2017  
## F = 1.3015, num df = 82, denom df = 177, p-value = 0.1509  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.9085131 1.9131403  
## sample estimates:  
## ratio of variances   
## 1.301535

#Retain null, variances are equal. Run student's t-test  
  
mohk\_ttest <- t.test(mohk\_2012, mohk\_2017, var.equal = TRUE)  
mohk\_ttest

##   
## Two Sample t-test  
##   
## data: mohk\_2012 and mohk\_2017  
## t = 4.0689, df = 259, p-value = 6.276e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.710776 7.795248  
## sample estimates:  
## mean of x mean of y   
## 77.25301 72.00000

#p-valye < 0.001, significant difference in lobster sizes  
  
#calculate effect size  
  
mohk\_effsize <- cohen.d(mohk\_2012, mohk\_2017)  
mohk\_effsize

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

#effect size is medium (0.54)

#Site 4: Isla Vista - MPA  
  
#Make vectors for the 2012 and 2017 values  
  
iv\_2012 <- mpa\_size\_2012 %>%   
 filter(SITE == "IVEE") %>%   
 pull(SIZE)  
  
iv\_2017 <- mpa\_size\_2017 %>%   
 filter(SITE == "IVEE") %>%   
 pull(SIZE)  
  
#Run an F-test for equal variance  
#H0: ratio of variances = 1  
#HA: ratio of variances NOT = 1  
  
iv\_ftest <- var.test(iv\_2012, iv\_2017)  
iv\_ftest

##   
## F test to compare two variances  
##   
## data: iv\_2012 and iv\_2017  
## F = 0.71311, num df = 25, denom df = 605, p-value = 0.307  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.4322948 1.3698611  
## sample estimates:  
## ratio of variances   
## 0.713111

#Retain null, variances are equal. Run student's t-test  
  
iv\_ttest <- t.test(iv\_2012, iv\_2017, var.equal = TRUE)  
iv\_ttest

##   
## Two Sample t-test  
##   
## data: iv\_2012 and iv\_2017  
## t = -1.885, df = 630, p-value = 0.0599  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -10.9750916 0.2246473  
## sample estimates:  
## mean of x mean of y   
## 66.07692 71.45215

#p-valye of 0.0599, no significant difference in lobster sizes  
  
#calculate effect size  
  
iv\_effsize <- cohen.d(iv\_2012, iv\_2017)  
iv\_effsize

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

#effect size is small (-0.377)

#Site 5: Naples Reef - MPA  
  
#Make vectors for the 2012 and 2017 values  
  
napl\_2012 <- mpa\_size\_2012 %>%   
 filter(SITE == "NAPL") %>%   
 pull(SIZE)  
  
napl\_2017 <- mpa\_size\_2017 %>%   
 filter(SITE == "NAPL") %>%   
 pull(SIZE)  
  
#Run an F-test for equal variance  
#H0: ratio of variances = 1  
#HA: ratio of variances NOT = 1  
  
napl\_ftest <- var.test(napl\_2012, napl\_2017)  
napl\_ftest

##   
## F test to compare two variances  
##   
## data: napl\_2012 and napl\_2017  
## F = 1.064, num df = 5, denom df = 111, p-value = 0.7685  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3966019 6.4626426  
## sample estimates:  
## ratio of variances   
## 1.064048

#Retain null, variances are equal. Run student's t-test  
  
napl\_ttest <- t.test(napl\_2012, napl\_2017, var.equal = TRUE)  
napl\_ttest

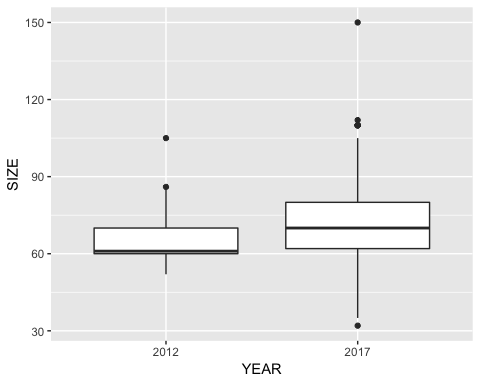
##   
## Two Sample t-test  
##   
## data: napl\_2012 and napl\_2017  
## t = -0.67636, df = 116, p-value = 0.5002  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -12.697051 6.232765  
## sample estimates:  
## mean of x mean of y   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

#p-valye of 0.5002, no significant difference in lobster sizes  
  
#calculate effect size  
  
napl\_effsize <- cohen.d(napl\_2012, napl\_2017)  
napl\_effsize

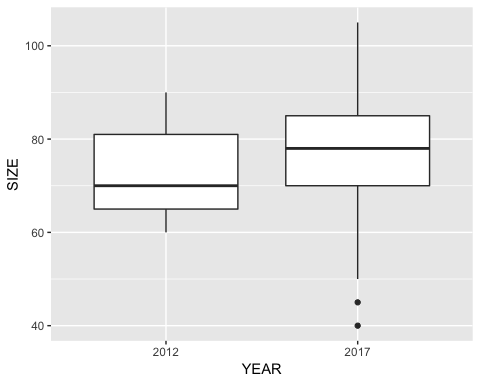
##   
## Cohen's d  
##   
## d estimate: -0.2834216 (small)  
## 95 percent confidence interval:  
## inf sup   
## -1.1141889 0.5473456

#effect size is small (-0.283)

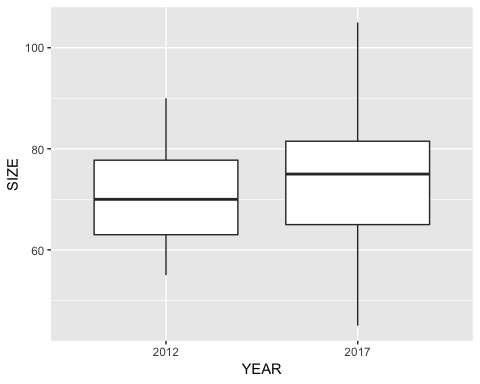
#Make boxplots for MPA sites  
  
#Isla Vista  
  
iv\_box <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE", YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE) %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
iv\_boxplot <- ggplot(iv\_box, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
iv\_boxplot



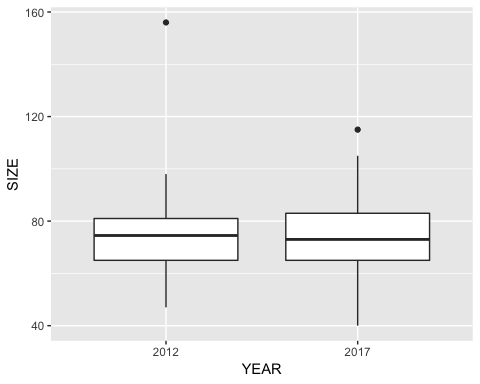
#Naples  
  
napl\_box <- lobster\_size\_tidy %>%   
 filter(SITE == "NAPL", YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE) %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
napl\_boxplot <- ggplot(napl\_box, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
napl\_boxplot



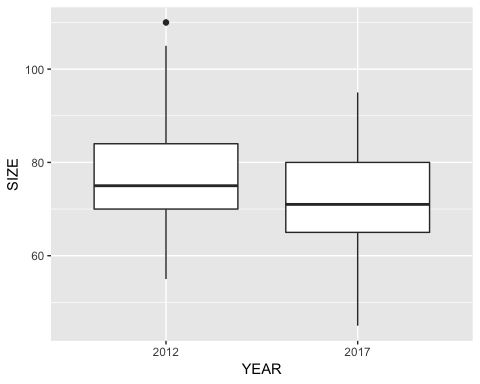
#Make boxplots for Non-MPA sites  
  
#Arroyo Quemado  
  
aq\_box <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE", YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE) %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
aq\_boxplot <- ggplot(aq\_box, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
aq\_boxplot



#Carpinteria  
  
carp\_box <- lobster\_size\_tidy %>%   
 filter(SITE == "CARP", YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE) %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
carp\_boxplot <- ggplot(carp\_box, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
carp\_boxplot



#Mohawk Reef  
  
mohk\_box <- lobster\_size\_tidy %>%   
 filter(SITE == "MOHK", YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE) %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
mohk\_boxplot <- ggplot(mohk\_box, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
mohk\_boxplot



# Part 4

#Make table with counts of legal and illegal lobsters for each site  
  
legal\_sizes <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE" | SITE == "NAPL" | SITE == "MOHK" | SITE == "AQUE" | SITE == "CARP") %>%   
 filter(YEAR == 2017) %>%   
 filter(SIZE != "-99999") %>%   
 select(YEAR, SITE, SIZE) %>%   
 mutate(legality = ifelse(SIZE > 83, "legal", "illegal")) %>%   
 count(SITE, legality) %>%   
 spread(legality, n) %>%   
 select(-SITE)  
  
rownames(legal\_sizes) <- c("AQUE", "CARP", "IVEE", "MOHK", "NAPL")

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

#Make table with proportions  
  
legal\_prop <- prop.table(as.matrix(legal\_sizes), 1)  
  
legal\_prop

## illegal legal  
## AQUE 0.7611940 0.2388060  
## CARP 0.7758865 0.2241135  
## IVEE 0.8052805 0.1947195  
## MOHK 0.8764045 0.1235955  
## NAPL 0.6875000 0.3125000

#Run chi square test on the legalilty table (original data, not prop table)  
  
legal\_x2 <- chisq.test(legal\_sizes)   
legal\_x2

##   
## Pearson's Chi-squared test  
##   
## data: legal\_sizes  
## X-squared = 17.178, df = 4, p-value = 0.001785