ESM206-assignment4

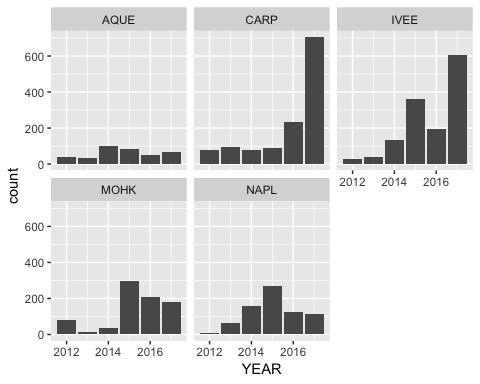
Claire Madden, Bridget Gibbons, Andrew Paterson

11/12/2018

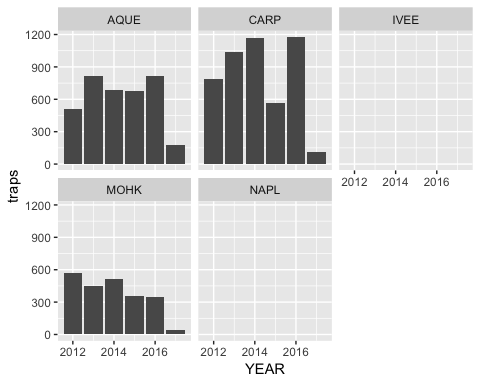
lobster\_size2 <- as.data.frame(lobster\_size)  
lobster\_size\_tidy <- expand.dft(lobster\_size2, freq = "COUNT")  
  
#If you guys don't want to change the site name here, no problem!  
size\_tidy <- lobster\_size\_tidy %>%   
 mutate(Site=  
 case\_when(  
 SITE == "NAPL" ~ "Naples Reef",  
 SITE == "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)

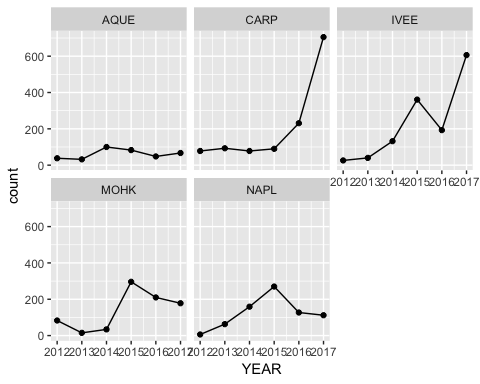
abundance\_summary <- lobster\_size %>%  
 group\_by(SITE, YEAR) %>%   
 summarize(  
 count = sum(COUNT))  
  
abundance\_column <- ggplot(abundance\_summary, aes( x = YEAR, y = count))+  
 geom\_col()+  
 facet\_wrap(~SITE)  
  
abundance\_column



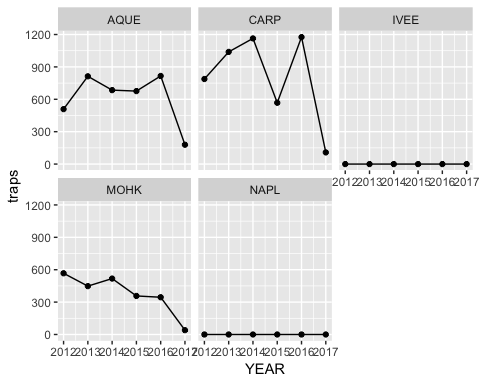
traps\_simple <- lobster\_traps %>%   
 filter(SITE == "AQUE" | SITE == "NAPL" | SITE == "MOHK" | SITE == "IVEE" | SITE == "CARP")  
  
traps\_summary <- traps\_simple %>%   
 group\_by(SITE, YEAR) %>%   
 summarize(  
 traps = sum(TRAPS))  
  
traps\_column <- ggplot(traps\_summary, aes( x = YEAR, y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



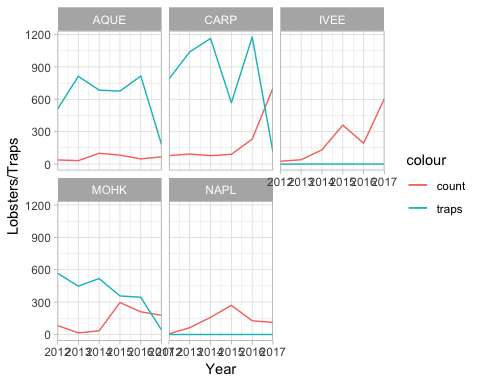
traps\_scatter <- ggplot(traps\_summary, aes(x = YEAR, y = traps))+  
 geom\_point()+  
 geom\_line()+  
 facet\_wrap(~SITE)  
  
traps\_scatter



# 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)

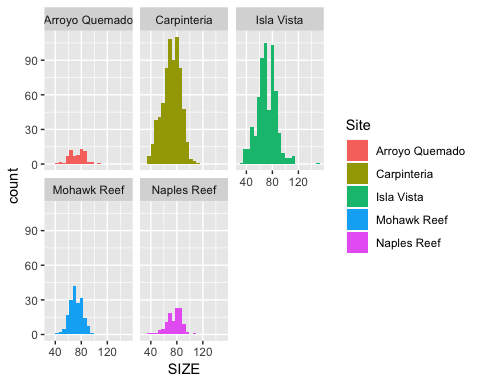
## Joining, by = c("SITE", "YEAR")

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

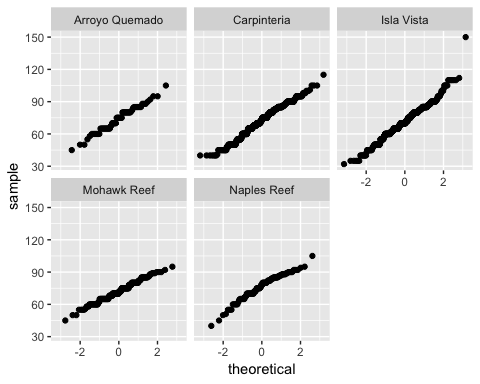


# Part 2 (Andrew)

#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

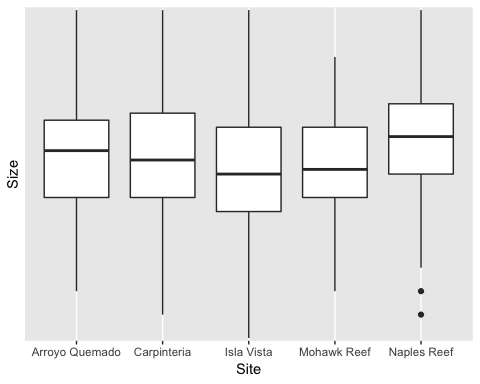
#We reject the null hypothesis of equal variances (p < .001) based on Levene's Test. However, we can examine the variances of each site to see if the largest is less than four times the smallest variance.   
  
carapace\_variances <- size\_2017 %>%   
 group\_by(Site) %>%   
 summarize(  
 Variance =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

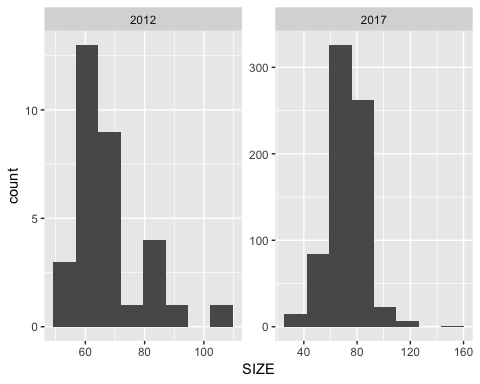
# Part 3 (Bridget)

## These chunks are lumping all the MPV sites together and then comparing 2012 and 2017, and then doing the same thing for non-MPV sites. If we need to do it site by site, this all needs to be changed

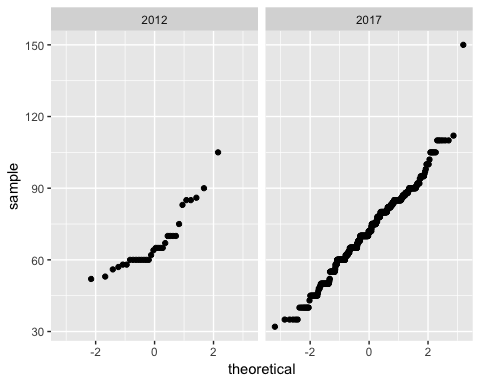
# MPA sites: Isla Vista, Naples Reef

# Non-MPA sites:Arroyo Quemado, Carpinteria, Mohawk Reef

#For the MPA sites  
  
mpa\_size <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE" | SITE == "NAPL") %>%   
 filter(YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE)  
   
   
#Need to test for normality to determine which kind of test to run  
  
mpa\_size\_hist <- ggplot(mpa\_size, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~YEAR, scale = "free")  
   
mpa\_size\_hist



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



#The data for 2017 looks normally distributed, but the qq plot for 2012 doesn't look linear. We have more than 30 samples for each year though so we can still use a t-test  
  
mpa\_2012 <- mpa\_size %>%   
 filter(YEAR == "2012") %>%   
 pull(SIZE)  
  
mpa\_2017 <- mpa\_size %>%   
 filter(YEAR == "2017") %>%   
 pull(SIZE)  
  
#Run an F-test to see which type of t-test we can use  
#H0: ratio of variances is equal to 1  
#HA: ratio of variances is NOT equal to 1  
  
mpa\_ftest <- var.test(mpa\_2012, mpa\_2017)  
mpa\_ftest

##   
## F test to compare two variances  
##   
## data: mpa\_2012 and mpa\_2017  
## F = 0.75323, num df = 31, denom df = 717, p-value = 0.3346  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.477719 1.341900  
## sample estimates:  
## ratio of variances   
## 0.7532319

#Based on result of F-test (p = 0.3346), we retain null hypothesis that variances are equal. We can use a student's t-test, which is more powerful than a Welch's test  
  
#H0 - difference in means is 0  
#HA - difference in means is NOT 0  
  
mpa\_ttest <- t.test(mpa\_2012, mpa\_2017, var.equal = TRUE)  
mpa\_ttest

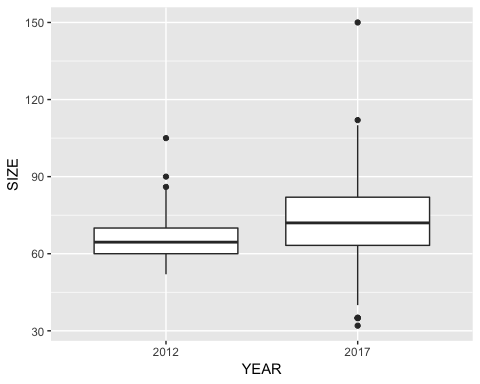
##   
## Two Sample t-test  
##   
## data: mpa\_2012 and mpa\_2017  
## t = -1.9159, df = 748, p-value = 0.05576  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -9.7644724 0.1189292  
## sample estimates:  
## mean of x mean of y   
## 67.37500 72.19777

#With p-value of 0.05576, we retain null hypothesis. There is not a significant difference in mean lobster size in MPAs between 2012 and 2017.  
  
#Calculate effect size  
  
mpa\_eff\_size <- cohen.d(mpa\_2012, mpa\_2017)  
mpa\_eff\_size

##   
## Cohen's d  
##   
## d estimate: -0.3461506 (small)  
## 95 percent confidence interval:  
## inf sup   
## -0.701270908 0.008969759

#There is a small effect size (-0.34)

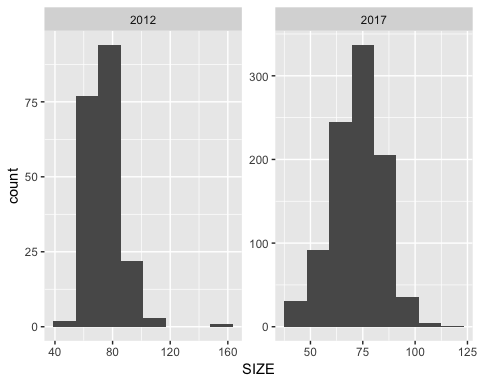
#Boxplot for MPA  
mpa\_size\_ordered <- mpa\_size %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
mpa\_size\_boxplot <- ggplot(mpa\_size\_ordered, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
mpa\_size\_boxplot



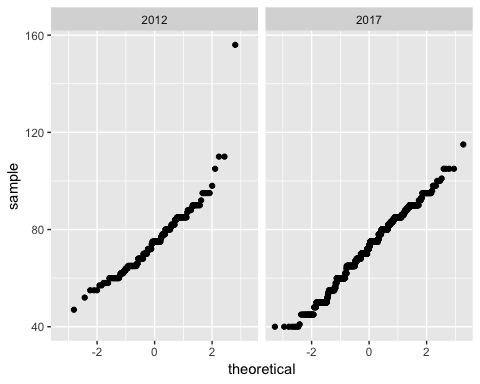
#Summary table with mean values  
  
mpa\_size\_summary <- mpa\_size\_ordered %>%  
 summarize(  
 size = mean(SIZE),  
 n = length(YEAR)  
 )  
  
mpa\_size\_summary

## # A tibble: 2 x 3  
## YEAR size n  
## <fct> <dbl> <int>  
## 1 2012 67.4 32  
## 2 2017 72.2 718

#For the non-mpa sites  
  
non\_mpa\_size <- lobster\_size\_tidy %>%   
 filter(SITE == "AQUE" | SITE == "MOHK" | SITE == "CARP") %>%   
 filter(YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE)  
   
   
#Need to test for normality to determine which kind of test to run  
  
non\_mpa\_size\_hist <- ggplot(non\_mpa\_size, aes(x = SIZE)) +   
 geom\_histogram(bins = 8) +  
 facet\_wrap(~YEAR, scale = "free")  
   
non\_mpa\_size\_hist



non\_mpa\_size\_qq <- ggplot (non\_mpa\_size, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~YEAR)  
  
non\_mpa\_size\_qq



#The data looks normally distributed  
  
non\_mpa\_2012 <- non\_mpa\_size %>%   
 filter(YEAR == "2012") %>%   
 pull(SIZE)  
  
non\_mpa\_2017 <- non\_mpa\_size %>%   
 filter(YEAR == "2017") %>%   
 pull(SIZE)  
  
#Run an F-test to see which type of t-test we can use  
#H0: ratio of variances is equal to 1  
#HA: ratio of variances is NOT equal to 1  
  
non\_mpa\_ftest <- var.test(non\_mpa\_2012, non\_mpa\_2017)  
non\_mpa\_ftest

##   
## F test to compare two variances  
##   
## data: non\_mpa\_2012 and non\_mpa\_2017  
## F = 0.99085, num df = 198, denom df = 949, p-value = 0.953  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8037718 1.2406929  
## sample estimates:  
## ratio of variances   
## 0.9908519

#Based on result of F-test (p = 0.953), we retain null hypothesis that variances are equal. We can use a student's t-test, which is more powerful than a Welch's test  
  
#H0 - difference in means is 0  
#HA - difference in means is NOT 0  
  
non\_mpa\_ttest <- t.test(non\_mpa\_2012, non\_mpa\_2017, var.equal = TRUE)  
non\_mpa\_ttest

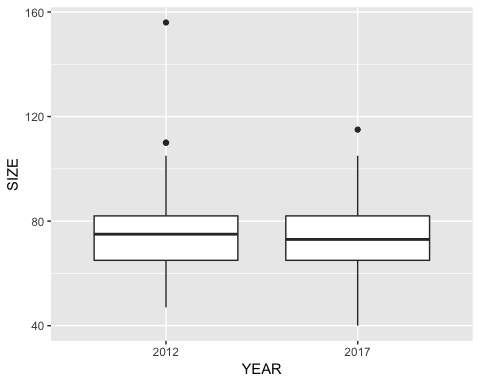
##   
## Two Sample t-test  
##   
## data: non\_mpa\_2012 and non\_mpa\_2017  
## t = 2.6973, df = 1147, p-value = 0.007093  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7143078 4.5265173  
## sample estimates:  
## mean of x mean of y   
## 74.92462 72.30421

#With p-value of 0.007, we reject the null hypothesis. There is a significant difference in mean lobster size in non-MPAs between 2012 and 2017.  
  
#Calculate effect size  
  
non\_mpa\_eff\_size <- cohen.d(non\_mpa\_2012, non\_mpa\_2017)  
non\_mpa\_eff\_size

##   
## Cohen's d  
##   
## d estimate: 0.2102816 (small)  
## 95 percent confidence interval:  
## inf sup   
## 0.05707948 0.36348365

#There is a small effect size (0.21)

#Boxplot for Non-MPA  
non\_mpa\_size\_ordered <- non\_mpa\_size %>%   
 mutate(YEAR = factor(YEAR)) %>%   
 group\_by(YEAR)  
  
non\_mpa\_size\_boxplot <- ggplot(non\_mpa\_size\_ordered, aes(x = YEAR, y = SIZE)) +  
 geom\_boxplot()  
  
non\_mpa\_size\_boxplot

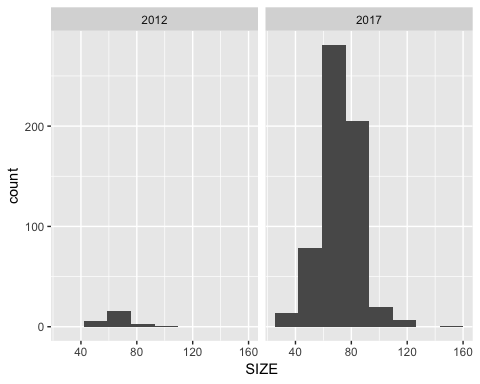


#Summary table with mean values  
  
non\_mpa\_size\_summary <- non\_mpa\_size\_ordered %>%  
 summarize(  
 size = mean(SIZE),  
 n = length(YEAR)  
 )  
  
non\_mpa\_size\_summary

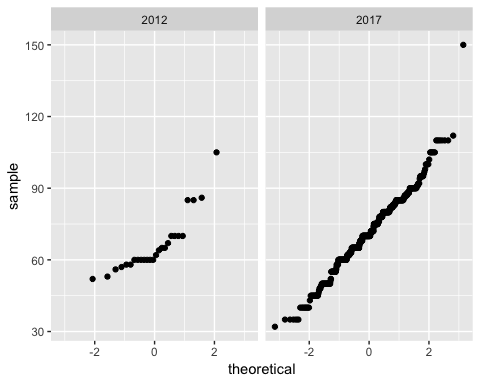
## # A tibble: 2 x 3  
## YEAR size n  
## <fct> <dbl> <int>  
## 1 2012 74.9 199  
## 2 2017 72.3 950

# The following chunks are comparing 2012 and 2017 for each site individually - doesn’t seem like these make sense, so probably go with the above analysis

#Isla Vista  
  
lobster\_sizes\_iv <- lobster\_size\_tidy %>%   
 filter(SITE == "IVEE") %>%   
 filter(YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE)  
  
#Create exploratory histogram and qq plot  
  
iv\_size\_hist <- ggplot(lobster\_sizes\_iv, aes(x = SIZE)) +  
 geom\_histogram(bins = 8) +  
 facet\_wrap(~YEAR)  
  
iv\_size\_hist



iv\_size\_qq <- ggplot(lobster\_sizes\_iv, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~YEAR)  
  
iv\_size\_qq

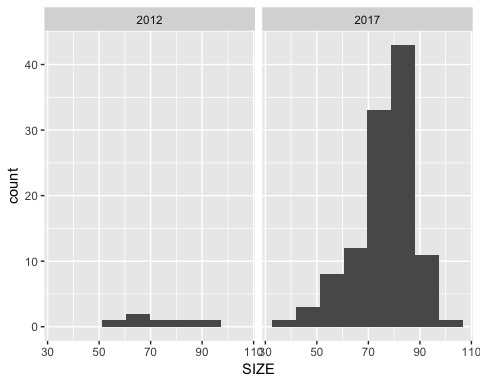


iv\_summary <- lobster\_sizes\_iv %>%   
 group\_by(YEAR) %>%   
 summarize(mean = mean(SIZE), n = length(SIZE))  
  
iv\_summary

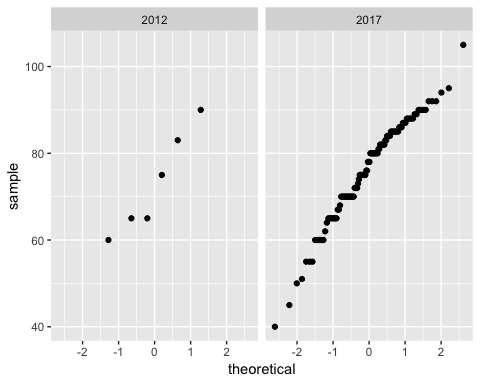
## # A tibble: 2 x 3  
## YEAR mean n  
## <int> <dbl> <int>  
## 1 2012 66.1 26  
## 2 2017 71.5 606

#QQ plot for 2012 does not look normal, only have 26 observations so can't use CLT

#Naples  
  
lobster\_sizes\_napl <- lobster\_size\_tidy %>%   
 filter(SITE == "NAPL") %>%   
 filter(YEAR == "2012" | YEAR == "2017") %>%   
 select(YEAR, SIZE)  
  
#Create exploratory histogram and qq plot  
  
napl\_size\_hist <- ggplot(lobster\_sizes\_napl, aes(x = SIZE)) +  
 geom\_histogram(bins = 8) +  
 facet\_wrap(~YEAR)  
  
napl\_size\_hist



napl\_size\_qq <- ggplot(lobster\_sizes\_napl, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~YEAR)  
  
napl\_size\_qq



napl\_summary <- lobster\_sizes\_napl %>%   
 group\_by(YEAR) %>%   
 summarize(mean = mean(SIZE), n = length(SIZE))  
  
napl\_summary

## # A tibble: 2 x 3  
## YEAR mean n  
## <int> <dbl> <int>  
## 1 2012 73 6  
## 2 2017 76.2 112

# 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