assignment4

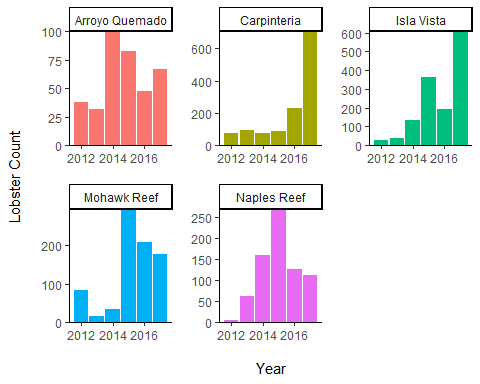
na

November 14, 2018

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_part 1\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
#column graphs of lobster abundance by year at each of the 5 sites  
size\_summary <- lsa\_filt %>%   
 group\_by(SITE, YEAR) %>%   
 filter(YEAR == "2012" | YEAR == "2013" | YEAR == "2014" | YEAR == "2015" | YEAR == "2016" | YEAR == "2017") %>%   
 summarize(  
 lobster\_count = length(SIZE)  
 )  
size\_summary

## # A tibble: 30 x 3  
## # Groups: SITE [?]  
## SITE YEAR lobster\_count  
## <chr> <int> <int>  
## 1 AQUE 2012 38  
## 2 AQUE 2013 32  
## 3 AQUE 2014 100  
## 4 AQUE 2015 83  
## 5 AQUE 2016 48  
## 6 AQUE 2017 67  
## 7 CARP 2012 78  
## 8 CARP 2013 93  
## 9 CARP 2014 78  
## 10 CARP 2015 90  
## # ... with 20 more rows

labels <- c(AQUE = "Arroyo Quemado", CARP = "Carpinteria", IVEE = "Isla Vista", MOHK = "Mohawk Reef", NAPL = "Naples Reef")  
abundance <- ggplot(size\_summary, aes(x = YEAR, y = lobster\_count))+  
 geom\_col(aes(fill = SITE)) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 theme\_classic() +  
 xlab("\nYear") +  
 ylab("Lobster Count\n") +  
 facet\_wrap(~SITE, scale = "free", labeller = labeller(SITE = labels)) +  
 theme(legend.position = "none") +  
 theme(panel.spacing = unit(1, "lines"))  
abundance

 **Figure 1. Abundance of the California Spiny Lobster at Five Long-Term Ecological Research Sites in the Santa Barbara Channel (2012-2017).** Lobster abundance data for Panulirus interruptus are collected by divers at five sites each year in late summer, before the start of the fishing season. These sites include Arroyo Quemado, Naples Reef, Mohawk Reef, Isla Vista, and Carpinteria. Naples Reef and Isla Vista are within marine protected areas; Arroyo Quemado, Mohawk, and Carpenteria are outside marine protected areas. Source: Reed, D. 2017. SBC LTER: Reef: Abundance, size and fishing effort for California Spiny Lobster (Panulirus interruptus), ongoing since 2012. Santa Barbara Coastal Long Term Ecological Research Project.

#Column graph of fishing pressure (trap buoys) by year at each of the 5 sites  
trap\_summary <- lt\_filt %>%   
 group\_by(SITE, YEAR) %>%   
 filter(YEAR == "2012" | YEAR == "2013" | YEAR == "2014" | YEAR == "2015" | YEAR == "2016" | YEAR == "2017") %>%   
 summarize(  
 trap\_count = sum(TRAPS)  
 )  
traps <- ggplot(lt\_filt, aes(x = YEAR, y = TRAPS))+  
 geom\_col(aes(fill = SITE)) +  
 scale\_y\_continuous(expand = c(0,0)) +  
 theme\_classic() +  
 xlab("\nYear") +  
 ylab("Number of Traps\n") +  
 facet\_wrap(~SITE, scale = "free", labeller = labeller(SITE = labels)) +  
 theme(legend.position = "none") +  
 theme(panel.spacing = unit(1, "lines"))  
   
traps

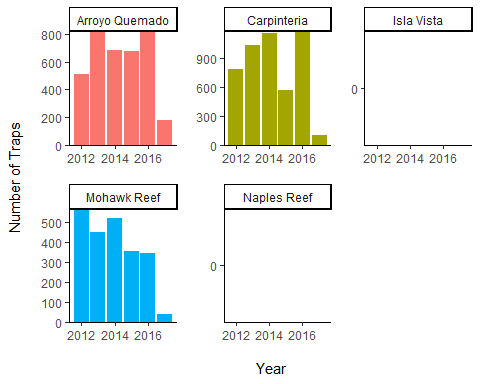
 **Figure 2. Lobster Trap Floats at Five Long-Term Ecological Research Sites in the Santa Barbara Channel (2012-2017).** Commerical lobster trap float counts are recorded every two to four weeks during the lobster fishing season (October to March). The abundance of lobster trap buoys are an indicator of fishing pressure. Since Naples Reef and Isla Vista are within Marine Protected Areas, there are no commerical trap floats deployed. Source: Reed, D. 2017. SBC LTER: Reef: Abundance, size and fishing effort for California Spiny Lobster (Panulirus interruptus), ongoing since 2012. Santa Barbara Coastal Long Term Ecological Research Project.

Table 1. Lobster abundance (counts) and fishing pressure (trap buoys) at five Long Term Ecological Research Sites in Santa Barbara Channel (2012 -2017).

lsa\_lt\_joined <- left\_join(size\_summary, trap\_summary, by = c("SITE", "YEAR"))  
lsa\_lt\_joined

## # A tibble: 30 x 4  
## # Groups: SITE [?]  
## SITE YEAR lobster\_count trap\_count  
## <chr> <int> <int> <int>  
## 1 AQUE 2012 38 509  
## 2 AQUE 2013 32 813  
## 3 AQUE 2014 100 685  
## 4 AQUE 2015 83 676  
## 5 AQUE 2016 48 816  
## 6 AQUE 2017 67 179  
## 7 CARP 2012 78 788  
## 8 CARP 2013 93 1039  
## 9 CARP 2014 78 1164  
## 10 CARP 2015 90 568  
## # ... with 20 more rows

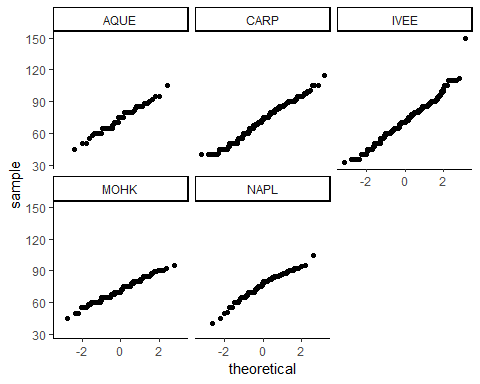
lsa\_lt\_table <- kable(lsa\_lt\_joined,   
 format = "markdown",  
 col.names = c("Site", "Year", "Lobster Count", "Trap Count")) %>%   
 kable\_styling(bootstrap\_options = c("border","condensed","striped"))

## Warning in kable\_styling(., bootstrap\_options = c("border", "condensed", :  
## Please specify format in kable. kableExtra can customize either HTML or  
## LaTeX outputs. See https://haozhu233.github.io/kableExtra/ for details.

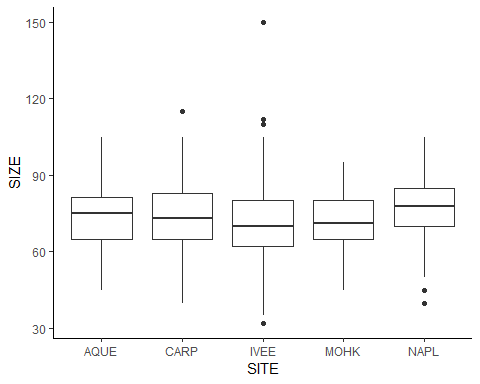
lsa\_lt\_table

|  |  |  |  |
| --- | --- | --- | --- |
| Site | Year | Lobster Count | Trap Count |
| AQUE | 2012 | 38 | 509 |
| AQUE | 2013 | 32 | 813 |
| AQUE | 2014 | 100 | 685 |
| AQUE | 2015 | 83 | 676 |
| AQUE | 2016 | 48 | 816 |
| AQUE | 2017 | 67 | 179 |
| CARP | 2012 | 78 | 788 |
| CARP | 2013 | 93 | 1039 |
| CARP | 2014 | 78 | 1164 |
| CARP | 2015 | 90 | 568 |
| CARP | 2016 | 231 | 1176 |
| CARP | 2017 | 705 | 108 |
| IVEE | 2012 | 26 | 0 |
| IVEE | 2013 | 40 | 0 |
| IVEE | 2014 | 132 | 0 |
| IVEE | 2015 | 361 | 0 |
| IVEE | 2016 | 193 | 0 |
| IVEE | 2017 | 606 | 0 |
| MOHK | 2012 | 83 | 567 |
| MOHK | 2013 | 15 | 448 |
| MOHK | 2014 | 34 | 518 |
| MOHK | 2015 | 296 | 357 |
| MOHK | 2016 | 210 | 345 |
| MOHK | 2017 | 178 | 40 |
| NAPL | 2012 | 6 | 0 |
| NAPL | 2013 | 63 | 0 |
| NAPL | 2014 | 159 | 0 |
| NAPL | 2015 | 270 | 0 |
| NAPL | 2016 | 127 | 0 |
| NAPL | 2017 | 112 | 0 |

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_part 2\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
  
# filter for only 2017  
lsa\_filt2017 <- lsa\_filt %>%  
 filter(YEAR == 2017)  
  
# look at site distributions  
qq\_site\_means <- ggplot(lsa\_filt2017, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~ SITE) +  
 theme\_classic()  
qq\_site\_means



box\_site\_means <- ggplot(lsa\_filt2017, aes(x = SITE, y = SIZE)) +  
 geom\_boxplot() +  
 theme\_classic()  
box\_site\_means



# convert site to factor in lsa\_filt  
lsa\_filt2017$SITE <- as.factor(lsa\_filt2017$SITE)  
  
# find site means, variance, etc.  
lsa\_site\_means <- lsa\_filt2017 %>%  
 group\_by(MPA, SITE) %>%  
 summarize(  
 mean\_size = mean(SIZE),  
 sd\_size = sd(SIZE),  
 var\_size = var(SIZE),  
 n = length(SIZE)  
 ) %>%  
 arrange(MPA) %>%  
 mutate(yloc = mean\_size + sd\_size + 3.5) # loc of annotation  
  
lsa\_site\_means$SITE <- factor(lsa\_site\_means$SITE, levels = c("IVEE","NAPL","AQUE","MOHK","CARP"))  
  
  
# levenes test for equal variances  
lsa\_levene <- leveneTest(SIZE ~ SITE, data = lsa\_filt2017)  
lsa\_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

lobster\_anova <- aov(SIZE ~ SITE, data = lsa\_filt2017)  
summary(lobster\_anova)

## 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

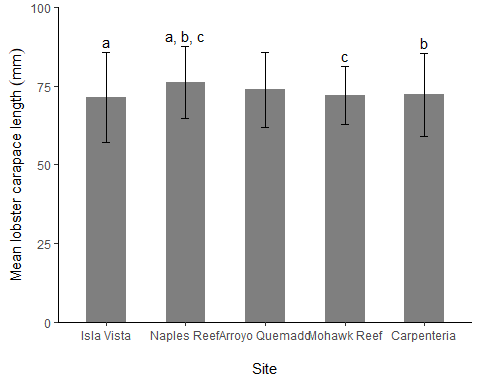
lobster\_tukey <- TukeyHSD(lobster\_anova)  
lobster\_tukey

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = SIZE ~ SITE, data = lsa\_filt2017)  
##   
## $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

# There is a significant difference between NAPL-CARP, NAPL-IVEE at alpha=.05  
  
# anova with unequal variances  
lsa\_aov\_size\_site <- oneway(lsa\_filt2017$SITE, y = lsa\_filt2017$SIZE, posthoc = 'games-howell', corrections = F, levene = T)  
lsa\_aov\_size\_site

## ### Oneway Anova for y=SIZE and x=SITE (groups: AQUE, CARP, IVEE, MOHK, NAPL)  
##   
## Omega squared: 95% CI = [0; .02], point estimate = .01  
## Eta Squared: 95% CI = [0; .01], point estimate = .01  
##   
## SS Df MS F p  
## Between groups (error + effect) 2354.51 4 588.63 3.42 .009  
## Within groups (error only) 285871.12 1663 171.9   
##   
##   
## ### Levene's test for homogeneity of variance:  
##   
## F[4, 1663] = 8.77, p < .001.  
##   
## ### Post hoc test: games-howell  
##   
## diff ci.lo ci.hi t df p  
## CARP-AQUE -1.67 -5.95 2.62 1.08 82.28 .814  
## IVEE-AQUE -2.44 -6.80 1.91 1.56 88.61 .526  
## MOHK-AQUE -1.90 -6.37 2.58 1.18 97.81 .765  
## NAPL-AQUE 2.34 -2.66 7.34 1.29 134.26 .696  
## IVEE-CARP -0.78 -2.87 1.31 1.02 1242.45 .848  
## MOHK-CARP -0.23 -2.57 2.11 0.27 379.43 .999  
## NAPL-CARP 4.00 0.73 7.27 3.38 162.35 .008  
## MOHK-IVEE 0.55 -1.94 3.03 0.60 447.17 .974  
## NAPL-IVEE 4.78 1.41 8.15 3.91 182.49 .001  
## NAPL-MOHK 4.23 0.71 7.76 3.30 201.12 .010

# y loc for annotations  
yloc\_i <- lsa\_site\_means %>%  
 filter(SITE == "IVEE") %>%  
 pull(yloc)  
  
yloc\_n <- lsa\_site\_means %>%  
 filter(SITE == "NAPL") %>%  
 pull(yloc)  
  
yloc\_c <- lsa\_site\_means %>%  
 filter(SITE == "CARP") %>%  
 pull(yloc)  
  
yloc\_m <- lsa\_site\_means %>%  
 filter(SITE == "MOHK") %>%  
 pull(yloc)  
  
  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_plot\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
lsa\_aov\_colplt <- ggplot(lsa\_site\_means, aes(x = SITE, y = mean\_size)) +  
 geom\_col(fill = "gray50", width = 0.5) +  
 theme\_classic() +  
 geom\_errorbar(aes(ymax =mean\_size + sd\_size, ymin = mean\_size - sd\_size), width = 0.1) +  
 scale\_y\_continuous(expand = c(0,0), limits = c(0,100)) +  
 labs(y=expression(Mean~lobster~carapace~length~(mm))) +  
 scale\_x\_discrete(labels = c("Isla Vista","Naples Reef","Arroyo Quemado","Mohawk Reef", "Carpenteria")) +  
 annotate("text", x = c(1), y = yloc\_i, label = "a") +  
 annotate("text", x = c(2), y = yloc\_n, label = "a, b, c") +  
 annotate("text", x = c(5), y = yloc\_c, label = "b") +  
 annotate("text", x = c(4), y = yloc\_m, label = "c") +  
 xlab("\nSite")  
lsa\_aov\_colplt



ggsave(filename = "colplotlobstersize.png", plot = lsa\_aov\_colplt,  
 scale = 1, width = 6, height = 4.5, units = "in",  
 dpi = 300)  
  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_table\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
lsa\_aov\_tabledf <- lsa\_site\_means %>%  
 ungroup() %>%  
 mutate(sitel = case\_when(  
 SITE == "IVEE" ~ "Isla Vista",  
 SITE == "NAPL" ~ "Naples Reef",  
 SITE == "AQUE" ~ "Arroyo Quemado",  
 SITE == "CARP" ~ "Carpinteria",  
 SITE == "MOHK" ~ "Mohawk Reef")  
 ) %>%  
 select(sitel, mean\_size, sd\_size, n)  
  
lsa\_aov\_table <- kable(lsa\_aov\_tabledf,   
 format = "markdown",  
 caption = "Table 2. Mean and standard deviation of lobster carapace length at five California sites in 2017. Source: Santa Barbara Coastal Long Term Ecological Research Project.",   
 col.names=c('Site',  
 'Mean carapace length (mm)',  
 'Standard deviation of carapace length (mm)',  
 'n'),  
 digits=2)  
lsa\_aov\_table

|  |  |  |  |
| --- | --- | --- | --- |
| Site | Mean carapace length (mm) | Standard deviation of carapace length (mm) | n |
| Arroyo Quemado | 73.90 | 11.89 | 67 |
| Carpinteria | 72.23 | 13.21 | 705 |
| Mohawk Reef | 72.00 | 9.28 | 178 |
| Isla Vista | 71.45 | 14.32 | 606 |
| Naples Reef | 76.23 | 11.39 | 112 |

**Figure ???. Lobster sizes along the Santa Barbara Coast in 2017.** Mean lobster carapace lengths (mm) recorded at five locations: Arroyo Quemado Reef (n = 67), Carpinteria Reef (n = 705), Mohawk Reef (n = 178), Isla Vista Reef (n = 606), and Naples Reef (n = 112). Error bars indicate +/- 1 standard deviation. A one-way ANOVA with post-hoc Tukey’s test revealed significant differences between mean length of lobsters in Naples Reef and Carpenteria and between Naples Reef and Isla Vista (one-way ANOVA, F[4, 1663] = 8.77, p < .001, with post-hoc Tukey’s).

Table 2. Mean and standard deviation of lobster carapace length at five California sites in 2017. Source: Santa Barbara Coastal Long Term Ecological Research Project.

The sample mean lobster size () in the *IVEE* were *NOT significantly* different between 2012 (66.1 ± 12.1 mm) and 2017 (71.5 ± 14.3 cm) according to a two-sample Student’s t-test (t(630) = -1.885, p = 0.060, = 0.05).

The sample mean lobster size () in the *NAPL* were *NOT significantly* different between 2012 (73 ± 11.7 mm) and 2017 (76.2 ± 11.4 cm) according to a two-sample Student’s t-test (t(116) = -0.676, p = 0.500, = 0.05).

The sample mean lobster size () in the *MOHK* were *significantly* different between 2012 (77.3 ± 10.6 mm) and 2017 (72 ± 9.28 cm) according to a two-sample Student’s t-test (t(259) = 4.069, p < 0.001, = 0.05). Additionally, the effect size is moderate (Cohen’s d = 0.54).

The sample mean lobster size () in the *CARP* were *NOT significantly* different between 2012 (74.4 ± 14.6 mm) and 2017 (72.2 ± 13.2 cm) according to a two-sample Student’s t-test (t(781) = 1.336, p = 0.182, = 0.05).

The sample mean lobster size () in the *AQUE* were *NOT significantly* different between 2012 (71 ± 10.2 mm) and 2017 (73.9 ± 11.9 cm) according to a two-sample Student’s t-test (t(103) = -1.262, p = 0.201, = 0.05).

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_part 4\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_#  
#Proportions of “legal” lobsters at the 5 sites in 2017  
#The legal minimum carapace size for lobster is 82.6 mm  
legal\_size <- lsa\_filt2017 %>%  
 mutate(  
 SIZE = case\_when(  
 SIZE <= 82.6 ~ "not legal",   
 SIZE > 82.6 ~ "legal"  
 )  
 ) %>%  
 count(SITE, SIZE) %>%  
 spread(SIZE, n) %>%   
 select(-SITE)  
   
rownames(legal\_size)<- c('Arroyo Quemado', 'Carpenteria', 'Isla Vista', 'Mohawk Reef', 'Naples Reef')

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

prop <- prop.table(as.matrix(legal\_size), 1)  
  
prop\_rounded <- round (prop, 2) #Round proportions to 2 digits  
prop\_rounded

## legal not legal  
## Arroyo Quemado 0.24 0.76  
## Carpenteria 0.25 0.75  
## Isla Vista 0.21 0.79  
## Mohawk Reef 0.13 0.87  
## Naples Reef 0.33 0.67

prop\_table <- kable(prop\_rounded,   
 format = "markdown",   
 caption = "Table 3. Proportion of lobsters that are above the legal minimum carapace size for lobster (82.6 mm) at five California sites in 2017. Source: Santa Barbara Coastal Long Term Ecological Research Project.",   
 col.names = c('Proportion Above Minimum Size', 'Proportion Below Min Size'))  
  
prop\_table

|  |  |  |
| --- | --- | --- |
|  | Proportion Above Minimum Size | Proportion Below Min Size |
| Arroyo Quemado | 0.24 | 0.76 |
| Carpenteria | 0.25 | 0.75 |
| Isla Vista | 0.21 | 0.79 |
| Mohawk Reef | 0.13 | 0.87 |
| Naples Reef | 0.33 | 0.67 |

#Run the chi-square test.  
legal\_chi <- chisq.test(prop\_rounded)

## Warning in chisq.test(prop\_rounded): Chi-squared approximation may be  
## incorrect

legal\_chi

##   
## Pearson's Chi-squared test  
##   
## data: prop\_rounded  
## X-squared = 0.11719, df = 4, p-value = 0.9983

A greater proportion of lobsters are below the legal minimum carapace size than above for all locations. The proportion of legal sized lobsters is smallest at Mohawk Reef as there is rougly 7 times more lobsters as this site that do not meet the legal minimum size. Naples Reef has the largest proportion of legal sized lobsters, roughly half that of the proportion of lobsters below the minimum size.

**Research Question**: Is there an association between site and proportion of “legal” lobsters?

#### Hypothesis Testing & Chi Square

Null Hypothesis: Location is independent of proportions of lobsters that are above the legal minimum. Alternative Hypothesis: Proportions of lobsters that are above the minimum size requirement are significantly different between sites.

**Conclusion**: Proportions of lobsters that are above the minimum size requirement do not differ significantly by site (2(4) = 0.117, p-value = 0.998, = 0.05).

#   
# legal\_chi$observed  
# round(legal\_chi$expected,0)  
#   
# legal\_chi$stdres #if standardized residuals are >2 this might be driving the significant finding  
#   
# legal\_expand <- data.frame(expand.grid(rownames(prop), colnames(prop)), value = c(prop)) # moves from a contigency table format to an expanded table format where we can expand either row or column and then by what value  
#   
# colnames(legal\_expand) <- c("Site","Legality","Proportion")  
# View(legal\_expand)  
#   
# stacked\_legal <- ggplot(legal\_expand, aes(x = Site, y = Proportion)) +  
# geom\_col(aes(fill = Legality), width = 0.5) +  
# theme\_classic() +  
# coord\_flip()  
#   
# stacked\_legal