Dataset & Analysis: Rocky Intertidal Experiment

Katie Wang

3/4/22

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(vegan)
```

```
## Warning: package 'vegan' was built under R version 4.1.2
```

```
## Loading required package: permute
```

```
## Warning: package 'permute' was built under R version 4.1.2
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-7
```

```r
library(infer)
theme_set(theme_classic())
```

```r
rocky_intertidal_unedited <- read_csv("RockyIntertidalData.csv")
```

```
## Rows: 32 Columns: 14
```

```
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr  (2): Zone, Student
## dbl (12): Index, X coordinate, Y coordinate, CA mussel, Buckshot barnacle, G...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
rocky_intertidal <- rocky_intertidal_unedited %>%
  filter(Index != 10 & Index != 11 & Index != 12) # tried to figure out how to
# do this filtering with %if% but it didn't work

# formatting it like dune and dune.env:
rocky <- rocky_intertidal %>%
  select('CA mussel',   'Buckshot barnacle',    'Goose barnacle',   'Thatched Barnacle',
```

```
        'Aggregate anemone', 'Solitary anemone',  'Lottia limpet',   'Periwinkle',   'Chitin')
rocky.env <- rocky_intertidal %>%
  select('Index',   'X coordinate', 'Y coordinate', 'Zone', 'Student')
```

Hypotheses

1. High-intertidal communities will have lower diversity in sessile animal species than mid-intertidal communities do, because high-intertidal zones are submerged for less of the day, therefore demanding greater desiccation resistance. The high-intertidal zone will have lower species richness because few marine species are adapted to desiccation, and longer periods of desiccation are less tolerable. The high-intertidal zone will also have lower evenness because there will be many individuals from common desiccation-resistance species and few from rare non-resistant species.

2. High-intertidal communities will have lower abundance of sessile animal species than mid-intertidal communities do, because water is more scarce in high-intertidal communities. This causes greater competition for water in high-intertidal communities, which means that fewer individuals can be supported.

Data description

We sampled a total of 32 quadrats (20 cm x 20 cm) at the Laguna Beach rocky intertidal zone: 16 in the mid-intertidal zone and 16 in the high-intertidal zone. For each sample, we randomized the coordinates of the sampling site, as well as one of the 16 squares in the quadrat. We recorded the index of the sample (numbered 1-32), the zone type (mid- or high-intertidal), which student collected the sample (which may be useful for explaining patterns in the data), and the number of each species encountered within the quadrat. For mussels and buckshot barnacles, which were very abundant, we estimated abundance within the quadrat by counting individuals within a single randomized square (mentioned earlier) and multiplying these numbers by 16. For the other species, we counted the exact number of individuals in the quadrat. Samples 10-12 are excluded from analysis due to severe overestimation of the number of limpets (a rare species).

```
head(rocky)
```

```
## # A tibble: 6 x 9
##    `CA mussel` `Buckshot barnacle` `Goose barnacle` `Thatched Barnacle`
##          <dbl>               <dbl>            <dbl>               <dbl>
## 1          224                1520                0                   0
## 2          144                 192                0                   0
## 3          160                2256                1                   0
## 4           80                1040                0                   0
## 5          288                1552               11                   0
## 6          352                1648                0                   0
## # ... with 5 more variables: Aggregate anemone <dbl>, Solitary anemone <dbl>,
## #   Lottia limpet <dbl>, Periwinkle <dbl>, Chitin <dbl>
```

```
head(rocky.env)
```

```
## # A tibble: 6 x 5
##    Index `X coordinate` `Y coordinate` Zone  Student
##    <dbl>          <dbl>          <dbl> <chr> <chr>
## 1      1            9.1            2.2 Mid   Ihlara
## 2      2           15.2            1.9 Mid   Ihlara
## 3      3            5.1            2.8 Mid   Ihlara
## 4      4           13.1            3.6 Mid   Ihlara
## 5      5           17.9            4   Mid   Kathleen
## 6      6           12.1            2.3 Mid   Kathleen
```
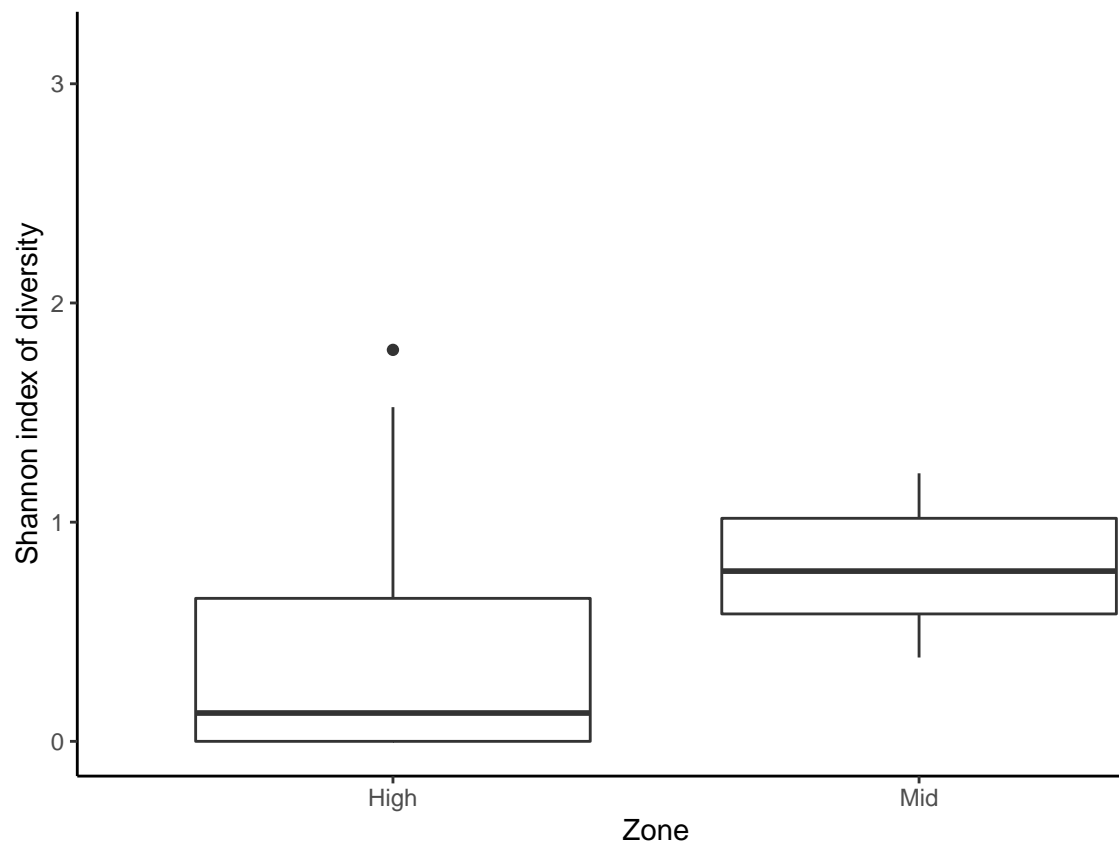
Data analysis

Hypothesis 1

```
rocky.diversity <- rocky.env %>%
  mutate(Shannon=diversity(rocky,index="shannon",base=2))

rocky.diversity %>%
  ggplot() +
    geom_boxplot(mapping = aes(x = Zone, y = Shannon)) +
    ylab("Shannon index of diversity") +
    ylim(0, log2(ncol(rocky))) #log base 2 of the number of species in total is
```


Species Diversity

```
# the max of Shannon index

# check for normality here
qqnorm(rocky.diversity$Shannon)
qqline(rocky.diversity$Shannon)
```

Normal Q–Q Plot
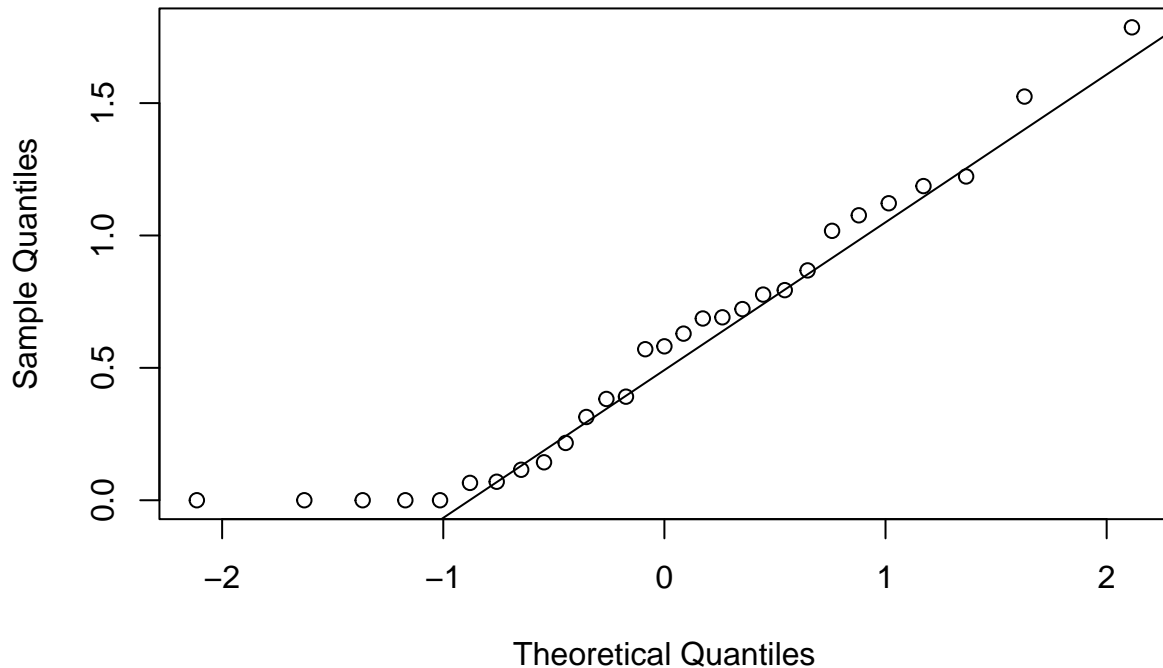

```
# does not look normally distributed, so I'll use Kruskal-Wallis test
kruskal.test(Shannon ~ Zone, data = rocky.diversity)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Shannon by Zone
## Kruskal-Wallis chi-squared = 6.5013, df = 1, p-value = 0.01078
```

I chose to analyze Shannon's index of diversity, rather than Simpson's index of diversity. Shannon's is more suited for detecting differences in rare species between groups, while Simpson's tells us more about differences in common species between groups. I'd like to focus on rare species because I'm more interested in species richness than in relative abundance of species within each zone.

The Kruskal-Wallis test for difference in Shannon diversity between zones yields p=0.01078 (a=0.05). This is a significant difference. The box plot shows that the mid-intertidal zone has higher diversity than the high-intertidal zone, so my hypothesis is supported.

```
rockySR.byzone <- specpool(rocky, pool=rocky.env$Zone) %>%
  rownames_to_column(var="Zone")
rockySR.byzone
```

Species Richness

```
##   Zone Species       chao  chao.se     jack1 jack1.se      jack2       boot
## 1 High       6    6.46875 1.251952   6.93750 0.937500   6.995833   6.484179
## 2  Mid       9   13.15385 6.664398  11.76923 2.098182  13.532051  10.184010
##     boot.se  n
```

```
## 1 0.7148239 16
## 2 1.0933099 13
```

```
rockySR.byzone %>%
  ggplot(aes(x=Zone,y=chao)) +
  geom_point() +
  geom_errorbar(aes(ymax=chao + chao.se, ymin=chao-chao.se), width=0.15) +
  ylim(0, 21) +
  xlab("Intertidal zone") +
  ylab("Chao1 species diversity estimation")
```
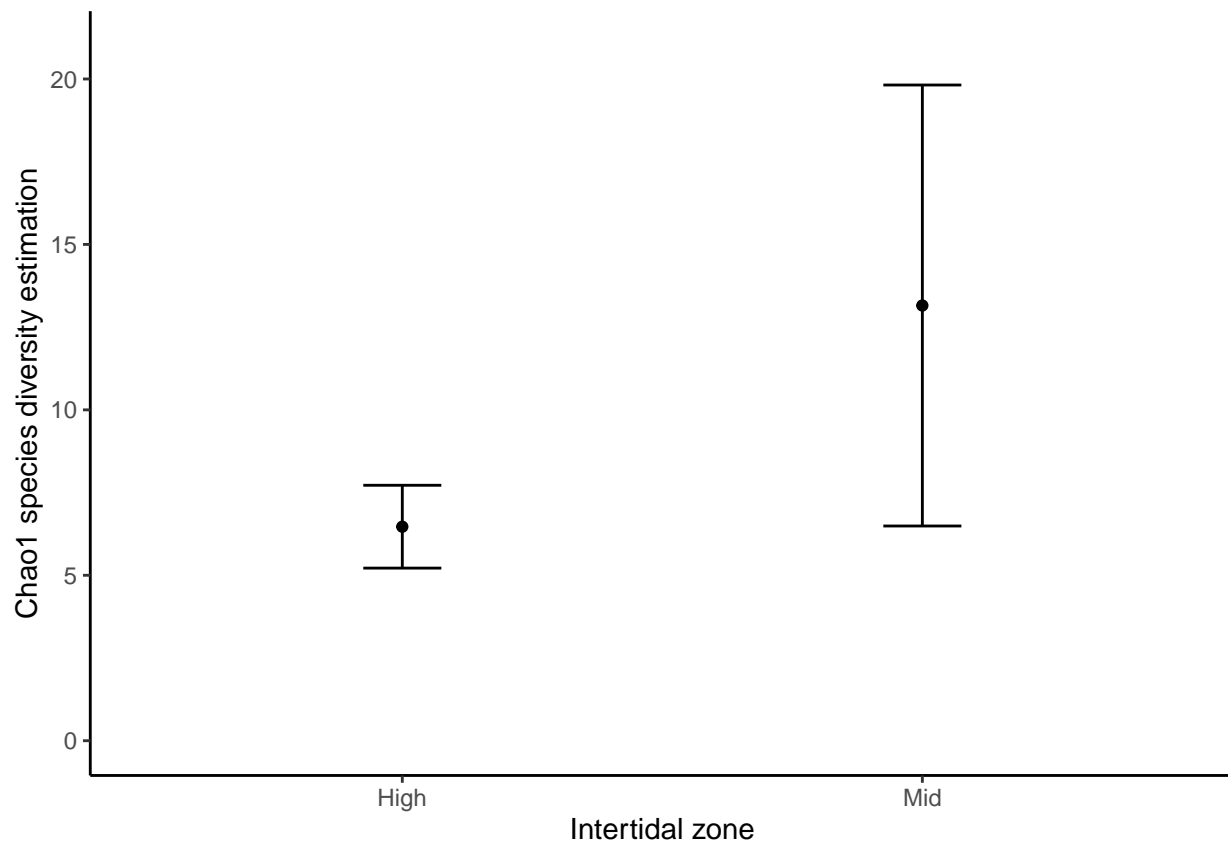


```
rich_v <- (rocky$'CA mussel'>0) + (rocky$'Buckshot barnacle'>0) +
  (rocky$'Goose barnacle'>0) + (rocky$'Thatched Barnacle'>0) +
  (rocky$'Aggregate anemone'>0) + (rocky$'Solitary anemone'>0) +
  (rocky$'Lottia limpet'>0) + (rocky$'Periwinkle'>0) + (rocky$'Chitin'>0)
# exploit addition of Boolean operators

rocky_intertidal <- rocky_intertidal %>%
  mutate(spec_count = rich_v)

rocky_intertidal %>%
  group_by(Zone) %>%
  summarize(mean_richness = mean(spec_count), median_richness = median(spec_count),
            min_richness = min(spec_count), max_richness = max(spec_count), num_samples = n())
```

```
## # A tibble: 2 x 6
##    Zone  mean_richness median_richness min_richness max_richness num_samples
```

```
##    <chr>          <dbl>          <dbl>          <int>          <int>          <int>
## 1 High            2.25              2              0              6             16
## 2 Mid             3.62              3              2              7             13
```

There is no significant difference in species richness between the two zones. The error bars on the plot of Chao1 species diversity overlap each other (though it should be mentioned that while the error bars for "Mid" overlap the mean of "High", the error bars for "High" do not overlap the mean of "Mid"). I wasn't sure how to find 95% confidence intervals for Chao1.

I've also included summary tables for the Chao1 statistics and for the species counts, grouped by zone.

```r
# adding a new column for evenness
rocky_intertidal <- rocky_intertidal %>%
  mutate(evenness = rocky.diversity$Shannon / log2(ncol(rocky)))

rocky_intertidal %>%
  ggplot() +
    geom_boxplot(mapping = aes(x = Zone, y = evenness)) +
    ylab("Evenness")
```
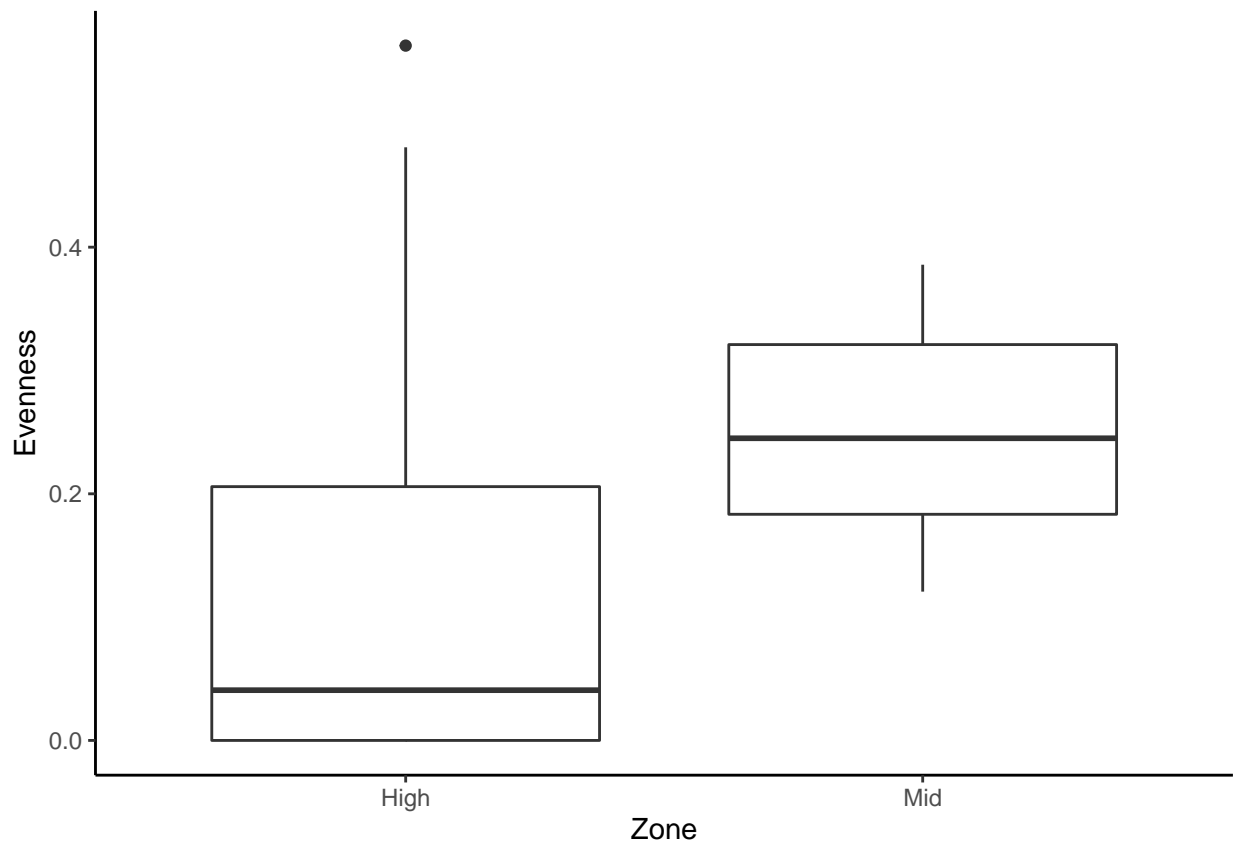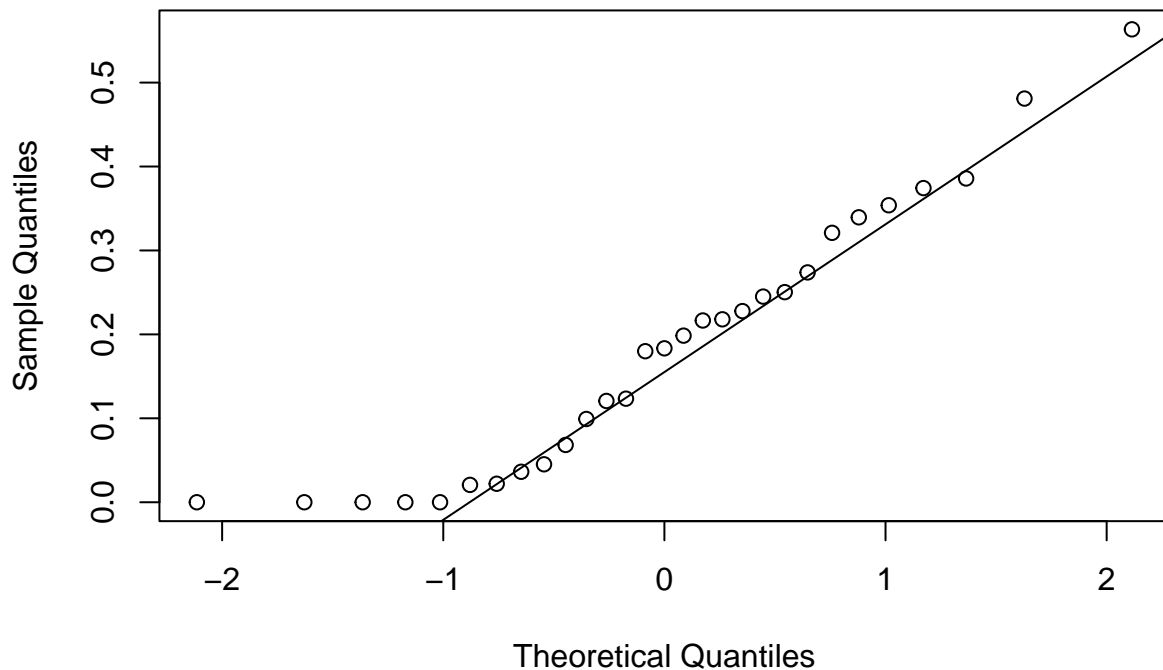

Evenness

```r
# check for normality here
qqnorm(rocky_intertidal$evenness)
qqline(rocky_intertidal$evenness)
```

Normal Q–Q Plot


```
# does not look normally distributed, so I'll use Kruskal-Wallis test
kruskal.test(evenness ~ Zone, data = rocky_intertidal)
```
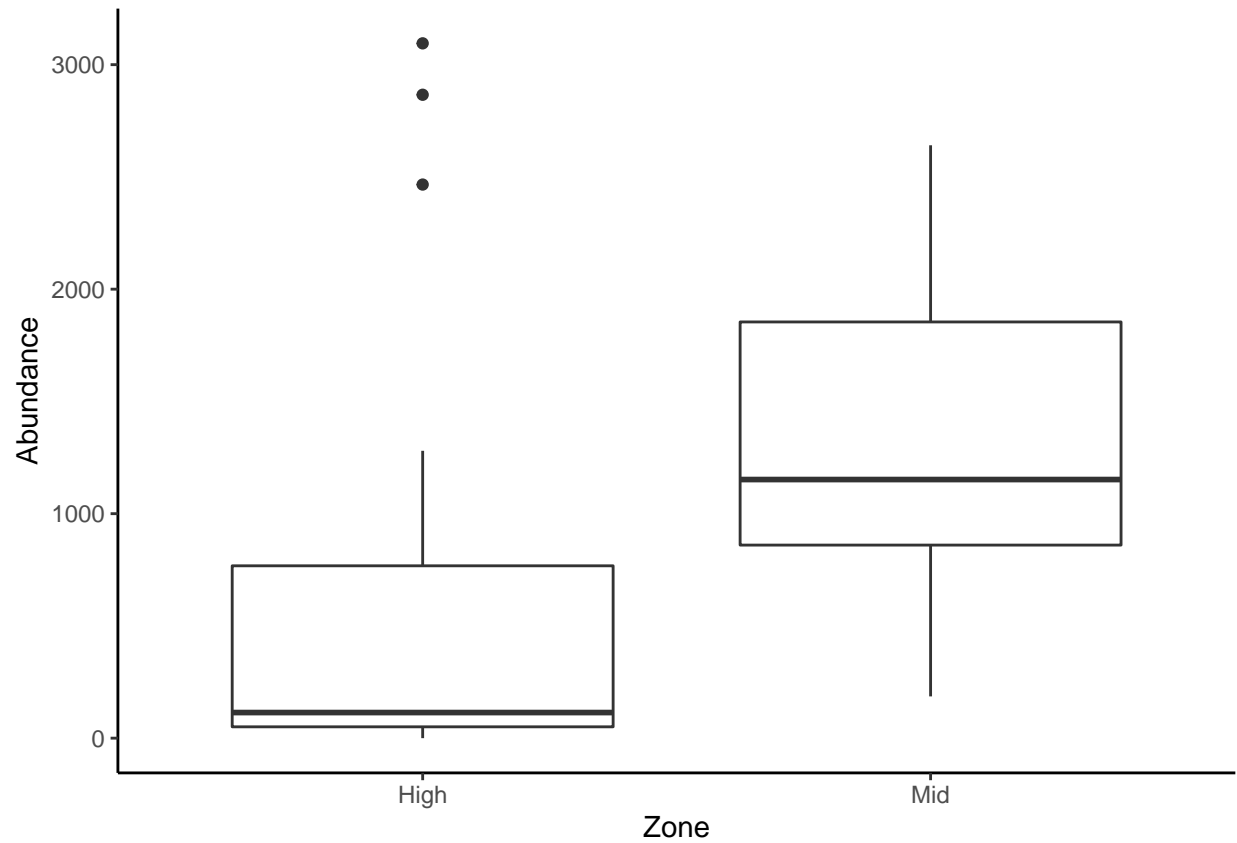
```
##
##  Kruskal-Wallis rank sum test
##
## data:  evenness by Zone
## Kruskal-Wallis chi-squared = 6.5013, df = 1, p-value = 0.01078
```

As expected, the Kruskal-Wallis test on difference in evenness between zones supports my hypothesis that there is a significant difference in evenness (p=0.01078, a=0.05), and that evenness is higher in the mid-intertidal zone (as seen in the box plot). This is because evenness is obtained by scaling diversity- you divide it by the log base 2 of the max number of species.

Hypothesis 2

```
# adding an abundance column
rocky_intertidal <- rocky_intertidal %>%
  mutate(total_abundances = rocky$'CA mussel' + rocky$'Buckshot barnacle' +
         rocky$'Goose barnacle' + rocky$'Thatched Barnacle' + rocky$'Aggregate anemone' +
         rocky$'Solitary anemone' + rocky$'Lottia limpet' + rocky$'Periwinkle' + rocky$'Chitin')

rocky_intertidal %>%
  ggplot() +
    geom_boxplot(mapping = aes(x = Zone, y = total_abundances)) +
    ylab("Abundance")
```
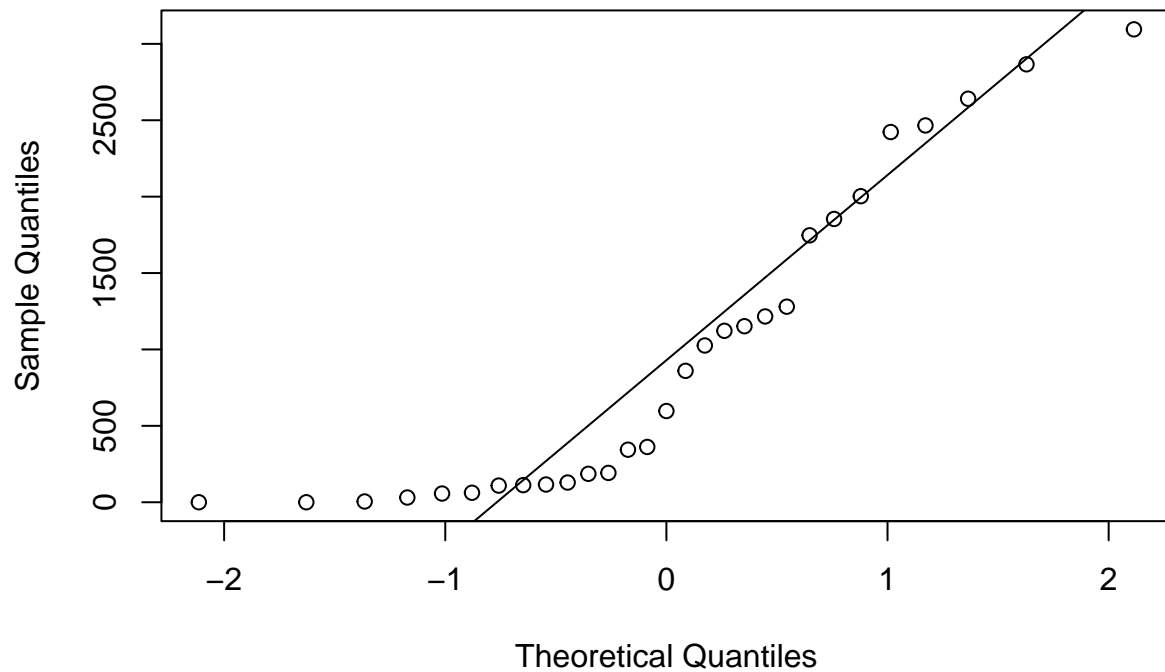
Abundance

```
# check for normality here
qqnorm(rocky_intertidal$total_abundances)
qqline(rocky_intertidal$total_abundances)
```

Normal Q–Q Plot


```
# does not look normally distributed, so I'll use Kruskal-Wallis test
kruskal.test(total_abundances ~ Zone, data = rocky_intertidal)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  total_abundances by Zone
## Kruskal-Wallis chi-squared = 5.6091, df = 1, p-value = 0.01787
```

The Kruskal-Wallis test supports the hypothesis that the abundances are different between zones (p=0.01787, a=0.05), with abundances higher in the mid-intertidal zone (as seen in the box plot).