

Individual Project Dataset & Analysis

4/17/22

Hypotheses

1. The Shannon diversity index is lower in the grassy zone because exotic grasses consume much of the soil water, competitively excluding the shrub species that are less resistant to drought.
2. Evenness in shrub species is lower in the grassy zone because the presence of grass increases selection for drought-resistant species, causing the drought-resistant species to have relatively higher abundances than the less drought-resistant species.
3. Shrub abundance is lower in the grassy zone because the grasses consume soil water, thus limiting the number of other plants (including shrubs) that can survive in the same soil.

Data description

```
shrub_data <- read_csv("BFS shrub data.csv")

## Rows: 20 Columns: 13

## -- Column specification -----
## Delimiter: ","
## chr (2): Zone, Date
## dbl (11): Index, N.coordinate, W.coordinate, CA.sage, Yerba.santa, Buckwheat...

##
## 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.

head(shrub_data)

## # A tibble: 6 x 13
##   Index N.coordinate W.coordinate Zone   Date   CA.sage Yerba.santa Buckwheat
##   <dbl>      <dbl>      <dbl> <chr> <chr>   <dbl>      <dbl>      <dbl>
## 1     1         34.1         118. East  4/1/22     0         0         0
## 2     2         34.1         118. East  4/1/22    11        21         2
## 3     3         34.1         118. West  4/1/22     1         6         3
## 4     4         34.1         118. West  4/6/22     0         3         2
## 5     5         34.1         118. West  4/6/22     0         4         0
## 6     6         34.1         118. East  4/6/22     0        17         0
## # ... with 5 more variables: Deerweed <dbl>, Showy.penstemon <dbl>,
## #   White.sage <dbl>, Bluewitch.nightshade <dbl>, Poison.oak <dbl>

# For use with diversity() from vegan:
shrub <- shrub_data %>%
  select(CA.sage, Yerba.santa, Buckwheat, Deerweed, Showy.penstemon, White.sage,
         Bluewitch.nightshade, Poison.oak)

# Visualizing species (and abundances) in each zone:
tidy_shrub <- shrub_data %>%
  select(Index, Zone, CA.sage, Yerba.santa, Buckwheat, Deerweed, Showy.penstemon,
```

```

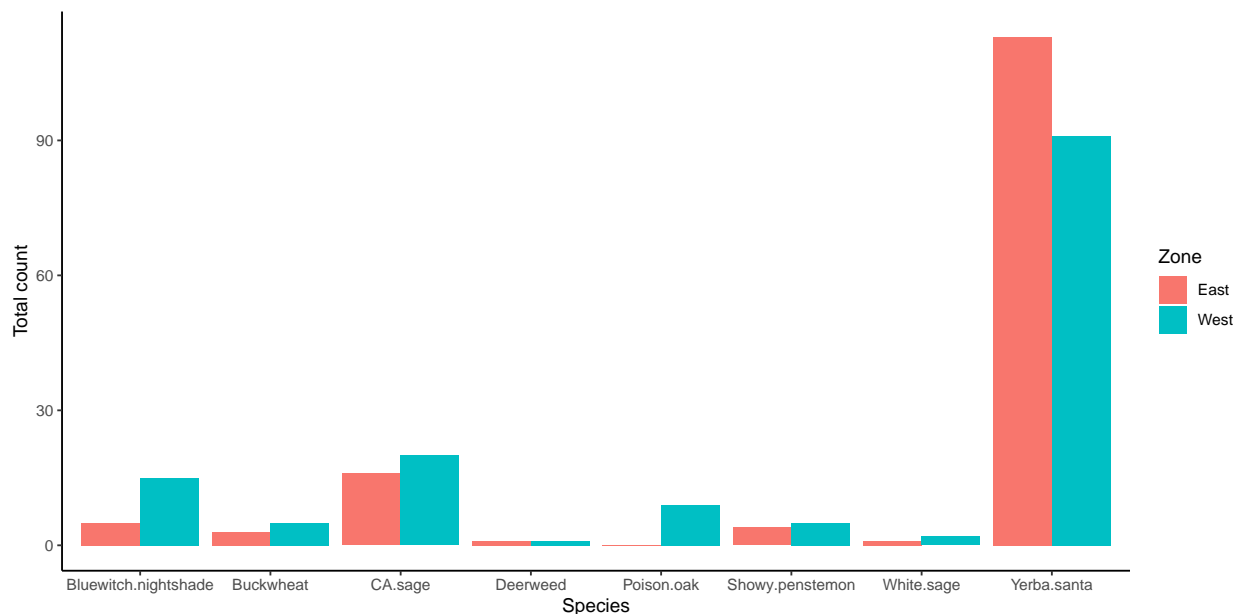
      White.sage, Bluewitch.nightshade, Poison.oak) %>%
pivot_longer(c(CA.sage, Yerba.santa, Buckwheat, Deerweed, Showy.penstemon,
      White.sage, Bluewitch.nightshade, Poison.oak),
      names_to = "Species", values_to = "Count")

summary_shrub <- tidy_shrub %>%
  group_by(Zone, Species) %>%
  summarize(total_obs = sum(Count))

## `summarise()` has grouped output by 'Zone'. You can override using the `.groups` argument.

summary_shrub %>% ggplot() +
  geom_col(aes(x=Species, y=total_obs, fill = Zone), position=position_dodge()) +
  ylab("Total count")

```



Over two weeks in April 2022, I sampled 8m x 3m belt transects in two 30 m x 30 m zones within the section of the Bernard Field Station that was burned in the 2013 fire- a zone with densely-growing grass (the east zone) and a zone with not much grass (the west zone). Ten transects were sampled per zone at randomly-generated, non-overlapping coordinates (in DD format) for a total of twenty samples. For each transect, I identified and counted all shrubs that overlapped or were contained by the transect lines. The data I recorded for each transect are as follows: the sample number (index), N and W coordinates in DD format, the zone, the date, and how many individuals of each species I counted.

Data analysis

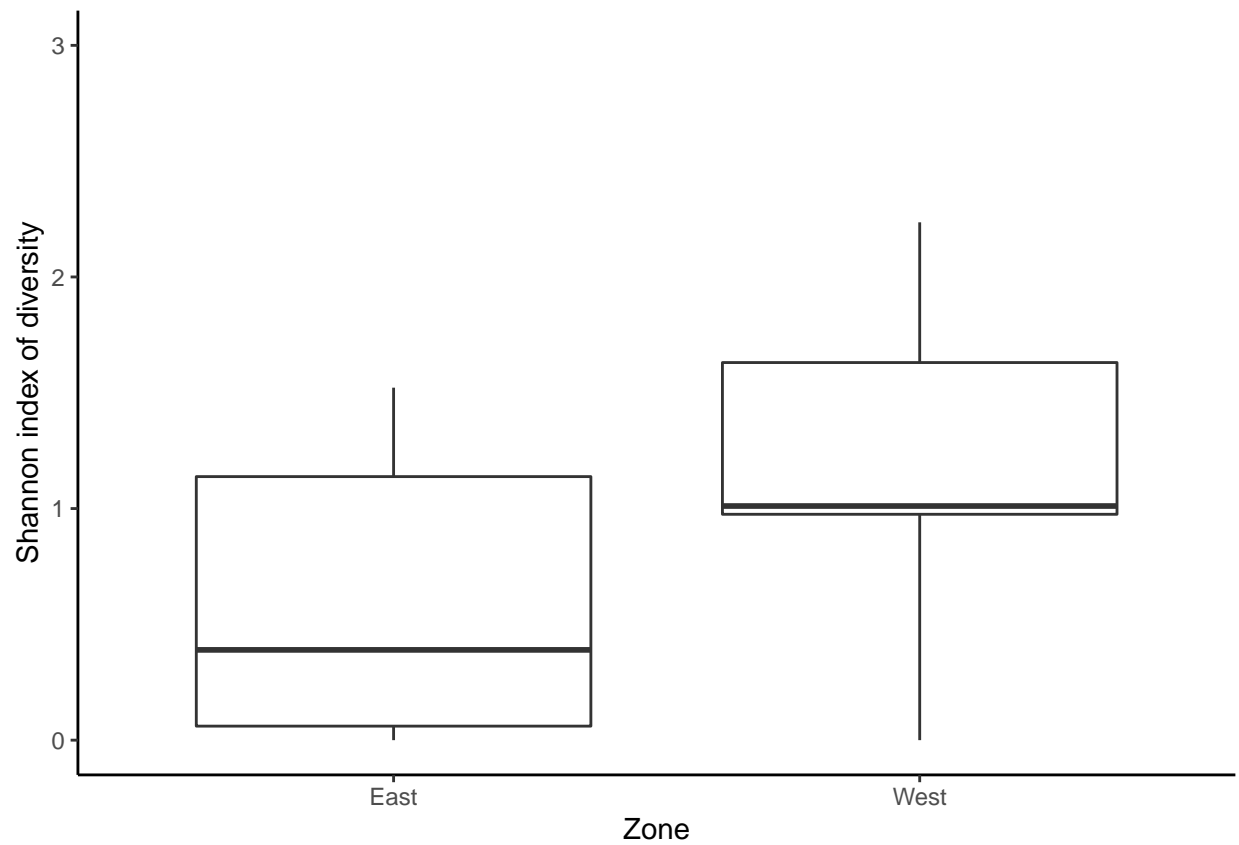
Hypothesis 1

```

shrub_data <- shrub_data %>%
  mutate(Shannon = diversity(shrub, index="shannon", base=2))

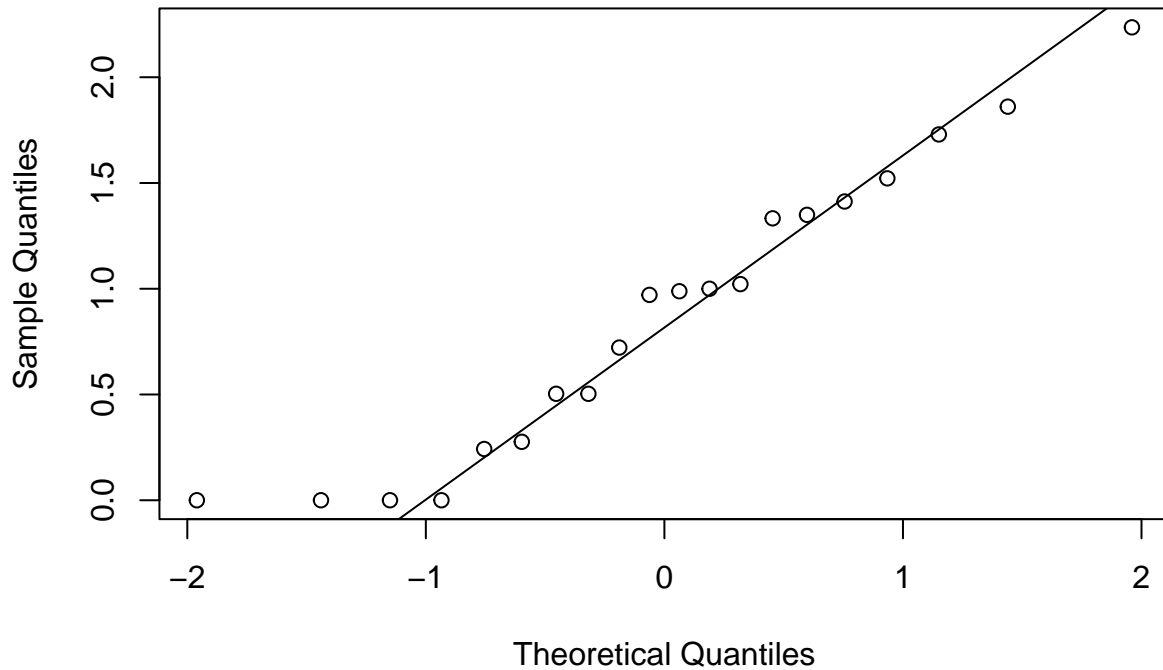
# boxplot
shrub_data %>% ggplot() +
  geom_boxplot(mapping = aes(x=Zone, y=Shannon)) +
  ylab("Shannon index of diversity") +
  ylim(0, log2(ncol(shrub)))

```



```
# normality check  
qqnorm(shrub_data$Shannon)  
qqline(shrub_data$Shannon)
```

Normal Q-Q Plot



```
# data isn't normally distributed, therefore use Kruskal-Wallis test
kruskal.test(Shannon ~ Zone, data = shrub_data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Shannon by Zone
## Kruskal-Wallis chi-squared = 3.182, df = 1, p-value = 0.07445
```

```
# summary statistics
shrub_data %>%
  group_by(Zone) %>%
  summarize(mean_Shannon=mean(Shannon),
            SE_Shannon = sd(Shannon)/sqrt(length(Shannon)))
```

```
## # A tibble: 2 x 3
##   Zone mean_Shannon SE_Shannon
##   <chr>      <dbl>      <dbl>
## 1 East      0.581      0.194
## 2 West      1.19      0.201
```

The results do not support my hypothesis. There is no significant difference in the Shannon diversity indices between the two zones (Kruskal-Wallis test, chi-squared=3.182, df=1, p=0.07445). The mean Shannon diversity indices were 0.58 for the east zone (SE=0.19) and 1.19 from the west zone (SE=0.20). Shannon diversity seems to be higher in the west zone (though not a statistically significant difference) and the standard errors for the two zones are similar.

Hypothesis 2

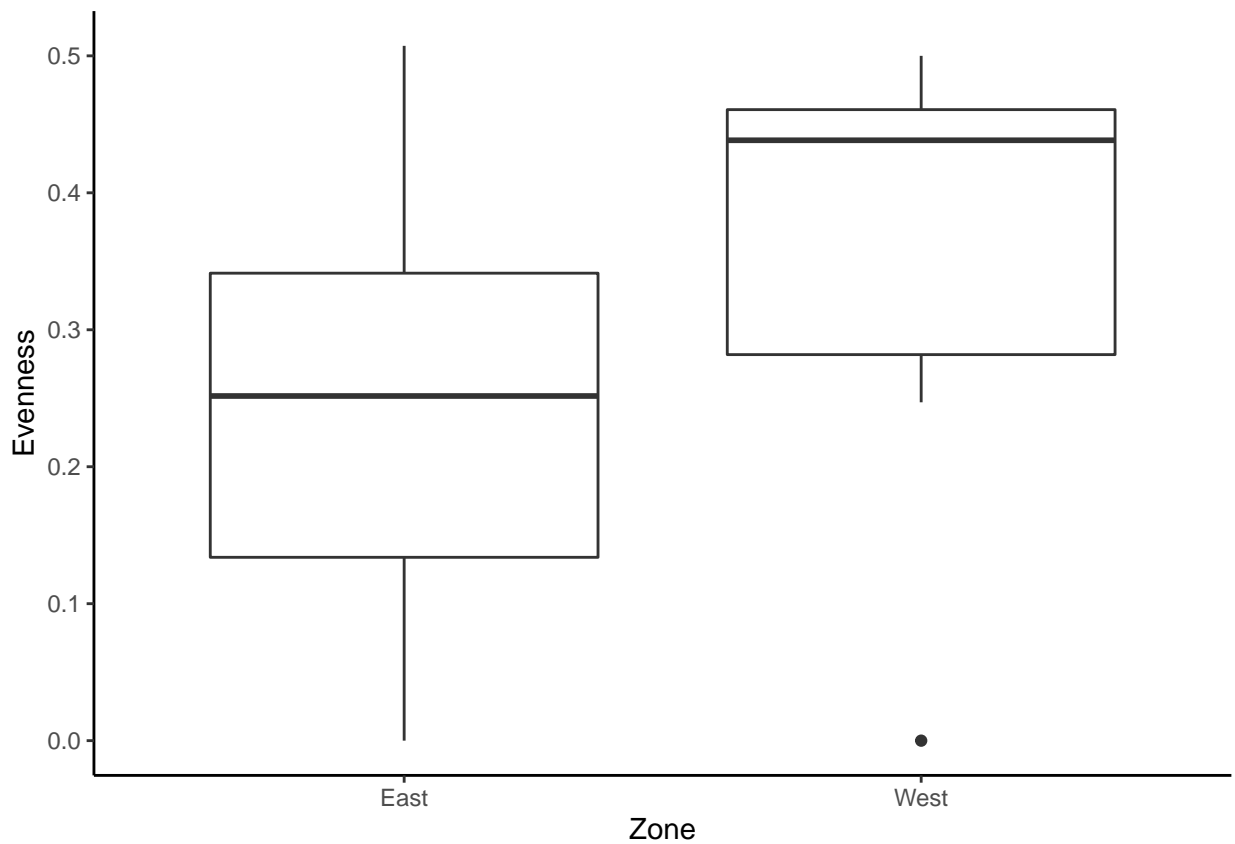
```
# To calculate evenness, first need species richness (species count) of samples
rich_v <- (shrub_data$CA.sage > 0) + (shrub_data$Yerba.santa > 0) +
  (shrub_data$Buckwheat > 0) + (shrub_data$Deerweed > 0) +
  (shrub_data$Showy.penstemon > 0) + (shrub_data$White.sage > 0) +
  (shrub_data$Bluewitch.nightshade > 0) + (shrub_data$Poison.oak > 0)

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

# Working with evenness
shrub_data <- shrub_data %>%
  mutate(evenness = Shannon/spec_count)

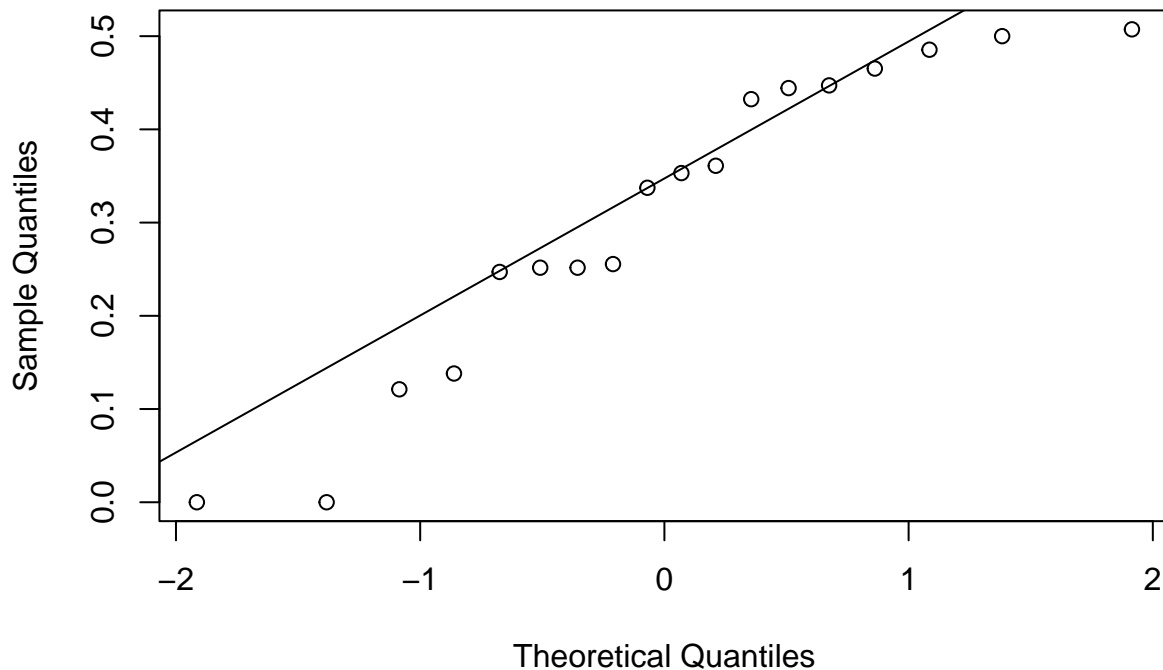
shrub_data %>% ggplot() +
  geom_boxplot(aes(x = Zone, y = evenness)) +
  ylab("Evenness")
```

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```



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

Normal Q-Q Plot



```
# doesn't seem normally distributed, therefore Kruskal-Wallis
kruskal.test(evenness ~ Zone, data = shrub_data)

##
##  Kruskal-Wallis rank sum test
##
## data:  evenness by Zone
## Kruskal-Wallis chi-squared = 2.4228, df = 1, p-value = 0.1196

# summary statistics
shrub_data %>%
  group_by(Zone) %>%
  filter(!is.na(evenness)) %>%
  summarize(mean_evenness = mean(evenness),
            SE_evenness = sd(evenness)/sqrt(length(evenness)))

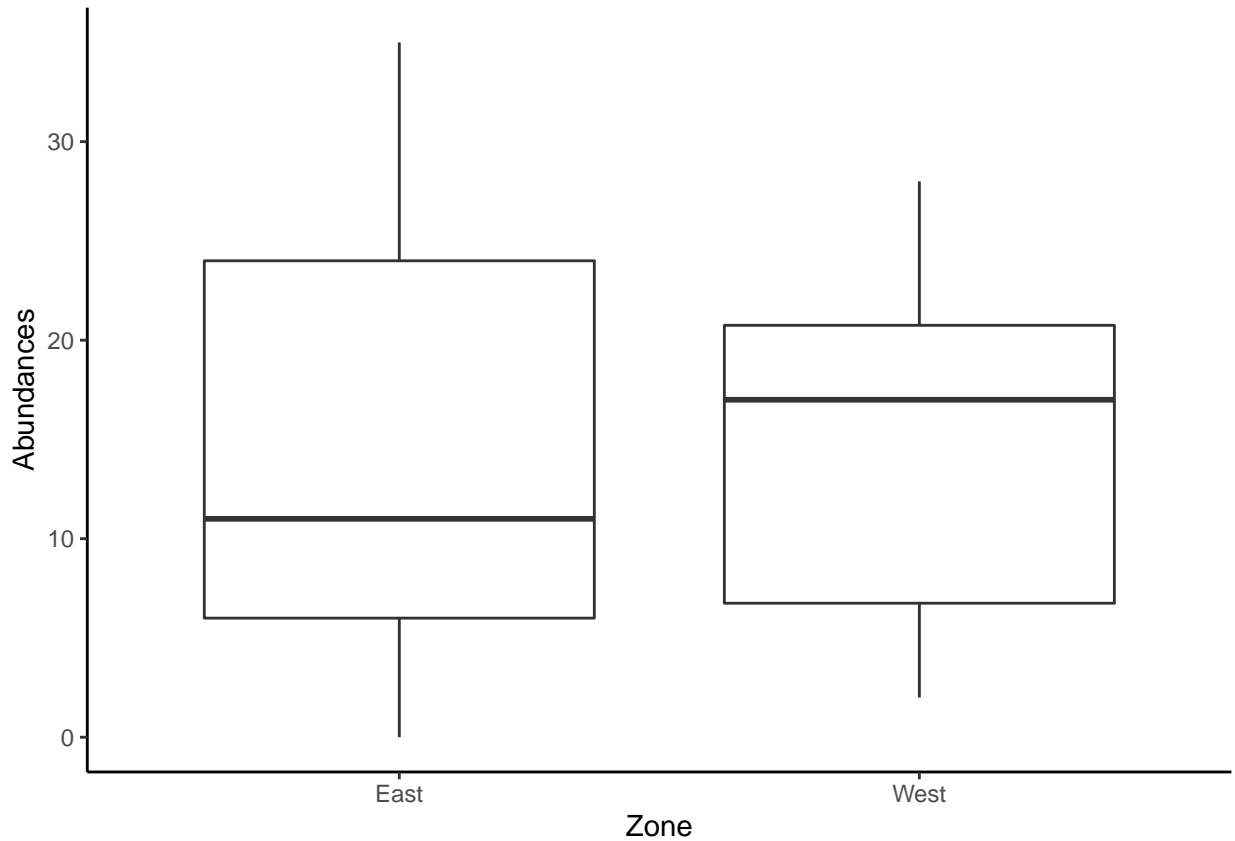
## # A tibble: 2 x 3
##   Zone mean_evenness SE_evenness
##   <chr>      <dbl>      <dbl>
## 1 East      0.245      0.0560
## 2 West      0.364      0.0494
```

The results do not support my hypothesis. There is no significant difference in evenness between the two zones (Kruskal-Wallis test, chi-squared=2.4228, df=1, p=0.1196). However, the mean evenness seems to be higher in the west zone, since the mean evenness in the west zone is 0.36 (SE = 0.049) while the mean evenness in the east zone is 0.25 (SE = 0.056). The standard errors of evenness are similar between zones.

Hypothesis 3

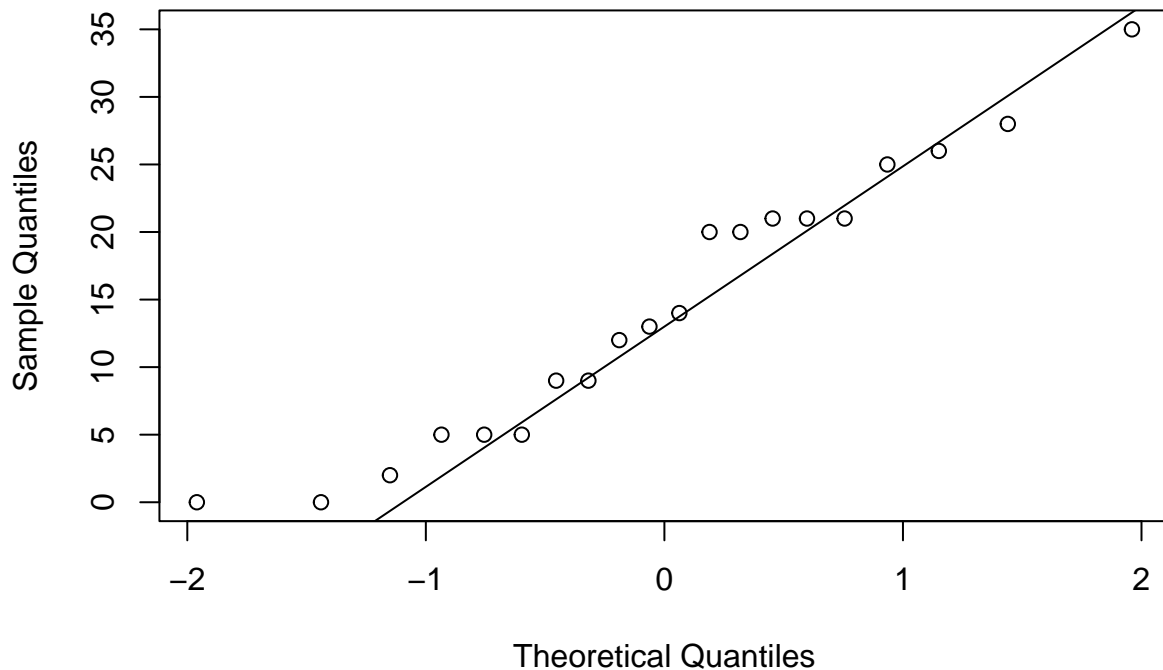
```
# adding an abundances column
shrub_data <- shrub_data %>%
  mutate(total_abundances = shrub_data$CA.sage + shrub_data$Yerba.santa +
    shrub_data$Buckwheat + shrub_data$Deerweed + shrub_data$Showy.penstemon +
    shrub_data$White.sage + shrub_data$Bluewitch.nightshade + shrub_data$Poison.oak)

shrub_data %>% ggplot() +
  geom_boxplot(aes(x=Zone, y=total_abundances)) +
  ylab("Abundances")
```



```
#normality check
qqnorm(shrub_data$total_abundances)
qqline(shrub_data$total_abundances)
```

Normal Q-Q Plot



#doesn't appear normally distributed; use Kruskal-Wallis

```
kruskal.test(total_abundances ~ Zone, data = shrub_data)
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: total_abundances by Zone
```

```
## Kruskal-Wallis chi-squared = 0.023048, df = 1, p-value = 0.8793
```

summary data

```
shrub_data %>%
```

```
  group_by(Zone) %>%
```

```
  summarize(mean_abundance = mean(total_abundances),
```

```
            SE_abundance = sd(total_abundances)/sqrt(length(total_abundances)))
```

```
## # A tibble: 2 x 3
```

```
##   Zone mean_abundance SE_abundance
```

```
##   <chr>      <dbl>      <dbl>
```

```
## 1 East      14.3        3.77
```

```
## 2 West      14.8        2.73
```

The results do not support my hypothesis. There is no significant difference in abundance between zones (Kruskal-Wallis test, chi-squared=0.023048, df=1, p=0.8793). The mean abundances for the two zones were similar: 14.8 for the west zone (SE = 2.73) and 14.3 for the east zone (SE = 3.77). The standard error for the east zone is higher than that of the west zone.