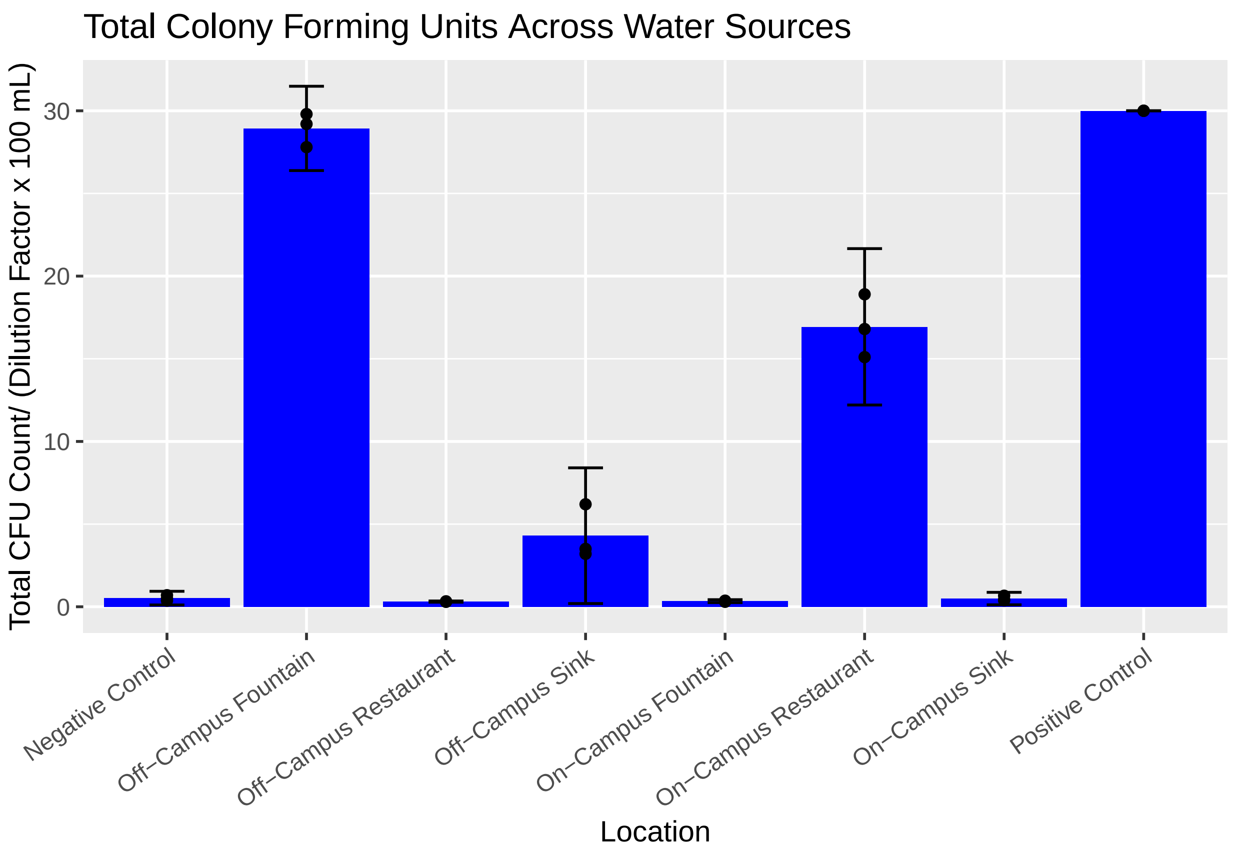
Section 3 Clusters

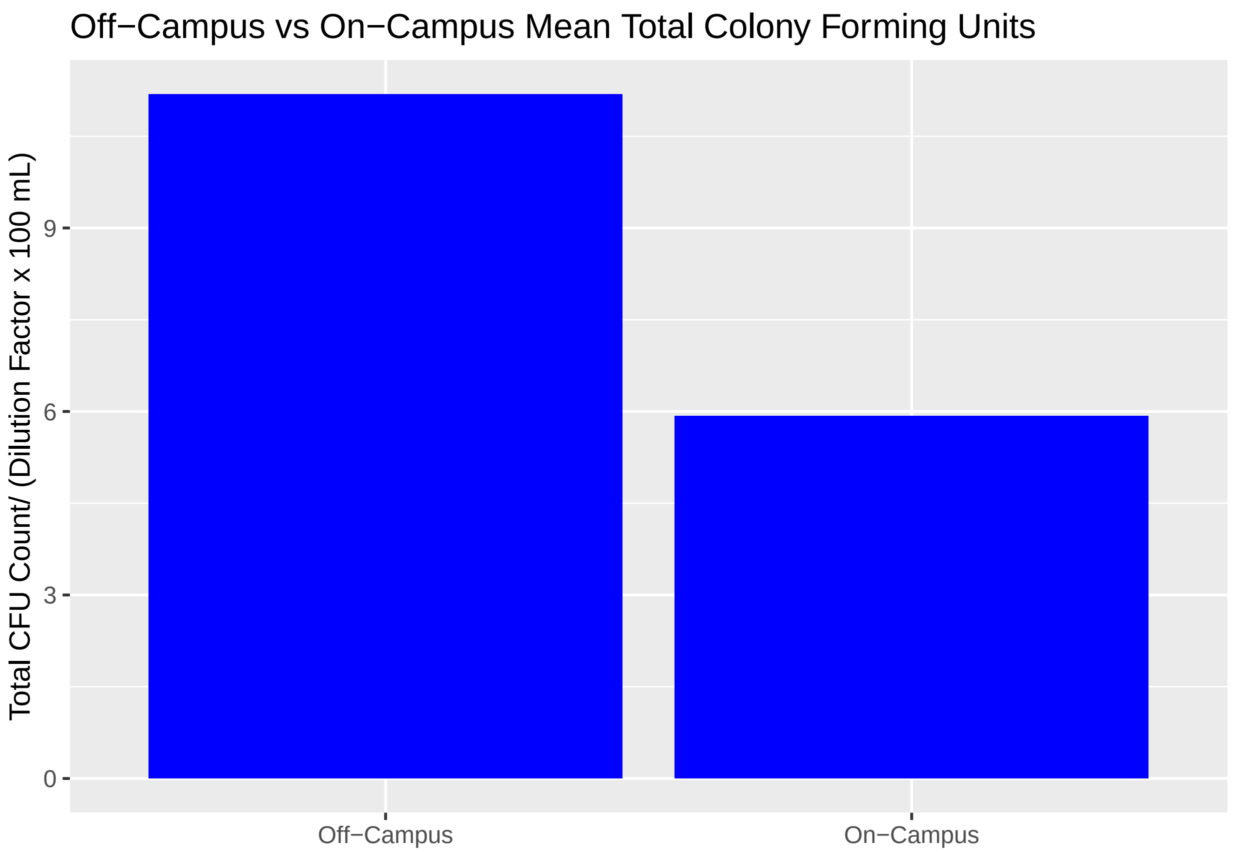
2022-11-19

```r
results <- data.frame(location = c("Negative Control", "Off-Campus Fountain",
                                   "On-Campus Fountain", "Off-Campus Sink",
                                   "On-Campus Sink", "Off-Campus Restaurant",
                                   "On-Campus Restaurant", "Positive Control"),
                      total_cfu_count_rep1 = c(0.37, 29.8, 0.35, 3.5,
                                               0.39, 0.30, 16.8, 30),
                      total_cfu_count_rep2 = c(0.50, 29.2, 0.37, 3.2,
                                               0.43, 0.33, 18.9, 30),
                      total_cfu_count_rep3 = c(0.70, 27.8, 0.30, 6.2,
                                               0.67, 0.31, 15.1, 30))

results$total_cfu_mean <- rowMeans(results[,2:4], na.rm = TRUE)
results <- results %>%
  mutate(total_cfu_sd = rowSds(as.matrix(.[c("total_cfu_count_rep1",
                                             "total_cfu_count_rep2",
                                             "total_cfu_count_rep3")]))) %>%
  mutate(total_cfu_se = total_cfu_sd/sqrt(3))

results <- results %>%
  mutate(campus = case_when(
    str_detect(location, "^On-Campus") ~ "On-Campus",
    str_detect(location, "^Off-Campus") ~ "Off-Campus",
    str_detect(location, "Control") ~ "Control")) %>%
  mutate(source = case_when(
    str_detect(location, "Fountain") ~ "Fountain",
    str_detect(location, "Sink") ~ "Sink",
    str_detect(location, "Restaurant") ~ "Restaurant",
    str_detect(location, "Control") ~ "Control"))
```

```r
t.score <- qt(0.025, df = 2, lower.tail = F)

results %>%
  ggplot() +
  geom_bar(aes(x = location, y = total_cfu_mean),
           stat = "identity", fill = "blue") +
  geom_point(aes(x = location, y = total_cfu_count_rep1)) +
  geom_point(aes(x = location, y = total_cfu_count_rep2)) +
  geom_point(aes(x = location, y = total_cfu_count_rep3)) +
  geom_errorbar(aes(x= location,
                    ymin = total_cfu_mean - (t.score * total_cfu_se),
                    ymax = total_cfu_mean + (t.score * total_cfu_se)),
                width = 0.25) +
  theme(axis.text.x = element_text(angle = 35, hjust = 1)) +
  labs(title = "Total Colony Forming Units Across Water Sources",
       x = "Location", y = "Total CFU Count/ (Dilution Factor x 100 mL)")
```



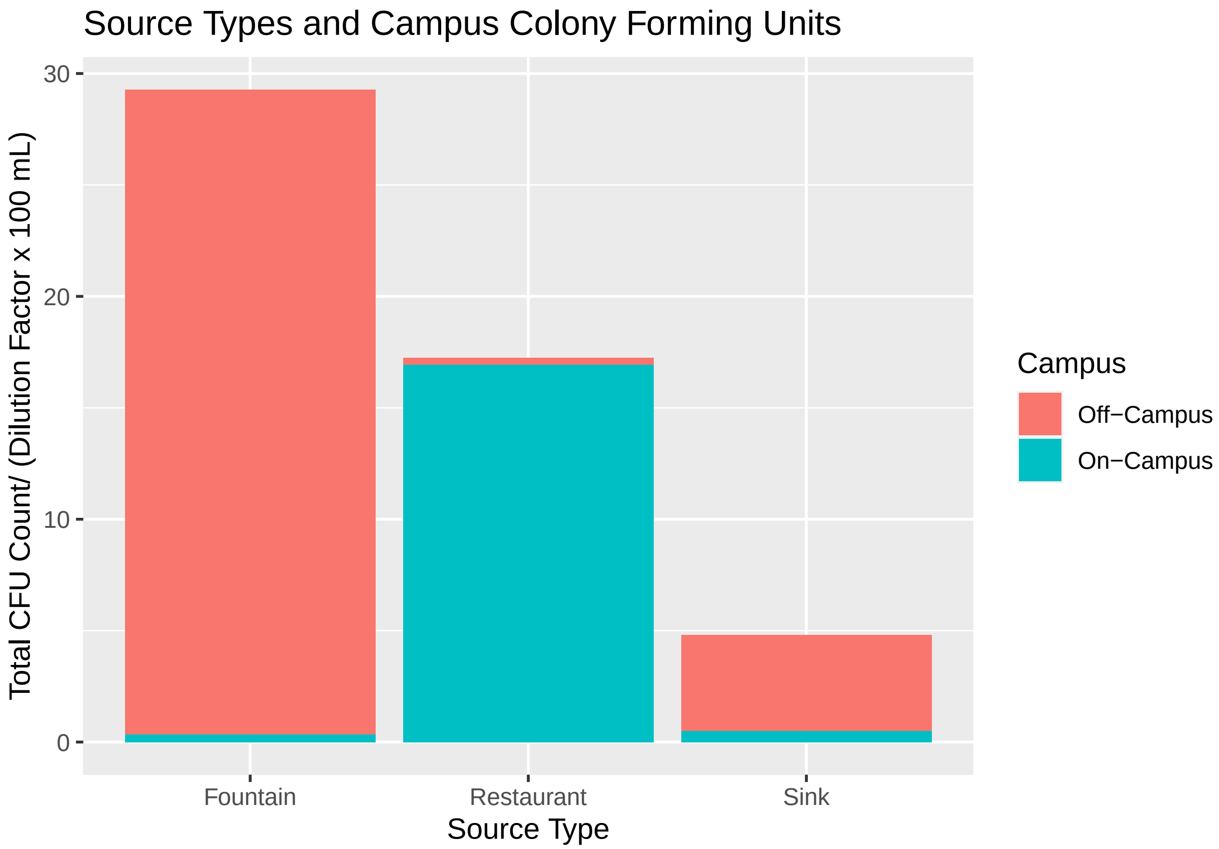
```
campus_averages <- results %>%
  filter(campus != "Control") %>%
  group_by(campus) %>%
  summarise(campus_mean_cfu = mean(total_cfu_mean))

campus_averages %>%
  ggplot(aes(x = campus, y = campus_mean_cfu, fill = source)) +
  geom_bar(position = "stack", stat="identity", fill = "blue") +
  theme(axis.title.x = element_blank()) +
  labs(title = "Off-Campus vs On-Campus Mean Total Colony Forming Units",
       y = "Total CFU Count/ (Dilution Factor x 100 mL)")
```



```
results %>%
  ungroup()
```

```
##                 location total_cfu_count_rep1 total_cfu_count_rep2
## 1       Negative Control                 0.37                 0.50
## 2   Off-Campus Fountain                 29.80                29.20
## 3    On-Campus Fountain                  0.35                 0.37
## 4        Off-Campus Sink                 3.50                 3.20
## 5         On-Campus Sink                 0.39                 0.43
## 6 Off-Campus Restaurant                  0.30                 0.33
## 7  On-Campus Restaurant                 16.80                18.90
## 8       Positive Control                30.00                30.00
##   total_cfu_count_rep3 total_cfu_mean total_cfu_sd total_cfu_se     campus
## 1                 0.70      0.5233333   0.16623277  0.095974534    Control
## 2                27.80     28.9333333   1.02632029  0.592546294 Off-Campus
## 3                 0.30      0.3400000   0.03605551  0.020816660  On-Campus
## 4                 6.20      4.3000000   1.65227116  0.953939201 Off-Campus
## 5                 0.67      0.4966667   0.15143756  0.087432514  On-Campus
## 6                 0.31      0.3133333   0.01527525  0.008819171 Off-Campus
## 7                15.10     16.9333333   1.90350554  1.098989435  On-Campus
## 8                30.00     30.0000000   0.00000000  0.000000000    Control
##       source
## 1    Control
## 2   Fountain
## 3   Fountain
## 4       Sink
## 5       Sink
## 6 Restaurant
## 7 Restaurant
## 8    Control
```

```
results %>%
  filter(campus != "Control") %>%
  ggplot(aes(x= source, y=total_cfu_mean, fill = campus)) +
  geom_bar(position = "stack", stat = "identity") +
  labs(title = "Source Types and Campus Colony Forming Units",
       x = "Source Type", y = "Total CFU Count/ (Dilution Factor x 100 mL)",
       fill = "Campus")
```



```
analysis_df  <- data.frame(campus = c("On-Campus", "On-Campus", "On-Campus",
                                      "On-Campus", "On-Campus", "On-Campus",
                                      "On-Campus", "On-Campus", "On-Campus",
                                      "Off-Campus", "Off-Campus", "Off-Campus",
                                      "Off-Campus", "Off-Campus", "Off-Campus",
                                      "Off-Campus", "Off-Campus", "Off-Campus"),
                           source = c("Fountain", "Fountain", "Fountain",
                                      "Sink", "Sink", "Sink",
                                      "Restaurant", "Restaurant", "Restaurant",
                                      "Fountain", "Fountain", "Fountain",
                                      "Sink", "Sink", "Sink",
                                      "Restaurant", "Restaurant", "Restaurant"),
                           location = c("On-Campus Fountain",
                                        "On-Campus Fountain",
                                        "On-Campus Fountain",
                                        "On-Campus Sink",
                                        "On-Campus Sink",
                                        "On-Campus Sink",
                                        "On-Campus Restaurant",
                                        "On-Campus Restaurant",
                                        "On-Campus Restaurant",
                                        "Off-Campus Fountain",
                                        "Off-Campus Fountain",
                                        "Off-Campus Fountain",
                                        "Off-Campus Sink",
                                        "Off-Campus Sink",
                                        "Off-Campus Sink",
                                        "Off-Campus Restaurant",
                                        "Off-Campus Restaurant",
                                        "Off-Campus Restaurant"),
                           total_cfu_count = c(0.35, 0.37, 0.30,
                                             0.39, 0.43, 0.67,
                                             16.80, 18.90, 15.10,
                                             29.8, 29.2, 27.8,
                                             3.50, 3.20, 6.20,
                                             0.30, 0.33, 0.31))

## ANOVA for location
summary(aov(total_cfu_count~location, data = analysis_df))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## location     5 2111.8   422.4     341 1.69e-12 ***
## Residuals   12   14.9     1.2                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## With p < 4.73x10$^{-13}$, we reject the null hypothesis that the means of all groups are the same

## Step-down pairwise tests to find significant pairs
pairs <- pairwise.t.test(analysis_df$total_cfu_count, analysis_df$location, p.adj = "bonferroni")
broom::tidy(pairs) %>%
  arrange(p.value)
```

```
## # A tibble: 15 x 3
##    group1                group2                 p.value
##    <chr>                 <chr>                    <dbl>
##  1 Off-Campus Restaurant Off-Campus Fountain   9.91e-12
##  2 On-Campus Fountain    Off-Campus Fountain   1.00e-11
##  3 On-Campus Sink        Off-Campus Fountain   1.07e-11
##  4 Off-Campus Sink       Off-Campus Fountain   5.86e-11
##  5 On-Campus Restaurant  Off-Campus Restaurant 5.92e- 9
##  6 On-Campus Restaurant  On-Campus Fountain    6.03e- 9
##  7 On-Campus Sink        On-Campus Restaurant  6.74e- 9
##  8 On-Campus Restaurant  Off-Campus Sink       1.38e- 7
##  9 On-Campus Restaurant  Off-Campus Fountain   2.48e- 7
## 10 Off-Campus Sink       Off-Campus Restaurant 1.33e- 2
## 11 On-Campus Fountain    Off-Campus Sink       1.40e- 2
## 12 On-Campus Sink        Off-Campus Sink       1.90e- 2
## 13 On-Campus Fountain    Off-Campus Restaurant 1   e+ 0
## 14 On-Campus Sink        Off-Campus Restaurant 1   e+ 0
## 15 On-Campus Sink        On-Campus Fountain    1   e+ 0
```

```
campus <- c(0.35, 0.37, 0.30, 0.39, 0.43, 0.67, 16.80, 18.90, 15.10)
off_campus <- c(29.8, 29.2, 27.8, 3.50, 3.20, 6.20, 0.30, 0.33, 0.31)
t.test(campus, off_campus, conf.level = 0.95)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  campus and off_campus
## t = -0.99726, df = 13.328, p-value = 0.3364
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.622852   6.105074
## sample estimates:
## mean of x mean of y 
##  5.923333 11.182222
```