Bacterial richness in public tap water and sinks of San Francisco’s public Pit Stop bathrooms

Andrae Ladores

April 4, 2022

# If you add any additional packages here, make sure they are  
# also listed in the DESCRIPTION file  
library("dplyr")

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library("tidyr")  
library("ggplot2")  
library("readr")  
library("ggmap")

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

library("tidycensus")  
library("tidyverse")

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ tibble 3.1.6 ✓ stringr 1.4.0  
## ✓ purrr 0.3.4 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

census\_api\_key("90f27dca0437203c57cb264c2b8efccbf7056737")

## To install your API key for use in future sessions, run this function with `install = TRUE`.

# Introduction

Throughout the world and including the United States, there are numerous individuals that are deprived of safe and reliable tap water. Cities like Newark, New Jersey and Flint, Michigan are only some examples of the water crisis in the United States (Pace et al. 2022). In California, there is a small percentage of public tap water systems that have been found to be in violation at least one point in their time since 2012 (Reese 2018). Further studies have revealed that water system violations in California transpire in low-income communities, where there are limited resources which makes it difficult to meet the standards, driving the inequitable access to safe tap water (Balazs et al. 2012, Balazs and Ray 2014). Besides access to safe public water, public restrooms are a necessity as well. They provide service to a mobile population and assists the individuals with greater needs (children, elderly, the homeless, etc.) (Stanwell-Smith 2010). In addition, public restrooms in high-income locations are known to display cleaner facilities and conditions compared to the restrooms in low and medium-income areas (Suen et al. 2019).

In this study, I focused on the bacterial richness in public tap water of San Francisco’s public restrooms (Pit Stop bathrooms) and how they vary regarding population density and the average neighborhood income. Additionally, I tested the pH levels between Outer Richmond’s Lake Stow and Lakeshore’s Lake Merced to investigate the pH levels between the lakes with differing green space. The study revolves around the use of field collected data across several districts of San Francisco and the use of pre-existing datasets relating to the study of different urban systems. The field collected data consisted of tap water samples of Pit Stop bathrooms along with swab samples of several locations on the sink in the same bathrooms. The publicly-available datasets consisted of information pertaining to San Francisco’s average income values along with San Francisco’s census data.

# Materials and Methods

## Location and Pit Stop Restrooms

As a class group, we decided which districts of San Francisco we wanted to study based on an official San Francisco district map. We concluded to observing Haight (HA), Marina (MA), Outer Richmond (OR), Civic Center (CC), Mission (MI), Bayview (BA), and Lakeshore (LA). We used Professor Zimmerman’s R code that randomly plotted points within a neighborhood to determine the exact position we should stop at via bus services.

Pit Stop bathrooms were the main public bathrooms studied and sampled. In neighborhoods where there were two Pit Stop bathrooms in close proximity to the predetermined bus stop, I sampled both bathrooms (B1 & B2), otherwise, I only sampled a single bathroom. San Francisco Public Work’s Pit Stop program supplies the public with safe and sanitized restrooms. Outside or near each unit contains used-needle disposals, animal waste stations, and a regular trash disposal. Inside each restroom contains a hands-free sink, soap, paper towels, toilet, and toilet paper. In addition, each unit gets automatically cleaned after each use and will get manually cleaned by a staff member periodically. Currently, there are 33 Pit Stop locations across 13 neighborhoods. The program utilizes street-cleaning data which determines the placement of each Pit Stops, placing the public restrooms where they are essential (Works 2014). There were some neighborhoods that did not contain a Pit Stop bathroom, in which case, I just found the nearest public restroom, which was often near or in public parks. It’s also important to note that these non Pit Stop restrooms contained bathroom facilities with both Men and Women options. In those cases, I sampled the Men’s side (B2) and I would have my female-identifying peer to assist in collecting my samples for the Women’s side (B1).

## TSA plates

The TSA plates used were prepared using 40g/L of tryptic soy agar and placed into an autoclavable glassware. Then, the addition of distilled and deionized water was added until the desired volume was reached. Once reached, the glasswared was autoclaved at 121°C for 30 minutes on liquid cycle. When it was ready, the product was poured into 100mm petri dishes to cool and stored in a closed bag at 4°C until eventually used.

## Tubes and Swabbing

For the data collection, I used 15mL falcon or centrifuge tubes containing 0.75mL phosphate buffered saline (PBS) 1x pH 7.4 which was autoclaved for 30 minutes at 121°C liquid cycle. Inside a Pit Stop bathroom, I used sterile cotton swabs which I would first dip into the PBS filled centrifuge tubes prior to swabbing three locations on the sink: the area around the spout (s1), the drain (s2), and the handle/arm rest area (s3). I would swab for 3-4 seconds and I would immediately place the swab back into its respective tube. For the negative control (c/control), I simply mimicked the same actions when swabbing the sink, but I simply swabbed the air with the same amount of time. For the water (w), I took an empty centrifuge tube without PBS and took a water sample from the automatic sink. For each Pit Stop unit, I had a total of 5 centrifuge tubes: the 3 sink samples, 1 control sample, and 1 water sample.

## Plating and Counting

Once I returned from the sample sites with my centrifuge tubes, I vortexed each centrifuge tube for 30-40 seconds. After vortexing, each tube sample (except the water sample) is aliquoted onto its respective TSA plate with a value of 0.075mL. The water sample is aliquoted with the same value of 0.075mL onto a TSA plate, but with 3 technical replicates. Per restroom, there are 7 total plates: 3 plates for the sink samples, 3 plates for the water sample, and 1 plate for the control. Once the samples are plated, approximately 7-12 sterilized glass beads are poured into the plates and each plate is shaken for 15-20 seconds or until the sample on the plate has adequately spread across the area. Once each plate has been thoroughly shaken and spread, the plates are incubated at around 37°C for 5 days.

After the incubation period, the plates are taken out for observation. Each plate is carefully observed and recorded for bacterial richness. I documented the morphology type of each bacteria, the count of each morphology type on the plate, and the overall total bacterial count present on each plate. After each plate is carefully studied and documented, the plate is properly discarded.

## Statistical Analysis

To analyze the data accumulated throughout the field work, I used boxplots, barplots, map figures, and scatterplots to illustrate my findings. For the boxplot figures created, I ran an ANOVA for the figures that contained three or more categorical variables to test for significance. For the boxplots that contained just two categorical variables, I used an independent t-test to test for significance. For the barplot figure, I subsetted the morphotype description names and their count, then I took the top ten morphotype count and plotted the results into a barplot to visualize the most abundant morphotypes. In terms of the scatterplots and regression lines, I used a paired t-test to determine significance. For all the statistical analysis to determine significance, I used a level of significance of 0.05.

## Field Sampling Design

Add text here.

## Data Analysis

# read in entire bathroom count data  
bathroom\_count <- read\_csv(  
 "data/raw\_data/2022-05-04\_2022\_spring\_urban\_ecology\_data\_Bathroom\_count.csv")

## Rows: 349 Columns: 15  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): sample\_group\_id, neighborhood, bathroom\_number, sample, sample\_typ...  
## dbl (4): mean\_median\_income, mean\_median\_population, morphotype\_count, colo...  
## date (1): date  
## time (1): time\_colony\_counted  
##   
## 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.

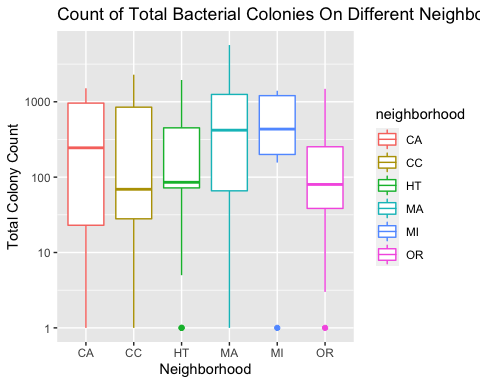
# read in entire bathroom sample data  
bathroom\_sample <- read\_csv(  
 "data/raw\_data/2022-05-03\_2022\_spring\_urban\_ecology\_data\_Bathroom\_sample.csv")

## Rows: 122 Columns: 14  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): sample\_group\_id, neighborhood, bathroom\_number, sample, sample\_typ...  
## dbl (2): lat\_dec\_degrees, long\_dec\_degrees  
## date (1): date  
## time (2): time\_sampled, time\_plated  
##   
## 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.

# create boxplot with total colony count vs. neighborhood district  
ggplot(data = bathroom\_count,  
 aes(x = neighborhood,  
 y = colony\_count\_per\_plate,  
 color = neighborhood)) +  
 geom\_boxplot() +  
 scale\_y\_log10() +  
 labs(x = "Neighborhood",  
 y = "Total Colony Count",  
 title = "Count of Total Bacterial Colonies On Different Neighborhoods",  
 scale\_fill\_manual = "Neighborhoods")

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 48 rows containing non-finite values (stat\_boxplot).



**Figure 1**: The total bacterial colony count was significantly different throughout the various San Francisco neighborhoods (ANOVA p < 0.05).

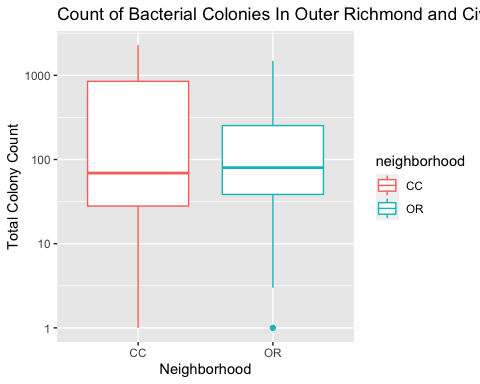
# statistical test to see if above results are significant.  
summary(lm(data = bathroom\_count, colony\_count\_per\_plate ~ neighborhood))

##   
## Call:  
## lm(formula = colony\_count\_per\_plate ~ neighborhood, data = bathroom\_count)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -958.8 -380.8 -195.5 125.8 4699.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 414.79 139.23 2.979 0.00311 \*\*  
## neighborhoodCC -16.66 191.59 -0.087 0.93077   
## neighborhoodHT -91.96 192.57 -0.478 0.63330   
## neighborhoodMA 544.03 168.03 3.238 0.00133 \*\*  
## neighborhoodMI 74.06 225.17 0.329 0.74243   
## neighborhoodOR -219.24 171.21 -1.280 0.20128   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 902.3 on 329 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.09832, Adjusted R-squared: 0.08461   
## F-statistic: 7.175 on 5 and 329 DF, p-value: 2.17e-06

# create boxplot for the total colony count for only Civic Center and  
# Outer Richmond neighborhoods.  
bathroom\_count %>%  
 filter(neighborhood %in% c("OR", "CC")) %>%  
 ggplot(aes(x = neighborhood,  
 y = colony\_count\_per\_plate,  
 color = neighborhood)) +  
 geom\_boxplot() +  
 scale\_y\_log10() +  
 labs(x = "Neighborhood",  
 y = "Total Colony Count",  
 title = "Count of Bacterial Colonies In Outer Richmond and Civic Center")

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 22 rows containing non-finite values (stat\_boxplot).



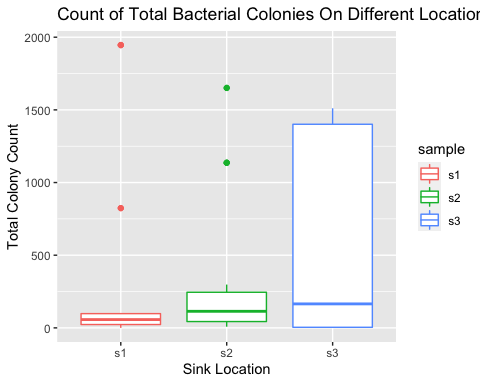
**Figure 2**: The total bacterial colony count was significantly different between Civic Center and Outer Richmond neighborhoods (t-test p < 0.05).

# statistical test to see if above results are significant.  
bathroom\_count %>%  
 filter(neighborhood %in% c("OR", "CC")) %>%  
 lm(data = ., colony\_count\_per\_plate ~ neighborhood) %>%  
 summary()

##   
## Call:  
## lm(formula = colony\_count\_per\_plate ~ neighborhood, data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -398.13 -292.13 -144.55 57.45 1889.87   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 398.13 72.23 5.512 1.89e-07 \*\*\*  
## neighborhoodOR -202.58 90.59 -2.236 0.0271 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 495.2 on 127 degrees of freedom  
## (9 observations deleted due to missingness)  
## Multiple R-squared: 0.03788, Adjusted R-squared: 0.03031   
## F-statistic: 5.001 on 1 and 127 DF, p-value: 0.02708

# create boxplot with total plate count and the different sink locations.  
bathroom\_count %>%  
 filter(sample %in% c("s1", "s2", "s3")) %>%  
 ggplot(aes(x = sample, y = colony\_count\_per\_plate, color = sample)) +  
 geom\_boxplot() +  
 labs(x = "Sink Location",  
 y = "Total Colony Count",  
 title =  
 "Count of Total Bacterial Colonies On Different Locations On The Sink")

## Warning: Removed 4 rows containing non-finite values (stat\_boxplot).



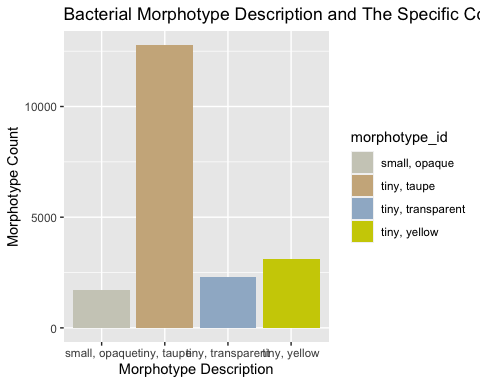
**Figure 3**: The total bacterial colony count was not significantly different between various sink locations (ANOVA p < 0.05).

# statistical test to see if above results are significant.  
bathroom\_count %>%  
 filter(sample %in% c("s1", "s2", "s3")) %>%  
 lm(data = ., colony\_count\_per\_plate ~ sample) %>%  
 summary()

##   
## Call:  
## lm(formula = colony\_count\_per\_plate ~ sample, data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -495.69 -330.69 -276.36 -86.22 1571.64   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 374.36 130.06 2.879 0.00548 \*\*  
## samples2 -43.14 175.20 -0.246 0.80631   
## samples3 121.32 200.43 0.605 0.54717   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 610 on 62 degrees of freedom  
## (4 observations deleted due to missingness)  
## Multiple R-squared: 0.01184, Adjusted R-squared: -0.02003   
## F-statistic: 0.3715 on 2 and 62 DF, p-value: 0.6912

# subset bathroom count data to only contain morphotype ID and morphotype count.  
# Then only have the top 10 morphotype counts.  
bathroom\_count\_table <- bathroom\_count %>%  
 select(morphotype\_id, morphotype\_count) %>%  
 arrange(desc(morphotype\_count)) %>%  
 head(10)

# create barplot that showcases the top morphotype IDs and their counts.  
bathroom\_count\_barplot <- ggplot(data = bathroom\_count\_table,  
 aes(x = morphotype\_id,  
 y = morphotype\_count,  
 fill = morphotype\_id)) +  
 geom\_bar(stat = "identity") +  
 scale\_fill\_manual(values = c("#CDCDC1",  
 "#CDB38B",  
 "#9FB6CD",  
 "#CDCD00")) +  
 labs(title = "Bacterial Morphotype Description and The Specific Counts",  
 x = "Morphotype Description",  
 y = "Morphotype Count")  
bathroom\_count\_barplot



**Figure 4**: Barplot of the top ten morphotype descriptions.

# load in base sf map  
sf\_basemap <- ggmap::get\_stamenmap(bbox = c(-122.533264,  
 37.700984,  
 -122.355307,  
 37.817383),  
 zoom = 13,  
 maptype = "terrain")

## Source : http://tile.stamen.com/terrain/13/1307/3165.png

## Source : http://tile.stamen.com/terrain/13/1308/3165.png

## Source : http://tile.stamen.com/terrain/13/1309/3165.png

## Source : http://tile.stamen.com/terrain/13/1310/3165.png

## Source : http://tile.stamen.com/terrain/13/1311/3165.png

## Source : http://tile.stamen.com/terrain/13/1307/3166.png

## Source : http://tile.stamen.com/terrain/13/1308/3166.png

## Source : http://tile.stamen.com/terrain/13/1309/3166.png

## Source : http://tile.stamen.com/terrain/13/1310/3166.png

## Source : http://tile.stamen.com/terrain/13/1311/3166.png

## Source : http://tile.stamen.com/terrain/13/1307/3167.png

## Source : http://tile.stamen.com/terrain/13/1308/3167.png

## Source : http://tile.stamen.com/terrain/13/1309/3167.png

## Source : http://tile.stamen.com/terrain/13/1310/3167.png

## Source : http://tile.stamen.com/terrain/13/1311/3167.png

## Source : http://tile.stamen.com/terrain/13/1307/3168.png

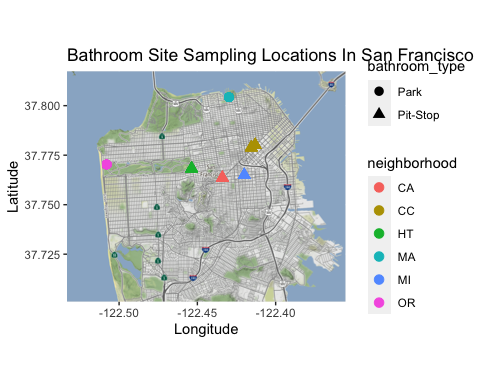
## Source : http://tile.stamen.com/terrain/13/1308/3168.png

## Source : http://tile.stamen.com/terrain/13/1309/3168.png

## Source : http://tile.stamen.com/terrain/13/1310/3168.png

## Source : http://tile.stamen.com/terrain/13/1311/3168.png

# plot basemap with each bathroom sample location for each neighborhood  
ggmap(sf\_basemap) +  
 geom\_point(data = bathroom\_sample,  
 aes(x = long\_dec\_degrees,  
 y = lat\_dec\_degrees,  
 color = neighborhood,  
 shape = bathroom\_type),  
 size = 3) +  
 labs(title = "Bathroom Site Sampling Locations In San Francisco",  
 x = "Longitude",  
 y = "Latitude")



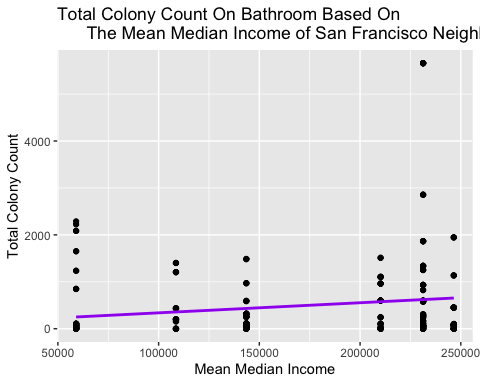
**Figure 5**: Sampled Bathroom Sites throughout San Francisco’s neighborhoods.

bathroom\_count %>%  
 ggplot(aes(y = colony\_count\_per\_plate, # plot sample type data  
 x = mean\_median\_income)) + # plot colony count data  
 geom\_point(color = "black") +  
 geom\_smooth(method = "lm", se = FALSE, color = "purple") +  
 labs(title = "Total Colony Count On Bathroom Based On  
 The Mean Median Income of San Francisco Neighborhoods",  
 x = "Mean Median Income",  
 y = "Total Colony Count")

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 14 rows containing non-finite values (stat\_smooth).

## Warning: Removed 14 rows containing missing values (geom\_point).



**Figure 6**: The total bacterial colony count was significantly different throughout the varied average income (p < 0.05).

# statistical test to see if above results are significant.  
summary(lm(data = bathroom\_count, colony\_count\_per\_plate ~ mean\_median\_income))

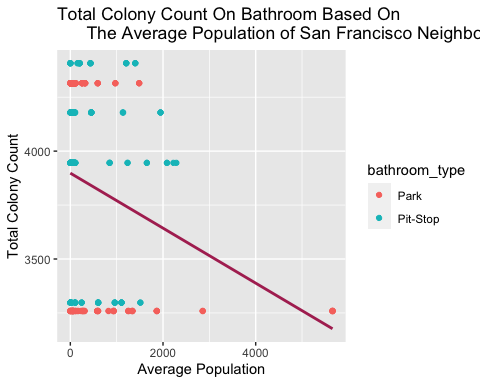
##   
## Call:  
## lm(formula = colony\_count\_per\_plate ~ mean\_median\_income, data = bathroom\_count)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -654.0 -473.8 -316.2 24.2 5036.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.235e+02 1.474e+02 0.838 0.40273   
## mean\_median\_income 2.152e-03 7.877e-04 2.732 0.00663 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 934.1 on 333 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.02192, Adjusted R-squared: 0.01898   
## F-statistic: 7.464 on 1 and 333 DF, p-value: 0.006631

bathroom\_count %>%  
 ggplot(aes(y = mean\_median\_population, # plot population  
 x = colony\_count\_per\_plate,  
 color = bathroom\_type)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE, color = "#B03060") +  
 labs(title = "Total Colony Count On Bathroom Based On  
 The Average Population of San Francisco Neighborhoods",  
 x = "Average Population",  
 y = "Total Colony Count")

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 14 rows containing non-finite values (stat\_smooth).

## Warning: Removed 14 rows containing missing values (geom\_point).



**Figure 7**: The total bacterial colony count was significantly different throughout the average population (p < 0.05).

# statistical test to see if above results are significant.  
bathroom\_count %>%  
 lm(data = ., colony\_count\_per\_plate ~ mean\_median\_population) %>%  
 summary()

##   
## Call:  
## lm(formula = colony\_count\_per\_plate ~ mean\_median\_population,   
## data = .)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -787.6 -433.8 -215.8 121.4 4870.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2410.1871 404.9331 5.952 6.71e-09 \*\*\*  
## mean\_median\_population -0.4979 0.1048 -4.750 3.02e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 914.1 on 333 degrees of freedom  
## (14 observations deleted due to missingness)  
## Multiple R-squared: 0.06346, Adjusted R-squared: 0.06065   
## F-statistic: 22.57 on 1 and 333 DF, p-value: 3.024e-06

# Results

## Microbial Richness

I observed that the overall bacterial richness throughout the sampled bathroom and their neighborhoods (Figure 5) was significantly different with an ANOVA p-value of 2.17E-6; p < 0.05 (Figure 1). Comparing two specific neighborhood’s bacterial richness, Outer Richmond and Civic Center/Tenderloin, with differing average income, I found that the bacterial richness was significantly different between Civic Center/Tenderloin and Outer Richmond area with a t-test p-value of 0.02708; p < 0.05 (Figure 2). Furthermore, I observed the relationship between the bacterial richness and the varying sink locations and found that the bacterial richness was not significantly different between the various sink locations (Figure 3; p-value = 0.6912; p < 0.05).

## Morphotypes

My barplot (Figure 4) to analyze the top morphotype count of the various morphotype identification descriptions and their respective bacterial abundance, found that the morphotype identification, tiny, taupe, had the most frequent and abundant morphotype count of 14,032. The second most abundant morphotype, tiny, yellow, amounted to a morphotype count of 3,116. That is a 10,916 difference between the most abundant morphotype, tiny, taupe, and the second most abundant morphotype, tiny, yellow. Furthermore, the morphotypes, tiny, transparent, totaled to the value of 2,288 and small, opaque, measured to a total of 1728.

## Average Income & Population

The scatterplot to measure the microbial richness along with the average income in San Francisco’s neighborhoods was significantly different with a p-value of 0.006631; p < 0.05 (Figure 6). Similarly, the relationship between the microbial richness and the average population throughout San Francisco was significant, it had a t-test value of p-value of 3.024E-6, p < 0.05 (Figure 7).

# Discussion

# Sources Cited

Balazs, C. L., R. Morello-Frosch, A. E. Hubbard, and I. Ray. 2012. Environmental justice implications of arsenic contamination in california’s san joaquin valley: A cross-sectional, cluster-design examining exposure and compliance in community drinking water systems. Environmental Health 11:1–12.

Balazs, C. L., and I. Ray. 2014. [The drinking water disparities framework: On the origins and persistence of inequities in exposure](https://doi.org/10.2105/AJPH.2013.301664). American Journal of Public Health 104:603–611.

Pace, C., C. Balazs, K. Bangia, N. Depsky, A. Renteria, R. Morello-Frosch, and L. J. Cushing. 2022. [Inequities in drinking water quality among domestic well communities and community water systems, california, 2011‒2019](https://doi.org/10.2105/AJPH.2021.306561). American Journal of Public Health 112:88–97.

Reese, P. 2018. Does your water district fail to comply with drinking water standards. Sacramento Bee.

Stanwell-Smith, R. 2010. [Public toilets down the drain? Why privies are a public health concern](https://doi.org/10.1016/j.puhe.2010.07.002). Public Health 124:613–616.

Suen, L. K., G. K. Siu, Y. P. Guo, S. K. Yeung, K. Y. Lo, and M. O’Donoghue. 2019. The public washroom-friend or foe? An observational study of washroom cleanliness combined with microbiological investigation of hand hygiene facilities. Antimicrobial Resistance & Infection Control 8:1–6.

Works, S. F. P. 2014. San francisco pit stop.