A Report on Lake Ontario's Microbes

Mark Watson

2025-02-19

Prepare the R environment

```
#load libraries/packages for file
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2

## Warning: package 'readr' was built under R version 4.4.2

## Warning: package 'forcats' was built under R version 4.4.2

## Warning: package 'lubridate' was built under R version 4.4.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Load in the Lake Ontario Data

```
# load in lake ontario microbial community data
sample_and_taxon <-
  read_csv("data/sample_and_taxon.csv")
```

```
## Rows: 71 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): sample_id, env_group
## dbl (13): depth, cells_per_ml, temperature, total_nitrogen, total_phosphorus...
##
## 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
#inspect data
glimpse(sample_and_taxon)
```

```
## Rows: 71
## Columns: 15
## $ sample_id         <chr> "May_12_B", "May_12_E", "May_12_M", "May_17_E", "May~
## $ env_group         <chr> "Deep", "Shallow_May", "Shallow_May", "Shallow_May",~
## $ depth             <dbl> 102.8, 5.0, 15.0, 5.0, 27.0, 5.0, 19.0, 135.0, 5.0, ~
## $ cells_per_ml      <dbl> 2058864, 4696827, 4808339, 3738681, 2153086, 3124920~
## $ temperature       <dbl> 4.07380, 7.01270, 6.13500, 5.99160, 4.66955, 5.97390~
## $ total_nitrogen    <dbl> 465, 465, 474, 492, 525, 521, 539, 505, 473, 515, 47~
## $ total_phosphorus  <dbl> 3.78, 4.39, 5.37, 4.67, 4.44, 3.71, 4.23, 4.18, 6.64~
## $ diss_org_carbon   <dbl> 2.478, 2.380, 2.601, 2.435, 2.396, 2.283, 2.334, 2.3~
## $ chlorophyll       <dbl> 0.05, 2.53, 3.20, 0.55, 0.48, 0.79, 0.44, 0.22, 3.44~
## $ Proteobacteria    <dbl> 0.4120986, 0.3389293, 0.2762080, 0.4351188, 0.410063~
## $ Actinobacteriota  <dbl> 0.1288958, 0.1861232, 0.2866884, 0.1910769, 0.280123~
## $ Bacteroidota      <dbl> 0.08065717, 0.23470807, 0.21659843, 0.21576244, 0.11~
## $ Chloroflexi       <dbl> 0.19463564, 0.08086689, 0.07032061, 0.08498357, 0.13~
## $ Verrucomicrobiota <dbl> 0.13249532, 0.10878214, 0.09991639, 0.05752092, 0.06~
## $ Cyanobacteria     <dbl> 2.482454e-04, 9.574640e-03, 1.262830e-02, 1.288730e-~
```

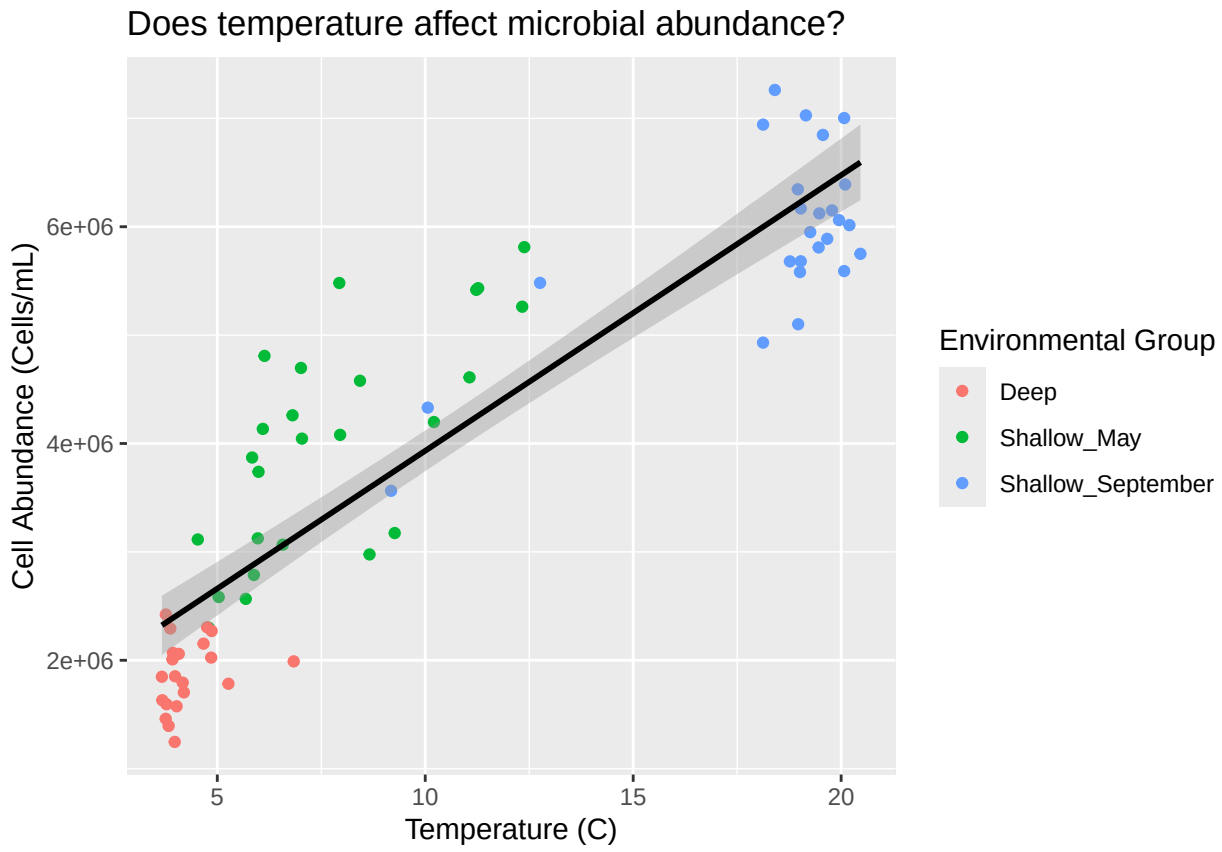
```r
sample_data <-
  read_csv("data/sample_data.csv")
```

```
## Rows: 71 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): sample_id, env_group
## dbl (7): depth, cells_per_ml, temperature, total_nitrogen, total_phosphorus,...
##
## 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.
```

Lake Ontario Microbial Abundance Versus Temperature

```r
# plot
# temp on x
# cel abundance on Y
# colored by env_group
# make it look nice
ggplot(data = sample_data) +
  aes(x = temperature) +
  labs(x = "Temperature (C)") +
  aes(y = cells_per_ml) +
  labs(y = "Cell Abundance (Cells/mL)")+
  geom_point(aes(color = env_group)) +
  labs(title = "Does temperature affect microbial abundance?") +
  geom_smooth(method = lm, color = "black") +
  labs(color = "Environmental Group")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



The above plot shows that:

- Temperature and Cell Abundance are positively correlated.
- Deep Samples are the coldest and have the fewest cells.
- Shallow Samples are warmer and have more cells.

The total number of samples is r n_samples. For this set of samples, temperature ranged from 3.7 to 20.5 Degrees Celsius.