

Data visualization practice

Experiential Data science for Undergraduate Cross-disciplinary Education

Dr. Kim Dill-McFarland, U. of British Columbia

Contents

Data visualization practice	1
Learning objectives	1
Setup	1
Data motivation	2
Practice	2

Data visualization practice

These problems are designed to help you practice concepts and functions covered in the ‘Data visualization tutorial’.

Learning objectives

- Define the grammar of graphics
- Create scatterplots using the `ggplot2` package
- Customize plot color, shape, axes, scales, and other attributes
- Represent subsets of data using facets

Setup

Open a new RStudio session, create a Project, and start a new R script to save you responses.

Install and load `tidyverse`

If you have not done so already, install the `tidyverse` packages. (More information available in our ‘RStudio tutorial’.)

```
install.packages("tidyverse")
```

*Please note that if you have **R v3.3 or older**, you may not be able to install **tidyverse**. In this case, you need to separately install each package within the **tidyverse**.*

For this practice, you will need to load:

R v3.4 or newer

```
library(tidyverse)
```

R v3.3 or older

```
library(readr)
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.5.2
```

```
library(ggplot2)
```

Load data

Load the pre-cleaned data from our GitHub.

```
pdat <- read.csv(  
  "https://raw.githubusercontent.com/EDUCE-UBC/workshop_data/master/Saanich_Data_clean.csv")
```

If you would like to learn more about Saanich Inlet and these data, checkout [our description](#) as well as how these data were cleaned in our ‘Data manipulation in R tutorial’.

Data motivation

As in the ‘Data visualization tutorial’, we will use plots to investigate the relationship between **oxygen** (O_2), **nitrate** (NO_3), and **hydrogen sulfide** (H_2S) in Saanich Inlet. We are particularly interested in these relationships, because oxygen is the most energetically favorable terminal electron acceptor for microbial metabolism in this system. However, oxygen is depleted in deeper waters as a result of the seasonal stratification.

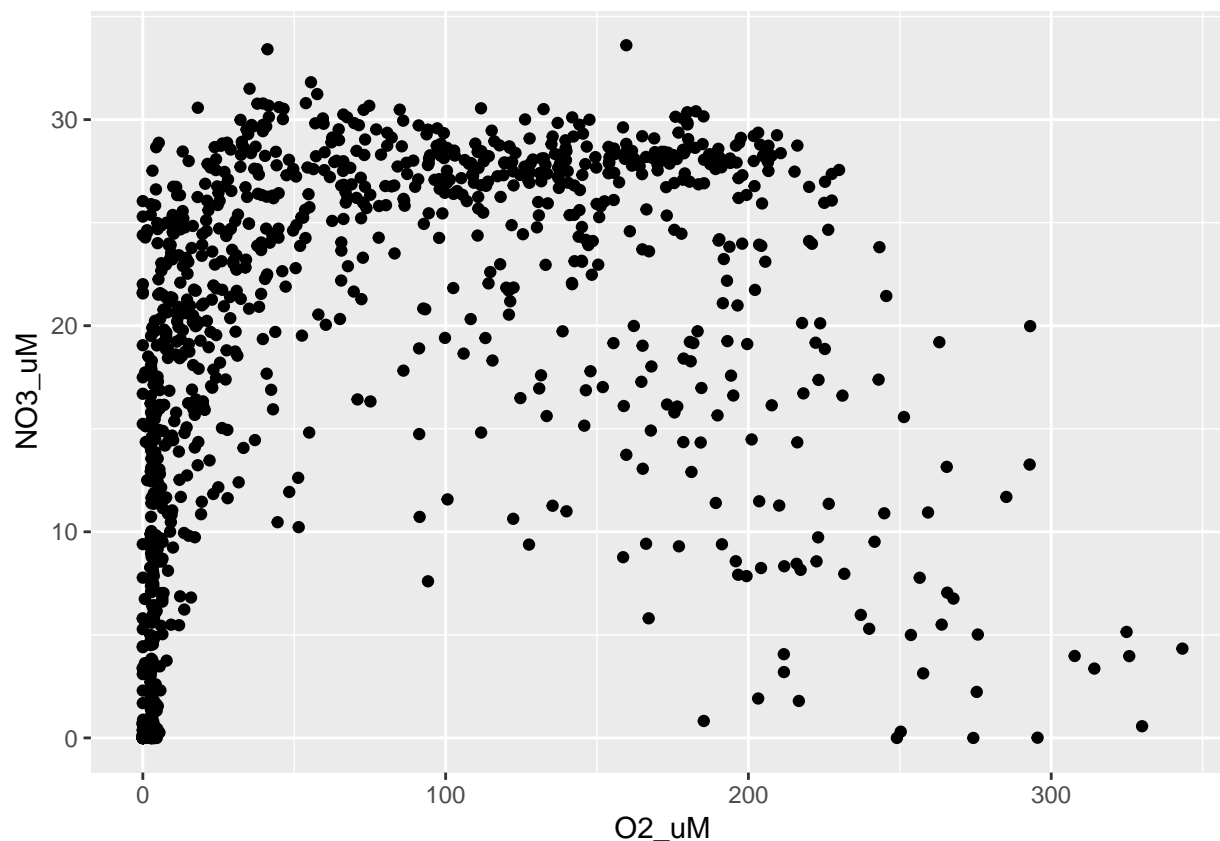
So, what are microbes using to survive? We will explore this further with the following practice exercises.

Practice

NOTE: Variable names containing O2 are O as in **O**xygen, not 0 as in zero.

color and shape

In our tutorial, we first explored the relationship between oxygen and nitrate below.



1. Continue this data exploration by plotting the other variable combinations below. Customize the shape and color of points in your plots.
 - oxygen vs. hydrogen sulfide
 - nitrate vs. hydrogen sulfide

size

2. Using the `size` option in `geom_point` (similar to `shape` and `color`), change the size of your data points from one of the plots in exercise 1. Try a few different values to create what you feel is the most readable figure.

themes

By default, ggplot uses a gray color scheme. However, there are a number of other themes that come with ggplot as well as many more in the additional package `ggthemes`. You can see many examples from [ggplot](#) or [ggthemes](#).

3. Choose a theme and apply it to one of your plots.
4. Further customize the theme using the `theme()` function to customize your axes labels. *Hint* Explore this function's help page with `?theme` to discover the available arguments.

facet_wrap

The above plots indicate that oxygen and nitrate might be related, but depth and other compounds, like hydrogen sulfide, might also play a role in this relationship. So, how might we further tease apart this plot?

5. Investigate if depth contributes to the relationship between oxygen and nitrate. We will look at only a couple of depths to simplify the plot.
- Filter to data to depths 10, 60, 100 and 200 m
 - Plot Oxygen vs Nitrate faceted by Depth

The result should be the following

