tidyverse exercise solutions

Kris Hong & Kim Dill-McFarland

version March 27, 2018

Contents

Setup

We will be working with the same data and packages as in the notes and main.R files.

```
library(tidyverse)
```

```
raw_data <- read_csv(file="Saanich_Data.csv",
                     col_names=TRUE,
                     na=c("", "NA", "NAN", "ND"))

dat <- raw_data
```

Because we will be manipulating the data further, first copy the data "dat" to practice data "pdat" so that what you do in the exercises does not impact the rest of the workshop.

```
pdat <- raw_data
```

dplyr

select and filter

Exercise. Using your pdat data:

1. select the Cruise, Date, Depth, PO4, and NO3 variables
2. filter the data to retain data on Cruise 72 where Depth is >= to 0.1

Your resulting pdat object should be a 8x5 data frame.

```
# 1
pdat <- select(pdat, Cruise, Date, Depth, PO4, WS_NO3)

# 2
pdat <- filter(pdat, Cruise == 72 & Depth >= 0.1)

# Check dimensions
dim(pdat)
```

```
## [1] 8 5
```

rename and mutate

To start fresh, remake the pdat object.

```
pdat <- raw_data
```

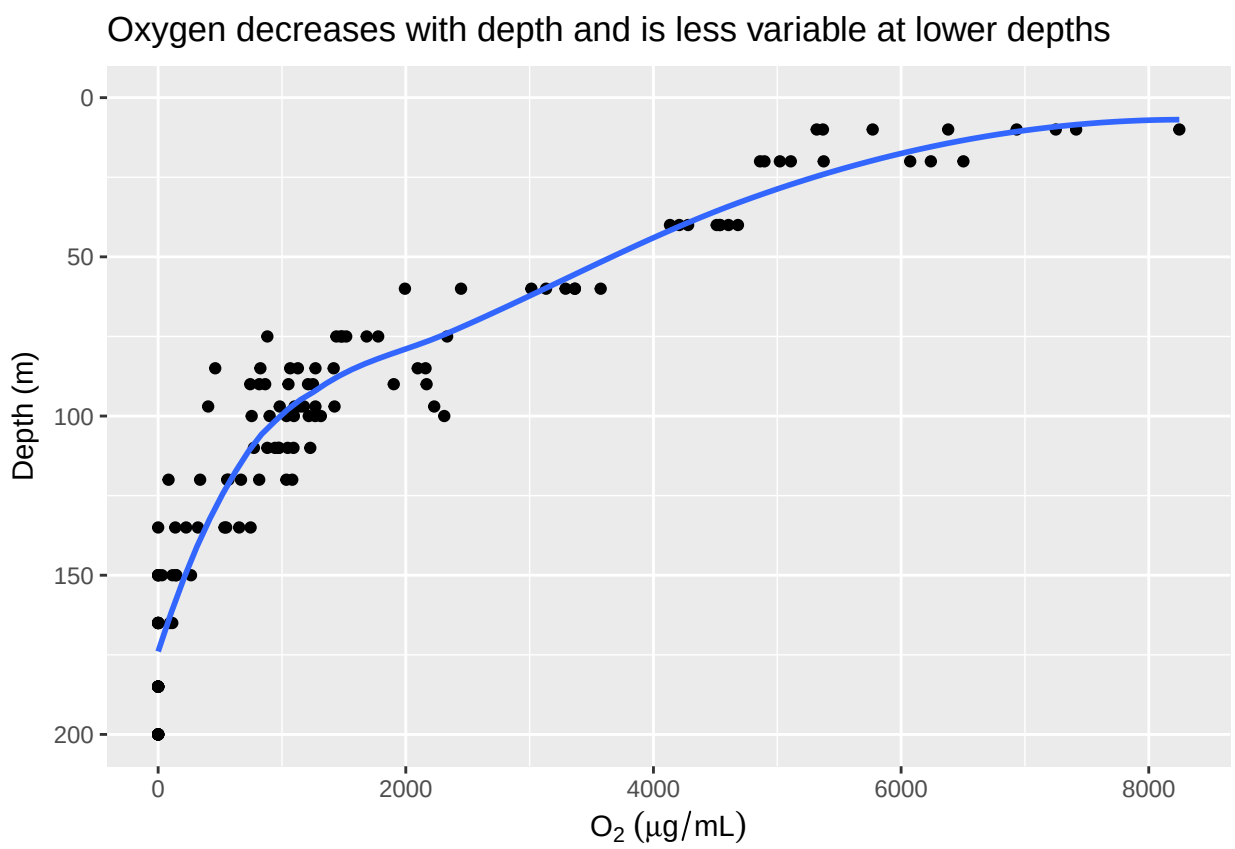
Exercise. Using the practice data (pdat):

1. Select the Date, Depth and O2 variables from pdat using select
2. Rename the O2 variable to Oxygen using rename
3. Keep August observations where Oxygen is nonmissing using filter, months, and !is.na
4. Transform Oxygen from micromoles/L to micrograms/L using mutate (multiply Oxygen by 32)
5. Run the provided ggplot() code to create a scatterplot of Oxygen vs Depth

```
# 1
pdat <- select(pdat, Date, Depth, WS_O2)

# 2
pdat <- rename(pdat, Oxygen=WS_O2)

# 3
pdat <- filter(pdat, months(Date) == "August")

# 4
pdat <- mutate(pdat, Oxygen=Oxygen*32)
```

Plug this modified pdat data into a ggplot function.

```
# 5
ggplot(pdat, aes(x=Oxygen, y=Depth*1000)) +
  geom_point() +
  geom_smooth(method="loess", se=FALSE) +
  scale_y_reverse(limits=c(200, 0)) +
  labs(x=expression(O[2]*" "*(mu*g/mL)),
       y="Depth (m)",
       title="Oxygen decreases with depth and is less variable at lower depths")
```



Pipes

Again start fresh.

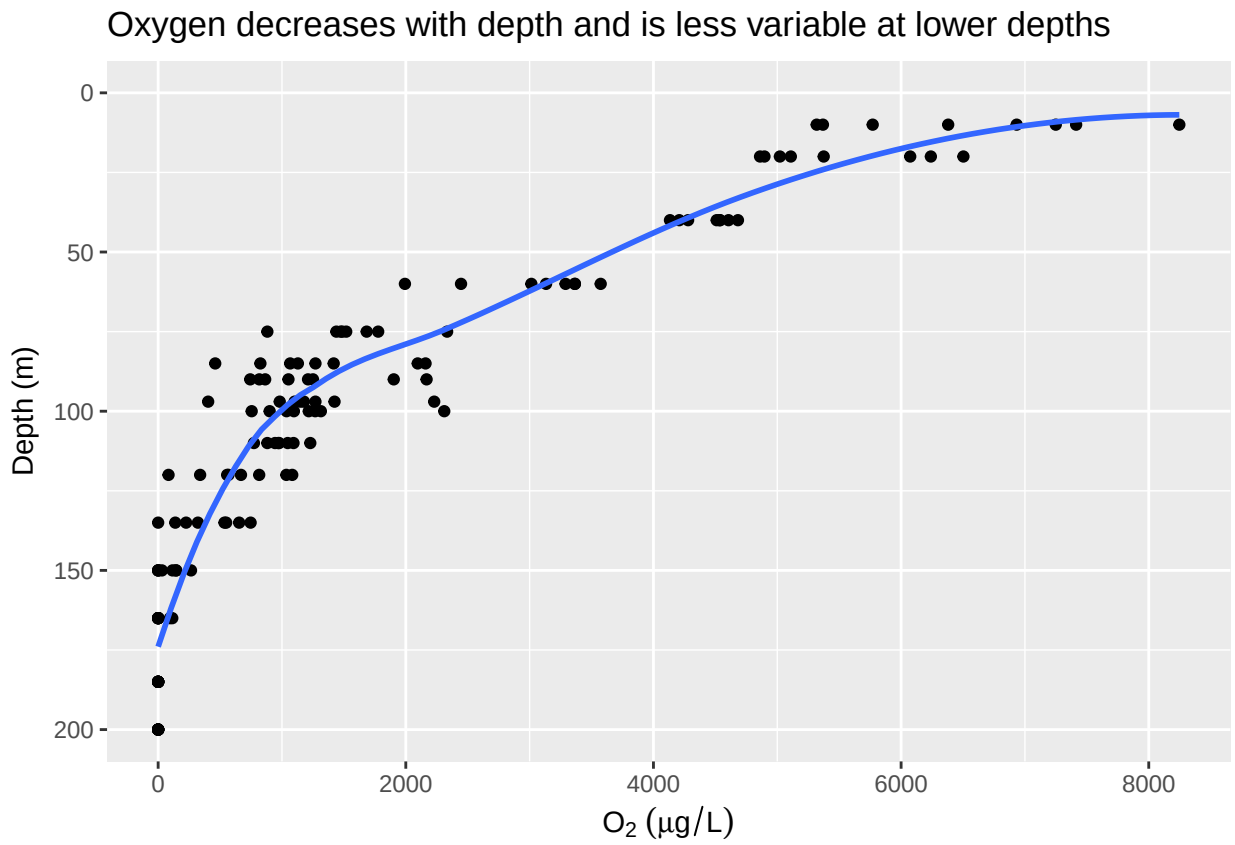
```
pdat <- raw_data
```

Exercise. Using your pdat data:

1. Rewrite your code from the previous exercise using pipes
2. Pipe your data into the ggplot function

```
#  1
pdat %>%
  select(Date, Depth, WS_O2) %>%
  rename(Oxygen=WS_O2) %>%
```

```
  filter(months(Date) == "August" & !is.na(Oxygen)) %>%
  mutate(Oxygen=Oxygen*32) %>%

# 2
  ggplot(aes(x=Oxygen, y=Depth*1000)) +
  geom_point() +
  geom_smooth(method="loess", se=FALSE) +
  scale_y_reverse(limits=c(200, 0)) +
  labs(x=expression(O[2]*" "*(mu*g/L)),
       y="Depth (m)",
       title="Oxygen decreases with depth and is less variable at lower depths")
```



Setup 2.0

We will want to use cleaned-up data in the remaining exercises. So use the following to modify the raw_data and create a new pdat object.

```
pdat <-
  raw_data %>%
  select(Cruise, Date, Depth,
         Temperature, Salinity, Density,
         WS_O2, WS_NO3, WS_H2S) %>%
```

```
  filter(Date >= "2008-02-01") %>%
  rename(O2=WS_O2, NO3=WS_NO3, H2S=WS_H2S) %>%
  mutate(Depth=Depth*1000)
```

summarise

Exercise. Using your new, cleaned-up pdat data:

1. Calculate median, interquartile range, and sample size of Temperature by depth

```
pdat %>%
  group_by(Depth) %>%
  summarise(Median_Temperature=median(Temperature, na.rm=TRUE),
            IQR_Temperature=IQR(Temperature, na.rm=TRUE),
            n=n())
```

```
## # A tibble: 24 x 4
##    Depth Median_Temperature IQR_Temperature     n
##    <dbl>              <dbl>           <dbl> <int>
##  1   10.              10.0            4.20     83
##  2   20.               9.40           3.31     83
##  3   40.               9.10           2.03     83
##  4   60.               8.56           1.42     83
##  5   75.               8.48           0.959    82
##  6   80.              NA             NA         1
##  7   85.               8.56           0.759    82
##  8   90.               8.60           0.752    82
##  9   97.               8.67           0.705    82
## 10  100.               8.70           0.644    83
## # ... with 14 more rows
```

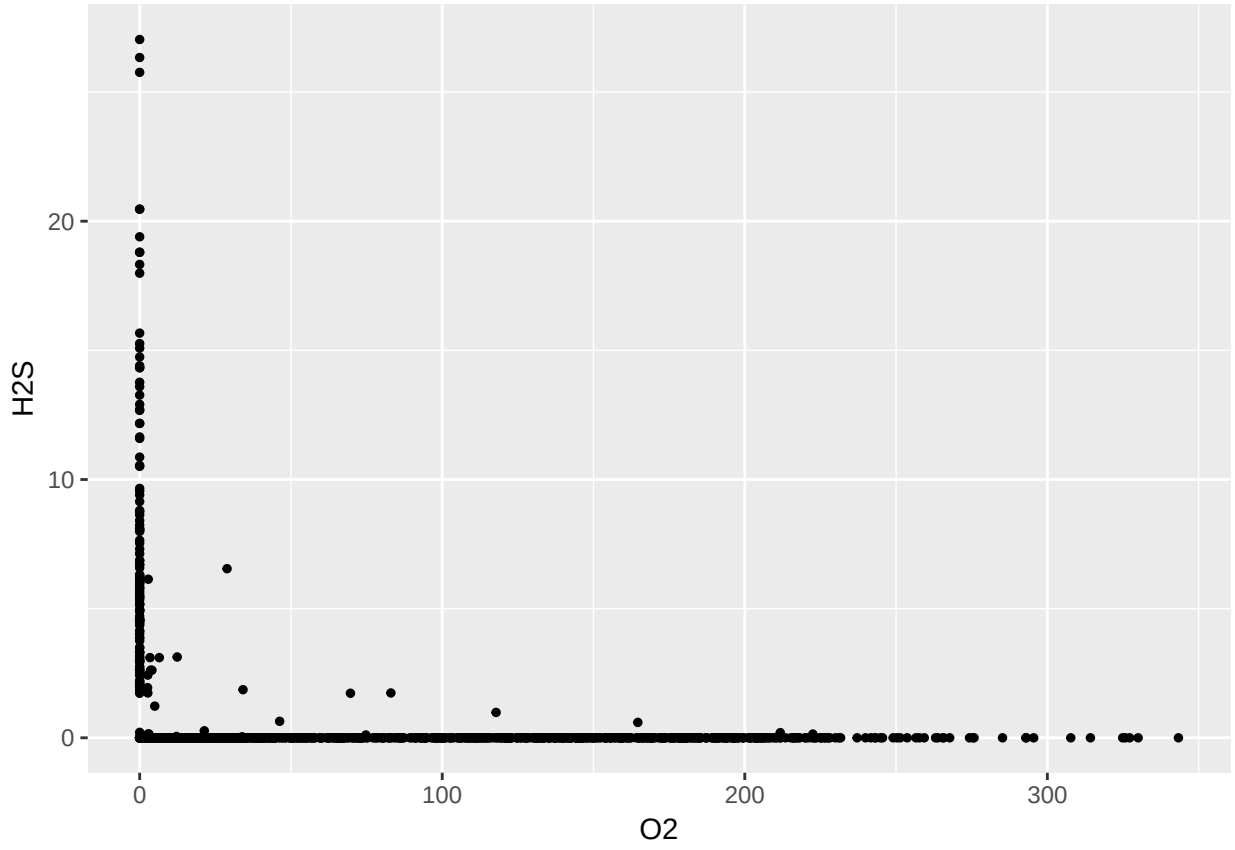
ggplot

geom_point

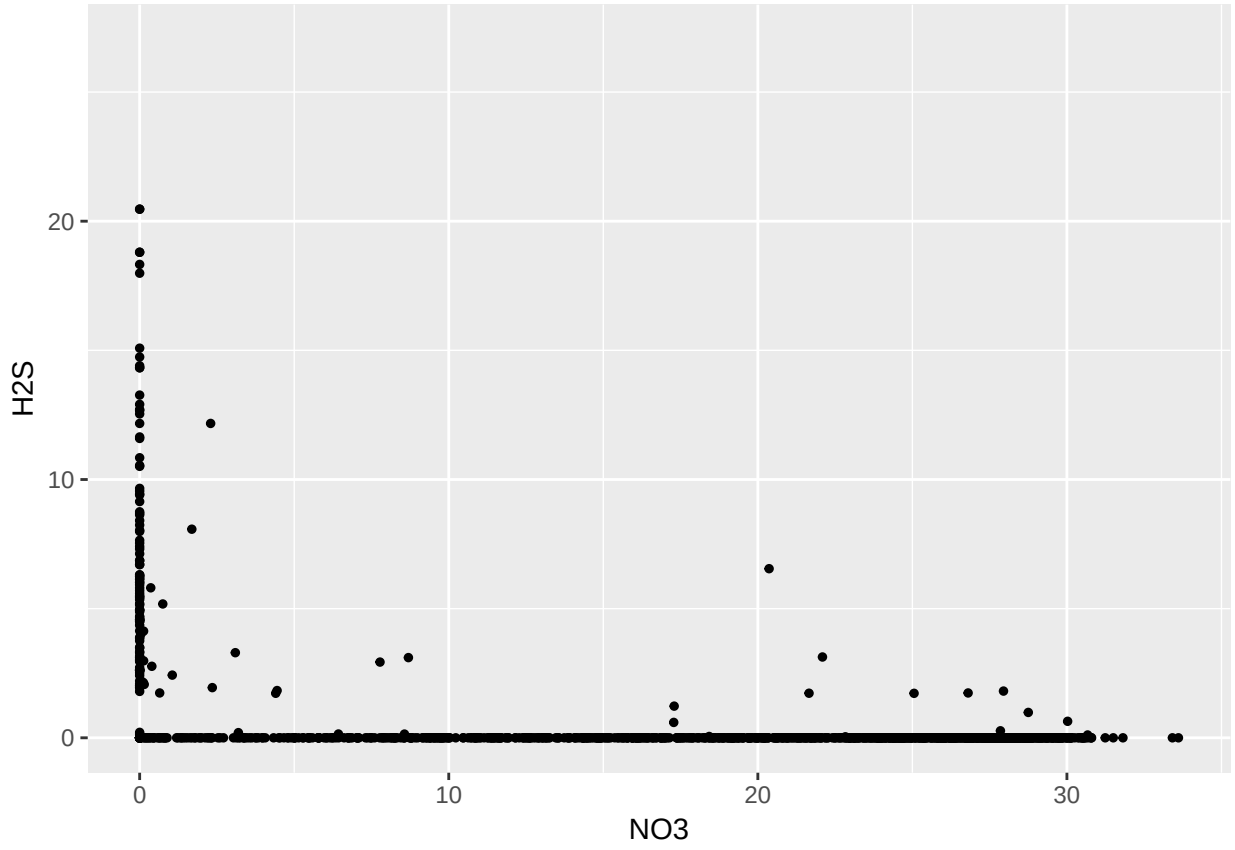
Exercise. Using your pdat data:

1. Investigate the relationship between O2 and H2S
2. Investigate the relationship between NO3 and H2S

```
# 1
pdat %>%
  ggplot(aes(x=O2, y=H2S)) +
  geom_point(size=1)
```




```
# 2
pdat %>%
  ggplot(aes(x=NO3, y=H2S)) +
  geom_point(size=1)
```

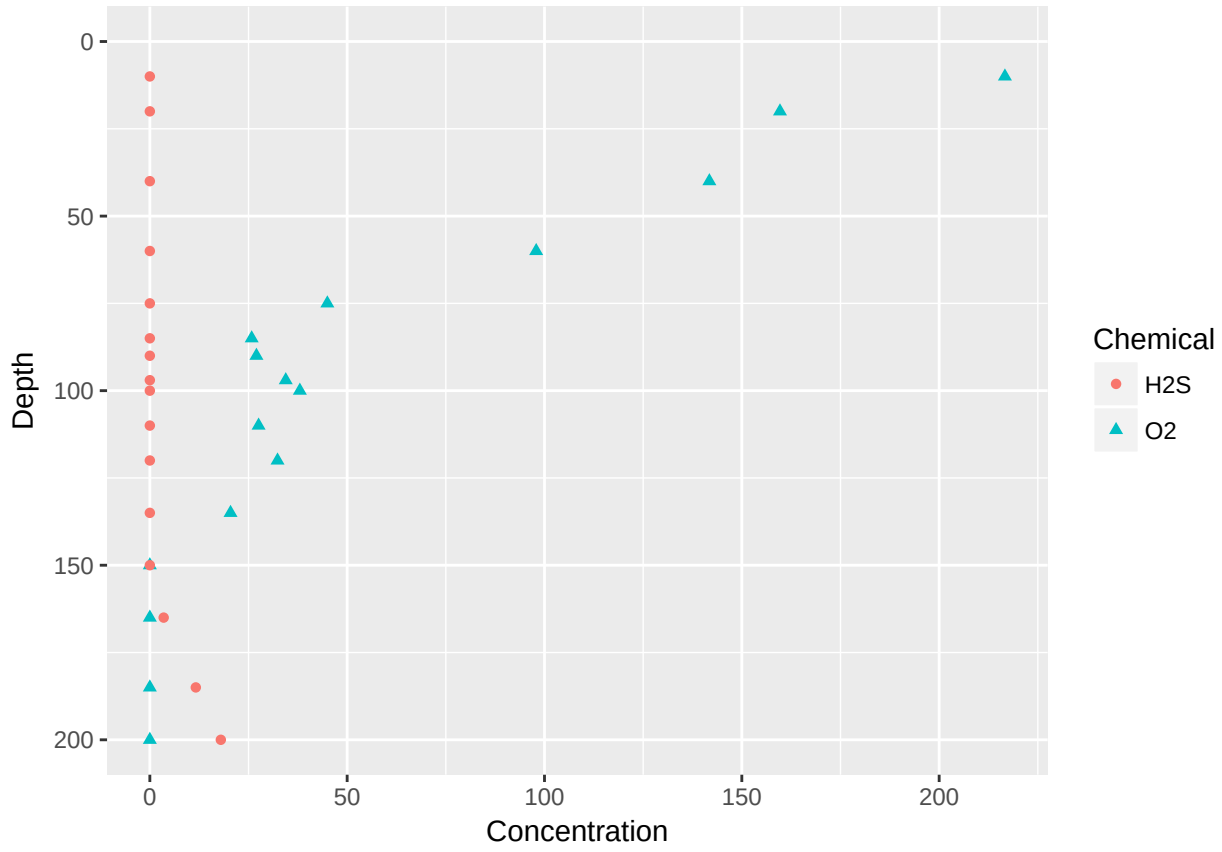


color and shape aesthetics

Exercise. Using your pdat data:

1. It may be difficult to differentiate between the different shapes in the previous plot so modify the following code to add colours to the shapes as well:

```
pdat %>%
  select(Cruise, Depth, O2, H2S) %>%
  filter(Cruise==72) %>%
  gather(key="Chemical", value="Concentration", -Cruise, -Depth) %>%

  ggplot(aes(x=Concentration, y=Depth, shape=Chemical, colour=Chemical)) +
  geom_point() +
  scale_y_reverse(limits=c(200, 0))
```



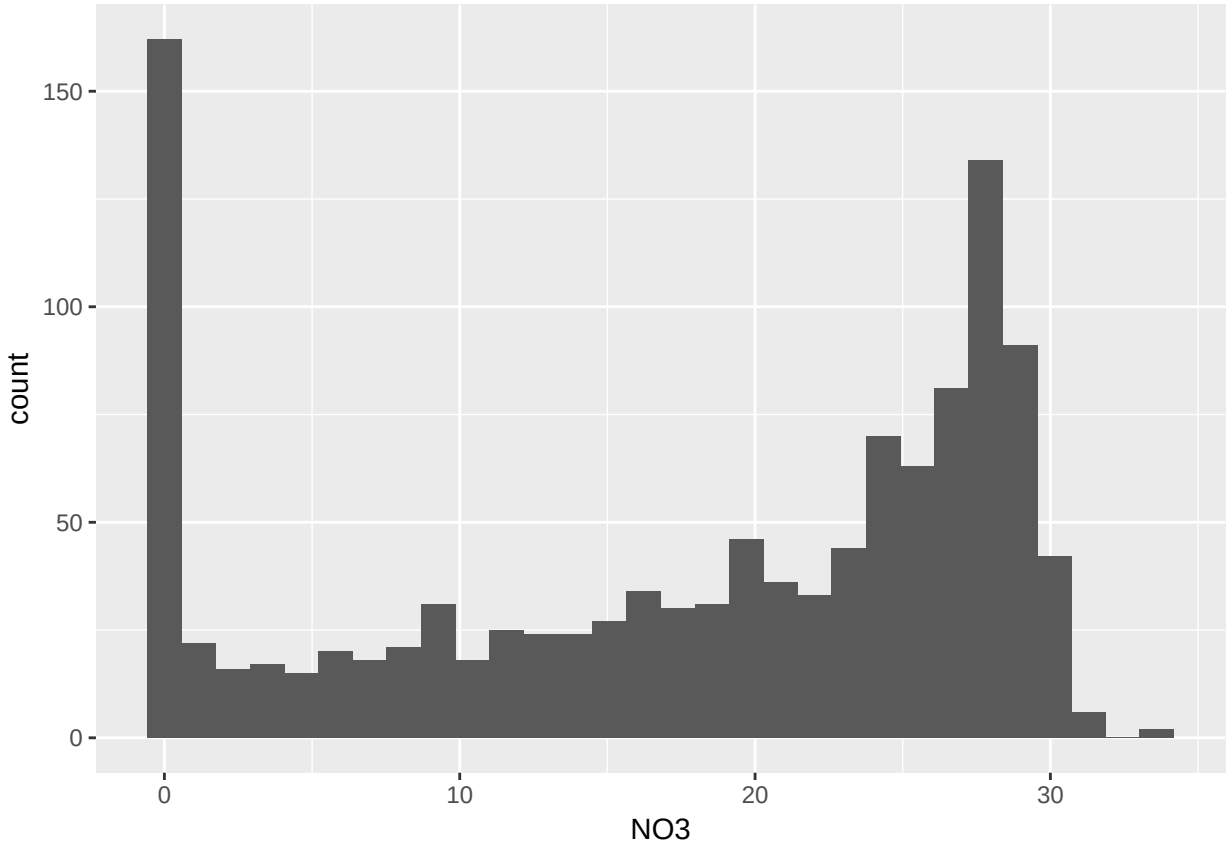
geom_histogram

Exercise. Using your pdat data:

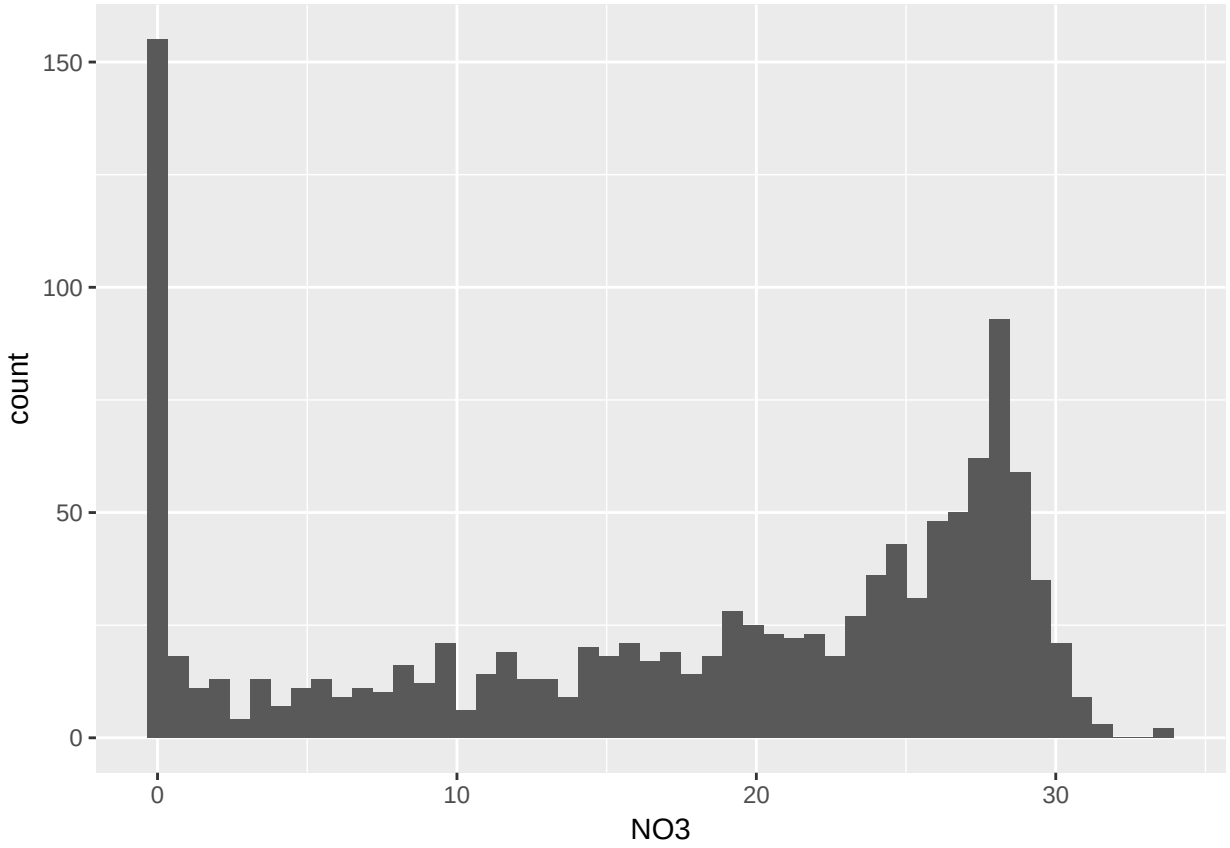
1. Investigate the distribution of nitrate across all depths
2. Test out different values for the bins argument ("bins=")

```
# 1
pdat %>%
  ggplot(aes(x=NO3))+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# 2
pdat %>%
  ggplot(aes(x=NO3)) +
  geom_histogram(bins=50)
```

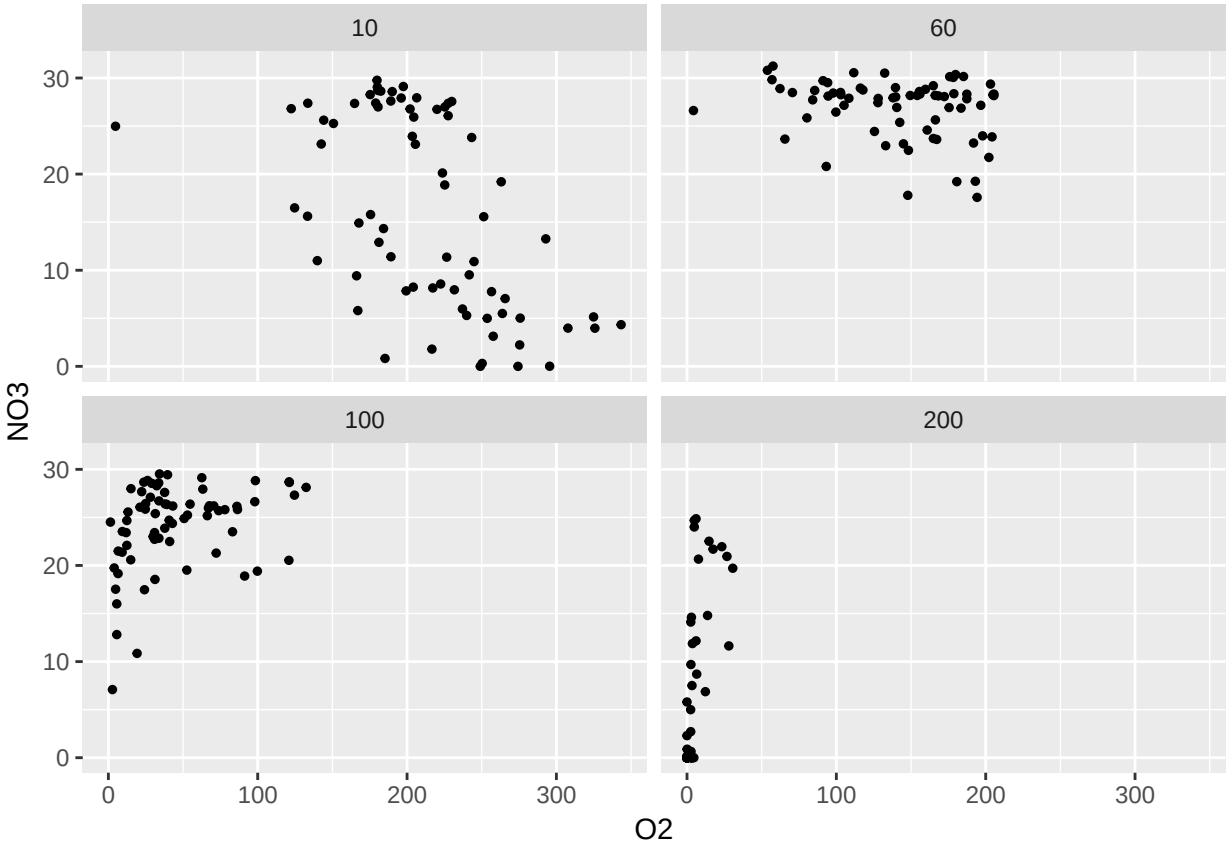


Facets

Exercise. Using your pdat data:

1. Filter to data at depths of 10, 60, 100 or 200
2. Plot Oxygen vs Nitrate faceted by Depth without providing arguments for dir or scales

```
pdat %>%
  filter(Depth %in% c(10, 60, 100, 200)) %>%

  ggplot(aes(x=O2, y=NO3)) +
  geom_point(size=1) +
  facet_wrap(~Depth, ncol=2)
```



Final plot

Setup a theme.

```
my_theme <-
  theme_bw() +
  theme(panel.grid.major=element_blank(),
        panel.grid.minor=element_blank())
```

Save the two plots as objects: p1 and p2.

```
p1 <-
  pdat %>%
  gather(key="Key", value="Value", -Cruise, -Date, -Depth, factor_key=TRUE) %>%

  ggplot(aes(x=Value, y=Depth)) +
  geom_point(size=1) +
  scale_y_reverse(limits=c(200, 0)) +
  facet_wrap(~Key, ncol=2, dir="v", scales="free_x") +
  my_theme +
  labs(x="",
       y="Depth (m)")
```

```
p2 <-
  pdat %>%
  filter(!is.na(H2S)) %>%
  arrange(H2S) %>%

  ggplot(aes(x=O2, y=NO3, colour=H2S)) +
  geom_point(size=2) +
  my_theme +
  labs(x="O2 in uM",
       y="NO3 in uM") +
  scale_colour_continuous(name="H2S in uM")
```

Use cowplot to create a multi-panel figure.

```
p <- cowplot::plot_grid(p1, p2, labels=c("A", "B"), rel_widths=c(2/5, 3/5))

p
```