Wednesday, March 5

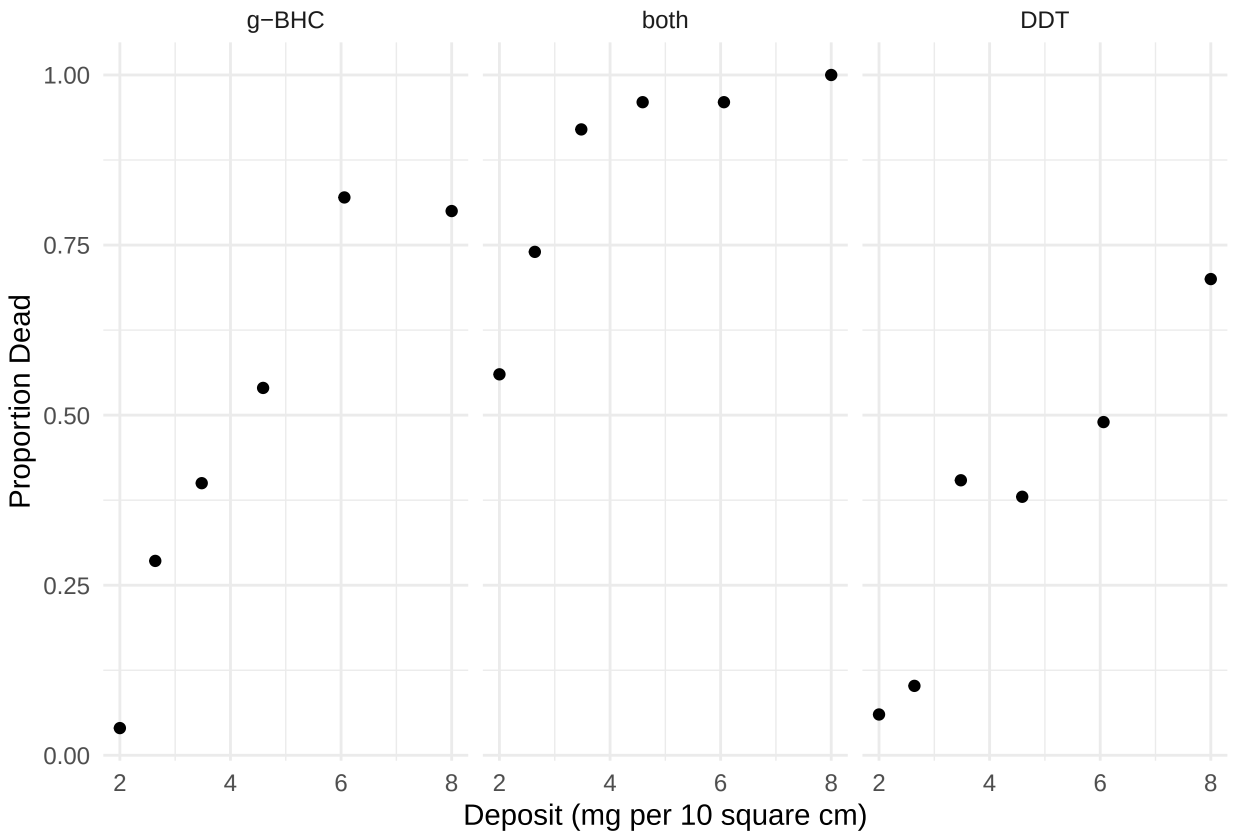
Odds Ratios Examples

Consider the following data from an experiment that investigated the effects of three insecticides on four beetles.

```
library(trtools) # contains the insecticide data frame
insecticide
```

```
   insecticide deposit deaths total
1          DDT    2.00      3    50
2          DDT    2.64      5    49
3          DDT    3.48     19    47
4          DDT    4.59     19    50
5          DDT    6.06     24    49
6          DDT    8.00     35    50
7        g-BHC    2.00      2    50
8        g-BHC    2.64     14    49
9        g-BHC    3.48     20    50
10       g-BHC    4.59     27    50
11       g-BHC    6.06     41    50
12       g-BHC    8.00     40    50
13        both    2.00     28    50
14        both    2.64     37    50
15        both    3.48     46    50
16        both    4.59     48    50
17        both    6.06     48    50
18        both    8.00     50    50
```

```
p <- ggplot(insecticide, aes(x = deposit, y = deaths/total)) +
  geom_point() + facet_wrap(~ insecticide) + theme_minimal() +
  labs(x = "Deposit (mg per 10 square cm)", y = "Proportion Dead")
plot(p)
```

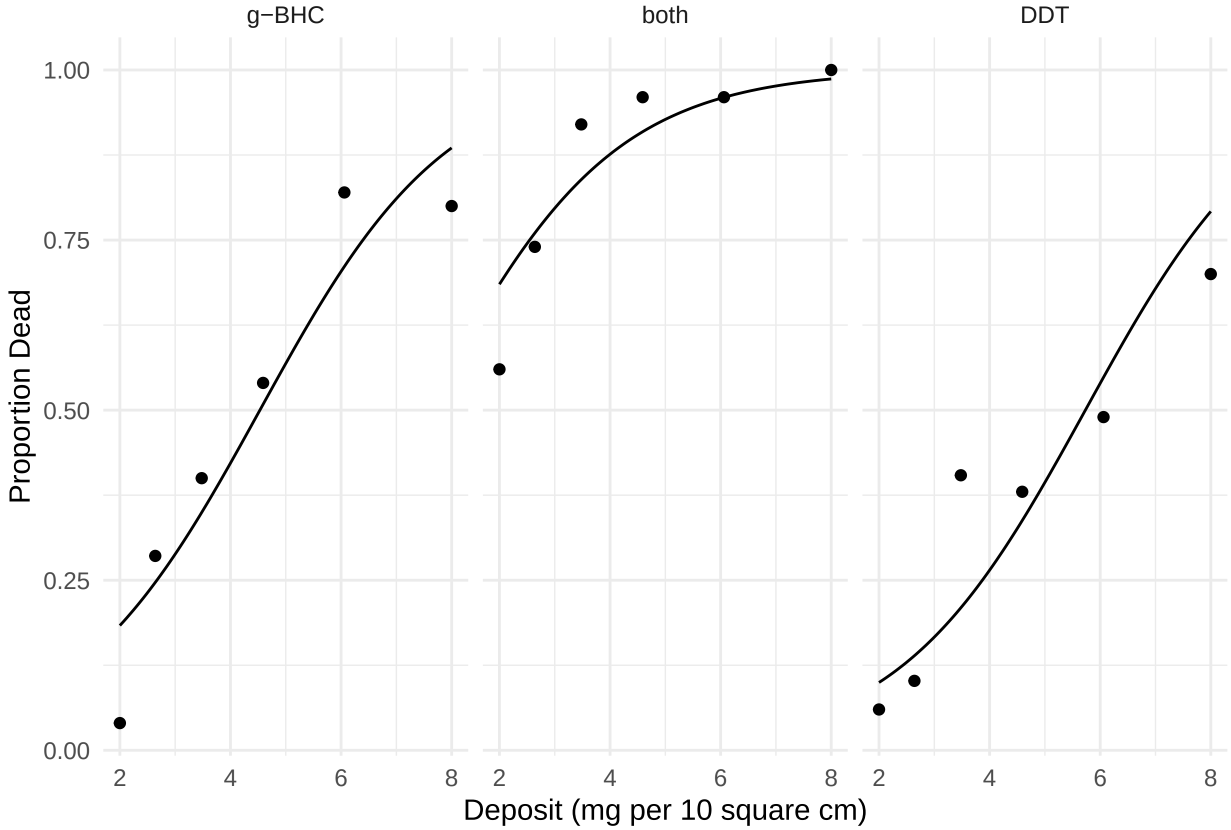


First consider an "additive" logistic regression model (i.e., a model with no interaction).

```r
m <- glm(cbind(deaths, total-deaths) ~ insecticide + deposit,
  family = binomial, data = insecticide)
summary(m)$coefficients
```

```
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       -2.673     0.2497  -10.71 9.55e-27
insecticideboth    2.270     0.2258   10.05 8.84e-24
insecticideDDT    -0.707     0.1973   -3.59 3.36e-04
deposit            0.590     0.0493   11.97 4.94e-33
```

```r
d <- expand.grid(deposit = seq(2, 8, length = 100),
  insecticide = c("DDT","g-BHC","both"))
d$yhat <- predict(m, newdata = d, type = "response")
p <- p + geom_line(aes(y = yhat), data = d)
plot(p)
```



A model for the *odds* of death can be written as

$$O_i = e^{\beta_0} e^{\beta_1 x_{i1}} e^{\beta_2 x_{i2}} e^{\beta_3 x_{i3}}$$

where x_{i1} and x_{i2} are indicator variables for the insecticides `both` and `DDT`, respectively, and x_{i3} is deposit. This can be written case-wise as

$$O_i = \begin{cases} e^{\beta_0} e^{\beta_3 d_i}, & \text{if the } i\text{-th observation of insecticide is g-BHC,} \\ e^{\beta_0} e^{\beta_1} e^{\beta_3 d_i}, & \text{if the } i\text{-th observation of insecticide is both,} \\ e^{\beta_0} e^{\beta_2} e^{\beta_3 d_i}, & \text{if the } i\text{-th observation of insecticide is DDT,} \end{cases}$$

and where $d_i = x_{i3}$ is the deposit. Note that we could also write the model as

$$O_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}),$$

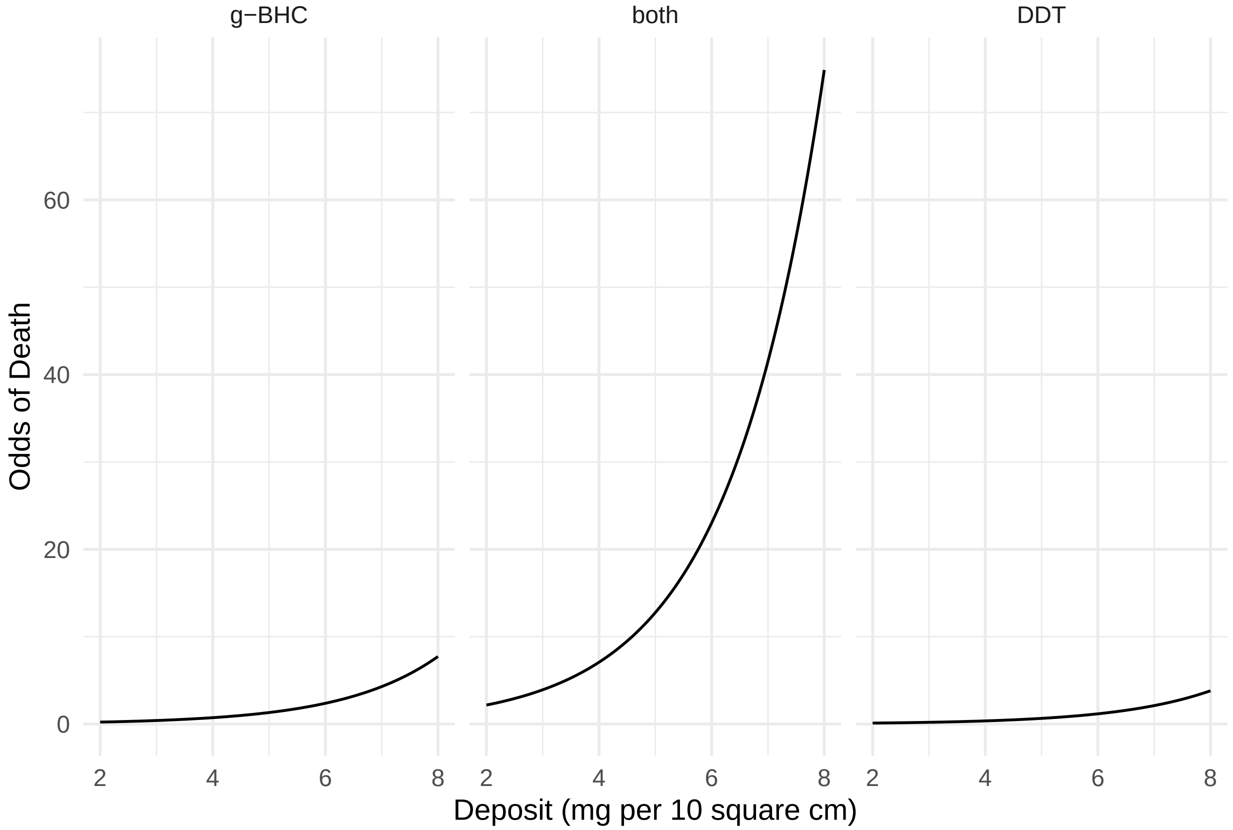
or

$$O_i = \begin{cases} \exp(\beta_0 + \beta_3 d_i), & \text{if the } i\text{-th observation of insecticide is g-BHC,} \\ \exp(\beta_0 + \beta_1 + \beta_3 d_i), & \text{if the } i\text{-th observation of insecticide is both,} \\ \exp(\beta_0 + \beta_2 + \beta_3 d_i), & \text{if the } i\text{-th observation of insecticide is DDT.} \end{cases}$$

We could plot the estimated *odds* of death as a function of deposit and insecticide type.

```r
d <- expand.grid(deposit = seq(2, 8, length = 100),
  insecticide = c("g-BHC","both","DDT"))
d$yhat <- predict(m, newdata = d, type = "response")
d$odds <- d$yhat / (1 - d$yhat) # computing the odds

p <- ggplot(d, aes(x = deposit, y = odds)) +
  geom_line() + facet_wrap(~ insecticide) + theme_minimal() +
  labs(x = "Deposit (mg per 10 square cm)", y = "Odds of Death")
plot(p)
```



It can be shown that the odds ratio for a one unit increase in deposit is e^{β_3} (regardless of insecticide used), and the odds ratio for comparing `both` with `g-BHC` is e^{β_1} (regardless of deposit amount). We can get these odds ratios as follows.

```
exp(cbind(coef(m), confint(m)))
```

```
                       2.5 % 97.5 %
(Intercept)     0.069 0.0418  0.111
insecticideboth 9.683 6.2753 15.225
insecticideDDT  0.493 0.3336  0.724
deposit         1.804 1.6419  1.992
```

But using `contrast` allows us to do this without having to figure out the parameterization.

```
# estimate the odds ratio for dose (one unit increase)
trtools::contrast(m,
  a = list(deposit = 3, insecticide = c("DDT","g-BHC","both")),
  b = list(deposit = 2, insecticide = c("DDT","g-BHC","both")),
  cnames = c("DDT","g-BHC","both"), tf = exp)
```

```
      estimate lower upper
DDT        1.8  1.64  1.99
g-BHC      1.8  1.64  1.99
both       1.8  1.64  1.99
```

```
# estimate the odds ratio for type of insecticide (both versus DDT)
trtools::contrast(m,
  a = list(deposit = c(2,5,8), insecticide = "both"),
  b = list(deposit = c(2,5,8), insecticide = "g-BHC"),
  cnames = c("2mg","5mg","8mg"), tf = exp)
```

```
    estimate lower upper
2mg     9.68  6.22  15.1
5mg     9.68  6.22  15.1
8mg     9.68  6.22  15.1
```

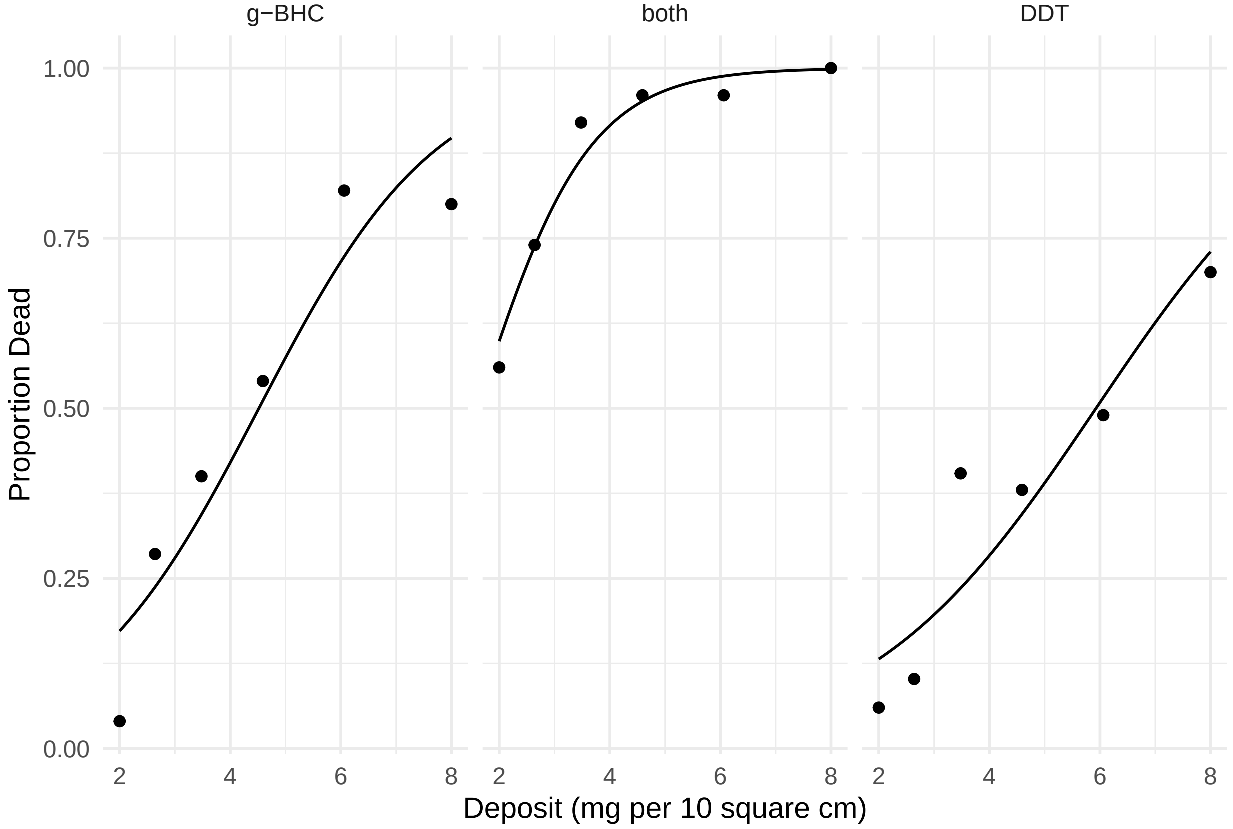
Now suppose we include an interaction between dose and type of insecticide.

```
m.int <- glm(cbind(deaths, total-deaths) ~ insecticide * deposit,
  family = binomial, data = insecticide)
summary(m.int)$coefficients
```

```
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -2.8109     0.3585 -7.8418 4.44e-15
insecticideboth           1.2258     0.6718  1.8247 6.80e-02
insecticideDDT           -0.0389     0.5072 -0.0768 9.39e-01
deposit                   0.6221     0.0779  7.9899 1.35e-15
insecticideboth:deposit   0.3701     0.2090  1.7711 7.65e-02
insecticideDDT:deposit   -0.1414     0.1038 -1.3630 1.73e-01
```

```
d <- expand.grid(deposit = seq(2, 8, length = 100),
  insecticide = c("DDT","g-BHC","both"))
d$yhat <- predict(m.int, newdata = d, type = "response")

p <- ggplot(insecticide, aes(x = deposit, y = deaths/total)) +
  geom_point() + facet_wrap(~ insecticide) + theme_minimal() +
  labs(x = "Deposit (mg per 10 square cm)", y = "Proportion Dead") +
  geom_line(aes(y = yhat), data = d)
plot(p)
```



```
# estimate the odds ratio for the effect of dose
trtools::contrast(m.int,
  a = list(deposit = 3, insecticide = c("DDT","g-BHC","both")),
  b = list(deposit = 2, insecticide = c("DDT","g-BHC","both")),
  cnames = c("DDT","g-BHC","both"), tf = exp)
```

```
      estimate lower upper
DDT       1.62  1.41  1.85
g-BHC     1.86  1.60  2.17
both      2.70  1.84  3.94
```

```
# estimate the odds ratio for the effect of type of insecticide (both versus g-BHC)
trtools::contrast(m.int,
  a = list(deposit = c(2,5,8), insecticide = "both"),
  b = list(deposit = c(2,5,8), insecticide = "g-BHC"),
  cnames = c("2mg","5mg","8mg"), tf = exp)
```

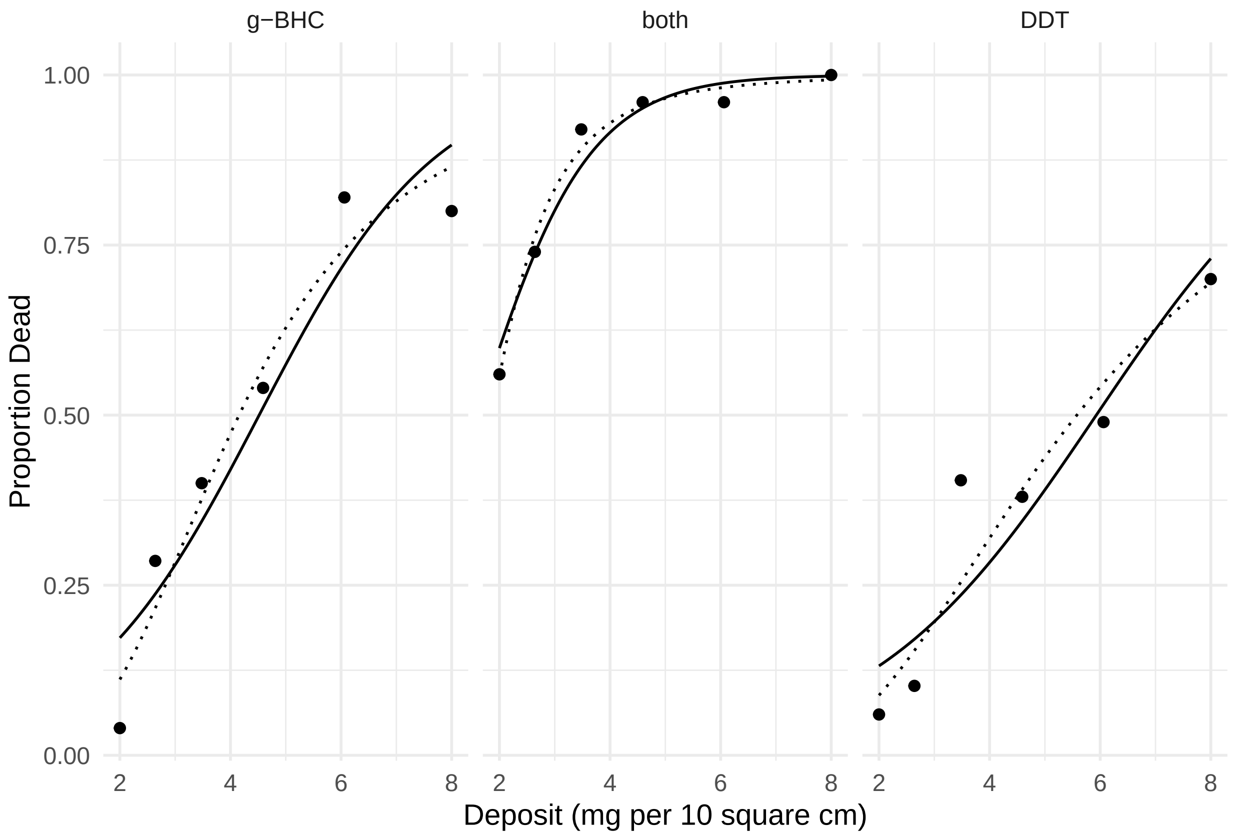
```
    estimate lower upper
2mg     7.14  3.79  13.5
5mg    21.68  8.29  56.7
8mg    65.80  7.96 544.1
```

Now consider a model where we use log transformation of dose.

```
m <- glm(cbind(deaths, total-deaths) ~ insecticide * log(deposit),
  family = binomial, data = insecticide)
summary(m)$coefficients
```

```
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -4.043      0.497  -8.131 4.27e-16
insecticideboth                 1.922      0.772   2.489 1.28e-02
insecticideDDT                  0.128      0.712   0.180 8.57e-01
log(deposit)                    2.838      0.339   8.367 5.93e-17
insecticideboth:log(deposit)    0.550      0.666   0.826 4.09e-01
insecticideDDT:log(deposit)    -0.560      0.468  -1.197 2.31e-01
```

```
d <- expand.grid(deposit = seq(2, 8, length = 100),
  insecticide = c("DDT","g-BHC","both"))
d$yhat <- predict(m, newdata = d, type = "response")
p <- p + geom_line(aes(y = yhat), data = d, linetype = 3)
plot(p)
```



Now the odds ratio shows the effect of *doubling* the dose.

```
# odds ratio for the effect of increasing dose from 1 to 2 (doubling)
trtools::contrast(m,
  a = list(deposit = 2, insecticide = c("DDT","g-BHC","both")),
  b = list(deposit = 1, insecticide = c("DDT","g-BHC","both")),
  cnames = c("DDT","g-BHC","both"), tf = exp)
```

```
      estimate lower upper
DDT       4.85  3.13  7.51
g-BHC     7.15  4.51 11.34
both     10.47  4.81 22.82
```

```
# odds ratio for the effect of increasing dose from 2 to 4 (doubling)
trtools::contrast(m,
  a = list(deposit = 4, insecticide = c("DDT","g-BHC","both")),
  b = list(deposit = 2, insecticide = c("DDT","g-BHC","both")),
  cnames = c("DDT","g-BHC","both"), tf = exp)
```

```
      estimate lower upper
DDT       4.85  3.13  7.51
g-BHC     7.15  4.51 11.34
both     10.47  4.81 22.82
```

```
# odds ratio for the effect of increasing dose from 2 to 3 (not doubling)
trtools::contrast(m,
  a = list(deposit = 3, insecticide = c("DDT","g-BHC","both")),
  b = list(deposit = 2, insecticide = c("DDT","g-BHC","both")),
  cnames = c("DDT","g-BHC","both"), tf = exp)
```

```
      estimate lower upper
DDT       2.52  1.95  3.25
g-BHC     3.16  2.41  4.14
both      3.95  2.50  6.23
```

Contrasts between insecticides can proceed in the usual way although the results are not quite the same as when we did not transform dose.

```
# odds ratio to compare two insecticides at three doses
trtools::contrast(m,
  a = list(deposit = c(2,5,8), insecticide = "both"),
  b = list(deposit = c(2,5,8), insecticide = "g-BHC"),
  cnames = c("2mg","5mg","8mg"), tf = exp)
```

```
    estimate lower upper
2mg     10.0  4.83  20.8
5mg     16.6  7.09  38.8
8mg     21.5  5.35  86.1
```

At some point we will want to visit the issue of how to evaluate/select models.

Binary/Bernoulli Logistic Regression Example

Consider the following data from a study that investigated the relationship between vasoconstriction and the rate and volume of air breathed by human subjects. Here the response variable is *binary* and thus has a *Bernoulli* distribution (a special case of the binomial distribution).

```
library(catdata)
data(vaso)
head(vaso)
```

```
     vol    rate vaso
1  1.308 -0.1924    1
2  1.253  0.0862    1
3  0.223  0.9163    1
4 -0.288  0.4055    1
5 -0.223  1.1632    1
6 -0.357  1.2528    1
```

Volume (`vol`) is the logarithm of volume in liters, and rate (`rate`) is the logarithm of liters per second. For this example I am going to transform these variables to the volume and rate in deciliters.

```
vaso$dvolume <- exp(vaso$vol)*10       # transform to deciliters
vaso$drate <- exp(vaso$rate)*10        # transform to deciliters per sec
```

I am also going to create a couple different versions of the response variable: one that is a character for plotting and one that is binary for modeling (note that the help file for `vaso` has coding on the `vaso` variable backwards).

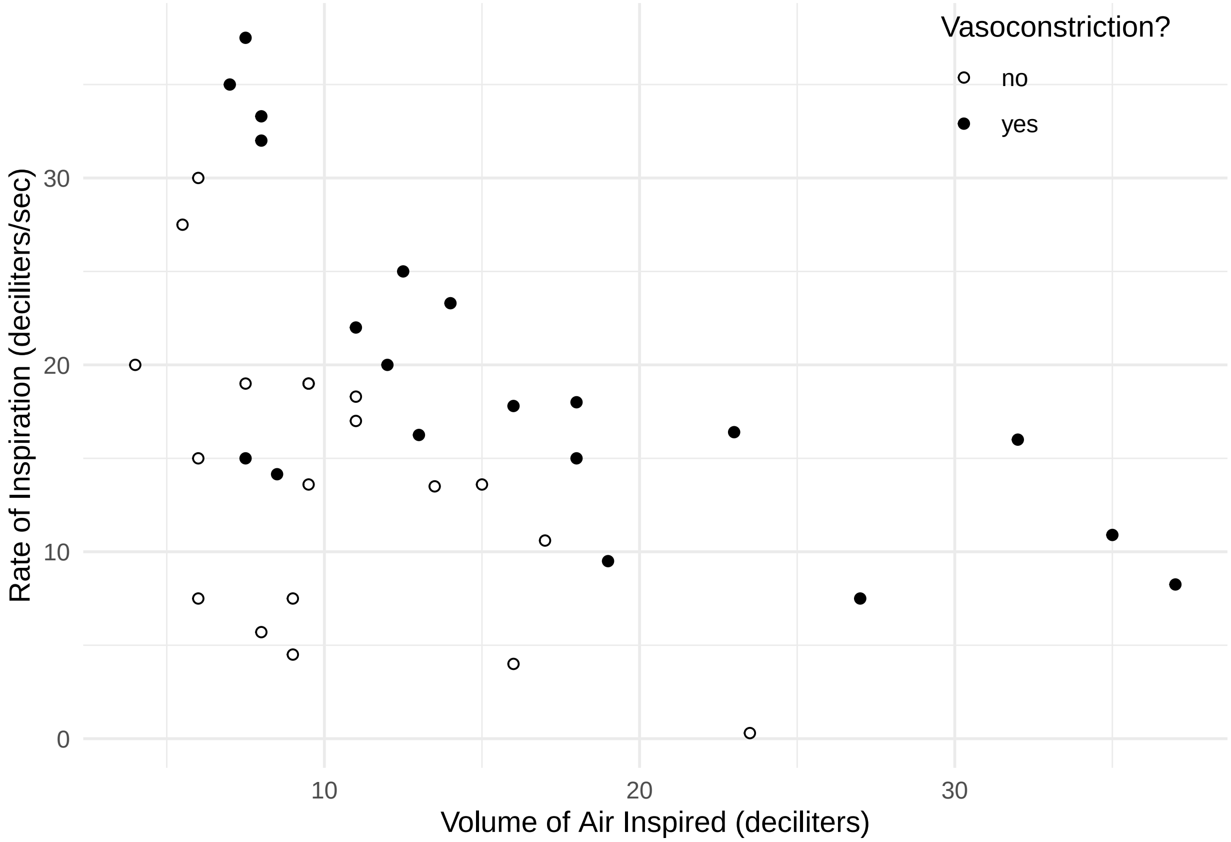
```
vaso$vasoconstriction <- ifelse(vaso$vaso == 1, "yes", "no")
vaso$y <- ifelse(vaso$vaso == 1, 1, 0) # create binary response
head(vaso)
```

```
     vol    rate vaso dvolume drate vasoconstriction y
1  1.308 -0.1924    1    37.0  8.25              yes 1
2  1.253  0.0862    1    35.0 10.90              yes 1
3  0.223  0.9163    1    12.5 25.00              yes 1
4 -0.288  0.4055    1     7.5 15.00              yes 1
5 -0.223  1.1632    1     8.0 32.00              yes 1
```

```
6 -0.357  1.2528    1     7.0 35.00              yes 1
```

Here is a scatterplot of volume and rate, with point color indicating vasoconstriction.

```r
p <- ggplot(vaso, aes(x = dvolume, y = drate)) +
  geom_point(aes(fill = vasoconstriction), shape = 21) +
  scale_fill_manual(values = c("white","black")) +
  labs(x = "Volume of Air Inspired (deciliters)",
    y = "Rate of Inspiration (deciliters/sec)",
    fill = "Vasoconstriction?") +
  theme_minimal() +
  theme(legend.position = "inside", legend.position.inside = c(0.85, 0.9))
plot(p)
```



If the response variable is *binary* (i.e., 0 or 1) then we can use `glm(y ~ ...)` rather than `glm(cbind(y, 1-y) ~ ...)`.

```r
m <- glm(y ~ dvolume + drate, family = binomial, data = vaso)
cbind(summary(m)$coefficients, confint(m))
```

```
            Estimate Std. Error z value Pr(>|z|)   2.5 % 97.5 %
(Intercept)   -9.530     3.2332   -2.95  0.00320 -17.559 -4.456
dvolume        0.388     0.1429    2.72  0.00658   0.165  0.739
drate          0.265     0.0914    2.90  0.00376   0.118  0.490
```

```r
exp(cbind(coef(m), confint(m)))
```

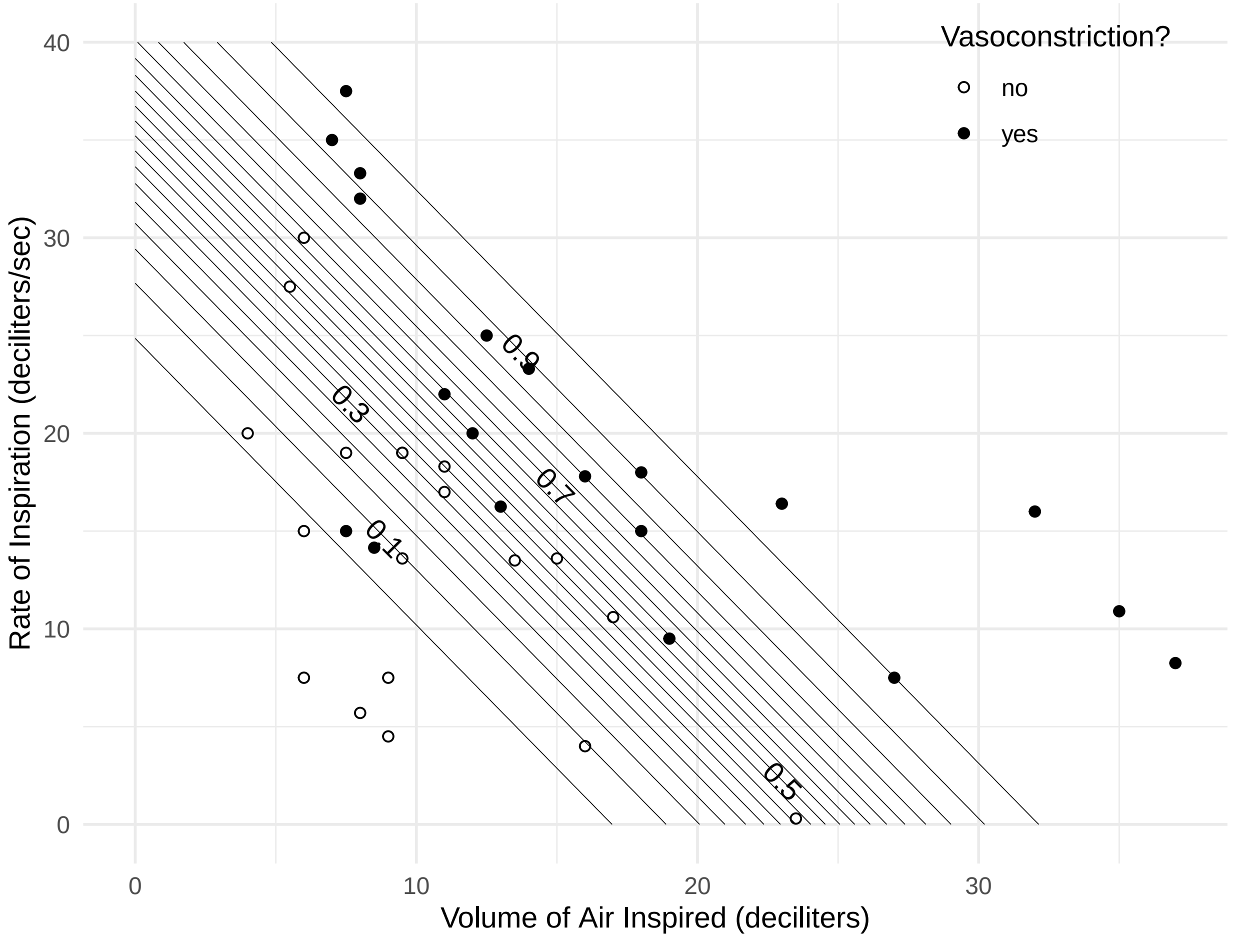
```
                       2.5 % 97.5 %
(Intercept) 7.27e-05 2.37e-08 0.0116
dvolume     1.47e+00 1.18e+00 2.0928
```

```
drate       1.30e+00 1.12e+00 1.6315
```

```
d <- expand.grid(dvolume = seq(0, 40, length = 100),
  drate = seq(0, 40, length = 100))
d$yhat <- predict(m, newdata = d, type = "response")

library(metR) # for geom_text_contour

p <- ggplot(vaso, aes(x = dvolume, y = drate)) +
  geom_point(aes(fill = vasoconstriction), shape = 21) +
  scale_fill_manual(values = c("white","black")) +
  geom_contour(aes(z = yhat), data = d, color = "black",
    linewidth = 0.15, breaks = seq(0.05, 0.95, by = 0.05)) +
  geom_text_contour(aes(z = yhat), data = d) +
  labs(x = "Volume of Air Inspired (deciliters)",
    y = "Rate of Inspiration (deciliters/sec)",
    fill = "Vasoconstriction?") +
  theme_minimal() +
  theme(legend.position = "inside", legend.position.inside = c(0.85, 0.9))
plot(p)
```



```
# odds ratio for the effect of volume
trtools::contrast(m, tf = exp,
  a = list(dvolume = 2, drate = c(10,20,30)),
  b = list(dvolume = 1, drate = c(10,20,30)),
```

```
  cnames = c(paste("at", c(10,20,30), "dl/sec")))
```

```
             estimate lower upper
at 10 dl/sec     1.47  1.11  1.95
at 20 dl/sec     1.47  1.11  1.95
at 30 dl/sec     1.47  1.11  1.95
```

```
# odds ratios for rate
trtools::contrast(m, tf = exp,
  a = list(drate = 2, dvolume = c(10,20,30)),
  b = list(drate = 1, dvolume = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl")))
```

```
         estimate lower upper
at 10 dl      1.3  1.09  1.56
at 20 dl      1.3  1.09  1.56
at 30 dl      1.3  1.09  1.56
```

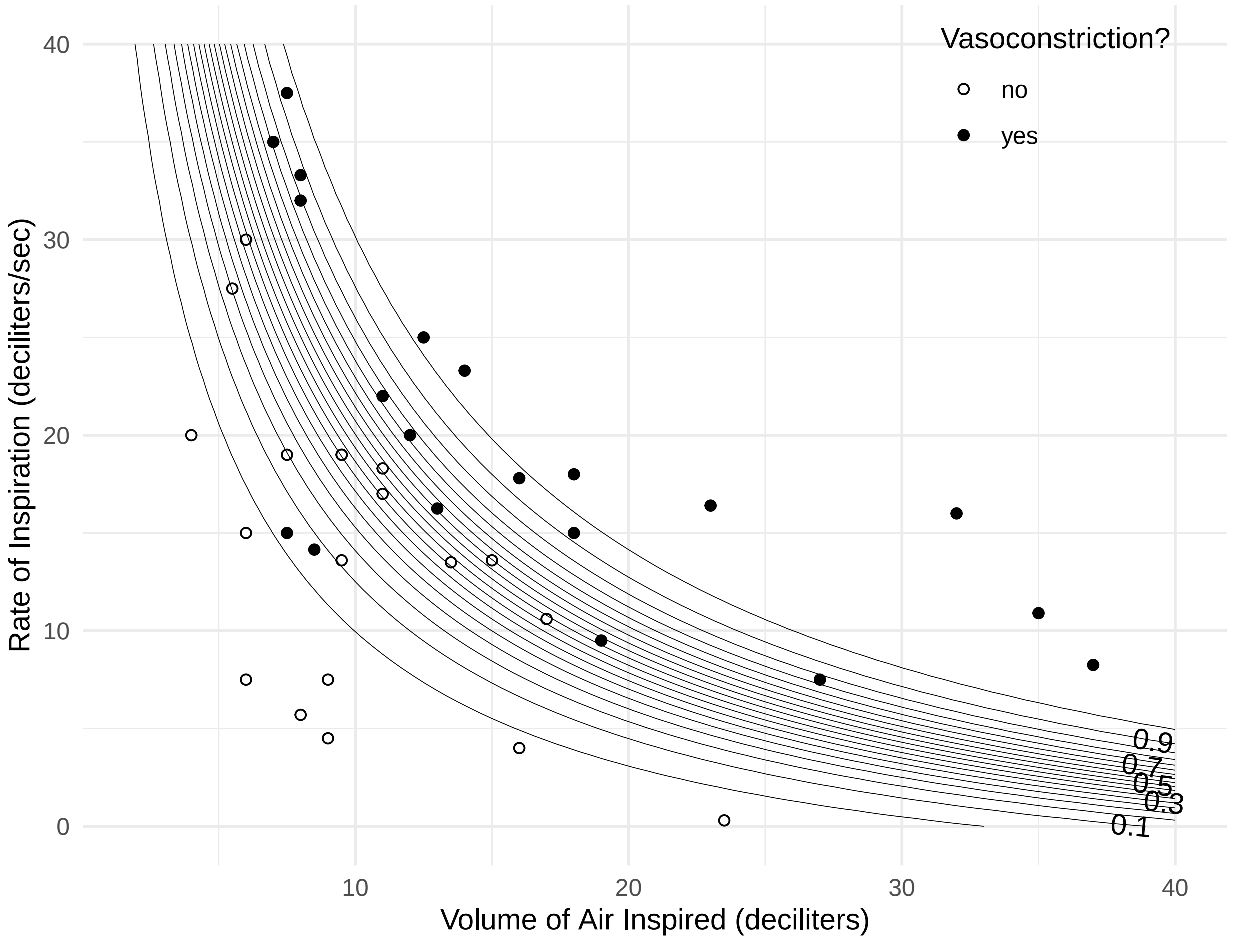
Now consider a model with a product term (i.e., “interaction”) for volume and rate.

```
m <- glm(y ~ dvolume + drate + dvolume*drate, family = binomial, data = vaso)
summary(m)$coefficients
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -7.1150     3.3485  -2.125   0.0336
dvolume         0.1264     0.2147   0.589   0.5561
drate           0.0511     0.1508   0.339   0.7346
dvolume:drate   0.0241     0.0166   1.449   0.1473
```

```
d <- expand.grid(dvolume = seq(0, 40, length = 100), drate = seq(0, 40, length = 100))

d$yhat <- predict(m, newdata = d, type = "response")
p <- ggplot(vaso, aes(x = dvolume, y = drate)) +
  geom_point(aes(fill = vasoconstriction), shape = 21) +
  scale_fill_manual(values = c("white","black")) +
  geom_contour(aes(z = yhat), data = d, color = "black",
    linewidth = 0.15, breaks = seq(0.05, 0.95, by = 0.05)) +
  geom_text_contour(aes(z = yhat), data = d) +
  labs(x = "Volume of Air Inspired (deciliters)",
    y = "Rate of Inspiration (deciliters/sec)",
    fill = "Vasoconstriction?") +
  theme_minimal() +
  theme(legend.position = "inside", legend.position.inside = c(0.85, 0.9))
plot(p)
```



```r
# odds ratios for the effect of volume
trtools::contrast(m, tf = exp,
  a = list(dvolume = 2, drate = c(10,20,30)),
  b = list(dvolume = 1, drate = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl/sec")))
```

```
             estimate lower upper
at 10 dl/sec     1.44  1.09  1.92
at 20 dl/sec     1.84  1.18  2.86
at 30 dl/sec     2.34  1.13  4.82
```

```r
# odds ratios for the effect of rate
trtools::contrast(m, tf = exp,
  a = list(drate = 2, dvolume = c(10,20,30)),
  b = list(drate = 1, dvolume = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl")))
```

```
         estimate lower upper
at 10 dl     1.34  1.10  1.64
at 20 dl     1.70  1.08  2.68
at 30 dl     2.17  1.01  4.65
```

Now about a model where we transform volume and rate to make it additive on the log scale?

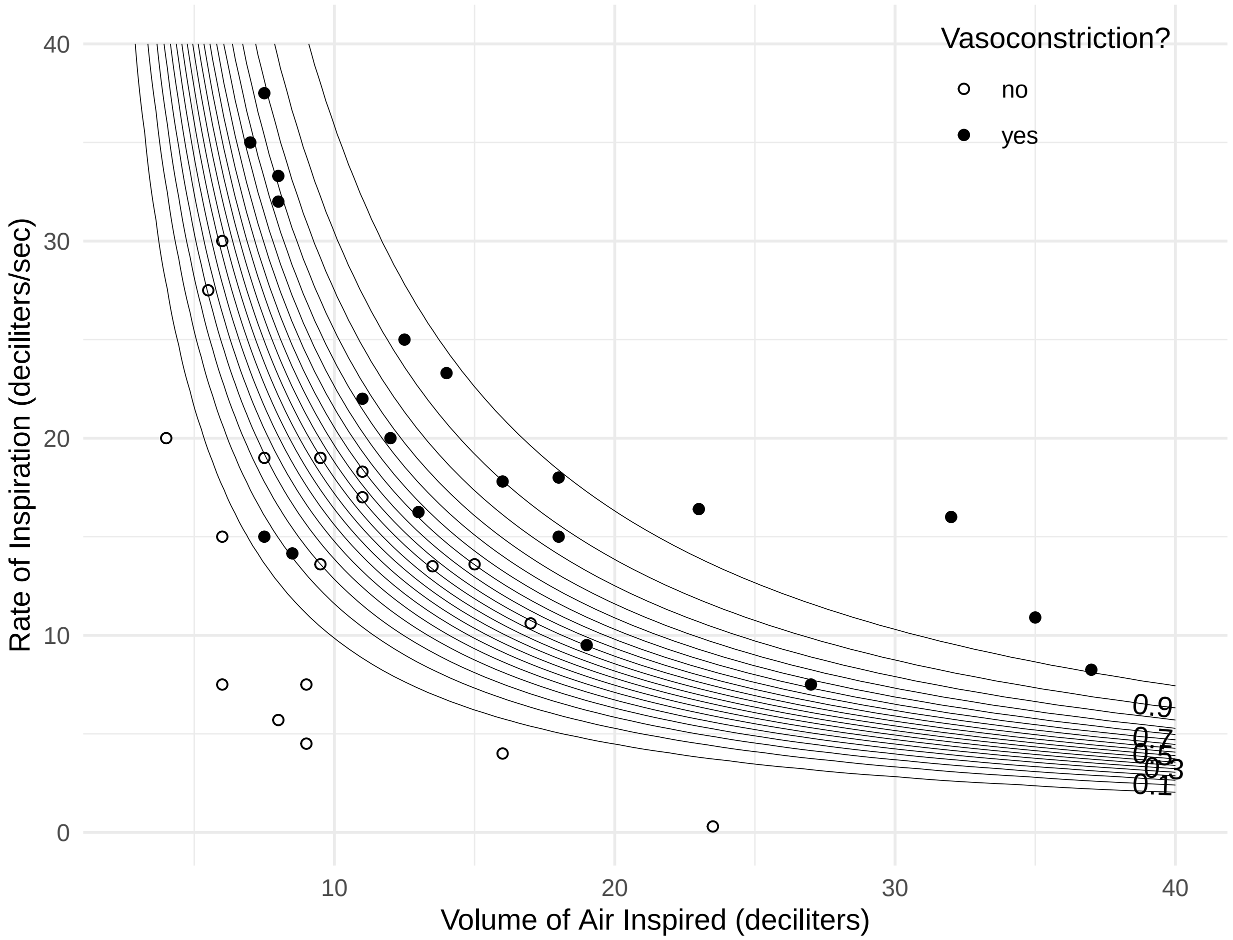
```r
m <- glm(y ~ log(dvolume) + log(drate), family = binomial, data = vaso)
exp(cbind(coef(m), confint(m)))
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
                         2.5 %   97.5 %
(Intercept)  1.02e-11 8.14e-22 1.44e-05
log(dvolume) 1.78e+02 9.91e+00 1.90e+04
log(drate)   9.57e+01 5.54e+00 8.40e+03
```

```r
d <- expand.grid(dvolume = seq(0, 40, length = 100),
  drate = seq(0, 40, length = 100))
d$yhat <- predict(m, newdata = d, type = "response")

p <- ggplot(vaso, aes(x = dvolume, y = drate)) +
  geom_point(aes(fill = vasoconstriction), shape = 21) +
  scale_fill_manual(values = c("white","black")) +
  geom_contour(aes(z = yhat), data = d, color = "black",
    linewidth = 0.15, breaks = seq(0.05, 0.95, by = 0.05)) +
  geom_text_contour(aes(z = yhat), data = d) +
  labs(x = "Volume of Air Inspired (deciliters)",
    y = "Rate of Inspiration (deciliters/sec)",
    fill = "Vasoconstriction?") +
  theme_minimal() +
  theme(legend.position = "inside", legend.position.inside = c(0.85, 0.9))
plot(p)
```



```
# odds ratios for the effect of volume
trtools::contrast(m, tf = exp,
  a = list(dvolume = 2, drate = c(10,20,30)),
  b = list(dvolume = 1, drate = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl/sec")))
```

```
             estimate lower upper
at 10 dl/sec     36.2  2.88   456
at 20 dl/sec     36.2  2.88   456
at 30 dl/sec     36.2  2.88   456
```

```
# odds ratios for the effect of rate
trtools::contrast(m, tf = exp,
  a = list(drate = 2, dvolume = c(10,20,30)),
  b = list(drate = 1, dvolume = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl")))
```

```
         estimate lower upper
at 10 dl     23.6  1.94   287
at 20 dl     23.6  1.94   287
at 30 dl     23.6  1.94   287
```

Doubling the volume or rate is a relatively large change. How about increasing it by only, say, 10% rather than 100%?

```
# odds ratios for the effect of volume
trtools::contrast(m, tf = exp,
  a = list(dvolume = 1.1, drate = c(10,20,30)),
  b = list(dvolume = 1.0, drate = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl/sec")))
```

```
             estimate lower upper
at 10 dl/sec     1.64  1.16  2.32
at 20 dl/sec     1.64  1.16  2.32
at 30 dl/sec     1.64  1.16  2.32
```

```
# odds ratios for the effect of rate
trtools::contrast(m, tf = exp,
  a = list(drate = 1.1, dvolume = c(10,20,30)),
  b = list(drate = 1.0, dvolume = c(10,20,30)),
  cnames = c(paste("at", c(10,20,30), "dl")))
```

```
         estimate lower upper
at 10 dl     1.54   1.1  2.18
at 20 dl     1.54   1.1  2.18
at 30 dl     1.54   1.1  2.18
```

Note that we'd get the same results for *any* 10% increase in volume or rate (e.g., from 2.0 to 2.2) because both are on the log scale.