ica_300: Graph Metric & RAVLT

```
library(tidyverse)
library(rstatix)
#library(ggpubr)
#library(car)
library(MASS)
```

Load data

```
data <- read.csv("../data/flat_nihtb-graph_metrics.csv")
```

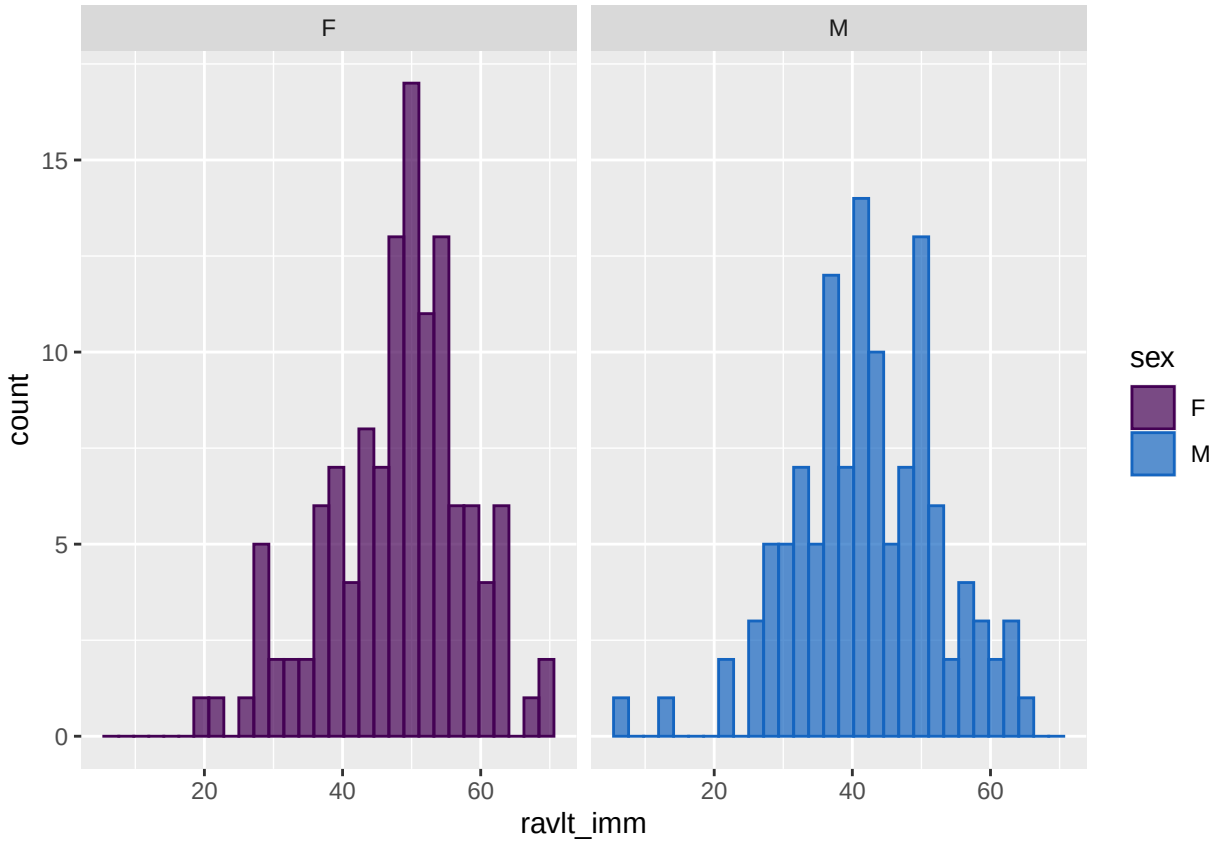
RAVLT

The RAVLT is a neuropsychological measure of working memory. A list of 15 words is presented across 5 trials. The list is read aloud and participants are asked to recall as many words as they heard. After trials 1-5, a new list of 15 words (List B) is presented, and then the participant is asked to recall those words. On Trial 6, the participant is asked to recall as many words from the first list. 30 Minutes after the completion of recall on List B, the participant is asked to again recall the words from List A.

RAVLT Immediate Learning

Immediate learning is scored as the sum from the scores of trials 1-5.

```
colors = c("#440154FF","#1565c0")
ggplot(data, aes(x = ravlt_imm, color = sex, fill = sex)) +
  geom_histogram(alpha = 0.7) +
  scale_color_manual(values = colors) +
  scale_fill_manual(values = colors) +
  facet_grid(col = vars(sex))
```



Mean and Variance of Immediate Learning

Mean does not equal variance which violates the assumption of poisson regression. Modeling with poisson also results in Variance larger than the mean (overdispersion). Overdispersion is also indicated when residual deviance is substantially larger than the degrees of freedom. Models rely on negative binomial GLMs. In negative binomial regression, the coefficients are interpreted as for every unit increase in the independent variable, the expected log count of the dependent variable will change by the value of the coefficient. For binary variables, the coefficients are interpreted in reference to the baseline variable. These can be interpreted as the expected log counts for variable A is X units higher/lower than in the expected log counts for the reference group. The constant tells us what is the expected log counts if all predictors are equal to zero.

```
data %>%
  summarise(count = n(),
            mean = mean(ravlt_imm, na.rm = TRUE),
            variance = var(ravlt_imm, na.rm = TRUE))
```

```
##   count     mean variance
## 1   255 45.22634 111.4486
```

Outliers using IQR method

Two outliers at the low end of the distribution have been identified.

```r
data %>%
  dplyr::select(., subject, ravlt_imm) %>%
  identify_outliers("ravlt_imm")
```

```
##            subject ravlt_imm is.outlier is.extreme
## 1 NDAR_INV7Y8BVHEX         7       TRUE      FALSE
## 2 NDAR_INVMN01YD5J        13       TRUE      FALSE
```

Variance is still larger than mean even after removing outliers

```r
data %>%
  filter(ravlt_imm > 13) %>%
  summarise(count = n(),
            mean = mean(ravlt_imm, na.rm = TRUE),
            variance = var(ravlt_imm, na.rm = TRUE))
```

```
##   count     mean variance
## 1   241 45.51867 101.8757
```

```r
data <- data %>%
  filter(ravlt_imm > 13)
```

Break down of participant sex

A Chi-square test of proportion reveals that the number of females in the sample is not significantly different than the number of males.

```r
count(data, vars = sex)
```

```
##   vars   n
## 1    F 125
## 2    M 116
```

```r
# Chi Square test of proportion
prop.test(x = 125, n = 241, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  125 out of 241, null probability 0.5
## X-squared = 0.26556, df = 1, p-value = 0.6063
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4537421 0.5829973
## sample estimates:
##         p
## 0.5186722
```

Modeling: Modularity

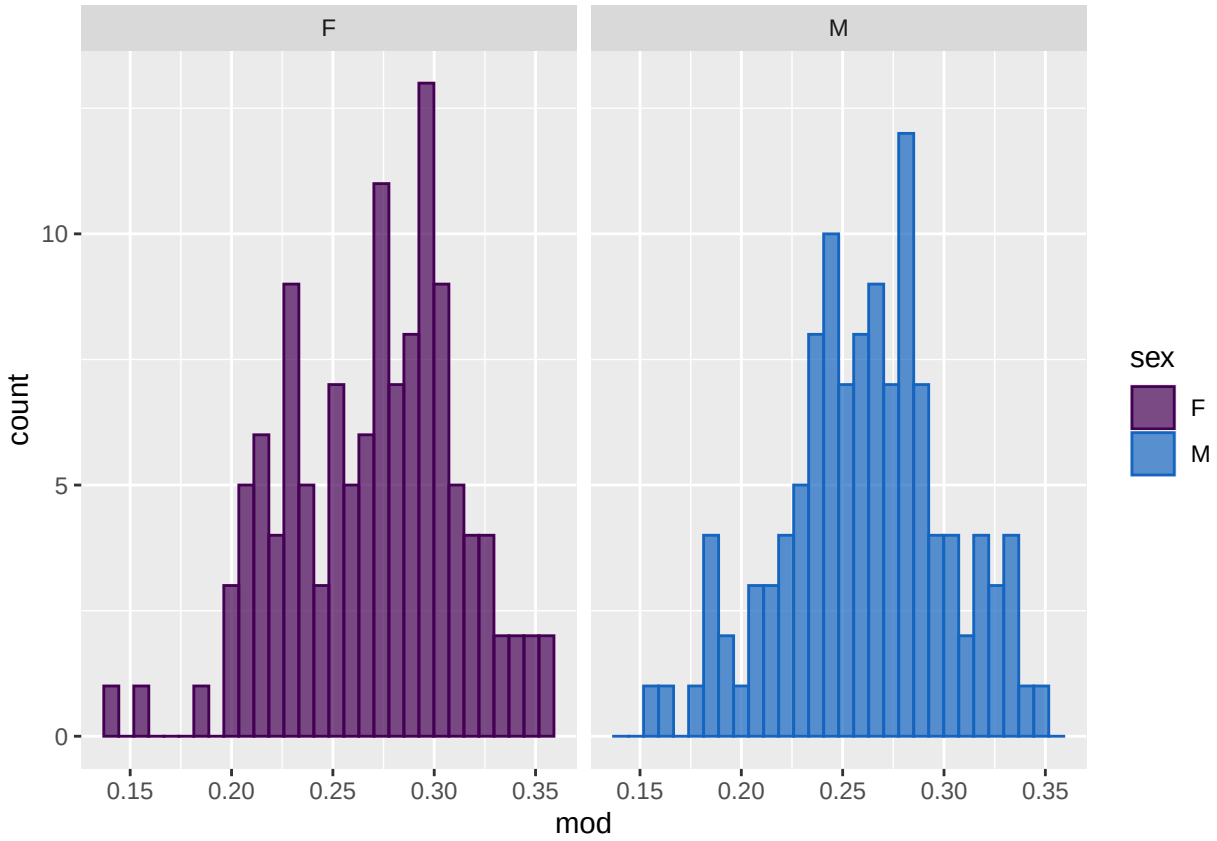
```
# Modularity
plot_and_model(data, "mod", "ravlt_imm")
```

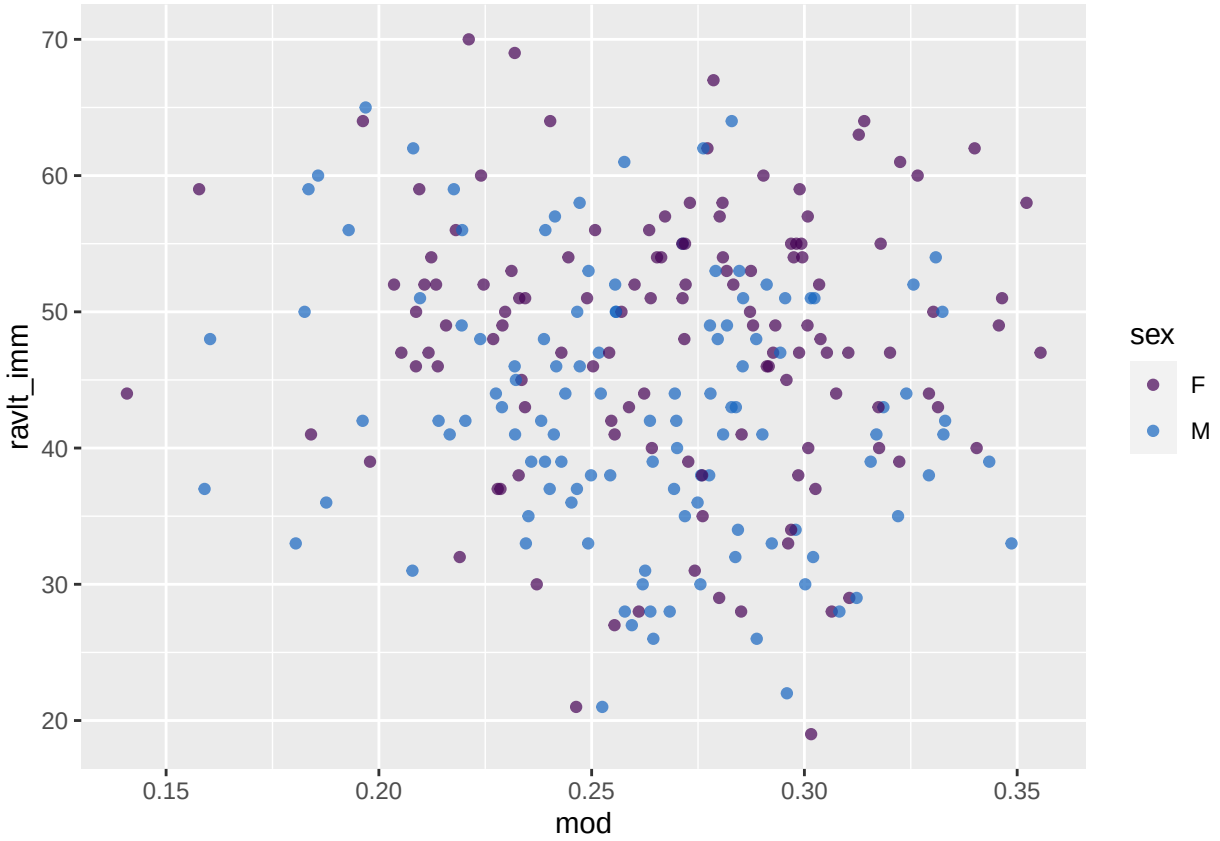
```
##
## Call:
## glm.nb(formula = f, data = arg_1, init.theta = 34.79224122, link = log)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.08998  -0.70471   0.03665   0.66802   2.02012
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.93627    0.09438  41.707   <2e-16 ***
## mod         -0.44556    0.35163  -1.267    0.205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(34.7922) family taken to be 1)
##
##     Null deviance: 248.19  on 240  degrees of freedom
## Residual deviance: 246.56  on 239  degrees of freedom
## AIC: 1810.4
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  34.79
##           Std. Err.:  5.71
##
##  2 x log-likelihood:  -1804.355
##
## [1] "Confidence Intervals:"

## Waiting for profiling to be done...

##                Estimate     2.5 %   97.5 %
## (Intercept)  3.9362677  3.752808 4.119932
## mod         -0.4455593 -1.129520 0.238032
##
## [1] "Incident Ratios:"
##               Estimate      2.5 %   97.5 %
## (Intercept) 51.227047 42.6406350 61.55508
## mod          0.640466  0.3231884  1.26875

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


F
M
count
10
5
0
0.15
0.20
0.25
0.30
0.35
mod
sex
F
M



Modeling: Clustering Coefficient

```
# Clustering
plot_and_model(data, "clust", "ravlt_imm")
```

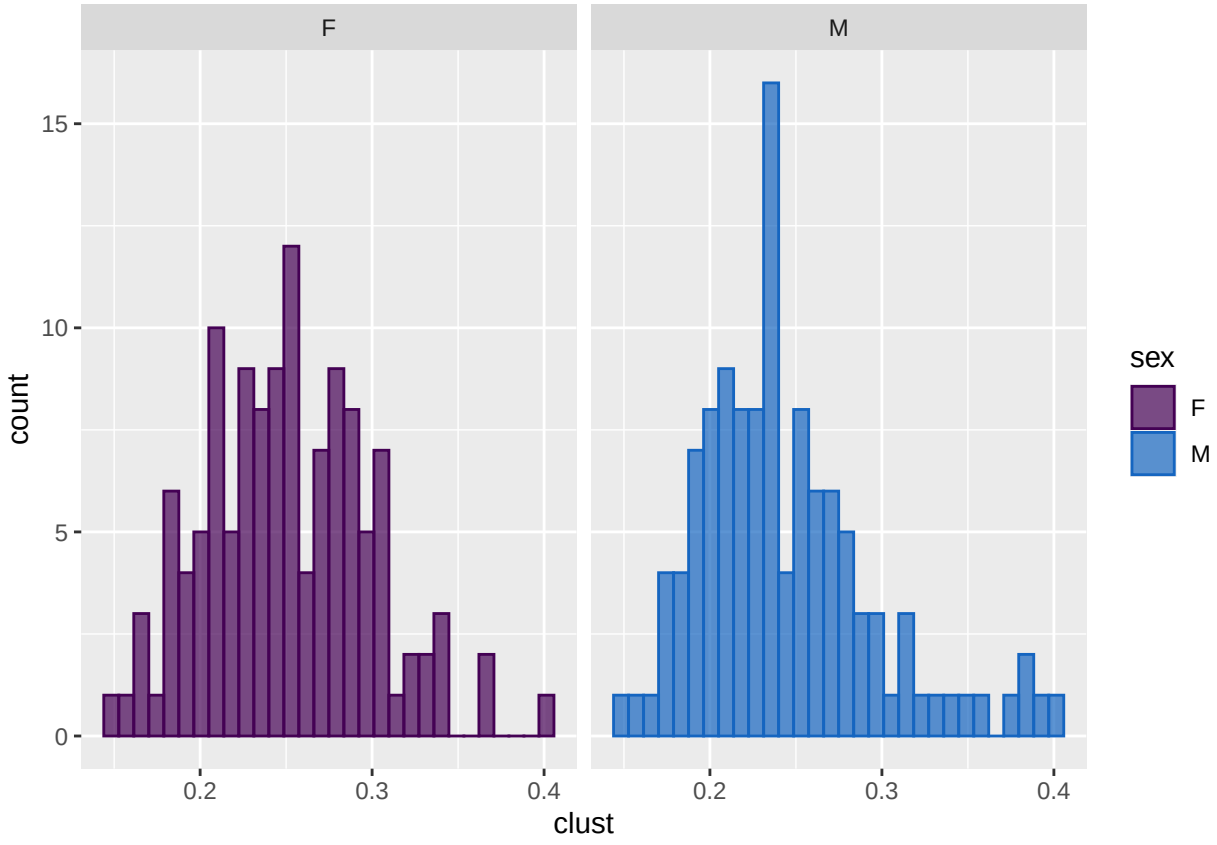
```
##
## Call:
## glm.nb(formula = f, data = arg_1, init.theta = 37.37939373, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2027  -0.6616   0.0122   0.7146   2.0470
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.58076    0.07201  49.728  < 2e-16 ***
## clust        0.95482    0.28409   3.361 0.000777 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(37.3794) family taken to be 1)
##
##     Null deviance: 258.03  on 240  degrees of freedom
## Residual deviance: 246.83  on 239  degrees of freedom
```

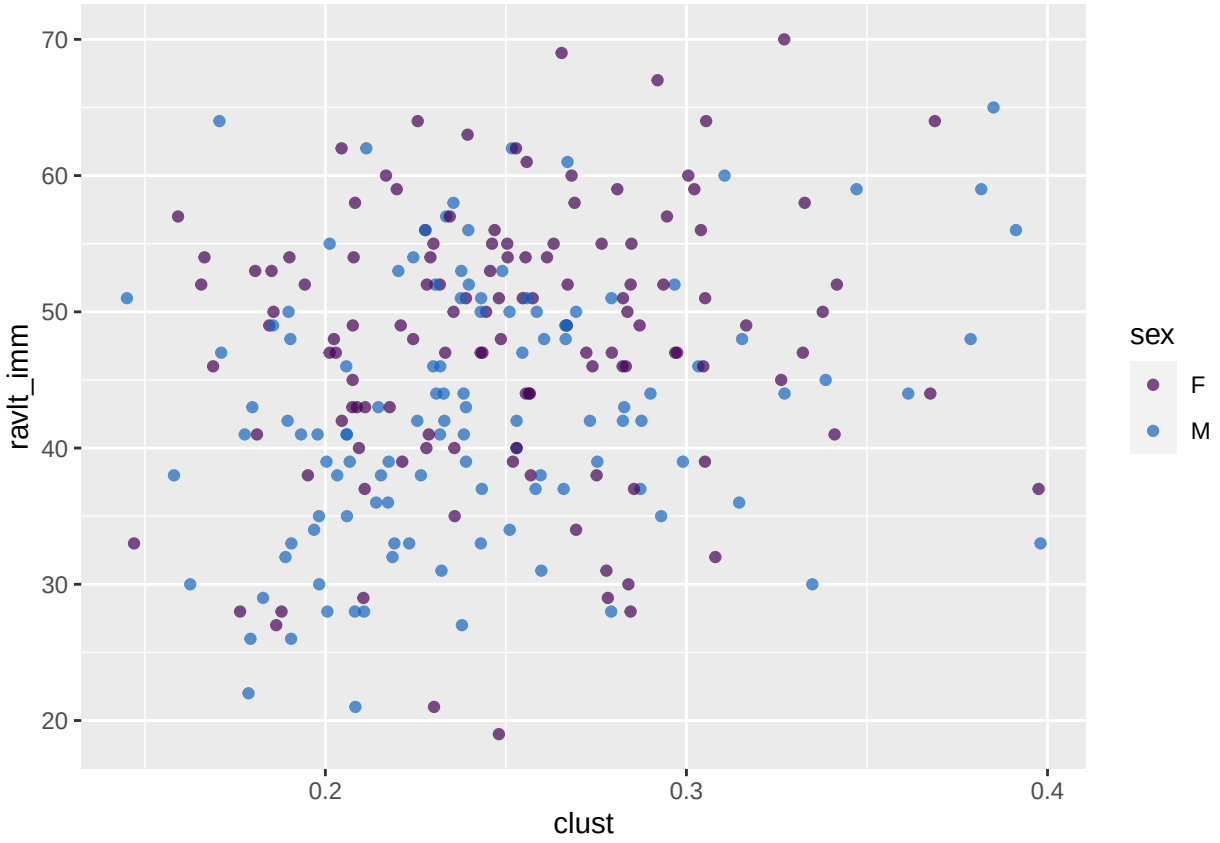
```
## AIC: 1801
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  37.38
##           Std. Err.:  6.34
##
##  2 x log-likelihood:  -1795.045
##
## [1] "Confidence Intervals:"

## Waiting for profiling to be done...

##             Estimate     2.5 %   97.5 %
## (Intercept) 3.580761 3.4388458 3.722582
## clust       0.954822 0.3953386 1.515041
##
## [1] "Incident Ratios:"
##              Estimate     2.5 %    97.5 %
## (Intercept) 35.900844 31.150983 41.371071
## clust        2.598208  1.484887  4.549609

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





Modeling: Characteristic Path Length

```
# Characteristic Path Length
plot_and_model(data, "cpath", "ravlt_imm")
```

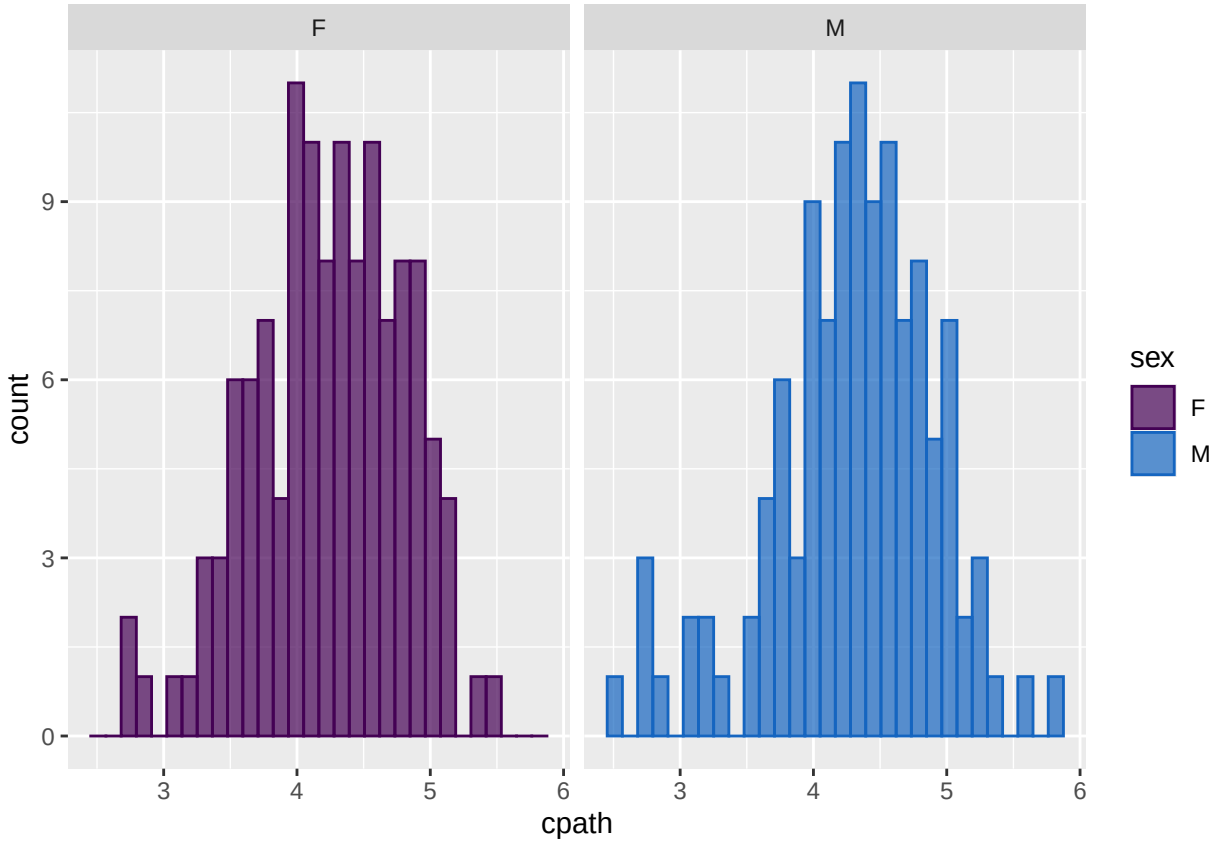
```
##
## Call:
## glm.nb(formula = f, data = arg_1, init.theta = 37.09065318, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1818  -0.7142   0.0345   0.7297   1.9820
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.14311    0.10331  40.102  < 2e-16 ***
## cpath       -0.07638    0.02404  -3.177  0.00149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(37.0907) family taken to be 1)
##
##     Null deviance: 256.96  on 240  degrees of freedom
## Residual deviance: 246.86  on 239  degrees of freedom
```

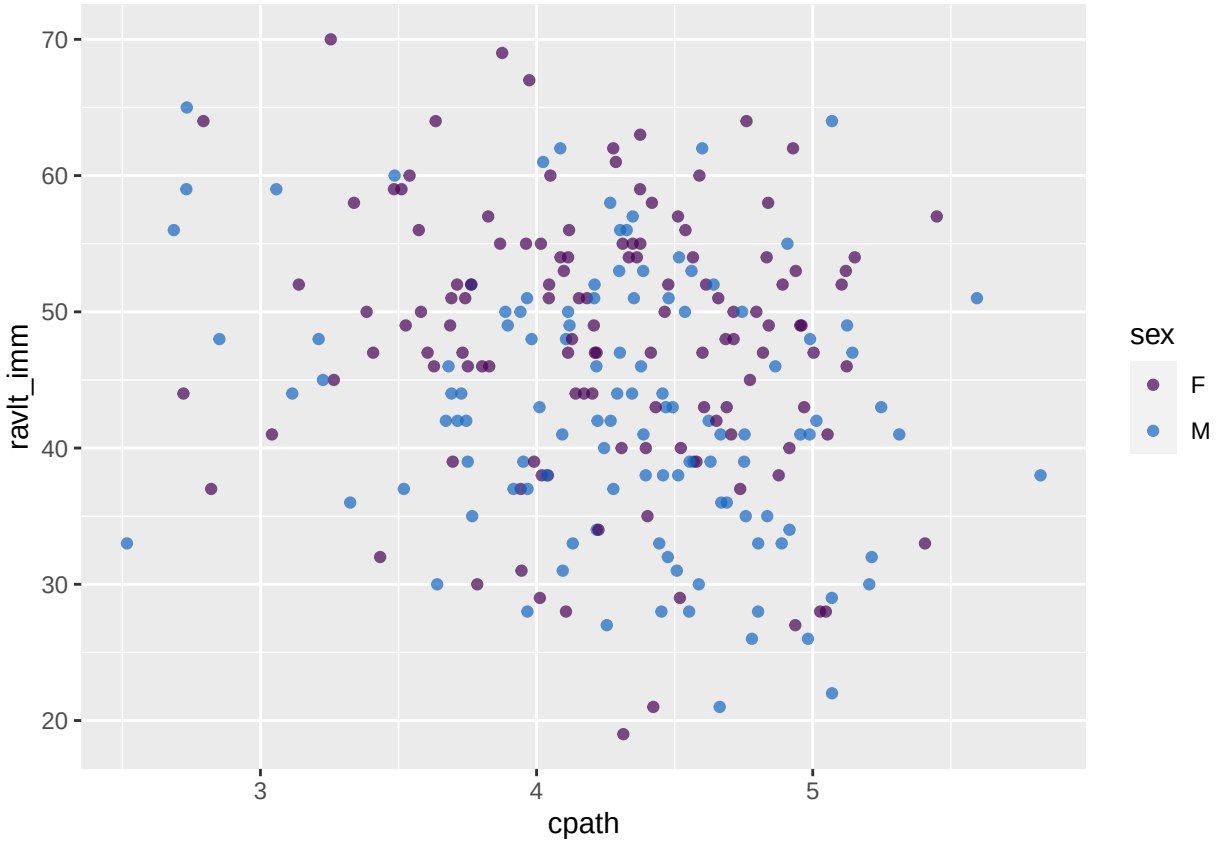
```
## AIC: 1802.1
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  37.09
##           Std. Err.:  6.27
##
##  2 x log-likelihood:  -1796.097
##
## [1] "Confidence Intervals:"

## Waiting for profiling to be done...

##                Estimate      2.5 %     97.5 %
## (Intercept)  4.14311258  3.9406626  4.3458642
## cpath       -0.07638452 -0.1235575 -0.0292626
##
## [1] "Incident Ratios:"
##                Estimate      2.5 %     97.5 %
## (Intercept) 62.9986045 51.4526806 77.1586874
## cpath        0.9264599  0.8837709  0.9711614

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





Modeling: Global Efficiency

```
# Global Efficiency
plot_and_model(data, "geff", "ravlt_imm")
```

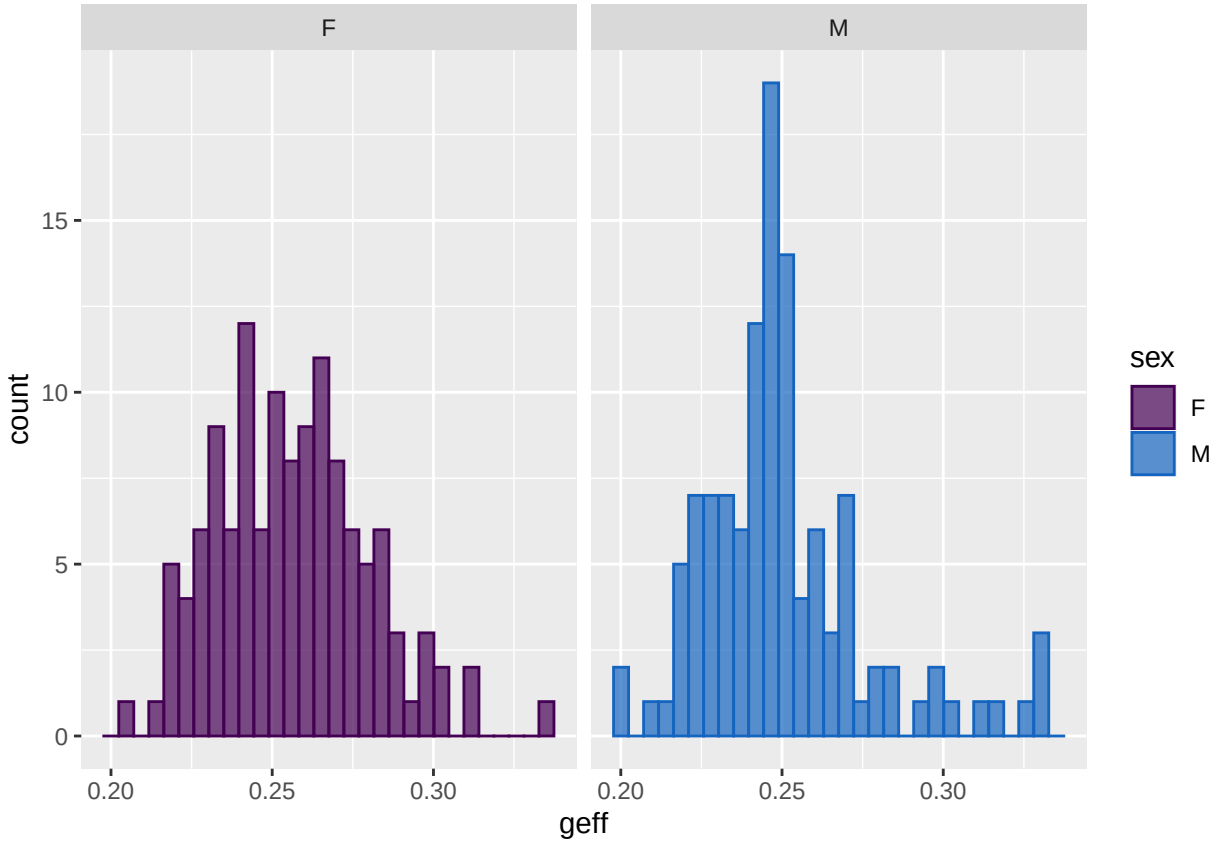
```
##
## Call:
## glm.nb(formula = f, data = arg_1, init.theta = 37.27988638, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2496  -0.7015  -0.0160   0.7012   1.9892
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.3402     0.1462  22.850  < 2e-16 ***
## geff          1.8815     0.5729   3.284  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(37.2799) family taken to be 1)
##
##     Null deviance: 257.66  on 240  degrees of freedom
## Residual deviance: 246.89  on 239  degrees of freedom
```

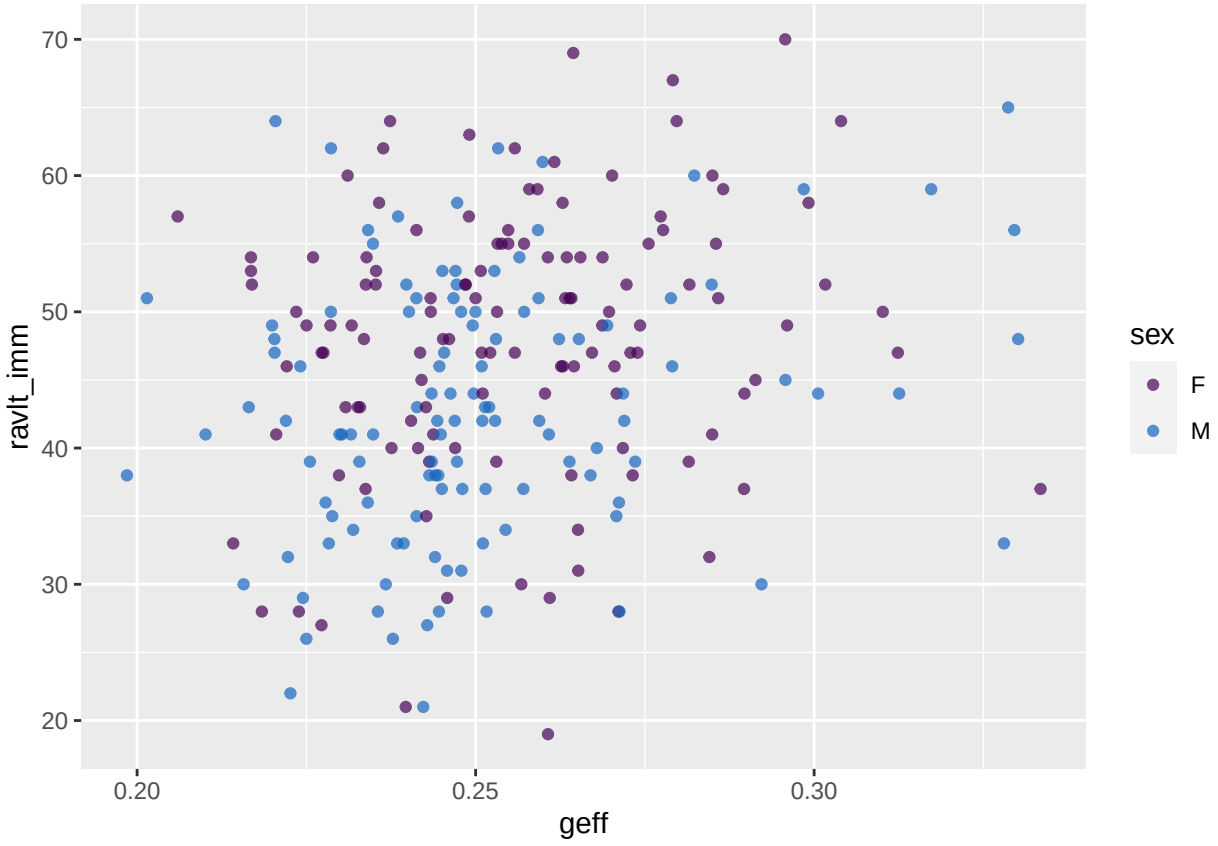
```
## AIC: 1801.5
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  37.28
##           Std. Err.:  6.32
##
##  2 x log-likelihood:  -1795.457
##
## [1] "Confidence Intervals:"

## Waiting for profiling to be done...

##             Estimate     2.5 %   97.5 %
## (Intercept) 3.340235 3.0530149 3.627116
## geff        1.881509 0.7572626 3.007460
##
## [1] "Incident Ratios:"
##              Estimate     2.5 %   97.5 %
## (Intercept) 28.225767 21.179101 37.60420
## geff         6.563405  2.132431 20.23594

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




70
60
50
40
30
20
ravlt_imm
0.20
0.25
0.30
geff
sex
F
M