Fenton & Tureson 573 Final Project

Laura Fenton & Kayla Tureson

2022-04-28

```
library(ggplot2)  # for plots
library(tinytex)
library(here)
library(readxl)  # for reading excel files
library(modelsummary)  # for summarizing data
library(rstan)
rstan_options(auto_write = TRUE)  # save compiled STAN object
options(mc.cores = 2)  # use two cores
library(posterior)
library(bayesplot)
library(dplyr)
library(kableExtra)
library(dagitty)
library(brms)
library(ggdag)  # render DAGs in ggplot style
theme_set(theme_classic() +
    theme(panel.grid.major.y = element_line(color = "grey92")))
```

Research Questions

> Is there a difference in amyloid accumulation between individuals in the control condition and the exercise intervention condition? Is there a difference in hippocampal atrophy between individuals in the control condition and the exercise intervention condition? Does VO2 max change mediate the relationship between treatment condition and hippocampal atrophy rate?

Variables

- `amyloid_accum`: accumulation of amyloid from pre (baseline) to post (wk52) intervention
- `Hippocampus_change`: hippocampal atrophy from (baseline) to post (wk52) intervention
- `VO2_change`: change in VO2 max from (baseline) to post (wk52) intervention
- `apx_group`: AExSupport = aerobic exercise condition, NoSupport = control condition

Import Data

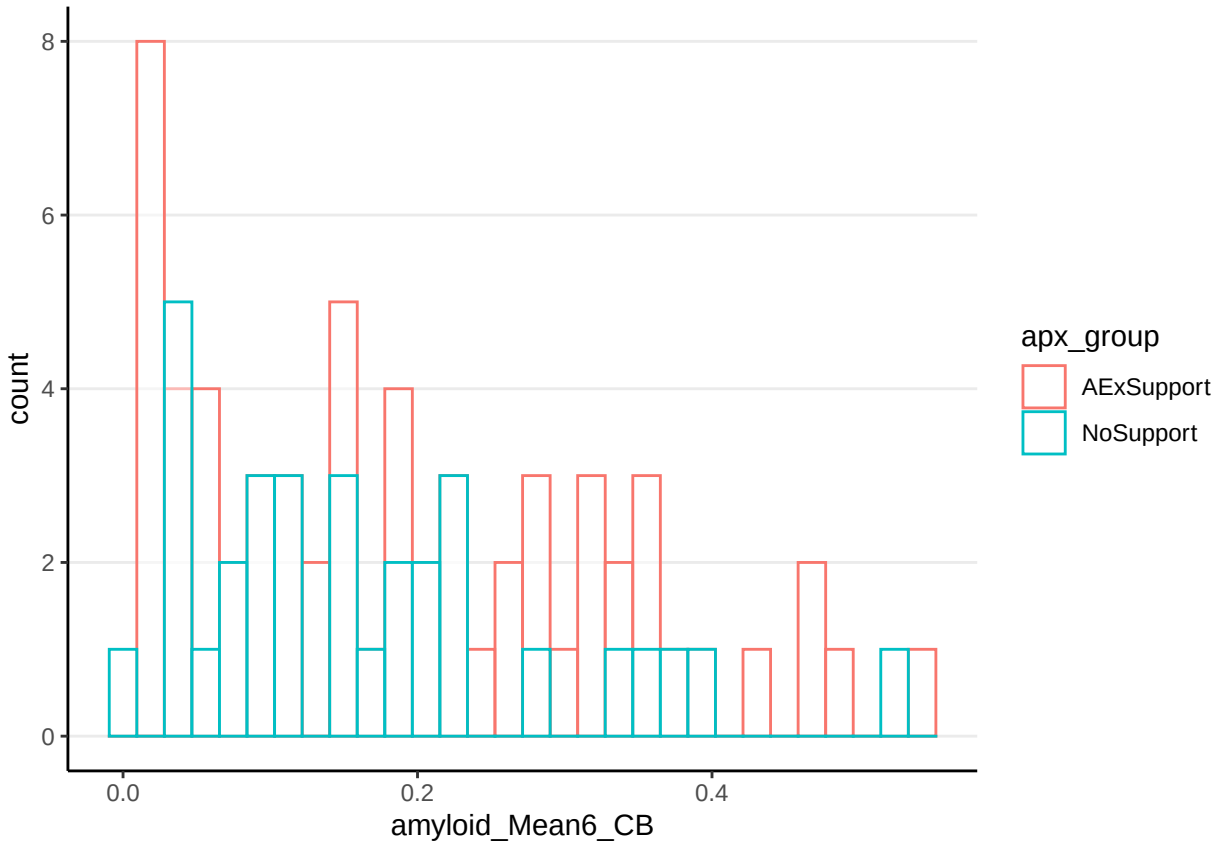
```
# load the data
exercise <- read_excel(here("APEx_FinalData_20201021.xlsx"))
```

```r
# subset only individuals who have baseline and follow up data
tt <- table(exercise$study_id)
exercise2 <- subset(exercise, study_id %in% names(tt[tt == 3]))

# log transform key variables
exercise2$amyloid_Mean6_CB <- log(exercise2$amyloid_Mean6_CB)
exercise2$Hippocampus_Vol_mL <- log(exercise2$Hippocampus_Vol_mL)
exercise2$apx_exppt_vo2mx_ml <- log(exercise2$apx_exppt_vo2mx_ml)

# create new variables: (amyloid_accum) for amyloid accumulation from baseline
# to week 52 (smaller values are better); (Hippocampus_change) for hippocampal neurodegeneration from b
exercise3 <-
exercise2 %>%
  group_by(study_id)  %>%
  mutate(amyloid_accum = amyloid_Mean6_CB[timept=="WK52"] -
           amyloid_Mean6_CB[timept=="BL"],
         Hippocampus_change = Hippocampus_Vol_mL[timept=="BL"] -
           Hippocampus_Vol_mL[timept=="WK52"],
         VO2_change = apx_exppt_vo2mx_ml[timept=="WK52"] -
           apx_exppt_vo2mx_ml[timept=="BL"])

# subset only the first entry of each participant
 exercise.subset <-
  exercise3 %>%
  group_by(study_id) %>%
  filter(row_number()==1)

# subset only the third entry of each participant
exercise.subset.wk52 <-
  exercise3 %>%
  group_by(study_id) %>%
  filter(row_number()==3)

# change the condition variable to a factor
exercise.subset$apx_group <- as.factor(exercise.subset$apx_group)

# remove NAs so that STAN will run later
exercise.subset <- na.omit(exercise.subset)

# create new variable for condition and set as integer
exercise.subset$apx_group_coded <- ifelse(exercise.subset$apx_group=="NoSupport", 0, ifelse(exercise.su

exercise.subset$apx_group_coded <- as.integer(exercise.subset$apx_group_coded)
```

Visualize data

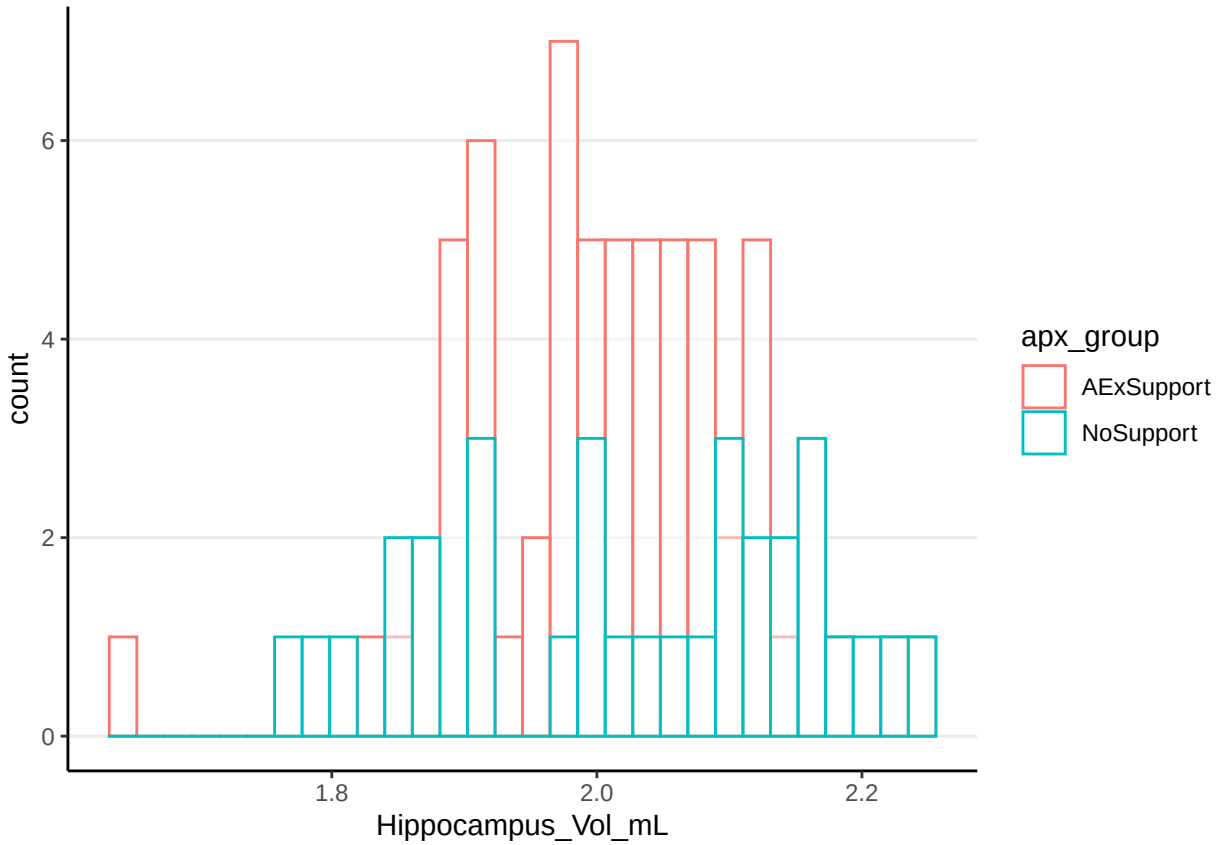
```r
# Visualize baseline variables
ggplot(exercise.subset, aes(x=amyloid_Mean6_CB, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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



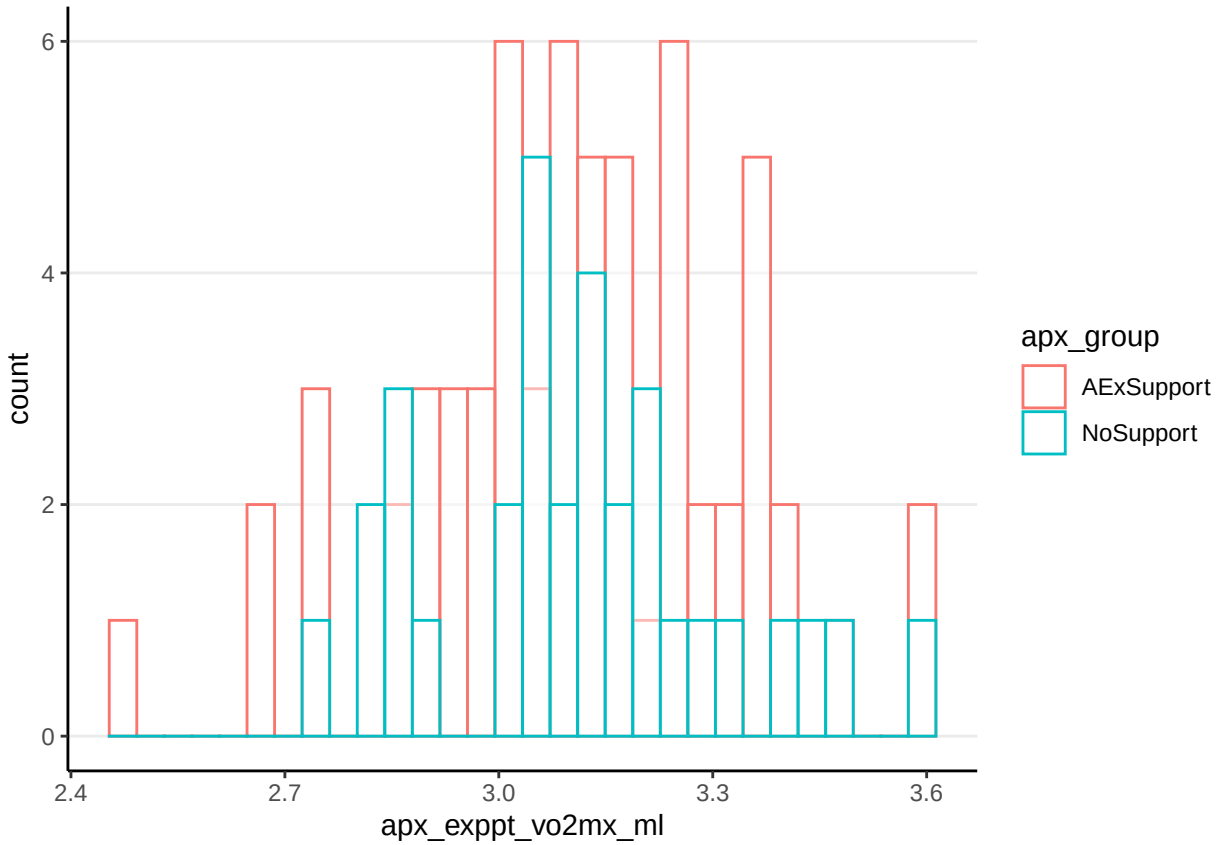

```
ggplot(exercise.subset, aes(x=Hippocampus_Vol_mL, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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




```
ggplot(exercise.subset, aes(x=apx_exppt_vo2mx_ml, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

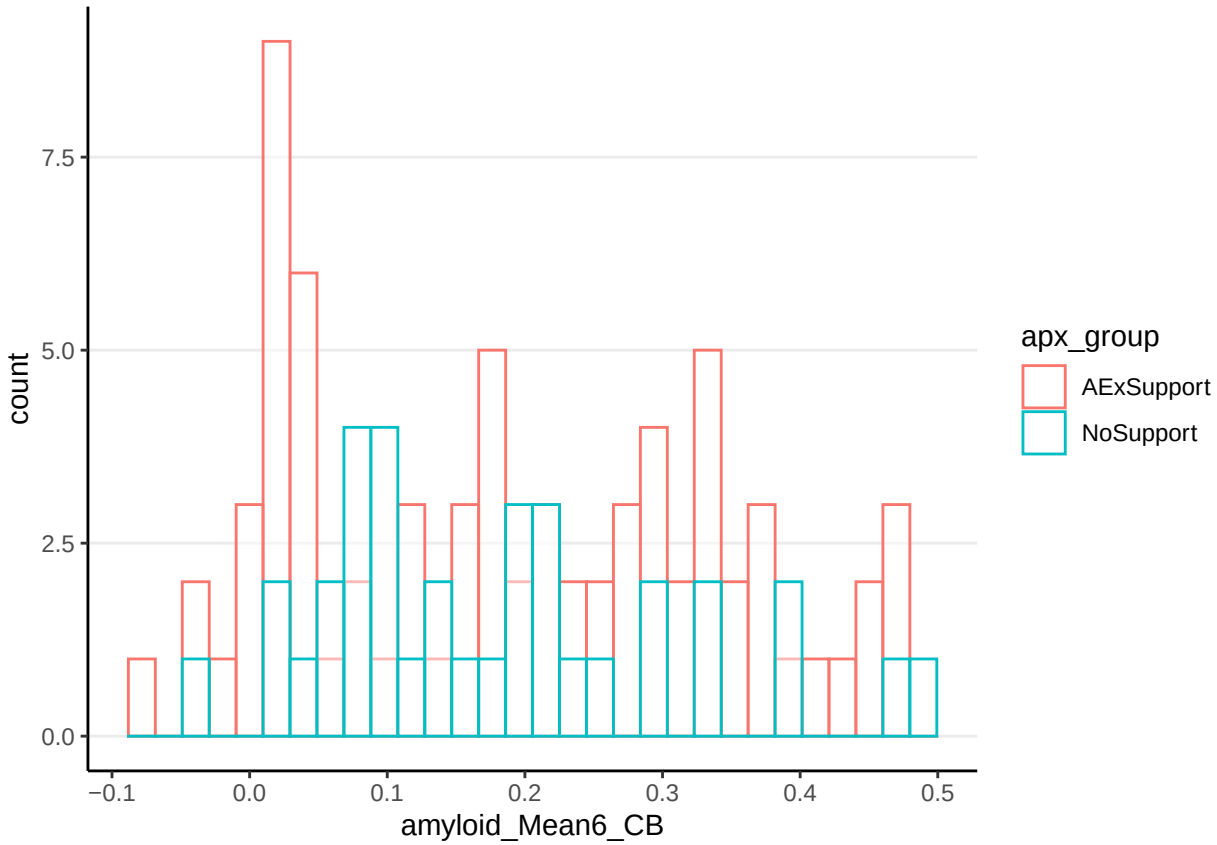
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#Visualize week 52 variables

ggplot(exercise.subset.wk52, aes(x=amyloid_Mean6_CB, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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

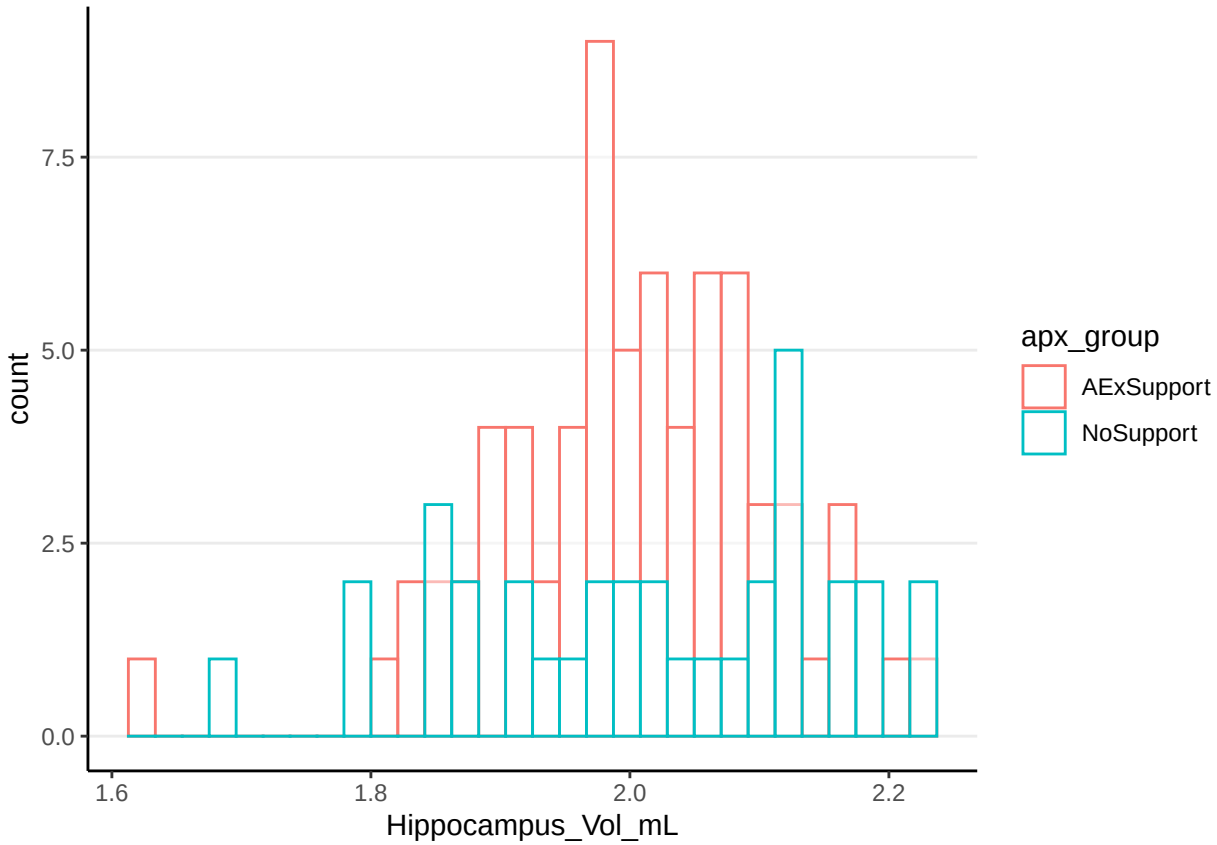
```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```




```
ggplot(exercise.subset.wk52, aes(x=Hippocampus_Vol_mL, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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

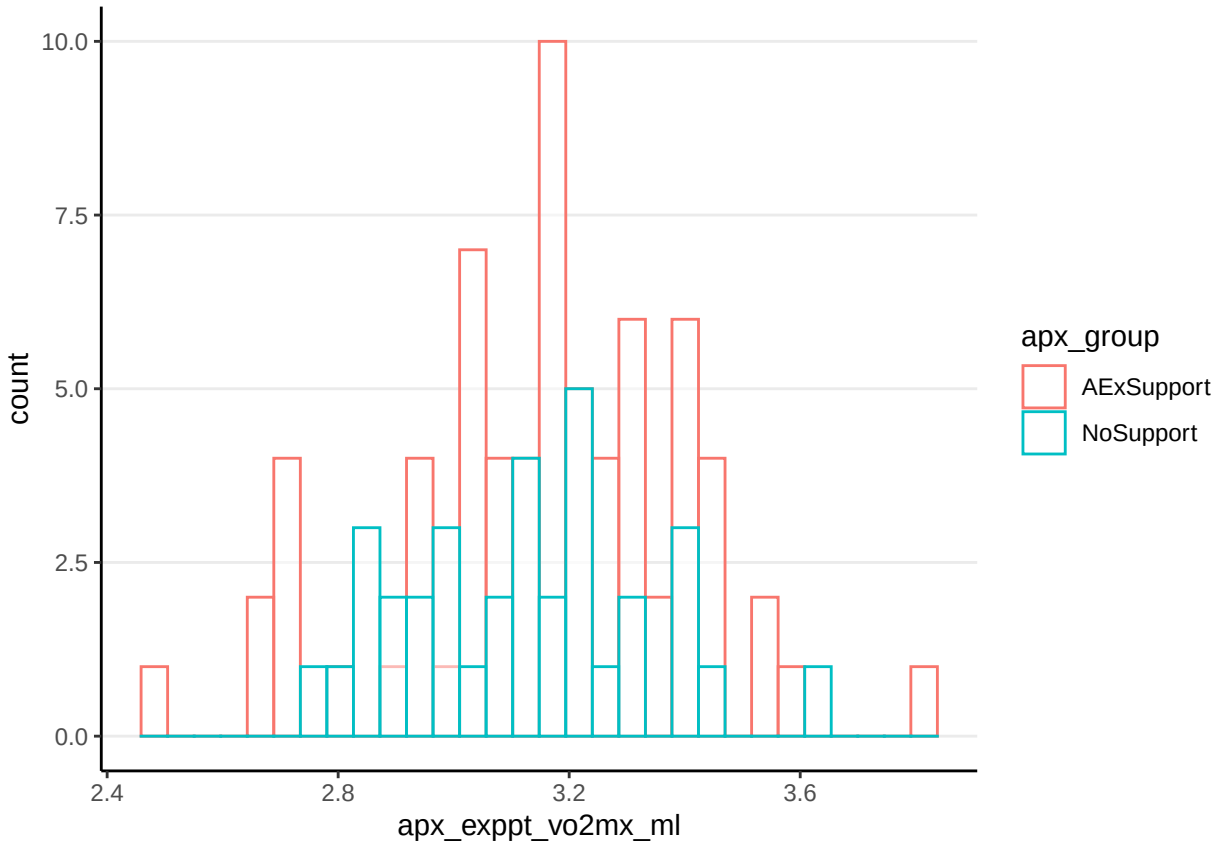
```
## Warning: Removed 7 rows containing non-finite values (stat_bin).
```




```
ggplot(exercise.subset.wk52, aes(x=apx_exppt_vo2mx_ml, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

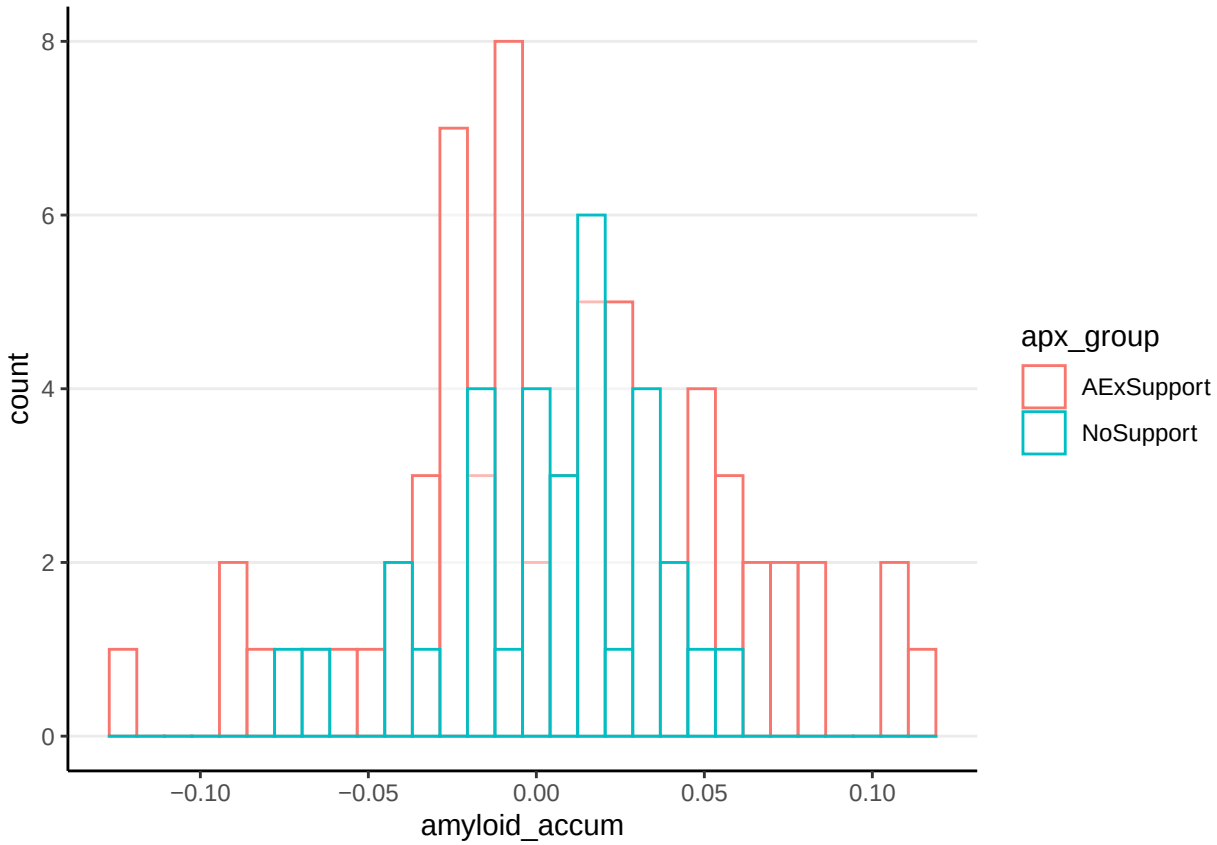
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 7 rows containing non-finite values (stat_bin).
```



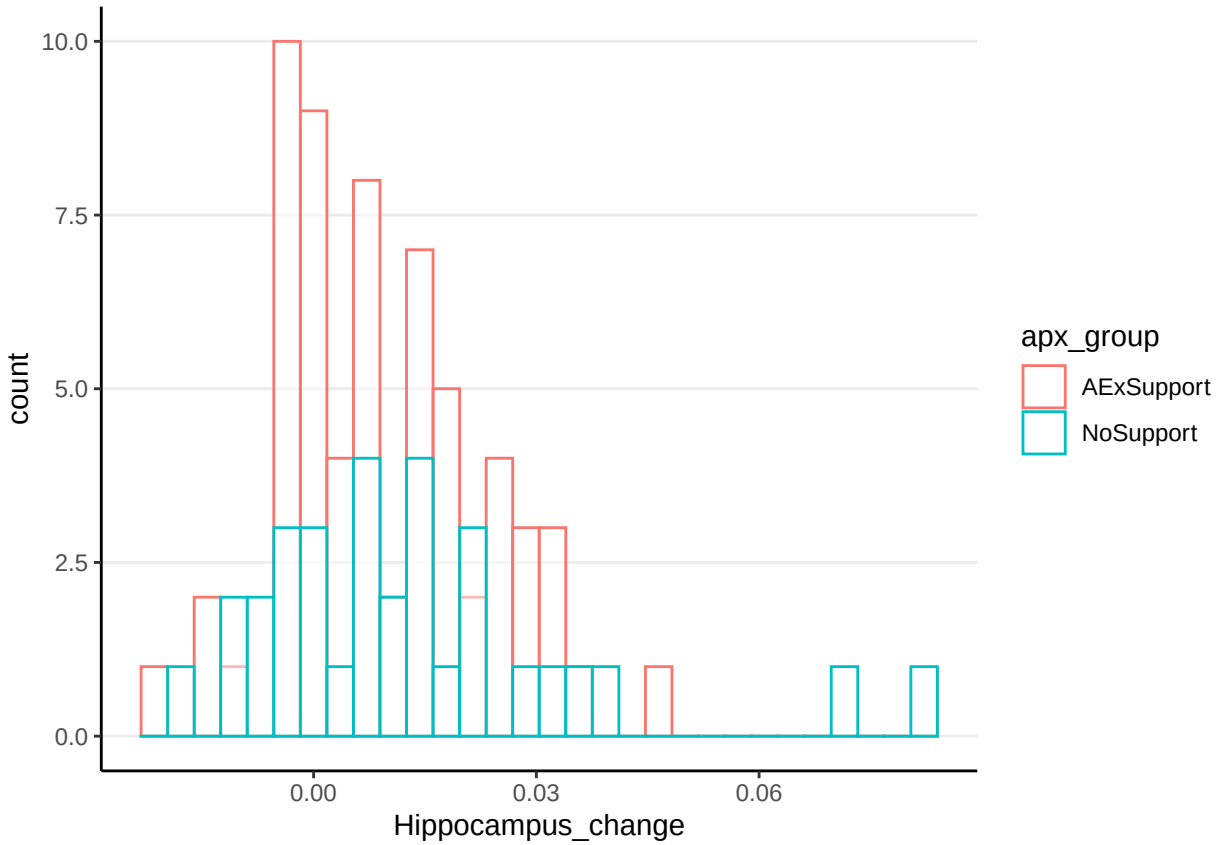

```
# Visualize change variables
ggplot(exercise.subset, aes(x=amyloid_accum, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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




```
ggplot(exercise.subset, aes(x=Hippocampus_change, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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



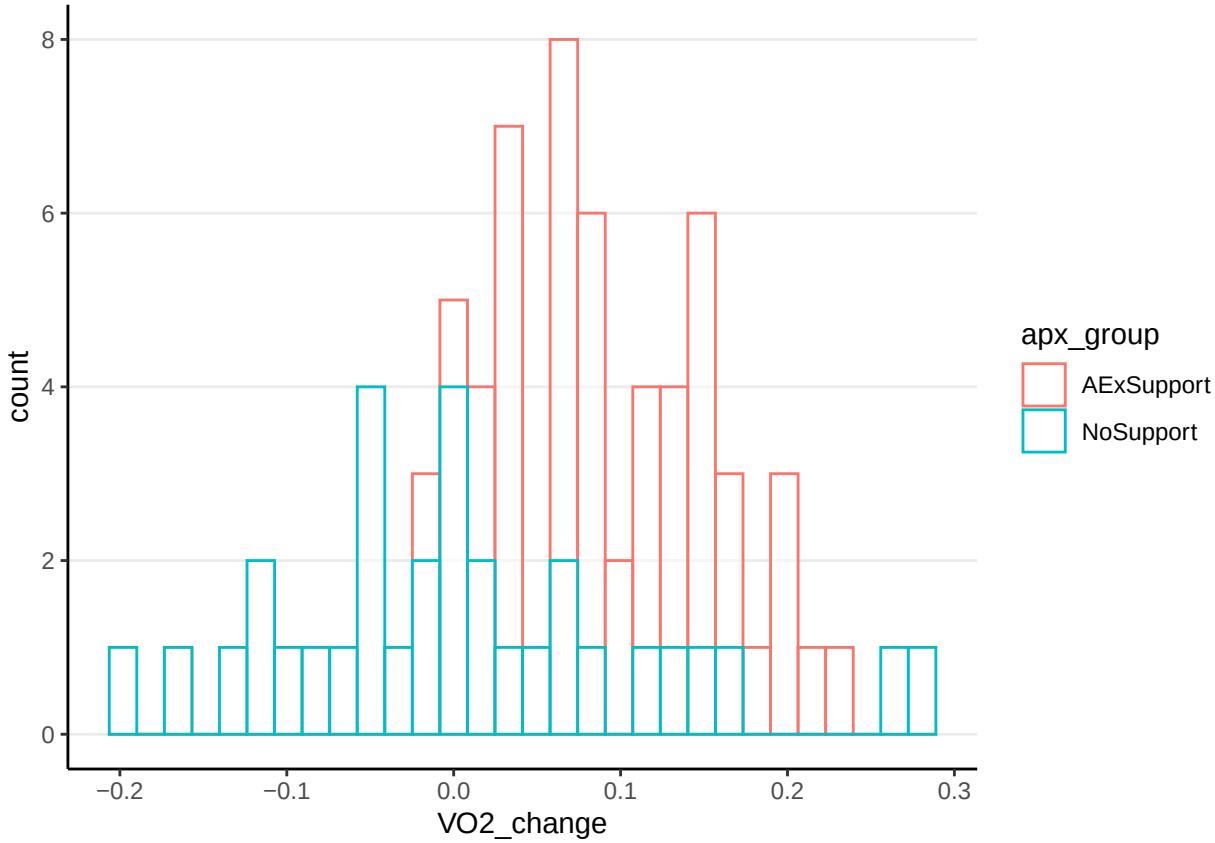

```
ggplot(exercise.subset, aes(x=VO2_change, color=apx_group)) +
  geom_histogram(fill="white", alpha=0.5, position="identity")
```

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

Table 1: Summary Statistics

		AExSupport	NoSupport
amyloid_accum	Mean	0.01	0.00
	SD	0.05	0.03
	Min	−0.12	−0.08
	Max	0.12	0.06
	Histogram		

SD = standard deviation



Variable Summary

```
## can't get N to work, not sure why
datasummary(amyloid_accum *
                (Mean + SD + Min + Max + Histogram) ~
                factor(apx_group),
            data = exercise.subset,
caption = "Summary Statistics") %>%
    # add table note
    add_footnote("SD = standard deviation", notation = "none")
```

Table 2: Summary Statistics

		AExSupport	NoSupport
Hippocampus_change	Mean	0.01	0.01
	SD	0.01	0.02
	Min	−0.02	−0.02
	Max	0.05	0.08
	Histogram		

SD = standard deviation

```
datasummary(Hippocampus_change *
                (Mean + SD + Min + Max + Histogram) ~
                factor(apx_group),
            data = exercise.subset,
caption = "Summary Statistics") %>%
    # add table note
    add_footnote("SD = standard deviation", notation = "none")
```

Model Amyloid Accumulation

Let Y_A = change in amyloid from baseline to week 52 (week 52 - baseline; smaller values indicate less accumulation), G = treatment condition Model:

$$Y_{Ai,G=0} \sim N(\mu_1, \sigma_1)$$
$$Y_{Ai,G=1} \sim N(\mu_2, \sigma_2)$$

Prior:

$$\mu_1 \sim N(3, 2)$$
$$\mu_2 \sim N(3, 2)$$
$$\sigma_1 \sim N^+(0, 2)$$
$$\sigma_2 \sim N^+(0, 2)$$

Running Stan

We used 4 chains, each with 4,000 iterations (first 2,000 as warm-ups).

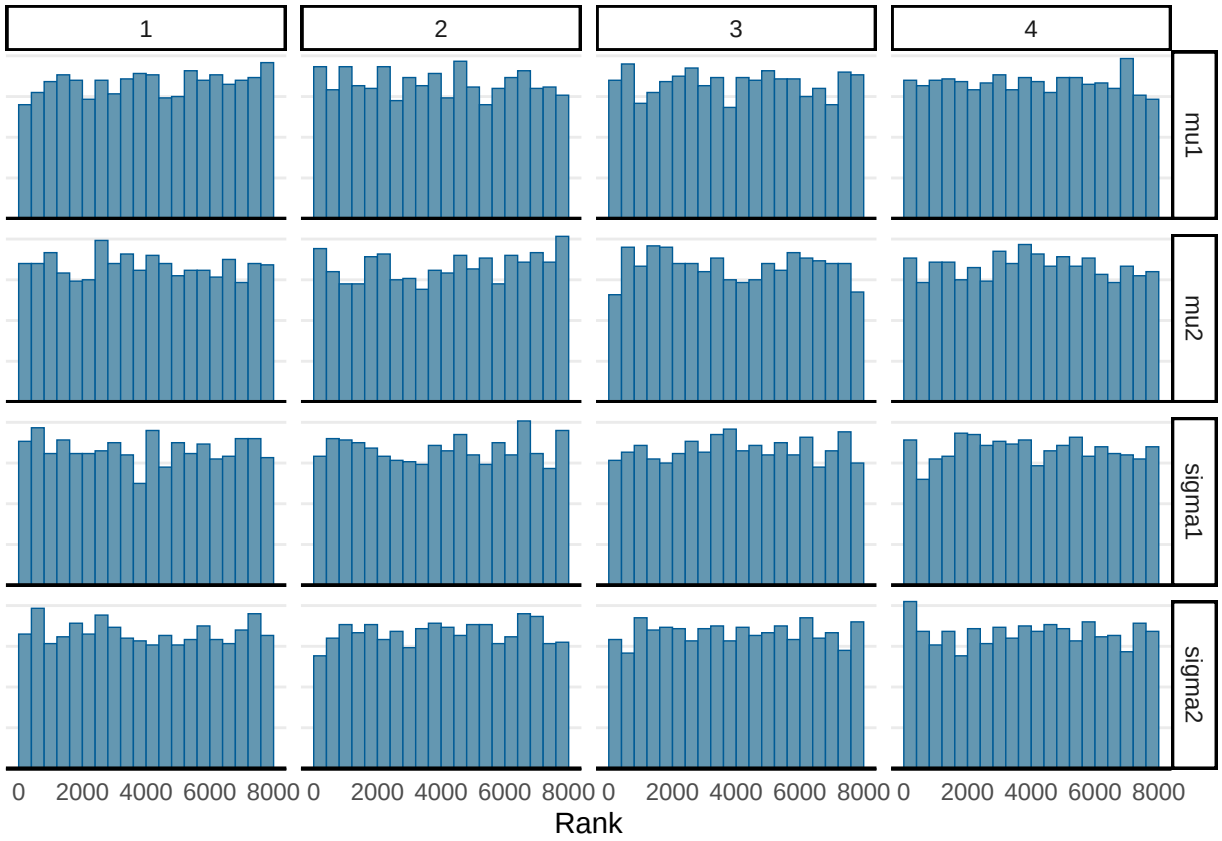
```
# Rescale data so all are greater than 0 and STAN will run (chose .13 based
# on min values from summary data)
exercise.subset$amyloid_accum_rescale <- exercise.subset$amyloid_accum + .13

# 1. form the data list for Stan
stan_dat <- with(exercise.subset,
    list(N1 = sum(apx_group == "NoSupport"),
         N2 = sum(apx_group == "AExSupport"),
         y1 = amyloid_accum_rescale[which(apx_group == "NoSupport")],
         y2 = amyloid_accum_rescale[which(apx_group == "AExSupport")])
)
```

```
# 2. Run Stan
m1 <- stan(
    file = here("robust_2group.stan"),
    data = stan_dat,
    seed = 1234,  # for reproducibility
    iter = 4000
)
```

Results

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m1, pars = c("mu1", "mu2", "sigma1", "sigma2"))
```



The following table shows the posterior distributions of μ_1, μ_2, σ_1, σ_2, and $\mu_2 - \mu_1$.

```
summ_m1 <- as_draws_df(m1) %>%
    subset_draws(variable = c("mu1", "mu2", "sigma1", "sigma2")) %>%
    mutate_variables(`mu2 - mu1` = mu2 - mu1) %>%
    summarise_draws()
knitr::kable(summ_m1, digits = 2)
```

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
mu1	0.14	0.14	0.01	0.01	0.13	0.15	1	8677.00	5878.90
mu2	0.14	0.14	0.01	0.01	0.13	0.15	1	9111.12	6053.73
sigma1	0.03	0.03	0.00	0.00	0.02	0.04	1	7549.52	5524.12
sigma2	0.05	0.05	0.01	0.00	0.04	0.06	1	7598.33	5195.63
mu2 - mu1	0.00	0.00	0.01	0.01	-0.01	0.02	1	9213.97	6289.59

The analysis showed that on average, there was no difference in amyloid accumulation from baseline to week 52 between individuals in the aerobic exercise condition and individuals in the education control condition. Individuals in both the control condition and intervention condition had a posterior mean of 0.14. The 90% CI's were identical for both groups. As a reminder, the actual values are .14 lower than those we see in the table, due to data transformation prior to running our model.

Model Hippocampal Neurodegeneration

Let Y_A = change in amyloid from baseline to week 52 (week 52 - baseline; smaller values indicate less accumulation), G = treatment condition Model:

$$Y_{Hi,G=0} \sim N(\mu_1, \sigma_1)$$
$$Y_{Hi,G=1} \sim N(\mu_2, \sigma_2)$$

Prior:

$$\mu_1 \sim N(3, 2)$$
$$\mu_2 \sim N(3, 2)$$
$$\sigma_1 \sim N^+(0, 2)$$
$$\sigma_2 \sim N^+(0, 2)$$

Running Stan

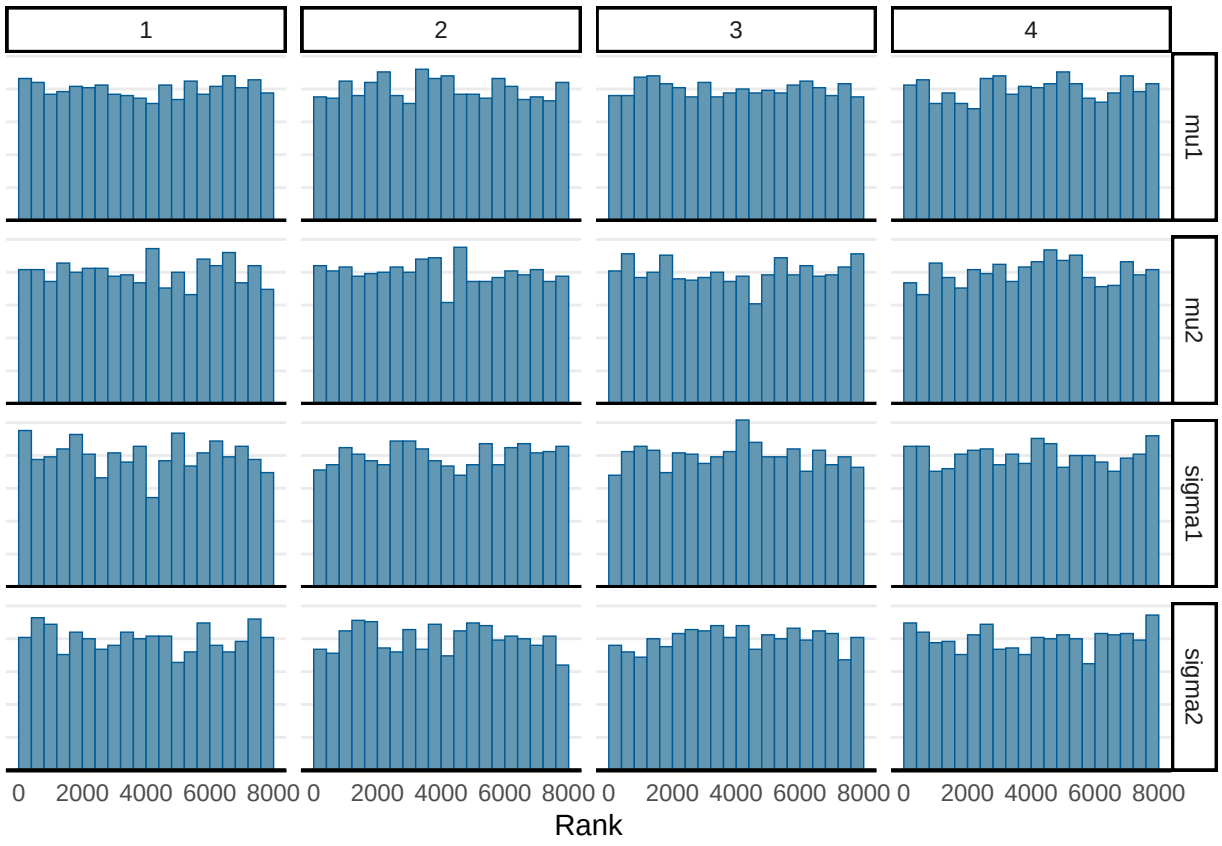
We used 4 chains, each with 4,000 iterations (first 2,000 as warm-ups).

```
# Rescale data so all are greater than 0 and STAN will run (chose .03 based
# on min values from summary data)
exercise.subset$Hippocampus_change_rescale <- exercise.subset$Hippocampus_change + .03

# 1. form the data list for Stan
stan_dat_2 <- with(exercise.subset,
    list(N1 = sum(apx_group == "NoSupport"),
         N2 = sum(apx_group == "AExSupport"),
         y1 = Hippocampus_change_rescale[which(apx_group == "NoSupport")],
         y2 = Hippocampus_change_rescale[which(apx_group == "AExSupport")])
)
# 2. Run Stan
m2 <- stan(
    file = here("robust_2group.stan"),
    data = stan_dat_2,
    seed = 1234,  # for reproducibility
    iter = 4000
)
```

Results

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m2, pars = c("mu1", "mu2", "sigma1", "sigma2"))
```



The following table shows the posterior distributions of μ_1, μ_2, σ_1, σ_2, and $\mu_2 - \mu_1$.

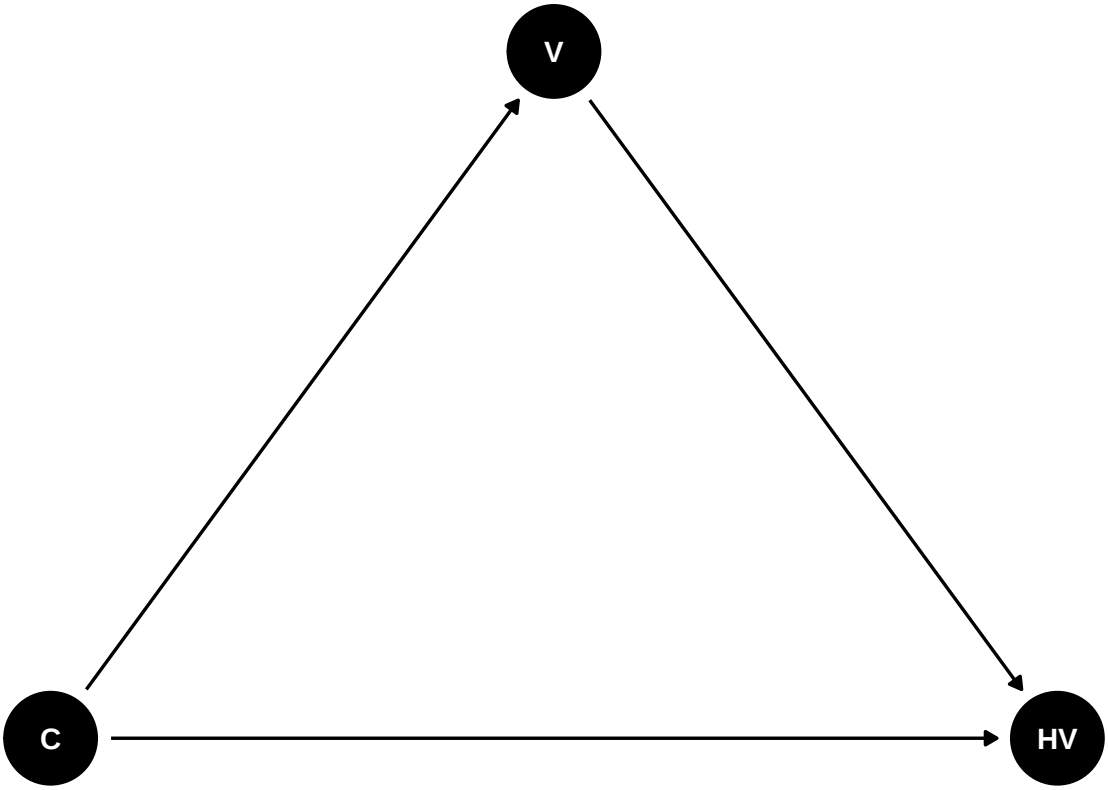
```
summ_m2 <- as_draws_df(m2) %>%
    subset_draws(variable = c("mu1", "mu2", "sigma1", "sigma2")) %>%
    mutate_variables(`mu2 - mu1` = mu2 - mu1) %>%
    summarise_draws()
knitr::kable(summ_m2, digits = 2)
```

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
mu1	0.04	0.04	0	0	0.04	0.05	1	7489.70	6024.99
mu2	0.04	0.04	0	0	0.04	0.04	1	9713.02	6581.52
sigma1	0.02	0.02	0	0	0.01	0.03	1	5663.46	5519.05
sigma2	0.01	0.01	0	0	0.01	0.02	1	5844.44	5162.64
mu2 - mu1	0.00	0.00	0	0	-0.01	0.00	1	8030.14	6406.76

The analysis showed that on average, there was no difference in hippocampal neurodegeneration from baseline to week 52 between individuals in the aerobic exercise condition and individuals in the education control condition. Individuals in the exercise and control conditions both had a posterior mean of 0.04, and the confidence intervals were identical. As a reminder, the actual values are .03 lower than those we see in the table, due to data transformation prior to running our model.

Mediation

```
dag1 <- dagitty(
    "dag{
      C -> HV; C -> V; V -> HV
    }"
)
coordinates(dag1) <- list(x = c(C = 0, V = 1, HV = 2),
                          y = c(C = 0, V = 1, HV = 0))
# Plot
ggdag(dag1) + theme_dag()
```




```
get_prior(Hippocampus_change ~ apx_group_coded + VO2_change,
          data = exercise.subset,
          family = gaussian(link = "identity"))
```

```
##                 prior     class            coef group resp dpar nlpar bound
##                (flat)         b
##                (flat)         b apx_group_coded
##                (flat)         b      VO2_change
##  student_t(3, 0, 2.5) Intercept
##  student_t(3, 0, 2.5)     sigma
##        source
##       default
```

```
##  (vectorized)
##  (vectorized)
##       default
##       default
```

```
get_prior(VO2_change ~ apx_group_coded,
          data = exercise.subset,
          family = gaussian(link = "identity"))
```

```
##                   prior     class            coef group resp dpar nlpar bound
##                  (flat)         b
##                  (flat)         b apx_group_coded
##  student_t(3, 0.1, 2.5) Intercept
##    student_t(3, 0, 2.5)     sigma
##         source
##       default
##  (vectorized)
##       default
##       default
```

```
prior_m <- prior(normal(0, 1), class = "b")
m_med <- brm(
        bf(Hippocampus_change ~ apx_group_coded + VO2_change) +
        bf(VO2_change ~ apx_group_coded) +
        set_rescor(FALSE),
       family = list(gaussian("identity"), gaussian("identity")),
       data = exercise.subset,
       prior = prior(normal(1,1), class = "b", resp = "VO2change") +
        prior(student_t(3, 0, 2.5), class = "sigma", resp = "VO2change") +
        prior(normal(1, 1), class = "b", resp = "Hippocampuschange"),
      seed = 1338,
      iter = 4000
)
```

```
## Compiling Stan program...
```

```
## Trying to compile a simple C file
```

```
## Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
## clang -mmacosx-version-min=10.13 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG   -I
## In file included from <built-in>:1:
## In file included from /Library/Frameworks/R.framework/Versions/4.1/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/inclu
## In file included from /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/inclu
## /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/include/Eigen/src/Core/util/
## namespace Eigen {
## ^
## /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/include/Eigen/src/Core/util/
## namespace Eigen {
##                 ^
##                 ;
## In file included from <built-in>:1:
```

```
## In file included from /Library/Frameworks/R.framework/Versions/4.1/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/inclu
## /Library/Frameworks/R.framework/Versions/4.1/Resources/library/RcppEigen/include/Eigen/Core:96:10: fa
## #include <complex>
##          ^~~~~~~~
## 3 errors generated.
## make: *** [foo.o] Error 1

## Start sampling
```

```
print(m_med)
```

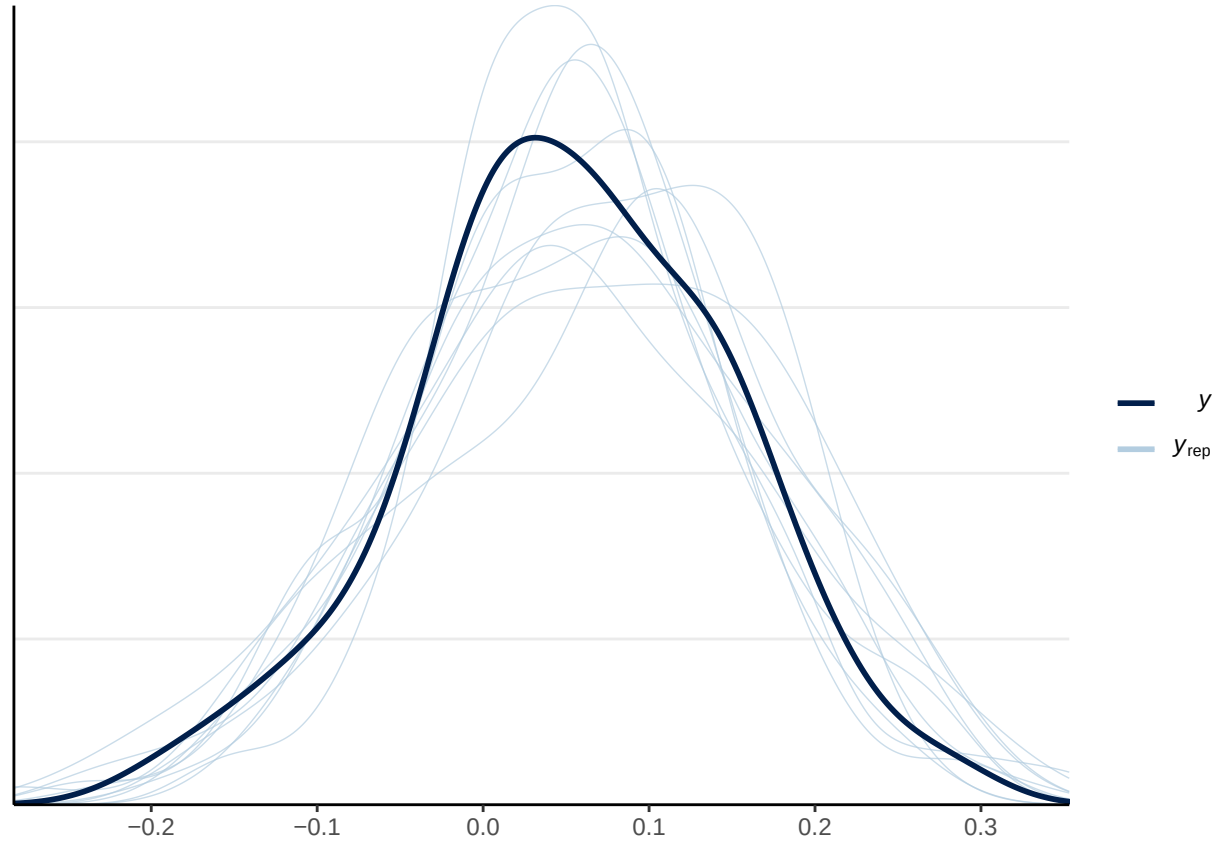
```
##  Family: MV(gaussian, gaussian)
##   Links: mu = identity; sigma = identity
##          mu = identity; sigma = identity
## Formula: Hippocampus_change ~ apx_group_coded + VO2_change
##          VO2_change ~ apx_group_coded
##    Data: exercise.subset (Number of observations: 95)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Population-Level Effects:
##                                 Estimate Est.Error l-95% CI u-95% CI Rhat
## Hippocampuschange_Intercept         0.01      0.00     0.01     0.02 1.00
## VO2change_Intercept                 0.01      0.02    -0.03     0.04 1.00
## Hippocampuschange_apx_group_coded  -0.00      0.00    -0.01     0.00 1.00
## Hippocampuschange_VO2_change        0.01      0.02    -0.03     0.04 1.00
## VO2change_apx_group_coded           0.07      0.02     0.03     0.11 1.00
##                                 Bulk_ESS Tail_ESS
## Hippocampuschange_Intercept        12755     6162
## VO2change_Intercept                13049     5965
## Hippocampuschange_apx_group_coded  12976     6687
## Hippocampuschange_VO2_change       11291     5605
## VO2change_apx_group_coded          13520     5867
##
## Family Specific Parameters:
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sigma_Hippocampuschange     0.02      0.00     0.01     0.02 1.00    11919
## sigma_VO2change             0.09      0.01     0.08     0.11 1.00    10546
##                         Tail_ESS
## sigma_Hippocampuschange     5792
## sigma_VO2change             5411
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

$$VO2change_i \sim N(\mu_i^v, \sigma^j v)$$
$$\mu_i^v = \beta_0^v + \beta_1 group_i$$
$$hippocampuschange_i \sim N(\mu_i^h, \sigma^h)$$
$$\mu_i^h = \beta_0^h + \beta_2 VO2change_i + \beta_3 group_i$$

$$\beta_0^v, \beta_0^h \sim N(0,1)$$
$$\beta_1, \beta_2, \beta_3 \sim N(0,1)$$
$$\sigma \sim t_3^+(0, 2.5)$$

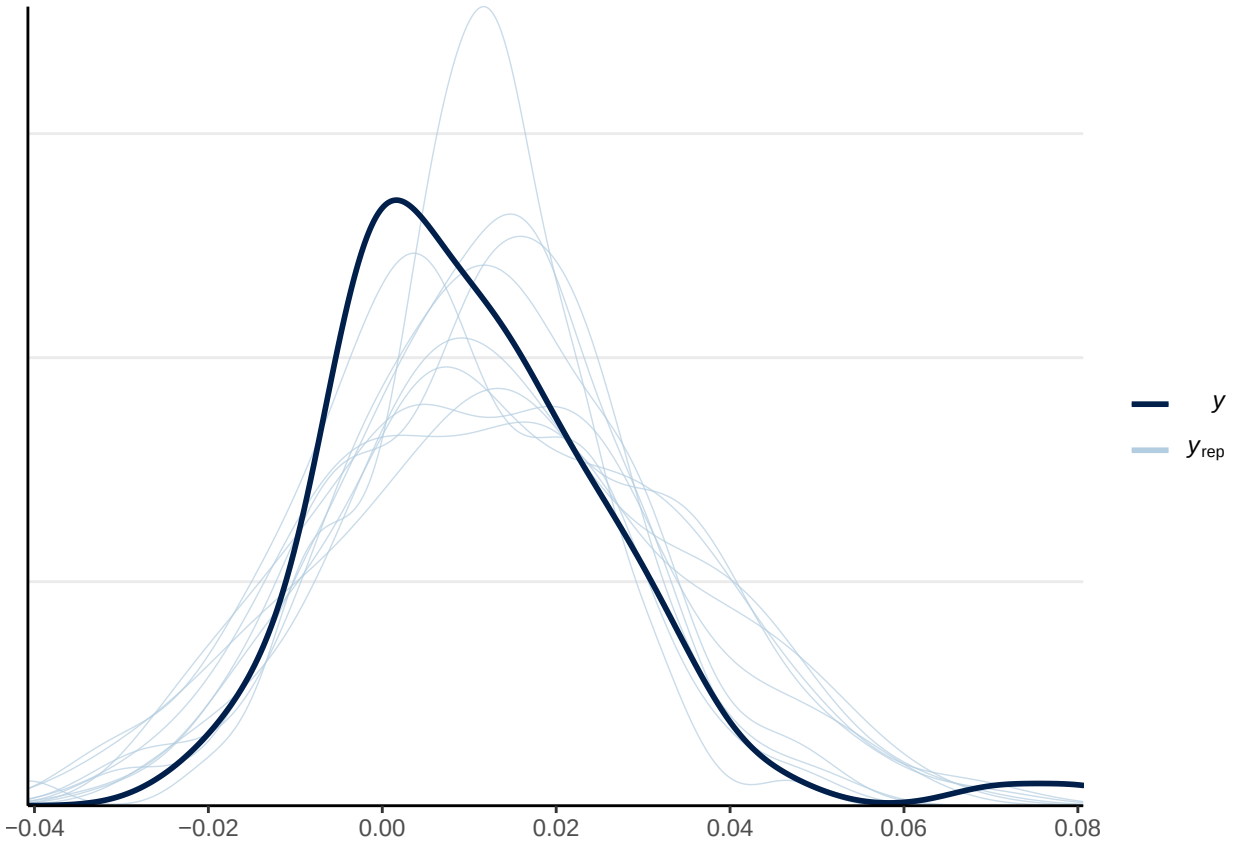
```r
pp_check(m_med, resp = "VO2change")
```

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



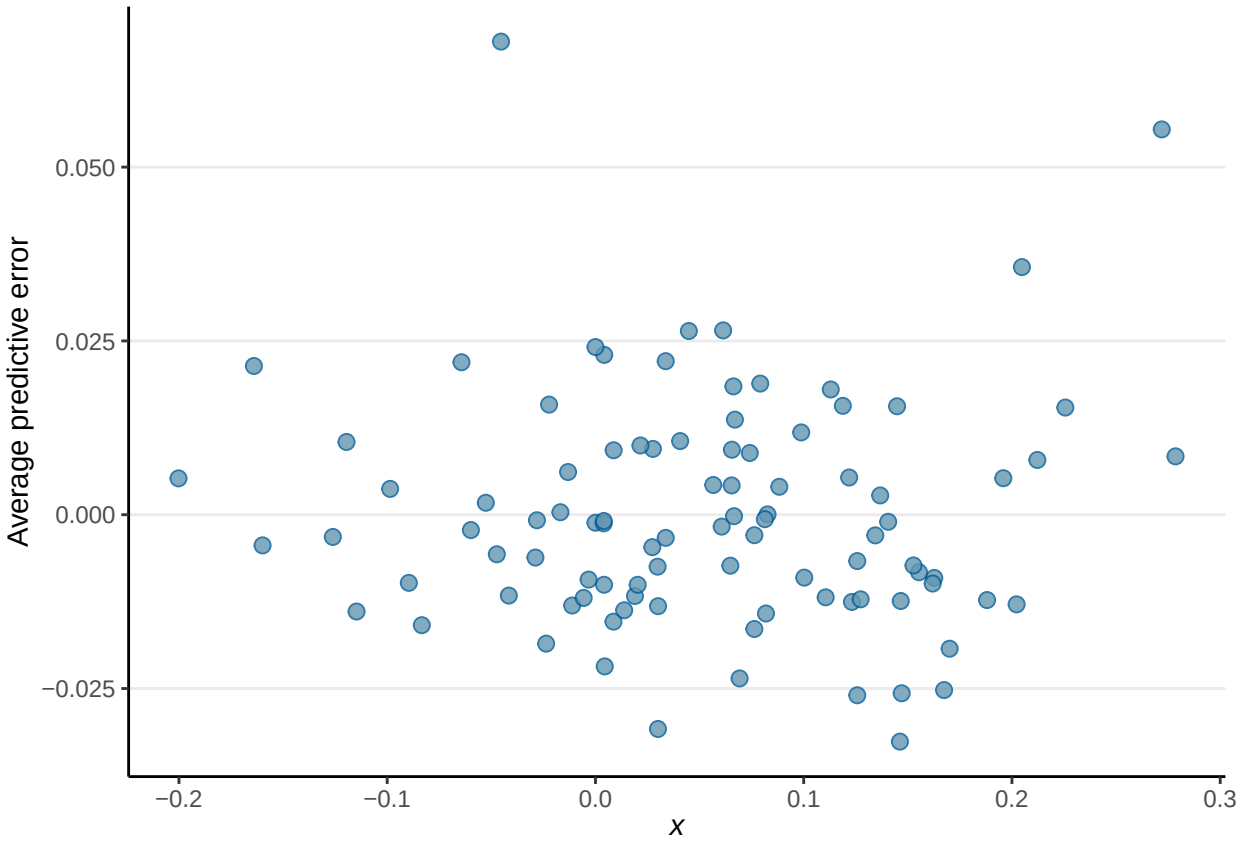

```r
pp_check(m_med, resp = "Hippocampuschange")
```

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```




```
pp_check(m_med, type = "error_scatter_avg_vs_x", resp = "Hippocampuschange",
x = "VO2_change")
```

```
## Using all posterior draws for ppc type 'error_scatter_avg_vs_x' by default.
```


Average predictive error
0.050
0.025
0.000
−0.025
−0.2
−0.1
0.0
0.1
0.2
0.3
x