2024

Example Analyses: Using the SGBA-5 with Simulated Data

Putman & Dogra

Example Analyses: Using the SGBA-5 with Simulated Data Table of Contents

1   Preamble

The example analyses presented in this document are a demonstration of how sex-and gender-based analysis (SGBA) can be conducted using results from the *Sex- and Gender-Based Analysis Tool – 5 item* (SGBA-5) by Putman and Dogra [DOI]. The examples do not cover all potential analyses that could be done using the SGBA-5 but does provide a solid foundation from which researchers can select from and build upon.

More information on the SGBA-5, instructions for its use, and rationale for are included in the SGBA-5's documentation [SUPPLEMENTARY MATERIAL URL]. Initial reliability and validity testing of the SGBA-5 are reported in the paper by Putman, Cole, & Dogra [DOI] and in the thesis work of Putman [PLACE HOLDER FOR THESIS].

2   Data structure

After collecting the data from a study that used the SGBA-5 you will have a dataset that looks

something like this:

Table A1: Example template for cleaned dataset with SGBA-5 results.

pt_id	sex	gen_id	gen_exp	gen_role	gen_rel	…
001	Male	87	81	67	98	…
002	Female	1	0	0	0	…
003	Female	16	22	45	23	…
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Where:

> **pt_id** = the participant identifier
>
> **sex** = the SGBA-5 categorical Biological Sex item (response options: male /
> female / intersex)
>
> **gen_id** = the SGBA-5 Gender Identity gendered aspect of health item (responses:
> 0 to 100 on a feminine to masculine scale; measured in mm if completed on
> paper)
>
> **gen_exp** = the SGBA-5 Gender Expression gendered aspect of health item
> (responses: 0 to 100 on a feminine to masculine scale; measured in mm if
> completed on paper)
>
> **gen_role** = the SGBA-5 Gender Role gendered aspect of health item (responses: 0
> to 100 on a feminine to masculine scale; measured in mm if completed on
> paper)
>
> **gen_rel** = the SGBA-5 Gender Relations gendered aspect of health item (responses:
> 0 to 100 on a feminine to masculine scale; measured in mm if completed on
> paper)

3 Simulated dataset (n = 30)

The simulated dataset (n = 30) was created for this example analysis. In addition to the variables above, it includes:

`sim_num` a simulated continuous variable (mean = 10, sd = 3)

`sim_ord` a simulated ordered variable, 7-point Likert scale (mean = 4, sd = 2)

`sim_binary` a simulated binary categorical variable, recorded as `yes` or `no`
(*probability* of `yes` = 0.67, *probability* of `no` = 0.33)

> The code for reproducing the simulated data in R is included at the end of this analysis and can be found in this document's associated GitHub repository: https://github.com/putman-a/SGBA-5_example_analysis.

4   Single-variable descriptives

Note: for conciseness, the following examples will only show results for two of the four gendered aspects of health items from the SGBA-5 (gender identity, and gender roles).

Visualize the distributions then report univariate statistics (e.g. a table 1).
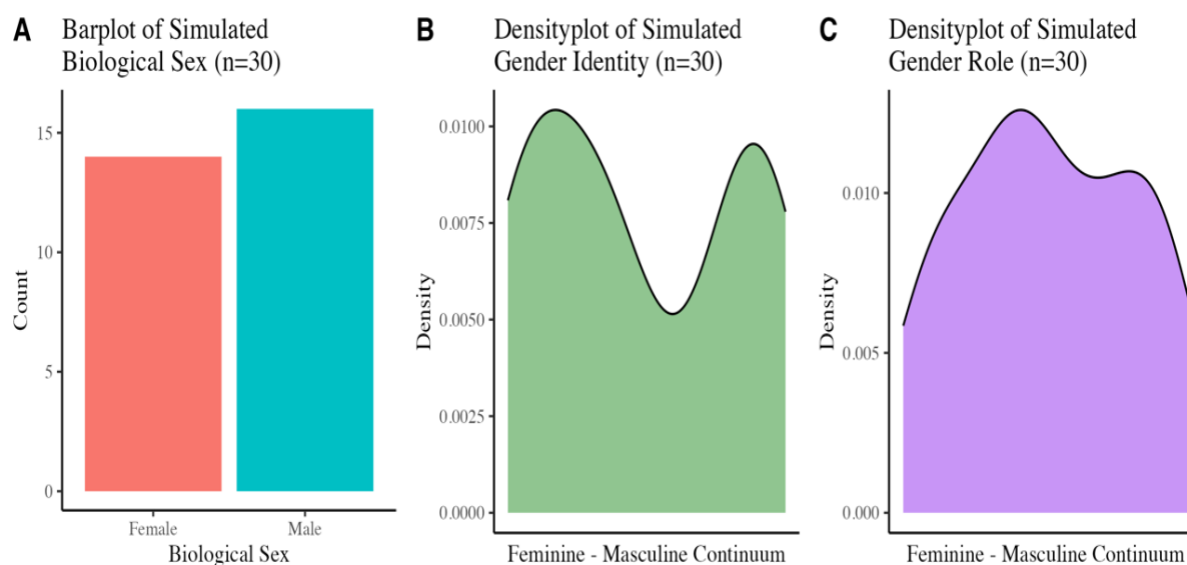
Figure A1: Descriptive plots of simulated sex & gender variables.

In *plot A of **Fig A1*** we see that there are more participants who report their biological sex as assigned as female at birth (n=18) than males (n=12). Looking at the density plots for gender identity (*plot B of **Fig A1***) and gender role (*plot C of **Fig A1***) gendered aspects of health we see that both variables are **bimodal** with one peak closer to the feminine side of the feminine-masculine continuum and one peak closer to masculine end of that continuum. Further, we can also see that overall, participants reported their gender identity and roles as being more feminine. Another potentially important aspect seen in these graphs is that the simulated participants'

distribution of self-reported gender identity is more clearly bimodal than the gender roles aspect of health.

Presently, there is ***no consensus on what descriptive statistics are most appropriate to report bimodal variables*** in health research (the typical mean(sd) or median(IQR) will not accurately represent that there is more than one peak in a bimodal variable's frequency distribution). When taken alongside the SGBA-5's assumption that the feminine-masculine continuum doesn't have a true 0 value, it is our suggestion that ***if researchers decide to report a single variable descriptive statistic for the gendered aspects of health items in the SGBA-5, that they should provide a nominal description of skew along the feminine-masculine continuum*** as their descriptive statistic rather than the numerical average (or other summary statistic). To determine skew of one of the gender variables, researchers suggest calculating the sample's mean score along the feminine-masculine continuum and then classifying the skew using the classifications described in **Table A2**. (**Table A2** assumes you have recorded the gendered aspects of health items as 0 being the most feminine score and 100 being the most masculine score).

Table A2: Potential interpretation of sample means for gendered aspect of health items.

Mean	Interpretation[a]
> 70	*"Skews masculine"*
55 to 70	*"More masculine than feminine"*
45 to 55	*"Not strongly skewed"*
30 to 45	*"More feminine than masculine"*
< 30	*"Skews feminine"*

[a]These classification guidelines are arbitrary and may not be appropriate in all circumstances.

For the simulated dataset represented in the density plots above, the mean score for the gender identity item was 50.2 and 46.8 for the gender role item. This means that when reporting

descriptive statistics on the simulated sample we would report that: "On the whole, the simulated sample was not strongly skewed on a feminine to masculine continuum for either the gender identity or gender role measures from the SGBA-5."

Taking all these together, an example of a sample characteristics table of the SGBA-5 items in the simulated dataset could look as displayed in **Table A3**.

Table A3: Simulated sample characteristics.

	Total (n=30)
Biological Sex (n(%))	
Female	14(47%)
Intersex	-
Male	16(53%)
Gendered Aspect of Health (skew)	
Gender Identity	*Not strongly skewed*
Gender Role	*Not strongly skewed*

5   Sex- and Gender-Based Analysis (SGBA)

The primary purpose of the SGBA-5 is to facilitate SGBA, which will be detailed in this section. Examples are provided comparing the biological sex and gendered aspect of health items to simulated health outcomes that are continuous, ordered (Likert scale), and categorical (binary).

5.1   SGBA with a continuous variable of interest

5.1.1   Biological sex

Starting with a visualization of the continuous variable's distribution disaggregated by sex like the density plot in **Fig A2**, calculate disaggregated summary statistics for the continuous variable disaggregated by sex (**Table A4**), then conduct a statistical test of difference in means (Welch's t-test for this example).
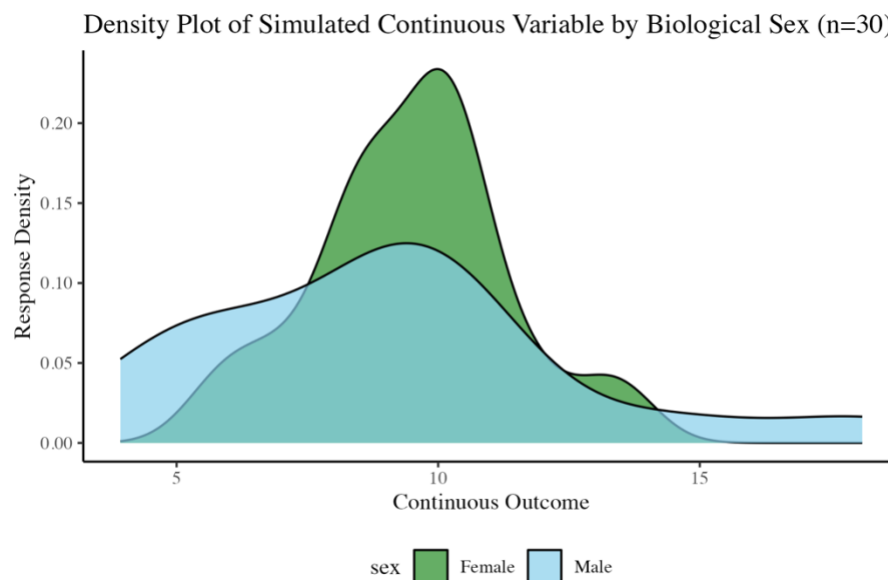
Figure A2: Density plot of simulated continuous variable by biological sex.

Table A4. Summary statistics of a simulated continuous variable disaggregated by sex.

	Female (n=14)	**Male** (n=16)
mean(sd)	9.5(1.9)	9.0(3.6)
median(IQR)	10(2)	9(4)

Welch's 2 sample t-test of biological sex and a continuous variable

$t = 0.44$, 95% CI = (-1.68, 2.58), $df = 23.27$, p > 0.65

∴ we do not have evidence to reject the null hypothesis that there is no significant difference between males and females for the continuous variables.

Looking at the plot and statistics shown above, we see that participants who self-identified their biological sex at birth as male demonstrated more variation (sd=3.6) than females (sd=1.9), but there was no significant difference in continuous variable means between males and females ($t = 0.44$, (95%CI: -1.68, 2.58), $df = 23.27$, p > 0.65).

5.1.2 *Gendered aspect of health: Gender identity*

To analyze the gendered aspects of health from SGBA-5 with a continuous variable of interest, we will generate scatter (***Fig A3 Plot A***) and 2D-density plots (***Fig A3 Plot B***) of the continuous variable by the feminine-masculine continuum, then calculate a Pearson correlation coefficient between each gendered aspect of health item and the continuous variable.
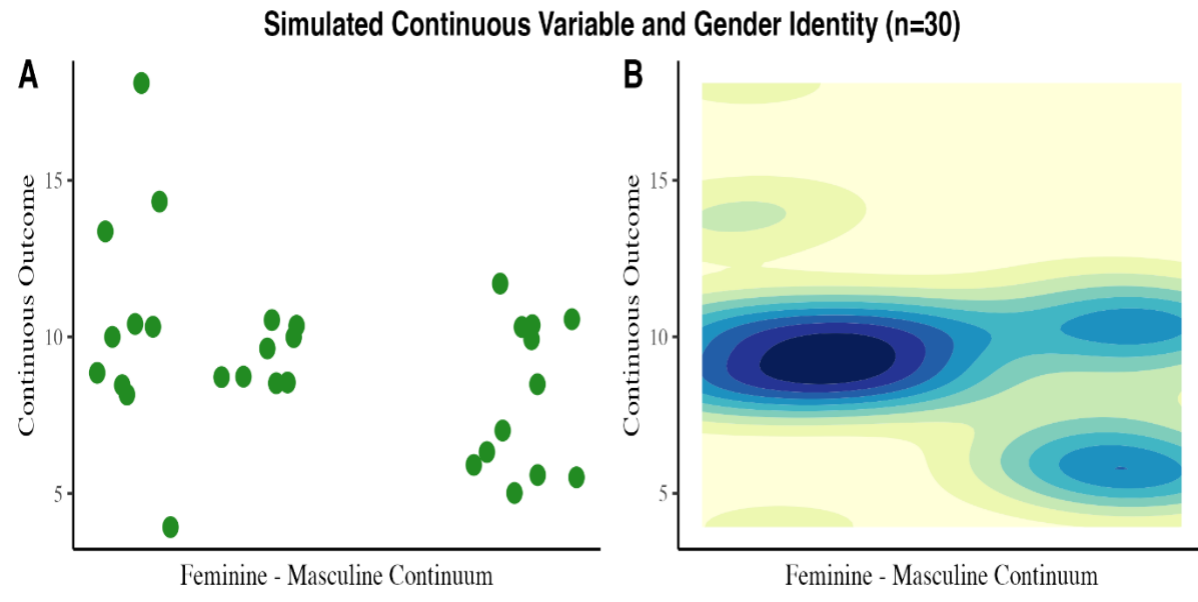
Figure A3: Scatter & 2D density plots of simulated continuous variable by gender identity.

The scatterplot and 2D-density plots show that in the simulated dataset there is a cluster of participants who rated their gender identity as more feminine than masculine and continuous variable scores of approximately 9, but there is not as clear of a cluster in participants who rated their gender identity as more masculine than feminine. Additionally, the simulated participants who reported their gender identity as being more strongly feminine or masculine had much wider variation in continuous outcomes scores than participants who rated their gender identity toward the middle of the feminine-masculine continuum.

> **Pearson's correlation coefficient (r) of gender identity and a continuous variable**
>
> $r = -0.36$, 95% CI = (-0.64, -0.01), $df = 28$, p = 0.048
>
> ∴ since the absolute r value is $|r|$ of 0.36 > 0.3 we found evidence of a weak relationship between differences in gender identity and the continuous outcome.

> *Interpretation of Pearson's r for gendered aspects of health items and a continuous variable*
>
> Since the gendered aspects of health items are ***not assumed to have a true 0 value,*** the results of a Pearson's *r* test can be interpreted in terms of the strength of the relationship (strong, moderate, weak) and direction but ***the exact value of the correlation coefficient should not be directly interpreted*** or assumed to be accurate to the degree a correlation coefficient would be when two continuous variables are being compared.

5.1.3 Gendered aspect of health: Gender roles
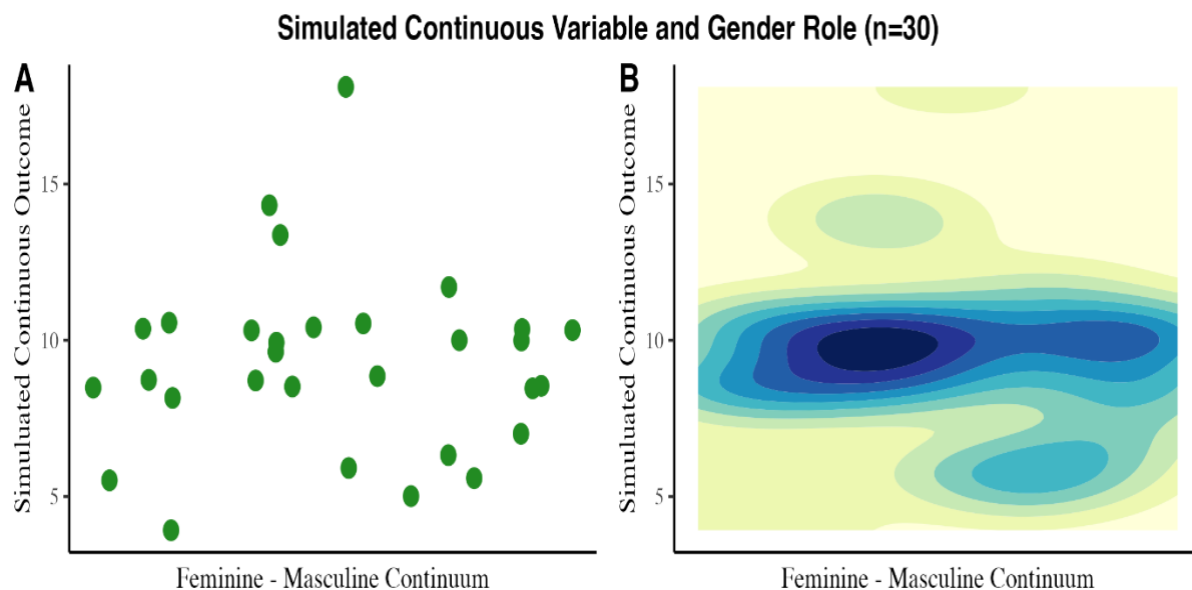
Figure A4: Scatter plot of simulated continuous variable by gender role.

The scatterplot (*Plot A of **Fig A4***) and 2D-density plot (*Plot B of **Fig A4***) show that in the simulated dataset there is a cluster of participants who rated their gender role as more feminine than masculine and continuous variable scores of approximately 9. In comparison to the scatterplot and

2D-density plots of the continuous variable (**Fig A3**) and the gender identity item, the simulated participants who rated their gender roles as less strongly masculine or feminine had continuous variable scores closer to the mean than those who rated their gender identity as strongly masculine than feminine.

> **Pearson's correlation coefficient (r) of gender role and a continuous variable**
>
> $r = 0.02$, 95% CI = (-0.34, 0.38), $df = 28$, p > 0.9
>
> ∴ since the absolute r value is $|r|$ of $0.02 < 0.3$ we did not find evidence of a meaningful relationship between differences in gender role and the simulated continuous outcome.

5.2   SGBA with an ordinal variable of interest

5.2.1   Biological sex

Visualization of the ordinal variable's distribution disaggregated by sex like the density plot in **Fig A5**, calculate disaggregated summary statistics for the ordinal variable disaggregated by sex in **Table A5**, then conduct a statistical test of difference, a chi-square test (X^2) in this example.

Bar Plot of Simulated Likert Outcome and Sex Assigned
at Birth(n=30)
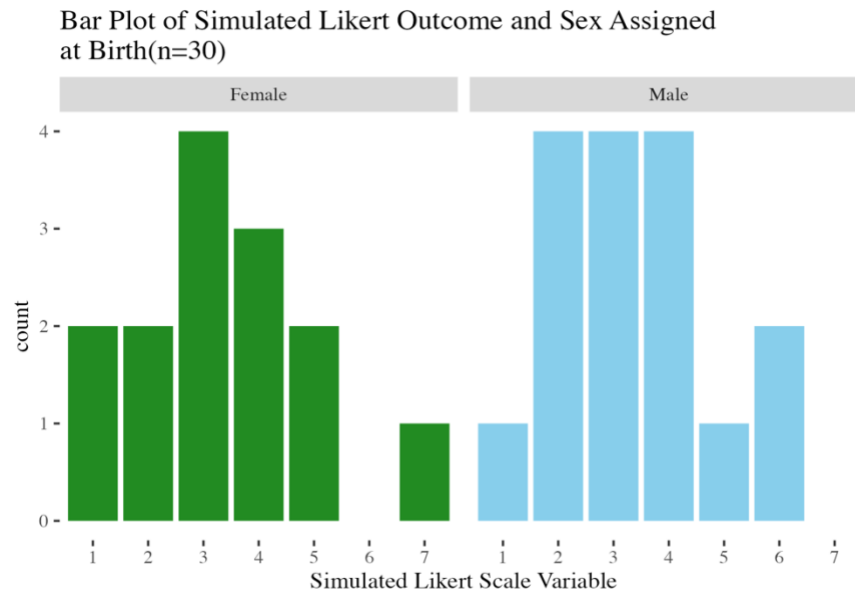
Figure A5: Density plot of simulated ordinal variable by biological sex.

Table A5: Summary statistics of a simulated ordinal variable disaggregated by sex.

	Female (n=14)	Male (n=16)
median(IQR)	3(2)	3(2)

Chi-square test of biological sex and an ordinal variable

$$X^2 = 4.36, df = 6, p > 0.6$$

∴ we do not have evidence* to reject the null hypothesis that there is no relationship between ordinal (Likert) scores and biological sex.

*As can be inferred from *Fig A5* above, the simulated sample does not have expected frequencies of at least 1 for the ordinal (Likert) scores of 6 or 7, one of the assumptions for chi-square tests. In this case the interpretation is still accurate, but more caution in interpretation would be needed if the chi-square test suggested that there was a statistically significant relationship between the simulated ordinal variable and biological sex variable.

Looking at the plot and statistics shown above we can see that there does not appear to be any significant differences in ordinal scores between male and females. Similarly, there was no evidence of a significant relationship found between ordinal scores for males and females.

5.2.2   Gendered aspect of health: Gender identity

To analyze the gendered aspects of health from SGBA-5 with an ordinal variable of interest, we will generate scatter (*Fig A6* *Plot A*) and 2D-density plots (*Fig A6* *Plot B*) of the ordinal (Likert) variable by the gender identity variable on a feminine-masculine continuum, then calculate a Pearson correlation coefficient between the gender identity item and the ordinal variable.
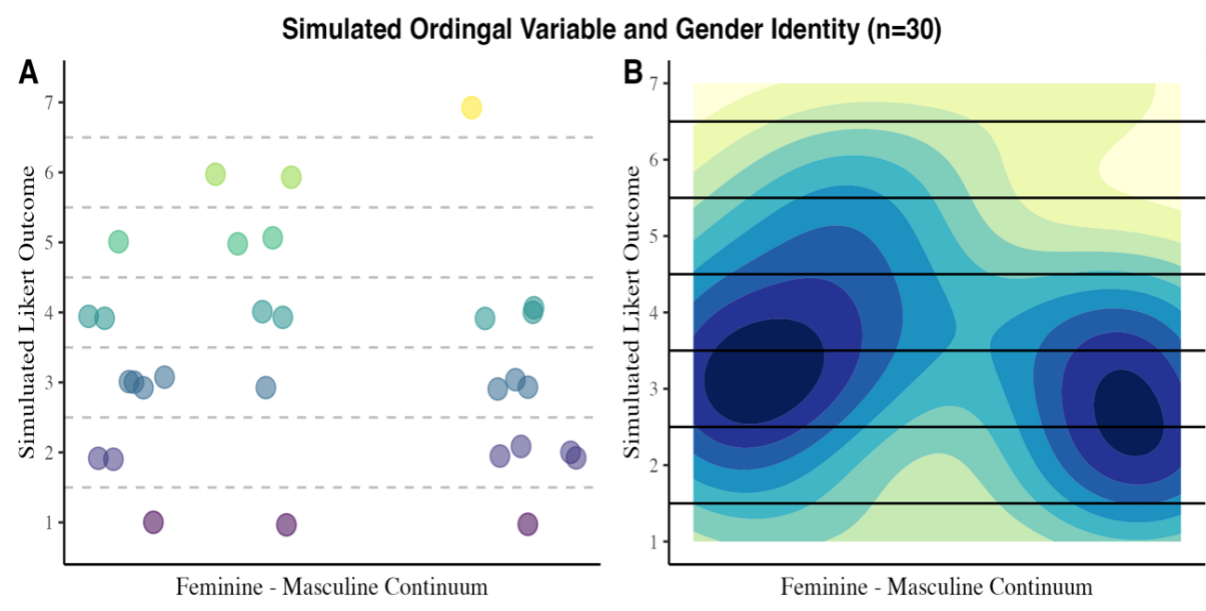
Figure A6: Scatter & 2D density plots of simulated ordinal variable by gender identity.

The scatterplot and 2D-density plots show that in the simulated dataset there is a cluster of participants who rated their gender identity as more feminine than masculine with ordinal variable scores of around 3 & 4, and there is a cluster in participants who rated their gender identity as more

masculine than feminine around ordinal scores of 2 & 3. Additionally, the simulated participants who reported their gender identity as being most strongly masculine showed less variation in ordinal scores.

Pearson's correlation coefficient (r) of gender identity and a continuous variable

$$r = -0.11, 95\%CI\ (-0.45, 0.26), df = 28, p > 0.5$$

∴ since the absolute r value is $|r|$ of 0.11 < 0.3 we did not find evidence of a relationship

between differences in gender identity and the ordinal outcome.

5.2.3 *Gendered aspect of health: Gender roles*
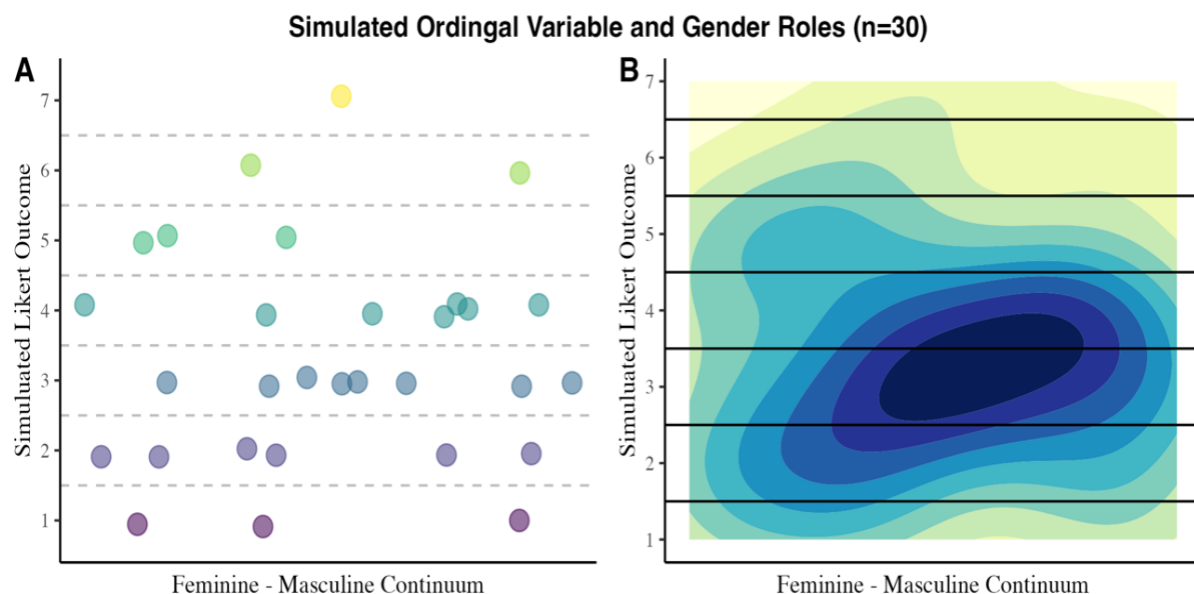
Figure A7: Scatter plot of simulated ordinal variable by gender role.

The scatterplot (*Plot A of **Fig A7***) and 2D-density plot (*Plot B of **Fig A7***) show that in the simulated dataset there is a cluster of participants who rated their gender role slightly more masculine than feminine with ordinal variable scores of 3 & 4. In comparison to the scatterplot and 2D-density

plots of the ordinal variable (***Fig A6***) and the gender identity item, the simulated participants who rated their gender roles as less strongly masculine or feminine had continuous variable scores closer to the ordinal variable's median than those who rated their gender identity as strongly masculine or feminine.

Pearson's correlation coefficient (r) of gender role and an ordinal variable

$$r = 0.04, 95\% CI\ (-0.32, 0.39), df = 28, p > 0.8$$

∴ since the absolute *r* value is |*r*| of 0.04 < 0.3 we did not find evidence of a meaningful relationship between differences in gender role and the simulated continuous outcome.

5.3 SGBA with a binary variable of interest

5.3.1 *Biological sex*

Visualization of the binary variable's distribution disaggregated by sex like the density plot in ***Fig A8***, present a frequency table for the ordinal binary and sex in **Table A6**, then conduct a statistical test of difference, a chi-square test (X^2) in this example.

Figure A8: Density plot of simulated binary variable by biological sex.

Table A6: Frequency Table of a simulated ordinal variable and sex.

	Female (n=14)	Male (n=16)
'yes'	9	11
'no'	5	5

Chi-square test of biological sex and an ordinal variable

$$X^2 < 0.01, df = 1, p = 1$$

$\therefore$ we do not have evidence to reject the null hypothesis that there is no relationship between binary outcome and biological sex.

Looking at the plot and statistics shown above we can see that there does not appear to be any significant differences in ordinal scores between male and females. Similarly, there was no evidence of a significant relationship found between ordinal scores for males and females.

5.3.2 *Gendered aspect of health: Gender identity*

To analyze the gendered aspects of health from SGBA-5 with an ordinal variable of interest, we will generate a density (*Fig A9*) of the binary variable by the gender identity variable on a feminine-masculine continuum, then calculate a t-test between the gender identity item and the binary variable.
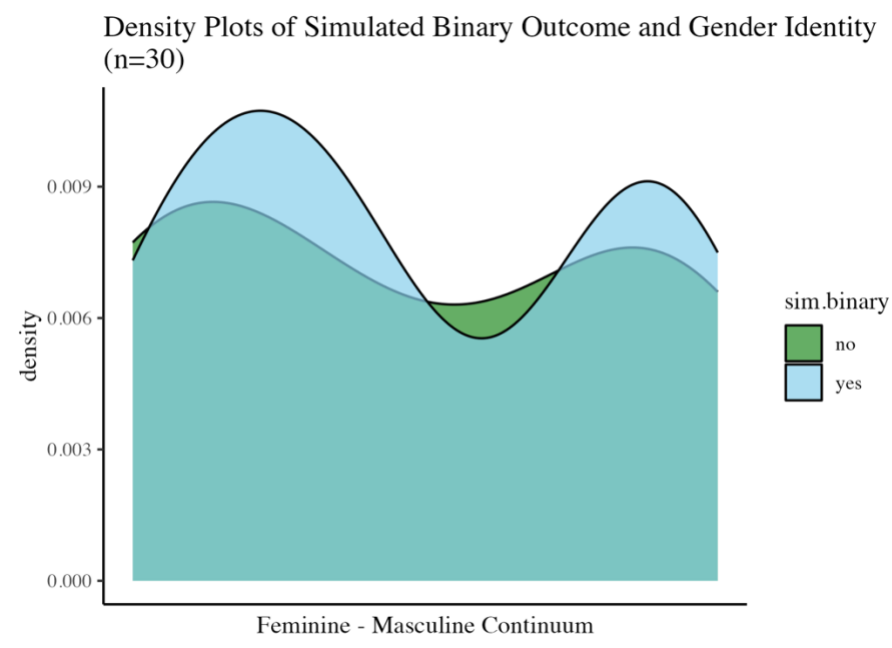
Figure A9: Density plot of simulated binary variable by gender identity.

The density plot show that in the simulated dataset there is a stronger bimodal for the 'yes' response of the binary outcome, but there is not much difference between participants who rated their gender identity as more masculine than feminine and vice versa.

Welch's 2 sample t-test of gender identity and a binary variable

$$t = -0.12, 95\% CI\ (-31.64, 28.34), df = 16.38, p > 0.9$$

$\therefore$ we do not have evidence to reject the null hypothesis that there is no significant difference between the simulated binary outcome and gender identity variables.

5.3.3   Gendered aspect of health: Gender roles
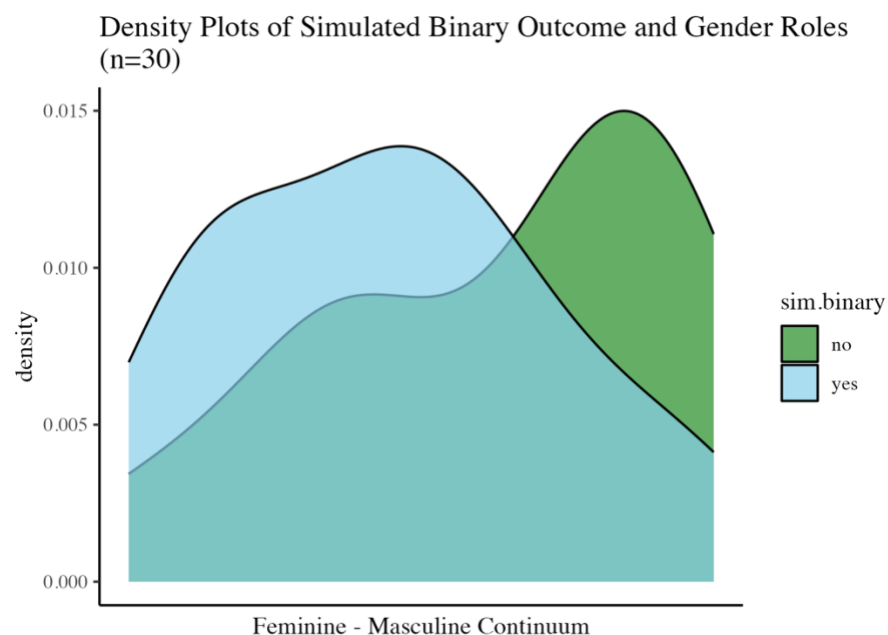
Figure A10: Scatter plot of simulated ordinal variable by gender role.

The density plot (***Fig A10***) show more participants with more masculine gender roles were more likely to also have the binary outcome 'no' and participants with the binary outcome 'yes' was

occurred more often in participants with less strong masculine or feminine gender roles. In comparison to the density plot of the binary variable (***Fig A9***) and the gender identity item, the simulated participants who rated their gender roles as more strongly masculine or feminine had more unimodal relations to the binary outcomes.

Welch's 2 sample t-test of gender identity and a binary variable

$$t = 1.53, 95\%CI\ (-5.77, 36.26), df = 16.71, p > 0.14$$

$\therefore$ we do not have evidence to reject the null hypothesis that there is no significant difference between the simulated binary outcome and gender identity variables.

---

6   R code used in example analysis

```
################################################################
The following R code scripts were created with R version 4.3.2 (2023-10-31) --
"Eye Holes" on Platform: aarch64-apple-darwin20 (64-bit). These scripts are
also available on GitHub at the following url but will not be maintained or
updated    for    future    R    versions:    https://github.com/putman-a/SGBA-
5_example_analysis
################################################################
```

6.1.1   Simulate data (n=30)

```
################################################################
##################################################################################
#####                      Simulate Data for SGBA                         #####
#####                                                                     #####
#####                          by: A Putman                               #####
##################################################################################
##### load libraries #####
library(tidyverse)

# set seed for reproducibility
set.seed(42)

##### write simulation functions #####
# function for simulating bimodal gender variables (loosely based on test-
retest sample)
binom.sim.bounded <- function(n, binom.prop, mean1, sd1, mean2, sd2,
      lower.lim = -Inf, upper.lim = Inf, rounding.decimal){
  # error checking
  if(binom.prop > 1 || binom.prop < 0){stop("binomial proportion should be
      between 0 and 1")}
  if(n < 1){stop("n must be greater than or equal to 1")}
  if(sd1 < 0 || sd2 < 0){stop("standard deviations must be non-negative")}
  # create index
  index <- rbinom(n, size = 1, prob = binom.prop)
  # simulate each peak using a bounded rnorm
  sim.sample.1 <- index * qnorm(runif(n, pnorm(lower.lim, mean1, sd1),
      pnorm(upper.lim, mean1, sd1)), mean1, sd1)
  sim.sample.0 <- (1-index) * qnorm(runif(n, pnorm(lower.lim, mean2, sd2),
      pnorm(upper.lim, mean2, sd2)), mean2, sd2)
  sim.sample <- sim.sample.0 + sim.sample.1
```

```r
  sim.sample <- round(sim.sample, rounding.decimal)
  sim.sample
}

# function for simulating Likert/ordinal outcome variable
ord.sim <- function(n, mean, sd, lower.lim = -Inf, upper.lim = Inf){
  sim <- qnorm(runif(n, pnorm(lower.lim, mean, sd), pnorm(upper.lim, mean,
      sd)), mean, sd)
  sim <- round(sim, 0)
  sim
}

##### simulate data with a sample size of 30 #####
# biological sex - categorical: female, intersex, male
sex <- factor(sample(c('Female', 'Male'), 30, replace = TRUE, prob=c(0.6,
      0.4)))

# gender identity - ordered: 0:100 (from feminine to masculine)
gen_id <- binom.sim.bounded(
  n = 30, binom.prop = 0.6, mean1 = 20, sd1 = 20, mean2 = 85, sd2 = 15,
  lower.lim = 0, upper.lim = 100, rounding.decimal = 0
  )

# gender role - ordered: 0:100 (from feminine to masculine)
gen_role <- binom.sim.bounded(
  n = 30, binom.prop = 0.6, mean1 = 25, sd1 = 20, mean2 = 80, sd2 = 20,
  lower.lim = 0, upper.lim = 100, rounding.decimal = 0
  )

# simulated outcome1 - continuous: mean of 10 with sd of 3
sim_num <- rnorm(n = 30, mean = 10, sd = 3)

# simulated outcome2 - ordered: 7-point Likert scale, mean of 4 with sd of 2
sim_ord <- ordered(ord.sim(n = 30, mean = 4, sd = 2, lower.lim = 1,
      upper.lim = 7))

# simulated outcome3 - binary category: achieved (yes, no)
sim_binary <- factor(sample(c('yes','no'), 30, replace = T, prob =
      c(0.67,0.33)))

# combine into dataframe
sim_data <- tibble(sex, gen_id, gen_role, sim_num, sim_ord, sim_binary)
```

```
# save simulated data frame
write_rds(sim_data, file = "~/[DESIREDFOLDER]/sim.data.RDS") # replace
[DESIREDFOLDER] with folder to save data set into


############################################################
```

6.1.2   SGBA of simulated data (n=30)

```
############################################################
################################################################
#####                    SGBA of Simulated Data                    #####
#####                                                               #####
#####                         by: A Putman                          #####
################################################################


##### load libraries and simulated data #####
# load libraries
library(tidyverse)
library(ggpubr)
library(extrafont)  # for producing plots with Times New Roman Font

# set seed for reproducibility
set.seed(42)

# load data set (from RDS if using previous script to simulate the data)
sim_df <- read_rds(file = '[SIMULATEDDATASET].RDS')  # replace
      [SIMULATEDDATASET] with location of data set

#####  store plot theme pre-sets  #####
# store plot aspects for easy reuse
plot.theme <-
  theme_classic() +
  theme(
    text = element_text(family = 'serif'),
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )

# default density plots and scale
plot.2d.density <- function(data, x, y){
  z <- ggplot(data = data, aes(x = x, y = y)) +
    geom_density_2d_filled(aes(fill = after_stat(level)),
                           bins = 9, contour_var = 'ndensity') +
```

```
        scale_fill_brewer(palette = 16, direction = 1,
                         labels = c('Low', '', '', '','','','','', 'High')) +
      guides(fill = guide_legend(title = 'Density of Occurrences',
                               direction = 'horizontal', nrow = 1,
                               label.position = 'bottom'))
  z
}



##### univariate plots #####
# biological sex bar plot
sex_bar <- ggplot(sim_df, aes(x = sex)) +
  geom_bar(aes(fill = sex), show.legend = F) +
  labs(title = 'Bar plot of Simulated\nBiological Sex (n=30)') +
  ylab('Count') + xlab('Biological Sex') +
  theme_classic() +
  theme(text = element_text(family = 'serif'))
sex_bar  # view plot

# gender identity density plot
gi_dens <- ggplot(sim_df, aes(x = gen_id)) +
  geom_density(alpha=.5, fill = 'forestgreen', colour = 'black') +
  labs(title = 'Density plot of Simulated\nGender Identity (n=30)') +
  ylab('Density') + xlab('Feminine - Masculine Continuum') +
  plot.theme
gi_dens  # view plot

# gender role density plot
grol_dens <- ggplot(sim_df, aes(x = gen_role)) +
  geom_density(alpha=.5, fill = 'purple2', colour = 'black') +
  labs(title = 'Densityplot of Simulated\nGender Role (n=30)') +
  ylab('Density') + xlab('Feminine - Masculine Continuum') +
  plot.theme
grol_dens  # view plot



##### SGBA #####

##### continuous variable by biological sex example #####
# density plot
con_sex_dens <- ggplot(sim_df, aes(x = sim_num, fill = sex)) +
  geom_density(alpha = 0.7) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
```

```r
  labs(title = 'Density Plot of Simulated Continuous Variable by Biological
       Sex (n=30)') +
  xlab('Continuous Outcome') + ylab('Response Density') +
  theme_classic() +
  theme(legend.position = 'bottom', text = element_text(family = 'serif'))
con_sex_dens  # view plot

# summary stats
con_sex_sum <- sim_df %>% group_by(sex) %>%
  summarise(
    n = n(),
    `mean continuous` = round(mean(sim_num),1),
    `SD continuous` = round(sd(sim_num),2),
    `median continuous` = round(median(sim_num),0),
    `IQR continuous` = round(IQR(sim_num),)
  ) %>%
  rename(`biological sex` = sex)
con_sex_sum    # view stats
# t test
con_sex_ttest <- t.test(sim_num ~ sex, data = sim_df)
con_sex_ttest   # view test results


##### continuous variable by gender identity example #####
# scatter plot
num_gi_scatter <- ggplot(data = sim_df, aes(x = gen_id, y = sim_num)) +
  geom_jitter(size = 3, colour = 'forestgreen') +
  #  labs(title = 'Scatterplot of Simulated Numerical Outcome\n& Gender
       Identity (n=30)') +
  ylab('Continuous Outcome') + xlab('Feminine - Masculine Continuum') +
  plot.theme
num_gi_scatter  # view plot

# 2D density plot
num_gi_density2d <- plot.2d.density(sim_df, gen_id, sim_num) +
  #  labs(title = '2D Density of Simulated Numerical\nOutcome & Gender
       Identity (n=30)') +
  ylab('Continuous Outcome') + xlab('Feminine - Masculine Continuum') +
  plot.theme
num_gi_density2d  # view plot

# Pearson's r
num_gi_cor <- cor.test(sim_df$sim_num, sim_df$gen_id, method='pearson')
```

```
num_gi_cor    # view test results


##### continuous variable by gender role example #####
# scatter plot
num_role_scatter <- ggplot(sim_df, aes(x = gen_role, y = sim_num)) +
  geom_jitter(size = 3, colour = 'forestgreen') +
  #  labs(title = 'Scatterplot of Simulated Numerical Outcome\n& Gender
      Role') +
  ylab('Simulated Continuous Outcome') + xlab('Feminine - Masculine
Continuum') +
  plot.theme
num_role_scatter  # view plot


# 2D density plot
num_role_density2d <- plot.2d.density(sim_df, gen_role, sim_num) +
  #  labs(title = '2D Density of Simulated Numerical\nOutcome & Gender Role
      (n=30)') +
  ylab('Simulated Continuous Outcome') + xlab('Feminine - Masculine
      Continuum') +
  plot.theme
num_role_density2d  # view plot


# Pearson's r
num_grol_cor <- cor.test(sim_df$sim_num, sim_df$gen_role, method='pearson')
num_grol_cor    # view test results



##### ordinal variable by biological sex example #####
# bar plot
ord_sex_bar <- ggplot(sim_df, aes(x = sim_ord, fill = sex)) +
  geom_bar() +
  facet_wrap(~ sex) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(title = 'Bar Plot of Simulated Likert Outcome and Sex Assigned\nat
      Birth(n=30)') +
  xlab('Simulated Likert Scale Variable') +
  theme(panel.grid = element_blank(),panel.background = element_blank(),
      legend.position = 'none', text = element_text(family = 'serif'))
ord_sex_bar  # view plot


# summary stats
ord_sex_sum <- sim_df %>% group_by(sex) %>%
```

```r
  summarise(
    n = n(),
    `median Likert` = round(median(as.numeric(sim_ord)),0),
    `IQR Likert` = round(IQR(as.numeric(sim_ord)),)
  ) %>%
  rename(`biological sex` = sex)
ord_sex_sum    # view stats


# chi-square test
ord_sex_chi <- chisq.test(sim_df$sim_ord,sim_df$sex)
ord_sex_chi    # view test results



##### ordinal variable by gender identity example #####
# jitter
ord_gi_jit <- ggplot(sim_df, aes(x = gen_id, y = sim_ord)) +
  geom_jitter(aes(colour = sim_ord, alpha = 0.7), size = 4, height = 0.1) +
  # labs(title = 'Scatterplot of Simulated Likert Outcome\n& Gender Identity
      (n=30)') +
  xlab('Feminine - Masculine Continuum') + ylab('Simulated Likert Outcome') +
  geom_hline(yintercept = c(1.5, 2.5, 3.5, 4.5, 5.5, 6.5), linetype='dashed',
      colour='grey') +
  plot.theme
ord_gi_jit  # view plot


# 2D Density
ord_gi_den <- plot.2d.density(sim_df, gen_id, as.numeric(sim_ord)) +
  scale_y_continuous(breaks = c(1,2,3,4,5,6,7), labels = c(1,2,3,4,5,6,7),
      limits = c(1,7)) +
  # labs(title = '2D Density of Simulated Likert Outcome\n& Gender Identity
      (n=30)') +
  xlab('Feminine - Masculine Continuum') + ylab('Simulated Likert Outcome') +
  geom_hline(yintercept = c(1.5, 2.5, 3.5, 4.5, 5.5, 6.5)) +
  plot.theme
ord_gi_den  # view plot


## Pearson's r
ord_gi_cor <- cor.test(as.numeric(sim_df$sim_ord), sim_df$gen_id,
      method='pearson')
ord_gi_cor    # view test results
```

```r
##### ordinal variable by gender role example #####
# jitter
ord_gr_jit <- ggplot(sim_df, aes(x = gen_role, y = sim_ord)) +
  geom_jitter(aes(colour = sim_ord, alpha = 0.7), size = 4, height = 0.1) +
  # labs(title = 'Scatterplot of Simulated Likert Outcome\n& Gender Role
      (n=30)') +
  xlab('Feminine - Masculine Continuum') + ylab('Simulated Likert Outcome') +
  geom_hline(yintercept = c(1.5, 2.5, 3.5, 4.5, 5.5, 6.5), linetype='dashed',
      colour='grey') +
  plot.theme
ord_gr_jit  # view plot


## 2D Density
ord_gr_den <- plot.2d.density(sim_df, gen_role, as.numeric(sim_ord)) +
  scale_y_continuous(breaks = c(1,2,3,4,5,6,7), labels = c(1,2,3,4,5,6,7),
      limits = c(1,7)) +
  # labs(title = '2D Density of Simulated Likert Outcome\n& Gender Role
      (n=30)') +
  xlab('Feminine - Masculine Continuum') + ylab('Simulated Likert Outcome') +
  geom_hline(yintercept = c(1.5, 2.5, 3.5, 4.5, 5.5, 6.5)) +
  plot.theme
ord_gr_den  # view plot


## Pearson's r
ord_gr_cor <- cor.test(as.numeric(sim_df$sim_ord), sim_df$gen_role,
      method='pearson')
ord_gr_cor   # view test results



##### binary variable by biological sex example #####
# bar plot
binary_sex_bar <- ggplot(sim_df, aes(x = sim_binary, fill = sex)) +
  geom_bar() +
  facet_wrap(~ sex) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(title = 'Bar Plot of Simulated Binary Outcome and Sex Assigned\nat
      Birth(n=30)') +
  xlab('Simulated Binary Variable') +
  theme(panel.grid = element_blank(),panel.background = element_blank(),
        legend.position = 'none', text = element_text(family = 'serif'))
binary_sex_bar  # view plot
```

```r
# frequency table
binary_sex_freq <- table(sim_df$sim_binary, sim_df$sex)
binary_sex_freq   # view stats


# chi-square test
binary_sex_chi <- chisq.test(sim_df$sim_binary,sim_df$sex)
binary_sex_chi    # view test results



##### binary variable by gender identity example #####
# bar plot
binary_gi_dens <- ggplot(sim_df, aes(x = gen_id, fill = sim_binary)) +
  geom_density(alpha = 0.7) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(title = 'Density Plots of Simulated Binary Outcome and Gender
      Identity\n(n=30)') +
  xlab('Feminine - Masculine Continuum') +
  theme_classic() +
  theme(
      text = element_text(family = 'serif'),
      axis.ticks.x = element_blank(),
      axis.text.x = element_blank()
  )
binary_gi_dens  # view plot


# t test
binary_gi_ttest <- t.test(gen_id ~ sim_binary, data = sim_df)
binary_gi_ttest   # view test results



##### binary variable by gender roles example #####
# bar plot
binary_grole_dens <- ggplot(sim_df, aes(x = gen_role, fill = sim_binary)) +
  geom_density(alpha = 0.7) +
  # facet_wrap(~ sim_binary) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(title = 'Density Plots of Simulated Binary Outcome and Gender
      Roles\n(n=30)') +
  xlab('Feminine - Masculine Continuum') +
  theme_classic() +
  theme(
    text = element_text(family = 'serif'),
    axis.ticks.x = element_blank(),
```

```
    axis.text.x = element_blank()
  )
binary_grole_dens  # view plot

# t test
binary_grole_ttest <- t.test(gen_role ~ sim_binary, data = sim_df)
binary_grole_ttest   # view test results
```