Earth Agency Project: Progress Report

Consulting Team A, group 3

March 20th, 2021

1. Decisions made in creating the working data frame

(a) We have combined *EarthAgency_Adults_R.csv* and *EarthAgency_Children_R.csv* into one data.frame: `AC_df`. We have done this so that we can make an Adult vs. children comparison. To combined the two data sets we have done b-d.

(b) For this version of the report, we have normalized by using a percentage to get the adults and children's `invitalscores` to have the same range $[0, 1]$. To do this we've calculated the percent-true $new = \frac{old}{5}$ for adults, and $new = \frac{old}{3}$ for children.

adult	0	1	2	3	4	5
normalized	0	0.2	0.4	0.6	0.8	1

child	0	1	2	3
normalized	0	0.33	0.67	1

(c) To normalize the adult and children's `inpsychscore`, we have also used a percent-true score.

adult	0	1	2	3	4	5
normalized	0	0.2	0.4	0.6	0.8	1

child	0	1	2	3
normalized	0	0.33	0.67	1

(d) We have added `FirstLang` to the children's records, and assigned every child a value of 1 for `FirstLang`. We're assuming that 1=using first language, and 0=not using first language.

(e) The 87th entry in the children's record's has a `MeanSeverity` of 2.67. But 2.67 is not a possible value (given that `MeanSeverity` is an average of 4 integer scores, and should thus be a multiple of 0.25). We've changed the `MeanSeverity` for that record to 2.5.

(f) The 41st children's record has no `BIoJtscore` or `AntJtscore`. We did not use this record in the data.frame.

(g) The 45th children's record has no `Agency_Language` or `SRFactsTotal`. We did not use this record in the data.frame.

(h) We have not included the independent variable `SRTotal` in the data frame. Our understanding is that this variable is measuring comprehension of the video and was designed to test if children had payed attention to the video. For the children it was a 4 question test. Of the 91 children, 4 had a score of 2, 28 a score of 3, and 59 a score of 4. And we have not filtered out any children from the data frame based on this comprehension check.

(i) There is also a `SRTotal` variable for the adults. We did not include it in the data frame. If it is also a comprehension check, it might be useful consider filtering out some of the adults based on the results. This is a table of the results:

score	6	7	8	9	10	11	12
count	1	1	15	18	40	37	20

(j) We have renamed the `Condition` levels to `Obj` (object), `Nat` (nature, animal, vitalist), `Per` (person, psychological), for ease of understanding and consistency while we were coding.

(k) For a better model fit, we have combined the original 13 levels of `MeanSeverity` into three levels, as follows:

original	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
combined	1	1	1	1	1	1	2	2	2	3	3	3	3

(l) For a better model fit, we have combined the original 5 levels of `BioJtscore` (the number of scenarios where the respondent used a biocentric justification), into three levels, as follows:

original	0	1	2	3	4
combined	1	1	1	2	3

(m) For a better model fit, we have combined the original 5 levels of `AntJtscore` (the number of scenarios where the respondent used an anthropocentric justification), into three levels, as follows:

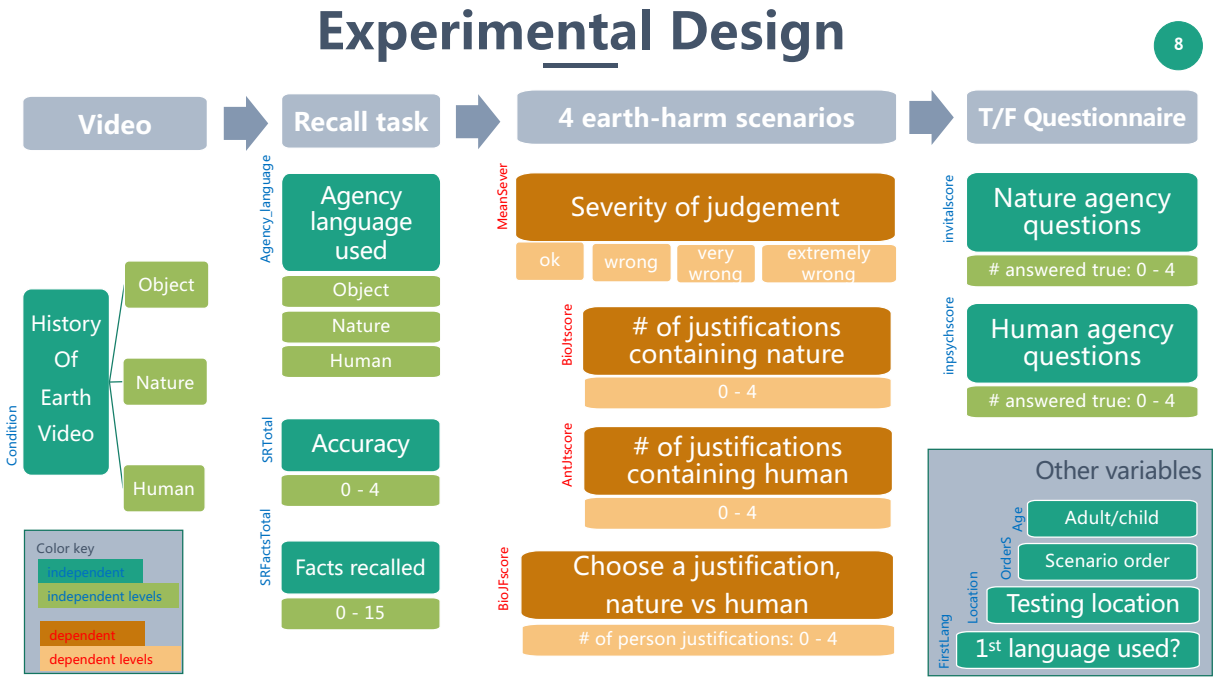
original	0	1	2	3	4
combined	1	1	1	2	3

(n) For a better model fit, we have combined the original 5 levels of `BioJFtotal` (the number of scenarios where the respondent chose biocentric when given a choice between biocentric and anthropocentric), into three levels, as follows:

original	0	1	2	3	4
combined	1	1	2	3	3

2. Experimental Design concerns

The chart below shows our understanding of the variables collected. The green are the independent variables with their levels, and the brown are the four dependent variables. As you have noted, the flow our your experiment design has intertwined the `Condition` variable with the `Agency_language`, `inpsychscore` and `invitalscore` variables. In particular, since the questionnaire was given after the participants watched the video, it's not clear that the `invitalscore` and `inpsychscore` variables are measuring the participants underlying beliefs, or the beliefs expressed in the video that they just watched. Also `Agency_language`, which is attempting to measure how the participants describe the video, overlaps with the perspective of the video watched. As such it is not clear that `Agency_language` is descriptive of the participant or of the video they just watched.



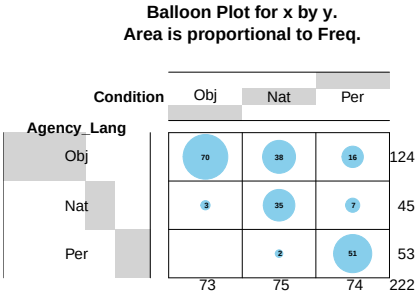
3. Correlation among independent variables

Because of our worry about the independence of the independent variables, we checked the correlations between the independent variables. The following chart (which is redundant across the diagonal), shows Pearson product-moment correlations between the numeric variables (`SRFactsTotal`, `invitalscore`, `inpsychscore`), a polychoric correlations between ordinal categorical variables (`Condition`, `Agency_Language`, `FirstLang` and `Age`), and polyserial correlations between numeric and ordinal variables. 1 or -1 is a strong correlation. 0 is no correlation.

	Condition	Agency_Language	SRFactsTotal	invitalscore	inpsychscore	postqs	FirstLang	Age
Condition	1.00	0.86	0.01	0.09	0.17	0.17	0.02	0.03
Agency_Language	0.86	1.00	0.10	0.05	0.26	0.19	0.14	0.18
SRFactsTotal	0.01	0.10	1.00	0.08	-0.33	-0.12	-0.19	-0.52
invitalscore	0.09	0.05	0.08	1.00	0.34	0.80	-0.13	-0.20
inpsychscore	0.17	0.26	-0.33	0.34	1.00	0.83	0.28	0.58
postqs	0.17	0.19	-0.12	0.80	0.83	1.00	0.03	0.17
FirstLang	0.02	0.14	-0.19	-0.13	0.28	0.03	1.00	0.87
Age	0.03	0.18	-0.52	-0.20	0.58	0.17	0.87	1.00

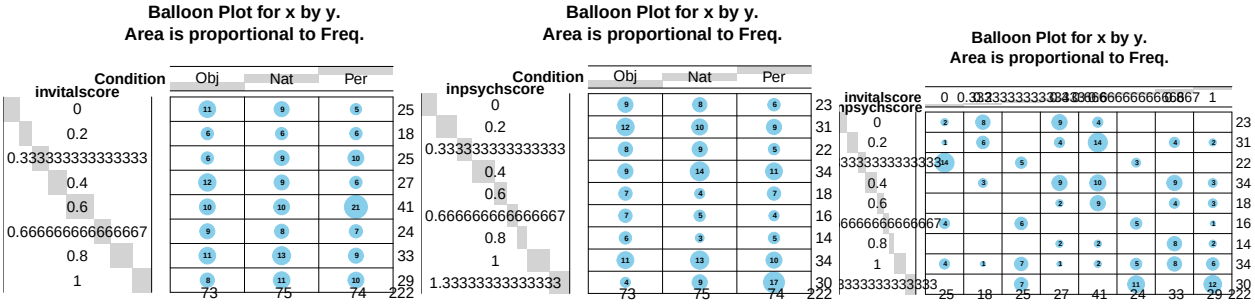
3a. Correlation between `Condition` and `Agency_language`

`Condition` and `Agency_language` had the strongest correlation and the p-value of the Pearson's chi-square test was effectively zero ($p = 7.3{*}10^{-36}$). Because of this, and because of our concerns about the experimental design, we have not included `Agency_language` in the model fits below. If we have time, we will explore using PCA (principal component analysis) to combine these two independent variables.



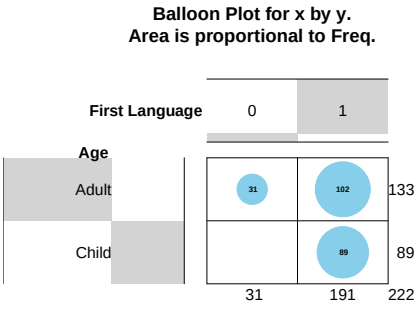
3b. Correlation between `Condition`, `invitalscore`, and `inpsychscore`

A Chi-square test shows that there is not a significant correlation between `Condition` and `invitalscore` (p=0.54), and there is a slightly significant correlation between `Condition` and `inpsychscore` (p=0.04). But the correlation between `invitalscore` and `inpsychscore` is highly significant ($p = 5.0 * 10^{-4}$). We have included both of these variables in the model fit below, and (if time permits) we are going to explore using PCA (principal component analysis) to combine these two independent variables (`invitalscore` and `inpsychsocre`).



3c. Correlation between `FirstLang` and `Age`

Because we are assuming that the 89 children in the study are all using their first language, there is a strong correlation between `FirstLang` and `Age` ($p = 2.4 * 10^{-6}$). But, for contextual reasons, we have left both of these variables in the models below. And when we ran the model without `FirstLang` it did not imporove the model fit.



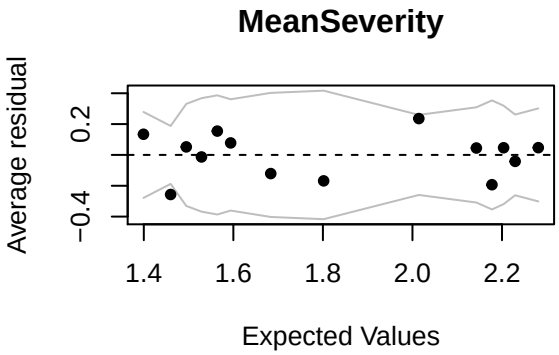
4a. Linear models

Just for MeanSever. AND, why is the ordinal model better?

```
Call:
lm(formula = MeanSever ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.68513 -0.44849  0.01683  0.41194  1.42652

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.623603   0.182912   8.876 2.81e-16 ***
ConditionNat  0.060888   0.109228   0.557   0.5778
ConditionPer  0.075715   0.111191   0.681   0.4966
SRFactsTotal  0.019409   0.015928   1.219   0.2243
invitalscore  0.001812   0.182530   0.010   0.9921
inpsychscore  0.036534   0.156307   0.234   0.8154
FirstLang1   -0.286629   0.136141  -2.105   0.0364 *
AgeChild      0.717942   0.134497   5.338 2.39e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6612 on 214 degrees of freedom
Multiple R-squared:  0.1969,    Adjusted R-squared:  0.1706
F-statistic: 7.494 on 7 and 214 DF,  p-value: 4.551e-08

Call:
lm(formula = MeanSever ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.70393 -0.54285 -0.00458  0.52012  1.41367

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.797236   0.191159   9.402  < 2e-16 ***
ConditionNat  0.066406   0.115996   0.572   0.5676
ConditionPer  0.021374   0.117591   0.182   0.8559
SRFactsTotal -0.004021   0.016260  -0.247   0.8049
invitalscore -0.435355   0.173242  -2.513   0.0127 *
inpsychscore  0.535110   0.133104   4.020 8.05e-05 ***
FirstLang1   -0.078933   0.138554  -0.570   0.5695
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7022 on 215 degrees of freedom
Multiple R-squared:  0.08994,   Adjusted R-squared:  0.06454
F-statistic: 3.541 on 6 and 215 DF,  p-value: 0.002304

Call:
```

```
lm(formula = MeanSever ~ SRFactsTotal + invitalscore + inpsychscore +
    FirstLang + Age, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.67627 -0.43069  0.02903  0.42958  1.37492

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.6596502  0.1748469   9.492  < 2e-16 ***
SRFactsTotal  0.0199637  0.0158374   1.261    0.209
invitalscore -0.0006204  0.1816413  -0.003    0.997
inpsychscore  0.0534142  0.1530991   0.349    0.728
FirstLang1   -0.2858515  0.1354608  -2.110    0.036 *
AgeChild      0.7117047  0.1331929   5.343 2.31e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6589 on 216 degrees of freedom
Multiple R-squared:  0.1949,    Adjusted R-squared:  0.1763
F-statistic: 10.46 on 5 and 216 DF,  p-value: 5.305e-09


Call:
lm(formula = MeanSever ~ Condition + SRFactsTotal + FirstLang +
    Age, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.66845 -0.44035  0.01934  0.41350  1.43854

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.64141    0.16045  10.230  < 2e-16 ***
ConditionNat  0.06300    0.10848   0.581   0.5620
ConditionPer  0.08176    0.10861   0.753   0.4524
SRFactsTotal  0.01896    0.01576   1.203   0.2304
FirstLang1   -0.28848    0.13529  -2.132   0.0341 *
AgeChild      0.73356    0.10571   6.939 4.53e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6583 on 216 degrees of freedom
Multiple R-squared:  0.1966,    Adjusted R-squared:  0.178
F-statistic: 10.57 on 5 and 216 DF,  p-value: 4.298e-09
```

4b. Binned residual plots



4c. ANOVA for linear model of Mean Severity

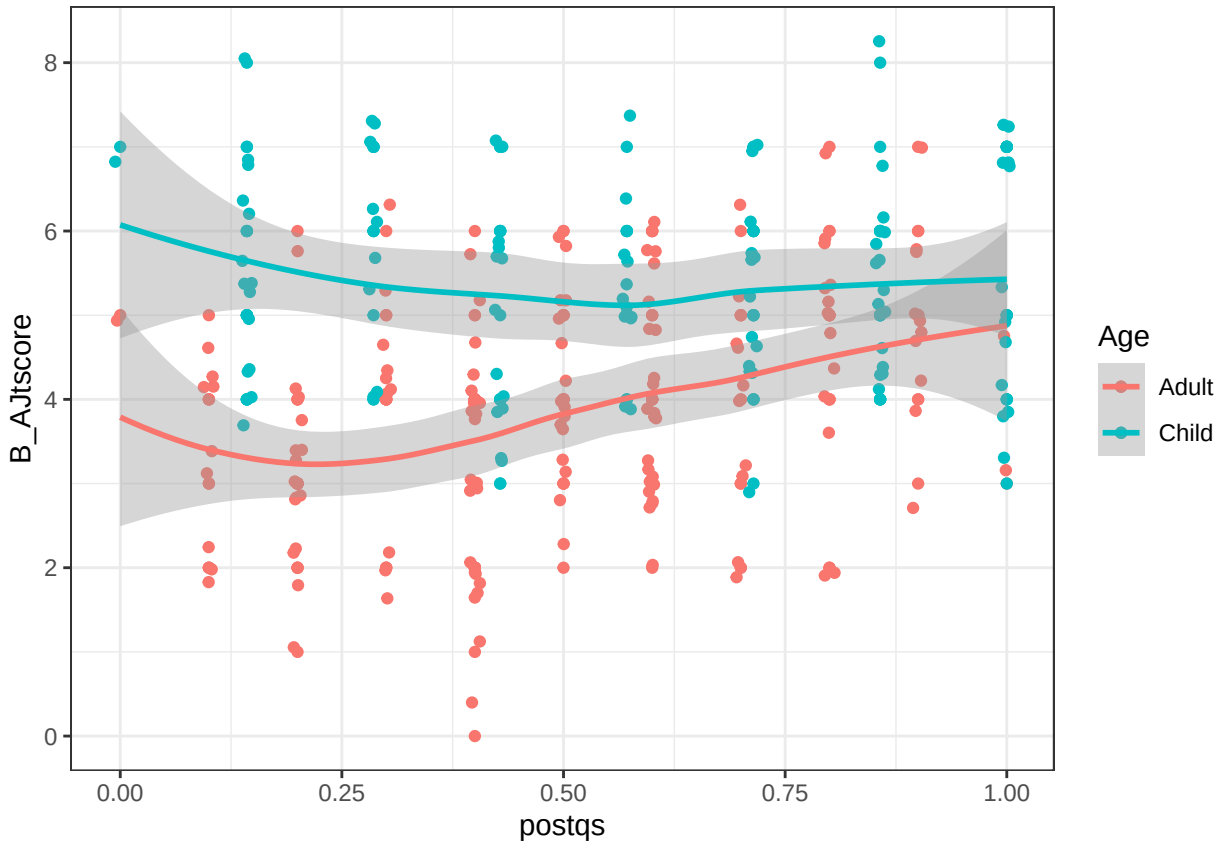
Age is significant. Condition and (invitalscore + inpsychscore) are not.

```
Analysis of Variance Table

Model 1: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang + Age
Model 2: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang
  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1    214  93.561
2    215 106.018 -1   -12.458 28.494 2.39e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analysis of Variance Table

Model 1: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang + Age
Model 2: MeanSever ~ SRFactsTotal + invitalscore + inpsychscore + FirstLang +
    Age
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    214 93.561
2    216 93.790 -2  -0.22946 0.2624 0.7694

Analysis of Variance Table

Model 1: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang + Age
Model 2: MeanSever ~ Condition + SRFactsTotal + FirstLang + Age
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    214 93.561
2    216 93.597 -2 -0.036273 0.0415 0.9594
```

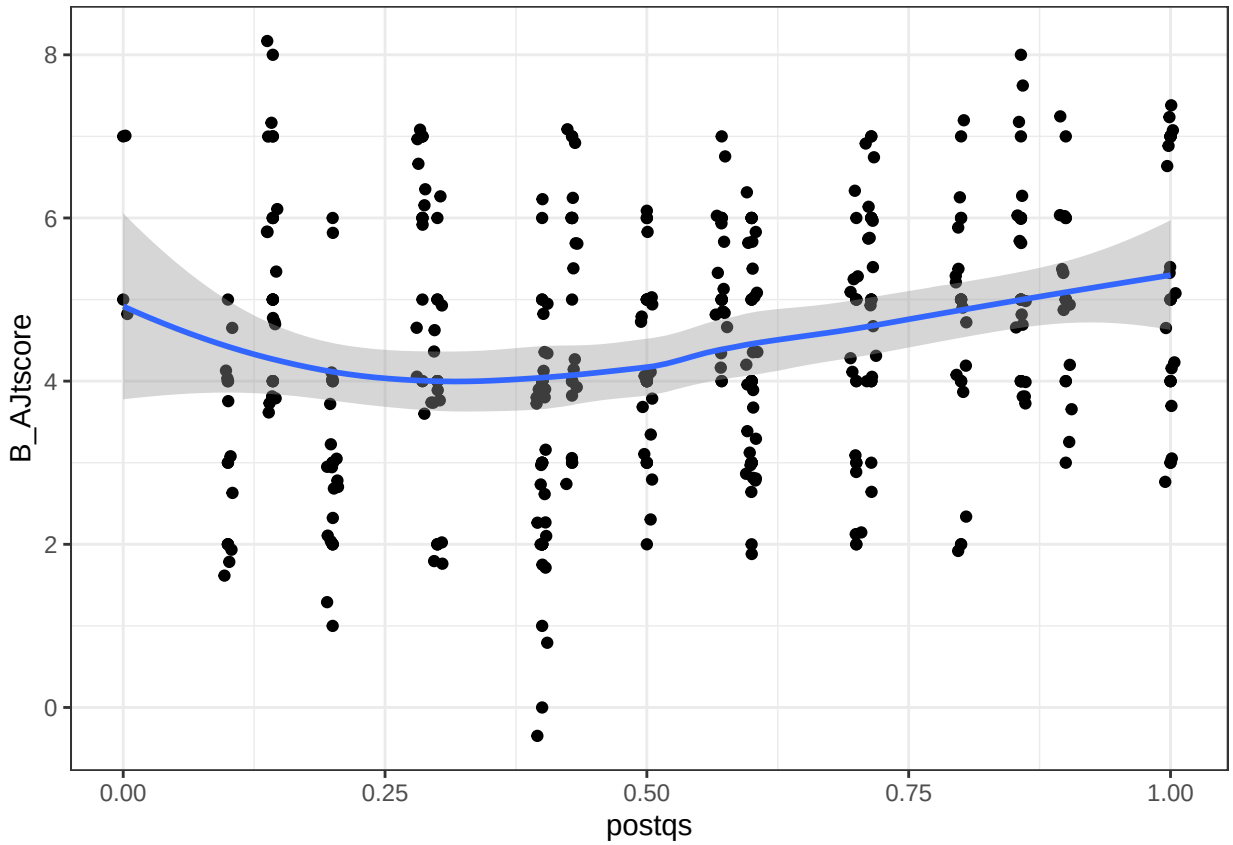
4d. What if we combine BioJtscore and AntJtscore (new variable `B_AJtscore`) along with combining invital and inpsych (new variable `postqs`)??

These new variables have been computed back in section 1. in the code chunk where we build the data frame `AC_df`. And this is what the scatter plot of just these two combined variables look like. We've included this scatterplot to explore the question of whether there is a relationship between how biocentric/anthropocentric the participants post questionnaire answer were compared to how strongly biocentric/anthropocentric their harm justifications were. Visually it looks like there is a relationship for the adults (positive and weak). Not so much for the children And combined weakly positive after the low `postqs` score are surpased. This analysis is merely EDA for us to get a sense of the relationship between this new dependent variable and the new independent variable.

```
ggplot(AC_df, aes(postqs, B_AJtscore, color=Age)) + geom_point() + geom_jitter() + theme_bw() + geom_sm
```




```
ggplot(AC_df, aes(postqs, B_AJtscore)) + geom_point() + geom_jitter() + theme_bw() + geom_smooth()
```



4e. Linear models with the new variables

```
lm.fit2 <- lm(B_AJtscore ~
                Condition + SRFactsTotal + postqs + FirstLang + Age,
              data = AC_df)
#drop Age
lm.fit2b <- lm(B_AJtscore ~
                Condition + SRFactsTotal + postqs + FirstLang,
              data = AC_df)
#drop Condition
lm.fit2c <- lm(B_AJtscore ~
                SRFactsTotal + postqs + FirstLang + Age,
              data = AC_df)
#drop postqs
lm.fit2d <- lm(B_AJtscore ~
                Condition + SRFactsTotal + FirstLang + Age,
              data = AC_df)

summary(lm.fit2)
```

```
##
## Call:
## lm(formula = B_AJtscore ~ Condition + SRFactsTotal + postqs +
##     FirstLang + Age, data = AC_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.5554 -0.9320  0.0113  0.9025  3.0124
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.35148    0.35382   9.472  < 2e-16 ***
## ConditionNat -0.20945    0.21074  -0.994  0.32140
## ConditionPer -0.23700    0.21300  -1.113  0.26709
## SRFactsTotal  0.07837    0.03061   2.560  0.01115 *
## postqs        0.98137    0.33200   2.956  0.00346 **
## FirstLang1   -0.42186    0.26229  -1.608  0.10922
## AgeChild      1.76288    0.20588   8.563 2.12e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.276 on 215 degrees of freedom
## Multiple R-squared:  0.3017, Adjusted R-squared:  0.2822
## F-statistic: 15.48 on 6 and 215 DF,  p-value: 9.699e-15
```

```
summary(lm.fit2b)
```

```
##
## Call:
## lm(formula = B_AJtscore ~ Condition + SRFactsTotal + postqs +
##     FirstLang, data = AC_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1734 -1.0240 -0.0545  1.0185  4.1204
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.82018    0.40386   9.459  < 2e-16 ***
## ConditionNat -0.15565    0.24336  -0.640  0.52312
## ConditionPer -0.25679    0.24608  -1.044  0.29787
## SRFactsTotal -0.03153    0.03211  -0.982  0.32725
## postqs        1.26493    0.38166   3.314  0.00108 **
## FirstLang1    0.29316    0.28727   1.021  0.30863
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.474 on 216 degrees of freedom
## Multiple R-squared:  0.06354,    Adjusted R-squared:  0.04187
## F-statistic: 2.931 on 5 and 216 DF,  p-value: 0.01393
```

```
summary(lm.fit2c)
```

```
##
## Call:
## lm(formula = B_AJtscore ~ SRFactsTotal + postqs + FirstLang +
##     Age, data = AC_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.6488 -0.9416  0.0126  0.8441  2.9049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.23982    0.34043   9.517  < 2e-16 ***
## SRFactsTotal  0.07673    0.03051   2.515  0.01262 *
## postqs        0.92595    0.32763   2.826  0.00515 **
## FirstLang1   -0.42181    0.26163  -1.612  0.10837
## AgeChild      1.76116    0.20545   8.572 1.92e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.274 on 217 degrees of freedom
## Multiple R-squared:  0.2969, Adjusted R-squared:  0.2839
## F-statistic:  22.9 on 4 and 217 DF,  p-value: 8.418e-16
```

```
summary(lm.fit2d)
```

```
##
## Call:
## lm(formula = B_AJtscore ~ Condition + SRFactsTotal + FirstLang +
##     Age, data = AC_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6989 -0.9783 -0.0514  0.9189  3.2543
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.85071    0.31643  12.169  < 2e-16 ***
## ConditionNat -0.16558    0.21394  -0.774   0.4398
## ConditionPer -0.14002    0.21420  -0.654   0.5140
## SRFactsTotal  0.07233    0.03109   2.327   0.0209 *
## FirstLang1   -0.44579    0.26682  -1.671   0.0962 .
## AgeChild      1.82358    0.20849   8.747 6.29e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.298 on 216 degrees of freedom
## Multiple R-squared:  0.2733, Adjusted R-squared:  0.2565
## F-statistic: 16.25 on 5 and 216 DF,  p-value: 1.351e-13
```

4f. ANOVA for combined justification dependent and combined questionnaire independent

Age continues to be significant. Condition still is not. BUT the new variable `postqs` which is a combination of `invitalscore` and `inpsychscore` is now significant.

```
Analysis of Variance Table

Model 1: B_AJtscore ~ Condition + SRFactsTotal + postqs + FirstLang +
    Age
```

```
Model 2: B_AJtscore ~ Condition + SRFactsTotal + postqs + FirstLang
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    215 349.82
2    216 469.12 -1    -119.3 73.322 2.124e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analysis of Variance Table

Model 1: B_AJtscore ~ Condition + SRFactsTotal + postqs + FirstLang +
    Age
Model 2: B_AJtscore ~ SRFactsTotal + postqs + FirstLang + Age
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    215 349.82
2    217 352.24 -2   -2.4182 0.7431 0.4769

Analysis of Variance Table

Model 1: B_AJtscore ~ Condition + SRFactsTotal + postqs + FirstLang +
    Age
Model 2: B_AJtscore ~ Condition + SRFactsTotal + FirstLang + Age
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    215 349.82
2    216 364.04 -1   -14.217 8.7377 0.003465 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Modeling

5a. Mean Severity Ordinal model with combined levels

```
original levels

   0 0.25  0.5 0.75    1 1.25  1.5 1.75    2 2.25  2.5 2.75    3
   1    2    6   18   17   21   22   24   29   23   24   16   19

combined levels

 1  2  3
65 75 82

polr(formula = comMeanSever ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

              Value Std. Error t value p value
ConditionNat  0.294      0.321   0.917   0.359
ConditionPer  0.259      0.319   0.814   0.416
SRFactsTotal  0.057      0.047   1.225   0.221
invitalscore -0.169      0.529  -0.319   0.750
inpsychscore  0.029      0.452   0.063   0.950
```

```
FirstLang1   -0.789      0.395  -1.999   0.046
AgeChild      2.040      0.408   4.999   0.000
1|2          -0.520      0.528  -0.986   0.324
2|3           1.175      0.533   2.204   0.028

              1  2  3
1            42 15  8
2            33 11 31
3            17 10 55
prop.correct 65 15 67

Misclassification error is: 0.5135135
```

5b. Biocentric Justification score, Ordinal model with combined levels

```
original levels

 0  1  2  3  4
19 57 61 62 23

combined levels

  1   2   3
137  62  23

polr(formula = comBioJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

              Value Std. Error t value p value
ConditionNat -0.058      0.356  -0.163   0.871
ConditionPer  0.187      0.358   0.522   0.602
SRFactsTotal  0.165      0.055   3.002   0.003
invitalscore -0.400      0.555  -0.720   0.472
inpsychscore  0.120      0.487   0.247   0.805
FirstLang1   -0.469      0.473  -0.992   0.321
AgeChild      2.042      0.458   4.457   0.000
1|2           1.679      0.628   2.672   0.008
2|3           3.605      0.673   5.355   0.000

               1  2 3
1            127 10 0
2             43 19 0
3              7 16 0
prop.correct  93 31 0

Misclassification error is: 0.3423423
```

5c. Anthropocentric Justification score, Ordinal model with combined levels

```
original levels

 0  1  2  3  4
 9 36 76 61 40

combined levels

  1   2   3
121  61  40

polr(formula = comAntJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

              Value Std. Error t value p value
ConditionNat -0.263      0.327  -0.805   0.421
ConditionPer -0.503      0.334  -1.506   0.132
SRFactsTotal  0.054      0.050   1.086   0.277
invitalscore  0.566      0.537   1.054   0.292
inpsychscore  0.447      0.462   0.969   0.333
FirstLang1    0.348      0.460   0.757   0.449
AgeChild      1.283      0.393   3.264   0.001
1|2           1.577      0.602   2.621   0.009
2|3           3.077      0.629   4.891   0.000

               1  2  3
1            106 11  4
2             45  8  8
3             24 10  6
prop.correct  88 13 15

Misclassification error is: 0.4594595
```

5d. Biocentric choice score, Ordinal model with combined levels

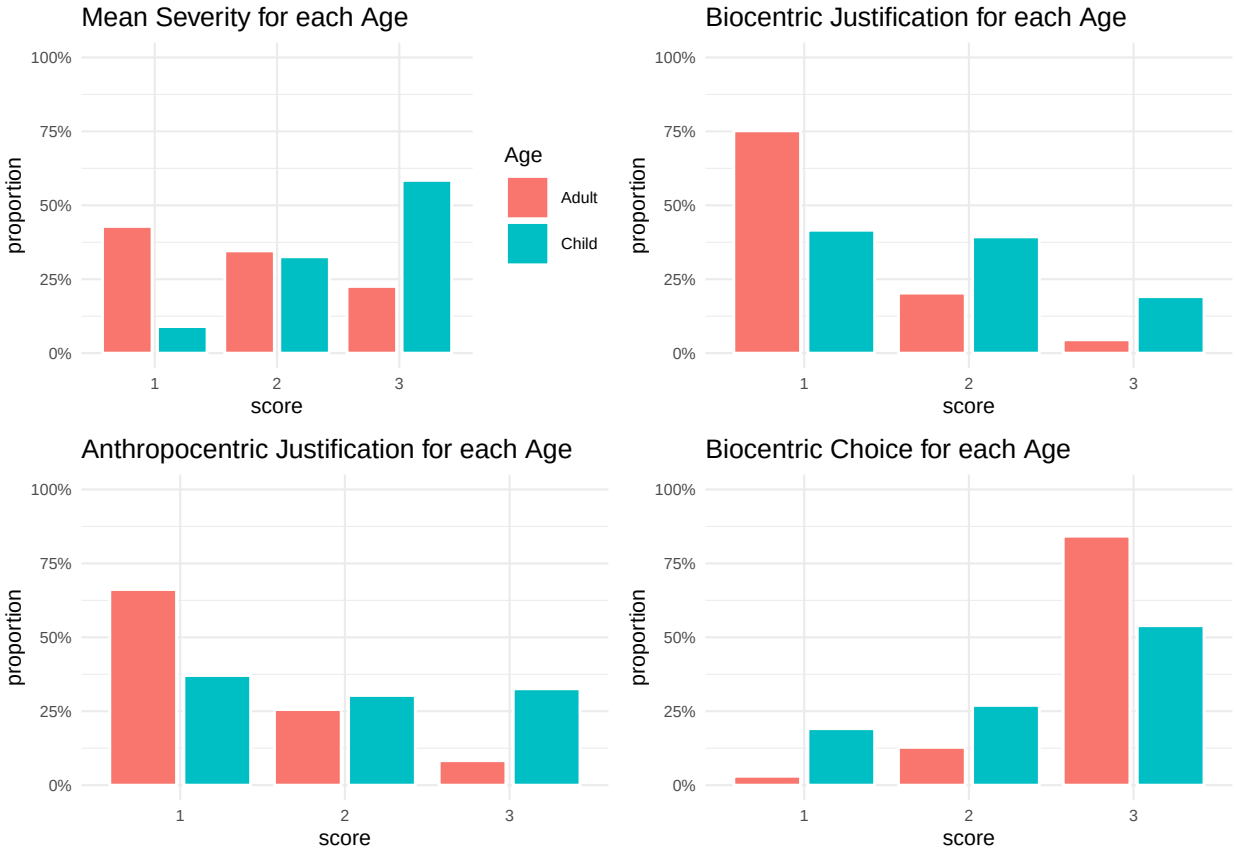
```
original levels

  0   1   2   3   4
  5  16  41  60 100

combined levels

  1   2   3
 21  41 160

polr(formula = comBioJFtotal ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

              Value Std. Error t value p value
ConditionNat  0.034      0.399   0.086   0.931
ConditionPer -0.290      0.402  -0.721   0.471
SRFactsTotal  0.192      0.064   2.982   0.003
invitalscore  1.075      0.646   1.665   0.096
inpsychscore -1.103      0.577  -1.912   0.056
FirstLang1    0.971      0.517   1.879   0.060
AgeChild     -0.879      0.495  -1.775   0.076
1|2          -1.505      0.650  -2.315   0.021
2|3           0.030      0.628   0.048   0.962

              1 2   3
1             3 5  13
2             0 3  38
3             1 2 157
prop.correct 14 7  98

Misclassification error is: 0.2657658
```

6. For `Age`, is there a significant difference in the dependent variables `MeanSeverity`, `BioJscore`, `AntJtscore`, and `BioJFtotal`?

Yes, there is a significant difference.

6a. Visualization

You can see the difference in the plots of the dependent variables colored by `Age`. Note: none of the other predictor variables are included in this visualization.



6b. ANOVA test

We did an ANOVA test using our ordinal models and found that when Age was included in the model there was significant difference from the model that excluded Age. Which means that knowing the age of the participant (adult vs. child) makes a difference to the modeling, and thus there is a significant difference between the dependent variable scores for adults and children. The p-values for these differences are `MeanSeverity` ($p = 4.8 * 10^{-09}$), `BioJscore` ($p = 5.3 * 10^{-08}$), `AntJtscore` ($p = 1.8 * 10^{-04}$), and `BioJFtotal` ($p = 0.015$).

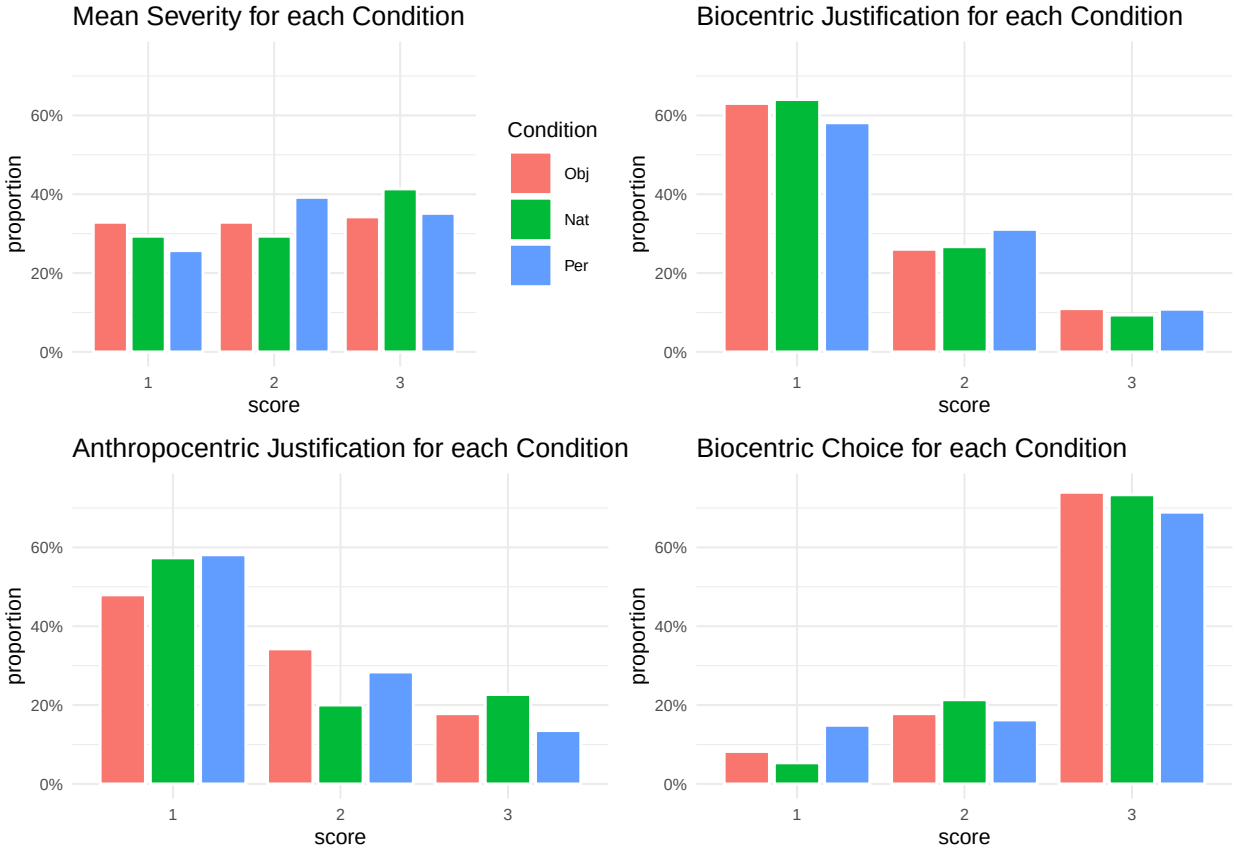
```
Likelihood ratio tests of ordinal regression models

Response: comMeanSever
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.     Pr(Chi)
1       214    464.381
2       213    437.946 1 vs 2     1 26.43508 2.72545e-07

Likelihood ratio tests of ordinal regression models

Response: comBioJtscore
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.     Pr(Chi)
1       214   377.4975
2       213   356.0743 1 vs 2     1 21.42323 3.68282e-06

Likelihood ratio tests of ordinal regression models

Response: comAntJtscore
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1       214   420.6297
2       213   409.7195 1 vs 2     1 10.91025 0.0009563364

Likelihood ratio tests of ordinal regression models

Response: comBioJFtotal
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
1       214   298.9507
2       213   295.7393 1 vs 2     1 3.211362 0.0731286
```

7. For `Condition`, is there a significant difference in the dependent variables `MeanSeverity`, `BioJscore`, `AntJtscore`, and `BioJFtotal`?

We did not find any evidence to conclude that `Condition` had an effect on any of the four dependent variables.

7a. Visualization

Looking at the distribution of Condition for each of the scores (for `MeanSeverity`, `BioJscore`, `AntJtscore`, and `BioJFtotal`) there is no clear visual pattern. Though, these plots do not include any of the other predictor variables.



7b. ANOVA test

The ANOVA test for our models did not show any significant difference when we included/excluded the `Condition` variable. The p-values for these differences are `MeanSeverity` ($p = 0.62$), `BioJscore` ($p = 0.75$), `AntJtscore` ($p = 0.28$), and `BioJFtotal` ($p = 0.68$).

```
Likelihood ratio tests of ordinal regression models

Response: comMeanSever
                                                                        Model
1             SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
1       215   438.9602
2       213   437.9460 1 vs 2     2 1.014252 0.6022239

Likelihood ratio tests of ordinal regression models

Response: comBioJtscore
                                                                        Model
1             SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df  LR stat.   Pr(Chi)
1       215   356.5968
2       213   356.0743 1 vs 2     2 0.5225667 0.7700627

Likelihood ratio tests of ordinal regression models

Response: comAntJtscore
                                                                        Model
1             SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
1       215   412.0154
2       213   409.7195 1 vs 2     2 2.295881 0.3172896

Likelihood ratio tests of ordinal regression models

Response: comBioJFtotal
                                                                        Model
1             SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df  LR stat.   Pr(Chi)
1       215   296.5409
2       213   295.7393 1 vs 2     2 0.8015664 0.6697952
```

7c. Confidence intervals of the model coefficients

We also looked at the 95% confidence intervals for the model coefficients. For each dependent variable, the `Condition` coefficient's confidence interval included zero. A coefficient of zero would mean that the predictor variable had no effect on the model. So a confidence interval that includes zero, means that we can not conclude that the predictor variable has an effect on the dependent variable.

```
polr(formula = comMeanSever ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

                   2.5 %      97.5 %
ConditionNat -0.33355000  0.92555983
ConditionPer -0.36454057  0.88662103
SRFactsTotal -0.03375177  0.14930412
invitalscore -1.20913269  0.87090414
inpsychscore -0.86075416  0.91457750
FirstLang1   -1.57021469 -0.01713478
AgeChild      1.25165280  2.85531281
```

```
polr(formula = comBioJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

                  2.5 %    97.5 %
ConditionNat -0.7576170 0.6416378
ConditionPer -0.5153472 0.8932125
SRFactsTotal  0.0588443 0.2746516
invitalscore -1.4908855 0.6889999
inpsychscore -0.8374278 1.0783297
FirstLang1   -1.3800502 0.4906933
AgeChild      1.1622761 2.9638546
```

```
polr(formula = comAntJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

                   2.5 %    97.5 %
ConditionNat -0.90907579 0.3769877
ConditionPer -1.16355269 0.1478639
SRFactsTotal -0.04319113 0.1524763
invitalscore -0.48490173 1.6254512
inpsychscore -0.45863584 1.3554089
FirstLang1   -0.52100071 1.3006549
AgeChild      0.51966945 2.0652035
```

```
polr(formula = comBioJFtotal ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

                   2.5 %     97.5 %
ConditionNat -0.75228618 0.81960216
ConditionPer -1.08536328 0.49827623
SRFactsTotal  0.06904114 0.32236839
invitalscore -0.18074106 2.35926132
inpsychscore -2.25074943 0.01967021
FirstLang1   -0.06264457 1.98212088
AgeChild     -1.86946992 0.08187747
```