All Groups Mapping

Kristin Walker

3/9/2021

Load Packages

Load the following libraries. If they are not installed, run install.packages("packagename")

```
library(rmarkdown)
library(ggplot2)
library(beeswarm)
library(MASS)
library(generalhoslem)
library(qwraps2)
library(quantreg)
library(olsrr)
library(car)
library(ggpubr)
library(dplyr)
library(tidyr)
library(lme4)
library(VGAM)
library(gridExtra)
library(sjPlot)
library(sjmisc)
library(sjlabelled)
library(olsrr)
```

Dataframes Setup

1. Import the data from a CSV file
    - Should have *224 rows*

```
setwd("~/Desktop/UConn Manuscripts/MappingPaper/AllGroupMapping")
Map_All <- read.csv("Mapping_Coding_KW_210309.csv", na.strings = "N/A")
dim(Map_All)
```

2. Subset data by age group (4;6-9;11)
    - Should have *220 rows* (includes children never tested on this task)

```
Map49 <- subset(Map_All, Map_All$Age_Rounded>=4.5 & Map_All$Age_Rounded<10 | is.na(Map_All$Age_Rounded))
dim(Map49)
```

3. Create a new dataframe from desired subset (e.g., children we are including in the analyses)
    - Should have ***190 rows***

```
Map_Inc <- subset(Map49, Map49$Including.in.Study == "Yes" & Map49$Coded. == "Yes" & Map49$Mapping_Inclu
dim(Map_Inc)
```

View(Map_Inc)

Participant Demographics

- Four Participant Groups
    - EE: English Early, hearing children exposed to English from birth
    - EL: English Later, D/HH children exposed to English "later"
    - AE: ASL Early, D/HH children exposed to ASL from birth (have at least 1 D/HH parent)
    - AL: ASL Later, D/HH children exposed to ASL "later" (have hearing parents)
- How many children were tested, but were not included in analyses?
    - 6 total: 3 had additional disabilities/suspected disabilities, 1 refused to participate, 1 due to technical difficulties (not filmed entirely), 1 Early ASL child tested in English after trying ASL (difficult to communicate with)

```
Map_Not_Inc <- subset(Map49, Map49$Tested == "Yes" & Map49$Mapping_Include. == "No")
nrow(Map_Not_Inc)
```

```
## [1] 6
```

- Does the status of including in study depend on which group children were in?
    - No, the status of inclusion in study is not dependent on group ($p > 0.05$).

```
EE_Inc <- subset(Map_Inc, Map_Inc$Group_4cat == "English Early")
EL_Inc <- subset(Map_Inc, Map_Inc$Group_4cat == "English Later")
AE_Inc <- subset(Map_Inc, Map_Inc$Group_4cat == "ASL Early")
AL_Inc <- subset(Map_Inc, Map_Inc$Group_4cat == "ASL Later")

EE_Not_Inc <- subset(Map_Not_Inc, Map_Not_Inc$Group_4cat == "English Early")
EL_Not_Inc <- subset(Map_Not_Inc, Map_Not_Inc$Group_4cat == "English Later")
AE_Not_Inc <- subset(Map_Not_Inc, Map_Not_Inc$Group_4cat == "ASL Early")
AL_Not_Inc <- subset(Map_Not_Inc, Map_Not_Inc$Group_4cat == "ASL Later")
rnames <- c("Including in Study","Not Including in Study")
cnames <- c("English Early","English Later","ASL Early","ASL Later")
Table_IncStudy <- matrix(c(nrow(EE_Inc),nrow(EL_Inc),nrow(AE_Inc),nrow(AL_Inc),nrow(EE_Not_Inc),nrow(EL_
Table_IncStudy
```

```
##                         English Early English Later ASL Early ASL Later
## Including in Study                 48            46        46        50
## Not Including in Study              0             0         4         2
```

```
chisq.test(Table_IncStudy)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Table_IncStudy
## X-squared = 7.1861, df = 3, p-value = 0.0662
```

- Demographic table information: descriptive statistics for total children and for each group (e.g., EE, EL, AE, AL)

```
#Recode Ethnicity to combine Unsure/Missing Categories
Map_Inc$Ethnicity <- dplyr::recode(as.character(Map_Inc$Ethnicity), 'Hispanic or Latino' = "Hispanic or

#Recode Race to combine Unsure/Missing Categories
Map_Inc$Race <- dplyr::recode(as.character(Map_Inc$Race), 'White' = "Caucasian", 'Black or African Amer

#Reorder Grade
Map_Inc$Grade. <- factor(Map_Inc$Grade., levels=c("Pre-Kindergarten", "Kindergarten", "1st", "2nd", "3r

#Reorder Ethnicity
Map_Inc$Ethnicity <- factor(Map_Inc$Ethnicity, levels=c("Not Hispanic or Latino", "Hispanic or Latino",

#Reorder Race
Map_Inc$Race <- factor(Map_Inc$Race, levels=c("Caucasian", "Asian", "African American", "American India

table1::label(Map_Inc$Age_Rounded) <- "Age (years)"
table1::label(Map_Inc$SES) <- "SES"
table1::label(Map_Inc$M.F) <- "Sex"
table1::label(Map_Inc$Race) <- "Race"
table1::label(Map_Inc$Ethnicity) <- "Ethnicity"
table1::label(Map_Inc$Grade.) <- "Grade"

table1::table1(~Age_Rounded + SES + M.F + Race + Ethnicity + Grade.| Group_4cat, data = Map_Inc)

#When were children tested?
Map_Inc$SchoolYear_Timing <- factor(Map_Inc$SchoolYear_Timing, levels=c("Beginning", "Middle", "End"))
table1::label(Map_Inc$SchoolYear_Timing) <- "When Children Were Tested"
table1::table1(~SchoolYear_Timing | Group_4cat, data = Map_Inc)
```

Analyses

General Performance

- Performing significantly greater than chance?
  - Yes for overall and four pariticpant groups, all $p < 0.001$

```
wilcox.test(Map_Inc$AvgCorrect_Total, mu = .25, alternative = "greater")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  Map_Inc$AvgCorrect_Total
## V = 18144, p-value < 2.2e-16
## alternative hypothesis: true location is greater than 0.25
```

```
wilcox.test(EE_Inc$AvgCorrect_Total, mu = .25, alternative = "greater")
```

```
## Warning in wilcox.test.default(EE_Inc$AvgCorrect_Total, mu = 0.25, alternative =
## "greater"): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  EE_Inc$AvgCorrect_Total
## V = 1176, p-value = 7.685e-10
## alternative hypothesis: true location is greater than 0.25
```

```
wilcox.test(AE_Inc$AvgCorrect_Total, mu = .25, alternative = "greater")
```

```
## Warning in wilcox.test.default(AE_Inc$AvgCorrect_Total, mu = 0.25, alternative =
## "greater"): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  AE_Inc$AvgCorrect_Total
## V = 1081, p-value = 1.585e-09
## alternative hypothesis: true location is greater than 0.25
```

```
wilcox.test(EL_Inc$AvgCorrect_Total, mu = .25, alternative = "greater")
```

```
## Warning in wilcox.test.default(EL_Inc$AvgCorrect_Total, mu = 0.25, alternative =
## "greater"): cannot compute exact p-value with ties
```

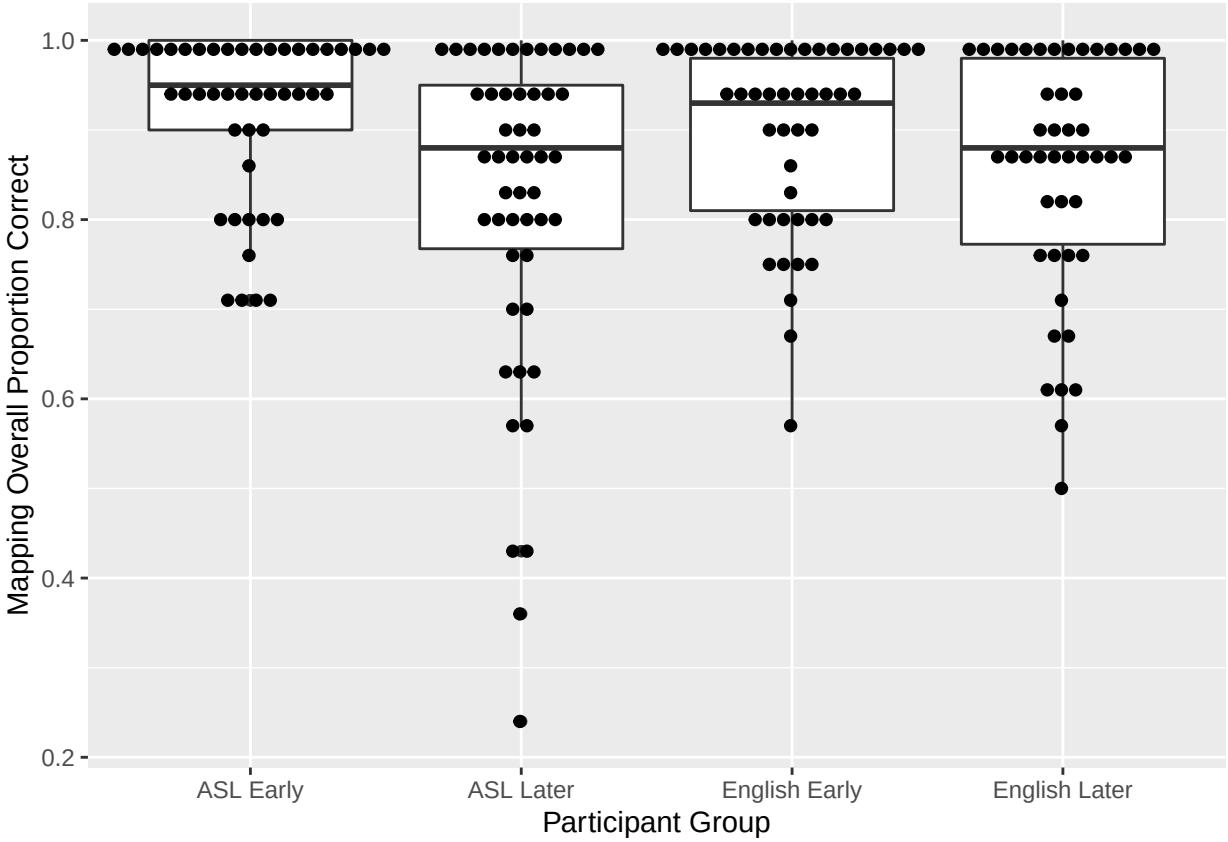
```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  EL_Inc$AvgCorrect_Total
## V = 1081, p-value = 1.734e-09
## alternative hypothesis: true location is greater than 0.25
```

```
wilcox.test(AL_Inc$AvgCorrect_Total, mu = .25, alternative = "greater")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  AL_Inc$AvgCorrect_Total
## V = 1274, p-value = 4.005e-10
## alternative hypothesis: true location is greater than 0.25
```

```
#https://github.com/emlini/SLaM-Give-N/blob/main/Carrigan_Coppola_Shusterman_GiveN_code.Rmd#L126
ggplot(data=Map_Inc, mapping = aes(x=Group_4cat, y=AvgCorrect_Total)) + geom_boxplot() + geom_dotplot(m
```

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



- Performance at or near ceiling (greater than or equal to 90%)
  - 58% of children performed at or near ceiling.

```
Ceil <- Map_Inc$AvgCorrect_Total
All <- .9 #ceiling performance
Var <- length(which(Ceil>= All))
N_O <- nrow(Map_Inc)
(Var/N_O)*100
```

```
## [1] 58.42105
```

QUESTION 1: What is the relationship between language experience and mapping skills?

- Helpful Tobit Model Link
- Create dataframe for tobit models (includes adding set size and mapping pair columns)
  - Should have ***1710 entries / 9 = 190 children***

– 9 rows for each child because we have 3 set sizes for each of the 3 different mapping pairs (Numeral-Word, Quantity-Word, Quantity-Numeral)

```
which(colnames(Map_Inc)=="AvgCorrect_Med_QW") #368
which(colnames(Map_Inc)=="AvgCorrect_Lrg_WN") #373
T_long <- pivot_longer(Map_Inc, cols = 368:373, values_to = "AvgCorr")
which(colnames(T_long)=="Item1_QN_Answer") #35
which(colnames(T_long)=="Item51_WN_Correct.") #343
T_short <- T_long[,-c(35:343)]
T_short <-  mutate(T_short, SetSize = case_when(grepl("Med", T_short$name) ~"Medium", grepl("Lrg", T_sh
T_short$SetSize <- as.factor(factor(as.character(T_short$SetSize), levels=c("Medium", "Large"), exclude=
T_short <- mutate(T_short, MapPair = case_when(grepl("QW", T_short$name) ~ "Quantity-Word", grepl("QN",
T_short$MapPair <- as.factor(factor(as.character(T_short$MapPair), levels=c("Numeral-Word","Quantity-Nu
T_short$MapPair_refQW <- as.factor(factor(as.character(T_short$MapPair), levels=c("Quantity-Word","Numer
```

View(T_short)

Tobit model #1 (ages 4;6-9;11)

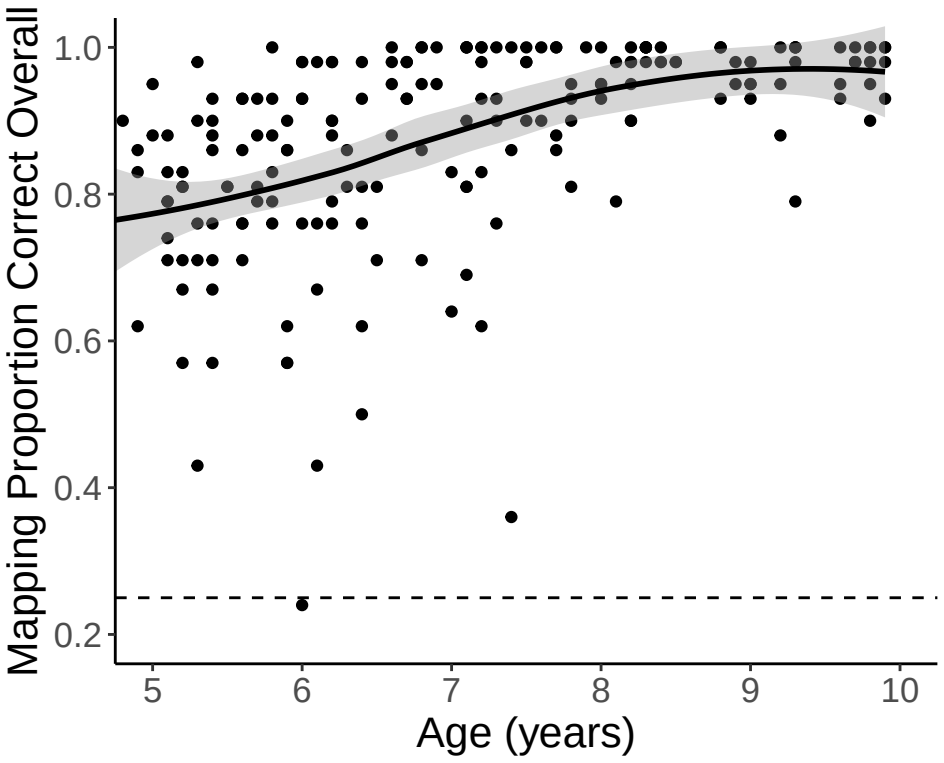
- Outcome variable: Overall Mapping Performance
- Predictors: Age + SES + Modality + Timing + Set Size + Map Pair + Timing x Modality
- Result: Hauck-Donner effect for Age due to ceiling effect

```
summary(Map_t<- vglm(AvgCorr ~  Age_Rounded + SES + Modality + Timing + SetSize + MapPair + Timing:Modal
```

```
##
## Call:
## vglm(formula = AvgCorr ~ Age_Rounded + SES + Modality + Timing +
##     SetSize + MapPair + Timing:Modality, family = tobit(Upper = 1),
##     data = T_short)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                0.4649676  0.0754402   6.163 7.12e-10 ***
## (Intercept):2               -1.1904061  0.0309445 -38.469  < 2e-16 ***
## Age_Rounded                  0.1236287  0.0087524  14.125  < 2e-16 ***
## SES                          0.0013840  0.0007527   1.839 0.065962 .
## ModalityEnglish              0.0020441  0.0339450   0.060 0.951982
## TimingLater                 -0.1761779  0.0312964  -5.629 1.81e-08 ***
## SetSizeLarge                -0.1465941  0.0219038  -6.693 2.19e-11 ***
## MapPairQuantity-Numeral     -0.2418398  0.0286535  -8.440  < 2e-16 ***
## MapPairQuantity-Word        -0.3232978  0.0283137 -11.418  < 2e-16 ***
## ModalityEnglish:TimingLater  0.1502937  0.0441575   3.404 0.000665 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: mu, loglink(sd)
##
## Log-likelihood: -427.3633 on 2270 degrees of freedom
##
## Number of Fisher scoring iterations: 19
##
```

```
## Warning: Hauck-Donner effect detected in the following estimate(s):
## 'Age_Rounded'
```

```
ggplot(data = Map_Inc, mapping = aes(x=Age_Rounded, y=AvgCorrect_Total)) + geom_point() + geom_smooth(m
```



Tobit model #2 (ages 4;6-7;11). Not including 8 to 9-year-olds to avoid Hauck-Donner effect.

- Should have ***810 entries / 6 = 135 children***

```
Map_58 <- subset(T_short, T_short$Age_Rounded <8)
nrow(Map_58)
```

```
## [1] 810
```

- Outcome variable: Overall Mapping Performance
- Predictors: Age + SES + Set Size + Map Pair + Modality + Timing + Timing x Modality
- Result: All predictors besides Modality are significant

Model 2a: Reference Group is Numeral-Word

```
summary(Map_58_I<- vglm(AvgCorr ~  Age_Rounded + SES + Modality + Timing + SetSize + MapPair + Modality
```

```
##
## Call:
## vglm(formula = AvgCorr ~ Age_Rounded + SES + Modality + Timing +
```

```
##     SetSize + MapPair + Modality:Timing, family = tobit(Upper = 1),
##     data = Map_58)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                0.3811567  0.1133900   3.361 0.000775 ***
## (Intercept):2               -1.1784486  0.0357074 -33.003  < 2e-16 ***
## Age_Rounded                  0.1326648  0.0153337   8.652  < 2e-16 ***
## SES                          0.0019221  0.0008322   2.310 0.020907 *
## ModalityEnglish              0.0119045  0.0380327   0.313 0.754275
## TimingLater                 -0.1697280  0.0367858  -4.614 3.95e-06 ***
## SetSizeLarge                -0.1598640  0.0246195  -6.493 8.39e-11 ***
## MapPairQuantity-Numeral     -0.2351472  0.0315320  -7.457 8.82e-14 ***
## MapPairQuantity-Word        -0.3115494  0.0312056  -9.984  < 2e-16 ***
## ModalityEnglish:TimingLater  0.1325587  0.0500617   2.648 0.008099 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: mu, loglink(sd)
##
## Log-likelihood: -336.7299 on 1610 degrees of freedom
##
## Number of Fisher scoring iterations: 12
##
## No Hauck-Donner effect found in any of the estimates
```

Model 2b: Reference Groups is Quantity-Word

```
summary(Map_58_I_ref<- vglm(AvgCorr ~  Age_Rounded + SES + Modality + Timing + SetSize + MapPair_refQW +
```

```
##
## Call:
## vglm(formula = AvgCorr ~ Age_Rounded + SES + Modality + Timing +
##     SetSize + MapPair_refQW + Modality:Timing, family = tobit(Upper = 1),
##     data = Map_58)
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                  0.0696073  0.1141071   0.610  0.54185
## (Intercept):2                 -1.1784486  0.0357074 -33.003  < 2e-16 ***
## Age_Rounded                    0.1326648  0.0153337   8.652  < 2e-16 ***
## SES                            0.0019221  0.0008322   2.310  0.02091 *
## ModalityEnglish                0.0119045  0.0380327   0.313  0.75427
## TimingLater                   -0.1697280  0.0367858  -4.614 3.95e-06 ***
## SetSizeLarge                  -0.1598640  0.0246195  -6.493 8.39e-11 ***
## MapPair_refQWNumeral-Word      0.3115494  0.0312056   9.984  < 2e-16 ***
## MapPair_refQWQuantity-Numeral  0.0764022  0.0285871   2.673  0.00753 **
## ModalityEnglish:TimingLater    0.1325587  0.0500617   2.648  0.00810 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: mu, loglink(sd)
##
```

```
## Log-likelihood: -336.7299 on 1610 degrees of freedom
##
## Number of Fisher scoring iterations: 12
##
## No Hauck-Donner effect found in any of the estimates
```

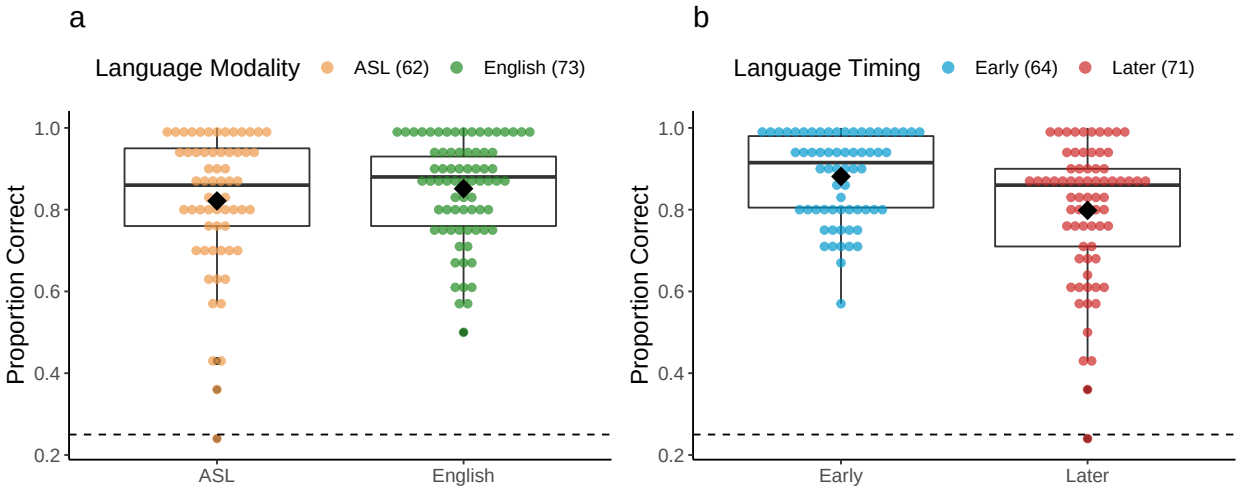
- Create dataframe for 4;6-7;11 year olds
  - Ensure has 135 children

```
Map_Inc_58 <- subset(Map_Inc, Map_Inc$Age_Rounded <8)
nrow(Map_Inc_58)
```

```
## [1] 135
```

```
M <- ggplot(data=Map_Inc_58, mapping = aes(x= Modality, y=AvgCorrect_Total)) + geom_boxplot() + geom_do
T <- ggplot(data=Map_Inc_58, mapping = aes(x= Timing, y=AvgCorrect_Total)) + geom_boxplot() + geom_dotpl
```

```
grid.arrange(M,T,nrow=1)
```



- Create HTML tobit model tables
  - Will need to edit after to add R-squared values

```
tab_model(Map_58_I, auto.label = FALSE, show.ci = FALSE, show.se = TRUE, show.stat = TRUE, show.est = TR
```

Reference group: Numeral-Word

Predictors

Estimate

Standard Error

t-statistic

p-value

Intercept 1

0.38

0.11

3.36

0.001

Intercept 2

-1.18

0.04

-33.00

<0.001

Age

0.13

0.02

8.65

<0.001

SES

0.00

0.00

2.31

0.021

Modality (English)

0.01

0.04

0.31

0.754

Timing (Later)

-0.17

0.04

-4.61

<0.001

Large

-0.16

0.02

-6.49

<0.001

Quantity-Numeral

-0.24

0.03

-7.46

<0.001

Quantity-Word

-0.31

0.03

-9.98

<0.001

Modality (English) x Timing (Later)

0.13

0.05

2.65

0.008

Observations

810

```
tab_model(Map_58_I_ref, auto.label = FALSE, show.ci = FALSE, show.se = TRUE, show.stat = TRUE, show.est
```

Reference group: Quantity-Word

Predictors

Estimate

Standard Error

t-statistic

p-value

Intercept 1

0.07

0.11

0.61

0.542

Intercept 2

-1.18

0.04

-33.00

<0.001

Age

0.13

0.02

8.65

<0.001

SES

0.00

0.00

2.31

0.021

Modality (English)

0.01

0.04

0.31

0.754

Timing (Later)

-0.17

0.04

-4.61

<0.001

Large

-0.16

0.02

-6.49

<0.001

Numeral-Word

0.31

0.03

9.98

<0.001

Quantity-Numeral

0.08

0.03

2.67

0.008

Modality (English) x Timing (Later)

0.13

0.05

2.65

0.008

Observations

810

- Calculate model fit
  - Log likelihood of null model: -455.3508
  - Log likelihood of our model: -336.7299
  - R-squared: 0.26

```
summary(Map_58_null<- vglm(AvgCorr ~  1, tobit(Upper = 1.0), data = Map_58))
```

```
##
## Call:
## vglm(formula = AvgCorr ~ 1, family = tobit(Upper = 1), data = Map_58)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  0.99276    0.01591   62.41   <2e-16 ***
## (Intercept):2 -1.00050    0.03932  -25.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: mu, loglink(sd)
##
## Log-likelihood: -455.3508 on 1618 degrees of freedom
##
## Number of Fisher scoring iterations: 6
##
## No Hauck-Donner effect found in any of the estimates
```

```
print(R2_Model2 <- 1 - (-336.7299 / -455.3508))
```

```
## [1] 0.2605044
```

- Obtain mapping performance means for mapping pairs and set sizes

```
QW_58 <- subset(Map_58, Map_58$MapPair == "Quantity-Word")
QN_58 <- subset(Map_58, Map_58$MapPair == "Quantity-Numeral")
NW_58 <- subset(Map_58, Map_58$MapPair == "Numeral-Word")
mean(QW_58$AvgCorr)
```

```
## [1] 0.7858889
```

```
mean(QN_58$AvgCorr)
```

```
## [1] 0.8147778
```

```
mean(NW_58$AvgCorr)
```

```
## [1] 0.9339259
```

```
Map_58_Med <- subset(Map_58, Map_58$SetSize == "Medium")
Map_58_Lrg <- subset(Map_58, Map_58$SetSize == "Large")
mean(Map_58_Med$AvgCorr)
```

```
## [1] 0.8798765
```

```
mean(Map_58_Lrg$AvgCorr)
```

```
## [1] 0.8098519
```

- Create subsetted dataframes for Timing and Modality (obtain ns and mean group performances)
  - Early: 64 children, M = 0.88
  - Later: 71 children, M = 0.80
  - English: 73 children, M = 0.85
  - ASL: 62 children, M = 0.82

```
Early_58 <- (subset(Map_Inc_58, Map_Inc_58$Timing == "Early"))
nrow(Early_58)
```

```
## [1] 64
```

```
mean(Early_58$AvgCorrect_Total)
```

```
## [1] 0.8810938
```

```
Later_58 <- (subset(Map_Inc_58, Map_Inc_58$Timing == "Later"))
nrow(Later_58)
```

```
## [1] 71
```

```
mean(Later_58$AvgCorrect_Total)
```

```
## [1] 0.7983099
```

```
English_58 <- (subset(Map_Inc_58, Map_Inc_58$Modality == "English"))
nrow(English_58)
```

```
## [1] 73
```

```
mean(English_58$AvgCorrect_Total)
```

```
## [1] 0.8513699
```

```
ASL_58 <- (subset(Map_Inc_58, Map_Inc_58$Modality == "ASL"))
nrow(ASL_58)
```

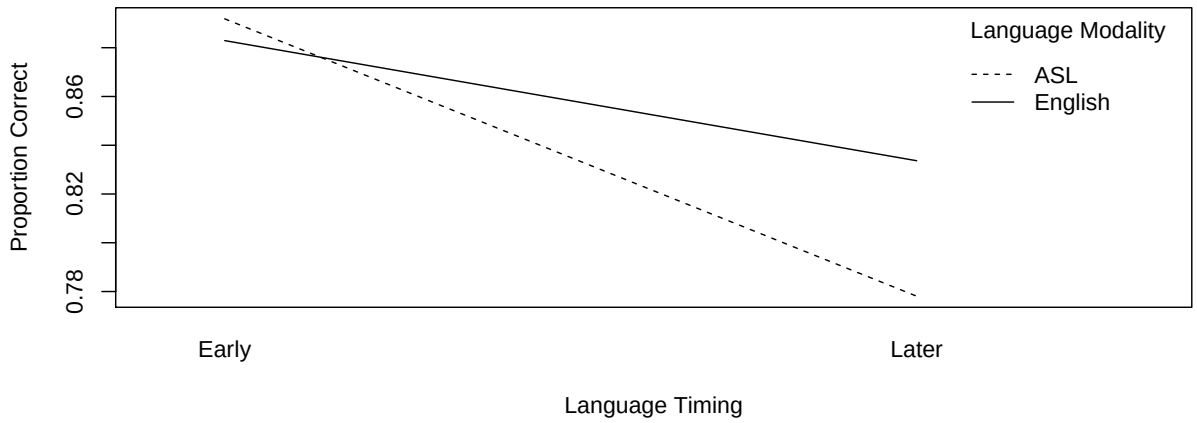
```
## [1] 62
```

```
mean(ASL_58$AvgCorrect_Total)
```

```
## [1] 0.8212903
```

```
interaction.plot(Map_58$Timing, Map_58$Modality, Map_58$AvgCorr,  ylab = "Proportion Correct", xlab = "
```

Explaining the Timing x Modality significant interaction



- Similar performances within Timing groups?
  - Early: Yes (p = 0.84). ASL Early (M=0.89) & English Early (M=0.88)
  - Later: Yes (p=0.28). ASL Later (M=0.77) & English Later (M=0.83)

```
wilcox.test(AvgCorrect_Total ~ Group_4cat, data = Early_58, exact = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  AvgCorrect_Total by Group_4cat
## W = 519, p-value = 0.8434
## alternative hypothesis: true location shift is not equal to 0
```

```r
mean(subset(Early_58, Group_4cat == "ASL Early")$AvgCorrect_Total)
```

```
## [1] 0.8860714
```

```r
mean(subset(Early_58, Group_4cat == "English Early")$AvgCorrect_Total)
```

```
## [1] 0.8772222
```

```r
wilcox.test(AvgCorrect_Total ~ Group_4cat, data = Later_58, exact = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  AvgCorrect_Total by Group_4cat
## W = 534.5, p-value = 0.2781
## alternative hypothesis: true location shift is not equal to 0
```

```r
mean(subset(Later_58, Group_4cat == "ASL Later")$AvgCorrect_Total)
```

```
## [1] 0.7679412
```

```r
mean(subset(Later_58, Group_4cat == "English Later")$AvgCorrect_Total)
```

```
## [1] 0.8262162
```

Perhaps children's age of language exposure within the Later group can explain the performance difference by modality (although not significant) within the Later group compared to the Early group.

- Age of first language exposure demographics for Later groups

```r
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, mean)
```

```
##      Group_4cat Age.of.Exposure..mo..Language
## 1     ASL Later                      43.01515
## 2 English Later                      21.39706
```

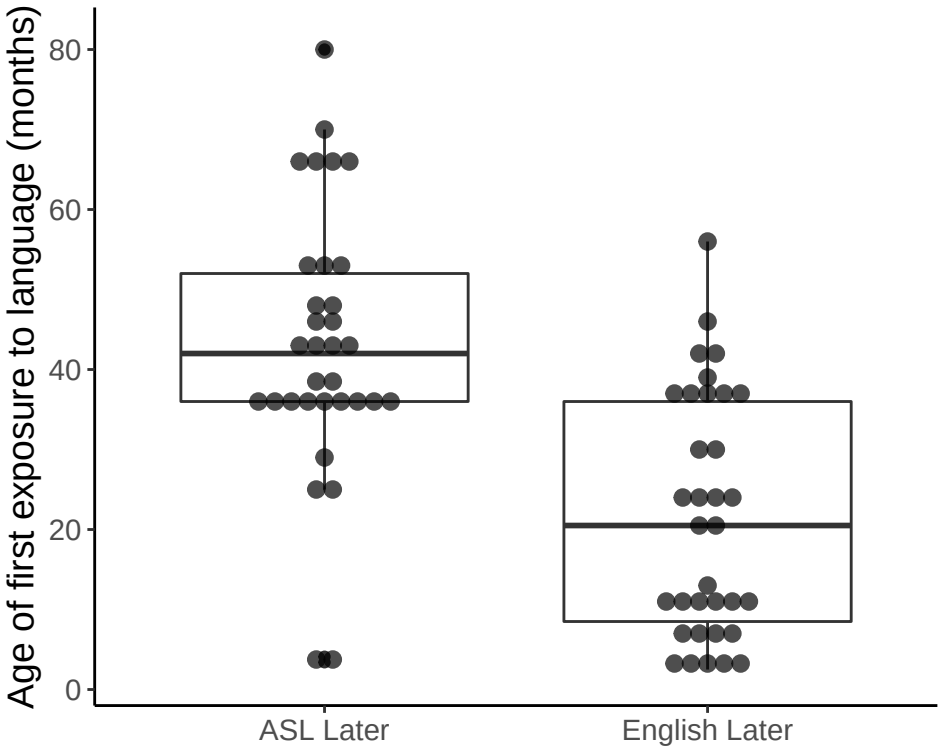
```r
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, sd)
```

```
##      Group_4cat Age.of.Exposure..mo..Language
## 1     ASL Later                      16.75407
## 2 English Later                      15.07244
```

```r
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, range)
```

```
##      Group_4cat Age.of.Exposure..mo..Language.1 Age.of.Exposure..mo..Language.2
## 1     ASL Later                             3.4                            80.0
## 2 English Later                             2.5                            56.0
```

```
ggplot(data=Later_58, mapping = aes(x= Group_4cat, y=Age.of.Exposure..mo..Language)) + geom_boxplot() +
```



- Does age of first language exposure differ between children exposed to language Later (e.g., ASL Later and English Later groups)?
  - Used Wilcoxon Sum Rank Test due to the data not being normally distributed (Shapiro-Wilk: $p < 0.05$)
  - Yes, ASL Later children are exposed to language significantly later than English Later children. $W = 927$, $p < 0.001$

```
shapiro.test(Later_58$Age.of.Exposure..mo..Language)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Later_58$Age.of.Exposure..mo..Language
## W = 0.95461, p-value = 0.01571
```

```
wilcox.test(Later_58$Age.of.Exposure..mo..Language ~ Later_58$Group_4cat, exact = FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  Later_58$Age.of.Exposure..mo..Language by Later_58$Group_4cat
## W = 927, p-value = 4.317e-06
## alternative hypothesis: true location shift is not equal to 0
```

```
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, mean)
```

```
##      Group_4cat Age.of.Exposure..mo..Language
## 1     ASL Later                      43.01515
## 2 English Later                      21.39706
```

```
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, sd)
```

```
##      Group_4cat Age.of.Exposure..mo..Language
## 1     ASL Later                      16.75407
## 2 English Later                      15.07244
```

```
aggregate(Age.of.Exposure..mo..Language ~ Group_4cat, data=Later_58, range)
```

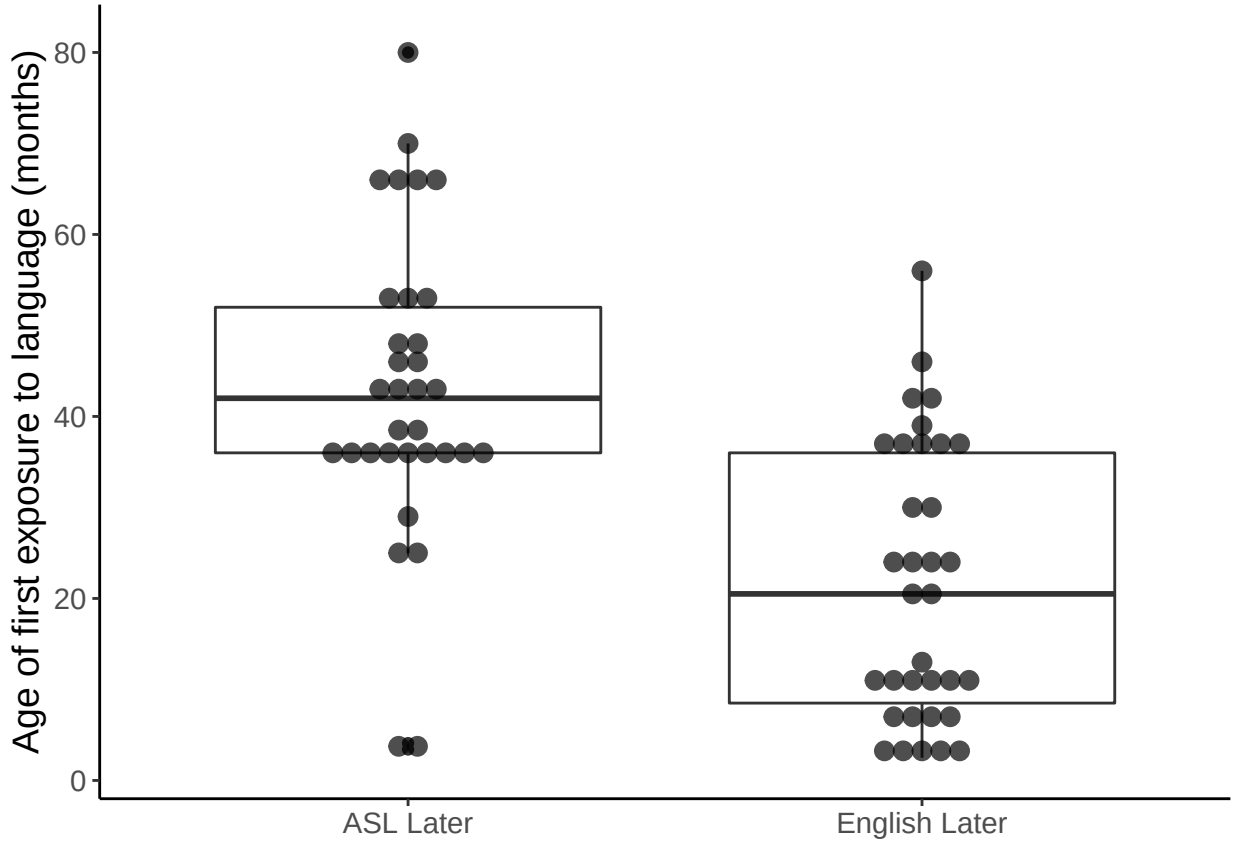
```
##      Group_4cat Age.of.Exposure..mo..Language.1 Age.of.Exposure..mo..Language.2
## 1     ASL Later                             3.4                            80.0
## 2 English Later                             2.5                            56.0
```

```
ggplot(data=Later_58, mapping = aes(x= Group_4cat, y=Age.of.Exposure..mo..Language)) + geom_boxplot() +
```

```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```

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

```
## Warning: Removed 4 rows containing non-finite values (stat_bindot).
```



```
which(colnames(Map_Inc)=="Sum_Med_QW") #352
```

Linear Regression Model

```
## [1] 352
```

```
which(colnames(Map_Inc)=="Sum_Lrg_WN") #357
```

```
## [1] 357
```

```
L_long <- pivot_longer(Map_Inc, cols = 352:357, values_to = "SumCorr")
which(colnames(L_long)=="Item1_QN_Answer") #35
```

```
## [1] 35
```

```
which(colnames(L_long)=="Item51_WN_Correct.") #343
```

```
## [1] 343
```

```
L_short <- L_long[,-c(35:343)]
L_short <-  mutate(L_short, SetSize = case_when(grepl("Med", L_short$name) ~"Medium", grepl("Lrg", L_sh
L_short$SetSize <- as.factor(factor(as.character(L_short$SetSize), levels=c("Medium", "Large"), exclude=
L_short <- mutate(L_short, MapPair = case_when(grepl("QW", L_short$name) ~ "Quantity-Word", grepl("QN",
L_short$MapPair <- as.factor(factor(as.character(L_short$MapPair), levels=c("Numeral-Word","Quantity-Nu
L_short$MapPair_refQW <- as.factor(factor(as.character(L_short$MapPair), levels=c("Quantity-Word","Nume
```

View(L_short)

```
#https://www.statmethods.net/stats/regression.html
Map_lm <- lm(SumCorr ~ Age_Rounded + SES + Modality + Timing + SetSize + MapPair + Timing:Modality, data
summary(Map_lm)
```

```
##
## Call:
## lm(formula = SumCorr ~ Age_Rounded + SES + Modality + Timing +
##     SetSize + MapPair + Timing:Modality, data = L_short)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.638 -0.567  0.195  0.799 31.687
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.8345753  0.3025243  12.675  < 2e-16 ***
## Age_Rounded                 0.3314836  0.0322029  10.294  < 2e-16 ***
## SES                         0.0009989  0.0032050   0.312   0.7554
## ModalityEnglish             0.0625429  0.1355981   0.461   0.6447
## TimingLater                -0.5310692  0.1266346  -4.194 2.96e-05 ***
```

```
## SetSizeLarge                 1.2929825  0.0897732  14.403  < 2e-16 ***
## MapPairQuantity-Numeral     -0.7684211  0.1099492  -6.989 4.72e-12 ***
## MapPairQuantity-Word        -0.9578947  0.1099492  -8.712  < 2e-16 ***
## ModalityEnglish:TimingLater  0.3910448  0.1811299   2.159   0.0311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.516 on 1131 degrees of freedom
## Multiple R-squared:  0.2739, Adjusted R-squared:  0.2688
## F-statistic: 53.33 on 8 and 1131 DF,  p-value: < 2.2e-16
```

```
#Multiple R-squared:  0.2739,   Adjusted R-squared:  0.2688
#F-statistic: 53.33 on 8 and 1131 DF,  p-value: < 2.2e-16

Map_lm_ref <- lm(SumCorr ~ Age_Rounded + SES + Modality + Timing + SetSize + MapPair_refQW + Timing:Moda
summary(Map_lm_ref)
```

```
##
## Call:
## lm(formula = SumCorr ~ Age_Rounded + SES + Modality + Timing +
##     SetSize + MapPair_refQW + Timing:Modality, data = L_short)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.638 -0.567  0.195  0.799 31.687
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2.8766806  0.3025243   9.509  < 2e-16 ***
## Age_Rounded                   0.3314836  0.0322029  10.294  < 2e-16 ***
## SES                           0.0009989  0.0032050   0.312   0.7554
## ModalityEnglish               0.0625429  0.1355981   0.461   0.6447
## TimingLater                  -0.5310692  0.1266346  -4.194 2.96e-05 ***
## SetSizeLarge                  1.2929825  0.0897732  14.403  < 2e-16 ***
## MapPair_refQWNumeral-Word     0.9578947  0.1099492   8.712  < 2e-16 ***
## MapPair_refQWQuantity-Numeral 0.1894737  0.1099492   1.723   0.0851 .
## ModalityEnglish:TimingLater   0.3910448  0.1811299   2.159   0.0311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.516 on 1131 degrees of freedom
## Multiple R-squared:  0.2739, Adjusted R-squared:  0.2688
## F-statistic: 53.33 on 8 and 1131 DF,  p-value: < 2.2e-16
```

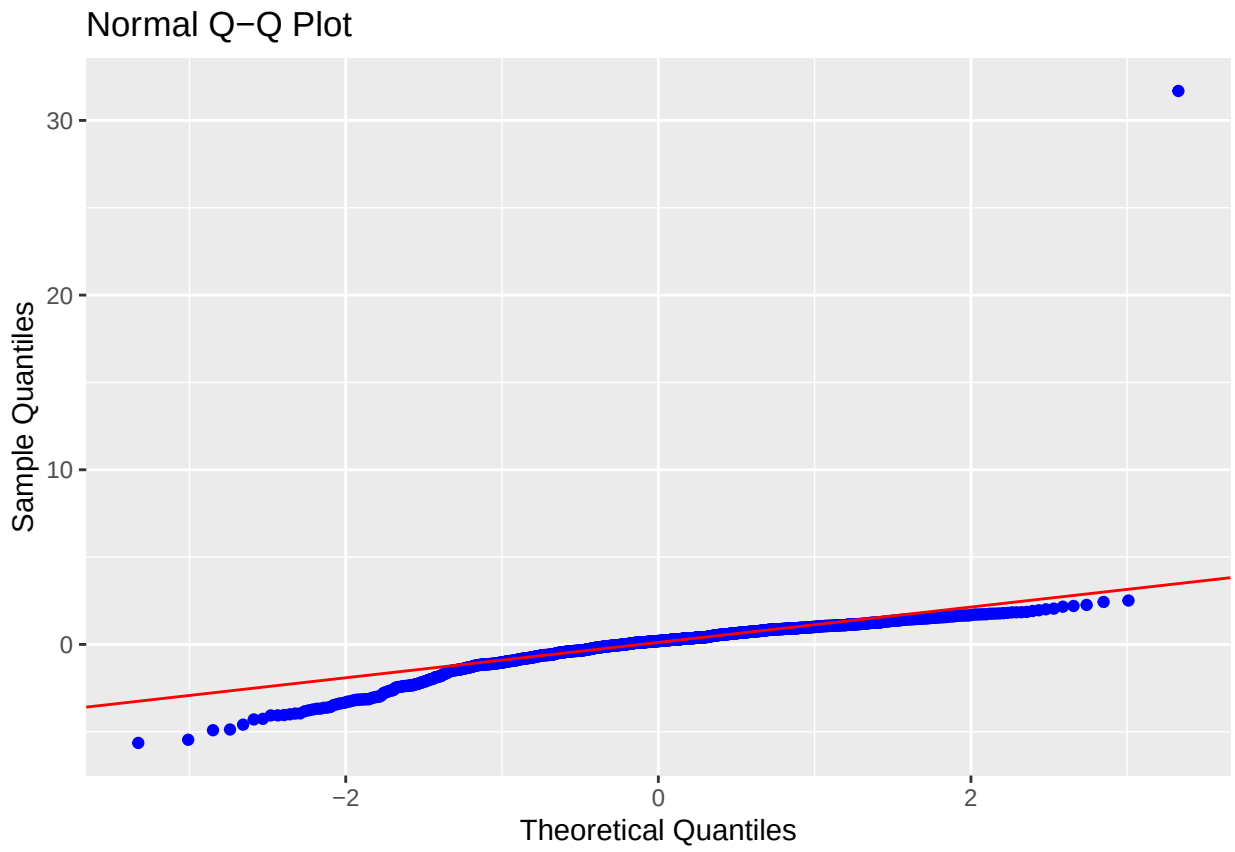
```
Map_lm_null <- lm(SumCorr ~ 1, data=L_short)

anova(Map_lm, Map_lm_null) #p < .001
```

```
## Analysis of Variance Table
##
## Model 1: SumCorr ~ Age_Rounded + SES + Modality + Timing + SetSize + MapPair +
##     Timing:Modality
## Model 2: SumCorr ~ 1
```

```
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   1131 2597.8
## 2   1139 3577.7 -8   -979.92 53.329 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#https://cran.rstudio.com/web/packages/olsrr/vignettes/residual_diagnostics.html
ols_plot_resid_qq(Map_lm)
```

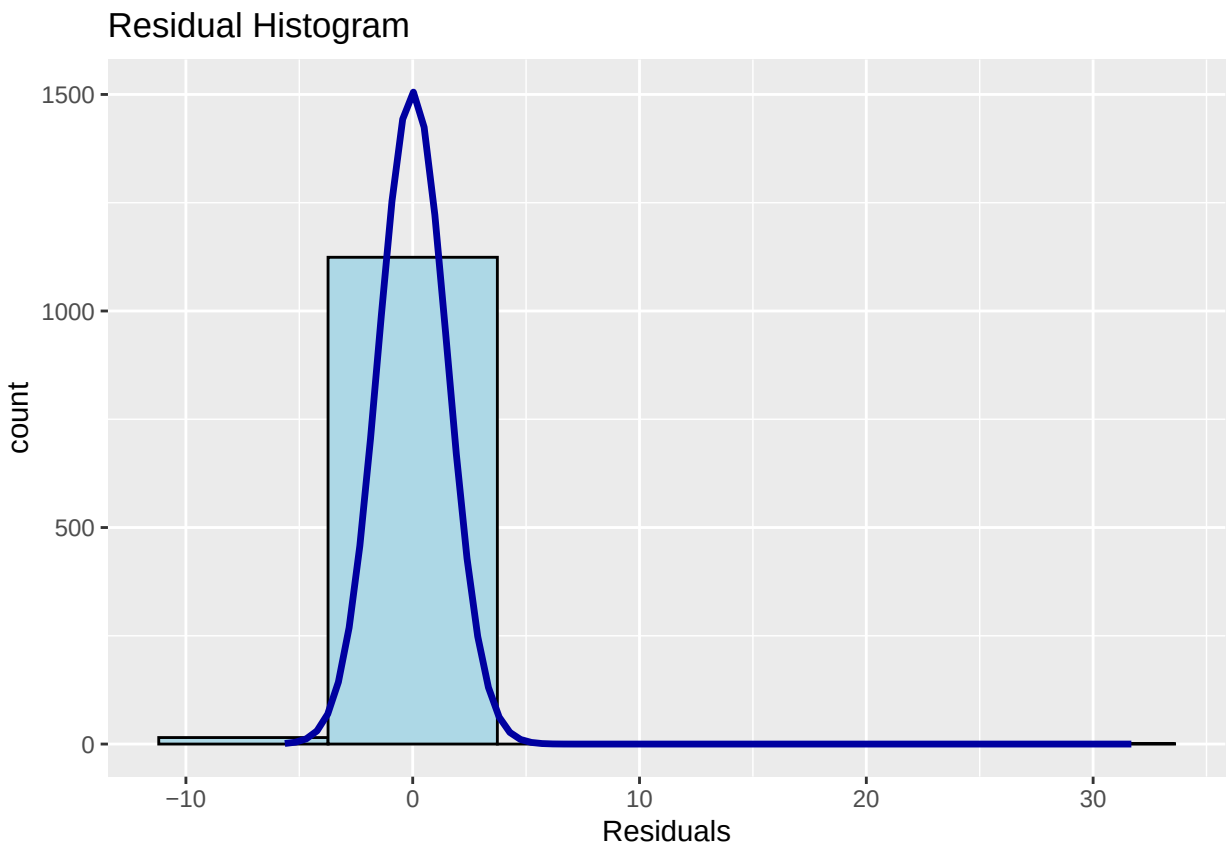



```
ols_test_normality(Map_lm)
```

```
## Warning in ks.test(y, "pnorm", mean(y), sd(y)): ties should not be present for
## the Kolmogorov-Smirnov test
```

```
## -----------------------------------------------
##        Test             Statistic       pvalue
## -----------------------------------------------
## Shapiro-Wilk              0.6562         0.0000
## Kolmogorov-Smirnov        0.1253         0.0000
## Cramer-von Mises         51.5972         0.0000
## Anderson-Darling         43.6089         0.0000
## -----------------------------------------------
```

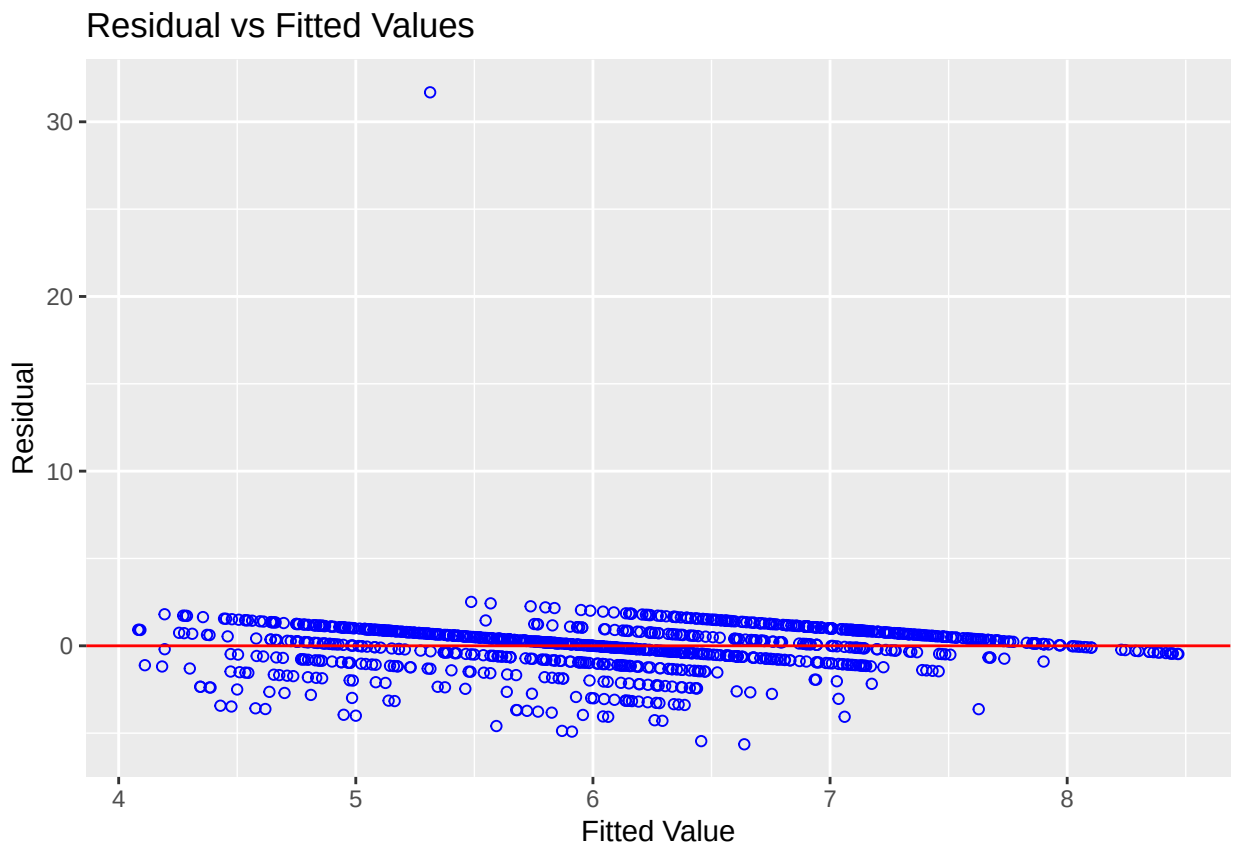
```
ols_plot_resid_hist(Map_lm)
```




```
ols_test_correlation(Map_lm)
```

```
## [1] 0.8063088
```

```
ols_plot_resid_fit(Map_lm)
```



OTHER CODES IN ANOTHER FILE... EDIT ONCE DECIDE TOBIT OR LINEAR!