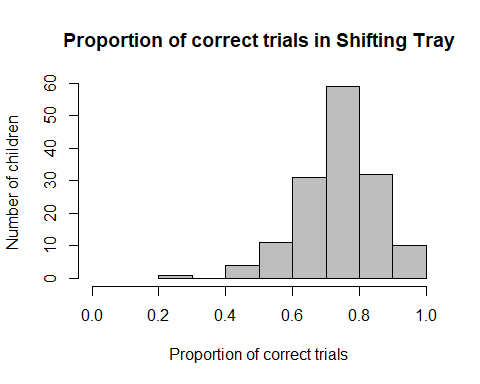
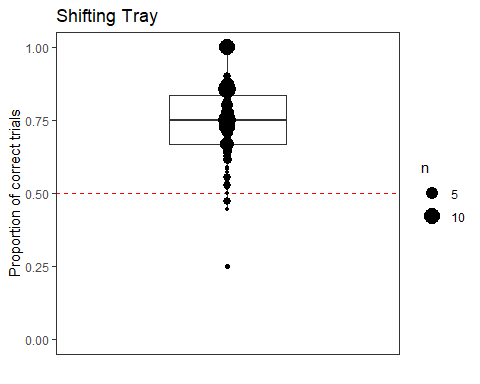
Shifting\_Tray

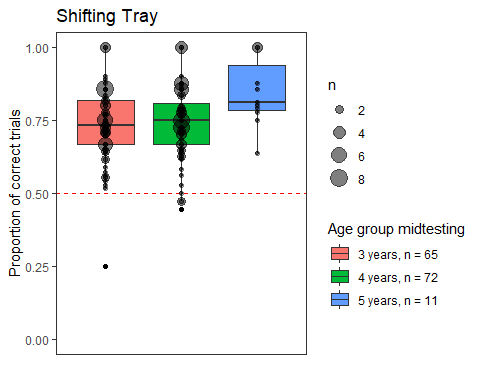
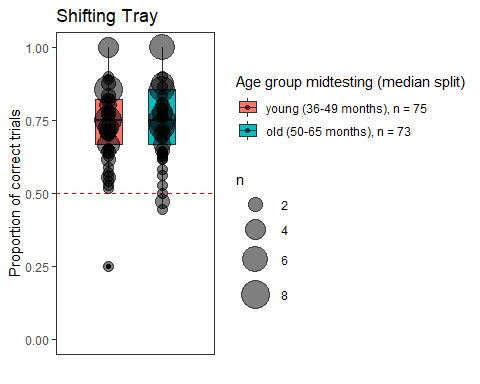
Eva Reindl

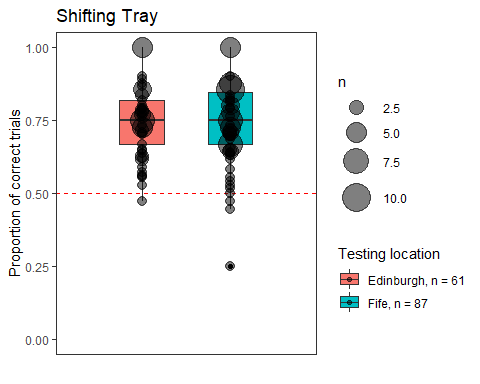
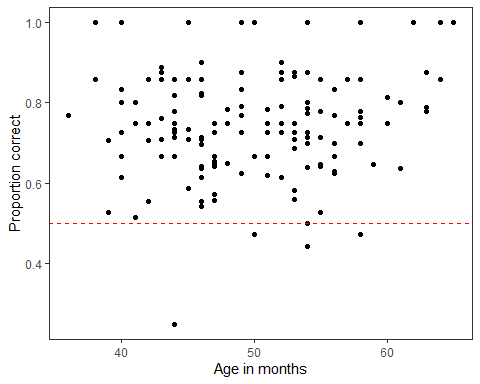
6 4 2020

Key highlights:

* 148 valid datapoints
* DV: Proportion of correct trial: **74.50% (SD = 12.67, range 25-100%)**
  + **not normally distributed**, W = 0.973, p = .005
  + Children’s performance (all age groups, young and old children) was **above chance**
  + 4- year-olds not better than 3-year-olds, older children not better than younger children
  + Effect of trial number on performance (but not of age or testing location; only marginally significant effect of BPVS score)

## [1] "R version 3.6.1 (2019-07-05)"

The entire dataset consists of 151 children.

# Dropouts

There are **3 dropouts** due to experimenter error (in 2 cases, the sticker was accidentally placed into the wrong tray and in one case the sticker was missing from the correct tray but the experimenter did not reward the child).

# Valid data

The final dataset consists of **148 children**.

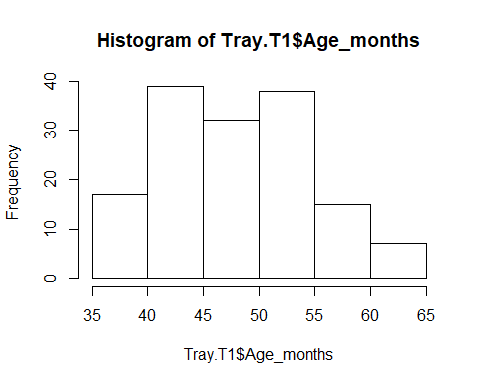
# Sample description

## Gender distribution

There are 77 females and 71 males.

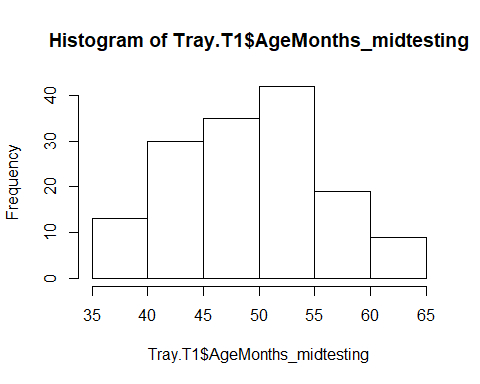
# Age

## Age at beginning of testing



At the beginning of testing, children taking part in the Shifting Tray task were on average 48.57 months (SD = 6.64, range 36-64) old. There were 68 3-year-olds, 73 4-year-olds, and 7 5-year-olds.

## Age in the middle of testing



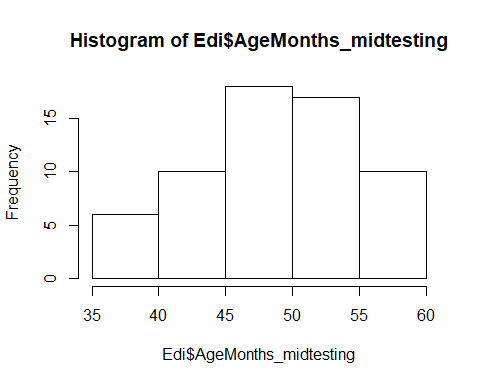
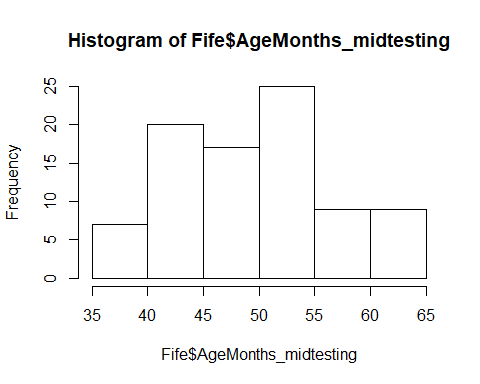
In the middle of testing, children taking part in the Shifting Tray task were on average **49.68 months (SD = 6.56, range 36-65)** old. There were

* 65 3-year-olds
* 72 4-year-olds
* 11 5-year-olds

## Age mediansplit (based on entire sample)

There were **75 young** and **73 old** children.

# Testing location

61 children were from Edinburgh, 87 children from Fife.

Edinburgh: M = 49.16 (SD = 5.76, range 36-58)

* 3y: 26
* 4y: 37

Fife: M = 50.04 (SD = 7.08, range 38-65)

* 3y: 39
* 4y: 37
* 5y: 11

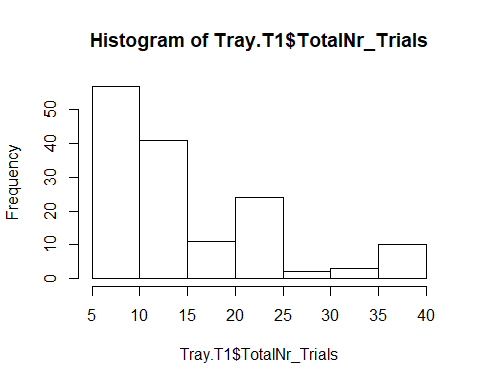
There is no difference in the age distribution between the two testing locations, two-sided Wilcoxon test, W = 2565, p = .731.

# Criterion reached

Of the 148 children, **139 (94%) reached the learning criterion**. 9 children (6%) failed to reach the learning criterion.

* 3y (n = 65): **61 reached criterion (94%)**
* 4y (n = 72): **67 reached criterion (93%)**
* 5y (n = 11): **11 reached criterion (100%)**
* young (n = 75): 71 reached criterion (95%)
* old (n = 73): 68 reached criterion (93%)

# Total number of trials



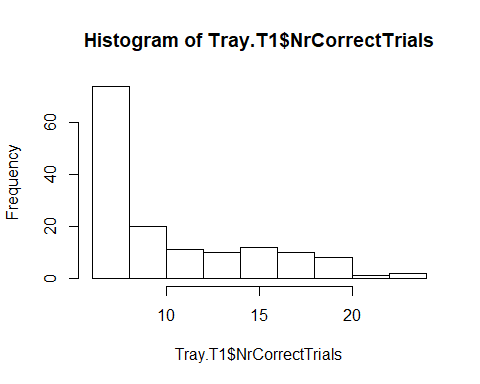
The average number of trials administered was 14.99 (SD = 8.57, range 6-36). 50% of the children were administered 12 or fewer trials. The distribution of administered trials is skewed.

* 3-year-olds: M = 15.48 (8.47, range 6-36)
* 4-year-olds: M = 15.37 (8.95, range 6-36)
* 5-year-olds: M = 9.64 (4.27, range 6-19)

Thus, numerically, 3- and 4-year-olds were perfoming equally well, while 5-year-olds were perfoming better.

* young children: 15.40 (SD = 8.26, range 6-36)
* old children: 14.57 (SD = 8.92, range 6-36) So using the median split, there was no difference between the groups.

# Number of correct trials



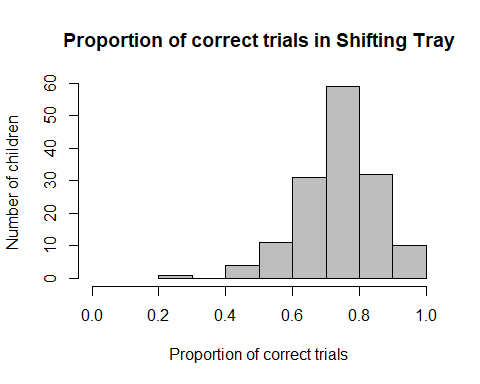
The average number of correct trials was **10.36 (SD = 4.52, range 6-24)**. 50% of the children had 8.5 or fewer trials correct. The distribution of correct trials is highly skewed.

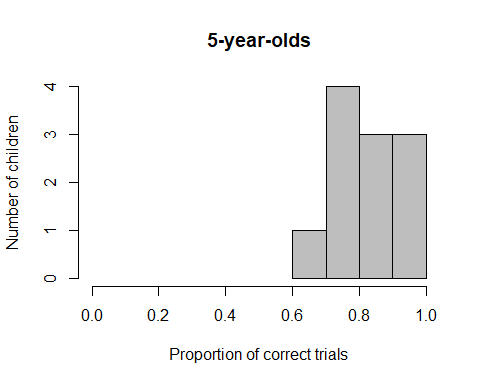
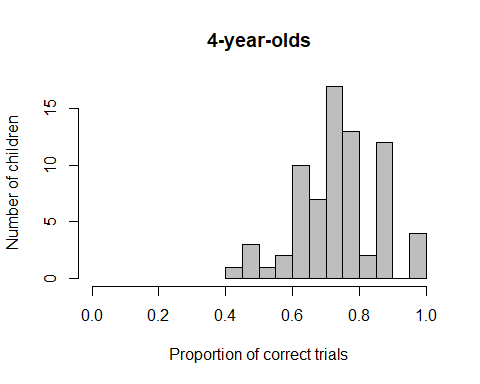
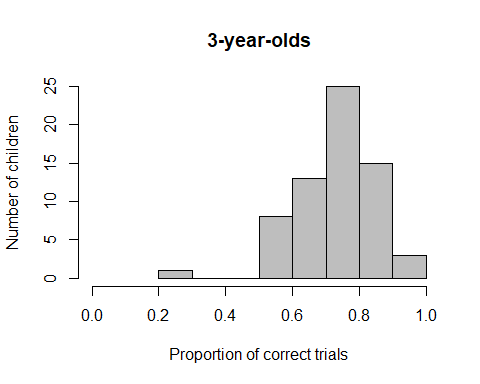
* 3-year-olds: M = 10.49 (4.28, range 6-20)
* 4-year-olds: M = 10.62 (4.84, range 6-24)
* 5-year-olds: M = 7.91 (3.11, range 6-15)

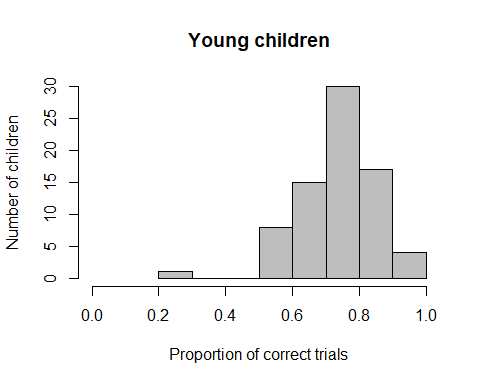
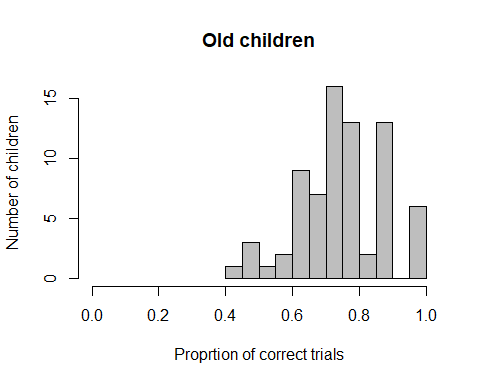
Thus, numerically, 3- and 4-year-olds were perfoming equally, while 5-year-olds were perfoming better.

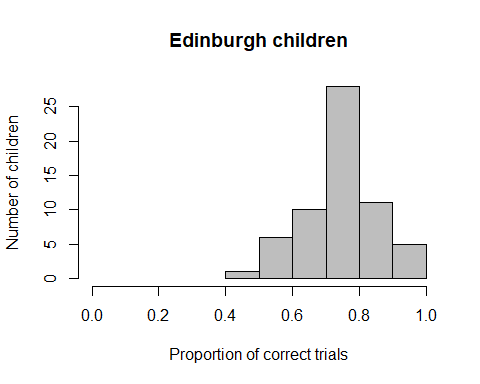
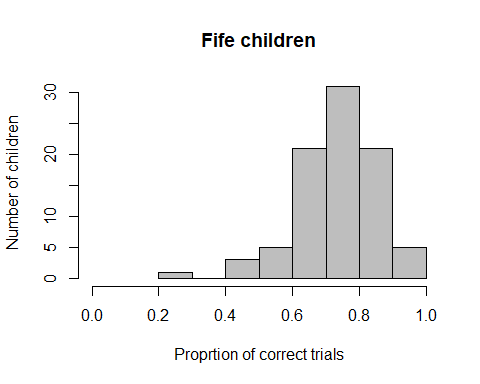
* young children: 10.59 (SD = 4.32, range 6-20)
* old children: 10.14 (SD = 4.73, range 6-24) So using the median split, there was no difference between the groups.

# DV: Proportion correct





The average proportion of correct trials was **74.50% (SD = 12.67, range 25-100%)**. 50% of the children had 75% or a smaller proportion of their trials correct. The DV is **not normally distributed**, W = 0.973, p = .005. Children’s **performance was significantly better than what would be expected by a chance performance (0.5)**, two-sided Wilcoxon test, V = 10789, p < .001.

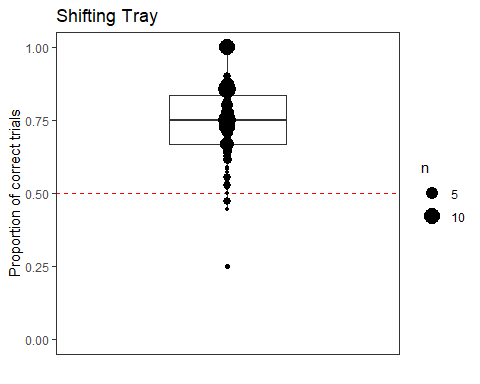
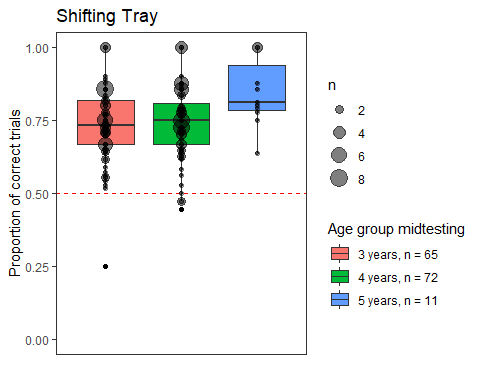
* 3-year-olds: **M = 73.17% (12.85, range 25-100%), performance significantly above chance** (0.5), V = 2108, p < .001
* 4-year-olds: **M = 74.17% (12.12, range 44-100%), performance significantly above chance** (0.5), t(71) = 16.915, p < .001
* 5-year-olds: **M = 84.53 (11.68, range 63.64-100%), performance significantly above chance** (0.5), t(10) = 9.807, p < .001

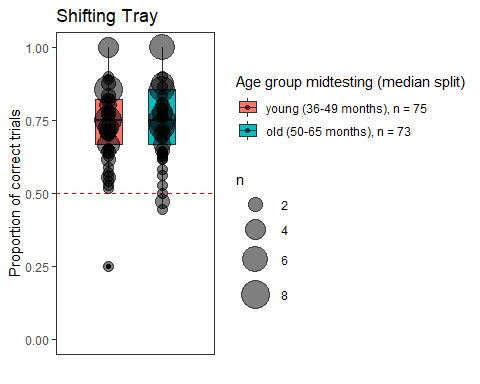
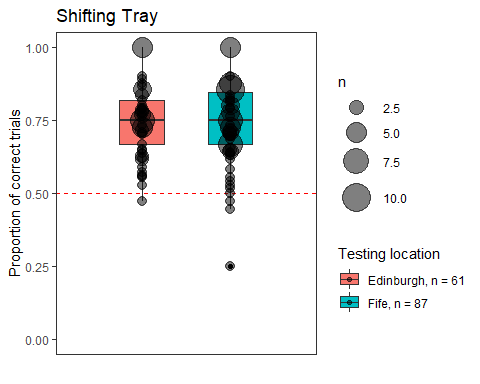
Thus, numerically, 3- and 4-year-olds were perfoming equally, while 5-year-olds were perfoming better. There was no difference between 3 and 4-year-olds, W = 2230.5, p = .319.

* young children: **73.82 (SD = 12.64, range 25-100%), performance significantly above chance** (0.5), V = 2809.5, p < .001
* old children: **75.20 (SD = 12.75, range 44-100%), performance significantly above chance** (0.5), t(72) = 16.889, p < .001

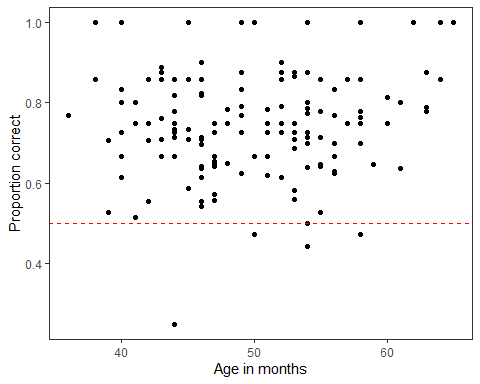
So using the median split, there was no difference between the groups, , W = 2902.5, p = .264.

* Edinburgh: **75.04 (SD = 12.17, range 47.22-100%), performance significantly above chance** (0.5), t(60) = 16.072, p < .001
* Fife: **74.12 (SD = 13.06, range 25-100%), performance significantly above chance** (0.5), V = 3687.5, p < .001

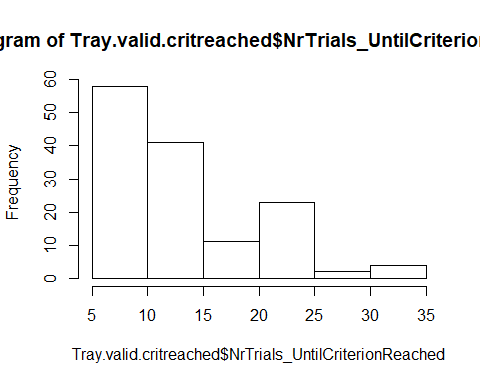
 

# Plot age as continuous variable against proportion correct



# Trials needed to criterion



**139 children (94%) reached the learning criterion**. Of those, children needed on average **13.47 trials (SD = 6.75, range 6-35)** to reach criterion. 50% of the children needed 11 or fewer trials, 75% od the children needed 17 or fewer trials. The distribution of this variable was skewed.

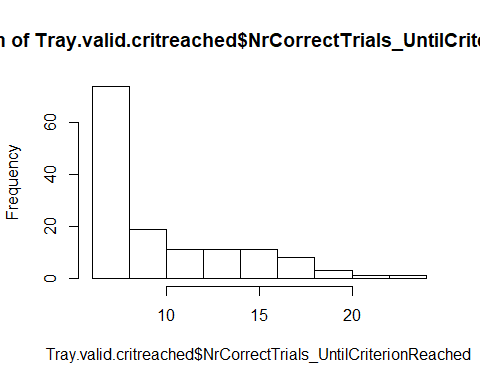
* **3-year-olds: M = 14.11** (6.80, range 6-34)
* **4-year-olds: M = 13.52** (6.91, range 6-35)
* **5-year-olds: M = 9.64** (4.27, range 6-19)

Thus, numerically, 3- and 4-year-olds were perfoming equally, while 5-year-olds were perfoming better.

* young children: 14.17 (SD = 6.73, range 6-34)
* old children: 12.76 (SD = 6.76, range 6-35)

So using the median split, there was only a slight better performance of older children.

# For those children who reached the criterion, how many correct trials did they have?



Children had on average 9.85 trials (SD = 4.08, range 6-24) correct. The distribution of this variable was highly skewed.

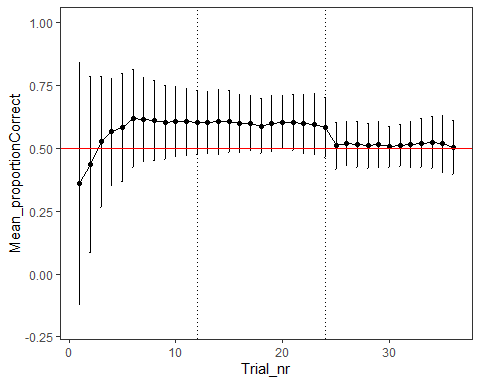
* 3-year-olds: M = 10.05 (3.85, range 6-19)
* 4-year-olds: M = 9.98 (4.39, range 6-24)
* 5-year-olds: M = 7.91 (3.11, range 6-15)

Thus, numerically, there was a slight age effect.

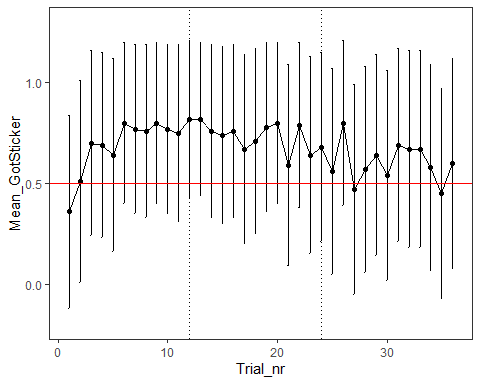
* young children: 10.15 (SD = 3.90, range 6-19)
* old children: 9.53 (SD = 4.27, range 6-24)

So using the median split, there was no difference between the groups.

# Cumulative proportion correct



# For each trial number, what is the mean success?



# Can children’s success be predicted by age?

res<-glmer(Trial\_Got\_sticker ~ z.age.midtesting + z.Trial\_no + z.age.midtesting:z.Trial\_no + (1+z.Trial\_no|ID), data=Tray.valid, family = binomial)#failed to converge

#increase number of iterations  
contr<-glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=10000000))  
  
res<-glmer(Trial\_Got\_sticker ~ z.age.midtesting + z.Trial\_no + z.age.midtesting:z.Trial\_no + (1+z.Trial\_no|ID), data=Tray.valid, family = binomial, control=contr)#singular fit

## boundary (singular) fit: see ?isSingular

#remove correlation between random effects  
res<-glmer(Trial\_Got\_sticker ~ z.age.midtesting + z.Trial\_no + z.age.midtesting:z.Trial\_no + (1|ID) + (0+z.Trial\_no|ID), data=Tray.valid, family = binomial, control=contr)#no singular fit warning

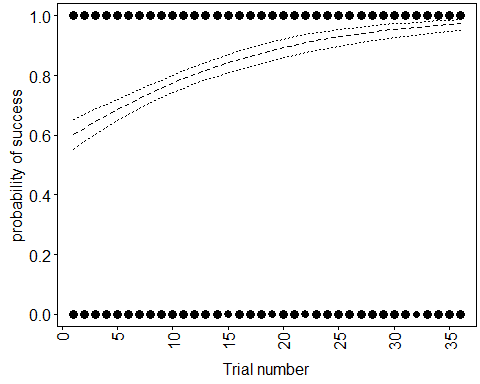
Trial number, age, and the interaction between trial number and age together can explain the data significantly better than a null model only containing an intercept, X2(3) = 54.32 p < .001.

The effect of interaction is not significant, X2(1) = 0.048, p = .827, therefore, we remove the term from the model. We aimed to include the non-transformed version of trial number, but then the model resultes in a singular fit warning, so we kept the z-transformed version.

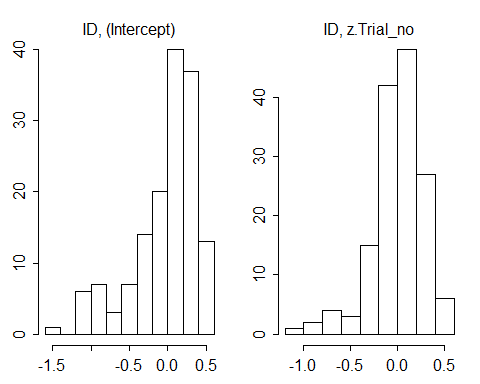
res<-glmer(Trial\_Got\_sticker ~ z.age.midtesting + z.Trial\_no + (1|ID) + (0+z.Trial\_no|ID), data=Tray.valid, family = binomial, control=contr)

Trial number and age can explain the data significantly better than a null model only containing an intercept, X2(2) = 54.276, p < .001. If we specify the null model including trial number, we find that age does not improve model fit, X2(1) = 1.212, p = .271.

There is a significant effect of trial, X2(1) = 53.349, p < .001, but no effect of age X2(1) = 1.212, p = .271.



## Model assumptions



#collinearity

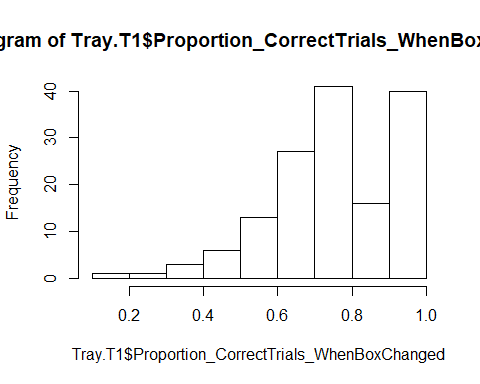
## z.age.midtesting z.Trial\_no   
## 1.000866 1.000866

source("glmm\_stability.r")  
m.stab=glmm.model.stab(model.res=res, contr=contr)  
m.stab$summary

## what orig min max  
## (Intercept) (Intercept) 1.2668376 1.25550896 1.2946535  
## z.age.midtesting z.age.midtesting 0.0821568 0.06050037 0.1049546  
## z.Trial\_no z.Trial\_no 0.7271453 0.70572490 0.7621157  
## ID@(Intercept) ID@(Intercept) 0.7872008 0.76733051 0.7984015  
## ID.1@z.Trial\_no ID.1@z.Trial\_no 0.7216677 0.68617314 0.7327763

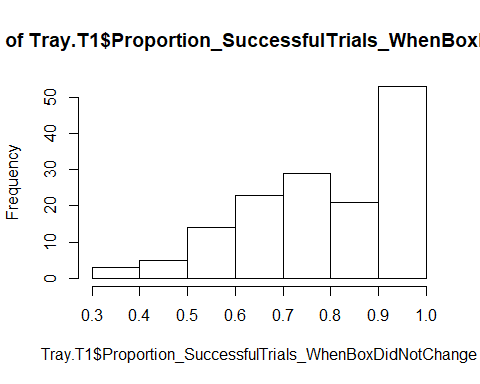
# Box change

## Proportion of correct trials out of all trials where a box change happened



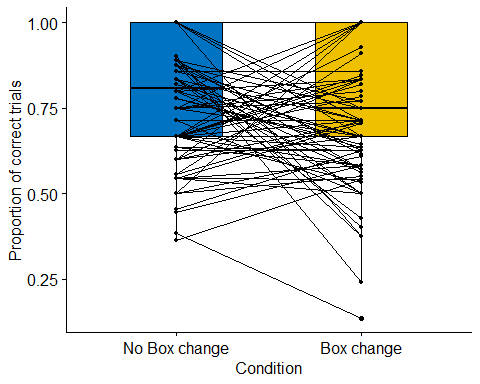
**When children just experienced a change in boxes** (i.e., the sticker was previously hidden in the yellow box and then it was hidden in the purple box, or vice versa), children had on average **76.42% (SD= 17.97, range 13.64-100%) of these trials correct**.

## Proportion of correct trials out of all trials where no box change happened



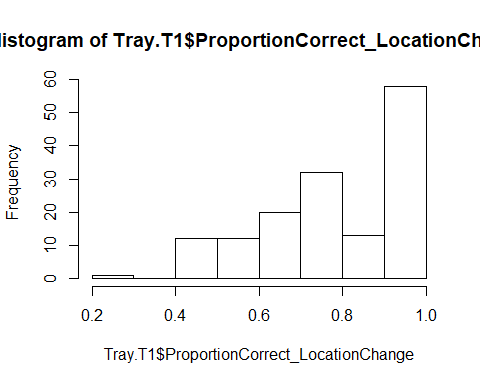
When children did **not experience a change in boxes** (i.e., the sticker was previously hidden in the yellow box and then it was again hidden there), children had on average **81.01% (SD= 17.47, range 36.36-100%) of these trials correct**.

Children’s **success rate when there was no box change was significantly higher compared to when there was a box change**, one-sided Wilcoxon test, V = 4467.5, p = .005.



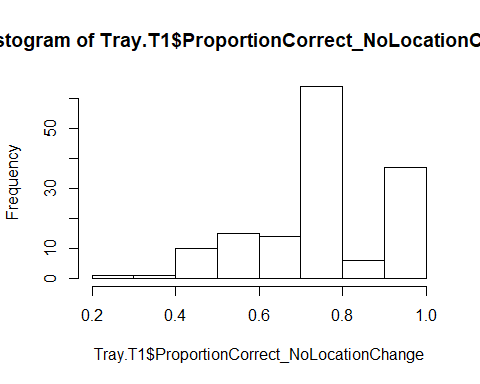
# Location change

## What was the proportion of correct trials out of all trials where a location change happened?



**When children just experienced a location change** in tray, children had on average **80.43% (SD= 18.46, range 21.05-100%) of these trials correct**.

## What was the proportion of correct trials out of all trials where no location change happened?



When children did **not experience a location change**, children had on average **77.24% (SD= 16.67, range 25-100%)** of these trials correct.

Children’s **success rate when there was no location change was not higher compared to when there was a location change**, one-sided Wilcoxon test, V = 2479.5, p = .968. Note, however, that a two-sided test reached marginal significance (V = 2479.5, p = .064).

