Picture Naming Study: Analysis

Abhilasha Kumar

November 16, 2017

1 Reading File

```
> pic = read.csv("Compiled_FINAL_CSV.csv", header = TRUE, sep = ",")
> pic_mainvariables = pic[, c(2,3,22,26,28,31,33,37,38,39, 8)]
> colnames(pic_mainvariables) = c("Subject", "Session", "Trial", "Object",
+                 "PictureType", "RT", "ObjectNo", "Name", "InvalidTrial",
+                 "Accuracy", "ItemCount")
> pic_mainvariables_valid = subset(pic_mainvariables,
+                                  pic_mainvariables$InvalidTrial == "0")
```

2 Making the z-scores

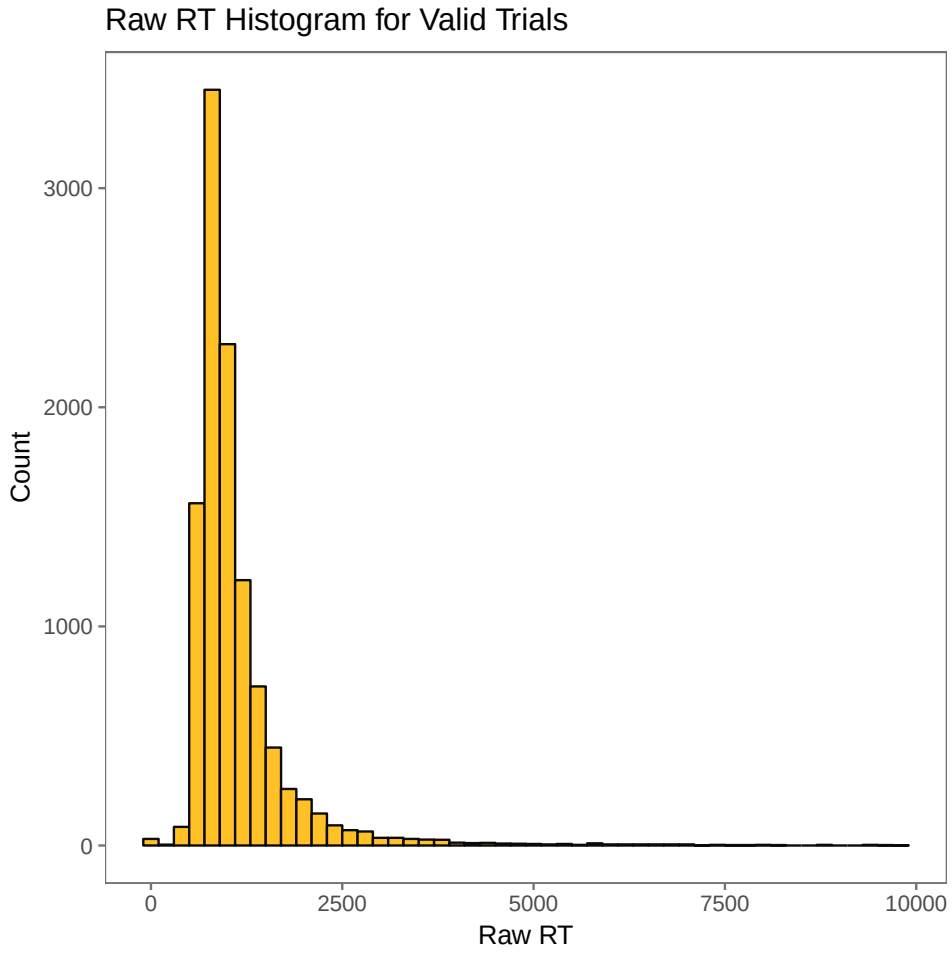
```
> library(dplyr)
> pic_firsttrim = pic_mainvariables_valid %>% filter( RT > 250 & RT < 5000)
> ## aggregate per subject all IVs and DVs
> meanRT = group_by(pic_firsttrim, Subject) %>%
+   summarise_at(vars(RT), mean)
> colnames(meanRT) = c("Subject", "MeanRT")
> sdRT = group_by(pic_firsttrim, Subject) %>%
+   summarise_at(vars(RT), sd)
> colnames(sdRT) = c("Subject", "sdRT")
> RT_agg = merge(meanRT, sdRT, by = "Subject")
> ## merge aggregate info with long data
> pic_z = merge(pic_firsttrim, RT_agg, by = "Subject", all.x = T)
> ## person and grand-mean centered scores using original and aggregate
> library(dplyr)
> pic_z = pic_z %>% mutate(zRT = (RT - MeanRT)/sdRT)
> ## checking: subject level means should be zero
>
> sub_pic = group_by(pic_z, Subject) %>%
+   summarise_at(vars(zRT), mean)
>
> #write.csv(pic_z, file="pic_z.csv")
```

3 Histograms for raw and z-RT

Raw RT

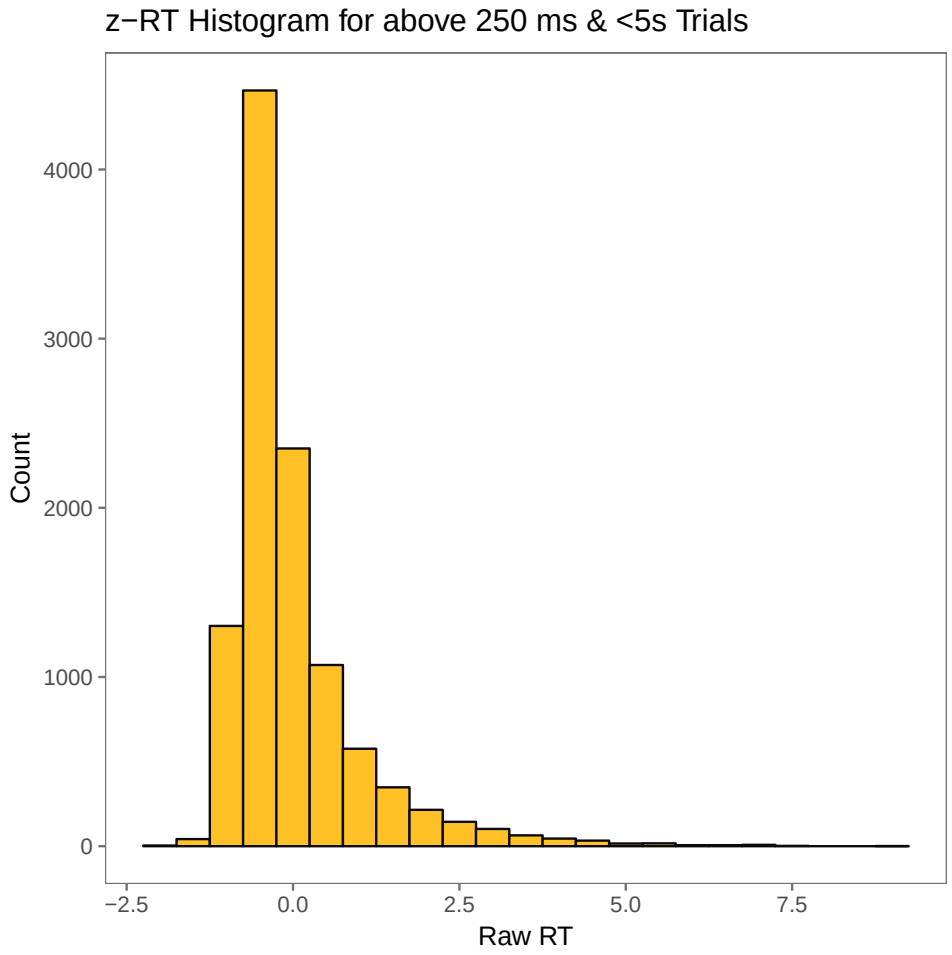
```
> library(ggplot2)
> library(ggthemes)
> ggplot(pic_mainvariables_valid, aes(x = RT))+
+ geom_histogram(binwidth = 200, color = "gray4", fill = "goldenrod1")+
+     theme_few()+
```

```
+    xlab("Raw RT") + ylab("Count") +
+    ggtitle("Raw RT Histogram for Valid Trials")
```



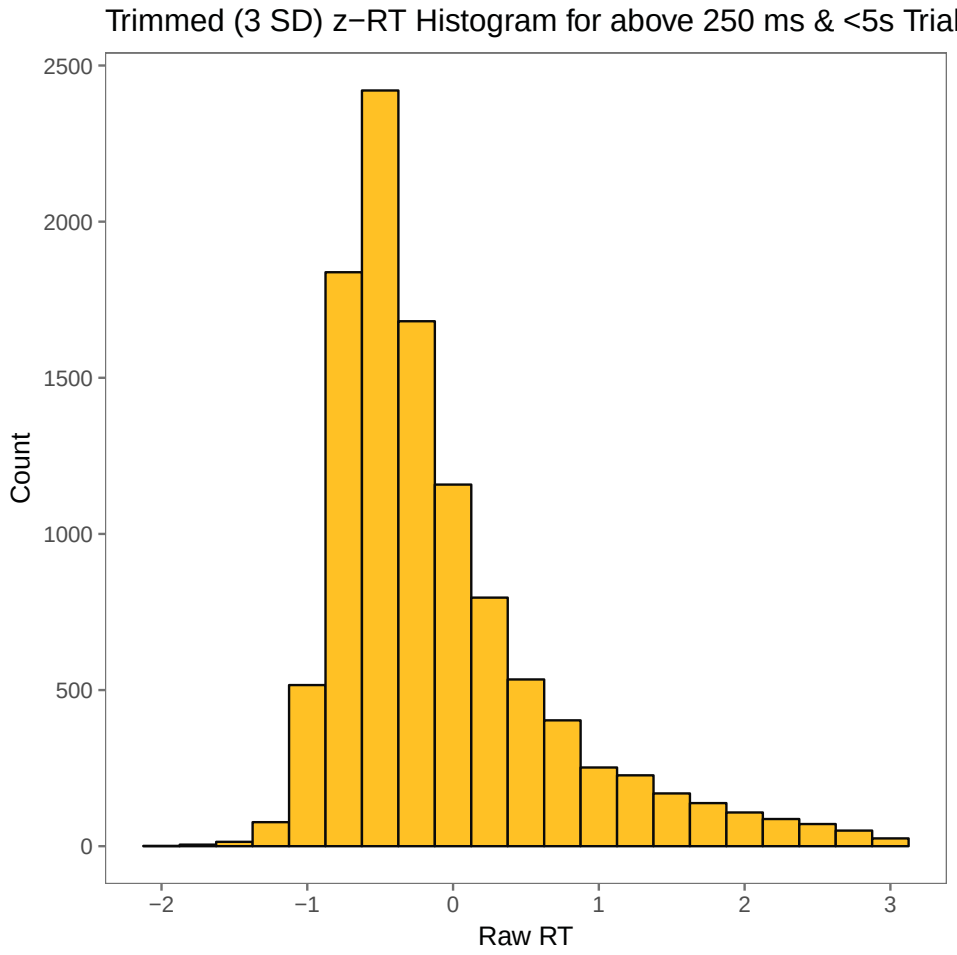
z RT

```
> ggplot(pic_z, aes(x = zRT))+
+ geom_histogram(binwidth = 0.5, color = "gray4", fill = "goldenrod1")+
+     theme_few()+
+    xlab("Raw RT") + ylab("Count") +
+    ggtitle("z-RT Histogram for above 250 ms & <5s Trials")
```



Trimming zRT

```
> ## trimming above and below 3 s.d.
> pic_z_trimmed = subset(pic_z, pic_z$zRT < 3 & pic_z$zRT > -3)
> ggplot(pic_z_trimmed, aes(x = zRT))+
+ geom_histogram(binwidth = 0.25, color = "gray4", fill = "goldenrod1")+
+     theme_few()+
+   xlab("Raw RT") + ylab("Count") +
+   ggtitle("Trimmed (3 SD) z-RT Histogram for above 250 ms & <5s Trials")
> ## now we need to repeat the z-scoring for these items -- in the next section
```



4 Repeat z-scoring after trimming

```
> library(dplyr)
> ## aggregate per subject all IVs and DVs
> meanRT_trim = group_by(pic_z_trimmed, Subject) %>%
+   summarise_at(vars(RT), mean)
> colnames(meanRT_trim) = c("Subject", "MeanRT_trim")
> sdRT_trim = group_by(pic_z_trimmed, Subject) %>%
+   summarise_at(vars(RT), sd)
> colnames(sdRT_trim) = c("Subject", "sdRT_trim")
> RT_agg_trim = merge(meanRT_trim, sdRT_trim, by = "Subject")
> ## merge aggregate info with long data
> new_pic_z = merge(pic_z_trimmed, RT_agg_trim, by = "Subject", all.x = T)
> ## person and grand-mean centered scores using original and aggregate
> library(dplyr)
> new_pic_z = new_pic_z %>% mutate(zRT_trim = (RT - MeanRT_trim)/sdRT_trim)
> ## checking: subject level means should be zero
>
> sub_pic = group_by(new_pic_z, Subject) %>%
+   summarise_at(vars(zRT), mean)
> write.csv(new_pic_z, file="final_pic_z.csv")
```

5 Excluding Subjects

```
> ## we exclude some subjects from all further RT analyses here
> library(dplyr)
> numitems = group_by(pic_mainvariables, Subject, ItemCount)%>%
+     summarise(count = n())
> ## At this point, we have 58 subjects: S1 and S2 have incorrect # items, so we remove that.
> ## And, S56, 59, 54, 57 and 60 removed for counterbalacing purposes at this point
>
> new_pic_z_final = new_pic_z %>% filter(!Subject %in% c(1,2,56,54,57,60,59))
> agg_sub = group_by(pic_mainvariables, Subject)%>%
+     summarise_at(vars(Accuracy), mean)
> # which(agg_sub$Accuracy < 0.51)
>
> ## Subject 4 is at 50% accuracy, so we remove them too
> new_pic_z_final =
+   subset(new_pic_z_final, new_pic_z_final$Subject != 4)
```

6 Aggregating RTs and Accuracy

```
> library(dplyr)
> agg_pic_validRT = group_by(new_pic_z_final, Subject, PictureType)%>%
+     summarise_at(vars(Accuracy, zRT), mean)
> agg_pic_validRT$Subject <- as.factor(agg_pic_validRT$Subject)
> agg_pic_validRT$PictureType <- as.factor(agg_pic_validRT$PictureType)
> pic_RT_aov <- aov(zRT ~ PictureType + Error(Subject/PictureType),
+                   data = agg_pic_validRT )
> summary(pic_RT_aov)
```

```
Error: Subject
          Df  Sum Sq  Mean Sq F value Pr(>F)
Residuals 49 0.08852 0.001807

Error: Subject:PictureType
            Df Sum Sq Mean Sq F value  Pr(>F)
PictureType  2 0.2558 0.12792    6.61 0.00203 **
Residuals   98 1.8967 0.01935
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> pic_acc_aov <- aov(Accuracy ~ PictureType + Error(Subject/PictureType),
+                    data = agg_pic_validRT)
> summary(pic_acc_aov)
```

```
Error: Subject
          Df Sum Sq  Mean Sq F value Pr(>F)
Residuals 49 0.3469 0.007079

Error: Subject:PictureType
            Df Sum Sq  Mean Sq F value Pr(>F)
PictureType  2 0.0031 0.001548   0.634  0.532
Residuals   98 0.2392 0.002441
```

7 Plotting Accuracy and RT

Plotting Accuracy

```
> library(Rmisc)
> agg_pic_plot_rmisc = summarySE(new_pic_z_final,
+                       measurevar = "Accuracy",
+                       groupvars = c("PictureType"))
> ggplot(agg_pic_plot_rmisc, aes(x = PictureType, y = Accuracy))+
+  geom_bar(stat = "identity", position = "dodge", width = 0.5, fill = "seagreen")+
+   geom_errorbar(aes(ymin = Accuracy - ci, ymax = Accuracy + ci),
+                 width=.05, position=position_dodge(.5)) +
+   theme_few()+
+   xlab("Picture Type") + ylab("Mean Accuracy") +
+   ggtitle("Effect of Picture Quality on Accuracy")
```

```
> agg_pic_validRT_plot_rmisc = summarySE(new_pic_z_final,
+                       measurevar = "zRT",
+                       groupvars = c("PictureType"))
> ggplot(agg_pic_validRT_plot_rmisc, aes(x = PictureType, y = zRT))+
+  geom_bar(stat = "identity", position = "dodge", width = 0.5, fill = "seagreen")+
+   geom_errorbar(aes(ymin = zRT - ci, ymax = zRT + ci),
+                 width=.05, position=position_dodge(.5)) +
+   theme_few()+
+   xlab("Picture Type") + ylab("Mean zRT") +
+   ggtitle("Effect of Picture Quality on Reaction Times")
```

8 Fetching AoA data and Merging

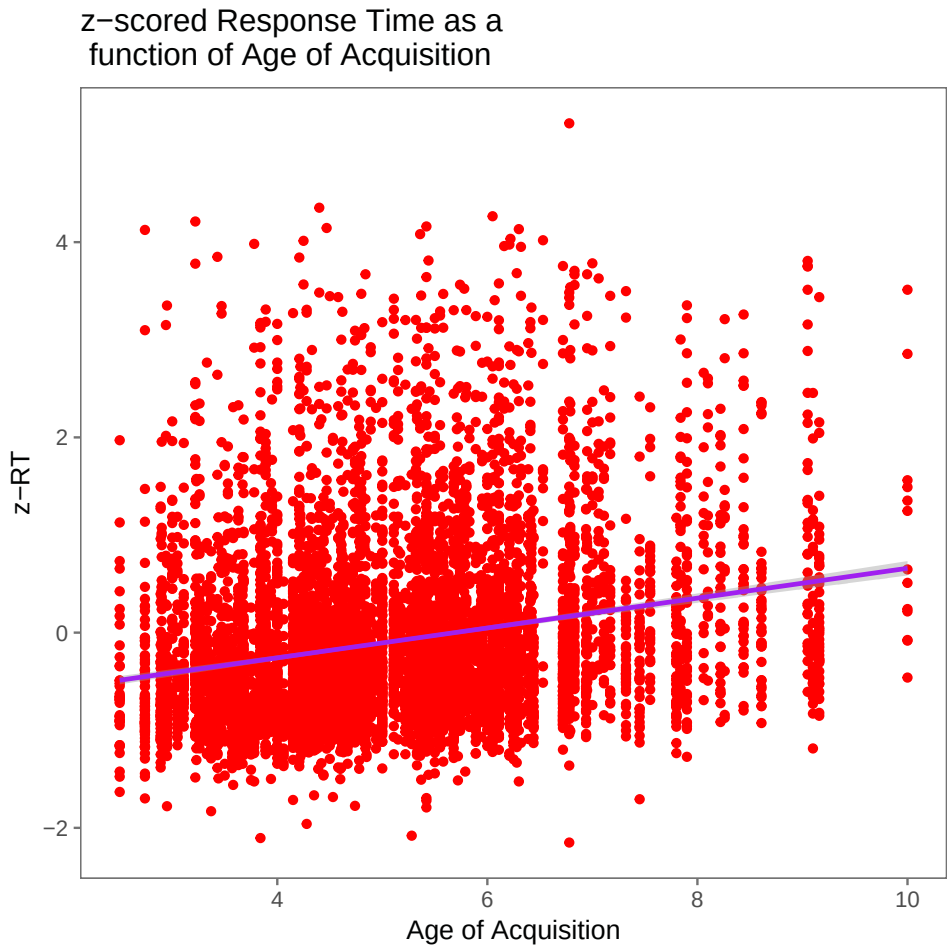
```
> AoA <- read.csv("AoA_51715_words.csv", header = TRUE, sep = ",")
```

```
> ## we need ONLY correct trials for AoA analyses.
> pic_z_correct = subset(new_pic_z_final, new_pic_z_final$Accuracy == '1')
> pic_withAoA_z = merge(pic_z_correct, AoA, by = "Object")
> pic_withAoA_z = pic_withAoA_z[, c(1:17, 27)]
> pic_withAoA_z = pic_withAoA_z[order(pic_withAoA_z$Subject),]
```

9 Actual Plots

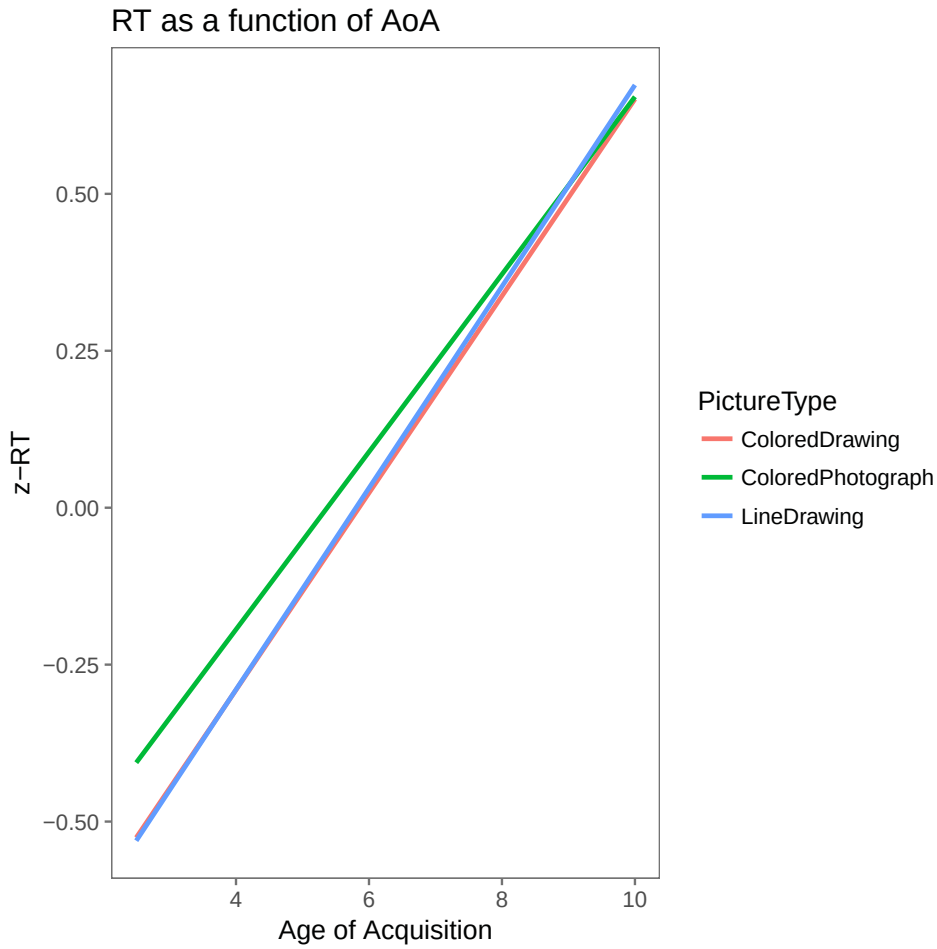
zRT and AoA

```
> ggplot(pic_withAoA_z, aes(x = AoA_Kup_lem, y = zRT_trim))+
+  geom_point(color = "red")+
+   geom_smooth(method = "lm", color = "purple")+
+  theme_few()+
+   xlab("Age of Acquisition") + ylab("z-RT") +
+   ggtitle("z-scored Response Time as a\n function of Age of Acquisition")
```



Raw Plot by Picture Type

```
> ggplot(pic_withAoA_z, aes(x = AoA_Kup_lem, y = zRT_trim,
+                           group = PictureType, color = PictureType))+
+  #geom_point(aes(color = PictureType))+
+   geom_smooth(method = "lm", se = FALSE)+
+  theme_few()+
+   xlab("Age of Acquisition") + ylab("z-RT") +
+   ggtitle("RT as a function of AoA")
```



10 Regressions

```
> library(lme4)
> m0 = lmer (data = pic_withAoA_z, zRT_trim ~ AoA_Kup_lem +
+              (1|Subject))
> summary(m0)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: zRT_trim ~ AoA_Kup_lem + (1 | Subject)
   Data: pic_withAoA_z

REML criterion at convergence: 20385.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.5968 -0.6486 -0.2789  0.3516  5.6596

Random effects:
 Groups   Name        Variance  Std.Dev.
 Subject  (Intercept) 2.582e-16 1.607e-08
 Residual             7.963e-01 8.924e-01
Number of obs: 7805, groups:  Subject, 50

Fixed effects:
            Estimate Std. Error t value
```

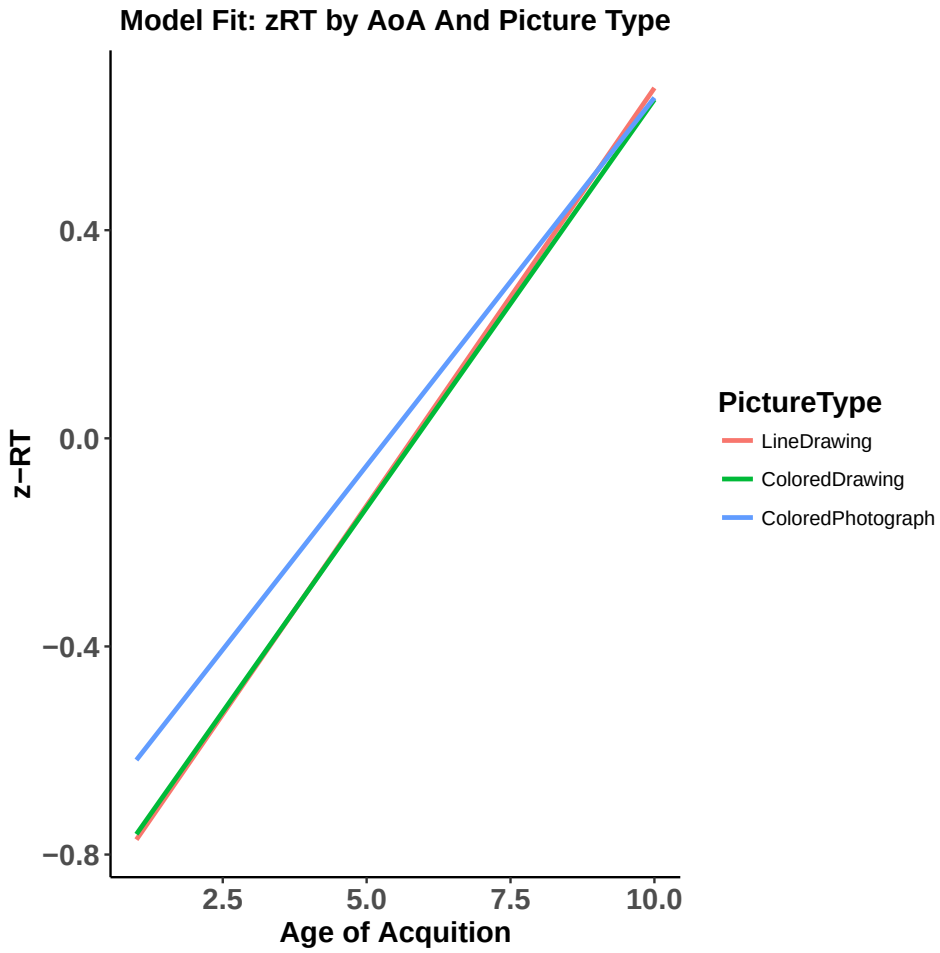
```
(Intercept) -0.869792   0.039182  -22.20
AoA_Kup_lem  0.152889   0.007427   20.59

Correlation of Fixed Effects:
            (Intr)
AoA_Kup_lem -0.966
```

```
> m1 = lmer (data = pic_withAoA_z, zRT_trim ~ AoA_Kup_lem*PictureType +
+              (1|Subject))
> summary(m1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula:
zRT_trim ~ AoA_Kup_lem * PictureType + (1 | Subject)
   Data: pic_withAoA_z

REML criterion at convergence: 20395.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.5869 -0.6487 -0.2763  0.3526  5.6868

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.0000   0.0000
 Residual             0.7953   0.8918
Number of obs: 7805, groups:  Subject, 50

Fixed effects:
                                          Estimate
(Intercept)                              -0.917774
AoA_Kup_lem                               0.156874
PictureTypeColoredPhotograph              0.158013
PictureTypeLineDrawing                   -0.014052
AoA_Kup_lem:PictureTypeColoredPhotograph -0.015434
AoA_Kup_lem:PictureTypeLineDrawing        0.003648
                                         Std. Error t value
(Intercept)                                0.067600 -13.577
AoA_Kup_lem                                0.012742  12.311
PictureTypeColoredPhotograph               0.095911   1.648
PictureTypeLineDrawing                     0.095784  -0.147
AoA_Kup_lem:PictureTypeColoredPhotograph   0.018136  -0.851
AoA_Kup_lem:PictureTypeLineDrawing         0.018150   0.201

Correlation of Fixed Effects:
            (Intr) AA_Kp_ PctTCP PctTLD AA_K_:PTC
AoA_Kup_lem -0.966
PctrTypClrP -0.705  0.681
PctrTypLnDr -0.706  0.682  0.497
AA_Kp_:PTCP  0.679 -0.703 -0.966 -0.479
AA_Kp_:PTLD  0.678 -0.702 -0.478 -0.966  0.493
```

11 Plotting Model Fits

```
> fixed.frame <-
+   data.frame(expand.grid(AoA_Kup_lem = seq(1,10,0.5),
+              PictureType = c("LineDrawing",
+                              "ColoredDrawing", "ColoredPhotograph"))) %>%
+   mutate(pred = predict(m1, newdata = ., re.form = NA))
> fixed.frame %>%
+   mutate(AoA = AoA_Kup_lem) %>%
+   ggplot(aes(x = AoA, y = pred, color = PictureType)) +
+     geom_line(size = 1) +
+     xlab("Age of Acquition") + ylab ("z-RT")+
+   ggtitle("Model Fit: zRT by AoA And Picture Type")+
+ theme_classic() +
+     theme(axis.text = element_text(face = "bold", size = rel(1.2)),
+           axis.title = element_text(face = "bold", size = rel(1.2)),
+           legend.title = element_text(face = "bold", size = rel(1.2)),
+           plot.title = element_text(face = "bold", size = rel(1.2), hjust = .5))
```



12 Analysing Data Item-Wise

We need to eliminate items that are <0.40 accuracy in any of the 3 picture types :

```
> item_acc = group_by(new_pic_z_final, Object, PictureType )%>%
+     summarise_at(vars(Accuracy), mean)
> ## so now we find the items with low accuracy
> low_acc_items = subset(item_acc, item_acc$Accuracy < 0.40)
> low_acc_items_list = unique(low_acc_items$Object)
```

```
> ## now we must eliminate these items from the new_pic_z_final file and re-run all regressions
>
> pic_z_itemsremoved = subset(new_pic_z_final, ! (new_pic_z_final$Object %in% low_acc_items_list))
> ## now we use this dataset to re-do our AoA analyses: so we first merge this with the AoA file
>
> item_withAoA_complete = merge(pic_z_itemsremoved, AoA, by = "Object")
> item_withAoA_complete = item_withAoA_complete[order(item_withAoA_complete$Subject),]
```

13 Removing Low-Accuracy items: Regressions

```
> library(lme4)
> m2 = lmer(data = item_withAoA_complete, zRT ~ AoA_Kup_lem +
+             (1|Subject))
> summary(m2)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: zRT ~ AoA_Kup_lem + (1 | Subject)
   Data: item_withAoA_complete

REML criterion at convergence: 17600.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4724 -0.6446 -0.2870  0.3407  4.6642

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.0000   0.000
 Residual             0.5098   0.714
Number of obs: 8126, groups:  Subject, 50

Fixed effects:
             Estimate Std. Error t value
(Intercept) -0.646391   0.030752  -21.02
AoA_Kup_lem  0.101144   0.005863   17.25

Correlation of Fixed Effects:
            (Intr)
AoA_Kup_lem -0.966
```

```
> m3 = lmer(data = item_withAoA_complete, zRT ~ AoA_Kup_lem*PictureType +
+             (1|Subject))
> summary(m3)
```

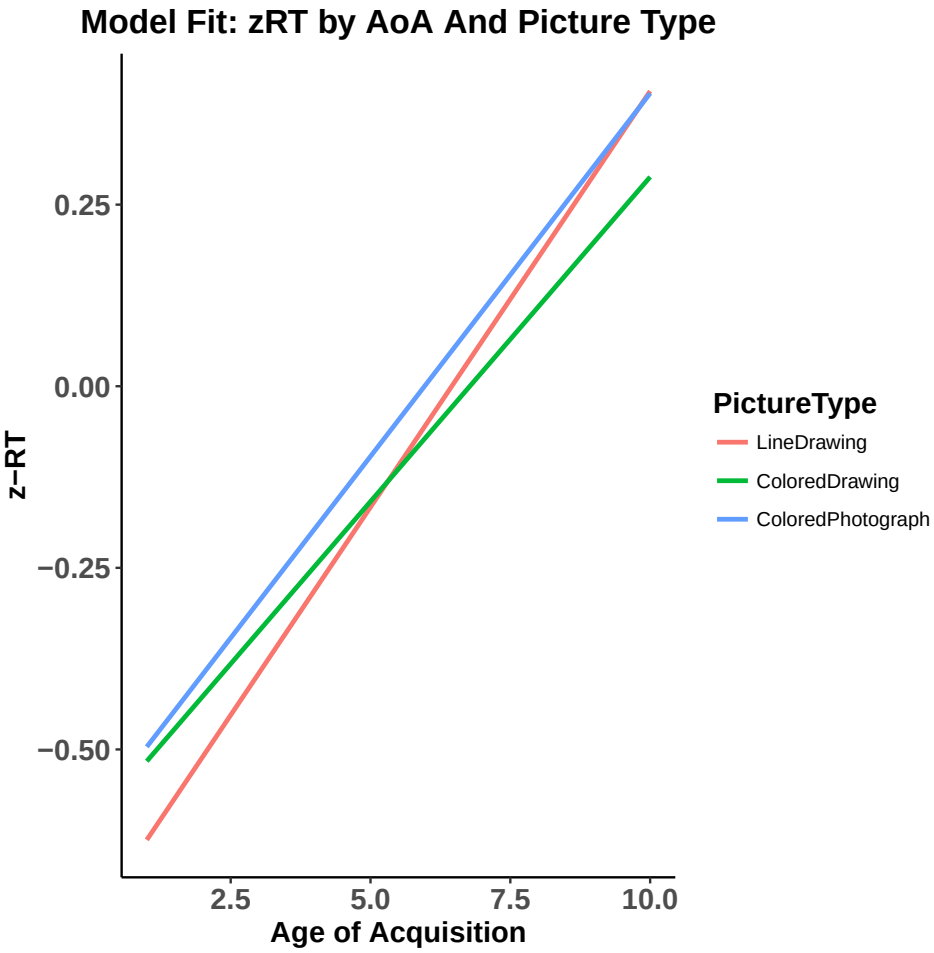
```
Linear mixed model fit by REML ['lmerMod']
Formula: zRT ~ AoA_Kup_lem * PictureType + (1 | Subject)
   Data: item_withAoA_complete

REML criterion at convergence: 17607.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4438 -0.6458 -0.2828  0.3441  4.6564
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.0000   0.0000
 Residual             0.5089   0.7134
Number of obs: 8126, groups:  Subject, 50

Fixed effects:
                                        Estimate
(Intercept)                             -0.60560
AoA_Kup_lem                              0.08937
PictureTypeColoredPhotograph             0.00917
PictureTypeLineDrawing                  -0.13347
AoA_Kup_lem:PictureTypeColoredPhotograph  0.01060
AoA_Kup_lem:PictureTypeLineDrawing        0.02516
                                        Std. Error t value
(Intercept)                                0.05315 -11.395
AoA_Kup_lem                                0.01011   8.841
PictureTypeColoredPhotograph               0.07510   0.122
PictureTypeLineDrawing                     0.07538  -1.771
AoA_Kup_lem:PictureTypeColoredPhotograph    0.01431   0.741
AoA_Kup_lem:PictureTypeLineDrawing          0.01436   1.752

Correlation of Fixed Effects:
            (Intr) AA_Kp_ PctTCP PctTLD AA_K_:PTC
AoA_Kup_lem -0.966
PctrTypClrP -0.708  0.684
PctrTypLnDr -0.705  0.681  0.499
AA_Kp_:PTCP  0.682 -0.707 -0.966 -0.481
AA_Kp_:PTLD  0.680 -0.704 -0.481 -0.966  0.497
```

Modified Plot

We also plot the curve again, to see if it's any different:

```
> fixed.frame.2 <-
+   data.frame(expand.grid(AoA_Kup_lem = seq(1,10,0.5),
+              PictureType = c("LineDrawing",
+                              "ColoredDrawing", "ColoredPhotograph"))) %>%
+   mutate(pred = predict(m3, newdata = ., re.form = NA))
> fixed.frame.2 %>%
+   mutate(AoA = AoA_Kup_lem) %>%
+   ggplot(aes(x = AoA, y = pred, color = PictureType)) +
+     geom_line(size = 1) +
+     xlab("Age of Acquisition") + ylab ("z-RT")+
+   ggtitle("Model Fit: zRT by AoA And Picture Type")+
+ theme_classic() +
+     theme(axis.text = element_text(face = "bold", size = rel(1.2)),
+           axis.title = element_text(face = "bold", size = rel(1.2)),
+           legend.title = element_text(face = "bold", size = rel(1.2)),
+           plot.title = element_text(face = "bold", size = rel(1.4), hjust = .5))
```


Model Fit: zRT by AoA And Picture Type
0.25
0.00
−0.25
−0.50
z−RT
2.5
5.0
7.5
10.0
Age of Acquisition
PictureType
LineDrawing
ColoredDrawing
ColoredPhotograph