

# Data analysis pilots

Nathanael Larigaldie

2024-01-19

## Contents

Dataset loading . . . . .	1
Dataset preparation . . . . .	1
Variables plot (violin & boxplot + density) . . . . .	1
Statistical models . . . . .	2

## Dataset loading

```
datasets      <- list.files(file.path("final_data"), pattern="*.csv", full.names = TRUE)
datasets      <- lapply(datasets, read_csv, show_col_types = FALSE)

d <- datasets[[1]]
for (i in seq(from=2, to=length(datasets)))
{
  d <- rbind(d, datasets[[i]])
}
```

## Dataset preparation

```
d <- d %>% mutate(Subject=rep(seq_len(length(datasets)), each=100))
```

Then we remove total dwell times  $\leq 1500$ ms, scale the Agreement feature and take its square, and log-transform Dwell.Time. We do the same with Agreement.Contacts

```
d_final <- d %>% filter(Dwell.Time > 1500) %>%
  mutate(Dwell.Time.Log = log(Dwell.Time),
         Agreement.Scaled = (Agreement-mean(Agreement))/sd(Agreement),
         Squared.Agreement.Scaled=Agreement.Scaled^2,
         Agreement.Contacts.Scaled = (Agreement.Contacts-mean(Agreement.Contacts))/sd(Agreement.Contacts),
         Squared.Agreement.Contacts.Scaled=Agreement.Contacts.Scaled^2,
         Index=as.factor(Index), Subject=as.factor(Subject))
```

## Variables plot (violin & boxplot + density)

```
violin_fixation <- ggplot(d_final, aes(x="Dwell time", y=Dwell.Time.Log)) +
  ylab("Dwell Time (log)") + theme_minimal() +
  theme(legend.position = "none") +
  geom_violin(width=1.2) + geom_boxplot(width=0.1, alpha=0.2)
```

```

violin_agreement <- ggplot(d_final, aes(x="Agreement", y=Agreement.Scaled)) +
  ylab("Agreement (scaled)") + theme_minimal() +
  theme(legend.position = "none") +
  geom_violin(width=1.2) + geom_boxplot(width=0.1, alpha=0.2)

violin_agreement_contacts <- ggplot(d_final, aes(x="Agreement (contacts)", y=Agreement.Contacts.Scaled)) +
  ylab("Agreement (contacts; scaled)") + theme_minimal() +
  theme(legend.position = "none") +
  geom_violin(width=1.2) + geom_boxplot(width=0.1, alpha=0.2)

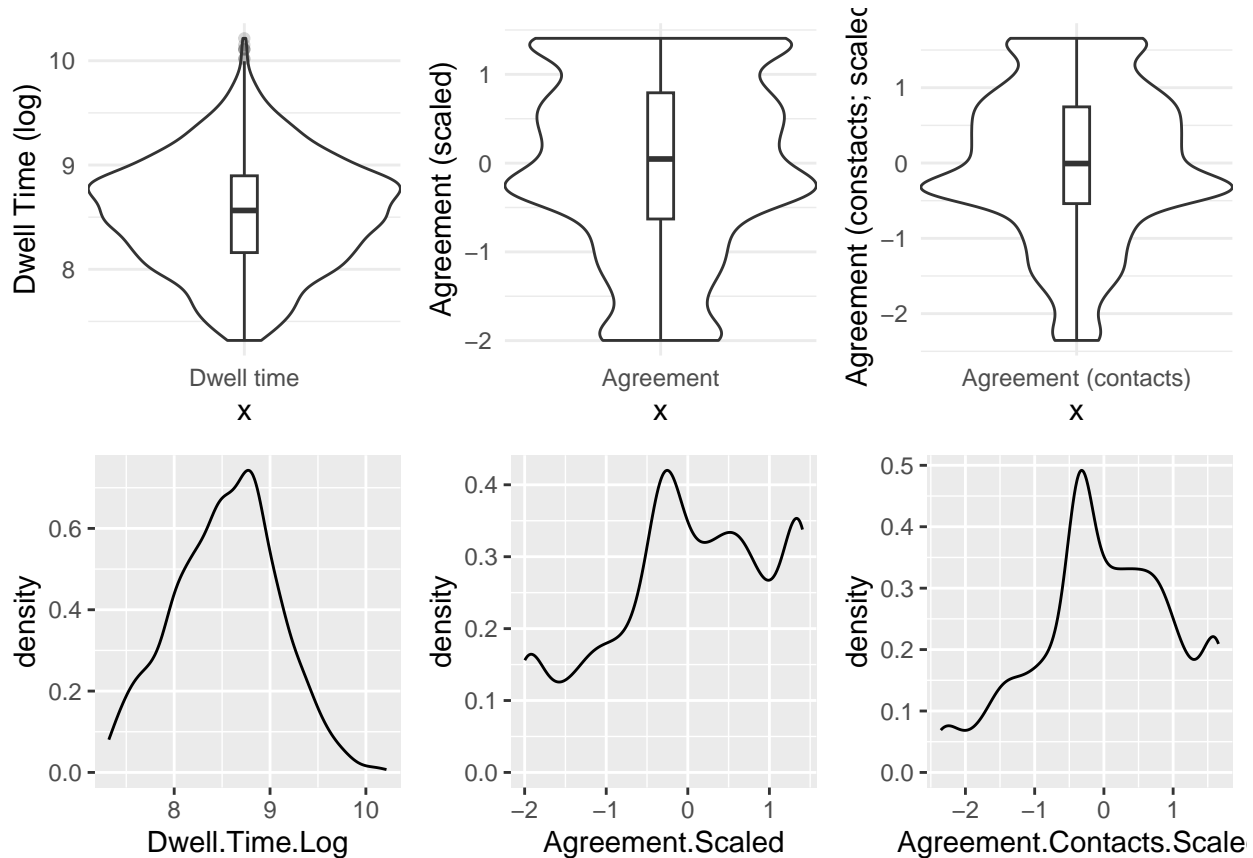
density_fixation <- ggplot(d_final, aes(x=Dwell.Time.Log)) + geom_density()

density_agreement <- ggplot(d_final, aes(x=Agreement.Scaled)) + geom_density()

density_agreement_contacts <- ggplot(d_final, aes(x=Agreement.Contacts.Scaled)) + geom_density()

grid.arrange(violin_fixation, violin_agreement, violin_agreement_contacts,
  density_fixation, density_agreement, density_agreement_contacts,
  nrow=2)

```



## Statistical models

### Agreement

```

res_mixed <- lmer(Dwell.Time.Log ~ Agreement.Scaled + Squared.Agreement.Scaled + Condition_Share + Agree
  (1 | Index) +

```

```

      (Agreement.Scaled + Squared.Agreement.Scaled + Condition_Share + Agreement.Scaled*Condition_Share)
data=d_final, control = lmerControl(optimizer = "bobyqa"))

## boundary (singular) fit: see help('isSingular')

## Warning: Model failed to converge with 2 negative eigenvalues: -1.3e-05
## -4.7e-04

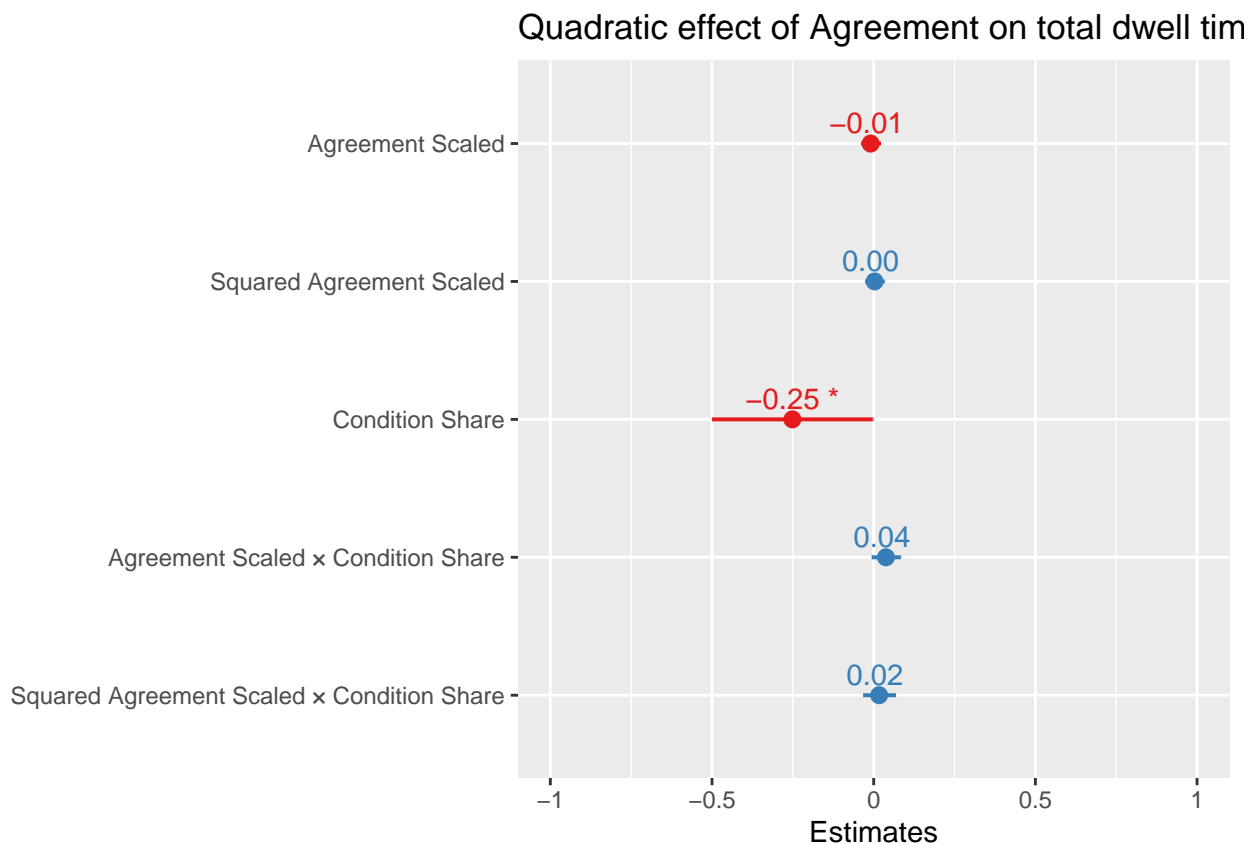
summary(res_mixed)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Dwell.Time.Log ~ Agreement.Scaled + Squared.Agreement.Scaled +
##      Condition_Share + Agreement.Scaled * Condition_Share + Squared.Agreement.Scaled *
##      Condition_Share + (1 | Index) + (Agreement.Scaled + Squared.Agreement.Scaled +
##      Condition_Share + Agreement.Scaled * Condition_Share + Squared.Agreement.Scaled *
##      Condition_Share | Subject)
## Data: d_final
## Control: lmerControl(optimizer = "bobyqa")
##
## REML criterion at convergence: 2820.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2546 -0.6495 -0.0344  0.6160  3.7541
##
## Random effects:
##      Groups      Name                Variance Std.Dev. Corr
##      Index      (Intercept)          0.0200991 0.14177
##      Subject    (Intercept)          0.0820210 0.28639
##              Agreement.Scaled        0.0002062 0.01436  0.45
##              Squared.Agreement.Scaled 0.0004584 0.02141 -0.84
##              Condition_Share         0.0990397 0.31471 -0.36
##              Agreement.Scaled:Condition_Share 0.0010536 0.03246  0.26
##              Squared.Agreement.Scaled:Condition_Share 0.0030229 0.05498  0.56
## Residual                        0.1823998 0.42708
##
##
##
##      -0.86
##      -0.53  0.53
##      -0.52  0.17  0.07
##      0.32 -0.51 -0.66  0.59
##
## Number of obs: 2256, groups:  Index, 100; Subject, 26
##
## Fixed effects:
##
##              Estimate Std. Error      df t value
## (Intercept)    8.603539   0.082744 12.749158 103.978
## Agreement.Scaled -0.009477   0.014720 38.665423  -0.644
## Squared.Agreement.Scaled 0.002797   0.014496 23.461014   0.193
## Condition_Share -0.251466   0.126338 23.200045 -1.990
## Agreement.Scaled:Condition_Share 0.037930   0.022303 38.706171   1.701
## Squared.Agreement.Scaled:Condition_Share 0.016859   0.025027 18.692890   0.674

```

```
##                                Pr(>|t|)
## (Intercept)                   <2e-16 ***
## Agreement.Scaled              0.5235
## Squared.Agreement.Scaled      0.8487
## Condition_Share               0.0585 .
## Agreement.Scaled:Condition_Share 0.0970 .
## Squared.Agreement.Scaled:Condition_Share 0.5088
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) Agrm.S Sq.A.S Cndt_S A.S:C_
## Agrmnt.Scld  0.058
## Sqrd.Agrm.S -0.479  0.170
## Conditn_Shr -0.635 -0.038  0.309
## Agrmn.S:C_S -0.042 -0.586 -0.098  0.068
## Sqr.A.S:C_S  0.273 -0.095 -0.556 -0.343  0.425
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
plot_model(res_mixed,
  #axis.labels=c("Agreement (Quadratic)", "Agreement"),
  show.values=TRUE, show.p=TRUE,
  title="Quadratic effect of Agreement on total dwell times")
```



```
quad_function <- function(x, alpha, beta1, beta2, beta3, beta4, beta5, condition)
{
```

```

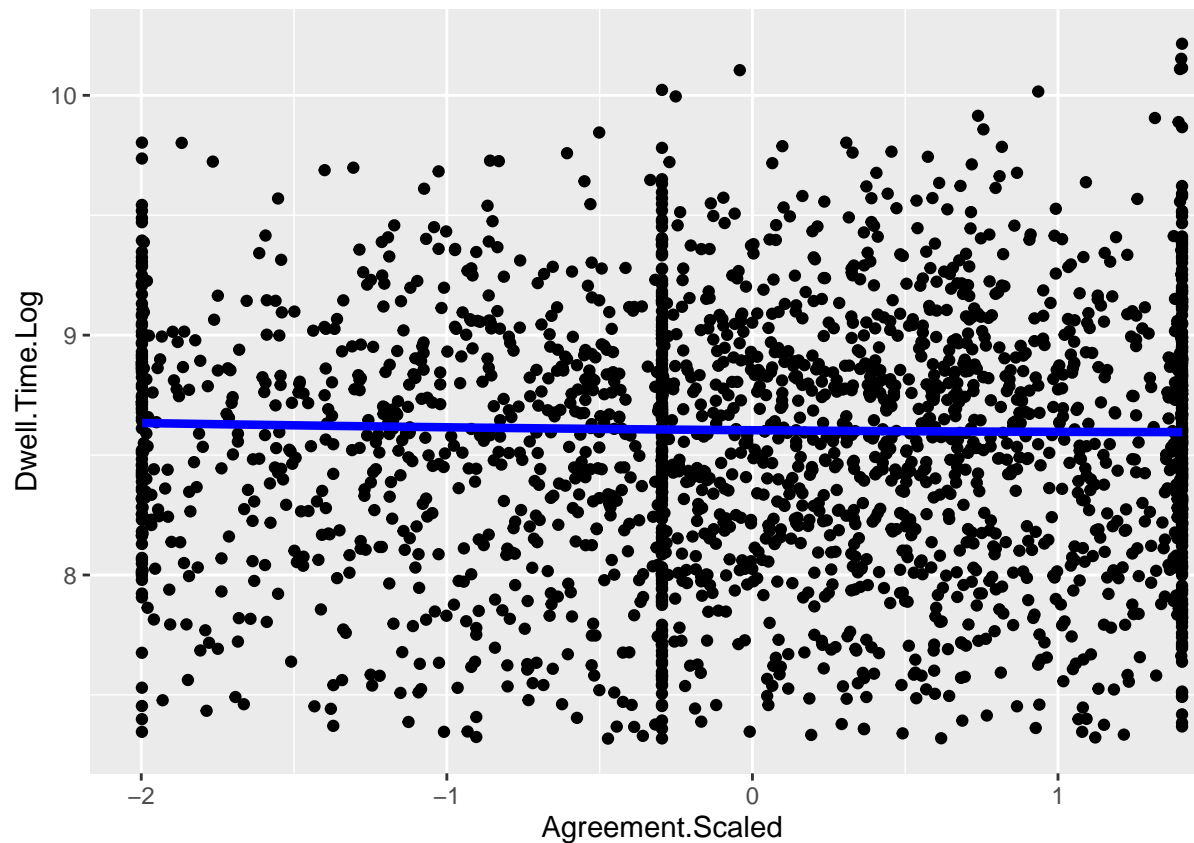
return(alpha + beta1*x + beta2*x^2 + beta3*condition + beta4*(condition*x) +
       beta5*(condition*x^2))
}

```

```

ggplot() +
  geom_point(data=d_final, aes(Agreement.Scaled, Dwell.Time.Log)) +
  geom_function(fun=quad_function,
               args=list(alpha=summary(res_mixed)[["coefficients"]][1],
                         beta1=summary(res_mixed)[["coefficients"]][2],
                         beta2=summary(res_mixed)[["coefficients"]][3],
                         beta3=summary(res_mixed)[["coefficients"]][4],
                         beta4=summary(res_mixed)[["coefficients"]][5],
                         beta5=summary(res_mixed)[["coefficients"]][6],
                         condition = 0),
               color="blue",
               linewidth=1.5)

```



Fixed effects plot

```

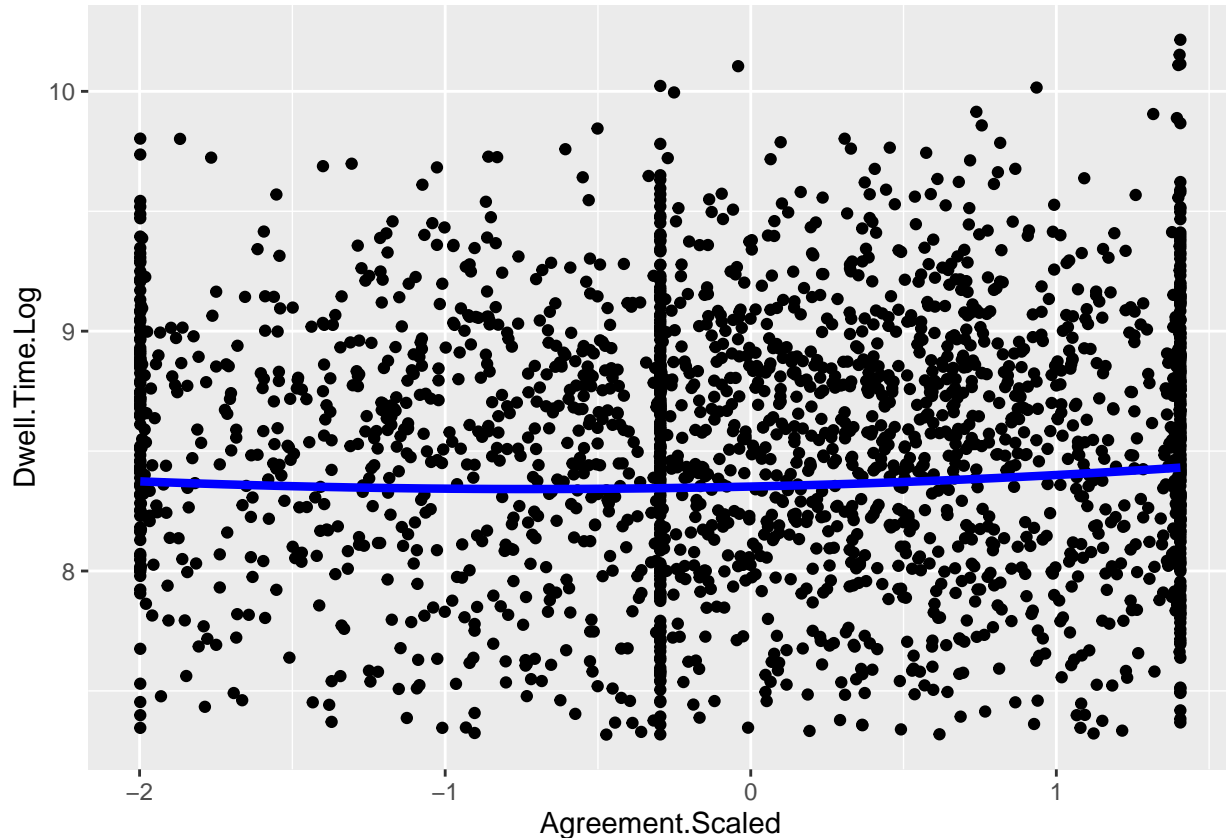
ggplot() +
  geom_point(data=d_final, aes(Agreement.Scaled, Dwell.Time.Log)) +
  geom_function(fun=quad_function,
               args=list(alpha=summary(res_mixed)[["coefficients"]][1],
                         beta1=summary(res_mixed)[["coefficients"]][2],
                         beta2=summary(res_mixed)[["coefficients"]][3],
                         beta3=summary(res_mixed)[["coefficients"]][4],
                         beta4=summary(res_mixed)[["coefficients"]][5],

```

```

beta5=summary(res_mixed)[["coefficients"]][6],
condition = 1),
color="blue",
linewidth=1.5)

```



Agreement (contacts)

```

res_mixed <- lmer(Dwell.Time.Log ~ Agreement.Contacts.Scaled + Squared.Agreement.Contacts.Scaled + Cond
              (1 | Index) +
              (Agreement.Contacts.Scaled + Squared.Agreement.Contacts.Scaled + Condition_Share + A
              data=d_final, control = lmerControl(optimizer = "bobyqa"))

```

```
## boundary (singular) fit: see help('isSingular')
```

```
## Warning: Model failed to converge with 4 negative eigenvalues: -3.6e-05
```

```
## -5.0e-05 -5.8e-05 -1.1e+01
```

```
summary(res_mixed)
```

```

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## Dwell.Time.Log ~ Agreement.Contacts.Scaled + Squared.Agreement.Contacts.Scaled +
##   Condition_Share + Agreement.Contacts.Scaled * Condition_Share +
##   Squared.Agreement.Contacts.Scaled * Condition_Share + (1 |
##   Index) + (Agreement.Contacts.Scaled + Squared.Agreement.Contacts.Scaled +
##   Condition_Share + Agreement.Contacts.Scaled * Condition_Share +

```

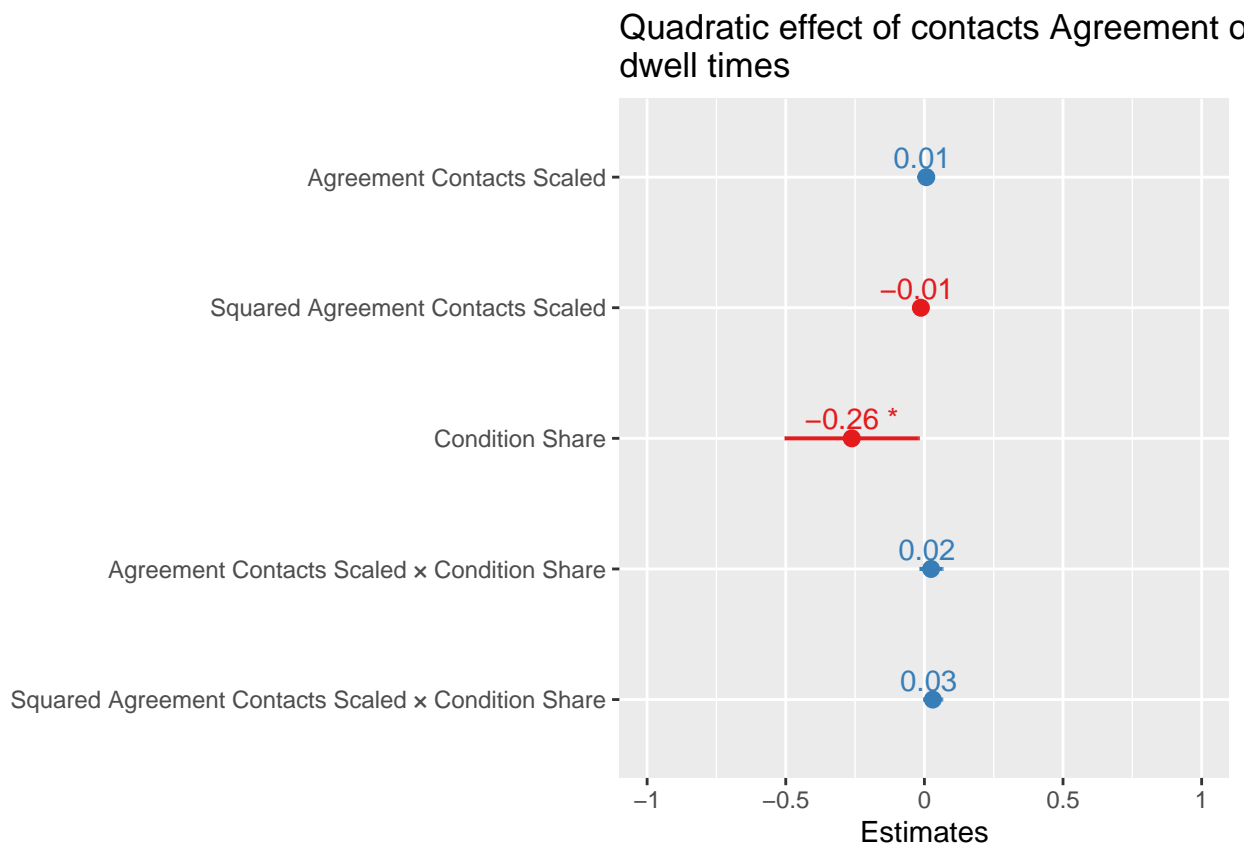
```

##      Squared.Agreement.Contacts.Scaled * Condition_Share | Subject)
##      Data: d_final
## Control: lmerControl(optimizer = "bobyqa")
##
## REML criterion at convergence: 2824.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2045 -0.6569 -0.0214  0.6181  3.7307
##
## Random effects:
##      Groups      Name                                Variance Std.Dev.
##      Index      (Intercept)                          2.015e-02 0.141950
##      Subject     (Intercept)                          7.082e-02 0.266116
##              Agreement.Contacts.Scaled                1.136e-05 0.003370
##              Squared.Agreement.Contacts.Scaled         3.750e-05 0.006124
##              Condition_Share                          1.353e-01 0.367824
##              Agreement.Contacts.Scaled:Condition_Share 7.142e-04 0.026724
##              Squared.Agreement.Contacts.Scaled:Condition_Share 2.013e-04 0.014188
##      Residual                                         1.834e-01 0.428298
##      Corr
##
##      1.00
##      0.50 0.50
##      -0.45 -0.45 -0.39
##      -0.41 -0.41 -0.04 -0.63
##      -0.49 -0.49 -0.41 -0.49 0.92
##
## Number of obs: 2256, groups:  Index, 100; Subject, 26
##
## Fixed effects:
##
##              Estimate Std. Error
## (Intercept)      8.618766   0.076945
## Agreement.Contacts.Scaled      0.006745   0.013970
## Squared.Agreement.Contacts.Scaled -0.012674   0.011433
## Condition_Share      -0.261895   0.123285
## Agreement.Contacts.Scaled:Condition_Share      0.024094   0.021005
## Squared.Agreement.Contacts.Scaled:Condition_Share 0.030468   0.017316
##
##              df t value Pr(>|t|)
## (Intercept)    12.830448 112.013 <2e-16
## Agreement.Contacts.Scaled    417.875079   0.483 0.6295
## Squared.Agreement.Contacts.Scaled     6.253790 -1.109 0.3084
## Condition_Share    22.393086 -2.124 0.0449
## Agreement.Contacts.Scaled:Condition_Share    94.336630   1.147 0.2543
## Squared.Agreement.Contacts.Scaled:Condition_Share  29.905233   1.760 0.0887
##
## (Intercept) ***
## Agreement.Contacts.Scaled
## Squared.Agreement.Contacts.Scaled
## Condition_Share *
## Agreement.Contacts.Scaled:Condition_Share
## Squared.Agreement.Contacts.Scaled:Condition_Share .
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) Ag.C.S Sq.A.C.S Cndt_S A.C.S:
## Agrmnt.Cn.S  0.030
## Sqrd.Ag.C.S -0.070  0.238
## Conditn_Shr -0.603 -0.019  0.041
## Agr.C.S:C_S -0.021 -0.610 -0.149  -0.273
## S.A.C.S:C_S  0.043 -0.159 -0.641  -0.260  0.354
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
plot_model(res_mixed,
            #axis.labels=c("Agreement (Quadratic)", "Agreement"),
            show.values=TRUE, show.p=TRUE,
            title="Quadratic effect of contacts Agreement on total dwell times")
```



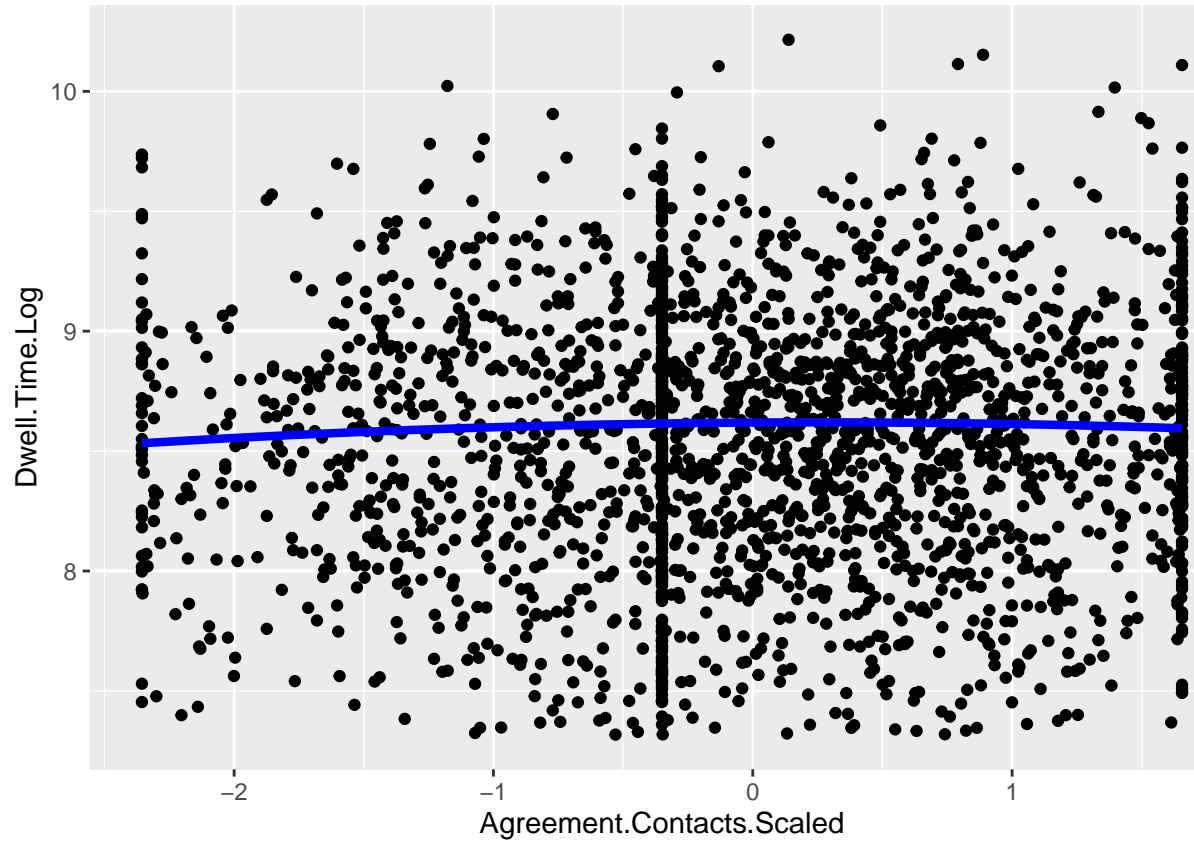
```
ggplot() +
  geom_point(data=d_final, aes(Agreement.Contacts.Scaled, Dwell.Time.Log)) +
  geom_function(fun=quad_function,
               args=list(alpha=summary(res_mixed)[["coefficients"]][1],
                         beta1=summary(res_mixed)[["coefficients"]][2],
                         beta2=summary(res_mixed)[["coefficients"]][3],
                         beta3=summary(res_mixed)[["coefficients"]][4],
                         beta4=summary(res_mixed)[["coefficients"]][5],
```



```

beta5=summary(res_mixed)[["coefficients"]][6],
condition = 0),
color="blue",
linewidth=1.5)

```

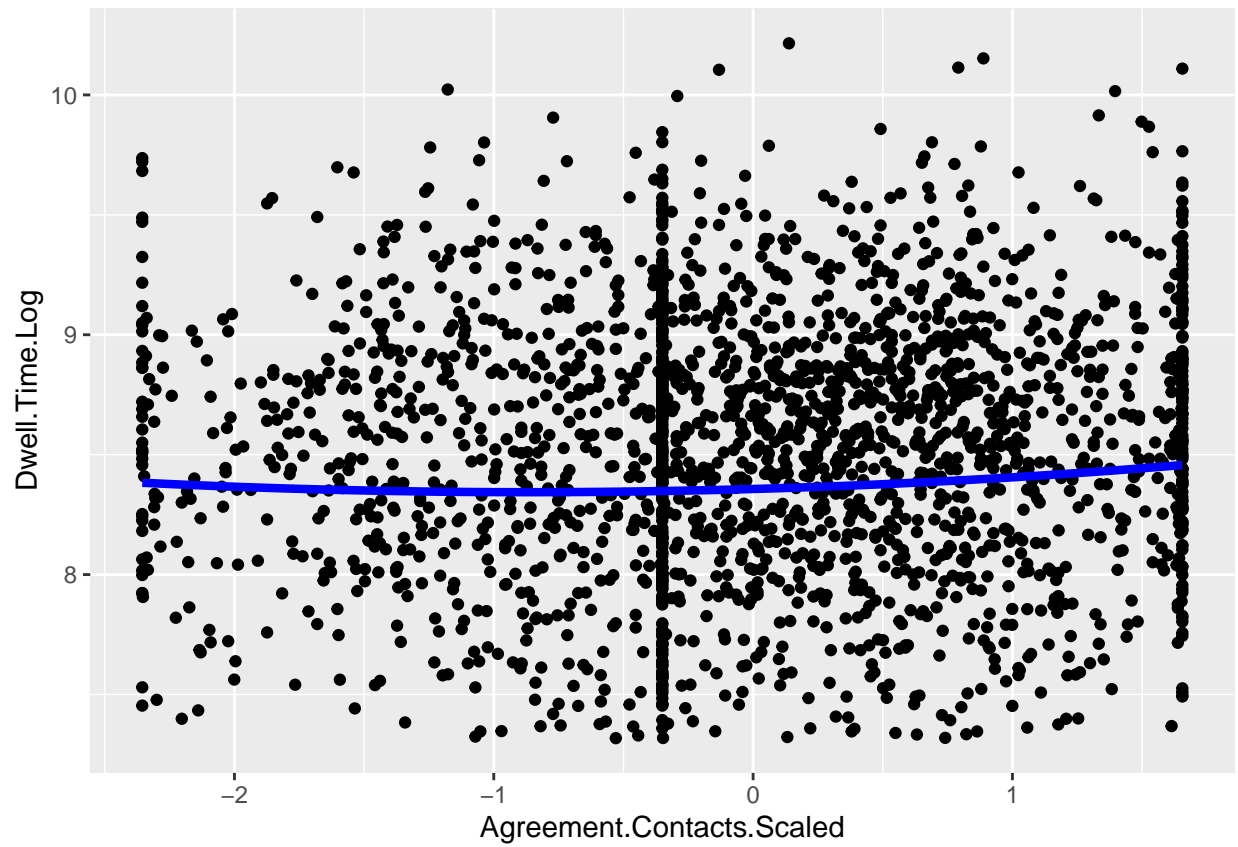


Fixed effects plot

```

ggplot() +
  geom_point(data=d_final, aes(Agreement.Contacts.Scaled, Dwell.Time.Log)) +
  geom_function(fun=quad_function,
    args=list(alpha=summary(res_mixed)[["coefficients"]][1],
              beta1=summary(res_mixed)[["coefficients"]][2],
              beta2=summary(res_mixed)[["coefficients"]][3],
              beta3=summary(res_mixed)[["coefficients"]][4],
              beta4=summary(res_mixed)[["coefficients"]][5],
              beta5=summary(res_mixed)[["coefficients"]][6],
              condition = 1),
    color="blue",
    linewidth=1.5)

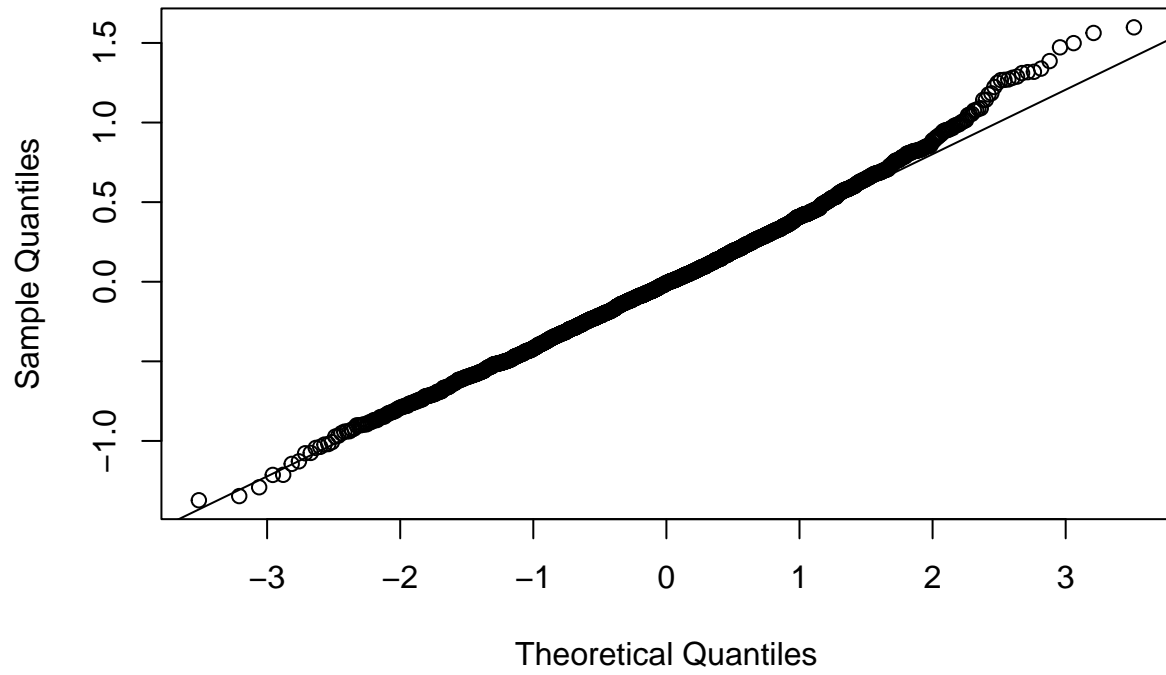
```



#### Residuals analysis

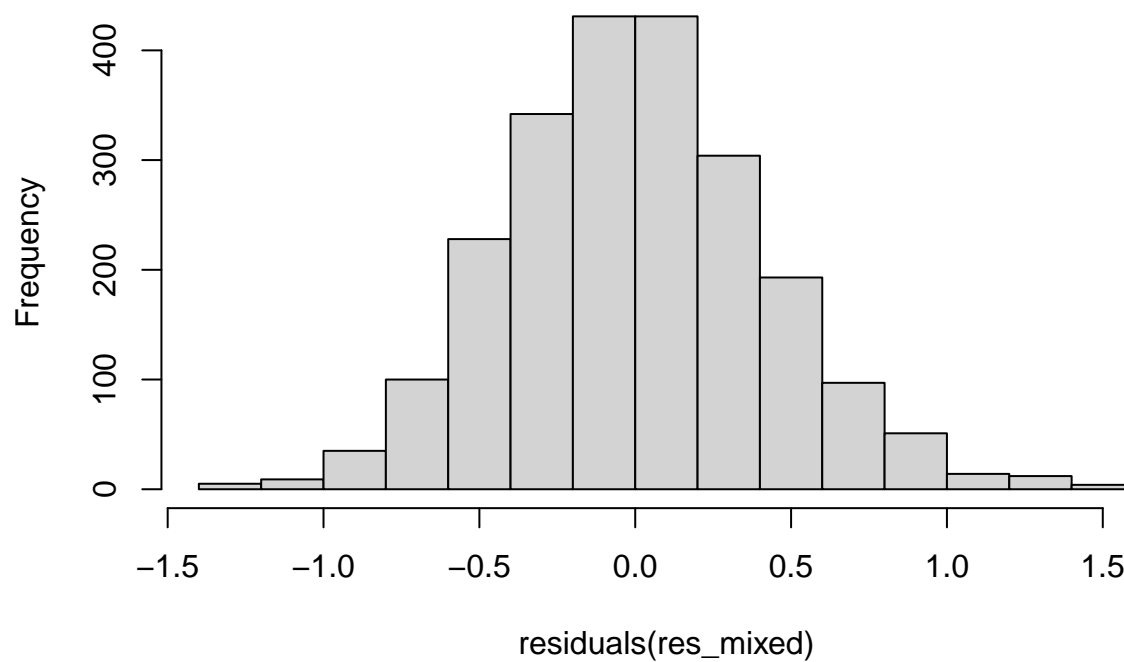
```
qqnorm(residuals(res_mixed))  
qqline(residuals(res_mixed))
```

Normal Q-Q Plot



```
hist(residuals(res_mixed))
```

**Histogram of residuals(res\_mixed)**



```
ks.test(x=residuals(res_mixed), y='pnorm')
```

```
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: residuals(res_mixed)  
## D = 0.20997, p-value < 2.2e-16  
## alternative hypothesis: two-sided
```