Impacts of Working with Different Age Groups on Early Childhood Teacher Earnings

Cassie Noble

Spring 2019

Introduction

Compensation is widely understood to be a major issue in the field of early childhood education (Workman, 2018). Previous research and initiatives on compensation parity in the field of early childhood have primarily focused on compensating preschool teachers more closely to their elementary school counterparts (Barnett & Kasmin, 2017). Recent interest in infant and toddler initiatives have spurred conversations around differences in compensation for preschool teachers vs. those who work with younger age groups.

The research question addressed in this analysis focuses on the impact of the age group an early childhood teacher works with (infants, toddlers, and/or preschoolers) on their hourly wage. Does working with infants vs. toddlers vs. preschoolers impact hourly wage? Does working with multiple groups impact wage in some way other than just the additive effect of each of those groups alone? I plan to test for any interaction effects between the three age groups tested.

Study Design

These data were collected from the Montana Early Childhood Project Practitioner Registry, a database of workforce data on individuals who work in the field of early care and education in the state of Montana (Early Childhood Project, 1999). The data collected for this analysis were focused to center-based early childhood teachers, who were current on the Practitioner Registry as of April 2019.

The response variable of interest is hourly wage (U.S. dollars). See Figure 1 for a plot of the distribution of teacher hourly wages (Phillips, 2017). As wages in this field tend to be low, we run into a lower bound of minimum wage ($8.30 in Montana at the time of data collection), which causes right-skewness in the wage data. A log-transformation was conducted on the wage variable, a distribution which can also be seen in the same plot.

The predictor variables of interest are the age groups a teacher works with. Age groups here include infants, toddlers, and preschoolers. One teacher may work with multiple age groups, so each teacher is assigned either true or false for each age group depending on whether they work with them or not. See Table 1 for counts of each age group combination. Note that this is not a balanced design as there are not equal numbers of observations for each group.

This is an observational study, as teachers were not assigned age groups. Information was also collected on the program the teacher works at, and the county they are employed in. County of employment was included as a group effect, to control for the fact that multiple teachers may work in one county. Teacher wages within the same county may be more similar than teachers from different counties. Teachers from the same program are also likely more similar, but there were not enough teachers per program, or programs per county, to include this effect.

Other variables studied include: length of employment in current position, age, average number of hours worked per week, and Career Path level. All participants of the Practitioner Registry are assigned a level on the Career Path based on education and experience. The Career Path spans from entry level “Membership” and includes those without a high school diploma, to “Level 10” which requires a doctorate in early childhood. Career Path level was treated as a factor in this analysis. These variables were all controlled for in the final model. Figure 2 shows a diagram of the measurement structure of this study, including teachers nested within counties, and all fixed effects in the model (Gordon, 2019).

All licensed providers are required to join the Practitioner Registry by July 2019, with a staggered implementation based on license expiration dates. These data on teachers represent more than one third of all licensed center-based teachers in Montana. The original data set can be assumed to be mostly representative of licensed center-based providers in Montana, as program license dates should not follow any sort of pattern. There is however an issue of missing wage data in the data set. Employment information collected is self-reported by the individual and subsequently verified by their employer. All data fields collected are required, except for wage which is optional to provide. See Table 2 for counts of if wage was reported or not (Xie, 2018). Figures 3-5 show proportions of wage reported by Career Path level, age, and length of employment.

Overall, 87.9% of subjects from the original data set provided wages. To deal with this issue of missing responses, an analysis was conducted to try to find any systematic predictors of reporting or not reporting wage. I fit a binomial generalized linear model to predict whether a teacher would report wage by their age, length of employment, or a simplified variable based on Career Path level (high or low based on whether college credit is required or not, respectively).

There was little to no evidence against the null hypothesis that age or Career Path level has no impact on reporting wage (p=0.978 for age, p=0.452 for Career Path level). There was strong evidence against the null that length of employment has no impact on reporting wage (z=-3.622, p=0.0002). From the plot in Figure 5, and the effects plot in Figure 6 (Fox & Weisberg, 2018), we can see that we may have an issue with missing responses from the teachers that have been employed in their current position for longer lengths of time. We proceed forward with analyses, keeping in mind that we may not be able to infer results to subjects employed for longer lengths of time.

The original data set contained 647 participants in 26 counties. After removing missing wages, we are left with 569 subjects in the same 26 counties.

Statistical Procedure Used

Summary statistics for wage by age group combinations are provided in Table 3 (Pruim, Kaplan, & Horton, 2017). Distributions of wages for each age group combination are shown in Figure 7. Figure 8 shows the same plots but for log-transformed wage distributions. The final model uses the log-transformed wage as the response, as that model better met model assumptions.

All analyses were conducted using the R programming language (R Core Team, 2019). A mixed model was fit to account for the hierarchical structure of this study design (Bates et al., 2015), and a type II ANOVA to compare the categorical predictors and interactions.

The Residuals vs. Fitted plot for the final model is provided in Figure 9. This plot shows some slight issues with non-constant variance but nothing too extreme. Figure 10 shows plots for assessing the normality of model residuals on the left, and normality of the county random effect on the right (Fox & Weisberg, 2011). The distribution for county meets the normality assumption. The model residuals show some slight right-skew. Figures 11 and 12 show the same diagnostic plots for the wage on the original scale. From these plots we can see that the log-transformation resulted in more constant variance, and less right-skew in the model residuals and county random effect.

Summary of Statistical Findings

The initial plan for this study was to do a step-down approach, by starting with the most complicated model possible (3-way interaction of infants by toddlers by preschoolers). However, there were not enough observations with reported wages in the infants by preschoolers group to include this interaction. This makes sense in context as it would be an unusual combination to work with infants and preschoolers but not toddlers.

In this case the most complicated model is one which includes interactions of infants by toddlers, and toddlers by preschoolers. This model was fit, and infants by toddlers had the largest p-value (t(5)=0.673, p=0.501), so this term was dropped, and a new model fit. In the new model, infants had the largest p-value (t(5)=-0.188, p=0.851) and so this term was dropped. At this point the toddlers by preschoolers interaction was the only remaining variable of interest and its p-value was sufficiently small to retain in the model ($F(1{,}545.12) = 6.347$, p-value $= 0.012$). The theoretical model for the final model is provided in Appendix A.

There is strong evidence against the null hypothesis of no interaction between working with toddlers and working with preschoolers on log-wages ($F(1{,}545.12) = 6.347$, p-value $= 0.012$), in a model that controls for age, length of employment in current position, Career Path level, average number of hours worked per week, and county to county variance. This suggests a need to retain this interaction in the model. Figure 13 shows the effects plot for all model components.

Figure 14 shows the toddler by preschooler interaction effect overlaid in one plot. This plot shows that there is very little difference in log wage for those who don’t work with preschoolers, regardless of whether or not they work with toddlers. The highest wages were for those who work with preschoolers but not toddlers, and the lowest wages were among those who work with both preschoolers and toddlers. A table of all results from the type II ANOVA test is also provided in Table 4.

The estimated correlation of wages between two individuals in the same county is 25.6%, after accounting for the fixed effects in the model (toddler by preschooler interaction, length of employment, age, Career Path level, and hours worked per week), indicating a moderate correlation. The fixed effects in this model account for 23.5% of the variance in wages, marginal $R^2 = 0.235$ (Barton, 2018). The random effect of county plus the fixed effects account for 43.1% of the variance in wages, conditional $R^2 = 0.431$.

Scope of Inference

This study did not involve random assignment of teacher to age group of children worked with. We can discuss relationships between the age group worked with and their hourly wage, but we cannot infer causation. The original data set was fairly representative of the population of center-based early childhood teachers in Montana. However, with missing responses, we must be careful with making inferences to larger populations. With systematic missing data from individuals employed for longer lengths of time, we are unable to make inferences to this group.

This analysis leads us to the conclusion that whether a teacher works with preschoolers or not impacts log-wages differently depending on whether a teacher works with toddlers or not, while controlling for individual fixed effects and county to county variation.

References

Barnett, W. S., & Kasmin, R. (2017). Teacher compensation parity policies and state-funded pre-k programs. Center for the Study of Child Care Employment, University of California, Berkeley.

Barton, K. (2018). MuMIn: Multi-Model Inference. R package version 1.42.1. URL: https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Early Childhood Project. (2019). Practitioner Registry [database]. Retrieved April, 2019, from http://mtecp.org/

Fox, J., & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Fox, J., & Weisberg, S. (2018). Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. Journal of Statistical Software, 87(9), 1-27.

Gordon, M. (2019). Gmisc: Descriptive Statistics, Transition Plots, and More. R package version 1.8. https://CRAN.R-project.org/package=Gmisc

Kuznetsova A., Brockhoff P. B., & Christensen, R. H. B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), 1-26. doi: 10.18637/jss.v082.i13. URL: http://doi.org/10.18637/jss.v082.i13

Phillips, N. (2017). yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". R package version 0.1.5. https://CRAN.R-project.org/package=yarrr

Pruim, R., Kaplan, D. T., & Horton, N. J. (2017). The mosaic Package: Helping Students to 'Think with Data' Using R. The R Journal, 9(1):77-102.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey, F., & Schafer, D. (2012). The statistical sleuth: a course in methods of data analysis. Cengage Learning.

Wickham, H., Hester, J., & Francois, R. (2018). readr: Read Rectangular Text Data. R package version 1.3.1. URL: https://CRAN.R-project.org/package=readr

Workman, S. (2018). When preschool teachers can't afford care for their own children. The Hechinger Report. URL: https://hechingerreport.org/opinion-when-preschool-teachers-cant-afford-care-for-their-own-children/

Xie, Y. (2018). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.21.

Appendix A: Theoretical Model

$$\mu\{LoggedHourlyWage_{ij}|Toddlers_{ij} * Preschoolers_{ij} + Age_{ij} +$$

$$EmployLengthYear_{ij} + Level_{ij} + HoursPerWeek_{ij}\} =$$

$$\beta_0 + \beta_1 * I_{Toddlers=TRUE_{ij}} + \beta_2 * I_{Preschoolers=TRUE_{ij}} + \beta_3 * Age +$$

$$\beta_4 * EmployLengthYear_{ij} + \beta_5 * I_{Level_{ij}} + \beta_6 * I_{HoursPerWeek_{ij}} +$$

$$\beta_7 * I_{Toddlers=TRUE_{ij}} * I_{Preschoolers=TRUE_{ij}} + County_i + \epsilon_{ij}$$

with $County_i \sim N(0, \sigma^2_{County})$ and $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_{ij}})$

where $I_{Toddlers=TRUE_{ij}} = \begin{cases} 0 & False \\ 1 & True \end{cases}$

and $I_{Preschoolers=TRUE_{ij}} = \begin{cases} 0 & False \\ 1 & True \end{cases}$

Appendix B: Figures and Tables

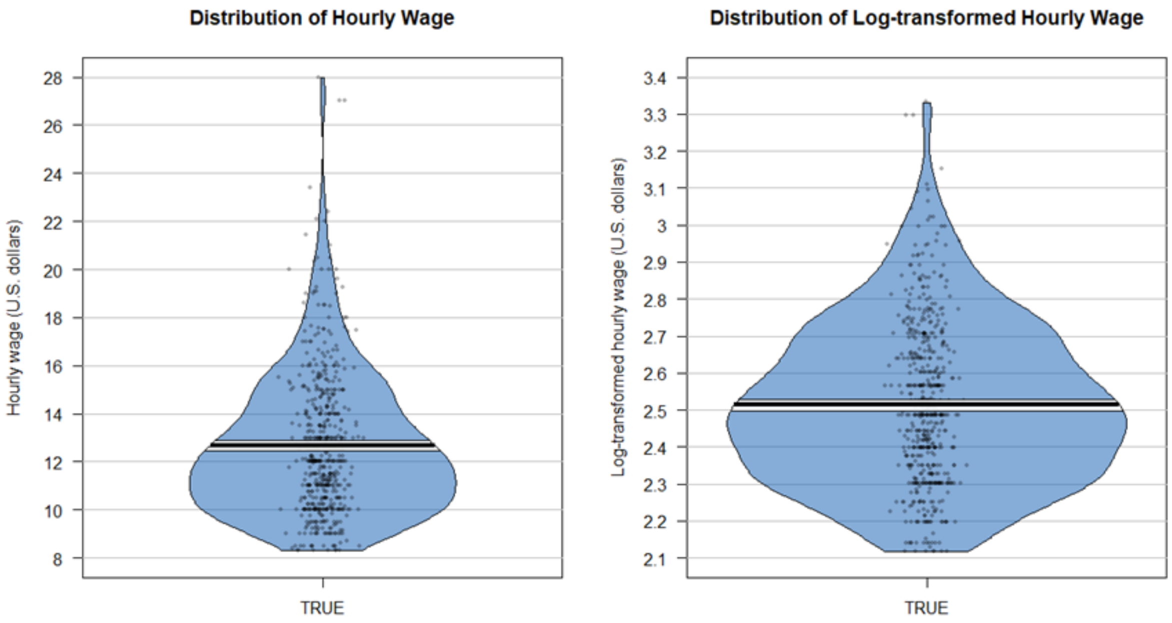


Figure 1. Distribution of early childhood teacher hourly wages. Untransformed wages are on the left with heavy right skew. Log-transformed wages are on the right with slight right skew.

		Preschoolers	
		FALSE	TRUE
Infants	Toddlers		
FALSE	FALSE	9	194
	TRUE	95	69
TRUE	FALSE	26	1
	TRUE	85	90

Table 1. Age groups of children worked with, for early childhood teachers that reported their hourly wage.

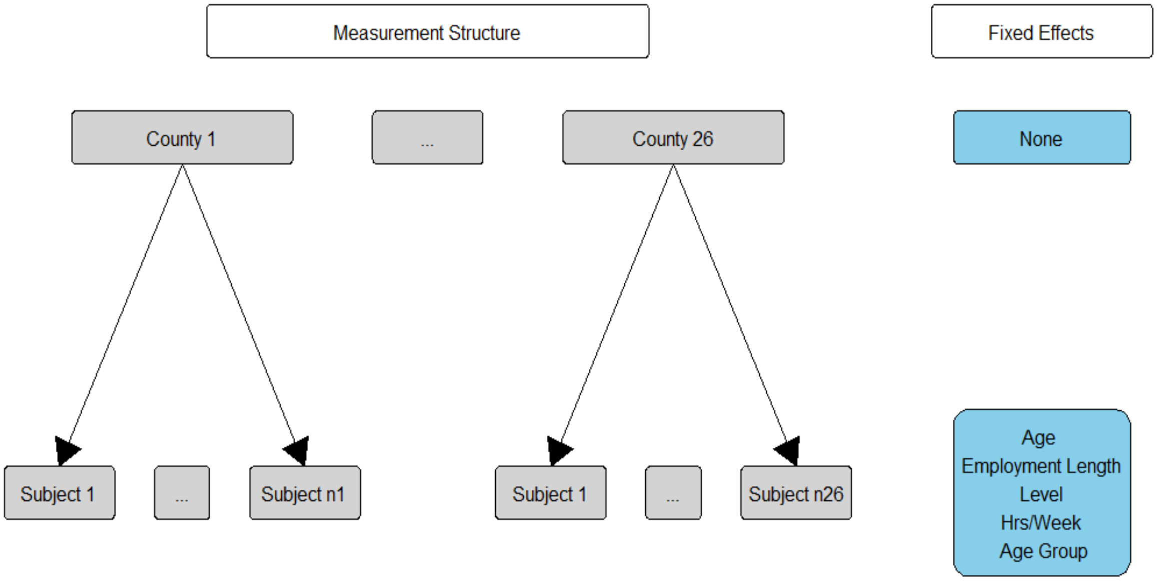


Figure 2. Diagram of study design with teachers nested within counties and fixed effects on teacher level.

Wage Reported	Count
FALSE	78
TRUE	569

Table 2. Counts of numbers of teachers that reported hourly wage vs. did not report.

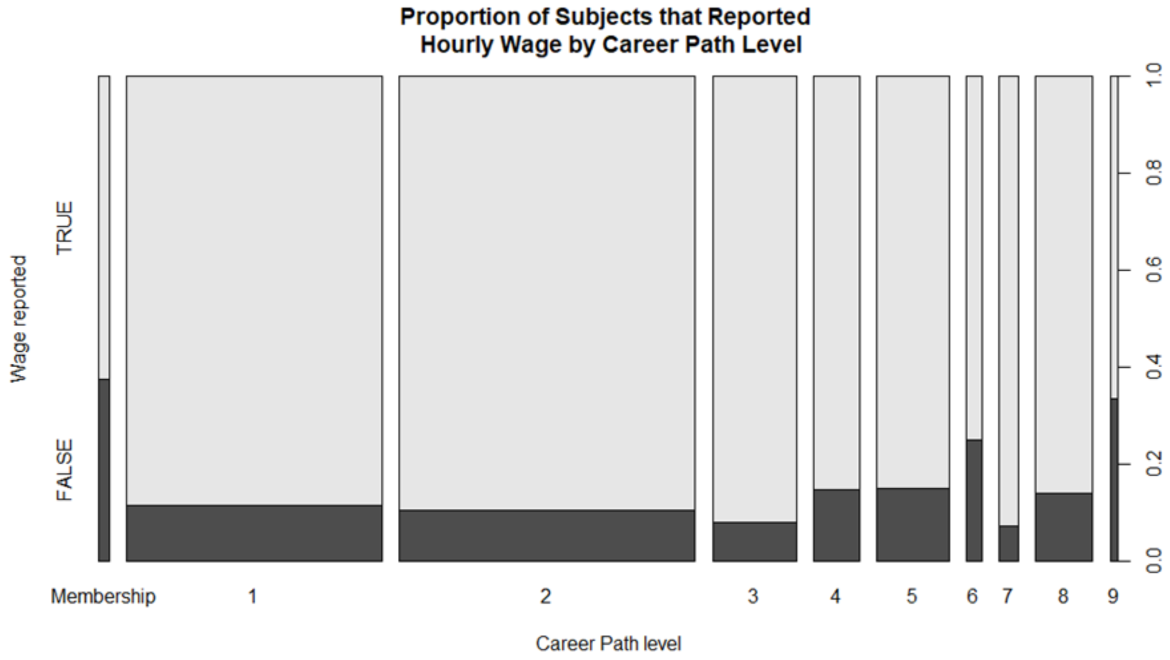


Figure 3. Proportion of teachers that reported hourly wage by their Career Path level.

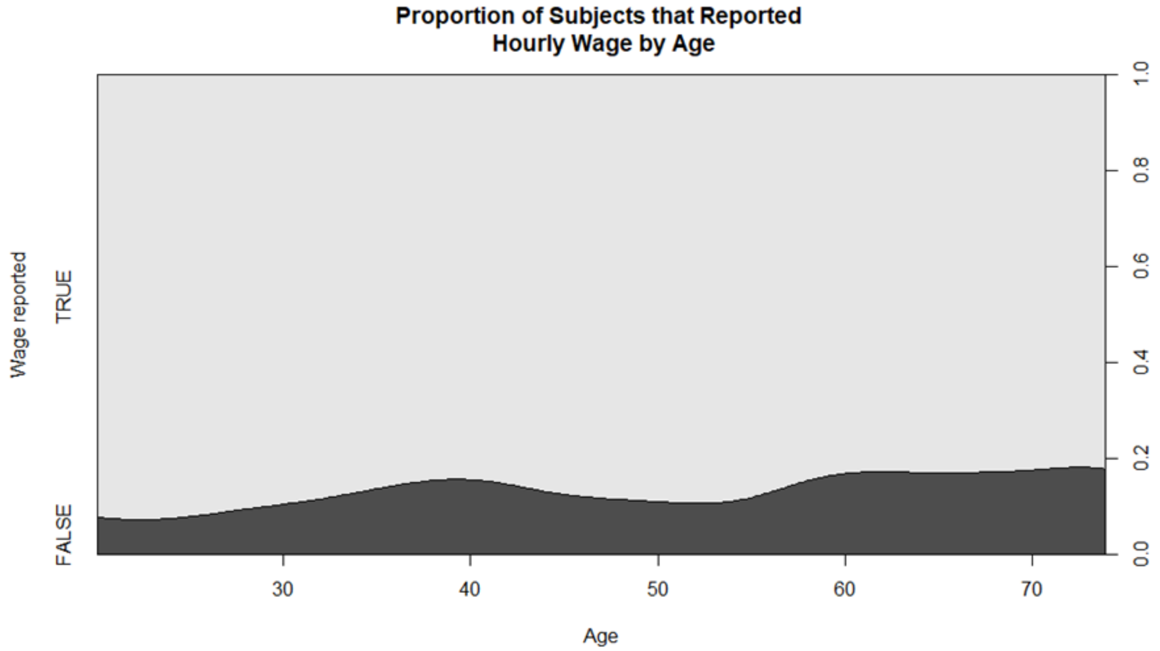


Figure 4. Proportion of teachers that reported hourly wage across age.

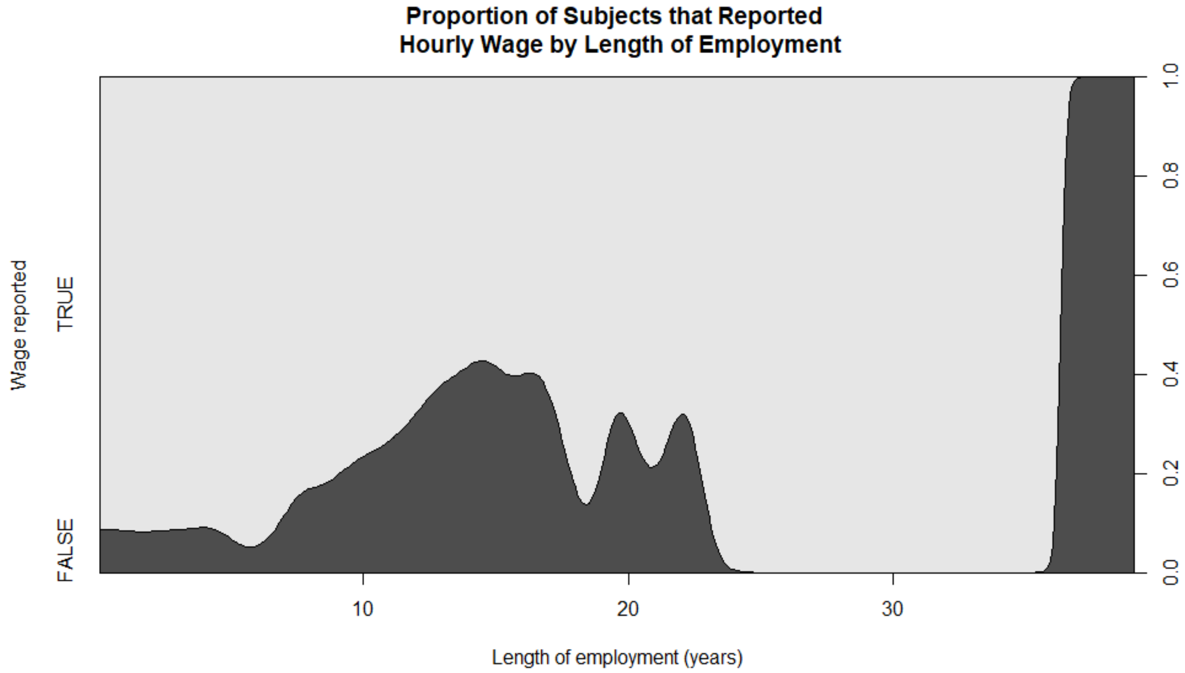


Figure 5. Proportion of teachers that reported hourly wage across length of employment in their current position.

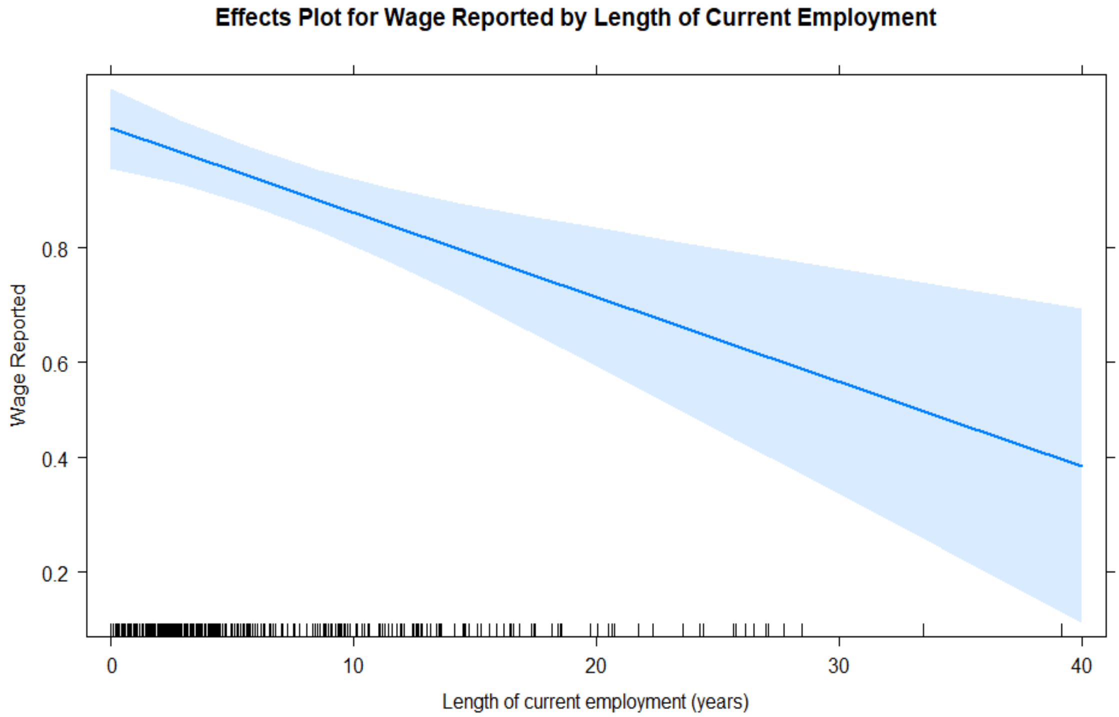


Figure 6. Effects plots for wage reported for length of employment.

Infant|Toddler|Preschooler	Min	Q1	median	Q3	max	mean	sd	n	missing
FALSE | FALSE | FALSE	10.0	11.11	12.00	15.00	20.00	13.36	3.23	9	2
TRUE | FALSE | FALSE	8.50	10.03	11.00	13.00	27.00	12.05	3.50	26	6
FALSE | TRUE | FALSE	8.30	10.50	11.50	13.26	22.00	12.08	2.32	95	12
TRUE | TRUE | FALSE	8.30	10.50	12.00	14.00	20.34	12.39	2.43	85	3
FALSE | FALSE| TRUE	8.30	11.48	13.48	15.79	28.00	13.87	3.23	194	25
TRUE | FALSE | TRUE	10.75	10.75	10.75	10.75	10.75	10.75	NA	1	1
FALSE | TRUE | TRUE	8.30	10.40	12.00	13.00	22.40	12.16	2.57	69	17
TRUE | TRUE | TRUE	8.30	9.51	10.80	12.42	27.00	11.51	2.95	90	12

Table 3. Summary statistics for hourly wage by each age group combination.

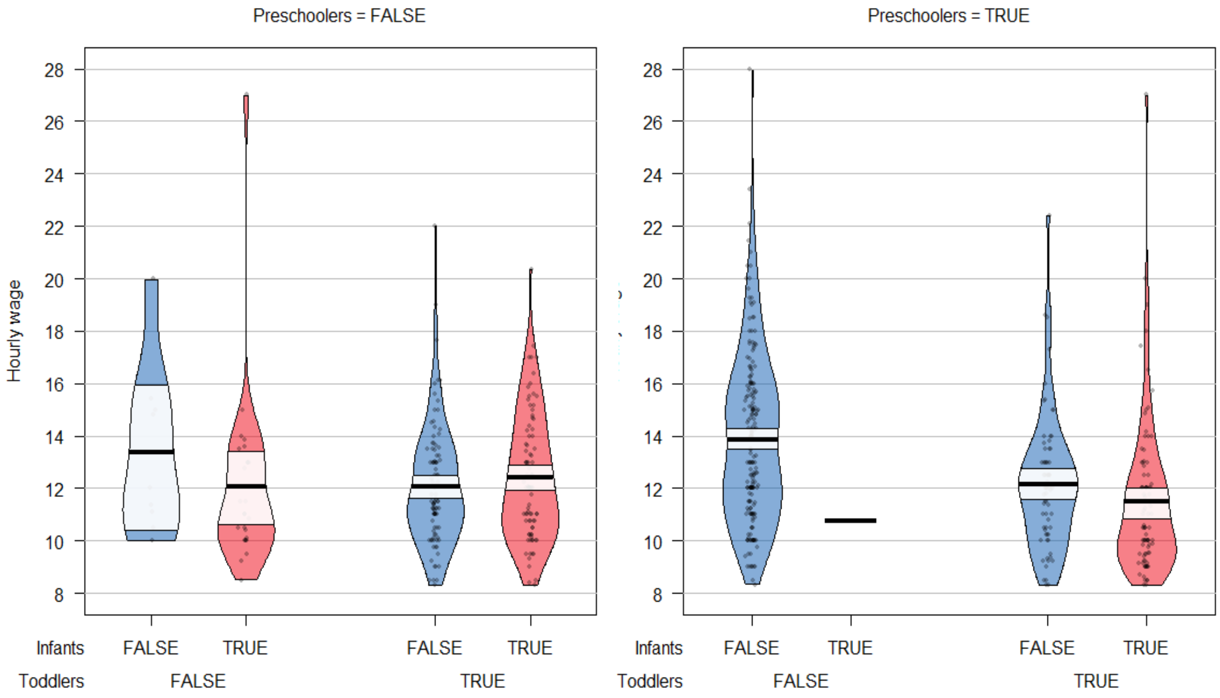


Figure 7. Untransformed hourly wage by age group worked with combinations.

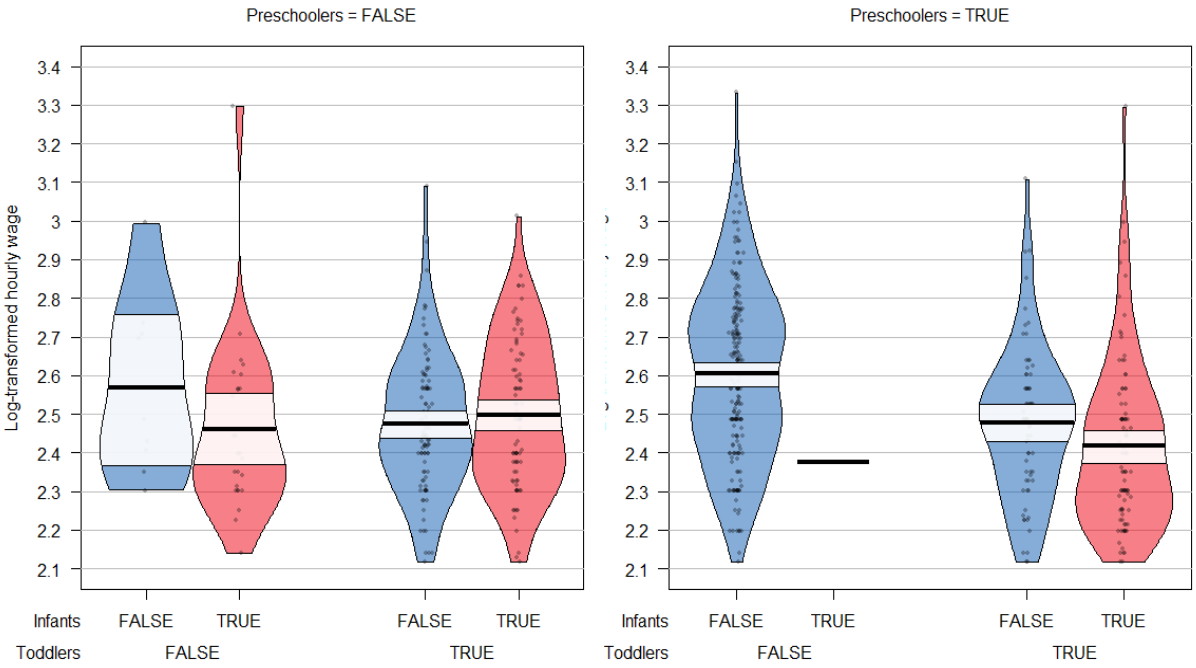


Figure 8. Log-transformed hourly wage by age group worked with combinations.

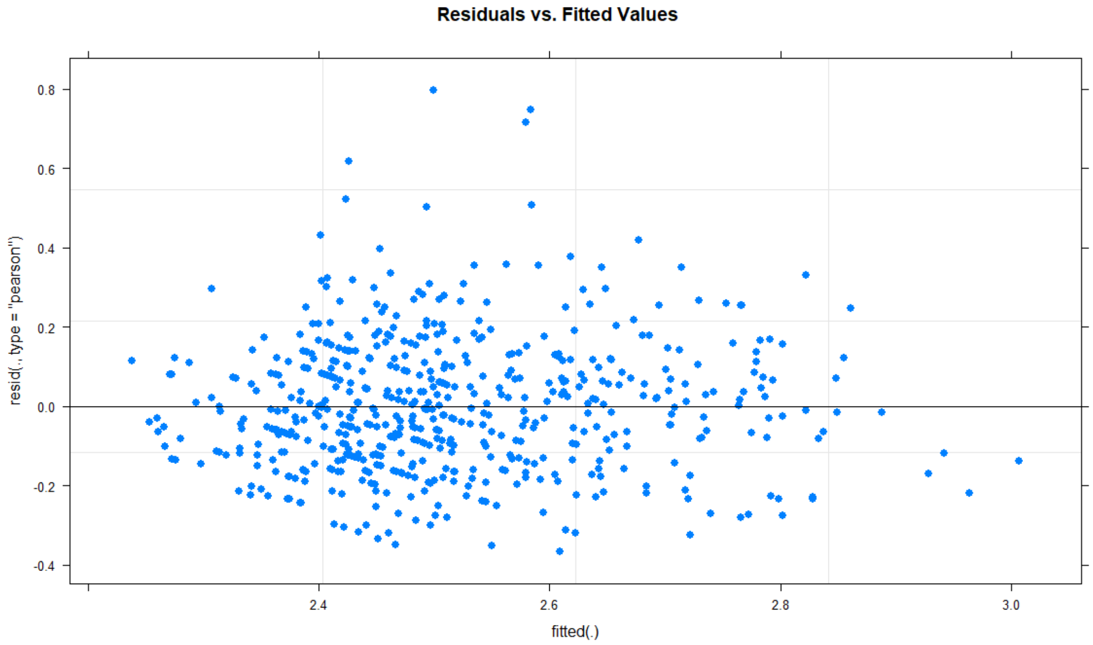


Figure 9. Plot of residuals vs. fitted values for final model.

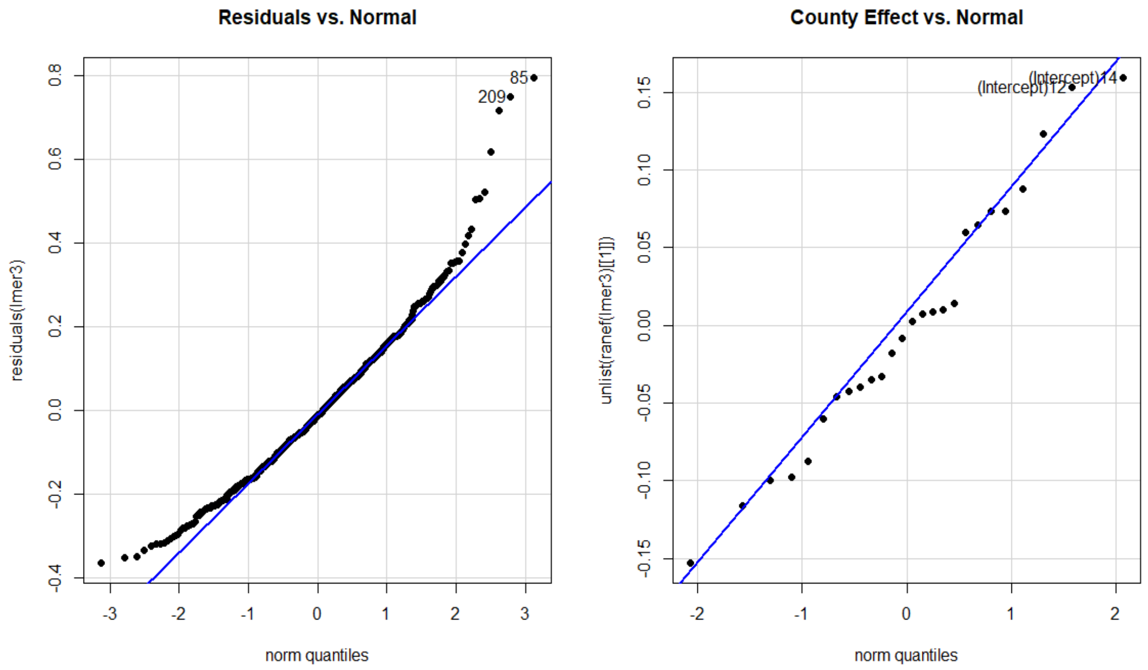


Figure 10. Plot of residuals vs. normal for final model.

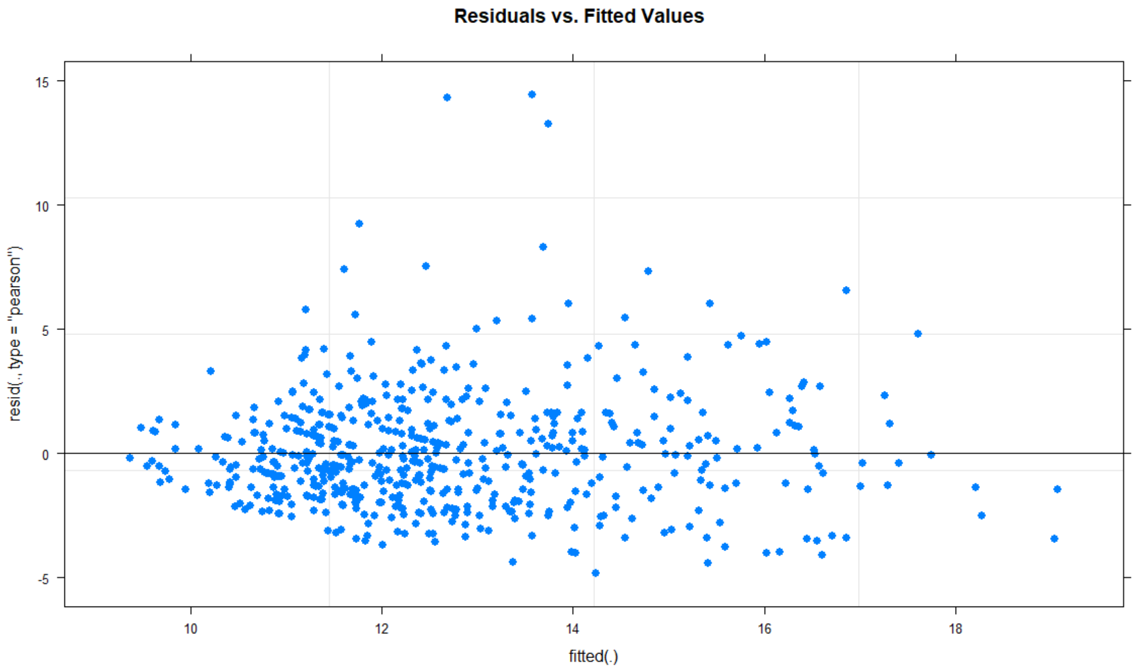


Figure 11. Plot of residuals vs. fitted values for untransformed wage model.

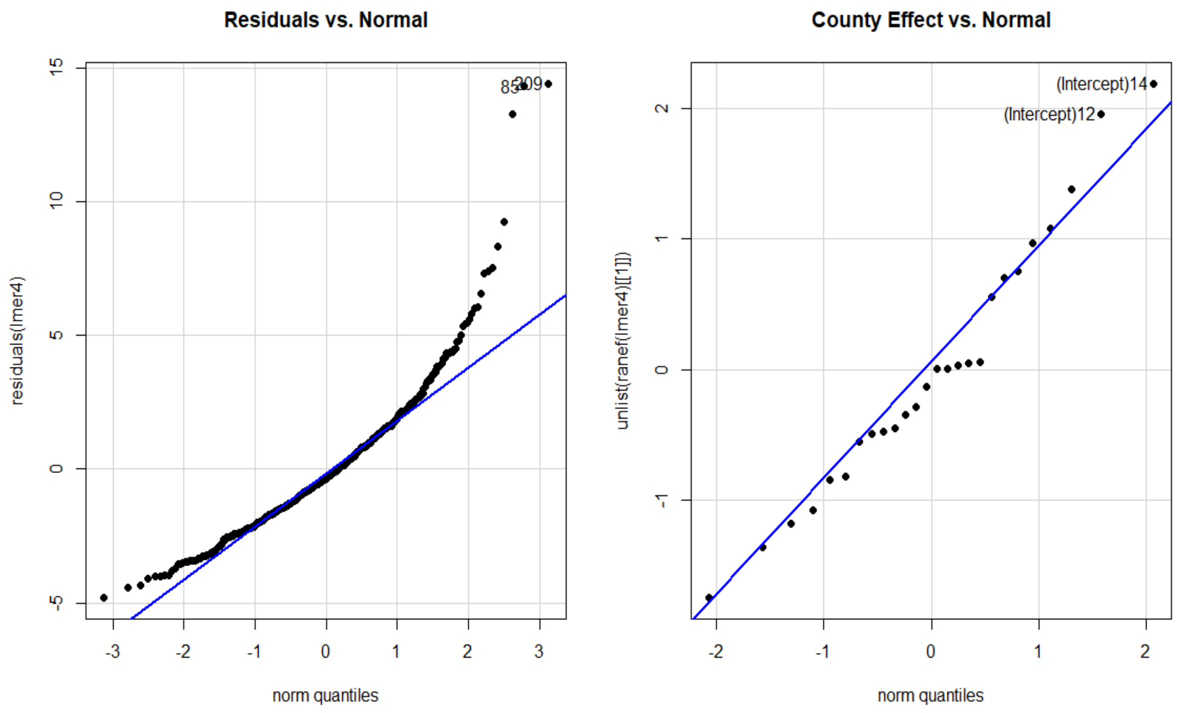


Figure 12. Plot of residuals vs. normal for untransformed wage model.

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Toddlers	0.580	0.580	1	542.975	19.272	0.000
Preschoolers	0.026	0.026	1	544.565	0.872	0.351
Age	0.022	0.022	1	538.981	0.737	0.391
EmployLengthYear	1.205	1.205	1	536.000	40.058	0.000
Level	2.647	0.294	9	537.478	9.774	0.000
HoursPerWeek	0.037	0.037	1	552.264	1.221	0.270
Toddlers:Preschoolers	0.191	0.191	1	545.120	6.347	0.0120

Table 4. Results from type II ANOVA for final model of log wage, including interaction of toddlers by preschooler while controlling for age, length of employment in current position, Career Path level, average number of hours worked per week, and county to county variance.

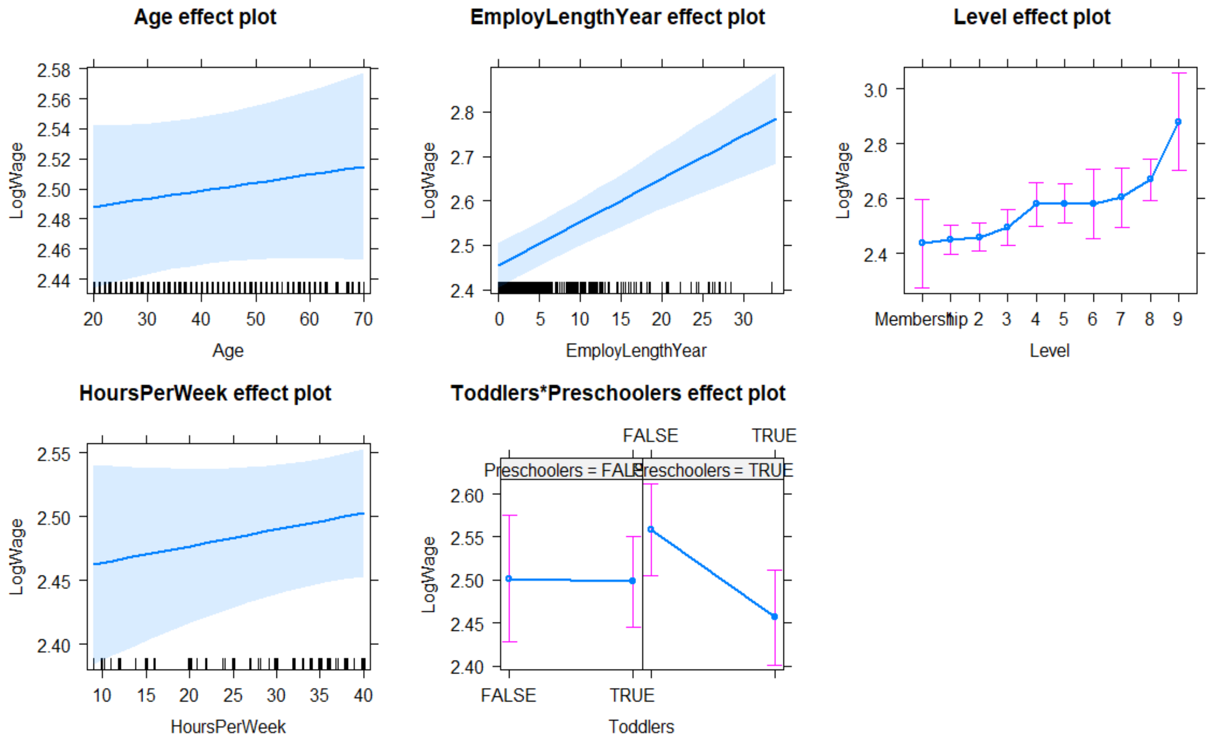


Figure 13. Effects plot for that includes all variables in final model.

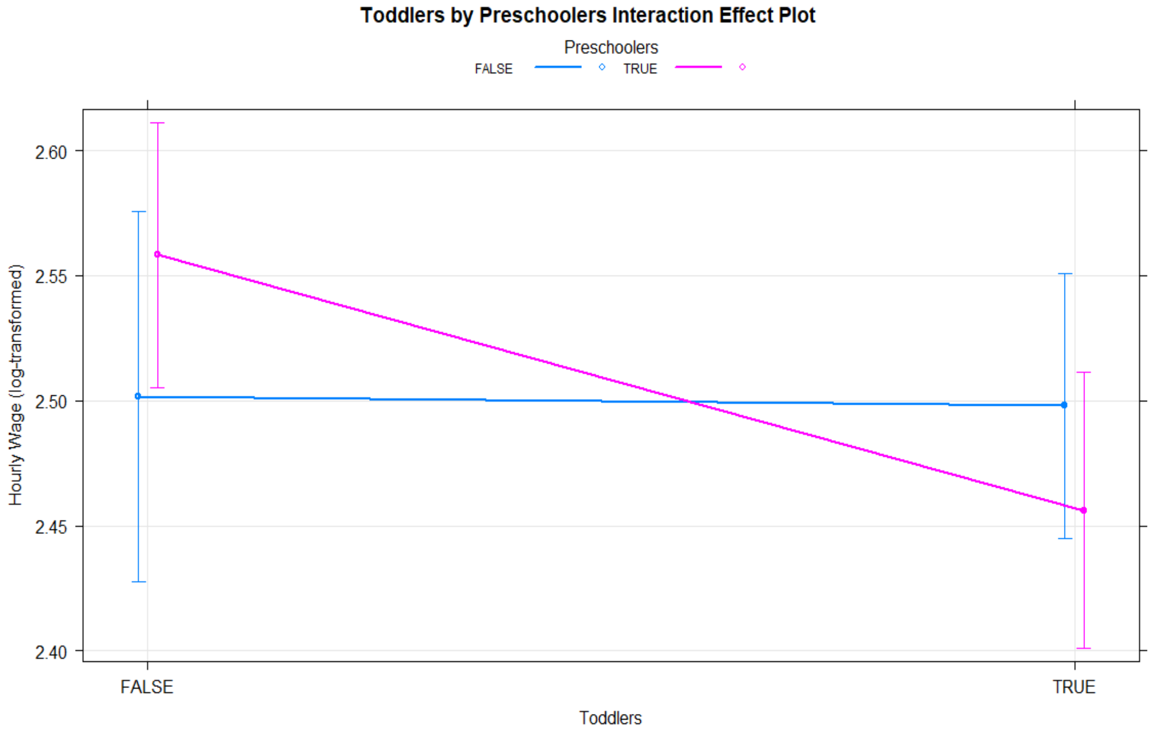
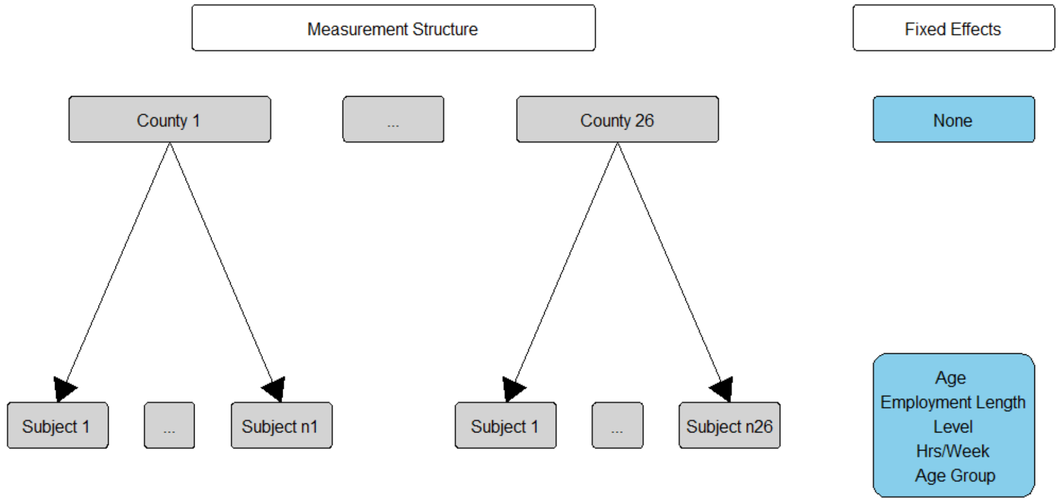


Figure 14. Effects plot for the toddlers by preschoolers interaction effect in the final model that controls for age, length of employment in current position, Career Path level, average number of hours worked per week, and county to county variance.

Appendix C: R code

```r
### study design diagram
library(grid)
library(Gmisc)
grid.newpage()

# set some parameters to use repeatedly
treeleft <- .2
treemid <-  .4
treeright <- .6
fixedcol <- 0.9
width <- .18
toplev <- 0.8
bottomlev <- 0.3
gp <- gpar(fill = "lightgrey")
gpf <- gpar(fill = "skyblue")

# label areas
(mainlabel <- boxGrob("Measurement Structure",
 x=treemid, y=0.95, box_gp = gpar(fill="white"), width = 2*width))

(fixedlabel <- boxGrob("Fixed Effects",
 x=fixedcol, y=0.95, box_gp = gpar(fill="white"), width = width))

# create boxes
(sub1 <- boxGrob("County 1",
 x=treeleft, y=toplev, box_gp = gp, width = width))

(sub_ellipses <- boxGrob("...",
 x=treemid, y=toplev, box_gp = gp, width = width/2))

(subI <- boxGrob("County 26",
 x=treeright, y=toplev, box_gp = gp, width = width))

(fixedtop <- boxGrob("None",
                    x=fixedcol, y=toplev, box_gp = gpf, width = width/1.25))
# observation level
width_obs <- width/2
(sub11 <- boxGrob("Subject 1",
 x=treeleft-width_obs-0.01, y=bottomlev, box_gp = gp, width = width_obs))

(sub1_ellipses <- boxGrob("...",
 x=treeleft, y=bottomlev, box_gp = gp, width = width_obs/2))

(sub1n_1 <- boxGrob("Subject n1",
 x=treeleft+width_obs+0.01, y=bottomlev, box_gp = gp, width = width_obs))

(subI1 <- boxGrob("Subject 1",
 x=treeright-width_obs-0.01, y=bottomlev, box_gp = gp, width = width_obs))
```

```
(subI_ellipses <- boxGrob("...",
 x=treeright, y=bottomlev, box_gp = gp, width = width_obs/2))

(subIn_I <- boxGrob("Subject n26",
 x=treeright+width_obs+0.01, y=bottomlev, box_gp = gp, width = width_obs))


(fixedbottom <- boxGrob("Age \n Employment Length \n Level \n Hrs/Week \n Age
Group",
                     x=fixedcol, y=bottomlev, box_gp = gpf, width = width/1.2
5))

# connect boxes
connectGrob(sub1, sub11, "v")
connectGrob(sub1, sub1n_1, "v")

connectGrob(subI, subI1, "v")
connectGrob(subI, subIn_I, "v")
```



```
### read in data and basic formatting
library(readr)

# read in data
workforce <- read_csv("TeacherWageAgeGroups.csv")

# refactor
workforce$County <- as.factor(workforce$County)
workforce$Infants <- as.factor(workforce$Infants)
workforce$Toddlers <- as.factor(workforce$Toddlers)
workforce$Preschoolers <- as.factor(workforce$Preschoolers)
```

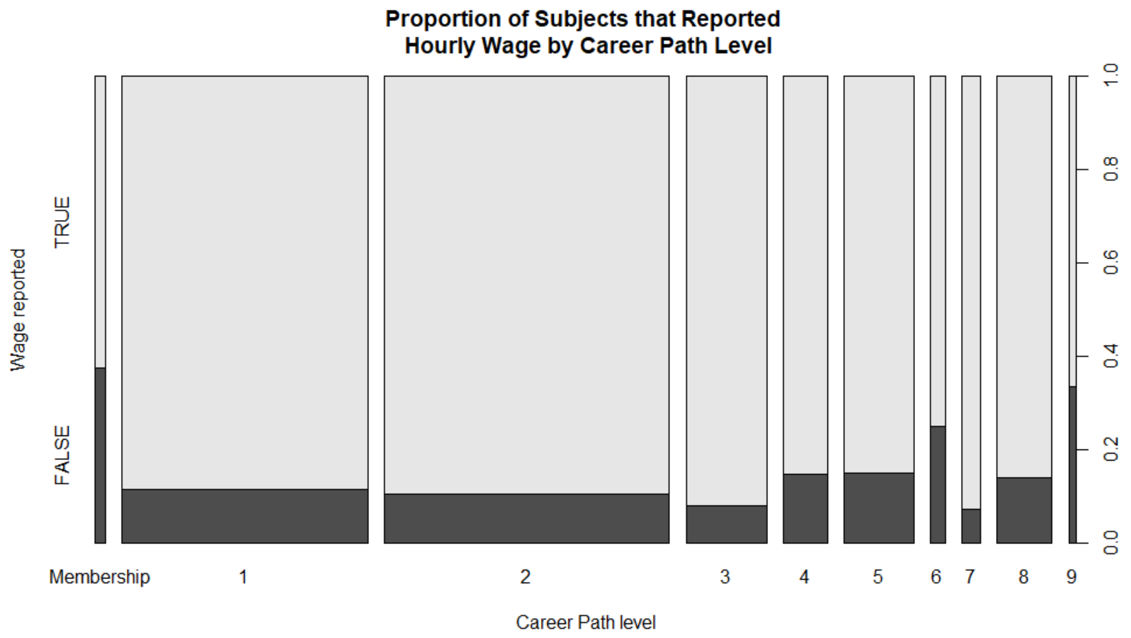
```
workforce$OrgIdF <- as.factor(workforce$OrgId)
workforce$Level <- as.factor(workforce$Level)
workforce$Level <- relevel(workforce$Level, "Membership")
levels(workforce$Level) <- c("Membership", "1", "2", "3", "4", "5",
                             "6", "7", "8", "9")

# split PR levels into high vs. low depending on if college credit required
workforce$LevelSimp <- "low"
for(i in 1:nrow(workforce)){
  if(workforce$Level[i] == 4 | workforce$Level[i] == 5 | workforce$Level[i] =
= 6 |
     workforce$Level[i] == 7 | workforce$Level[i] == 8 | workforce$Level[i] =
= 9){
    workforce$LevelSimp[i] <- "high"
  }
}

workforce$LevelSimp <- as.factor(workforce$LevelSimp)
workforce$LevelSimp <- relevel(workforce$LevelSimp, "low")
levels(workforce$LevelSimp)

## [1] "low"  "high"

### deal with missing wage data

# create variable for missing wage
workforce$WageReported <- TRUE

for(i in 1:nrow(workforce)){
  if(workforce$HourlyWage[i] == 0){
    workforce$WageReported[i] <- FALSE
  }
}

workforce$WageReported <- as.factor(workforce$WageReported)
levels(workforce$WageReported)

## [1] "FALSE" "TRUE"

# convert 0s to NAs for missing wage data
for(i in 1:nrow(workforce)){
  if(workforce$HourlyWage[i] == 0){
    workforce$HourlyWage[i] <- NA
  }
}

# subset to remove missing wages
workforceSub <- subset(workforce, !is.na(HourlyWage))
```

```
### view figures and tables surrounding missing wage data
library(knitr)
library(yarrr)

# table wage reported counts
reportWages <- table(workforce$WageReported)
kable(reportWages, format = "markdown", col.names =
        c("Wage Reported", "Count"))
```

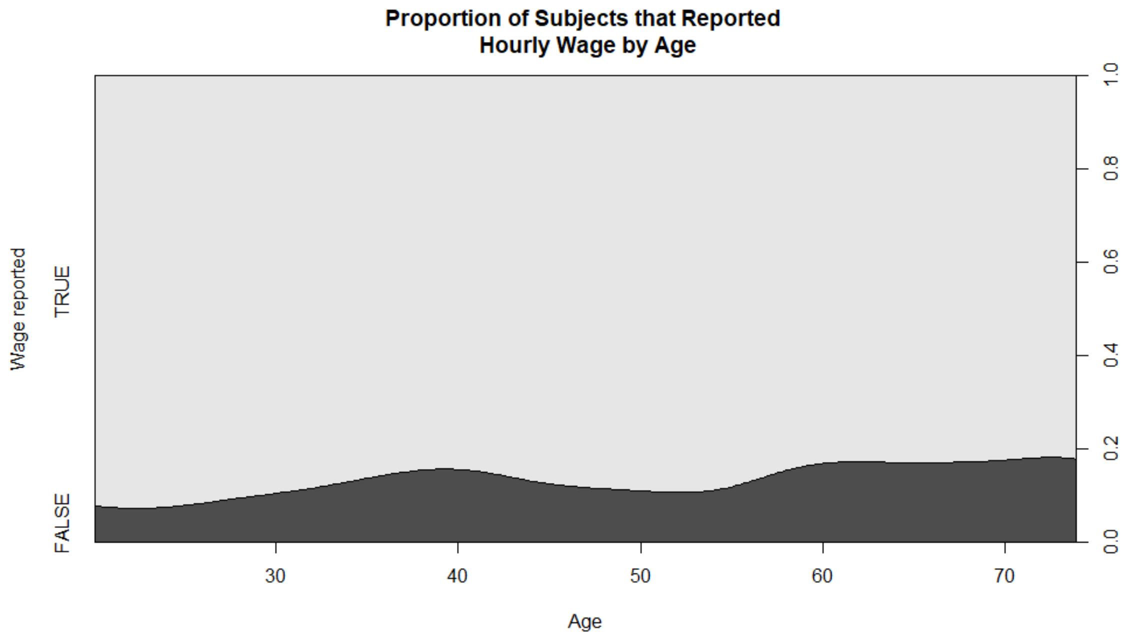
Wage Reported	Count
FALSE	78
TRUE	569

```
# percent reported wage
reportWages[2] / nrow(workforce)

##      TRUE
## 0.8794436

# sample sizes
nrow(workforce)

## [1] 647

unique(workforce$County)

##  [1] Cascade         Lewis and Clark Gallatin        Flathead
##  [5] Missoula        Ravalli         Fergus          Yellowstone
##  [9] Lincoln         Silver Bow      Park            Hill
## [13] Richland        Rosebud         Jefferson       Big Horn
## [17] Lake            Stillwater      Dawson          Deer Lodge
## [21] Blaine          Carbon          Beaverhead      Madison
## [25] Sheridan        Glacier
## 26 Levels: Beaverhead Big Horn Blaine Carbon Cascade Dawson ... Yellowston
e

nrow(workforceSub)

## [1] 569

unique(workforceSub$County)

##  [1] Cascade         Lewis and Clark Flathead        Missoula
##  [5] Ravalli         Fergus          Yellowstone     Gallatin
##  [9] Lincoln         Silver Bow      Hill            Richland
## [13] Rosebud         Big Horn        Lake            Stillwater
## [17] Dawson          Park            Deer Lodge      Jefferson
## [21] Blaine          Carbon          Beaverhead      Madison
## [25] Sheridan        Glacier
## 26 Levels: Beaverhead Big Horn Blaine Carbon Cascade Dawson ... Yellowston
e
```

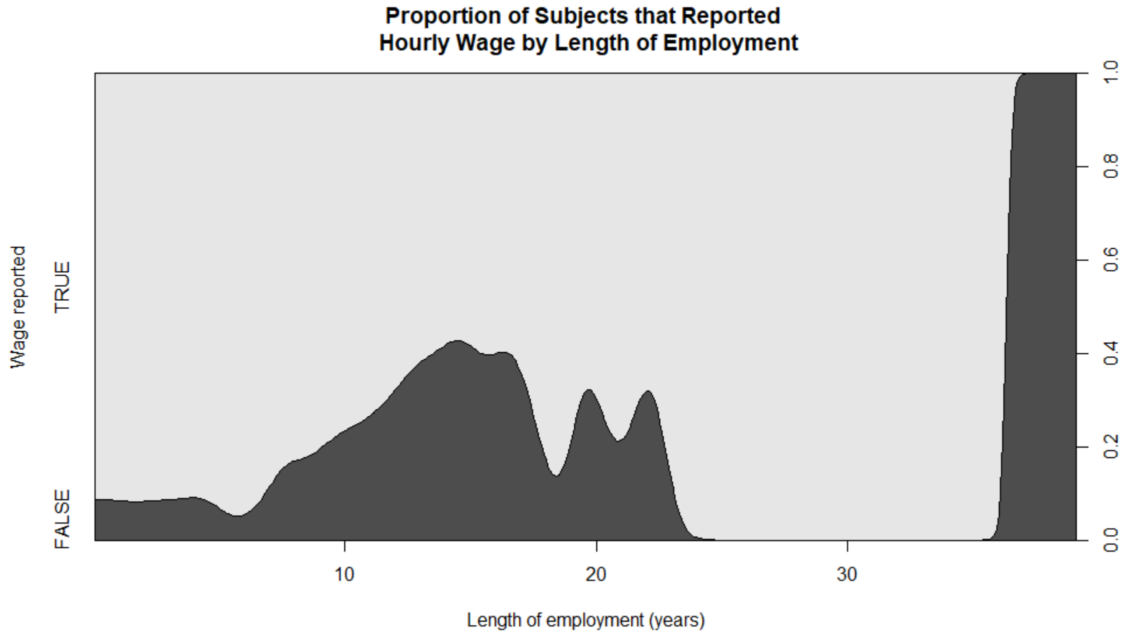
```r
# registry level
plot(WageReported ~ Level, data = workforce,
     xlab = "Career Path level",
     ylab = "Wage reported",
     main = "Proportion of Subjects that Reported \n Hourly Wage by Career Pa
th Level")
```



```r
# table wage reported by level
levelTable <- xtabs(~ WageReported + Level, data = workforce)
ftable(levelTable)

##              Level Membership   1   2   3   4   5   6   7   8   9
## WageReported
## FALSE                       3  22  23   5   5   8   3   1   6   2
## TRUE                        5 169 199  58  29  46   9  13  37   4

# age
cdplot(WageReported ~ Age, data = workforce, ylab = "Wage reported",
       main = "Proportion of Subjects that Reported \n Hourly Wage by Age")
```

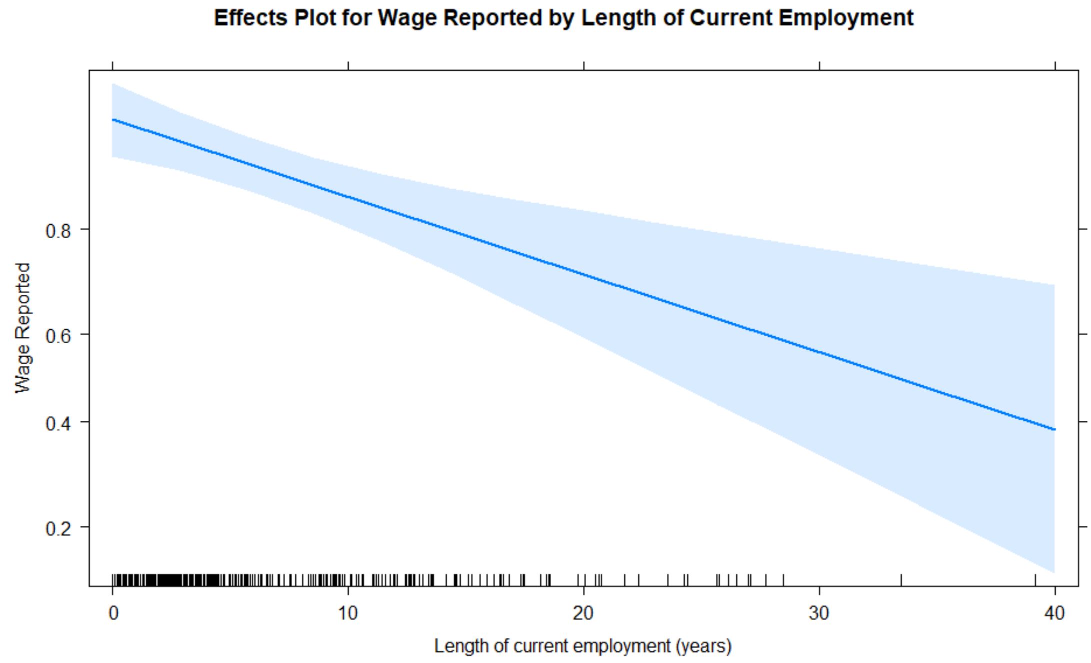


```
# length of employment
cdplot(WageReported ~ EmployLengthYear, data = workforce,
       ylab = "Wage reported",
       xlab = "Length of employment (years)",
       main = "Proportion of Subjects that Reported \n Hourly Wage by Length
of Employment")
```

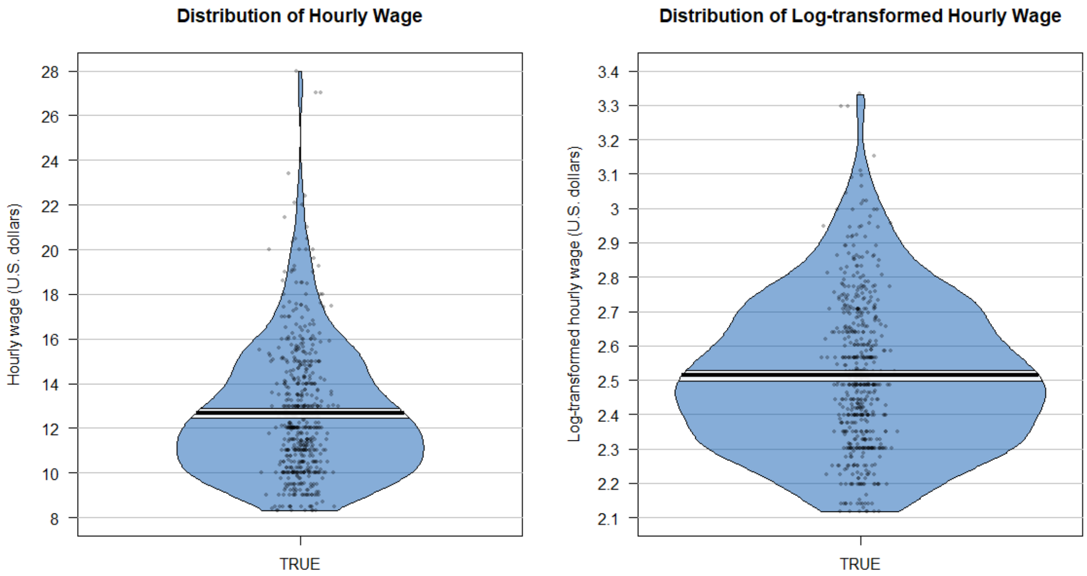


```
### find any systematic predictors for missingness of wage data
```

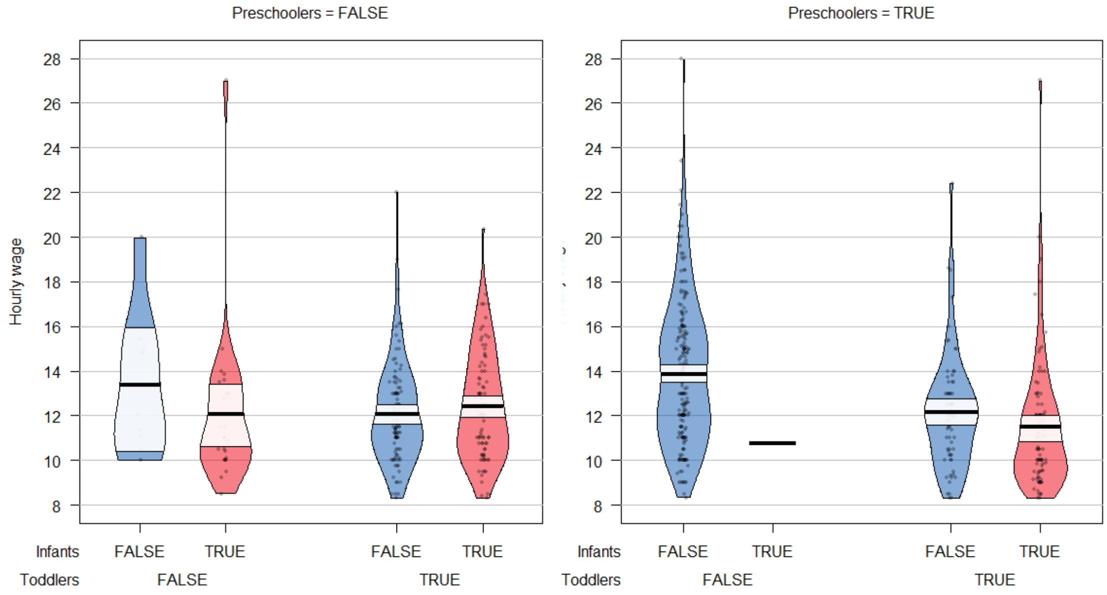
```
# wage reported by variables of possible interest
glm1 <- glm(WageReported ~ Age + EmployLengthYear + LevelSimp,
            data = workforce, family = "binomial")
summary(glm1)

##
## Call:
## glm(formula = WageReported ~ Age + EmployLengthYear + LevelSimp,
##     family = "binomial", data = workforce)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2454   0.4185   0.4496   0.4878   1.2442
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       2.4499260  0.3994366   6.133  8.6e-10 ***
## Age               0.0002836  0.0103046   0.028 0.978046
## EmployLengthYear -0.0721774  0.0199270  -3.622 0.000292 ***
## LevelSimphigh    -0.2055301  0.2736684  -0.751 0.452641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 476.23  on 646  degrees of freedom
## Residual deviance: 458.56  on 643  degrees of freedom
## AIC: 466.56
##
## Number of Fisher Scoring iterations: 4

# effects plot
library(effects)
plot(allEffects(glm1)[2], ylab = "Wage Reported",
     xlab = "Length of current employment (years)",
     main = "Effects Plot for Wage Reported by Length of Current Employment")
```



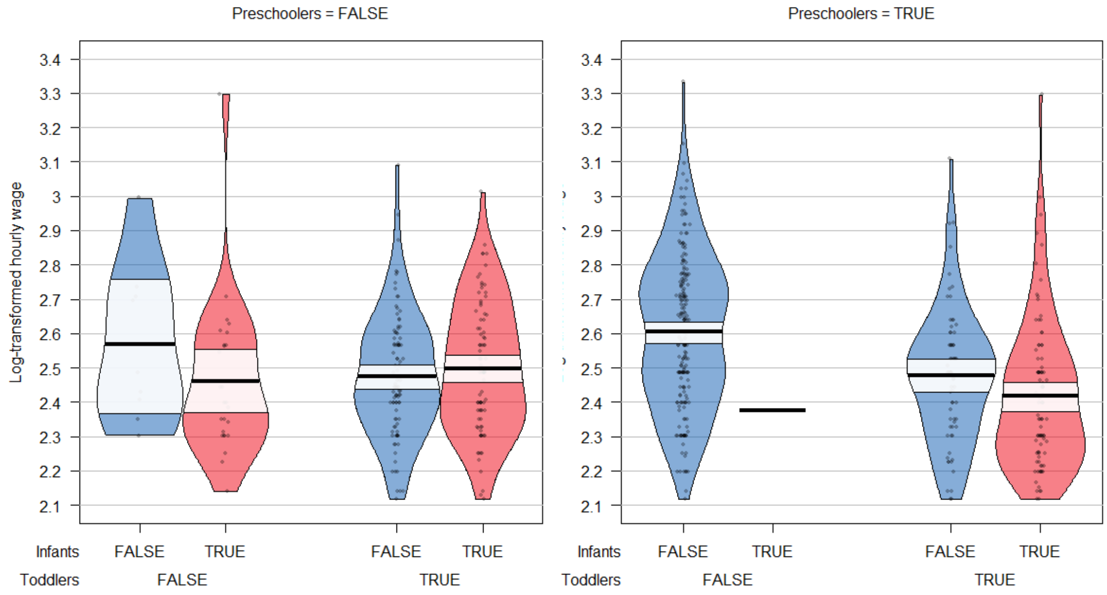
```
### tables and plots of response and predictors
library(mosaic)

# table age group counts
ageGroupTable1 <- xtabs(~ Infants + Toddlers + Preschoolers,
                        data = workforceSub)
ftable(ageGroupTable1)

##                   Preschoolers FALSE TRUE
## Infants Toddlers
## FALSE   FALSE                      9  194
##         TRUE                      95   69
## TRUE    FALSE                     26    1
##         TRUE                      85   90

# table each age group by wage
favstats(HourlyWage ~ Infants + Toddlers + Preschoolers, data = workforce)

##   Infants.Toddlers.Preschoolers   min      Q1 median      Q3   max
## 1             FALSE.FALSE.FALSE 10.00 11.1100 12.000 15.0000 20.00
## 2              TRUE.FALSE.FALSE  8.50 10.0250 11.000 13.0000 27.00
## 3              FALSE.TRUE.FALSE  8.30 10.5000 11.500 13.2550 22.00
## 4               TRUE.TRUE.FALSE  8.30 10.5000 12.000 14.0000 20.34
## 5              FALSE.FALSE.TRUE  8.30 11.4775 13.475 15.7875 28.00
## 6               TRUE.FALSE.TRUE 10.75 10.7500 10.750 10.7500 10.75
## 7               FALSE.TRUE.TRUE  8.30 10.4000 12.000 13.0000 22.40
## 8                TRUE.TRUE.TRUE  8.30  9.5050 10.800 12.4175 27.00
##       mean       sd  n missing
## 1 13.36000 3.233570  9       2
## 2 12.04885 3.499566 26       6
## 3 12.08189 2.316078  95       12
## 4 12.39259 2.426471  85        3
## 5 13.86918 3.225693 194       25
## 6 10.75000       NA   1        1
## 7 12.16087 2.574214  69       17
## 8 11.51256 2.946528  90       12

# transform response
workforceSub$LogWage <- log(workforceSub$HourlyWage)

# plot untransformed wage data
par(mfrow = c(1, 2))
pirateplot(HourlyWage ~ WageReported, xlab = "",
           ylab = "Hourly wage (U.S. dollars)",
           main = "Distribution of Hourly Wage",
           data = workforceSub, theme = 3)

# plot transformed wage data
pirateplot(LogWage ~ WageReported, xlab = "",
           ylab = "Log-transformed hourly wage (U.S. dollars)",
           main = "Distribution of Log-transformed Hourly Wage",
           data = workforceSub, theme = 3)
```



```
# plot untransformed wages by age groups
par(mfrow = c(1, 1))
pirateplot(HourlyWage ~ Infants + Toddlers + Preschoolers,
           data = workforceSub, theme = 3,
           ylab = "Hourly wage")
```



```r
# plot transformed wage by age groups
pirateplot(LogWage ~ Infants + Toddlers + Preschoolers,
           data = workforceSub, theme = 3,
           ylab = "Log-transformed hourly wage")
```



```r
### fit mixed models
library(lme4)
library(lmerTest)

# wage by 2-way interactions that include toddlers
lmer1 <- lmer(LogWage ~ Infants * Toddlers + Toddlers * Preschoolers +
              Age + EmployLengthYear + Level + HoursPerWeek + (1|County),
              data = workforceSub)
summary(lmer1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LogWage ~ Infants * Toddlers + Toddlers * Preschoolers + Age +
##     EmployLengthYear + Level + HoursPerWeek + (1 | County)
##    Data: workforceSub
##
## REML criterion at convergence: -246.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1081 -0.6949 -0.0743  0.5792  4.5866
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  County   (Intercept) 0.01012  0.1006
##  Residual             0.03020  0.1738
## Number of obs: 569, groups:  County, 26
##
## Fixed effects:
##                                 Estimate Std. Error         df t value
## (Intercept)                    2.357e+00  1.087e-01  5.475e+02  21.681
## InfantsTRUE                   -4.497e-02  6.433e-02  5.355e+02  -0.699
## ToddlersTRUE                  -3.739e-02  5.895e-02  5.321e+02  -0.634
## PreschoolersTRUE               2.374e-02  5.764e-02  5.316e+02   0.412
## Age                            5.372e-04  6.247e-04  5.371e+02   0.860
## EmployLengthYear               9.673e-03  1.541e-03  5.341e+02   6.275
## Level1                         1.523e-02  7.958e-02  5.295e+02   0.191
## Level2                         2.440e-02  7.944e-02  5.298e+02   0.307
## Level3                         6.119e-02  8.204e-02  5.303e+02   0.746
## Level4                         1.438e-01  8.511e-02  5.294e+02   1.690
## Level5                         1.462e-01  8.304e-02  5.315e+02   1.761
## Level6                         1.455e-01  9.884e-02  5.329e+02   1.472
## Level7                         1.667e-01  9.252e-02  5.294e+02   1.802
## Level8                         2.318e-01  8.441e-02  5.312e+02   2.746
## Level9                         4.440e-01  1.173e-01  5.284e+02   3.784
## HoursPerWeek                   1.413e-03  1.185e-03  5.504e+02   1.192
## InfantsTRUE:ToddlersTRUE       4.518e-02  6.713e-02  5.360e+02   0.673
## ToddlersTRUE:PreschoolersTRUE -6.510e-02  6.170e-02  5.330e+02  -1.055
##                               Pr(>|t|)
## (Intercept)                    < 2e-16 ***
## InfantsTRUE                   0.484832
## ToddlersTRUE                  0.526163
## PreschoolersTRUE              0.680597
## Age                           0.390223
## EmployLengthYear              7.24e-10 ***
## Level1                        0.848259
## Level2                        0.758856
## Level3                        0.456121
## Level4                        0.091691 .
## Level5                        0.078838 .
## Level6                        0.141682
## Level7                        0.072094 .
## Level8                        0.006238 **
## Level9                        0.000172 ***
## HoursPerWeek                  0.233612
## InfantsTRUE:ToddlersTRUE      0.501202
## ToddlersTRUE:PreschoolersTRUE 0.291917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# reduce model (drop infants*toddlers)
lmer2 <- lmer(LogWage ~ Toddlers * Preschoolers + Infants +
              Age + EmployLengthYear + Level + HoursPerWeek + (1|County),
              data = workforceSub)
summary(lmer2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## LogWage ~ Toddlers * Preschoolers + Infants + Age + EmployLengthYear +
##     Level + HoursPerWeek + (1 | County)
##    Data: workforceSub
##
## REML criterion at convergence: -250
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1091 -0.6940 -0.0799  0.5804  4.5931
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  County   (Intercept) 0.01030  0.1015
##  Residual             0.03015  0.1736
## Number of obs: 569, groups:  County, 26
##
## Fixed effects:
##                                Estimate Std. Error         df t value
## (Intercept)                   2.330e+00  1.008e-01  5.426e+02  23.109
## ToddlersTRUE                 -4.749e-03  3.342e-02  5.440e+02  -0.142
## PreschoolersTRUE              5.401e-02  3.596e-02  5.437e+02   1.502
## InfantsTRUE                  -3.525e-03  1.873e-02  5.420e+02  -0.188
## Age                           5.333e-04  6.243e-04  5.380e+02   0.854
## EmployLengthYear              9.729e-03  1.538e-03  5.350e+02   6.324
## Level1                        1.486e-02  7.952e-02  5.305e+02   0.187
## Level2                        2.377e-02  7.938e-02  5.307e+02   0.300
## Level3                        5.994e-02  8.196e-02  5.313e+02   0.731
## Level4                         1.433e-01  8.504e-02  5.304e+02   1.685
## Level5                         1.464e-01  8.298e-02  5.324e+02   1.764
## Level6                         1.449e-01  9.876e-02  5.338e+02   1.467
## Level7                         1.682e-01  9.241e-02  5.304e+02   1.821
## Level8                         2.321e-01  8.434e-02  5.321e+02   2.752
## Level9                         4.445e-01  1.172e-01  5.293e+02   3.792
## HoursPerWeek                   1.311e-03  1.175e-03  5.512e+02   1.115
## ToddlersTRUE:PreschoolersTRUE -9.538e-02  4.214e-02  5.429e+02  -2.263
##                               Pr(>|t|)    
## (Intercept)                    < 2e-16 ***
## ToddlersTRUE                  0.887048    
## PreschoolersTRUE              0.133648    
## InfantsTRUE                   0.850801    
## Age                           0.393319    
## EmployLengthYear              5.38e-10 ***
## Level1                        0.851847    
## Level2                        0.764674    
## Level3                        0.464935    
## Level4                        0.092643 .  
## Level5                        0.078344 .  
## Level6                        0.142838    
## Level7                        0.069232 .  
## Level8                        0.006129 ** 
## Level9                        0.000167 ***
## HoursPerWeek                  0.265340    
## ToddlersTRUE:PreschoolersTRUE 0.024000 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### fit final model

# reduce model (drop infants)
lmer3 <- lmer(LogWage ~ Toddlers * Preschoolers + Age + EmployLengthYear +
                Level + HoursPerWeek + (1|County), data = workforceSub)
summary(lmer3)
```

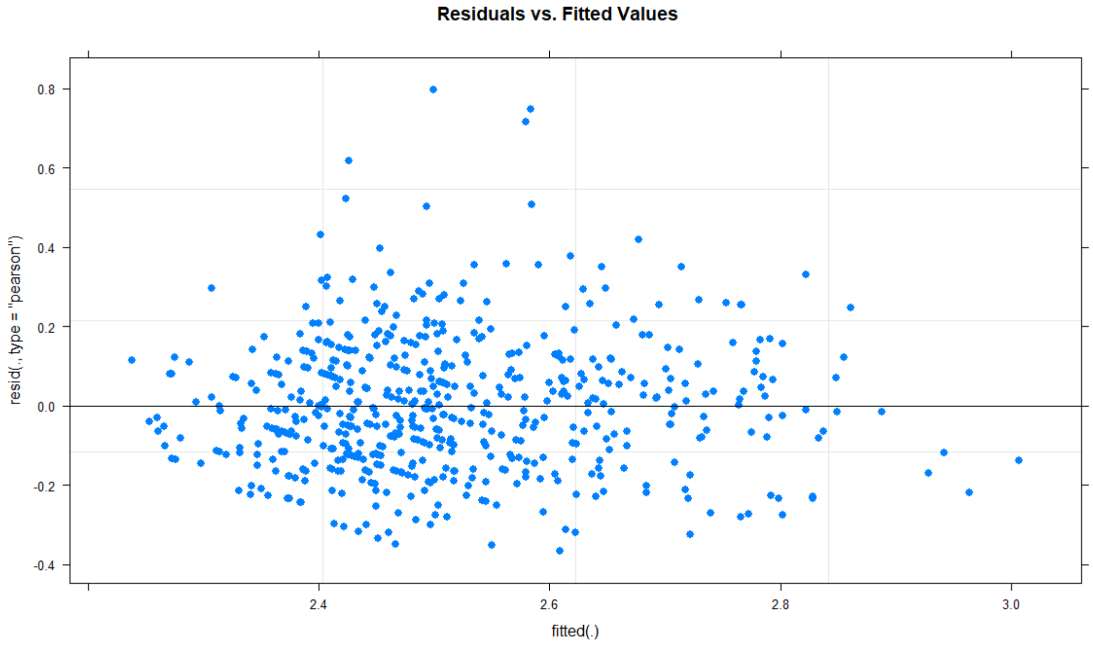
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: LogWage ~ Toddlers * Preschoolers + Age + EmployLengthYear +  
##     Level + HoursPerWeek + (1 | County)
##    Data: workforceSub
## 
## REML criterion at convergence: -256
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -2.1091 -0.7030 -0.0742  0.5793  4.5896 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  County   (Intercept) 0.01033  0.1016
##  Residual             0.03009  0.1735
## Number of obs: 569, groups:  County, 26
## 
## Fixed effects:
##                                  Estimate Std. Error         df t value
## (Intercept)                     2.328e+00  9.978e-02  5.429e+02  23.328
## ToddlersTRUE                   -3.698e-03  3.291e-02  5.451e+02  -0.112
## PreschoolersTRUE                5.658e-02  3.319e-02  5.455e+02   1.705
## Age                             5.353e-04  6.236e-04  5.390e+02   0.858
## EmployLengthYear                9.727e-03  1.537e-03  5.360e+02   6.329
## Level1                          1.495e-02  7.944e-02  5.314e+02   0.188
## Level2                          2.409e-02  7.928e-02  5.317e+02   0.304
## Level3                          6.014e-02  8.188e-02  5.323e+02   0.735
## Level4                          1.439e-01  8.491e-02  5.315e+02   1.694
## Level5                          1.466e-01  8.289e-02  5.334e+02   1.768
## Level6                          1.449e-01  9.867e-02  5.347e+02   1.469
## Level7                          1.689e-01  9.225e-02  5.314e+02   1.832
## Level8                          2.327e-01  8.420e-02  5.331e+02   2.763
## Level9                          4.443e-01  1.171e-01  5.302e+02   3.794
## HoursPerWeek                    1.295e-03  1.172e-03  5.523e+02   1.105
## ToddlersTRUE:PreschoolersTRUE -9.834e-02  3.904e-02  5.451e+02  -2.519
##                               Pr(>|t|)    
## (Intercept)                    < 2e-16 ***
## ToddlersTRUE                  0.910575    
## PreschoolersTRUE              0.088839 .  
## Age                           0.391043    
## EmployLengthYear              5.21e-10 ***
## Level1                        0.850767    
## Level2                        0.761362    
## Level3                        0.462937    
## Level4                        0.090805 .  
## Level5                        0.077630 .  
## Level6                        0.142513    
## Level7                        0.067585 .  
## Level8                        0.005917 ** 
## Level9                        0.000166 ***
## HoursPerWeek                  0.269686    
## ToddlersTRUE:PreschoolersTRUE 0.012044 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anovaResults <- anova(lmer3, type="II")
kable(anovaResults, format = "markdown")
```

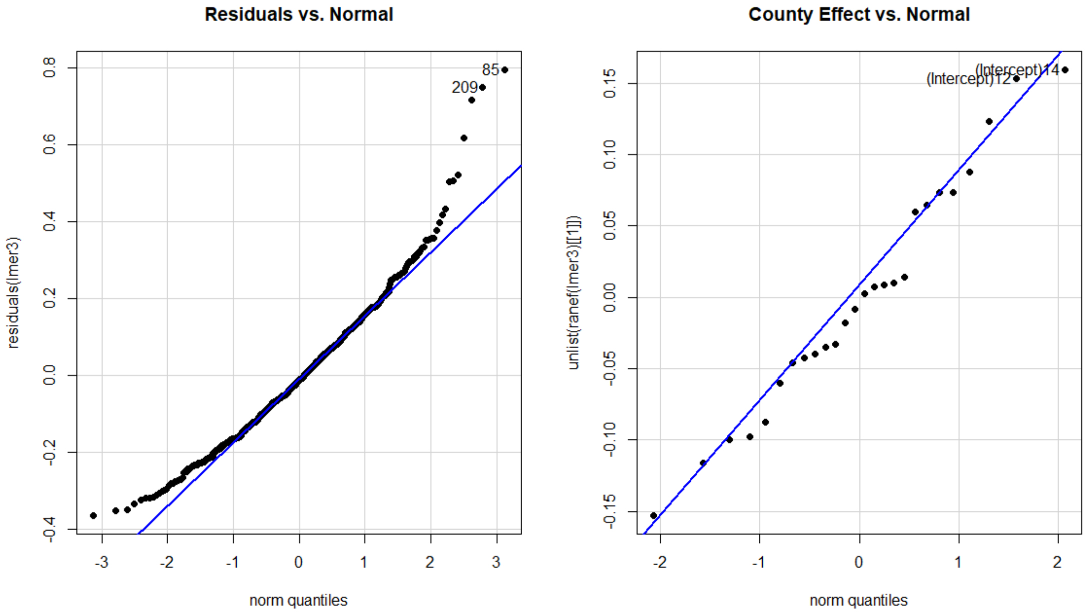
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Toddlers	0.5799354	0.5799354	1	542.9754	19.2718299	0.0000136

Preschoolers	0.0262466	0.0262466	1	544.5649	0.8722015	0.3507605
Age	0.0221746	0.0221746	1	538.9814	0.7368830	0.3910428
EmployLengthYear	1.2054291	1.2054291	1	536.0000	40.0576055	0.0000000
Level	2.6471090	0.2941232	9	537.4779	9.7740066	0.0000000
HoursPerWeek	0.0367366	0.0367366	1	552.2636	1.2207938	0.2696860
Toddlers:Preschoolers	0.1909896	0.1909896	1	545.1202	6.3467734	0.0120443

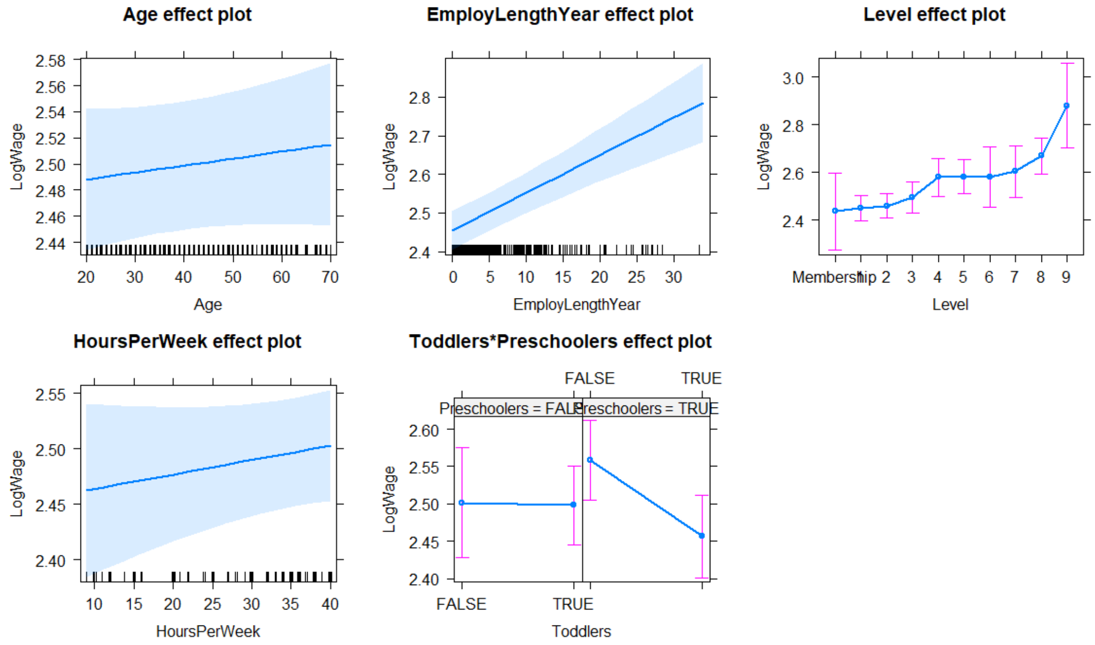
```
### assess model diagnostics
library(car)
# constant variance
plot(lmer3, pch=16, main = "Residuals vs. Fitted Values")
```



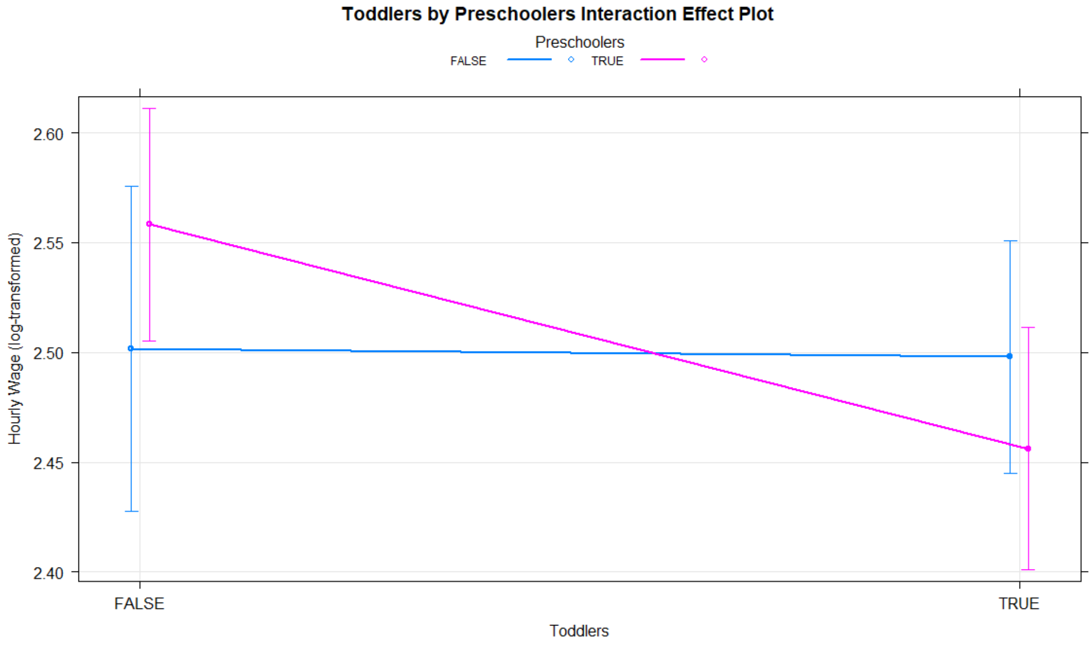
```
par(mfrow = c(1, 2))
# normality of residuals
q1 <- qqPlot(residuals(lmer3), envelope=F, pch=16, main = "Residuals vs. Normal")
# normality of random effect
q1 <- qqPlot(unlist(ranef(lmer3)[[1]]), envelope=F, pch=16,
             main = "County Effect vs. Normal")
```



```
### effects plot
par(mfrow = c(1, 1))
library(effects)
plot(allEffects(lmer3))
```



```
plot(allEffects(lmer3)[5], multiline=T, ci.style="bars", grid=T,
     ylab = "Hourly Wage (log-transformed)",
     main = "Toddlers by Preschoolers Interaction Effect Plot")
```



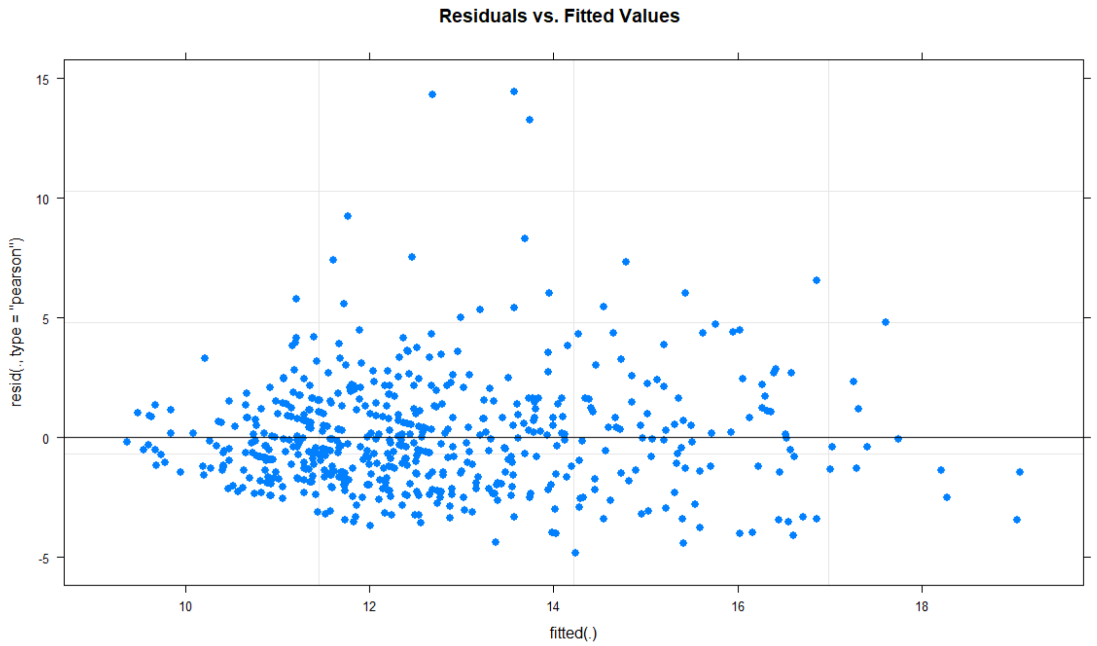
```
### get confidence intervals
confint(lmer3)
```

```
##                                       2.5 %       97.5 %
## .sig01                         0.0637391832  0.149517657
## .sigma                         0.1614196579  0.181812795
## (Intercept)                    2.1334567565  2.520567299
## ToddlersTRUE                  -0.0673527953  0.060134986
## PreschoolersTRUE              -0.0076033956  0.121035336
## Age                           -0.0006704722  0.001749748
## EmployLengthYear               0.0067389622  0.012698886
## Level1                        -0.1387456565  0.169044886
## Level2                        -0.1293248874  0.177804025
## Level3                        -0.0983466525  0.218800526
## Level4                        -0.0203939644  0.308585215
## Level5                        -0.0138037504  0.307352178
## Level6                        -0.0458687889  0.336862240
## Level7                        -0.0094444701  0.348113725
## Level8                         0.0698410320  0.396156989
## Level9                         0.2176288808  0.671318214
## HoursPerWeek                  -0.0009759552  0.003590702
## ToddlersTRUE:PreschoolersTRUE -0.1740808553 -0.022850374
```

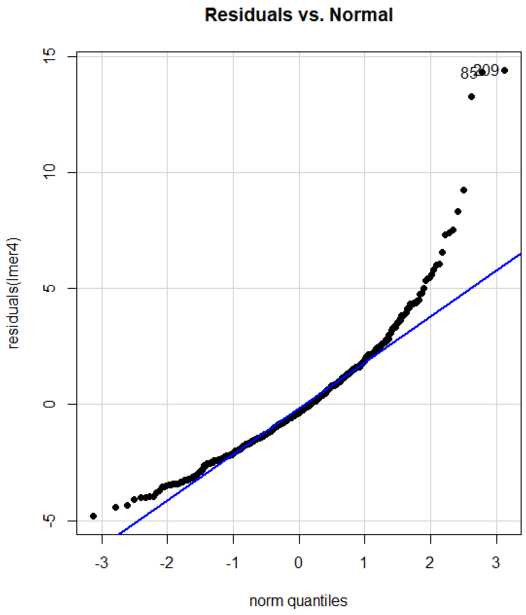
```
### get r-squared
library(MuMIn)
r.squaredGLMM(lmer3)
```

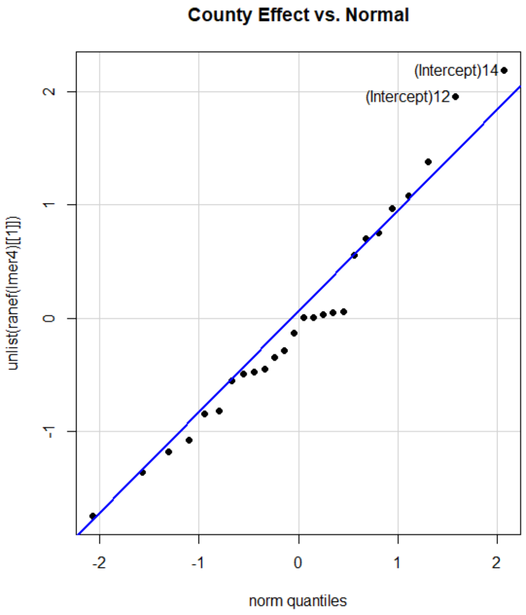
```
##            R2m       R2c
## [1,] 0.2352519 0.4306542
```

```
### get icc
# icc = county variance / (county variance + residual variance)
icc <- 0.01033 / (0.01033 + 0.03009)
icc

## [1] 0.2555666

### fit same model with untransformed wage
lmer4 <- lmer(HourlyWage ~ Toddlers * Preschoolers + Age + EmployLengthYear +
                Level + HoursPerWeek + (1|County), data = workforceSub)
anova(lmer4, type="II")

## Type II Analysis of Variance Table with Satterthwaite's method
##                        Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)    
## Toddlers               104.90 104.901     1 544.37 17.8976 2.735e-05 ***
## Preschoolers             1.64   1.641     1 545.90  0.2800   0.59690    
## Age                      3.24   3.237     1 540.35  0.5523   0.45772    
## EmployLengthYear       231.22 231.225     1 536.96 39.4503 6.955e-10 ***
## Level                  487.19  54.132     9 538.80  9.2357 4.423e-13 ***
## HoursPerWeek             3.91   3.910     1 552.81  0.6671   0.41442    
## Toddlers:Preschoolers   28.03  28.027     1 545.84  4.7818   0.02919 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lmer4)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: HourlyWage ~ Toddlers * Preschoolers + Age + EmployLengthYear +  
##     Level + HoursPerWeek + (1 | County)
##    Data: workforceSub
## 
## REML criterion at convergence: 2655.6
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.9910 -0.6277 -0.1417  0.4783  5.9597 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  County   (Intercept) 1.592    1.262   
##  Residual             5.861    2.421   
## Number of obs: 569, groups:  County, 26
## 
## Fixed effects:
##                                   Estimate Std. Error         df t value
## (Intercept)                      10.319646   1.385330 547.546251   7.449
## ToddlersTRUE                     -0.145479   0.458644 545.500066  -0.317
## PreschoolersTRUE                  0.750543   0.462562 546.026867   1.623
## Age                               0.006462   0.008695 540.353062   0.743
## EmployLengthYear                  0.134643   0.021437 536.960364   6.281
## Level1                          0.362634   1.108503 532.292154   0.327
## Level2                          0.340263   1.106216 532.664544   0.308
## Level3                          0.790705   1.142376 533.363772   0.692
## Level4                          1.898934   1.184696 532.420565   1.603
## Level5                          1.855176   1.156442 534.591227   1.604
## Level6                          2.101051   1.376414 536.085096   1.526
## Level7                          2.295669   1.287143 532.196409   1.784
## Level8                          3.300509   1.174675 534.229430   2.810
## Level9                          6.396472   1.634390 530.956860   3.914
## HoursPerWeek                    0.013319   0.016307 552.806915   0.817
## ToddlersTRUE:PreschoolersTRUE  -1.189511   0.543969 545.837627  -2.187
##                               Pr(>|t|)    
## (Intercept)                   3.68e-13 ***
## ToddlersTRUE                  0.751217    
## PreschoolersTRUE              0.105257    
## Age                           0.457722    
## EmployLengthYear              6.96e-10 ***
## Level1                        0.743692    
## Level2                        0.758513    
## Level3                        0.489139    
## Level4                        0.109552    
## Level5                        0.109258    
## Level6                        0.127483    
## Level7                        0.075068 .  
## Level8                        0.005140 ** 
## Level9                        0.000103 ***
## HoursPerWeek                  0.414420    
## ToddlersTRUE:PreschoolersTRUE 0.029186 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

### model diagnostics
# constant variance
plot(lmer4, pch=16, main = "Residuals vs. Fitted Values")
```


Residuals vs. Fitted Values

```r
par(mfrow = c(1, 2))
# normality of residuals
q1 <- qqPlot(residuals(lmer4), envelope=F, pch=16,
             main = "Residuals vs. Normal")
# normality of random effect
q1 <- qqPlot(unlist(ranef(lmer4)[[1]]), envelope=F, pch=16,
             main = "County Effect vs. Normal")
```


Residuals vs. Normal


County Effect vs. Normal