R Code

2025-06-05

```
#Load Packages
pacman::p_load("tidyverse", "lme4", "lmerTest", "mlmhelpr", "haven", "sjPlot", "quest", "mlmhelpr", "lattice", "interaction
s", "ggeffects", "lmtest", "sandwich", "multiwayvcov", "stargazer", "psych", "scales", "merTools", "emmeans", "HLMdiag", "sj
Plot", "stringr", "mice", "naniar", "car", "patchwork")
```

```
#Load Data
soep <- readRDS("C:/Users/lenn0/AppData/Local/Temp/834af560-b32b-440a-afdf-938e22b26caa_SOEP-CORE.v39eu_R_EN.zip.caa/R_EN/so
epdata/pequiv.rds")
```

```
#Data Wrangling

#Select variables that will be used in the subsequent analyses
soep_sub <- soep %>%
  dplyr::select(x11101ll, x11102, d11101, d11102ll, d11104,i11110, syear, d11108) %>%
  arrange(x11101ll)

#Assign meaningful names to the variables
soep_sub <- soep_sub %>%
  rename(id_id = x11101ll) %>%
  rename(hh_id = x11102) %>%
  rename(age = d11101) %>%
  rename(gender = d11102ll) %>%
  rename(marital_status_unfinished = d11104) %>%
  rename(income = i11110) %>%
  rename(year = syear) %>%
  rename(education = d11108)

#Count units on each level
data.frame(
  level1_observations = nrow(soep_sub),
  level2_individuals = length(unique(soep_sub$id_id)),
  level3_households = length(unique(soep_sub$hh_id)))
```

```
##   level1_observations level2_individuals level3_households
## 1             1148926             179412             66909
```

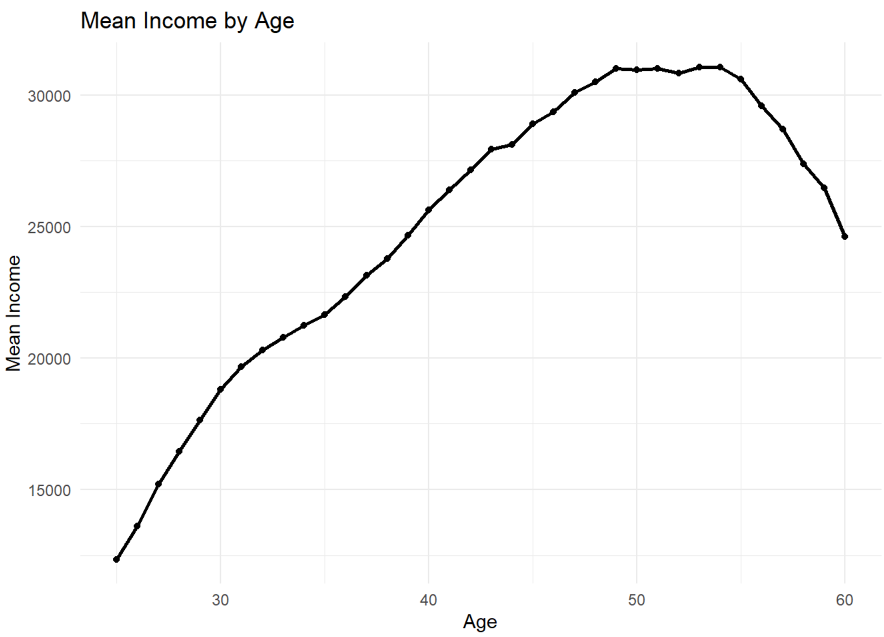
```
#Data Wrangling

#Code variables as numeric
soep_sub$income <- as.numeric(soep_sub$income)
soep_sub$age <- as.numeric(soep_sub$age)
soep_sub$gender <- as.numeric(soep_sub$gender)
soep_sub$marital_status_unfinished <- as.numeric(soep_sub$marital_status_unfinished)
soep_sub$education <- as.numeric(soep_sub$education)

#Filter the dataset for NAs and values that are relevant for the subsequent analyses
soep_fil <- soep_sub %>%
  filter(income >= 0) %>%
  filter(age >= 25 & age <= 60) %>%
  filter(gender > 0) %>%
  filter(marital_status_unfinished > 0) %>%
  filter(education != -1)


#Filter individuals between 25 and 60
soep_fil1 <- soep_sub %>%
  filter(age >= 25 & age <= 60)

#Recode gender so that male = 1 and female = 0
soep_fil <- soep_fil %>%
  mutate(male = case_when(
    gender == 1 ~ 1,
    gender == 2 ~ 0,
    TRUE ~ NA_real_))

#Recode marital status so that single = 0 and married = 1
soep_fil <- soep_fil %>%
  mutate(marital_status = marital_status_unfinished - 1) %>%
  mutate(marital_status = case_when(
    marital_status == 0 ~ 1,
    marital_status == 1 ~ 0,
    TRUE ~ marital_status ))


#Recode marital status so that single = 0 and married = 1
soep_fil <- soep_fil %>%
  mutate(education = education - 1)

soep_fil$education <- as.character(soep_fil$education)

#Create a new variable for which single = 0, married = 1, and other = 2
soep_fil <- soep_fil %>%
  mutate(marital_status_reduced = case_when(
    marital_status == 0 ~ 0,
    marital_status == 1 ~ 1,
    marital_status > 1 ~ 2
  ))

#Create a quadratic polynomial of age
soep_fil <- soep_fil %>%
  mutate(age_squared = age^2)

#Create a variable representing the logarithm of income
soep_fil <- soep_fil %>%
  mutate(log_income = case_when(
    income > 0  ~ log(income),
    income == 0 ~ 0,
    TRUE        ~ NA_real_))
```

```
#Data Wrangling

#Center age and age-squared
soep_fil <- soep_fil %>%
  mutate(age_c = age - mean(age, na.rm = T)) %>%
  mutate(age_squared_c = age_squared - mean(age_squared, na.rm = T))
```

```
#Data Wrangling

#Compute mean income for each age group (by years)
soep_summary1 <- soep_fil %>%
  group_by(age) %>%
  summarise(mean_income_age = mean(income, na.rm = T), .groups = "drop")


#Compute mean income for each age group conditioned on gender
soep_summary3 <- soep_fil %>%
  group_by(age, male) %>%
  summarise(mean_income_age = mean(income, na.rm = TRUE), .groups = "drop")


#Compute mean income for each age group conditioned on gender and marital status
soep_summary5 <- soep_fil %>%
  mutate(gender_marital_status = case_when(
    male == 0 & marital_status_reduced == 0 ~ "female_single",
    male == 0 & marital_status_reduced == 1 ~ "female_married",
    male == 0 & marital_status_reduced == 2 ~ "female_other",
    male == 1 & marital_status_reduced == 0 ~ "male_single",
    male == 1 & marital_status_reduced == 1 ~ "male_married",
    male == 1 & marital_status_reduced == 2 ~ "male_other")) %>%
  group_by(age, gender_marital_status) %>%
  summarise(mean_income_age_marital_status_gender = mean(income, na.rm = TRUE), .groups = "drop")
```

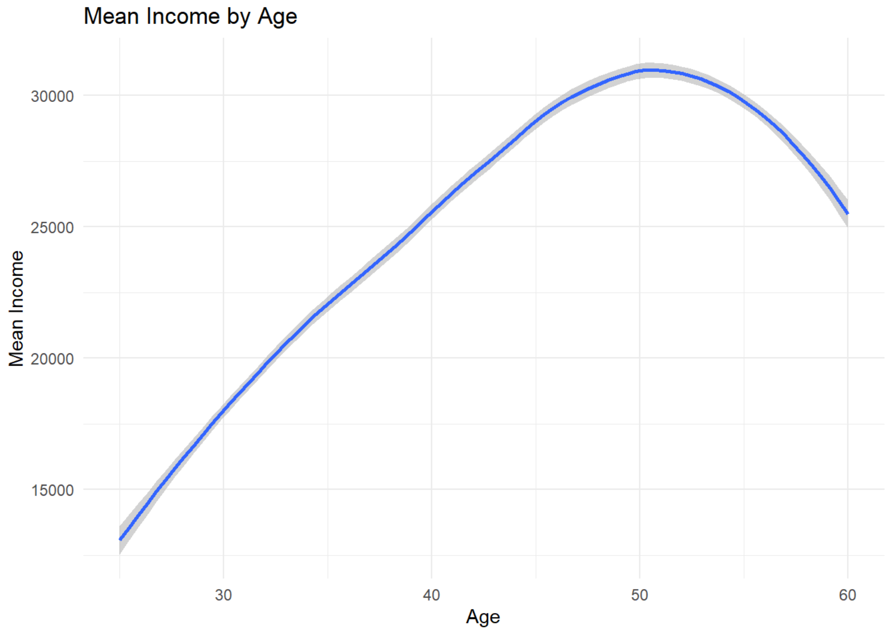
```
#Graphical Analysis

#Mean income
ggplot(soep_summary1, aes(x = age, y = mean_income_age))+
  geom_line(size =1) +
  geom_point() +
  labs(
    title = "Mean Income by Age",
    x = "Age",
    y = "Mean Income") +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

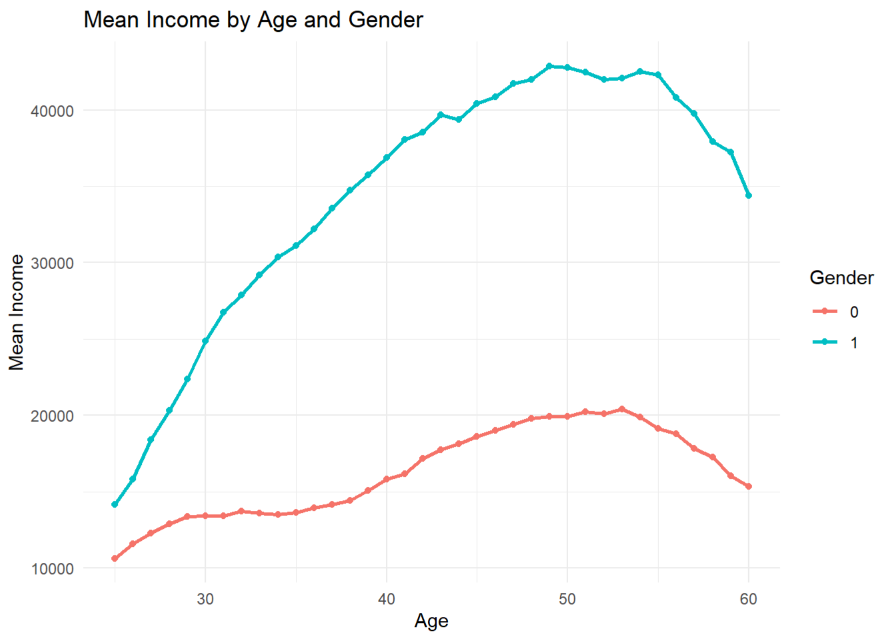



```
ggplot(soep_summary1, aes(x = age, y = mean_income_age))+
  geom_smooth(size =1) +
  labs(
    title = "Mean Income by Age",
    x = "Age",
    y = "Mean Income") +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

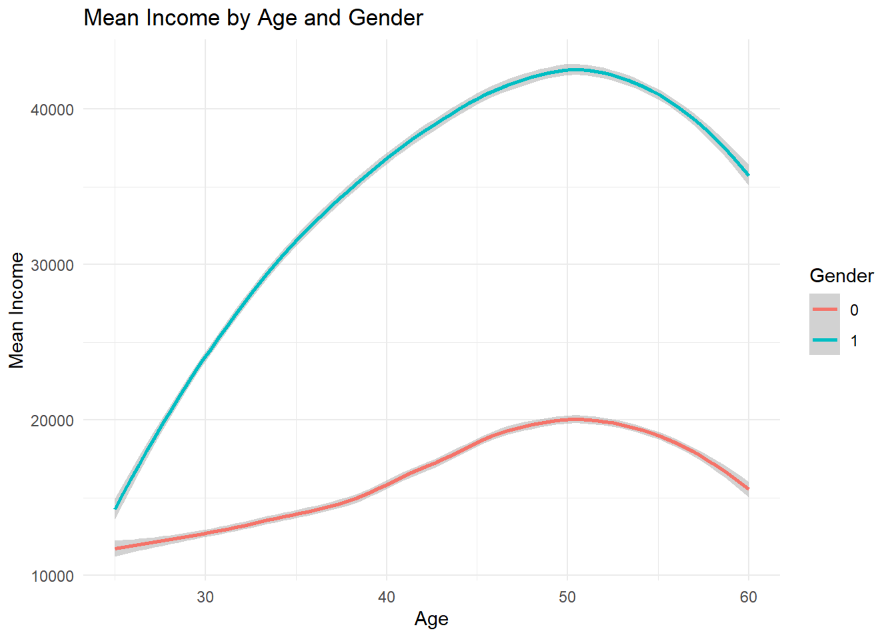



```
#Mean income by age and gender
ggplot(soep_summary3, aes(x = age, y = mean_income_age, color = as.factor(male))) +
  geom_line(size = 1) +
  geom_point() +
  labs(
    title = "Mean Income by Age and Gender",
    x = "Age",
    y = "Mean Income",
    color = "Gender") +
  theme_minimal()
```

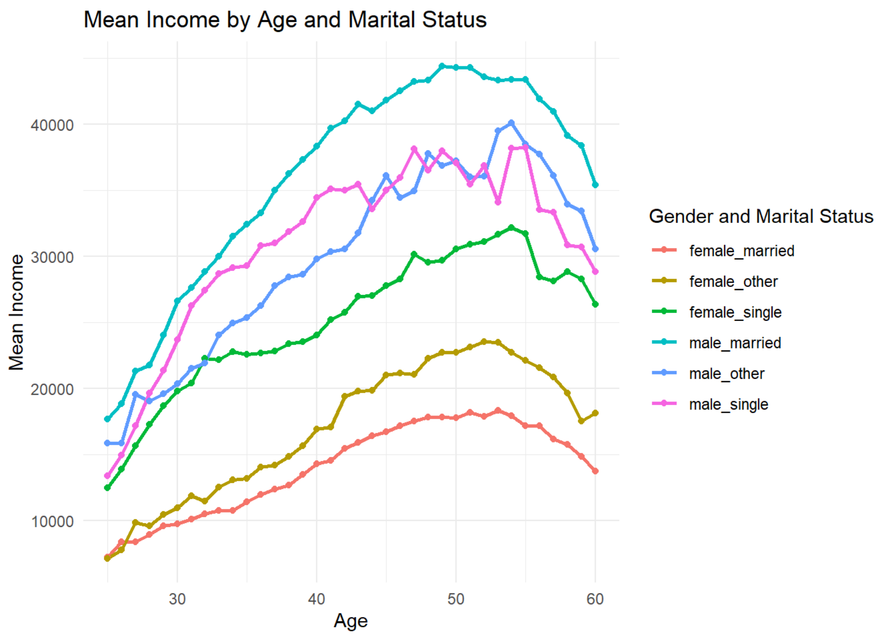



```
ggplot(soep_summary3, aes(x = age, y = mean_income_age, color = as.factor(male))) +
  geom_smooth(size = 1) +
  labs(
    title = "Mean Income by Age and Gender",
    x = "Age",
    y = "Mean Income",
    color = "Gender") +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

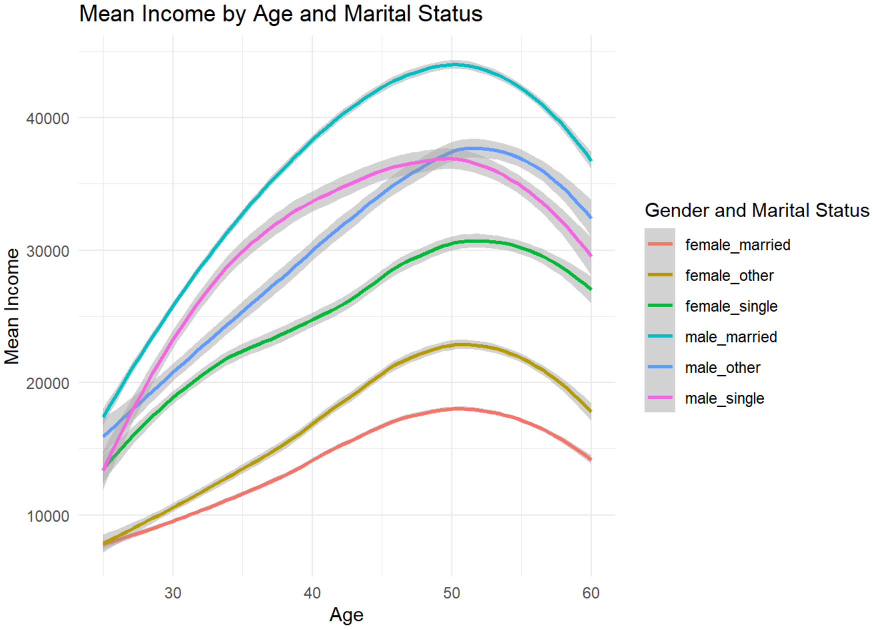



```
#Mean income by age, gender, and marital status
ggplot(soep_summary5, aes(x = age, y = mean_income_age_marital_status_gender, color = as.factor(gender_marital_status))) +
  geom_line(size = 1) +
  geom_point() +
  labs(
    title = "Mean Income by Age and Marital Status",
    x = "Age",
    y = "Mean Income",
    color = "Gender and Marital Status") +
  theme_minimal()
```




```
ggplot(soep_summary5, aes(x = age, y = mean_income_age_marital_status_gender, color = as.factor(gender_marital_status))) +
  geom_smooth(size = 1) +
  labs(
    title = "Mean Income by Age and Marital Status",
    x = "Age",
    y = "Mean Income",
    color = "Gender and Marital Status") +
  theme_minimal()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
#Filter the dataset for individuals that went from unmarried to married during the course of the panel
ids_changed <- soep_fil %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(prev_marital_status = lag(marital_status_reduced)) %>%
  filter(prev_marital_status == 0 & marital_status_reduced == 1) %>%
  distinct(id_id)

soep_changed <- soep_fil %>%
  filter(id_id %in% ids_changed$id_id)
```

```
#Determine the year in which the individual got married
soep_d <- soep_fil %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(prev_marital_status = lag(marital_status_reduced),
         transition = prev_marital_status == 0 & marital_status_reduced == 1) %>%
  filter(transition) %>%
  slice(1) %>%
  dplyr::select(id_id, transition_year = year)


#Center the variable so that 0 = year of marriage
soep_final <- soep_fil %>%
  inner_join(soep_d, by = "id_id") %>%
  mutate(relative_year = year - transition_year)

#Once married individuals are permanetly coded as married
soep_final <- soep_final %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(married = cummax(marital_status_reduced %in% c(1, 2)))
```

```
#Count observations on each level
data.frame(
  level1_observations = nrow(soep_final),
  level2_individuals = length(unique(soep_final$id_id)),
  level3_households = length(unique(soep_final$hh_id)))
```

```
##   level1_observations level2_individuals level3_households
## 1               64997               4885              4730
```

```
#Check skew
describe(soep_final$income)
```

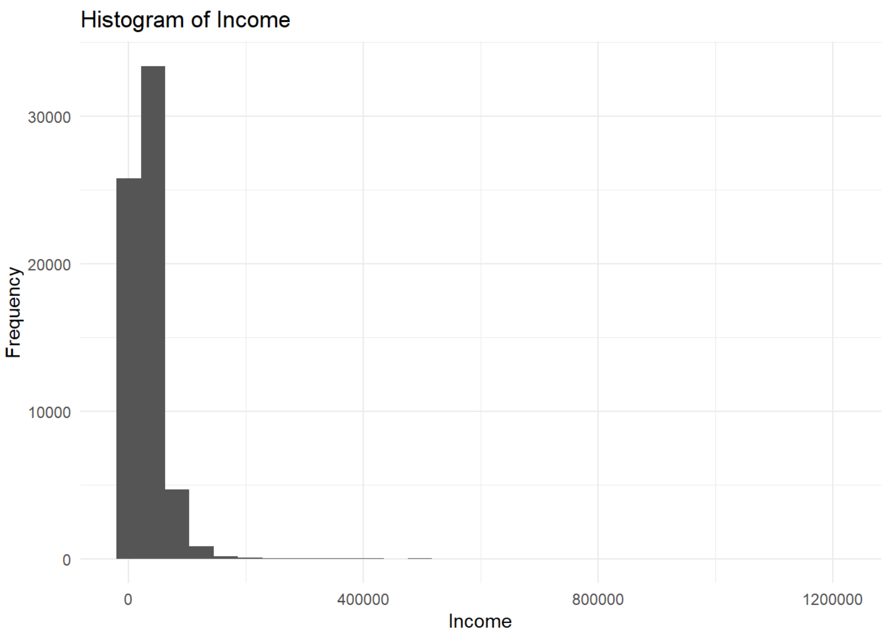
```
##    vars     n     mean       sd median  trimmed      mad min     max   range
## X1    1 64997 29828.82 27721.38  25695 26405.71 21032.16   0 1199988 1199988
##    skew kurtosis     se
## X1  4.7    94.07 108.73
```

```
describe(soep_final$log_income)
```

```
##    vars     n mean   sd median trimmed mad min max range skew kurtosis   se
## X1    1 64997 9.07 3.15  10.15    9.92 0.8   0  14    14 -2.3      3.9 0.01
```

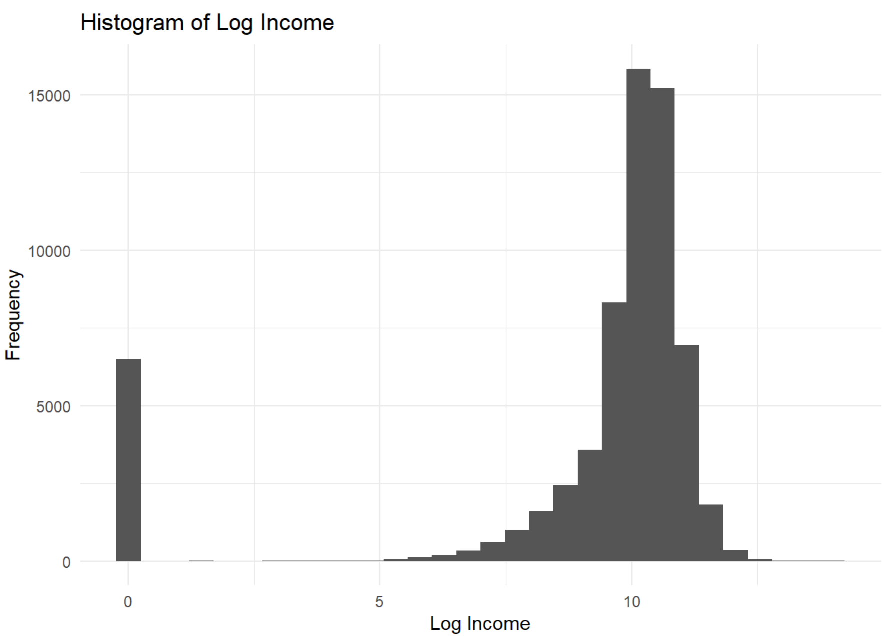
```
#Histogram of income
ggplot(soep_final, aes(x = income)) +
  geom_histogram() +
  labs(
    title = "Histogram of Income",
    x = "Income",
    y = "Frequency"
  ) +
  theme_minimal()
```

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




```
#Histogram of log income
ggplot(soep_final, aes(x = log_income)) +
  geom_histogram() +
  labs(
    title = "Histogram of Log Income",
    x = "Log Income",
    y = "Frequency"
  ) +
  theme_minimal()
```

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



```
soep_final <- soep_final %>%
  mutate(relative_year_squared = relative_year^2)


#Standardize variables
soep_final <- soep_final %>%
  mutate(
    age_sd = (age_c) / sd(age_c, na.rm = TRUE),
    age_squared_sd = (age_squared_c) / sd(age_squared_c, na.rm = TRUE),
    married_sd = (married - mean(married, na.rm = TRUE)) / sd(married, na.rm = TRUE),
    male_sd = (male - mean(male, na.rm = TRUE)) / sd(male, na.rm = TRUE),
    relative_year_sd = (relative_year - mean(relative_year, na.rm = TRUE)) / sd(relative_year, na.rm = TRUE),
    relative_year_squared_sd = (relative_year_squared - mean(relative_year_squared, na.rm = TRUE)) / sd(relative_year_square
d, na.rm = TRUE))

#Center transition_year
soep_final <- soep_final %>%
  mutate(transition_year = transition_year - 1985)


#Convert the variables "male" and "married" to factors
soep_final <- soep_final %>%
  ungroup() %>%
  mutate(
    male = factor(male, levels = c(0, 1), labels = c("Female", "Male")),
    married = factor(married, levels = c(0, 1), labels = c("No", "Yes"))
  )

#Create a new time variable with 0 = start of the panel
soep_final <- soep_final %>%
  mutate(year_1984 = year - 1984)

soep_final$education <- factor(soep_final$education)
```

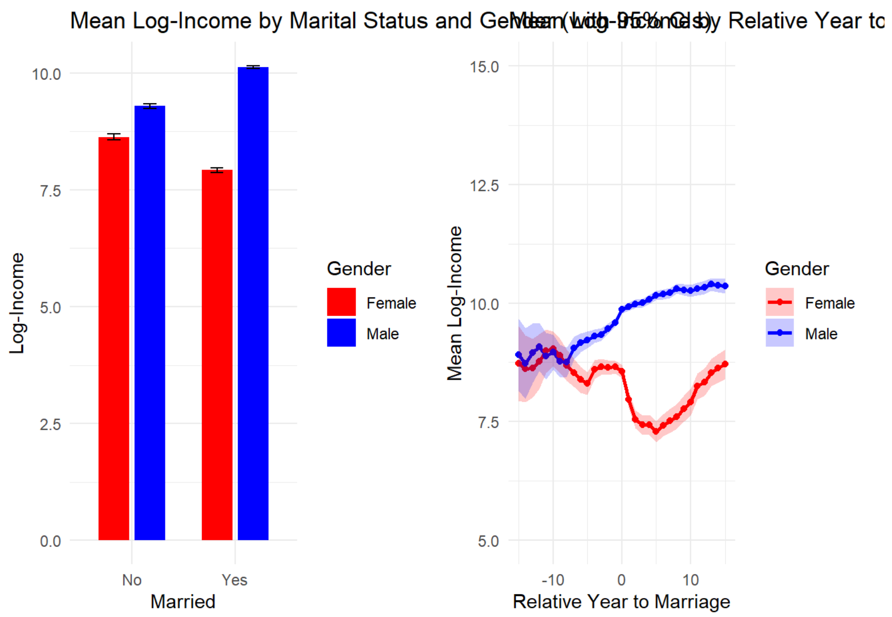
```
# Compute mean log income and CIs
soep_k <- soep_final %>%
  group_by(male, married) %>%
  summarise(
    mean_log_income = mean(log_income, na.rm = TRUE),
    se_log = sd(log_income, na.rm = TRUE) / sqrt(n()),
    ci_lower_log = mean_log_income - 1.96 * se_log,
    ci_upper_log = mean_log_income + 1.96 * se_log,
    .groups = "drop"
  )

# Plot with CIs
p1 <- ggplot(soep_k, aes(x = as.factor(married), y = mean_log_income, fill = as.factor(male))) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
  geom_errorbar(
    aes(ymin = ci_lower_log, ymax = ci_upper_log),
    position = position_dodge(width = 0.7),
    width = 0.25,
    color = "black",
    linewidth = 0.5
  ) +
  scale_x_discrete(labels = c("0" = "Not Married", "1" = "Married")) +
  scale_fill_manual(
    values = c("Female" = "red", "Male" = "blue"),
    labels = c("Female", "Male")
  ) +
  labs(
    title = "Mean Log-Income by Marital Status and Gender (with 95% CIs)",
    x = "Married",
    y = "Log-Income",
    fill = "Gender"
  ) +
  theme_minimal()


soep_l <- soep_final %>%
  group_by(relative_year, male) %>%
  summarise(
    log_mean_income = mean(log_income, na.rm = TRUE),
    sd_log_income = sd(log_income, na.rm = TRUE),
    n = sum(!is.na(log_income)),
    se = sd_log_income / sqrt(n),
    ci_lower = log_mean_income - 1.96 * se,
    ci_upper = log_mean_income + 1.96 * se
  ) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'relative_year'. You can override using the
## `.groups` argument.
```

```
p2 <- ggplot(soep_l, aes(x = relative_year, y = log_mean_income, color = as.factor(male))) +
  geom_line(size = 1) +
  geom_point() +
  geom_ribbon(
    aes(ymin = ci_lower, ymax = ci_upper, fill = as.factor(male)),
    alpha = 0.2,
    color = NA) +
  scale_x_continuous(limits = c(-15, 15)) +
  scale_y_continuous(limits = c(5, 15)) +
  scale_color_manual(
    values = c("Female" = "red", "Male" = "blue"),
    labels = c("Female", "Male")) +
  scale_fill_manual(
    values = c("Female" = "red", "Male" = "blue"),
    labels = c("Female", "Male")) +
  labs(
    title = "Mean Log-Income by Relative Year to Marriage (95% CIs)",
    x = "Relative Year to Marriage",
    y = "Mean Log-Income",
    color = "Gender",
    fill = "Gender") +
  theme_minimal()


#Combine and plot the pllts for the graphical analysis
combined_plot <- p1 + p2
combined_plot
```

```
## Warning: Removed 65 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

```
## Warning: Removed 65 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
#OLS baseline model
ols0 <- lm(log_income ~ year_1984, data = soep_final)
summary(ols0)
```

```
##
## Call:
## lm(formula = log_income ~ year_1984, data = soep_final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4699  0.3131  1.1270  1.5306  4.6215
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.435119   0.033732  250.06   <2e-16 ***
## year_1984   0.027232   0.001346   20.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.145 on 64995 degrees of freedom
## Multiple R-squared:  0.006257,   Adjusted R-squared:  0.006242
## F-statistic: 409.2 on 1 and 64995 DF,  p-value: < 2.2e-16
```

```
#OLS final Model 1
ols1 <- lm(log_income ~ year_1984 + age_sd + age_squared_sd + education + married*male + transition_year, data = soep_final)
summary(ols1)
```

```
##
## Call:
## lm(formula = log_income ~ year_1984 + age_sd + age_squared_sd +
##     education + married * male + transition_year, data = soep_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3323   0.0012   0.6928   1.5314   5.1979
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.506856   0.064039 101.608   <2e-16 ***
## year_1984            0.031226   0.002292  13.624   <2e-16 ***
## age_sd               0.080388   0.035021   2.295   0.0217 *
## age_squared_sd      -0.022182   0.026647  -0.832   0.4052
## education1           1.544496   0.048540  31.819   <2e-16 ***
## education2           2.326532   0.050337  46.219   <2e-16 ***
## marriedYes          -1.033450   0.042951 -24.061   <2e-16 ***
## maleMale             0.720245   0.040921  17.601   <2e-16 ***
## transition_year     -0.002289   0.002124  -1.078   0.2812
## marriedYes:maleMale  1.541858   0.049539  31.124   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.938 on 64987 degrees of freedom
## Multiple R-squared:  0.1326, Adjusted R-squared:  0.1325
## F-statistic:  1104 on 9 and 64987 DF,  p-value: < 2.2e-16
```

```
clustered_se <- cluster.vcov(ols1, cluster = soep_final[, c("id_id", "hh_id")])

#OLS final Model 1 with clustered se
ols1_clustered_se <- coeftest(ols1, vcov = clustered_se)
summary(ols1_clustered_se)
```

```
##     Estimate          Std. Error          t value             Pr(>|t|)
##  Min.   :-1.03345   Min.   :0.005548   Min.   :-12.72343   Min.   :0.0000
##  1st Qu.: 0.00609   1st Qu.:0.066494   1st Qu.: -0.02927   1st Qu.:0.0000
##  Median : 0.40032   Median :0.084783   Median :  7.27671   Median :0.0000
##  Mean   : 1.16937   Mean   :0.096361   Mean   :  7.36450   Mean   :0.1786
##  3rd Qu.: 1.54384   3rd Qu.:0.154412   3rd Qu.: 12.44133   3rd Qu.:0.2721
##  Max.   : 6.50686   Max.   :0.199016   Max.   : 32.69512   Max.   :0.7321
```

```
stargazer(ols1, ols1_clustered_se, type = "text")
```

```
##
## ==========================================================
##                              Dependent variable:
##                     --------------------------------------
##                             log_income
##                               OLS             coefficient
##                                                  test
##                               (1)                 (2)
## ----------------------------------------------------------
## year_1984                    0.031***           0.031***
##                              (0.002)             (0.006)
##
## age_sd                       0.080**              0.080
##                              (0.035)             (0.088)
##
## age_squared_sd                -0.022             -0.022
##                              (0.027)             (0.065)
##
## education1                   1.544***           1.544***
##                              (0.049)             (0.173)
##
## education2                   2.327***           2.327***
##                              (0.050)             (0.176)
##
## marriedYes                  -1.033***           -1.033***
##                              (0.043)             (0.081)
##
## maleMale                     0.720***           0.720***
##                              (0.041)             (0.072)
##
## transition_year               -0.002             -0.002
##                              (0.002)             (0.006)
##
## marriedYes:maleMale          1.542***           1.542***
##                              (0.050)             (0.098)
##
## Constant                     6.507***           6.507***
##                              (0.064)             (0.199)
##
## ----------------------------------------------------------
## Observations                  64,997
## R2                            0.133
## Adjusted R2                   0.132
## Residual Std. Error      2.938 (df = 64987)
## F Statistic         1,103.850*** (df = 9; 64987)
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

```
#OLS final Model 2
ols2 <- lm(log_income ~ year_1984 + education + relative_year_sd*male + relative_year_squared_sd*male , data = soep_final)
summary(ols2)
```

```
##
## Call:
## lm(formula = log_income ~ year_1984 + education + relative_year_sd *
##     male + relative_year_squared_sd * male, data = soep_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6723   0.0088   0.7186   1.6041   5.6926
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      5.837556   0.060972   95.74  < 2e-16 ***
## year_1984                        0.023592   0.001589   14.85  < 2e-16 ***
## education1                       1.552302   0.048618   31.93  < 2e-16 ***
## education2                       2.331263   0.050410   46.25  < 2e-16 ***
## relative_year_sd                -0.072211   0.021301   -3.39 0.000699 ***
## maleMale                         1.781322   0.023385   76.17  < 2e-16 ***
## relative_year_squared_sd         0.066745   0.019691    3.39 0.000700 ***
## relative_year_sd:maleMale        0.412033   0.026733   15.41  < 2e-16 ***
## maleMale:relative_year_squared_sd -0.320318   0.026727  -11.98  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.955 on 64988 degrees of freedom
## Multiple R-squared:  0.1228, Adjusted R-squared:  0.1227
## F-statistic:  1137 on 8 and 64988 DF,  p-value: < 2.2e-16
```

```
clustered_se2 <- cluster.vcov(ols2, cluster = soep_final[, c("id_id", "hh_id")])

#OLS final Model 2 with clustered se
ols2_clustered_se <- coeftest(ols2, vcov = clustered_se2)
summary(ols2_clustered_se)
```

```
##     Estimate          Std. Error          t value          Pr(>|t|)
##  Min.   :-0.32032   Min.   :0.004017   Min.   :-8.266   Min.   :0.000e+00
##  1st Qu.: 0.02359   1st Qu.:0.038752   1st Qu.: 2.166   1st Qu.:0.000e+00
##  Median : 0.41203   Median :0.043932   Median : 9.026   Median :0.000e+00
##  Mean   : 1.29025   Mean   :0.085640   Mean   : 9.656   Mean   :1.418e-02
##  3rd Qu.: 1.78132   3rd Qu.:0.171975   3rd Qu.:13.381   3rd Qu.:4.000e-09
##  Max.   : 5.83756   Max.   :0.199184   Max.   :29.307   Max.   :9.728e-02
```

```
stargazer(ols2, ols2_clustered_se, type = "text")
```

```
##
## ========================================================================
##                                            Dependent variable:
##                                     ------------------------------------
##                                           log_income
##                                              OLS              coefficient
##                                                                  test
##                                              (1)                 (2)
## ------------------------------------------------------------------------
## year_1984                                 0.024***            0.024***
##                                           (0.002)             (0.004)
##
## education1                                1.552***            1.552***
##                                           (0.049)             (0.172)
##
## education2                                2.331***            2.331***
##                                           (0.050)             (0.174)
##
## relative_year_sd                         -0.072***            -0.072*
##                                           (0.021)             (0.044)
##
## maleMale                                  1.781***            1.781***
##                                           (0.023)             (0.064)
##
## relative_year_squared_sd                  0.067***             0.067**
##                                           (0.020)             (0.031)
##
## relative_year_sd:maleMale                 0.412***            0.412***
##                                           (0.027)             (0.044)
##
## maleMale:relative_year_squared_sd        -0.320***           -0.320***
##                                           (0.027)             (0.039)
##
## Constant                                  5.838***            5.838***
##                                           (0.061)             (0.199)
##
## ------------------------------------------------------------------------
## Observations                               64,997
## R2                                          0.123
## Adjusted R2                                 0.123
## Residual Std. Error                  2.955 (df = 64988)
## F Statistic                     1,137.052*** (df = 8; 64988)
## ========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

```
#Random Intercept
ml1 <- lmer(log_income ~ year_1984 + (1|hh_id) + (1|id_id), data = soep_final, REML = F)
summary(ml1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  305272.6  305318.0 -152631.3  305262.6     64992
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8936 -0.0280  0.0887  0.2887  4.0013
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 3.318    1.821
##  hh_id    (Intercept) 2.587    1.608
##  Residual             5.200    2.280
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 8.048e+00  5.282e-02 1.412e+04  152.37   <2e-16 ***
## year_1984   3.540e-02  1.529e-03 6.155e+04   23.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr)
## year_1984 -0.692
```

```
icc(ml1)
```

```
##                grps   icc
## 1 id_id (Intercept) 0.299
## 2 hh_id (Intercept) 0.233
## 3          Residual 0.468
```

```
#+key level 1 predictor
ml2 <- lmer(log_income ~ year_1984 + married  + (1|hh_id) + (1|id_id), data = soep_final, REML = F)
summary(ml2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + married + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  305255.3  305309.8 -152621.6  305243.3     64991
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8843 -0.0313  0.0897  0.2907  4.0285
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 3.321    1.822
##  hh_id    (Intercept) 2.606    1.614
##  Residual             5.196    2.280
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  8.023e+00  5.316e-02  1.389e+04 150.921  < 2e-16 ***
## year_1984    3.944e-02  1.782e-03  5.798e+04  22.125  < 2e-16 ***
## marriedYes  -1.205e-01  2.736e-02  6.414e+04  -4.404 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) y_1984
## year_1984  -0.644
## marriedYes  0.105 -0.514
```

```
#+ level 1 control variables
ml3 <- lmer(log_income ~ year_1984 + age_sd + age_squared_sd + education + married + (1|hh_id) + (1|id_id), data = soep_fina
l, REML = F)
summary(ml3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  304149.2  304240.1 -152064.6  304129.2     64987
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8412 -0.0429  0.0909  0.2964  4.0239
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 3.355    1.832
##  hh_id    (Intercept) 2.256    1.502
##  Residual             5.125    2.264
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                 Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)     7.789e+00  1.134e-01  1.355e+04  68.685  < 2e-16 ***
## year_1984       1.737e-02  2.436e-03  1.992e+04   7.128 1.05e-12 ***
## age_sd          3.641e-01  5.839e-02  1.310e+04   6.235 4.66e-10 ***
## age_squared_sd -2.217e-01  4.479e-02  1.215e+04  -4.951 7.50e-07 ***
## education1      6.084e-01  9.121e-02  2.051e+04   6.670 2.62e-11 ***
## education2      2.136e+00  9.969e-02  1.849e+04  21.427  < 2e-16 ***
## marriedYes     -2.885e-01  2.994e-02  6.219e+04  -9.637  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 age_sd ag_sq_ edctn1 edctn2
## year_1984   -0.491
## age_sd       0.314 -0.635
## age_sqrd_sd -0.248  0.593 -0.990
## education1  -0.715 -0.040  0.025 -0.033
## education2  -0.670 -0.069  0.010 -0.018  0.861
## marriedYes  -0.145 -0.044 -0.378  0.347 -0.008 -0.029
```

```
#+key level 2 predictor
ml4 <- lmer(log_income ~ year_1984 + age_sd + age_squared_sd + education + married + male + (1|hh_id) + (1|id_id), data = so
ep_final, REML = F)
summary(ml4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + male + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  303309.8  303409.7 -151643.9  303287.8     64986
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8583 -0.0510  0.0866  0.3069  4.0339
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 2.350    1.533
##  hh_id    (Intercept) 2.345    1.531
##  Residual             5.132    2.265
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                 Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)     6.602e+00  1.139e-01  1.151e+04  57.965  < 2e-16 ***
## year_1984       2.283e-02  2.352e-03  1.874e+04   9.707  < 2e-16 ***
## age_sd          2.764e-01  5.549e-02  1.260e+04   4.981 6.42e-07 ***
## age_squared_sd -1.734e-01  4.243e-02  1.185e+04  -4.086 4.42e-05 ***
## education1      7.006e-01  8.715e-02  1.746e+04   8.039 9.65e-16 ***
## education2      2.171e+00  9.508e-02  1.585e+04  22.832  < 2e-16 ***
## marriedYes     -2.611e-01  2.966e-02  6.136e+04  -8.804  < 2e-16 ***
## maleMale        1.698e+00  5.452e-02  3.215e+03  31.140  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 age_sd ag_sq_ edctn1 edctn2 mrrdYs
## year_1984   -0.488
## age_sd       0.314 -0.623
## age_sqrd_sd -0.249  0.583 -0.990
## education1  -0.682 -0.039  0.025 -0.033
## education2  -0.642 -0.068  0.011 -0.020  0.859
## marriedYes  -0.143 -0.062 -0.370  0.340 -0.008 -0.027
## maleMale    -0.329  0.079 -0.050  0.034  0.010  0.020  0.025
```

```
#+ level 2 control variables
ml4a <- lmer(log_income ~ year_1984 + age_sd + age_squared_sd + education + married + male + transition_year + (1|hh_id) +
(1|id_id), data = soep_final, REML = F)
summary(ml4a)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + male + transition_year + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  303299.8  303408.8 -151637.9  303275.8     64985
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8828 -0.0518  0.0873  0.3076  4.0473
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 2.345    1.531
##  hh_id    (Intercept) 2.338    1.529
##  Residual             5.132    2.265
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                  Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)      6.783e+00  1.254e-01  8.453e+03  54.088  < 2e-16 ***
## year_1984        2.726e-02  2.677e-03  3.413e+04  10.183  < 2e-16 ***
## age_sd           2.440e-01  5.624e-02  1.297e+04   4.339 1.45e-05 ***
## age_squared_sd  -1.559e-01  4.270e-02  1.212e+04  -3.652 0.000261 ***
## education1       6.928e-01  8.714e-02  1.746e+04   7.951 1.97e-15 ***
## education2       2.170e+00  9.503e-02  1.584e+04  22.833  < 2e-16 ***
## marriedYes      -2.783e-01  3.008e-02  6.446e+04  -9.254  < 2e-16 ***
## maleMale         1.693e+00  5.448e-02  3.215e+03  31.079  < 2e-16 ***
## transition_year -1.388e-02  4.021e-03  7.127e+03  -3.453 0.000558 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 age_sd ag_sq_ edctn1 edctn2 mrrdYs maleMl
## year_1984   -0.188
## age_sd       0.211 -0.618
## age_sqrd_sd -0.175  0.564 -0.989
## education1  -0.631 -0.047  0.029 -0.036
## education2  -0.584 -0.062  0.012 -0.021  0.859
## marriedYes  -0.198 -0.134 -0.332  0.313 -0.003 -0.026
## maleMale    -0.309  0.058 -0.045  0.031  0.011  0.020  0.028
## transitn_yr -0.420 -0.478  0.166 -0.118  0.028  0.004  0.167  0.024
```

```
#+random slope
ml5 <- lmer(log_income ~ year_1984 + age_sd + age_squared_sd + education + married + male + transition_year + (1| hh_id) +
(1 + married | id_id), data = soep_final, REML = FALSE)
summary(ml5)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + male + transition_year + (1 | hh_id) + (1 + married |      id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  299617.9  299745.0 -149794.9  299589.9     64983
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2342 -0.0477  0.0752  0.2689  4.4872
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id_id    (Intercept) 2.910    1.706
##           marriedYes  3.961    1.990    -0.54
##  hh_id    (Intercept) 1.907    1.381
##  Residual             4.533    2.129
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                   Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)      6.762e+00  1.221e-01  8.062e+03  55.373  < 2e-16 ***
## year_1984        3.234e-02  2.677e-03  2.965e+04  12.077  < 2e-16 ***
## age_sd           2.208e-01  5.384e-02  1.142e+04   4.100 4.16e-05 ***
## age_squared_sd  -1.416e-01  4.070e-02  1.085e+04  -3.479 0.000504 ***
## education1       9.001e-01  8.630e-02  1.409e+04  10.430  < 2e-16 ***
## education2       2.194e+00  9.396e-02  1.297e+04  23.353  < 2e-16 ***
## marriedYes      -2.519e-01  4.144e-02  7.085e+03  -6.080 1.27e-09 ***
## maleMale         1.516e+00  5.200e-02  2.967e+03  29.157  < 2e-16 ***
## transition_year -2.005e-02  3.854e-03  7.223e+03  -5.202 2.02e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 age_sd ag_sq_ edctn1 edctn2 mrrdYs maleMl
## year_1984   -0.199
## age_sd       0.218 -0.637
## age_sqrd_sd -0.183  0.583 -0.989
## education1  -0.641 -0.042  0.029 -0.037
## education2  -0.593 -0.057  0.011 -0.021  0.852
## marriedYes  -0.236 -0.047 -0.245  0.232 -0.001 -0.017
## maleMale    -0.305  0.062 -0.048  0.035  0.011  0.022  0.023
## transitn_yr -0.383 -0.510  0.200 -0.150  0.027  0.001  0.103  0.021
```

```
#+cross-level interaction
ml6 <- lmer(log_income ~  year_1984 + age_sd + age_squared_sd + education + married*male + transition_year + (1| hh_id) + (1
+ married | id_id), data = soep_final, REML = F)
summary(ml6)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married * male + transition_year + (1 | hh_id) + (1 + married |      id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  299350.9  299487.1 -149660.4  299320.9     64982
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2362 -0.0517  0.0728  0.2742  4.4812
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id_id    (Intercept) 2.803    1.674
##           marriedYes  3.621    1.903    -0.53
##  hh_id    (Intercept) 1.888    1.374
##  Residual             4.535    2.130
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                       Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)          7.104e+00  1.233e-01  8.205e+03  57.634  < 2e-16 ***
## year_1984            3.182e-02  2.669e-03  2.977e+04  11.921  < 2e-16 ***
## age_sd               2.253e-01  5.367e-02  1.146e+04   4.197 2.73e-05 ***
## age_squared_sd      -1.444e-01  4.058e-02  1.089e+04  -3.559 0.000374 ***
## education1           9.021e-01  8.597e-02  1.407e+04  10.493  < 2e-16 ***
## education2           2.193e+00  9.360e-02  1.296e+04  23.435  < 2e-16 ***
## marriedYes          -8.886e-01  5.572e-02  5.817e+03 -15.948  < 2e-16 ***
## maleMale             8.753e-01  6.467e-02  3.118e+03  13.535  < 2e-16 ***
## transition_year     -1.959e-02  3.838e-03  7.253e+03  -5.103 3.42e-07 ***
## marriedYes:maleMale  1.209e+00  7.272e-02  4.678e+03  16.625  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 age_sd ag_sq_ edctn1 edctn2 mrrdYs maleMl trnst_
## year_1984   -0.197
## age_sd       0.216 -0.635
## age_sqrd_sd -0.181  0.582 -0.989
## education1  -0.632 -0.042  0.029 -0.037
## education2  -0.585 -0.057  0.011 -0.021  0.852
## marriedYes  -0.282 -0.031 -0.185  0.174 -0.004 -0.012
## maleMale    -0.341  0.055 -0.042  0.031  0.006  0.018  0.425
## transitn_yr -0.377 -0.510  0.199 -0.150  0.027  0.001  0.074  0.013
## mrrdYs:mlMl  0.167 -0.009  0.005 -0.004  0.004 -0.001 -0.686 -0.599  0.005
```

```
#Safe estimated marginal effects
emm <- emmeans(ml6, ~ married * male)
```

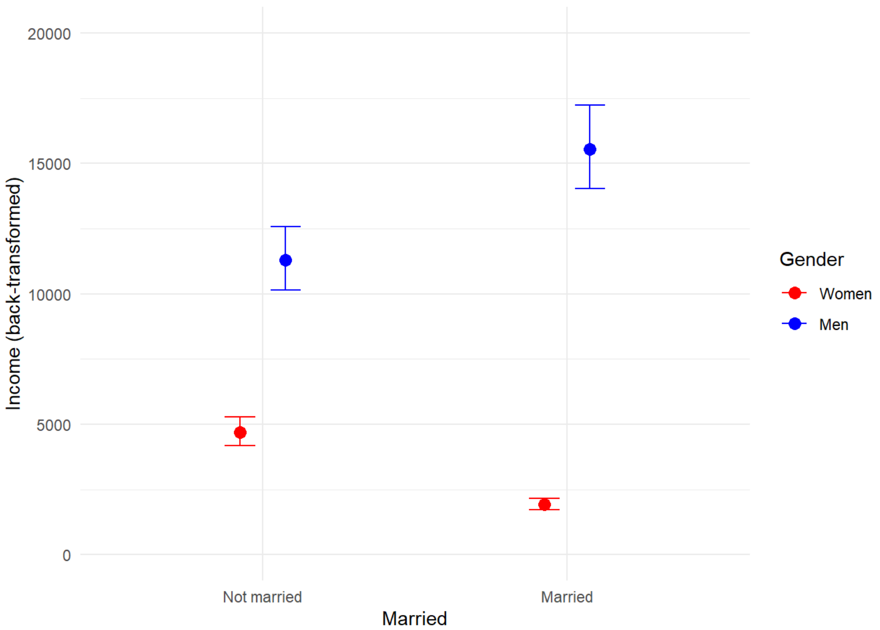
```
## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.
## To enable adjustments, add the argument 'pbkrtest.limit = 64997' (or larger)
## [or, globally, 'set emm_options(pbkrtest.limit = 64997)' or larger];
## but be warned that this may result in large computation time and memory use.
```

```
## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.
## To enable adjustments, add the argument 'lmerTest.limit = 64997' (or larger)
## [or, globally, 'set emm_options(lmerTest.limit = 64997)' or larger];
## but be warned that this may result in large computation time and memory use.
```

```
# Get married effect within each gender:
contrast(emm, method = "pairwise", by = "male")
```

```
## male = Female:
##  contrast estimate     SE  df z.ratio p.value
##  No - Yes    0.889 0.0557 Inf  15.948  <.0001
##
## male = Male:
##  contrast estimate     SE  df z.ratio p.value
##  No - Yes   -0.320 0.0532 Inf  -6.021  <.0001
##
## Results are averaged over the levels of: education
## Degrees-of-freedom method: asymptotic
```

```
#Safe as dataframe
emm_df <- as.data.frame(emm)


#Back-transformed the Log-transformed income (and confidence intervals)
emm_df <- emm_df %>%
  mutate(
    emmean_expo = exp(emmean),
    lower_expo = exp(asymp.LCL),
    upper_expo = exp(asymp.UCL)
  )


#Plot the EMM
ggplot(emm_df, aes(x = factor(married), y = emmean_expo, color = factor(male))) +
  geom_point(position = position_dodge(0.3), size = 3) +
  geom_errorbar(aes(ymin = lower_expo, ymax = upper_expo),
                position = position_dodge(0.3), width = 0.2) +
  scale_y_continuous(limits = c(0, 20000)) +
  labs(
    x = "Married",
    y = "Income (back-transformed)",
    color = "Gender"
  ) +
  scale_color_manual(
    values = c("Female" = "red", "Male" = "blue"),
    labels = c("Women", "Men")
  ) +
  scale_x_discrete(labels = c("Not married", "Married")) +
  theme_minimal()
```




```
#+ key Level 1 predictors
rm1 <- lmer(log_income ~ year_1984 + relative_year_sd + relative_year_squared_sd + (1|hh_id) + (1|id_id), data = soep_final,
REML = F)
summary(rm1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + relative_year_sd + relative_year_squared_sd +
##     (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  305144.1  305207.7 -152565.1  305130.1     64990
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.9129 -0.0316  0.0862  0.2905  3.9826
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 3.305    1.818
##  hh_id    (Intercept) 2.515    1.586
##  Residual             5.195    2.279
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                            Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)               8.486e+00  8.460e-02  1.497e+04 100.300  < 2e-16 ***
## year_1984                 1.777e-02  3.117e-03  2.187e+04   5.701 1.20e-08 ***
## relative_year_sd          1.794e-01  1.963e-02  3.584e+04   9.137  < 2e-16 ***
## relative_year_squared_sd -8.440e-02  1.100e-02  6.277e+04  -7.671 1.74e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 rltv__
## year_1984   -0.894
## reltv_yr_sd  0.747 -0.829
## rltv_yr_sq_  0.181 -0.212 -0.071
```

```
#+level 1 control variables
rm1a <- lmer(log_income ~ year_1984  + education + relative_year_sd + relative_year_squared_sd + (1|hh_id) + (1|id_id), data
= soep_final, REML = F)
summary(rm1a)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  304226.6  304308.4 -152104.3  304208.6     64988
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8932 -0.0342  0.0883  0.2925  3.9481
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 3.380    1.839
##  hh_id    (Intercept) 2.190    1.480
##  Residual             5.135    2.266
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                            Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)               7.573e+00  1.166e-01  1.750e+04  64.958  < 2e-16 ***
## year_1984                 1.200e-02  3.083e-03  2.111e+04   3.893 9.94e-05 ***
## education1                6.213e-01  9.100e-02  2.032e+04   6.828 8.85e-12 ***
## education2                2.103e+00  9.958e-02  1.823e+04  21.123  < 2e-16 ***
## relative_year_sd          1.415e-01  1.946e-02  3.486e+04   7.273 3.59e-13 ***
## relative_year_squared_sd -5.305e-02  1.097e-02  6.283e+04  -4.834 1.34e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__
## year_1984   -0.623
## education1  -0.691 -0.015
## education2  -0.647 -0.045  0.860
## reltv_yr_sd  0.556 -0.820 -0.022 -0.051
## rltv_yr_sq_  0.118 -0.215  0.000  0.051 -0.078
```

```
#+key level 2 predictor
rm2 <- lmer(log_income ~ year_1984  + education + relative_year_sd + relative_year_squared_sd + male + (1|hh_id) + (1|id_i
d), data = soep_final, REML = F)
summary(rm2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  303351.6  303442.4 -151665.8  303331.6     64987
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.9318 -0.0444  0.0837  0.3031  3.9661
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id_id    (Intercept) 2.342    1.530
##  hh_id    (Intercept) 2.296    1.515
##  Residual             5.140    2.267
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                           Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)              6.463e+00  1.164e-01  1.423e+04  55.528  < 2e-16 ***
## year_1984                1.725e-02  2.976e-03  1.841e+04   5.796 6.89e-09 ***
## education1               7.092e-01  8.682e-02  1.731e+04   8.168 3.34e-16 ***
## education2               2.133e+00  9.482e-02  1.563e+04  22.497  < 2e-16 ***
## relative_year_sd         1.143e-01  1.900e-02  3.232e+04   6.017 1.80e-09 ***
## relative_year_squared_sd -5.675e-02  1.096e-02  6.301e+04  -5.179 2.24e-07 ***
## maleMale                 1.724e+00  5.413e-02  3.208e+03  31.859  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__ rlt___
## year_1984   -0.618
## education1  -0.662 -0.013
## education2  -0.621 -0.043  0.858
## reltv_yr_sd  0.544 -0.811 -0.023 -0.050
## rltv_yr_sq_  0.113 -0.208 -0.002  0.048 -0.092
## maleMale    -0.297  0.066  0.007  0.016 -0.058 -0.009
```

```
#+random slope
rm3 <- lmer(log_income ~  year_1984 + education + relative_year_sd + relative_year_squared_sd + male + (1| hh_id) + (1 + rel
ative_year_sd | id_id), data = soep_final, REML = FALSE)
summary(rm3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  296056.6  296165.6 -148016.3  296032.6     64985
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.7466 -0.0522  0.0649  0.2614  5.0445
##
## Random effects:
##  Groups   Name             Variance Std.Dev. Corr
##  id_id    (Intercept)      2.509    1.584
##           relative_year_sd 1.187    1.089    0.12
##  hh_id    (Intercept)      1.858    1.363
##  Residual                  4.113    2.028
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                            Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)               6.428e+00  1.310e-01  8.434e+03  49.084  < 2e-16 ***
## year_1984                 1.425e-02  3.629e-03  6.839e+03   3.926 8.73e-05 ***
## education1                9.154e-01  9.215e-02  1.249e+04   9.935  < 2e-16 ***
## education2                2.198e+00  9.931e-02  1.178e+04  22.131  < 2e-16 ***
## relative_year_sd          1.317e-01  2.512e-02  1.009e+04   5.242 1.62e-07 ***
## relative_year_squared_sd -5.183e-02  1.322e-02  4.095e+04  -3.921 8.83e-05 ***
## maleMale                  1.669e+00  5.384e-02  3.205e+03  30.993  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__ rlt___
## year_1984   -0.689
## education1  -0.636  0.010
## education2  -0.589 -0.028  0.856
## reltv_yr_sd  0.480 -0.637 -0.028 -0.046
## rltv_yr_sq_  0.048 -0.103  0.003  0.060 -0.111
## maleMale    -0.286  0.085  0.010  0.019 -0.057 -0.005
```

```
#+random slope
rm4 <- lmer(log_income ~ year_1984  + education + relative_year_sd + relative_year_squared_sd + male + (1| hh_id) + (1 + rel
ative_year_squared_sd | id_id), data = soep_final, REML = FALSE)
summary(rm4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_squared_sd | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  298523.3  298632.3 -149249.7  298499.3     64985
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.3383 -0.0504  0.0702  0.2631  4.3546
##
## Random effects:
##  Groups   Name                     Variance Std.Dev. Corr
##  id_id    (Intercept)              2.4425   1.5628
##           relative_year_squared_sd 0.8475   0.9206   0.12
##  hh_id    (Intercept)              2.1781   1.4758
##  Residual                          4.3474   2.0850
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                           Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)              6.638e+00  1.246e-01  1.115e+04  53.257  < 2e-16 ***
## year_1984                9.869e-03  3.356e-03  1.082e+04   2.940  0.00329 **
## education1               7.271e-01  9.022e-02  1.444e+04   8.060 8.26e-16 ***
## education2               2.129e+00  9.764e-02  1.337e+04  21.806  < 2e-16 ***
## relative_year_sd         6.544e-02  2.097e-02  2.271e+04   3.121  0.00181 **
## relative_year_squared_sd -5.208e-02  1.739e-02  5.308e+03  -2.994  0.00276 **
## maleMale                 1.717e+00  5.397e-02  3.208e+03  31.811  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__ rlt___
## year_1984   -0.658
## education1  -0.646 -0.005
## education2  -0.600 -0.041  0.860
## reltv_yr_sd  0.542 -0.768 -0.023 -0.047
## rltv_yr_sq_  0.086 -0.118 -0.003  0.030 -0.068
## maleMale    -0.289  0.076  0.008  0.017 -0.063 -0.006
```

```
#+random slope
rm4b <- lmer(log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd + male + (1| hh_id) + (1 + rel
ative_year_sd + relative_year_squared_sd| id_id), data = soep_final, REML = F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00378561 (tol = 0.002, component 1)
```

```
summary(rm4b)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd + relative_year_squared_sd |
##     id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  293856.5  293992.7 -146913.2  293826.5     64982
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.3671 -0.0568  0.0547  0.2327  4.5175
##
## Random effects:
##  Groups   Name                     Variance Std.Dev. Corr
##  id_id    (Intercept)              2.546    1.596
##           relative_year_sd         1.945    1.395     0.26
##           relative_year_squared_sd 1.291    1.136    -0.21 -0.55
##  hh_id    (Intercept)              1.699    1.303
##  Residual                          3.642    1.908
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                           Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)              6.471e+00  1.303e-01  7.811e+03  49.675  < 2e-16 ***
## year_1984                1.184e-02  3.597e-03  6.255e+03   3.293 0.000996 ***
## education1               1.079e+00  9.211e-02  1.121e+04  11.715  < 2e-16 ***
## education2               2.201e+00  9.921e-02  1.062e+04  22.188  < 2e-16 ***
## relative_year_sd         4.987e-02  2.931e-02  6.041e+03   1.701 0.088906 .
## relative_year_squared_sd -5.172e-03  2.186e-02  3.217e+03  -0.237 0.813013
## maleMale                 1.525e+00  5.269e-02  3.072e+03  28.936  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__ rlt___
## year_1984   -0.689
## education1  -0.640  0.013
## education2  -0.591 -0.027  0.851
## reltv_yr_sd  0.429 -0.539 -0.027 -0.043
## rltv_yr_sq_  0.009 -0.071  0.002  0.034 -0.393
## maleMale    -0.284  0.087  0.009  0.021 -0.049 -0.006
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00378561 (tol = 0.002, component 1)
```

```
#+cross-level interaction
rm5 <- lmer(log_income ~ year_1984 + education + relative_year_sd*male + relative_year_squared_sd*male + (1| hh_id) + (1 + r
elative_year_sd + relative_year_squared_sd| id_id), data = soep_final, REML = F)
summary(rm5)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log_income ~ year_1984 + education + relative_year_sd * male +
##     relative_year_squared_sd * male + (1 | hh_id) + (1 + relative_year_sd +
##     relative_year_squared_sd | id_id)
##    Data: soep_final
##
##       AIC       BIC    logLik -2*log(L)  df.resid
##  293576.7  293731.1 -146771.4  293542.7     64980
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.3683 -0.0621  0.0545  0.2380  4.4786
##
## Random effects:
##  Groups   Name                     Variance Std.Dev. Corr
##  id_id    (Intercept)              2.544    1.595
##           relative_year_sd         1.804    1.343     0.26
##           relative_year_squared_sd 1.229    1.109    -0.21 -0.52
##  hh_id    (Intercept)              1.683    1.297
##  Residual                          3.641    1.908
## Number of obs: 64997, groups:  id_id, 4885; hh_id, 4730
##
## Fixed effects:
##                                   Estimate Std. Error         df t value
## (Intercept)                      6.407e+00  1.299e-01  7.876e+03  49.325
## year_1984                        1.091e-02  3.584e-03  6.291e+03   3.045
## education1                       1.078e+00  9.180e-02  1.125e+04  11.744
## education2                       2.197e+00  9.889e-02  1.066e+04  22.219
## relative_year_sd                -3.448e-01  3.807e-02  4.719e+03  -9.056
## maleMale                         1.695e+00  5.364e-02  3.221e+03  31.602
## relative_year_squared_sd         2.923e-01  3.125e-02  3.266e+03   9.356
## relative_year_sd:maleMale        7.568e-01  4.784e-02  3.412e+03  15.818
## maleMale:relative_year_squared_sd -5.647e-01  4.282e-02  3.213e+03 -13.190
##                                 Pr(>|t|)
## (Intercept)                      < 2e-16 ***
## year_1984                        0.00233 **
## education1                       < 2e-16 ***
## education2                       < 2e-16 ***
## relative_year_sd                 < 2e-16 ***
## maleMale                         < 2e-16 ***
## relative_year_squared_sd         < 2e-16 ***
## relative_year_sd:maleMale        < 2e-16 ***
## maleMale:relative_year_squared_sd  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) y_1984 edctn1 edctn2 rltv__ maleMl rlt___ rl__:M
## year_1984   -0.688
## education1  -0.640  0.013
## education2  -0.591 -0.027  0.851
## reltv_yr_sd  0.347 -0.402 -0.023 -0.032
## maleMale    -0.285  0.083  0.010  0.020 -0.159
## rltv_yr_sq_ -0.019 -0.047  0.005  0.024 -0.432  0.096
## rltv_yr_s:M -0.029 -0.018  0.004 -0.002 -0.656  0.186  0.357
## mlMl:rlt___  0.035 -0.004 -0.006 -0.001  0.328 -0.137 -0.726 -0.488
```

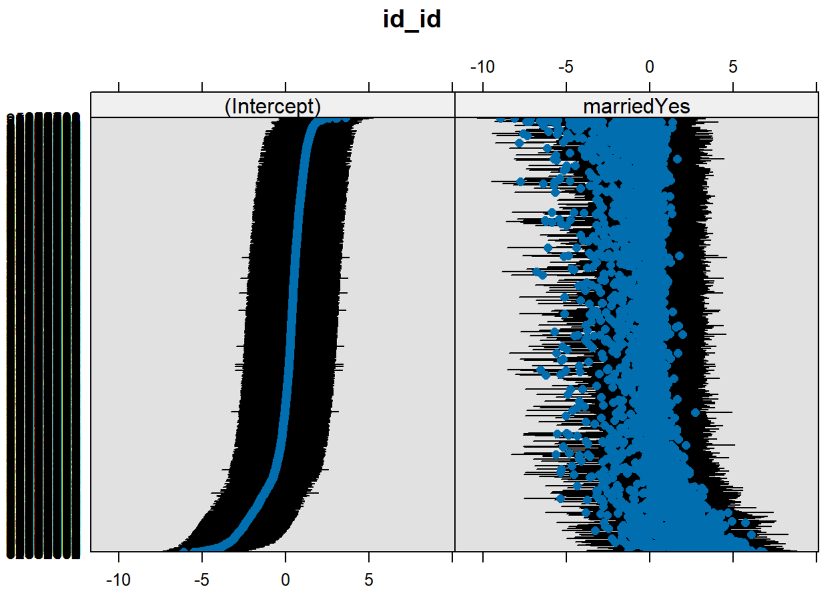
```
#SD for relative year
sd(soep_final$relative_year)
```

```
## [1] 8.059238
```

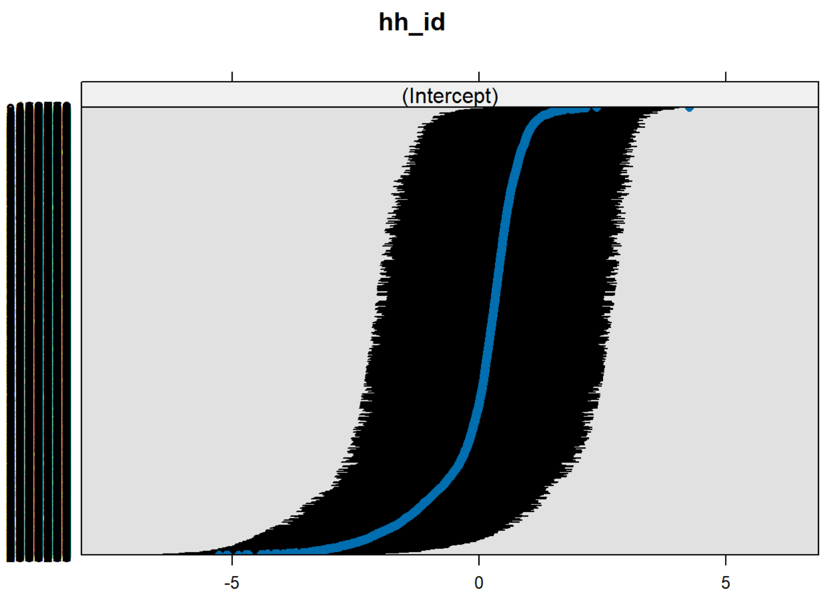
```
# # Create prediction data
# new_data <- expand.grid(
#   relative_year_sd = seq(-1.5, 1.5, by = 0.1),
#   male = factor(c("Female", "Male")),
#   year_1984 = c(0),
#   education = "0"
# )

# new_data$relative_year_squared_sd <- new_data$relative_year_sd^2

# # Ensure factors match model
# new_data$male <- factor(new_data$male, levels = levels(soep_final$male))
# new_data$education <- factor(new_data$education, levels = levels(soep_final$education))

# # Predict with standard errors
# pred <- predict(rm5, newdata = new_data, re.form = NA, se.fit = TRUE)

# # Add predicted values and 95% CI
# new_data$predicted_log_income <- pred$fit
# new_data$se <- pred$se.fit
# new_data$lower <- new_data$predicted_log_income - 1.96 * new_data$se
# new_data$upper <- new_data$predicted_log_income + 1.96 * new_data$se

# # Back-transform log-transformed income variable (and confidence intervals)
# new_data$predicted_income <- exp(new_data$predicted_log_income)
# new_data$lower_income <- exp(new_data$lower)
# new_data$upper_income <- exp(new_data$upper)

# # Plot the trajectories (back-transformed logged income)
# ggplot(new_data, aes(x = relative_year_sd, y = predicted_income, color = male, fill = male)) +
#   geom_line(size = 1.2) +
#   geom_ribbon(aes(ymin = lower_income, ymax = upper_income), alpha = 0.2, color = NA) +
#   labs(
#     title = "Predicted Income Trajectories Over Time",
#     x = "Standardized Relative Year",
#     y = "Predicted Income",
#     color = "Gender",
#     fill = "Gender"
#   ) +
#   scale_color_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   scale_fill_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   theme_minimal(base_size = 14) +
#   theme(
#     axis.text.y = element_blank(),
#     axis.ticks.y = element_blank())

# # Predict the trajectories (logged income)
# ggplot(new_data, aes(x = relative_year_sd, y = predicted_log_income, color = male, fill = male)) +
#   geom_line(size = 1.2) +
#   geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2, color = NA) +
#   labs(
#     title = "Predicted Log Income Trajectories Over Time",
#     x = "Standardized Relative Year",
#     y = "Predicted Log Income",
#     color = "Gender",
#     fill = "Gender"
#   ) +
#   scale_color_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   scale_fill_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   theme_minimal(base_size = 14)

# ggplot(new_data, aes(x = relative_year_sd, y = predicted_income, color = male, fill = male)) +
#   geom_line(size = 1.2) +
#   geom_ribbon(aes(ymin = lower_income, ymax = upper_income), alpha = 0.2, color = NA) +
#   labs(
#     title = "Predicted Income Trajectories Over Time",
#     x = "Standardized Relative Year",
#     y = "Predicted Income",
#     color = "Gender",
#     fill = "Gender"
#   ) +
#   scale_color_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   scale_fill_manual(values = c("red", "blue"), labels = c("Women", "Men")) +
#   theme_minimal(base_size = 14) +
#   theme_minimal(base_size = 14)
```

```
#Inspect variance components
dotplot(ranef(ml5))
```

```
## $id_id
```

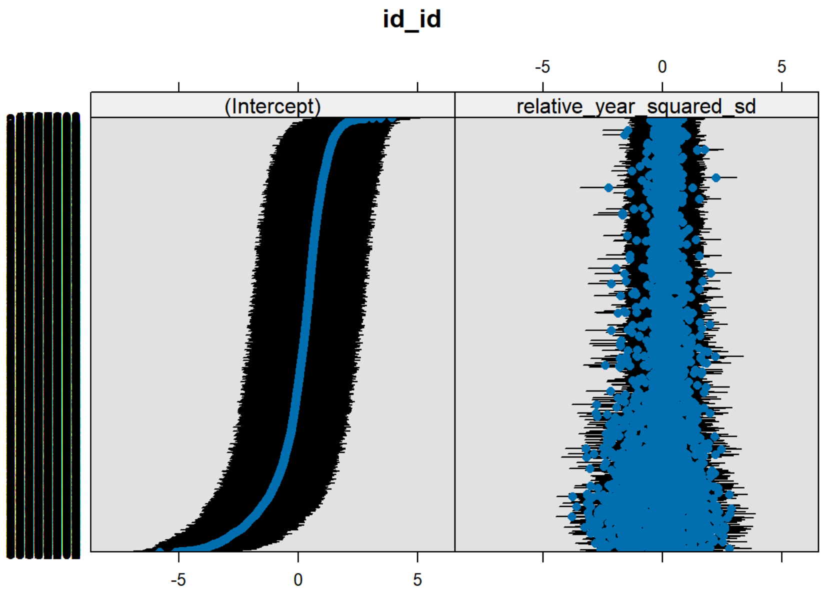



```
##
## $hh_id
```

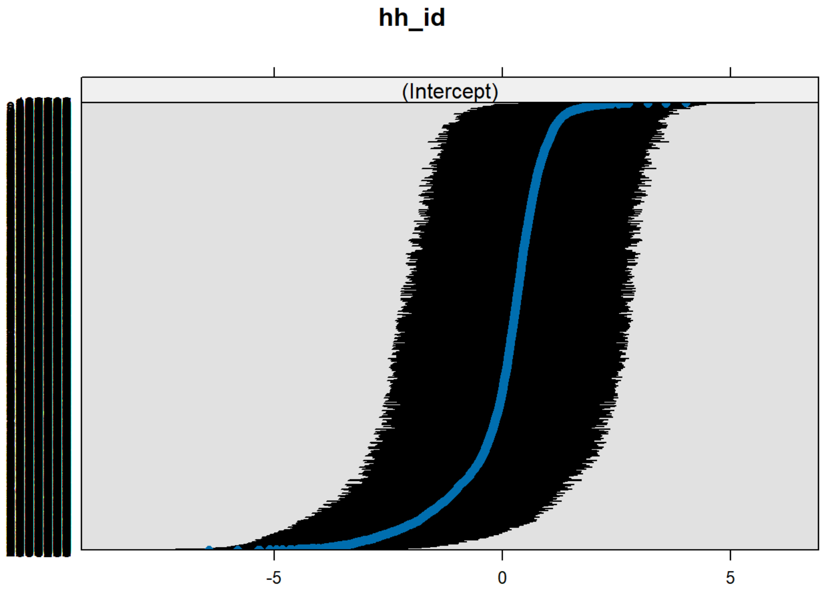



```
dotplot(ranef(rm4))
```

```
## $id_id
```

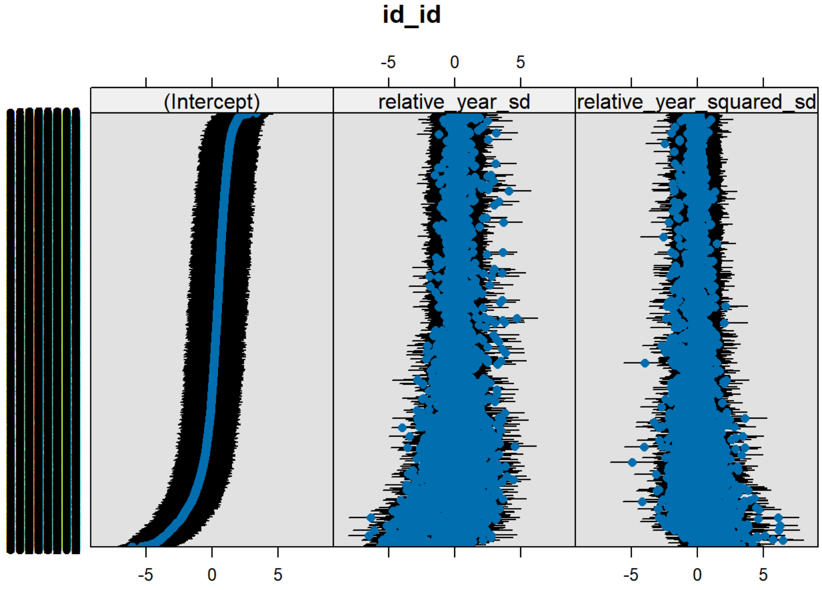



```
##
## $hh_id
```

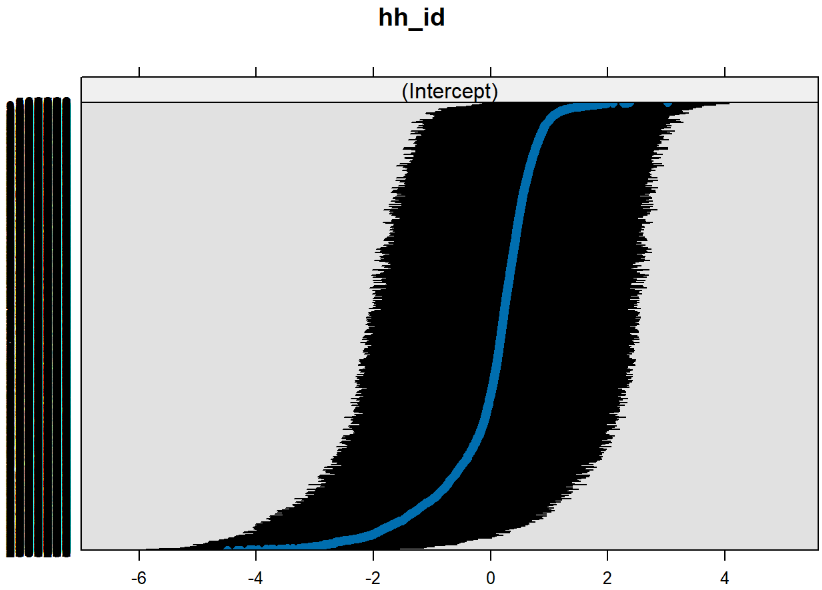



```
dotplot(ranef(rm4b))
```

```
## $id_id
```




```
##
## $hh_id
```




```
#Likelihood-Ratio Tests to compare the models with and without the random slopes
lrtest(ml5, ml4a)
```

```
## Likelihood ratio test
##
## Model 1: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + male + transition_year + (1 | hh_id) + (1 + married |
##     id_id)
## Model 2: log_income ~ year_1984 + age_sd + age_squared_sd + education +
##     married + male + transition_year + (1 | hh_id) + (1 | id_id)
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1  14 -149795
## 2  12 -151638 -2  3686  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(rm2, rm3)
```

```
## Likelihood ratio test
##
## Model 1: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 | id_id)
## Model 2: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd | id_id)
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1  10 -151666
## 2  12 -148016  2  7299  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(rm2, rm4)
```

```
## Likelihood ratio test
##
## Model 1: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 | id_id)
## Model 2: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_squared_sd | id_id)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1  10 -151666
## 2  12 -149250  2 4832.3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(rm3, rm4b)
```

```
## Likelihood ratio test
##
## Model 1: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd | id_id)
## Model 2: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd + relative_year_squared_sd |
##     id_id)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1  12 -148016
## 2  15 -146913  3 2206.1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(rm4, rm4b)
```

```
## Likelihood ratio test
##
## Model 1: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_squared_sd | id_id)
## Model 2: log_income ~ year_1984 + education + relative_year_sd + relative_year_squared_sd +
##     male + (1 | hh_id) + (1 + relative_year_sd + relative_year_squared_sd |
##     id_id)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1  12 -149250
## 2  15 -146913  3 4672.8  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Variance in the effect of marriage status on income explained by gender
((as.data.frame(VarCorr(ml5))[2,4])-(as.data.frame(VarCorr(ml6))[2,4])) /
(as.data.frame(VarCorr(ml5))[2,4])
```

```
## [1] 0.08591941
```

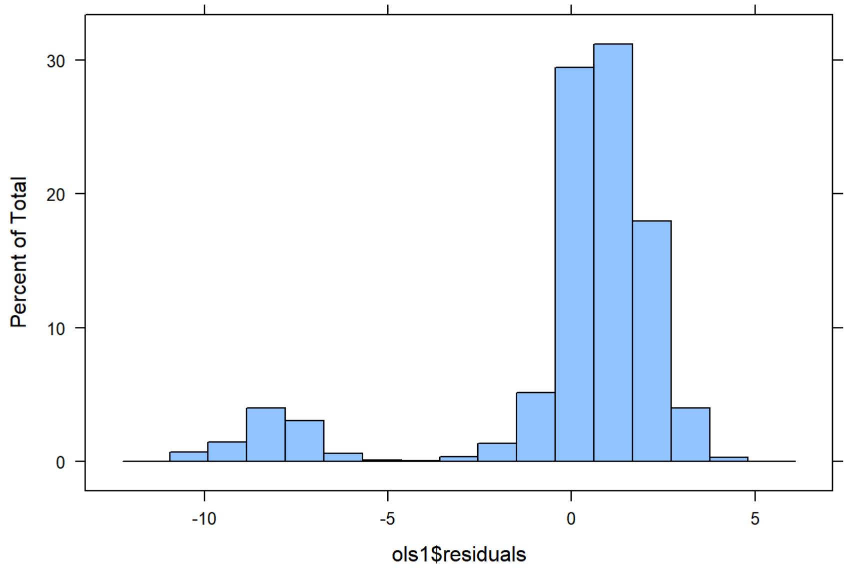
```
# Extract variance components as data frames
vc_ml5a <- as.data.frame(VarCorr(ml5))
vc_ml6 <- as.data.frame(VarCorr(ml6))

# Extract the random slope variance for 'marriedYes' at the individual level ("id_id")
slope_var_ml5a <- vc_ml5a[vc_ml5a$grp == "id_id" & vc_ml5a$var1 == "marriedYes" & is.na(vc_ml5a$var2), "vcov"]
slope_var_ml6 <- vc_ml6[vc_ml6$grp == "id_id" & vc_ml6$var1 == "marriedYes" & is.na(vc_ml6$var2), "vcov"]

# Calculate proportional reduction in variance
reduction_married <- (slope_var_ml5a - slope_var_ml6) / slope_var_ml5a

# Show result
reduction_married
```

```
## [1] 0.08591941
```

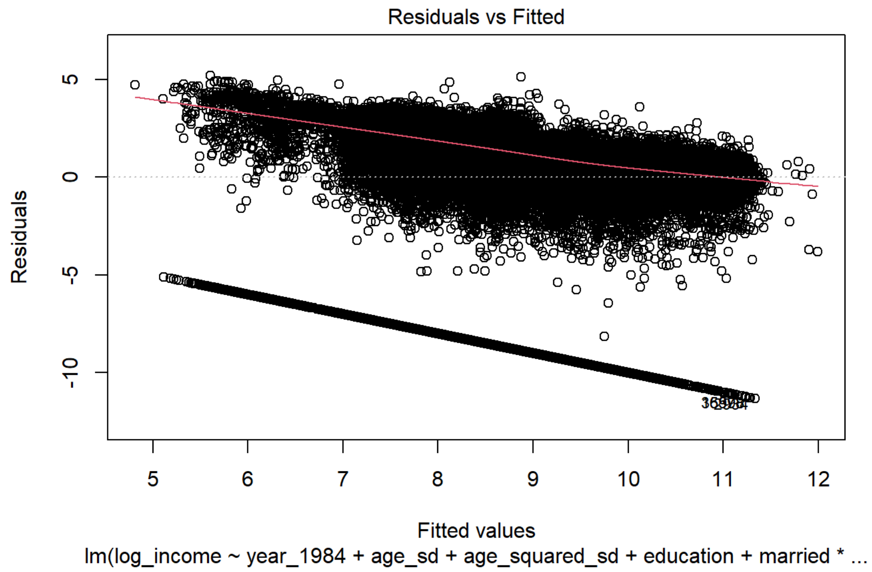
```
# Extract random effect variance components as data frames
vc_rm4b <- as.data.frame(VarCorr(rm4b))
vc_rm5 <- as.data.frame(VarCorr(rm5))

# Filter to individual-level random slopes (usually group = "id_id")
slope_var_ry <- vc_rm4b[vc_rm4b$grp == "id_id" & vc_rm4b$var1 == "relative_year_sd" & is.na(vc_rm4b$var2), "vcov"]
slope_var_ry2 <- vc_rm4b[vc_rm4b$grp == "id_id" & vc_rm4b$var1 == "relative_year_squared_sd" & is.na(vc_rm4b$var2), "vcov"]

slope_var_ry_new <- vc_rm5[vc_rm5$grp == "id_id" & vc_rm5$var1 == "relative_year_sd" & is.na(vc_rm5$var2), "vcov"]
slope_var_ry2_new <- vc_rm5[vc_rm5$grp == "id_id" & vc_rm5$var1 == "relative_year_squared_sd" & is.na(vc_rm5$var2), "vcov"]

# Compute proportional reduction in variance for both slopes
reduction_ry <- (slope_var_ry - slope_var_ry_new) / slope_var_ry
reduction_ry2 <- (slope_var_ry2 - slope_var_ry2_new) / slope_var_ry2

reduction_ry + reduction_ry2
```

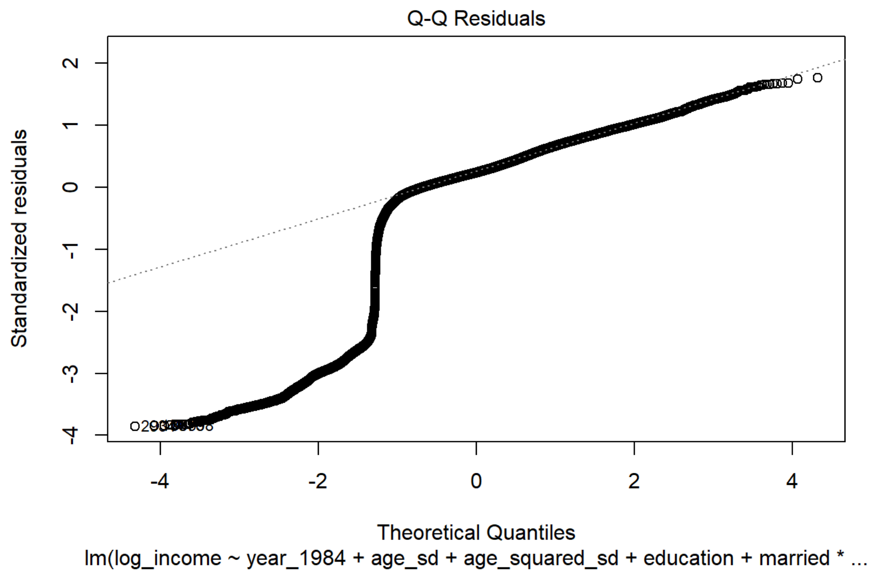
```
## [1] 0.1204257
```

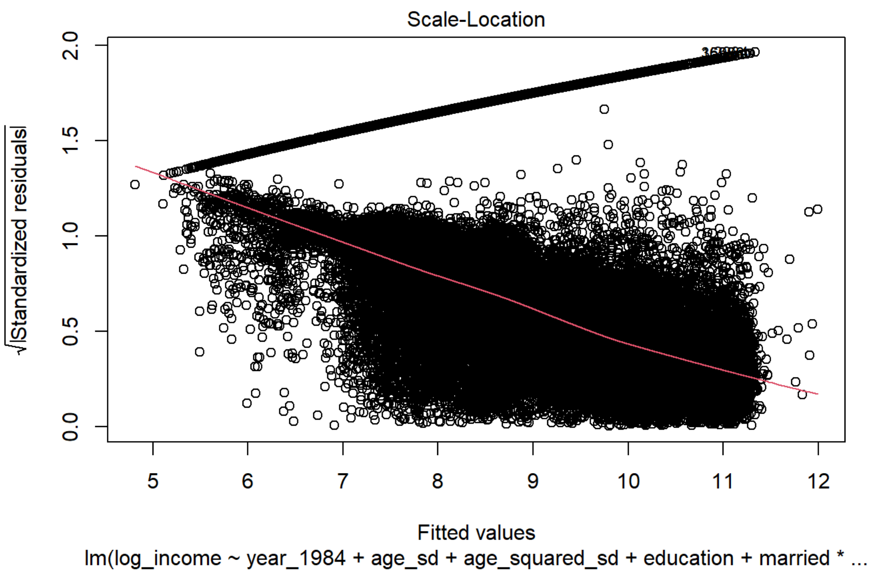
```
#Assumptions OLS

histogram(ols1$residuals)
```

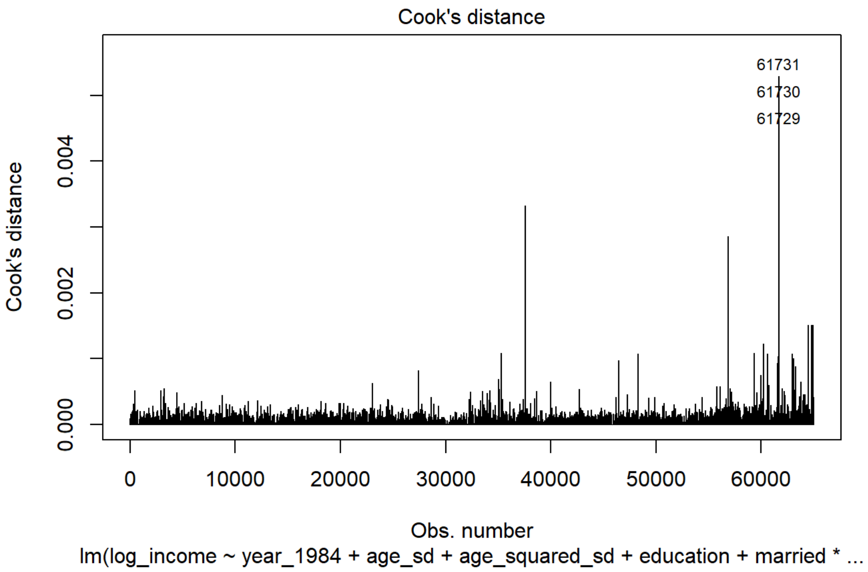


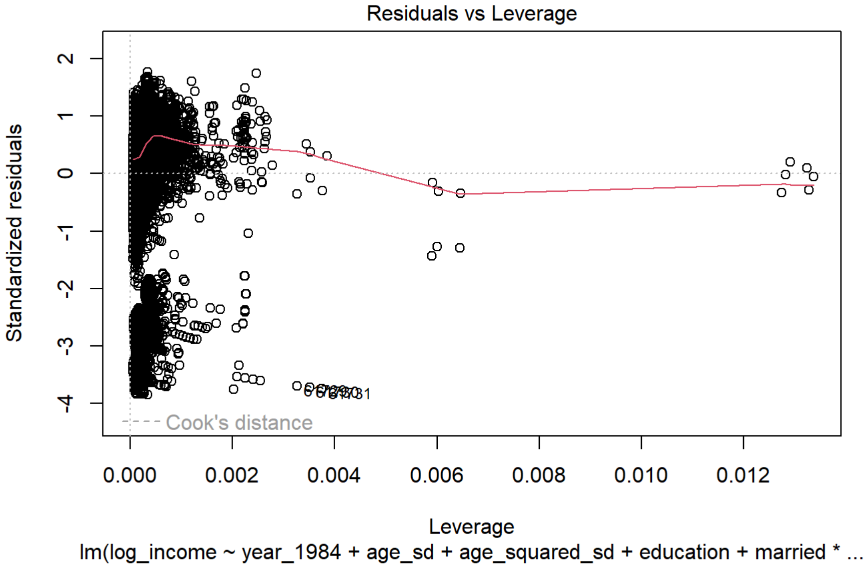

```
plot(ols1, which = 1:5)
```


Residuals vs Fitted
Residuals
5
0
-5
-10
5 6 7 8 9 10 11 12
Fitted values
lm(log_income ~ year_1984 + age_sd + age_squared_sd + education + married * ...


Q-Q Residuals
Standardized residuals
2 1 0 -1 -2 -3 -4
-4 -2 0 2 4
Theoretical Quantiles
lm(log_income ~ year_1984 + age_sd + age_squared_sd + education + married * ...


Scale-Location
√|Standardized residuals|
2.0 1.5 1.0 0.5 0.0
5 6 7 8 9 10 11 12
Fitted values
lm(log_income ~ year_1984 + age_sd + age_squared_sd + education + married * ...





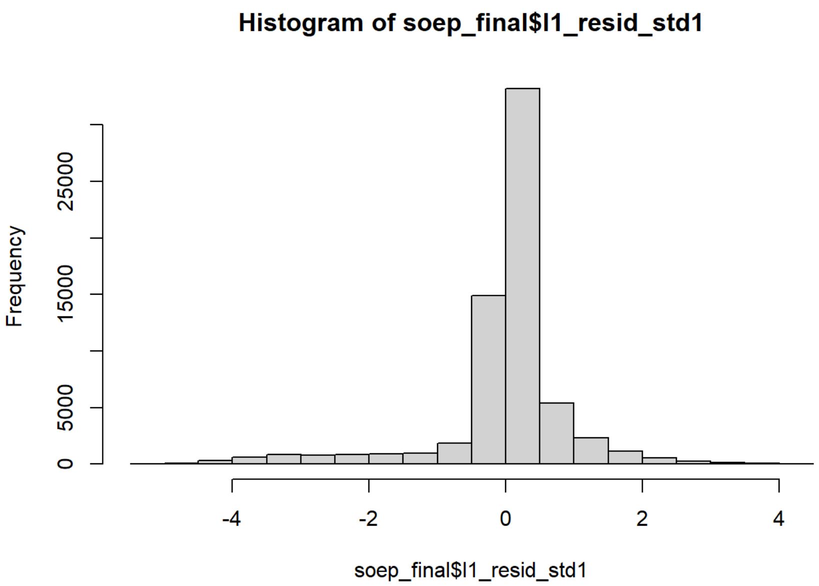
#Assumptions multilevel models

#Model 1

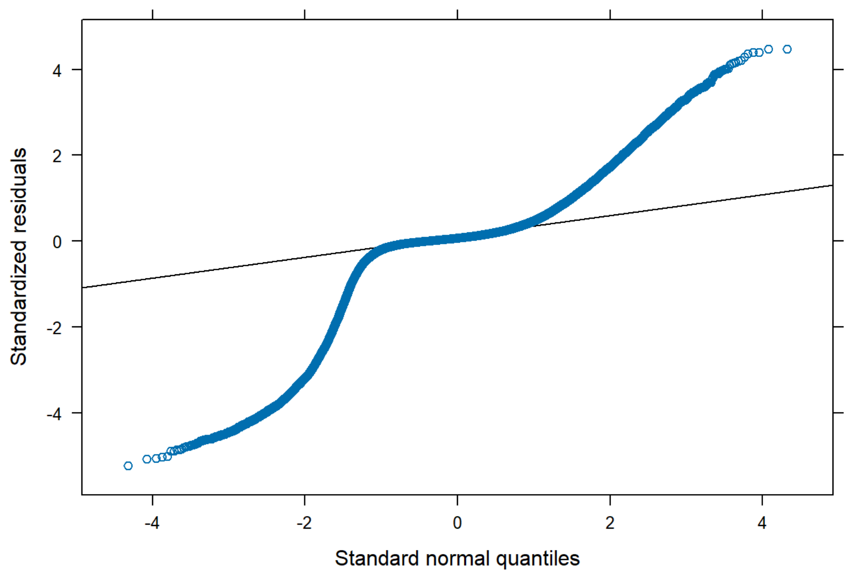
#Level 1

#Normality

```
#Level 1 residuals

#Extract L1 residuals
soep_final$l1resid <- residuals(ml6)

#Extract standardized residuals
soep_final$l1_resid_std1 <- resid(ml6, type = "pearson", scale = TRUE)

#Histogram L1 residuals
hist(soep_final$l1_resid_std1)
```

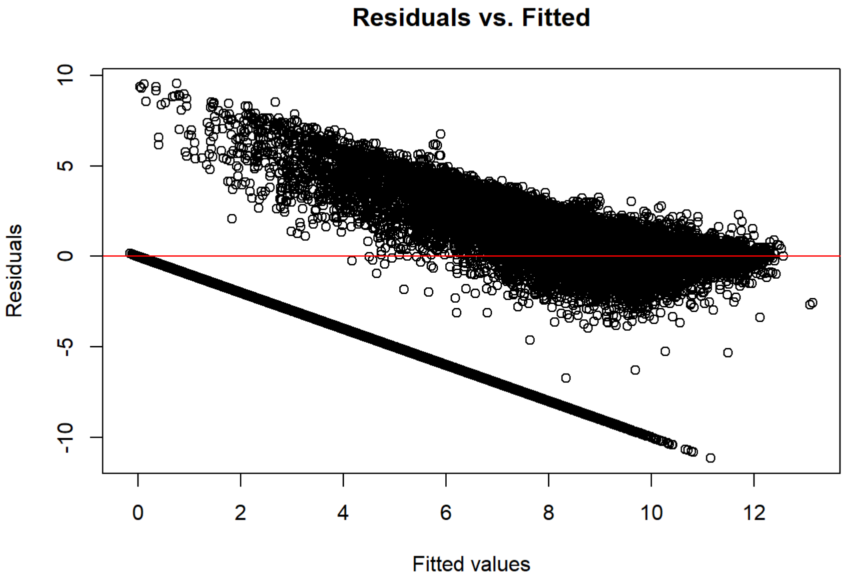



```
#QQ Plot
qqmath(ml6)
```



#Homoskedasticity

```
plot(fitted(ml6), resid(ml6),
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs. Fitted")
abline(h = 0, col = "red")
```

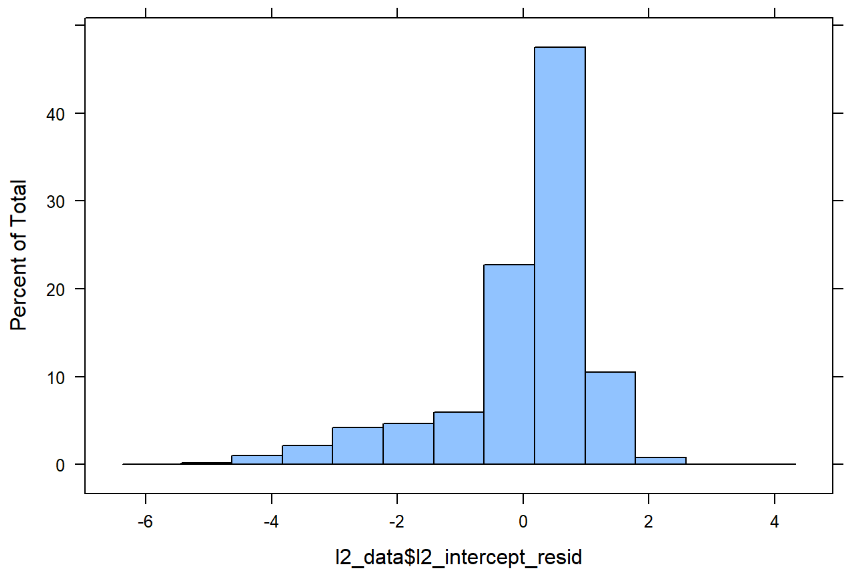


#Level 2

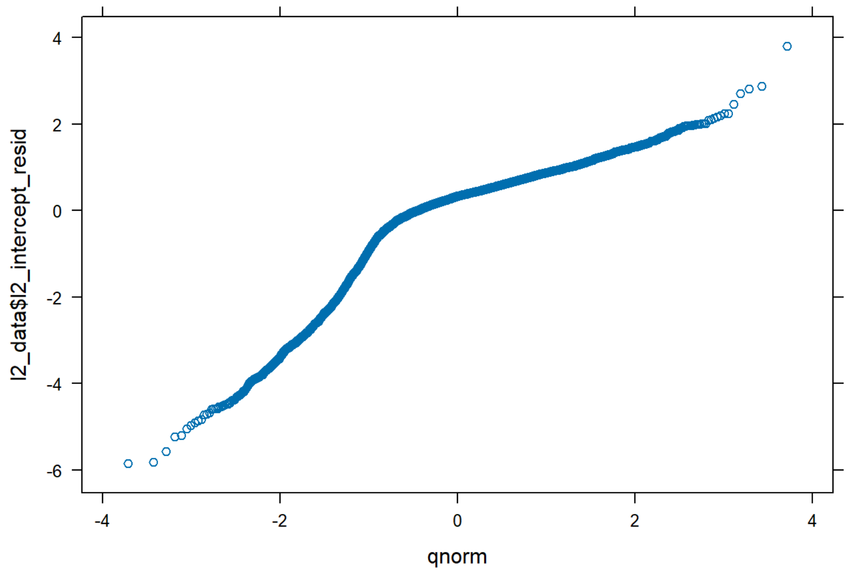
```
#Level 2 residuals

l2_data <- soep_final %>%
group_by(id_id) %>%
mutate(log_income_mean = mean(log_income, na.rm = T)) %>%
dplyr::select(id_id, log_income_mean, male) %>%
unique()

#Random Intercept and Random Slope residuals
l2_data$l2_intercept_resid = ranef(ml6)$id_id [, 1]
l2_data$l2_slope_resid = ranef(ml6)$id_id [, 2]
```

#Normality

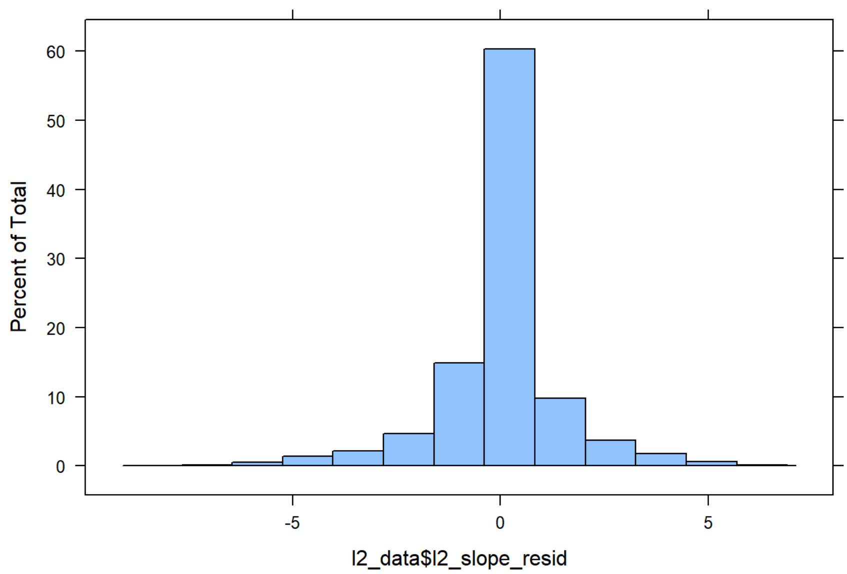
```
#Histogram
histogram(l2_data$l2_intercept_resid)
```



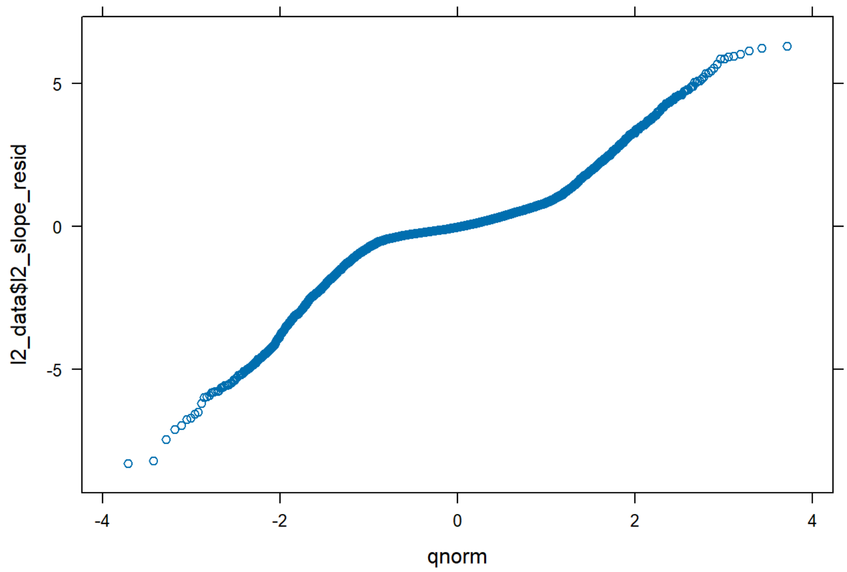

```
#QQ Plot
qqmath(l2_data$l2_intercept_resid)
```




```
#Histogram
histogram(l2_data$l2_slope_resid)
```



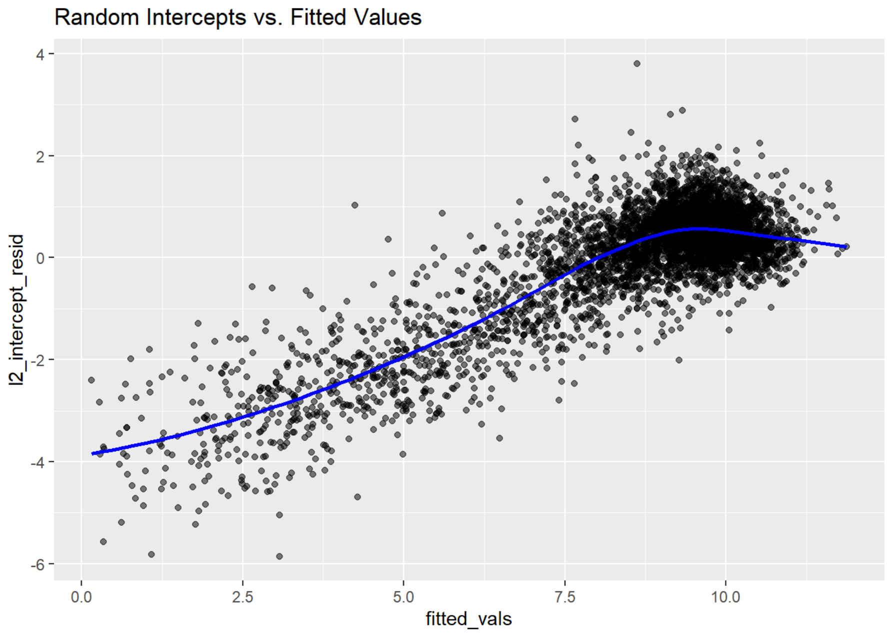

```
#QQ Plot
qqmath(l2_data$l2_slope_resid)
```



#Homoskedasticity

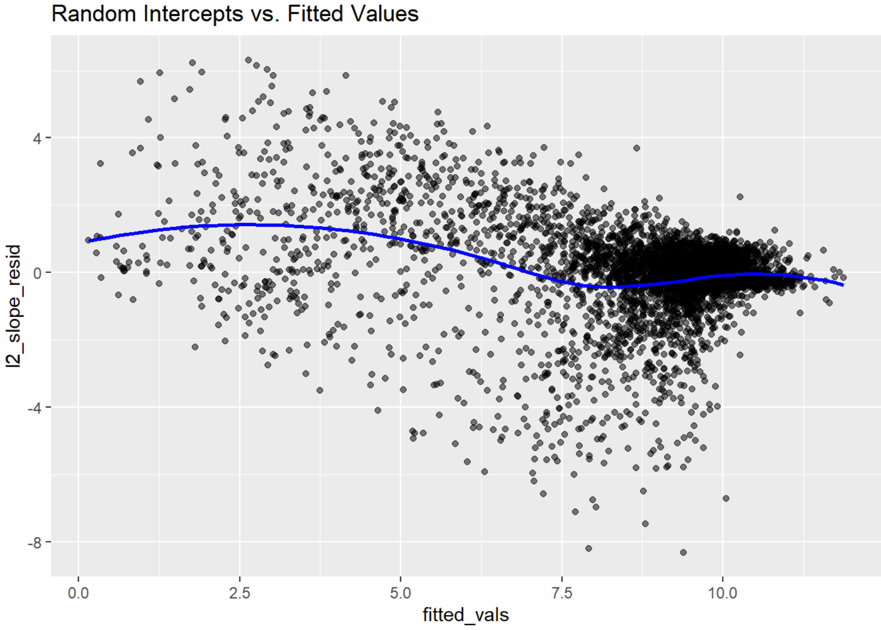
```
# Random intercepts
l2_data$fitted_vals <- fitted(ml6)[match(l2_data$id_id, soep_final$id_id)]

ggplot(l2_data, aes(x = fitted_vals, y = l2_intercept_resid)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Random slopes
ggplot(l2_data, aes(x = fitted_vals, y = l2_slope_resid)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



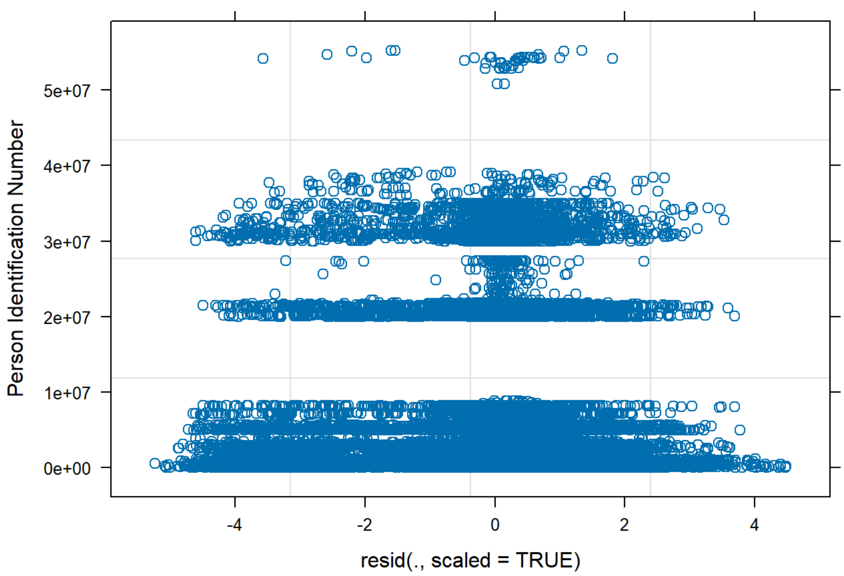

```
#Check whether the variance of the L1 residual errors is the same in all groups
soep_final$abs.l1resid<-soep_final$l1residuals
```

```
## Warning: Unknown or uninitialised column: `l1residuals`.
```

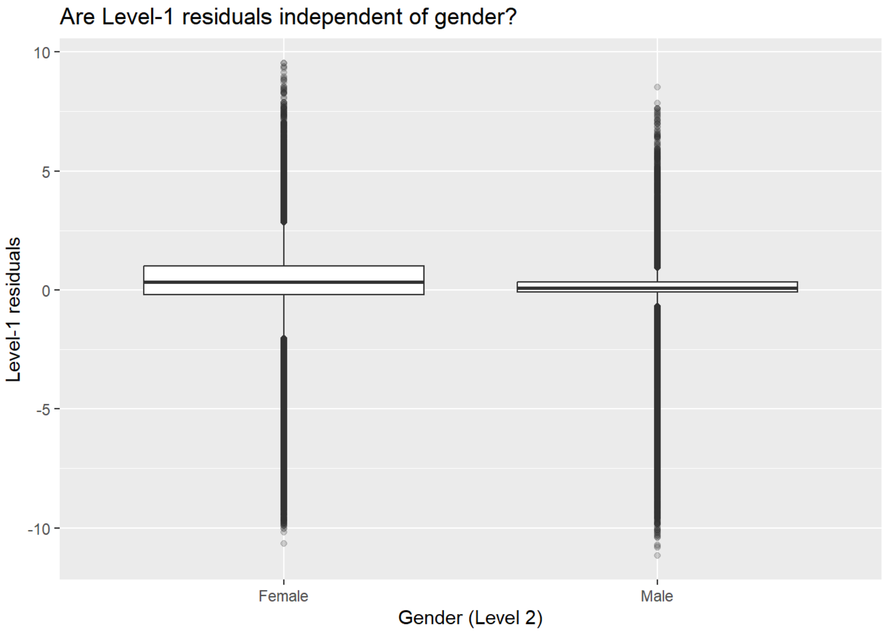
```
soep_final$abs.l1resid <- abs(soep_final$l1resid)
soep_final$abs.l1resid2 <- soep_final$abs.l1resid^2

soep_final$abs.l1resid2<-soep_final$abs.l1resid^2


Levene.Model.F <- lm(abs.l1resid2 ~ id_id, data=soep_final)
#ANOVA of the squared residuals
anova(Levene.Model.F) #displays the results
```

```
## Analysis of Variance Table
##
## Response: abs.l1resid2
##              Df  Sum Sq Mean Sq F value Pr(>F)
## id_id         1       9   9.196  0.0663 0.7968
## Residuals 64995 9014157 138.690
```

```
plot(ml6, id_id ~ resid(., scaled=TRUE))
```



```
# Independence of L1 residuals from level 2 predictors

soep_final$l1resid <- resid(ml6)

ggplot(soep_final, aes(x = male, y = l1resid)) +
  geom_boxplot(outlier.alpha = 0.2) +
  labs(x = "Gender (Level 2)",
       y = "Level-1 residuals",
       title = "Are Level-1 residuals independent of gender?")
```




```
leveneTest(l1resid ~ male, data = soep_final)
```

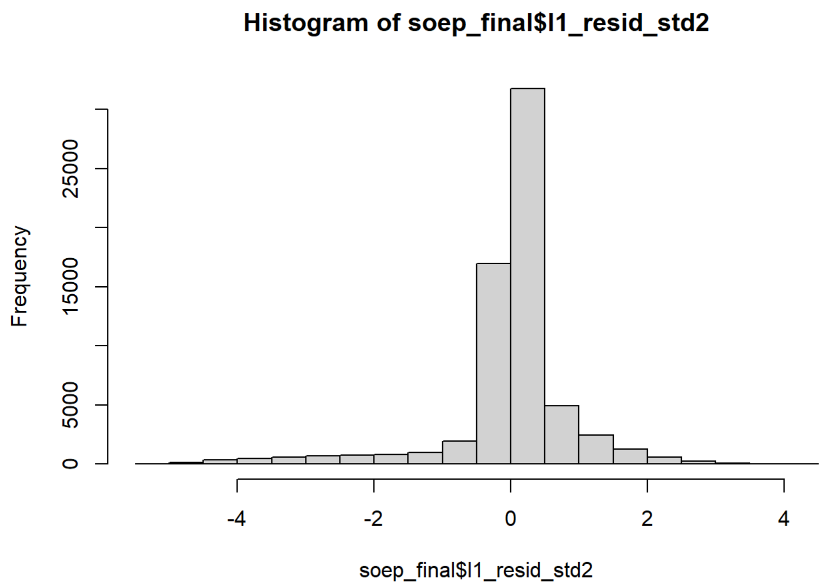
```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group     1  4384.9 < 2.2e-16 ***
##       64995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Model 2

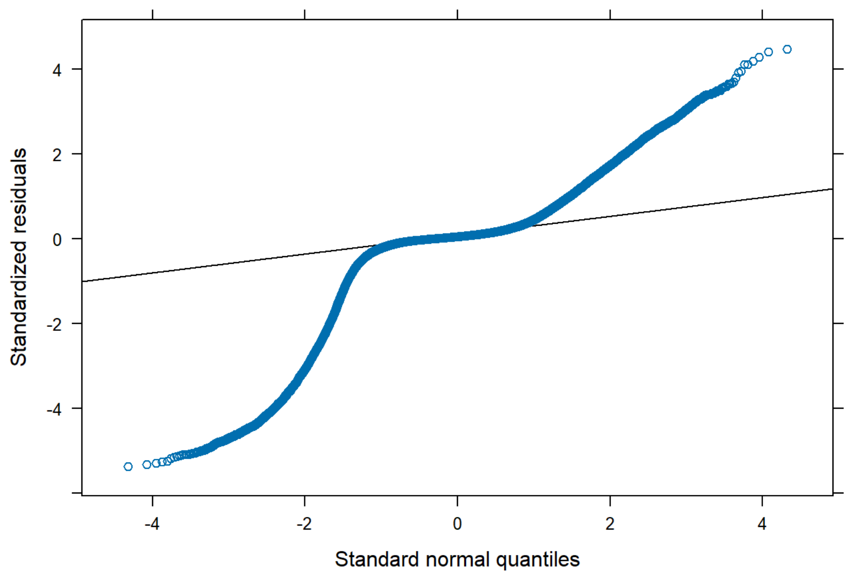
#Level 1

#Normality

```
#Level 1 residuals

#Extract L1 residuals
soep_final$l1resid1 <- residuals(rm5)

#Extract standardized residuals
soep_final$l1_resid_std2 <- resid(rm5, type = "pearson", scale = TRUE)

#Histogram L1 residuals
hist(soep_final$l1_resid_std2)
```

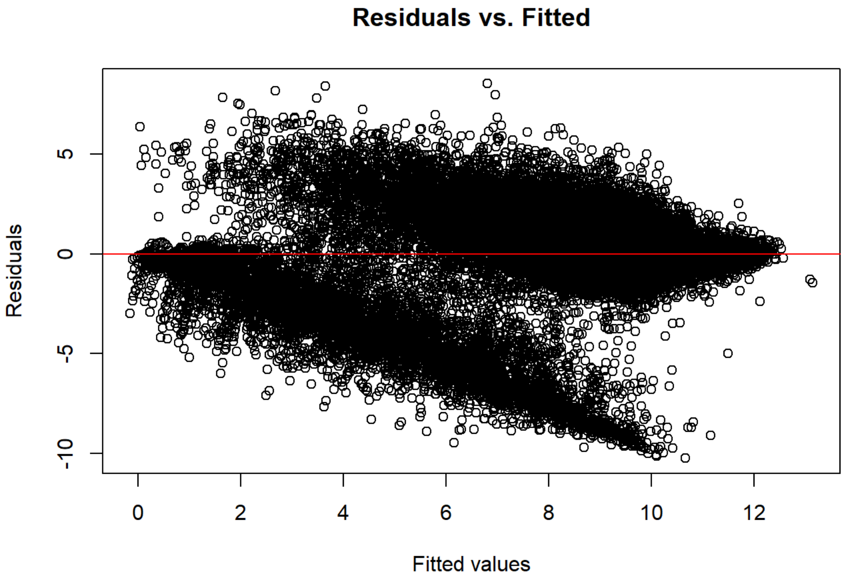



```
#QQ Plot
qqmath(rm5)
```



#Homoskedasticity

```
plot(fitted(ml6), resid(rm5),
     xlab = "Fitted values", ylab = "Residuals",
     main = "Residuals vs. Fitted")
abline(h = 0, col = "red")
```

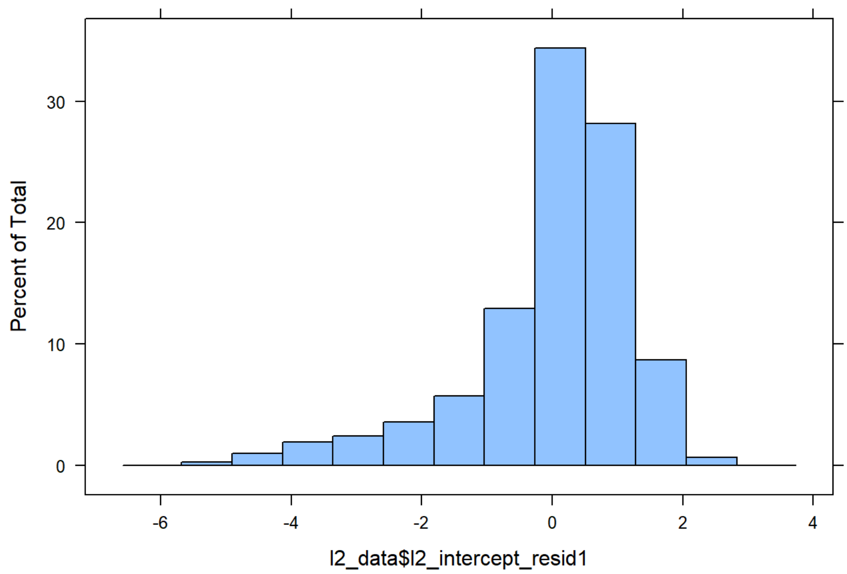


#Level 2

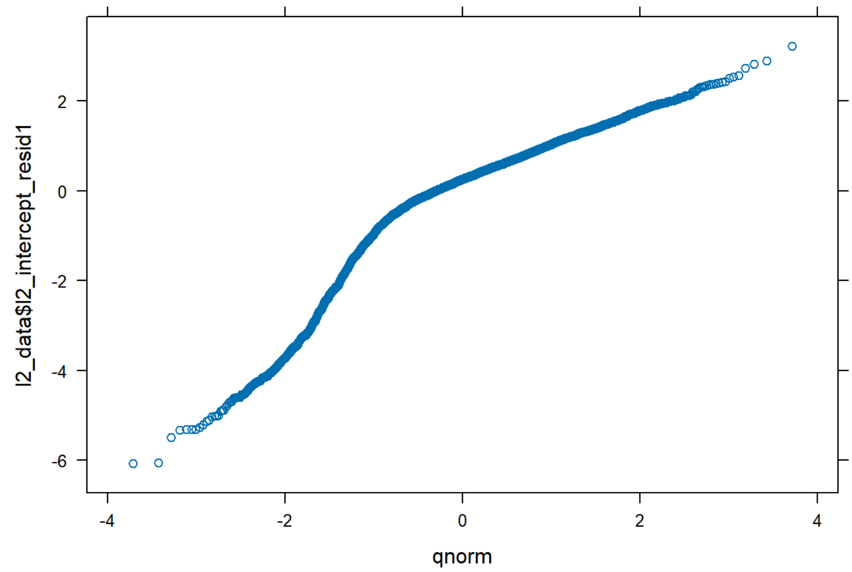
```
#Level 2 residuals

l2_data <- soep_final %>%
group_by(id_id) %>%
mutate(log_income_mean = mean(log_income, na.rm = T)) %>%
dplyr::select(id_id, log_income_mean, male) %>%
unique()

#Random Intercept and Random Slope residuals
l2_data$l2_intercept_resid1 = ranef(rm5)$id_id [, 1]
l2_data$l2_slope_resid1 = ranef(rm5)$id_id [, 2]
l2_data$l2_slope_resid2 = ranef(rm5)$id_id[, 3]
```

#Normality

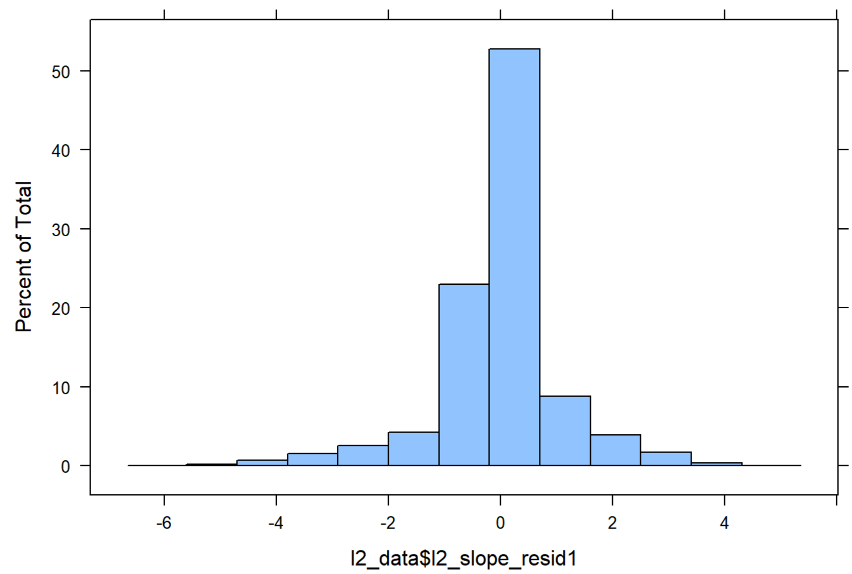
```
#Histogram
histogram(l2_data$l2_intercept_resid1)
```



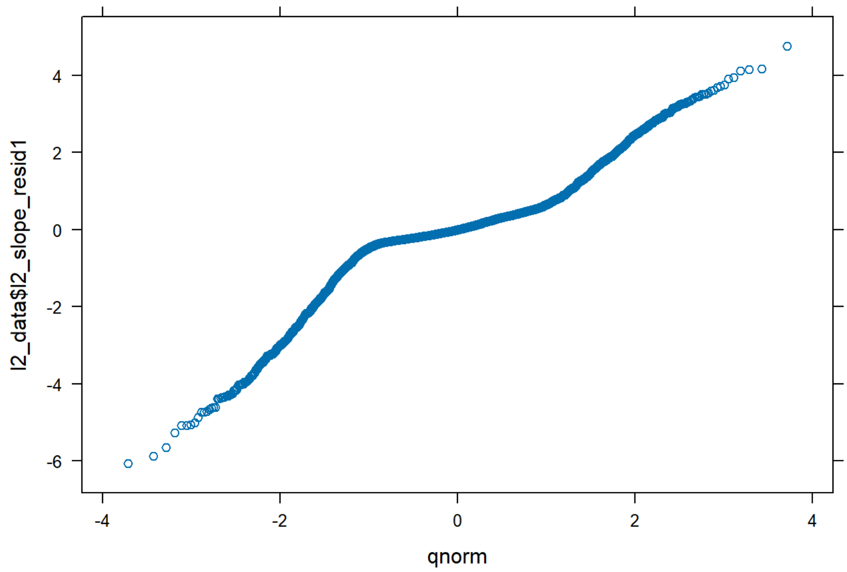

```
#QQ Plot
qqmath(l2_data$l2_intercept_resid1)
```



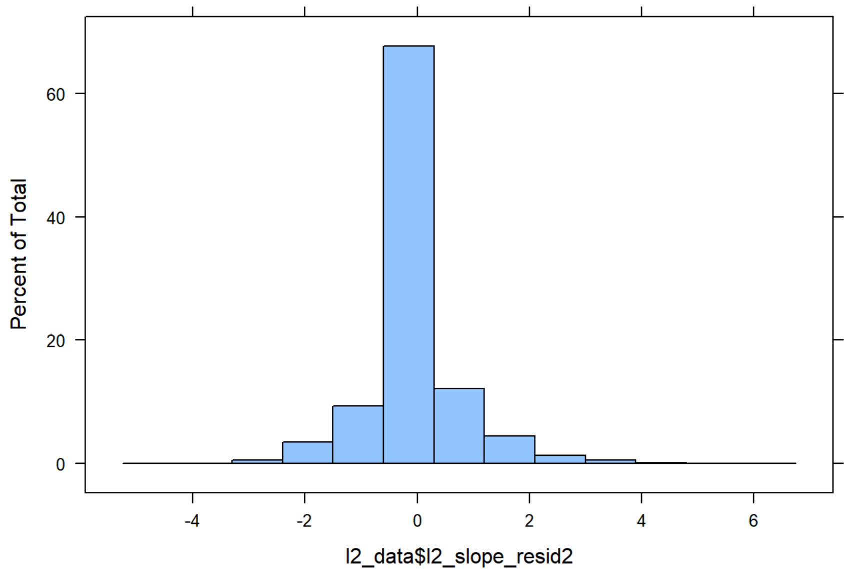

```
#Histogram
histogram(l2_data$l2_slope_resid1)
```



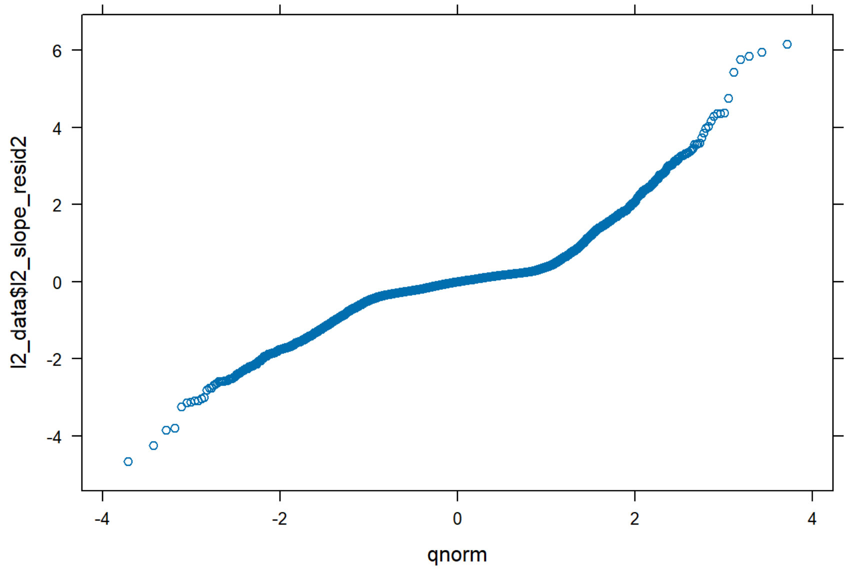

```
#QQ Plot
qqmath(l2_data$l2_slope_resid1)
```




```
#Histogram
histogram(l2_data$l2_slope_resid2)
```



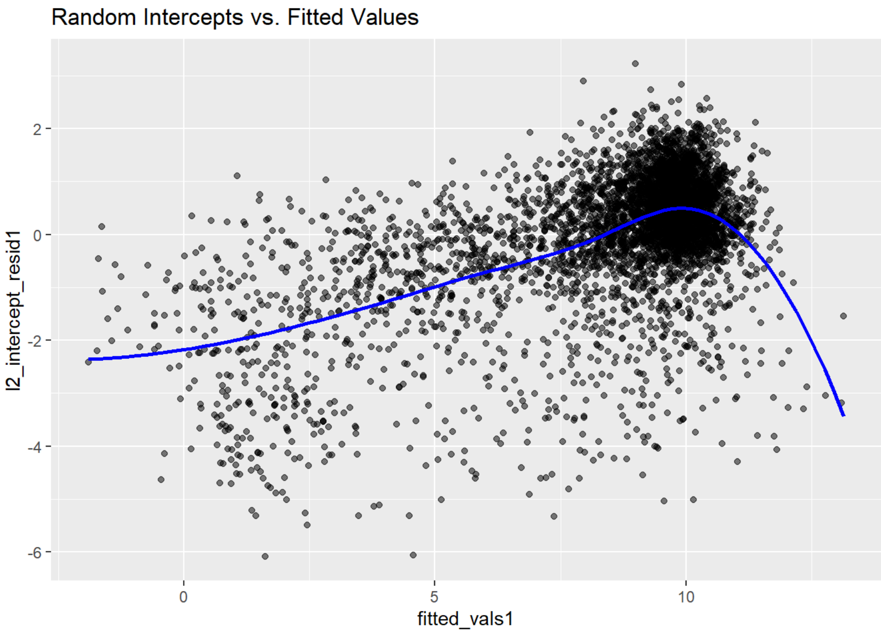

```
#QQ Plot
qqmath(l2_data$l2_slope_resid2)
```



#Homoskedasticity

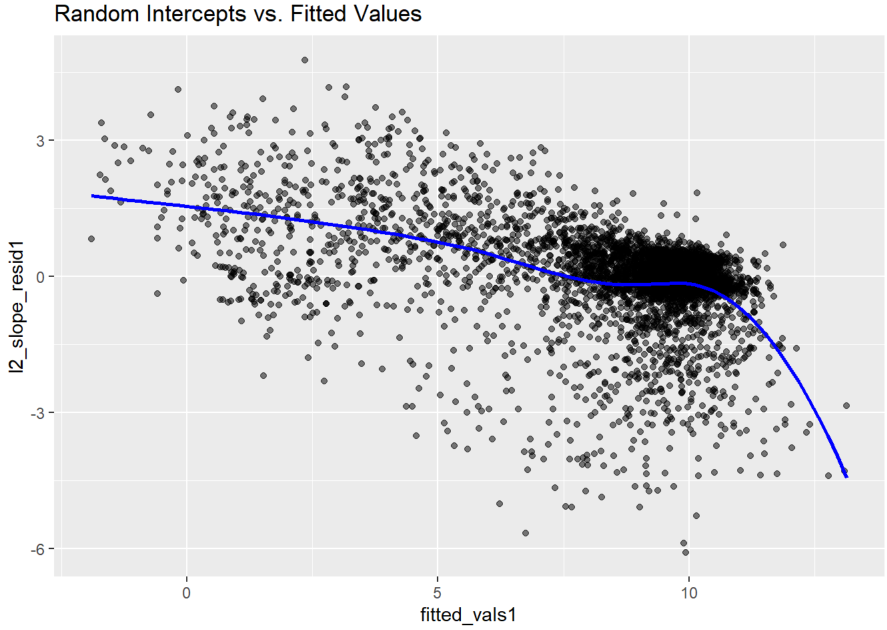
```
# Random intercepts
l2_data$fitted_vals1 <- fitted(rm5)[match(l2_data$id_id, soep_final$id_id)]

ggplot(l2_data, aes(x = fitted_vals1, y = l2_intercept_resid1)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



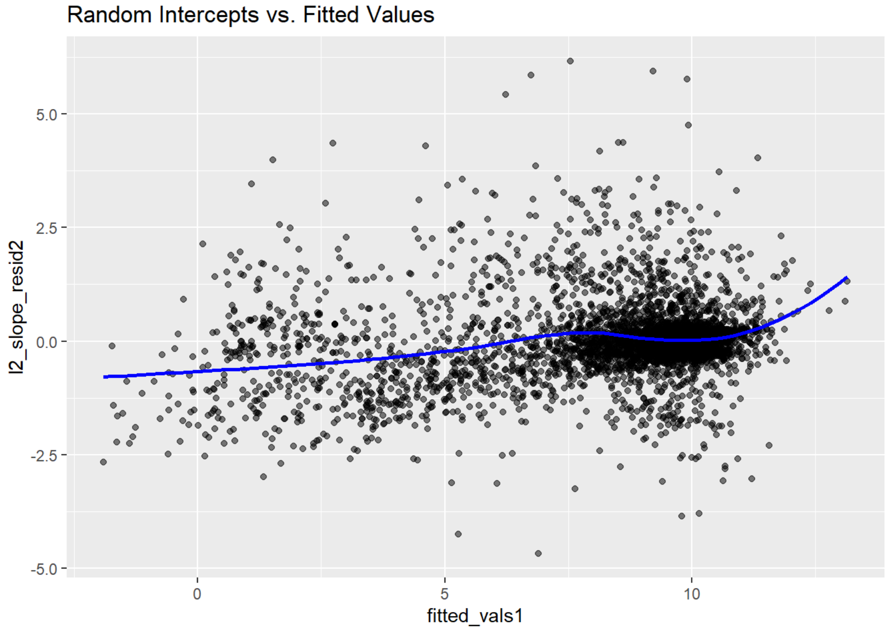

```
# Random slopes
ggplot(l2_data, aes(x = fitted_vals1, y = l2_slope_resid1)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Random slopes
ggplot(l2_data, aes(x = fitted_vals1, y = l2_slope_resid2)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

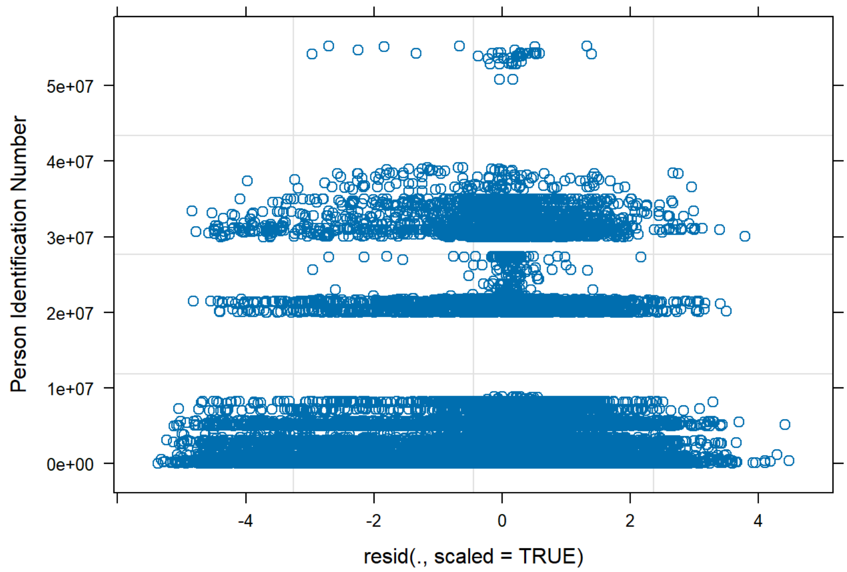
```
## `geom_smooth()` using formula = 'y ~ x'
```



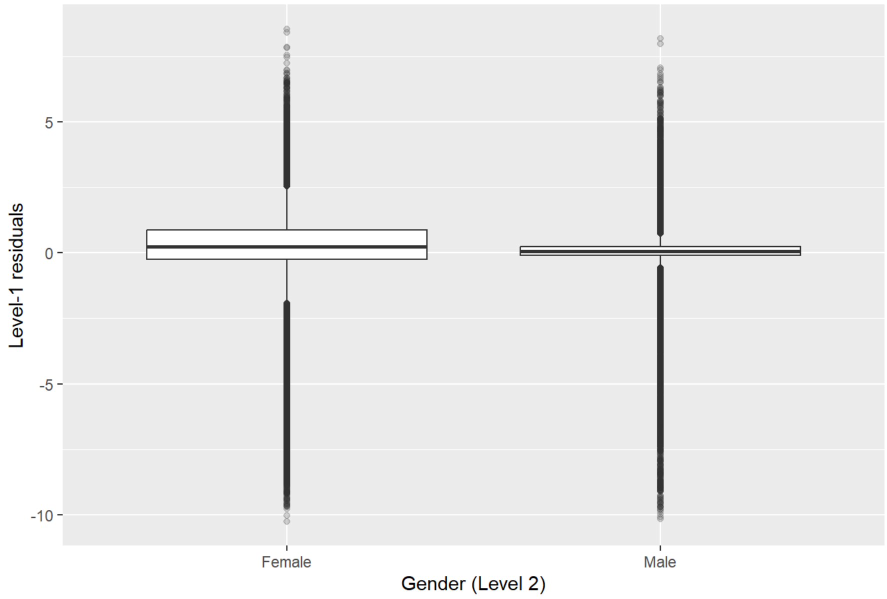

```
#Check whether the variance of the L1 residual errors is the same in all groups
soep_final$l1resid123 <- resid(rm5)

soep_final$abs_l1resid1 <- abs(soep_final$l1resid123)
soep_final$sq_l1resid  <- soep_final$abs_l1resid1^2


Levene.Model.F1 <- lm(sq_l1resid ~ id_id, data=soep_final)
#ANOVA of the squared residuals
anova(Levene.Model.F1) #displays the results
```

```
## Analysis of Variance Table
##
## Response: sq_l1resid
##              Df  Sum Sq Mean Sq F value Pr(>F)
## id_id         1       2   2.286  0.0254 0.8734
## Residuals 64995 5846838  89.958
```

```
plot(rm5, id_id ~ resid(., scaled=TRUE))
```



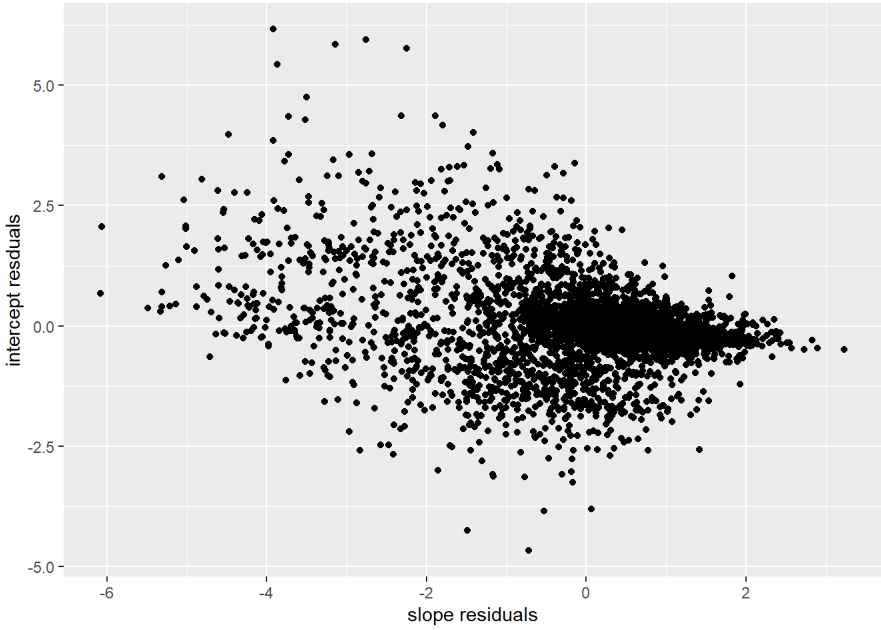

```
# Independence of L1 residuals from Level 2 predictors
ggplot(soep_final, aes(x = male, y = l1resid123)) +
  geom_boxplot(outlier.alpha = 0.2) +
  labs(x = "Gender (Level 2)",
       y = "Level-1 residuals",
       title = "Are Level-1 residuals independent of gender?")
```




```
leveneTest(l1resid123 ~ male, data = soep_final)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value    Pr(>F)
## group     1  4582.6 < 2.2e-16 ***
##       64995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Intercept residuals and Slope residuals relation
l2_data %>%
 ggplot(mapping = aes(x = l2_intercept_resid1, y = l2_slope_resid2)) +
 geom_point() +
 labs(x = "slope residuals", y = "intercept resduals")
```




```
cor.test(l2_data$l2_slope_resid2, as.numeric(l2_data$l2_intercept_resid1))
```

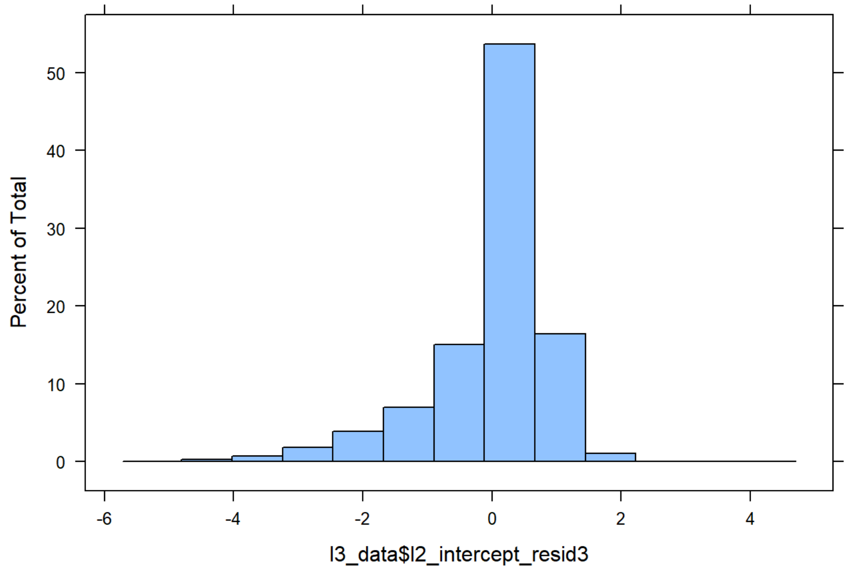
```
##
##  Pearson's product-moment correlation
##
## data:  l2_data$l2_slope_resid2 and as.numeric(l2_data$l2_intercept_resid1)
## t = -21.195, df = 4883, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3157268 -0.2643612
## sample estimates:
##       cor
## -0.290253
```

#Level 3

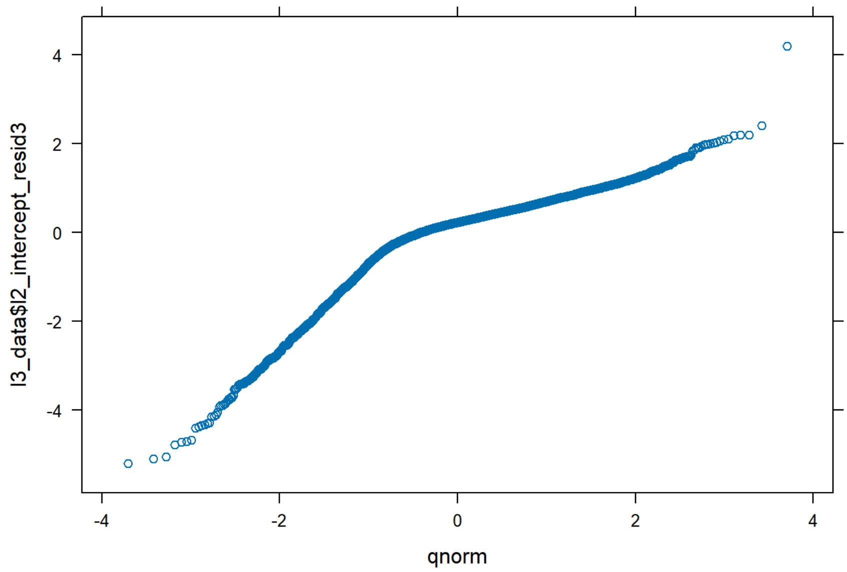
```
#Level 2 residuals

l3_data <- soep_final %>%
group_by(hh_id) %>%
mutate(log_income_mean = mean(log_income, na.rm = T)) %>%
dplyr::select(hh_id, log_income_mean) %>%
unique()

#Random Intercept and Random Slope residuals
l3_data$l2_intercept_resid3 = ranef(m16)$hh_id [, 1]
l3_data$l2_intercept_resid4 = ranef(rm5)$hh_id [, 1]
```

#Normality

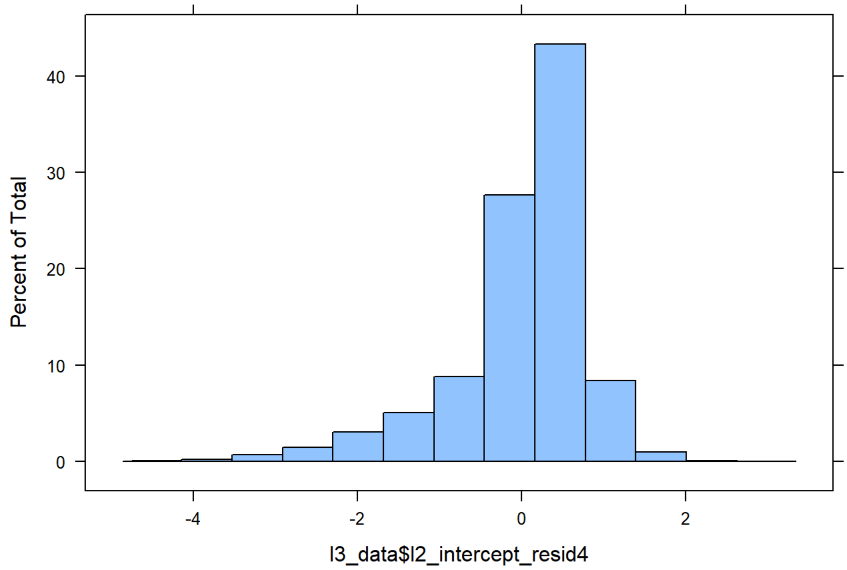
```
#Histogram
histogram(l3_data$l2_intercept_resid3)
```



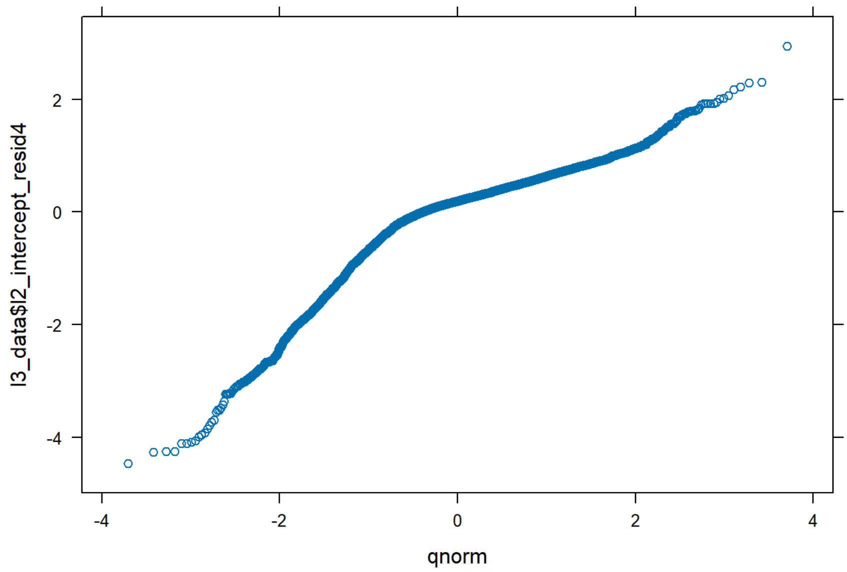

```
#QQ Plot
qqmath(l3_data$l2_intercept_resid3)
```




```
#Histogram
histogram(l3_data$l2_intercept_resid4)
```



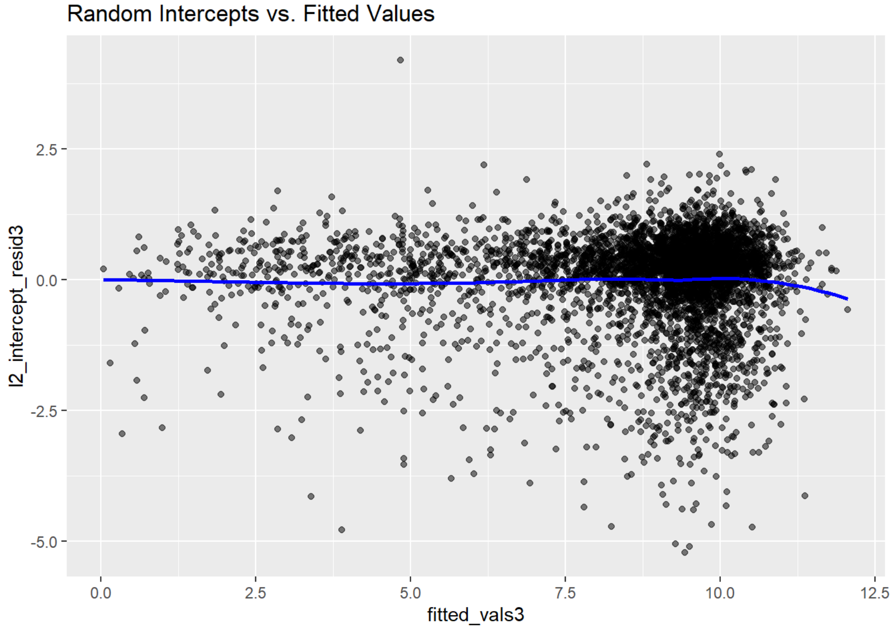

```
#QQ Plot
qqmath(l3_data$l2_intercept_resid4)
```



#Homoskedasticity

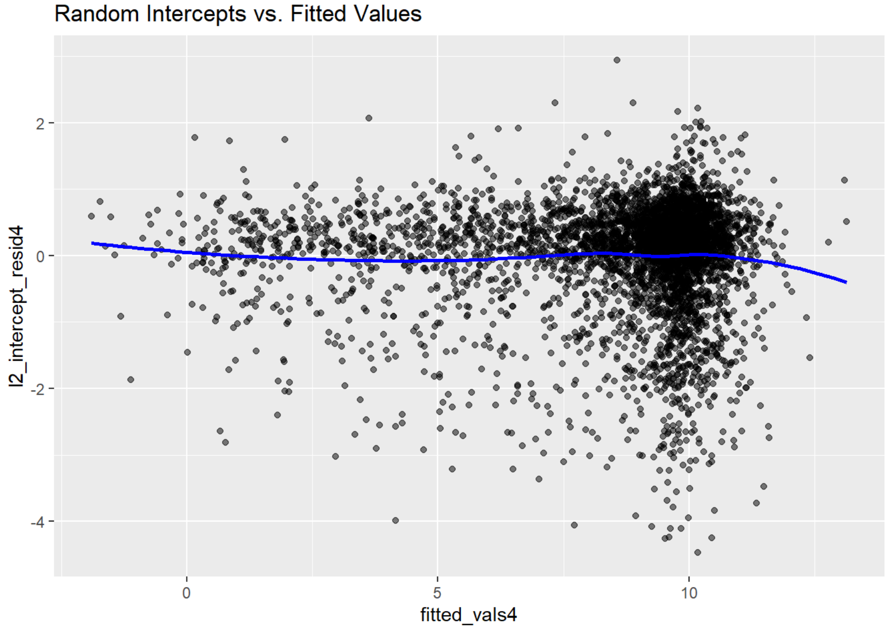
```
# Random intercepts
l3_data$fitted_vals3 <- fitted(ml6)[match(l3_data$hh_id, soep_final$hh_id)]
l3_data$fitted_vals4 <- fitted(rm5)[match(l3_data$hh_id, soep_final$hh_id)]

ggplot(l3_data, aes(x = fitted_vals3, y = l2_intercept_resid3)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(l3_data, aes(x = fitted_vals4, y = l2_intercept_resid4)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "loess", color = "blue", se = FALSE) +
  labs(title = "Random Intercepts vs. Fitted Values")
```

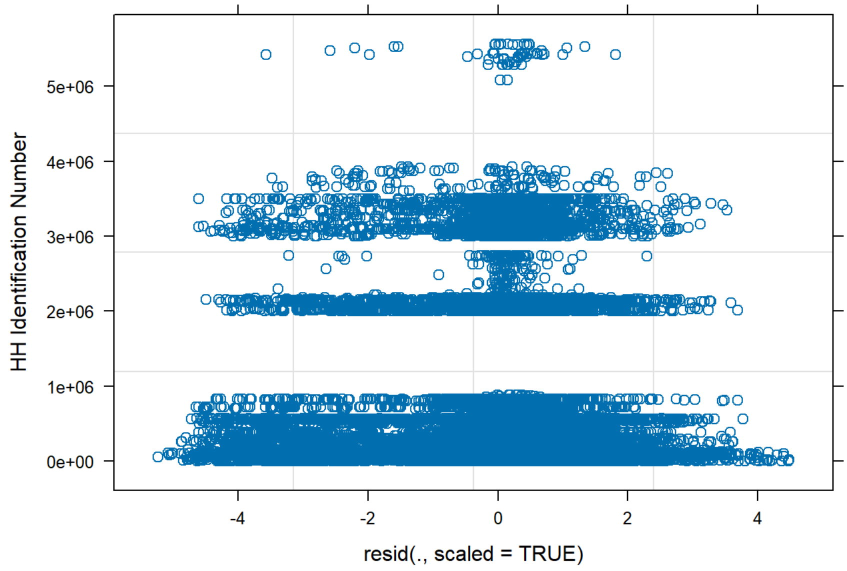
```
## `geom_smooth()` using formula = 'y ~ x'
```




```
#Check whether the variance of the L1 residual errors is the same in all groups
soep_final$l1resid3 <- resid(ml6)
soep_final$l1resid4 <- resid(rm5)

soep_final$abs_l1resid3 <- abs(soep_final$l1resid3)
soep_final$sq_l1resid3  <- soep_final$abs_l1resid3^2

soep_final$abs_l1resid4 <- abs(soep_final$l1resid4)
soep_final$sq_l1resid4  <- soep_final$abs_l1resid4^2



Levene.Model.F3 <- lm(sq_l1resid3 ~ hh_id, data=soep_final)
#ANOVA of the squared residuals
anova(Levene.Model.F3) #displays the results
```

```
## Analysis of Variance Table
##
## Response: sq_l1resid3
##              Df  Sum Sq Mean Sq F value Pr(>F)
## hh_id         1      44  44.103   0.318 0.5728
## Residuals 64995 9014122 138.689
```

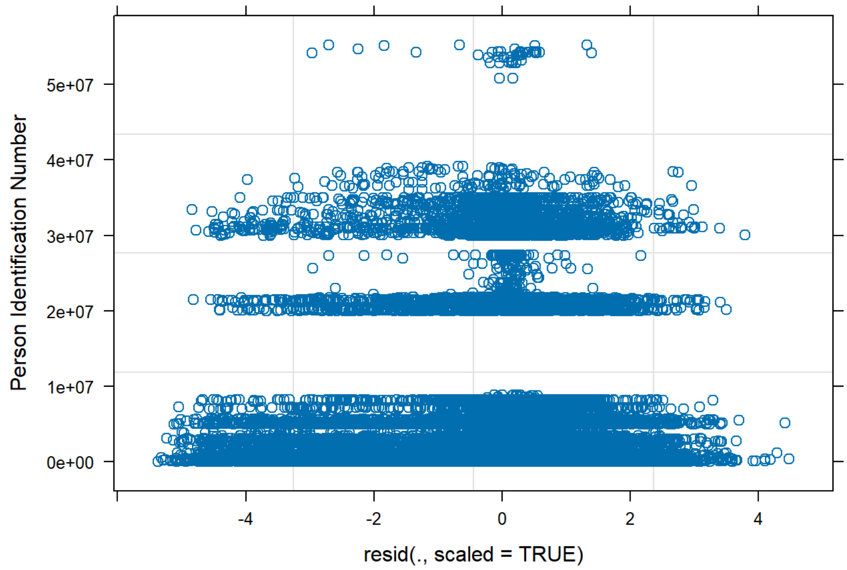
```
plot(ml6, hh_id ~ resid(., scaled=TRUE))
```




```
Levene.Model.F4 <- lm(sq_l1resid4 ~ hh_id, data=soep_final)
#ANOVA of the squared residuals
anova(Levene.Model.F4) #displays the results
```

```
## Analysis of Variance Table
##
## Response: sq_l1resid4
##              Df  Sum Sq Mean Sq F value Pr(>F)
## hh_id         1       0   0.206  0.0023 0.9618
## Residuals 64995 5846840  89.958
```

```
plot(rm5, id_id ~ resid(., scaled=TRUE))
```



##Missing Data

```
soep_fil1 <- soep_sub %>%
  filter(age >= 25 & age <= 60)

#Recode gender so that male = 1 and female = 0
soep_fil1 <- soep_fil1 %>%
  mutate(male = case_when(
    gender == 1 ~ 1,
    gender == 2 ~ 0,
    TRUE ~ NA_real_))

#Recode marital status so that single = 0 and married = 1
soep_fil1 <- soep_fil1 %>%
  mutate(marital_status = marital_status_unfinished - 1) %>%
  mutate(marital_status = case_when(
    marital_status == 0 ~ 1,
    marital_status == 1 ~ 0,
    TRUE ~ marital_status ))

#Create a new variable for which single = 0, married = 1, and other = 2
soep_fil1 <- soep_fil1 %>%
  mutate(marital_status_reduced = case_when(
    marital_status == 0 ~ 0,
    marital_status == 1 ~ 1,
    marital_status > 1 ~ 2
  ))
```

```
#Filter the dataset for individuals that went from unmarried to married during the course of the panel
ids_changed1 <- soep_fil1 %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(prev_marital_status = lag(marital_status_reduced)) %>%
  filter(prev_marital_status == 0 & marital_status_reduced == 1) %>%
  distinct(id_id)

soep_changed1 <- soep_fil1 %>%
  filter(id_id %in% ids_changed$id_id)
```

```
#Determine the year in which the individual got married
soep_d1 <- soep_fil1 %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(prev_marital_status = lag(marital_status_reduced),
         transition = prev_marital_status == 0 & marital_status_reduced == 1) %>%
  filter(transition) %>%
  slice(1) %>%
  dplyr::select(id_id, transition_year = year)


#Center the variable so that 0 = year of marriage
soep_final1 <- soep_fil1 %>%
  inner_join(soep_d, by = "id_id") %>%
  mutate(relative_year = year - transition_year)

soep_final1 <- soep_final1 %>%
  arrange(id_id, year) %>%
  group_by(id_id) %>%
  mutate(married = cummax(marital_status_reduced %in% c(1, 2)))

soep_final_NA <- soep_final1 %>%
  dplyr::select(id_id, hh_id, income, age, year, education, gender, married, transition_year, relative_year)


soep_final_NA <- soep_final_NA %>%
  mutate(
    income = ifelse(income < 0, NA, income),
    gender = ifelse(gender < 0, NA, gender),
    age = ifelse(age < 0, NA, age),
    education = ifelse(education < 0, NA, education))
```

```
#Count observations on each level
data.frame(
  level1_observations = nrow(soep_final_NA),
  level2_individuals = length(unique(soep_final_NA$id_id)),
  level3_households = length(unique(soep_final_NA$hh_id)))
```

```
##   level1_observations level2_individuals level3_households
## 1               66523               4885              4759
```

```
#Total number of missings
sum(is.na(soep_final_NA))
```

```
## [1] 1514
```

```
#Number of missings per variable
colSums(is.na(soep_final_NA))
```

```
##           id_id           hh_id          income             age            year 
##               0               0             132               0               0 
##       education          gender         married transition_year   relative_year 
##            1382               0               0               0               0
```

```
#Pattern of missing values
md.pattern(soep_final_NA, plot = F)
```

```
##       id_id hh_id age year gender married transition_year relative_year income
## 65016     1     1   1    1      1       1               1             1      1
## 1375      1     1   1    1      1       1               1             1      1
## 125       1     1   1    1      1       1               1             1      0
## 7         1     1   1    1      1       1               1             1      0
##           0     0   0    0      0       0               0             0    132
##       education     
## 65016         1    0
## 1375          0    1
## 125           1    1
## 7             0    2
##            1382 1514
```

```
#Proportion of missing values
sum(is.na(soep_final_NA)) / (nrow(soep_final_NA) * ncol(soep_final_NA)) * 100
```

```
## [1] 0.2275905
```

```
#Number of missings on income variable
sum(is.na(soep_final_NA$income))
```

```
## [1] 132
```

```
#Number of missings on education variable
sum(is.na(soep_final_NA$education))
```

```
## [1] 1382
```

```
#Proportion of missings on income variable
sum(is.na(soep_final_NA$income)) / nrow(soep_final_NA) * 100
```

```
## [1] 0.1984276
```

```
#Proportion of missings on education variable
sum(is.na(soep_final_NA$education)) / nrow(soep_final_NA) * 100
```

```
## [1] 2.077477
```

```
#Missings on income by gender
soep_final_NA %>%
  mutate(income_missing = is.na(income)) %>%
  group_by(gender) %>%
  summarise(
    percent_missing_income = mean(income_missing) * 100,
    n = n())
```

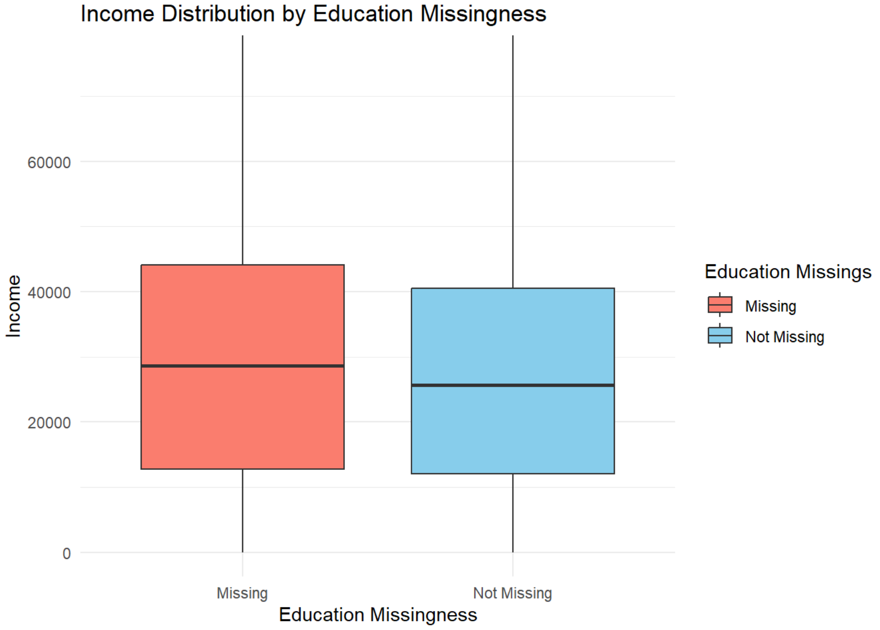
```
## # A tibble: 2 × 3
##   gender percent_missing_income     n
##    <dbl>                  <dbl> <int>
## 1      1                  0.247 36040
## 2      2                  0.141 30483
```

```
#Missings on education by gender
soep_final_NA %>%
  mutate(education_missing = is.na(education)) %>%
  group_by(gender) %>%
  summarise(
    percent_missing_education = mean(education_missing) * 100,
    n = n())
```

```
## # A tibble: 2 × 3
##   gender percent_missing_education     n
##    <dbl>                     <dbl> <int>
## 1      1                      2.60 36040
## 2      2                      1.46 30483
```

```
#Income of observations with NA on education vs no NA
soep_final_NA %>%
  mutate(
    education_missing = ifelse(is.na(education), "Missing", "Not Missing")
  ) %>%
  ggplot(aes(x = education_missing, y = income, fill = education_missing)) +
  geom_boxplot(outlier.shape = NA) +
  coord_cartesian(ylim = c(0, quantile(soep_final_NA$income, 0.95, na.rm = TRUE))) +
  labs(
    x = "Education Missingness",
    y = "Income",
    title = "Income Distribution by Education Missingness",
    fill = "Education Missings"
  ) +
  scale_fill_manual(values = c("Not Missing" = "skyblue", "Missing" = "salmon")) +
  theme_minimal()
```

```
## Warning: Removed 132 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
#Little's MCAR test
soep_final_NA_test <- soep_final_NA %>%
  dplyr::select(education, income, id_id, age, relative_year)

mcar_test(soep_final_NA_test)
```

```
## # A tibble: 1 × 4
##   statistic    df p.value missing.patterns
##       <dbl> <dbl>   <dbl>            <int>
## 1      380.    11       0                4
```

##Tables

```
#Descriptive Statistics

soep_desc <- soep_final %>%
  dplyr::select(log_income, married, relative_year, male, age, age_squared, transition_year, education)

soep_desc <- as.data.frame(soep_desc)

soep_desc$married <- as.numeric(soep_desc$married)
soep_desc$male <- as.numeric(soep_desc$male)
soep_desc$education <- as.numeric(soep_desc$education)

soep_desc <- soep_desc %>%
  mutate(married = case_when(
    married == 1 ~ 0,
    married == 2 ~ 1,
    TRUE ~ NA_real_)) %>%
  mutate(male = case_when(
    male == 1 ~ 0,
    male == 2 ~ 1,
    TRUE ~ NA_real_))


soep_desc <- soep_desc %>%
  mutate(education0 = case_when(
    education == 1 ~ 1,
    education == 2 ~ 0,
    education == 3 ~ 0,
    TRUE ~ NA_real_))

soep_desc <- soep_desc %>%
  mutate(education1 = case_when(
    education == 1 ~ 0,
    education == 2 ~ 1,
    education == 3 ~ 0,
    TRUE ~ NA_real_))

soep_desc <- soep_desc %>%
  mutate(education2 = case_when(
    education == 1 ~ 0,
    education == 2 ~ 0,
    education == 3 ~ 1,
    TRUE ~ NA_real_))

soep_desc <- soep_desc %>%
  dplyr::select(-c(education))


#USE LATEX AND CHANGE N MANUALLY
stargazer(soep_desc, style = "asr", digits=1, title = "Table 1 Descriptives", out = "table1.html", covariate.labels =
        c("log Income", "Marital Status", "Relative Year to Marriage","Gender" , "Age", "Age Squared", "Year of Marriage",
"Less than High School", "High School", "More than High School"))
```

```
##
## % Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com
## % Date and time: Thu, Jun 26, 2025 - 10:28:22 PM
## \begin{table}[!htbp] \centering
##   \caption{Table 1 Descriptives}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lccccc}
## \\[-1.8ex]\hline \\[-1.8ex]
## Statistic & \multicolumn{1}{c}{N} & \multicolumn{1}{c}{Mean} & \multicolumn{1}{c}{St. Dev.} & \multicolumn{1}{c}{Min} &
\multicolumn{1}{c}{Max} \\
## \hline \\[-1.8ex]
## log Income & 64,997 & 9.1 & 3.2 & 0.0 & 14.0 \\
## Marital Status & 64,997 & 0.7 & 0.5 & 0 & 1 \\
## Relative Year to Marriage & 64,997 & 3.5 & 8.1 & $-$30 & 34 \\
## Gender & 64,997 & 0.5 & 0.5 & 0 & 1 \\
## Age & 64,997 & 36.8 & 8.1 & 25 & 60 \\
## Age Squared & 64,997 & 1,417.6 & 645.4 & 625 & 3,600 \\
## Year of Marriage & 64,997 & 18.8 & 9.9 & 0 & 37 \\
## Less than High School & 64,997 & 0.1 & 0.2 & 0 & 1 \\
## High School & 64,997 & 0.6 & 0.5 & 0 & 1 \\
## More than High School & 64,997 & 0.3 & 0.5 & 0 & 1 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
#Presentation


tab_model(
  ml1, ml2, ml3, ml4, ml4a, ml5, ml6,
  dv.labels = "Log Income",
  show.re.var = TRUE,
  show.ci = FALSE,
  file = "M1.html",
  pred.labels = c("Intercept", "Year", "Age", "Age Squared",
                  "High School", "More Than High School",
                  "Married", "Male",
                  "Year of Marriage", "Married:Male"))
```

	Log Income													
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Intercept	8.05	**<0.001**	8.02	**<0.001**	7.79	**<0.001**	6.60	**<0.001**	6.78	**<0.001**	6.76	**<0.001**	7.10	**<0.001**
Year	0.04	**<0.001**	0.04	**<0.001**	0.02	**<0.001**	0.02	**<0.001**	0.03	**<0.001**	0.03	**<0.001**	0.03	**<0.001**
Age			-0.12	**<0.001**	-0.29	**<0.001**	-0.26	**<0.001**	-0.28	**<0.001**	-0.25	**<0.001**	-0.89	**<0.001**
Age Squared					0.36	**<0.001**	0.28	**<0.001**	0.24	**<0.001**	0.22	**<0.001**	0.23	**<0.001**
High School					-0.22	**<0.001**	-0.17	**<0.001**	-0.16	**<0.001**	-0.14	**0.001**	-0.14	**<0.001**
More Than High School					0.61	**<0.001**	0.70	**<0.001**	0.69	**<0.001**	0.90	**<0.001**	0.90	**<0.001**
Married					2.14	**<0.001**	2.17	**<0.001**	2.17	**<0.001**	2.19	**<0.001**	2.19	**<0.001**
Male							1.70	**<0.001**	1.69	**<0.001**	1.52	**<0.001**	0.88	**<0.001**
Year of Marriage									-0.01	**0.001**	-0.02	**<0.001**	-0.02	**<0.001**
Married:Male													1.21	**<0.001**
Random Effects														
σ^2	5.20		5.20		5.12		5.13		5.13		4.53		4.53	
τ_{00}	3.32 id_id		3.32 id_id		3.36 id_id		2.35 id_id		2.34 id_id		2.91 id_id		2.80 id_id	
	2.59 hh_id		2.61 hh_id		2.26 hh_id		2.34 hh_id		2.34 hh_id		1.91 hh_id		1.89 hh_id	
τ_{11}											3.96 id_id.marriedYes		3.62 id_id.marriedYes	
ρ_{01}											-0.54 id_id		-0.53 id_id	
ICC	0.53		0.53		0.52		0.48		0.48		0.52		0.52	
N	4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id	
	4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id	
Observations	64997		64997		64997		64997		64997		64997		64997	
Marginal R^2 / Conditional R^2	0.009 / 0.536		0.011 / 0.538		0.063 / 0.553		0.121 / 0.541		0.120 / 0.540		0.105 / 0.574		0.125 / 0.578	

```
tab_model(
  ml1, rm1, rm1a, rm2, rm3, rm4, rm4b, rm5,
  dv.labels = "Log Income",
  show.re.var = TRUE,
  show.ci = FALSE,
  file = "M2.html",
  pred.labels = c("Intercept","Year",
                  "High School", "More Than High School",
                  "Relative Year to Marriage", "Male", "Relative Year to Marriage Squared",
                  "Relative Year to Marriage:Male", "Relative Year to Marriage Squared:Male"))
```

	Log Income													
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Intercept	8.05	**<0.001**	8.49	**<0.001**	7.57	**<0.001**	6.46	**<0.001**	6.43	**<0.001**	6.64	**<0.001**	6.47	**<0.0**
Year	0.04	**<0.001**	0.02	**<0.001**	0.01	**<0.001**	0.02	**<0.001**	0.01	**<0.001**	0.01	**0.003**	0.01	**0.00**
High School			0.18	**<0.001**	0.14	**<0.001**	0.11	**<0.001**	0.13	**<0.001**	0.07	**0.002**	0.05	0.08
More Than High School			-0.08	**<0.001**	-0.05	**<0.001**	-0.06	**<0.001**	-0.05	**<0.001**	-0.05	**0.003**	-0.01	0.81

Relative Year to Marriage					0.62	**<0.001**	0.71	**<0.001**	0.92	**<0.001**	0.73	**<0.001**	1.08	**<0.0**
Male					2.10	**<0.001**	2.13	**<0.001**	2.20	**<0.001**	2.13	**<0.001**	2.20	**<0.0**
Relative Year to Marriage Squared							1.72	**<0.001**	1.67	**<0.001**	1.72	**<0.001**	1.52	**<0.0**
Relative Year to Marriage:Male														
Relative Year to Marriage Squared:Male														
Random Effects														
σ^2	5.20		5.19		5.14		5.14		4.11		4.35		3.64	
τ_{00}	3.32 id_id		3.30 id_id		3.38 id_id		2.34 id_id		2.51 id_id		2.44 id_id		2.55 id_id	
	2.59 hh_id		2.51 hh_id		2.19 hh_id		2.30 hh_id		1.86 hh_id		2.18 hh_id		1.70 hh_id	
τ_{11}									1.19 id_id.relative_year_sd		0.85 id_id.relative_year_squared_sd		1.94 id_id.relative_year_	
													1.29 id_id.relative_year_square	
ρ_{01}									0.12 id_id		0.12 id_id		0.26 id_id.relative_year_	
													-0.21 id_id.relative_year_square	
ICC	0.53		0.53		0.52		0.47		0.57		0.55		0.64	
N	4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id		4730 hh_id	
	4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id		4885 id_id	
Observations	64997		64997		64997		64997		64997		64997		64997	
Marginal R^2 / Conditional R^2	0.009 / 0.536		0.007 / 0.532		0.056 / 0.547		0.117 / 0.536		0.110 / 0.618		0.113 / 0.604		0.088 / 0.674	