Contraceptive Methods and Age

Jessica Spencer

September 2024

Age on Contraceptive Prevalence

This data set is from a 1987 National Indonesia Contraceptive Prevalence Survey. All observations are married women who were definitely not pregnant or did not know yet. Questions on the survey covered topics regarding socio-economic status and general demographics.

1. Exploratoration of Dataset
2. Distribution and Correlation Visualization
3. Relative Odds of Contraceptive Method
4. Predictive Strength of Age on Contraceptive Method

```
library('ggvis')
library('tidyverse')
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()       masks stats::filter()
## x dplyr::lag()          masks stats::lag()
## x ggplot2::resolution() masks ggvis::resolution()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library('ggplot2')

df = read.csv('data/1987 Indonesia Contraception Prevalence Study.csv')
head(df)
```

```
##   Age Education Partner.Education Number.of.Children Religion...Islam
## 1  24         2                 3                  3                1
## 2  45         1                 3                 10                1
## 3  43         2                 3                  7                1
## 4  42         3                 2                  9                1
## 5  36         3                 3                  8                1
## 6  19         4                 4                  0                1
##   Currently.working Husband.Occupation Standard.of.Living Media.Exposure
## 1                 1                  2                  3              0
```

```
## 2                 1                 3                 4              0
## 3                 1                 3                 4              0
## 4                 1                 3                 3              0
## 5                 1                 3                 2              0
## 6                 1                 3                 3              0
##   Contraceptive.Method.Used
## 1                         1
## 2                         1
## 3                         1
## 4                         1
## 5                         1
## 6                         1
```

Variable Information

Variable Information:

```
Age - age of the woman
Education - level of education woman has received (1=low, 4=high)
Partner Education - level of education partner has received (1=low, 4=high)
Number of Children - number of kids mothered by woman
Religion=Islam - woman that identify as Muslim (0=No, 1=Yes)
Currently Working - woman is currently employed (0=Yes, 1=No)
Husbands Occupation - Not specified (categorical 1-4)
Standard of Living - based on the standard of living index (1=low, 4=high)
Media exposure - quality of media exposure (0=Good, 1=Not good)
Contraceptive Method Used - 1=No-use, 2=Long-term, 3=Short-term
```

Exploratory Data Analysis

- Description of Dataframe
- Missingness Check
- Distribution of Variables

```
print("Rows x Columns:")
```

```
## [1] "Rows x Columns:"
```

```
print(dim(df))
```

```
## [1] 1473   10
```

```
print("Feature Type:")# Data types of the variables
```

```
## [1] "Feature Type:"
```

```
print(sapply(df, class))
```

```
##                         Age                   Education         Partner.Education
##                   "integer"                   "integer"                 "integer"
##         Number.of.Children            Religion...Islam         Currently.working
##                   "integer"                   "integer"                 "integer"
##         Husband.Occupation          Standard.of.Living            Media.Exposure
##                   "integer"                   "integer"                 "integer"
## Contraceptive.Method.Used
##                   "integer"
```

```
print("Missing Values Per Feature:") #Checking for missing values in the columns
```

```
## [1] "Missing Values Per Feature:"
```

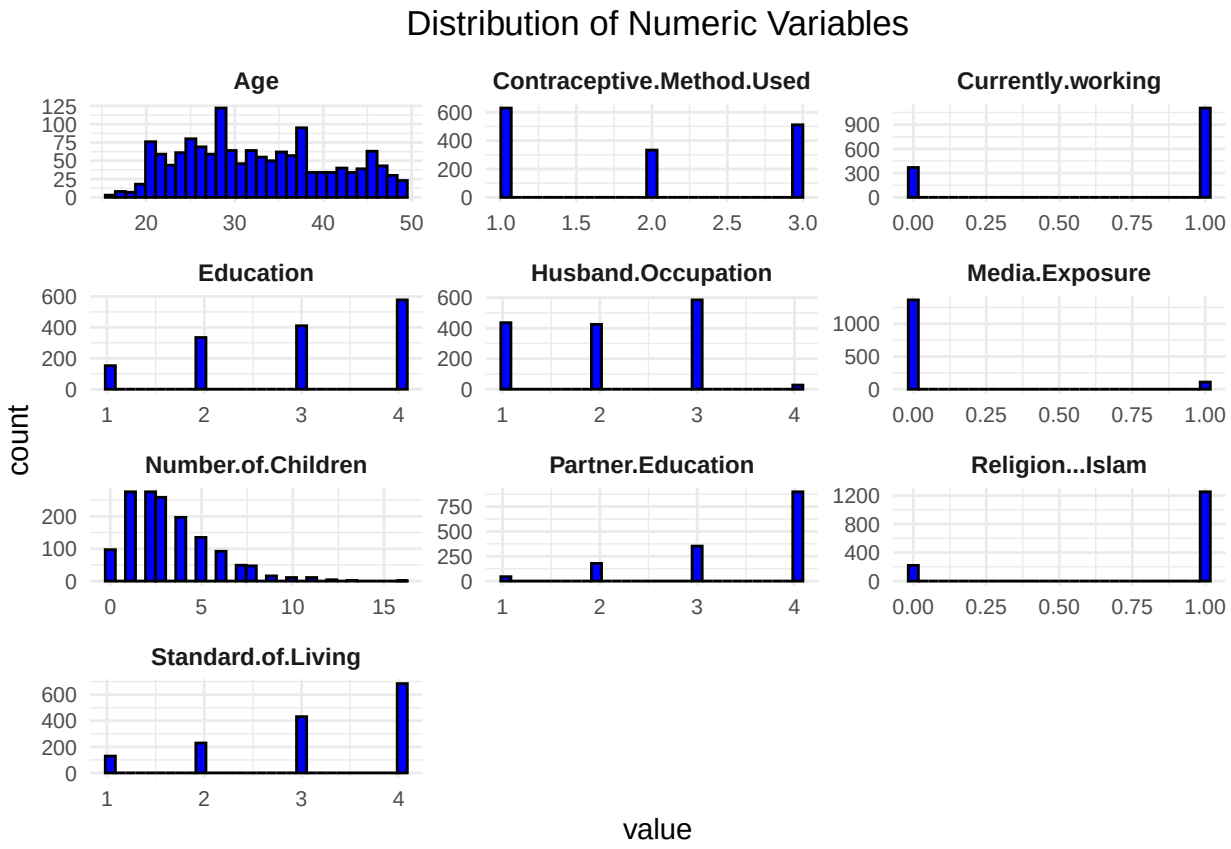
```
print(colSums(is.na(df)))
```

```
##                         Age                   Education         Partner.Education
##                           0                           0                         0
##         Number.of.Children            Religion...Islam         Currently.working
##                           0                           0                         0
##         Husband.Occupation          Standard.of.Living            Media.Exposure
##                           0                           0                         0
## Contraceptive.Method.Used
##                           0
```

```
library(tidyverse)
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.3.3
```

```
create_numeric_histograms <- function(df, ncol = 3) {
 df %>%
   select(where(is.numeric)) %>%
   pivot_longer(cols = everything()) %>%
   ggplot(aes(x = value)) +
   geom_histogram(bins = 30, fill = "blue", color = "black") +
   facet_wrap(~ name, scales = "free", ncol = ncol) +
   theme_minimal() +
   theme(
     strip.text = element_text(face = "bold"),
     axis.text = element_text(size = 8),
     plot.title = element_text(hjust = 0.5)
   ) +
   labs(title = "Distribution of Numeric Variables")
}

create_numeric_histograms(df)
```

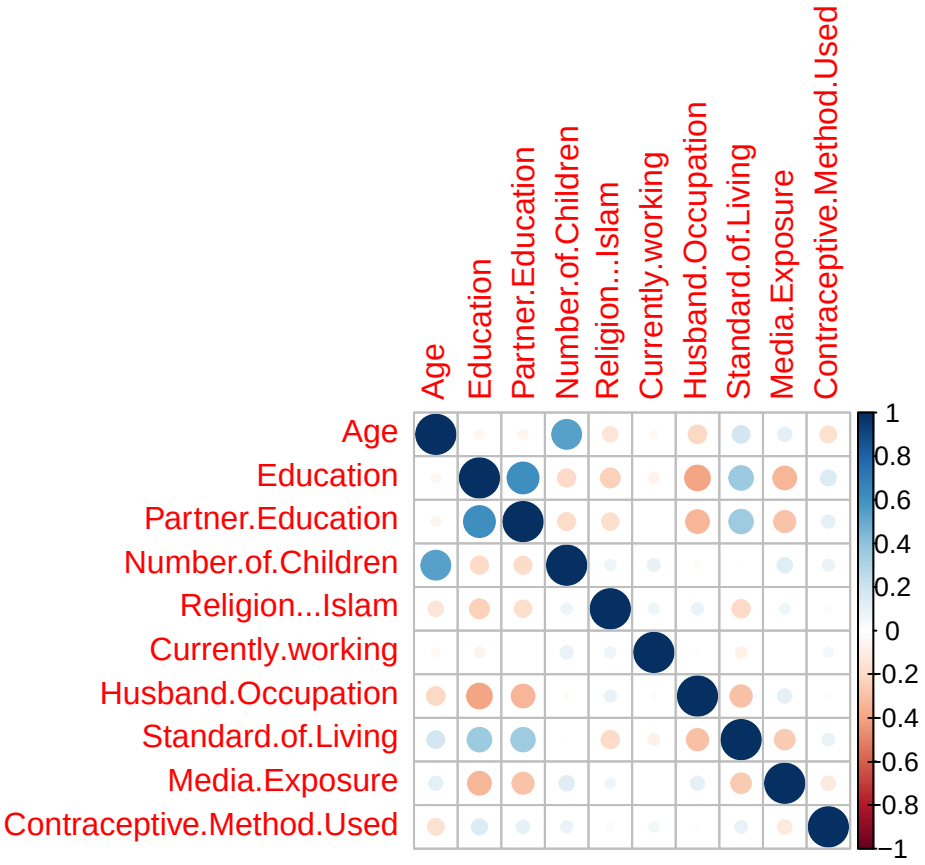



```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.95 loaded
```

```
cor_matrix <- df %>%select(where(is.numeric))  %>% cor()

corrplot(cor_matrix)
```

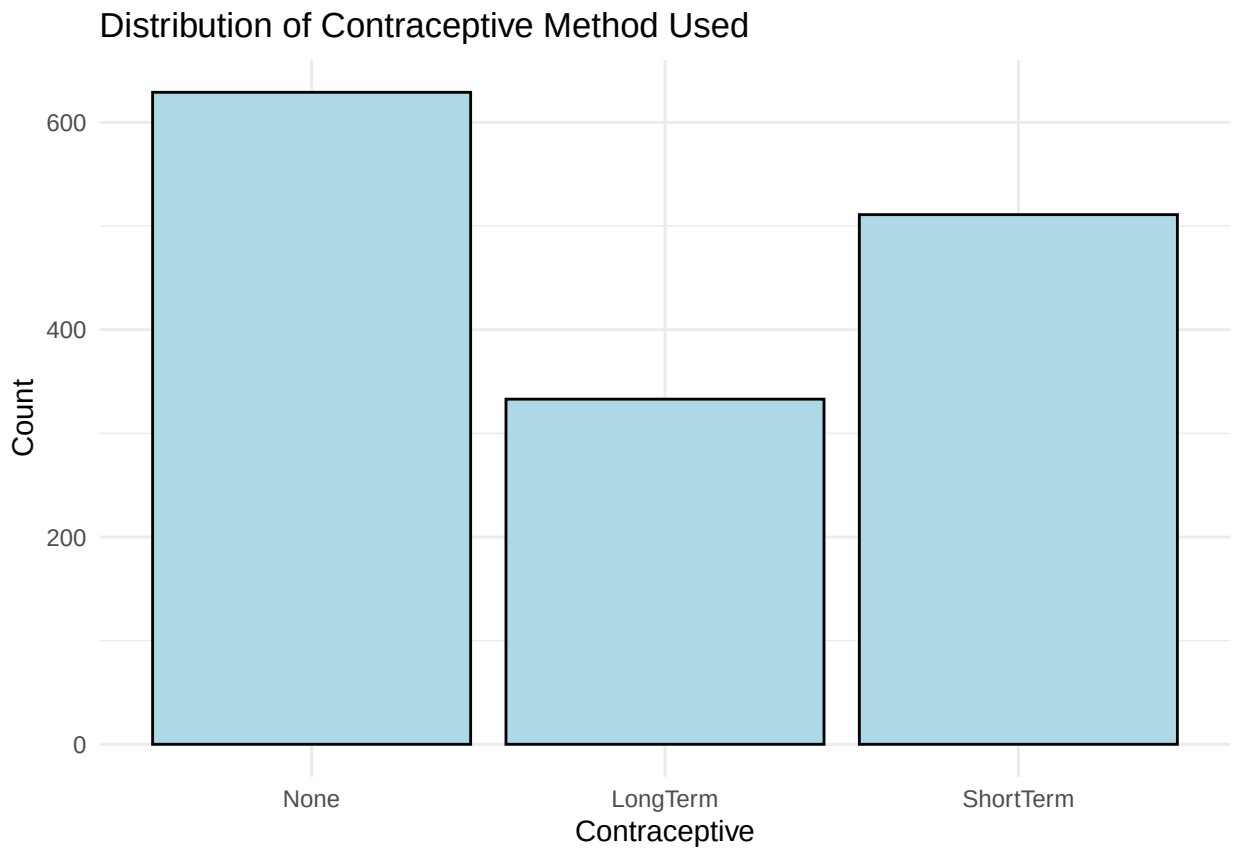


Impacts of Age on Contraceptive Method Used

Contraceptive Method Used: * 1 = No-use * 2 = Long-term * 3 = Short-term

```
df$contraceptive.method <- as.factor(df$Contraceptive.Method.Used)
#levels(df$contraceptive.method)
levels(df$contraceptive.method) <- c("None", "LongTerm", "ShortTerm")

ggplot(df, aes(x = contraceptive.method)) +
 geom_histogram(stat = "count", fill = "lightblue", color = "black") +
 theme_minimal() +
 labs(
   title = "Distribution of Contraceptive Method Used",
   x = "Contraceptive",
   y = "Count"
 )
```

```
## Warning in geom_histogram(stat = "count", fill = "lightblue", color = "black"):
## Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```




```
dplyr::count(df, contraceptive.method, sort = TRUE)
```

```
##   contraceptive.method   n
## 1                 None 629
## 2            ShortTerm 511
## 3             LongTerm 333
```

There are 3 unequal classes of contraceptive use. Most women (629 participants) in this sample do not use contraceptives, followed by short term contraceptive use (511 participants), and finally, about twenty-two percent of these women use long term contraceptives (333 participants).

Relative Odds of Contraception Method

```
library(nnet)
model1 = multinom(df$Contraceptive.Method.Used~1)
```

```
## # weights:  6 (2 variable)
## initial  value 1618.255901
## final  value 1571.363231
## converged
```

```
summary(model1)
```

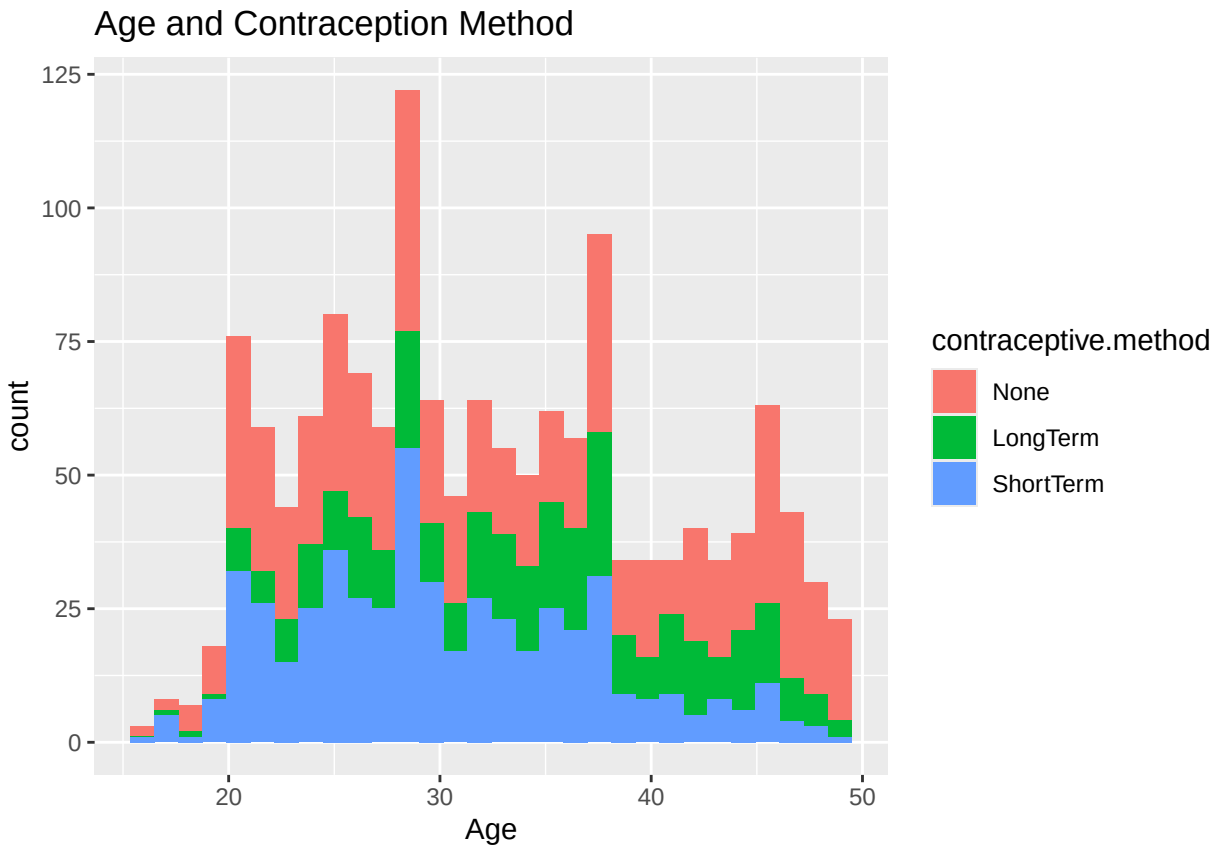
```
## Call:
## multinom(formula = df$Contraceptive.Method.Used ~ 1)
##
## Coefficients:
##   (Intercept)
## 2  -0.6359864
## 3  -0.2077754
##
## Std. Errors:
##   (Intercept)
## 2  0.06777021
## 3  0.05955488
##
## Residual Deviance: 3142.726
## AIC: 3146.726
```

Based on a null "mlogit" model, the relative odds of certain form of Contraceptive Use relative to neither are: * P(*Long-Term Contraception*)/P(No Use)=exp(-0.6359864)= *0.529* * P(*Short-Term Contraception*)/P(No Use)=exp(-0.2077754)= *0.812*

Predictive Strength of Age

```
df$contraceptive.method <- as.factor(df$Contraceptive.Method.Used)
#levels(df$contraceptive.method)
levels(df$contraceptive.method) <- c("None", "LongTerm", "ShortTerm")
#df = df %>%group_by(contraceptive.method) %>%mutate(Frequency = n())

ggplot(df, aes(Age, fill = contraceptive.method)) +
  geom_histogram() + ggtitle("Age and Contraception Method")
```

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




```
model2=multinom(contraceptive.method ~ Age, data = df)
```

```
## # weights:  9 (4 variable)
## initial  value 1618.255901
## final  value 1538.602530
## converged
```

```
summary(model2)
```

```
## Call:
## multinom(formula = contraceptive.method ~ Age, data = df)
##
## Coefficients:
##           (Intercept)         Age
## LongTerm    -1.114493  0.01411370
## ShortTerm    1.374399 -0.04975711
##
## Std. Errors:
##           (Intercept)         Age
## LongTerm    0.2881547 0.008224668
## ShortTerm   0.2491427 0.007654223
##
## Residual Deviance: 3077.205
## AIC: 3085.205
```

```
z <- summary(model2)$coefficients/summary(model2)$standard.errors
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
```

```
##            (Intercept)          Age
## LongTerm  1.098712e-04 8.615819e-02
## ShortTerm 3.457904e-08 7.999579e-11
```

With a p-value of 0.05 , the Age term is significant.

For Readability

```
beta.mat<-coef(model2)
beta.mat
```

```
##           (Intercept)         Age
## LongTerm    -1.114493  0.01411370
## ShortTerm    1.374399 -0.04975711
```

Recovering Relative probabilities Log(P(LongTerm)/P(None))=-1.114493+0.01411370*Age

Log(P(ShortTerm)/P(None))=1.374399-0.04975711*Age

P(ShortTerm)/P(None) = 0.9337401

```
print("odds ratio of long term contraception to no contraception for a 29 year old:")
```

```
## [1] "odds ratio of long term contraception to no contraception for a 29 year old:"
```

```
exp(1.114493+(0.01411370*29))
```

```
## [1] 4.589588
```

```
print("odds ratio of short term contraception to no contraception for an 18 year old:")
```

```
## [1] "odds ratio of short term contraception to no contraception for an 18 year old:"
```

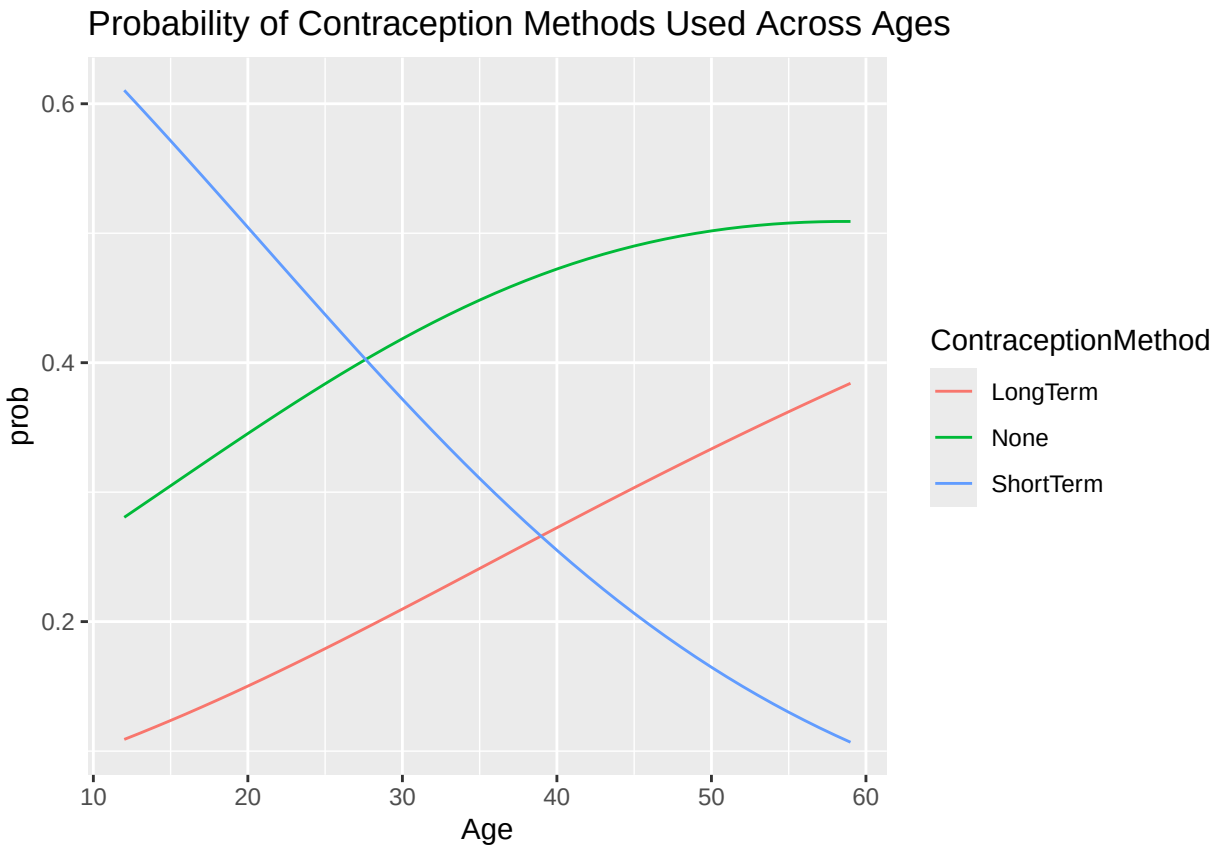
```
exp(1.374399+ (-0.04975711*18))
```

```
## [1] 1.614089
```

Graphing the probability of Contraceptive Method use across ages:

```
newdata = as.data.frame(matrix(0,48,1))
names(newdata) = c("Age")
newdata[,1] = 12:59
pred = predict(model2,newdata=newdata, type="probs")

new_predict <- cbind(newdata, pred)
#head(new_predict)
prob_plot <- new_predict %>%
  pivot_longer(2:4, names_to = "ContraceptionMethod", values_to = "prob")
#head(prob_plot)
```

```
ggplot(prob_plot, aes(x=Age, y=prob, group=ContraceptionMethod)) +
  geom_line(aes(color=ContraceptionMethod)) +
  ggtitle("Probability of Contraception Methods Used Across Ages")
```



Interestingly, it is predicted that using short-term contraception is much higher for younger participants, at greater than 40% probability of usage for those under 20 years old. As the age in the sample increases, predicted long-term contraceptive use increases, and short-term declines. Predicted use of no contraception surpasses short term around the age of 30.

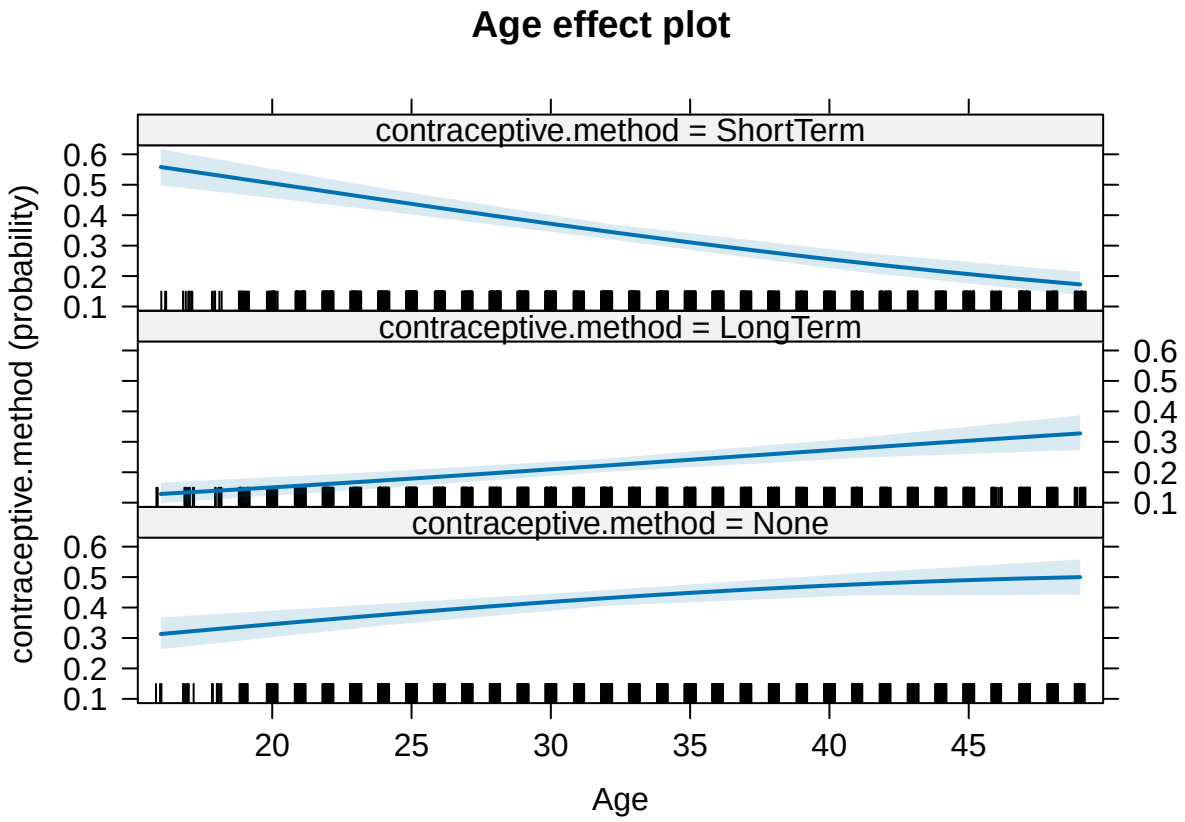
Following is the same chart, but broken out into contraceptive method:

```
library(effects)
```

```
## Loading required package: carData
```

```
## Warning in check_dep_version(): ABI version mismatch:
## lme4 was built with Matrix ABI version 1
## Current Matrix ABI version is 0
## Please re-install lme4 from source or restore original 'Matrix' package

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
plot(Effect("Age",model2))
```



Education and Contraceptive Type

Education and Partner's Education are strongly correlated. First I will add them separately, then as an interaction term if appropriate.

```
head(df)
```

```
##   Age Education Partner.Education Number.of.Children Religion...Islam
## 1  24         2                 3                 3                1
## 2  45         1                 3                10                1
## 3  43         2                 3                 7                1
## 4  42         3                 2                 9                1
## 5  36         3                 3                 8                1
## 6  19         4                 4                 0                1
```

```
##   Currently.working Husband.Occupation Standard.of.Living Media.Exposure
## 1                 1                  2                  3              0
## 2                 1                  3                  4              0
## 3                 1                  3                  4              0
## 4                 1                  3                  3              0
## 5                 1                  3                  2              0
## 6                 1                  3                  3              0
##   Contraceptive.Method.Used contraceptive.method
## 1                         1                 None
## 2                         1                 None
## 3                         1                 None
## 4                         1                 None
## 5                         1                 None
## 6                         1                 None
```

```
model3=multinom(contraceptive.method ~ Age + Education + Partner.Education, data = df)
```

```
## # weights:  15 (8 variable)
## initial  value 1618.255901
## iter  10 value 1472.478497
## final  value 1468.563869
## converged
```

```
summary(model3)
```

```
## Call:
## multinom(formula = contraceptive.method ~ Age + Education + Partner.Education,
##     data = df)
##
## Coefficients:
##           (Intercept)         Age Education Partner.Education
## LongTerm   -3.9048594  0.01882748 0.8859641       -0.03227119
## ShortTerm   0.3977791 -0.04657558 0.2677880        0.03383691
##
## Std. Errors:
##           (Intercept)         Age  Education Partner.Education
## LongTerm    0.4763635 0.008795481 0.10249389        0.12404645
## ShortTerm   0.3636809 0.007702697 0.07612793        0.09253532
##
## Residual Deviance: 2937.128
## AIC: 2953.128
```

Adding these terms has reduced the effect of age significantly.

The Partner Education variable seems weak. It is a very small coefficient, with a relatively large standard error.

it seems like education of the women has a large positive effect on the relative odds of choosing Long Term contraception over no contraception. It has a smaller positive effect on the relative odds of choosing short term education.

Interestingly, the education of the partner has a small negative effect on the relative odds of choosing long term contraception, and a small positive effect on relative odds of choosing short term contraception. However, the effect of this is very small, with a huge confidence interval. With a significance test we find that this does not pass the alpha level (0.05). Therefore, I'm dropping this variable

```
z <- summary(model3)$coefficients/summary(model3)$standard.errors
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
```

```
##            (Intercept)          Age    Education Partner.Education
## LongTerm  2.220446e-16 3.230743e-02 0.0000000000         0.7947450
## ShortTerm 2.740610e-01 1.478808e-09 0.0004354596         0.7146152
```

Education and Age are a significant term, but Partner.Education is not (alpha level 0.05). A Dropping Partner Education

```
model4=multinom(contraceptive.method ~ Age + Education, data = df)
```

```
## # weights:  12 (6 variable)
## initial  value 1618.255901
## iter  10 value 1468.708898
## final  value 1468.706497
## converged
```

```
summary(model4)
```

```
## Call:
## multinom(formula = contraceptive.method ~ Age + Education, data = df)
##
## Coefficients:
##           (Intercept)         Age Education
## LongTerm   -3.9684874  0.01884403 0.8699694
## ShortTerm   0.4668048 -0.04662449 0.2842928
##
## Std. Errors:
##           (Intercept)         Age  Education
## LongTerm    0.4088587 0.008795574 0.08121775
## ShortTerm   0.3106686 0.007699104 0.06119004
##
## Residual Deviance: 2937.413
## AIC: 2949.413
```

```
z <- summary(model4)$coefficients/summary(model4)$standard.errors
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
```

```
##           (Intercept)          Age    Education
## LongTerm    0.0000000 3.215775e-02 0.000000e+00
## ShortTerm   0.1329472 1.396925e-09 3.383303e-06
```

```
newdata = as.data.frame(matrix(0,4,2))
names(newdata) = c("Education","Age")
newdata[,1] = c(1,2,3,4)#1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)
newdata[,2] = 25#c(15,25,35,45,15,25,35,45,15,25,35,45,15,25,35,45)
pred = predict(model4,newdata=newdata, type="probs")
```

```r
new_predict <- cbind(newdata, pred)
#head(new_predict)
prob_plot <- new_predict %>%
  pivot_longer(3:5, names_to = "ContraceptionMethod", values_to = "prob")
head(prob_plot)
```

```
## # A tibble: 6 x 4
##   Education   Age ContraceptionMethod   prob
##       <dbl> <dbl> <chr>                <dbl>
## 1         1    25 None                 0.577
## 2         1    25 LongTerm             0.0417
## 3         1    25 ShortTerm            0.381
## 4         2    25 None                 0.488
## 5         2    25 LongTerm             0.0841
## 6         2    25 ShortTerm            0.428
```

```r
newdata = as.data.frame(matrix(0,4,2))
names(newdata) = c("Education","Age")
newdata[,1] = c(1,2,3,4)#1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)
newdata[,2] = 25#c(15,25,35,45,15,25,35,45,15,25,35,45,15,25,35,45)
pred = predict(model4,newdata=newdata, type="probs")

new_predict <- cbind(newdata, pred)
#head(new_predict)
prob_plot <- new_predict %>%
  pivot_longer(3:5, names_to = "ContraceptionMethod", values_to = "prob")
head(prob_plot)
```

```
## # A tibble: 6 x 4
##   Education   Age ContraceptionMethod   prob
##       <dbl> <dbl> <chr>                <dbl>
## 1         1    25 None                 0.577
## 2         1    25 LongTerm             0.0417
## 3         1    25 ShortTerm            0.381
## 4         2    25 None                 0.488
## 5         2    25 LongTerm             0.0841
## 6         2    25 ShortTerm            0.428
```

Average Predicted Probabilities for Education:

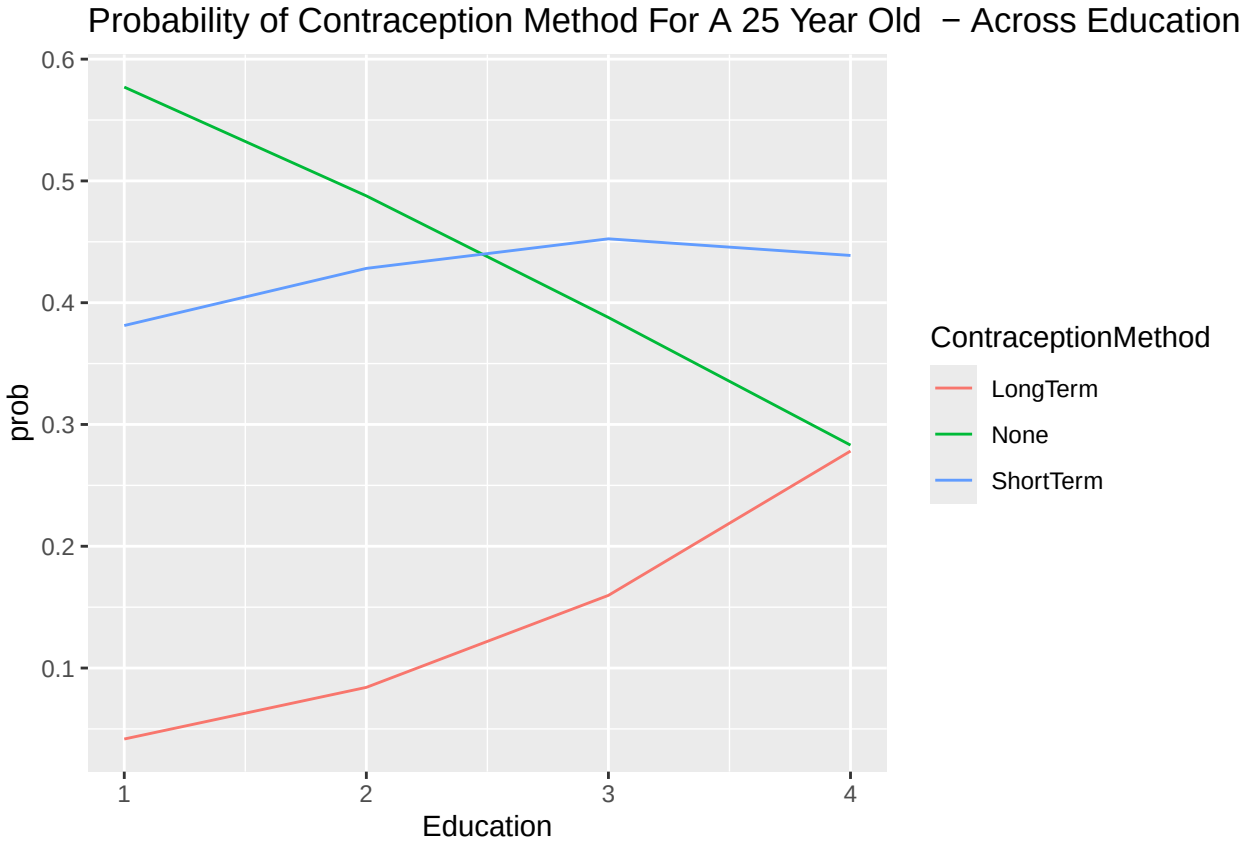
```r
newd <- data.frame(Education = rep(c(1,2,3,4), each = 48), Age = rep(c(12:59),
    4))

pp.ed <- cbind(newd, predict(model4, newdata = newd, type = "probs", se = TRUE))

by(pp.ed[, 3:5], pp.ed$Education, colMeans)
```

```
## pp.ed$Education: 1
##       None   LongTerm  ShortTerm
## 0.64928554 0.06153994 0.28917452
```

```
## ------------------------------------------------------------
## pp.ed$Education: 2
##      None  LongTerm ShortTerm
## 0.5507942 0.1247513 0.3244545
## ------------------------------------------------------------
## pp.ed$Education: 3
##      None  LongTerm ShortTerm
## 0.4294132 0.2310356 0.3395512
## ------------------------------------------------------------
## pp.ed$Education: 4
##      None  LongTerm ShortTerm
## 0.2984758 0.3785432 0.3229810
```

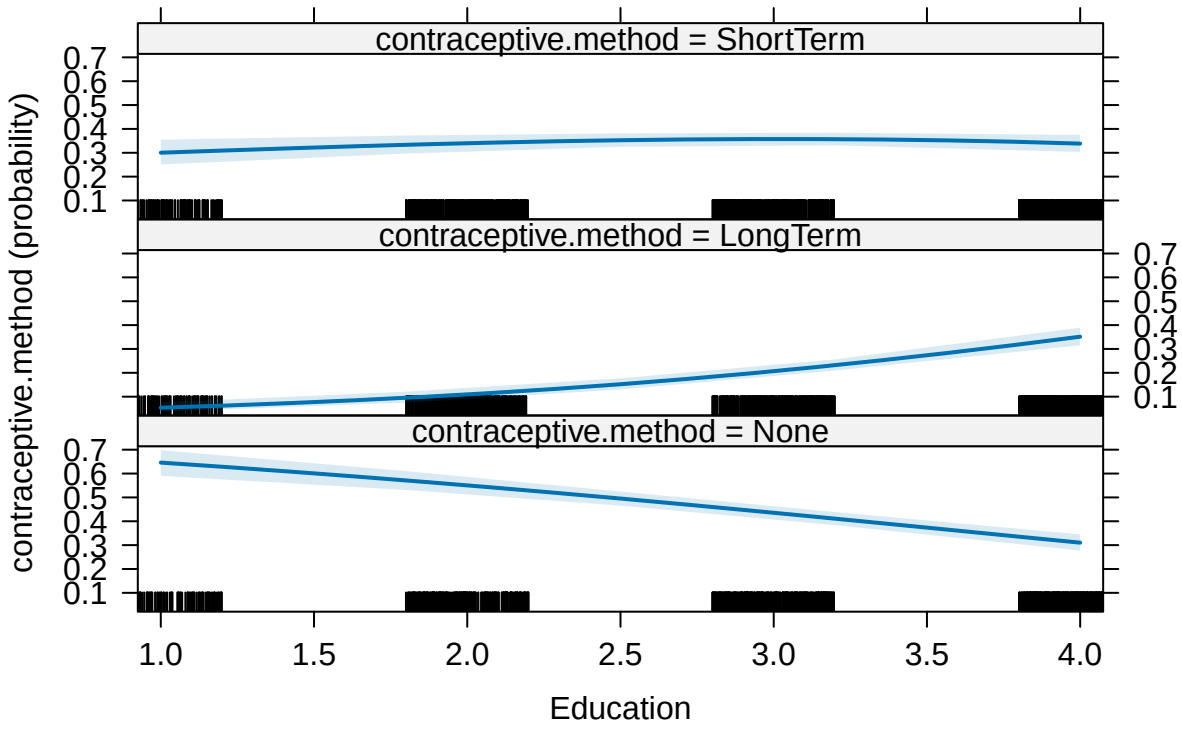
As age and education increase, so does the probability of using long term contraception.

```
ggplot(prob_plot, aes(x=Education, y=prob, group=ContraceptionMethod)) +
  geom_line(aes(color=ContraceptionMethod)) +
  ggtitle("Probability of Contraception Method For A 25 Year Old  - Across Education Levels")
```



There is a steep negative association between education and the use of no Contraception, and a positive association between education and long term contraceptive choice for a 25 year old similar to a participant in this dataset. The probability of short term contraception use rises slightly with higher education.

```
plot(Effect("Education",model4))
```




```
library(reshape2)
```

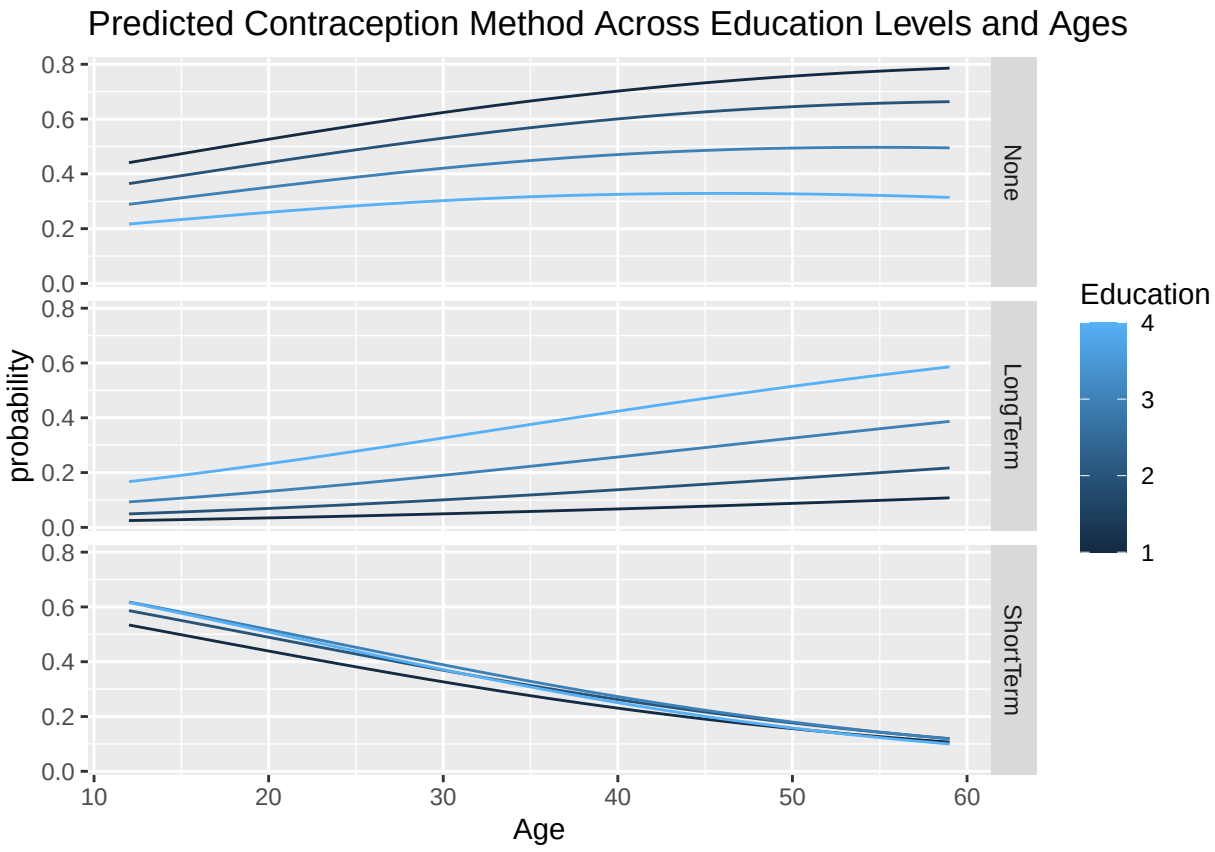
```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
lpp <- melt(pp.ed, id.vars = c("Education", "Age"), value.name = "probability")
tail(lpp)
```

```
##     Education Age  variable probability
## 571         4  54 ShortTerm  0.12943616
## 572         4  55 ShortTerm  0.12298340
## 573         4  56 ShortTerm  0.11680065
## 574         4  57 ShortTerm  0.11088134
## 575         4  58 ShortTerm  0.10521860
## 576         4  59 ShortTerm  0.09980529
```

```
## plot predicted probabilities across write values for each level of ses
## facetted by program type
ggplot(lpp, aes(x = Age, y = probability, color = Education)) + geom_line(aes(group=Education)) + facet_
  ggtitle("Predicted Contraception Method Across Education Levels and Ages")
```


Predicted Contraception Method Across Education Levels and Ages
probability
0.0
0.2
0.4
0.6
0.8
None
LongTerm
ShortTerm
10
20
30
40
50
60
Age
Education
4
3
2
1