Homework 3

Nathalie Hanner

21.05.2020

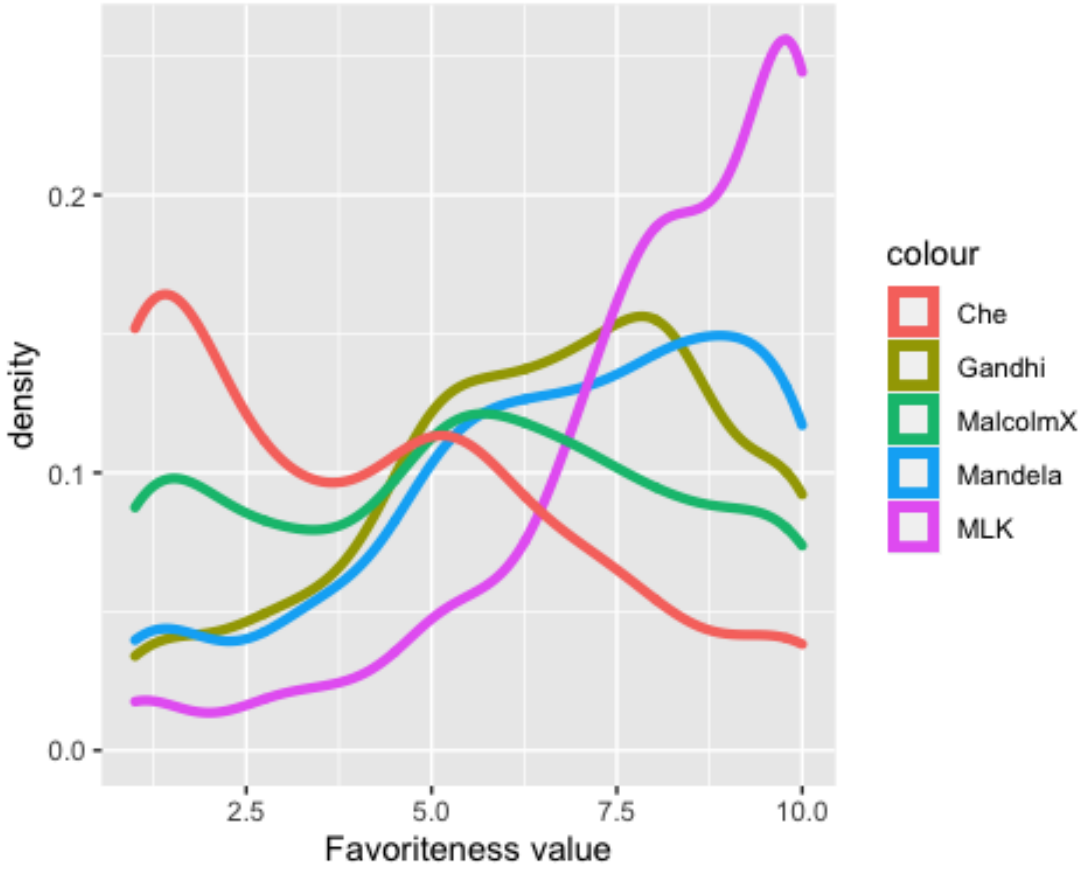
Quantitative Methods in Political Science
Leiden University

1.

```
rawData <- read.csv("https://raw.githubusercontent.com/babakrezaee/MethodsCourses/master/LeidenUniv_MAQM2020/Datasets/R%26P_RezaeeAsadzadeh_MethodsCourse.csv")

library(ggplot2)

ggplot(data=rawData)+
  geom_density(aes(x=Gandhi, colour= "Gandhi"), lwd=1.5)+
  geom_density(aes(x=Mandela, colour= "Mandela"), lwd=1.5)+
  geom_density(aes(x=MLK, colour= "MLK"), lwd=1.5)+
  geom_density(aes(x=MalcomX, colour= "MalcolmX"), lwd=1.5)+
```



```r
  geom_density(aes(x=Che, colour= "Che"), lwd=1.5)+
  labs(x = "Favoriteness value")

library(jtools)

ml_OLS1=lm(Nonviolent_effective~Mandela, data=rawData)
ml_OLS2=lm(Nonviolent_effective~Mandela+MalcomX, data=rawData)

summ(ml_OLS1)
```

```
## MODEL INFO:
## Observations: 404 (32 missing obs. deleted)
## Dependent Variable: Nonviolent_effective
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,402) = 27.73, p = 0.00
## R² = 0.06
## Adj. R² = 0.06
##
## Standard errors: OLS
## ------------------------------------------------
##                      Est.   S.E.   t val.      p
## ----------------- ------ ------ -------- ------
## (Intercept)          5.41   0.31    17.50   0.00
## Mandela              0.23   0.04     5.27   0.00
## ------------------------------------------------
```

```r
summ(ml_OLS2)
```

```
## MODEL INFO:
## Observations: 392 (44 missing obs. deleted)
## Dependent Variable: Nonviolent_effective
## Type: OLS linear regression
##
## MODEL FIT:
## F(2,389) = 15.07, p = 0.00
## R² = 0.07
## Adj. R² = 0.07
##
## Standard errors: OLS
## ------------------------------------------------
##                       Est.   S.E.   t val.      p
## ----------------- ------- ------ -------- ------
## (Intercept)          5.42   0.32    16.96   0.00
## Mandela              0.27   0.05     5.10   0.00
## MalcomX             -0.06   0.05    -1.18   0.24
## ------------------------------------------------
```

When looking at OLS1, the estimated coefficient for Mandela is 0.23 which indicates a positive relationship between a positive view on Mandela and nonviolent resistance. However, the slope is relatively small. In addition, the value for R^2 is very small with only 0.06 which means that only 6% of the observed variation can actually be explained by the input variables. Put differently, when trying to determine what factors influence support for nonviolent resistance as our Dependent Variable, the model's explanatory power is low as the likelihood of a view of Mandela being a main indicator of support for nonviolent resistance is only at 6%.

OLS 2 shows the same relationship but controlling for views on Malcolm X and nonviolent resistance. While the the values linking Mandela to nonviolent resistance is similar to the observation in OLS1 but slightly higher at 0.27, the estimated coefficient for Malcolm X is negative at -0.06 which indicates negative correlation, thereby implying that people advocating Malcom X are less likely to support nonviolent resistance.

Despite the low R^2 which tells us that no matter the variables, they do not explain much about support for nonviolent resistance overall, the results are statistically significant. We can see this by means of the low p-value which is 0, indicating that the correlation we wanted to test is probably true. A low p-value helps reject the null hypothesis which would imply that there is no correlation. We want to reject this null hypothesis which is possible with the case at hand. This does not "prove" our hypothesis but gives strong arguments in favour of it. Again in different words, the p-value shows us whether a chosen variable is useful to explain our case or not. Just because a models explanatory power is low, it does not mean that variables of minor impact in general are not contributing to a solution wherefore they can still be statistically significant.

Our image of how the view of a popular figure in the strive towards equality is linked to their support for nonviolent or violent resistance seems to comply with our expectation, namely that the convictions of the named leaders are to a certain degree represented by the person who was asked in the sample: However, this remains an assumption since famous quotes of leaders can easily be misinterpreted or abused. For instance Malcolm X was famous for his blatant rhetoric whereas Mandela is more known for hin peaceful speeches. Overall, these words can be interpreted more extremely by followers, leading to more or less extreme results in statistical samples.

In short, the variables deliver statistical significance but the explanatory power (or level of confidence), remains very low wherefore the model helps little to explain what determines support for nonviolent resistance.

2.

OLS1: Nonviolent_effective = 5.41 + .23 × 4 + ε Nonviolent_effective = 6.33

OLS2: Nonviolent_effective = 5.42 + .27 × 4 − .06 × 9 + ε Nonviolent_effective = 5.96

The result of OLS1 shows that with a value of 6.33, support for nonviolent resistance when having a low view on Mandela is still above the intercept, that is above the average or mean value of support for nonviolent resistance among the participants. When looking at OLS2 where we add a high view on Malcolm X, the result is with 5.96 lower as compared to OLS1. This indicates that while a low view on Mandela does not directly indicate low support for nonviolent resistance, the support is lower if this view goes together with a high view on Malcolm X.

3.

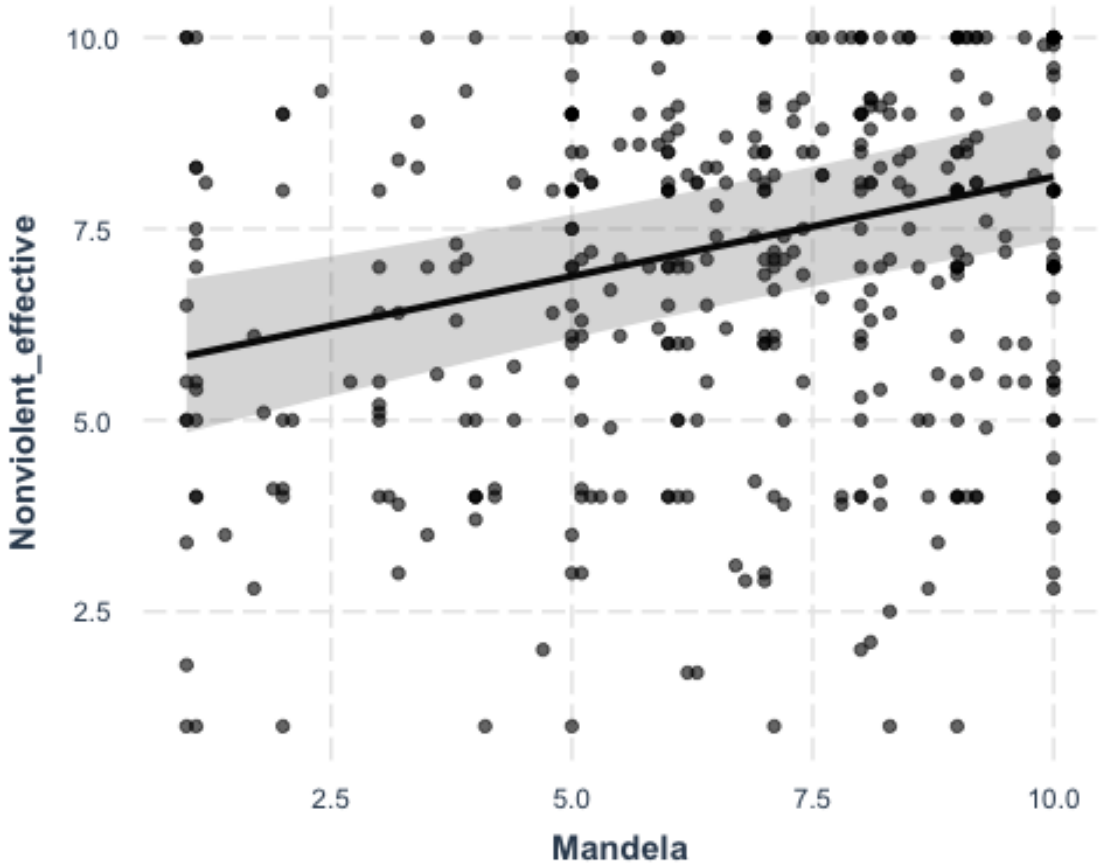
```
ml_OLS3=lm(Nonviolent_effective~Mandela+MalcomX+Marriage+Age+Gender, data=raw
Data)
summ(ml_OLS3)
```

```
## MODEL INFO:
## Observations: 392 (44 missing obs. deleted)
## Dependent Variable: Nonviolent_effective
## Type: OLS linear regression
##
## MODEL FIT:
## F(5,386) = 6.63, p = 0.00
## R² = 0.08
## Adj. R² = 0.07
##
## Standard errors: OLS
## ------------------------------------------------
##                         Est.   S.E.   t val.      p
## -------------------- ------- ------ -------- ------
## (Intercept)             6.03   0.83     7.25   0.00
## Mandela                 0.26   0.05     4.85   0.00
## MalcomX                -0.05   0.05    -1.14   0.26
## MarriageSingle         -0.16   0.40    -0.40   0.69
## Age                    -0.01   0.02    -0.30   0.76
## GenderMale             -0.40   0.24    -1.68   0.09
## ------------------------------------------------
```

The explanatory power of this correlation overall is very low because the value of R^2 is at only 8%. Accordingly, the variation of responses is high and it is difficult to obtain a reliable image of the factors impacting support for nonviolent resistance. However, as we assume that a variety of factors has an impact, we can still account for the statistical significance of other variables which will be discussed below.

Support for nonviolent resistance controlling for participant's view on Mandela:

```
effect_plot(ml_OLS3, pred = Mandela, interval = TRUE, plot.points = TRUE)
```

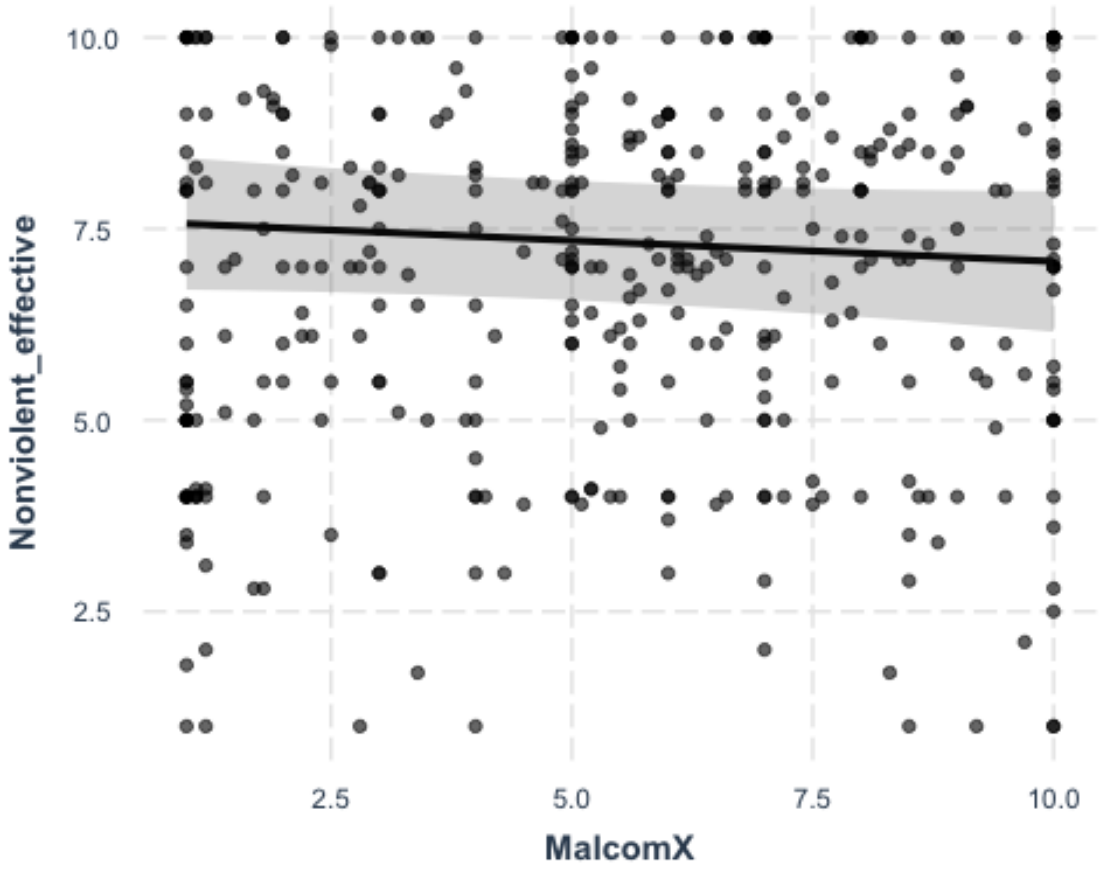


The plot shows that the higher participant's think of Mandela, the higher is the likelihood that they support nonviolent resistance. We can see what has been outlined before that the slope is positive, thereby illustrating a positive correlation between Mandela and nonviolent_effective.

As explained earlier, the statistical relevance with a p-value of 0 for this correlation is high.

Support for nonviolent resistance controlling for participant's view on Malcolm X:

```
effect_plot(ml_OLS3, pred = MalcomX, interval = TRUE, plot.points = TRUE)
```

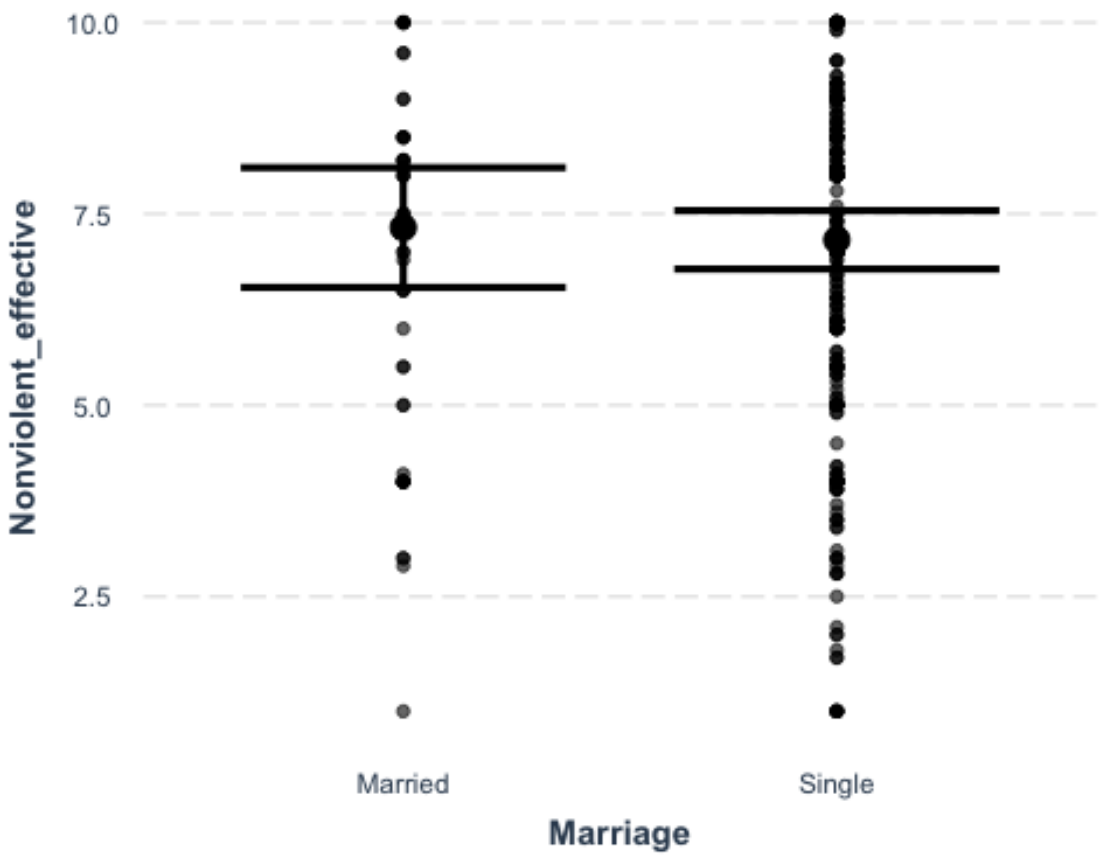


In contrast to the previous graph, the graph controlling for Malcolm X shows a negative impact as seen by the flatter slope. This indicates that the change between a weak and a strong opinion about Malcolm X in relation to the support for nonviolent resistance is small. The intercept of the line is higher than in the previous graph which shows how a weak view on Malcolm X is linked to high support for nonviolent resistance. Since this is a control variable and therefore added to the data on Mandela, the function has shifted upwards. The same explanation accounts for the flatter slope. A high view on Malcolm X is associated with lower support for nonviolent resistance. Added to the results for Mandela, this decreases the overall support which is why the end of the line is lower than it is when looking at Mandela only.

The statistical significance here is low with a p-value of 0.26 which is too high to offer a valid explanation on support for nonviolent resistance. This, however, is not surprising, taking into account the personality and nature of Malcolm X and his actions.

Support for nonviolent resistance controlling for participant's marital status:

```
effect_plot(ml_OLS3, pred = Marriage, interval = TRUE, plot.points = TRUE)
```

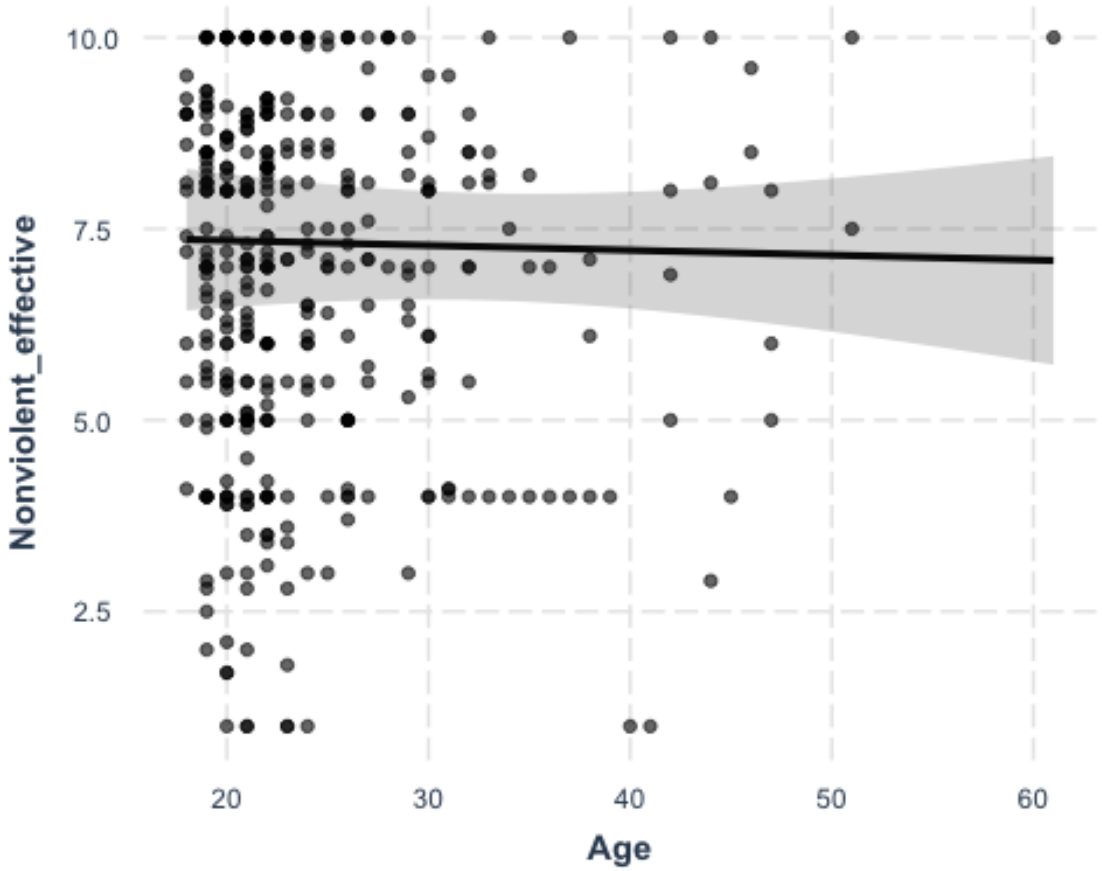


When controlling for the marital status of the respondends, we can see that a larger part of married people would support nonviolent resistance. The largest amount of single people to support nonviolent resistance overlaps with the results of married people but is smaller. This is observable as the right boxplot showing the responses of singles is smaller than the right one mapping the results for married people. Overall, it is also evident that single people offer a greater variety of responses, whereas married people seem relatively clear on their view with a high support for nonviolent resistance.

The statistical significance for this control variable is extremely low with a p-value of 69%. This means with a chance of 69%, our hypothesis to find a correlation here is rejected. Marital status accordingly does not offer a good explanation on support for nonviolent resistance.

Support for nonviolent resistance controlling for participant's age:

```
effect_plot(ml_OLS3, pred = Age, interval = TRUE, plot.points = TRUE)
```

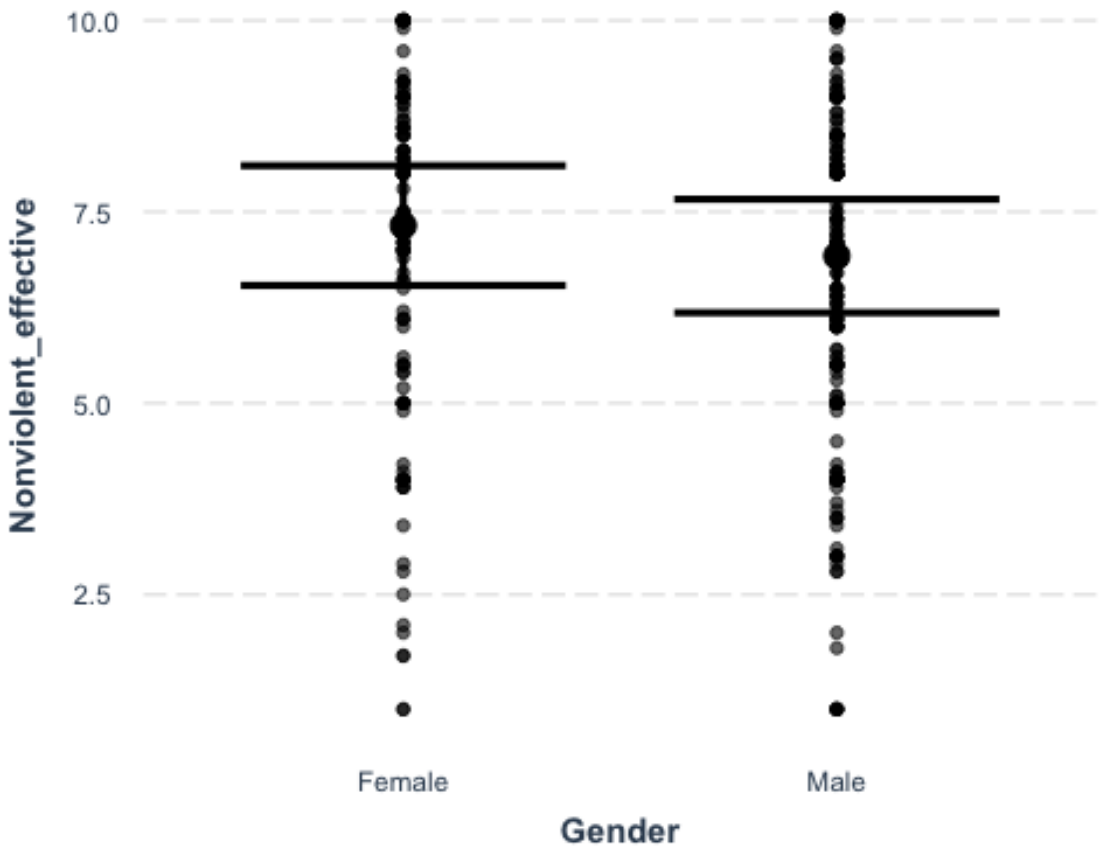


The scatterplot shows a number of dots and a straight line which indicates that age is hardly correlated with support for nonviolent resistance. The plot shows that the largest number of participants is aged between 20 and thirty and a small number of people between 30 and 40. Above 40, there are very few respondents. The slight downward slope of the line seems to portray a decrease in support for nonviolent resistance. However, due to the few respondents in high age groups, this observation is hardly representative. The grey areas surrounding the line are relatively big, which also shows that the level of confidence in the results is not too precise.

When controlling for age, the p-value is the highest among variables we are looking at. With a chance of 76%, this indicator can be rejected as explanation. Accordingly, the opinion on nonviolent resistance does not change much as people get older.

Support for nonviolent resistance controlling for participant's gender:

```
effect_plot(ml_OLS3, pred = Gender, interval = TRUE, plot.points = TRUE)
```



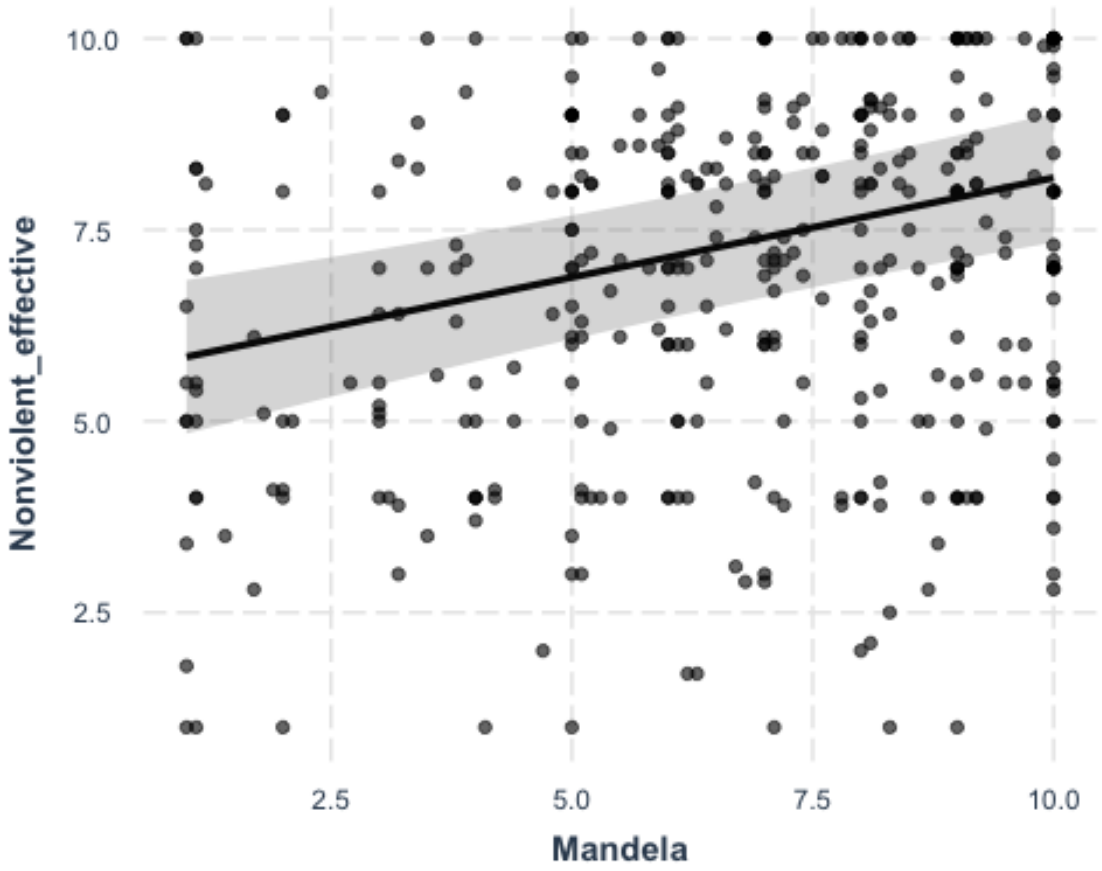
When controlling for gender, we can see that the two boxes of the boxplot are approximately the same size, yet the one showing the responses of females is higher which portrays a higher support of nonviolent resistance among females. The responses of males partly overlap but are overall lower, showing a smaller support. When looking at the dots, we can also see more in the levels below 5 on the male column. Few females have yielded a below 5 result. Therefore, females are more likely to support nonviolent resistance than males.

Here, we find a low p-value of 9%. It is not as low as we wish for p-values to be, but low enough to observe a relatively significant impact of gender on the opinion of nonviolent resistance. This variable therefore offers some valuable insights into relevant factors that can account for support for nonviolent resistance.

4.

```
library(broom)
library(sjPlot)

effect_plot(ml_OLS3, pred = Mandela, interval = TRUE, plot.points = TRUE)
```



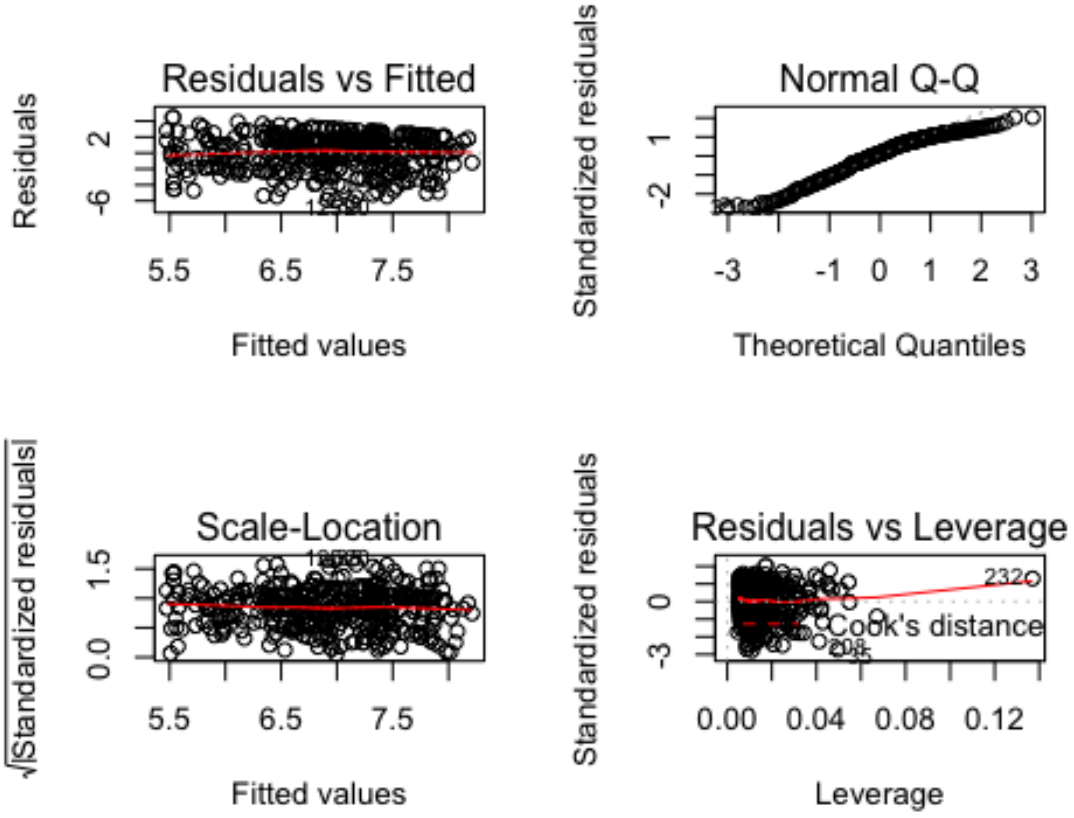

```
OLS3_broomed <- augment(ml_OLS3)

head(OLS3_broomed)
```

```
## # A tibble: 6 x 14
##   .rownames Nonviolent_effe… Mandela MalcomX Marriage   Age Gender .fitted
##   <chr>                <dbl>   <dbl>   <dbl> <fct>    <int> <fct>    <dbl>
## 1 1                     8.20    6.90    7.60 Married     33 Male      6.80
## 2 2                    10       9       1    Single      27 Male      7.59
## 3 3                     9.5    10      10    Single      30 Male      7.34
## 4 4                     4       4       4    Single      34 Female    6.48
## 5 5                     9       5      10    Married     29 Male      6.21
## 6 7                     7       3.8     5.30 Single      35 Male      5.95
## # … with 6 more variables: .se.fit <dbl>, .resid <dbl>, .hat <dbl>,
## #   .sigma <dbl>, .cooksd <dbl>, .std.resid <dbl>
```

```
par(mfrow = c(2,2))

plot(ml_OLS3)
```



Residuals vs. Fitted: The linearity assumption is fulfilled. We see a scatterplot and a relatively straight line through the sample. Accordingly, we have a linear relationship at hand.

Normal Q-Q: The residuals also appear to be normally distributed because they portray a relatively clear, positively sloped, line. Considering that an approximately 45° line indicates a good fit, despite the minor deviation towards the upper part of the line, this assumption counts fulfilled as well.

Scale-location: Here we check for homoskedacity, meaning the homogeneity of variance of the residuals, which is our desired outcome. In other words, we aspire to observe a relatively even spread of the sample results around a horizontal line. The graph shows that this requirement is met as we see a very even horizontal line around which all our points representing survey data is located.

Residuals vs. Leverage: At this point we control for outliers that might have an impact on the results of our regression analysis and thereby impact the accuracy of our results. Here, we see a slight upward slope towards the end of the red line. However, we again aspire to have a line that is as straight and horizontal as possible. The reason for this are outliers. In order

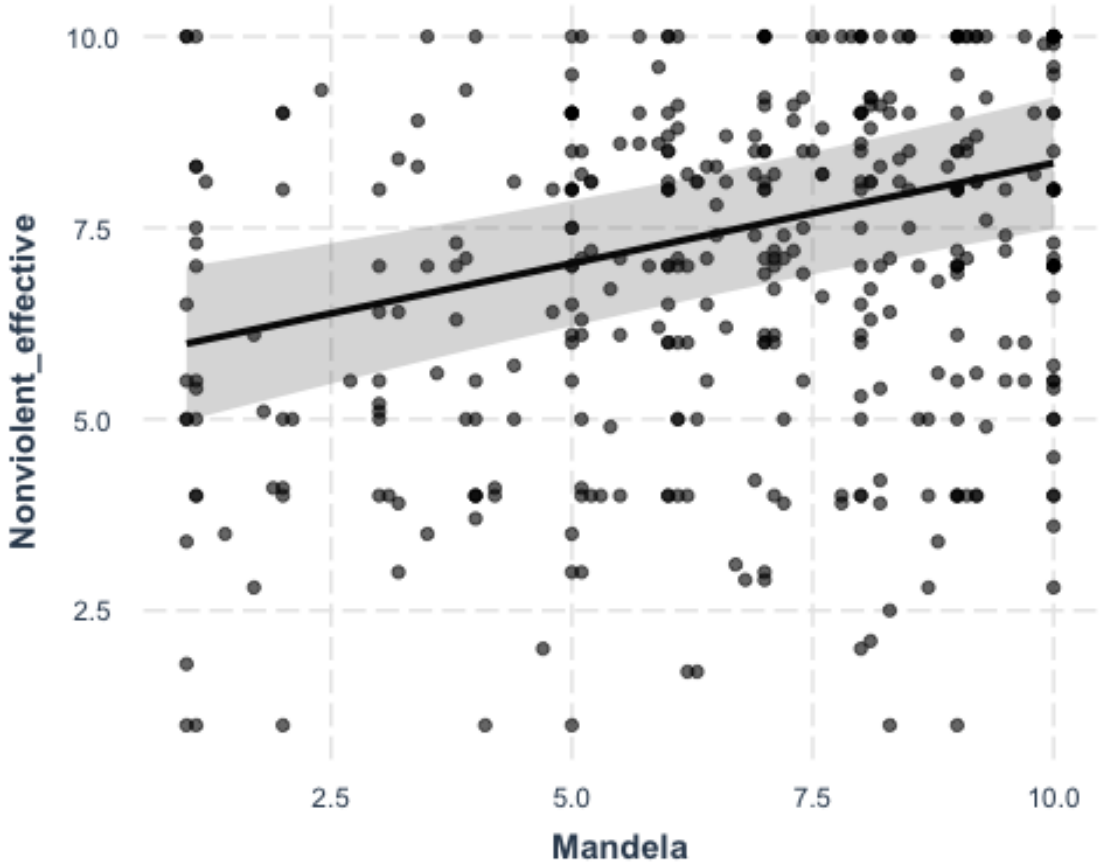
to fix this problem, we remove the outliers and run the analysis again without these few distorting values from our sample. The fixed plot can be seen below.

Remove outliers to fix the plots:

```
rawData_noOutliers <- rawData [-c(35, 208, 232), ]

ml_OLS4=lm(Nonviolent_effective~Mandela+MalcomX+Marriage+Age+Gender, data=raw
Data_noOutliers)

effect_plot(ml_OLS4, pred = Mandela, interval = TRUE, plot.points = TRUE)
```



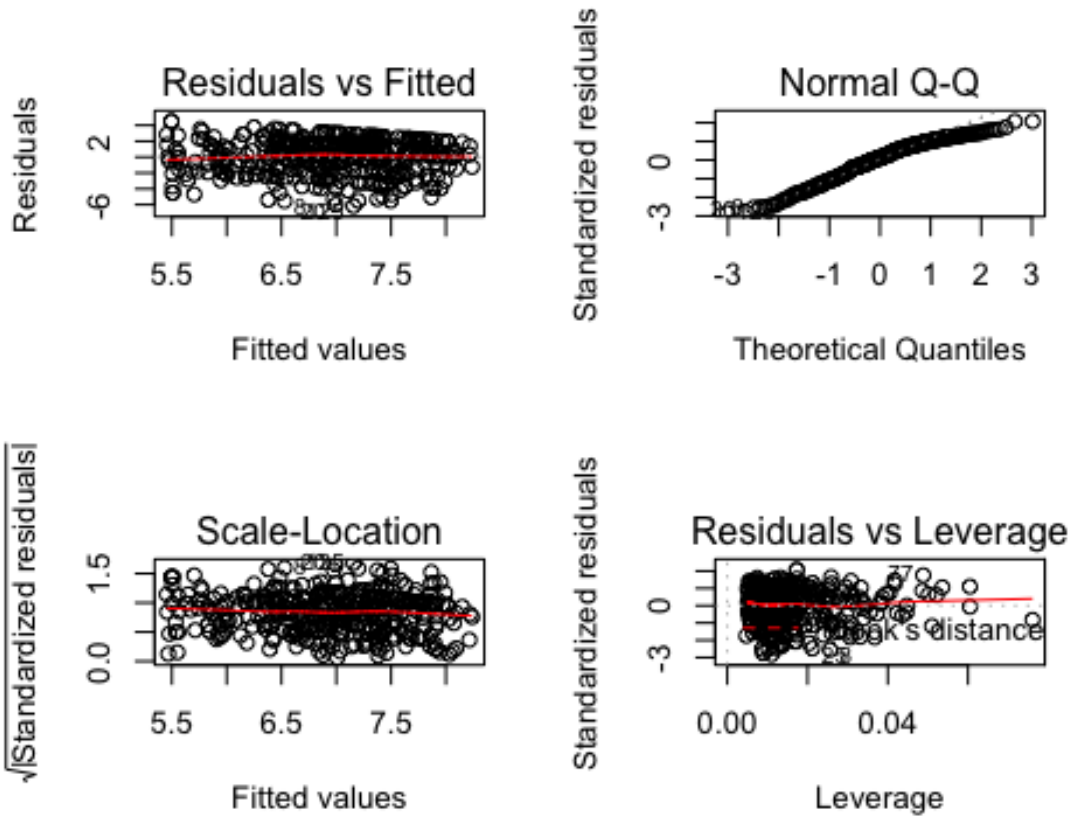
```
OLS4_broomed <- augment(ml_OLS4)

head(OLS4_broomed)
```

```
## # A tibble: 6 x 14
##   .rownames Nonviolent_effe… Mandela MalcomX Marriage   Age Gender .fitted
##   <chr>                <dbl>   <dbl>   <dbl> <fct>    <int> <fct>    <dbl>
## 1 1                     8.20    6.90    7.60 Married     33 Male      6.91
## 2 2                    10       9       1    Single      27 Male      7.56
## 3 3                     9.5    10      10    Single      30 Male      7.33
## 4 4                     4       4       4    Single      34 Female    6.46
## 5 5                     9       5      10    Married     29 Male      6.33
## 6 7                     7       3.8     5.30 Single      35 Male      5.90
## # … with 6 more variables: .se.fit <dbl>, .resid <dbl>, .hat <dbl>,
## #   .sigma <dbl>, .cooksd <dbl>, .std.resid <dbl>
```

```
par(mfrow = c(2,2))

plot(ml_OLS4)
```



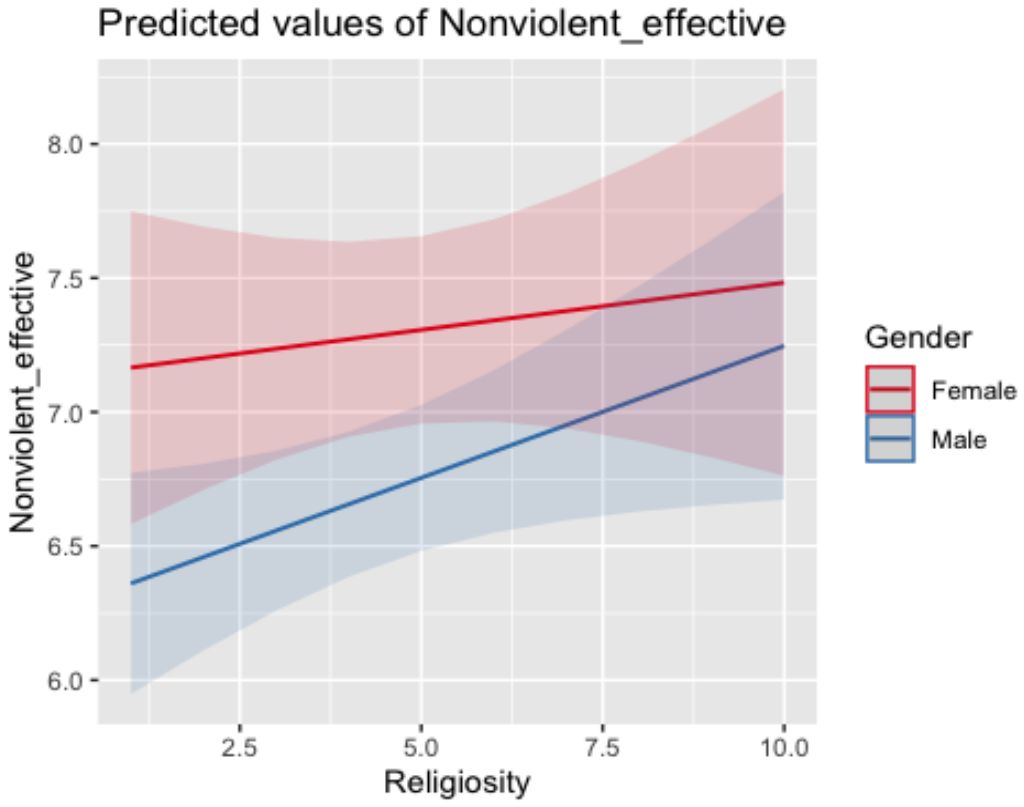
Now that we removed the outliers, we can see that the red line in the *Residuals vs. Leverage* plot has flattened and portrays a more accurate representation of our findings. In order to achieve this, we had to run the regression again with OLS4 which made use of the new data set excluding the outliers.

5.

```
ml_OLS5=lm(Nonviolent_effective~Religiosity+Gender+Religiosity*Gender, data=r
awData)
summ(ml_OLS5)
```

```
## MODEL INFO:
## Observations: 435 (1 missing obs. deleted)
## Dependent Variable: Nonviolent_effective
## Type: OLS linear regression
##
## MODEL FIT:
## F(3,431) = 3.99, p = 0.01
## R² = 0.03
## Adj. R² = 0.02
##
## Standard errors: OLS
## ------------------------------------------------------------
##                                 Est.   S.E.   t val.      p
## ---------------------------- ------- ------ -------- ------
## (Intercept)                     7.13   0.35    20.40   0.00
## Religiosity                     0.04   0.06     0.57   0.57
## GenderMale                     -0.87   0.43    -2.03   0.04
## Religiosity:GenderMale          0.06   0.08     0.81   0.42
## ------------------------------------------------------------
```

```
library(sjPlot)
plot_model(ml_OLS5, type = "pred", terms = c("Religiosity", "Gender"))
```



To understand the overall results here and the significant difference to what happens when exchanging *gender* with *marriage* in this function, I also explain what is observable in the initial plot.

The plot shows the support for nonviolent resistance of females (red) and males (blue), controlling for religiosity, that means how their support respectively changes the more religious they are. The line indicating the results of female participants is significantly higher than the line of male participants which shows that overall, females show higher support for nonviolent resistance than males. This finding, we have already seen earlier when controlling for gender in exercise 3.

The important difference here is that we can see that the two lines are getting to closer to each other the higher the participant's score on relgiosity. In other words, we can see that both males and females are more likely to support nonviolent resistance when they are more religious. Males that are not or only a bit religious are siginificantly less likely to support nonviolent resistance than are women with the same scores whereas both males and females scoring high in religiosity are also scoring high in their support of nonviolent resistance.

We also observe a large confidence interval which is shown by the light blue and pink areas surrounding the male and female line respectively. The confidence interval for females is larger overall. It tells us that if the project was repeated on a new sample with new participants, the results of new responses would most likely be within these shaded areas. In this sense, we can also observe that the overlap of the confidence intrevals is small for low levels of religiosity and nonviolent resistance support but big for higher levels of both. This again indicates that the responses/opinions on support vor nonviolent resistance of males and females overlap if they are more religious, meaning that overall, people that are more religious are generally more likely to be in favour of nonviolent resistance.

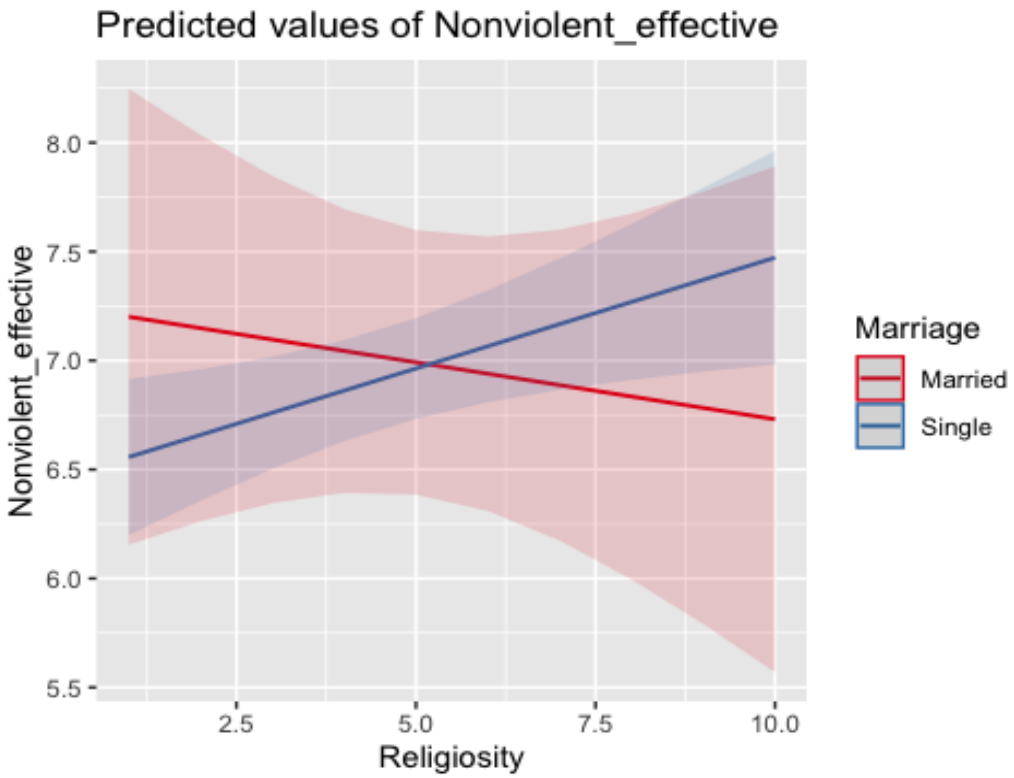
The p-value for religiousity: GenderMale is relatively high with a value of 0.42 which decreases the statistical significance of the correlation. Only the p-value for gender is low enough with 0.04 that we can deduce statistical significance from it whereas religion seems relatively unsignficant at a value of 0.057. This explains why the overall significance is imapcted.

Since the R^2 is also very low, indicating that the explanatory power of the model is not very strong. Only 2% of the elements that can impact support for nonviolent resistance are explicable with this model.

```r
ml_OLS6=lm(Nonviolent_effective~Religiosity+Marriage+Religiosity*Marriage, da
ta=rawData)
summ(ml_OLS6)
```

```
## MODEL INFO:
## Observations: 435 (1 missing obs. deleted)
## Dependent Variable: Nonviolent_effective
## Type: OLS linear regression
##
## MODEL FIT:
## F(3,431) = 2.23, p = 0.08
## R² = 0.02
## Adj. R² = 0.01
##
## Standard errors: OLS
## ------------------------------------------------------------------
##                                    Est.   S.E.   t val.      p
## ------------------------------ ------- ------ -------- ------
## (Intercept)                        7.25   0.62    11.67   0.00
## Religiosity                       -0.05   0.10    -0.50   0.62
## MarriageSingle                    -0.80   0.66    -1.21   0.23
## Religiosity:MarriageSingle         0.15   0.11     1.38   0.17
## ------------------------------------------------------------------
```

```r
plot_model(ml_OLS6, type = "pred", terms = c("Religiosity", "Marriage"))
```



When looking at how religiousity impacts the support for nonviolent resistance of married and single people, the results are somewhat more surprising as the plot illustrates. The lines showing responses of married and single people cross approximately in the middle, portraying an interesting overlap of opinions as well as developments in opposite directions. Previously, when controlling for religion on gender, the correlation was positive in both cases and just increased in different intensity wheras here we observe opposing developments.

We have earlier seen when controlling only for marital status that a higher percentage of married people would support nonviolent resistance, the impact of religiousity as a countering effect. As the red line showing responses of married people displays, those less religious are more supportive of nonviolence than are highly religious ones. We can see that be means of the negative slope.

In contrast, single people with low levels of religiousity are less likely to support nonviolent resistance, yet their support rises significantly when they are more religious.

The confidence interval of the red line portraying married people is extremely wide and overlaps entirely with the confidence interval of the blue line for single people. Accordingly, the variance of possible answers of married people is very high and there seems to be no clear-cut pattern of opinions. This is also shown by the R^2 which is only at 2% and therefore indicates that this combination of variables is not suitable to explain support for nonviolent resistance. The p-values of all three variables, Religiousity, Marriage and the combination of Religiousiy and Marriage are high. While they are lower for the latter two at 0.23 and 0.17 respectively, they are still to high to account for statistical significance.

Overall, the results portrayed in this graph are interesting but turned out to be not necessarily representative or reliable for our case.