Final Project

Christopher Byrd

2024-12-13

Contents

Section 1: Hypothesis

The goal of this research paper is to determine how attendance of church and gender affects opinion on seperation of church and state.

$Null_1$: There will be no significant difference in opinion on church and state seperation based upon sex.
$Research_1$: There will be a significant difference in opinion on church and state seperation based upon sex.

$Null_2$: There will be no significant difference in opinion on church and state seperation based upon church attendance.
$Research_2$: There will be a significant difference in opinion on church and state seperation based upon church attendance.

$Null_3$: There will be no significant difference in opinion on church and state seperation based upon sex and church attendance.
$Research_3$: There will be a significant difference in opinion on church and state seperation based upon sex and church attendance.

I got a little carried away on this. Sorry Dr. K!

Section 2: Data

The data being utilized is from the Baylor Survey of Religious Life . Participants asked to respond to the question listed in the following code book:

Question or Description	Possible Answers
What is your preferred gender?	1) Female 2) Male
The federal government should enforce strict separation of church and state.	1) Strongly Disagree 2) Disagree 3) Agree 4) Strongly Agree 8) Undecided*
How often do you attend religious services at a place of worship?	1) Never 2) Less than once a year 3) Once or twice a year 4) Several times a year 5) Once a month 6) 2 to 3 times a month 7) About once a week 8) Several times a week**

*Note: The original code book says 5 is undecided. This is incorrect.
**Note: No responses were recorded.

Section 3: Method

To evaluate the data, an analysis of variance (ANOVA) was calculated to see where variance might lie in the model and whether said variance was significant.

```
aovResults <- aov(formula = churchState ~ sex + attendance, df)
anova(aovResults)
```

Next, I ran `cor()` to determine Pearson's r and see where correlation lies within the variables. I indexed 1:3 as to avoid running `cor()` on fitted values and residuals from `lm()`, as they are also contained within the `df`. This is only necessary when running the test a second time after starting up an R instance.

```
cor(df[1:3])
```

Following this, a linear model was calculated as church and state opinion regressed upon sex and attendance.

```
lmresult <- lm(formula = churchState ~ sex + attendance, df)
summary(lmresult)
```

After this, I created 3 graphs, with the regression for sex and attendance depicted below. An individual graph for church and state regressed upon sex and attendance separately are included.

```
ggplot(df, aes(x = fitted, y = churchState)) +
  geom_count()+
  scale_size_area()+
  geom_smooth(method = 'lm', color = '#679df5') +
  labs(title = "Opinion on Seperation of Church & State\
       Based Upon Fitted Values") +
  xlab("Fitted values (see section 2-4 for more info)") +
  ylab("Opinion on Seperation of Church & State") +
  stat_poly_eq(use_label(c("eq", "R2")), rr.digits = 4,vjust = 1.3) +
  theme_pander() +
  theme(plot.margin = margin(t = 15, r = 15, b = 15, l = 15, unit = "pt"))
```

Lastly, I performed all these analyses a second time, this time omitting the response 8 for church and state from the tests. Responses of 8 only make up 6.5 of responses and cause a noticeable skew to the mean in the way they are encoded. More explanation will be provided in section 5.

Section 4: Results

Results are as follows and are broken up by inclusion or omission of undecided responses.

Section 4.1: With Undecided Discussed first are tables 1 and 2. When regressed upon sex and attendance, our ANOVA shows statistically significant variation in the model caused by attendance, but not sex. Pearson's r demonstrates an extremely weak, indirect correlation between church and state when considering sex or attendance.

Moving to the linear regression, table 3 shows only attendance is a significant predictor within the model when both sex and attendance are considered.

Next, tables 4 and 5 are discussed, where each individual independent variable is regressed. When only attendance is considered, the results are identically significant. When sex is considered alone, it swaps from having a weak direct relationship to having a weak indirect relationship. This indirect relationship is still insignificant.

Tables 3-5 are represented graphically in figures 1 through 3.

Section 4.2: With Undecided Omitted Discussed next are tables 6 and 7. When regressed upon sex and attendance, our ANOVA shows statistically significant variation in the model caused by both independent variables. Pearson's r demonstrates a weak, indirect correlation between church and state when considering sex, and a moderate indirect relationship when considering attendance.

Moving to the linear regression, table 8 shows both variables are significant predictors within the model when both sex and attendance are considered.

Next, tables 9 and 10 are discussed, where each individual independent variable is regressed. When only attendance is considered, the results are identically significant. When sex is considered alone, it's indirect relationship moderately increases in strength, but becomes more significant.

Tables 8-10 are represented graphically in figures 4 through 6.

Section 5: Discussion and Conclusion

From these results, when considering undecided responses, we can accept Null_1, Research_2, and Research_3. Thus, the converses are rejected.

When undecided responses are omitted, we can accept Research_1, Research_2, and Research_3 while rejecting their corresponding null hypotheses.

As to why I decided to run my analyses twice, I was curious how the undecided results affected the overall picture. The answer is quite significantly. They skewed both strength of correlation and correlation direction. I think coding them as an 8 was not the best idea, since this introduces a larger difference in mean than if they had been coded as a number "closer to the pack".

Section 6: Graphics and Visuals

Section 6.1: Visuals Containing Undecided

Section 6.1.1: Tables

```
## Analysis of Variance Table
##
## Response: churchState
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## sex           1    0.07   0.070    0.03    0.8626
## attendance    1   61.26  61.264   26.04 3.868e-07 ***
## Residuals  1235 2905.55   2.353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 1: ANOVA table.

```
##                      sex  churchState  attendance
## sex          1.000000000 -0.004874326  0.06787925
## churchState -0.004874326  1.000000000 -0.14369797
## attendance   0.067879253 -0.143697972  1.00000000
```

Table 2: Correlation matrix with Pearson's r.

```
##
## Call:
## lm(formula = churchState ~ sex + attendance, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6196 -1.0092 -0.0964  0.4676  5.0062
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.60420    0.08449  42.660  < 2e-16 ***
## sex          0.01536    0.08844   0.174    0.862
## attendance  -0.08720    0.01709  -5.103 3.87e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.534 on 1235 degrees of freedom
## Multiple R-squared:  0.02067,    Adjusted R-squared:  0.01909
## F-statistic: 13.04 on 2 and 1235 DF,  p-value: 2.5e-06
```

Table 3: Linear regression based upon fitted values.

```
##
## Call:
## lm(formula = churchState ~ attendance, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6124 -1.0034 -0.0904  0.4746  4.9966
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.61241    0.06999  51.611  < 2e-16 ***
## attendance  -0.08700    0.01704  -5.105 3.83e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.533 on 1236 degrees of freedom
## Multiple R-squared:  0.02065,    Adjusted R-squared:  0.01986
## F-statistic: 26.06 on 1 and 1236 DF,  p-value: 3.828e-07
```

Table 4: Linear regression based upon church attendance.

```
##
## Call:
## lm(formula = churchState ~ sex, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3416 -1.3263 -0.3263  0.6584  4.6737
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.34160    0.06768  49.372   <2e-16 ***
## sex         -0.01527    0.08912  -0.171    0.864
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.549 on 1236 degrees of freedom
## Multiple R-squared:  2.376e-05,  Adjusted R-squared:  -0.0007853
## F-statistic: 0.02937 on 1 and 1236 DF,  p-value: 0.864
```

Table 5: Linear regression based upon sex.

Section 6.1.2: Graphs

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

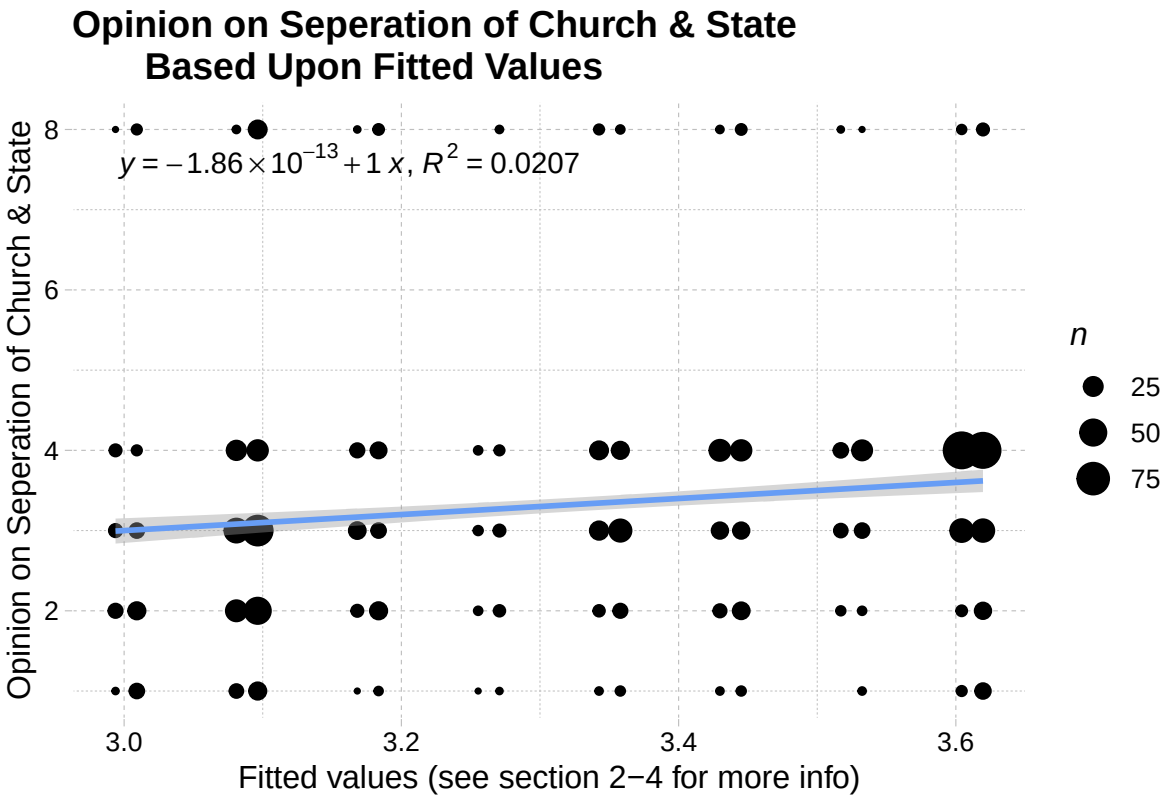

Figure 1: Scatter plot with linear regression based upon fitted values.

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

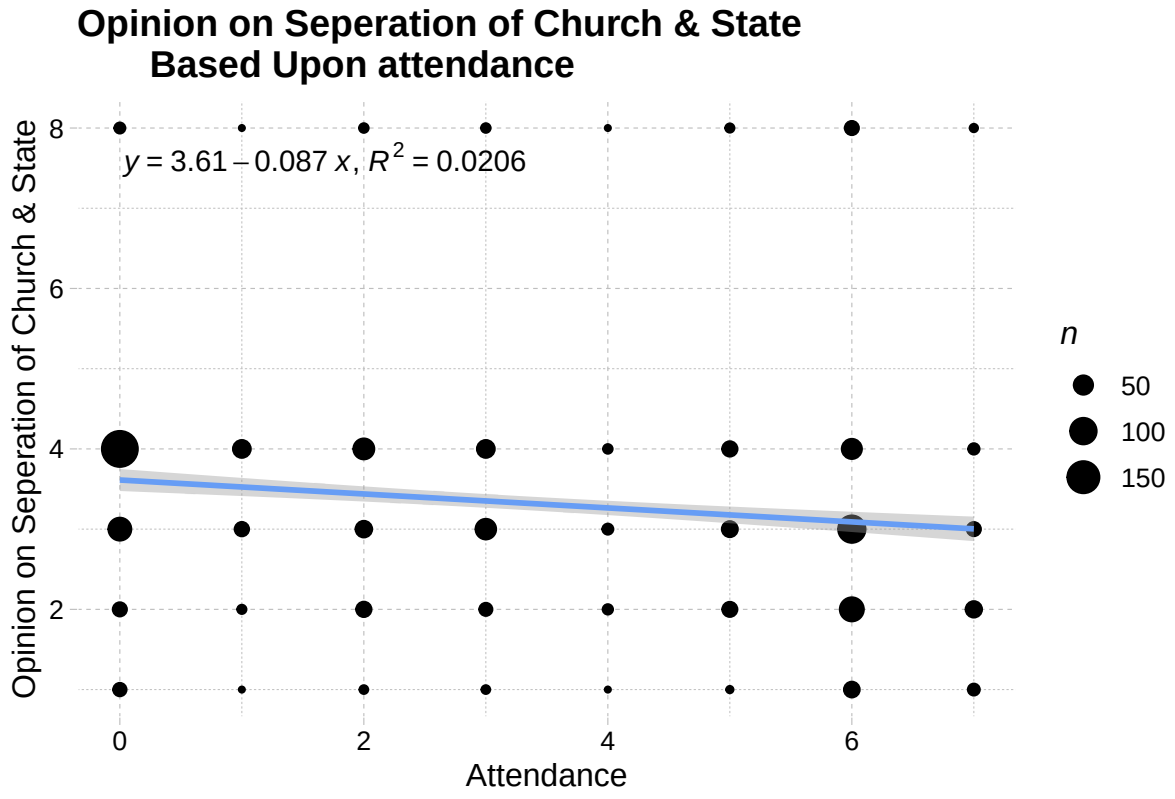

Figure 2: Scatter plot with linear regression based upon attendance.

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

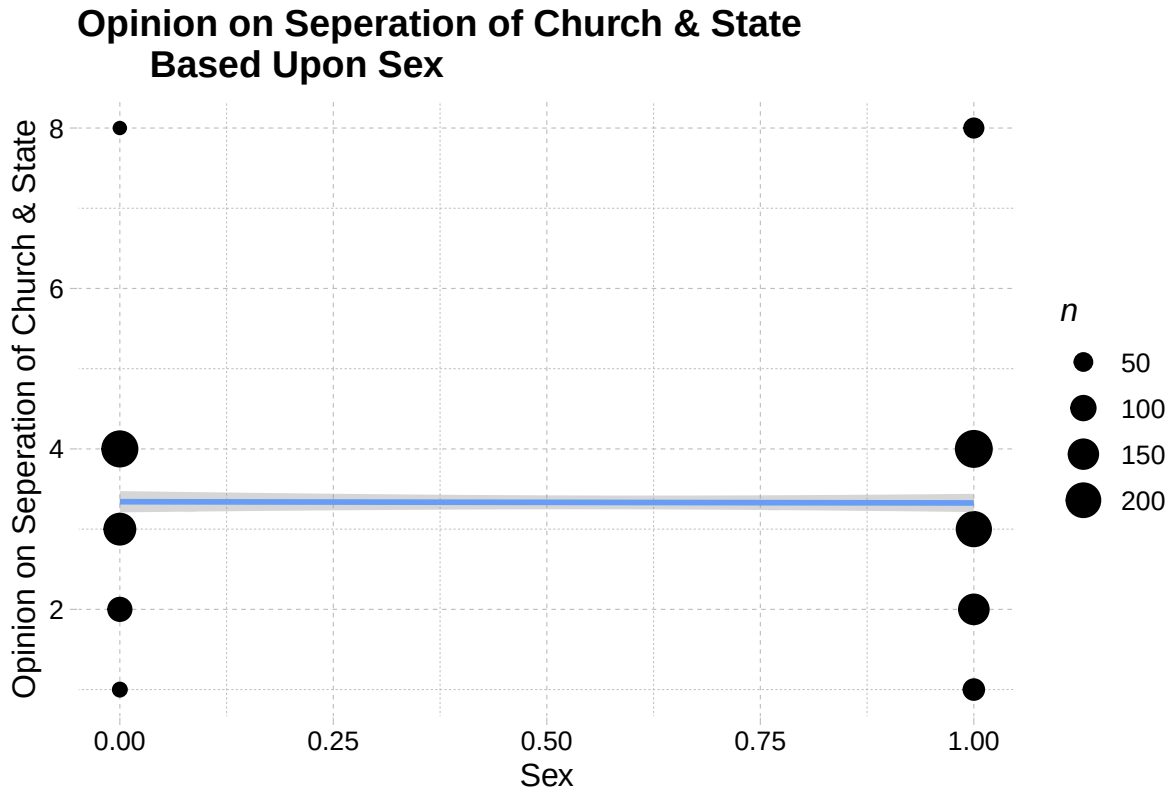

Figure 3: Scatter plot with linear regression based upon sex

Section 6.2: Visuals with Undecided Omitted

Section 6.2.1: Tables

```
## Analysis of Variance Table
##
## Response: churchState
##              Df Sum Sq Mean Sq F value    Pr(>F)
## sex           1  15.31  15.310  18.608 1.743e-05 ***
## attendance    1 114.14 114.142 138.725 < 2.2e-16 ***
## Residuals  1154 949.51   0.823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 6: ANOVA table.

```
##                     sex churchState  attendance
## sex          1.00000000  -0.1191218  0.05128999
## churchState -0.11912181   1.0000000 -0.33093379
## attendance   0.05128999  -0.3309338  1.00000000
```

Table 7: Correlation matrix with Pearson's r.

```
##
## Call:
## lm(formula = churchState ~ sex + attendance, data = dfs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5084 -0.5723  0.2282  0.6911  1.5505
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.50845    0.05130  68.397  < 2e-16 ***
## sex         -0.19956    0.05388  -3.704 0.000222 ***
## attendance  -0.12277    0.01042 -11.778  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9071 on 1154 degrees of freedom
## Multiple R-squared:   0.12,  Adjusted R-squared:  0.1185
## F-statistic: 78.67 on 2 and 1154 DF,  p-value: < 2.2e-16
```

Table 8: Linear regression based upon fitted values.

```
##
## Call:
## lm(formula = churchState ~ attendance, data = dfs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4018 -0.6532  0.2220  0.5982  1.4715
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.40176    0.04268   79.71   <2e-16 ***
## attendance  -0.12475    0.01047  -11.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9121 on 1155 degrees of freedom
## Multiple R-squared:  0.1095, Adjusted R-squared:  0.1087
## F-statistic:   142 on 1 and 1155 DF,  p-value: < 2.2e-16
```

Table 9: Linear regression based upon church attendance.

```
##
## Call:
## lm(formula = churchState ~ sex, data = dfs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.13745 -0.90534  0.09466  0.86255  1.09466
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  3.13745    0.04283  73.252  < 2e-16 ***
## sex         -0.23211    0.05692  -4.077 4.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9596 on 1155 degrees of freedom
## Multiple R-squared:  0.01419,    Adjusted R-squared:  0.01334
## F-statistic: 16.63 on 1 and 1155 DF,  p-value: 4.866e-05
```

Table 10: Linear regression based upon sex.

Section 6.2.2: Graphs

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

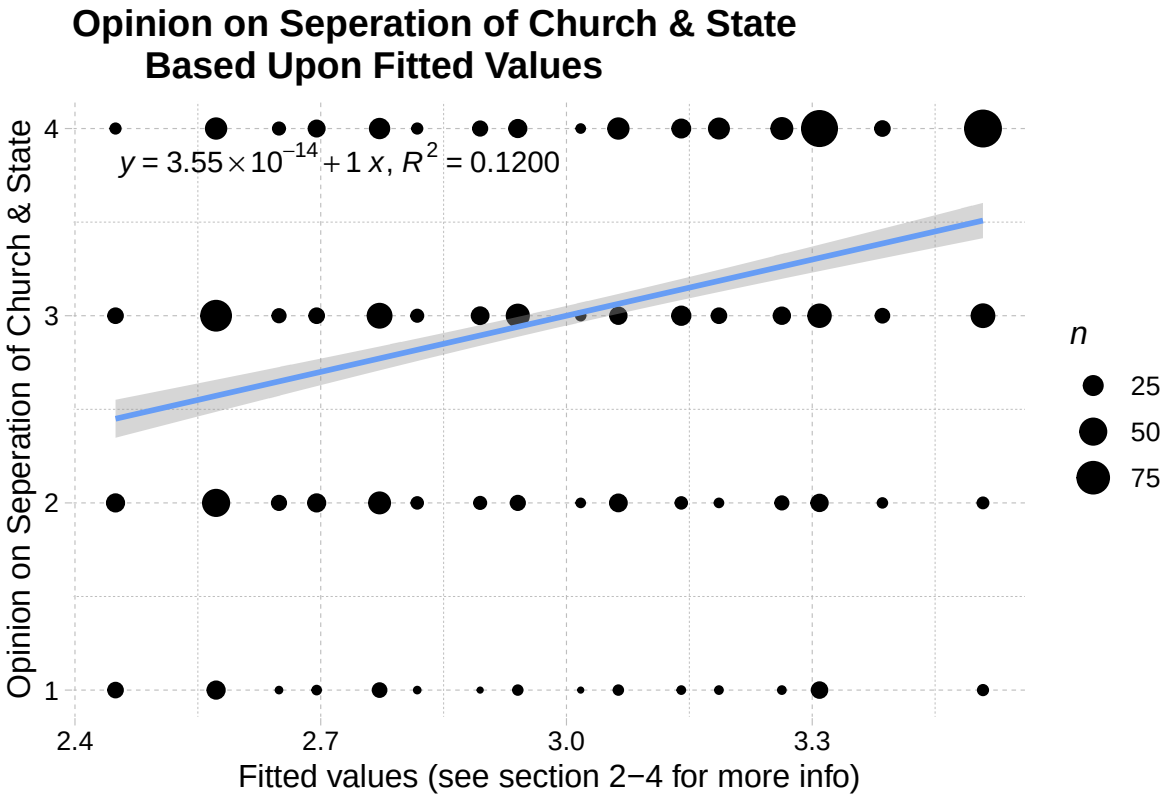

Figure 4: Scatter plot with linear regression based upon fitted values.

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

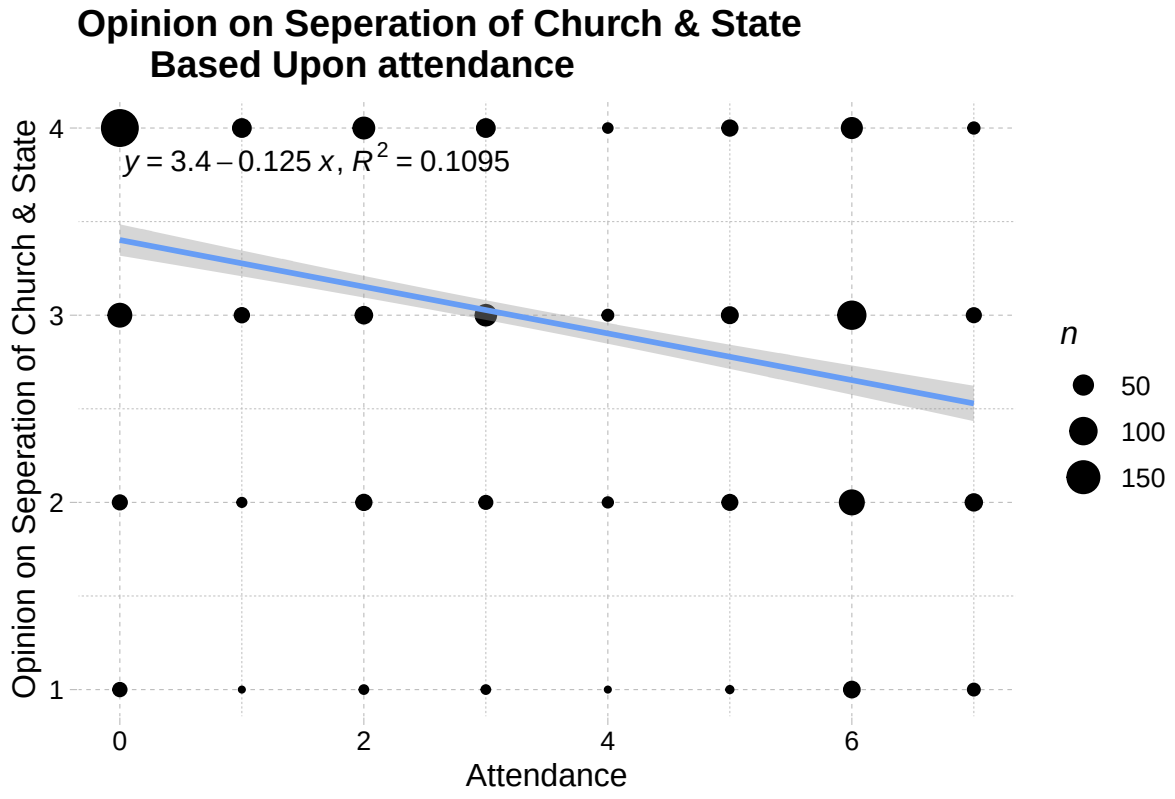

Figure 5: Scatter plot with linear regression based upon attendance.

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

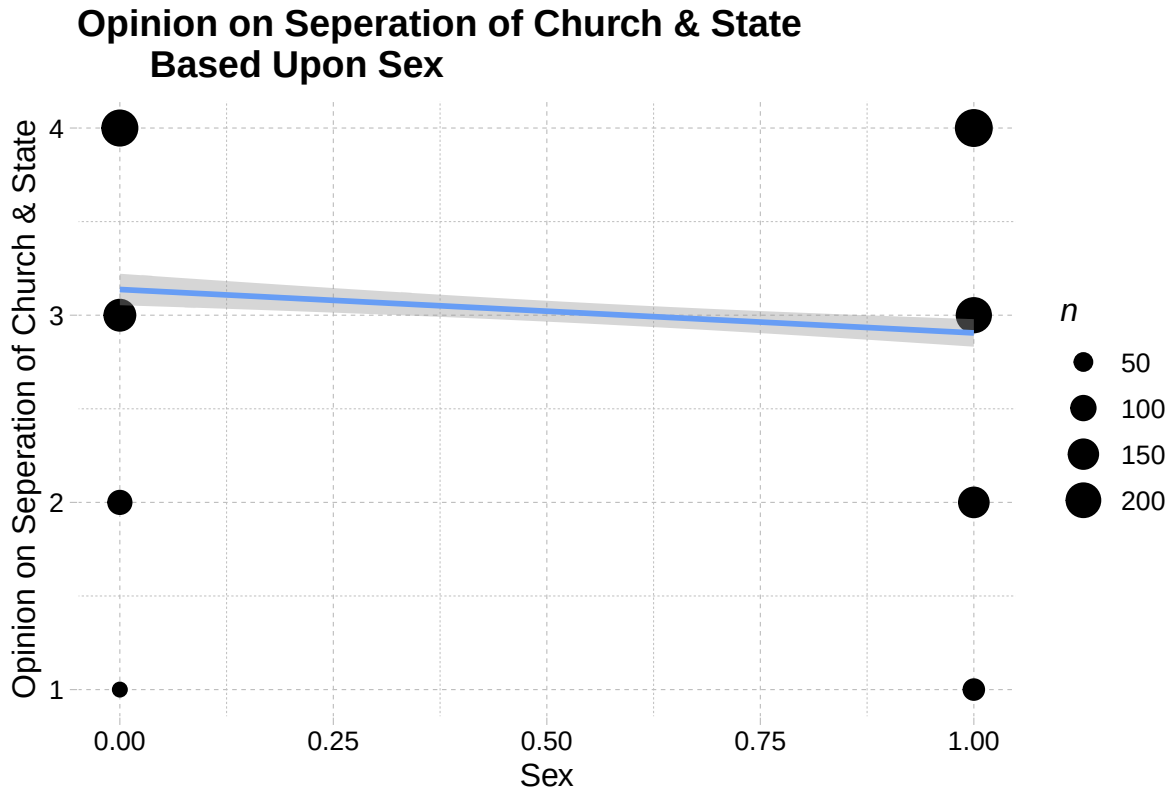

Figure 6: Scatter plot with linear regression based upon sex