MANOVA and ANOVA Analysis on Crime in Ohio

In this project, I apply MANOVA and ANOVA tests to data on perception of crime in Ohio. Ultimately, we want to answer the question of whether females and males answer questions differently and whether people of different races (white/black/other) answer questions on crime differently.

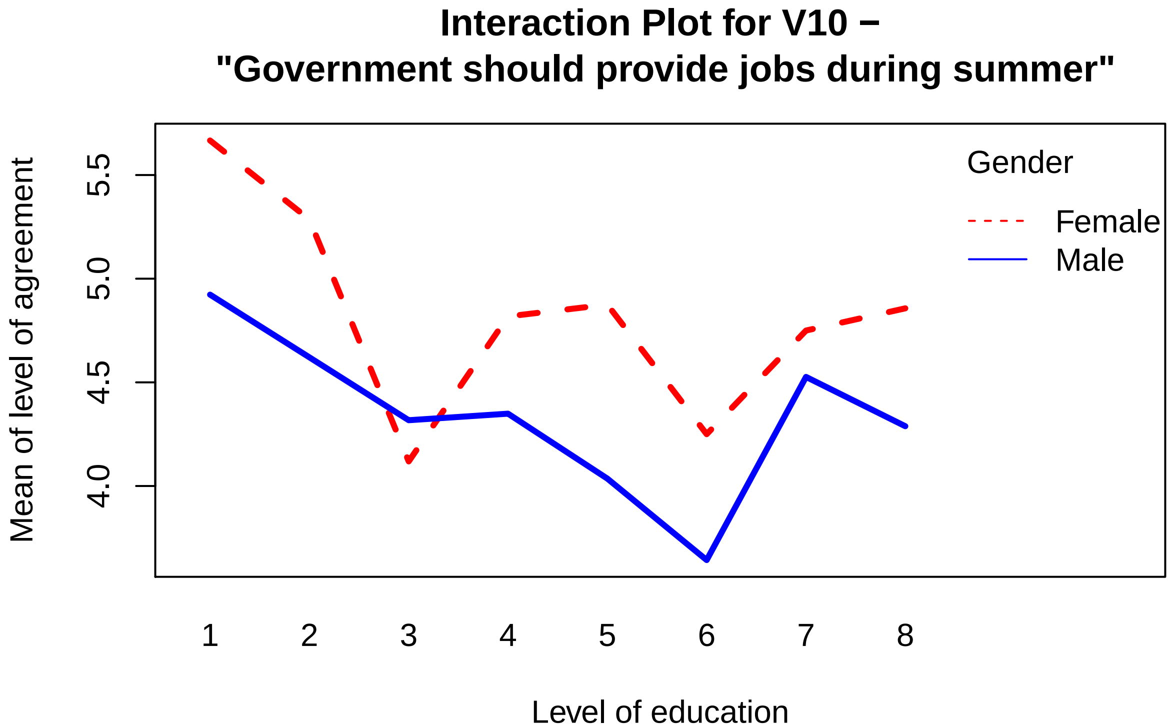
```
# read in the data
crime <- read.csv("Data/ohiocrime.csv")
head(crime)
```

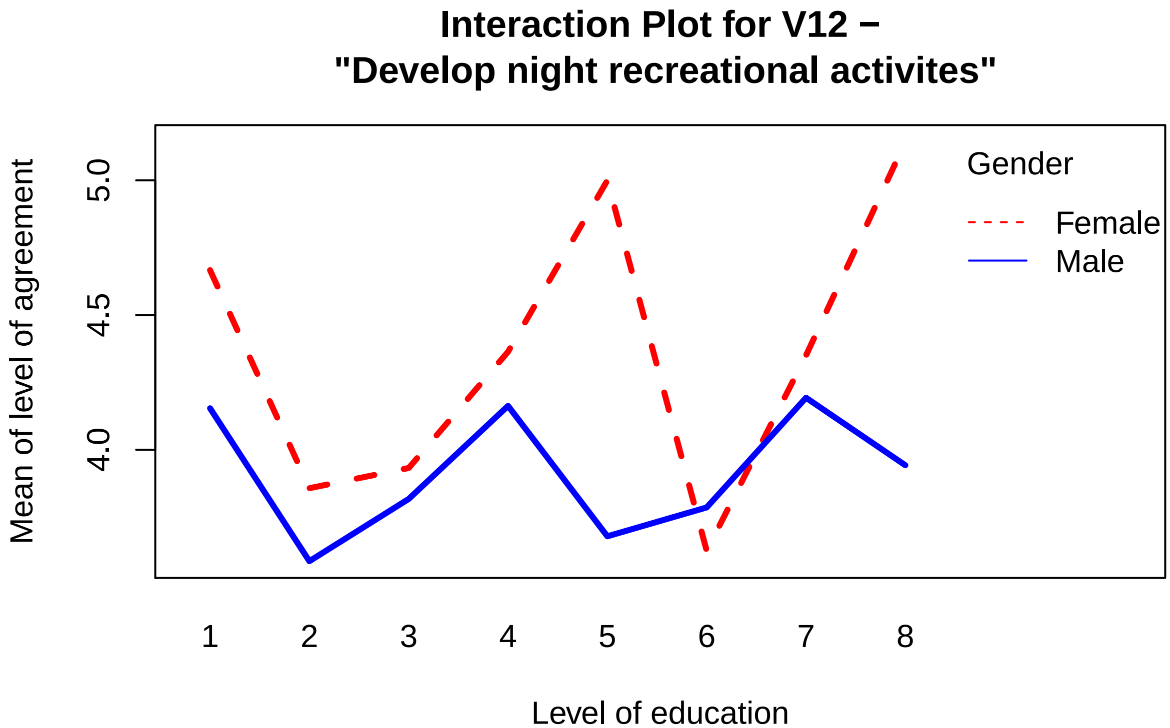
```
##   V10 V12 V16 V23 V64 V67 V70 V71 V72 V73 V86 V87
## 1   4   4   4   5   2   4   1   1   5   2   7   5
## 2   5   1   5   3   2   4   1   1   5   2   5   2
## 3   6   6   6   6   6   1   0   1   2   2   5   1
## 4   2   1   1   1   4   5   1   1   7   2   8   6
## 5   2   1   1   5   2   4   1   1   6   2   7   5
## 6   4   2   3   4   3   5   0   1   7   2   8   1
```

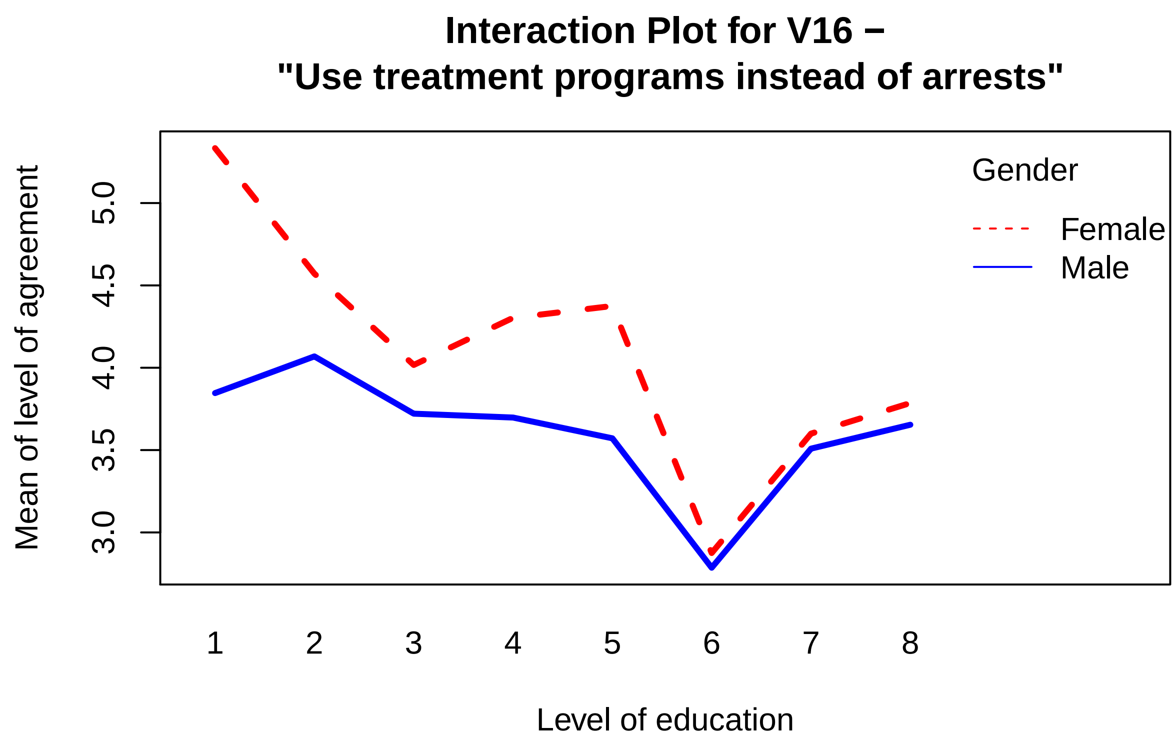
```
crime <- na.omit(crime)

crime$V70[crime$V70 == 0] <- "Female"
crime$V70[crime$V70 == 1] <- "Male"
crime$V70 <- as.factor(crime$V70)

crime$V71[crime$V71 == 1] <- "White"
crime$V71[crime$V71 == 2] <- "Black"
crime$V71[crime$V71 == 3] <- "Other"
crime$V71 <- as.factor(crime$V71)
```

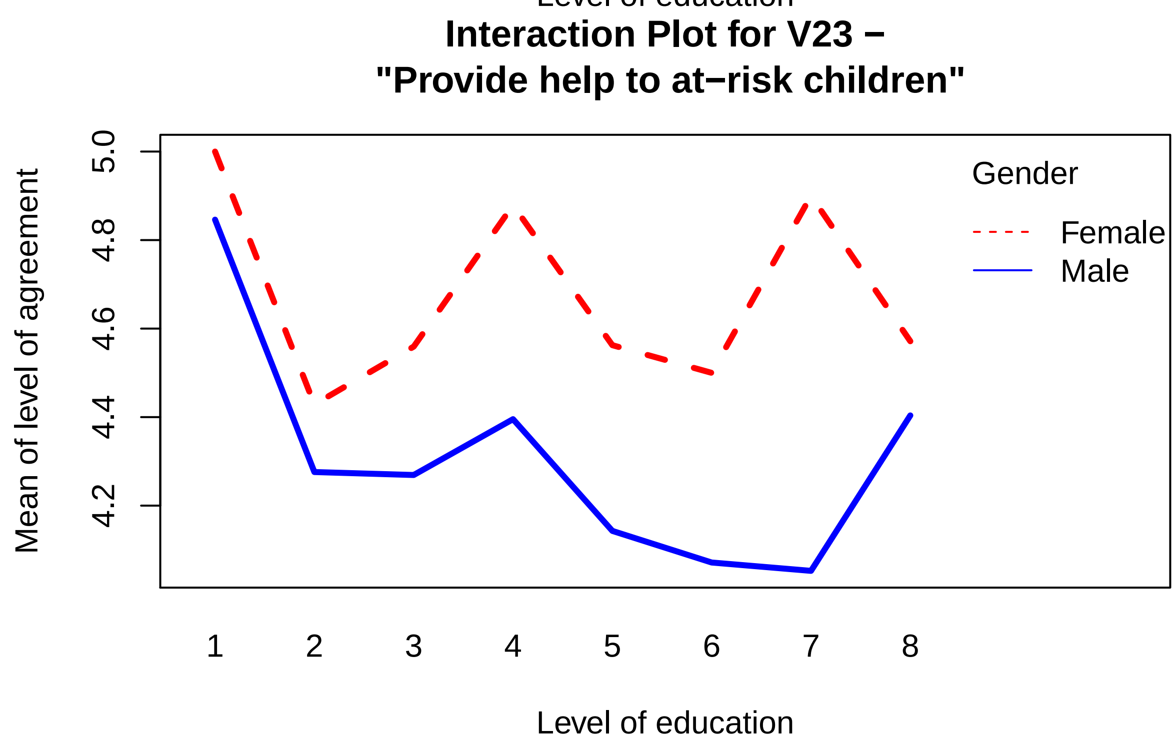
```
# generate interaction plots

interaction.plot(crime$V72, crime$V70, crime$V10,
 lwd=3,col=c("red","blue"),xlab="Level of education",
 ylab = "Mean of level of agreement", trace.label = "Gender",
 main="Interaction Plot for V10 -
 \"Government should provide jobs during summer\"")

interaction.plot(crime$V72, crime$V70, crime$V12,
 lwd=3,col=c("red","blue"),xlab="Level of education",
 ylab = "Mean of level of agreement", trace.label = "Gender",
 main="Interaction Plot for V12 -
 \"Develop night recreational activites\"")

interaction.plot(crime$V72, crime$V70, crime$V16,
 lwd=3,col=c("red","blue"),xlab="Level of education",
 ylab = "Mean of level of agreement", trace.label = "Gender",
 main="Interaction Plot for V16 -
 \"Use treatment programs instead of arrests\"")

interaction.plot(crime$V72, crime$V70, crime$V23,
 lwd=3,col=c("red","blue"),xlab="Level of education",
 ylab = "Mean of level of agreement", trace.label = "Gender",
 main="Interaction Plot for V23 -
 \"Provide help to at-risk children\"")
```

In general across all 4 responses, we see that females respond with higher levels of agreement to the questions than males. Thus, I argue that there does seem to be a difference in mean agreement levels between men and women and that women have a higher mean. Additionally, there seems to be a multimodal distribution across the different responses and education levels, with responders having very low (i.e., below 2), medium (i.e., between around 4) and very high (i.e., above 7) having similar responses.










```r
# fit linear model
model <- manova(as.matrix(crime[,1:4])~ crime$V70 + crime$V72 + crime$V70*crime$V72)

# get univariate results
summary.aov(model)
```

```
##  Response V10 :
##                        Df Sum Sq Mean Sq F value  Pr(>F)
```

```
## crime$V70             1   5.21  5.2062  3.1522 0.07644 .
## crime$V72             1   0.07  0.0722  0.0437 0.83452
## crime$V70:crime$V72   1   3.32  3.3188  2.0094 0.15695
## Residuals           496 819.20  1.6516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V12 :
##                      Df  Sum Sq Mean Sq F value  Pr(>F)
## crime$V70             1   13.92 13.9225  5.1642 0.02348 *
## crime$V72             1    8.80  8.7977  3.2633 0.07145 .
## crime$V70:crime$V72   1    4.36  4.3571  1.6161 0.20423
## Residuals           496 1337.20  2.6960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V16 :
##                      Df  Sum Sq Mean Sq F value   Pr(>F)
## crime$V70             1   15.33 15.3300  7.0594 0.008139 **
## crime$V72             1   12.17 12.1726  5.6054 0.018288 *
## crime$V70:crime$V72   1    2.30  2.3007  1.0595 0.303841
## Residuals           496 1077.11  2.1716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V23 :
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## crime$V70             1  16.99 16.9945 11.3650 0.0008068 ***
## crime$V72             1   0.45  0.4524  0.3026 0.5825247
## crime$V70:crime$V72   1   0.87  0.8660  0.5791 0.4470219
## Residuals           496 741.69  1.4953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# get multivariate results
summary.manova(model)
```

```
##                      Df   Pillai approx F num Df den Df   Pr(>F)
## crime$V70             1 0.029827   3.7892      4    493 0.004788 **
## crime$V72             1 0.024965   3.1557      4    493 0.014062 *
## crime$V70:crime$V72   1 0.011290   1.4074      4    493 0.230302
## Residuals           496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary.manova(model,test="Wilks")
```

```
##                      Df   Wilks approx F num Df den Df   Pr(>F)
## crime$V70             1 0.97017   3.7892      4    493 0.004788 **
## crime$V72             1 0.97504   3.1557      4    493 0.014062 *
## crime$V70:crime$V72   1 0.98871   1.4074      4    493 0.230302
## Residuals           496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Univariate Results: Here, I make the following observations for each response variable:

V10 (Summer jobs provided by state government) For V10, there are differences between levels of education (V72), but no overall observed differences due to gender (V70). There is no evidence of an interaction effect.

V12 (Recreation programs) For V12, there are differences between genders (V70), but no overall observed differences due to level of education. There is no evidence of an interaction effect.

V16 (Treatment programs instead of arrest) For V16, there are differences due to both levels of education (V72) and gender (V70). There is no evidence of an interaction effect.

V23 (Help for families for at-risk children) The results for V23 are similar to V12 (only level of education has differences).

Multivariate Results: Overall, there are differences between genders (V70) and between education levels (V72). The multivariate statistics for both Pillai's Trace and for Wilks' Lambda were significant. However, in both tests there is no evidence for an interaction effect between gender and education level.

```
contrasts(crime$V71, 1) <- matrix(c(0, 1, 1), 3, 1)
contrasts(crime$V71)
```

```
##       [,1]
## Black    0
## Other    1
## White    1
```

```
model_null <- lm(cbind(V10, V12, V16, V23) ~ 1, crime)
model <- lm(cbind(V10, V12, V16, V23) ~ V71, crime)
model_null
```

```
##
## Call:
## lm(formula = cbind(V10, V12, V16, V23) ~ 1, data = crime)
##
## Coefficients:
##              V10    V12    V16    V23
## (Intercept)  4.420  4.038  3.776  4.400
```

```
model
```

```
##
## Call:
## lm(formula = cbind(V10, V12, V16, V23) ~ V71, data = crime)
##
## Coefficients:
##              V10       V12       V16       V23
## (Intercept)   5.03226   4.09677   4.22581   4.83871
## V711         -0.65273  -0.06266  -0.47954  -0.46771
```

```
anova(model_null, model)
```

```
## Analysis of Variance Table
##
## Model 1: cbind(V10, V12, V16, V23) ~ 1
## Model 2: cbind(V10, V12, V16, V23) ~ V71
##   Res.Df Df Gen.var.   Pillai approx F num Df den Df  Pr(>F)
## 1    499      1.6974
## 2    498 -1   1.6912 0.022626   2.8648      4    495 0.02288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

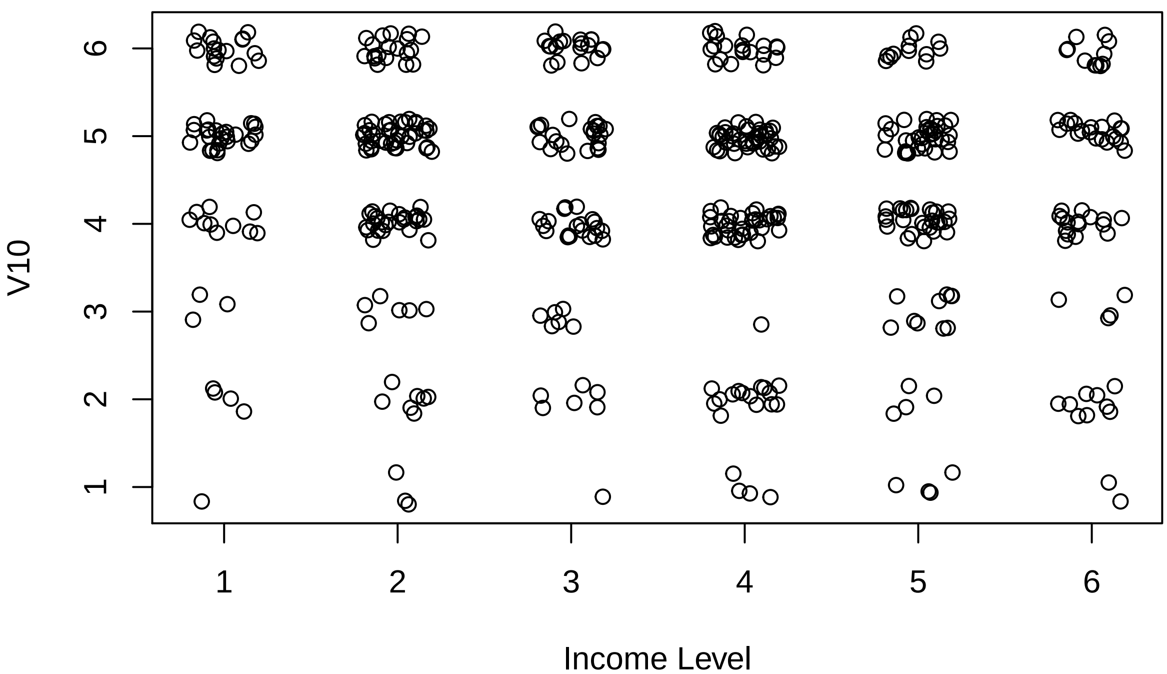
Here, we set contrasts using the contrasts() function, such that only one contrast is used to examine differences between the race groups. We wish to examine the difference in mean (multivariate) responses to the questions, between Black and non-Black (White and Other) groups. Hence I set the contrast to include the White and Other groups, so that the Black mean (multivariate) responses are accounted for in the overall model intercept.

There is a significant difference between the mean (multivariate) responses of these two groups if, and only if, the model that accounts for the difference in means (using the contrasts) causes a statistically significant reduction in the residual sum of squares (as indicated by the test statistics) from the intercept-only null model, which does not account for such differences.

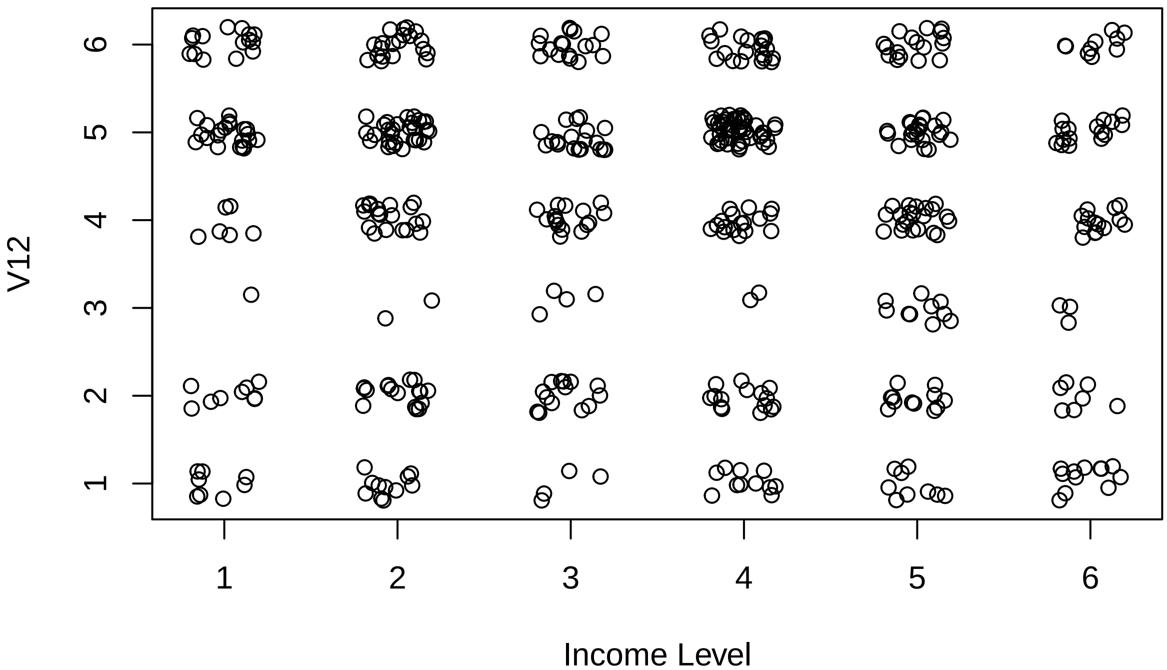
Hence, we can test the null hypothesis that the mean (multivariate) responses of the two groups are equal by using the anova() function to perform a comparison of the two models. In this case, we find that the mean (multivariate) responses between the Black and non-Black groups are significant at $\alpha = 0.05$.

This method of applying multivariate contrasts in R appears to be easily generalisable, if it is correct.

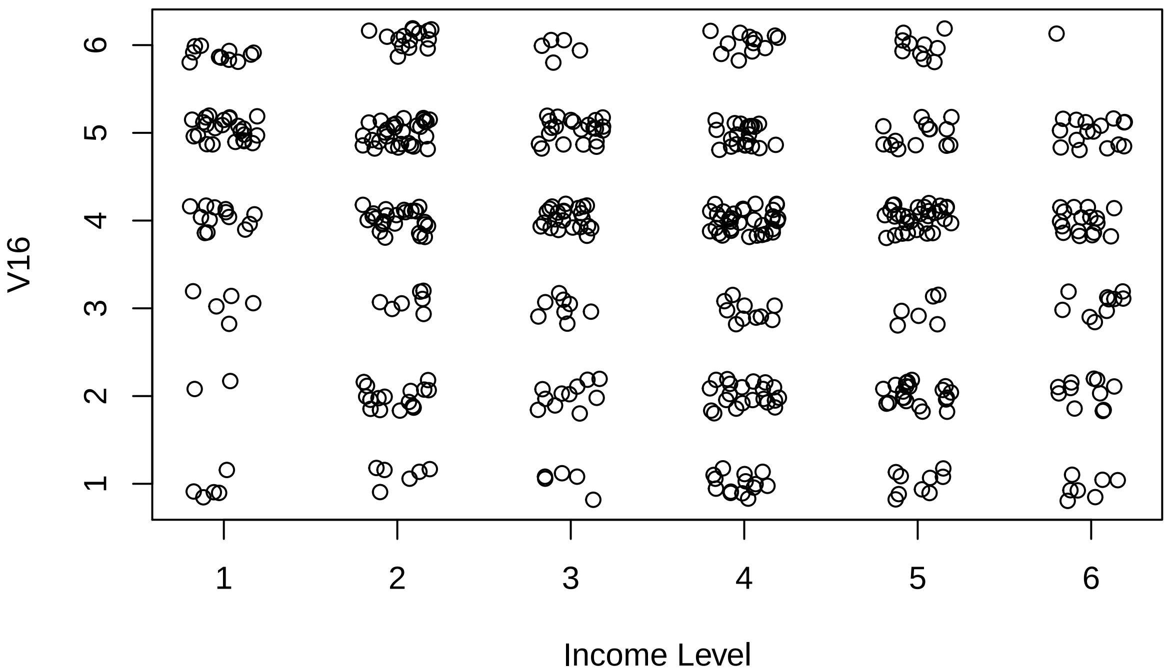
```
crime$V72 <- as.factor(crime$V72)
linear_model <- lm(as.matrix(crime[,1:4])~ crime$V70 + crime$V72 + crime$V70*crime$V72 + crime$V87)

plot(jitter(crime$V87), jitter(crime[, 1]), xlab = "Income Level", ylab = "V10")
```



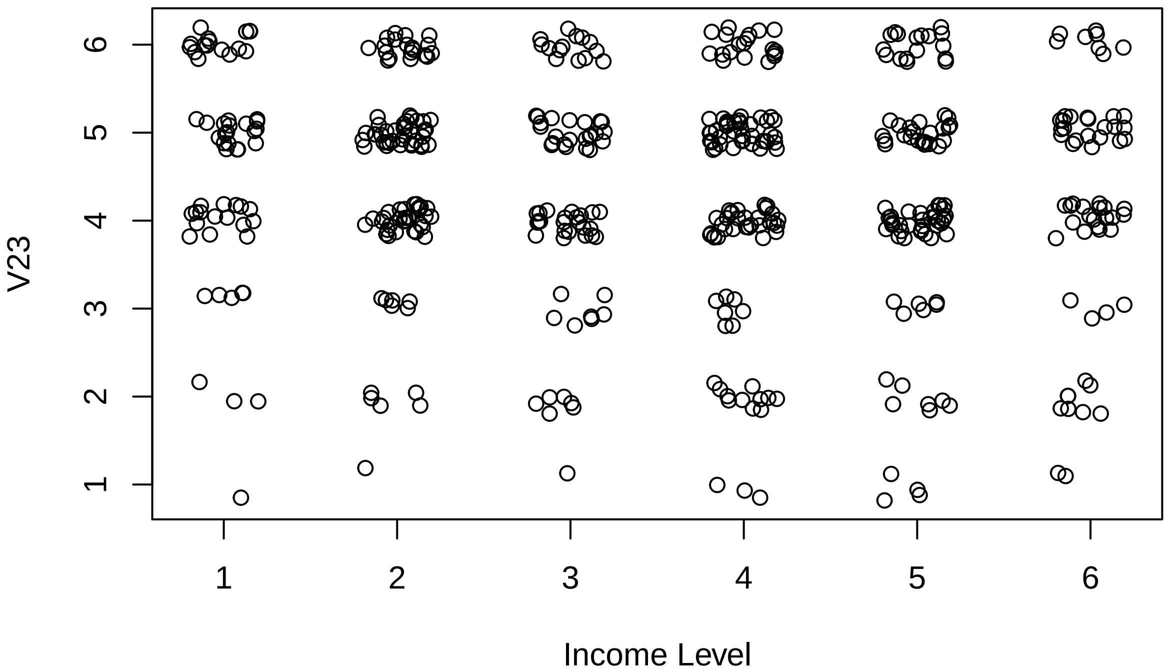

```
plot(jitter(crime$V87), jitter(crime[, 2]), xlab = "Income Level", ylab = "V12")
```




```r
plot(jitter(crime$V87), jitter(crime[, 3]), xlab = "Income Level", ylab = "V16")
```




```r
plot(jitter(crime$V87), jitter(crime[, 4]), xlab = "Income Level", ylab = "V23")
```



In general, there are no linear relationships between income level and the response variables. For there seems to be no discernible pattern with respect to the density of the points plotted. However, income level and response V16 ("Use treatment programs instead of arrests") seem to be negatively correlated to a small extent.