HW4: MANOVA

2025-06-14


```
library(MASS)
library(biotools)
```

```
## ---
## biotools version 4.3
```

```
library(klaR)
library(car)
```

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

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(ggExtra)
library(heplots)
```

```
## Loading required package: broom
```

```
## Warning in rgl.init(initValue, onlyNULL): RGL: unable to open X11 display
```

```
## Warning: 'rgl.init' failed, will use the null device.
## See '?rgl.useNULL' for ways to avoid this warning.
```

```
##
## Attaching package: 'heplots'
```

```
## The following object is masked from 'package:biotools':
##
##     boxM
```

```
library(readr)
library(caret)
```

```
## Loading required package: lattice
```

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Registered S3 methods overwritten by 'vegan':
##   method      from
##   plot.rda    klaR
##   predict.rda klaR
##   print.rda   klaR
```

```
##
## Attaching package: 'vegan'
```

```
## The following object is masked from 'package:caret':
##
##     tolerance
```

```
## The following object is masked from 'package:klaR':
##
##     rda
```

1 Forest Fire Dataset

1.1 Dataset Intro & Cleaning

The dataset from UCI Machine Learning Respository (link:%20https://archive.ics.uci.edu/dataset/547/algerian+forest+fires+dataset) includes 244 instances that regroup a data of two regions of Algeria,namely the Bejaia region located in the northeast of Algeria and the Sidi Bel-abbes region located in the northwest of Algeria.

- **region** (categorical): Bejaia or Sidi-Bel Abbès.
- **day** (integer): Day of month

- **month** (categorical): Month of year. From June to September.
- **year** (categorical): Calendar year
- **Temperature** (integer): Temperature at noon (unit: C)
- **RH** (integer): Relative humidity (unit: %)
- **Ws** (integer): Wind speed (unit: km/h)
- **Rain** (continuous): Accumulated rainfall (unit: mm)
- **FFMC** (continuous): Fine Fuel Moisture Code. Measures moisture of surface litter (needles, grass). Higher = drier = more flammable).
    - Depends on temperature, RH, wind, rain, and previous FFMC.
- **DMC** (continuous): Duff Moisture Code. Measures moisture in loosely compacted organic layers (duff) below the surface. Higher DMC -> drier duff -> easier sustained fire.
    - Depends on rain, temperature, RH, and previous DMC
- **DC** (continuous): Drought Code. Measures moisture in deep, compact organic matter. Higher DC -> deeper, long-term drought conditions -> deep-burning fires possible.
    - Depends on rain, temperature, and previous DC
- **ISI** (continuous): Initial Spread Index. Measures potential fire spread rate immediately after ignition.Higher ISI -> faster spread.
    - Depends on FFMC, wind
- **BUI** (continuous): Buildup Index. Measures amount of fuel available for combustion. Higher BUI -> more fuel -> more intense fire if ignited.
    - Depends on DMC, DC
- **FWI** (continuous): Fire Weather Index. Measures general fire intensity (combines ISI and BUI). Higher FWI -> higher potential fire severity.
- **Classes** (categorical):; Fire occurrence (`not fire` / `fire`)

1.1.1 Data Cleaning

```
df <- read_csv("algerian_forest_fires.csv", skip = 1)
```

```
## Rows: 244 Columns: 15
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (4): DC, FWI, Classes, Region
## dbl (11): day, month, year, Temperature, RH, Ws, Rain, FFMC, DMC, ISI, BUI
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df <- df[complete.cases(df), ]
df$day <- as.integer(df$day)
df$month <- factor(df$month, levels = 6:9,
                   labels = c("Jun","Jul","Aug","Sep"))
df$year <- as.factor(df$year)
df$DC <- as.numeric(df$DC)
df$FWI <- as.numeric(df$FWI)
df$Classes <- trimws(df$Classes)
df$Classes <- factor(df$Classes)
df$Region <- factor(df$Region)

df$sqWs <- sqrt(df$Ws + 1)
df$tRain <- 1 / ((df$Rain + 1)^3)
df$tFFMC <- (df$FFMC)^3
df$logDMC <- log(df$DMC + 1)
df$logDC <- log(df$DC + 1)
df$logFWI <- log(df$FWI + 1)
```

In this assignment:

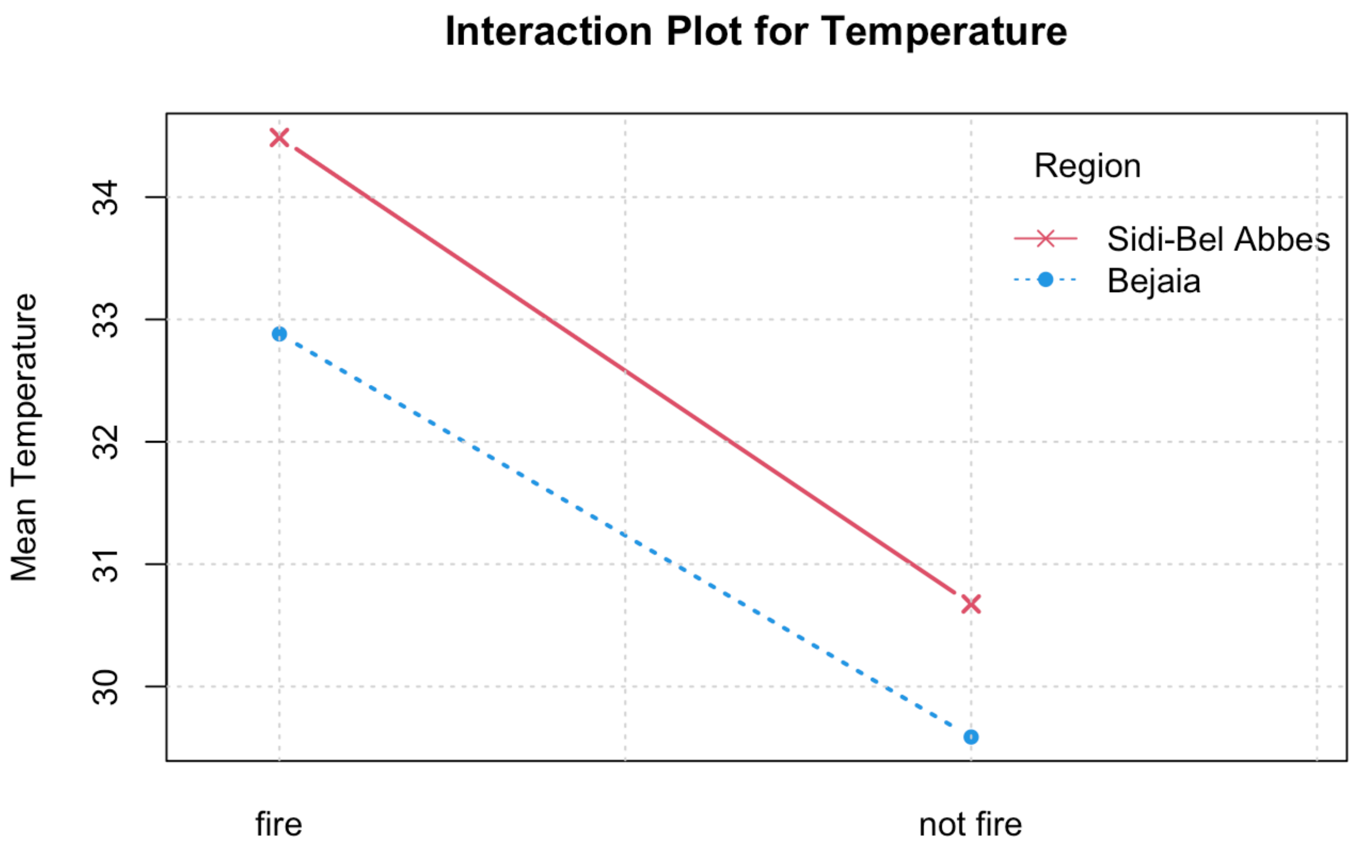
- Continuous response variables: our three basic meteorological measurements `Temperature` , `RH` , and `Ws` .
- Categorical predictors: `Classes` (fire vs. not fire), `Region` (Bejaia vs. Sidi-Bel Abbes), and `month` (Jun, Jul, Aug, Sep).
- One or more additional continuous predictors: any other continuous variables in our list.

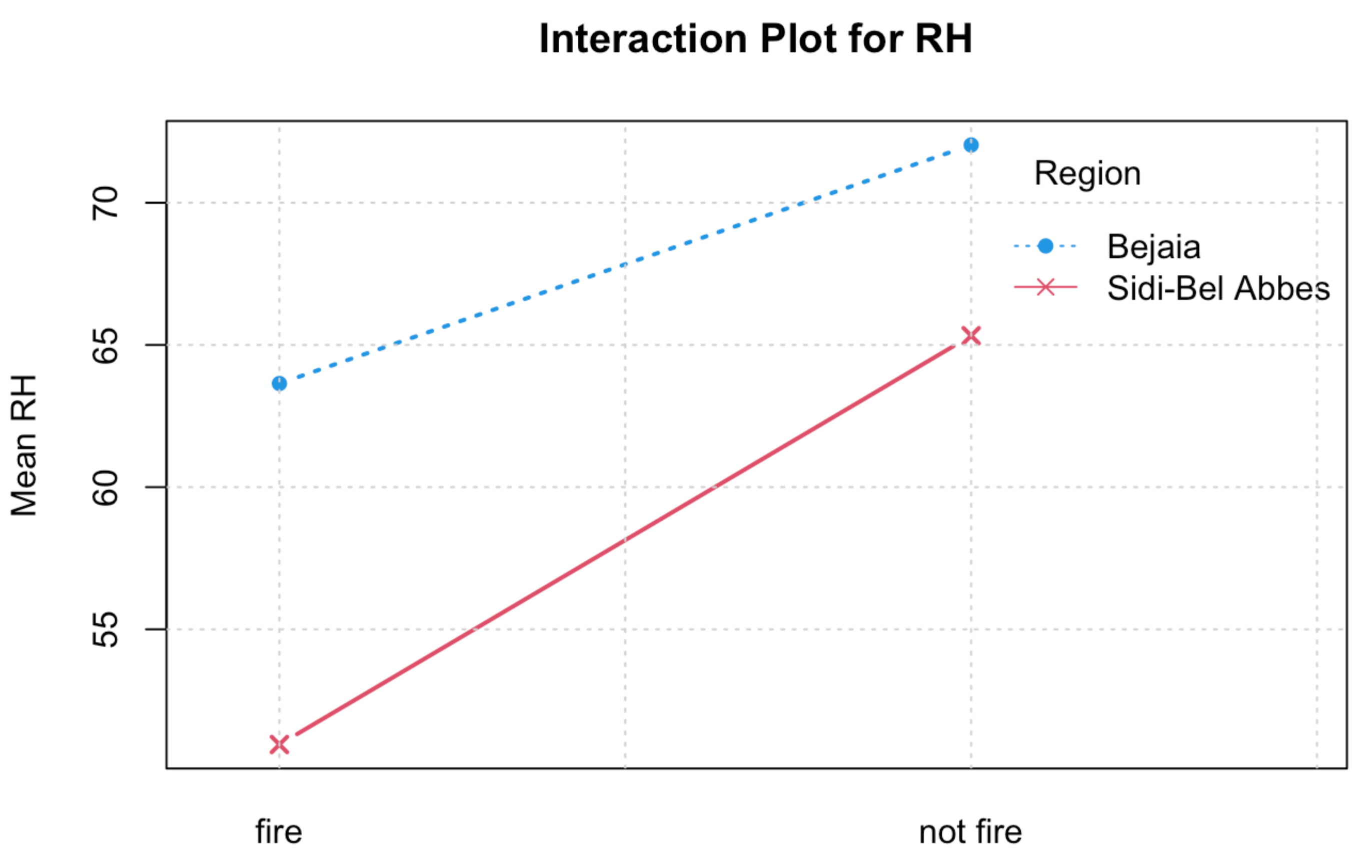
1.2 Interaction Plots

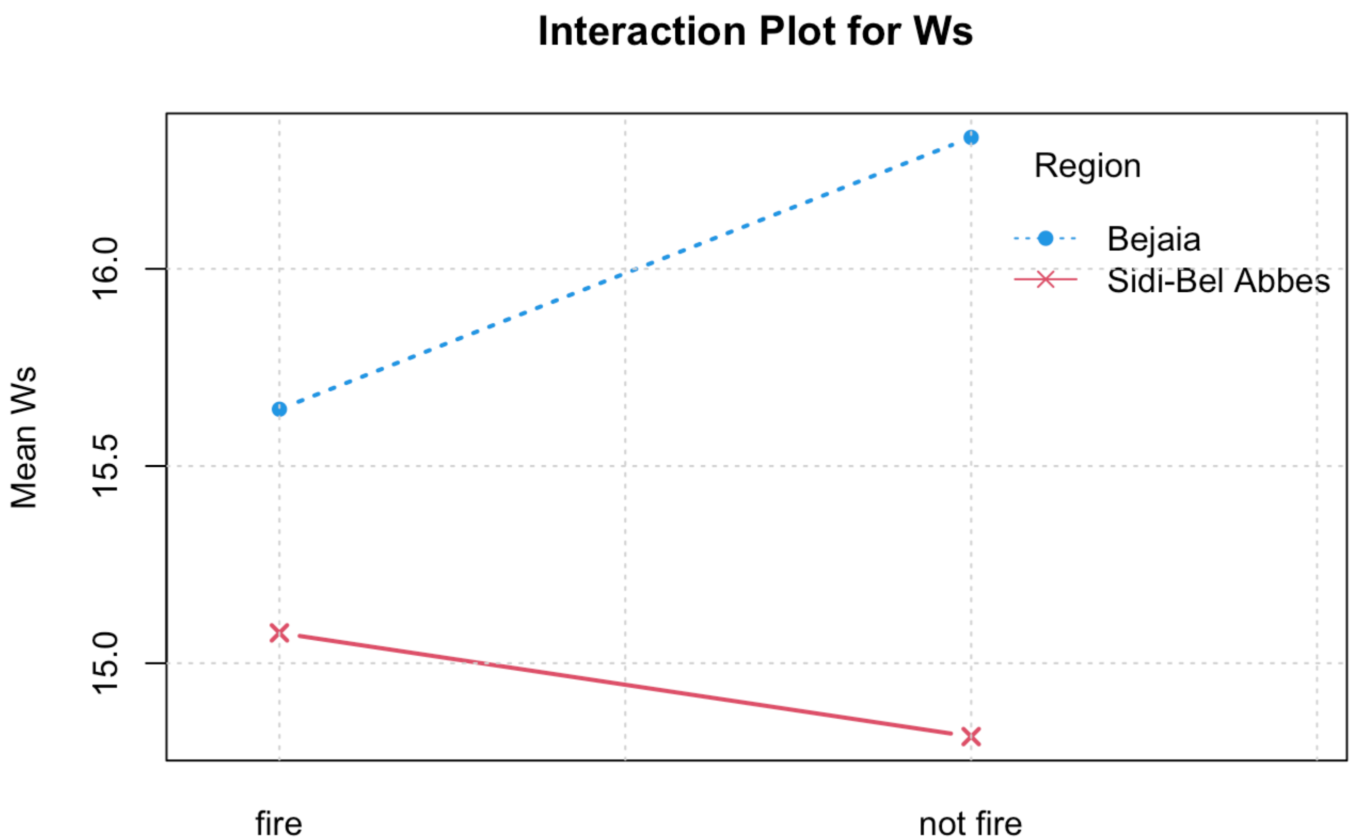
1.2.1 By categorical predictors `Classes` and `Region`

We start by creating interaction plots for `Temperature` , `RH` , and `Ws` , grouped by the first two categorical predictors `Classes` and `Region` , across all months.

```r
resp.vars <- c("Temperature","RH", "Ws")

for (v in resp.vars) {
    interaction.plot(
      x.factor = df$Classes,
      trace.factor = df$Region,
      response = df[[v]],
      type = 'b', lwd = 2,
      trace.label = "Region",
      lty = c(3, 1), col = c(4, 2), pch = c(16, 4),
      xlab = " ",
      ylab = paste("Mean", v),
      main = paste("Interaction Plot for", v),
    )
  grid()
}
```


Interaction Plot for Temperature
Region
Sidi-Bel Abbes
Bejaia
Mean Temperature
34
33
32
31
30
fire
not fire


Interaction Plot for RH
Region
Bejaia
Sidi-Bel Abbes
Mean RH
70
65
60
55
fire
not fire



Observations:

1. In both regions, temperatures tend to be higher on fire days than on non-fire days. Sidi-Bel Abbes generally records higher temperatures than Bejaia. The lines are nearly parallel, suggesting little to none interaction between Classes and Region.

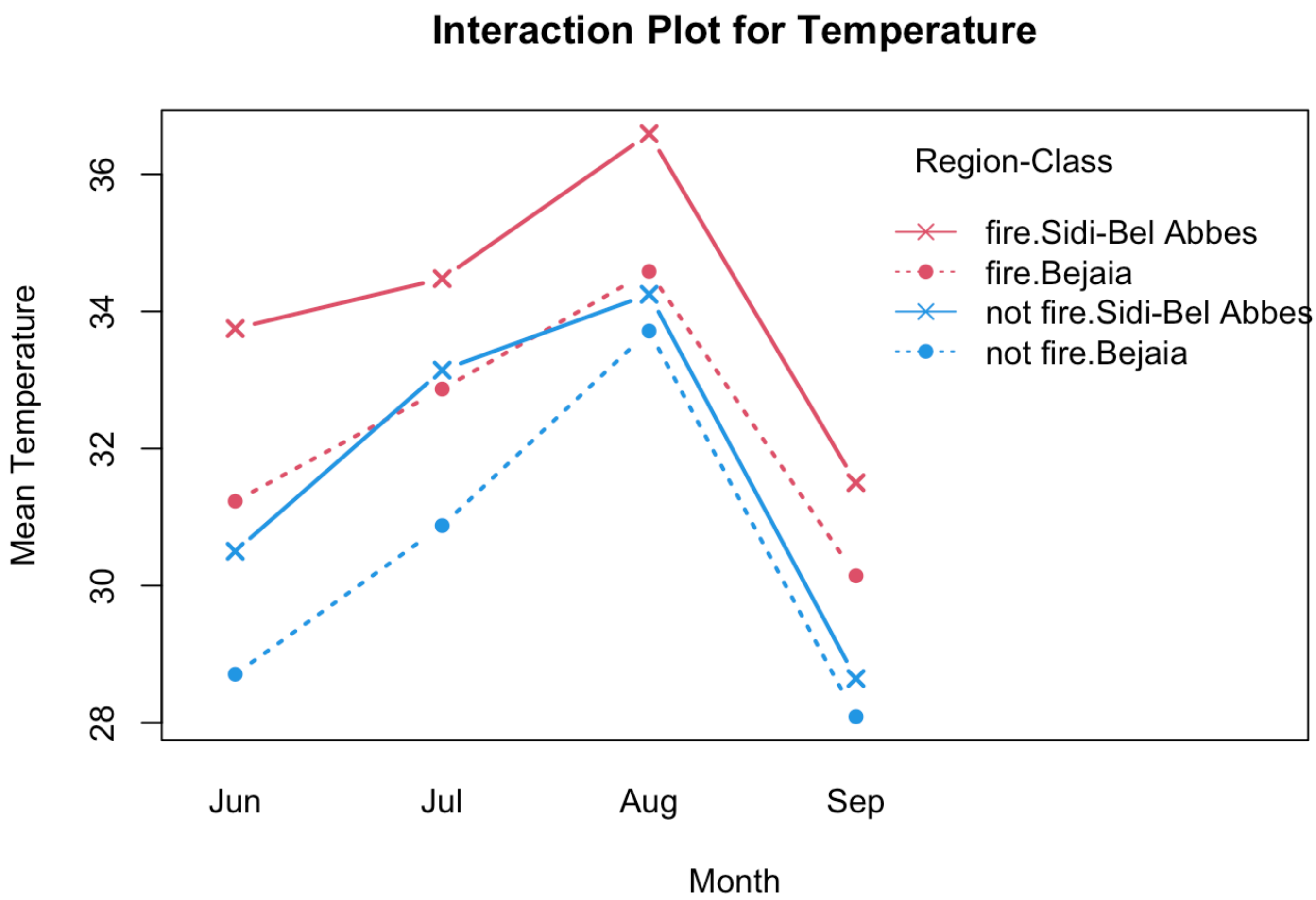
2. RH is lower on fire days, especially in Sidi-Bel Abbes. A possible interaction is observed: the RH difference between fire and not fire is greater in Sidi-Bel Abbes.

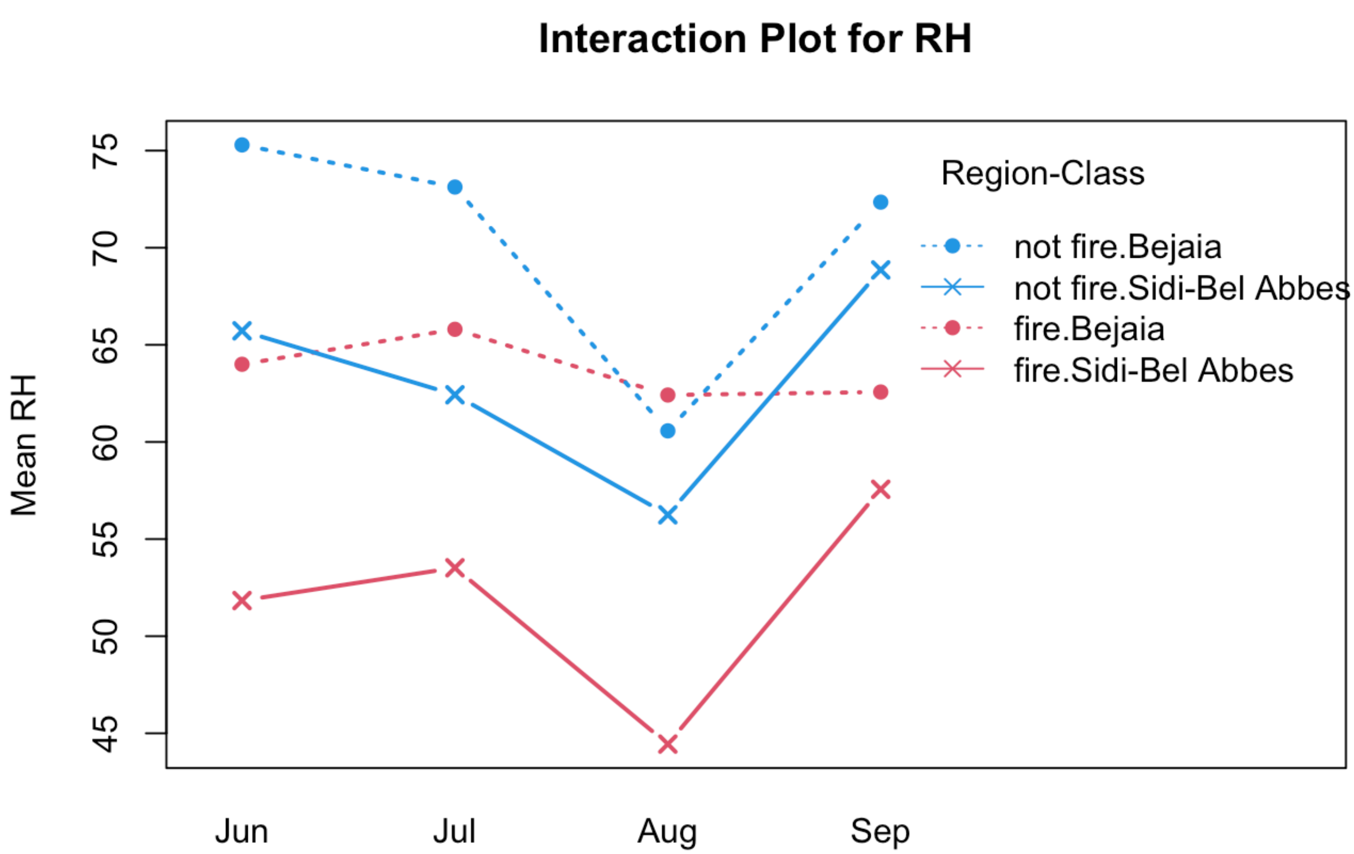
3. In Sidi-Bel Abbes, mean wind speed is slightly lower on non-fire days than on fire days. In Bejaia, it's the opposite: wind speed is actually higher on non-fire days than on fire days. This seems a bit counter-intuitive: as it makes more sense for fire risk to increase with higher the wind speeds, the faster the fire spreads. This negative relationship might suggests that wind speed is not a major contributor to fire occurances in the region.

1.2.2 By all 3 ategorical predictors

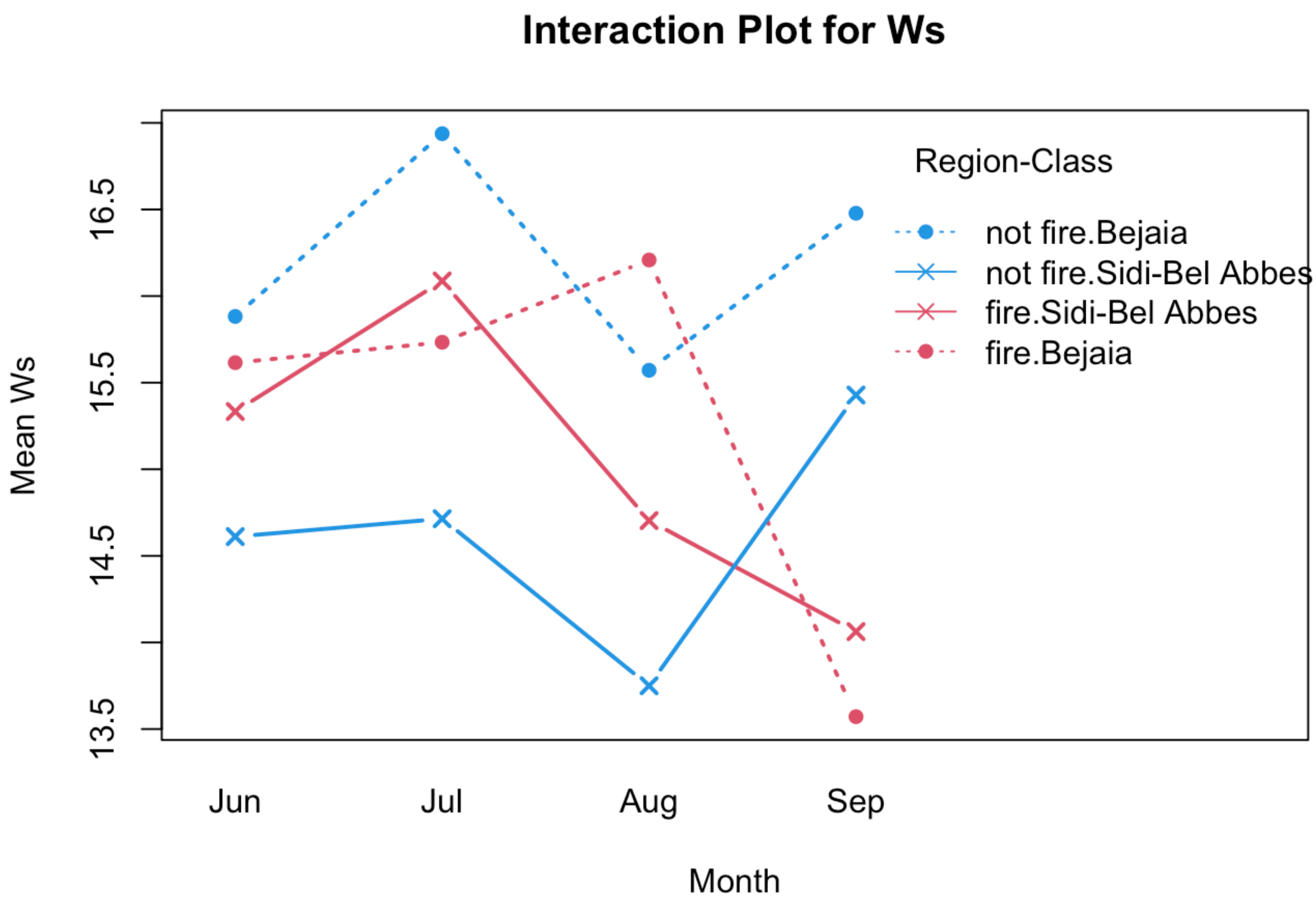
Out of curiosity, I made interaction plots for the response variables grouped by all three categorical predictors `Classes` , `Region` , and `months` , where Classes = fire is in red, Classes = not fire in blue, Region = Bejaia is in dashed line, and Region = Sidi-Bel Abbes is in solid line.

```r
for (v in resp.vars) {
    interaction.plot(
      x.factor = df$month,
      trace.factor = interaction(df$Classes, df$Region),
      response = df[[v]],
      type = 'b', lwd = 2,
      trace.label = "Region-Class",
      lty = c(3,3, 1,1), col = c(2, 4, 2, 4), pch = c(16, 16, 4, 4),
      xlab = "Month",
      ylab = paste("Mean", v),
      main = paste("Interaction Plot for", v),
    )
}
```


Interaction Plot for Temperature
Mean Temperature
28
30
32
34
36
Jun
Jul
Aug
Sep
Month
Region-Class
fire.Sidi-Bel Abbes
fire.Bejaia
not fire.Sidi-Bel Abbes
not fire.Bejaia


Interaction Plot for RH
Mean RH
45
50
55
60
65
70
75
Jun
Jul
Aug
Sep
Region-Class
not fire.Bejaia
not fire.Sidi-Bel Abbes
fire.Bejaia
fire.Sidi-Bel Abbes

Month



Observations:

1. Temperatures peak in August across all groups. Sidi-Bel Abbes seems to be more fire prone and generally records higher temperatures than Bejaia. For both regions, temperatures tend to be at least one degree higher on fire days than on non-fire days.

2. RH is generally inversely related to fire occurrence: fire groups (red) tend to have lower humidity, especially Sidi-Bel Abbes in August (~45%). Bejaia is more humid than Sidi-Bel Abbes. Bejaia—not fire shows the highest RH throughout (~75% in June). This pattern support the hypothesis that drier air contributes to fire events, especially during midsummer.

A possible outlier is observed: the RH is higher under fire conditions than non-fire conditions in Bajaia in August, whereas in all other region–class combinations and months, RH consistently drops by approximately 10% from not fire to fire,

3. Wind speed patterns are more variable and region-dependent. Generally, Bejaia seems to be windier than Sidi-Bel Abbes.

1.3 Two-Way MANOVA

```
options(contrasts = c("contr.sum","contr.poly"))
mod_mva <- lm(
  cbind(Temperature, RH, Ws) ~ Classes *Region, data = df
)

summary(Anova(mod_mva, type = 3), univariate = TRUE)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##             Temperature        RH        Ws
## Temperature   2220.3684 -4706.159 -540.3526
## RH           -4706.1593 36708.699 1654.7711
## Ws            -540.3526  1654.771 1833.5755
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##             Temperature       RH       Ws
## Temperature    236427.3 466722.2 114607.2
## RH             466722.2 921338.8 226241.8
## Ws             114607.2 226241.8  55555.4
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1   0.99682 24784.08      3    237 < 2.22e-16 ***
## Wilks             1   0.00318 24784.08      3    237 < 2.22e-16 ***
## Hotelling-Lawley  1 313.72253 24784.08      3    237 < 2.22e-16 ***
## Roy               1 313.72253 24784.08      3    237 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH         Ws
## Temperature   733.06133 -2348.1411 -43.972028
## RH          -2348.14115  7521.5628 140.851144
## Ws            -43.97203   140.8511   2.637623
##
## Multivariate Tests: Classes
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.2755645 30.05043      3    237 < 2.22e-16 ***
## Wilks             1 0.7244355 30.05043      3    237 < 2.22e-16 ***
```

```
## Hotelling-Lawley  1 0.3803851 30.05043      3    237 < 2.22e-16 ***
## Roy               1 0.3803851 30.05043      3    237 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   105.25526 -758.3203 -81.55311
## RH           -758.32027 5463.3815 587.55618
## Ws            -81.55311  587.5562  63.18839
##
## Multivariate Tests: Region
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1386396 12.71538      3    237 9.8157e-08 ***
## Wilks             1 0.8613604 12.71538      3    237 9.8157e-08 ***
## Hotelling-Lawley  1 0.1609542 12.71538      3    237 9.8157e-08 ***
## Roy               1 0.1609542 12.71538      3    237 9.8157e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes:Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature    3.905114 -45.08995   7.16896
## RH           -45.089947 520.62588 -82.77557
## Ws             7.168960 -82.77557  13.16069
##
## Multivariate Tests: Classes:Region
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1 0.0258281 2.094519      3    237 0.10162
## Wilks             1 0.9741719 2.094519      3    237 0.10162
## Hotelling-Lawley  1 0.0265129 2.094519      3    237 0.10162
## Roy               1 0.0265129 2.094519      3    237 0.10162
##
##  Type III Sums of Squares
##                 df Temperature        RH         Ws
## (Intercept)      1  2.3643e+05 921338.84 55555.4047
## Classes          1  7.3306e+02   7521.56     2.6376
## Region           1  1.0526e+02   5463.38    63.1884
## Classes:Region   1  3.9051e+00    520.63    13.1607
## residuals      239  2.2204e+03  36708.70  1833.5755
##
##  F-tests
```

```
##                 Temperature      RH      Ws
## (Intercept)        25448.98 5998.58 7241.45
## Classes               78.91   48.97    0.34
## Region                11.33   35.57    8.24
## Classes:Region         0.42    3.39    1.72
##
##  p-values
##                 Temperature RH         Ws
## (Intercept)     < 2.22e-16  < 2.22e-16 < 2.22e-16
## Classes         < 2.22e-16  2.6002e-11 0.55819465
## Region          0.00088866  8.7923e-09 0.00447376
## Classes:Region 0.51738702  0.06684634 0.19153751
```

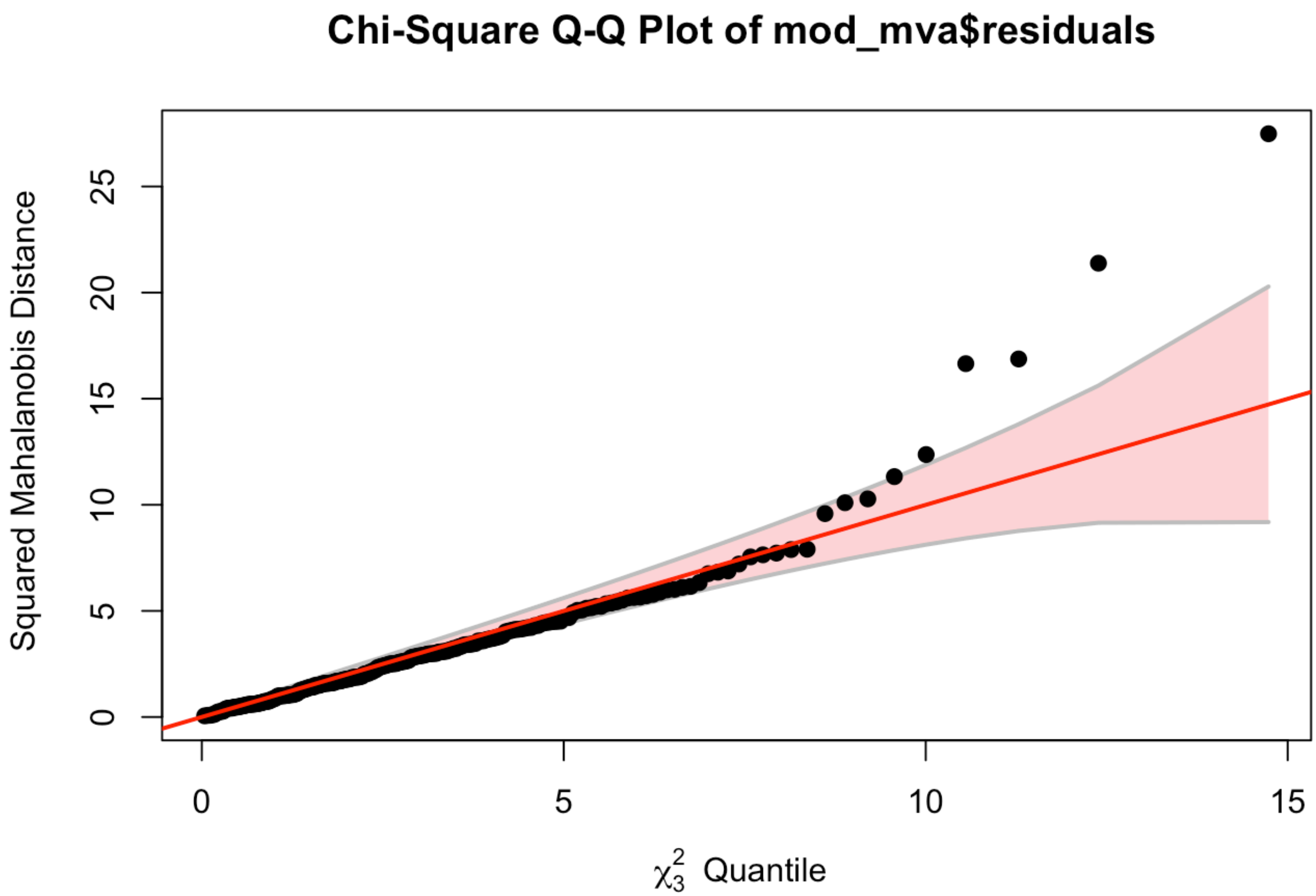
1. Multivariate tests

All the tests show highly significant main effects for both `Classes` and `Region` (p-values << 0.001), i.e. fire vs non-fire differ multivariately, so do Sidi-Bel Abbes vs Bejaia. However, there is no significant `Classes × Region` interaction, i.e. there is no strong evidence that the fire–non-fire difference itself changes between regions.

2. Univariate follow-ups

- Fire presence (`Classes`) is significantly associated with higher temperature and lower relative humidity (p-value << 0.01), but not significantly associated with wind speed (p = 0.558).
- `Region` significantly affects all three meteorological variables, i.e. Bejaia and Sidi-Bel Abbes have different climates/microclimates.
- None of the individual responses show a significant Classes×Region term at alpha = 0.05 (RH is borderline at $p \approx 0.067$), matching the non-significant multivariate interaction.

```
cqplot(mod_mva$residuals, label = "Residuals MANOVA")
```



In the chi-square qq plot, most points track the line quite closely except a few pointing a little upwards towards the tail, but this is good enough for us to say that the multivariate-normality assumption is satisfied.

1.3.1 Three-way MANOVA

Since the two-way MANOVA for Region x Classes is not statistically significant, I decided to add month as a third factor. By doing so, I wanted to ask if the shift in joint distribution of Temperature, RH and Wind attributed to `Classes` and `Region` also change over the seasonal cycle.

```
mod_mva3way <- lm(
  cbind(Temperature, RH, Ws) ~ Classes * Region * month,
  data = df
)

summary(Anova(mod_mva3way, type = 3), univariate = TRUE)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##               Temperature         RH         Ws
## Temperature     1431.0946 -3324.380 -586.9317
## RH             -3324.3801 32901.958 1515.0685
```

```
## Ws             -586.9317  1515.068 1726.7833
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##             Temperature       RH        Ws
## Temperature   190477.65 370048.4  90842.15
## RH            370048.44 718907.7 176482.64
## Ws             90842.15 176482.6  43324.23
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1    0.9973 27442.93      3    225 < 2.22e-16 ***
## Wilks             1    0.0027 27442.93      3    225 < 2.22e-16 ***
## Hotelling-Lawley  1  365.9057 27442.93      3    225 < 2.22e-16 ***
## Roy               1  365.9057 27442.93      3    225 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH         Ws
## Temperature   214.73606 -903.0670 -25.660381
## RH           -903.06699 3797.8251 107.914075
## Ws            -25.66038  107.9141   3.066346
##
## Multivariate Tests: Classes
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1620981 14.50928      3    225 1.1302e-08 ***
## Wilks             1 0.8379019 14.50928      3    225 1.1302e-08 ***
## Hotelling-Lawley  1 0.1934571 14.50928      3    225 1.1302e-08 ***
## Roy               1 0.1934571 14.50928      3    225 1.1302e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   115.80863 -691.2048 -66.89517
## RH           -691.20483 4125.4623 399.26442
## Ws            -66.89517  399.2644  38.64102
##
```

```
## Multivariate Tests: Region
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1256448  10.7775      3    225 1.2069e-06 ***
## Wilks             1 0.8743552  10.7775      3    225 1.2069e-06 ***
## Hotelling-Lawley  1 0.1437000  10.7775      3    225 1.2069e-06 ***
## Roy               1 0.1437000  10.7775      3    225 1.2069e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: month
##
## Sum of squares and products for the hypothesis:
##             Temperature         RH       Ws
## Temperature   633.62565 -989.70150 48.58209
## RH           -989.70150 2083.23480 34.93185
## Ws             48.58209   34.93185 27.09281
##
## Multivariate Tests: month
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.3893691 11.28549      9 681.0000 < 2.22e-16 ***
## Wilks             3 0.6212120 13.14959      9 547.7415 < 2.22e-16 ***
## Hotelling-Lawley  3 0.5927269 14.73036      9 671.0000 < 2.22e-16 ***
## Roy               3 0.5624664 42.55996      3 227.0000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes:Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH         Ws
## Temperature    3.974519 -32.80782   9.197735
## RH           -32.807819 270.81337 -75.923044
## Ws             9.197735 -75.92304  21.285170
##
## Multivariate Tests: Classes:Region
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0279156 2.153795      3    225 0.094333 .
## Wilks             1 0.9720844 2.153795      3    225 0.094333 .
## Hotelling-Lawley  1 0.0287173 2.153795      3    225 0.094333 .
## Roy               1 0.0287173 2.153795      3    225 0.094333 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes:month
```

```
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   14.309106 -66.40631 -8.703779
## RH           -66.406313 350.08024 48.683197
## Ws            -8.703779  48.68320 58.200541
##
## Multivariate Tests: Classes:month
##                  Df test stat approx F num Df   den Df   Pr(>F)
## Pillai            3 0.0491600 1.260580      9 681.0000 0.255094
## Wilks             3 0.9513511 1.260010      9 547.7415 0.255978
## Hotelling-Lawley  3 0.0506000 1.257503      9 671.0000 0.256905
## Roy               3 0.0362476 2.742732      3 227.0000 0.043984 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Region:month
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   12.004969 -57.49608 -2.105804
## RH           -57.496077 443.79159 49.422918
## Ws            -2.105804  49.42292  9.843280
##
## Multivariate Tests: Region:month
##                  Df test stat  approx F num Df   den Df  Pr(>F)
## Pillai            3 0.0242358 0.6162591      9 681.0000 0.78376
## Wilks             3 0.9759051 0.6129820      9 547.7415 0.78643
## Hotelling-Lawley  3 0.0245454 0.6099987      9 671.0000 0.78910
## Roy               3 0.0149135 1.1284560      3 227.0000 0.33832
##
## ------------------------------------------
##
## Term: Classes:Region:month
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature    6.713160 -25.30107 -6.758531
## RH           -25.301071 242.46485 30.175037
## Ws            -6.758531  30.17504  7.542430
##
## Multivariate Tests: Classes:Region:month
##                  Df test stat  approx F num Df   den Df  Pr(>F)
## Pillai            3 0.0130245 0.3299399      9 681.0000 0.96515
## Wilks             3 0.9870151 0.3277180      9 547.7415 0.96585
## Hotelling-Lawley  3 0.0131156 0.3259475      9 671.0000 0.96654
## Roy               3 0.0085968 0.6504900      3 227.0000 0.58340
##
```

```
##  Type III Sums of Squares
##                      df Temperature        RH         Ws
## (Intercept)           1  1.9048e+05 718907.72 43324.2285
## Classes               1  2.1474e+02   3797.83     3.0663
## Region                1  1.1581e+02   4125.46    38.6410
## month                 3  6.3363e+02   2083.23    27.0928
## Classes:Region        1  3.9745e+00    270.81    21.2852
## Classes:month         3  1.4309e+01    350.08    58.2005
## Region:month          3  1.2005e+01    443.79     9.8433
## Classes:Region:month  3  6.7132e+00    242.46     7.5424
## residuals           227  1.4311e+03  32901.96  1726.7833
##
##  F-tests
##                     Temperature      RH      Ws
## (Intercept)            30213.53 4959.95 5695.33
## Classes                   11.35   26.20    0.13
## Region                     6.12    9.49    5.08
## month                    100.51   14.37    1.19
## Classes:Region             0.63    0.62    0.93
## Classes:month              0.76    2.42    7.65
## Region:month               1.90    1.02    1.29
## Classes:Region:month       0.35    0.56    0.33
##
##  p-values
##                     Temperature RH         Ws
## (Intercept)         < 2.22e-16  < 2.22e-16 < 2.22e-16
## Classes             5.7533e-07  6.5502e-07 0.93949273
## Region              0.00050652  6.2592e-06 0.02516277
## month               < 2.22e-16  0.00019222 0.31539600
## Classes:Region      0.42802408  0.60092349 0.42562534
## Classes:month       0.51956923  0.12154775 0.00614145
## Region:month        0.16896363  0.38427470 0.25651488
## Classes:Region:month 0.78560528  0.64354283 0.80330136
```

1. Multivariate results

When testing Temperature, RH and Wind Speed jointly, all three main effects (`Classes` , `Region` and `month`) are highly significant (each $p << 0.01$), but none of the two-or-three-way interactions reach significance on Pillai's trace (all $p > 0.05$). Roy's largest-root test for the fire×month term, however, is $p \approx 0.044$, indicating that there is at least one specific linear combination of (T, RH, Ws) in which the fire×month interaction is detectable.

Follow up univariately: in the univariate F-tests, only Ws has a significant Classes×month effect (F=7.65, p≈0.006), which is the exact response driving Roy's result.

2. Univariate follow-ups

Looking at each response by itself, Temperature and RH behave as expected: the three response variables each explain a highly significant portion of their variability, but none of their interactions do.

On the other hand, for wind speed, region remains significant ($p \approx 0.025$), month does not ($p \approx 0.32$), yet the Classes × month F-test for wind has $p \approx 0.006$, which is significant. In other words, in some months the wind speed effect on fire days is larger compared to non-fire days. `Ws` is the exact response driving Roy's result on the Classes×month effect in the multivariate results.

1.4 Contrasts

```
options(contrasts = c("contr.treatment", "contr.poly"))
contrasts(df$Region)
```

```
##                  Sidi-Bel Abbes
## Bejaia                        0
## Sidi-Bel Abbes                1
```

```
contrasts(df$Classes)
```

```
##            not fire
## fire              0
## not fire          1
```

Therefore, `Region1` is 1 for Sidi-Bel Abbes (0 for Bejaia), and `Classes1` is 1 for not-fire days (0 for fire days).

1.4.1 Multivariate contrasts

```
rownames(coef(mod_mva))
```

```
## [1] "(Intercept)"      "Classes1"         "Region1"          "Classes1:Region1"
```

```
# Test 1
linearHypothesis(mod_mva, "Region1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   105.25526 -758.3203 -81.55311
## RH           -758.32027 5463.3815 587.55618
## Ws            -81.55311  587.5562  63.18839
##
## Sum of squares and products for error:
##             Temperature        RH        Ws
## Temperature   2220.3684 -4706.159 -540.3526
## RH           -4706.1593 36708.699 1654.7711
## Ws            -540.3526  1654.771 1833.5755
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1386396 12.71538      3    237 9.8157e-08 ***
## Wilks             1 0.8613604 12.71538      3    237 9.8157e-08 ***
## Hotelling-Lawley  1 0.1609542 12.71538      3    237 9.8157e-08 ***
## Roy               1 0.1609542 12.71538      3    237 9.8157e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This tests $H_0 : \mu_B = \mu_S$. All four multivariate test statistics agree with $p << 0.01$, so we reject H_0 and conclude that the two regions have significantly different joint means of `(Temperature, RH, Ws)`.

```
# Test 2
linearHypothesis(mod_mva, "Classes1 + 0.5*Classes1:Region1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             Temperature         RH         Ws
## Temperature   565.80576 -1633.5908 -74.566426
## RH          -1633.59083  4716.4932 215.287715
## Ws            -74.56643   215.2877   9.826962
##
## Sum of squares and products for error:
##             Temperature        RH        Ws
## Temperature   2220.3684 -4706.159 -540.3526
## RH           -4706.1593 36708.699 1654.7711
## Ws            -540.3526  1654.771 1833.5755
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.2142208 21.53715      3    237 2.2685e-12 ***
## Wilks             1 0.7857792 21.53715      3    237 2.2685e-12 ***
## Hotelling-Lawley  1 0.2726221 21.53715      3    237 2.2685e-12 ***
## Roy               1 0.2726221 21.53715      3    237 2.2685e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This tests whether the average fire-effect across both regions is zero, i.e.

$$H_0: \frac{(\mu_{fire,B} - \mu_{nonfire,B}) + (\mu_{fire,S} - \mu_{nonfire,S})}{2} = 0$$

The resulting p-values are << 0.01, so we reject H_0. Even when averaging the fire vs. non-fire difference across regions, the effect remains highly significant.

```
# Test 3
linearHypothesis(mod_mva,"Classes1 + Classes1:Region1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   330.59345 -841.7929 -69.17514
## RH           -841.79289 2143.4643 176.14124
## Ws            -69.17514  176.1412  14.47458
##
## Sum of squares and products for error:
##             Temperature        RH        Ws
## Temperature   2220.3684 -4706.159 -540.3526
## RH           -4706.1593 36708.699 1654.7711
## Ws            -540.3526  1654.771 1833.5755
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1315616 11.96789      3    237 2.5255e-07 ***
## Wilks             1 0.8684384 11.96789      3    237 2.5255e-07 ***
## Hotelling-Lawley  1 0.1514922 11.96789      3    237 2.5255e-07 ***
## Roy               1 0.1514922 11.96789      3    237 2.5255e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This tests whether fire vs. non-fire differ in Sidi-Bel Abbes alone. The results suggest that we reject H_0. In Sidi-Bel Abbes, fire days differ significantly from non-fire days.

```
# Test 4
linearHypothesis(mod_mva, "Classes1:Region1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature    3.905114 -45.08995   7.16896
## RH           -45.089947 520.62588 -82.77557
## Ws             7.168960 -82.77557  13.16069
##
## Sum of squares and products for error:
##             Temperature        RH        Ws
## Temperature   2220.3684 -4706.159 -540.3526
## RH           -4706.1593 36708.699 1654.7711
## Ws            -540.3526  1654.771 1833.5755
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1 0.0258281 2.094519      3    237 0.10162
## Wilks             1 0.9741719 2.094519      3    237 0.10162
## Hotelling-Lawley  1 0.0265129 2.094519      3    237 0.10162
## Roy               1 0.0265129 2.094519      3    237 0.10162
```

None of the test results were significant. This confirms again that there is no significant Classes×Region interaction. In other words, fire vs non-fire is comparable in both regions, and the three basic meteorological measurements could be good universal indicators for fire prediction, regardless of region.

1.4.2 Univariate Contrasts

Performing the univariate followup based on the first multivariate contrast (test 1) we have performed:

```
options(contrasts = c("contr.sum","contr.poly"))
mod.T <- lm(Temperature ~ Classes * Region, data = df)
mod.RH <- lm(RH ~ Classes * Region, data = df)
mod.Ws <- lm(Ws ~ Classes * Region, data = df)

names(coef(mod.T))
```

```
## [1] "(Intercept)"      "Classes1"         "Region1"          "Classes1:Region1"
```

```
linearHypothesis(mod.T, "Region1 = 0")
```

```
##
## Linear hypothesis test:
## Region1 = 0
##
## Model 1: restricted model
## Model 2: Temperature ~ Classes * Region
##
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1    240 2325.6
## 2    239 2220.4  1    105.25 11.33 0.0008887 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
linearHypothesis(mod.RH, "Region1 = 0")
```

```
##
## Linear hypothesis test:
## Region1 = 0
##
## Model 1: restricted model
## Model 2: RH ~ Classes * Region
##
##   Res.Df   RSS Df Sum of Sq     F    Pr(>F)
## 1    240 42172
## 2    239 36709  1    5463.4 35.57 8.792e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
linearHypothesis(mod.Ws, "Region1 = 0")
```

```
##
## Linear hypothesis test:
## Region1 = 0
##
## Model 1: restricted model
## Model 2: Ws ~ Classes * Region
##
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    240 1896.8
## 2    239 1833.6  1    63.188 8.2364 0.004474 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The univariate confirm that all three responses differ significantly by region, specially `Temperature` and `RH`.

```r
linearHypothesis(mod.T, "Classes1:Region1 = 0")
```

```
##
## Linear hypothesis test:
## Classes1:Region1 = 0
##
## Model 1: restricted model
## Model 2: Temperature ~ Classes * Region
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    240 2224.3
## 2    239 2220.4  1    3.9051 0.4203 0.5174
```

```r
linearHypothesis(mod.RH, "Classes1:Region1 = 0")
```

```
##
## Linear hypothesis test:
## Classes1:Region1 = 0
##
## Model 1: restricted model
## Model 2: RH ~ Classes * Region
##
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    240 37229
## 2    239 36709  1    520.63 3.3896 0.06685 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

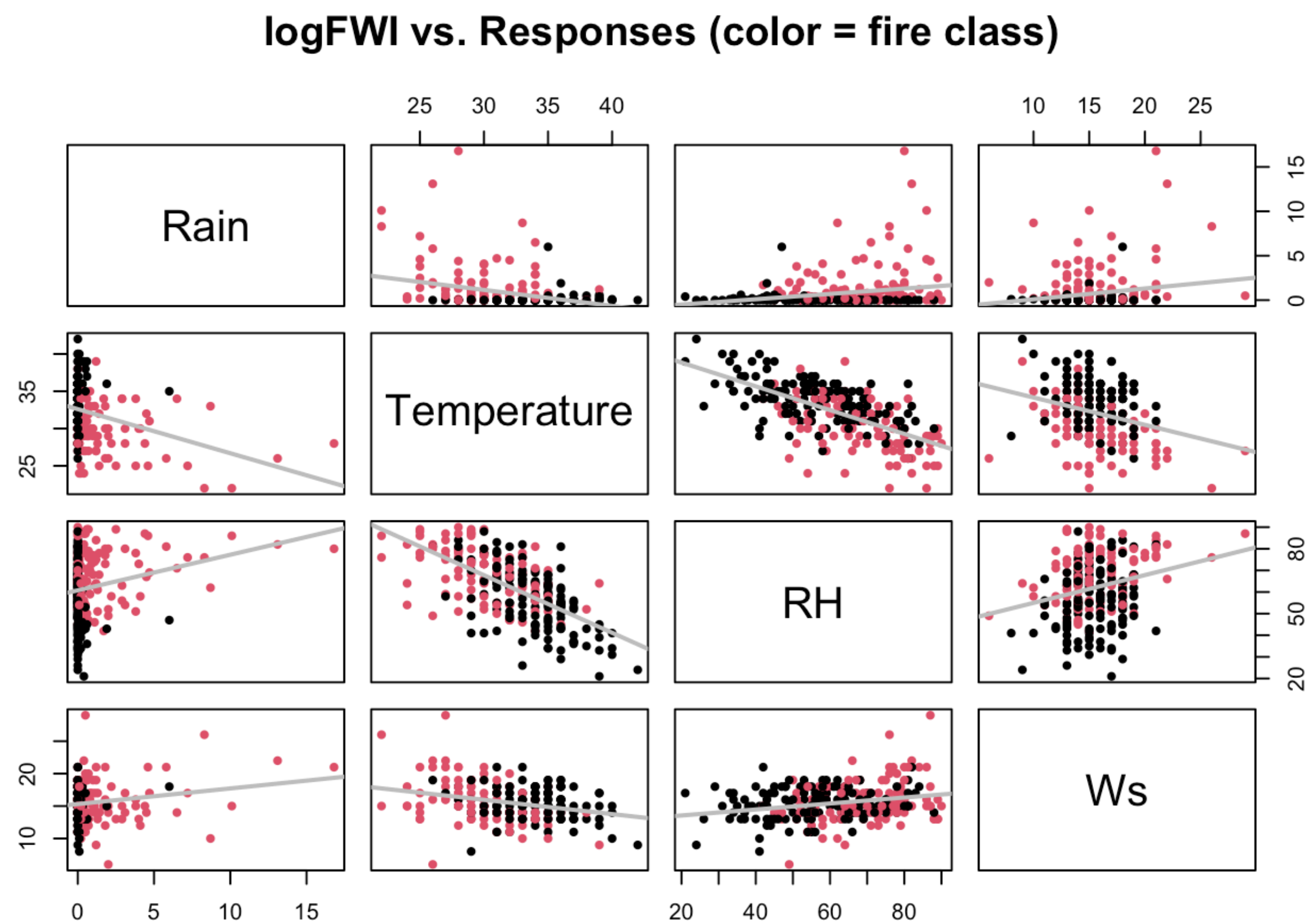
```
linearHypothesis(mod.Ws, "Classes1:Region1 = 0")
```

```
##
## Linear hypothesis test:
## Classes1:Region1 = 0
##
## Model 1: restricted model
## Model 2: Ws ~ Classes * Region
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    240 1846.7
## 2    239 1833.6  1    13.161 1.7154 0.1915
```

Zooming in to the univariate breakdown of Region x Classes interaction, none of these univariate p-values drops below 0.05. The slight signal in RH ($p \approx 0.067$) matches what we saw before but both temperature and wind show no region-specific fire effect.

1.5 Model with Added Variable

```
pairs(df[c("Rain","Temperature","RH","Ws")],
      pch = 20,
      col = as.numeric(df$Classes),
      main = "logFWI vs. Responses (color = fire class)",
      panel = function(x, y, ...) {
    points(x, y, ...)
    abline(lm(y ~ x), col = 'grey', lwd = 2)
  }
)
```



In the pair-plots, as `Rain` increases, average temperature tends to drop, relative humidity rises, and wind speed slightly increase). Considering that rainfall data is always non-negative, and its distribution tend to be naturally highly right skewed, and there isn't too much to do to change that (after several testing), we can loosely accept that the variables are linearly associated.

```
mod_mva2 <- manova(
  cbind(Temperature, RH, Ws) ~ Classes * Region + Rain,
  data = df
)

summary(Anova(mod_mva2, type = 3), univariate = TRUE)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##              Temperature         RH        Ws
## Temperature    2151.9642 -4566.988 -482.7096
## RH            -4566.9875 36425.547 1537.4935
## Ws             -482.7096  1537.493 1785.0008
##
## ------------------------------------------
```

```
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##             Temperature       RH        Ws
## Temperature   199266.70 387262.8  94573.09
## RH            387262.81 752621.9 183797.08
## Ws             94573.09 183797.1  44884.91
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1    0.9962 20602.26      3    236 < 2.22e-16 ***
## Wilks             1    0.0038 20602.26      3    236 < 2.22e-16 ***
## Hotelling-Lawley  1  261.8932 20602.26      3    236 < 2.22e-16 ***
## Roy               1  261.8932 20602.26      3    236 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes
##
## Sum of squares and products for the hypothesis:
##             Temperature          RH         Ws
## Temperature   482.25473 -1624.76762  24.658738
## RH          -1624.76762  5474.01544 -83.077920
## Ws             24.65874   -83.07792   1.260855
##
## Multivariate Tests: Classes
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.2175733 21.87524      3    236 1.5536e-12 ***
## Wilks             1 0.7824267 21.87524      3    236 1.5536e-12 ***
## Hotelling-Lawley  1 0.2780751 21.87524      3    236 1.5536e-12 ***
## Roy               1 0.2780751 21.87524      3    236 1.5536e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature   108.80216 -774.5074 -84.42426
## RH           -774.50743 5513.3259 600.97351
## Ws            -84.42426  600.9735  65.50840
##
## Multivariate Tests: Region
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.1421553 13.03602      3    236 6.5918e-08 ***
```

```
## Wilks             1 0.8578447 13.03602      3    236 6.5918e-08 ***
## Hotelling-Lawley  1 0.1657121 13.03602      3    236 6.5918e-08 ***
## Roy               1 0.1657121 13.03602      3    236 6.5918e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Rain
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH        Ws
## Temperature    68.40418 -139.1717 -57.64301
## RH           -139.17174  283.1519 117.27760
## Ws            -57.64301  117.2776  48.57476
##
## Multivariate Tests: Rain
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0454227 3.743279      3    236 0.011759 *
## Wilks             1 0.9545773 3.743279      3    236 0.011759 *
## Hotelling-Lawley  1 0.0475841 3.743279      3    236 0.011759 *
## Roy               1 0.0475841 3.743279      3    236 0.011759 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Classes:Region
##
## Sum of squares and products for the hypothesis:
##             Temperature        RH         Ws
## Temperature    4.494424 -48.98714   7.432159
## RH           -48.987142 533.93721 -81.007096
## Ws             7.432159 -81.00710  12.290115
##
## Multivariate Tests: Classes:Region
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1 0.0256567 2.071472      3    236 0.10468
## Wilks             1 0.9743433 2.071472      3    236 0.10468
## Hotelling-Lawley  1 0.0263323 2.071472      3    236 0.10468
## Roy               1 0.0263323 2.071472      3    236 0.10468
##
##  Type III Sums of Squares
##                 df Temperature        RH         Ws
## (Intercept)      1  1.9927e+05 752621.90 44884.9124
## Classes          1  4.8225e+02   5474.02     1.2609
## Region           1  1.0880e+02   5513.33    65.5084
## Rain             1  6.8404e+01    283.15    48.5748
## Classes:Region   1  4.4944e+00    533.94    12.2901
## residuals      238  2.1520e+03  36425.55  1785.0008
```

```
##
##  F-tests
##                 Temperature      RH      Ws
## (Intercept)        22038.23 4917.54 5984.65
## Classes               53.34   35.77    0.17
## Region                12.03   36.02    8.73
## Rain                   7.57    1.85    6.48
## Classes:Region         0.50    3.49    1.64
##
##  p-values
##                 Temperature RH         Ws
## (Intercept)     < 2.22e-16  < 2.22e-16 < 2.22e-16
## Classes         4.2068e-12  8.0953e-09 0.68216229
## Region          0.00062022  7.2193e-09 0.00343621
## Rain            0.00640764  0.17506110 0.01156286
## Classes:Region 0.48148058  0.06301886 0.20175279
```

```
summary.aov(mod_mva2)
```

```
##  Response Temperature :
##                  Df  Sum Sq Mean Sq F value    Pr(>F)
## Classes           1  848.17  848.17 93.8052 < 2.2e-16 ***
## Region            1  112.92  112.92 12.4885 0.0004918 ***
## Rain              1   67.81   67.81  7.5001 0.0066362 **
## Classes:Region    1    4.49    4.49  0.4971 0.4814806
## Residuals       238 2151.96    9.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response RH :
##                  Df Sum Sq Mean Sq F value    Pr(>F)
## Classes           1   9938  9937.6 64.9309 3.750e-14 ***
## Region            1   6043  6042.7 39.4822 1.563e-09 ***
## Rain              1    270   269.8  1.7631   0.18551
## Classes:Region    1    534   533.9  3.4887   0.06302 .
## Residuals       238  36426   153.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response Ws :
##                  Df  Sum Sq Mean Sq F value   Pr(>F)
## Classes           1    9.36   9.363  1.2484 0.264993
## Region            1   56.64  56.642  7.5522 0.006453 **
## Rain              1   49.45  49.445  6.5927 0.010852 *
## Classes:Region    1   12.29  12.290  1.6387 0.201753
## Residuals       238 1785.00   7.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multivariate:

After adjusting for `Rain`, `Classes` and `Region` remain highly significant jointly on `(Temperature, RH, Ws)`. `Rain` also shows a small but significant multivariate effect ($p \approx 0.012$), i.e. adding it to the model explains extra variation across T, RH and Ws. The Classes × Region interaction stays non-significant.

Univariate:

- `Temperature ~ Rain`:

  Fire days remain much hotter than non-fire days, S-B Abbes remains cooler than Bejaia, and each extra millimeter of rain cools the day significantly.

- `RH ~ Rain`:

  Fires are still drier and S-B Abbes more humid, but rainfall itself no longer shows a clear univariate effect on RH. This suggests that the big humidity contrasts are driven by `Classes` and `Region` rather than by rainfall.

- `Ws ~ Rain`:

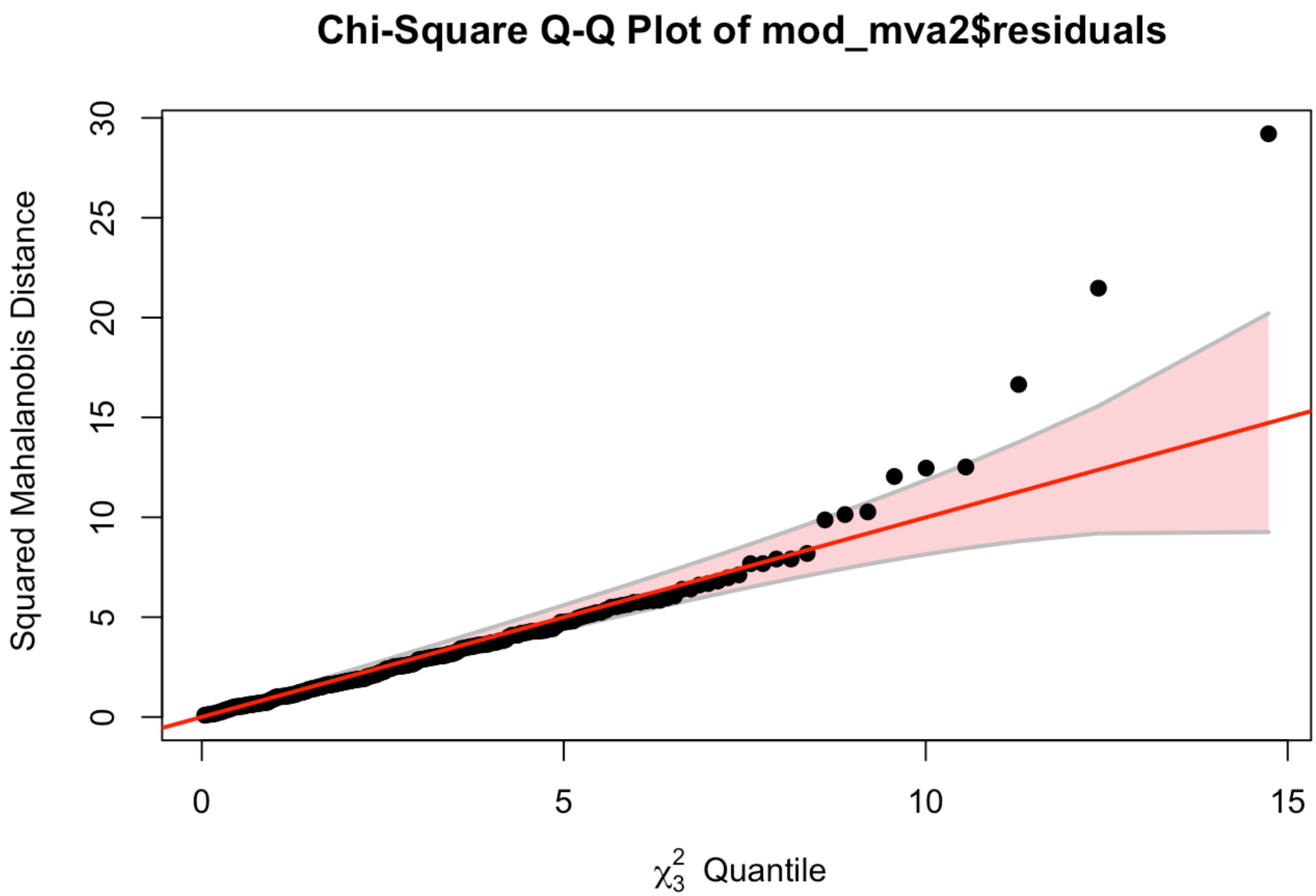
  Wind still does not differ by fire class, but remains higher in S-B Abbes and now also increases modestly with rainfall.

Also note that after we account for rain, the Classes F–stat for Temperature drops from about 78.9 (mod_mva) to 53.3 (mod_mva2), and for RH from about 48.97 to 35.8. This shrinking means part of what we originally attributed to fire effects was actually just days with lower rainfall.

In contrast, the Region F–statistics stay right around 11–12 for Temperature and 35–36 for RH, so the climate difference between Bejaia and Sidi-Bel Abbes isn't explained away by rainfall.

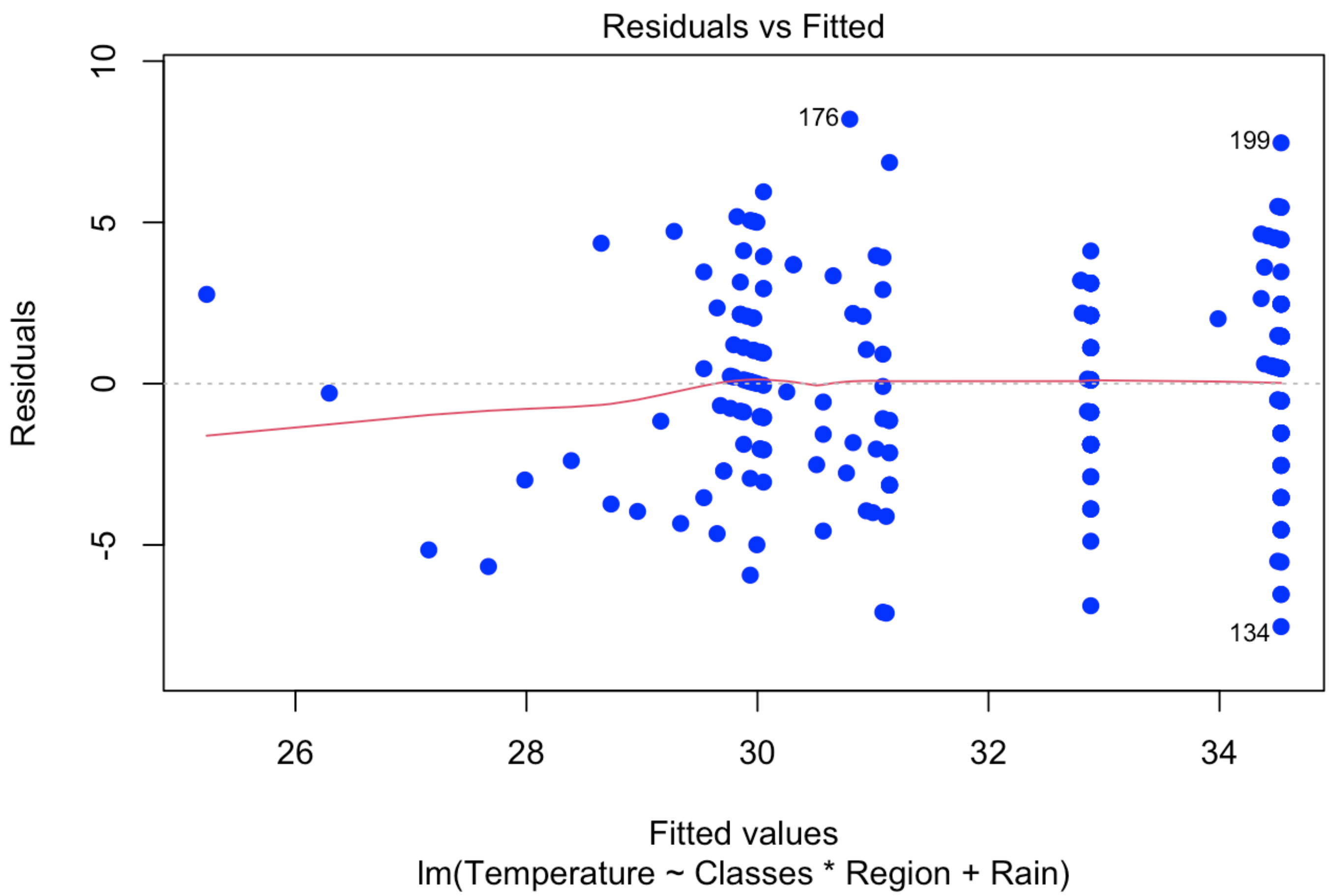
1.6 Model Assumptions

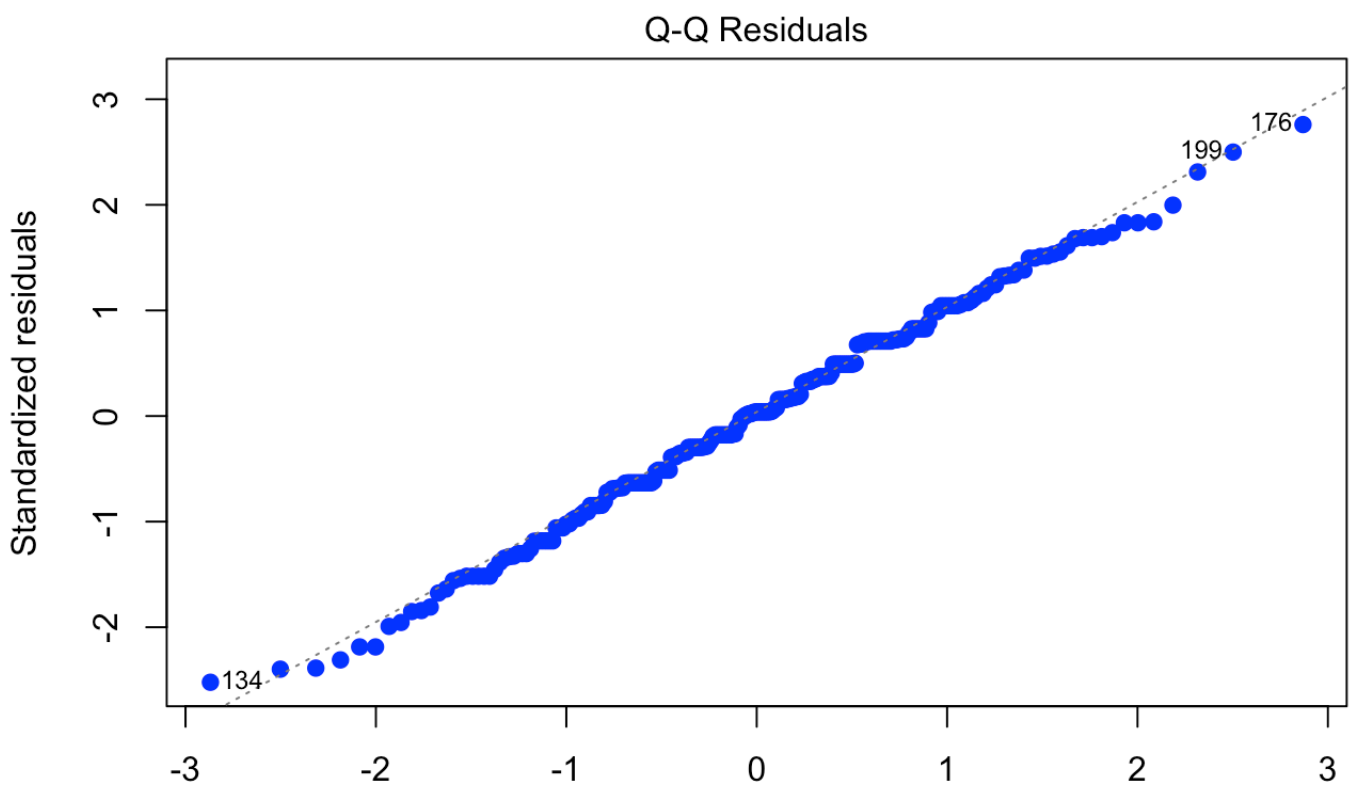
```
mod_mva2 <- manova(cbind(Temperature, RH, Ws) ~ Classes*Region + Rain, data = df)

cqplot(mod_mva2$residuals, label = "MANOVA residuals")
```



The chi-square qq plot seems to show reasonably multivariately normal distribution of the residuals. Still, considering the handful of points lying above the upper band, we could try to look at the residuals vs. fitted plots for our dependent variables and perform a box-cox transformation.

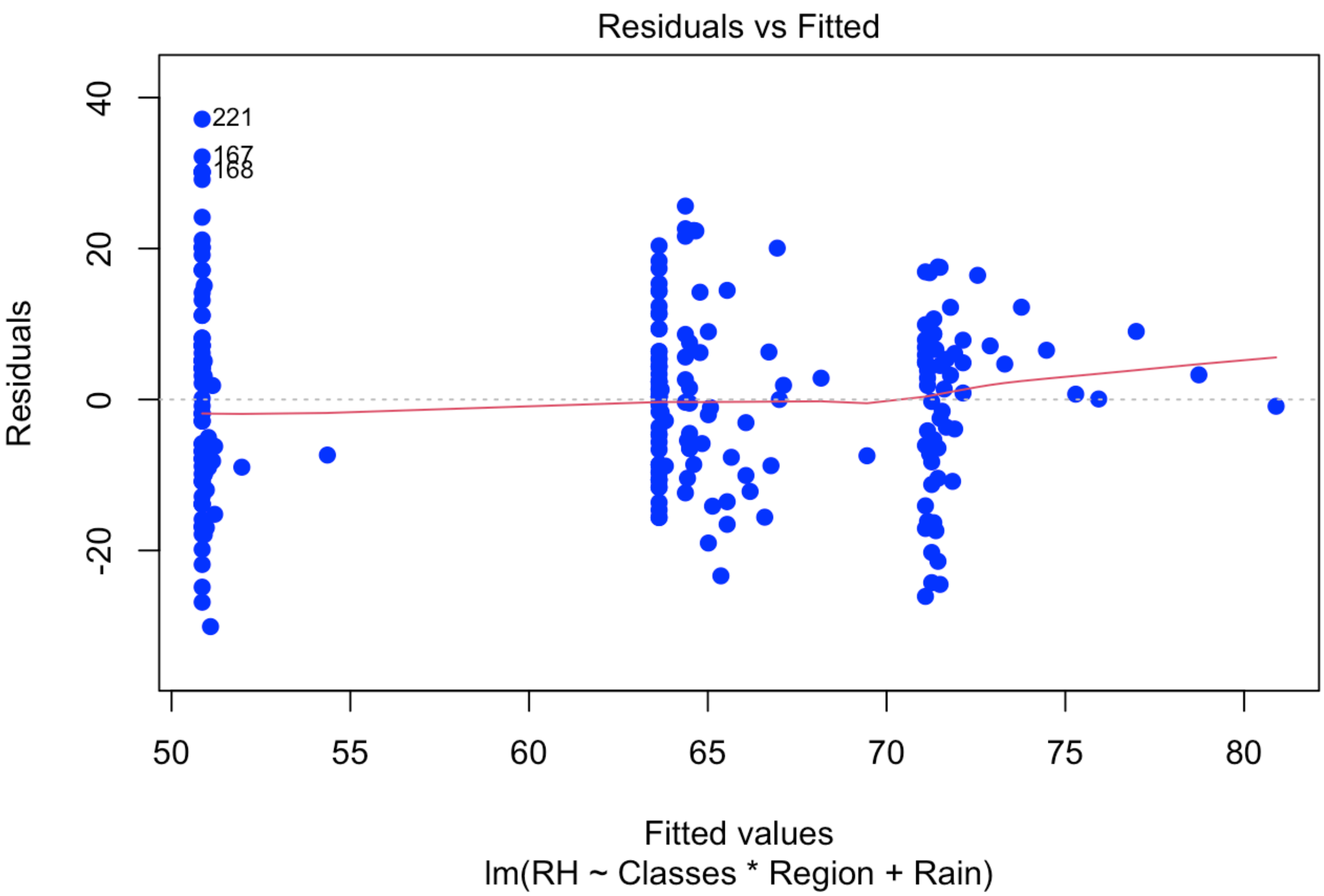
```
mod.T2 <- lm(Temperature ~ Classes * Region + Rain, data = df)
plot(mod.T2, which = c(1,2), pch = 19, col = 'blue')
```

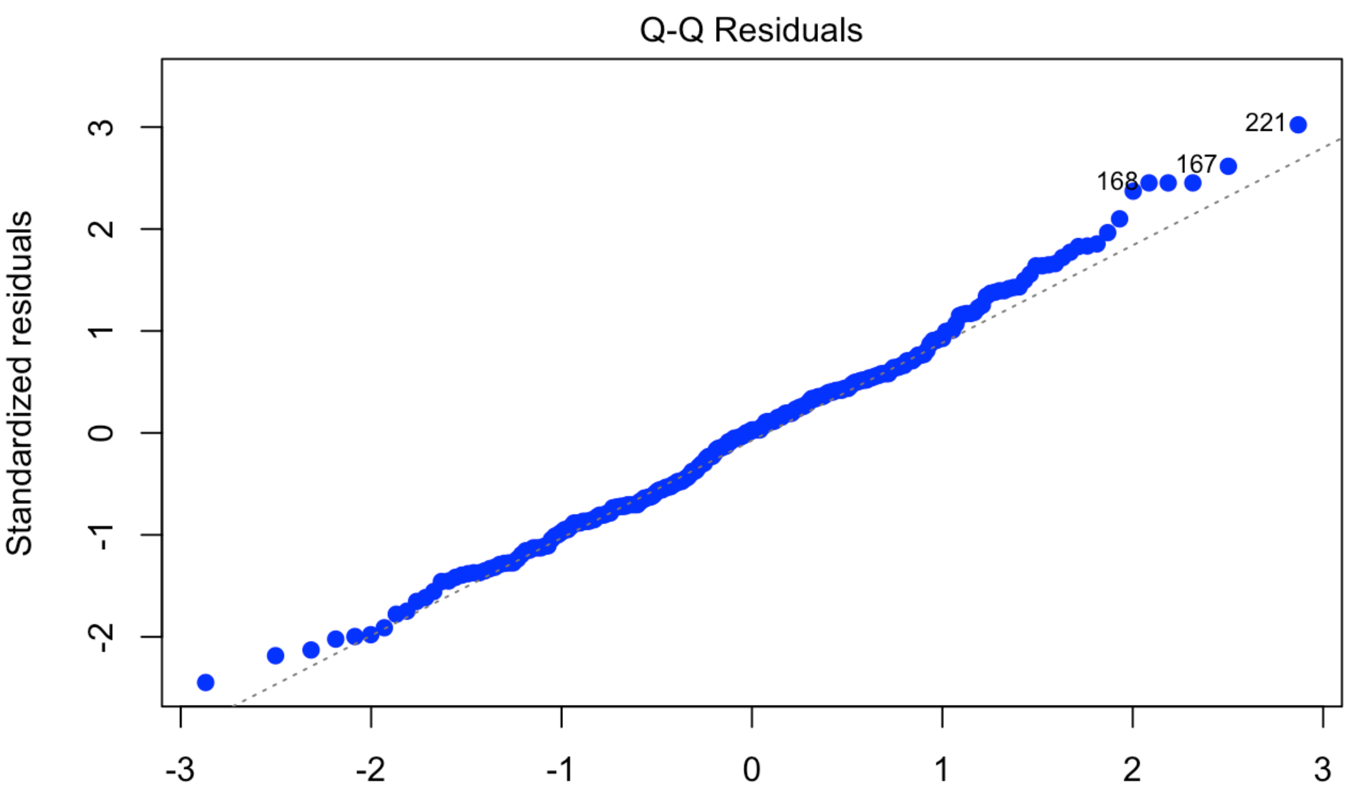

Residuals vs Fitted
Residuals
Fitted values
lm(Temperature ~ Classes * Region + Rain)
176
199
134
10
5
0
-5
26
28
30
32
34


Q-Q Residuals
Standardized residuals
176
199
134
3
2
1
0
-1
-2
-3

Theoretical Quantiles
lm(Temperature ~ Classes * Region + Rain)

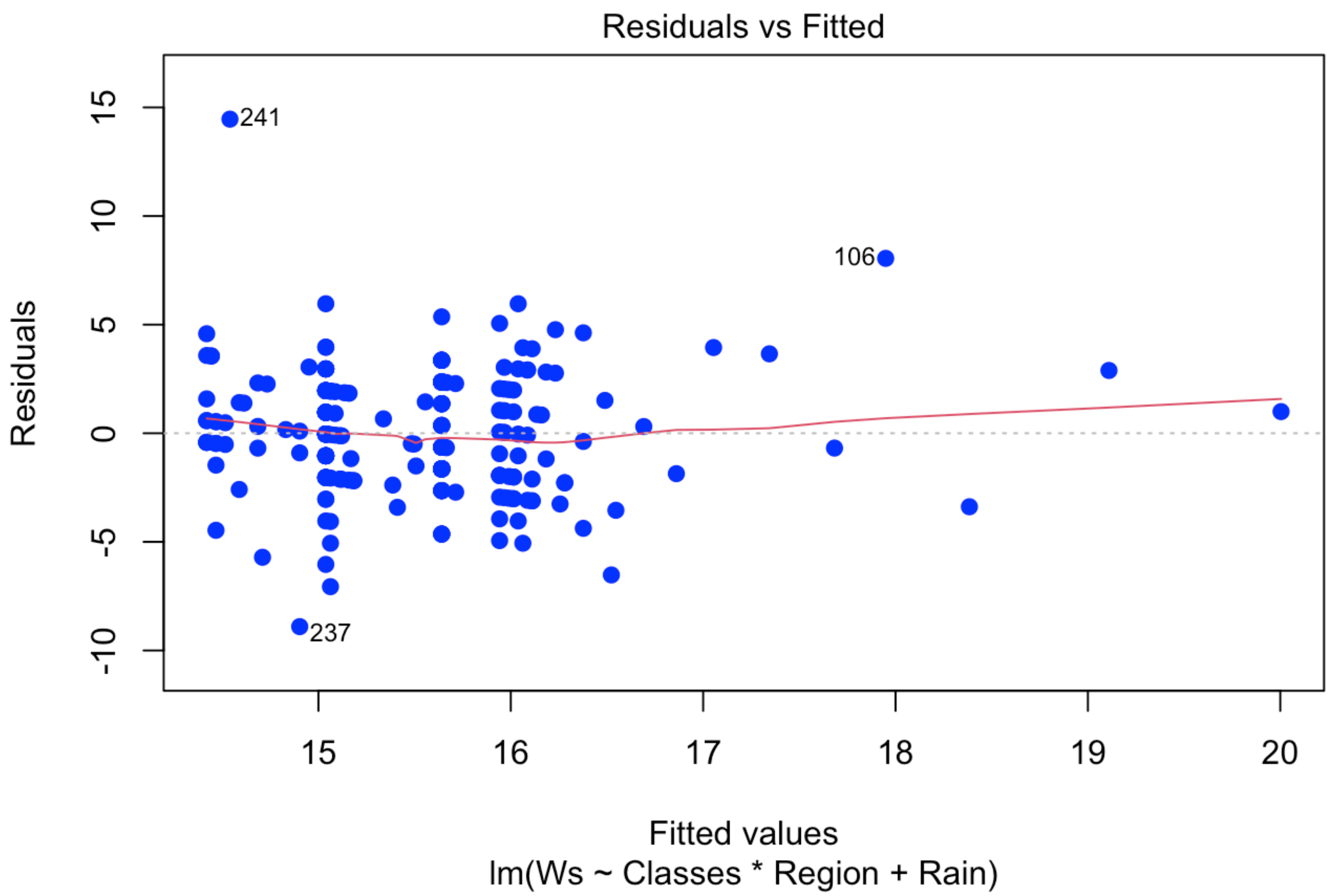
```
mod.RH2 <- lm(RH ~ Classes * Region + Rain, data = df)
plot(mod.RH2, which = c(1,2), pch = 19, col = 'blue')
```

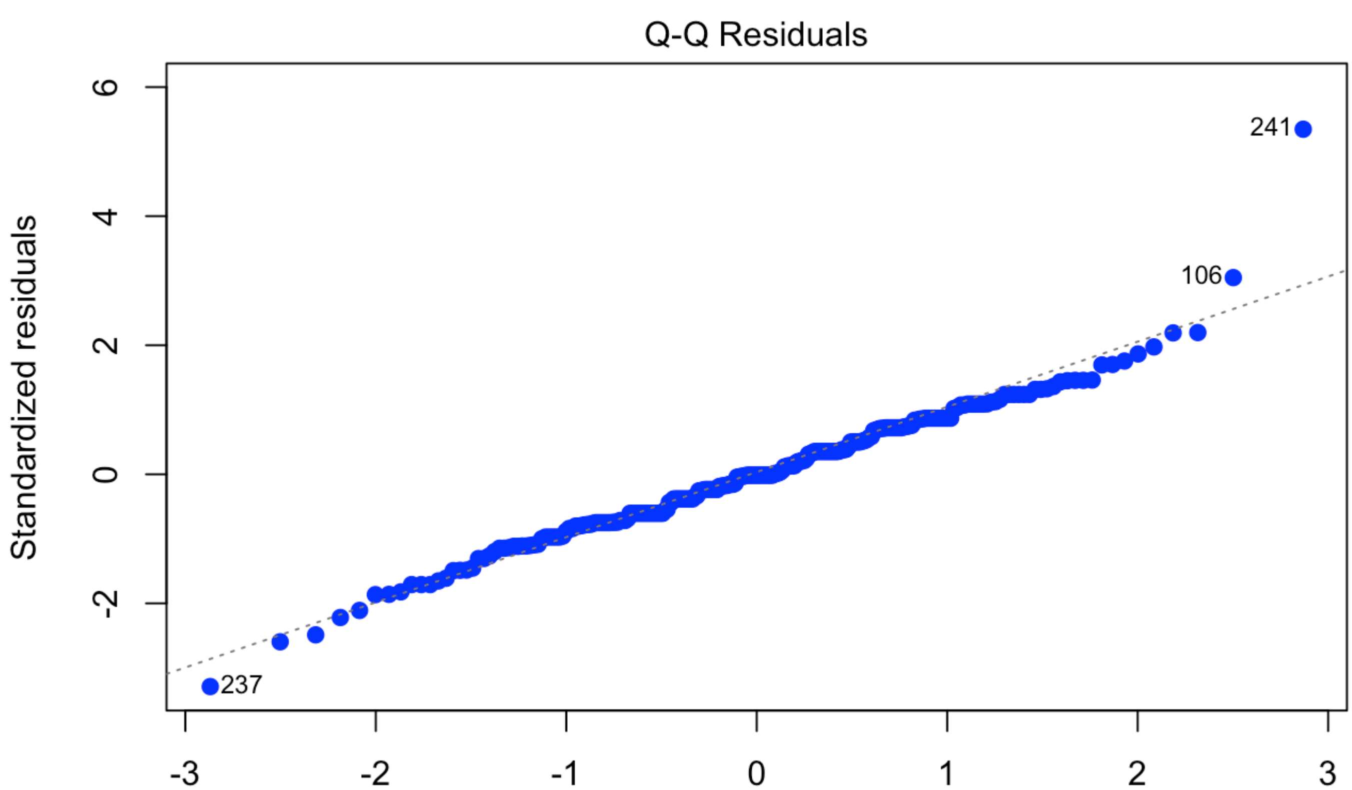

Residuals vs Fitted
Residuals
Fitted values
lm(RH ~ Classes * Region + Rain)
221
167
168
40
20
0
-20
50
55
60
65
70
75
80


Q-Q Residuals
Standardized residuals
221
167
168
3
2
1
0
-1
-2
-3

Theoretical Quantiles
lm(RH ~ Classes * Region + Rain)

```
mod.Ws2 <- lm(Ws ~ Classes * Region + Rain, data = df)
plot(mod.Ws2, which = c(1,2), pch = 19, col = 'blue')
```


Residuals vs Fitted
Residuals
Fitted values
lm(Ws ~ Classes * Region + Rain)
241
106
237
15
10
5
0
-5
-10
15
16
17
18
19
20


Q-Q Residuals
Standardized residuals
241
106
237
6
4
2
0
-2
-3
-2
-1
0
1
2
3

Theoretical Quantiles
lm(Ws ~ Classes * Region + Rain)

There is some heteroskedasticity and uneven spread in the Residuals vs. Fitted plot for all there variabels.

```
bcT  <- powerTransform(mod.T2)
(lamT <- bcT$x[which.max(bcT$y)])
```

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

```
bcH  <- powerTransform(mod.RH2)
(lamH <- bcH$x[which.max(bcH$y)])
```

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

```
bcW  <- powerTransform(mod.Ws2)
(lamW <- bcW$x[which.max(bcW$y)])
```

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

Interestingly, for each margin, the box–cox log-likelihood is maximized at $\lambda = 1$, i.e. no power transformation. This probably means that the residuals are heavy tailed and aren't skewed in a way a simple power could fix.

Moreover, because Pillai's trace is fairly robust to mild normality violations, the MANOVA conclusions should still remain strong.

1.7 MRPP

```
(mrpp1 <- mrpp(df[,c("Temperature","RH","Ws")], df$Classes))
```

```
##
## Call:
## mrpp(dat = df[, c("Temperature", "RH", "Ws")], grouping = df$Classes)
##
## Dissimilarity index: euclidean
## Weights for groups:  n
##
## Class means and counts:
##
##       fire  not fire
## delta 17.42    16
## n     137   106
##
## Chance corrected within-group agreement A: 0.09042
## Based on observed delta 16.8 and expected delta 18.47
##
## Significance of delta: 0.001
## Permutation: free
## Number of permutations: 999
```

The average within-group distance in the 3d space of `Temperature`, `RH` and `Ws` was 16.8, lower than the expected value of 18.47 if fire and non-fire days were drawn from the same distribution. The resulting of A is ~0.09 and permutation p-value is 0.001, meaning fire status explains roughly 9% of the total variation (pretty weak) and that the clustering of days by fire vs. non-fire is highly unlikely by chance.

1.8 Conclusion

Meteorological conditions differ meaningfully between fire and non-fire days. Patterns are consistent across regions and months, with temperature and humidity as the strongest indicators. Wind speed seems to show inconsistent associations and contributs less to the multivariate separation. Rainfall adds modest explanatory power but doesn't override the main effects. Lastly, the overall clustering of fire vs. non-fire days, while statistically significant, was relatively weak.

Given these limitations, I thought MANOVA did not yield particularly novel insights for this dataset. To further explore MANOVA, I repeated this procedure on the loaner datasets, hoping to see stornger group separation.

---

2 Loaner Dataset: Ohio Crime

```
crime <- read_csv("ohiocrimehm.csv")
```

```
## Rows: 559 Columns: 12
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (12): V10, V12, V16, V23, V64, V67, V70, V71, V72, V73, V86, V87
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Our four dependent variables are:

- GovtJobs (V10): During the summer, the state government should help provide jobs for inner-city youths from poor families.
- RecreationPrograms (V12): Develop recreation programs, like midnight basketball, so that youths will have something to do instead of wandering the streets at night.
- DrugTreatment (V16): Instead of arresting people who use drugs, get them into treatment programs that can help get them off drugs.
- FamilyHelp (V23): Provide help to families and their children as soon as a child shows signs that he or she might later get into trouble with the law.

And 2 predictors (plus their interaction term):

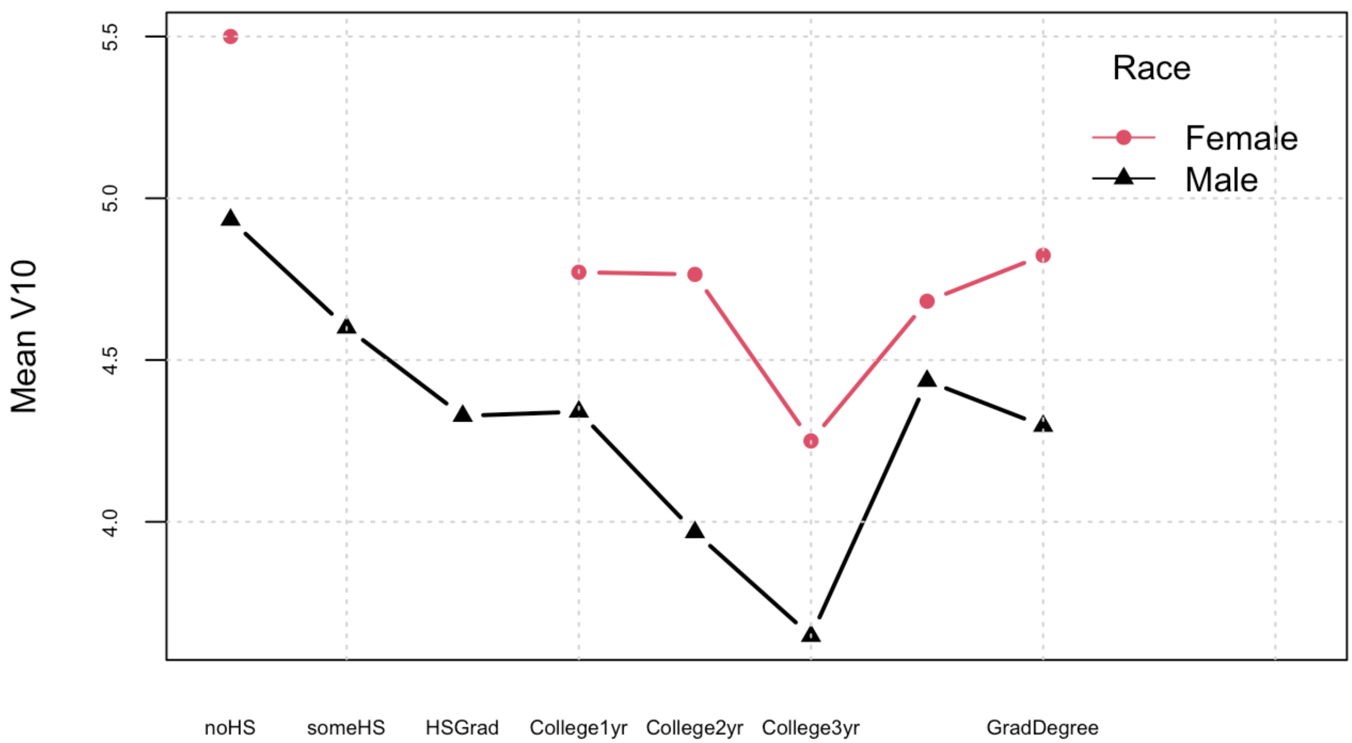
- Gender (V70): 0 = female, 1 = male
- Education (V72):
  - 1 = never went to high school,
  - 2 = went to high school but did not graduate,
  - 3 = graduated from high school,
  - 4 = finished one year of college (or post-high school training),
  - 5 = finished two years of college,
  - 6 = finished three years of college,
  - 7 = graduated from college,
  - 8 = finished one or more years of graduate school
- Gender x Education interaction

```
responses <- c("V10","V12","V16","V23")
names(responses) <- c(
  "GovtJobs",
  "RecreationPrograms",
  "DrugTreatment",
  "FamilyHelp"
)

crime$Gender <- factor(crime$V70, levels=c(0,1), labels=c("Female","Male"))
crime$Race <- factor(crime$V71, levels=c(1,2,3),
                     labels=c("White","Black","Other"))
crime$Education <-  factor(crime$V72, levels=1:8,
                           labels=c("noHS", "someHS", "HSGrad",
                                    "College1yr", "College2yr", "College3yr",
                                    "CollegeGrad","GradDegree"))
```

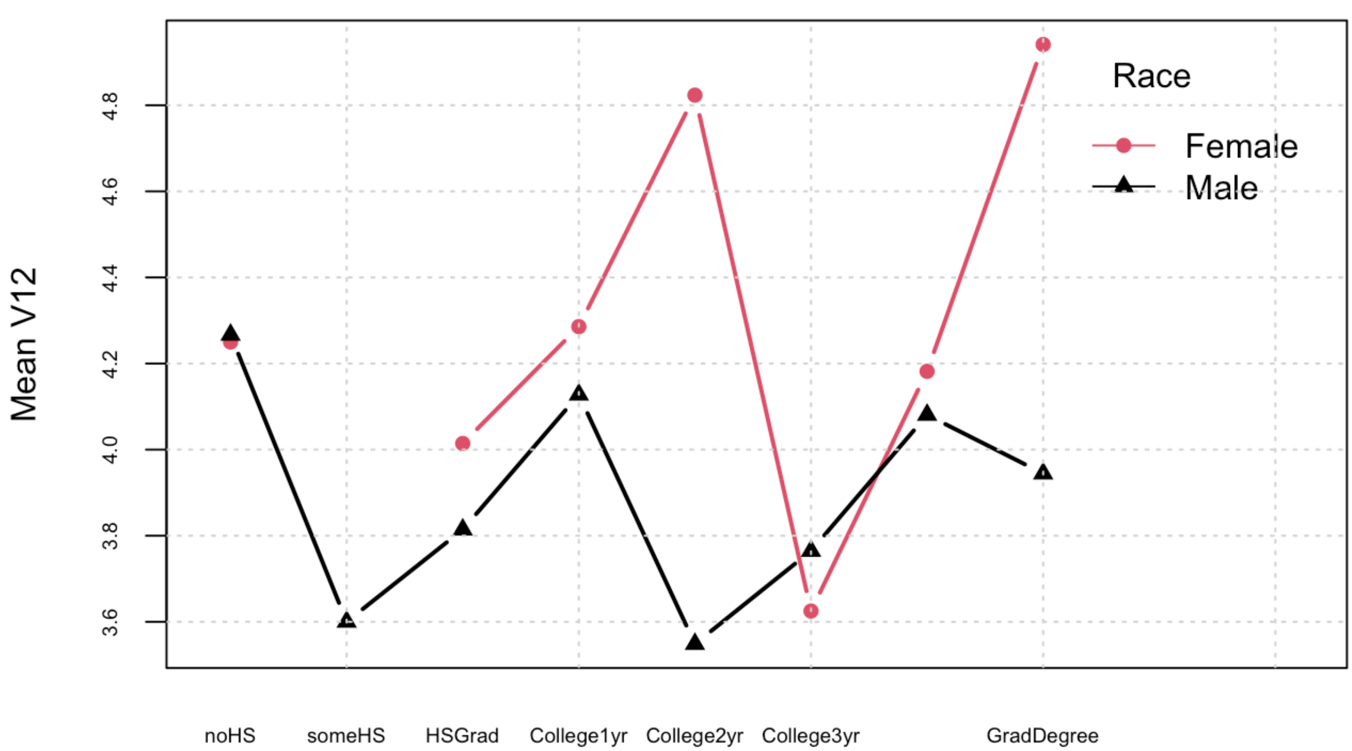
2.1 Interaction Plots

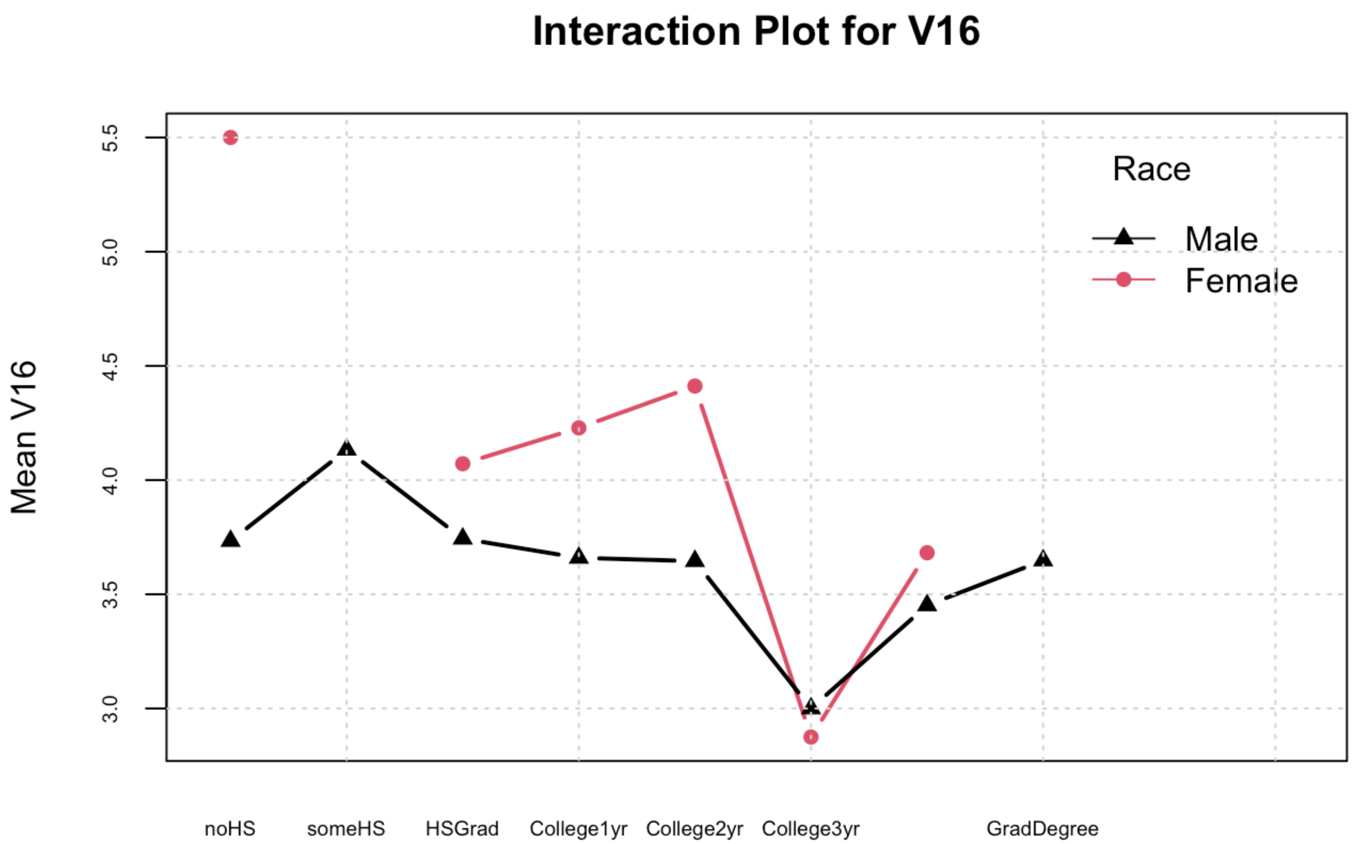
```
for (v in responses) {
  interaction.plot(
    x.factor = crime$Education,
    trace.factor = crime$Gender,
    response = crime[[v]],
    type = "b", lwd = 2,
    lty = 1, col = c(2, 1, 4), pch = 16:18,
    trace.label = "Race",
    xlab = '',
    ylab = paste("Mean", v),
    main = paste("Interaction Plot for", v),
    cex.axis = 0.6
  )
    grid()
}
```

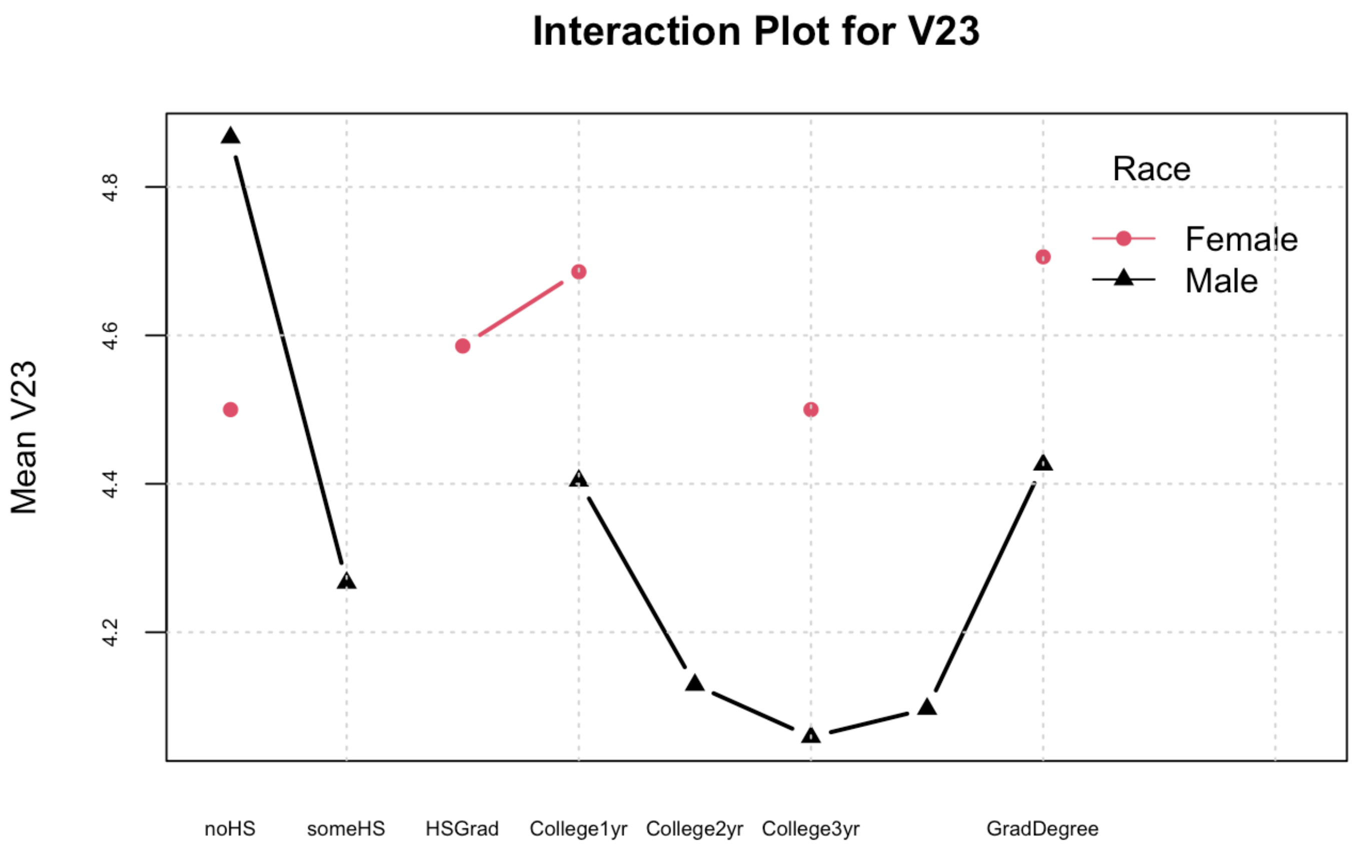
Interaction Plot for V10



Interaction Plot for V12




Interaction Plot for V16
Mean V16
5.5
5.0
4.5
4.0
3.5
3.0
Race
Male
Female
noHS
someHS
HSGrad
College1yr
College2yr
College3yr
GradDegree



For V10, females generally exhibit higher mean scores than males, with both genders following relatively parallel trends across education levels, indicating limited interaction.

In both V12 and V16, the patterns are more irregular, though females still tend to report higher mean scores. Male responses show less variability and consistently lower values, suggesting potential gender differences.

In V23, data for females seem to be limited, but among males, there is strong support for family involvement among those with no high school education. However, this support declines among those with 2–3 years of college education.

Overall, the presence of intersecting lines in these plots suggests interaction effects between gender and education. However, there seem to be a consistent dip in scores at the College3yr level across variables and genders. This may reflect not just genuine attitude differences but also survey-related reasons, for example interpretation ambiguity or response tendencies within this education subgroup.

2.2 Two-Way MANOVA

```r
options(contrasts = c("contr.sum","contr.poly"))

crime_mva <- lm(
  cbind(V10, V12, V16, V23) ~ Education *Gender, data = crime
)
```

2.2.1 Multivariate Results

```r
summary(Anova(crime_mva, type = 3), univariate = TRUE)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##          V10      V12      V16      V23
## V10 5626.638 5070.704 4757.690 5486.543
## V12 5070.704 4569.699 4287.612 4944.452
## V16 4757.690 4287.612 4022.938 4639.231
## V23 5486.543 4944.452 4639.231 5349.937
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1   0.91252 1374.308      4    527 < 2.22e-16 ***
## Wilks             1   0.08748 1374.308      4    527 < 2.22e-16 ***
## Hotelling-Lawley  1  10.43118 1374.308      4    527 < 2.22e-16 ***
## Roy               1  10.43118 1374.308      4    527 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Education
##
## Sum of squares and products for the hypothesis:
##           V10        V12        V16      V23
## V10 29.474873 13.3177167 21.5957023 5.926393
## V12 13.317717 21.1965724  0.2303029 4.988018
## V16 21.595702  0.2303029 49.9892539 2.081062
```

```
## V23  5.926393  4.9880180  2.0810618 3.231626
##
## Multivariate Tests: Education
##                  Df test stat approx F num Df   den Df    Pr(>F)
## Pillai            7 0.0901985 1.746717     28 2120.000 0.0091167 **
## Wilks             7 0.9120955 1.755370     28 1901.548 0.0086513 **
## Hotelling-Lawley  7 0.0938807 1.761939     28 2102.000 0.0082081 **
## Roy               7 0.0550295 4.166522      7  530.000 0.0001766 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Gender
##
## Sum of squares and products for the hypothesis:
##          V10      V12      V16       V23
## V10 19.20871 17.31075 18.12485 11.174032
## V12 17.31075 15.60032 16.33399 10.069958
## V16 18.12485 16.33399 17.10215 10.543535
## V23 11.17403 10.06996 10.54353  6.500125
##
## Multivariate Tests: Gender
##                  Df test stat approx F num Df den Df    Pr(>F)
## Pillai            1 0.0289081 3.922021      4    527 0.0037885 **
## Wilks             1 0.9710919 3.922021      4    527 0.0037885 **
## Hotelling-Lawley  1 0.0297687 3.922021      4    527 0.0037885 **
## Roy               1 0.0297687 3.922021      4    527 0.0037885 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Education:Gender
##
## Sum of squares and products for the hypothesis:
##            V10         V12         V16        V23
## V10 14.4746321 11.53835665  3.27713852  0.6236801
## V12 11.5383567 24.41095765  0.05604968  0.0196499
## V16  3.2771385  0.05604968 14.34669471 -4.9049821
## V23  0.6236801  0.01964990 -4.90498206  6.0417494
##
## Multivariate Tests: Education:Gender
##                  Df test stat approx F num Df   den Df   Pr(>F)
## Pillai            7 0.0593362 1.140062     28 2120.000 0.279521
## Wilks             7 0.9418382 1.138088     28 1901.548 0.281986
## Hotelling-Lawley  7 0.0605158 1.135752     28 2102.000 0.284486
## Roy               7 0.0253478 1.919188      7  530.000 0.064482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##  Type III Sums of Squares
##                  df      V10      V12      V16       V23
## (Intercept)        1 5626.638 4569.699 4022.938 5349.9372
## Education          7   29.475   21.197   49.989    3.2316
## Gender             1   19.209   15.600   17.102    6.5001
## Education:Gender   7   14.475   24.411   14.347    6.0417
## residuals        530  845.269 1418.549 1158.744  804.9742
##
##  F-tests
##                     V10    V12     V16    V23
## (Intercept)      3528.01 243.91 1840.06 503.20
## Education          18.48   1.13   22.86   0.30
## Gender             12.04   0.83    7.82   0.61
## Education:Gender    9.08   1.30    6.56   0.57
##
##  p-values
##                  V10        V12        V16        V23
## (Intercept)      < 2.22e-16 < 2.22e-16 < 2.22e-16 < 2.22e-16
## Education        2.0424e-05 0.34168434 2.2554e-06 0.95208548
## Gender           0.00056188 0.56043493 0.00534780 0.74671260
## Education:Gender 0.00271376 0.24658261 0.01069308 0.78187967
```

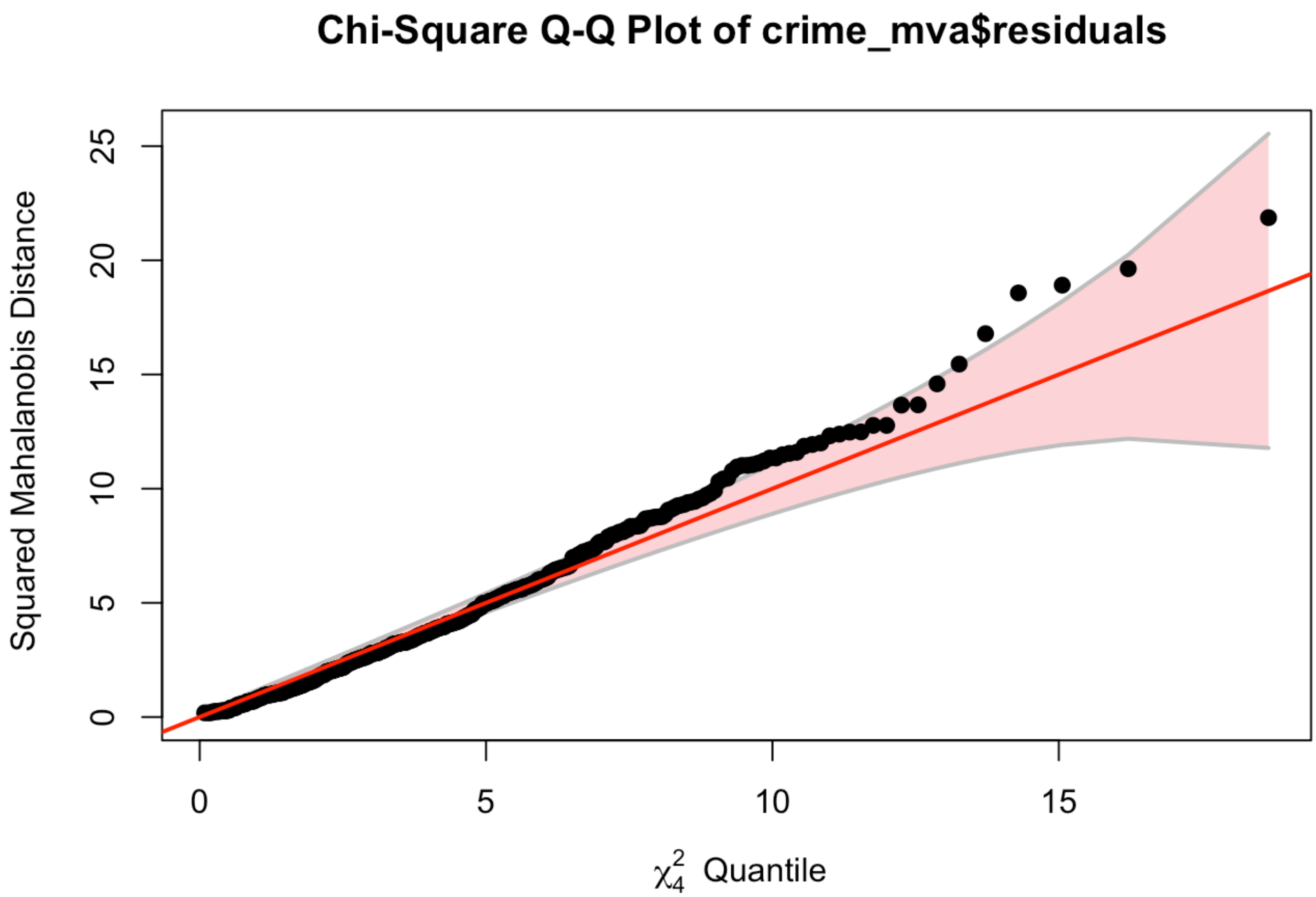
Interpretation:

Both `Education` and `Gender` show significant multivariate effect (p < 0.01), confirming that attitudes differ individually among different education levels and the two gender.

However, the Education × Gender interaction is not significant (p > 0.40), i.e. no evidence that the gender gap changes across education levels. As with the forest fire data, in this dataset, Roy's test again comes close ($p \approx 0.064$), corresponding to the univariate results where Education × Gender is significant for V10 and V16, but not much from V12 or V23.

Other univariate followups suggest that for V10 and V12, Education, Gender, and their interactions are highly significant. For V12 and V23, however, no significant effects were found, i.e. consistant responses across groups.

```
cqplot(crime_mva$residuals, label = "Residuals MANOVA")
```



All data points seem to lie within the bounds, so the residuals are multivariately normally distributed, so assumptions of the model are met.

2.3 Contrasts

```
options(contrasts = c("contr.treatment", "contr.poly"))
contrasts(crime$Gender)
```

```
##        Male
## Female    0
## Male      1
```

```
contrasts(crime$Education)
```

```
##              someHS HSGrad College1yr College2yr College3yr CollegeGrad
## noHS              0      0          0          0          0           0
## someHS            1      0          0          0          0           0
## HSGrad            0      1          0          0          0           0
## College1yr        0      0          1          0          0           0
## College2yr        0      0          0          1          0           0
## College3yr        0      0          0          0          1           0
## CollegeGrad       0      0          0          0          0           1
## GradDegree        0      0          0          0          0           0
##              GradDegree
## noHS                  0
## someHS                0
## HSGrad                0
## College1yr            0
## College2yr            0
## College3yr            0
## CollegeGrad           0
## GradDegree            1
```

```
rownames(coef(crime_mva))
```

```
##  [1] "(Intercept)"        "Education1"         "Education2"
##  [4] "Education3"         "Education4"         "Education5"
##  [7] "Education6"         "Education7"         "Gender1"
## [10] "Education1:Gender1" "Education2:Gender1" "Education3:Gender1"
## [13] "Education4:Gender1" "Education5:Gender1" "Education6:Gender1"
## [16] "Education7:Gender1"
```

```
linearHypothesis(crime_mva, "Gender1 = 0")
```

```
## 
## Sum of squares and products for the hypothesis:
##          V10      V12      V16       V23
## V10 19.20871 17.31075 18.12485 11.174032
## V12 17.31075 15.60032 16.33399 10.069958
## V16 18.12485 16.33399 17.10215 10.543535
## V23 11.17403 10.06996 10.54353  6.500125
## 
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
## 
## Multivariate Tests: 
##                  Df test stat approx F num Df den Df    Pr(>F)   
## Pillai            1 0.0289081 3.922021      4    527 0.0037885 **
## Wilks             1 0.9710919 3.922021      4    527 0.0037885 **
## Hotelling-Lawley  1 0.0297687 3.922021      4    527 0.0037885 **
## Roy               1 0.0297687 3.922021      4    527 0.0037885 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This tests overall gender effect. The results shows that there is a significant multivariate gender gap.

```
linearHypothesis(crime_mva, "Education2 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##            V10        V12       V16        V23
## V10  6.1161708 -2.3670829  8.897320 -0.8829047
## V12 -2.3670829  0.9161094 -3.443445  0.3417021
## V16  8.8973204 -3.4434446 12.943116 -1.2843798
## V23 -0.8829047  0.3417021 -1.284380  0.1274524
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0222346  2.99602      4    527 0.018327 *
## Wilks             1 0.9777654  2.99602      4    527 0.018327 *
## Hotelling-Lawley  1 0.0227402  2.99602      4    527 0.018327 *
## Roy               1 0.0227402  2.99602      4    527 0.018327 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result shows that high-school graduation vs. no high-school shift the joint profile of attitudes in a statistically significant way.

```
linearHypothesis(crime_mva, "Education7 - Education3 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##           V10        V12       V16        V23
## V10  4.770419  4.0207009 -5.740616  1.1859195
## V12  4.020701  3.3888080 -4.838422  0.9995405
## V16 -5.740616 -4.8384216  6.908129 -1.4271090
## V23  1.185919  0.9995405 -1.427109  0.2948179
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0183233 2.459153      4    527 0.044589 *
## Wilks             1 0.9816767 2.459153      4    527 0.044589 *
## Hotelling-Lawley  1 0.0186653 2.459153      4    527 0.044589 *
## Roy               1 0.0186653 2.459153      4    527 0.044589 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a small but statistically significant shift in the combined attitude vector when going from SomeCollege to GradDegree.

```
linearHypothesis(crime_mva, "Gender1 + Education7:Gender1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##           V10      V12       V16       V23
## V10 2.1946681 1.482640 0.7030148  5.020247
## V12 1.4826402 1.001619 0.4749320  3.391501
## V16 0.7030148 0.474932 0.2251957  1.608128
## V23 5.0202469 3.391501 1.6081284 11.483686
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df   Pr(>F)
## Pillai            1 0.0148352 1.983972      4    527 0.095629 .
## Wilks             1 0.9851648 1.983972      4    527 0.095629 .
## Hotelling-Lawley  1 0.0150586 1.983972      4    527 0.095629 .
## Roy               1 0.0150586 1.983972      4    527 0.095629 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This test looks at the gender gap specifically at the top education level. It turns out that the male–female difference among those with a graduate degree is not significant at alpha = .05 despite a weak trend ($p \approx .096$).

```
linearHypothesis(crime_mva, "Education2:Gender1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##             V10         V12         V16          V23
## V10  0.77089767  0.62552971  0.62427820 -0.057496661
## V12  0.62552971  0.50757374  0.50655823 -0.046654531
## V16  0.62427820  0.50655823  0.50554475 -0.046561189
## V23 -0.05749666 -0.04665453 -0.04656119  0.004288333
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat  approx F num Df den Df Pr(>F)
## Pillai            1 0.0013739 0.1812589      4    527 0.9481
## Wilks             1 0.9986261 0.1812589      4    527 0.9481
## Hotelling-Lawley  1 0.0013758 0.1812589      4    527 0.9481
## Roy               1 0.0013758 0.1812589      4    527 0.9481
```

Based on the resulting p-values, there is no evidence that the male–female gap at HSGrad differs from the male–female gap at noHS.

```
linearHypothesis(crime_mva, "Education6:Gender1 - Education3:Gender1 = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##            V10        V12        V16         V23
## V10  2.2400933 -1.2663431 -1.4528299  0.46559125
## V12 -1.2663431  0.7158742  0.8212967 -0.26320256
## V16 -1.4528299  0.8212967  0.9422441 -0.30196282
## V23  0.4655913 -0.2632026 -0.3019628  0.09677062
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            1 0.0069285 0.9191999      4    527 0.45236
## Wilks             1 0.9930715 0.9191999      4    527 0.45236
## Hotelling-Lawley  1 0.0069768 0.9191999      4    527 0.45236
## Roy               1 0.0069768 0.9191999      4    527 0.45236
```

The gender gap at CollegeGrad is not significantly different from the gender gap at HSGrad.

Overall interaction test:

```
linearHypothesis(crime_mva, c(
  "Education1:Gender1 = 0",
  "Education2:Gender1 = 0",
  "Education3:Gender1 = 0",
  "Education4:Gender1 = 0",
  "Education5:Gender1 = 0",
  "Education6:Gender1 = 0",
  "Education7:Gender1 = 0"
))
```

```
##
## Sum of squares and products for the hypothesis:
##            V10         V12         V16        V23
## V10 14.4746321 11.53835665  3.27713852  0.6236801
## V12 11.5383567 24.41095765  0.05604968  0.0196499
## V16  3.2771385  0.05604968 14.34669471 -4.9049821
## V23  0.6236801  0.01964990 -4.90498206  6.0417494
##
## Sum of squares and products for error:
##          V10       V12       V16      V23
## V10 845.2686  484.0144  332.7784 304.6463
## V12 484.0144 1418.5489  354.8061 383.3887
## V16 332.7784  354.8061 1158.7445 241.4170
## V23 304.6463  383.3887  241.4170 804.9742
##
## Multivariate Tests:
##                  Df test stat approx F num Df   den Df   Pr(>F)
## Pillai            7 0.0593362 1.140062     28 2120.000 0.279521
## Wilks             7 0.9418382 1.138088     28 1901.548 0.281986
## Hotelling-Lawley  7 0.0605158 1.135752     28 2102.000 0.284486
## Roy               7 0.0253478 1.919188      7  530.000 0.064482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

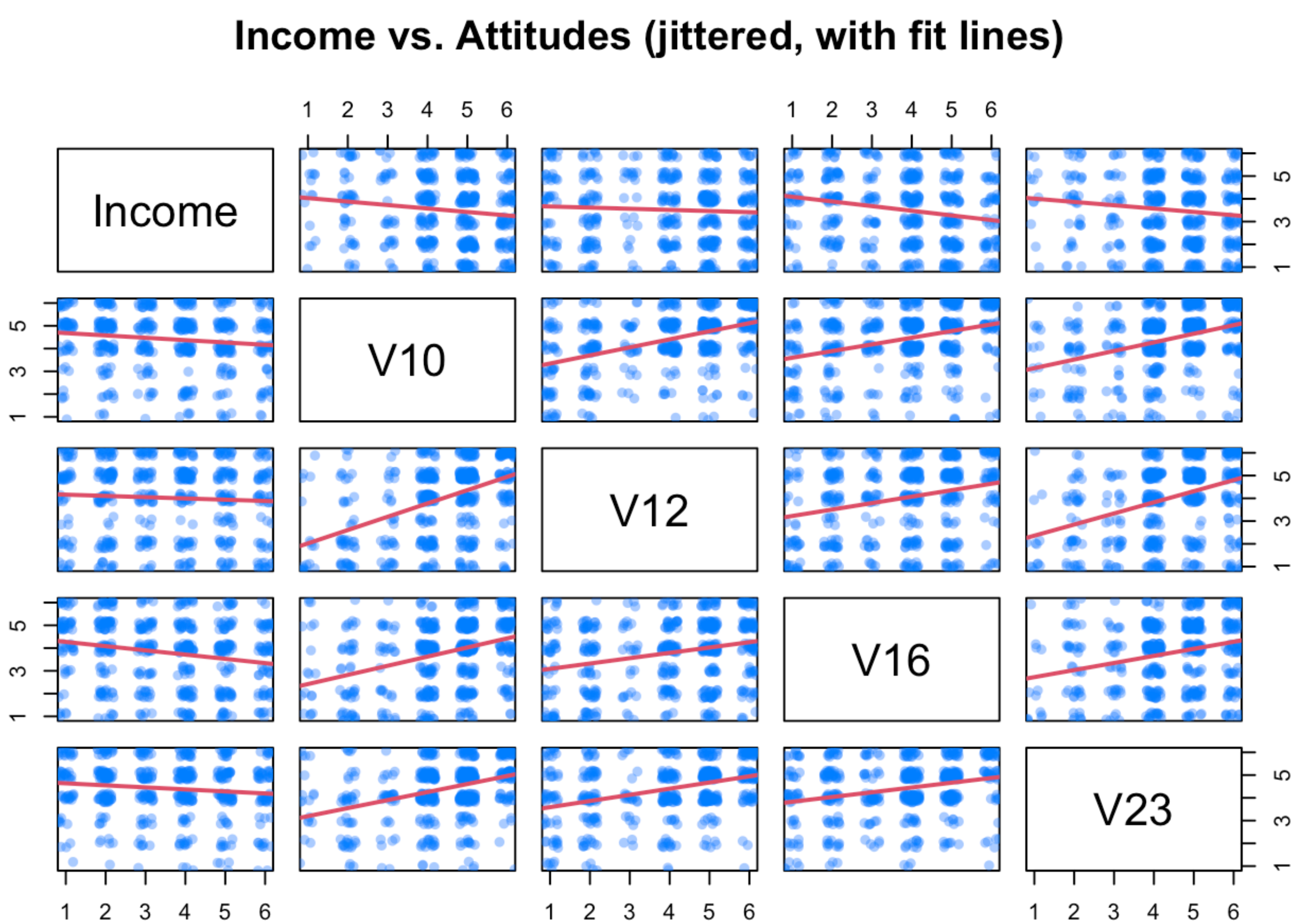
Jointly, there is no significant interaction.

Putting everything together, Gender has a significant main effect across all four response variables. Big gaps in Education (HSGrad vs noHS) produces significantly different responses, smaller educaitonal gaps (someHS vs noHS) produces little difference.

Overall, support for crime-prevention measures changes with gender and to some extent with completing high school, but the gender gap itself does not depend on a person’s education level.

2.4 Model with Added Variable

```r
crime$Income <- crime$V87

options(contrasts = c("contr.treatment","contr.poly"))

j <- function(x) jitter(x, factor = 1)

panel_jitter <- function(x, y, ...) {
  xi <- j(x); yi <- j(y)
  points(xi, yi, pch = 16, col = rgb(0, 0.5, 1, 0.4), ...)
  ab <- coef(lm(y ~ x))
  abline(ab, col = 2, lwd = 2)
}

pairs(
  crime[c("Income","V10","V12","V16","V23")],
  panel = panel_jitter,
  main = "Income vs. Attitudes (jittered, with fit lines)",
  labels = c("Income","V10","V12","V16","V23")
)
```



Because of the categorical nature of the data, it is hard to tell if the variables are linearly associated.

```
crime_mod2 <- lm(
  cbind(V10, V12, V16, V23) ~ Gender * Education + Income, data = crime
)

summary(crime_mod2, test = "Pillai")
```

```
## Response V10 :
##
## Call:
## lm(formula = V10 ~ Gender * Education + Income, data = crime)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6623 -0.4831  0.2068  0.8031  2.4239
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     5.58615    0.63245   8.832   <2e-16 ***
## GenderMale                     -0.51728    0.72221  -0.716   0.4742
## EducationsomeHS                -0.04962    0.77375  -0.064   0.9489
## EducationHSGrad                -1.10276    0.65439  -1.685   0.0926 .
## EducationCollege1yr            -0.53563    0.67204  -0.797   0.4258
## EducationCollege2yr            -0.44732    0.71090  -0.629   0.5295
## EducationCollege3yr            -1.01310    0.78157  -1.296   0.1955
## EducationCollegeGrad           -0.47306    0.69675  -0.679   0.4975
## EducationGradDegree            -0.28791    0.71830  -0.401   0.6887
## Income                         -0.08615    0.04232  -2.036   0.0423 *
## GenderMale:EducationsomeHS     -0.18467    0.88021  -0.210   0.8339
## GenderMale:EducationHSGrad      0.65503    0.74851   0.875   0.3819
## GenderMale:EducationCollege1yr  0.07535    0.77755   0.097   0.9228
## GenderMale:EducationCollege2yr -0.28124    0.82296  -0.342   0.7327
## GenderMale:EducationCollege3yr -0.04892    0.90203  -0.054   0.9568
## GenderMale:EducationCollegeGrad 0.31801    0.79013   0.402   0.6875
## GenderMale:EducationGradDegree -0.06720    0.80645  -0.083   0.9336
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.262 on 503 degrees of freedom
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.06814,    Adjusted R-squared:  0.0385
## F-statistic: 2.299 on 16 and 503 DF,  p-value: 0.002905
##
##
## Response V12 :
##
## Call:
## lm(formula = V12 ~ Gender * Education + Income, data = crime)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.0888 -1.5092  0.3668  1.1877  2.6241
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.33356    0.81728   5.302 1.71e-07 ***
## GenderMale                      -0.03831    0.93327  -0.041    0.967
## EducationsomeHS                 -0.05189    0.99986  -0.052    0.959
## EducationHSGrad                 -0.10345    0.84562  -0.122    0.903
## EducationCollege1yr              0.25543    0.86843   0.294    0.769
## EducationCollege2yr              0.92234    0.91865   1.004    0.316
## EducationCollege3yr             -0.39521    1.00997  -0.391    0.696
## EducationCollegeGrad             0.29422    0.90037   0.327    0.744
## EducationGradDegree              1.14134    0.92821   1.230    0.219
## Income                          -0.08356    0.05468  -1.528    0.127
## GenderMale:EducationsomeHS      -0.44970    1.13744  -0.395    0.693
## GenderMale:EducationHSGrad      -0.08884    0.96726  -0.092    0.927
## GenderMale:EducationCollege1yr  -0.12948    1.00478  -0.129    0.898
## GenderMale:EducationCollege2yr  -1.24358    1.06345  -1.169    0.243
## GenderMale:EducationCollege3yr   0.21365    1.16563   0.183    0.855
## GenderMale:EducationCollegeGrad  0.01932    1.02104   0.019    0.985
## GenderMale:EducationGradDegree  -1.11314    1.04212  -1.068    0.286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.631 on 503 degrees of freedom
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.04708,    Adjusted R-squared:  0.01676
## F-statistic: 1.553 on 16 and 503 DF,  p-value: 0.07739
##
##
## Response V16 :
##
## Call:
## lm(formula = V16 ~ Gender * Education + Income, data = crime)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8759 -1.3336  0.3349  1.1064  2.9746
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      5.64387    0.73304   7.699 7.31e-14 ***
## GenderMale                      -1.55425    0.83707  -1.857  0.06393 .
## EducationsomeHS                 -0.62412    0.89680  -0.696  0.48679
## EducationHSGrad                 -1.17076    0.75846  -1.544  0.12331
## EducationCollege1yr             -0.95283    0.77891  -1.223  0.22180
## EducationCollege2yr             -0.82828    0.82396  -1.005  0.31526
## EducationCollege3yr             -2.22937    0.90587  -2.461  0.01419 *
## EducationCollegeGrad            -1.56548    0.80756  -1.939  0.05312 .
```

```
## EducationGradDegree             -1.47893    0.83254  -1.776  0.07627 .
## Income                          -0.14387    0.04905  -2.933  0.00351 **
## GenderMale:EducationsomeHS       0.96065    1.02020   0.942  0.34684
## GenderMale:EducationHSGrad       1.32174    0.86756   1.524  0.12826
## GenderMale:EducationCollege1yr   1.06378    0.90121   1.180  0.23840
## GenderMale:EducationCollege2yr   0.81875    0.95384   0.858  0.39109
## GenderMale:EducationCollege3yr   1.74060    1.04549   1.665  0.09656 .
## GenderMale:EducationCollegeGrad  1.67262    0.91580   1.826  0.06838 .
## GenderMale:EducationGradDegree   1.73128    0.93471   1.852  0.06458 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.463 on 503 degrees of freedom
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.07387,    Adjusted R-squared:  0.04442
## F-statistic: 2.508 on 16 and 503 DF,  p-value: 0.00104
##
##
## Response V23 :
##
## Call:
## lm(formula = V23 ~ Gender * Education + Income, data = crime)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7173 -0.5513  0.1075  0.8161  2.0245
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.56689    0.61075   7.477 3.39e-13 ***
## GenderMale                       0.39246    0.69743   0.563    0.574
## EducationsomeHS                  0.05853    0.74720   0.078    0.938
## EducationHSGrad                  0.21732    0.63193   0.344    0.731
## EducationCollege1yr              0.49230    0.64898   0.759    0.448
## EducationCollege2yr              0.20046    0.68651   0.292    0.770
## EducationCollege3yr              0.18395    0.75475   0.244    0.808
## EducationCollegeGrad             0.62120    0.67284   0.923    0.356
## EducationGradDegree              0.40072    0.69365   0.578    0.564
## Income                          -0.06689    0.04086  -1.637    0.102
## GenderMale:EducationsomeHS      -0.57595    0.85001  -0.678    0.498
## GenderMale:EducationHSGrad      -0.65830    0.72283  -0.911    0.363
## GenderMale:EducationCollege1yr  -0.84786    0.75087  -1.129    0.259
## GenderMale:EducationCollege2yr  -0.78045    0.79472  -0.982    0.327
## GenderMale:EducationCollege3yr  -0.80511    0.87108  -0.924    0.356
## GenderMale:EducationCollegeGrad -1.20366    0.76302  -1.577    0.115
## GenderMale:EducationGradDegree  -0.61052    0.77878  -0.784    0.433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.219 on 503 degrees of freedom
```

```
##   (39 observations deleted due to missingness)
## Multiple R-squared:  0.04712,    Adjusted R-squared:  0.01681
## F-statistic: 1.555 on 16 and 503 DF,  p-value: 0.07688
```

```
summary.aov(crime_mod2)
```

```
##  Response V10 :
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## Gender             1   9.88  9.8843  6.2055 0.01306 *
## Education          7  28.76  4.1088  2.5795 0.01277 *
## Income             1   6.71  6.7061  4.2102 0.04070 *
## Gender:Education   7  13.23  1.8904  1.1868 0.30851
## Residuals        503 801.19  1.5928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V12 :
##                   Df  Sum Sq Mean Sq F value  Pr(>F)
## Gender             1   17.21 17.2053  6.4686 0.01128 *
## Education          7   20.11  2.8735  1.0803 0.37458
## Income             1    5.82  5.8172  2.1871 0.13980
## Gender:Education   7   22.96  3.2796  1.2330 0.28265
## Residuals        503 1337.89  2.6598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V16 :
##                   Df  Sum Sq Mean Sq F value   Pr(>F)
## Gender             1   16.75 16.7497  7.8278 0.005342 **
## Education          7   36.37  5.1961  2.4283 0.018731 *
## Income             1   17.94 17.9351  8.3818 0.003955 **
## Gender:Education   7   14.80  2.1138  0.9879 0.439156
## Residuals        503 1076.30  2.1398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response V23 :
##                   Df Sum Sq Mean Sq F value    Pr(>F)
## Gender             1  19.49 19.4926 13.1228 0.0003213 ***
## Education          7   7.15  1.0219  0.6879 0.6823514
## Income             1   4.45  4.4515  2.9968 0.0840399 .
## Gender:Education   7   5.85  0.8363  0.5630 0.7860845
## Residuals        503 747.15  1.4854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## 39 observations deleted due to missingness
```

For V10 and V16, we see significant negative association with income. Effects from Education and Gender still remain significant, but interaction term is no longer significant.

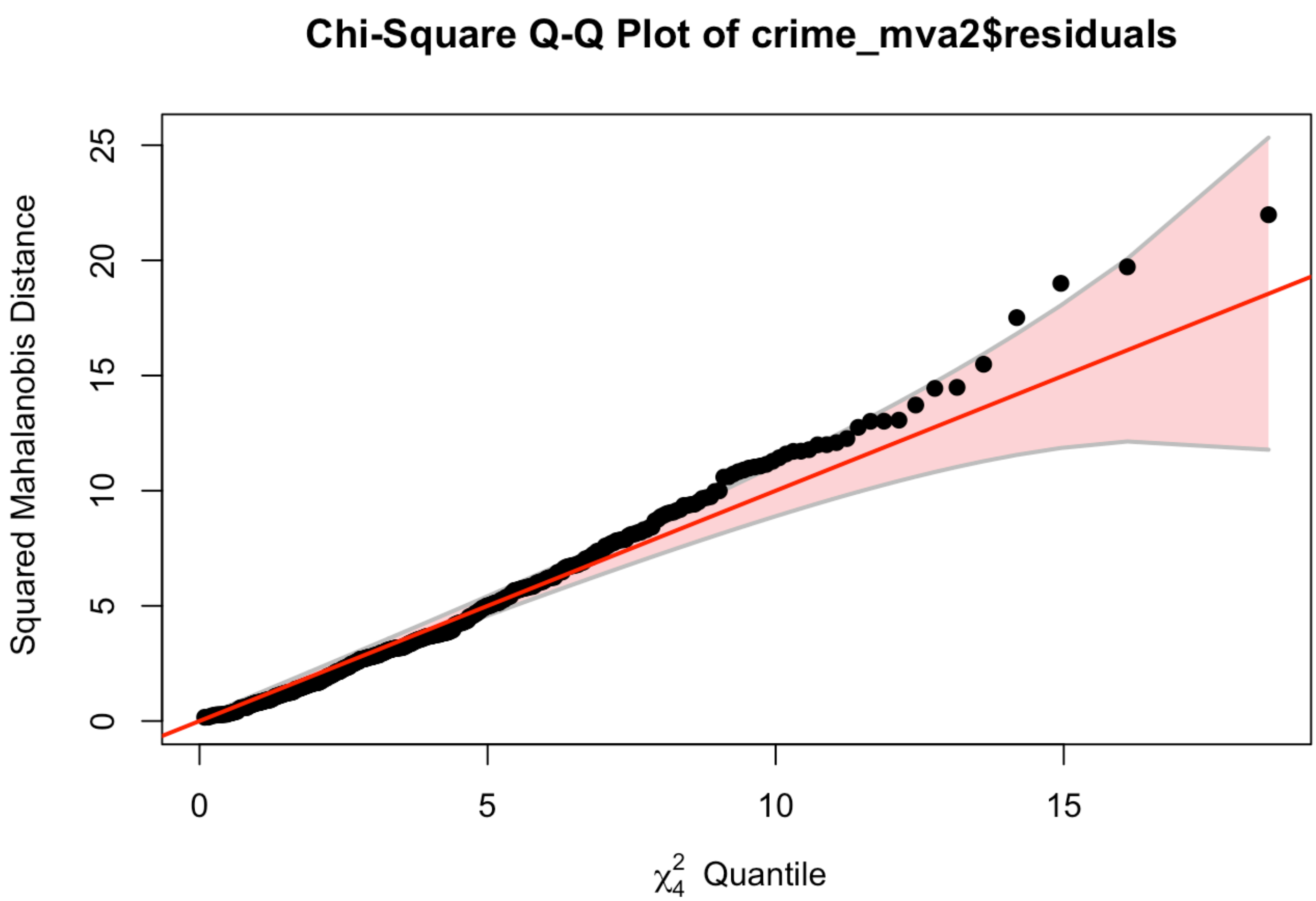
V12 is very stable and no predictors are statistically significant.

V23 shows significant difference in responses across genders.

Overall, adding the income term doesn't change the overall pattern of significance in the multivariate tests, but it explains additional variation, especially in V10 and V16.

2.5 Model Assumptions

```
crime_mva2 <- manova(
  cbind(V10, V12, V16, V23) ~ Gender * Education + Income, data = crime)

cqplot(crime_mva2$residuals, label = "MANOVA residuals")
```



Model assumptions are well-met.

2.6 p-value Adjustment

```r
mod_gender <- manova(
  cbind(V10, V12, V16, V23) ~ Gender,
  data = crime
)

summary(mod_gender, test = "Pillai")
```

```
##             Df   Pillai approx F num Df den Df   Pr(>F)
## Gender       1 0.030128   4.2014      4    541 0.002332 **
## Residuals 544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
aov_gender <- summary.aov(mod_gender)

p_gender <- sapply(
  aov_gender,
  function(x) x["Gender", "Pr(>F)"]
)

p_adj_gender <- p.adjust(p_gender, method = "holm")

gender_contrasts <- data.frame(
  P_value    = p_gender,
  Holm_Adj_P = p_adj_gender
)
print(gender_contrasts)
```

```
##                    P_value  Holm_Adj_P
##  Response V10 0.011028302 0.015269863
##  Response V12 0.005089954 0.015269863
##  Response V16 0.005303752 0.015269863
##  Response V23 0.001343217 0.005372868
```

Even after adjusting for multiple comparisons, all four items show a statistically significant gender difference. In each case, the mean support among women differs from men, so gender is a robust predictor. The smallest adjusted p is for V23 (“FamilyHelp”), indicating the strongest gender gap there.