Advanced graphics and data analysis

Research Methods in Psychology

Department of Psychology
Northwestern University
Evanston, Illinois USA

November, 2011

Advanced graphics and data analysis

Creating the data

1. A 1000 cases were simulated using our simulation experiment.
    - This was not a realistic number for a normal experiment but is done to show analytical techniques.
    - All Independent Variables and Subject Variables were set to random values.
2. By allowing everything to vary this way, we are not systematically controlling for various artifacts such as time of term.
    - Time of term is probably related to various state and trait variables.
3. Once simulated, the data were copied into the clipboard and read using read.clipboard()

Describing the data for 1000 simulated subjects

1. Every time you start R you need to make the psych package active using the `library` command.
2. Then read in the data (typically by copying it to the clipboard and then reading the clipboard).
3. Then do some descriptive statistics.

```
> library(psych)    #only necessary the first time you are starting R up.
> sim.data <- read.clipboard()
> describe(sim.data, skew=FALSE)    #don't bother to show skew and kurtosis
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	se
snum	1	1000	500.50	288.82	500.5	500.50	370.65	1	1000	999	9.13
sex	2	1000	1.50	0.50	1.5	1.50	0.74	1	2	1	0.02
drug	3	1000	0.53	0.50	1.0	0.54	0.00	0	1	1	0.02
time	4	1000	14.76	4.30	15.0	14.71	5.93	8	22	14	0.14
anxiety	5	1000	4.07	2.02	4.0	4.07	1.48	−2	11	13	0.06
impulsivity	6	1000	5.22	2.03	5.0	5.22	1.48	0	12	12	0.06
arousal	7	1000	60.29	42.60	60.0	60.04	45.22	−57	195	252	1.35
tension	8	1000	24.42	13.64	22.0	23.26	13.34	3	74	71	0.43
performance	9	1000	63.24	11.19	64.0	63.69	11.86	23	87	64	0.35
cost	10	1000	1.00	0.00	1.0	1.00	0.00	1	1	0	0.00

Find the correlations

As part of the descriptions, see what goes with what. Find the correlations using `cor`. If there are missing data specify use="pairwise" or use the `corr.test` function.

```
> round(cor(sim.data[2:9]),2)  #round off the correlations to 2 decimals
```

	sex	drug	**time**	anxiety	impulsivity	arousal	tension	performance
sex	1.00	−0.03	−0.06	0.03	−0.03	−0.06	0.00	−0.06
drug	−0.03	1.00	0.01	−0.02	0.01	0.39	0.60	0.33
time	−0.06	0.01	1.00	−0.03	0.02	0.36	−0.02	0.31
anxiety	0.03	−0.02	−0.03	1.00	−0.03	0.00	0.72	−0.01
impulsivity	−0.03	0.01	0.02	−0.03	1.00	−0.11	−0.01	−0.12
arousal	−0.06	0.39	0.36	0.00	−0.11	1.00	0.23	0.81
tension	0.00	0.60	−0.02	0.72	−0.01	0.23	1.00	0.18
performance	−0.06	0.33	0.31	−0.01	−0.12	0.81	0.18	1.00

Correlations may be tested for significance by comparing them to the confidence interval of a correlation of 0 for this many participants. For most reasonable sample sizes, this will be $\frac{1.96}{\sqrt{N-2}}$. Thus, for 1000 participants, anything greater than (absolute value) .062 has less than 1/20 chance of happening by chance.

Test for significance of correlations. Use both the Fisher test (for single correlations) and the Holm test for multiple comparisons.

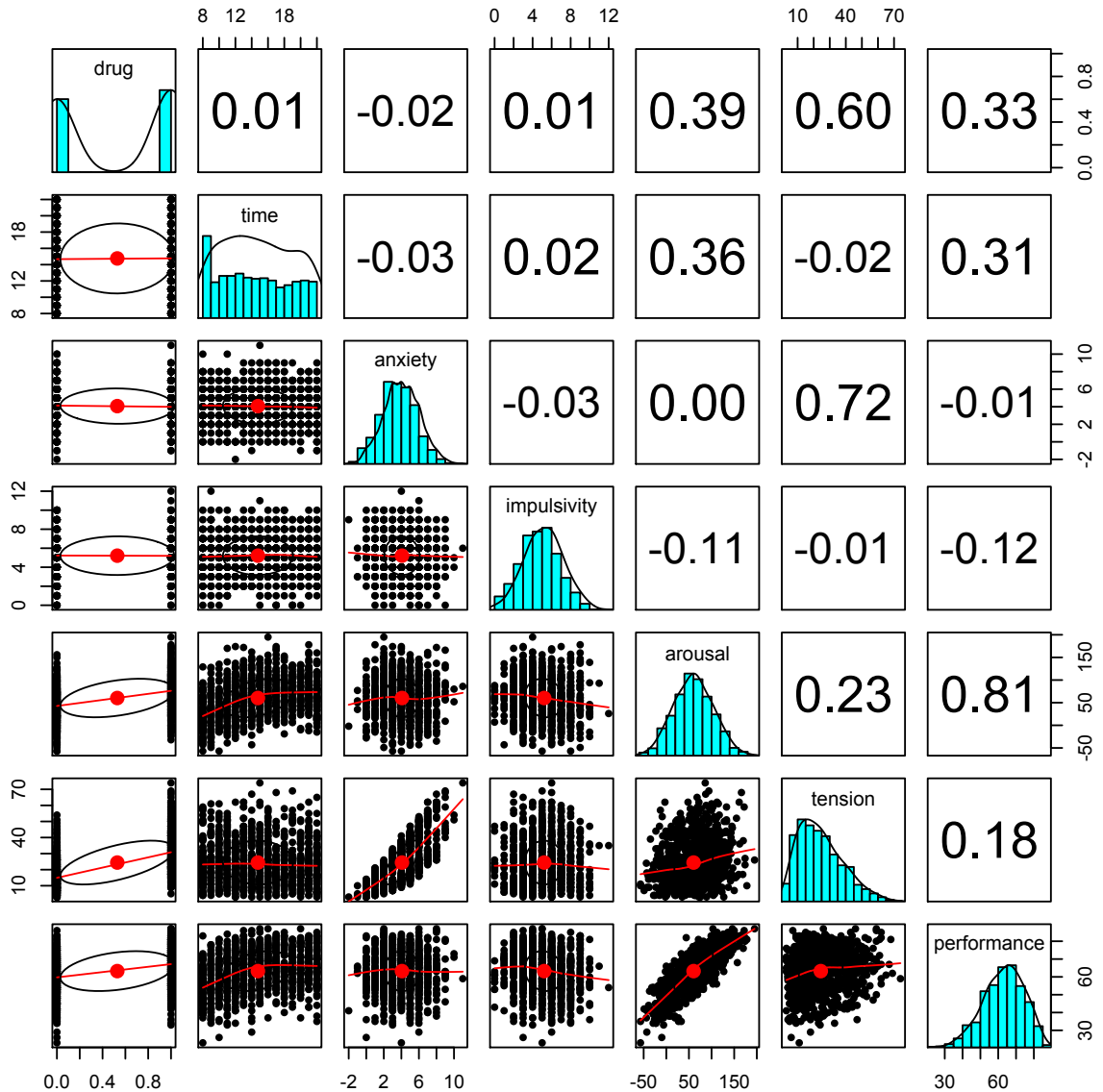
`corr.test` finds correlations using the "pairwise" option, counts the number of cases per pair, and then tests the correlations for significance. Two tests are applied. The first is just the normal *Fisher* test for significance. The second adjusts for the number of correlations being tested by using the *Holm* adjustment.

```
> corr.test(sim.data[2:9])

Call:corr.test(x = sim.data[2:9])
Correlation matrix
            sex  drug  time anxiety impulsivity arousal tension performance
sex        1.00 -0.03 -0.06    0.03       -0.03   -0.06    0.00       -0.06
drug      -0.03  1.00  0.01   -0.02        0.01    0.39    0.60        0.33
time      -0.06  0.01  1.00   -0.03        0.02    0.36   -0.02        0.31
...
performance -0.06  0.33  0.31   -0.01       -0.12    0.81    0.18        1.00

Probability values (Entries above the diagonal are adjusted for multiple tests)
            sex drug time anxiety impulsivity arousal tension performance
sex        0.00 1.00 0.68    1.00         1.0    0.68       1         0.9
drug       0.28 0.00 1.00    1.00         1.0    0.00       0         0.0
time       0.04 0.65 0.00    1.00         1.0    0.00       1         0.0
anxiety    0.27 0.45 0.39    0.00         1.0    1.00       0         1.0
impulsivity 0.34 0.74 0.61    0.37         0.0    0.01       1         0.0
arousal    0.04 0.00 0.00    0.90         0.0    0.00       0         0.0
tension    0.89 0.00 0.49    0.00         0.8    0.00       0         0.0
performance 0.06 0.00 0.00    0.66         0.0    0.00       0         0.0
```

Show the correlations graphically using `pairs.panels`

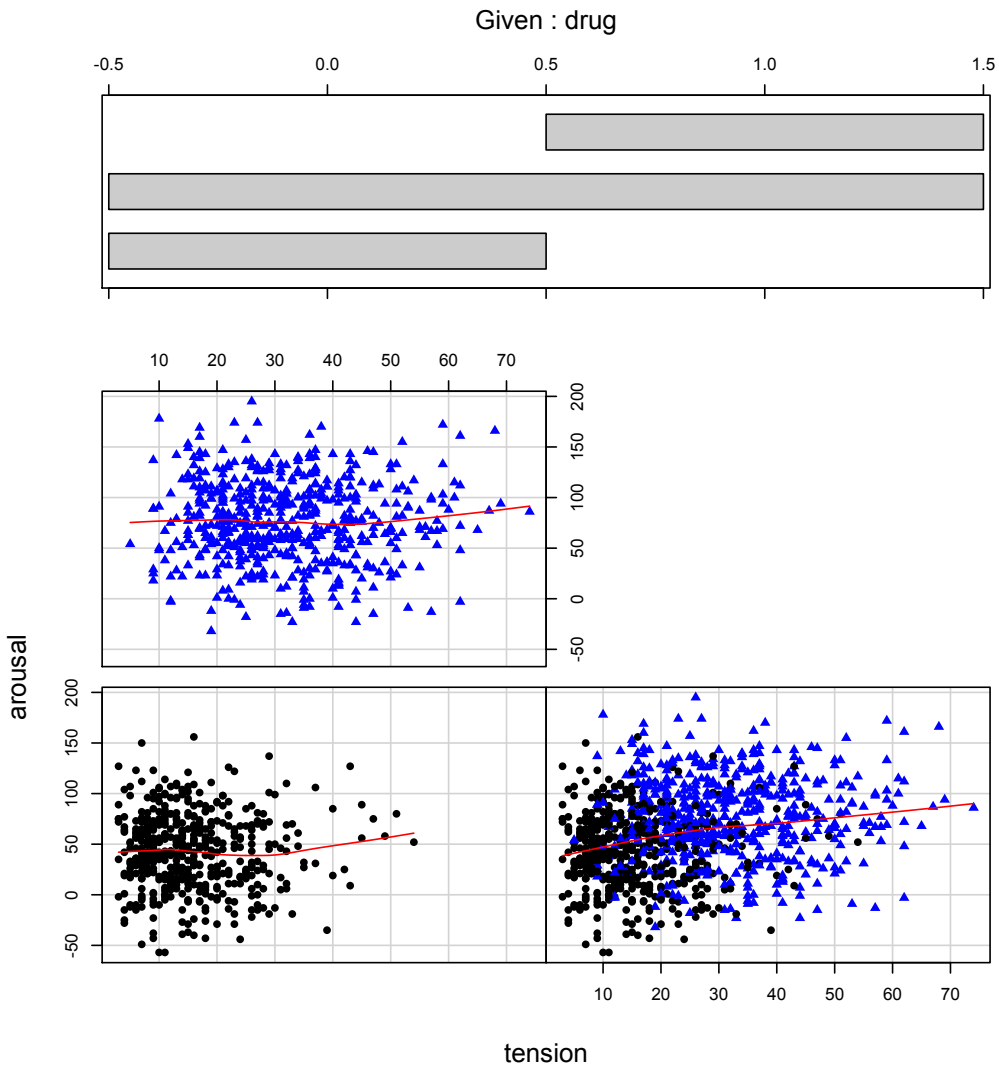


Using graphics to explore relationships

1. The basic graphic techniques we have discussed show bivariate plots with locally smoothed (loess) regressions.
2. More complicated graphics are possible by doing multi-panel graphs.
   - We did this for the analysis in terms of showing regressions depending upon third variables.
3. Also possible to explore this by doing "coplots"
   - These plot multipanel graphs showing bivariate relationships in each graph.
4. Consider the example of tension and arousal, do the covary just because of caffeine?

A coplot of arousal x tension for two drug levels

```
coplot(arousal ~ tension | drug, data=sim.data, panel = panel.smooth, pch=16+sim.data$drug,
col=c("black","blue")[sim.data$drug+1])
```



We may also do this with the by command as well as linear modeling

```
with(sim.data,cor(tension,arousal))

[1] 0.2310705

> by(sim.data,sim.data$drug,function(x)
                cor(x$tension,x$arousal))

sim.data$drug: 0
[1] 0.000401661
------------------------------------
sim.data$drug: 1
[1] -0.004094199
```

1. First find the zero order correlation
2. Then find it within each of two conditions
3. The correlation has vanished!

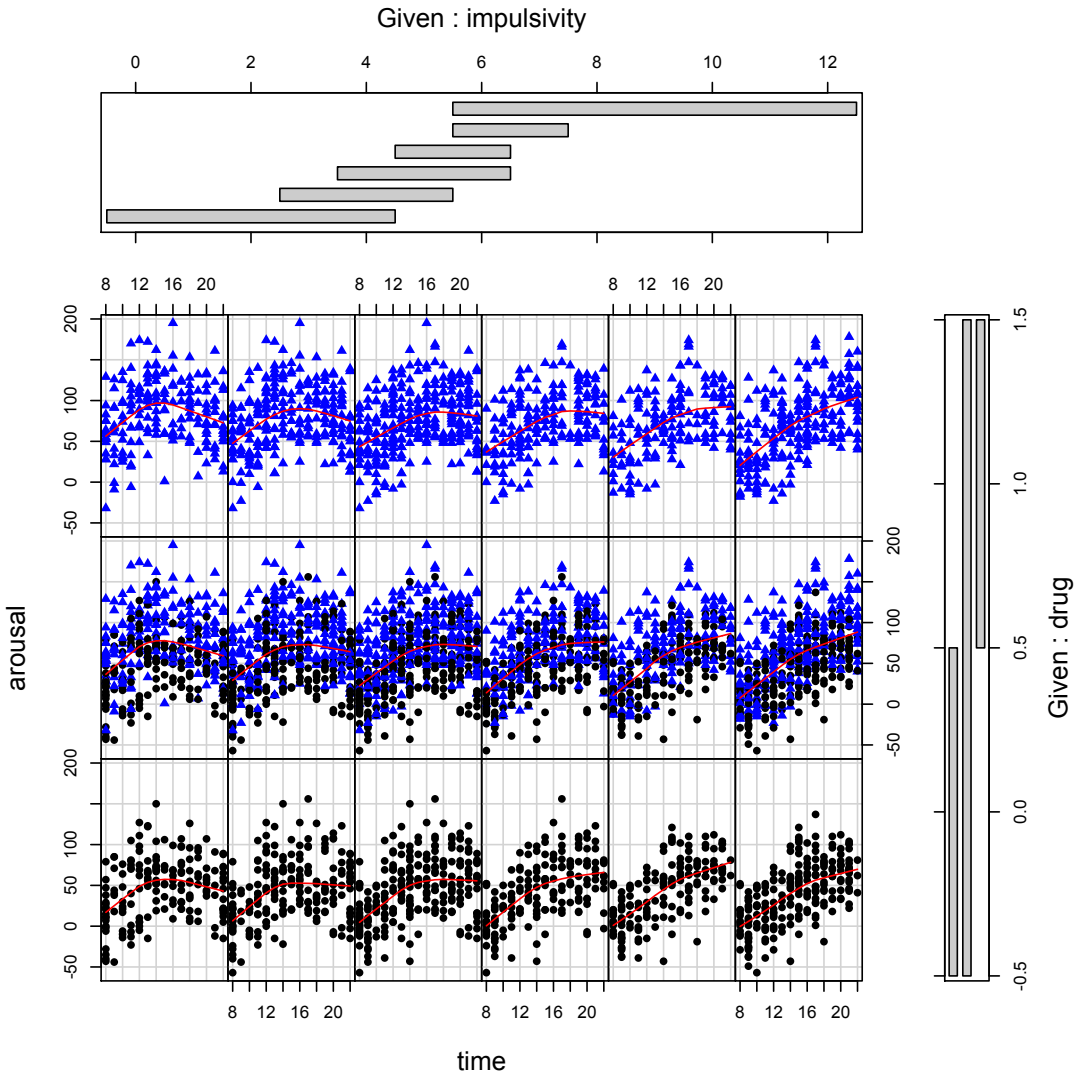
That the arousal - tension effect is mediated by the caffeine effect is also shown in a linear model.

```
> lm(tension ~ arousal + drug,data=sim.data)

Call:
lm(formula = tension ~ arousal + drug, data = sim.data)
Coefficients:
(Intercept)      arousal         drug
 15.8209666   -0.0006802   16.3271145
```

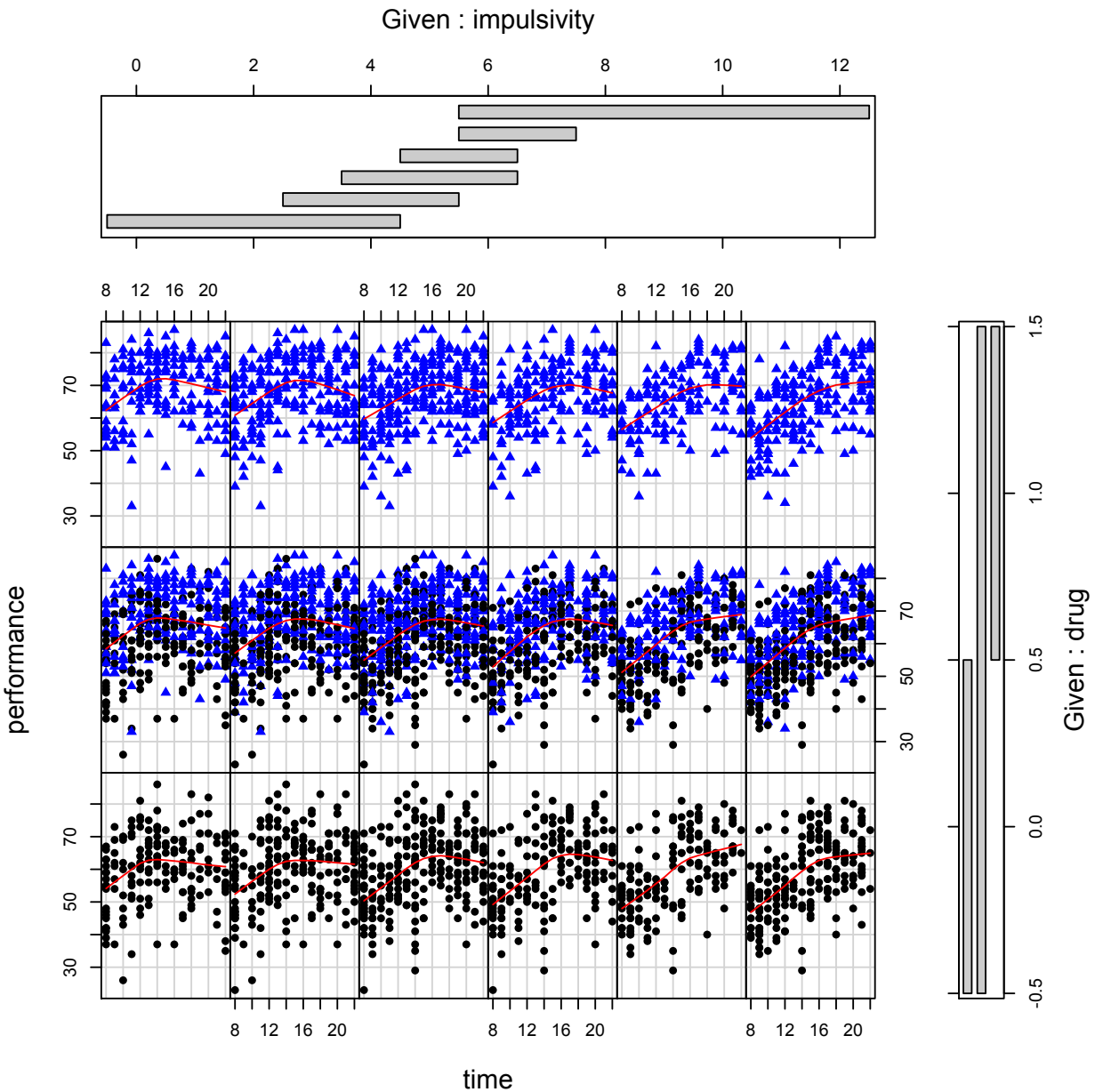
The effect of impulsivity and time and drug on arousal

```
coplot(arousal ~ time | impulsivity+ drug,data=sim.data,panel =
panel.smooth,pch=16+sim.data$drug, col = c("black","blue")[sim.data$drug+1])
```



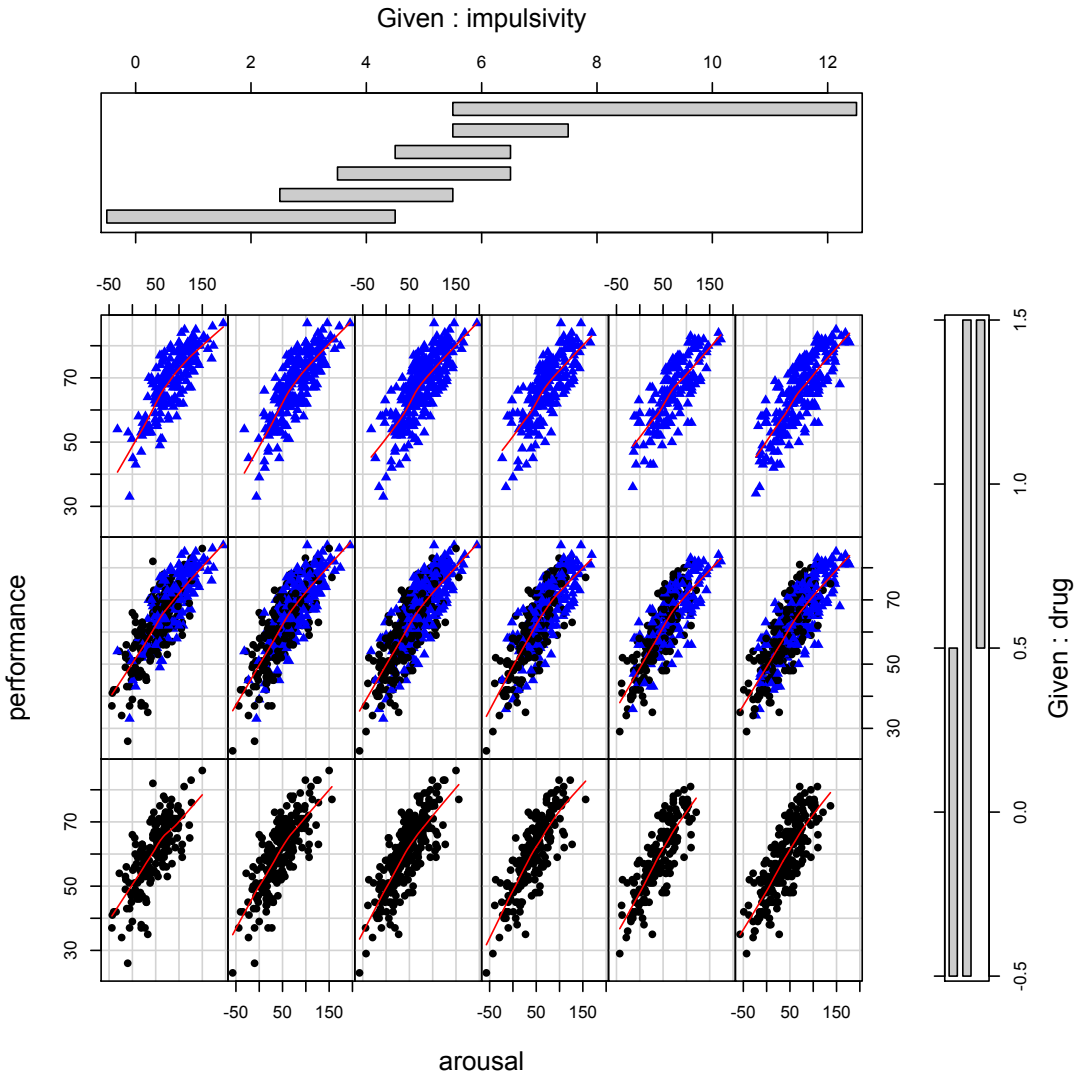
1. Clear effect of caffeine on arousal (the blues are higher than the blacks).
2. Subtle time of day effect showing interaction with impulsivity.
   - Note the shape of the time x arousal line changes by impulsivity level.

The effect of impulsivity and time and drug on performance



1. Clear effect of caffeine on performance (the blues are higher than the blacks).
2. Subtle time of day effect showing interaction with impulsivity.
   - Note the shape of the time x performance line changes by impulsivity level.

The effect of arousal on performance examining drug and impulsivity



1. The interactions with drug and impulsivity have vanished.
2. Arousal level drives the performance effect
   - Note the shape of the of the arousal x performance line does not change by drug or impulsivity.

Mediation versus moderation

1. Moderation is an interaction.
   - The effect of impulsivity *interacts* with time of day.
   - The effect of impulsivity is *moderated* by the time of day.
2. Mediation specifies an intervening causal influence
   - The effects of caffeine and impulsivity on performance are *mediated* by arousal.
   - Caffeine and impulsivity affect arousal.
   - Arousal affects performance.
   - Holding arousal constant makes the effects of caffeine or time of day vanish.

Testing for mediation: part 1 – Mediator not included

```
> mod1 <- lm(performance ~ drug * impulsivity * time, data = cen.data)
> summary(mod1)

Call:
lm(formula = performance ~ drug * impulsivity * time, data = cen.data)

Residuals:
     Min       1Q   Median       3Q      Max
-33.278   -6.516    0.521    7.147   26.514

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            -0.01915    0.30621  -0.063    0.950
drug                    7.31855    0.61348  11.930  < 2e-16 ***
impulsivity            -0.73627    0.15096  -4.877 1.25e-06 ***
time                    0.82135    0.07131  11.517  < 2e-16 ***
drug:impulsivity       -0.38704    0.30123  -1.285    0.199
drug:time              -0.17223    0.14285  -1.206    0.228
impulsivity:time        0.24044    0.03497   6.876 1.09e-11 ***
drug:impulsivity:time   0.08125    0.06977   1.164    0.245
---
Signif. codes:  0 Ô***Õ 0.001 Ô**Õ 0.01 Ô*Õ 0.05 Ô.Õ 0.1 Ô Õ 1

Residual standard error: 9.68 on 992 degrees of freedom
Multiple R-squared: 0.2575, Adjusted R-squared: 0.2523
F-statistic: 49.15 on 7 and 992 DF,  p-value: < 2.2e-16
```

Testing for mediation: part 2 – mediator included

```
> mod2 <- lm(performance ~ drug * impulsivity * time + arousal, data = cen.data)
> summary(mod2)

Call:
lm(formula = performance ~ drug * impulsivity * time + arousal,
    data = cen.data)

Residuals:
     Min       1Q   Median       3Q      Max
-22.6573  -4.3852  -0.0704   4.6756  22.4502

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -0.003139   0.207105  -0.015   0.9879
drug                     0.518495   0.459808   1.128   0.2597
impulsivity             -0.183312   0.103366  -1.773   0.0765 .
time                     0.099036   0.052625   1.882   0.0601 .
arousal                  0.205363   0.005984  34.316   <2e-16 ***
drug:impulsivity        -0.219262   0.203794  -1.076   0.2822
drug:time               -0.070183   0.096659  -0.726   0.4680
impulsivity:time         0.027301   0.024451   1.117   0.2644
drug:impulsivity:time   -0.060582   0.047370  -1.279   0.2012
---
Signif. codes:  0 Ô***Õ 0.001 Ô**Õ 0.01 Ô*Õ 0.05 Ô.Õ 0.1 Ô Õ 1

Residual standard error: 6.547 on 991 degrees of freedom
Multiple R-squared: 0.6607, Adjusted R-squared: 0.658
F-statistic: 241.2 on 8 and 991 DF,  p-value: < 2.2e-16
```