Silly interaction example

Using data about condiments, and temperature and pressure

James Twose

```
library(sjPlot)
library(sjmisc)
library(ggplot2)
library(interactions)
```

```
## Example 1 - categorical
# Read in categorical data
int_cat_df <- read.csv("Interactions_Categorical.csv")

# Show summary and head of the data set
summary(int_cat_df); str(int_cat_df); head(int_cat_df)
```

```
##    Enjoyment             Food              Condiment
##  Min.   : 52.31   Hot Dog  :40   Chocolate Sauce:40
##  1st Qu.: 62.33   Ice Cream:40   Mustard        :40
##  Median : 78.00
##  Mean   : 77.32
##  3rd Qu.: 90.92
##  Max.   :102.62
```

```
## 'data.frame':    80 obs. of  3 variables:
##  $ Enjoyment: num  81.9 84.9 90.3 89.6 97.7 ...
##  $ Food     : Factor w/ 2 levels "Hot Dog","Ice Cream": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Condiment: Factor w/ 2 levels "Chocolate Sauce",..: 2 2 2 2 2 2 2 2 2 2 ...
```

```
##   Enjoyment    Food Condiment
## 1  81.92696 Hot Dog   Mustard
## 2  84.93977 Hot Dog   Mustard
## 3  90.28648 Hot Dog   Mustard
## 4  89.56180 Hot Dog   Mustard
## 5  97.67683 Hot Dog   Mustard
## 6  83.61713 Hot Dog   Mustard
```

```
# run a linear regression with the interaction and enjoyment outcome
int_cat_lm <- lm(Enjoyment ~ Food*Condiment, data=int_cat_df)

# show a summary of the lm output (coefficients, t-values, p-values etc.)
summary(int_cat_lm)
```

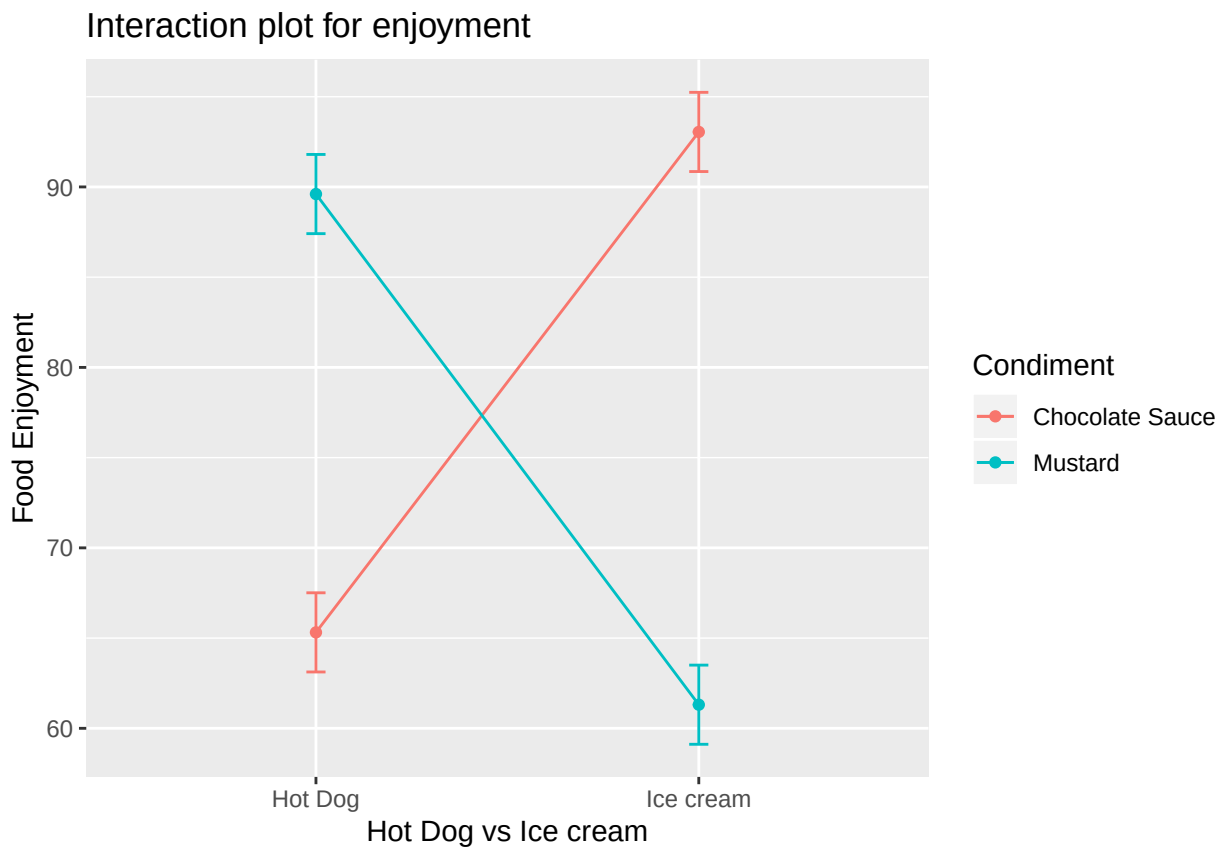
```
##
## Call:
## lm(formula = Enjoyment ~ Food * Condiment, data = int_cat_df)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -9.068 -3.068 -0.407  2.802 13.015
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     65.317      1.120   58.34   <2e-16 ***
## FoodIce Cream                   27.731      1.583   17.52   <2e-16 ***
## CondimentMustard                24.289      1.583   15.34   <2e-16 ***
## FoodIce Cream:CondimentMustard -56.028      2.239  -25.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.007 on 76 degrees of freedom
## Multiple R-squared:  0.8935, Adjusted R-squared:  0.8892
## F-statistic: 212.4 on 3 and 76 DF,  p-value: < 2.2e-16
```

```
# show the anova of the lm output (f-values, p-values etc.)
anova(int_cat_lm)
```

```
## Analysis of Variance Table
##
## Response: Enjoyment
##                Df  Sum Sq Mean Sq  F value    Pr(>F)
## Food            1     1.6     1.6   0.0637  0.801362
## Condiment       1   277.5   277.5  11.0711  0.001353 **
## Food:Condiment  1 15695.8 15695.8 626.1534 < 2.2e-16 ***
## Residuals      76  1905.1    25.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

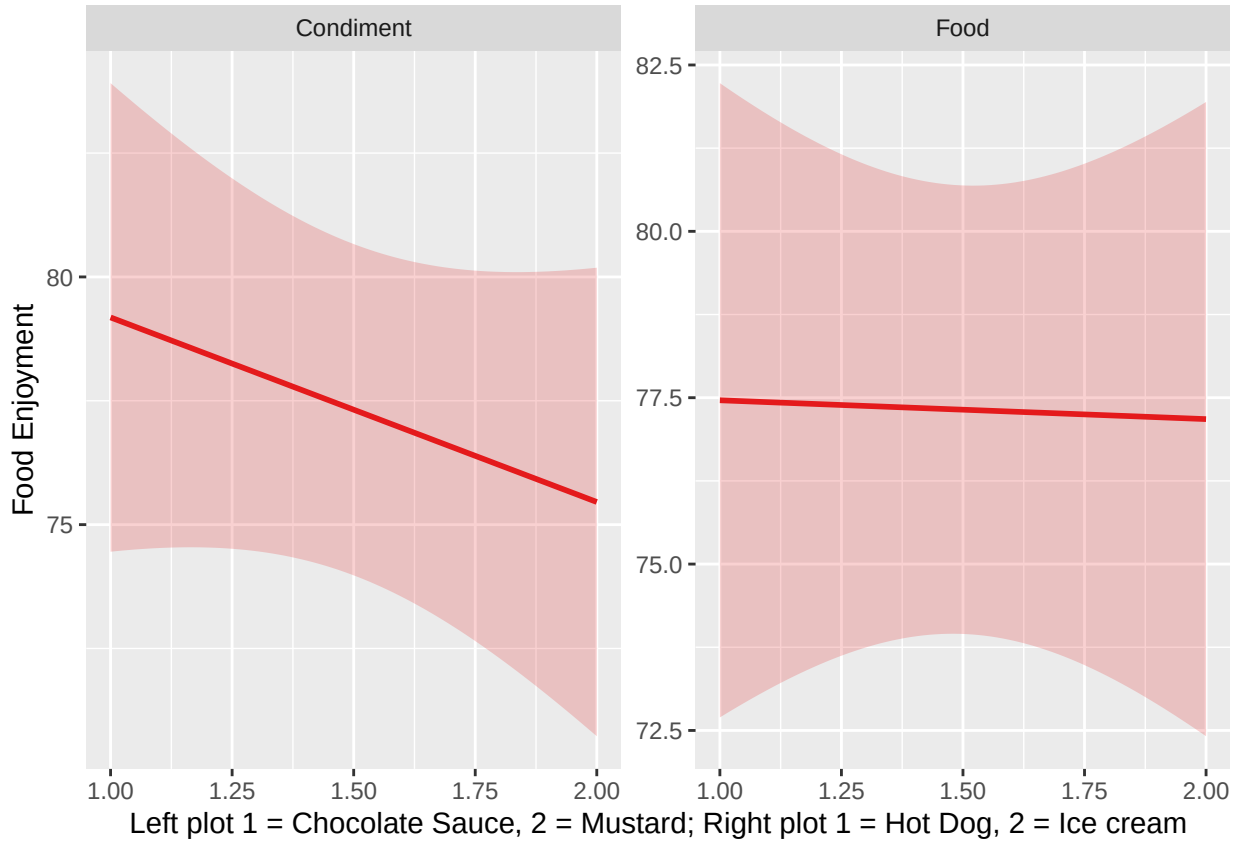
```
# here you can see that when looking at explained variance, the food:condiment interaction
# is considerably larger than either of the main effects of food or condiment.

# Be lazy and use the plot_model function to calc means and confidence intervals
p <- plot_model(int_cat_lm, type = "pred", terms = c("Food", "Condiment"))
p$data$x <- factor(p$data$x, levels = c(1,2), labels = c("Hot Dog", "Ice cream"))

# plot the interaction
ggplot(data=p$data, aes(x=x, y=predicted, group=group, colour=group)) +
  geom_line() +
  geom_point() +
  geom_errorbar(aes(ymin=conf.low, ymax=conf.high), width=.05) +
  labs(x = "Hot Dog vs Ice cream", y = "Food Enjoyment", colour = "Condiment") +
  ggtitle("Interaction plot for enjoyment")
```




```
# plot the main effects to compare
p <- plot_model(int_cat_lm, type = "slope", show.loess=F)
```

```
## Warning: Interaction terms are not supported by this plot type. Output for
## interaction terms may be inappropriate.
```

```
p + labs(x = "Left plot 1 = Chocolate Sauce, 2 = Mustard; Right plot 1 = Hot Dog, 2 = Ice cream", y = "
```




```
# Based on the interaction plot and the main effects plot you can see that you would have missed
# a large amount of explained variance if you had not included the interaction


## Example 2 - continuous
# Read in continuous data
int_con_df <- read.csv("Interactions_Continuous.csv")

# Show summary and head of the data set
summary(int_con_df); str(int_con_df); head(int_con_df)
```

```
##     Strength       Temperature        Pressure          Time
##  Min.   : 67.22   Min.   : 93.24   Min.   :63.68   Min.   :31.08
##  1st Qu.: 88.44   1st Qu.:103.38   1st Qu.:68.28   1st Qu.:32.08
##  Median : 95.52   Median :106.05   Median :71.78   Median :32.90
##  Mean   : 92.46   Mean   :105.29   Mean   :71.06   Mean   :33.40
##  3rd Qu.: 98.44   3rd Qu.:107.31   3rd Qu.:73.00   3rd Qu.:34.74
##  Max.   :103.22   Max.   :113.46   Max.   :81.10   Max.   :38.10

## 'data.frame':    29 obs. of  4 variables:
##  $ Strength   : num  100.7 97.8 88.4 85.4 93.2 ...
##  $ Temperature: num  100.6 109.5 104 99.4 107.2 ...
##  $ Pressure   : num  81.1 72.4 74.6 65 67.4 ...
##  $ Time       : num  33.3 32.9 35.3 35 32.8 ...

##    Strength Temperature Pressure  Time
```

```
## 1 100.66667      100.59    81.10 33.32
## 2  97.77778      109.50    72.38 32.92
## 3  88.44444      103.98    74.62 35.32
## 4  85.44444       99.39    65.04 35.00
## 5  93.18519      107.25    67.42 32.80
## 6  95.51852      103.38    68.28 32.56
```

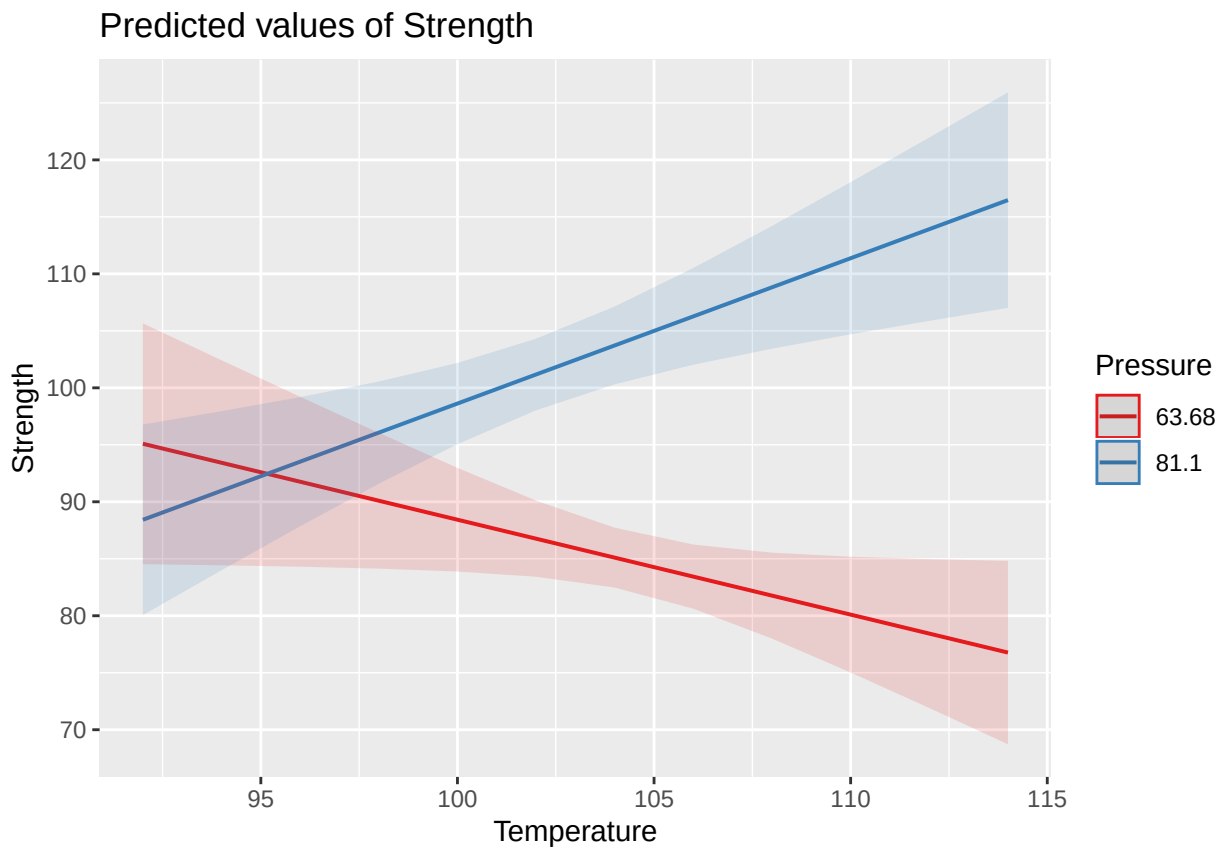
```
# run a linear regression with the interaction and enjoyment outcome
int_con_lm <- lm(Strength ~ Time + Temperature*Pressure, data=int_con_df)

# show a summary of the lm output (coefficients, t-values, p-values etc.)
summary(int_con_lm)
```

```
##
## Call:
## lm(formula = Strength ~ Time + Temperature * Pressure, data = int_con_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.974 -1.855  0.229  1.391  6.235
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1064.46035  330.58328   3.220  0.00366 **
## Time                   -4.77386    0.32576 -14.655  1.8e-13 ***
## Temperature            -8.53821    3.15715  -2.704  0.01238 *
## Pressure              -11.51515    4.48315  -2.569  0.01686 *
## Temperature:Pressure    0.12100    0.04337   2.790  0.01016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.824 on 24 degrees of freedom
## Multiple R-squared:  0.905,  Adjusted R-squared:  0.8891
## F-statistic: 57.13 on 4 and 24 DF,  p-value: 6.446e-12
```

```
# show the anova of the lm output (f-values, p-values etc.)
anova(int_con_lm)
```

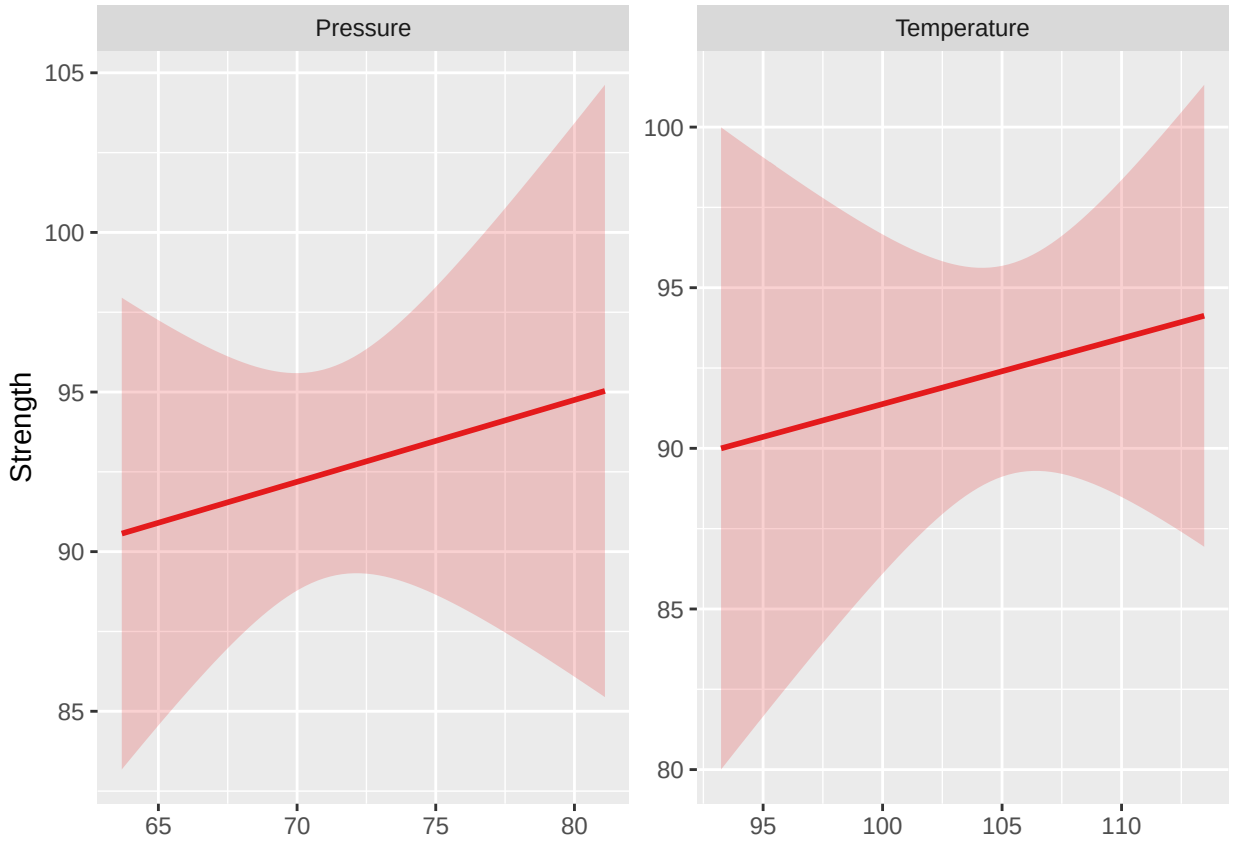
```
## Analysis of Variance Table
##
## Response: Strength
##                      Df  Sum Sq Mean Sq  F value    Pr(>F)
## Time                  1 1451.12 1451.12 181.9376 1.071e-12 ***
## Temperature           1    0.02    0.02   0.0020   0.96482
## Pressure              1  309.34  309.34  38.7838 1.952e-06 ***
## Temperature:Pressure  1   62.08   62.08   7.7839   0.01016 *
## Residuals            24  191.42    7.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Be lazy and use the plot_model function to plot means and confidence intervals
plot_model(int_con_lm, type = "int", terms = c("Temperature", "Pressure"))
```



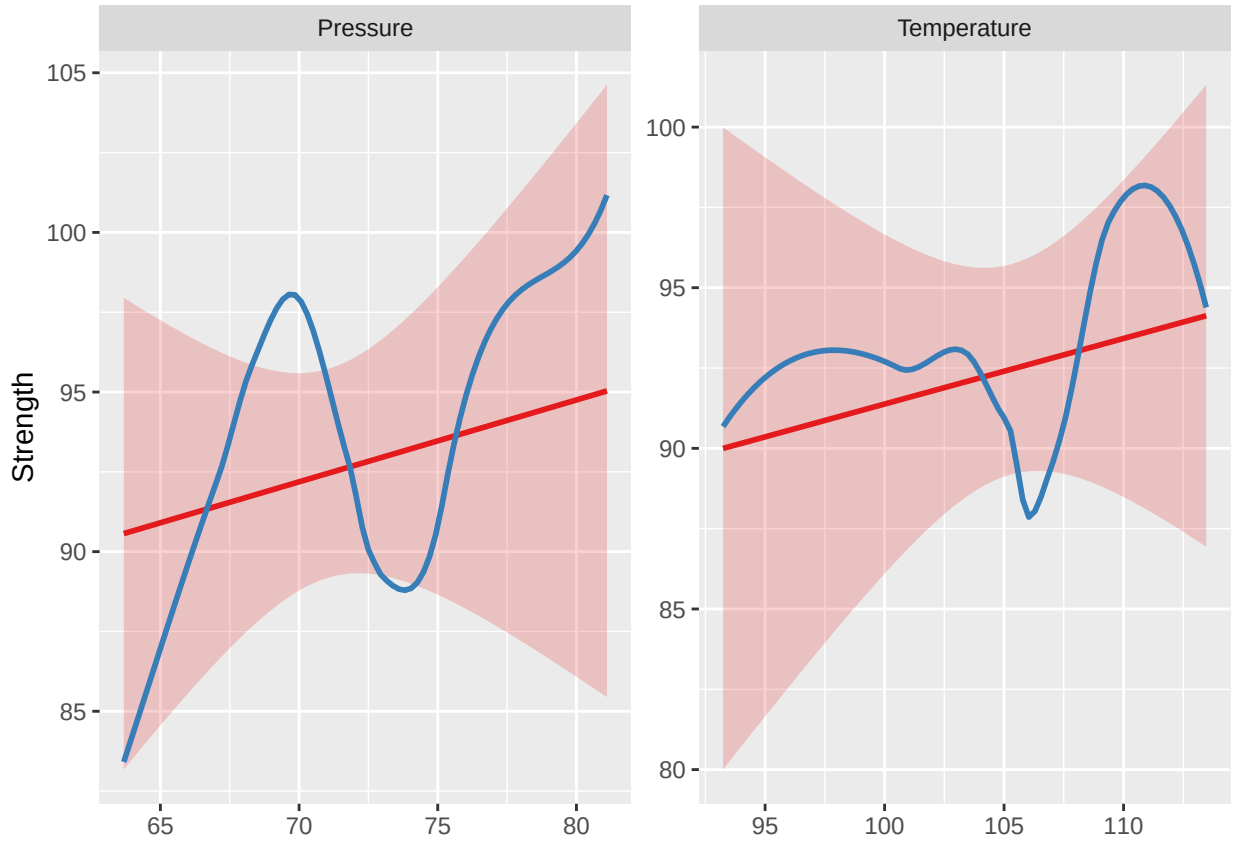

```
# plot the main effects to compare
plot_model(int_con_lm, type = "slope", terms = c("Temperature", "Pressure"), show.loess=F)
```

```
## Warning: Interaction terms are not supported by this plot type. Output for
## interaction terms may be inappropriate.
```




```
plot_model(int_con_lm, type = "slope", terms = c("Temperature", "Pressure"), show.loess=TRUE)
```

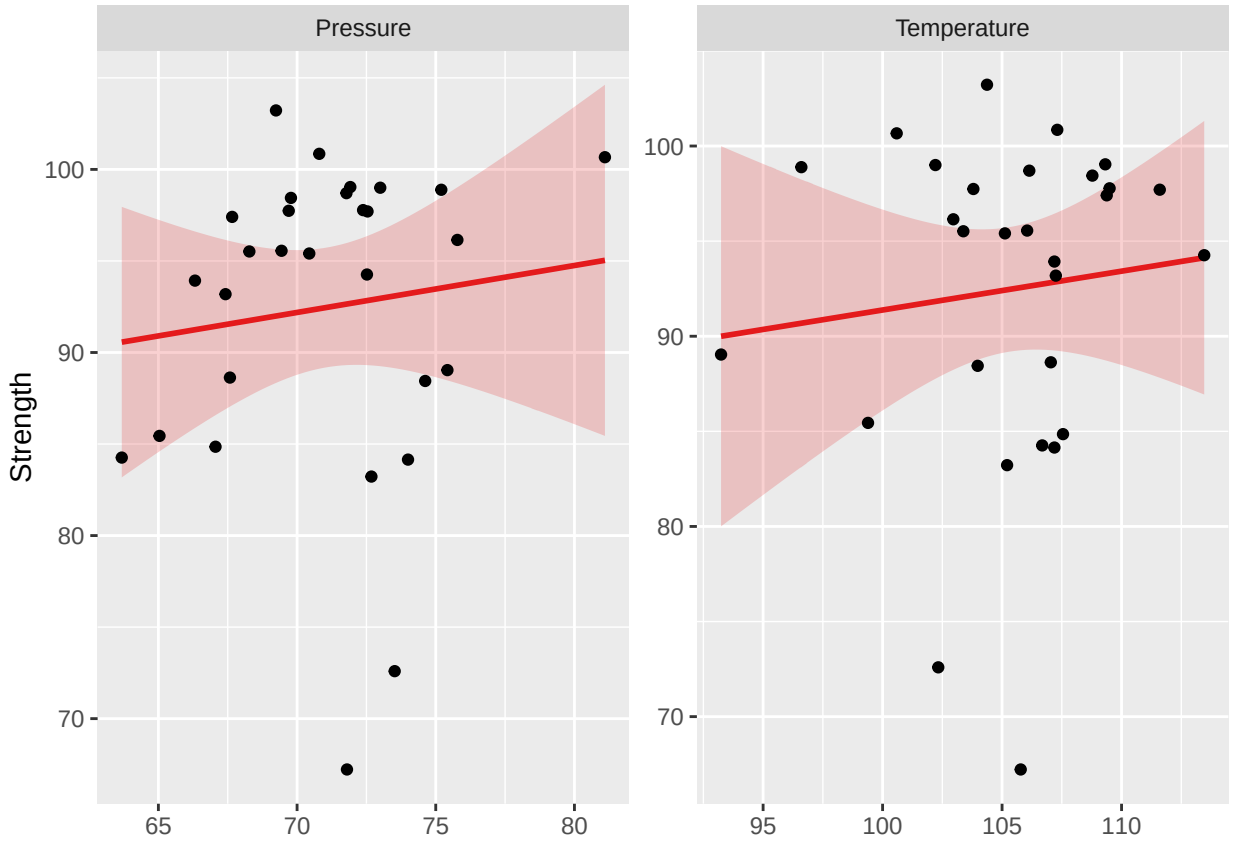
```
## Warning: Interaction terms are not supported by this plot type. Output for
## interaction terms may be inappropriate.
```




```
# Although the loess lines look a bit interesting...

# Let's plot the raw figures against this "Linear trend"
p <- plot_model(int_con_lm, type = "slope", terms = c("Temperature", "Pressure"), show.loess=F)
```

```
## Warning: Interaction terms are not supported by this plot type. Output for
## interaction terms may be inappropriate.
```

```
p + geom_point()
```




```
# hmm maybe we should have used a non linear approach on this data?
```