Solution to warpbreaks exercise
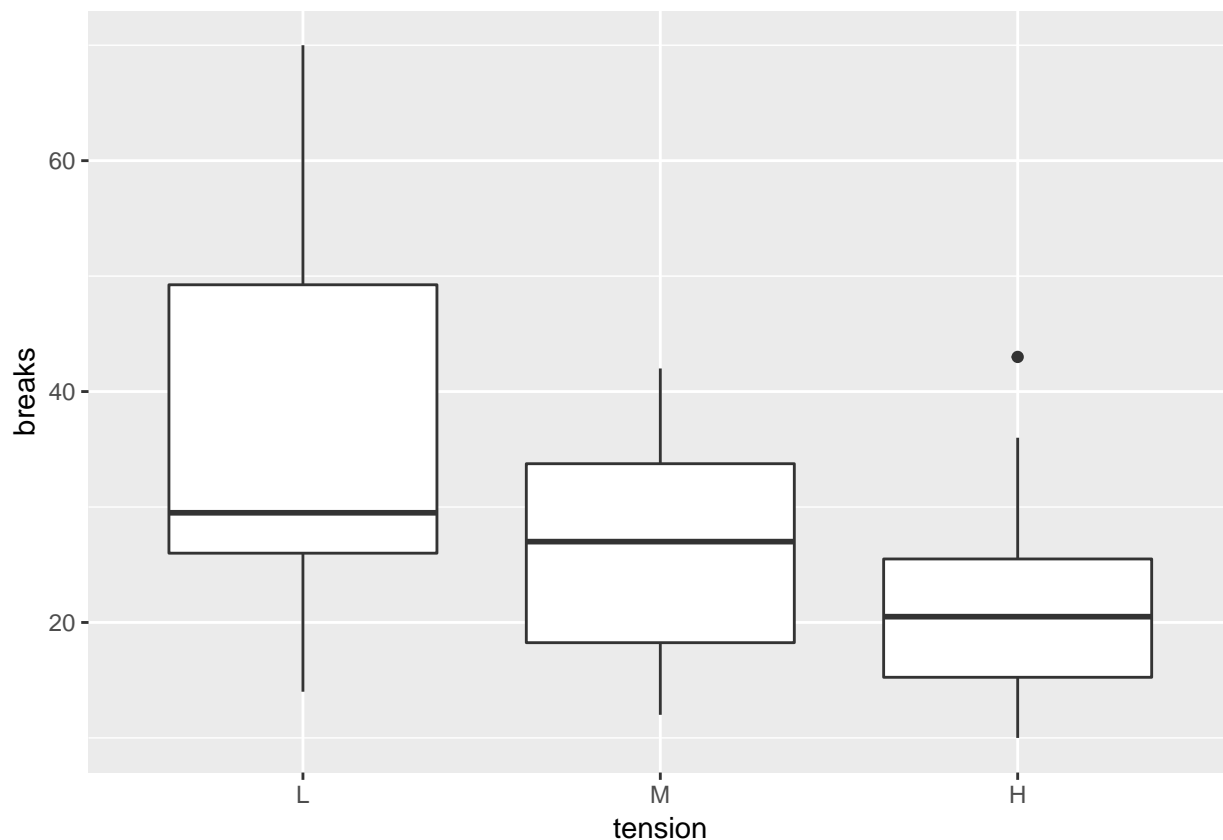
The number of breaks in yarn during weaving

The following dataset contains the results of an experiment where the number of yarn breaks during weaving are measured for two different types of wool at three different levels of string tension.

```
data(warpbreaks)
head(warpbreaks)
```

```
##   breaks wool tension
## 1     26    A       L
## 2     30    A       L
## 3     54    A       L
## 4     25    A       L
## 5     70    A       L
## 6     52    A       L
```

Perform a one way analysis of variance (ANOVA) of the `warpbreaks` data with `breaks` as response and `tension` as factor/explanatory variable. Remember to first plot the data.

```
gf_boxplot(breaks~tension,data=warpbreaks)
```


Then make a model in `R` using `lm`.

```
model1<-lm(breaks~tension ,data=warpbreaks)
summary(model1)
```

```
##
## Call:
## lm(formula = breaks ~ tension, data = warpbreaks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.389  -8.139  -2.667   6.333  33.611
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    36.39       2.80  12.995  < 2e-16 ***
## tensionM      -10.00       3.96  -2.525 0.014717 *
## tensionH      -14.72       3.96  -3.718 0.000501 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.88 on 51 degrees of freedom
## Multiple R-squared:  0.2203, Adjusted R-squared:  0.1898
## F-statistic: 7.206 on 2 and 51 DF,  p-value: 0.001753
```

Write down the prediction equation.

- We take low tension to be the reference group and define dummy variables:

$$z_1 = \begin{cases} 1 & \text{if medium tension,} \\ 0 & \text{otherwise.} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{if high tension,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\hat{y} = \alpha + \beta_1 z_1 + \beta_2 z_2$$
$$\hat{y} = 36.39 - 10.00 * z_1 - 14.72 * z_2$$

Use the prediction equation to calculate the predicted mean number of breaks for each tension level.

$$\hat{y} = \begin{cases} 36.39, & \text{for low tension,} \\ 26.39, & \text{for medium tension,} \\ 21.67, & \text{for high tension,} \end{cases}$$

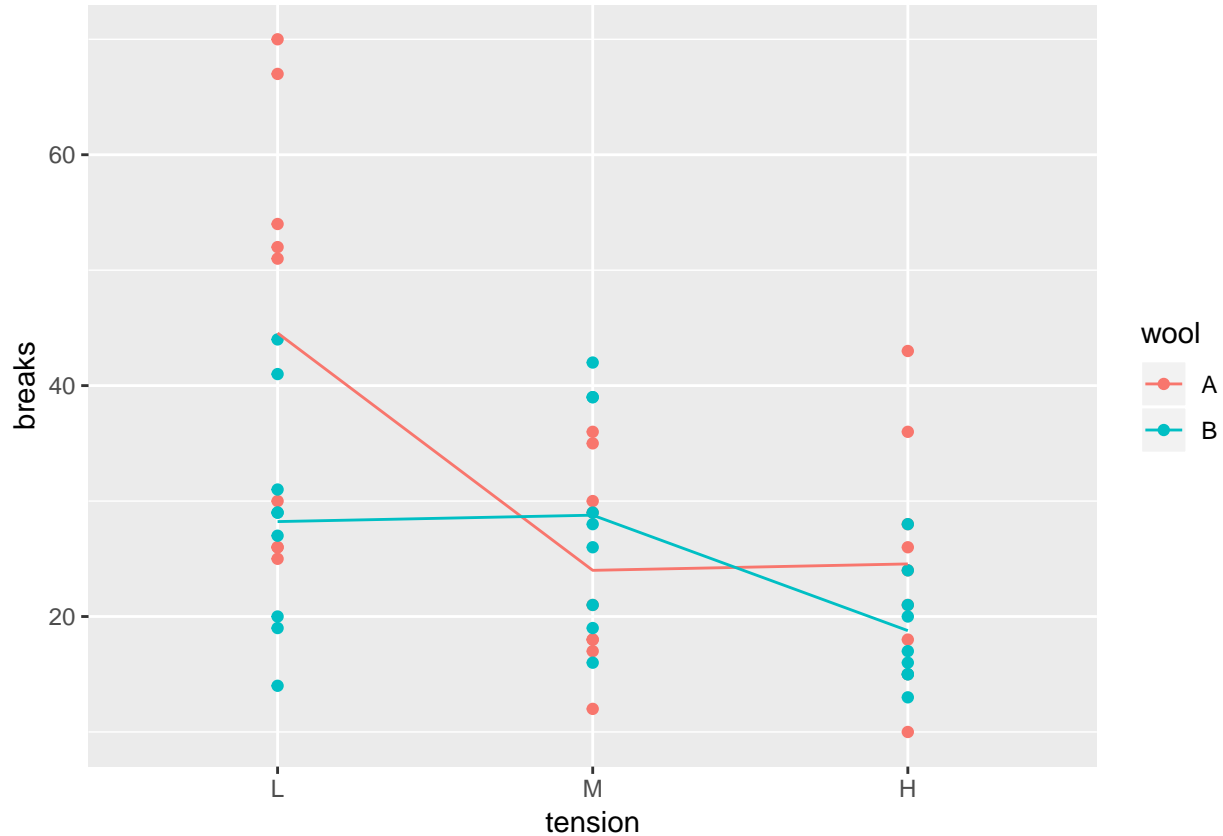Is the mean number of breaks the same for all tensions?

- The F-test at the bottom of the output provides a test for no effect of tension. This is rejected with $p = 0.001753$.

Next, perform a two way analysis of variance with both wool type and tension as factors and no interaction. Remember to make plot(s).

- We plot the sample mean for all combinations of wool and tension:

```r
gf_point(breaks ~ tension, color = ~ wool, data = warpbreaks) %>% gf_line(breaks ~ tension, group = ~woo
```

```
## No summary function supplied, defaulting to `mean_se()
```


- The lines are not parallel, suggesting an interaction. The model without interaction is fitted:

```r
model2<-lm(breaks ~ tension + wool, data=warpbreaks)
summary(model2)
```

```
##
## Call:
## lm(formula = breaks ~ tension + wool, data = warpbreaks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.500  -8.083  -2.139   6.472  30.722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.278      3.162  12.423  < 2e-16 ***
## tensionM     -10.000      3.872  -2.582 0.012787 *
## tensionH     -14.722      3.872  -3.802 0.000391 ***
## woolB         -5.778      3.162  -1.827 0.073614 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.62 on 50 degrees of freedom
## Multiple R-squared:  0.2691, Adjusted R-squared:  0.2253
## F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
```

Write down the prediction equation.

- Let
$$w = \begin{cases} 1 & \text{if wool has type B,} \\ 0 & \text{otherwise.} \end{cases}$$

  Then
$$\hat{y} = 39.3 - 10.00 * z_1 - 14.72 * z_2 - 5.778 * w$$

Use the prediction equation to calculate the predicted mean number of breaks for each combination of wool and tension.

- wool A, low tension: 39.3

- wool A, med. tension: 29.3

- wool A, high tension: 24.6

- wool B, low tension: 33.5

- wool B, med. tension: 23.5

- wool B, high tension: 18.8

Investigate if there is an interaction between wool and tension.

- We fit the model with interaction and make an F-test:

```
model3 <- lm(breaks ~ tension * wool, data = warpbreaks)
anova(model2, model3)
```

```
## Analysis of Variance Table
##
## Model 1: breaks ~ tension + wool
## Model 2: breaks ~ tension * wool
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     50 6747.9
## 2     48 5745.1  2    1002.8 4.1891 0.02104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that there is a significant interaction (at the 0.05 level) with $p = 0.021$.