Analyzing Breaks in Yarn by Tension

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.2.2

##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(broom)  
library(pixiedust)  
library(ggplot2)

# Purpose

To illustrate the reproducibility tools in R Statistical Software by performing a mock analysis of analyzing the number of warp breaks observed on 9 looms with independent factors wool and tension.

# The warpbreaks data

The warpbreaks dataset, available in the datasets package, is a data frame with 54 rows and 3 columns. There are 2 levels of the wool variable and 3 levels of the tension variable, with 9 observations at each combination.

For the purpose of our demonstration, we are going to select only the first six observations at each combination as our initial data set. We will then introduce the remaining three observations at each combination as if they had been collected after the first analysis.

#\* Note: Let me apologize here. The code that follows   
#\* is a little bit ahead of ourselves and I won't spend   
#\* a lot of time explaining it in this demonstration.   
#\* We'll come back to these concepts in a future meeting.  
  
Initial\_Breaks <- warpbreaks %>%  
 group\_by(wool, tension) %>%  
 mutate(observation\_number = 1:n()) %>%  
 filter(observation\_number %in% 1:6)  
  
Final\_Breaks <- warpbreaks %>%  
 group\_by(wool, tension) %>%  
 mutate(observation\_number = 1:n()) %>%  
 ungroup()

#\* At first is might seem strange to make a copy of the   
#\* data here. Hopefully we'll see the value of this later  
#\* in the demonstration  
Breaks <- Initial\_Breaks

*From this point on, we'll assume we're performing the analysis as if we didn't know there was more data coming later.*

# Exploratory Data Analysis

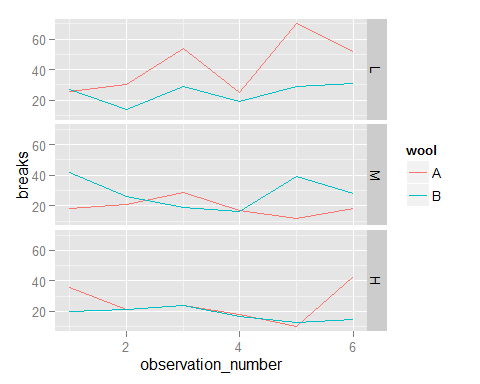
The data were gathered in a balanced design with 6 observations per wool-tension group. A total of 36 observations were made. A summary of the number of breaks in each group is provided in the table below.

Breaks %>%  
 group\_by(wool, tension) %>%  
 summarise(Mean = mean(breaks),  
 SD = sd(breaks),  
 Min = min(breaks),  
 P25 = quantile(breaks, probs = .25),  
 Median = median(breaks),  
 P75 = quantile(breaks, probs = .75),  
 Max = max(breaks)) %>%  
 as.data.frame()

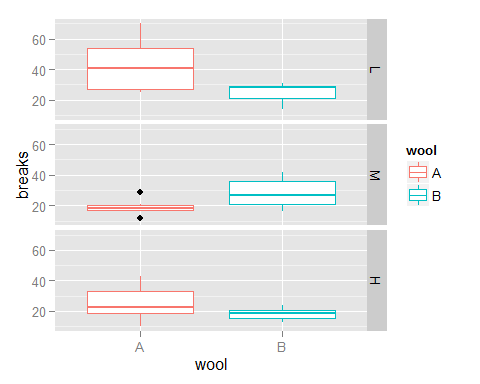
## wool tension Mean SD Min P25 Median P75 Max  
## 1 A L 42.83333 18.508557 25 27.00 41.0 53.50 70  
## 2 A M 19.16667 5.636193 12 17.25 18.0 20.25 29  
## 3 A H 25.33333 12.127105 10 18.75 22.5 33.00 43  
## 4 B L 24.83333 6.765107 14 21.00 28.0 29.00 31  
## 5 B M 28.33333 10.443499 16 20.75 27.0 36.25 42  
## 6 B H 18.33333 4.082483 13 15.50 18.5 20.75 24

And a few plots for the number of breaks by group

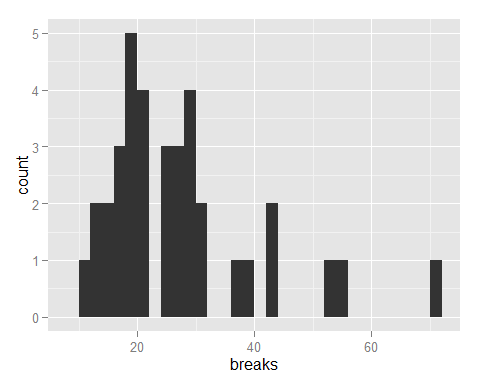
ggplot(Breaks,  
 aes(x = observation\_number,  
 y = breaks,  
 colour = wool)) +   
 geom\_line() +   
 facet\_grid(tension ~ .)



ggplot(Breaks,  
 aes(x = wool,  
 y = breaks,  
 colour = wool)) +   
 geom\_boxplot() +   
 facet\_grid(tension ~ .)



ggplot(Breaks,  
 aes(x = breaks)) +   
 geom\_histogram(binwidth = 2)



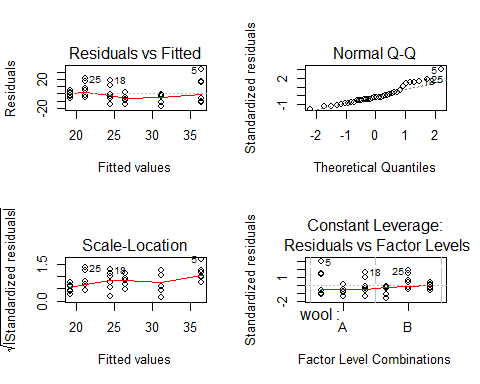
# Analysis

To evaluate the effect of wool and tension on the number of breaks observed, we perform an analysis of variance to see if there are differences between groups.

fit <- aov(breaks ~ wool + tension, data = Breaks)  
(results <- tidy(fit))

## term df sumsq meansq statistic p.value  
## 1 wool 1 250.6944 250.6944 1.749709 0.19528894  
## 2 tension 2 997.3889 498.6944 3.480613 0.04288229  
## 3 Residuals 32 4584.8889 143.2778 NA NA

#\* Check the model diagnostics  
par(mfrow=c(2, 2))  
plot(fit)



An ordinary least squares / linear regression model provides us with estimates of the change in breaks between levels of the tension variable.

(lm\_result <- tidy(lm(fit)))

## term estimate std.error statistic p.value  
## 1 (Intercept) 36.472222 3.989957 9.141007 1.946292e-10  
## 2 woolB -5.277778 3.989957 -1.322766 1.952889e-01  
## 3 tensionM -10.083333 4.886679 -2.063433 4.726183e-02  
## 4 tensionH -12.000000 4.886679 -2.455656 1.967763e-02

# Conclusions

Based on the results of the analysis of variance, we find insufficient evidence to claim that wool is associated with a difference in yarn breaks (p = 0.2). A statistically significant association between tension and yarn breaks was observed (p = 0.043). The results of the linear regression suggest that breaks decrease as tension incrases.