Sweave Intro

First we define a figure hook:

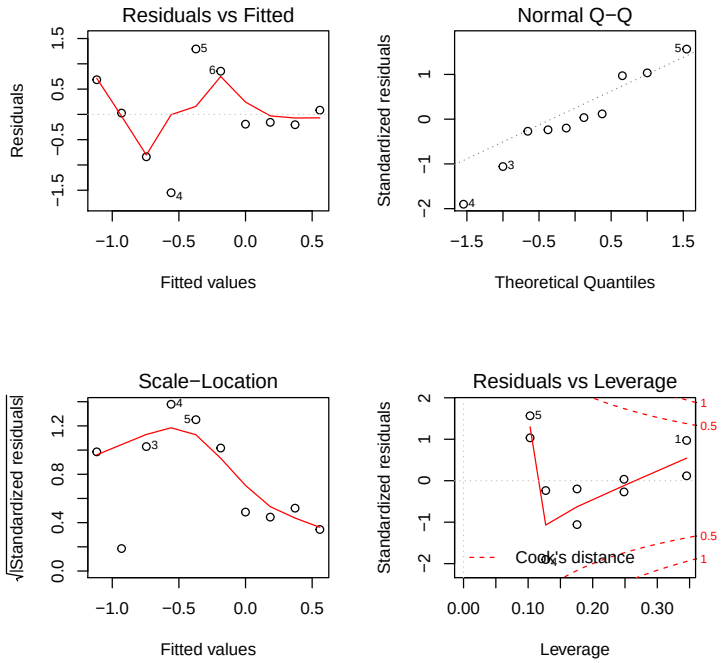
```
> options(SweaveHooks = list(fig = function() par(mfrow=c(2,2))))
```

Then we setup variable definitions without actually evaluating them

```
> x <- 1:10
> y <- rnorm(x)
```

Then we put the pieces together:

```
> x <- 1:10
> y <- rnorm(x)
> lm1 <- lm(y~x)
> summary(lm1)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-1.54911 -0.20161 -0.06565  0.53485  1.29361

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.30234    0.59592  -2.185   0.0603 .
x            0.18601    0.09604   1.937   0.0888 .
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 0.8723 on 8 degrees of freedom
Multiple R-squared:  0.3192,        Adjusted R-squared:  0.2341
F-statistic: 3.751 on 1 and 8 DF,  p-value: 0.08878

> plot(lm1)
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

Iris Data

Consider the classic `iris` data set. The data frame contains measurements of petal and sepals from 150 Iris flowers of three related species: (50 Iris setosa, 50 Iris virginica, and 50 Iris versicolor). A linear regression model of sepal length as a function of the sepal width can be fitted in R using the command

```
> lm1 = lm(Sepal.Length ~ Sepal.Width, data=iris)
> lm1

Call:
lm(formula = Sepal.Length ~ Sepal.Width, data = iris)

Coefficients:
(Intercept)  Sepal.Width
     6.5262      -0.2234
```

Tests for significance of the coefficients are shown in Table 1.

```
> xtable(lm1, caption="Linear regression model for iris data.",
+ label="tab:coef1")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.5262	0.4789	13.63	0.0000
Sepal.Width	-0.2234	0.1551	-1.44	0.1519

Table 1: Linear regression model for iris data.

It seems that sepal width is not a very good predictor of sepal length. However, the model performs far better when the data is seperated by species, as seen in a second linear model:

```
> lm2 = lm(Sepal.Length ~ Sepal.Width:Species + Species - 1, data=iris)
> lm2

Call:
lm(formula = Sepal.Length ~ Sepal.Width:Species + Species - 1,
    data = iris)

Coefficients:
               Speciessetosa              Speciesversicolor
                      2.6390                         3.5397
            Speciesvirginica     Sepal.Width:Speciessetosa
                      3.9068                         0.6905
Sepal.Width:Speciesversicolor  Sepal.Width:Speciesvirginica
                      0.8651                         0.9015
```

With coefficients displayed in Table 2.

```
> xtable(lm2, caption="Linear regression model for iris data, by species",
+ label="tab:coef2")
```

A scatter plot including regression lines from both experiments is shown in Figure 1.

	Estimate	Std. Error	t value	Pr(>|t|)
Speciessetosa	2.6390	0.5715	4.62	0.0000
Speciesversicolor	3.5397	0.5580	6.34	0.0000
Speciesvirginica	3.9068	0.5827	6.71	0.0000
Sepal.Width:Speciessetosa	0.6905	0.1657	4.17	0.0001
Sepal.Width:Speciesversicolor	0.8651	0.2002	4.32	0.0000
Sepal.Width:Speciesvirginica	0.9015	0.1948	4.63	0.0000

Table 2: Linear regression model for iris data, by species

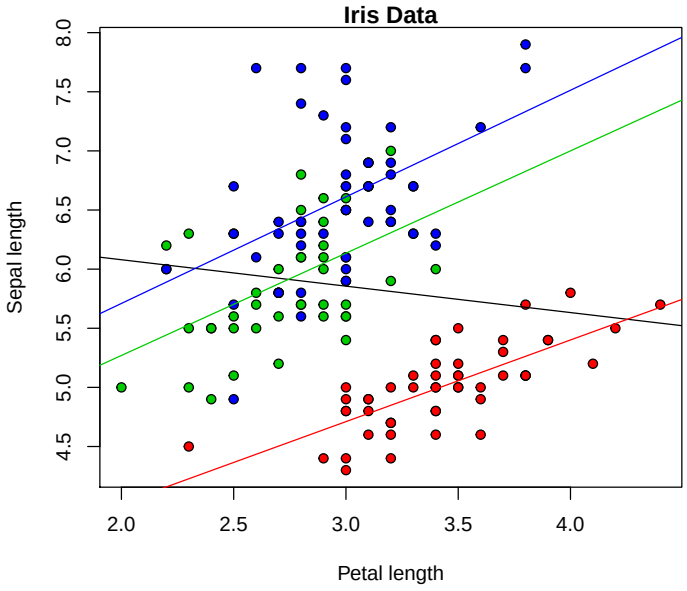


Figure 1: Iris sepal length vs width, broken down by species.