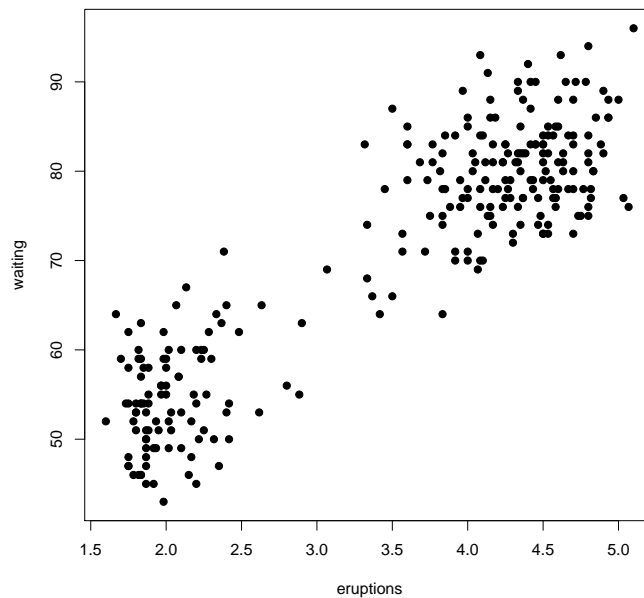


**Background** Consider the Old Faithful data. The data consist of two variables: waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming. There are 272 consecutive observations (date unknown).

```
> data(faithful)
> faithful
> attach(faithful)
> names(faithful)
> plot(eruptions, waiting)
> apply(faithful,2,mean)
eruptions  waiting
3.487783  70.897059
> apply(faithful,2,sd)
eruptions  waiting
1.141371  13.594974
```



## Correlation

```
> cor(eruptions,waiting)
[1] 0.9008112

> cor.test(eruptions,waiting)
t = 34.089, df = 270, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8756964 0.9210652
sample estimates:
      cor
0.9008112
```

## Regression

```
> faith.lm <- lm(waiting ~ eruptions)
```

```
> summary(faith.lm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	33.4744	1.1549	28.98	<2e-16 ***
eruptions	10.7296	0.3148	34.09	<2e-16 ***

Residual standard error: 5.914 on 270 degrees of freedom

Multiple R-squared: 0.8115, Adjusted R-squared: 0.8108

F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16

```
> anova(faith.lm)
```

Analysis of Variance Table

Response: waiting

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
eruptions	1	40644	40644	1162.1	< 2.2e-16 ***
Residuals	270	9443	35		

```
> plot(fitted(faith.lm), resid(faith.lm), pch=19, xlab="fitted values",  
       ylab="residuals", main="Old Faithful data")  
> abline(h=0)
```

