Unit 3 Tutorial

Cheatsheet for fitting a linear model in R

Load data

Recall the mortality and hard water data from yesterday. We have recordings of the mortality rate (deaths per 100,000 population) and concentration of calcium in drinking water (parts per million) in 61 large towns in England and Wales

```
mortality_water <- read_csv("https://mhc-stat140-2017.github.io/data/sdm4/Hard_water_Derby.csv")
```

Scatter plot

```
ggplot(data = mortality_water, mapping = aes(x = Calcium, y = Mortality)) +
  geom_point()
```
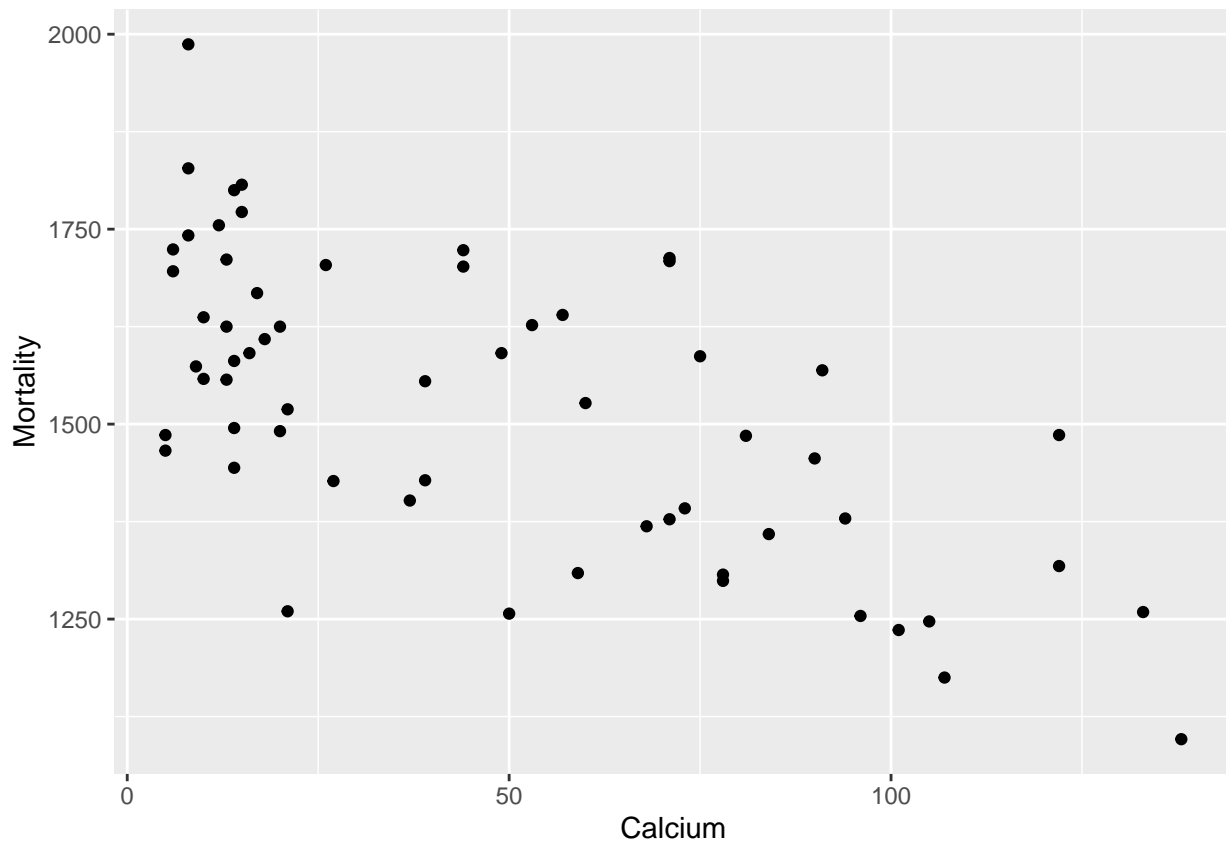
Fit a linear model

We fitted a linear model between the mortality rate `Mortality` and concentration of calcium `Calcium` using the `lm` function.

```
linear_fit <- lm(Mortality ~ Calcium, data = mortality_water)
```

View the summary of `lm`

View summary of linear model fit

```
summary(linear_fit)
```

```
##
## Call:
## lm(formula = Mortality ~ Calcium, data = mortality_water)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -348.61  -114.52    -7.09   111.52   336.45
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1676.3556    29.2981  57.217  < 2e-16 ***
## Calcium       -3.2261     0.4847  -6.656 1.03e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143 on 59 degrees of freedom
## Multiple R-squared:  0.4288,	Adjusted R-squared:  0.4191
## F-statistic:  44.3 on 1 and 59 DF,  p-value: 1.033e-08
```

intercept

slope

Predict

One of the towns in our sample had a measured Calcium concentration of 71. What is the predicted value for the mortality rate in that town?

```
predict_data <- data.frame( Calcium = 71
)
predict(linear_fit, newdata = predict_data)
```

```
##        1
## 1447.303
```