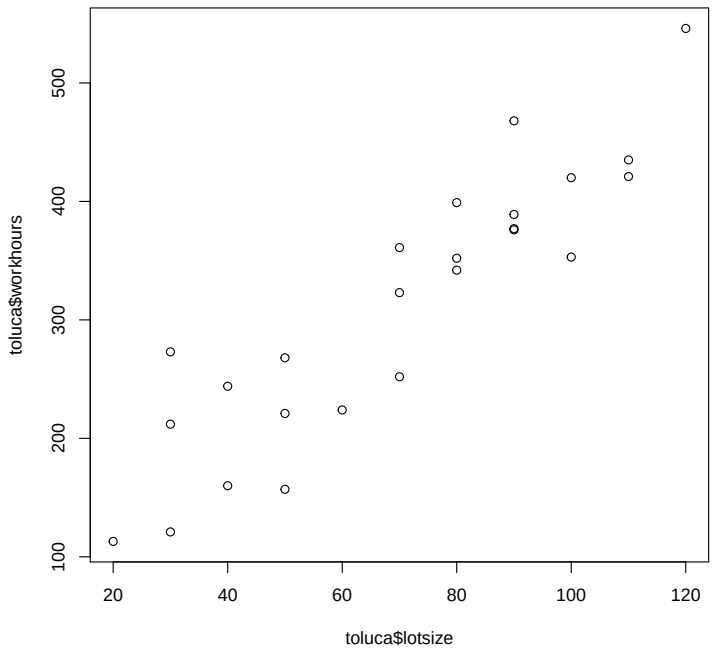
2.1.1: R: Simple Linear Regression

Stat 5100: Dr. Bean

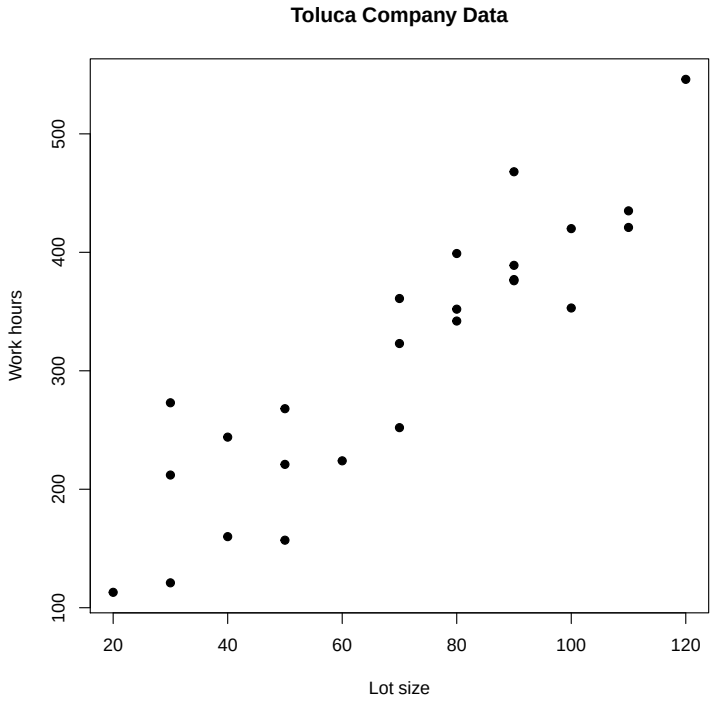
Example: The Toluca Company makes replacement parts for refrigeration equipment. For a certain part, it takes some time to set up the production process, and then the production of a given lot size can begin. As part of a cost improvement program, the company wished to better understand the relationship between the lot size (X) and the total work hours (Y). Data were reported for 25 representative lots of varying size.

```
library(stat5100)
data(toluca)

# Make a scatterplot of work hours and lotsize
plot(toluca$lotsize, toluca$workhours)
```




```
# Give it some labels
plot(toluca$lotsize, toluca$workhours,
     xlab = "Lot size", ylab = "Work hours", main = "Toluca Company Data",
     pch = 19)
```




```
# View some summary and correlation statistics
summary(toluca)
```

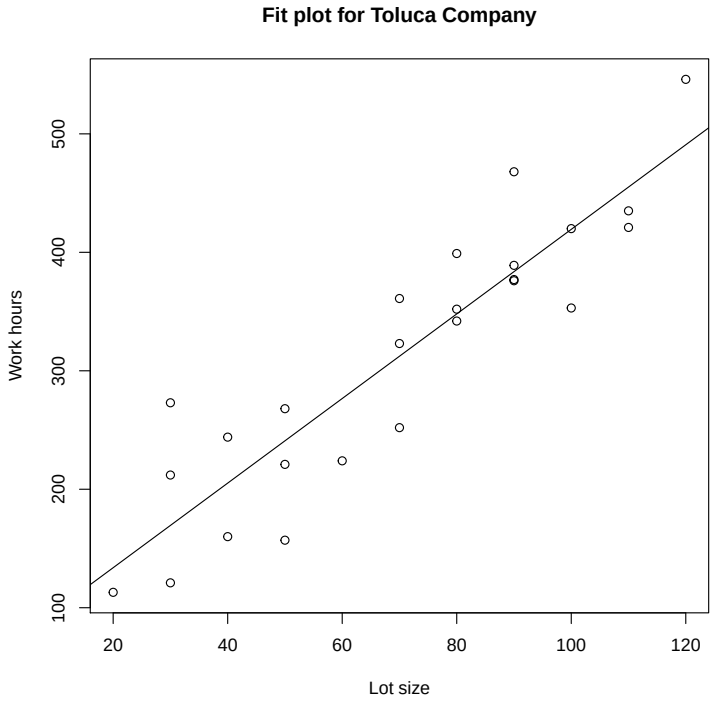
```
##     lotsize      workhours
##  Min.   : 20   Min.   :113.0
##  1st Qu.: 50   1st Qu.:224.0
##  Median : 70   Median :342.0
##  Mean   : 70   Mean   :312.3
##  3rd Qu.: 90   3rd Qu.:389.0
##  Max.   :120   Max.   :546.0
```

```
cor(toluca)
```

```
##             lotsize workhours
## lotsize   1.0000000 0.9063848
## workhours 0.9063848 1.0000000
```

```
# Fit a simple linear model with Y = workhours and X = lotsize
toluca_lm <- lm(workhours ~ lotsize, data = toluca)
```

```
# Look at a fit plot for the linear model
stat5100::fit_plot(toluca_lm, main = "Fit plot for Toluca Company",
         xlab = "Lot size", ylab = "Work hours")
```




```
# Look at the ANOVA table and coefficient estimates
toluca_lm

##
## Call:
## lm(formula = workhours ~ lotsize, data = toluca)
##
## Coefficients:
## (Intercept)      lotsize
##       62.37         3.57

anova(toluca_lm)

## Analysis of Variance Table
##
## Response: workhours
##           Df Sum Sq Mean Sq F value    Pr(>F)
## lotsize    1 252378  252378  105.88 4.449e-10 ***
## Residuals 23  54825    2384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Look at a sample of predicted values
sample_predicted <- cbind(toluca, pred_workhours = toluca_lm$fitted.values)
head(sample_predicted)

##   lotsize workhours pred_workhours
## 1      80       399       347.9820
## 2      30       121       169.4719
## 3      50       221       240.8760
## 4      90       376       383.6840
## 5      70       361       312.2800
## 6      60       224       276.5780
```