HW2

Raphael Nash

9/3/2017

```r
library("ggplot2")
```

p69 # 12

For the scenarios presented in Problems 9.17, identify a problem worth studying and list the variables that affect the behavior you have identified. Which variables would be neglected completely? Which might be considered as constants initially? Can you identify any submodels you would want to study in detail? Identify any data you would want collected.

A company with a fleet of trucks faces increasing maintenance costs as the age and mileage of the trucks increase.

Identify a problem:

Total Cost of ownership of the truck

What variables to be considered:

- Cost of truck
- Engine Size
- Maintence schedule
- Fuel type and effeciency
- Type of driving city vs highway
- Reliability of make/model

What vairables could be neglected:

- Type of financing
- Cost of employee's

Initial Constants:

- Routine Maintence Schedule
- Insurance cost

Submodels:

- Miles per galon based on driving conditions
- Non Scheduled maintence based on brand/driving
- Depreciate per month

Data needing to be collected:

- Non scheduled maintence
- Full useage
- Driving logs of driving types and conditions

p 79 # 11

determine whether the data set supports the stated proportionality model.

$$y \propto x^3$$

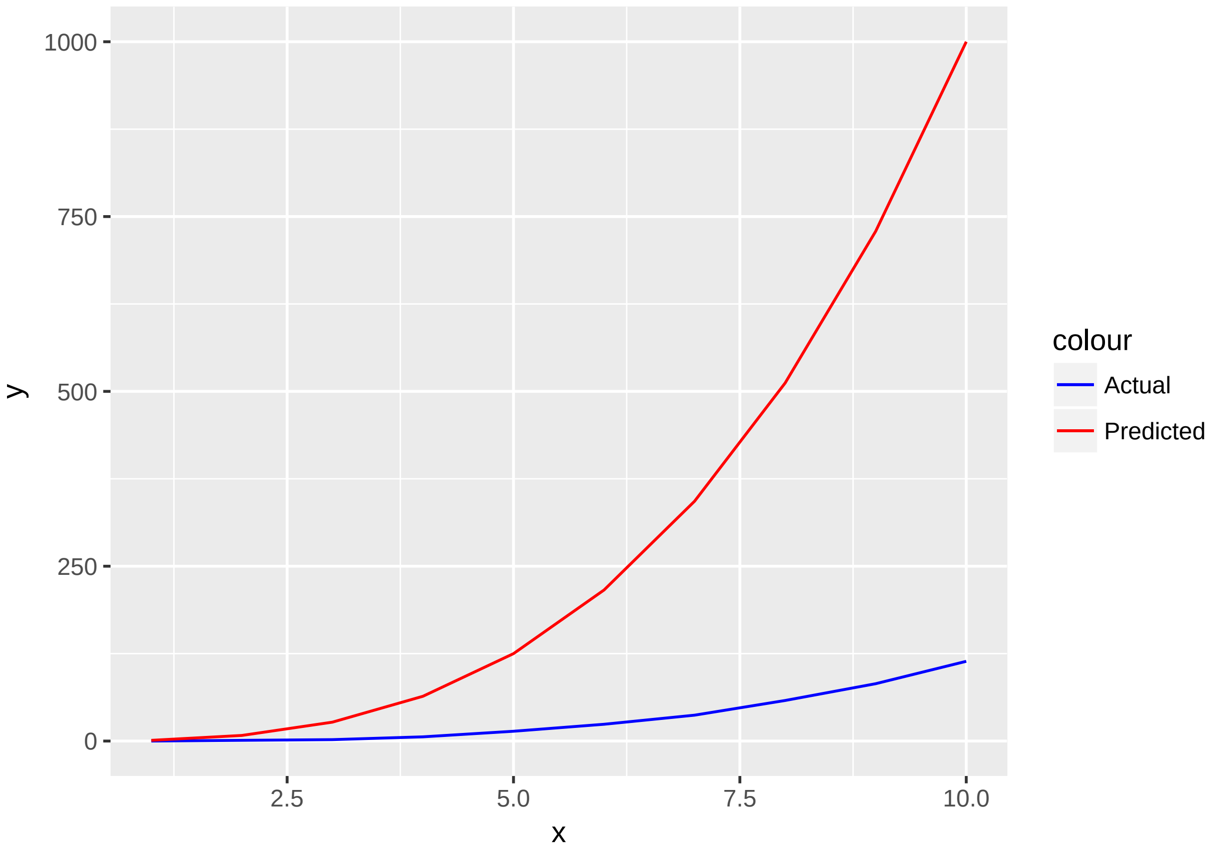
```r
problem_data <- data.frame(
  y = c(0,1,2,6,14,24,37,58,82,114),
  x = c(1,2,3,4,5,6,7,8,9,10 )
  )

problem_data
```

```
##      y  x
## 1    0  1
## 2    1  2
## 3    2  3
## 4    6  4
## 5   14  5
## 6   24  6
## 7   37  7
## 8   58  8
## 9   82  9
## 10 114 10
```

```r
problem_data["y_predicted"] = problem_data$x^3
problem_data
```

```
##      y  x y_predicted
## 1    0  1           1
## 2    1  2           8
## 3    2  3          27
## 4    6  4          64
## 5   14  5         125
## 6   24  6         216
## 7   37  7         343
## 8   58  8         512
## 9   82  9         729
## 10 114 10        1000
```

```r
ggplot(problem_data) + geom_line(aes(x=x, y=y , color = "blue" )) +
  geom_line(aes(x=x, y=y_predicted , color = "red" )) +
  scale_color_manual(labels = c("Actual", "Predicted"), values = c("blue", "red"))
```



Is there a ratio of the Y to Y predicted?

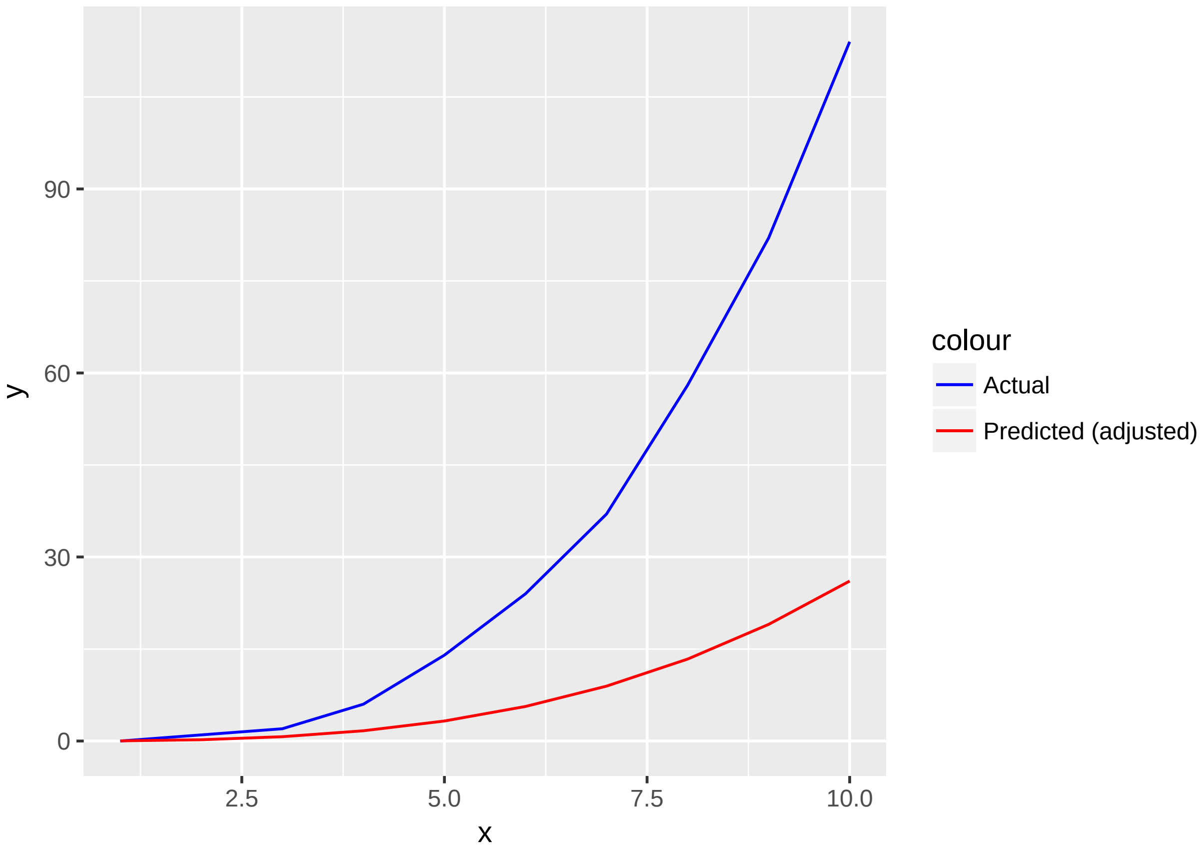
```
coef = mean (problem_data[,"y"]^(1/3)/problem_data[,"y_predicted"]  )
coef
```

```
## [1] 0.02607751
```

```
problem_data["y_predicted_adjusted"] <- problem_data$x^3 * coef
problem_data
```

```
##      y  x y_predicted y_predicted_adjusted
## 1    0  1           1           0.02607751
## 2    1  2           8           0.20862009
## 3    2  3          27           0.70409279
## 4    6  4          64           1.66896068
## 5   14  5         125           3.25968884
## 6   24  6         216           5.63274231
## 7   37  7         343           8.94458617
## 8   58  8         512          13.35168547
## 9   82  9         729          19.01050529
## 10 114 10        1000          26.07751069
```

```
ggplot(problem_data) + geom_line(aes(x=x, y=y , color = "blue" )) +
geom_line(aes(x=x, y=y_predicted_adjusted , color = "red" )) +
  scale_color_manual(labels = c("Actual", "Predicted (adjusted) "), values = c("blue", "red")) +
  ylab("y")
```



I would say the proportionality $y \propto x^3$ does not really hold up.

p94 # 4

Lumber Cutters Lumber cutters wish to use readily available measurements to estimate the number of board feet of lumber in a tree. Assume they measure the diameter of the tree in inches at waist height. Develop a model that predicts board feet as a function of diameter in inches. Use the following data for your test:

```
lumber_data = data.frame(
      x= c(17,19,20,23,25,28,32,38,39,41),
      y= c(19,25,32,57,71,113,123,252,259,294))
lumber_data
```

```
##     x   y
## 1  17  19
## 2  19  25
## 3  20  32
## 4  23  57
## 5  25  71
## 6  28 113
## 7  32 123
## 8  38 252
## 9  39 259
## 10 41 294
```

The variable x is the diameter of a ponderosa pine in inches, and y is the number of board feet divided by 10.

The variable x is the diameter of a ponderosa pine in inches, and y is the number of board feet divided by 10.

a. Consider two separate assumptions, allowing each to lead to a model. Completely analyze each model.

geometric proportionality is based on $f(x) = \pi r^2 h$

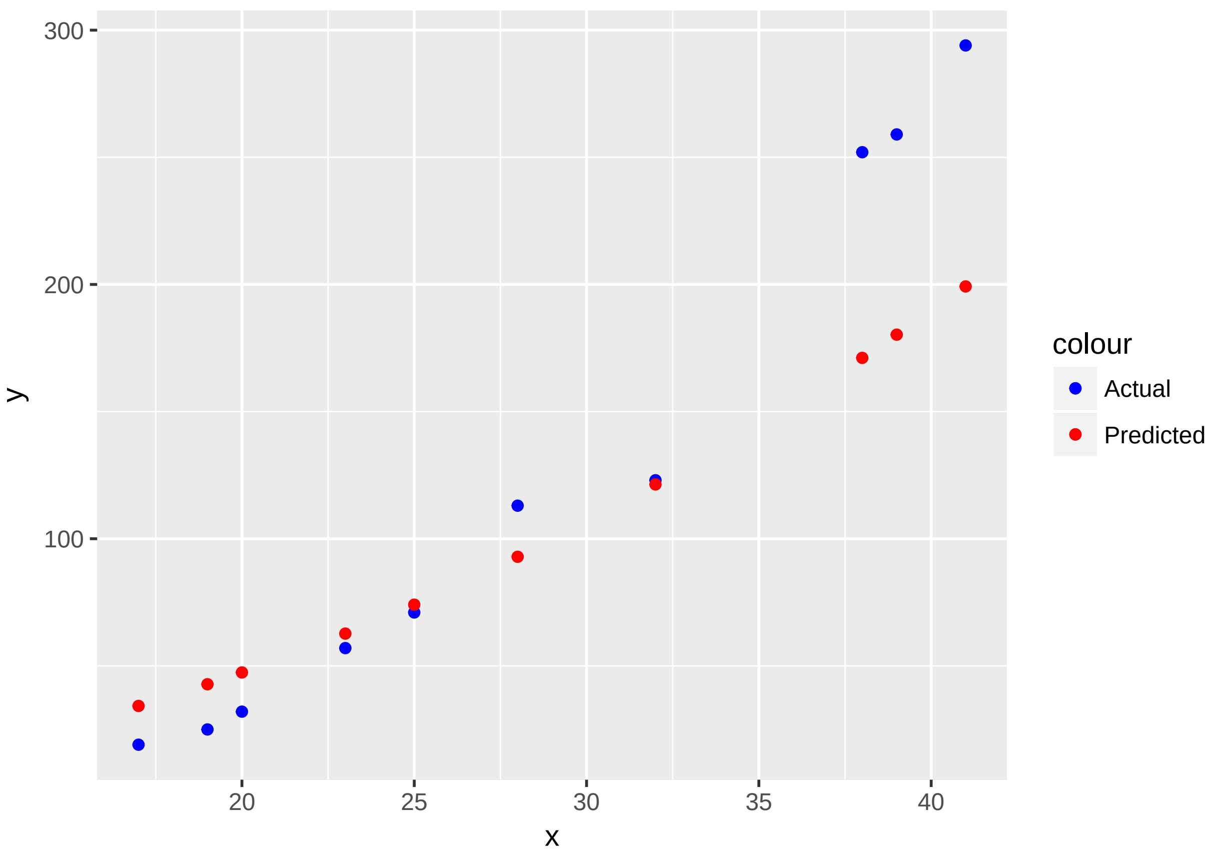
i. Assume that all trees are right-circular cylinders and are approximately the same height.

In this scenario $y \propto x^2$ since on the r^2 term is considard

```
part_i <- lumber_data

coef <-   mean ((part_i[,"y"]^(1/2))/part_i[,"x"]  )^2
coef
```

```
## [1] 0.118511
```

```
part_i["y_predicted"] <- coef * (part_i["x"] ) ^2


ggplot(part_i) + geom_point(aes(x=x, y=y , color = "blue" )) +
geom_point(aes(x=x, y=y_predicted , color = "red" )) +
  scale_color_manual(labels = c("Actual", "Predicted"), values = c("blue", "red")) +
  ylab("y")
```



ii. Assume that all trees are right-circular cylinders and that the height of the tree is propor-

tional to the diameter.

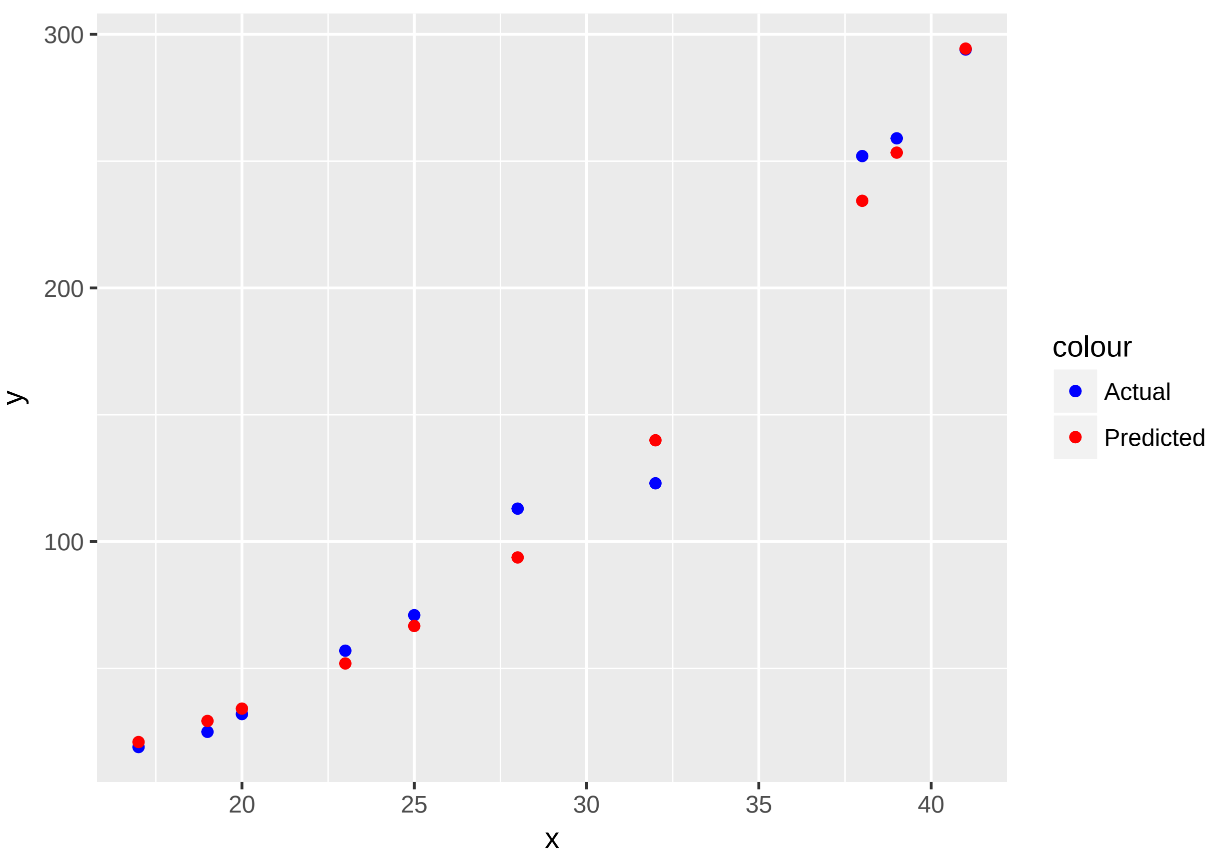
```
part_ii <- lumber_data

coef <-   mean (part_ii[,"y"]^(1/3)/part_ii[,"x"]  )^3
coef
```

```
## [1] 0.004271082
```

```
part_ii["y_predicted"] <- coef * (part_ii["x"] ) ^3

part_ii
```

```
##     x   y          x
## 1  17  19  20.98383
## 2  19  25  29.29535
## 3  20  32  34.16866
## 4  23  57  51.96625
## 5  25  71  66.73566
## 6  28 113  93.75879
## 7  32 123 139.95481
## 8  38 252 234.36281
## 9  39 259 253.35631
## 10 41 294 294.36724
```

```
ggplot(part_ii) + geom_point(aes(x=x, y=y , color = "blue" )) +
geom_point(aes(x=x, y=y_predicted , color = "red" )) +
  scale_color_manual(labels = c("Actual", "Predicted"), values = c("blue", "red")) +
  ylab("y")
```



```
part_ii
```

```
##     x   y         x
## 1  17  19  20.98383
## 2  19  25  29.29535
## 3  20  32  34.16866
## 4  23  57  51.96625
## 5  25  71  66.73566
## 6  28 113  93.75879
## 7  32 123 139.95481
## 8  38 252 234.36281
## 9  39 259 253.35631
## 10 41 294 294.36724
```

** Which model appears to be better? Why? Justify your conclusions.**

Part II where the height is proportional to the radius seems to make poor sense for 2 reasons.

1. This makes sense from what I know about trees older trees are taller and have a larger trunk radius.
2. The graph of actual vs predicted boord feet seems to fit better on that model.