Project 1 – FIN654

Bill Foote

January 17, 2017

Problem and Questions

What is the impact of terms and conditions of commercial loans on loan rates? xxxx

Data Preparation

Data is from FRED .

. . . Data dictionary. . .

xx	xx	xxx

some more words to talk about preparation

```
x.data <- read.csv("data/commloans.csv")
head(x.data, n = 5)
```

```
##        date prepaypenalty maturity rate size volume
## 1  4/1/2003          16.5      124 3.77  449  11406
## 2  7/1/2003          18.1       70 3.09  356  14586
## 3 10/1/2003          44.9       48 2.83  532  21022
## 4  1/1/2004          30.4       87 3.06  602  21472
## 5  4/1/2004          23.5       68 2.97  600  22359
```

```
tail(x.data, n = 5)
```

```
##         date prepaypenalty maturity rate size volume
## 50  7/1/2015          16.9       76 2.30 1405  30586
## 51 10/1/2015          11.7       77 2.31 1534  36840
## 52  1/1/2016          13.6       66 2.43 1317  36316
## 53  4/1/2016          20.6       93 2.63 1227  24803
## 54  7/1/2016          14.5       66 2.41 1460  40682
```

```
summary(x.data)
```

```
##        date    prepaypenalty      maturity          rate
##  1/1/2004: 1   Min.   : 8.80   Min.   : 40.00   Min.   :2.240
##  1/1/2005: 1   1st Qu.:16.95   1st Qu.: 68.25   1st Qu.:2.482
##  1/1/2006: 1   Median :20.70   Median : 89.00   Median :2.825
##  1/1/2007: 1   Mean   :23.06   Mean   : 95.28   Mean   :3.652
##  1/1/2008: 1   3rd Qu.:29.93   3rd Qu.:112.25   3rd Qu.:4.197
##  1/1/2009: 1   Max.   :51.90   Max.   :396.00   Max.   :7.410
##  (Other) :48
##       size            volume
##  Min.   : 356.0   Min.   :11406
##  1st Qu.: 639.5   1st Qu.:15451
##  Median : 824.5   Median :18670
##  Mean   : 881.7   Mean   :20824
##  3rd Qu.:1017.8   3rd Qu.:24258
##  Max.   :1715.0   Max.   :40682
```

##

Short discussion about data. Note anomalies.

Data Analysis

Data Exploration

. . . lead in description . . .

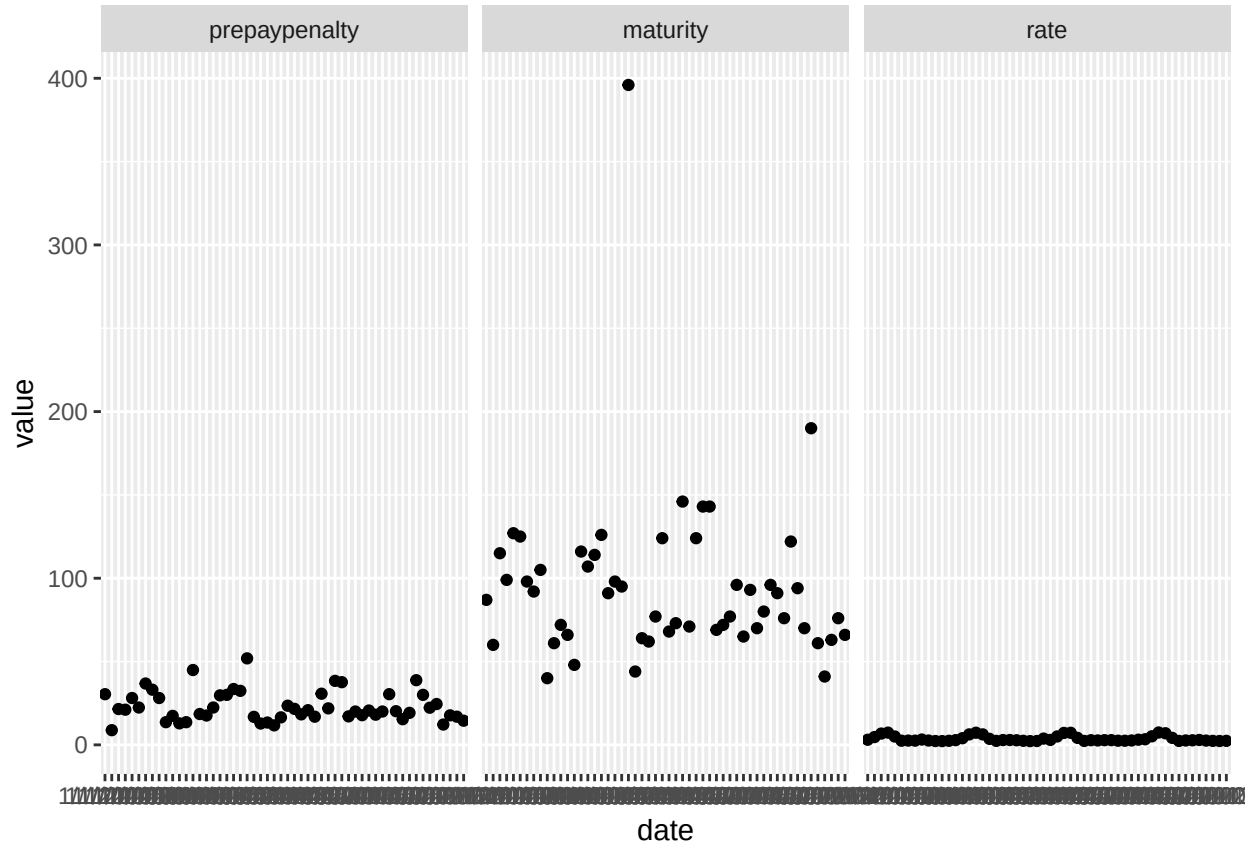
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(reshape2)
```

```
## Loading required package: reshape2
```

```
# Use melt() from reshape2 to build data frame with data as id and values of variables
x.melted <- melt(x.data[, c(1:4)], id = "date")
ggplot(data = x.melted, aes(x = date, y = value)) +
  geom_point() + facet_wrap(~variable, scales = "free_x")
```



commentary on the plots.

Relational Analysis

. . . lead in ..

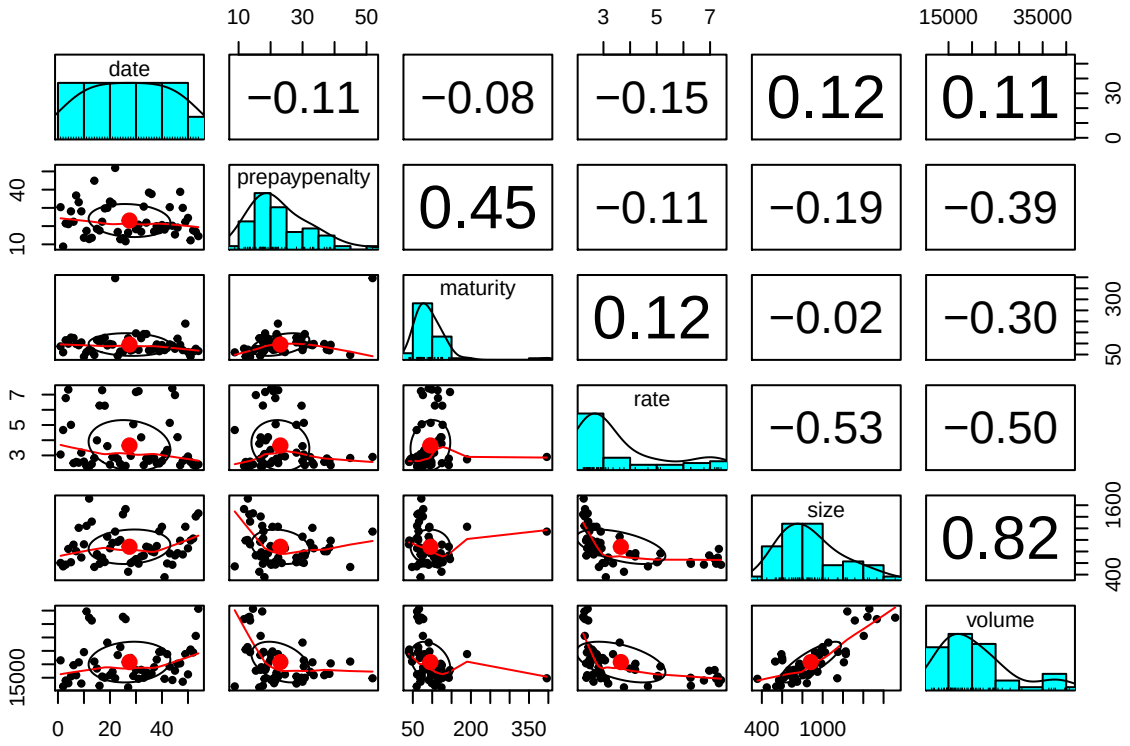
```
require(psych)
```

```
## Loading required package: psych
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
pairs.panels(x.data)
```



Note any observations here.

Let's regress `rate` on the rest of the variables in `x.data`. To do this we form a matrix of independent variables (predictor or explanatory variables) in the matrix `X` and a separate vector `y` for the dependent (response) variable `rate`. We recall that the `1` vector will produce a constant intercept in the regression model.

```
y <- as.vector(x.data[,"rate"])
X <- as.matrix(cbind(1, x.data[,c("prepaypenalty", "maturity", "size", "volume")]))
head(y)
```

```
## [1] 3.77 3.09 2.83 3.06 2.97 3.36
```

```
head(X)
```

```
##      1 prepaypenalty maturity size volume
## [1,] 1          16.5      124  449  11406
## [2,] 1          18.1       70  356  14586
## [3,] 1          44.9       48  532  21022
## [4,] 1          30.4       87  602  21472
## [5,] 1          23.5       68  600  22359
## [6,] 1          20.0       80  593  23780
```

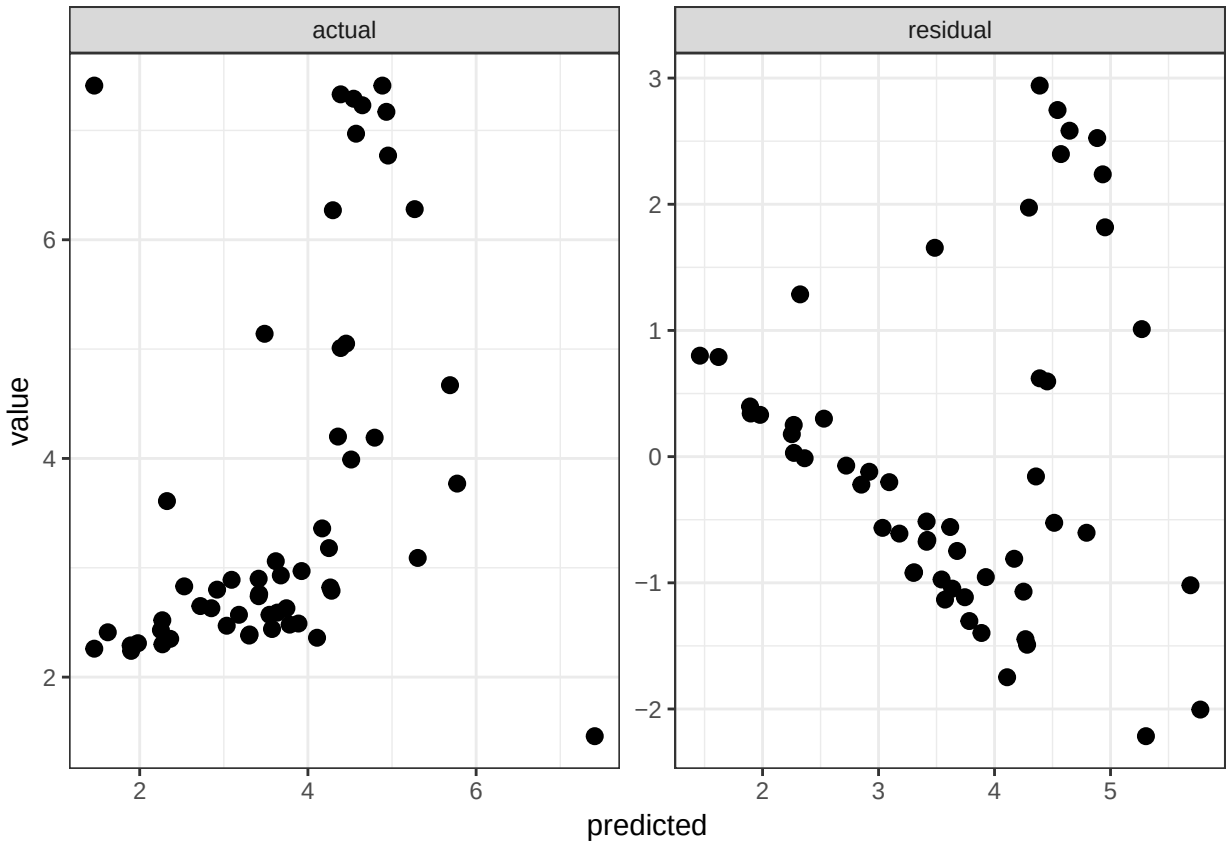
- Explain the code used to form `y` and `X`.
- Calculate the $\hat{\beta}$ coefficients and interpret their meaning.

```
XTX.inverse <- solve(t(X) %*% X)
(beta.hat <- XTX.inverse %*% t(X) %*% y )
```

```
##                         [,1]
## 1               7.771438e+00
## prepaypenalty -6.968996e-02
## maturity       6.399952e-03
## size          -2.041351e-03
## volume        -6.347851e-05
```

- Calculate actual and predicted `rates` and plot using this code.

```
# Insert comment here
require(reshape2)
require(ggplot2)
actual <- y
predicted <- X%*%beta.hat
residual <- actual - predicted
results <- data.frame(actual = actual, predicted = predicted, residual = residual)
# Insert comment here
min_xy <- min(min(results$actual), min(results$predicted))
max_xy <- max(max(results$actual), max(results$predicted))
# Insert comment here
plot.melt <- melt(results, id.vars = "predicted")
# Insert comment here
plot.data <- rbind(plot.melt, data.frame(predicted = c(min_xy, max_xy), variable = c("actual", "actual"
# Insert comment here
p <- ggplot(plot.data, aes(x = predicted, y = value)) + geom_point(size = 2.5) + theme_bw()
p <- p + facet_wrap(~variable, scales = "free")
p
```



- Insert explanatory comments into the code chunk to document the work flow for this plot.
- Interpret the graphs of actual and residual versus predicted values of `rate`.
- Calculate the standard error of the residuals. Interpret its meaning.

```
e <- y - X %*% beta.hat
(e.sse <- t(e) %*% e)
```

```
##           [,1]
## [1,] 88.62341
```

```
(n <- dim(X)[1])
```

```
## [1] 54
```

```
(k <- nrow(beta.hat))
```

```
## [1] 5
```

```
(e.se <- (e.sse / (n - k))^0.5)
```

```
##           [,1]
## [1,] 1.344857
```

. . . note any analytical insights

Observations and Recommmendations

What we observe especially insights.

Any recommendations for management in setting terms and conditions of loans.

Sources

Google, CRAN for ggplot2, reshape2, FRED, discussions with whomever. Other sources.