initial

Abhinav Tripathy

## Libraries

library(GGally)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

## Setup

a <- read.csv("../datasets/A.csv")  
dim(a)

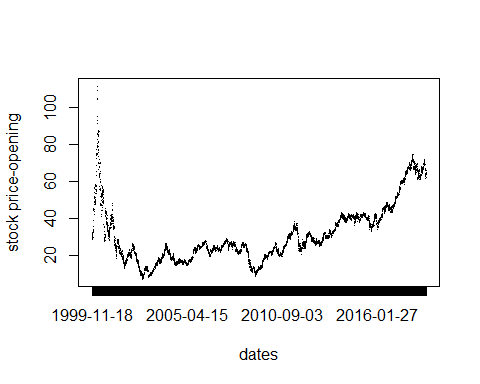
## [1] 4771 7

head(a, 5)

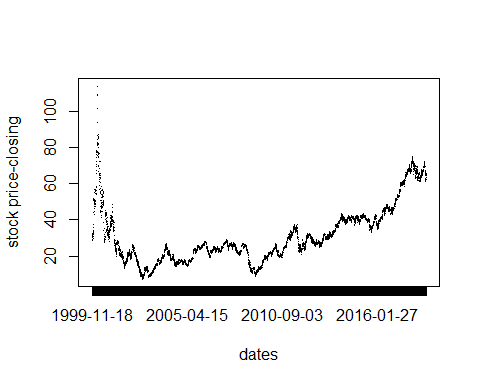
## date volume open close high low adjclose  
## 1 2018-11-02 2272100 65.80 65.93 66.61 65.44 65.93  
## 2 2018-11-01 1882400 65.07 65.22 65.34 64.58 65.22  
## 3 2018-10-31 3820700 64.35 64.79 65.98 64.19 64.79  
## 4 2018-10-30 3076700 62.35 63.64 63.72 61.68 63.64  
## 5 2018-10-29 2919900 63.65 62.08 64.17 61.13 62.08

## Plots

plot(a$date, a$open, xlab="dates", ylab="stock price-opening")



plot(a$date, a$close, xlab="dates", ylab="stock price-closing")



## Linear Regression

fit.1 <- lm(close ~ open, data=a)   
summary(fit.1)

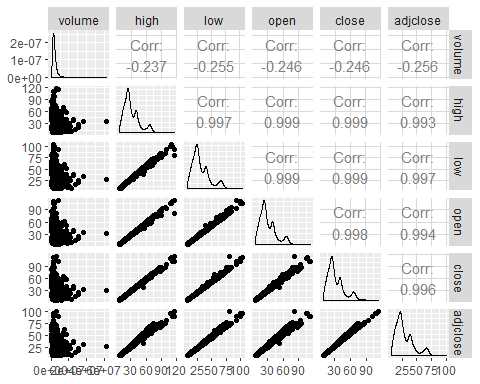
##   
## Call:  
## lm(formula = close ~ open, data = a)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.178 -0.274 0.005 0.280 33.008   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0862181 0.0320447 2.691 0.00716 \*\*   
## open 0.9976425 0.0009258 1077.617 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9788 on 4769 degrees of freedom  
## Multiple R-squared: 0.9959, Adjusted R-squared: 0.9959   
## F-statistic: 1.161e+06 on 1 and 4769 DF, p-value: < 2.2e-16

coef(fit.1)

## (Intercept) open   
## 0.08621812 0.99764254

## GG Pairs

a.gg <- a[, c("volume", "high", "low", "open", "close", "adjclose")]  
ggpairs(data=a.gg)



## Testing linear regression model accuracy

test.data <- a[1,c("close", "open")]  
pred.1 <- predict(fit.1, test.data, interval = "prediction", level = 0.95)  
summary(pred.1)

## fit lwr upr   
## Min. :65.73 Min. :63.81 Min. :67.65   
## 1st Qu.:65.73 1st Qu.:63.81 1st Qu.:67.65   
## Median :65.73 Median :63.81 Median :67.65   
## Mean :65.73 Mean :63.81 Mean :67.65   
## 3rd Qu.:65.73 3rd Qu.:63.81 3rd Qu.:67.65   
## Max. :65.73 Max. :63.81 Max. :67.65

# Subtracting actual value from the linear model prediction  
residual <- test.data$close - pred.1[,1]  
print(residual)

## [1] 0.1988997