

clothing

Chris Parrish

January 9, 2016

clothing

references:

- Cannon, et al., Stat2, chapter 03, example 3.17

Import the data.

```
data <- read.csv("Clothing.csv", header=TRUE)
head(data)
```

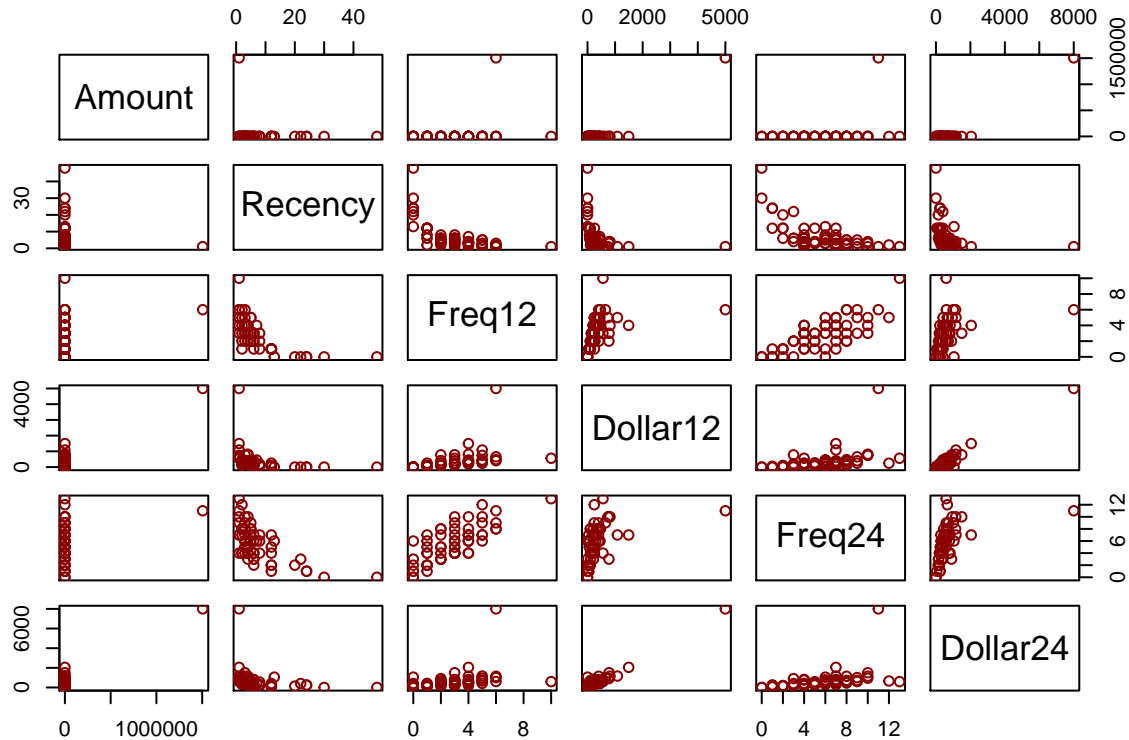
```
##   ID Amount Recency Freq12 Dollar12 Freq24 Dollar24 Card
## 1  1     0     22     0         0      3      400     0
## 2  2     0     30     0         0      0       0     0
## 3  3     0     24     0         0      1     250     0
## 4  4    30     6     3        140     4     225     0
## 5  5    33    12     1         50     1      50     0
## 6  6    35    48     0         0     0       0     0
```

```
dim(data)
```

```
## [1] 60 8
```

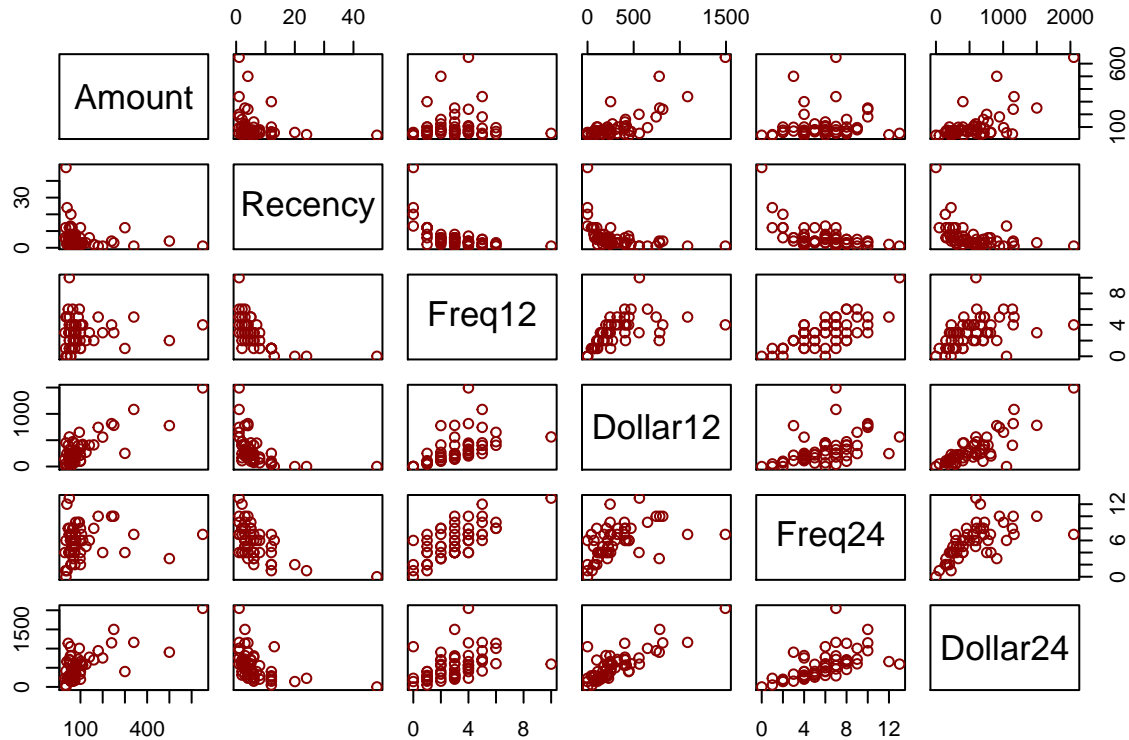
Scatterplot matrix.

```
pairs(~ Amount + Recency + Freq12 + Dollar12 + Freq24 + Dollar24, data=data, col="darkred")
```



Clean data.

```
data <- data[-c(1:3, 60), ] # remove 4 records
pairs(~ Amount + Recency + Freq12 + Dollar12 + Freq24 + Dollar24, data=data, col="darkred")
```



```
dim(data)
```

```
## [1] 56 8
```

Matrix of correlation coefficients.

```
with(data,
  round(cor(cbind(Amount, Recency, Freq12, Dollar12, Freq24, Dollar24)), 3))
```

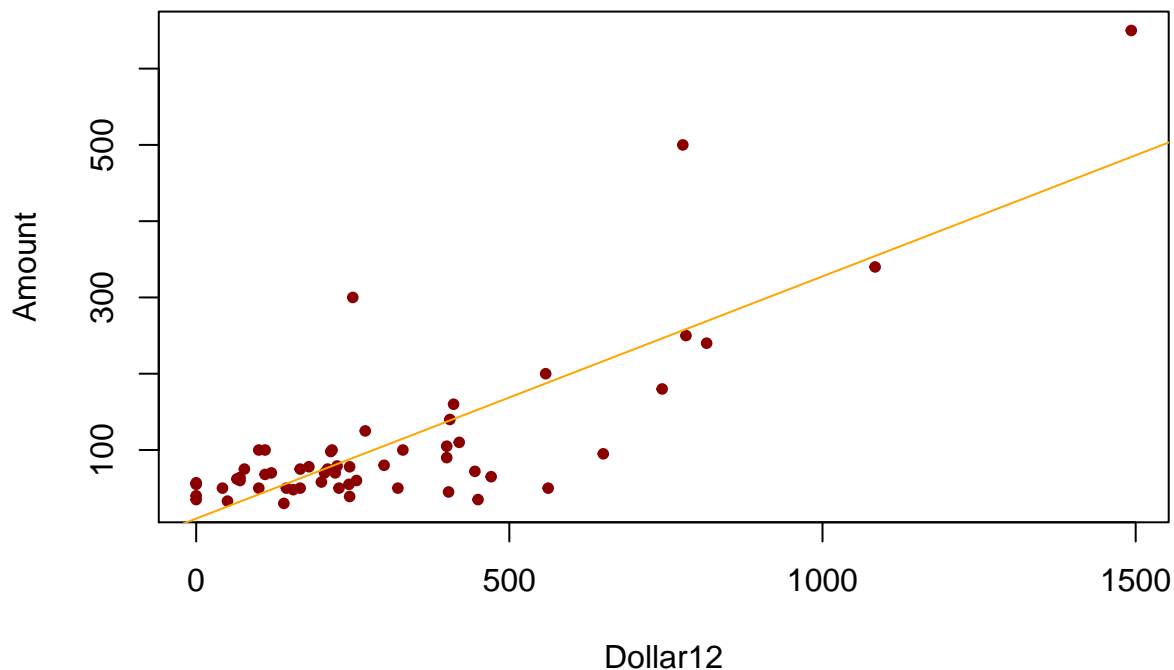
```
##           Amount Recency Freq12 Dollar12 Freq24 Dollar24
## Amount      1.000  -0.221  0.052   0.804  0.102   0.677
## Recency    -0.221   1.000 -0.584  -0.454 -0.549  -0.432
## Freq12      0.052  -0.584  1.000   0.556  0.710   0.421
## Dollar12    0.804  -0.454  0.556   1.000  0.485   0.827
## Freq24      0.102  -0.549  0.710   0.485  1.000   0.596
## Dollar24    0.677  -0.432  0.421   0.827  0.596   1.000
```

Simple linear regression.

```
clothing.lm1 <- lm(Amount ~ Dollar12, data=data)
options(show.signif.stars=FALSE)
summary(clothing.lm1)
```

```
##
## Call:
## lm(formula = Amount ~ Dollar12, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -138.54  -31.55   -3.85   25.34  243.18
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.0756    13.3783   0.753   0.455
## Dollar12      0.3176     0.0320   9.925 8.93e-14
##
## Residual standard error: 67.37 on 54 degrees of freedom
## Multiple R-squared:  0.6459, Adjusted R-squared:  0.6393
## F-statistic: 98.5 on 1 and 54 DF,  p-value: 8.929e-14
```

```
plot(Amount ~ Dollar12, data=data,
      pch=20, col="darkred")
abline(clothing.lm1, col="orange")
```



Multiple linear regression.

```
clothing.lm2 <- lm(Amount ~ Dollar12 + Dollar24 + Recency, data=data)
summary(clothing.lm2)
```

```
##
## Call:
## lm(formula = Amount ~ Dollar12 + Dollar24 + Recency, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -126.522 -24.098 0.247 23.652 237.852
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -23.05236 21.59290 -1.068 0.2906
## Dollar12 0.32724 0.05678 5.764 4.53e-07
## Dollar24 0.02151 0.04202 0.512 0.6110
## Recency 2.86718 1.37573 2.084 0.0421
##
## Residual standard error: 65.91 on 52 degrees of freedom
## Multiple R-squared: 0.6736, Adjusted R-squared: 0.6548
## F-statistic: 35.78 on 3 and 52 DF, p-value: 1.097e-12
```

Full model.

```
clothing.lm3 <- lm(Amount ~ Freq12 + Dollar12 + Freq24 + Dollar24 + Recency + Card, data=data)
summary(clothing.lm3)
```

```
##
## Call:
## lm(formula = Amount ~ Freq12 + Dollar12 + Freq24 + Dollar24 +
## Recency + Card, data = data)
##
## Residuals:
## Min 1Q Median 3Q Max
## -63.799 -12.218 -3.334 7.299 156.822
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 104.251935 19.834341 5.256 3.20e-06
## Freq12 -32.353539 5.187870 -6.236 1.01e-07
## Dollar12 0.429683 0.041325 10.398 5.43e-14
## Freq24 -5.173593 3.619661 -1.429 0.159
## Dollar24 0.001756 0.031850 0.055 0.956
## Recency -1.345963 0.971053 -1.386 0.172
## Card 14.624409 14.575770 1.003 0.321
##
## Residual standard error: 40.83 on 49 degrees of freedom
## Multiple R-squared: 0.882, Adjusted R-squared: 0.8675
## F-statistic: 61.02 on 6 and 49 DF, p-value: < 2.2e-16
```

Another model, balancing simplicity and explanatory power.

```
clothing.lm4 <- lm(Amount ~ Dollar12 + Freq12, data=data)
summary(clothing.lm4)
```

```
##
## Call:
## lm(formula = Amount ~ Dollar12 + Freq12, data = data)
##
## Residuals:
## Min 1Q Median 3Q Max
```

```
## -67.161 -17.147 -1.061 11.434 150.629
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  73.89763   10.46860    7.059 3.62e-09
## Dollar12     0.44315    0.02337   18.959 < 2e-16
## Freq12      -34.42587    3.56139   -9.666 2.72e-13
##
## Residual standard error: 40.91 on 53 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.867
## F-statistic: 180.3 on 2 and 53 DF,  p-value: < 2.2e-16
```

Create a new regressor.

```
data <- data[data$Freq12 != 0, ]
dim(data)
```

```
## [1] 52  8
```

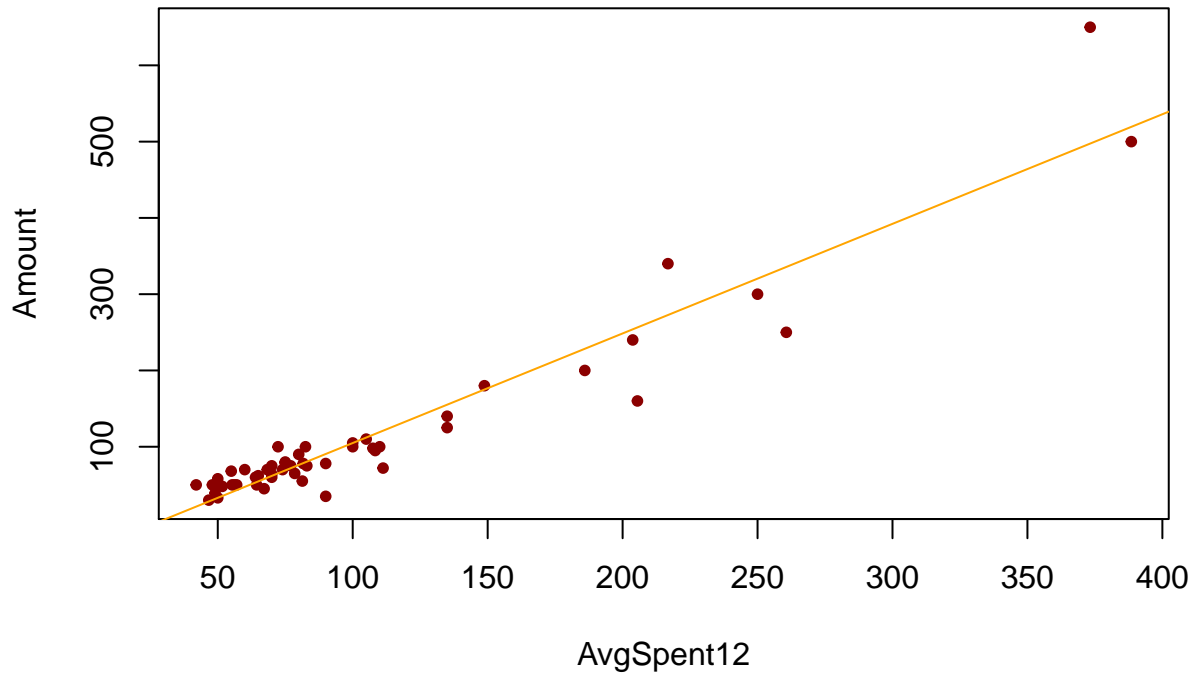
```
data$AvgSpent12 <- with(data, Dollar12 / Freq12)
head(data)
```

```
##   ID Amount Recency Freq12 Dollar12 Freq24 Dollar24 Card AvgSpent12
## 4   4     30      6      3     140      4     225    0  46.66667
## 5   5     33     12      1      50      1      50    0  50.00000
## 7   7     35      5      5     450      6     415    0  90.00000
## 8   8     39      2      5     245     12     661    1  49.00000
## 10 10     45      3      6     403      8    1138    0  67.16667
## 11 11     48      6      3     155      4     262    0  51.66667
```

```
clothing.lm5 <- lm(Amount ~ AvgSpent12, data=data)
summary(clothing.lm5)
```

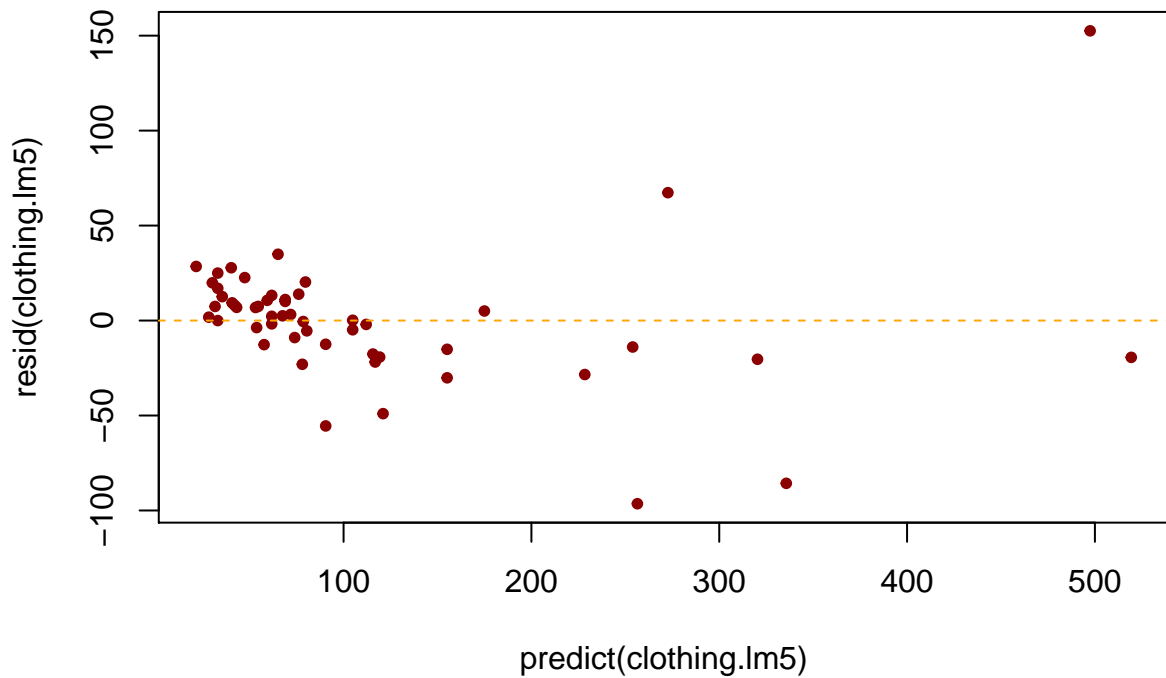
```
##
## Call:
## lm(formula = Amount ~ AvgSpent12, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -96.439 -14.230   2.011  11.446 152.536
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -38.8254     8.3438  -4.653 2.43e-05
## AvgSpent12   1.4368     0.0642  22.380 < 2e-16
##
## Residual standard error: 35.02 on 50 degrees of freedom
## Multiple R-squared:  0.9092, Adjusted R-squared:  0.9074
## F-statistic: 500.9 on 1 and 50 DF,  p-value: < 2.2e-16
```

```
plot(Amount ~ AvgSpent12, data=data,  
     pch=20, col="darkred")  
abline(clothing.lm5, col="orange")
```



Residuals for the new regressor.

```
plot(predict(clothing.lm5), resid(clothing.lm5),  
     pch=20, col="darkred")  
abline(h=0, col="orange", lty="dashed")
```

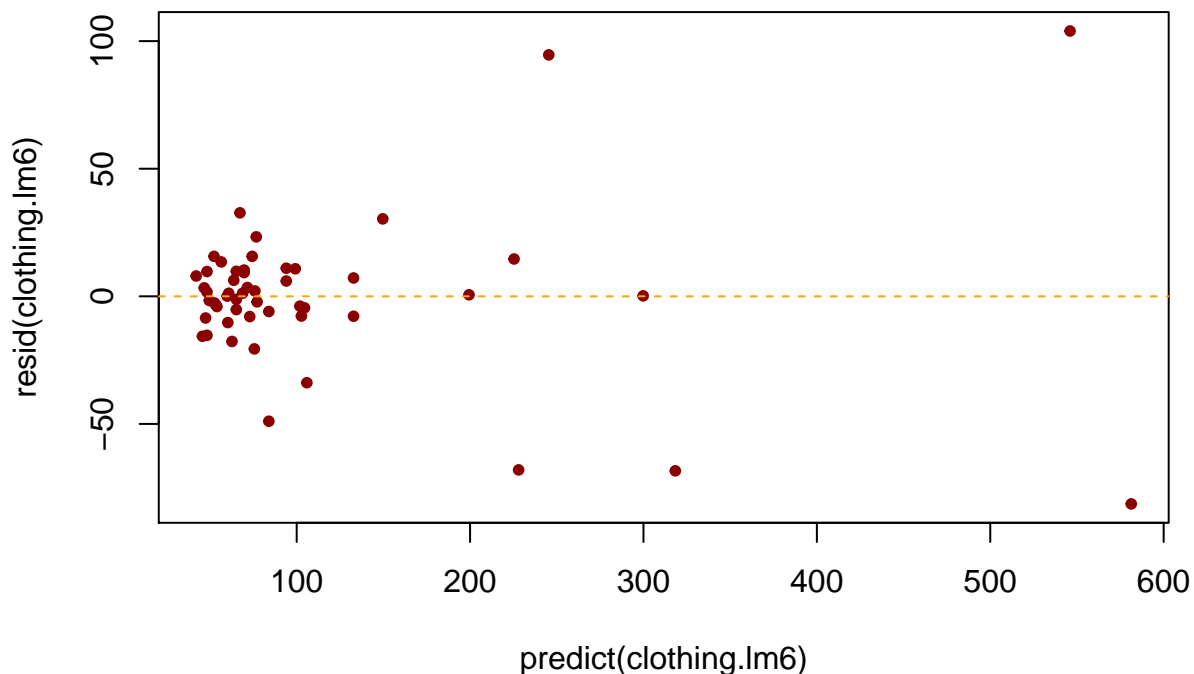


Add a quadratic term.

```
clothing.lm6 <- lm(Amount ~ AvgSpent12 + I(AvgSpent12^2), data=data)
summary(clothing.lm6)
```

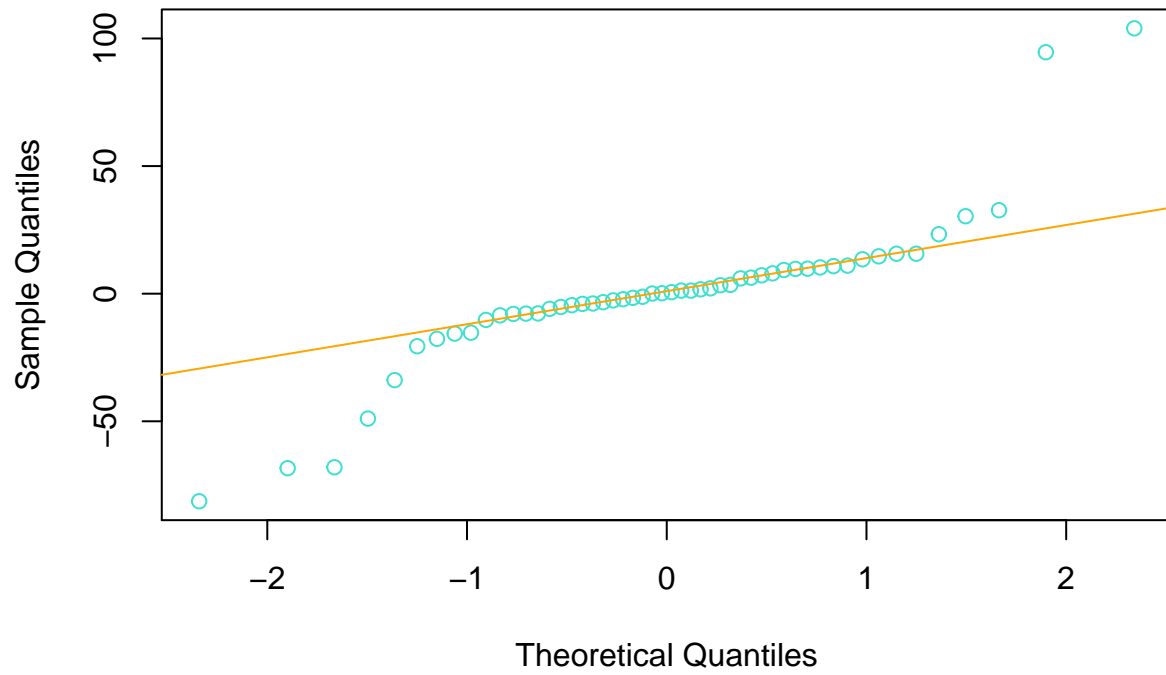
```
##
## Call:
## lm(formula = Amount ~ AvgSpent12 + I(AvgSpent12^2), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -81.332  -7.752   0.389   9.734 103.968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.402e+01  1.457e+01   0.963  0.34046
## AvgSpent12   5.709e-01  2.145e-01   2.661  0.01050
## I(AvgSpent12^2) 2.289e-03  5.477e-04   4.180  0.00012
##
## Residual standard error: 30.37 on 49 degrees of freedom
## Multiple R-squared:  0.9331, Adjusted R-squared:  0.9304
## F-statistic: 341.7 on 2 and 49 DF,  p-value: < 2.2e-16
```

```
plot(predict(clothing.lm6), resid(clothing.lm6),
      pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```



```
qqnorm(resid(clothing.lm6),
       col="turquoise")
qqline(resid(clothing.lm6),
       col="orange")
```

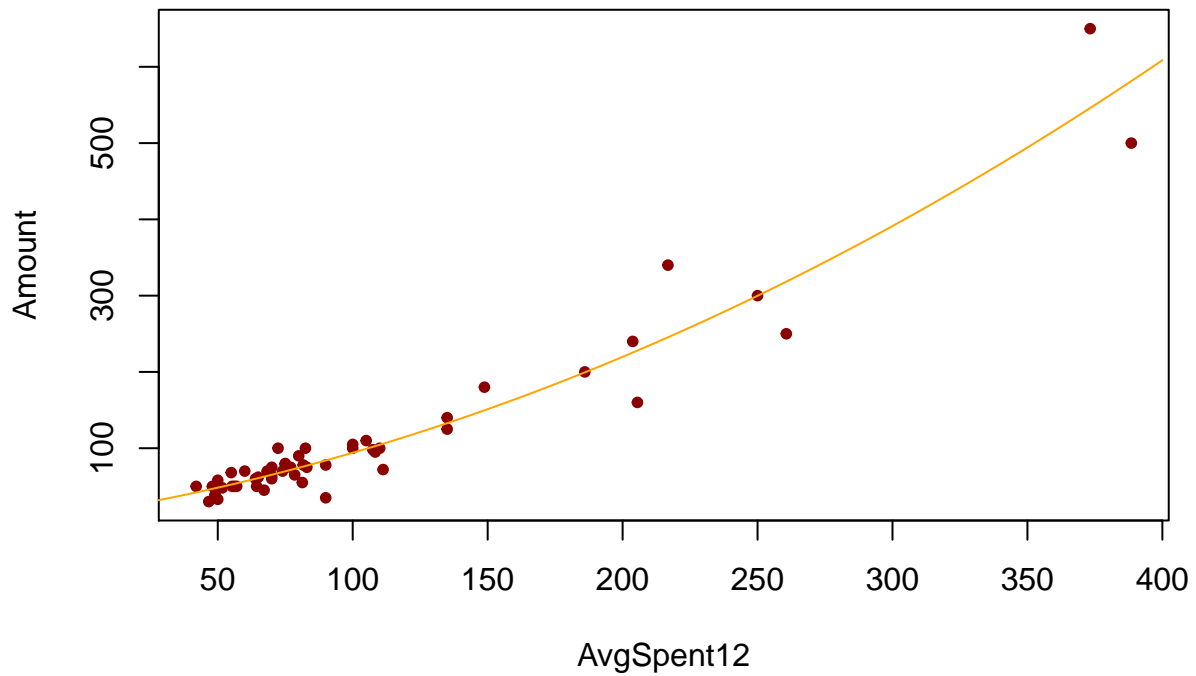
Normal Q-Q Plot



Final model.

$$\widehat{Amount} = 14.022 + 0.571 \text{ AvgSpent12} + 0.002 \text{ AvgSpent12}^2$$

```
plot(Amount ~ AvgSpent12, data=data,
     pch=20, col="darkred")
amount <- function(avgSpent12){
  a <- 14.02
  b <- 0.5709
  c <- 0.002289
  amt <- a + b * avgSpent12 + c * avgSpent12^2
  return(amt)
}
curve(amount, from=0, to=400,
      col="orange", add=TRUE)
```

Effect plot.

```
library(alr4)  
plot(effect("AvgSpent12", clothing.lm6))
```

AvgSpent12 effect plot

