Predictive Analytics homework - Conjoint Analysis (Preferences for Mobile Communication Services)

Godfred Somua-Gyimah

January 10, 2017

Contents

Note: In order to run this demo, the following R packages must be installed in your R environment:

- ggplot2: visualization
- conjoint: marketing conjoint analysis

1. Read in Dataset

```
# Clean the environment
rm(list = ls())
# Read data file
df <- read.csv("mobile_services_ranking.csv")
```

2. Understand Data

```
# Show head
head(df)
```

```
##           brand startup monthly  service       retail       apple
## 1        "AT&T"  "$100"  "$100"  "4G NO"  "Retail NO"  "Apple NO"
## 2     "Verizon"  "$300"  "$100"  "4G NO" "Retail YES" "Apple YES"
## 3 "US Cellular"  "$400"  "$200"  "4G NO"  "Retail NO"  "Apple NO"
## 4     "Verizon"  "$400"  "$400" "4G YES" "Retail YES"  "Apple NO"
## 5     "Verizon"  "$200"  "$300"  "4G NO"  "Retail NO"  "Apple NO"
## 6     "Verizon"  "$100"  "$200" "4G YES"  "Retail NO" "Apple YES"
##         samsung      google ranking
## 1  "Samsung NO"  "Nexus NO"      11
## 2 "Samsung YES"  "Nexus NO"      12
## 3 "Samsung YES"  "Nexus NO"       9
## 4  "Samsung NO"  "Nexus NO"       2
## 5 "Samsung YES" "Nexus YES"       8
## 6  "Samsung NO" "Nexus YES"      13
```

```
# Show the structure of the data frame
str(df)
```

```
## 'data.frame':    16 obs. of  9 variables:
##  $ brand  : Factor w/ 4 levels "\"AT&T\"","\"T-Mobile\"",..: 1 4 3 4 4 4 3 1 1 2 ...
##  $ startup: Factor w/ 4 levels "\"$100\"","\"$200\"",..: 1 3 4 4 2 1 3 4 2 4 ...
##  $ monthly: Factor w/ 4 levels "\"$100\"","\"$200\"",..: 1 1 2 4 3 2 3 3 4 1 ...
##  $ service: Factor w/ 2 levels "\"4G NO\"","\"4G YES\"": 1 1 1 2 1 2 2 1 2 2 ...
##  $ retail : Factor w/ 2 levels "\"Retail NO\"",..: 1 2 1 2 1 1 1 2 1 1 ...
##  $ apple  : Factor w/ 2 levels "\"Apple NO\"",..: 1 2 1 1 1 2 2 2 2 2 ...
##  $ samsung: Factor w/ 2 levels "\"Samsung NO\"",..: 1 2 2 1 2 1 1 1 2 2 ...
##  $ google : Factor w/ 2 levels "\"Nexus NO\"",..: 1 1 1 1 2 2 1 2 1 2 ...
##  $ ranking: int  11 12 9 2 8 13 7 4 5 16 ...
```

```
# Show summary statistics
summary(df)
```

```
##           brand     startup    monthly      service           retail
##  "AT&T"       :4   "$100":4   "$100":4   "4G NO" :8   "Retail NO" :8
##  "T-Mobile"   :4   "$200":4   "$200":4   "4G YES":8   "Retail YES":8
##  "US Cellular":4   "$300":4   "$300":4
##  "Verizon"    :4   "$400":4   "$400":4
##
##
##          apple            samsung          google     ranking
##  "Apple NO" :8   "Samsung NO" :8   "Nexus NO" :8   Min.   : 1.00
```

```
##  "Apple YES":8   "Samsung YES":8   "Nexus YES":8   1st Qu.: 4.75
##                                                   Median : 8.50
##                                                   Mean   : 8.50
##                                                   3rd Qu.:12.25
##                                                   Max.   :16.00
```

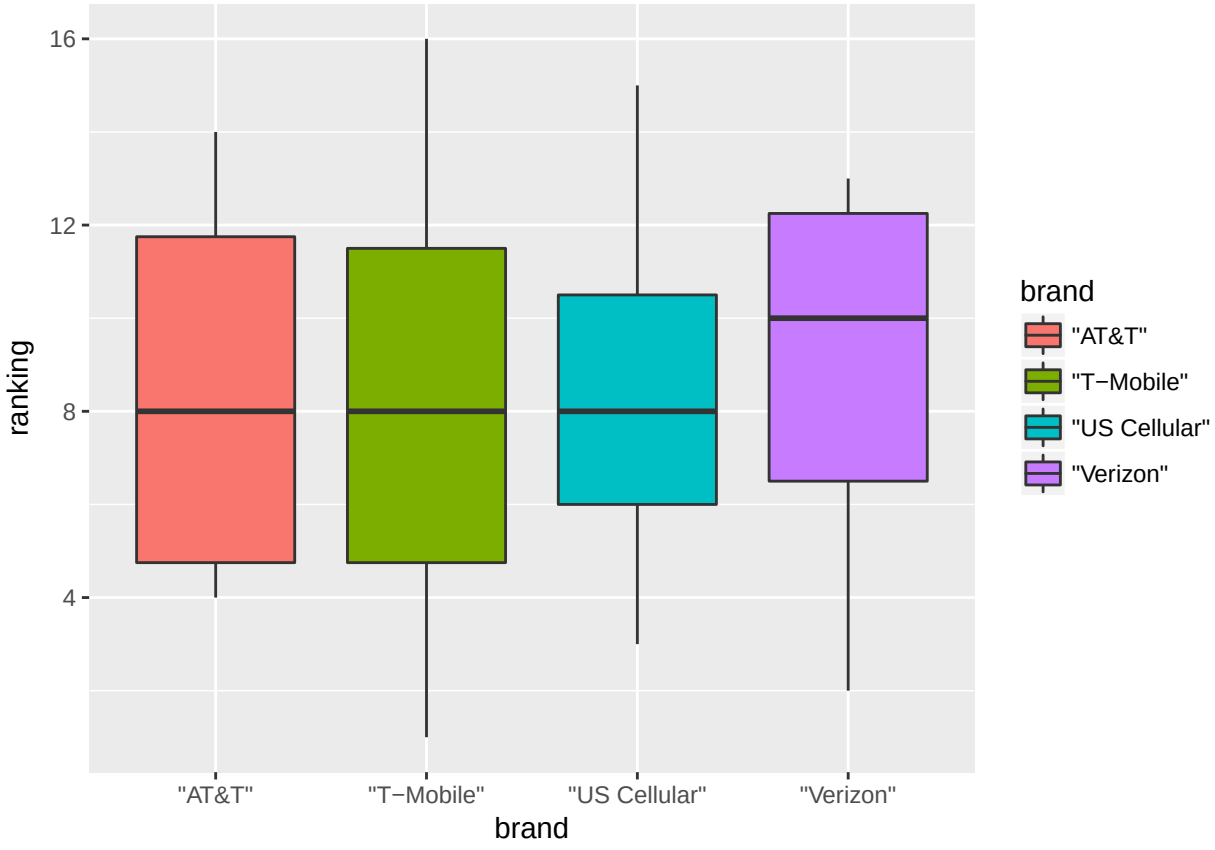
3. Exploratory Data Analysis

We can draw a couple of plots to explore the data.

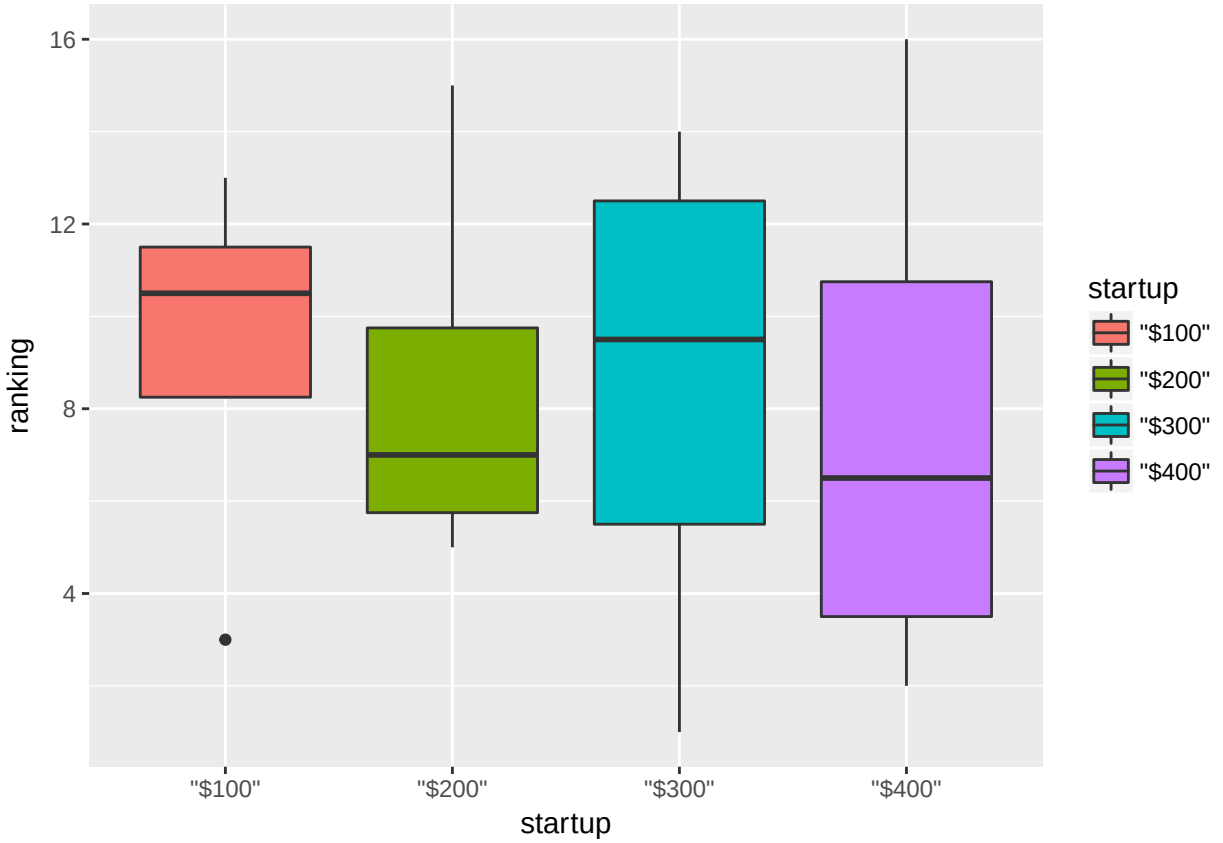
First, load ggplot2 package.

```
# Load ggplot2 package for plotting
library(ggplot2)
```

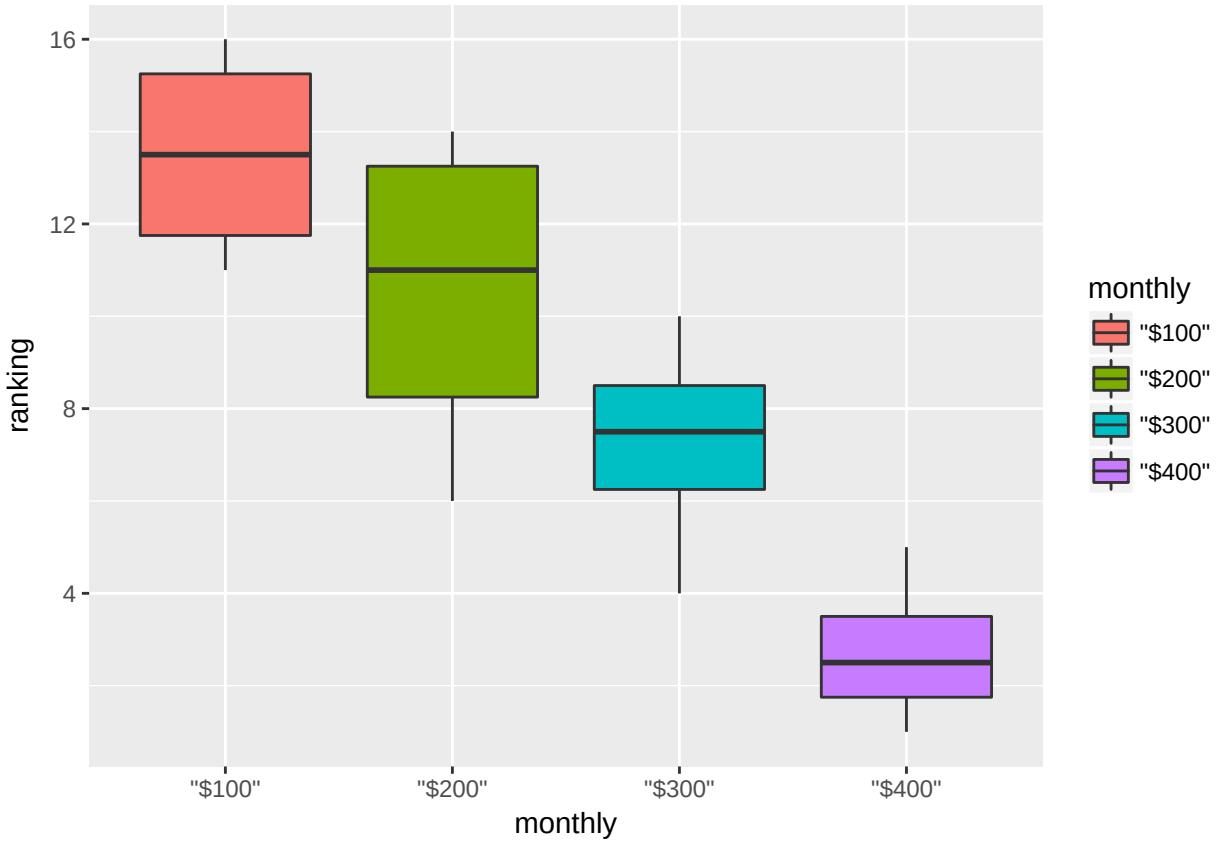
```
# Draw box plot of ranking by brand
ggplot(df,aes(x=brand, y=ranking, fill=brand)) +
  geom_boxplot()
```




```
# Draw box plot of ranking by startup cost
ggplot(df,aes(x=startup, y=ranking, fill=startup)) +
  geom_boxplot()
```




```
# Draw box plot of ranking by monthly cost
ggplot(df,aes(x=monthly, y=ranking, fill=monthly)) +
  geom_boxplot()
```



From the above box plot, we can see it clearly that ranking is higher with lower monthly cost.

4. Regression Analysis of the Main Effect Model

```
# Fit linear regression model using main effects only (no interaction terms)
fit <- lm(ranking ~.,data= df)

print(summary(fit))
```

```
##
## Call:
## lm(formula = ranking ~ ., data = df)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10
## -0.125  0.125  0.125 -0.125 -0.125  0.125 -0.125  0.125  0.125 -0.125
##     11     12     13     14     15     16
## -0.125 -0.125  0.125  0.125  0.125 -0.125
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.112e+01  4.841e-01  22.980   0.0277 *
## brand"T-Mobile"        -2.500e-01  3.536e-01  -0.707   0.6082
## brand"US Cellular"     -6.629e-16  3.536e-01   0.000   1.0000
```

```
## brand"Verizon"        2.500e-01  3.536e-01   0.707   0.6082
## startup"$200"        -7.500e-01  3.536e-01  -2.121   0.2804
## startup"$300"        -7.500e-01  3.536e-01  -2.121   0.2804
## startup"$400"        -1.500e+00  3.536e-01  -4.243   0.1474
## monthly"$200"        -3.000e+00  3.536e-01  -8.485   0.0747 .
## monthly"$300"        -6.250e+00  3.536e-01 -17.678   0.0360 *
## monthly"$400"        -1.075e+01  3.536e-01 -30.406   0.0209 *
## service"4G YES"       3.500e+00  2.500e-01  14.000   0.0454 *
## retail"Retail YES"   -5.000e-01  2.500e-01  -2.000   0.2952
## apple"Apple YES"     -5.000e-01  2.500e-01  -2.000   0.2952
## samsung"Samsung YES"  2.250e+00  2.500e-01   9.000   0.0704 .
## google"Nexus YES"     1.500e+00  2.500e-01   6.000   0.1051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5 on 1 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.989
## F-statistic: 97.07 on 14 and 1 DF,  p-value: 0.0794
```

```
# Show coefficient rounding to the nearest hundredth
formatC(fit$coefficients[-1], format = 'f', digits = 2)
```

```
##      brand"T-Mobile"   brand"US Cellular"       brand"Verizon"
##              "-0.25"              "-0.00"               "0.25"
##        startup"$200"        startup"$300"        startup"$400"
##              "-0.75"              "-0.75"              "-1.50"
##        monthly"$200"        monthly"$300"        monthly"$400"
##              "-3.00"              "-6.25"             "-10.75"
##      service"4G YES"   retail"Retail YES"     apple"Apple YES"
##               "3.50"              "-0.50"              "-0.50"
## samsung"Samsung YES"    google"Nexus YES"
##               "2.25"               "1.50"
```

5. Traditional Conjoint Analysis

```
# Load the conjoint package
library(conjoint)
```

```
## Warning: package 'conjoint' was built under R version 3.3.3
```

```
##
## This is package 'modeest' written by P. PONCET.
## For a complete list of functions, use 'library(help = "modeest")' or 'help.start()'.
```

5.1. Prepare Data

```
# Create levels
level = c(levels(df$brand),
          levels(df$startup),
          levels(df$monthly),
          levels(df$service),
          levels(df$retail),
```

```
          levels(df$apple),
          levels(df$samsung),
          levels(df$google))

# Remove double quote in the levels
level <- gsub('"','',level)

level
```

```
##  [1] "AT&T"        "T-Mobile"    "US Cellular" "Verizon"     "$100"
##  [6] "$200"        "$300"        "$400"        "$100"        "$200"
## [11] "$300"        "$400"        "4G NO"       "4G YES"      "Retail NO"
## [16] "Retail YES"  "Apple NO"    "Apple YES"   "Samsung NO"  "Samsung YES"
## [21] "Nexus NO"    "Nexus YES"
```

```
# Create profiles
profiles <- df[1:8]

profiles
```

```
##            brand startup monthly  service       retail       apple
## 1          "AT&T"  "$100"  "$100"  "4G NO"  "Retail NO"  "Apple NO"
## 2       "Verizon"  "$300"  "$100"  "4G NO" "Retail YES" "Apple YES"
## 3  "US Cellular"  "$400"  "$200"  "4G NO"  "Retail NO"  "Apple NO"
## 4       "Verizon"  "$400"  "$400" "4G YES" "Retail YES"  "Apple NO"
## 5       "Verizon"  "$200"  "$300"  "4G NO"  "Retail NO"  "Apple NO"
## 6       "Verizon"  "$100"  "$200" "4G YES"  "Retail NO" "Apple YES"
## 7  "US Cellular"  "$300"  "$300" "4G YES"  "Retail NO" "Apple YES"
## 8          "AT&T"  "$400"  "$300"  "4G NO" "Retail YES" "Apple YES"
## 9          "AT&T"  "$200"  "$400" "4G YES"  "Retail NO" "Apple YES"
## 10    "T-Mobile"  "$400"  "$100" "4G YES"  "Retail NO" "Apple YES"
## 11 "US Cellular"  "$100"  "$400"  "4G NO" "Retail YES" "Apple YES"
## 12    "T-Mobile"  "$200"  "$200"  "4G NO" "Retail YES" "Apple YES"
## 13    "T-Mobile"  "$100"  "$300" "4G YES" "Retail YES"  "Apple NO"
## 14 "US Cellular"  "$200"  "$100" "4G YES" "Retail YES"  "Apple NO"
## 15    "T-Mobile"  "$300"  "$400"  "4G NO"  "Retail NO"  "Apple NO"
## 16         "AT&T"  "$300"  "$200" "4G YES" "Retail YES"  "Apple NO"
##           samsung      google
## 1   "Samsung NO"  "Nexus NO"
## 2  "Samsung YES"  "Nexus NO"
## 3  "Samsung YES"  "Nexus NO"
## 4   "Samsung NO"  "Nexus NO"
## 5  "Samsung YES" "Nexus YES"
## 6   "Samsung NO" "Nexus YES"
## 7   "Samsung NO"  "Nexus NO"
## 8   "Samsung NO" "Nexus YES"
## 9  "Samsung YES"  "Nexus NO"
## 10 "Samsung YES" "Nexus YES"
## 11 "Samsung YES" "Nexus YES"
## 12  "Samsung NO"  "Nexus NO"
## 13 "Samsung YES"  "Nexus NO"
## 14  "Samsung NO" "Nexus YES"
## 15  "Samsung NO" "Nexus YES"
## 16 "Samsung YES" "Nexus YES"
```

```
# Create ratings of respondent
ratings = df[9]

head(ratings)
```

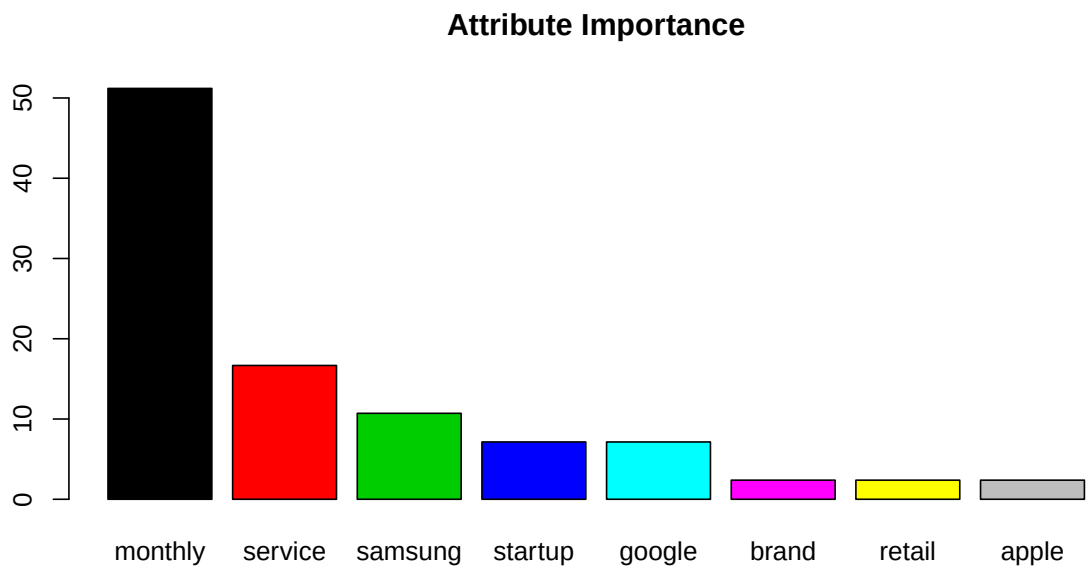
```
##   ranking
## 1      11
## 2      12
## 3       9
## 4       2
## 5       8
## 6      13
```

5.2. Calculate Attribute Importance

```
# Calculate the importance of each attribute
im = caImportance(y=ratings, x=profiles)
names(im) = names(profiles)
im
```

```
##   brand startup monthly service  retail   apple samsung  google
##    2.38    7.14   51.19   16.67    2.38    2.38   10.71    7.14
```

```
barplot(sort(im, decreasing = TRUE),col=1:length(im), main = "Attribute Importance")
```



5.3. Calculate Part Worths

```
pw <- caUtilities(y=ratings, x=profiles, z=level)
```

```
##
## Call:
## lm(formula = frml)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10
## -0,125  0,125  0,125 -0,125 -0,125  0,125 -0,125  0,125  0,125 -0,125
##     11     12     13     14     15     16
## -0,125 -0,125  0,125  0,125  0,125 -0,125
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          8,500e+00  1,250e-01  68,000  0,00936 **
## factor(x$brand)1     1,374e-16  2,165e-01   0,000  1,00000
## factor(x$brand)2    -2,500e-01  2,165e-01  -1,155  0,45437
## factor(x$brand)3    -1,202e-16  2,165e-01   0,000  1,00000
## factor(x$startup)1   7,500e-01  2,165e-01   3,464  0,17891
## factor(x$startup)2   8,240e-16  2,165e-01   0,000  1,00000
## factor(x$startup)3  -2,794e-16  2,165e-01   0,000  1,00000
## factor(x$monthly)1   5,000e+00  2,165e-01  23,094  0,02755 *
## factor(x$monthly)2   2,000e+00  2,165e-01   9,238  0,06865 .
## factor(x$monthly)3  -1,250e+00  2,165e-01  -5,774  0,10918
## factor(x$service)1  -1,750e+00  1,250e-01 -14,000  0,04540 *
## factor(x$retail)1    2,500e-01  1,250e-01   2,000  0,29517
## factor(x$apple)1     2,500e-01  1,250e-01   2,000  0,29517
## factor(x$samsung)1  -1,125e+00  1,250e-01  -9,000  0,07045 .
## factor(x$google)1   -7,500e-01  1,250e-01  -6,000  0,10514
## ---
## Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
##
## Residual standard error: 0,5 on 1 degrees of freedom
## Multiple R-squared:  0,9993, Adjusted R-squared:  0,989
## F-statistic: 97,07 on 14 and 1 DF,  p-value: 0,0794
```

```
names(pw) = c('intercept',level)

# Show part worths rounding to the nearest hundredth
formatC(pw[-1], format = 'f', digits = 2)  # Do not show intercept
```

```
##        AT&T    T-Mobile US Cellular     Verizon        $100        $200
##      "0.00"     "-0.25"     "-0.00"      "0.25"      "0.75"      "0.00"
##        $300        $400        $100        $200        $300        $400
##     "-0.00"     "-0.75"      "5.00"      "2.00"     "-1.25"     "-5.75"
##       4G NO      4G YES   Retail NO  Retail YES    Apple NO   Apple YES
##     "-1.75"      "1.75"      "0.25"     "-0.25"      "0.25"     "-0.25"
##  Samsung NO Samsung YES    Nexus NO   Nexus YES
##     "-1.12"      "1.12"     "-0.75"      "0.75"
```

```
# Show coefficient rounding to the nearest hundredth
formatC(fit$coefficients[-1], format = 'f', digits = 2)
```

```
##      brand"T-Mobile"   brand"US Cellular"       brand"Verizon"
##              "-0.25"              "-0.00"               "0.25"
##        startup"$200"        startup"$300"        startup"$400"
##              "-0.75"              "-0.75"              "-1.50"
```

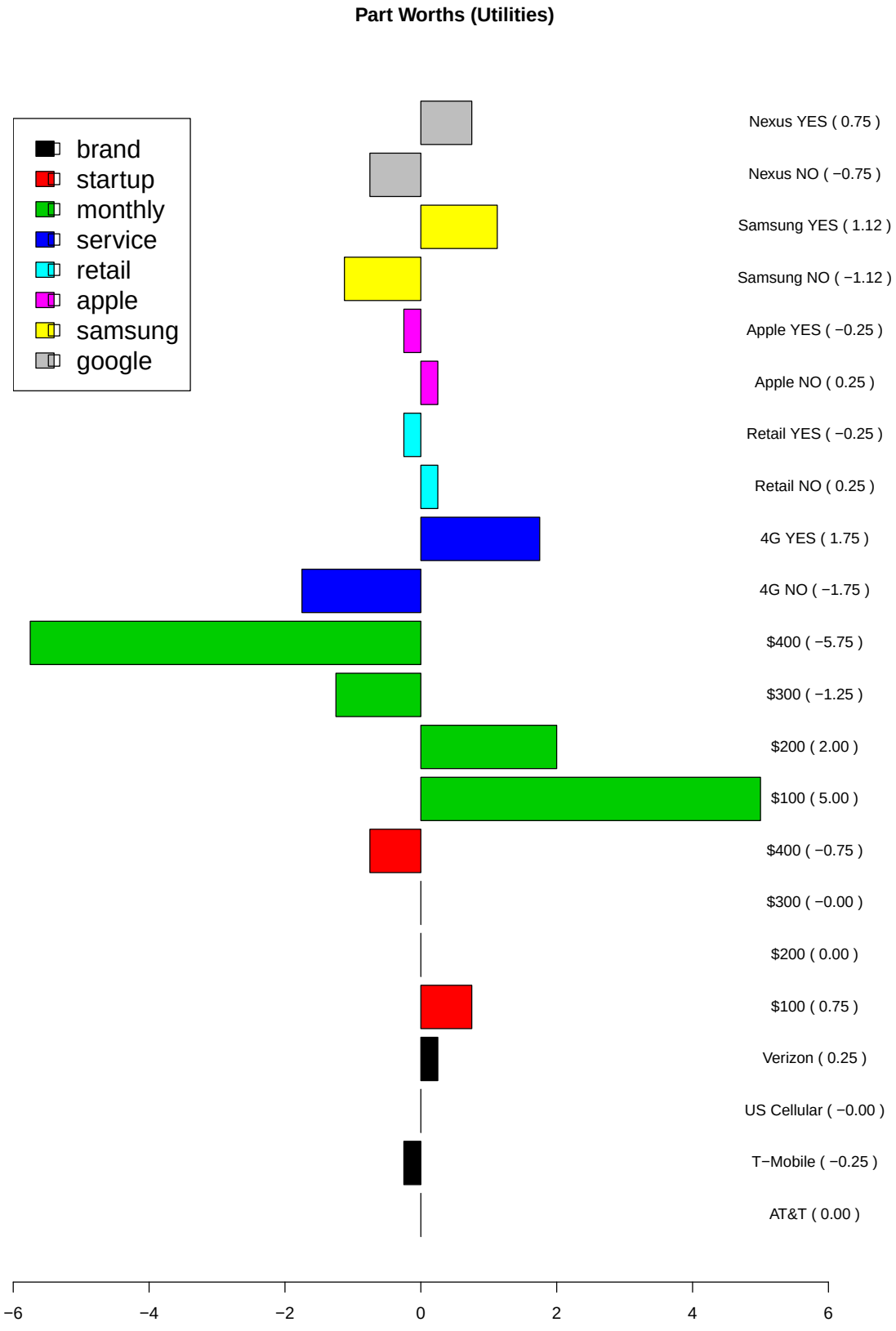
```
##        monthly"$200"        monthly"$300"        monthly"$400"
##              "-3.00"              "-6.25"             "-10.75"
##      service"4G YES"   retail"Retail YES"     apple"Apple YES"
##               "3.50"              "-0.50"              "-0.50"
## samsung"Samsung YES"    google"Nexus YES"
##               "2.25"               "1.50"
```

Compare the above part worths and OLS coefficients, can you find out their relationship?

```
#  Create a vector that assigns the same color value to levels under a same attribute
cols = c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,7,8,8)

# Draw barplot of part worths
bp <- barplot(pw[-1],las = 1, horiz = TRUE, xlim = c(-6, 7),axisnames = FALSE,
        col = cols, main = "Part Worths (Utilities)")

text(x= 5.8, y= bp, cex = 0.9,
     labels= paste(level,
                   "(",
                   formatC(pw[-1], format = 'f', digits = 2),
                   ')'))
# Add legend
legend(-6, 26,names(profiles), pch =0, cex = 1.5, fill = 1:8)
```


Part Worths (Utilities)
brand
startup
monthly
service
retail
apple
samsung
google
Nexus YES (0.75)
Nexus NO (−0.75)
Samsung YES (1.12)
Samsung NO (−1.12)
Apple YES (−0.25)
Apple NO (0.25)
Retail YES (−0.25)
Retail NO (0.25)
4G YES (1.75)
4G NO (−1.75)
$400 (−5.75)
$300 (−1.25)
$200 (2.00)
$100 (5.00)
$400 (−0.75)
$300 (−0.00)
$200 (0.00)
$100 (0.75)
Verizon (0.25)
US Cellular (−0.00)
T−Mobile (−0.25)
AT&T (0.00)
−6
−4
−2
0
2
4
6