Coffee Sales

Claudia Clinchard

2025-08-29

```
knitr::opts_chunk$set(echo = TRUE)
setwd("~/Desktop/Projects/Coffee_sales")
#load necessary libraries
library(readxl)
library(Hmisc)
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:Hmisc':
##
##     src, summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(leaps)
library(ggplot2)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
#read in data and examine
coffee <- read_excel("Coffee_sales.xlsx") |>
  mutate(across(where(is.character), as.factor))
head(coffee)
```

```
## # A tibble: 6 × 12
##   date                datetime            hour_of_day cash_type card
money
##   <dttm>              <dttm>                    <dbl> <fct>     <fct>
<dbl>
## 1 2024-03-01 00:00:00 2024-03-01 10:15:50          10 card      ANON-0000…
38.7
## 2 2024-03-01 00:00:00 2024-03-01 12:19:22          12 card      ANON-0000…
38.7
## 3 2024-03-01 00:00:00 2024-03-01 12:20:18          12 card      ANON-0000…
38.7
## 4 2024-03-01 00:00:00 2024-03-01 13:46:33          13 card      ANON-0000…
28.9
## 5 2024-03-01 00:00:00 2024-03-01 13:48:14          13 card      ANON-0000…
38.7
## 6 2024-03-01 00:00:00 2024-03-01 15:39:47          15 card      ANON-0000…
33.8
## # ℹ 6 more variables: coffee_name <fct>, Time_of_Day <fct>, Weekday <fct>
,
## #   Month_name <fct>, Weekdaysort <dbl>, Monthsort <dbl>
```

```r
colnames(coffee)
```

```
##  [1] "date"        "datetime"    "hour_of_day" "cash_type"   "card"
##  [6] "money"       "coffee_name" "Time_of_Day" "Weekday"     "Month_name"
## [11] "Weekdaysort" "Monthsort"
```

```r
Hmisc::describe(coffee) #descriptives
```

```
## coffee
##
##  12  Variables      3636  Observations
## ---------------------------------------------------------------------------
------
## date
##          n    missing   distinct       Info       Mean    pMedian        G
md
##       3636          0        381          1 2024-09-30  1.728e+09   109880
19
##        .05        .10        .25        .50        .75        .90        .
95
## 2024-03-26 2024-04-24 2024-07-03 2024-10-07 2025-01-08 2025-02-26 2025-03-
11
##
## lowest : 2024-03-01 2024-03-02 2024-03-03 2024-03-04 2024-03-05
## highest: 2025-03-19 2025-03-20 2025-03-21 2025-03-22 2025-03-23
```

```
## ------------------------------------------------------------------------
------
## datetime
##                    n              missing            distinct
Info
##                 3636                    0                3636
1
##                 Mean              pMedian                  Gmd
.05
## 2024-10-01 02:35:30           1.728e+09            10987103 2024-03-26 18:
34:35
##                  .10                 .25                 .50
.75
## 2024-04-24 14:39:08 2024-07-03 16:54:06 2024-10-07 02:55:12 2025-01-08 07:
55:20
##                  .90                 .95
## 2025-02-26 15:18:43 2025-03-11 10:58:33
##
## lowest : 2024-03-01 10:15:50 2024-03-01 12:19:22 2024-03-01 12:20:18 2024-
03-01 13:46:33 2024-03-01 13:48:14
## highest: 2025-03-23 10:34:54 2025-03-23 14:43:37 2025-03-23 14:44:16 2025-
03-23 15:47:28 2025-03-23 18:11:38
## ------------------------------------------------------------------------
------
## hour_of_day
##        n  missing distinct     Info     Mean  pMedian      Gmd      .05
##     3636        0       17    0.995    14.17       14    4.863        8
##      .10      .25      .50      .75      .90      .95
##        9       10       14       18       20       21
##
## Value          6     7     8     9    10    11    12    13    14    15
16
## Frequency      5    91   235   247   349   294   249   227   228   245   2
82
## Proportion 0.001 0.025 0.065 0.068 0.096 0.081 0.068 0.062 0.063 0.067 0.0
78
##
## Value         17    18    19    20    21    22
## Frequency    245   220   235   171   197   116
## Proportion 0.067 0.061 0.065 0.047 0.054 0.032
##
## For the frequency table, variable is rounded to the nearest 0
## ------------------------------------------------------------------------
------
## cash_type
##        n  missing distinct
##     3636        0        2
##
## Value       card  cash
## Frequency   3547    89
```

```
## Proportion 0.976 0.024
## ---------------------------------------------------------------------------
------
## card
##        n  missing distinct
##     3547       89     1316
##
## lowest : ANON-0000-0000-0001 ANON-0000-0000-0002 ANON-0000-0000-0003 ANON-
0000-0000-0004 ANON-0000-0000-0005
## highest: ANON-0000-0000-1312 ANON-0000-0000-1313 ANON-0000-0000-1314 ANON-
0000-0000-1315 ANON-0000-0000-1316
## ---------------------------------------------------------------------------
------
## money
##        n  missing distinct     Info     Mean  pMedian      Gmd      .05
##     3636        0       20    0.971    31.75    31.84    5.506    23.02
##      .10      .25      .50      .75      .90      .95
##    25.96    27.92    32.82    35.76    37.72    37.72
##
## Value      18.12 21.06 23.02 24.00 25.00 25.96 27.92 28.90 29.00 30.00 30.
86
## Frequency     36    59   208    14     4   459   334    93     8    11   3
89
## Proportion 0.010 0.016 0.057 0.004 0.001 0.126 0.092 0.026 0.002 0.003 0.1
07
##
## Value      32.82 33.80 34.00 35.00 35.76 37.72 38.70 39.00 40.00
## Frequency    490    51     7     8  1014   273   128    18    32
## Proportion 0.135 0.014 0.002 0.002 0.279 0.075 0.035 0.005 0.009
##
## For the frequency table, variable is rounded to the nearest 0
## ---------------------------------------------------------------------------
------
## coffee_name
##        n  missing distinct
##     3636        0        8
##
## Americano (578, 0.159), Americano with Milk (824, 0.227), Cappuccino (501,
## 0.138), Cocoa (243, 0.067), Cortado (292, 0.080), Espresso (134, 0.037), H
ot
## Chocolate (282, 0.078), Latte (782, 0.215)
## ---------------------------------------------------------------------------
------
## Time_of_Day
##        n  missing distinct
##     3636        0        3
##
## Value      Afternoon   Morning     Night
## Frequency       1231      1221      1184
## Proportion     0.339     0.336     0.326
```

```
## ------------------------------------------------------------------------
------
## Weekday
##        n  missing distinct
##     3636        0        7
##
## Value        Fri   Mon   Sat   Sun   Thu   Tue   Wed
## Frequency    544   561   482   434   520   585   510
## Proportion 0.150 0.154 0.133 0.119 0.143 0.161 0.140
## ------------------------------------------------------------------------
------
## Month_name
##        n  missing distinct
##     3636        0       12
##
## Value        Apr   Aug   Dec   Feb   Jan   Jul   Jun   Mar   May   Nov   O
ct
## Frequency    196   272   259   423   201   237   227   525   267   259   4
26
## Proportion 0.054 0.075 0.071 0.116 0.055 0.065 0.062 0.144 0.073 0.071 0.1
17
##
## Value        Sep
## Frequency    344
## Proportion 0.095
## ------------------------------------------------------------------------
------
## Weekdaysort
##        n  missing distinct     Info     Mean  pMedian      Gmd
##     3636        0        7    0.979    3.848        4    2.257
##
## Value          1     2     3     4     5     6     7
## Frequency    561   585   510   520   544   482   434
## Proportion 0.154 0.161 0.140 0.143 0.150 0.133 0.119
##
## For the frequency table, variable is rounded to the nearest 0
## ------------------------------------------------------------------------
------
## Monthsort
##        n  missing distinct     Info     Mean  pMedian      Gmd      .05
##     3636        0       12    0.991    6.395      6.5     3.99        1
##      .10      .25      .50      .75      .90      .95
##        2        3        6       10       11       12
##
## Value          1     2     3     4     5     6     7     8     9    10
11
## Frequency    201   423   525   196   267   227   237   272   344   426   2
59
## Proportion 0.055 0.116 0.144 0.054 0.073 0.062 0.065 0.075 0.095 0.117 0.0
71
```

```
##
## Value         12
## Frequency    259
## Proportion 0.071
##
## For the frequency table, variable is rounded to the nearest 0
## ----------------------------------------------------------------------
------
```

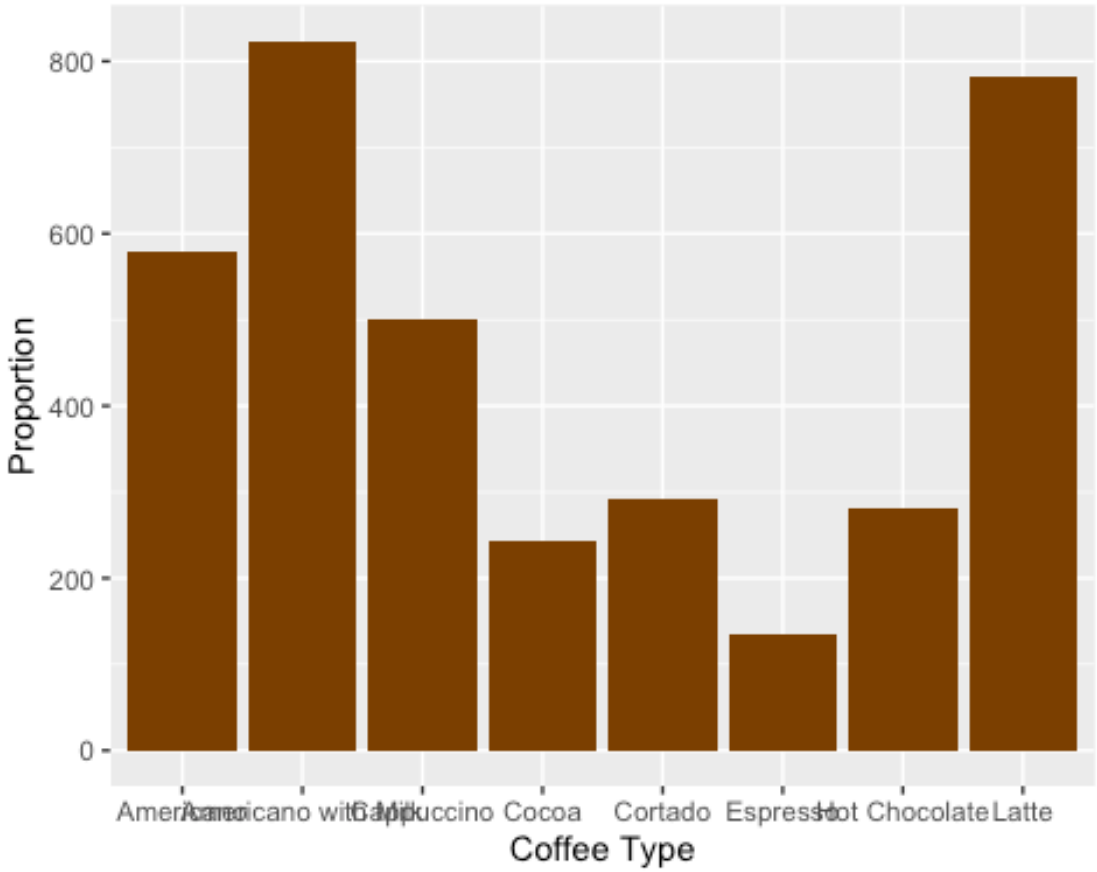
```
str(coffee) #check type
```

```
## tibble [3,636 × 12] (S3: tbl_df/tbl/data.frame)
##  $ date       : POSIXct[1:3636], format: "2024-03-01" "2024-03-01" ...
##  $ datetime   : POSIXct[1:3636], format: "2024-03-01 10:15:50" "2024-03-01
12:19:22" ...
##  $ hour_of_day: num [1:3636] 10 12 12 13 13 15 16 18 19 19 ...
##  $ cash_type  : Factor w/ 2 levels "card","cash": 1 1 1 1 1 1 1 1 1 1 ...
##  $ card       : Factor w/ 1316 levels "ANON-0000-0000-0001",..: 1 2 2 3 4
5 6 7 8 8 ...
##  $ money      : num [1:3636] 38.7 38.7 38.7 28.9 38.7 33.8 38.7 33.8 38.7
33.8 ...
##  $ coffee_name: Factor w/ 8 levels "Americano","Americano with Milk",..: 8
7 7 1 8 2 7 2 4 2 ...
##  $ Time_of_Day: Factor w/ 3 levels "Afternoon","Morning",..: 2 1 1 1 1 1 1
3 3 3 ...
##  $ Weekday    : Factor w/ 7 levels "Fri","Mon","Sat",..: 1 1 1 1 1 1 1 1 1
1 ...
##  $ Month_name : Factor w/ 12 levels "Apr","Aug","Dec",..: 8 8 8 8 8 8 8 8
8 8 ...
##  $ Weekdaysort: num [1:3636] 5 5 5 5 5 5 5 5 5 5 ...
##  $ Monthsort  : num [1:3636] 3 3 3 3 3 3 3 3 3 3 ...
```

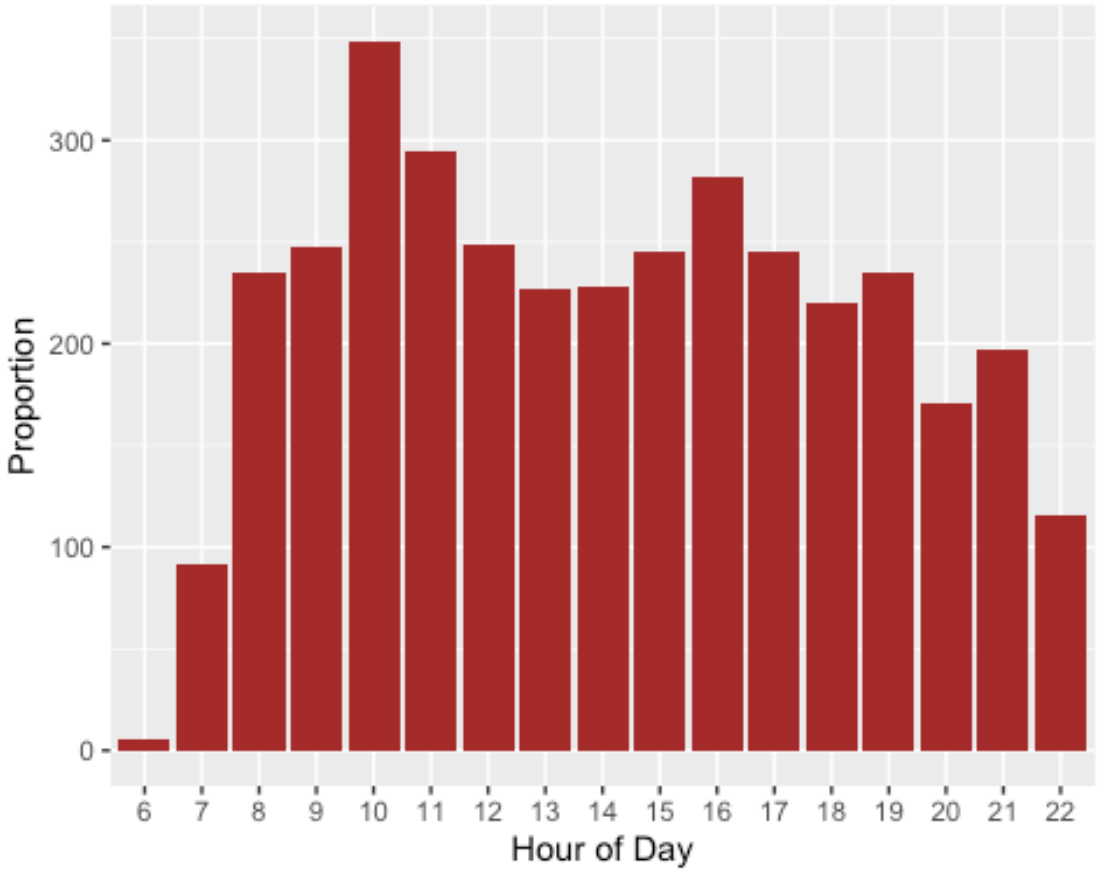
```
coffee$hour_of_day <- as.factor(coffee$hour_of_day )
sapply(coffee, nlevels)
```

```
##        date    datetime hour_of_day   cash_type        card       money
##           0           0          17           2        1316           0
## coffee_name Time_of_Day     Weekday  Month_name Weekdaysort   Monthsort
##           8           3           7          12           0           0
```

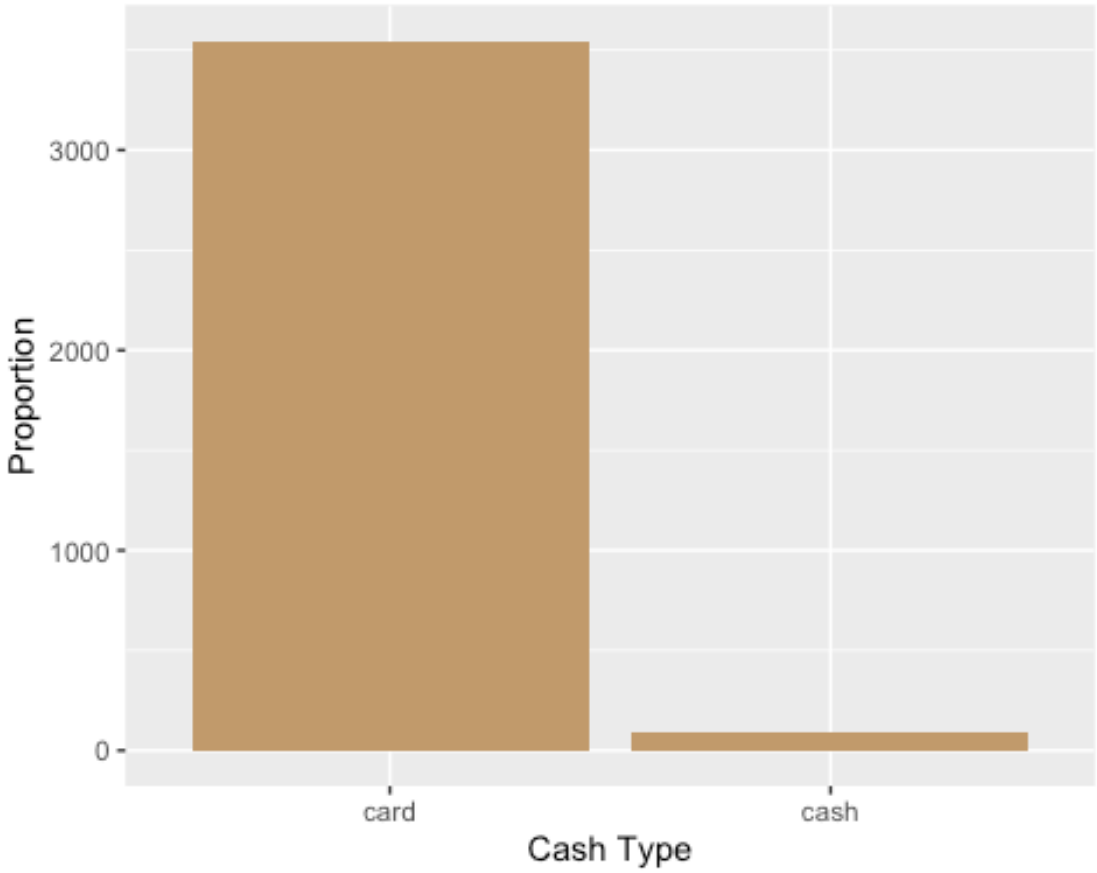
```
# Create bar charts to show proportions

coffee %>%
  ggplot(aes(x = coffee_name, y = after_stat(count))) +
  geom_bar(fill = "#7B3F00") +
  labs(y = "Proportion", x = "Coffee Type")
```



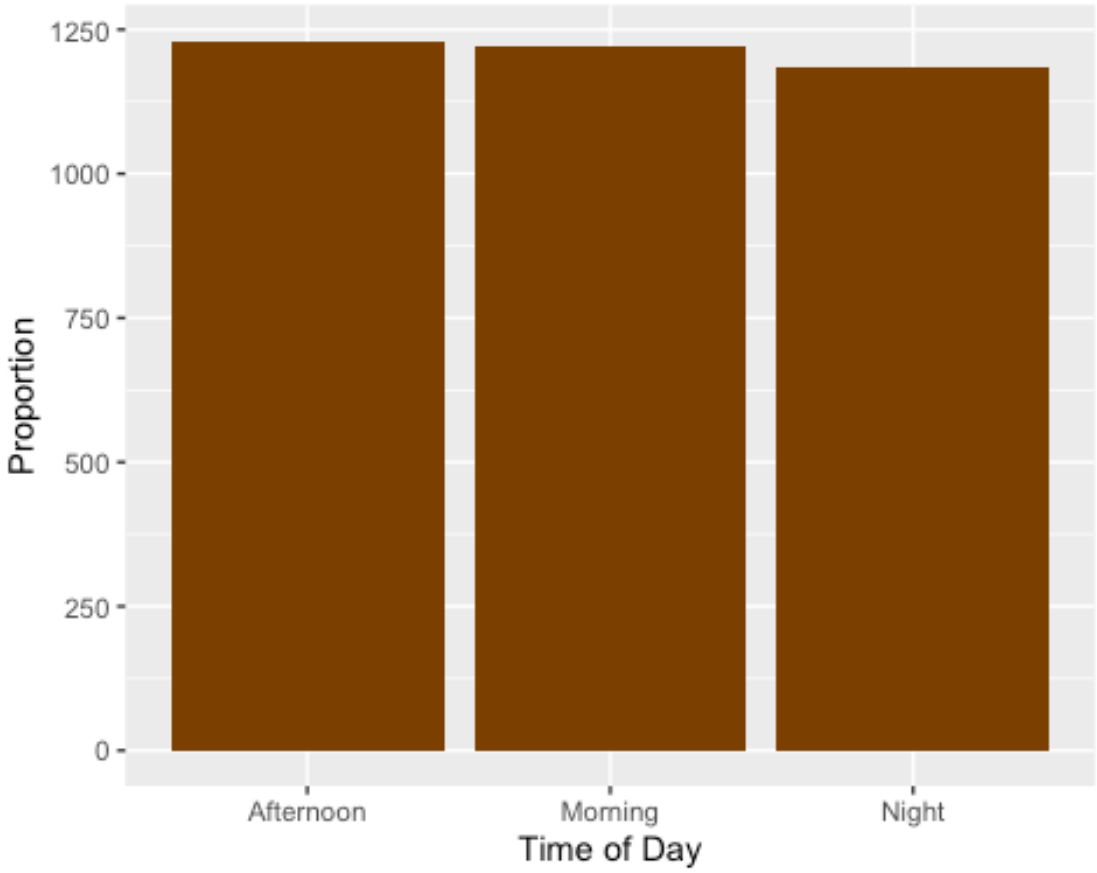

```
coffee %>%
  ggplot(aes(x = hour_of_day, y = after_stat(count))) +
  geom_bar(fill = "#A52A2A") +
  labs(y = "Proportion", x = "Hour of Day")
```



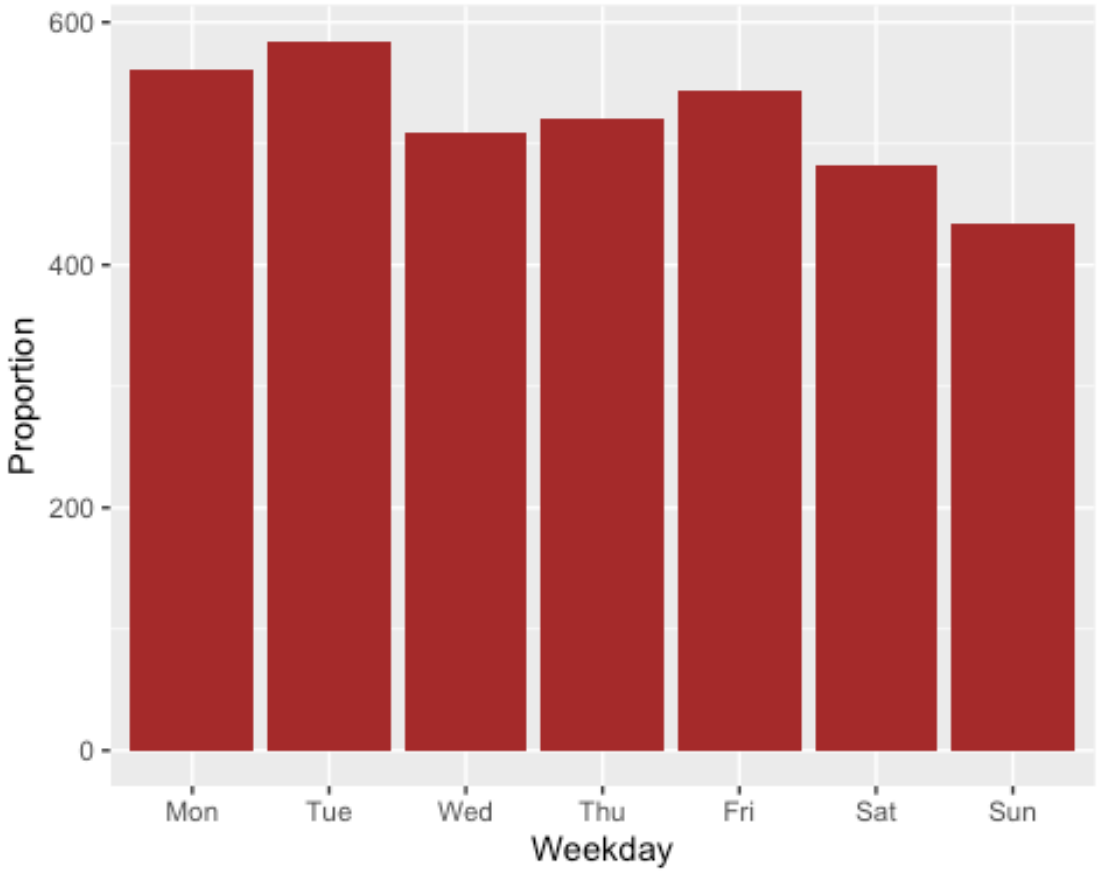

```
coffee %>%
  ggplot(aes(x = cash_type, y = after_stat(count))) +
  geom_bar(fill = "#C19A6B") +
  labs(y = "Proportion", x = "Cash Type")
```



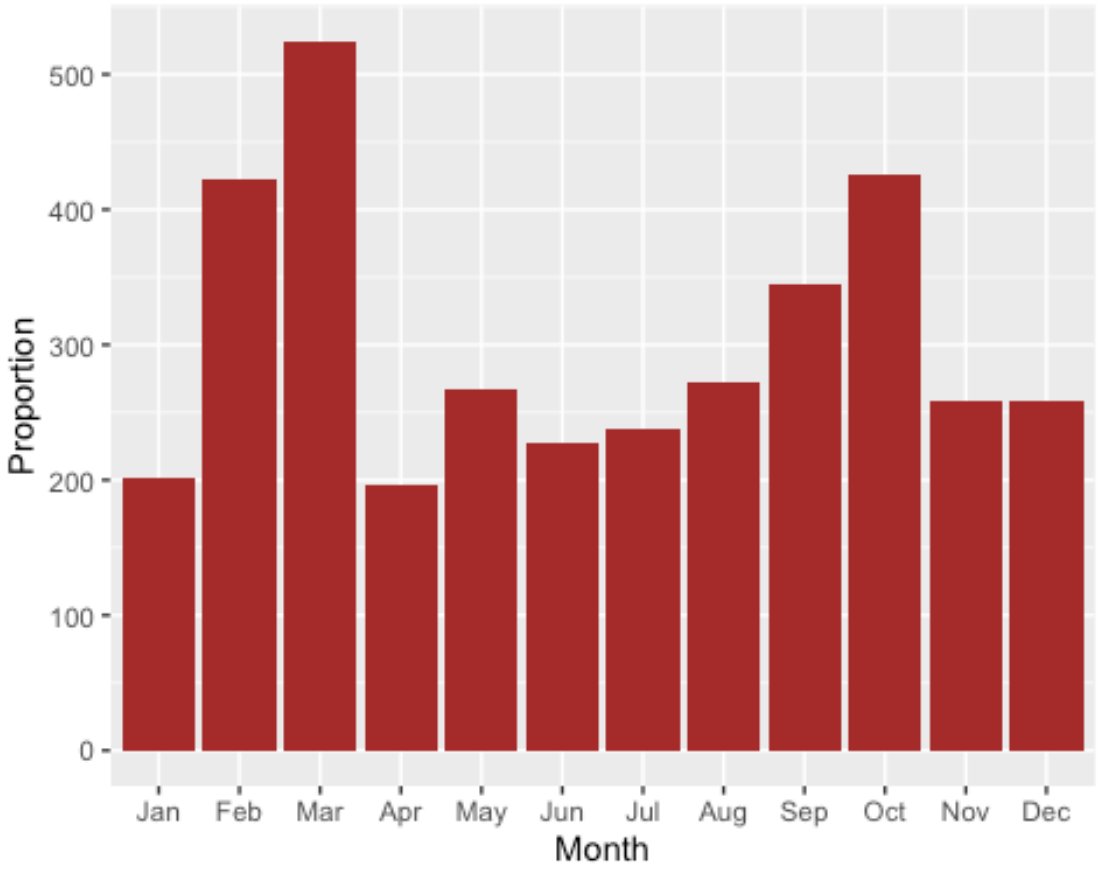

```
coffee %>%
  ggplot(aes(x = Time_of_Day, y = after_stat(count))) +
  geom_bar(fill = "#7B3F00") +
  labs(y = "Proportion", x = "Time of Day")
```




```
coffee$Weekday <- factor(coffee$Weekday, levels = c("Mon", "Tue", "Wed", "Thu
", "Fri", "Sat", "Sun"))
coffee %>%
  ggplot(aes(x = Weekday, y = after_stat(count))) +
  geom_bar(fill = "#A52A2A") +
  labs(y = "Proportion", x = "Weekday")
```




```r
coffee$Month_name <- factor(coffee$Month_name, levels = c("Jan", "Feb", "Mar"
, "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))
coffee %>%
  ggplot(aes(x = Month_name, y = after_stat(count))) +
  geom_bar(fill = "#A52A2A") +
  labs(y = "Proportion", x = "Month")
```




```
#Q1: What are the most important predictors when it comes to price?
#stepwise regression
full_mod <- lm(money ~ hour_of_day + cash_type + coffee_name + Time_of_Day +
Weekday + Month_name, data = coffee)
summary(full_mod)

##
## Call:
## lm(formula = money ~ hour_of_day + cash_type + coffee_name +
##     Time_of_Day + Weekday + Month_name, data = coffee)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3567 -0.2320 -0.0483  0.1209  4.3153
##
## Coefficients: (2 not defined because of singularities)
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   25.645216   0.334405  76.689  < 2e-16 ***
## hour_of_day7                   0.071920   0.337161   0.213    0.831
## hour_of_day8                   0.050867   0.331896   0.153    0.878
## hour_of_day9                  -0.048117   0.331707  -0.145    0.885
## hour_of_day10                  0.334347   0.330704   1.011    0.312
## hour_of_day11                  0.143590   0.331332   0.433    0.665
```

```
## hour_of_day12                    0.343650   0.331706   1.036    0.300
## hour_of_day13                    0.360206   0.331873   1.085    0.278
## hour_of_day14                    0.270522   0.331749   0.815    0.415
## hour_of_day15                    0.120046   0.331488   0.362    0.717
## hour_of_day16                    0.307500   0.331310   0.928    0.353
## hour_of_day17                    0.232654   0.331439   0.702    0.483
## hour_of_day18                    0.294557   0.332106   0.887    0.375
## hour_of_day19                    0.291602   0.332028   0.878    0.380
## hour_of_day20                    0.169620   0.333712   0.508    0.611
## hour_of_day21                    0.126837   0.333312   0.381    0.704
## hour_of_day22                    0.423108   0.335846   1.260    0.208
## cash_typecash                    1.964582   0.081920  23.982  < 2e-16 ***
## coffee_nameAmericano with Milk   5.012150   0.040706 123.130  < 2e-16 ***
## coffee_nameCappuccino            9.889347   0.045679 216.495  < 2e-16 ***
## coffee_nameCocoa                 9.795944   0.057125 171.481  < 2e-16 ***
## coffee_nameCortado               0.293612   0.053672   5.470  4.8e-08 ***
## coffee_nameEspresso             -4.703701   0.070536 -66.685  < 2e-16 ***
## coffee_nameHot Chocolate         9.917046   0.054781 181.030  < 2e-16 ***
## coffee_nameLatte                 9.892567   0.041291 239.583  < 2e-16 ***
## Time_of_DayMorning                     NA         NA      NA       NA
## Time_of_DayNight                       NA         NA      NA       NA
## WeekdayTue                      -0.026541   0.043443  -0.611    0.541
## WeekdayWed                      -0.035230   0.045053  -0.782    0.434
## WeekdayThu                      -0.040851   0.044772  -0.912    0.362
## WeekdayFri                       0.009731   0.044355   0.219    0.826
## WeekdaySat                       0.042835   0.045983   0.932    0.352
## WeekdaySun                      -0.005094   0.047310  -0.108    0.914
## Month_nameFeb                    0.028122   0.063605   0.442    0.658
## Month_nameMar                    1.118593   0.061499  18.189  < 2e-16 ***
## Month_nameApr                    2.467791   0.075238  32.800  < 2e-16 ***
## Month_nameMay                    1.885148   0.069117  27.275  < 2e-16 ***
## Month_nameJun                    1.947274   0.071755  27.138  < 2e-16 ***
## Month_nameJul                   -2.284526   0.070524 -32.394  < 2e-16 ***
## Month_nameAug                   -2.928784   0.068395 -42.822  < 2e-16 ***
## Month_nameSep                   -2.647035   0.065460 -40.438  < 2e-16 ***
## Month_nameOct                    0.016222   0.062915   0.258    0.797
## Month_nameNov                    0.001979   0.069245   0.029    0.977
## Month_nameDec                   -0.009027   0.069112  -0.131    0.896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7303 on 3594 degrees of freedom
## Multiple R-squared:  0.9782, Adjusted R-squared:  0.978
## F-statistic:  3937 on 41 and 3594 DF,  p-value: < 2.2e-16
```

The NAs for time_of_day indicate a potential multicollinearity, probably with hour of day.

```
#Perform a Chi-square test
contingency_table_hour_time <- table(coffee$hour_of_day, coffee$Time_of_Day)
contingency_table_hour_time
```

```
##
##      Afternoon Morning Night
##   6          0       5     0
##   7          0      91     0
##   8          0     235     0
##   9          0     247     0
##   10         0     349     0
##   11         0     294     0
##   12       249       0     0
##   13       227       0     0
##   14       228       0     0
##   15       245       0     0
##   16       282       0     0
##   17         0       0   245
##   18         0       0   220
##   19         0       0   235
##   20         0       0   171
##   21         0       0   197
##   22         0       0   116
```

```
chi_square_hour_time <- chisq.test(contingency_table_hour_time)
```

```
## Warning in chisq.test(contingency_table_hour_time): Chi-squared approximat
ion
## may be incorrect
```

```
chi_square_hour_time
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table_hour_time
## X-squared = 7272, df = 32, p-value < 2.2e-16
```

```
#Let's remove hour of day

full_mod2 <- lm(money ~ + cash_type + coffee_name + Weekday + Time_of_Day + M
onth_name, data = coffee)
summary(full_mod2)
```

```
##
## Call:
## lm(formula = money ~ +cash_type + coffee_name + Weekday + Time_of_Day +
##     Month_name, data = coffee)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3374 -0.2181 -0.0238  0.0841  4.4583
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                     25.9139422  0.0684211 378.742  < 2e-16 ***
## cash_typecash                    1.9802737  0.0823471  24.048  < 2e-16 ***
## coffee_nameAmericano with Milk   5.0057017  0.0408044 122.676  < 2e-16 ***
## coffee_nameCappuccino            9.8830809  0.0458336 215.630  < 2e-16 ***
## coffee_nameCocoa                 9.7781591  0.0573313 170.555  < 2e-16 ***
## coffee_nameCortado               0.2853874  0.0539868   5.286 1.32e-07 ***
## coffee_nameEspresso             -4.7079629  0.0710491 -66.264  < 2e-16 ***
## coffee_nameHot Chocolate         9.9202282  0.0550183 180.308  < 2e-16 ***
## coffee_nameLatte                 9.8891122  0.0413919 238.914  < 2e-16 ***
## WeekdayTue                      -0.0246388  0.0437592  -0.563    0.573
## WeekdayWed                      -0.0246380  0.0452819  -0.544    0.586
## WeekdayThu                      -0.0373324  0.0450740  -0.828    0.408
## WeekdayFri                       0.0127268  0.0445290   0.286    0.775
## WeekdaySat                       0.0545793  0.0461465   1.183    0.237
## WeekdaySun                       0.0068974  0.0474683   0.145    0.884
## Time_of_DayMorning              -0.1413870  0.0303413  -4.660 3.28e-06 ***
## Time_of_DayNight                -0.0289976  0.0308801  -0.939    0.348
## Month_nameFeb                    0.0374102  0.0636929   0.587    0.557
## Month_nameMar                    1.1354185  0.0616499  18.417  < 2e-16 ***
## Month_nameApr                    2.5134302  0.0751563  33.443  < 2e-16 ***
## Month_nameMay                    1.8980183  0.0695500  27.290  < 2e-16 ***
## Month_nameJun                    1.9407303  0.0718231  27.021  < 2e-16 ***
## Month_nameJul                   -2.2735113  0.0710149 -32.015  < 2e-16 ***
## Month_nameAug                   -2.9331765  0.0688366 -42.611  < 2e-16 ***
## Month_nameSep                   -2.6505856  0.0658523 -40.250  < 2e-16 ***
## Month_nameOct                    0.0157993  0.0633468   0.249    0.803
## Month_nameNov                    0.0136413  0.0696321   0.196    0.845
## Month_nameDec                    0.0008563  0.0695487   0.012    0.990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7366 on 3608 degrees of freedom
## Multiple R-squared:  0.9778, Adjusted R-squared:  0.9776
## F-statistic:  5873 on 27 and 3608 DF,  p-value: < 2.2e-16

both_step <- step(full_mod2, direction = 'both', scope = formula(full_mod2))

## Start:  AIC=-2195.28
## money ~ +cash_type + coffee_name + Weekday + Time_of_Day + Month_name
##
##               Df Sum of Sq   RSS     AIC
## - Weekday      6         3  1960 -2202.0
## <none>                      1958 -2195.3
## - Time_of_Day  2        13  1971 -2175.2
## - cash_type    1       314  2271 -1656.7
## - Month_name  11      9136 11094  4090.1
## - coffee_name  7     69509 71466 10871.2
##
## Step:  AIC=-2202
## money ~ cash_type + coffee_name + Time_of_Day + Month_name
```

```
##
##               Df Sum of Sq   RSS     AIC
## <none>                      1960 -2202.0
## + Weekday      6         3  1958 -2195.3
## - Time_of_Day  2        13  1974 -2181.4
## - cash_type    1       315  2276 -1661.9
## - Month_name  11      9144 11104  4081.4
## - coffee_name  7     69609 71570 10864.5
```

```
summary(both_step)
```

```
##
## Call:
## lm(formula = money ~ cash_type + coffee_name + Time_of_Day +
##     Month_name, data = coffee)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3779 -0.2104 -0.0323  0.0805  4.4260
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  25.914439   0.061813 419.239  < 2e-16 ***
## cash_typecash                 1.983629   0.082286  24.106  < 2e-16 ***
## coffee_nameAmericano with Milk  5.006651   0.040735 122.908  < 2e-16 ***
## coffee_nameCappuccino          9.883933   0.045764 215.976  < 2e-16 ***
## coffee_nameCocoa               9.778838   0.057181 171.015  < 2e-16 ***
## coffee_nameCortado             0.288743   0.053864   5.361 8.82e-08 ***
## coffee_nameEspresso           -4.712276   0.070935 -66.431  < 2e-16 ***
## coffee_nameHot Chocolate       9.918291   0.054940 180.529  < 2e-16 ***
## coffee_nameLatte               9.888638   0.041367 239.048  < 2e-16 ***
## Time_of_DayMorning            -0.144187   0.030230  -4.770 1.92e-06 ***
## Time_of_DayNight              -0.034801   0.030741  -1.132    0.258
## Month_nameFeb                  0.032814   0.063534   0.516    0.606
## Month_nameMar                  1.134756   0.061584  18.426  < 2e-16 ***
## Month_nameApr                  2.513781   0.075035  33.501  < 2e-16 ***
## Month_nameMay                  1.893257   0.069415  27.275  < 2e-16 ***
## Month_nameJun                  1.942431   0.071607  27.126  < 2e-16 ***
## Month_nameJul                 -2.276208   0.070836 -32.134  < 2e-16 ***
## Month_nameAug                 -2.931563   0.068703 -42.670  < 2e-16 ***
## Month_nameSep                 -2.649189   0.065580 -40.396  < 2e-16 ***
## Month_nameOct                  0.011780   0.063205   0.186    0.852
## Month_nameNov                  0.018451   0.069508   0.265    0.791
## Month_nameDec                  0.002605   0.069300   0.038    0.970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7365 on 3614 degrees of freedom
## Multiple R-squared:  0.9777, Adjusted R-squared:  0.9776
## F-statistic:  7552 on 21 and 3614 DF,  p-value: < 2.2e-16
```

```r
red_mod <- lm(money ~ cash_type + coffee_name + Time_of_Day +
                Month_name, data = coffee)
summary(red_mod)
```

```
##
## Call:
## lm(formula = money ~ cash_type + coffee_name + Time_of_Day +
##     Month_name, data = coffee)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3779 -0.2104 -0.0323  0.0805  4.4260
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    25.914439   0.061813 419.239  < 2e-16 ***
## cash_typecash                   1.983629   0.082286  24.106  < 2e-16 ***
## coffee_nameAmericano with Milk  5.006651   0.040735 122.908  < 2e-16 ***
## coffee_nameCappuccino           9.883933   0.045764 215.976  < 2e-16 ***
## coffee_nameCocoa                9.778838   0.057181 171.015  < 2e-16 ***
## coffee_nameCortado              0.288743   0.053864   5.361 8.82e-08 ***
## coffee_nameEspresso            -4.712276   0.070935 -66.431  < 2e-16 ***
## coffee_nameHot Chocolate        9.918291   0.054940 180.529  < 2e-16 ***
## coffee_nameLatte                9.888638   0.041367 239.048  < 2e-16 ***
## Time_of_DayMorning             -0.144187   0.030230  -4.770 1.92e-06 ***
## Time_of_DayNight               -0.034801   0.030741  -1.132    0.258
## Month_nameFeb                   0.032814   0.063534   0.516    0.606
## Month_nameMar                   1.134756   0.061584  18.426  < 2e-16 ***
## Month_nameApr                   2.513781   0.075035  33.501  < 2e-16 ***
## Month_nameMay                   1.893257   0.069415  27.275  < 2e-16 ***
## Month_nameJun                   1.942431   0.071607  27.126  < 2e-16 ***
## Month_nameJul                  -2.276208   0.070836 -32.134  < 2e-16 ***
## Month_nameAug                  -2.931563   0.068703 -42.670  < 2e-16 ***
## Month_nameSep                  -2.649189   0.065580 -40.396  < 2e-16 ***
## Month_nameOct                   0.011780   0.063205   0.186    0.852
## Month_nameNov                   0.018451   0.069508   0.265    0.791
## Month_nameDec                   0.002605   0.069300   0.038    0.970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7365 on 3614 degrees of freedom
## Multiple R-squared:  0.9777, Adjusted R-squared:  0.9776
## F-statistic:  7552 on 21 and 3614 DF,  p-value: < 2.2e-16
```

```r
AIC(full_mod2)
```

```
## [1] 8125.239
```

```r
AIC(red_mod)
```

```
## [1] 8118.52
```

```r
#To answer Q3

modpred <- lm(money ~ hour_of_day + Weekday, data = coffee)
summary(modpred)
```

```
##
## Call:
## lm(formula = money ~ hour_of_day + Weekday, data = coffee)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.0469  -3.9076   0.7681   3.6993  10.1282
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.91932    2.14989  13.917   <2e-16 ***
## hour_of_day7   2.55819    2.20364   1.161   0.2458
## hour_of_day8   0.08776    2.16837   0.040   0.9677
## hour_of_day9   0.31595    2.16740   0.146   0.8841
## hour_of_day10  1.74875    2.16203   0.809   0.4187
## hour_of_day11  0.35690    2.16539   0.165   0.8691
## hour_of_day12  1.04782    2.16843   0.483   0.6290
## hour_of_day13  1.58026    2.17016   0.728   0.4666
## hour_of_day14  2.09163    2.16967   0.964   0.3351
## hour_of_day15  2.03518    2.16819   0.939   0.3480
## hour_of_day16  2.83247    2.16561   1.308   0.1910
## hour_of_day17  2.74605    2.16801   1.267   0.2054
## hour_of_day18  3.12789    2.17081   1.441   0.1497
## hour_of_day19  4.16844    2.16954   1.921   0.0548 .
## hour_of_day20  3.34706    2.17845   1.536   0.1245
## hour_of_day21  3.08302    2.17453   1.418   0.1563
## hour_of_day22  2.54123    2.19101   1.160   0.2462
## WeekdayTue    -0.09947    0.28378  -0.351   0.7260
## WeekdayWed    -0.40866    0.29420  -1.389   0.1649
## WeekdayThu    -0.40446    0.29244  -1.383   0.1667
## WeekdayFri    -0.06553    0.28971  -0.226   0.8211
## WeekdaySat    -0.24108    0.29963  -0.805   0.4211
## WeekdaySun     0.04178    0.30832   0.136   0.8922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.792 on 3613 degrees of freedom
## Multiple R-squared:  0.05724,    Adjusted R-squared:  0.0515
## F-statistic: 9.972 on 22 and 3613 DF,  p-value: < 2.2e-16
```