View All View Keyframes

Overview of R

- Over the past decade, companies have been using a variety of statistical packages, including SPSS, SAS, Minitab, Stata, and others.
- Recently, companies have begun to migrate to a powerful statistical and data mining package called **R**.
- R is an open source software.
  - Free
  - Cost of other packages has driven new users to R.
- R is a language, but we will mostly operate it using a graphical user interface.

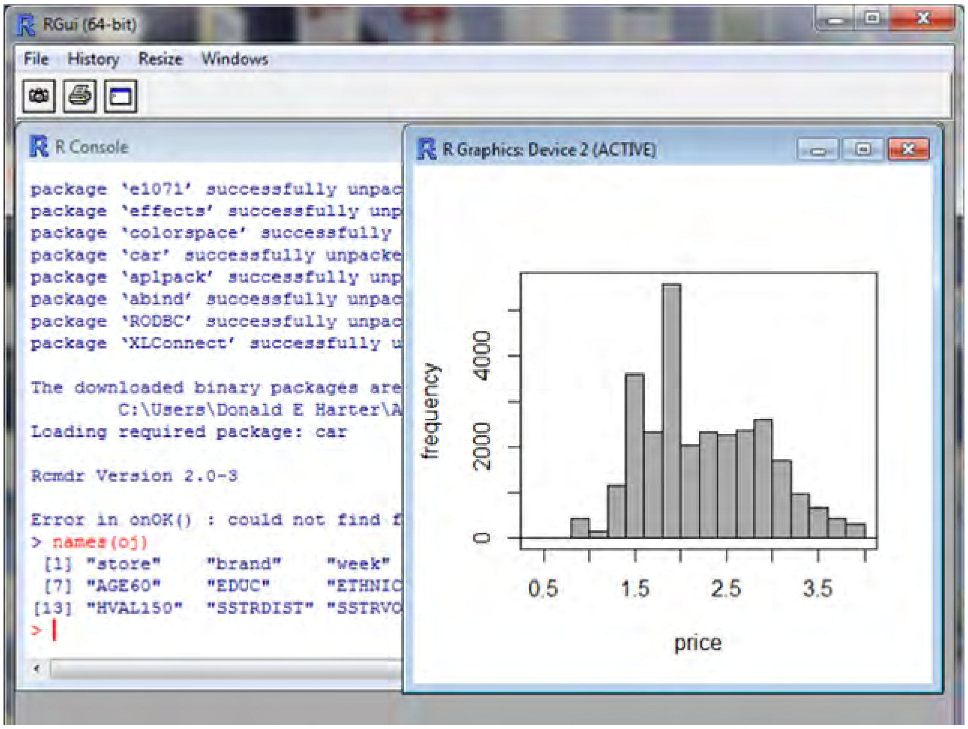
7 of 7

View All View Keyframes

Visualization

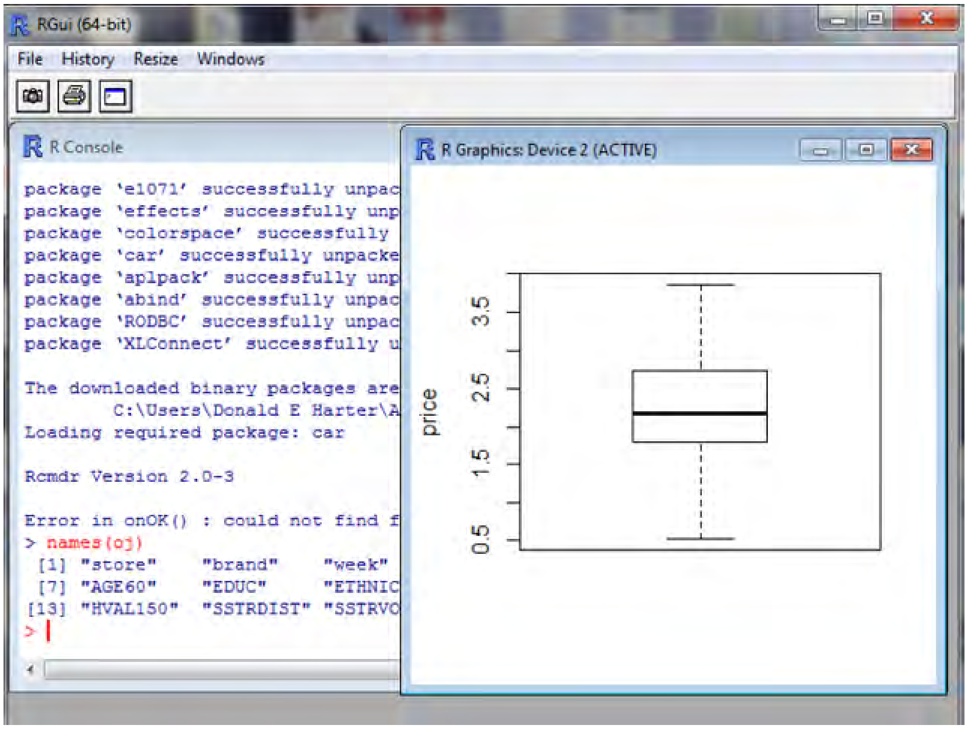
- R has the ability to generate histograms, boxplots, scatterplots, and XY plots similar to Excel.
- Graphs from R can differentiate by brand, product category, or other variable.
    - Gives greater insight into patterns

4 of 39



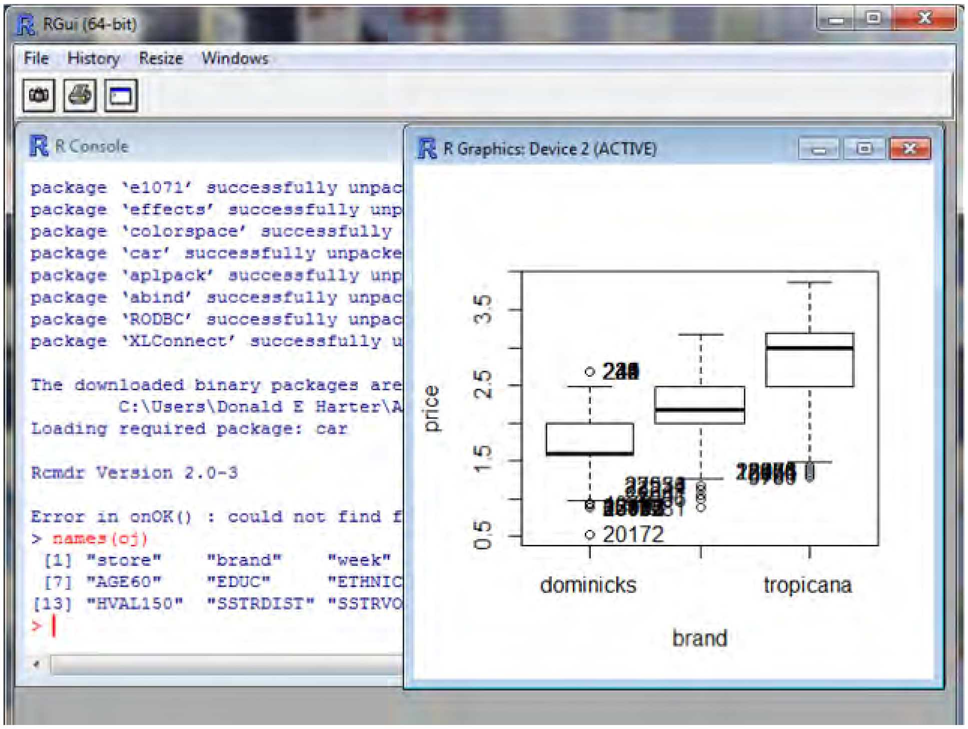
```
package 'e1071' successfully unpac
package 'effects' successfully unp
package 'colorspace' successfully
package 'car' successfully unpacke
package 'aplpack' successfully unp
package 'abind' successfully unpac
package 'RODBC' successfully unpac
package 'XLConnect' successfully u

The downloaded binary packages are
        C:\Users\Donald E Harter\A
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find f
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNIC
[13] "HVAL150"  "SSTRDIST" "SSTRVO
> 
```

5 of 39



```
package 'e1071' successfully unpac
package 'effects' successfully unp
package 'colorspace' successfully
package 'car' successfully unpacke
package 'aplpack' successfully unp
package 'abind' successfully unpac
package 'RODBC' successfully unpac
package 'XLConnect' successfully u

The downloaded binary packages are
        C:\Users\Donald E Harter\A
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find f
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNIC
[13] "HVAL150"  "SSTRDIST" "SSTRVO
> 
```

10 of 39



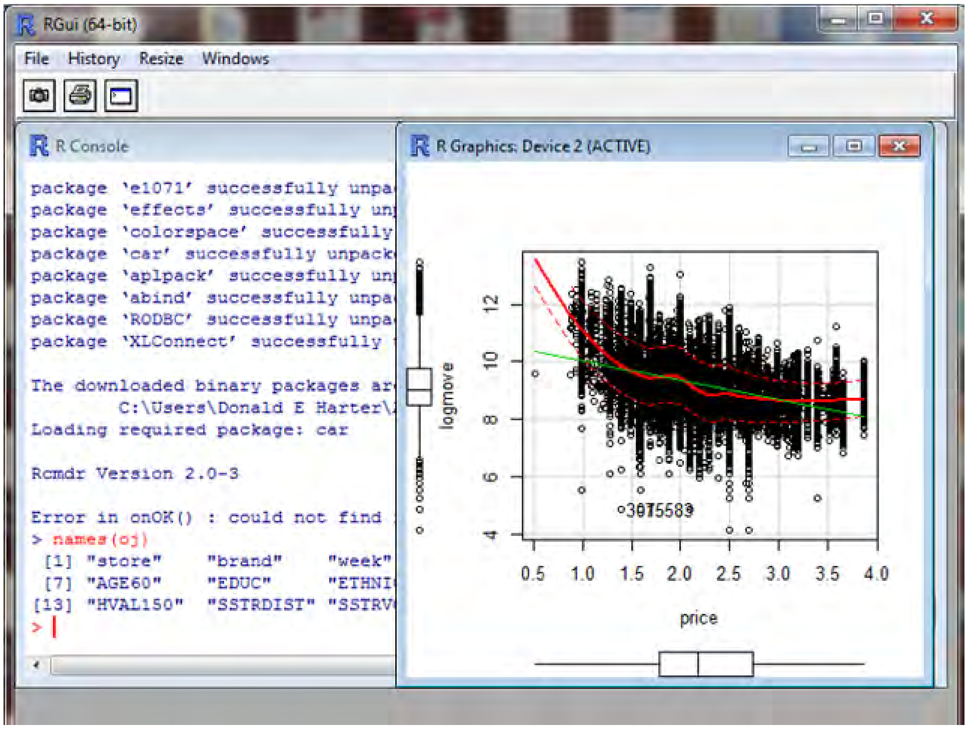
```
package 'e1071' successfully unpac
package 'effects' successfully unp
package 'colorspace' successfully
package 'car' successfully unpacke
package 'aplpack' successfully unp
package 'abind' successfully unpac
package 'RODBC' successfully unpac
package 'XLConnect' successfully u

The downloaded binary packages are
        C:\Users\Donald E Harter\A
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find f
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNIC
[13] "HVAL150"  "SSTRDIST" "SSTRVO
> 
```

19 of 39

Histograms and Boxplots

- **Histograms** show frequency of data within intervals or bins
  - Easy to develop in Excel or R.
- **Boxplots** identify:
  - Maximum and minimum—whiskers
  - 25%-ile and 75%-ile—box
  - Median—center bar
- Boxplots in R can also differentiate by brands.

23 of 39



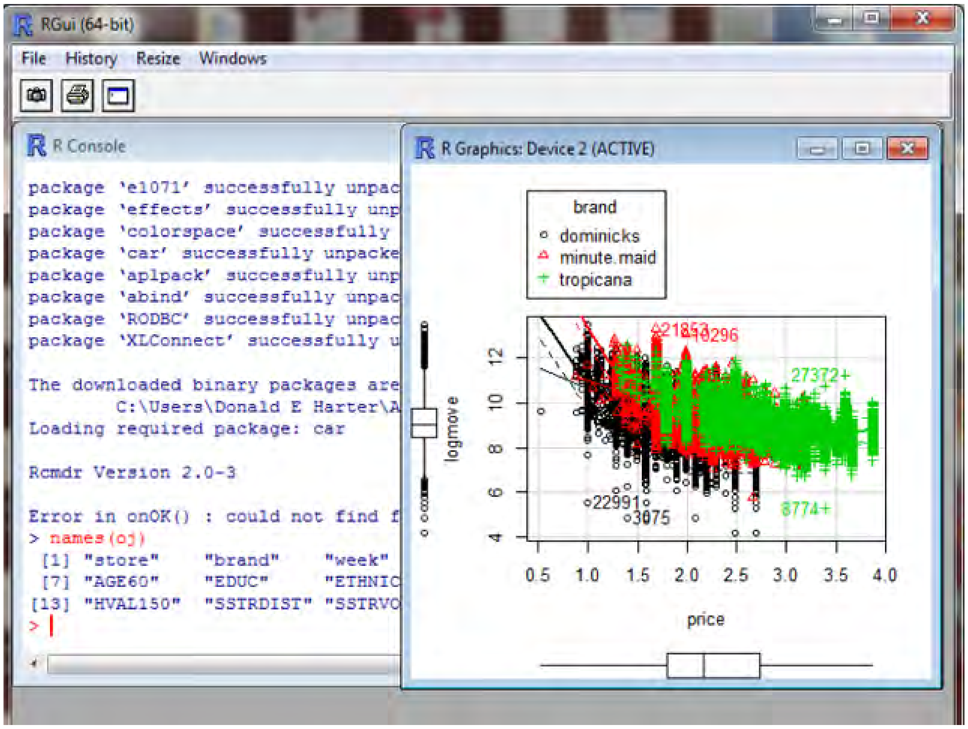
```
package 'e1071' successfully unpa
package 'effects' successfully un
package 'colorspace' successfully
package 'car' successfully unpack
package 'aplpack' successfully un
package 'abind' successfully unpa
package 'RODBC' successfully unpa
package 'XLConnect' successfully

The downloaded binary packages ar
        C:\Users\Donald E Harter\
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNI
[13] "HVAL150"  "SSTRDIST" "SSTRV
>
```

24 of 39

Scatterplots

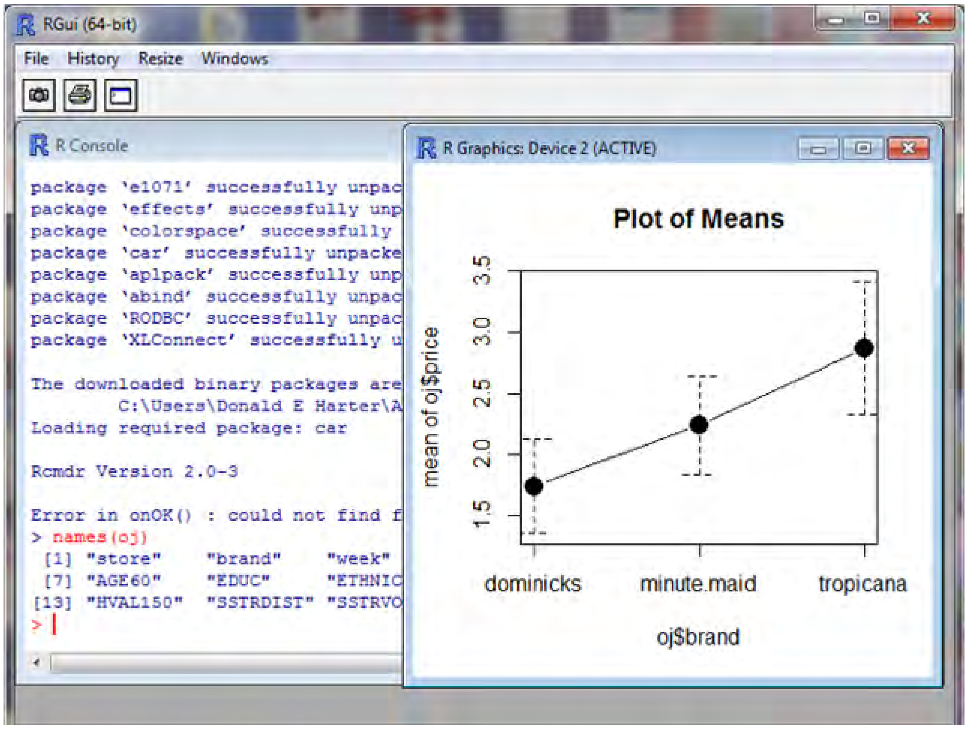
- **Scatterplots** offer not only trend line, but price by price.
- From example:
  - Solid red line—average sales by price
  - Green line—trend line
  - Dotted red line—standard deviation
- Scatterplots of log sales by price can also be categorized by brand.

30 of 39



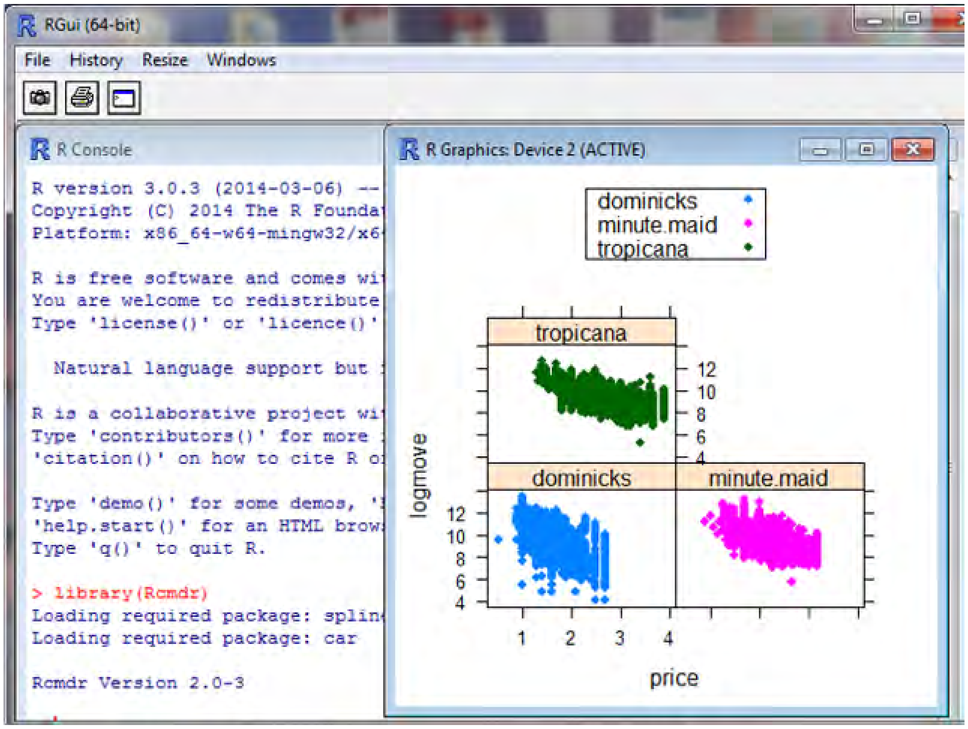
```
package 'e1071' successfully unpac
package 'effects' successfully unp
package 'colorspace' successfully
package 'car' successfully unpacke
package 'aplpack' successfully unp
package 'abind' successfully unpac
package 'RODBC' successfully unpac
package 'XLConnect' successfully u

The downloaded binary packages are
        C:\Users\Donald E Harter\A
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find f
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNIC
[13] "HVAL150"  "SSTRDIST" "SSTRVO
>
```

31 of 39



```
package 'e1071' successfully unpac
package 'effects' successfully unp
package 'colorspace' successfully
package 'car' successfully unpacke
package 'aplpack' successfully unp
package 'abind' successfully unpac
package 'RODBC' successfully unpac
package 'XLConnect' successfully u

The downloaded binary packages are
        C:\Users\Donald E Harter\A
Loading required package: car

Rcmdr Version 2.0-3

Error in onOK() : could not find f
> names(oj)
 [1] "store"    "brand"    "week"
 [7] "AGE60"    "EDUC"     "ETHNIC
[13] "HVAL150"  "SSTRDIST" "SSTRVO
>
```

36 of 39



```
R version 3.0.3 (2014-03-06) --
Copyright (C) 2014 The R Founda
Platform: x86_64-w64-mingw32/x6

R is free software and comes wi
You are welcome to redistribute
Type 'license()' or 'licence()'

  Natural language support but

R is a collaborative project wi
Type 'contributors()' for more
'citation()' on how to cite R o

Type 'demo()' for some demos, '
'help.start()' for an HTML brow
Type 'q()' to quit R.

> library(Rcmdr)
Loading required package: splin
Loading required package: car

Rcmdr Version 2.0-3
```

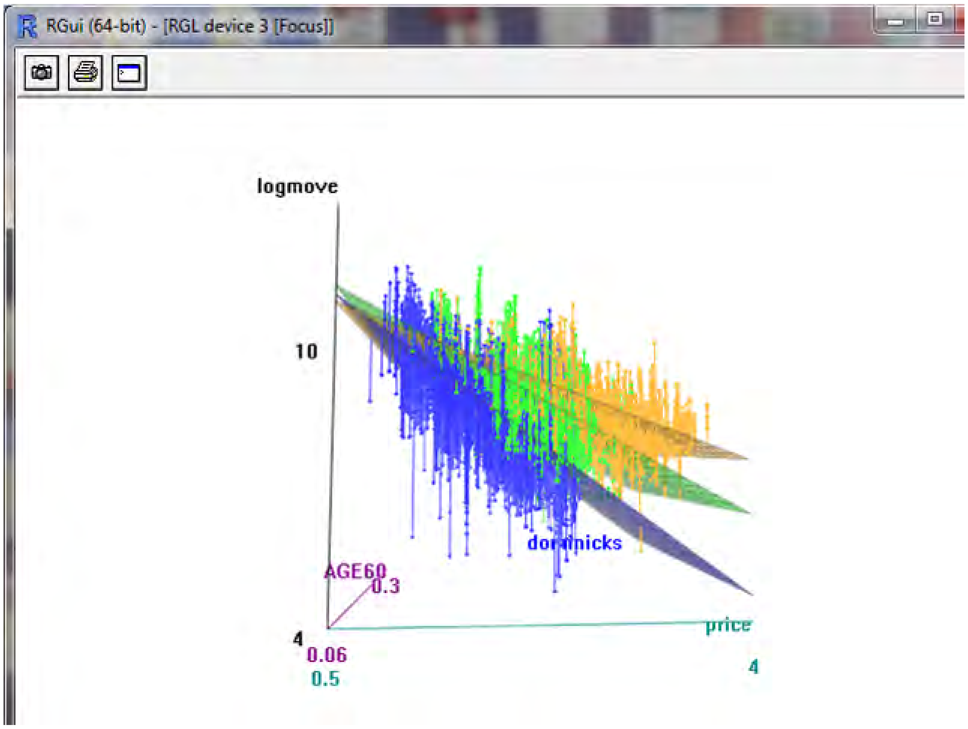
38 of 39

Plot of Means and XY Plots

- **Plot of means** is where the average price by brand is plotted.
- **XY plots** display the effect of price on sales per brand by separating and placing plots next to each other.

39 of 39

View All View Keyframes



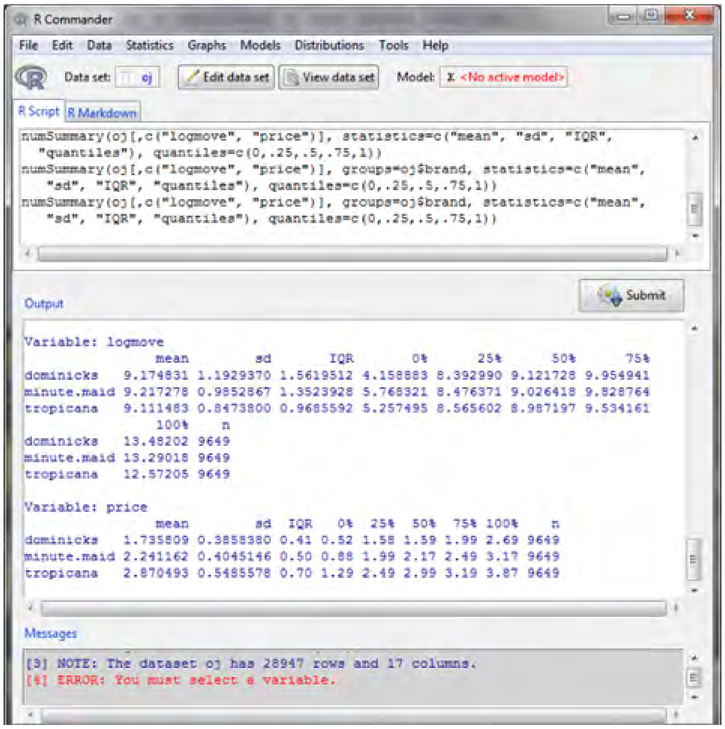
5 of 13

3-D Visualization

- R can perform more sophisticated 3-D graphing than Excel.
  - Plot the planes—helps find patterns
  - Rotate the graphs—offers different senses of parameters.
- 3-D allows you to see regression lines and planes to identify patterns.

13 of 13

View All View Keyframes


R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: oj
Edit data set
View data set
Model: Σ <No active model>
R Script
R Markdown
Submit
Output
Messages

```
numSummary(oj[,c("logmove", "price")], statistics=c("mean", "sd", "IQR",
  "quantiles"), quantiles=c(0,.25,.5,.75,1))
numSummary(oj[,c("logmove", "price")], groups=oj$brand, statistics=c("mean",
   "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
numSummary(oj[,c("logmove", "price")], groups=oj$brand, statistics=c("mean",
   "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

```
Variable: logmove
                mean        sd       IQR       0%      25%      50%      75%
dominicks   9.174831 1.1929370 1.5619512 4.158883 8.392990 9.121728 9.954941
minute.maid 9.217278 0.9852867 1.3523928 5.768321 8.476371 9.026418 9.828764
tropicana   9.111483 0.8473800 0.9685592 5.257495 8.565602 8.987197 9.534161
                100%    n
dominicks   13.48202 9649
minute.maid 13.29018 9649
tropicana   12.57205 9649

Variable: price
                mean        sd  IQR   0%  25%  50%  75% 100%    n
dominicks   1.735809 0.3858380 0.41 0.52 1.58 1.59 1.99 2.69 9649
minute.maid 2.241162 0.4045146 0.50 0.88 1.99 2.17 2.49 3.17 9649
tropicana   2.870493 0.5485578 0.70 1.29 2.49 2.99 3.19 3.87 9649
```

```
[3] NOTE: The dataset oj has 28947 rows and 17 columns.
[4] ERROR: You must select a variable.
```

4 of 13

Statistical Summaries

- Similar to Excel, R can calculate descriptive statistics.
- R has the additional ability to calculate descriptive statistics by brand.
- Stratify and characterize data while performing analysis.

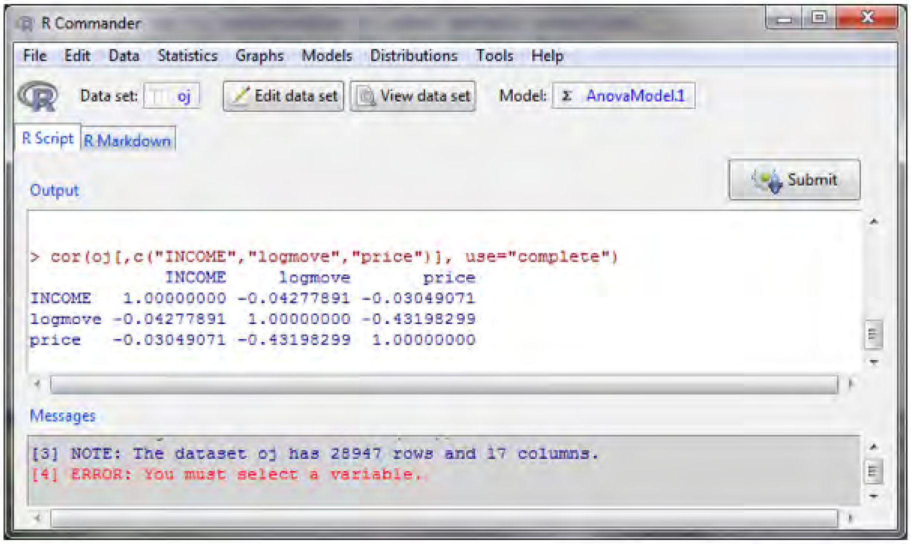
13 of 13

View All View Keyframes

Correlations

- Measure how two variables are related
- Positively correlated: one variable increases, the other one also increases
- Negatively correlated: one variable increases, the other one decreases

4 of 11

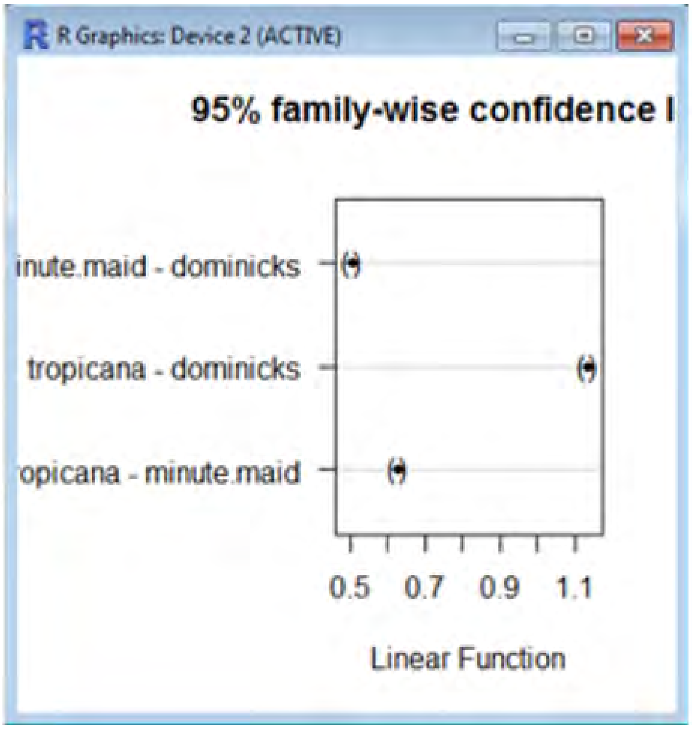


```
> cor(oj[,c("INCOME","logmove","price")], use="complete")
             INCOME     logmove       price
INCOME   1.00000000 -0.04277891 -0.03049071
logmove -0.04277891  1.00000000 -0.43198299
price   -0.03049071 -0.43198299  1.00000000
```

```
[3] NOTE: The dataset oj has 28947 rows and 17 columns.
[4] ERROR: You must select a variable.
```

5 of 11

View All View Keyframes



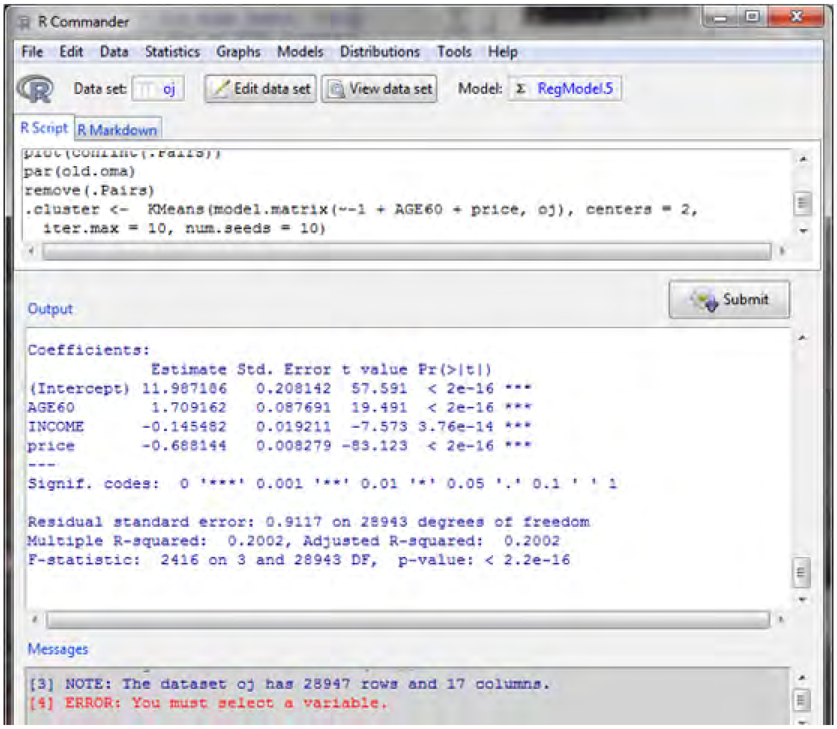
8 of 11

Analysis of Variance

- Store managers want to compare product placement positions, sales, and prices of different products or brands.
- Analysis of variance (ANOVA) is used to compare the averages for different categories or brands.
- ANOVA calculates the average of each item, combines with standard deviation/variance, and determines if they are statistically different from one another.
- ANOVA is a valuable technique for product placement.

11 of 11

View All View Keyframes



```
plot(confint(.Pairs))
par(old.oma)
remove(.Pairs)
.cluster <-  KMeans(model.matrix(~-1 + AGE60 + price, oj), centers = 2,
  iter.max = 10, num.seeds = 10)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.987186   0.208142  57.591  < 2e-16 ***
AGE60        1.709162   0.087691  19.491  < 2e-16 ***
INCOME      -0.145482   0.019211  -7.573 3.76e-14 ***
price       -0.688144   0.008279 -83.123  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9117 on 28943 degrees of freedom
Multiple R-squared:  0.2002, Adjusted R-squared:  0.2002
F-statistic:  2416 on 3 and 28943 DF,  p-value: < 2.2e-16
```

```
[3] NOTE: The dataset oj has 28947 rows and 17 columns.
[4] ERROR: You must select a variable.
```

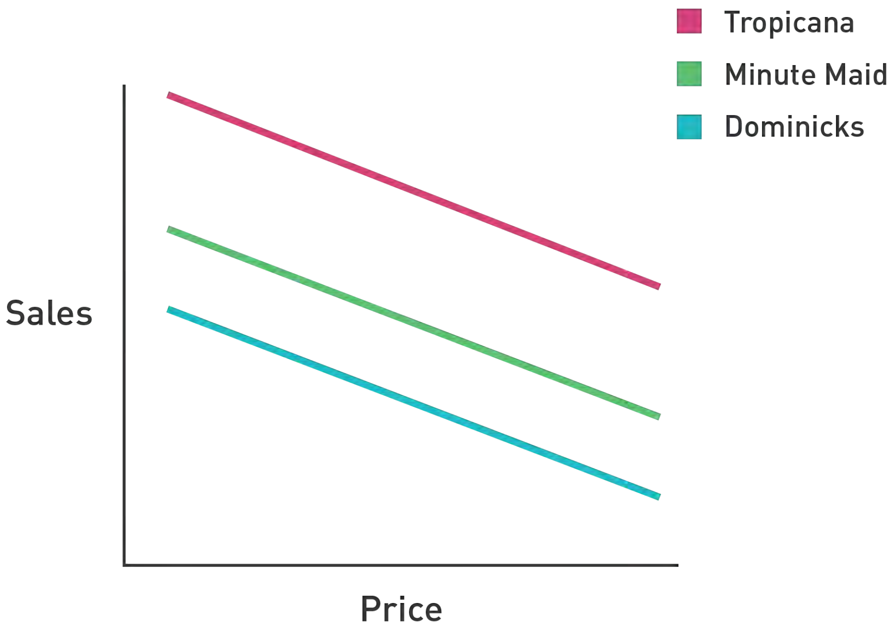
4 of 9

Regression

- R can perform linear regression.
    - Output is similar to Excel's multivariate regression output
- Analysis produces coefficients, statistical significance, and R-squared values.
    - Sets foundation to view more complex regressions

9 of 9

View All View Keyframes



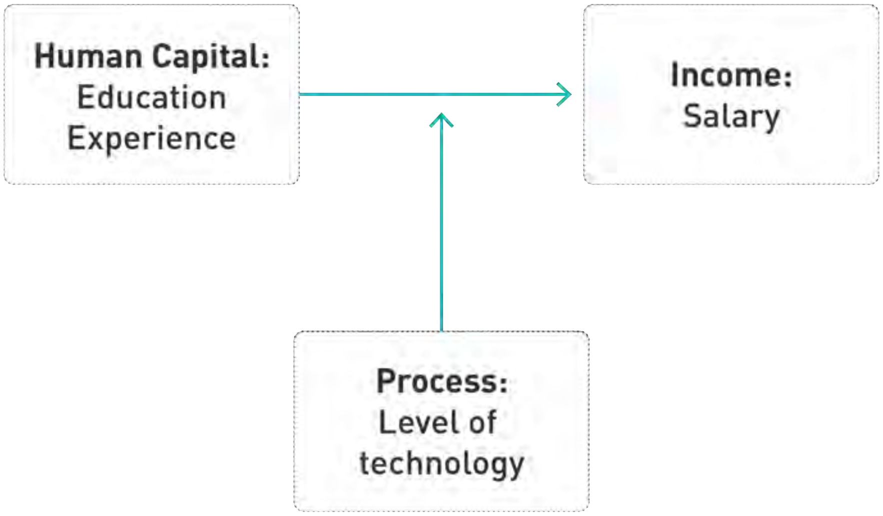
5 of 10

Regression with Dummy Variables

- Dummy variables change the intercept of some items to be different from others.
    - E.g., one brand will always be higher in sales than another
    - Reflected in different intercepts or heights of the regression line
- This allows you to look at price differentiation in different product markets.

10 of 10

View All View Keyframes

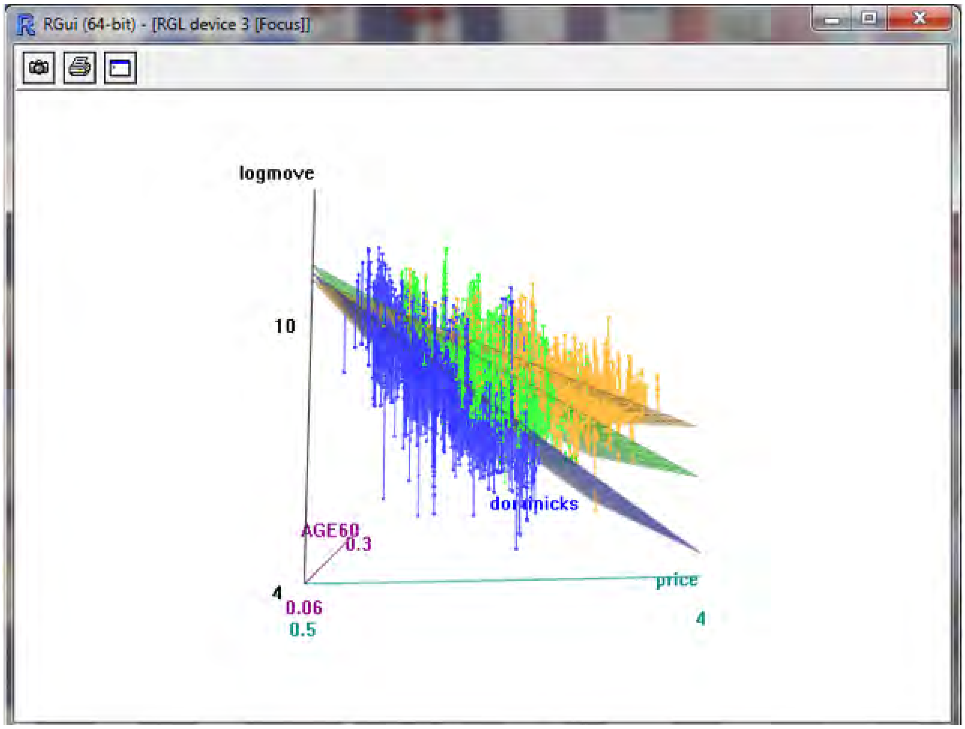


10 of 19

Regression with Moderating Effects

- Dummy variables can only change the intercepts.
- Moderating effects allow the slope of the line to change.
  - Recognizes that two variables might interact, magnifying their effects
- Education and experience both have a positive effect on income.
  - When an experienced, educated professional learns technology, the technology is the moderating effect.
- **Moderating effect:** An interaction of variables that leverage each other
- The moderating effect acts as a catalyst to accelerate the effect of certain variables.

18 of 19


RGui (64-bit) - [RGL device 3 [Focus]]
logmove
10
4
dominicks
AGE60
0.3
0.06
0.5
price
4

19 of 19