

Rotman

**Master of
Management
Analytics**

INTRO TO JMP - PART 1

Bootcamp (<https://tdmdal.github.io/mma-jmp/>)

September 11, 2020 Prepared by Jay / [TDMDAL](#)



Rotman School of Management
UNIVERSITY OF TORONTO

What is JMP

- Predictive analytics software from SAS Institute
 - data manipulation
 - [visualization](#)
 - statistical & machine learning modeling
 - reporting
- Intuitive point-and-click interface for beginners
- Flexible and extensible for advanced users
 - JMP Scripting Language (JSL) for automating or extending point-and-click functionality
 - Connect to the richness of SAS: retrieve SAS data and submit SAS code
 - Connect to Database engine, Matlab, R, Python, Excel, Web API, etc.

See key features of JMP Pro at [JMP Pro website](#).

Why JMP

- Get you started as quickly as possible
 - **amazing** intuitive point-and-click interface
 - quickly get things done
- JMP is a great tool
 - sufficient for many analytics tasks
 - good [visualization](#) for reporting/presentation
 - good for initial exploration even you plan to use more advanced tools later

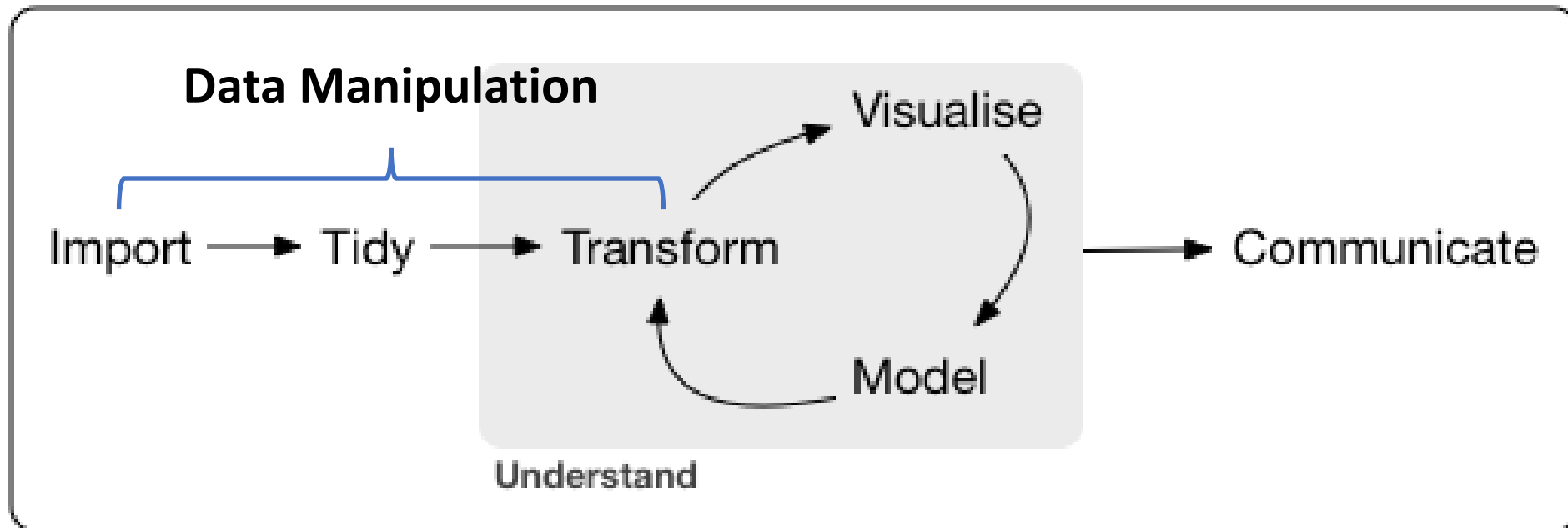
Learning by Doing

- This workshop: task/problem based
 - I show you how to do a task
 - You will try yourself on a variation of the task
 - you will discover JMP on your own
- Learning Resources
 - [JMP Documentation Library](#)
 - Getting started document: [Discovering JMP](#)
 - Basics: [Using JMP](#) (data table), [Basic Analysis](#) and [Essential Graphing](#)
 - Specific topics: [Fitting Linear Models](#), [Predictive and Specialized Modeling](#), etc.
 - [JMP Academic Home](#)
 - [JMP Learning Library](#)

Plan

- Session 1
 - **Workflow overview**
 - Basic data manipulation
- Session 2
 - Join data tables
 - JMP graphing
- Session 3
 - Modelling
 - JMP Journal
 - JMP Scripting Language

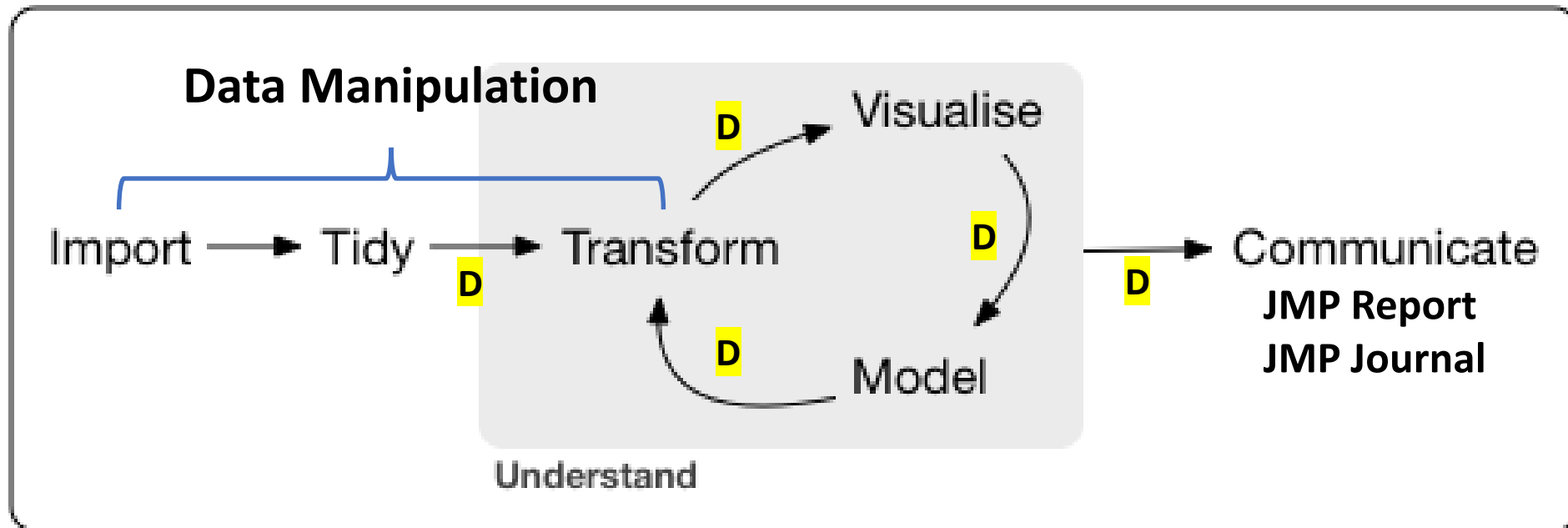
Overview: Analytics Software Typical Workflow



point-and-click & programming

Concept and graph adapted from <https://r4ds.had.co.nz/introduction.html>

Overview: JMP Typical Workflow



point-and-click & programming

D: Dynamic Link

Overview: A Simple Workflow Demo

- A simple example
 - import data
 - fit y by x (Profit by # Employ; linear model; removing an outlier)
- Data (data/basics/companies_mma.csv)

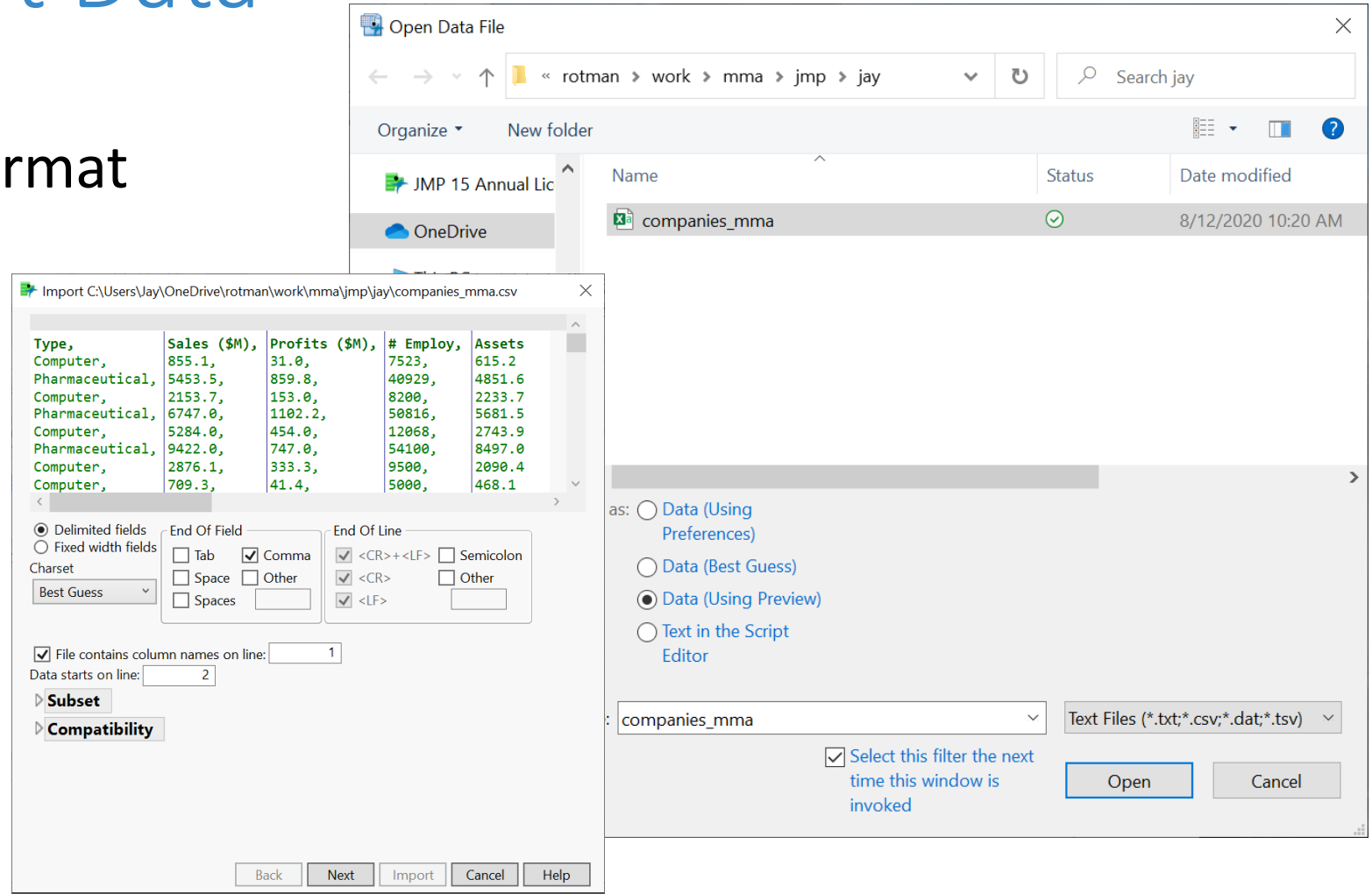
```
Type, Sales ($M), Profits ($M), # Employ, Assets  
Computer, 855.1, 31.0, 7523, 615.2  
Pharmaceutical, 5453.5, 859.8, 40929, 4851.6  
Computer, 2153.7, 153.0, 8200, 2233.7  
...
```


Your Turn (Hands-on; 5mins)

- Repeat the demo I just did
 - import the dataset to a JMP data table
 - fit y by x (Profit by # Employ)
 - remove an outlier; redo the fit
 - save the analysis script in the data table
- Challenge: run a multiple linear regression
 - predict Profits using number of employees (# Employ) and Sales

Review: Import Data

- Support many file format
 - csv
 - Excel
 - json
 - many more



<https://www.jmp.com/support/help/en/15.2/#page/jmp/import-your-data.shtml#>

Review: Data Table

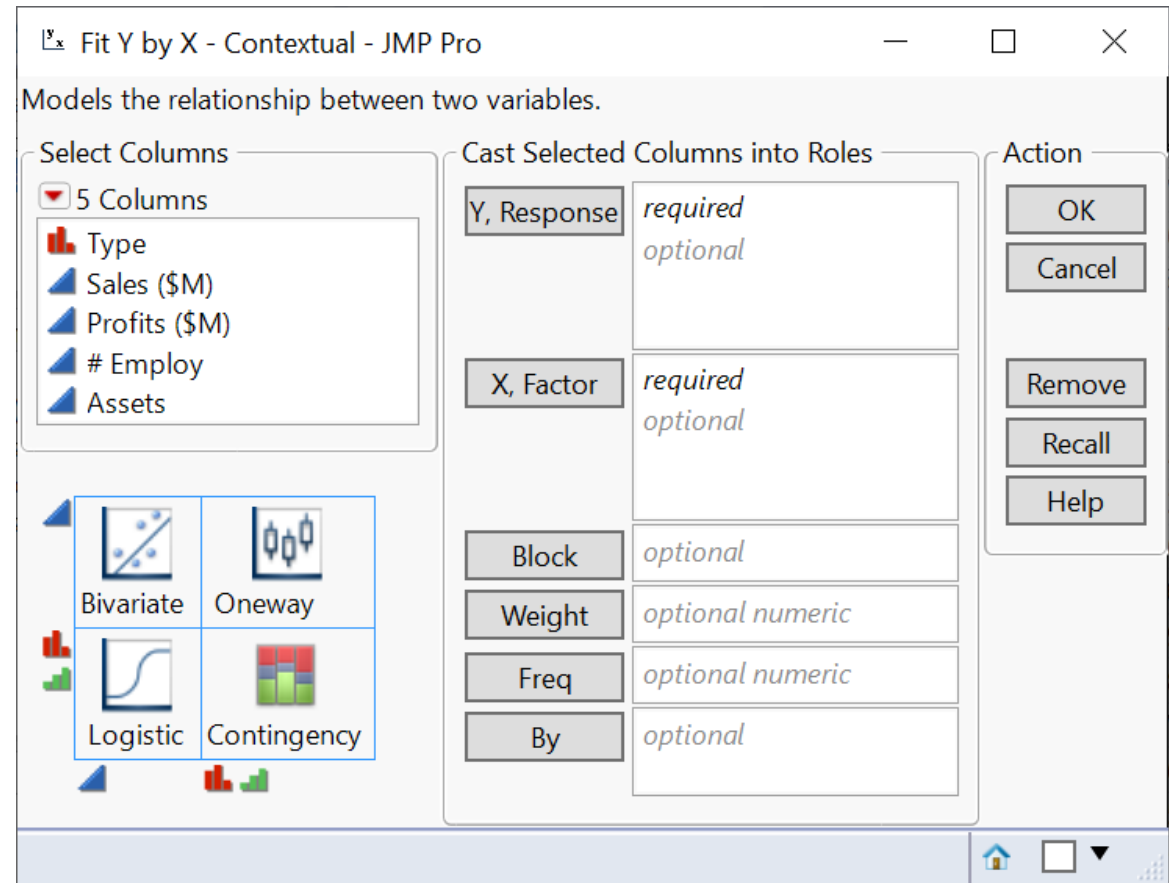
- Three panels on the left
 - table & JMP Script
 - columns
 - rows
- Column Info
- Excluding rows

The screenshot displays the JMP Pro interface with a data table named 'companies_mma'. The table has columns for 'Type', 'Sales (\$M)', 'Profits (\$M)', '# Employ', and 'Assets'. A dialog box titled '# Employ - JMP Pro' is open, showing the column name '# Employ' and its properties: Data Type is 'Numeric', Modeling Type is 'Continuous', and Format is 'Best' with a width of 12. The dialog also includes options for 'Lock', 'Use thousands separator (,)', and 'Column Properties'.

	Type	Sales (\$M)	Profits (\$M)	# Employ	Assets
16	Pharmaceutical	969.2	227.4	3418	784
17	Pharmaceutical	6698.4	1495.4	34400	6756.7
18	Computer	5956	412	56000	4500
19	Pharmaceutical	5903.7	681.1	42100	8324.8
20	Computer	2959.3	252.8	31404	5611.1
21	Pharmaceut				
22	Computer				
23	Pharmaceut				
24	Computer				
25	Computer				
26	Computer				
27	Computer				
28	Computer				
29	Pharmaceut				
30	Computer				
31	Pharmaceut				
32	Computer				

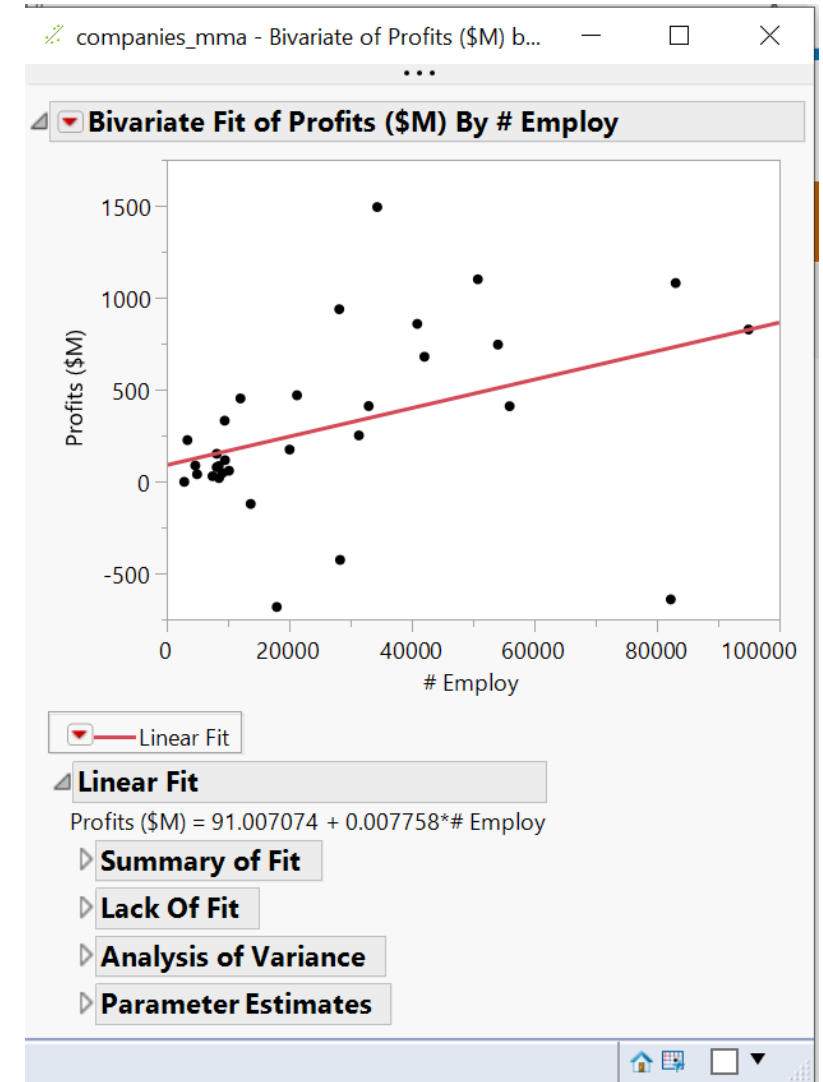
Review: Fit Y by X Platform

- Relationships between two variables
- Model depends on Y and X variable types
 - 4 main types
 - many model choices under bivariate model



Review: Analysis Report

- Analysis reports are dynamic
 - a report and its underlying data are linked
- Many actions are available under the red triangle buttons
 - redo analysis
 - save script



Review: Save Analysis

- Save data table
- Capture a script for analysis report

The screenshot shows the JMP Pro interface. The main window displays a table with columns 'Type' and 'Sales (\$M)'. The data is as follows:

Row	Type	Sales (\$M)
1	Computer	855.7
2	Pharmaceutical	5453.5
3	Computer	2153.7
4	Pharmaceutical	674.7
5	Computer	528.4
6	Pharmaceutical	942.2
7	Computer	2876.3
8	Computer	709.3
9	Computer	2952.7
10	Computer	784.7
11	Computer	1324.3
12	Pharmaceutical	4175.6
13	Computer	1189.5
14	Computer	873.6
15	Pharmaceutical	9844
16	Pharmaceutical	969.2
17	Pharmaceutical	6698.4
18	Computer	5956
19	Pharmaceutical	5903.7

The secondary window, 'Bivariate Fit of Profits (\$M) By # Employ', shows a scatter plot of Profits (\$M) versus # Employ. A red regression line is fitted to the data points. The context menu is open, and the 'Save Script' option is selected, showing a sub-menu with options: 'To Data Table...', 'To Journal', 'To Script Window', 'To Report', and 'To Clipboard'.

Review: Fit Model - Multiple Linear Reg

- Analyze > Fit Model

The screenshot shows the 'Fit Model - JMP Pro' dialog box. The 'Model Specification' section is active, showing the following settings:

- Select Columns:** 5 Columns (Type, Sales (\$M), Profits (\$M), # Employ, Assets)
- Pick Role Variables:** Y (highlighted in yellow), Profits (\$M) (optional)
- Weight:** optional numeric
- Freq:** optional numeric
- Validation:** optional
- By:** optional
- Personality:** Standard Least Squares (highlighted in yellow)
- Emphasis:** Effect Leverage
- Buttons:** Help, Run, Recall, Remove
- Keep dialog open:**
- Construct Model Effects:** Add (highlighted in yellow), Cross, Nest, Macros (dropdown), Degree (2), Attributes (dropdown), Transform (dropdown), No Intercept (checkbox)

Plan

- Session 1
 - Workflow overview
 - **Basic data manipulation**
- Session 2
 - Join data tables
 - JMP graphing
- Session 3
 - Modelling
 - JMP Journal
 - JMP Scripting Language

Data Manipulation - Overview

- Three levels
 - Columns
 - col + row operations
 - JMP creates a new table
 - Rows
 - Tables

The screenshot shows the JMP Pro interface for a data table named 'companies_mma'. The main window displays a table with the following data:

	Type	Sales (\$M)	Profits (\$M)	# Employ	Assets
1	Computer	855.1	31	7523	615.2
2	Pharmaceutical	5453.5	859.8	40929	4851.6
3	Computer	2153.7	153	8200	2233.7
4	Pharmaceutical	6747	1102.2	50816	5681.5
5	Computer	5284	454	12068	2743.9
6	Pharmaceutical	9422	747	54100	8497
7	Computer	2876.1	333.3	9500	2090.4
8	Computer	709.3	41.4	5000	468.1
9	Computer	2952.1	-680.4	18000	1860.7
10	Computer	784.7	89	4708	955.8
11	Computer	1324.3	-119.7	13740	1040.2
12	Pharmaceutical	4175.6	939.5	28200	5848
13	Computer	11899	829	95000	10075
14	Computer	873.6	79.5	8200	808
15	Pharmaceutical	9844	1082	83100	7919
16	Pharmaceutical	969.2	227.4	3418	784
17	Pharmaceutical	6698.4	1495.4	34400	6756.7
18	Computer	5956	412	56000	4500
19	Pharmaceutical	5903.7	681.1	42100	8324.8

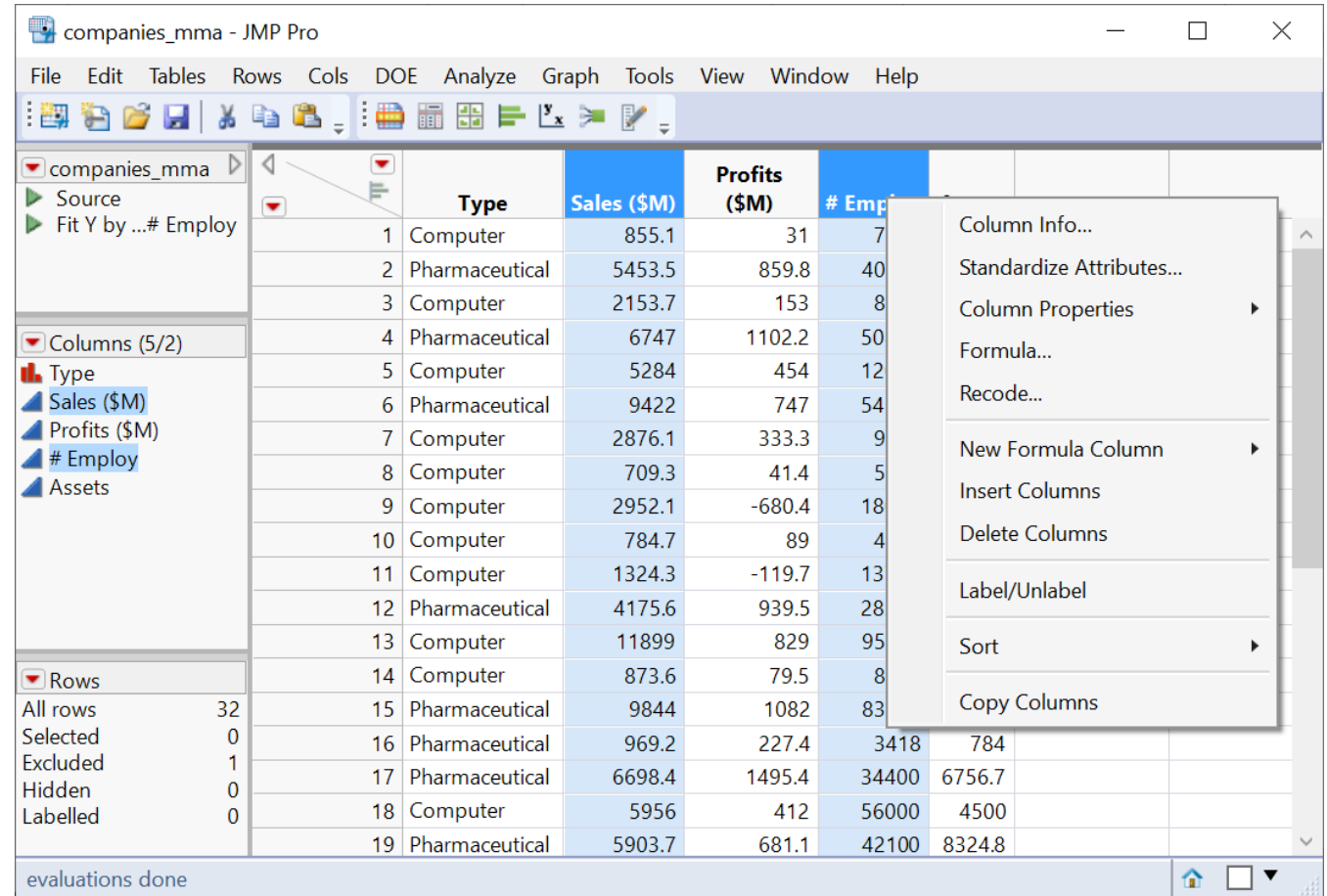
The interface also shows a sidebar with a tree view of the data structure, including 'companies_mma', 'Source', 'Fit Y by ...# Employ', 'Columns (5/0)', and 'Rows'. The 'Columns' section lists 'Type', 'Sales (\$M)', 'Profits (\$M)', '# Employ', and 'Assets'. The 'Rows' section shows 'All rows' (32), 'Selected' (0), 'Excluded' (1), 'Hidden' (0), and 'Labelled' (0).

Data Manipulation - Basics

- **Cols**
 - select columns
 - add new columns
- Rows
 - order/sort rows (see table operation)
 - filter rows (have tried excluding certain rows)
- Tables
 - subset a table
 - sort a table
 - aggregate/summarize (by group)

Select Columns

- Select/deselect columns
 - along with doing analysis
 - in data table window
- Drop/delete columns
 - in data table window
 - Tables -> Subset

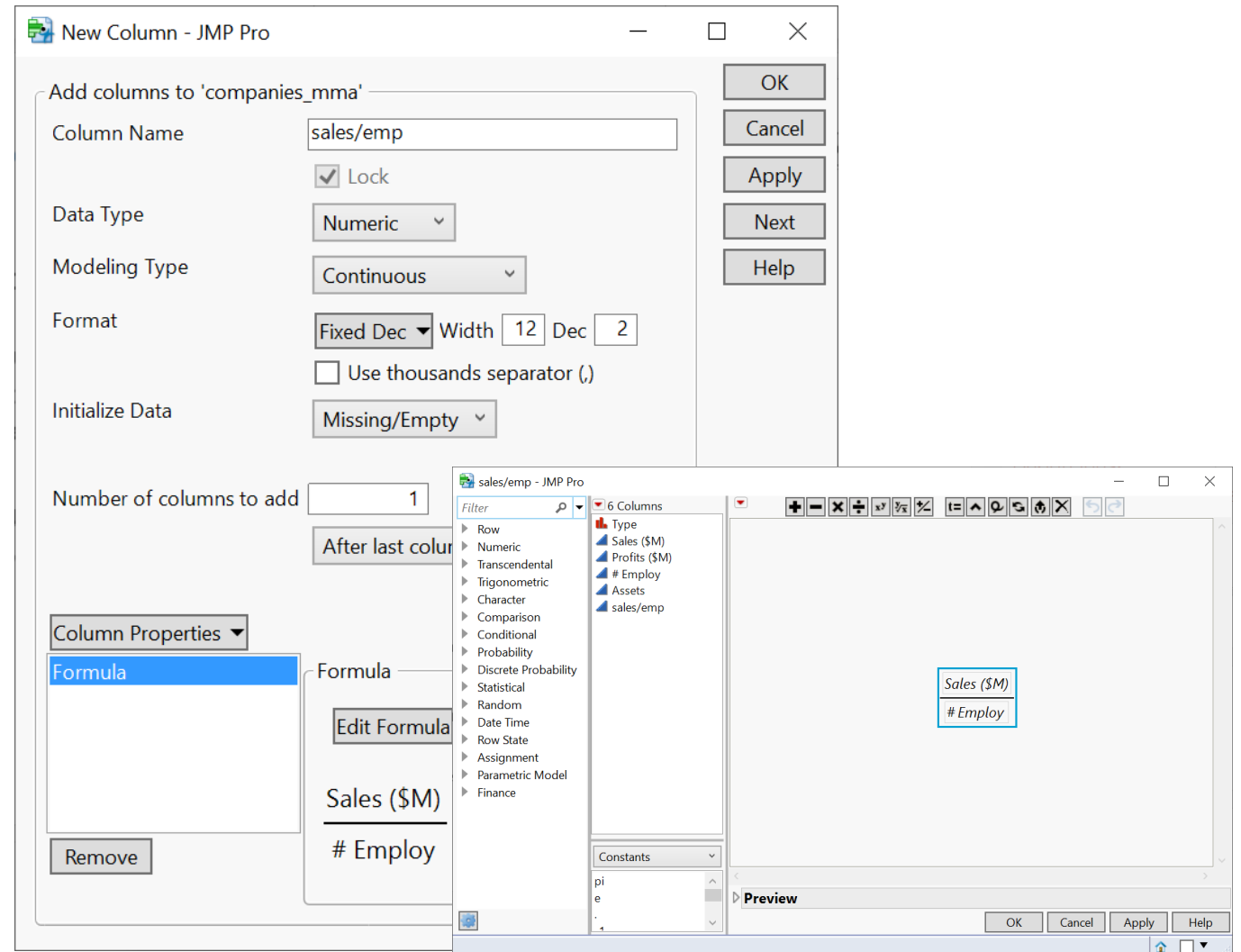


The screenshot shows the JMP Pro interface with a data table named 'companies_mma'. The table has columns: Type, Sales (\$M), Profits (\$M), # Employ, and Assets. A context menu is open over the 'Sales (\$M)' column, showing options like 'Column Info...', 'Standardize Attributes...', 'Column Properties', 'Formula...', 'Recode...', 'New Formula Column', 'Insert Columns', 'Delete Columns', 'Label/Unlabel', 'Sort', and 'Copy Columns'. The 'Columns' panel on the left shows 'Sales (\$M)', 'Profits (\$M)', and '# Employ' selected. The 'Rows' panel shows 32 rows, with 0 selected, 1 excluded, 0 hidden, and 0 labelled.

	Type	Sales (\$M)	Profits (\$M)	# Employ	Assets
1	Computer	855.1	31	7	
2	Pharmaceutical	5453.5	859.8	40	
3	Computer	2153.7	153	8	
4	Pharmaceutical	6747	1102.2	50	
5	Computer	5284	454	12	
6	Pharmaceutical	9422	747	54	
7	Computer	2876.1	333.3	9	
8	Computer	709.3	41.4	5	
9	Computer	2952.1	-680.4	18	
10	Computer	784.7	89	4	
11	Computer	1324.3	-119.7	13	
12	Pharmaceutical	4175.6	939.5	28	
13	Computer	11899	829	95	
14	Computer	873.6	79.5	8	
15	Pharmaceutical	9844	1082	83	
16	Pharmaceutical	969.2	227.4	3418	784
17	Pharmaceutical	6698.4	1495.4	34400	6756.7
18	Computer	5956	412	56000	4500
19	Pharmaceutical	5903.7	681.1	42100	8324.8

Create a New Column (Demo)

- Create a new column
 - Col name: sales/emp
 - Data type: Numeric
 - Col property: Formula
 - Formula: Sales (\$M) / # Employ
- A note on column names
 - JMP is flexible with col names
 - special symbols in col names usually not a good idea
- Formula editor



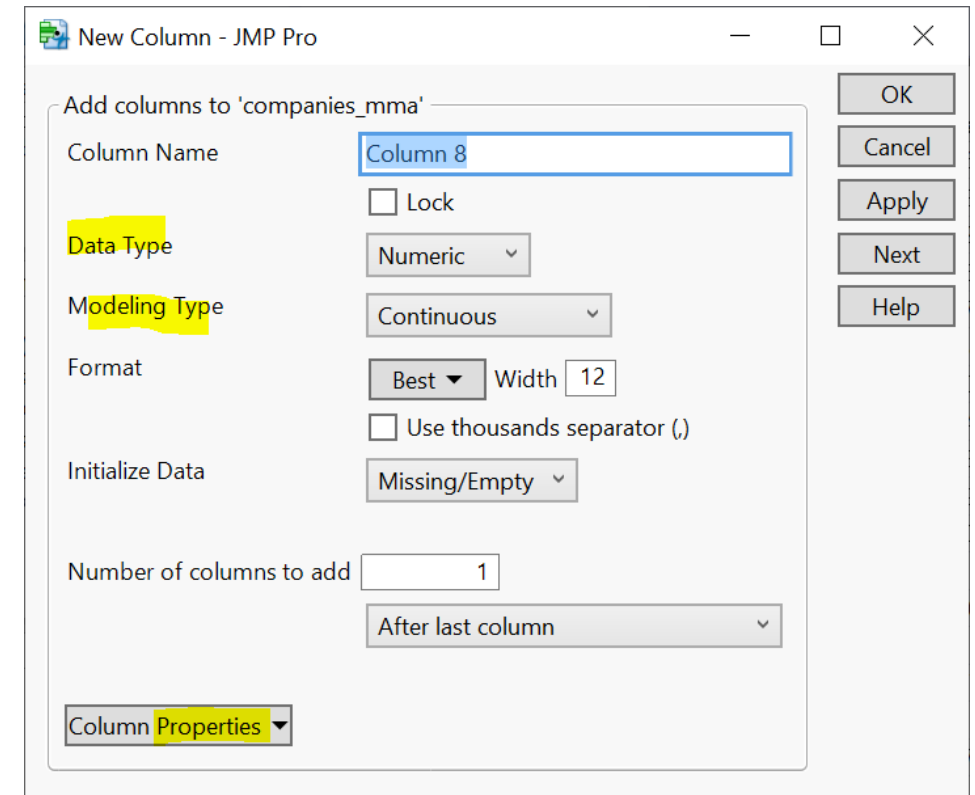
Your Turn (Hands-on)

- Create a new column
 - Col name: size
 - Data type: Character
 - Col property: formula
- It's a categorical variable
 - if # Employ < 20000, then size = "small"
 - if 20000 <= #Employ < 40000, then size = "medium"
 - if #Employ >= 40000, then size = "big"
- Save the result

The image shows two overlapping windows from the JMP Pro software. The top window is titled 'New Column - JMP Pro' and is used for creating a new column. The 'Column Name' field is set to 'size', the 'Data Type' is 'Character', and the 'Modeling Type' is 'Nominal'. The 'Initialize Data' is set to 'Missing/Empty'. The 'Number of columns to add' is set to 1, and the location is 'After last column'. The bottom window is titled 'Column 6 - JMP Pro' and shows the 'Column Properties' dialog for the newly created column. The 'Formula' tab is selected, and the formula editor contains the following conditional logic:
$$\text{If} \left(\begin{array}{l} \# \text{Employ} < 20000 \Rightarrow \text{"small"} \\ 20000 \leq \# \text{Employ} < 40000 \Rightarrow \text{"medium"} \\ \text{else} \Rightarrow \text{"big"} \end{array} \right)$$
 The 'Filter' pane on the left shows the 'If' function selected under the 'Conditional' category. The 'Preview' pane at the bottom right shows the resulting formula.

Review: Create a Column

- Data type
 - Numeric
 - Character
- Modelling type
 - Continuous
 - Ordinal (e.g. age: 1, 2, 3, ...; Monday, Tuesday, ...)
 - Nominal (e.g. male, female)
- Column Property
 - Formula



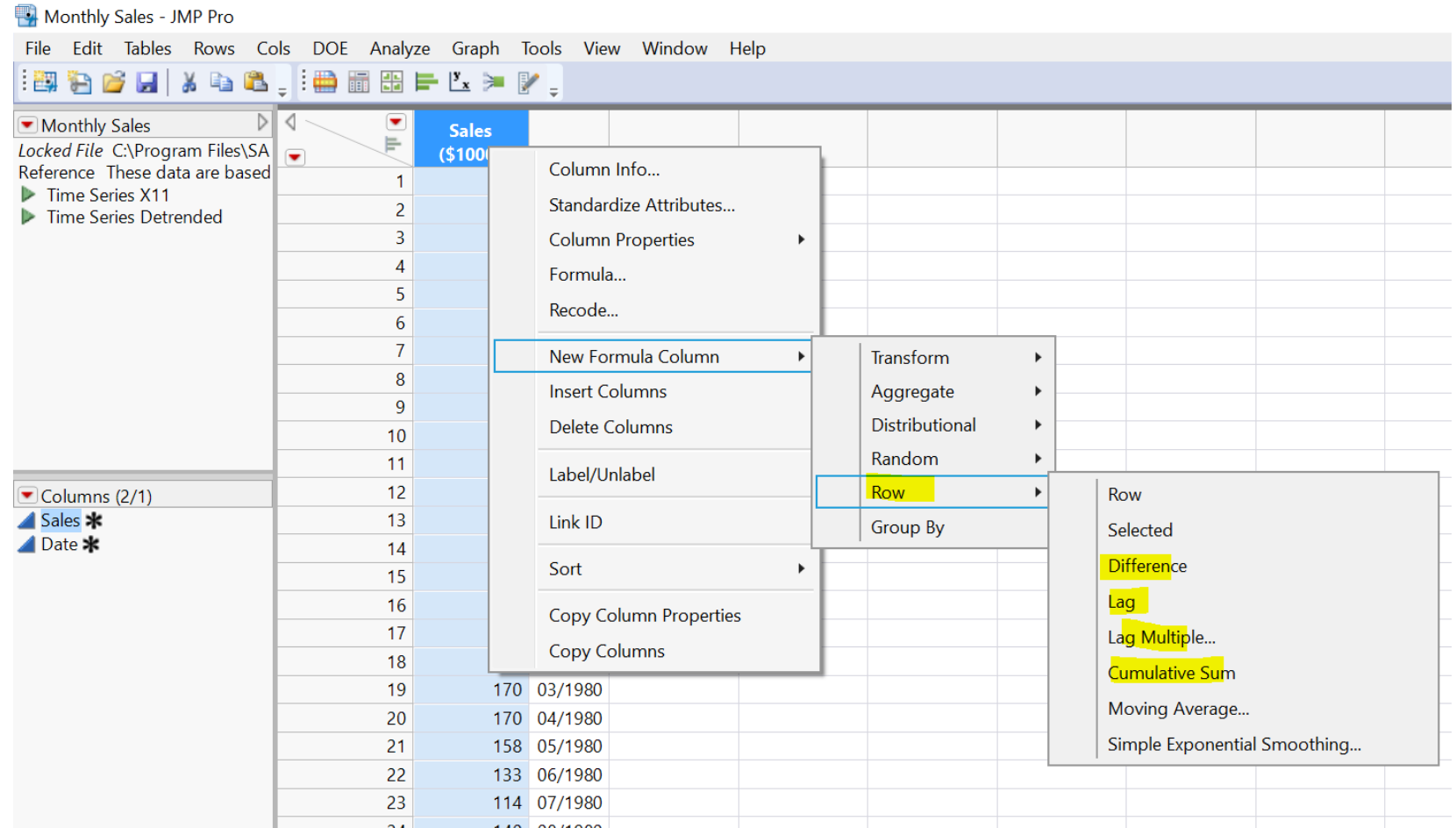
Review: Create a Column Shortcut (Demo)

The screenshot shows the JMP Pro interface with a data table named 'companies_mma'. The table has columns for 'Type', 'Sales', 'Profits', 'sales/emp', and 'size'. A context menu is open over the 'Sales' column, and the 'New Formula Column' option is selected. The sub-menu for 'New Formula Column' is also visible, showing options like 'Transform', 'Combine', 'Aggregate', 'Distributional', 'Random', 'Row', and 'Group By'.

Row	Type	Sales	Profits	sales/emp	size		
1	Computer			0.1136647614	small		
2	Pharmaceutical			0.1332429329	big		
3	Computer			0.2626463415	small		
4	Pharmaceutical			0.1327731423	big		
5	Computer			0.437852171	small		
6	Pharmaceutical			0.1741589649	big		
7	Computer						
8	Computer						
9	Computer						
10	Computer						
11	Computer						
12	Pharmaceutical						
13	Computer						
14	Computer						
15	Pharmaceutical						
16	Pharmaceutical	969.2	227.4	3418	784	0.283557636	small
17	Pharmaceutical	6698.4	1495.4	34400	6756.7	0.1947209302	medium
18	Computer	5956	412	56000	4500	0.1063571429	big
19	Pharmaceutical	5002.7	681.1	42100	8224.8	0.1402204028	big

Review: Create a Column Shortcut

- Some useful transformations for time series data manipulation

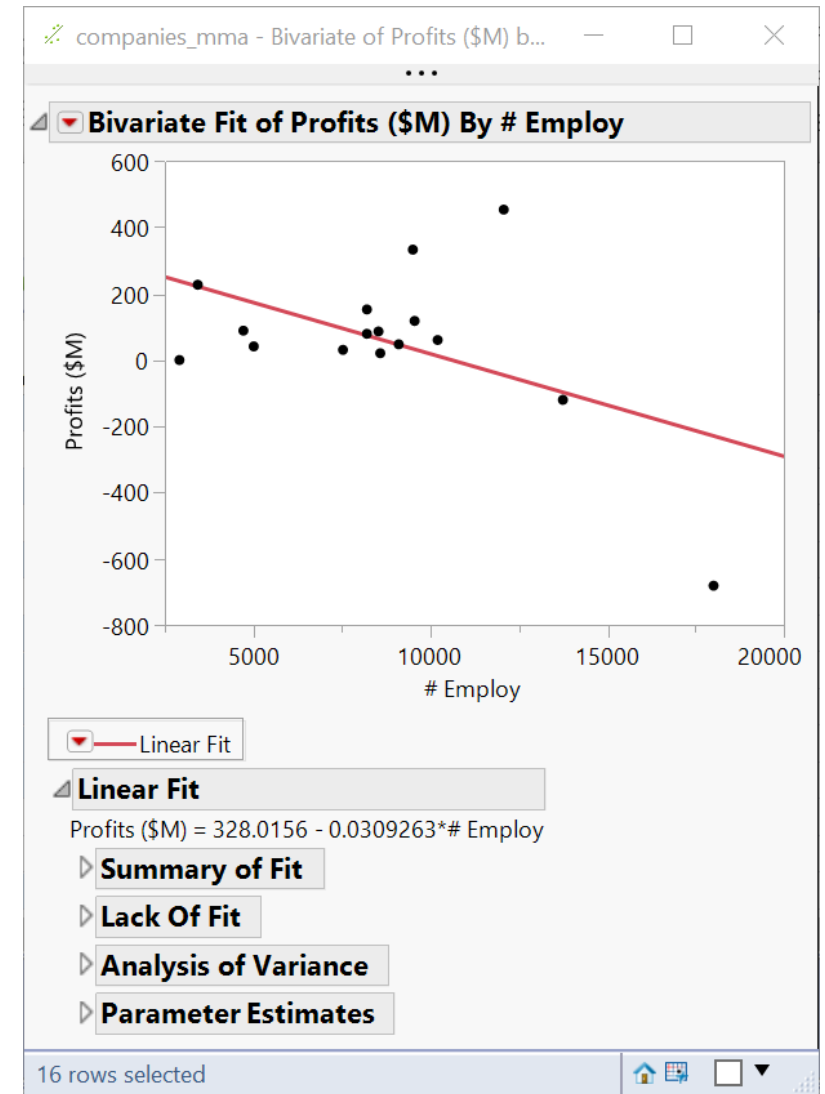
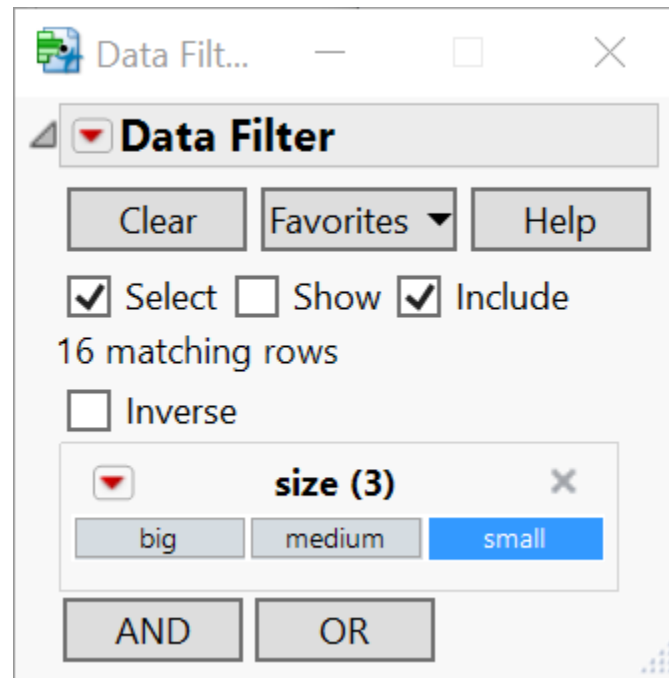


Data Manipulation - Basics

- Cols
 - select columns
 - add new columns
- **Rows**
 - order/sort rows (see table operation)
 - **filter rows** (have tried excluding certain rows)
- Tables
 - subset a table
 - sort a table
 - aggregate/summarize (by group)

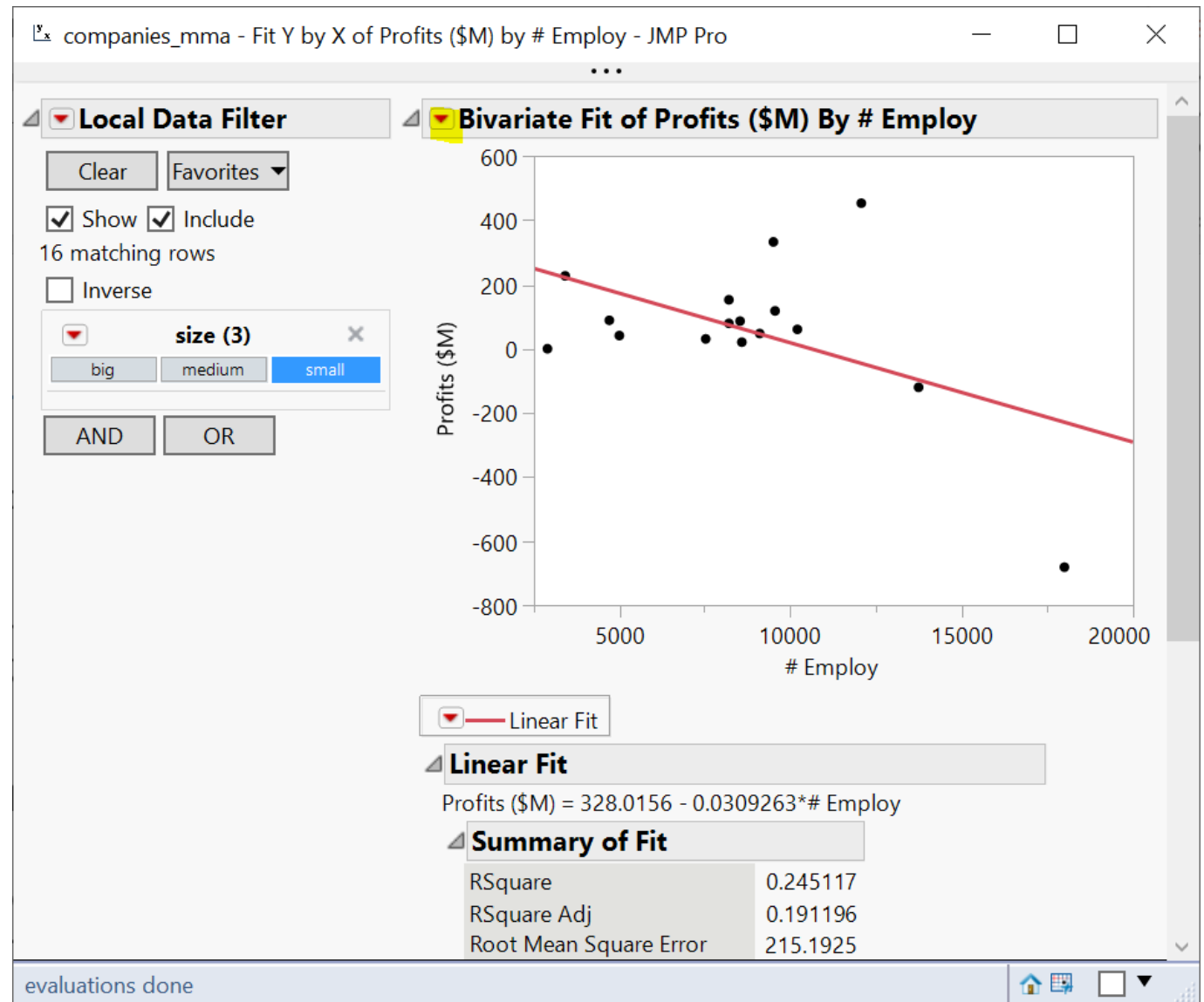
Filter Rows (Demo)

- Data Filter (global)
 - Affects all linked analysis report
- Rows > Data Filter
 - filter on the **size** col



Filter Rows (Demo)

- Local Data Filter
 - only affects the report where the filter is enabled



Your Turn (Hands-on)

- Create a data filter based on size & Type
 - either global or local
- Explore the filter in a linear regression analysis
- Save the filter script
- Explore adding a continuous col into the filter

Data Filter

Clear Favorites Help

Select Show Include
14 matching rows

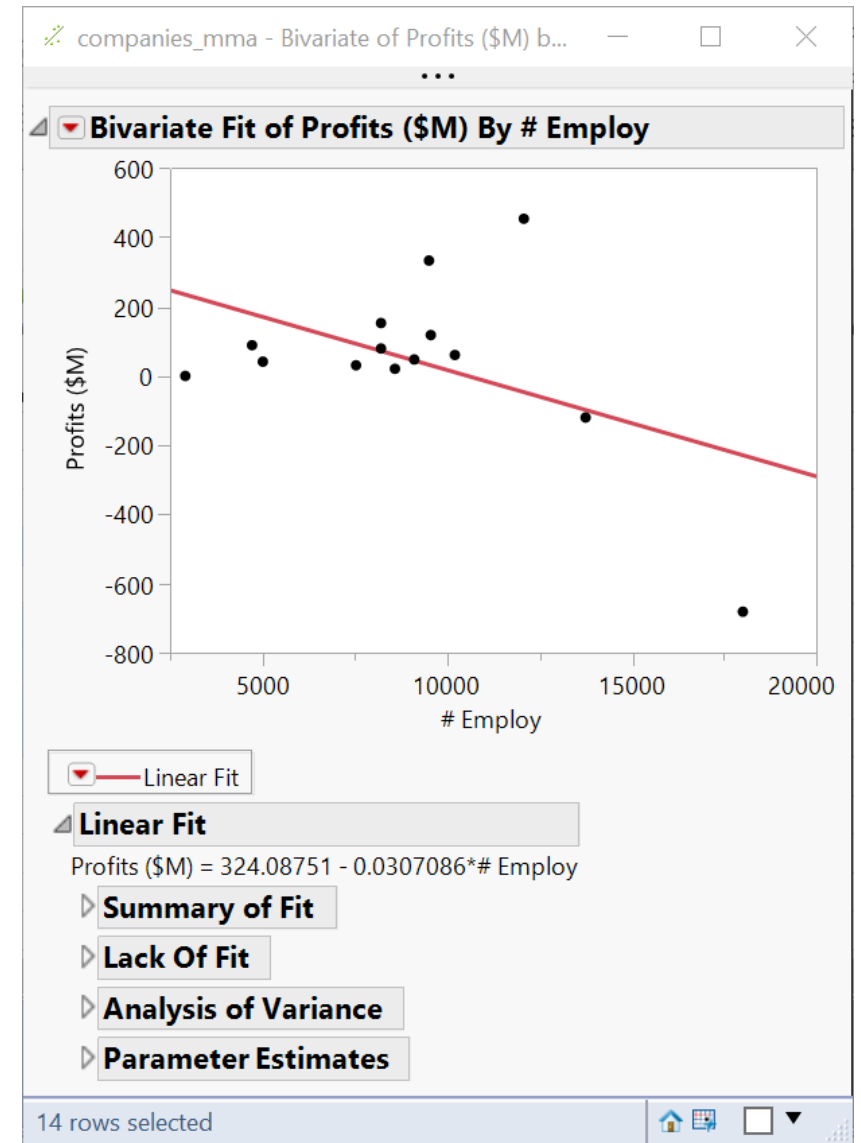
Inverse

size (3)
big medium **small**

Type (2)
Computer Pharmaceutical

AND OR

evaluations done



Filter Rows (Demo)

- Rows > Row Selection > Select Where...

The screenshot shows the JMP Pro interface with the 'Rows' menu open. The 'Row Selection' option is selected, and the 'Select Where...' dialog box is displayed. The dialog box contains the following options:

- Select All Rows
- Select Excluded
- Select Hidden
- Select Labeled
- Select Where... (Ctrl+Shift+W)
- Select Matching Cells
- Select All Matching Cells
- Select Randomly...
- Select Dominant...
- Select Duplicate Rows
- Name Selection in Column...

The background data table is partially visible, showing columns for various metrics and categories.

-680.4	18000	1860.7	0.1640055556	small
89	4708	955.8	0.1666737468	small
-119.7	13740	1040.2	0.0963828239	small
939.5	28200	5848	0.148070922	medium
829	95000	10075	0.1252526316	big
79.5	8200	808	0.1065365854	small
1082	83100	7919	0.1184596871	big
227.4	3418	784	0.283557636	small
495.4	34400	6756.7	0.1947209302	medium
412	56000	4500	0.1063571429	big

Data Manipulation - Basics

- Cols
 - select columns
 - add new columns
- Rows
 - order/sort rows (see table operation)
 - filter rows (have tried excluding certain rows)
- **Tables**
 - subset a table
 - sort a table
 - aggregate/summarize (by group)

Subset a table (demo)

The screenshot shows the JMP Pro interface with the 'Tables' menu open, highlighting the 'Subset' option. The 'Subset - JMP Pro' dialog box is open, showing the following settings:

- Subset by
- Rows:
 - All rows
 - Selected Rows
 - Random - sampling rate :
 - Random - sample size :
 - Stratify
- Columns:
 - All columns Selected columns
 - Keep by columns
- Output table name:
- Link to original data table
- Copy formula
- Suppress formula evaluation
- Keep dialog open
- Save Script to Source Table

The background data table is partially visible, showing columns for Sales (\$M), Profits (\$M), # Employ, and Assets.

	Sales (\$M)	Profits (\$M)	# Employ	Assets
	855.1	31	7523	615.2
utical	5453.5	859.8	40929	4851.6
	2153.7	153	8200	2233.7
utical	6747	1102.2	50816	5681.5
	5284	454	12068	2743.9
utical	9422	747	54100	8497
	2876.1	333.3	9500	2090.4
	709.3	41.4	5000	468.1
	2952.1	-680.4	18000	1860.7
	784.7	89	4708	955.8
	1324.3	-119.7	13740	1040.2
utical	4175.6	939.5	28200	5848
	11899	829	95000	10075
	873.6	79.5	8200	808
	9844	1082	83100	7919
	969.2	227.4	3418	784
	6698.4	1495.4	34400	6756.7
	5956	412	56000	4500
	5903.7	681.1	42100	8324.8

Sort Table / Order Rows (demo)

- Order rows by
 - a single column
 - multiple columns

The screenshot displays the JMP Pro interface with the 'Sort' menu open. The 'Sort' option is highlighted, and a tooltip reads 'Sort rows by specified columns.' The background shows a data table with columns: Type, Sales (\$M), Profits (\$M), # Employ, Assets, sales/emp, and size. The 'Sort - JMP Pro' dialog box is open, showing the following configuration:

- Select Columns:** 7 Columns (Type, Sales (\$M), Profits (\$M), # Employ, Assets, sales/emp, size)
- Action:** By (size), Remove (Sales (\$M) optional)
- Copy formula
- Suppress formula evaluation
- Replace table
- Output table name: []
- Keep dialog open
- Save Script to Source Table

Type	Sales (\$M)	Profits (\$M)	# Employ	Assets	sales/emp	size
Pharmaceutical	7523	615.2	40929	4851.6	0.1136647614	small
Pharmaceutical	5453.5	859.8	8200	2233.7	0.2626463415	small
Computer	6747	1102.2	50816	5681.5	0.1327731423	big

Your Turn (Hands-on)

- What's the maximum Sales (\$M) a “small” firm makes?

Aggregate / Summarize (by Group) (Demo)

- Tables > Summary
- Find average profit by Type

companies_mma - JMP Pro

File Edit **Tables** Rows Cols DOE Analyze Graph

Summary

Request Summary Statistics by Grouping Columns.

Select Columns

7 Columns

- Type
- Sales (\$M)
- Profits (\$M)
- # Employ
- Assets
- sales/emp
- size

Include marginal statistics

For quantile statistics, enter value (%)

25

statistics column name format

stat(column)

Output table name:

Link to original data table

Prompt to save when closing summary tables

Keep dialog open

Save Script to Source Table

Statistics

Mean(Profits (\$M)) optional

Group

Type optional

Subgroup

optional

Freq

optional

Weight

optional

Action

OK

Cancel

Remove

Recall

Help

Columns

- Type
- Sales (\$M)
- Profits (\$M)
- # Employ
- Assets
- sales/emp
- size

Rows

All rows	32	15	Pharmaceutical
Selected	0	16	Pharmaceutical
Excluded	1	17	Pharmaceutical
Hidden	0	18	Computer
Labelled	0	19	

evaluations done

Your Turn (Hands-on)

- Find total sales by Type and Size

The screenshot displays the JMP Pro interface. The main window, titled "companies_mma By (Type, size) - JMP Pro", shows a data table with the following columns: Type, size, N Rows, and Sum(Sales (\$M)). The data is grouped by Type and size, with rows numbered 1 through 6. The Summary dialog box is open, showing the "Request Summary Statistics by Grouping Columns" options. The "Select Columns" list includes Type, Sales (\$M), Profits (\$M), # Employ, Assets, sales/emp, and size. The "Statistics" dropdown is set to "Sum(Sales (\$M))". The "Group" dropdown is set to "Type" and "size". The "Subgroup" dropdown is set to "optional". The "Freq" and "Weight" dropdowns are also set to "optional". The "Output table name" field is empty. The "Link to original data table" checkbox is checked.

	Type	size	N Rows	Sum(Sales (\$M))
1	Computer	big	3	18951.9
2	Computer	medium	2	6037.7
3	Computer	small	14	24612.8
4	Pharmaceutical	big	5	37370.2
5	Pharmaceutical	medium	5	21305.3
6	Pharmaceutical	small	2	2167.5

Summary - JMP Pro

Request Summary Statistics by Grouping Columns.

Select Columns

7 Columns

- Type
- Sales (\$M)
- Profits (\$M)
- # Employ
- Assets
- sales/emp
- size

Include marginal statistics

For quantile statistics, enter value (%)

25

statistics column name format

stat(column)

Output table name:

Link to original data table

Prompt to save when closing summary tables

Keep dialog open

Save Script to Source Table

Statistics: Sum(Sales (\$M)) optional

Group: Type, size optional

Subgroup: optional

Freq: optional

Weight: optional

Action: OK, Cancel, Remove, Recall, Help

Aggregate / Summarize (by Group) (Demo)

- Analyze > Tabulate
- Tabulate (red triangle) > Make into data table

The screenshot shows the JMP Pro interface with the 'Analyze' menu open and 'Tabulate' selected. The 'Tabulate' dialog box is open, showing a list of columns and statistics. A preview table shows the resulting data structure.

Columns (7/0)

- Type
- Sales (\$M)
- Profits (\$M)
- # Employ
- Assets
- sales/emp
- size

Statistics

- N
- Mean
- Std Dev
- Min
- Max
- Range
- % of Total
- N Missing
- N Categories
- Sum
- Sum Wgt
- Variance
- Std Err
- CV
- Median
- Geom... Mean
- Interq... Range
- Quantiles
- Column %
- Row %
- All

Preview Table:

Type	size	Sum
Computer	big	18951.9
	medium	6037.7
	small	24612.8
Pharmaceutical	big	37370.2
	medium	21305.3
	small	2167.5

1 row has been excluded.