**Rotman**

# INTRO TO JMP – PART 1

Bootcamp

August 2, 2023   Prepared by Jay / TDMDAL

# What is JMP

- Predictive analytics software from SAS Institute
  - data manipulation
  - visualization
  - statistical & machine learning modeling
  - reporting

- Intuitive point-and-click interface for beginners
  - Analytics with zero/minimum coding

- Flexible and extensible for advanced users
  - JMP Scripting Language (JSL) for automating or extending point-and-click functionality
  - Connect to Database engine, Matlab, R, Python, Excel, Web API, etc.
  - Connect to the richness of SAS: retrieve SAS data and submit SAS code

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 or rapid prototyping
    - even you plan to use more advanced tools later

- Guide you to learn other tools
  - How do I achieve this JMP data manipulation/modeling/visualization in Python or R?
  - Try to separate the concepts from the software

# Learning by Doing

- This workshop is 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

- We will mostly focus
  - On learning the tools to do
    - data manipulation
    - visualization and report
    - simple modeling
  - Not on statistical or modeling theories

# Learning Resources (jmp.com)

- JMP Academic (Students) 👍
  - JMP Basics (great beginner videos!)
  - Learning Resources for Students (learning library, academic webinar, etc.)

- JMP User Community

- JMP Documentation Library (many examples using included datasets)
  - Getting started document: Discovering JMP
  - Basics
    - Using JMP (data table)
    - Basic Analysis
    - Essential Graphing
  - Specific topics
    - Fitting Linear Models
    - Predictive and Specialized Modeling

# Plan

- Session 1
  - **Workflow overview**
  - Basic data manipulation


- Session 2
  - Join data tables
  - JMP graphing


- Session 3
  - Modelling
  - JMP Journal
  - JMP Scripting Language

# Overview: A Typical Workflow



**Data Manipulation**
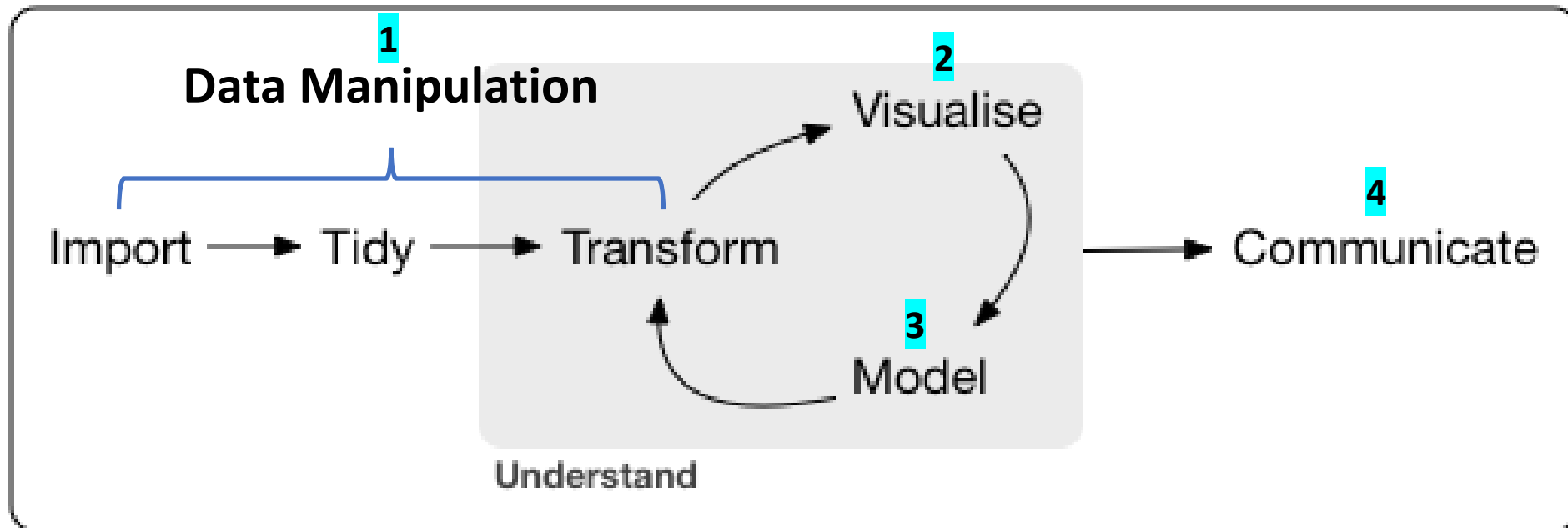
**point-and-click & programming**

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

# Overview: JMP Dynamic Link

**Data Manipulation**

Import → Tidy → **D** Transform → **D** Visualise

**D** Visualise → **D** Model

Model → **D** Transform

Understand

Transform/Visualise/Model → **D** → Communicate

**JMP Report**
**JMP Journal**
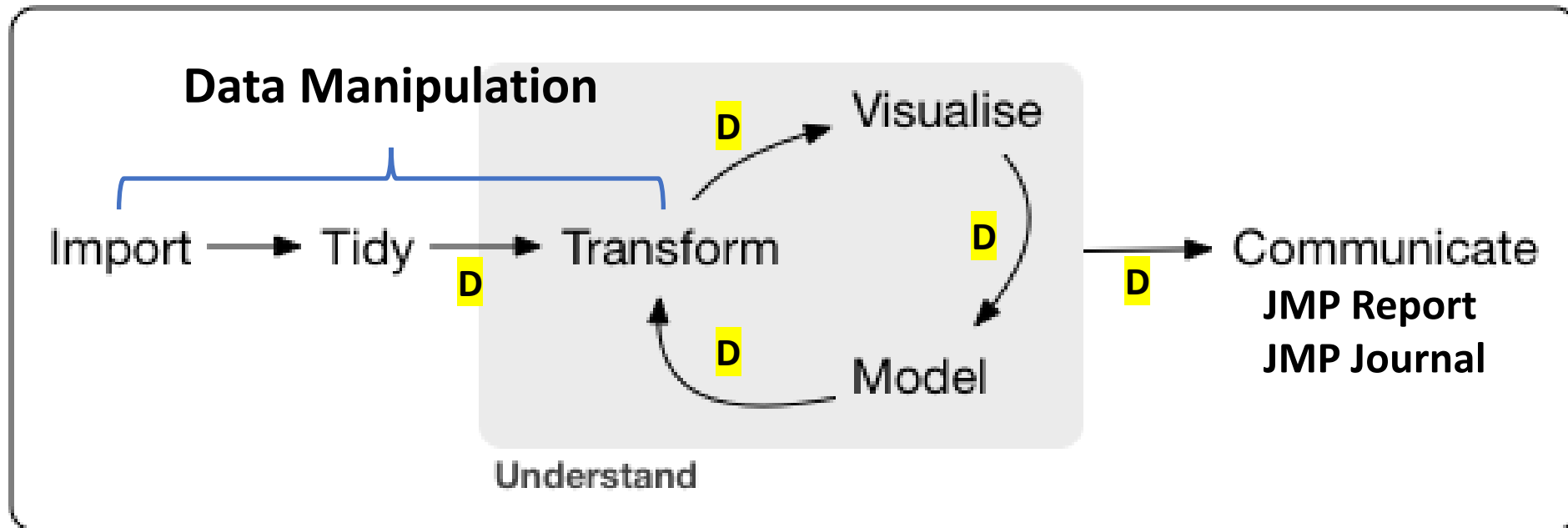
**point-and-click & programming**

**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)
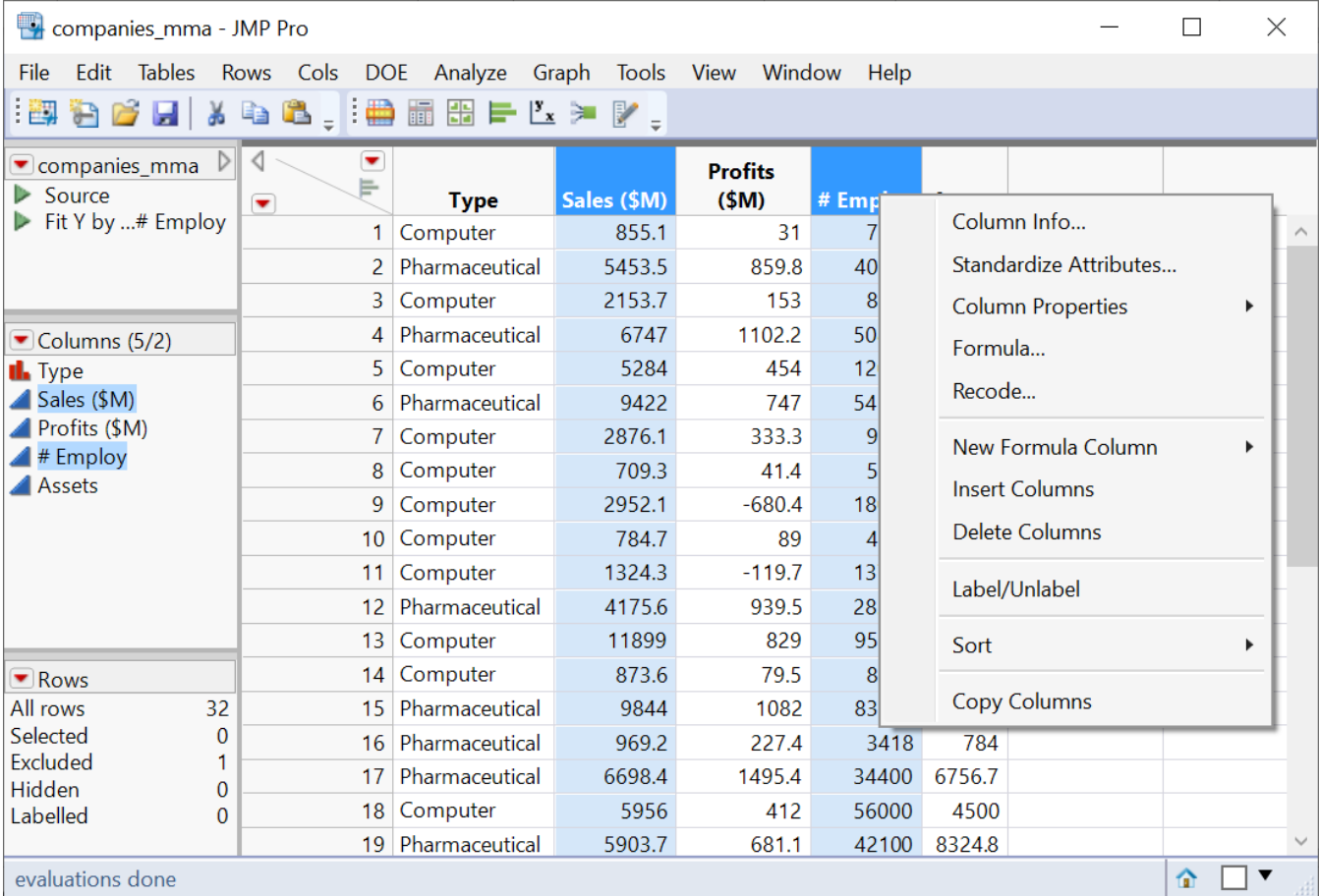
- Data (data/basics/companies_mma.csv)

```
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)
  - exclude 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/17.1/#page/jmp/import-your-data.shtml#
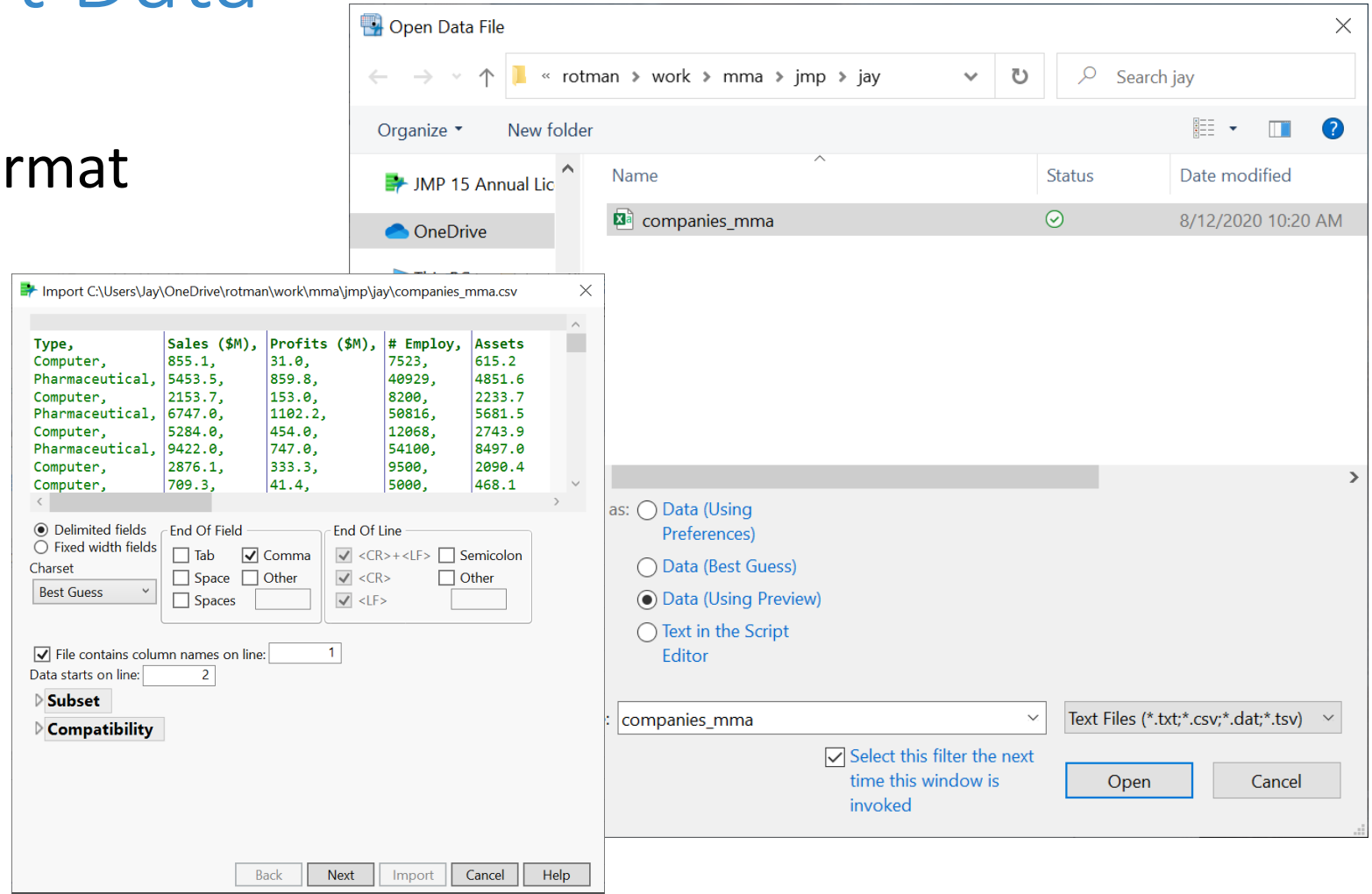
# Review: Data Table

- Three panels on the left
  - Table (JMP Script)
  - Columns
  - Rows

- Column Info

- Excluding rows



https://www.jmp.com/support/help/en/17.1/#page/jmp/work-with-data-tables.shtml#

# 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
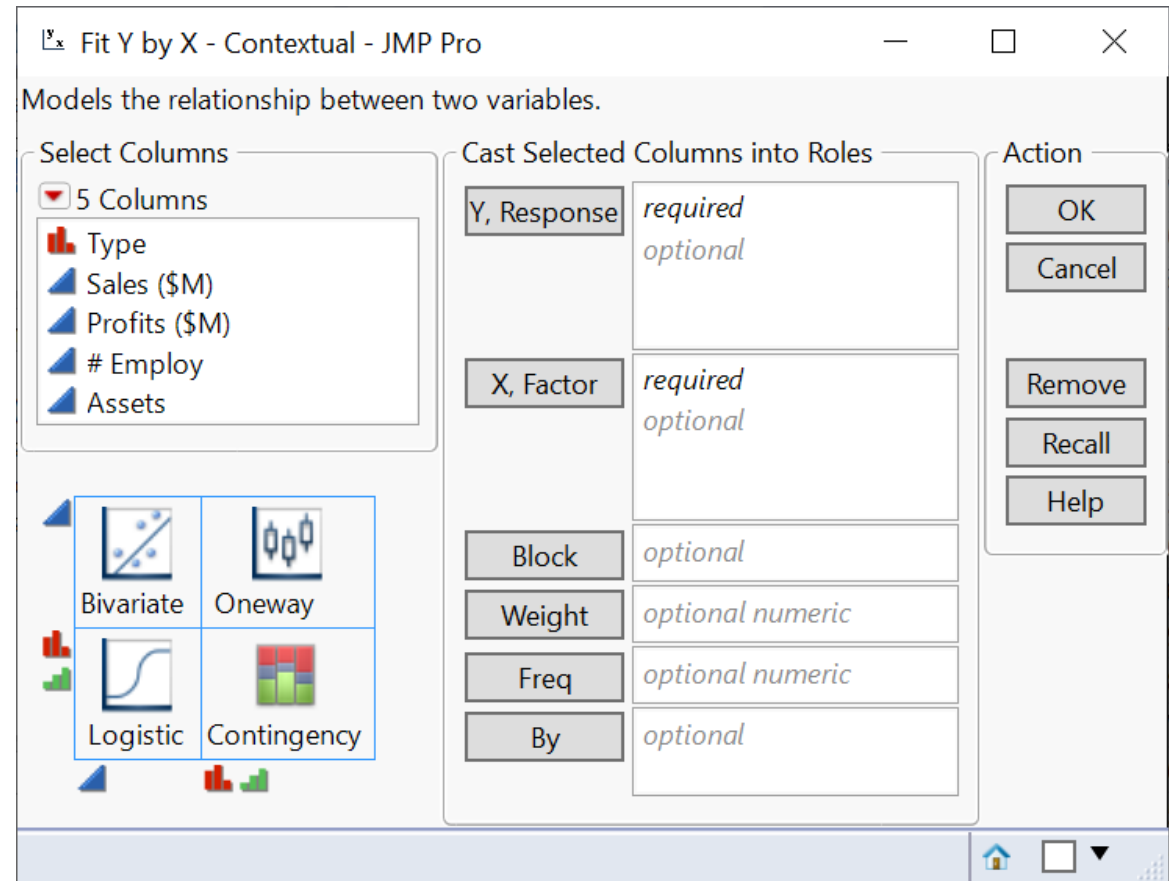
# Review: Analysis Report

- 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

- Right click gives many actions too
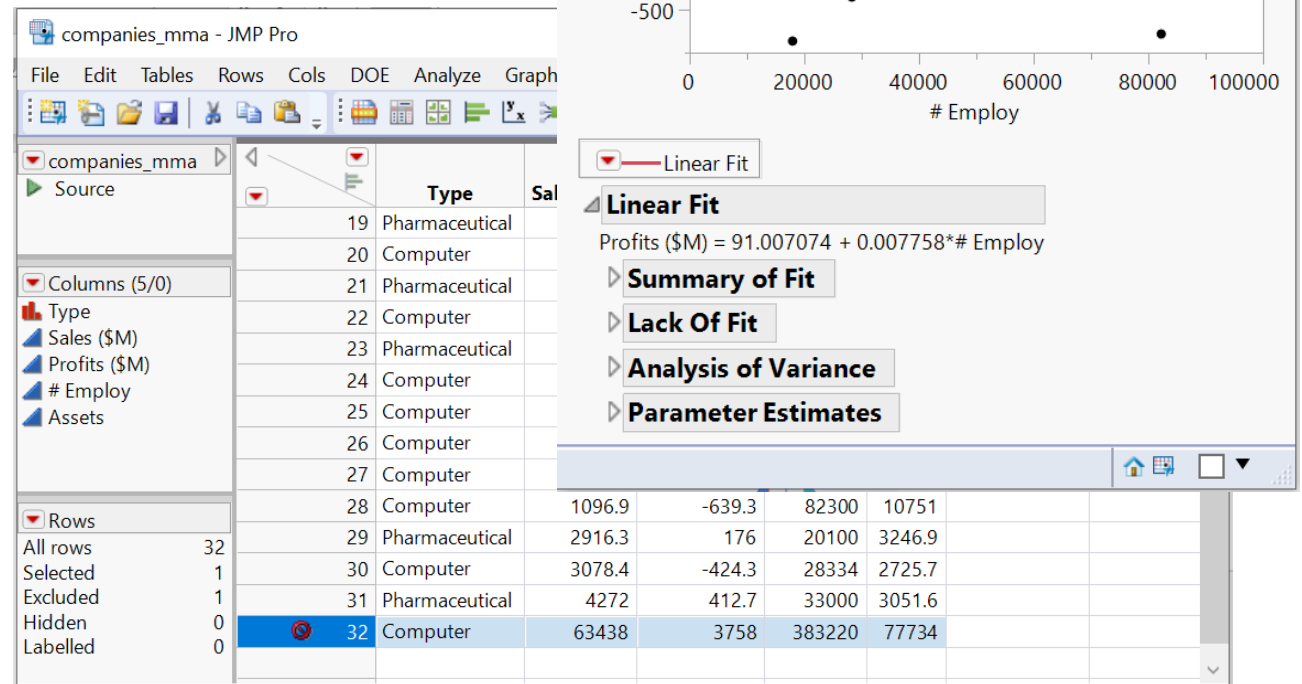
# Review: Save Analysis

- Save data table

- Capture a script for analysis report

# Review: Fit Model - Multiple Linear Reg

- Analyze > Fit Model

# 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
  - Rows
  - Tables
    - col + row operations
    - JMP creates a new table

# Data Manipulation - Basics

- **Cols**
  - select or de-select columns
  - drop/delete columns
  - add new columns


- Rows
  - order/sort rows (see table operation)
  - filter rows (we have tried excluding certain rows)


- Tables
  - subset a table
  - sort a table
  - aggregate/summarize (by group)

# Select and Delete Columns

- Select or deselect columns
  - along with doing analysis
  - in data table window

- Drop/delete columns
  - in data table window
  - Tables -> Subset
    - A new data table will be created

# 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
  - In general, special symbols in col names is not a good idea

- Formula editor
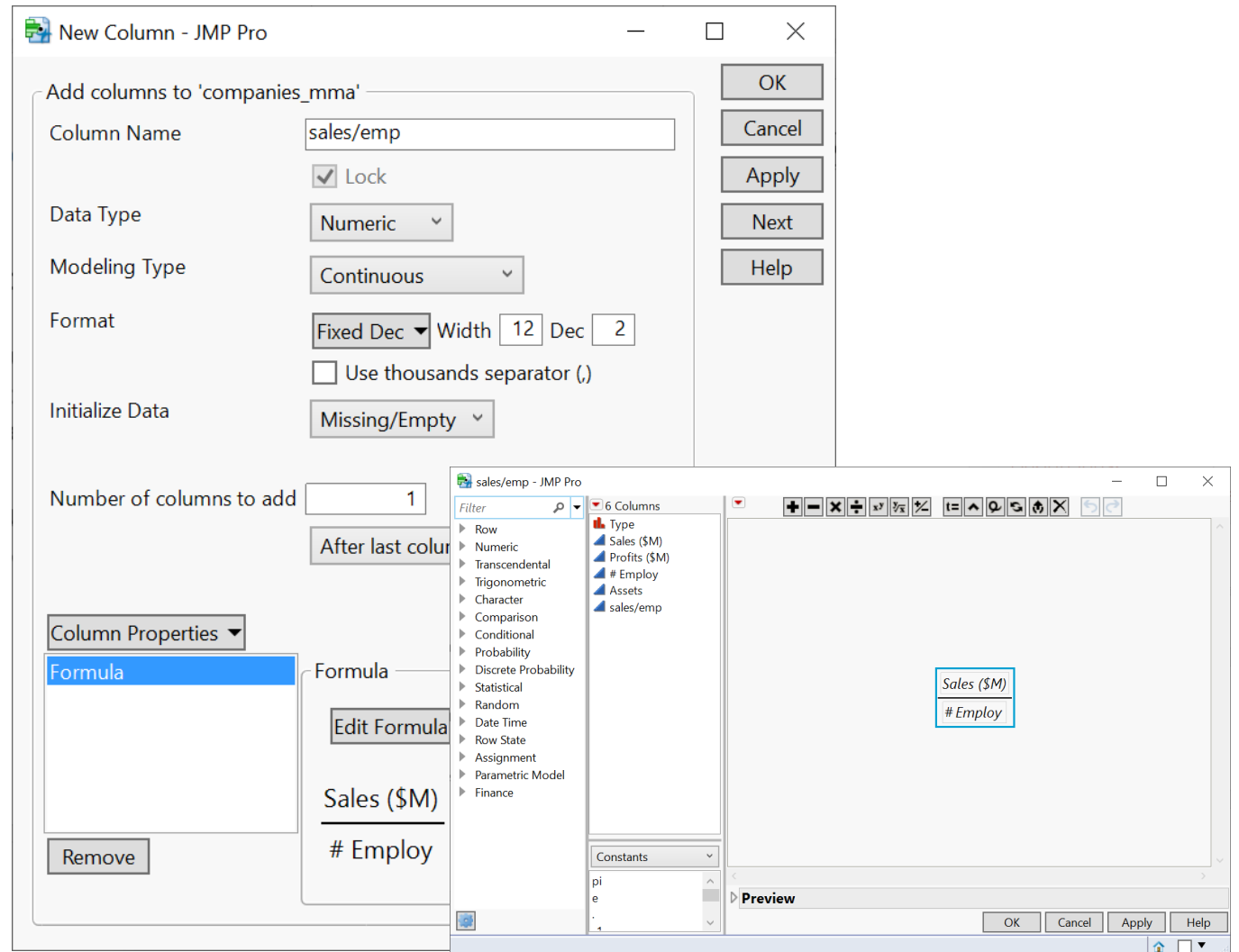
# Your Turn (Hands-on)

- 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

# 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

https://www.jmp.com/support/help/en/17.1/#page/jmp/about-the-column-info-window.shtml#

# Review: Create a Column Shortcut (Demo)

# Review: Create a Column Shortcut

- Some useful transformations for time series manipulation

# Data Manipulation - Basics

- Cols
    - select or de-select columns
    - drop/delete columns
    - add new columns


- **Rows**
    - order/sort rows (see table operation)
    - filter rows (we 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



https://www.jmp.com/support/help/en/17.1/#page/jmp/the-data-filter.shtml#
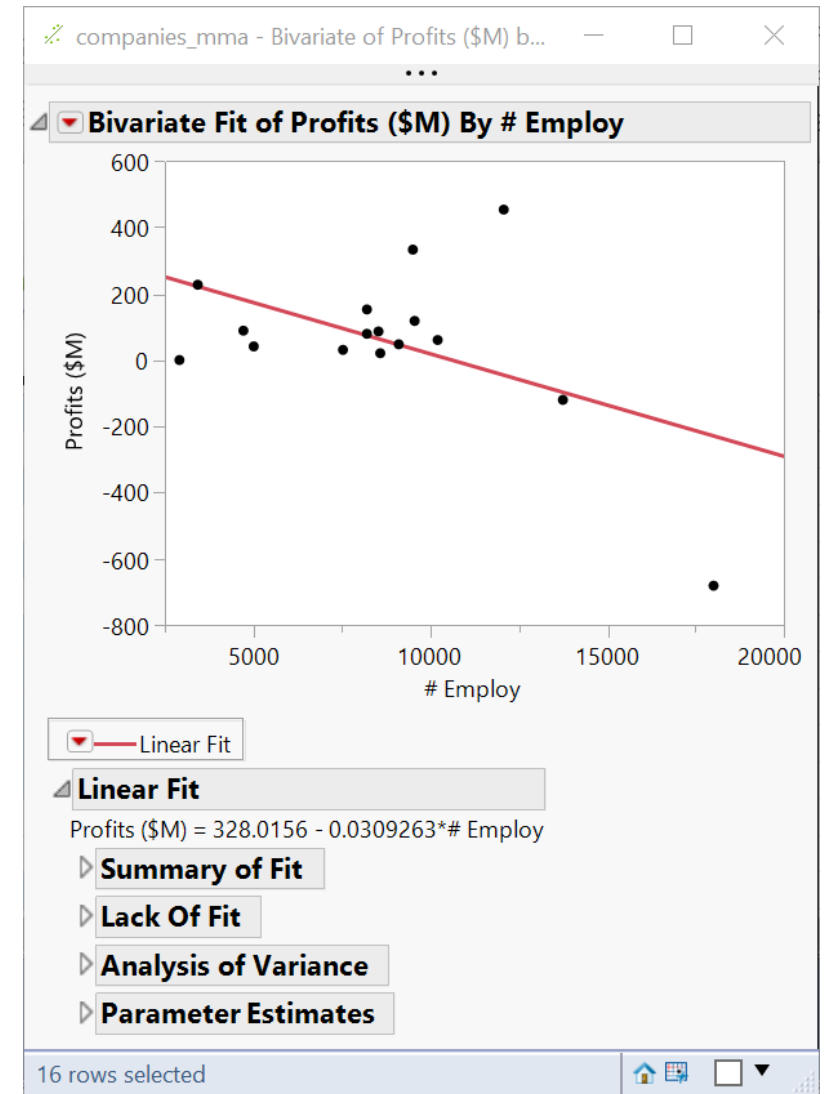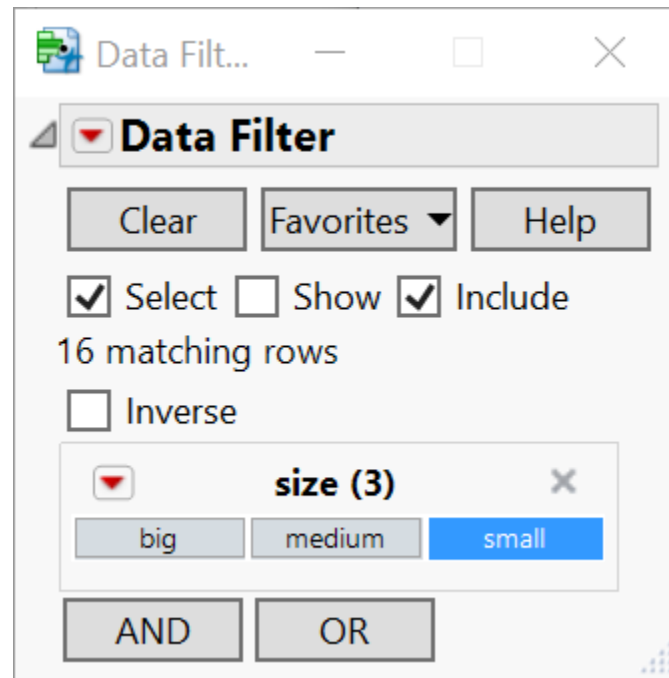
# Filter Rows (Demo)

- Local Data Filter
  - only affects the report where the filter is enabled
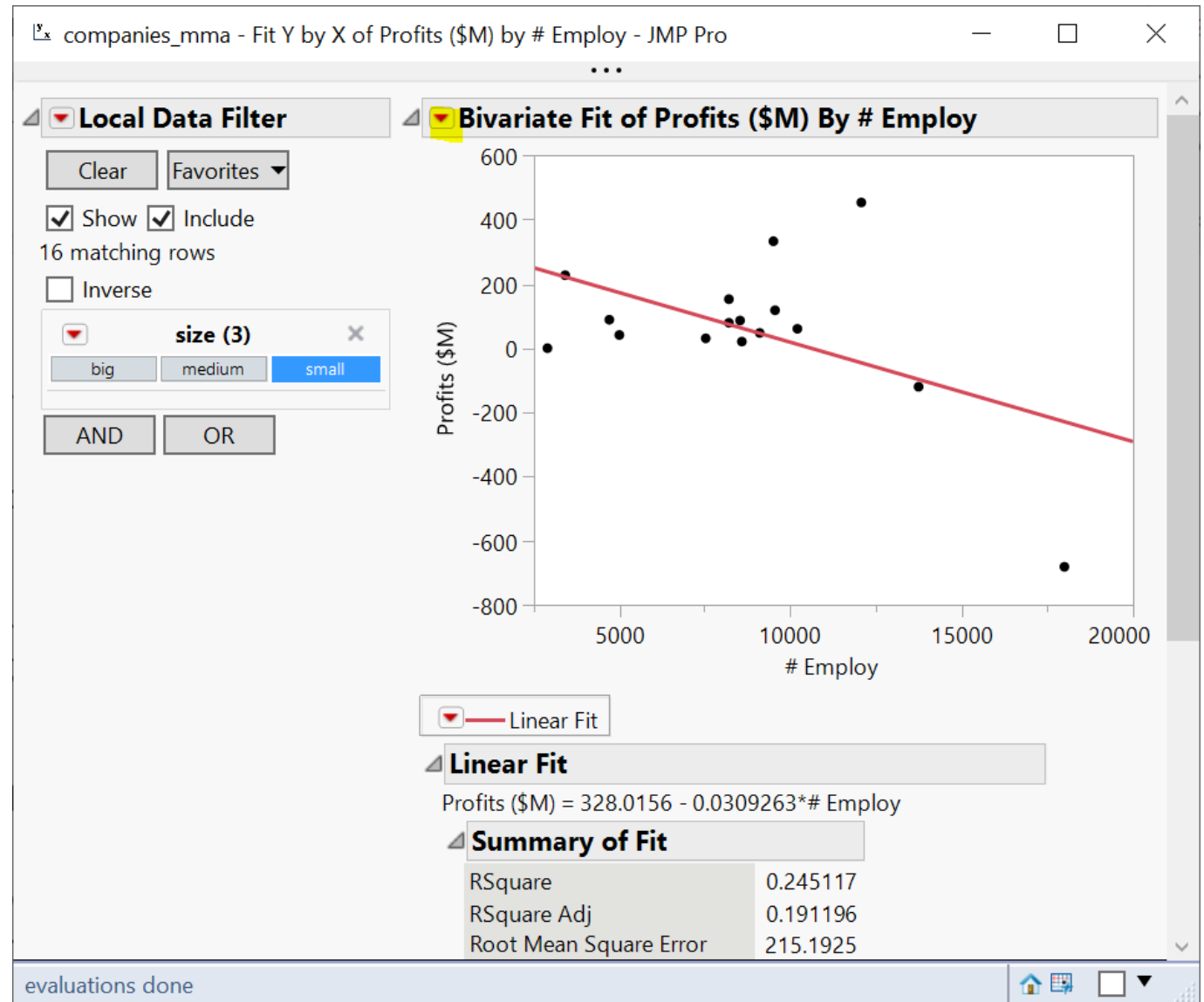
# Your Turn (Hands-on)

- 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
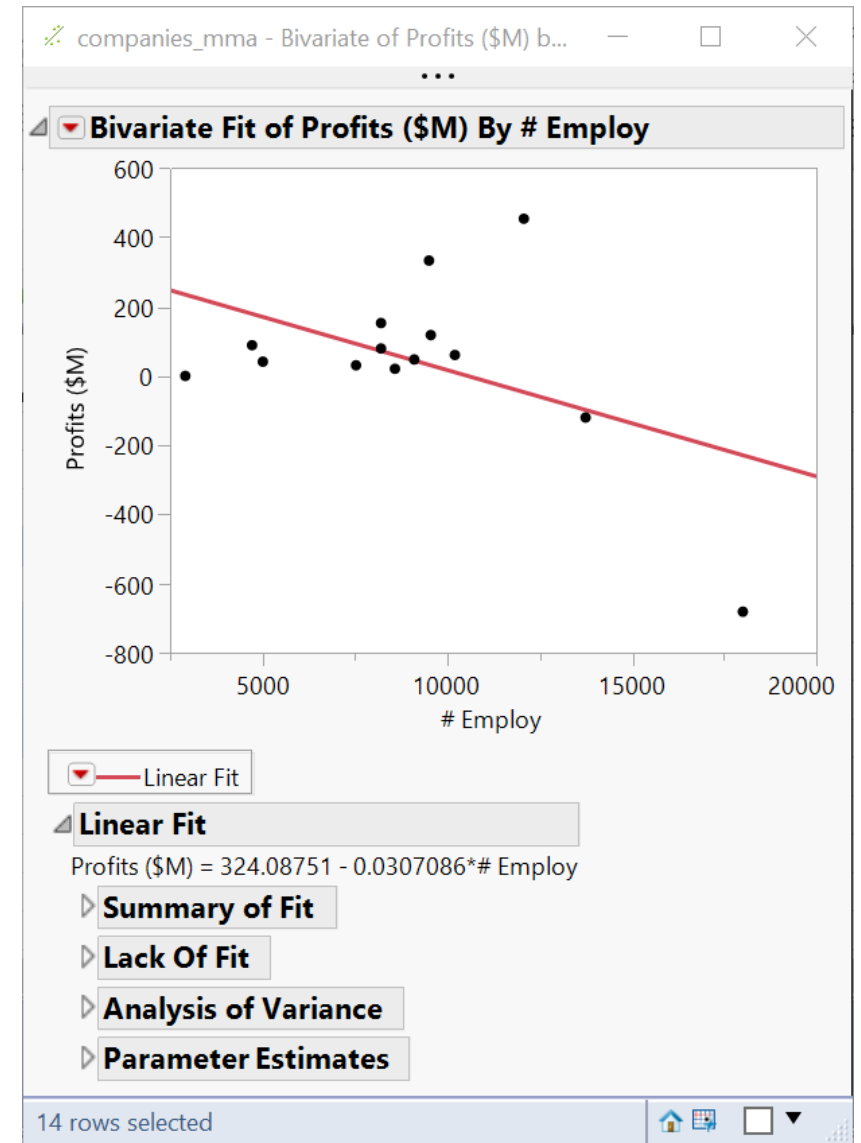
# Filter Rows (Demo)

- Rows > Row Selection > Select Where…

# 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)

# Sort Table / Order Rows (demo)

- Order rows by
  - a single column
  - multiple columns

# 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**



https://www.jmp.com/support/help/en/17.1/index.shtml#page/jmp/summarize-your-data.shtml

# Your Turn (Hands-on)

- Find total sales by Type and Size

# Aggregate / Summarize (by Group) (Demo)

- Analyze > Tabulate

- Tabulate (red triangle) > Make into data table