

JMP12: The art of data analysis

Marie Filteau
July 21st 2015

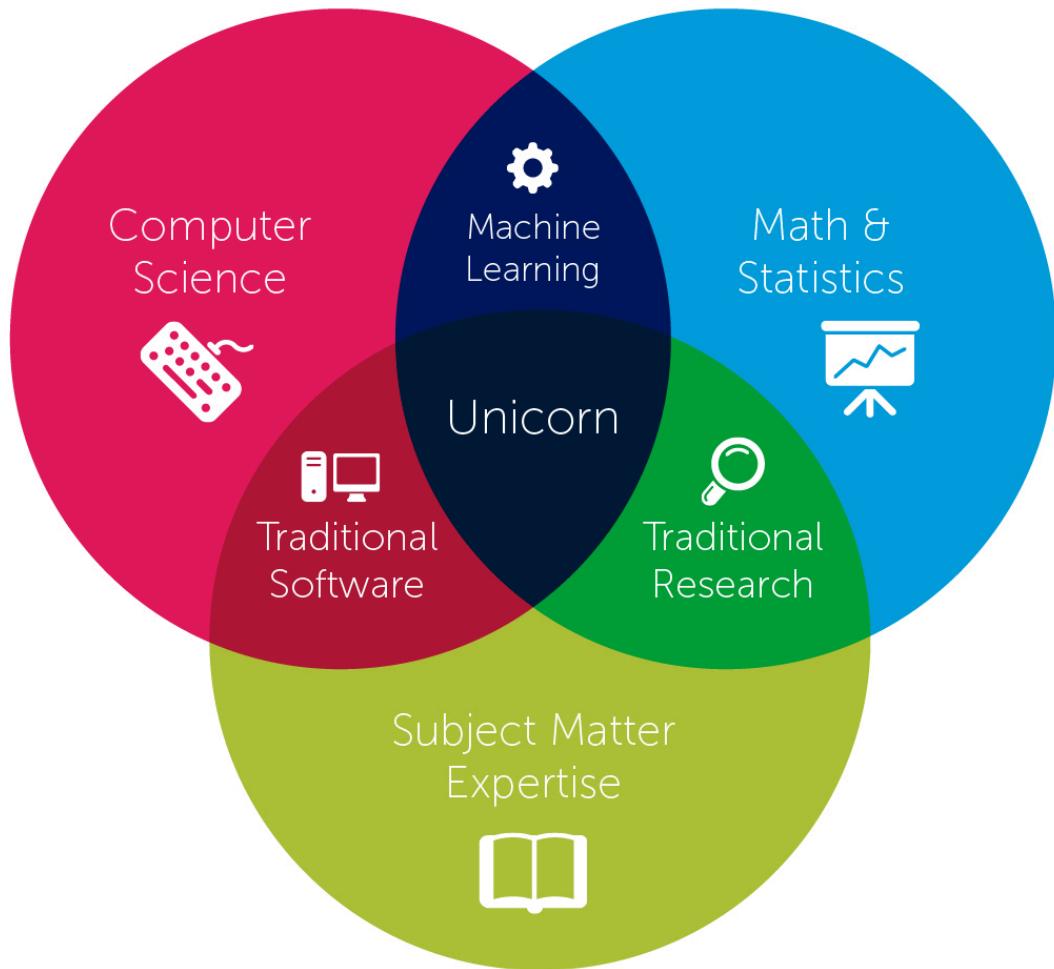
Disclaimer

Any views or opinions presented in this workshop are solely those of the author and do not necessarily represent those of the company.

The author declares that there are no conflicts of interest.



Data Science



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image,
provided that this copyright notice remains intact.

For those of us who
are not unicorns...

There's
jmp
Statistical Discovery.™ From SAS.

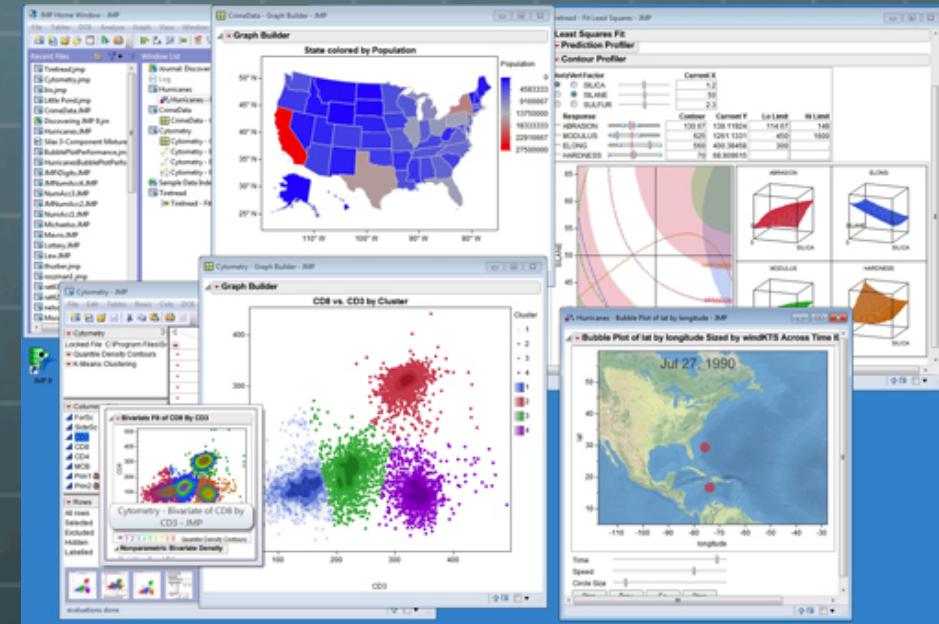
Why JMP?

Pros:

Renown company (SAS)
Graphical interface
Quick insight into your data
Full of options!

Cons:

Commercial (\$)
Not made for network analysis
Separate version for Genomic analysis



Today's layout

1. Demo of JMP functionalities
 1. Data import
 2. Organizing your data
 3. Graph builder
 4. Filter
 5. Tabulate
 6. Univariate statistical analysis (e.g. t-test, fisher's test)
 7. Multivariate analysis (e. g. PCA, clustering)
 8. Formulas / scripts
2. Tutorials
 1. Build in tutorials
 2. Analyze a real dataset (Part of...)

Data import

- .XLS , .TXT, .CSV (preview or best guess)

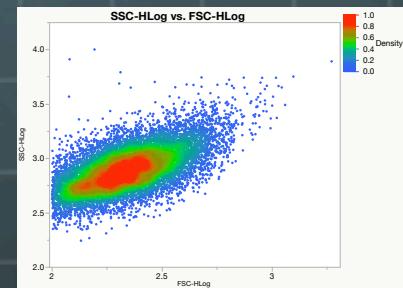
- You can always copy/paste..
 - > edit >Paste with column names

- .FCS (Cytometer)

- Check the data type



Categorical
Continuous
Ordinal



Organizing data

- What is what?
- Each data point should have its own row
- Each data point should be annotated (sample, condition, replicate, etc.)
- Make a design file beforehand
 - >Tables > join (cartesian)
 - Or with > DOE
- “antibiotics MICs.jmp”, “cancer1.jmp”, “cancer2.jmp” > Tables



Graph builder

- The best feature of JMP!
- > Graph > Graph Builder
- Explore your organized data with graphs
- “Big class.jmp”
- Open more than one graph builder at a time
- Export in Power Point, illustrator or directly as an image



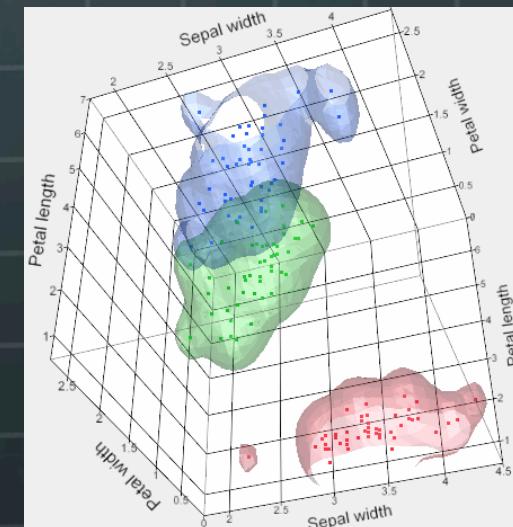
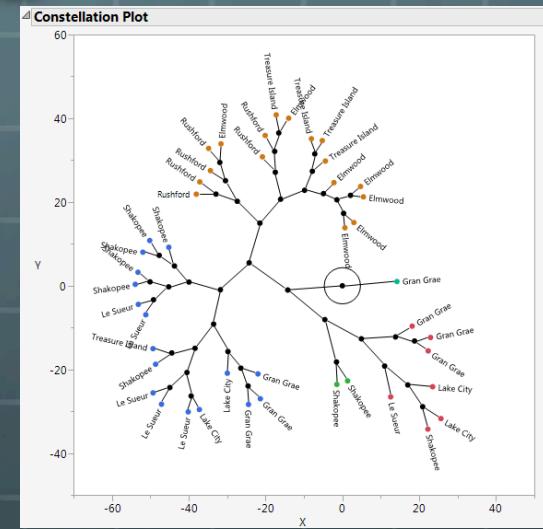
JMP Ipad App

Other Graphs

- Some graphs are hidden in analysis platforms
 - e.g. heatmaps

- 3D scatterplot (not publication quality)

- Bubble plots “**Air Traffic.jmp**”



Clean and Filter

- > columns > utilities > recode
- Row states: selected, excluded, hidden, color, markers
- > Row > filter
- Local data filter (**air traffic.jmp**)

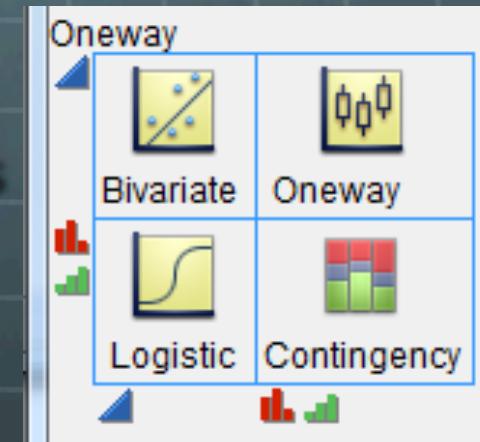


Tabulate

- Use the tabulate function to construct a summary table
- E.g. Find the mean of a group
 - “Big class.jmp”
 - > analysis >tabulate

Statistical tests

- You want a p-value?
- > Analyze > fit X by Y
- The type of data dictates the type of analysis
- Anova: “drug toxicity.jmp”
- Fisher’s test: “alcohol.jmp”
- Books that explain de statistics under Help > Books



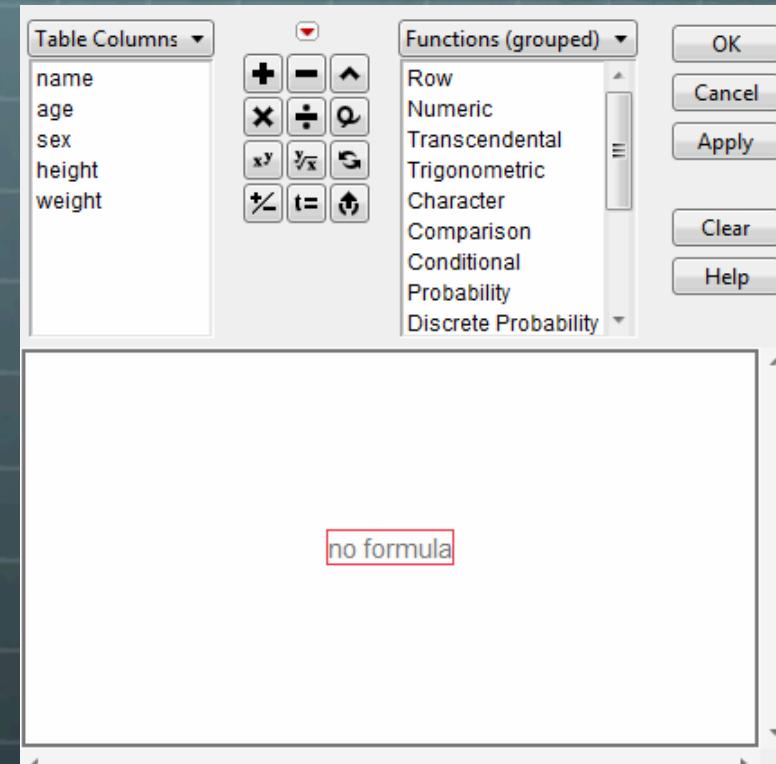
Know what you are doing if you are reporting statistics!

Multivariate analysis

- > analysis > multivariate method > Clustering
 - Hierarchical
 - Kmeans : “Cytometry.jmp”
- > analysis > multivariate method > principal component
 - PCA
- > analysis > multivariate method > multivariate
 - correlations

Formulas

- Create a new column
 - >column info > formula
- Directly in graph builder
- Normalization :
 - > Distribution > save

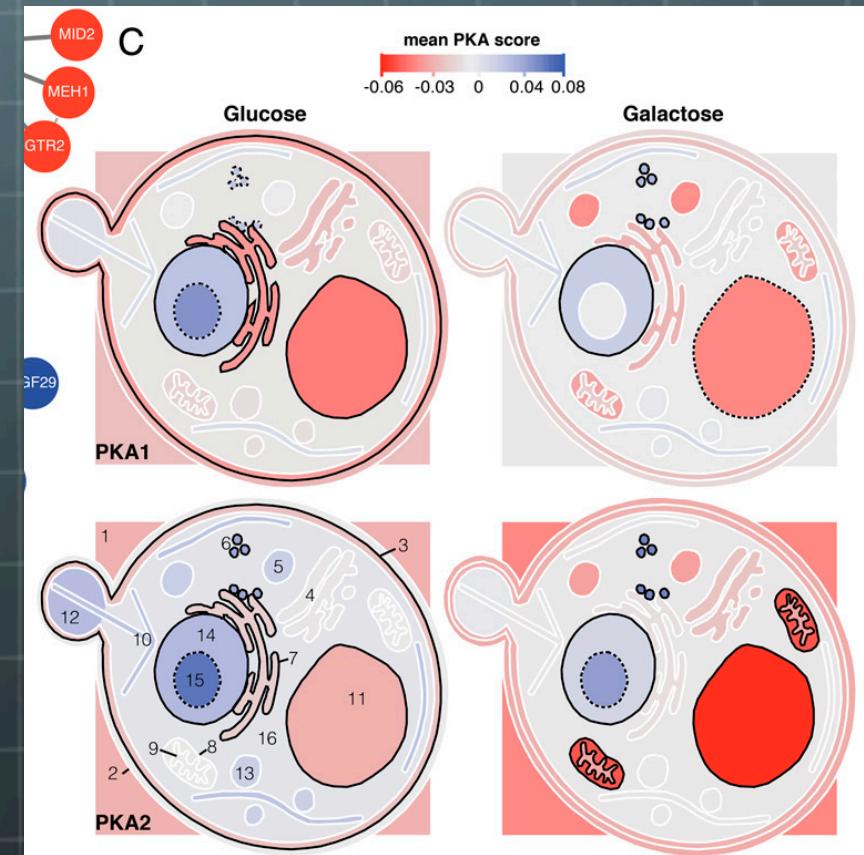
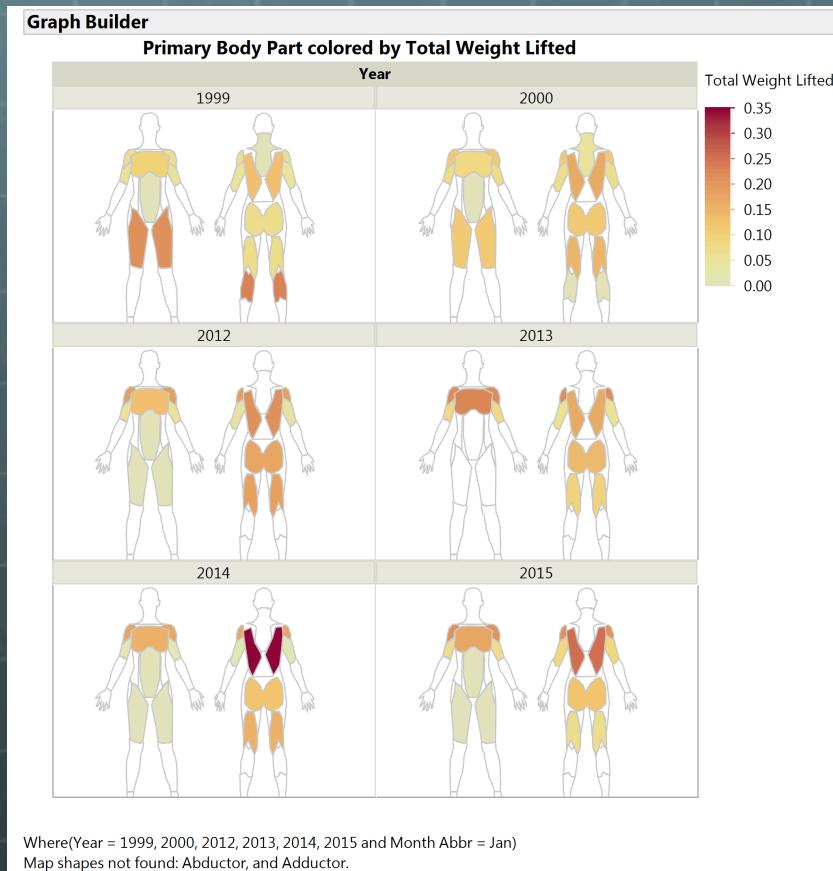


Scripts

- Save script to table
- Re-run the script to get your previous analysis / graph
- Watch out for row states!
- Lots of possibilities... go to >[Help](#) for more

Find more at

http://wwwjmp.com/en_us/software/jmp.html



Tutorials

Let get to work!

Build-in tutorials

Go to >Help > tutorials > tutorials directory

Follow the instructions

Quick help document:

https://www.dropbox.com/s/socxjr4h7p6s1be/2014_05%20-%20OPG%20Combined%20for%20Printing%20-%20Version%204.pdf?dl=0

Analyze this dataset



Download these files:

IMG_0908.txt

IMG_0909.txt

IMG_0910.txt

IMG_0911.txt

URL: https://github.com/enormandeau/club_bioinfo_ibis

IMG_0912.txt

IMG_0913.txt

IMG_0914.txt

IMG_0915.txt

screen design.jmp

Analyze this dataset

- 9 files:

- 8 .txt files

- “**IntDenBackSub**” is the size of colonies as shown on the picture

- 1 Design file that contains the annotation:

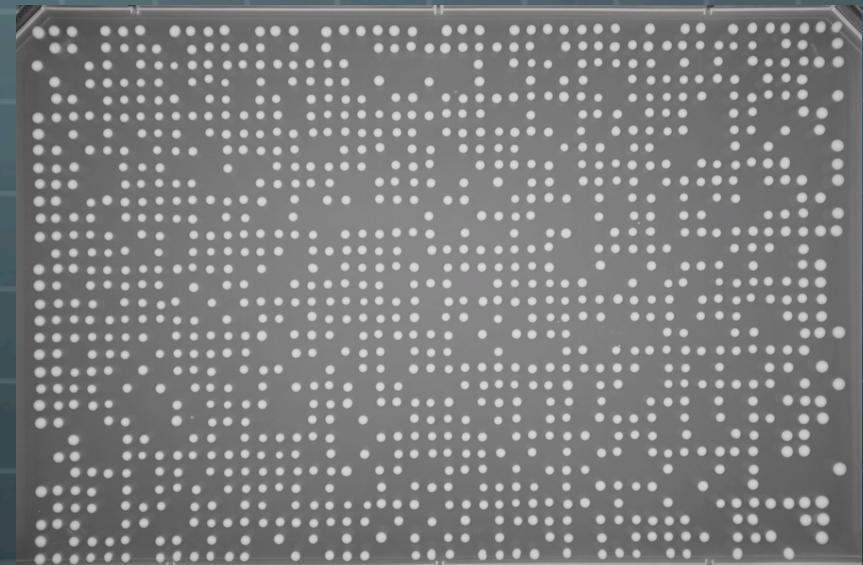
- 8 pictures of plates

- 1536 positions on each array

- 2 different arrays

- 4 replicate of each array

1536 colony array



Analyze this dataset

- ➊ Open the 8 **IMG_XXXX.txt** files in JMP
- ➋ Under **>Table > Concatenate**
 - ➌ Add the tables to concatenate
 - ➍ The order is not important
 - ➎ Do not add the first table twice!
 - ➏ Check the “**Create source column**” box
- ➐ You should get a new table with 12,288 lines

Analyze this dataset

- ➊ Open “screen design.jmp”
- ➋ Recode the “Row” columns with numbers
- ➌ Set the “Row” and “Col” columns as character in both tables: right click on the column title > column info
- ➍ Go to Tables > Update
 - ➎ Check “Match columns”
 - ➏ “Picture” with “Source Table”
 - ➏ “Row” with “Row”
 - ➏ “Col” with “Col”

Analyze this dataset

Play around with the data

Graph builder

-  right click on selected “Row” and “Col” > character > concatenate
-  Scroll down to the new variable, right click > add to data table, rename the new colum “position”

Multivariate methods > cluster

-  Change “data as usual” for “data is stacked”
-  Y column: “IntDenbacksub”
-  Object ID: “Source Table”
-  Attribute ID: “Position”
-  Launch analysis
-  >Hierarchical clustering > two way clustering
-  Notice anything?

Analyze this dataset

- Can you rank the Strains (“**ORF**”) by colony size? (“**Int DenBackSub**”)
- Which grows best?
- Pick top 30, copy the ORF and paste the list here in the “**analyse**” box:

<http://yeastmine.yeastgenome.org/yeastmine/begin.do>

Is there a GO term enrichment?

GRACIAS SPASSIRO
ARIGATO MURUN
SHUKURIA TASHAKKUR ATU
JIUSPAKAR TASHAKKUR ATU
TRAYEPUCH SHUKURIA
MEDEKHKEE YAQHANYELAY
GOZAIMASHITA TASHAKKUR ATU
EFCHARISTO MAKEE
FAKAJAU KOMAPSUMJIDA
LAM MEHRBANI SUKSAMA
PALDIES EKHMET
MERCI BOLZİN
SHUKRIA BİYAN
TINGKI