CMPT 155: Computer Applications for Life Sciences

Lecture 12: Importing, Sorting and Parsing Data; Matrix Operations

Ivan E. Perez

April 26, 2022

Presentation Outline

Homework & Administrative Schedule

- Homeworks:
  - #7 Due: Friday, April 29rd at 6pm
  - #8 Due: Friday, May 6th at 6pm
- Final Exam Review: Tuesday, May 3rd at 6pm
- Mock Final Exam: Wednesday, May 4th
- **Final Exams**:
  - Section 01 (8am) Final Exam: May 9th 11am - 1pm
  - Section 02 (9am) Final Exam: May 10th 11 am - 1pm

Importing Data

Data can be imported from:

- text files (e.g., .txt, .csv)
- database connections (e.g., MySQL, MSAccess)

Importing from Text Files

Importing Data From Text

- right (Ctrl) -click a text file and try opening it with Excel
- In the Data Tab go to (Get Data) followed by From Text
- The Text to Columns wizard should start up.

Parsing Data

Data can be parsed using the Text to Columns wizard.

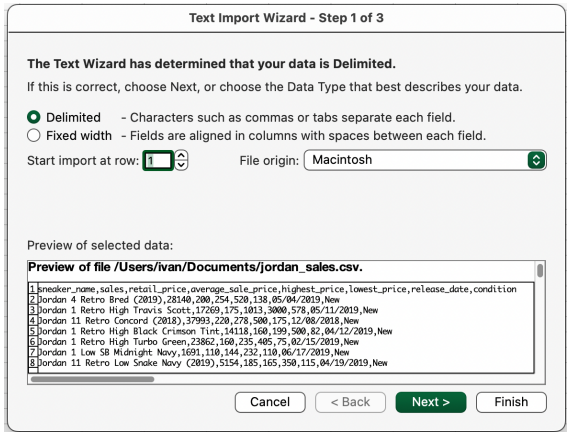

Figure: Step 1: Specify the way you wanted to delimit (i.e., separate/find breaks) in your data.

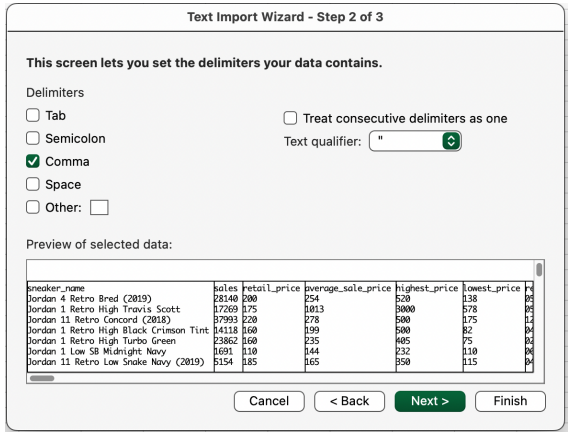

Figure: Step 2: Apply the delimiter that makes sense for your raw data. In this case 'Commas', ',' is our delimiter.

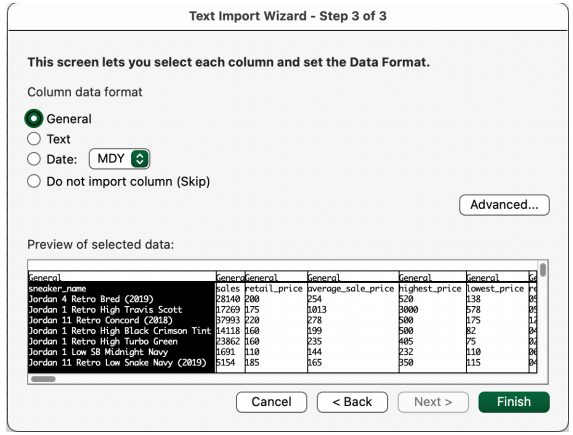

Figure: Step 3: Verify that data has been parsed correctly, and add final touches and/or Advanced options.

Concatenating Data

Cells can be concatenaed by using the '&' operator or using CONCAT().

	A	B	C	D	E	F	G	H
1	sneaker_name	sales	retail_price	average_sale_price	highest_price	lowest_price	release_date	condition
2	Jordan 4 Retro Bred (2019)	28140	200	254	520	138	5/4/19	New
3	Jordan 1 Retro High Travis Scott	17269	175	1013	3000	578	5/11/19	New
4								
5	Concatedated using							
6	CONCAT	=CONCAT(A2:H2)	Jordan 4 Retro Bred (2019)28140200254520138435895/4/19New					
7	&	=A2&B2&C2&D2 &E2&F2&G2&H2	Jordan 4 Retro Bred (2019)28140200254520138435895/4/19New					

Sorting Data

Data can be sorted in the Home → Sort & Filter Menu:

- manually by using 'Custom Sort' wizard.
- by A-Z by using 'Sort A to Z' or 'Sort Z to A'.
- automatically selecting Filter icon and using the Filter submenus.

Filtering Data

Data can be filtered and sorted using the Autofilter button in the Data Tab.

Example 1: NY data

Restated from p.45

1. Open: NYC Open Data Search 311 Service Requests 2010 to Present.
2. Click Export → CSV. A very long download should start.
   - If you want to get a feel for the data try using a snippet of this dataset called NYCOpenData311Sample.csv
3. Try importing this data using the Data import wizard.
4. Try answering the following questions:
   a. How Many 311 requests were filed under the Department of Transportation (DOT), and how many were filed under the NYPD?
   b. How many complaints did each Borough(Communitys) have?
   c. What Type of complaint was the most common?
   d. What were the Unique Keys, and Descriptors of the complaints **not** associated with Noise?

Example 1: Solution

Question No.	Answer
a	DOT = 2; NYPD = 12
b	Manhattan = 5; Bronx = 3; Brooklyn =4; Queens =2
c	Noise Complaint

d.

Unique Key	Descriptor
997177	Pothole
53995389	With License Plate
53994527	Blocked Hydrant
53999207	Plate Condition - Noisy
54000934	Blocked Hydrant

Exercise 1: Jordan Sales

1. Import the file 'JordanSales.csv'.
2. Use Autofiler to Create Filter Criteria.
3. Answer the following Questions about the data set.
   - What is the average sneaker sales price for release years 2014 through 2019?
   - How many options does a customer have if they want a sneaker from 2019 with an average retail price between \$175-\$250?
   - Based on your taste, what release year would you buy from and how much would you be willing to pay for Jordans?

Matrices

Matrices are arrays of numbers *m*-rows and *n*-columns. Similar to how we performed operations on cells with single values, certain operations and be applied to matrices. Matrices can be labled using capital letters.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 7 & 9 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 4 & 5 & 8 \\ 10 & 11 & 12 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix Operations

Addition: $+$

- Matrices must be the same size
- Size: $(m \times n) - (m \times n)$
- Output Size: $m \times n$
- Example: $\mathbf{A} + \mathbf{A} = 2\mathbf{A}$

Subtraction: $-$

- Matrices must be the same size
- Size: $(m \times n) - (m \times n)$
- Output Size: $m \times n$
- Example: $\mathbf{A} - \mathbf{A} = 0$

Matrix Operations: Continued

Multiplication: MMULT()

- Number of rows in first matrix **MUST** equal number of columns in the second matrix.
- Size: $(m \times n) \cdot (n \times r)$
- Output Size: $m \times r$
- Example: $\mathbf{A} \cdot \mathbf{B} = \mathbf{AB}$

Determinant: MDETERM()

- Square matrices only
- Size: $n \times n$
- Output Size: Single Value
- Example: $\det(\mathbf{C}) = -2$

Matrix Operations: Continued

Inverse: MINVERSE()

- Square nonsingular matrices only.
- Size: $n \times n$
- Output Size: $n \times n$
- Example: $\mathbf{C}^{-1}$

Transpose: TRANSPOSE()

- All Matrices
- Size: $m \times n$
- Output Size: $n \times m$
- Example: $\mathbf{A}^{\mathsf{T}}$

Solving Linear Equations with Matrices

Linear equations with three unknowns can take the form:

$$\begin{aligned} 10x + 12y + 15z &= 40 \\ 11x + 12y + 14z &= 80 \\ 3x + 4y + 4z &= 24 \end{aligned}$$

Expressing this equation using matrices we get,

$$\mathbf{AX} = \mathbf{b}$$

Where

$$\mathbf{A} = \begin{bmatrix} 10 & 12 & 15 \\ 11 & 12 & 14 \\ 3 & 4 & 4 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 40 \\ 80 \\ 24 \end{bmatrix}$$

Example 2: Solution

The solution to this equation is

$$\mathbf{AXA}^{-1} = \mathbf{bA}^{-1}$$
$$\cancel{\mathbf{A}}\mathbf{XA}\cancel{^{-1}} = \mathbf{bA}^{-1}$$
$$\mathbf{X} = \mathbf{bA}^{-1}$$

We can express this solution in Excel by:

1. Writing out arrays for **A** and **b**.
2. Using MINVERSE() on the selection for **A** to derive $\mathbf{A}^{-1}$.
3. Using MMULT() to multiply **b** by $\mathbf{A}^{-1}$.

Example 2: Solution

	A	B	C	D	E
1	Matrix A				Matrix b
2	10	12	15		40
3	11	12	14		80
4	3	4	4		24
5					
6	Matrix A INV				Matrix b(A^-1)
7	-0.5	0.75	-0.75		22
8	-0.125	-0.3125	1.5625		7.5
9	0.5	-0.25	-0.75		-18

Further Reading

The topics covered in the lecture can be found in *Compter Applications for Life Sciences* p.39-46 and p. 85-90