Egad! It’s Excel

Data Boot Camp

Lessons 1.2 and 1.3

Quick Refresher

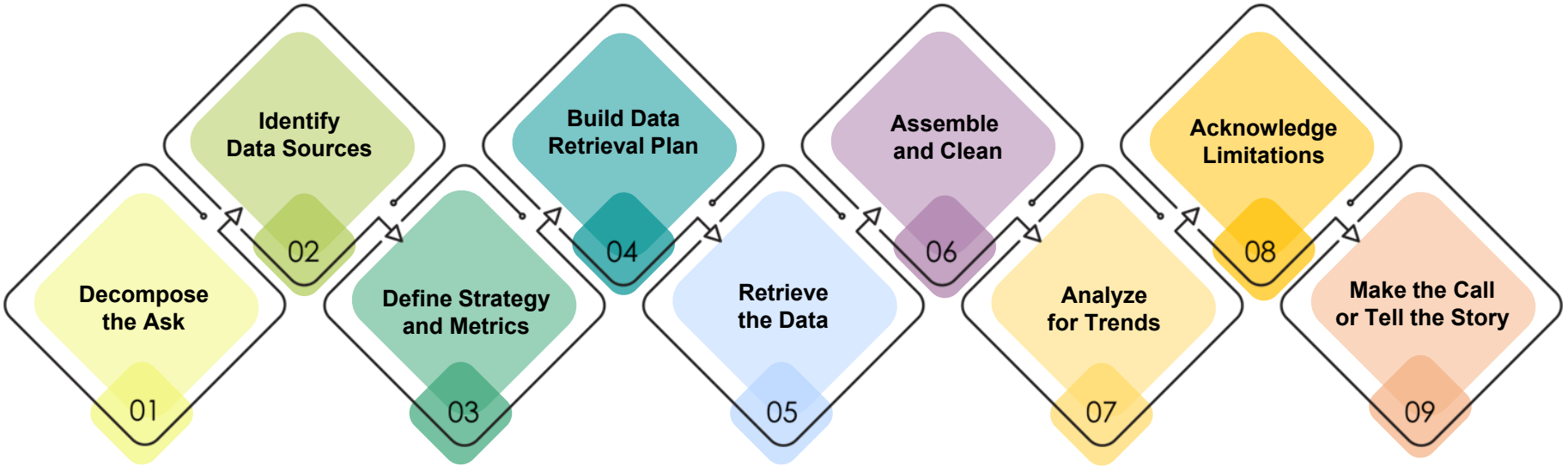
Data analytics is about what **two** things?

Fundamentally, data analytics is about **storytelling** and **truth-telling.**

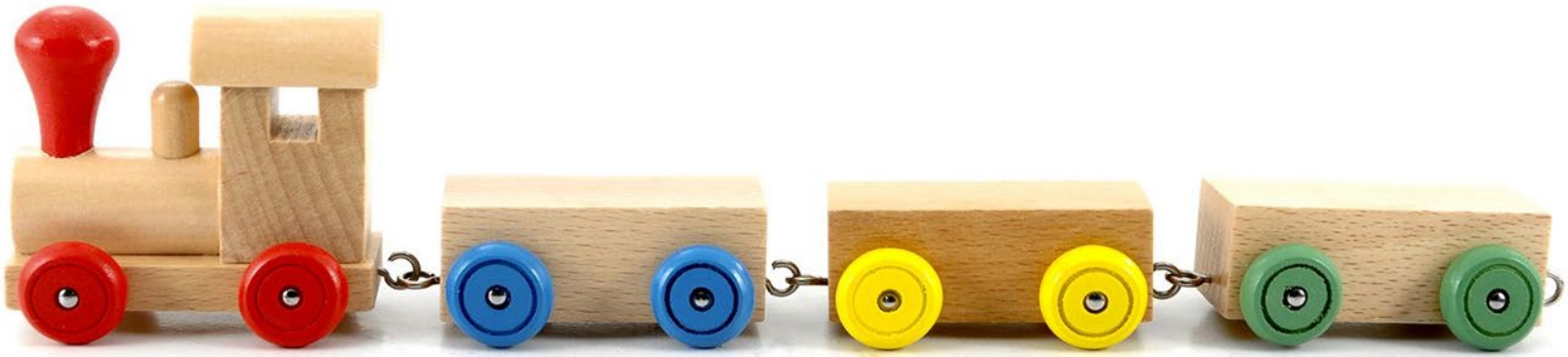
What are the steps in the **Analytics Paradigm**?

Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



Let’s Start with the Basics



Formulas

Ooh…Coding! (Sort Of)

Excel has introduced you to a sort of proto-programming. When you write scripts, you will rely on functions (methods) that do something to or with arguments.

Ooh…Coding! (Sort Of)

When we reference a range or a set of ranges, Excel is given a set of variable inputs. Excel will determine the actual values of these inputs prior to executing the function.

Ooh…Coding! (Sort Of)

What about this example?

Which is the function?

Which are the arguments?

```
= SUM( AVG(F4:F6), AVG(G4:G6) )
```

Ooh…Coding! (Sort Of)

What about this example?
Which is the function?
Which are the arguments?

It depends.

```
= SUM( AVG(F4:F6), AVG(G4:G6) )
```

Ooh…Coding! (Sort Of)

What about this example?
Which is the function?
Which are the arguments?

The **AVG functions** take the provided ranges as their arguments.

```
= SUM( AVG(F4:F6) , AVG(G4:G6) )
```

Ooh…Coding! (Sort Of)

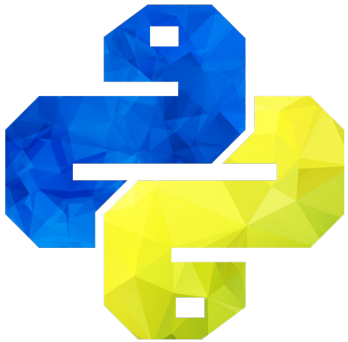
What about this example?
Which is the function?
Which are the arguments?

This is a **nested function**. We'll be doing plenty of complex nests in this class.

```
= SUM( AVG(F4:F6), AVG(G4:G6) )
```

You Can Code Too!

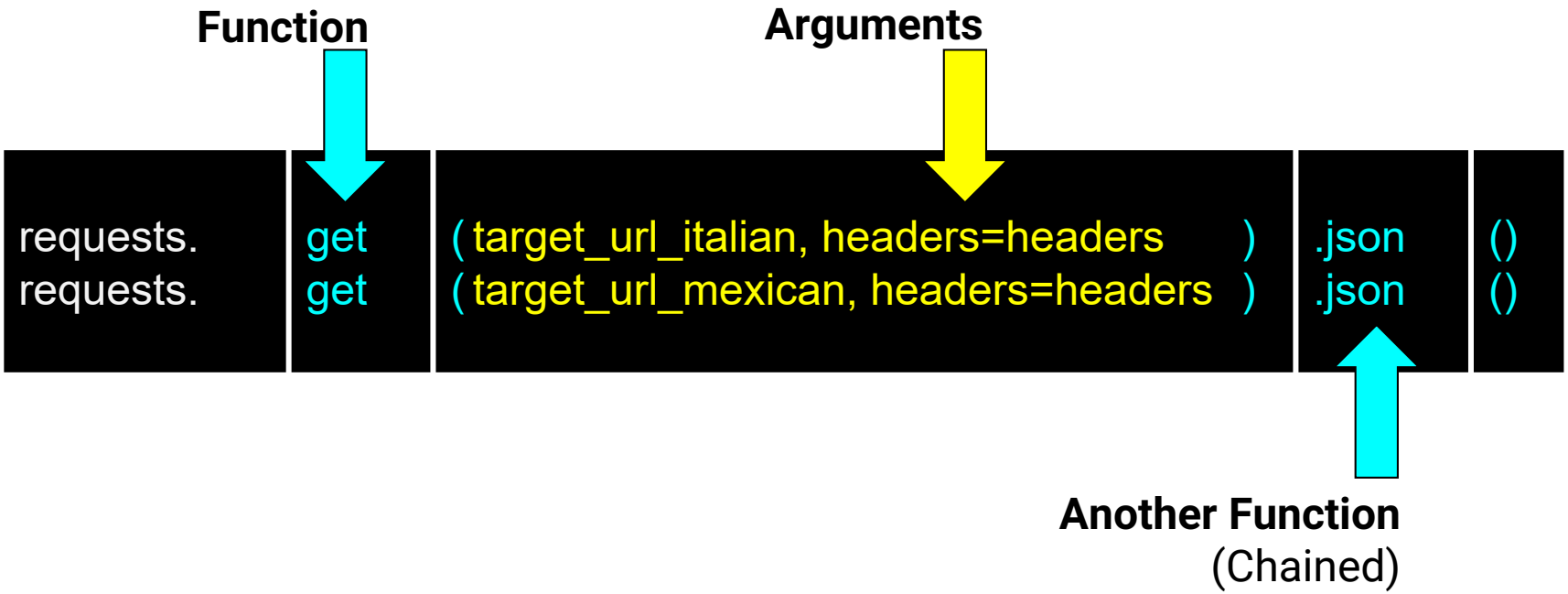
Here’s a Python snippet from the last class.



```
requests.get(target_url_italian,          headers=headers).json()
requests.get(target_url_mexican,          headers=headers).json()
```

You Can Code Too!

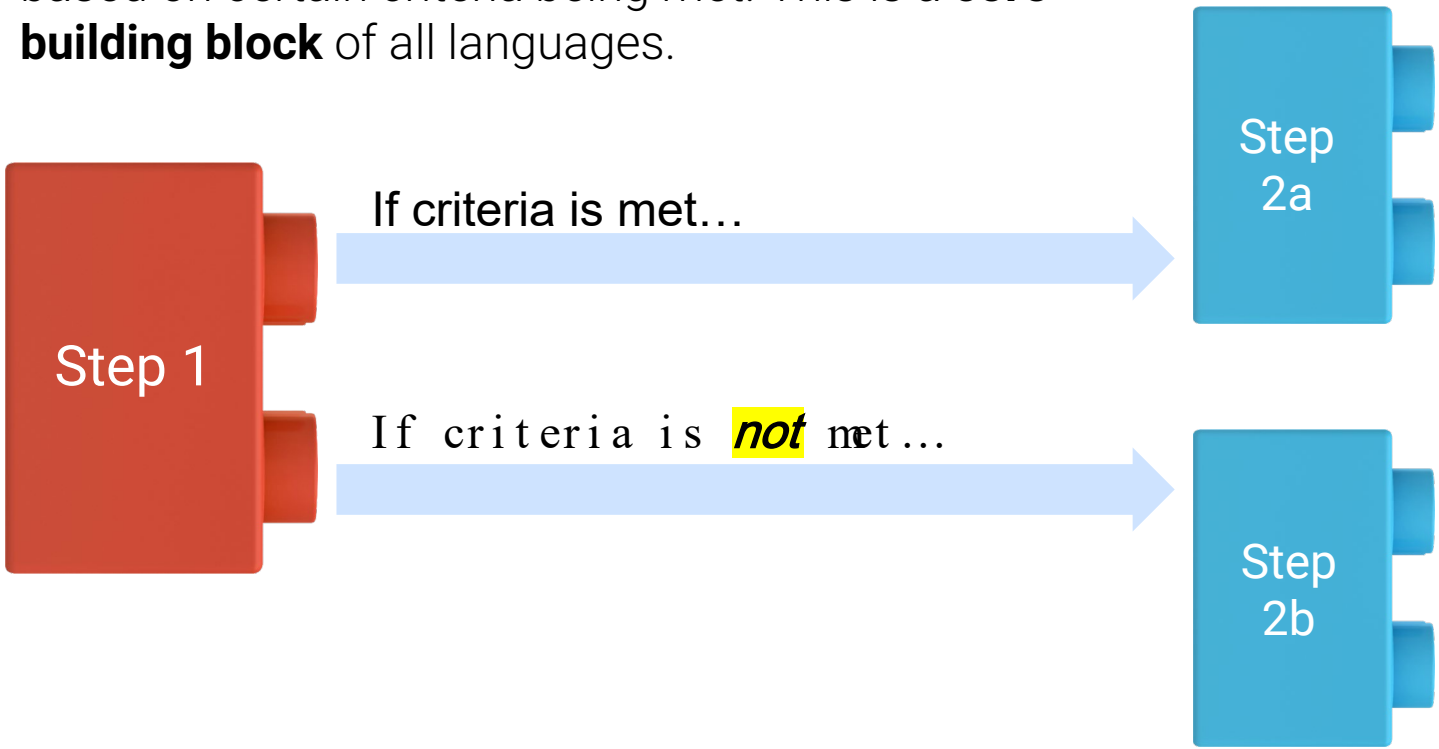
Syntax and capabilities may differ across technologies and platforms, but fundamental concepts remain the same.



Conditionals

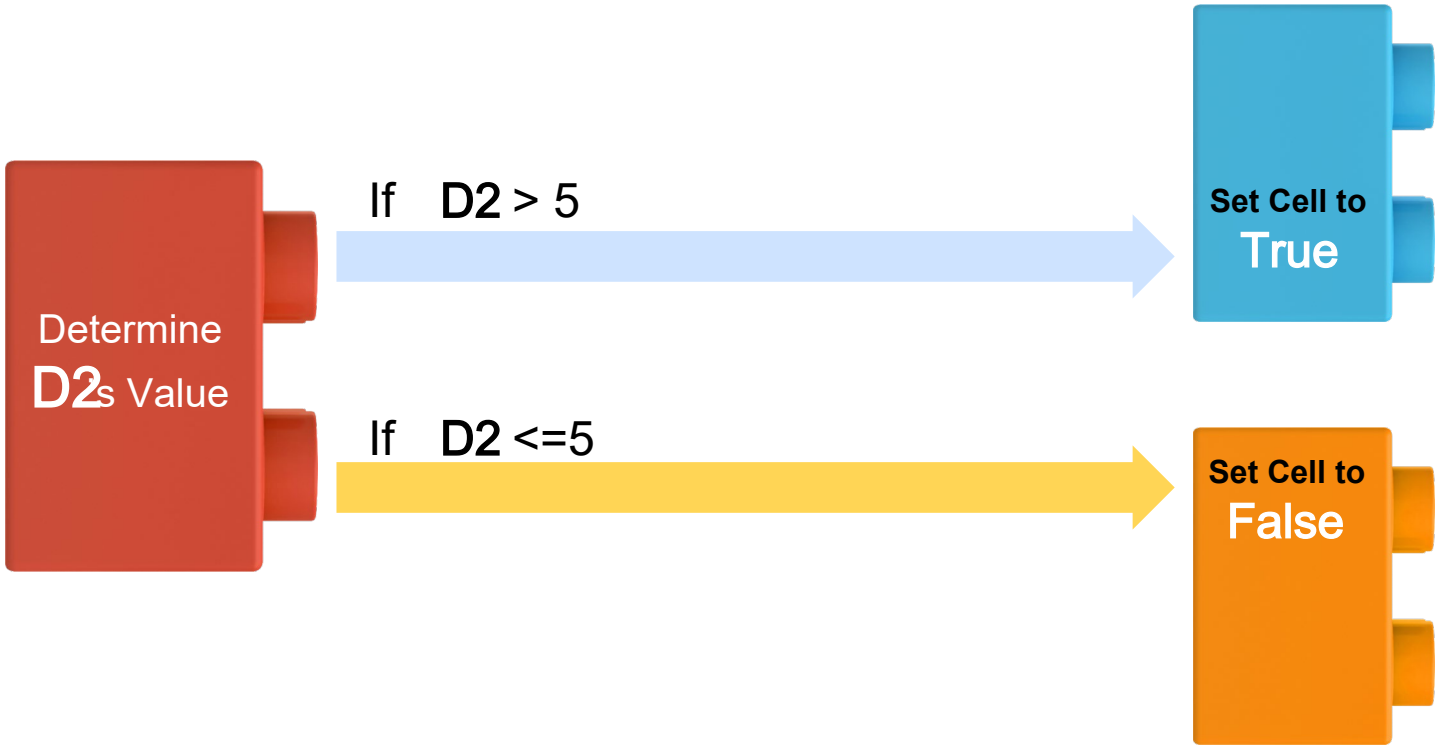
Conditionals: If This, Then That

Conditionals present a way to control the flow of logic based on certain criteria being met. This is a **core building block** of all languages.



Conditionals: If This, Then That

=IF(D2>5, TRUE, FALSE)



But what if we want to **combine** conditions?

AND, NOT, OR

Ooh…Coding! (Sort Of)

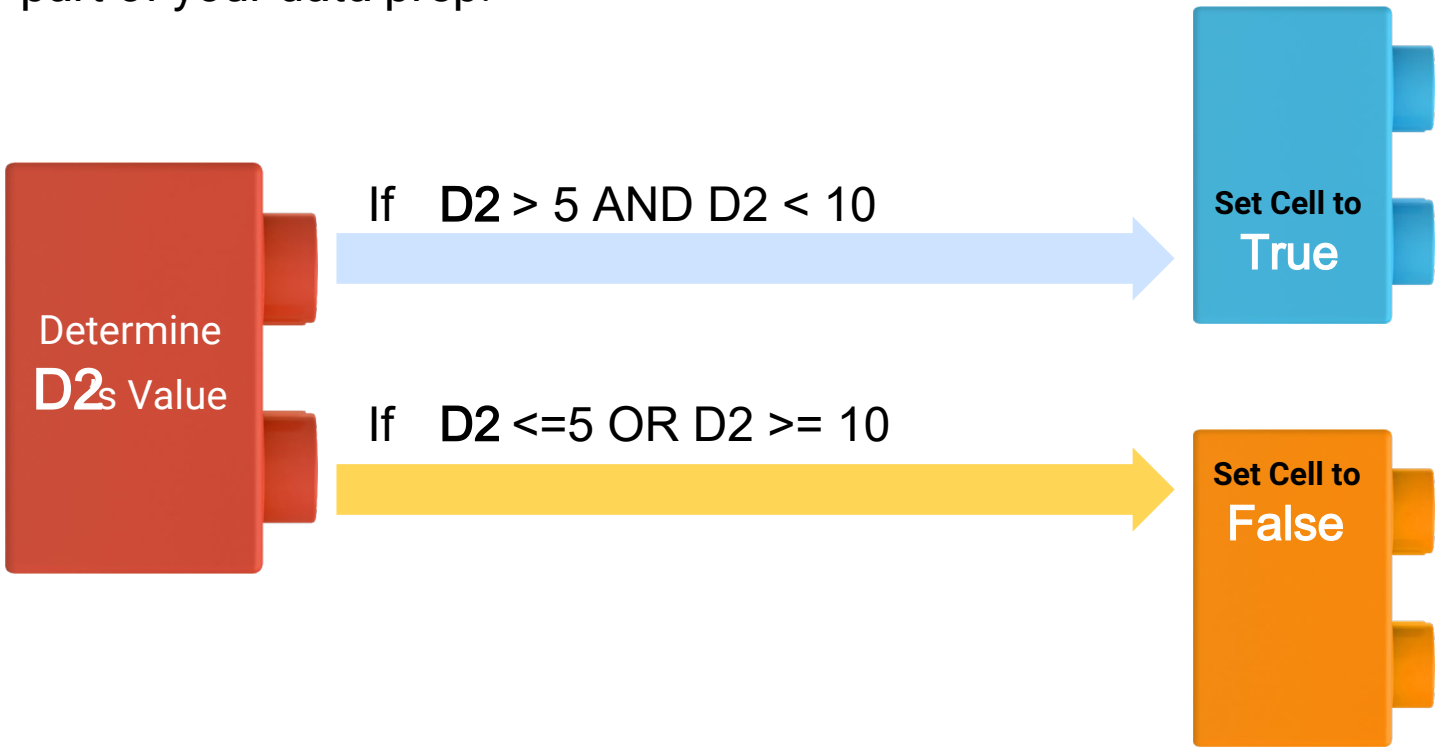
But what if we want
to **combine** conditions?

AND, NOT, OR

```
=IF( AND(D2>5, D2<10),TRUE,FALSE)
```

Conditionals: If This, Then That

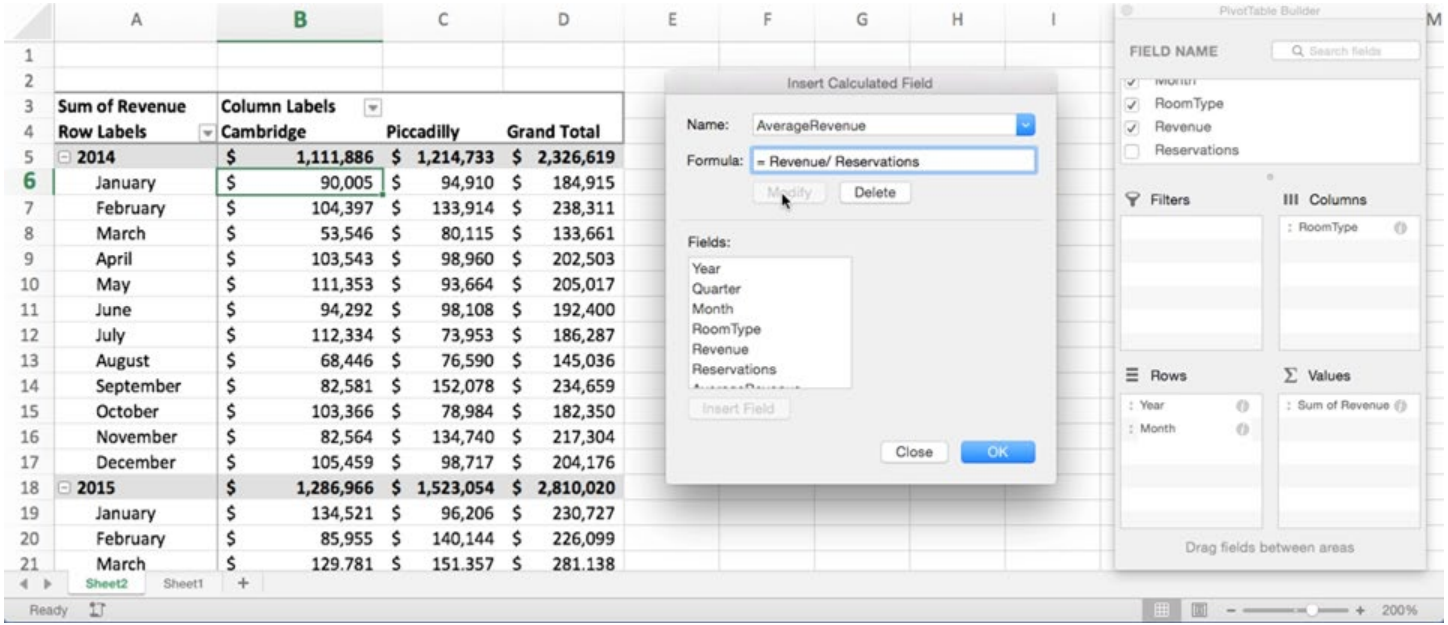
Nesting conditionals can quickly become a very convoluted (but necessary) part of your data prep.



Pivot Tables

Get Pivot With It

Pivot tables are one of the most important data visualization concepts to master in this class. (Don’t worry. They are a cinch to deal with.)



Get Pivot With It

In essence, a pivot table is a **summative** analytic tool that allows us to perform aggregate functions that allow any combination of fields. (The term *pivot table* comes from the fact that we are pivoting along a data axis).

Seller	Qty. Sold	Date
Joseph	$42.50	1/1/17
Jacob	$65.00	1/3/17
Jacob	$5.25	1/6/17
Joseph	$125.00	1/6/17
Jacob	$3.50	1/7/17
Matt	$32.00	1/9/17

Seller	Total Sold
Joseph	$167.50
Jacob	$73.75
Matt	$32.00

Word to the Wise: Keep It Flat!

Modern Business Intelligence (BI) tools like Tableau, Sisense, and Salesforce work best if data is stored in flat CSVs—meaning column headers represent fields (vertically) on the spreadsheet. This is largely because all of these technologies heavily utilize pivot tables as a tool for their visualizations. **Don't try to confuse this simplicity. "Spreadsheet magic" is a nightmare to analyze.**

B	C	D	E	F	G	H
DateTime	Week #	Section?	Pace	Academic Support	Self-Mastery	Instructor Er
2016-09-11T04:00:00.000Z	18	RCB0503FSF - CCC	3	5	5	4
2016-09-11T05:00:00.000Z	6	UT0726FSF	3	5	3	4
2016-09-12T04:00:00.000Z	11	UCF062016FSF	4	4	3	5
2016-09-12T04:00:00.000Z	23	UCF0329FSF	2	4	5	1
2016-09-12T04:00:00.000Z	9	UNC0712FSF	3	4	4	3
2016-09-12T04:00:00.000Z	23	UCF0328FSF	4	3	2	3
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	5	4	4	5
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	5	5	4	5
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	2	4	4	4
2016-09-12T04:00:00.000Z	11	UCF062016FSF	4	5	4	5
2016-09-12T04:00:00.000Z	13	UCF061416FSF	4	5	1	5

Lookups

Look It Up with Lookups

Assume this table is gigantic. How would we **retrieve** the population of a specific planet for use in another formula?

Planet	Population
Zeelo	5020
Merinoa	380
Cardboard Box	2
…	…
Asteroid 9	95

Look It Up with Lookups

Assume this table is gigantic. How would we **retrieve** the population of a specific planet for use in another formula?

=vlookup(<value>, <full table>,
<column to retrieve>,<match parameter>)

Planet	Population
Zeelo	5020
Merinoa	380
Cardboard Box	2
…	…
Asteroid 9	95

Look It Up with Lookups

What will this yield?
=vlookup(“Asteroid 9”, Planets, 3, FALSE)

Planet	Population	Species
Zeelo	5020	Zoltans
Merinoa	380	Murphies
Cardboard Box	2	Hambones
…	…	
Asteroid 9	95	Asterisks

Look It Up with Lookups

What will this yield?
=vlookup(“Asteroid 9”, Planets, 3, FALSE)

Planet	Population	Species
Zeelo	5020	Zoltans
Merinoa	380	Murphies
Cardboard Box	2	Hambones
…	…	
Asteroid 9	95	Asterisks

Asterisks

Questions?