Day 2

Egad! It’s Excel

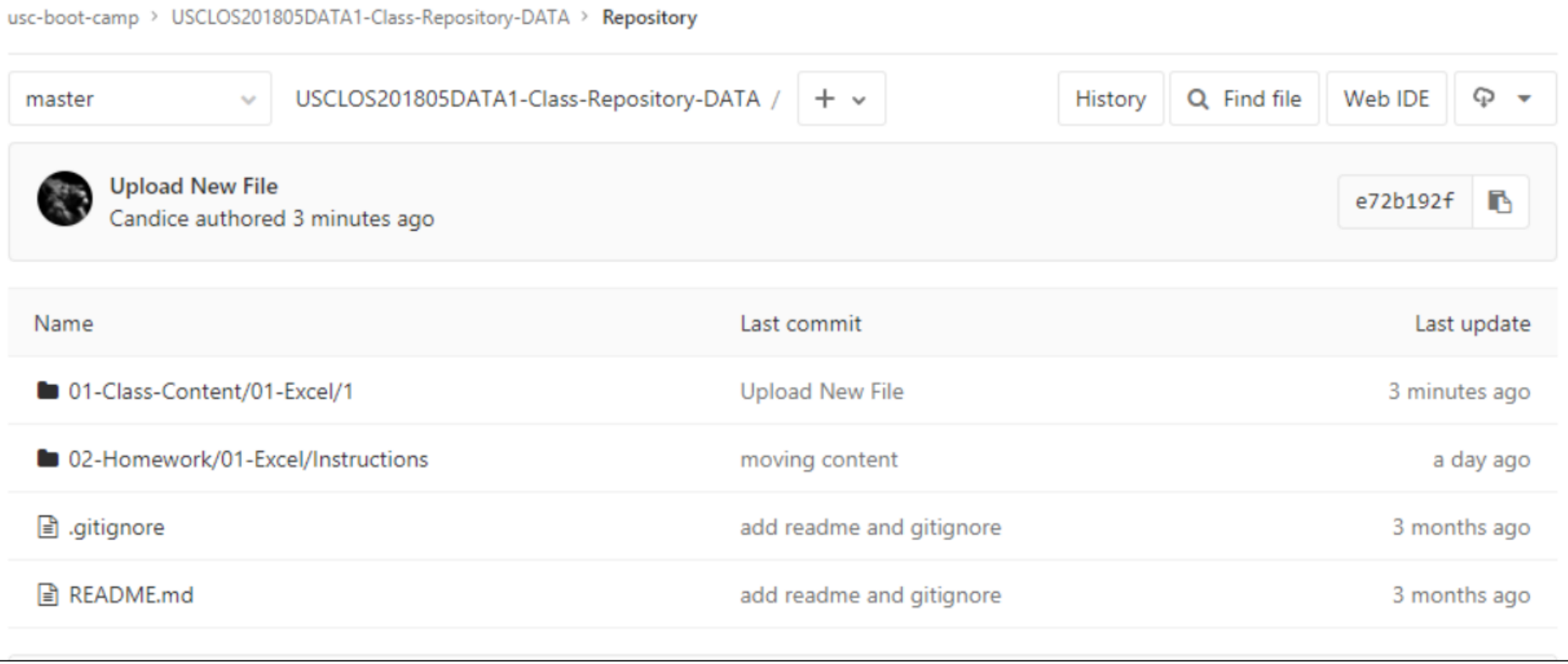
The Data Bootcamp

Objectives

- Basic Excel navigation and functionality
- Gain familiarity with the value of Pivot Tables and the steps for their utilization.
- Gain comfort utilizing VLookups and Hlookups
- Understand how to implement conditional formatting based on logical rules

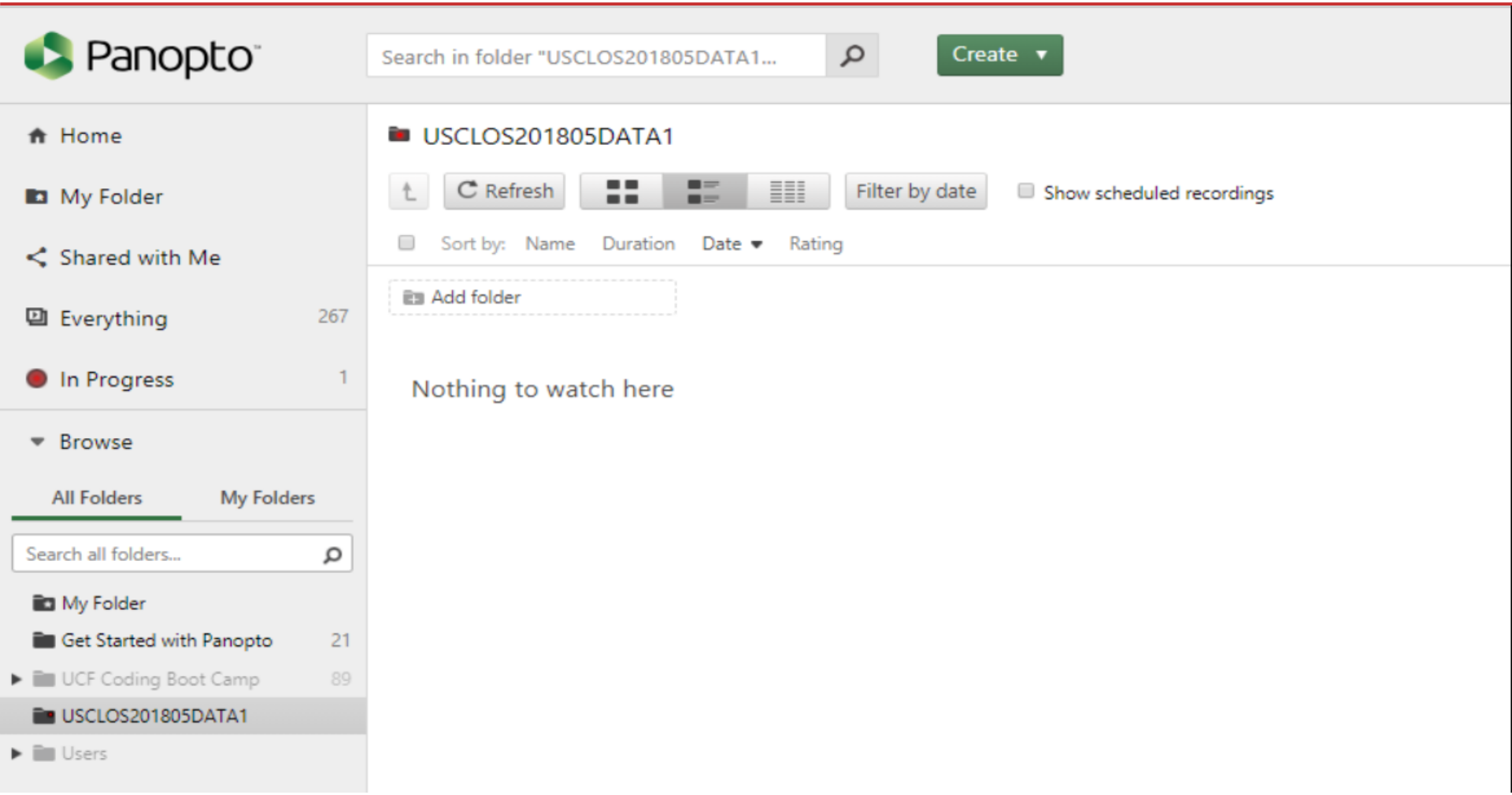
Admin Stuff

Class Git Repository



All Class Content and Homework will be here:
https://usc.bootcampcontent.com/usc-boot-camp/USCLOS201805DATA1-Class-Repository-DATA.git

Class Videos (UPDATE)



Class Videos will be automatically uploaded here:
https://codingbootcamp.hosted.panopto.com/Panopto/Pages/Sessions/List.aspx#folderID=%22ee8745c8-c7ff-4d38-b766-a8d3001005ad%22&folderSets=3

Free Python Material:

- Social Learning
  - https://www.sololearn.com/Course/Python
- Daily Newsletter:
  - https://realpython.com/python-tricks/
- Learn by Doing Project:
  - https://knightlab.northwestern.edu/2014/06/05/five-mini-programming-projects-for-the-python-beginner/
- Practical Business Python:
  - http://pbpython.com/
- Learn by watching video:
  - https://www.fullstackpython.com/best-python-videos.html

Deep Learning

- Daily Digest:
  - https://medium.com/tag/deep-learning
  - https://www.kdnuggets.com/author/gregory-piatetsky
  - https://www.analyticsvidhya.com/
- Certifications:
  - https://www.coursera.org/specializations/deep-learning
  - http://www.openculture.com/2017/05/artificial-intelligence-a-free-online-course-from-mit.html
  - https://www.coursera.org/learn/deep-learning-business

Homework Assignment #1

You will be analyzing thousands of Kickstarter projects to look for funding trends across goal targets and topics.

Homework Assignment #1

Due: Next Saturday

Recommended Target: Thursday of Next Week

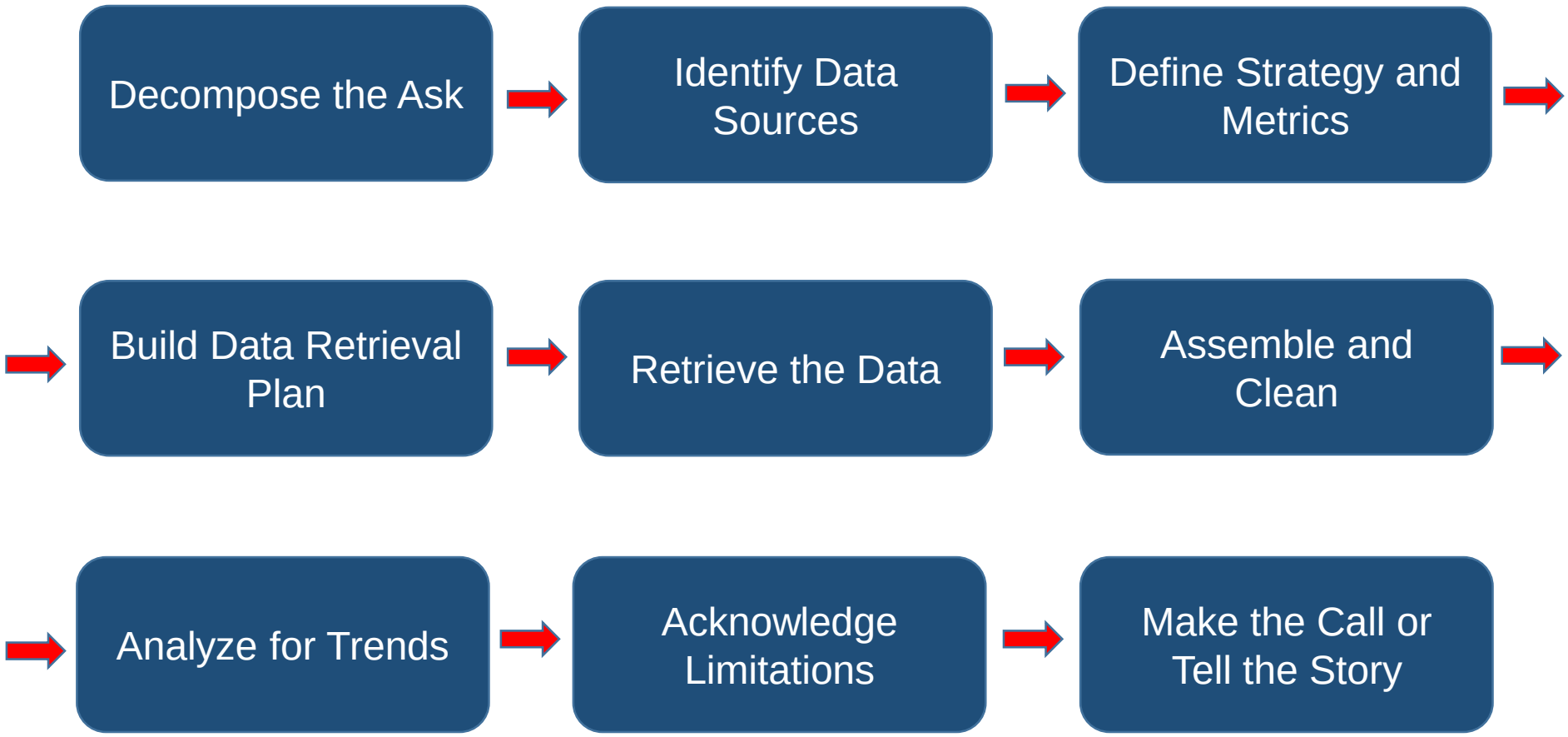
Quick Refresher

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

Truth-Telling
&
Story Telling

What are the steps in the
Analytics Paradigm?

Analytics Paradigm

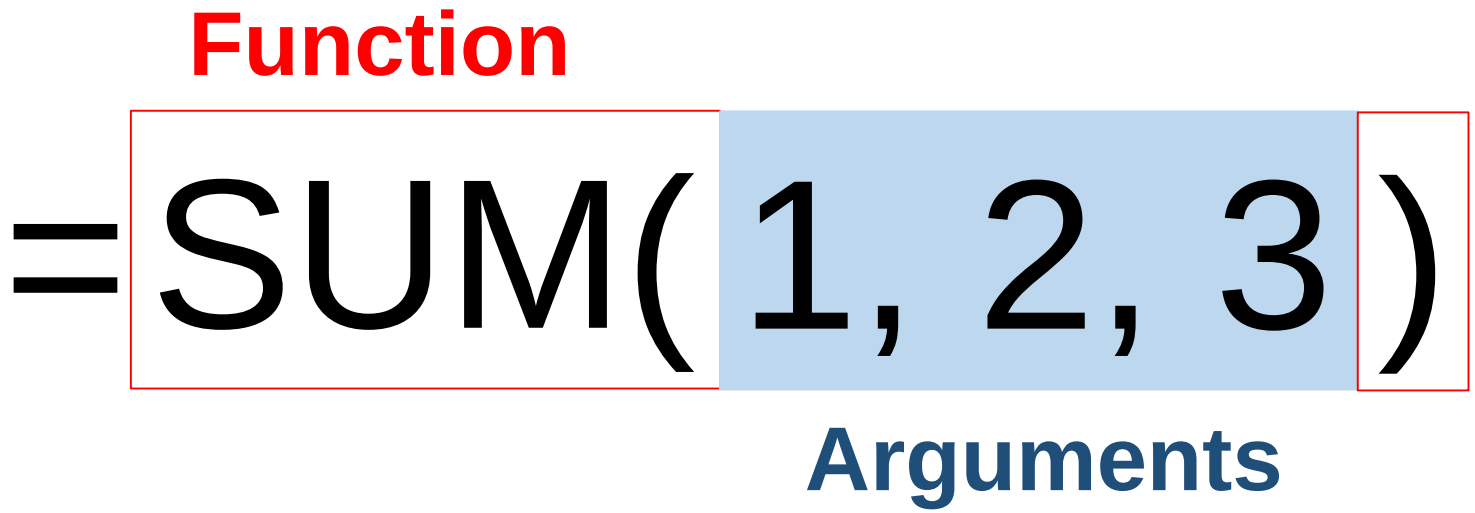


Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.

Let's Start with the Basics

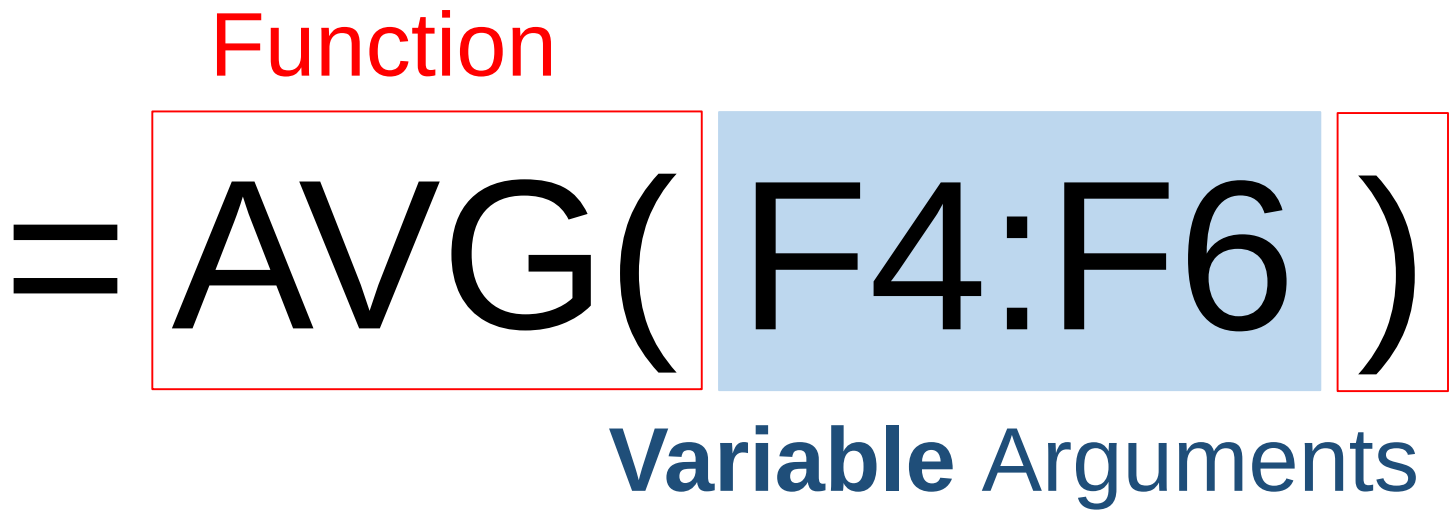
Formulas

Ooh… Coding! (Sort Of)



In a way, Excel has introduced you to a sort of proto-programming. Throughout your time writing scripts you will rely on **functions** (methods) that do *something* to or with **arguments**.

Ooh… Coding! (Sort Of)



When we reference a set of range, Excel is being given a set of **variable** inputs. It 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?

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

It Depends…

Ooh… Coding! (Sort Of)

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

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

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

Ooh… Coding! (Sort Of)

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

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

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

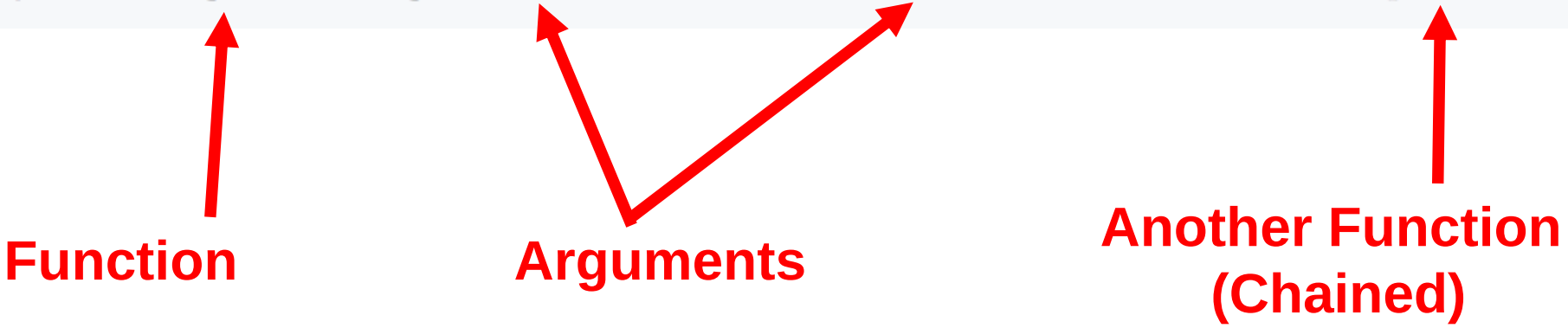
You Can Code Too!

Python Snippet from Last Class

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

Python Snippet from Last Class

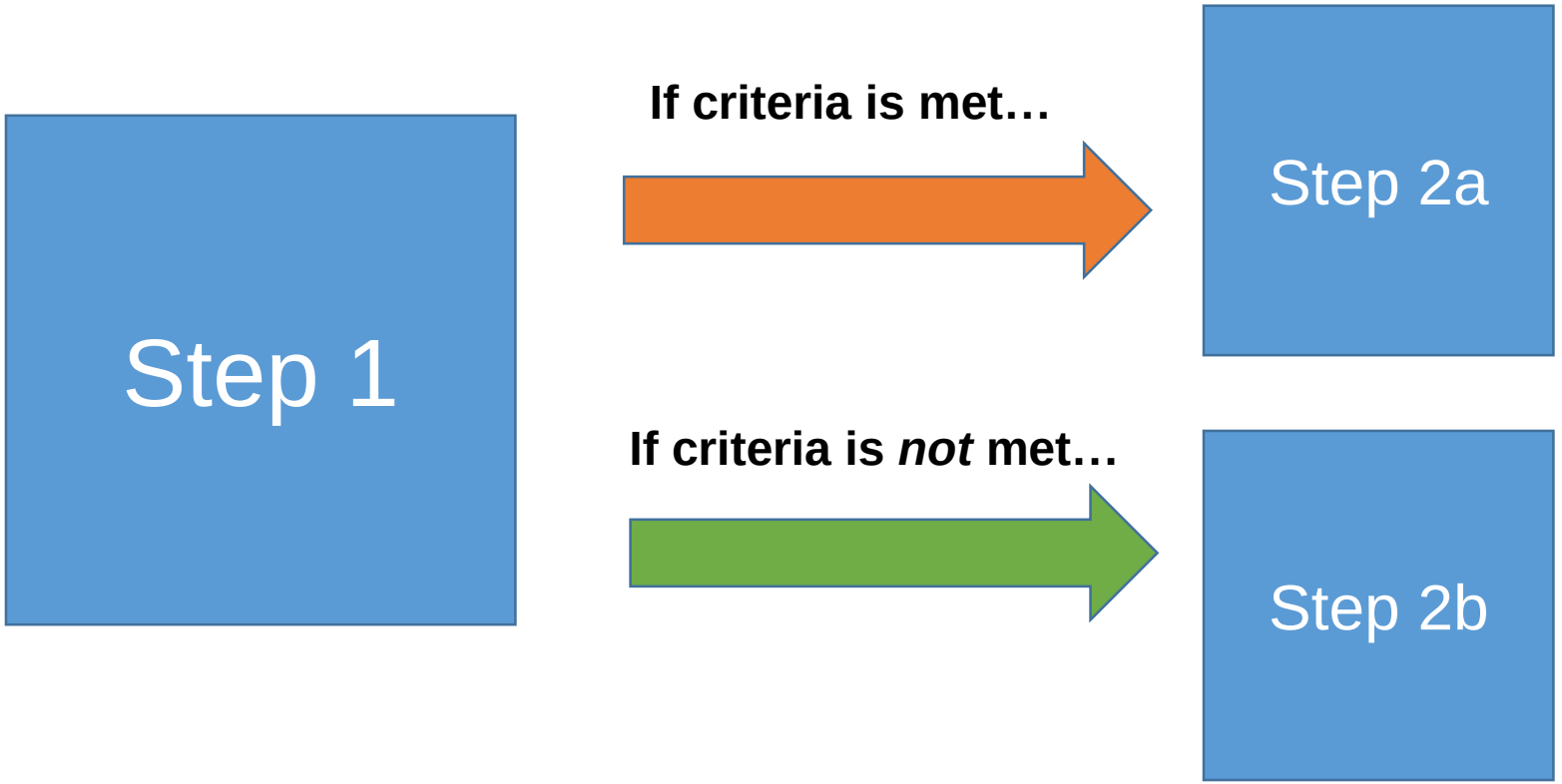
```
requests.get(target_url_italian, headers=headers).json()
requests.get(target_url_mexican, headers=headers).json()
```



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

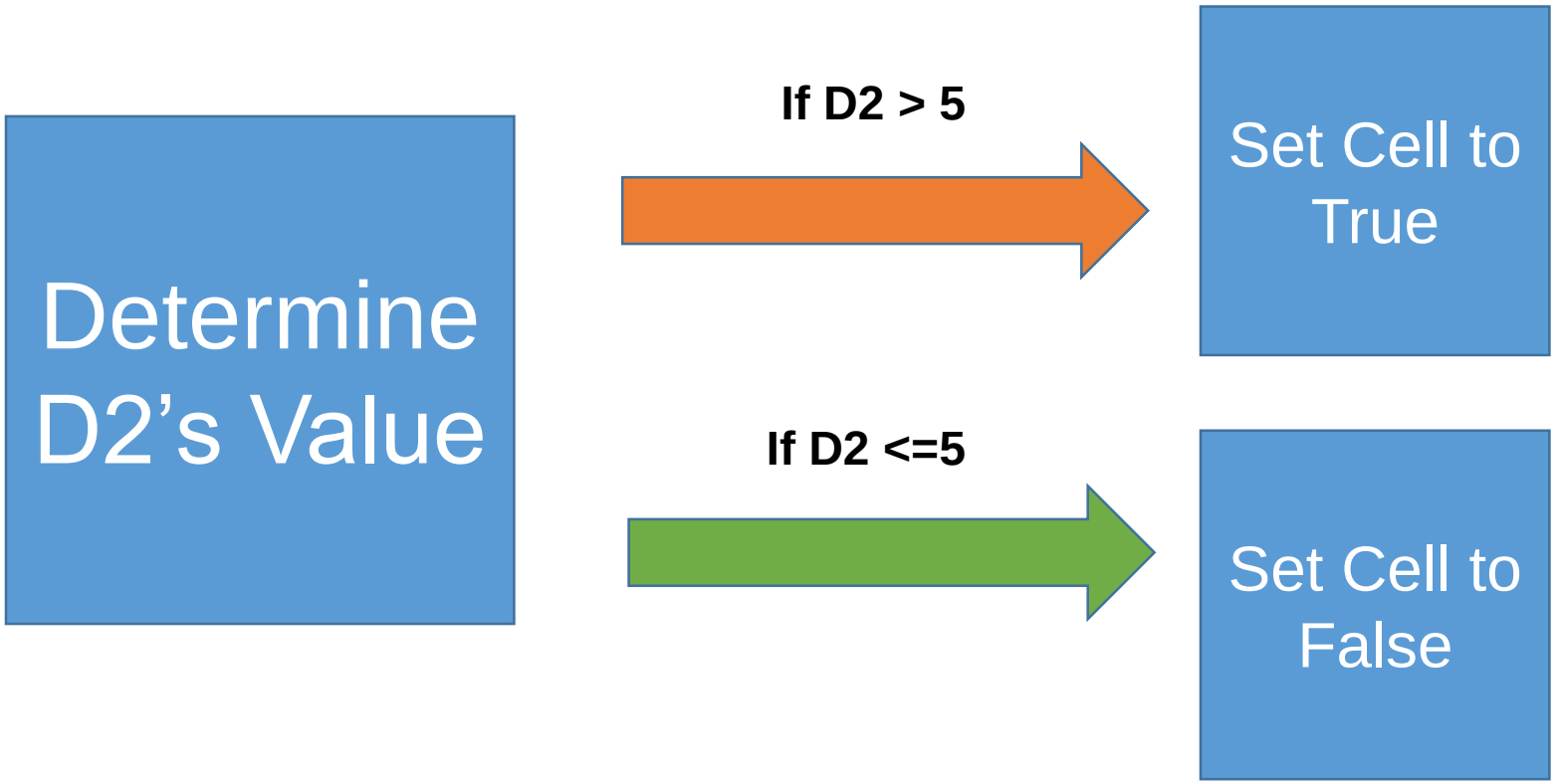
Conditionals

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* in all languages.

Conditionals: If This… Then That



=IF(D2>5,TRUE,FALSE)

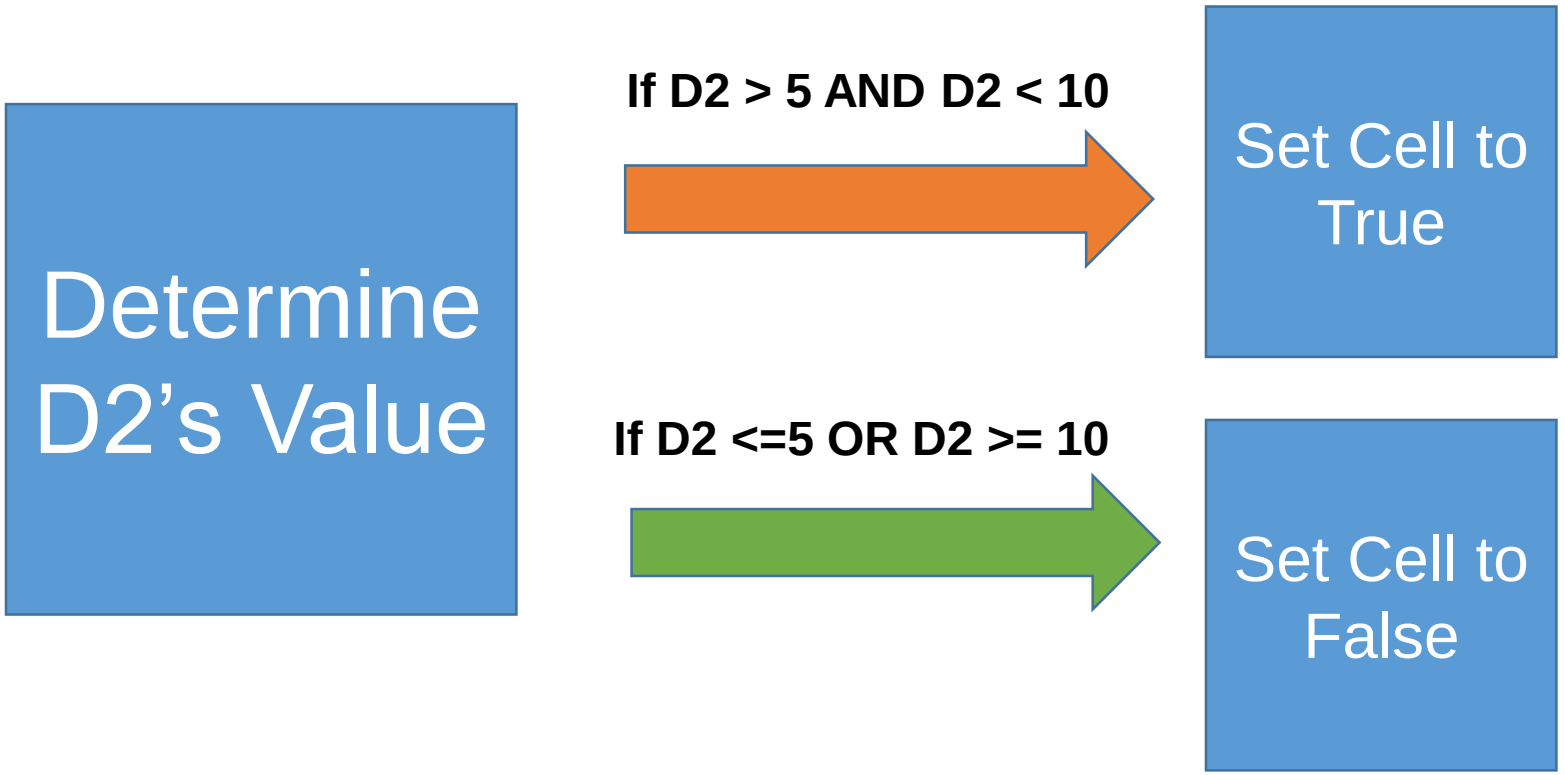
But what if… we wanted to combine conditions?

AND, NOT, OR

Conditionals: If This… Then That

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

Conditionals: If This… Then That



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

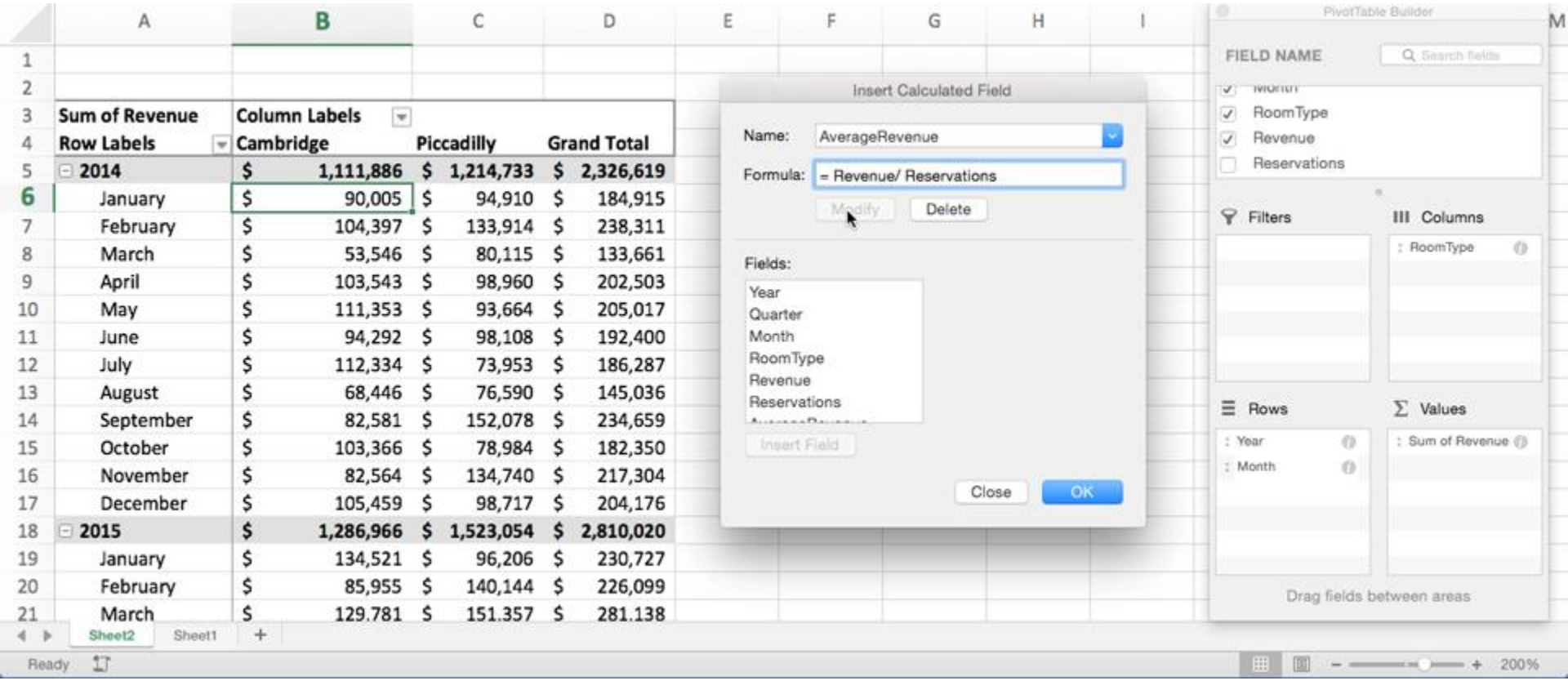
Demo Time!

(04-Stu_GradeBook – 08-McDonalds)

BREAK

Pivot Tables

Get Pivot With It



Sum of Revenue	Column Labels		
Row Labels	Cambridge	Piccadilly	Grand Total
2014	$ 1,111,886	$ 1,214,733	$ 2,326,619
January	$ 90,005	$ 94,910	$ 184,915
February	$ 104,397	$ 133,914	$ 238,311
March	$ 53,546	$ 80,115	$ 133,661
April	$ 103,543	$ 98,960	$ 202,503
May	$ 111,353	$ 93,664	$ 205,017
June	$ 94,292	$ 98,108	$ 192,400
July	$ 112,334	$ 73,953	$ 186,287
August	$ 68,446	$ 76,590	$ 145,036
September	$ 82,581	$ 152,078	$ 234,659
October	$ 103,366	$ 78,984	$ 182,350
November	$ 82,564	$ 134,740	$ 217,304
December	$ 105,459	$ 98,717	$ 204,176
2015	$ 1,286,966	$ 1,523,054	$ 2,810,020
January	$ 134,521	$ 96,206	$ 230,727
February	$ 85,955	$ 140,144	$ 226,099
March	$ 129.781	$ 151.357	$ 281.138

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

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	$32.0

In essence, Pivot tables are a **summative** analytic tool that allows us to perform aggregate functions that along any combination of fields.

(The name comes from the fact that we are pivoting along a data axis)

Words to the Wise – Keep It Flat!

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

- Modern 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 beneath their visualizations.

- Don’t try to confuse this simplicity. “Spreadsheet magic” is a nightmare to analyze.

Demo Time!

(09-PivotTables, 10-TopSongs)

Lookups

Look It Up with Lookups

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

Assume this table is gigantic…

How would we **retrieve** the population of a specific planet for use in another formula?

Look It Up with Lookups

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

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>)

What Will This Yield?

Planets

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

=vlookup(“Asteroid 9”, Planets, 3)

What Will This Yield?

Planets

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

=vlookup(“Astroid 9”, Planets, 3)

The Asterisks

Demo Time!

(11-Lookups, 12-ProductPivot)

Questions / Discussion