Poli 5D Social Science Data Analytics

Functions in Excel

Shane Xinyang Xuan
ShaneXuan.com

January 18, 2017

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Roberts	M	1600-1800 (SSB 299)
Jason Bigenho	Th	1000-1200 (Econ 116)
Shane Xuan	Th	1200-1400 (SSB 332)

Supplemental Materials

UCLA STATA starter kit

`http://www.ats.ucla.edu/stat/stata/sk/`

Princeton data analysis

`http://dss.princeton.edu/training/`

Road map

Here is what we did last week:

- Variable types
- Longitudinal/Cross sectional data
- Unit of analysis (esp. for time series and cross sectional data)
- Excel shortcuts
- Functions:
  1. IF function
  2. FIND function
  3. LEFT function

Road map

Here is what we did last week:

- Variable types
- Longitudinal/Cross sectional data
- Unit of analysis (esp. for time series and cross sectional data)
- Excel shortcuts
- Functions:
  1. IF function
  2. FIND function
  3. LEFT function

Today we are going to discuss more about functions in Excel:

1. Statistical functions: AVERAGE, MEDIAN, MIN, MAX, and COUNTIF functions
2. Lookup functions: MATCH and VLOOKUP functions

Road map

Here is what we did last week:

- Variable types
- Longitudinal/Cross sectional data
- Unit of analysis (esp. for time series and cross sectional data)
- Excel shortcuts
- Functions:
  1. IF function
  2. FIND function
  3. LEFT function

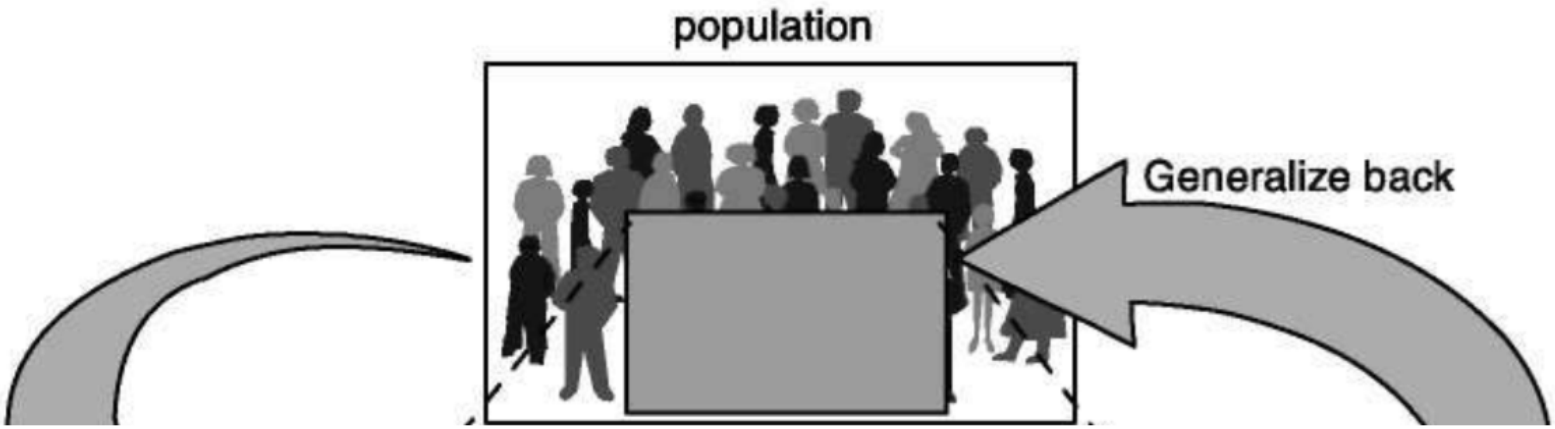
Today we are going to discuss more about functions in Excel:

1. Statistical functions: AVERAGE, MEDIAN, MIN, MAX, and COUNTIF functions
2. Lookup functions: MATCH and VLOOKUP functions

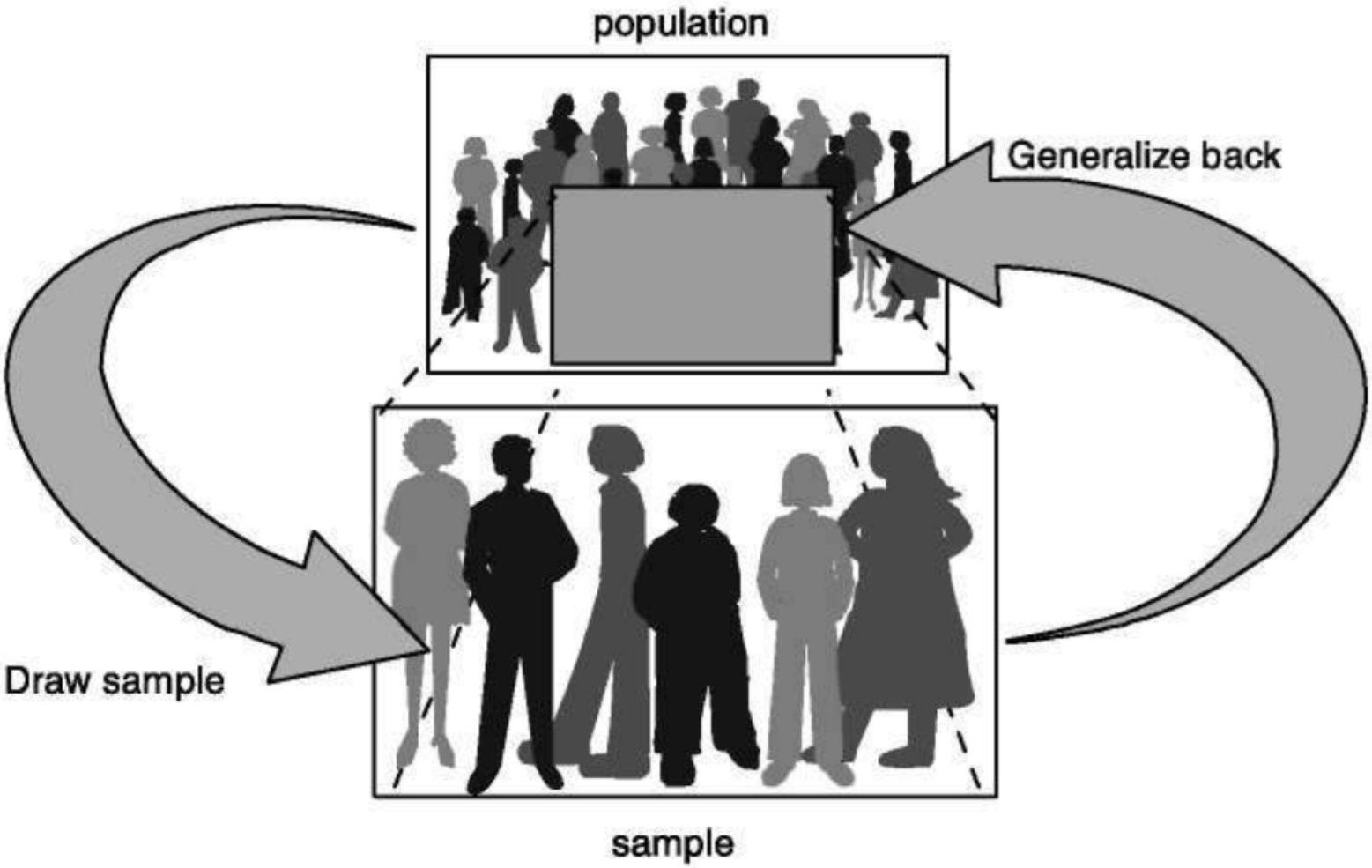
We will also be discussing some conceptual topics:

1. Sample and population
2. Bias

Sample and population



Sample and population



Sample and population

- Population
  - A collection of objects or individuals
- Sample
  - A (hopefully representative) slice from the population
- Population parameter is any summary of the population
- Sample statistic is any summary of the sample

Sample and population

- Example 1: Hite mailed out 100,000 fifteen-page questionnaires to women who were members of a wide variety of organizations across the U.S. Questionnaires were actually sent to the leader of each organization. The leader was asked to distribute questionnaires to all members. Each questionnaire contained 127 open-ended questions with many parts and follow-ups. Part of Hite's directions read as follows: "Feel free to skip around and answer only those questions you choose." Approximately 4500 questionnaires were returned. What is the population? What is the sample?
- Population: All American women
- Sample: The 4,500 women who responded

Bias

- Example 1: Problems with the previous example?
- Example 2: We want to study savings and investment decisions of adult Americans. The sample is UCSD undergraduates. Are there any problems with it?

Bias

- Example 1: Problems with the previous example?
- Example 2: We want to study savings and investment decisions of adult Americans. The sample is UCSD undergraduates. Are there any problems with it?

Bias includes

- Sampling bias
  - Selection bias
  - Undercoverage bias
- Response bias

Statistical concepts

- Average
- Median
- Minimum/Maximum

AVERAGE/MEDIAN functions

The AVERAGE function calculates the average value from a collection of numbers

- Syntax: AVERAGE (number1, number2, ...)
- For example: AVERAGE(A1:A4)

The MEDIAN function calculates the median of the values from the specified range

- Syntax: MEDIAN (number1, number2, ...)
- For example: MEDIAN(A1:A4)

Other common statistical functions

Similarly, you can use

- MAX(number1, number2, ...)
- MIN(number1, number2, ...)
- SUM(number1, number2, ...)
- ROUND(number, numDigits), where numDigits is the number of decimal places

Other common statistical functions

Similarly, you can use

- MAX(number1, number2, ...)
- MIN(number1, number2, ...)
- SUM(number1, number2, ...)
- ROUND(number, numDigits), where numDigits is the number of decimal places

Quiz: What will the result be?

ROUND(2.718282, 2)

Turn your quiz in!

Other common statistical functions

Similarly, you can use

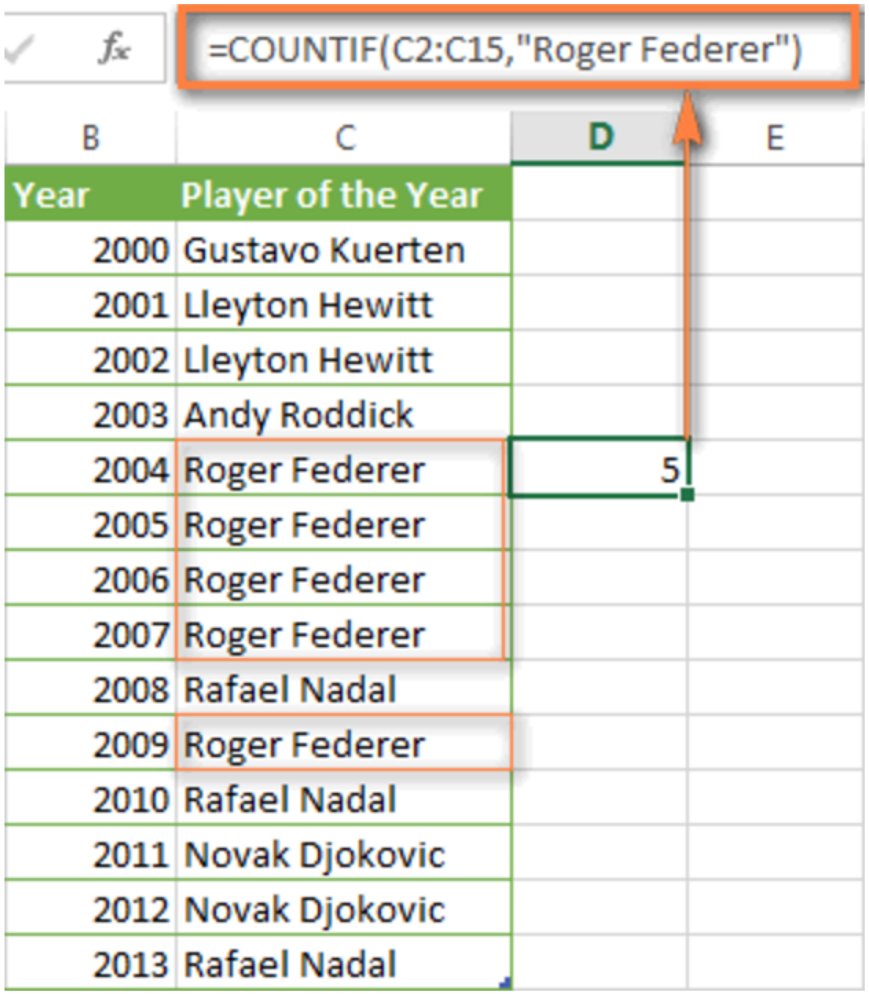
- MAX(number1, number2, ...)
- MIN(number1, number2, ...)
- SUM(number1, number2, ...)
- ROUND(number, numDigits), where numDigits is the number of decimal places

Quiz: What will the result be?
ROUND(2.718282, 2)
Turn your quiz in!
It should be 2.72.

Statistical function: COUNT, COUNTIF, ...

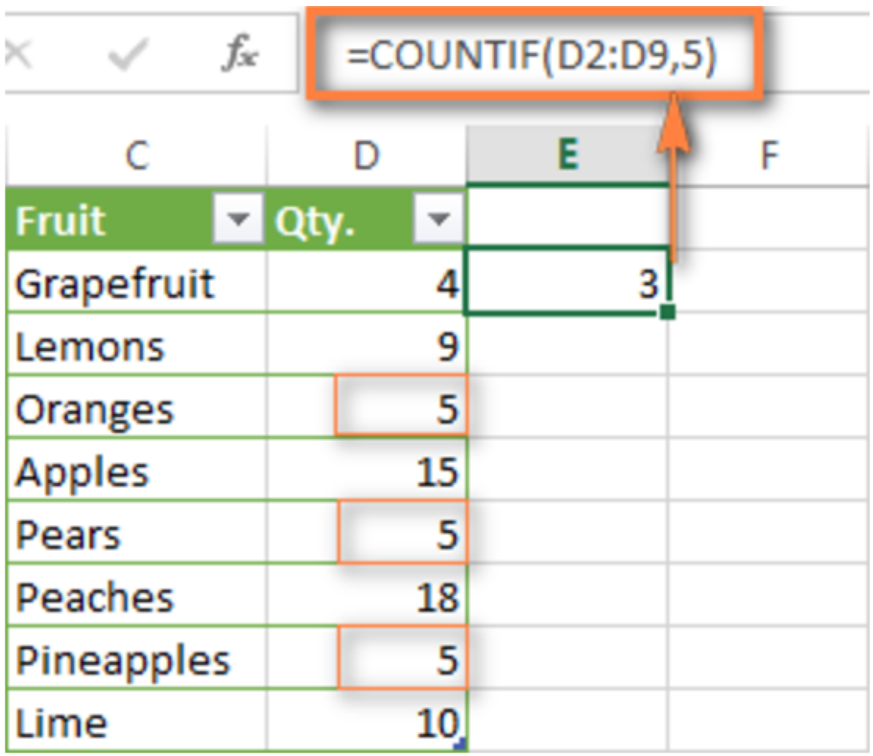
- COUNT: How many unique items are included in a range
  - Syntax: COUNT(*range*)
- COUNTIF: Calculate the number of cells in a range that match the criteria
  - Syntax: COUNTIF(*range*, *criteria*)
  - Example: Next 2 slides!
- AVERAGEIF; AVERAGEIFS; SUMIF; SUMIFS; COUNTIFS

COUNTIF: Examples



B	C	D	E
Year	Player of the Year		
2000	Gustavo Kuerten		
2001	Lleyton Hewitt		
2002	Lleyton Hewitt		
2003	Andy Roddick		
2004	Roger Federer	5	
2005	Roger Federer		
2006	Roger Federer		
2007	Roger Federer		
2008	Rafael Nadal		
2009	Roger Federer		
2010	Rafael Nadal		
2011	Novak Djokovic		
2012	Novak Djokovic		
2013	Rafael Nadal		

COUNTIF: Examples

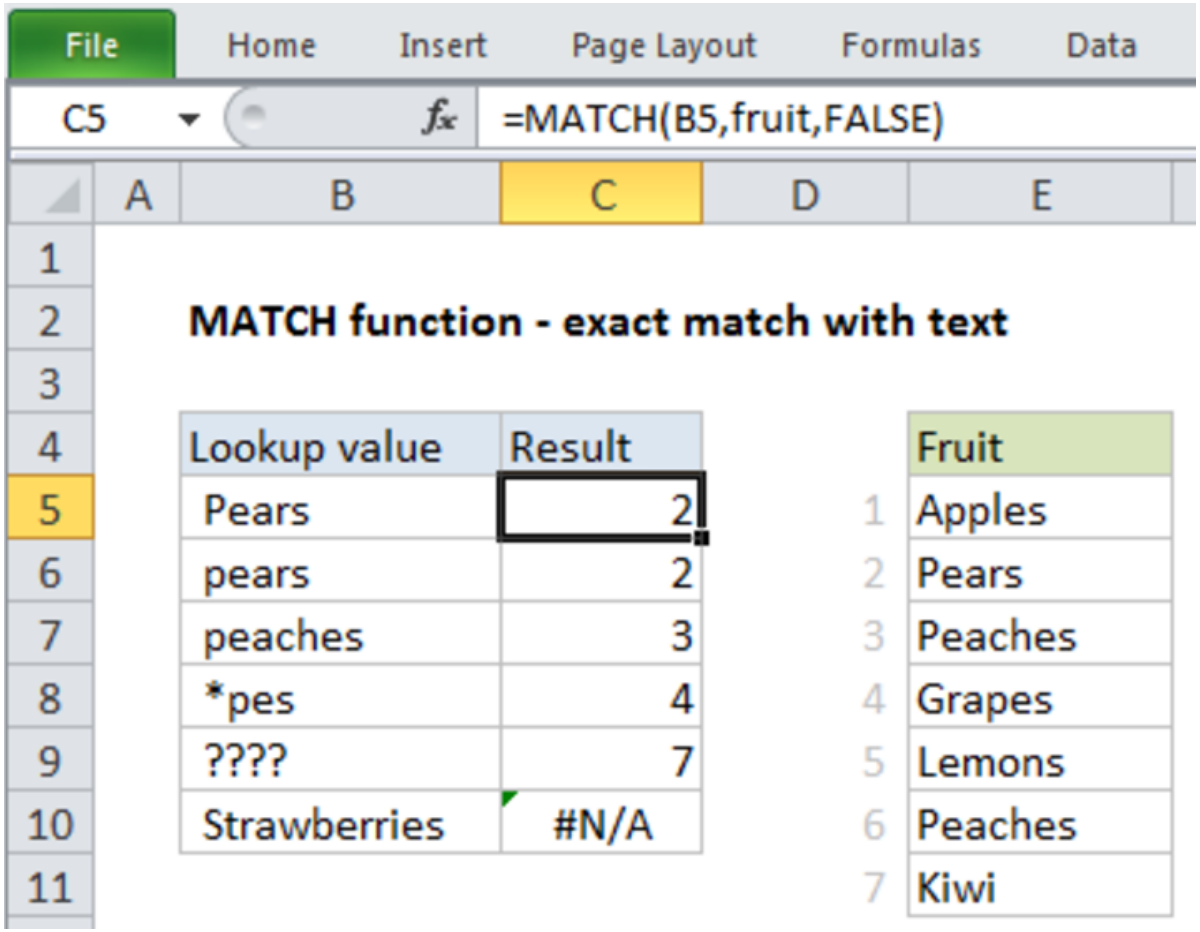


Lookup functions

The MATCH function returns the position of a value in a given range:

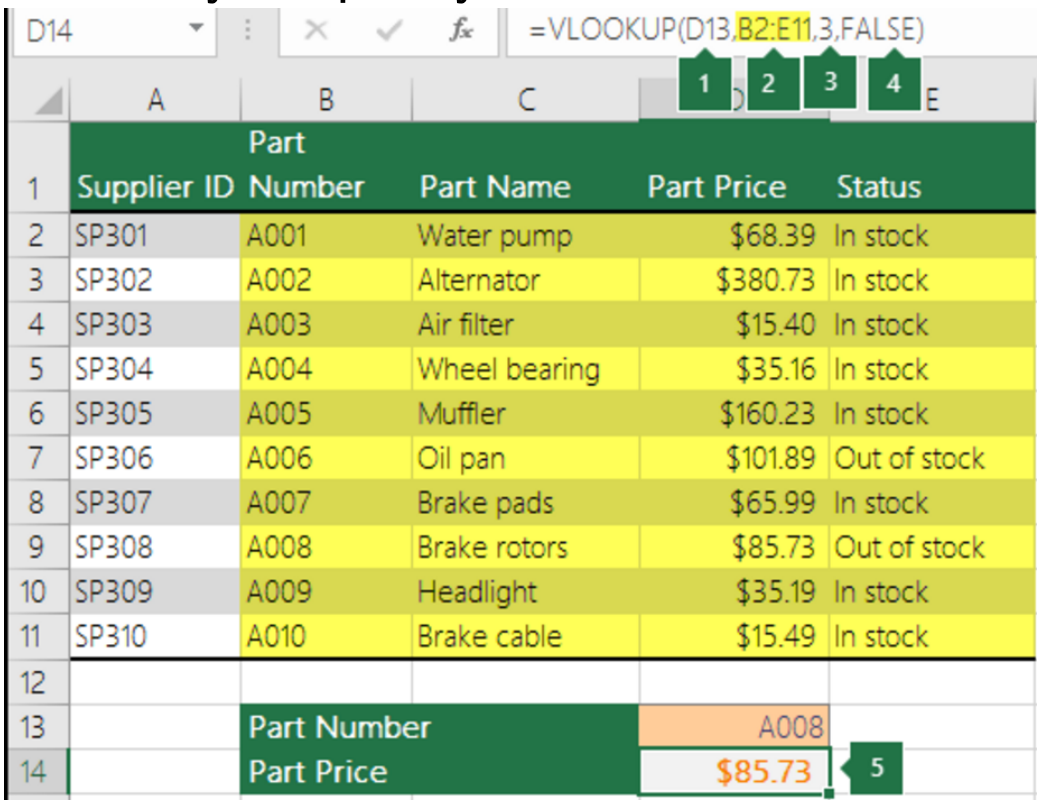
- Scenario: Among the range E4:E9, I want to look for A2
- Example: MATCH(A2, E4:E9, 0)
- Returns the position

Lookup functions



Lookup functions

The VLOOKUP function looks for a value in the leftmost column of a table, and then returns a value in the same row from another column you specify



	A	B	C	D	E
1	Supplier ID	Part Number	Part Name	Part Price	Status
2	SP301	A001	Water pump	$68.39	In stock
3	SP302	A002	Alternator	$380.73	In stock
4	SP303	A003	Air filter	$15.40	In stock
5	SP304	A004	Wheel bearing	$35.16	In stock
6	SP305	A005	Muffler	$160.23	In stock
7	SP306	A006	Oil pan	$101.89	Out of stock
8	SP307	A007	Brake pads	$65.99	In stock
9	SP308	A008	Brake rotors	$85.73	Out of stock
10	SP309	A009	Headlight	$35.19	In stock
11	SP310	A010	Brake cable	$15.49	In stock
12					
13		Part Number		A008	
14		Part Price		$85.73	

Lookup functions

D14 =VLOOKUP(D13,B2:E11,3,FALSE)

	A	B	C	D	E
1	Supplier ID	Part Number	Part Name	Part Price	Status
2	SP301	A001	Water pump	$68.39	In stock
3	SP302	A002	Alternator	$380.73	In stock
4	SP303	A003	Air filter	$15.40	In stock
5	SP304	A004	Wheel bearing	$35.16	In stock
6	SP305	A005	Muffler	$160.23	In stock
7	SP306	A006	Oil pan	$101.89	Out of stock
8	SP307	A007	Brake pads	$65.99	In stock
9	SP308	A008	Brake rotors	$85.73	Out of stock
10	SP309	A009	Headlight	$35.19	In stock
11	SP310	A010	Brake cable	$15.49	In stock
12					
13		Part Number		A008	
14		Part Price		$85.73	

1. D13 the value you want to look up
2. B2 to E11 (highlighted in yellow in the table) is the range where the lookup value is located
3. 3 is the column number (in the range) that contains the return value; in our case, it is "Part Price"
4. FALSE makes sure that the return will be an exact match
5. Output is 85.73

Lookup functions

Other REALLY important functions

- HLOOKUP
- INDEX
- CHOOSE

Wrap up

- All functions (including last week) that we talked about are important. Any questions?
- Pace of the section: Too fast? Too slow?
- HW1 due on 1/25
- Start early because I will NOT answer any emails starting from 1/24 after noon