



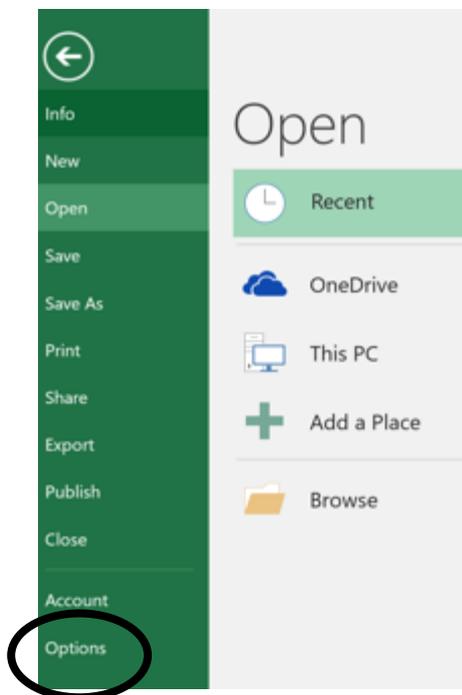
Basic Data Analysis Using MS Excel

Teresa McCoy
Assistant Director, University of Maryland Extension
Evaluation & Assessment

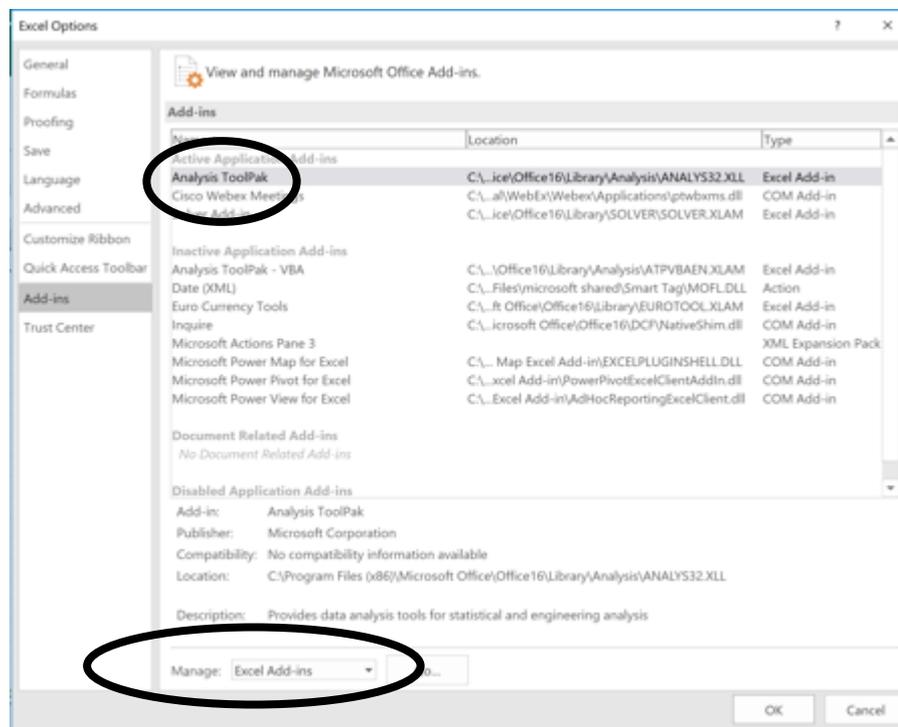


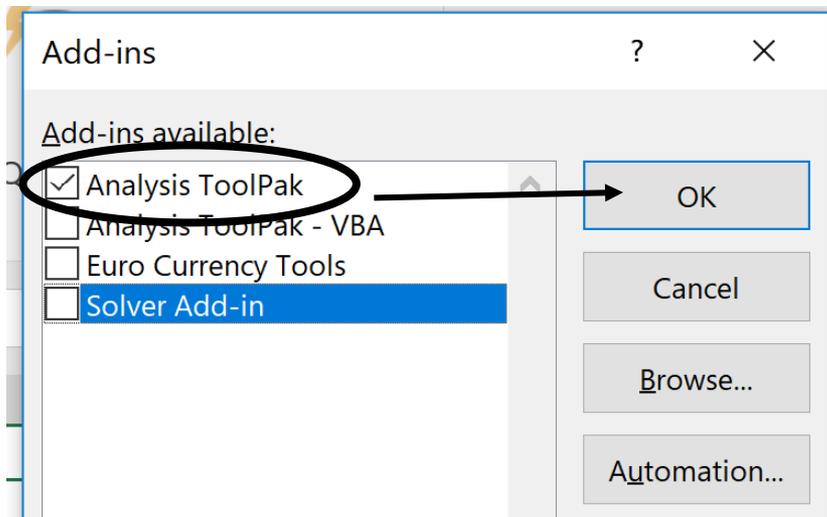
First Things First: Toolpak

Click on “File” then on “Options”



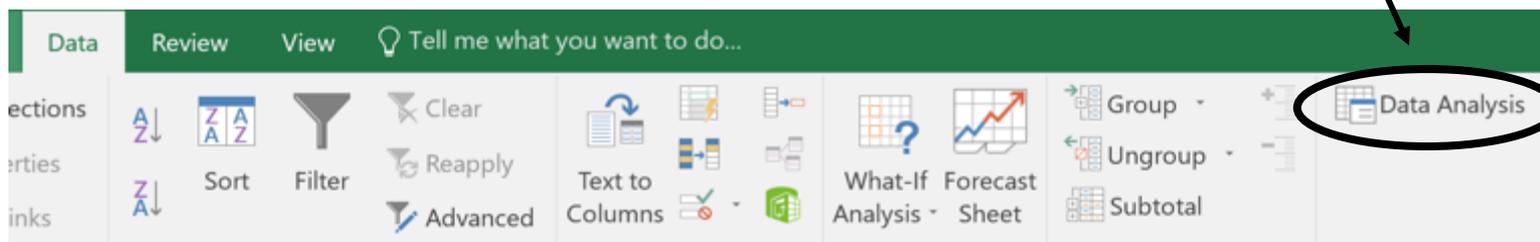
Click on “Add Ins” then on “Manage ... Go”





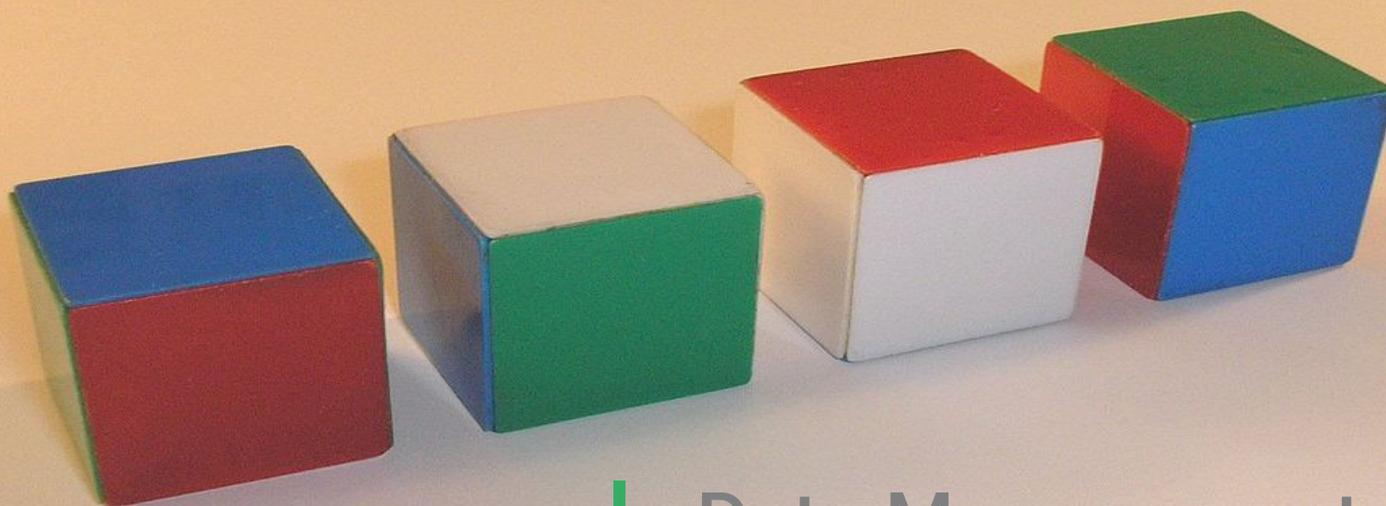
Click on the box, “Analysis Toolpak,” and Then click on “OK”

You will see the “Data Analysis” option appear on the right-hand side of your ribbon.





Overview



- I. Data Management
- II. Data Manipulation**
- III. Data Analysis**
- IV. Data Communication



Preface

Relax and just listen

Beginner pace

Exposure to the Excel platform

Exposure to T-Tests

Tips & Tricks

Support Materials



Data Manipulation



1) Sort

2) Filter

Tip: Auto- or Flash-fill



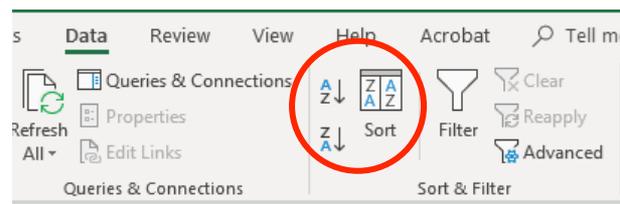
Sort

	A	B	C	D	E	F	G	
1	State	County	Age	Grade	Gender	Race	4-H Hours	
2	CT	Hartford	9	4	1	1	1	
3	CT	Hartford	15	12	1	7	1	
4	CT	Hartford	17	2	2	2	3	
5	CT	Tolland	17	10	1	7	2	
6	CT	Tolland	16	11	1	7	3	
7	CT	Tolland	14	4	1	3	6	
8	CT	Tolland	11	8	2	2	6	
9	CT	New London	6	12	2	3	1	
10	CT	New London	10	9	2	1	3	
11	CT	New London	17	8	2	5	5	
12	CT	New London	15	3	2	6	6	
13	CT	Litchfield	15	6	2	4	4	
14	CT	Litchfield	14	8	2	2	2	
15	CT	Litchfield	14	6	2	1	5	
16	CT							
17	CT	State	County	Age	Grade	Gender	Race	4-H Hours
18	CT	Hartford	9	4	1	1	1	
19	CT	Hartford	15	12	1	7	1	
20	CT	Hartford	17	2	2	2	3	
21	CT	Tolland	17	10	1	7	2	
22	CT	Tolland	16	11	1	7	3	
23	CT	Tolland	14	4	1	3	6	
24	CT	Tolland	11	8	2	2	6	
25	CT	New London	6	12	2	3	1	
26	CT	New London	10	9	2	1	3	
27	CT	New London	17	8	2	5	5	
28	CT	New London	15	3	2	6	6	
29	CT	Litchfield	15	6	2	4	4	
30	CT	Litchfield	14	8	2	2	2	
31	CT	Litchfield	14	6	2	1	5	
32	CT	Litchfield	10	7	1	2	6	
33	CT	Fairfield	12	2	2	7	2	
34	CT	Fairfield	9	10	2	4	2	
35	CT	Fairfield	5	3	1	3	4	
36	CT	Fairfield	10	7	1	5	5	
37	CT	Windham	11	10	1	6	3	
38	CT	Windham	16	6	1	5	6	
39	CT	Windham	12	2	1	3	3	
40	CT	Middlesex	9	6	2	2	2	
41	CT	Middlesex	13	5	1	5	3	
42	CT	Middlesex	12	11	2	3	1	

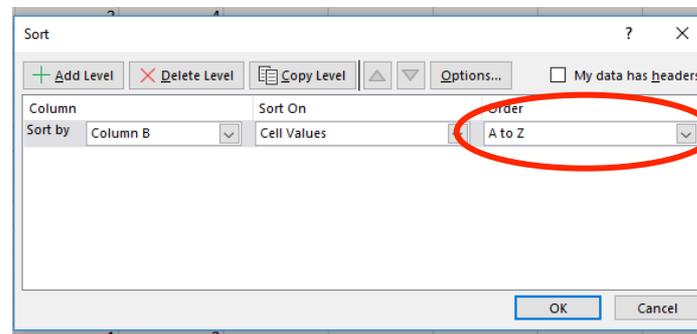
Select the data in my spreadsheet.

Go to "Data" tab.

Select "Sort."

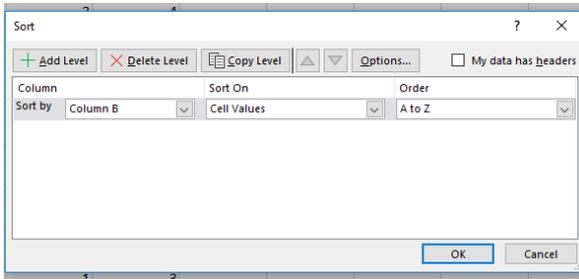


I'm going to sort by Column B (county) in alphabetical order.



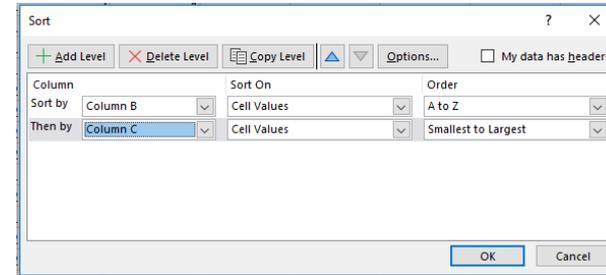


Sorted by Column A alphabetically



	A	B	C	D	E	F	G
1	State	County	Age	Grade	Gender	Race	4-H Hours
2	CT	Fairfield	12	2	2	7	2
3	CT	Fairfield	9	10	2	4	2
4	CT	Fairfield	5	3	1	3	4
5	CT	Fairfield	10	7	1	5	5
6	CT	Hartford	9	4	1	1	1
7	CT	Hartford	15	12	1	7	1
8	CT	Hartford	17	2	2	2	3
9	CT	Litchfield	15	6	2	4	4
10	CT	Litchfield	14	8	2	2	2
11	CT	Litchfield	14	6	2	1	5
12	CT	Litchfield	10	7	1	2	6
13	CT	Middlesex	9	6	2	2	2
14	CT	Middlesex	13	5	1	5	3
15	CT	Middlesex	12	11	2	3	1
16	CT	New London	6	12	2	3	1
17	CT	New London	10	9	2	1	3
18	CT	New London	17	8	2	5	5
19	CT	New London	15	3	2	6	6
20	CT	Tolland	17	10	1	7	2
21	CT	Tolland	16	11	1	7	3
22	CT	Tolland	14	4	1	3	6
23	CT	Tolland	11	8	2	2	6
24	CT	Windham	11	10	1	6	3
25	CT	Windham	16	6	1	5	6
26	CT	Windham	12	2	1	3	3

Sorted by Column B alphabetically and then by Column C smallest to largest

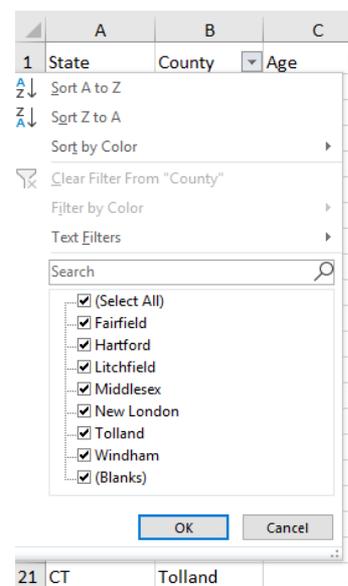
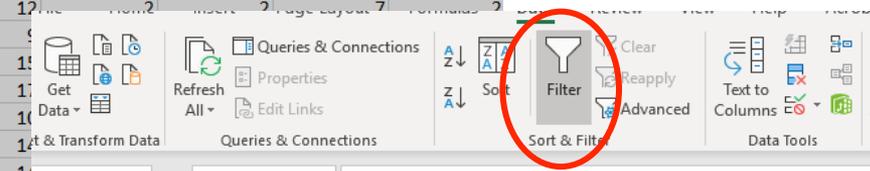


	A	B	C	D	E	F	G
1	State	County	Age	Grade	Gender	Race	4-H Hours
2	CT	Fairfield	5	3	1	3	4
3	CT	Fairfield	9	10	2	4	2
4	CT	Fairfield	10	7	1	5	5
5	CT	Fairfield	12	2	2	7	2
6	CT	Hartford	9	4	1	1	1
7	CT	Hartford	15	12	1	7	1
8	CT	Hartford	17	2	2	2	3
9	CT	Litchfield	10	7	1	2	6
10	CT	Litchfield	14	8	2	2	2
11	CT	Litchfield	14	6	2	1	5
12	CT	Litchfield	15	6	2	4	4
13	CT	Middlesex	9	6	2	2	2
14	CT	Middlesex	12	11	2	3	1
15	CT	Middlesex	13	5	1	5	3
16	CT	New London	6	12	2	3	1
17	CT	New London	10	9	2	1	3
18	CT	New London	15	3	2	6	6
19	CT	New London	17	8	2	5	5
20	CT	Tolland	11	8	2	2	6
21	CT	Tolland	14	4	1	3	6
22	CT	Tolland	16	11	1	7	3
23	CT	Tolland	17	10	1	7	2
24	CT	Windham	11	10	1	6	3
25	CT	Windham	12	2	1	3	3
26	CT	Windham	16	6	1	5	6



Sort & Filter

	A	B	C	D	E	F	G
1	State	County	Age	Grade	Gender	Race	4-H Hours
2	CT	Fairfield	5	3	1	3	4
3	CT	Fairfield	9	10	2	4	2
4	CT	Fairfield	10	7	1	5	5
5	CT	Fairfield	12	2	2	7	2
6	CT	Hartford	9	4	1	1	1
7	CT	Hartford	15	12	1	7	1
8	CT	Hartford	17	2	2	2	3
9	CT	Litchfield	10	7	1	2	6
10	CT	Litchfield	14	8	2	2	2
11	CT	Litchfield	14	6	2	1	5
12	CT	Litchfield	15	6	2	4	4
13	CT	Middlesex	9	6	2	2	2
14	CT	Middlesex	12	11	2	3	1
15	CT	Middlesex	13	5	1	5	3
16	CT	New London	6	12	2	3	1
17	CT	New London	10	9	2	1	3
18	CT	New London	15	3	2	6	6
19	CT	New London	17	8	2	5	5
20	CT	Tolland	11	8	2	2	6
21	CT	Tolland	14	4	1	3	6
22	CT	Tolland	16	11	1	7	3
23	CT	Tolland	17	10	1	7	2
24	CT	Windham	11	10	1	6	3
25	CT	Windham	12	2	1	3	3
26	CT	Windham	16	6	1	5	6



State	County	Age	Grade	Gender	Race	4-H Hours
CT	Hartford	9	4	1	1	1
CT	Hartford	15	12	1	7	1
CT	Hartford	17	2	2	2	3



Auto or Flash Fill: Can fill in the same value or formula or a sequence

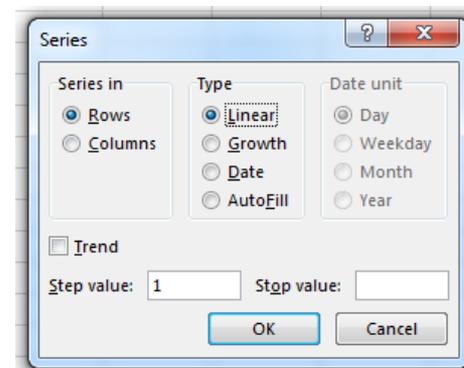
	A	B	C	D
258	MA26115	Mt. Airy	2015	0
259	MA26215	Mt. Airy	2015	0
260	MA26315	Mt. Airy	2015	0
261	MA26415	Mt. Airy	2015	0
262	MA26515	Mt. Airy	2015	0
263	MA26615	Mt. Airy	2015	0
264	MA26715	Mt. Airy	2015	0
265	MA26815	Mt. Airy	2015	0
266	MA26915	Mt. Airy	2015	0
267	MA27015	Mt. Airy	2015	0

Grab the handle at the bottom right corner. Hold and drag

266	MA26915	Mt. Airy
267	MA27015	Mt. Airy
268	MA27016	
269	MA27017	
270	MA27018	
271	MA27019	
272	MA27020	

266	MA26915	Mt. Airy
267	MA27015	Mt. Airy
268	MA27016	Mt. Airy
269	MA27017	Mt. Airy
270	MA27018	Mt. Airy
271	MA27019	Mt. Airy
272	MA27020	Mt. Airy

There is also a “Fill” option on the Home Menu Ribbon where you can set up multiple fill options.





III. Data Analysis

Descriptive Statistics:

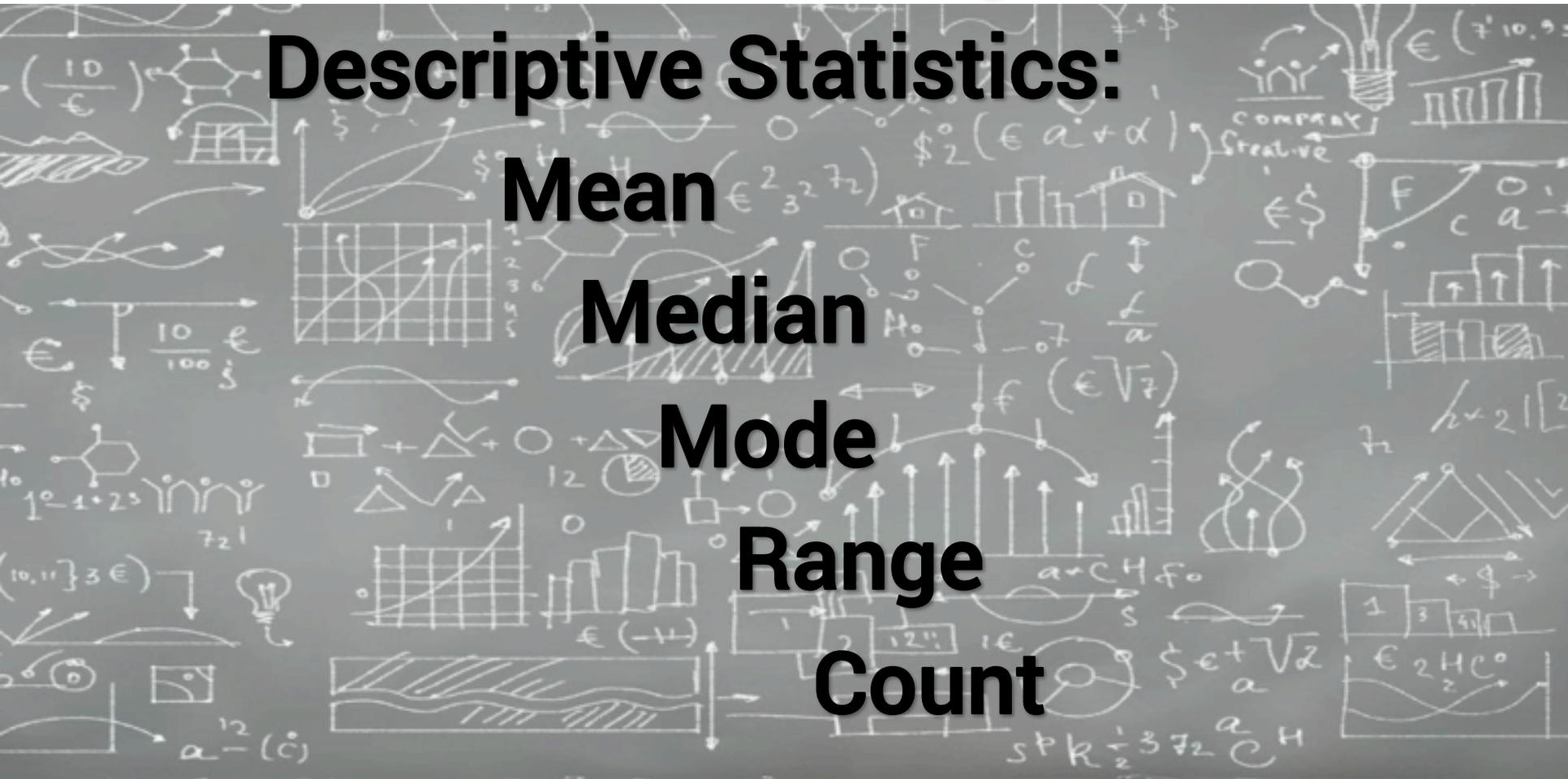
Mean

Median

Mode

Range

Count

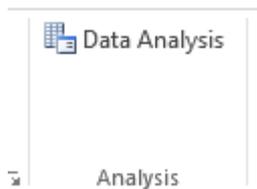




Caveat

Multiple ways to go about your analysis.

1) Toolpak



2) Formulas



3) Manual entry

149	CT	Middlesex	9
150	CT	Middlesex	8
151	CT	Middlesex	6
152			=sum(C134:C151)



Descriptive Stats Using Toolpak

	A	B	C	D	E	F	G
2	Hartford	15	9	1	2	2	
3	Hartford	9					
4	Hartford	9					
5	Hartford	15					
6	Hartford	17					
7	Hartford	15					
8	Hartford	17					
9	Hartford	5					
0	Hartford	5					
1	Hartford	7					
2	Hartford	14					
3	Hartford	13					
4	Hartford	12					
5	Hartford	11					
6	Hartford	14					
7	Hartford	5					
8	Hartford	6					
9	Hartford	16					
0	Hartford	6					

Descriptive Statistics

Input
 Input Range:
 Grouped By: Columns Rows
 Labels in first row

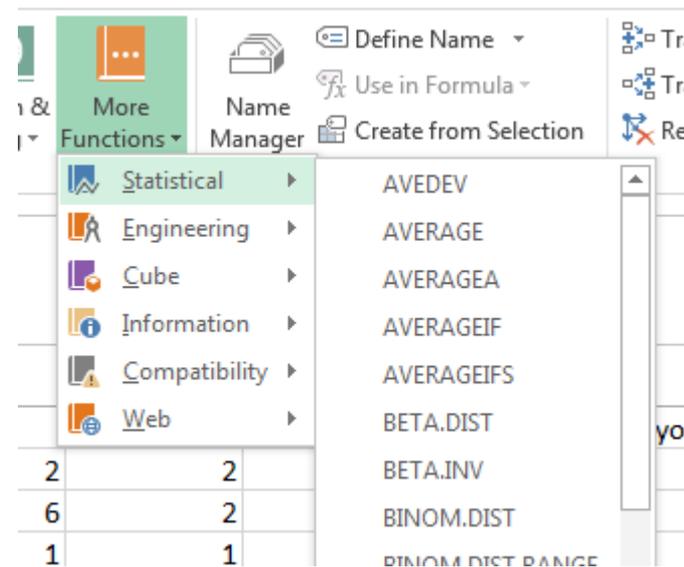
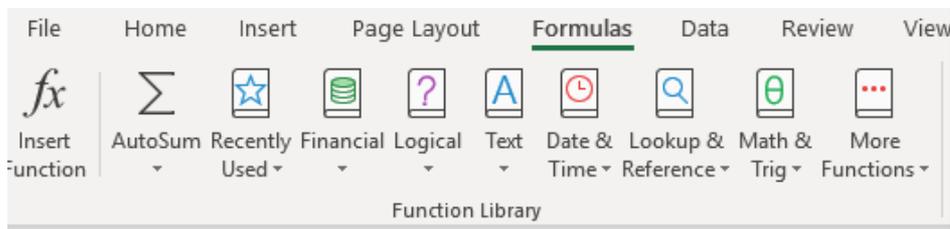
Output options
 Output Range:
 New Worksheet Ply:
 New Workbook
 Summary statistics
 Confidence Level for Mean: %
 Kth Largest:
 Kth Smallest:

	15
Mean	11.590604
Standard Error	0.33041792
Median	12
Mode	14
Standard Dev	3.91119914
Sample Varia	15.2974787
Kurtosis	-1.1611581
Skewness	0.1057812
Range	13
Minimum	5
Maximum	18
Sum	1727
Count	149
Confidence L	0.63318506

N=149



Using Formulas from the Formula Tab



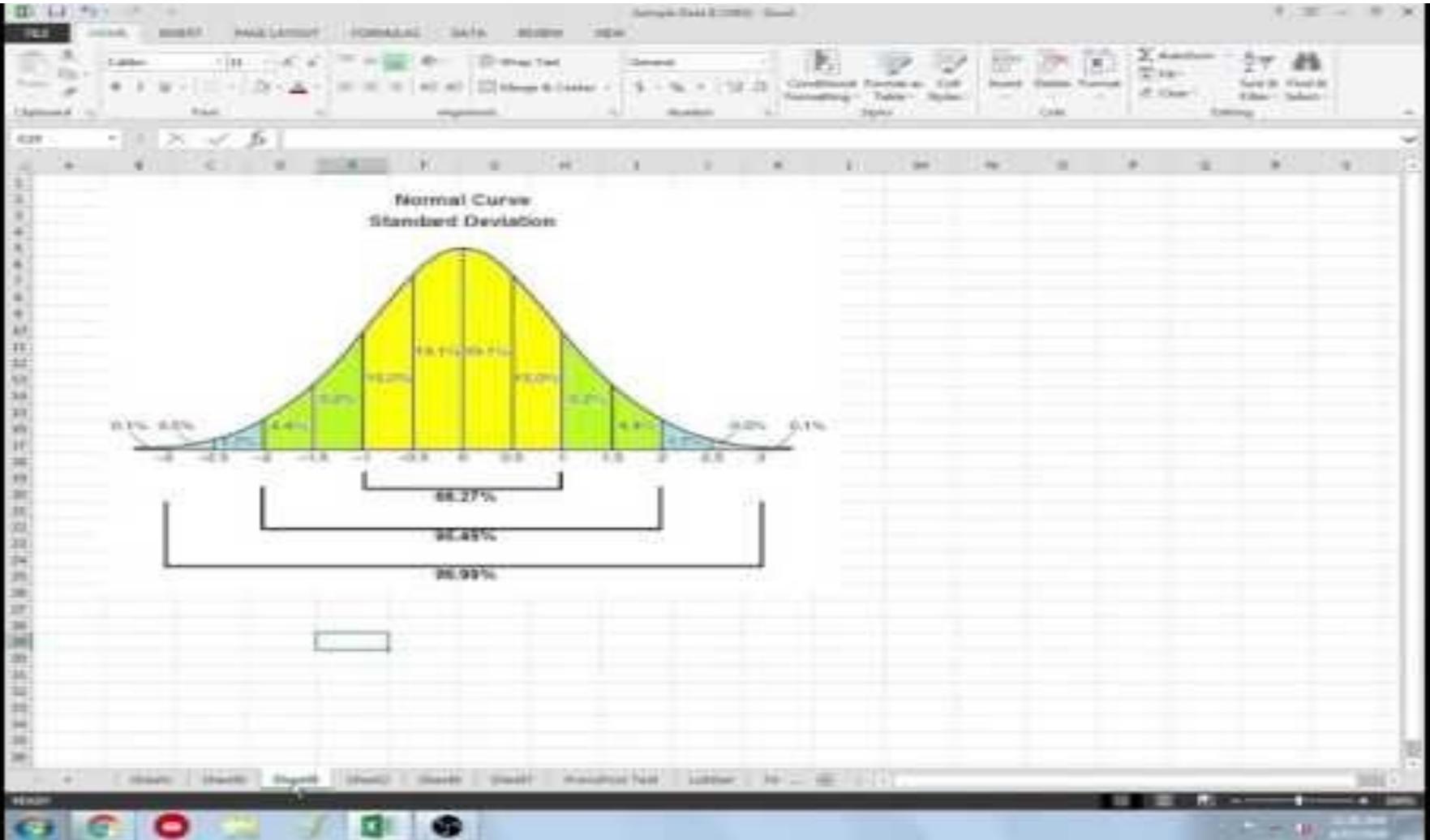
P-Value shows as .36

1	# of Hours			
2	Boy	Girl		
3		2	2	0.364016
4		1	3	
5		1	5	



III. Data Analysis

Infer (Inferential Statistics):
Paired & Un-paired T-Tests
P-Value (probability value)





Review Main Points

- T-test looks for significant difference between two sets of data.
- Eliminates our bias.
- Science wants a 95% confidence level.
- T-test returns the p-value (probability).
- Want a p-value that is $<.05$
- Because that says the difference in the data is NOT attributable to random chance.



Review Main Points

**If the results are not
attributable to chance,
Then our program intervention
worked.**



Question

Is there a significant difference between girls and boys in the number of hours spent in 4-H activities each week?

Open-ended; therefore a two-tailed t-test



Question

Do boys spend more time each week in 4-H activities than girls?

Closed-ended; one direction; therefore a one-tailed t-test



Paired or Unpaired Data

	Student ID	Gender	Hours
1			
2	CT1	1	2
3	CT2	2	2
4	CT3	1	1
5	CT4	1	1
6	CT5	2	3
7	CT6	1	4
8	CT7	2	5
9	CT8	2	2
0	CT9	1	3
1	CT10	1	5
2	CT11	2	2
3	CT12	1	6
4	CT13	1	6
5	CT14	2	1
6	CT15	1	6
7	CT16	2	1
8	CT17	2	2
9	CT18	1	1
0	CT19	2	3
1	CT20	2	4
2	CT21	2	2
3	CT22	2	4
4	CT23	1	1
5	CT24	1	5
6	CT25	2	2
7	CT26	2	3
8	CT27	2	5
9	CT28	2	6

Paired



Paired or Unpaired Data

Average # of Hours Spent in 4-H Activities/Week				
Boys	Girls			
2	2			
1	3			
1	5			
4	2			
3	2			
5	1			
6	1			
6	2			
6	3			
1	4			
1	2			
5	4			
3.416667	2.583333			

Unpaired



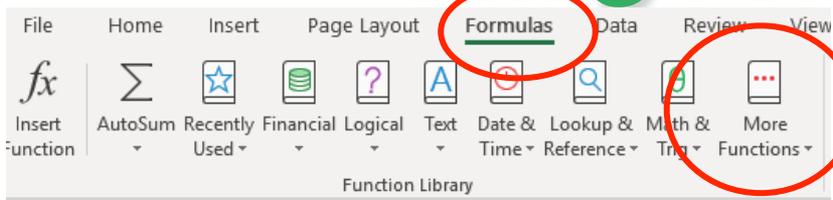
Let's go back to a question

Is there a significant difference between girls and boys in the number of hours spent in 4-H activities each week?

Open-ended; therefore a two-tailed t-test



Two-tailed T-Test on Unpaired Data Using Formula Bar



# of Hours	
Boy	Girl
2	2
1	3
1	5
4	2
3	2
5	1
6	1
6	2
6	3
1	6
2	1
3	4
3	1
2	2
4	5
5	4
5	3
2	2
3	3
7	2
5	5

Function Arguments

T.TEST

Array1 = array

Array2 = array

Tails = number

Type = number

=

Returns the probability associated with a Student's t-Test.

Array1 is the first data set.

Formula result =

[Help on this function](#) OK Cancel



T-Test Result of $p=.36$

1	# of Hours		
2	Boy	Girl	
3	2	2	0.364016
4	1	3	
5	1	5	

Can we be confident that there is a significant difference between girls and boys in the number of hours spent in 4-H activities each week?

No



Two-tailed T-Test on Unpaired Data Using Toolpak

G	H	I	J	K
t-Test: Two-Sample Assuming Unequal Variances				
	<i>Variable 1</i>	<i>Variable 2</i>		
Mean	3.630769	3.388235		
Variance	2.424038	2.859384		
Observations	65	85		
Hypothesized Mean Difference	0			
df	143			
t Stat	0.910645			
P(T<=t) one-tail	0.182007			
t Critical one-tail	1.655579			
P(T<=t) two-tail	0.364015			
t Critical two-tail	1.976692			



Remember,

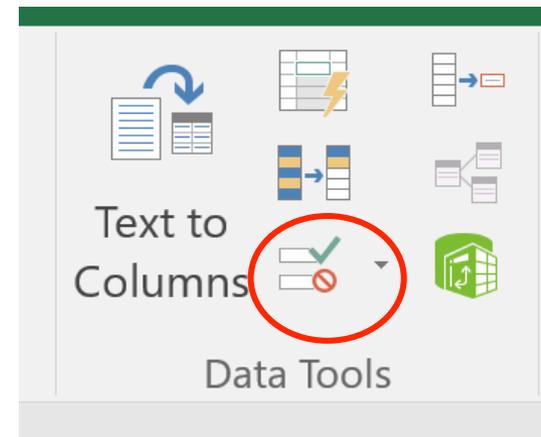
- Look back at the video.
- Don't be intimidated.
- Ask for help.



I. Data Management

Data Entry Quality Tool = Data Validation

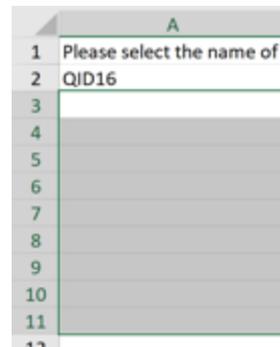
You can use data validation to restrict the type of data or the values that users enter into a cell. One of the most common data validation uses is to [create a drop-down list](#).



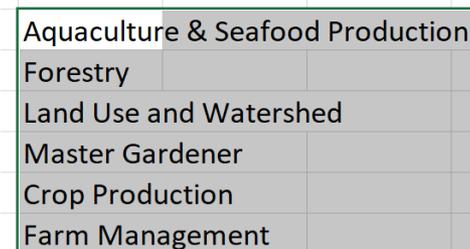


Data Validation

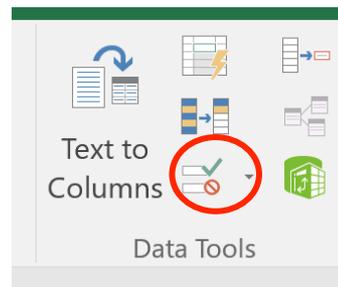
- 1) Select the cells that you want to include in the data validation.
- 2) Choose an empty space in your spreadsheet and type in the answers you will accept.
- 3) Click on your data validation option.



This is the area on the sheet for data entry.



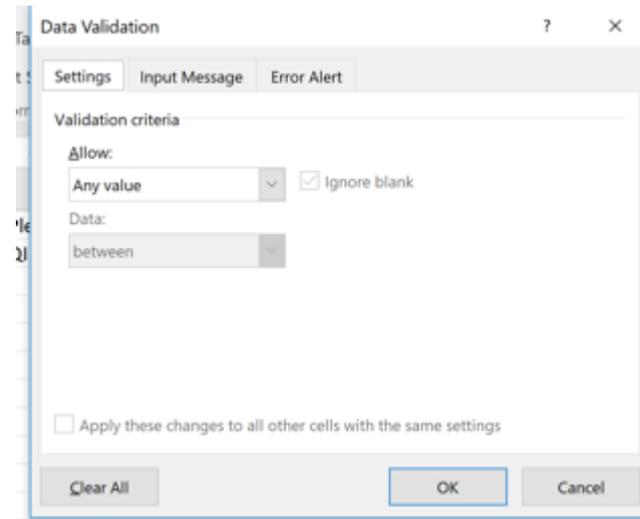
These are the only entries I will allow in the cells.





Data Validation

- 4) The data validation dialog box will open.
- 5) In "Allow," choose the drop-down box and choose "List."
- 6) You will then have to tell Excel what list to use.
- 7) Go back to the area of the sheet where you typed in the allowable entries and drag across those boxes.



Aquaculture & Seafood Production			
Forestry			
Land Use and Watershed			
Master Gardener			
Crop Production			
Farm Management			



Data Validation

- Users can only choose what you have listed.
- This eliminates user error in data entry.

C	D
Please select the workshop	The presenter was
QID19	QID11
	<input type="text"/>

Please choose a focus area
 Please choose one of the focus areas from this list. Click on the upside down triangle to see the list.