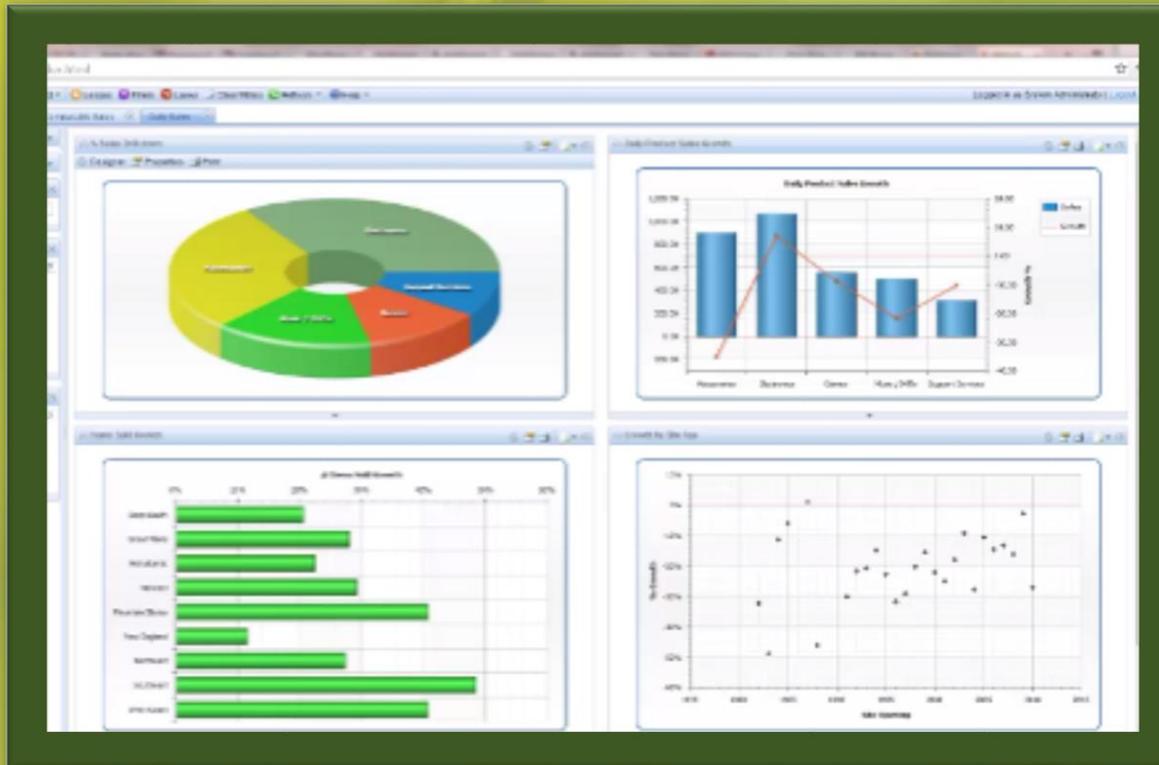


**The Giro d'Italia\* celebrates its centenary edition in 2017. What other cycling races have reached such an important milestone?**

- The Tour de France
- Ronde van Vlaanderen,
- Paris-Roubaix,
- Giro di Lombardia

\*The Giro d'Italia or as we in America would say, the Tour of Italy, (aka Corsa Rosa or Giro), is an annual multiple-stage bicycle race primarily held in Italy, while also occasionally passing through nearby countries.



# Using Excel for Analysis

Micki Harris, MBA  
 Forevermore, llc

# Agenda

- Using Excel for Analysis
- Why Use Excel?
- Limitations of Excel
- Statistical Functions in Excel
- Additional Resources

# Using Excel for Analysis

The aim of this webinar is to provide an introduction to using Microsoft Excel data analysis.

Excel has several statistical analysis functions and other tools that can be used to run descriptive statistics and to perform several different and useful inferential statistical tests.

Most of the functionality that I will refer to is available in Excel versions 2007 – 2016.

- I will be using Excel 2016 to demonstrate today's Excel features.

I am assuming that you are already familiar with the basics of using Excel.

- If you have never used Excel, I will list a few resources at the end of this webinar.

# Why Use Excel?

With so many specialist software packages available, why use Excel for statistical analysis?

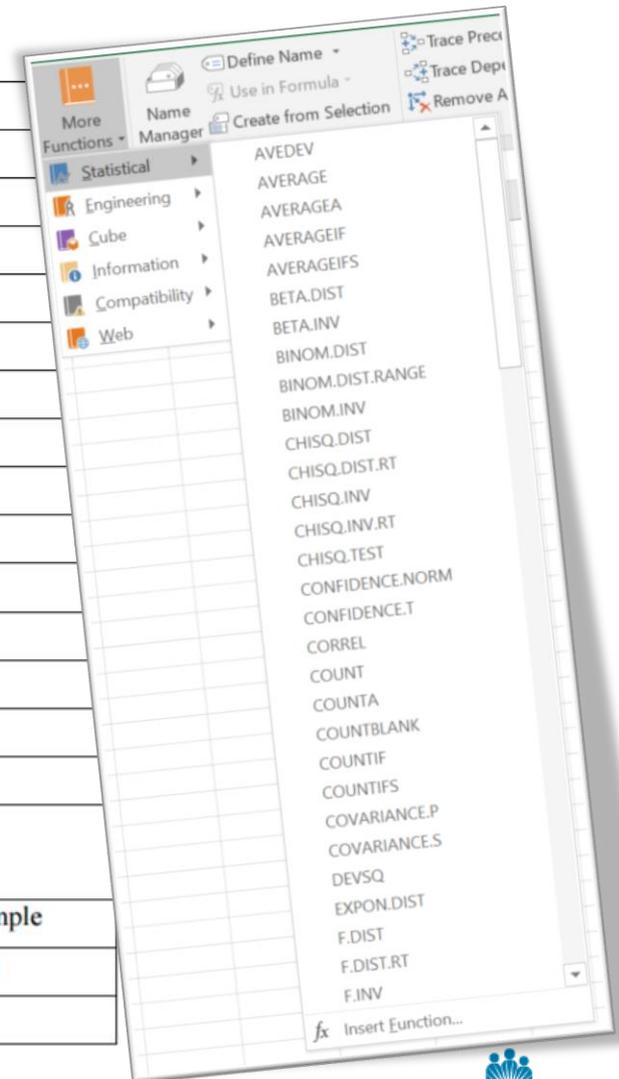
- Convenience
- Cost
- Remove the need to learn a software program
- Excel integrates easily into other Microsoft Office software products, which can be helpful when preparing reports or presentations.

# Limitations of Excel

- Excel does have limitations for statistical analysis.
  - It remains first and foremost a spreadsheet package.
  - It does not cover many of the more advanced statistical techniques that are used in research.
- Problems of using Excel for statistical analysis include:
  - Missing values are handled inconsistently, and sometimes incorrectly
  - There is no record of how an analysis was accomplished

# Built-in Statistical Functions in Excel That Are Especially Relevant to Data Analysis

Function name	Description
AVERAGE	Returns the arithmetic mean (average) of the given numbers
CHISQ.DIST.RT	Returns the right-tailed probability for the chi squared distribution
CHISQ.TEST	Returns the $p$ -value for the chi-squared test of association
CONFIDENCE.T	Returns the margin of error for a confidence interval for the mean
COUNT	Counts the number of cells in a range that contain numbers
COUNTIF	Counts the number of cells in a range that meet a given condition
COUNTA	Counts the number of cells in a range that are not empty
KURT	Returns the kurtosis of a dataset
MAX	Returns the maximum value of the given numbers
MEDIAN	Returns the median of the given numbers
MIN	Returns the minimum value of the given numbers
MODE.SNGL	Returns the mode of the given numbers
PEARSON	Returns the Pearson correlation coefficient ( $r$ ) of two variables
SKEW	Returns the skewness of a dataset
STDEV.P	Returns the standard deviation of the given numbers, based on the population
STDEV.S	Returns the standard deviation of the given numbers, based on a sample
VAR.P	Returns the variance of the given numbers, based on the population
VAR.S	Returns the variance of the given numbers, based on a sample



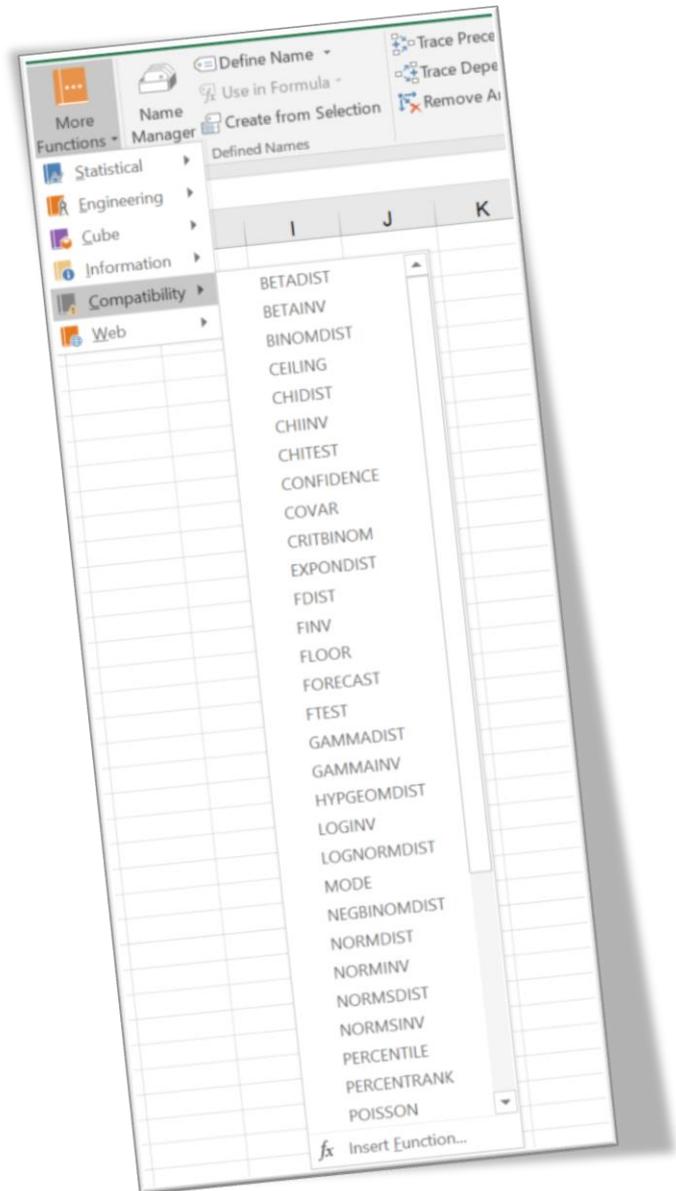
# Statistical Functions pre-2010

If you are creating a spreadsheet for users of older Excel versions, you will want to use the older functions.

You will not find them on the Statistical Functions menu. They have their own menu.

To find it, select

Formulas - More functions - Compatibility



# Example Worksheets

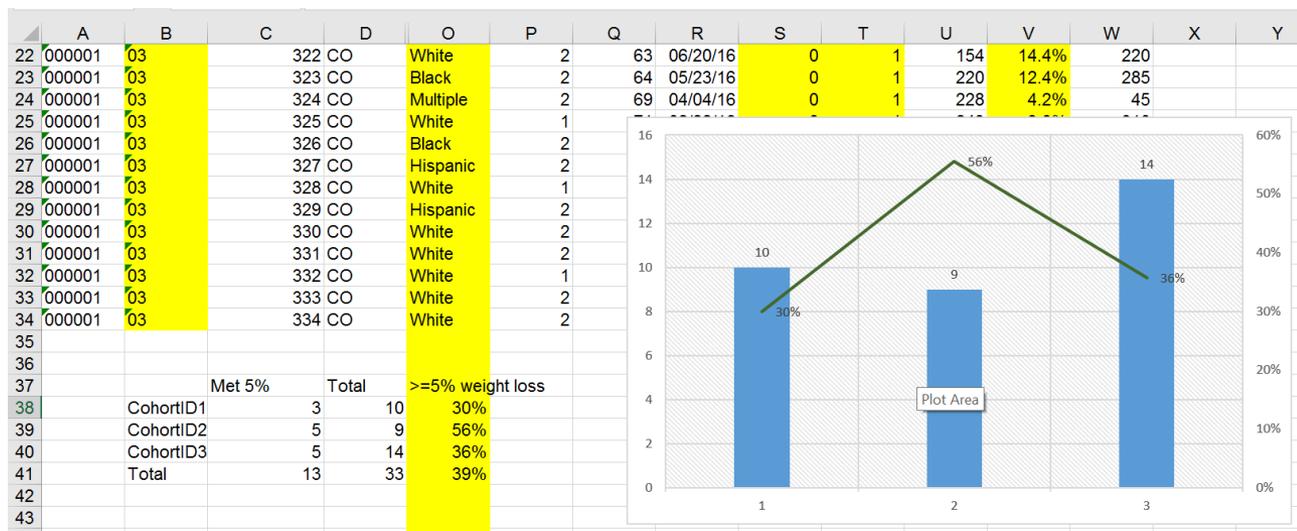
- Diabetes Prevention Data Collection Template
- Coalition Survey

1	A	B	C	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
	ORGCODE	Cohort ID	PARTICIP	ASIAN	BLACK	NRHOPI	WHITE	Race	SEX	HEIGHT	DATE	init_class_flg	last_class_flg	WEIGHT	pct_wt_loss	PAI	HealthInsCo	HealthInsOther	Referral_all	REPROV	ProvName	ClinicName
2	000001	01	101	2	2	2	2	White	2	62	11/01/15	1	0	165	0.0%	nr	99		Provider	1 Fred Flinstone	Great Health Clinic	
3	000001	01	101	2	2	2	2	White	2	62	11/05/15	0	0	165	0.0%	nr						
4	000001	01	101	2	2	2	2	White	2	62	11/15/15	0	0	164	0.0%	nr						
5	000001	01	101	2	2	2	2	White	2	62	11/29/15	0	0	162	1.8%	nr						
6	000001	01	101	2	2	2	2	White	2	62	12/06/15	0	0	160	3.0%	nr						
7	000001	01	101	2	2	2	2	White	2	62	12/13/15	0	0	160	3.0%	nr						
8	000001	01	101	2	2	2	2	White	2	62	12/27/15	0	0	160	3.0%	nr						
9	000001	01	101	2	2	2	2	White	2	62	01/10/16	0	0	159	3.6%	nr						
10	000001	01	101	2	2	2	2	White	2	62	01/17/16	0	0	158	4.2%	nr						
11	000001	01	101	2	2	2	2	White	2	62	01/24/16	0	0	157	4.8%	nr						
12	000001	01	101	2	2	2	2	White	2	62	01/31/16	0	0	158	4.2%	nr						
13	000001	01	101	2	2	2	2	White	2	62	02/14/16	0	0	157	4.8%	nr						
14	000001	01	101	2	2	2	2	White	2	62	02/21/16	0	0	158	4.2%	nr						
15	000001	01	101	2	2	2	2	White	2	62	03/06/16	0	0	155	6.1%	nr						
16	000001	01	101	2	2	2	2	White	2	62	03/20/16	0	0	156	5.5%	nr						
17	000001	01	101	2	2	2	2	White	2	62	04/03/16	0	0	155	6.1%	nr						
18	000001	01	101	2	2	2	2	White	2	62	05/01/16	0	0	154	6.7%	nr						
19	000001	01	101	2	2	2	2	White	2	62	05/29/16	0	0	152	7.9%	nr						
20	000001	01	101	2	2	2	2	White	2	62	07/03/16	0	1	146	11.5%	nr						
21	000001	01	102	2	2	2	2	White	2	66	11/01/15	1	0	139	0.0%	nr	99	Provider	1 Greta Garbo	Spring Valley Clinic		
22	000001	01	102	2	2	2	2	White	2	66	11/08/15	0	0	138	0.7%	nr						
23	000001	01	102	2	2	2	2	White	2	66	11/15/15	0	0	134	3.6%	nr						
24	000001	01	102	2	2	2	2	White	2	66	11/29/15	0	0	132	5.0%	nr						
25	000001	01	102	2	2	2	2	White	2	66	12/06/15	0	0	134	3.6%	nr						
26	000001	01	102	2	2	2	2	White	2	66	12/13/15	0	0	135	2.9%	nr						
27	000001	01	102	2	2	2	2	White	2	66	12/27/15	0	0	131	5.8%	nr						
28	000001	01	102	2	2	2	2	White	2	66	01/10/16	0	0	130	6.5%	nr						
29	000001	01	102	2	2	2	2	White	2	66	01/17/16	0	0	132	5.0%	nr						

1	A	B	C	What is the primary focus area of the coalition work? Check all that apply.									
	Survey Date	CoalitionID	RespondentID	Focus_ActiveLiving	Focus_Asthma	Focus_BuiltEnv	Focus_CancerPrevention	Focus_ClimateEnv	Focus_Nutrition	Focus_HealthEquity	Focus_HealthCommunities	Focus_HealthCo	
2								0=Not Selected					
3	Feb-17 Site1		1	1	0	1	0	0	0	1	1	1	
4	Feb-17 Site1		2	1	0	1	0	1	1	0	1	1	
5	Feb-17 Site1		3	1	0	1	0	0	0	1	1	1	
6	Feb-17 Site1		4	1	0	1	0	0	0	0	1	1	
7	Feb-17 Site1		5	1	0	1	0	0	0	1	1	1	
8	Feb-17 Site1		6	1	0	1	0	0	1	1	1	1	
9	Feb-17 Site1		7	1	0	1	0	1	1	1	1	1	
10	Feb-17 Site1		8	1	0	0	0	0	0	1	1	1	
11	Feb-17 Site1		9	1	0	1	0	0	0	1	1	1	

# Diabetes Prevention Data Collection

- How many people in a cohort (and across cohorts) met 5% weight loss?
- Features of Excel we will use:
  - Countifs
  - Countif
  - Simple formula
  - Combo chart



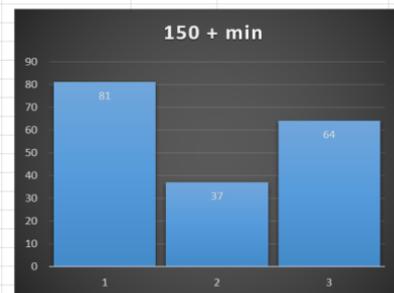
# Diabetes Prevention Data Collection

- How many people in a cohort (and across cohorts) met 150 minutes of physical activity per week?

- Features of Excel we will use:

- If
- Countifs
- Countif
- Simple formula
- AVERAGEIFS
- AVERAGEIF
- AVERAGE (Mean)
- Median
- Mode

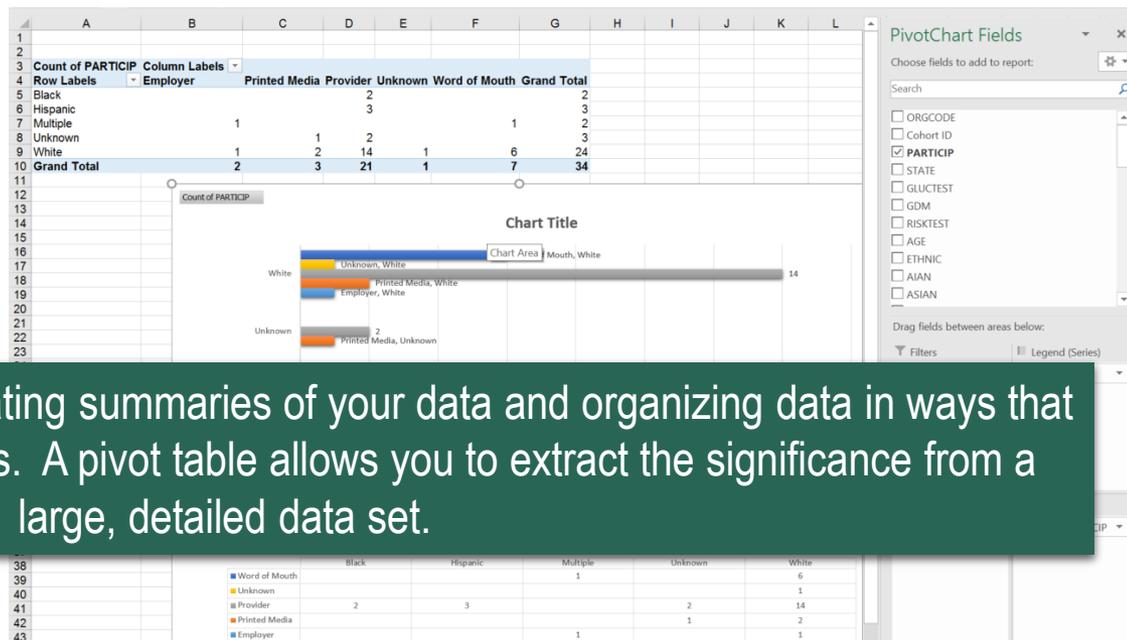
	A	B	C	D	E	F	G	H	I	J	K
1	ORGCODE	Cohort ID	PARTICIP	PA							
2	000001	01	101	0	No						
3	000001	01	101	120	No						
4	000001	01	101	120	No						
5	000001	01	101	120	No						
6	000001	01	101	240	Yes	Cohort	150 + min	All participates	% of overall with 150+	Mean of those 150+ min	Mean for all within Cohort
7	000001	01	101	345	Yes	1	81	167	48.50	304.06	170.89
8	000001	01	101	240	Yes	2	37	97	38.14	274.16	128.35
9	000001	01	101	240	Yes	3	64	184	34.78	266.58	116.30
10	000001	01	101	0	No	Totals	182	448			
11	000001	01	101	180	Yes	Mean Overall	139.26				
12	000001	01	101	300	Yes	Median	120				
13	000001	01	101	0	No	Mode	0				
14	000001	01	101	280	Yes						
15	000001	01	101	200	Yes						
16	000001	01	101	0	No						
17	000001	01	101	150	Yes						
18	000001	01	101	420	Yes						
19	000001	01	101	0	No						
20	000001	01	101	360	Yes						
21	000001	01	102	0	No						
22	000001	01	102	0	No						
23	000001	01	102	335	Yes						
24	000001	01	102	240	Yes						
25	000001	01	102	195	Yes						
26	000001	01	102	480	Yes						
27	000001	01	102	325	Yes						
28	000001	01	102	325	Yes						
29	000001	01	102	358	Yes						
30	000001	01	102	590	Yes						



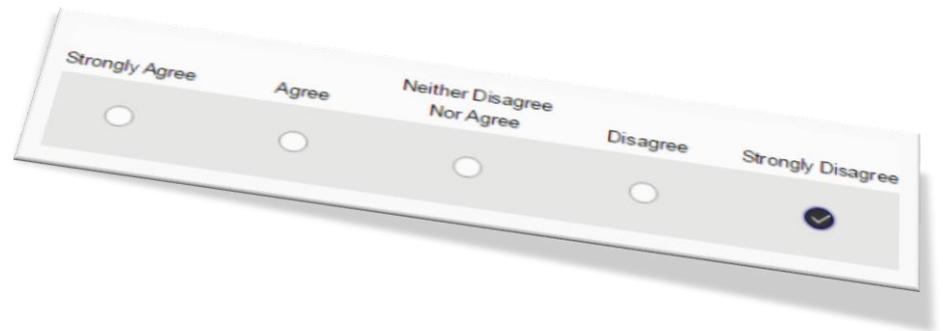
# Diabetes Prevention Data Collection

- Is there a relationship between referral source and ethnicity/race?
- Features of Excel we will use:

- Pivot Table
- Pivot Chart



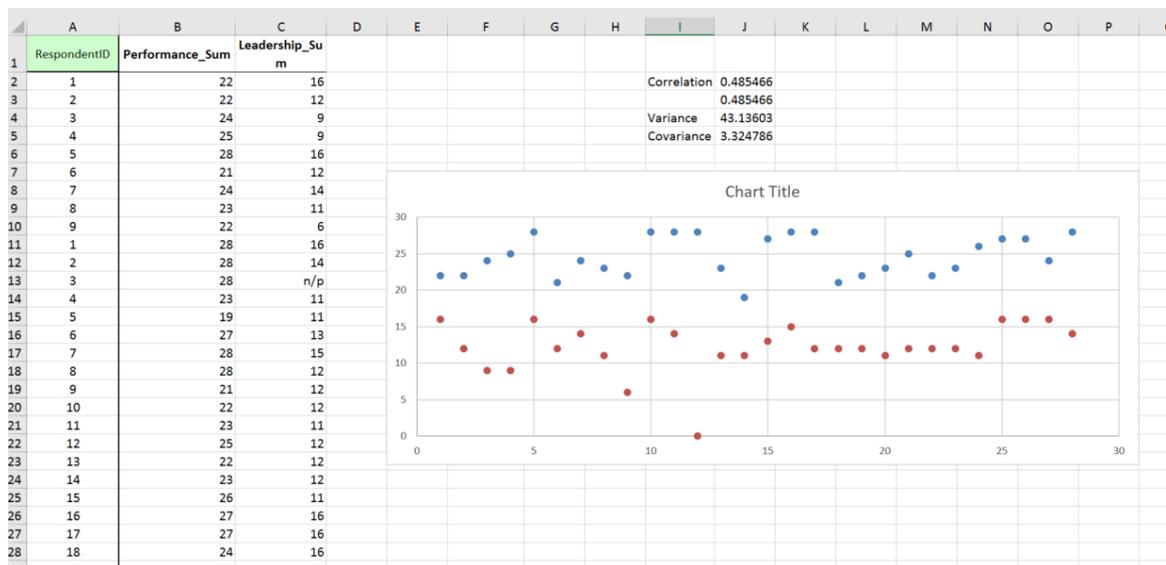
# Coalition Survey



- Is there a relationship between length of time active in the community and overall rating on coalition performance?
  - We are trying to find a relationship between one of the descriptive questions and a subscale.
- Features of Excel we will use:
  - Pivot Table
  - Pivot Chart

# Coalition Survey

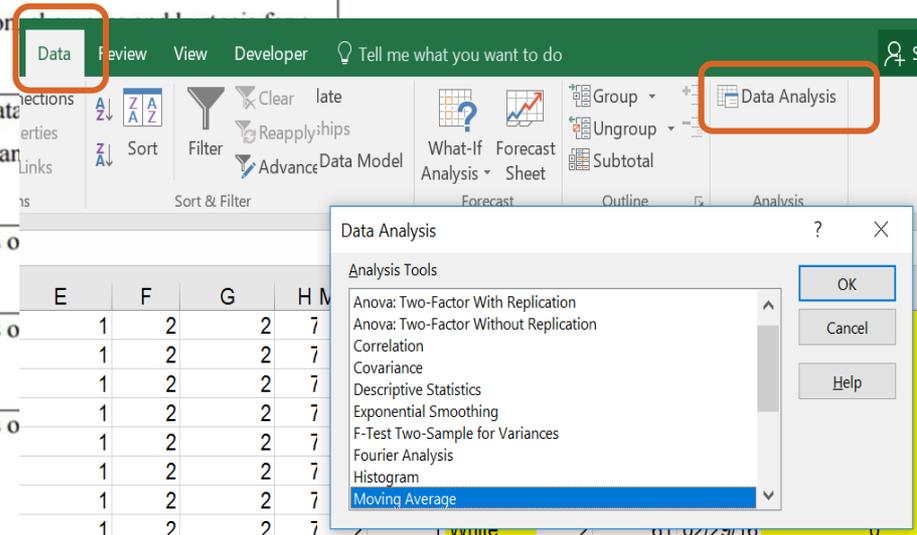
- Is there a relationship between coalition performance and coalition leadership?
  - We will try to find a relationship between scales.
- Features of Excel we will use:
  - Scatter Chart
  - Also look at:
    - Correlation
    - Variance
    - Covariance



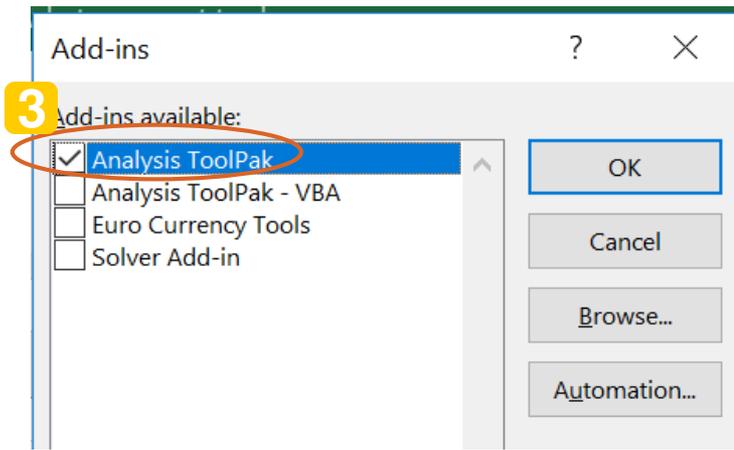
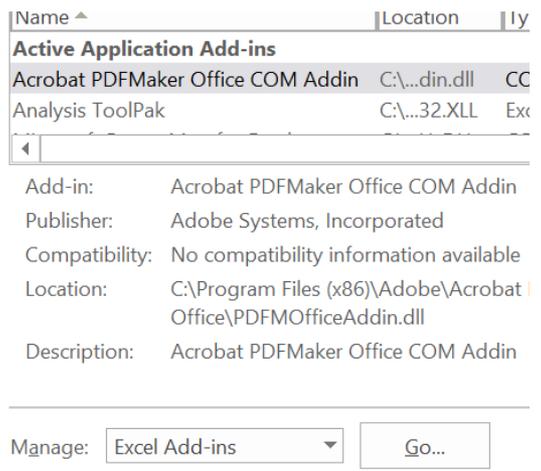
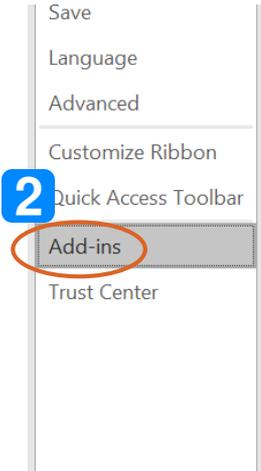
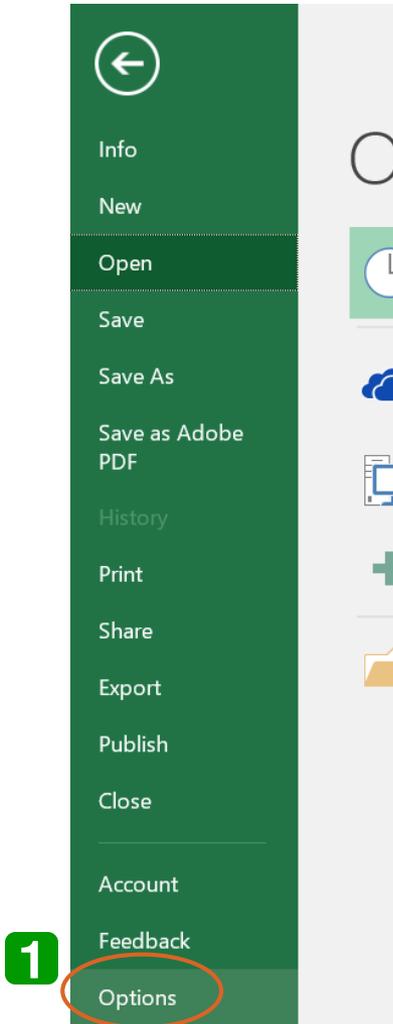
# Data Analysis ToolPak

- The Data Analysis ToolPak is an Excel add-in. It contains more extensive functions, including some useful inferential statistical tests.

Function name	Description
Anova: Single Factor	Performs a one-way analysis of variance (ANOVA)
Correlation	Creates a correlation matrix showing the Pearson correlation coefficient ( $r$ ) for each pair of variables of $N$ cases selected
Descriptive statistics	Calculates a range of univariate descriptive statistics, including measures of central tendency, dispersion, and shape for a single variable
Histogram	Generates a histogram for a range of data and a table of the data on which the histogram is based (includes Pareto analysis)
$t$ -Test: Paired Two-Sample for Means	Performs a $t$ -test to compare the means of two samples
$t$ -Test: Two-Sample Assuming Equal Variances	Performs a $t$ -test to compare the means of two samples assuming equal variances
$t$ -Test: Two-Sample Assuming Unequal Variances	Performs a $t$ -test to compare the means of two samples assuming unequal variances



# How to Add the Data Analysis ToolPak



# Excel for Mac

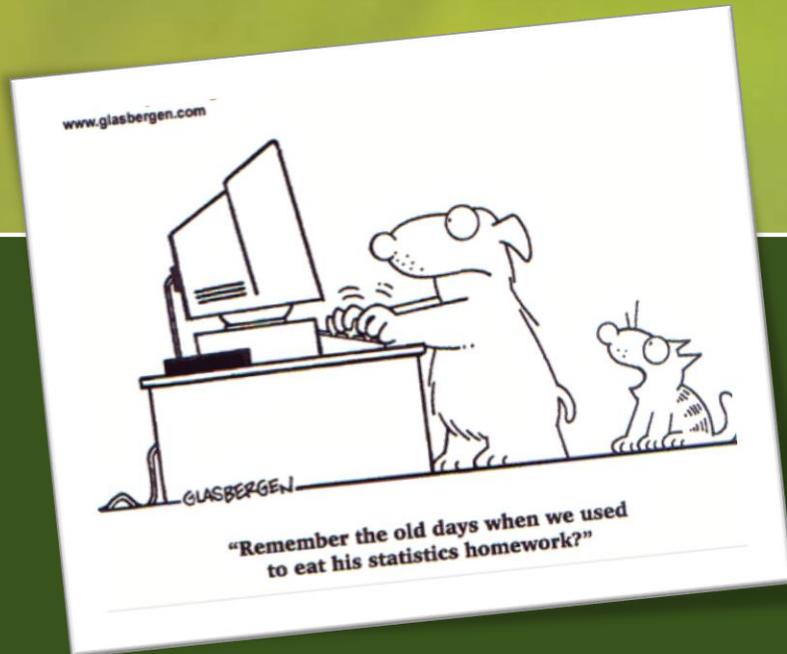
- An important addition in Excel 2016 for Mac users – the Analysis Tool Pak is available.
- If you are using a version of Excel prior to 2016, the data analysis ToolPak add-in is not available within Excel. An alternative is a free downloadable add-in called StatPlus:mac LE.

# Resources

- Statistical functions (reference)
  - <https://support.office.com/en-us/article/Statistical-functions-reference-624dac86-a375-4435-bc25-76d659719ffd>
- StatPlus:mac Le
  - <https://www.analystsoft.com/en/products/statplusmacle/>
- Need a brush up on Excel. Two sites I really like (both free):
  - <http://www.gcflearnfree.org/topics/office2013/>
  - <https://support.office.com/en-gb/article/Excel-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb?ui=en-US&rs=en-GB&ad=GB>
- Statistics How To (You Tube videos)
  - [https://www.youtube.com/channel/UCs3lhN8VOA\\_5WxpAgbSmFkg](https://www.youtube.com/channel/UCs3lhN8VOA_5WxpAgbSmFkg)



# Questions?



Have an  
**Excellent Day!**  
The Evaluation Team



Please take the two question survey that will appear in a pop-up window when you exit from the webinar

# Evaluation Team

- ❑ Cheryl Kelly (Evaluation Investigator)
- ❑ Marisa Allen
- ❑ Bre Barela
- ❑ Lisa Harner
- ❑ Denise Hartsock
- ❑ Bonnie Leeman-Castillo
- ❑ Carmen Luna
- ❑ Erica Morse