

Dating Data!

INTRODUCTION TO STATISTICS IN SPREADSHEETS



Ted Kwartler
Data Dude

Mean Age

<i>fx</i>	=AVERAGE		
	AVERAGE		
	Numerical average value in a dataset, ignoring text.		
1			
2	AVERAGEA		
3	AVERAGEIF		
4	AVERAGEIFS		
5	AVERAGE.WEIGHTED		
6	3	47	
7	4	48	
8	5	48	
9	9	55	
10	8	76	
11	10	96	
12			



Maximum & Minimum Age

Maximum Value Spreadsheet Formula

```
MAX(...)
```

Minimum Value Spreadsheet Formula

```
MIN(...)
```



Social behavior

Counting when

```
COUNTIF(A1:A10, <3)
```

Quartiles

```
QUARTILE(A1:B10, 4)
```

Correlation

```
CORREL(A1:A10, B1:B10)
```

Let's practice!

INTRODUCTION TO STATISTICS IN SPREADSHEETS

Visuals & Distributions

INTRODUCTION TO STATISTICS IN SPREADSHEETS

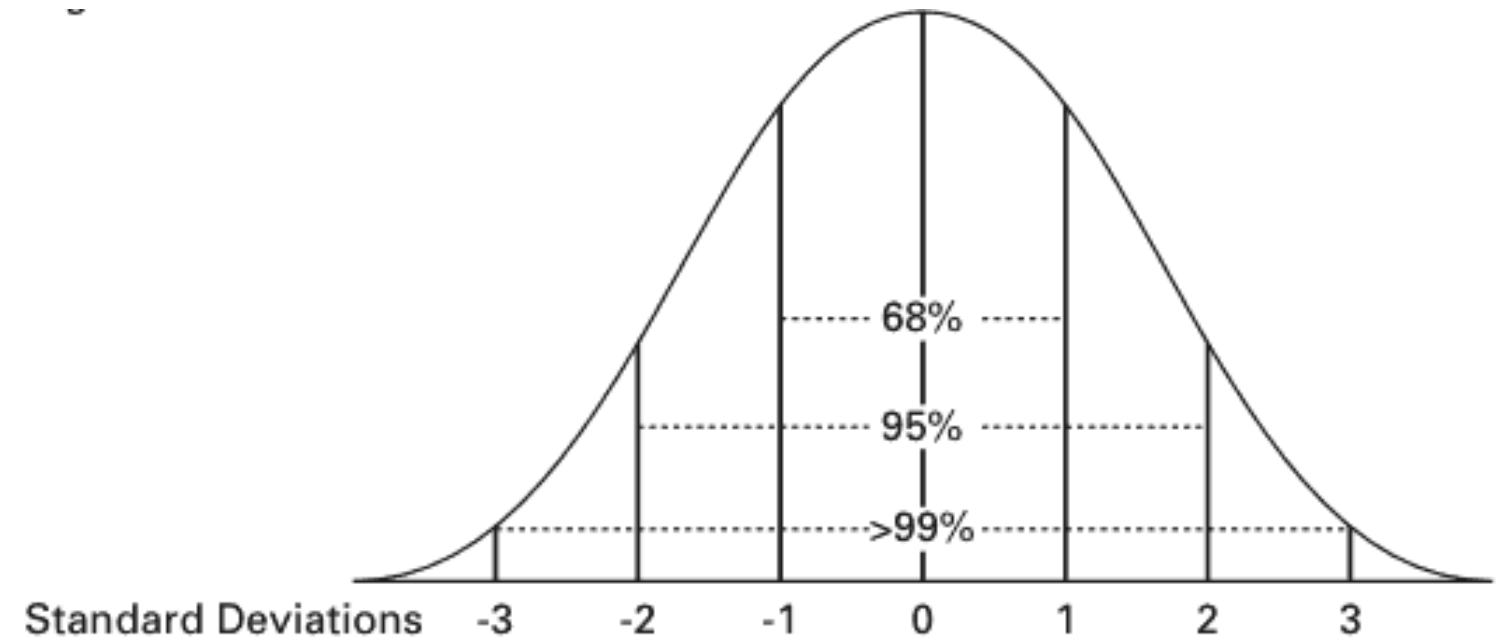


Ted Kwartler
Data Dude

Distribution Stats

Variance

- *how far the data is spread out*
- ** Standard Deviation
- *unit of dispersion relative to the mean*
- *square root of variance*
- ** Z-Score
- *the number of standard deviations a point is from mean*

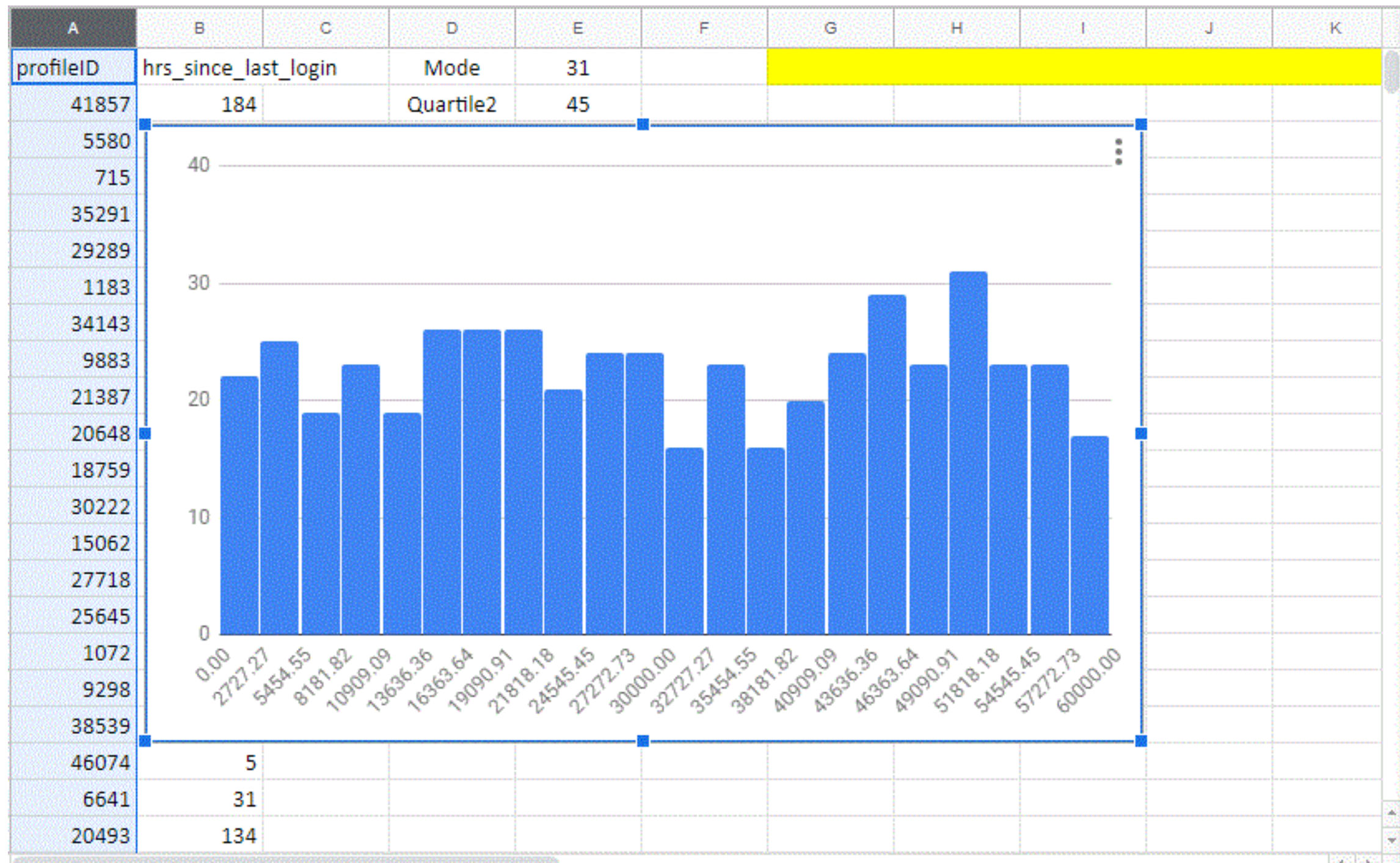


Visualizing Distributions

Histogram Considerations

- Determine an appropriate number of buckets of bins
- If using sampled data
 - Make sure data is randomly sampled
 - *at least* 20+ observations in the sample





Setup
Customize

Chart type

Histogram chart

Data range

A1:A1000

X-AXIS

Add X-axis

SERIES

123 profileID

Add Series

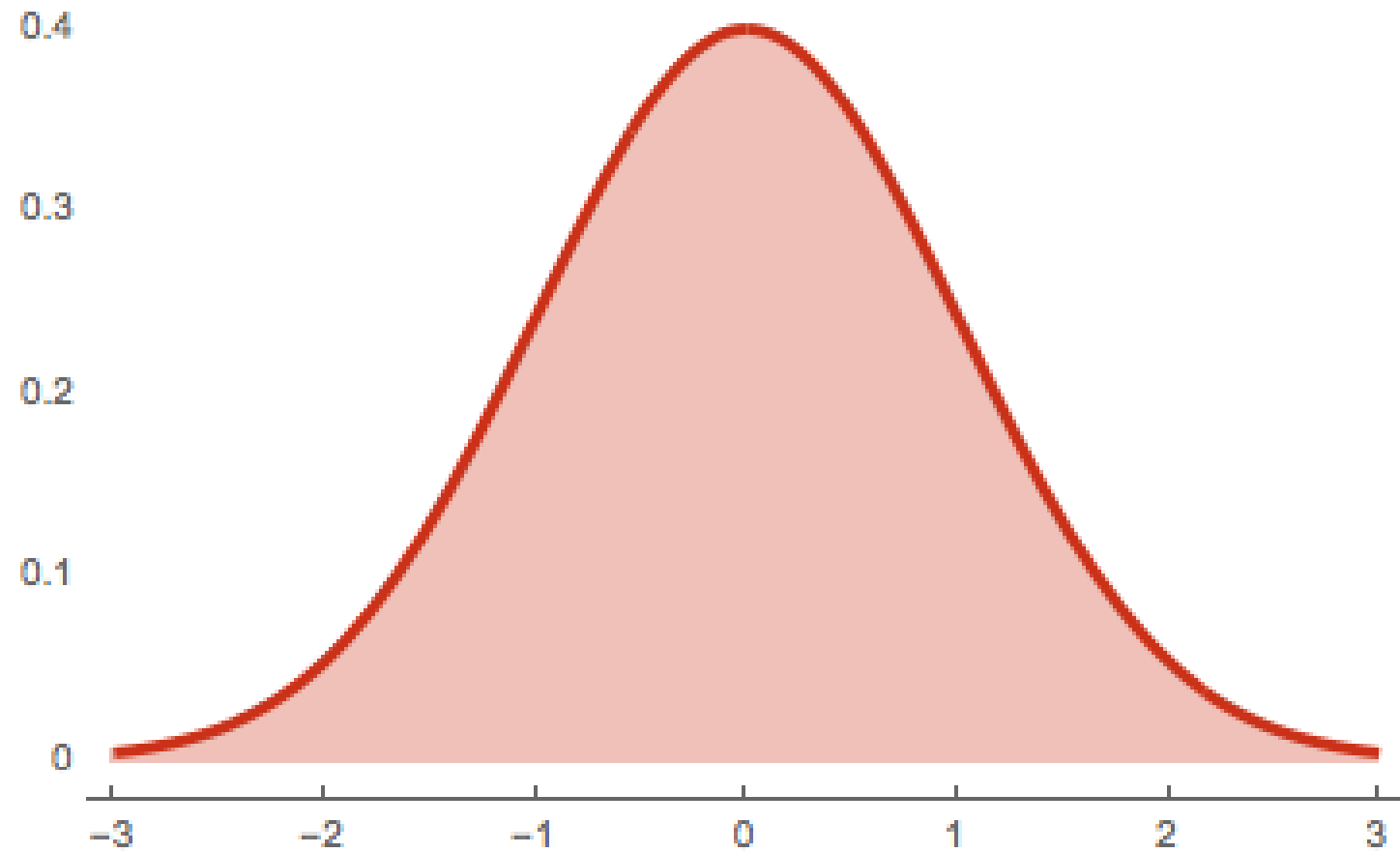
SORT

☐ Switch rows / columns

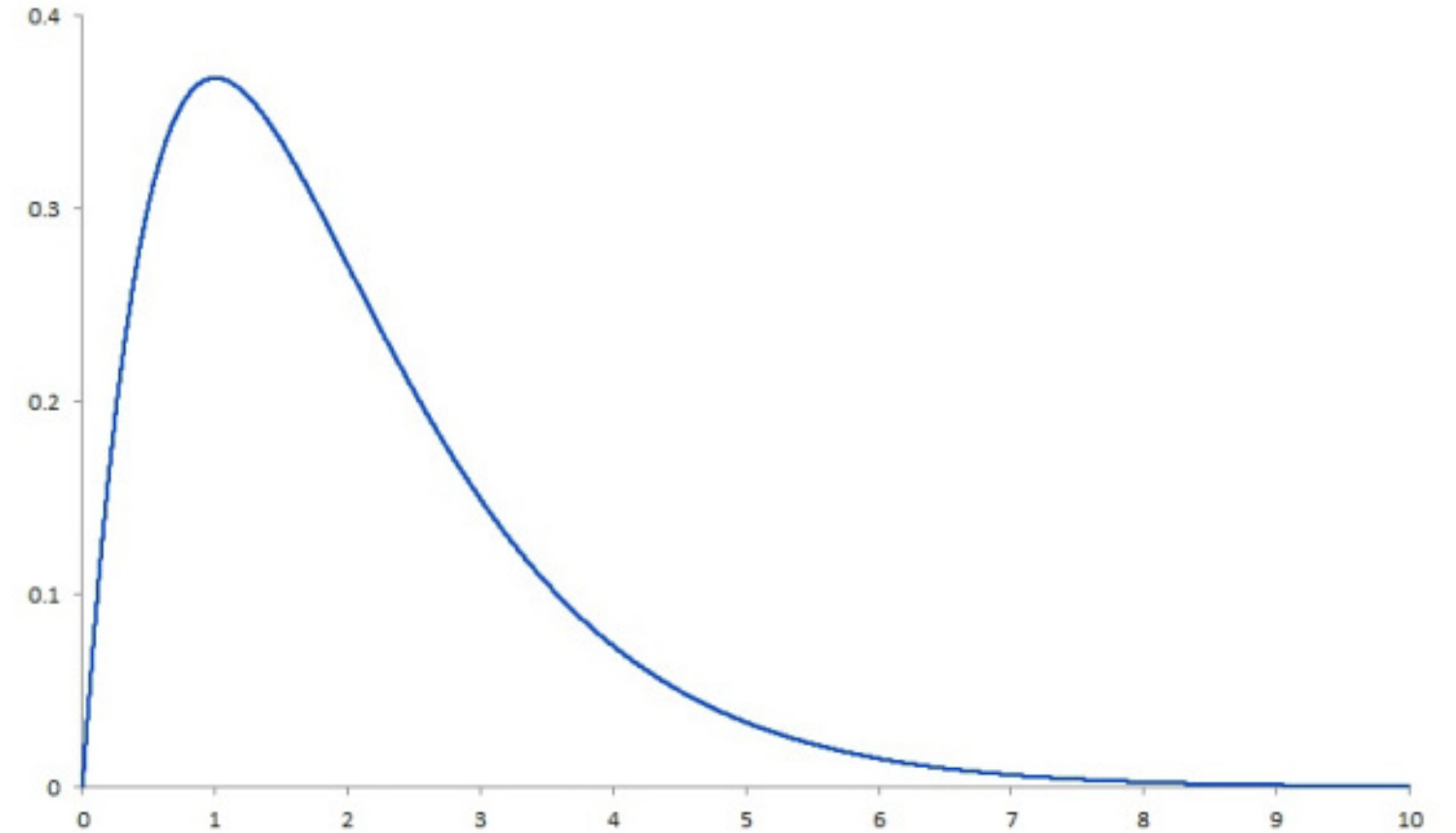
☐ Use row 1 as headers

Comparing Distributions

Normal Distribution



Skewed Distribution



Normal Stats

Skewness

- A measure of symmetry in the distribution
- Look for values between 2 & -2

```
SKEW(A1:A10)
```

Kurtosis

- Is the peak near the mean of distribution?
- Look for values between 2 & -2

```
KURT(A1:A10)
```

Let's practice!

INTRODUCTION TO STATISTICS IN SPREADSHEETS

Tipping the scale to positive correlation

INTRODUCTION TO STATISTICS IN SPREADSHEETS



Ted Kwartler
Data Dude

Correlation tips from one to negative one



Correlation = 1

=- Positive: as one var so does the other #

Correlation = 0

- No relationship among variables

Correlation = -1

- Negative or opposite relationship

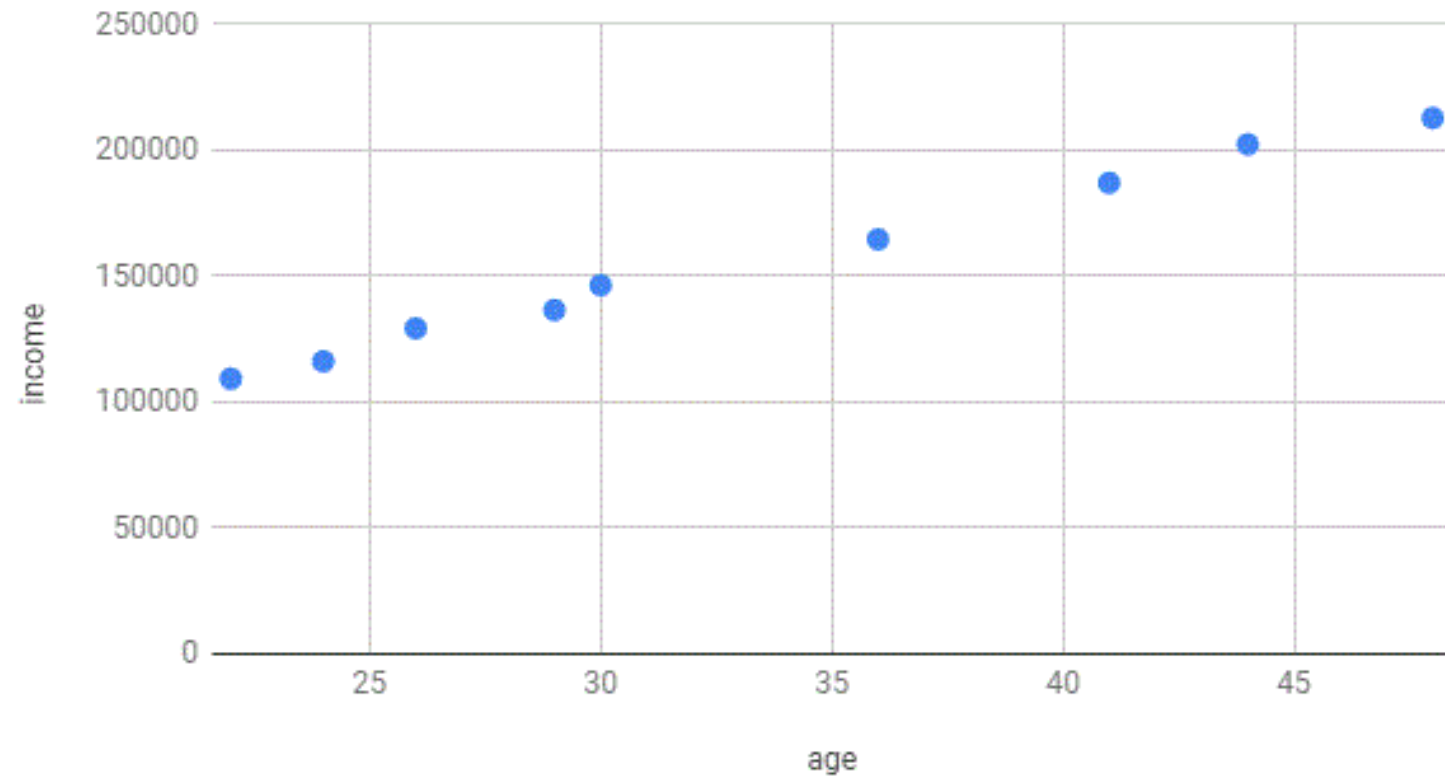
Lifeguards for Ice Cream



Captain Obvious:

Ice cream DOES NOT cause drowning!

Correlation Visualized



Correlation

0.99

Let's practice!

INTRODUCTION TO STATISTICS IN SPREADSHEETS

More complex statistical (& profile) relationships

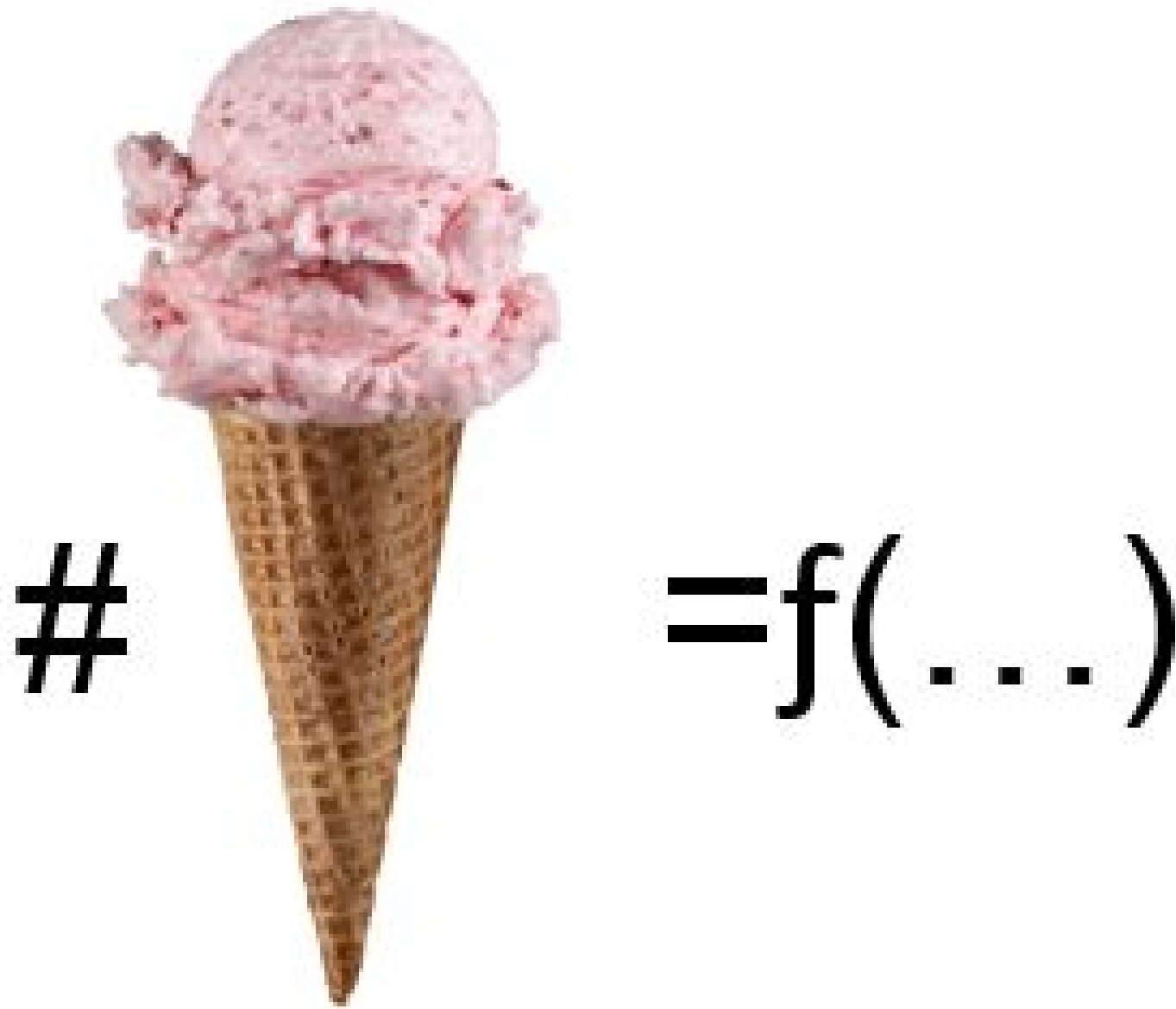
INTRODUCTION TO STATISTICS IN SPREADSHEETS



Ted Kwartler
Data Dude



Predicting Ice Cream Cones



Temperature as an input



=f(temperature)

Cones as a function of temp, price & saturday



#

=f(temperature + price + saturday)



#

= β + β *temperature + β *day + β *price + error

Multiple Regression in sheets

Pseudo Formula

```
LINEST(y-Variables, x-Variables)
```

Example

```
LINEST(A1:A10, B1:D10)
```

Deconstructing the relationship

1	x-Temp	x-Price	y-Cones
2	93	2.4	800
3	90	2.6	650
...
499	89	2.4	470
500	92	2.4	730

```
LINEST(C2:C500,A2:B500)
```


The tricky sheets results

```
LINEST(D2:D500,A2:C500)
```

x-Temp	x-Price	INTERCEPT
0.25	-0.5	7

```
LINEST(D2:D5,A2:C5)
```

x-Temp	x-Price	x-Sat.	INTERCEPT
0.25	-0.5	3	7

Equation

```
Cones = (0.25 * temp) + (-0.5 * price) + 7
```

Example Day

```
Cones = (0.25 * 100) + (-0.5 * $2.40) + 7
```

```
Cones = 25 + (-1.2) + 7
```

```
Cones = 30.8
```

Count Pivot Table

Height	Blue	Brown	Total
Tall	20	25	45
Short	35	20	55
Total	55	45	100

Expected Frequency Table

Height	Blue	Brown	Total
Tall	<code>(totalBlue * totalTall) / (grandTotal)</code>	<code>(45 * 45)/100</code>	45
Short	30.25	??	55

Chi-Squared Test

Pivot

Height	Blue	Brown	Total
Tall	20	25	45
Short	35	20	55
Total	55	45	100

Expected Frequency

Height	Blue	Brown	Total
Tall	24.75	20.25	45
Short	30.25	24.75	55

```
CHISQ.TEST(OriginalCountData,  
ExpectedFrequencyData)
```

Acceptance or rejection for dating profiles

```
CHISQ.TEST(OriginalCountData, ExpectedFrequencyData)
```

```
Result = 0.055
```

Let's practice!

INTRODUCTION TO STATISTICS IN SPREADSHEETS

Congratulations!

INTRODUCTION TO STATISTICS IN SPREADSHEETS



Ted Kwartler
Data Dude

Congratulations!

INTRODUCTION TO STATISTICS IN SPREADSHEETS