

Defining error and uncertainty

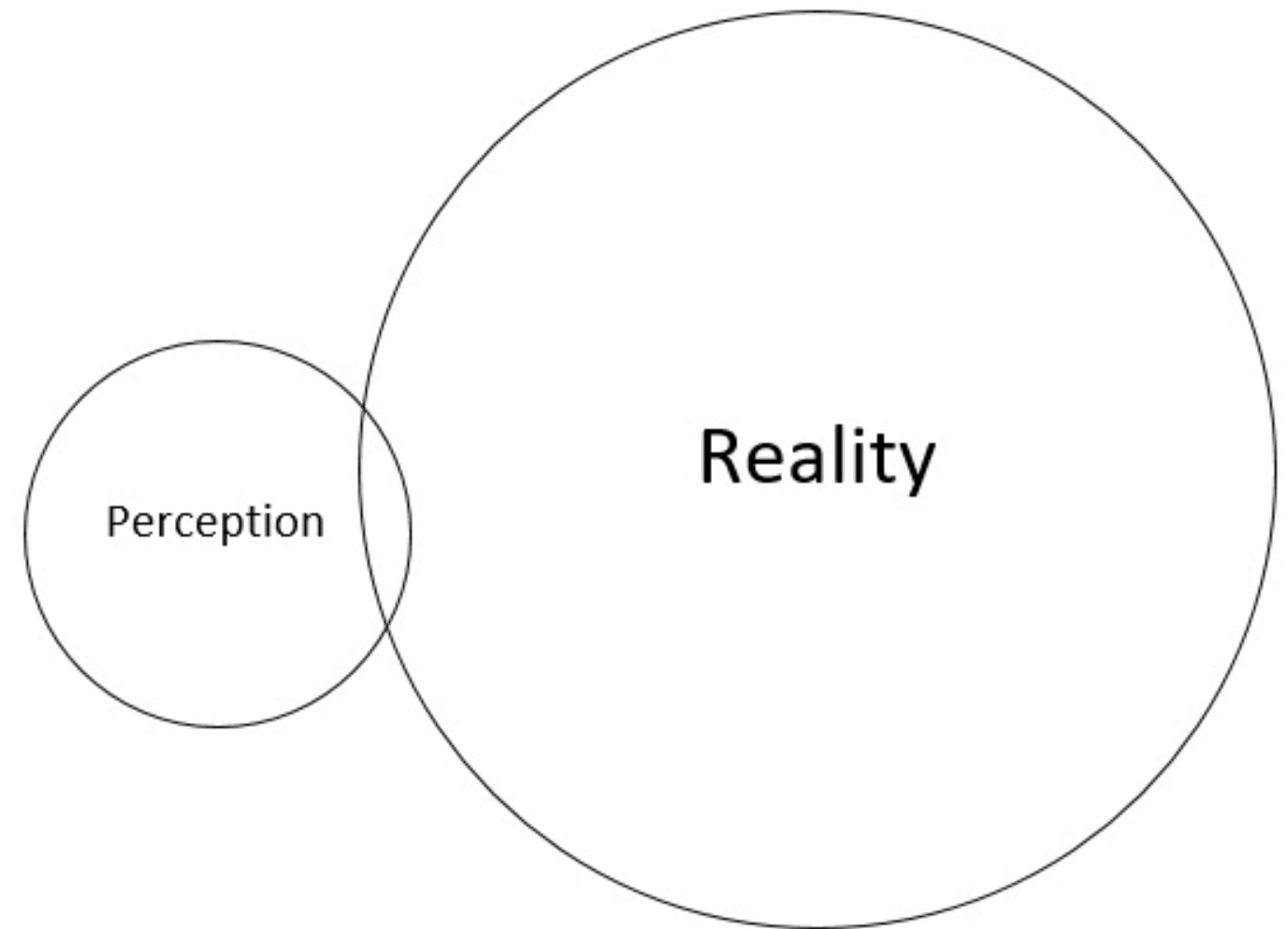
ERROR AND UNCERTAINTY IN SPREADSHEETS



Evan Kramer
Instructor

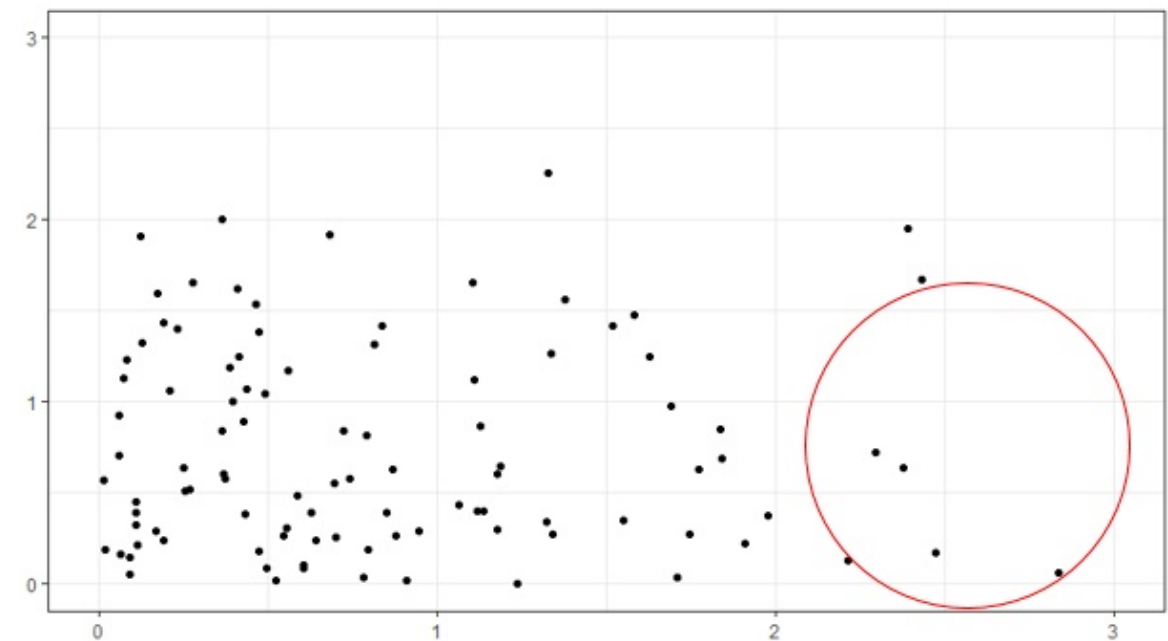
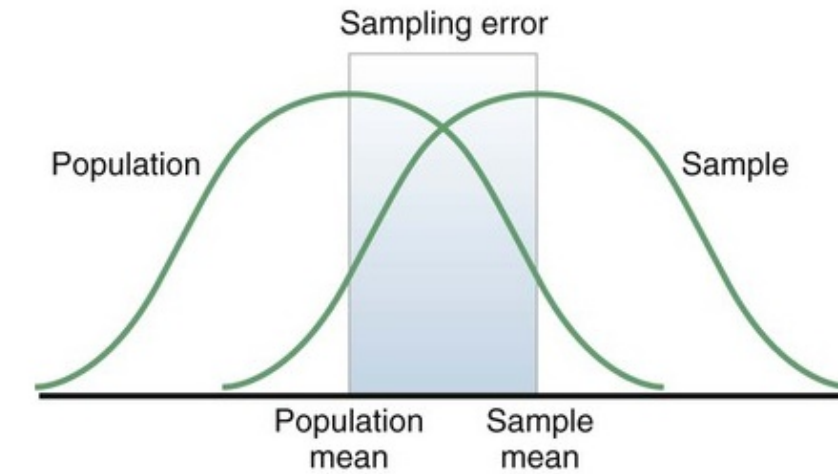
We make errors

- Errors in judgment
 - How or why things happened
 - Biases, assumptions, psychology
- Errors in statistics
 - Measures of variation or uncertainty
 - Incorrect conclusions about the data



Why care about error?

- Example
 - New cologne launch
- Variation
 - Due to chance
 - "Statistical significance"
- Sampling
 - Representative of population



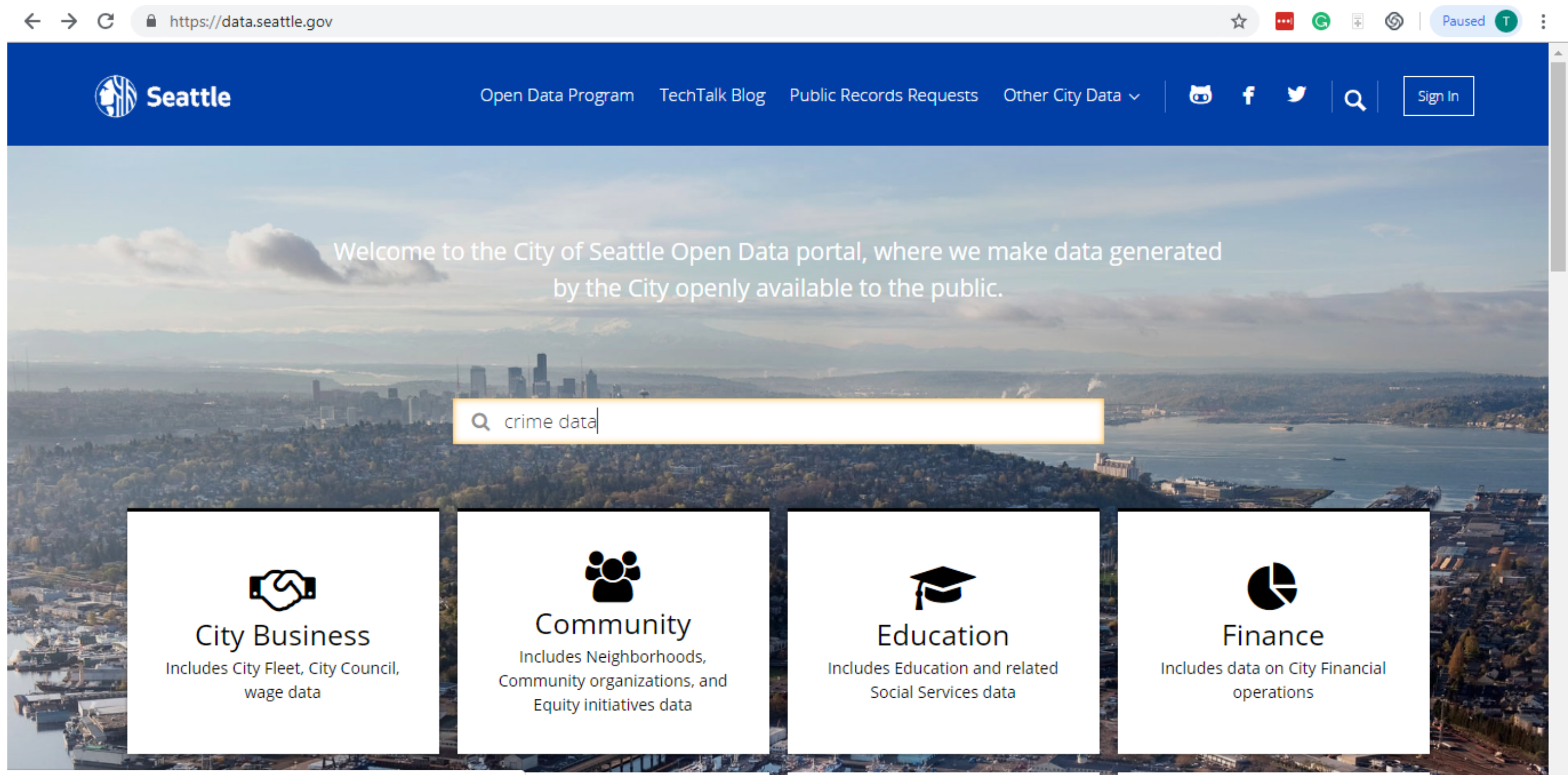
¹ Nurse Key. 2017. "Chapter 15: Sampling." February 17. Accessed Feb. 4, 2019 <https://nursekey.com/sampling/>

Types of statistical errors

- **Type I error**
 - "False positive"
 - Sampling
 - Data quality
 - Design
- **Type II error**
 - "False negative"
 - Sampling
 - Variation
 - Measurement

Measures of central tendency

- `AVERAGE()`
 - Mean
- `MEDIAN()`
 - Midpoint
- `MODE()`
 - Most common value
- `STDEV()`
 - Average distance from the average



¹ <https://data.seattle.gov>

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

IF functions

ERROR AND UNCERTAINTY IN SPREADSHEETS



Evan Kramer
Instructor

Grouping functions

- `UNIQUE()` function
 - Returns distinct values
- `=UNIQUE(C:C)`
 - Returns list of precincts

fx | =unique(C:C)

	A	B	C	D	E	F	G
1	Date Occurred	Time Occurred	Precinct	Offense			Precinct
2	1/1/2008	0	WEST	THEFT-CARPROWL			WEST
3	1/1/2008	0	SOUTH	CHILD-OTHER			SOUTH
4	1/1/2008	0	UNKNOWN	SEXOFF-OTHER			UNKNOWN
5	1/1/2008	0	SOUTHWEST	CHILD-ABUSED-NOFORCE			SOUTHWEST
6	1/1/2008	0	NORTH	THEFT-OTH			NORTH
7	1/1/2008	0	SOUTH	CHILD-OTHER			EAST
8	1/1/2008	0	NORTH	BURGLARY-NOFORCE-RES			
9	1/2/2008	2352	EAST	THEFT-OTH			
10	1/2/2008	2347	SOUTHWEST	NARC-POSSESS-OTHER			
11	1/2/2008	2328	WEST	NARC-SELL-COCAINE			
12	1/2/2008	2307	WEST	NARC-POSSESS-COCAINE			
13	1/2/2008	2228	SOUTHWEST	DUI-LIQUOR			
14	1/2/2008	2200	EAST	THEFT-CARPROWL			
15	1/2/2008	2145	NORTH	THEFT-CARPROWL			
16	1/2/2008	2130	SOUTHWEST	THEFT-CARPROWL			
17	1/2/2008	2100	NORTH	THEFT-CARPROWL			
18	1/2/2008	2057	NORTH	TRESPASS			
19	1/2/2008	2000	WEST	BURGLARY-NOFORCE-RES			
20	1/2/2008	1900	SOUTHWEST	THEFT-CARPROWL			

Counting functions

- Counts of unique values
 - `COUNTIF()` function
 - range
 - criterion
- `=COUNTIF(C:C, G2)`
- `=COUNTIF(B:B, ">2000")`

	A	B	C	D	E	F	G
1	Date Occurred	Time Occurred	Precinct	Offense		Precinct	# of Offenses
2	1/1/2008	0	WEST	THEFT-CARPROWL		WEST	165
3	1/1/2008	0	SOUTH	CHILD-OTHER		SOUTH	66
4	1/1/2008	0	UNKNOWN	SEXOFF-OTHER		UNKNOWN	5
5	1/1/2008	0	SOUTHWEST	CHILD-ABUSED-NOFORCE		SOUTHWEST	45
6	1/1/2008	0	NORTH	THEFT-OTH		NORTH	157
7	1/1/2008	0	SOUTH	CHILD-OTHER		EAST	62
8	1/1/2008	0	NORTH	BURGLARY-NOFORCE-RES			
9	1/2/2008	2352	EAST	THEFT-OTH			
10	1/2/2008	2347	SOUTHWEST	NARC-POSSESS-OTHER			
11	1/2/2008	2328	WEST	NARC-SELL-COCAINE			
12	1/2/2008	2307	WEST	NARC-POSSESS-COCAINE			
13	1/2/2008	2228	SOUTHWEST	DUI-LIQUOR			
14	1/2/2008	2200	EAST	THEFT-CARPROWL			
15	1/2/2008	2145	NORTH	THEFT-CARPROWL			
16	1/2/2008	2130	SOUTHWEST	THEFT-CARPROWL			
17	1/2/2008	2100	NORTH	THEFT-CARPROWL			
18	1/2/2008	2057	NORTH	TRESPASS			
19	1/2/2008	2000	WEST	BURGLARY-NOFORCE-RES			
20	1/2/2008	1900	SOUTHWEST	THEFT-CARPROWL			
21	1/2/2008	1830	WEST	CHILD-OTHER			
22	1/2/2008	1826	NORTH	THEFT-SHOPLIFT			
23	1/2/2008	1800	NORTH	THEFT-CARPROWL			
24	1/2/2008	1634	SOUTH	NARC-POSSESS-MARIJU			
25	1/2/2008	1615	WEST	THEFT-SHOPLIFT			
26	1/2/2008	1544	NORTH	ROBBERY-STREET-WEAPON			

Other IF functions

- `AVERAGEIF()`
 - Criteria range
 - Criterion
 - Range
- `=AVERAGEIF(C:C, G2, B2:B)`
 - `C:C` : Column containing precincts
 - `G2` : Name of precinct of interest
- `SUMIF()`
 - Criteria range
 - Criterion
 - Range
- `=SUMIF(F2:F7, "WEST", G2:G7)`

Multiple IF conditions

- `COUNTIFS()` function
 - Pairs of ranges and criteria
 - Allows multiple criteria
 - Different criteria in same range
- **Example:** Crimes in West precinct after 8:00
 - `=COUNTIFS(C:C, G2, B:B, ">2000")`

	A	B	C	D	E	F	G	H	I
1	Date Occurred	Time Occurred	Precinct	Offense		Precinct	# of Offenses	Avg. Time of Offense	# Evening Offenses
2	1/1/2008	0	WEST	THEFT-CARPROWL		WEST	165	1508	32
3	1/1/2008	0	SOUTH	CHILD-OTHER		SOUTH	66	1135	5
4	1/1/2008	0	UNKNOWN	SEXOFF-OTHER		UNKNOWN	5	1156	2
5	1/1/2008	0	SOUTHWEST	CHILD-ABUSED-NOFORCE		SOUTHWEST	45	1348	10
6	1/1/2008	0	NORTH	THEFT-OTH		NORTH	157	1426	39
7	1/1/2008	0	SOUTH	CHILD-OTHER		EAST	62	1284	11
8	1/1/2008	0	NORTH	BURGLARY-NOFORCE-RES					
9	1/2/2008	2352	EAST	THEFT-OTH					
10	1/2/2008	2347	SOUTHWEST	NARC-POSSESS-OTHER					
11	1/2/2008	2328	WEST	NARC-SELL-COCAINE					
12	1/2/2008	2307	WEST	NARC-POSSESS-COCAINE					
13	1/2/2008	2228	SOUTHWEST	DUI-LIQUOR					
14	1/2/2008	2200	EAST	THEFT-CARPROWL					
15	1/2/2008	2145	NORTH	THEFT-CARPROWL					
16	1/2/2008	2130	SOUTHWEST	THEFT-CARPROWL					
17	1/2/2008	2100	NORTH	THEFT-CARPROWL					
18	1/2/2008	2057	NORTH	TRESPASS					
19	1/2/2008	2000	WEST	BURGLARY-NOFORCE-RES					
20	1/2/2008	1900	SOUTHWEST	THEFT-CARPROWL					
21	1/2/2008	1830	WEST	CHILD-OTHER					
22	1/2/2008	1826	NORTH	THEFT-SHOPLIFT					
23	1/2/2008	1800	NORTH	THEFT-CARPROWL					
24	1/2/2008	1634	SOUTH	NARC-POSSESS-MARIJU					
25	1/2/2008	1615	WEST	THEFT-SHOPLIFT					
26	1/2/2008	1544	NORTH	ROBBERY-STREET-WEAPON					
27	1/2/2008	1445	WEST	THEFT-CARPROWL					
28	1/2/2008	1430	WEST	THEFT-CARPROWL					
29	1/2/2008	1410	EAST	BURGLARY-NOFORCE-NONRES					
30	1/2/2008	1351	NORTH	THEFT-BUILDING					
31	1/2/2008	1340	SOUTH	THEFT-OTH					
32	1/2/2008	1250	EAST	BURGLARY-NOFORCE-NONRES					

Multiple IF conditions

- **AVERAGEIFS()** function
 - Range
 - Pairs of ranges and criteria
 - Allows multiple criteria

	B	C	D	E	F	G	H	I
1	Time Occurred	Precinct	Offense		Precinct	# of Offenses	Avg. Time of Offense	Avg. Time of Morning Offenses
2	0	WEST	THEFT-CARPROWL		WEST	165	1508	666.3255814
3	0	SOUTH	CHILD-OTHER		SOUTH	66	1135	542.5483871
4	0	UNKNOWN	SEXOFF-OTHER		UNKNOWN	5	1156	7.5
5	0	SOUTHWEST	CHILD-ABUSED-NOFORCE		SOUTHWEST	45	1348	437.1875
6	0	NORTH	THEFT-OTH		NORTH	157	1426	446.0681818
7	0	SOUTH	CHILD-OTHER		EAST	62	1284	519.625
8	0	NORTH	BURGLARY-NOFORCE-RES					
9	2352	EAST	THEFT-OTH					
10	2347	SOUTHWEST	NARC-POSSESS-OTHER					
11	2328	WEST	NARC-SELL-COCAINE					
12	2307	WEST	NARC-POSSESS-COCAINE					
13	2228	SOUTHWEST	DUI-LIQUOR					
14	2200	EAST	THEFT-CARPROWL					
15	2145	NORTH	THEFT-CARPROWL					
16	2130	SOUTHWEST	THEFT-CARPROWL					
17	2100	NORTH	THEFT-CARPROWL					
18	2057	NORTH	TRESPASS					
19	2000	WEST	BURGLARY-NOFORCE-RES					
20	1900	SOUTHWEST	THEFT-CARPROWL					
21	1830	WEST	CHILD-OTHER					
22	1826	NORTH	THEFT-SHOPLIFT					
23	1800	NORTH	THEFT-CARPROWL					
24	1634	SOUTH	NARC-POSSESS-MARIJU					
25	1615	WEST	THEFT-SHOPLIFT					
26	1544	NORTH	ROBBERY-STREET-WEAPON					
27	1445	WEST	THEFT-CARPROWL					
28	1430	WEST	THEFT-CARPROWL					
29	1410	EAST	BURGLARY-NOFORCE-NONRES					
30	1351	NORTH	THEFT-BUILDING					
31	1340	SOUTH	THEFT-OTH					

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

Correlation

ERROR AND UNCERTAINTY IN SPREADSHEETS



Evan Kramer
Instructor

Correlation

- Relationship between two groups
- Correlation coefficient (r)
 - Varies between -1 and 1
 - Positive: Increase/decrease together
 - Negative: Opposite increases/decreases
 - Zero: No relationship

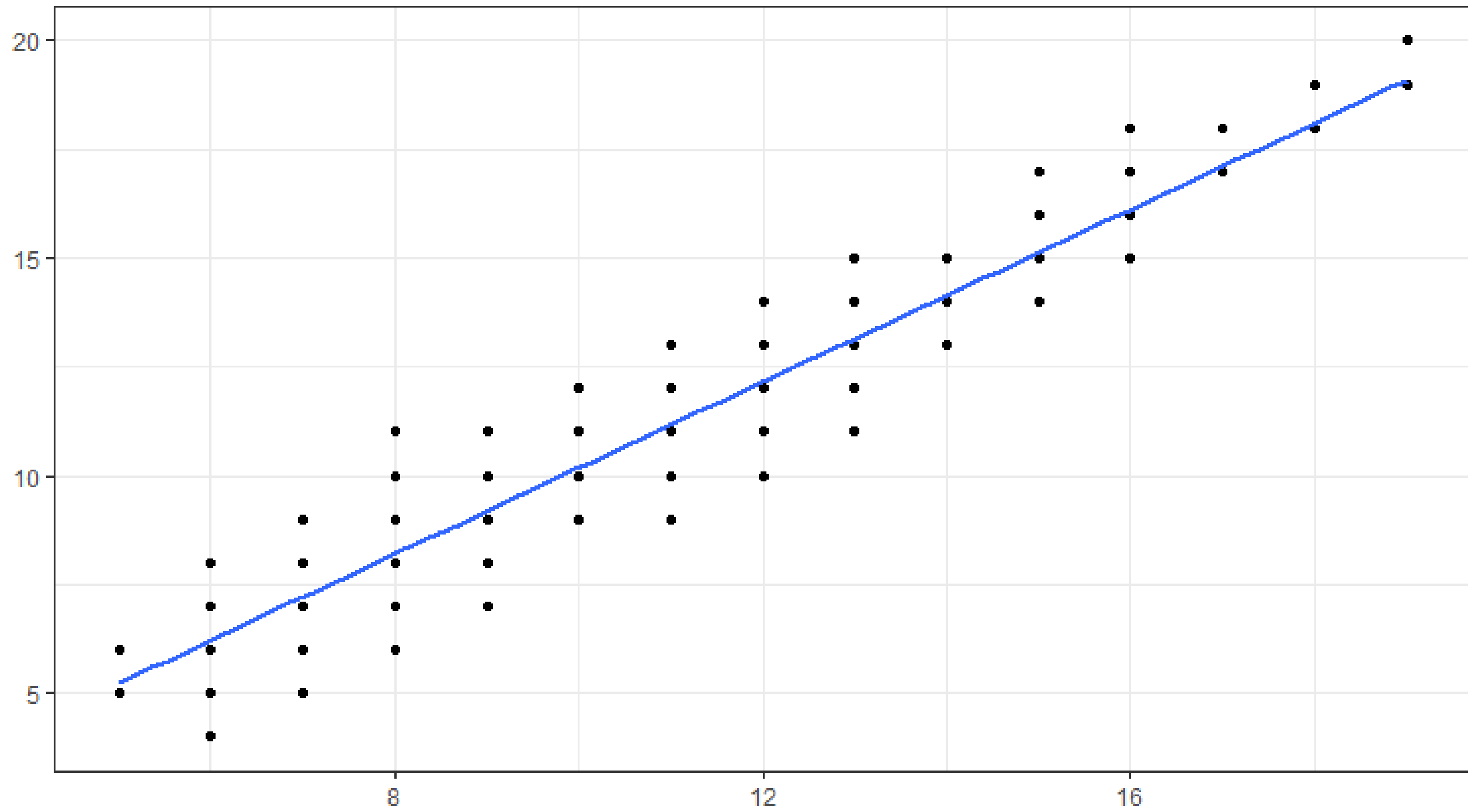
The CORREL() function

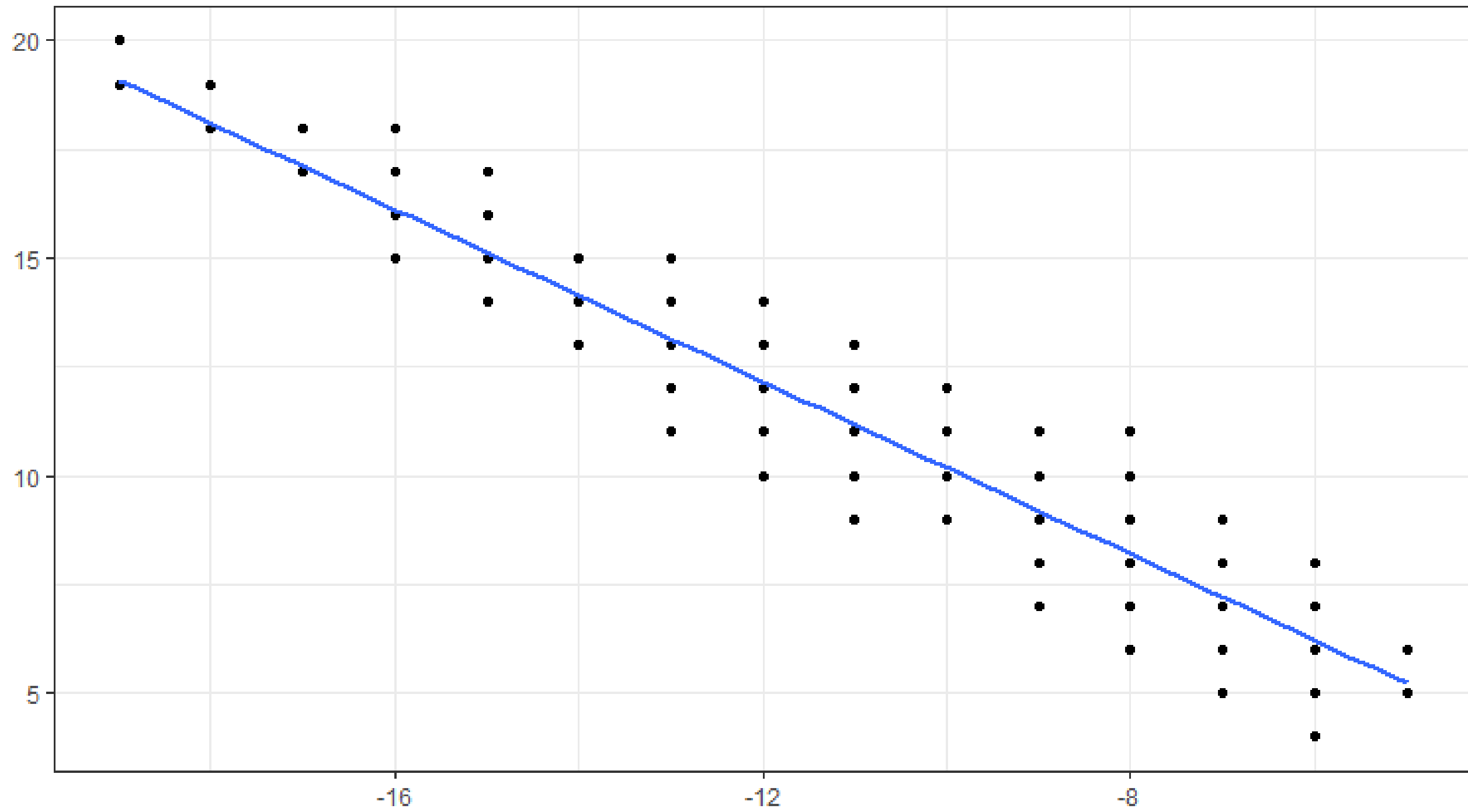
- CORREL() function
 - Two numeric data ranges
- Correlate values in columns G and H
 - =CORREL(G:G,H:H)

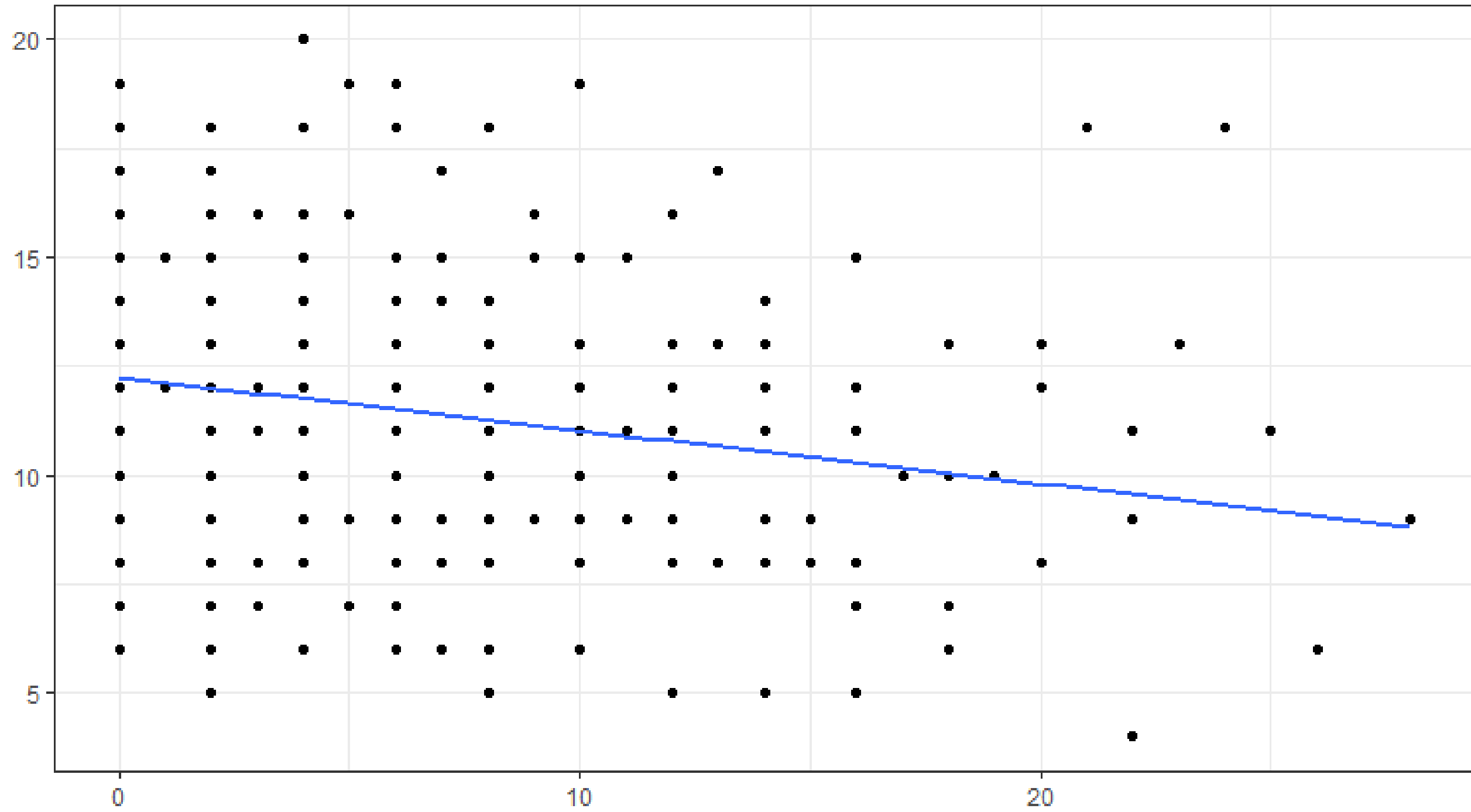
I	J	K	L	M	N
Number of Offenses					
UNKNOWN	SOUTHWEST	NORTH	EAST		WEST-SOUTHWEST
1	9	42	6		0.693627901
0	0	1	1		
1	0	1	0		
0	1	0	0		
1	3	13	6		
0	3	7	1		
0	1	1	0		
0	0	0	2		
0	1	4	3		
0	2	3	3		
0	0	2	2		

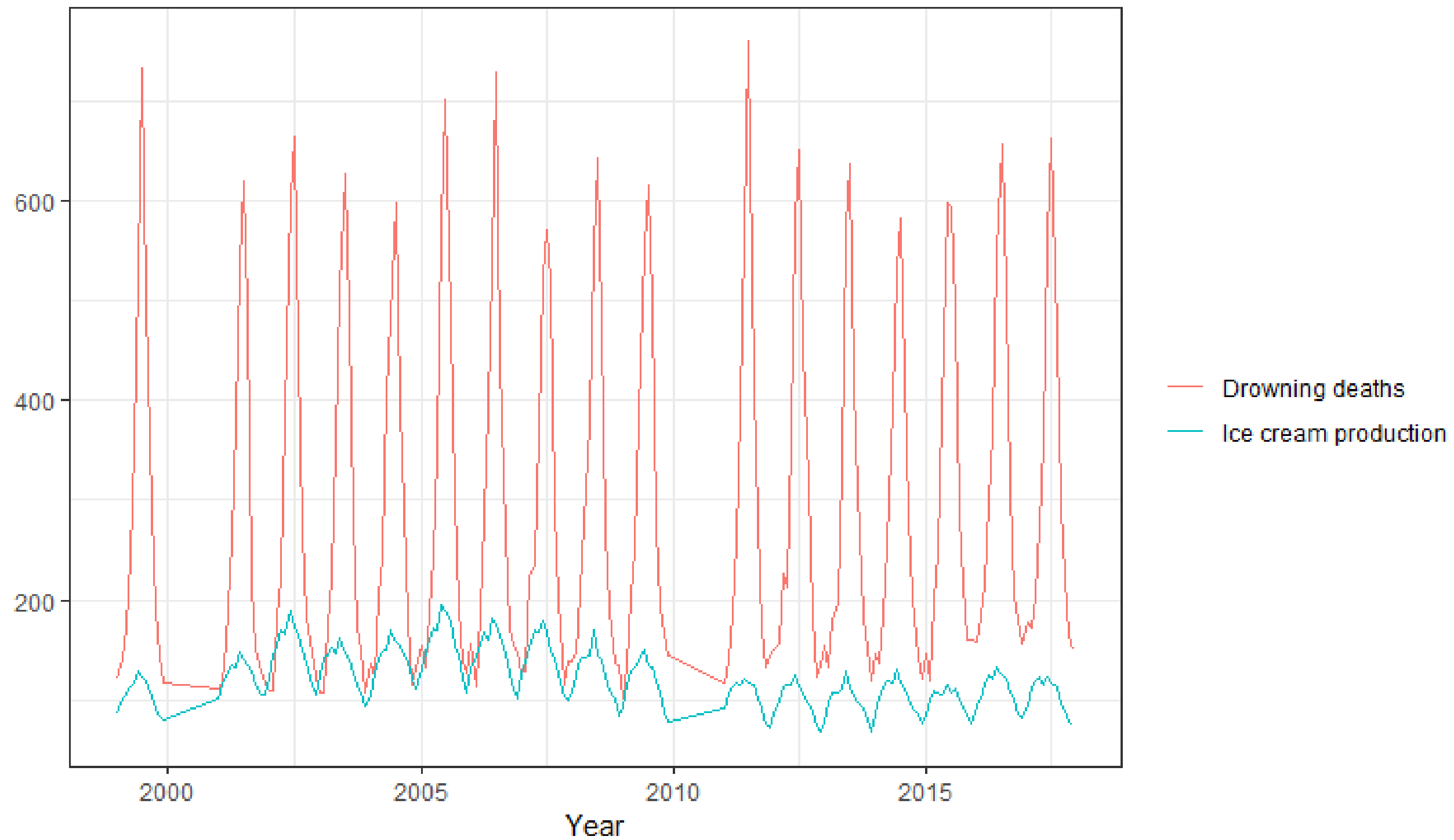
Evaluating strength of relationships

- Weak
 - -0.3 to 0.3
- Moderate
 - -0.3 to -0.7 **or** 0.3 to 0.7
- Strong
 - -0.7 to -1.0 **or** 0.7 to 1.0









Problems with correlation

- Example
 - Ice cream consumption and drowning deaths
- Causation
 - A causes B?
 - B causes A?
- Third variable problem
 - C causes both A and B?

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS