Outliers

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

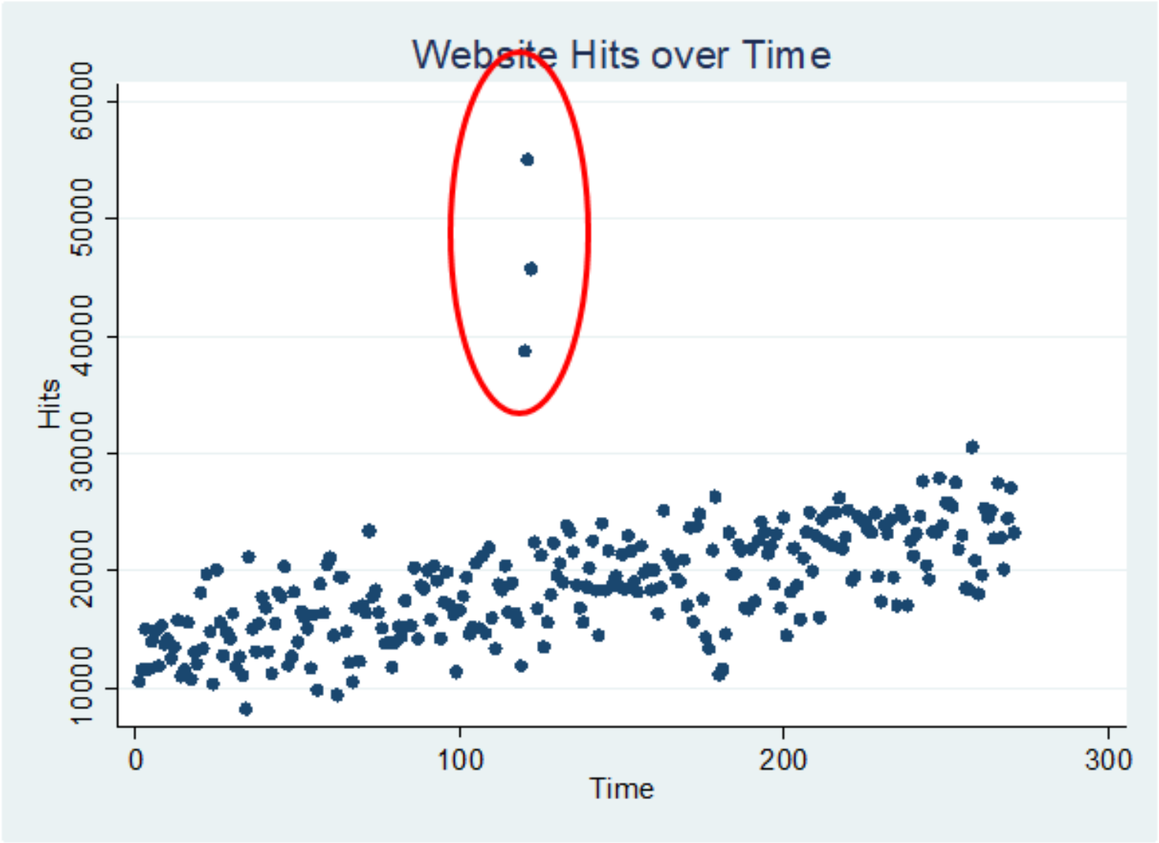
Defining risk

- Exposure to danger
  - Likelihood
    - How frequently occurring
  - Consequences
    - Severity
- **Example:** earthquakes
  - Rare
  - Severe

		Consequence		
		Minor	Moderate	Severe
Likelihood	Likely			
	Possible			
	Rare			

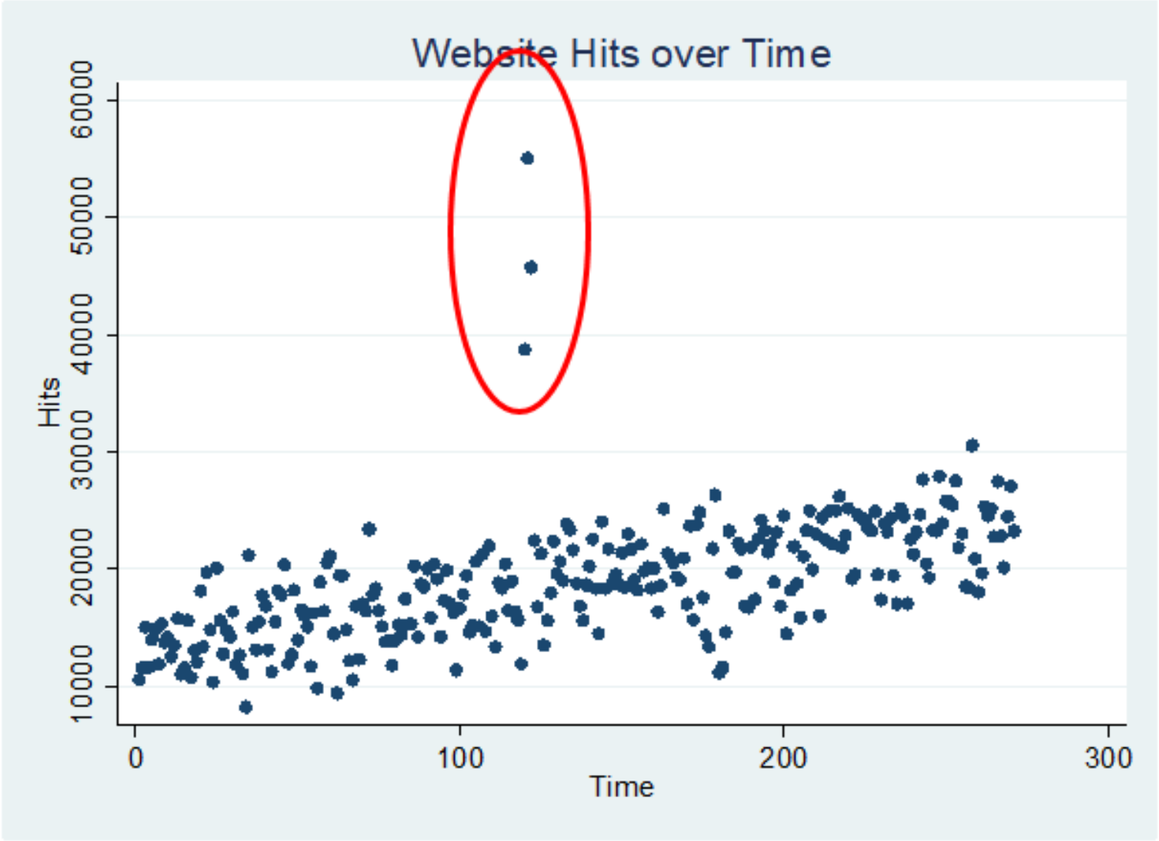
Defining outliers

- Outside normal range
  - Skew data
  - Low likelihood
  - (Potentially) severe
- Examples
  - Home prices
  - Website hits



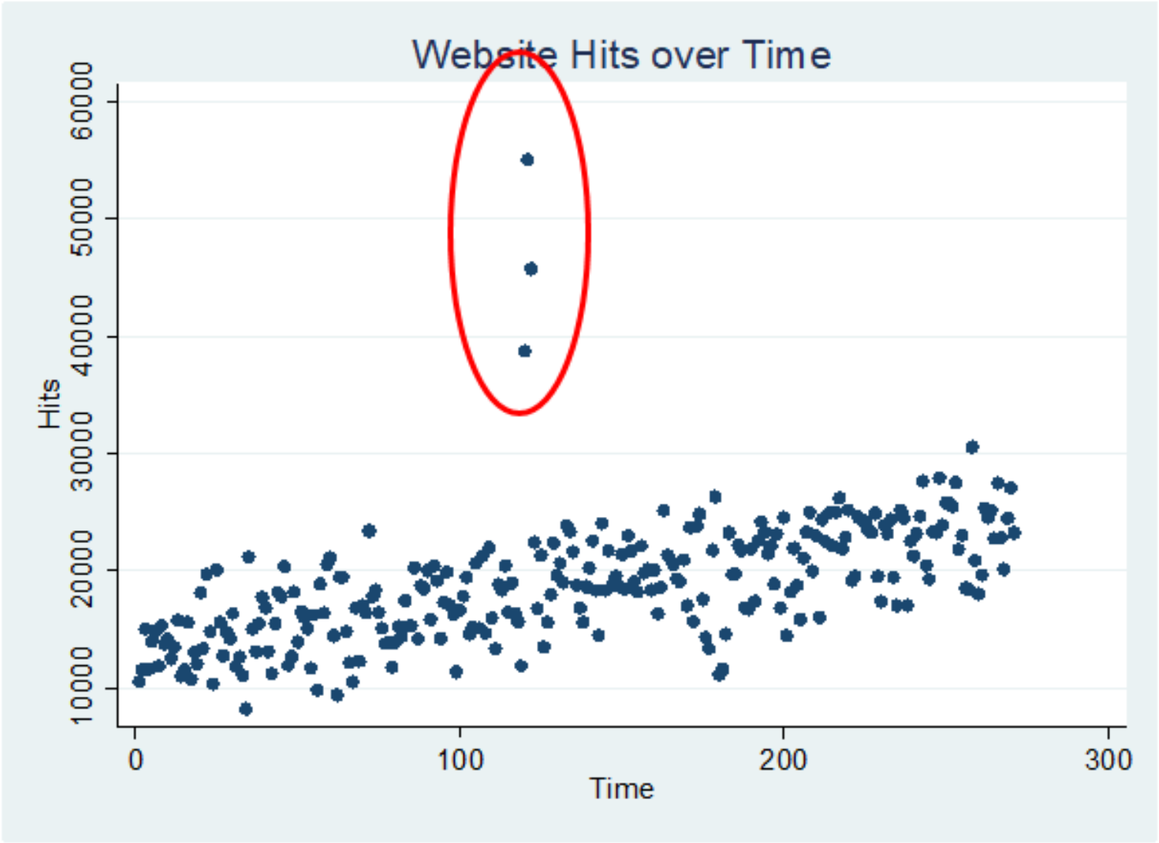
SORTing outliers

- `SORT()` function
  - `range` : Cells to sort
  - `sort_column` : Column to sort by
  - `is_ascending` : Whether to sort low to high



FILTERing outliers

- `FILTER()` function
  - `range` : Cells to filter
  - `condition1` : Filter to apply
    - `>` , `<=` , etc.
    - `A2:A20 = "WEST"`



	A	B	C	D	E
1	Precinct	# Vehicles	# Injured	# Predicted Injured	Absolute Deviation
2	WEST	2	0	0	0
3	MADISON	2	0	0	0
4	NORTH	2	0	0	0
5	SOUTH	2	0	0	0
6	EAST	2	0	0	0
7	HERMITAGE	3	0	0	0
8	NORTH	2	0	1	1
9	HERMITAGE	2	0	0	0
10	EAST	2	0	1	1
11	CENTRAL	2	0	0	0
12	HERMITAGE	1	0	0	0
13	CENTRAL	3	1	0	1
14	CENTRAL	2	0	1	1
15	CENTRAL	2	0	0	0
16	MIDTOWN	2	1	0	1
17	CENTRAL	2	0	0	0
18	SOUTH	2	0	0	0
19	SOUTH	2	0	0	0

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

Sparklines

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

Sparklines

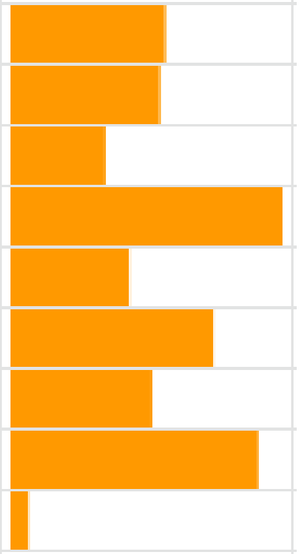
- `SPARKLINE()` function
  - `data` : The cell(s) to plot
  - `options` : Options to configure the plot

Precinct	# Vehicles	
WEST	137	
MADISON	132	
NORTH	84	
SOUTH	240	
EAST	105	
HERMITAGE	179	
CENTRAL	125	
MIDTOWN	219	
UNKNOWN	16	

Options

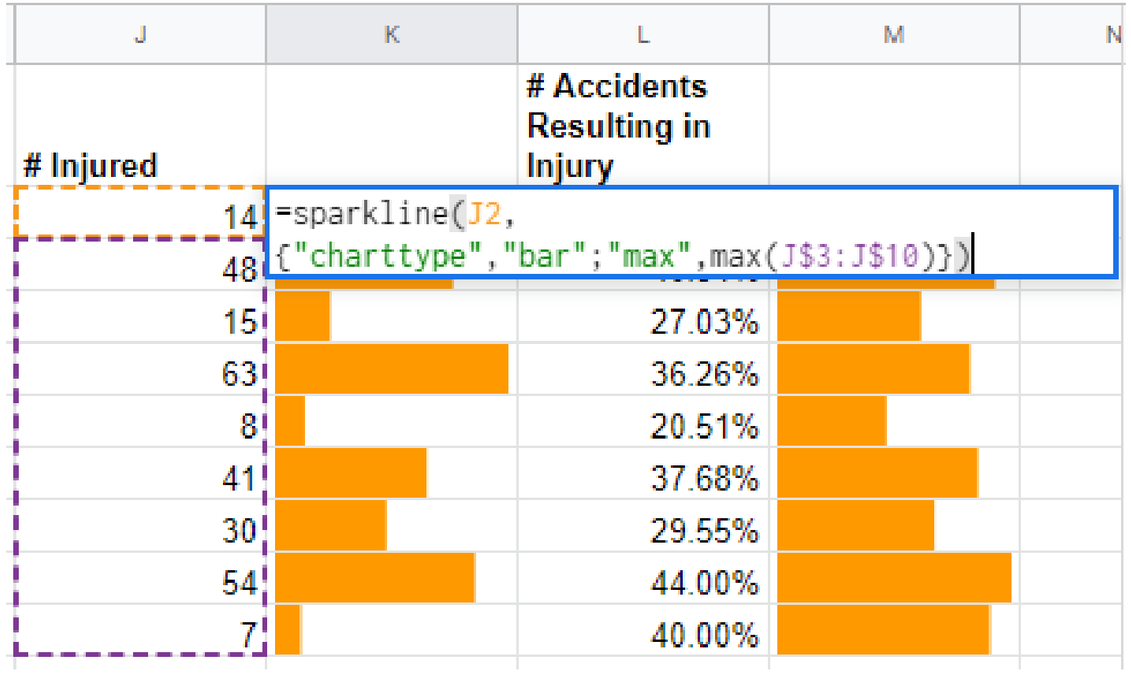
- `charttype`
  - bar, line, column, etc.
- `max`
  - upper limit

Precinct	# Vehicles
WEST	137
MADISON	132
NORTH	84
SOUTH	240
EAST	105
HERMITAGE	179
CENTRAL	125
MIDTOWN	219
UNKNOWN	16



Examples

- `=SPARKLINE(data, [options])`



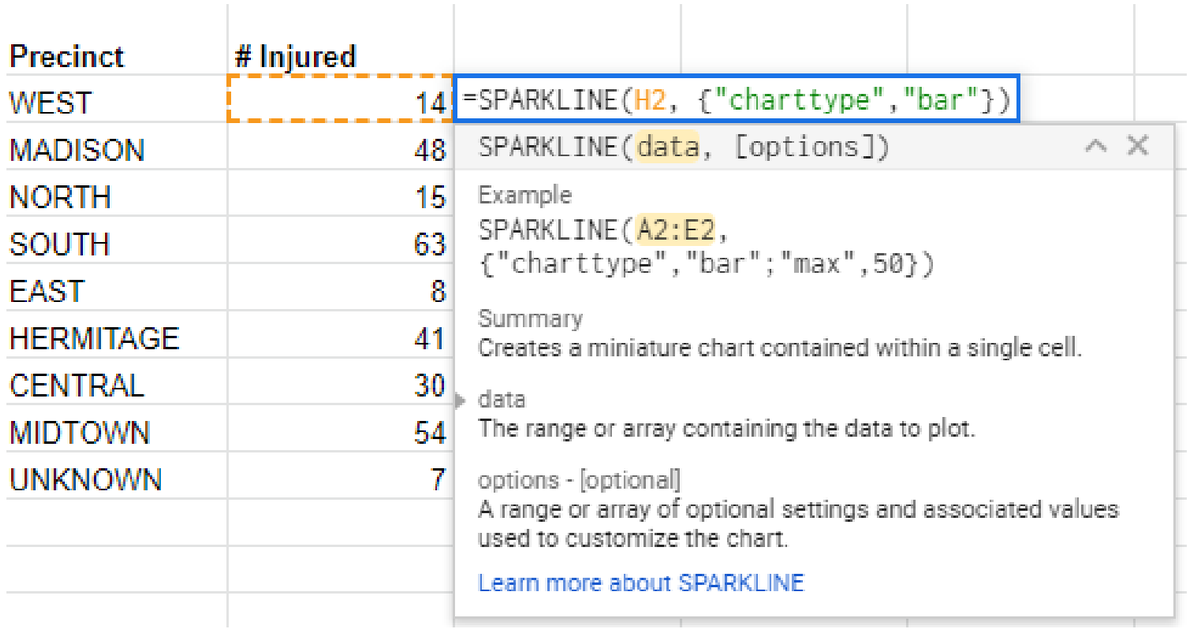
Examples

- `=SPARKLINE(H2, [options])`

Precinct	# Vehicles	
WEST	137	
MADISON	132	
NORTH	84	
SOUTH	240	
EAST	105	
HERMITAGE	179	
CENTRAL	125	
MIDTOWN	219	
UNKNOWN	16	

Examples

- `=SPARKLINE(H2, {"charttype", "bar"})`
  - Curly braces
  - [Option name], [option]



Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

What's the worst that could happen?

ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

datacamp

A likely story

- Risk `=` Likelihood `X` consequences

		Consequence		
		Minor	Moderate	Severe
Likelihood	Likely			
	Possible			
	Rare			

Who wants to be a millionaire

- **Example:** lottery
  - Consequences: `$1,000,000`
  - Likelihood: `1 / 11,688,053.52`

Another risk example

- **Consequence:** actual or predicted injuries
- **Likelihood:** crash frequency

	A	B	C	D	E	F
1	**Precinct**	**# Vehicles**	**# Injured**	**Consequence**	**Likelihood**	**Risk**
2	WEST	2	0	0.430810801	3.65%	0.01572302194
3	MADISON	2	0	0.430810801	32.58%	0.1403398822
4	NORTH	2	0	0.430810801	9.52%	0.0410296001
5	SOUTH	2	0	0.430810801	19.17%	0.0825720702
6	EAST	2	0	0.430810801	0.00%	0
7	HERMITAGE	3	0	0.6884078513	13.97%	0.09614634794
8	NORTH	2	0	0.430810801	9.52%	0.0410296001
9	HERMITAGE	2	0	0.430810801	13.97%	0.06016910629
10	EAST	2	0	0.430810801	0.00%	0
11	CENTRAL	2	0	0.430810801	18.40%	0.07926918739
12	HERMITAGE	1	0	0.1732137508	13.97%	0.02419186463
13	CENTRAL	3	1	0.6884078513	18.40%	0.1266670446
14	CENTRAL	2	0	0.430810801	18.40%	0.07926918739
15	CENTRAL	2	0	0.430810801	18.40%	0.07926918739
16	MIDTOWN	2	1	0.430810801	15.07%	0.06491669605
17	CENTRAL	2	0	0.430810801	18.40%	0.07926918739
18	SOUTH	2	0	0.430810801	19.17%	0.0825720702
19	SOUTH	2	0	0.430810801	19.17%	0.0825720702
20	SOUTH	1	0	0.1732137508	19.17%	0.03319930223

Review: useful functions

- `FORECAST()` : Predicts values
- `SUMIF()` : Add values from rows that meet criteria

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS

Risky business

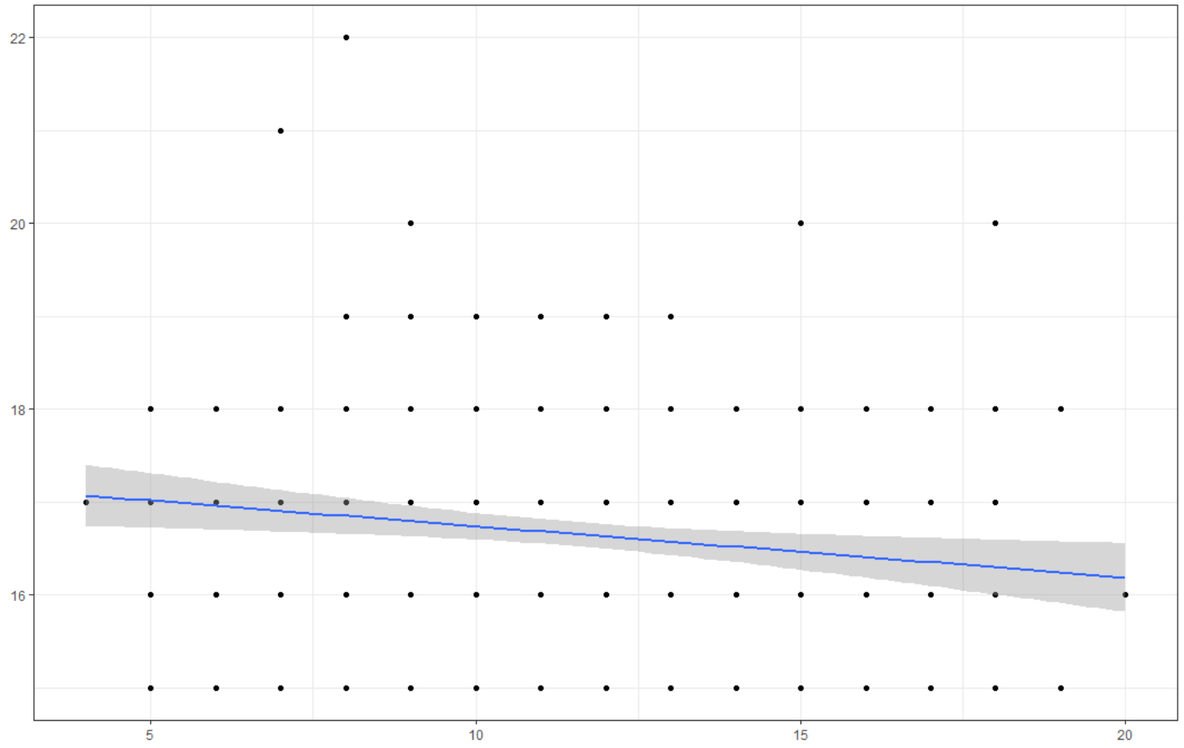
ERROR AND UNCERTAINTY IN SPREADSHEETS

Evan Kramer
Instructor

datacamp

Bring the noise

- Noise
  - Unexplained variation



Random numbers

- Multiply predictions
- `=RANDBETWEEN(low, high)`

Random numbers

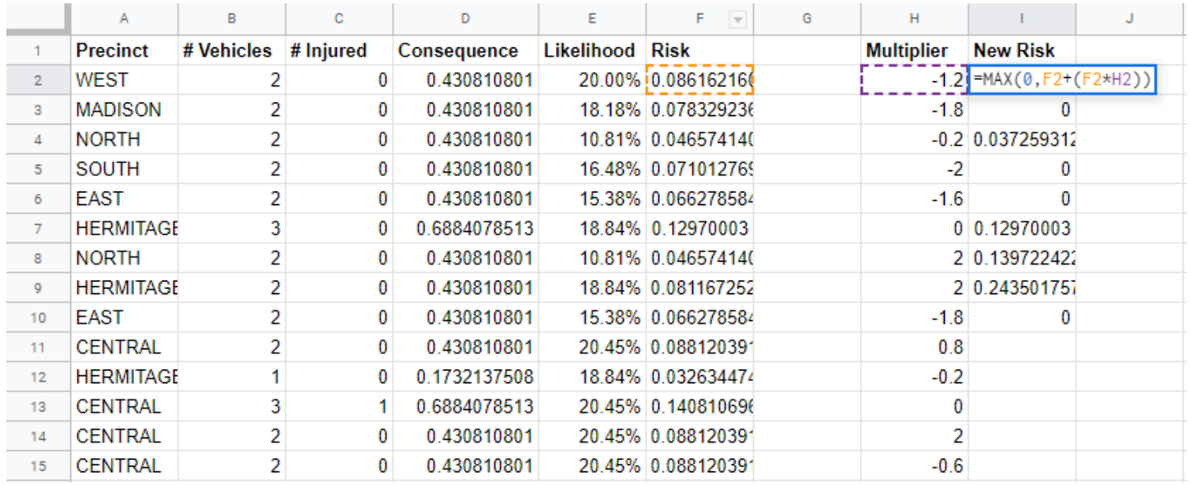
- `RANDBETWEEN(-100, 100)`
  - Too much variation
- `RANDBETWEEN(-100, 100)/50`

Too random?

- **Challenge:** Results "refresh"
- **Solution:** Paste as values

Adjusting data

- Multiply and add product to actual data
  - **Problem:** Negative numbers
  - **Solution:** `MAX()` function
    - E.g., `MAX(0,F2+(F2*H2))`



	A	B	C	D	E	F	G	H	I	J
1	Precinct	# Vehicles	# Injured	Consequence	Likelihood	Risk		Multiplier	New Risk	
2	WEST	2	0	0.430810801	20.00%	0.086162166		-1.2	=MAX(0,F2+(F2*H2))	
3	MADISON	2	0	0.430810801	18.18%	0.078329236		-1.8	0	
4	NORTH	2	0	0.430810801	10.81%	0.046574140		-0.2	0.037259312	
5	SOUTH	2	0	0.430810801	16.48%	0.071012769		-2	0	
6	EAST	2	0	0.430810801	15.38%	0.066278584		-1.6	0	
7	HERMITAGE	3	0	0.6884078513	18.84%	0.12970003		0	0.12970003	
8	NORTH	2	0	0.430810801	10.81%	0.046574140		2	0.139722422	
9	HERMITAGE	2	0	0.430810801	18.84%	0.081167252		2	0.243501757	
10	EAST	2	0	0.430810801	15.38%	0.066278584		-1.8	0	
11	CENTRAL	2	0	0.430810801	20.45%	0.088120391		0.8		
12	HERMITAGE	1	0	0.1732137508	18.84%	0.032634474		-0.2		
13	CENTRAL	3	1	0.6884078513	20.45%	0.140810696		0		
14	CENTRAL	2	0	0.430810801	20.45%	0.088120391		2		
15	CENTRAL	2	0	0.430810801	20.45%	0.088120391		-0.6		

Framing effect

- Choose between:
  - Sure gain of $250
    - (84% of respondents)

- Choose between:
  - 25% chance of $1000, 75% chance of $0
    - (16% of respondents)

Framing effect

- Choose between:
  - Sure loss of $750
    - (13% of respondents)

- Choose between:
  - 75% chance of losing $1000, 25% chance of losing $0
    - (87% of respondents)

Framing effect

- **Conclusion:** loss aversion

Let's practice!

ERROR AND UNCERTAINTY IN SPREADSHEETS