



**Education Conversations
of the Future**
Quality Future Focussed Education



How teaching statistics will have to adapt to prepare learners to manage data in the future

Herholdt Bezuidenhout



Dr. Erna Lampen



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY



**Education Conversations
of the Future**
Quality Future Focussed Education



Questions:

- Do our current approach in teaching statistics in schools prepare our learners adequately for the the emerging job market?
- How can we adapt our teaching of statistics to prepare learners?

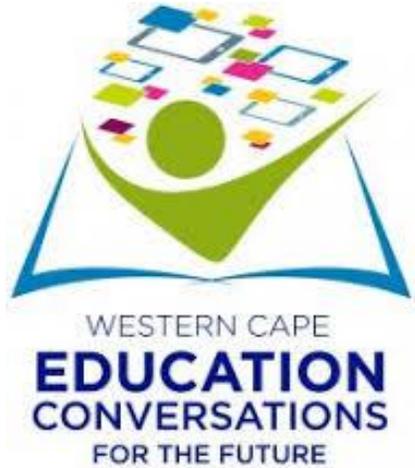
Herholdt Bezuidenhout



Dr. Erna Lampen



Building a new future on



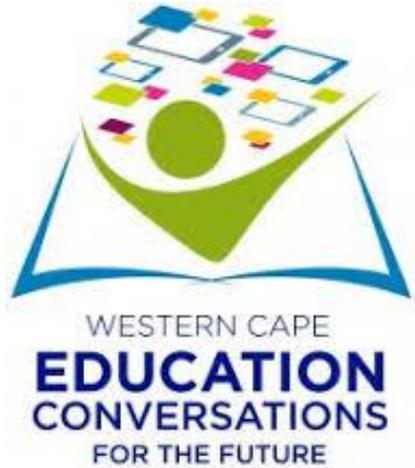
Future
Data Science?

Discussion points:

- Background
- Limitations
- Possibilities
- Way forward

Past
Statistics

Building a new future on



Future
Data Science?

Discussion points:

- Background
- Limitations
- Possibilities
- Way forward

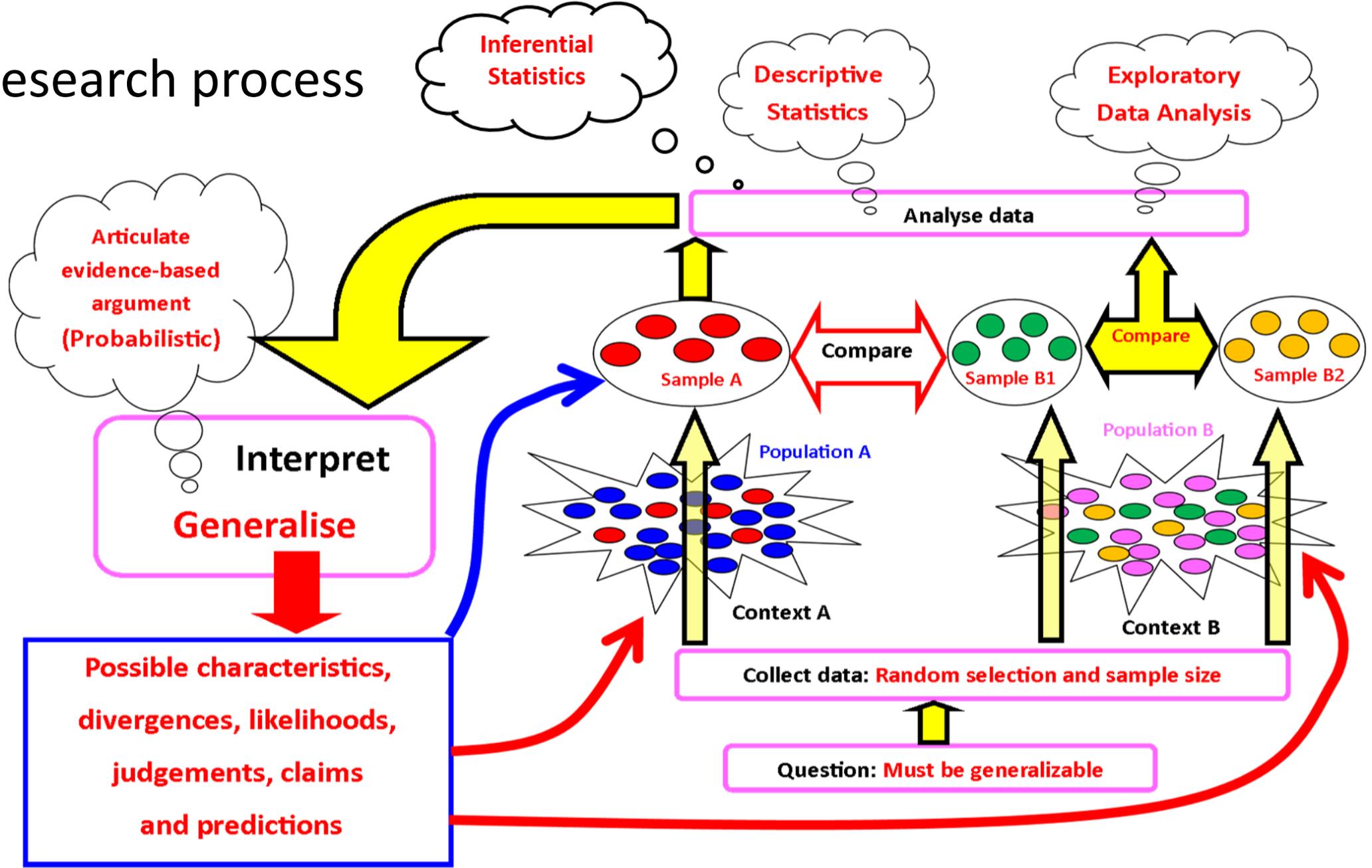
Past
Statistics

Historically



Where or how do I get the best yield?

Statistical research process



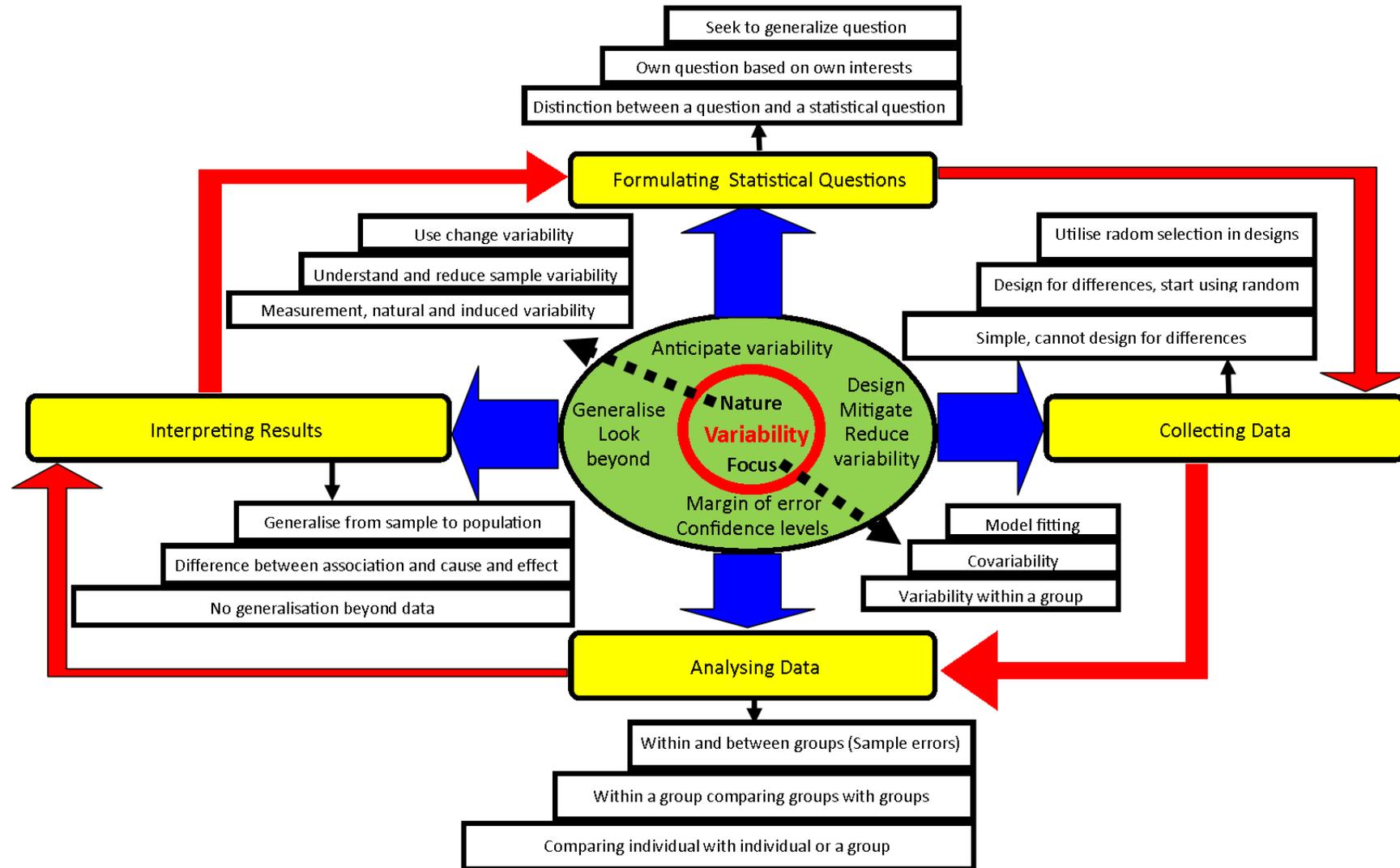
The following is a possible list of causes of variation:

- a) Measurement error
- b) Process variation
- c) Induced variation
- d) Sample variation
- e) Natural variation
- f) Multiple variations



Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-k-12 curriculum framework*. Alexandria: American Statistical Association.

GUIDELINES FOR ASSESSMENT AND INSTRUCTION IN STATISTICS EDUCATION (GAISE)



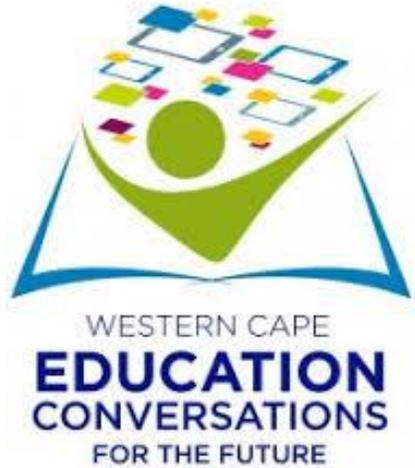
Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-k-12 curriculum framework*. Alexandria: American Statistical Association.

Importance of statistics today:

- Numbers govern our lives. We find **numbers** in the newspapers, reports at work and **all around us** humans are **bombarded with quantitative information** ranging from results from experiments to crime and other statistics.
- Modern man **needs sound statistical reasoning** to cope with the quantitative information as a citizen, employee or family member in order to have a healthy, happy and productive life.
- It is not only required from them to **make informed personal choices** but also to **improve quality** and **enhance productivity** as effective employees at the workplace.
- However, humans as consumers of information must be **highly sceptical of the reliability** of other people or organisations' claims and must be able to **confront those authorities confidently**.

(Franklin, et al., 2007, pp. 2-3; Gal, 2004, pp. 4-5).

Building a new future on

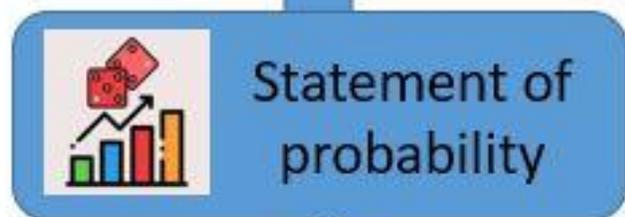


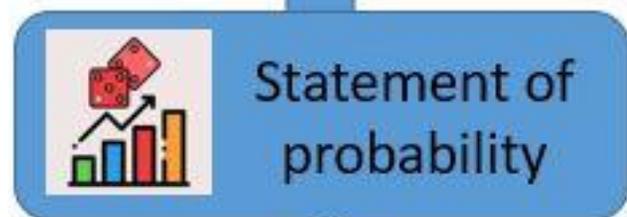
Past
Statistics

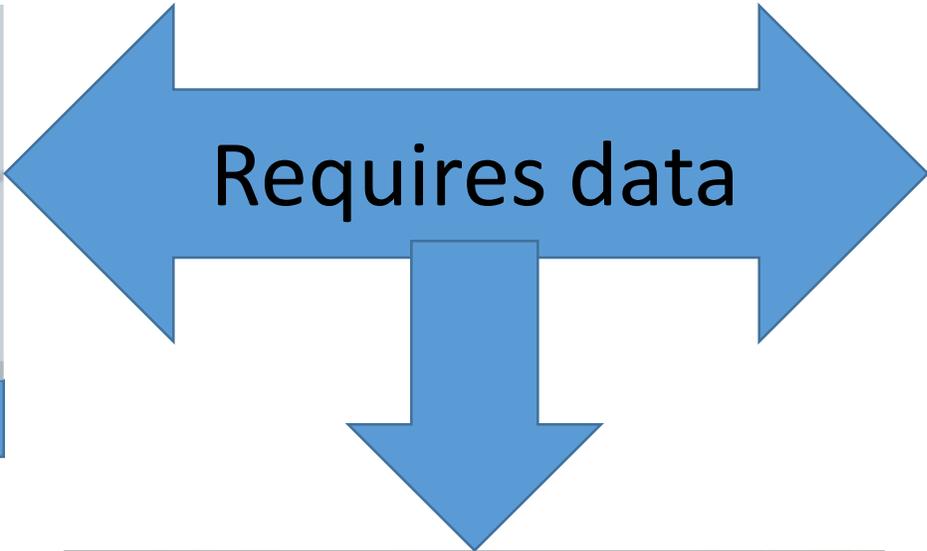
Future
Data Science?

Discussion points:

- Background
- Limitations
- Possibilities
- Way forward





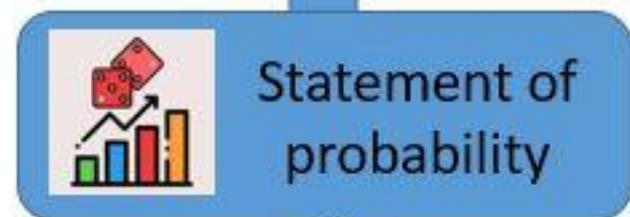


Internal



External







KING CRAB JOHN R. MANNING 1968

TEMPERATURES (°C)		CORRECTED	KING CRAB		DATE		
(M) SURFACE	(M) BOTTOM	SURFACE TEMPERATURE	TOTAL	MOLDS (NO)		SPERMATOPHYTES (NO)	
		%		> 145mm	< 145mm	> 100mm	< 100mm
0.00	0.00						
27	2.0	31.826	7	5	0	1	0
31	3.1	32.048	1	1	0	0	0
31	3.1	31.762	62	5	2	42	6
2.3	2.4	31.794	3	0	2	1	0
2.3	1.8	31.738	5	3	2	0	0
2.4	1.4	31.671	26	8	2	14	2
2.3	1.4	31.647	47	4	23	17	3
2.1	1.4	31.578	118	7	50	22	57
2.3	1.0	31.898	571	58	238	263	162
1.9	1.5	31.878	651	23	238	219	133
2.1	1.4	31.587	0	0	0	0	0
2.1	1.4	31.847	111	20	20	21	15



LIMITATIONS



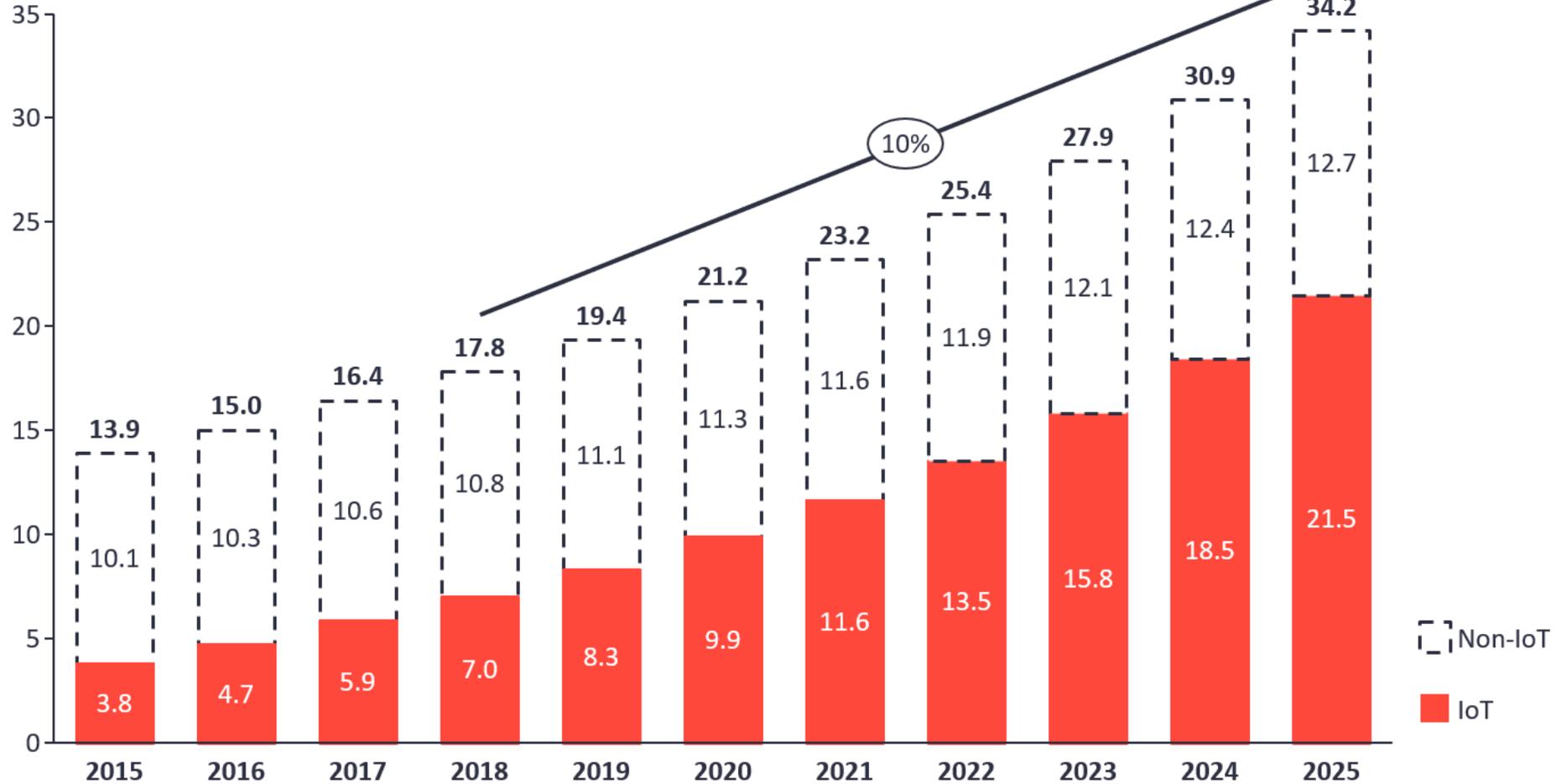
© Can Stock Photo - csp23152715

2019 *This Is What Happens In An Internet Minute*



Total number of active device connections worldwide

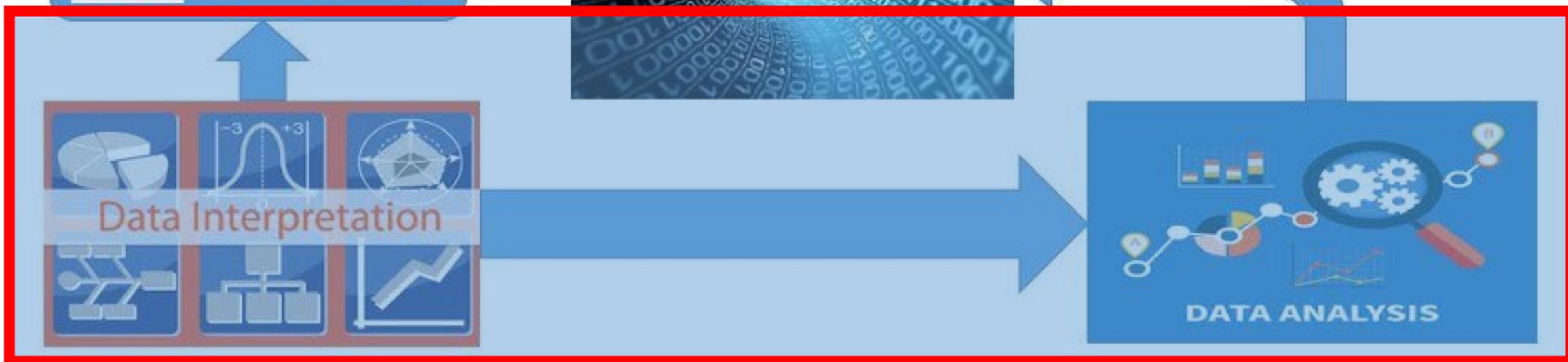
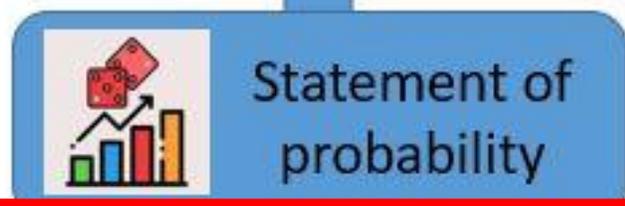
Number of global active Connections (installed base) in Bn

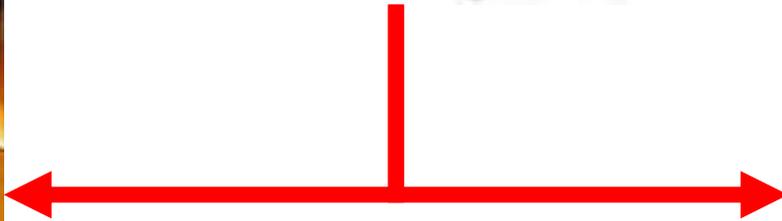
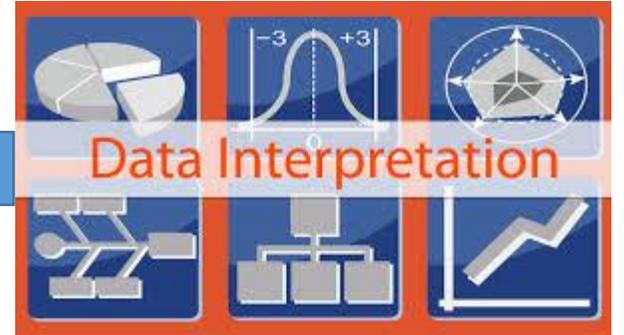
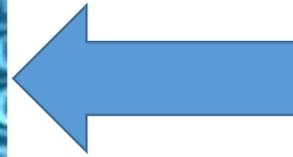
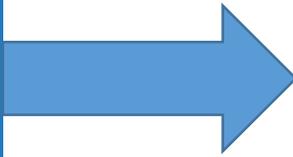


Note: Non-IoT includes all mobile phones, tablets, PCs, laptops, and fixed line phones. IoT includes all consumer and B2B devices connected – see IoT break-down for further details
 Source: IoT Analytics Research 2018

Big Data: Google Server “Farm”

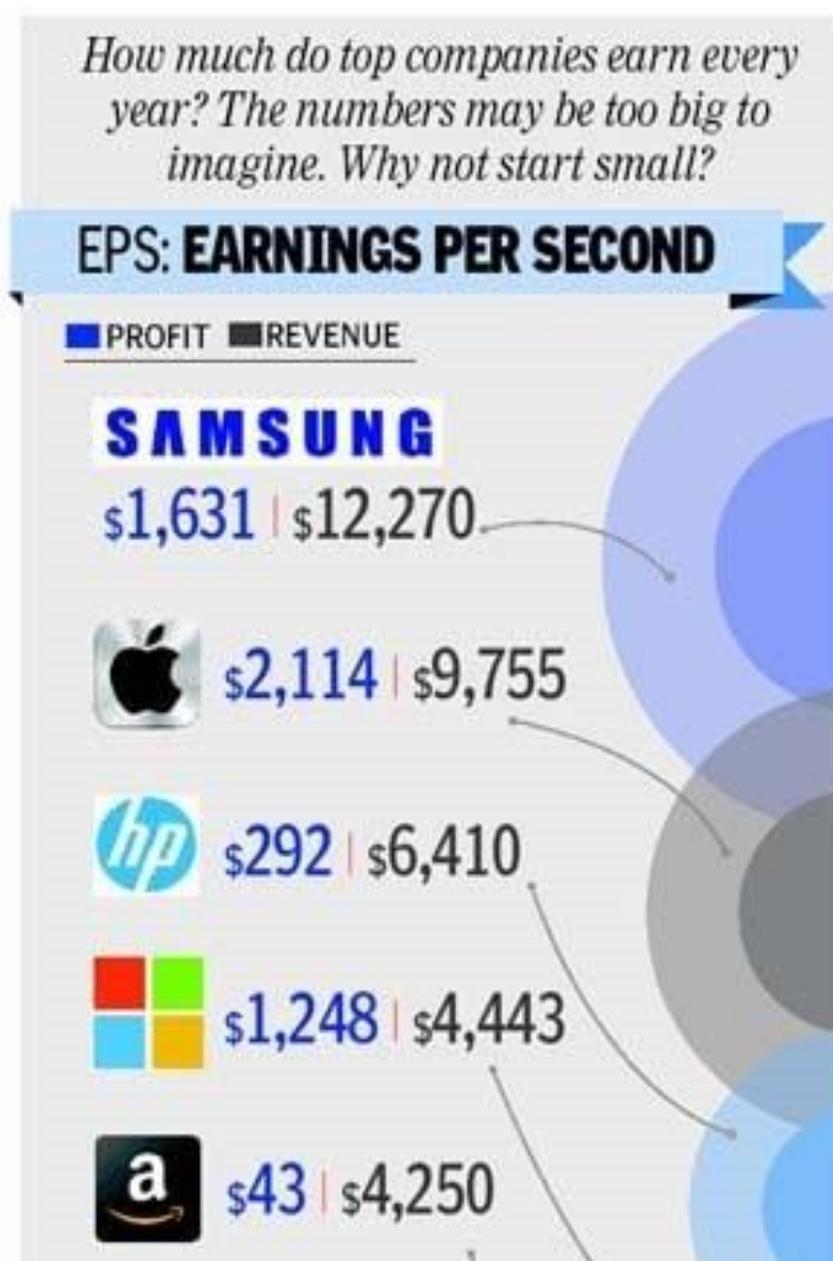




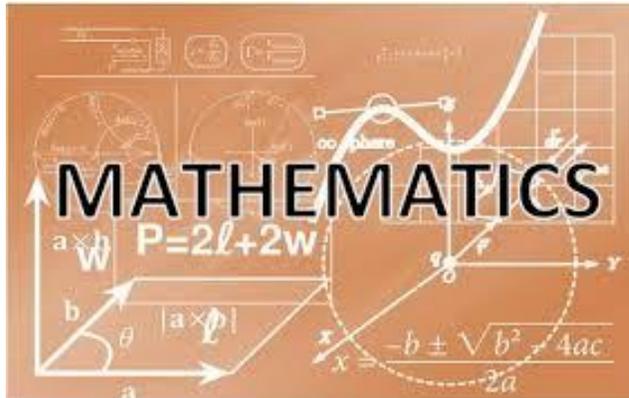
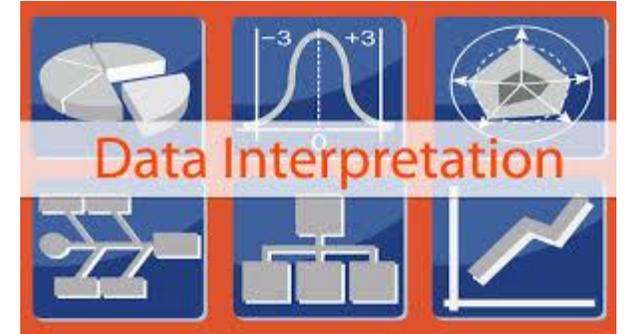
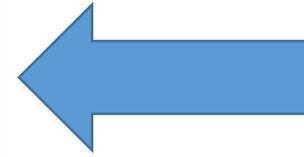




Currency



<https://moneyconnexion.com/per-second-earnings-of-top-companies.htm>



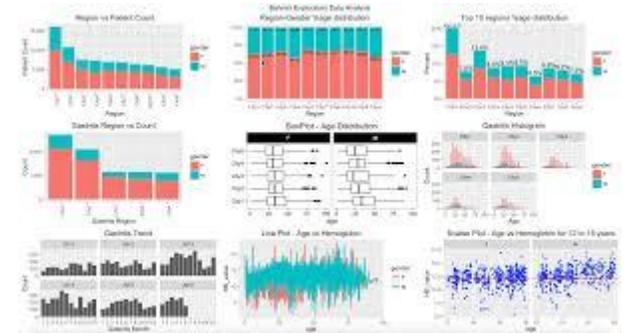
Descriptive
Statistics

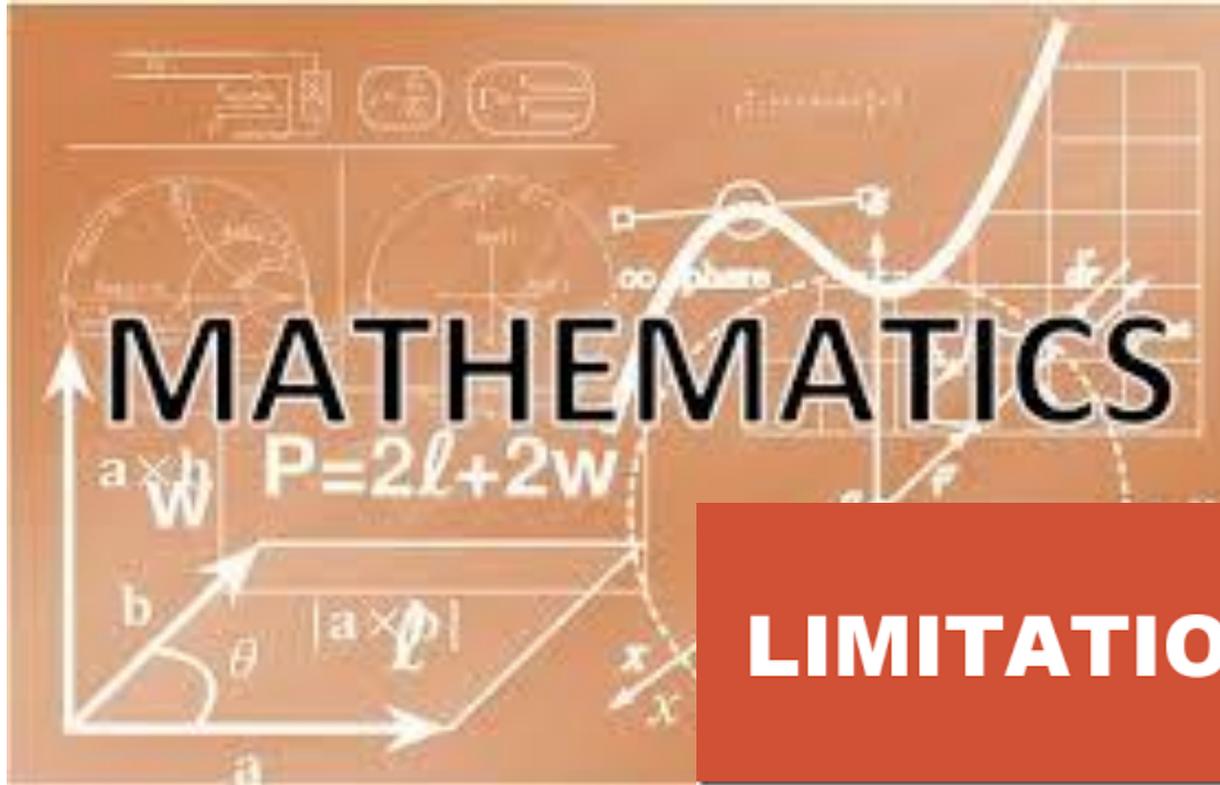
- Organise
- Summarise
- Simplify
- Describe and present data

Inferential
Statistics

- Generalise from samples to populations
- Hypothesis testing
- Making predictions

Exploratory
Data
Analysis





“Statistics is a science,
not a branch of mathematics,
but uses mathematical models
as essential tools.”

–John Tukey

LIMITATIONS

- The all present idea of variability are at the root for the existence of statistics. (Non-deterministic answers.)
- For statistics data are not just numbers but numbers within a context.
- We may use mathematical processes to determine when and how things happen but to answer the “why” question we need statistical reasoning.

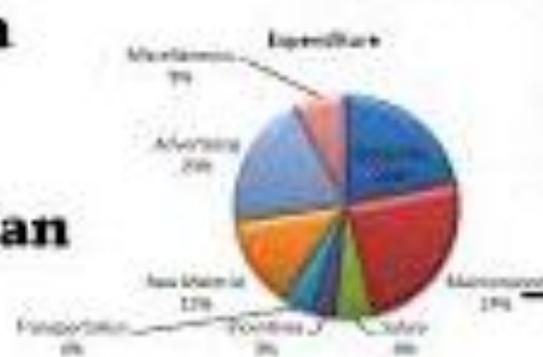
Descriptive Statistics

- Organise
- Summarise
- Simplify
- Describe and present data

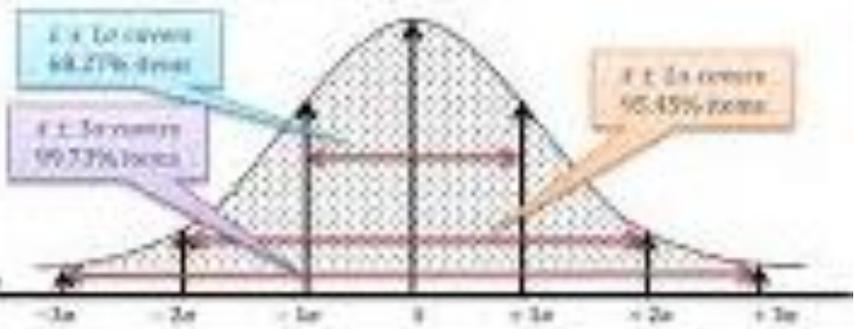
- Describe the basic features of the data
- Provide simple summaries about the sample and the measures.
- Simple graphics analysis
- Quantitative analysis of data.

Descriptive Statistics

Mean

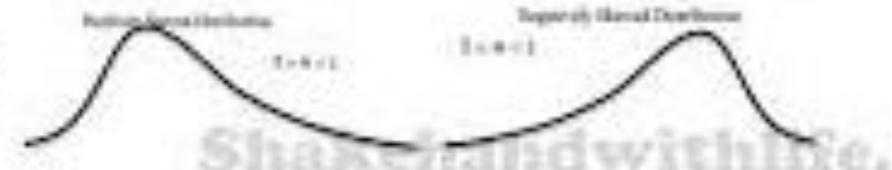


Median



Mode

$$\text{Std. Dev. } \sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$



Descriptive Statistics

- Organise
- Summarise
- Simplify
- Describe and present data

The Anscombe's quartet (1973 by the statistician Francis Anscombe)

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

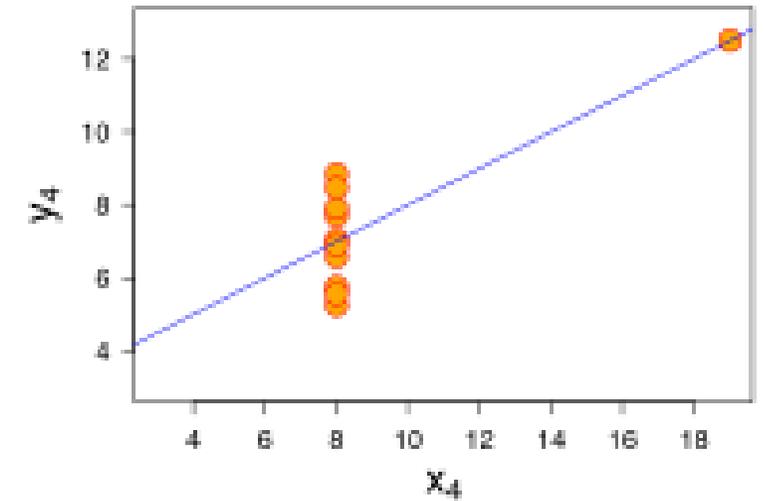
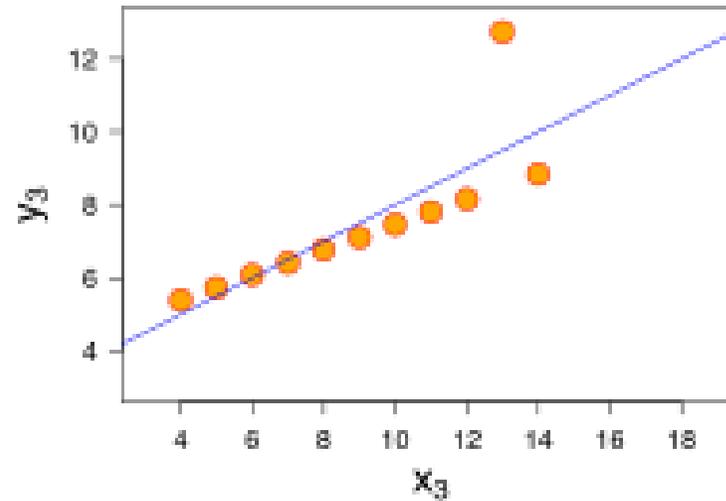
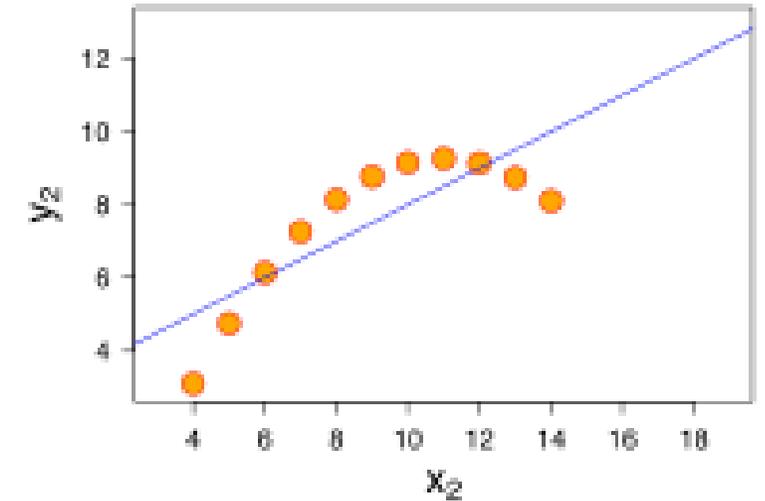
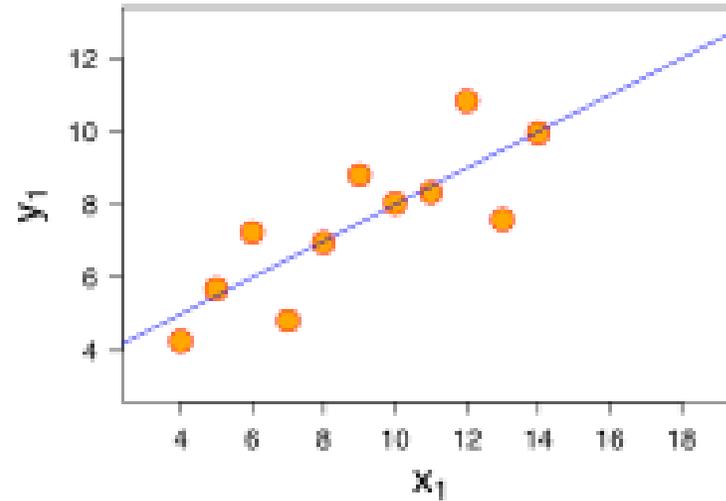
Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	±0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places

Descriptive Statistics

- Organise
- Summarise
- Simplify
- Describe and present data

LIMITATIONS

Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	± 0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places



Exploratory Data Analysis



Data detectives

- Statisticians historically started by exploring and based on their exploration they confirmed the underlying statistics.
- The confirmation techniques and procedures became less flexible and the exploration started to take less priority.
- The best way to learn statistics is to begin with exploration.

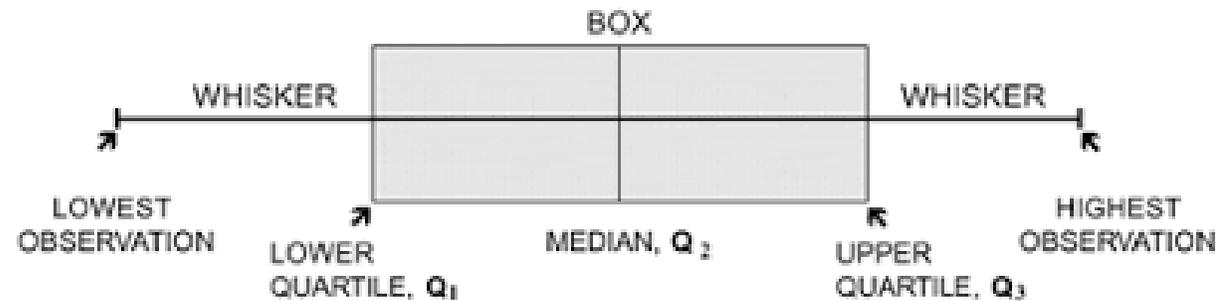


**John
Tukey**

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley Publishing.

Exploratory Data Analysis

Data detective tools

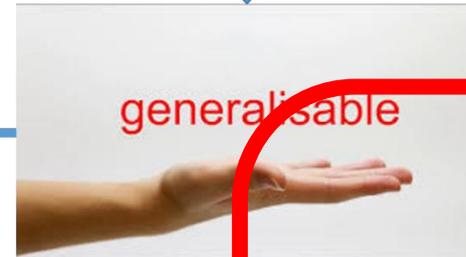
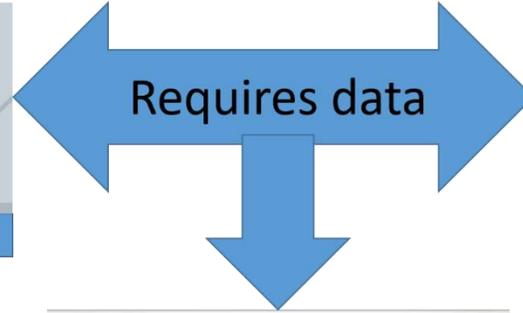


stem	leaf
5	6
6	7, 7, 9
7	2, 4, 7, 7, 8
8	1, 2, 2, 3, 4, 8
9	0, 2, 3, 4

Key: 5 | 6 = 56%

Tukey's introduction of simple data tools, such as stem and leaf plots and boxplots, paved the way for students at all levels to analyse real data interactively without having to spend hours on the underlying theory, calculations, and complicated procedures

LIMITATIONS



EDA

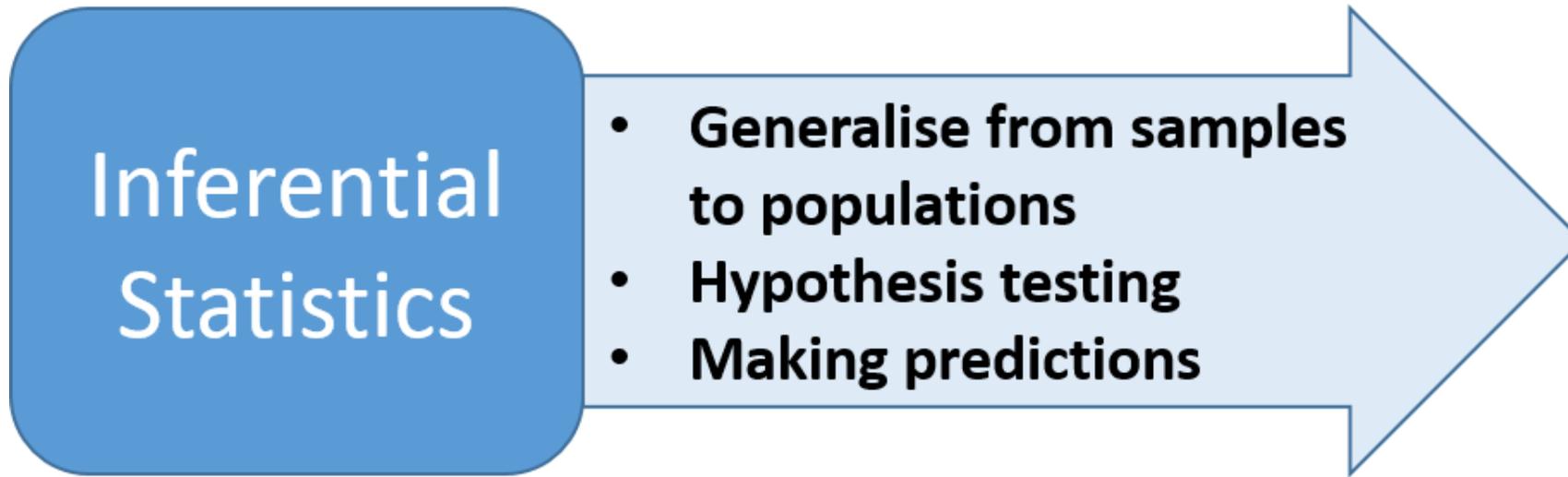
Focus within the data



External

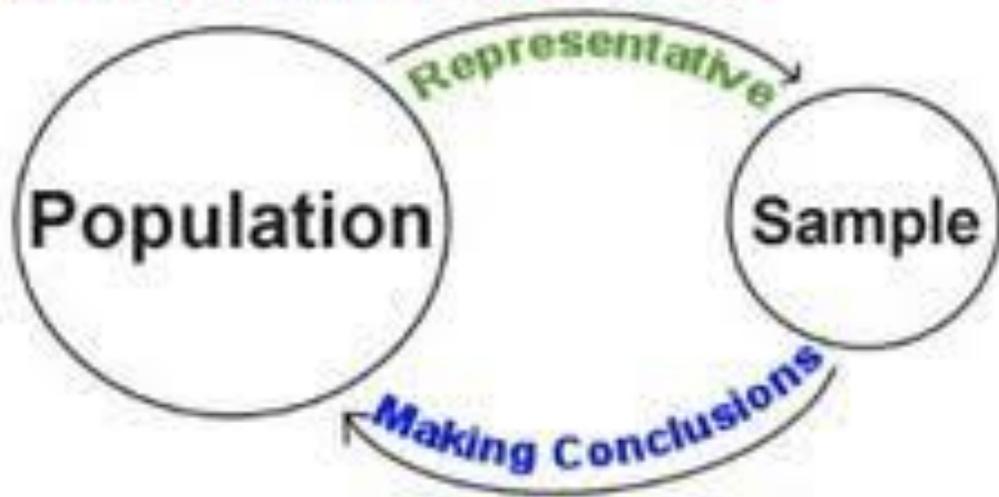


Not extending
beyond the
data



Parametric approaches

Inferential Statistics



Certainty and chance: What is the probability that the proposed cause can take place as a possible event?

Confidence levels: To what level can I trust the inference made about the cause? What provides confidence that the inference is possible and probable?

Inferential Statistics

- **Generalise from samples to populations**
- **Hypothesis testing**
- **Making predictions**

LIMITATIONS

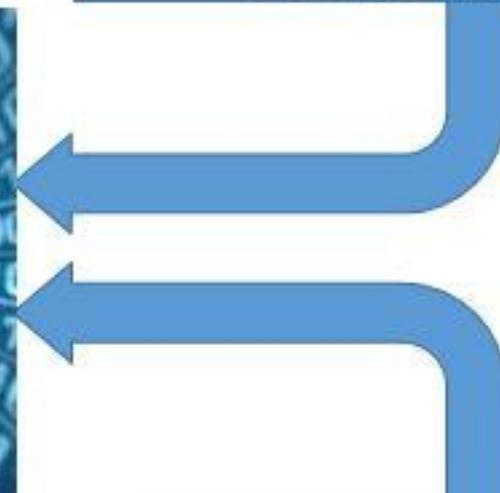
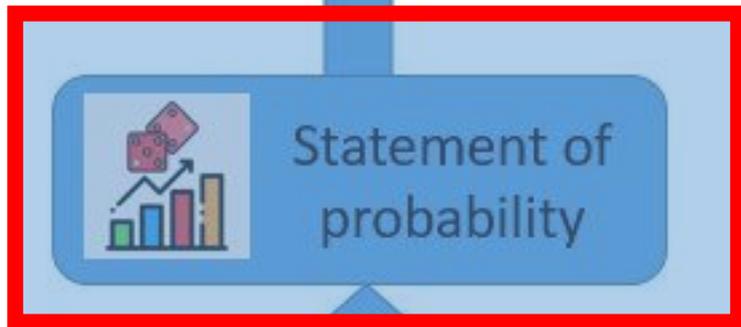
- **Complicated**
- **Reliance on parameters**
- **Exclude the context**

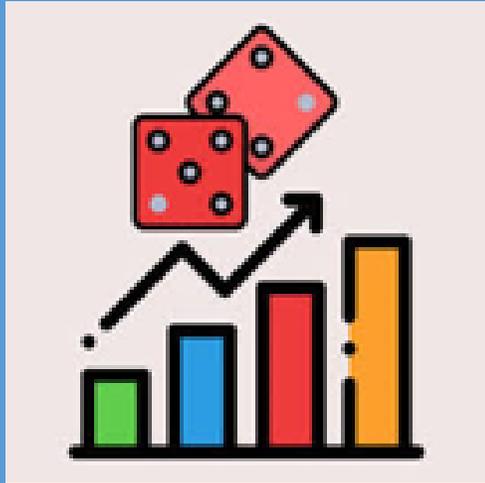


Artificial Intelligence Machine Learning

The scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on **patterns** and **inference** instead.

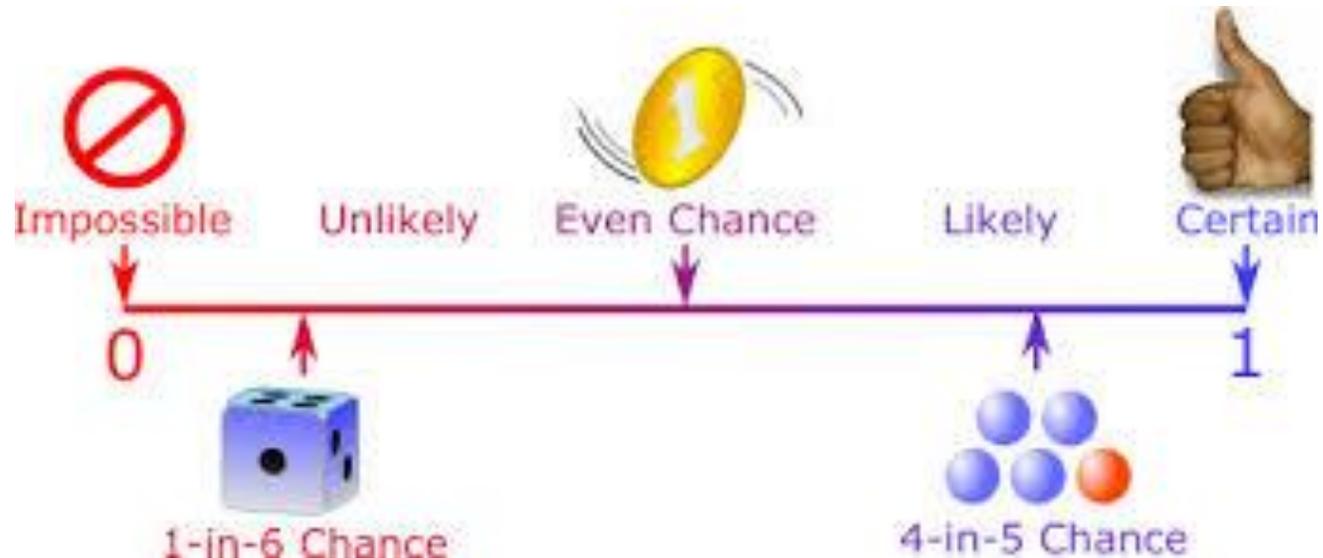
Data mining – unsupervised learning – Predictive analytics





Statement of probability

The answer to my question is non-deterministic





Statement of probability

LIMITATIONS

Which Sides Are Greater Than 3?



Number of Favorable Outcomes = 3

How Many Possibilities Are There?



Total Number of Possibilities = 6

Number of Favorable Outcomes / Total Number of Outcomes = $3/6 = 50\%$

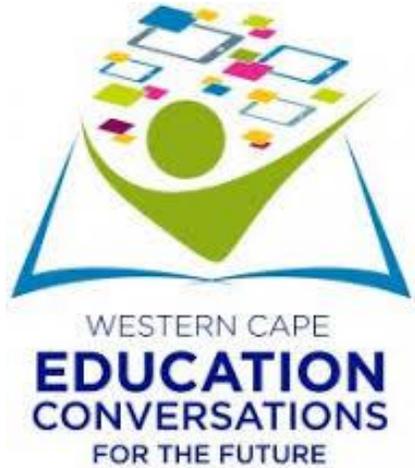


A man chooses a card at random from a pack of playing cards. What is the probability that the card is a Queen?



- Taught in the context of games of chance or options
- Probability is examined in paper 1 and statistics in paper 2

Building a new future on

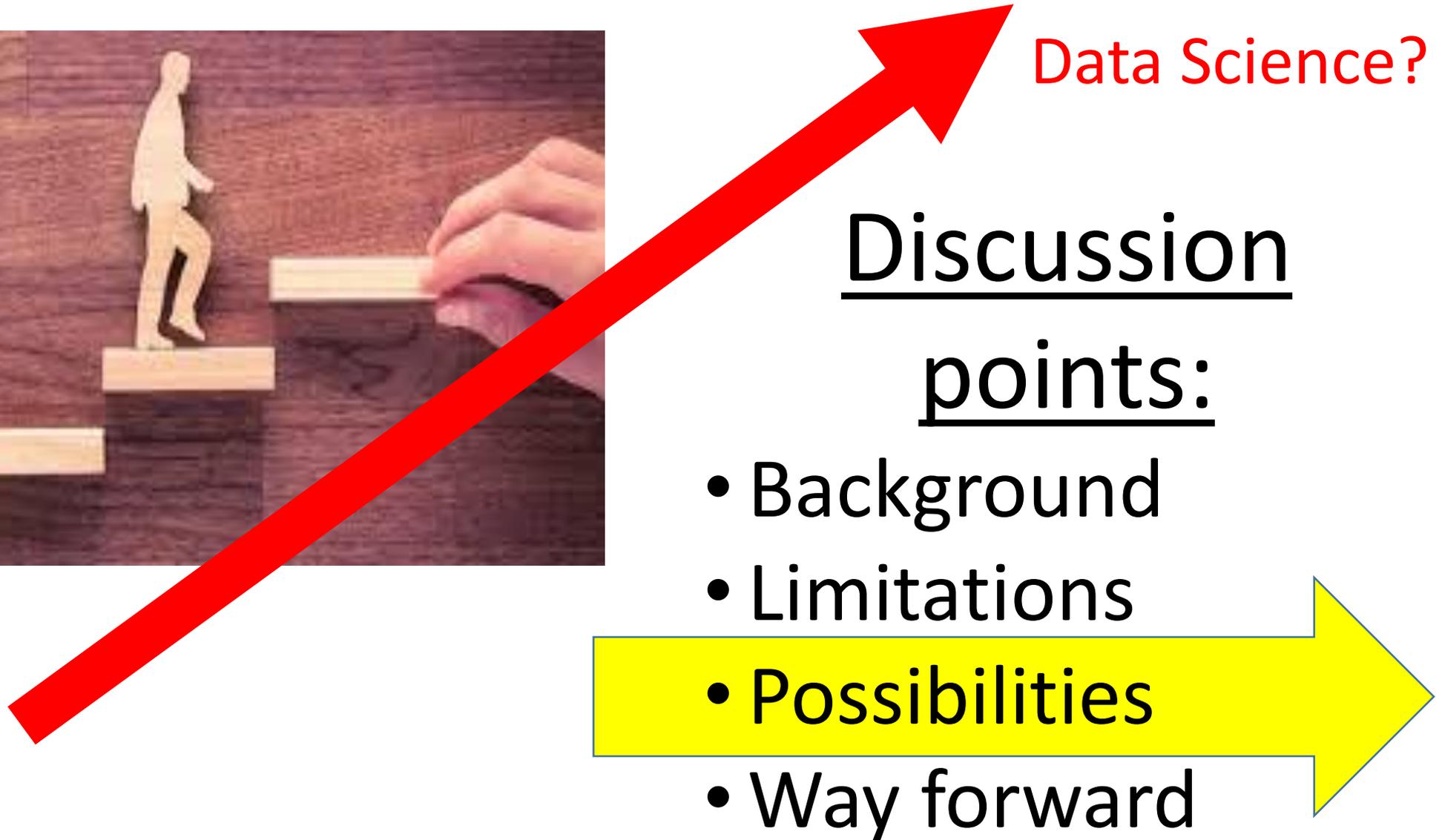


Future
Data Science?

Discussion points:

- Background
- Limitations
- Possibilities
- Way forward

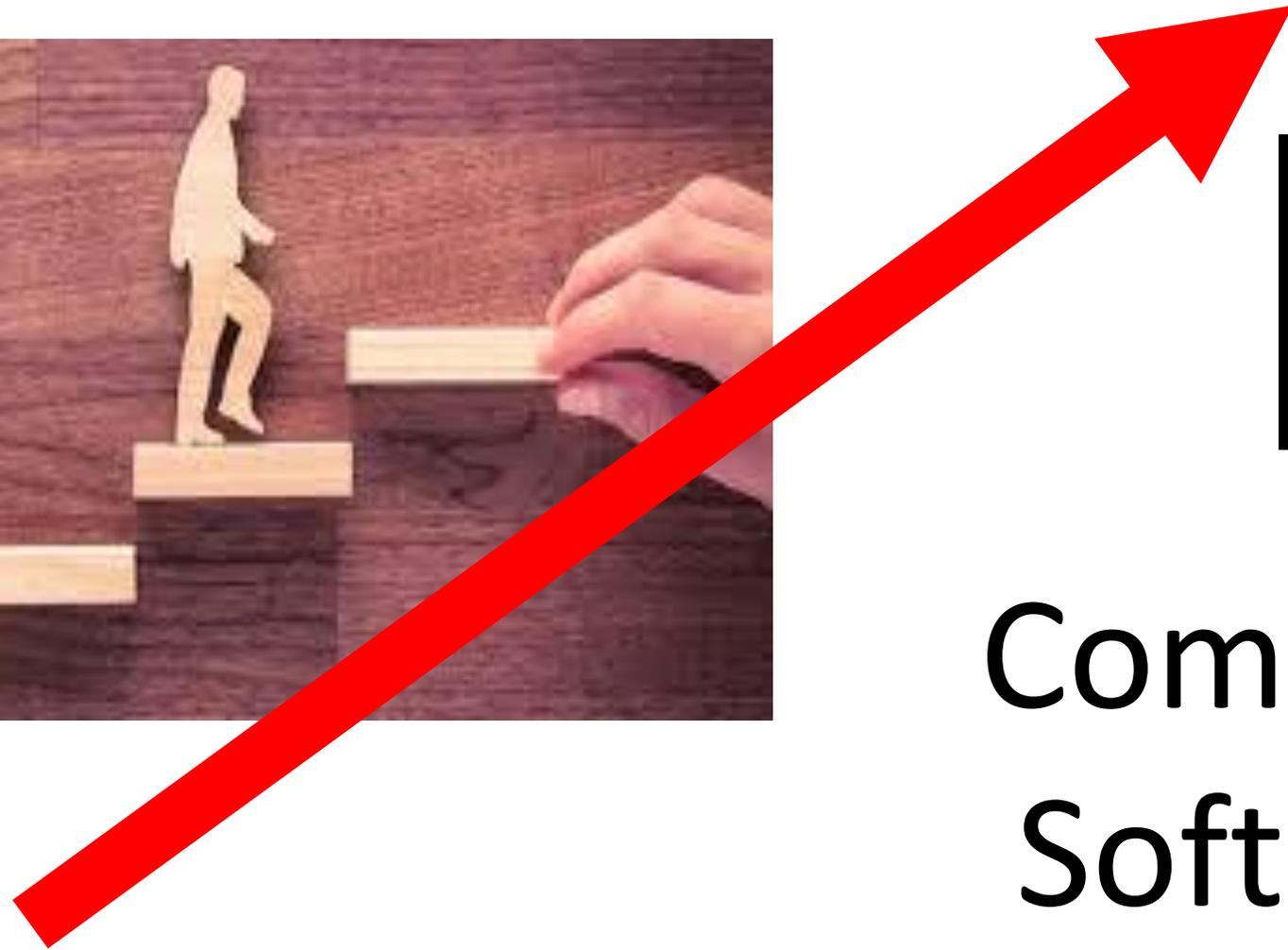
Past
Statistics



Building a new future on



Past
Statistics



Future
Data Science?



Computer
Software
Developments

Purpose build training software

Click Drag
Think

TinkerPlots - [Backpack Report.tp]

File Edit Object Data Window Help

Cards Table Plot Sampler Text Separate Order Stack Ref. Line Divider Hats Counts Averages Label Key

PURPOSE

This report is an example of how students might use TinkerPlots to analyze data and to write a report. It is based loosely on a report written by students at Germantown Academy, available at www.ga.k12.pa.us/Academics/LS/5th/Backpack/Index.htm.

Student with backpack. Photo from Germantown Academy site.

ARE BACKPACKS GIVING US A PAIN?

PURPOSE OF STUDY (From Germantown Academy site)

"A team of fifth and twelfth graders at Germantown Academy have worked together to conduct an important study. We have examined the student backpack weights in relation to student body weights in all three divisions of our school."

"Experts say that for every 20 lbs. of body weight, you should only carry 3 lbs in your backpack. This equals 15%. Other recommendations go as low as 6% of your body weight. Doctors say that heavy loads can cause shoulder pain or lower-back pain."

METHOD

We collected our data in four different grades:

Grade

Seven

Five

Three

One

Backpack

Options

3% 24% 34% 24% 8% 4% 3% 1%

Grade

Five Sev... One Thr...

count, ordered by Gra...

PercentWeight

Graph 2 (at left) shows the percentage of students who are carrying more than the recommended 15% of their weight. Overall, about 40% of students carry too much and are in danger of developing back problems. About 8% carry

Graph 3 (below) makes it easier to see the problem at each grade level. Most students in grades 1 and 3 seem to be ok. But in grade 5, about half of the students are carrying too much, and the majority of 7th graders carry more than the recommended weight for their size.

CONCLUSION

Students in 5th and 7th grade carry more than they should in their backpacks. We think this is because teachers load us with homework assigned from textbooks the size of dictionaries.

Options

14.9606

Grade

Seven

Five

Three

One

PercentWeight





Situation:

A farmer made a statement that last year was a very dry year. He based it on the rainfall figures for the last 5-years:

- 2014: 49 inches
- 2015: 52 inches
- 2016: 46 inches
- 2017: 58 inches
- 2018: 32 inches

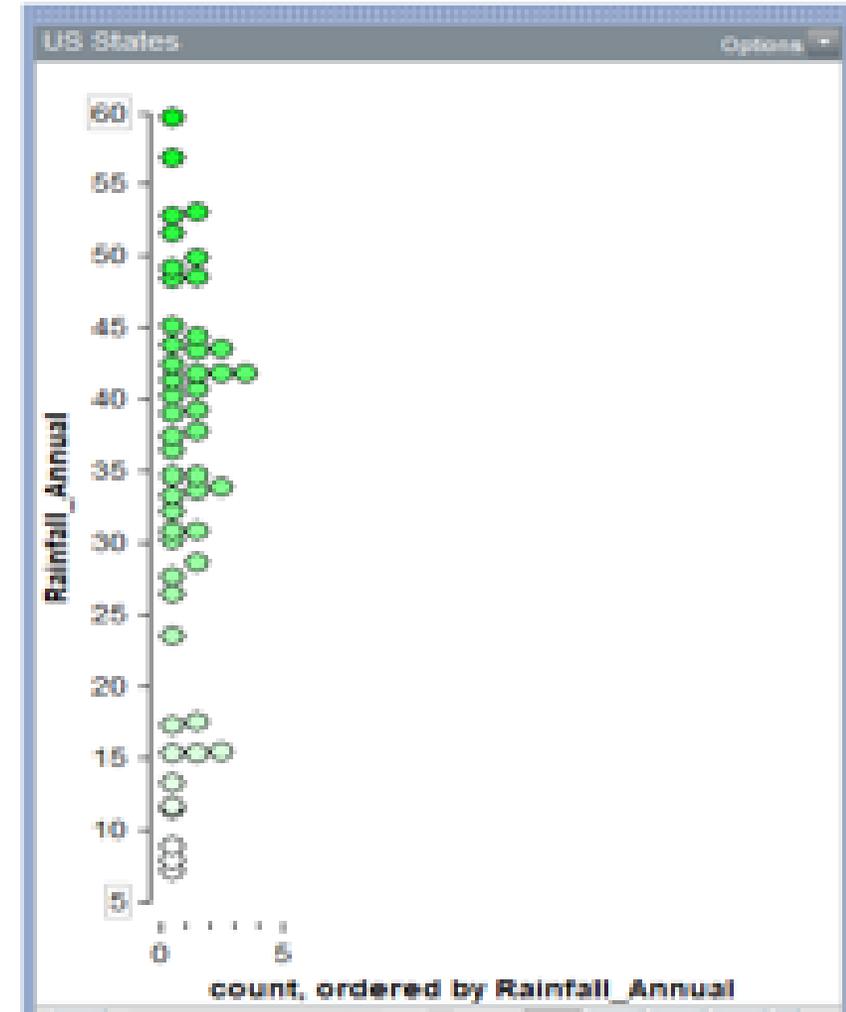
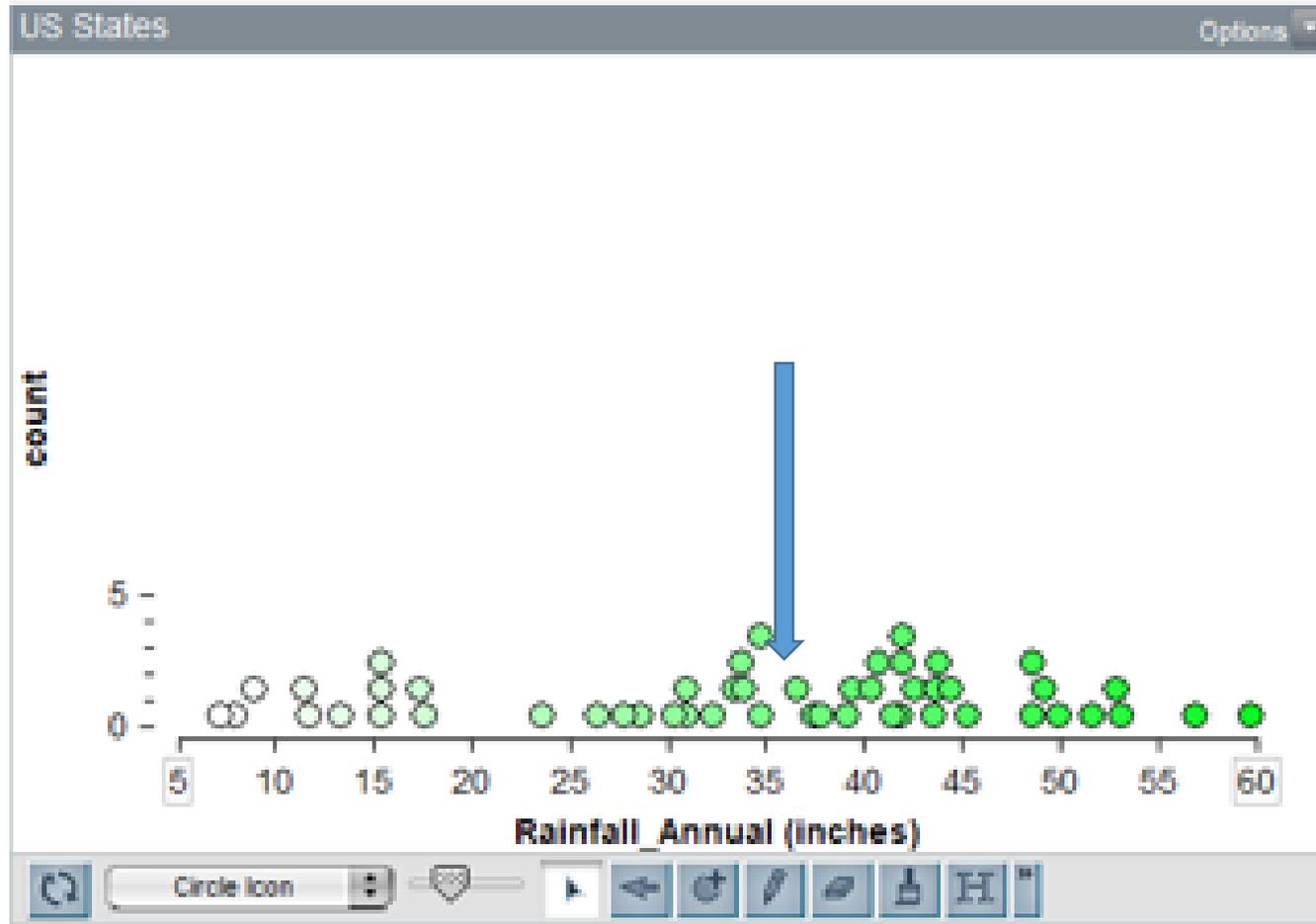
Do you think he has a valid argument?

Let us turn to data to consider his statement:

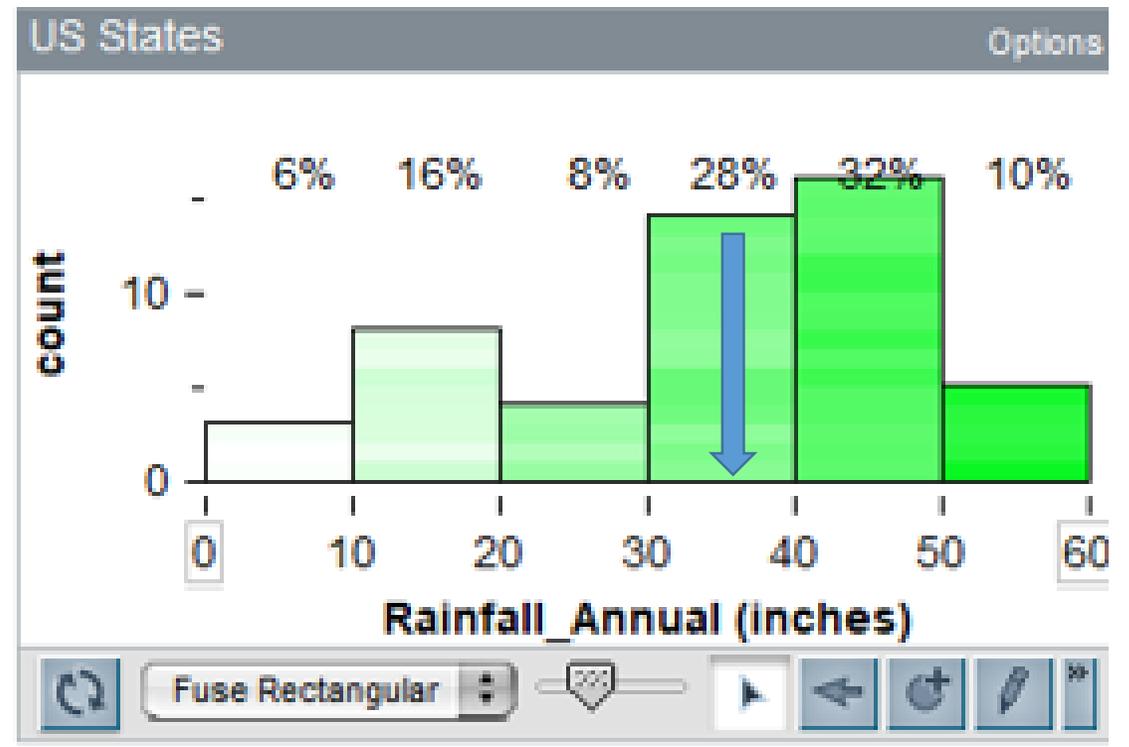
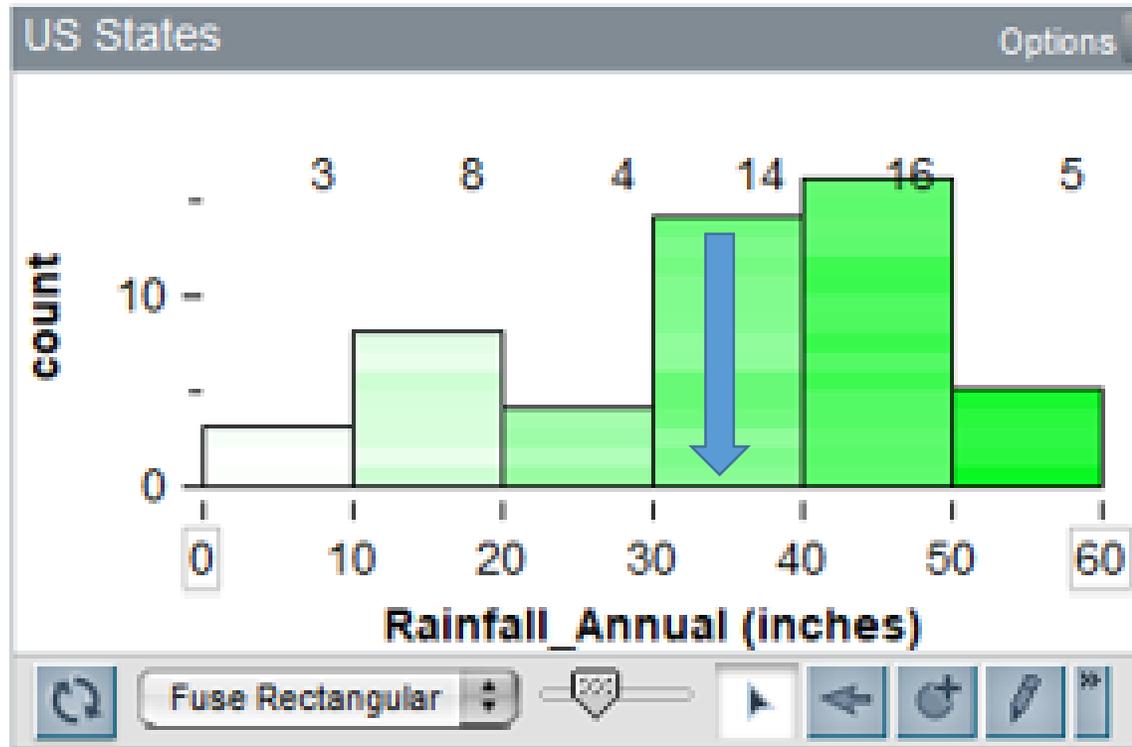
- We collected the data for annual rainfall over a 100-year period for the area
- The data was then placed into the software for analysis
- Various graphical representations were created

See if you agree with the farmer.

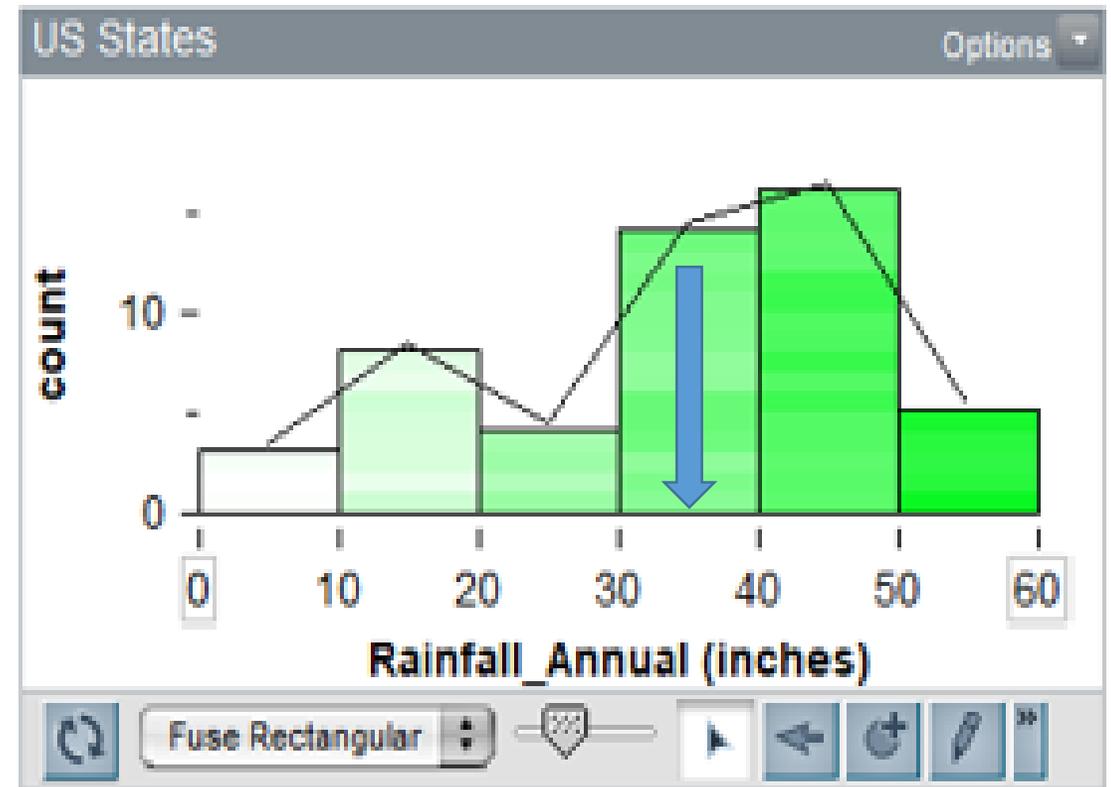
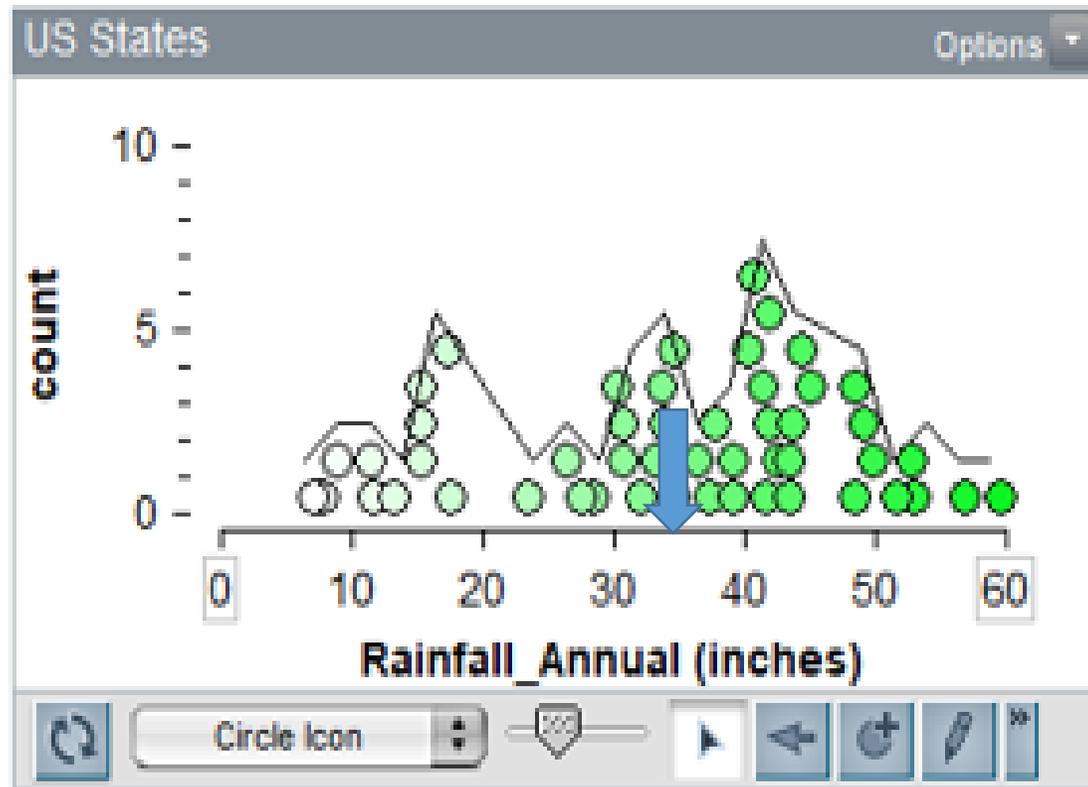
Dot plot (left) and stem plot (right)



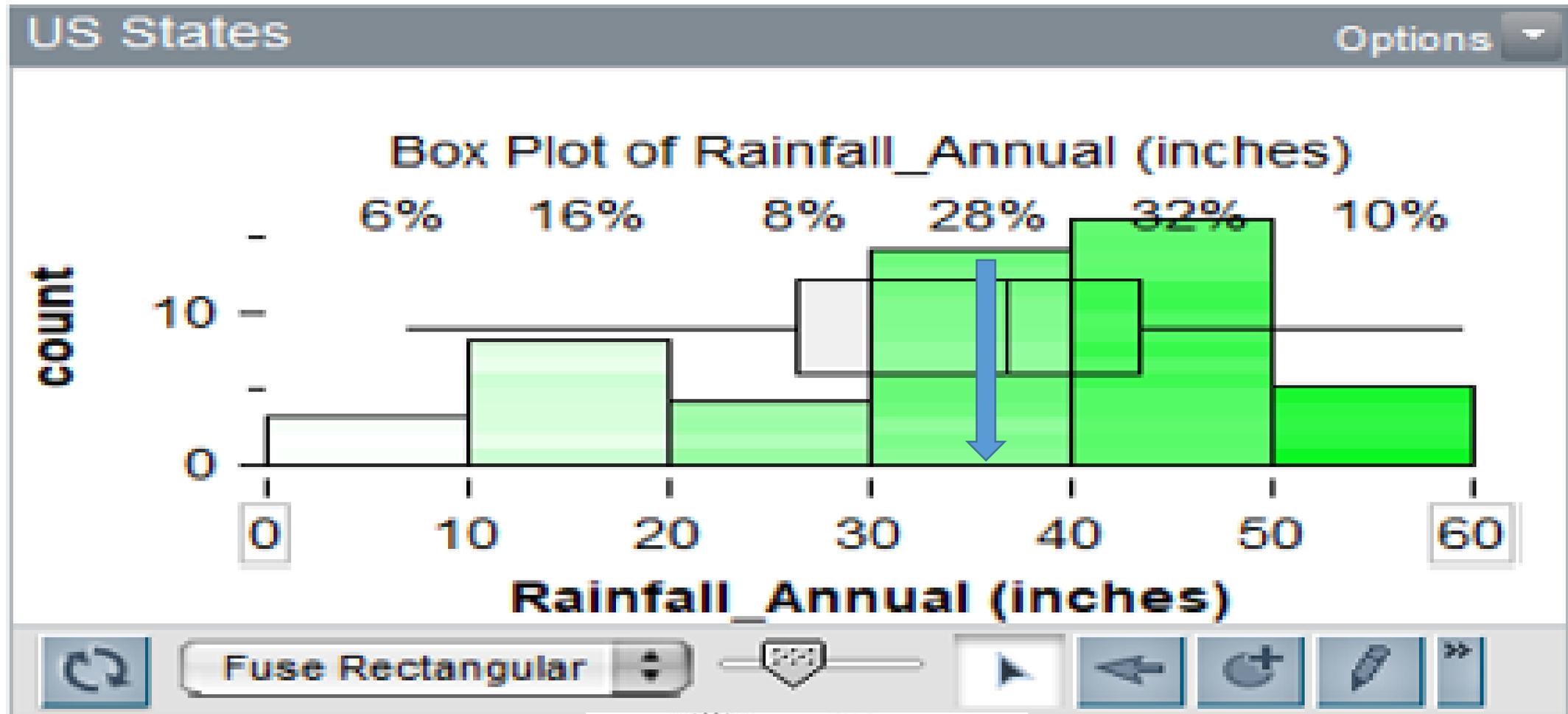
Frequency and relative frequency histogram



Using the "connect stacks" feature



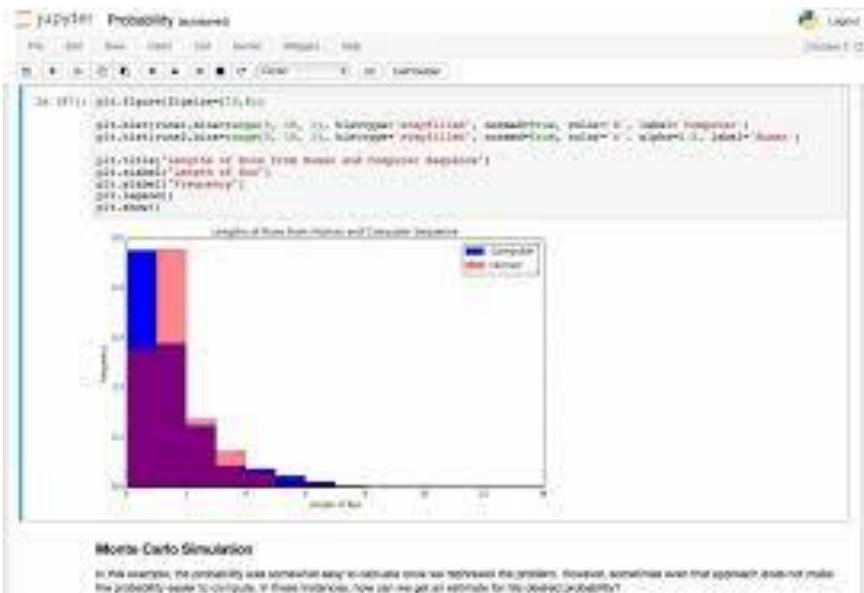
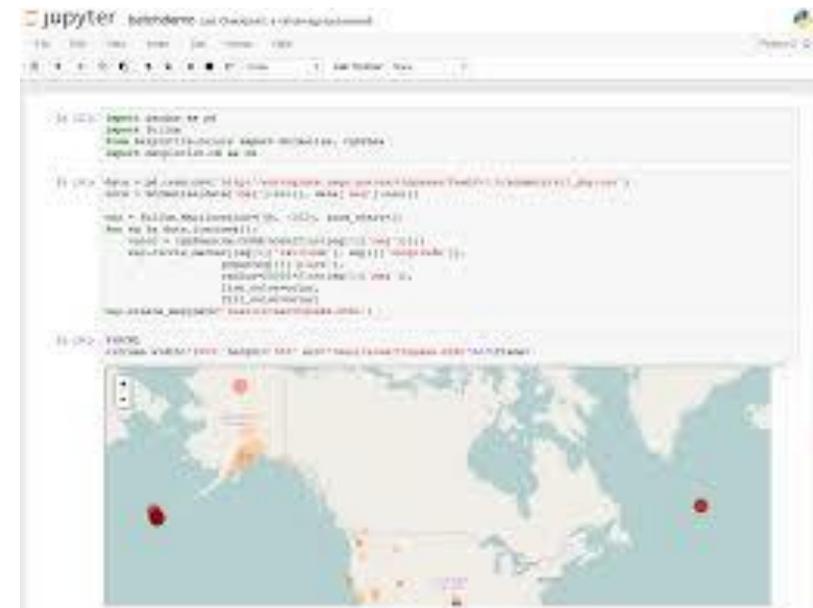
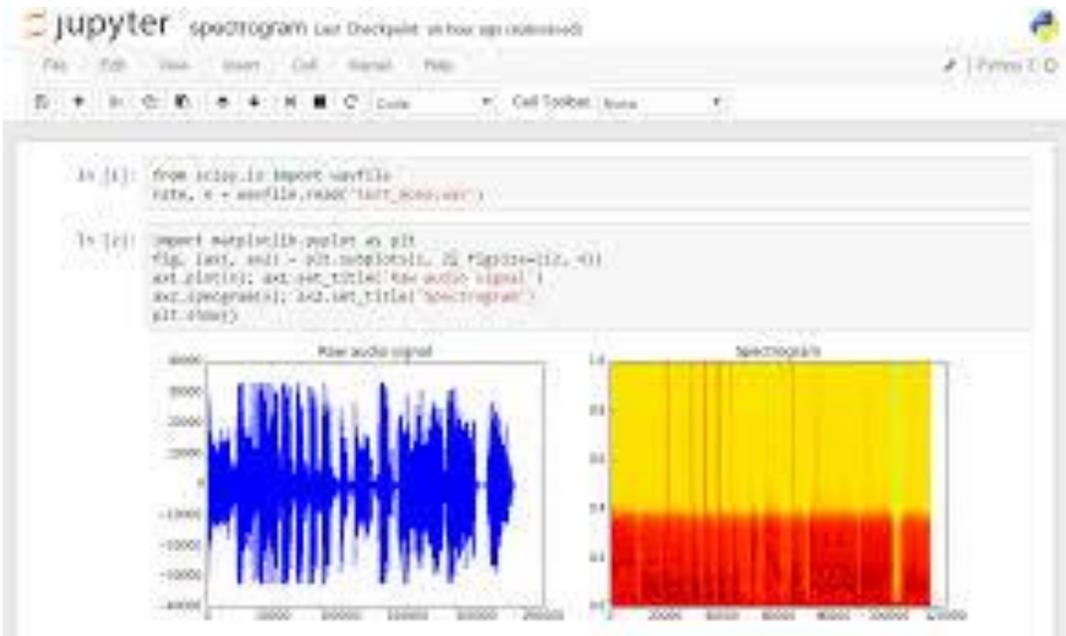
Combined relative frequency histogram and box-and-whisker plot





Power BI





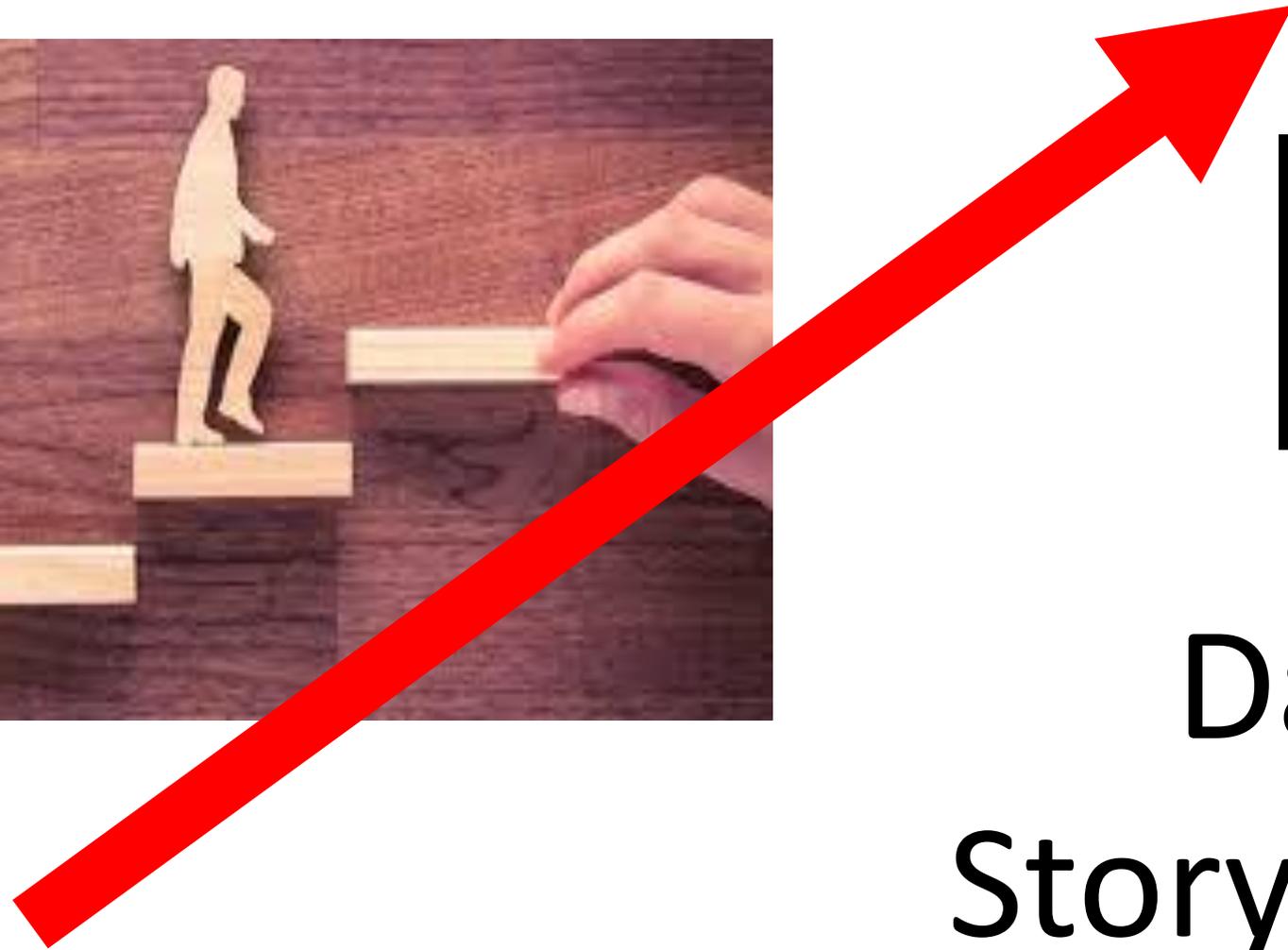
Coding
meets
data



Building a new future on



Past
Statistics

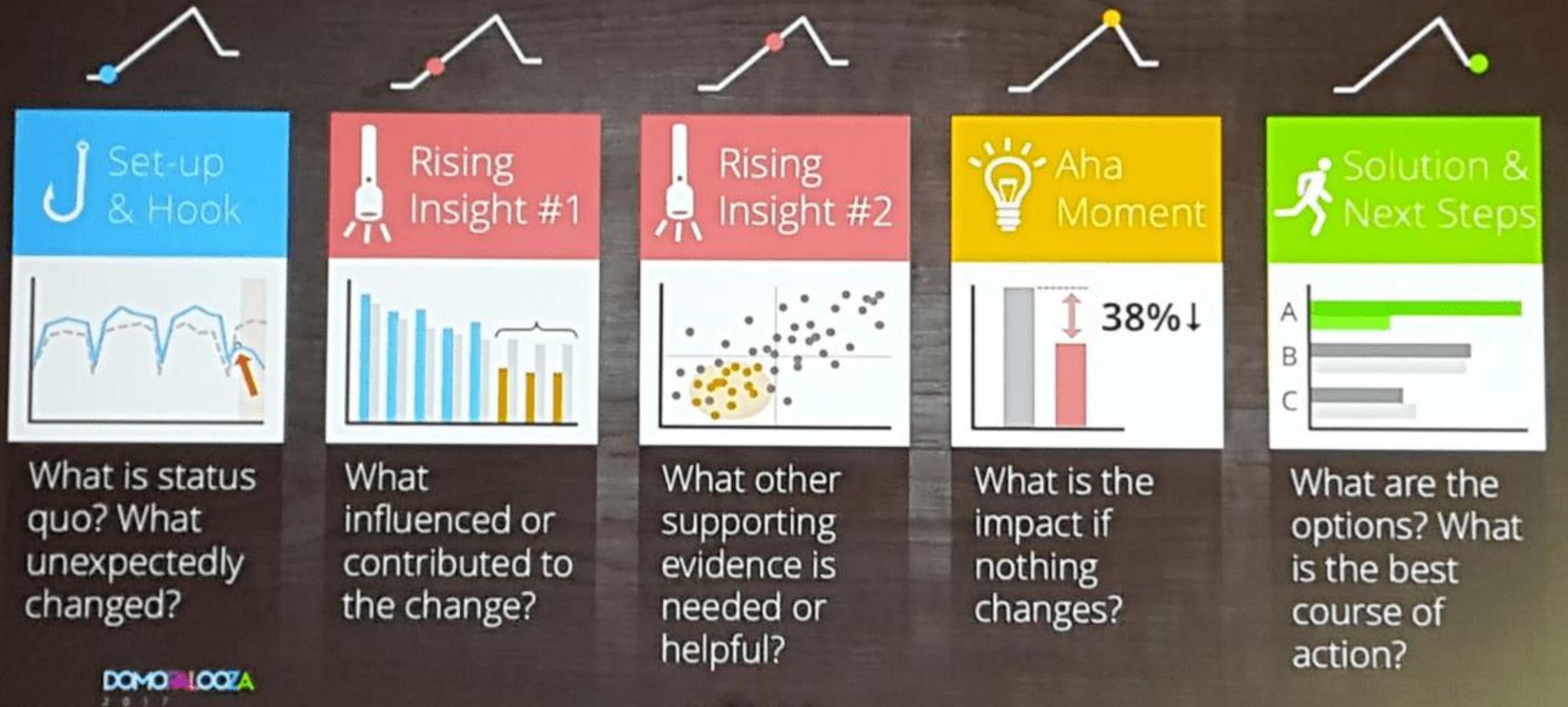


Future
Data Science?

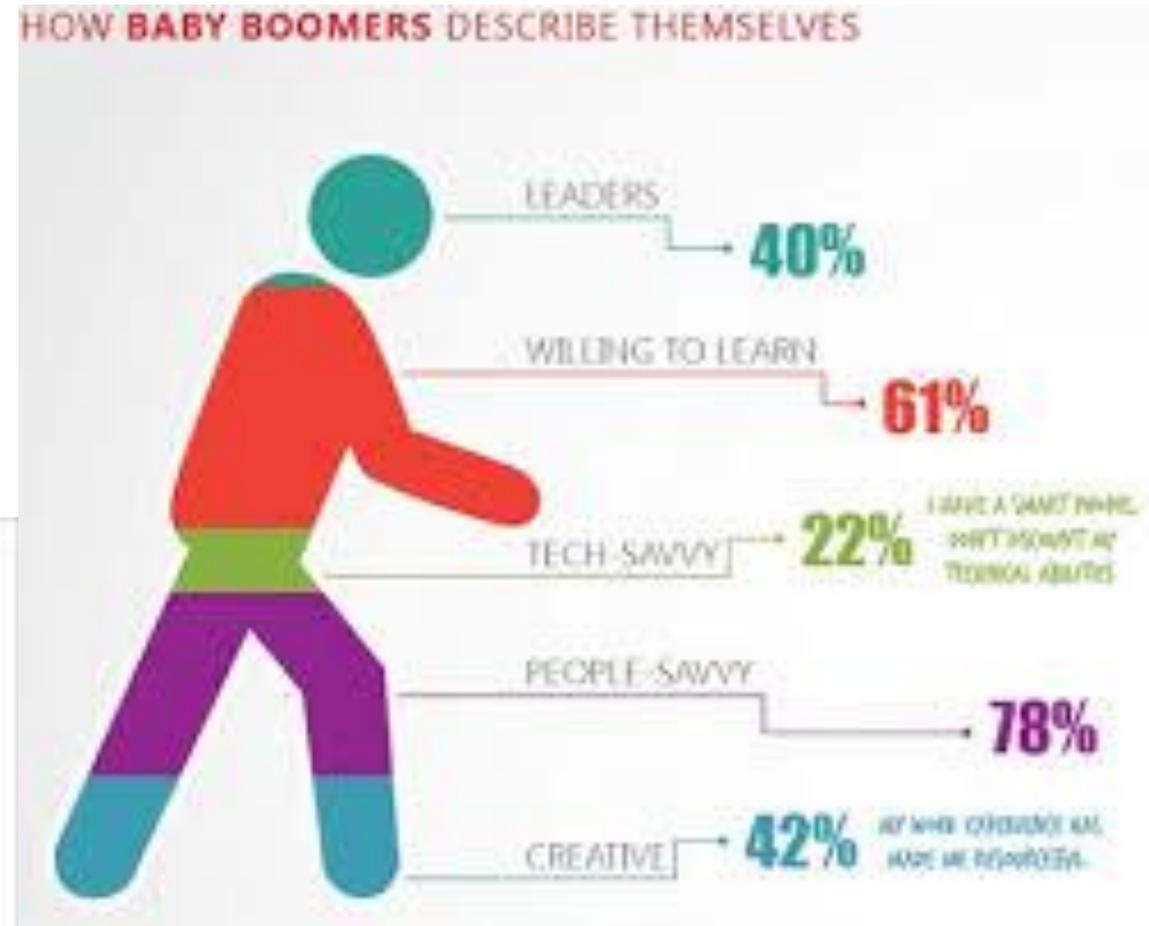


Data
Storytelling

DATA STORYTELLING ARC IN ACTION



Infographics



**Design
meets
Mathematics**

Building a new future on

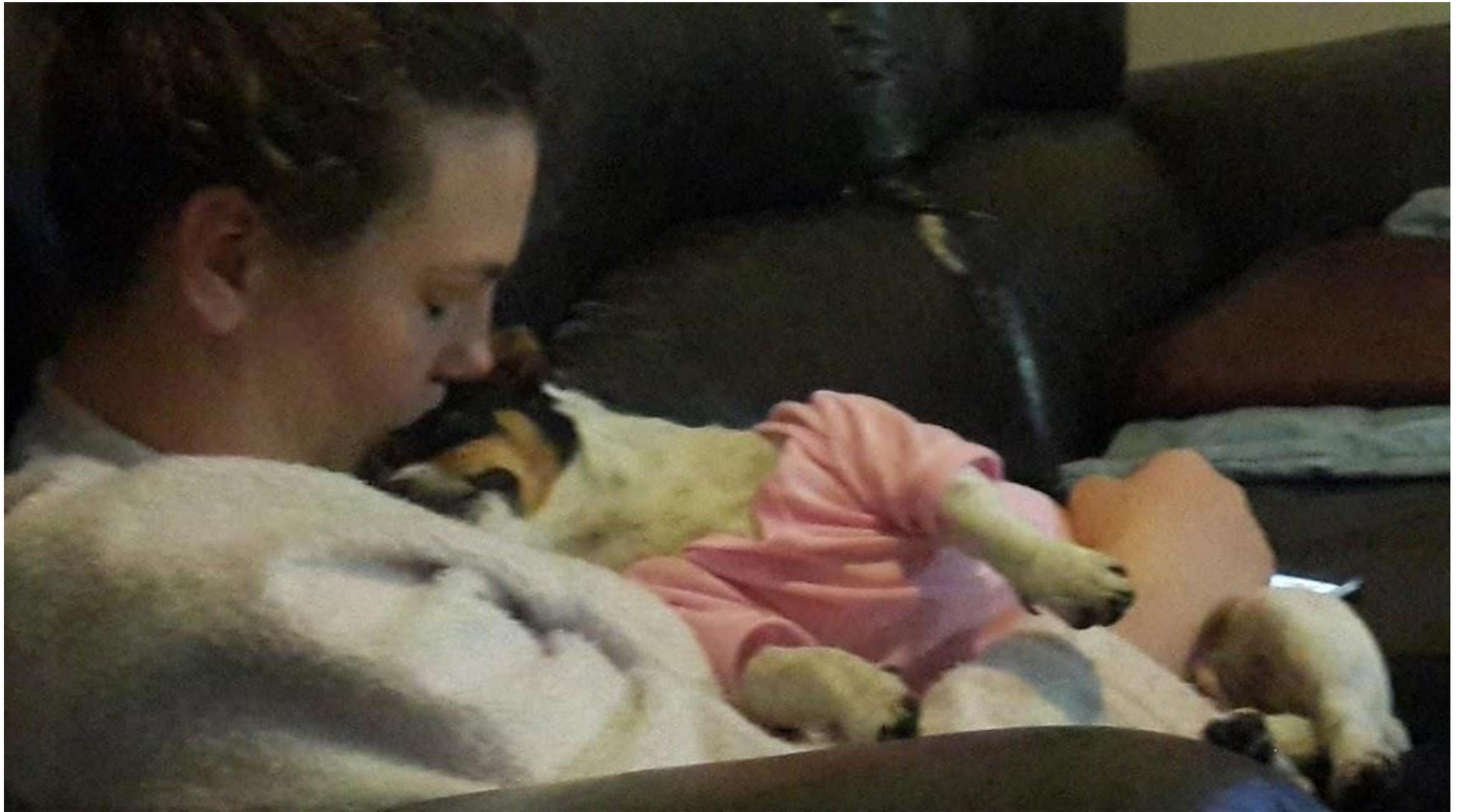


Past
Statistics

Informal Inferential Reasoning:
Modelling, simulation or randomisation

Future
Data Science?



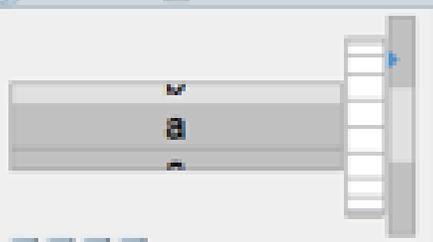


Example of the difference between theoretical and experiential probability

RUN Medium **Options**

Repeat 2 **Attr1**

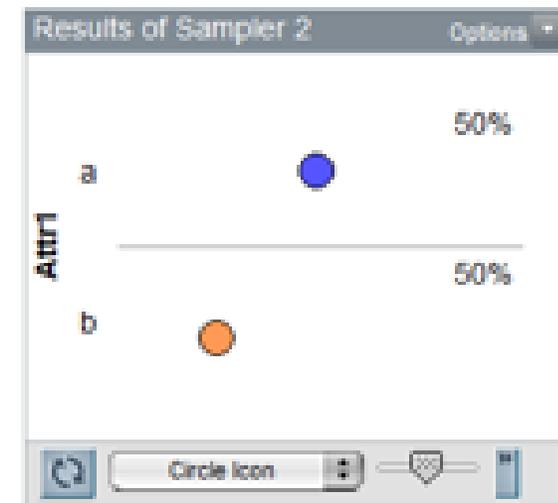
Draw 1



Mixer Stacks Spinner Bars Curve Counter

Results of Sampler 2 **Options**

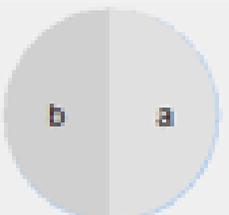
	Attr1	<new>
		
1	a	
2	b	



RUN Fastest **Options**

Repeat 1000 **Attr1**

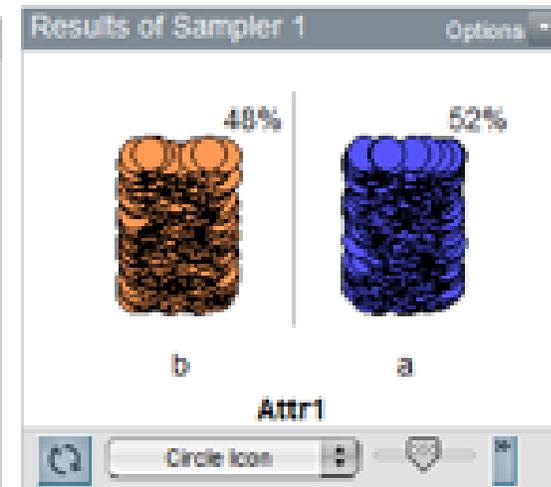
Draw 1



Mixer Stacks Spinner Bars Curve Counter

Results of Sampler 1 **Options**

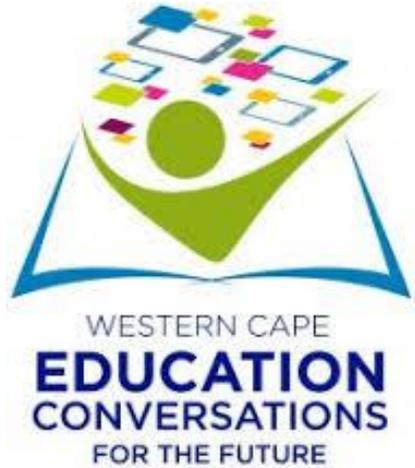
	Attr1	<new>
		
995	a	
996	a	
997	b	
998	a	
999	b	
1000	a	



Modelling, simulation or randomisation can be used as a useful device for inference:

- First, you set up a model of a process in which chance is the only factor influencing the outcomes.
- The second stage is repetition—you repeat the process.
- During the third phase you plot the distribution of the summary statistics in order to determine how likely the actual result or one even more extreme would be.
- Finally, you reach—or infer—a conclusion.
- If the probability of getting a summary statistic as extreme as that from your actual data is small, conclude that chance isn't a reasonable explanation. If the probability isn't small, conclude that you can reasonably attribute the result to chance alone.

Building a new future on

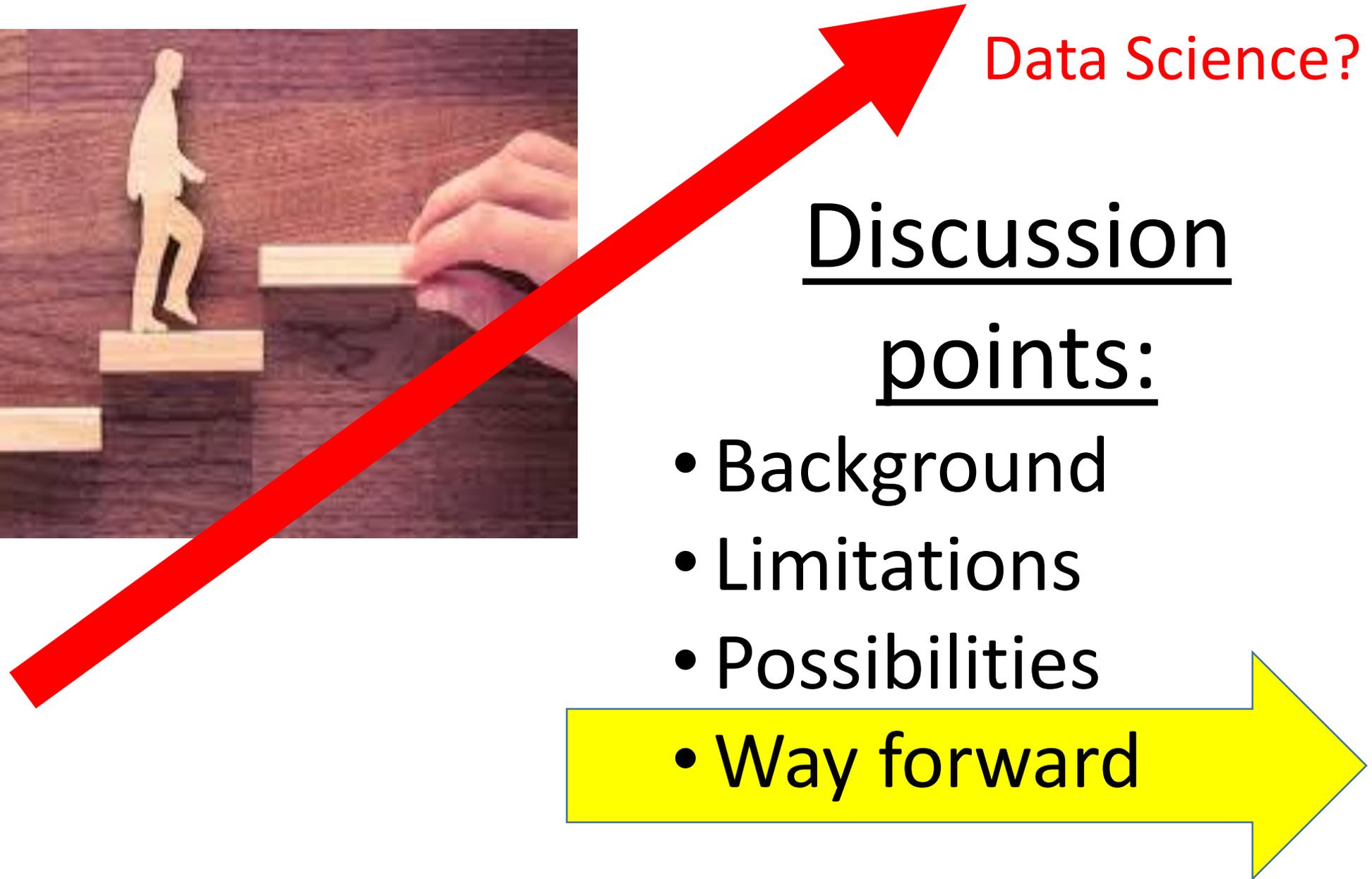


Past
Statistics

Future
Data Science?

Discussion points:

- Background
- Limitations
- Possibilities
- Way forward



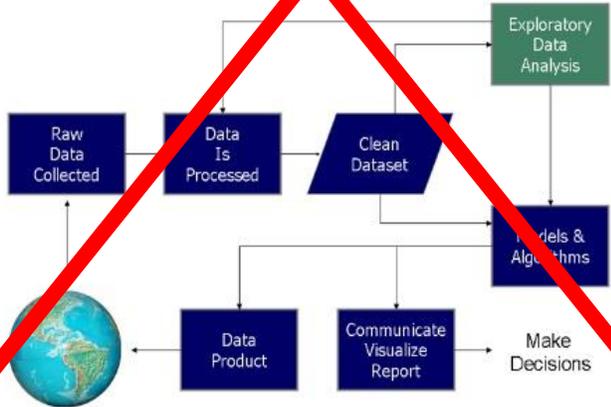
Data Science: process

Developing Data Detectives to become Data Scientists

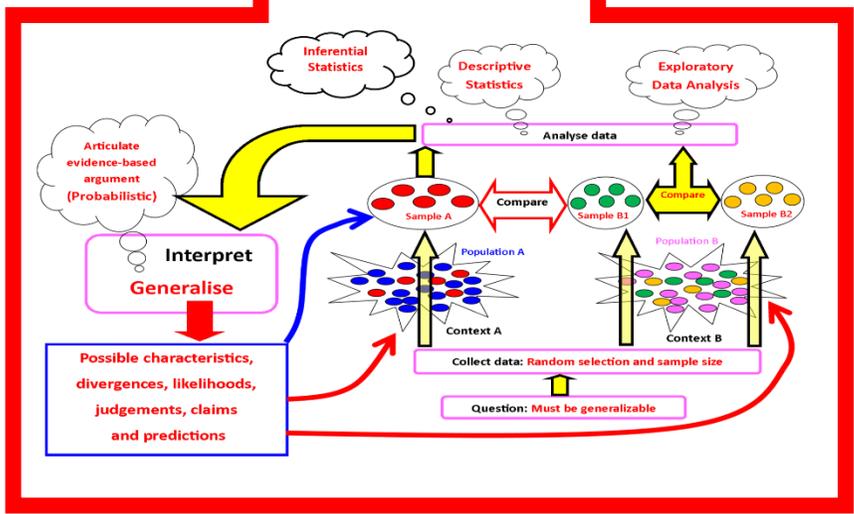
Future: A standalone subject?
Data Science as a subject on
it's own.

OR

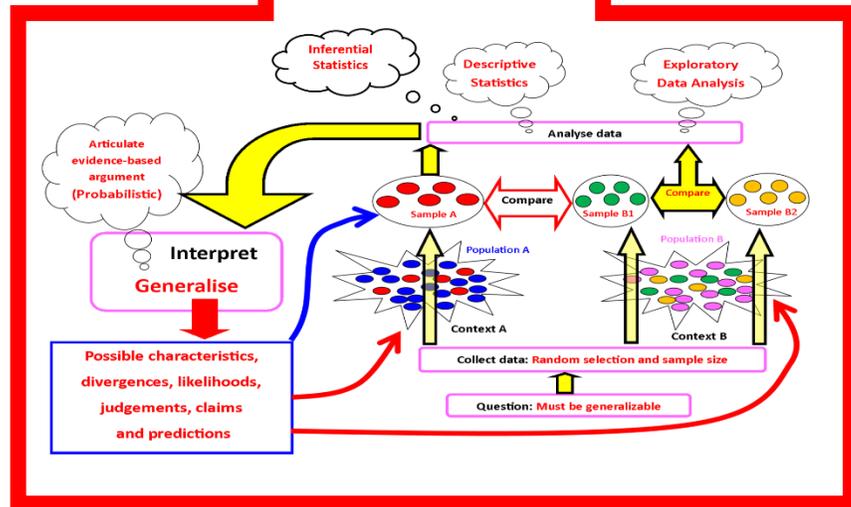
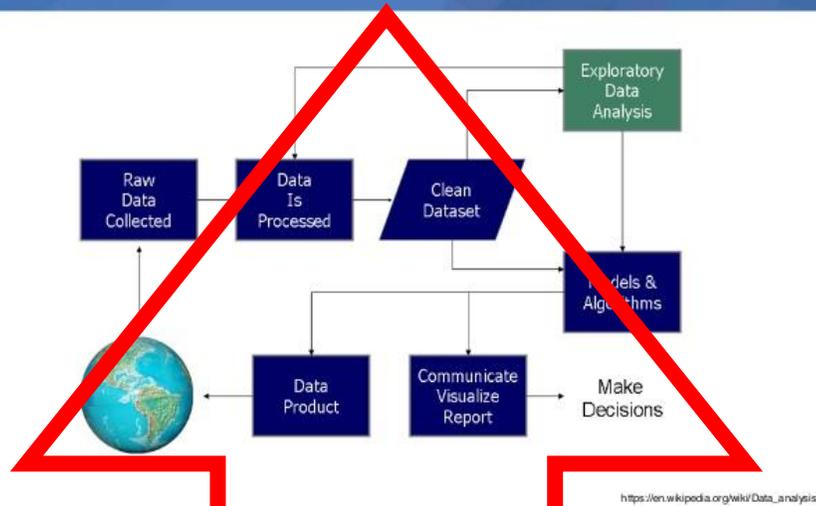
Integrated with subjects
where measurement are
made.



https://en.wikipedia.org/wiki/Data_analysis



Data Science: process



Developing Data Detectives to become Data Scientists

Development focus:

- Informal inferential reasoning
- Modeling, simulation and randomisation
- Probabilistic reasoning
- Use of technology
- Design and data storytelling
- Coding and data



**Education Conversations
of the Future**
Quality Future Focussed Education



*Thank
you*

A close-up illustration of a fountain pen nib, showing the gold-colored metal and the black barrel.

Herholdt Bezuidenhout
bezuidenhouth@hsgrabouw.co.za

Dr. Erna Lampen
ernalampen@sun.ac.za



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY