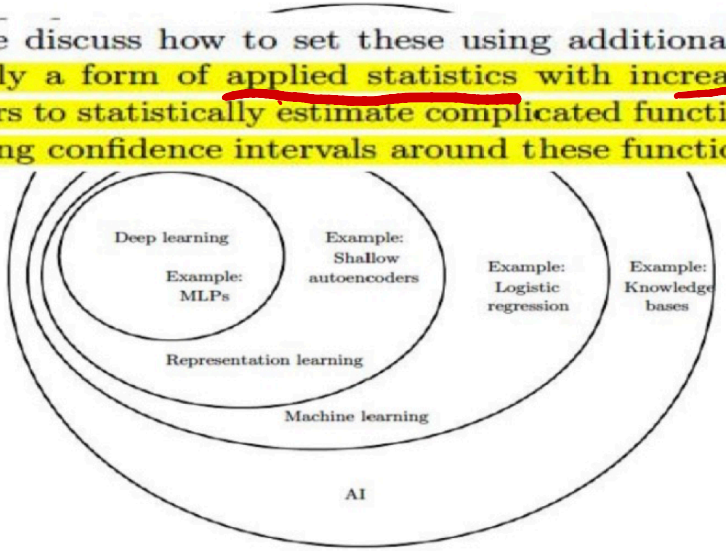
Theme 0

① WHAT IS IT ABOUT

STATISTICS, MACHINE LEARNING, PROBABILITY

- LEARNING FROM DATA
  - → PREDICTION ←
  - → INFERENCE
  - → CAUSALITY
  - → UNCERTAINTY QUANTIFICATION ←

itself; we discuss how to set these using additional data. Machine learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions; we therefore present the



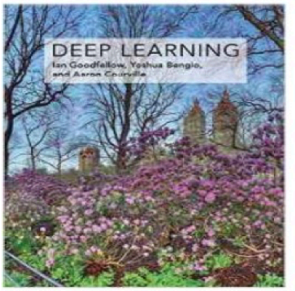


Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

1. DATA VIZ & SUMMARIZATION
2. FROM RANDOMIZATION TO RANDOMNESS
3. WHAT IS PROBABILITY & WHAT IS USEFUL FOR
4. THE CALCULUS OF PROBABILITY
5. MORE MODELS FOR MORE DATA
6. JOINT DISTRIBUTIONS, INDEPENDENCE, MAXIMUM LIKELIHOOD
7. EXPECTATION
8. CONCENTRATION, INEQUALITIES, LIMIT THEOREMS
9. STATISTICAL LEARNING

REFERENCES

- LECTURE NOTES
- ART OF STATISTICS (BOOK)
- VARIOUS ONLINE SOURCES (DATA)
- WIKI

APPROACH

- START FROM DATA & QUESTIONS
- ONLY TEACH USEFUL THINGS
- COLOUR CODING
  - FUNDAMENTAL RESULT IN STATS
  - FUNDAMENTAL RESULT IN MATH

EVALUATION

- 1 / 2 / 3 LEVELS
- SEE BLACKBOARD

THEME 1 : DATA VIZ & SUMMARIZATION

- SHIPMAN'S DEAD PATIENTS
- VOX POPULI
- BAR & BOX PLOT
- MEANS & QUANTILES
- SCATTER PLOT & CORRELATION

SIMPLEST TYPE OF DATA : BINARY

BINARY

→ SUMMARY OF BINARY DATA : FREQUENCY

A.S

Event	Percentage in 10,267 people allocated placebo	Percentage in 10,269 people allocated statin	% (relative) risk reduction in those allocated statins
Heart attack	11.8	8.7	27%
Stroke	5.7	4.3	25%
Death from any cause	14.7	12.9	13%

Table 4.1
The outcomes at five years in the Heart Protection Study, according to treatments allocated to patients. The absolute reduction in the risk of a heart attack was 11.8 − 8.7 = 3.1%. So out of 1,000 people taking a statin, around 31 heart attacks were prevented – this means that around 30 people had to take a statin for five years to prevent one heart attack.

SUMMARIZATION (AT THIS POINT) FOR BINARY

- FREQUENCY :

  NUMBER OF OBSERVATIONS IN EACH CATEGORY

- PERCENTAGE = FREQUENCY / TOTAL

VISUALIZATIONS OF SUMMARIES OF BINARY DATA

- ALREADY SEEN: TABLE OF FREQ's / PERC's
- INFOGRAMS

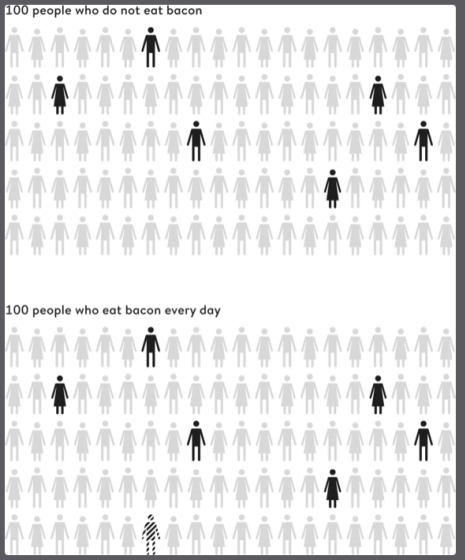
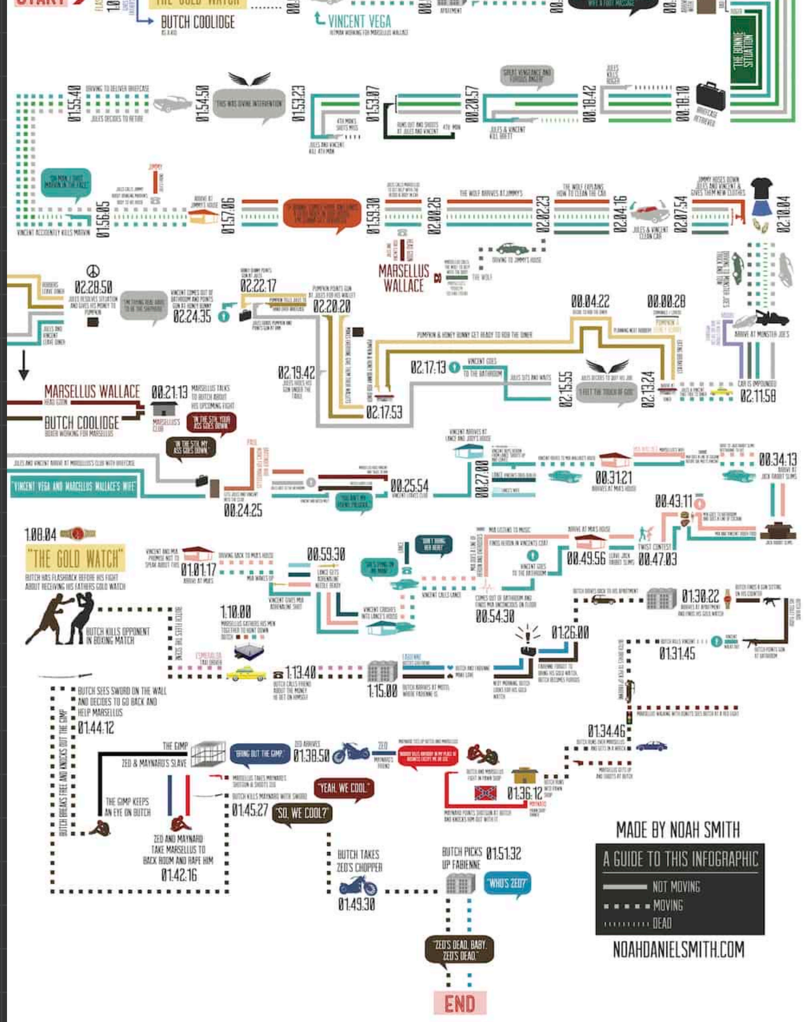


Figure 1.4
Bacon sandwich example using a pair of icon arrays, with randomly scattered icons showing the incremental risk of eating bacon every day. Of 100 people who do not eat bacon, 6 (solid icons) develop bowel cancer in the normal run of events. Of 100 people who eat bacon every day of their lives, there is 1 additional (striped) case.[fn5]

A.S

TWO BINARY VAR'S :


START
BUTCH COOLIDGE
VINCENT VEGA
THE BONNIE SITUATION
"THIS WAS DIVINE INTERVENTION"
01:55:40
01:54:50
01:53:23
01:53:07
00:20:57
00:18:42
00:18:10
01:56:05
01:57:06
01:59:30
02:00:26
02:02:23
02:04:16
02:07:54
02:10:04
MARSELLUS WALLACE
02:28:50
02:24:35
02:22:17
02:20:20
02:19:42
02:17:53
02:17:13
02:15:55
02:13:24
02:11:50
00:04:22
00:00:20
"I FELT THE TOUCH OF GOD"
MARSELLUS WALLACE
00:21:13
BUTCH COOLIDGE
"VINCENT VEGA AND MARCELLUS WALLACE'S WIFE"
00:24:25
00:25:54
00:27:00
00:31:21
00:34:13
00:43:11
1:00:04
"THE GOLD WATCH"
BUTCH KILLS OPPONENT IN BOXING MATCH
01:01:17
00:59:30
00:49:56
00:47:03
00:54:30
1:10:00
01:26:00
01:30:22
01:31:45
1:13:40
1:15:00
BUTCH SEES SWORD ON THE WALL AND DECIDES TO GO BACK AND HELP MARSELLUS
01:44:12
THE GIMP
ZED & MAYNARD'S SLAVE
"BRING OUT THE GIMP"
01:38:50
01:34:46
"YEAH, WE COOL"
01:45:27
"SO, WE COOL?"
01:36:12
THE GIMP KEEPS AN EYE ON BUTCH
ZED AND MAYNARD TAKE MARSELLUS TO BACK ROOM AND RAPE HIM
01:42:16
BUTCH TAKES ZED'S CHOPPER
01:49:30
BUTCH PICKS UP FABIENNE
01:51:32
"WHO'S ZED?"
"ZED'S DEAD, BABY. ZED'S DEAD."
END
MADE BY NOAH SMITH
A GUIDE TO THIS INFOGRAPHIC
NOT MOVING
MOVING
DEAD
NOAHDANIELSMITH.COM

CATEGORICAL DATA

Event	Percentage in 10,267 people allocated placebo	Percentage in 10,269 people allocated statin	% (relative) risk reduction in those allocated statins
Heart attack	11.8	8.7	27%
Stroke	5.7	4.3	25%
Death from any cause	14.7	12.9	13%

Table 4.1
The outcomes at five years in the Heart Protection Study, according to treatments allocated to patients. The absolute reduction in the risk of a heart attack was 11.8 − 8.7 = 3.1%. So out of 1,000 people taking a statin, around 31 heart attacks were prevented – this means that around 30 people had to take a statin for five years to prevent one heart attack.

BINARY IS CATEGORICAL WITH TWO LEVELS

HERE ALSO SEE CATEGORICAL WITH 4 LEVELS