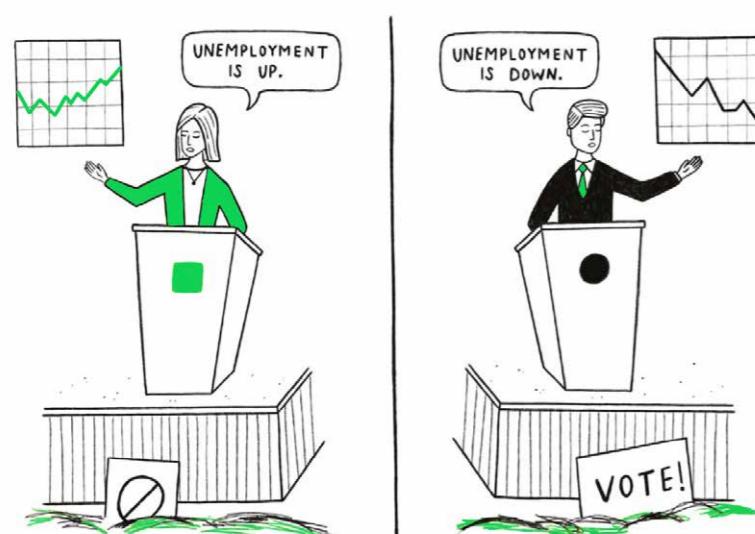


# Data Fallacies to Avoid



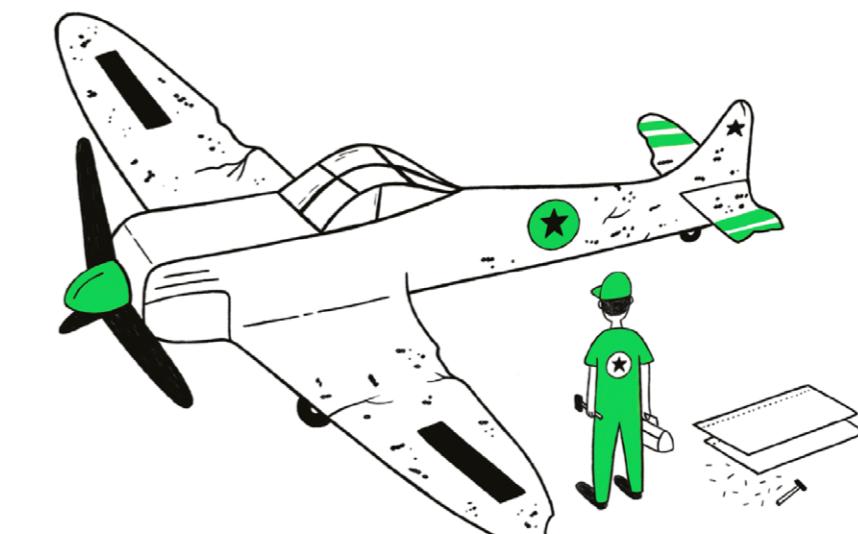
## Cherry Picking

Selecting results that fit your claim and excluding those that don't.



## Data Dredging

Repeatedly testing new hypotheses against the same set of data, failing to acknowledge that most correlations will be the result of chance.



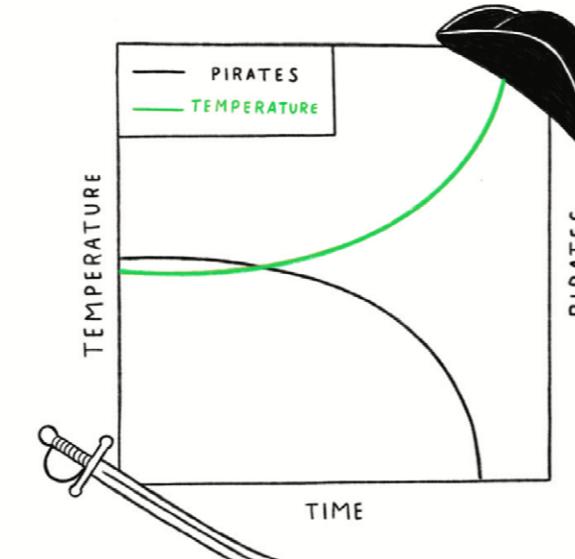
## Survivorship Bias

Drawing conclusions from an incomplete set of data, because that data has 'survived' some selection criteria.



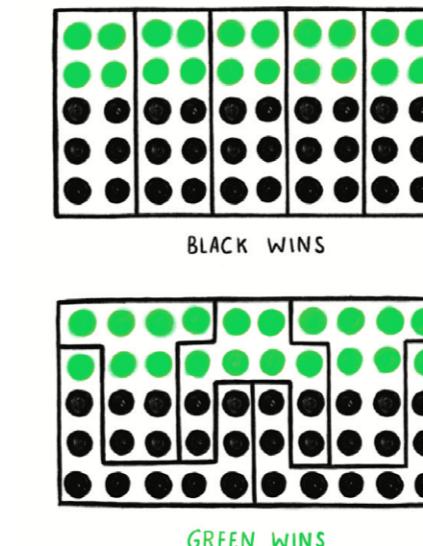
## Cobra Effect

Setting an incentive that accidentally produces the opposite result to the one intended. Also known as a Perverse Incentive.



## False Causality

Falsely assuming when two events appear related that one must have caused the other.



## Gerrymandering

Manipulating the geographical boundaries used to group data in order to change the result.



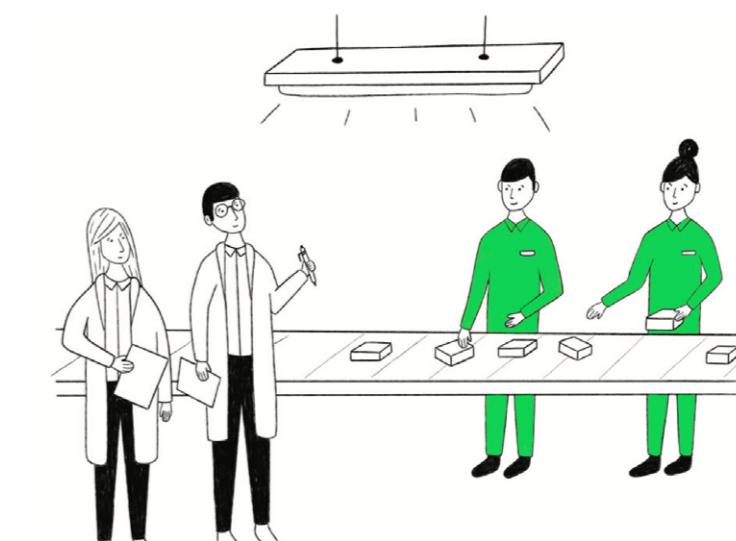
## Sampling Bias

Drawing conclusions from a set of data that isn't representative of the population you're trying to understand.



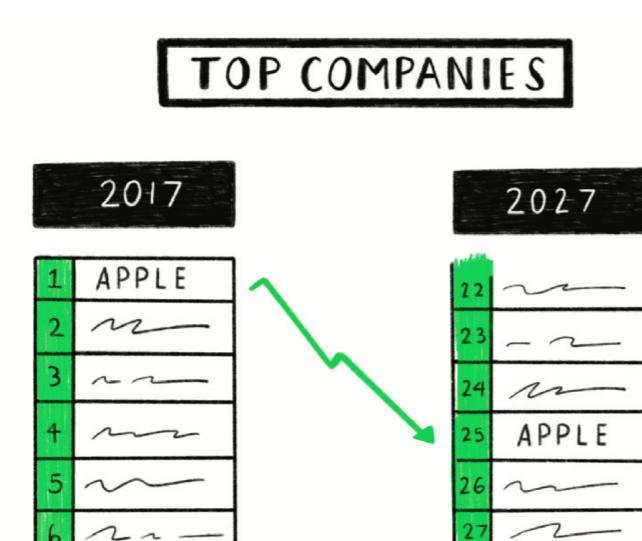
## Gambler's Fallacy

Mistakenly believing that because something has happened more frequently than usual, it's now less likely to happen in future (and vice versa).



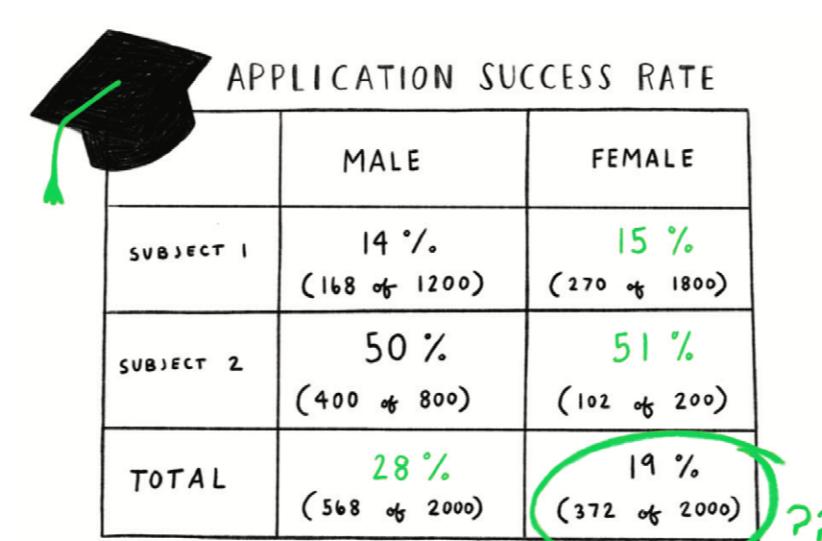
## Hawthorne Effect

The act of monitoring someone can affect their behaviour, leading to spurious findings. Also known as the Observer Effect.



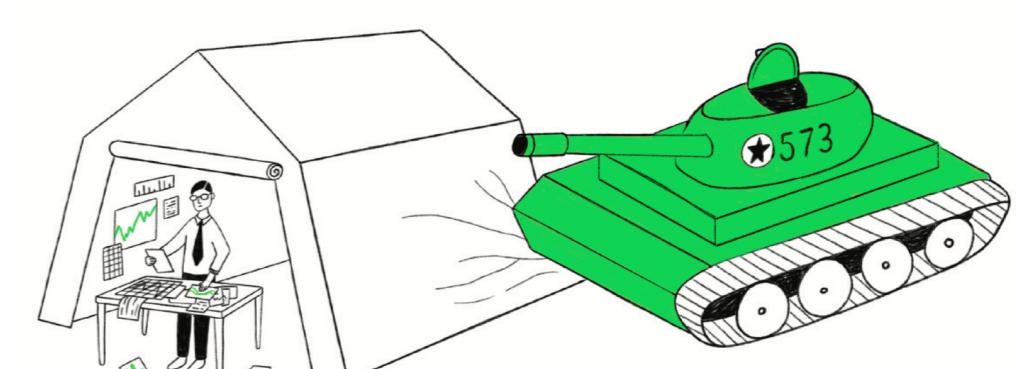
## Regression Towards the Mean

When something happens that's unusually good or bad, it will revert back towards the average over time.



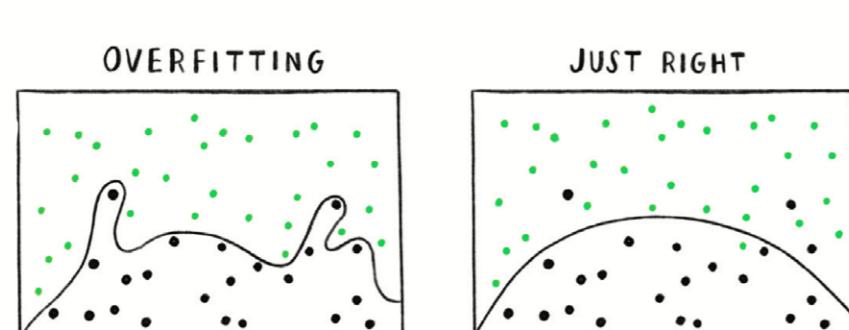
## Simpson's Paradox

When a trend appears in different subsets of data but disappears or reverses when the groups are combined.



## McNamara Fallacy

Relying solely on metrics in complex situations and losing sight of the bigger picture.



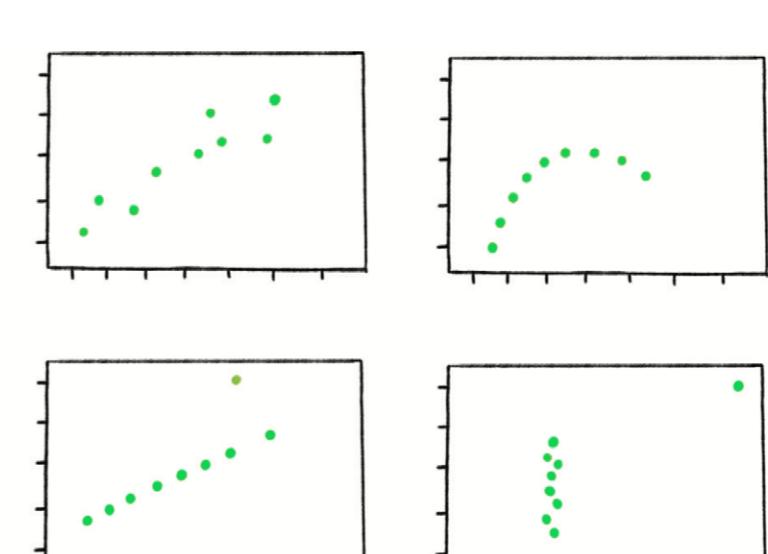
## Overfitting

Creating a model that's overly tailored to the data you have and not representative of the general trend.



## Publication Bias

Interesting research findings are more likely to be published, distorting our impression of reality.



## Danger of Summary Metrics

Only looking at summary metrics and missing big differences in the raw data.