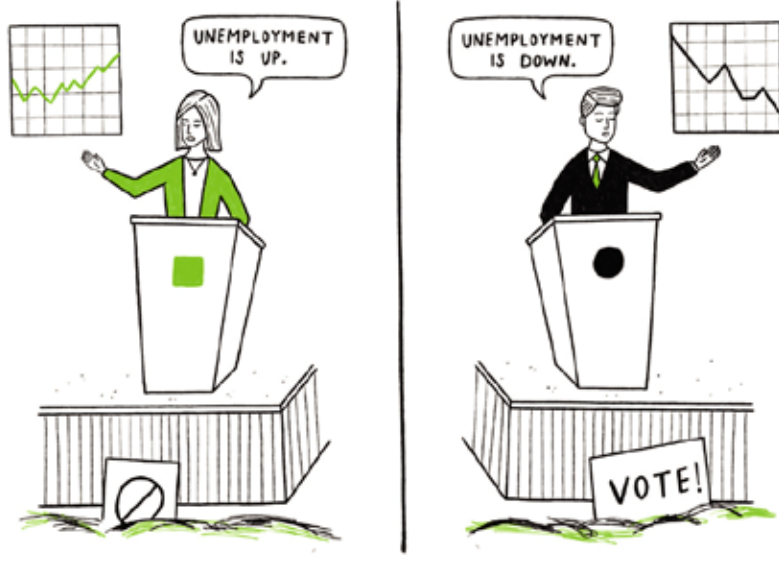


15 Datenirrtümer, die du vermeiden solltest



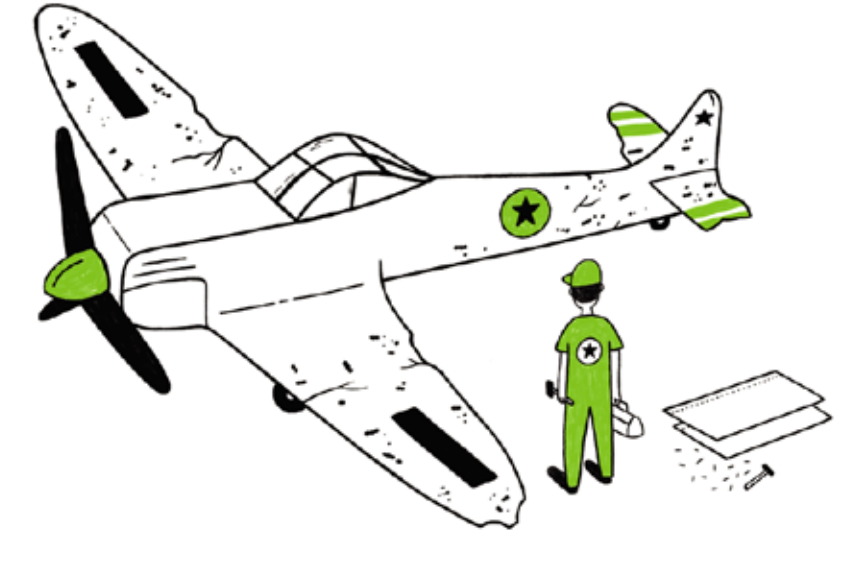
Cherry Picking

Auswahl von Ergebnissen, die zu deiner Behauptung passen und Ausschluss derer, die nicht passen.



Datenfischen

Wiederholtes Testen neuer Hypothesen gegen die gleichen Daten, wobei nicht berücksichtigt wird, dass die meisten Korrelationen zufällig entstehen.



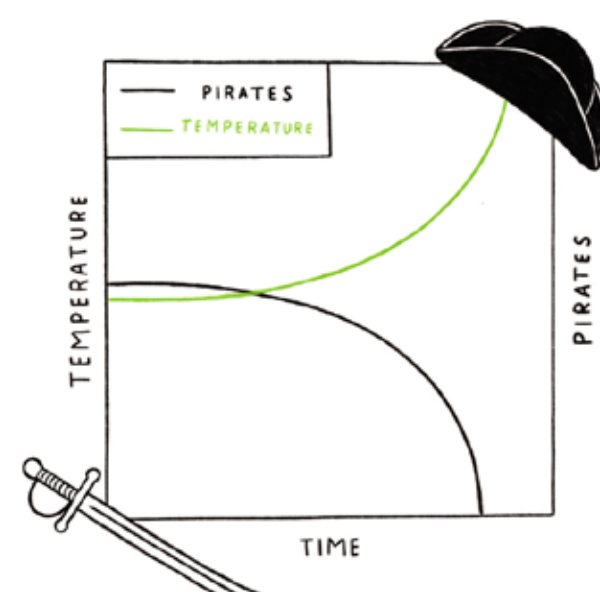
Survivorship-Bias

Das Ziehen von Schlussfolgerungen aus einem unvollständigen Datensatz, weil diese Daten einige Auswahlkriterien "überlebt" haben.



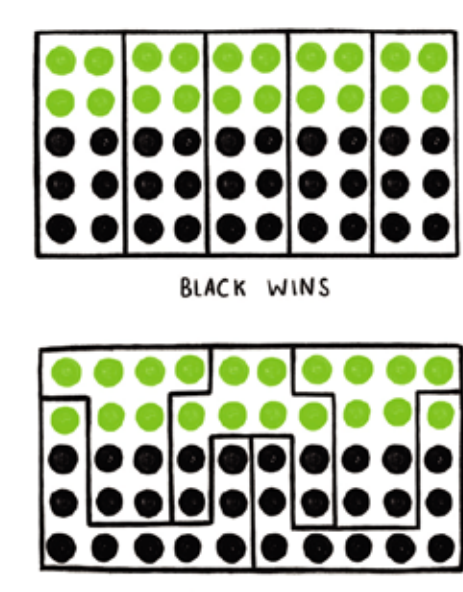
Kobra-Effekt

Setzen eines Anreizes, der versehentlich das gegenteilige Ergebnis erzeugt. Wird auch als Fehlanreiz bezeichnet.



Falsche Kausalität

Die fälschliche Annahme, dass wenn zwei Ereignisse zusammenhängend erscheinen, das eine das andere verursacht haben muss.



Gerrymandering

Manipulation der geografischen Grenzen, die zur Gruppierung von Daten verwendet werden, um das Ergebnis zu verändern.



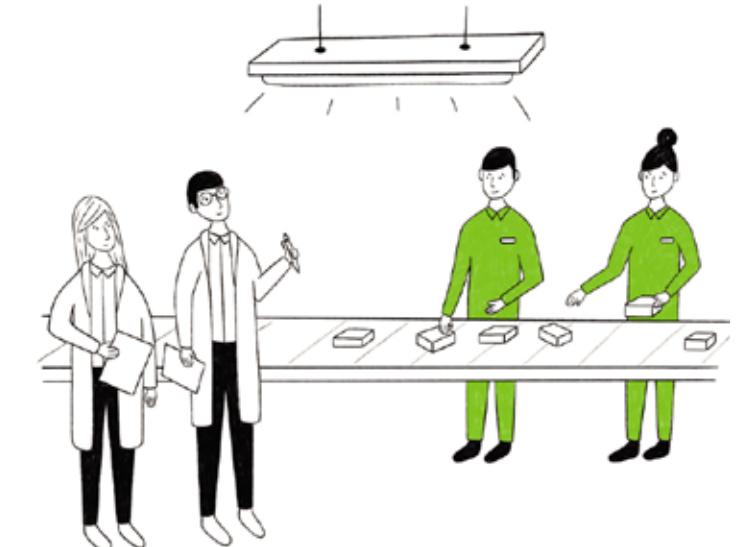
Stichproben-Bias

Ziehen von Schlussfolgerungen aus einem Datensatz, der nicht repräsentativ für die Grundgesamtheit ist, die man zu verstehen versucht.



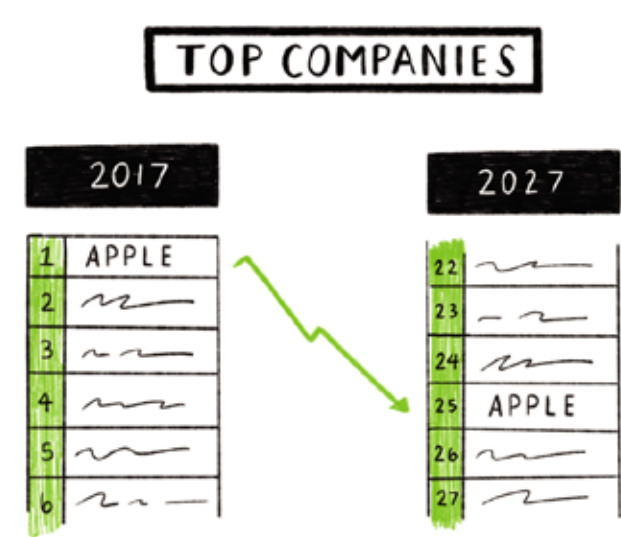
Spilerfehlschluss

Irrglaube, wenn etwas häufiger als gewöhnlich passiert, es nun weniger wahrscheinlich ist, dass es in Zukunft passiert (und umgekehrt).



Hawthorne-Effekt

Der Akt des Beobachtens einer Person kann deren Verhalten beeinflussen, was zu beeinflussten Ergebnissen führt. Auch als Beobachtereffekt bekannt.



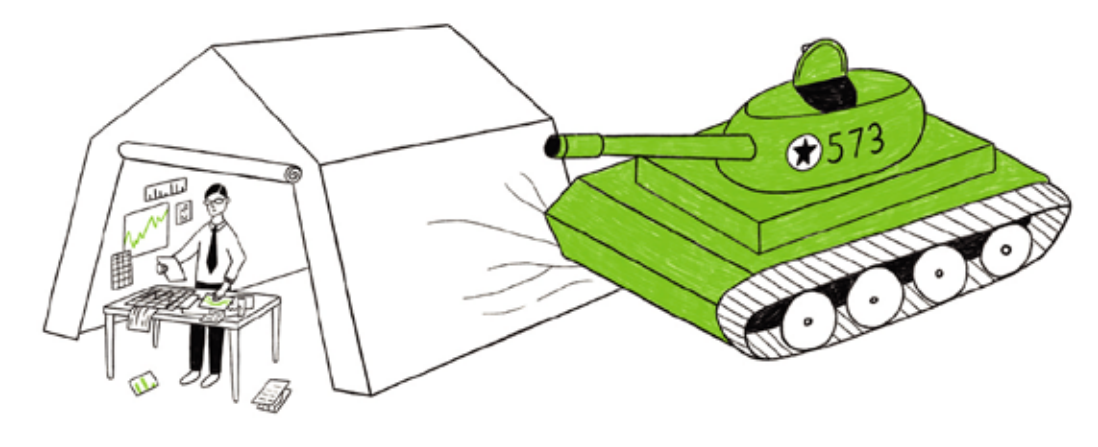
Regression zum Mittelwert

Etwas ungewöhnlich Gutes oder Schlechtes bewegt sich mit der Zeit wieder zum Durchschnitt.

APPLICATION SUCCESS RATE		
	MALE	FEMALE
SUBJECT 1	14 % (168 of 1200)	15 % (270 of 1800)
SUBJECT 2	50 % (400 of 800)	51 % (102 of 200)
TOTAL	28 % (568 of 2000)	19 % (372 of 2000) ??

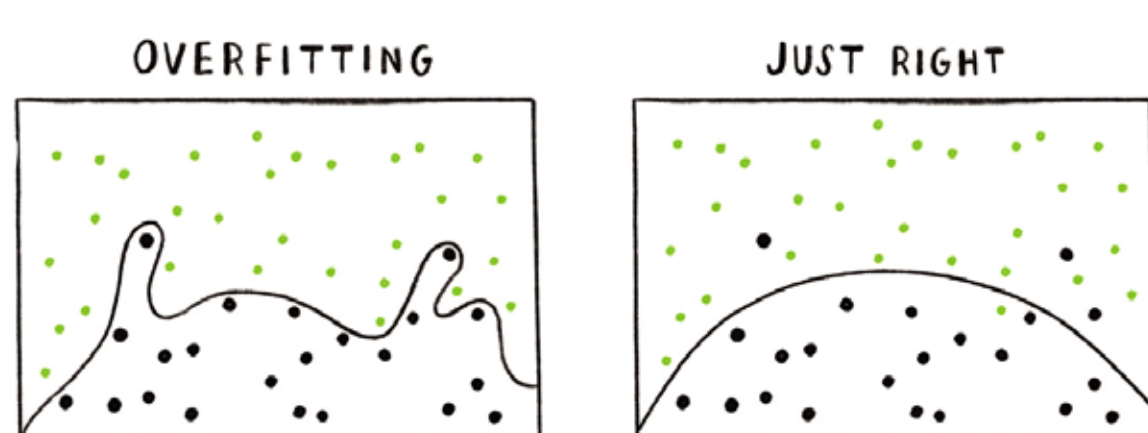
Simpson-Paradoxon

Wenn ein Trend in verschiedenen Teilmengen von Daten auftritt, aber verschwindet oder gegenteilig wird, wenn die Gruppen kombiniert werden.



McNamara-Täuschung

Sich in komplexen Situationen ausschließlich auf Metriken verlassen und den Blick für das große Ganze verlieren.



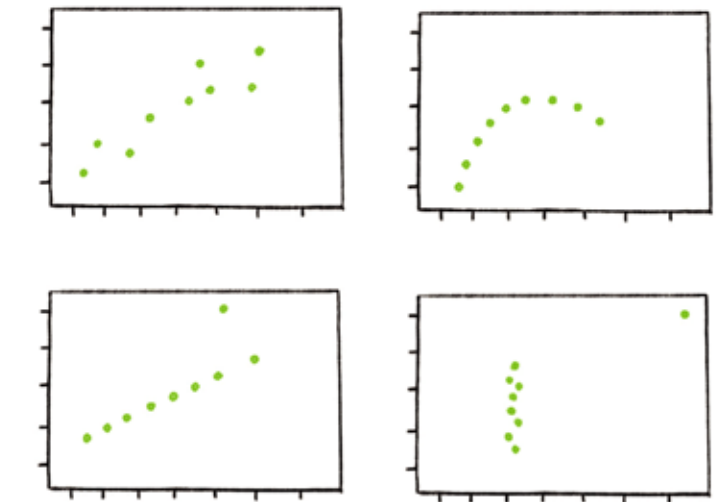
Overfitting

Erstellen eines Modells, das zu sehr auf die vorhandenen Daten zugeschnitten und nicht repräsentativ für den allgemeinen Trend ist.



Publikationsbias

Interessante Forschungsergebnisse werden mit höherer Wahrscheinlichkeit veröffentlicht und verzerren unseren Eindruck von der Realität.



Gefahr zusammenfassender Metriken

Nur auf zusammenfassende Metriken schauen und große Unterschiede in den Rohdaten übersehen.