

N° **301**

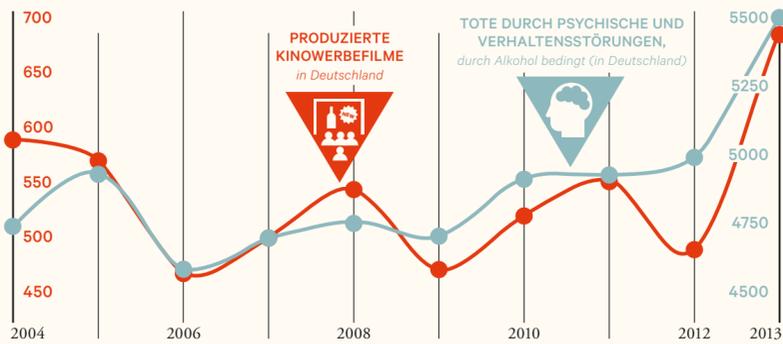
Die Themen der letzten Grafiken:
300 Crowdfunding
299 Formel-1-Wagen
298 Aufbau eines Skis
 Weitere Grafiken im Internet:
www.zeit.de/grafik

Es ist nicht so, wie es scheint

Guck mal, die Grafik! Wie ähnlich die beiden Kurven verlaufen, da muss es doch einen Zusammenhang geben ...
 Nein, muss es nicht. Aber solche Trugschlüsse sind ein häufiges Phänomen. Wir haben in den unterschiedlichsten Statistiken nach solchen Scheinkorrelationen gesucht. Und viel dabei gelacht

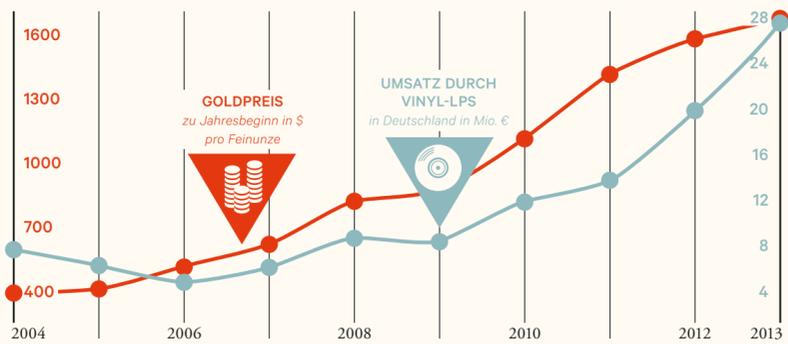
IRREFÜHRENDE WERBUNG

Können Werbefilme verantwortlich sein für den Tod von Menschen?
 Korrelationskoeffizient: 0,792 (siehe Erklärung unten rechts)



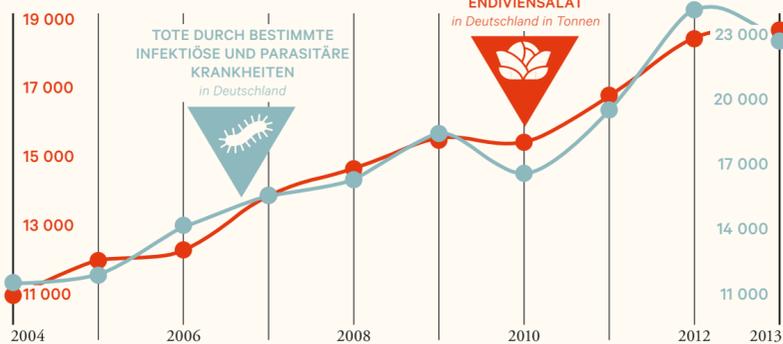
GOLDENE SCHALLPLATTEN

Treibt der Verkauf von Vinylplatten den Goldpreis in die Höhe?
 Korrelationskoeffizient: 0,905



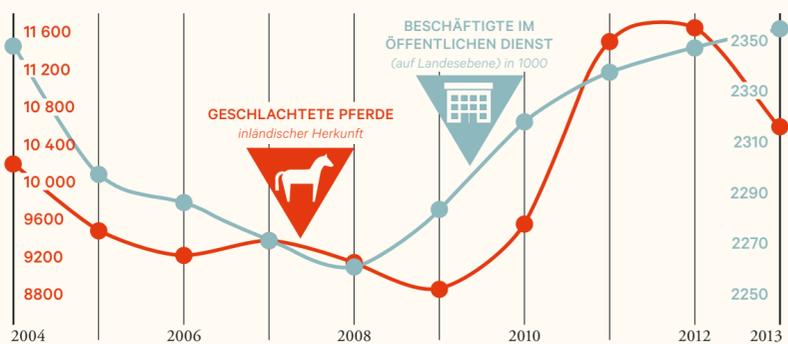
GIFTIGES GRÜN

Ist Endiviensalat verantwortlich für Infektionskrankheiten?
 Korrelationskoeffizient: 0,981



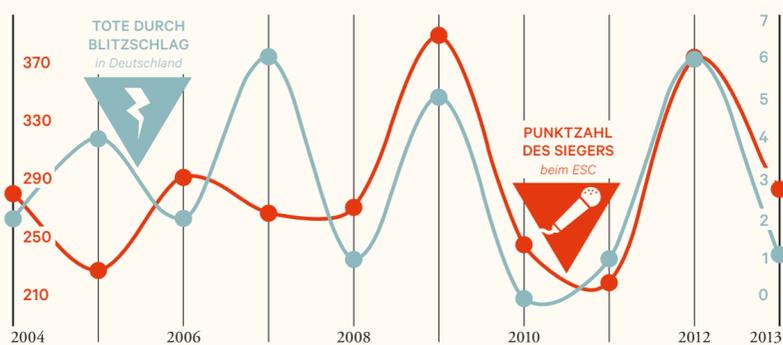
TÖDLICHER AMTSSCHIMMEL

Nimmt die Anzahl der Beschäftigten im öffentlichen Dienst zu, weil mehr Pferde geschlachtet werden?
 Korrelationskoeffizient: 0,821



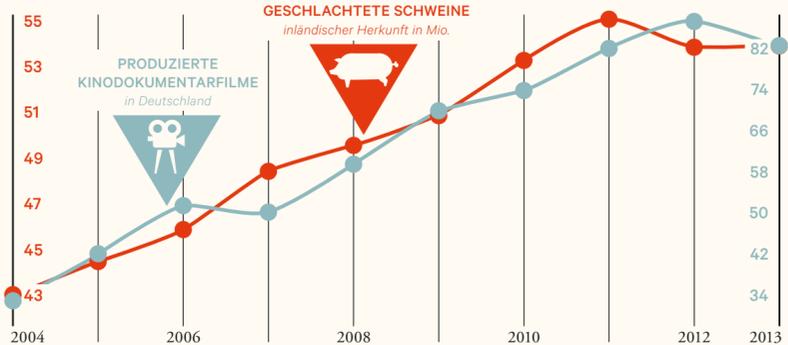
EINSCHLAGENDER ERFOLG

Was hat die Punktzahl des Siegers beim Eurovision Song Contest mit Toden durch Blitzschlag zu tun?
 Korrelationskoeffizient: 0,571



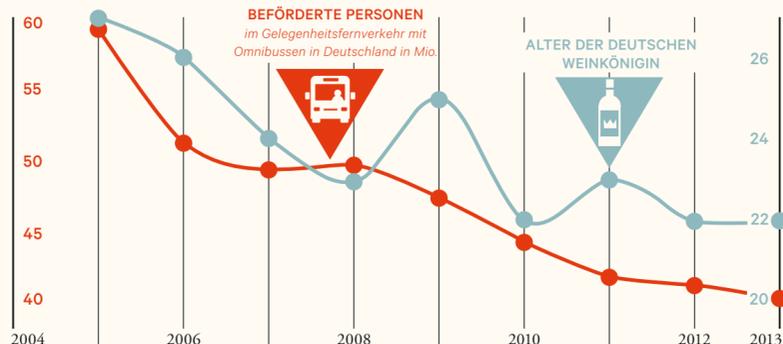
SCHWEINISCHE FILME

Können Dokumentarfilme schuld sein am Tod von Schweinen?
 Korrelationskoeffizient: 0,974



JUNGE REISEHINDERNISSE

Fahren immer weniger Menschen Bus, weil die Weinköniginnen jünger werden?
 Korrelationskoeffizient: 0,879



So ähnlich: Der Korrelationskoeffizient

Statistiker kennen einen Wert für die Ähnlichkeit zweier Kurven, den **Korrelationskoeffizienten**. Er gibt an, ob es einen linearen Zusammenhang zwischen zwei Datenreihen gibt. Er nimmt Werte zwischen -1 und +1 an. **Minus eins** steht für einen umgekehrten Zusammenhang – der eine Wert nimmt genau in dem Maß zu, in dem der andere abnimmt. **Plus eins** steht für zwei Zahlenreihen, deren Verlauf sich exakt in Deckung bringen lässt.

Aber auch aus starken Korrelationen wie in unseren Beispielen darf man nicht schließen, dass es einen **kausal Zusammenhang** gibt (dass also der Anstieg von A die **Ursache** für den Anstieg von B ist). Es kann eine versteckte dritte Größe geben, die mit beiden zusammenhängt. Zum Beispiel gibt es, wenn mehr Sonnenbrillen verkauft werden, auch mehr Sonnenbrände. Eine Sonnenbrille zu tragen erzeugt aber natürlich keinen Sonnenbrand – vielmehr lässt die stärkere Sonneneinstrahlung im Sommer beide Größen ansteigen. Ein beliebtes Beispiel ist auch der Zusammenhang zwischen der **Zahl der Störche und der Geburtenziffer**. Sie korrelieren zwar, doch es ist die Verstädterung, die beide Werte langfristig sinken lässt.

Ein gleichförmiger Verlauf zweier Kurven kann aber auch **völlig zufällig** sein, wie die Beispiele auf dieser Seite zeigen. Insbesondere wenn es sich um stetig steigende oder abfallende Zahlenreihen handelt, entstehen leicht **»Scheinkorrelationen«**. Und umgekehrt gilt: Auch eine geringe Korrelation schließt nicht sicher aus, dass zwei Größen etwas miteinander zu tun haben.

Grafik:
 Christoph Drösser (Diagramme),
 Jelka Lerche

Recherche:
 Christoph Drösser

Quellen:
 BVMI;
 finanzen.net;
 Statistisches Bundesamt;
 Wikipedia;
 inspiriert durch »Spurious Correlations« (tylervigen.com)

Methodik:
 Vergleich von je zehn (bzw. neun) Datenpunkten pro Größe aus denselben Jahren