Statistik zwischen Wissenschaft und Scharlatanerie

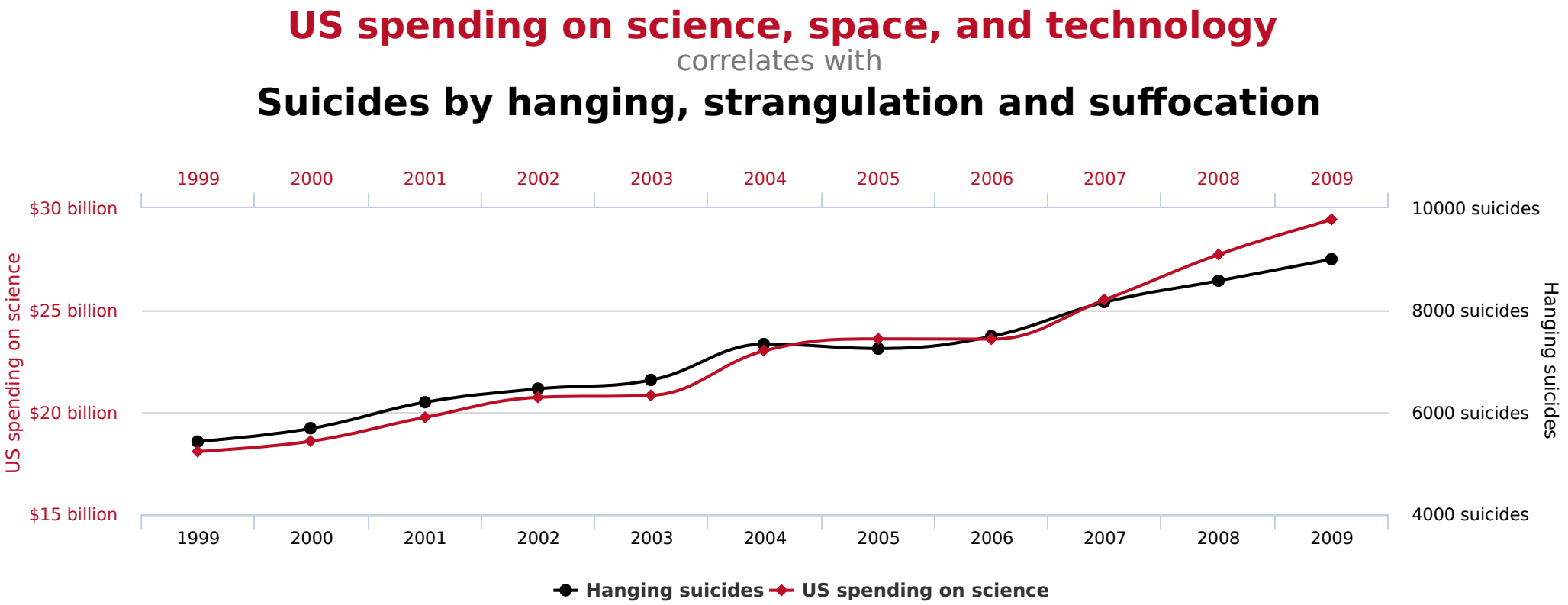
V 0.1 170704 Hattingen
Jochim Selzer
Jselzer@vorratsdatenspeicherung.de
https://cryptoparty.in/cryptopartykbn
a57a 4e28 a59d 0385 d33d 6301 763d ce1b 65f4 c445

Big Data an der Tankstelle

- Alle tanken Diesel.
- Alle Opel-Fahrer tanken Diesel.
- Alle Fahrer blauer Opel ab Baujahr 2005 tanken Diesel

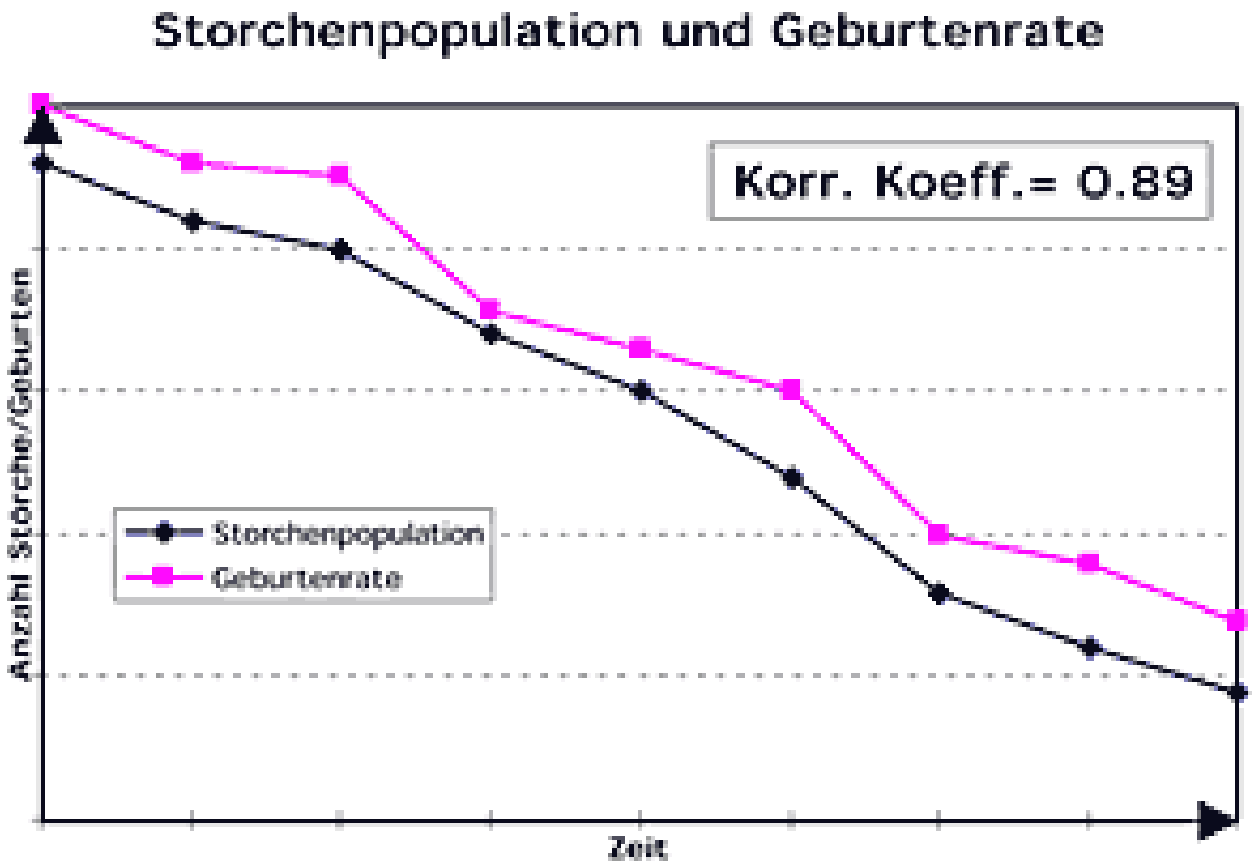
- Viktor Mayer-Schönberger: Big Data – die Revolution, die unser Leben verändern wird
- Wir brauchen die Zusammenhänge nicht verstehen, weil sie trotzdem wahr sind.
- Ignaz Philipp Semmelweis (1819-1865), Gynäkologe
-

Zweifelhafte Zusammenhänge



http://www.tylervigen.com/spurious-correlations

Der Storch und die Kinder



https://de.wikipedia.org/wiki/Scheinkorrelation

Big Data bei Spiegel Online

- David Kriesel: SpiegelMining – Reverse Engineering von Spiegel-Online
-

Eine einfache Analyse

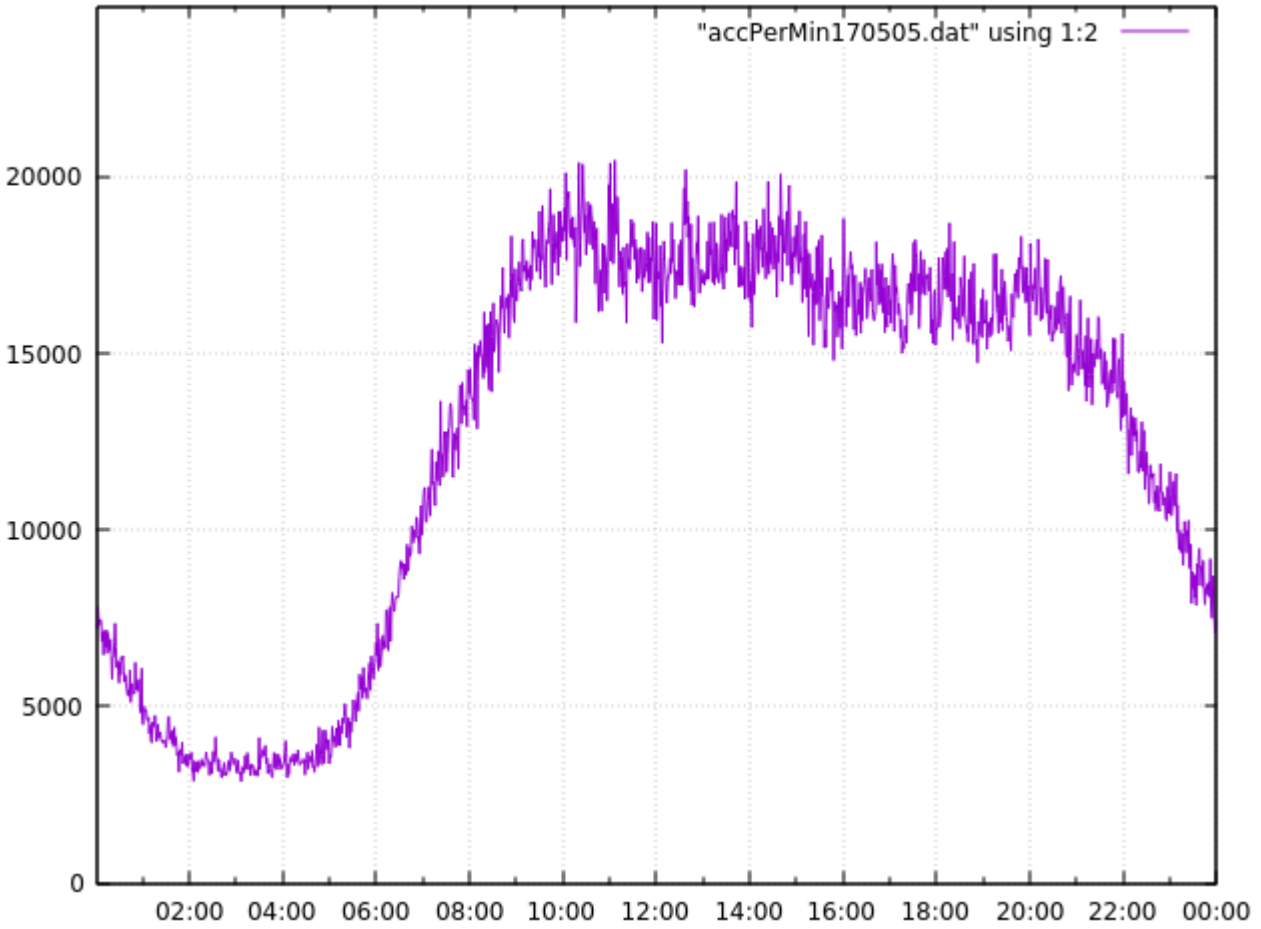
- Webserver-Zugriffslogs
- Enthalten nur IP-Adressen und Zugriffszeiten
- Frage: Finde mögliche Angreifer

Rohdaten

```
217.186.228.96 04/05/2017:23:59:59
78.49.146.182 04/05/2017:23:59:59
178.5.185.179 05/05/2017:00:00:00
178.5.185.179 04/05/2017:23:59:58
217.186.228.96 04/05/2017:23:59:59
88.73.181.251 05/05/2017:00:00:00
62.159.102.26 04/05/2017:23:59:59
109.45.3.3 04/05/2017:23:59:59
88.73.181.251 05/05/2017:00:00:00
91.10.148.178 04/05/2017:23:59:59
178.5.185.179 05/05/2017:00:00:00
95.91.252.205 05/05/2017:00:00:00
217.186.228.96 05/05/2017:00:00:00
178.5.185.179 05/05/2017:00:00:00
217.246.170.83 04/05/2017:23:59:59
178.5.185.179 05/05/2017:00:00:00
178.5.185.179 05/05/2017:00:00:00
178.5.185.179 05/05/2017:00:00:00
88.73.181.251 05/05/2017:00:00:00
178.5.185.179 05/05/2017:00:00:00
```

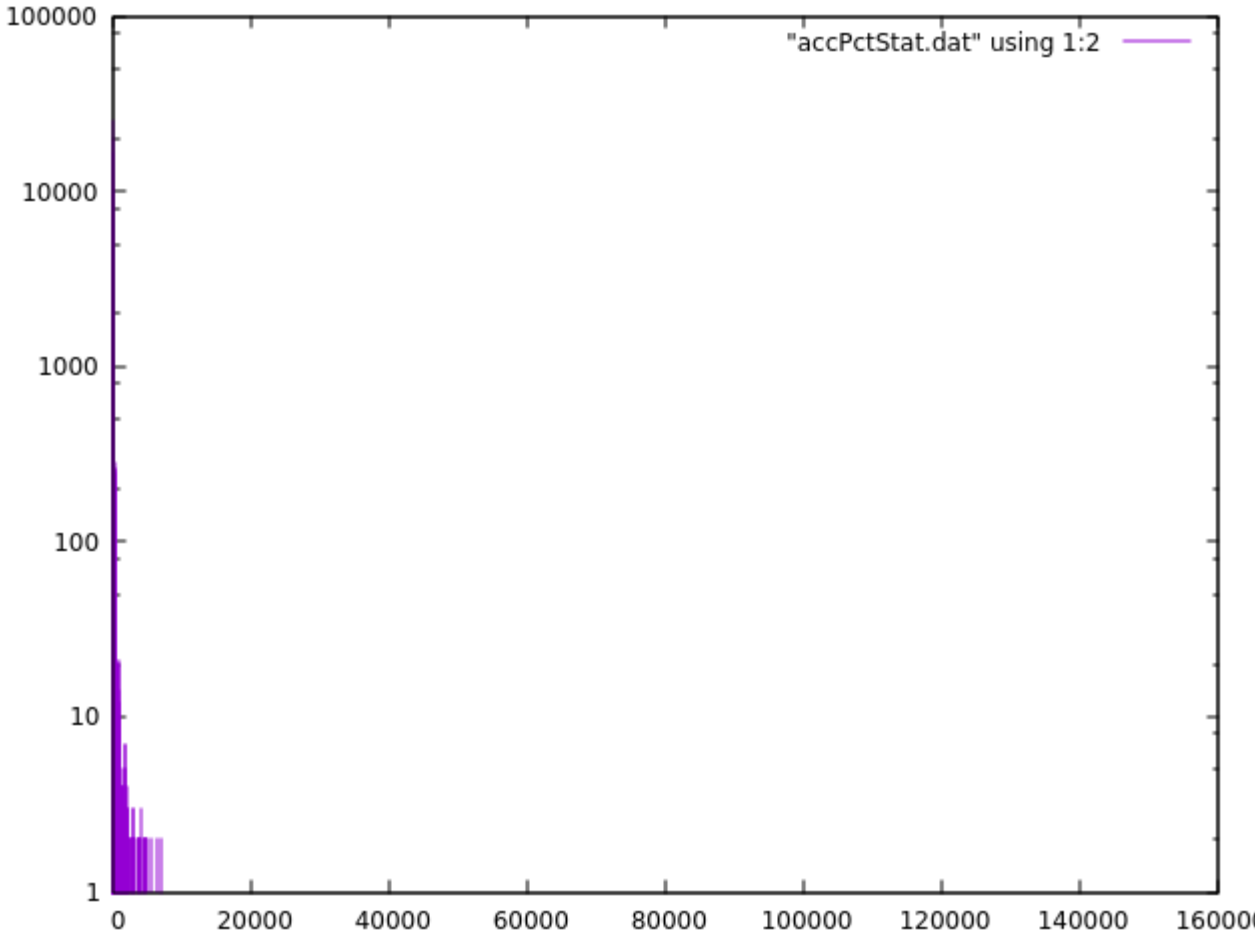
Insgesamt 18.205.753 Datensätze

Zugriffe pro Minute



Wie oft greift ein User zu?

Zugriffe pro IP	Anzahl IPs
33	4115
31	4388
29	4505
32	6301
2	6393
28	7097
27	9560
1	16675
25	19529
26	25191



Die meisten Leute haben 20 bis 30 Seitenzugriffe

Normales Verhalten

- Aktiv zwischen 10 und 20 Uhr
- Kommt aus Mitteleuropa, wahrscheinlich DE oder AT
- Ungefähr 30 Aufrufe
- Wir suchen also Adressen aus dem Ausland mit (z.B.) mehr als 1000 Aufrufen

Mögliche Verdächtige

46.61.242.132	1097	RU
188.13.160.224	1134	IT
81.244.34.94	1294	BE
193.3.141.124	1309	DK
109.115.24.200	1477	IT
85.203.13.34	1687	FR
95.27.176.165	1728	RU
95.195.140.27	1830	SE
37.152.82.60	2130	ES
194.72.50.58	2765	GB
81.45.78.19	2772	ES
193.15.96.218	2934	SE
217.195.251.43	3222	NL
46.226.218.20	6749	GB
159.205.69.249	8338	PL
94.48.141.104	9797	SA
2.254.128.24	12812	SE
217.17.137.242	**19265**	**NL**

Verdächtige untersucht

```
$ nmap -Pn 217.17.137.242

Starting Nmap 7.01 ( https://nmap.org ) at 2017-
07-04 14:02 CEST
Nmap scan report for 217.17.137.242
Host is up (0.0065s latency).
Not shown: 998 filtered ports
PORT     STATE SERVICE
53/tcp   open  domain
8080/tcp open  http-proxy
```