

Datenvisualisierung in der Wissenschaft

Die Integrität der Information

Dr. Cédric Scherer

RPTU Nachwuchsring
3., 10. und 17. Juli 2025



Information

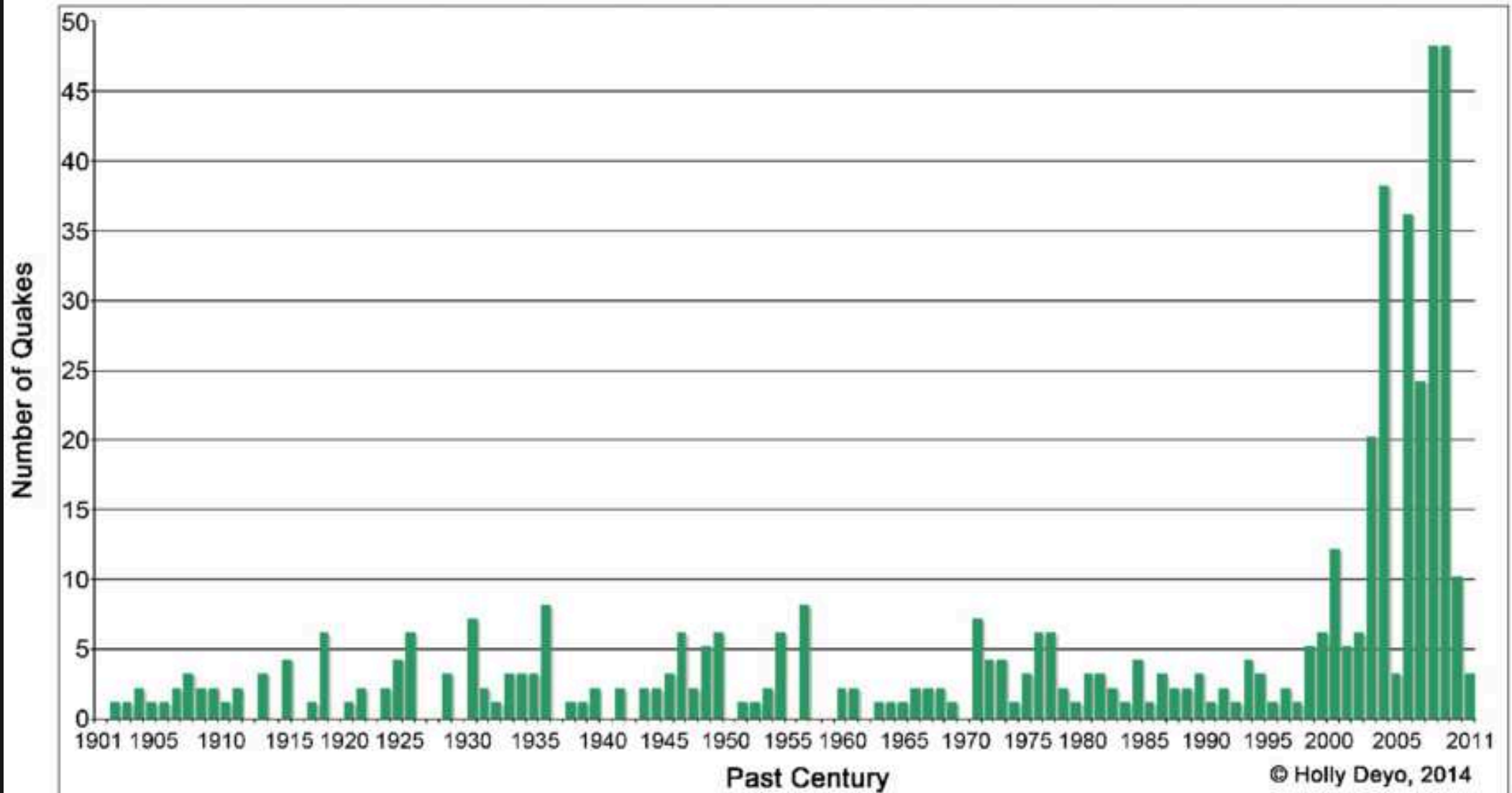
Integrität der Daten



**Daten bilden die Realität nie
vollständig oder perfekt ab.**



USGS Worldwide Deadly & Destructive Earthquakes Between Magnitudes 6 and 8
reconstructed from [http:// earthquake.usgs.gov/regional/world/historical.php](http://earthquake.usgs.gov/regional/world/historical.php)

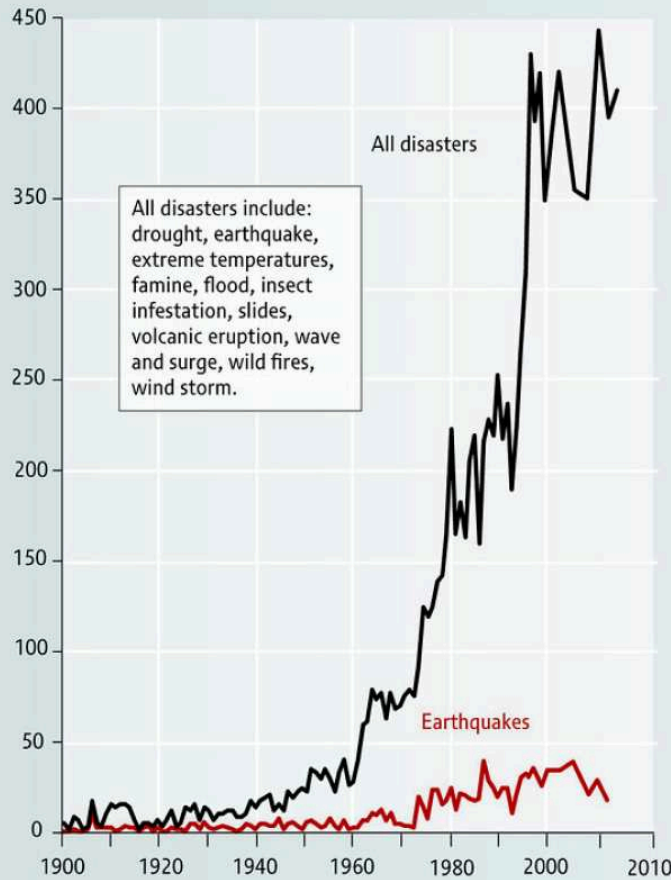


Quelle: Poole (2018)



Number of disasters

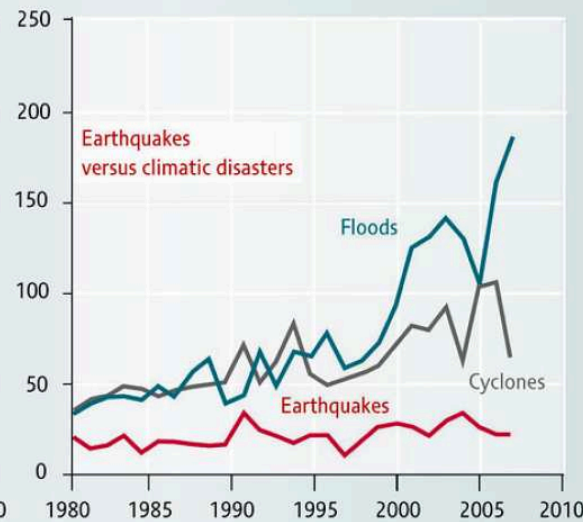
per year



Source: CRED Annual Disaster Statistical Review 2006, 2007.

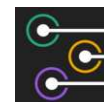
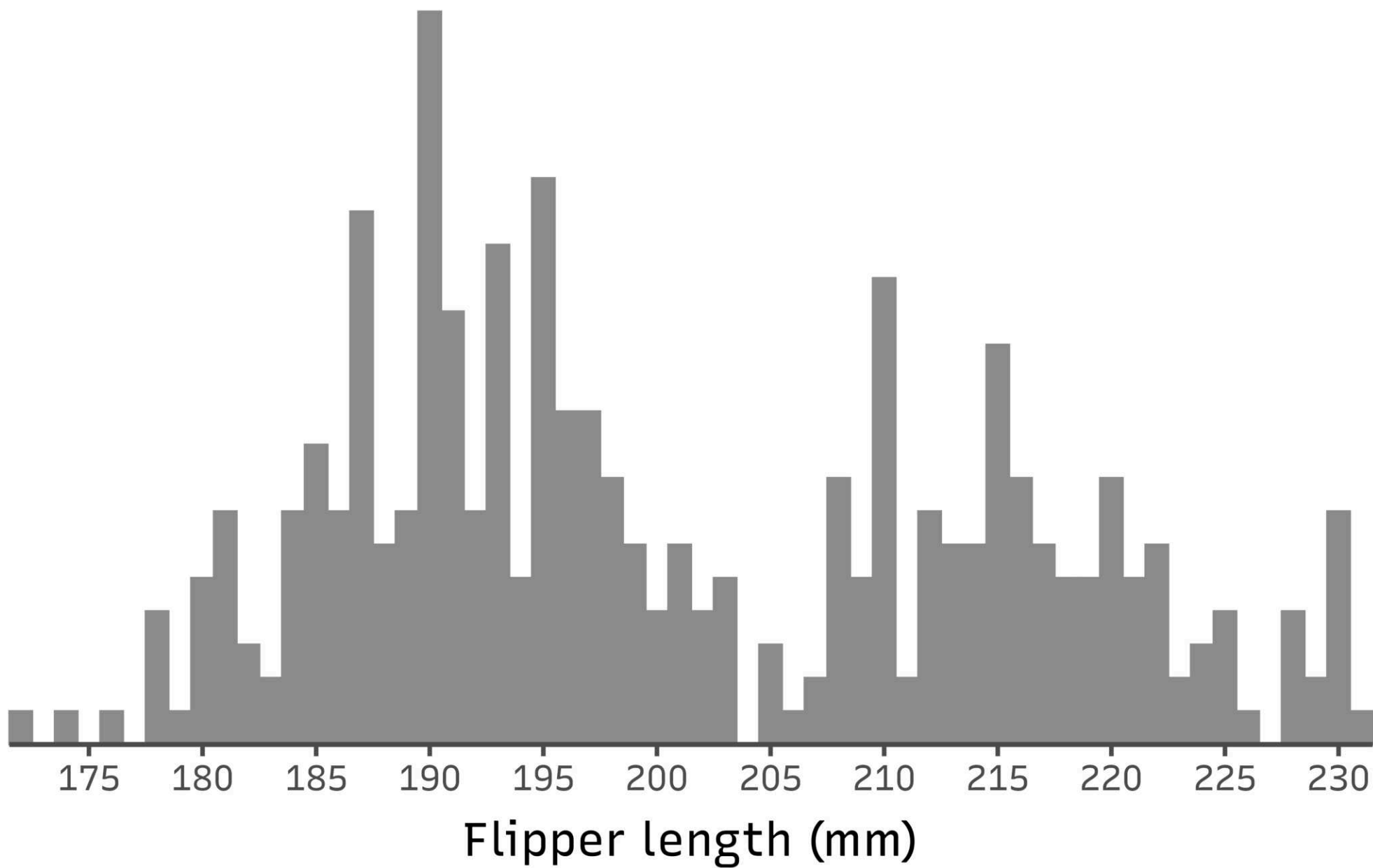
Trends in number of reported disasters

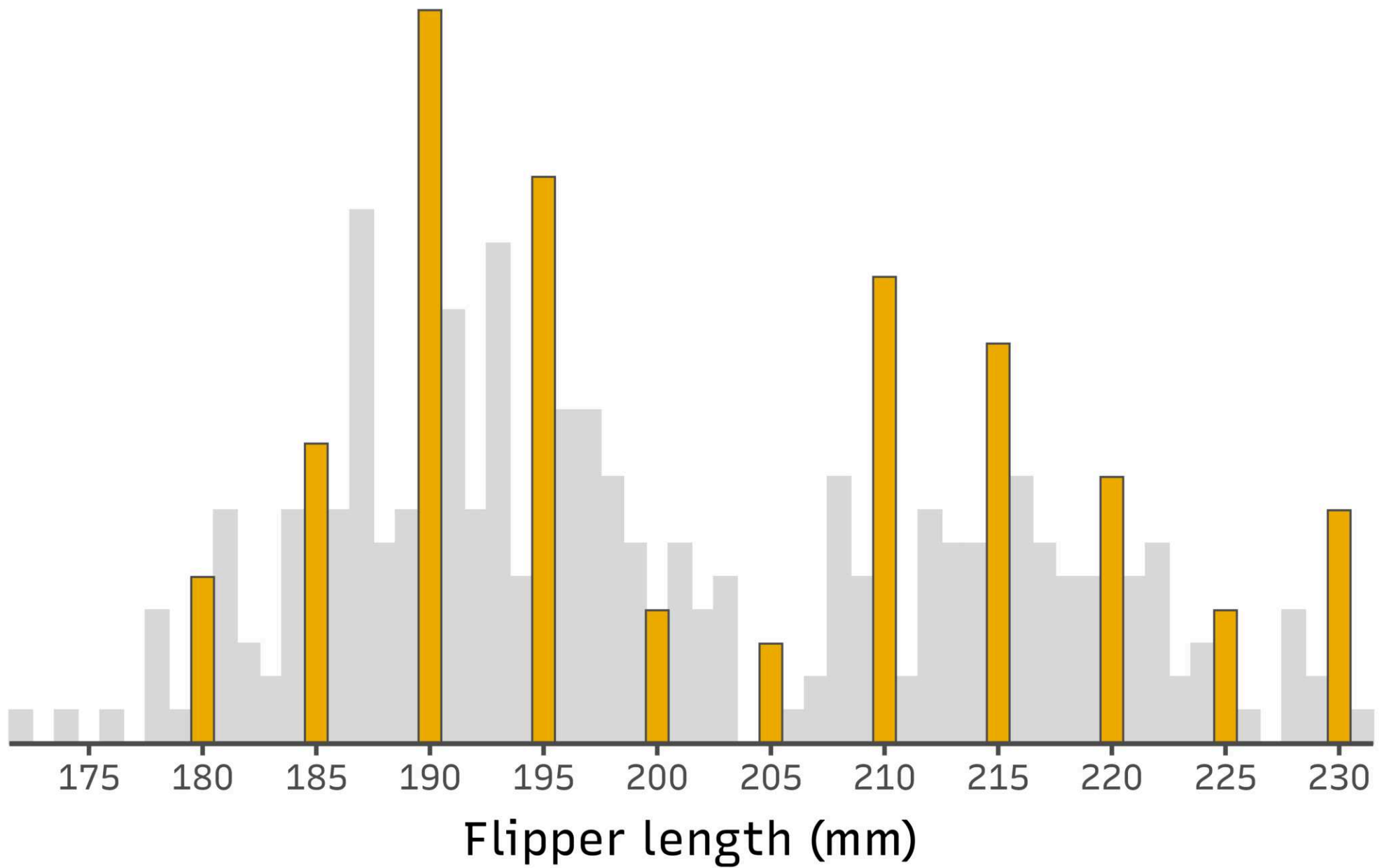
Much of the increase in the number of hazardous events reported is probably due to significant improvements in information access and also to population growth, but the number of floods and cyclones reported is still rising compared to earthquakes. Is global warming affecting the frequency of natural hazards?



“Much of the increase of hazardous events reported is probably due to **significant improvements in information access.**”







Scientists rename human genes to stop Microsoft Excel from misreading them as dates



Illustration by Alex Castro / The Verge

/ Sometimes it's easier to rewrite genetics than update Excel

by **James Vincent**

Aug 6, 2020, 2:44 PM GMT+2



James Vincent is a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

Update October 24th, 2023, 11:25AM ET: Microsoft has updated Excel on Windows and macOS, adding a toggle to turn off automatic data conversion. The original version of this article continues below.

Quelle: [The Verge](#)



[Comment](#) | [Open access](#) | Published: 23 August 2016

Gene name errors are widespread in the scientific literature

[Mark Ziemann](#), [Yotam Eren](#) & [Assam El-Osta](#) [Genome Biology](#) **17**, Article number: 177 (2016) | [Cite this article](#)**161k** Accesses | **144** Citations | **2590** Altmetric | [Metrics](#)

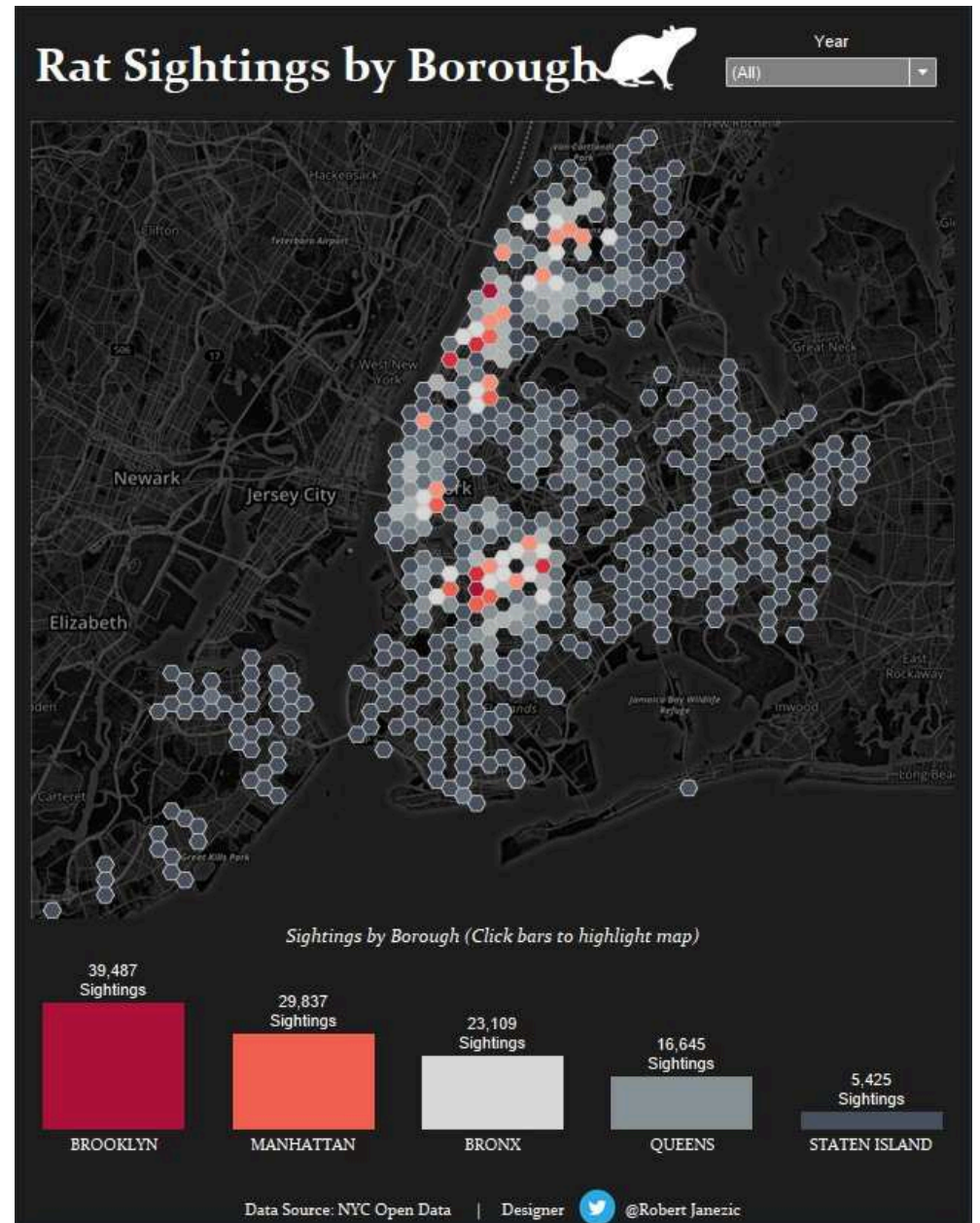
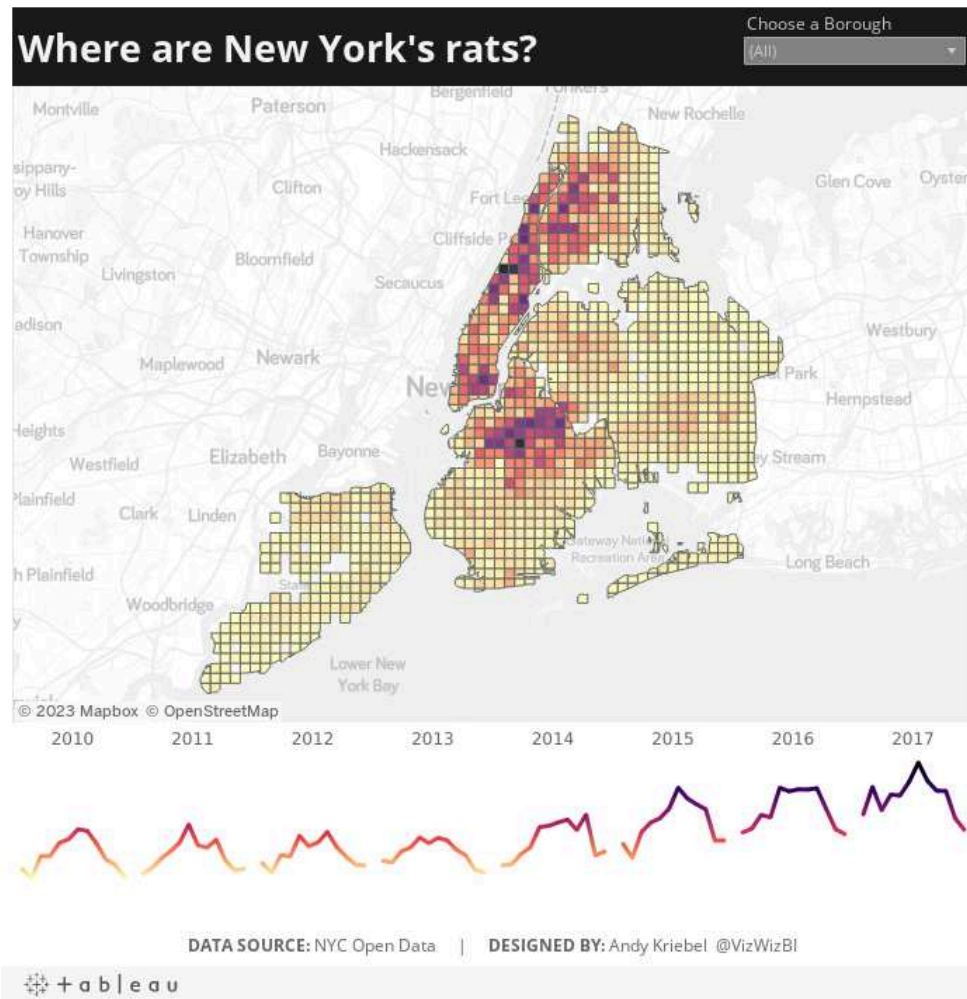
Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

Ziemann, Eren & El-Osta (2016) doi: [10.1186/s13059-016-1044-7](https://doi.org/10.1186/s13059-016-1044-7)

Cédric Scherer Data Visualization & Information Design





Integrität der Information

- **Datenqualität:**

- Schätzung, Präzision und Fehler
- Fehler und Irrtümer bei der Berechnung
- unvollständige Daten und fehlende Werte
- zusammengefasste und relative Daten

- **Nur ein Teilabbild:**

- historischer oder gegenwärtiger Zustand
- nicht Straftaten, sondern gemeldete Straftaten



Übungsteil



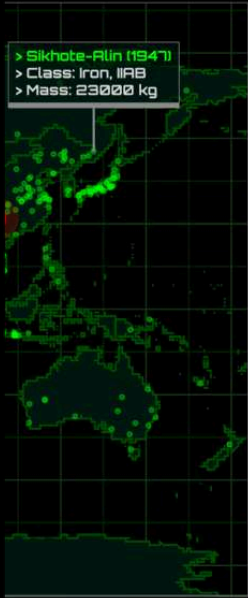
Übung

Die in der folgenden – und nicht ganz ernst gemeinten – Visualisierung gezeigten Daten haben **zwei Probleme**.

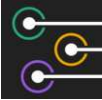
Identifiziere und erkläre sie.

Wie kann man mit diesen Problemen umgehen?





administration (NASA)



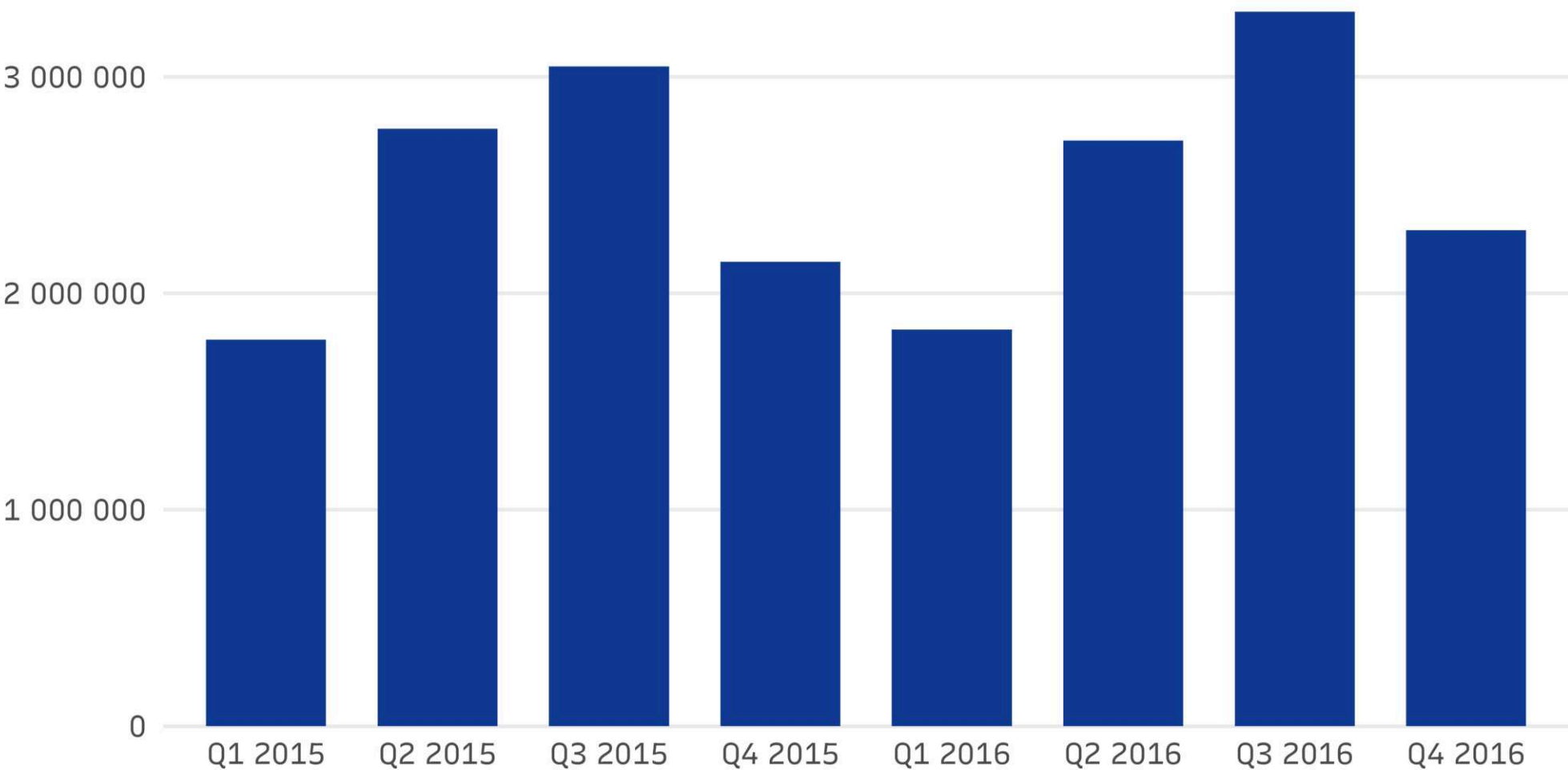
Übung

Schaue dir die folgende Zeitreihen nacheinander an. Notiere zu jeder Abbildung vor der Betrachtung der nächsten:

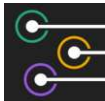
- Welche **Muster** oder andere interessante **Details** gibt es?
- Ist der **Detailgrad sinnvoll** oder wäre eine andere Auflösung besser?
- Gibt es **Auffälligkeiten** in den Daten?
- Was könnten **Ursachen** für Muster und andere Auffälligkeiten sein? Liste **potentielle Gründe** sowie wie man diese überprüfen könnte.



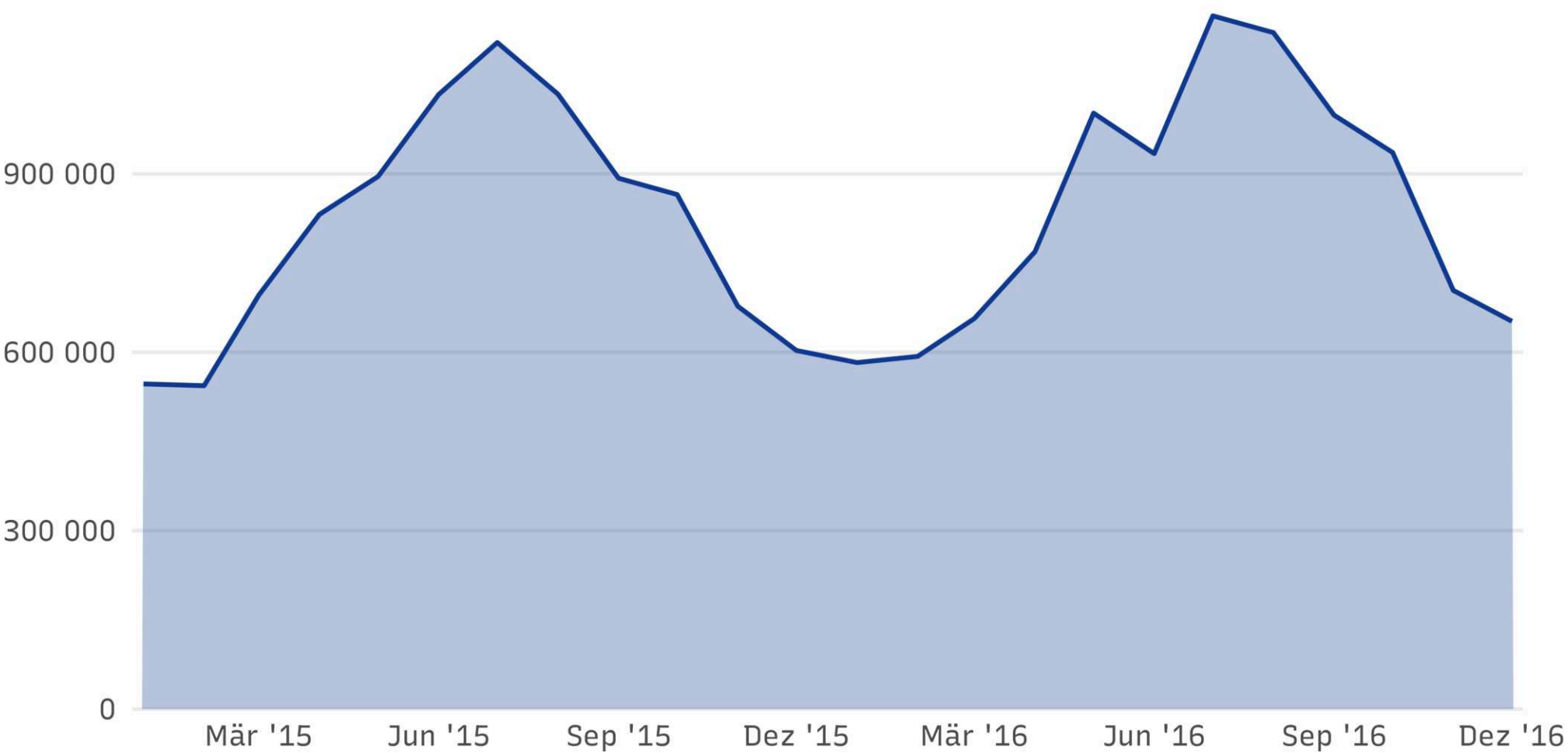
Registrierte TfL-Fahrradmieten in London pro Quartal



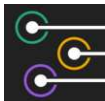
Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



Registrierte TfL-Fahrradmieten in London pro Monat



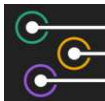
Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



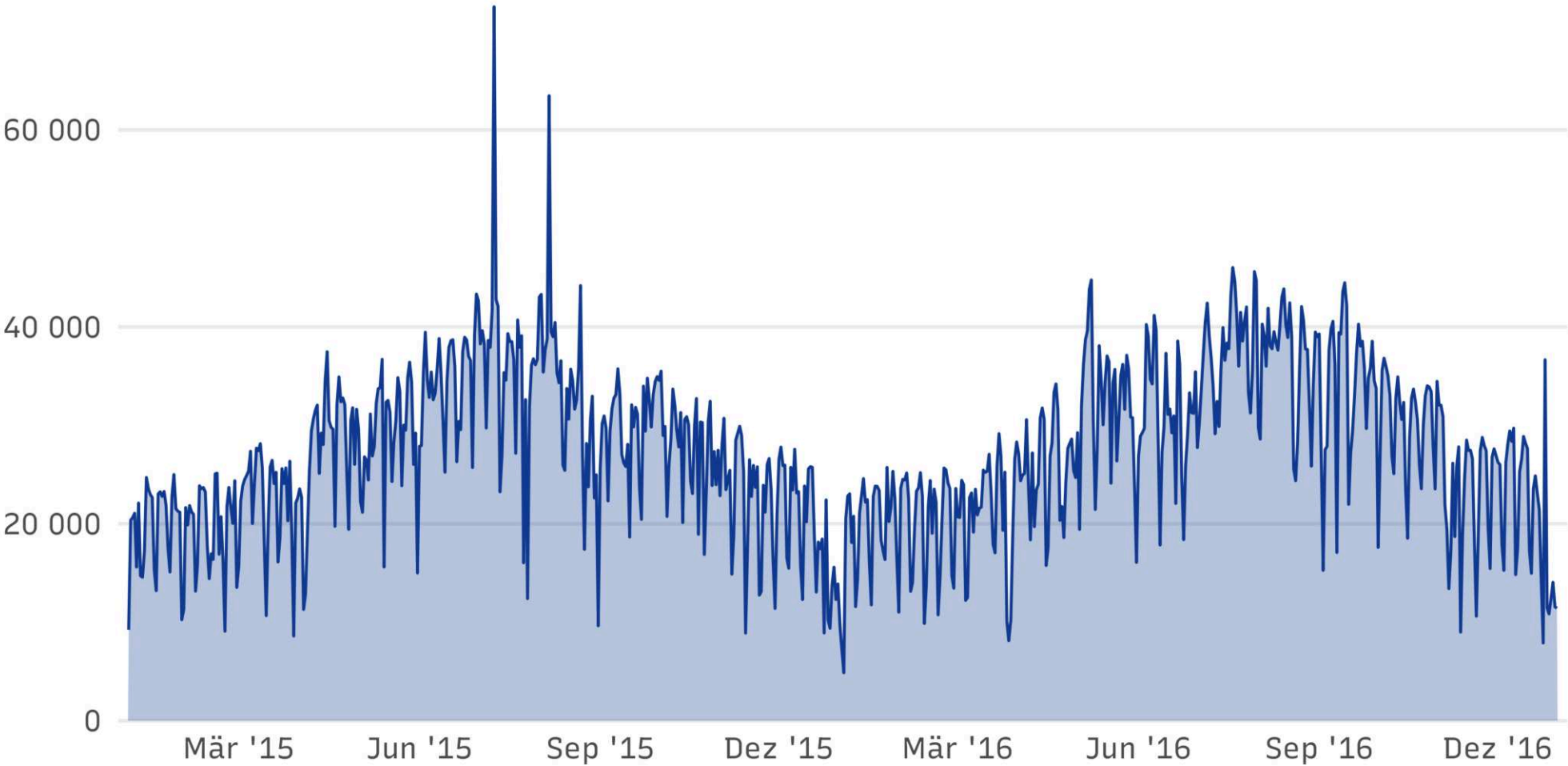
Registrierte TfL-Fahrradmieten in London pro Woche



Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



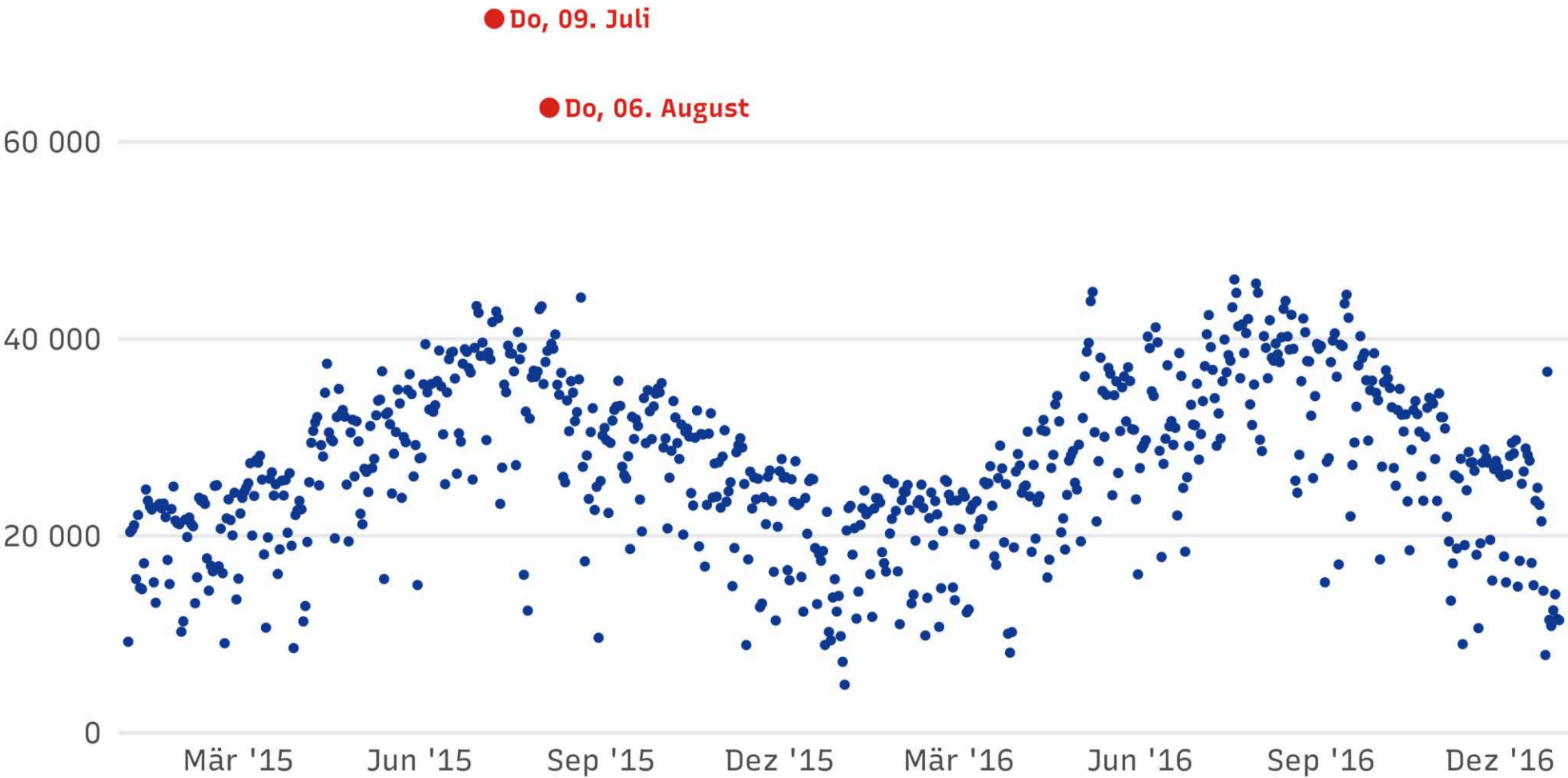
Registrierte TfL-Fahrradmieten in London pro Tag



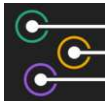
Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



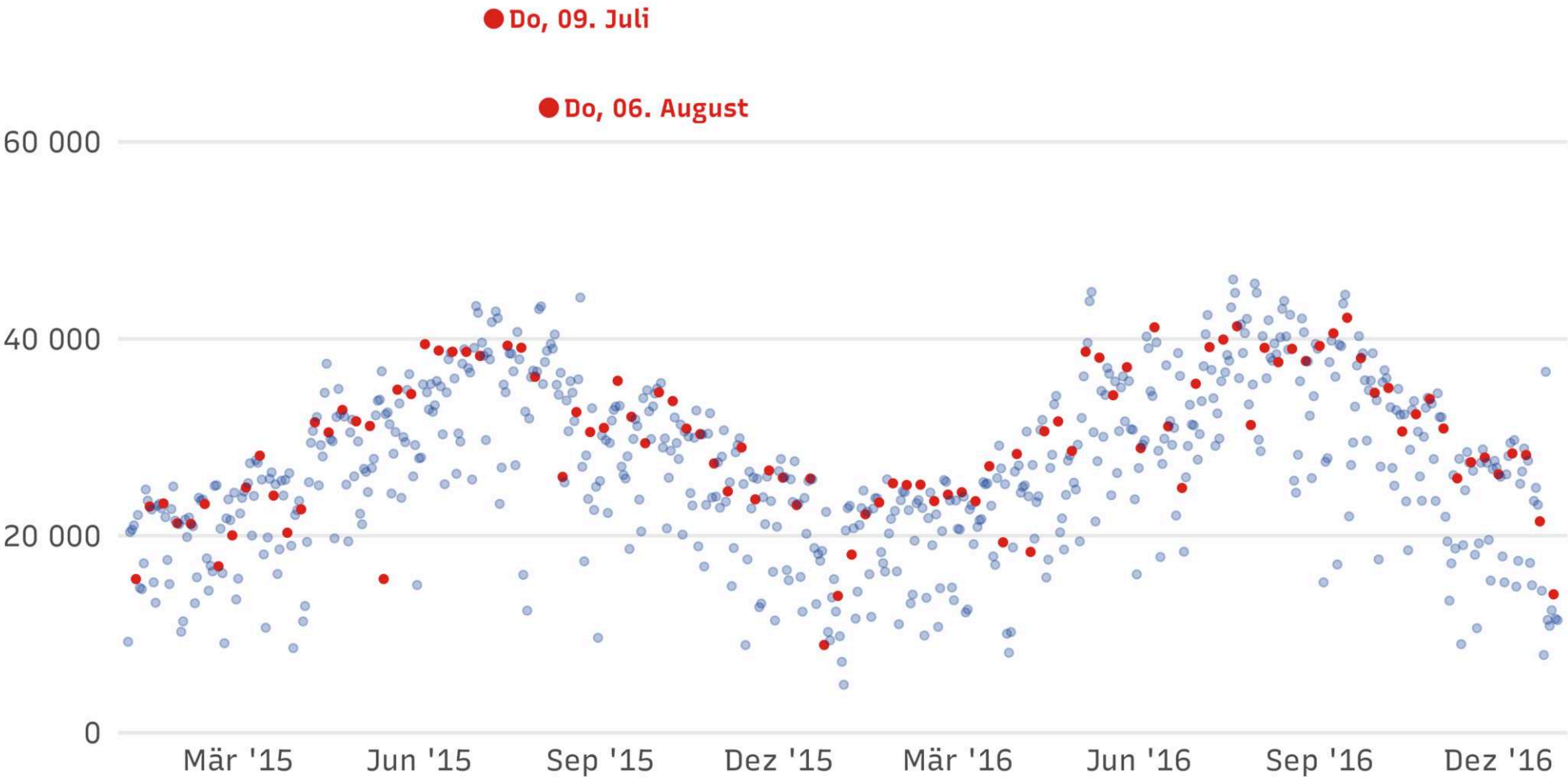
Registrierte TfL-Fahrradmieten in London pro Tag



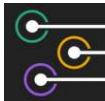
Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



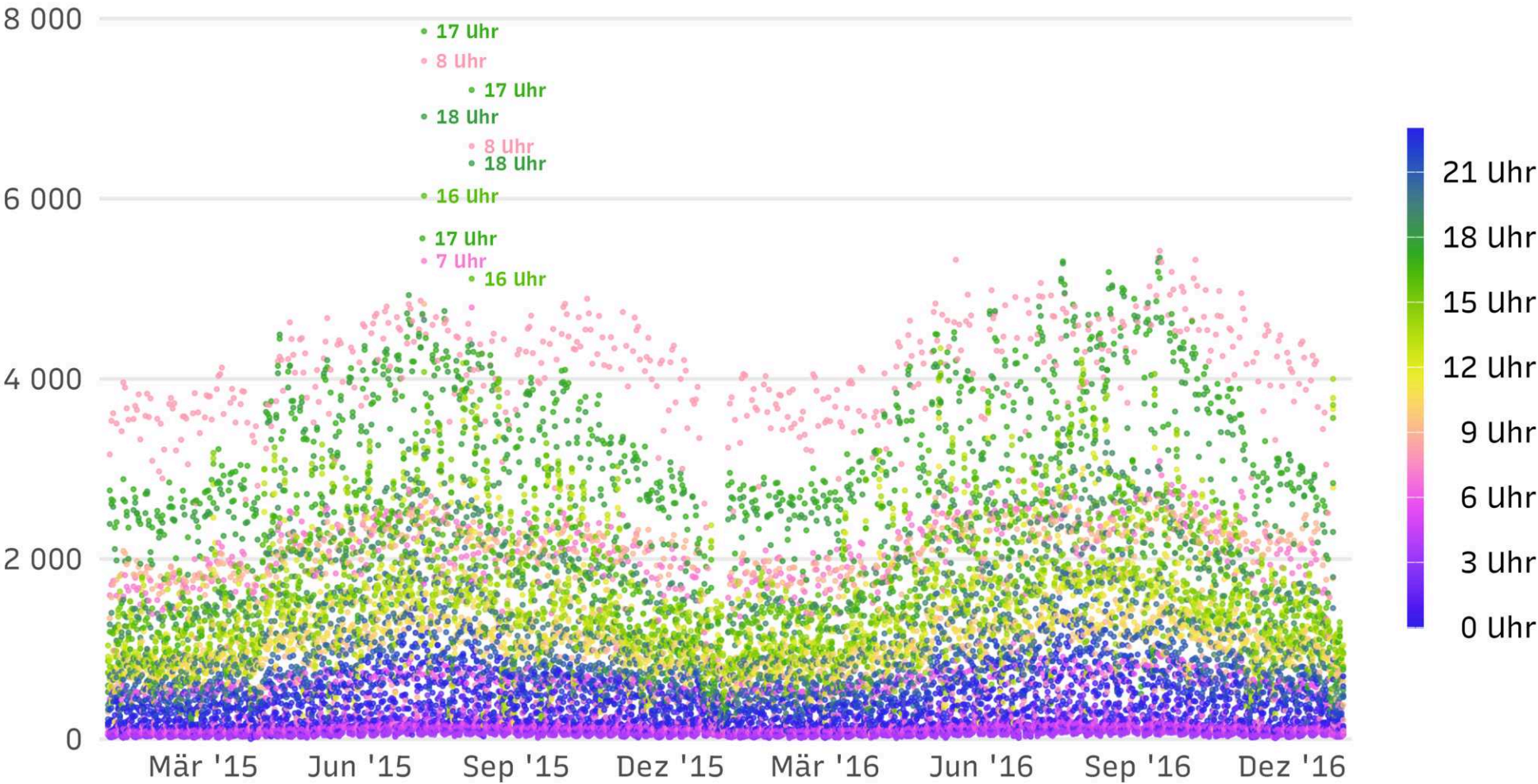
Registrierte TfL-Fahrradmieten in London pro Tag



Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data



Registrierte TfL-Fahrradmieten in London pro Stunde



Quelle: Transport for London, 2015-2016 | Powered by TfL Open Data

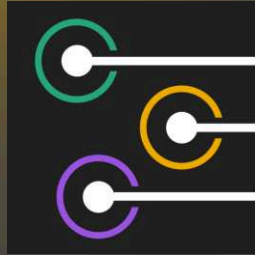


Projektarbeit

- Überlege, ob es bei deinen Daten ebenfalls zu Fehlern bei der Aufnahme, Probleme durch versteckte Mustern, “Guesstimation” oder ähnlichem gekommen sein kann.
- Wie gehst du, in diesem Fall und generell, mit “outliern”, “missing data” und bereits aggregierten Daten um?
- Wie stellst du sicher, dass die Daten ausreichend geprüft werden?
- Sind dir Fälle bekannt, bei denen es aus etwaigen Gründen zu Datenproblemen kam?



Dankeschön!



CÉDRIC SCHERER
Data Visualization & Information Design

www.cedricscherer.com

