

Geschichten in Daten finden und erzählen

24:Stunden:Zukunft - DJV Jung

30.03.2019

Kernpunkte Datenjournalismus

“Einen Datensatz interviewen”

- Mit Thesen nach Daten suchen
- Fragen beantworten
- Visualisieren
- Geschichte schreiben - drehen etc.
- Weiterdrehen können

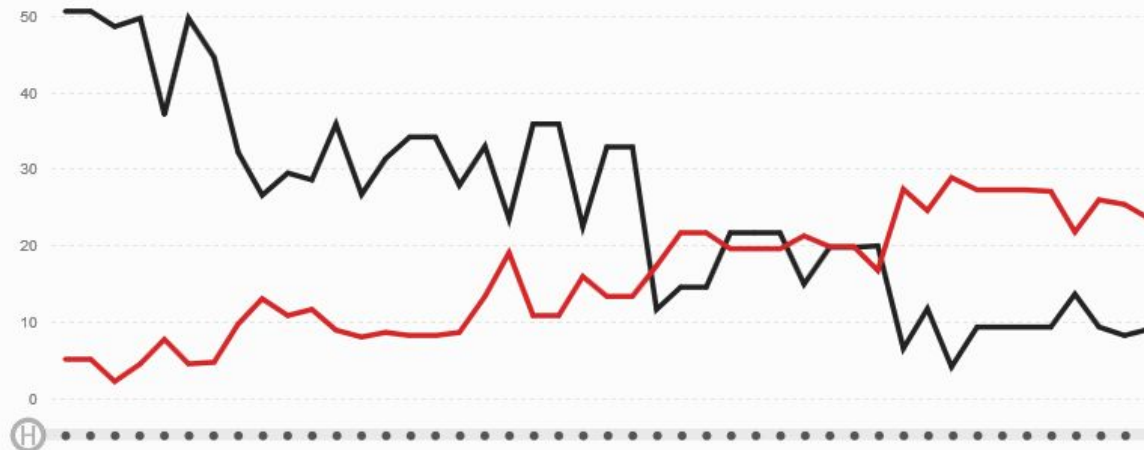
Datenjournalismus-Beispiele

- Berliner Morgenpost: M29 - Berlins Buslinie der großen Unterschiede
- BR Data / Spiegel Online: Erhöhtes Risiko
- Periscopic: On the Trump Emoto-Coaster
- rbb: An diesen Berliner Straßen herrscht Abgasalarm
- Spiegel Online: So pendelt Deutschland zur Arbeit
- SZ: Deutschlands Mietmarkt ist kaputt
- Zeit Online: Straßenbilder - Mozart, Marx & ein Diktator

- Roseneck
- Topfizer Str.
- Joe eph/Lochim-Platz
- Delbrückstr.
- Herdtstr.
- Lynenstr.
- Bismarckplatz
- Rathenauplatz
- S. Helenesee
- Kurtfriesenhamm
- Lehniner Platz/Schaubühne
- U. Adenauerplatz
- Oliver Platz
- Bleibergstr.
- U. Ullrichstr.
- U. Kurtfriesenhamm
- Europa Center
- An der Wittenbergplatz
- Schillstr.
- Litgowplatz
- Hilfeshausplatz
- Gedenksäule
- Potsdamer Brücke
- U. Menschenlehn-Grünhof
- S. Antonier Bahnhof
- Wilhelmstr./Kochplatz
- Charlottenstr./Checkpoint Charlie
- Waldespark
- Alexanderplatz
- U. Monbijouplatz
- Oranienplatz
- Anhalterplatz
- Holocaustdenkmal
- Spreewaldplatz
- U. Goltzplatz
- Gröbner Str.
- Pflügerstr.
- Sonnenallee
- U. Hermannplatz
- U. Hermannplatz/Urbanstr.
- Sonnenallee/Pannierstr.

CDU- und Linke-Wähler

Stimmen der Anwohner entlang der Buslinie M29, in Prozent

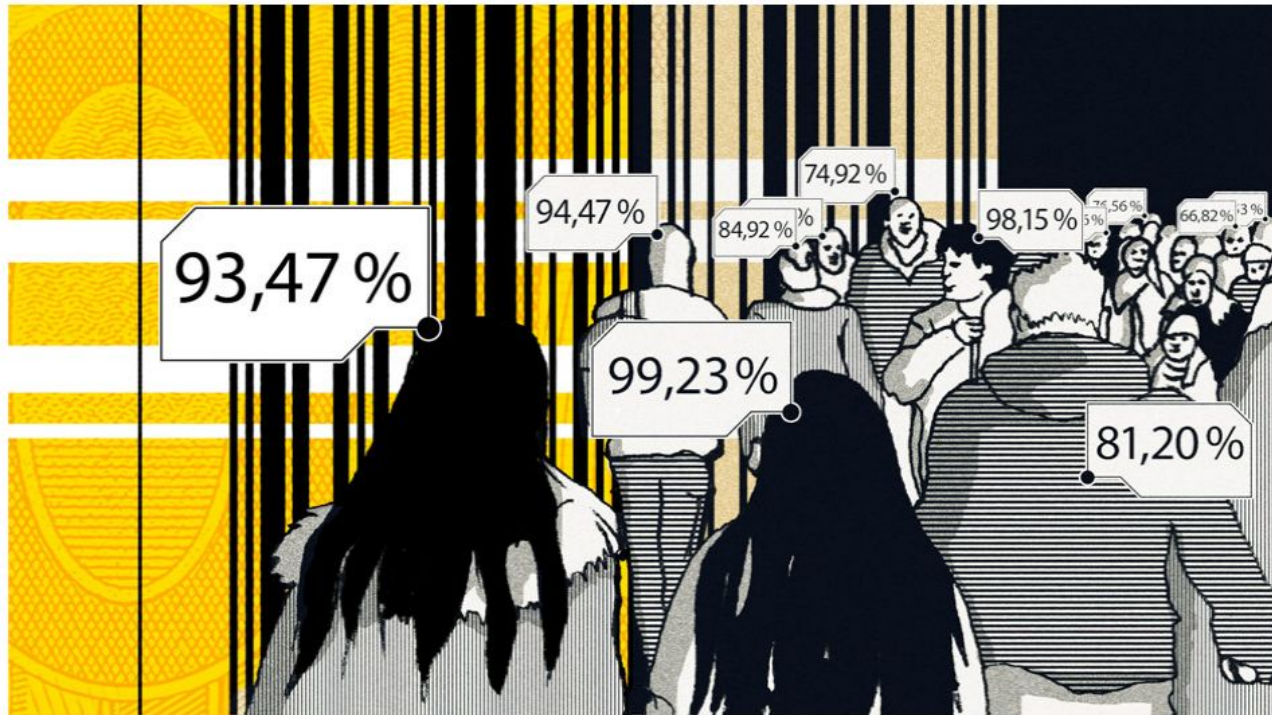


interaktiv.morgenpost.de/m29

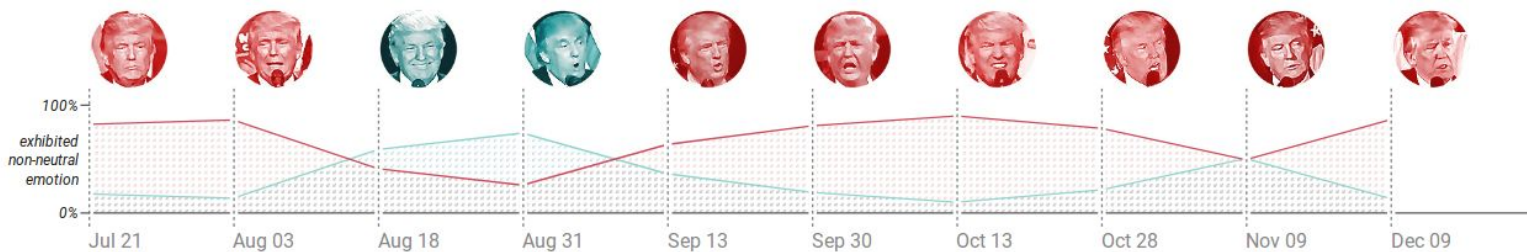
Erhöhtes Risiko

Scoring-Firmen berechnen unsere Kreditwürdigkeit, die bekannteste ist die Schufa. Jeden Tag stützen sich Banken, Provider und Online-Händler hunderttausendfach auf ihre Bewertung. Doch eine Datenanalyse von BR und Spiegel zeigt: Viele Menschen werden unverschuldet zum Risikofall erklärt. Einblicke in eine Blackbox.

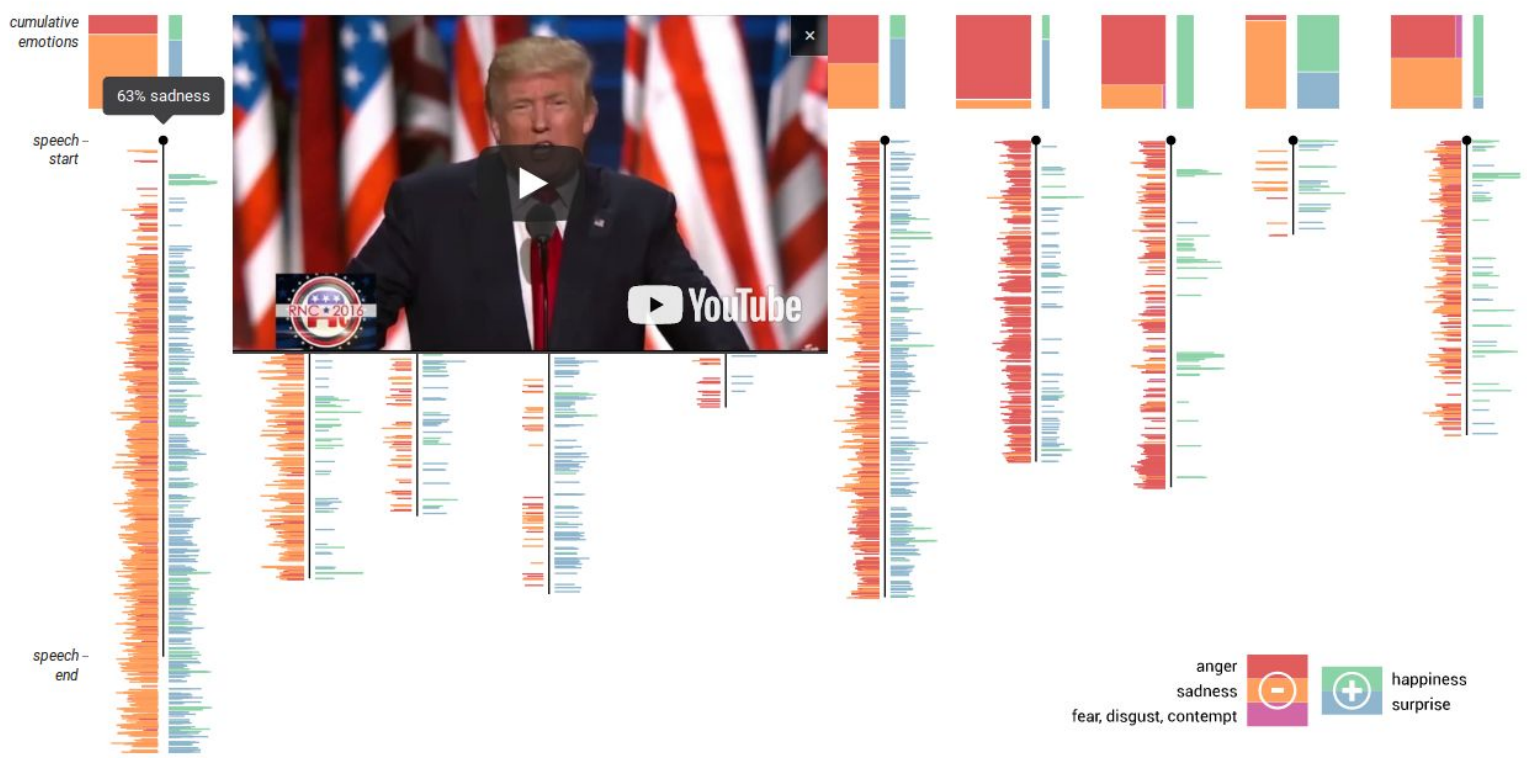
Eine Recherche von BR Data und SPIEGEL Data



web.br.de/interaktiv/erhoehtes-risiko



City	Date	Attendees	Key Event
Cleveland OH	Jul 21	21,000 attendees	Republican National Convention
Daytona FL	Aug 03	10,000 attendees	Trump claims general election is rigged
Charlotte NC	Aug 18	5,000 attendees	Day after Trump hires new campaign manager
Phoenix AZ	Aug 31	4,200 attendees	Immigration focus; Trump meets with Mexico's president
Des Moines IA	Sep 13	1,600 attendees	Attacks Hillary for 'deplorables' comment
Detroit MI	Sep 30	6,000 attendees	USA Today declares Trump unfit for presidency
Cincinnati OH	Oct 13	17,000 attendees	Numerous allegations surface of sexual assault
Manchester NH	Oct 28	2,000 attendees	FBI reopens Clinton probe after new emails found
New York NY	Nov 09	2,800 attendees	Victory speech - private, invitation-only event
Baton Rouge LA	Dec 09	5,000 attendees	CIA says Russia intervened to help Trump win



emotions.periscopic.com

#abgasalarm



Video: Abendschau | 12.12.17 | Robin Avram | Sudiogast: Regine Günther | Bild: rbb/C. Winkler/M. Springmann

Stickoxid-Messreihe von rbb|24 und TU Berlin

An diesen Berliner Straßen herrscht Abgasalarm

12.12.17 | 08:45 Uhr

Mehr Berliner als bislang bekannt sind zu hoher Stickoxid-Belastung ausgesetzt - das zeigen Messungen von rbb|24 und der TU Berlin. Das Problem ist flächendeckend: An 73 Standorten liegen die Werte über dem Grenzwert. Unsere Datenanalyse zeigt die Belastung in Ihrer Nähe.

[rbb24.de/politik/thema/2017/abgasalarm/beitraege/abgasalarm-Stickoxid-Werte-in-Berlin-flaechendeckend-zu-hoch.html](https://www.rbb24.de/politik/thema/2017/abgasalarm/beitraege/abgasalarm-Stickoxid-Werte-in-Berlin-flaechendeckend-zu-hoch.html)



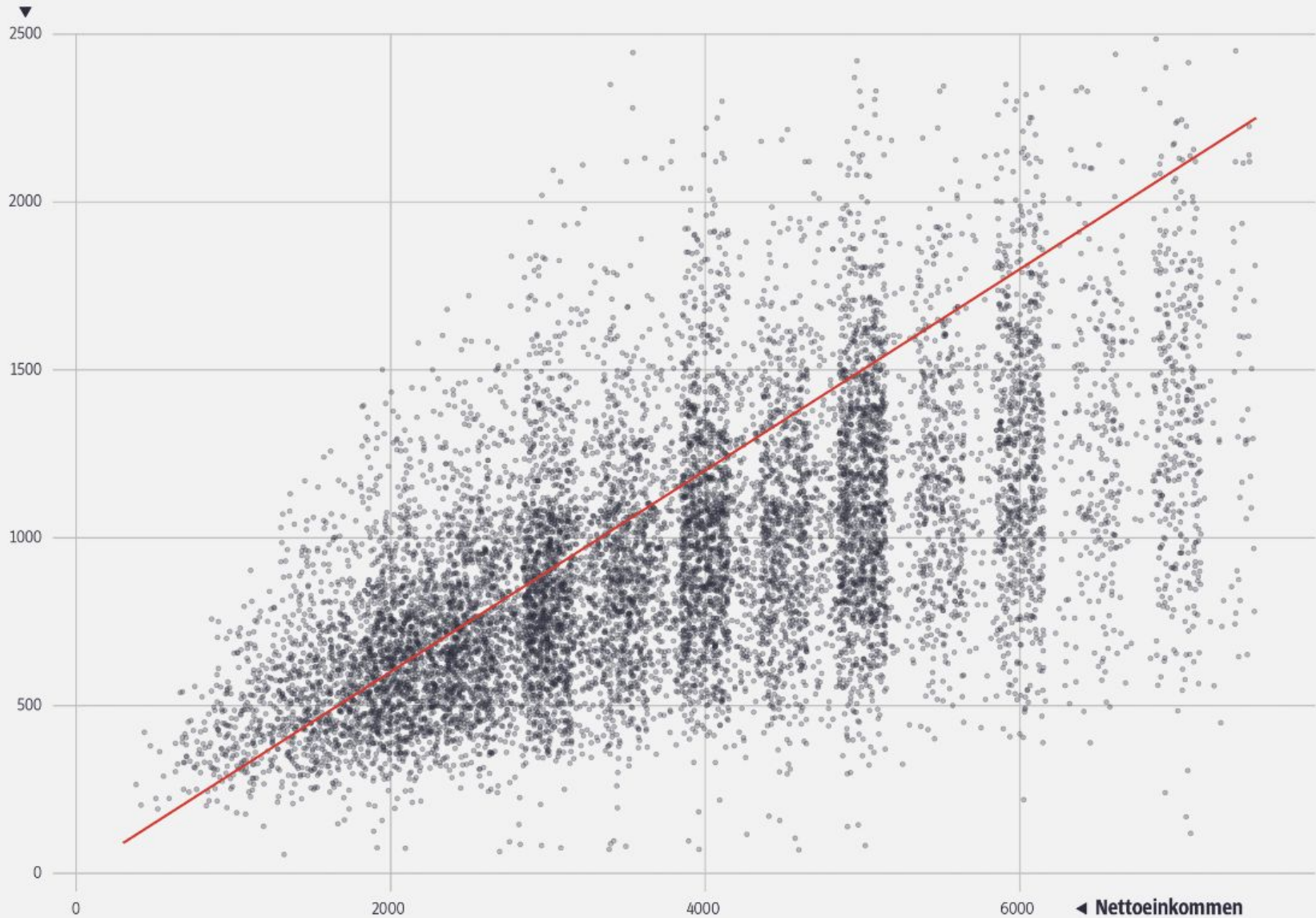
Einsteigen, bitte!

Noch nie gab es in Deutschland so viele Pendler wie heute. Sind Sie einer von ihnen? Dann ist das *Ihre* Geschichte.

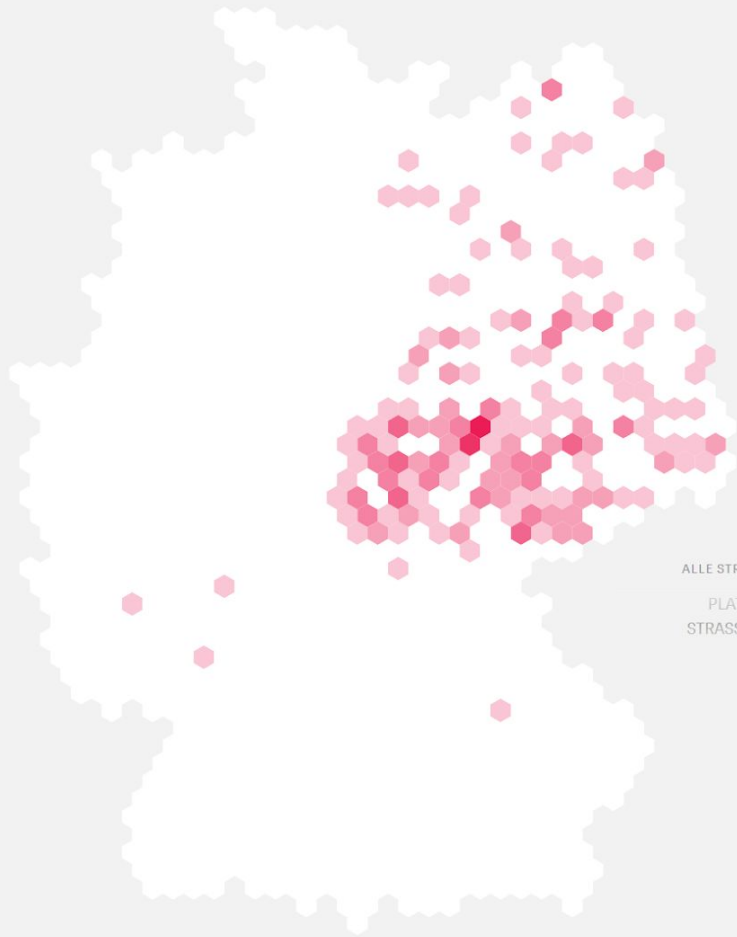


[spiegel.de/wirtschaft/verkehr-so-pendelt-deutschland-zu-arbeit-a-1187172.h](https://www.spiegel.de/wirtschaft/verkehr-so-pendelt-deutschland-zu-arbeit-a-1187172.h)

Kaltmiete mit Nebenkosten



projekte.sueddeutsche.de/artikel/wirtschaft/miete-wohnen-in-der-krise-e687627/



ALLE STRASSENAMEN MIT ...
PLATZ DER **EINHEIT**SGEMEINDE
STRASSE DER **EINHEIT**
EINHEIT german

Straßenbilder

Mozart, Marx und ein Diktator

25. JANUAR 2018

Straßennamen zeichnen Bilder unserer Erinnerung.
Ob Römer, Kaiser, Sozialisten: Alle hinterließen Spuren.
Wir zeigen Muster, die sich in den 450.000 Namen
verbergen.

zeit.de/feature/strassenverzeichnis-strassennamen-herkunft-deutschland-infografik

Stärken und Schwächen

Datenjournalismus - Stärken

- Evidenzbasierte Berichterstattung
- Durch Analyse viele Geschichten in Datensatz auffindbar und weiterdrehbar
- Stärkerer Zuschnitt auf Nutzer möglich (“Regelmäßige Staus in deiner Umgebung”)
- Visuelles Erzählen statt Textwüsten
- Teamwork zwischen davor oft getrennten Bereichen (Journalisten, IT, Grafik)

Datenjournalismus - Schwächen

- Bestimmte Daten oft schwer zu bekommen
- Viele Tools für viele verschiedene Aufgaben
- Gefahr, Fehlschlüsse zu ziehen
- Schlecht umgesetzte Visualisierung lassen Nutzer ratlos zurück
- Zeitaufwand eher hoch
- Teamwork: Journalisten allein kommen an bestimmter Stelle an Grenzen

Das Datenportal für Deutschland

Open Government: Verwaltungsdaten transparent, offen und frei nutzbar

Nach Datensätzen, Dokumenten und Apps suchen

Suchen

Erweiterte Suche

Kartensuche

Veranstaltungen und Termine zu Open Data

Sind Sie interessiert an Veranstaltungen und Terminen rund um Open Data? In unserer Terminübersicht finden Sie eine laufend aktualisierte Übersicht über Veranstaltungen in Deutschland oder auch darüber hinaus. Gucken Sie doch mal [rein](#).

Stöbern Sie in diesen Kategorien



Bevölkerung
(1221)



Bildung und
Wissenschaft
(1376)



Geographie, Geologie
und Geobasisdaten
(2231)



Gesetze und Justiz
(349)



Gesundheit
(766)



Infrastruktur, Bauen
und Wohnen
(3521)



Kultur, Freizeit, Sport
und Tourismus
(198)

Wo und wie finde ich Daten?

Datenportale

- Daten werden bereitgestellt (“Open Data”)
- Staatliche Stellen (v.a. Statistische Ämter), NGOs, Internationale Organisationen
- Beispiele: govdata.de, Eurostat, regionalstatistik.de - künftig: datengui.de
- Daten per Stichwort suchen, herunterladen
- Oder googeln: (+ filetype:xls site:abc.de)

Informationsfreiheit

- Informationsanspruch an Ämter und Behörden (nach Bundes-IFG oder Landesgesetz - manche Bundesländer noch immer ohne)
- Behörde hat ein bis zwei Monate Zeit
- Auch Daten anfragbar
- Hilfreich dabei: [fragdenstaat.de](https://www.fragdenstaat.de)

Websites abgrasen

- Die Tabelle ist brauchbar, aber weit und breit kein Download-Menü
- Lösung: So genanntes Scraping - Inhalte über Browser-Plugin oder Skripte speichern
- Browser-Plugin - einfach (**Tabletools2**)
- Skript: Viel Einarbeitung oder Programmierer

Klassisch: Anrufen

- Viele Datensätze sind nicht auf Websites zu finden - doch es gibt sie durchaus
- Gerade in statistischen Behörden freuen sich Mitarbeiter über Nachfrage und wenn Journalisten Daten wollen
- Anruf genügt, oft bekommt man noch hilfreiche Zusatz-Infos

Wie bearbeite ich Daten?

Daten-Szenarien

- Bestfall: Daten sauber und als Excel-Datei
- Okay: Excel, aber dreckig
- Mittel: PDF, aber sauber
- Unschön: Gescanntes PDF
- Auch unschön: Daten auf Website, nicht herunterladbar

Wichtigste Werkzeuge

- Excel bzw. Libre Office - Alltagswerkzeug
- PDFs umwandeln mit cometdocs.com oder Abbyy Finereader (OCR) - abbyy.de
- Visualisierung
 - Charts & Karten: Datawrapper - datawrapper.de
 - Geo-Tool QGIS (mit Lernkurve) - qgis.org
- Umfassende Datensäuberung mit Open Refine - openrefine.org

**Wie komme ich zur
Geschichte?**

Muster finden und erkennen

- Mit konkreter Frage starten - überprüfen
- Suche nach geographischen Mustern
- Suche nach Ähnlichkeiten / Unterschieden
- Top 10 als Ausgangspunkt - evtl. Bottom 10
- Pivot Tabellen: Aggregation
- Abweichung vom Mittelwert (z.B. bei Durchschnittsalter auf Landkreisebene)

Wie kläre ich, ob was dran ist?

- Frage(n) an Daten beantwortet?
- Ergebnis(se) herausarbeiten
- Im Datensatz gegenchecken
 - Ergebnisse in größerem Datensatz “nachbauen”
 - Zwischenergebnisse reproduzieren und summieren
- Mit offiziellen Statistik-Veröffentlichungen abgleichen: Ergebnis identisch?
- Experten Ergebnis zeigen und besprechen

Nach dem Überprüfen

- Wenn alles passt, Daten für Veröffentlichung aufbereiten
- (Immer Kopien der Originale behalten)
- Wie visualisieren? Seiten wie **Dataviz Catalogue** geben Hilfe
- Visualisierung und Text/Radio-/FS-Beitrag

Praktischer Teil

Remove to improve the **data tables** edition

Created by **Darkhorse Analytics**

www.darkhorseanalytics.com

Mit Excel und Co. arbeiten

Grundwerkzeug Excel & Co

Was kann Excel, was ich nicht kann?

- Ans große Ganze denken
(Summen, Produkte etc.)
- Unwichtiges ausblenden (Filtern)
 - Summe etc. des Gefilterten
- Wichtiges hervorheben
(Bedingte Formatierung)
- Prozentrechnen

Möglichkeiten von Excel & Co

- Suchen & Ersetzen
 - (Umweg über Word!)
- Teile (rechts, links, frei) extrahieren
- Zusammenfügen aus mehreren Zellen
- Pivot Tables helfen beim strukturierten Durchsuchen / Erforschen

Tabellenkalkulations-Basics

- Mehrere Datenblätter
- Jede Zelle hat “Adresse” aus
 - Angabe für Spalte (Buchstaben) und
 - Angabe für Zeile (Ziffern)
 - [A2] = 1. Spalte, 2. Zeile
- Formeln beginnen immer mit [=...]
- “Enter” oder “Tabulator” zur nächsten Zelle
- “F2”, um Zelle und Bezüge anzuzeigen

Excel-Praxis

Datei: parteispenden_2016.xlsx

Ziele

- Teile darstellen und berechnen
- Ergebnisse finden + anschaulich machen
- Einfache Rechenoperationen

Excel: Filtern

Große Datenmengen schnell überschaubar,
wenn nur wesentliche Teile durch Filter
dargestellt werden (Vorteil: nichts gelöscht)

[Menü “Daten”, Untermenü “Filter”]

- Autofilter - wenige Klicks
- Spezialfilter - Kombination von Kriterien

Excel: Filtern II

- Auto-Filter mit Dropdown-Menü
 - z.B. nur bestimmte Postleitzahlen, Städte oder Bundesländer
 - Auch mehrere z.B. PLZ durch Häkchen wählbar
- Benutzerdefinierte Filter (in Auto-Filter)
 - z.B. um “Frankfurt”, Frankfurt/Main, Frankfurt a.M., Frankfurt a. Main, Frankfurt (Main) zu finden und durch “Frankfurt” zu ersetzen

Excel: Bedingte Formate

In einer Spalte lassen sich bestimmte Werte farblich hervorheben mit Bedingter Formatierung

Beispieldatei: "Parteispenden"

- Ganze Spalte markieren
- Unter "Format" bzw. "Start"* , "Bedingte Formatierung"
- Wert festlegen, z.B. "größer oder gleich" 25000, Farbe für Zellschattierung unter "Muster" anklicken
- Alle Werte über 25000 farblich hinterlegt

* Je nach Excel-Version: Vor 2007 "Format", ab da "Start"

Libre Office: Bedingte Formate

Bei Libre / Open Office sind bedingte Formate etwas sperriger anzulegen

- Ganze Spalte markieren
- Unter “Format”, “Bedingte Formatierung”, “Bedingung”
- Wert festlegen, z.B. “größer oder gleich” 25000, “Neue Vorlage” für Einfärbung, Namen geben und im Reiter “Hintergrund” Farbe für Zellschattierung auswählen, Vorschau zeigt Färbung an
- Alle Werte über 25000 farblich hinterlegt
- Weitere Einfärbungen über Knopf “Hinzufügen”

Excel: Pivot (Parteispenden)

Weiter anwenden auf Parteispenden 2016

- “Einfügen”, “PivotTabelle”, “PivotTable erstellen”
- Unter “PivotTable-Felder” “Betrag” auswählen, es landet automatisch unter “Werte” als “Summe von Betrag”
- In “Zeilen” “Partei” ziehen
- Ggf. in Spalten “Pers./Inst.”
- Bereichsfilter z.B. für “Ort” - aber auch als weitere Zeile möglich, per Drag&Drop Reihenfolge bestimmbar

Excel: Pivot (Parteispenden) II

- Trick, um sowohl Prozente als auch absolute Zahlen anzeigen zu lassen:
 - “Betrag” ein zweites Mal in “Werte” ziehen,
 - anschließend dort im Kontextmenü auf “Wertfeldeinstellungen” gehen
 - den Tab “Werte anzeigen als” wählen und dort z.B. “% des Gesamtergebnisses” nehmen - je nach Kontext auch % der Zeilen- oder Spalte sinnvoll

1 Daten hochladen ✓

2 Prüfen & Beschreiben ✓

3 Visualisieren

4 Veröffentlichen & Einbetten

Diagr.-typ

Anpassen

Beschreiben

Design



Balken



Balken
(separat)



Balken
(gestapelt)



Bullet Balken



Dot-Plot



Intervalle



Pfeile



Säulen



Säulen
(gruppiert)



Säulen
(gestapelt)



Flächen



Linien



Torten-
diagramm



Donut-
diagramm



Mehrere
Torten



Mehrere
Donuts



Scatterplot



Sitzverteilung



Kurze Tabelle



Lange Tabelle

Archivierte Diagr.-typen:

Parteispenden 2016

SPD 6.372.437 €

Grüne 5.168.142 €

CDU 5.086.242 €

Die Linke 2.759.514 €

FDP 1.571.462 €

CSU 1.473.316 €

AfD 328.735 €

Quelle: Deutscher Bundestag

MODUS

Web

Print

GRÖSSE

548 x 285



FARBBLIND-TEST



Norm



Deut



Prot



Trit

Praxis: Datawrapper

Über Datawrapper

- Gezielt vereinfacht in Möglichkeiten
- Verwendet Best Practices der Visualisierung
- Kostenloser Account ermöglicht Einbindung in Website - bei mehr als 10.000 Aufrufen kostenpflichtiger Account nötig
- Anpassung an Corporate Design möglich
- Account erstellen unter **datawrapper.de** mit direkt abrufbarer Mailadresse!

Ablauf: Datawrapper

Vier simple Grundschritte

- Einfügen der Daten
- Überprüfen + evtl. Anpassen
- Auswahl der Grafik-Möglichkeiten
 - Enthält einige Unterschritte
- Fertige interaktive Grafik mit Einbett-Code

(Bei den Karten-Optionen (“Neue Karte”)
weicht die Bedienung etwas ab)

Übung: Charts erstellen mit Daten

Spenden nach Parteien (und Typ)

Ziel der Übung: Daten in verschiedenen Typen visualisieren

(Dafür über Excels Pivot-Funktion Spenden nach Parteien + Typ zusammenfassen)

Datei: parteispenden_2016.xlsx

Datawrapper: Bedienung

- Daten per Copy & Paste (oder Excel-Import) reinladen, Spalten ausblendbar oder um Suffix (z.B. €) erweiterbar
- Verschiedene Balken-, Säulen-, Kreis-, Linien-Diagramme u.a. auswählbar
 - Besonderheiten: Dot-Plot, Scatterplot, “Mehrere Torten/Donuts”
- Tipp: “Transponieren”, wenn Darstellung nicht klappt
- Bei “Anpassen” finden sich je nach Diagramm-Typ wenige oder sehr viele Optionen
- Bei “Beschreiben” gibt es Überschrift, Teaser-Text (“Beschreibung”) und Fußnoten (“Hinweise”)

Datawrapper: Anpassen

- Nach Einfügen der Daten möglich, einzelne Spalten auszublenden
- Gleich Diagrammtyp “Balken” möglich
- Unter “Anpassen” Zahlen-Spalte für Visualisierung
- Auch “Gruppierung” möglich, falls weitere Spalte mit sinnvollem Merkmal vorhanden, z.B. Europa, Amerika
- “Darstellung” besonders wichtig, einzelne Farben vergebbar, z.B. jede Partei individuell, “Dickere Balken” sowie “Balkenhintergrund einfärben” - für Einordnung geeignet

Datawrapper: Beschreiben

- Titel: Steht oben
- Beschreibung: folgt darauf
- Hinweise: Fußnote, erscheint unten kursiv
- Datenquelle: Angabe der Quelle, nicht verlinkt außer Feld für Link wird ausgefüllt
- Element hervorheben: variiert je nach Chart-Typ, in manchen Fällen klare Hervorhebung, in anderen nicht

Übung: Flächen einfärben mit Daten

Durchschnittsalter auf Kreisebene

Ziel der Übung: Karte von Landkreisen
erstellen mit Regional-Datensatz

Datei: durchschnittsalter_kreise_2016.xlsx

Datawrapper: Flächen einfärben

Grundablauf

- Nach “New Map” / “Neue Karte”
“Choroplethen-Karte” auswählen
- “Deutschland Distrikte” suchen und wählen
- Daten hochladen per “Importiere”
- Farbskala auswählen

Datawrapper: Flächen - Details

Reiter “Anpassen”

- Spalte für Einfärbung, Farbschema und Farb-Abstufung (“Stops”) auswählbar
- “Tooltips” - über Menü Infos auswählen, die beim Darüberbewegen angezeigt werden
- “Kartenbeschriftung” zeigt Namen für z.B. Kreis an
- Legende wird automatisch generiert, ist abwählbar; Bezeichnung erstellbar, Position unter “Anordnung”

Übung: Punkte auf Karte anzeigen

Parteispenden visualisieren

Ziel der Übung: Auf Karte einzelne Punkte inklusive Infofeld darstellen können

(Auswahl der höchsten 100 Spenden)

Datei: parteispenden_2016_geo_top100.xlsx

Datawrapper: Karten erstellen

Vier simple Grundschritte

- Beginn mit Auswahl der Karte bzw. Region
- Daten i.d.R. mit “Importiere deinen Datensatz” einfügen
 - Dabei mehrere Spalten möglich, auch Länder-Name
- Zuordnen: ISO-Code- und Werte-Spalte auswählen
- Feintuning der Karte
 - Tooltips, Legende, Beschreibung
- Fertige interaktive Karte mit Einbett-Code

Datawrapper: Daten importieren

- Nach “New Map” / “Neue Karte”
“Symbolkarte” auswählen
- Im nächsten Schritt Hintergrund auswählen,
 - z.B. Deutschland, Bundesländer, Landkreise
- “Importiere deinen Datensatz” (u. links)
 - Per Copy & Paste oder als CSV (nur 501 Einträge inkl. Dachzeile möglich) - sinnvoll aber viel weniger
 - Spalte für Längen- und Breitengrad auswählen

Datawrapper: Punkte darstellen

Bereich “Beschreiben & Visualisieren”

- Mehrere Grundoptionen unter “Anpassen”
 - Symbol-Form (u.a. Punkt, Kreis, Hexagon)
 - Spalte für Größe des Symbols
 - Farbe des Symbols
 - Gruppierung
 - Kartenlegende
- “Beschreiben” für Überschrift, Quellenangabe etc.
- “Design” für Grundanpassung/Auswahl

Datawrapper: Ressourcen

Datawrapper-Blog: blog.datawrapper.de

- What to consider when creating choropleth maps:
blog.datawrapper.de/choroplethmaps
- How to choose a color palette for choropleth maps:
blog.datawrapper.de/how-to-choose-a-color-palette-for-choropleth-maps
- Introducing Cartograms in Datawrapper:
blog.datawrapper.de/cartograms

Datawrapper River: river.datawrapper.de

Visualisierungen zum ausdrücklichen Weiterverwenden

Michael Hörz

www.michael-hoerz.de

hoerz@michael-hoerz.de

twitter.com/data_meining