

Michael Hörz

Datenjournalistische Nutzung üblicher und unüblicher urbaner Daten

Dieser Beitrag beschäftigt sich mit der systematischen journalistischen Nutzung üblicher und unüblicher Datenquellen in Berlin. Als übliche sind statistische Daten zu verstehen, die etwa das Amt für Statistik Berlin-Brandenburg zur Verfügung stellt, ebenso solche Daten aus den einzelnen Senatsbehörden. Unübliche Quellen sind etwa Internetangebote wie das Portal Airbnb. Vorgestellt wird zum einen das Projekt Airbnb vs. Berlin, das Studenten der FH Potsdam umgesetzt haben, zum anderen das Projekt sowird-berlin.de, das Volontäre der Evangelischen Journalistenschule in Berlin realisiert haben.

Michael Hörz

Politikwissenschaftler (M. A.), seit 2012 freier Datenjournalist und Trainer, davor Online-Redakteur bei Rundfunk Berlin-Brandenburg und Tagesspiegel. Schwerpunkte: Geodaten, Transparenzdaten, Open Data, Informationsfreiheit.

✉ hoerz@michael-hoerz.de

Schlüsselwörter:

Datenjournalismus – Visualisierung – neue Datenquellen – Airbnb

Was ist Datenjournalismus?

Die Branche operierte für längere Zeit ohne eine spezifische Definition, demzufolge ist das Verständnis von Datenjournalismus recht breit. Verwendet wird der Begriff seit 2005, im breiteren Kontext seit 2010. In einer minimalen Definition geht es darum, „mit Daten Journalismus zu betreiben“. Die Abgrenzung zu anderen Bereichen des Journalismus ergibt sich daraus, dass Journalisten mit strukturierten Daten arbeiten (Kayser-Bril 2015). Das bloße Erwähnen einer oder weniger statistischer Werte in einem größeren Fließtext begreifen Datenjournalisten nicht als Beispiel für ihr Metier.

Eine erweiterte Definition liefern die Journalismusforscher Royal und Blasingame auf Basis von Literaturlauswertung und der systematischen Befragung von Datenjournalisten. Demzufolge ist Datenjournalismus „a process by which analysis and presentation of data are employed to better inform and engage the public. Its roots are in the fields of computer-assisted and investigative reporting, but data journalism products may add engagement through customization and user contribution made possible by web development and programming techniques.“ (Royal und Blasingame 2015).

Grundsätzlich zeichnet sich Datenjournalismus durch die Präsentationsform und die Analyse und Einordnung aus: Essenziell sind Visualisierungen, überwiegend in interaktiver Form, verbunden mit Analysen und Bereitstellen des Kontexts. Eine reine Darstellung von Datenpunkten auf einer Karte oder eine aufwändige Visualisierung fällt nicht unter Journalismus. Zwar fügt die Möglichkeit, die datenjournalistischen Ergebnisse selbst in vielfältiger Weise interaktiv zu explorieren, einen Mehrwert hinzu, doch das Aufzeigen der Geschichte in und hinter den Daten spielt die wesentliche Rolle.

Die Veröffentlichung der genauen Datenquelle inklusive Verlinkung gehört zur guten Praxis des Datenjournalismus (z. B. New York Times, Los Angeles Times, Guardian, Zeit Online, Spiegel Online). Im besten Fall veröffentlichen Datenjournalismus-Teams auch die Daten in der konkret von ihnen verwendeten Form. Damit ermöglichen sie einerseits, ihre Arbeit zu überprüfen, andererseits können andere von der Vorarbeit profitieren. Insbesondere wenn ein Medium nicht allgemein verfügbare Daten nutzt oder sogar selbst Daten erhoben hat, ist der Mehrwert für Dritte erheblich. Darüber hinaus können Nachnutzer solcher vorbearbeiteter Daten Zeit sparen, denn die Erfahrung zeigt, dass im Datenjournalismus oft ein erheblicher Anteil (bis zu 50 Prozent) für die Umwandlung und

Säuberung von Daten verwendet werden muss. Häufig liegen die gewünschten Daten nicht in maschinenlesbarer Form vor und/oder sind schlecht und inkonsistent formatiert.

Projekt A: „Airbnb vs. Berlin“

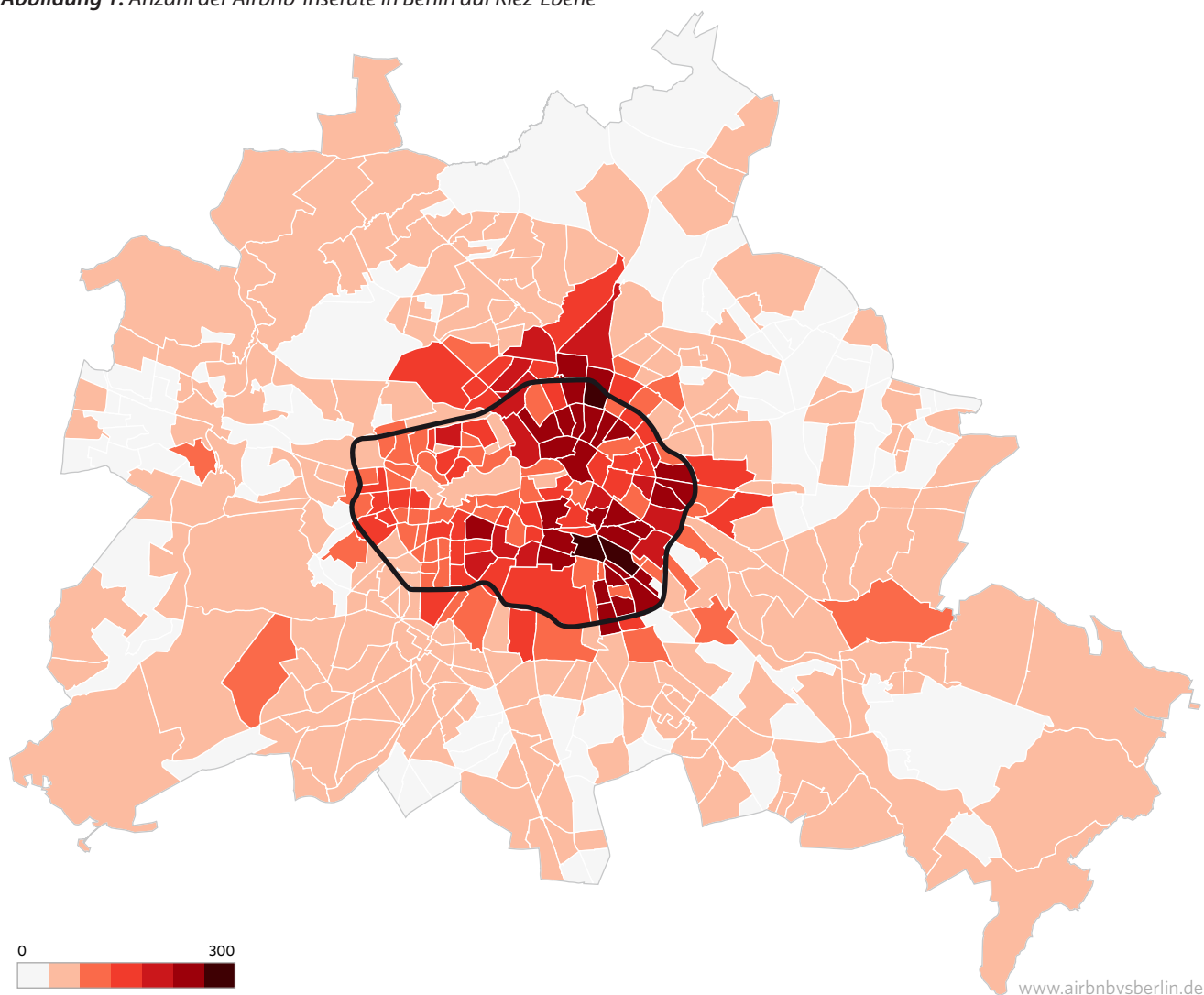
Rahmen

Das Projekt „Airbnb vs. Berlin“ entstand im Wintersemester 2014/15 in der Lehrveranstaltung „From Reading to Exploring – Visuelles Geschichtenerzählen im Datenjournalismus“ am Bereich Interfacedesign der FH Potsdam. Dozenten waren Jan-Erik Stange (Akademischer Mitarbeiter) und Michael Hörz als Lehrbeauftragter. Die Studierenden bekamen in der ersten Semesterhälfte Einführungen in Datenjournalismus und Datenvisualisierung, Praktiker wie etwa Sascha Venohr von Zeit Online hielten Gastvorträge. In der zweiten Semesterhälfte ging es um die Umsetzung konkreter Projekte. Dazu fanden zwei Workshops mit zahlreichen Datenjournalismus-Experten statt, die in einer Art „Speeddating“ alle studentischen Teams berie-

ten. Am Ende des Seminars entwickelten die Teams insgesamt sechs datenjournalistische Projekte, von denen hier dasjenige mit städtestatistischem Hintergrund herausgegriffen wird.

Das Projekt nähert sich der Debatte über die Vermietung von Privatwohnungen über Internetportale wie Airbnb mittels einer Datenauswertung. Der Titel lautet „Airbnb vs. Berlin – Was sagen die Daten?“. Die Studenten sind Alsino Skworonek, Lucas Vogel und Jonas Parnow. Ihr Projekt veröffentlichten sie unter www.airbnbvsberlin.de, wegen des starken internationalen Interesses auch in einer englischen Version unter www.airbnbvsberlin.com. Das Veröffentlichen auf einer dezidierten Website in Kombination mit auf anderen Websites einbettbaren interaktiven Grafiken hat das Projekt sehr erfolgreich gemacht. Zahlreiche, auch überregionale Medien haben über das Projekt berichtet. Sowohl Gegner wie Befürworter von Airbnb verweisen auf „Airbnb vs. Berlin“. Die Seite ist für mehrere Preise nominiert worden und wurde bereits ausgezeichnet. Grundsätzlich empfiehlt es sich, das Projekt online anzusehen, da viele Details nur in der interaktiven Version nutzbar sind.

Abbildung 1: Anzahl der Airbnb-Inserate in Berlin auf Kiez-Ebene



Die Karte zeigt die Anzahl aller Airbnb-Inserate in Berlin auf Kiez-Ebene. Je dunkler die Einfärbung, desto mehr Angebote gibt es. Zur Orientierung ist der „Große Hundekopf“ (S-Bahn-Ring) eingezeichnet.

Daten

Die Website airbnbsberlin.de verwendet als Hauptbestandteil Daten des Portals airbnb.com. Das Unternehmen stellt diese Daten nicht zum Download bereit, weder als gesammelte historische Daten („Dump“) noch über eine konfigurierbare Web-Schnittstelle (API). Da jedoch ein Kernelement von Airbnb das Anzeigen aktueller Übernachtungsangebote im Browser und in Smartphone-Apps ist, bekamen die Studenten beim zweiten Projektentwicklungs-Workshop den Tipp, die Abfragen mittels der Anwendung „Charles“ zu untersuchen. Damit ließ sich sehr schnell die interne API-Struktur von Airbnb erschließen, sodass die Studenten zu einem Stichtag am Ende Januar 2015 sämtliche Angebote für Berlin mit allen verfügbaren Daten herunterladen konnten.

Beachtet werden muss, dass Airbnb nicht die exakte Geoposition der inserierten Wohnungen übermittelt, sondern eine Anonymisierung vornimmt: Eine zufällige Koordinate in einem Radius von ca. 200 Metern um die eigentliche Position der Wohnung wird angegeben. Demzufolge sind Abweichungen etwa bei der Zuordnung zu einem kleinräumigen Gebiet möglich.

Als weitere Daten zur Kontextualisierung dienten die Umriss der Ebene Planungsräume der Lebensweltlich Orientierten Räume (LOR). Diese sind verfügbar beim Amt für Statistik Berlin-Brandenburg in dessen Open-Data-Bereich.

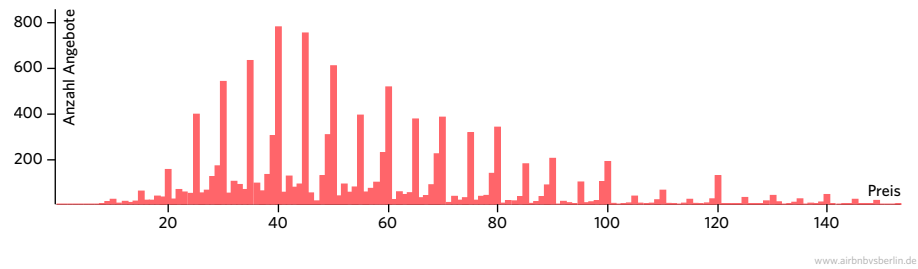
Analysen und visuelle Aufbereitungen

Die Website nimmt zahlreiche Analysen vor, unter anderem zeigt eine Choroplethenkarte die Zahl der Angebote je Lebensweltlich Orientiertem Raum. Auffällig ist dabei eine Konzentration auf innerstädtische Bereiche und auf bestimmte Kieze. Die meisten Angebote gibt es im Reuterkiez in Neukölln (476), in der Umgebung des Helmholtzplatzes in Prenzlauer Berg (345) und im Gräfekiez in Kreuzberg (314). Die Studenten errechneten für die drei Spitzenreiter eine Dichte von 16 bis 17 Angeboten pro 1.000 Einwohner. Für die Choroplethenkarte verwendeten die Studenten das Online-Tool CartoDB (siehe Abbildung 1).

Auch die Preisverteilung der Angebote ist aufschlussreich. So kostet ein durchschnittliches Angebot in Berlin 55 Euro pro Nacht, gut 7.000 Angebote (von 11.000) kosten weniger als 55 Euro pro Nacht. Das folgende Histogramm zeigt die Preisverteilung (siehe Abbildung 2).

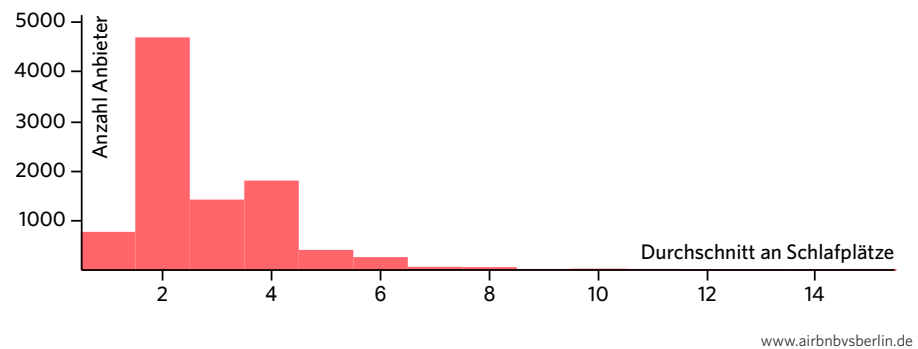
Eine Einordnung mit weiteren deutschen Großstädten zeigt eine ähnliche Preisverteilung, in Hamburg, Köln und Frankfurt

Abbildung 2: Preisverteilung der Berliner Airbnb-Angebote



Die Grafik zeigt die Preisverteilung der Berliner Airbnb-Angebote in Euro. Für jeden Preis (x-Achse) wird die Anzahl der angebotenen Unterkünfte dargestellt (y-Achse).

Abbildung 3: Anbietern nach Anzahl der bereitgestellten Schlafplätze



Die Grafik zeigt die Menge an Anbietern, die eine bestimmte Anzahl an Schlafplätze zur Verfügung stellen.

gibt es die meisten Wohnungen für etwas weniger als 50 Euro pro Nacht, der Durchschnittspreis ist ähnlich (55–60 Euro). Einzig München reißt nach oben aus, hier finden sich mehr höherpreisige Übernachtungsangebote, der Durchschnittspreis liegt mit 71 Euro ebenfalls höher.

Die Daten zeigen auch eine deutlich andere Nutzung durch die Anbieter als ursprünglich von Airbnb intendiert, nämlich ungenutzten Raum der eigenen Wohnung zu vermieten. So bieten in Berlin zehn Prozent aller Anbieter mehr als eine Wohnung an (rund 1.200 Anbieter). Die zehn größten Anbieter haben 240 Wohnungen im Angebot. Eine tendenziell gewerbliche Nutzung legt auch die Zahl der angebotenen Schlafplätze nahe. Während 46 Prozent der Anbieter einen oder zwei Schlafplätze offerieren, hat der Rest mehr Schlafplätze im Angebot. Zehn Anbieter vermieten sogar 10 Schlafplätze. Die Grafik zeigt allerdings, dass der größere Teil zwischen drei und sechs Plätze bereitstellt (siehe Abbildung 3).

Mithilfe der Geopositionen ermittelten die Studierenden, dass sich die Wohnungen der Top-10-Vermieter über ganz Berlin verteilten – einige hatten ein enges Netz innerhalb eines eingegrenzten Gebiets, zahlreiche verteilten ihre Wohnungen aber auch über die ganze Stadt. Darüber hinaus ließ sich feststellen, dass die (Groß-)Vermieter offenbar Wohnungen in noch nicht ganz so populären Gegenden bevorzugen, wo sich Immobilien günstiger und in größerer Zahl erwerben lassen.

Ebenfalls die Geopositionen nutzten die Studenten für eine Analyse besonders populärer Straßen. Sie kartierten – mit

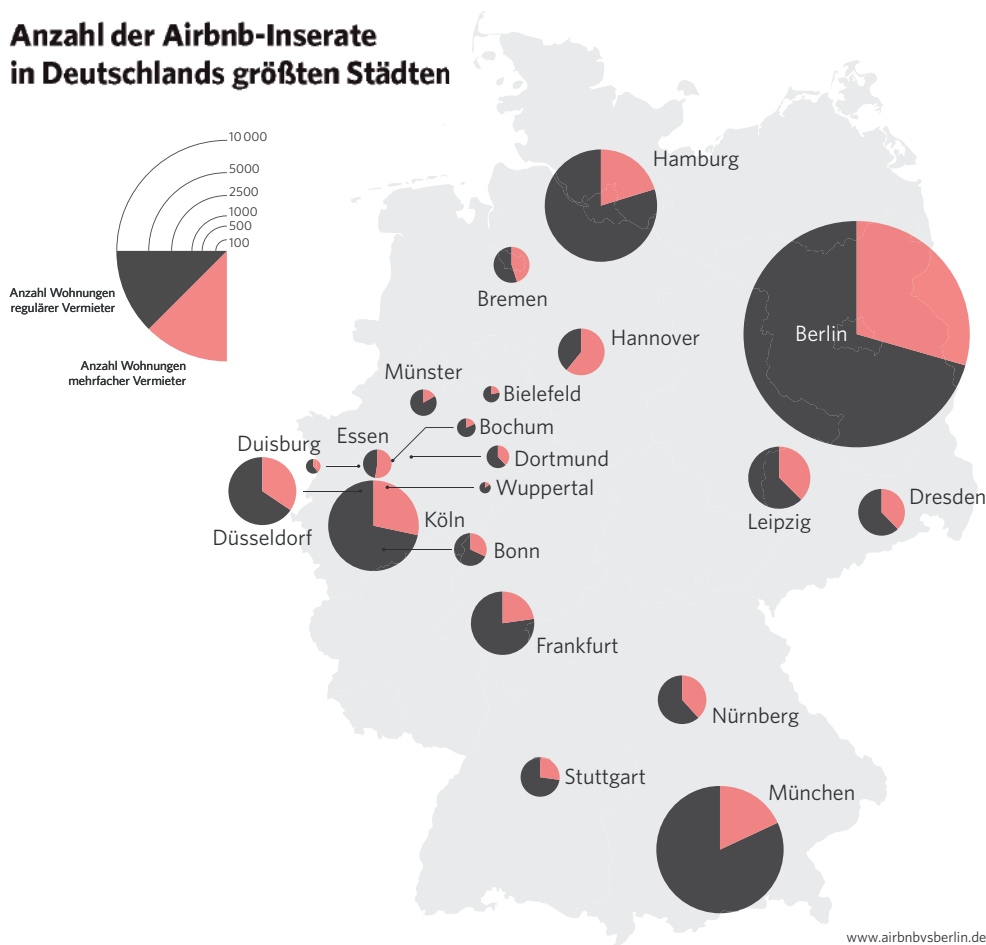
Abbildung 4: Straßen in Berlin mit mehr als 20 Airbnb-Angeboten



Die Karte zeigt alle „Airbnb-Straßen“ in Berlin mit mehr als 20 Angeboten. Je dunkler die Farbe, desto mehr Angebote.

Abbildung 5: Anzahl der Airbnb-Inserate in den zwanzig größten deutschen Städten

Anzahl der Airbnb-Inserate in Deutschlands größten Städten



Die Grafik zeigt einerseits mit der Größe der Kreise die Anzahl der Inserate der zwanzig größten deutschen Städte. Die Aufteilung in den Kreisen zeigt andererseits das Verhältnis von regulär vermieteten Wohnungen und mutmaßlich professionell betriebenen Inseraten.

den gegebenen Einschränkungen – sämtliche Angebote auf Straßenebene. Dabei konnten die Studenten feststellen, dass sich Ferienwohnungen auf touristische Hotspots konzentrieren, die „Szenebezirke“ Mitte, Prenzlauer Berg, Neukölln und Kreuzberg-Friedrichshain. Allein in Sonnenalle und Weserstraße in Neukölln gibt es 200 Inserate (siehe Abbildung 4). Zur Gesamteinordnung verglichen die Studenten für die zwanzig größten Städte Deutschlands die jeweilige Gesamtzahl von Inseraten mit der Anzahl von gewerblichen Inseraten (Anbieter mit mindestens zwei Inseraten). Berlin sticht absolut in beiden Kategorien heraus, mit absolut über 10.000 Angeboten als auch dem Anteil gewerblicher Einheiten (29 Prozent). Auffällig ist dabei auch, dass München absolut gesehen mehr Angebote hat als das deutlich größere Hamburg – der gewerbliche Anteil liegt in beiden Fällen ähnlich hoch (18 und 20 Prozent). Ebenfalls auffällig sind die Städte Hannover und Essen, die mit 60 und 52 Prozent Gewerbeanteil herausstechen. Allerdings ist hier auch die Betrachtung der absoluten Angebote wichtig, denn diese beträgt bei Hannover 446, bei Essen 167. In Frankfurt, Düsseldorf, Leipzig, Dresden und Nürnberg werden zwischen 400 und 1.000 Wohnungen offeriert, unterhalb der obersten zehn sinkt die Zahl rasch. Die letzten fünf auf der Liste bewegen sich zwischen 105 (Dortmund) und 26 Inseraten (Wuppertal) (siehe Abbildung 5).

Vor allem aber die Relation der Inserate zur Einwohnerzahl ist aufschlussreich: Auch hier liegt Berlin mit etwa 300 Inseraten pro 100.000 Einwohnern an der Spitze, während Hamburg mit gut 150 Inseraten pro 100.000 Einwohnern nur halb so viele Angebote hat. München liegt mit etwa 235 Inseraten pro 100.000 Einwohnern an zweiter Stelle.

Einordnung

Als neue Datenquelle ist Airbnb interessant – auch wenn die Daten nicht unmittelbar angeboten werden. Die Studenten berichteten allerdings, dass die Firma entspannt mit dem Auslesen über die API umging. So fragte Airbnb direkt nach der Veröffentlichung, woher die Daten stammten, und bekam die Antwort: „Von Ihrer Seite“. Das Unternehmen ließ es dabei bewenden, und lud die Studenten zu sich ein, um ihr Projekt vorzustellen.

Aus statistischer Sicht handelt es sich zwar um eine Vollerhebung zu einem Stichtag, jedoch gibt es dabei mehrere Caveats. So ist nicht eindeutig auszumachen, ob die Inserate an diesem Tag wirklich im Angebot waren oder nur als grundsätzlich existent angezeigt wurden. Weiterhin ist eine einzelne Tagesprobe nicht sicher repräsentativ. Dazu ist das systematische Erheben der Daten über einen längeren Zeitraum nötig, was die Studenten inzwischen auch tun. Für eine weitergehende Analyse stünde also das Material zur Verfügung. Ebenfalls interessant könnte eine ausführlichere lokale Untersuchung in Städten wie Berlin oder München sein.

Anfang Dezember 2015 hat Airbnb in New York eine erhebliche Menge seiner Daten in anonymisierter Form zur Verfügung gestellt. Noch immer ist kein Online-Abruf möglich, sondern ein Besuch der New Yorker Niederlassung der Firma nötig (Isaac 2015). Trotzdem ist dies eine Änderung der bisherigen Firmenpolitik, die auch für Berlin als Stadt mit den Abstand meisten Angeboten Hoffnung macht.

Projekt B: „So wird Berlin“

Rahmen

„So wird Berlin“ ist zwischen Mai und Juni 2015 an der Evangelischen Journalistenschule in Berlin entstanden. Der 11. Volontärsjahrgang (2015–2017) durchlief im Rahmen seiner Grundausbildung einen dreiwöchigen Kurs „Datenjournalismus“, den der Autor mit der Designerin Cornelia Geissler leitete. Eine solche Dauer ist für ein datenjournalistisches Modul bislang eine Ausnahme. Der Zeitrahmen ermöglichte eine ausführliche Recherche, Analyse und Visualisierung der Daten.

Der Fokus lag auf Berlin um das Jahr 2030, was eine datenbasierte Betrachtung erschwerte – schließlich beruhen die meisten datenjournalistischen Geschichten auf der Auswertung vergangener oder gegenwärtiger Daten. Durch intensive Recherche konnten die Volontäre trotzdem genügend Datenquellen erschließen. Insgesamt veröffentlichten die Volontäre auf der eigens eingerichteten Website www.sowirdberlin.de sechs Geschichten, die die folgende Themen abdeckten:

- Klima („So warm – Wo es 2100 so heiß wie am Mittelmeer wird“)
- Bildung („So knapp – Wie viele Lehrer bis 2023 an den Schulen fehlen werden“)
- Nahverkehr („So mobil – Wo bis 2030 neue Bahnhöfe entstehen“)
- Arbeitsmarkt („So beschäftigt – Welche Jobs bis 2030 entstehen und wegfallen“)
- Demographie („So viele – Welche Kieze bis 2030 schrumpfen oder wachsen“)
- Bauen („So bebaut – Wie viele Kleingärten bis 2020 Wohnungen weichen müssen“)

Das Interaktiv-Team der Berliner Morgenpost unterstützte das Projekt, setzte einige Geschichten ausführlicher um und veröffentlichte diese in Print und Online.

Vorgestellte Teilprojekte

Diese Darstellung wird auf zwei Themen fokussieren, da eine ausführliche Darstellung aller Themen im Rahmen der Veröffentlichung nicht möglich ist. Vorgestellt werden die Themen Klima („So warm“) und Demographie („So viele“). Am Klima-Thema beteiligt waren Florian Zimmer-Amrhein und Florian Haenes, am Demographie-Thema Max Dinkelaker, Martin Nejezchleba, Timo Nicolas und Gesa Steeger.

Daten – „So warm“

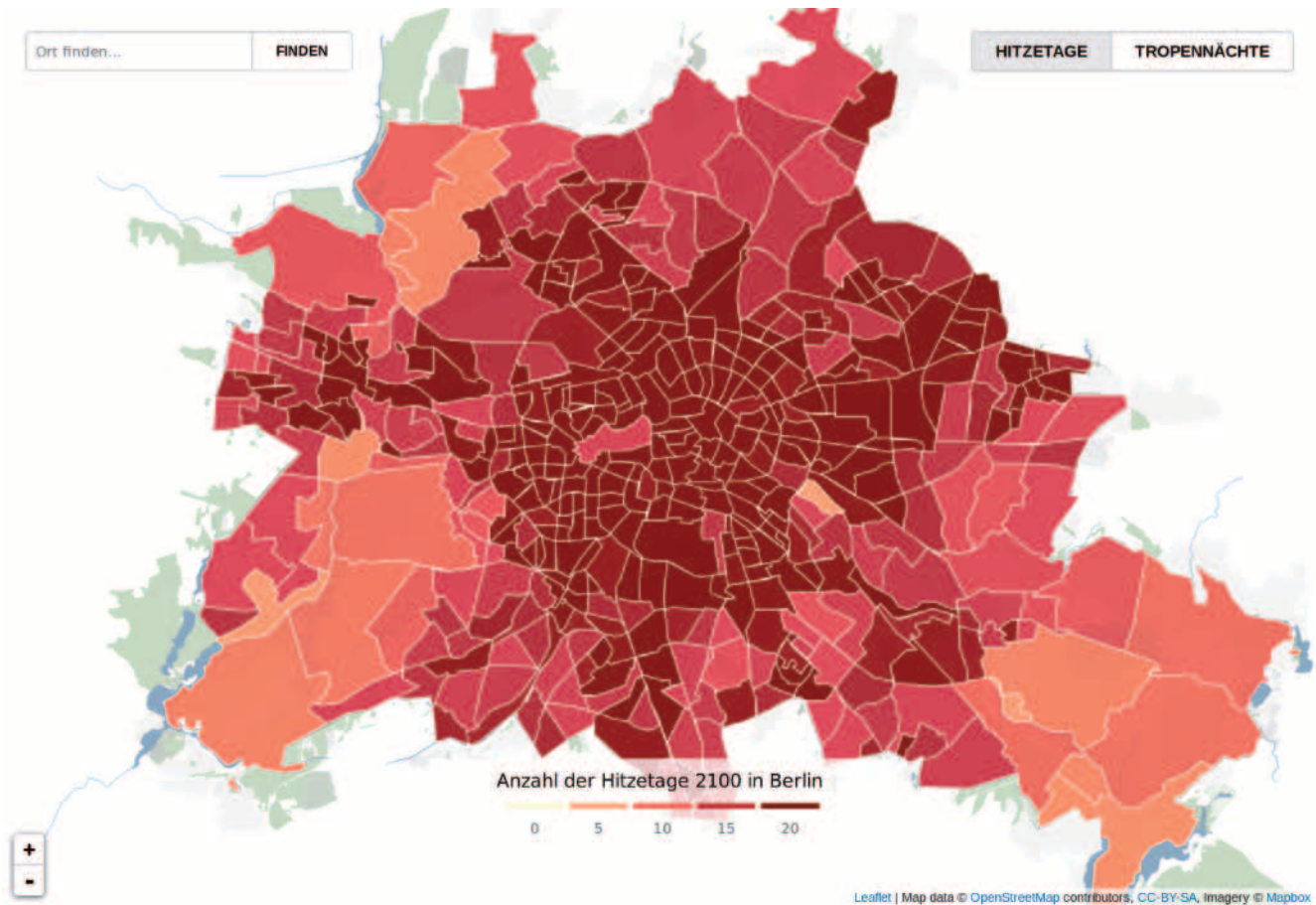
Für Berlin gibt es verschiedene Klimaszenarien und Klimaprojektionen. Im Mittelpunkt des Artikels über Berlins künftiges Klima stehen Daten, die die Berliner Senatsverwaltung für Stadtentwicklung und Umwelt in ihrem Stadtentwicklungsplan (StEP) Klima verwendet. Die Daten wurden von der GEO NET Umweltconsulting erfasst und modelliert (Senatsverwaltung für Stadtentwicklung 2011). Die Klimaexperten verwendeten dazu ein Raster von 25x25 Metern. Oberflächenbeschaffenheit und Gebäudehöhe als wichtige Parameter gingen in die Modellierung ein. Die Daten betrachten Berlin in den Zeiträumen 2001–2010, 2045–2055 und 2091–2099. Mehrere Ausprägungen wurden für die Zeiträume berechnet, unter anderem Sommertage (maximale Temperatur über 25°C), heiße

Tage (maximale Temperatur über 30°C) und Tropennächte (Minimaltemperatur über 20°C).

Die Daten sind nicht zum direkten Herunterladen erhältlich, sondern wurden den Volontären von der Senatsverwaltung zur Verfügung gestellt. Maßgabe war hierbei, die Temperatur-

werte nicht im 25-Meter-Raster zu veröffentlichen, sondern auf die granularste Ebene der Lebensweltlich Orientierten Räume (LOR) umzurechnen (447 Planungsräume). Grund dafür waren nach Verständnis der Volontäre und des Autors mögliche Rückschlüsse auf die künftige Hitzebelastung einzelner Immobilien,

Abbildung 6: Durchschnittliche Zahl der Hitzetage

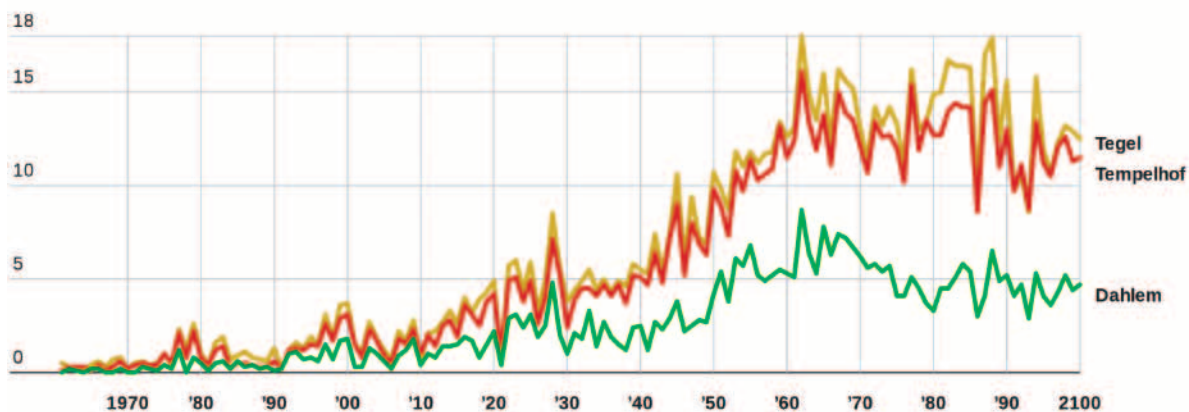


Die Karte zeigt die durchschnittliche Anzahl an Hitzetagen pro Jahr für die Zeiträume 2001 bis 2010 sowie 2090 bis 2099. Je dunkler die Kieze eingefärbt sind, desto höher ist die Zahl der Tage.

Abbildung 7: Anzahl der Nächte, an denen die Temperatur von 20°C nicht unterschritten wird

Tropennächte

Anzahl der Nächte, an denen 20°C nicht unterschritten werden



Anzahl der Nächte, an denen 20°C nicht unterschritten werden (= Tropennächte). Die Projektion bis ins Jahr 2100 beruht auf Messdaten der drei Wetterstationen in Tegel, Tempelhof und Dahlem.

was deren Marktwert negativ beeinflussen könnte. Für die Daten wurde mit QGIS der Durchschnittswert für jeden LOR errechnet. Verwendet wurden am Ende nur die Werte für heiße Tage und Tropennächte.

Analyse und Visuelle Aufbereitung – „So warm“

Zentrales Element der Darstellung und Analyse war eine interaktive Choroplethenkarte, die es für jeden der 447 Planungsräume ermöglicht, die Zahl der Hitzetage und Tropennächte im Jahr 2100 zu sehen und mit den Werten für 2001–2010 zu vergleichen. Die Planungsräume wurden im journalistischen Kontext als „Kiez“ bezeichnet, da dies näher am Verständnis der alltäglichen Nutzer ist. Essenzielle Funktion ist die straßengenaue Suche des Kiezes, da ein gewöhnlicher Nutzer keinen Einblick hat, in welchem Planungsraum er wohnt (siehe Abb. 6). Bei der Darstellung mithilfe der Choroplethenkarte zeigt sich, dass im Zeitraum 2090–2099 die Kieze in der Regel doppelt so viele heiße Tage erleben wird wie am Anfang des Jahrhunderts. Die absolute Zahl ist in den Innenstadt-Kiezen deutlich höher als in den Randlagen, am wenigsten betroffen sind durch den ausgleichenden Effekt waldreiche Gebiete. Im Kiez Wilhelmstraße in Mitte etwa steigt die Zahl der Hitzetage von 11 auf 25, während im Grunewald die Zahl von drei auf neun steigt. Deutlicher ausgeprägt sind die Unterschiede bei den Tropennächten, hier zeigen sich die eigentlichen Aufheizungseffekte bei Gebäuden und versiegelten Flächen. Im Beusselkiez (Mitte) etwa steigt die Zahl der Nächte mit einer Mindesttemperatur von 20 Grad von vier auf 15, während die Zahl in vielen grüneren Gebieten von null auf vier oder fünf steigt. Bei zahlreichen innerstädtischen Kiezen tendiert Zahl der Tropennächte 2090–2099 gegen zehn bis 15.

Außerdem verwendeten die Volontäre eine weitere Simulation, die für drei Berliner Klimastationen (Tegel, Tempelhof, Dahlem) existiert. Sie basiert auf einem weiteren Klimamodell. Die Entwicklung von Hitzetagen zwischen 1960 und 2100 zeigt auf Liniendiagrammen einen grundsätzlichen Anstieg spätestens ab 2050, der für alle drei Stationen annähernd gleich ist. Bei den Tropennächten hingegen steigt die Zahl für alle drei Stationen, jedoch entwickeln sich Tegel und Tempelhof nahezu identisch, während die Zahlen für Dahlem weit unter denen der anderen beiden Stationen bleiben. Dies bestätigt letztlich die grundsätzliche Tendenz des anderen, granularen Modells (siehe Abbildung 7).

Für die datenjournalistische Aufbereitung solch komplexer Daten ist die Einordnung durch Experten von großer Wichtigkeit. Die Volontäre führten ein Videointerview mit dem Physiker Matthias Lüdeke vom Potsdam-Institut für Klimafolgenforschung, der erklärte, wie das Klima der Zukunft berechnet wird. Ebenso sprachen sie mit dem Klimatologen Wilfried Endlicher vom Geografischen Institut der Berliner Humboldt Universität, der erläuterte, wie sich Städte dem Klimawandel anpassen können. Grundsätzlich möglich wäre auch eine Betrachtung der Sommertage statt der Hitzetage möglich gewesen. Dies hätte unter Umständen durch die stärker divergierende Zahl pro Kiez eine ausdifferenziertere Analyse ermöglicht.

Daten – „So viele“

In einem deutlichen kleineren Zeitraum bewegte sich die Bevölkerungsprognose für Berlin. Sie ist für die Jahre 2011 bis

2030 angelegt und beruht auf dem Einwohnermelderegister. Die Daten liegen auf Ebene der 60 Prognoseräume (größte Granularität der Lebensweltlich Orientierten Räume) vor. Für jeden der Prognoseräume ist die Anzahl der Bewohner aufgeschlüsselt nach einzelnen Kalenderjahren und Altersjahren, etwa Anzahl der 67-Jährigen in Friedrichshain-Ost im Jahr 2021.

Die verwendete Bevölkerungsprognose (von 2012) war bei Umsetzung des Projekts die aktuellste, jedoch hat die Senatsverwaltung für Stadtentwicklung eine Korrektur der Zahlen nach oben angekündigt, da die Zuwanderung nach Berlin zwischen 2012 und 2015 höher war als erwartet.

Analyse und Visuelle Aufbereitung – „So viele“

Die Volontäre arbeiteten in der Analysephase die für sie relevanten Kerntendenzen der Bevölkerungsprognose heraus, nämlich die Veränderungen bei Älteren (65+) und Kleinkindern (0–6 Jahre). Dazu verglichen sie die absoluten Werte für 2011 und 2030 und berechneten die prozentualen Unterschiede. Während es in nahezu allen Berliner Prognoseräumen im Jahr 2030 deutlich mehr Ältere geben wird, sinkt in den meisten Räumen die Zahl der Kleinkinder. Nur in wenigen Gebieten werden der Prognose zufolge mehr Kinder leben. Das Team verwendete für eine erste räumliche Analyse das kostenlose Kartenwerkzeug CartoDB.

Auch bei diesem Artikel unterstützten die Entwickler des Morgenpost-Interaktiv-Teams die Volontäre, so dass diese auf eine leistungsfähige interaktive Karte zurückgreifen konnten. Diese Choroplethenkarte ermöglicht eine straßengenaue Suche und einen Vergleich zwischen der Gesamtentwicklung, der Veränderung bei Senioren und bei Kleinkindern (siehe Abbildung 8).

Besonders auffällig sind die Unterschiede im Bezirk Lichtenberg, wo im Lichtenberg-Süd – als nur einem von sieben Räumen in ganz Berlin – die Zahl der Kinder steigen wird, während in Hohenschönhausen-Nord (ebenfalls Lichtenberg) die Kinderzahl um mehr als 15 Prozent sinken wird. Zugleich nimmt die Zahl der Senioren in jenem Gebiet um über 90 Prozent zu. Von diesen Erkenntnissen ausgehend fokussierten die Volontäre auf zwei Aspekte: Sie untersuchten zum einen die gegensätzlichen Teile Lichtenbergs und anderer Bezirke und fragten Bewohner, wie sie zur künftigen Entwicklung stehen. Zum andern fiel ihnen die steigende Zahl der über 99-Jährigen in Berlin auf – der Prognose zufolge sollen sich ihre Zahl im Jahr 2030 auf fast 4.000 Personen fast verdoppeln (2011: 1.737). Die Volontäre führten daher Videointerviews mit drei 100-jährigen Berlinerinnen und schlüsselten auf, in welchen Gebieten 100-Jährige vor allem leben werden.

Einordnung Teilprojekte

Während das Teilprojekt „So warm“ vor allem von einer sehr komplexen Materie und eher zu vielen Daten geprägt war, standen den Volontären bei „So viele“ wenig granulare Daten zur Verfügung, die an der Grenze der Verwendbarkeit waren. Die Volontäre beider Teilprojekte haben jedoch die Herausforderungen aus datenjournalistischer Sicht gut bewältigt und angemessene Umsetzungen gefunden. Wo die Daten nicht „individuell“ genug waren, fanden sie interessante Interviewpartner, während das Klima-Teilprojekt durch das recht

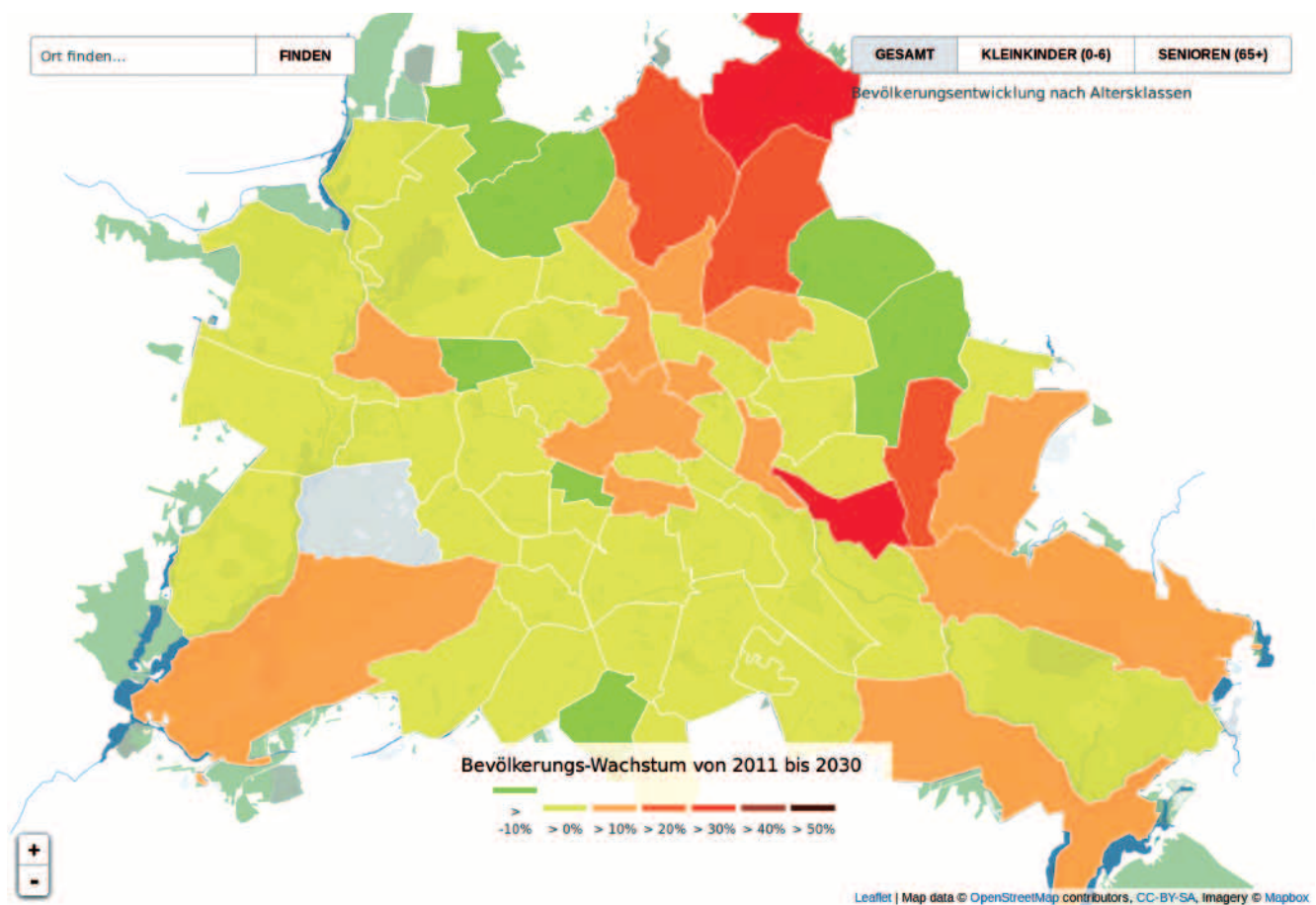
granulare Datenniveau den Nutzern einen guten Bezug zur Alltagsumgebung ermöglichte. Zudem demonstrierten die Volontäre des Klima-Teilprojekts, dass engagierte Gespräche mit den Entscheidern in der Senatsverwaltung den Zugang zu Daten ermöglichen, die nicht öffentlich zugänglich sind.

Ausblick

Die beiden Projekte „Airbnb vs. Berlin“ und „So wird Berlin“ zeigen, dass auch komplexere Daten einem Publikum jenseits der Fachöffentlichkeit vermittelt werden können. Wichtig sind dabei hinreichende Erklärungen und Einordnungen und gute

visuelle Aufbereitungen. Zugute kommen Journalisten und Designern die steigende Leistungsfähigkeit von Webbrowsern und Daten und Visualisierungen als Trendthemen. Momentan liegt eine große Herausforderung in der Datenakquise, die alles andere als standardisiert ist. Doch auch nicht-öffentliche Unternehmen wie Airbnb unterliegen einem gewissen Transparenzdruck, der dafür sorgen könnte, dass Interessierte einen Zugang zu Daten erhalten. Eine große Chance könnten nach Auffassung des Autors projektbezogene Kooperationen zwischen Statistik-Spezialisten und Datenjournalisten sein. Beide Seiten profitieren hier von der jeweiligen Spezialisierung, das Ergebnis wäre eine größere Öffentlichkeit.

Abbildung 8: Vorausberechnete Bevölkerungsentwicklung



Die Karte zeigt, wie stark die jeweiligen Stadtteile (LOR-Prognoseräume) im Jahr 2030 im Vergleich zu 2011 schrumpfen oder wachsen werden. Je dunkler die Stadtteile eingefärbt sind, desto stärker das Wachstum. Grün bedeutet ein Schrumpfen.

Literatur

Kayser-Bril, Nicolas (2015): Datajournalism. Blogpost auf: <http://blog.nkb.fr/datajournalism/>
Isaac, Mike (2015): Airbnb Releases Trove of New York City Home-Sharing Data. In: New York Times, 1. Dezember 2015. Online: <http://www.nytimes.com/2015/12/02/technology/airbnb-releases-trove-of-new-york-city-home-sharing-data.html>

Senatsverwaltung für Stadtentwicklung (2011): Untersuchungen zum Klimawandel in Berlin – Dokumentation der im Rahmen des Stadtentwicklungsplan (StEP) Klima durchgeführten Modellrechnungen. Online: http://www.stadtentwicklung.berlin.de/download/StEP_Klima/SenStadt_StEP_Klima_Fachbeitrag_Klima-modellierung.pdf

Royal, Cindy, Blasingame, Dale (2015): Data journalism: An explication. In: Multimedia Edition of #ISOJ, the official research journal of ISOJ International Symposium on Online Journalism), Volume 5, Issue 1 Spring 2015. Online: <https://isojournal.wordpress.com/2015/04/15/data-journalism-an-explication/>