

DDJ in der Praxis bei Spiegel Online

@PatrickStotz

@datenlese

Das Team



DER SPIEGEL

Verschiedene Projekt-Typen

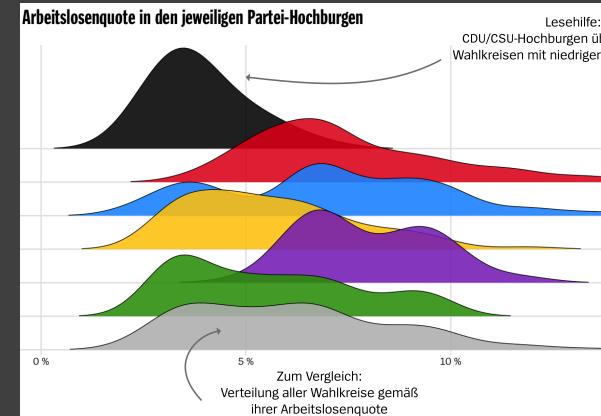
Leuchttürme



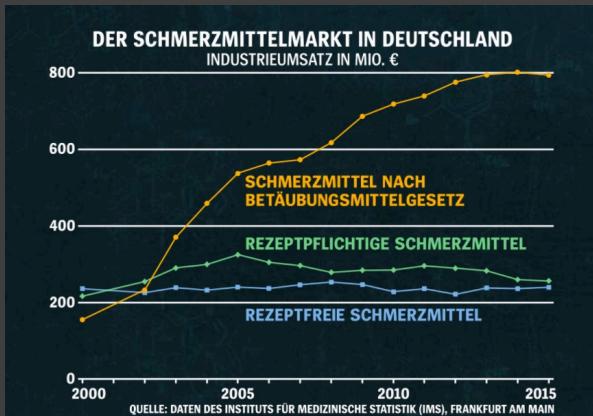
Artikelserie (aus Datenrecherche)



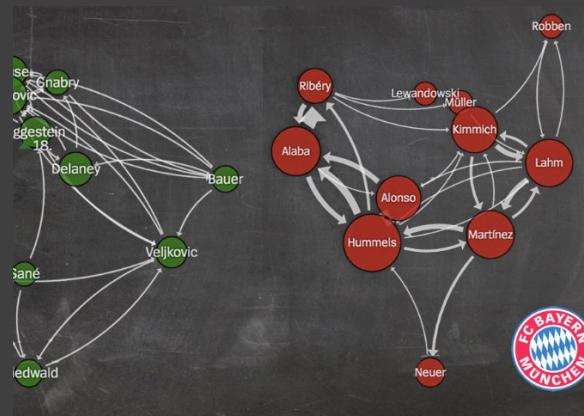
Tages-/wochenaktuelle Geschichten



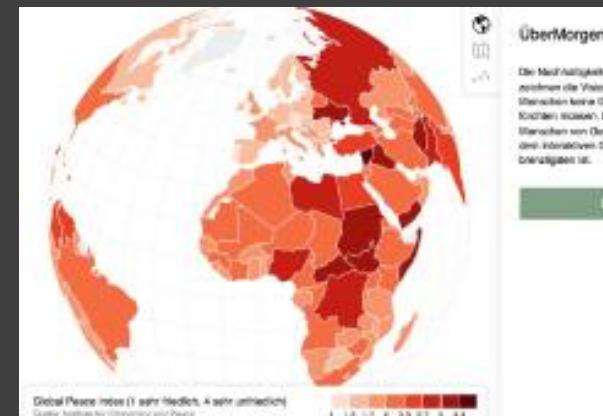
Recherche-Unterstützung



Infrastruktur - DDJ für den Alltag



Experimente - Expedition Übermorgen



Beispiele und Workflows

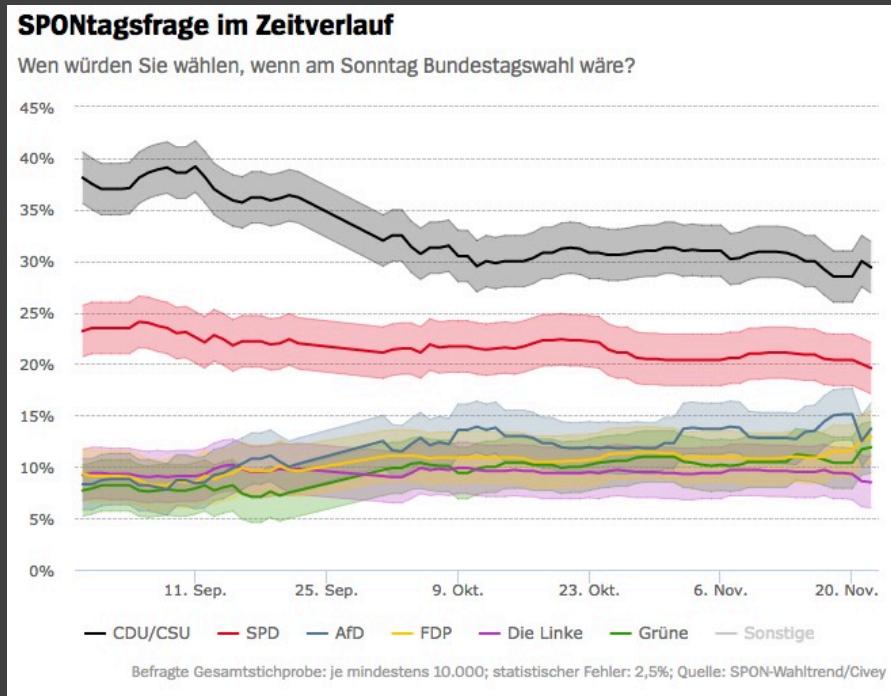
(nach steigender Komplexität sortiert)

Kleine Recherchen, Analysen & Einschätzungen



- Daten recherchieren
- Quellen kombinieren
- Zentrale Punkte identifizieren
- Visualisierungsformen wählen
- Abstimmung mit Redaktion und Grafik
- (Rohform der) Grafik erstellen

Kleine Recherchen, Analysen & Einschätzungen



- „Erziehungsarbeit“ in Sachen Unsicherheit...
- ...bei der Erstellung der Visualisierung
- ...als Sparringspartner für Aufmachung & Formulierung

Was ist das Besondere an der Civey-Methode?

Das Meinungsforschungsinstitut [Civey](#) arbeitet mit einem mehrstufigen Verfahren. Zuerst werden alle Umfragen in einem Netzwerk aus mehr als 12.000 Websites ausgespielt ("Riversampling"). Online kann jeder an den Befragungen teilnehmen und wird mit seinen Antworten im repräsentativen Ergebnis berücksichtigt, sofern er sich registriert hat. Aus diesen Nutzern zieht Civey eine quotierte Stichprobe, unter anderem nach den Merkmalen Alter, Geschlecht und Bevölkerungsdichte. In einem dritten Schritt werden die Ergebnisse nach weiteren Faktoren und Wertehaltungen gewichtet, um Verzerrungen zu korrigieren und Manipulationen zu verhindern.

- ⊕ Warum ist eine Registrierung nötig?
- ⊕ Wie werden die Ergebnisse repräsentativ?
- ⊕ Erreicht man online überhaupt genügend Teilnehmer?
- ⊕ Woran erkenne ich die Güte eines Ergebnisses?
- ⊕ Was bedeutet es, wenn sich die farbigen Bereiche in den Grafiken überschneiden?
- ⊕ Was passiert mit meinen Daten?

- Transparenz fördern
- Methodische Details im Artikel durch Erklärkasten im FAQ-Format

Scraping

Was passiert auf dem Fußball-Transfermarkt? Viele schreiben vom „Transfer-Irrsinn“ - ist das wirklich alles so irrational?

Wir haben untersucht?

- Die Entwicklung der Transfersummen über Zeit
- Vergleich der Ligen
- Durchschnittstransfer vs. Superstars

Antwort:

- Umsatzentwicklung im Fußball beachten!
- Der Fußball hat seine eigene Inflation...und die kann man berechnen

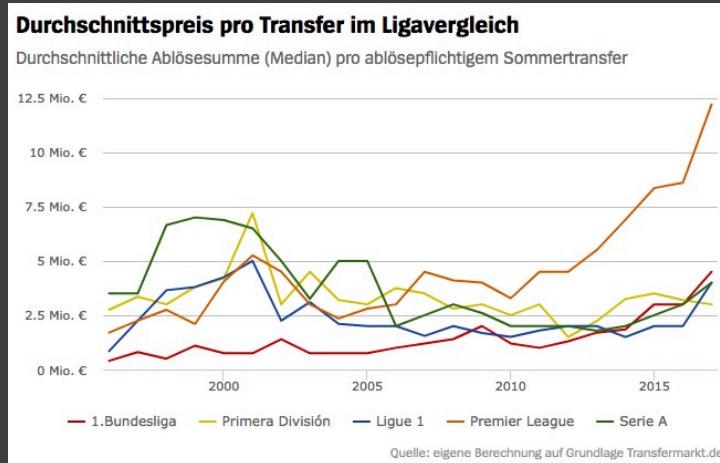
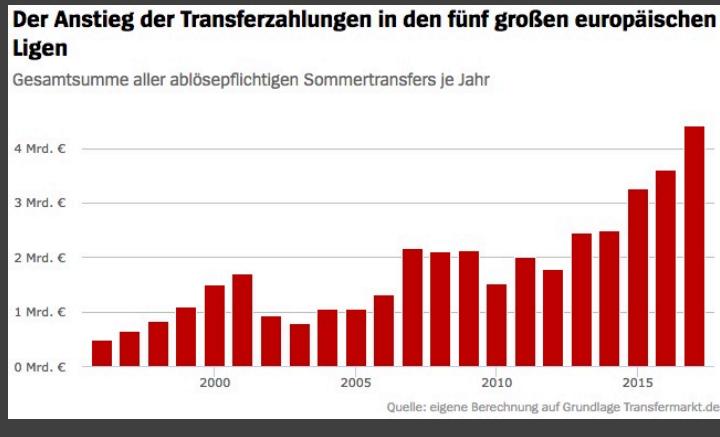


Scraping

- Quelle: Transfermarkt.de
- Daten ändern sich täglich
- Technik: R + rvest + selector gadget + tidyverse
- Analyse + Artikel fertig vorbereitet, nach Schließung des Transfermarkts in ca. 15 Minuten auf den aktuellen Stand gebracht

The screenshot shows the transfermarkt.de website interface. At the top, there's a navigation bar with links like NEWS, TV, TRANSFERS & GERÜCHTE, MARKTWERTE, WETTBEWERBE (which is highlighted in blue), FOREN, MEIN TM, LIVE, and ANMELDEN. Below the navigation is a breadcrumb trail: Deutschland > 1.Bundesliga > Verein > Spieler. The main content area is titled "1.BUNDESLIGA" and displays various statistics: Ligahöhe: 1.Liga - Deutschland, Ligagröße: 18 Mannschaften, Spieler: 520, Legionäre: 274 Spieler (52%), and more. To the right, it shows a "Gesamtmarktwert: 2,89 Mrd. €". Below this is a "TRANSFERS 17/18" section with filters for Saison (17/18), Transferzeitpunkt (egal), and Transfers (Leihen, Vereinsinterne Wechsel). It lists transfers for FC Bayern München, including Corentin Tolisso, Kingsley Coman, Niklas Süle, Serge Gnabry, Sebastian Rudy, and Tom Starke. The "Ausgaben: 90,50 Mio. €" column is highlighted with a red border. To the right, there's a "TRANSFERBILANZ 17/18" section with statistics for Abgänge (125), Zugänge (111), and Gesamtbilanz (Gesamtbilanz: -101.435.000 €, Bilanz pro Verein: -5.635.278 €, Bilanz pro Spieler: -1.431.400 €). At the bottom right, there's a "Top-Gerüchte" section featuring a player profile for Sandro Wagner.

Aufbereitung mit hauseigenen / adaptierten Tools



- Unsere Standard-Lösung für (interaktive) Diagramme: Highcharts
- + einfache Diagramme im SPON-Look in 5 Minuten
 - + recht große Gestaltungsfreiheit
 - + Responsivität out-of-the-box
 - Interface für „normale Redakteure“ kaum zu bedienen
 - Laufen wir alle in Legacy-Probleme rein? Was passiert mit unseren Inhalten in 2/3/5 Jahren?

Aufbereitung mit hauseigenen / adaptierten Tools

Sortier- und filterbare Tabellen als sehr einfache Form der interaktiven Darstellung.

Leser suchen immer:

- ihren Verein
- ihren Lieblingsspieler
- ihre Gemeinde
- ihren Arzt

Die 500 teuersten Bundesliga-Transfers unter Berücksichtigung der Fußball-Inflation*							
Einträge pro Seite: ▾ 20		Suche					
Name	Inflationsbereinigte Ablöse in Mio. €	Ablösesumme in Mio. €	Jahr	Von	Nach	Rang	
Ousmane Dembélé	105,00	105,00	2017	Borussia Dortmund	FC Barcelona	1	
Roy Makaay	92,75	19,75	2003	Deportivo La Coruna	FC Bayern München	2	
Kevin De Bruyne	85,32	74,00	2015	VfL Wolfsburg	Manchester City	3	
Edin Dzeko	80,29	37,00	2010	VfL Wolfsburg	Manchester City	4	
Javi Martínez	79,33	40,00	2012	Athletic Bilbao	FC Bayern München	5	
Evanilson	70,44	15,00	2003	AC Parma	Borussia Dortmund	6	
Mario Gómez	58,35	30,00	2009	VfB Stuttgart	FC Bayern München	7	
Mario Götze	57,91	37,00	2013	Borussia Dortmund	FC Bayern München	8	
Márcio Amoroso	57,58	25,50	2001	AC Parma	Borussia Dortmund	9	
Manuel Neuer	56,28	30,00	2011	FC Schalke 04	FC Bayern München	10	
Jörg Heinrich	52,56	12,60	1998	Borussia Dortmund	AC Florenz	11	
Diego	52,51	27,00	2009	SV Werder Bremen	Juventus Turin	12	
Aleksandr Hleb	52,39	15,00	2005	VfB Stuttgart	FC Arsenal	13	
Leroy Sané	49,96	50,00	2016	FC Schalke 04	Manchester City	14	
Julian Draxler	49,58	43,00	2015	FC Schalke 04	VfL Wolfsburg	15	
André Schürrle	48,61	32,00	2014	FC Chelsea	VfL Wolfsburg	16	
Roberto Firmino	47,27	41,00	2015	TSG 1899 Hoffenheim	FC Liverpool	17	
Arjen Robben	46,68	24,00	2009	Real Madrid	FC Bayern München	18	
Toni Kroos	45,57	30,00	2014	FC Bayern München	Real Madrid	19	
Granit Xhaka	44,96	45,00	2016	Borussia Mönchengladbach	FC Arsenal	20	
Einträge 1 bis 20 von 500							
◀		1	2	3	4	5	… 25 ➤

Karten



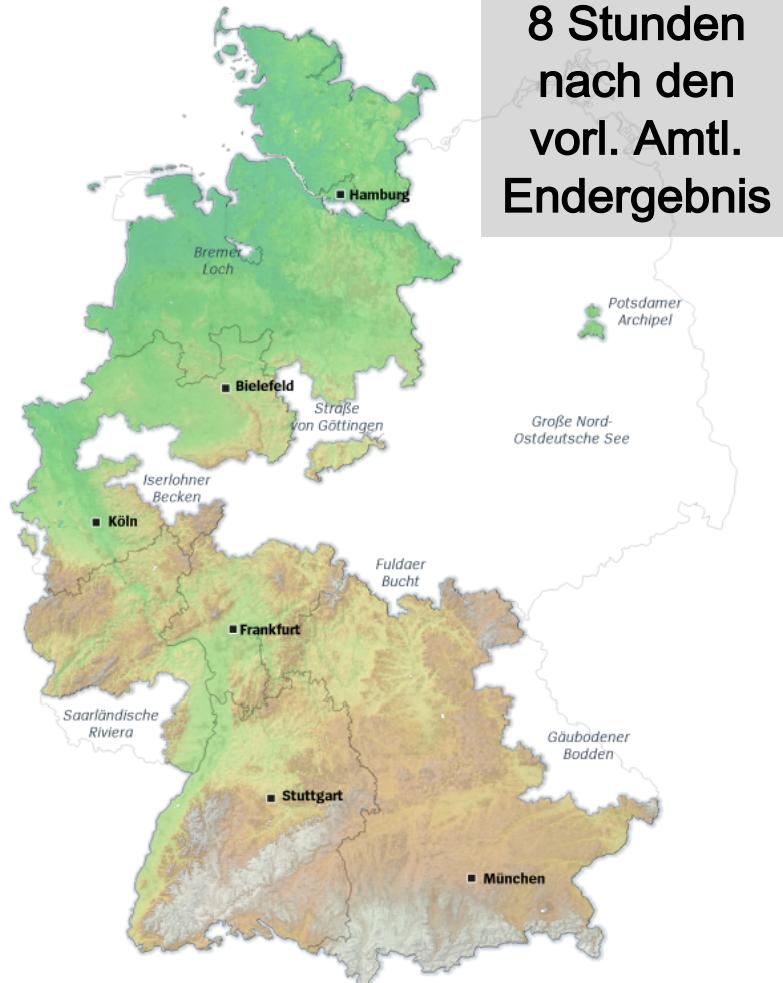
Das wollen wir für die #BTW17 auch

Dynamisch verknüpft mit den Wahldaten + schnell bitte

Die Jamaika-Republik

Wo Union, FDP und Grüne eine Mehrheit der Zweitstimmen haben

8 Stunden nach den vorl. Amtl. Endergebnis



56,2 Mio
Einwohner



5%
Arbeitslose



67%
Christe



35%
Hochschulreife



22.124 €
Netto-Einkommen
pro Kopf

Karten

QGIS 2.18.2 - Merkel_Schulz_Kartenstil

QGIS mächtiges Tool
Karten sind eine eigene Welt

Layers Panel

- Topmap
 - Staatsgrenzen_D_2500k
 - Google Labels
 - Koalitionen
 - KOAL_GROKO
 - KOAL_SCHWEGELB
 - KOAL_ROTGRUEN
 - KOAL_ROTROTGRUEN
 - KOAL_AMPEL
 - KOAL_JAMAICA
 - Border
 - Water
 - KOAL_JAMAICA_2 OGRGeoJS...
 - VG250_LAN
 - KOAL_UNION_ABS
 - Basemap
 - geonames
 - Bing Satellite
 - Topography
 - Topography Kopie
 - Hillshade

Browser Panel

- Project home
- Home
- Favourites
 - /Users/stotz/Desktop
 - Bildschirmfoto 2017-09-29 um 13.45.58.j
 - bttw17.est.west.xlsx
 - Ohne Titel.shp
 - sachsen.json
 - ToDos.txt
 - /Users/stotz/Desktop/QGIS_workshop
 - /Users/stotz/Documents
 - /Volumes/bereiche/SPONDatenlese
 - 00 - Geodaten
 - admin_Grenzen_D_2010_Geodatenz
 - dokumentation
 - vg250_ebenen-historisch
 - de1012
 - vg250_bld.shp
 - vg250_gem.shp
 - vg250_krs.shp
 - vg250_lan.shp
 - vg250_uta.shp
 - vg250_vwg.shp
 - quellenvermerk.txt
 - admin_Grenzen_D_2014_Geodatenz
 - admin_Grenzen_D_2015_Geodatenz
 - admin_Grenzen_D_2016_Geodatenz
 - admin_Grenzen_D_2017_Geodatenz
 - admin_Grenzen_EU_2015_Nuts_1-3
 - admin_Grenzen_Welt_2015_Natural
 - admin_Grenzen_Welt_2015_Natural
 - admin_Grenzen_Welt_2015_Natural

Karten

QGIS Composer Edit View Layout Atlas Settings Window Help KOAL_JAMAICA

Die Jamaika-Republik
Wo Union, FDP und Grüne eine Mehrheit der Zweitstimmen haben

56,2 Mio Einwohner 5% Arbeitslose 67% Christen 35% Hochschulreife 22.124 € Netto-Einkommen pro Kopf

SPIEGEL ONLINE

Items Command history

Command history

Composition Item properties Atlas generation

Label

Main properties

22.124 €

Render as HTML Insert an expression...

Appearance

Font color: black

Horizontal margin: 1,00 mm

Vertical margin: 1,00 mm

Horizontal alignment: Center

Vertical alignment: Top

Position and size

Rotation

Frame

Background

Item ID

Rendering

Variables

Variable	Value
Global	
qgis_os_...	osx
qgis_rele...	2.10.0
qgis_rele...	Las Palmas
qgis_vers...	2.18.2
qgis_vers...	21802
user_acco...	stotzP
user_full...	StotzP

x: 255.207 mm y: 231.339 mm page: 1 134.5%

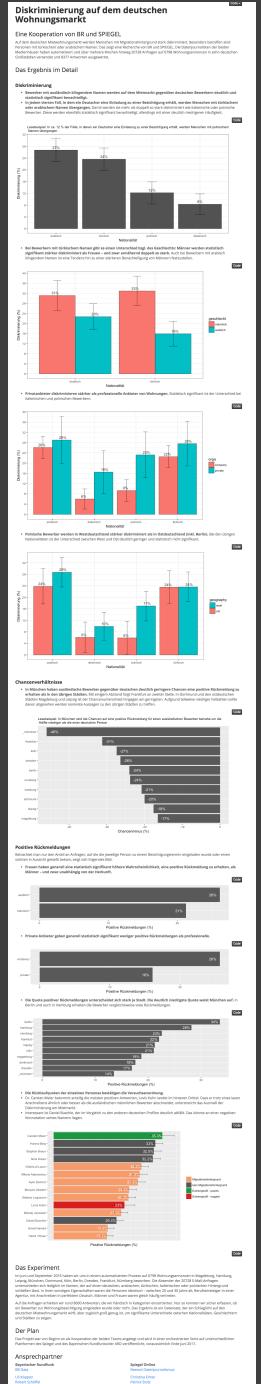
Analysen mit Notebooks

Was sind Notebooks?

- Eine Art Forschungs-/Recherche-Tagebuch, gepflegt von Datenjournalisten, mit Rohdaten verknüpft und doch für jeden lesbar.
- Gibt es in der R- und Python-Welt (R Notebook, jupyter ipython Notebooks)

Wo liegen die Vorteile?

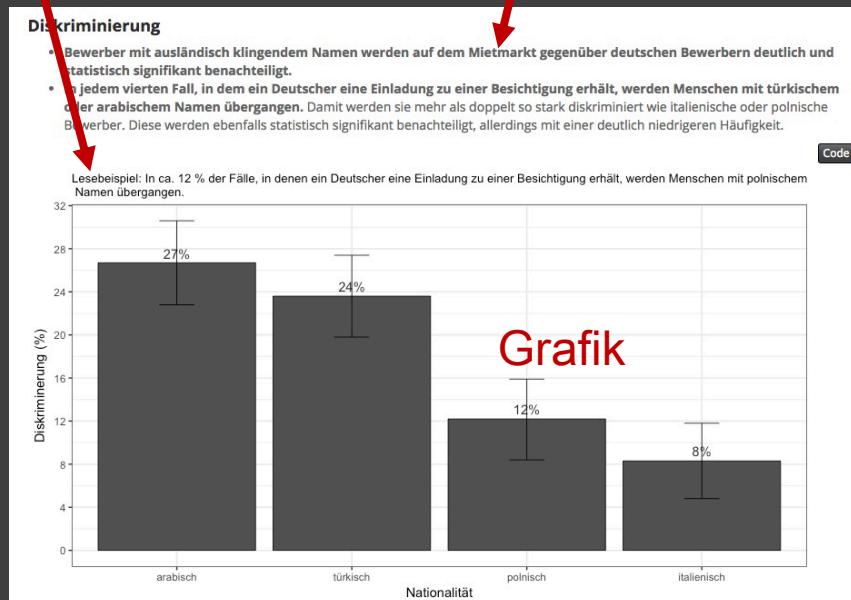
- Kommunikation mit (anderen) Redaktionen
- Alles in einem Dokument. Immer aktuell. In jedem Browser lesbar.
- Hervorragend geeignet um Methodik transparent und reproduzierbar zu veröffentlichen



Analysen mit Notebooks

Formulierungs-vorschlag

Lesebeispiel /
Erläuterung



Code-Schnipsel

Das Experiment

Im Juni und September 2016 haben wir uns in einem automatisierten Prozess auf 6798 Wohnungsannoncen in Magdeburg, Hamburg, Leipzig, München, Dortmund, Köln, Berlin, Dresden, Frankfurt, Nürnberg beworben. Die Absender der 20728 E-Mail-Anfragen unterschieden sich lediglich im Namen, der auf einen deutschen, arabischen, türkischen, italienischen oder polnischen Hintergrund schließen lässt. In ihren sonstigen Eigenschaften waren die Personen identisch – zumindest farblich gesehen – als Wohnungssucher in einer Agentur, mit Anschreiben in perfektem Deutsch. Männer und Frauen waren gleich häufig vertreten.

Auf die Anfragen erhielten wir rund 8000 Antworten, die wir händlich in Kategorien eingesortiert. Nur so konnten wir sicher erfassen, ob ein Bewerber zur Wohnungsbesichtigung eingeladen wurde oder nicht. Das Ergebnis ist ein Datensatz, der ein Schlaglicht auf den deutschen Mietwohnungsmarkt wirft, aber zugleich groß genug ist, um signifikante Unterschiede zwischen Nationalitäten, Geschlechtern und Städten zu zeigen.

Der Plan

Das Projekt war von Beginn an als Kooperation der beiden Teams angelegt und wird in einer orchestrierten Serie auf unterschiedlichen Plattformen des Spiegel und des Bayerischen Rundfunks/der ARD veröffentlicht, voraussichtlich Ende Juni 2017.

Ansprechpartner

Bayerischer Rundfunk

BR Data

Uli Köppen
Robert Schöffel
Oliver Schnuck

Spiegel Online
Ressort Datenjournalismus

Christina Elmer
Patrick Stotz
Achim Tack

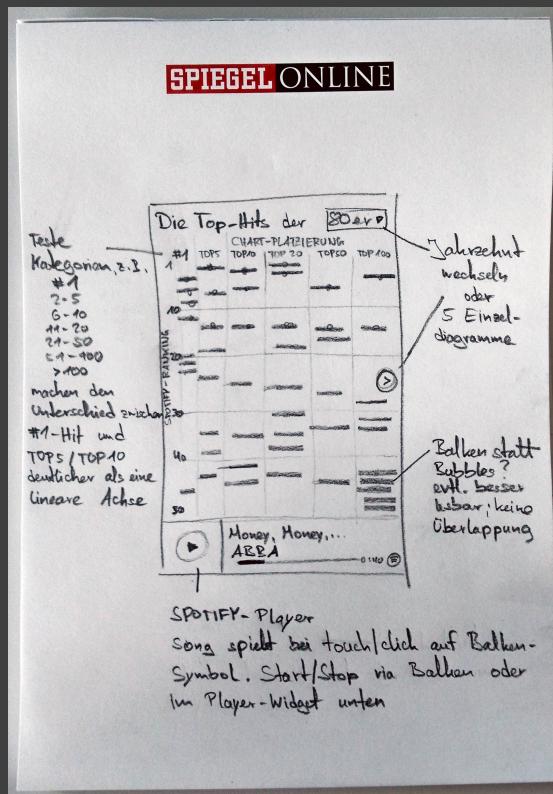
Kontakt

Hintergrund

Eigene Visualisierungen mit D3.js

Frage: Was sind eigentlich die beliebtesten Hits (in Deutschland) der vergangenen Jahrzehnte?

Skizze: mobile first, interaktiv, da muss man Musik spielen können



Umsetzung: läuft, wird geklickt, mit externen Widgets/Anbietern (hier Spotify) arbeiten 😊