



Dokumentation

85.386.227

births over 21 years

when did **that** happen?

von Lars Obist

Einleitung

Grundidee

Bei dem Brainstorming für das Thema zu der Datenvisualisierung wollte ich in jedem Fall etwas visualisieren, welches etwas schönes sowie fröhliches visualisiert.

Somit sollte das ausgewählte Thema keine reinen Zahlen, wie bei einer möglichen Datenvisualisierung zur Entwicklung der Wirtschaft eines Landes, oder Problematiken in der Gesellschaft darstellen, da es in den Bereichen genügend Datenvisualisierungen gibt und man somit nichts wirklich neues visualisieren würde.

Schlussendlich bin ich auf die Idee gekommen, die Verteilung von Geburten über das Jahr zu analysieren und visualisieren.

Thema

Jeder hat sich vermutlich bereits gefragt, wie viele Menschen wohl an dem selben Tag mit einem Geburtstag haben. Viele werden sich vermutlich auch bereits gefragt haben, wieso sich viele Geburtstage in seinem Umfeld in einem bestimmten Zeitraum ballen.

Unter dem Arbeitstitel „Circle of life“ habe ich mehrere Datenvisualisierungen entwickelt, um die Verteilung von Geburten sowie andere daraus entstehende Auffälligkeiten hervorzuheben und zu visualisieren.

Speziell stelle ich die Frage „when did that happen?“ welche auf die 85 Millionen Geburten in einem Zeitraum von 21 Jahre anspielen. Ich habe die Webseite auf Englisch verfasst, damit Nutzer aus der gesamten Welt diese Visualisierung verstehen und ansehen können.

Mithilfe der entwickelten Datenvisualisierungen möchte ich Nutzern die Möglichkeit geben, nachzusehen, ob die persönlichen Annahmen und das eigene Empfinden auch repräsentativ ist oder es lediglich im eigenen Umfeld so ist.

Datensatz

Auf der Suche nach einem passenden Datensatz habe ich einen Artikel von FiveThirtyEight mit dem Titel „Some People Are Too Superstitious To Have A Baby On Friday The 13th“ gefunden.

FiveThirtyEight ist eine US-amerikanische Nachrichtenseite, welche besonders viel mit Daten und Statistiken arbeitet. Sie gehört dem US-Fernsehsender ABC News an.

In dem Bericht stellen sie die Daten bereit, welche sie für den Bericht benutzt haben. Die Erhebungen zeigen über einen großen Zeitraum die von den US Behörden aufgezeichneten Geburten.

Diese Datenerhebungen wurden von der „Social Security Administration“ sowie dem „Centers of Disease Control and Prevention’s National Center for Health Statistics“ bereitgestellt und in separaten Datensätzen gespeichert.

Wie repräsentativ sind die Daten?

Die Geburten wurden über einen Zeitraum von 21 Jahren in Datensätzen gespeichert und spiegeln lediglich die Zahlen aus den Vereinigten Staaten Amerikas dar.

Zwischen 1994 und 2014 wurden insgesamt 7670 Einträge gemacht und umfasst in der Summe 85.386.227 aufgezeichnete Geburten.

Anhand dieser Daten kann ich für meine Datenvisualisierung eine repräsentative Erhebung machen.

Einheiten

Beide Datensätze haben die identische Struktur. In den Datensätzen finden sich folgende Einheiten:

Jahr, Monat, Tag des Monats, Wochentag, Anzahl der Geburten an dem Tag

Daraus ergebende Visualisierungen

Durch die Struktur der Datensätze ergab sich die Möglichkeit, zusätzlich zu der angedachten Übersicht weitere Visualisierungen anzufertigen.

Dazu mussten die Datensätze jedoch zunächst einmal vereint und bereinigt werden, da beide in dem Zeitraum 2000 bis 2003 Daten über die Anzahl von Geburten haben.

Durch die Ausarbeitung und Bereinigung der Datensätze ergaben sich folgende Fragestellungen und Möglichkeiten, die Anzahl der Geburten zu analysieren und zu visualisieren.

Wie verteilen sich 85.386.227 Geburten durchschnittlich über ein Kalenderjahr?

Haben alle Wochentage durchschnittlich eine ähnliche Anzahl von Geburten?

Sind die Geburten innerhalb eines Jahres über die Monate konstant?

Bei der Ausarbeitungen der obigen Fragestellungen wurden beabsichtigt Durchschnittswerte berechnet, da sonst die Daten innerhalb der Tagesübersicht, der Wochentagsübersicht und der Monatsübersicht durch den 29. Februar, welcher alle 4 Jahre durch das Schaltjahr in den Datensätzen auftritt, verfälscht dargestellt werden würden.

Eine weitere Fragestellung die sich durch den Datensatz ergeben hat ist:

Wie entwickelt sich die Zahlen über die Jahre?

Lediglich bei dieser Ausarbeitung wurden Summen verwendet, wenn auch bei den Schaltjahren ein Tag mehr in die Berechnung eingeflossen ist.

Bei allen Ausarbeitungen wurde auf Besonderheiten geachtet, die durch die Visualisierungen Hervorgehoben werden.

Moodboards

Durch die Ausarbeitung der Datensätze wurde klar, mit welchen Datenpunkten gearbeitet werden kann.

Visualisierungen

Um mir einen Eindruck zu machen, welche Optionen mir zum visualisieren der Daten zur Verfügung stehen, habe ich ein Moodboard für mögliche Ansätze erstellt.

So hab ich mir für die Verteilung über das Jahr inspiration von folgenden Visualisierungen genommen:

Webseite

Für die Webseite habe ich bezüglich des Designs, der Farbe und der Schrift ebenfalls drei Moodboards entwickelt, in der eine Sammlung von Inspirationen gesammelt wurden. Im weiteren Designprozess habe ich dieses Moodboard als Referenz verwendet und mir daraus gewisse Elemente ausgesucht, welche die Atmosphäre aufbaut welche zur Thematik von Geburtenzahlen passt.

Visuelle Ideen

Bei genauerer Analyse des Moodboards wird klar, dass die Inhalte auf der Webseite sehr modern und farbfroh werden sollen. Jedoch soll dies nicht zu überladen wirken, da es seriös und mit Leichtigkeit die Daten darstellen soll.

Die Nutzung soll möglichst eindeutig und simpel sein um zu Gewährleisten, dass der Fokus des Nutzers auf den Visualisierungen ist und nicht bei der Steuerung der Seite.

Alle Elemente auf der Webseite und auch innerhalb den Visualisierungen sollen abgerundet sein. Dies schafft eine freundliche und vertrauenswürdige Umgebung, was ideal zur Thematik passt, da scharfe Ecken im Gegensatz mit Vorsicht begegnet werden.

Visualisierungen

Alle Visualisierungen sollen sehr minimalistisch gestaltet werden.

Um die 366 Datenpunkte, die sich bei der Verteilung von den Geburten über das Jahr ergeben, möglichst kompakt und simpel darzustellen, habe ich mich entschieden eine Heatmap zu verwenden.

So werden die vielen Datenpunkte über die Vertikale auf die Monate und über die horizontale auf die Tage des Monats aufgeteilt. Somit entsteht eine Tabellenähnliche Darstellung mit 12 Zeilen und 31 Spalten.

Um diese Anzeige minimalistisch darzustellen werden jedoch keine Daten in Form von Zahlen gezeigt. Dies soll durch die Erstellung von Zahlenbereichen, welche durch verschiedene Farbstufen verdeutlicht werden sollen, minimiert werden.

Hervorhebungen der Heatmap werden durch eine separate Tabelle sowie durch kurze Texte verdeutlicht.

Für die Visualisierungen von der Wochentags- und Monatsverteilung sowie der Jahresentwicklung

wurden Säulendiagramme verwendet.

Um die ebenfalls so minimalistisch wie möglich Darzustellen, wurde auf die aktive Verwendung der Y-Achse verzichtet.

Ebenso wird, wie auch bei der Heatmap, auf die Anzeige von Zahlen verzichtet.








Wenn der Nutzer bei einer der vier Datenvisualisierungen genaue Informationen zu dem Datenpunkt erhalten möchte, wird er dies durch die Interaktionsmöglichkeiten in der Darstellung tun können. So werden ihm dann die genauen Zahlen und Daten dargestellt.

Layout

Farben

Für die Visualisierung der vielen verschiedenen Datenpunkte wurde eine Varianz aus verschiedenen Farbtönen gewählt. So ist es möglich für die unterschiedlichen Häufigkeiten von Geburten in der Heatmap eine andere Farbstufen zu nutzen.

Explizit wurden folgende Farbtöne verwendet:

-  #ffe6b5 - bis zu 10.000 Geburten
-  #ffd480 - zwischen 10.000 und 10.400
-  #ffb974 - zwischen 10.400 und 10.800
-  #ff9c68 - zwischen 10.800 und 11.200
-  #ff805c - zwischen 11.200 und 11.600
-  #ff644f - zwischen 11.600 und 12.000
-  #ff4542 - über 12.000 Geburten

Die Farbabstufung der Orangetöne hat den Zweck, eine gewisse Zahlenspanne zu markieren. Somit ist es dem Nutzer ohne genaue Analyse möglich, erste Eindrücke und grobe Werte von bestimmten Datenpunkten zu assoziieren.

Möchte er genauere Informationen über einen Datenpunkt erhalten, so kann der dies mit durch die Interaktion erfahren.

So werden die Tage, an denen durchschnittlich weniger Geburten stattgefunden haben mit einem stark verblassten Orange verdeutlicht. Tage, an denen durchschnittlich viele Geburten stattgefunden haben, werden durch ein sehr kräftigen Orangeton visualisiert. Die Zahlenbereiche zwischen den beiden Extremen werden weitere Farbabstufungen verdeutlicht.

Bei den Säulendiagrammen wurden diese Farben ebenfalls verwendet. Es wurde sich jedoch dagegen entschieden bestimmte Zahlen durch einen stärkeren spezifischen Farbton hervorzuheben, da dies bereits durch die Höhe der Säule geschehen ist.

Die Farben wurden nun in einem Verlauf angewandt, so dass der in der Höhe der Säule der Farbton kräftiger wird.

Ähnlich wie in den Säulendiagrammen wird der

Farbverlauf auch in dem Logo angewandt. In diesem Falle ist es jedoch ein horizontaler statt vertikaler Verlauf.

Es wurden Orangetöne gewählt, da ich keine unbeabsichtigte Assoziation zu Geschlechtern bilden wollte. So wird durch ein Blau oder ein Lila vermutlich die Erwartung aufgebaut, dass hinter dem Datenpunkt auch eine Erläuterung stecken könne, ob mehr Jungs oder Mädchen an dem Tag geboren wurden. Dies hat der Datensatz jedoch nicht hervorgehoben, weshalb die Datenpunkte Geschlechtsneutral verdeutlicht werden sollten.

Typografie

Für die Überschriften auf der Webseite habe ich die Schrift IBM Plex Sans ausgewählt. Neutral, aber dennoch leicht schafft sie eine freundliche Atmosphäre. Zudem ergreift sie durch ihre sowohl scharfen Kanten und Runden das Auge des Betrachters.

Etwas besonderes dieser Schrift ist, dass sie eine Variable Font ist. Somit besteht die Möglichkeit, die Schrift durch Code in ihren Eigenschaften dynamisch zu verändern, welches auf dieser Seite angewandt wird.

Durch die Veränderung in den Eigenschaften entsteht Beweglichkeit, welches auch mit Veränderungen assoziiert werden kann, was oftmals durch Geburten passiert.

Ebenfalls wird dadurch die Verteilung von Geburten aufgegriffen, indem dickere Schrift mit einem höheren Durchschnitt von Geburten in assoziiert werden kann.

Für Beschreibungen oder Erläuterungen habe ich mich für Open Sans entschieden.

Sie ist eine Schrift, die selbst in kleiner Form sehr leserlich ist. Durch ihre offenen Formen schafft sie ein neutrales, aber vor allem freundliches Erscheinungsbild, welches ideal zum Rest des Designs sowie der Thematik passt.

Layout

Hintergrund

Für den Hintergrund habe ich mich dazu entschieden, keinen eintönigen Farbton zu verwenden, sondern einen Gradienten in dem mehreren Farben ineinander Verlaufen.

Dieser Gradient besteht aus Hell- sowie Dunkelgrün und einem Gelbton.

Ebenfalls wurde hier darauf geachtet, dass die Farben keine Assoziation zu Geschlechtern aufzubauen. Der gesamte Hintergrund ist mit den Farbtönen der Visualisierungen abgestimmt, um ihn nicht zu sehr mit den Farben der Datenvisualisierungen konkurrieren zu lassen. So steht der Fokus des Nutzers auf den Farben der Visualisierungen und nicht auf dem Hintergrund.

Die genauen Farbtöne für den Gradienten sind:

#81FF72, #ECFF4F, #48FF8B



Glasmorphismus

Ich habe mich dazu entschieden Glasmorphismus auf der Webseite anzuwenden. Das heißt, dass die Inhalte der Webseite auf mattem, verschwommenen Glas sind.

Überschriften sind durch ihre dicke der Schrift direkt auf der ersten Ebene des Glas, da dort die Leserlichkeit sichergestellt ist.

Durch die Verwendung von einer weitere Ebene des Glas wird sicher gestellt, dass Inhalte bei denen die Leserlichkeit zu gewährleisten ist, besser zu erkennen sind.

Dieser Effekt gibt dem Hintergrund die Möglichkeit in seiner Erscheinung etwas präsenter zu sein, als wenn die Inhalte der Webseite auf soliden Flächen vor dem Hintergrund gezeigt werden.

Durch die Leichtigkeit der Elemente und der vielen Farben soll eine fröhliche Stimmung erzeugt werden, welche oftmals bei Geburten herrscht.

Layout

Typografie

Für die Überschriften auf der Webseite habe ich die Schrift IBM Plex Sans ausgewählt. Neutral, aber dennoch leicht schafft sie eine freundliche Atmosphäre. Zudem ergreift sie durch ihre sowohl scharfen Kanten und Runden das Auge des Betrachters.

Etwas besonderes dieser Schrift ist, dass sie eine Variable Font ist. Somit besteht die Möglichkeit, die Schrift durch Code in ihren Eigenschaften dynamisch zu verändern, welches auf dieser Seite angewandt wird.

Durch die Veränderung in den Eigenschaften entsteht Beweglichkeit, welches auch mit Veränderungen assoziiert werden kann, was oftmals durch Geburten passiert.

Ebenfalls wird dadurch die Verteilung von Geburten aufgegriffen, indem dickere Schrift mit einem höheren Durchschnitt von Geburten in assoziiert werden kann.

Für Beschreibungen oder Erläuterungen habe ich mich für Open Sans entschieden.

Sie ist eine Schrift, die selbst in kleiner Form sehr leserlich ist. Durch ihre offenen Formen schafft sie ein neutrales, aber vor allem freundliches Erscheinungsbild, welches ideal zum Rest des Designs sowie der Thematik passt.

Navigation

Struktur

Ich habe ich dazu entschieden, die visuelle Sprache auch in den Aufbau der Webseite einzubinden. Durch den Minimalismus der Webseite entsteht eine gewisse Leichtigkeit, die ebenfalls durch die Struktur zu spüren sein soll.

So wurde aktiv die Entscheidung getroffen, alle Elemente in einer Webseite ohne Verlinkungen auf andere Seiten darzustellen. Durch die Entscheidung für einen Onepager ist auch entschieden worden, dass kein Menü auf der Seite notwendig ist.

Um die Seite jedoch nicht unendlich lang werden zu lassen, werden Bereiche zusammengeführt. Der Startbereich hat mit einer kleiner Darstellung einer Zahl und einer kurzen Fragestellung zu der Zahl Thematisch einen kleinen Einblick gewährleistet.

Durch den Einsatz der variablen Font ist ebenfalls bereits einen kleiner Eindruck vermittelt worden, dass durch das Interagieren mit Elementen neue Ansichten und Inhalte zum vorscheinen kommen.

Im Anschluss wird die Hauptvisualisierung der Seite, also die Tagesübersicht gezeigt.

Da es jedoch viele Datenpunkte zum Entdecken gibt, ist die Entscheidung getroffen worden auf der selben Ebene der Navigation, also sozusagen horizontal, bestimmte Datenpunkte hervorzuheben.

So wird auf der selben Ebene thematisch tiefer in die Daten eingetaucht. Explizit wird dort auf die 10 Tage eingegangen, an denen sowohl die meisten als auch wenigsten Geburten durchschnittlich stattgefunden haben. Ebenfalls werden Besonderheiten, die einem beim Betrachten der Heatmap auffallen könnten, etwas genauer hervorgehoben und analysiert.

Scrollt man weiter nach unten kommt man zu den anderen Visualisierungen. Diese werden jedoch nicht alle vertikal nacheinander gezeigt, sondern ähnlich wie bei der ersten Ebene angezeigt.

Aufgrund dessen, dass bei jeder der Visualisierun-

gen die Daten anhand von Säulendiagrammen verdeutlicht werden, habe ich es für sinnvoll betrachtet, sie auf einer Ebene zu bündeln.

Abschließend ist zu Ende der Seite die eine Ebene, bei der ein wenig Kontext zu dem Datensatz gegeben wird, als auch auf die Quelle des Datensatzes verwiesen wird.

Interaktionen

Zu Beginn der Seite hat der Nutzer die Möglichkeit durch seine Mausinteraktion die Zahlen aus „85.386.227“ in ihrer breite und dicke zu verändern. Dies soll dem Betrachter auf einer spielerischen Art und Weise darauf hinweisen, dass es auf dieser Webseite um die Verteilung von Geburten geht.

Alle weiteren Interaktionen auf der Webseite folgen dem identischen Prinzip: Das durchstöbern mit der Maus gibt dem Nutzer die Möglichkeit mehr zu erfahren.

Geht ein Nutzer mit der Maus über einen Datenpunkt, so bekommt er für diesen nun genaue Zahlen. Vorher konnte der Nutzer lediglich durch die Farbe des Datenpunkts in der Heatmap oder der Höhe bei den Säulendiagrammen erahnen um welche Zahl es sich handeln könnte.

Will der Nutzer mehr über das Thema herausfinden muss er über die Menüführung innerhalb der Ebene dies auswählen.

Das selbe gilt für die Ebene, auf der die Säulendiagramme angeordnet sind.

Realisierung Prototyp

Dieses Projekt wurde mithilfe von HTML, CSS und JavaScript programmiert und ist für die Bildschirme eines PCs, Laptops und eines Tablets optimiert.

so dass der Nutzer den Bereich besser erkennen kann.

Besonderheiten

Um das Erlebnis der Seite spezieller zu machen, habe ich mich entschieden, das Scroll-Verhalten zu verändern.

Das Verhalten wird so manipuliert, dass der Nutzer immer an bestimmten Positionen landet. Dies nennt sich Hard-Scroll Effekt.

Sobald der Nutzer nach unten scrollt, landet der Nutzer nicht wie üblich bei dem Endpunkt wo er aufgehört hat zu scrollen, sondern praktisch eine ganze Bildschirmhöhe weiter unten.

So werden alle Inhalte und Visualisierungen zu jedem Zeitpunkt auf dem Bildschirm zentriert dargestellt und sind nie abgeschnitten.

Die Datenvisualisierungen wurden mithilfe der JavaScript Library „ApexCharts.js“ dargestellt. Hierfür werden die von mir angepassten Datensätze in dem JavaScript Dokument eingebunden und durch das von den bereitgestellten Skript visualisiert.

Das Design der Visualisierungen wurde jedoch an mein Design angepasst. So wurden je nach Datenvisualisierung die Farben, Formen und Achsen angepasst.

Ebenfalls wurde der Tooltip, also die Anzeige von genauen Zahlen des spezifischen Datenpunktes, die man beim Halten über dem Punkt erhält, verändert. Dort wird ebenfalls der Effekt des Glas-Morphismus angewandt, welcher auf der Seite zu finden ist.

Eine weitere, von der Library bereitgestellte, Interaktion ist das Hervorheben von mehreren Datenpunkten innerhalb der Heatmap. Hierfür muss der Nutzer lediglich mit der Maus über einen Punkt der Legende hovern. In dem Fall werden dann alle anderen Zahlenbereiche in ihre Farbe reduziert,

Fazit

Das Endergebnis spiegelt in meinen Augen vieles dar, was auch in den Lehrveranstaltungen „Interaktive Datenvisualisierung“ gelehrt wurde.

Komplexe Datensätze so simpel wie möglich darzustellen, wie es in der Geschichte oftmals fantastisch vorgemacht wurde, ist keine einfache Arbeit. Durch die Reduktion auf die wichtigsten und relevantesten Elemente werden dem Betrachter von Beginn an die nicht zwingend notwendigen Inhalte verborgen.

So muss der Mensch nicht immer eine Menge von Zahlen lesen, um zu verstehen wie es um eine Thematik aussieht. Dies kann ebenfalls durch Interaktionen und gestalterische Mittel geschehen.

All dies habe ich versucht in meinem Projekt anzuwenden. So verzichte ich, sofern sie nicht zwingend nötig sind, auf Legenden.

Der Nutzer verliert jedoch durch den Verzicht keine Informationen, da diese durch andere Wege in die Visualisierung eingebunden werden. So wird durch Farbcodierung oder der Höhe von Säulendiagrammen dem Nutzer bereits klar, wie diese Datenpunkte sich im Vergleich zu anderen Werten einordnen lassen.

Möchte der Nutzer dann genauere Informationen zu einem Datenpunkt, so profitiert er durch die eingebauten Interaktionen.

Durch das hovern über eine Instanz bekommt er die genaue Zahl des jeweiligen Datenpunktes angezeigt.

Den Ansatz der Simplifizierung habe ich auch auf die Webseite übertragen. Ebenso wie bei den Interaktionen der Datenvisualisierung bekommt der Nutzer durch das scrollen oder auswählen von anderen Menüpunkten in der Ebene die Chance, weitere Informationen zur Visualisierung oder auch weitere Visualisierungen zu erhalten.

Die Webseite wird dadurch nicht unendlich lang, wodurch der Nutzer möglichst lange auf der Seite bleiben mag, da er nicht den Eindruck bekommt, von Informationen überladen zu werden.

Vielmehr bekommt der Nutzer so die Möglichkeit

sowohl bei den Visualisierungen als auch bei der Webseite sich Interaktiv tiefer mit dem Thema zu befassen, falls er durch den ersten Eindruck bereits Interesse bekommen hat.



Ein Projekt von:

