

# Visualisierung von geobasierten Daten aus einem E-Commercesystem

TOBIAS CURTH



MASTERARBEIT

eingereicht am  
Fachhochschul-Masterstudiengang

Interactive Media

in Hagenberg

im Juni 2015

© Copyright 2015 Tobias Curth

Diese Arbeit wird unter den Bedingungen der *Creative Commons Lizenz Namensnennung–NichtKommerziell–KeineBearbeitung Österreich* (CC BY-NC-ND) veröffentlicht – siehe <http://creativecommons.org/licenses/by-nc-nd/3.0/at/>.

# Erklärung

Ich erkläre eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den benutzten Quellen entnommenen Stellen als solche gekennzeichnet habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Hagenberg, am 29. Juni 2015

Tobias Curth

# Inhaltsverzeichnis

<b>Erklärung</b>	<b>iii</b>
<b>Kurzfassung</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Zielsetzung . . . . .	2
1.2 Gliederung . . . . .	3
<b>2 Kartografie</b>	<b>5</b>
2.1 Begriffsdefinition . . . . .	5
2.2 Kartentypen . . . . .	7
2.2.1 Topografische Karte . . . . .	7
2.2.2 Thematische Karte . . . . .	8
2.3 Abbildungstechniken . . . . .	8
2.3.1 Räumliche Dimension . . . . .	9
2.3.2 Visuelle Variablen . . . . .	9
2.3.3 Modell von geografischen Phänomene . . . . .	12
2.3.4 Choroplethenkarte . . . . .	13
2.3.5 Proportional Symbol . . . . .	14
2.3.6 Isoplethenkarte . . . . .	15
2.3.7 Punktstreukarte . . . . .	15
2.3.8 Heatmap . . . . .	16
2.4 Statische vs. dynamische Karte im WWW . . . . .	17
2.5 Digitale Geodaten und -formate . . . . .	18
<b>3 Erweitertes Hintergrundwissen</b>	<b>20</b>
3.1 Electronic Commerce . . . . .	20
3.2 Big Data . . . . .	21
3.3 Data-Mining und Visual Analytics . . . . .	21
3.4 Geoinformationssysteme . . . . .	22
3.5 Geokodierung . . . . .	22
3.6 INSPIRE . . . . .	23



<b>4</b>	<b>Eigener Entwurf</b>	<b>24</b>
4.1	Anforderungen und Voraussetzungen . . . . .	24
4.1.1	E-Commerce-System . . . . .	24
4.1.2	Primärdaten . . . . .	25
4.1.3	Sekundärdaten . . . . .	27
4.1.4	Interaktivität und Anpassungsmöglichkeiten . . . . .	28
4.1.5	Visualisierungstypen . . . . .	29
4.1.6	Echtsystem . . . . .	29
4.2	Benutzeroberfläche . . . . .	29
4.2.1	Karte . . . . .	30
4.2.2	Optionenbereich . . . . .	31
<b>5</b>	<b>Umsetzung</b>	<b>33</b>
5.1	Verwendete Frameworks und Technologien . . . . .	33
5.2	Repräsentation von geografischen Daten . . . . .	35
5.3	Architektur und technische Ausführung . . . . .	36
5.4	Implementierung . . . . .	37
5.4.1	Request . . . . .	40
5.4.2	Response . . . . .	41
5.5	Nutzungsmöglichkeiten für Echtsysteme . . . . .	42
5.5.1	Adaption der PHP-Schnittstelle . . . . .	42
5.5.2	Adaption des Datenbankschemas mithilfe von Views . . . . .	42
<b>6</b>	<b>Evaluierung</b>	<b>43</b>
6.1	Teilnehmer und Durchführung . . . . .	43
6.2	Ergebnisse . . . . .	44
6.2.1	Zielgruppenanalyse . . . . .	44
6.2.2	Datenanalysetool . . . . .	45
6.2.3	Visualisierungsarten . . . . .	46
<b>7</b>	<b>Résumé</b>	<b>49</b>
	<b>Quellenverzeichnis</b>	<b>51</b>
	Literatur . . . . .	51
	Online-Quellen . . . . .	53

# Kurzfassung

Daten über Websitebesucher aus einem E-Commercesystem, z.B. einem Webshop, werden in aller Regel mit Diagrammen oder Tabellen dargestellt. Sobald geobasierte Informationen zu einem Datensatz zur Verfügung stehen, können diese auch auf einer Karte visualisiert werden. Marketinganalysten nutzen solche Datenverkehrsanalysetools, um neues Wissen über die Websitebesucher zu generieren und somit auf den Markt besser und schneller reagieren zu können. So lässt sich daraus u. a. ableiten, welche Produkte in dem Webshop profitabel sind oder welche Region besonders häufig ein bestimmtes Produkt kauft. Auch kann die Herkunft der Besucher so einfach ermittelt werden.

Problematischer dagegen ist das Auffinden von Mustern, Abhängigkeiten oder versteckten Informationen. Diese Phänomene sind mithilfe von Diagrammen oder Tabellen oftmals nur schwer zu konstatieren. Dagegen hat die Darstellung von Informationen auf Karten den Vorteil, dass Muster durch den Menschen ohne Weiteres erkannt werden können. Bedingt durch die Abbildung der Informationen auf der Karte ist gegeben, dass rasch erkannt werden kann, in welcher Region ein Datensatz entstand.

Diese Masterthesis beschäftigt sich mit der Frage, ob und welche Kartenvisualisierungsarten sich zum Analysieren von Informationen aus einem E-Commercesystem eignen. Zuerst werden Grundlagen zur Kartographie vorgestellt sowie wesentliche Kartenvisualisierungen aufgezeigt. Auch werden typische Abbildungstechniken für die Kartenherstellung erörtert. Das darauffolgende Kapitel widmet sich weiteren Grundbegriffen bzw. -technologien. Für die Evaluierung von geeigneten Visualisierungsarten wurde ein Prototyp erstellt. Dieser kann Informationen aus einem fiktiven Webshop mit verschiedenen Darstellungen figurieren. Die Arbeit beschreibt die gegebenen Anforderungen, Konzeption sowie die Umsetzung des Prototypen.

# Abstract

Data about website visitors from an e-commerce system, such as a web shop, are usually displayed with charts and tables. Once spatial data is available they can also be visualized on an online map. Marketing analysts use such web analytic tools to generate new knowledge about the site visitors and thus be able to respond better and faster to the market. It can be derived, which products are most profitable in the web shop or which region most frequently buys a particular product. Even the origin of the visitors can be easily determined.

However, more problematic is the discovery of patterns, dependencies or hidden information. These phenomena are often difficult to be stated using charts or tables. In contrast, the presentation of information on maps has the advantage that pattern can be recognized by humans without further ado. Due to the way of imaging of the information on the maps, it is easily to identify in which region a data record occurred.

This master thesis discusses the question whether and which map visualization types are suitable for analyzing information from an e-commerce system. First of all, basics about the topic cartography are presented and essential map visualizations are shown. The following chapter is devoted to other fundamental concepts and technologies. For the evaluation of appropriate visualization types, a prototype was implemented. The prototype figures information from a fictitious web shop with various visualization types. This work describes the given requirements, design concepts and realization of the implementation of the prototype.

# Kapitel 1

## Einleitung

Heutzutage sind Informationen über Websitebesucher eine essentielle Komponente in der elektronischen Geschäftsabwicklung. Besonders bei E-Commercesystemen, sprich Orten, an denen Konsumenten Produkte oder Dienstleistungen über das World Wide Web kaufen, können diese zum Erfolg des Unternehmens beitragen. Das Internet hat mittlerweile einen sehr entscheidenden Einfluss auf das Kaufverhalten des Menschen. Durch die stetig wachsende Wichtigkeit des Onlinehandels für die Unternehmen hat sich auch das Online-Marketing zu einer unerlässlichen Kernaufgabe beim operativen Geschäft entwickelt.

Dabei bestimmen nicht mehr nur Search-Engine Optimierungen (kurz SEO) eine entscheidende Rolle, um neue Kunden zu gewinnen, sondern vielmehr das schnelle Reagieren auf die neuen Gegebenheiten des Online-marktes. Im Gegensatz zu herkömmlichen Medien, wie z. B. Fernsehen oder Zeitungen, können bei einer Website eine Vielzahl an Informationen von Besuchern gemessen und ausgewertet werden. So kann heute nicht mehr nur das Verhalten des Besuchers auf der Website selbst protokolliert werden, auch Fakten, die von anderen Websites ausgewertet wurden, können in den Prozess der Wissensgenerierung einfließen.

Letztendlich müssen diese Massen an Daten visualisiert und ausgewertet werden. Aus dieser Anforderung entstand auch das Fachgebiet der Marketinganalysten. Diese *durchforsten* die Informationen nach Merkmalen, um neue Erkenntnisse und schlussendlich neues Wissen zu generieren.

Inzwischen gibt es eine Vielfalt an, teils unterschiedlichen, sogenannten Datenverkehrsanalysewerkzeugen. Diese machen das Aufzeichnen der Daten einfach und schnell. So muss z. B. bei Google Analytics, eines der bekanntesten Datenverkehrsanalysewerkzeuge, lediglich ein kleiner Teil Code in die Website eingefügt werden. Indes ist auch das Auswerten der Daten, dank flexibler und anpassbarer Berichte und Diagramme, einfacher geworden.

Problematischer dagegen wird die stetig wachsende Menge an Informationen. Wurde früher nur das Verhalten aufgezeichnet, werden heutzutage

immer mehr Details des Benutzer protokolliert. So kann z. B. auch ermittelt werden, wie lange ein Kunde tatsächlich benötigt hat, um sich das Produkt letztendlich zu kaufen. Dies ist besonders interessant, da sich teilweise Kunden sehr lange über andere Websites informieren und sich erst nach der Produktrecherche zum Kauf entscheiden. Werbung kann so mit Hilfe dieser Informationen auf Informationsplattformen personalisiert angezeigt werden und infolgedessen den Umsatz steigern.

Marketinganalysten visualisieren Informationen meist mit Hilfe von individuell anpassbaren Diagrammen oder Tabellen in einem Analyseprogramm. Sobald geografische Daten vorhanden sind, z. B. der Standort des Besuchers, können diese auch auf einer Karte angezeigt werden. Mittlerweile können Daten mit verschiedenen Visualisierungsarten dargestellt werden, gleichwohl nicht jede Darstellung für jede Art von Information gleich sinnvoll ist. Die Herausforderung ist, eine möglichst optimale Darstellung zu finden, sodass neues Wissen generiert werden kann. Wird die Darstellung von Informationen aus einem Webshop betrachtet, so ergeben sich auch hieraus Anforderungen. So müssen Informationen oftmals zuallererst gruppiert werden. Würde jede einzelne Information auf der Karte dargestellt werden, würde diese schnell unübersichtlich werden und schlussendlich die Generierung von neuem Wissen negativ beeinträchtigen.

Ein weiteres Problem bei der Analyse von Daten ist, dass Muster, Abhängigkeiten oder versteckte Informationen oftmals nur schwer erkennbar sind. Abgesehen davon, dass heutzutage diese zwar durch Algorithmen auch erkannt werden können, so ist die Mustererkennung durch einen Analysten dennoch teilweise genauer. Ein Algorithmus kann lediglich aus bekannten und definierten Regeln Muster erkennen, sodass der Algorithmus nur so viel Wert ist wie die vorhandenen Regeln. Menschen dagegen nutzen ihre eigenen Erfahrungen und können auch einfacher Neues erlernen. Besonders bei unbekannten bzw. neuen Phänomenen ist dies hilfreich, da auch ohne vorhandene Erfahrung diese erkundet und daraus neues Wissen abgeleitet werden kann.

## 1.1 Zielsetzung

Wie eben beschrieben, werden in der Gegenwart Informationen aus einem Webshop oftmals mithilfe von Diagrammen oder Tabellen visualisiert. Versteckte Informationen oder Muster sind damit jedoch oftmals nur mit viel Erfahrung erkennbar. Die Herausforderung für Marketinganalysten besteht darin, eine optimale Visualisierung der vorhandenen Informationen herauszufinden, um schlussendlich daraus neue Erkenntnisse zu gewinnen. Sind geobasierte Daten, z. B. der Standort von Besuchern, vorhanden, so können diese auch auf einer Karte dargestellt werden. Im Gegensatz dazu kann ein Mensch grafische Muster leichter erkennen, als diese aus Zahlen oder einem

Diagramm abzuleiten. Auch sind Verteilungen oder Ausnahmen ohne großen Aufwand auf einer Karte ersichtlich.

Ziel dieser Masterthesis besteht darin, Visualisierungsarten für die Darstellung von Informationen aus einem Webshop zu evaluieren. Dabei sind stets geobasierte Informationen in den Datensätzen vorhanden, sodass diese auch auf einer Karte dargestellt werden können. Die Evaluierung verschiedener Visualisierungsarten soll Aufschluss geben, ob eine Darstellung von Informationen auf einer Karte auch für Marketinganalysten hilfreich sein kann. Die Kartendarstellung der Informationen soll nicht die üblichen Diagramme bzw. Tabellen aus einem Datenverkehrsanalysewerkzeug ersetzen, sondern lediglich eine weitere Option zur Wissensgenerierung ermöglichen.

Das Ergebnis der Evaluierung soll zeigen, welche Visualisierungsarten von Karten generell für die Beantwortung von üblichen Fragen bei der Datenanalyse eines Webshops geeignet sind. Durch die Gegenüberstellung verschiedener Visualisierungsarten können Tendenzen abgeleitet werden. Des Weiteren soll das Ergebnis offenbaren, ob die gestellten Fragen nicht auch durch ein Diagramm hätten beantwortet werden können. Abschließend sollen die Visualisierungen einzeln bewertet werden, um Vor- und Nachteile zu evaluieren.

Für die Erreichung des Zieles ist ein Prototyp zu entwickeln, der Informationen aus einem Webshop auf einer Karte visualisieren kann. Die Visualisierung soll über verschiedene Parameter konfiguriert werden können. Auch soll die Möglichkeit vorhanden sein, Informationen aus verschiedenen Datenquellen anzuzeigen.

## 1.2 Gliederung

Kapitel 2 beschreibt Grundlagen der *Kartografie*. Dabei wird der Begriff aus Sicht der Wissenschaft definiert und abgegrenzt. Auch werden wesentliche historische Meilensteine betrachtet. Ein Fokus liegt dabei auf der modernen Zeit, da sich heutzutage insbesondere digitale Karten durchgesetzt haben. Des Weiteren werden Merkmale von topografischen sowie thematischen Kartentypen diskutiert.

Möglichkeiten zur Datenvisualisierung auf (digitalen) Karten werden im Abschnitt *Abbildungstechniken* erläutert. Im Zuge dessen wird die *räumliche Dimension* sowie der Term *visuelle Variablen* beschrieben. Ein weiterer Abschnitt widmet sich dem *Modell von geografischen Phänomenen*. Außerdem werden unterschiedliche Kartogramme, wie z.B. Choroplethenkarte, Proportional Symbol, Isoplethenkarte, Punktreuekarte oder Heatmap, durch Beispiele exemplifiziert. Zusätzlich wird der Unterschied zwischen statischen und dynamischen Karten im World Wide Web erörtert. Im letzten Abschnitt werden wesentliche digitale Geodaten und -formate beschrieben.

Das dritte Kapitel informiert über weitere Grundlagen sowie Begriffs-

definitionen, wie z. B. Electronic Commerce, Big Data, Data-Mining sowie Visual Analytics. Anschließend werden Geoinformationssysteme, inklusive deren Anwendungsgebiete, verdeutlicht. Einen Einblick in die Geokodierung befindet sich in Abschnitt 3.5. Die Richtlinie des europäischen Parlaments und Rates zur *Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft* wird im letzten Abschnitt expliziert.

Das Kapitel 4 mit dem Titel *eigener Entwurf* liefert einen Einblick in die wesentlichen Anforderungen und Voraussetzungen des entwickelten Prototypen. Hier wird zudem das Konzept der Benutzeroberfläche beschrieben.

Die *Umsetzung* des Prototypen wird in Kapitel 5 erörtert. Dabei werden u. a. die *verwendeten Frameworks und Technologien* beschrieben. Auch wird die *Repräsentation von geografischen Daten* in diesem Prototypen exemplifiziert. Außerdem wird die grundlegende Architektur sowie technische Ausführung diskutiert. Zum Schluss werden wesentliche Teile der Implementierung erörtert.

Für die *Evaluierung* wurde ein Fragebogen erstellt. Dieser wird in Kapitel 6, inklusive der gegebenen Anforderungen sowie der Durchführung, vorgestellt. Die Ergebnisse der einzelnen Fragen werden anschließend separat betrachtet und vorgestellt.

Abgerundet wird diese Arbeit im Kapitel 7 mit einer Zusammenfassung über die Ergebnisse durch die Evaluierung des Prototypen und schließt mit einem kurzen möglichen Ausblick für die Zukunft ab. Der Anhang enthält ein Verzeichnis aller verwendeten Literaturquellen.

Auf der beigelegten DVD kann u. a. der Fragebogen und die Ergebnisse der Evaluierung sowie der Quellcode des Prototypen eingesehen werden. Auch befindet sich dort eine digitale Version dieser Masterthesis.

## Kapitel 2

# Kartografie

In diesem Kapitel werden wesentliche Grundlagen und Begrifflichkeiten zur Kartografie erläutert. Zu Beginn wird in Abschnitt 2.1 der Begriff definiert. In Abschnitt 2.2 werden Merkmale von topografischen sowie thematischen Karten beschrieben. Abbildungstechniken bei der Erstellung von Karten werden in Abschnitt 2.3 erörtert.

### 2.1 Begriffsdefinition

Kartografie ist die *geografische Wissenschaft und Kunst Karten zu gestalten, zu konstruieren und zu erzeugen* [26] und beinhaltet dabei diverse Prozesse zur Herstellung und Vertrieb von Karten (*engl. map*). Außerdem wird sie auch als Wissenschaft der menschlichen Kommunikation verstanden [14]. Ziel der Kartografie ist, Fakten und Ideen mithilfe von Zeichnungen, Wörtern und Symbolen zu kommunizieren. Die Detailtiefe kann dabei je nach Zielgruppe und Funktion variieren.

Nach Kraak [19] werden Karten verwendet, um Geodaten, auch Geoobjekte genannt, zu visualisieren. Daten von Positionen und Attributen von Objekten bzw. Phänomene beziehen sich in aller Regel auf die Erde.

Karten helfen Benutzern raumbezogene Zusammenhänge besser zu verstehen. Insbesondere können Informationen über Entfernungen, Richtungen und Flächengrößen abgeleitet, Muster erkannt und Beziehungen quantifiziert werden. Die Abschnitte 2.3, 2.5 sowie 5.2 beschreiben Möglichkeiten zur Repräsentation von Geodaten bzw. digitale Geoformate.

Historisch betrachtet, gibt es Karten seit über 8000 Jahren. Vor allem anfangs wurden diese erstellt, um anderen erkundete Orte zu beschreiben [14]. Dies erforderte Wissen über die reale Welt sowie über die Übertragung der Informationen auf eine Karte. Misra [26] et al. unterteilt die Entwicklung in der Kartografie in vier Epochen: Antike (bis 400 n. Chr.), Mittelalter (400 bis 1500 n. Chr.), Frühneuzeit (1500 bis 1900 n. Chr.) sowie Moderne (ab 1900 n. Chr.). Jedoch ist der Entwicklungsstand je nach vorhandenem



(technologischen) Wissen unterschiedlich (gewesen), weshalb eine Epoche nicht immer eindeutig abgrenzbar ist.

Seit den 1980er Jahren hat sich in der Verarbeitung von digitalen Geodaten eine Eigendynamik entwickelt. Die Art und Weise wie Benutzer Karten nutzen konnten, änderte sich erheblich. Durch die Entwicklung in der Digitalisierung von Informationen durch Computer entstanden folglich auch *on-screen* Karten. In Folge dessen konnten Abfragen, meist basierend auf einer Datenbank, generiert werden und einfache analytische Funktionen realisiert werden. In den 1980ern entstand hierfür der Oberbegriff *Geoinformationssystem* (engl. *geographic information system*), kurz *GIS* (eine genauere Begriffserklärung befindet sich in Abschnitt 3.4). Nun konnten auch verschiedene Quellen zeitgleich genutzt werden, was neue Möglichkeiten in der Datenmanipulation, -analyse und -visualisierung ermöglichte [19].

Die Verbreitung und Weiterentwicklung des Internets – insbesondere in den 1990er Jahren<sup>1</sup> – brachte die nächste Revolution in der Kartographie. Mit dem Zugang zu interaktiven Karten, wie z. B. Google Maps<sup>2</sup> oder OpenStreetMap<sup>3</sup>, konnte jedermann professionelle Kartenanwendungen nutzen, eigene Daten hinzufügen oder verbreiten.

Heutzutage sind Kartendienste ein tagtäglicher Begleiter. Sie unterstützen uns Menschen in vielen Aktivitäten und Situationen. Eine Navigation zu einem bestimmten Ort ist dank eines satellitengestützten Systems keine Schwierigkeit mehr.

Auch in den Medien, z. B. im Fernsehen oder in der Zeitung, werden räumliche Informationen immer häufiger mit (thematischen) Karten dargestellt. So werden Übersichtskarten von Krisengebieten oder auch der Wetterbericht über solch eine Form der Informationsdarstellung präsentiert. Karten visualisieren eine Selektion von komplexen Echtzeiten in einer abstrakten Art. Jedoch sollten diese nicht nur als *Instrument zur Darstellung von Geodaten* gesehen werden, viel mehr bietet es durch Interaktionen mit Daten die Möglichkeit, Muster, Beziehungen und Trends zu erforschen [17].

Friendly und Denis [12] erstellten über *Meilensteine in der thematischen Kartographie, statistische Grafiken und Datenvisualisierung* eine interaktive Zeitleiste. Auf ihrer Website können weitere Erfindungen bzw. Entdeckungen, samt zusätzlicher Beschreibungen, abgerufen werden.

---

<sup>1</sup>Hier ist speziell die Entstehung des WWWs (World-Wide-Web) gemeint, welches von Tim Berners-Lee 1989 am CERN entwickelt wurde. Im WWW befinden sich verlinkte Dokumente (auch Websites genannt), die über das Internet abrufbar sind. Lee war einer der Pioniere, die einen erheblichen Einfluss auf die Digitalisierung von Daten und Informationen hatten. Siehe: <http://home.web.cern.ch/topics/birth-web>.

<sup>2</sup><http://maps.google.com>

<sup>3</sup><http://openstreetmap.org>

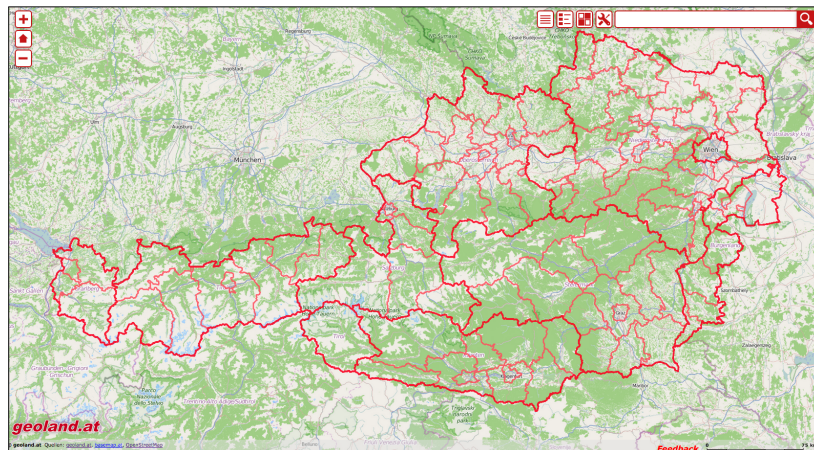
## 2.2 Kartentypen

Die Kategorisierung von Kartentypen kann in unterschiedlicher Art erfolgen. Einerseits ist es möglich, diese nach der Skalierung (klein, mittel oder groß) zu charakterisieren, andererseits kann die Art der Information bzw. des Zwecks herangezogen werden. Letzteres ergibt drei Kategorien: allgemeine (engl. *general reference map*), thematische (engl. *thematic or statistical map*) und spezielle Karten [26, 31].

Topologische Karten (siehe Abschnitt 2.2.1) und Atlanten werden dabei als allgemeine Karten deklariert. Thematische Karten (siehe Abschnitt 2.2.2) hingegen stellen Daten bzw. Informationen, wie z. B. die Bevölkerungsdichte oder das Wahlergebnis in einer Region, dar. Spezialkarten werden für Menschen mit Beeinträchtigung hergestellt [26]. Aufgrund fehlender Normen bzw. Kategorisierung von Kartentypen sind folgend die zwei wesentlichen aus Misra et al. detaillierter beschrieben.

### 2.2.1 Topografische Karte

Eine topografische Karte ist eine detaillierte und naturgetreue Abbildung von Geländeformen und natürlichen, sichtbaren Merkmalen der Erdoberfläche [5]. Merkmale, z. B. Straßen, Gebäude, Umrisse von Ortschaften, Gewässer, Berge oder Vegetation, erweitern die Karte mit zusätzlichen Informationen [15].



**Abbildung 2.1:** Beispiel für eine topografische Karte: Bundesrepublik Österreich mit Verwaltungsgrenzen (rote Linien). Erstellt mit *geoland.at*, *INSPIRE Austria*.

Außerdem befinden sich oftmals weitere Orientierungshilfen auf der Karte, wie ein Kompass oder Höhenlinien, sodass sich der Betrachter schneller zurecht findet. Aufgrund ihrer allgemeinen Bedeutung werden topografische

Karten meist von öffentlichen Behörden herausgegeben. In Österreich ist hierfür u. a. das Bundesamt für Eich- und Vermessungswesen zuständig [4].

Ein Beispiel für eine topografische Karte zeigt Abbildung 2.1. Die Karte stellt die Bundesrepublik Österreich, inklusive ihrer Verwaltungsgrenzen (rote Linien), dar. Auf der Website von *geoland.at* können diverse Merkmale, z. B. Naturschutz-, Bildungs- und Wasserinformationen, auf der Karte hinzugefügt werden.

### 2.2.2 Thematische Karte

Eine thematische Karte stellt Informationen zu einem Thema bzw. einer Thematik grafisch dar. Der Unterschied zu topografischen Karten besteht darin, dass zusätzlich zu natürlichen und sichtbaren Merkmalen, wie z. B. Bergen und Flüssen, weitere, spezifische Daten bzw. Attribute abgebildet werden [31]. Auch ist es möglich, bestimmte Merkmale abstrakt darzustellen, um den Fokus mehr auf die Thematik zu richten. Olbrich [28] et al. definiert eine thematische Karte als *eine Sammlung unterschiedlichster Zeichen und Formen, die im Idealfall zu einem Ganzen verschmelzen*.

In heutigen Kartenanwendungen müssen Nutzer üblicherweise mit Einstellungsmöglichkeiten experimentieren bevor ihre Anforderungen erfüllt und ihre Fragen beantwortet werden. Aufgrund der, in aller Regel unlimitierten, Anzahl an unterschiedlichen (Visualisierungs-) Optionen, z. B. Farben, Größe oder Selektierung der Daten, sind Nutzer oftmals überfordert bzw. finden es schwer, die geeignetste Darstellung zu determinieren. Ghamdi [13] gibt zusätzlich an, dass nur ein geschulter Nutzer imstande ist optimale Entscheidungen (Optionen) zu treffen.

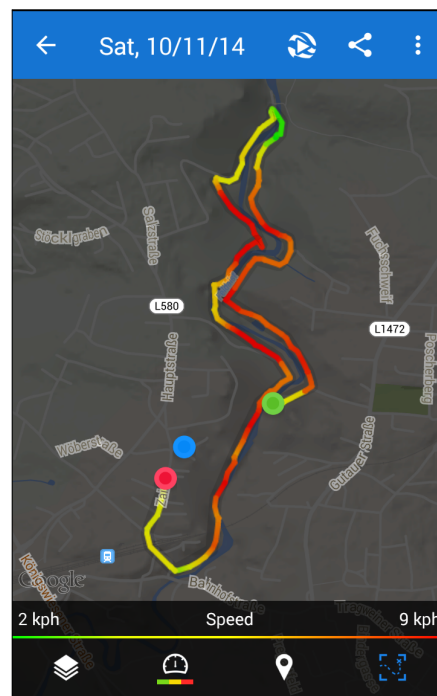
Aufgrund ihrer oftmals komplexen Darstellung sind thematische schwieriger zu interpretieren als topografische Karten. Daher ist es ratsam, nur wichtige und relevante Informationen darzustellen. Als ein Beispiel für eine thematische Karte zeigt die Abbildung 2.2 die Android App *Runtastic*<sup>4</sup>. Die App protokolliert sportliche Aktivitäten auf Basis von GPS-Daten. Auf der Verlaufskarte wird die aufgezeichnete Strecke farblich visualisiert. Je nach Einstellung kann dies die Geschwindigkeit oder auch die aktuelle Höhe bei der ausgeübten Aktivität sein.

## 2.3 Abbildungstechniken

Geodaten sind räumliche Elemente, die zusätzlich geometrische Informationen besitzen sowie zeitlichen Veränderungen unterliegen können [21]. Die folgenden Abschnitte beschreiben weitere wesentliche Merkmale von Geodaten. Außerdem werden typische Darstellungarten vorgestellt, die vor allem bei thematischen Karten genutzt werden.

---

<sup>4</sup><https://www.runtastic.com/>



**Abbildung 2.2:** Thematische Karte am Beispiel der Android App *Runtastic*. Eine zuvor gelaufene und aufgezeichnete Strecke wird farblich dargestellt. Die Legende im unteren Bereich zeigt die Zuordnung Farbe - Geschwindigkeit auf der Karte an. Auch Start- (rot), Zielpunkt (grün) sowie der aktuelle Aufenthaltsort (blau) werden farblich gekennzeichnet.

### 2.3.1 Räumliche Dimension

Generell können grafische Phänomene u. a. als Punkt (engl. *point*), Linie (engl. *linear*) oder Fläche (engl. *areal*) abgebildet werden [21, 31]. Punkte werden auch als nulldimensional bezeichnet, da sie keine Flächenausdehnung haben. Linien dagegen haben eine Dimension (Länge) und bestehen aus einer Menge von verbundenen und nicht geschlossenen Punkten. Flächen sind zweidimensional und haben sowohl eine Länge als auch eine Breite. Sie werden wiederum mittels Punkten definiert, jedoch müssen diese eine Fläche umgeben [20, 31]. Slocum beschreibt noch 2½- und dreidimensionale Phänomene (Körper), die jedoch aufgrund ihrer teils komplexen Spezifizierung in dieser Arbeit nicht weiter beschrieben werden.

### 2.3.2 Visuelle Variablen

Der Term *visuelle Variablen* ist allgemein üblich, um die verschiedenen perzipierten Unterschiede von Kartensymbolen zu beschreiben, die geografische Phänomene darstellen. 1983 wurde der Term von Jacques Bertin entwickelt

und mehrmals, u. a. von DiBiase oder MacEachren, modifiziert. Nachfolgend werden gebräuchliche Symbole aus Slocum [31] et al. beschrieben. Generell werden diese in zwei Kategorien unterteilt: quantitativ oder qualitativ. Während quantitative Symbole stets eine Ordnungszahl, Intervall oder Verhältnis reflektieren, werden qualitative für ein nominelles Messniveau verwendet. Abbildung 2.3 verdeutlicht die beschriebenen visuellen Variablen grafisch. De Lange [21] gibt zu bedenken, dass die Bedeutung der Zeichen sowohl beim Sender (z. B. Kartenhersteller) als auch beim Empfänger (z. B. Kartenleser) identisch sein müssen, um Fehlinterpretationen zu vermeiden.

### **Abstand**

Ein Abstand (auch Textur genannt) wird dazu genutzt, um quantitativ Distanzänderungen zwischen Markierungen innerhalb der Symbole darzustellen. Dabei symbolisiert ein geringerer Abstand in aller Regel einen höheren Datenwert.

### **Größe**


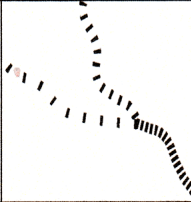
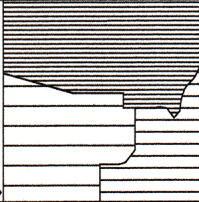


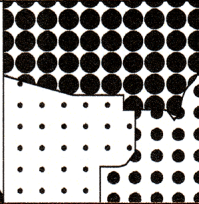


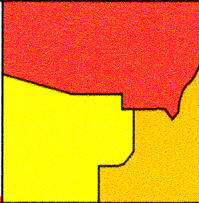

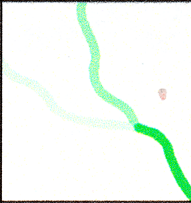
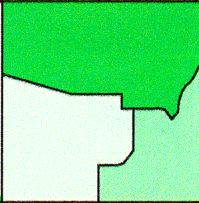


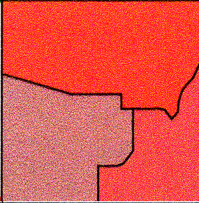

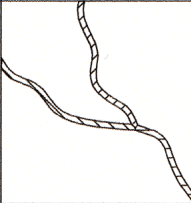
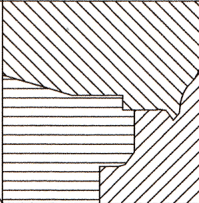

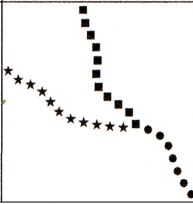
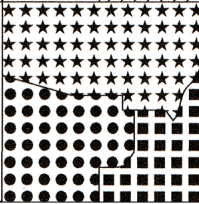
Auch die Größe der Symbole wird häufig für die Darstellung von Phänomenen verwendet. Je größer ein Symbol dargestellt wird, desto höher ist der zu repräsentierende quantitative Datenwert.

### **Farbton, Farbhelligkeit und Farbsättigung**

Farbton, Farbhelligkeit und Farbsättigung (engl. *hue, lightness and saturation*) sind allgemein bekannt als die Basiskomponenten von Farben. Der Farbton ist dabei der am häufigsten genutzte Parameter, um unterschiedliche Farben zu kreieren. Farbhelligkeit bezieht sich auf die Intensität der Leuchtkraft. Je weniger Helligkeit eine Farbe hat, desto dunkler (schwärzer) ist sie. Hingegen wirkt sich die Farbsättigung auf den Weißanteil der Farbe aus. Aufgrund der Kombinationsmöglichkeiten der drei Parameter können Farben gut quantitative sowie auch teils qualitative Unterschiede zwischen zwei Symbolen darstellen. Jedoch ist zu beachten, dass Farben sowohl von Mensch-zu-Mensch als auch von Bildschirm-zu-Bildschirm unterschiedlich wahrgenommen werden könnten, weshalb stets eine eindeutige Werterkennung vorhanden sein sollte.

### **Form und Orientierung**

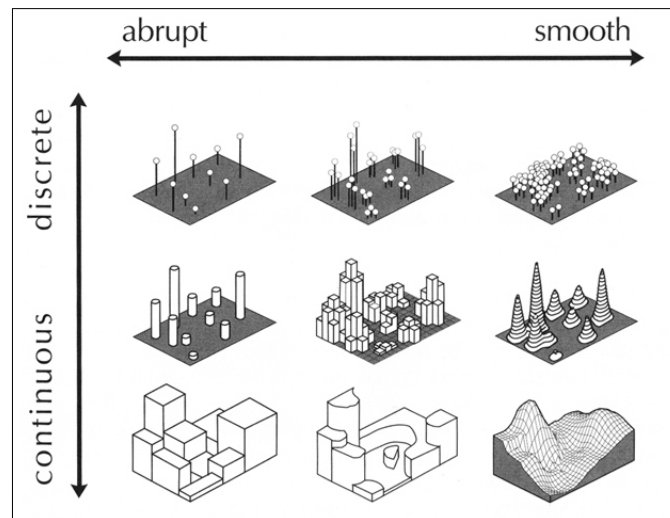
Form und Orientierung stellen qualitative Unterschiede von Symbolen dar. So können diese z. B. die Flussrichtung angeben oder auch eine Symbolkategorie repräsentieren. Zum Beispiel ist durch die Nutzung von unterschiedlichen Symbolformen eine Abgrenzung zwischen Gemeinde- und Bundeslandgrenzen möglich.

	Point	Linear	Areal
Spacing			
Size			
Hue			
Lightness			
Saturation			
Orientation			
Shape			

**Abbildung 2.3:** Die beschriebenen visuellen Variablen wie sie auf einer Karte als Punkt, Linie oder Fläche genutzt werden können [31].

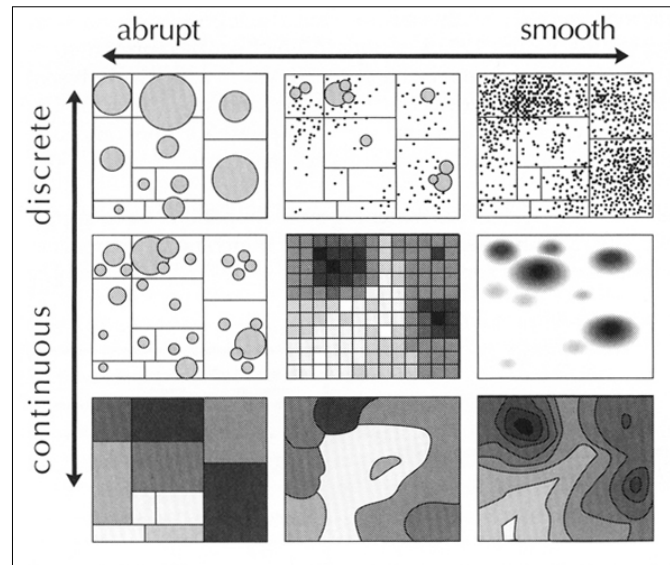
### 2.3.3 Modell von geografischen Phänomene

In der Statistik sind die Begriffe *diskret* und *kontinuierlich* verbreitet, um Datentypen zu beschreiben. MacEachren und DiBiase bezogen 1991 die beiden Begriffe auf geografische Phänomene. *Diskrete Phänomene* kommen bei eindeutig definierbaren Stellen vor.



**Abbildung 2.4:** Modell von geografischen Phänomene sowie deren Visualisierungsmöglichkeiten: Verteilung von geografischen Phänomene entlang der Achsen diskret-kontinuierlich (*discrete-continuous*) und abrupt-gleichmäßig (*abrupt-smooth*) [23]. (Teil 1/2)

Beispielsweise kann für jeden Einwohner einer Stadt zu einem bestimmten Zeitpunkt eine spezifische, räumlich voneinander getrennte, Position festgelegt werden. Währenddessen weisen *kontinuierliche Phänomene* keine bestimmte Position auf, sondern sie beziehen sich auf das Ganze (z. B. eine Region). Des Weiteren können diskrete und kontinuierliche Phänomene als *abrupt* oder *gleichmäßig* (engl. *smooth*) beschrieben werden. Während abrupte Phänomene sich umschlagend ändern, passiert dies bei gleichmäßigen schrittweise. Ein Beispiel für ein abruptes und kontinuierliches Phänomen ist die Darstellung von Wählerstimmen, da jede Region (z. B. ein Wahlkreis) einen bestimmten Wert (z. B. Anzahl Stimmen für eine Partei) hat und sich dieser abrupt an einer Regionsgrenze ändert. In Gegensatz dazu ist der durchschnittliche Niederschlag in einer Region gleichmäßig und kontinuierlich [23, 31]. Abbildungen 2.4 und 2.5 verdeutlichen die genannten Begriffe und deren Kombinationen anhand von Beispielen. *Teil 1* beschreibt die Verteilung der Phänomene entlang der Achsen diskret-kontinuierlich und abrupt-gleichmäßig. *Teil 2* zeigt dazugehörige symbolische Methoden und Visualisierungsmöglichkeiten.



**Abbildung 2.5:** Modell von geografischen Phänomene sowie deren Visualisierungsmöglichkeiten: Zeigt dazugehörige symbolische Methoden und Visualisierungsmöglichkeiten. Die vier Eckkarten wurden mit der üblichen Bezeichnung beschriftet [23]. (Teil 2/2)

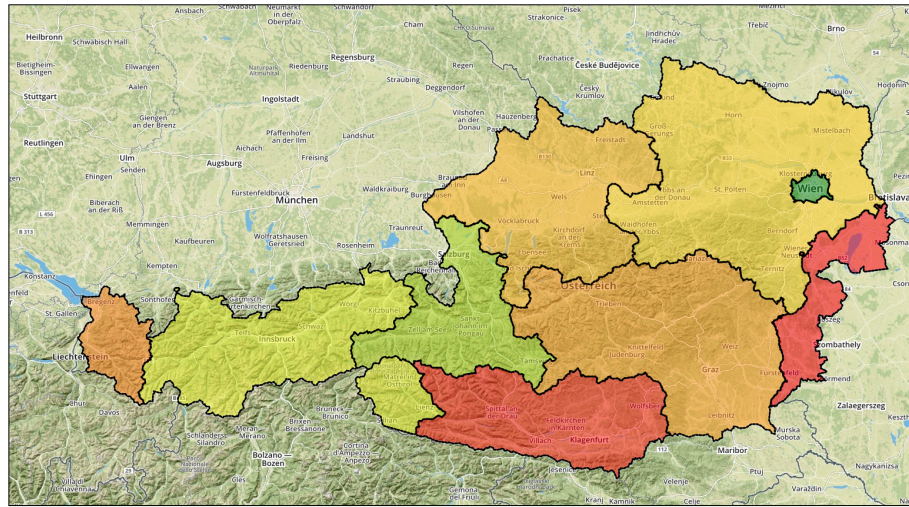
### 2.3.4 Choroplethenkarte

Eine Choroplethenkarte, auch Flächenkartogramm, ist eine thematische Karte, welche gesammelte und typischerweise gruppierte Daten (z. B. nach Regionen) als Fläche durch Farben, Schattierungen oder Schraffierungen abhängig zum darzustellenden Wert abbildet. Das Ziel einer Choroplethenkarte ist, eine fundierte und gültige Wirkung der darzustellenden Daten zu vermitteln.

Nach Slocum [31] funktioniert eine Choroplethenkarte optimal, wenn sich die Werte abrupt an Gruppengrenzen ändern (vgl. Abschnitt 2.3.3). Genauso gut kann jedoch eine einzelne Gruppe hervorgehoben werden, sofern die anderen nahezu identische Werte aufweisen.

Aufgrund der bedingten Gruppierung von Daten kann keine Verteilung innerhalb einer Gruppe dargestellt werden. Je nach Art und Weise der Gruppierung (z.B. durch Mittelwertbildung, Median oder Summierung) kann das Ergebnis bzw. der Wert unterschiedlich sein. Abbildung 2.6 zeigt beispielhaft eine Choroplethenkarte. In diesem Beispiel werden die einzelnen Flächen (hier: Bundesländer von Österreich) anhand eines akkumulierten Wertes gefärbt.

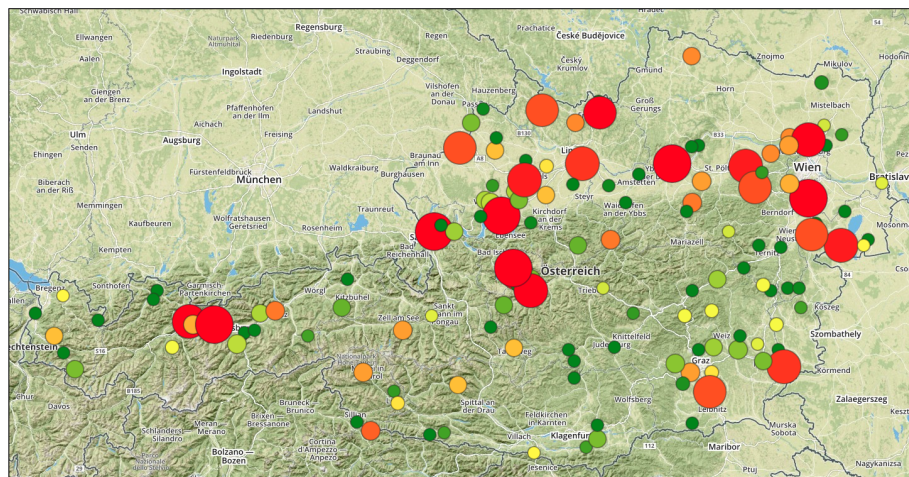




**Abbildung 2.6:** Choroplethenkarte: Standardisierte Werte werden als gefärbte Flächen dargestellt. Kleine Werte werden grün dargestellt, mittlere gelb und hohe rot. Zu sehen ist der Prototyp aus Kapitel 4.

### 2.3.5 Proportional Symbol

Bei einer Karte mit *Proportional Symbols* (zu deutsch etwa Verhältnissymbole) werden die Symbole in der Größe proportional zu ihrem repräsentierenden Wert dargestellt. Jedes Symbol kann entweder einen tatsächlichen oder konzeptionellen standardisierten Punkt symbolisieren.



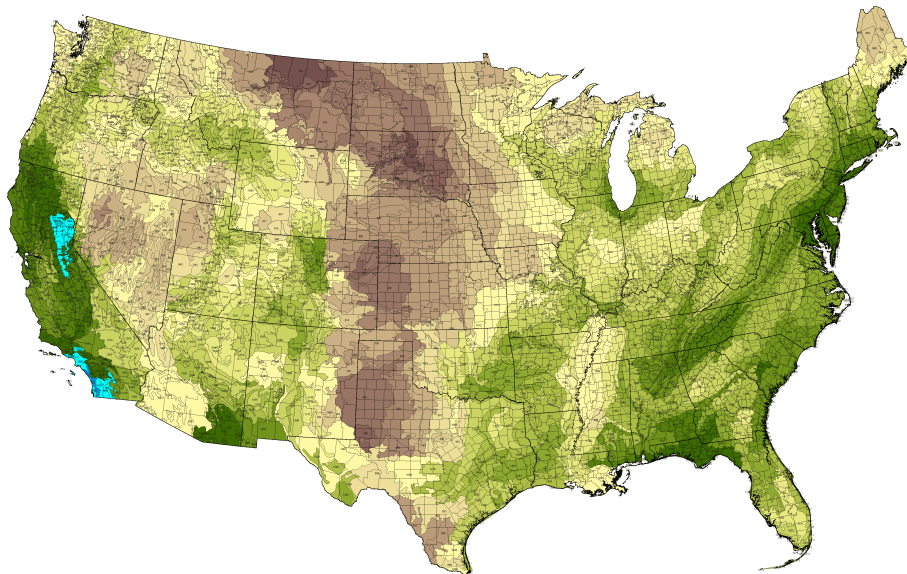
**Abbildung 2.7:** Proportional Symbol: Jedes Phänomen wird als skaliertes und gefärbtes Punkt dargestellt. Kleine Werte werden grün, mittlere gelb und große Werte rot dargestellt. Zu sehen ist der Prototyp aus Kapitel 4.

Während tatsächliche Symbole genau an einen vorkommenden Punkt referenzieren, können konzeptionelle auch auf einen Mittelpunkt von einer Gruppe von Werten verweisen. Visuell werden solche Symbole oftmals mit Kreisen dargestellt. Je höher der abzubildende Wert ist, desto größer ist auch der Kreis. Im Gegensatz zu einer Choroplethenkarte können Verhältnissymbole den genauen Ursprung des Wertes anzeigen.

Abbildung 2.7 zeigt eine Verhältnissymbolkarte, wobei die Symbole nach ihrem repräsentierenden Wert skaliert sowie gefärbt (von klein = grün, über mittel = gelb, bis groß = rot) sind.

### 2.3.6 Isoplethenkarte

Eine Isoplethenkarte, auch Konturkarte genannt, besteht aus einer Menge von interpolierten Isolinien, welche aus Stichprobenpunkten von bekannten Werten generiert werden. Dies erfordert, wie auch bei einer Choroplethenkarte, standardisierte Werte durch Gruppenbildung. Abbildung 2.8 zeigt beispielhaft wie so eine Karte aussehen kann.



**Abbildung 2.8:** Isoplethenkarte: Beispiel von einer Schätzung über die Zahl der einheimischen Arten pro 10.000 km<sup>2</sup> [30].

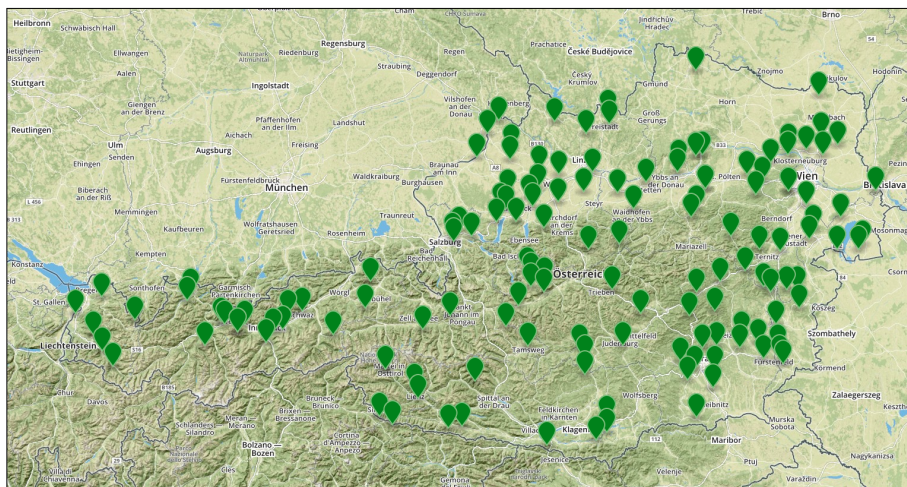
### 2.3.7 Punktstreukarte

Im Gegensatz zur Verhältnissymbolkarte wird bei einer Punktstreukarte (engl. *Dot-Map*) für jeden vorhandenen Datensatz ein Symbol dargestellt. Sollten hierbei Daten interpoliert werden, so wird ein Mittelpunkt gebildet oder es wird an der Stelle dargestellt, wo der Wert am ehesten lokalisiert



ist. Generell zeigt eine Punktstreukarte ein sehr genaues Abbild davon wo die vorhandenen Daten stammen. Jedoch erfordert dies auch Zusatzinformationen, an welcher genauen Stelle diese auftreten. Manchmal ist dies nicht eindeutig identifizierbar (Wo ist der Mittelpunkt einer Stadt?). Genauso sollte beachtet werden, dass sobald zu viele Werte dargestellt werden, sich dies auf die Übersichtlichkeit auswirken kann.

Abbildung 2.9 demonstriert, wie solch eine Karte aussehen kann. Zu beachten ist, dass als Symbol bei diesem Beispiel grüne Marker verwendet wurden anstatt, wie sonst üblich, Punkte.

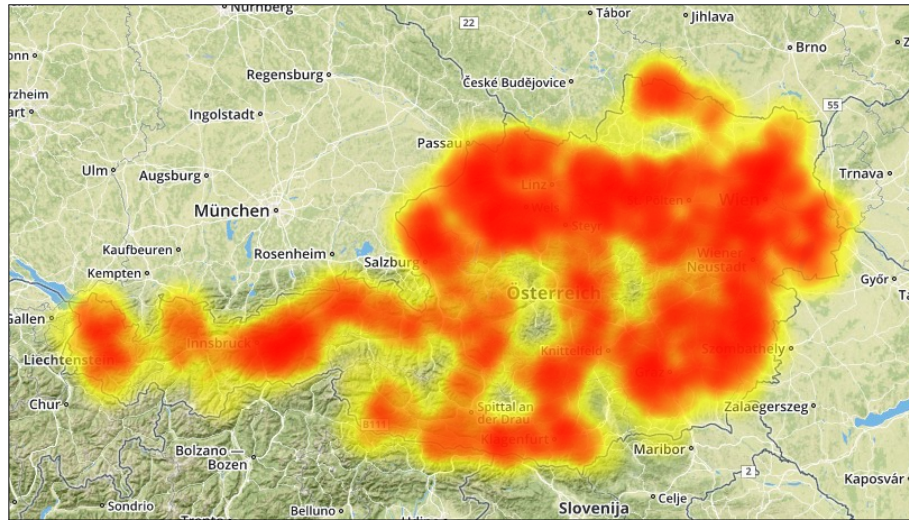


**Abbildung 2.9:** Punktstreukarte: Für jedes Phänomen wird ein Symbol dargestellt, wobei dieses Beispiel grüne *Marker* nutzt. Zu sehen ist der Prototyp aus Kapitel 4.

### 2.3.8 Heatmap

Heatmaps sind zweidimensionale grafische Visualisierungen von Daten, wobei dabei die Werte farblich dargestellt werden. Die übliche Farbskala bezieht sich oftmals auf Temperaturen (z. B. von grün = kalt bis rot = heiß), weshalb Heatmaps generell mit wenig Aufwand zu verstehen sind. Auch die dargestellten Werte sind aufgrund ihrer Farbe durchaus erkennbar. Wie bei einer Punktstreukarte oder Verhältnissymbolkarte können Werte direkt bei ihrem tatsächlichen Vorkommen angezeigt werden. Typischerweise werden Heatmaps für Ergebnisse aus Benutzerstudien, z. B. Augenbewegung oder Klick-Verhalten, genutzt [3].

Abbildung 2.10 zeigt eine Heatmapkarte. Basierend auf vorhandenen Daten werden Flächen erzeugt. Je nach dem Datenwert variiert die Farbe der Fläche.



**Abbildung 2.10:** Heatmap: Basierend auf den vorhandenen Daten wird eine Heatmap erzeugt. Rote Flächen bedeuten, dass Datensätze an dieser Stelle vorhanden sind. Zu sehen ist der Prototyp aus Kapitel 4.

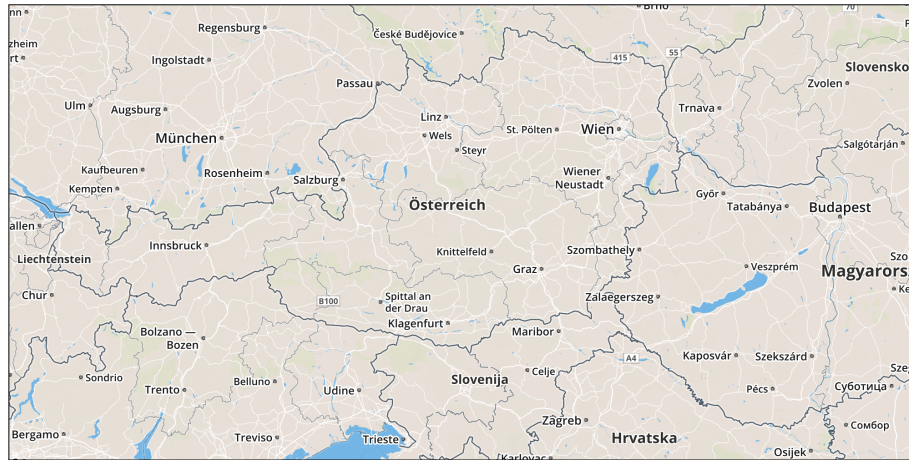
## 2.4 Statische vs. dynamische Karte im WWW

Statische Karten haben einen zuvor definierten Kartenausschnitt, inklusive sämtlicher dargestellten Informationen, der vom Benutzer selbst nicht änderbar ist. Technisch gesehen entspricht das einem Bild, z. B. ein Abbild aus einem Atlas oder einer Kopie von einer dynamisch erzeugten Karte. Jedoch ist es nicht möglich zusätzliche Informationen einzublenden.

Dynamische Karten hingegen können von einem Benutzer individuell angepasst werden. So ist es möglich, den aktuellen Kartenausschnitt zu verändern, Details zu vergrößern oder weitere Dateninformationen, wie die aktuelle Verkehrslage, zur Karte hinzuzufügen. Dynamische Karten haben sich mittlerweile weitreichend durchgesetzt. Dies geschah einerseits dadurch, dass Computer mit immer größeren Datenbeständen in geringerer Zeit operieren konnten. Andererseits ist es oft nicht möglich zu wissen, was ein Nutzer genau herausfinden möchte.

Dennoch ist es möglich, dass sowohl statische als auch dynamische Karten interaktiv sein können. Interaktiv bedeutet, dass z. B. über Weblinks auf eine weitere Quelle (z. B. Karte, Website oder Grafik) referenziert wird. Solche Karten werden gemeinhin auch als *anklickbare Karten* bezeichnet [18, 33].

Beispielhaft zeigt Abbildung 2.11 eine statische Karte. Diese wurde mit dem Framework von Mapbox erstellt. Über die Adresszeile kann u. a. das Ausgabeformat, die angezeigte Position sowie die Vergrößerungsstufe angegeben werden.



**Abbildung 2.11:** Beispiel einer statischen Karte. Die Karte wurde mit dem Framework *Mapbox Static Maps*<sup>5</sup> erstellt und zeigt einen Kartenausschnitt von Österreich. Aufgrund dessen, dass die Karte statisch erzeugt wurde, kann weder der Kartenausschnitt, noch die Vergrößerungsstufe verändert werden.

## 2.5 Digitale Geodaten und -formate

Geodaten (engl. *geospatial data*), auch raumbezogene Daten oder Geoinformationen, sind Daten für Geoinformationssysteme mit geografischen Informationen, wie Objektbeschreibungen, Positionen (bezogen auf die Realwelt) und Beziehungen zueinander. Merkmale, z. B. Häuser, Straßen, Felder oder Berge, die eine eindeutige Position in der Realwelt haben, können mithilfe von einem geografischen Dateiformat (*GIS file format*) repräsentiert werden [9, 19, 22].

Üblicherweise werden derartige Dateiformate von einer staatlichen Behörde, gemeinnützigen Organisation oder einem Softwarehersteller definiert und veröffentlicht [2]. Hierbei ist vor allem das Open Geospatial Consortium<sup>6</sup> (OGC) zu nennen, das das Ziel hat, die Entwicklung von komplexen raumbezogenen Informationen auf Basis allgemeingültiger Standards zum Zweck der (Daten-) Interoperabilität<sup>7</sup> festzulegen und dadurch für alle Arten von Anwendungen zugänglich zu machen [6]. So wurde die verbreitete *Geography Markup Language*, *GML*<sup>8</sup> von OGC spezifiziert.

GML ist eine auf XML basierende Auszeichnungssprache, welche zum Austausch von raumbezogenen Merkmalen (*Features*) dient [7]. Weitere Stan-

<sup>6</sup><http://www.opengeospatial.org>

<sup>7</sup>Bartelme [2] beschreibt Interoperabilität als *die Fähigkeit zur Kommunikation, zur Ausführung von Programmen und zum Austausch von Daten zwischen verschiedenen funktionalen Einheiten in einer Art und Weise, die von Anwendern wenige oder gar keine Kenntnisse über die Besonderheiten dieser Einheit erfordert*.

<sup>8</sup><https://www.iso.org/obp/ui/#iso:std:32554:en>

dards befinden sich auf der Website des Konsortiums. Abschnitt 5.2 beschreibt eine weitere Sprache, welche basierend auf der JavaScript-Notation Daten strukturiert speichert.

Programmcode 2.1 zeigt beispielhaft, wie eine Koordinate im GML-Format dargestellt werden kann. Ein weiteres Beispiel für die Darstellung im GeoJSON-Format befindet sich in Abschnitt 5.2.

```
1 <gml:Point gml:id="p21"  
2   srsName="http://www.opengis.net/def/crs/EPSG/0/4326">  
3   <gml:coordinates>45.67, 88.56</gml:coordinates>  
4 </gml:Point>
```

**Programm 2.1:** Beispielcode für die Repräsentation von Koordinaten im GML-Format.

## Kapitel 3

# Erweitertes Hintergrundwissen

In diesem Kapitel werden weitere Begrifflichkeiten definiert, welche Voraussetzung für diese Masterthesis sind. Abschnitt 3.1 beschreibt das Schlagwort *Electronic Commerce*. Die Begriffe *Big Data*, *Data-Mining* sowie *Visual Analytics* werden in Abschnitt 3.2 bzw. 3.3 erläutert. Die Begriffe Geoinformationssysteme sowie Geokodierung werden in den Abschnitten 3.4 bzw. 3.5 spezifiziert. Zum Schluss wird in Abschnitt 3.6 die Richtlinie der Europäischen Union zur *Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft*, kurz INSPIRE, beschrieben.

### 3.1 Electronic Commerce

In diesem Abschnitt wird der Begriff *Electronic Commerce*, auch E-Commerce genannt, definiert. Die genaue Definition von E-Commerce verändert bzw. erweitert sich aufgrund des ständigen Wandels im Bereich von elektronischen Wirtschaftssystemen und der Kommunikation häufig. Grundsätzlich werden unter dem Begriff elektronische sowie meist geschäftliche Interaktionen verstanden, die von zwei Akteuren durchgeführt werden. Akteure können z. B. Unternehmen (Business), Personen (Consumer) oder Behörden (Government) sein. Auch nicht kommerzielle Interaktionen werden mittlerweile in diesem Bereich eingegliedert. So werden u. a. auch Aktivitäten im Umfeld der Unterhaltung, Kommunikation, Forschung oder Weiterbildung als Teil von E-Commerce angesehen, vorausgesetzt die Interaktion wird über das Internet durchgeführt. Die Auslieferung des Endproduktes kann auch außerhalb des Internets durchgeführt werden [36].

Sämtliche Bereiche der elektronischen Handlung sind nicht mehr eine eigenständige Anwendung, sondern vielmehr ein Aspekt des gesamten (elektronischen) Geschäftsprozesses. Gesammelte Informationen aus Web-Shops werden u. a. dazu verwendet, kundengerecht Produkte anzupassen, den zu-

künftigen Bedarf zu ermitteln sowie Business-Strategien zu entwickeln. Heutzutage suchen Verbraucher speziell nach Produkt- oder Unternehmensinformationen online, sodass ein alternativer Marketingkanal entstanden ist [25].

## 3.2 Big Data

*Big Data* bezeichnet extrem große und komplexe Datenmengen, die nicht mehr mit konventionellen Werkzeugen der Datenverarbeitung in effizienter Zeit verarbeitet bzw. ausgewertet werden können. Die primären Herausforderungen beruhen dabei meist auf der Akquisition, Speicherung, Suche, Zugriff, Analyse sowie Darstellung von Daten [27].

Inzwischen ist der Begriff *Big Data* ein Synonym für weitere Konzepte, wie z. B. Business Intelligence, Analyse oder Data Mining. Heutzutage werden möglichst viele Daten erfasst, um darauf basierend neue Erkenntnisse durch Analysen zu gewinnen. Daten können aus unterschiedlichen Quellen stammen, z. B. durch Sensoren über das aktuelle Wetter, on- und offline Bestellungen, digitale Fotos und Videos oder GPS-Daten von einem Smartphone [1, 8].

Das Schlagwort Big Data ist auch negativ belastet. Aufgrund der stetig besseren und schnelleren Algorithmen zur Datenanalyse fühlen sich viele Personen überwacht, transparent und machtlos gegenüber Unternehmen oder dem Staat. Mittlerweile wird fast alles aufgezeichnet und ausgewertet, was technisch möglich ist [32]. So kann anhand der Interessen der jeweiligen Person Werbung gezielter angezeigt werden. Auch Bewegungsprofile können durch das Mitschneiden der aktuellen Position erstellt werden. Besonders ist hierbei der Begriff *Vorratsdatenspeicherung* zu nennen, der für viele Personen einer *Totalüberwachung* der Menschen gleich kommt [29].

Die Europäische Union hat in Richtlinie 2006/24/EG über die Vorratsdatenspeicherung erlassen, dass Daten zum Zwecke der *Ermittlung, Feststellung und Verfolgung von schweren Straftaten* erhoben werden dürfen, jedoch die Privatsphäre von natürlichen Personen geschützt sein muss [34].

## 3.3 Data-Mining und Visual Analytics

**Data-Mining:** *Data-Mining* ist sinngemäß der Prozess des Suchens von (versteckten) Mustern in einer großen Datenmenge. Die Suche wird dabei (semi-) automatisiert über computergestützte Systeme durchgeführt. Das Hauptziel von Data Mining ist durch das Analysieren von Daten, Probleme zu lösen, Vorhersagen zu berechnen und neues Wissen zu generieren [37].

Der Begriff Data-Mining wird oftmals synonym für die Wissensentdeckung in Datenbanken (engl. *Knowledge Discovery in Databases, KDD*) verwendet, wenngleich Data-Mining lediglich ein Teilbereich (Analyse von Daten) von KDD ist. Verfahren aus Data-Mining werden u. a. im Bereich der



Statistik, Datenbanken oder im maschinellen Lernen angewendet [11].

**Visual Analytics:** Heutzutage werden Daten in immer größerer Kapazität, jedoch in immer kürzerer Zeit gesammelt. Doch genau diese große Flut an Daten in einer adäquaten Zeit zu analysieren wurde inzwischen für Datenanalysten und Decision-Makers zu einer neuen Herausforderung. *Visual Analytics* (zu deutsch etwa *visuelle Analyse*) ist ein Ansatz, um einerseits die Stärken der automatischen Datenanalyse (Data-Mining), andererseits mit den Fähigkeiten des Menschen so zu verknüpfen, sodass schnell Muster visuell erkannt werden können. In der Datenanalyse werden durch geeignete Interaktionsmechanismen Daten analysiert und exploriert, sodass neues Wissen gewonnen werden kann. Visual Analytics ist eine Kombination aus verwandten Forschungsgebieten einschließlich der Visualisierung, Data-Mining sowie der Statistik [16].

### 3.4 Geoinformationssysteme

Ein Geoinformationssystem (GIS) ist ein computergestütztes System, welches geografische und räumliche Daten erfasst, speichert, manipuliert, analysiert, verwaltet und visualisiert [17, 22]. Es umfasst sämtliche benötigte Anwendungen, Datenbanken, Prozesse, Soft- und Hardware. GIS gilt als eine der wichtigsten neuen Technologien mit dem Potenzial viele Aspekte der Gesellschaft, durch eine bessere Möglichkeit Entscheidungen zu treffen (engl. *Decision making*) und Probleme zu lösen (engl. *Problem solving*), zu revolutionieren.

In der heutigen Zeit sind viele Herausforderungen (z.B. Überbevölkerung, Umweltschutz, Naturkatastrophen) zu bewältigen. Ein GIS kann diese Probleme sichtbar machen und Einblicke in die teils komplexe Ursache geben sowie dem Entscheidungsträger bei der Analyse nach einer Lösung unterstützen. Aber auch Marketinganalysten können mittels eines GIS Kundenanalysen durchführen, um neue Kunden gewinnen zu können.

Das Akronym GIS wird teilweise auch für die Wissenschaft *Geographical Information Science* (*GIScience*) verwendet.

### 3.5 Geokodierung

Geokodierung (engl. *Geocoding*) ist der Prozess, um übereinstimmende Roh-Addressinformationen mit einem digitalen räumlichen Datensatz abzugleichen. Als Ergebnis werden alle Koordinaten, die von Interesse sein könnten, retourniert. Eine Koordinate besteht dabei aus dem zugeordneten Breiten- und Längengrad. Vorbedingung der Geokodierung ist das Akquirieren von (Echt-) Daten aus Primär- und Sekundärquellen. Je umfangreicher und detaillierter der bestehende Datensatz ist, desto erfolgreicher und korrekter ist

das Ergebnis der Geokodierung. Aufgrund stetiger Änderungen der Daten (z. B. durch Erschließung neuer Straßen) ist eine regelmäßige Aktualisierung des Datenbestandes unabdingbar. Auch ist es möglich, dass die genutzten Roh-Daten fehlerbehaftet sind, was sich auf eine schlechtere bzw. ungenauere Trefferquote auswirkt [22].

### 3.6 INSPIRE

INSPIRE (Infrastructure for Spatial Information in the European Community) ist eine Richtlinie (2007/2/EG vom 14. März 2007) des europäischen Parlaments und Rates zur *Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft* [35]. Ziel ist u. a., die Probleme der Verfügbarkeit, Qualität, Organisation, Zugänglichkeit und das gemeinsame Nutzen von Geodaten zu verbessern. Ein primärer Aspekt der Richtlinie ist die Schaffung einer Infrastruktur, die den Prozess zur Entscheidungsfindung auf Umweltfragen in Bezug auf politische Konzepte und Maßnahmen unterstützt.

Zur Erreichung dieser Ziele sollen Mitgliedsstaaten der Europäischen Union ihre Geodateninfrastruktur (GDI) so auslegen, dass Geodaten auf der *optimal geeigneten Ebene gespeichert, zugänglich gemacht und verwaltet werden können*. Des Weiteren ist definiert, dass die (Geo-) Daten für verschiedene Nutzer und Anwendungen öffentlich bereitgestellt werden.

Unter Punkt 17 heißt es weiter, dass *Netzdienste erforderlich [sind], um Geodaten auf den verschiedenen Verwaltungsebenen in der Gemeinschaft gemeinsam nutzen zu können. Über diese Netzdienste sollte es möglich sein, Geodaten zu ermitteln, umzuwandeln, abzurufen und herunterzuladen und Geodatendienste sowie Dienste des elektronischen Geschäftsverkehrs in Anspruch zu nehmen*.

Österreich hat dafür u. a. die Informationswebsite *INSPIRE Austria*<sup>1</sup> sowie u. a. das Geoportal des Bundesamtes für Eich- und Messwesen (BEV)<sup>2</sup> veröffentlicht, um die Richtlinie umzusetzen. Die europäische Kommission hat außerdem die Plattform *INSPIRE Geoportal*<sup>3</sup> publiziert, worüber auch Geodaten zur Verfügungen gestellt werden.

---

<sup>1</sup><http://www.inspire.gv.at/>

<sup>2</sup><http://www.austrianmap.at/>

<sup>3</sup><http://inspire-geoportal.ec.europa.eu/>

## Kapitel 4

# Eigener Entwurf

Das folgende Kapitel liefert in Abschnitt 4.1 einen Einblick in die Anforderungen und Voraussetzungen des entwickelten Prototypen. Außerdem wird in Abschnitt 4.2 die Designkonzeption des Entwurfs vorgestellt.

### 4.1 Anforderungen und Voraussetzungen

Ziel des Projektes ist es, raumbezogene Daten aus einem E-Commerce-System auf einer interaktiven Karte darzustellen und durch visuelle Datenanalyse neue Erkenntnisse über das Verhalten der Benutzer zu eruieren. Folgend werden wesentliche Anforderungen und Voraussetzungen an den Prototypen spezifiziert. Unterschieden wird im Prototyp zwischen Primär- und Sekundärdaten. Während Primärdaten Informationen zu dem System selbst (z. B. Besucher und Bestellungen) enthalten, stammen Sekundärdaten aus externen Quellen. Diese erweitern das vorhandene Wissen um allgemeine und meist collagierte Informationen, die z. B. von einer Behörde oder Organisation veröffentlicht wurden.

#### 4.1.1 E-Commerce-System

In der Regel ist der Aufbau eines E-Commerce-Systems komplex. Das heißt, dass u. a. die Datenbankstruktur verschachtelt und teilweise wenig ersichtlich ist. Beispielsweise hat der weitverbreitete Webshop *Magento – Community Edition*<sup>1</sup> in der Standardinstallation über 300 Entitäten. Alleine aus diesem Grund ist es schwierig die gesammelten Daten aus dem Shop zu extrahieren, konvertieren und darzustellen. Des Weiteren wäre solch ein Prototyp auch ausschließlich auf ein System zugeschnitten, was nicht Ziel des Projektes ist. Der Prototyp soll Daten aus einem unspezifischen System visualisieren, weshalb ein fiktives erforderlich ist.

---

<sup>1</sup><http://magento.com/products/overview#community>

### 4.1.2 Primärdaten

Die Datenbasis des Projektes besteht aus einer Menge von vordefinierten Entitäten sowie deren Eigenschaften. In diesem Abschnitt sind die drei Primärentitäten (*Besucher*, *Bestellung* und *Kunde*) beschrieben. Dabei gibt die Domäne an, welcher Datenwert für ein Attribut festgelegt wurde. In den Tabellen symbolisiert die letzte Spalte jeweils einen möglichen Beispielwert.

#### Besucher

Eine der wichtigsten Datenquellen zu einem E-Commerce-System sind Informationen über Besucher (engl. *visitor*). Mittlerweile ist die Reichweite an verfügbaren Fakten zu einem Besucher groß. So kann nicht nur der aktuelle Standort des Besuchers eruiert werden, auch weitere Eigenschaften wie das Geschlecht, Alter und sogar die Interessen des Besuchers sind ermittelbar. Die Daten stammen dabei oftmals aus Angaben von anderen Plattformen, jedoch durch das Protokollieren und Mitschneiden der Benutzereingaben und des -verhaltens können diese aggregiert und collagiert werden.

Die Entität *Besucher* umfasst diese grundlegenden Informationen, sodass diese vom Prototypen verwendet werden können. Tabelle 4.1 beschreibt die Mindestanforderungen.

<i>Attribut</i>	<i>Domäne</i>	<i>Beispielwert</i>
Position des Besuchers	Koordinate	N 16.37, O 48.209
Besuchtes Produkt	Zeichenkette	Produkt A
Aktuelles Alter des Besuchers	Ganzzahl	55
Zeitpunkt des Aufrufs	Datum + Zeit	2012-12-06 23:34:16
Geschlecht des Besuchers	[männlich, weiblich]	männlich
Dauer des Aufenthalts auf der Seite	Sekunden	15
Genutztes Endgerät des Besuchers	[Desktop-PC, Mobil]	Mobil
Insgesamt aufgerufene Seiten	Ganzzahl	3

**Tabelle 4.1:** Mindestanforderungen an die Entität *Besucher*.

### Bestellung

Sobald Produkte bestellt werden, wird auch Umsatz generiert. Daher sind Bestellungen essentiell für den Erfolg eines Onlineshops. Von Bedeutung sind hier die Absatzzahlen einzelner Produkte. Auch der Ort der Abwicklung kann entscheidend sein. Einige Produkte verkaufen sich in bestimmten Regionen besser als in anderen.

Zum Beispiel wird Streusalz häufiger in Bergregionen benötigt. Aufgrund dessen, dass eine Bestellung von einem Besucher bzw. bestehenden Kunden ausgeführt wird (eine grundsätzliche Annahme ist, dass keine Bestellungen über ein computergestütztes System getätigt werden können), sind einige Fakten identisch mit dem eines Besuchers.

Die Entität *Bestellung* enthält diese grundlegenden Daten. Tabelle 4.2 beschreibt die Mindestanforderungen.

<i>Attribut</i>	<i>Domäne</i>	<i>Beispielwert</i>
Gekauftes Produkt	Zeichenkette	Produkt A
Quantität des Produktes	Ganzzahl	2
Position des Käufers	Koordinate	N 16.37 O 48.209
Zeitpunkt des Kaufes	Datum + Zeit	2012-12-06 23:34:16
Gesamtpreis des gekauften Produktes	Dezimalzahl	123,45
Aktuelles Alter des Käufers	Ganzzahl	55
Geschlecht des Käufers	[männlich, weiblich]	männlich
Anzahl vergangene Tage bis zum Abschluss des Kaufes	Ganzzahl	3

**Tabelle 4.2:** Mindestanforderungen an die Entität *Bestellung*.

### Kunde

Kunden haben in der Vergangenheit mindestens ein Produkt über das E-Commerce-System bestellt. Zufriedene Kunden kaufen oftmals wiederholt über das System, weshalb Stammkunden sehr wertvoll sind. Diese Kunden sollten möglichst lange gebunden werden, gleichwohl dies durchaus viel Marketing erfordert.

Die Entität *Kunde* enthält diese grundlegenden Informationen, wie z. B. die Anzahl an getätigten Bestellungen oder den summierten Umsatz von einem Kunden. Tabelle 4.3 beschreibt die Mindestanforderungen.

<i>Attribut</i>	<i>Domäne</i>	<i>Beispielwert</i>
Vorname des Kunden	Zeichenkette	Max
Nachname des Kunden	Zeichenkette	Mustermann
Position des Kunden	Koordinate	N 16.37 O 48.209
Anzahl aller gekauften Produkte	Ganzzahl	3
Gesamtpreis aller gekauften Produkte	Dezimalzahl	123,45
Aktuelles Alter des Kunden	Ganzzahl	55
Geschlecht des Kunden	[männlich, weiblich]	männlich

**Tabelle 4.3:** Mindestanforderungen an die Entität *Kunde*.

### 4.1.3 Sekundärdaten

Anders als Primärdaten stammen Sekundärdaten aus einer externen Datenquelle. Sie bieten u. a. allgemeine Informationen zu Regionen, z. B. das durchschnittliche Einkommen oder Verkaufsverhalten.

Externe Daten werden ebenfalls in der Datenbank collagiert, sodass diese im Prototypen einfach dargestellt werden können. Derweil sind lediglich länderspezifische Informationen zum Land Österreich zu integrieren. Diese werden im folgenden Abschnitt beschrieben.

### Länderspezifische Informationen

Die Entität *Länderspezifische Informationen* enthält sekundäre Daten aus einer externen Datenquelle. Das Attribut *Koordinaten* besteht aus einem *JavaScript-Array* (bzw. genauer ein JSON-Objekt<sup>2</sup>), welches die geometrischen Informationen (z. B. Eckpunkte) des umfassenden Gebietes beschreibt. Abschnitt 5.2 und 5.4.2 diskutieren die Implementierung bzw. Verwendung dieses Objektes.

Zur Vereinfachung wird sogleich das Zielformat (Koordinaten im GeoJSON Format) genutzt, sodass die Performanz der Datenverarbeitung durch die Minimierung an Datentransformationsschritten erhöht wird. JSON, bzw. für die Repräsentation von räumlichen Informationen GeoJSON, ist dank der schlanken Datengröße zum Maßstab beim Datenaustausch über das WWW geworden, weshalb auch dieses Projekt die Technologie einsetzt.

Die Attribute *Einwohner*, *Einkommen*, *Haushaltsgröße* sowie *Verkaufsverhalten* können unterschiedliche Ausprägungen haben. *Einwohner* und

<sup>2</sup><http://json.org>

*Einkommen* sind jeweils vom Geschlecht abhängig, wobei auch die Summe beider Ausprägungen vorhanden sein muss. *Einwohner* hat zusätzlich die Information über die genaue Anzahl der Personen pro Dekade. Die *Haushaltsgröße* hat die Ausprägungen *maximal eine, fünf und mehr* und *durchschnittliche Anzahl* im Haushalt lebender Personen sowie *Haushalt ohne Kinder* und *durchschnittliche Anzahl an Kindern im Haushalt*. Tabelle 4.4 beschreibt die Mindestanforderungen.

<i>Attribut</i>	<i>Domäne</i>	<i>Beispielwert</i>
Name des Gebietes	Zeichenkette	Oberösterreich
ISO-Code des Gebietes	Zeichenkette	AT-1
Koordinaten des Gebietes	JSON	—
Geometrischer Type	[Polygon, MultiPolygon]	Polygon
Einwohner	Ganzzahl	1.625.485
Einkommen	Ganzzahl	27.189
Haushaltseinkommen	Ganzzahl	55.456
Durchschnittsalter	Dezimalzahl	41,12
Arbeitslosenquote	Dezimalzahl	7,12
Arbeiterquote	Dezimalzahl	75,12
Haushaltsgröße	Dezimalzahl	3,5
Verkaufsverhalten	Dezimalzahl	10,59

**Tabelle 4.4:** Mindestanforderungen an die Entität *Länderspezifische Informationen*.

#### 4.1.4 Interaktivität und Anpassungsmöglichkeiten

Der Prototyp ist so zu realisieren, dass dieser durchweg interaktiv genutzt werden kann. Das bedeutet, dass u. a. der Report der Primär- und Sekundärdatenquelle angegeben werden kann. Ferner soll es möglich sein, Filteroptionen auf den einzelnen Attributen der Daten anzuwenden bzw. die Bildung von Regions-Clustern zu ermöglichen.

Auch die Visualisierungsart der Primärdaten kann vom Benutzer gewählt werden. Des Weiteren ist die Karte interaktiv zu gestalten, d. h. die Karte wird dynamisch geladen und kann vom Benutzer u. a. via eines Zooms vergrößert werden. Ebenfalls kann der dargestellte Kartenausschnitt verändert werden (vgl. Abschnitt 2.4).

#### 4.1.5 Visualisierungstypen

Es sollen verschiedene Visualisierungstypen zur Verfügung stehen. Verpflichtend ist dabei die Darstellung einzelner Datensätze als Symbol (Marker und Kreis), Heatmap und Choroplethenkarten.

Bei Markern und Kreisen soll es möglich sein, diese aufgrund eines Wertes zu färben. Marker sollen zusätzlich gruppiert werden können. Die Heatmap stellt lediglich das Vorkommen der Datensätze dar. Choroplethenkarten collagieren die gegebenen Datensätze einer Region (Bundesland) und färben die umspannende als Fläche je nach gruppierten Wert ein. Die funktionelle Art der Gruppierung kann definiert werden, mindestens erforderlich ist jedoch die Gruppenfunktion *Summierung*, *Durchschnitt*, *Anzahl* und *Mittelwertbildung*.

#### 4.1.6 Echtsystem

In Abschnitt 4.1.1 wurde beschrieben, dass ein fiktives E-Commerce-System erschaffen werden soll. Jedoch ist zu beachten, dass es ohne großen Aufwand möglich sein soll, den Prototypen mit einem Echtsystem zu verknüpfen. Dies betrifft insbesondere das Konzept des Datenschemas.

Zum Beispiel kann der Prototyp so implementiert werden, dass Echtdaten lediglich in das System konvertiert werden müssen. Zwei Möglichkeiten zur Integration eines Echtsystems werden im Abschnitt 5.5 erörtert.

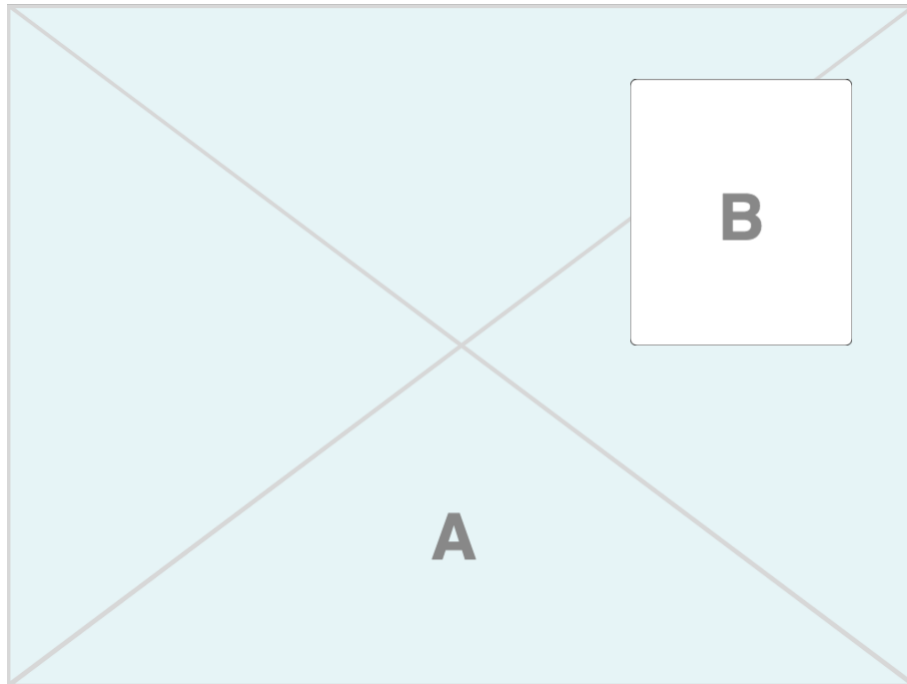
### 4.2 Benutzeroberfläche

Die Benutzeroberfläche besteht aus zwei Grundelementen: einer *Karte* und einem *Optionenbereich*. Die Karte wird über den ganzen Bildschirm dargestellt, um vom größtmöglichen Platz des Bildschirms zu profitieren. Der Optionenbereich ist rechts oben lokalisiert, sodass dieser für den Benutzer leicht erreichbar ist, trotz alledem keinen allzu großen Platz bedarf.

Generell ist das Konzept ausschließlich für einen Desktop-PC optimiert. Eine mobile Nutzung (Smartphone oder Tablet) ist derzeit nicht vorgesehen, da die Karten meist detailreich und somit die zu erwartende transferierte Datenmenge hoch sein wird.

Abbildung 4.1 zeigt ein Wireframe-Modell mit den zwei beschriebenen Grundelementen.



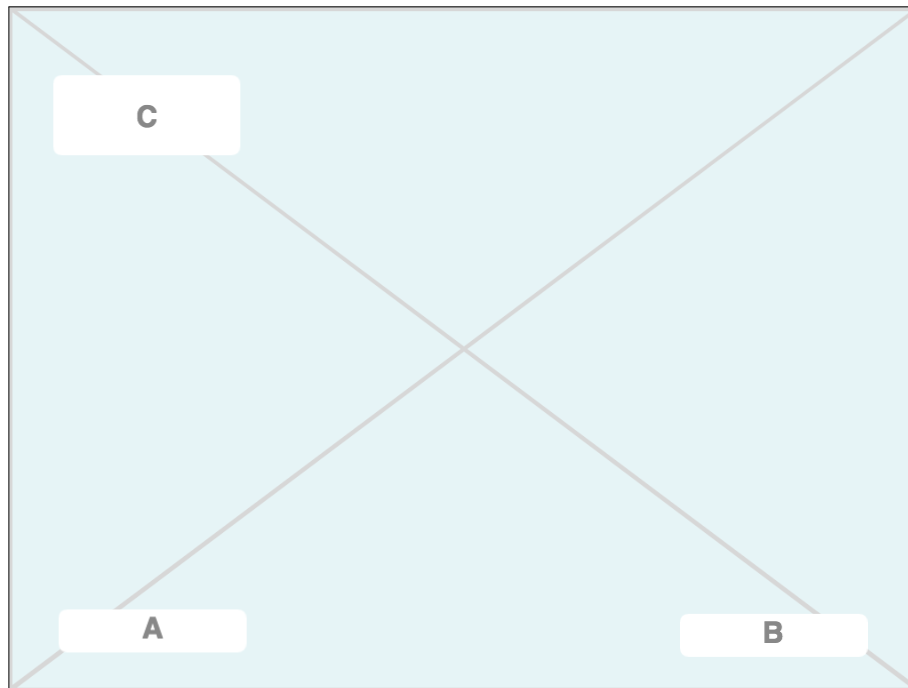


**Abbildung 4.1:** Konzept über die Grundelemente von der Benutzeroberfläche als Wireframe-Modell. Die Karte (**A**) füllt ganzflächig den Bildschirm, sodass der größtmögliche Platz ausgenutzt wird. Der Optionenbereich (**B**) ist oben rechts platziert.

#### 4.2.1 Karte

Wie im vorherigen Abschnitt beschrieben, wird die Karte über den ganzen Bildschirm angezeigt. Wie in Abschnitt 4.1.4 definiert, kann die Karte via Eingaben modifiziert werden. Hierzu zählen u. a. das Zoomen und das Verändern des Kartenausschnitts. Zusätzlich können drei weitere Bereiche genutzt werden.

Abbildung 4.2 spezifiziert die Platzierung dieser Bereiche. *A* beinhaltet die Legende aus der Primärquelle. *B* umfasst die Sekundärquelle. Der Bereich *C* ist für weitere Informationen vorgesehen, z. B. können Detailinformationen über ein ausgewähltes Datenobjekt angezeigt werden.



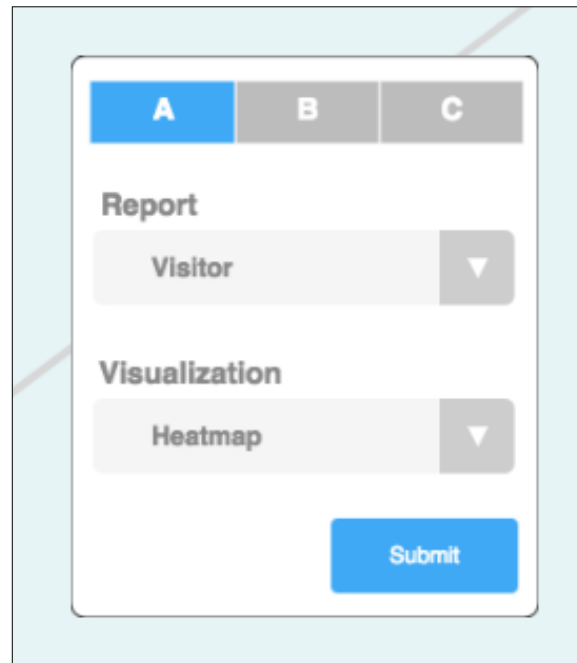
**Abbildung 4.2:** Konzept der Legendenplatzierung als Wireframe-Modell. **A** umfasst die Legende für die Primärquelle, **B** für die Sekundärquelle. Der Bereich **C** ist für zusätzliche Informationen über einzelne selektierte Datenobjekte vorgesehen.

### 4.2.2 Optionenbereich

Im Optionenbereich werden sämtliche Visualisierungseinstellungen vorgenommen.

Im Tab *A* kann unter *Report* der Bericht aus der Primärquelle ausgewählt werden. Unter *Visualization* befinden sich die Visualisierungsmöglichkeiten, wie z. B. die Option Heatmap oder Choroplethenkarte. Der Tab *B* enthält Filter für die Primär- und Sekundärquelle, sodass lediglich eine Selektion von Datensätzen geladen und angezeigt wird. Weitere Einstellungen für die Sekundärquelle befinden sich im Tab *C*.

Abbildung 4.3 verdeutlicht den Aufbau des Optionenbereichs.



**Abbildung 4.3:** Konzept des Optionenbereichs als Wireframe-Modell. Tab **A** zeigt eine Auswahl der Primärquelle sowie der Visualisierungsoptionen. Tab **B** beinhaltet Filter für die Daten aus der Primärquelle. Die Sekundärquelle wird in Tab **C** ausgewählt.

# Kapitel 5

## Umsetzung

Dieses Kapitel widmet sich wesentlichen Details und Besonderheiten bei der Umsetzung des Prototypen. Zu Beginn sind in Abschnitt 5.1 zentrale Frameworks und Technologien erläutert. Für die Umsetzung wurde das Datenformat *GeoJSON* verwendet, welches in Abschnitt 5.2 vorgestellt wird. Eine Übersicht über die Architektur des Prototypen gibt Abschnitt 5.3 wieder. Besonderheiten der Implementierung, inklusive einer Beschreibung der Requests und Responses der Datenübertragung, werden in Abschnitt 5.4 beschrieben. Zuletzt sind Nutzungsmöglichkeiten für die Integration von Echt-systemen erörtert. Der vollständige Programmcode, inklusive einer Installationsanleitung, befindet sich auf dem beiliegenden Datenträger.

### 5.1 Verwendete Frameworks und Technologien

Im Folgenden werden wesentliche Frameworks und Technologien vorgestellt, die in der Implementierung des Prototypens verwendet wurden und den angeführten Anforderungen aus Abschnitt 4.1 entsprechen.

**Mapbox:** *Mapbox*<sup>1</sup>, gegründet 2010, ist eine quelloffene Plattform, um digital anpassbare und skalierbare Karten in einer (Business-) Web-Applikation zu nutzen. Mittlerweile ist MapBox neben *Google Maps*<sup>2</sup> einer der *Big Player* im Bereich der *Online Maps*-Anbieter [24]. Die Plattform unterstützt diverse Technologien, wie JavaScript, iOS und Android. Das Application Programming Interface (kurz API) bietet zahlreiche Möglichkeiten, um Karten zu designen, erstellen und zu publizieren. Als Kartenbasis nutzt Mapbox das Open-Data Projekt *OpenStreetMap*<sup>3</sup>. Mapbox dient als Kartenbasis für den Prototypen.

---

<sup>1</sup><http://mapbox.com>

<sup>2</sup><http://maps.google.com>

<sup>3</sup><http://openstreetmap.org>

**Leaflet:** *Leaflet*<sup>4</sup> ist eine quelloffene, moderne und leichtgewichtige JavaScript-Bibliothek für interaktive Karten. Bei der Bibliothek steht Simplizität, Performanz und Benutzerfreundlichkeit im Vordergrund, gleichwohl alle weit verbreitenden Plattformen (Desktop und Mobil) unterstützt werden. Funktionalität ist über Plugins erweiterbar. Mapbox verwendet diese Bibliothek, weshalb der Prototyp Funktionen von Leaflet einsetzt.

**jQuery:** *jQuery*<sup>5</sup> ist eine freie, plattformunabhängige und quelloffene JavaScript-Bibliothek. Sie dient als clientseitige Scriptsprache zur Erstellung von Webanwendungen. Heutzutage ist jQuery eines der am häufigsten eingesetzten JavaScript-Bibliotheken. Die Hauptaugenmerke der Bibliothek liegen beim Selektieren und Manipulieren von DOM-Elementen, Animationen, Events sowie browserunabhängige asynchrone Kommunikation mit Servern (Ajax). Durch den modularen Aufbau können einfach leistungsstarke, dynamische Websites erstellt werden. Besonders für die Kommunikation mit der Datenbank sowie das Erstellen und Manipulieren der DOM-Elemente wurde das Framework im Prototypen genutzt.

**Bootstrap:** *Bootstrap*<sup>6</sup> ist ein populäres HTML-, CSS- und JavaScript-Framework für das Entwickeln von responsiven und *mobile-first* Webprojekten. Basierend auf einer einheitlichen Codebasis skaliert Bootstrap einfach und effizient Websites und Applikationen für Smartphones, Tablets oder Desktop-PCs mithilfe von *CSS-Media-Queries*. Zusätzlich enthält es Features wie ein Gridsystem, Themes oder vorgefertigte HTML-Komponenten. Die Grundstruktur des HTML-Codes wurde mithilfe des Frameworks erstellt.

**RainbowVis-JS:** *RainbowVis-JS*<sup>7</sup> ist eine quelloffene JavaScript-Bibliothek zur farblichen Visualisierung von Daten. Die Bibliothek generiert aufgrund eines zuvor definierten Farbspektrums sowie eines Wertebereichs Farben. Das Farbspektrum muss mit mindestens zwei Farben definiert werden. Diese werden auf den Wertebereich abgebildet, sodass ein Farbverlauf entsteht. Sobald eine Farbe berechnet werden soll, wird je nach angegebenem Wert die dazugehörige Farbe aus dem Farbverlauf retourniert. Der Prototyp verwendet die berechneten Farben u. a. für das Färben von Symbolen oder der Choroplethenkarten.

---

<sup>4</sup><http://leafletjs.com>

<sup>5</sup><http://jquery.com>

<sup>6</sup><http://getbootstrap.com>

<sup>7</sup><http://github.com/anomal/RainbowVis-JS>

## 5.2 Repräsentation von geografischen Daten

Für das Speichern von einfachen geografische Daten gibt es mittlerweile eine Vielzahl von Datenformaten (vgl. Abschnitt 2.5). Durchgesetzt hat sich heutzutage vor allem das Format *GeoJSON*<sup>8</sup>. GeoJSON ist ein offenes und standardisiertes Dateiformat, welches ein Merkmal oder eine Menge von einfachen, geografische Merkmalen mit ihren nicht-räumlichen Attributen strukturiert abbildet. Es unterstützt primitive geometrische Formen wie z. B. Punkte, Linien oder Polygone sowie komplexere wie z. B. ein Multi-Polygon oder eine Kollektion. GeoJSON ist technisch ein JSON-Objekt, wovon auch der Name des Datenformats abgeleitet wird. Daher besteht solch ein Objekt aus einer Menge von *Name-Wert-Paaren*, wobei ein Name stets aus einer Zeichenkette und der Wert entweder aus einer Zahl, Objekt, Feld, Zeichenkette oder eines der Literale *true/false* bzw. *null* bestehen muss. Ein Feld besteht wiederum aus einem der eben beschriebenen Name-Wert-Paaren. Die genaue Spezifikation von GeoJSON befindet sich u. a. online auf deren Website<sup>9</sup>. Der Prototyp nutzt dieses Datenformat für den Austausch und die Repräsentation der geometrischen Daten.

Folgend ist ein Beispiel für ein GeoJSON-Objekt beschrieben (s. Programm 5.1). Das Beispiel definiert ein einfaches Merkmal (zu erkennen an dem Attribut *type* in Zeile 2). Geometrische Eigenschaften werden im Attribut *geometry* angegeben. Das Merkmal besteht aus einem geometrischen Punkt (Zeile 4), wobei dieser durch Koordinaten (Zeile 5) spezifiziert wird. Zusätzliche und nicht-räumliche Beschreibungen zu dem Objekt werden in dem Attribut *properties* aufgeführt, wodurch der Punkt zusätzlich einen Namen erhält (Zeile 8).

```
1 {  
2   "type": "Feature",  
3   "geometry": {  
4     "type": "Point",  
5     "coordinates": [ 14.51366901397705, 48.36822124792832 ]  
6   },  
7   "properties": {  
8     "name": "FH Oberösterreich, Campus Hagenberg"  
9   }  
10 }
```

**Programm 5.1:** Beispielcode für ein GeoJSON-Objekt.

---

<sup>8</sup><http://geojson.org>

<sup>9</sup><http://geojson.org/geojson-spec.html>

### 5.3 Architektur und technische Ausführung

Wie aus Abbildung 5.1 ersichtlich besteht die Architektur des Prototypen lediglich aus einigen wenigen Komponenten. Die *Applikation* umfasst sämtlichen Code, der u. a. für die Interaktion mit den anderen Komponenten benötigt wird. Zum Beispiel werden hierüber Daten aus dem E-Commerce-System und der Karte geladen, aufbereitet und dargestellt. Des Weiteren enthält diese die *Business Logic* des Prototypen und ist daher auch das Herzstück der Anwendung.

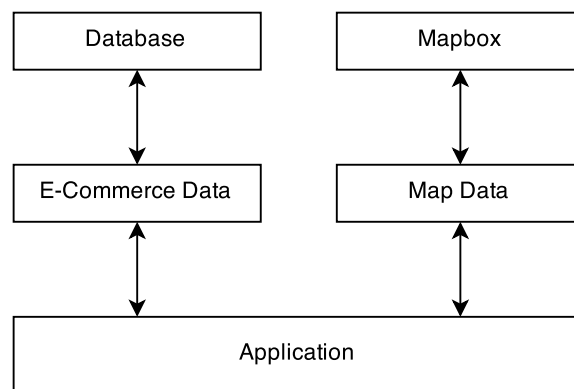


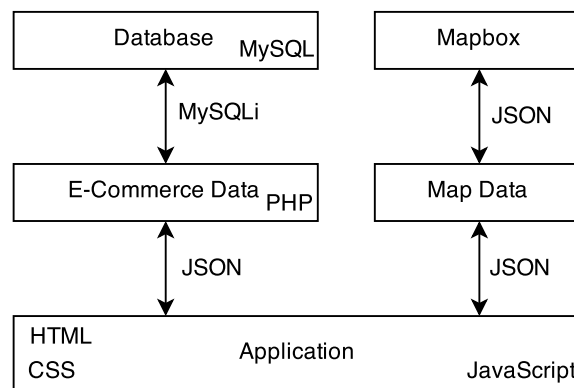
Abbildung 5.1: Aufbau der Architektur des Prototypen.

In einer relationalen (MySQL) Datenbank werden *E-Commerce Daten* gespeichert, welche über eine Schnittstelle zugänglich sind. Die Speicherung der Daten in einer Datenbank hat den Vorteil, dass weitere Datensätze einfach erweiterbar sind. Durch die Abstraktion kann auch eine andere, z. B. eine NoSQL-Datenbank angebunden werden. Die meisten E-Commerce-Systeme setzen jedoch eine relationale Datenbank ein, weshalb deren Daten in der relationalen migriert bzw. integriert werden könnten. Daher wurde bei dem Prototypen eine relationale Datenbank eingesetzt, gleichwohl eine schemalose eine bessere Performanz erzielen würde. Dieser Aspekt wurde jedoch nicht weiter beachtet, da ausschließlich die Visualisierung von geografischen Daten im Fokus steht.

Das API von *Mapbox* wird über die Komponente *Application* angesteuert und konfiguriert. Da der Prototyp derzeit ausschließlich Daten aus Österreich darstellt, wird initial die Karte auf die Bundesrepublik zentriert. Als Darstellungsart der Karte selbst hat sich die Terrainansicht bewährt, da diese meist eine gleichmäßige Farblandschaft einsetzt und dementsprechend wenig von den visualisierten Daten ablenkt. Des Weiteren werden nur einige wenige Details dargestellt, wie z. B. Städtenamen und Grenzen.

Der Prototyp wurde in der Auszeichnungssprache *Hypertext Markup Lan-*

*guage*<sup>10</sup> (HTML) sowie *JavaScript*<sup>11</sup> realisiert. *Hypertext Preprocessor*<sup>12</sup> (PHP) dient zum Datenaustausch zwischen Applikation und Datenbank. Wie bei Websites üblich ist das Design mithilfe von *Cascading Stylesheets*<sup>13</sup> (CSS) entwickelt worden. Zusammenfassend stellt Abbildung 5.2 die wesentlichen verwendeten Sprachen der einzelnen Komponenten sowie deren Datenaustauschformate dar.



**Abbildung 5.2:** Ausschnitt der verwendeten (Programmier-) Sprachen oder Technologien sowie Datenaustauschformate der einzelnen Komponenten des Prototypen.

## 5.4 Implementierung

Die Kommunikation der einzelnen Komponenten wurde mit dem bewährten Konzept der asynchronen Datenübertragung *Asynchronous JavaScript and XML* (kurz AJAX) implementiert (s. Programmcode 5.2). Dies bedeutet, dass sämtliche Serveranfragen (engl. Requests) asynchron ausgeführt werden. Der Browser lädt somit die angefragten Daten im Hintergrund, sodass der Nutzer stets eine reaktive Website auffindet, obwohl große Datenmengen im Hintergrund transferiert werden. In der Javascript-Datei *script.js* werden Daten aus der Datenbank mittels der jQuery-Funktion *getJSON* geladen. Anschließend werden diese in die Mapbox-Instanz integriert und, sofern notwendig, aufbereitet.

Zur besseren Veranschaulichung ist nachfolgend ein Programmcode angegeben, welcher eine PHP-Datei asynchron lädt (Zeile 1) und die retournierten Daten in ein Layer von Mapbox (Zeile 6) integriert. Optional können Parameter (Zeile 2-4) angegeben werden, um die Datenbankabfrage zu

<sup>10</sup><http://www.w3.org/html/>

<sup>11</sup><http://www.w3.org/standards/webdesign/script.html/>

<sup>12</sup><http://php.net>

<sup>13</sup><http://www.w3.org/Style/CSS/>



adaptieren. Die Funktion *dataLayer* visualisiert, basierend auf den Einstellungen aus dem Optionenmenü, die Daten (Zeile 7). So werden hier u. a. die Marker, Symbole oder Choroplethenkarte erstellt. Abschnitt 5.4.1 beschreibt verwendete Parameter der Requests für den Prototypen. Abschnitt 5.4.2 veranschaulicht mögliche Serverantworten (Response).

```
1 $.getJSON("PATH_TO_PHP_FILE.php", {
2     parameter1: value1,
3     parameter2: value2
4     // ,[...]
5 }, function (geojson) {
6     layer.setGeoJSON(geojson);
7     dataLayer(data);
8 });
```

**Programm 5.2:** Beispielcode für das asynchrone Laden der darzustellenden Daten

Wie oben angedeutet, werden innerhalb der Funktion *dataLayer* die Daten visualisiert. Folgend ist eine vereinfachte Variante für das Anzeigen von gefärbten proportionalen Symbolen (in diesem Fall Kreise) beschrieben. Ein Layer wird mittels der Mapbox-Funktion *L.geoJson* sowie den darzustellenden Daten erstellt (s. Programmcode 5.3). Der zweite Parameter gibt eine Funktion an, wie die Daten visualisiert werden sollen. Die Größe des Symbols resultiert aus dem zu repräsentierenden Datenwert.

Zur Vereinfachung wird ein weiterer, prozentualer Wert (*density*) angegeben, wobei der höchste 100% und der kleinste darzustellende Wert 0% entspricht. Dadurch kann ein Verhältnis zwischen aktuellem und maximalem Wert berechnet werden.

Die Funktion *L.circleMarker* (Zeile 6) erstellt letztendlich das Symbol, wobei u. a. die Farbe über die Funktion *getColor* ermittelt wird sowie die Größe (Radius) definiert wird. *addLayer* (Zeile 15 und 16) fügt den Layer mit den zuvor instanzierten Symbolen der Karte hinzu, sodass diese angezeigt werden. Die Legende wird über die Funktion *legendDataLayer.update* (Zeile 21) erstellt bzw. aktualisiert.

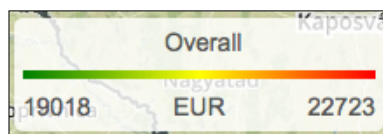
```

1 var layer = L.geoJson(data, {
2   pointToLayer: function (feature, latlng) {
3     var radius = feature.properties.density || 5;
4     var density = density || 1;
5
6     var marker = L.circleMarker(latlng, {
7       radius: radius,
8       fillColor: getColor(density)
9       // ,[...]
10    });
11
12    return marker;
13  }
14 });
15 circle.addLayer(layer);
16 map.addLayer(circle);
17
18 var max = data["features"][0]["properties"]['max'];
19 var min = data["features"][0]["properties"]['min'];
20
21 legendDataLayer.update(max, min, legendText, legendUnit);

```

**Programm 5.3:** Beispielcode für das Erstellen von Markern

Je nach Visualisierungsart ist die Datenprojektion unterschiedlich. Zum Beispiel können Marker oder eine Choroplethenkarte dargestellt werden. Diese Projektion wird in der Legende beschrieben, sodass Nutzer diese leichter interpretieren können. Zum besseren Verständnis ist nachfolgend ein Beispiel einer Legende angegeben. Abbildung 5.3 beschreibt beispielhaft das (gefilterte) Gesamteinkommen.



**Abbildung 5.3:** Beispiel für die Darstellung einer Legende.

Eine Legende gibt aus dem visualisierten Datenbestand das Wertspektrum (kleinster bis höchster darzustellender Wert) an, sowie deren Maßeinheit. Des Weiteren kann eine Überschrift angezeigt werden. Der Farbverlauf wird mittels einer Schleife erstellt, wobei wiederum die Funktion *getColor* eine Farbe anhand eines Wertes berechnet. Da wie oben beschrieben der kleinste 0% und der größte Wert 100% entspricht, kann zwischen diesen beiden iteriert werden.

### 5.4.1 Request

Wie zuvor in Abschnitt 5.4 beschrieben, werden sowohl die Primär- als auch Sekundärdaten aus der Datenbank über eine PHP-Schnittstelle geladen und über das GeoJSON-Format retourniert. Tabelle 5.1 stellt obligatorische sowie optionale Parameter der derzeit implementierten Requests für den Prototypen dar. In Abschnitt 5.4.2 wird der mögliche Response zu solch einem Request beschrieben.

<i>Parameter</i>	<i>Datentyp</i>	<i>V</i>	<i>O</i>	<i>U</i>	<i>C</i>
report	S	o	o	o	
type	S				o
grouped	B	o	o	o	
groupedType	S	+	+	+	
dateFrom	D	+	+		
dateTo	D	+	+		
sex	E	+	+	+	+
ageFrom	I	+	+	+	+
ageTo	I	+	+	+	+
product	S	+	+		

**Tabelle 5.1:** Übersicht über mögliche obligatorische und optionale Parameter für die Primär- und Sekundärdaten.

V = Besucher, O = Bestellung, U = Kunde, C = landesspezifische Daten (Sekundärdaten).

B = Boole'scher Wert, D = Datum, E = Enumeration, I = Ganzzahl, S = Zeichenkette

o = obligat, + = optional

Parameter werden über das PHP-Array `$_REQUEST` ausgelesen, weshalb diese sowohl über die HTTP-Methode *GET* als auch *POST* angegeben werden können. *report* und *type* definieren das gewünschte Attribut, welches als Wert zurückgegeben werden soll. *grouped* und *groupedType* werden für das Gruppieren der Daten, z. B. für die Choroplethenkarte, verwendet. *\*From* und *\*To* beschreiben jeweils ein Wertebereich. *sex* kann entweder männlich (m) oder weiblich (f) sein. Ist ein verpflichtender Parameter nicht festgelegt, so wird jeweils ein definierter Standardwert genutzt.

### 5.4.2 Response

Wie in den vorausgehenden Abschnitten beschrieben, besteht der Response (Antwort vom Server) zum Laden der Primär- sowie Sekundärdaten aus einem GeoJSON-Objekt (vgl. Abschnitt 5.2). Abhängig vom Request besteht der Response aus einer Menge an Merkmalen. Ein Merkmal ist entweder vom Typ *Point*, *Polygon* oder *MultiPolygon*. Des Weiteren besteht ein Merkmal je nach Typ aus einem oder mehreren Koordinaten. Koordinaten werden stets als Tupel von Längen- und Breitengrad angegeben.

```
1 {  
2   "type": "FeatureCollection",  
3   "features": [  
4     {  
5       "type": "Feature",  
6       "geometry":  
7         {  
8           "type": "Polygon",  
9           "coordinates":  
10          [  
11            [  
12              [16.3481314, 48.2904798],  
13              [16.3535122, 48.2836886],  
14              [16.3559554, 48.2833477]  
15              //,[...]  
16            ]  
17          ]  
18        },  
19        "properties":  
20        {  
21          "name": "Wien",  
22          "max": 22723,  
23          "min": 19018,  
24          "value": 21280,  
25          "density": 61  
26        }  
27      }  
28      //,[...]  
29    ]  
}
```

**Programm 5.4:** Beispielcode für ein Response zu einem Request aus Abschnitt 5.4.1

Im Programmcode 5.4 wird ein vereinfachter Response für die Region *Wien* dargestellt. Der grundsätzliche Aufbau der verschiedenen Responses ist nahezu identisch. In Zeile 2 wird definiert, dass mehrere Merkmale beschrieben werden (*FeatureCollection* spezifiziert eine folgende Kollektion von Objekten). Merkmale werden im Attribut *features* angeführt. Der Typ

des geografischen Merkmals wird in Zeile 8 angegeben. Die weiteren, nicht-räumlichen Attribute variieren je nach vorangegangenen Request.

## 5.5 Nutzungsmöglichkeiten für Echtsysteme

Der Prototyp ist derzeit mit keinem Echtsystem verbunden. Wie in Abschnitt 4.1.6 spezifiziert, soll darauf geachtet werden, dass ein echtes E-Commerce-System in den Prototypen integriert werden kann. Dieser Abschnitt beschreibt solch Möglichkeiten der Systemintegration.

### 5.5.1 Adaption der PHP-Schnittstelle

Eine Möglichkeit besteht darin, die PHP-Schnittstelle zu modifizieren. Dies bedeutet, dass die Schnittstelle zwischen Applikation und Datenbank so angepasst wird, dass diese an ein Echtsystem gekoppelt ist.

Dazu muss sämtlicher (My-) SQL-Code (innerhalb der *geojson\_*- sowie der *Core.php*-Dateien) adaptiert werden und auf das Datenbankschema des Systems angeglichen werden. Hierzu zählen u. a. die Statements für die Datenabfrage sowie die Berechnung des minimalen sowie maximalen Wertes. Sollten sich die Verbindungsparameter bzw. der Datenbanktreiber geändert haben, so ist eine Adaption der *init*-Methode in der *Core.php* obligatorisch. Bei der Berechnung für die Gruppierung von Daten zu einer Region sind ebenfalls Änderungen erforderlich.

### 5.5.2 Adaption des Datenbankschemas mithilfe von Views

Leichter als die Adaption der PHP-Schnittstelle ist das Anwenden von *Datenbank-Views* (Sicht). Ein View ist eine virtuelle Tabelle, die aus anderen Tabellen abgeleitet wurde. Solche Tabellen werden in aller Regel nicht physisch gespeichert. Der Vorteil ist, dass das dahinterliegende Datenbankschema unabhängig von der View ist, solange diese das Schema kennt. Oftmals werden Views eingesetzt, damit Benutzer nur einen gewissen Teil der Tabelle einsehen (und bearbeiten) können [10].

Dieses Konzept lässt sich auch für die Integration nutzen, da hierbei lediglich die Views erstellt werden müssten, um die Daten vom Echtsystem auslesen zu können. Sollte sich das Datenbankschema nicht ändern, wären keine weiteren Modifikationen notwendig.

## Kapitel 6

# Evaluierung

Ziel der Evaluierung ist, herauszufinden, welche Visualisierungsarten für die Darstellung von Informationen aus einem Webshop geeignet sind. Daher liegt ein Hauptaugenmerk auch auf der Bewertung dieser verschiedenen Darstellungen. Dafür wurden typische Fragestellungen aus der Datenanalyse für einen Webshop erstellt. Die Teilnehmer sollen mithilfe der Darstellungen die Fragen beantworten. Das Ergebnis soll aufzeigen, welche der Visualisierungsarten sich für die Beantwortung der Fragen eignen.

Nicht inkludiert im Fragebogen ist die Erhebung von demographischen Daten der Teilnehmer. Es ist davon auszugehen, dass weder das Alter noch das Geschlecht einen Einfluss auf das Ergebnis haben. Die Bewertung des Fragebogens durch die Teilnehmer soll rein subjektiv sein und das eigene Empfinden widerspiegeln. Jedoch wurde zu Beginn des Fragenkatalogs ermittelt, ob die Teilnehmer Erfahrung mit der Erstellung von Websites haben. Durch diese qualitative Datenerhebung soll die Aussagekraft der Evaluierung erhöht werden.

Abschließend ist jede Visualisierungsart einzeln zu bewerten. Dabei werden verschiedene Kenngrößen anhand einer Notenskala eingestuft. Zusätzlich können die Teilnehmer weitere positive sowie negative Ausprägungen der Darstellung beschreiben.

### 6.1 Teilnehmer und Durchführung

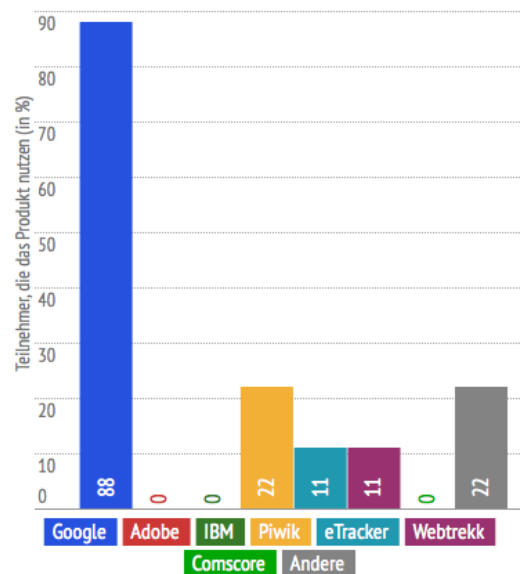
Die qualitative Evaluierung des Prototypen erfordert zunächst Testpersonen, die sich mit der professionellen Erstellung bzw. Wartung von Websites auseinandersetzen. Daher wurde die Evaluierung lediglich an Personen aus dieser Zielgruppe versendet. Insgesamt teilgenommen an der Befragung haben 14 Personen, 12 davon haben den Fragebogen vollständig beantwortet. Die Durchführungsdauer betrug zwei Wochen.

## 6.2 Ergebnisse

In den folgenden Abschnitten werden Teile der Ergebnisse aus der Evaluierung resümiert. Die einzelnen Bewertungen des Fragebogens befinden sich auf der beiliegenden DVD zu dieser Masterthesis.

### 6.2.1 Zielgruppenanalyse

13 der 14 Teilnehmer haben angegeben, Websites professionell zu erstellen. Die Frage, wie lange sich die Teilnehmer schon mit Websites auseinander setzen wurde unterschiedlich beantwortet. Von den 12 Angaben gaben fünf Personen weniger als ein Jahr, drei zwischen einem und sechs und die übrigen mehr als sieben Jahre an. Aufgrund dessen, dass vier Teilnehmer angaben an mindestens 4 Tagen mit der Erstellung von Websites zu tun haben, ist davon auszugehen, dass diese ihre berufliche Tätigkeit auch in diesem Bereich fällt. Dies spiegelt sich auch bei der Frage nach der Anzahl der bislang erstellten Websites wider. Hier haben 5 Teilnehmer angegeben, bisher mehr als 11 Websites erstellt zu haben.



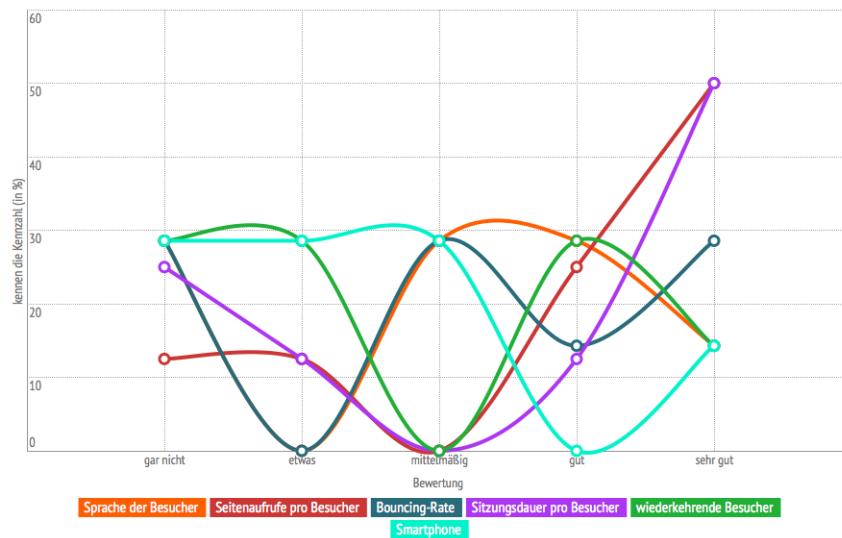
**Abbildung 6.1:** Verteilung der genutzten Datenanalysetools Hersteller bzw. Produkte.

Neun von zwölf Teilnehmer gaben zusätzlich an, ein Datenanalysetool zum Protokollieren der Besucheraktivitäten zu nutzen. Dabei wird besonders häufig *Google Analytics* (8 von 9) eingesetzt. Die Open-Source Software *Piwik* nutzen zwei, *eTracker* und *Webtrekk* jeweils ein Teilnehmer. Die vollständige Verteilung wird in Abbildung 6.1 gezeigt. Auffällig ist, dass die

meisten Teilnehmer (6 von 9) nur einmal pro Woche ein Datenanalysetool einsetzen. Dennoch finden vier Teilnehmer diese *hilfreich* oder sogar *sehr hilfreich*.

### 6.2.2 Datenanalysetool

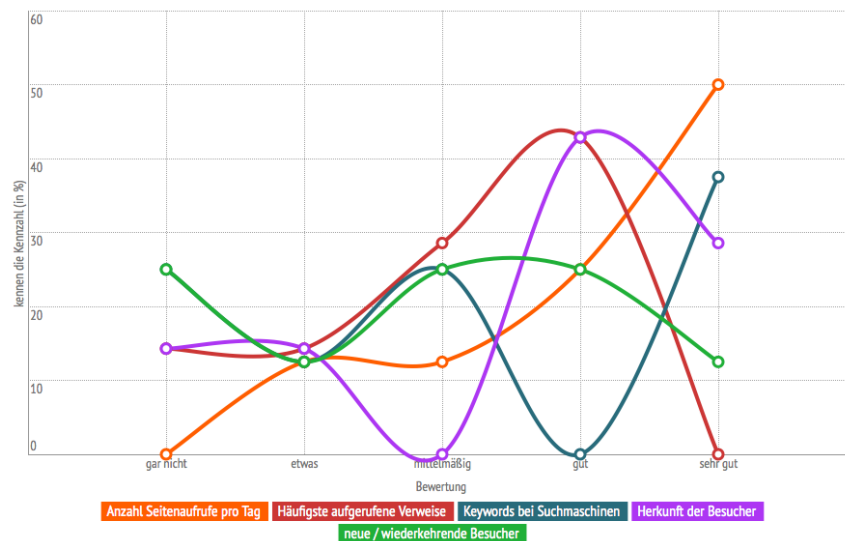
Besonders gut kennen die Teilnehmer die Kennzahl *Anzahl der Seitenaufrufe pro Tag* für eine beliebige Website, für die sich auch Zugang zu einem Datenanalysetool haben. 75% der Teilnehmer gaben an, diesen Wert *gut* oder *sehr gut* zu kennen. Auch die Herkunft der Besucher ist allgemein *gut* (43%) oder *sehr gut* (29%) bekannt. Die Liniendiagramme 6.2 und 6.3 zeigen die Verteilung der Bewertung für alle gefragten Kennzahlen. Für eine bessere Übersichtlichkeit wurden die Kennzahl auf zwei Diagramme aufgeteilt.



**Abbildung 6.2:** Verteilung der Bewertung auf die Frage, *wie gut die Teilnehmer verschiedene Kennzahlen einer Website aus einem Datenanalysetool kennen*. (Teil 1/2)

Fast 38% haben sich schon einmal Daten auf einer Karte anzeigen lassen. Alle Teilnehmer fanden dabei diese Art der Darstellung für die Beantwortung ihrer Fragen *hilfreich* (1) oder *sehr hilfreich* (2).





**Abbildung 6.3:** Verteilung der Bewertung auf die Frage, *wie gut die Teilnehmer verschiedene Kennzahlen einer Website aus einem Datenanalysetool kennen*. (Teil 2/2)

### 6.2.3 Visualisierungsarten

Abbildung 6.4 zeigt für die evaluierten Visualisierungsarten die jeweilige Verteilung aller abgegebenen Bewertungen. Ermittelt wurde jeweils ein Mittelwert.

#### Choroplethenkarte

Durchwegs gut bewertet wurden Choroplethenkarten. Häufig konnten mit dieser Darstellung die Fragen eindeutig beantwortet werden. Nur Diagramme konnten bei dieser Erhebung besser abschneiden. Choroplethenkarten werden allgemein ein guter Abstand, Anordnung, Farbwahl und Größe der Symbole attestiert. Etwas schlechter beurteilt wurde ausschließlich die Übersichtlichkeit und Eindeutigkeit, gleichwohl beide Ausprägungen im Durchschnitt noch mittelmäßige Bewertungen erhielten. So haben z.B. bei der Frage nach der Eindeutigkeit zwar 30% der Teilnehmer diese Darstellung als *schlecht* oder *eher schlecht* beurteilt. Auf der anderen Seite meinten jedoch 70%, dass diese Darstellung eine *gute* oder gar *sehr gute* Eindeutigkeit aufweist.

Generell würden knapp 60% der Befragten diese Art der Visualisierung auch in einem Datenanalysetool verwenden. Besonders die schnelle Informationserfassung sowie die Erkennung einer Region sind positiv wahrgenommen worden. Als teilweise verwirrend wurde der zu geringe Farbunterschied bei nahezu identischen Werten interpretiert. Auch wurde angegeben, dass

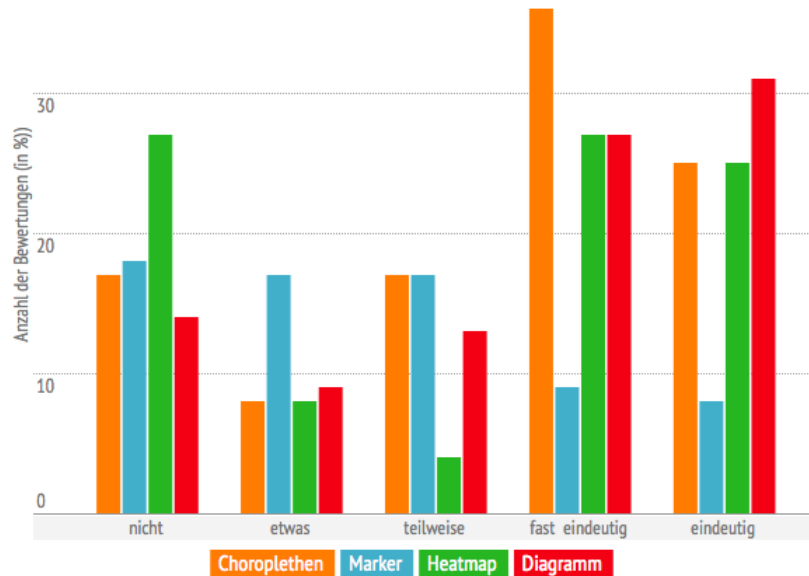


Abbildung 6.4: Verteilung aller abgegebenen Bewertung.

sich diese Darstellung wohl nur bei größere Flächen eignen würde. Insgesamt wurde diese unter allen Visualisierungsarten als zweitbeste aufgefasst.

### Marker

Wie in Abbildung 6.4 ersichtlich, wurden Marker unterschiedlich bewertet. Positiv bewertet wurde u. a. die Größe der Symbole (90% fanden diese *gut* oder *sehr gut*). 33% der Teilnehmer empfanden auch die Übersichtlichkeit *gut*, 17% sogar *sehr gut*. Auch der Abstand der Symbole wurde im Durchschnitt angenehm eingeschätzt, nur 18% der Teilnehmer evaluierten diese Ausprägung als *eher schlecht*. Hingegen negativ beurteilt wurde das gewählte Farbschema der Symbole. Nur fünf der 12 Bewertenden fanden diese mindestens *gut*.

Bemerkenswert ist, dass trotz der allgemein guten Bewertung diese Visualisierungsart nur 17% der Teilnehmer auch in einem Datenanalysetool nutzen würden. Die schlechte Beurteilung wird auch durch die offenen Fragen bestätigt. Hier wird insbesondere die nicht gegebene Eindeutigkeit bemängelt. Auch ist bei großen Datenmengen die Qualität der Symbole nicht ausreichend, weshalb die Karte als unübersichtlich wahrgenommen wird.

### Heatmap

Die Darstellung als Heatmap wurde im allgemeinen mittelmäßig aufgenommen. Immerhin gaben 40% der Teilnehmer an, dass Heatmaps eine *gute* oder *sehr gute* Symbolgröße haben. Bemängelt wurde hingegen besonders die Eindeutigkeit der dargestellten Informationen. 70% der Befragten fanden diese *eher schlecht* oder *schlecht*. Dennoch konnten einige Fragen *fast eindeutig* oder sogar *eindeutig* beantwortet werden.

Trotz alledem würden nur 25% aller Teilnehmer diese Art der Darstellung in einem Datenanalysetool nutzen. Bemängelt wurden vor allem die ungenauen Informationen sowie die schlechte Farbwahl.

### Diagramm

Die Visualisierungsart Diagramme wurden mit am Besten bewertet. Die gegebenen Fragen können oft eindeutig beantwortet werden. Positiv sind die Werte bei der Eindeutigkeit (81% *sehr gut*) sowie Übersichtlichkeit (92% *gut* oder *sehr gut*) der genutzten Symbole. Den referenzierenden Ort auf der Karte zu finden ist mangels der nicht vorhandenen Karte nicht möglich. Dennoch reichte der Regionsname für knapp 30% der Teilnehmer. Wie auch bei den Kartendarstellungen sollten auch Diagramme farblich hervorgehoben werden. Die schwarz-weiße Farbgebung wurde als schlecht kritisiert.

Diagramme haben mit fast 67% den höchsten Wert bei der Verwendung in einem Datenanalysetool. Dass diese Art der Darstellung heutzutage häufig verwendet wird, ist berechtigt. Aber sobald auf geobasierte Informationen referenziert wird, fehlt die visuelle Zuordnung.

# Kapitel 7

## Résumé

Informationen aus einem E-Commercesystem können mit verschiedenen Darstellungen visualisiert werden. Datenanalysetools, wie z. B. Google Analytics, nutzen dabei häufig Tabellen oder Diagramme. Doch diese Darstellungen erschweren die Interpretation von Daten, die sich auf einen bestimmten Ort beziehen. Der Grund ist recht eindeutig: es fehlt die örtliche Zuordnung, da diese Darstellungen ausschließlich den Namen des Ortes angeben können.

Die Visualisierung von geobasierten Informationen auf einer Karte kann dem Betrachter genau diese Referenzpunkte geben. Der Betrachter kann u. a. selbst den Kartenausschnitt oder den Detailgrad definieren, sodass Daten visuell einfacher analysiert werden können. E-Commercesysteme, wie z. B. ein Webshop, generieren viele Daten, die an einem bestimmten Ort entstehen. Dies liegt vor allem daran, dass grundsätzlich Menschen mit einem Webshop interagieren und sich diese an einem definierten Ort aufhalten. Datenanalysten können anhand des Standortes des Benutzers auch ortsbezogene Marketingmaßnahmen steuern.

Der implementierte Prototyp kann substanzielle Informationen aus einem fiktiven Webshop mit unterschiedlichen Visualisierungsarten darstellen. Dabei kann der Betrachter die Datenquelle sowie deren Eingrenzung bestimmen. Zusätzlich ist es möglich, aus verschiedenen Sekundärquellen weitere Informationen der Karte hinzuzufügen.

Für die Evaluierung wurden übliche Fragen von einem Marketinganalysten für solch ein System erstellt. Die Teilnehmer beurteilten, wie eindeutig die gegebenen Fragen mithilfe der Darstellungen beantwortet werden konnten. Um Vor- und Nachteile der unterschiedlichen Visualisierungsarten zu ermitteln, wurde abschließend jede einzeln bewertet. Der Fragebogen wurde an Teilnehmern durchgeführt, die bereits Erfahrung mit der Erstellung und Wartung von Websites haben.

Das Ergebnis besagt, dass nicht alle Daten sich gleichermaßen gut auf einer Karte repräsentieren lassen. Die Verwendung einer Karte anstatt z. B. eines Diagramms ist auch von der Fragestellung sowie des vorhandenen Da-

tenbestandes abhängig. Generell konnten die Fragen mit einem Diagramm mehrfach eindeutig beantworten. Bemängelt wurde jedoch, dass teilweise der örtliche Bezug fehlte. Auch die Darstellung als Choroplethenkarte wurde als nützlich bewertet. Mehr als die Hälfte aller Befragten würden die Choroplethenkarte auch in einem Datenanalysetool verwenden. Unter allen evaluierten Kartenvisualisierungen stellte dies den höchsten Wert dar.

Sowohl Marker als auch Heatmaps bekamen divergente Bewertungen. Es ist daher davon auszugehen, dass diese von den Teilnehmern ungleich wahrgenommen wurden. Besonders bei diesen Darstellungen wurde deutlich, dass die qualitative Nutzung der Visualisierung auch von der jeweiligen Fragestellung abhängt. Obwohl die Darstellung mit Markern summarisch ein schlechtes Endergebnis erzielte, ergeben sich daraus Anwendungsgebiete. Marker haben den Vorteil, dass diese sehr genau auf einen Ort referenzieren. Allerdings ist so ein genaues Abbild der Informationen bei einem Webshop nicht ausschlaggebend, sondern vielmehr ein akkumuliertes Gesamtbild einer Region.

Im Allgemeinen gilt, dass die Farbwahl und Anordnung der Symbole einen starken Einfluss auf die Qualität der Visualisierung haben. Beim implementierten Prototypen können diese Aspekte derzeit nicht verändert werden. Es wurde außerdem angegeben, dass teilweise nicht eindeutig ist, was die Farben genau ausdrücken sollen.

Zusammenfassend kann festgestellt werden, dass die Darstellung von geobasierten Informationen auf einer Karte durchaus hilfreich sein kann, gleichwohl dies abhängig von der Fragestellung ist. Besonders die Choroplethenkarte eignet sich für viele Fragestellungen aus dem Bereich der E-Commercesysteme. Ich empfehle, für den Prototypen weitere Darstellungsoptionen zu entwickeln, sodass der Benutzer zusätzliche Anpassungsmöglichkeiten hat.

Besonders die Farbgestaltung der Symbole scheint nicht trivial zu sein. Zum Beispiel könnte verifiziert werden, welche Farbgestaltungen bei einer Choroplethenkarte unterstützend wirkt. Darüber hinaus gilt es genauer festzustellen, welche Auswirkungen sich auf die Interpretation der Darstellung bei korrelierenden Informationen ergeben. Diese Aspekte wurden in der Evaluierung bislang kaum beachtet.

# Quellenverzeichnis

## Literatur

- [1] Marcos D. Assunção u. a. „Big Data Computing and Clouds: Trends and Future Directions“. In: *Journal of Parallel and Distributed Computing* 79–80 (2015), S. 3–15 (siehe S. 21).
- [2] Norbert Bartelme. *Geoinformatik: Modelle, Strukturen, Funktionen*. 4. Aufl. Berlin & Heidelberg: Springer-Verlag, 2005 (siehe S. 18).
- [3] Agnieszka (Aga) Bojko. „Informative or Misleading? Heatmaps Deconstructed“. In: *Human-Computer Interaction. New Trends*. Hrsg. von Julie A. Jacko. Berlin & Heidelberg: Springer-Verlag, 2009, S. 30–39 (siehe S. 16).
- [9] Max J. Egenhofer, Harvey J. Miller und Christian Freksa. *Geographic Information Science*. Berlin & Heidelberg: Springer-Verlag, 2004 (siehe S. 18).
- [10] Ramez A. Elmasri und Shamkant B. Navathe. *Grundlagen von Datenbanksystemen*. 3. Aufl. München: Pearson Studium, 2009 (siehe S. 42).
- [11] Martin Ester und Jörg Sander. *Knowledge Discovery in Databases: Techniken und Anwendungen*. Berlin & Heidelberg: Springer-Verlag, 2000 (siehe S. 22).
- [13] Ali M. Al-Ghamdi. „Optimising the Selection of a Number of Choropleth Map Classes“. In: *Thematic Cartography for the Society*. Hrsg. von Temenoujka Bandrova, Milan Konecny und Sisi Zlatanova. Cham, Schweiz: Springer International Publishing, 2014, S. 61–77 (siehe S. 8).
- [14] J. B. Harley. „The Map and the Development of the History of Cartography“. In: *The History of Cartography: Cartography in Prehistoric, Ancient and Medieval Europe and the Mediterranean*. Hrsg. von J. B. Harley und David Woodward. Chicago, USA: University Of Chicago Press, 1987. Kap. 1, S. 1–42 (siehe S. 5).
- [15] John N. Hatzopoulos. *Topographic Mapping – Covering the Wider Field of Geospatial Information Science & Technology (GIS & T)*. Florida, USA: Universal-Publishers, 2008 (siehe S. 7).

- [16] Daniel A. Keim u. a. „Visual Analytics: Scope and Challenges“. In: *Visual Data Mining*. Hrsg. von Simeon J. Simoff, Michael H. Böhlen und Arturas Mazeika. Berlin & Heidelberg: Springer-Verlag, 2008, S. 76–90 (siehe S. 22).
- [17] Menno-Jan Kraak. „Playing with Maps Explore, Discover, Learn, Categorize, Model, Analyse, Explain, Present Geographic and Non-geographic Data“. In: *Proceedings of the Conference on Information Visualization*. (London). Washington, USA: IEEE Computer Society, Juli 2006, S. 291–296 (siehe S. 6, 22).
- [18] Menno-Jan Kraak und Allan Brown. *Web Cartography – developments and prospects*. London, UK: Taylor & Francis, 2001 (siehe S. 17).
- [19] Menno-Jan Kraak und Ferjan Ormeling. *Cartography: visualization of spatial data*. 3. Aufl. New York, USA: Pearson Education Limited, 2011 (siehe S. 5, 6, 18).
- [20] John Krygier und Denis Wood. *Making Maps: A Visual Guide to Map Design for GIS*. 2. Aufl. New York, USA: The Guilford Press, 2011 (siehe S. 9).
- [21] Norbert de Lange. *Geoinformatik*. 3. Aufl. Berlin & Heidelberg: Springer-Verlag, 2013 (siehe S. 8–10).
- [22] Paul A. Longley u. a. *Geographic Information Systems and Science*. 2. Aufl. New York: John Wiley & Sons, 2005 (siehe S. 18, 22, 23).
- [23] Alan M. Maceachren und David Dibiase. „Animated Maps of Aggregate Data: Conceptual and Practical Problems“. In: *Cartography and Geographic Information Systems* 18 (1991), S. 221–229 (siehe S. 12, 13).
- [25] Amir Manzoor. *E-Commerce*. Saarbrücken: Lambert Academic Publishing, 2010 (siehe S. 21).
- [26] Rameshwar P. Misra und A. Ramesh. *Fundamentals of Cartography*. New Delhi, Indien: Concept Publishing Company, 1969 (siehe S. 5, 7).
- [27] Frank .J. Ohlhorst. *Big Data Analytics: Turning Big Data into Big Money*. New York, USA: John Wiley & Sons, 2013 (siehe S. 21).
- [28] Gerold Olbrich, Michael Quick und Jürgen Schweikart. *Desktop Mapping*. 3. Aufl. Berlin & Heidelberg: Springer-Verlag, 2002 (siehe S. 8).
- [31] Terry A. Slocum u. a. *Thematic Cartography and Geovisualization*. 3. Aufl. New York, USA: Pearson Education Limited, 2013 (siehe S. 7–13).

- [34] Europäische Union. *Richtlinie 2006/24/EG des Europäischen Parlaments und des Rates vom 15. März 2006 über die Vorratsspeicherung von Daten, die bei der Bereitstellung öffentlich zugänglicher elektronischer Kommunikationsdienste oder öffentlicher Kommunikationsnetze erzeugt oder verarbeitet werden, und zur Änderung der Richtlinie 2002/58/EG*. Amtsblatt der Europäischen Union, L 105/54. URL: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:DE:PDF> (siehe S. 21).
- [35] Europäische Union. *Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE)*. Amtsblatt der Europäischen Union, L 108/1. URL: <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002> (siehe S. 23).
- [36] Andrew B. Whinston, Soon-Yong Choi und Dale O. Stahl. *The Economics of Electronic Commerce*. New York, USA: Pearson Education Limited, 1997 (siehe S. 20).
- [37] Ian H. Witten, Eibe Frank und Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. 3. Aufl. Burlington, USA: Elsevier, 2011 (siehe S. 21).

## Online-Quellen

- [4] Forschung und Wirtschaft Bundesministerium für Wissenschaft. URL: [http://www.bev.gv.at/portal/page?\\_pageid=713,1605147%5C&\\_dad=portal%5C&\\_schema=PORTAL](http://www.bev.gv.at/portal/page?_pageid=713,1605147%5C&_dad=portal%5C&_schema=PORTAL) (besucht am 10.03.2015) (siehe S. 8).
- [5] Natural Resources Canada. URL: <http://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/maps/9803#1> (besucht am 13.03.2015) (siehe S. 7).
- [6] Open Geospatial Consortium. URL: <http://www.opengeospatial.org/ogc/historylong> (besucht am 28.03.2015) (siehe S. 18).
- [7] Open Geospatial Consortium. URL: <http://www.opengeospatial.org/standards/gml> (besucht am 28.03.2015) (siehe S. 18).
- [8] Edd Dumbill. *What is big data? — An introduction to the big data landscape*. URL: <http://radar.oreilly.com/2012/01/what-is-big-data.html> (besucht am 04.05.2015) (siehe S. 21).
- [12] Michael Friendly und Daniel J. Denis. *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. URL: <http://www.datavis.ca/milestones/> (besucht am 04.04.2015) (siehe S. 6).



- [24] Tim Maly. *A Cloudless Atlas — How MapBox Aims to Make the World's 'Most Beautiful Map'*. URL: <http://www.wired.com/2013/05/a-cloudless-atlas/> (besucht am 26.04.2015) (siehe S. 33).
- [29] Arbeitskreis Vorratsdaten Österreich. URL: <https://www.akvorrat.at/about> (besucht am 04.05.2015) (siehe S. 21).
- [30] Greg J. Schmidt, Misako Nishino und John Kartes. *Density Gradient Map Samples Produced From BONAP's Floristic Synthesis*. URL: <http://www.bonap.org/diversity/diversity/diversity.html> (besucht am 03.04.2015) (siehe S. 15).
- [32] Joel Stein. *Data Mining: How Companies Now Know Everything About You*. URL: <http://content.time.com/time/magazine/article/0,9171,2058205,00.html> (besucht am 04.05.2015) (siehe S. 21).
- [33] David Totman. URL: <http://www.pwmag.com/computerized-maintenance-management-system/static-web-maps-vs-dynamic-web-gis.aspx> (besucht am 03.04.2015) (siehe S. 17).