

**Abgabe der Projektarbeit im Seminar:  
„Informationsvisualisierung und Sprache“  
SoSe 2020 bei Dr. André Calero Valdez  
Vorgelegt von  
Gapinaya Ganeshabalan (349603)  
und Lena Pohlmann (412443)**

## **Visualisierungskonzept**

### **Definition Visualisierungsziel:**

Das Visualisierungsziel zum Projekt “Black Education Matters” verfolgt einige übergeordnete Ziele. Zum einen ist es die verständliche Einsicht und Analyse der Daten und zum anderen die Kommunikation relevanter Inhalte sowie die Erregung von Aufmerksamkeit. Informationen und Daten werden dem Datensatz entnommen und Zusammenhänge miteinander verknüpft, die vorher nicht offensichtlich waren. Hierdurch wird es dem Rezipienten ermöglicht, sich mit weniger kognitivem Aufwand mit diesen auseinanderzusetzen. Des Weiteren sollen Nicht-Experten sowie Experten die Bedeutsamkeit von allgemein zugänglicher Bildung auf eine Art und Weise nähergebracht werden, die für jeden verständlich ist. Zudem soll ihre Meinungsbildung in eine bestimmte Richtung gelenkt werden. Es soll auf das Thema aufmerksam gemacht und aufgeklärt werden, um mehr Bewusstsein für das Thema und schließlich eine Handlungsbeeinflussung zu erzielen. Daher werden die Zusammenhänge im Datensatz so einfach wie möglich, aber dennoch sehr informativ dargestellt, um einerseits mehr Übersicht über die Daten und andererseits Informativität zu gewährleisten.

### **Kurzzusammenfassung Projekt**

Für das Projekt wurden verschiedene Datensätze aus dem Format TidyTuesday auf der Internetseite Github gewählt. „College tuition, diversity, and pay“ sowie „African American Achievements“ bieten aktuelle Datensätze aus dem Jahr 2020 und „School Diversity“ wurde 2019 zur Verfügung gestellt. Innerhalb des ersten Datensets wurde sich auf aktuelle Studienkosten, ihre Entwicklung und die Diversität an amerikanischen Universitäten und Colleges konzentriert. Vergleichend dazu wurde diesen Daten die Diversität in amerikanischen Schulen aus dem Datensatz „School Diversity“ gegenübergestellt. Um die Erkenntnisse der vorherigen Datensätze in einem weiteren Kontext zu verordnen, wurde außerdem der Datensatz zu den Errungenschaften von Afroamerikanern verwendet.

Der Fokus des Projekts liegt auf der Darstellung des Zusammenspiels von Diversität und Studiengebühren im amerikanischen Hochschulsystem. Die Datensätze zur Diversität haben viele thematische Überschneidungen und ähneln sich sowohl in ihrem zeitlichen als auch in ihrem geografischen Bezug und lassen sich somit gut vergleichen. Von einer Zusammenführung der Datensätze wurde jedoch abgesehen, da die Daten unterschiedlich formatiert wurden. Der Datensatz zu den Errungenschaften von Afroamerikanern dagegen hat zwar einen entfernteren thematischen Bezug und zeigt gänzlich andere Variablen, unterstützt jedoch die Aussagen der vorherigen Daten in Bezug auf das Visualisierungsziel.

### **Definition Zielgruppe**

Das Projekt richtet sich zunächst besonders an Studieninteressierte sowie Universitäten und schließlich an die Politik, um gerade innerhalb dieser Gruppen das Visualisierungsziel zu erreichen. Dabei lässt sich sagen, dass der Bildungsstand vor allem zwischen dem eines Studienanfängers und einer promovierten und bereits arbeitenden Person liegt. Auch die Allgemeinbevölkerung spielt jedoch eine Rolle, da eine möglichst breite Masse erreicht werden soll. Der Bildungsstand kann hier jedoch stark variieren. Aufgrund dieser heterogenen, breiten Zielgruppe werden Aspekte wie Alter, Geschlecht und Herkunft nicht weiter eingegrenzt.

Wichtig ist vor allem, auch in Bezug zu dem Thema des Projektes, die Inhalte dieser breit aufgestellten Zielgruppe zugänglich zu machen sowie unterschiedlich detaillierte Informationen bereitzustellen, um jeweils eine Identifizierung mit dem Thema zu ermöglichen.

### **Begründung und Erklärung zu den Entscheidungen bzgl. des Storytellings zur Zielerreichung**

Zu Beginn und zum Schluss dienen die Texte zur Aufmerksamkeitserregung und Orientierung sowie zur Betonung der Dringlichkeit und Bedeutsamkeit des Themas. Die Texte in den einzelnen Unterkapiteln dienen dazu, die Diagramme besser zu verstehen und die aus den Diagrammen resultierenden Aussagen zu stützen. Damit soll das Verständnis der wichtigsten Informationen für die gesamte Zielgruppe garantiert und die weitere Orientierung im Gesamtkontext gesichert werden.

Auf der sprachlichen Ebene wurden alle verwendeten Textbausteine in einer alltäglichen und einfachen Sprache gehalten und keine Fachbegriffe sowie ein einfacher Satzbau verwendet. Die Forderungen der Webseite sollen klar zum Vorschein kommen und unmissverständlich beim Rezipienten ankommen. Begriffe aus den Datensätzen, die in ihrer Definition nicht klar abgegrenzt sind, wurden im Text näher beschrieben und erklärt. Dadurch soll der gesamten Zielgruppe die wichtigsten Inhalte des Projektes zugänglich gemacht werden. Das Projekt ist außerdem in englischer Sprache verfasst, da große Teile der Zielgruppe, wie die Universitäten oder die Politik auf die USA begrenzt sind und so außerdem eine breitere Masse erreicht werden kann.

Wir haben uns für eine Webseite entschieden, da der Großteil der Gesamtbevölkerung zunächst im World Wide Web nach Informationen sucht und das Herunterladen einer App auch einen Mehraufwand für den Rezipienten bedeutet. Hinzu kommt, dass wir überwiegend Wissen weitergeben möchten und kaum Interaktionen beabsichtigen. Die Webseite ist interaktiv genug, um Aufmerksamkeit zu erregen, aber nicht in solchem Maße interaktiv, als dass eine App benötigt wird. Der Rezipient kann durch das interaktive Diagramm selbst tätig werden, sodass er sich direkt angesprochen fühlt, seine Aufmerksamkeit steigt und er motiviert wird, weiterzulesen. Insgesamt sind die Visualisierungen aber größtenteils "author-driven", da nur eine vereinzelte Interaktion möglich ist und die Diagramme ansonsten starke, eindeutige Aussagen vermitteln und so zur Meinungssteuerung beitragen. Da es verschiedene Kapitel zum Thema gibt, hätte auch ein Poster nicht ausgereicht, da es inhaltlich den Rahmen übertroffen hätte und damit den Zweck eines Posters verfehlt hätte.

Die Anordnung der Diagramme folgt dem inhaltlichen Handlungsstrang des Projektes. Ein roter Faden zieht sich also nicht nur durch die Diagramme, sondern wird vom sprachlichen Teil zusätzlich unterstützt. Anfangs wird auf das Problem aufmerksam gemacht. Von einem allgemeinen Überblick über die Diversität von Universitäten und Colleges gelangt man zu detaillierten Informationen der einzelnen Staaten und zum weiteren Vergleich zu der Diversität an Schulen. Im Folgenden werden diese Informationen in Bezug zu den Studienkosten gesetzt. Anschließend wird durch die letzte Visualisierung wieder auf den allgemeinen Kontext geschlossen und die Erzählung abgeschlossen. Anhand der Diagramme werden Dringlichkeit und Bedeutsamkeit des Problems aufgezeigt und im Fazit werden auf sprachlicher Ebene Lösungsansätze präsentiert. Insgesamt wurde innerhalb der Diagramme auf komplexere Visualisierungen, wie beispielsweise durch die Verwendung von Box-Plots, verzichtet. Die Diagramme sind möglichst intuitiv und einfach gehalten und haben zusätzlich die Option, sowohl einen kurzen Überblick zu erhalten als auch weitere Informationen zu entdecken. So wird sowohl das Verständnis und die Aufmerksamkeit der weniger gebildeten Rezipienten gesichert, als auch mehr Informationen für die gebildeteren oder besonders interessierten Rezipienten zur Verfügung gestellt. Ein Beispiel hierfür ist das Diagramm zu den Mittelwerten der Studiengebühren in verschiedenen US-Staaten. Man erkennt recht einfach, wo der Großteil der Mittelwerte liegt und dass es starke Schwankungen innerhalb der Staaten gibt.

Es gibt aber außerdem die Option, für die einzelnen Staaten zu schauen, wo sich der jeweilige Mittelwert genau befindet und was die Preisspanne ist.

### **Erklärung der Wahl der Channels und Marks bei den einzelnen Visualisierungen**

Für die erste Visualisierung wurde ein horizontales (Channel: Position) Balkendiagramm (Mark: Area) gewählt, um die Werte von A im Gesamten und nach Gruppen vergleichbar (Channel: Length) und die Kategorie-Bezeichnungen gut lesbar zu gestalten. Die weitere Aufteilung der Balken in die einzelnen US-Staaten (Channel: Hue) ermöglicht sowohl einen Überblick über die gesammelten Daten als auch präzisere Details. Um die Vergleichbarkeit zu steigern und einen größeren Effekt bezüglich des Visualisierungsziels sicherzustellen, wurden die Werte der Größe nach sortiert und der höchste Wert („White“) an oberster Stelle aufgezählt. Die Werte nehmen in Leserichtung ab und so wird, vor allem im direkten Vergleich zum höchsten Wert, ein Abstieg und somit geringe Diversität suggeriert. Die Farbe dient als Codierung für die Gruppen. Die Farbgebung würde man zwar zunächst für mindestens ordinale bis Verhältnis-skalierte anstelle von nominalen Werten erwarten, jedoch ist eine differenzierte Farbwahl aufgrund der hohen Anzahl an US-Staaten schwierig und würde die Harmonie der Visualisierung stören. Als erste Visualisierung ist die Ästhetik jedoch besonders wichtig, um das Interesse einer breiten Gruppe von Rezipienten zu wecken. Der negative Effekt der schweren Differenzierung wird außerdem durch die interaktive Funktion der Visualisierung aufgehoben, die es ermöglicht, die Werte für einzelne Staaten durch einen einfachen Doppelklick isoliert anzuzeigen. Diese Funktion soll außerdem durch die Ansprache individueller Interessen des Rezipienten und Interaktion Aufmerksamkeit erregen, eine stärkere Identifizierung mit dem Thema ermöglichen und weiteres Leseinteresse wecken.

Für die zweite Visualisierung (Diversität an Schulen) wurde ebenfalls ein Balkendiagramm (Mark: Area) gewählt, um die Werte im Gesamten und nach Gruppen vergleichbar (Channel: Length) zu machen.

Die Visualisierungen der Studienkosten (2/4 Jahre) verwenden die gleichen Marks und Channels für mehr Vergleichbarkeit. Dabei haben wir uns für ein Balkendiagramm (Mark: Lines, Channel: Length, Position (vertical)) entschieden, um nicht nur die Werte zwischen den Diagrammen vergleichbar zu machen, sondern auch die Gruppen innerhalb der Diagramme. Diese Gruppen wurden jeweils durch Farbe kodiert (Channel: Hue). Die Werte wurden außerdem bewusst nicht nach Größe geordnet, damit man die Staaten aufgrund ihrer alphabetischen Reihenfolge schneller ausfindig machen kann.

Die Visualisierung der Mittelwerte der Studienkosten ist ein Liniendiagramm (Mark: Lines, Channel: Length, Position (vertical)). Die Mittelwerte werden als Punkte dargestellt (Mark: Points). Zweck des Diagramms ist, dass man anhand der Linien sehen kann wie die Kosten innerhalb eines Staates variieren können, da man den höchsten und niedrigsten Wert ablesen kann. Zudem sieht man, ob der Mittelwert durch eventuelle Ausreißer beeinflusst wurde oder nicht.

Für die Visualisierung der Studienkosten über Zeit wurde ein Punktediagramm (Mark: Points, Channel: Position (both)) gewählt, um anzuzeigen, in welchen Abständen die Studienkosten gestiegen sind. Die Werte steigen von links nach rechts, um sowohl einen zeitlichen Verlauf, als auch einen Anstieg zu fokussieren. Die verschiedenen Gruppen werden durch die Farbkodierung (Channel: Hue) gekennzeichnet.

Für die letzte Visualisierung wurde ein Flächendiagramm gewählt (Mark: Area, Channel: Position (vertical)), um sowohl die Entwicklung über eine Zeitspanne hinweg als auch die jeweiligen Anteile der Gruppen darzustellen. Die Farbe dient als Kodierung der Gruppen (Channel: Hue).

Das Farbkonzept innerhalb aller Visualisierungen (Channel: Hue) orientiert sich an einigen Farben der Regenbogenflagge. Besonders bekannt als Symbol für die LGBTQ+ Szene wird diese aufgrund fehlender Copyright-Regelungen weltweit auch in vielen weiteren Diversitäts-Kontexten genutzt und gilt als Zeichen für Toleranz, Akzeptanz und Vielfalt. Auch ihre Entwicklung durch das Hinzufügen von schwarzen und braunen Farbbalken spiegelt das Thema des Projektes wider. Der Kontrast zwischen den Farben ist groß genug, um nominale Variablen deutlich voneinander abzugrenzen. Die Farbe Rot, die ebenfalls zur Regenbogenflagge gehört, wurde bewusst ausgelassen, um negative Assoziationen zu vermeiden. Einzige Ausnahme stellt die erste Visualisierung dar, die rötliche Farbtöne verwendet und die einzelnen Farben aufgrund der hohen Anzahl an Staaten und ihrer alphabetischen Nähe zueinander nicht so klar voneinander getrennt sind wie bei den restlichen Visualisierungen. Die Farben wurden so ausgewählt, dass sie außerdem zum gesamten Farbkonzept der Webseite passen und sich so ohne Störung integrieren.

## Literatur, Quellen, Lizenzen

### Datensätze:

- [https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/tuition\\_cost.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/tuition_cost.csv)
- [https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/historical\\_tuition.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/historical_tuition.csv)
- [https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/diversity\\_school.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-03-10/diversity_school.csv)
- <https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-06-09/firsts.csv>
- [https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2019/2019-09-24/school\\_diversity.csv](https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2019/2019-09-24/school_diversity.csv)

### Literatur:

An ausgewählten Stellen der Internetseite wird das vermittelte Wissen auch in Aussagen aus der folgenden Literatur gestützt. Diese Links wurden unter dem Reiter "Listing of sources and literature" auch auf der Webseite eingefügt. Außerdem wurde weitere Literatur zum Erstellen dieses Dokuments verwendet. Für die Erstellung des Kapitels zu den Channels und Marks und dem Design wurden außerdem zwei Vorlesungseinheiten von Dr. Calero Valdez (2020) verwendet. Zum Story-Telling wurden Informationen von Prof. Schröder (Zuyd Hogeschool) in einer weiteren Vorlesungseinheit mit einbezogen.

- Allen, Drew; Wolniak, Gregory C. (2015): Exploring the Effects of Tuition Increases on Racial/Ethnic Diversity at Public Colleges and Universities. URL: <http://s18.middlebury.edu/ECON0210B/STUDENTS/Group%20Research%20Proposal/Garrett,%20Greene/PS/PS15/Exploring%20the%20Effects%20of%20Tuition%20Increases%20on%20Racial-Ethnic%20Diversity%20at.pdf> (letzter Zugriff am 13.08.2020)
- Möller Digital (2020): Amerikanisches Schulsystem. Was sind die Besonderheiten im Schulsystem der USA? URL: <https://www.usa-info.net/usa-wiki/amerikanisches-schulsystem/#amerikanisches-schulsystem-8211-die-verschiedenen-schulformen-der-usa> (letzter Zugriff am 12.08.2020)
- Möller, Torsten (2015): Visualization. URL: [http://vda.univie.ac.at/Teaching/Vis/15s/LectureNotes/01\\_intro.pdf](http://vda.univie.ac.at/Teaching/Vis/15s/LectureNotes/01_intro.pdf) (letzter Zugriff am 12.08.2020)
- laenderdaten.info (2020): Durchschnittliches Einkommen weltweit. URL: <https://www.laenderdaten.info/durchschnittseinkommen.php> (letzter Zugriff am 20.08.2020)

- Lex, Alexander (2014): Visualization and Design. The Visualization Alphabet: Marks and Channels. URL: [http://comp.eng.ankara.edu.tr/files/2019/10/03-marks\\_channels.pdf](http://comp.eng.ankara.edu.tr/files/2019/10/03-marks_channels.pdf) (letzter Zugriff am 12.08.2020)
- Paulsen, Michael B.; St. John, Edward P. (2015): Social Class and College Costs. Examining the Financial Nexus Between College Choice and Persistence. URL: [https://www.researchgate.net/profile/Edward\\_St\\_John/publication/200031231\\_Social\\_Class\\_and\\_College\\_Costs\\_Examining\\_the\\_Financial\\_Nexus\\_Between\\_College\\_Choice\\_and\\_Persistence/links/5564e38608ae06101abdfec3.pdf](https://www.researchgate.net/profile/Edward_St_John/publication/200031231_Social_Class_and_College_Costs_Examining_the_Financial_Nexus_Between_College_Choice_and_Persistence/links/5564e38608ae06101abdfec3.pdf) (letzter Zugriff am 15.08.2020)

Die verwendeten Bilder auf der Webseite wurden selbst erstellt via canva.com und PowerPoint.