

Dokumentation Studienarbeit

Informationsvisualisierung SoSe 2022 – Juliana Kühn

Data Transformation:

Nach dem Laden der Excel-Tabellen und dem Einbinden verschiedener Bibliotheken werden zunächst paar Eingriffe gemacht um die erste leere Zeile zu löschen sowie leere Einträge in den Spalten „Jahr“ und „Orte“ aufzufüllen.

Um mir einen ersten Überblick über die Informationen zu beschaffen und zur Erleichterung welche Merkmale ich brauche, habe ich der Tabelle zwei neue Spalten hinzugefügt, die jeweils einmal die Anzahl der Medaillen pro Sportart und einmal die Anzahl der Medaillen pro Bezirk beinhalten. Mit diesen Spalten habe ich zwei Visualisierungen ausprobiert. Zum ersten ein Liniendiagramm, welches zeigt, wie viele Medaillen in der jeweiligen Sportart gewonnen wurde. Zum zweiten habe ich ein Balkendiagramm ausprobiert, das auch anzeigt, aus welchen Bezirken wie viele Medaillen kommen.

Nach dem ersten Einblick, habe ich eine neue Tabelle erstellt, diese aber nur die Merkmale beinhalten soll, die ich für mich am wichtigsten empfinde. Das wären die Merkmale „Jahr“, „Orte“, „Bezirk“, „Geburtsort“, „Medaille“, „Sportart“ und „Disziplin“.

Nachdem das Sonderzeichen geändert wurde, kann man mit der Methode „str.split“ zwei Wörter, die sich in einem Eintrag befinden, trennen und jeweils eins davon nutzen. Da in der Spalte Medaille nicht nur immer „Gold“, „Silber“ oder „Bronze“ steht, sondern auch mal eine Zahl davor oder andere, nicht relevante Informationen, habe ich ein Array mit den benötigten Wörtern erstellt. Mit diesem Array konnte ich dann im nächsten Schritt mit „str.extract“ genau diese Wörter, die im Array stehen, herausfiltern. Nun enthält die Spalte „Medaille“ nur die Wörter, die wichtig sind.

Als nächstes erstelle ich für jeden Bezirk mit der Methode „.loc“, das alle Zeilen mit bestimmten Inhalt nimmt, eine Tabelle und sortiere die Werte nach der Spalte „Jahr“.

Daraufhin wird für jeden Bezirk eine Tabelle erstellt, diese werden alle miteinander zu einer neuen finalen Tabelle verknüpft. Anschließend werden neue Variablen für die Spalten „Latitude“, „Longitude“ und „Bezirk“ aus der Längen- und Breitengrad Tabelle erstellt. Diese Spalten werden danach in die neue Tabelle mit eingefügt, damit man für jeden Geburtsort die Koordinaten besitzt. Die finale Tabelle ist nun vollständig und kann jetzt als CSV Datei exportiert werden.

Visual Mapping:

1. Für das Merkmal „Bezirk“ wird eine Karte von Bayern verwendet. Somit bekommt man einen Überblick welche Geburtsorte in welchem Bezirk liegen, zudem weiß man wie viele Medaillen in welchen Bezirken gewonnen wurden.
2. Das Merkmal „Geburtsort“ wird als Kreis mit entsprechenden Längen- und Breitengrad Koordinaten auf der Landkarte im jeweiligen Bezirk dargestellt.

In Kombination mit der Karte der Bezirke, weiß man wo sich die Geburtsorte in Bayern befinden.

3. Gleichzeitig wird mit dem Kreis auch das Merkmal „Medaille“ in der jeweiligen Medaillen Farbe dargestellt (Gold, Silber, Bronze). Gab es aus einem Ort z.B. eine Gold- und eine Silbermedaille, werden zwei Kreise dementsprechend gezeichnet. Die farbigen Kreise sind perfekt dafür geeignet um auf einem Blick zu erkennen, was für eine Art von Medaille in welchem Geburtsort in welchem Bezirk gewonnen wurde.
4. Das Merkmal „Sportart“ wird per Mausklick als Text auf der Seite dargestellt. Es werden alle Sportarten, für den jeweiligen Bezirk angezeigt. Somit hat man mit den farbigen Kreisen eine gute Übersicht, in welcher Sportart die Medaillen gewonnen wurden.

Implementierung:

1. Eine Karte von Bayern wird als Bild für den Hintergrund eingefügt. Auf dieser Karte sind die Bezirke farblich gekennzeichnet und auch die Grenzen sind gut sichtbar. Namen der Bezirke wurden als schlichter Text separat eingefügt.
2. In der Funktion draw() wird jede Zeile der Tabelle mithilfe der For-Schleife durchgegangen und die Einträge der jeweiligen Zeilen in Variablen gespeichert. Dafür wurde bereits vorher eine Funktion definiert, die dann für jeden Bezirk individuell verwendet werden kann. In der Funktion werden für jeden Bezirk (das per if-Anweisung geprüft wird) nun die Kreise für die Geburtsorte gezeichnet, dafür werden die Längen- und Breitengrade (die in den Variablen long und lat gespeichert sind) mit der Methode „map“ in den Variablen x und y gespeichert, damit diese exakt auf der Karte von Bayern dargestellt werden.
3. Da die Kreise jedoch auch gleichzeitig die jeweiligen Medaillen darstellen, wird innerhalb der if-Bezirk-Prüfung nochmal per if-Anweisung geprüft, ob es sich bei der jeweiligen Zeile um einen Gold-, Silber- oder Bronzemedaille Eintrag handelt. Daraufhin wird ein Kreis mit den entsprechenden Koordinaten der Geburtsorte und mit der passenden Farbe für die betroffene Medaille gezeichnet. Goldmedaillen fallen etwas größer aus, da die Silber- und Bronzemedallien in den Goldmedaillen mit reingezeichnet werden um den Überblick zu behalten.

Nachdem die Geburtsort- und Medaillenkreise gezeichnet wurden, wird wieder in der if-Bezirk-Prüfung per if-Anweisung geprüft, ob sich der Mauszeiger innerhalb vom Kreis mit einem Abstand von 10px befindet. Wenn das der Fall ist, wird ein Textfeld bei der aktuellen Position über den Mauszeiger erstellt, welches den Namen vom jeweiligen Geburtsort anzeigt.

4. Damit die Sportarten angezeigt werden, muss zuerst die Funktion „mousePressed“ implementiert werden, damit bei einem Mausklick der boolean von der Variablen „clicked“ von false auf true geändert wird und bei erneutem Klick von true wieder auf false. Dann wird in der Funktion drawSport() geprüft, ob sich der Mauszeiger innerhalb bestimmter x- und y-Koordinaten, die die Grenze der Bezirke bilden,

befindet und wenn das der Fall ist, werden bei einem Mausklick die Sportarten der jeweiligen Bezirke an der Seite der Map dargestellt.

Satz von Tufte:

„Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.“

In meiner Visualisierung soll der Betrachter auf einem Blick sehen, aus welchen bayerischen Bezirken und Orten die errungenen Medaillen - egal ob Gold, Silber oder Bronze - stammen und mit nur einem Mausklick sieht er auch sofort in welchen Sportarten diese geholt wurden.

Alles in einem bekommt der Betrachter all diese Informationen innerhalb kürzester Zeit ohne den Überblick zu verlieren.

Bedienungsanleitung:

Folgende Interaktionen sind vorhanden: Mouseover, Mouseclick

Mouseover: mit dem Mauszeiger über die Medaillen gehen um sich die Namen der Geburtsorte anzeigen zu lassen

Mouseclick: mit der Maus auf den Bezirk-Bereich klicken, um sich die Sportarten, in denen die Medaillen aus diesem Bezirk gewonnen wurden, anzeigen zu lassen