Visualisierung:

- Kartenbasiert

- Job Angebote als Datenset

- Interaktion zwischen Karte und anderen Visualisierungen

- Python 3 als Sprache

- Frameworks: Altair, Geopandas

- Dokumentation unseres Projektes in dieser Datei

- Source Code und Projekt in Form eines interaktiven Jupyter Notebooks

**Karte von Stellenangeboten in den USA**

Für unser Projekt planten wir zunächst mit der github jobs API (<https://jobs.github.com/api> ) eine Stellenmarktkarte zu erstellen. Wir mussten leider schnell feststellen, dass die API für diesen Zweck/den Umfang des Projekts ungeeignet ist, da mehr Arbeitsaufwand in das Crawling gesteckt werden müsste als in die Visualisierung.

Wir entschlossen uns deshalb einen anderen aber ähnlichen Datensatz zu verwenden: Ein Auszug aus dem Monster.com Jobangebot Datensatz. Dieser Auszug enthält 22.000 Jobangebote in den USA. (<https://www.kaggle.com/PromptCloudHQ/us-jobs-on-monstercom#monster_com-job_sample.csv>)

Ziel ist es, die Erkundung des Arbeitsmarktes auf einer interaktiven Karte zu erleichtern. Auf diese Weise erhalten Sie einen Überblick über die verschiedenen Faktoren, die für die Erkundung des Datensatzes interessant sind, so wie Anzahl der Jobangebote in den verschiedenen Bundesstaaten, Gehälter und Job-Sektoren.

**What, Why, How**

* What: Jobangebote in den USA
* Why: Die Erkundung des Jobmarktes und seiner Differenzen in den verschiedenen Bundesstaaten
* How: Nutzung geographischer Daten auf Karte, Mappen der Werte auf Farben, bzw. Encoden der Werte als Position (Scatterplot) und Mappen der Zugehörigkeit als Farbe (Legende).

Juxtaposition zwischen verschiedenen Daten (verschiedene Scatterplots mit Brushing & Linking), bzw. zwischen gleichen Daten und verschiedenen Darstellungen (Scatterplot und Choropleth-Karte).

**Überlegungen**

Der Datensatz enthält zwar eine vielzahl an Positionsdaten, allerdings nur in der Form eines von den Nutzern frei eingetragenen ‘location’ Feldes, welches die Platzierung der Daten auf einer Karte erschwert. Longitute und Latitude sind nicht gegeben, aber in vielen Fällen der Name des Bundesstaates. Da dies ein Visualisierungs- und kein Datenscience-Projekt ist, haben wir uns entschieden, die Aufbereitung des Datensatzes so einfach wie möglich zu halten.

Hierfür haben wir die ‘location’ auf Nennung der Bundesstaaten (bzw. deren Abkürzungen) gefiltert, was die Projektion der Daten auf eine Choropleth-Karte erlaubt.

Die Darstellung auf der Karte macht am meisten Sinn für Daten in Relation zur Bevölkerungszahl der Bundesstaaten (hiermit vermeiden wir auch Fehler die durch Verwendung von Daten in absoluten Werten auftreten würden) - hierfür haben wir den Datensatz kollabiert (alle Felder eines Bundesstaats jeweils aufsummiert, gezählt oder Durchschnitt berechnet).

Für die Darstellung des nicht kollabierten Datensatzes eignet sich die Karte hingegen nicht/nur bedingt, da die ‘location’ eben nur auf Bundesstaaten beschränkt ist, d.h. keine interessanten Beobachtungen über geographische Verteilung im Detail gemacht werden können. Besser eignen sich hierfür Scatterplots, welche man per Brush-and-link an die Karte koppelt.

Dies erlaubt sowohl Juxtaposition (Die Gegenüberstellung der Daten in zwei verschiedenen Ansichten, bzw. verschiedener Daten in gleichen Ansichten), sowie die Reduktion des Datensatzes (Einschränkung auf Daten des aktuell selektierten Bundesstaates).

Wir erstellen zudem eine verlinktes Diagramm, welches das hineinzoomen in einen Teilbereich der Daten erlaubt.

Zu guter letzt erstellen wir eine Map-Ansicht, bei der die Auswahl eines Staates alle Daten Highlighted, welche einen höheren (oder gleichen) Wert haben wie der aktuell ausgewählte Bundesstaat.

**Dependencies:**

Altair

Vega

Pandas