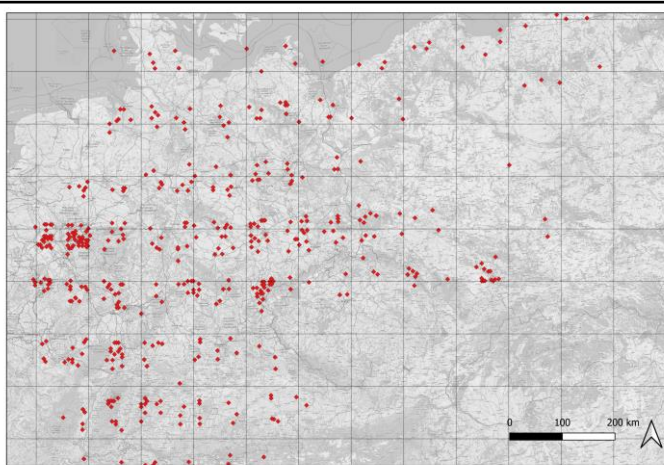


Untersuchungen zur Clusterung der digitalisierten Locations des Bombers-Baedeker

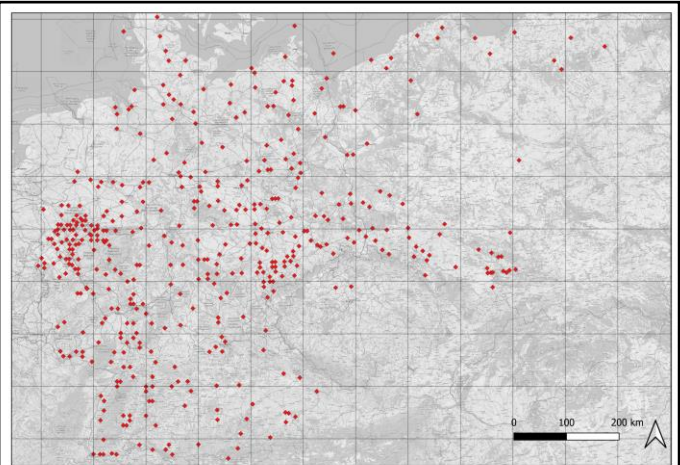
Diese Ausarbeitung basiert auf dem Projekt *Bomber's Baedeker – vom Text zum Bild zur Datenquelle* des IEG Mainz. Hierbei wurde das zweibändige Druckwerk *The Bomber's Baedeker. A Guide to the Economic Importance of German Towns and Cities* digitalisiert, via OCR maschinenlesbar gemacht und in einen XML Datensatz überführt.

Bei der Kartierung ist eine Clusterung der Orte/Locations aufgefallen.

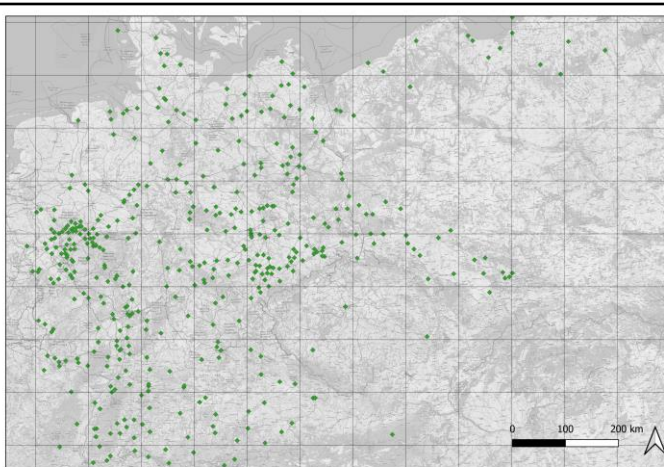
Diese Ausarbeitung untersucht mögliche Ursachen und bietet mögliche Lösungsansätze



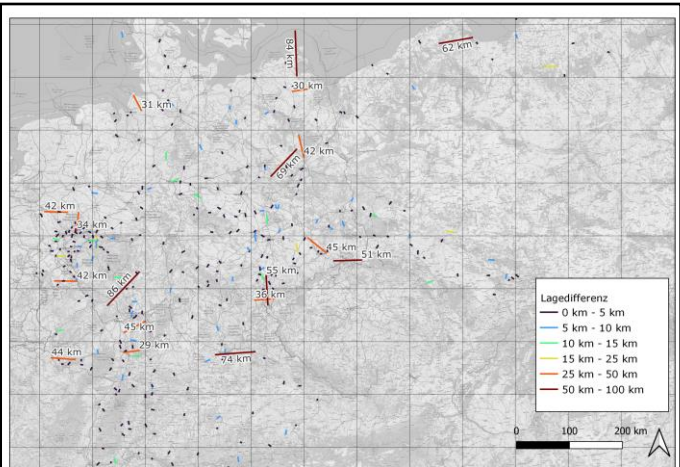
Ausgangspunkte der Untersuchung. Eine Clusterung ist zu erkennen. Zu erklären dadurch dass in der Druckausgabe des *Bomber Baedeker's* die Koordinaten in GMS-Format angegeben wurden, diese jedoch im durch die Texterkennung als Dezimalgrad übernommen wurden.



Fehlerkorrigierte Punkte der Digitalisierung. Die Clusterung/Häufung der Punkte ist nicht eindeutig anzusprechen.



Karte der in *Bombers Baedeker* genannten Orte mit Ortsdaten aus *OpenStreetMap*. Diese wurden durch Geocodierung innerhalb von QGIS und Abgleich mit der OSM Datenbank gewonnen.



Differenzlinien zwischen den korrigierten Koordinaten des *Bomber Baedeker's* und den Geocodierten Koordinaten von *OpenStreetMap*. Zu erkennen dass die Abweichung, bis auf wenige Ausnahmen, unter 15km liegt.