

Kreis Unna – Automatisiert erstellter Grundstücksmarktbericht mit Rstudio

Aufbau und Datengrundlage

Statistische Auswertungen und Dokumentationen von Erhebungen (auch mit Raumbezug) sind aufwendige, fehleranfällige und intensive Tätigkeiten in der Verwaltung.

Häufig wird viel Arbeit in das Visualisieren von Zahlenkolonnen, die Interpretation und die Formulierung von anschaulichen Texten investiert. Sachbearbeitende sind dabei Zahlenkünstler, Schriftsteller und sogar Designer zugleich, wenn sie ziel- und nutzerorientierte Diagramme und Layouts erstellen. Dabei können regelmäßig wiederkehrende Auswertungen und insbesondere das Druck-Layout im Zeitalter der IT vereinfacht oder vollständig automatisiert werden.

Alljährlich wird durch die Geschäftsstelle des Gutachterausschusses für Grundstückswerte im Kreis Unna der Grundstücksmarktbericht für seinen Zuständigkeitsbereich (ohne die Städte Unna und Lünen) erstellt. Dabei handelt es sich um ein Dokument, das zahlreiche statistische Auswertungen für den jeweiligen Berichtszeitraum, der ein Kalenderjahr umfasst, enthält. Sein Ziel ist es, Umsatz- und Preisentwicklungen darzustellen und über das Preisniveau auf dem Grundstücksmarkt zu informieren sowie die für die Wertermittlung erforderlichen Daten bereitzustellen. Grundlage dafür ist die Zentrale Kaufpreissammlung (ZKPS), in der die Daten der ausgewerteten Kaufverträge aller Grundstücksverkäufe im Zuständigkeitsbereich anonymisiert geführt werden. Zielgruppe sind – neben Sachverständigen für Immobilienbewertung – die freie Wirtschaft, die öffentliche Verwaltung mit Bedarf nach Kenntnissen über den Grundstücksmarkt sowie interessierte Bürgerinnen und Bürger. Der Grundstücksmarktbericht ist ein wesentlicher Beitrag zur Transparenz auf dem Markt.

Ergänzt wird er unterjährig durch einen Trendbericht nach dem ersten Halbjahr, der ähnliche Informationen in reduzierter Anzahl basierend auf den bisher im jeweiligen Kalenderjahr vorliegenden Daten präsentiert.

Der Grundstücksmarktbericht umfasst auf etwa 60 Seiten ungefähr 40 Tabellen, 30 Diagramme und ausführliche Erläuterungen, die in Abhängigkeit der Daten der ZKPS bzw. der Entwicklung der Daten (z. B. Abwärts- oder Aufwärtstrend) visualisiert bzw. interpretiert werden. Die Veröffentlichung erfolgt in der Regel gegen Ende Februar eines Jahres.

Der Trendbericht hat einen weniger offiziellen Charakter und umfasst demgegenüber einen bewusst reduzierten Inhalt. Er wird üblicherweise am Anfang des dritten Quartals veröffentlicht und stellt Entwicklungen und Tendenzen des Jahres dar. Er hat etwa ein Dutzend Diagramme, die überwiegend Sachwertfaktoren für die im Zuständigkeitsbereich liegenden Kommunen abbilden, wenige Tabellen und im Vergleich zum Grundstücksmarktbericht stark begrenzte Beschreibungen. Dies geht einher damit, die Entwicklungen des Grundstücksmarktes ohne den gleichen zeitlichen und personellen Aufwand wie beim Grundstücksmarktbericht darzustellen.

Beide Berichte besitzen die üblichen formalen Elemente wie einen Titel, ein Inhaltsverzeichnis, eine Einführung sowie Informationen zu Kontakten zur Geschäftsstelle, zu anderen Geschäftsstellen und Gutachtern und das Impressum.

Die Auswertung der Daten aus der ZKPS, die Visualisierung von Diagrammen und Tabellen erfolgte auf klassischer Weise mithilfe gängiger Tabellenkalkulations-Software. Auch das Layout wurde mit üblicher Textverarbeitung erstellt.

Kartographische Darstellungen spielten bisher keine große Rolle, obwohl es sich anbietet, statistische Informationen je Raumeinheit auch in Choroplethenkarten oder Kartogrammen anschaulich zu visualisieren und eine Verteilung innerhalb eines administrativen Gebietes nachvollziehen zu können.

Im Jahr 2022 wurde damit begonnen, die Auswertung der Daten zu automatisieren. Grundlage dazu war die Open-Source-Software RStudio bzw. zukünftig Posit.

RStudio ist eine Entwicklungsumgebung für die Programmiersprache R (bzw. für Textbausteine LaTeX). In RStudio kann ein R-Markdown-Skript entwickelt werden, welches reproduzierbare Ergebnisse in verschiedenen Formaten wie PDF, Word-Dokument oder auch HTML exportieren kann. Dazu werden beschreibender Text in Form von Textbausteinen und R-Code zusammengefügt. So können verschiedene Datenquellen wie bereits vorliegende Tabellen, Datenbanken o. ä. angebunden werden, auf deren Grundlage dann die Statistischen Auswertungen und die Zusammenfassung in einem Bericht erfolgen. Dabei werden Diagramme, Grafiken oder Tabellen sowie der gesamte Seitenaufbau gestaltet. Verändern sich die Quelldaten, passen sich diese dann beim Neuausführen dieses erstellten Scripts an.

Es ist zwar zunächst Zeit und Arbeit aufzuwenden, das notwendige Skript zu entwickeln. Dazu sind intensive Beschäftigungen mit dem Programmaufbau und den verwendeten Programmiersprachen notwendig. Dieser Aufwand wird jedoch spätestens beim ersten Datenexport im Folgejahr ausgeglichen, wenn zu diesem Zeitpunkt keine oder nur geringfügige Nachbesserungen zu erfolgen haben.

Bei der Entwicklung des Skripts wird zunächst der Nachbau des bestehenden Grundstücksmarktberichts angestrebt. Bei dem Nachbau werden vereinzelt Grafiken verändert.

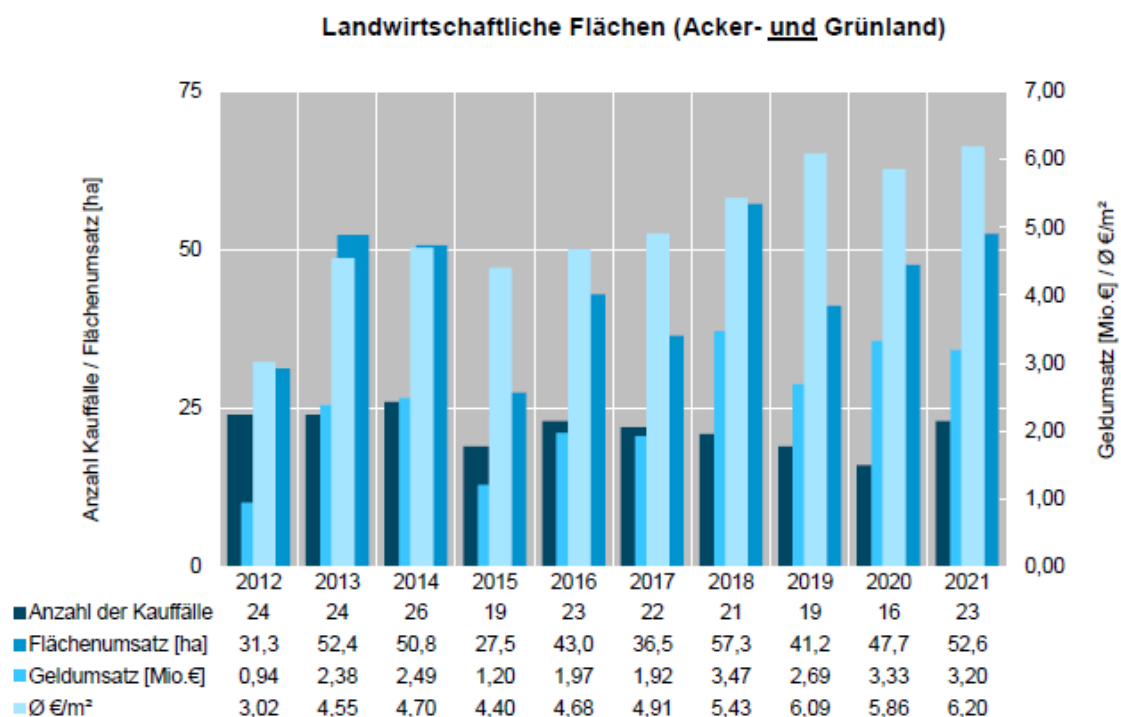


Abbildung 1: Landwirtschaftliche Flächen, ursprüngliche Grafik (Quelle: Kreis Unna)

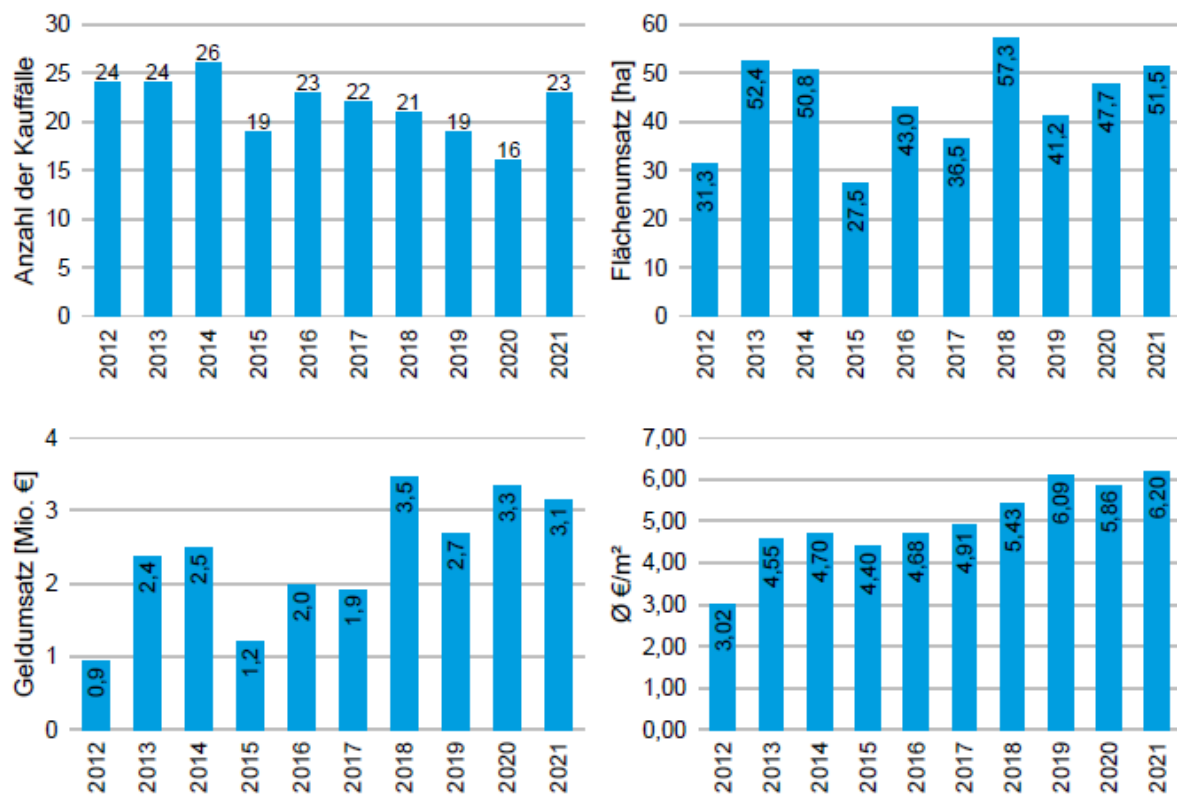


Abbildung 2: Landwirtschaftliche Flächen, neue Darstellung (Quelle: Kreis Unna)

An anderer Stelle werden Tabellen zu Grafiken umgewandelt, um auf den ersten Blick eine Verteilung der Werte zwischen verschiedenen Gemeinden treffen zu können.

Städte und Gemeinden	Kaufpreise für Wohnbauflächen in Euro/m²			Anzahl
	von	bis	Ø	
Bergkamen	130	319	239	20
Bönen	91	177	137	7
Fröndenberg/Ruhr	102	261	176	14
Holzwickede	190	576	306	5
Kamen	110	430	257	28
Schwerte	139	419	270	10
Selm	197	275	231	5
Werne	193	250	244	20
Kreis Unna	91	576	235	109

Abbildung3: Kaufpreise für Wohnbauflächen, ursprünglich als Tabelle dargestellt (Quelle: Kreis Unna)

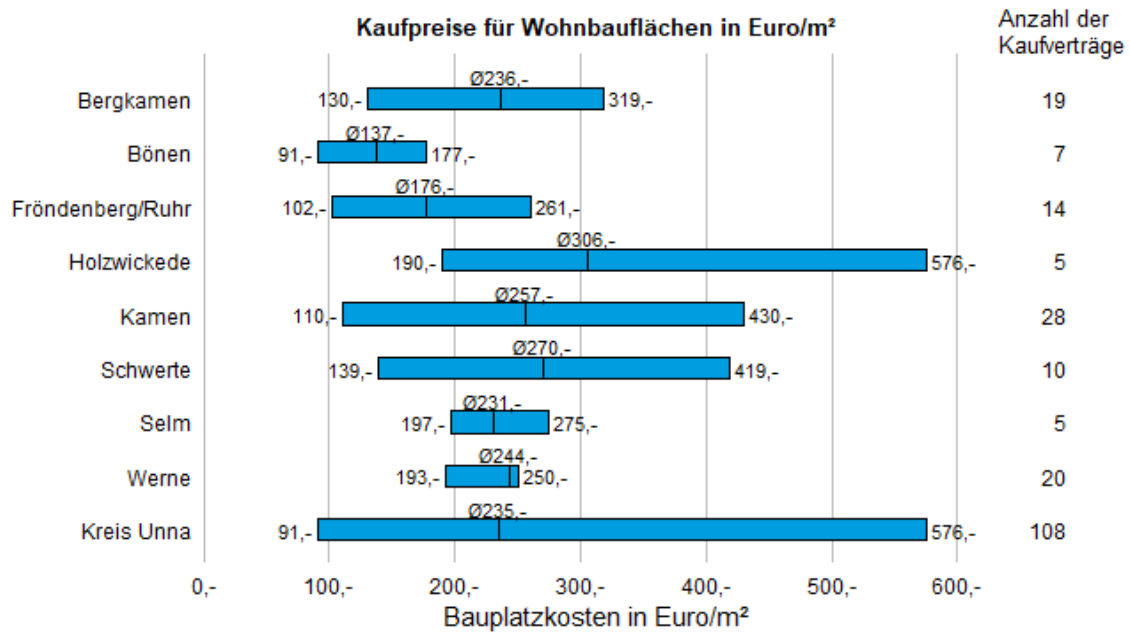


Abbildung 4: Kaufpreise für Wohnbauflächen, neue Darstellung (Quelle: Kreis Unna)

```

694 # teilmärkte auswählen
695 Grafik_Bauplaetze %%%
696 ggplot(aes(x = as.numeric(l_gemeinde), y = Bp1_Kosten_m2)) +
697   coord_flip() +
698   # Titel
699   ggtitle("Kaufpreise für Wohnbauflächen in Euro/m²") +
700   # Beschriftung der Achsen
701   ylab("Bauplatzkosten in Euro/m²") +
702   xlab("") +
703   # Balkendiagramm
704   geom_crossbar(aes(ymin = Min_Kosten_m2,
705                     ymax = Max_Kosten_m2),
706                 colour = "black", size = 0.2, width = 0.4, fill = params$Farbe1) +
707   # Beschriftung Min Max Mittel
708   geom_text(aes(y = Min_Kosten_m2,
709                 label = paste0(round(Min_Kosten_m2, 0), "-"),
710                 size = 3.0, hjust = 1) +
711   geom_text(aes(y = Max_Kosten_m2,
712                 label = paste0(" ", round(Max_Kosten_m2, 0), "-"),
713                 size = 3.0, hjust = 0) +
714   geom_text(aes(y = Bp1_Kosten_m2,
715                 label = paste0("Ø", round(Bp1_Kosten_m2, 0), "-"),
716                 size = 3.0, hjust = 0.5, nudge_x = 0.4) +
717   # Diskrete Werte als Zahl dargestellt, um eine zweite x-Achse erstellen zu können
718   scale_x_reverse(breaks = 1:length(Grafik_Bauplaetze$l_gemeinde),
719                 labels = Grafik_Bauplaetze$l_gemeinde,
720                 sec.axis = sec_axis(., name = "Anzahl der Kaufverträge",
721                                   breaks = 1:length(Grafik_Bauplaetze$Anzahl),
722                                   labels = Grafik_Bauplaetze$Anzahl)
723   ) +
724   scale_y_continuous(breaks = Y_Breaks,
725                     expand = c(0,0),
726                     # Beschriftung ändern
727                     labels = label_number(suffix = "-"),
728                     # Restdivision (wenn zur letzten einteilung weniger als 30 dann + 25, für Beschriftung überhalb des Diagramms)
729                     limits = c(0, ifelse(MaxBaupl % 100 > 70 | MaxBaupl % 100 == 0, max(Y_Breaks) + 25, max(Y_Breaks)))
730   ) +
731   # vordefiniertes Thema wählen
732   Grafik_Thema_Flip +
733   theme(plot.margin = unit(c(1,1,1,0), "cm"))

```

Abbildung 5: R-Code für die in Abbildung 4 dargestellte Grafik (Quelle: Kreis Unna)

Texte werden mit dem Textsatzsystem LaTeX erstellt. So können Textbausteine mit passender Formatierung in das Skript eingebunden werden.

```

11 Die zentrale Aufgabe der Gutachterausschüsse ist es, für Transparenz auf dem Immobilienmarkt zu sorgen. Somit gehören zu ihren wesentlichen Aufgaben:
12 \vspace[-\topsep]
13 \begin{itemize}[nosep]
14 \item die Führung und Auswertung der Kaufpreissammlung
15 \item die Ermittlung der für die Wertermittlung erforderlichen Daten
16 \item die Ermittlung und Veröffentlichung von Boden- und Immobilienrichtwerten
17 \item die Veröffentlichung eines Grundstücksmarktberichts
18 \item die Erstattung von Gutachten über den Verkehrswert von Immobilien
19 \end{itemize}

```

Die zentrale Aufgabe der Gutachterausschüsse ist es, für Transparenz auf dem Immobilienmarkt zu sorgen.

Somit gehören zu ihren wesentlichen Aufgaben:

- die Führung und Auswertung der Kaufpreissammlung
- die Ermittlung der für die Wertermittlung erforderlichen Daten
- die Ermittlung und Veröffentlichung von Boden- und Immobilienrichtwerten
- die Veröffentlichung eines Grundstücksmarktberichts
- die Erstattung von Gutachten über den Verkehrswert von Immobilien

Abbildung 6: Beispiel aus einem LaTeX-Textbaustein mit Ergebnis (Quelle: Kreis Unna)

In die Textbausteine kann ebenfalls R-Code eingebaut werden, sodass Variablen direkt ausgefüllt werden und Text sich auf die verändernden Werte bezieht.

```

1 # Variablen für den Text
2 ""{r Gesamtumsatz Text}
3 AktuellesJahr <- param$JahrGMD-1
4 VorherigesJahr <- param$JahrGMD-2
5
6 AktuelleAnzahl <- kapitel_3 %>% filter(nrJahr==param$JahrGMD-1) %>% $Gesamtanzahl
7 VorherigeAnzahl <- kapitel_3 %>% filter(nrJahr==param$JahrGMD-2) %>% $Gesamtanzahl
8 Prozent <- round((AktuelleAnzahl-VorherigeAnzahl)/VorherigeAnzahl*100,digits = 0)
9
10 UmsatzNeu <- round(kapitel_3 %>% filter(nrJahr==param$JahrGMD-1) %>% $Gesamtumsatz/100000,digits = 0)
11 UmsatzAlt <- round(kapitel_3 %>% filter(nrJahr==param$JahrGMD-2) %>% $Gesamtumsatz/100000,digits = 0)
12 ""
13
14 # Text
15 Im Jahre {r AktuellesJahr} wurden im Zuständigkeitsbereich des Gutachterausschusses {r prettyNum(AktuelleAnzahl, big.mark = ",", decimal.mark = ",")}
16 Kauffälle ausgewertet. Dies ist gegenüber {r VorherigesJahr} {(r prettyNum(VorherigeAnzahl, big.mark = ",", decimal.mark = ","))} ein
17 {r ifelse(Prozent > 0, paste("Anstieg um", Prozent), paste("Rückgang um", -Prozent))}%. Der Umsatz ist von {r UmsatzAlt} Mio. EUR auf {r UmsatzNeu} Mio. EUR
18 im Jahr {r AktuellesJahr} {(r ifelse(UmsatzNeu > UmsatzAlt, "gestiegen", "gesunken"))}.
19
20
21 # Variablen entfernen
22 ""{r Gesamtumsatz Text en}
23 rm(AktuellesJahr, VorherigesJahr, AktuelleAnzahl, VorherigeAnzahl, Prozent, UmsatzAlt, UmsatzNeu)
24 ""

```

Im Jahre 2021 wurden im Zuständigkeitsbereich des Gutachterausschusses 2.776 Kauffälle ausgewertet. Dies ist gegenüber 2020 (2.825) ein Rückgang um 2%. Der Umsatz ist von 693 Mio. EUR auf 875 Mio. EUR im Jahr 2021 gestiegen.

Abbildung 7: Beispiel LaTeX-Textbaustein mit R-Code und Ergebnis (Quelle: Kreis Unna)

RStudio wird lokal auf dem Arbeitsplatz-PC installiert. Viele Funktionen, die beispielsweise für die Erstellung bestimmter Diagrammtypen, Tabellen-Formatierungen o. ä. aus dem zuvor beschriebenen Grundstücksmarktbericht benötigt werden, sind nur mithilfe konkreter sogenannter R-Pakete nutzbar, die basierend auf dem gewünschten Ergebnis recherchiert und installiert werden müssen. Falls der Download von Zip-Dateien durch eine interne Firewall blockiert wird, kann eine in RStudio integrierte Download-Funktion für diese Pakete nicht genutzt werden. Gleiches gilt für häufig wiederkehrende Aktualisierungen der R-Pakete. Für das Textsatzsystem gibt es eigene Pakete, die ebenfalls als Zip-Datei bereitgestellt werden. Im Vorfeld der Bearbeitung ist der Workflow der Erstinstallation und der stetigen Aktualisierung daher mit der zuständigen IT abzustimmen. Erfolgen die Updates der Pakete nicht, kommt es schlimmstenfalls dazu, dass ein Auswertungsprozess nicht abgeschlossen wird.

Die bearbeitende Person benötigt vor einem konkreten Projekt einen Einstieg in RStudio, die Programmiersprache R und Latex. Dies kann mithilfe von kommerziellen Einstiegskursen sowie Literatur geschehen, die zwar oftmals kostenfrei verfügbar, jedoch meist in englischer Sprache gehalten ist. Dabei sind insbesondere die Unterschiede zwischen RStudio (Umgebung), R (Programmiersprache), R-Markdown (Skript) und Latex (Textsatzsystem) nachzuvollziehen. Als Übung können zunächst einfache Daten berechnet und in Diagrammen dargestellt werden.

Eine wichtige, aber nicht selbstverständliche Voraussetzung für eine erfolgreiche Auswertung dieser Größenordnung ist eine einheitlich ausgefüllte, saubere Datenbank. Alle vorliegenden Daten sollten im Vorfeld intensiv auf ihre Einheitlichkeit, Vollständigkeit und Aktualität geprüft werden.

Ein Ziel der Umstellung der Erstellung des Grundstücksmarktberichts auf automatisierte Prozesse ist die anschauliche (räumliche) Präsentation der Daten in Form von Dashboards. Auf diesen können die regelmäßig aktualisierten Daten von Nutzenden schnell überschaut und visuell erfasst werden.

Der Grundstücksmarktbericht soll weiterhin jährlich erscheinen, allerdings wesentlich früher als bisher. Die Trendberichte sollen ab sofort quartalsweise erscheinen. Da die technischen Möglichkeiten nun existieren, wird der inhaltliche Umfang des Trendberichtes ausführlicher sein.

Neben dem Grundstücksmarktbericht und Trendbericht existieren in der Verwaltung weitere periodisch erstellte Produkte, die Auswertungen auf Basis von Datenbanken oder anderen statistischen Informationen darstellen. Es bietet sich an, mithilfe optimierter Programmierungen auch diese Erarbeitung zu automatisieren und periodisch wiederkehrende zeitaufwändige Auswertungen zu reduzieren.

Das Team des Geodatenmanagements profitierte in diesem Projekt von der fachlichen Nähe zu den Arbeitsthemen in der Geschäftsstelle des Gutachterausschusses und Erfahrungen im Scripting aus vergangenen Projekten unter dem Einsatz von GIS. Räumliche und organisatorisch kurze Dienstwege zwischen den Sachgebieten haben mit schnellen Absprachen und Informationsabfragen dazu beigetragen, das Projekt erfolgreich abzuschließen. Langfristig soll ein größerer Fokus auf kartographische Visualisierungen gelegt werden, für die in RStudio ebenfalls hinreichend R-Pakete beispielsweise zum Austausch mit der Software QGIS zur Verfügung stehen. Hier können wiederum interaktive Präsentationen mithilfe von Leaflet erzeugt werden – Methoden und Techniken, die im Bereich Geodatenmanagement geübte Praxis sind.