

Finale Präsentation - Hinweise

Datenanalyse und -visualisierung mit R
Hochschule Bonn-Rhein-Sieg
Wintersemester 2022-23

Pekka Sagner M.Sc.

Grundsätzliche Idee

Ihre Präsentation soll zeigen, dass Sie sich mit den Inhalten der Vorlesung vertraut gemacht haben. Sie sollen in einer eigenen Datenanalyse das Gelernte anwenden. Die finale Präsentation soll Sie darüber hinaus motivieren, sich weitere Konzepte der Datenanalyse und -visualisierung mit R, die über die Vorlesungsinhalte hinausgehen, anzueignen.

Termin

Der Termin für die finale Präsentation ist voraussichtlich der **13.01.2023**.

Inhalt der Präsentation

Folgendes müssen Sie für eine erfolgreiche Abschlusspräsentation tun:

1. Laden Sie sich einen Datensatz Ihrer Wahl herunter und untersuchen Sie diesen in R. Sie sollten die Wahl des Datensatzes begründen können. Der gewählte Datensatz sollte den Eignungskriterien (siehe unten) entsprechen.
2. Finden Sie eine **Story in den Daten**. Stellen Sie sicher, dass die Story auch wirklich stimmt. Eine gute Story ist spannend und enthält Neues, möglicherweise Unerwartetes.
3. Nutzen Sie R, um Abbildungen zu erstellen, mit denen Sie Ihre Story erzählen. Sie können so viele Abbildungen erstellen, wie Sie möchten. Ihre finale Präsentation dauert jedoch **zwischen 15 und 20 Minuten**. Limitieren Sie die Zahl der Abbildungen in Ihrer finalen Präsentation also entsprechend (*Empfehlung: vier bis sechs Abbildungen*). Sie sollten jedoch **mindestens drei verschiedene Abbildungstypen** in Ihre Präsentation aufnehmen.
4. Speichern Sie die Abbildungen für Ihre Präsentation in einem geeigneten Format, z.B. als png-, jpg- oder pdf-Datei. Erstellen Sie Ihre finale Präsentation mit einem Programm Ihrer Wahl, z. B. Microsoft Powerpoint.
5. Senden Sie Ihr Datenanalyseprojekt bis zum **11.01.2023** an »sagner@iwkoeln.de«. Idealerweise erstellen Sie einen zip-Ordner. Folgende Dateien sollten bereitgestellt werden:
 - Ihr R-Projekt-Ordner: enthält R-Skript zur Datenanalyse, die analysierten Daten und die erstellten Abbildungen.

- Ihre finale Präsentation als pdf-Datei. Diese kann auch im R-Projektordner gespeichert sein.
6. Präsentieren Sie Ihre spannende Datenanalyse. Der Fokus sollte auf der Story liegen, nicht auf den Daten selbst. Resümieren Sie Ihre Erfahrungen bei der Datenanalyse kurz. Zum Beispiel: Welche Konzepte mussten Sie sich zusätzlich aneignen? Welche Besonderheiten brachte die Arbeit mit dem Datensatz mit sich?

Nützliche Hinweise

Daten

Auswahl eines geeigneten Datensatzes

Ein Datensatz ist für Ihre Analyse geeignet, wenn er

1) **spannende Daten enthält**,

- Was macht einen Datensatz spannend? Sie finden ihn spannend! Nutzen Sie bei der Auswahl ihres Datensatzes die Freiheit, die Ihnen geboten wird. Sie können Sich mit einem Thema beschäftigen, dass **Sie** interessant finden. Die Wahl des Datensatzes kann aber beispielsweise auch auf Basis der Art der analysierten Daten gewählt werden. Nehmen Sie sich bei der Auswahl Ihres Datensatzes Zeit. Mehr unter **Mögliche Datenquellen**.

2) **ausreichend Informationsgehalt bietet**

- Ein Datensatz, der nur eine Variable enthält, genügt nicht, um eine informative Präsentation zu gestalten. Grundsätzlich gilt: je mehr Informationen ihr gewählter Datensatz enthält, desto mehr Möglichkeiten für spannende inhaltliche Schlüsse, Visualisierungen und Analyseschritte bieten sich.

3) und ein **Mindestmaß an Komplexität** aufweist.

- Dieses Eignungskriterium hängt direkt mit dem obigen zusammen. Hier ist nicht gemeint, dass die Daten besonders unordentlich, lückenhaft oder schwierig zu interpretieren sind. Teil ihrer Präsentation sollte dennoch sein, welche Schwierigkeiten bei der Analyse des Datensatzes aufkamen.

Mögliche Datenquellen

Grundsätzlich stehen Ihnen **alle Daten** offen. Vielleicht haben Sie schon eine Idee, mit welchem Thema Sie sich beschäftigen möchten. Die Suche nach einer seriösen Datenquelle für ein Thema ist häufig der Schritt vor dem ersten Schritt einer Datenanalyse. Falls Sie sich lieber auf Basis bestehender Datensätze inspirieren lassen möchten, finden Sie hier eine Liste möglicher Datenquellen.

Absolute Empfehlung: **TidyTuesday-Datensätze**

- Beschreibung: ***A weekly social data project in R:** A weekly data project aimed at the R ecosystem. As this project was borne out of the R4DS Online Learning Community and the R for Data Science textbook, an emphasis was placed on understanding how to summarize and arrange data to make meaningful charts with ggplot2, tidy, dplyr, and other tools in the tidyverse ecosystem.*
 - Inhalt: Eine wilde Sammlung von fast 200 Datensätzen zu den verschiedensten Themen.
 - Tipp: Unter **#tidytuesday** auf Twitter teilen die TeilnehmerInnen ihre erstellten Visualisierungen, in der Regel zusammen mit dem entsprechenden Code. Dies bietet eine hervorragende Lernresource und jede Menge Inspiration für Ihre eigenen Analysen.

Administrative Daten

- Administrative Daten bieten einen hohen Grad an Verlässlichkeit. Gesammelt werden diese auf regionaler Ebene. Mit der Arbeit mit Daten aus administrativen Quellen ist in der Regel auch eine gewisse Einarbeitungszeit in die Struktur der jeweiligen Datenbank verbunden. Darüber hinaus ist Fachwissen gefragt, um die Relevanz und Konsistenz der Daten einzuordnen. Für Deutschland finden Sie regionale Daten beispielsweise unter:
 - [inkar.de](#): *In welchen Regionen werden besonders viele neue Wohnungen gebaut? Wo gibt es die meisten Kindergartenplätze? Wie weit ist es bis zum nächsten Supermarkt oder zur nächsten Apotheke? INKAR, der interaktive Online-Atlas des Bundesinstituts für Bau-, Stadt- und Raumforschung veranschaulicht die Lebensbedingungen in Deutschland und Europa. Mehr als 700 Indikatoren ermöglichen Stadt-Land-Vergleiche ebenso wie Auswertungen über die letzten zwei Jahrzehnte.*
 - [regionalstatistik.de](#): *Reginaldatenbank des Statistischen Bundesamts.*
- Auf europäischer Ebene unter:
 - [Eurostat database](#): *Datenbank des Statistischen Amtes der EU. Das Statistische Amt der Europäischen Union, kurz Eurostat oder ESTAT, ist die Verwaltungseinheit der Europäischen Union zur Erstellung amtlicher europäischer Statistiken und hat ihren Sitz in Luxemburg.*
- Auf multinationaler Ebene unter:
 - [UN data](#): *Datenbank der Vereinten Nationen. Die Organisation der Vereinten Nationen, auch UNO, sind ein zwischenstaatlicher Zusammenschluss von 193 Staaten und als globale internationale Organisation ein uneingeschränkt anerkanntes Völkerrechtssubjekt.*
 - [OECD database](#): *Datenbank der OECD. Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung ist eine internationale Organisation mit 38 Mitgliedstaaten, die sich der Demokratie und Marktwirtschaft verpflichtet fühlen. Die meisten Mitglieder gehören zu den Ländern mit hohem Pro-Kopf-Einkommen und gelten als entwickelte Länder.*