

rstatsZH

Lars Schöbitz

2024-08-14

Table of contents

Willkommen	3
Zielgruppe	4
Anna	4
Carlos	5
Fatima	5
Thomas	6
Lernziele	7
Abschliessende Beurteilung	8
Zeitaufwand	9
1 Module (noch nicht finalisiert)	10
References	11

Willkommen

Willkommen zu rstatsZH - Data Science mit R. Dieser Kurs richtet sich an Mitarbeiter:innen der Verwaltung des Kantons Zürich. Die Kosten für den Kurs betragen 1950.00 CHF pro Person für ein 14-wöchiges Programm mit einem Gesamtaufwand von etwa 50 Stunden. Das gesamte Programm wird online mittels Zoom Videokonferenzen durchgeführt.

Seit 2021 wird der Kurs jeweils im März und September durch das Statistische Amt Kanton Zürich angeboten. Im Jahr 2024 wurde der Kurs nach 8 Durchführungen überarbeitet und auf den neuesten Stand gebracht. Der Kurs wird von Lars Schöbitz, Lars Schöbitz GmbH, entwickelt und durchgeführt.

Zielgruppe

Die folgenden Learner Personas (Wilson 2019) beschreiben für welche Zielgruppe dieser Kurs konzipiert wurde. Falls du dich nicht in einer der folgenden Personas wiederfindest, ist das kein Problem. Der Kurs ist generell für Beginner:innen konzipiert für alle Personen geeignet, die Interesse an der Datenanalyse mit R haben und ihre Fähigkeiten in diesem Bereich aufbauen möchten.

Anna

Allgemeiner Hintergrund:

Anna ist eine 36-jährige Projektmanagerin im Gesundheitsamt der Kantonsregierung. Sie hat einen Master-Abschluss in Public Health und arbeitet seit über einem Jahrzehnt im öffentlichen Sektor. Zu ihren Aufgaben gehört die Koordination verschiedener Gesundheitsinitiativen und die Evaluierung von Programmergebnissen.

Relevantes Vorwissen:

Anna ist versiert im Umgang mit Excel und nutzt es ausgiebig für Datenanalysen und Berichterstellung. Sie verfügt aufgrund ihres akademischen Hintergrunds über Grundkenntnisse in Statistik, hat jedoch noch keine Programmiersprachen für die Datenanalyse verwendet und keine Erfahrung mit Systemen für Versionskontrolle.

Wahrnehmung der eigenen Bedürfnisse:

Anna hat das Bedürfnis ihre Datenanalysen reproduzierbar und öffentlich zu machen. Sie möchte durch offene Daten und offenen Code mehr Transparenz schaffen und effizienter werden. Sie glaubt, dass das Erlernen von R und Datenvisualisierungstechniken ihr dabei helfen wird, Daten effektiver für Nutzende zu präsentieren.

Besondere Überlegungen:

Anna hat einen vollen Terminkalender mit mehreren laufenden Projekten, daher benötigt sie einen flexiblen Lernplan und die Möglichkeit an einzelnen Modulen nicht teilzunehmen diese jedoch über eine Aufnahme nachzuholen. Sie lernt am liebsten anhand praktischer Beispiele aus ihrem Fachgebiet, wie zum Beispiel der Analyse von Gesundheitsdaten.

Carlos

Allgemeiner Hintergrund:

Carlos ist ein 50-jähriger Wirtschaftswissenschaftler, der in einer leitenden Funktion der Finanzabteilung der Kantonsregierung arbeitet. Er hat einen Dokortitel in Wirtschaftswissenschaften und ist seit über 20 Jahren in der Politikanalyse und Wirtschaftsprognose tätig.

Relevantes Vorwissen:

Carlos ist sehr erfahren im Umgang mit Excel für komplexe ökonomische Modellierung und Datenmanipulation. Er hat durch Kurse, die er während seiner Promotion besucht hat, einige Berührungspunkte mit Programmiersprachen, aber er hat diese Fähigkeiten in seiner derzeitigen Position nicht aktiv genutzt.

Wahrnehmung der eigenen Bedürfnisse:

Carlos möchte seine Fähigkeiten um moderne datenwissenschaftliche Tools erweitern. Er ist besonders daran interessiert, R für parameterisierte Analysen und Git für die Versionskontrolle zu erlernen, um gemeinsam mit seinem Team an Projekten zu arbeiten.

Besondere Überlegungen:

Carlos legt Wert auf einen strukturierten Lernansatz mit klaren Zielen. Er kann selbstständig lernen, schätzt aber auch die Möglichkeit zu diskutieren und zusammenzuarbeiten. Datenschutz ist ihm wichtig, da sein Team mit sensiblen Finanzdaten arbeitet, welche nicht öffentlich zugänglich sein dürfen.

Fatima

Allgemeiner Hintergrund:

Fatima ist eine 28-jährige Datenanalystin in der Bildungsdirektion. Sie hat einen Master-Abschluss in Mathematik und arbeitet seit fünf Jahren mit Bildungsdaten.

Relevantes Vorwissen

Sie verfügt über solide Grundkenntnisse in Mathematik und Statistik und arbeitet bereits mit base R, welches sie in einem Kurs während ihres Studiums kennengelernt hat.

Wahrnehmung der eigenen Bedürfnisse:

Fatima hat von den R Tidyverse Packages gehört und möchte diese nutzen, da ihr gesagt wurde es würde das Programmieren mit R erleichtern. Sie ist die einzige Person in ihrem Team mit Programmierkenntnissen und möchte ihre Teamkolleg:innen motivieren R zu lernen.

Besondere Überlegungen:

Fatima ist wissbegierig und bevorzugt interaktive, praktische Lernerfahrungen. Sie profitiert von visuellen Hilfsmitteln und Beispielen aus der Praxis, die in direktem Zusammenhang mit ihrer Arbeit im Bildungsbereich stehen. Während den Kursmodulen möchte sie die Möglichkeit haben, Fragen zu stellen und sich mit anderen Lernenden auszutauschen.

Thomas

Allgemeiner Hintergrund:

Thomas ist ein 42-jähriger IT-Spezialist aus Deutschland. Er hat einen Master-Abschluss in Informatik und ist seit einem Jahr bei der im Amt für Informatik tätig, wo er sich auf IT-Systeme und Datenmanagement konzentriert.

Relevante Vorkenntnisse:

Thomas hat umfangreiche Programmierkenntnisse in Sprachen wie C++, Python und SQL. Er ist mit Versionskontrollsystemen vertraut und verwendet Git in seinen Projekten. Er hat wenig Erfahrung mit R und Datenvisualisierung.

Einschätzung der eigenen Bedürfnisse:

Thomas ist daran interessiert, seine datenwissenschaftlichen Fähigkeiten mit R zu erweitern um für die steigenden Bedürfnisse an R Support in der Kantonsverwaltung gerüstet zu sein. Er möchte aktiv an Open-Source-Projekten teilnehmen und die R Community unterstützen.

Besondere Überlegungen:

Thomas fühlt sich mit technischen Inhalten wohl und hat Spaß an der Lösung von Problemen. Er bevorzugt es in online Kursen für sich alleine zu arbeiten und seine Kamera und Mikrofon auszuschalten. Thomas fällt es schwer Dialekt zu verstehen und zu sprechen, daher bevorzugt er Kurse in Hochdeutsch oder Englisch.

Lernziele

Die folgenden Übergeordneten Lernziele sind für diesen Kurs definiert:

1. Teilnehmer:innen werden in der Lage sein, eine Reihe gängiger datenwissenschaftlicher Werkzeuge (R, RStudio IDE, Git, GitHub, R tidyverse packages, Quarto) zu verwenden, um die Ergebnisse von Datenanalyseprojekten zu veranschaulichen und öffentlich zu kommunizieren.
2. Teilnehmer:innen können das Quarto-Dateiformat und den visuellen Bearbeitungsmodus der RStudio IDE zu verwenden, um technisch-wissenschaftliche Dokumente mit Zitaten, Fußnoten, Querverweisen, Abbildungen und Tabellen zu erstellen.
3. Teilnehmer:innen können Git Versionskontrolle verwenden um: (1) Änderungen an Dateien zu GitHub zu übertragen und von GitHub zu übernehmen, (2) Branches zu nutzen und Pull Requests auf GitHub zu öffnen, (3) mittels des GitHub Issue Tracker innerhalb eines Projekts zu kommunizieren und an Aufgaben zusammenzuarbeiten.

Abschliessende Beurteilung

Am Ende des Kurses können die Teilnehmer:innen ein Problem mittels eines Abschlussprojektes lösen, welches über GitHub mit der Öffentlichkeit geteilt wird. Das Output dieses Projektes ist ein zweiseitiger Bericht, welcher mit Quarto erstellt wird und Text (in Markdown Syntax) und Code (in R Syntax) enthält. Der Bericht ist basierend auf offenen Daten mit Variablen für eine Zeitreihe (z.B. Jahr), eine Kategorie (z.B. Gemeinde) und einer Metrik (z.B. Einwohnerzahl). Der Bericht enthält eine Visualisierung der Daten erstellt mittels R package ggplot2 und einer zusammenfassende Tabelle erstellt mit R package gt basierend auf einer deskriptiven Analyse (z.B. Mittelwert, Standardabweichung, Median, etc.). Der Bericht ist parametrisiert für die Zeitreihe (z.B. Jahr) und einer Kategorie (z.B. Gemeinde) und kann mittels einer selbst erstellten Funktion und eines for loop als HTML und PDF Output erstellt werden.

Zeitaufwand

Der Kurs besteht aus 10 Modulen von 2.5 Stunden pro Modul. Zusätzlich werden zwei Wochen für die Vorbereitung auf den Kurs und zwei Wochen für die Fertigstellung eines Abschlussprojektes angesetzt womit sich der gesamte Kurs über einen Zeitraum von 14 Wochen streckt.

Pro Modul wird von den Teilnehmer:innen ein zusätzlicher Zeitaufwand von 1.5 Stunden erwartet. Für die Fertigstellung des Abschlussprojektes in den zwei Wochen nach dem Kurs werden 4 Stunden erwartet. Dies führt zu folgendem Zeitaufwand:

- Vorbereitung: 2 Wochen à 1 Stunde = 2 Stunden
- Teilnahme Kursmodule: 10 Module à 2.5 Stunden = 25 Stunden
- Übungen: 10 Module à 1.5 Stunden = 15 Stunden
- Abschlussprojekt: 2 Wochen à 4 Stunden = 8 Stunden

Insgesamt wird von den Teilnehmer:innen ein Zeitaufwand von 50 Stunden erwartet.

1 Module (noch nicht finalisiert)

Die folgenden 10 Module mit einer Dauer von 2.5 Stunden pro Modul werden über einen Zeitraum von 10 Wochen gelehrt:

modul	titel
Modul 01	Willkommen zum Kurs & Werkzeugkiste
Modul 02	Explorative Datenanalyse durch Visualisierung & Digitaler Arbeitsplatz (DAP)
Modul 03	Daten Transformation mit dplyr & Kollaboratives Arbeiten mit GitHub I
Modul 04	Daten Import & Daten Management
Modul 05	Bedingte Anweisungen & Datumswerte
Modul 06	Daten Typen & Vektoren & For Loops
Modul 07	Datenformen lang und breit & Daten zusammenfügen
Modul 08	Mit Text Daten arbeiten & Kollaboratives Arbeiten mit GitHub II
Modul 09	Erstellung und Veröffentlichung analytischer Artikel mit Quarto und GitHub Pages
Modul 10	Funktionen schreiben & Parameterisierung von Berichten

References

Wilson, Greg. 2019. *Teaching Tech Together: How to Make Lessons That Work and Build a Teaching Community Around Them*. Chapman & Hall/CRC Press.