

AUTNES Data Dashboard- Dokumentation

Kurzdokumentation – Stand: 8. Mai 2025

Inhalt

Projektüberblick.....	1
1. Datenaufbereitung (<code>autnes_prep_combined.R</code>).....	2
Inputs	2
Outputs	2
2. Wählerwanderungen (<code>sankey_nrvote.R</code>).....	3
Inputs	3
Outputs	3
3. AUTNES Data Dashboard in RShiny (<code>flex_dash.R</code>).....	3
Inputs	3
Outputs	3
Instruktionen: Hinzufügen neuer Welle an Paneldaten	4
Instruktionen: Hinzufügen neuer Wahltermine.....	4
Instruktionen: Hinzufügen neuer Variablen	4
Ebene 0 – Fragenformulierung (<code>mappings_variable_text.xlsx</code>).....	4
Ebene 1 – Datenaufbereitung (<code>autnes_prep_combined.R</code>).....	4
Ebene 2 – Metadaten & Plotlogik im Dashboard (<code>flex_dash.R</code>).....	4
Lizenz und Wartung	5

Projektüberblick

Die vorliegende Pipeline führt vom Rohdatensatz der AUTNES Online Panel Study (Wellen 2017 – 2024) bis zur interaktiven Web-Applikation, dem AUTNES Data Dashboard.

Sie umfasst drei zentral verknüpfte R-Scripts/Dateien:

1. `autnes_prep_combined.R` – Datenaufbereitung
2. `sankey_nrvote.R` – Aufbereitung & Visualisierung der Wählerwanderungen
3. `flex_dash.Rmd` – Shiny/Flexdashboard-App

Gemeinsam erzeugen sie bereinigte Datensätze im RDS-Format, ein interaktives Sankey-Diagramm (HTML) sowie das Dashboard selbst und bilden damit sämtliche Arbeitsschritte von der Datenharmonisierung bis zur Veröffentlichung ab.

1. Datenaufbereitung (autnes_prep_combined.R)

Das Skript liest den kompletten AUTNES-Rohdatensatz ein, harmonisiert Variablennamen und Antwortkategorien über 24 Panelwellen, recodiert verbale Antworten in numerische Skalen und berechnet anschließend

- gewichtete Antwort-Verteilungen pro Welle,
- gewichtete Mittelwerte (inkl. 95 %-Konfidenz-intervalle) pro Welle,
- sowie Mittelwerte nach Subgruppen (Alter, Geschlecht, Bildungsgrad).

Anzumerken ist, dass in diesem Skript Konsolidierungen der Daten zu Zwecken der konsistenten Darstellung gemacht werden. Dazu zählen:

- Vereinheitlichung der Variablennamen – jedes Item erhält einen `basename + _w{Nr.}`, damit in allen 24 Wellen dieselbe Logik gilt (z.B. `vertrauen1_w7`).
- Zusammenführen funktional gleicher Variablen
 - Wirtschaftslage: `wirtschaftslage4_w*` → `wirtschaftslage3_w*`
 - PTV ÖVP: Alias-Spalten `ptv3`, `ptv4` werden in `ptv2` überführt.
- Harmonisierung von Antwort-kategorien
 - Zusammenlegung verschieden geschriebener Ausprägungen (z. B. „trifft gar nicht zu“ ↔ „stimme gar nicht zu“, „NEOS“ ↔ „Neos“).
 - Setzen sämtlicher Nicht-Antworten (keine Angabe, weiß nicht ...) auf NA.
- Rekodierung in numerische Skalen – verbale Kategorien werden für Mittelwert-Berechnungen in durchgängige Zahlenwerte übersetzt (0-10, 1-5, ...).
- Entfernung inkompatibler Items – Variablen mit abweichender Skala (z. B. `meinung13_w8`, `meinung16_w8`) werden ausgeschlossen.
- Zusammenfassung seltener oder nicht vergleichbarer Gruppen
 - Altersangaben in fünf Altersklassen (16-29, 30-39, ...),
 - Bildungsabschlüsse in gering / mittel / hoch,
 - Ausschluss unterrepräsentierter Gender-Kategorien (Divers / Möchte ich nicht sagen).

Inputs

- Datendatei **AUTNES_OPS_2017-20XX_w1-XX_DE.dta**
- Hart codierte Mapping-Listen (Label ↔ Basename, Antwort → numerischer Wert)
- Parameter `num_wave = XX` zur Definition der Anzahl der Wellen sowie Pfaddefinitionen

Outputs

- **dat_percentage.rds** – Antwortverteilungen
- **dat_means.rds** – Mittelwerte-Zeitreihen
- **dat_grouped.rds** – Mittelwerte nach soziodemographischen Gruppen

Diese drei RDS-Dateien stellen die Hauptdatenquelle des Dashboards dar.

2. Wählerwanderungen (`sankey_nrvote.R`)

Das Skript rekodiert für die Wahljahre 2013, 2017, 2019 und 2024 alle relevanten Wahl-variablen aus bis zu 23 Panelwellen in sieben Standard-kategorien (ÖVP, SPÖ, FPÖ, GRÜNE, NEOS, SONSTIGE/UNGÜLTIG, NICHTWAHL), berücksichtigt Nichtwahlindikatoren, passt die Stichprobe mit Raking-Gewichten (`anesrake`) an die amtlichen Stimmanteile an und erstellt schließlich die Datenstrukturen für ein Sankey-Diagramm (`links` und `nodes`) inklusive Farbuordnung und Tooltips.

Inputs

- Datendatei **AUTNES_OPS_2017-20XX_w1-XX_DE.dta**
- Zielverteilungen (Stimmanteile) der Wahljahre 2013, 2017, 2019, 2024

Outputs

- **`dat_links.rds`** und **`dat_nodes.rds`** – werden im Dashboard geladen
- **`sankey_diagram.html`** – eigenständige interaktive Grafik (optional, not used further)

3. AUTNES Data Dashboard in RShiny (`flex_dash.R`)

Dieses R Markdown-File definiert das Flexdashboard-Layout, lädt die zuvor erzeugten RDS-Dateien und stellt vier interaktive Ansichten bereit:

1. **Verteilung der Antworten** – gestapelte Flächenplots für eine Variable
2. **Mittelwerte** – Zeitreihen mehrerer Variablen
3. **Gruppierte Mittelwerte** – Vergleich nach Alter, Geschlecht, Bildung
4. **Wählerwanderungen** – Sankey-Diagramm der Parteiströme

Die App synchronisiert Datums-Slider zwischen Tabs, blendet optional Wahltermine und Konfidenzintervalle ein, zeigt Frageformulierungen kontextabhängig an. Auf Mobilgeräten wird die App unterdrückt und eine Informationsmeldung angezeigt.

Inputs

- **`dat_percentage.rds`**, **`dat_means.rds`**, **`dat_grouped.rds`** (aus Script 1)
- **`dat_links.rds`**, **`dat_nodes.rds`** (aus Script 2)
- **`*_election_dates`** (codiert als Variablen)
- **`mappings_variable_text.xlsx`** (Frageformulierungen)
- **`mobile-warning.html`** (Hinweis für Mobilgeräte)
- **`fa6.html`** (Einbindung des FontAwesome Bluesky Icon)
- Logo-PNG, internal CSS

Outputs

- Shiny-Dashboard

Instruktionen: Hinzufügen neuer Welle an Paneldaten

Inkrementieren der Variablen `num_wave` zur Bestimmung der Anzahl der Wellen sowie Überprüfung aller Pfade zur Datendatei (`MAIN`, `DATA`, `dat`) in der Datei `autnes_prep_combined.R`.

Instruktionen: Hinzufügen neuer Wahltermine

Adaption des Vektors, der die Daten der entsprechenden Wahl hält.

In Datei `flex_dash.R`: `austrian_election_dates`,
`presidential_election_dates`, `europa_election_dates`

Instruktionen: Hinzufügen neuer Variablen

In diesem Abschnitt sollen die Änderungen in den Dateien geschildert werden, die fürs Hinzufügen neuer Variablen erforderlich sind.

Ebene 0 – Fragenformulierung (`mappings_variable_text.xlsx`)

In der Excel-Datei sind neue Zeilen mit den entsprechenden Informationen zu Variablenamen, Prompt, Item und deren Vereinigung anzulegen.

Ebene 1 – Datenaufbereitung (`autnes_prep_combined.R`)

1. Label und Basename ergänzen

Im Block „*Define Labels and Basenames*“ ist der exakte Fragetext im Vektor `labels` und der passende Kurzname im Vektor `basenames` an identischer Stelle hinzuzufügen.

Beispiel:

```
labels <- c(..., "VERTRAUEN: DEM RECHNUNGSHOF")
basenames <- c(..., "vertrauen9")
```

2. Antwortskala numerisch abbilden

Hat die Frage eine bisher unbekannte Skala, ist in der Funktion `map_category_to_numeric()` ein neuer *case_when*-Zweig zu ergänzen, der jede verbale Kategorie auf eine Zahl abbildet.

3. Variablenvereinigung prüfen

Besitzt die Frage Aliase (z. B. andere Schreibungen derselben Partei), füge im Abschnitt `fact_collapse` den Join hinzu, damit identische Kategorien nicht doppelt auftauchen.

4. Prüfung auf Daten in einem neuen Jahr

Werden Daten einer Welle eines neuen Jahres eingefügt, ist unter „Date Format and Median Date Calculation“ zu definieren, ob es neue Spalten für jahresabhängige Variablen gibt, bzw. ob Werte der Vorjahre übernommen werden sollen.

Ebene 2 – Metadaten & Plotlogik im Dashboard (`flex_dash.R`)

5. Antwortkonfiguration anlegen

Im Block `category_cfg` (Helper Functions) einen neuen Eintrag für die Variable ergänzen:

- **Regex-Key**, z. B. "**^VERTRAUEN: DEM RECHNUNGSHOF**"
- **y_levels** – Reihenfolge der verbalen Ausprägungen
- **y_levels_numeric** – Mapping der Ausprägungen auf Zahlen
- **stack_dir** und **pos** – Richtung bzw. Stack-Position der Flächen

6. Regex-Listen erweitern

Die Regex anschließend in

- **category_regex** (Tab *Verteilung und gr. Mittelwerte*) und
- **category_regex_mittelwert** (Tab *Mittelwerte*)
eintragen, damit die Frage in den Drop-Down-Menüs auftaucht.

7. Kategorie-Liste ergänzen

Falls eine neue Themenrubrik entsteht, den Namen in **variable_categories_mittelwert** ergänzen.

8. Farben anpassen (im Falle von Parteibezug oder J/N-Frage)

- parteibezogene Variablen → Palette `party_colors` erweitern
- Ja/Nein-Fragen → Palette `yes_no_colors` anpassen, falls nötig

Lizenz und Wartung

Das AUTNES Data Dashboard einschließlich aller R-Skripte wird unter der Creative-Commons-Lizenz CC BY 4.0 Intl. bereitgestellt. Damit ist jede Nutzung, Änderung und Weiterverbreitung zulässig, sofern sichtbar auf die Autorinnen und Autoren sowie auf die Quelle „AUTNES Online Panel Study — Data Dashboard“ verwiesen wird. Auch eine kommerzielle Verwendung ist gestattet, solange die Namensnennung erfolgt. Für eingebundene Fremddaten, etwa den AUTNES-Paneldatensatz oder das Logo, gelten deren eigenen Lizenzbestimmungen.

Der Quellcode wird in einem Git-Repository versioniert; Änderungen an Aufbereitungsskripten oder der Dashboard-Logik werden aussagekräftig dokumentiert. Feedback und Rückmeldungen an: philipp.thannen@tuwien.ac.at oder pthannen@ethz.ch