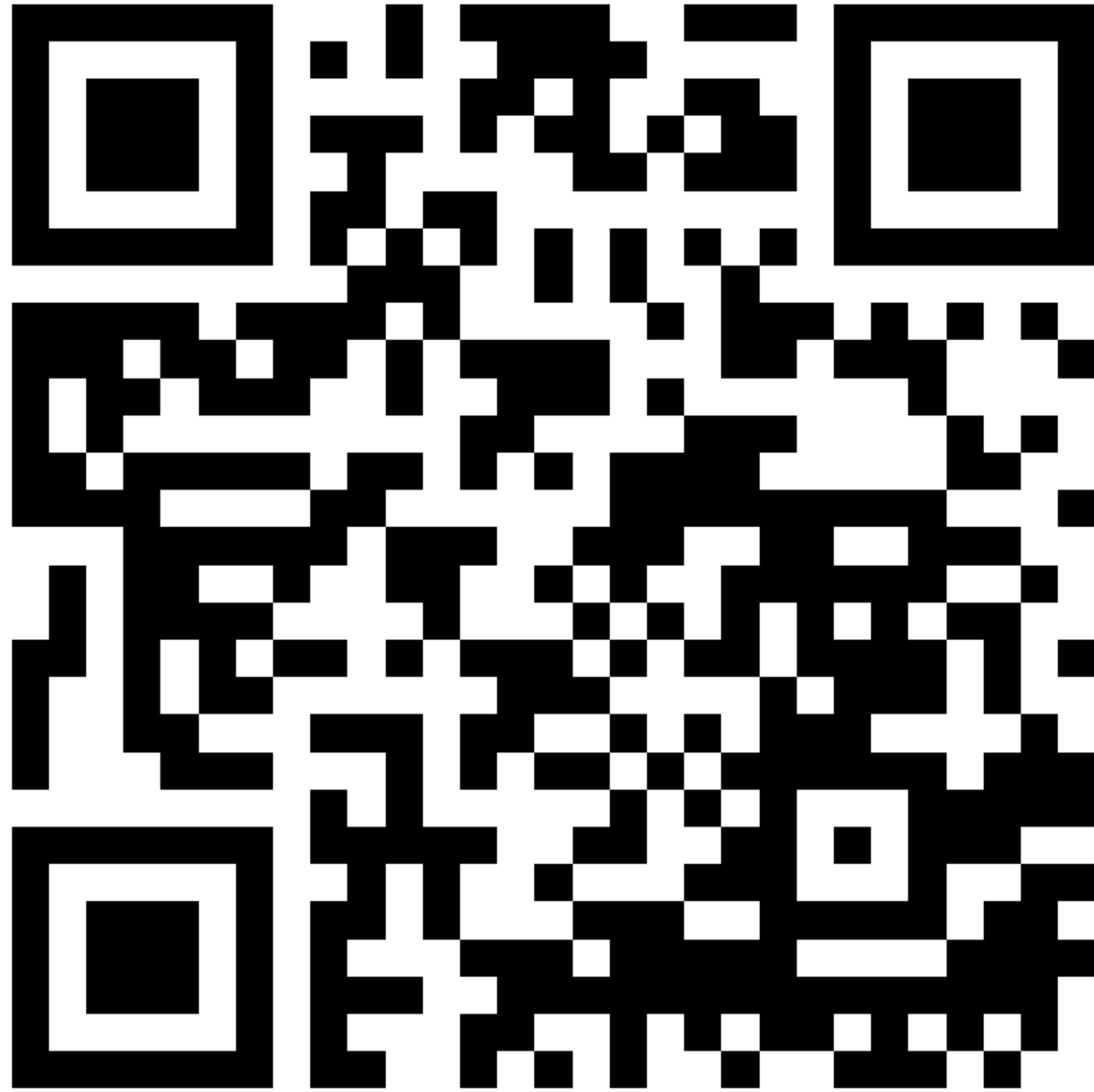


Praxis 1 | Datenrauferei



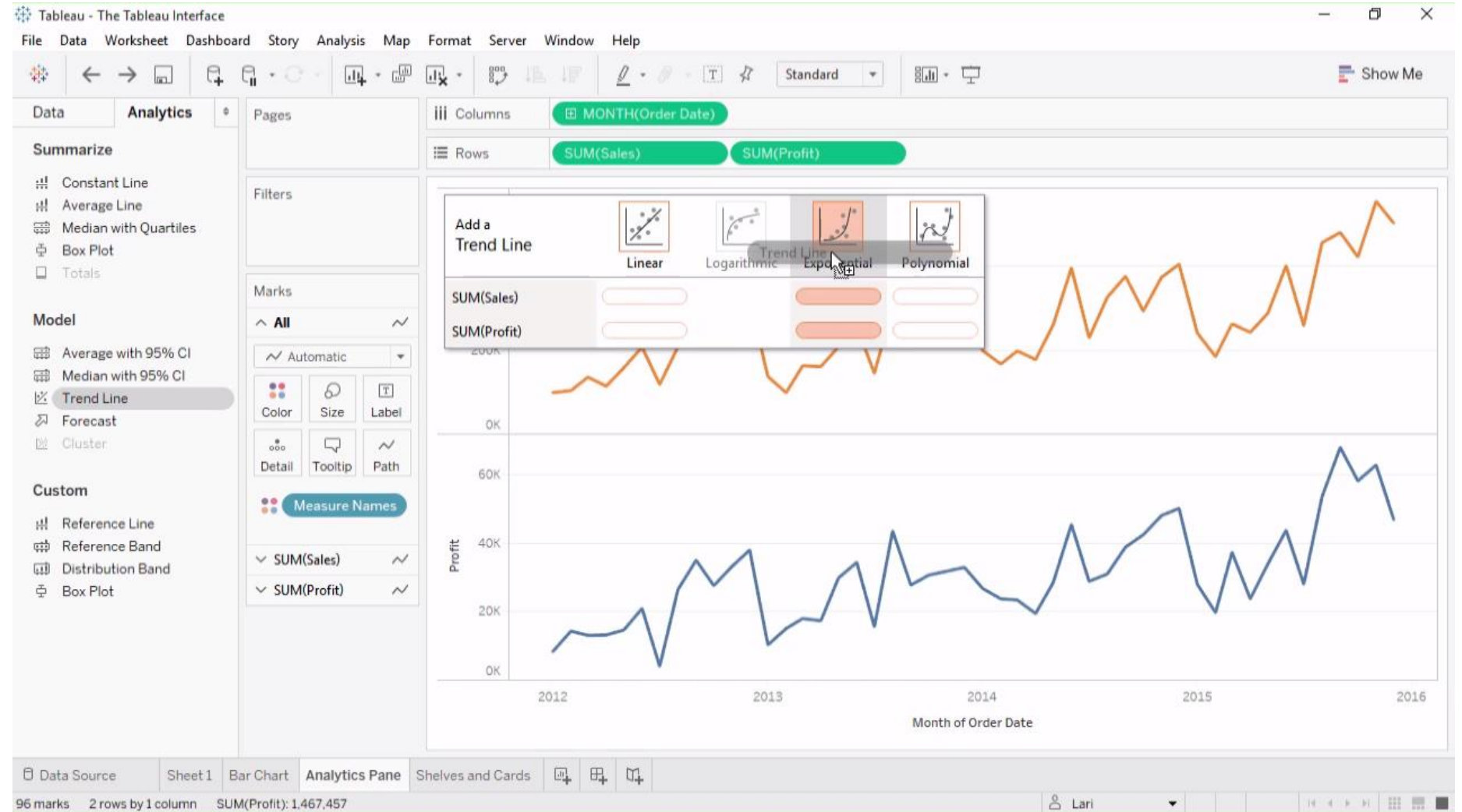
https://github.com/nzzdev/ddj_sfgz

Das richtige Tool

(gibt es nicht)



Data Reporting | Die (teuren) Platzhirsche



[illegible]

News
17.08.2021
Lesedauer ca. 5
Minuten
[Drucken](#)
[Teilen](#)

GENFORSCHUNG

Autokorrektur schreibt Fehler in Gendaten

Vor einer Sache sind Genetiker gewarnt worden: Tabellenkalkulationen. Offenbar umsonst, denn immer noch finden sich Fehler im Datenanhang von rund einem Drittel der Veröffentlichungen. Sie gehen auf das Konto von Excel und Co.

von [Dyani Lewis](#)



© INK MEDIA / STOCK.ADOBE.COM (AUSSCHNITT)

EXKLUSIVE ÜBERSETZUNG AUS **nature**

In Internet-Listenartikel und Twitter-Threads werden peinliche Autokorrektur-Fehler seit jeher gern herumgezeigt. Für Genforscher sind sie dagegen die Pest.

«Die im Prinzip lang bekannte Fehlerkette nimmt ihren Lauf, sobald die abgekürzte Form eines Gennamens falsch als Datum erkannt und von dem Tabellenprogramm automatisch »korrigiert« wird. **So werden dann wie von selbst aus *SEPT4* (Septin 4) und *MARCH1* (membranassoziierter Ring-CH-Finger 1) zum Beispiel 4-Sep und 1-Mar, also der 4. September und 1. März.»**

Vor- und Nachteile von Tabellenkalkulatoren

Vorteile

- Einfach
- Verfügbar oder kostenlos nutzbar
- Kombiniert Datenbereinigung und Datenrecherche

Nachteile

- Fehleranfällig
- Verschiedene Tabellen zusammenführen wird schnell kompliziert
- Nicht reproduzierbar
- Keine Code-Dokumentation möglich

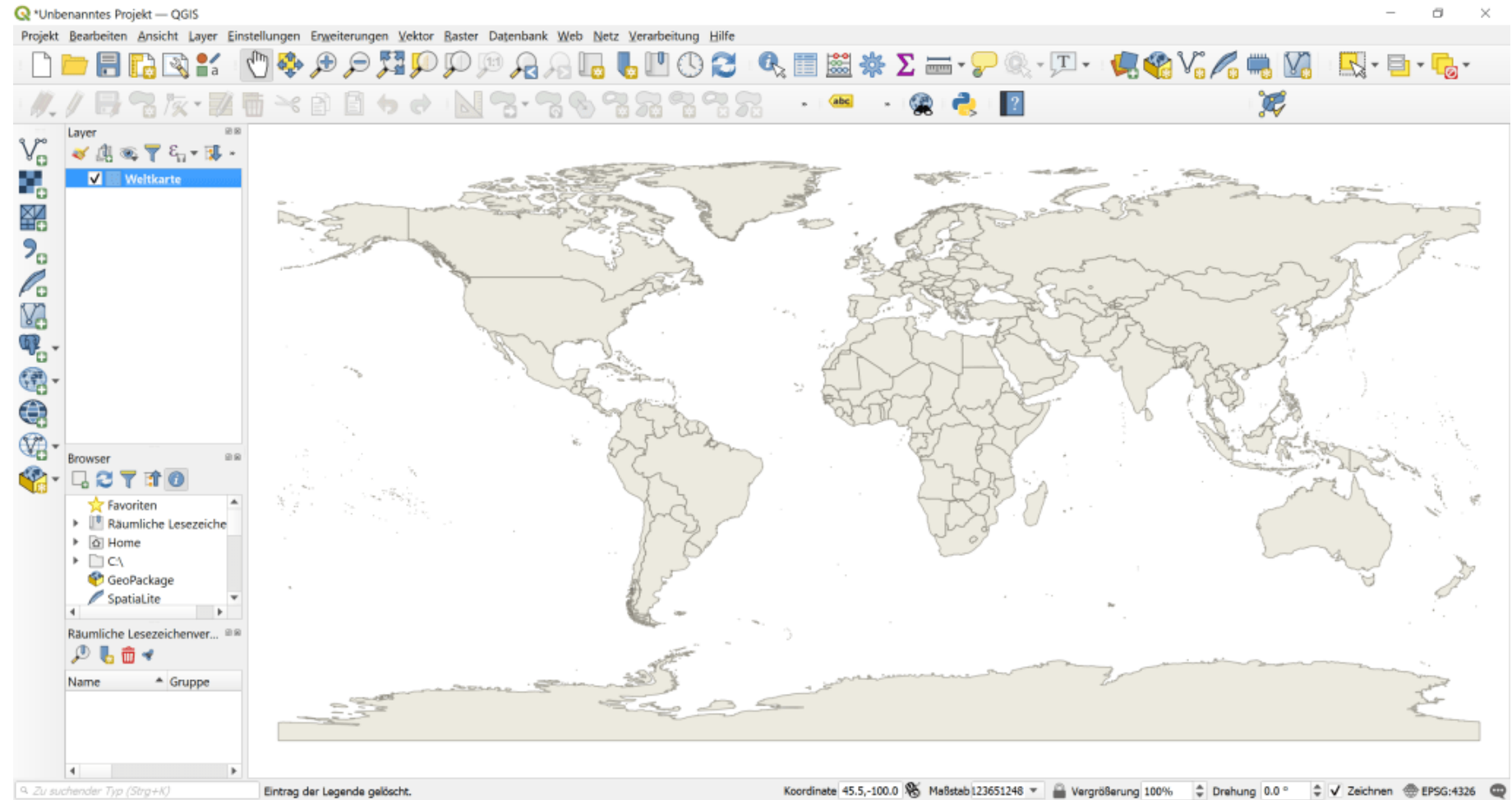
Die wichtigsten Excel/Sheets-Funktionen

- **Summieren**
=summe()
- **Zählen**
=anzahl()
- **Mittelwert**
=mittelwert()
- **Summieren, wenn Bedingung eintrifft**
=summewenn()
- **Werte aus einer anderen Tabelle übertragen**
=sverweis()

Koordinaten statt Tabellen | GIS-Software



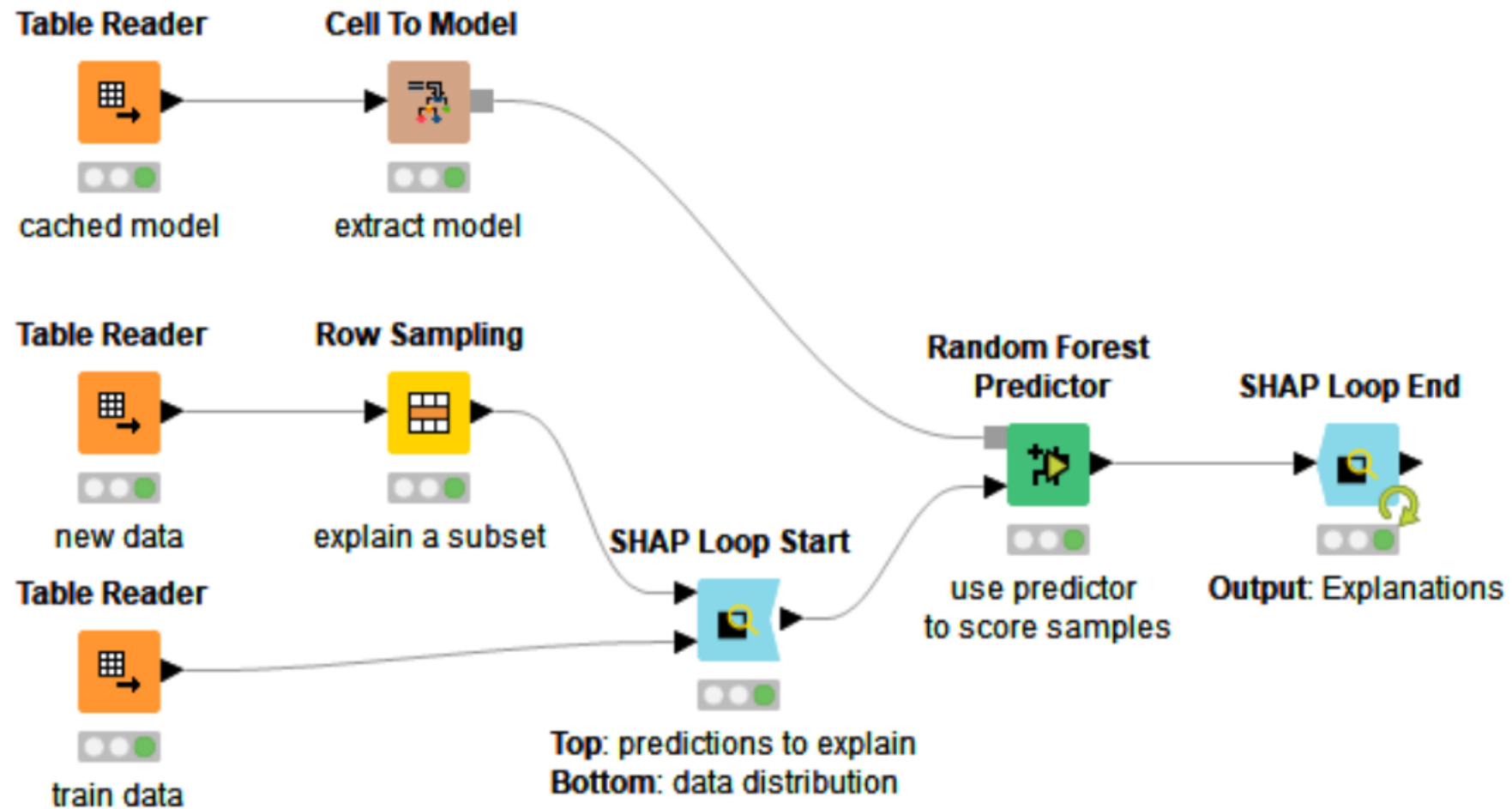
ArcGIS



Kostenlos:

- QGIS: <https://www.qgis.org/de/site/>
- Unfolded Studio: <https://www.unfolded.ai/studio/>

Der Underdog | Knime



Für Erwachsene | R und Python



jupyter tutorial Last Checkpoint: 3 minutes ago (autosaved) Python 3 Logout

File Edit View Insert Cell Kernel Widgets Help

Run Code

PyCon 2018: Using pandas for Better (and Worse) Data Science

GitHub: <https://github.com/justmarkham/pycon-2018-tutorial>

```
In [1]: import matplotlib.pyplot as plt
import pandas as pd
pd.__version__
```

```
Out[1]: '0.24.1'
```

Dataset: Stanford Open Policing Project ([video](#))

```
In [2]: # ri stands for Rhode Island
ri = pd.read_csv('police.csv')
```

```
In [3]: # what does each row represent?
ri.head()
```

```
Out[3]:
```

	stop_date	stop_time	county_name	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search
0	2005-01-02	01:55	NaN	M	1985.0	20.0	White	Speeding	Speeding	
1	2005-01-18	08:15	NaN	M	1965.0	40.0	White	Speeding	Speeding	
2	2005-01-23	23:15	NaN	M	1972.0	33.0	White	Speeding	Speeding	
3	2005-02-20	17:15	NaN	M	1986.0	19.0	White	Call for Service	Other	

Was man mit R & Python tun kann

Komplexe Analysen

Automatisierung

Machine Learning

Visualisieren

Web Scraping

Bots

NLP

R oder Python lernen

R Studio

Wichtige Komponenten

- Tidyverse
- ggplot

Python

Wichtige Komponenten

- Jupyter Notebook (Notebook-Ansatz)
- Pandas (Datenverarbeitung)
- Matplotlib (zum Plotten)

Das «richtige» Tool gibt es nicht

Praxisbeispiel 1

These definieren

Im Wallis leben die ältesten Nationalräte

Existieren Daten?

Ja: Postadressen -> <https://www.parlament.ch/de/organe/adressen>

Sind die Daten verlässlich?

Stichproben vornehmen. Urheber kontaktieren

Liegen die Daten bereinigt vor?

Ja (Beispiel)

Auswerten

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Natürliche Bevölkerungsbewegung, 2019															T 01.02.04.05
2	Kantone	Lebendgeburten		Todesfälle		Heiraten		Scheidungen		Eintragung der Partnerschaft		Auflösung der Partnerschaft ¹		Anerken- nungen ²	Adoptionen	
3		absolut	je 1000 Einwohner	absolut	je 1000 Einwohner	absolut	je 1000 Einwohner	absolut	je 1000 Einwohner	absolut	je 1000 Einwohner	absolut	je 1000 Einwohner	absolut	absolut	
4																
5																
6	Total	86 172	10.0	67 780	7.9	38 974	4.5	16 885	2.0	674	0.1	200	0.0	22 256	462	
7	Genferseeregion	17 095	10.4	11 721	7.1	6 761	4.1	3 551	2.2	152	0.1	47	0.0	5 393	92	
8	Waadt	8 607	10.7	5 454	6.8	3 147	3.9	1 708	2.1	75	0.1	19	0.0	2 798	36	
9	Wallis	3 138	9.1	2 898	8.4	1 464	4.2	711	2.1	21	0.1	7	0.0	845	21	
10	Genf	5 350	10.7	3 369	6.7	2 150	4.3	1 132	2.3	56	0.1	21	0.0	1 750	35	
11	Espace Mittelland	18 402	9.8	16 328	8.7	8 081	4.3	3 609	1.9	128	0.1	35	0.0	5 193	104	
12	Bern	9 933	9.6	9 667	9.3	4 575	4.4	1 935	1.9	67	0.1	18	0.0	2 503	47	
13	Freiburg	3 498	10.9	2 099	6.6	1 341	4.2	626	2.0	28	0.1	5	0.0	1 170	22	
14	Solothurn	2 592	9.5	2 311	8.4	1 259	4.6	486	1.8	21	0.1	6	0.0	629	21	
15	Neuenburg	1 670	9.5	1 585	9.0	635	3.6	392	2.2	9	0.1	5	0.0	637	11	
16	Jura	709	9.6	666	9.1	271	3.7	170	2.3	3	0.0	1	0.0	254	3	
17	Nordwestschweiz	11 647	10.0	9 541	8.2	5 419	4.6	2 297	2.0	74	0.1	23	0.0	2 583	57	
18	Basel-Stadt	2 060	10.5	2 063	10.6	966	4.9	394	2.0	17	0.1	4	0.0	560	3	
19	Basel-Landschaft	2 543	8.8	2 481	8.6	1 271	4.4	512	1.8	24	0.1	6	0.0	592	15	
20	Aargau	7 044	10.3	4 997	7.3	3 182	4.7	1 391	2.0	33	0.0	13	0.0	1 431	39	
21	Zürich	16 587	10.8	11 305	7.4	7 950	5.2	3 100	2.0	184	0.1	62	0.0	4 074	70	
22	Ostschweiz	11 652	9.9	9 723	8.2	5 674	4.8	2 193	1.9	67	0.1	16	0.0	2 324	65	
23	Glarus	413	10.2	416	10.3	174	4.3	58	1.4	4	0.1	0	*	104	3	
24	Schaffhausen	724	8.8	809	9.8	403	4.9	163	2.0	6	0.1	0	*	157	2	
25	Appenzell A. Rh.	558	10.1	484	8.7	276	5.0	111	2.0	2	0.0	2	0.0	124	4	
26	Appenzell I. Rh.	174	10.8	134	8.3	79	4.9	16	1.0	1	0.1	0	*	30	0	
27	St. Gallen	5 269	10.3	4 047	7.9	2 527	5.0	976	1.9	32	0.1	10	0.0	963	29	
28	Graubünden	1 630	8.2	1 748	8.8	855	4.3	342	1.7	10	0.1	1	0.0	415	7	
29	Thurgau	2 884	10.4	2 085	7.5	1 360	4.9	527	1.9	12	0.0	3	0.0	531	20	
30	Zentralschweiz	8 295	10.2	5 924	7.3	3 884	4.8	1 412	1.7	42	0.1	14	0.0	1 821	44	
31	Luzern	4 357	10.6	3 104	7.5	1 975	4.8	671	1.6	22	0.1	7	0.0	1 002	22	
32	Uri	342	9.4	307	8.4	175	4.8	47	1.3	2	0.1	0	*	79	3	
33	Schwyz	1 536	9.6	1 107	6.9	737	4.6	307	1.9	9	0.1	2	0.0	317	6	
34	Obwalden	347	9.2	279	7.4	189	5.0	54	1.4	1	0.0	0	*	65	3	
35	Nidwalden	384	8.9	336	7.8	187	4.3	89	2.1	2	0.0	1	0.0	97	2	
36	Zug	1 329	10.4	791	6.2	621	4.9	244	1.9	6	0.0	4	0.0	261	8	
37	Tessin	2 494	7.1	3 238	9.2	1 205	3.4	723	2.1	27	0.1	3	0.0	868	30	
38	* Entfällt, weil trivial oder Begriffe nicht anwendbar															
39	¹ Ausschiesslich gerichtlich aufgelöste Partnerschaften															
40	² Einschliesslich gerichtliche Feststellungen der Vaterschaft und Anerkennungen vor Gericht															
41	Quellen: BEVNAT, STATPOP															
42	© BFS															
43																
44	Auskunft: Bundesamt für Statistik (BFS), Sektion Demografie und Migration, info.dem@bfs.admin.ch, Tel. 058 463 67 11															
45																
46																

Tabelle 1: Demographische Indikatoren der Kantone der Schweiz, 2019														
Kanton	Geburten		Todesfälle		Heiraten		Scheidungen		Partnerschaften		Auflösungen		Adoptionen	
	absolut	je 1000	absolut	je 1000	absolut	je 1000	absolut	je 1000	absolut	je 1000	absolut	je 1000	absolut	je 1000
Genferseeregion	17 095	10.4	11 721	7.1	6 761	4.1	3 551	2.2	152	0.1	47	0.0	92	
Waadt	8 607	10.7	5 454	6.8	3 147	3.9	1 708	2.1	75	0.1	19	0.0	36	
Wallis	3 138	9.1	2 898	8.4	1 464	4.2	711	2.1	21	0.1	7	0.0	21	
Genf	5 350	10.7	3 369	6.7	2 150	4.3	1 132	2.3	56	0.1	21	0.0	35	
Espace Mittelland	18 402	9.8	16 328	8.7	8 081	4.3	3 609	1.9	128	0.1	35	0.0	104	
Bern	9 933	9.6	9 667	9.3	4 575	4.4	1 935	1.9	67	0.1	18	0.0	47	
Freiburg	3 498	10.9	2 099	6.6	1 341	4.2	626	2.0	28	0.1	5	0.0	22	
Solothurn	2 592	9.5	2 311	8.4	1 259	4.6	486	1.8	21	0.1	6	0.0	21	
Neuenburg	1 670	9.5	1 585	9.0	635	3.6	392	2.2	9	0.1	5	0.0	11	
Jura	709	9.6	666	9.1	271	3.7	170	2.3	3	0.0	1	0.0	3	
Nordwestschweiz	11 647	10.0	9 541	8.2	5 419	4.6	2 297	2.0	74	0.1	23	0.0	57	
Basel-Stadt	2 060	10.5	2 063	10.6	966	4.9	394	2.0	17	0.1	4	0.0	3	
Basel-Landschaft	2 543	8.8	2 481	8.6	1 271	4.4	512	1.8	24	0.1	6	0.0	15	
Aargau	7 044	10.3	4 997	7.3	3 182	4.7	1 391	2.0	33	0.0	13	0.0	39	
Zürich	16 587	10.8	11 305	7.4	7 950	5.2	3 100	2.0	184	0.1	62	0.0	70	
Ostschweiz	11 652	9.9	9 723	8.2	5 674	4.8	2 193	1.9	67	0.1	16	0.0	65	
Glarus	413	10.2	416	10.3	174	4.3	58	1.4	4	0.1	0	*	3	
Schaffhausen	724	8.8	809	9.8	403	4.9	163	2.0	6	0.1	0	*	2	
Appenzell A. Rh.	558	10.1	484	8.7	276	5.0	111	2.0	2	0.0	2	0.0	4	
Appenzell I. Rh.	174	10.8	134	8.3	79	4.9	16	1.0	1	0.1	0	*	0	
St. Gallen	5 269	10.3	4 047	7.9	2 527	5.0	976	1.9	32	0.1	10	0.0	29	
Graubünden	1 630	8.2	1 748	8.8	855	4.3	342	1.7	10	0.1	1	0.0	7	
Thurgau	2 884	10.4	2 085	7.5	1 360	4.9	527	1.9	12	0.0	3	0.0	20	
Zentralschweiz	8 295	10.2	5 924	7.3	3 884	4.8	1 412	1.7	42	0.1	14	0.0	44	
Luzern	4 357	10.6	3 104	7.5	1 975	4.8	671	1.6	22	0.1	7	0.0	22	
Uri	342	9.4	307	8.4	175	4.8	47	1.3	2	0.1	0	*	3	
Schwyz	1 536	9.6	1 107	6.9	737	4.6	307	1.9	9	0.1	2	0.0	6	
Obwalden	347	9.2	279	7.4	189	5.0	54	1.4	1	0.0	0	*	3	
Nidwalden	384	8.9	336	7.8	187	4.3	89	2.1	2	0.0	1	0.0	2	
Zug	1 329	10.4	791	6.2	621	4.9	244	1.9	6	0.0	4	0.0	8	
Tessin	2 494	7.1	3 238	9.2	1 205	3.4	723	2.1	27	0.1	3	0.0	30	

//Notiz//

1. Überblick verschaffen. Wer ist am längsten im Amt? Filterfunktion
2. Ständeräte aus Zürich
3. Alle jünger 30 einfärben
4. Alter berechnen
5. Grafiken einfügen

Weitere Thesen. Stimmen sie?

- 15 Prozent der Nationalratsmitglieder stammen aus Zürich
- Die SVP-Fraktion («V») hat die dienstältesten Mitglieder

Andere Datenformate als Excel (.xlsx)

Kommagetrennte Textdatei (.csv)

Vorgehen: Importieren in Excel/Google Sheets

```
geoRegion,datum,entries,sumTotal,timeframe_7d,offset_last7d,sumTotal_last7d,timeframe_14d,offset_las
CH,2020-02-24,1,1,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,540
CH,2020-02-25,1,2,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,540
CH,2020-02-26,10,12,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-02-27,10,22,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-02-28,10,32,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-02-29,13,45,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-03-01,11,56,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-03-02,30,86,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,5
CH,2020-03-03,33,119,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-04,61,180,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-05,62,242,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-06,73,315,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-07,49,364,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-08,68,432,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE,
CH,2020-03-09,192,624,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE
CH,2020-03-10,210,834,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALSE
CH,2020-03-11,333,1167,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-12,356,1523,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-13,432,1955,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-14,417,2372,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-15,326,2698,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-16,1060,3758,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FAL
CH,2020-03-17,1082,4840,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FAL
CH,2020-03-18,1206,6046,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FAL
CH,2020-03-19,835,6881,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-20,1138,8019,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FAL
CH,2020-03-21,691,8710,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-22,547,9257,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FALS
CH,2020-03-23,1461,10718,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FA
CH,2020-03-24,1242,11960,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FA
CH,2020-03-25,1068,13028,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FA
CH,2020-03-26,1114,14142,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FA
CH,2020-03-27,1305,15447,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FA
CH,2020-03-28,722,16160,FALSE,811274,0,FALSE,793134,0,FALSE,756507,0,FALSE,30944,0,FALSE,52691,0,FAL
```

JavaScript-Objekt (.json)

Vorgehen: Versuchen, auf eine Hierarchie zu bringen:

<https://www.convertcsv.com/json-to-csv.htm>

```
"events": [  
  {  
    "shipmentNumber": "996002075340319210",  
    "eventCode": "PARCEL.*.1.1201",  
    "timestamp": "2020-02-04T10:10:59+01:00",  
    "zip": "462000",  
    "city": "Härkingen Paketzentrum",  
    "country": "CH",  
    "subEventId": null,  
    "subEventAdditionalInfo": null,  
    "eventCodeType": "INTERNAL",  
    "externalMetadata": {}  
  },  
  {  
    "shipmentNumber": "996002075340319210",  
    "eventCode": "PARCEL.*.1.4000",  
    "timestamp": "2020-02-05T07:36:28+01:00",  
    "zip": "613070",  
    "city": "Willisau Distributionshub",  
    "country": "CH",  
    "subEventId": null,  
    "subEventAdditionalInfo": null,  
    "eventCodeType": "INTERNAL",  
    "externalMetadata": {}  
  },  
  {  
    "shipmentNumber": "996002075340319210",  
    "eventCode": "PARCEL.*.1.600",  
    "timestamp": "2020-02-04T06:12:09+01:00",  
    "zip": null,  
    "city": null,  
    "country": "CH",  
    "subEventId": null
```

Spatial Data (.geojson, .shp, .gpkg, ...)

Vorgehen: Mit GIS-Software wie QGIS oder Unfolded öffnen

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "stroke": "red"
      },
      "geometry": {
        "type": "Polygon",
        "coordinates": [
          [
            [
              8.528459072113037,
              47.380727869134866
            ],
            [
              8.532922267913818,
              47.37930395943437
            ],
            [
              8.537063598632812,
              47.3779671928823
            ],
            [
              8.5410333267974854,
              47.37713896181911
            ],
            [
              8.528459072113037,
              47.380727869134866
            ]
          ]
        ]
      }
    }
  ]
}
```


Praxisbeispiel 2

These: Seit dieser Woche sinken die Todesfallzahlen der Corona-Pandemie.

Daten (am Seitenende «Daten als .csv»):

<https://www.covid19.admin.ch/de/overview>

Datenbeschreibung:

<https://www.covid19.admin.ch/api/data/documentation>

Tools

Excel / Sheets

Praxisbeispiel 3

Wo in der Schweiz leben besonders viele ältere Menschen?

Daten:

Gemeindeporträt

<https://www.bfs.admin.ch/bfs/de/home/statistiken/regionalstatistik/regionale-portraets-kennzahlen/gemeinden.html>

Gemeindegrenzen

<https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/administrative-grenzen/generalisierte-gemeindegrenzen.html>

Tools:

- Excel/Sheets
- Unfolded Studio

Wenn Daten als PDF vorliegen

Beispiel:

<https://www.mundipharma.ch/wp-content/uploads/Zuwendungen-an-medizinische-Fachkreise-2020.pdf>

Tools:

- Wenn maschinenlesbar: Tabula
<https://tabula.technology/>
- Wenn gerastert (Bild): Abbyy Finereader (kostenpflichtig)
<https://pdf.abbyy.com/de/>

Daten aus Karten «leihen»

Beispiel:

<https://defikarte.ch/>

Tools:

- WebInspector von Chrome (oder Safari) GeoJSON exportieren
- In QGIS öffnen und gewünscht exportieren

