

Why every data journalist should work with Python and Jupyter Notebooks

Barnaby Skinner,
Ressortleiter NZZ Visuals
20. May 2021

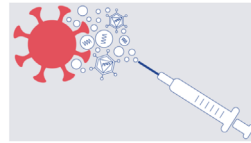
NZZ Visuals

++ Data aggregation ++ Cleaning ++ Analysis ++ Plotting ++ Mapping
++ Scrollytelling ++ Interactives ++ Illustration ++ Writing

Was ist wahr und was falsch: 11 Mythen zu Folgen, Risiken und Nebenwirkungen der Coronavirus-Impfung

Es kursieren viele Behauptungen rund ums Impfen allgemein und um die Covid-19-Impfung. Wir klären auf, was dahintersteckt.

Illustration: Kathrin, Lena Dählmann, Apple Lenzke
28.01.2021, 13:00 Uhr



Es gibt viele Vorurteile und Mythen, die derzeit über Impfungen generell und die Covid-19-Impfung im Besonderen kursieren. Was ist wahr? Welche Nebenwirkungen gibt es, und welche Spekulationen sind wissenschaftlich nicht belegt? Wir klären elf gängige, teils halbwahre und teils falsche Aussagen auf.

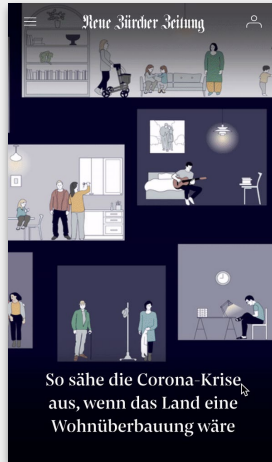
Diese Staaten entscheiden die US-Präsidentschaftswahl

Joe Biden liegt einen Monat vor den US-Wahlen in den Umfragen vor Donald Trump – aber noch ist nichts entschieden. Wir zeigen in diesem Erklärstück auf, wie wichtige «Swing States» in der Vergangenheit gewählt haben und was für Trump und was für Biden spricht.

Joana Kelén, Alexandra Kohler, Adina Renner, Manuel Roth

05.10.2020, 05:30 Uhr

Am 3. November 2020 wählen die Amerikaner ihren nächsten Präsidenten und entscheiden, ob der Republikaner Donald Trump im Amt bleibt oder der Demokrat Joe Biden der nächste Präsident der Vereinigten Staaten wird.



Nidwalden liegt im Impfvergleich vorne

Anzahl verabreichter Impfdosen pro 100 Einwohner, nach Kanton

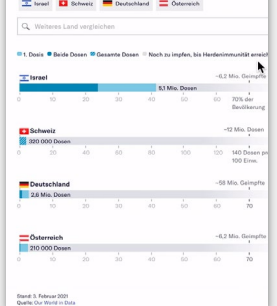


Die Zahlen beziehen sich auf die verabreichten Impfdosen, nicht auf genehmigte Personen. Eine Person muss im Normalfall zwei Dosen erhalten. Stand: 6.3.2021
Quelle: [BFS](#), [NZZ](#)

Mit rund 370 000 verabreichten Impfdosen positioniert sich die Schweiz noch immer weit entfernt von den weltweiten Spitzenreitern, rückt aber leicht auf. Im internationalen Vergleich der verimpften Dosen pro 100 Einwohner belegt die Schweiz unter den Ländern mit mindestens einer Million Einwohnern derzeit aber immerhin Platz 6.

So schreitet die Impfkampagne weltweit voran

Geographische Personen (in Prozent der Bevölkerung) und verabreichte Impfdosen (auf 100 Einwohner angepasste Länder)



Stand: 3. Februar 2021
Quelle: Our World in Data

NZZ Visuals



**Barnaby
Skinner**
Ressortleiter Visuals



Kaspar Manz
Leiter Grafik



**Kiara Leona
Nagy**
Leiterin Editorial Tech



**Christian
Kleeb**
Designer



Anja Lemcke
Designerin



**Eugen
Fleckenstein**
Designer



Joana Kelen
Designerin



Manuel Roth
Softwareentwickler



Philip Küng
Softwareentwickler



Nikolai Thelitz
Datenjournalist



Jonas Oesch
Designer



**Alexandra
Kohler**
Datenjournalistin



Adina Renner
Designerin



Florian Seliger
Datenjournalist



Nicolas Staub
Softwareentwickler



**Charlotte
Eckstein**
Designerin



Franco Gervasi
Softwareentwickler

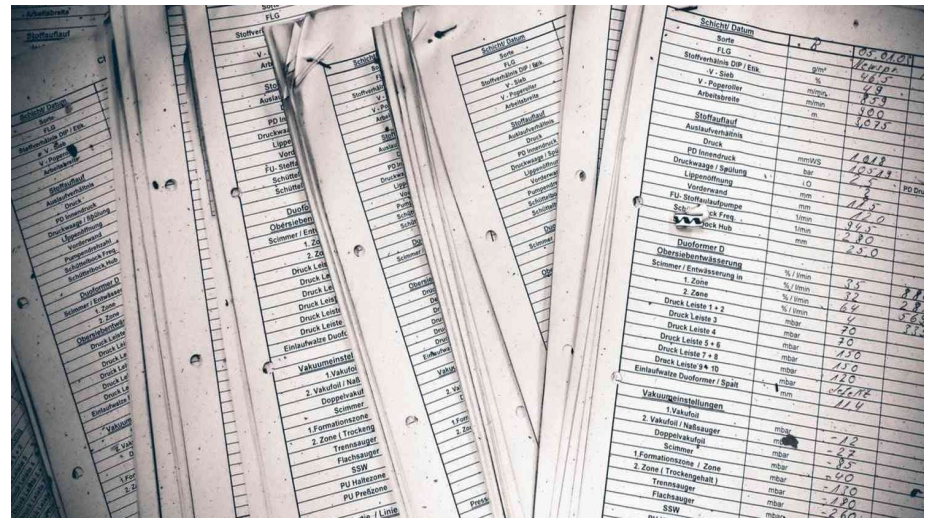


Simon Haas
Datenjournalist

Messy Data

REL006A: generation 0									
BLOUNT et al. 2 sequenced at MT single end, 35 or 38 bp									
Terwilliger et al. 2 Population: Ara-3									
STRAIN	Genotype	Clade	Population	Marker	Run	Sequencing date	Clade	generation	mutator
1	Z56464	20000	C1(C2)	Ara-3	None	SPR036205	20.7 unknown	REL2181A	5000
2	REL10979	40000	C5(H)	Ara-3	+	SPR036203	30.2 +	REL360A	1000
3	REL10988	40000	C2	Ara-3	+	SPR036203	30.2 minus	REL764B	None
4	Z56409	10000	UC	Ara-3	None	SPR036202	87.3 unknown	REL1166A	2000
5	Z56357	30000	C2	Ara-3	None	SPR036203	111.2 unknown	REL7176B	15000
6	Z56316	30000	C1	Ara-3	None	SPR036203	113.9 unknown	REL1070A	1500
7	Z56458	20000	C1(C2)	Ara-3	None	SPR036204	126.8 unknown	REL4538A	10000 UC
8	Z56446	15000	UC	Ara-3	None	SPR036203	141.1 unknown	REL366B	1000
9	Z56409	5000	unknown	Ara-3	None	SPR036201	144.2 unknown	REL1070B	1500
10	Z56467	20000	C1(C2)	Ara-3	None	SPR036205	unknown	REL1166B	2000
11	Z56477	25000	C1	Ara-3	None	SPR036207	unknown	REL764A	500
12	Z56483	25000	C3	Ara-3	None	SPR036208	unknown	REL1166C	50000 C2(H)
13	Z56399	31000	C1	Ara-3	None	SPR036204	minus	REL1166A	50000 C2(H)
14	Z56399	31000	C2	Ara-3	None	SPR036207	minus	REL1166A	50000 C2(H)
15	Z56399	31000	C3	Ara-3	None	SPR036209	+	REL1166A	50000 C2(H)
16	Z56399	31000	C4	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
17	Z56399	31000	C5	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
18	Z56399	31000	C6	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
19	Z56399	31000	C7	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
20	Z56399	31000	C8	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
21	Z56399	31000	C9	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
22	Z56399	31000	C10	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
23	Z56399	31000	C11	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
24	Z56399	31000	C12	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
25	Z56399	31000	C13	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)
26	Z56399	31000	C14	Ara-3	None	SPR036202	plus	REL1166A	50000 C2(H)

	A	B
1	Customer Name	
3	John K. Doe Jr.	Doe, John
4	Mr. Doe, John	Doe, John
5	Jane A. Smith	Smith, Jane
6	MS. Jane Smith	Smith, Jane
7	Smith, Jane	Smith, Jane
8	Dr Anthony R Von Fange III	Von Fange, Anthony
9	Peter Tyson	Tyson, Peter
10	Dan E. Williams	Williams, Dan
11	James Davis Sr.	Davis, James
12	James J. Davis	Davis, James
13	Mr. Donald Edward Miller	Miller, Donald
14	Miller, Donald	Miller, Donald
15	Rajesh Krishnan	Krishnan, Rajesh



	A	B	C	D	E	F
2	Company Name		Invoice Date	Delivery Date	Amounts	
4	Jenny		01.09.2007	1900/01/00	2057	
5	Jenny Total		01.11.2007	1900/01/00	2669	
6	Sam			1900/01/01	1426	
7				1998/01/01 01.01.1998	1185	
8				1998/01/01	2359	
9				1998/01/01 01.06.1998	1886	
10				1998/01/01	2359	
11				2000/07/01 01.07.2000	2486	
12	Sam Total				9342	
13						
14						
15						
16						
17						
18	Peter		2000/01/01	1900/01/00	2385	
19			1975/04/01	1900/01/00	0	
20			2000/04/01	1900/01/00 0.000		
21			2005/06/01	1900/01/00 7.293 07		
22			1993/07/01	1900/01/00 42.717 42		
23			1993/07/01 01.07.1993	55.872.63		
24			01.08.2000	1900/01/00 40.176.80		
25			01.09.2000	1900/01/00	1585	
26			01.10.2001	1900/01/00	1384	
27			01.10.2004	01.10.2004	01518	

Enter Python (and Jupyter Notebooks)

Tools

```
In [1]: %matplotlib inline
import pandas as pd
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
import requests
from pyaxis import pyaxis
#from google.colab import files
```

Data source

```
In [3]: #Quelle
EXAMPLE_URL = 'https://www.bfs.admin.ch/bfsstatic/dam/assets/12407536/mastk...'

px = pyaxis.parse(EXAMPLE_URL, encoding='ISO-8859-2')
df = pd.DataFrame(px['DATA'])
```

Import

```
In [4]: #Data cleaning
df['DATA'] = df['DATA'].astype(int)
```

Data

```
In [5]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 365904 entries, 0 to 365903
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Unfallfolge                          365904 non-null object
1   Verwendetes Verkehrsmittel           365904 non-null object
2   Art der Verkehrsteilnahme            365904 non-null object
3   Geschlecht                           365904 non-null object
4   Altersklasse                         365904 non-null object
5   Verwendung eines Schutzsystems        365904 non-null object
6   Jahr                                 365904 non-null object
7   DATA                                365904 non-null int64
dtypes: int64(1), object(7)
memory usage: 22.3+ MB
```

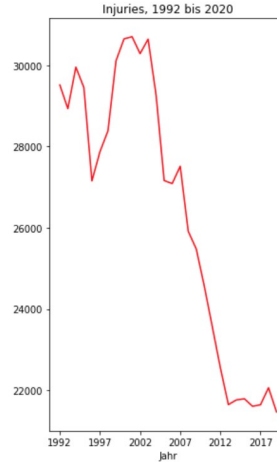
Enter Matplotlib

1 Development of injuries suffered in car accidents in Switzerland, 1992 - 2020?

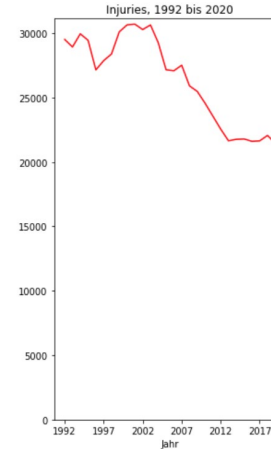
```
#Visualisieren  
df.groupby('Jahr')['DATA'].sum().plot()  
  
<AxesSubplot: xlabel='Jahr'>
```



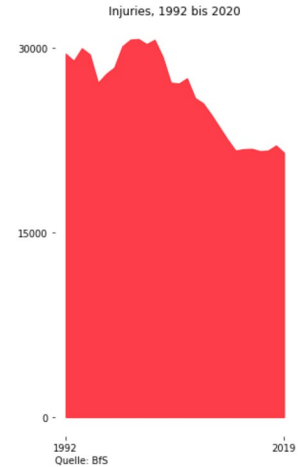
Injuries in Accidents on Swiss Roads
have decreased



Injuries in Accidents on Swiss Roads
have decreased



Injuries in Accidents on Swiss Roads
have decreased



Automate

Imports

```
import matplotlib inline
import pandas as pd
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
import geopandas
from geopandas import GeoDataFrame
import requests
from pyaxis import pyaxis
```

```
#Quelle
EXAMPLE_URL = 'https://www.bfs.

px = pyaxis.parse(EXAMPLE_URL,
df = pd.DataFrame(px['DATA']))
```

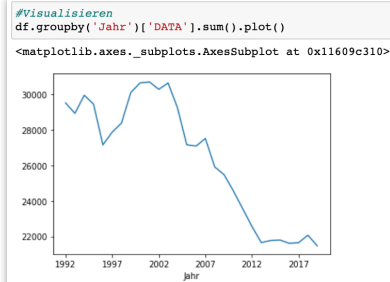
Analyse

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 365904 entries, 0 to 365903
Data columns (total 8 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   Unfallfolge                 365904 non-null object
1   Verwendetes Verkehrsmittel  365904 non-null object
2   Art der Verkehrsteilnahme   365904 non-null object
3   Geschlecht                  365904 non-null object
4   Altersklasse                365904 non-null object
5   Verwendung eines Schutzsystems 365904 non-null object
6   Jahr                        365904 non-null object
7   DATA                       365904 non-null int64
dtypes: int64(1), object(7)
memory usage: 22.3+ MB
```

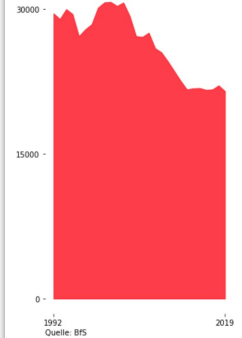
```
#Visualisieren
df.groupby('Jahr')['DATA'].sum().plot()
```

Visualise

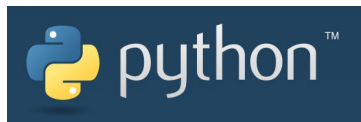


Style

Rückgang der Verletzungen
Verletzte Personen auf Schweizer Strassen,
1992 bis 2019



Beautiful Open Source



Thanks

@barjack

barnaby.skinner@nzz.ch

Barnaby Skinner,
Ressortleiter NZZ Visuals
20. May 2021