A Data Validation Infrastructure for R


Daniel Franke
Introducing A Data Validation Infrastructure
For R: The {validate} Package

Data Validation

	SEX	AGE	LANG_INTERVIEW	POB	PARTY	LEFT_RIGHT_0_10	CHILDREN
1	Dona	41	Castellà	Catalunya	PSC	Extrema esquerra	2
2	Home	59	Castellà	Altres comunitats autònomes	Podemos	5	2
3	Dona	18	Català	Catalunya	Podemos	2	Cap
4	Dona	23	Castellà	Catalunya	Cap	5	Cap
5	Dona	74	Català	Catalunya	Cap	5	2
6	Dona	89	Castellà	Altres comunitats autònomes	PSC	2	8

Data Validation

	SEX	AGE	LANG_INTERVIEW	POB	PARTY	LEFT_RIGHT_0_10	CHILDREN
1	Dona	41	Castellà	Catalunya	PSC	Extrema esquerra	2
2	Home	59	Castellà	Altres comunitats autònomes	Podemos	5	2
3	Dona	18	Català	Catalunya	Podemos	2	Cap
4	Dona	23	Castellà	Catalunya	Cap	5	Cap
5	Dona	74	Català	Catalunya	Cap	5	2
6	Dona	89	Castellà	Altres comunitats autònomes	PSC	2	8

- Technical checks: Age vs. Children
- Checks on subject: Party vs Ideology
- Metadata: Interview Language

Data Validation

	SEX	AGE	LA	ILDREN
1	Dona	41		2
2	Home	59		2
3	Dona	18		Cap
4	Dona	23		Cap
5	Dona	74		2
6	Dona	89		8

	SEX	AGE	LANG_INTERVIEW	POB	PARTY	LEFT_RIGHT_0_10	CHILDREN
1	Dona	41	Castellà	Catalunya	PSC	Extrema esquerra	2
2	Home	59	Castellà	Altres comunitats autònomes	Podemos	5	2
3	Dona	18	Català	Catalunya	Podemos	2	Cap
4	Dona	23	Castellà	Catalunya	Cap	5	Cap
5	Dona	74	Català	Catalunya	Cap	5	2
6	Dona	89	Castellà	Altres comunitats autònomes	PSC	2	8
7	Dona	30	Castellà	Catalunya	Cap	2	Cap
8	Dona	53	Castellà	Resta del món	Cap	4	3
9	Home	25	Castellà	Catalunya	Cap	5	Cap
10	Home	67	Català	Catalunya	Junts per Catalunya	3	2
11	Home	51	Català	Catalunya	ERC	2	3
12	Home	30	Castellà	Catalunya	PSC	No ho sap	Cap
13	Dona	38	Castellà	Resta del món	Cap	6	4
14	Dona	35	Castellà	Catalunya	Cap	7	1
15	Dona	26	Català	Catalunya	Barcelona en Comú	4	Cap
16	Dona	68	Català	Catalunya	Junts per Catalunya	5	1
17	Dona	28	Castellà	Resta del món	PSC	5	1
18	Dona	40	Català	Catalunya	ERC	3	1
19	Home	39	Castellà	Catalunya	C's	6	2
20	Dona	69	Castellà	Altres comunitats autònomes	PSC	5	4
21	Dona	40	Català	Catalunya	ERC	3	2
22	Home	82	Castellà	Altres comunitats autònomes	Cap	5	4
23	Home	40	Castellà	Resta del món	PSC	8	Cap
24	Home	44	Castellà	Catalunya	ERC	2	2
25	Dona	20	Castellà	Catalunya	Cap	No ho sap	Cap
26	Dona	62	Castellà	Altres comunitats autònomes	Podemos	4	2
27	Dona	71	Castellà	Altres comunitats autònomes	PSC	5	2
28	Home	27	Castellà	Catalunya	Cap	5	1
29	Dona	52	Castellà	Catalunya	PPC	5	Cap
30	Dona	55	Castellà	Resta del món	Cap	5	2
31	Dona	74	Castellà	Altres comunitats autònomes	PSC	Extrema esquerra	3
32	Home	73	Català	Catalunya	PSC	5	1
33	Home	29	Català	Altres comunitats autònomes	Podemos	2	Cap
34	Home	26	Català	Catalunya	CUP	5	Cap
35	Home	73	Català	Catalunya	Cap	4	3
36	Home	38	Català	Catalunya	ERC	2	3
37	Home	64	Català	Catalunya	Catalunya en Comú Podem	3	2

Definition

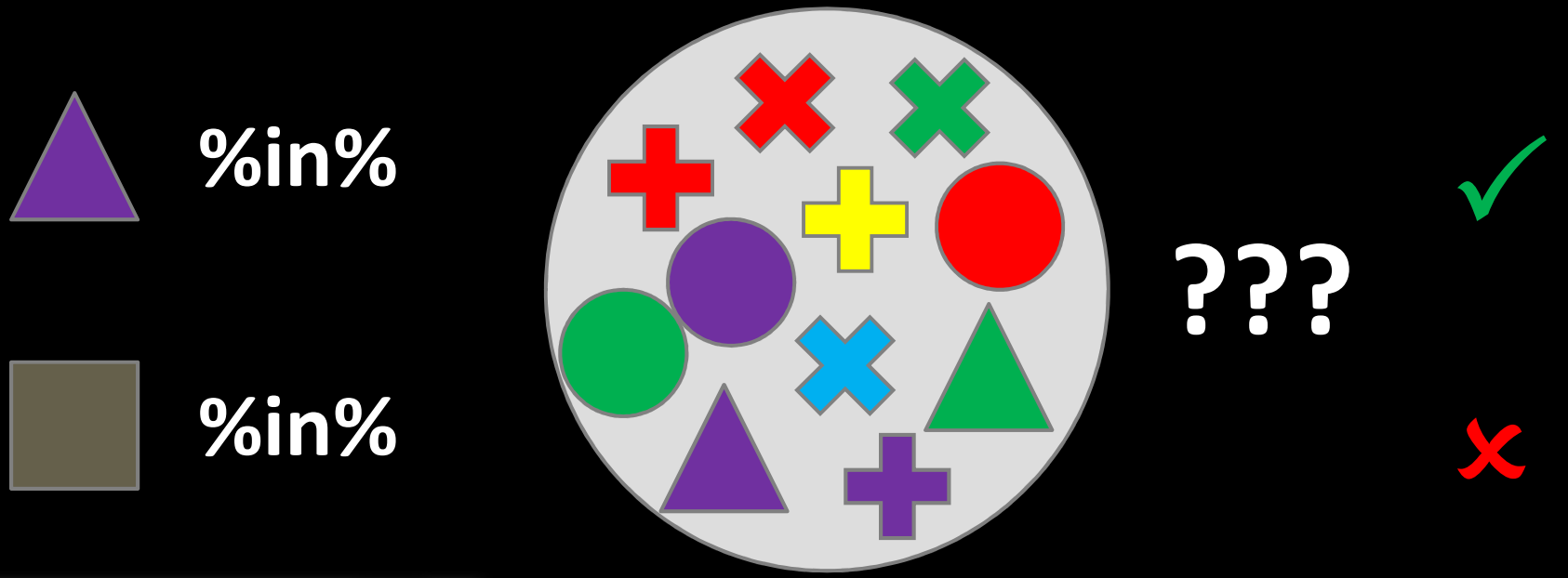
- Data Validation:
  - „Checking if a combination of values is member of a set of acceptable value combinations“

European Statistical System

Definition

- Data Validation:
  - „Checking if a combination of values is member of a set of acceptable value combinations“

European Statistical System

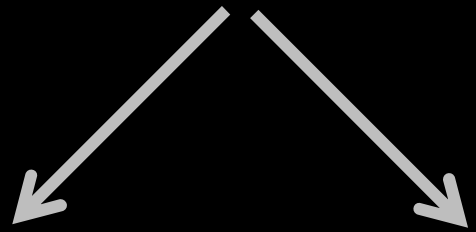


Definition

- Data Validation:
  - „Checking if a combination of values is member of a set of acceptable value combinations“

European Statistical System

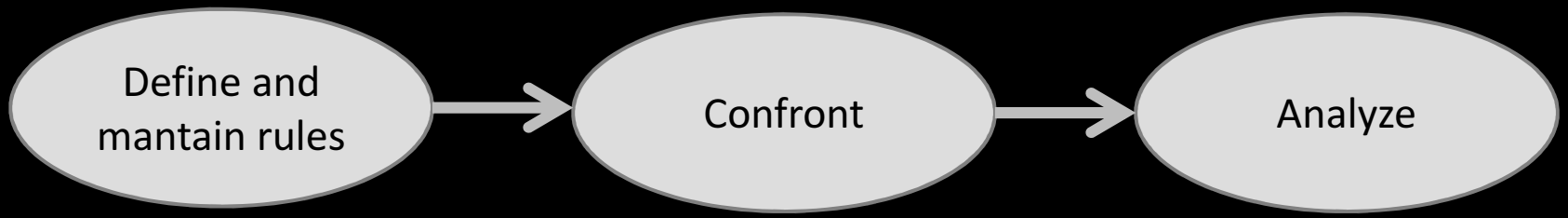
- Data validation function returns a boolean vector



TRUE **FALSE**

The {validate} Package

- Domain Specific Language



`validator()`

- free text
- yaml
- data frame
- ...

`confront()`

`summary()`
`values()`
`plot results`

Centre d'Estudis d'Opinió (CEO)
Baròmetre d'Opinió Política. 2a onada 2020
Survey Duration: 25-30 minutes
n=2000
Total population 18+ years resident in Catalonia

LET'S CODE

Links

- Centre d'Estudis d'Opinió (CEO). *Baròmetre d'Opinió Política: 2a onada 2020*. http://ceo.gencat.cat/ca/barometre/detall/index.html?id=7688
- van der Loo, Mark & de Jonge, Edwin (2019). Data Validation Infrastructure for R. *Journal of Statistical Software*. https://CRAN.R-project.org/package=validate
- Daniel Franke https://www.linkedin.com/in/daniel-f-12a750164/