



ARS TOSCANA
agenzia regionale di sanità

Generazione di grandi quantità di tabelle e grafici in formato standard per pubblicazioni cartacee

Matilde Razzanelli, Mirko Monnini, Cristina Orsini, Elena
Marchini, Simone Bartolacci, Roberto Berni, Rosa Gini

Agenzia regionale di sanità della Toscana

matilde.razzanelli@ars.toscana.it

Indice

I natali

Il metodo

Prospettive di utilizzo e sviluppo

Indice

I natali

Il metodo

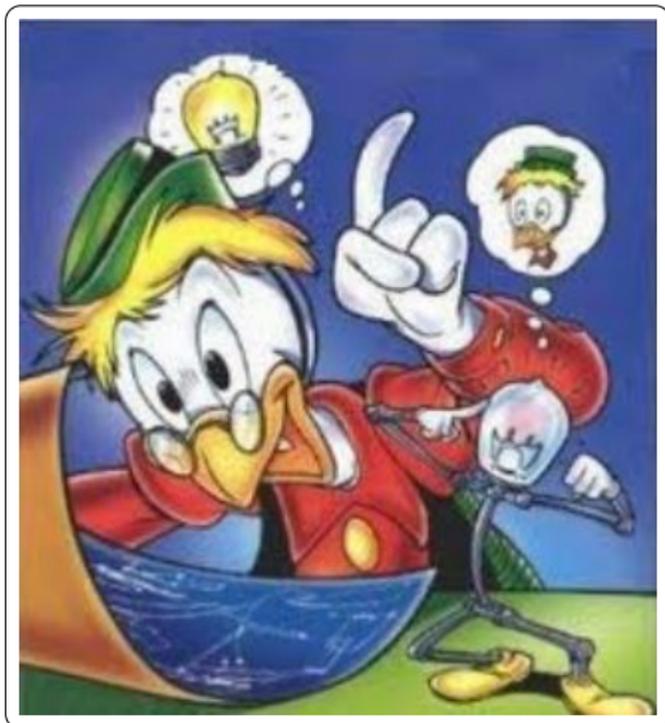
Prospettive di utilizzo e sviluppo

La procedura nasce perché

- ARS ha spesso bisogno di generare **grandi quantità** di grafici e tabelle per pubblicazioni cartacee
- questo materiale deve rispondere a dei requisiti di **uniformità** in termini di
 - dimensioni
 - colori (gamma e modalità di codifica (RGB,CMYK...))
 - label (Nomi delle ASL, Regioni ...)
 - formattazione

grafici : plotregion, presenza e caratteristiche degli assi, dei valori ecc;
tabelle : label, grassetto, griglia
- la normalizzazione del materiale prodotto dai singoli ricercatori richiede **molto lavoro di editing manuale**

L'idea: la generazione in automatico



L'idea: la generazione in automatico



Usiamo *Stata* e \LaTeX

L'idea: la generazione in automatico



creiamo un procedura

Usiamo *Stata* e $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$

L'idea: la generazione in automatico

che genera gli oggetti **finiti**



creiamo un procedura

Usiamo *Stata* e $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$

Indice

I natali

II metodo

Isolare le caratteristiche standard e variabili

Definire un *data model standard*

Realizzare le procedure

 I grafici

 Le tabelle

Prospettive di utilizzo e sviluppo

Indice

I natali

Il metodo

Isolare le caratteristiche standard e variabili

Definire un *data model standard*

Realizzare le procedure

Prospettive di utilizzo e sviluppo

Caratteristiche standard e variabili

- Standard
 - larghezza massima degli oggetti
 - tipo e dimensione dei font
 - colori
 - altre caratteristiche:
 - ordinamento dei valori in grafici e tabelle (es: ASL per codice, regioni per valore degli indicatori ecc)
 - ordine di comparsa di specifici valori (es: Maschi, Femmine, Totale)
- Variabili
 - tipologie di grafici e tabelle ammesse
 - numero massimo di dimensioni ammesse nelle tabelle (celle unite) e nei grafici

Indice

I natali

Il metodo

Isolare le caratteristiche standard e variabili

Definire un *data model standard*

Realizzare le procedure

Prospettive di utilizzo e sviluppo

Il data model

La procedura ha bisogno di dati in formato standard.

Abbiamo quindi definito le informazioni che erano necessarie suddividendole in 2 archivi:

Metadati : Codice univoco del grafico/tabella, Titolo/Caption, Tipo di grafico, Livelli...

Dati : Codice univoco del grafico/tabella, Livelli, Valori ...

Indice

I natali

II metodo

Isolare le caratteristiche standard e variabili

Definire un *data model standard*

Realizzare le procedure

I grafici

Le tabelle

Prospettive di utilizzo e sviluppo

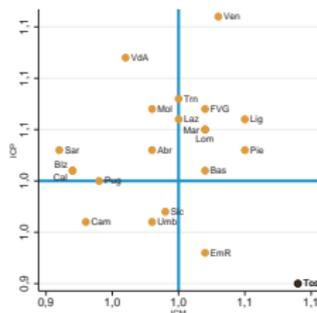
I grafici

La procedura viene gestita da un unico file *do* che

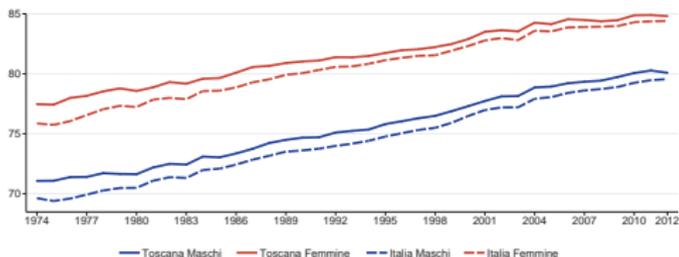
- *carica i dati e i metadati*
- *elabora i metadati in modo da derivarne i parametri necessari per la caratterizzazione degli oggetti da realizzare ed in particolare ne deriva la **tipologia di grafico***
- *per ciascuna tipologia attiva uno specifico file *do* che realizza, salva e stampa il grafico*

*I parametri generali (colori, dimensioni, aspetti della graph region ecc, sono definiti in *global* raccolte in un altro file *do**

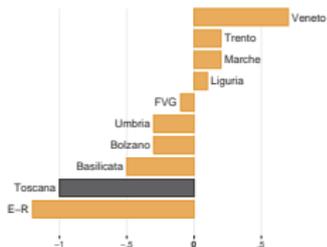
Figure: Alcuni esempi di grafici generati dalla procedura



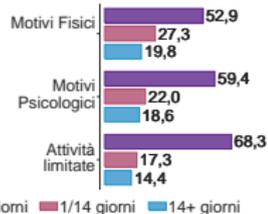
(a) Scatter, con etichette che ruotano intorno al valore nel caso più valori si sovrappongano



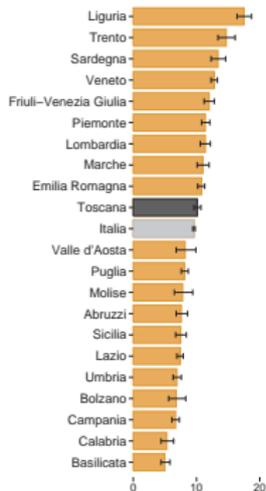
(b) Linee, con colori e tratteggi predefiniti



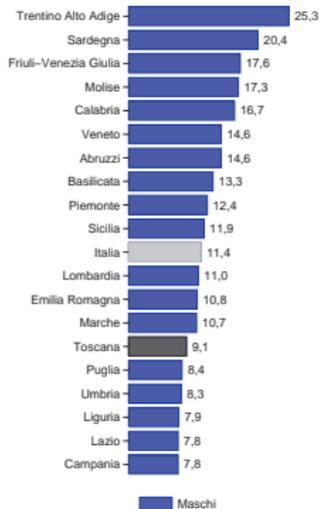
(a) Barre con valori negativi



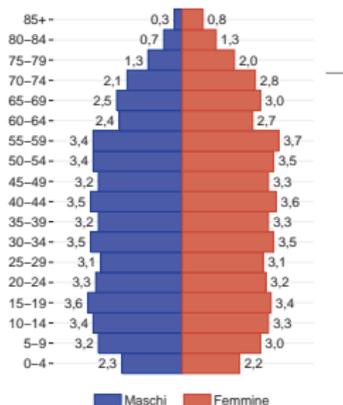
(b) Barre appaiate orizzontali



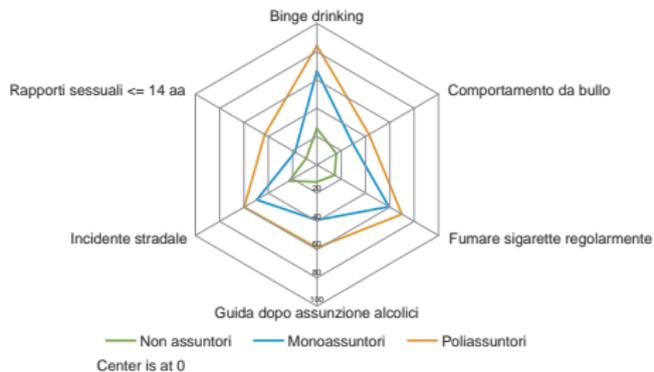
(c) Barre con asse graduato e gli intervalli di confidenza



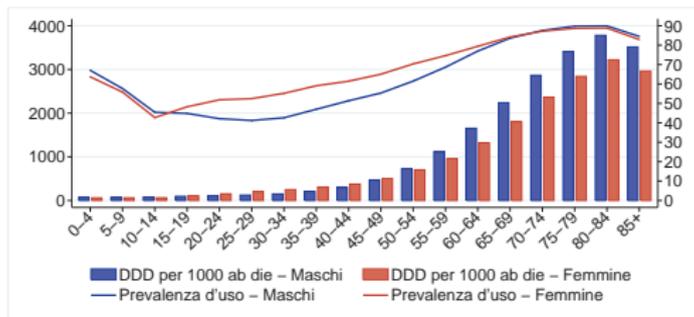
(d) Barre, grafico combinato per genere



(e) Piramide dell'età



(f) Radar



(g) Istogramma e linee con doppio asse verticale

Alcuni problemi generali

Rispondere alle esigenze degli standard fissi ha richiesto la risoluzione di alcuni problemi:

- **dimensione del font:** in *Stata* la dimensione del font è proporzionale alla dimensione maggiore del grafico. Avendo noi posto una larghezza massima abbiamo dovuto trovare il modo di scalare correttamente il font in modo che rimanesse di una dimensione data anche quando la lunghezza del grafico superava la larghezza
- **colori e tipologia di linee:** abbiamo parametrizzato l'uso di determinati colori e linee che dovevano rimanere costanti in tutta la pubblicazione (maschi- femmine; Italia-Toscana ecc)
- **colori in quadricromia**

Alcuni problemi specifici

- **altezza delle barre costante** parametrizzata per essere indipendente (come il font) dalla dimensione del grafico stesso.
- **la distanza tra gruppi di barre** nelle barre appaiate.
- **aspetto della legenda** parametrizzazione del numero di righe/colonne della legenda in base alla lunghezza delle label
- **xlabel/ylabel** parametrizzazione del numero di righe in cui distribuire le label testuali perchè il grafico non venga sproorzionato rispetto ad esse (ylabel) o perchè non si sovrappongono (xlabel)
- **xscale/yscale** parametrizzazione delle scale da utilizzare in base ai range di valori e delle etichette
- **la posizione delle etichette** negli scatter: rotazione delle etichette nei casi di punti sovrapposti

Un caso a parte: le Mappe

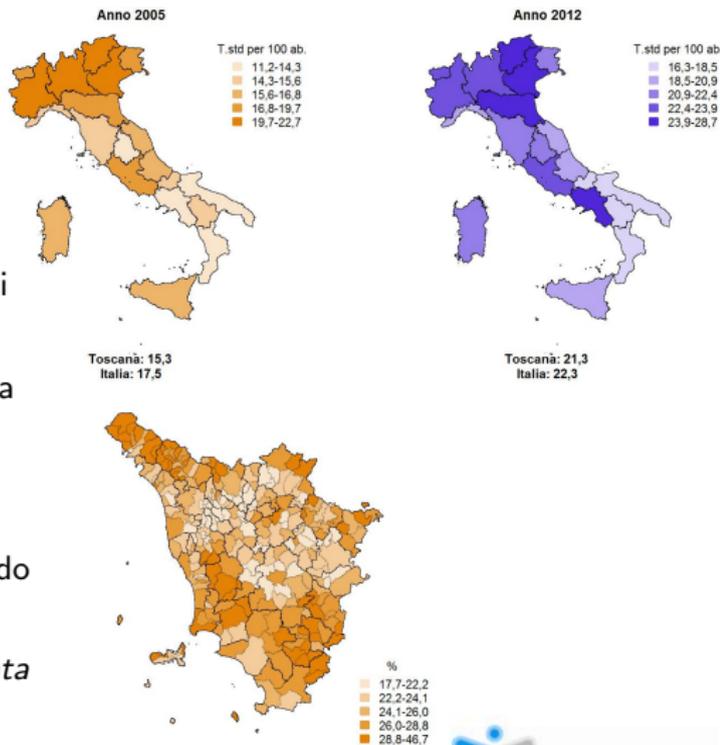
Per le mappe abbiamo sfruttato un *modulo R* già sviluppato in ARS.

Come?

Stata prepara in apposite cartelle

- i dataset con dati e metadati già pre-elaborati
- dei file *r* generati da *Stata* (file di testo con estensione *.r*)
 - l'identificativo del grafico da realizzare
 - i parametri relativi a colori, dimensioni...
- lancia il *modulo R* con il comando *shell*

In prospettiva passeremo all'uso di *Stata* per la realizzazione delle mappe



Le tabelle

La procedura viene gestita da un unico file.do che

- carica i dati e i metadati
- elabora i metadati in modo da derivarne i parametri necessari per la caratterizzazione degli oggetti da realizzare ed in particolare ne deriva **il numero di colonne**
- successivamente un unico file *do* produce la tabella scritta in \LaTeX e inserita in un ambiente *standalone*
- l'output finale è un *pdf* già della giusta dimensione che può essere inserito nella pubblicazione come immagine

Le tabelle

La realizzazione delle tabelle avviene

- manipolando i dati in modo da generare un dataset *Stata* analogo a quello della tabella finale,
- manipolando ulteriormente il codice *Stata* che genera il codice \LaTeX della tabella in modo da gestire:
 - *multirow*
 - grassetto
 - larghezza delle colonne
- generando, tramite manipolazione dei metadati, le intestazioni delle colonne \LaTeX
- stampando poi la tabella tramite l'uso del comando *listtex*

Alcuni esempi di tabelle

AUSL	Casi Incidenti			Casi Prevalenti		
	Maschi	Femmine	Totale	Maschi	Femmine	Totale
1 - Massa	682	658	1.340	4.361	5.574	9.935
2 - Lucca	753	705	1.458	4.849	5.982	10.831
3 - Pistoia	955	904	1.859	6.160	7.666	13.826
4 - Prato	753	720	1.473	4.851	6.103	10.954
5 - Pisa	1.101	1.049	2.150	7.087	8.898	15.985
6 - Livorno	1.208	1.131	2.339	7.743	9.592	17.335
7 - Siena	919	865	1.784	5.982	7.341	13.323
8 - Arezzo	1.155	1.058	2.213	7.484	8.984	16.468
9 - Grosseto	787	735	1.522	5.060	6.233	11.293
10 - Firenze	2.760	2.663	5.423	17.818	22.607	40.425
11 - Empoli	763	712	1.475	4.953	6.039	10.992
12 - Viareggio	542	536	1.078	3.466	4.540	8.006
Toscana	12.378	11.736	24.114	79.814	99.559	179.373

(h) Tabella a 1 livello orizzontale e 2 verticali, con impostazione in automatico dei grassetti

	Titolo di Studio	Sottopeso	Normopeso	Sovrappeso	Obeso
Maschi	Elementare/nessuno	0,0	38,9	47,9	13,2
	Licenza media inferiore	0,0	40,0	50,0	10,0
	Licenza media superiore	1,8	53,3	38,1	6,8
	Laurea	0,7	46,1	48,1	5,1
	Elementare/nessuno	2,7	43,9	39,4	14,0
Femmine	Licenza media inferiore	4,2	57,1	29,6	9,1
	Licenza media superiore	6,9	67,7	19,9	5,5
	Laurea	5,2	72,4	17,8	4,6

(i) Tabella a 2 livelli orizzontali e 1 verticali, con multirow

Alcuni esempi di tabelle

	Specie animale	Macellati		Morti		Abbattuti		Totale	
		Esaminati	Positivi	Esaminati	Positivi	Esaminati	Positivi	Esaminati	Positivi
2009	Bovini e bufalini	886	0	1.015	0	0	0	1.901	0
	Ovini	569	0	1.317	4	456	5	2.342	9
	Caprini	72	0	120	0	22	0	214	0
2010	Bovini e bufalini	873	0	896	0	0	0	1.769	0
	Ovini	484	1.251	3	0	0	0	1.735	3
	Caprini	122	0	225	0	0	0	347	0
2011	Bovini e bufalini	707	0	867	0	0	0	1.574	0
	Ovini	480	0	1.237	7	150	16	1.903	23
	Caprini	87	0	143	0	0	0	230	0
2012	Bovini e bufalini	480	0	857	0	0	0	1.337	0
	Ovini	552	1	1.112	1	136	13	1.800	15
	Caprini	216	1	168	0	13	0	397	1
2013	Bovini e bufalini	320	0	950	0	0	0	1.270	0
	Ovini	408	0	1.188	2	9	2	1.605	4
	Caprini	352	0	181	0	1	0	534	0

(j) Tabella a 2 livelli orizzontali e 2 verticali, con multirow

Alcuni problemi risolti

- **Tabella Standalone:** utilizzo del pacchetto \LaTeX *Standalone* per generare una tabella ritagliata utilizzabile come figura
- **Dimensionamento delle colonne di intestazione di riga:** ha richiesto un utilizzo avanzato del codice \LaTeX per relativizzare le dimensioni delle colonne alla lunghezza del testo che deve contenere.
- **Automatizzazione dell'utilizzo del grassetto**

Indice

I natali

Il metodo

Prospettive di utilizzo e sviluppo

Utilizzo e sviluppo

La procedura sviluppata può essere usata per generazione di grafici e tabelle di buona qualità

- per pubblicazioni o presentazioni formali
- per report generati in automatico con *Stata*

Stiamo lavorando ad un'**interfaccia** avanzata che:

- consenta l'inserimento/scelta da maschera dei metadati (salvarti poi in apposito file *do*)
- consenta l'inserimento dei dati in formato non standard
- trasformi i dati nel formato richiesto dalla procedura
- lanci la procedura e realizza l'oggetto

Questo faciliterebbe l'utilizzo della procedura e ne allargherebbe ulteriormente la platea soprattutto per la realizzazione dei grafici più complessi (barre con linee, mappe ecc)

Riferimenti

- Gini R. **“Gestione di processi complessi da Stata”**
<http://econpapers.repec.org/paper/bocisug09/07.htm>
- Gini R., Pasquini J. **“Automatic generation of documents”**
http://econpapers.repec.org/article/tsjstataj/v_3a6_3ay_3a2006_3ai_3a1_3ap_3a22-39.htm