

La statistica per i curiosi

Giovanni A. Barbieri – Istat
IV giornata per la diffusione della statistica
Firenze – Centro Didattico Polivalente Morgagni

Swivel

La statistica per i curiosi



25 maggio 2007

G. A. Barbieri

2

I fenomeni collettivi (1)

- Stai guidando nella nebbia e la macchina davanti a te inchioda
- Ti attacchi al clacson e pensi “ma che fai? Levati di mezzo, pericolo pubblico”
- Poi la nebbia si solleva, e vedi che davanti a te, a perdita d’occhio, c’è una fila di macchine ferme
- Di colpo, pensare che quello davanti a te sia un cretino o che gli si è rotta la macchina diventa un’idiozia
- Non serve più la psicologia o la competenza di un meccanico
- Il livello a cui devi cercare la spiegazione diventa un altro: che cosa ha causato l’ingorgo? Un incidente? Lavori in corso più avanti? Un temporale?
- È necessario fare un’**astrazione**. Passare da una spiegazione causale individuale e locale, allo studio di un **fenomeno collettivo**

I fenomeni collettivi (2)

- La macchina davanti a te è indubbiamente la causa immediata del tuo stop, ma al tempo stesso è **irrilevante** per il fenomeno che vuoi capire, è una pedina in un gioco più grande di lei
- Capire la causa di quello che sta succedendo richiede di **spostarsi di livello**. Il ruolo delle singole macchine resta reale, ma per capire davvero devi prendere in considerazione delle strutture più grandi e le loro interazioni più astratte
- Qui entra in gioco la **statistica**

Statistica e termodinamica

- Un momento chiave nello sviluppo della statistica come scienza è stato la sua applicazione alla **termodinamica**
- Pensiamo – tanto per restare in tema – a quello che succede nel motore: scoppia la scintilla, il gas scaldandosi si dilata e *quindi* spinge il pistone
- Questo è quello che accade al livello della termodinamica
- Quello che accade alle singole molecole del gas, al livello micro resta assolutamente **vero**, ma è **irrilevante**

Meccanica statistica

- La meccanica statistica è l'applicazione della **statistica**, che include strumenti matematici per gestire insiemi formati da numerosi elementi, al campo della **meccanica**, che si occupa del moto di oggetti quando soggetti a una forza, e in particolare dello studio di sistemi composti da molte particelle (sistemi termodinamici come i gas perfetti). Per lo studio di questi sistemi è necessario l'approccio statistico per 2 motivi, uno pratico e uno teorico:
 - Lo studio di un sistema con N particelle non interagenti richiede la soluzione di N equazioni differenziali (le equazioni del moto di ogni particella). Ma una mole di gas contiene un numero di Avogadro di particelle, cioè circa 6×10^{23} particelle. La difficoltà nel determinare la dinamica di un sistema a molte particelle non sta solo nell'elevato numero di equazioni da risolvere, ma anche nell'impossibilità di risolverle analiticamente
 - Per risolvere le 10^{23} equazioni è necessario stabilire le condizioni iniziali (posizione e quantità di moto all'istante di tempo iniziale). Questo è un problema insolubile: il principio di indeterminazione di Heisenberg ci dice che non è possibile conoscere con esattezza e contemporaneamente posizione e quantità di moto di una particella
- La meccanica statistica fornisce un **modello** per collegare le proprietà di atomi singoli e molecole alle proprietà macroscopiche dei materiali che vediamo nella vita quotidiana, spiegando dunque la termodinamica (**livello macro**) come un risultato di statistica e meccanica (**livello micro**)

Che cos'è la statistica

- Scienza che ha per oggetto lo studio dei **fenomeni collettivi** suscettibili di **misurazione** e di **descrizione quantitativa**
 - Specialmente quando il numero degli individui interessato è talmente elevato da escludere la possibilità o la convenienza di seguire le vicende di ogni singolo individuo
 - Si perviene alla formulazione di leggi di media che governano tali fenomeni, dette **leggi statistiche**
 - basandosi sulla **raccolta di un grande numero di dati** inerenti ai fenomeni in esame, e partendo da **ipotesi** più o meno direttamente suggerite dall'esperienza o da **analogie** con altri fenomeni già noti
 - mediante l'applicazione di **metodi matematici** fondati sul calcolo delle probabilità
 - Spesso la raccolta dei dati viene limitata a un **campione** più ristretto, opportunamente predeterminato in modo da rappresentare fedelmente le caratteristiche generali
- Concepita inizialmente come attività descrittiva di certi fatti sociali e in particolare come attività amministrativa dello Stato, ha via via ampliato i suoi confini, fino a diventare una vera e propria «**scienza del collettivo**», disciplina con finalità non solo descrittive dei fenomeni sociali e naturali, ma orientata anche a finalità di ricerca

Alcuni punti fermi

- Rapporto tra “**statistica**” e “**fenomeni collettivi**”: il tipo di incertezza che la statistica cerca di tenere sotto controllo è insita nel numero degli individui o dei fenomeni
- Semplificare la complessità attraverso un **modello** – basato sul calcolo delle probabilità – capace di **conservare gran parte dell’informazione riducendo drasticamente la molteplicità** dei fenomeni
- Misurabilità
- Rilevanza

Misurabilità

- **Misura:** Il valore numerico attribuito a una grandezza, ottenuto ed espresso come **rapporto** tra la grandezza data e un'altra della stessa specie assunta come unità (*unità di misura*), e determinato con opportuni metodi o strumenti di misurazione
- **Misurazione:** L'operazione del misurare, consistente nel confrontare una determinata grandezza fisica con la sua unità di misura, allo scopo di determinare il valore (o misura) della grandezza stessa [...]
 - **diretta** (o fondamentale, o relativa), quella che consente di determinare direttamente la misura di una grandezza (senza quindi far ricorso alla misurazione di altre grandezze), **confrontandola con un campione** (per es., la misurazione della massa di un corpo mediante una bilancia, sull'altro piatto della quale vengono poste masse campione)
 - **indiretta** (o derivata), quella in cui la grandezza da misurare dipende, secondo una **relazione funzionale nota**, da altre grandezze misurate direttamente (per es., la misurazione della velocità media di un corpo come rapporto tra lo spazio percorso e il tempo impiegato a percorrerlo)

Rilevanza

- Il fatto, la caratteristica di essere rilevante, cioè di notevole importanza o anche gravità, soprattutto riguardo a determinati fini
- *Relevance is term used to describe how pertinent, connected, or applicable some information is to a given matter*
 - Concetto importante ma elusivo:
 - La soluzione di un problema richiede l'individuazione *ex ante* degli elementi rilevanti che possono concorrere alla sua soluzione
 - Ma la logica convenzionale non lo cattura (una proposizione falsa implica tutte le altre proposizioni, ma non tutte sono rilevanti)
 - **La rilevanza è funzione di un obiettivo** (*goal dependent*): un elemento (oggetto o proposizione) è rilevante per un obiettivo se e solo se è essenziale all'interno di un piano per conseguirlo (Gorayska e Lindsay)
 - Soddisfa sia le esigenze della logica sia quelle delle attività di problem-solving
 - È **definita empiricamente**, non sulla base dello stato delle conoscenze o delle credenze

Modello

- Un modello astratto (o concettuale) è una costruzione teorica che rappresenta processi fisici, biologici o sociali, con un insieme di **variabili** e un insieme di **relazioni logiche e quantitative** tra loro
- In questa accezione, il modello consente di ragionare all'interno di uno schema logico **astratto** e **semplificato**:
 - Astratto (idealizzato) perché il modello può formulare ipotesi esplicite di cui è noto che – a un certo livello di dettaglio – sono false
 - Semplificato perché ciò consente di pervenire a soluzioni ragionevolmente accurate, trascurando la complessità implicita nel grande numero di variabili e attori del processo modellizzato

42: l'importanza dei metadati

- Questa è una **storia** che illustro spesso per spiegare che i dati senza metadati non servono a niente
- Secondo la *Guida galattica per gli autostoppisti* di Douglas Adams (trilogia in cinque parti pubblicata in Italia da Mondadori), gli scienziati di una specie di esseri super-intelligenti multi-dimensionali costruirono il più grande computer di tutto lo spazio e di tutti i tempi, *Deep Thought*, affinché calcolasse la risposta ultima della questione fondamentale sulla vita, l'universo e tutto quanto. Dopo sette milioni e mezzo di anni di calcolo, il computer diede la risposta: 42
- “42!” - gridò Loonquawl, il capo degli scienziati - “Tutto qui il risultato di sette milioni e mezzo di anni di lavoro?”
- “Ho controllato con scrupolo: la risposta è esatta” - rispose il computer - “Ma se devo essere onesto, il problema è che non avete mai saputo quale fosse la domanda!”

La classificazione dei bassotti

- In un famoso racconto, una creatura di Cortázar decide di classificare in modo definitivo i bassotti
- Individuato il primo gruppo, formato da 8 bassotti, si accorge che deve a sua volta suddividerlo in tre sottogruppi - “bassotti baffuti, bassotti tipo pugile e bassotti stile segretario di ministero” – composti rispettivamente di 3, 3 e 2 bassotti
- Separatili sulla base della nuova suddivisione, si rende conto però che il primo sottogruppo non è omogeneo, “perché due bassotti baffuti appartenevano al tipo roditore, mentre quello che restava era senza alcun dubbio un bassotto di taglio giapponese”
- Messo da parte quest’ultimo, si accinge ad annotare le caratteristiche del sottogruppo dei due roditori nella cartella dei suoi lavori scientifici, quando si girano di profilo: “mentre il primo roditore era un bassotto brachicefalo, l’altro bassotto metteva in evidenza un cranio molto più adatto per appenderci un cappello che per calzarlo”
- “Fu così – conclude Cortázar – che il sottogruppo le si dissolse tra le mani; quanto al resto, non vale neppure la pena di parlarne”

La mappa dell'impero

- Tutti citano questo racconto di Borges a memoria, ed è difficile da ritrovare nella sua sterminata produzione. Ecco qui:
- “In quell’Impero l’arte della cartografia raggiunse tale perfezione che la mappa d’una sola provincia occupava tutta la città, e la mappa dell’Impero tutta una provincia. Col tempo codeste mappe smisurate non soddisfecero e i colleghi dei cartografi eressero una mappa dell’Impero che uguagliava in grandezza l’Impero e coincideva puntualmente con esso. Meno dedite allo studio della cartografia, le generazioni successive compresero che quella vasta mappa era inutile e non senza empietà l’abbandonarono alle inclemenze del sole e degl’inverni. Nei deserti dell’Ovest rimangono lacere rovine della mappa, abitate da animali e mendichi; in tutto il Paese non è altra reliquia delle discipline geografiche.”

Somiglianze di famiglia (1)

- Due brani di Wittgenstein (*Ricerche filosofiche*):
- Consideratelo un commento a *La classificazione dei bassotti*.
- “Considera, ad esempio, i processi che chiamiamo “giochi”. Intendo giochi da scacchiera, giochi di carte, giochi di palla, gare sportive, e via scorrendo. Che cosa è comune a tutti questi giochi? – **non dire: “deve esserci qualcosa di comune a tutti, altrimenti non si chiamerebbero ‘giochi’ “– ma guarda se ci sia qualcosa di comune a tutti.** – Infatti, se li osservi, non vedrai certamente qualche cosa che sia comune a *tutti*, ma vedrai somiglianze, parentele, e anzi ne vedrai tutta una serie. Come ho detto: **non pensare, ma osserva!** – Osserva, ad esempio, i giochi da scacchiera, con le loro molteplici affinità. Ora passa ai giochi di carte: qui trovi molte corrispondenze con quelli della prima classe, ma molti tratti comuni sono scomparsi, altri ne sono subentrati. Se ora passiamo ai giochi di palla, qualcosa di comune si è conservato, ma molto è andato perduto. Sono tutti *‘divertenti’*? Confronta il gioco degli scacchi con quello della tria oppure c’è dappertutto un perdente o un vincente o una competizione tra giocatori? Pensa allora ai solitari. Nei giochi con la palla c’è vincere e perdere; ma quando un bambino getta la palla contro un muro e la riacchiappa, questa caratteristica è sparita. Considera quale parte abbiano abilità e fortuna. E quanto sia differente l’abilità negli scacchi da quella nel tennis. Pensa ora ai girotondi: qui c’è l’elemento del divertimento, ma quanti degli altri tratti caratteristici sono scomparsi! E così possiamo passare in rassegna molti altri gruppi di giochi. Veder somiglianze emergere e sparire. E il risultato di questo esame suona: **Vediamo una rete complicata di somiglianze che si sovrappongono e si incrociano a vicenda.** Somiglianze in grande e in piccolo.”

Somiglianze di famiglia (2)

- “Non posso caratterizzare queste somiglianze meglio che con l’espressione “somiglianze di famiglia”; infatti le varie somiglianze che sussistono tra i membri di una famiglia si sovrappongono e si incrociano nello stesso modo: corporatura, tratti del volto, colore degli occhi, modo di camminare, temperamento, ecc. ecc. – E dirò: i ‘giochi’ formano una famiglia. E nello stesso modo formano una famiglia, ad esempio, i vari tipi di numeri. Perché chiamiamo una certa cosa ‘numero’? Forse perché ha una – diretta – parentela con qualcosa che finora si è chiamato numero; e in questo modo, possiamo dire, acquisisce una parentela indiretta con altre cose che chiamiamo anche così. Ed estendiamo il nostro concetto di numero così come, nel tessere un filo, intrecciamo fibra con fibra. E **la robustezza del filo non è data dal fatto che una fibra corre per tutta la sua lunghezza ma dal sovrapporsi di molte fibre una all’altra**. Se però qualcuno dicesse: “Dunque c’è qualcosa di comune a tutte queste formazioni, – vale a dire la disgiunzione di tutte queste comunanze” – io risponderei: qui ti limiti a giocare con una parola. Allo stesso modo si potrebbe dire: un qualcosa percorre tutto il filo, – cioè l’ininterrotto sovrapporsi di queste fibre.”

Swivel



- Che cos'è Swivel?
- Swivel è un sito dove le persone curiose possono esplorare qualunque tipo di dati
- Swivel è nato in California il 6 dicembre 2006
- È ancora un'anticipazione:
“Come anticipazione, non è ancora del tutto rifinita, ma piena fino all'orlo di buoni dati saporiti.
Lasciatevi guidare dal vostro amore per i dati”

Com'è?

- Giusto per darvi un'idea ...

The screenshot shows the Swivel website interface. At the top, there is a navigation bar with links for Home, Graphs, Data, People, and Upload. A search bar is also present. Below the navigation, a banner reads "Swivel is a place where curious people explore all kinds of data." and "As a preview it's rough around the edges, yet brimming with tasty data goodies. May your love for data guide you."

The main content area features a "Spotlight" section with the title "Running On Empty". It displays a scatter plot of "City MPG" (Y-axis, 10 to 40) versus "Engine Size (L)" (X-axis, 2.5 to 6.5). The plot shows a negative correlation between engine size and city MPG. A red line is drawn through the data points, and a white circle highlights a cluster of points in the top-left corner, labeled "E". A white circle highlights a cluster of points in the bottom-right corner, labeled "F". A white circle highlights a cluster of points in the middle-right area, labeled "V2".

Source: [Environmental Protection Agency](#)

This great scatter plot by [Solarwind](#) shows the average city miles per gallon for 2007 SUVs versus their engine size in liters. Check out all of the great [data](#). Notice the hybrid SUVs are outliers in the top left corner.

Below the spotlight, there is a "Featured Graphs" section with a link to "Paper Doesn't Grow on Trees".

On the right side, there are two sections: "2617 Data Sets" and "1201980 Graphs". The "Data Sets" section lists three items: "Pirates versus global temperature" (Contributed by [ehamiter](#) about 6 hours ago), "Top 100 World Newspapers" (Contributed by [m4tts1m](#) about 17 hours ago), and "Distribution of Pee Frequency" (Contributed by [Eximio](#) 1 day ago). The "Graphs" section lists two items: "Austria by Albania" (Created by [seema](#) from [Paper Consumption](#) data 27 days ago) and "Saudi Arabia and Chile" (Created by [seema](#) from [Paper Consumption](#) data 27 days ago).

Swivel è un sito dove le persone curiose esplorano i dati

- Swivel ti permette di esplorare i dati e condividere con altri le tue conclusioni. Swivel ha dati sulla politica, l'economia, il tempo, gli sport, gli affari e altri ancora
- Swivel è pieno di roba. Gli autori la pensano così:
 - **Esplora** i dati, dai più comuni ai più oscuri. Cerca quello che vuoi, oppure naviga semplicemente tra i grafici a colori, i data-set e le opinioni
 - **Confronta** i prezzi della benzina e i sondaggi sulla politica del governo, gli avvistamenti degli UFO con le vendite di iPod. Potresti trovare una coincidenza folle o fare una scoperta importante
 - **Condividi** le tue scoperte con il post di un grafico sul tuo blog o inviando un link per email ai tuoi amici o ai compagni di scuola
 - **Carica** le informazioni che ti interessano, descrivile, scegli i colori del tuo grafico, aggiungi un'immagine o una foto
- Se sei **uno curioso dei dati**, Swivel è il posto per te

Esplora: lo so solo se lo vedo

- C'è spesso che **non sai che cosa stai cercando finché non lo vedi**. Swivel ti dice che, se stiamo parlando di dati, è proprio questo il punto
- In **Swivel** gironzolare tra tabelle di dati e grafici è **facile** come navigare sul web: un click dopo l'altro
 - Esplora, **imparando** mentre lo fai
 - **Naviga** tra i grafici con una scelta vastissima d'argomenti
 - **Ordina** dati e grafici secondo i tuoi interessi (i più visitati, le novità, i più apprezzati ...)
 - **Salva** quelli che ti interessano di più in una lista che gestisci tu
- Clicca da un dato all'altro, guarda il lavoro dei tuoi autori preferiti, curiosa tra i grafici del giorno: con **Swivel** esplorare i dati è **divertente**

Browse

Most Viewed

Most Discussed

Most Recent

By Time

Today

This Week

This Month

All Time

Popular Tags

All

000

001

abortion

age

all

birth

by

campaign

capita

causes

column

consumption

contributions

crime

cut

Most Viewed Data Sets (All Time)



Major League Baseball Teams

Views: 7784
Created: about 2 months ago
Created By: [Brian Mulloy](#)



Airline On Time Performance

Views: 6111
Created: about 3 months ago
Created By: [seema](#)



Bush Ratings

Views: 4443
Created: about 3 months ago
Created By: [Dmitry](#)



US Electricity Costs

Views: 132
Created: about 3 months ago
Created By: [Dmitry](#)



Brent Crude Oil

Views: 127
Created: about 4 months ago
Created By: [Dmitry](#)



Per Capita Beverage

Views: 90
Created: about 3 months ago
Created By: [seema](#)



Confronta: una coincidenza? O qualcosa di più?

- Se ti è capitato di chiederti **se e come due cose diverse sono correlate** – ad esempio, felicità e salute, crimine e povertà, prezzo delle case e clima – Swivel è fatto per te, e tu sei fatto per Swivel
- Con Swivel, puoi confrontare qualunque tipo di dati con pochi click e Swivel ti dirà se e con quale intensità i dati sono correlati
 - **Confronta** qualunque tipo di dati in un grafico solo
 - Osserva l'**indicatore di correlazione** per vedere l'intensità del collegamento
 - Trova nuove **idee** grazie ai **suggerimenti**
- Se vedi qualcosa di interessante, potrebbe essere soltanto una coincidenza o qualcosa di più. Con Swivel, confrontare i dati e metterli in relazione è **facile**: puoi sperimentare tutti i confronti che ti vengono in mente, finché **trovi** qualcosa d'interessante

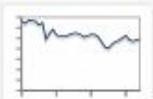
Add an item to compare

coffee

Related Recent Saved Search

Showing: 1 - 12 of 15 matched graphs.

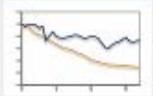
◀ 1 2 ▶



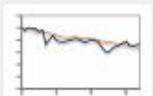
Coffee



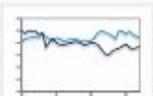
Coffee



Whole Milk and Coffee



Milk and Coffee



Tea and Coffee



Coffee and Bottled Water

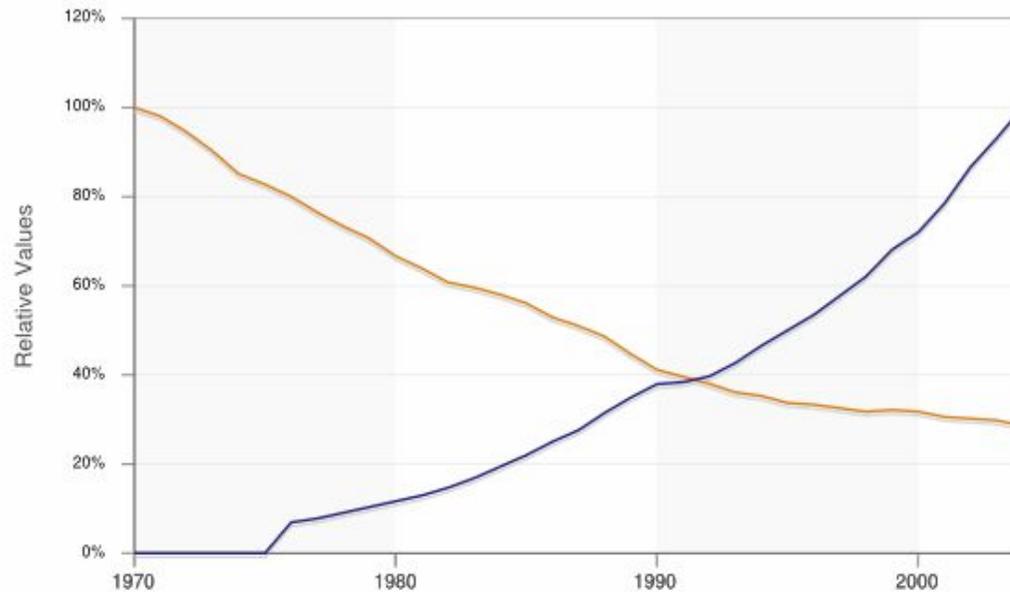


Coffee and Carbated Soft Drinks

Whole Milk x Bottled Water x

Whole Milk and Bottled Water

● Per Capita Beverage Whole Milk ● Per Capita Beverage Bottled Water



View Edit Blog Full Screen Add to favorites Feature

Graph Data Table

How related are these items to one another?



↑ Whole Milk
↓ Bottled Water

Condividi

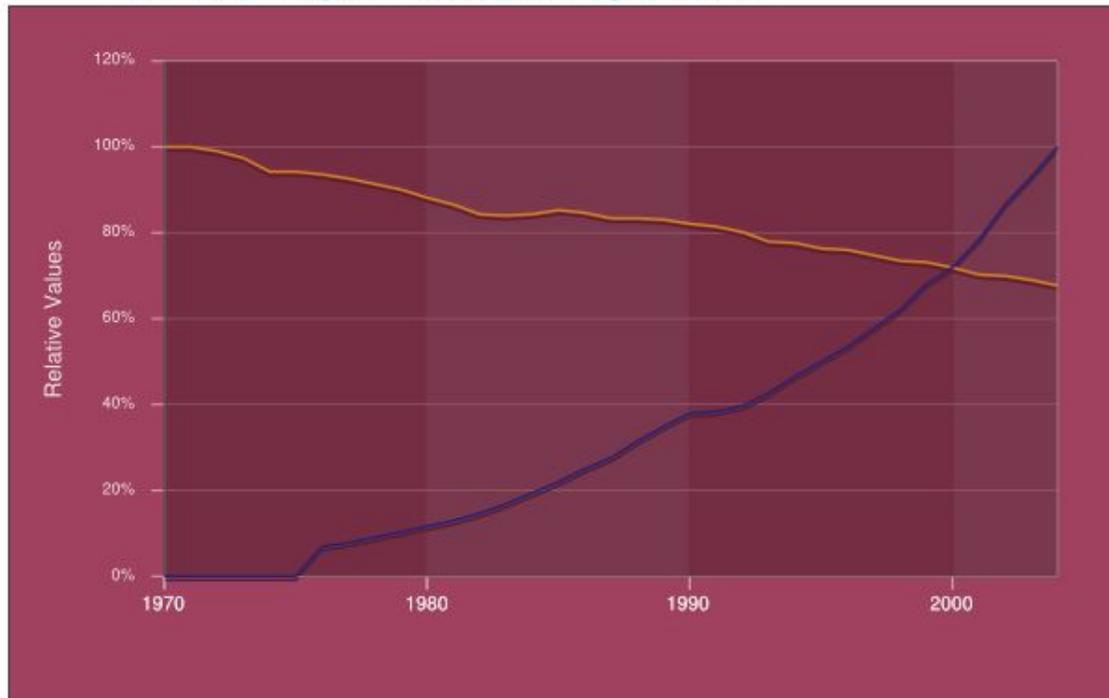
- Quando hai trovato qualcosa di interessante, condividilo con gli altri. Con Swivel è facile aggiungere un grafico al tuo blog o mandarlo per email. Quando gli altri lo vedono, il grafico non è soltanto un'immagine. Possono cliccarlo e cominciare a loro volta un nuovo ciclo – esplorare, confrontare, condividere:
 - Aggiungi un grafico al tuo **blog**
 - Invia un link per **email** a chi vuoi
 - Gli altri possono **cliccare sul tuo grafico** e cominciare a esplorare e confrontare i dati sottostanti
 - Tieni **traccia** delle visite: quante persone e per quante volte hanno visto e cliccato il tuo lavoro?

Save Changes [Cancel](#)

Type: **Line** | Scale: **Relative to Range** | Date Range: **All** Jan 01 1970 to Jan 01 2004 | Time Scale: **Yearly** | Display: **Average**

Bottled Water Takes Market from Cows **ok** **cancel**

● Per Capita Beverage Milk ● Per Capita Beverage Bottled Water



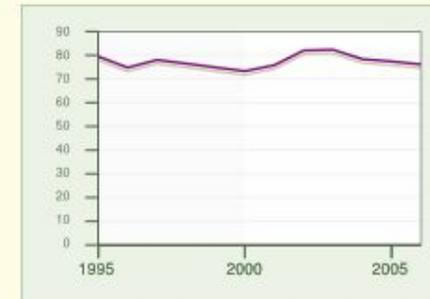
View Compare Blog Full Screen

Description

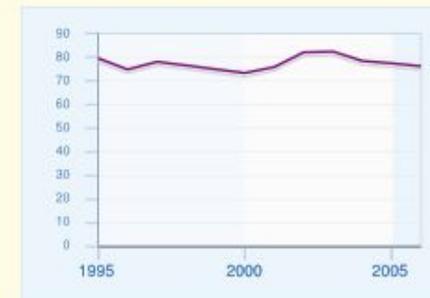
This is an interesting side effect of all the bottled water we have been drinking.

Pick a Graph Style

Graph Styles [Custom Templates](#)



Blueprint:



Black:



Carica

- Cerchi dei dati e non li trovi? Puoi caricarli tu stesso su Swivel. Una parte del divertimento sta qui:
 - Scegli i **colori**
 - **Descrivi** i dati
 - Mettici un'**etichetta**
 - Scegli una **foto** per illustrarli meglio
 - Tieni **traccia** delle visite
- Su Swivel puoi diventare la **persona di riferimento** per i dati che ti interessano di più

Describe the data.

Name and Description

Name

Description

Source

Name (ex. US Census, NASDAQ, Heard on Radio)

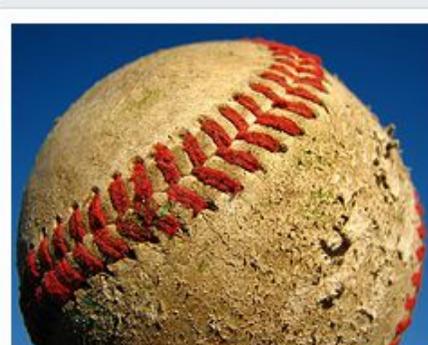
Web site (ex. http://census.gov, http://nasdaq.com)

Where's this from? Please take a few moments to let folks know where you got your data. Even if it's just 'I heard it on the radio,' when you share your source it makes your data more trustworthy. Plus, it's nice to give credit where credit is due.

Tags

toledo mudhens minorleaguebaseball mudhens

Image



[clear](#)

Please click on a photo below.

The image you select will help bring your data to life. We are big fans of Flickr, an awesome photo-sharing Web site, and have included a little photo browser below so you can find just the right image.

Go nuts, you can change it up later.

flickr GAMMA **Search photos**



[flickr](#)



[flickr](#)



[flickr](#)

A conti fatti

- **Swivel è un sito Web site per chi è curioso di esplorare i dati**
 - **Divertente:** un caleidoscopio di tabelle e grafici; tutti gli strumenti del web 2.0 (commenti, classifiche, comunità degli utenti, crocevia del dibattito); uno spazio per le opinioni più azzardate e controverse: bugie, dannate bugie e Swivel
 - **Esplorazione dei dati:** dati d'ogni tipo, l'aiuto per trovarli (dati simili, popolari, nuovi, raccomandati ...) confrontarli (risultati immediati e interattivi) personalizzarli (il tuo Swivel e le tue tracce)
 - **Sito web:** facile come un tostapane, facile come internet: non un grigio magazzino aziendale, ma un negozio di prelibatezze!
 - **Una comunità di curiosi:** gente come te, che si fa delle domande e prova a dare delle risposte, gente che cerca le storie dietro i numeri!

GRAZIE DELL'ATTENZIONE!

Giovanni A. Barbieri
barbieri@istat.it