
Fonti, metodi e strumenti per l'analisi del territorio

Metodologia della ricerca socio-economica locale

Giovanni A. Barbieri

12-13 dicembre 2007

Rappresentare i dati: un breve corso di bricolage

- Scrivere una relazione
- Scrivere di numeri
- Tabelle
- Grafici

Scrivere una relazione utilizzando dati statistici

- Che cos'è una relazione?
 - testo di carattere prevalentemente tecnico
 - strutturato (formato comune, spesso codificato)
 - per facilitare lettura e comprensione
 - per trasmettere contenuto informativo
- Perché corredare le argomentazioni di dati:
 - avanzare un'ipotesi
 - sostenere una tesi
- Progettare la relazione:
 - Presentare materiale e argomentazioni in modo ordinato
 - Utilizzare un linguaggio semplice e chiaro

If you fail to plan, you plan to fail!

Fasi della preparazione di una relazione

- Definire il contenuto
- Pianificare il lavoro
- Raccogliere le informazioni rilevanti
- Organizzare e strutturare le informazioni
- Redigere una prima stesura
- Rileggere, verificare la correttezza delle informazioni, procedere alla stesura finale

Definire il contenuto

- Quale tesi volete presentare nella vostra relazione?
- Con quali argomentazioni la volete sostenere?

Definire gli obiettivi aiuta a comunicare con maggiore chiarezza, ma anche a essere più selettivi nel raccogliere e presentare le informazioni

Pianificare il lavoro

- Perché è essenziale:
 - dedicare tempo sufficiente a ciascuna fase
 - produrre una relazione chiara, concisa ed efficace
- Stendete una "scaletta", articolata per punti
- Traducete i punti in fasi
- Assegnatevi delle scadenze!

La raccolta e la selezione delle informazioni statistiche, e la revisione finale assorbono più tempo

Raccogliere le informazioni

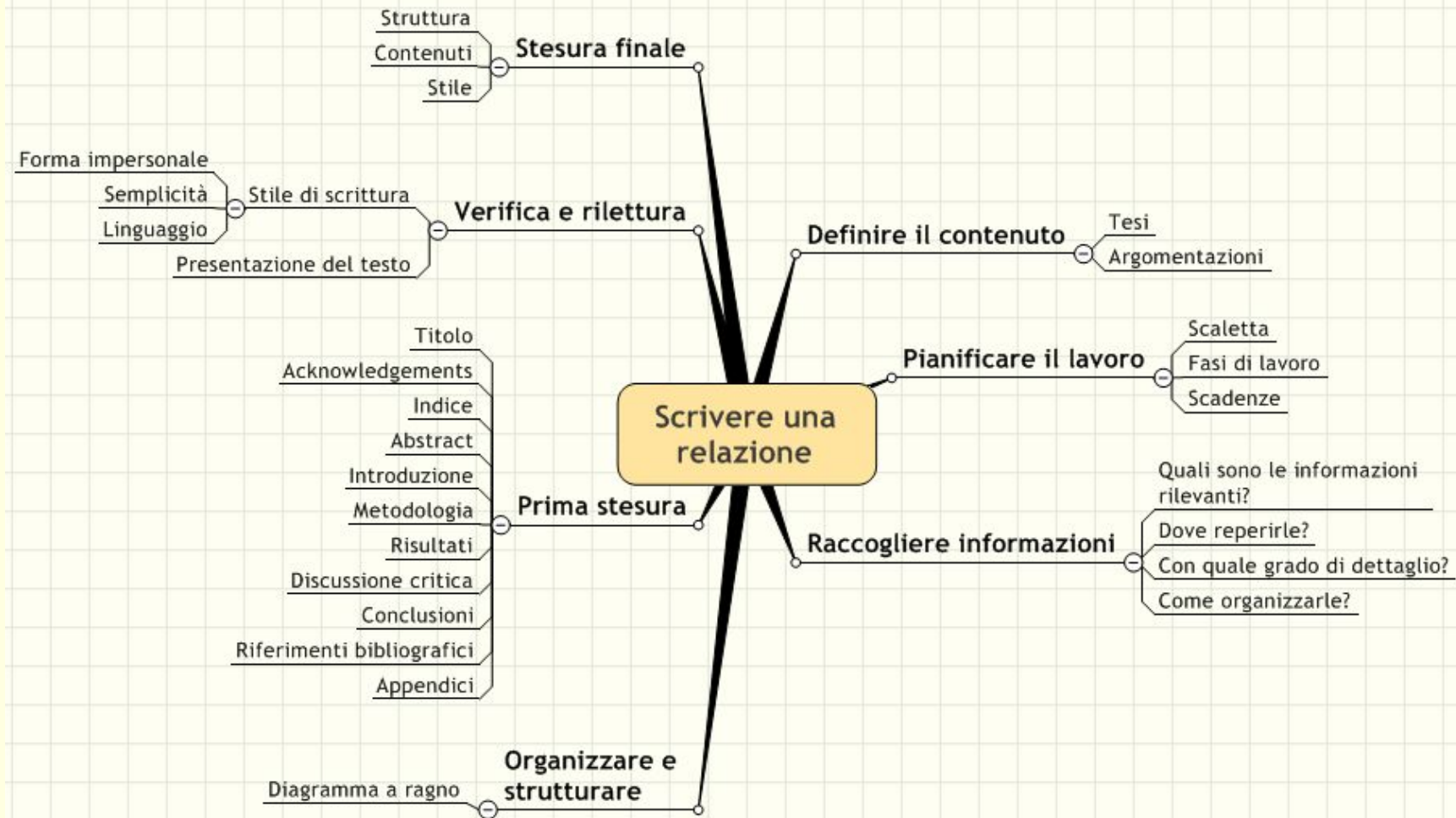
- Iniziare ponendosi una serie di domande:
 - quali sono le informazioni rilevanti?
 - dove reperirle?
 - con quale grado di dettaglio?
 - come organizzarle?

Preparare una lista delle informazioni da raccogliere e un calendario delle attività

Organizzare e strutturare le informazioni

- Diagrammi a ragno
 - scrivete il tema principale della relazione al centro della pagina
 - annotate i temi e le parole chiave attinenti al vostro argomento a partire dal centro, via via collegando e dettagliando le idee come i rami di un albero
 - ogni idea può essere sottolineata o posta al centro di un nuovo cerchio
 - quando avete finito, sottolineate i punti fondamentali e ordinate gli argomenti
 - in questo modo, le idee più importanti costituiscono gli snodi principali dell'argomentazione, e le altre le articolazioni subordinate
 - seguendo questo procedimento, dovrete essere in grado di veder emergere la struttura della relazione

Diagramma a ragno



Strutturare la relazione (1)

- Titolo
- Ringraziamenti e riconoscimenti (*acknowledgements*)
- Indice
- Sommario o *abstract*
- Introduzione
- Metodologia
- Risultati
- Discussione critica
- Conclusioni
- Riferimenti bibliografici
- Appendici

Strutturare la relazione (2)

- Introduzione
 - Obiettivi, limiti dell'esposizione, metodi d'indagine e di elaborazione, stato dell'arte delle ricerche pertinenti la materia trattata, sviluppi futuri
- Metodologia
 - In che modo sono stati raccolti i dati? Sono stati utilizzati dei questionari o delle interviste dirette? Quali fonti sono state consultate e utilizzate? Quali analisi sono state condotte? Come sono stati sintetizzati i dati? Quali indicatori sono stati costruiti?
- Risultati
 - La presentazione dei risultati è la parte della relazione in cui testo, tabelle e figure dovranno essere meglio integrate
- Discussione critica
 - I punti essenziali dell'argomentazione in sequenza logica
 - Interpretazione dei risultati ottenuti.
 - Difficoltà incontrate e limiti di validità dei risultati e delle interpretazioni
 - Confutazione delle possibili obiezioni
 - Altre teorie o interpretazioni
- Conclusioni
 - Non deve mai presentare informazioni nuove rispetto a quelle esposte nelle parti precedenti

Stile di scrittura

- Forma impersonale
 - “si raccomanda la lettura...”, piuttosto che “raccomando la lettura...”
 - Consentito l’uso del plurale (“noi” invece che “io”)
 - Evitare le forme passive
- Semplicità
 - Evitare un linguaggio complesso o involuto
 - Al centro l’illustrazione quantitativa dei fatti rilevanti
 - Ridurre al minimo necessario l’uso di “tecnicismi”
 - Poche abbreviazioni (glossario delle abbreviazioni utilizzate in appendice)
- Uso del linguaggio
 - Niente i giudizi di valore
 - Tenere separati la presentazione dei risultati dell’analisi e la loro discussione e interpretazione
 - No all’uso fantasioso o enfatico del linguaggio, soprattutto nell’aggettivazione
 - No alle locuzioni introduttive inutili
 - Limitare il ricorso al condizionale o a espressioni dubitative
 - No alle virgolette per prendere le distanze da un’affermazione che fate voi stessi

Presentazione del testo

- Utilizzare un carattere leggibile e usato correntemente
- Usate un "corpo" sufficientemente grande per il testo e uno più piccolo per le note
- Margini sufficientemente ampi (85 battute per riga)
- Strutturare le intestazioni dei capitoli e dei paragrafi
- Segnalare l'inizio del capoverso
- I capoversi sono l'unità di base del pensiero scritto. Ogni capoverso, quindi, dovrebbe contenere integralmente un passaggio logico del testo (tre-cinque periodi, 120 parole)
- Scrivete i vostri periodi in modo che ognuno contenga di norma un solo concetto o una sola affermazione nella frase principale e una, o al massimo due, subordinate.
- Kiss: *keep it short and simple*.
- Date un titolo e numerate tutte le tabelle e le figure
- Utilizzate unità di misura standard

Revisione e stesura finale

- La revisione deve prendere in considerazione gli aspetti relativi a:
 - struttura
 - contenuti
 - stile
- Fate una lista di controllo
- In conclusione
 - Quali erano i punti principali della relazione? Li avete presentati con chiarezza?
 - Quali sono le parti che vi sembrano riuscite meglio? Che cosa potete fare per portare anche le altre allo stesso livello?
 - Che cosa ha ancora bisogno di essere migliorato? Che cosa si gioverebbe di un'ulteriore messa a punto?
- Siate perfezionisti!

Rappresentare i dati: un breve corso di bricolage

- Scrivere una relazione
- Scrivere di numeri
- Tabelle
- Grafici

Perché scrivere di numeri? Perché introdurre riferimenti quantitativi in un testo?

- Per il lettore, i dati che presentate devono essere la risposta a una domanda che avete posto
 - Mettete i dati nel loro contesto (*statistical story-telling*)
 - Iniziate con un'affermazione, un'esposizione o l'illustrazione di una tesi
 - Poi esponete dati a sostegno o confutazione
- Per raccontare la storia, dovete essere i primi a sapere perché includete uno o più elementi quantitativi
 - I dati illustrano quanto un fenomeno è grande/frequente o piccolo/raro?
 - Illustrano una tendenza alla stabilità o al cambiamento?
 - Rappresentano valori tipici o inusuali? (importanza dei valori di riferimento)

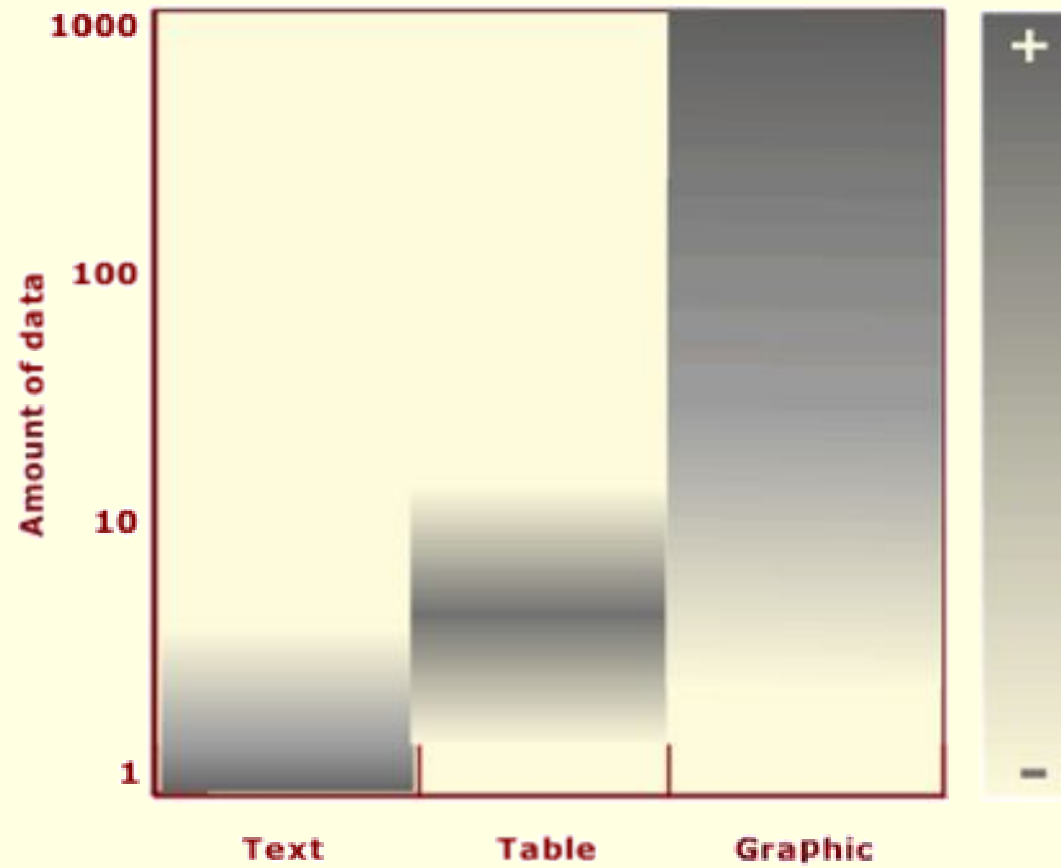
Scrivere di numeri: gli attrezzi del mestiere

- Testo
 - In un testo, i dati sono presentati all'interno di una struttura narrativa. Questo limita il numero di informazioni quantitative che possono essere inserite in un periodo o in un capoverso
- Tabelle
 - Scopo essenziale di una tabella è di semplificare la presentazione di dati statistici, che altrimenti dovrebbero essere illustrati (soltanto) in forma narrativa
- Grafici
 - Idealmente, una rappresentazione grafica deve trasmettere al lettore informazioni che non sarebbero altrettanto efficacemente presentate con una tabella o con una descrizione testuale
- Secondo Edward Tufte e Juan C. Dürsteler:
 - Una frase di testo non consente informazione che su 2-3 dati
 - Una tabella per 3-20 dati
 - Un grafico per più di 20 dati

Lo stesso concetto in una tabella...

	OK
Sentence	Relations between 2 or 3 data at most.
Table	Between 3 and 20 (approx)
Graphics	From 20 on. Specially useful for large amounts of data

... e in un grafico



Sette regole d'oro

- Stabilire il contesto
- Scegliere esempi semplici e plausibili
- Scegliere gli strumenti adeguati
- Definire la terminologia
- Analisi e interpretazione
- Specificare direzione e grandezza di un'associazione
- Riassumere e generalizzare

Stabilire il contesto

- Le W (chi, che cosa, quando, dove):
 - Male: "ci furono 25 milioni di morti"
 - Meglio: "nel XIV secolo, 25 milioni di persone morirono in Europa"
 - Ottimo: "quando la peste nera colpì l'Europa nella seconda metà del XIV secolo, uccise 25 milioni di persone, giovani e vecchi, nelle città e nelle campagne: circa un quarto della popolazione europea dell'epoca"
- Unità e scale di misura
 - Un aspetto del "che cosa"
 - Metadati!

Esercizio 1

- Descrivere la dimensione relativa delle tre città in una frase

Table 2A. Population of three largest cities worldwide, 1995

City	Population (millions)
Sao Paulo	16.5
Mexico City	16.6
Tokyo	27.0

Source: Population Reference Bureau, "World Population: More than Just Numbers" (Washington, DC: Population Reference Bureau, 1999).

Risposte 1

- “Nel 1995, la città più popolosa del mondo, contava 27 milioni di abitanti. Con una popolazione di circa 16,5 milioni ciascuna, la seconda e la terza per dimensione, Città del Messico e San Paolo, si attestavano al 60 per cento circa di Tokyo”

Esercizio 2

Table 2B. Final medal standings of the top four countries, 2002 Winter Olympic Games

Country	Gold	Silver	Bronze	Total
Germany	12	16	7	35
United States	10	13	11	34
Norway	11	7	6	24
Canada	6	3	8	17

Scegliere esempi semplici e plausibili

- Obiettivo degli esempi (stabilire l'importanza dell'argomento, comparare risultati, illustrare implicazioni)
 - Male: "nel 2001, la temperatura media di New York è stata di 13 gradi"
 - Meglio: "nel 2001, la temperatura media di New York è stata di 13 gradi, 1,5 al di sopra della media"
 - Ottimo: "nel 2001, la temperatura media di New York è stata di 13 gradi, 1,5 al di sopra della media: il settimo anno più caldo degli ultimi cento anni"
- Criteri di scelta:
 - Semplicità (KISS): si misura sulla necessità di definire i termini usati
 - Plausibilità: le differenze tra gruppi o le variazioni nel tempo devono essere sensate

Scegliere gli strumenti adeguati

- Quanti numeri?
- Quanto tempo?
- Quanto è importante la precisione?
- Combinare gli strumenti

Quanti numeri

- La regola
 - Una frase di testo non consente informazione che su 2-3 dati
 - Una tabella per 3-20 dati
 - Un grafico per più di 20 dati
- Un esempio: l'andamento della disoccupazione per regione e classe d'età
 - Dare un quadro generale: una frase
 - A dicembre 2005, il tasso di disoccupazione in Italia è salito al 5,8%, dal 3,9% di un anno prima. Anche nelle cinque ripartizioni, il tasso di disoccupazione mostra una crescita analoga
 - Dati per dieci anni, quattro classi d'età e 20 regioni: una tabella o un grafico

Quanto tempo

- Se il tempo è limitato (una presentazione, un memo per il capo), meglio una lista (“elenchi puntati e numerati”)
- Nelle presentazioni, oltre alle liste sono efficaci i grafici (se illustrati)
 - Evitate invece le tabelle (difficili da leggere e da seguire)
- Questi principi non si applicano alle relazioni e agli articoli scientifici

Quanto è importante la precisione?

- Se lo scopo è quello di mostrare i risultati di una elaborazione a un pubblico tecnico (ad esempio, i coefficienti e i test di un modello), è appropriata la presentazione di una tabella dettagliata e un livello di precisione di 3-4 cifre significative
- Un grafico è più efficace ed economico per mostrare la dimensione di una differenza o la direzione di una tendenza, ma al prezzo di una perdita di precisione (è difficile accertare i valori puntuali)
- Per le tabelle (e per i commenti testuali) è opportuno applicare la regola di Howard Wainer (un'autorità nella presentazione dei dati)
 - Nella maggior parte delle tabelle non c'è ragione per presentare più di due cifre decimali
 - La stessa tesi è sostenuta da altri: "The major problem with numerical data is that we cannot manipulate long numbers in our heads. The way to overcome this problem is to round each number to two effective digits" (Ehrenberg, *A Primer in Data Reduction*)

Combinare gli strumenti: integrare testo, tabelle e grafici

Mai una tabella o un grafico fini a sé stessi

- Come regola generale, ogni tabella o grafico dovrebbe essere accompagnato da 10-15 righe di testo direttamente riferito al contenuto della tabella o del grafico
- Richiami nel testo
- Di regola, non duplicate le informazioni (oltre a essere inutile, è fuorviante: induce il lettore a chiedersi se non gli è sfuggito qualcosa)
 - Eccezione: una frase e/o un grafico nel corpo del testo, una tabella di dettaglio in appendice
- Diversi strumenti per diversi formati

Esercizio 3

- Per ciascuna delle situazioni, spiegate se usereste una frase, una tabella o un grafico:
 - Statistiche su cinque tipi di inquinanti nelle 10 città più importanti per una relazione al parlamento
 - Tendenza delle quotazioni di tre azioni nell'arco di un anno per una pagina Web
 - Informazione agli altri impiegati della vostra società sul cambiamento delle tariffe di spedizione
 - Statistiche nazionali sul numero dei non-assicurati tra i lavoratori a tempo pieno e parziale per la parte introduttiva di un articolo sugli effetti delle tipologie di occupazione sulla copertura assicurativa nella città di New York

Risposte 3

- Una tabella, per organizzare e illustrare 50 valori
- Un grafico (linee multiple) per illustrare tendenze approssimative
- Un testo (un breve memo)
- Un testo (alcune frasi)

Definire la terminologia

- Definire termini, acronimi e simboli: perché?
 - Non definire termini tecnici (costo opportunità, modello lineare gerarchico ...) e gli acronimi (OLS, EUSILC ...) comporta il rischio di perdere il proprio pubblico
 - Termini che hanno più di un significato
 - Differenza significativa
 - Analisi della regressione
 - α
 - Differenti termini per lo stesso concetto (ad esempio, in diverse discipline)

Sono veramente necessari i termini tecnici?

Quando evitare il gergo

- Quando il testo è destinato a un pubblico di non addetti ai lavori
- Quando un termine viene usato una volta sola
- Nell'introduzione o nelle conclusioni di un paper scientifico (distrarre dai punti principali)
- Un esempio (un ponte tra due piccoli centri):
 - Male: "il rischio relativo di cedimento del materiale C è pari a 0,78" (che materiale? Quale dura di più?)
 - Meglio: "in condizioni di simulazione controllata, il materiale migliore (C) dura 1,28 volte più del secondo (B)"
 - Ottimo: "in situazioni che simulano le condizioni meteorologiche, il volume e il tonnellaggio del traffico previsto tra Cesole e Marcaria, il materiale migliore (C) ha una vita media attesa di 64 anni, contro i 50 della seconda scelta tecnologica (materiale B)"

Quando usare il gergo (e parafrasarlo)

- Quanto un termine tecnico è usato ripetutamente
- Quando il termine o il simbolo ha più significati diversi
- Quando si fa riferimento a un concetto per cui si usano termini diversi in diverse discipline
- Quando i vantaggi della sintesi eccedono i suoi costi (è spesso il caso delle formule)
- Un esempio:
 - Male: "nel 1999, il tasso lordo di mortalità (TLM) per la Svezia era di 11 morti per 1000 abitanti e per il Ghana di 10 morti per 1000 abitanti. Ci si aspetterebbe che la Svezia (uno dei paesi più industrializzati) abbia una mortalità più bassa del Ghana (paese meno sviluppato). Il motivo risiede nella struttura per età della popolazione. Pertanto è opportuno calcolare la speranza di vita per i due paesi, applicando tassi di mortalità specifici per classe d'età [...] La speranza di vita è di 78 anni per la Svezia e di 58 per il Ghana"
 - Meglio (per un pubblico non tecnico): "nel 1999, gli abitanti del Ghana potevano sperare di vivere in media fino all'età di 58 anni, contro i 78 degli svedesi. La differenza nella speranza di vita riflette tassi di mortalità molto più bassi in Svezia"
 - Ottimo (per un paper scientifico): "nel 1999, il tasso lordo di mortalità (TLM) per la Svezia era di 11 morti per 1000 abitanti e per il Ghana di 10 morti per 1000 abitanti, dando l'illusione di probabilità di sopravvivenza leggermente migliori per il Ghana. Tuttavia, la Svezia ha una quota di popolazione molto più elevata nelle classi d'età più anziane (il 17% ha 65 anni o più, contro il 3% in Ghana) e gli anziani hanno tassi di mortalità più elevati. Questa differenza eleva il tasso di mortalità medio per la Svezia. La speranza di vita – una misura della mortalità che tiene conto delle differenze nella distribuzione per età – mostra che le probabilità di sopravvivenza sono molto migliori in Svezia, dove l'abitante medio può sperare di vivere 78 anni, che in Ghana (58)"

Esercizio 4

- “Una popolazione che cresce del 2% l’anno ha un tempo di raddoppio di 35 anni”

Analisi e interpretazione

- Riportare i valori assoluti è soltanto il primo passo
 - Non mettere i dati nel loro contesto “scarica” questo compito sul lettore, che potrebbe non essere in grado di effettuare i confronti rilevanti
- È comunque necessario, anche se non sufficiente
 - Una variazione del 30% della densità della popolazione di cervi (da 5,0 a 6,5 per km² o da 1.000 a 1.300?)
- Un esempio:
 - Male: “nel 1998, la spesa sanitaria totale negli Stati Uniti ammonta a oltre 1.100 miliardi di dollari” (in confronto a che?!)
 - Meglio: “nel 1998, la spesa sanitaria totale negli Stati Uniti ammonta a oltre 1.100 miliardi di dollari, pari a 4.198 \$ per abitante (uomo, donna o bambino)”
 - Ottimo (per richiamare l’attenzione sulla tendenza): “tra il 1990 e il 1998, il costo totale della sanità negli Stati Uniti è passato da 699 a 1.100 miliardi di dollari, con un aumento del 65%. Nello stesso periodo, la quota della spesa sanitaria sul prodotto interno lordo (PIL) è cresciuta dal 12,0 al 13,1%”
 - Ottimo (per collocare gli Stati Uniti nel contesto internazionale): “durante gli anni Novanta, negli Stati Uniti la spesa sanitaria media è stata di 4.108 \$ per abitante, con una quota sul prodotto interno lordo (PIL) del 13,0% - la più alta al mondo nel medesimo periodo. La Svizzera – al secondo posto nella graduatoria della spesa sanitaria pro capite – ha speso in media 3.835 \$ per abitante, il 10,4% del PIL. Nessun altro paese ha superato i 3.000 \$”

Specificare direzione e grandezza di un'associazione

- Scrivere di numeri comporta in genere la descrizione delle relazioni tra due o più variabili
- Per interpretare un'associazione, è necessario conoscerne la forma e la dimensione
- Direzione dell'associazione:
 - Positiva/diretta vs. negativa/inversa
 - Per le variabili nominali (le cui categorie non hanno un ordine inerente) la direzione dell'associazione comporta il riferimento alle categorie con il valore più alto e più basso
 - Un esempio:
 - Male: "l'appartenenza religiosa è negativamente associata all'abitudine al fumo"
 - Meglio: "tra le confessioni religiose, i mormoni hanno minore probabilità di essere fumatori, mentre i battisti hanno la più elevata"

Dimensione dell'associazione

- Esempio:
 - Male: " mortalità ed età sono correlate" (in che direzione? Quanto?)
 - Meglio: "la mortalità cresce al crescere dell'età"
 - Ottimo: "tra gli anziani, la mortalità grosso modo raddoppia per ogni classe quinquennale d'età successiva"
- Specificare direzione e grandezza dell'associazione aiuta a chiarire la discussione delle ipotesi:
 - Male: "ci si attende che i sintomi differiscano tra il gruppo trattato e quello di controllo"
 - Meglio: "ci si attende che i pazienti trattati con il farmaco A presentino sintomi meno forti di quelli che ricevono il placebo"

Riassumere e generalizzare

- I dati presentati (nel testo o in tabelle e grafici) servono a illustrare un argomento o a presentare elementi a sostegno di una tesi
 - Non riassumere e generalizzare – individuando pattern – “scarica” questo compito sul lettore, che potrebbe non essere in grado di farlo
- Due errori comuni:
 - Commentare nel testo tutti i dati di una tabella (ridondanza)
 - Scelta arbitraria di alcuni valori, senza porsi il problema se individuano un pattern rilevante
- Riassumere serve a riportare l’attenzione all’argomento della relazione
 - I prezzi delle abitazioni crescono, diminuiscono o restano stabili?
 - Cambiano nel tempo come ci si aspetterebbe dal mutare delle condizioni economiche?
 - Emergono disparità regionali significative?
- Il lettore può trovare i dettagli nelle tabelle!

GEE: generalizzare, esemplificare, eccezioni

- Jane E. Miller (*The Chicago Guide to Writing about Numbers*) propone l'approccio GEE: *Generalization, Example, Exceptions*
- Generalizzare: individuare una regolarità e descriverla in termini generali
- Esemplificare: scegliere un esempio per illustrare la regolarità
- Eccezioni: spiegare e mostrare le principali eccezioni alla regola

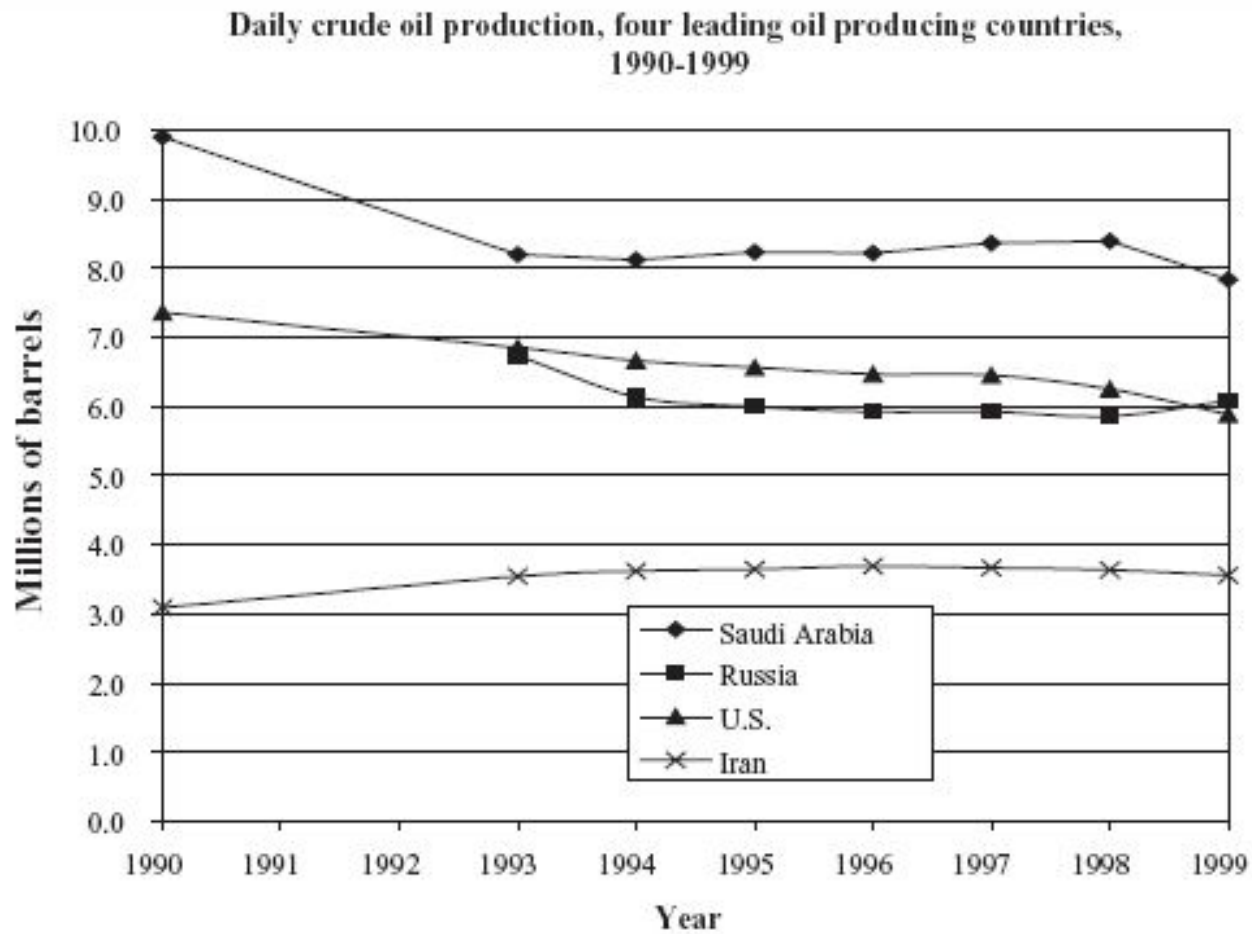
GEE in sei mosse (1-3)

1. Rappresentate i dati in una tabella e/o un grafico
 - A fini d'analisi, anche se non intendete metterlo nella versione definitiva del lavoro
2. Individuate le dimensioni del confronto
 - In una tabella, leggendo i dati per riga lungo ogni colonna, e per colonna lungo ogni riga
 - In un grafico a linee multiple, seguendo la tendenza (lungo le linee) e le differenze tra oggetti (tra le linee)
3. Scegliete un esempio rappresentativo
 - Se disponete di totali o misure riassuntive, è una buona idea partire da lì
 - Altrimenti scegliete il valore più rappresentativo (per esempio, la classe modale) o un gruppo di particolare interesse per la vostra ricerca
 - Verificate (fasi 4. e 5.) che l'esempio sia veramente rappresentativo!

GEE in sei mosse (4-6)

4. Caratterizzate il pattern
 - Prendete appunti
 - Direzione
 - Tendenze
 - Costanza o cambiamento?
 - Lineare, in accelerazione o diminuzione?
 - Monotonica o cambia direzione?
 - Differenziali
 - Valori massimo e minimo
 - Posizioni relative rispetto a massimo/minimo
 - Dimensione
 - Può essere utile fare i calcoli su un foglio elettronico
 - Significatività statistica
 - Può emergere un pattern in sé
5. Individuate le eccezioni
 - Nella direzione
 - Nella dimensione
 - Nella significatività
6. Redigete il testo

GEE: un esempio (1)

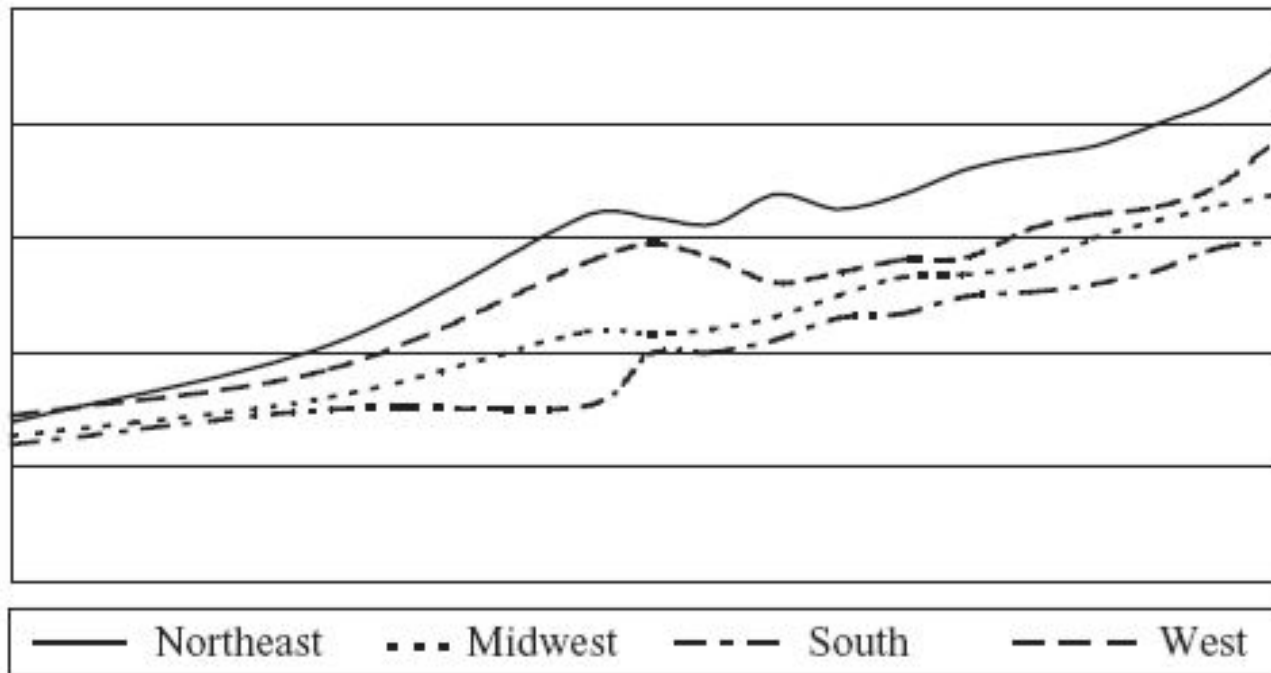


GEE: un esempio (2)

- Generalizzare
 - Il grafico rappresenta l'evoluzione della produzione giornaliera di greggio nei quattro principali paesi produttori nel corso degli anni Novanta
 - Durante tutto il decennio, l'Arabia Saudita ha sempre detenuto la posizione del maggior produttore, seguita da Russia, Stati Uniti e Iran
 - Tuttavia, la tendenza alla diminuzione della produzione giornaliera nei tre principali produttori, insieme alla produzione stabile fatta registrare dall'Iran, ha fatto sì che tra il 1990 e il 1999 le distanze tra paesi si siano ridotte
- Esempificare
 - Nel 1990, la produzione di greggio dell'Arabia Saudita superava del 30% quella degli Stati Uniti e sopravanzava di tre volte quella dell'Iran (10, 7 e 3 milioni di barili al giorno, rispettivamente)
 - Nel 1999, il distacco si era ridotto al 25% rispetto a Stati Uniti e Russia, e a circa due volte rispetto all'Iran
- Eccezioni
 - La tendenza alla riduzione della produzione giornaliera è costante in tutto il periodo per gli Stati Uniti, mentre in Arabia Saudita a una prima diminuzione all'inizio del decennio (da 10 a 8 milioni di barili al giorno), segue una fase di lieve ripresa della produzione giornaliera fino al 1998, seguita da una nuova contrazione che, nel 1999, porta la produzione media al di sotto degli 8 milioni di barili
 - Al contrario, tra il 1998 e il 1999 la Russia fa registrare un aumento della produzione media giornaliera, che la porta a sopravanzare gli Stati Uniti nella graduatoria dei maggiori produttori

GEE: un altro esempio (1)

Median sales price of new one-family homes,
by region, United States, 1980–2000



GEE: un altro esempio (2)

- Generalizzare
 - Tra il 1980 e il 2000, il prezzo mediano di un'abitazione unifamiliare nuova ha seguito un andamento sostanzialmente crescente in ciascuna delle quattro grandi ripartizioni censuarie
 - Questa tendenza si è interrotta, e in alcuni casi invertita, con prezzi stabili o calanti, intorno al 1990, per poi riprendere la traiettoria ascendente
 - Per la maggior parte del periodo analizzato, i prezzi più elevati si sono registrati nel Northeast, seguito da West, Midwest e South
- Esempificare
 - Per seguire la tendenza
 - Ad esempio, nel Northeast, il prezzo mediano di un'abitazione unifamiliare nuova è più che triplicato, passando da 69.500 \$ nel 1980 a 227.400 \$ nel 2000
 - Per seguire le disparità regionali
 - Nel 2000, i prezzi medi di un'abitazione unifamiliare nuova erano di 227.400, 196.400, 169.700 e 148.000 \$ nel Northeast, West, Midwest e South, rispettivamente
- Eccezioni
 - In tre delle quattro ripartizioni, i prezzi sono cresciuti per tutti gli anni Ottanta
 - Nel South, tuttavia, i prezzi delle abitazioni non hanno cominciato a salire che nel 1990, per poi crescere a un tasso analogo a quello delle altre ripartizioni

Le sette regole d'oro: una lista di controllo

- Stabilite il contesto: le W
- Scegliete esempi e analogie efficaci
 - Esempi semplici e familiari, che il vostro lettore comprenda e possa applicare a un contesto abituale
 - Scegliete confronti realistici in condizioni ordinarie
- Scegliete un lessico adeguato al vostro pubblico
 - Definite i termini e citate i sinonimi utilizzati in campi affini
 - Sostituite i termini tecnici e i simboli matematici ovunque appropriato
- Decidete la combinazione di testo, tabelle e grafici
 - Quanti dati dovete presentare?
 - Quanto tempo ha il vostro pubblico per comprenderli
 - Sono necessari valori precisi?
- Presentate, analizzate e interpretate i dati
 - Presentate i dati ed esplicitate lo scopo per cui lo fate
 - Analizzateli, interpretateli e riconduceteli al tema della relazione
- Specificate direzione e dimensione dell'associazione tra variabili
 - Se è un andamento, è in crescita o diminuzione?
 - Se differenze tra gruppi o regioni, individuate il valore massimo e collocate gli altri rispetto a quello
- Per descrivere una regolarità che emerge da molti dati, riassumete la regolarità senza ripetere tutti i valori
 - Individuate una generalizzazione che si attaglia alla maggior parte dei dati
 - Riportate alcuni valori esemplificativi dalla tabella o dal grafico
 - Descrivete le eccezioni

Rappresentare i dati: un breve corso di bricolage

- Scrivere una relazione
- Scrivere di numeri
- **Tabelle**
- Grafici

Perché e quando le tabelle?

- Scopo essenziale di una tabella è di semplificare la presentazione di dati statistici, che altrimenti dovrebbero essere illustrati (soltanto) in forma narrativa
- Quando una tabella è ben costruita?
 - Una tabella ben costruita deve presentare i dati in modo più efficiente che in forma testuale
 - Requisiti:
 - deve presentare dati **dotati di senso**
 - i dati **non** devono essere **ambigui**
 - deve trasmettere informazioni in modo **efficiente**

Significato

- Il **significato** dei dati ha a che fare con il rapporto tra i dati stessi e gli argomenti principali del testo o dell'analisi di riferimento

Ambiguità

- L'**ambiguità** dei dati dipende in larga misura dal testo contenuto nella tabella
- L'informazione riportata nella tabella – il titolo, le intestazioni dei dati (testate e fiancate), le note e la citazione delle fonti – deve essere integralmente comprensibile in sé, senza l'ausilio della testo di riferimento
- Testate e fiancate devono definire esattamente che cosa ogni numero della tabella rappresenta
- Il riferimento alla fonte dei dati deve dare al lettore un'indicazione (il più possibile precisa) sull'affidabilità e l'ambito di validità dei dati

Efficienza

- L'**efficienza** della presentazione è ciò che consente al lettore di trarre una molteplicità di conclusioni interessanti da un insieme vasto di informazioni
- La rapidità con cui il lettore è in grado di "metabolizzare" l'informazione presentata, individuare le relazioni critiche tra i dati e trarre conclusioni dotate di senso dipendono dalla struttura della tabella

Creare una tabella auto-contenuta

- I lettori di un testo di analisi di dati statistici tendono a essere di due tipi:
 - quelli che scorrono il testo saltando la maggior parte dei grafici e delle tabelle
 - quelli che si soffermano su tabelle e grafici ignorando il testo
- Un'analisi ben fatta deve essere in grado di soddisfare entrambi i tipi di lettori, che devono essere messi in condizione di cogliere i punti essenziali dell'analisi quale che sia il loro "stile" di lettura
- Per creare una tabella auto-contenuta è necessario che essa presenti queste informazioni:
 - Scopo della tabella
 - Contesto dei dati (le W)
 - Collocazione delle variabili dentro la tabella
 - Codifica e unità di misura dei singoli dati
 - Tipo di statistiche o modelli statistici
 - Fonti dei dati
 - Definizioni dei termini rilevanti e delle abbreviazioni

Ambiguità

- Titolo, intestazioni di riga e colonna (fiancate e testate) e le note devono definire precisamente il significato di ogni dato puntuale
- Se si utilizzano tassi o rapporti, definire chiaramente denominatore e numeratore (se possibile ambiguità, anche la formula adottata)
- Particolare attenzione alle unità di misura adottate
- Statistiche espresse in migliaia / rapporti per mille unità di popolazione:
 - Vittime di incidenti stradali (migliaia).
 - Vittime di incidenti stradali (per 100.000 abitanti).
 - Vittime di incidenti stradali (per 100 milioni di veicoli-chilometro)

Variabili, casi, dati

- Una **variabile** fa riferimento al concetto che si vuole misurare; per esempio: il tasso di criminalità, il reddito pro capite, il PIL. Nella maggior parte delle tabelle, le variabili sono riportate nelle colonne.
- I **casi** fanno riferimento alle unità geografiche, ai gruppi, al tempo, alle unità di analisi. Di solito, i casi sono riportati nelle righe (con molte eccezioni). In realtà, anche i casi sono variabili.
- Un **dato** – o dato puntuale – è il **valore** assunto dalla **variabile** in un **caso** particolare

Anatomia d'una tabella

Numerazione progressiva

Tab. 2

Attuazione dell'Obiettivo 1 a fine 2003, 2000-2006

Titolo

Testata

Stato membro	Valori assoluti (M€)			Rapporti percentuali		
	Contributo	Impegni	Pagamenti	Imp./Con.	Pag./Con.	Pag./Imp.
BE BELGIO	644.91	414.91	154.74	64.3	24.0	37.3
DK DANIMARCA	-	-	-	-	-	-
DE GERMANIA	20.707.00	12.120.00	7.854.82	58.5	37.9	64.8
GR GRECIA	21.415.00	10.420.66	4.796.66	48.7	22.4	46.0
ES SPAGNA	39.548.00	22.750.75	16.135.46	57.5	40.8	70.9
FR FRANCIA	3.948.00	2.321.68	972.10	58.8	24.6	41.9
IE IRLANDA	3.060.94	2.178.00	1.340.87	71.2	43.8	61.6
IT ITALIA	21.638.18	13.114.03	5.426.65	60.6	25.1	41.4
LU LUSSEMBURGO	-	-	-	-	-	-
NL PAESI BASSI	125.93	86.93	24.97	69.0	19.8	28.7
AT AUSTRIA	271.00	156.88	112.98	57.9	41.7	72.0
PT PORTOGALLO	19.177.47	12.213.00	8.109.34	63.7	42.3	66.4
FI FINLANDIA	948.00	548.00	305.27	57.8	32.2	55.7
SE SVEZIA	748.00	432.93	301.89	57.9	40.4	69.7
GB REGNO UNITO	6.053.97	3.729.97	1.424.21	61.6	38.5	38.2
ALTRO (1)	531.00	421.00	83.56	79.3	19.7	19.8
EUR 15	138.817.40	80.968.72	47.043.53	58.3	31.7	58.1

Fiancata

Fonte

Fonte: Commissione europea

Richiamo nota

Dati

Ai prezzi dell'anno di adozione dei provvedimenti di programmazione
(1) Programmi di cooperazione interregionale

12-13.12 2007

Note

Giovanni A. Barbieri

57

Lo stile Istat

Tavola 3.2 - Occupati e sottoccupati per sesso e settore di attività economica - Anno 2004

SETTORI	Valori in migliaia		Composizioni percentuali		Tassi di sottoccupazione
	Occupati	Sottoccupati	Occupati	Sottoccupati	
MASCHI					
Agricoltura	683	40	5,0	7,3	5,9
Industria in senso stretto	3.572	126	26,2	23,0	3,5
Costruzioni	1.726	107	12,7	19,6	6,2
Servizi	7.641	275	56,1	50,3	3,6
Totale	13.622	547	100,0	100,0	4,0
FEMMINE					
Agricoltura	307	17	3,5	3,8	5,5
Industria in senso stretto	1.464	62	16,7	13,9	4,2
Costruzioni	107	5	1,2	1,1	4,7
Servizi	6.905	361	78,6	81,1	5,2
Totale	8.783	445	100,0	100,0	5,1
TOTALE					
Agricoltura	990	57	4,4	5,7	5,8
Industria in senso stretto	5.036	188	22,5	19,0	3,7
Costruzioni	1.833	112	8,2	11,3	6,1
Servizi	14.546	635	64,9	64,0	4,4
Totale	22.404	992	100,0	100,0	4,4

Fonte: Istat, Rilevazione continua sulle forze di lavoro

Titolo

- Numerate progressivamente le tabelle
- Oggetto (<unità d'analisi> per <variabili di testata> e <variabili di fiancata>)
- Tipo di statistica (se soltanto uno nella tabella – altrimenti nelle testate o fiancate)
 - Univariate: distribuzione, composizione, medie ...
 - Bivariate
 - Risultati di un modello
- Contesto (le W)
 - Restrizioni temporale, spaziali, rispetto ai gruppi rappresentati ...
- Unità di misura (se soltanto una nella tabella – altrimenti nelle testate o fiancate)

Etichette di fiancata e di testata

- Ogni dato viene individuato dall'etichetta di fiancata e di testata (oltre che dall'oggetto nel titolo)
- Se ci sono più unità di misura, vanno specificate nelle etichette di testata e fiancata rilevanti
- Evitate le abbreviazioni (se necessario, usate le note)
- Se necessario, dividete la tabella in "pannelli", individuati da un titolo di pannello (nell'esempio: maschi, femmine, totale)
- L'uso dei pannelli (oltre che di testate e fiancate multiple) consente di aggiungere dimensioni (ad esempio, per anni diversi o per diverse caratteristiche della popolazione)
- Organizzate le tabelle per blocchi collegati concettualmente (si risparmiano tabelle e fatica al lettore)
- Per le fiancate: uso dei rientri
- Per le testate: titoli che raggruppano più colonne ("spanners")

Attenzione al significato

- I dati dotati di significato sono quelli che misurano qualcosa di importante rispetto ai casi analizzati
- In genere, le statistiche basate su tassi, rapporti e misure pro capite sono più dotate di senso dei valori assoluti
- Attenzione alle misure che riflettono soltanto la dimensione dei casi (per esempio, la popolazione)

Valori assoluti e rapporti

Murders* in Ten Largest US Cities, 1998		Murder Rates* in Ten Largest US Cities, 1998	
Chicago	703	Detroit	43,0
New York	633	Chicago	25,6
Detroit	430	Philadelphia	23,3
Los Angeles	426	Dallas	23,1
Philadelphia	338	Phoenix	15,1
Houston	254	Houston	14,1
Dallas	252	Los Angeles	11,8
Phoenix	185	New York	8,6
San Antonio	89	San Antonio	8,1
San Diego	42	San Diego	3,5
*Murder and non-negligent manslaughter		*Murder and non-negligent manslaughter per 100.000 population	
Source: Statistical Abstract 2000, CD-Rom, table 332; and Bureau of Justice Statistics: http://www.ojp.usdoj.gov/bjs/data/cities92.wk1			

La dimensione temporale

- Due punti nel tempo sono in genere meglio di uno (il confronto è informativo)
- Intervalli di tre, cinque o dieci anni consentono analisi più affidabili (fluttuazioni casuali) – a meno che si voglia misurare un effetto prima e dopo un evento
- Valore medio riferito a un triennio o un quinquennio per isolare effetti di fluttuazioni casuali (ma attenzione agli effetti di selezione)

La dimensione temporale: Murder Rates in Ten Largest US Cities, 1995-98

	1995	1998	Net Change
Detroit	47,6	43,0	- 4,6
Chicago	30,0	25,6	- 4,4
Philadelphia	28,2	23,3	- 4,9
Dallas	26,5	23,1	- 3,3
Los Angeles	24,5	11,8	-12,7
Phoenix	19,7	15,1	- 4,6
Houston	18,2	14,1	- 4,1
New York	16,1	8,6	- 7,5
San Antonio	14,2	8,1	- 6,1
San Diego	7,9	3,5	- 4,4

*Murder and non-negligent manslaughter per 100.000 population

Sources: Statistical Abstract 2000, CD-Rom, table 332; and Bureau of Justice Statistics:
<http://www.ojp.usdoj.gov/bjs/data/cities92.wk1>

Uso delle percentuali

- Le percentuali si possono calcolare in almeno due modi e rappresentano spesso una fonte di confusione
- Le percentuali di “composizione” o di “distribuzione” sono meno utili di quelle basate su rapporti
- Particolarmente vero quando le categorie sono arbitrarie

Uso delle percentuali

Distribution of Poverty Families by age of householder, 1999		Poverty Rate of Families by age of householder, 1999	
Age of householder	%	Age of householder	%
18 to 24 years old	14	18 to 24 years old	31
25 to 34 years old	29	25 to 34 years old	16
35 to 44 years old	27	35 to 44 years old	10
45 to 54 years old	12	45 to 54 years old	6
55 to 64 years old	9	55 to 64 years old	6
65 years old and over	10	65 years old and over	6
Total	100	All families	9

Source: US Statistical Abstract 2001, CD-Rom. Table 686

Percentuali e variazioni

- Il calcolo delle variazioni riferite a percentuali, tassi e rapporti è fonte di grande confusione, sia nelle tabelle, sia nei commenti testuali
- Calcolare la variazione percentuale di una percentuale è problematico perché l'uso di una determinata misura o del suo complemento è arbitraria
- Consideriamo i tassi di occupazione e disoccupazione
 - Se il tasso di disoccupazione scende dal 10 al 5%, si tratta di una variazione percentuale in diminuzione del 50%
 - Il tasso di occupazione – che ne è il complemento – è salito dal 90 al 95%: una crescita in termini percentuali pari soltanto al 5,6%
 - Nelle statistiche del lavoro, l'indicatore più comunemente usato è il tasso di disoccupazione, piuttosto che quello di occupazione
 - Nella misurazione della partecipazione al mercato del lavoro, si usa più frequentemente il rapporto tra forze di lavoro e popolazione, piuttosto che quello tra non-forze di lavoro e popolazione

Ordinamento dei dati

- Ordinate i dati sulla base della variabile più rilevante ai fini dell'analisi
- La maggior parte delle tabelle pubblicate nelle fonti statistiche presenta i dati ordinati alfabeticamente o (nel caso di unità territoriali) geograficamente
- Raramente l'ordinamento alfabetico coincide con la variabile più importante
- Coerenza!

Ordinamento dei dati

Youth Television Watching	Percent of 9-year-olds who watch more than 5 hours of television per weekday	Youth Television Watching	Percent of 9-year-olds who watch more than 5 hours of television per weekday
Canada	14,9	United States	21,5
Denmark	6,0	Spain	17,5
Finland	6,1	Canada	14,9
France	5,5	Netherlands	12,6
Germany	4,4	Ireland	11,8
Ireland	11,8	Italy	9,2
Italy	9,2	Finland	6,1
Netherlands	12,6	Denmark	6,0
Spain	17,5	France	5,5
Sweden	4,7	Sweden	4,7
United States	21,5	Germany	4,4

Source: Uri Bronfenbarger, et. al. *The State of Americans* (New York: The Free Press, 1996)
 from: William Bennett, *The Index of Leading Cultural Indicators* (New York: Broadway Books, 1999), p. 230

Decimali e arrotondamenti

- Limitare il numero dei decimali a quelli necessari a rappresentare i dati con due o tre cifre significative
- Non aggiungere il segno di percentuale (o quello della valuta, nel caso di importi monetari) accanto alle cifre
- L'indicazione va riportata nel titolo della tabella (se si applica a tutti i valori nella tabella) o nell'intestazione della variabile

Decimali e arrotondamenti: Income, Poverty and Education, by Race and Ethnic Origin: 1998

	Median Family Income	Poverty Rate (%)	High School Graduates (%)
White	49.023	10,1	84,3
Asian*	52.826	12,5	84,7
Hispanic**	29.608	25,6	56,1
Black	29.404	26,1	77,0

*Asian and Pacific Islander

**Hispanics many be of any race

Source: Statistical Abstract 2000, Tables 737, 755
and 250.

Decimali e arrotondamenti: la stessa tavola secondo Wainer e Ehrenberg

	Median Family Income	Poverty Rate (%)	High School Graduates (%)
White	49.000	10	84
Asian*	53.000	13	85
Hispanic**	30.000	26	56
Black	29.000	26	77

*Asian and Pacific Islander

**Hispanics many be of any race

Source: Statistical Abstract 2000, Tables 737, 755 and 250.

Un'altra variazione: sbagliata!

	White	Asian*	Hispanic**	Black
Median Family Income	49.023	52.826	29.608	29.404
Poverty Rate (%)	10,1	12,5	25,6	26,1
High School Graduates (%)	84,3	84,7	56,1	77

*Asian and Pacific Islander

**Hispanics many be of any race

Source: Statistical Abstract 2000, Tables 737, 755 and 250.

Non disorientare il lettore!

- Dati dello stesso genere devono essere presentati nella medesima colonna
- Proporre dati di diverso tipo nella stessa colonna disorienta

Welcome to Farmer City!

Founded	1837
Population	2114
Elevation	965 ft.
Total	4916

Riferimenti temporali

- Nelle tabelle in cui i riferimenti temporali caratterizzano le colonne, disponeteli in colonne adiacenti, ordinati da sinistra a destra in serie crescente
- Se invece sono disposti in fiancata, ordinarli in modo che il periodo più recente sia all'ultimo posto in basso

Riferimenti temporali: tabelle o grafici?

- I dati in serie storica, quando sono più di cinque, sono più comprensibili su un grafico che su una tabella
- I grafici illustrano una tendenza in modo più efficace di una tabella, anche se al costo di una perdita di precisione
- Un'alternativa è quella di rinunciare alla presentazione in tabella dei dati annuali, scegliendo una periodizzazione diversa

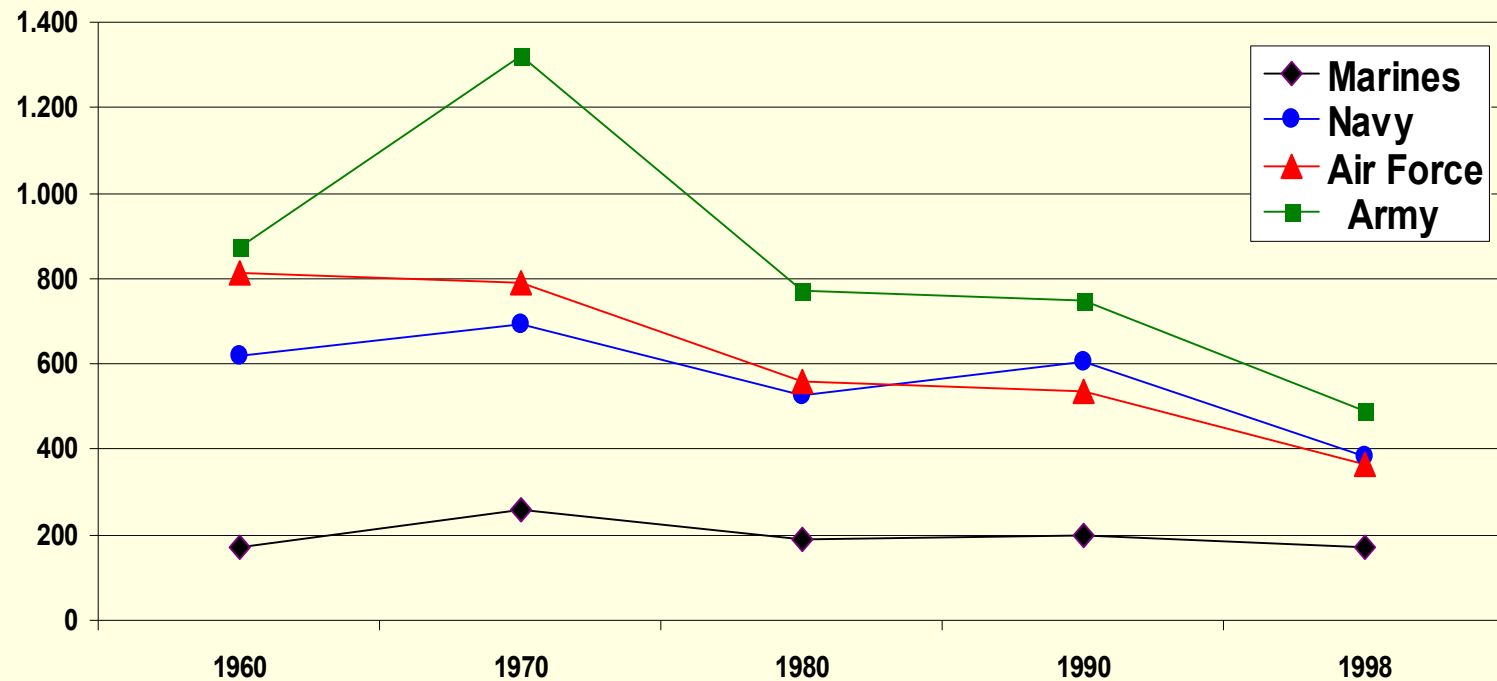
Riferimenti temporali: US Armed Forces: Active Duty Personnel (1000s)

	Army	Air Force	Navy	Marines	Total
1960	871	814	618	171	2.474
1970	1.320	791	693	260	3.063
1980	773	558	527	188	2.046
1990	746	535	605	197	2.083
1998	492	363	381	173	1.409

Source: Bennet, *Index of Leading Cultural Indicators*, pp. 202-203

O meglio un grafico?

Figure 1: US Armed Forces: Active Duty Personnel



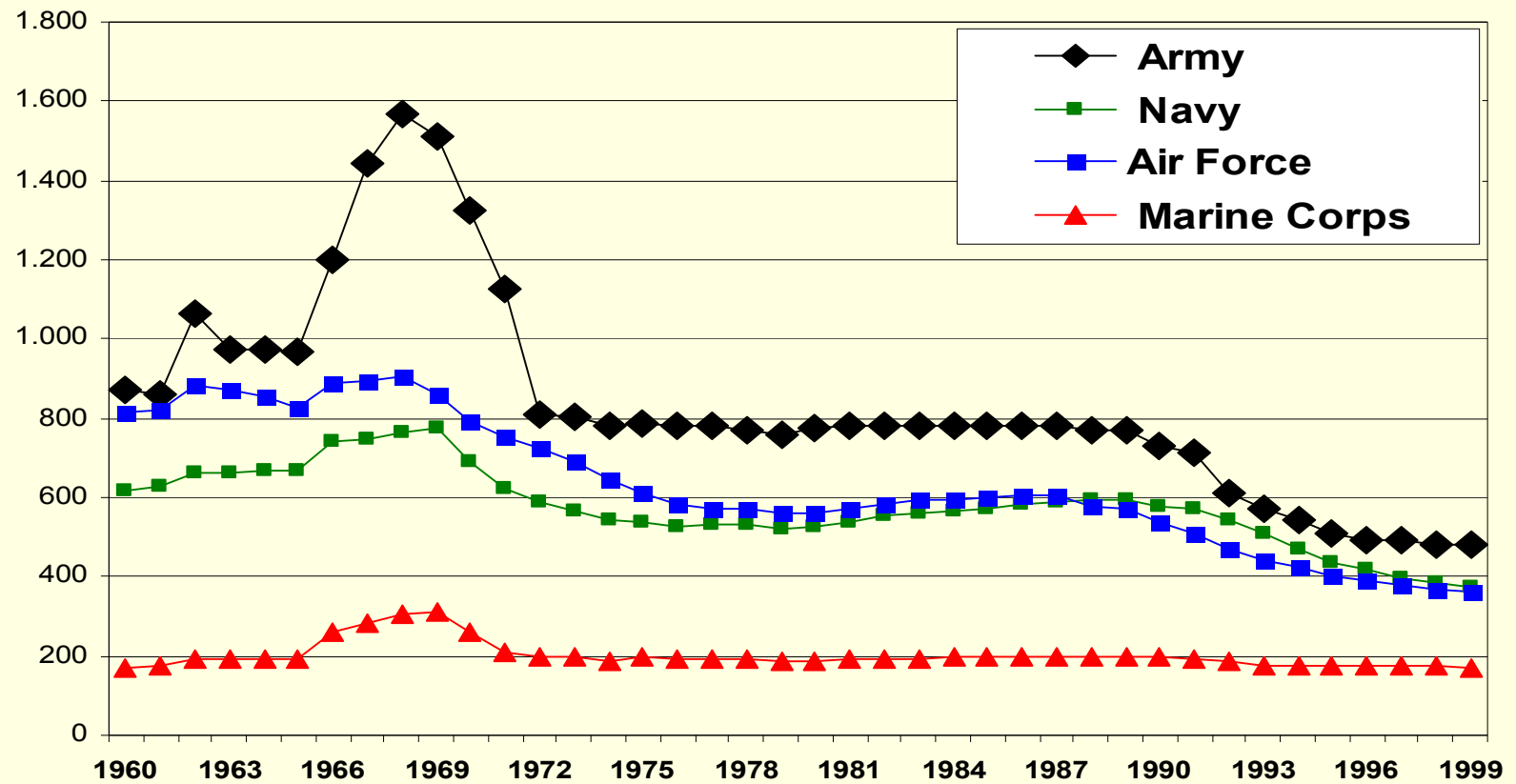
E in questo caso? Tabella...

	Army	Air Force	Navy	Marines	Total
1960	873	815	617	171	2.476
1961	859	821	626	177	2.483
1962	1.066	884	664	191	2.805
1963	976	869	664	190	2.699
1964	973	857	666	190	2.686
1965	969	825	670	190	2.654
1966	1.200	887	743	262	3.092
1967	1.442	897	750	285	3.374
1968	1.570	905	764	307	3.546
1969	1.512	862	774	310	3.458
.....
1995	509	400	435	175	1.519
1996	491	389	417	175	1.472
1997	492	377	396	174	1.439
1998	484	367	382	173	1.407
1999	479	361	373	173	1.386

Source: U.S. Dept. of Defense, Selected Manpower Statistics, annual.

0 grafico?

US Department of Defense Manpower
(active duty personnel in thousands)



Coerenza e regole di stile

- Quando un testo o una pubblicazione contiene più di una tabella, la struttura dovrebbe essere coerente nelle diverse tabelle:
 - Stesso carattere
 - Stesso stile delle fiancate e testate
 - Stesse cornici
 - Stesso ordinamento
- Per questo esistono i manuali di stile

Struttura e cornici

- Nella maggior parte delle tabelle che abbiamo visto oggi viene utilizzata una struttura semplice e comune:
 - Righe orizzontali (eventualmente di diverso spessore) sotto le intestazioni delle colonne e al termine del corpo dei dati
 - Altre linee separano la fonte e le note e possono essere utilizzate per separare settori diversi della tabella ed eventualmente i totali
- In genere non sono necessarie fincature verticali, perché è sufficiente l'allineamento dei valori numerici
- Per le tabelle è preferibile utilizzare un carattere "a bastone" (cioè privo di "grazie") come Arial o Verdana, mentre per i testi si preferiscono in genere i caratteri dotati di grazie (come Times o Garamond).
- Il titolo è di norma in grassetto, che va invece evitato nelle tabelle (a meno che i testi siano molto brevi)
- La fonte (o le fonti) dei dati e le note (se presenti) vanno elencate in calce al corpo dei dati

Combinare più tabelle

- Evitare di stipare troppi dati o troppi diversi tipi di dati in una sola tabella
- Questo principio generale, tuttavia, non deve scoraggiare la ricerca di occasioni per combinare più tabelle elementari in una più complessa, soprattutto se ciò consente di individuare relazioni interessanti tra i dati

Alcune considerazioni finali

- Quali dati includere (distinguate la vostra base di dati, l'output dei programmi di elaborazione, le vostre tabelle di lavoro dalle tabelle che volete includere nel lavoro, che devono presentare **tutti** e **solo** i dati rilevanti alla vostra argomentazione)
- Scala e numero di decimali (attenti alle convenzioni: interi e valori monetari)
- Allineamento (etichette di riga a sinistra, di colonna centrate, numeri allineati correttamente)
- Disposizione nella pagina verticale/orizzontale (dipende dal numero di righe e colonne!)
- Progettate le tabelle con carta e penna!

Tabelle: una lista di controllo

- Titolo
 - Scopo della tabella
 - Contesto dei dati (le W)
- Etichette di riga e di colonna
 - Individuate i contenuti (variabili e casi)
 - Specificate le unità di misura, se necessario
- Fonte
- Note
- Struttura e organizzazione
 - Utilizzate rientri, spanners e pannelli
 - Organizzate logicamente righe e colonne (per le tabelle nel testo)
 - Usate il numero minimo possibile di cifre significative (in funzione dei vostri obiettivi e delle convenzioni)
 - Siate coerenti nell'uso degli stili e dei formati
- Verificate che la tabella sia comprensibile indipendentemente dal testo

Rappresentare i dati: un breve corso di bricolage

- Scrivere una relazione
- Scrivere di numeri
- Tabelle
- **Grafici**

Perché e quando i grafici?

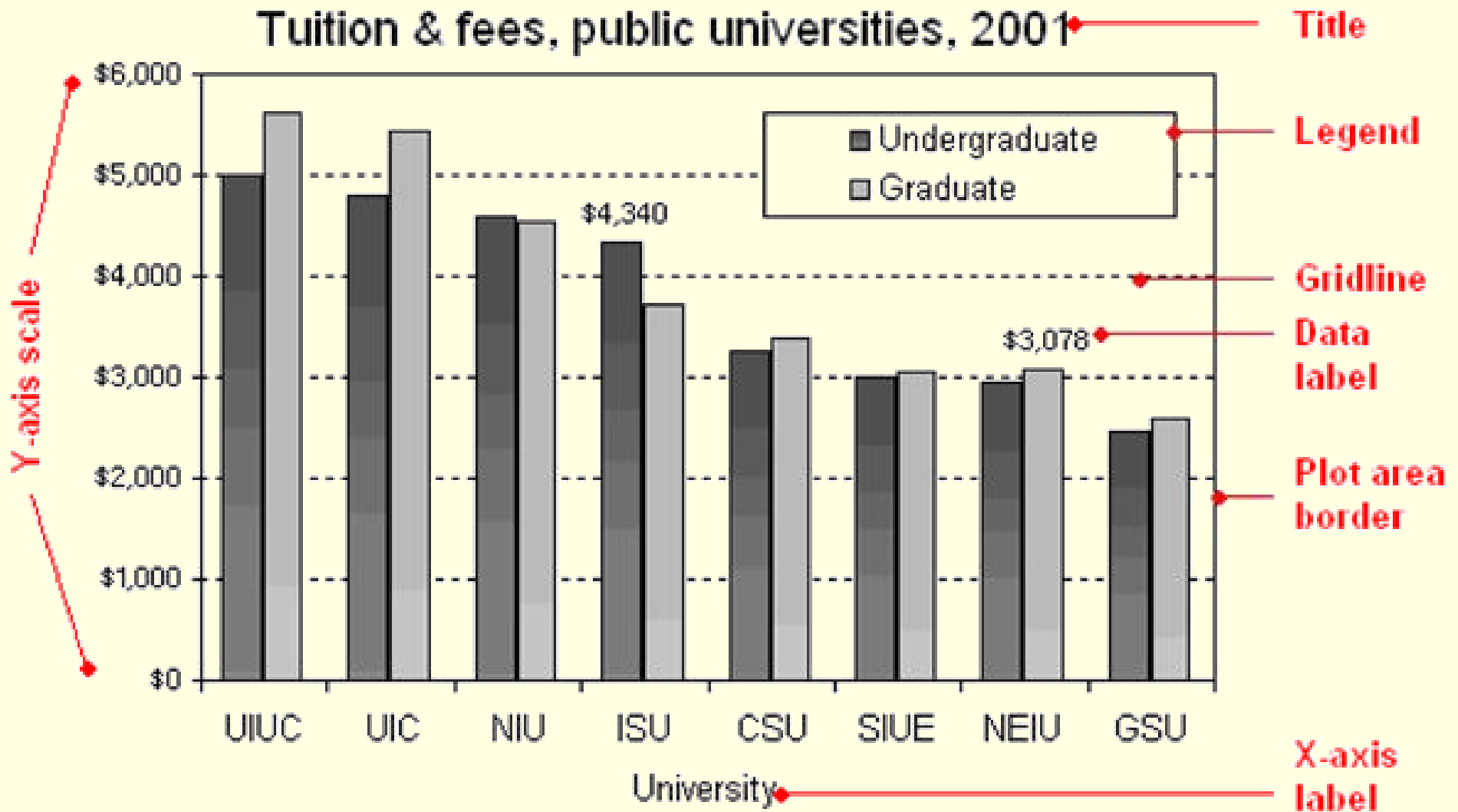
- Una rappresentazione grafica offre una presentazione visiva di dati che altrimenti sarebbero presentati in una tabella
 - a sua volta, una tabella presenta dati che altrimenti sarebbero descritti in un testo
- Idealmente, una rappresentazione grafica deve trasmettere al lettore informazioni che non sarebbero altrettanto efficacemente presentate con una tabella o con una descrizione testuale
- Quando un grafico è ben costruito?
 - deve presentare dati **dotati di senso**
 - i dati **non** devono essere **ambigui**
 - deve trasmettere informazioni in modo **efficiente**

Evitare ogni ambiguità

- L'ambiguità di un grafico è legata all'incapacità di definire con precisione ciò che i dati rappresentano
- Ogni punto in un grafico a dispersione, ogni linea di una serie storica, ogni barra di un grafico a barre rappresenta un dato
- Compito della meta-informazione del grafico spiegare esattamente che cosa ciascuno dei numeri rappresenta
 - Se un numero in un grafico è 33,5, il testo che lo accompagna – nel titolo, nelle intestazioni degli assi, nelle etichette dei dati, nella legenda, nelle eventuali note – deve rispondere alla domanda *"33,5 che cosa?"*

Anatomia d'un grafico

Major Components of a Chart



Come creare un buon grafico

- Regole simili a quelle delle tabelle
 - Date un titolo che consenta di comprendere il grafico indipendentemente da testo e tabelle
 - Create un grafico auto-contenuto (contesto, unità di misura, etichette, fonti, note, definizioni e abbreviazioni)
 - Progettate il grafico in modo che illustri le regolarità presentate nella relazione
- Titolo
 - Contesto (le W)
- Titolo ed etichette degli assi
- Legenda
- Etichette dei dati (soltanto se necessarie)

Rappresentazioni grafiche e immaginazione

- Una buona rappresentazione grafica richiede qualcosa di più di una buona tabella
- Accanto al rigore scientifico, anche una certa dose di creatività
- È necessario conoscere e comprendere profondamente i propri dati, ma anche “figurarsi”, “immaginare” come il lettore visualizzerà gli elementi grafici rappresentati

Distorsione e distrazione

- Una buona rappresentazione grafica comunica informazioni sui dati che non sono evidenti in una tabella
- Una cattiva rappresentazione grafica **distorce** o **nasconde** l'informazione contenuta nei dati:
 - **Distorsione**: gli elementi di un grafico non offrono una rappresentazione corretta del contenuto dei dati
 - **Distrazione**: gli elementi di un grafico distraggono il lettore dalla percezione del contenuto dei dati

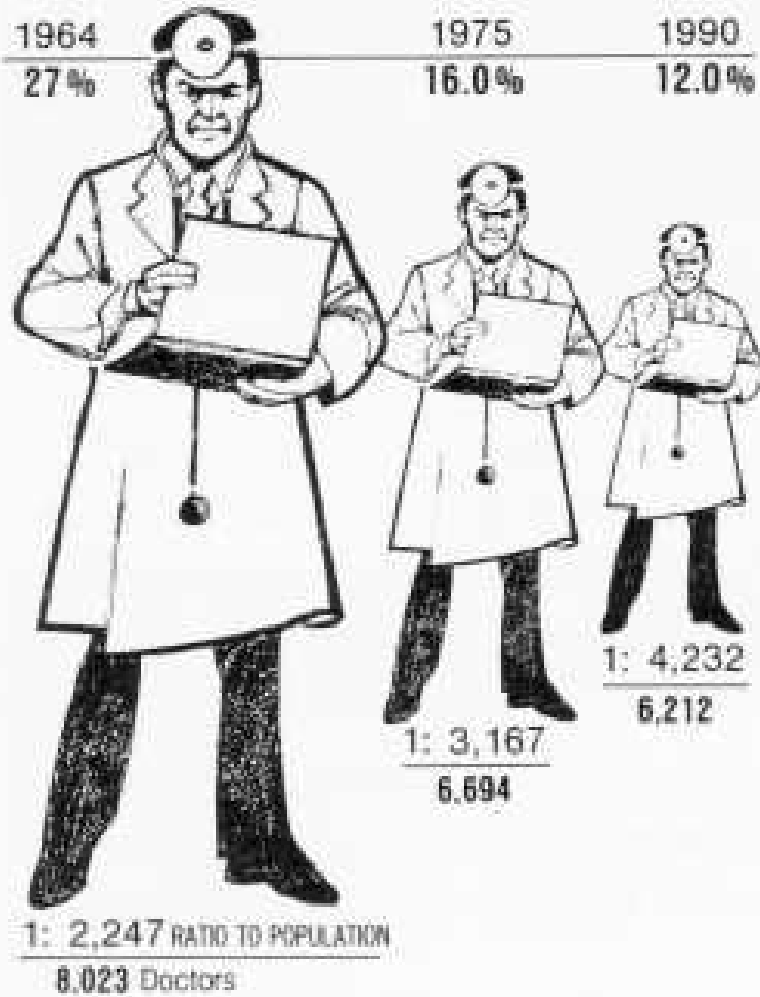
Distorsione e pittogrammi

- Pittogrammi bi-dimensionali, in cui l'altezza dell'oggetto è proporzionale alla misura del dato
- Se tanto l'altezza quanto la larghezza dell'oggetto variano proporzionalmente alla misura del dato, si esagerano le differenze di grandezza
- Il lettore tende a percepire l'area della figura, piuttosto che una dimensione lineare (in questo caso l'altezza), come rappresentativa del dato
- Quando è rappresentato in due dimensioni un oggetto tridimensionale, siamo addirittura portati istintivamente a considerarne il volume

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



Los Angeles Times, August 5, 1979, p. 3-

Discussione

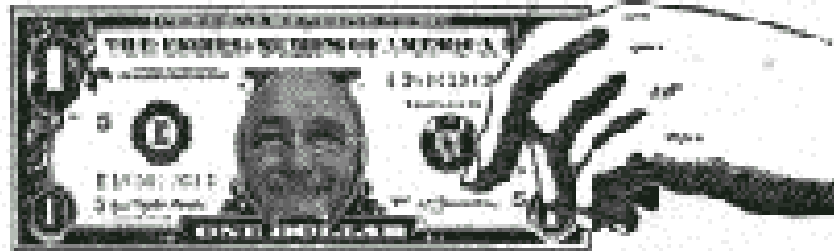
- Nel grafico, tutte le immagini hanno le stesse proporzioni
- Il medico del 1990 è alto poco meno della metà di quello del 1964
- Riflettete un momento:
 - se il primo fosse alto 160 centimetri, il secondo sarebbe alto più di tre metri
 - ma se il primo pesasse 60 chili (una misura del volume), il secondo – a parità di struttura fisica – quanti chili peserebbe?
- Il “fattore menzogna” (vedi oltre) è 2,8



1958 — EISENHOWER: \$1.00



1963 — KENNEDY: 94¢



1968 — JOHNSON: 83¢



1973 — NIXON: 64¢



1975 — CARTER: 44¢

(August)

Purchasing Power of the Diminishing Dollar

Source: Labor Department

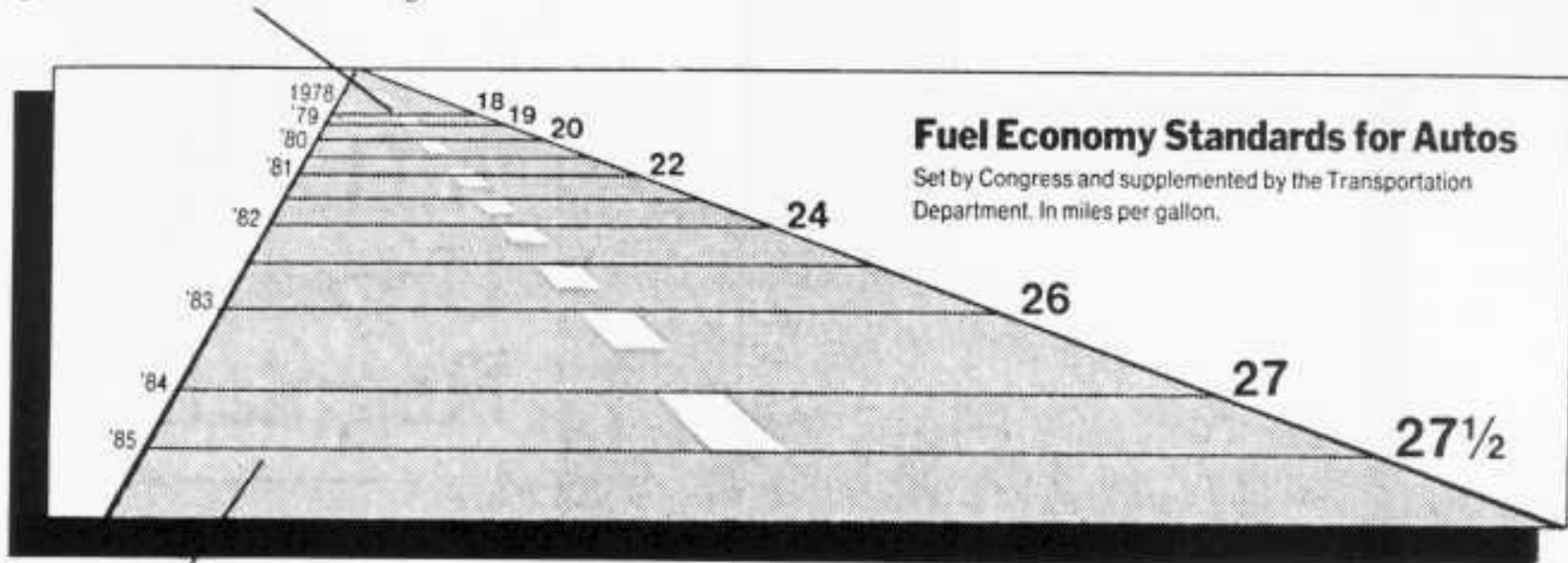
Discussione

- La larghezza delle banconote è utilizzata per rappresentare l'erosione del valore della moneta dovuto all'inflazione, da un dollaro nel 1958 a 46 cent nel 1973
- Ma il disegnatore ha ridotto sia l'altezza sia la larghezza delle banconote: in termini di area, il dollaro del 1958 è cinque volte più grande di quello del 1973.
- La distorsione – misurata da quello che Tufte chiama “fattore menzogna”, definito come il rapporto tra l'effetto mostrato nel grafico e il dato reale – è di cinque a due

Il principio di Tufte e il “fattore menzogna”

- Le rappresentazioni grafiche si fondano sull'assunto che un valore numerico è rappresentato visivamente dalla grandezza di un elemento grafico
- “La rappresentazione di un numero, misurata fisicamente sulla superficie del grafico stesso, deve essere direttamente proporzionale alle quantità rappresentate” (Tufte)
- La violazione di questo principio è misurata dal “**fattore menzogna**”, definito come il rapporto tra la misura dell'effetto mostrato graficamente e la misura del dato

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

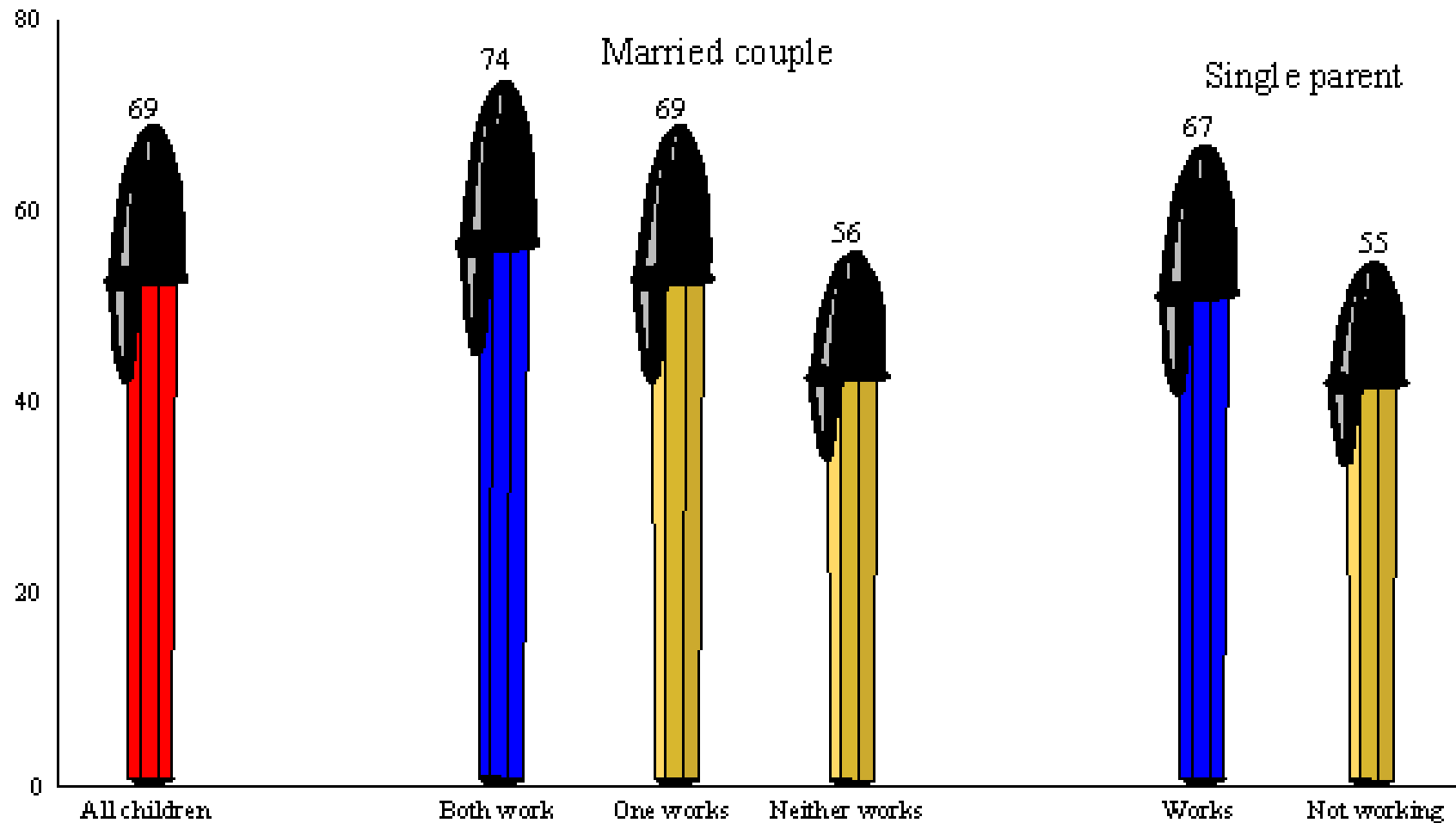
New York Times, August 9, 1978, p. D-2.

- Il grafico mostra l'evoluzione nel tempo degli standard di consumo energetico massimo stabiliti dal Ministero dei trasporti statunitense
- Tra il 1978 e il 1985, gli standard sono passati da 18 a 27,5 miglia per gallone, un incremento del 53%
- L'incremento rappresentato sul grafico – dato dal rapporto tra le linee relative ai due anni – è del 783%
- Il fattore menzogna è quindi $(783/53)$ pari a 14,8!

Dalla distorsione è difficile sfuggire

- Se si prova a non distorcere i dati, il risultato è spesso quello di distorcere l'immagine
- Inevitabile quando si cambia una sola dimensione in un'immagine bidimensionale
- La conseguenza: si perde l'effetto che si cercava
- Tre esempi tratti da una pubblicazione dell'US Census Bureau

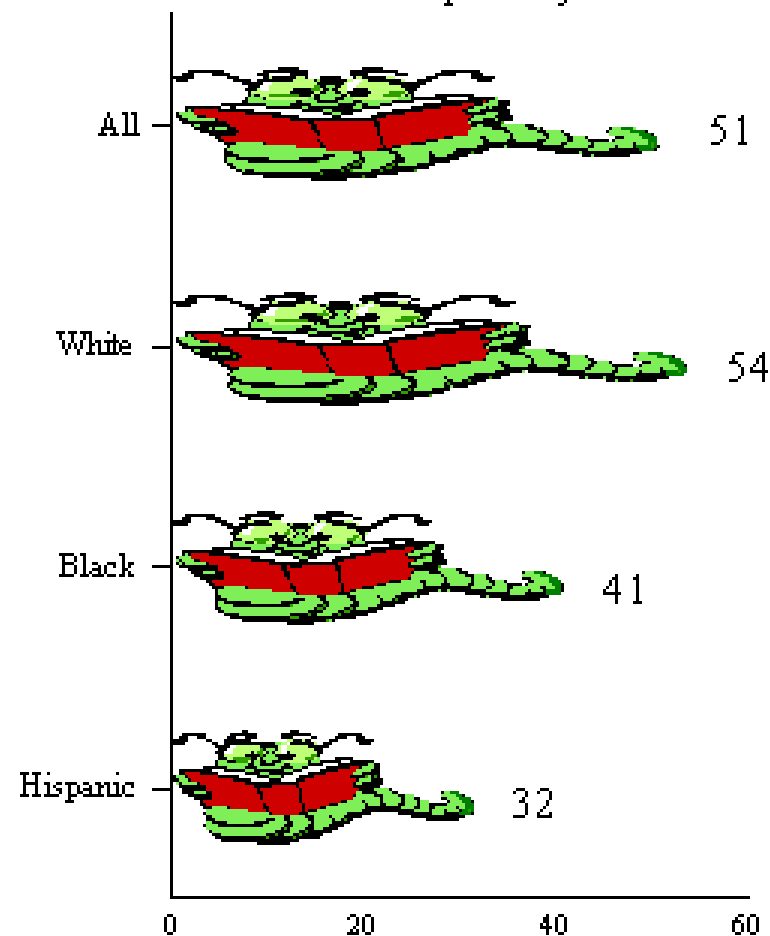
10. Children 12-17 Years Old Who are “On-track”,
 by Marital Status and Work Status of Parent(s)
 (Percent of children “on-track”)



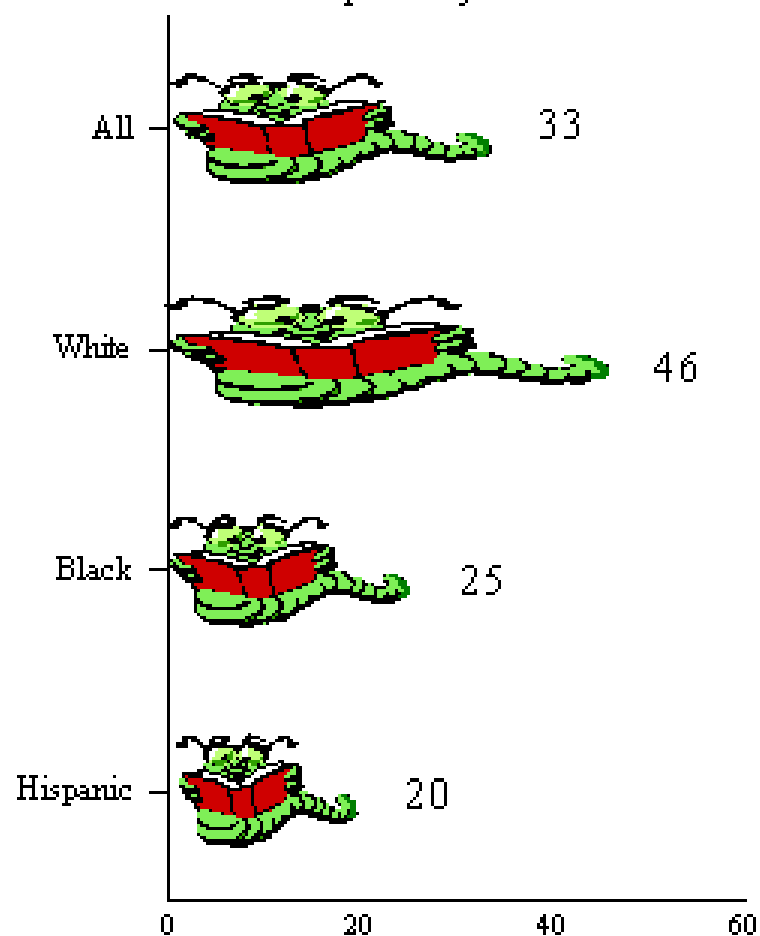
4. Times per Week 3-5 Year Olds Are Read to by Race / Ethnic Group* of Parent and Poverty Status

(Percent of children read to by their parents)

Percent above or on the poverty level



Percent below the poverty level



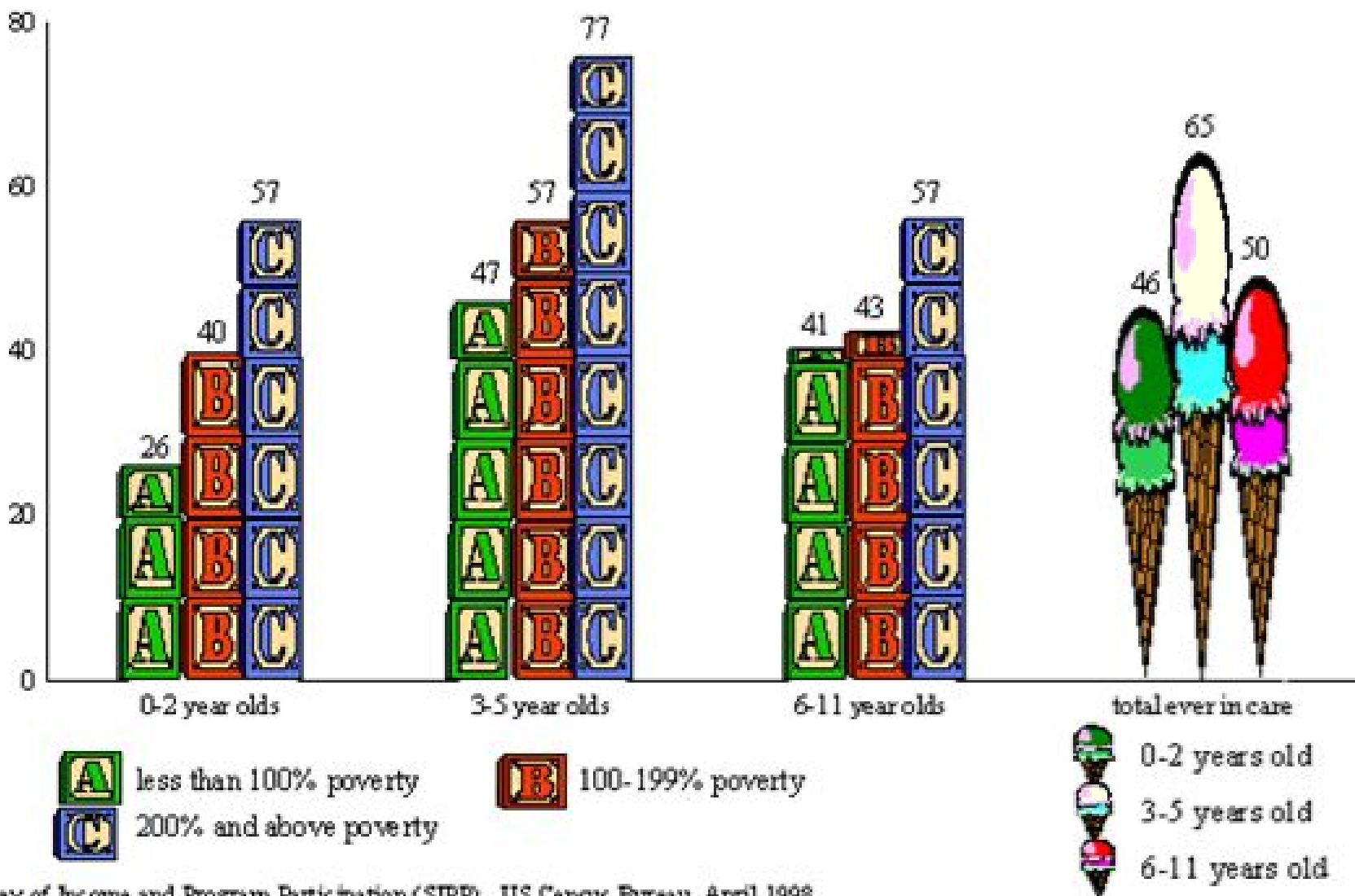
7+ times per week

* White and Black races exclude people of Hispanic origin.

Survey of Income and Program Participation (SIPP), US Census Bureau, April 1998

1. Children Ever in Child Care by Poverty Status and Age Group

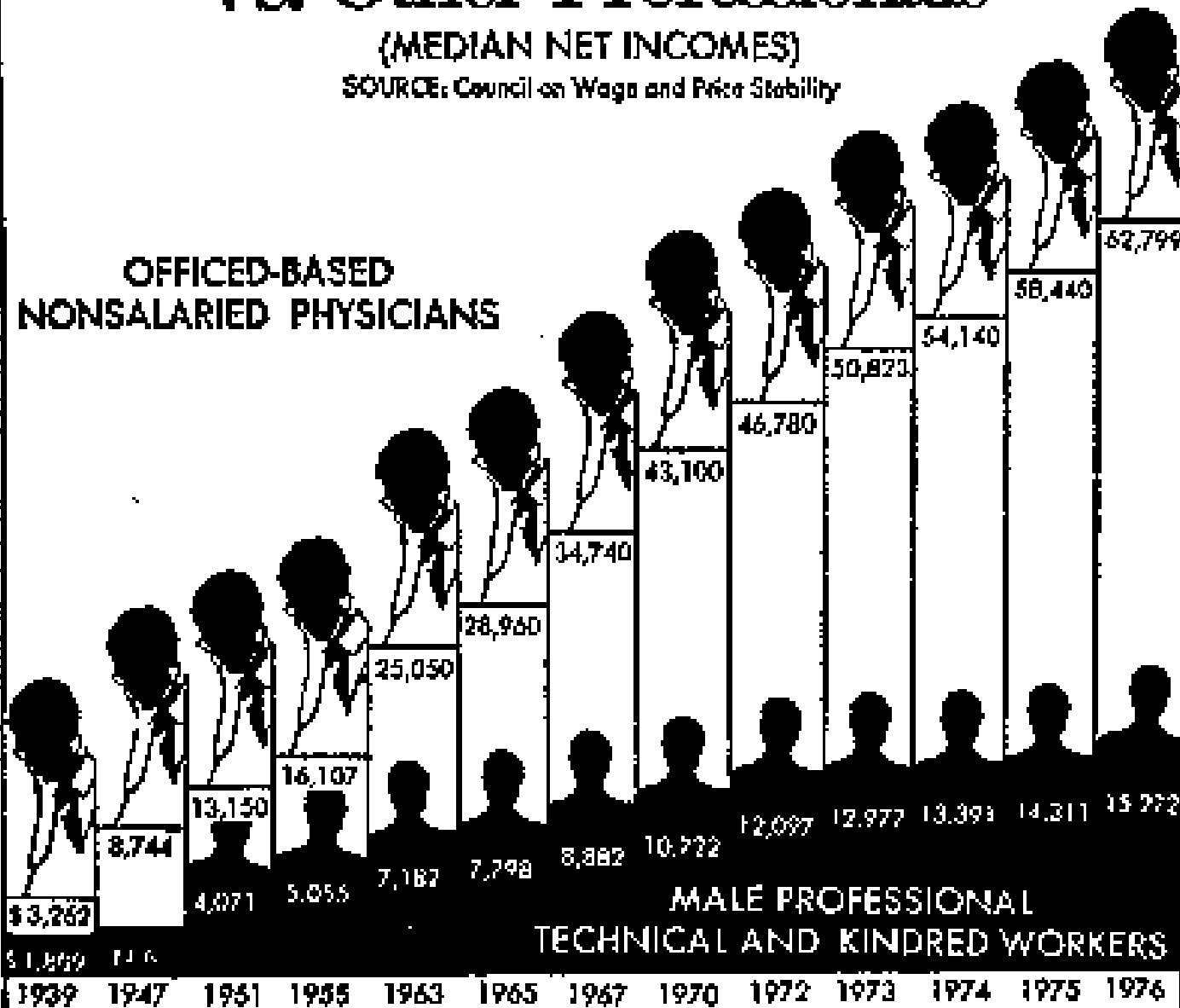
(Percent of children ever in child care)



Incomes of Doctors Vs. Other Professionals

(MEDIAN NET INCOMES)

SOURCE: Council on Wage and Price Stability



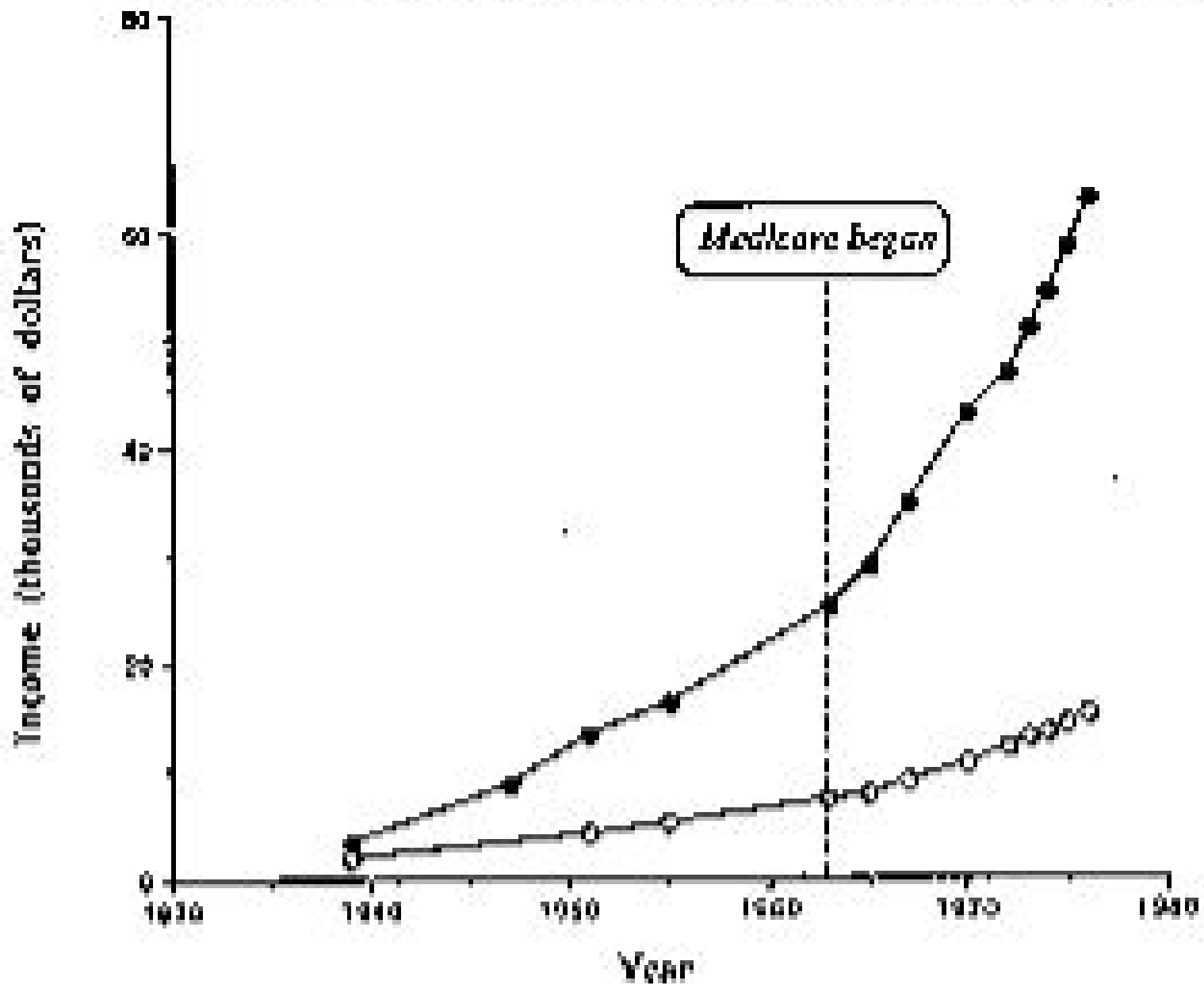
Scale elastiche

- Un modo più insidioso di distorcere i dati consiste nel cambiare scala all'interno del grafico
- Nell'esempio (dal *Washington Post*) è rappresentata la crescita del reddito dei medici tra 1939 e 1976
- All'apparenza la crescita è lineare, con un rallentamento negli anni più recenti
- In realtà, all'inizio sono rappresentati intervalli pluriennali, e poi annuali
- La scala delle ascisse è quasi nascosta!

Come ridisegnarlo?

- Rispettando la scala delle ascisse (distanze eguali per intervalli temporali eguali) emerge un quadro interpretativo diverso:
 - il reddito dei medici ha un andamento esponenziale
 - quello degli altri professionisti, lineare
- Introducendo il riferimento all'entrata in vigore dell'assistenza sanitaria (medicare), il quadro cambia ancora:
 - crescita lineare
 - ma articolata in due sotto-periodi

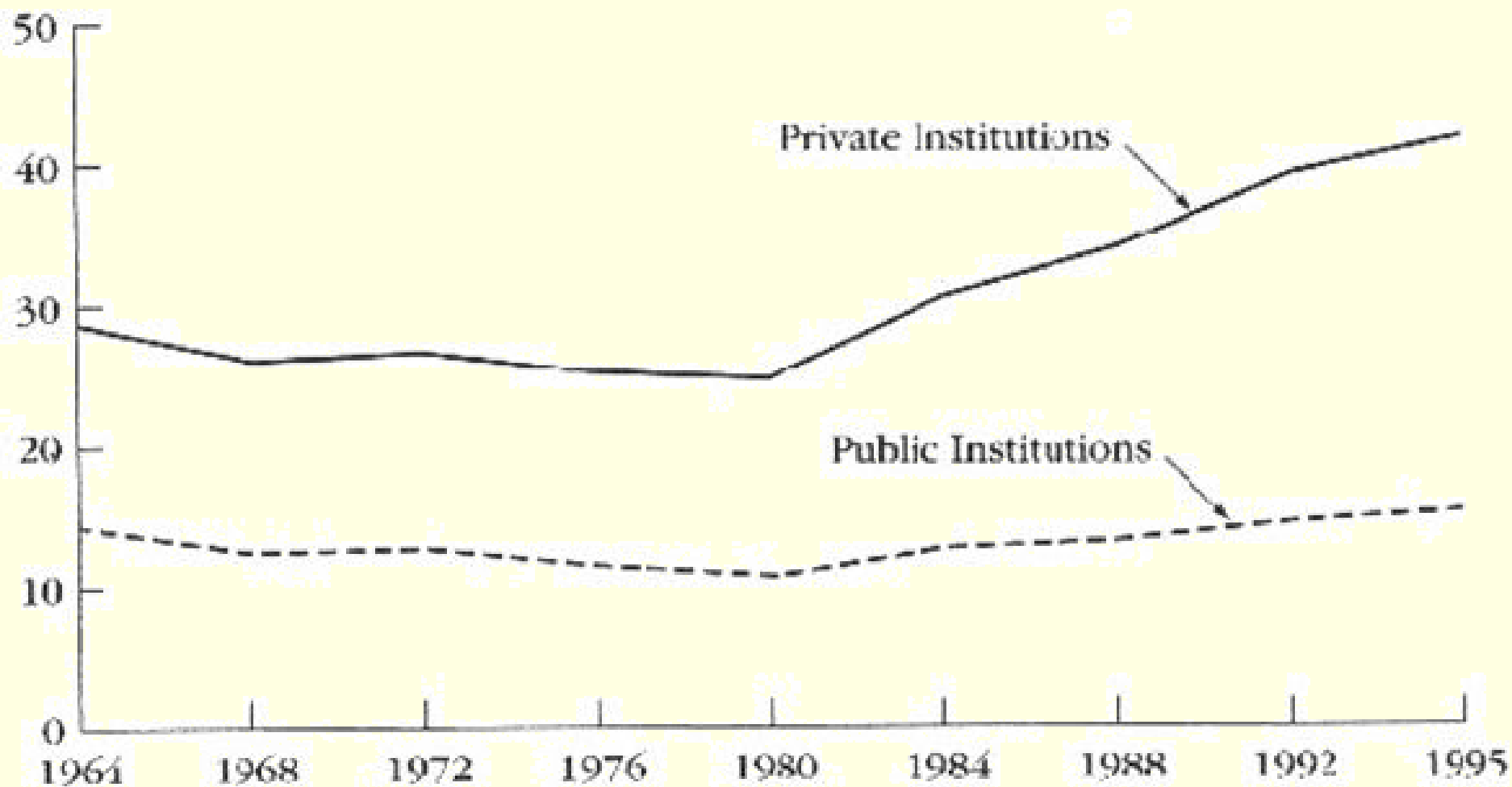
Physicians' income has grown exponentially since 1939
Whereas other professionals' income has gone up linearly



Fogli elettronici e distorsione

- Con la diffusione dei grafici ottenuti dai fogli elettronici, questo tipo di distorsione è diventata meno comune, almeno nei testi scientifici (sui giornali, questi errori sono ancora molto frequenti ...)
- L'arte di mentire con i grafici è ancora viva e vegeta, anche se è diventata più tecnologica e predominano altre forme di grafici ingannevoli

FIGURE 9-12 ■ Average Tuition, Room, and Board as a Percentage of Median Family Income, 1964–1995



SOURCE: U.S. Dept. of Education.

L'interpretazione degli autori...

- Ci sono evidenze empiriche che il costo dell'istruzione universitaria non sia aumentato in misura così importante [...]
- Nel grafico è rappresentato il costo medio di tasse scolastiche, vitto e alloggio come percentuale del reddito familiare mediano tra il 1964 e il 1995
- Mentre il costo delle università private è aumentato notevolmente, quello delle università pubbliche è rimasto costante
- Ciò indica che l'aumento dei costi associati all'istruzione universitaria è ragionevole se comparato con l'andamento dei livelli di reddito familiare

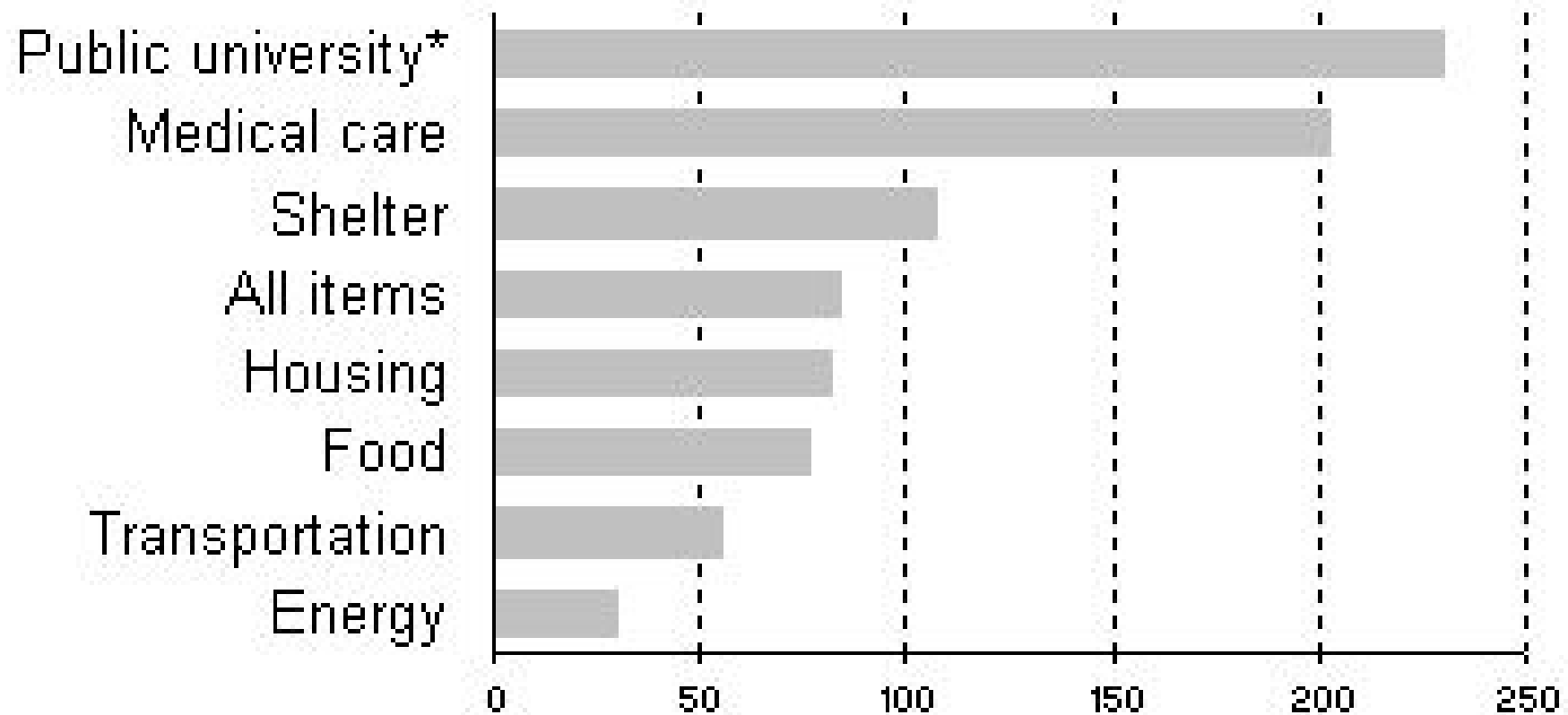
... e la sua critica

1. I costi sono “deflazionati” usando il reddito familiare mediano piuttosto che con l’indice dei prezzi al consumo; ma soprattutto dopo il 1982, il primo è cresciuto molto più rapidamente di quest’ultimo
2. Rappresentare nel medesimo grafico i dati relativi all’istruzione universitaria pubblica e privata amplia la scala sulla quale sono presentati i dati relativi all’istruzione universitaria pubblica, il cui andamento appare di conseguenza “compresso”. Non è facile percepirlo dal grafico, in cui l’andamento dei costi dell’istruzione universitaria pubblica appare quasi orizzontale, ma l’incidenza di tali costi sul reddito familiare mediano è passato dall’11 al 15 per cento, cioè quasi di un terzo
3. La scala temporale parte dalla metà degli anni Sessanta, ma la crescita effettiva dei costi dell’istruzione ha luogo a partire dal 1980

Un modo alternativo di guardare ai dati

- Confrontando la dinamica dei costi dell'istruzione universitaria con quella di altre grandezze economiche, emerge un quadro completamente differente
- Tra il 1981 e il 1999 – durante la vita degli attuali studenti universitari – i costi dell'istruzione universitaria pubblica sono aumentati più di quelli relativi ad altri settori dell'economia
- Persino più di quelli dell'assistenza sanitaria pubblica, a proposito della quale nello stesso testo citato in precedenza si affermava che "in campo medico la lievitazione dei costi è stata sostenuta e costante"

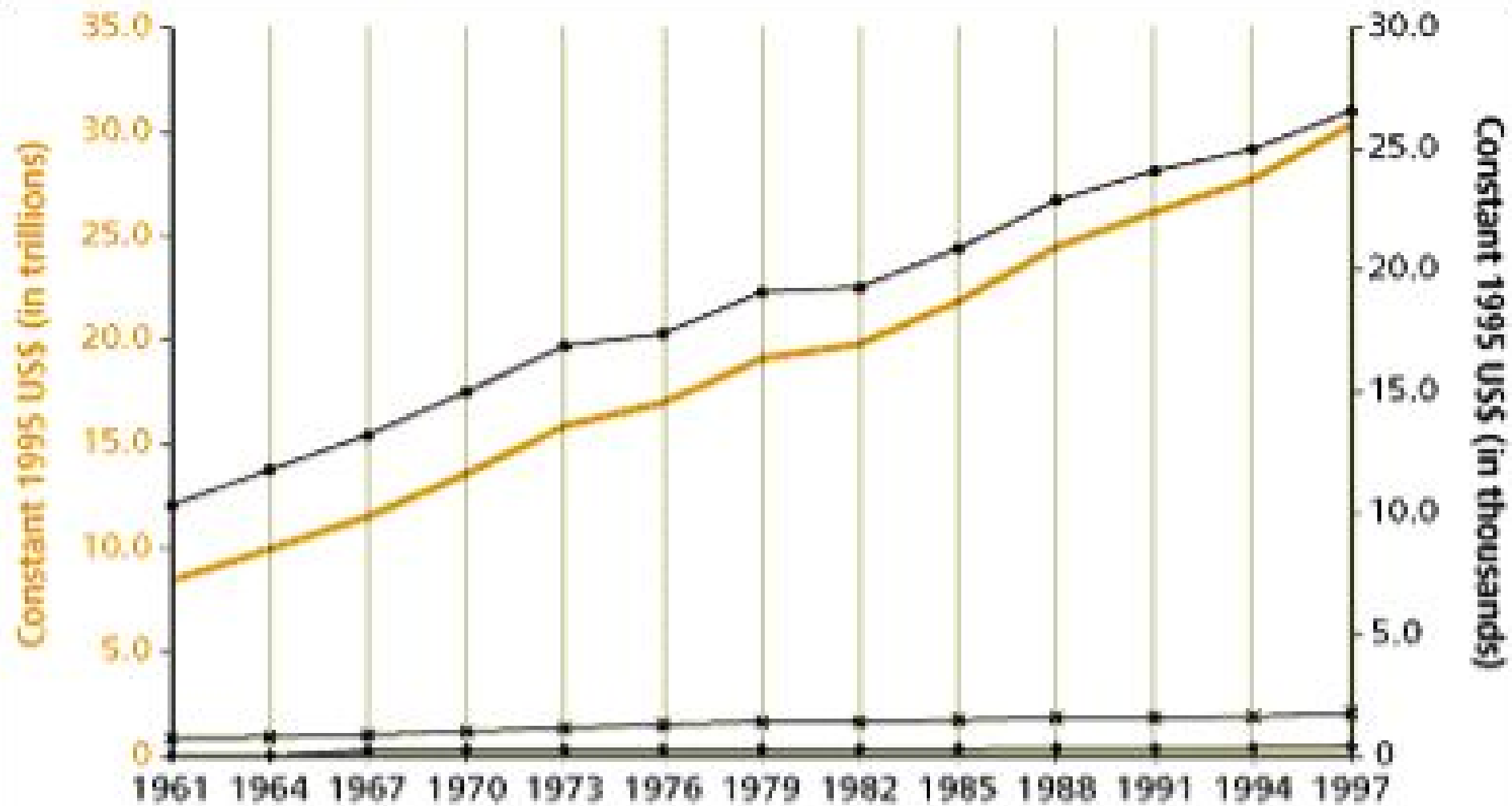
% Increase in Prices, by Economic Sector, 1981-1999



*Public 4-year universities, tuition, room and board
source: US Statistical Abstract 2000 CD-Rom, table 311

Il divario tra paesi ricchi e poveri

- Il divario tra Paesi poveri e Paesi ricchi si va allargando?
- Il Pnl *pro capite* dei paesi ricchi è più che raddoppiato (passando da 12.000 a 26.000 dollari)
- Per i paesi a reddito basso o medio è impossibile capire in che misura il Pnl *pro capite* è cresciuto: più o meno del doppio? O forse è triplicato?



GROSS NATIONAL PRODUCT

— World

GROSS NATIONAL PRODUCT (per capita)

—●— High-income countries

—■— Middle-income countries

—◆— Low-income countries

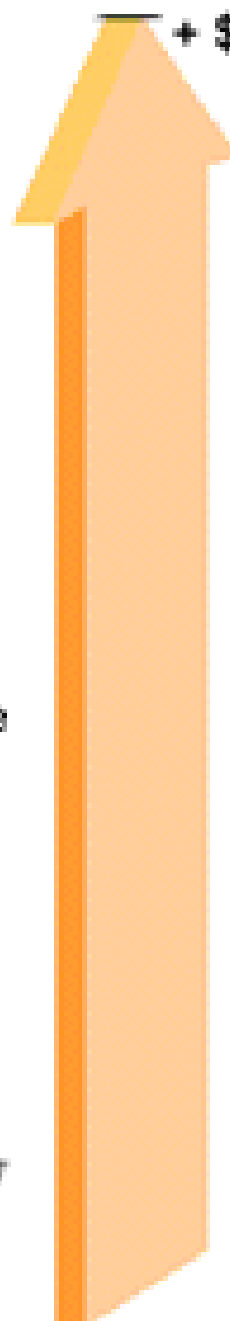
GNP change
per person
(\$US)
1990-1997
for donor
nations

+ \$7,878

ODA change
per person
(\$US)
1990-1997
for donor
nations

1990

- \$18



Un altro esempio di distorsione

- Apparentemente i dati sono rappresentati graficamente dall'altezza delle frecce, ma non è così:
 - l'aumento di 7.878 dollari del Pnl *pro capite* dei paesi donatori è rappresentato dalla freccia di destra (e dal trattino orizzontale che ne misura l'altezza)
 - nella parte sinistra del grafico la diminuzione di 18 dollari dell'Oda (*Official Assistance to Development*) *pro capite* è correttamente rappresentato dal trattino, *ma non dalla dimensione della freccia!*

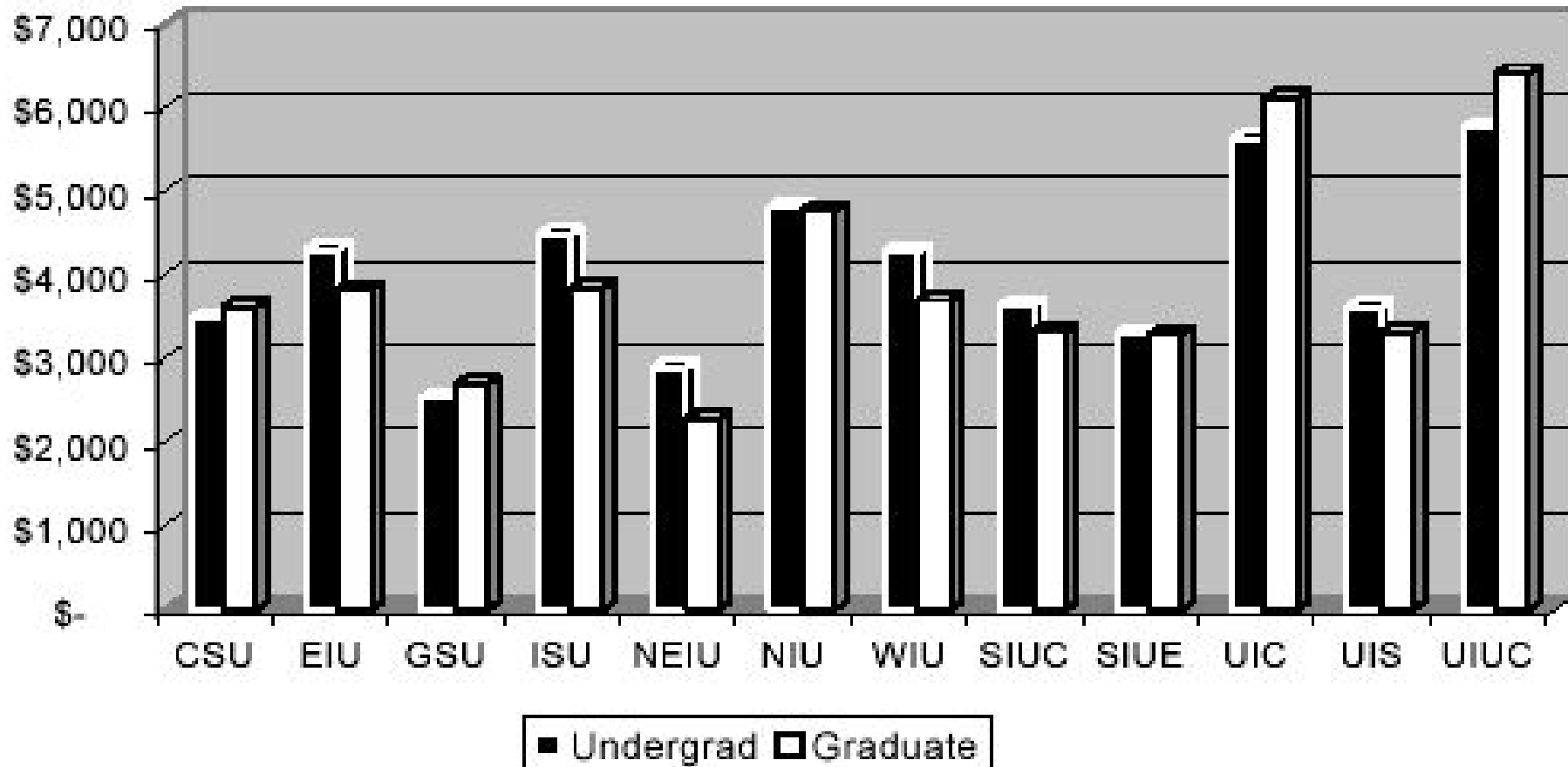
La regola aurea di Edward Tufte

- La regola aurea di una rappresentazione grafica efficiente: *minimizzare il rapporto inchiostro/dati*
- È una regola molto simile a quella presente in molti manuali di scrittura:
 - Una frase non deve contenere parole non necessarie e un paragrafo non deve contenere frasi superflue, allo stesso modo in un disegno non deve contenere linee non necessarie e una macchina non deve avere parti superflue (Strunk e White, *The Elements of Style*)

Distrazione e fogli elettronici

- Pacchetti grafici dei fogli elettronici → opzioni tridimensionali
- **Nessuna funzione utile**: aggiungono inchiostro al grafico e, in genere, rendono più laborioso percepire i valori che il grafico intende rappresentare
- A peggiorare la situazione, spesso si aggiungono le altre opzioni del foglio elettronico:
 - aggiungere una prospettiva
 - grafici fantasiosi (a torta, ad anello, a radar, a bolle, a cilindri, a coni, a piramide, ...)

Tuition & fees, public universities, 2001



Perché eliminare gli elementi di distrazione

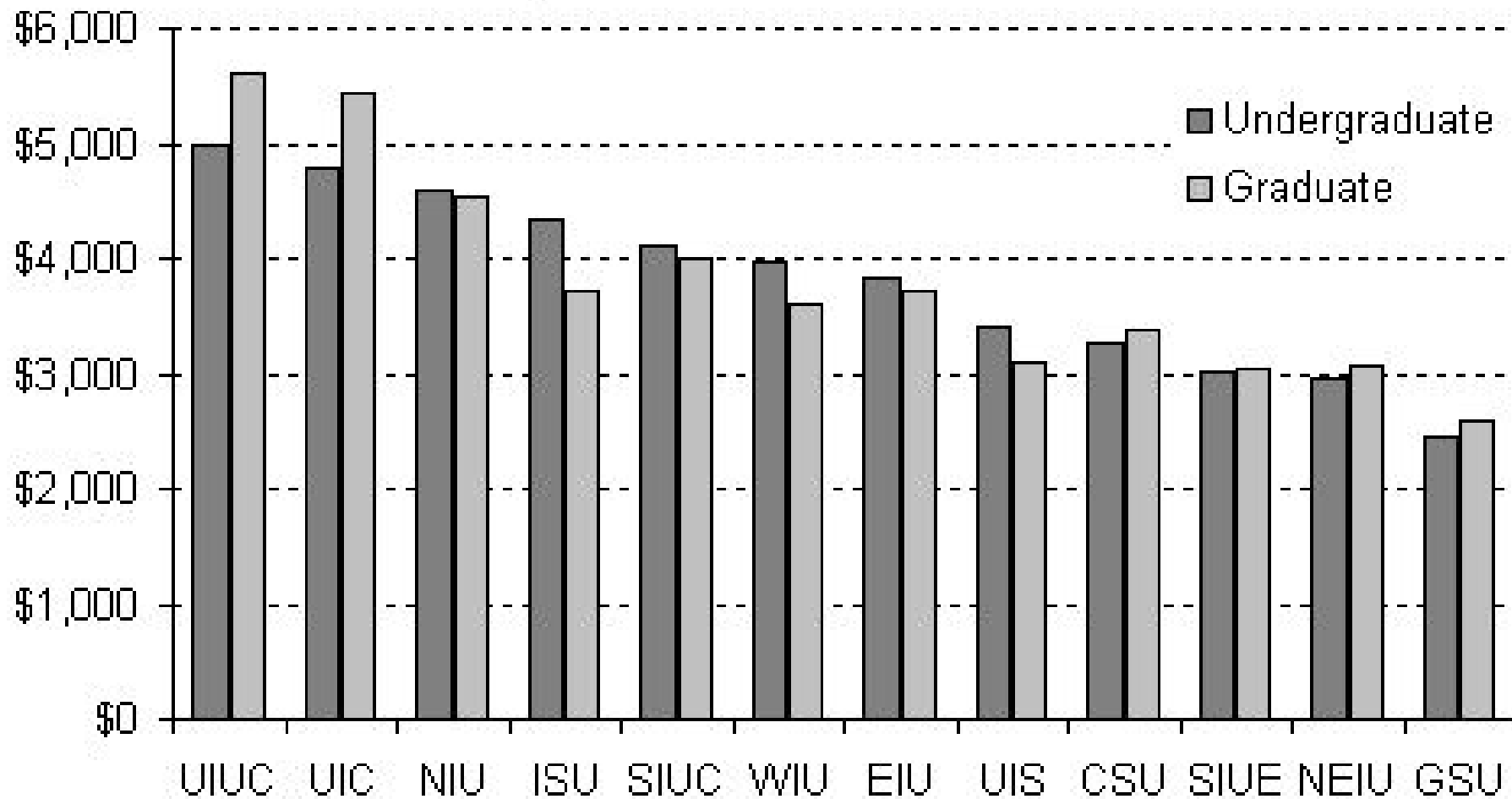
- Il corollario della regola aurea di Tufte è: *far vedere i dati*
- Troppo spesso, elementi grafici e di testo irrilevanti e superflui dominano il grafico, distraendo dalla sua funzione fondamentale, che è quella di presentare **visivamente** i dati
- Una buona rappresentazione grafica massimizza la presentazione dei dati e minimizza la rappresentazione degli elementi non necessari alla presentazione dei dati

Come eliminare gli elementi di distrazione

- Rimozione dell'effetto tridimensionale → permette di cogliere i valori con maggiore precisione
- Inserimento della legenda all'interno dell'area del grafico → consente di disporre di un asse delle ordinate più lungo, con più spazio per i dati
- Adozione dell'ordinamento basato su una delle variabili → facilita le comparazioni
- De-enfatizzare la griglia → induce a concentrarsi sulla lunghezza relativa delle barre

Il grafico ne guadagna in leggibilità ed efficacia

Tuition & fees, public universities, 2001

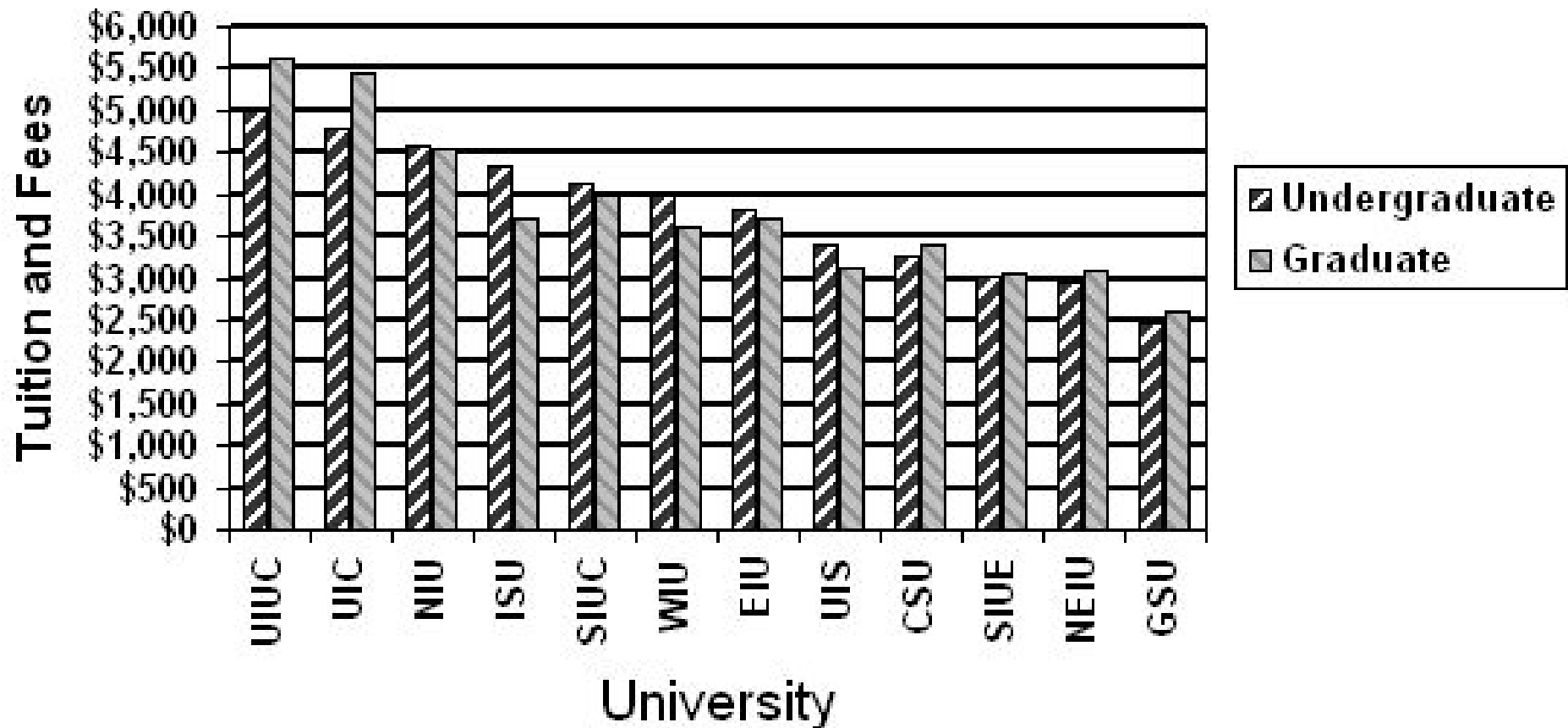


Al contrario...

- Se tutti questi elementi di contorno sono enfatizzati
 - ➔ perdita di leggibilità
 - ➔ distrazione dai dati

Over-emphasizing non-data elements:

Tuition & fees, public universities, 2001



Mostrare più dati

- Un modo di minimizzare il rapporto inchiostro/dati è quello di mostrare più dati
- Troppi dati in un grafico solo – per esempio, troppe linee in un grafico che presenta serie storiche – possono portare a un sovraffollamento del grafico
- L'errore opposto è molto più frequente: l'uso di più grafici quando uno solo sarebbe più che sufficiente
 - Se si fosse "spezzato" il grafico in due parti – una dedicata ai diplomati e una ai laureati – si sarebbe costretto il lettore a operare confronti *tra* grafici, invece che *all'interno* del grafico: cosa da evitare quanto più possibile

Ciarpame grafico

- Non è necessario fare quello che si può fare!
- È controproducente!
- È brutto!

CHART I

COMPARATIVE CONCENTRATION OF INCOME IN THE MAJOR WESTERN NATIONS (CIRCA 1980)

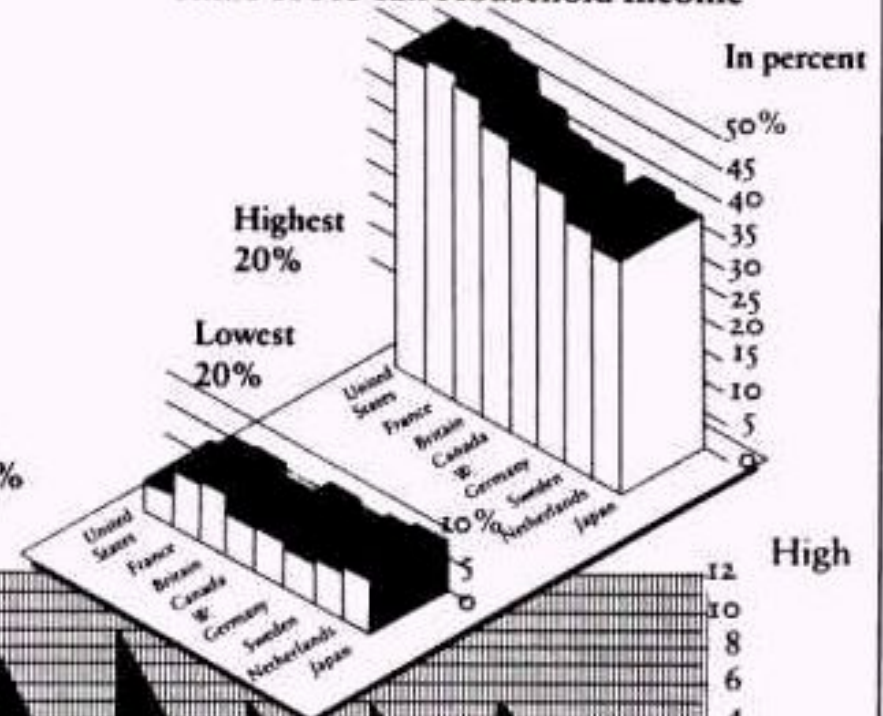
Pre-Tax Income Distribution in Industrial Nations



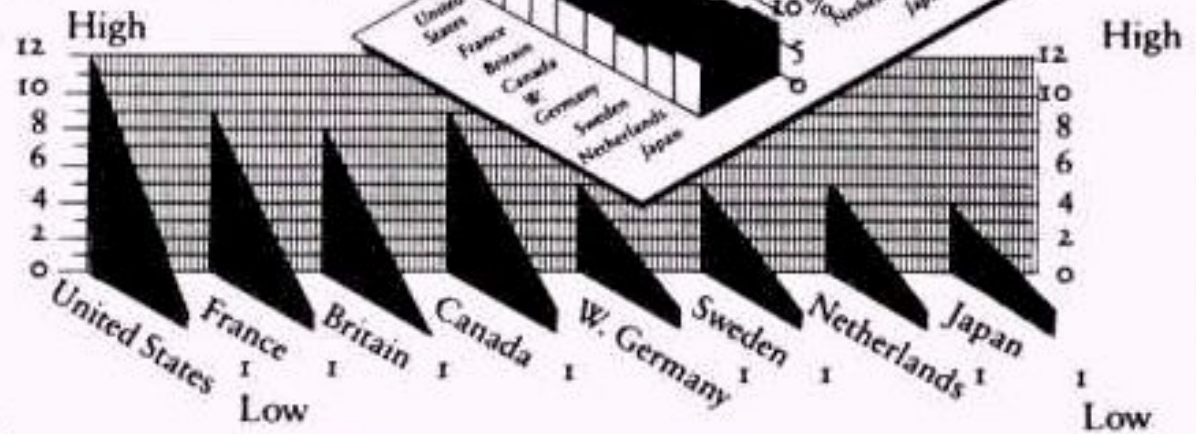
Note: Figures for household income generally show a greater gap than those for family income cited for the United States and Canada in the story and definitions of "household" vary somewhat among countries. Figures are for various years in the late 1970's and early 1980's.

Sources are individual governments and the World Bank.
 Source: *Los Angeles Times*, October 21, 1984
 Note: The ratios for Japan changed significantly in the late 1980s.

Share of Pre-Tax Household Income



The Ratio Between Highest and Lowest 20%



Caratteri spuri

- La cartina del mondo è assolutamente irrilevante
- Due tipi di grafici tridimensionali, completamente diversi, e presentati in una prospettiva diversa
- I nomi dei paesi sono ripetuti tre volte;
- Per rappresentare 24 dati, si utilizzano una trentina di numeri per le scale
- I paesi sono presentati senza ordine apparente (neppure in ordine alfabetico)
- Per separare i paesi nel grafico più in basso viene utilizzata la I maiuscola

Conclusione: sarebbe stata meglio una tabella!

Pre-Tax income Distribution in Industrial Nations

	Share of Pre-tax Household Income		Ratio: Top to bottom shares
	Top income quintile	Bottom income quintile	
United States	45	4	12
Canada	42	4	9
France	47	5	9
Britain	45	6	8
W. Germany	39	8	5
Sweden	38	8	5
Netherlands	37	7	5
Japan	36	9	4

*data estimated from chart.

Le famiglie di grafici

- Quattro grandi famiglie
 - grafici a dispersione
 - serie storiche (linee)
 - grafici a barre
 - grafici a torta
- Varianti e combinazioni
- La scelta tra i tipi di grafici dipende dai dati da rappresentare e dai rapporti che si vogliono mettere in luce

Quale grafico usare

- I grafici sono efficienti solo se i dati sono numerosi (regola generale)
- Soltanto se il confronto non è evidente in una tabella
- Grafici a dispersione e serie storiche
 - per mostrare relazioni tra due (o più) variabili
 - richiedono almeno due variabili quantitative
- Grafici a barre e a torta per le variabili categoriali
- A volte si usano i grafici a barre per le serie storiche

Grafici a barre e tabelle

- In genere i grafici a barre, soprattutto se rappresentano una sola variabile, non sono efficienti
- Anche se si minimizza l'inchiostro (via assi, scala, griglia, bordi ...)
- In realtà, le stesse barre non sono necessarie

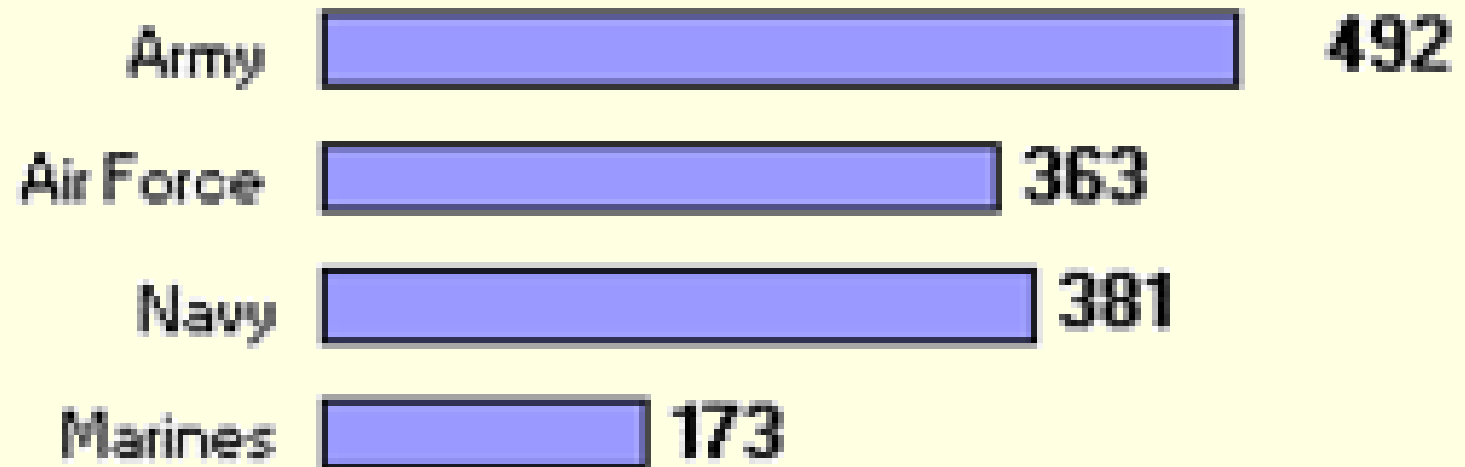
Ricordate la regola generale

qui ci sono soltanto quattro dati!

(giustificato soltanto se la dimensione relativa dei valori può sfuggire al lettore)

Un grafico inutile

Active Duty Personnel, 1998 (millions)



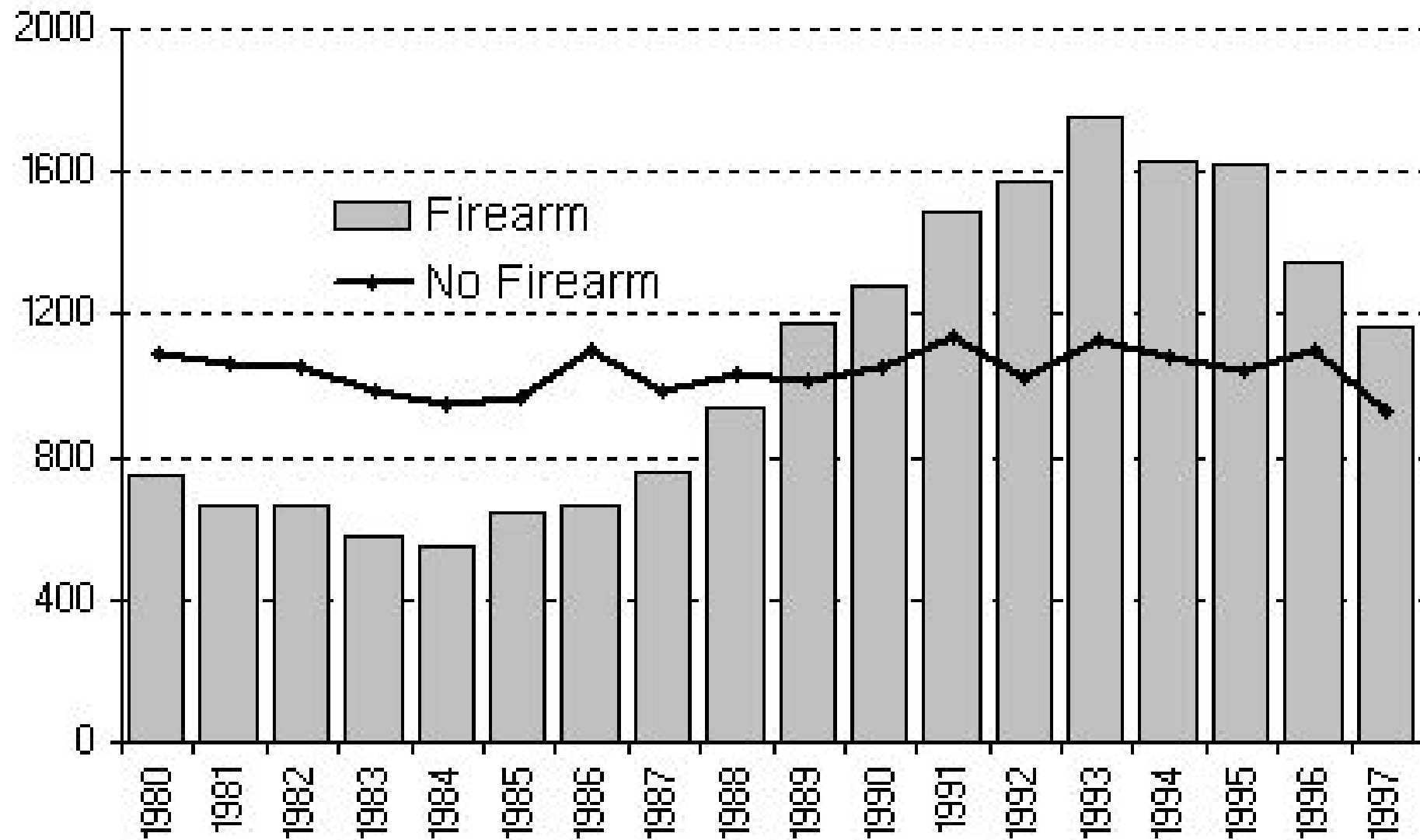
Grafici a barre e torte

- Le torte sono ancora più inefficienti
- Inoltre, sono esposte al rischio di distorsione
- Tutto ciò che può essere rappresentato in una torta può essere rappresentato anche in un grafico a barre

Serie storiche

- Il modo più efficiente di rappresentare grandi quantità di dati
- Agevolano confronti significativi
- Linee, ma anche grafici a barre

Juvenile homicide victims by firearm involvement



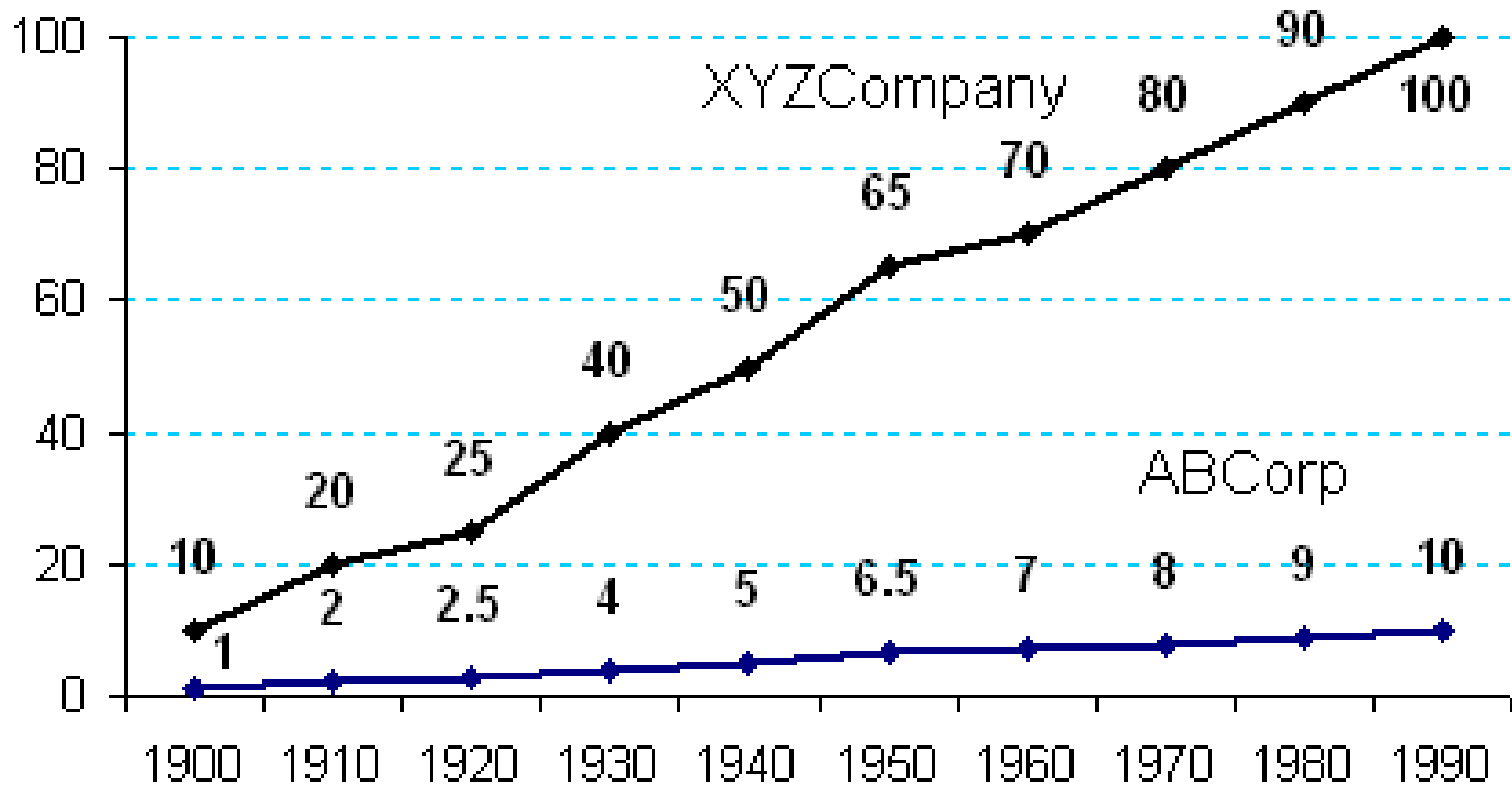
Standard e regole generali

- In genere il tempo va rappresentato in ascissa, da sinistra a destra
- Minimizzare l'inchiostro (regola generale)
- Evitare di sovraffollare il grafico
- Se si usa un grafico a barre, i punti che rappresentano valori in serie temporale devono essere adiacenti
- Se si usano le linee, devono essere chiaramente distinguibili quelle che rappresentano variabili diverse
- Se si usano variabili che rappresentano valori monetari, è in genere opportuno usare serie deflazionate ed espresse in valori *pro capite* (o in quota del PIL)

Problemi di scala

- Se sono rappresentate sullo stesso grafico due serie di grandezza molto diversa, la più piccola appare “piatta” (anche a parità di tasso di variazione)
- Può essere opportuno introdurre una seconda scala (a destra)
 - ma la differenza di grandezza lo deve giustificare
 - e c'è il rischio un forte rischio di **distorsione**

Figure 3: Stock Prices: Two Hypothetical Companies



Un caso dubbio

Figure 4a: Poverty and Unemployment 1959-00

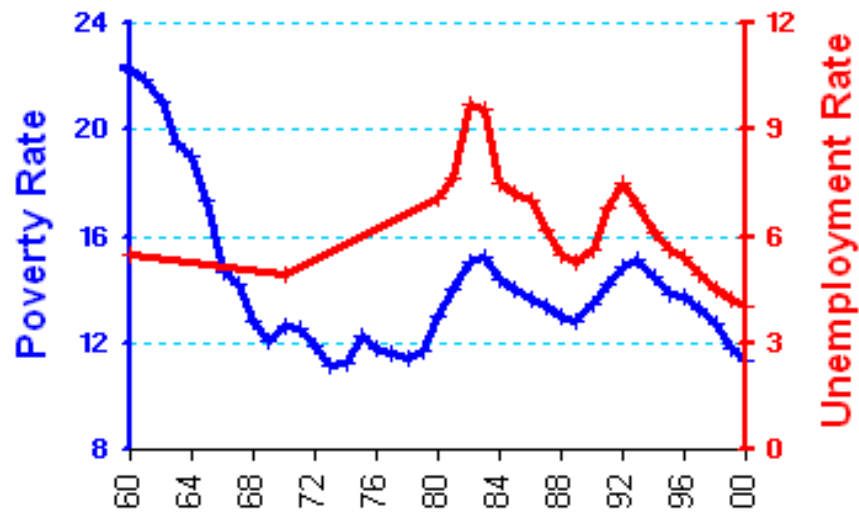
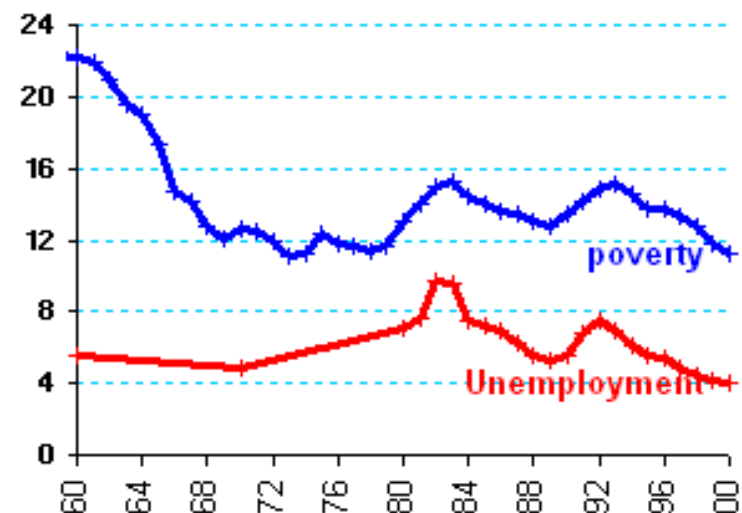
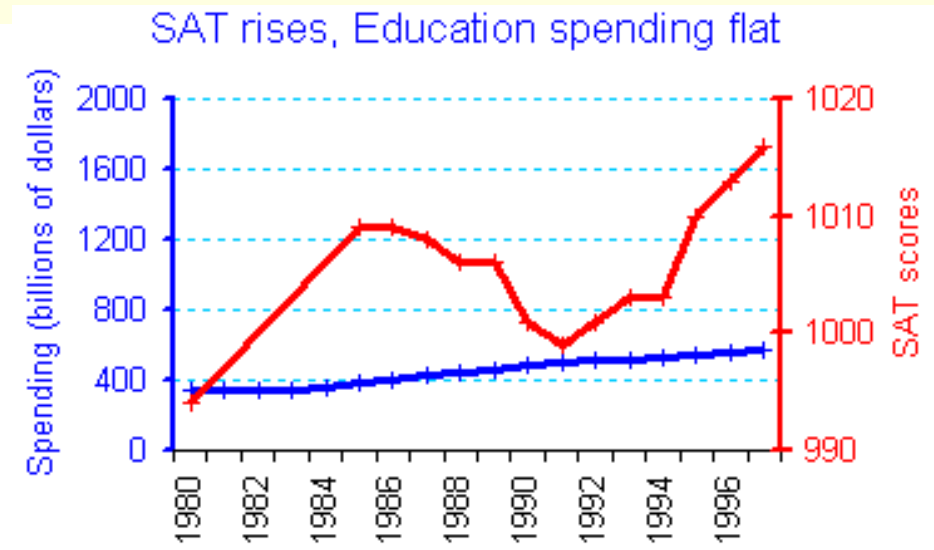
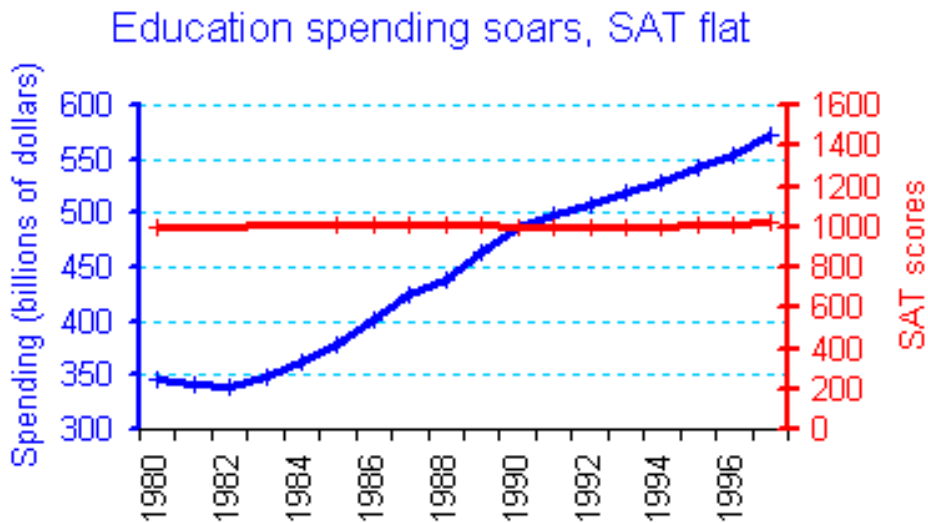


Figure 4b: Poverty and Unemployment 1959-00



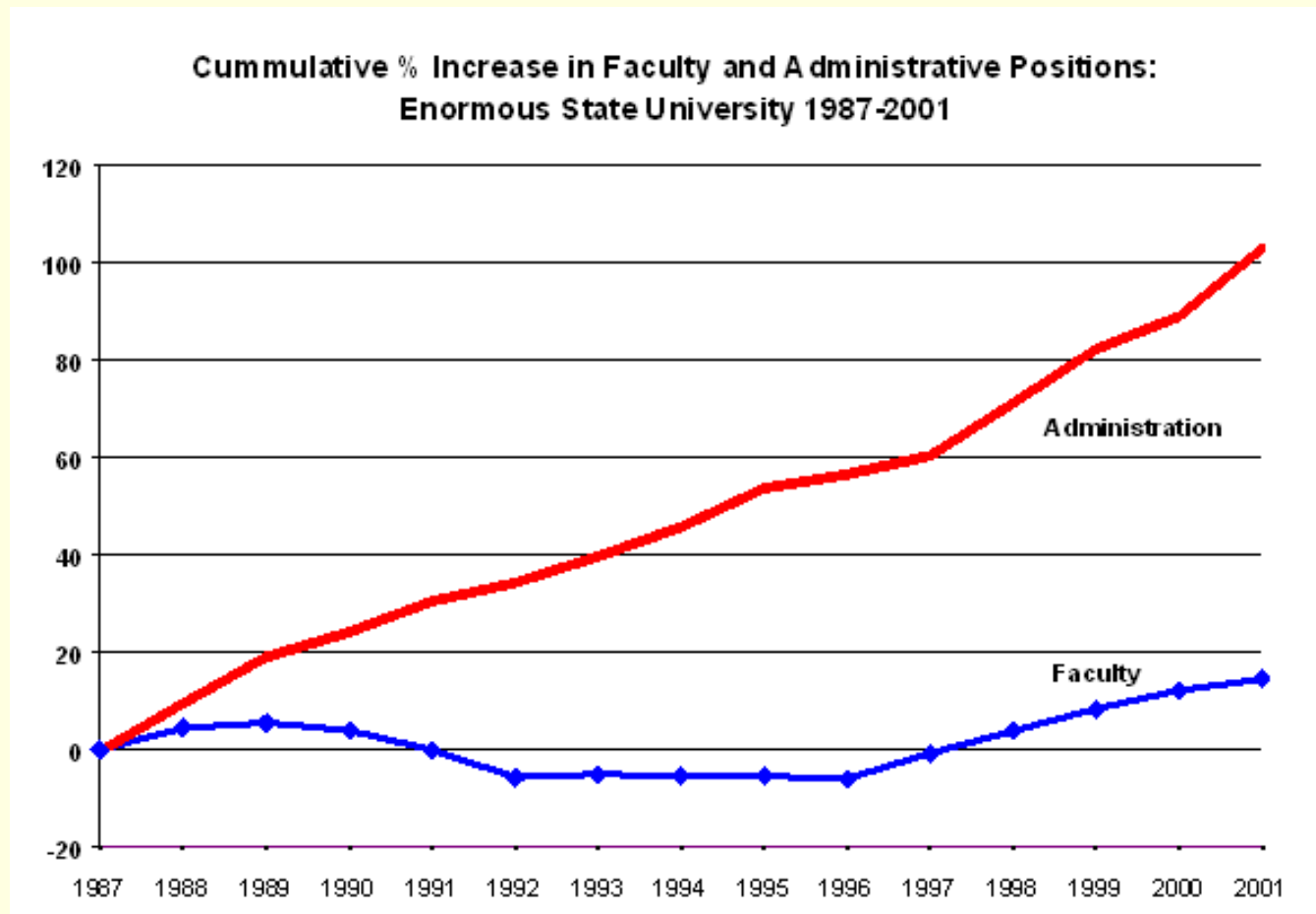
Scale arbitrarie e distorsione



Numeri indici

- L'uso dei numeri indici permette di "ri-scalare" le serie, evitando gran parte dei problemi citati

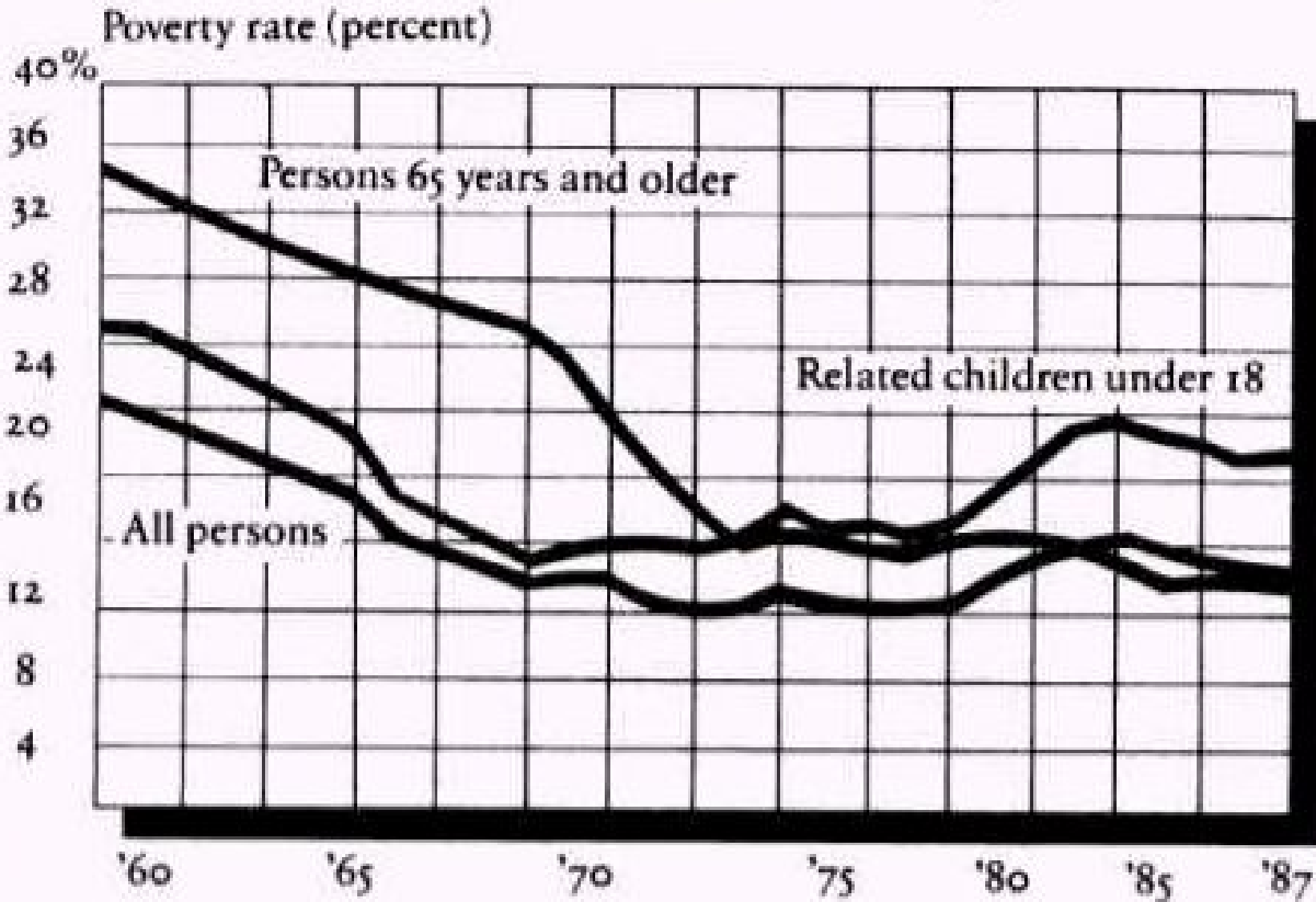
Re-scaling



Linee che si incrociano

- Quando sono rappresentate molte serie, spesso è difficili distinguerle, soprattutto se si incrociano
- Utilizzare i colori, ma tenere conto del fatto che la stampa potrebbe poi essere in bianco e nero
- Utilizzare artifici grafici (tratteggi, puntini, eccetera)

CHART 8
AGE AND WEALTH
The New Arithmetic of Poverty

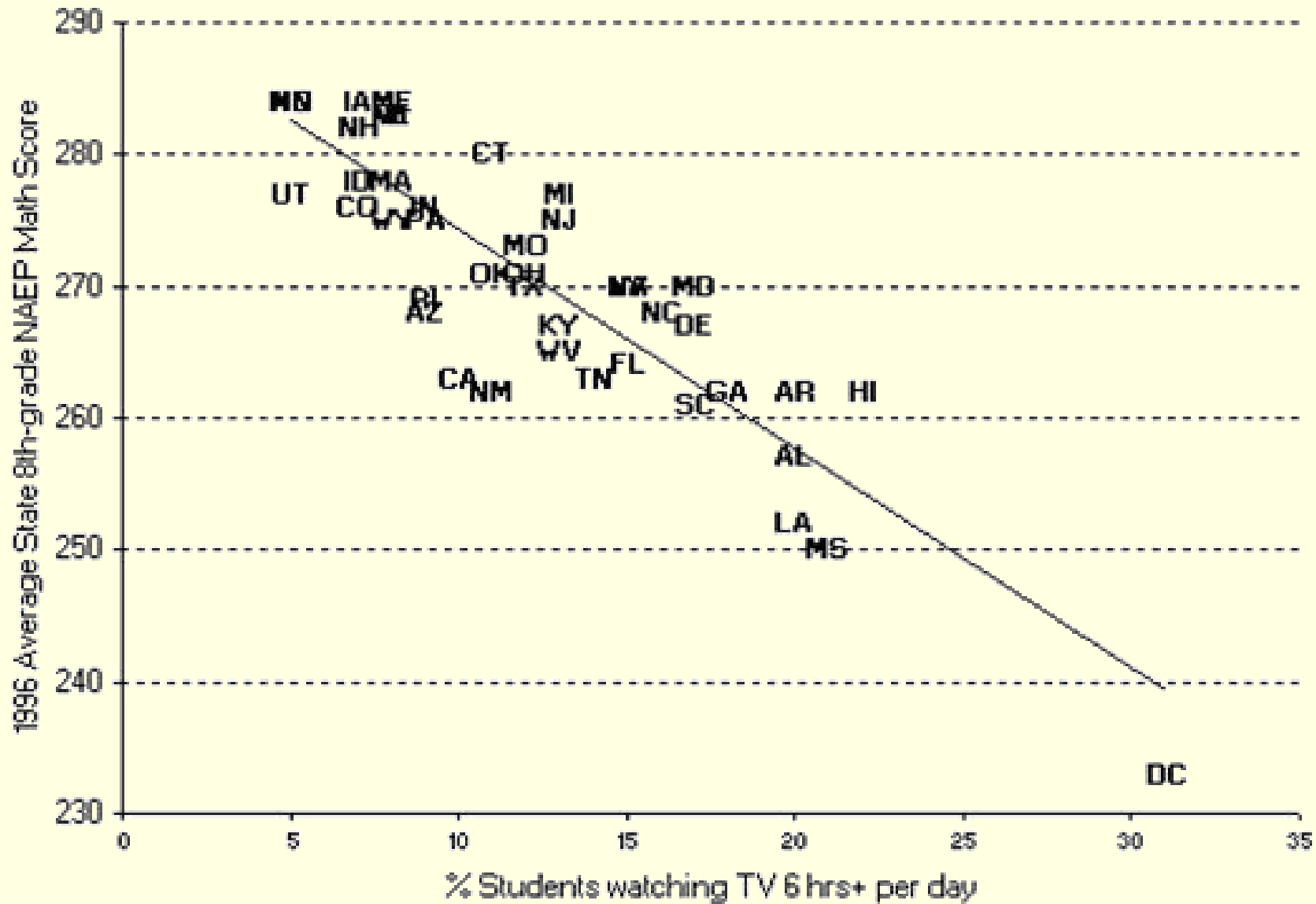


Source: U.S. Bureau of the Census

Grafici a dispersione

- Uno dei modi più efficienti per la rappresentazione grafica dei dati
- Particolarmente adatto a illustrare la relazione tra due variabili quantitative

State Math Scores and Students TV Viewing Habits



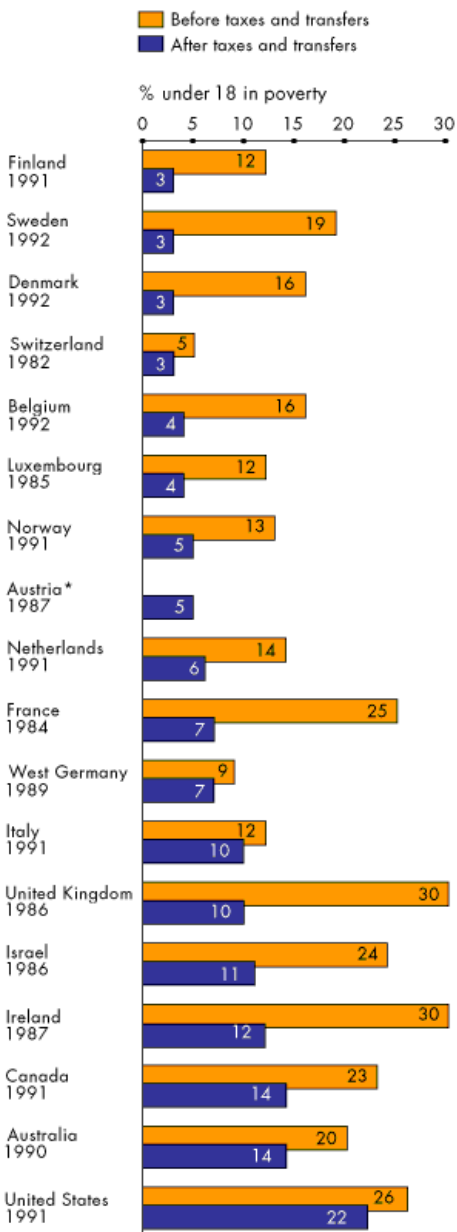
Standard e regole generali

- Variabili quantitative
- Individuare chiaramente le variabili (sulle intestazioni degli assi)
- Definire i casi (che cosa rappresentano i punti sul grafico: etichette individuali o legenda)
- La scala degli assi deve permettere la presentazione di tutti i dati
- Se si suggerisce, o è implicita, una relazione causale tra le variabili, la convenzione è di porre la variabile indipendente in ascissa e la dipendente in ordinata
- Si può aggiungere una retta di regressione
- In genere, la sovrapposizione dei punti non è un problema

Grafici a barre

- Quando non sono usati per le serie storiche, spesso sono poco efficienti rispetto a una tabella

Child poverty before and after government intervention

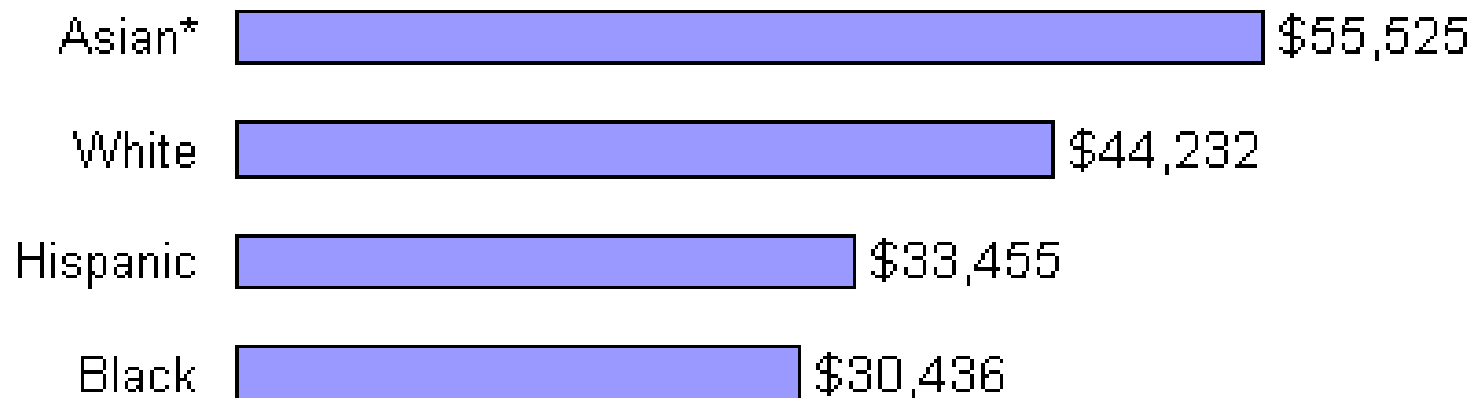


* Data not available on income before taxes and transfers.

- 18 casi, 36 variabili
- Ordinamento significativo
- Griglia eliminata perché superflua (anche l'asse delle ascisse, dato che i valori sono riportati)
- Confronto tra le barre blu (posizione degli USA)
- Confronto tra blu e arancio (non occorre conoscere il significato esatto dei trasferimenti per cogliere il senso)
- Gruppi di paesi (paesi nord-europei in testa, non europei in coda)
- E tuttavia una tabella ...

I grafici a barre sono inutili?

Median Household Income, by Race and Ethnicity: 1999



*Asian or Pacific Islander

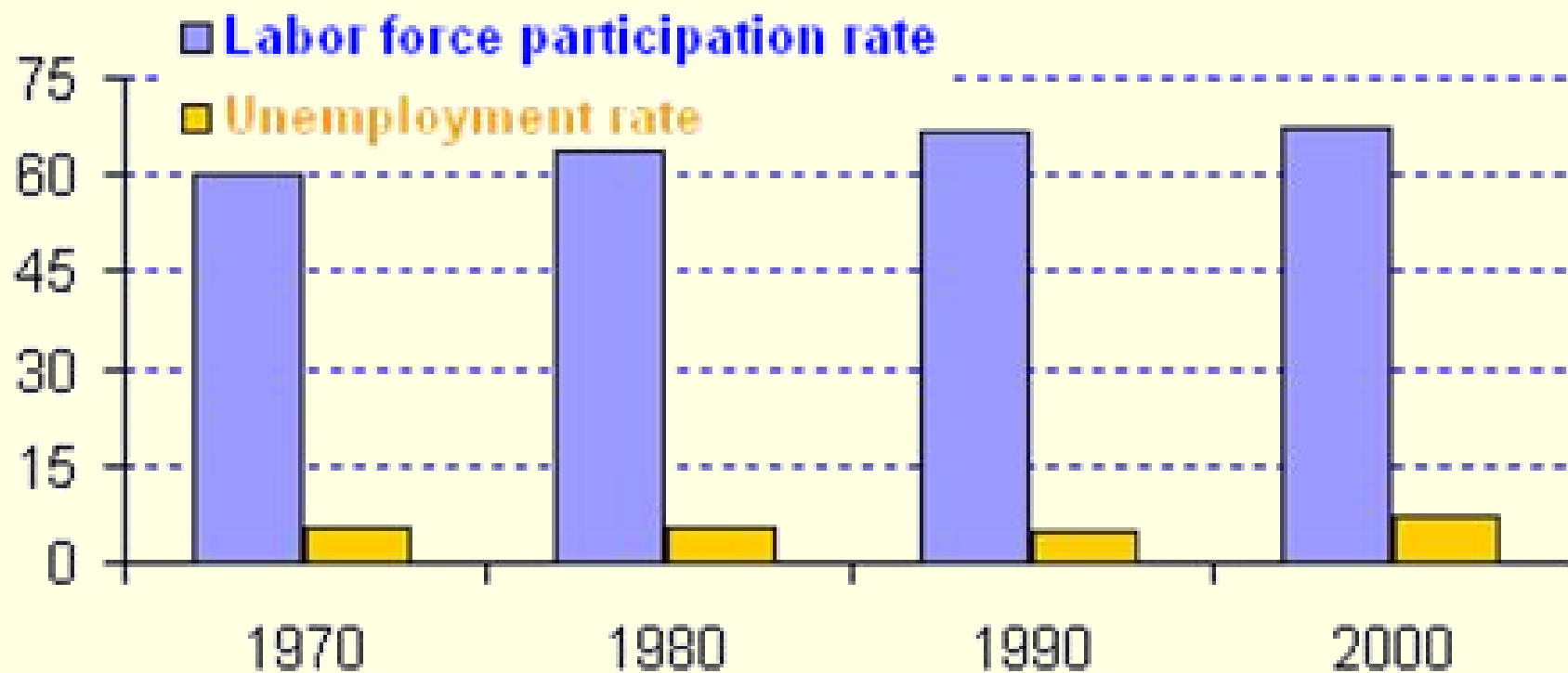
Bureau of Labor Statistics:

<http://www.bls.census.gov/cps/ads/2001/susenote.htm>

Standard e regole generali

- Minimizzare il rapporto inchiostro/dati (perché i dati sono in genere pochi)
- Mai grafici a barre tridimensionali
- Linee sottili (soprattutto nella griglia)
- Tra i quattro e i sei dati sull'asse delle ordinate (se no, girare il grafico)
- Se i dati sono pochi, la scala è inutile
- Non usare pittogrammi (distrazione)

Labor force participation and unemployment, 1970-00



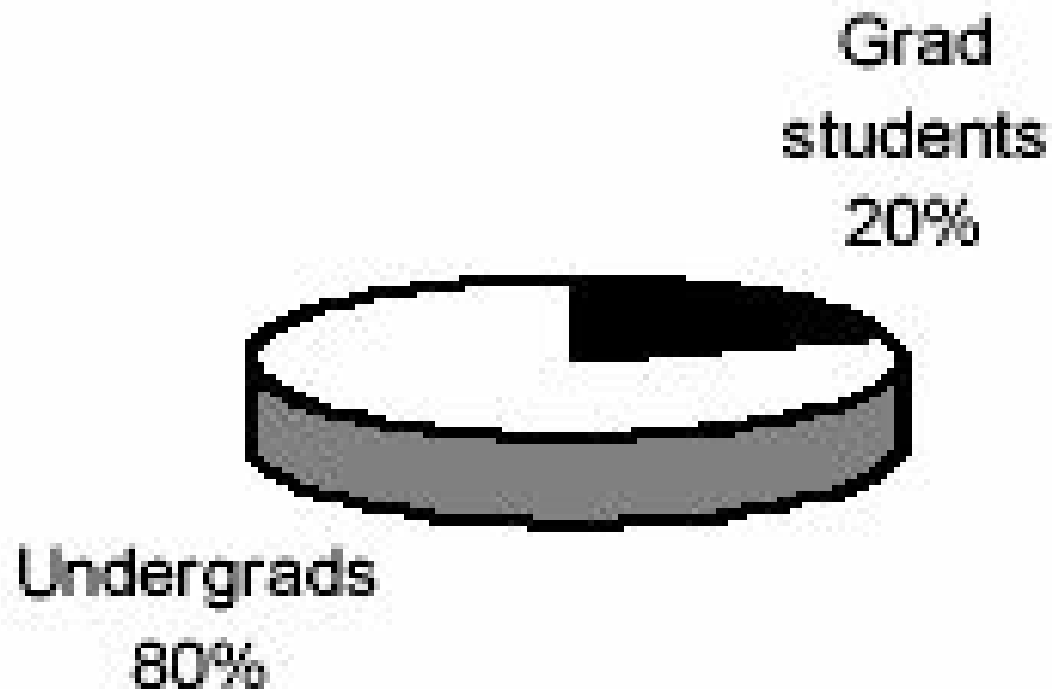
source: US Statistical Abstract CD-Rom 2001, table 567

Problemi di scala e di colore

- Quando ci sono più variabili, e le differenze di scala sono rilevanti, si verificano gli stessi problemi che emergono con le serie storiche
- Il colore è particolarmente importante:
 - Preferire i colori pastello
 - Per linee e grafici a dispersione, invece, meglio i colori primari

Le famigerate torte

2001 Enrollment by student level

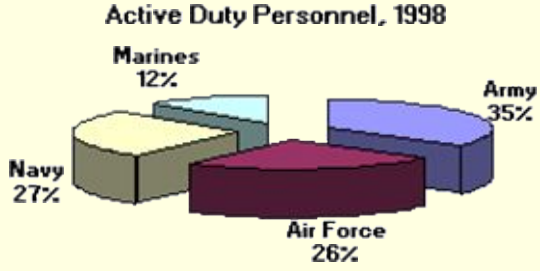
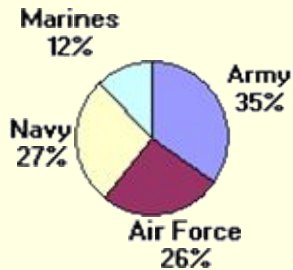
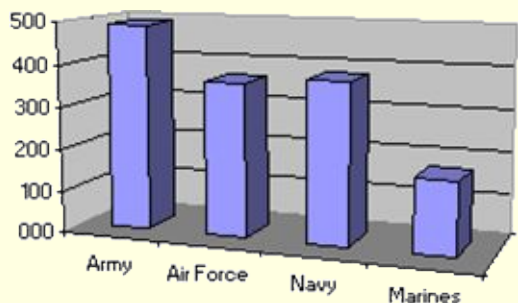
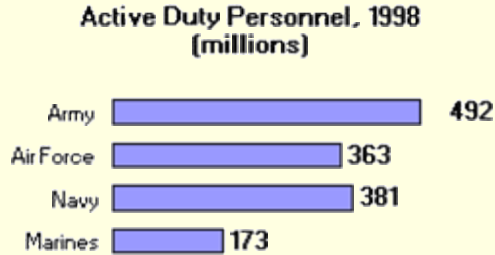


Perché è bene non usarle

- Il rapporto inchiostro/dati è ancora più sfavorevole che nei grafico a barre
- È più difficile valutare visivamente l'area di un settore circolare che l'area di un rettangolo (in cui una dimensione è fissa!)
- Mai le torte tridimensionali (lo spessore esagera percettivamente la dimensione delle "fette" piccole)

Conclusione:

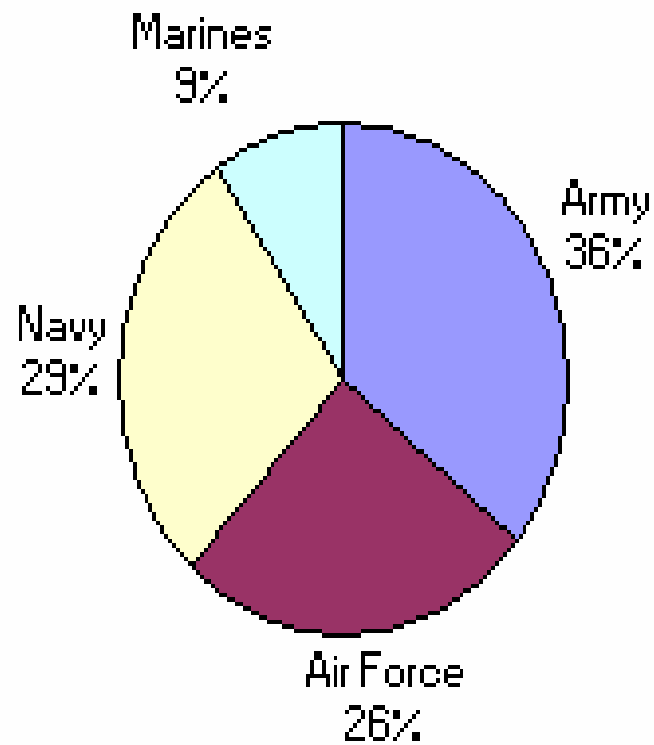
tutto quello che si può rappresentare su una torta si può rappresentare anche in un grafico a barre

<p style="text-align: center;">Active Duty Personnel, 1998</p>  <p style="text-align: center;"> Army 35% Air Force 26% Navy 27% Marines 12% </p>	<p style="text-align: center;">Active Duty Personnel, 1998</p>  <p style="text-align: center;"> Army 35% Air Force 26% Navy 27% Marines 12% </p>
<p>a. 3-D Pie chart, exploded</p>	<p>b. 2-D Pie</p>
<p style="text-align: center;">Active Duty Personnel, 1998 (millions)</p>  <p style="text-align: center;"> Army 492 Air Force 363 Navy 381 Marines 173 </p>	<p style="text-align: center;">Active Duty Personnel, 1998 (millions)</p>  <p style="text-align: center;"> Army 492 Air Force 363 Navy 381 Marines 173 </p>
<p>c. 3-D Column Bar</p>	<p>d. Simple 2-D Bar</p>

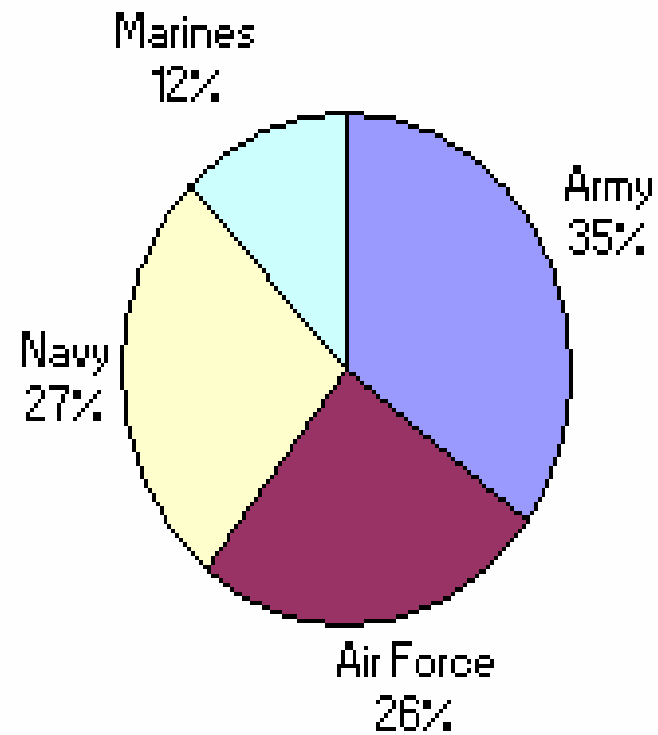
Discussione

- Notate quanto meno inchiostro il grafico a barre bidimensionale utilizza rispetto a quello tridimensionale
- L'eliminazione della scala delle ordinate e la sua sostituzione con le etichette dei dati riduce i numeri da 6 a 4 e fa guadagnare in precisione
- Come sono ordinati i dati?
- Meglio una tabella?

Due torte son peggio di una!



1990



1998