Visualisation d'information

Jean-Yves Antoine

http://www.info.univ-tours.fr/~antoine/

Interaction Homme-Machine

Chapitre 1.1 — Types de graphiques et type de données

… Ou le camembert n'est pas fait pour tout

INTRODUCTION - Objectifs

1.1.1. Notions

1.1.1.1. Types d’information
1.1.1.2. Graphiques pour distributions statistiques
1.1.1.3. Graphiques pour séries continues ou temporelles
1.1.1.4. Graphiques pour données hiérarchiques ou en réseau
1.1.1.5. Graphiques pour données spatialisées

1.1.2. Pratiques

1.1.2.1. Choix d’une rechnique de visualisation adaptée à un type d’information
1.1.2.3. Recommandations ergonomiques générales : affichage de données
1.1.2.2. Recommandations ergonomiques propres à chaque type de données

TYPES D'INFORMATION & VISUALISATION

- **Visualisation** = multiples choix de graphiques
- **Visualisation réussie** = complémentarité entre objectifs de communication, type de données et … type de visualisation retenu

Types de données

- **Nominale** — Nom de catégories, caractérisation qualitative : non ordonnée

  Exemples — géonyme ou toponyme
  nom de maladie
  opinion subjective {*attirant*, *utile*, *cher, …*}

- **Ordinales** — Non quantifiée mais ordonnée

  Exemples : — {Printemps, Eté, Automne, Hiver}
  {glacial, froid, tiède, chaud, bouillant}

- **Quantifiées** — Valeurs discrètes (entiers), continues (réels), intervalles

- **Temps** — Vu comme une donnée ordinaire (ordinale ou quantifiée) ou comme une dimension spécifique : représentation adaptée

TYPES D'INFORMATION

Typologie d'information

[Shneiderman 1996, Andrews 2002]

Information complexe : plusieurs dimensions correspondant à des types de données potentiellement différentes (nominale, ordinale, quantifiée)

- Unidimensionnelle — *Ne considère pas la dimension qui peut servir à ordonner les données*
- Bidimensionnelle
- Tridimensionnelle — *Relations entre dimensions ?*
- Multidimensionnelle — *Visualisation simultanée impossible*
- Temporelle — *Animation ou pas*
- Spatiale et spatio-temporelle — *Cours de SIG*
- Relationnelle — *Hiérarchie, graphe, réseau*

INFORMATION TYPES & VISUALISATION

Typology of information: Standford's vizualisation zoo

[Heer and al. 2010]

Statistical Distributions **Discrete series**	Histograms (bar charts), pie-charts, scatter plots, box & whisker plot, q-q,
Time series **Continuous series**	Index chart, stacked-data graph, horizon graph, small-multiple graph
Hierarchies	Node-link diagrams, tree graphs, tree maps
Networks	Force directed graphs, arc diagrams, Matrix views
Maps	Maps, cartograms, chloropleth

DISPLAYS FOR STATISTICAL DISTRIBUTIONS

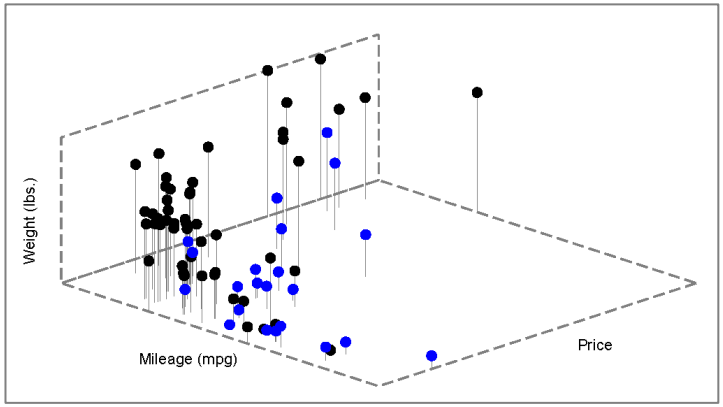
Graphs (graphiques)

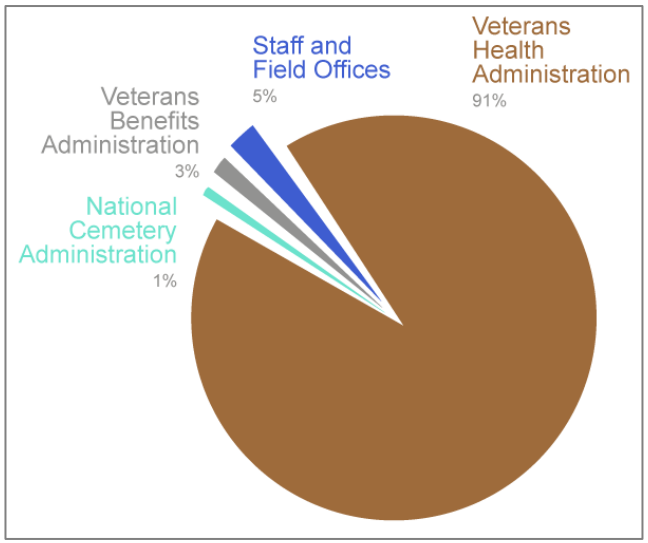
[Kosslyn, 1989 ; Heer and al. 2010]

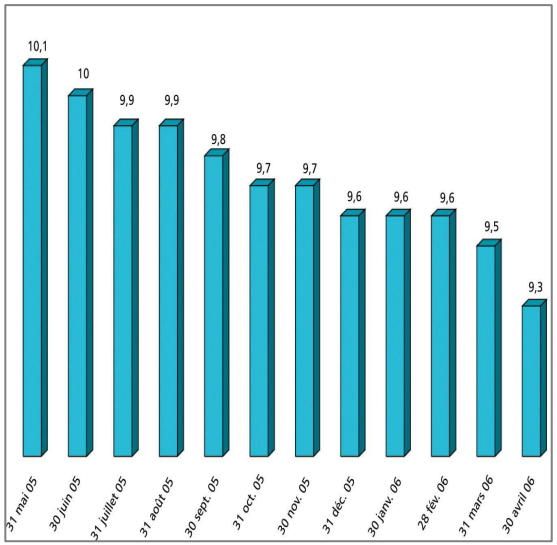
- At least two dimensions (1+1)
- More than 3D : additional visual variable

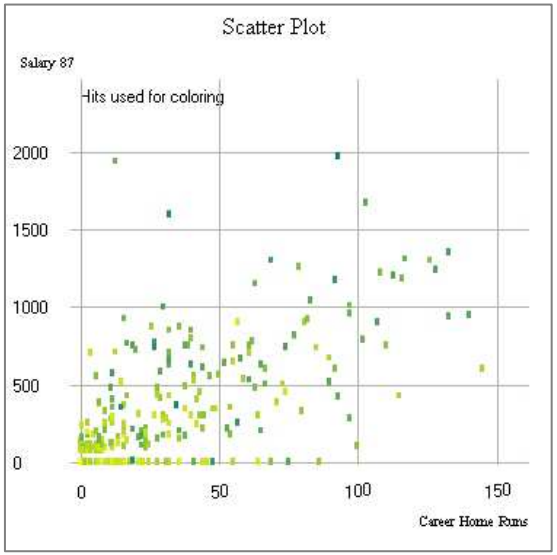
Graphs for statistical analysis

- Scatter-plot (*diagramme de dispersion*)
- Bar-chart (*histogramme*)
- Pie-graphs (*camemberts*)







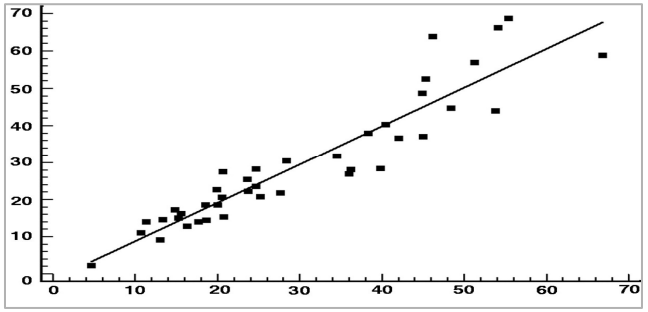


DISPLAYS FOR STATISTICAL DISTRIBUTIONS

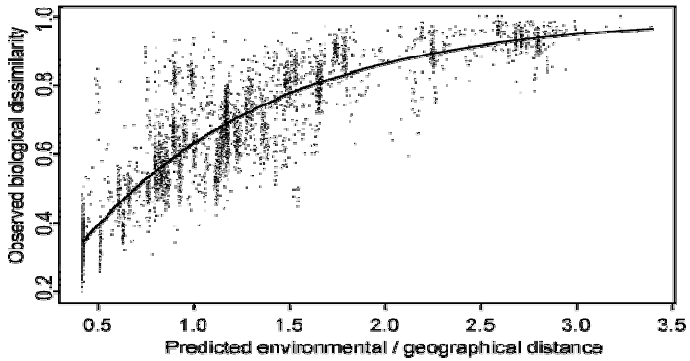
Scatter plot

- **Aim** – visual detection of correlations between **two continuous variables** : analysis of the shape of the plots distributions.

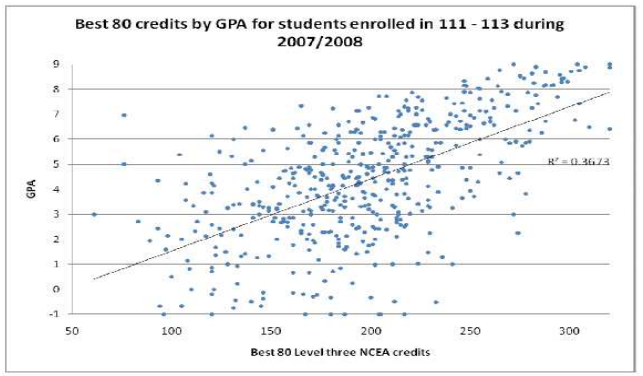
↳**High correlation** – linear mapping between axes ⇨ regression line



↳ **Non linear correlation** – curvature of the pattern of the plots ⇨ consider logarithmic scales



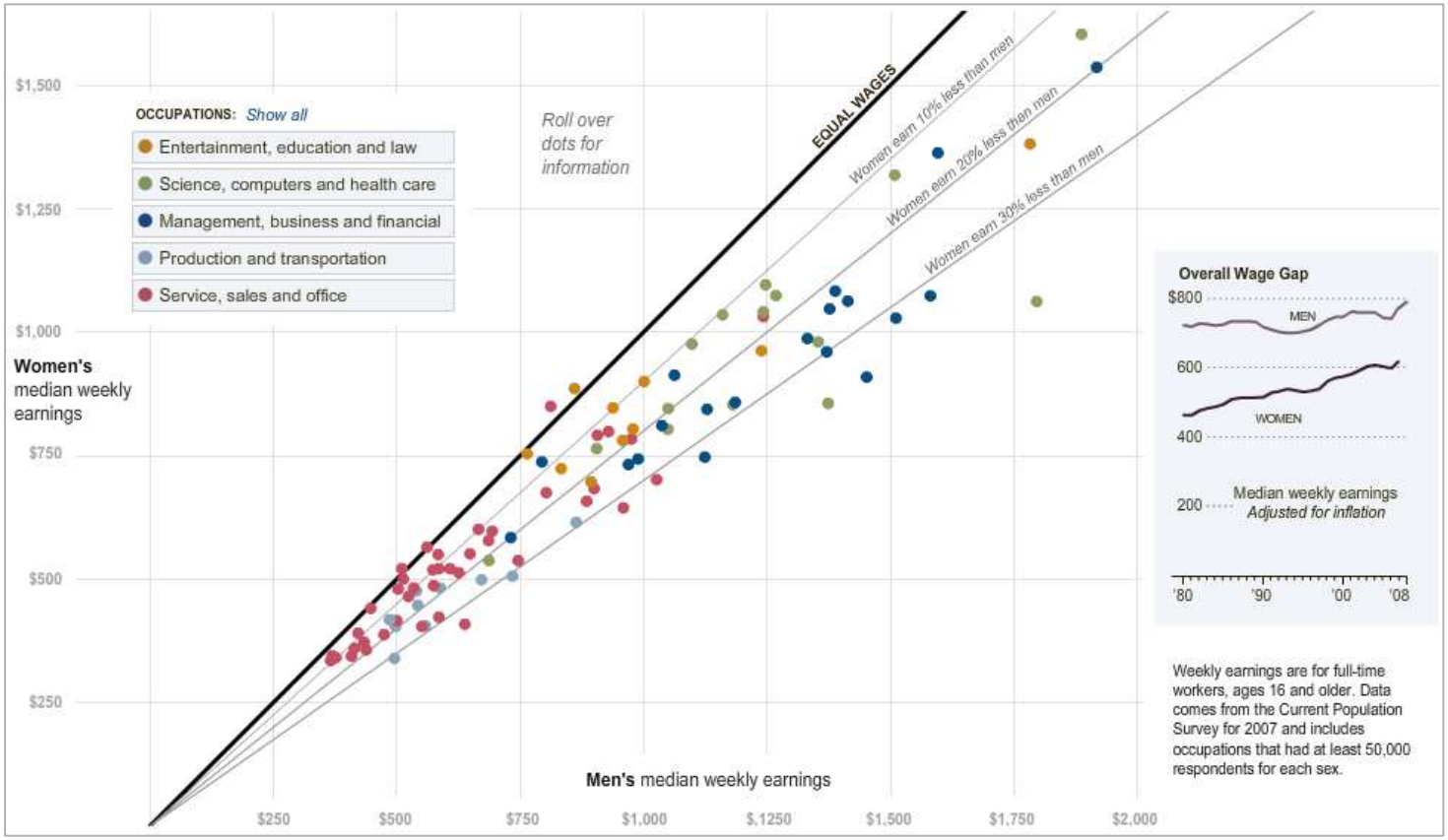
↳ **Low correlation** – spherical, rectangular or irregular distribution



DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Scatter plot

- **Aim (2)** – visual detection of particular data, or specific distributions ⇨ isolated dotted-points
- **Interpretation** – consider color to highlight particular data

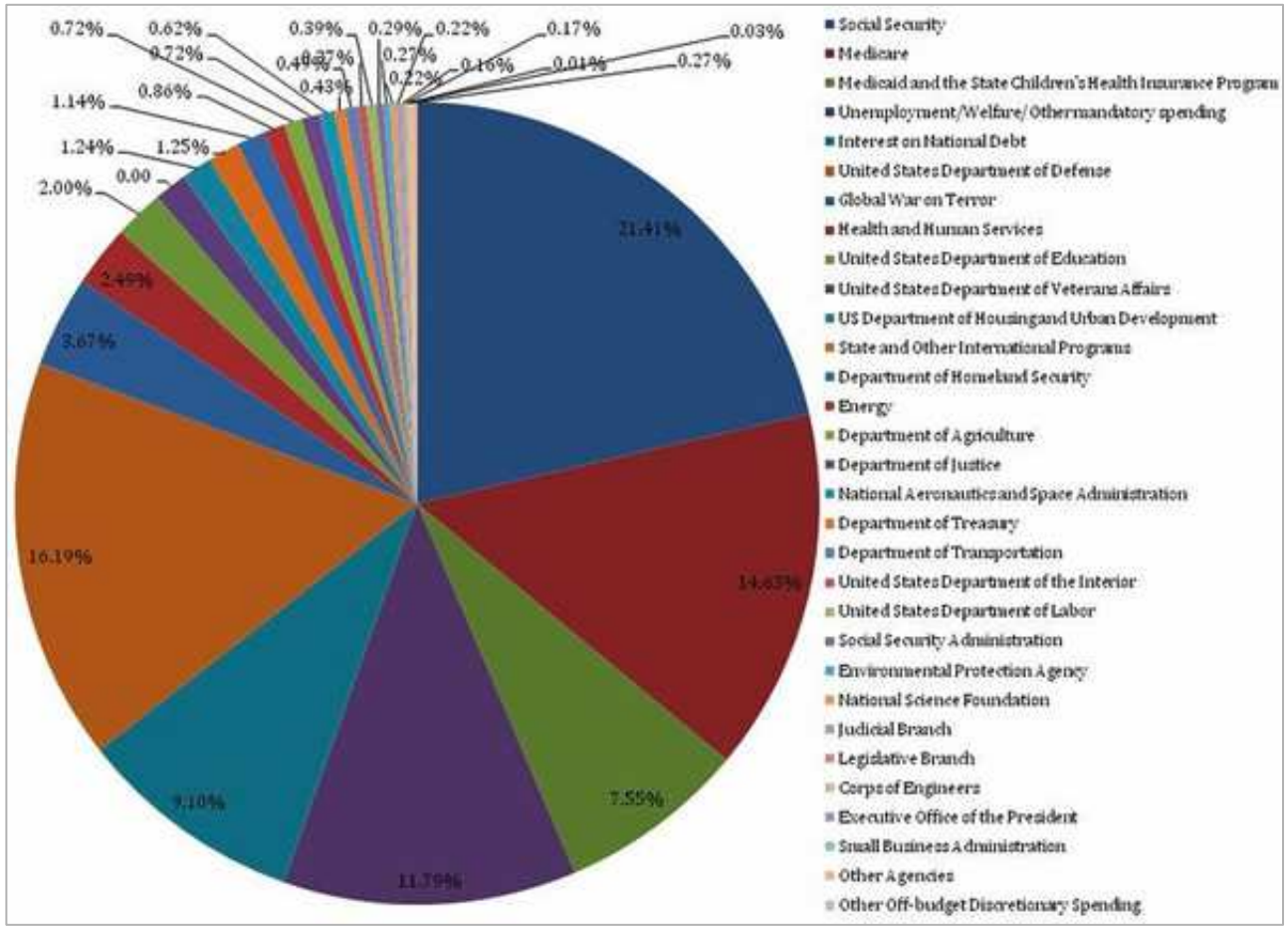


[Hannah Fairfield and Graham Roberts, *New York Times*, 2010]

DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts

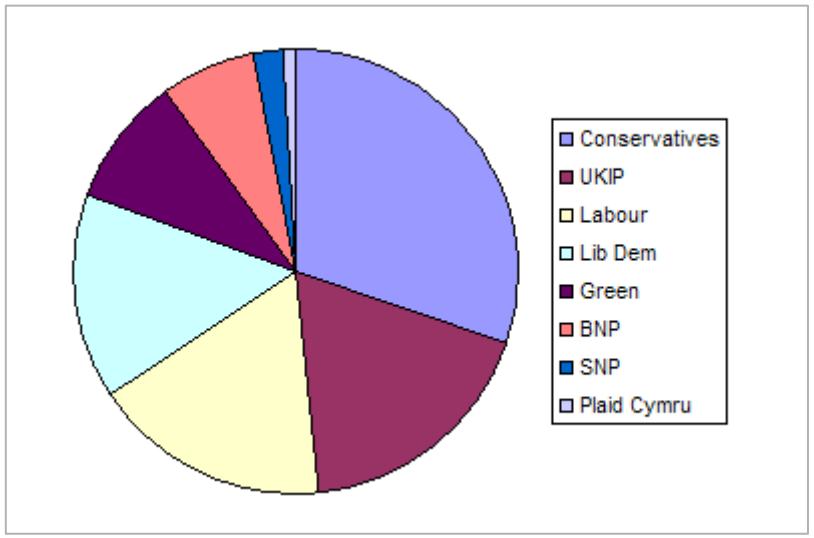
- **Aim** – visual detection of **differences on proportion** among several classes
- **Limitation** – easy interpretation only with a few data

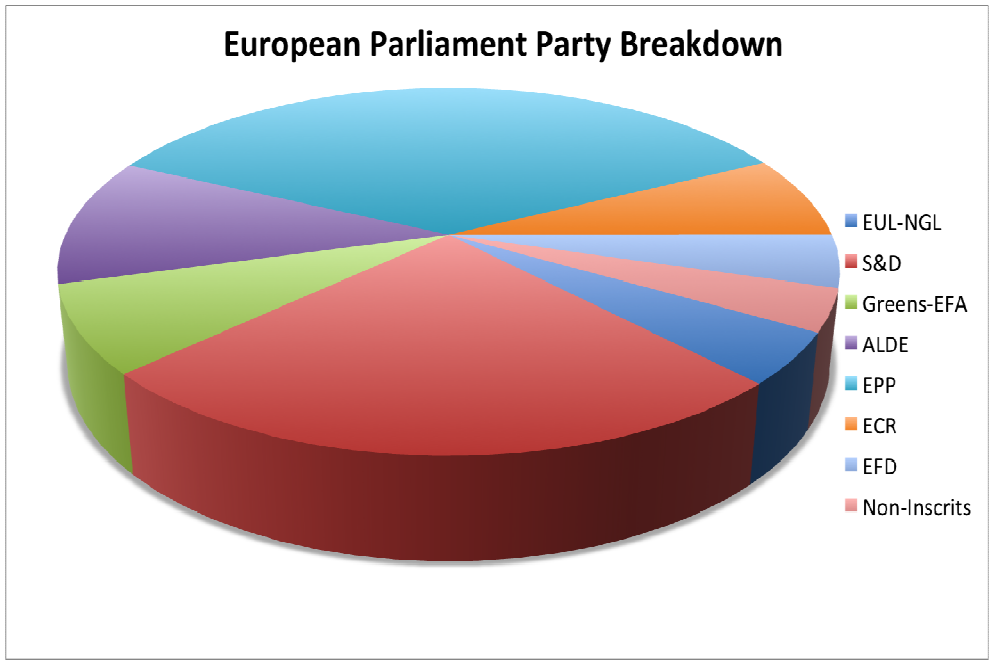


DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts

- **3D pie charts** are even more dangerous than 2D pie-charts

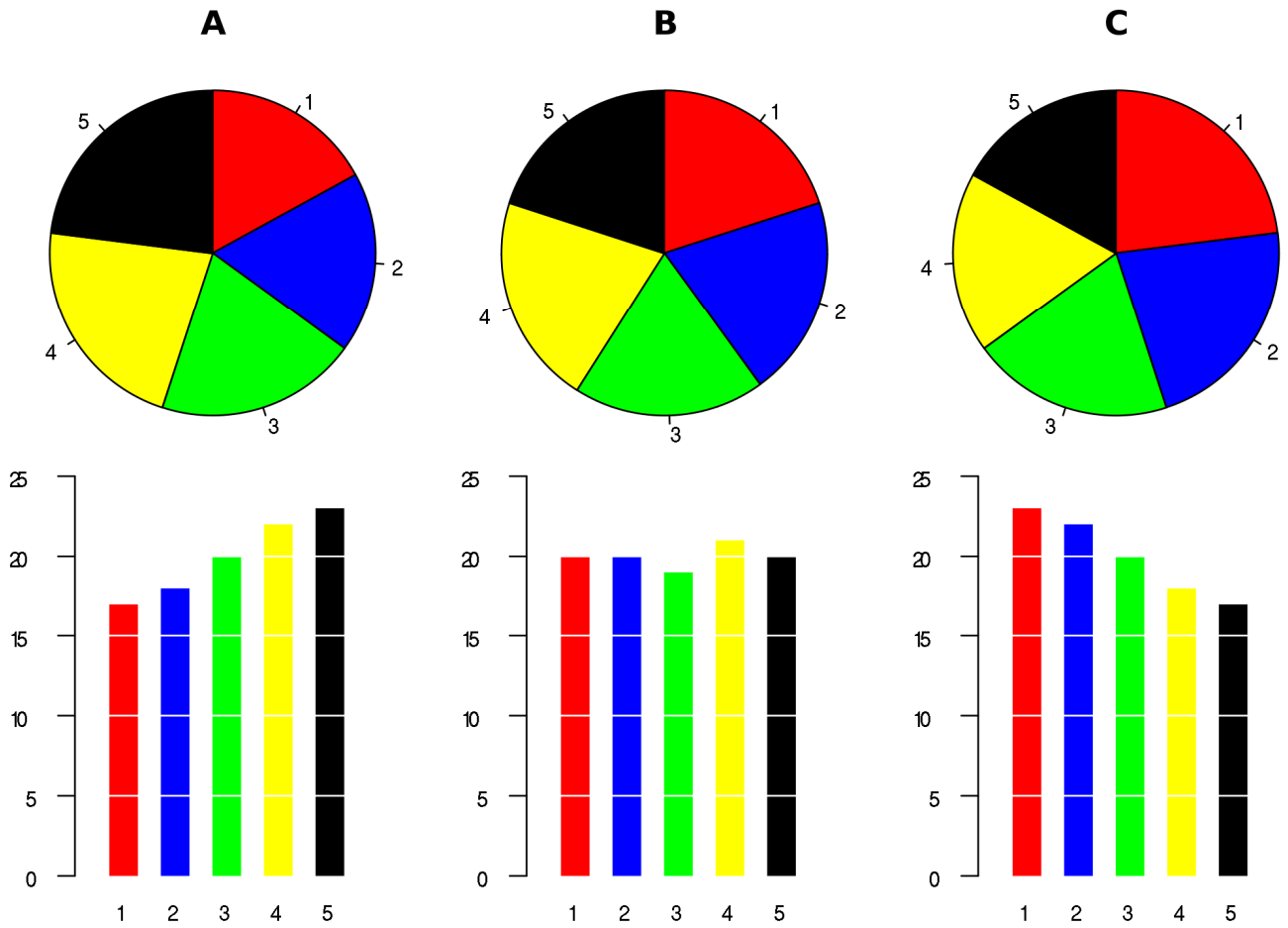




DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts

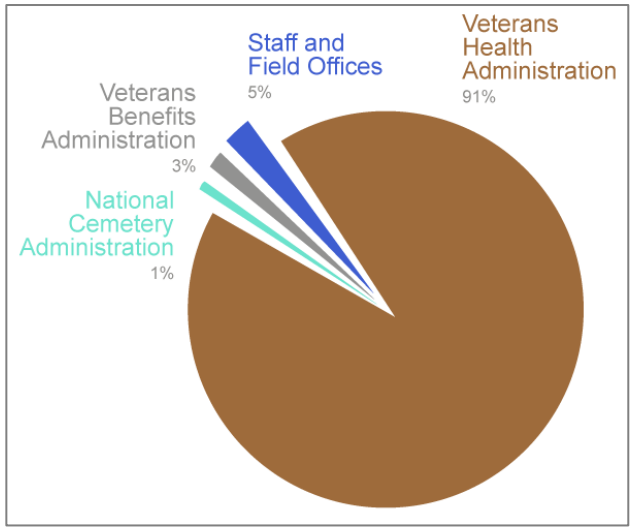
- **Use with caution** – close numeric values are not easily comparable

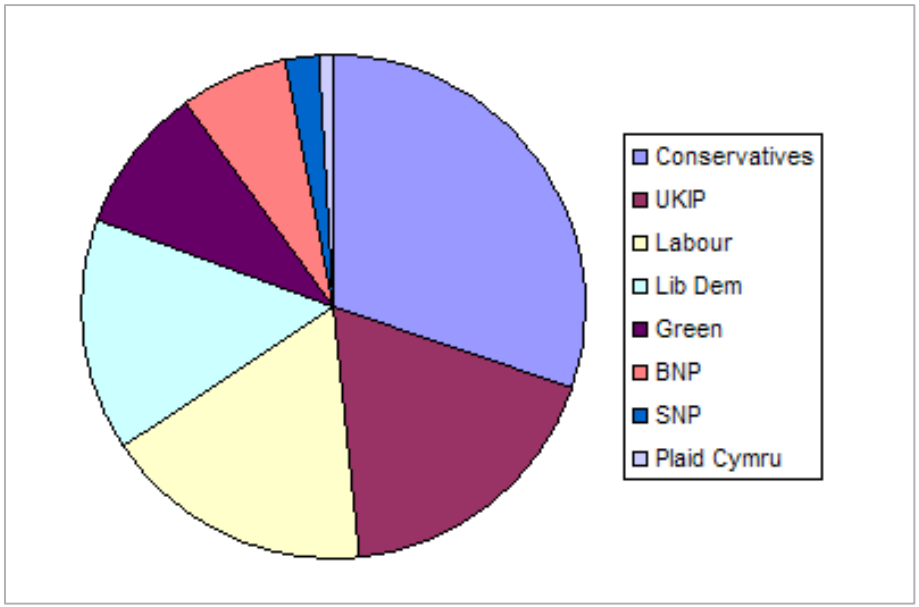


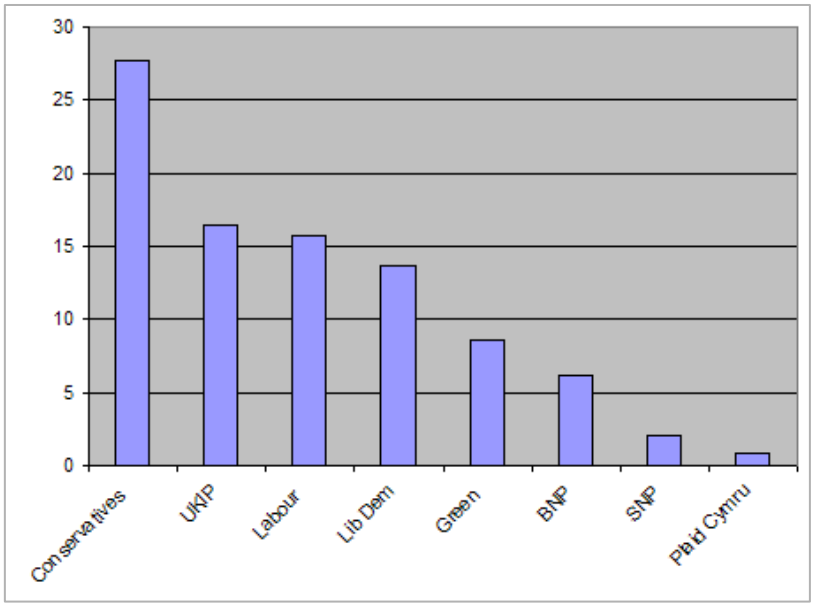
DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts : guidelines

- Use **only if you need to show how are distributed 100%** of a data.
- Better works with only **2 to 4 classes**
- Always consider **bar charts** as an alternative



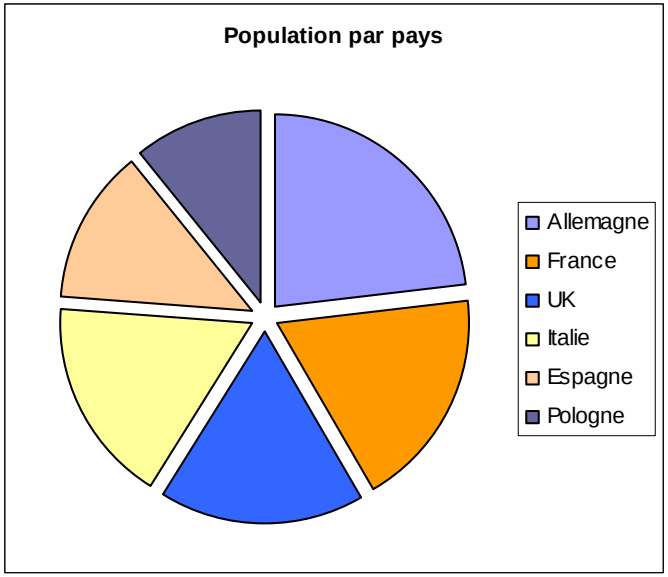


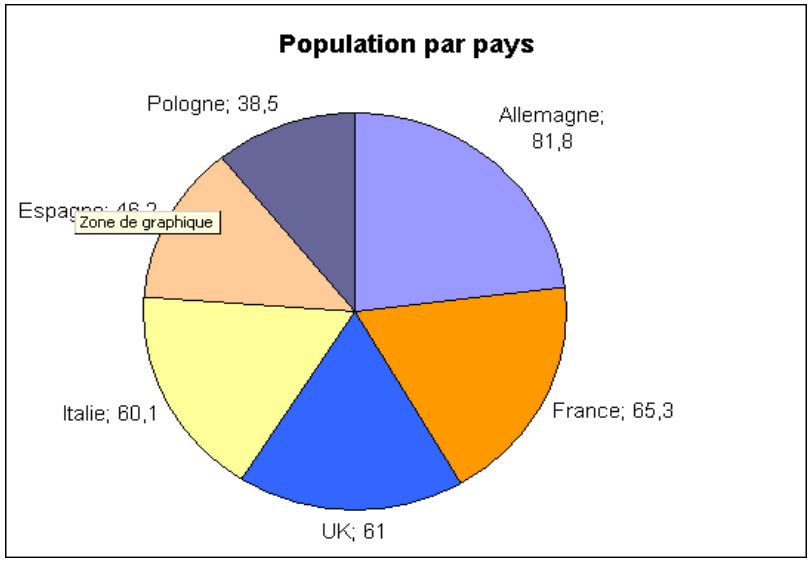


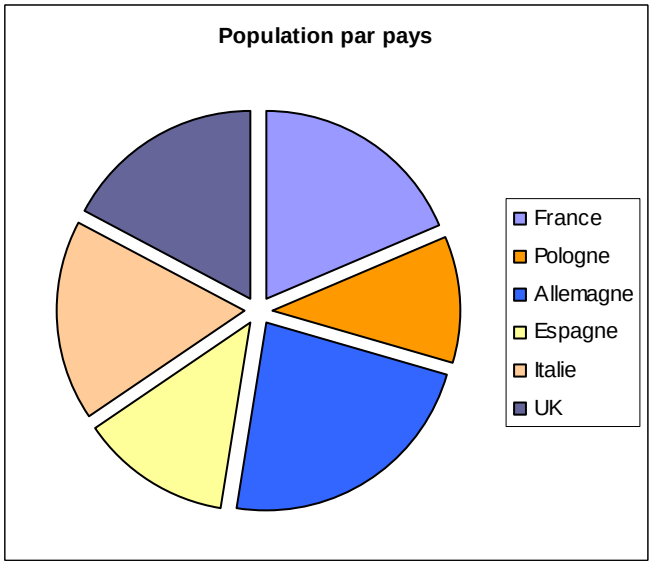
DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts : guidelines

- **No axe** : **add labels** to show the value of each class, if it is important for the interpretation.
- **Labels near the slices** rather than a separate legend (particularly with more than 3 classes)
- **Order segment by value** to ease comparisons
- Careful attention to **color coding** (chapter 1.2)



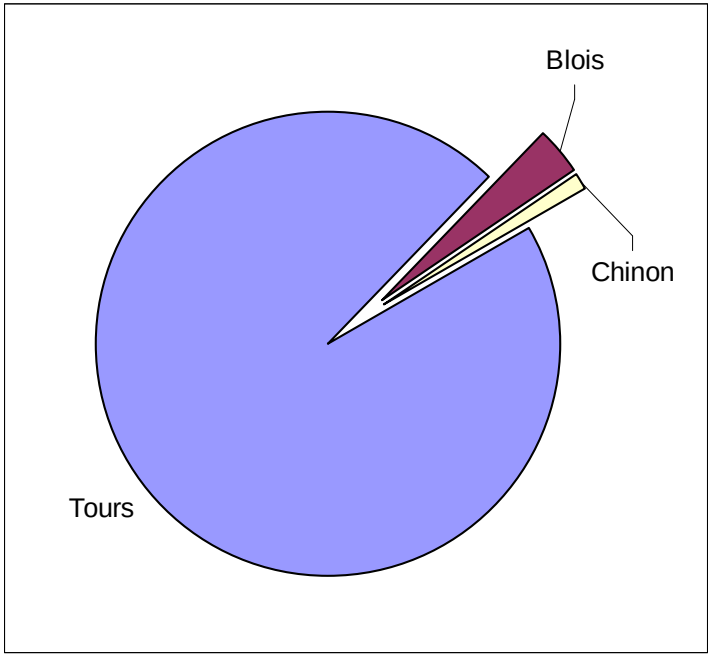


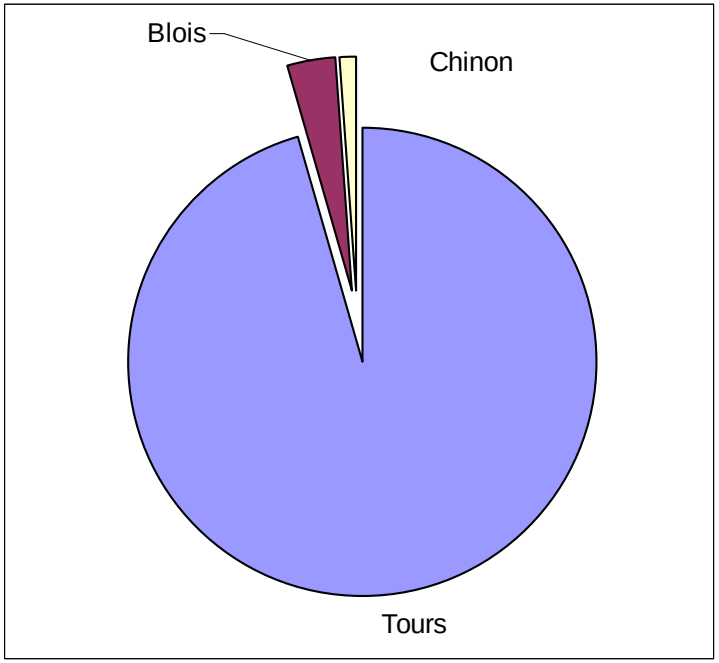


DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts : guidelines

- **Avoid** 12-o'clock position for the slices radii [Hollands, 2003]

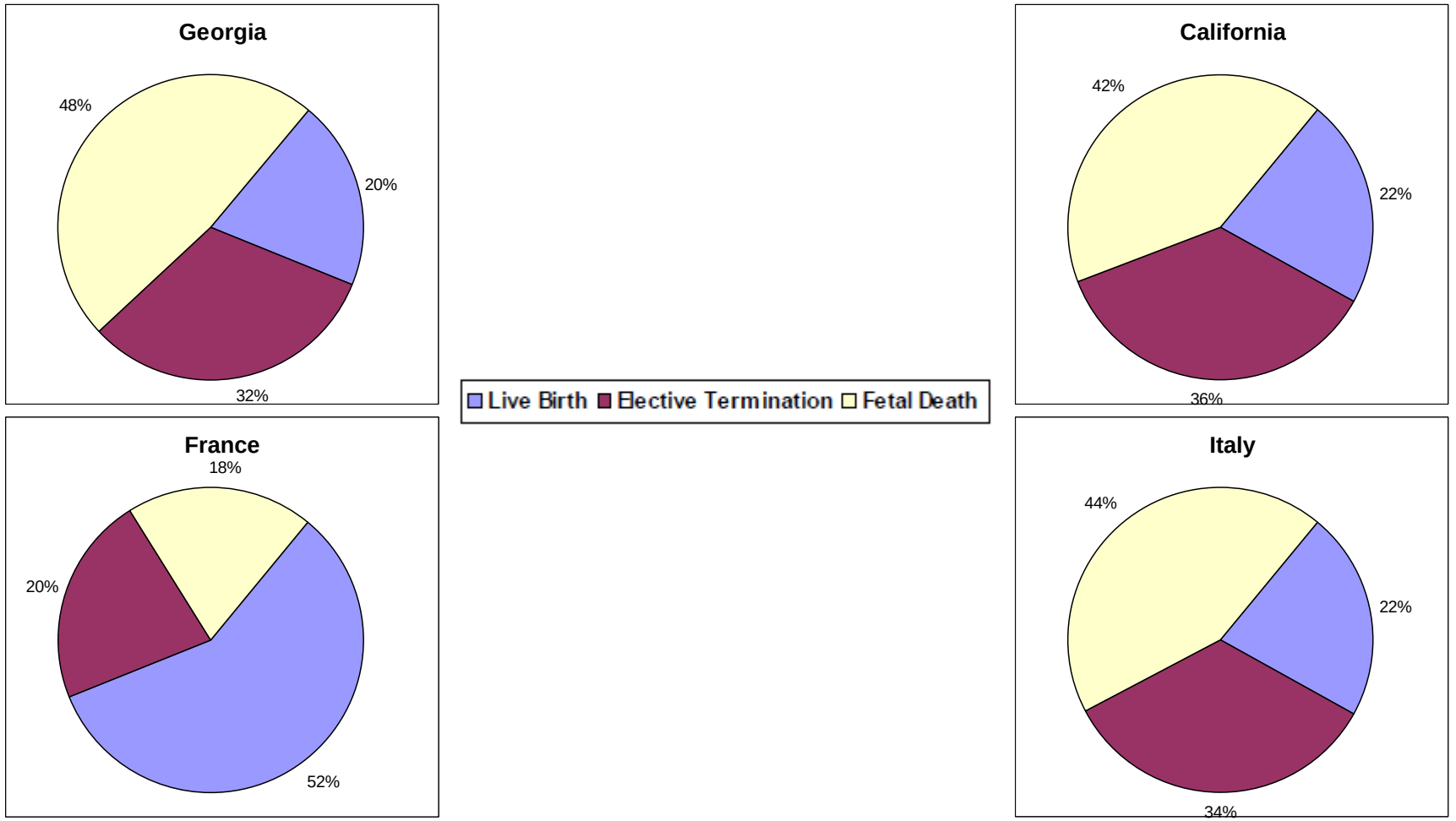




DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Pie charts and multiple data sets

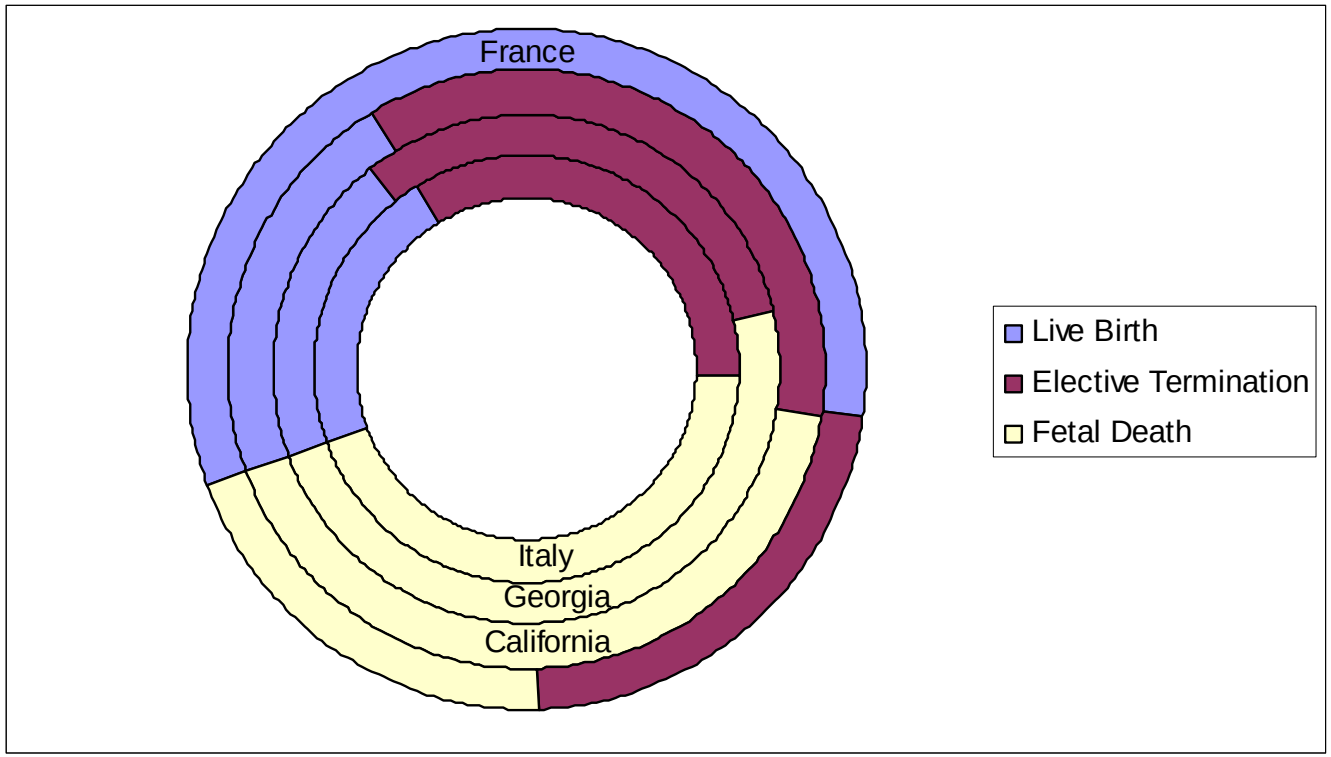
- Comparison of multiple pie charts is very difficult even impossible…



DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Annular pie chart

- **Aim** – Visual comparison of proportions between multiple data sets
- **Limitation** – Works only with restricted numbers of data and when data distributions permit a clear comparison

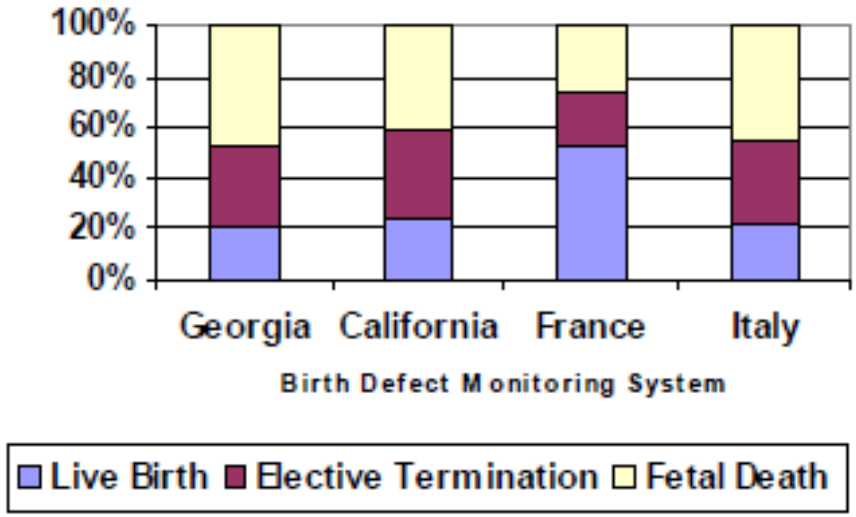


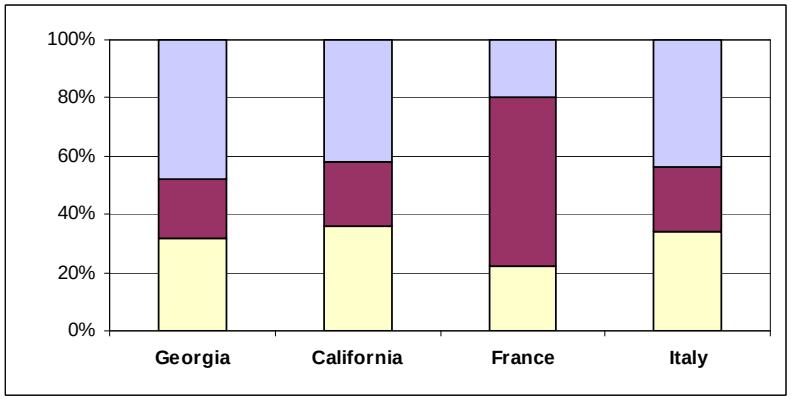
- **Guidelines** - careful attention to slices and rings ordering
  - guidelines for pie chart remain relevant

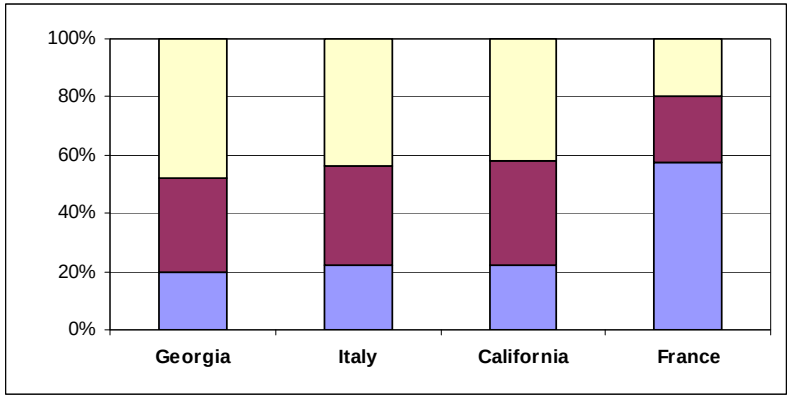
DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Stacked-bar charts

- **Aim** – Visual **comparison of proportions between multiple data sets**.
- **Guidelines** – Consider different stacks and sets ordering to ease interpretation



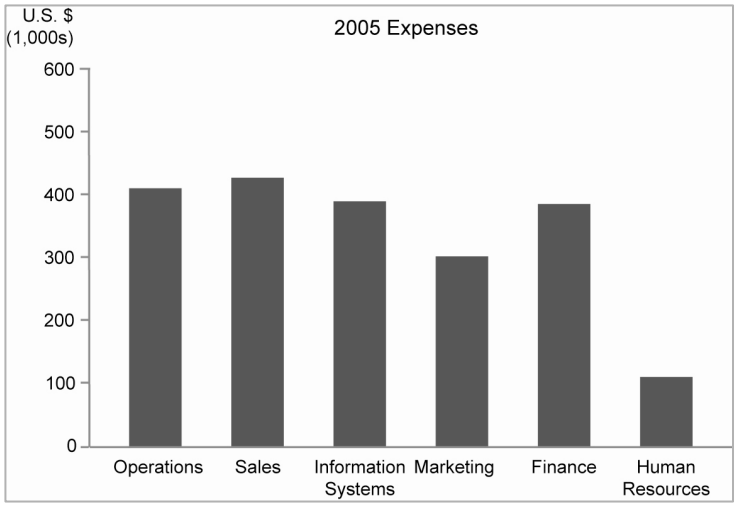


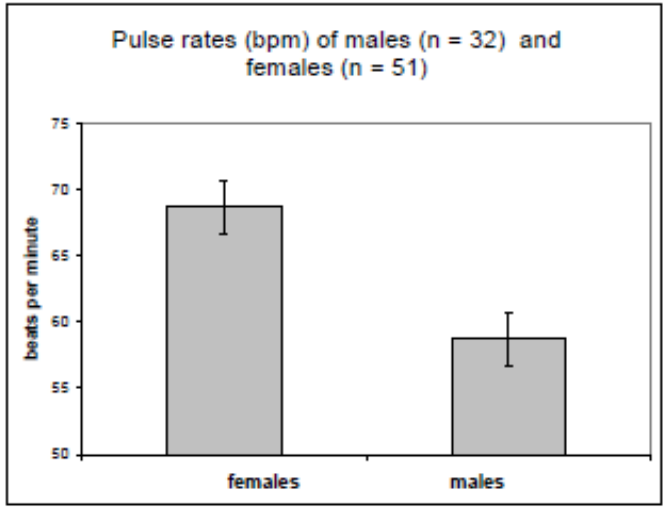


DISPLAYS FOR STATISTICAL DISTRIBUTIONS

Bar-charts (*histogrammes*)

- **Aim** – Visual detection **of differences between individual values**.
- **Data** – **Two variables ore more**, one of them must be discrete (classes)
- **Data : discrete series** – not restricted to statistical distributions : any kind of numerical values related to discrete classes (nominal, ordinal).
- **Interest** - **Values clearly separated** : usable with **large sets of data**
  - **Vertical bars** : differences of values distinguishable quickly
  - **Axes** : numerical values rather accessible without label.

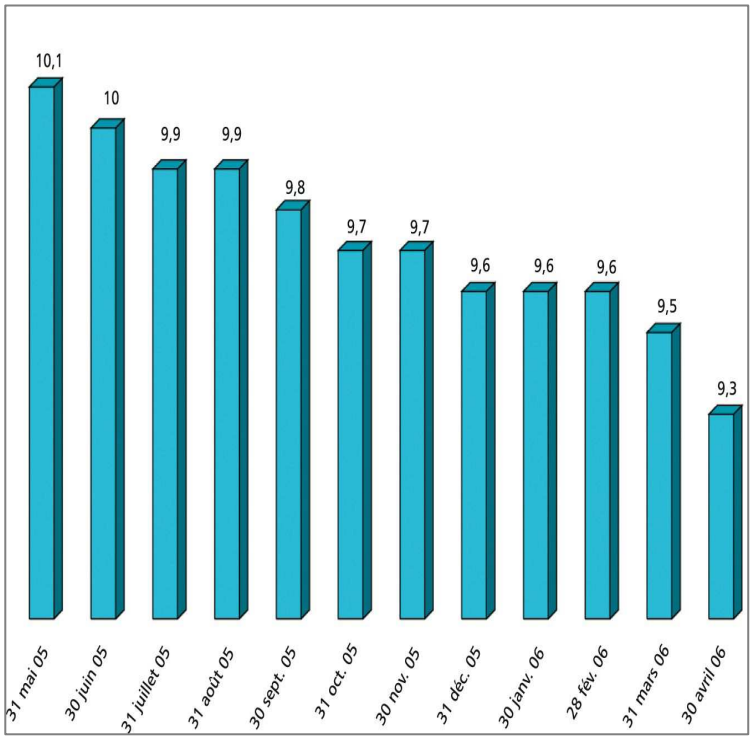


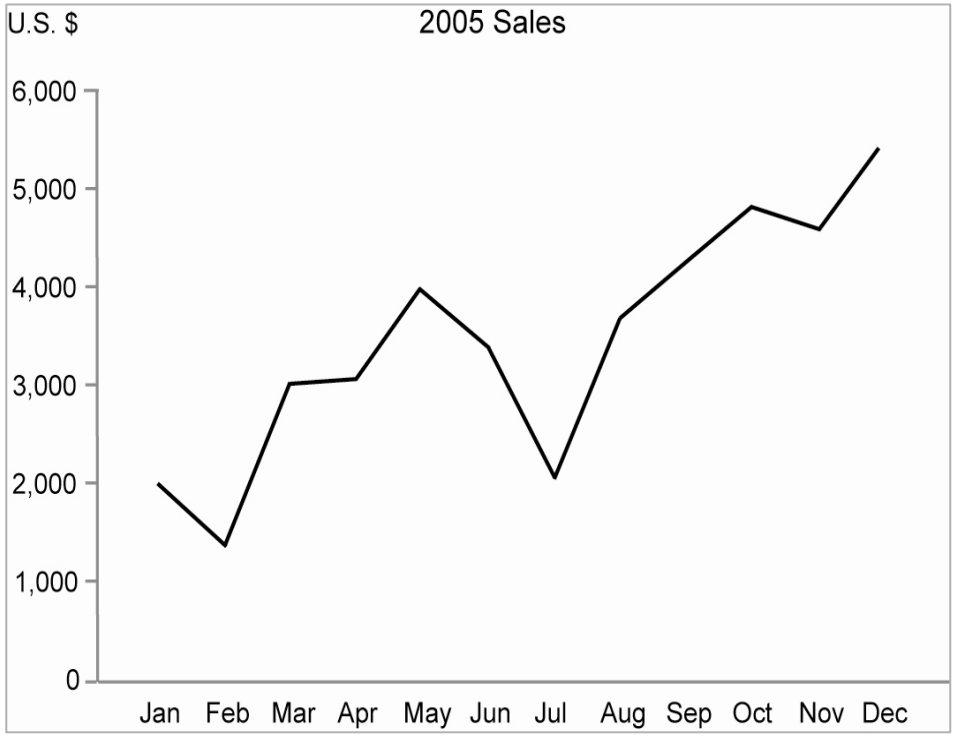


DISPLAYS FOR TIME AND CONTINUOUS SERIES

Times and continuous series

- **Bat charts** – usually **not adapted** for representing time series
- **Index charts** (or **line-charts**) – **well adapted** to time or continuous series : successive values are connected with lines, what show better an overall evolution-shape or a continuous process

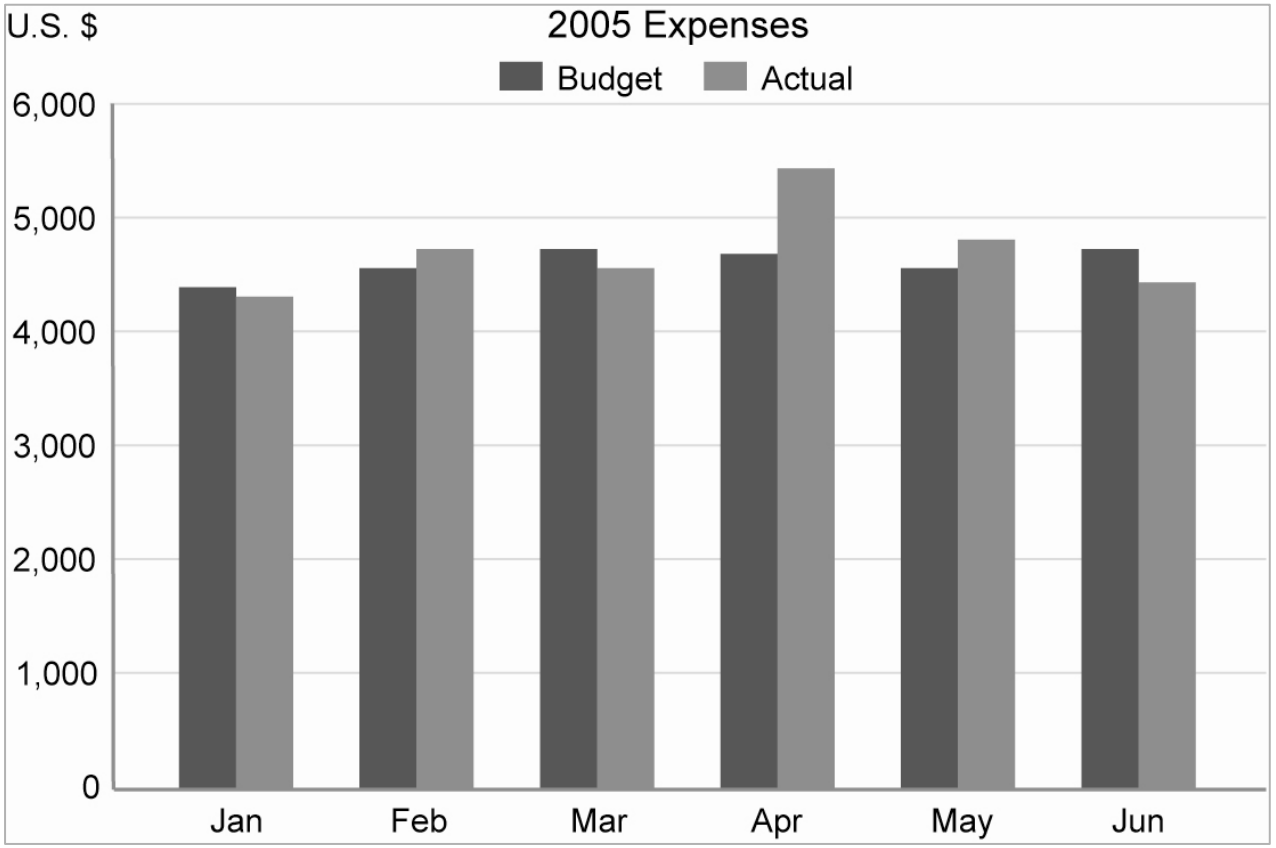




DISPLAYS FOR TIME AND CONTINUOUS SERIES

Bar-charts (*histogrammes*)

Case study : 3 variables : double comparison with a time serie

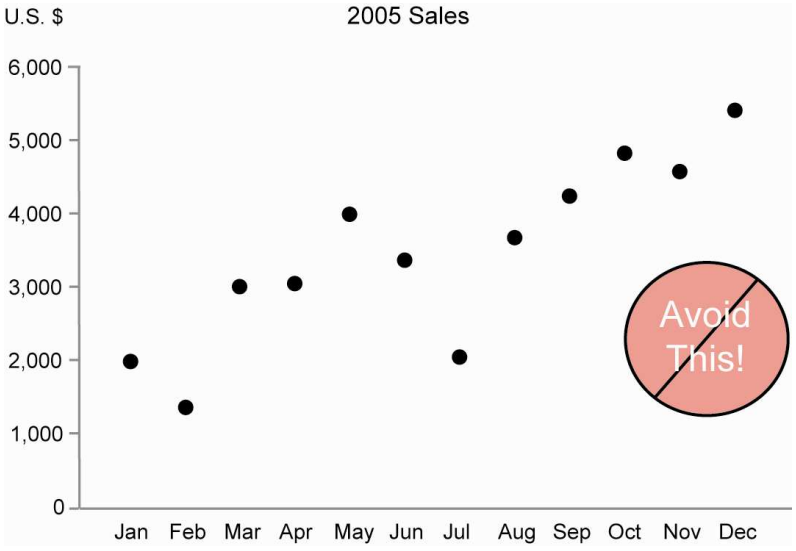


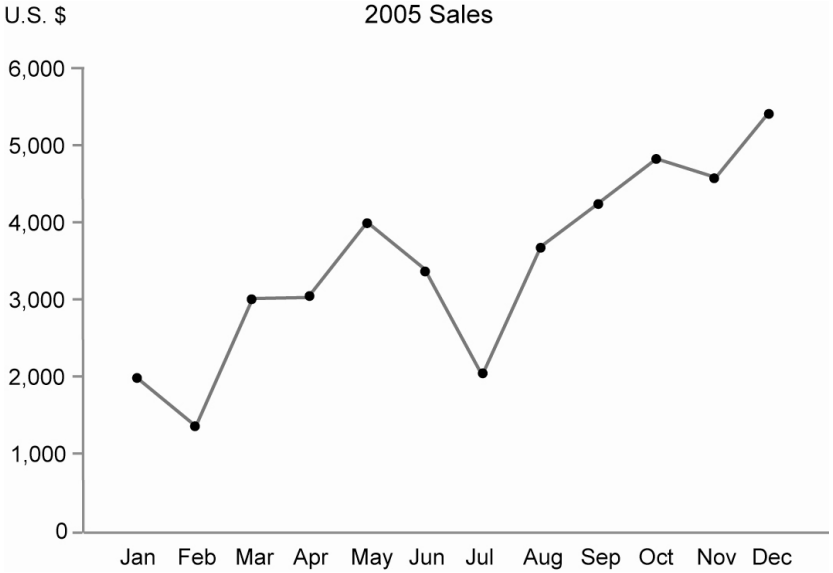
Main communicative aim : comparison between provisional and actual budget ⇨ bar-chart is adapted while we are considering a time serie.

DISPLAYS FOR TIME AND CONTINUOUS SERIES

Point graph

- Two dimensions – X and Y axis
- **Constraint** – Discrete values on X axis
- **Constraint** – **Not suitable for time series** : consider line-graphs, where the connections between values show better the temporal evolution.



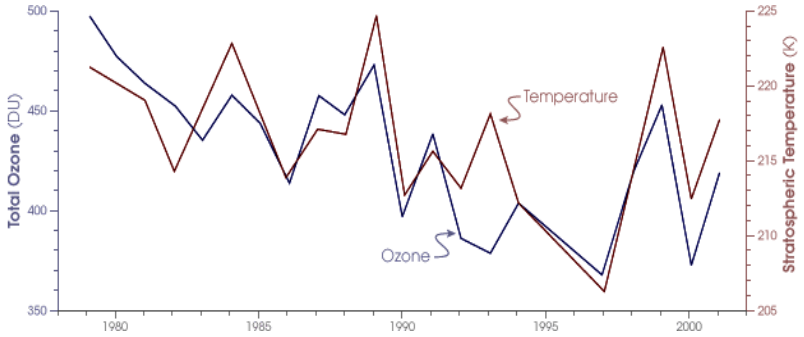


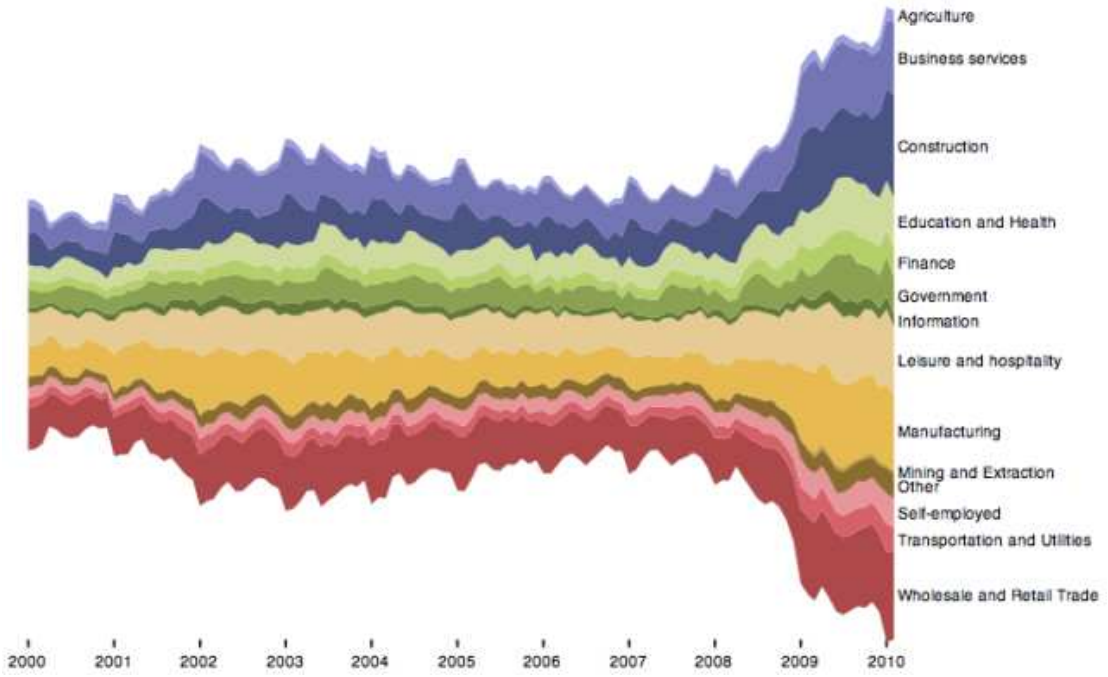
DISPLAYS FOR TIME SERIES

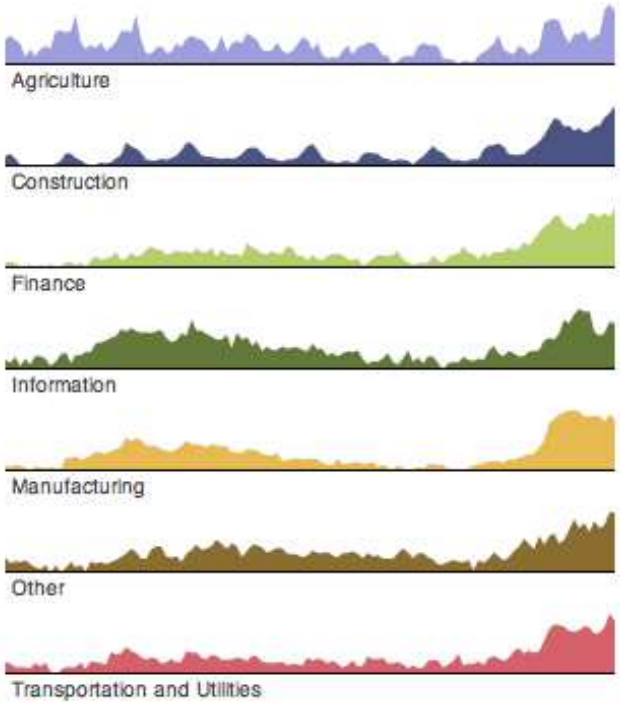
Graphs for times series

[Kosslyn, 1989 ; Heer and al. 2010]

- Layer-graphs or index-charts
- Stacked-graphs / steams graphs
- Small-multiple graphs



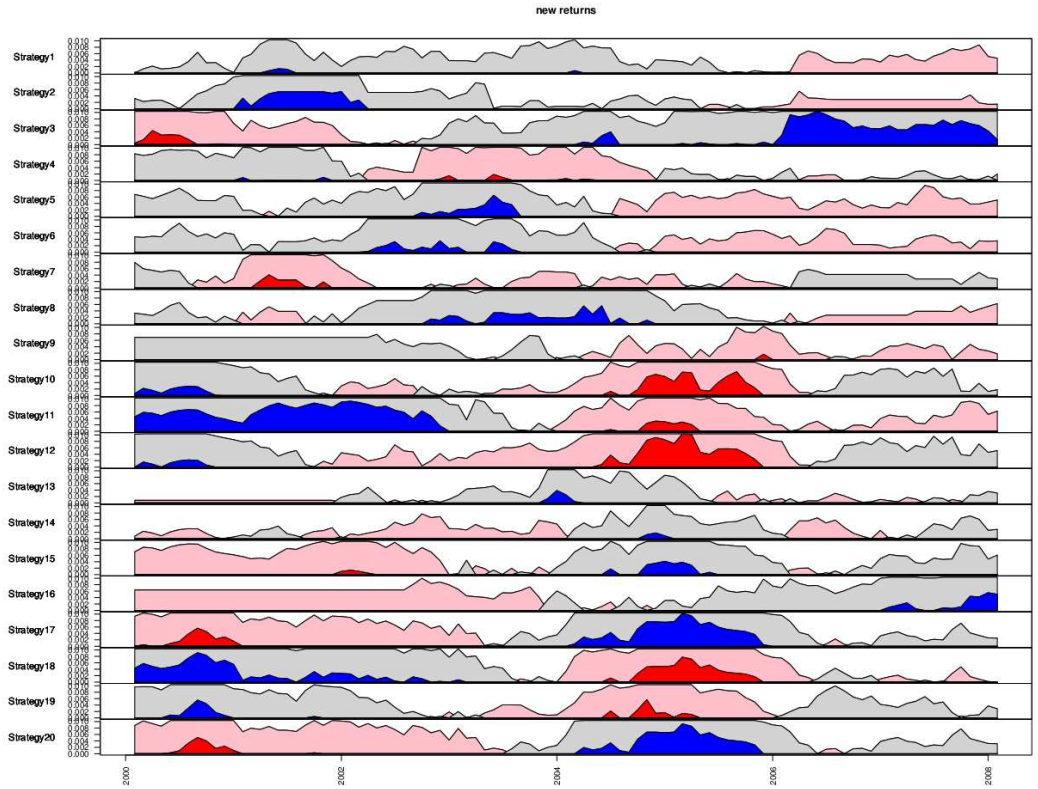




DISPLAYS FOR TIME SERIES

Graphs for times series

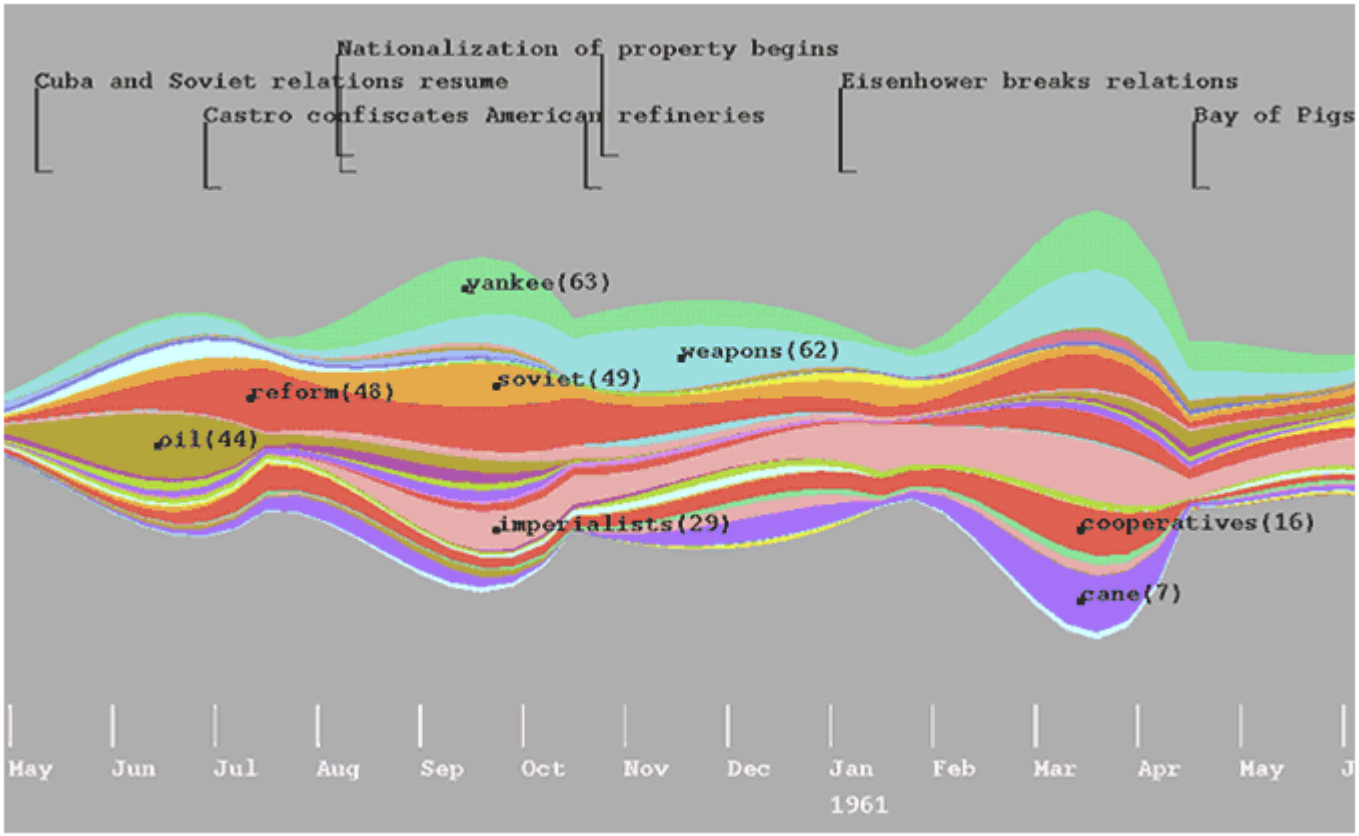
- Horizon graphs : densified small-multiple graphs



DISPLAYS FOR TIME SERIES

Layer-graphs or index charts without times series

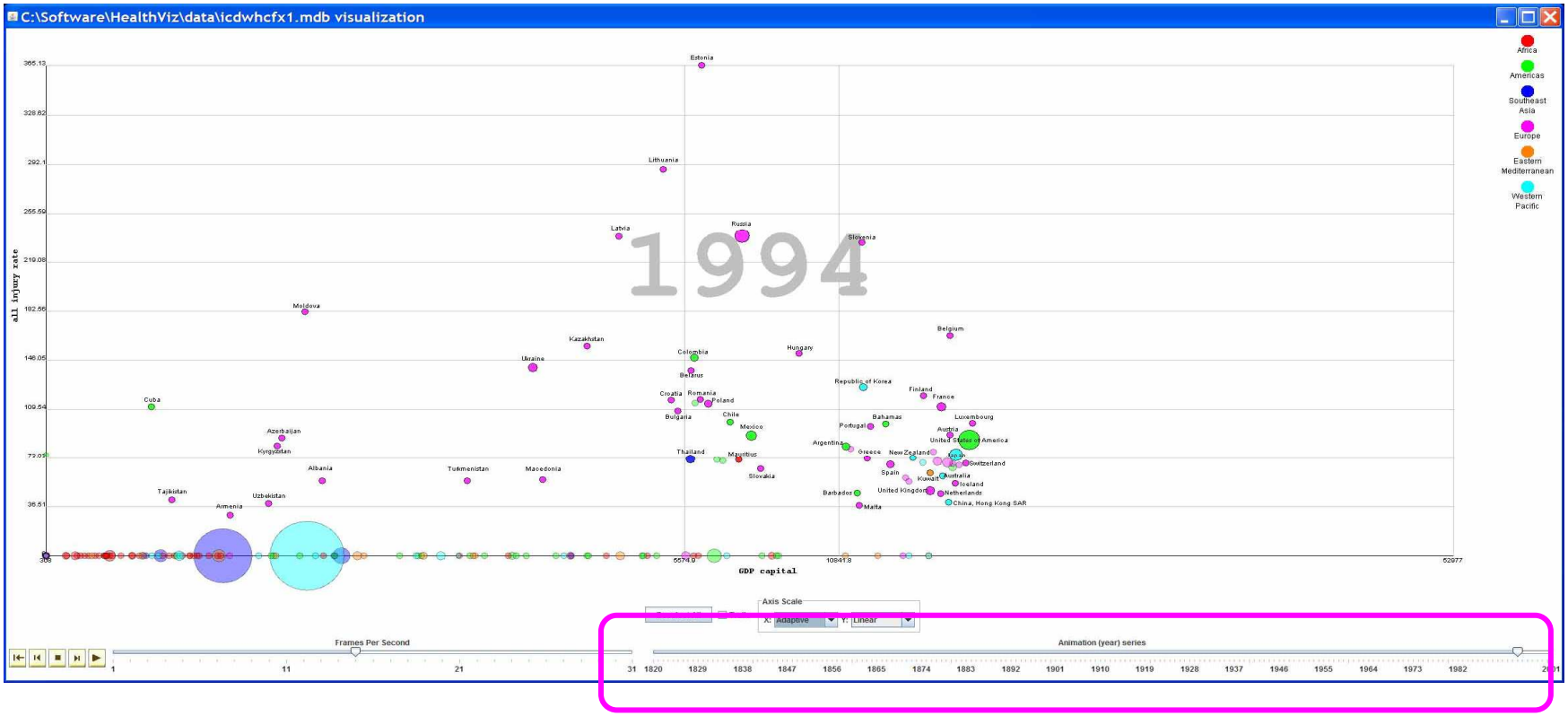
- **X-axis** : quantitative data or familiar ordering (example : chronological order)



DISPLAYS FOR TIME SERIES

Timeline

Consider animation for time series



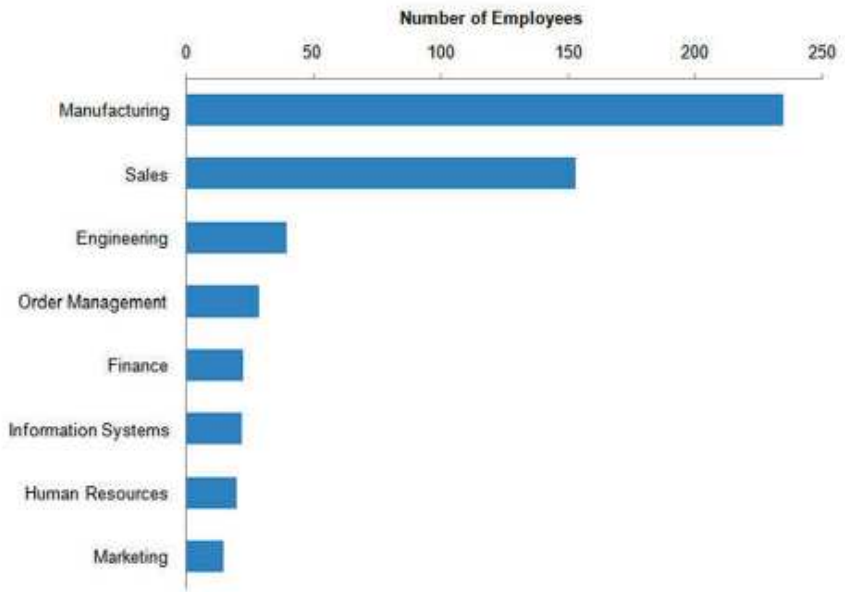
Animation : careful use !

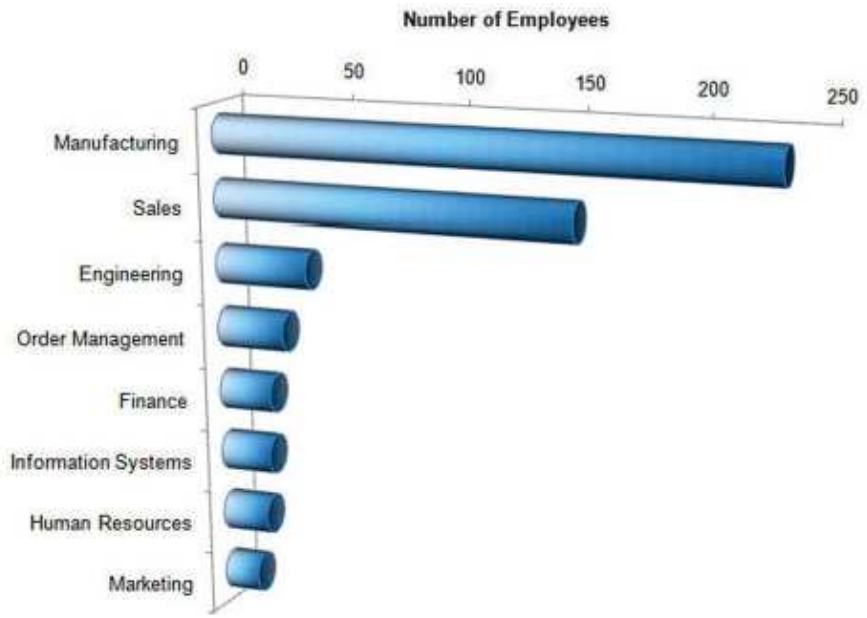
- **Congruence principle** – Natural correspondence between animation and information
- **Apprehension principle** – Animation slow and clear enough for understanding

AFFICHAGES DE DONNEES NUMERIQUES

Recommandations évidentes … mais toujours bonnes à rappeler

- Tout graphique doit avoir un titre, que l'on soit sur une interface interactive ou un document de synthèse produit d'un tableau de bord
- Chaque variable doit être décrite (légende, label axial…)
- Tout axe doit être gradué, en précisant les valeurs extrêmes
- **Rendu 3D à éviter** : perte en lisibilité vs. attractivité à peine meilleure [Few 08]



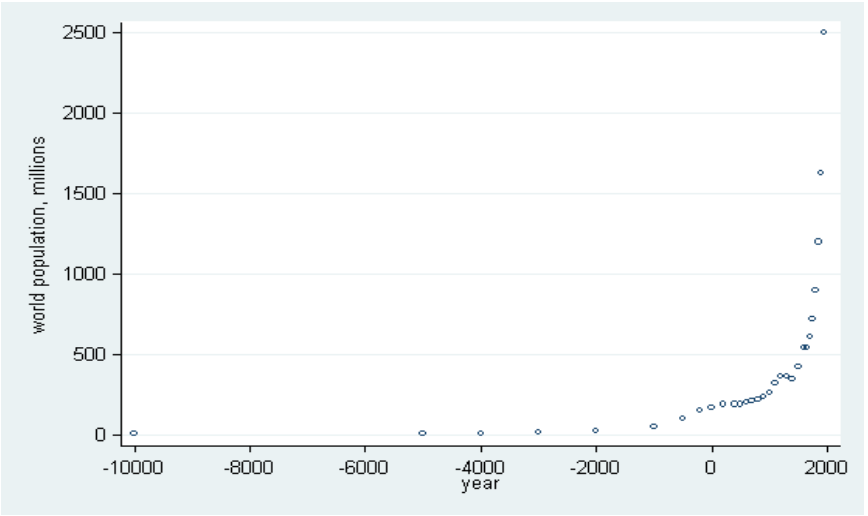


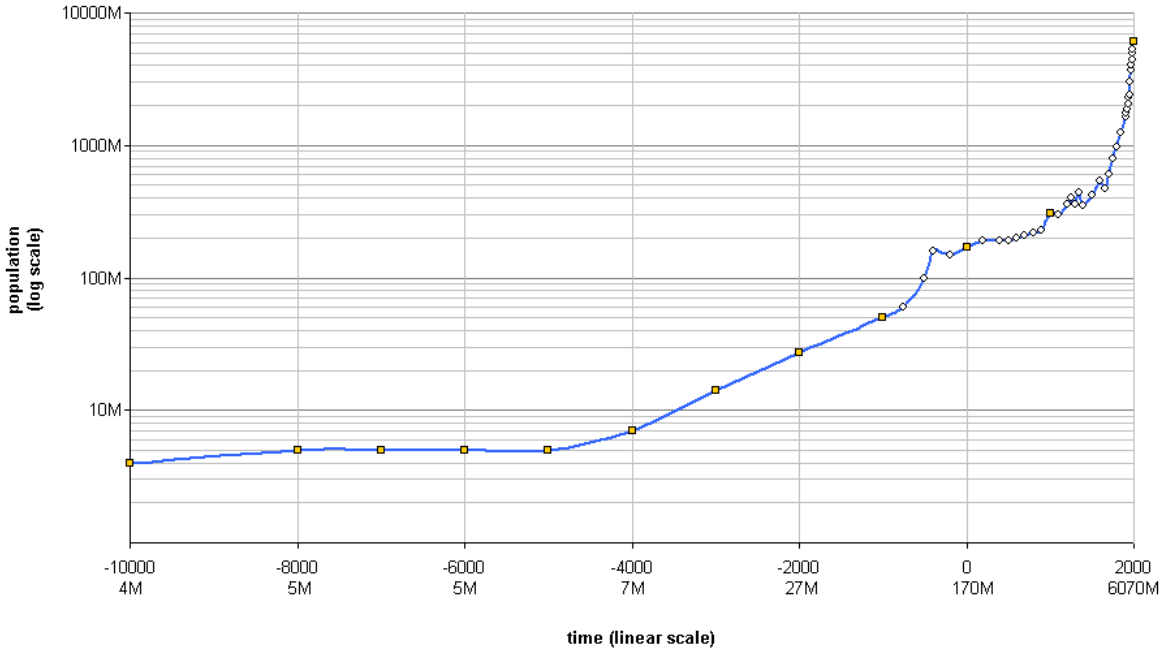
AFFICHAGES DE DONNEES NUMERIQUES

Echelle de valeurs

Règle – adapter l'échelle aux données, au problèmes et au but communicationnel ⇨ échelle linéaire ou logarithmique

Example – World population



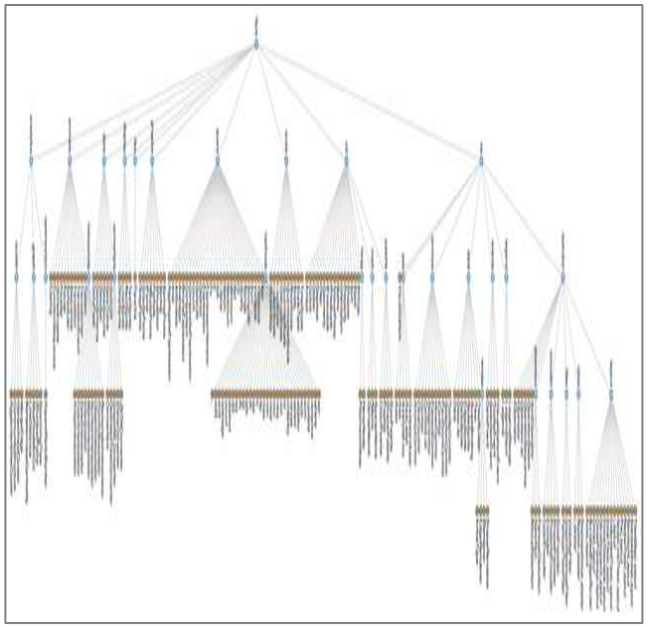


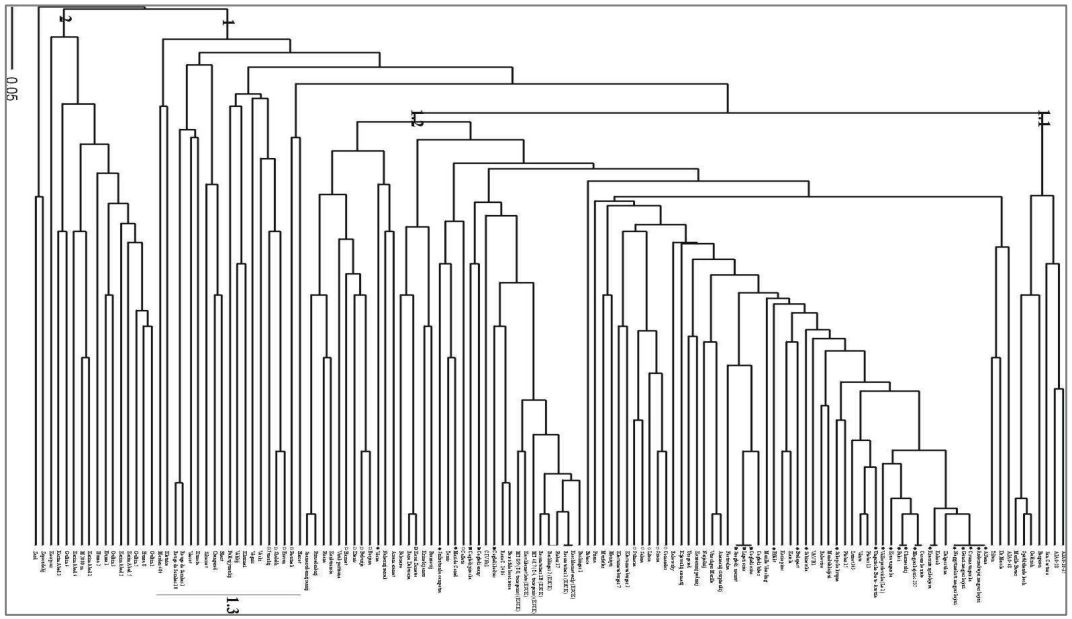
DISPLAYS FOR HIERARCHICAL DATA

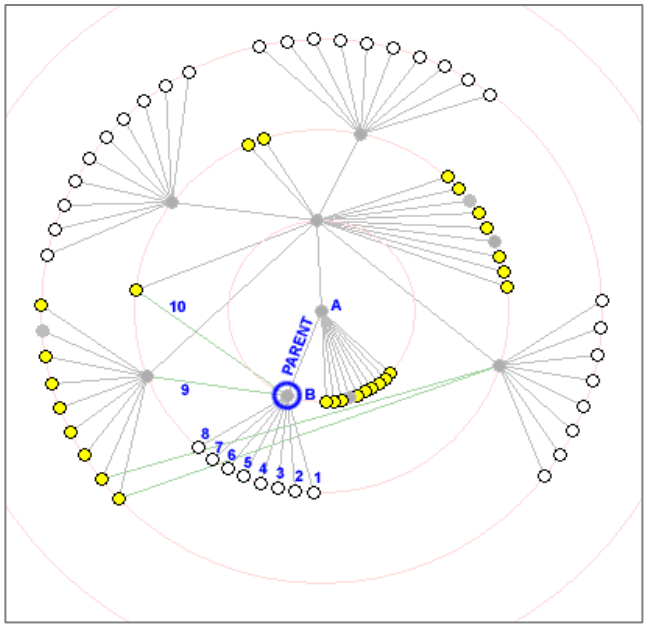
Charts *(graphes)* [Kosslyn, 1989 ; Heer and al. 2010]

- Relations between items and groups of items
- Not restricted to nominal data

- **Charts for hierarchical data**
  - Tree (leaf-node or hyperbolic),
  - Dendogram
  - *Adjency Tree, Tree Map* (chap. 1.3)





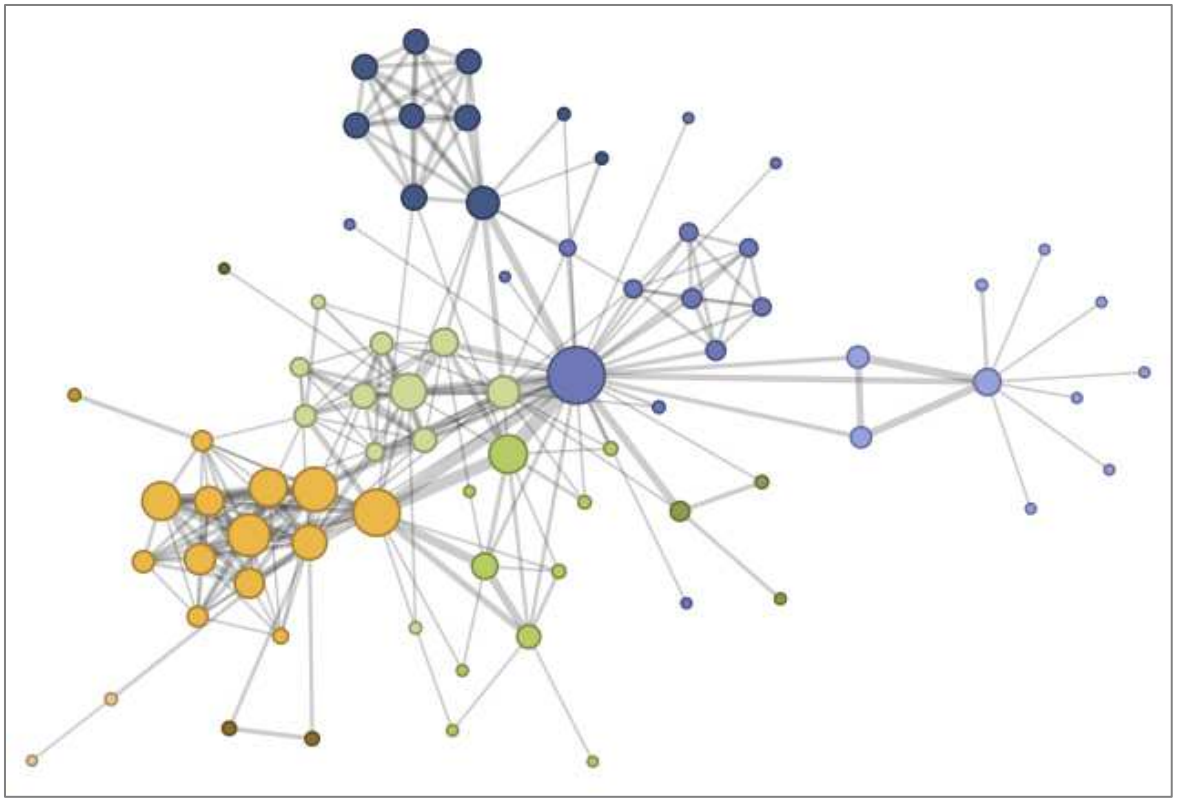


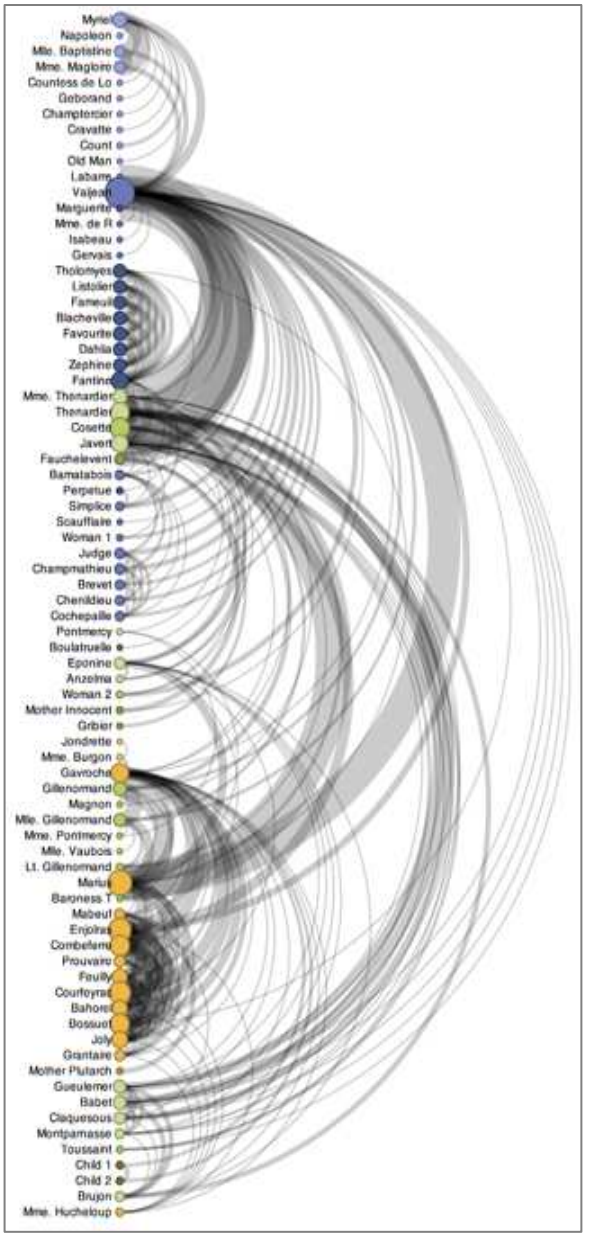
DISPLAYS FOR NETWORKS

Charts for networks

[Heer and al. 2010]

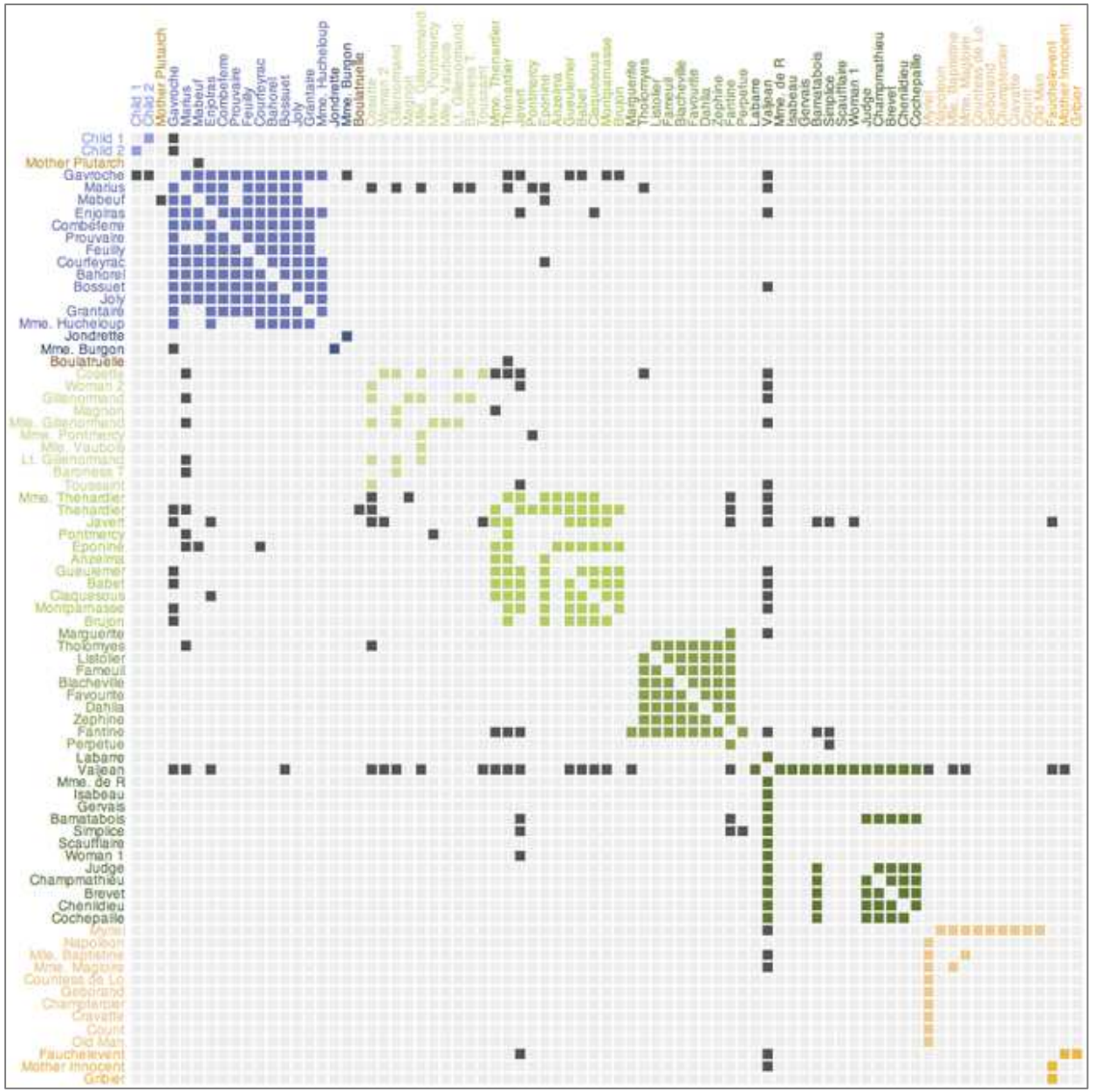
- **force-directed network** – consider links thickness to represent additional information
- **arc diagrams** – interactive ordering of the nodes





DISPLAYS FOR NETWORKS

Matrix representation



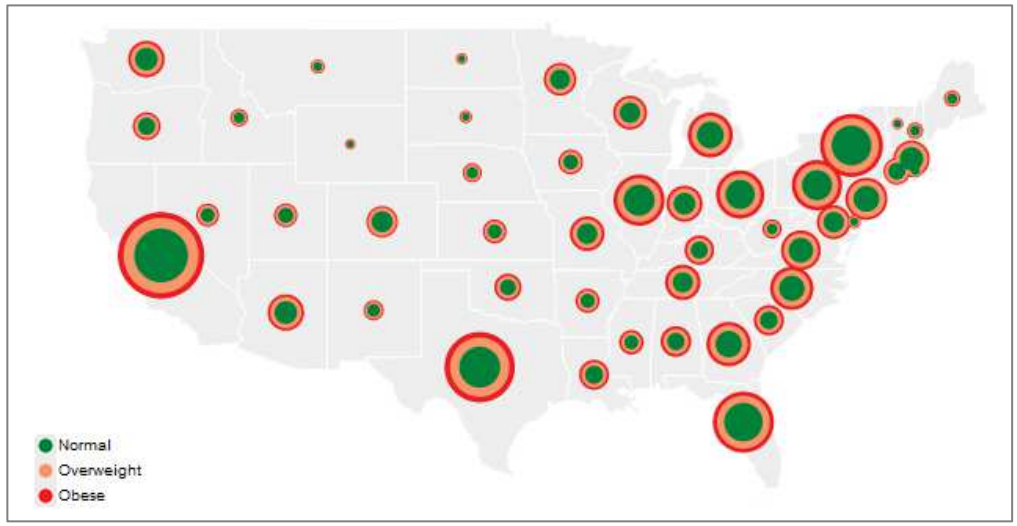
DISPLAYS FOR SPATIAL DATA

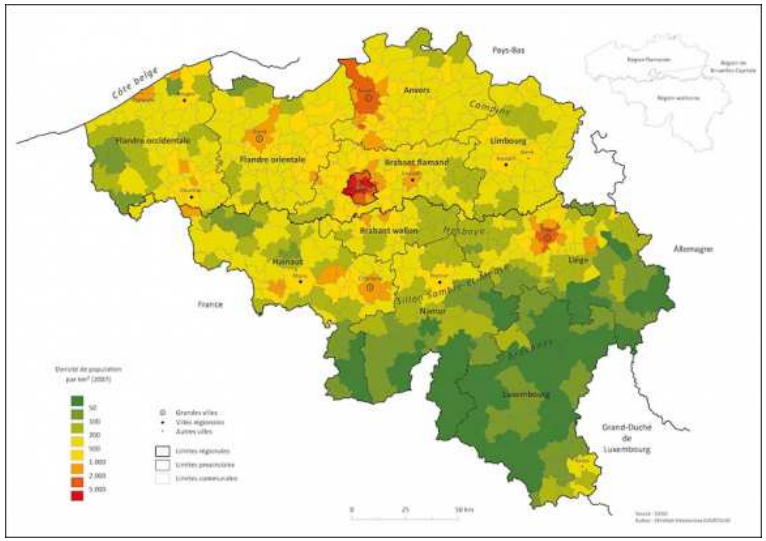
Maps for discrete data

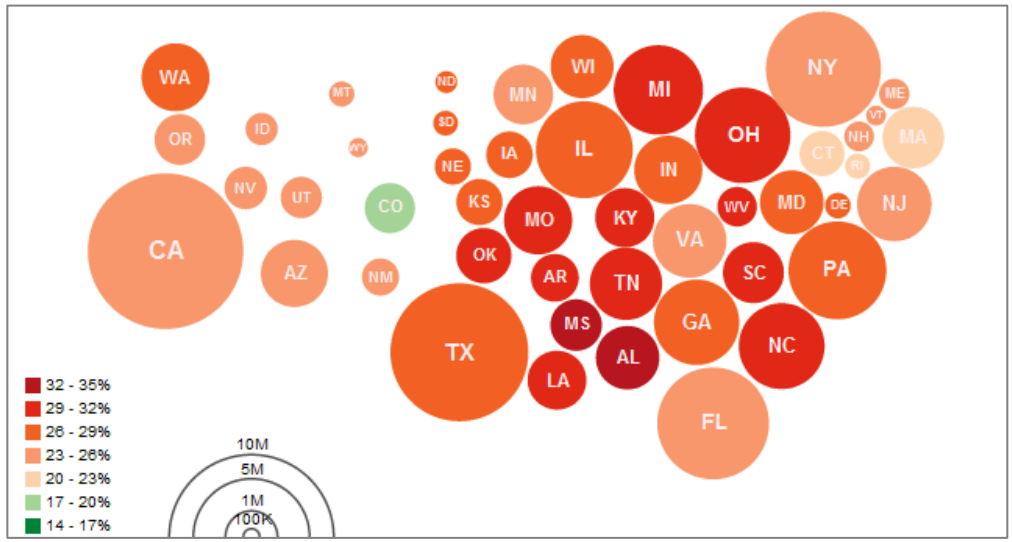
When the precise location of the observations is not important

Example : average value on a area

- choropleths (*area map)*
- graduates symbol maps
- cartograms



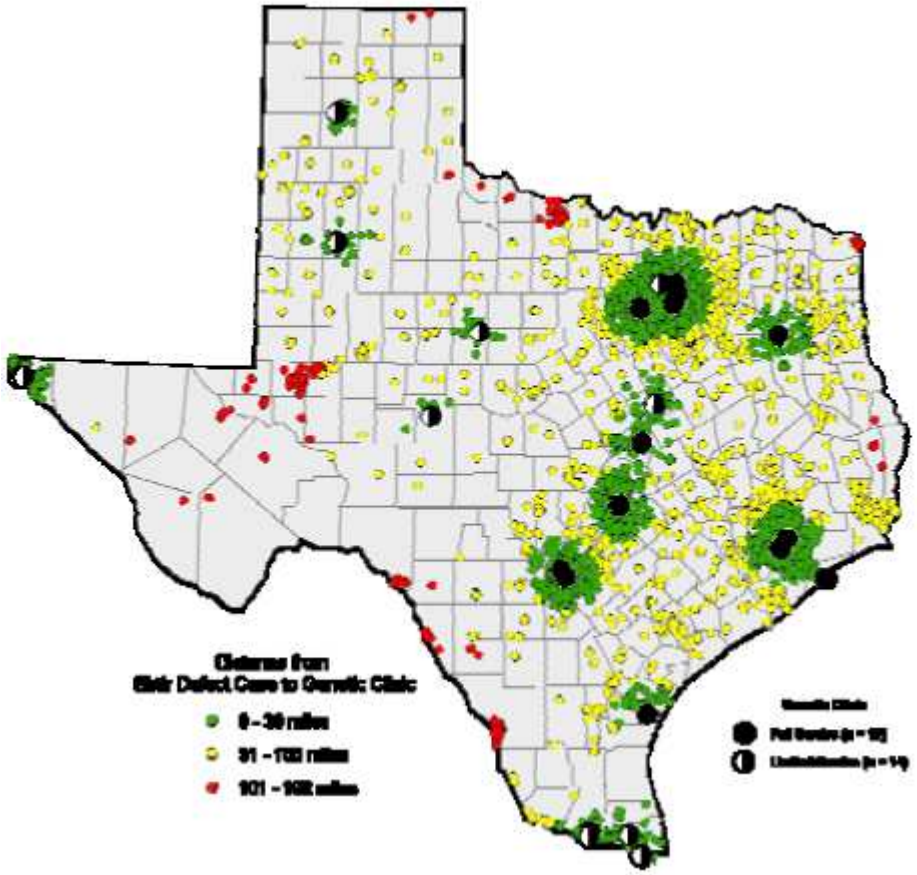




DISPLAYS FOR SPATIAL DATA

Maps for discrete data

Spot Map – When the precise location of the observation points is important

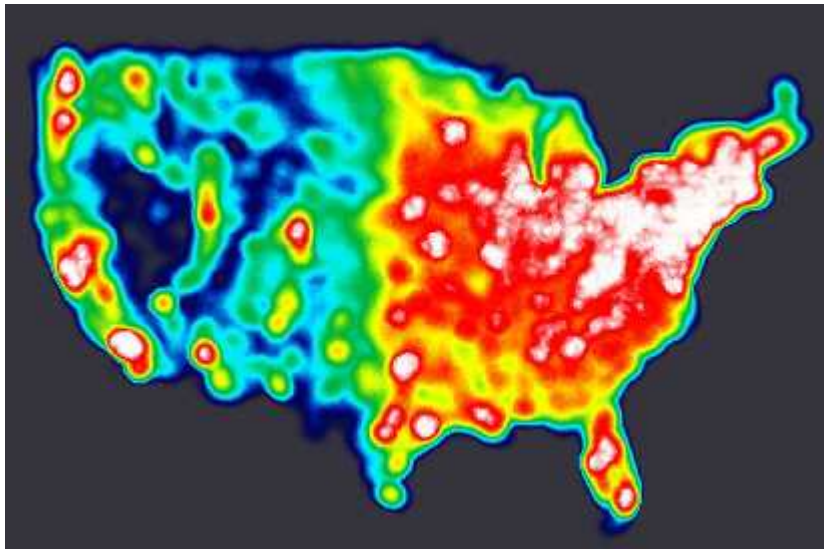


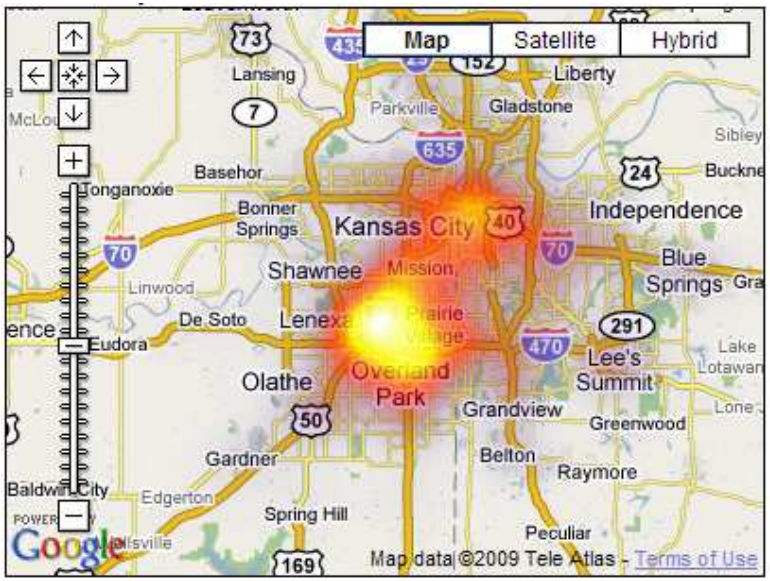
DISPLAYS FOR SPATIAL DATA

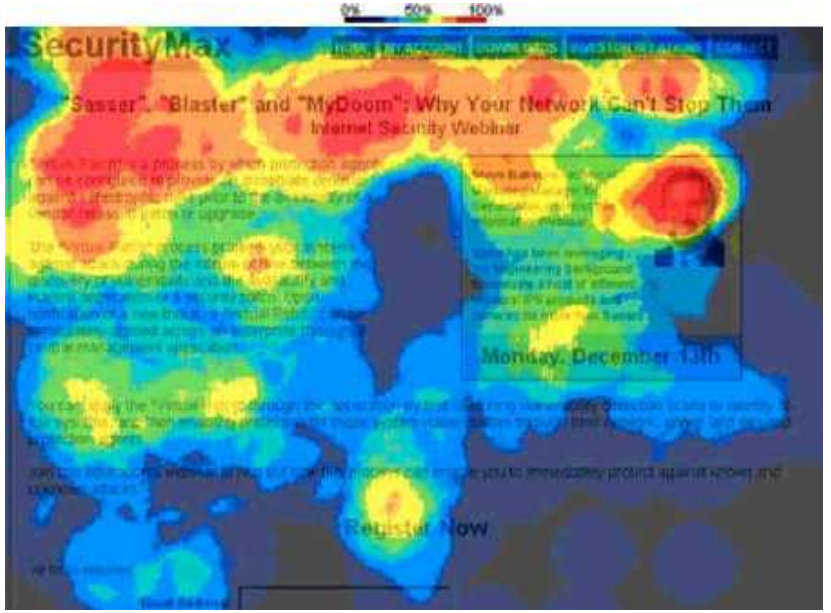
Maps for continuous data

- Heat maps (*cartes de densité*)

More details in the GIS Module







SYMBOLIC DISPLAYS

Diagrams

[Kosslyn, 1989]

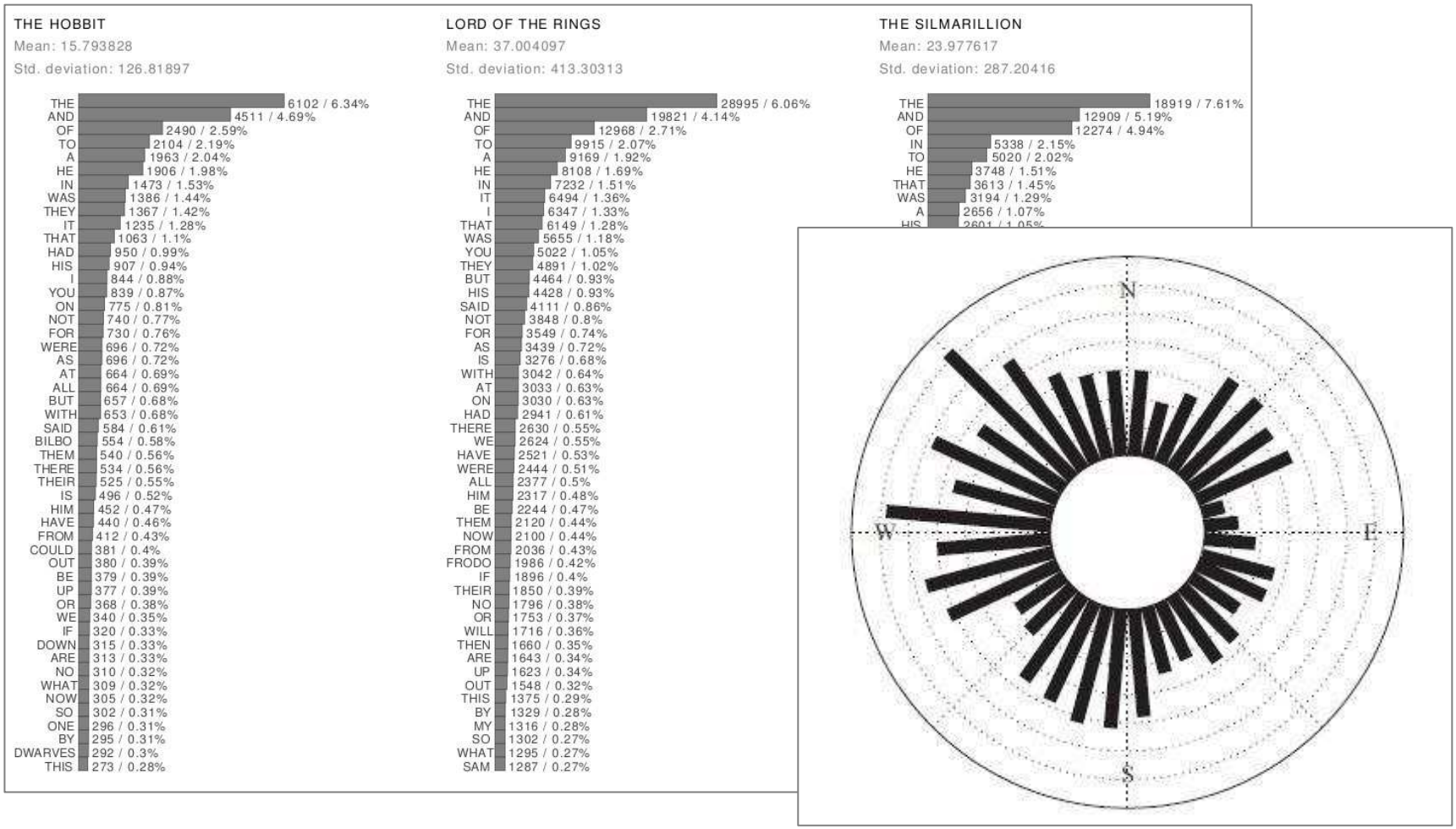
- Synthetic combination of schemata & symbols
- Could be related to information visualization
  (quantitative, ordinal, nominal)
- Be innovative !
- Respect visual and cognitive principles !

Word cloud (*Wordle*)

TYPES D'INFORMATION & VISUALISATION

Cas d'étude : fréquence des mots dans un texte

- **Affichage quantitatif** Histogrammes

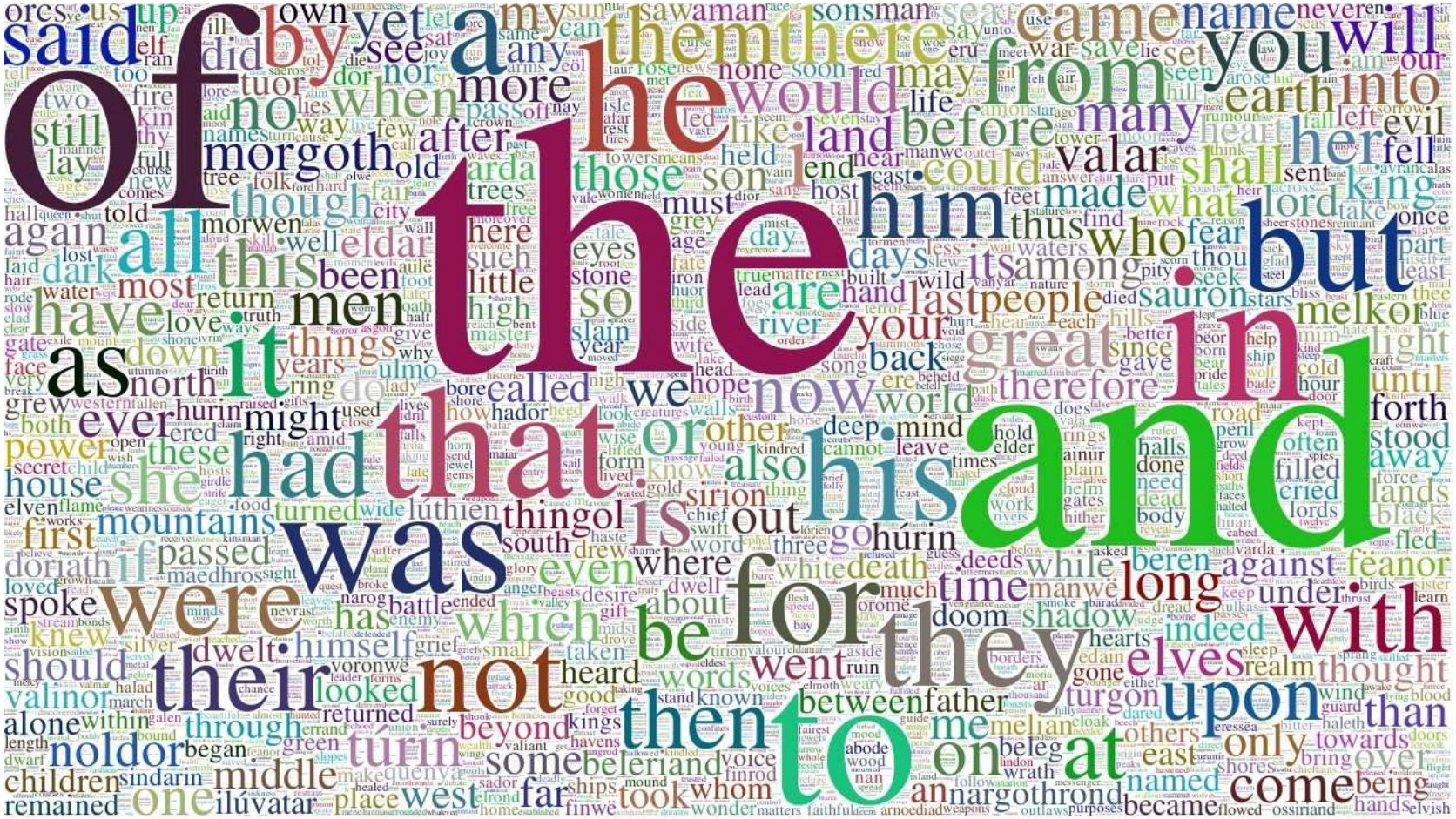


TYPES D'INFORMATION & VISUALISATION

Cas d'étude : fréquence des mots dans un texte

- **Affichage plutôt ordinal**

Nuage de mots (*words cloud*)



TYPES D'INFORMATION & VISUALISATION

Synthèse

If Data Are:		And These Conditions Apply:		Then Choose:
Proportions		< 6 data points	1 series	Pie chart (Sample Figure 4)
			>1 series	100% stacked bars (Sample Figure 5)
		6+ data points	1+ series	Consider combining data point categories or table. (Sample Figure 5)
Time Series		Numbers of Cases		Line chart (Sample Figure 6)
Data with discrete categories				Bar chart (Sample Figures 7, 8)
Place	Number of cases	Not readily identified on map		Bar chart (Sample Figure 11)
		Readily identified on map	Specific site important	Spot map (Sample Figure 9)
			Specific site unimportant	Area map (Sample Figure 10)
	Rates			Area map (Sample Figure 12)

[NBDPN 2008]

BIBLIOGRAPHIE

Ressources en ligne

- Heer J., Bostock M., Ogievetsky V. (2010) A tour through the vizualisation zoo. [http://hci.stanford.edu/jheer/files/zoo/]
- Information is Beautiful [http://www.informationisbeautiful.net/], un blog très visuel sur la conception de graphiques, d'idées sur la visualisation de données : enloy !

Ouvrages de référence

- Schneiderman B. (1986) The eyes have it: a task by data type taxonomy for visual information. Proc. *IEEE Symposium on Visual Language, VL'96*. Boulder, CO. www.mat.ucsb.edu/~g.legrady/academic/courses/11w259/schneiderman.pdf

Travaux cités

- Few S. (2008) Practical rules for using colors in charts. *Visual Business Intelligence Newsletter*.
- Hollands JG (2003) The classification of graphical elements. *Canadian Journal of Experimental Psychology*. **57**(1), 38-47.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, **3**, 185-225
- NBDPN (collectif, 2008) What Type of Chart or Graph Should I Use?, In *NBDPN Guidelines for Conducting Birth Defects Surveillance*, Appendix 11.4