Majeure Data

Session I

29 janvier - 1 février

~~Théorie 17h30~~ -> 10h30

~~1. Prise de décision et systèmes d'aide à la décision~~

~~2. Concepts de Business Intelligence et capacités de plate-forme~~

3. Visualization des données et conception de tableaux de bord

4. Business Performance Management Systems

5. BI-Projet de maturité, de stratégie et de synthèse

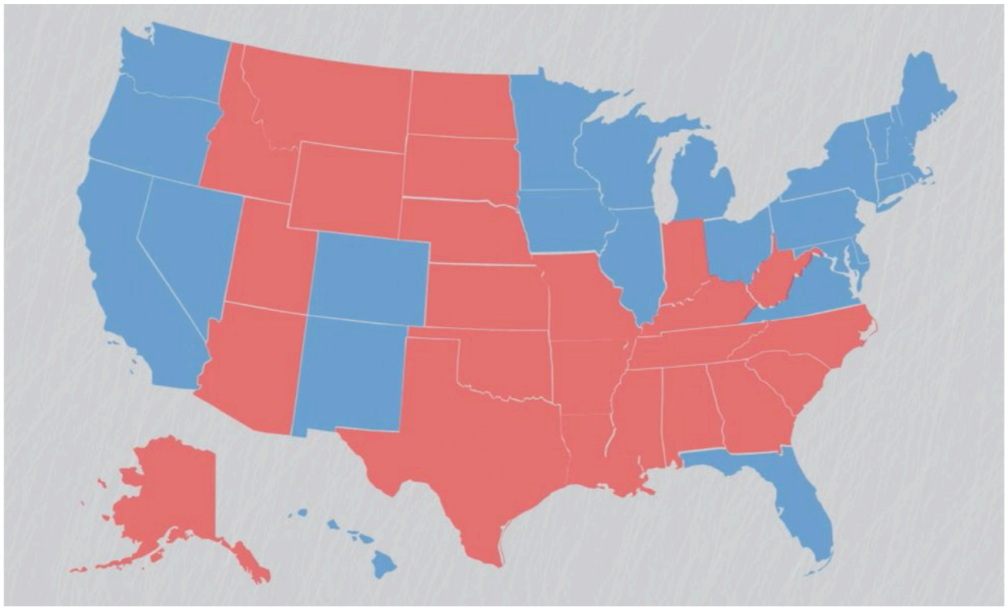
TP 10h30

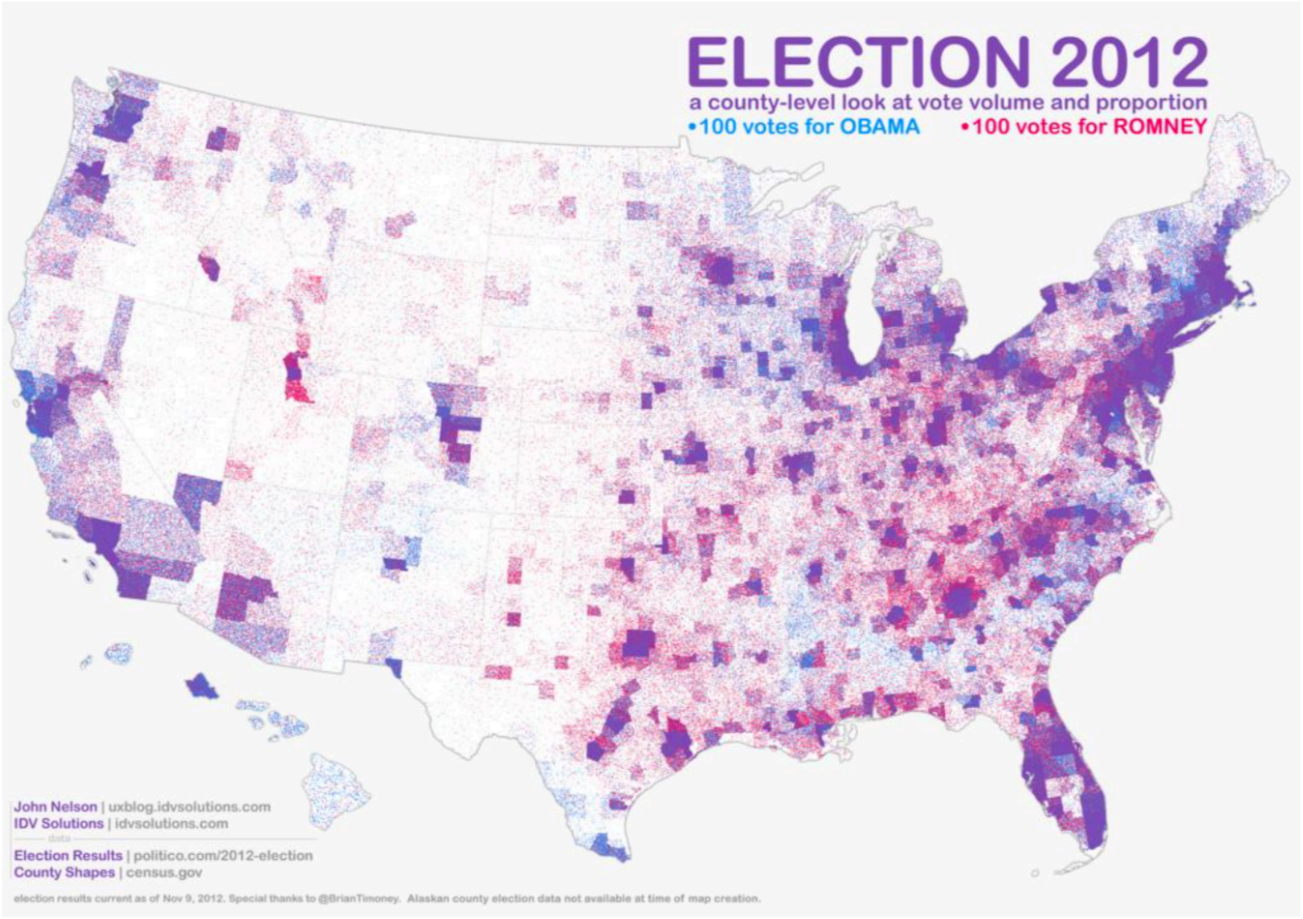
https://goo.gl/GkQ1H6

Zinnya DEL VILLAR

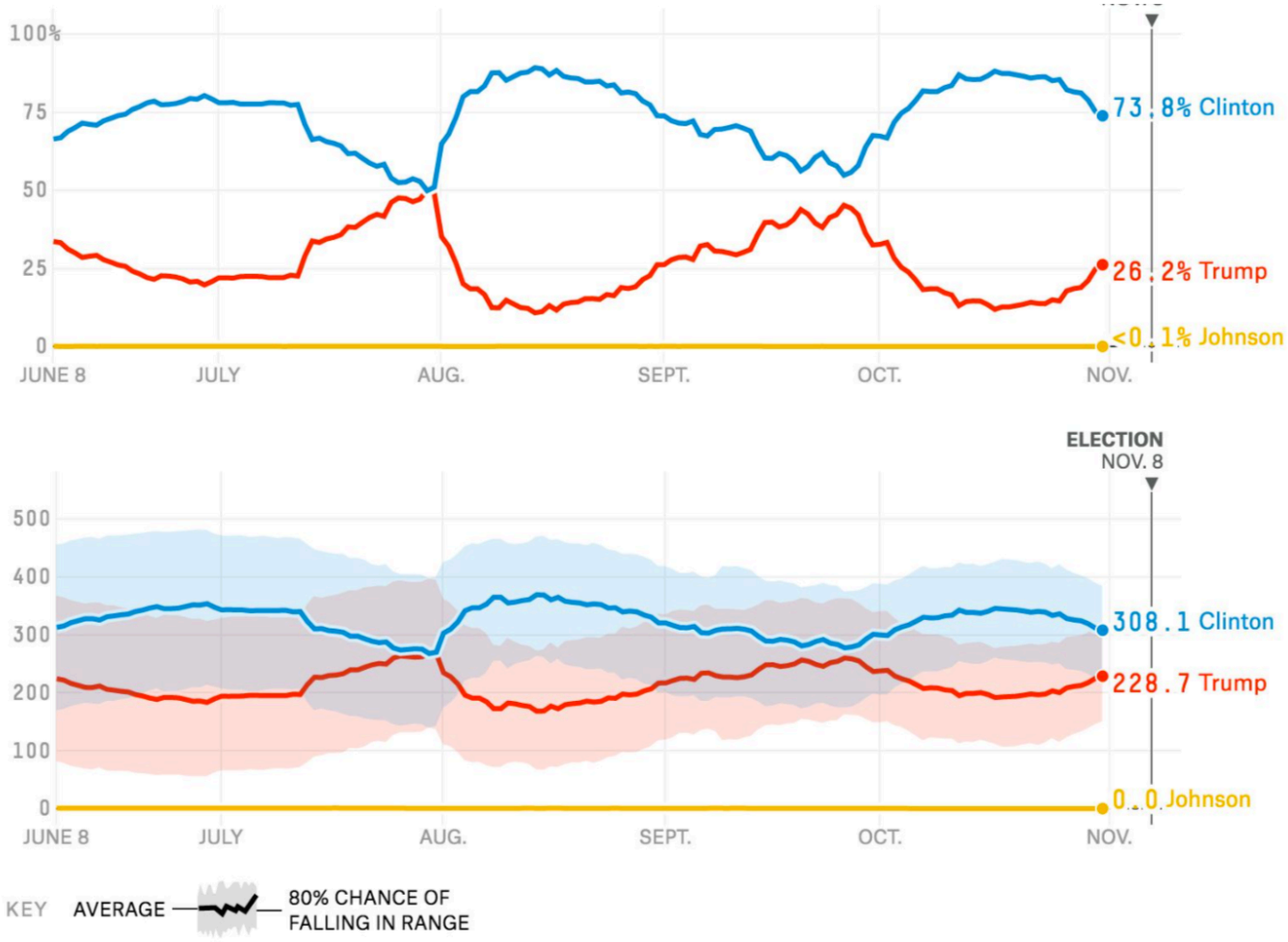
Critique

- À qui s'adresse la visualisation ?
  -> 1 proposition
- À quelle question la visualisation permet elle de répondre ?
  -> 1 proposition
- Pourquoi (n')aimez vous (pas) cette visualisation ?
  -> 2 raisons
- Quelles améliorations apporter ?
  -> 3 propositions





https://www.flickr.com/photos/idvsolutions/8182119174/sizes/k/in/photostream/



http://projects.fivethirtyeight.com/2016-election-forecast/

Les données

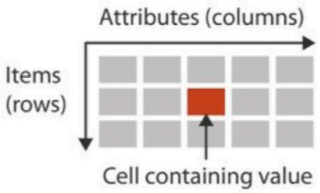
À la base de toute visualisation

Un bon designer de visualisation doit connaître :

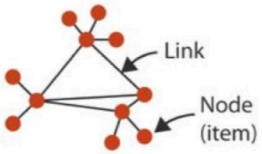
- Les propriétés des données
- Les méta-données associées
- Ce que les gens veulent tirer des données

Types de jeux de donnés

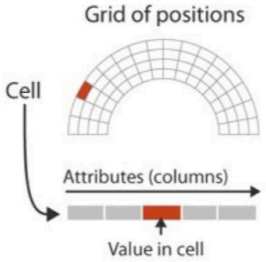
➔ Tables



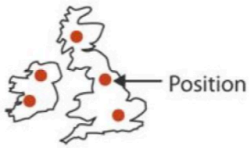
➔ Networks



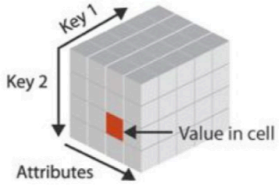
➔ Fields (Continuous)



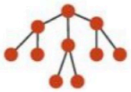
➔ Geometry (Spatial)



➔ *Multidimensional Table*



➔ *Trees*



-> Ce qu'on veut visualiser

Type de données de base

Unités fondamentales

Constituent les jeux de donnés

- Item / élément
- Lien
- Attribut
- Position
- Grille

—

Exemple item (élément)/attribut

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box		7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	[illegible]	5-Low	Wrap Bag	0.56	1/20/05
69	[illegible]5	4-Not Specified	Small Pack	0.44	6/6/05
69	[illegible]5	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Élément et attribut

Élément :

- Entité individuelle, discrète.
- Ex: un patient, une voiture

Attribut :

- Propriété mesurée ou observée
- Ex: taille, pression sanguine (patient), vitesse (voiture)

Lien, Position et Grille

Lien

- Relation entre deux éléments
- Ex : “amitié sur Facebook”

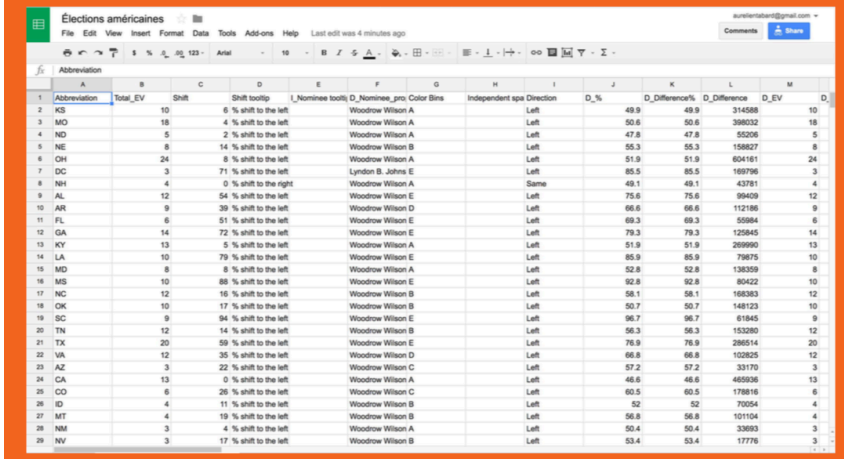
Position

- Données spatiales (en 2D ou 3D)
- Ex : latitude/longitude

Grille

- Stratégie d'échantillonnage pour données continues
- Ex: positions de stations météo

Données



Abbreviation	Total_EV	Shift	Shift tooltip	I_Nominee tooltip	D_Nominee_prop	Color Bins	Independent spa	Direction	D_%	D_Difference%	D_Difference	D_EV
KS	10	6	% shift to the left		Woodrow Wilson A			Left	49.9	49.9	314588	10
MO	18	4	% shift to the left		Woodrow Wilson A			Left	50.6	50.6	398032	18
ND	5	2	% shift to the left		Woodrow Wilson A			Left	47.8	47.8	55206	5
NE	8	14	% shift to the left		Woodrow Wilson B			Left	55.3	55.3	158827	8
OH	24	8	% shift to the left		Woodrow Wilson A			Left	51.9	51.9	604161	24
DC	3	71	% shift to the left		Lyndon B. Johns E			Left	85.5	85.5	169796	3
NH	4	0	% shift to the right		Woodrow Wilson A			Same	49.1	49.1	43781	4
AL	12	54	% shift to the left		Woodrow Wilson E			Left	75.6	75.6	99409	12
AR	9	39	% shift to the left		Woodrow Wilson D			Left	66.6	66.6	112186	9
FL	6	51	% shift to the left		Woodrow Wilson E			Left	69.3	69.3	55984	6
GA	14	72	% shift to the left		Woodrow Wilson E			Left	79.3	79.3	125845	14
KY	13	5	% shift to the left		Woodrow Wilson A			Left	51.9	51.9	269990	13
LA	10	79	% shift to the left		Woodrow Wilson E			Left	85.9	85.9	79875	10
MD	8	8	% shift to the left		Woodrow Wilson A			Left	52.8	52.8	138359	8
MS	10	88	% shift to the left		Woodrow Wilson E			Left	92.8	92.8	80422	10
NC	12	16	% shift to the left		Woodrow Wilson B			Left	58.1	58.1	168383	12
OK	10	17	% shift to the left		Woodrow Wilson B			Left	50.7	50.7	148123	10
SC	9	94	% shift to the left		Woodrow Wilson E			Left	96.7	96.7	61845	9
TN	12	14	% shift to the left		Woodrow Wilson B			Left	56.3	56.3	153280	12
TX	20	59	% shift to the left		Woodrow Wilson E			Left	76.9	76.9	286514	20
VA	12	35	% shift to the left		Woodrow Wilson D			Left	66.8	66.8	102825	12
AZ	3	22	% shift to the left		Woodrow Wilson C			Left	57.2	57.2	33170	3
CA	13	0	% shift to the left		Woodrow Wilson A			Left	46.6	46.6	465936	13
CO	6	26	% shift to the left		Woodrow Wilson C			Left	60.5	60.5	178816	6
ID	4	11	% shift to the left		Woodrow Wilson B			Left	52	52	70054	4
MT	4	19	% shift to the left		Woodrow Wilson B			Left	56.8	56.8	101104	4
NM	3	4	% shift to the left		Woodrow Wilson A			Left	50.4	50.4	33693	3
NV	3	17	% shift to the left		Woodrow Wilson B			Left	53.4	53.4	17776	3

Exercice : goo.gl/cZYJkK

Trouver à quoi correspond :

- Un item / un élément / une variable (indépendante)
- Un attribut / une dimension / une variable (dépendante) / une feature
- Les clés

Ou est définie la sémantique de la table ?

Type d'échelles

Nominale (catégoriel)

- Fruits: pommes, oranges, ...

Ordinale (ordonné)

- Qualité d'un frigo: A+, A++, A+++ ...
- Peut être compté et ordonné mais pas mesuré

Intervalle (zéro arbitraire)

- Dates, longitude, latitude

Ratio (zero fixé)

- Le zéro a un sens (rien)
- Mesure physique : poid, longueur, ...

Type d'échelles

Nominale (catégoriel)

- Opérations : =, ≠

Ordinale (ordonné)

- Opérations : =, ≠, >, <

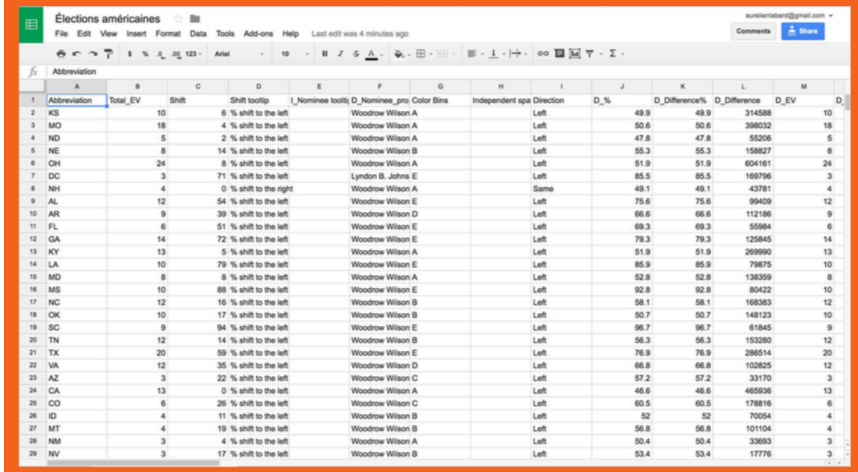
Intervalle (zéro arbitraire) ex : [1989 – 1999] + [2002 – 2012]

- Opérations : =, ≠, >, <, +, − peut mesurer les distances

Ratio (zero fixé) ex : 10kg / 5kg

- Opérations: =, ≠, >, <, +, −, ×, ÷ peut mesurer les proportions

Données


Élections américaines

File Edit View Insert Format Data Tools Add-ons Help Last edit was 4 minutes ago

Comments Share

Abbreviation

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Abbreviation	Total_EV	Shift	Shift tooltip	I_Nominee tools	D_Nominee_pro	Color Bins	Independent spa	Direction	D_%	D_Difference%	D_Difference	D_EV
2	KS	10	6	% shift to the left		Woodrow Wilson	A		Left	49.9	49.9	314588	10
3	MO	18	4	% shift to the left		Woodrow Wilson	A		Left	50.6	50.6	398032	18
4	ND	5	2	% shift to the left		Woodrow Wilson	A		Left	47.8	47.8	55206	5
5	NE	8	14	% shift to the left		Woodrow Wilson	B		Left	55.3	55.3	158827	8
6	OH	24	8	% shift to the left		Woodrow Wilson	A		Left	51.9	51.9	604161	24
7	DC	3	71	% shift to the left		Lyndon B. Johns	E		Left	85.5	85.5	169796	3
8	NH	4	0	% shift to the right		Woodrow Wilson	A		Same	49.1	49.1	43781	4
9	AL	12	54	% shift to the left		Woodrow Wilson	E		Left	75.6	75.6	99409	12
10	AR	9	39	% shift to the left		Woodrow Wilson	D		Left	66.6	66.6	112186	9
11	FL	6	51	% shift to the left		Woodrow Wilson	E		Left	69.3	69.3	55984	6
12	GA	14	72	% shift to the left		Woodrow Wilson	E		Left	79.3	79.3	125945	14
13	KY	13	5	% shift to the left		Woodrow Wilson	A		Left	51.9	51.9	269990	13
14	LA	10	79	% shift to the left		Woodrow Wilson	E		Left	85.9	85.9	79875	10
15	MD	8	8	% shift to the left		Woodrow Wilson	A		Left	52.8	52.8	138359	8
16	MS	10	88	% shift to the left		Woodrow Wilson	E		Left	92.8	92.8	80422	10
17	NC	12	16	% shift to the left		Woodrow Wilson	B		Left	58.1	58.1	168383	12
18	OK	10	17	% shift to the left		Woodrow Wilson	B		Left	50.7	50.7	148123	10
19	SC	9	94	% shift to the left		Woodrow Wilson	E		Left	96.7	96.7	61845	9
20	TN	12	14	% shift to the left		Woodrow Wilson	B		Left	56.3	56.3	153280	12
21	TX	20	59	% shift to the left		Woodrow Wilson	E		Left	76.9	76.9	286514	20
22	VA	12	35	% shift to the left		Woodrow Wilson	D		Left	66.8	66.8	102825	12
23	AZ	3	22	% shift to the left		Woodrow Wilson	C		Left	57.2	57.2	33170	3
24	CA	13	0	% shift to the left		Woodrow Wilson	A		Left	46.6	46.6	465936	13
25	CO	6	26	% shift to the left		Woodrow Wilson	C		Left	60.5	60.5	178816	6
26	ID	4	11	% shift to the left		Woodrow Wilson	B		Left	52	52	70054	4
27	MT	4	19	% shift to the left		Woodrow Wilson	B		Left	56.8	56.8	101104	4
28	NM	3	4	% shift to the left		Woodrow Wilson	A		Left	50.4	50.4	33693	3
29	NV	3	17	% shift to the left		Woodrow Wilson	B		Left	53.4	53.4	17776	3

Exercice : goo.gl/cZYJkK

Trouver un type de données :

- Nominal / Catégoriel
- Ordinal / Ordonné
- Interval
- Ratio

Modèle de données vs. conceptuel

Modèle de données (description bas niveau)

- Flottants : 32.5, 54.0, -17.3

Modèle conceptuel (construction mentale)

- Température

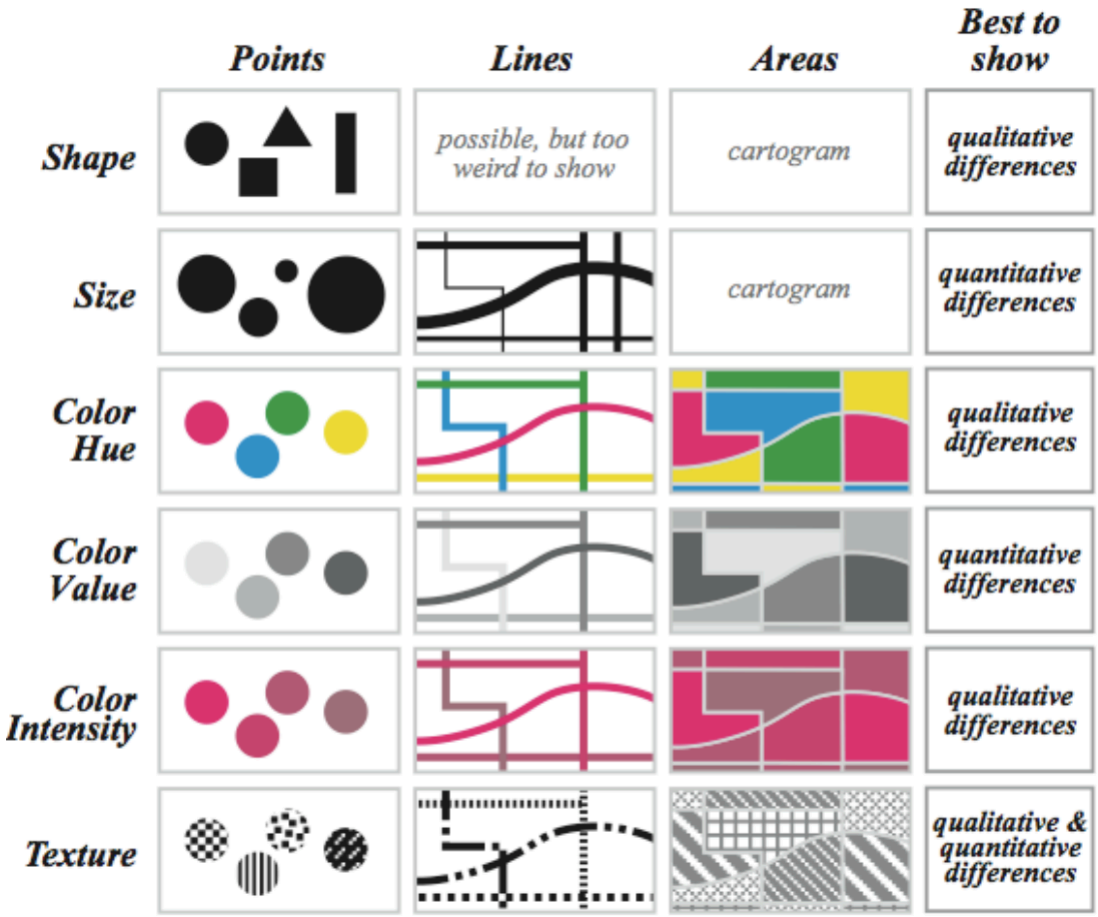
Type de données

- Valeur continue avec 1 chiffre significatif (Quantitatif)
- Chaud - tiède - froid - glacé (Ordinal)
- Brulé / pas brulé (Nominal)

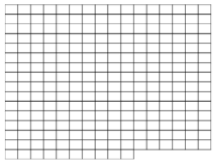
—

Les variables de Jacques Bertin

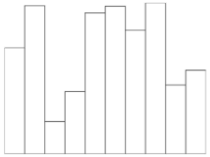
Cartographe français,
auteur de la sémiologie graphique



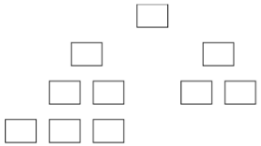
BASIC DATAVIZ TEMPLATES



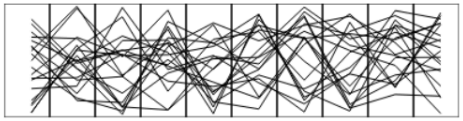
grid / isotype



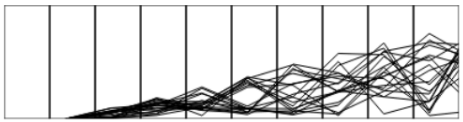
bar chart



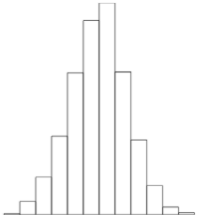
tree layout



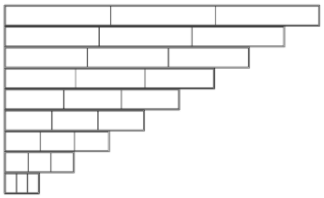
parallel coordinates



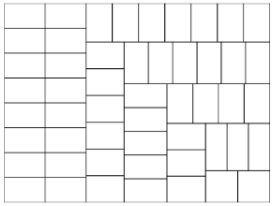
line chart



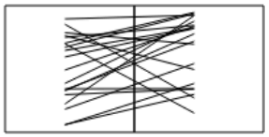
histogram



stacked chart



treemap



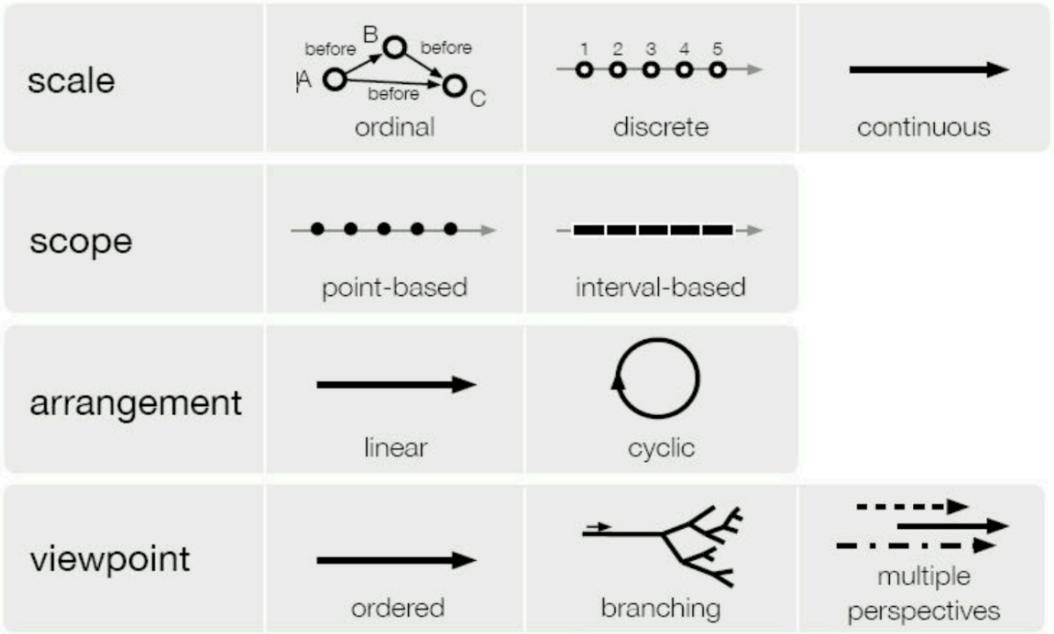
slope graph

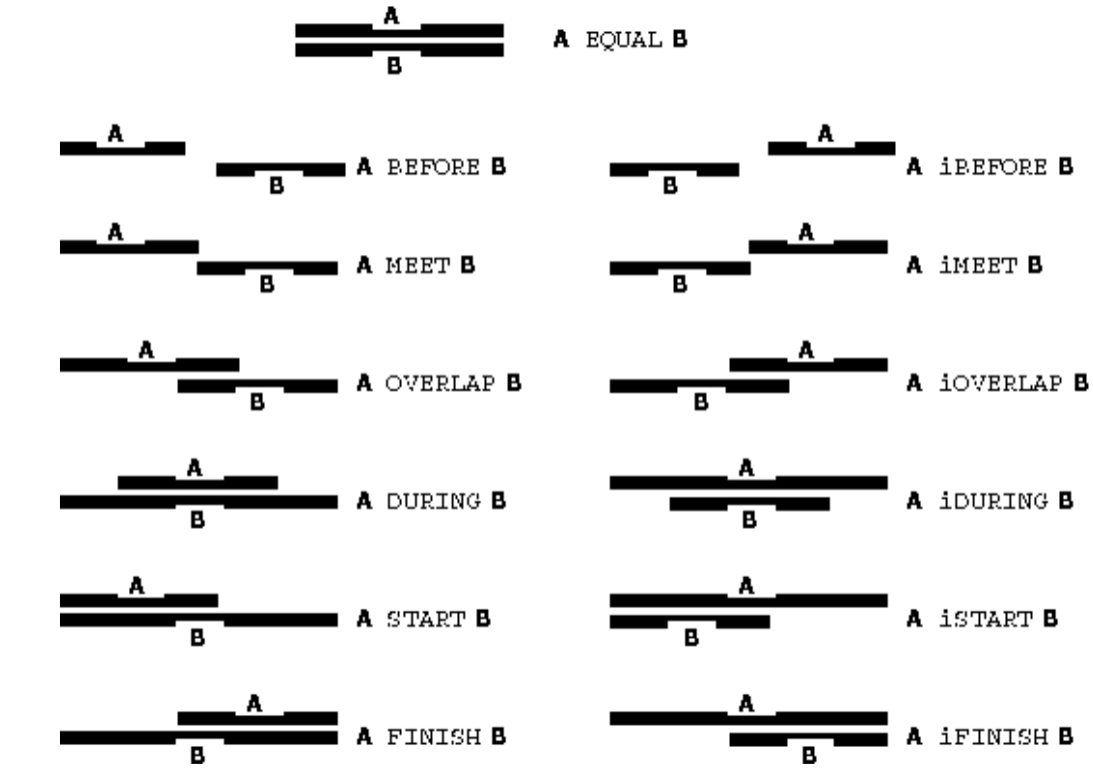
LIST OF CHARTS TEMPLATES

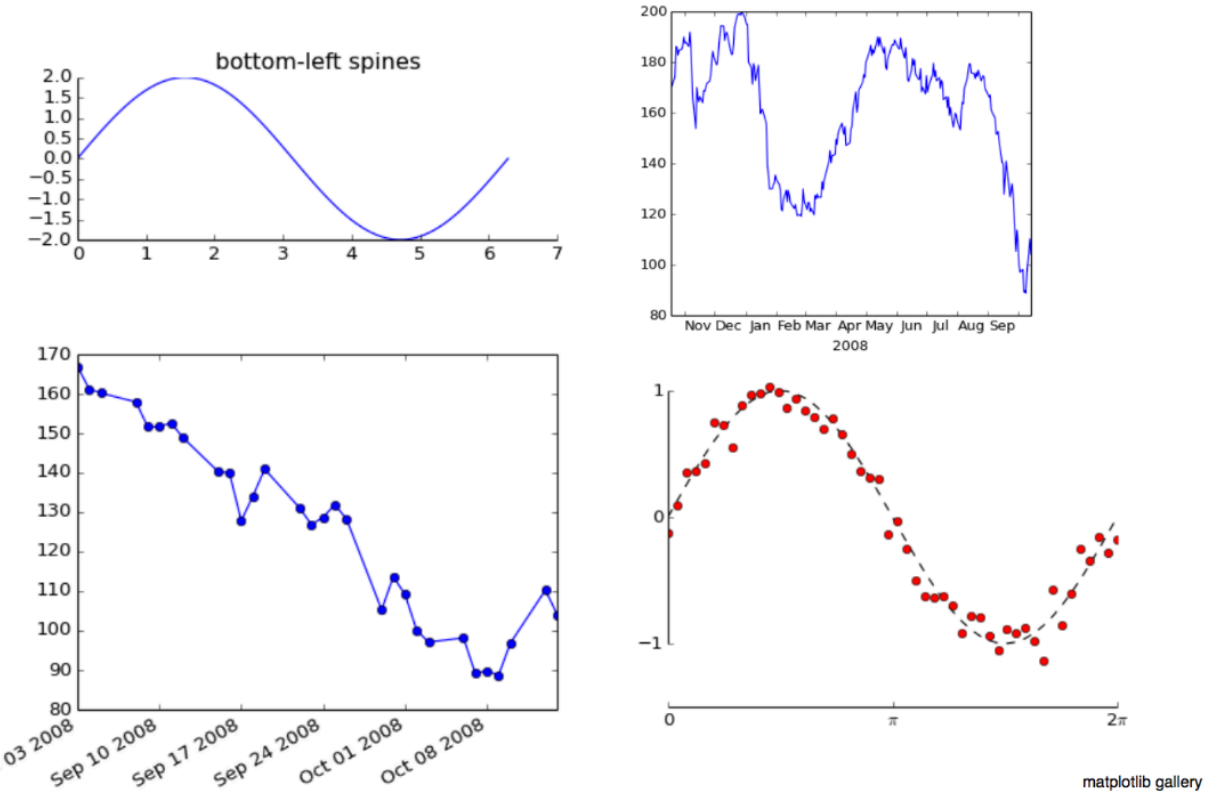
Datavisualization: temps

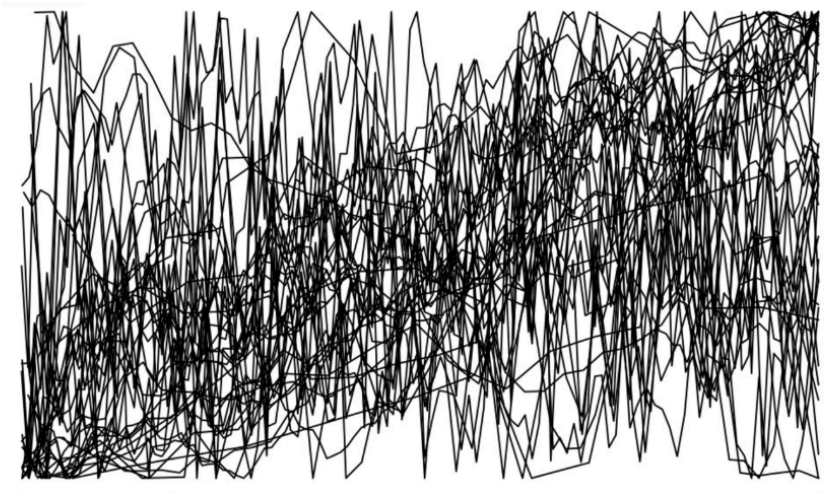
Qu'est ce que c'est le temps?

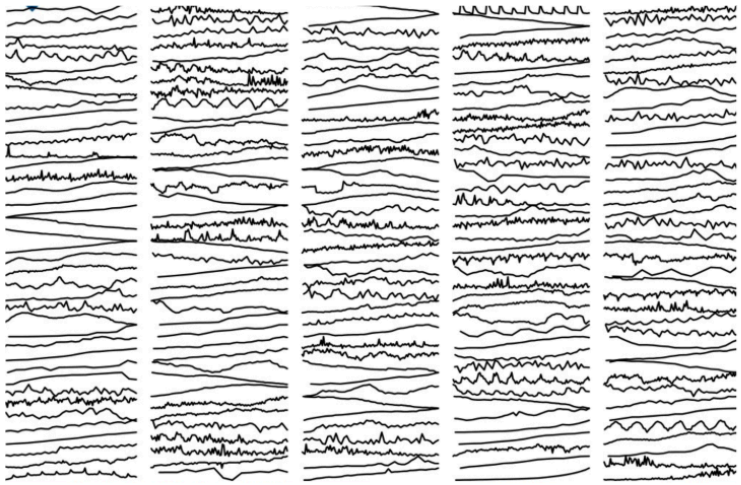
Visualization du temps

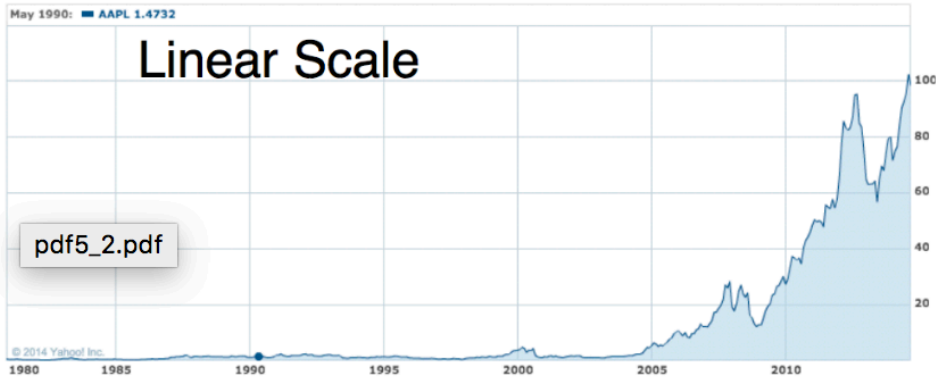



A
B
A EQUAL B
A BEFORE B
A iBEFORE B
A MEET B
A iMEET B
A OVERLAP B
A iOVERLAP B
A DURING B
A iDURING B
A START B
A iSTART B
A FINISH B
A iFINISH B

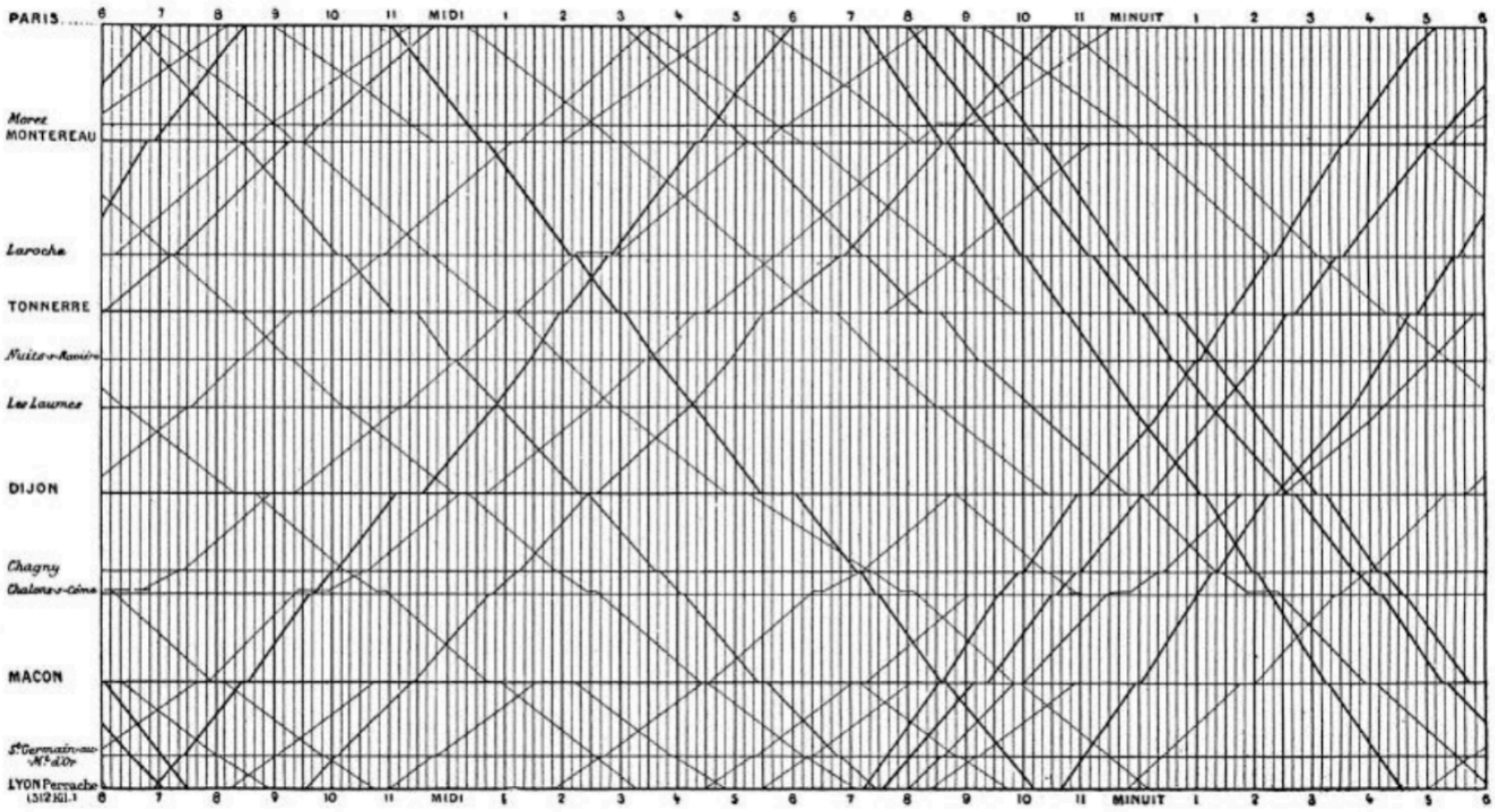

bottom-left spines
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
1
2
3
4
5
6
7
200
180
160
140
120
100
80
Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
2008
170
160
150
140
130
120
110
100
90
80
03 2008
Sep 10 2008
Sep 17 2008
Sep 24 2008
Oct 01 2008
Oct 08 2008
1
0
−1
0
π
2π
matplotlib gallery



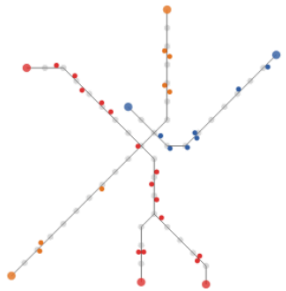



May 1990: AAPL 1.4732
Linear Scale
pdf5_2.pdf
© 2014 Yahoo! Inc.
1980
1985
1990
1995
2000
2005
2010
100
80
60
40
20


May 1990: AAPL 1.4732
Log Scale
© 2014 Yahoo! Inc.
1980
1985
1990
1995
2000
2005
2010
100
80
60
40
20


PARIS
6 7 8 9 10 11 MIDI 1 2 3 4 5 6 7 8 9 10 11 MINUIT 1 2 3 4 5 6
Morez
MONTEREAU
Laroche
TONNERRE
Nuits-s-Ravière
Les Laumes
DIJON
Chagny
Chalons-s-Saône
MACON
St Germain-au-Mt d'Or
LYON Perrache
6 7 8 9 10 11 MIDI 1 2 3 4 5 6 7 8 9 10 11 MINUIT 1 2 3 4 5 6

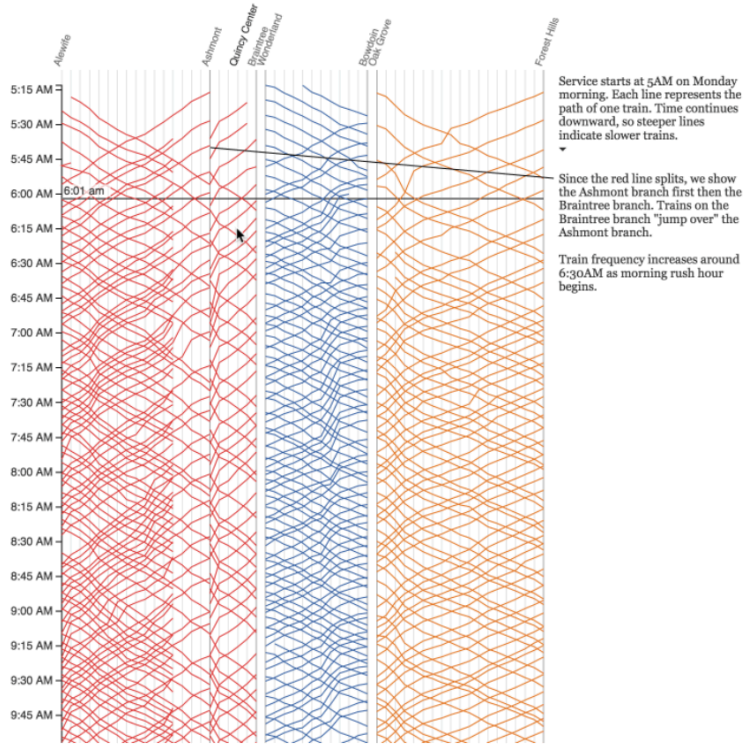
Subway Trips on Monday February 3, 2014



Locations of each train on the red, blue, and orange lines at 6:01 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

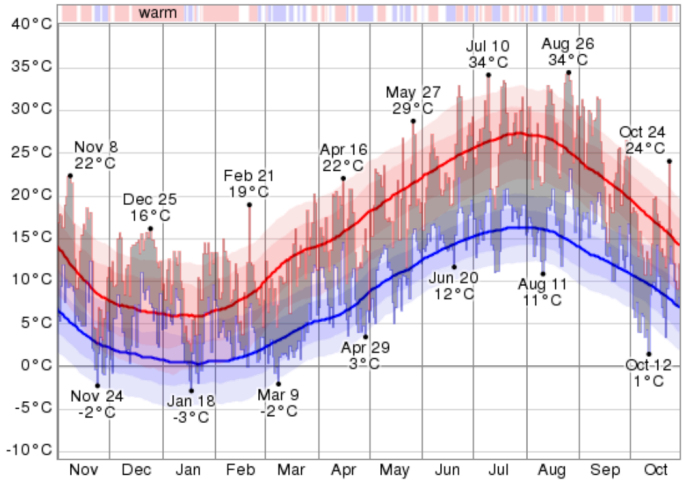
See the morning rush-hour, midday lull, afternoon rush-hour, and the evening lull.



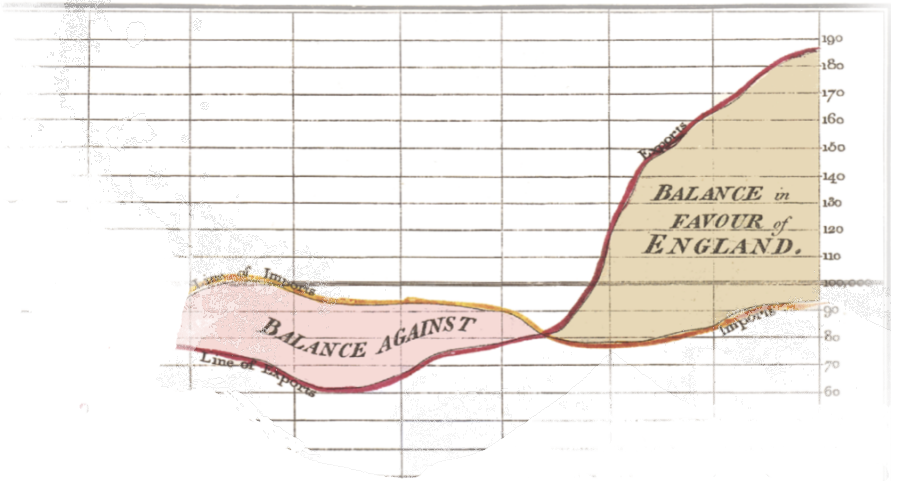
Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

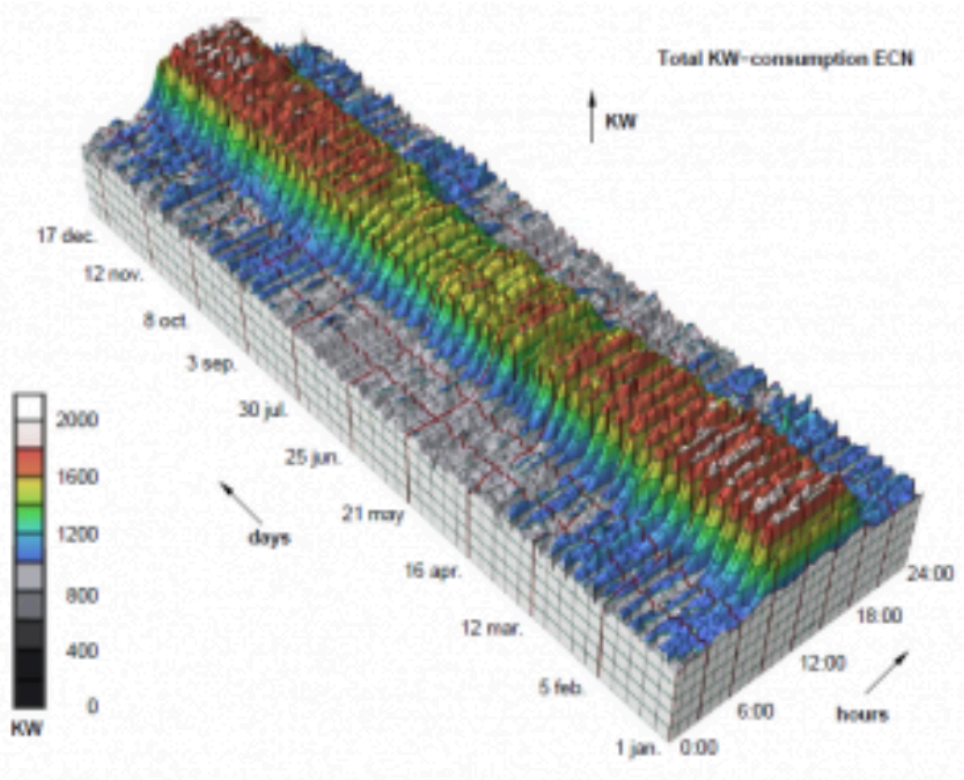
Train frequency increases around 6:30AM as morning rush hour begins.


warm
40°C
35°C
30°C
25°C
20°C
15°C
10°C
5°C
0°C
-5°C
-10°C
Nov 8
22°C
Dec 25
16°C
Nov 24
-2°C
Jan 18
-3°C
Feb 21
19°C
Mar 9
-2°C
Apr 16
22°C
Apr 29
3°C
May 27
29°C
Jun 20
12°C
Jul 10
34°C
Aug 11
11°C
Aug 26
34°C
Oct 24
24°C
Oct 12
1°C
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct


SELL
SELL
BUY
SELL
BUY
BUY
www.stock-trading-infocentre.com
August
September
October
November
December
2010


Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.
BALANCE in FAVOUR of ENGLAND.
BALANCE AGAINST
Line of Imports
Line of Exports
Exports
Imports

3D


Total KW-consumption ECN
KW
17 dec.
12 nov.
8 oct.
3 sep.
30 jul.
25 jun.
21 may
16 apr.
12 mar.
5 feb.
1 jan.
days
0:00
6:00
12:00
18:00
24:00
hours
2000
1600
1200
800
400
0
KW

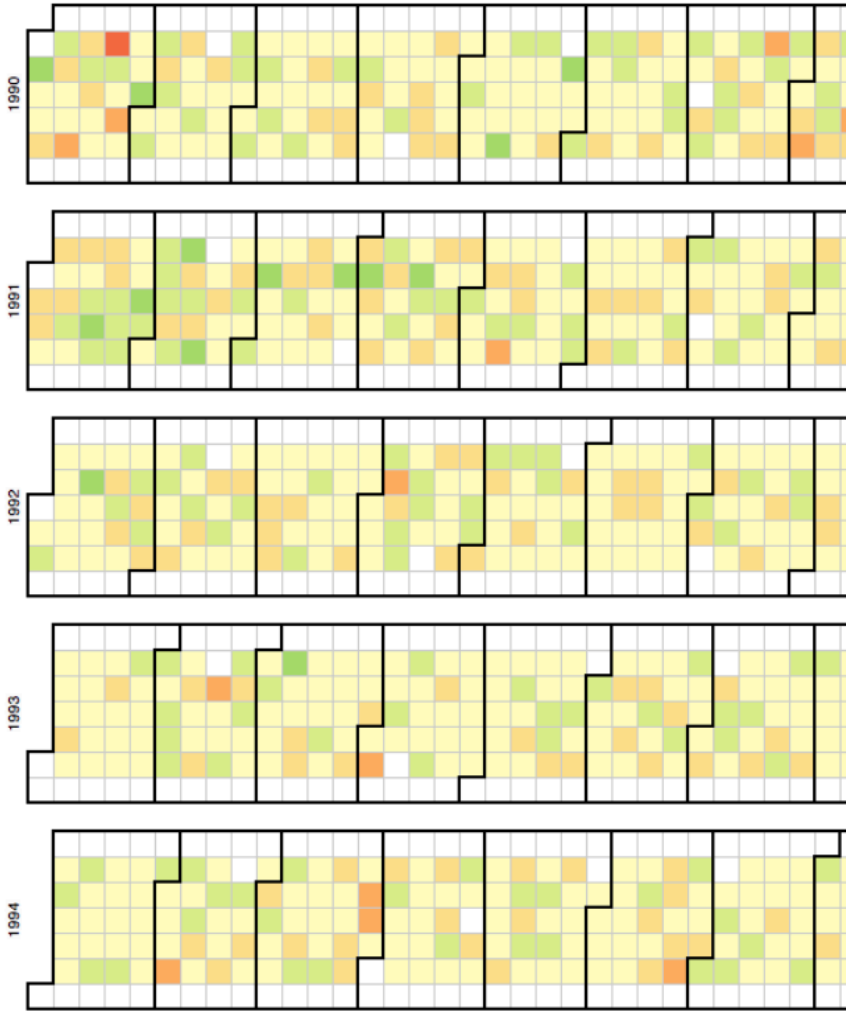
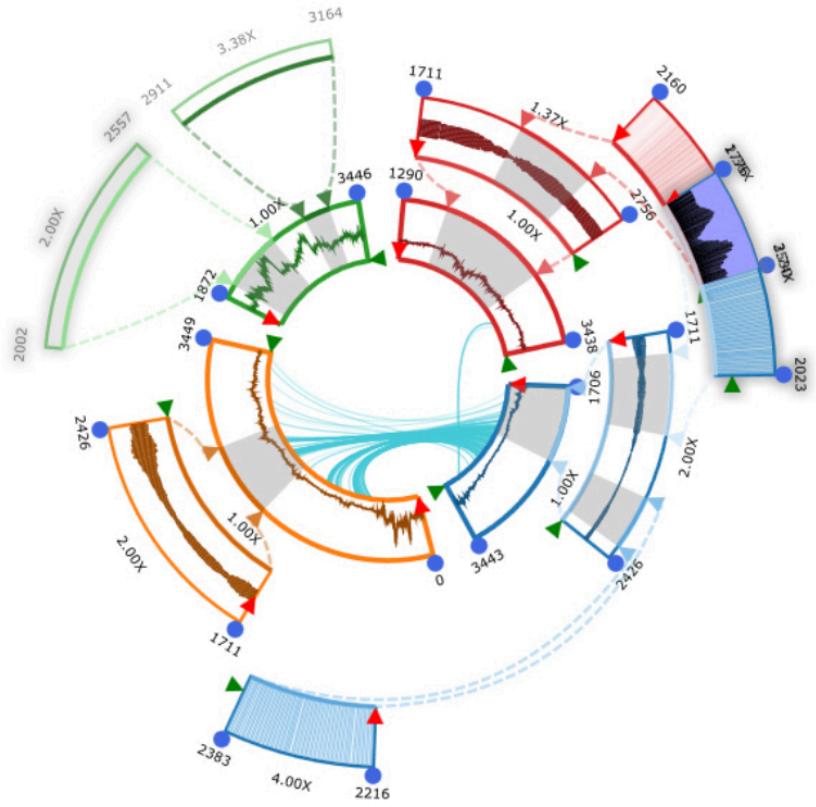
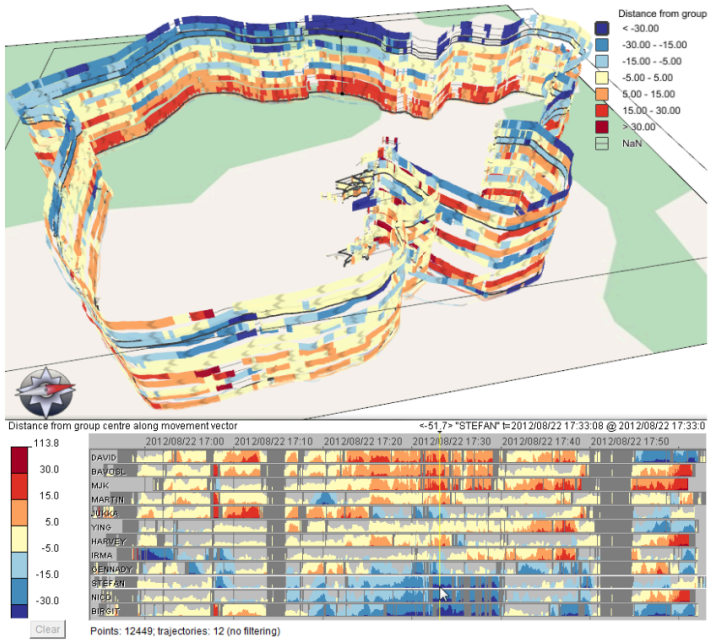

1990
1991
1992
1993
1994


Chart
Map
How to use
Share graph
Full screen
lin
Life expectancy (years)
80
75
70
65
60
55
50
45
40
35
30
25
20
15
2012
400
1 000
2 000
4 000
10 000
20 000
40 000
100 000
Income per person (GDP/capita, PPP$ inflation-adjusted)
log
Play
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Trails
Terms of use
Color
Geographic regions
Select
Afghanistan
Albania
Algeria
Angola
Antigua and Barb...
Argentina
Armenia
Aruba
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Deselect all
Size
Population, total
1.36 B
0
© Google 2008

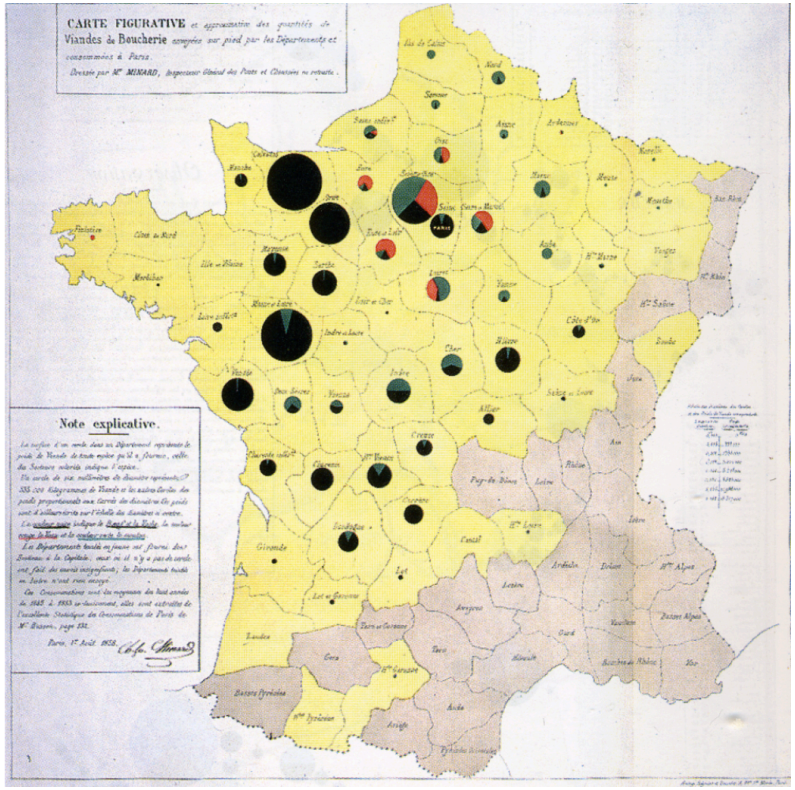


KronoMiner: Using Multi-Foci Navigation for the Visual Exploration of Time-Series Data
https://www.youtube.com/watch?v=U0IN7vfrxi0

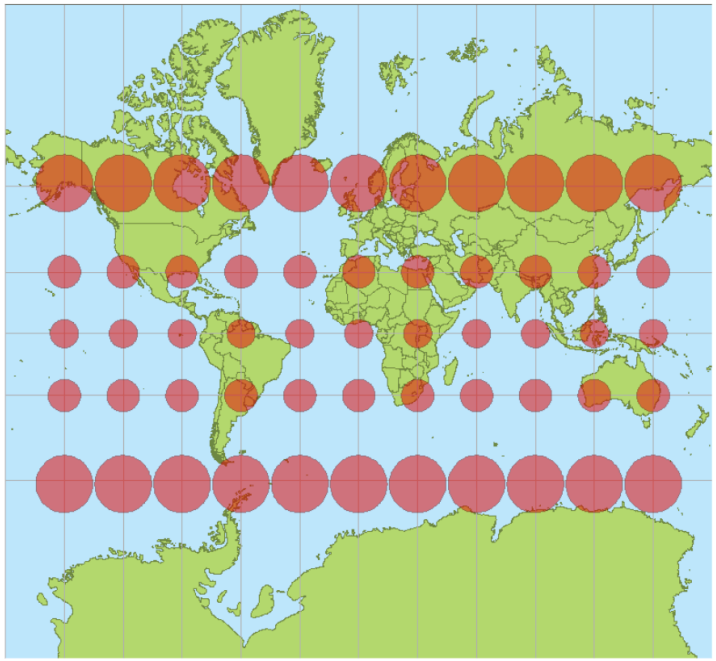
Espace & temps



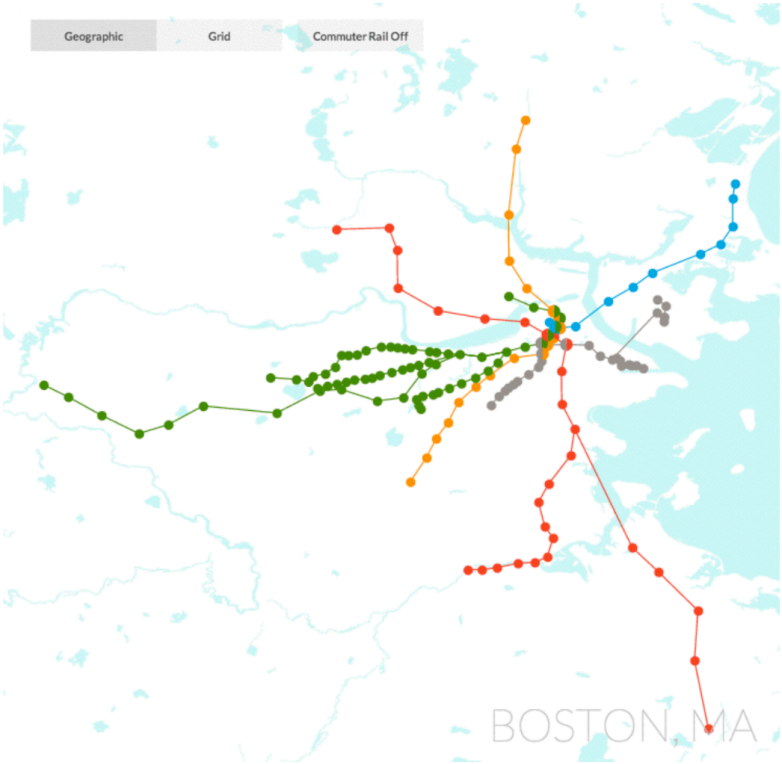
Datavisualization: geo-map



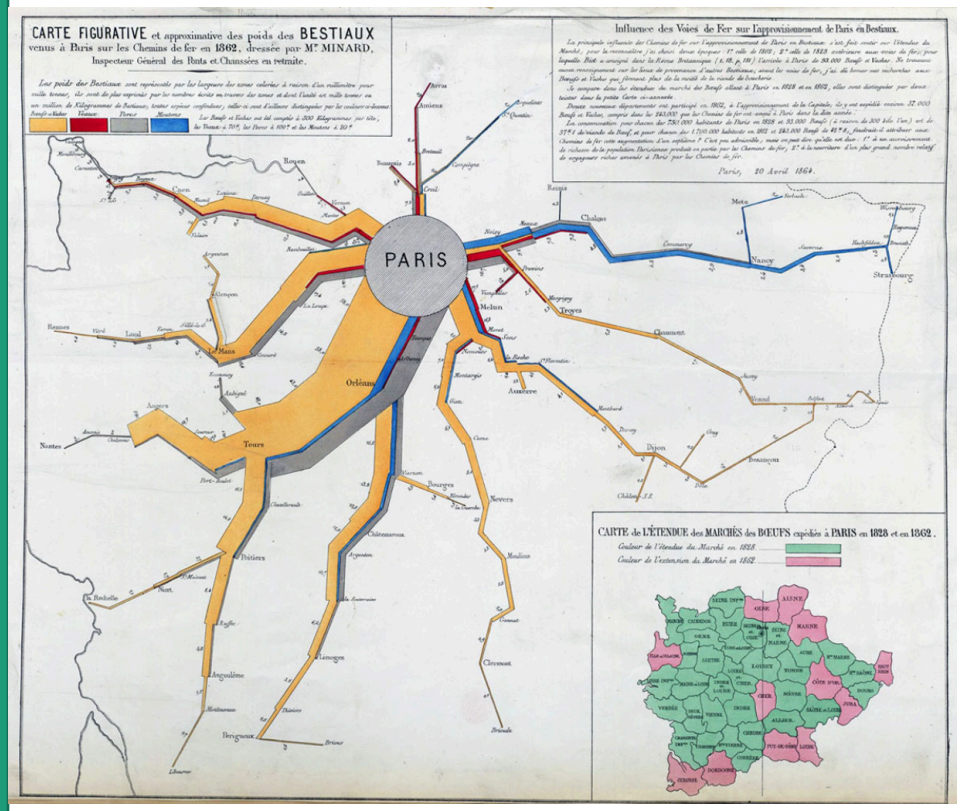
Problèmes avec les projections



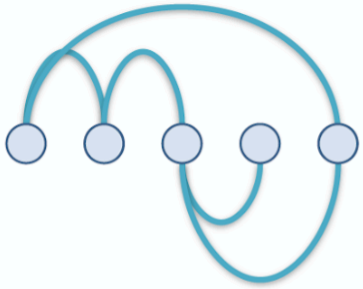
Datavisualization: Geo+Network



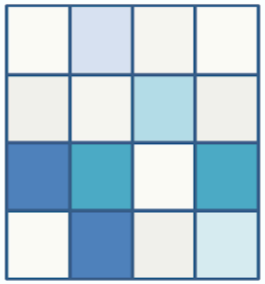
Datavisualization: Geo+Flow



Datavisualization: Arbres et graphes

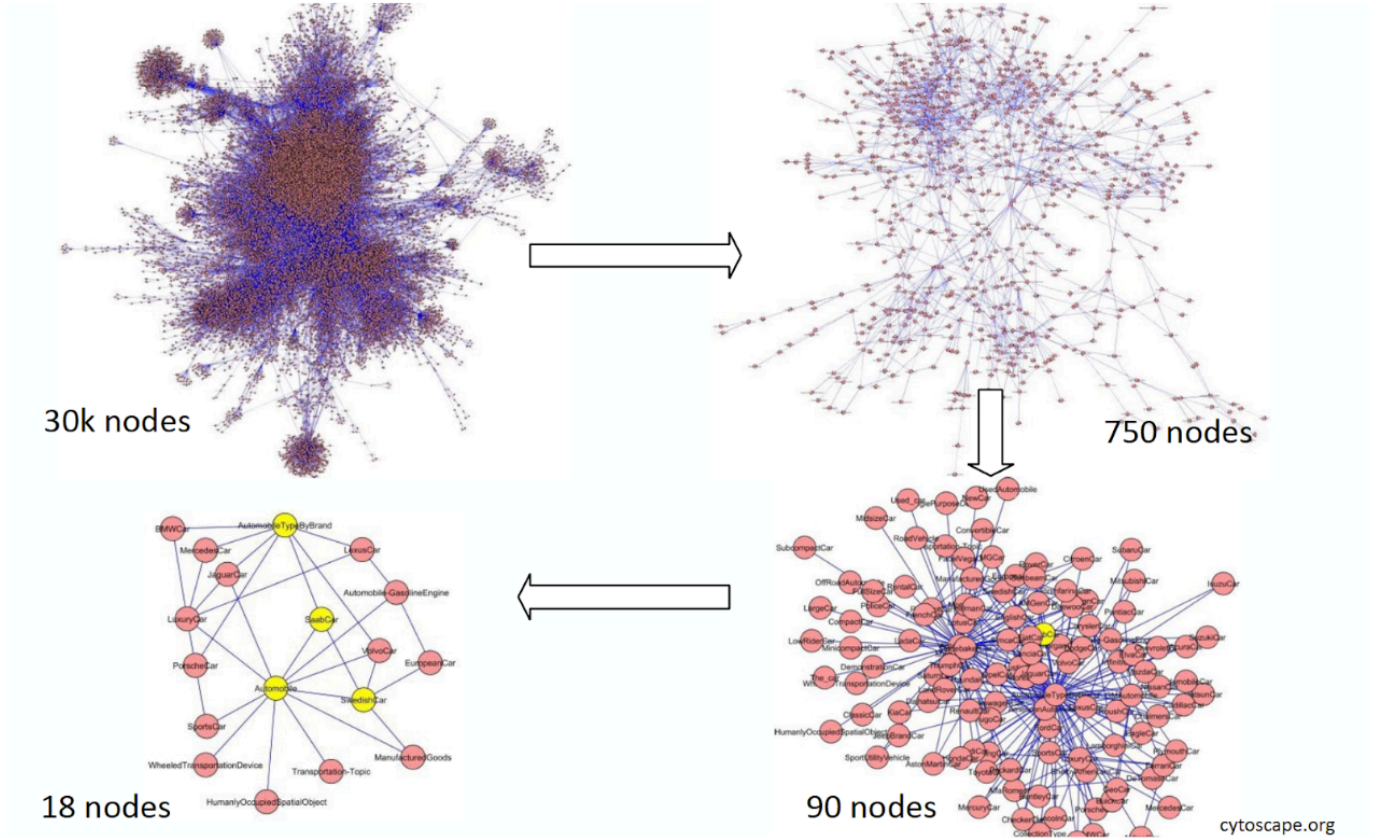


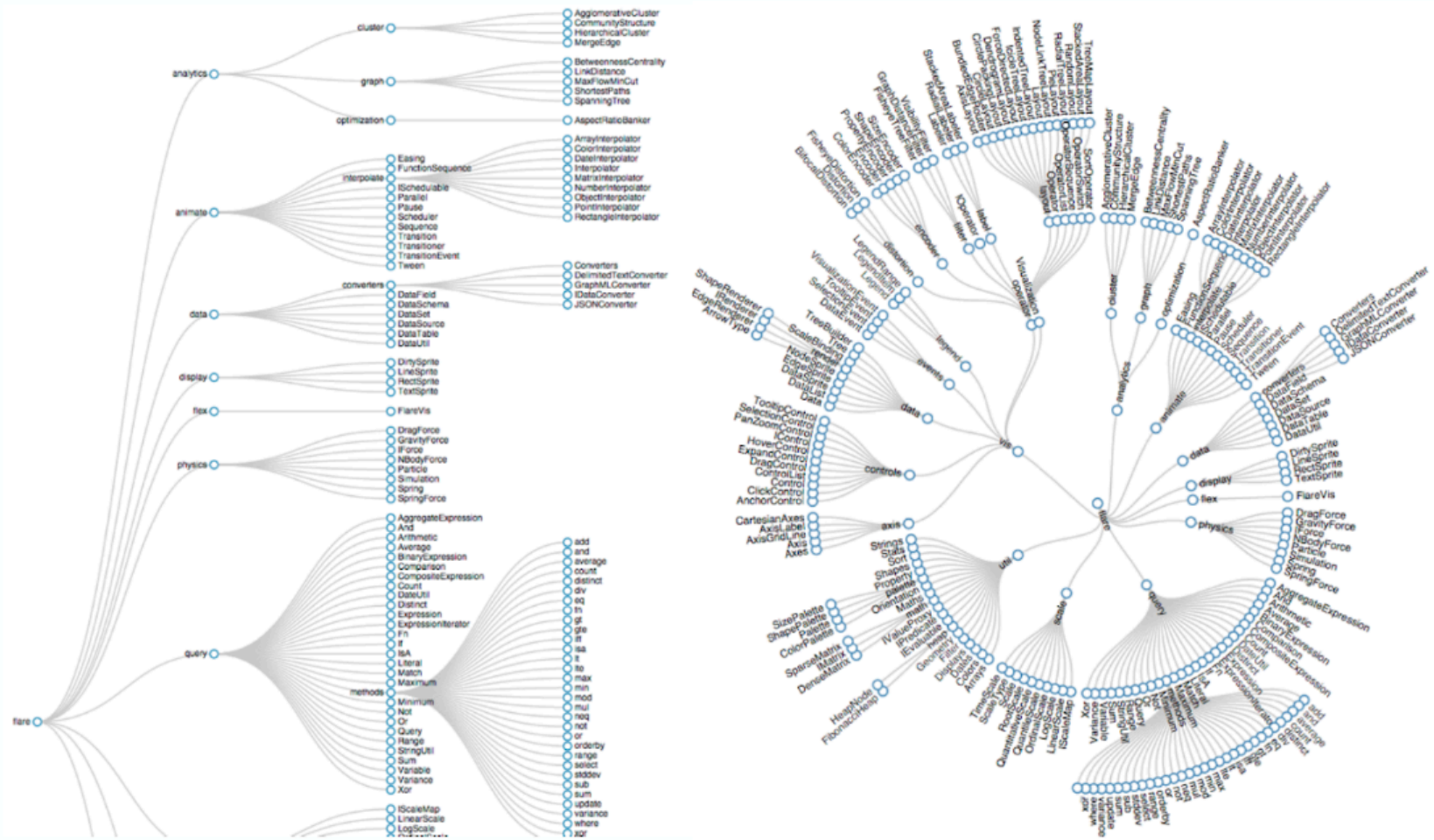
Explicit
(Node-Link)



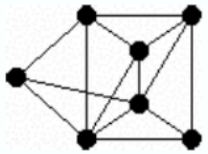
Matrix

Implicit

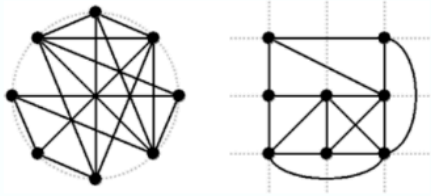

30k nodes
750 nodes
18 nodes
90 nodes
cytoscape.org


flare
analytics
cluster
AgglomerativeCluster
CommunityStructure
HierarchicalCluster
MergeEdge
graph
BetweennessCentrality
LinkDistance
MaxFlowMinCut
ShortestPaths
SpanningTree
optimization
AspectRatioBanker
animate
interpolate
ArrayInterpolator
ColorInterpolator
DateInterpolator
Interpolator
MatrixInterpolator
NumberInterpolator
ObjectInterpolator
PointInterpolator
RectangleInterpolator
Easing
FunctionSequence
ISchedulable
Parallel
Pause
Scheduler
Sequence
Transition
Transitioner
TransitionEvent
Tween
data
converters
Converters
DelimitedTextConverter
GraphMLConverter
IDataConverter
JSONConverter
DataField
DataSchema
DataSet
DataSource
DataTable
DataUtil
display
DirtySprite
LineSprite
RectSprite
TextSprite
flex
FlareVis
physics
DragForce
GravityForce
IForce
NBodyForce
Particle
Simulation
Spring
SpringForce
query
AggregateExpression
And
Arithmetic
Average
BinaryExpression
Comparison
CompositeExpression
Count
DateUtil
Distinct
Expression
ExpressionIterator
Fn
If
IsA
Literal
Match
Maximum
Minimum
Not
Or
Query
Range
StringUtil
Sum
Variable
Variance
Xor
methods
add
and
average
count
distinct
div
eq
fn
gt
gte
if
isa
lt
lte
max
min
mod
mul
neq
not
or
orderby
range
select
stddev
sub
sum
update
variance
where
xor
IScaleMap
LinearScale
LogScale

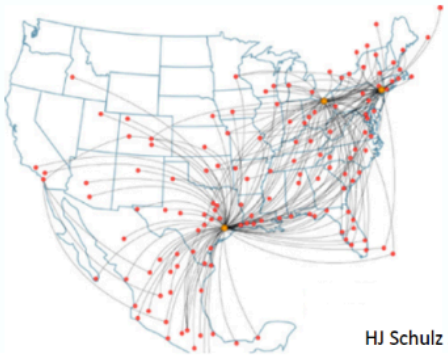
Free



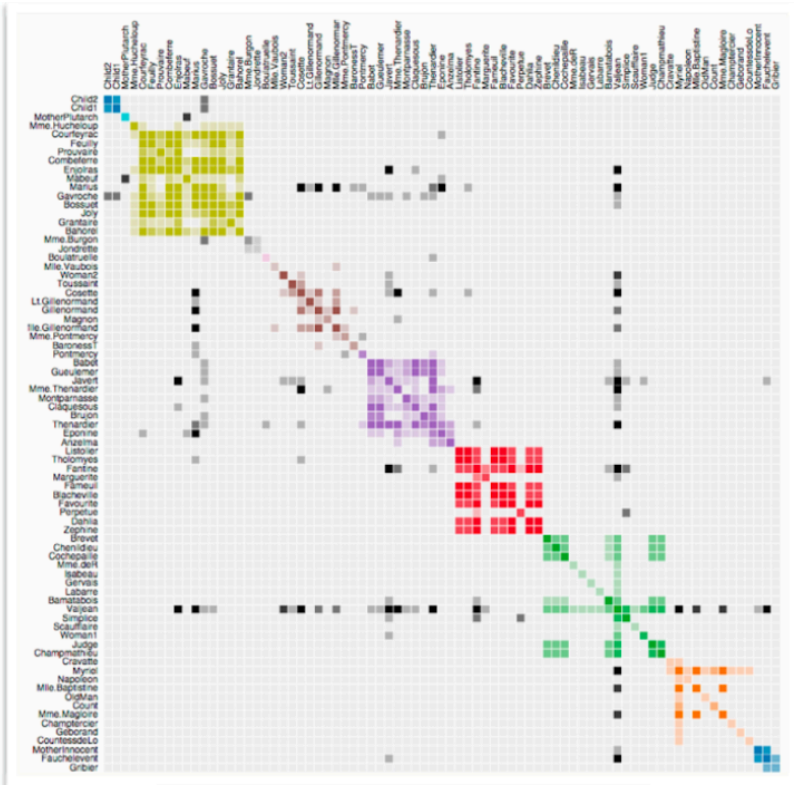
Styled



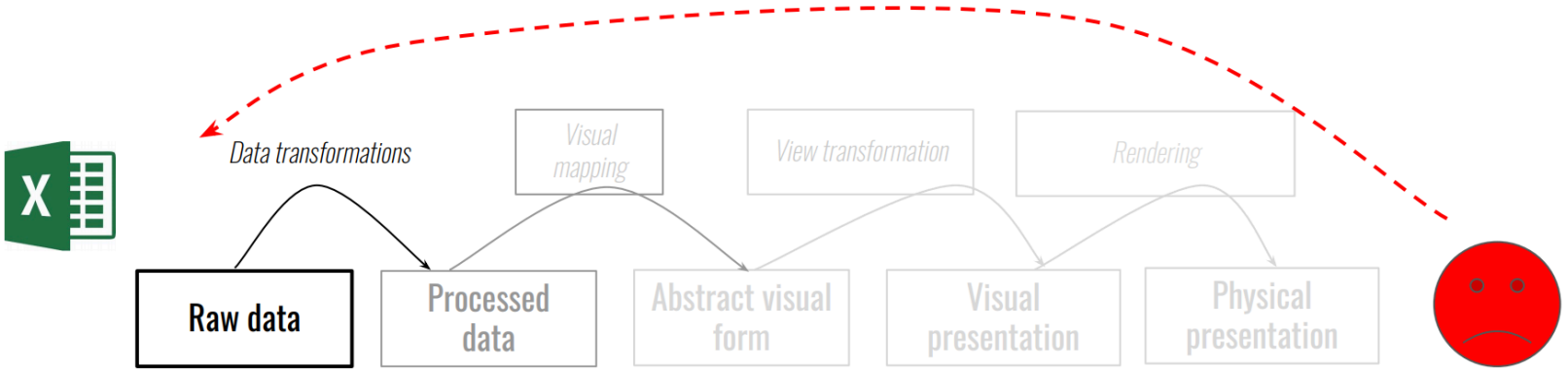
Fixed



HJ Schulz 2006



Processus iterative



Merci