

Welcome all

I am Jérôme Baton (jerome.baton@epita.fr)

I have been retrieving and presenting data for years,
mostly a Java+backend specialist, Android, ...
Speaker

Now :

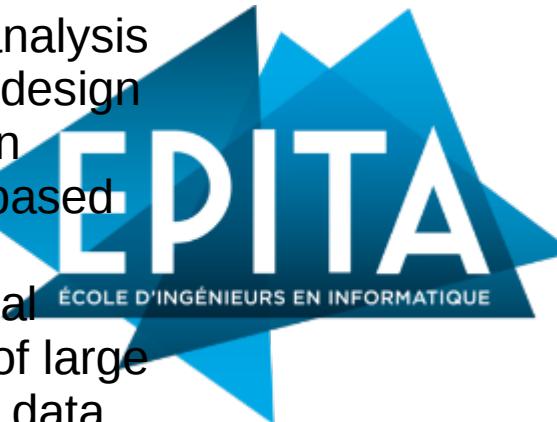
ProudFessor + enterTRainer (several subjects, many places),
Author (« Learning Neo4j 3.x », Packt Publishing, 2017)

Data Visualization

DATAVIZ

The major goals of this course are to understand how visual representations can help in the analysis and understanding of complex data, how to design effective visualizations, and how to create an interactive visualization using modern web-based frameworks.

It explores the underlying theory and practical concepts in creating visual representations of large amounts of data. It covers the core topics in data visualization: data representation, visualization toolkits, scientific visualization, medical visualization, information visualization, flow visualization, and volume rendering techniques



Apologies in advance for my accent

I might pronounce your names badly
or even the lectures

Please bear with my puns and jokes

Today, we'll start with theory, softwares
Other lectures will be about practice, projects and grading

Planning : 16 hours - A packed week

Today 17/2 Theory

18/2 Exercises

19/2 Quizz (40%)

Project BYOD Bring your own data :)
(no duplicates allowed) (60%)

20/2 Presentations

YOU

Are to take notes

As everything is not in the slides

Are not allowed to take
pictures/movies/streams/sound
captations



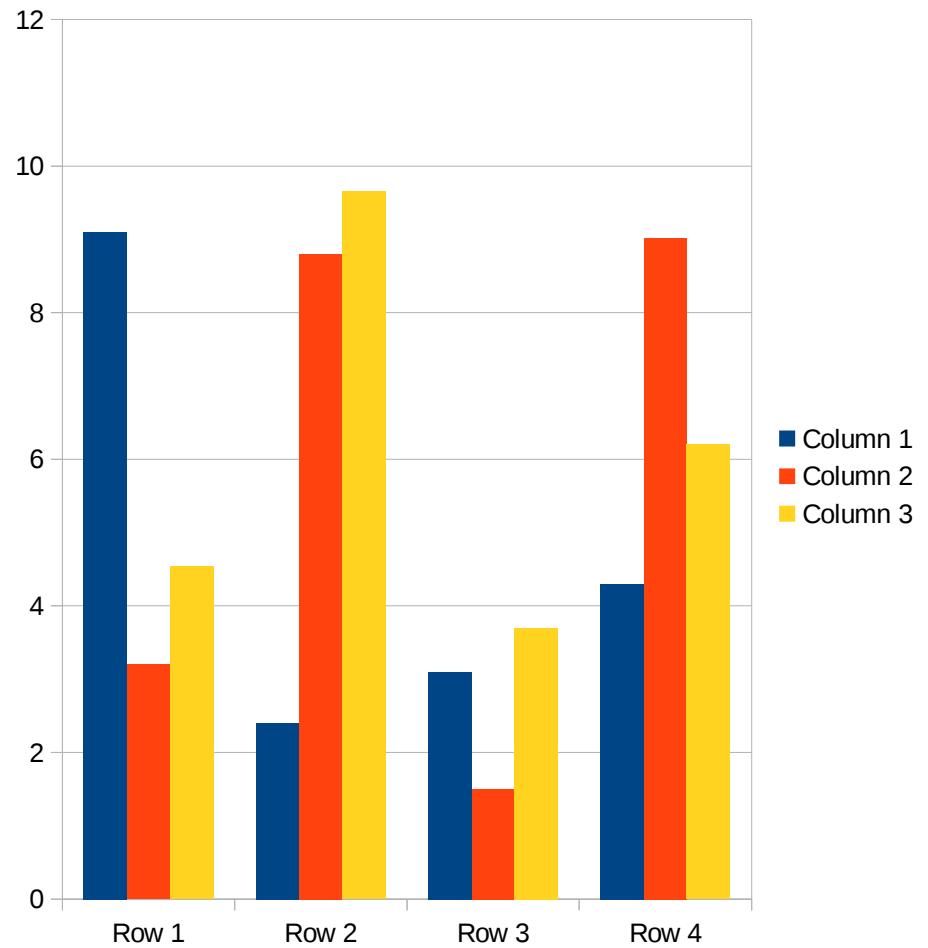
What is dataviz ?

In short data vizualisation

Like this ---->

What for ?

Get a better understanding of the collected data

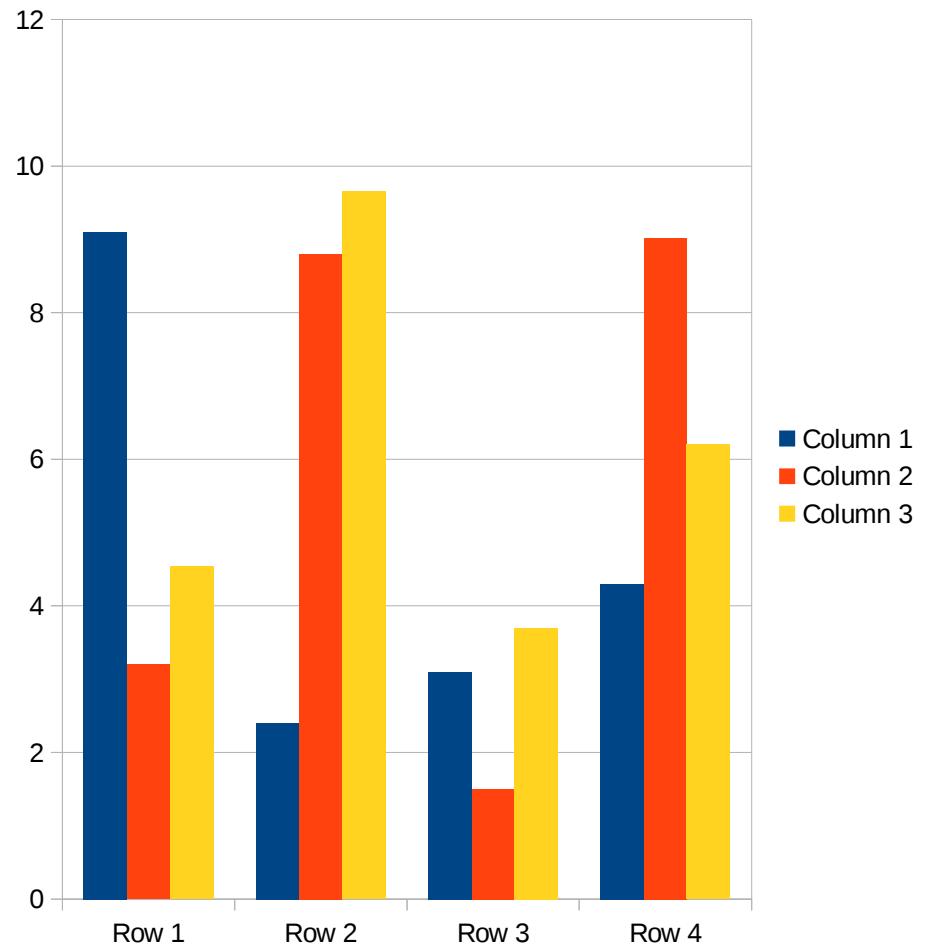


What is dataviz ?

In short data
vizualisation

What for ?

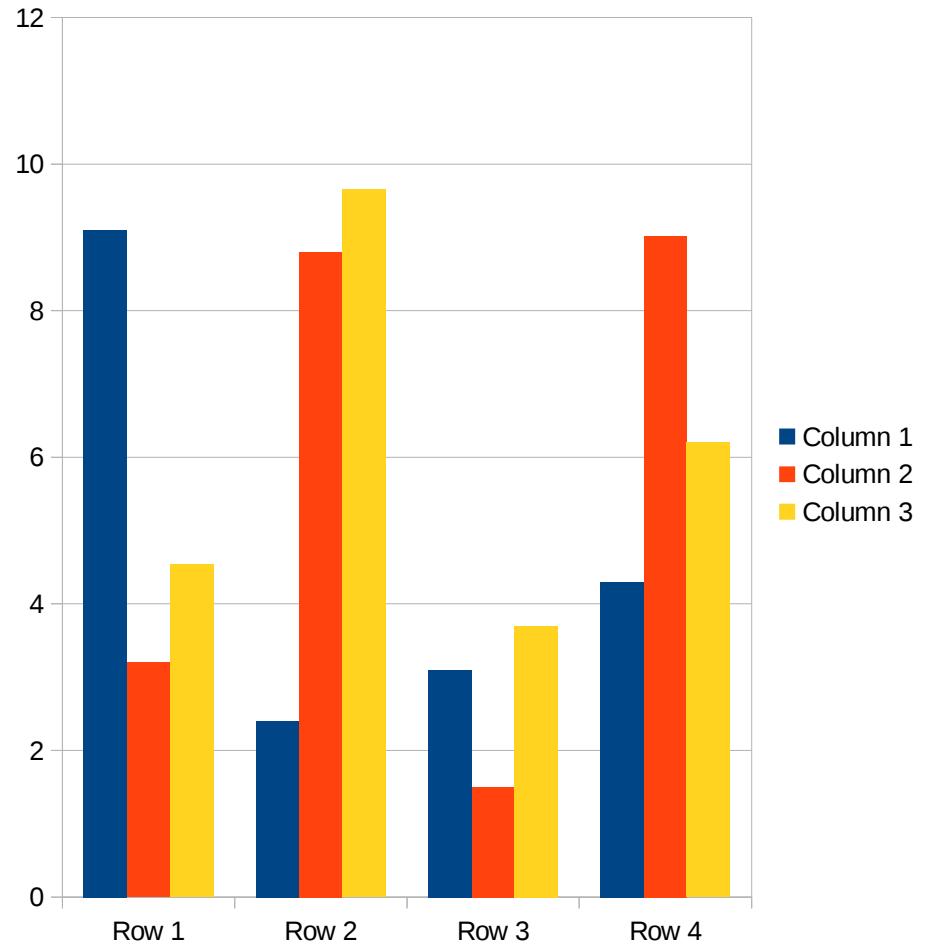
Get a better
understanding of
the collected data



What is dataviz ?

In short data
vizualisation

What for ?
Get a better
understanding of the
collected data



What is this dataviz ?

In short data
vizualisation

Could be
sales data,

could be IOT-generated data

Could be a random pic

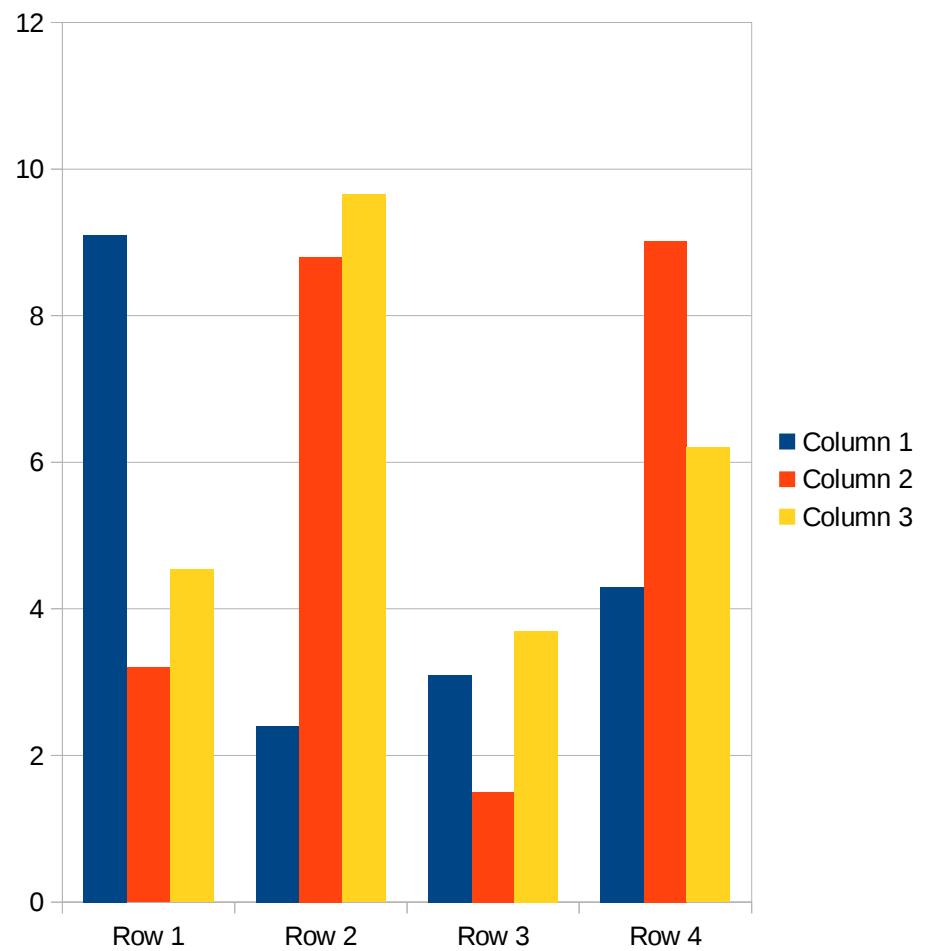
What for ?

Get a better
understanding of the
collected data

What is this dataviz ?

**This could be a chart of
the sales on three
continents in four months**

Or many different things



Out of all of our logical capabilities, we understand things best through the processing of visual information.

When data is represented visually, the probability of understanding complex builds and numbers increases

Mario Döbler and Tim Großmann

Visualizing data has many advantages, such as the following:

- **Complex data can be easily understood**
- **A simple visual representation of outliers, target audiences, and future markets can be created**
- **Storytelling can be done using dashboards and animations**
- **Data can be explored through interactive visualizations**

**One could represent many different data
With the same chart**

**Beware not to use the same chart again and
again**

The power of charts

DEMO

This ...

**Value
4 M**

Is different of



The first conveys an information

The second conveys an information and eases interpretation

KEY SENTENCE

If there is one thing you should remember from this lecture in 20 years, it is this :

KEY SENTENCE

If there is one thing you should remember from this lecture in 20 years, it is this :

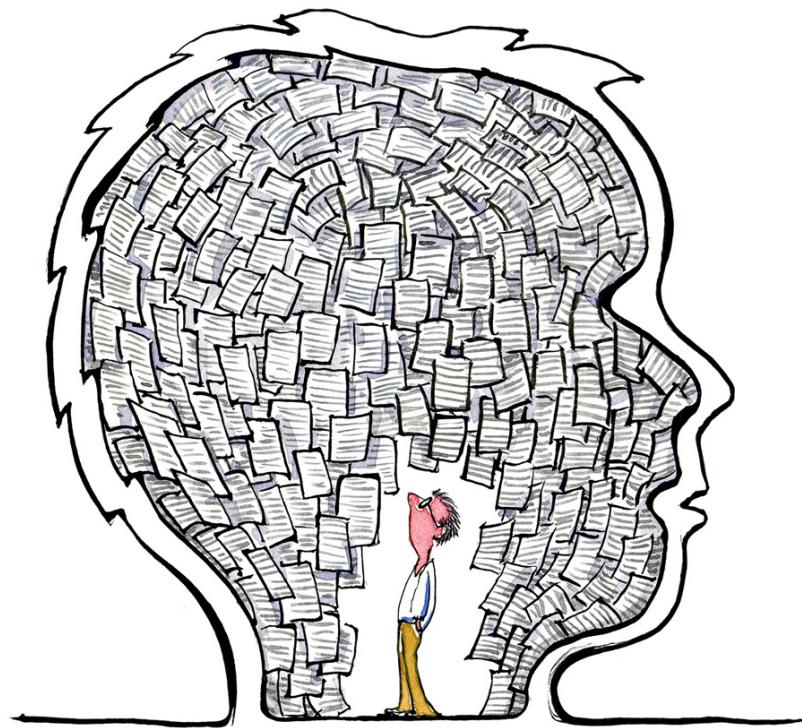
A PICTURE IS WORTH A THOUSAND WORDS !

(sorry I screamt)

, about, above, across, act, active, activity, add, adder, after, again, ago, ago, all, all, alone, along, already, always, am, amou
n, and, angry, another, answer, any, anyone, anything, anytime, appear, apple, are, area, arm, army, around, arrive, art, as, ask, at,
attack, aunt, autumn, away, baby, base, back, bad, bag, ball, bank, basket, bath, be, bean, bear, beautiful, beer, bed, bedroom, behave,
before, begin, behind, bell, below, besides, best, better, between, big, bird, birth, birthday, bit, bite, black, bleed, block, blood, blow, blue
board, boat, body, boil, bone, book, border, born, borrow, both, bottle, bottom, bowl, box, boy, branch, brave, bread, break, breakfast,
reathe, bridge, bright, bring, brother, brown, brush, build, burn, business, bus, busy, but, buy, by, cake, call, can, candle, cap, car, car
are, careful, careless, carry, case, cat, catch, central, century, certain, chair, chance, change, chase, cheap, cheese, chicken, child,
children, chocolate, choice, choose, circle, city, class, clever, clean, clear, climb, clock, cloth, clothes, cloud, cloudy, close, coffee,
oat, coin, cold, collect, colour, comb, comfortable, common, compare, come, complete, computer, condition, continue, control, cook,
ool, copper, corn, corner, correct, cost, contain, count, country, course, cover, crash, cross, cry, cup, cupboard, cut, dance,
angerous, dark, daughter, day, dead, decide, decrease, deep, deer, depend, desk, destroy, develop, die, different, difficult, dinner,
irection, dirty, discover, dish, direction, do, dog, door, double, down, draw, dream, dress, drink, drive, drop, dry, duck, dust, duty, each
ar, early, earn, earth, east, easy, eat, education, effect, egg, eight, either, electric, elephant, else, empty, end, enemy, enjoy, enough,
nter, equal, entrance, escape, even, evening, event, ever, every, everyone, exact, everybody, examination, example, except, excited,
xercise, expect, expensive, explain, extremely, eye, face, fact, fail, fall, false, family, famous, far, farm, father, fast, fat, fault, fear,
bed, feel, female, fever, few, fight, fill, film, find, fine, finger, finish, fire, first, fit, five, fix, flag, flat, float, floor, flour, flower, fly, fold,
ool, foot, football, for, force, foreign, forest, forget, forgive, fork, form, fox, four, free, freedom, freeze, fresh, friend, friendly, from, front
uit, full, fun, funny, furniture, further, future, game, garden, gate, general, gentleman, get, gift, give, glad, glass, go, goat, god, gold,
ood, goodbye, grandfather, grandmother, grass, grave, great, green, grey, ground, group, grow, gun, hair, half, hall, hammer, hand,
appen, happy, hard, hat, hate, have, he, head, healthy, hear, heavy, hello, help, heart, heaven, height, help, hen, her, here, hers, hide,
igh, hill, him, his, hit, hobby, hold, hole, holiday, home, hope, horse, hospital, hot, hotel, house, how, hundred, hungry, hour, hurry,
usband, hurt, I, ice, idea, if, important, increase, inside, into, introduce, invent, iron, invite, is, island, it, its, jelly, job, join, juice,
imp, just, keep, key, kill, kind, king, knife, knee, knife, know, ladder, lady, lamp, land, large, last, late, lately, laugh, lazy, lead
leaf, learn, leave, leg, left, lend, length, less, lie, like, little, library, life, light, like, lip, lot, listen, little, live, lock, lonely, long
look, lose, lot, love, low, lower, luck, machine, main, make, male, man, many, map, mark, market, marry, matter, may, me, meal, mean
measure, meat, medicine, men, mental, mention, method, middle, milk, million, mind, minute, miss, mistake, mix, model, modern,
moment, money, monkey, moon, moon, morning, miss, mother, mountain, much, move, much, music, must, my, name, narrow,
ation, nature, near, nearly, neck, need, needle, neighbour, out, out, net, news, new, news, newspaper, nice, night, nine, no, noble
oise, none, nor, north, nose, not, nothing, notice, now, number, obey, object, ocean, of, off, offer, office, often, oil, old, on, one, only,
pen, opposite, or, orange, order, other, our, out, outside, over, own, page, pain, paint, pair, pan, paper, parent, park, part, partner, part
ass, past, path, pay, peace, pen, pencil, people, pepper, per, perfect, period, person, petrol, photograph, piano, pick, picture, piece, pig
in, pink, place, plane, plant, plastic, plate, play, please, pleased, plenty, pocket, point, poison, police, polite, pool, poor, popular,
osition, possible, potato, pour, power, present, press, pretty, prevent, price, prince, prison, private, prize, probably, problem, produce,
romise, proper, protect, provide, public, pull, punish, pupil, push, put, queen, question, quick, quiet, quite, radio, rain, rainy, raise, reac
ead, ready, real, really, receive, record, red, remember, remind, remove, rent, repair, repeat, reply, report, rest, restaurant, result,
eturn, rice, rich, ride, right, ring, rise, road, rob, rock, room, round, rubber, rude, rule, ruler, run, rush, sad, safe, sail, salt, same, sand,
ave, say, school, science, scissors, search, seat, second, see, seem, sell, send, sentence, serve, seven, several, sex, shade, shadow,
hake, shape, share, sharp, she, sheep, sheet, shelf, shine, ship, shirt, shoe, shoot, shop, short, should, shoulder, shout, show, sick,
ide, signal, silence, silly, silver, similar, simple, single, since, sing, sink, sister, sit, six, size, skill, skin, skirt, sky, sleep, slip, slow,
moke, small, smell, smile, smoke, snow, so, soap, sock, soft, some, someone, something, sometimes, son, soon, sorry, sound, soupe,
outh, space, speak, special, speed, spell, spend, spoon, sport, spread, spring, square, stamp, stand, star, start, station, stay, steal,
team, step, still, stomach, stone, stop, store, storm, story, strange, street, strong, structure, student, study, stupid, subject, substance,
uccessful, such, sudden, sugar, suitable, summer, sun, sunny, support, sure, surprise, sweet, swim, sword, table, take, talk, tall, taste,
taxi, tea, teach, team, tear, telephone, television, tell, ten, tennis, terrible, test, than, that, the, their, then, there, therefore, these, thick,
thin, thing, think, third, this, though, threat, three, tidy, tie, title, to, today, toe, together, tomorrow, tonight, too, tool, tooth, top, total,
ouch, town, train, tram, travel, tree, trouble, true, trust, twice, try, turn, type, uncle, under, understand, unit, until, up, use, useful, usual
usually, vegetable, very, village, voice, visit, wait, wake, walk, want, warm, wash, waste, watch, water, way, we, weak, wear, weather,
redding, week, weight, welcome, well, west, wet, what, wheel, when, where, which, while, white, who, why, wide, wife, wild, will, win,
wind, window, wine, winter, wire, wise, wish, with, without, woman, wonder, word, work, world, worry, worst, write, wrong, year, yes,
esterday, yet.

A Picture Is Worth A Thousand Words

Is it ?



Can a picture say more
than a thousand words?

HikingArtist.com

1000 words =?

Two pages when the font type is Arial, the font size is 12pt, and it has single spacing.

Four pages when the font type is Times New Roman,
the font size is 12pt, and it has double spacing.
These are the common requirements for both
high school and college writing.

Today's DataViz is graphical

**It has evolved with technical capacities
and its audience**

From this

A	B	C	D	E
1	Principal	\$50,000		
2	Rate	13.0%		
3	Years	5		
4	Payment	1,137.65		
5				
6	Year	Begin Bal.	End Bal.	Total Paid
7	1	50,000.00	42,406.26	13,651.84
8	2	42,406.26	33,764.33	13,651.84
9	3	33,764.33	23,929.53	13,651.84
10	4	23,929.53	12,737.22	13,651.84
11	5	12,737.22	0.00	13,651.84
12				914.63
13				
14				
15				
16				
17				
18				
19				
20				

This is high-end software in the 80's

(back in the days) Dataviz was for company boards

It used to be reports

Litterally piles of paper

Then reports with charts

Still paper piles

Or just charts in a presentation

(also invention of 'death by powerpoint')

Dashboards

On demand

Now : all this plus animated charts, 3D

Tomorrow charts in VR ?

We are viewers, not readers anymore

Lets not compare time spent reading books now and before Mainstream TV happened.

We want information quick

The Hurry Generations

Because of living with remote controls.

Attention span : size matters

From

From an ocean of data (Internet)

Or data lakes :) (company zone)

(back in the days) Dataviz was for company boards

It used to be reports

Litterally piles of paper

Then reports with charts

Still paper piles

Or just charts in a presentation

(also invention of 'death by powerpoint')

Dashboards

On demand

Now : all this plus animated charts, 3D

Tomorrow charts in VR ?

From https://en.wikipedia.org/wiki/Data_lake

A data lake is a system or repository of data stored in its natural/raw format, [1] usually object blobs or files. A data lake is usually a single store of all enterprise data including raw copies of source system data and transformed data used for tasks such as reporting, visualization, advanced analytics and machine learning. A data lake can include structured data from relational databases (rows and columns), semi-structured data (CSV, logs, XML, JSON), unstructured data (emails, documents, PDFs) and binary data (images, audio, video). [2]

A data swamp is a deteriorated and unmanaged data lake that is either inaccessible to its intended users or is providing little value.[3]

To which I'll append 'data dump'

Dataviz conveys a message

If you produce a chart

You want it to illustrate a point

However its difficult to get understood

DataViz must be adapted to the audience

The how

Is audience reading LTR or RTL ?

Some colors convey a different meaning in different cultures !

The what

Dont show board's salaries to a company salarymen

Dataviz must be adapted to the era

Dont represent booming business numbers with explosions.

Avoid anything that can be stamped as not politically correct

However ...

Some rules stay valid

Show ordered data

Use a legend

Create a title

Keep user's attention

The RED color

Red – Fire: Red is an auspicious as well as a popular color in Chinese culture. It symbolizes luck, happiness and joy. It also represents summer season and the Southern Direction. It is the color worn by brides, since it is believed to be an auspicious color for warding off evil. Red also represents vitality, celebration and fertility in traditional Chinese color symbolism.

Today, westerners take red to be a representation of China's Revolution or political system.

However, this is definitely not the case since ancient Chinese roots certainly did not factor in the revolution for it. If you are doing business in China, make use of bit of red in marketing, but not too much.

I like red too – Jérôme, France (hence the slides)

Red is the enemy, USSR – says every world map representation in the 80s

Colors meanings (west world only ?)



Colors

So, choose wisely !

A beautiful palette

is not friendly to the color blind

BIG ISSUE

Not easy to satisfy everybody when you have a global audience (web site).

Because people tend to see things.

Things they love

Things they hate

Not mandatorily what you show, but what they interpret

Communication.

Example : this course

What I know

What I thought

What I wrote

What you read

What you take away

All these can be different from one another

Interpretation

Communication is difficult

**Charts are a way to communicate,
to give an information disguised as a scientific
proof**

ANTIPATTERN : ONE DATAVIZ TO RULE THEM ALL !

Predicate :



ONE DATAVIZ CAN EXPRESS ANY DATA !

It is Camembert viz (Eng : pie chart)

Or so seem to think many, many journalists.

FYI : Camembert is from Normandy

Real camembert is made from unpasteurized milk

ONE DATAVIZ TO RULE THEM ALL !

Predicate :

ONE DATAVIZ CAN EXPRESS ANY DATA !

It is Camembert viz (Eng : pie chart) !!

Experience replies :

No, not at all



Same data

Can be represented in many ways

So many chart types to choose from

Same data

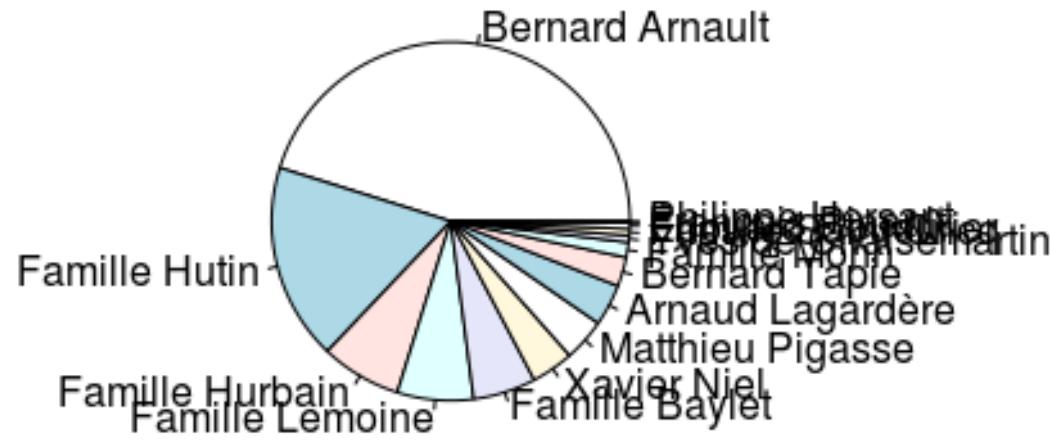
Can be represented in many ways

So many chart types to choose from

The role of charts is to make data show a meaning
within a glance

Pie - Most known & badly used

We will see later what this data is



Your comments ?

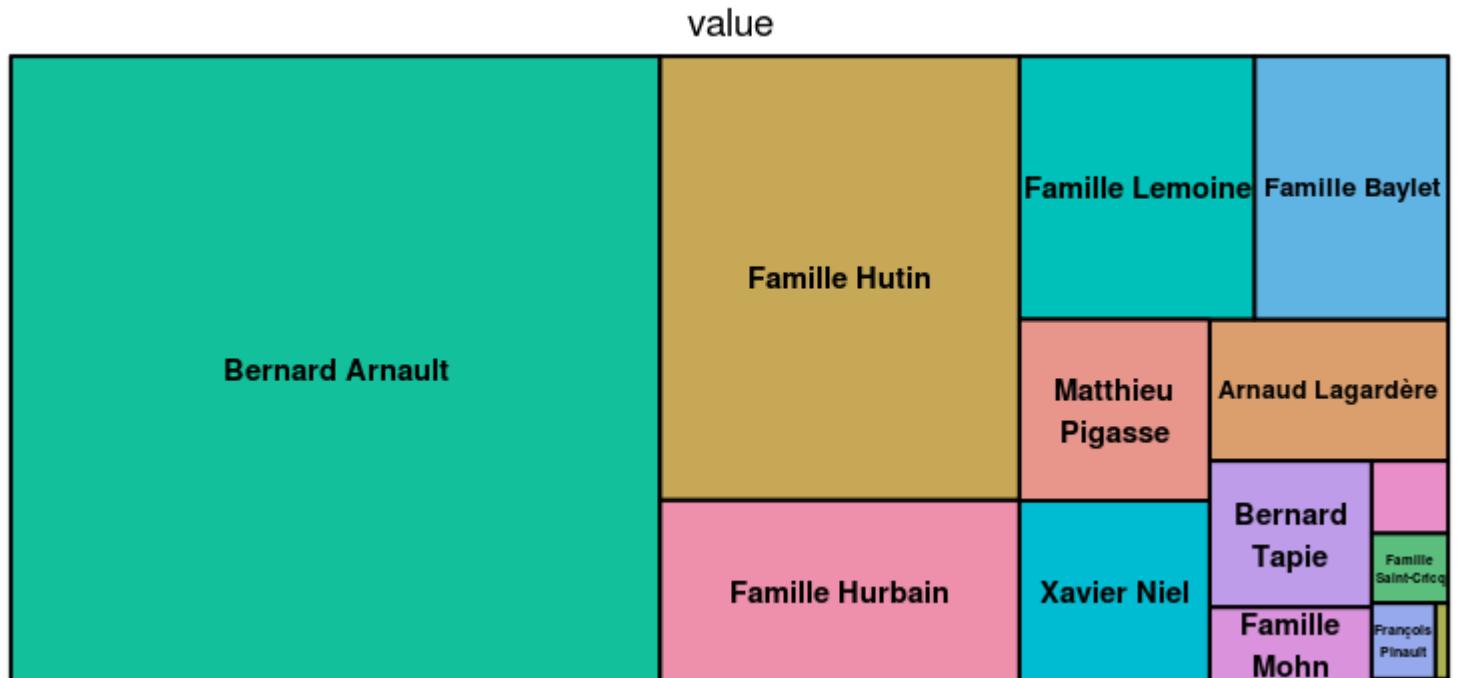
Source : https://github.com/wadael/presse_fr

Did I hear ‘unreadable’ ?

Indeed, unreadable

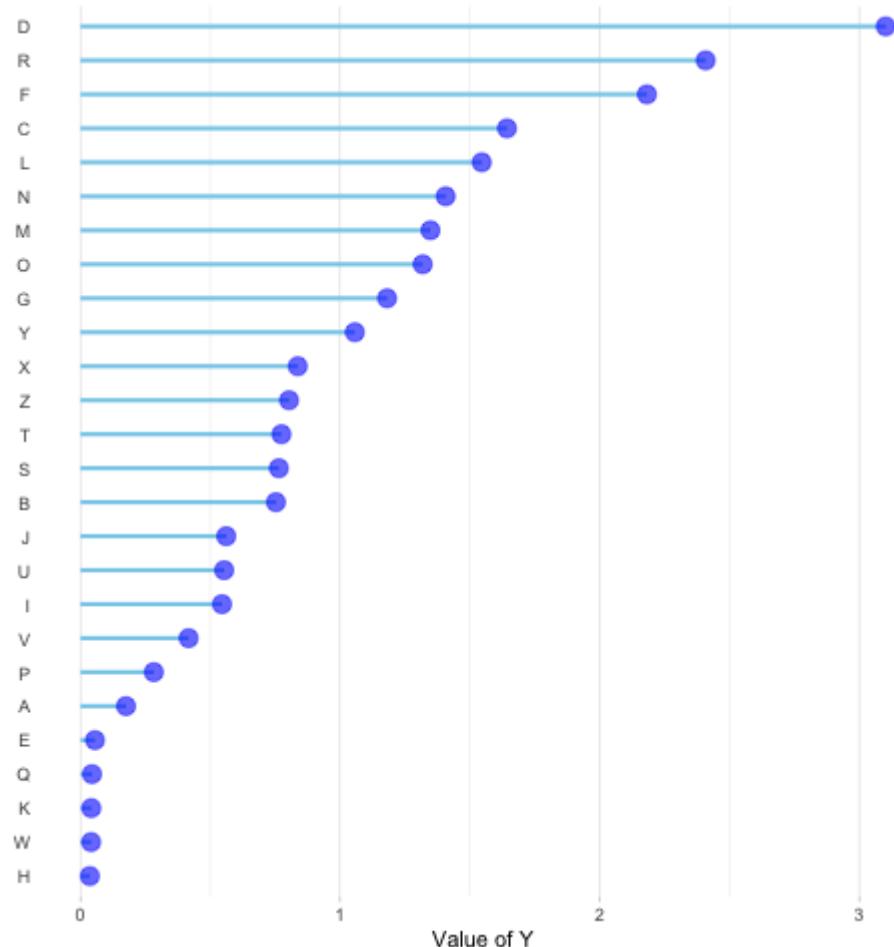
5 points for Gryffindor

This chart is better (treemap)



Lollipop

**Gives a better insight
Than pie**



! Do not forget to sort !

Plenty of fish in the ocean

To find the good dataviz, you need to select a few factors

*** nature of information (numeric, map ...)**

*** how many parameters is there ?**

1 : business per year

2 : business per year per business line

3 : business per year per business line per continent

*** how many dimensions ?**

/!\ Color scheme

Is all data shown ?

Charts are used to fool people

They are displayed as proofs of anything

And just like numbers can say anything and its contrary, charts are the same

How to fool people

Scales

Sometimes scale is given and not respected

Linear and logarithmic produce different results

No origin in the chart

A scale must start at zero and be linear !

Because this is implied for most readers

Units

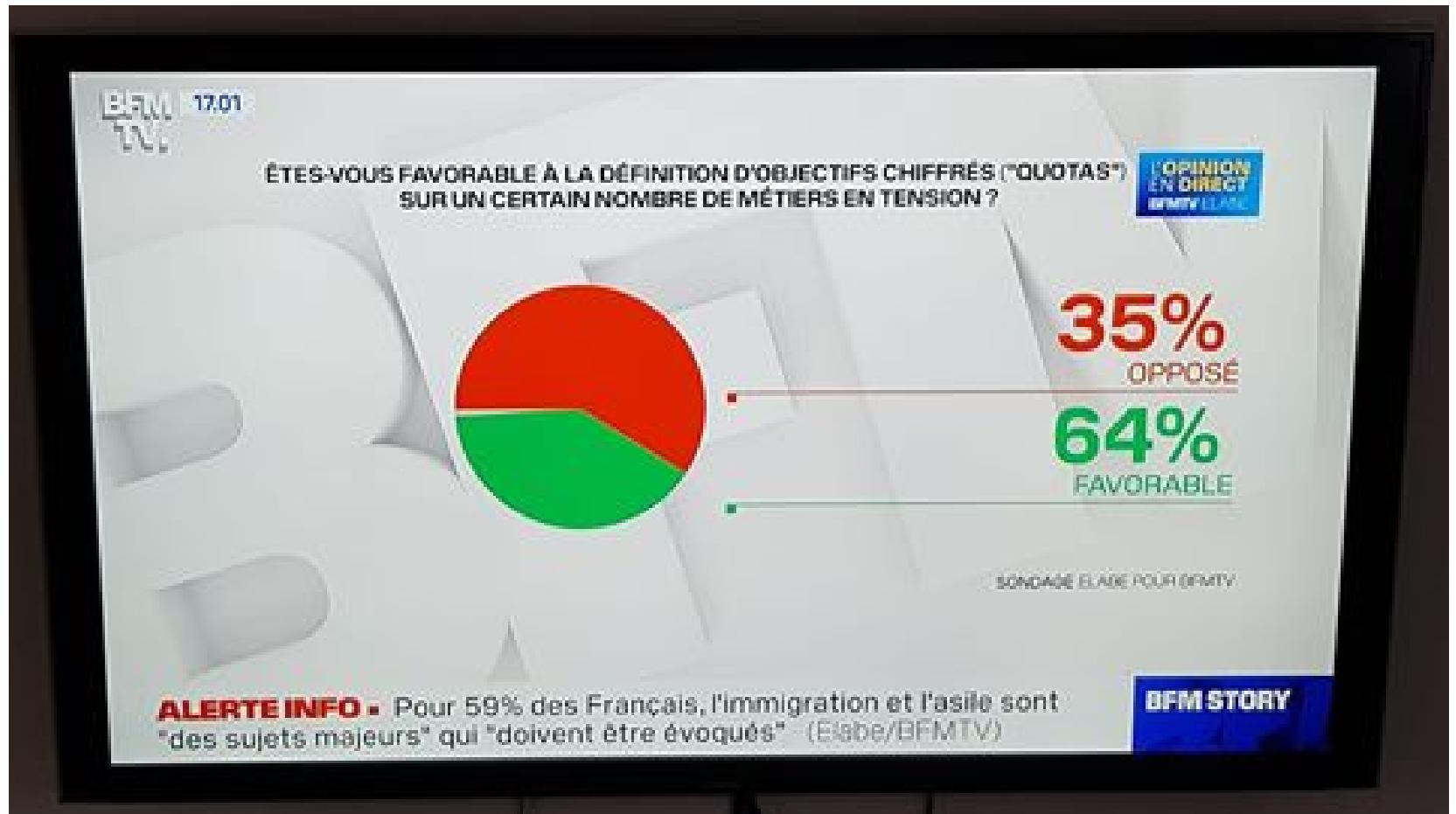
Different units on the same axis

Examples of this

On every election day in France, I see lying charts

Like 40 % being more than the half of a half-disc representing the new chamber of Parliament repartition.

With the % written on it



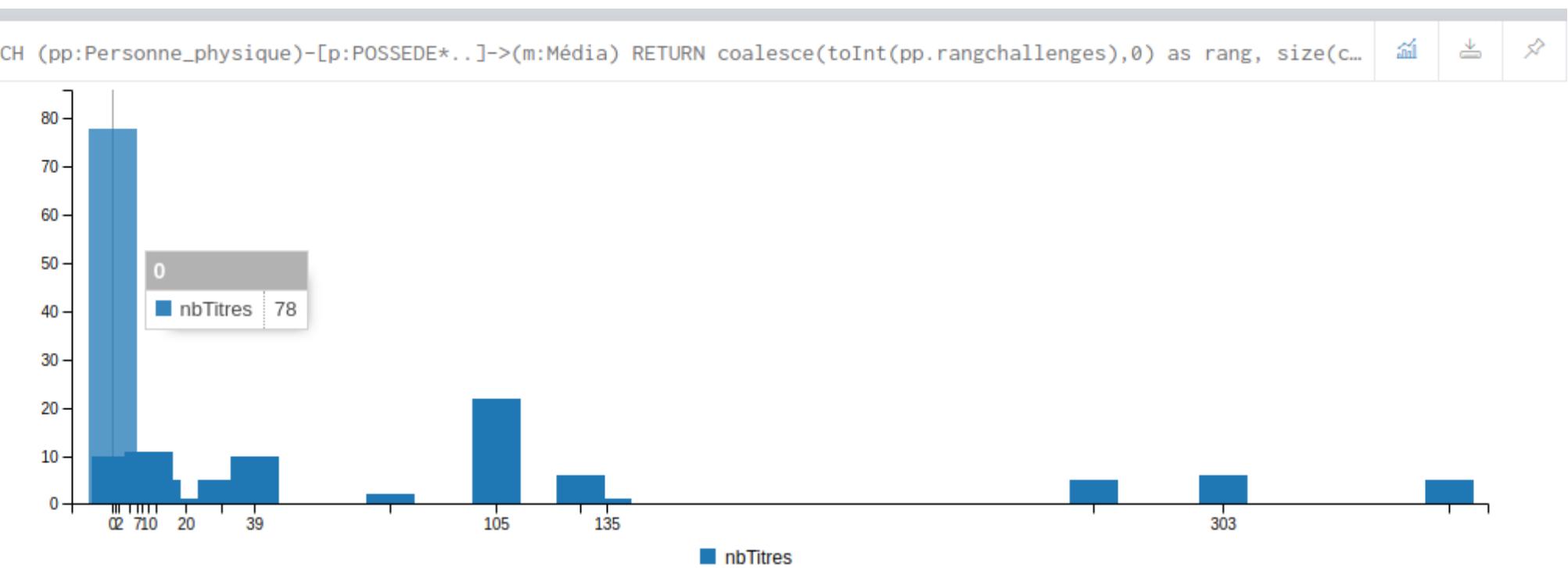
Local FoxNews at it again !

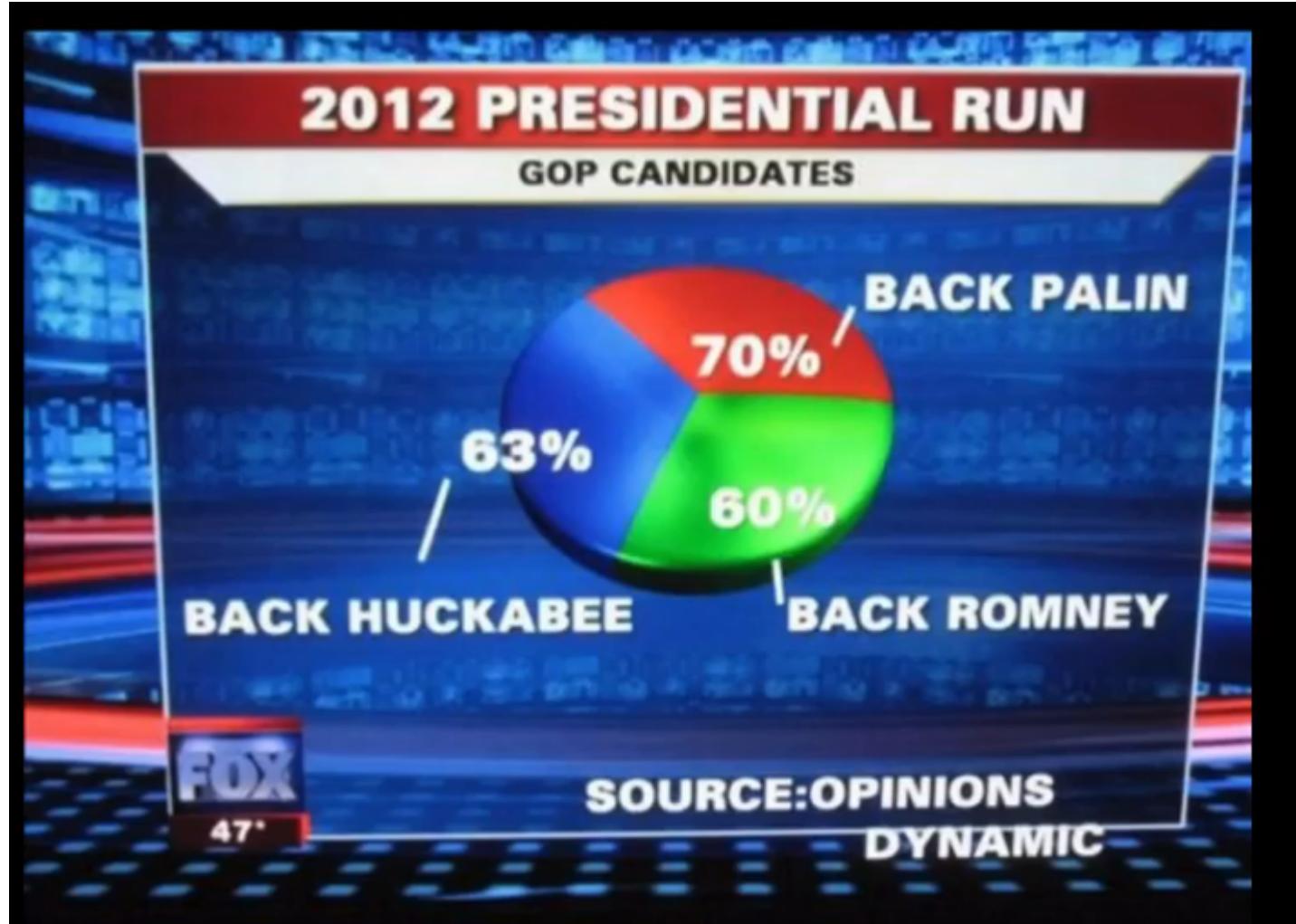


Clearly 48 % is more than a half :)

**Also, color choice is never innocent
(green is support government)**

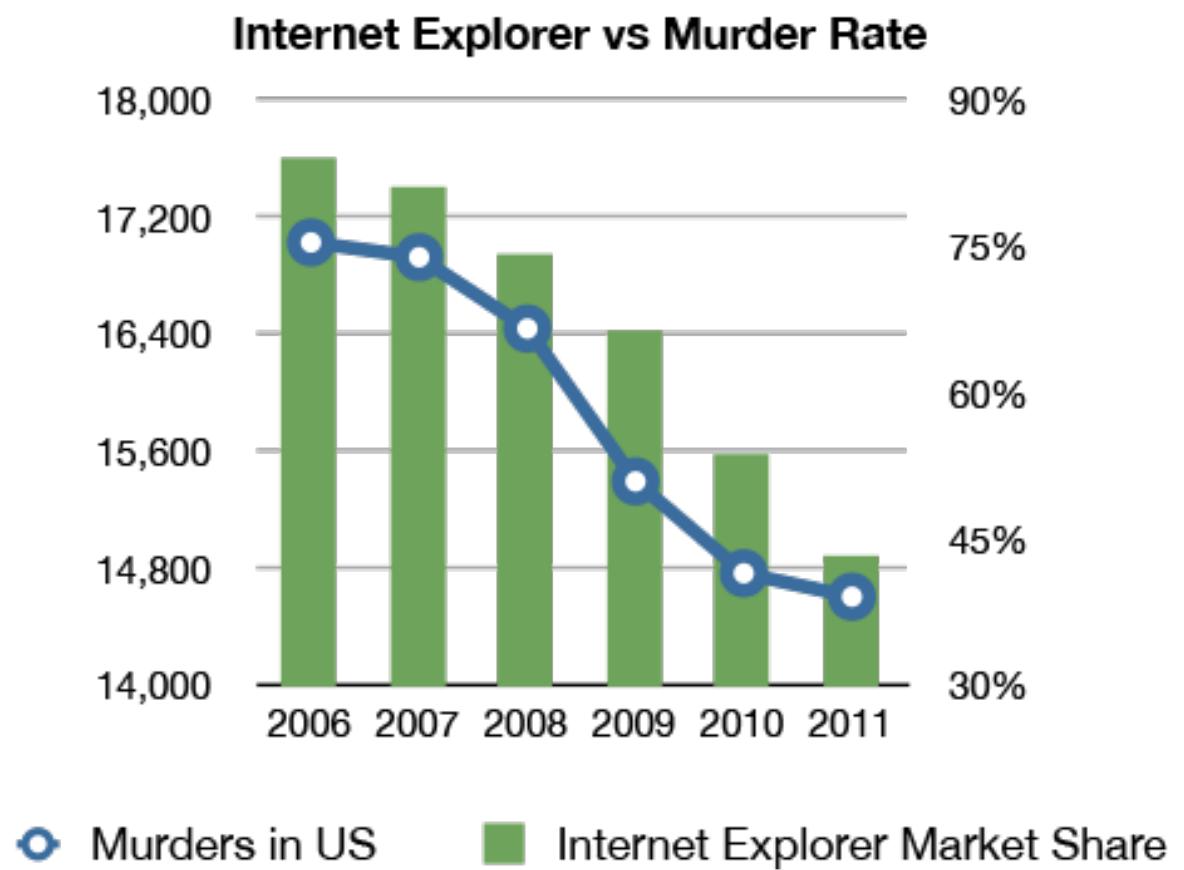
One for which I can only blame myself



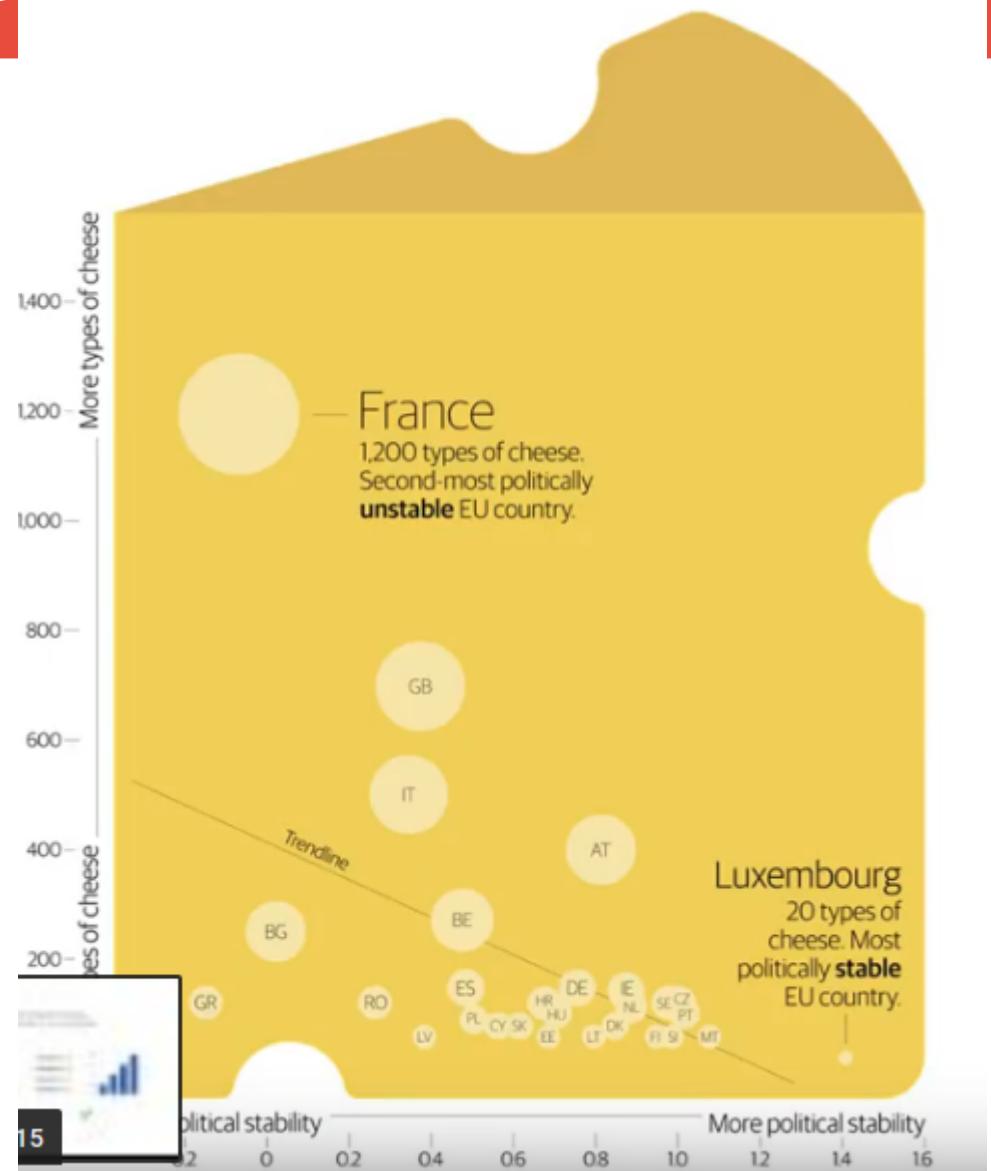


One more faulty chart

Numbers can be used to prove anything



Not bad, just too rich



Lets have a new look to the first slide.

It was BAAAD

find why

Data Visualization

DATAVIZ

The major goals of this course are to understand how visual representations can help in the analysis and understanding of complex data, how to design effective visualizations, and how to create an interactive visualization using modern web-based frameworks.

It explores the underlying theory and practical concepts in creating visual representations of large amounts of data. It covers the core topics in data visualization: data representation, visualization toolkits, scientific visualization, medical visualization, information visualization, flow visualization, and volume rendering techniques



- **Title repetition**
- **Unreadable block of text**
 - Not easy to read on screen
 - No spacing
 - Same term repeated ad
 - leads to → no idea what was on the slide as soon as slide changes
- **Logo in an unexpected place**
 - Top Left is more common

That was just a warning

Presenting efficiently is another skill

**However some items are shared between
charting and powerpointing.**

(this is outside of course' scope)

Rules of thumb

A few



Overloading

DO NOT OVERLOAD the chart

Ask yourself

« Do you really think this needs 5 lines ? »

« Where is my area ? » (hidden)

« how to read a bubble-filled bubble chart ? »

« should I compare life expectancy in the last 50 years for all the known countries in the same chart ? »

Organize logically

Choose how you sort data.

Do not go with defaults



Help comparisons

Why two line charts looking alike if you can draw two lines in one chart ?



No overplotting

Keep it readable



Use another pair of eyes

**Ask someone who did not work on your viz
how (s)he understands your viz**

Plenty of fish in the ocean

Some people on the Internet have made a terrific work in classifying the existing data visualizations.

They made the flowchart to decide which chart to use.

Lets have a look at a

<https://datavizcatalogue.com/>

<https://www.data-to-viz.com/>

<https://www.r-graph-gallery.com/all-graphs/>

Credits

http://pre15.deviantart.net/f493/th/pre/i/2012/142/1/9/a_picture_is_worth_a_thousand_words_by_shadowxomega5-d50oo8p.png

<https://hikingartist.files.wordpress.com/2012/07/a-picture-is-worth-a-thousand-words-3.jpg>

<https://www.outsidethebeltway.com/farewell-to-lotus-1-2-3/>

<https://mathix.org/linux/archives/12851>

https://assets.lls.fr/pages/7363852/HIS.2de.AP.graphique_greve_.jpg

Recommended book

Data Visualization with Python

Mario Döbler and Tim Großmann - Packt Publishing

Exercises

List the steps to create the best chart possible

Get DASH

Install

Get sample project

Adapt it (or get Sakila DB)

Lets pick a few dataviz and for each one, lets find

A case when it is perfect (explain why)

A case when it is a bad choice (explain differently)

2 dataviz per student, please synch yourselves

Will be shared with group later