Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2020/2021

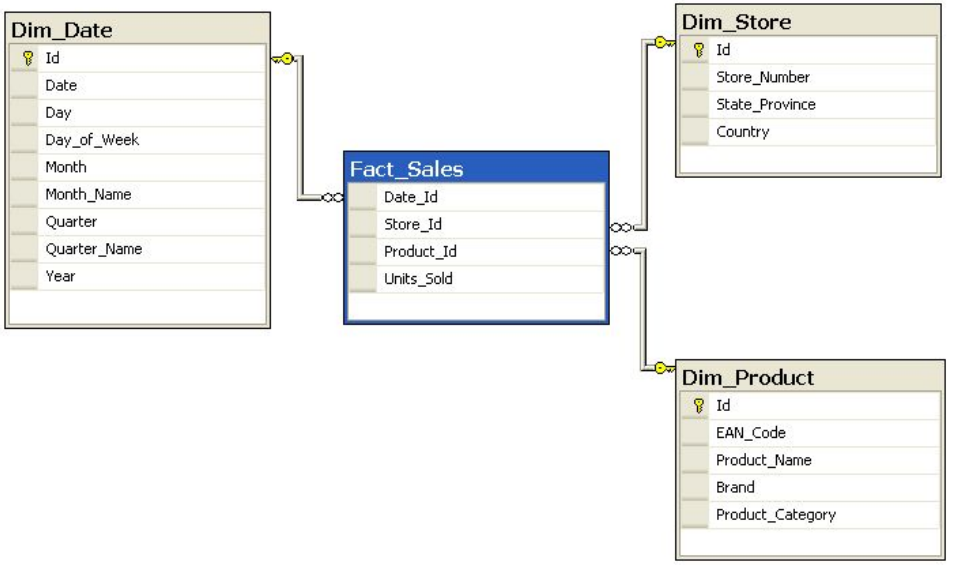
FACULTY DIEGO PAJARITO

Star schema

Describing objects through multiple dimensions

One object multiple dimensions

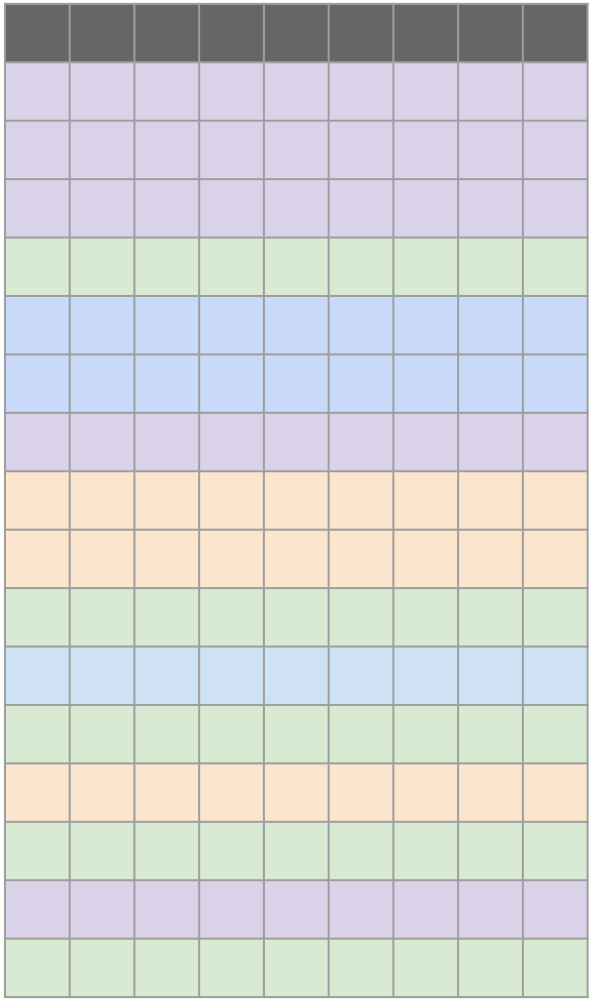
This is the simplest style of a data mart schema.
The approach is also used to develop data warehouses and work with multi dimensional data.



Source: https://en.wikipedia.org/wiki/Star_schema

Aggregation

Reducing rows and columns in a meaningful way

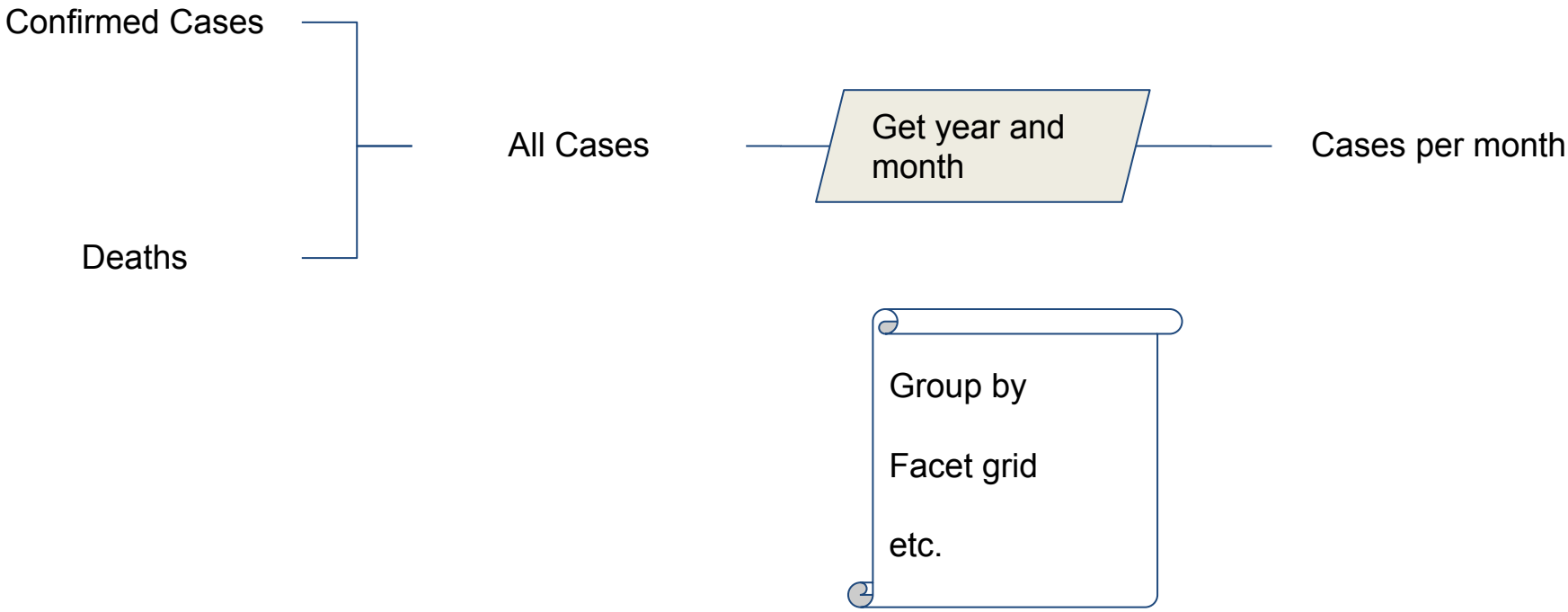


Group by and functions

group	f(x)_1	f(x)_2	...	f(x)_n

mean():	*Compute mean of groups*
sum():	*Compute sum of group values*
size():	*Compute group sizes*
count():	*Compute count of group*
std():	*Standard deviation of groups*
var():	*Compute variance of groups*
sem():	*Standard error of the mean of groups*
describe():	*Generates descriptive statistics*
first():	*Compute first of group values*
last():	*Compute last of group values*
nth() :	*Take nth value, or a subset if n is a list*
min():	*Compute min of group values*
max():	*Compute max of group values*

Source: https://cmdlinetips.com/2019/10/pandas-groupby-13-functions-to-aggregate/


Confirmed Cases
Deaths
All Cases
Get year and month
Cases per month
Group by
Facet grid
etc.

Merge

Turning two or more data sets into a single one

id	Name	Color
1	Name_a	Black
2	Name_b	White
3	Name_c	Gray
4	Name_d	Blue

id	Transparency	Intensity
1	80%	High
3	75%	Low

Left Join

id	Name	Color	Transparency	Intensity
1	Name_a	Black	80%	High
2	Name_b	White		
3	Name_c	Gray	75%	Low
4	Name_d	Blue		

Full Join

id	Name	Color	Transparency	Intensity
1	Name_a	Black	80%	High
3	Name_c	Gray	75%	Low

When individual data parts are stored in different sets (files, layers, dataframes, etc.)

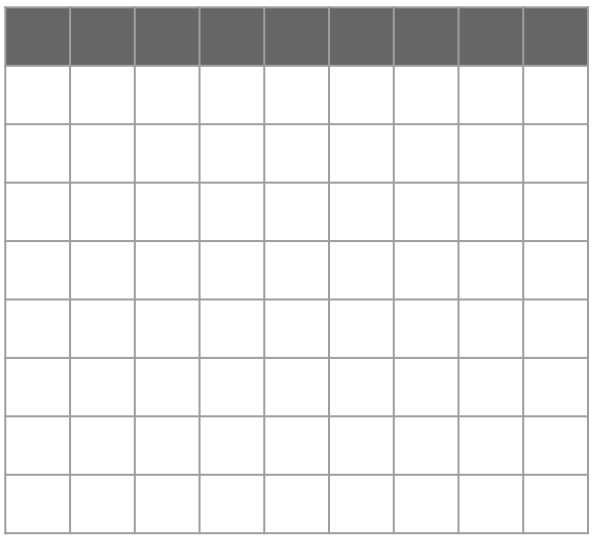
If you plan to combine two or more data parts into a single analysis

A common field or identificator is mandatory
Numbers are preferred as identificators
Text can make the operation more difficult

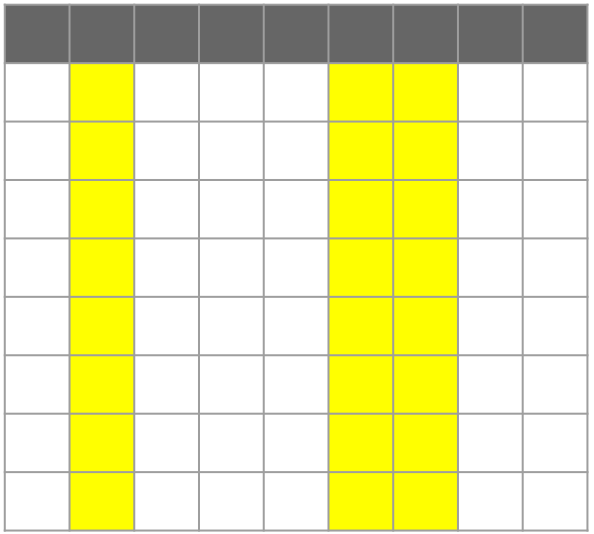
Available in (almost) all software (Excell, QGIS, pandas, etc.)

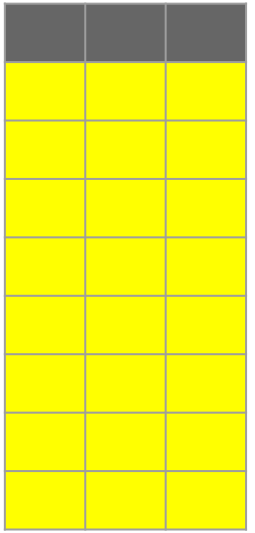
Slice and Dice

Extracting only the elements you need

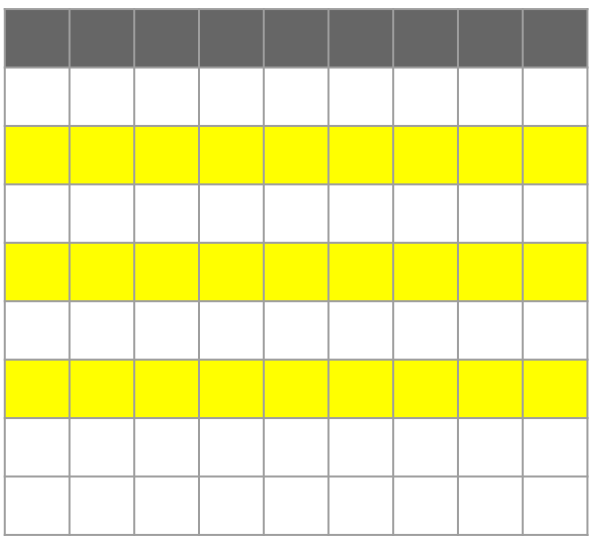


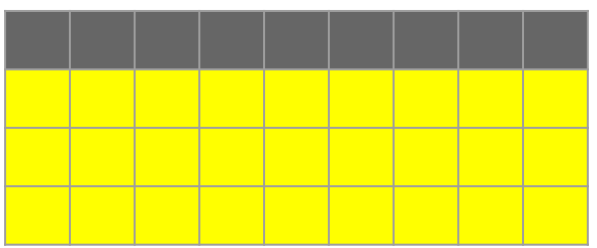
Column names



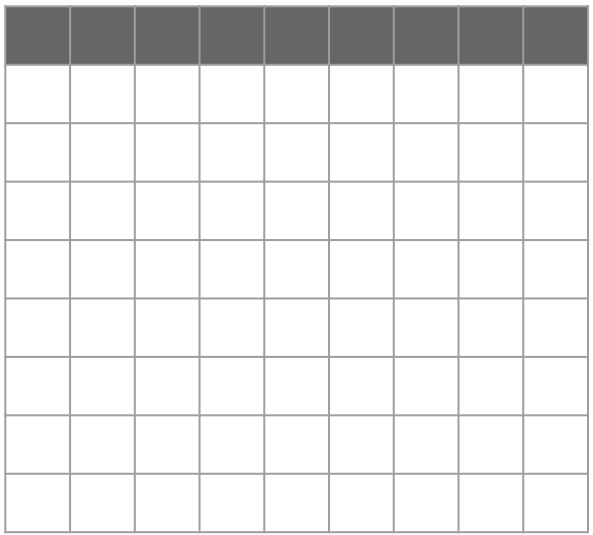


Conditionals

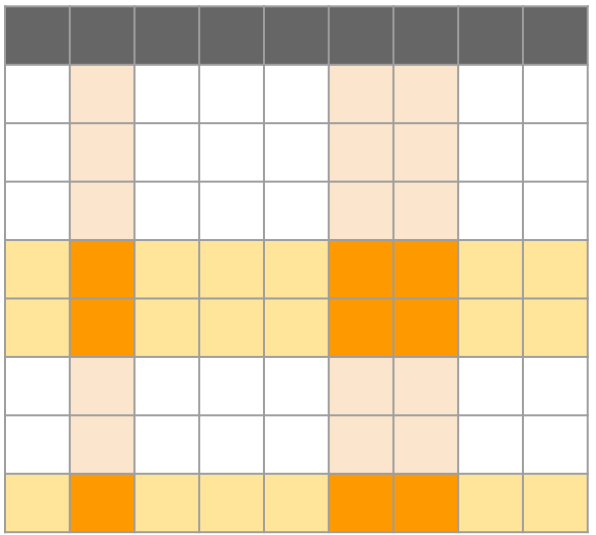


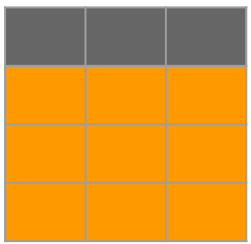


How slice and dice operations work?



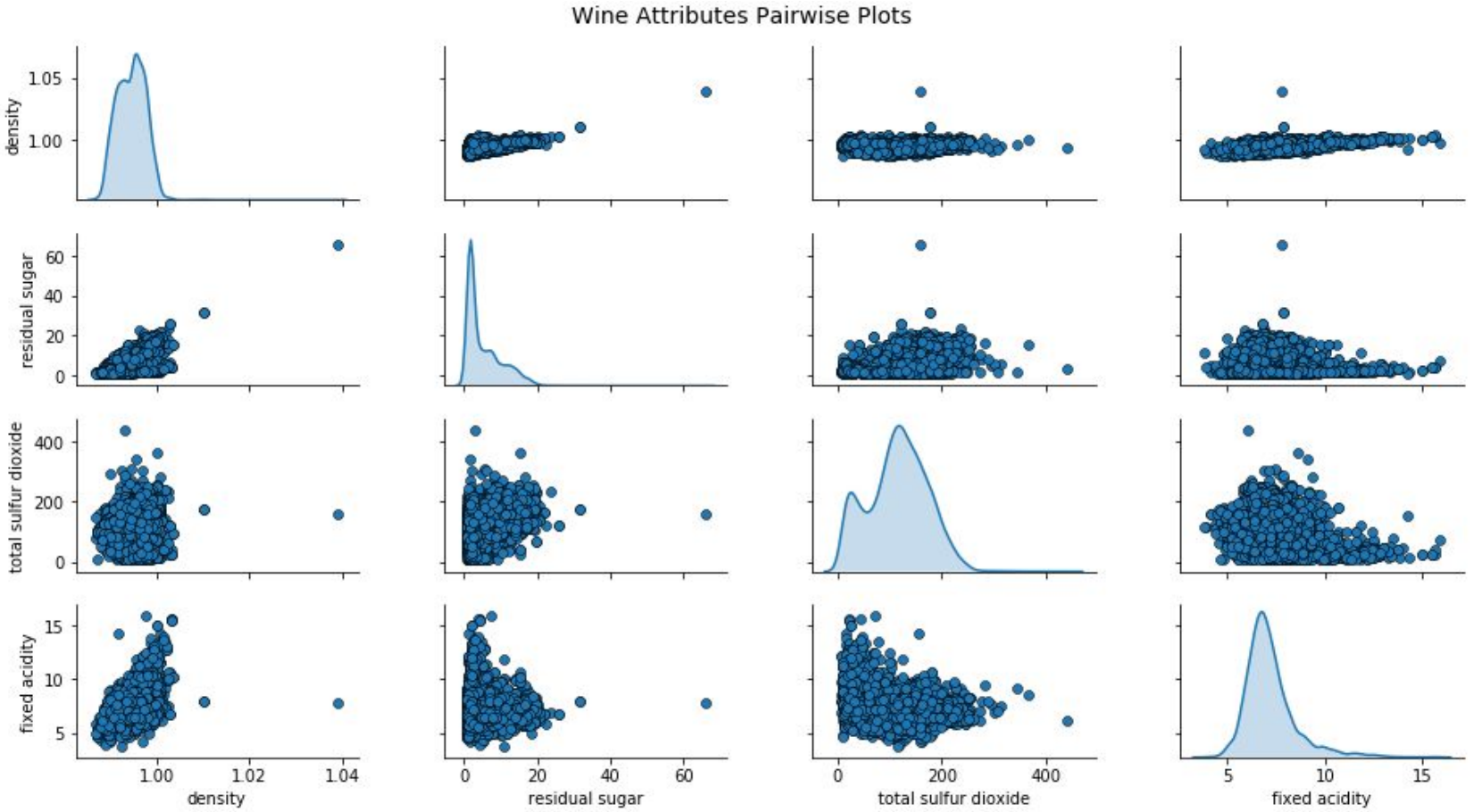
Column and conditionals





2D datavis

Multidimensional data visualisation



Source: http://wiki.gis.com/wiki/index.php/Clip

Source: http://wiki.gis.com/wiki/index.php/Clip

Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2020/2021

FACULTY DIEGO PAJARITO