Institute for advanced architecture of Catalonia | BARCELONA

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

FACULTY DIEGO PAJARITO

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.)

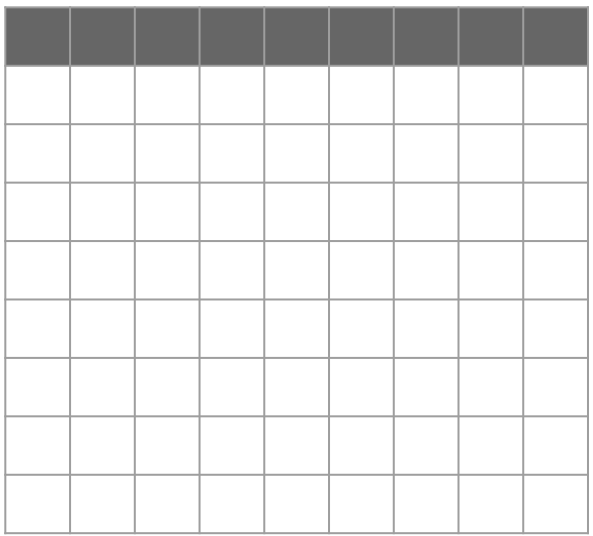
If you plan to combine two or more data parts into a single analysis

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

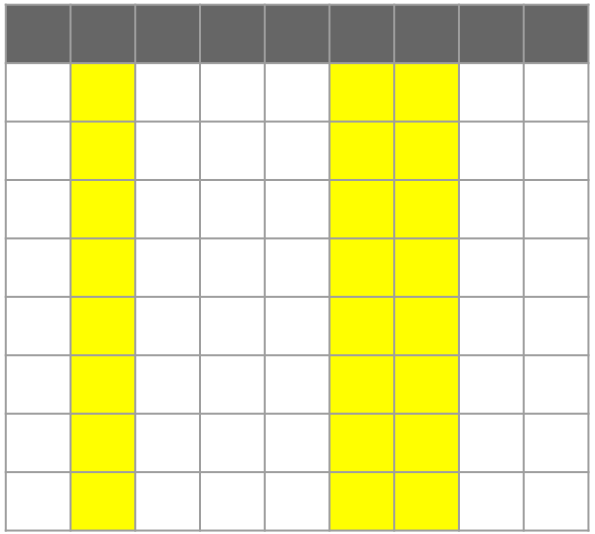
Available in (almost) all software (Excell, QGIS, pandas, etc.)

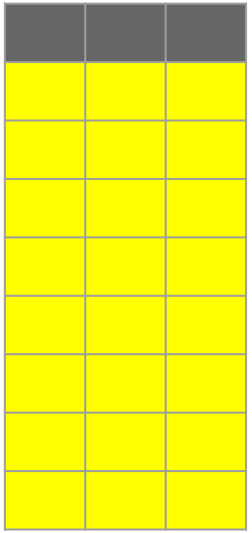
Slice and Dice

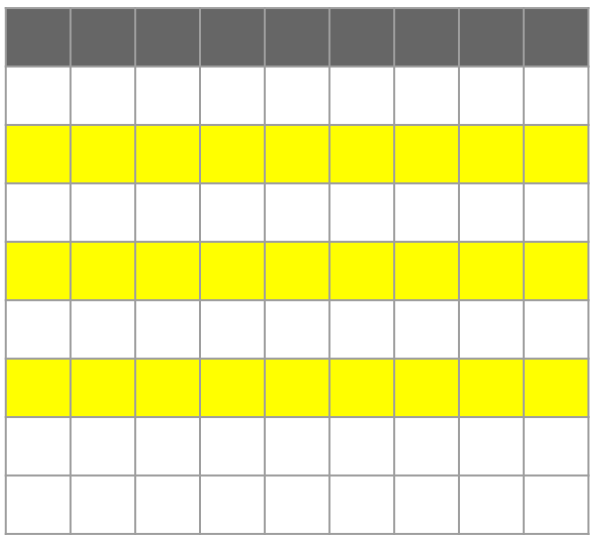
Extracting only the elements you need



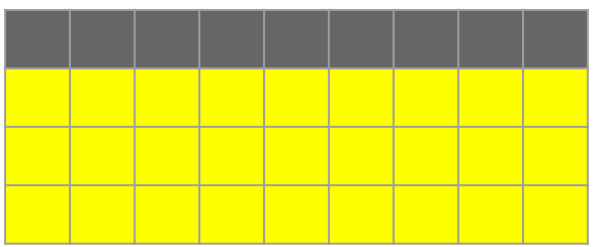
Column names

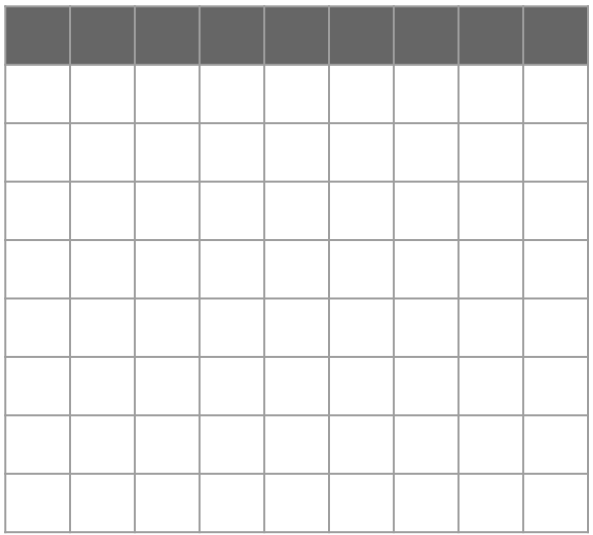




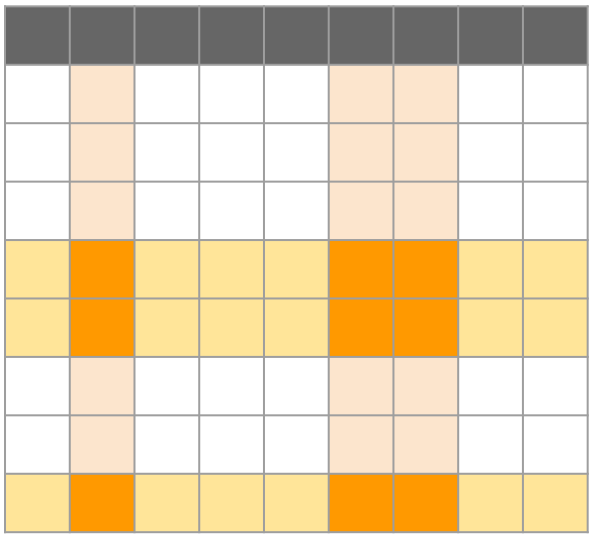


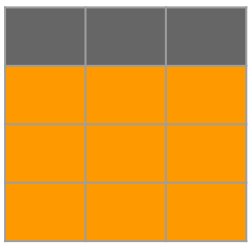
Conditionals

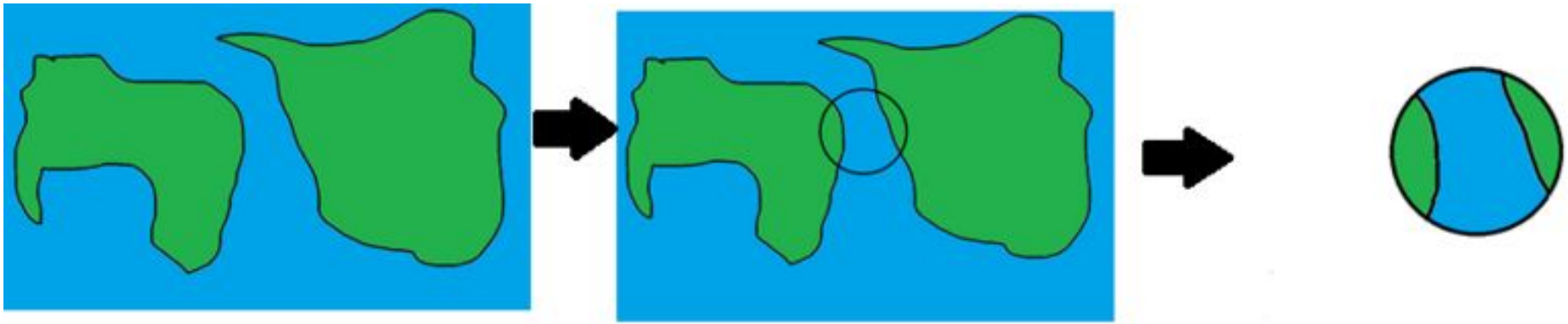




Column and conditionals







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

When massive datasets are no easy to handle

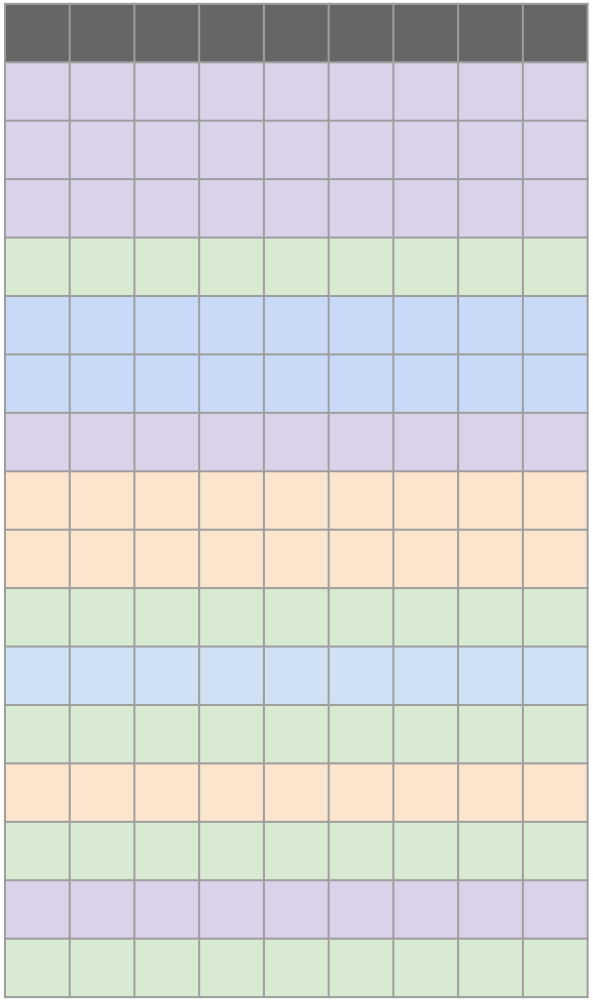
To simplify existing data structures and summarise results

A clear description of contents to select elements

A definition for conditions to select data

Aggregation

Turning multiple rows into less and more meaningful ones

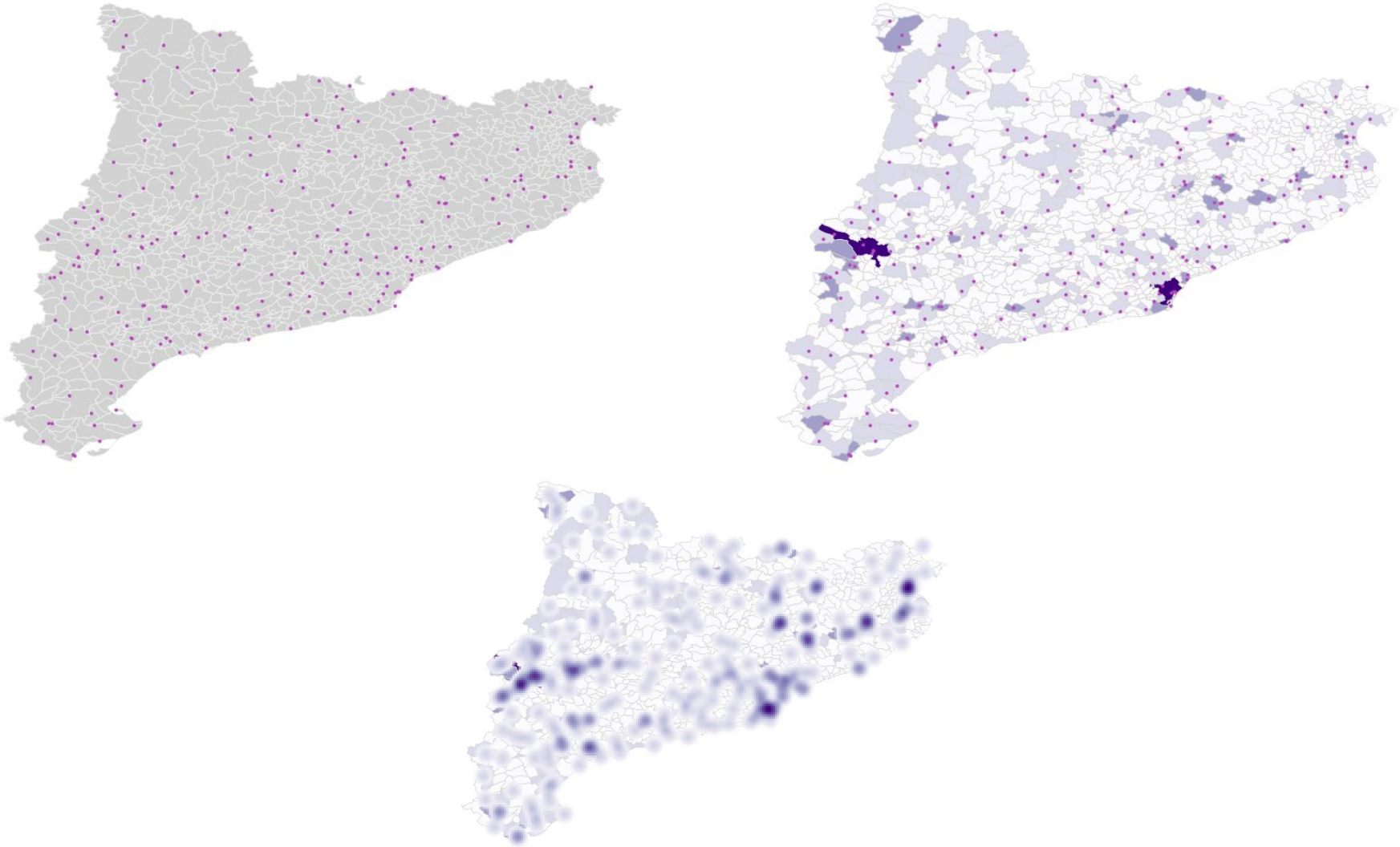


Group by and functions

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/



Count points in polygons

Count lines in polygons

Clip

Dissolve

Spatial clusters

Heatmaps

Zonal statistics

...

Source: https://www.qgistutorials.com/en/docs/points_in_polygon.html
https://geopandas.org/aggregation_with_dissolve.html

Institute for advanced architecture of Catalonia | BARCELONA

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

FACULTY DIEGO PAJARITO