Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2021/2022

FACULTY DIEGO PAJARITO

Understanding flow, density and distribution

Multidimensional Data

A single object can be described using multiple variables

In econometrics, having more than three dimensions to describe a single phenomenon generates a panel a multidimensional data panel.

For advanced architecture, datasets are commonly build with tens of dimensions.

How many options do we have to analyse and understand these dimensions?

Definition from: https://en.wikipedia.org/wiki/Multidimensional_panel_data

Some Facts

microdata.worldbank.org	3,695 surveys
	3,478,375 variables
	2,609 citations
opendata-ajuntament.barcelona.cat	522 datasets

To provide an experience handling common tasks of big data, data science or data analytics.

The course provides a practical perspective of the main activities developed for urban analytics. From data collection, ingestion, analysis and visualization, the students will experience the workflow while getting their hands on extracting information from massive datasets.

Getting familiar with data repositories for Big Data

Spatial and temporal dimensions

Multidimensional data management and visualisation

Data visualisation and descriptive statistics

Studio integration through pandas

Source Code

Examples of the tasks performed during the course

Gallery / Dashboard / Portfolio

Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2021/2022

FACULTY DIEGO PAJARITO

Flow

Connecting endpoints through data analysis

Open Data **BCN**

My First Dashboard Last Data Imported: 3/19/2018 Re-import Data

Configure | Open in Browser | Export | Publish | Help

38.72	27	30	1.03
gestation_weeks	mother_age	father_age	plurality

Variable Importance

Gestation Weeks - Histogram

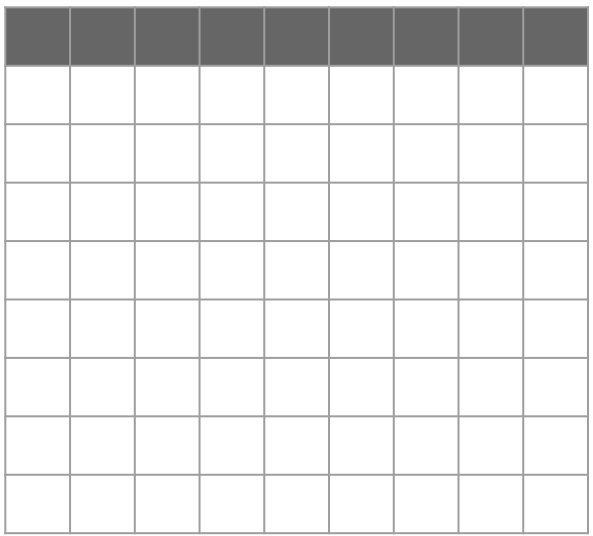
Analytics 1 - Summary by Class

mother_race_name	Class	F Score	Accuracy Rate
American Indian	TRUE	0.15068493151	0.87608850821
American Indian	FALSE	0.93316907915	0.87608850821
Black	TRUE	0.22128360977	0.84882
Black	FALSE	0.91628365433	0.84882
Chinese	TRUE	0.10288386594	0.926213218
Chinese	FALSE	0.9615243189	0.926213218
Filipino	TRUE	0.18138875768	0.88435101768
Filipino	FALSE	0.9377849043	0.88435101768
Hawaiian	TRUE	0.22680412371	0.87001733102
Hawaiian	FALSE	0.92904446547	0.87001733102
Japanese	TRUE	0.14005602241	0.91096287703

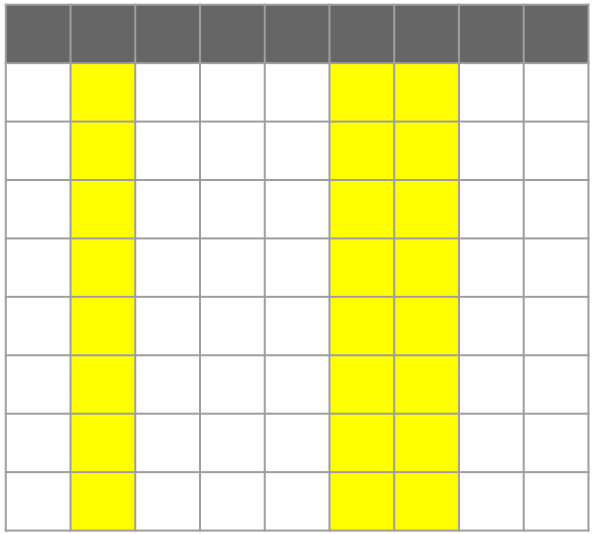
US State - Gestation Week

Mother Age Distribution

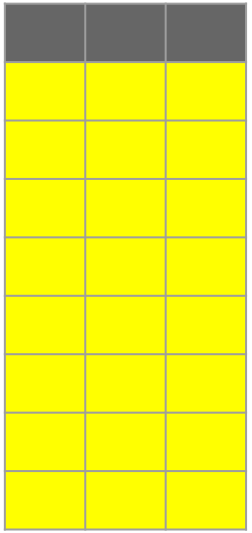
Mother Age x Father Age



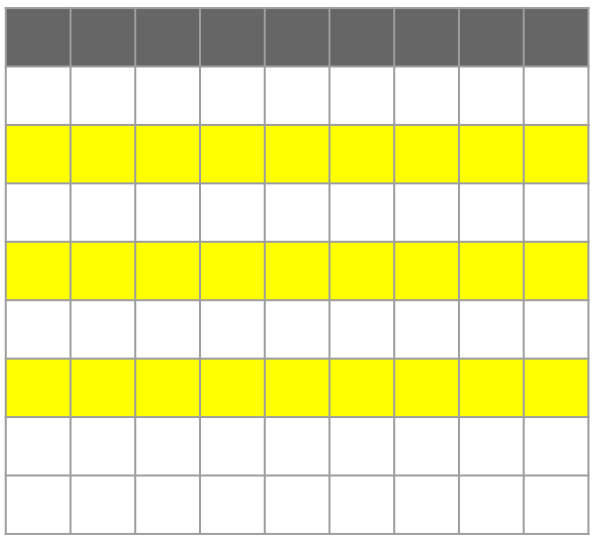
Column names →



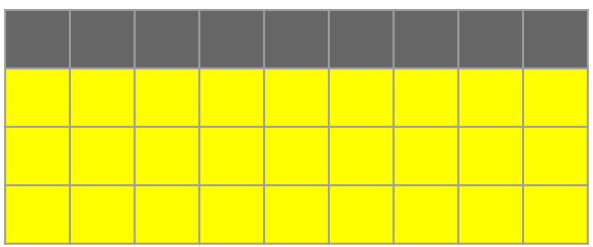
→

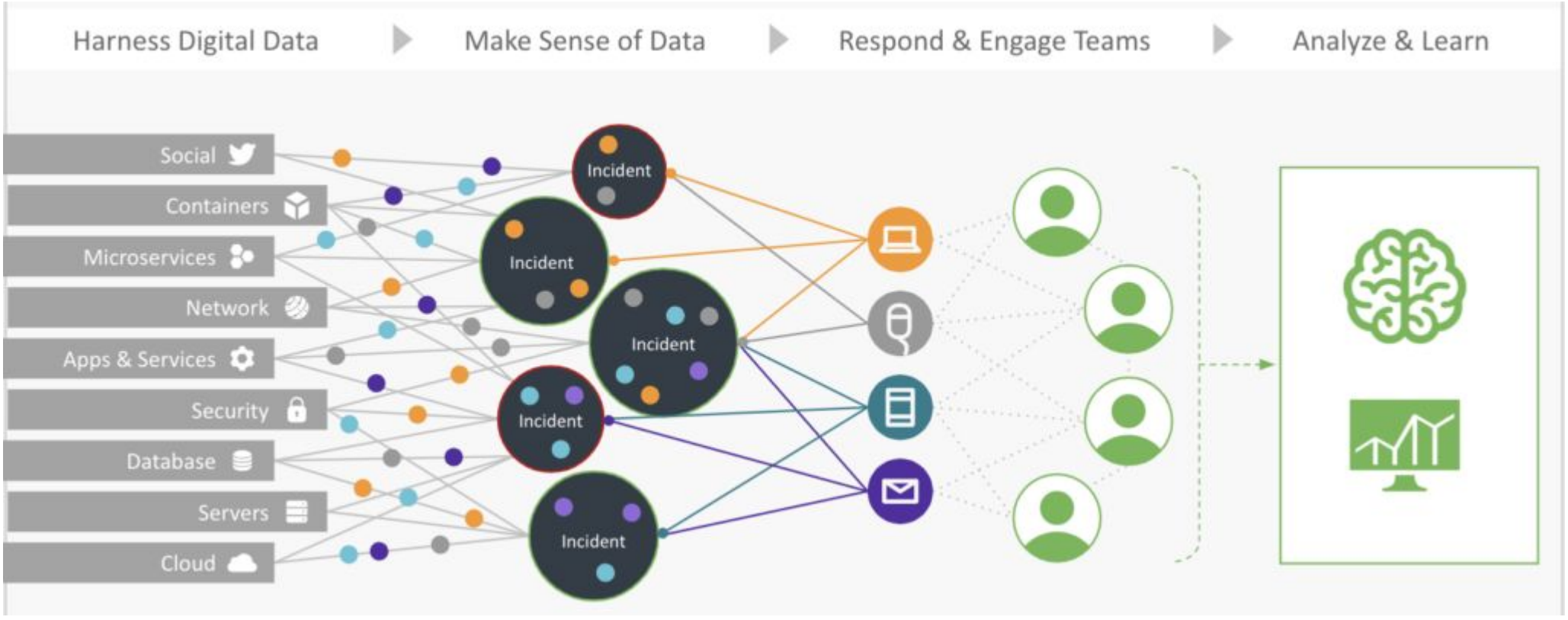


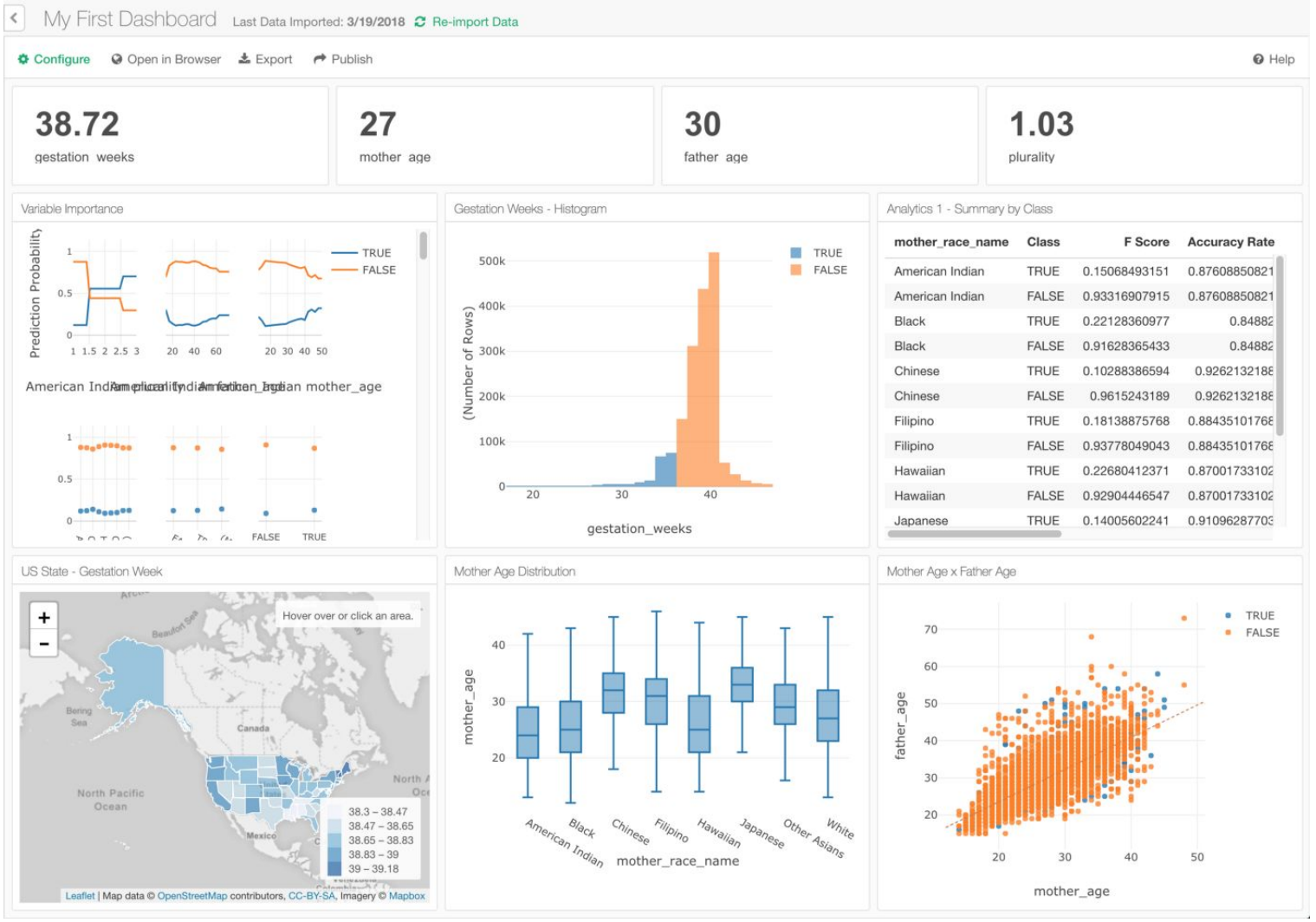
Conditionals →



→

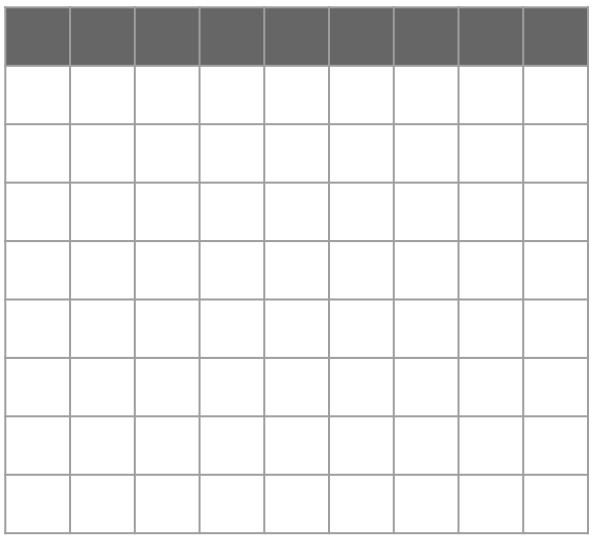



Harness Digital Data
Make Sense of Data
Respond & Engage Teams
Analyze & Learn
Social
Containers
Microservices
Network
Apps & Services
Security
Database
Servers
Cloud
Incident
Incident
Incident
Incident
Incident

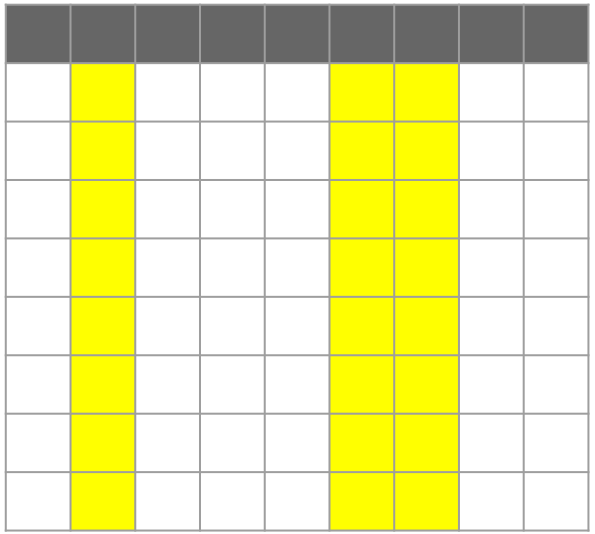

My First Dashboard
Last Data Imported: 3/19/2018
Re-import Data
Configure
Open in Browser
Export
Publish
Help
38.72
gestation_weeks
27
mother_age
30
father_age
1.03
plurality
Variable Importance
Prediction Probability
TRUE
FALSE
American Indian
mother_age
Gestation Weeks - Histogram
(Number of Rows)
gestation_weeks
Analytics 1 - Summary by Class
mother_race_name | Class | F Score | Accuracy Rate
American Indian | TRUE | 0.15068493151 | 0.87608850821
American Indian | FALSE | 0.93316907915 | 0.87608850821
Black | TRUE | 0.22128360977 | 0.84882
Black | FALSE | 0.91628365433 | 0.84882
Chinese | TRUE | 0.10288386594 | 0.92621321
Chinese | FALSE | 0.9615243189 | 0.92621321
Filipino | TRUE | 0.18138875768 | 0.88435101768
Filipino | FALSE | 0.93778049043 | 0.88435101768
Hawaiian | TRUE | 0.22680412371 | 0.87001733102
Hawaiian | FALSE | 0.92904446547 | 0.87001733102
Japanese | TRUE | 0.14005602241 | 0.91096287703
US State - Gestation Week
Hover over or click an area.
Bering Sea
Canada
North Pacific Ocean
Mexico
38.3 – 38.47
38.47 – 38.65
38.65 – 38.83
38.83 – 39
39 – 39.18
Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox
Mother Age Distribution
mother_age
American Indian
Black
Chinese
Filipino
Hawaiian
Japanese
Other Asians
White
mother_race_name
Mother Age x Father Age
father_age
mother_age
TRUE
FALSE

Density

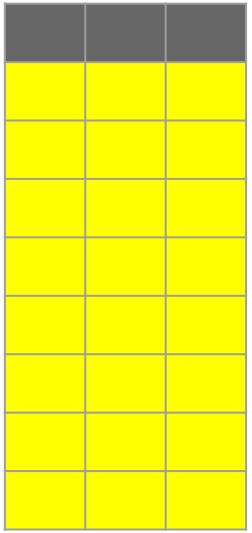
Finding where data values concentrate



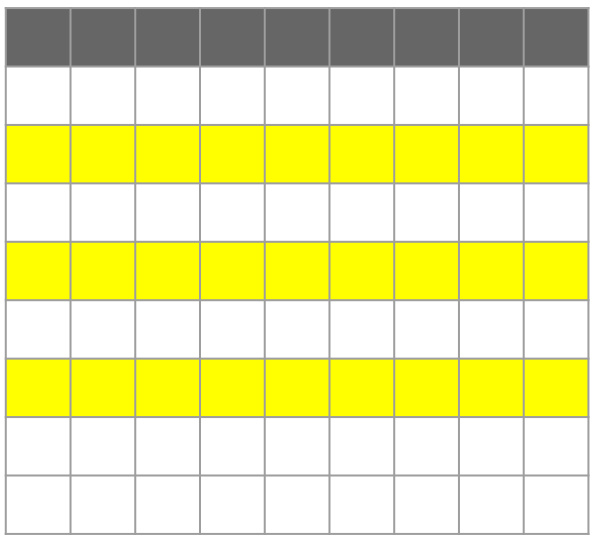
Column names →



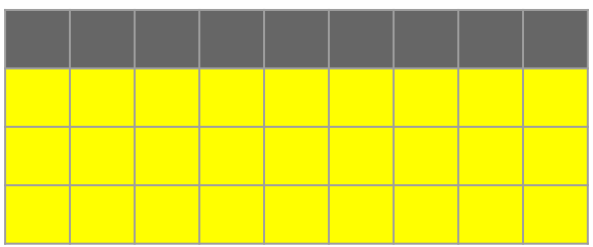
→



Conditionals →



→



	species	sepal_length	sepal_width	petal_length	petal_width
0	setosa	5.1	3.5	1.4	0.2
1	setosa	4.9	3.0	1.4	0.2
2	setosa	4.7	3.2	1.3	0.2
3	setosa	4.6	3.1	1.5	0.2
4	setosa	5.0	3.6	1.4	0.2
50	versicolor	7.0	3.2	4.7	1.4
51	versicolor	6.4	3.2	4.5	1.5
52	versicolor	6.9	3.1	4.9	1.5
53	versicolor	5.5	2.3	4.0	1.3
54	versicolor	6.5	2.8	4.6	1.5
100	virginica	6.3	3.3	6.0	2.5
101	virginica	5.8	2.7	5.1	1.9
102	virginica	7.1	3.0	5.9	2.1
103	virginica	6.3	2.9	5.6	1.8
104	virginica	6.5	3.0	5.8	2.2

SUM

SUM

SUM

species	sepal_length	sepal_width	petal_length	petal_width
setosa	24.3	16.4	7.0	1.0
versicolor	32.3	14.6	22.7	7.2
virginica	32.0	14.9	28.4	10.5

Source:
https://towardsdatascience.com/how-to-use-the-split-apply-combine-strategy-in-pandas-groupby-29e0eb44b62e

Pivot

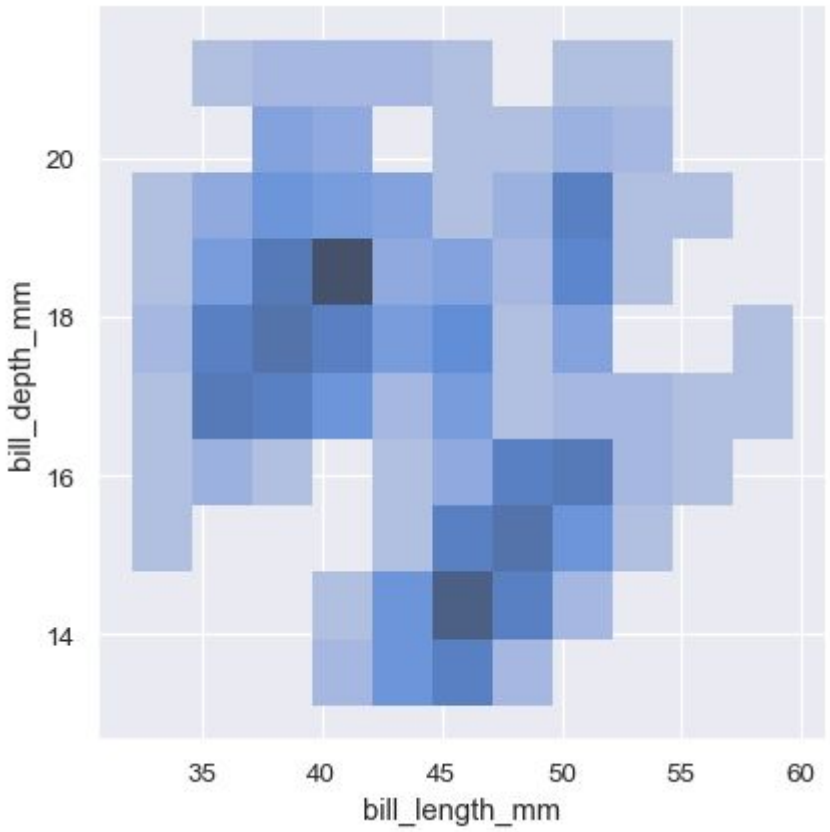
df

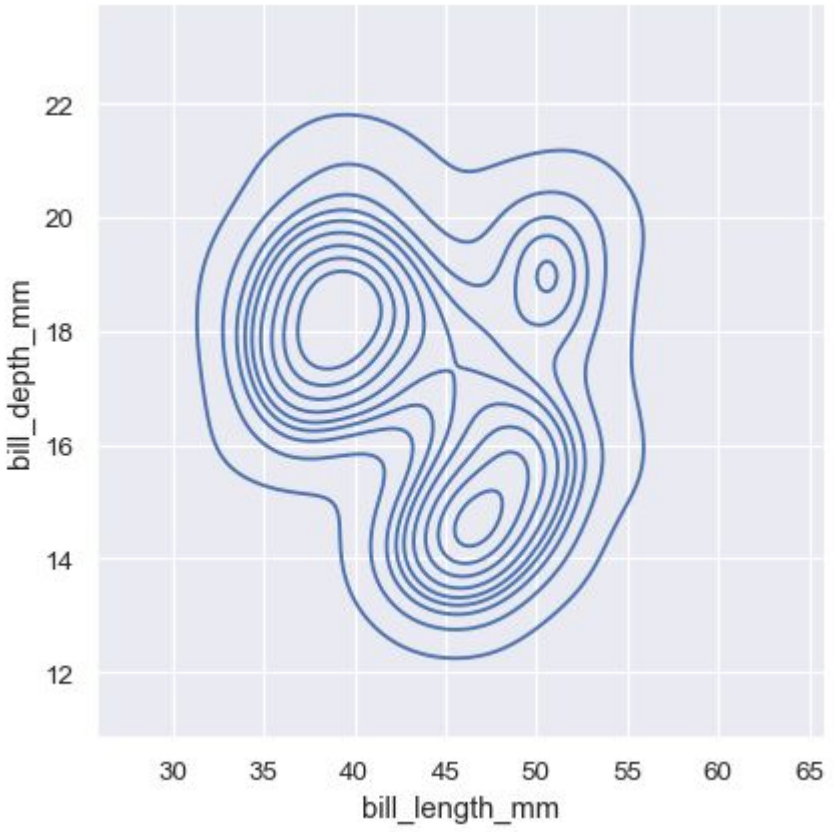
	foo	bar	baz	zoo
0	one	A	1	x
1	one	B	2	y
2	one	C	3	z
3	two	A	4	q
4	two	B	5	w
5	two	C	6	t

```
df.pivot(index='foo',
         columns='bar',
         values='baz')
```

bar	A	B	C
foo			
one	1	2	3
two	4	5	6

Source:
https://pandas.pydata.org/pandas-docs/stable/user_guide/reshaping.html





Source: https://seaborn.pydata.org/tutorial/distributions.html

Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2021/2022

FACULTY DIEGO PAJARITO