

# What is DATA?

Data Journalism E-Course (Session-3 Aug 11, 2021)



**1** What is  
Data  
Science?

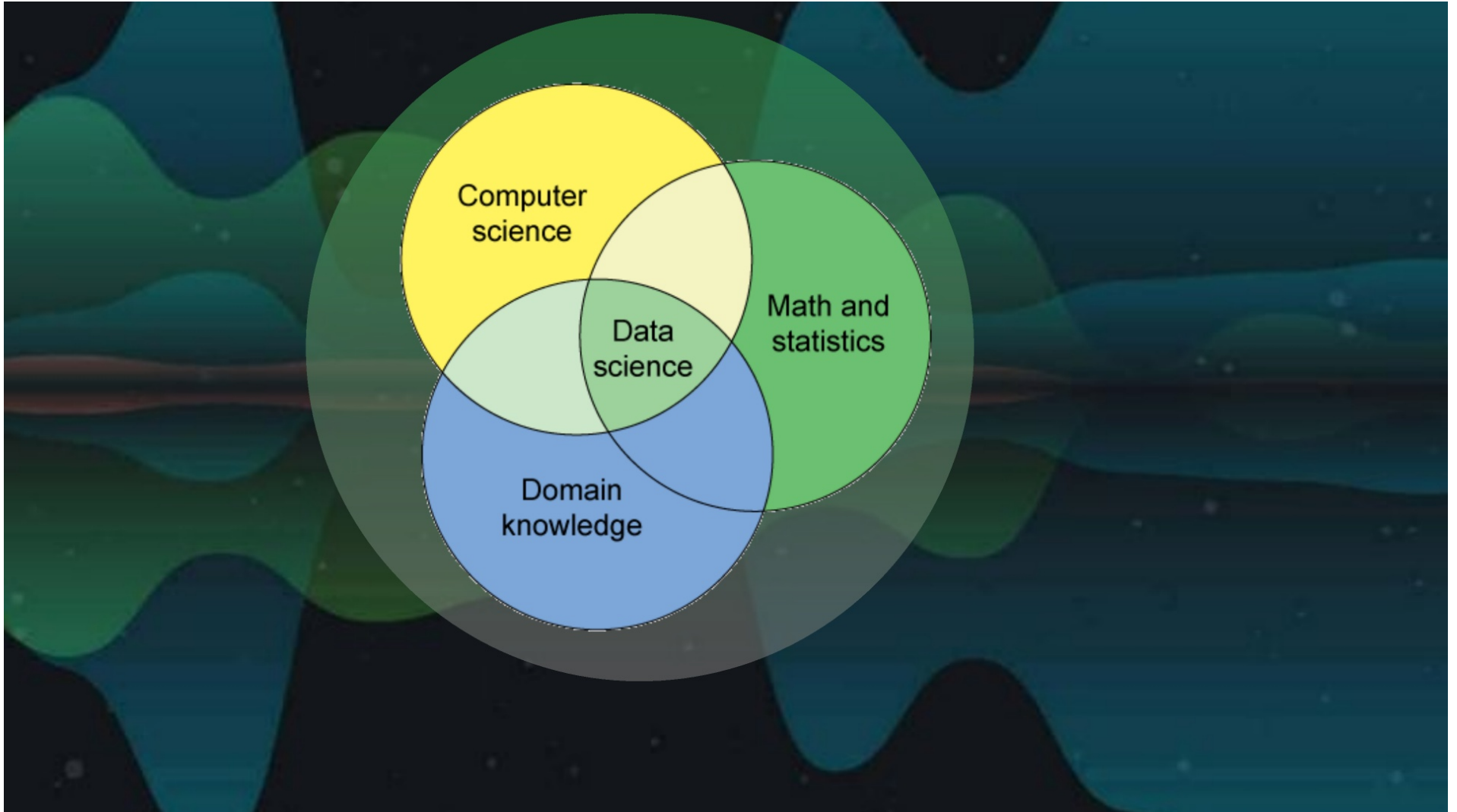
**2** What is  
DATA?

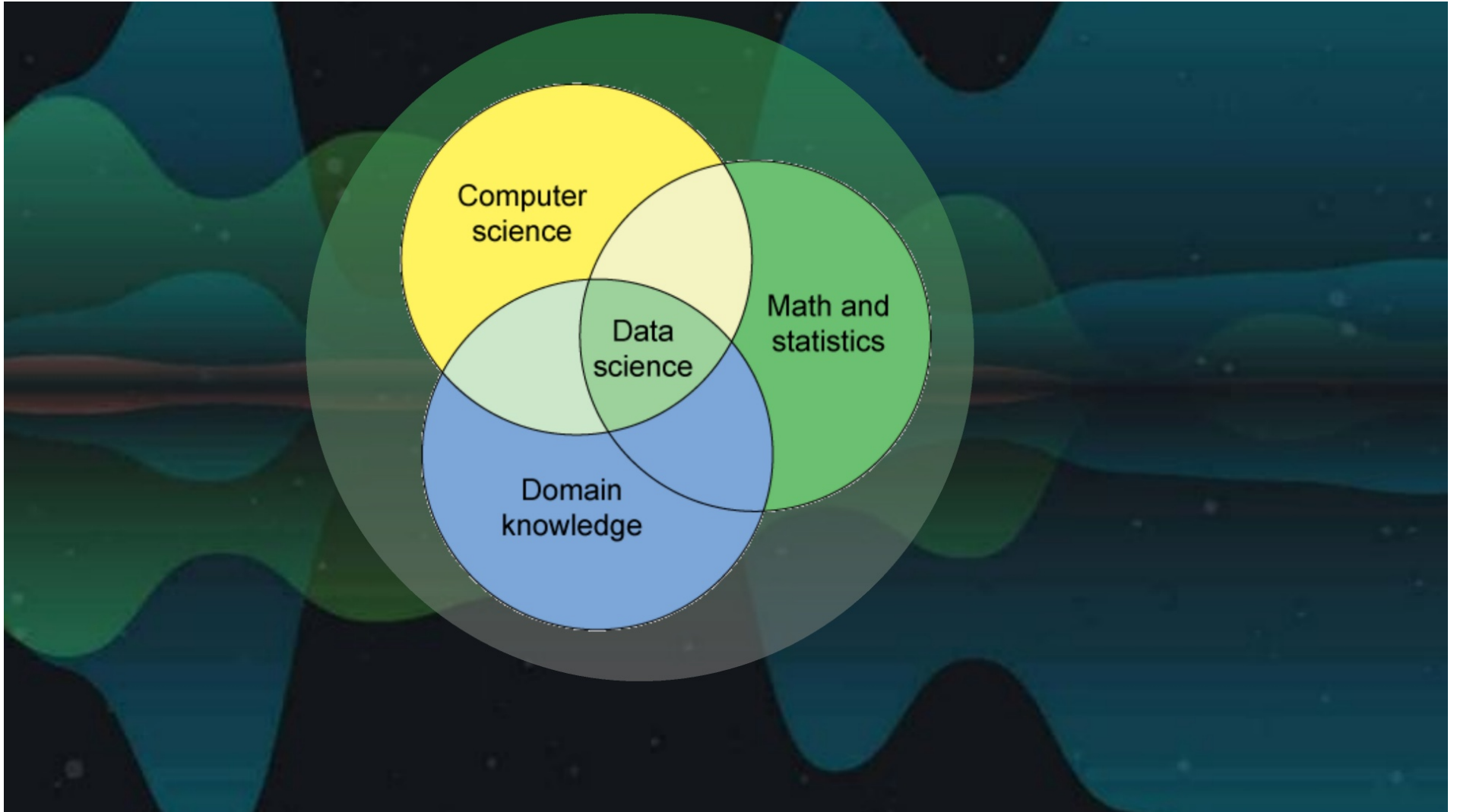
**3** Data  
Pipeline

**4** USE  
CASES

By ALI SROUR

Data Scientist and AI Consultant  
[/in/srouali](#)





# What is DATA?

Data Journalism E-Course (Session-3 Aug 11, 2021)



**1** What is  
Data  
Science?

**2** What is  
DATA?

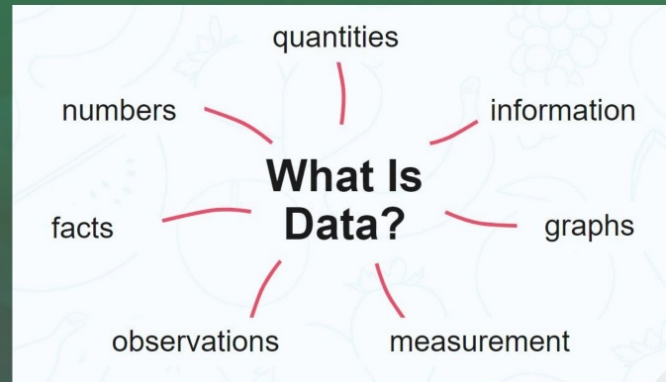
**3** Data  
Pipeline

**4** USE  
CASES

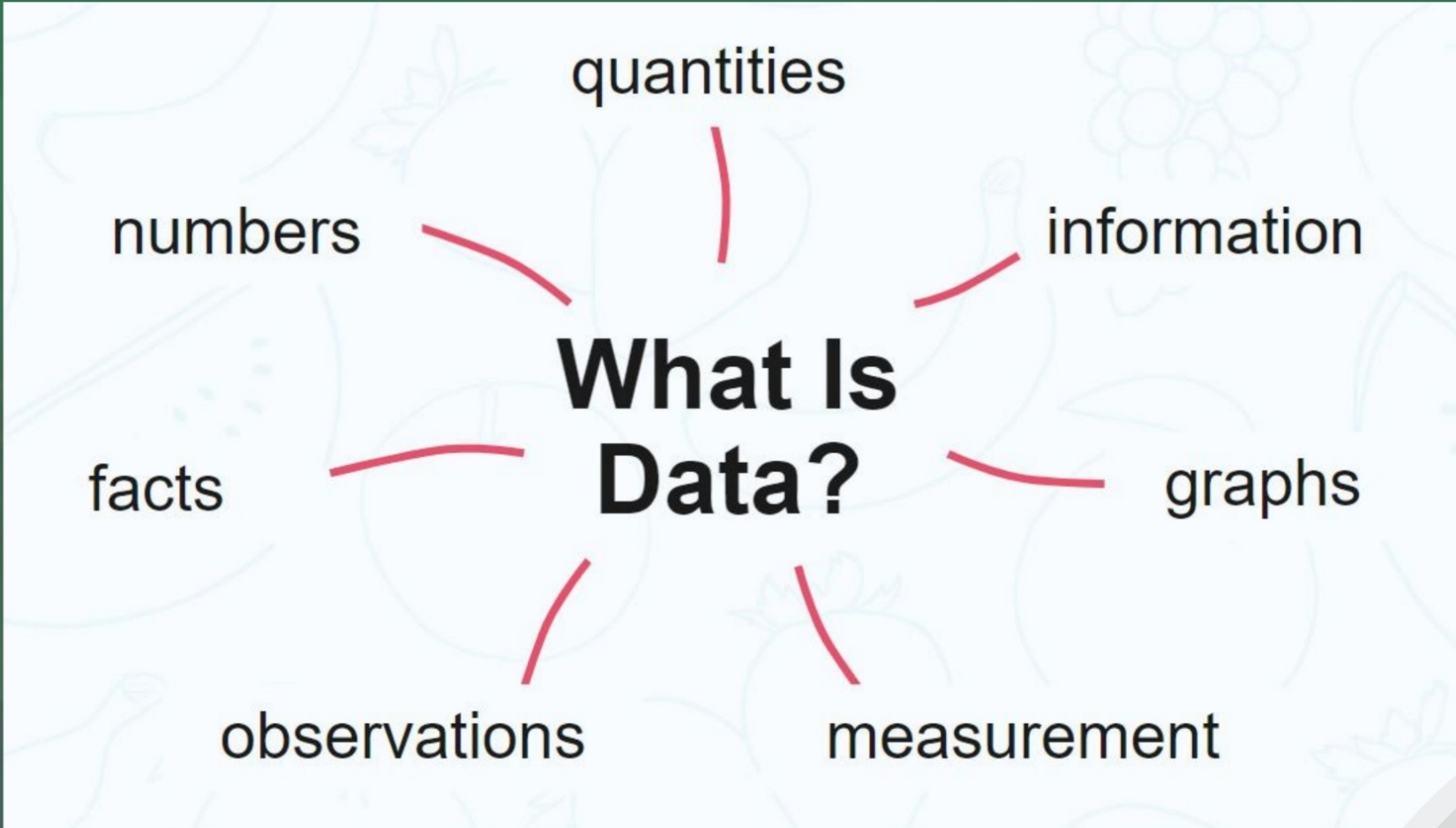
By ALI SROUR

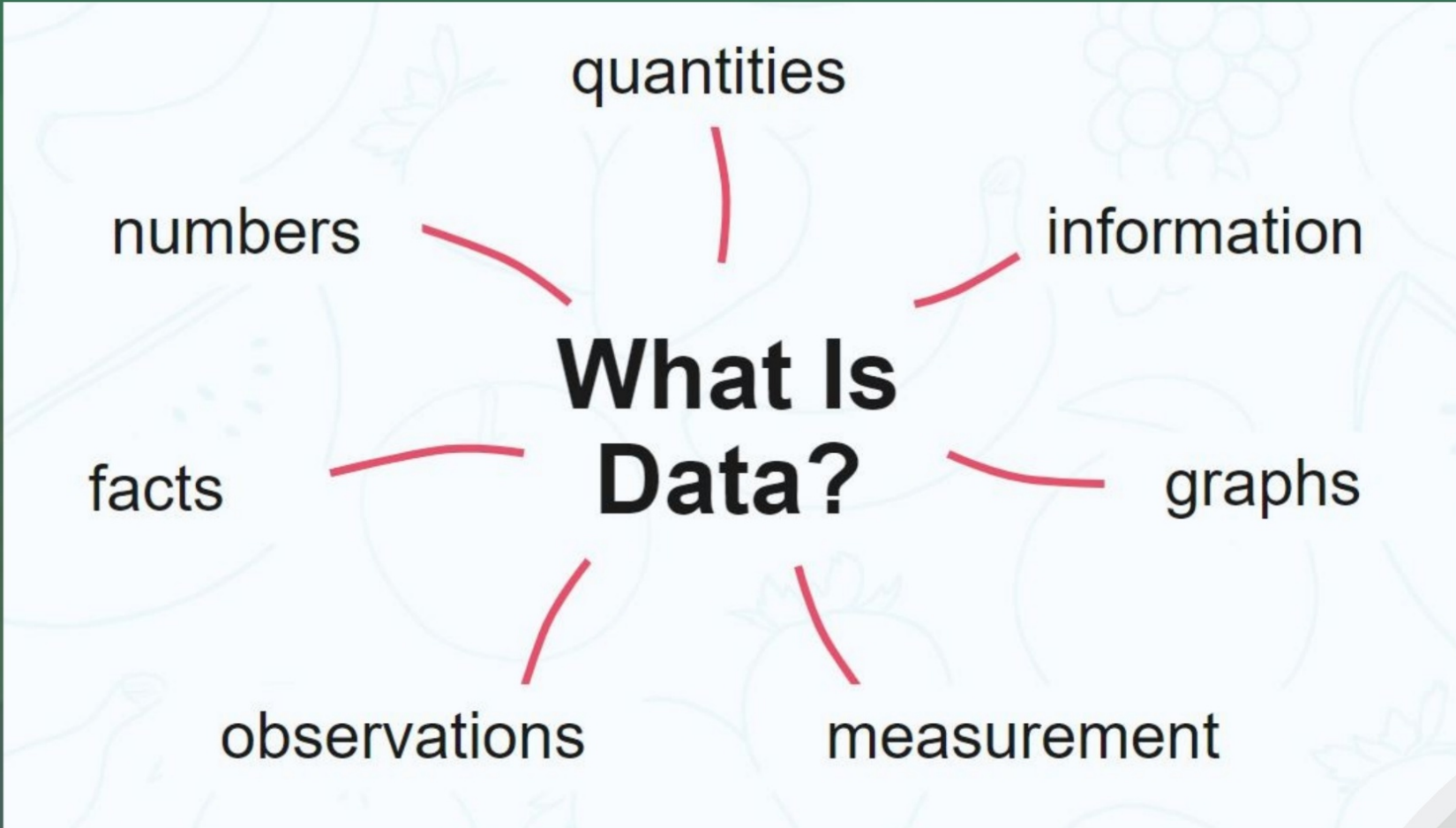
Data Scientist and AI Consultant  
[/in/srouali](#)

# DATA Formats

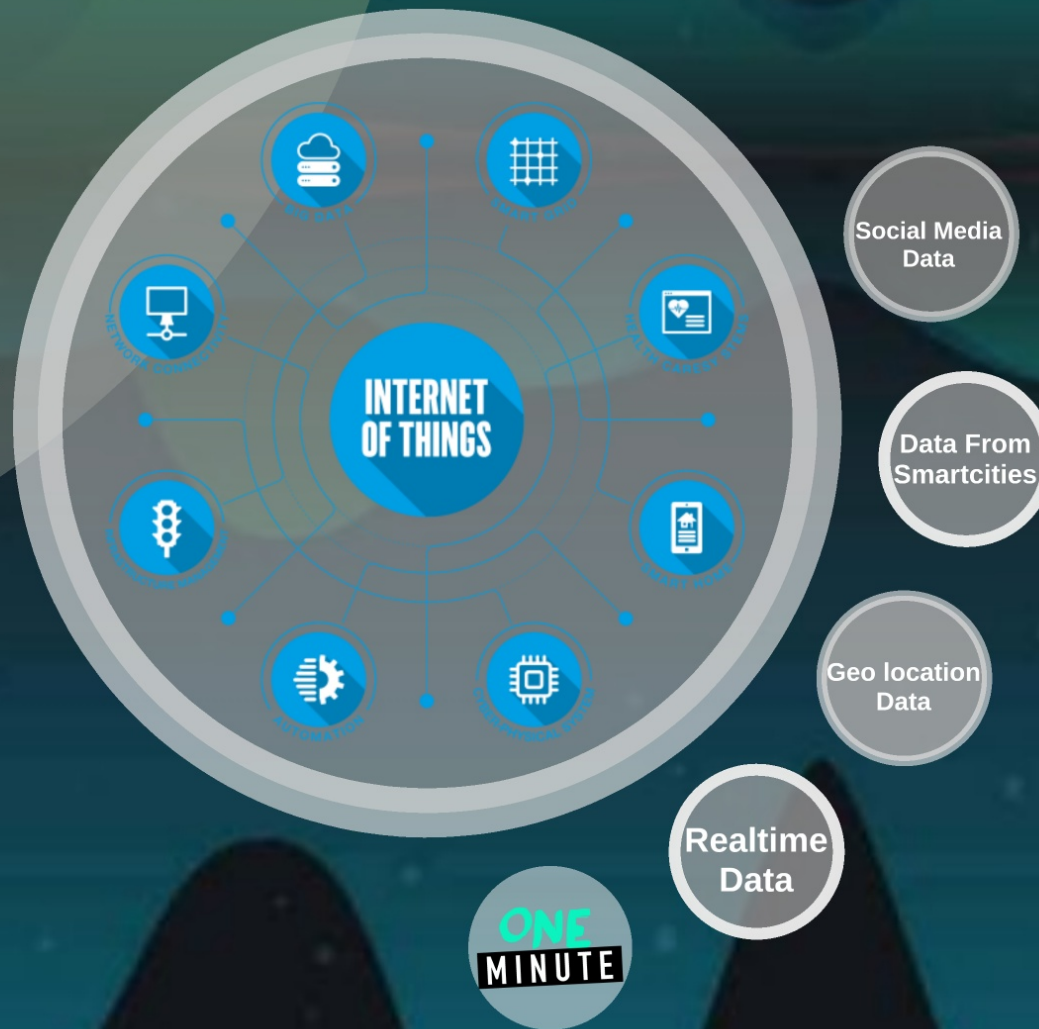


Sources





# Examples





# Social Media Data



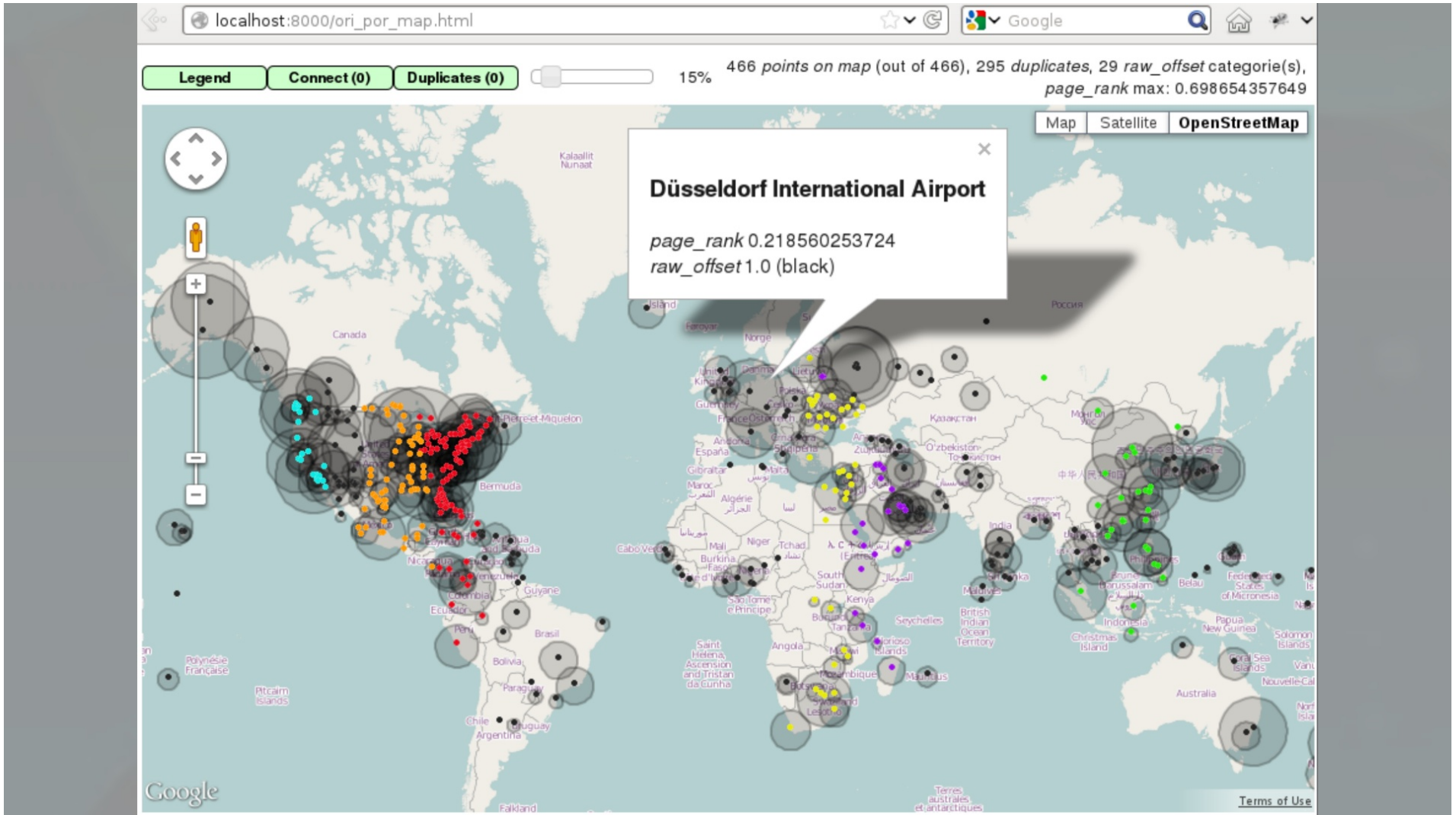
# Data From SC





# Geo-location Data







Realtime Data

Something is  
Happening  
**NOW**

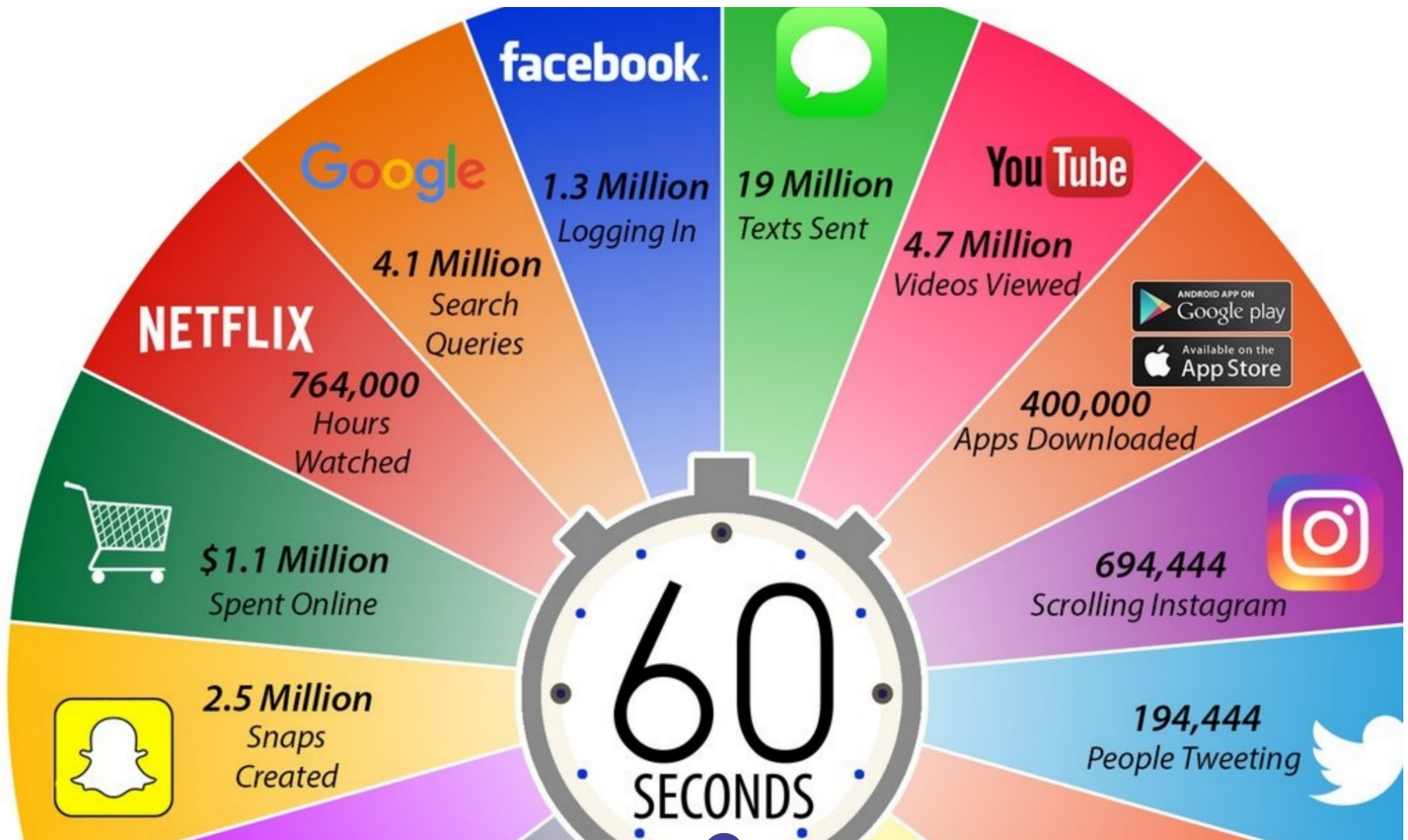
# 2020 *This Is What Happens In An Internet Minute*

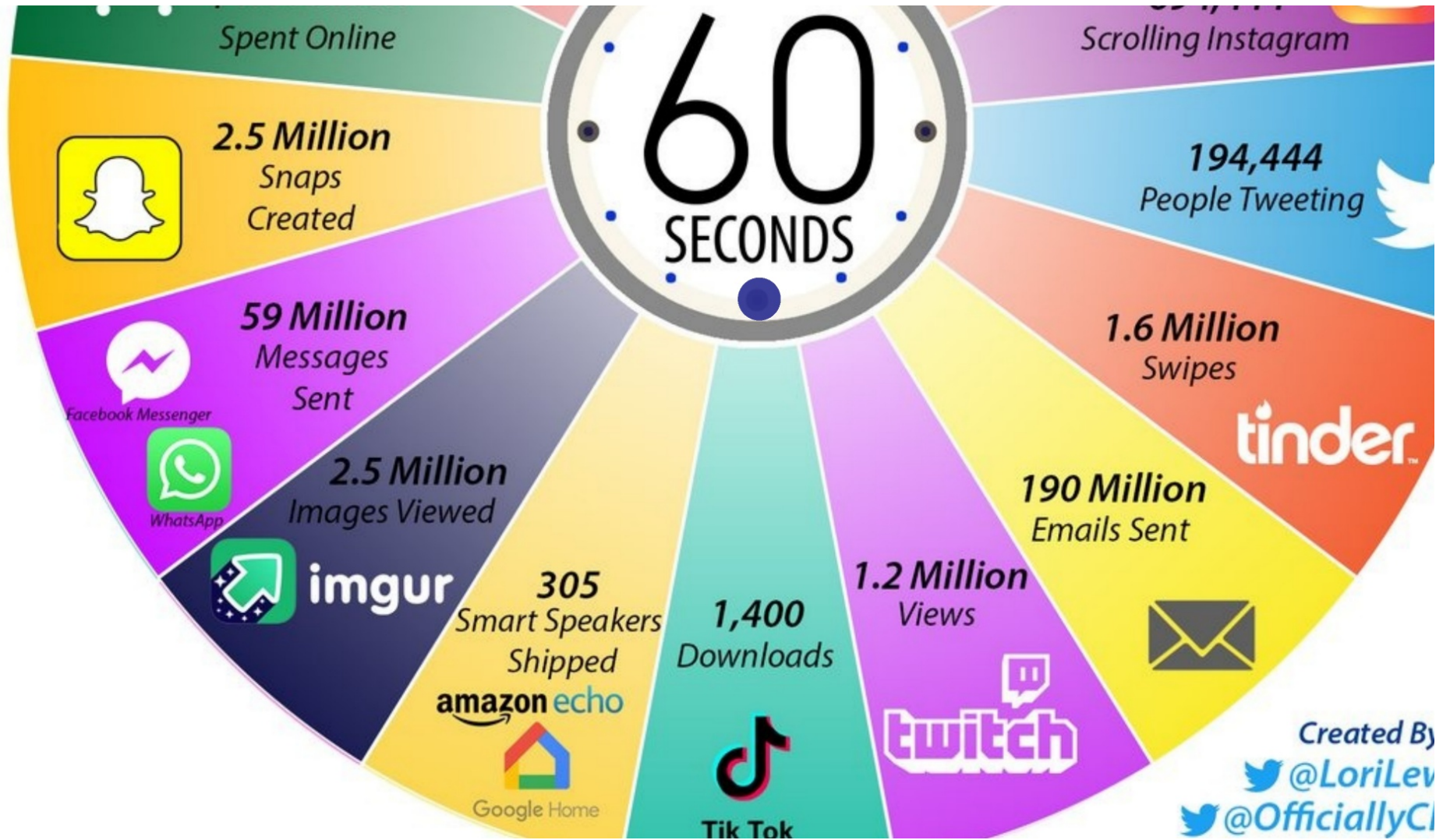


# 2020 *This Is What Happens In An Internet Minute*



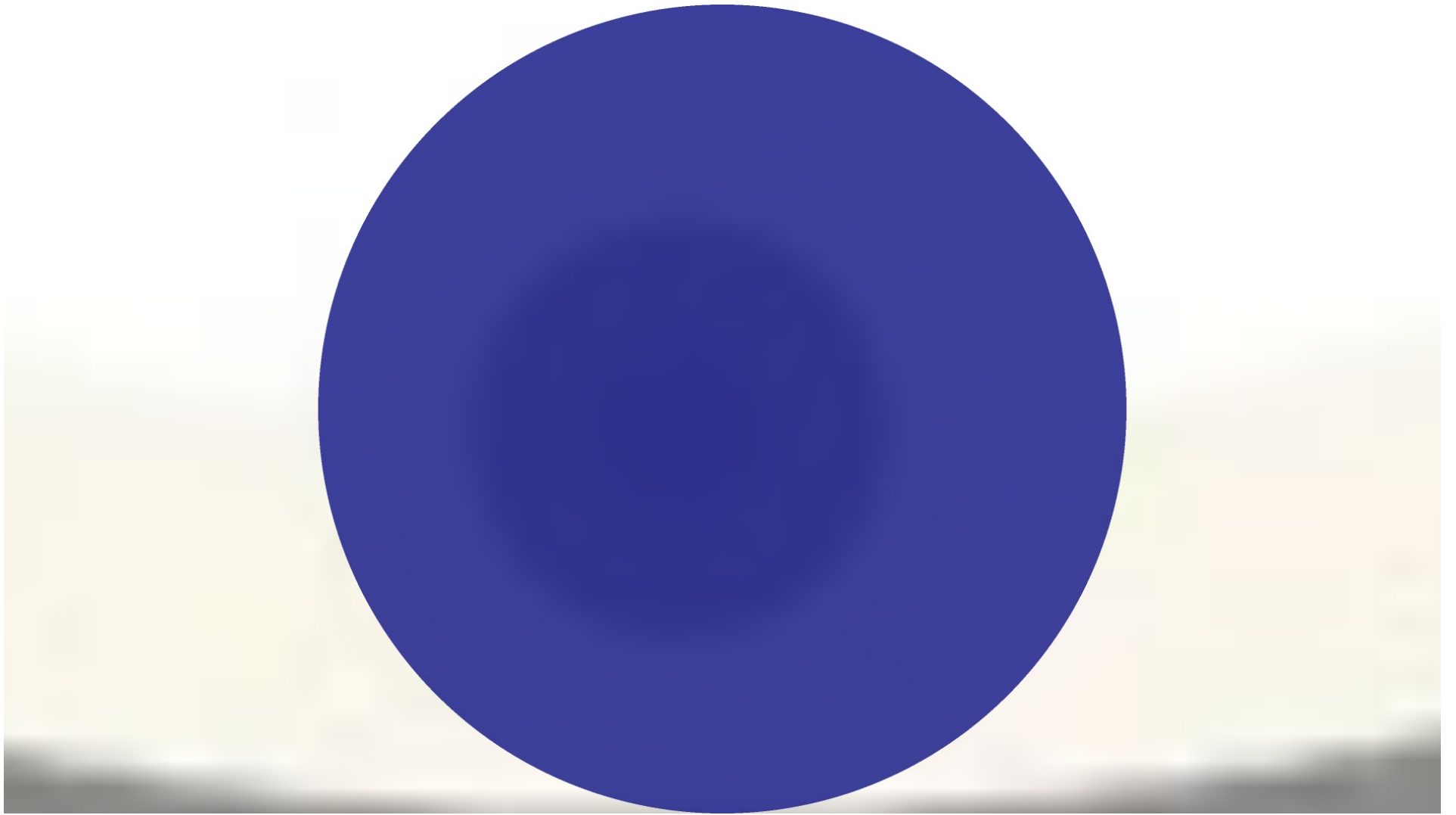






# 2020 *This Is What Happens In An Internet Minute*







Data is **EVERYWHERE..**

Data is **EVERYWHERE..**

We don't really care about "big data",

Data is **EVERYWHERE..**

We don't really care about "big data",

We want **USEFUL** data

# 2020 *This Is What Happens In An Internet Minute*





# What is DATA?

Data Journalism E-Course (Session-3 Aug 11, 2021)



**1** What is  
Data  
Science?

**2** What is  
DATA?

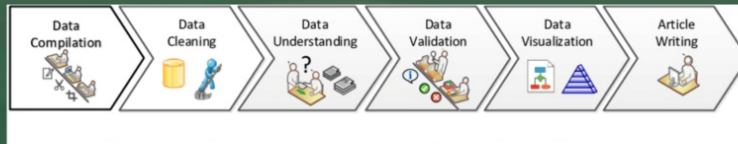
**3** Data  
Pipeline

**4** USE  
CASES

By ALI SROUR

Data Scientist and AI Consultant  
[/in/srouali](#)

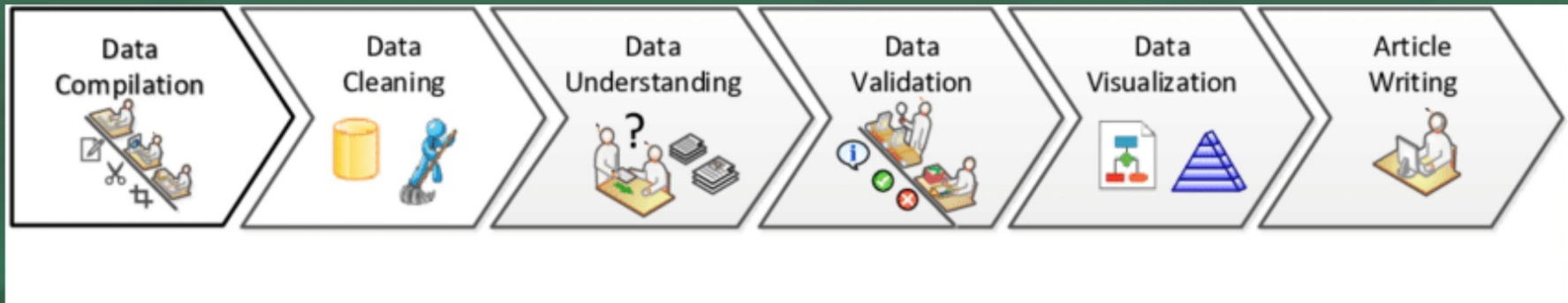
# Data Driven Workflow



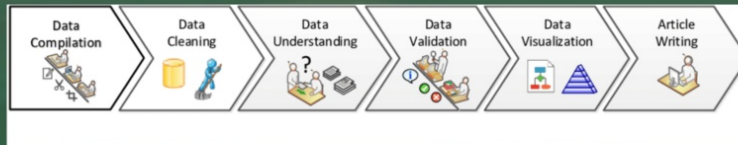
Traditional  
VS  
Data Journalism

DATA  
PIPELINE

KEY WINS



# Data Driven Workflow



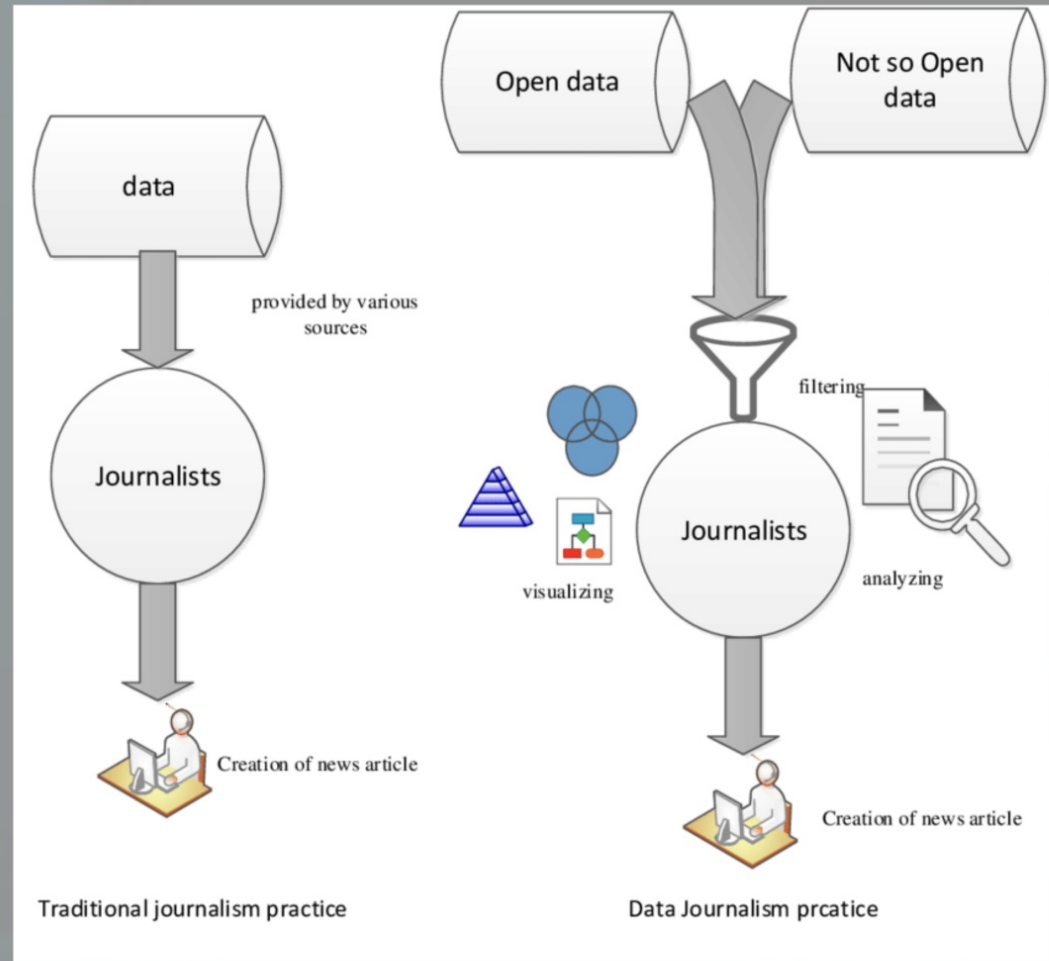
Traditional  
VS  
Data Journalism

DATA  
PIPELINE

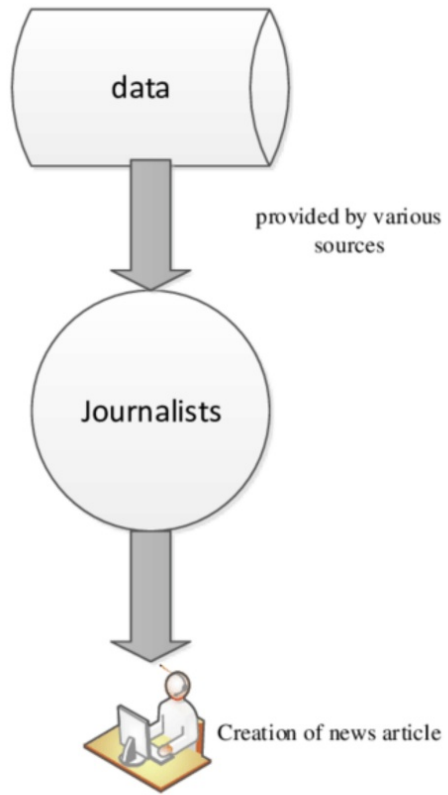
KEY WINS

# Comparison

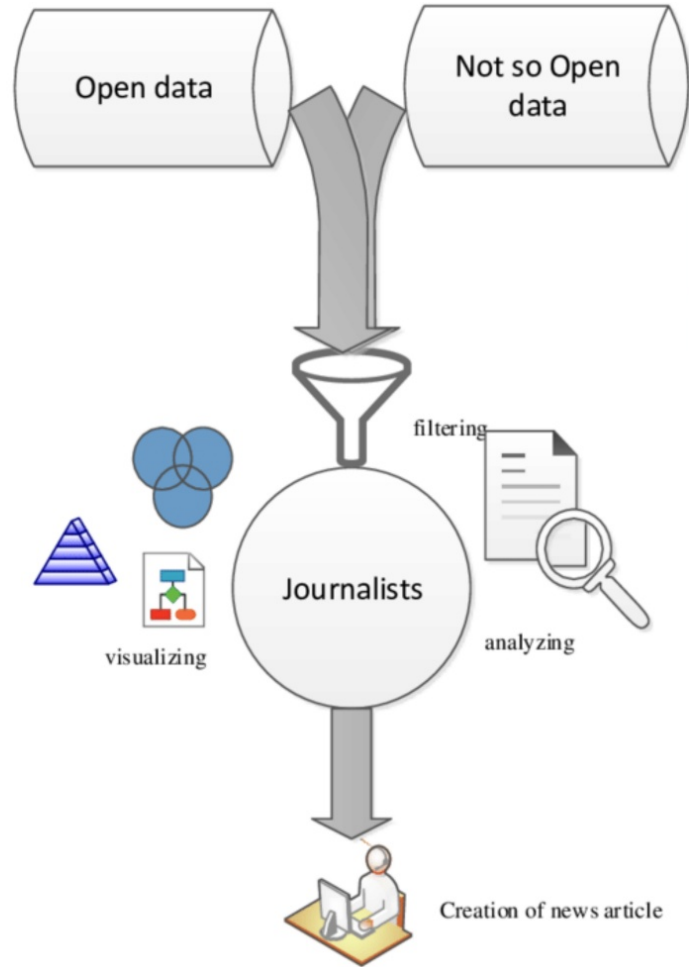
## Traditional vs DATA Journalism Practice



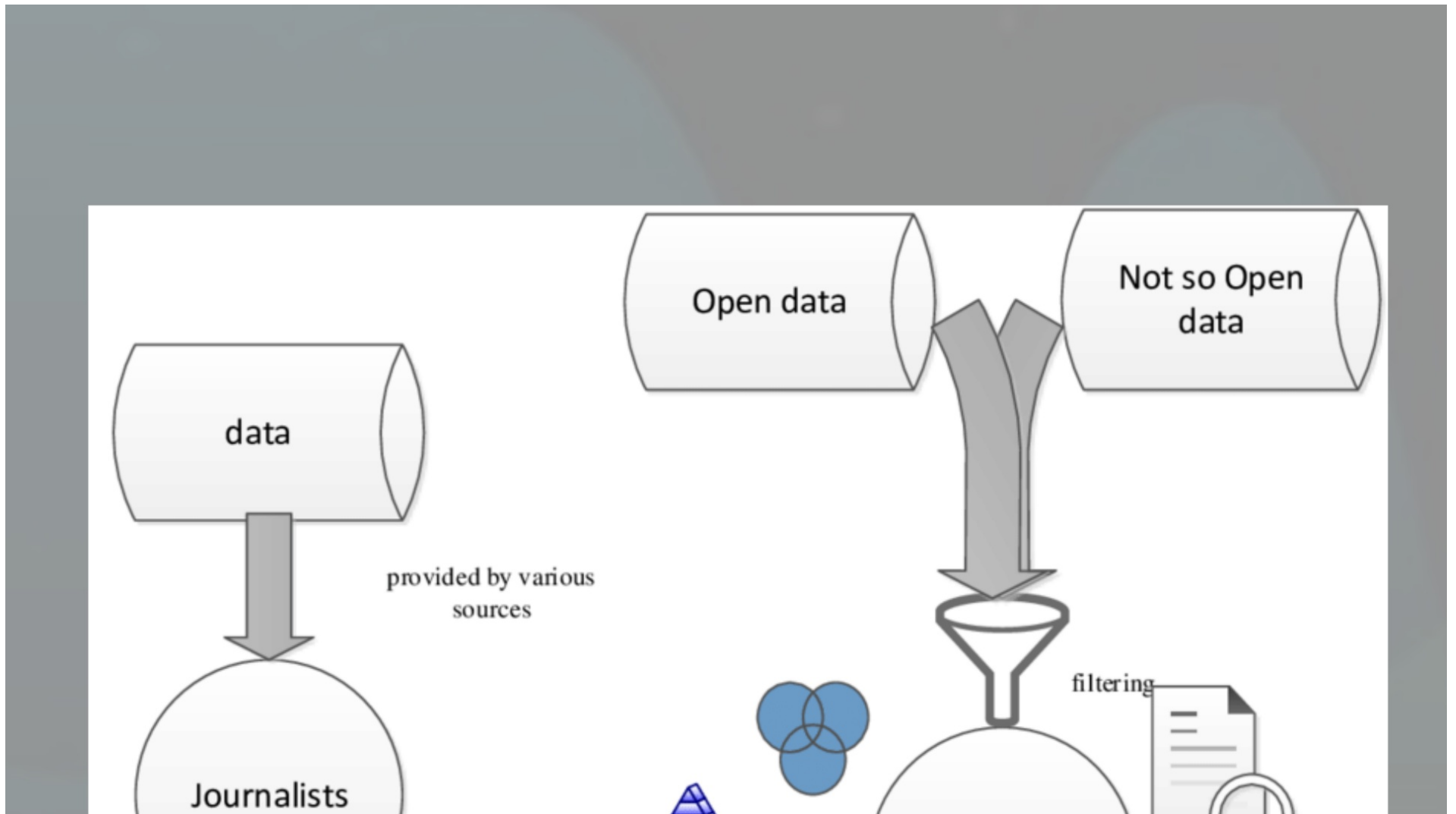
on  
s  
alism



Traditional journalism practice



Data Journalism practice





Traditional journalism practice

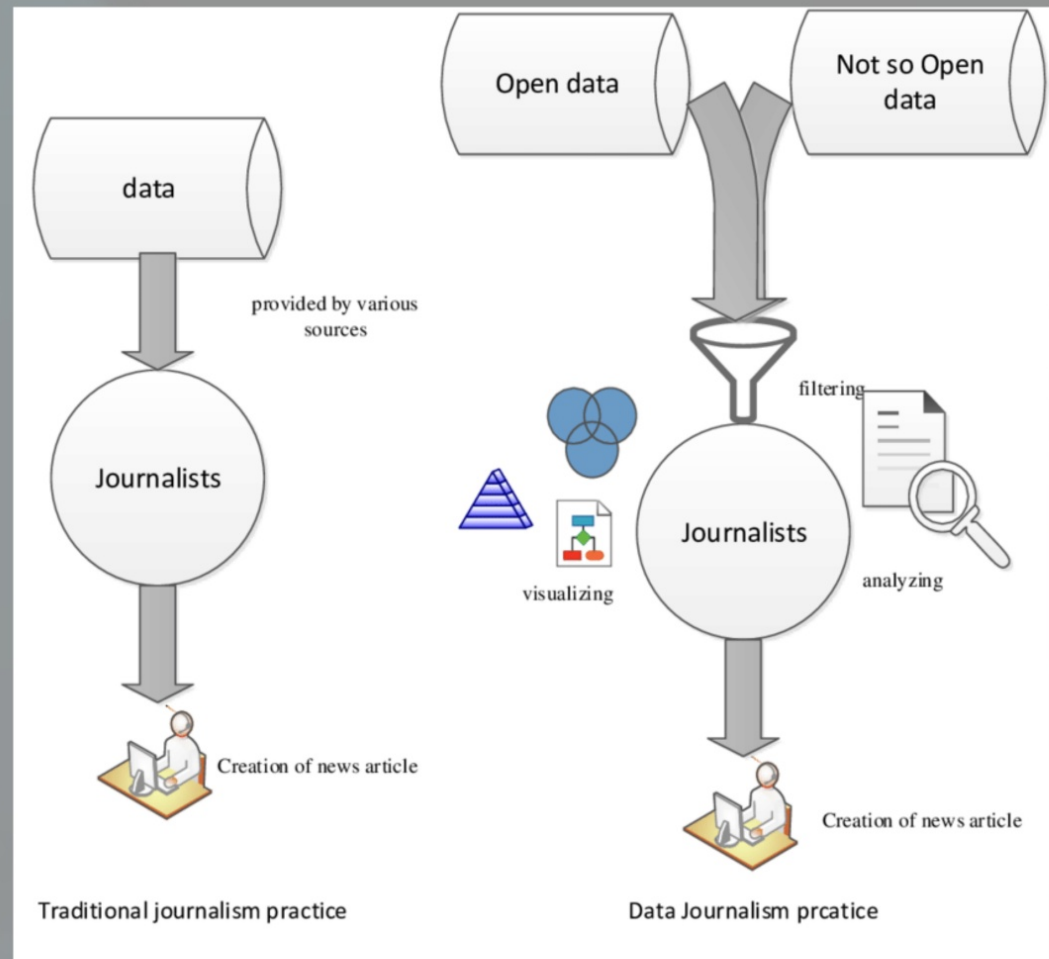


Data Journalism practice



# Comparison

## Traditional vs DATA Journalism Practice



# DATA Pipeline



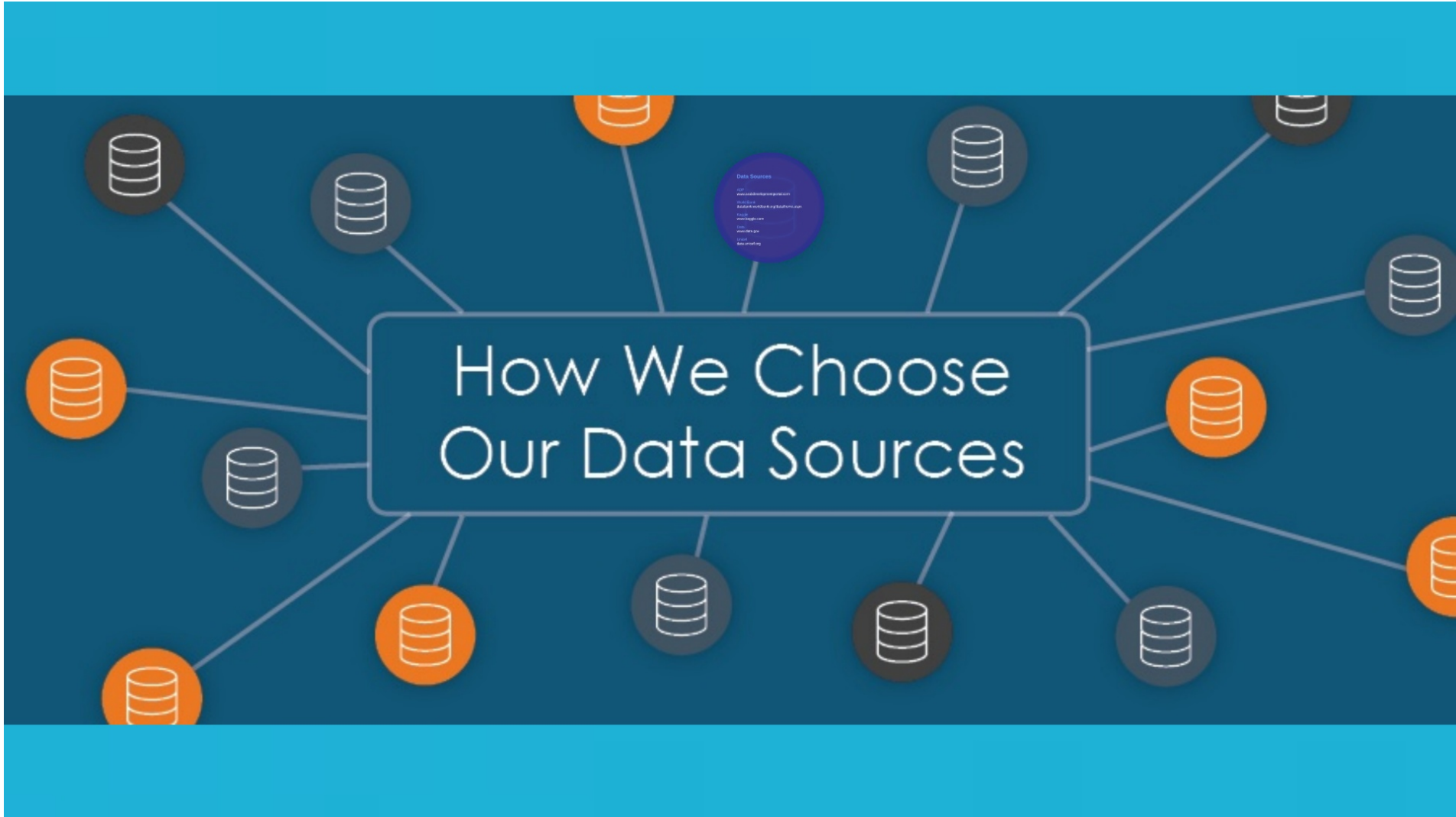


# SUSTAINABLE DEVELOPMENT GOALS



# DATA Pipeline





## Data Sources

ADP

[www.arabdevelopmentportal.com](http://www.arabdevelopmentportal.com)

World Bank

[databank.worldbank.org/data/home.aspx](http://databank.worldbank.org/data/home.aspx)

Kaggle

[www.kaggle.com](http://www.kaggle.com)

Data

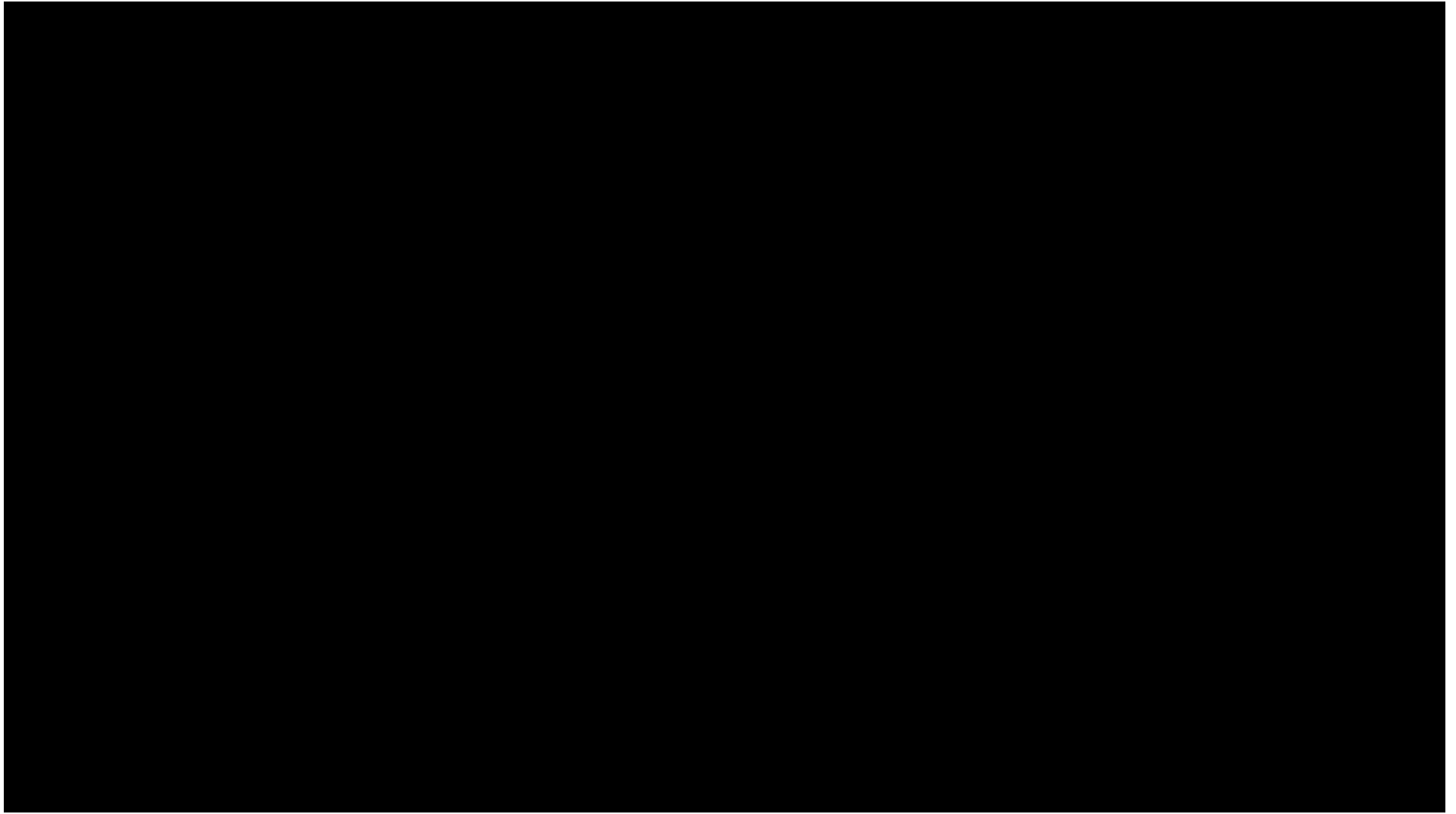
[www.data.gov](http://www.data.gov)

Unicef

[data.unicef.org](http://data.unicef.org)

# DATA Pipeline







## Step 1: Identify your Data Source

## Step 1: Identify your Data Source

Government

## Step 1: Identify your Data Source

Government

Organizations

## Step 1: Identify your Data Source

Government

Organizations

Science

## **Step 1: Identify your Data Source**

Government

Organizations

Science

## **Step 2: Getting data in the format you need it**

## **Step 1: Identify your Data Source**

Government

Organizations

Science

## **Step 2: Getting data in the format you need it**

Json, text, excel, csv, quantitative, ..etc

# DATA Pipeline





Find and  
record errors

Perform edits  
or note  
exceptions

Acceptable or  
unacceptable



# DATA Pipeline





## BAD DATA IS



Duplicate  
data



Missing  
information

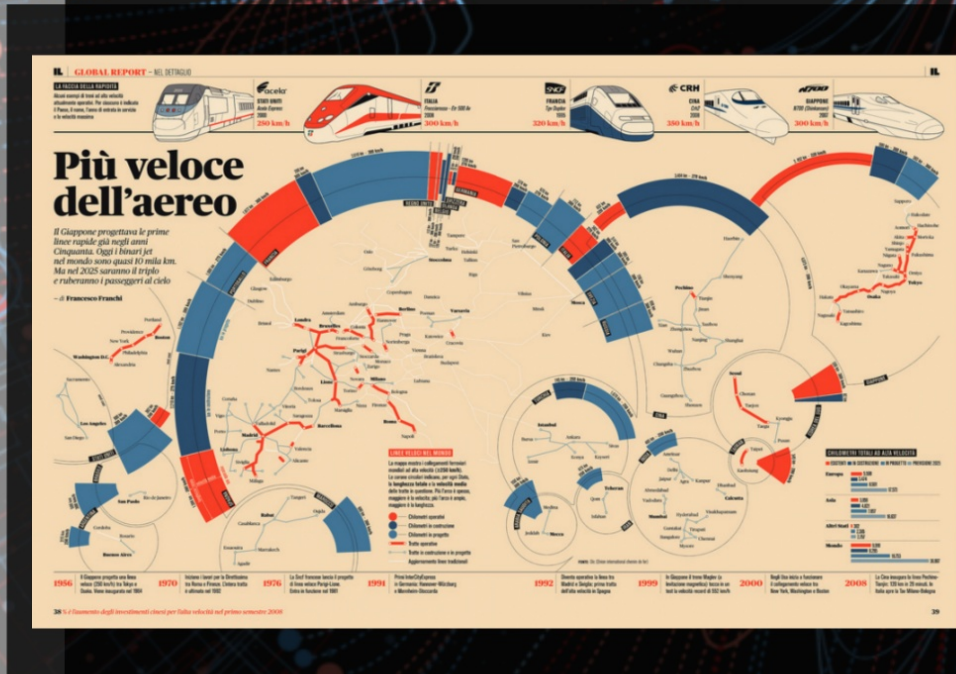


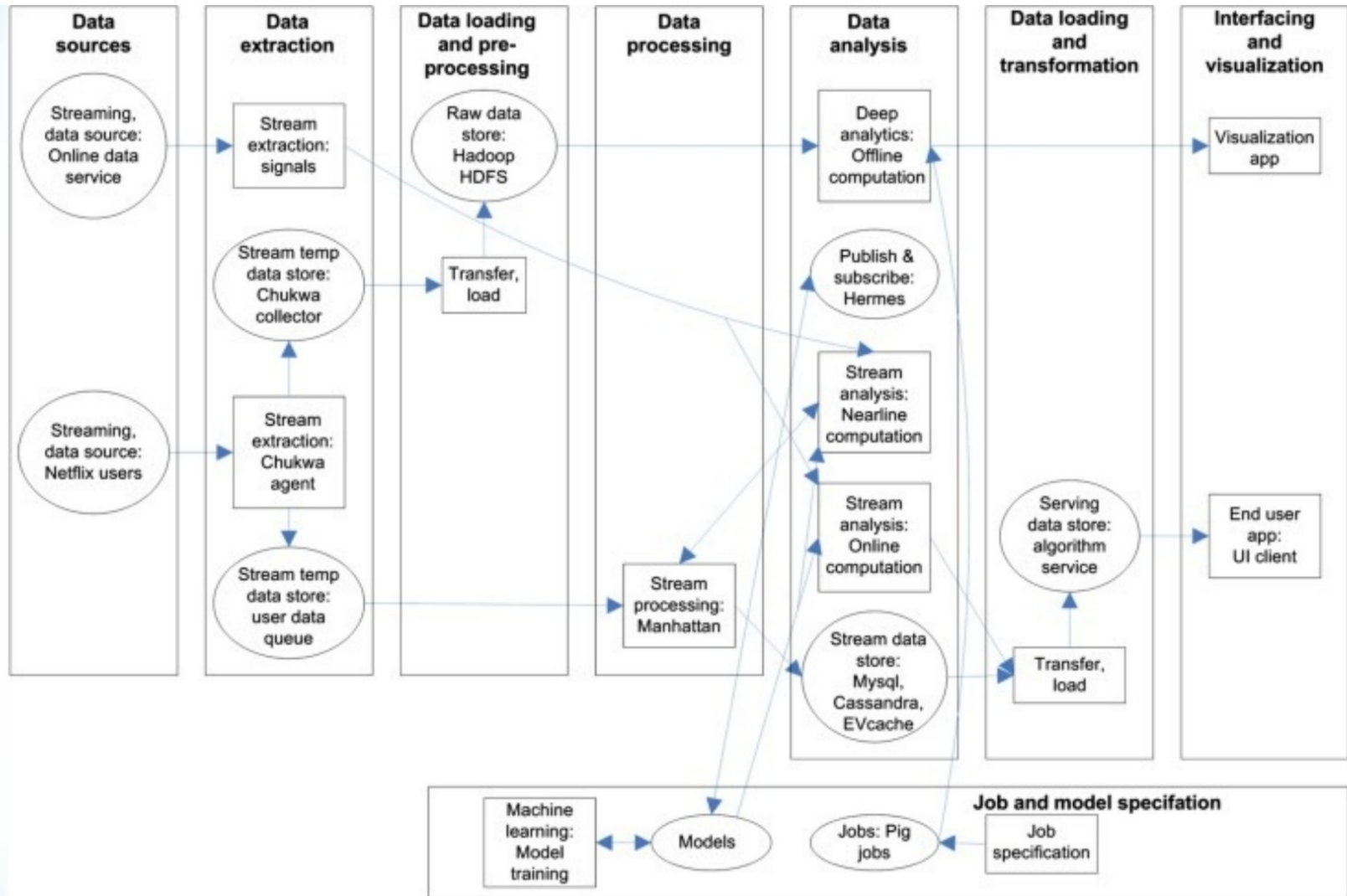
Inaccurate  
information



Incorrect  
information

# DATA Pipeline

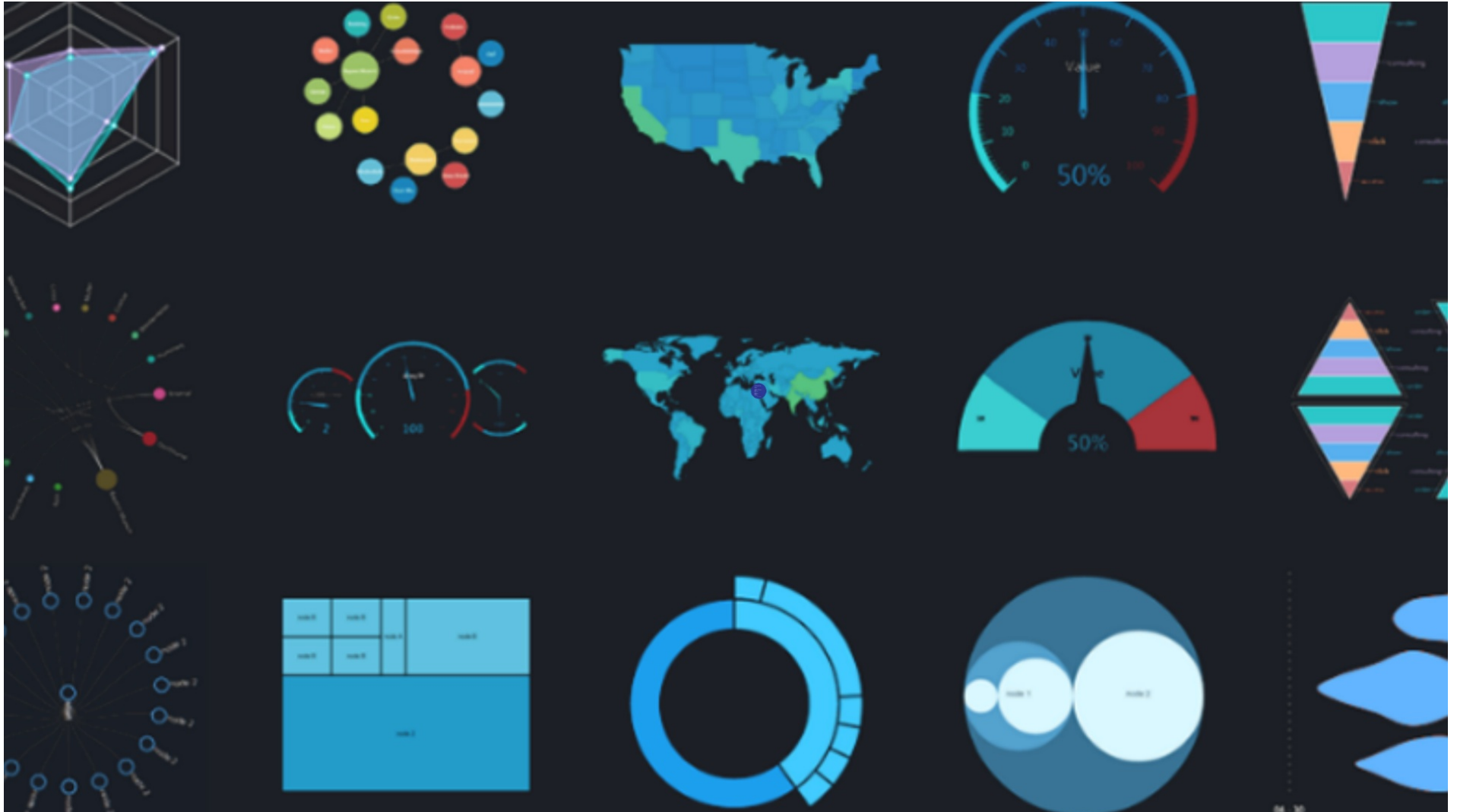




# DATA Pipeline









## GeoMapping Tools

**CARTO**

[www.carto.com](http://www.carto.com)

**Map Business Online**

[www.mapBusinessOnline.com](http://www.mapBusinessOnline.com)

**ArcGIS**

[www.arcgis.com](http://www.arcgis.com)

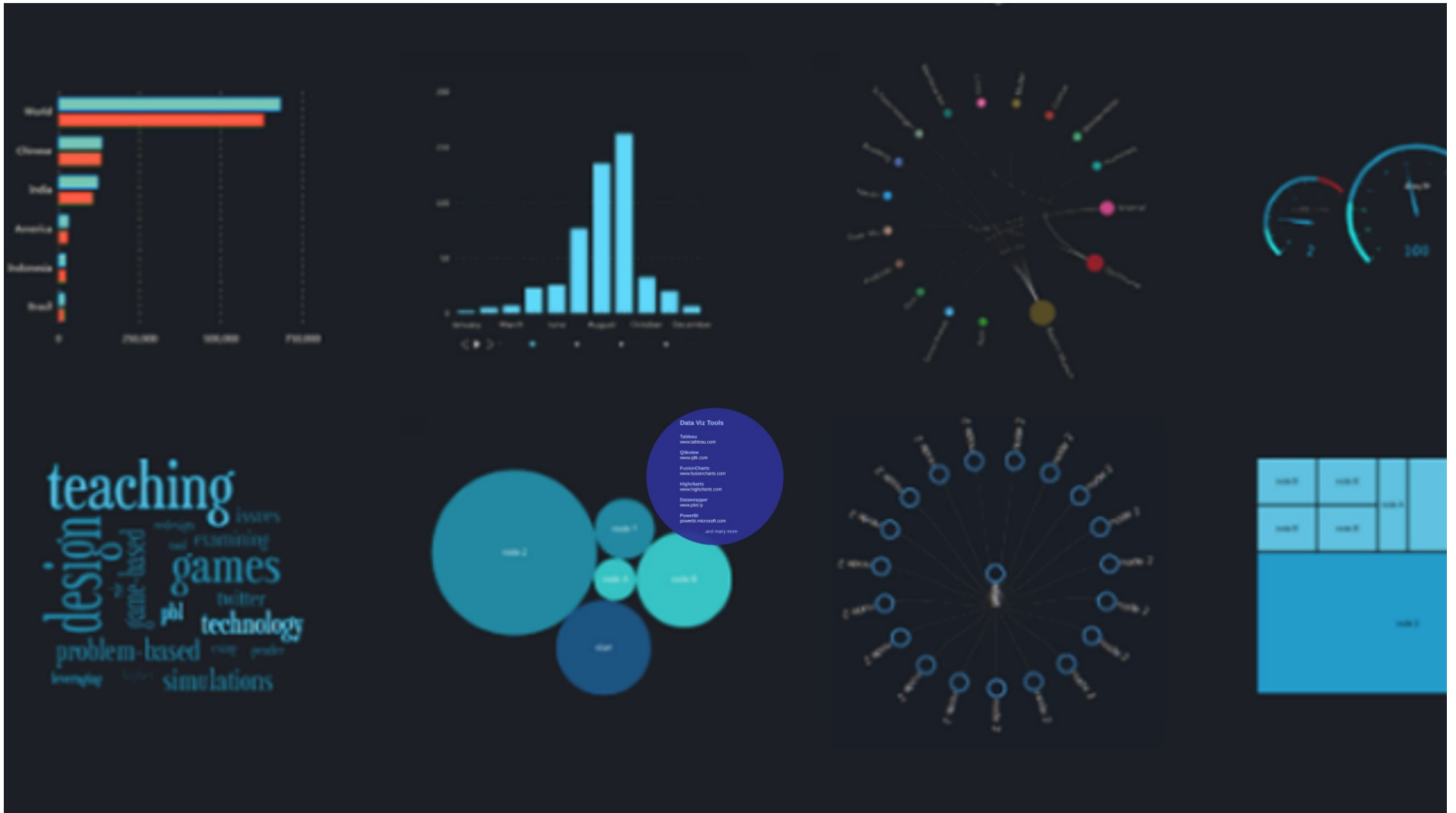
**eSpatial**

[www.espatial.com](http://www.espatial.com)

**Caliper**

[www.caliper.com](http://www.caliper.com)

..and many more



## Data Viz Tools

**Tableau**  
[www.tableau.com](http://www.tableau.com)

**Qlikview**  
[www.qlik.com](http://www.qlik.com)

**FusionCharts**  
[www.fusioncharts.com](http://www.fusioncharts.com)

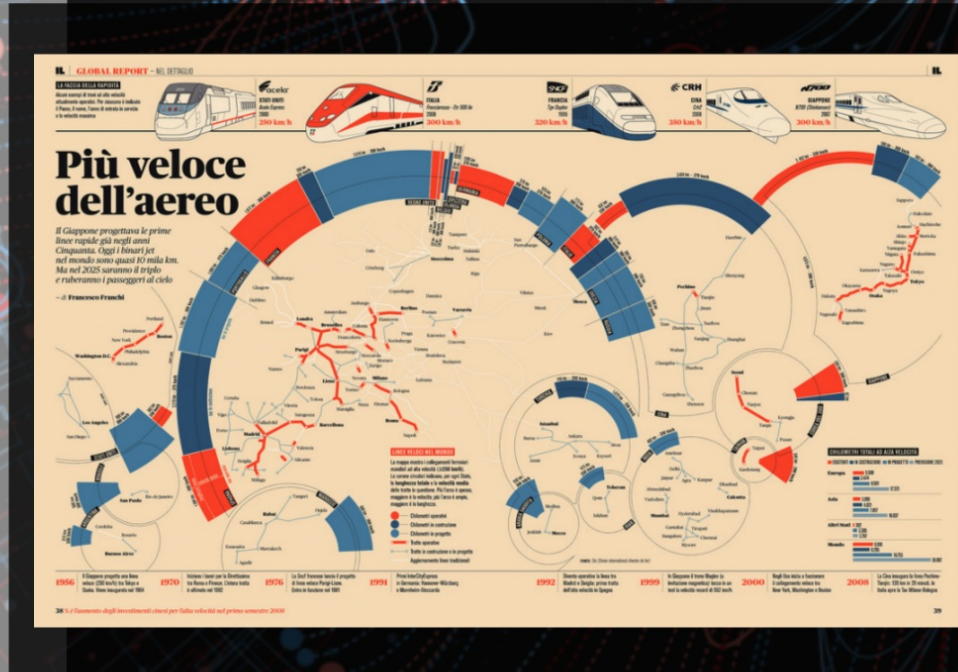
**Highcharts**  
[www.highcharts.com](http://www.highcharts.com)

**Datawrapper**  
[www.plot.ly](http://www.plot.ly)

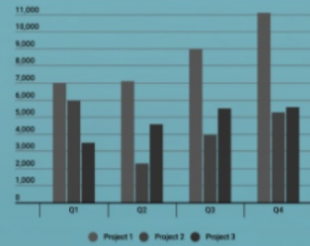
**PowerBI**  
[powerbi.microsoft.com](http://powerbi.microsoft.com)

..and many more

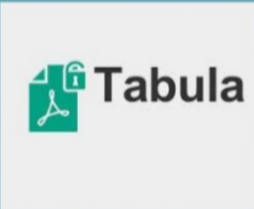
# DATA Pipeline



# Key wins



**CARTODB**  
Geospatial database on the cloud



## Math and Theory

- Statistics, Linear Algebra, Optimization, Time Series, etc.

## Applied Algorithms

- Machine Learning, Data Structures, Parallel Algorithms, etc.

## Engineering and Technologies

- Storage and computing platforms, statistical tools ,etc.

## Domain Expertise

- Text, Finance, Images, Econometrics etc.

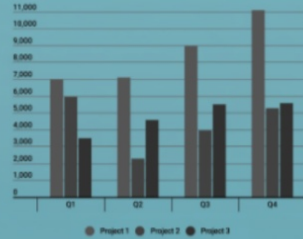
## Art

- Visualization, Infographics

## Best practices and hacks

- Handle missed values in data, transform and represent data, etc.

# Key wins



**CARTODB**  
Geospatial database on the cloud



+tableau++public



**TILEMILL**

## Math and Theory

- Statistics, Linear Algebra, Optimization, Time Series, etc.

## Applied Algorithms

- Machine Learning, Data Structures, Parallel Algorithms, etc.

## Engineering and Technologies

- Storage and computing platforms, statistical tools ,etc.

## Domain Expertise

- Text, Finance, Images, Econometrics etc.

## Art

- Visualization, Infographics

## Best practices and hacks

- Handle missed values in data, transform and represent data, etc.

# What is DATA?

Data Journalism E-Course (Session-3 Aug 11, 2021)



**1** What is  
Data  
Science?

**2** What is  
DATA?

**3** Data  
Pipeline

**4** USE  
CASES

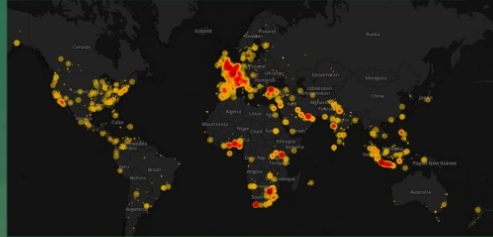
By ALI SROUR

Data Scientist and AI Consultant  
[/in/srouali](#)





## Nelson Mandela's Death Tweets

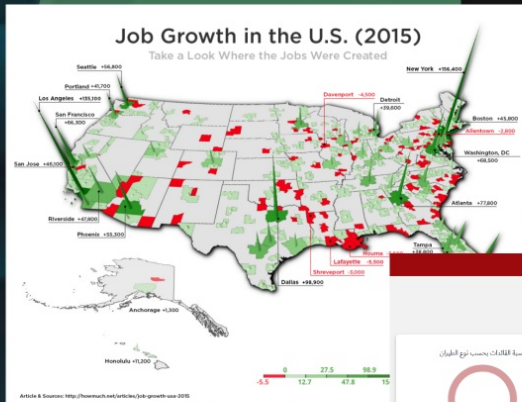




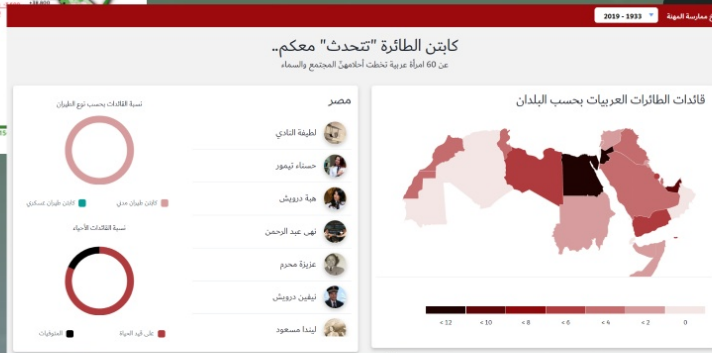
## Nelson Mandela's Death Tweets



## Mining Job Portals

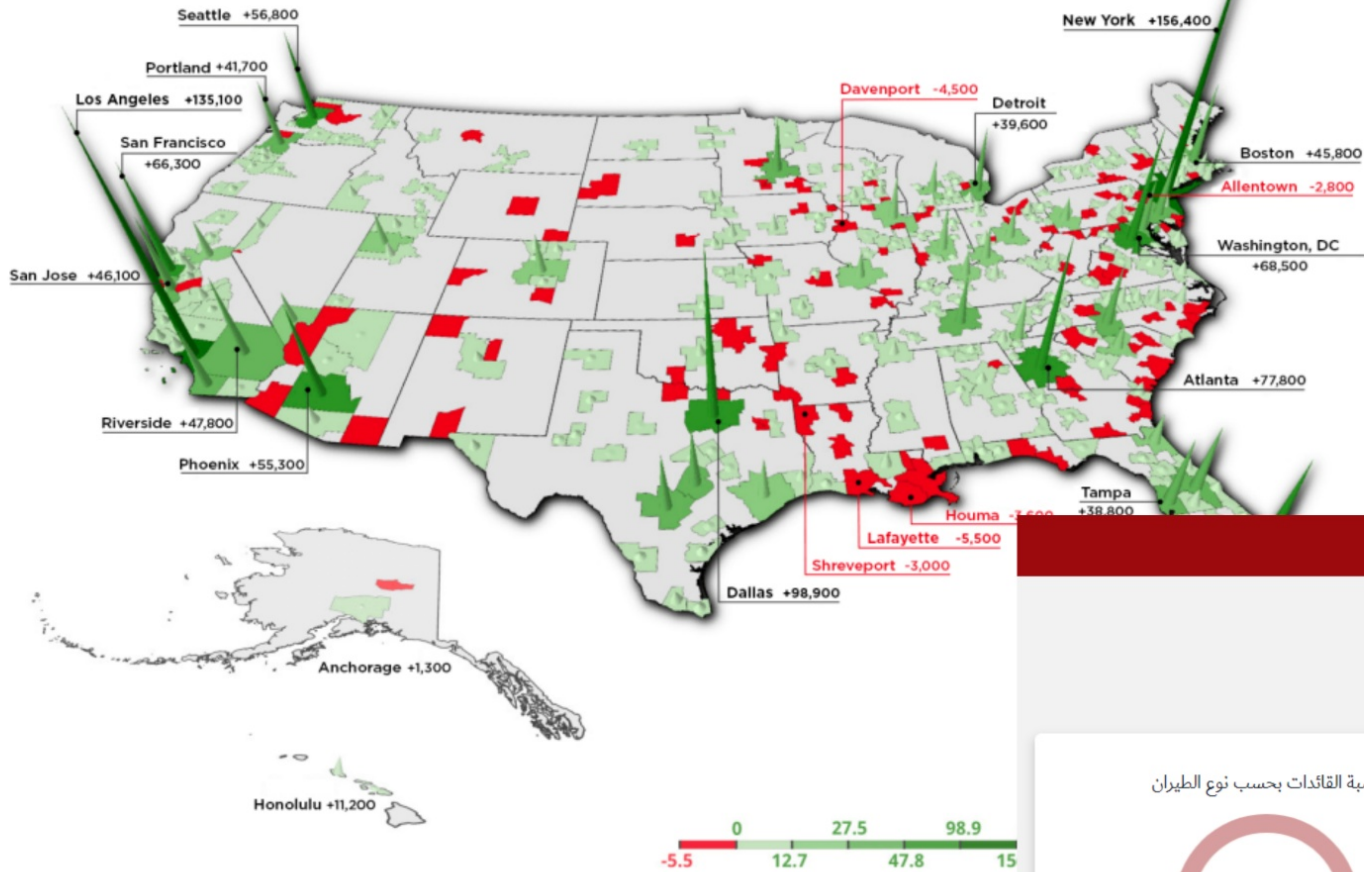


## Arab Pilots Story



# Job Growth in the U.S. (2015)

Take a Look Where the Jobs Were Created



Article & Sources: <http://howmuch.net/articles/job-growth-usa-2015>

نسبة القاءات بحسب نوع الطيران



# What is DATA?

Data Journalism E-Course (Session-3 Aug 11, 2021)



**1** What is  
Data  
Science?

**2** What is  
DATA?

**3** Data  
Pipeline

**4** USE  
CASES

By ALI SROUR

Data Scientist and AI Consultant  
[/in/srouali](#)