



Advanced usages of flux

Emanuele Della Valle

Prof. @ Politecnico di Milano

Founder & Partner @ Quantia Consulting

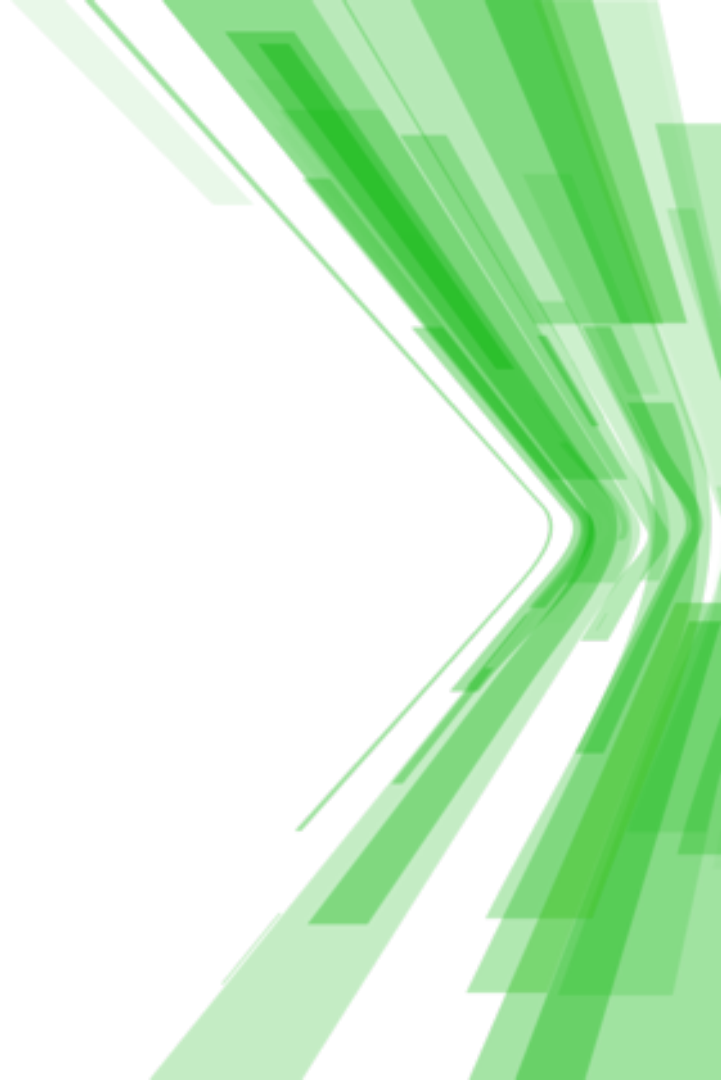
Marco Balduini

Founder & CEO @ Quantia Consulting

Senior Researcher @ Politecnico di Milano



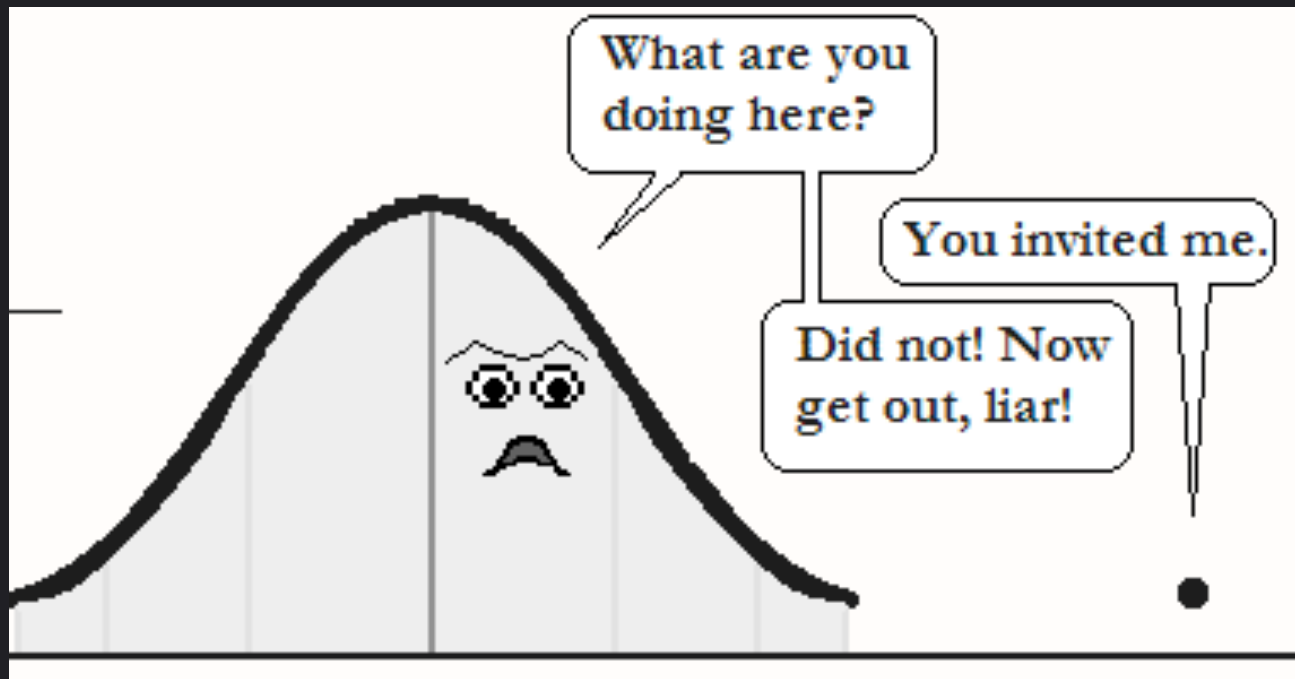
Anomaly Detection



Anomaly Detection

- Anomaly detection is a challenging data analysis task
- Anomalies can represent
 - spurious data to clean out
 - security breaches to notify
 - abnormal patterns to detect
- This generality makes anomaly detection a powerful tool used in network security, remote sensing, fault detection, and many other domains

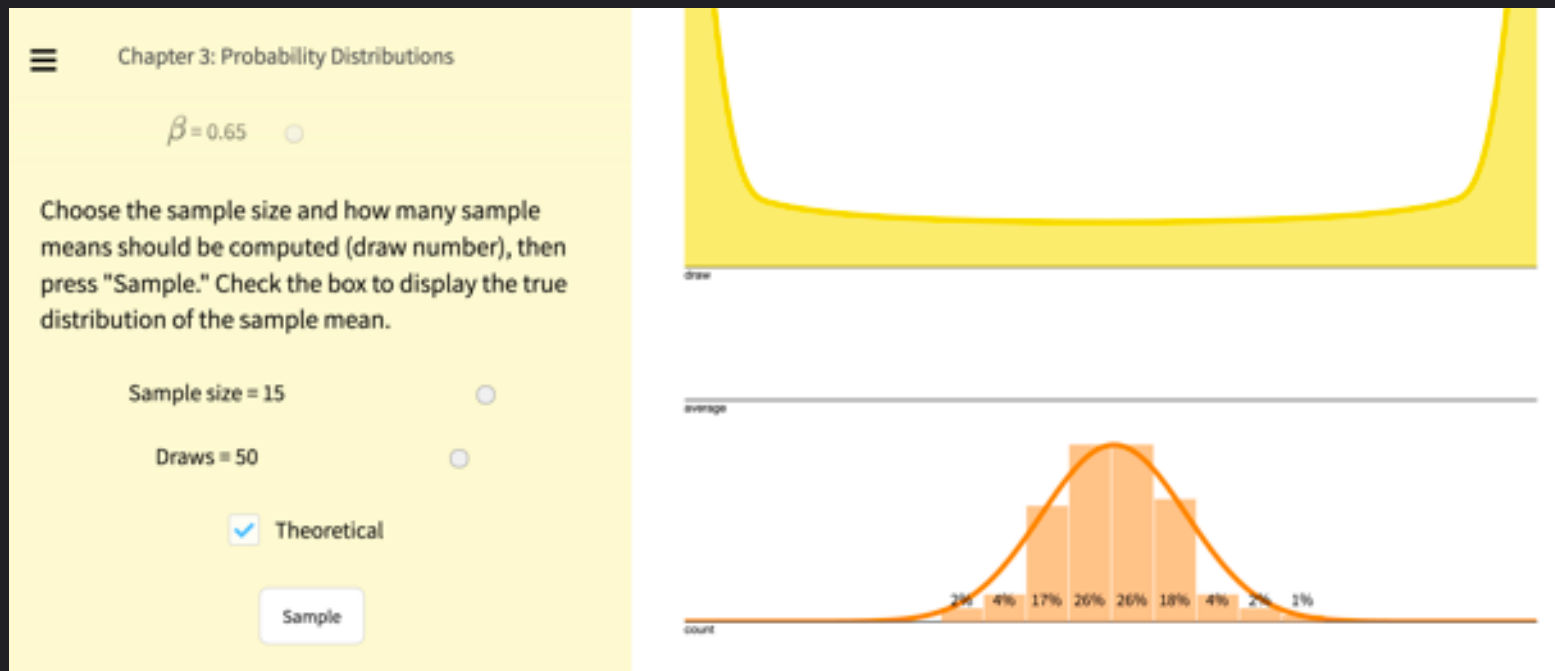
Anomaly Detection – the intuition



[image: <https://cerijayne.files.wordpress.com/2011/10/outliersss.png>]

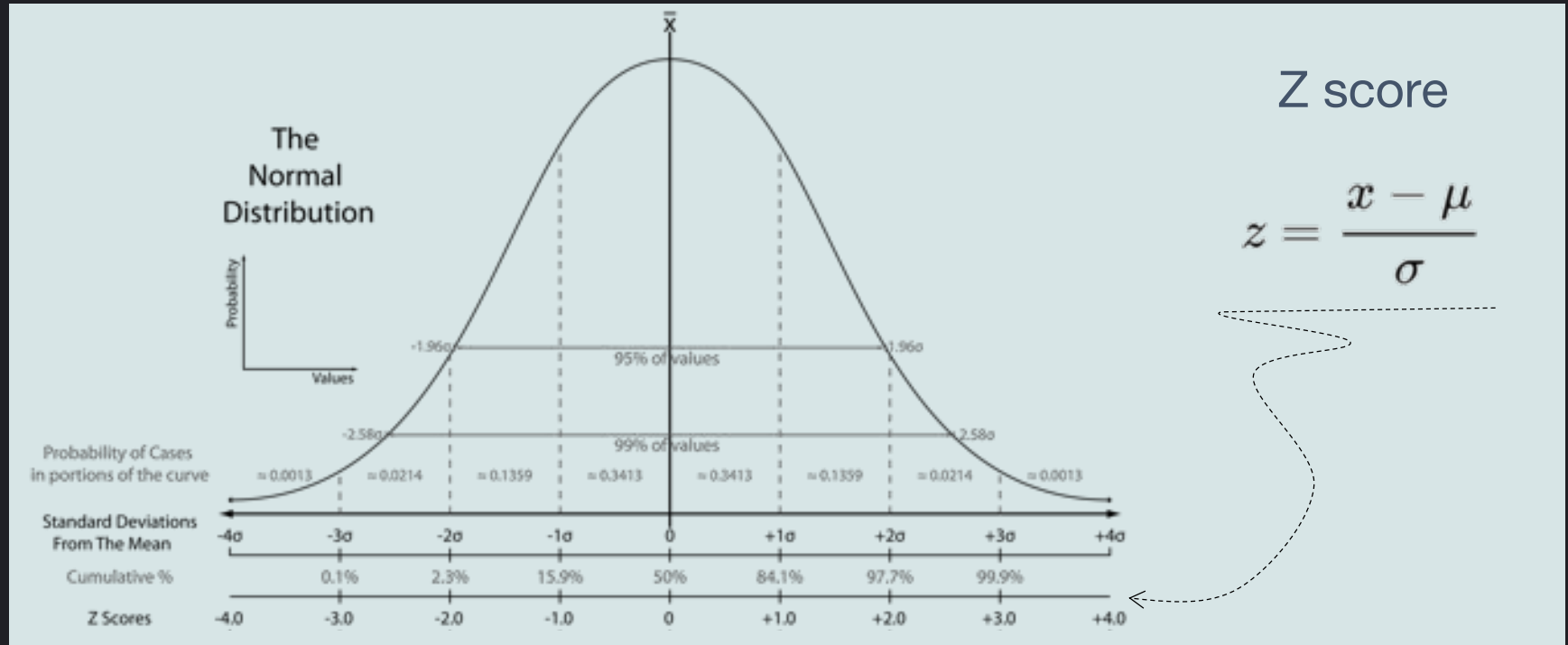
Considering the distribution of the data point in a time series, anomalous points are **outliers**

Anomaly Detection – a bit of theory

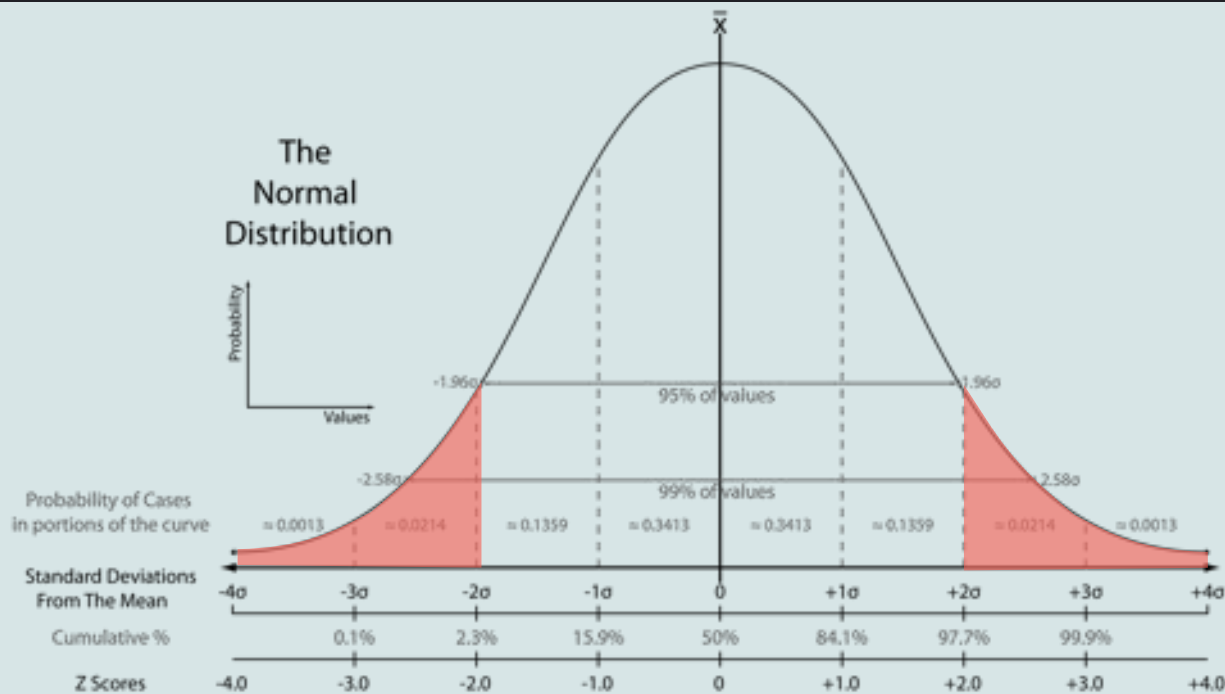


[src: <https://seeing-theory.brown.edu/probability-distributions/index.html#section3>]

Let's operationalize this intuition using the Z score



Let's operationalize this intuition using the Z score



Z score

$$z = \frac{x - \mu}{\sigma}$$

Anomalies

$$\text{Abs}(z) > 2$$

4,6% of the data

Let's get dirty!



13

Task

- Compute the zscore of the CPU user usage
 - What do we need?
 - Long time mean (60 mins)
 - Long time std (60 mins)
 - Short time mean (1 min)
- Find if the mean of CPU user usage in the last minute is anomalous

Take home message

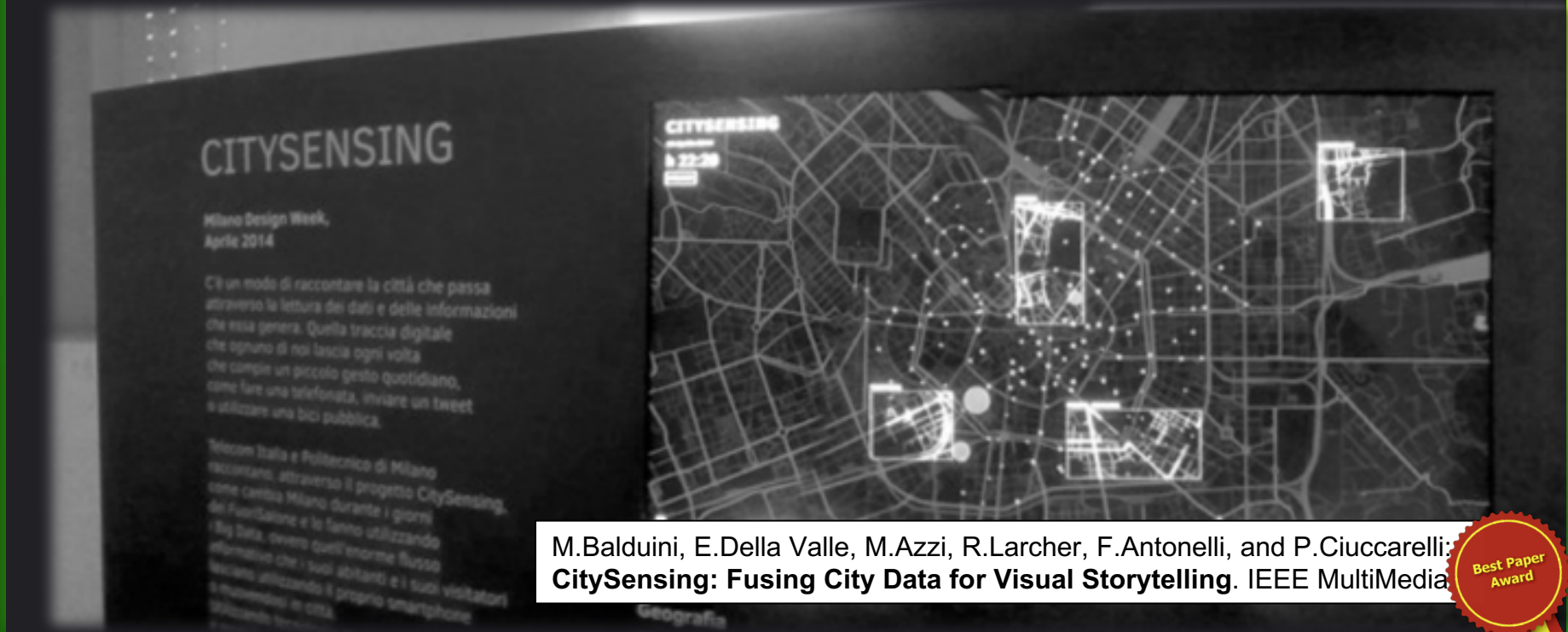
You learnt enough Flux to build an advanced continuous data analytics pipeline

A Network Security example

- Finding anomalies in the number of times users inserts a wrong password when they log into a system
 - Refuse the access to a user that inserts the right password at the first time if she gets in average the right password after 3 attempts
- Much more effective than any rule based approach



City scale crowd monitoring (Milan Design Week 13.4.2013)

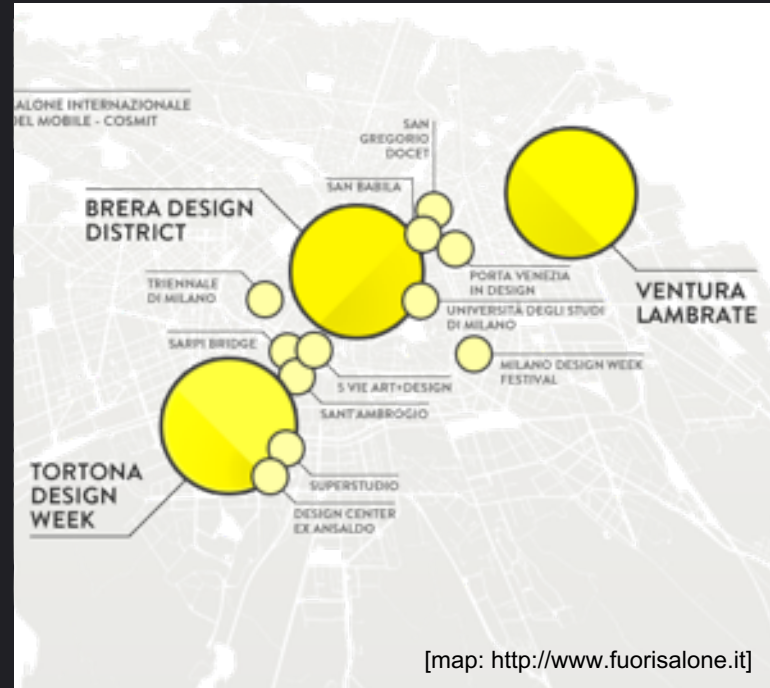


M.Balduini, E.Della Valle, M.Azzi, R.Larcher, F.Antonelli, and P.Ciuccarelli:
CitySensing: Fusing City Data for Visual Storytelling. IEEE MultiMedia

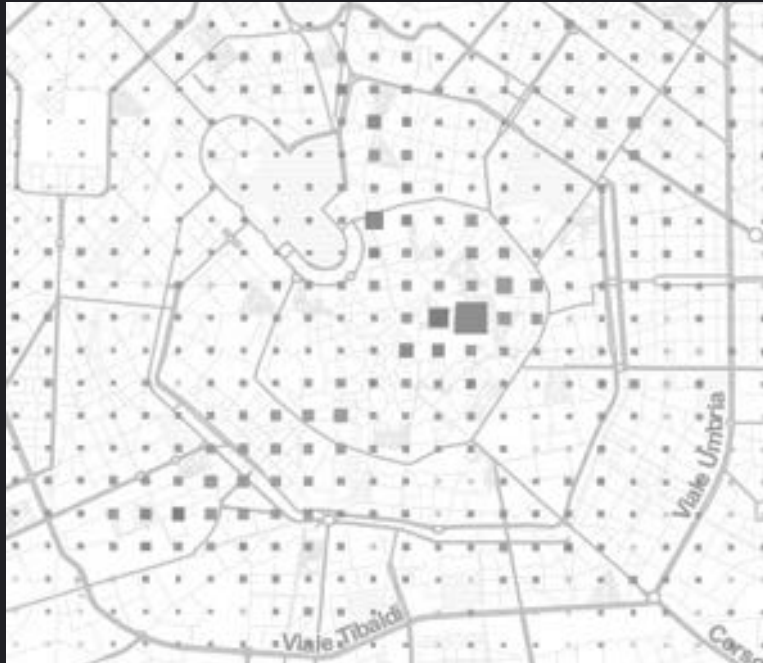
What's Milan Design Week?

The Milan Design Week (MDW) is a city-scale event

- held yearly in Milan,
- featuring around 1,200 events
- in 500+ places spread across the city and
- attracting about half a million people from all over the world.



City scale crowd monitoring (Milan Design Week 13.4.2013)



Brera
Design
District

Tortona
Design
Week

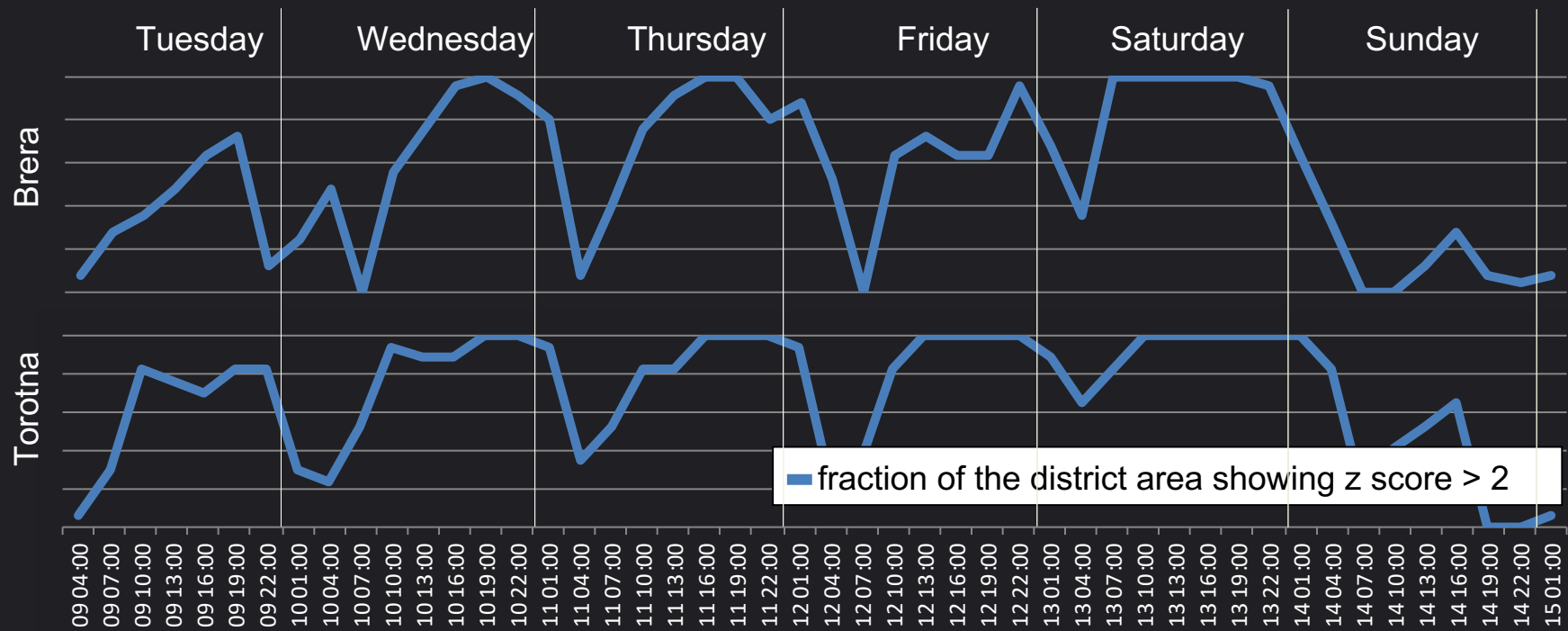


– number of mobile calls

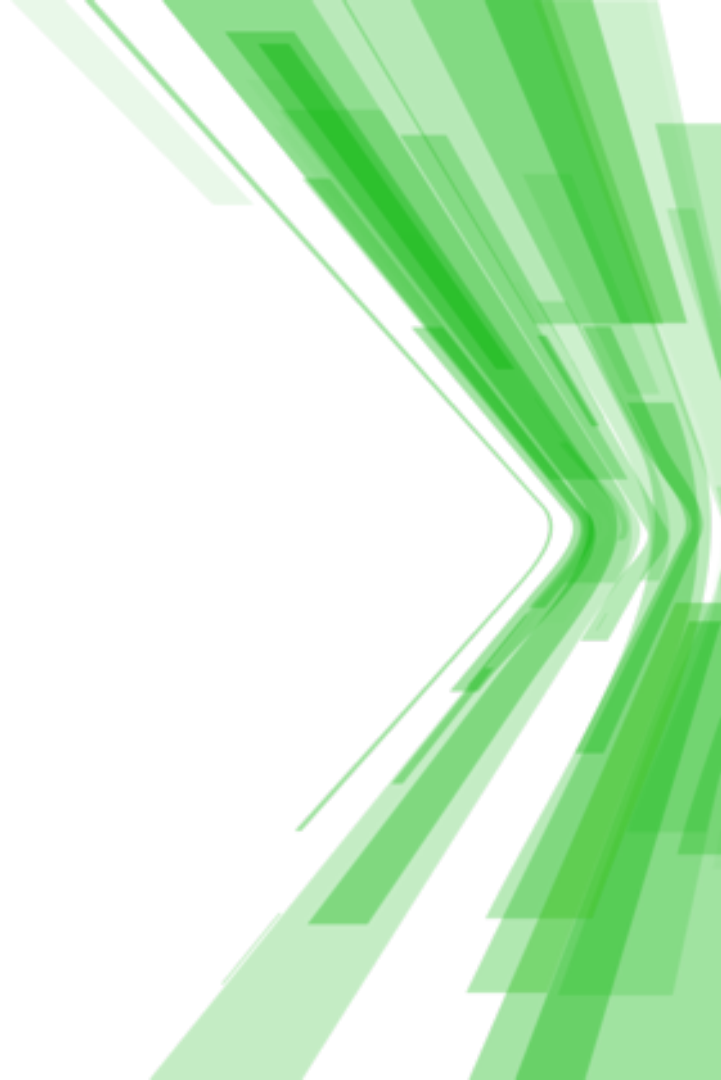
E. Della Valle

z-score of mobile calls

City scale crowd monitoring (Milan Design Week 13.4.2013)



Advanced Time Series Analytics



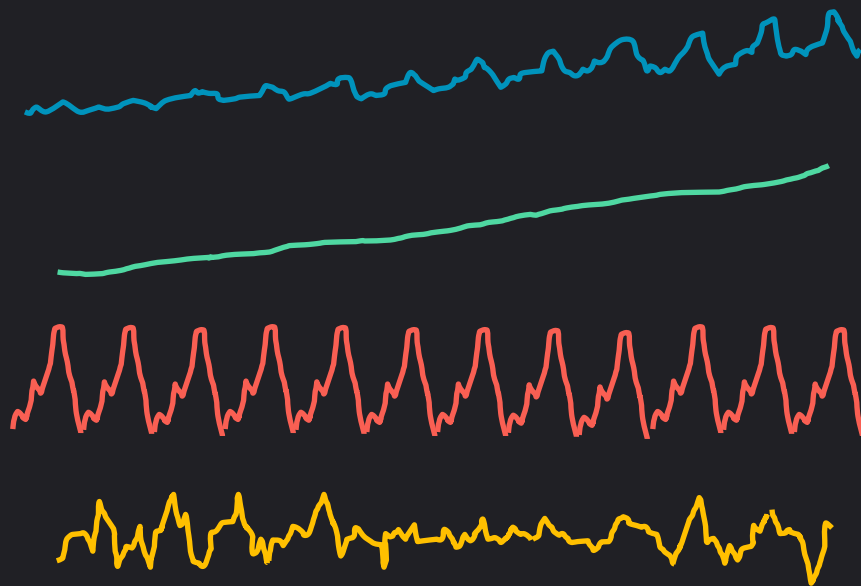
Holt-Winter's method

Given a **time series**

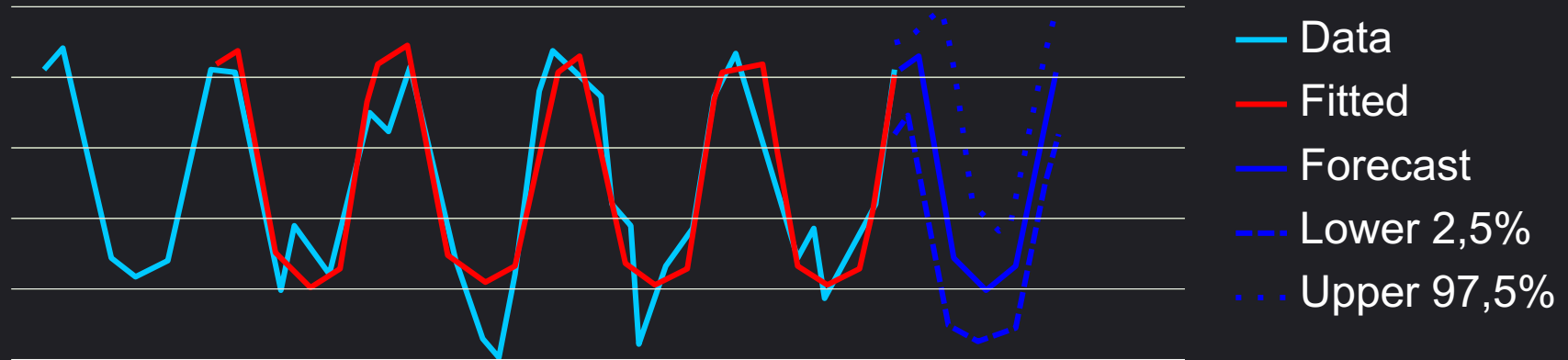
It is often possible to extract

- a **trend**: a long term variation of the time series.
- a **seasonality**: a short term recurring pattern

What is left is call **residual**



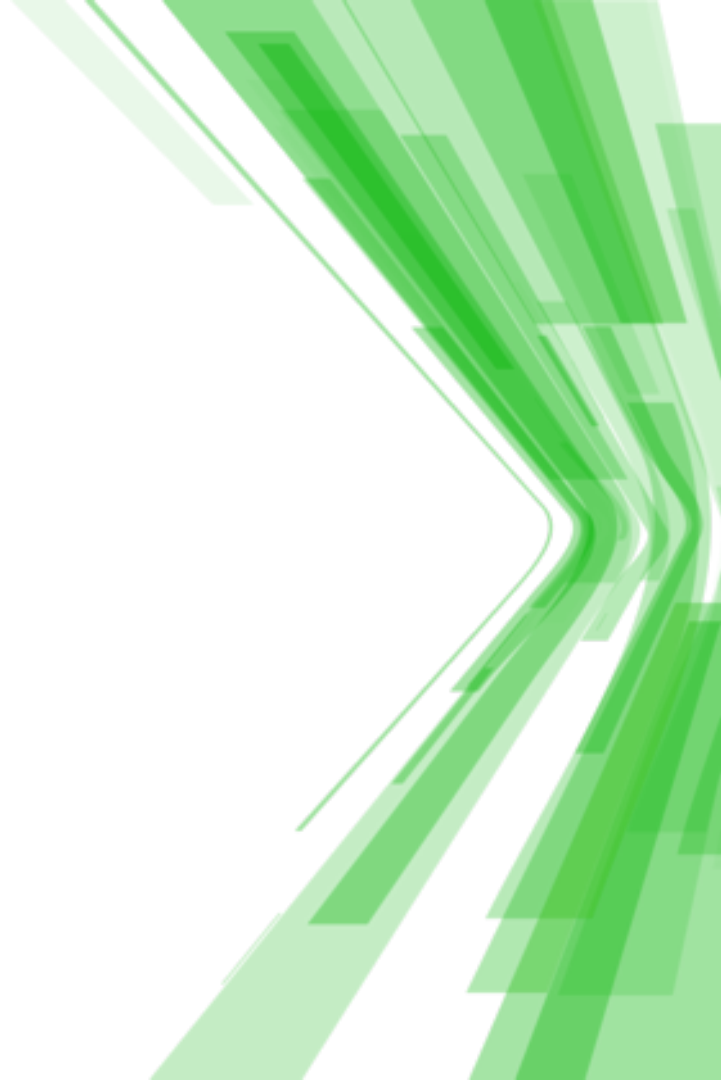
Forecasting with Holt-Winter's method



Holt-Winter's in flux



Time Series Enrichment



Time Series Enrichment – the intuition

- Combine data and perform additional calculations as needed

```
review,  
user_id=4,channel=web  
stars=4,message="Surprisingly good!"  
1569232800000000000  
review,  
user_id=3,channel=iOS  
stars=1,message="Worst. Product. Ever!"  
1569233700000000000
```



USER_ID	GENDER	STATUS
4	Male	platinum
3	Female	bronze

enrichedReview,
channel=web,gender=Male,status=platinum stars=4,message="Sur... good!" 1569232800000000000

enrichedReview,
channel=iOS,gender=Female,status=bronze stars=1,message="Wor...ever!" 1569233700000000000

Time Series Enrichment in flux

The screenshot displays the InfluxDB Flux query editor interface. At the top, a table shows the result of a query, with columns representing different attributes and their values. Below the table, a Flux script is visible, detailing the data sources and the enrichment process. The script uses the `sql.from` function to fetch data from a MySQL database and the `from(bucket: "training")` function to access data from the 'training' bucket. It then uses `range`, `filter`, and `map` functions to process the data, and finally joins the two data sources using `join` and drops unnecessary columns with `drop`.

false	false	true	true	true	true
string	string	string	string	dateTime:RFC3339	dateTime:RFC3339
CLUB_STATUS	GENDER	_field	_measurement	_start	_stop
bronze	Female	stars	review	2019-10-01T12:00:00Z	2019-10-01T12:05:00Z
platinum	Male	stars	review	2019-10-01T12:00:00Z	2019-10-01T12:05:00Z

```
1
2
3 userInfo = sql.from(
4   driverName: "mysql",
5   dataSourceName: "mysql://root:demo@tcp(127.0.0.1:3306)/demo",
6   query: "SELECT ID AS float_user_id, GENDER, CLUB_STATUS FROM CUSTOMERS;"
7 )
8
9 review = from(bucket: "training")
10  |> range(start: 2019-10-01T12:00:00Z, stop: 2019-10-01T12:05:00Z)
11  |> filter(fn: (r) => r._measurement == "review")
12  |> filter(fn: (r) => r._field == "stars")
13  |> map(fn: (r) => ({ r with float_user_id: int(v: r.user_id) })
14
15
16 join(tables: {t1: userInfo, t2: review}, on: ["float_user_id"])
17  |> drop(columns: ["float_user_id", "user_id"])
18
```

What (else) can I do with Flux?



What (else) can I do with flux?

- Map/Reduce
- Derivative, Difference, Integral
- Histograms
- Regular expression matching
- String functions
- Go Math library parity
- Quantile
- Conditional Expressions
- Pivot tables
- Cross-measurement calculations
- Conditional expressions
- Sorting
- Data Generation
- In-language test assertions
- ...
- And more every day!



Advanced usages of flux

Emanuele Della Valle Prof. @ Politecnico di Milano & Partner @ Quantia Consulting

Marco Balduini Founder & CEO @ Quantia Consulting & Senior Researcher @ Politecnico di Milano

Where's the pizza?

