Towards Democratizing Relational Data Visualization

SIGMOD 2019 Tutorial
June 30, 2019
Amsterdam, The Netherlands

Nan Tang
Qatar Computing Research Institute
HBKU, Qatar Foundation

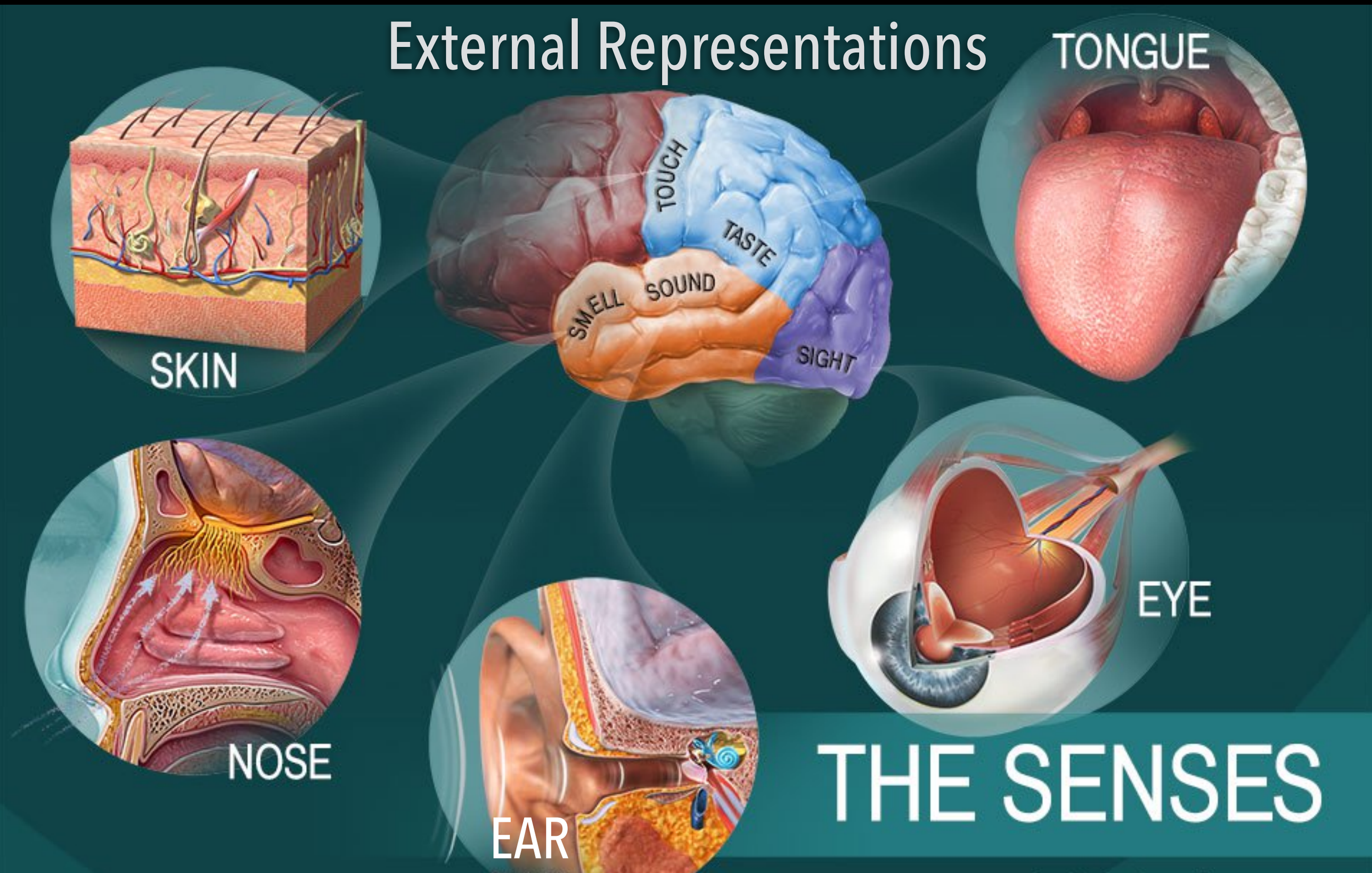
Eugene Wu
Computer Science
Columbia University

Guoliang Li
Computer Science and Technology
Tsinghua University

Outline

- **Nan:** Fundamentals and State-of-the-art (25-30 minutes)
  - ***why*** data visualization is so successful for *human-in-the-loop* data analytics
  - ***what*** are data visualizations
  - ***how*** have data visualizations been used
- **Eugene:** Efficient, Effective and Interactive Visualizations (60-65 minutes)
- **Guoliang:** Recommendation (~60 minutes)
- **Nan:** Uncertainty, collaborative, and immersive data visualizations (~30 minutes)

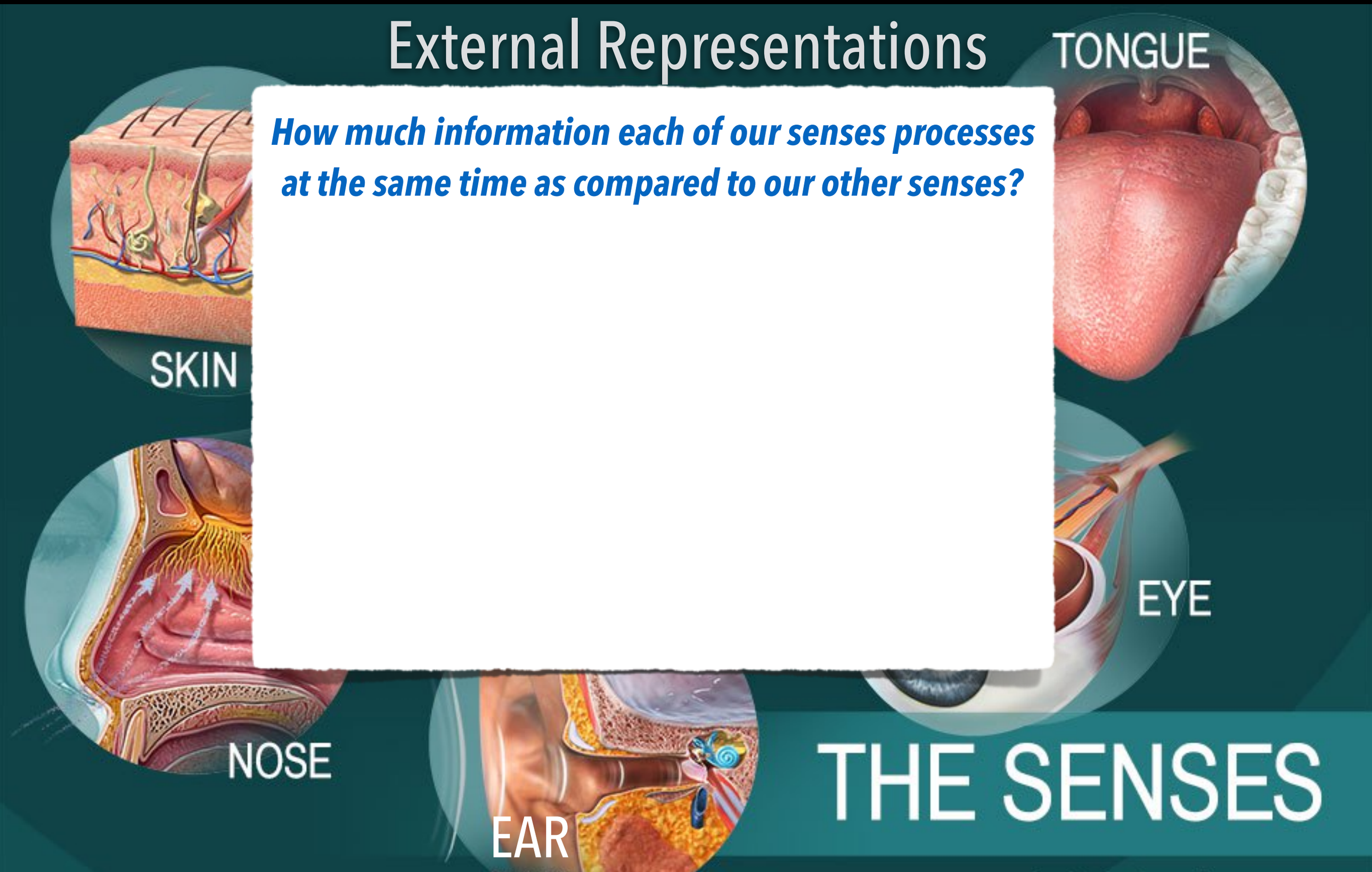
Sight > The Other Senses ?



Sight > The Other Senses ?

External Representations

How much information each of our senses processes at the same time as compared to our other senses?



Sight > The Other Senses ?

External Representations

How much information each of our senses processes at the same time as compared to our other senses?

Neuroscience and Cognitive Psychology

L.D. Rosenblum, Harold Stolovitch, Erica Keeps

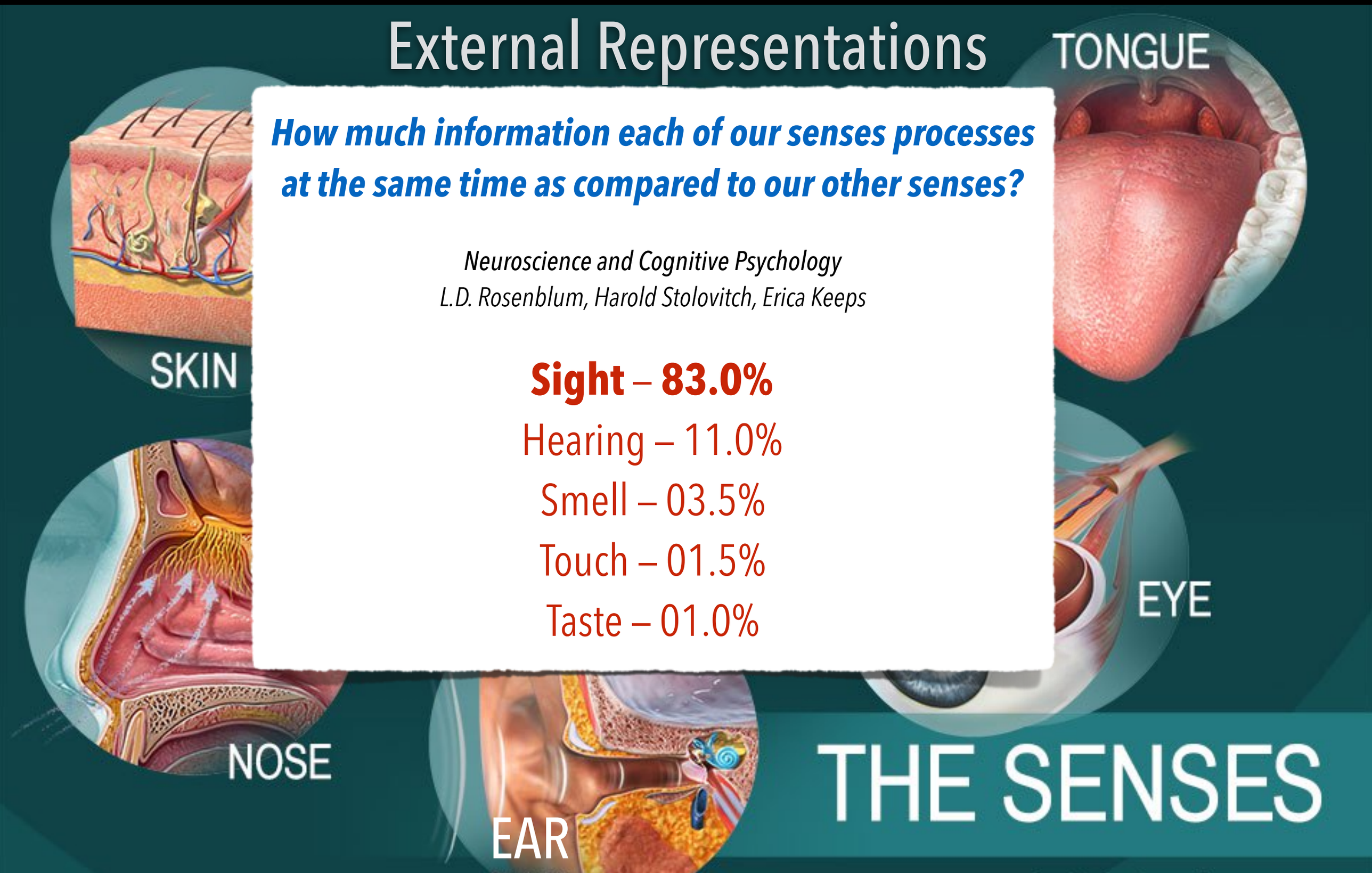
Sight – 83.0%

Hearing – 11.0%

Smell – 03.5%

Touch – 01.5%

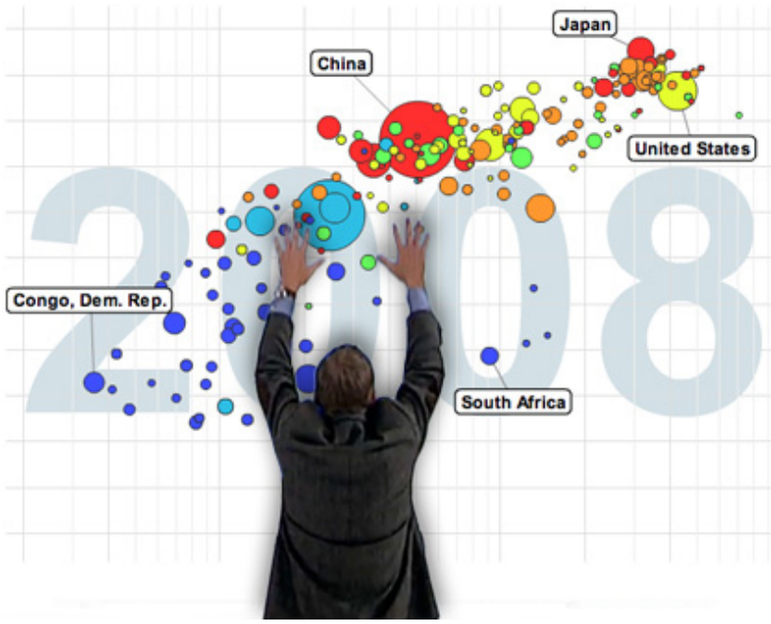
Taste – 01.0%



State-of-the-art

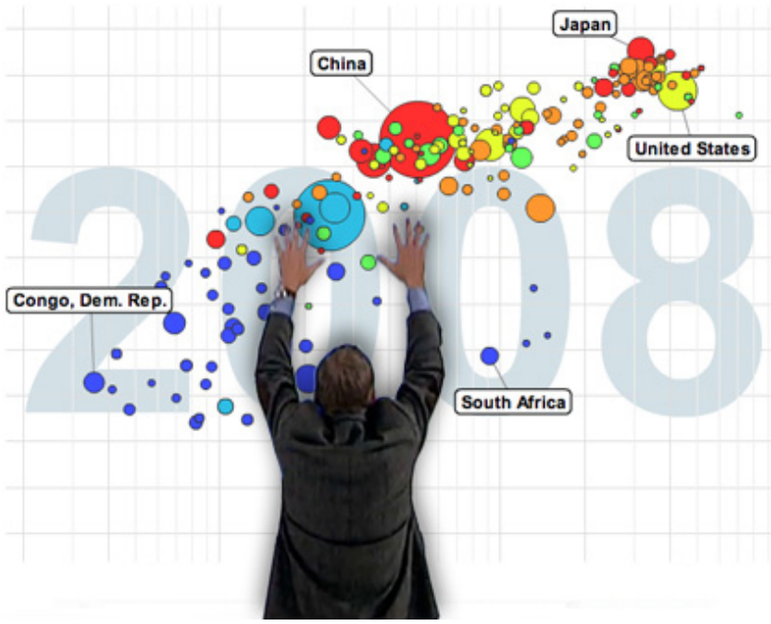
State-of-the-art

Storytelling

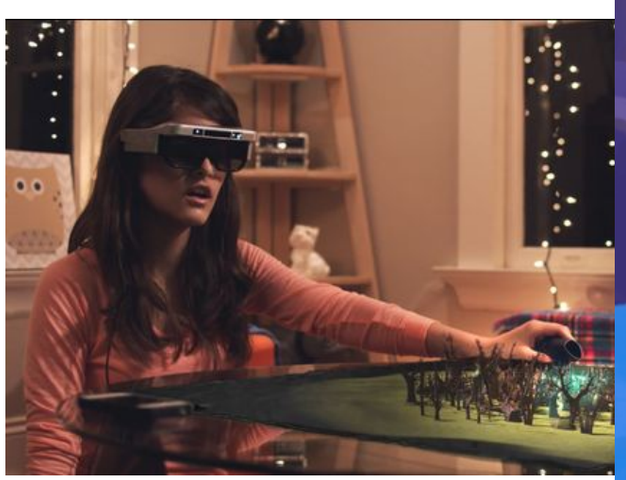


State-of-the-art

Storytelling

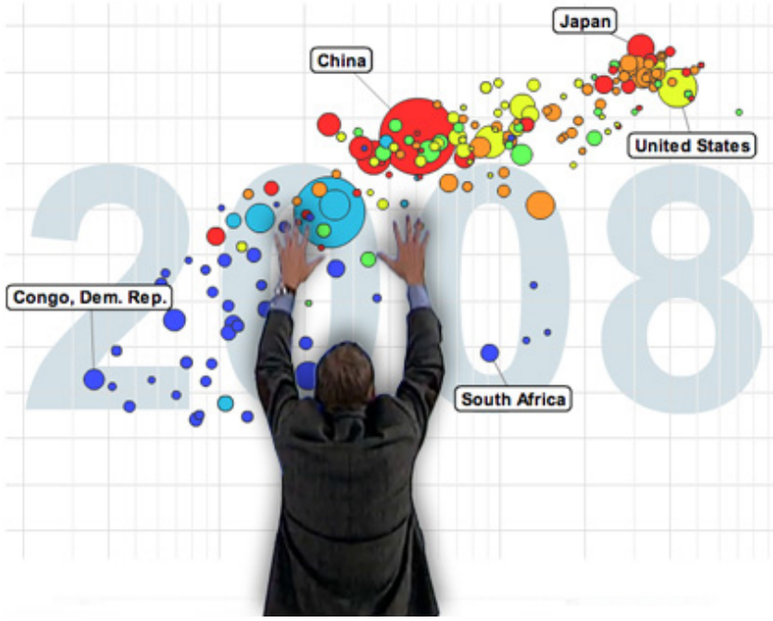


Virtual/Augmented/Mixed Reality

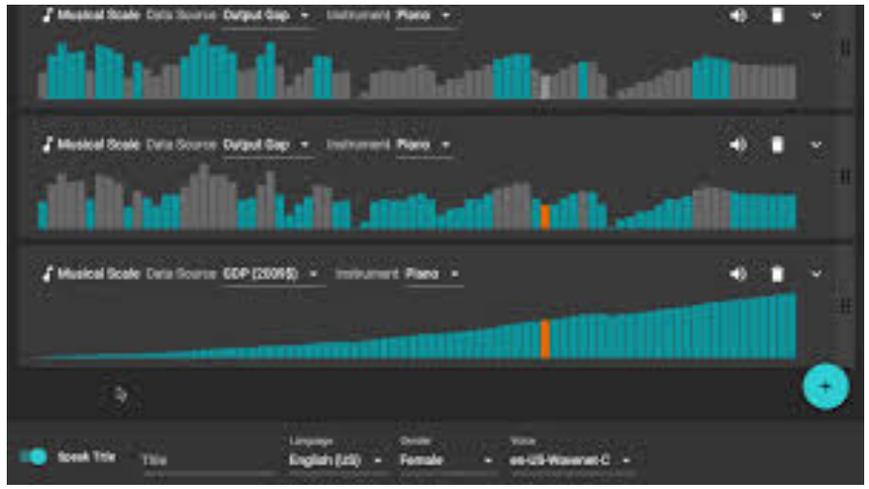


State-of-the-art

Storytelling



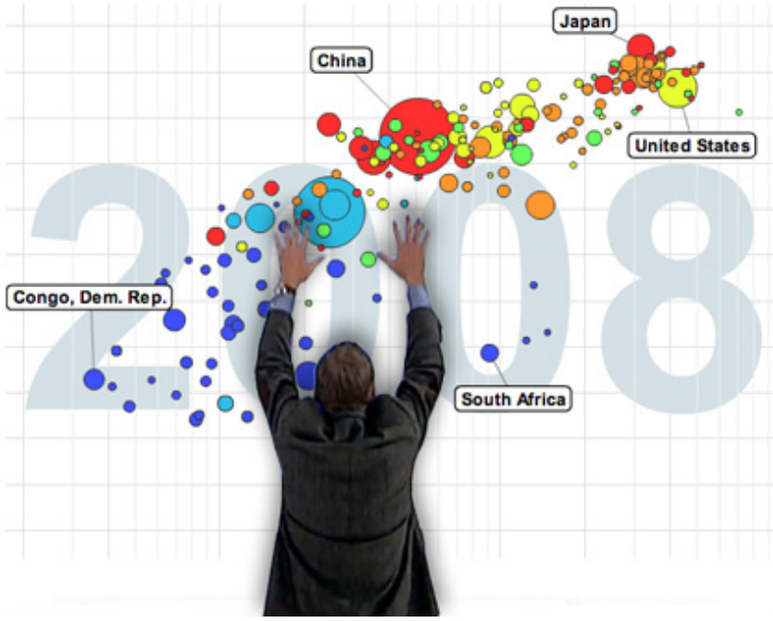
Sonification



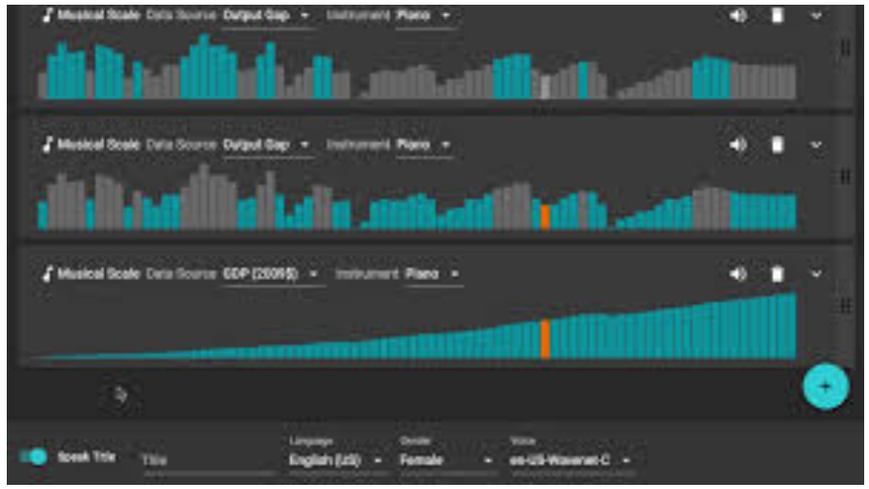
Virtual/Augmented/Mixed Reality

State-of-the-art

Storytelling



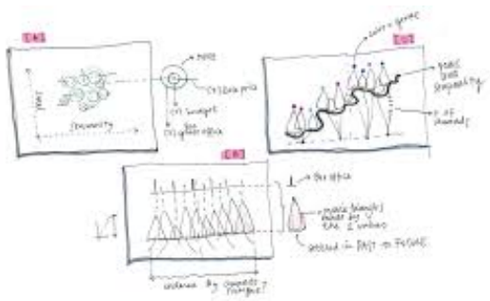
Sonification



Virtual/Augmented/Mixed Reality

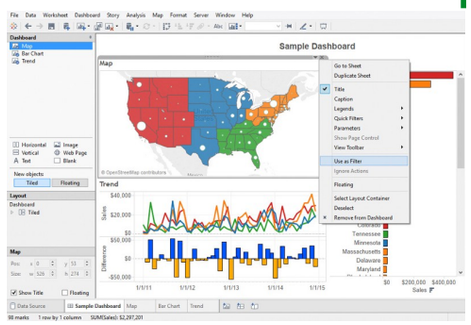
Physicalization

What and How



human

human



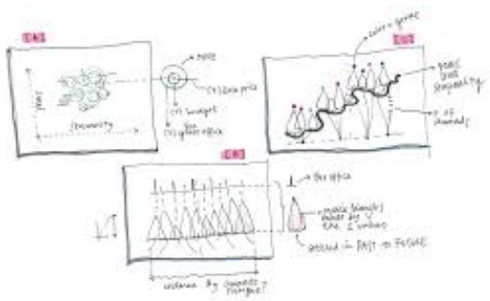
machine

human

machine

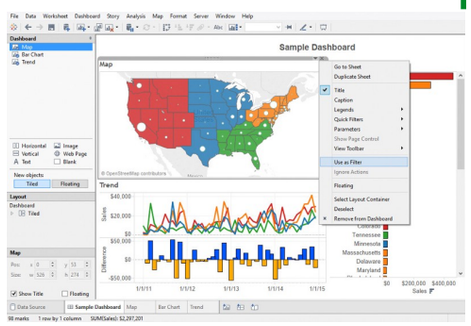
machine

What and How



human

human



machine

human

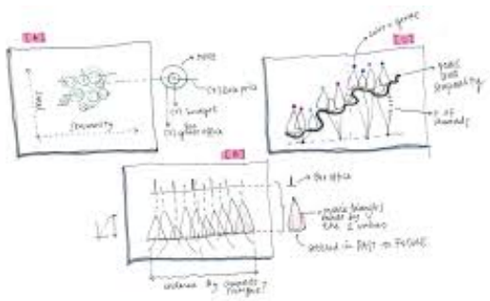
machine

X

machine

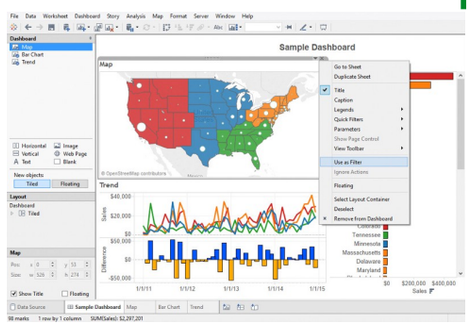
Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out **tasks** more **effectively**. – *Tamara Munzner at UBC*

What and How



human

human



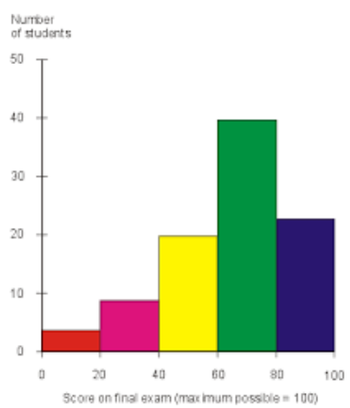
machine

human

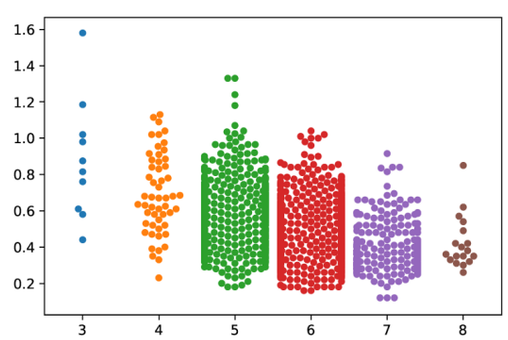
machine

machine

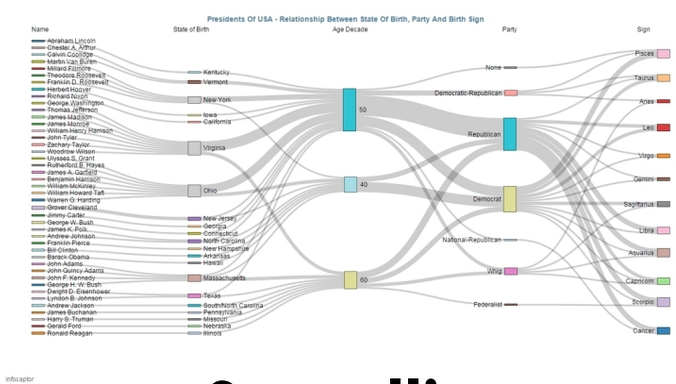
Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out **tasks** more **effectively**. – *Tamara Munzner at UBC*



Understanding

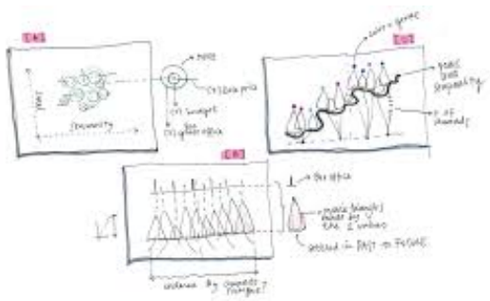


Exploratory



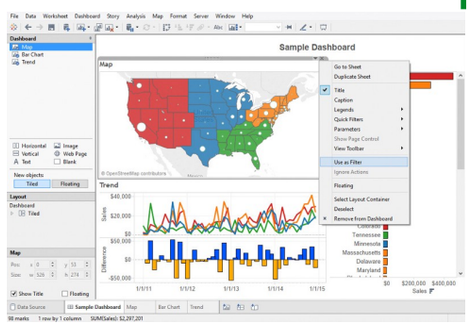
Storytelling

What and How



human

human



machine

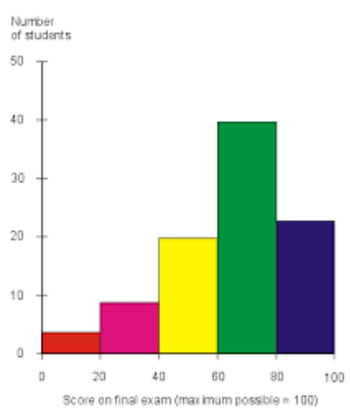
human

machine

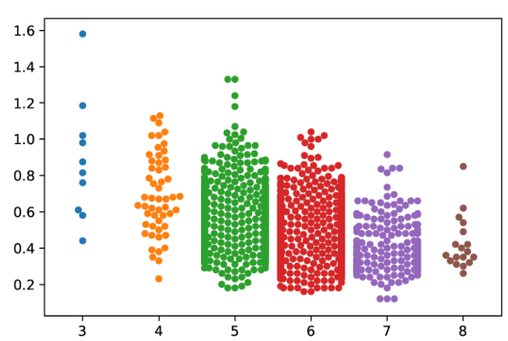
X

machine

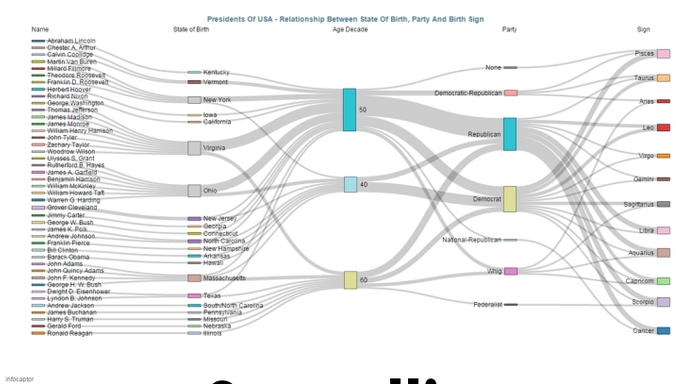
Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out **tasks** more **effectively**. – *Tamara Munzner at UBC*



Understanding



Exploratory



Storytelling

*Making **Human-in-the-loop** Data Analytics (Science) More **Effective***

History

Michael Friendly, "Milestones in the history of thematic cartography, statistical graphics, and data visualization"

Pre-17 Century
Early maps and diagrams

1600-1699
Measurement and theory

1700-1799
New graphic forms

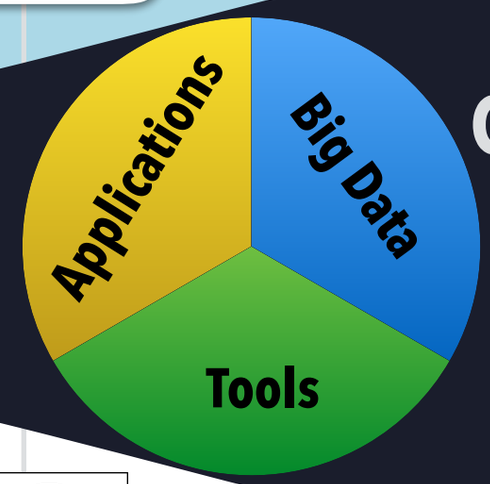
1800-1850
Beginning of modern graphics

1850-1900
The golden age of statistical graphics

1900-1950
The modern dark ages

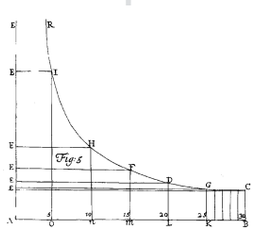
1950-1975
Re-birth of data visualization

1975- now
High-D, interactive, and dynamic

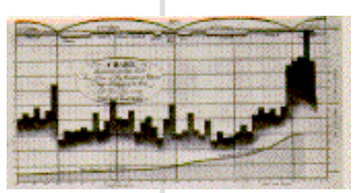


Cognitive Science
Computer Graphics
Visualization
Data Mining
Database

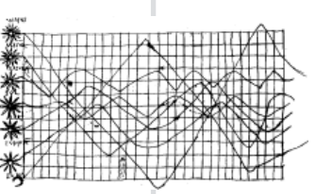
6200 BC
Konya town map



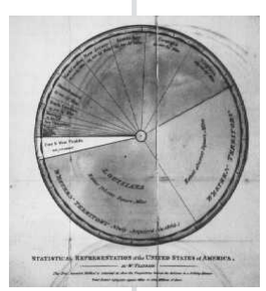
1684
Barometric pressure vs. altitude



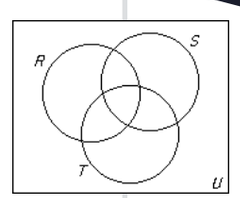
1786
Bar/line chart: Economic data, England



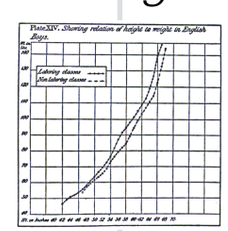
950
Changing values (positions of sun)



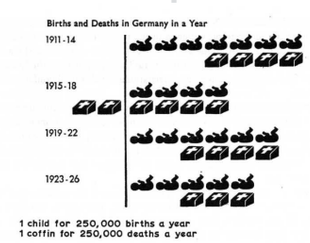
1801
Pie chart for part-whole relations



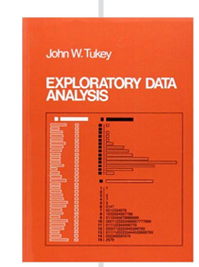
1880
Ven Diagram



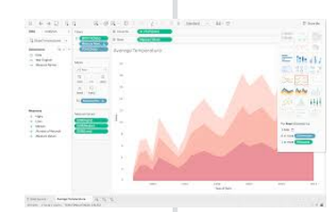
1880
Regression curve



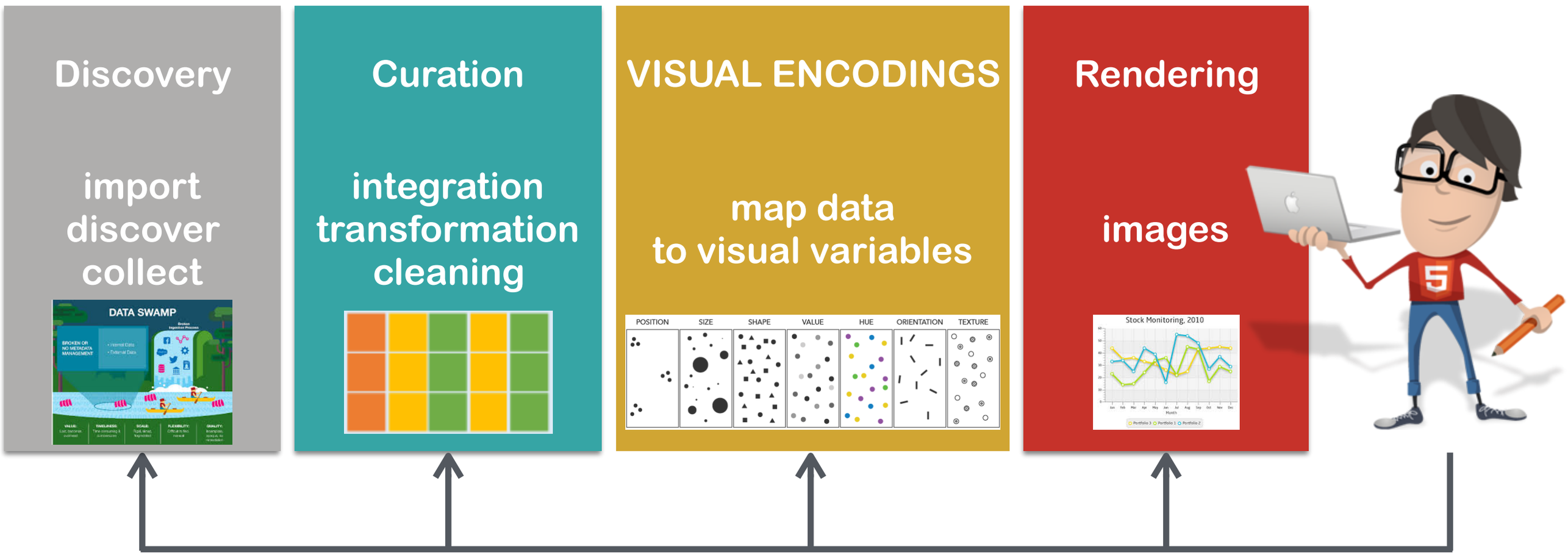
1924
births/ deaths in Germany



1970-1972

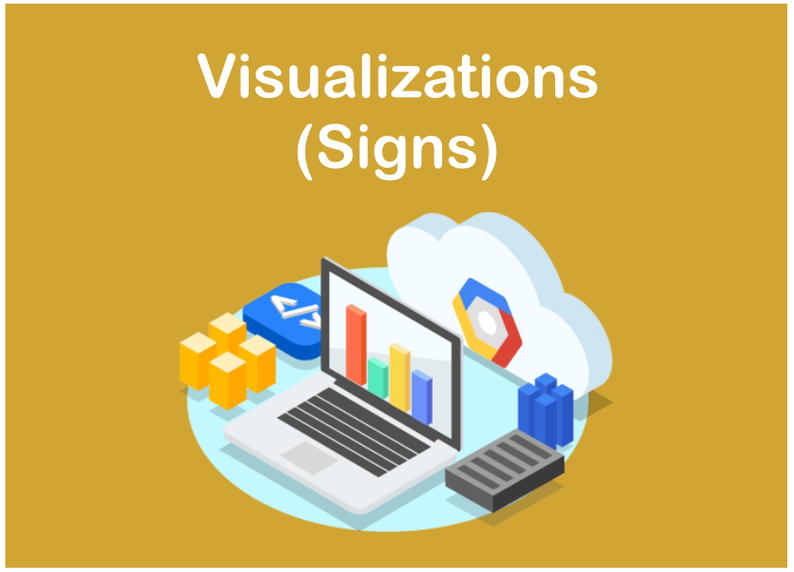


The Visualization Pipeline



Mapping Data to Visualizations

encode →



decode →

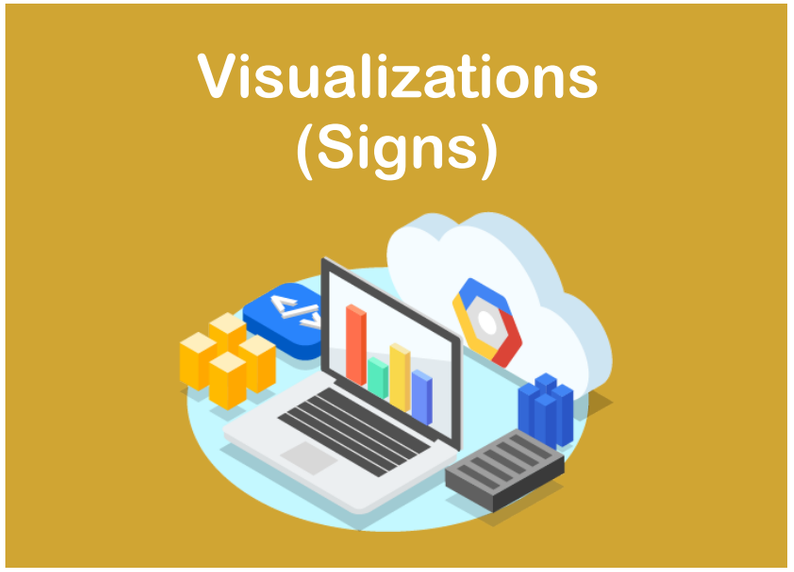
the "most effective" visualization

visual language
is a sign system

Jacques Bertin

Mapping Data to Visualizations

encode →



decode →

interact ↔

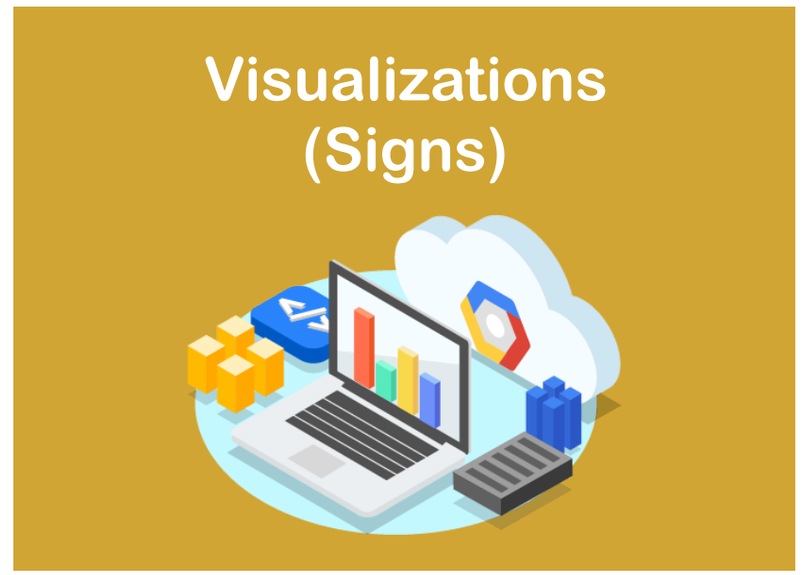
the "most effective" visualization

visual language
is a sign system

Jacques Bertin

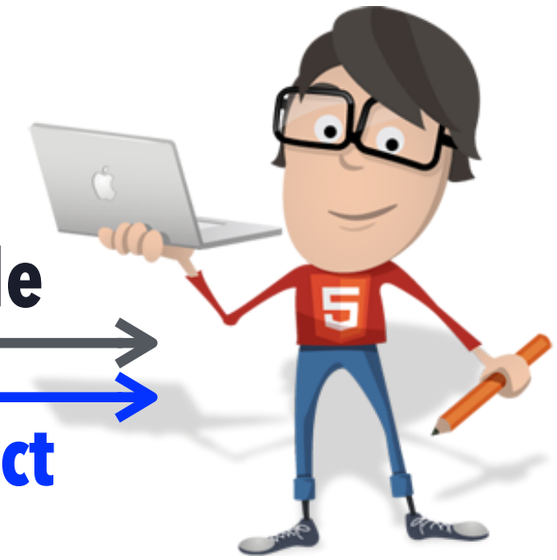
Mapping Data to Visualizations

encode



decode

interact



the "most effective" visualization

visual language
is a sign system

Jacques Bertin

Tools or Languages are needed

Characterizing Data and Visualizations

- Nominal

  - members of certain classes -

  - USA, Qatar, Netherlands
- Ordinal

  - related by order -

  - tiny, small, medium, large
  - Days: Mon, Tue, …, Sun
- Quantitative

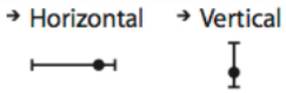
  - numerical values -

  - 2.3, 4.56, 0.8
  - Physical measurements: temperature

Characterizing Data and Visualizations

- Nominal
  - members of certain classes -
  - USA, Qatar, Netherlands
- Ordinal
  - related by order -
  - tiny, small, medium, large
  - Days: Mon, Tue, …, Sun
- Quantitative
  - numerical values -
  - 2.3, 4.56, 0.8
  - Physical measurements: temperature

Marks

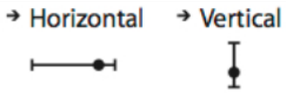
Points	
Lines	
Areas	
Position (x 2)	
Size	
Shape	
Value	
Colour	
Orientation	
Texture	

Visual Variables (Channels)

Characterizing Data and Visualizations

- Nominal
  - members of certain classes -
  - USA, Qatar, Netherlands
- Ordinal
  - related by order -
  - tiny, small, medium, large
  - Days: Mon, Tue, …, Sun
- Quantitative
  - numerical values -
  - 2.3, 4.56, 0.8
  - Physical measurements: temperature

map →

Points	
Lines	
Areas	
Position (x 2)	
Size	
Shape	
Value	
Colour	
Orientation	
Texture	

Marks

Visual Variables (Channels)

Characterizing Data and Visualizations

- Nominal
  - members of certain classes -
  - USA, Qatar, Netherlands
- Ordinal
  - related by order -
  - tiny, small, medium, large
  - Days: Mon, Tue, …, Sun
- Quantitative
  - numerical values -
  - 2.3, 4.56, 0.8
  - Physical measurements: temperature

map →

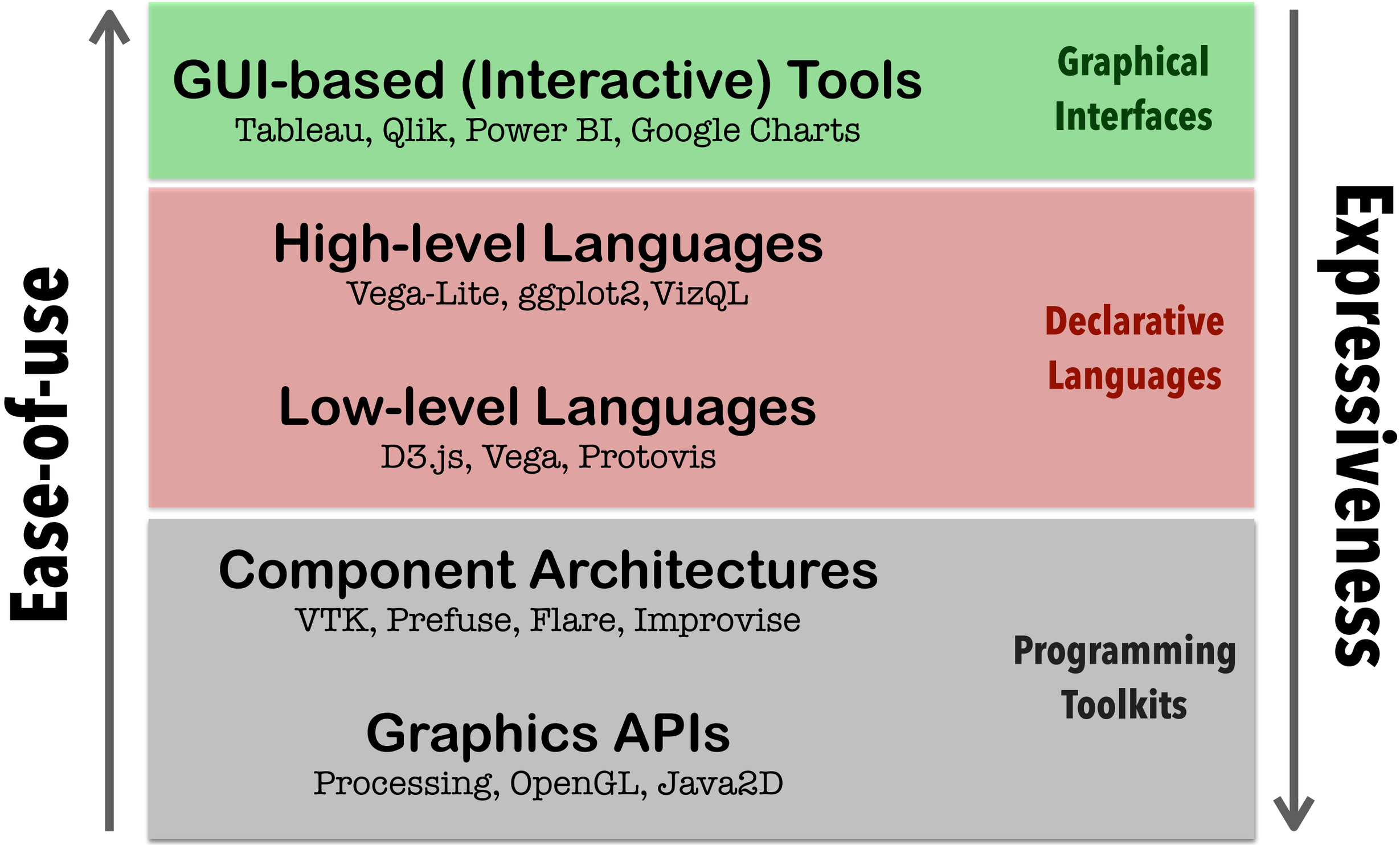
Points	
Lines	
Areas	
Position (x 2)	Horizontal Vertical
Size	
Shape	
Value	
Colour	
Orientation	
Texture	

Marks

Visual Variables (Channels)

https://jenniewblog.wordpress.com/2016/03/08/marks-and-channels-chapter5/

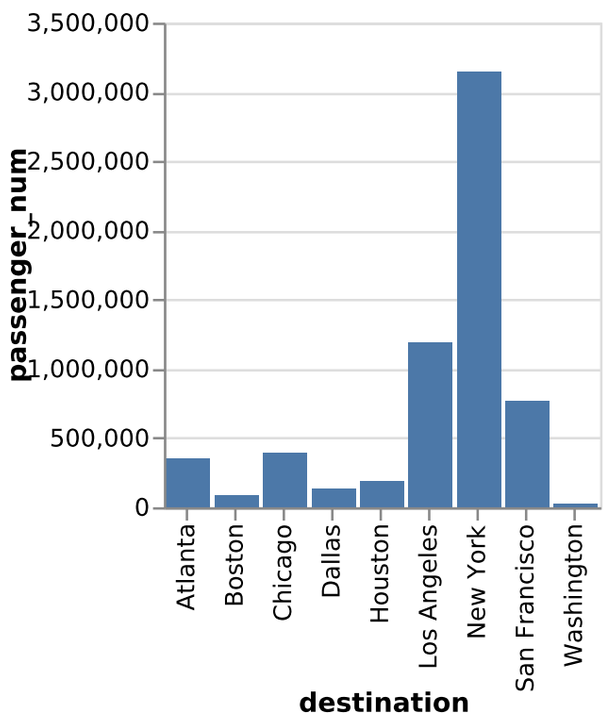
A Visualization Tool Stack



Vega-Lite and Vega

Vega-Lite

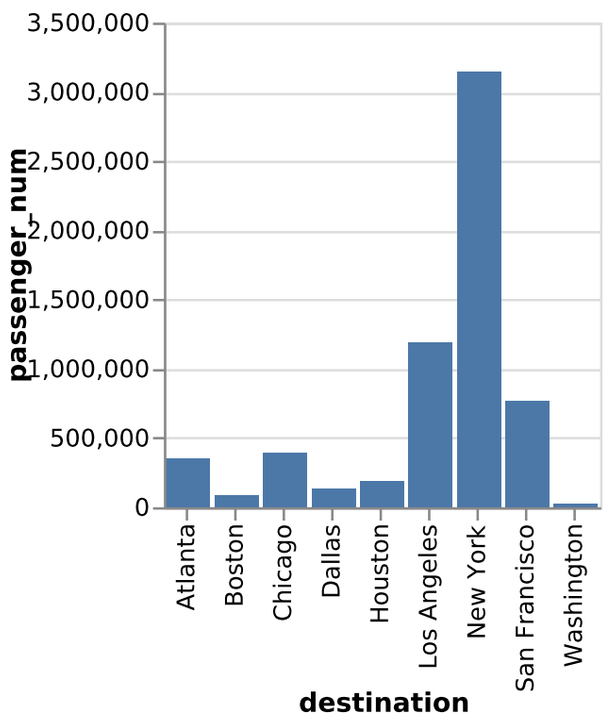
```
{
  "data": [
    {"name": "table", "url": "/data/flight_statistics.json"}
  ],

  "mark": "bar",
  "encoding": {
    "x": {"field": "destination", "type": "ordinal"},
    "y": {"field": "passenger_num", "type": "quantitative"}
  }
}
```



Vega-Lite and Vega

Vega-Lite

```
{
  "data": [
    {"name": "table", "url": "/data/flight_statistics.json"}
  ],

  "mark": "bar",
  "encoding": {
    "x": {"field": "destination", "type": "ordinal"},
    "y": {"field": "passenger_num", "type": "quantitative"}
  }
}
```



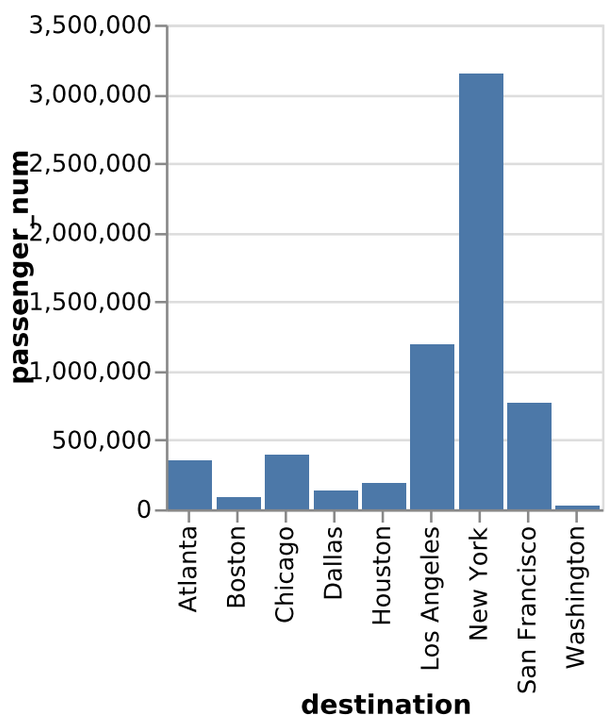
Vega

```
{
  "width": 600,
  "height": 200,
  "padding": 5,

  "data": [
    {"name": "table", "url": "/data/flight_statistics.json"}
  ],

  "scales": [
    {
      "name": "xscale",
      "type": "band",
      "domain": {"data": "table", "field": "destination"},
      "range": "width",
      "padding": 0.05,
      "round": true
    },
    {
      "name": "yscale",
      "domain": {"data": "table", "field": "passenger_num"},
      "nice": true,
      "range": "height"
    }
  ],

  "axes": [
    { "orient": "bottom", "scale": "xscale" },
    { "orient": "left", "scale": "yscale" }
  ],

  "marks": [
    {
      "type": "rect",
      "from": {"data":"table"},
      "encode": {
        "enter": {
          "x": {"scale": "xscale", "field": "destination"},
          "width": {"scale": "xscale", "band": 1},
          "y": {"scale": "yscale", "field":
"passenger_num"},
          "y2": {"scale": "yscale", "value": 0}
        }
      }
    },
    {
      "type": "text",
      "encode": {
        "enter": {
          "align": {"value": "center"},
          "baseline": {"value": "bottom"},
          "fill": {"value": "#333"}
        }
      }
    }
  ]
}
```

Vega-Lite and Vega

Vega-Lite

```
{
  "data": [
    {"name": "table", "url": "/data/flight_statistics.json"}
  ],

  "mark": "bar",
  "encoding": {
    "x": {"field": "destination", "type": "ordinal"},
    "y": {"field": "passenger_num", "type": "quantitative"}
  }
}
```



Vega

```
{
  "width": 600,
  "height": 200,
  "padding": 5,
```

Data + Transforms

```
"data": [
  {"name": "table", "url": "/data/flight_statistics.json"}
],
```

Scales

```
"scales": [
  {
    "name": "xscale",
    "type": "band",
    "domain": {"data": "table", "field": "destination"},
    "range": "width",
    "padding": 0.05,
    "round": true
  },
  {
    "name": "yscale",
    "domain": {"data": "table", "field": "passenger_num"},
    "nice": true,
    "range": "height"
  }
],
```

Guides

```
"axes": [
  { "orient": "bottom", "scale": "xscale" },
  { "orient": "left", "scale": "yscale" }
],
```

Marks

```
  "marks": [
    {
      "type": "rect",
      "from": {"data":"table"},
      "encode": {
        "enter": {
          "x": {"scale": "xscale", "field": "destination"},
          "width": {"scale": "xscale", "band": 1},
          "y": {"scale": "yscale", "field":
"passenger_num"},
          "y2": {"scale": "yscale", "value": 0}
        }
      }
    },
    {
      "type": "text",
      "encode": {
        "enter": {
          "align": {"value": "center"},
          "baseline": {"value": "bottom"},
          "fill": {"value": "#333"}
        }
      }
    }
  ]
}
```

GUI-based (Interactive) Interface

Mutual Intelligibility and Shared Understanding

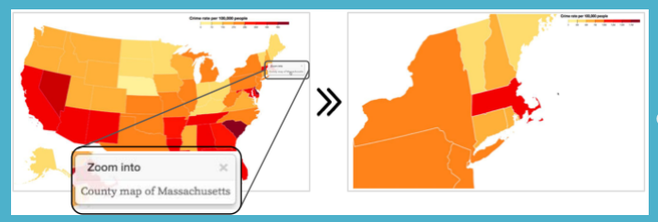
{ People }

Data/View Specification
visualize
filter
sort
derive

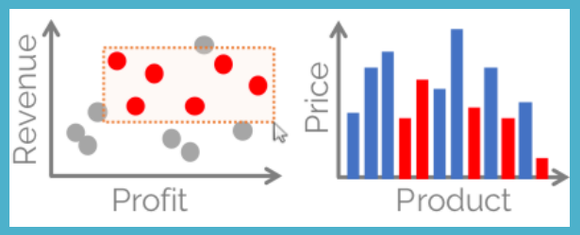
View Manipulation
select
navigate
coordinate
organize

Process and Provenance
record
annotate
share
guide

"Kyrix: Interactive Visual Data Exploration at Scale. CIDR 2019.

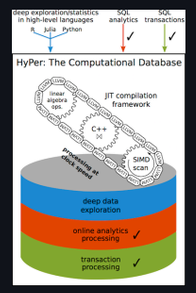


Ermac: Combining design and performance in a data visualization management system. CIDR 2017.



{ Machines }

declarative language
data + transforms
mapping



HyPer

Optimizer

Executor

PostgreSQL

SQLite

GUI-based (Interactive) Interface

Mutual Intelligibility and Shared Understanding

{ People }

Data/View Specification
visualize
filter
sort
derive

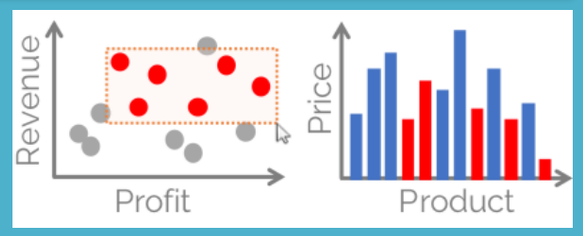
View Manipulation
select
navigate
coordinate
organize

Process and Provenance
record
annotate
share
guide

"Kyrix: Interactive Visual Data Exploration at Scale.
CIDR 2019.

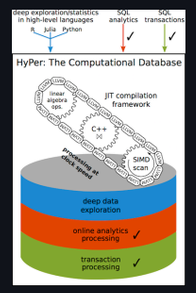
Civilizer 2.0, VLDB 2019 demo

Ermac: Combining design and performance in a data visualization management system.
CIDR 2017.



{ Machines }

declarative language
data + transforms
mapping



HyPer

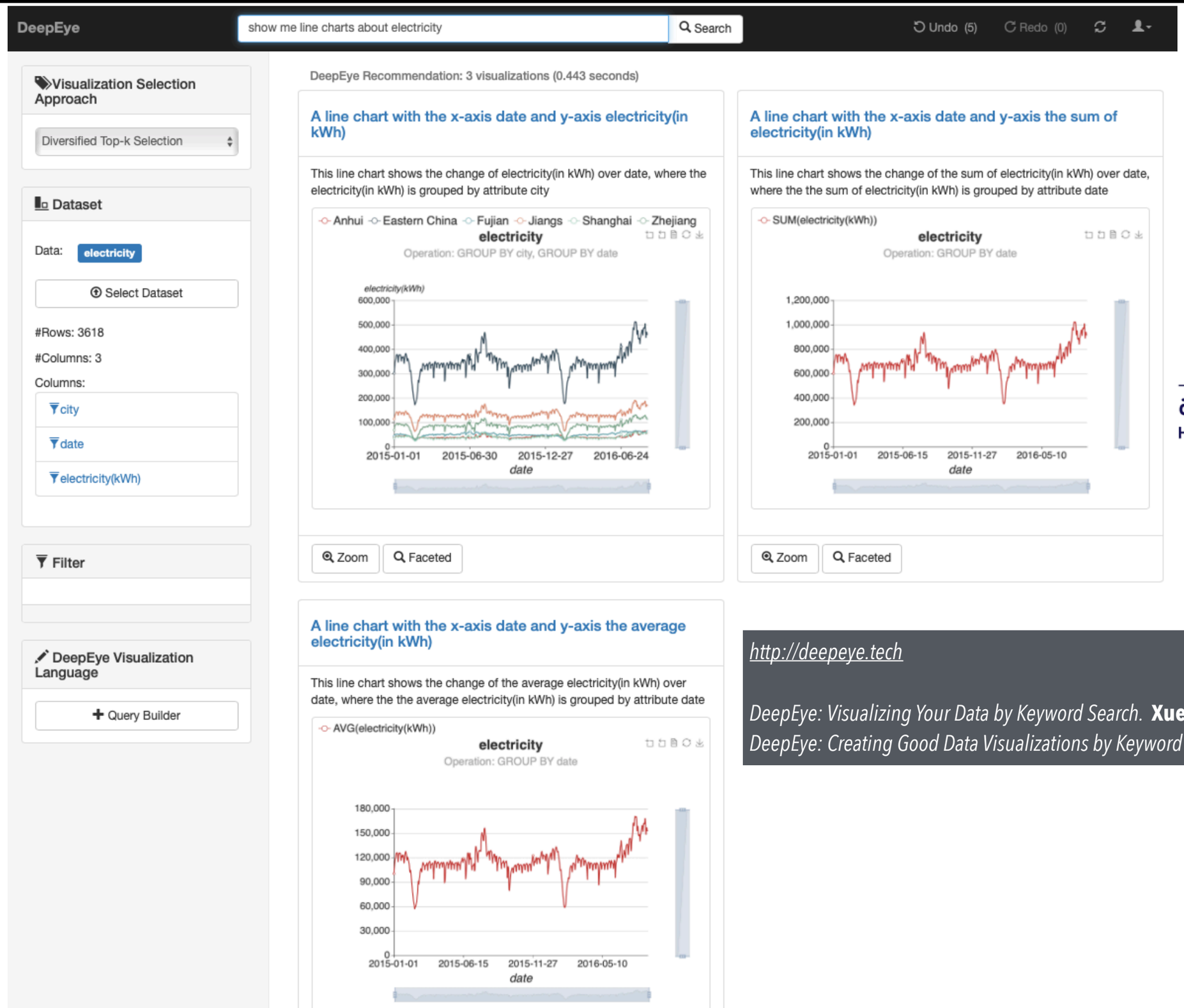
Optimizer

Executor

PostgreSQL

SQLite

Keyword (Under-specified)

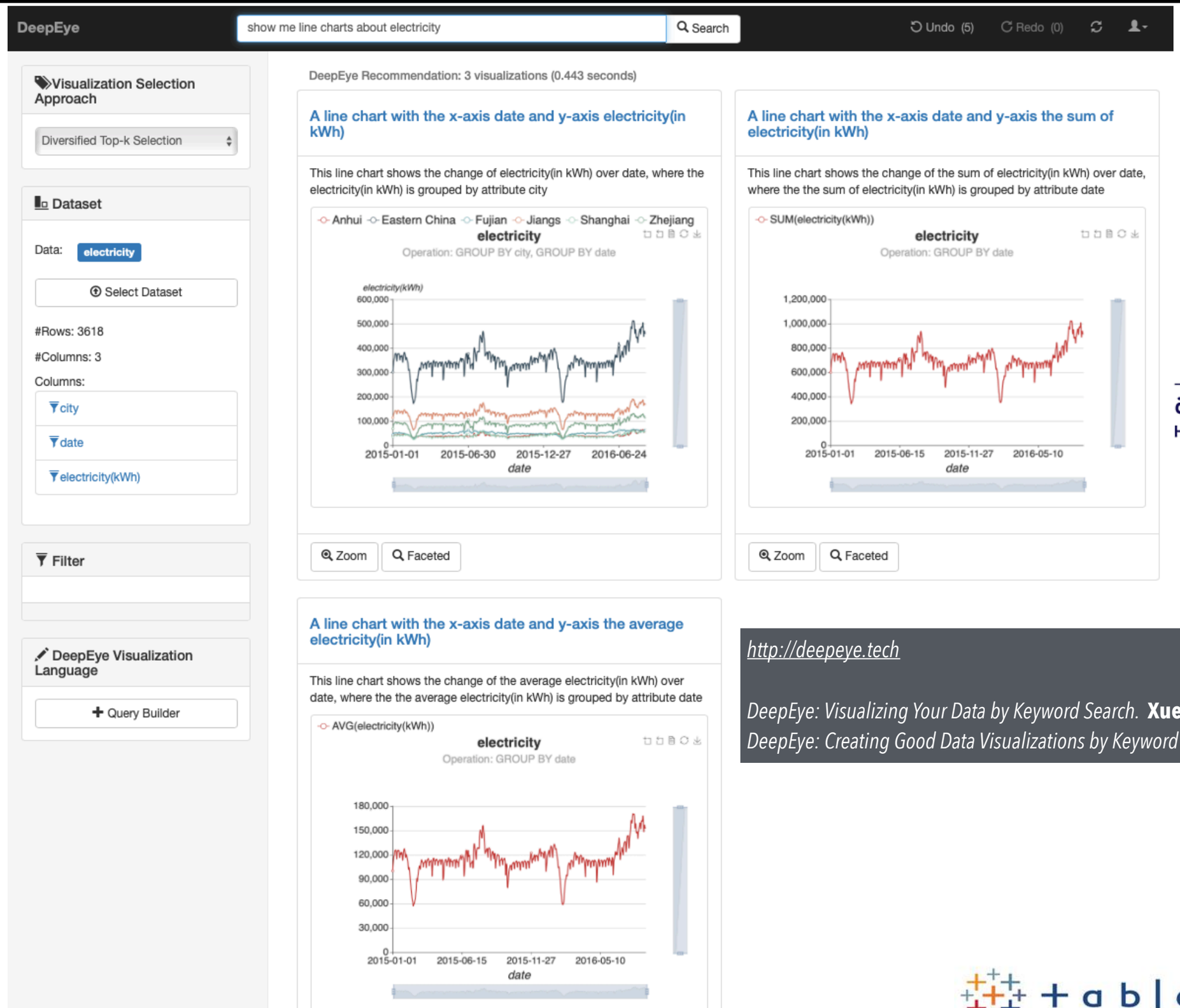


QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

http://deepeye.tech

DeepEye: Visualizing Your Data by Keyword Search. **Xue di et al.**, EDBT (vision) 2018.
DeepEye: Creating Good Data Visualizations by Keyword Search. **Yuyu Luo et al.**, *SIGMOD Demo 2018.*

Keyword (Under-specified)



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

http://deepeye.tech

DeepEye: Visualizing Your Data by Keyword Search. **Xue di et al.**, EDBT (vision) 2018.
DeepEye: Creating Good Data Visualizations by Keyword Search. **Yuyu Luo et al.**, *SIGMOD Demo 2018.*

tableau **Ask Data**

Further Readings

- Tamara Munzner, "***Visualization Analysis & Design***", Tutorial on VIS 2017
- Tamara Munzner, "***Data Visualization Pitfalls to Avoid***", Tutorial
- Jeffrey Heer, "***Data Visualization***", University of Washington, Lecture CSE 442
- Jacques Bertin, "***Semiology of Graphics: Diagrams, Networks, Maps***". 1967
- Leland Wilkinson, "***The Grammar of Graphics***", 1999
- Scott Murray, "***Interactive Data Visualization for the Web***", 2013
- Jeff Johnson, "***Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules***", Morgan Kaufmann, 2010
- Stanley Smith Stevens, "***Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects***", Wiley, 1975
- Colin Ware, "***Visual Thinking for Design***", Morgan Kaufmann, 2008

- Enrico Bertini and Moritz Stefaner, "***Data stories***", podcast
- Amy Cesal, Mollie Pettit and Elijah Meek, "***Data Visualization Society***", a slack channel