github repo

Doing Data Visualisation in Python with Altair, A Primer

Cagatay Turkay

Associate Professor,
Centre for Interdisciplinary Methodologies
University of Warwick

DSSGxUK 2022, University of Warwick, 29 June 2022

Associate Professor @ CIM at University of Warwick

Senior Lecturer in at giCentre, City, University of London (2014 - 2019)

PhD in Visualisation, University of Bergen, NO

MSc & BSc in Computer Science (Sabanci University, TR & METU, TR)

Research on ...

Methods: Interactive visual analytics techniques

Integrating **Computational Tools** and **Visualisation**

New **interaction** forms (e.g., natural lang. + vis)

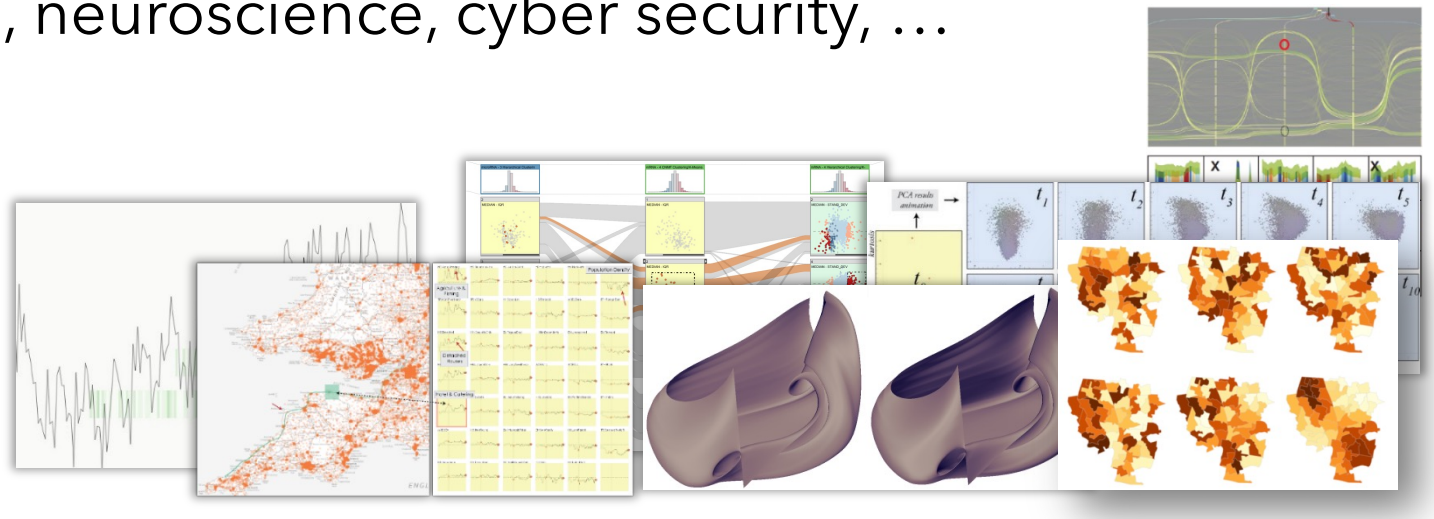
Theoretical Foundations: Perceptual/Cognitive mechanisms of Visualisation

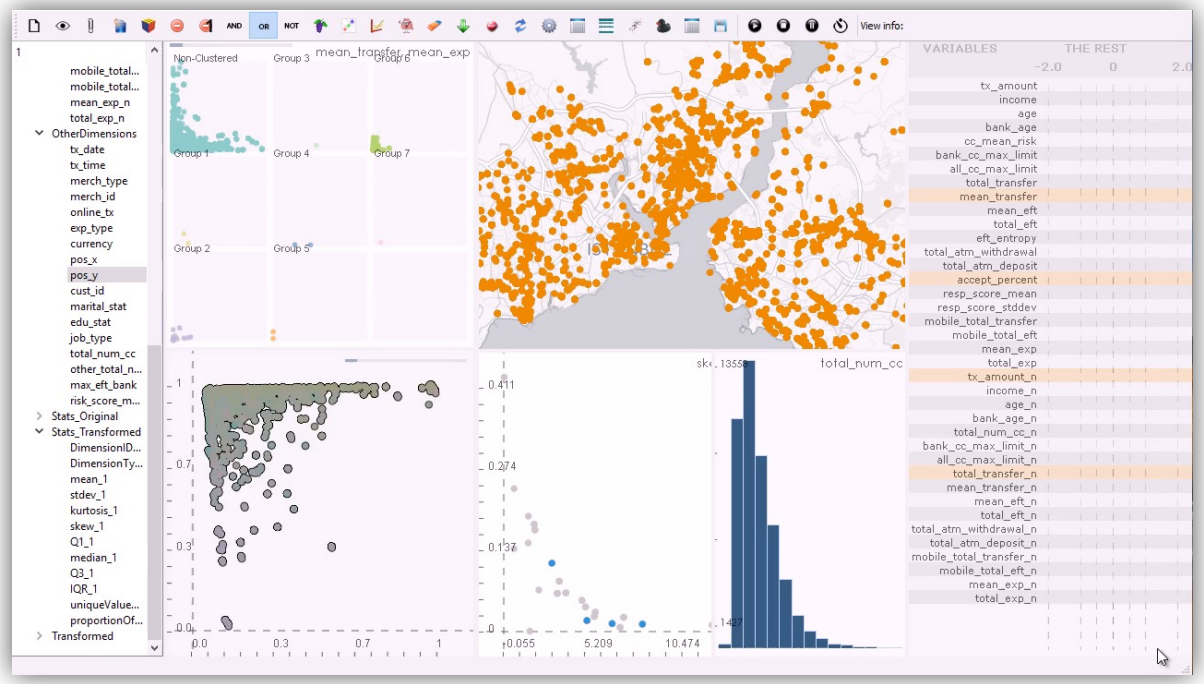
Applications in ...

Biology, transport, intelligence, neuroscience, cyber security, …

Teaching on ...

(Visual) Data Science



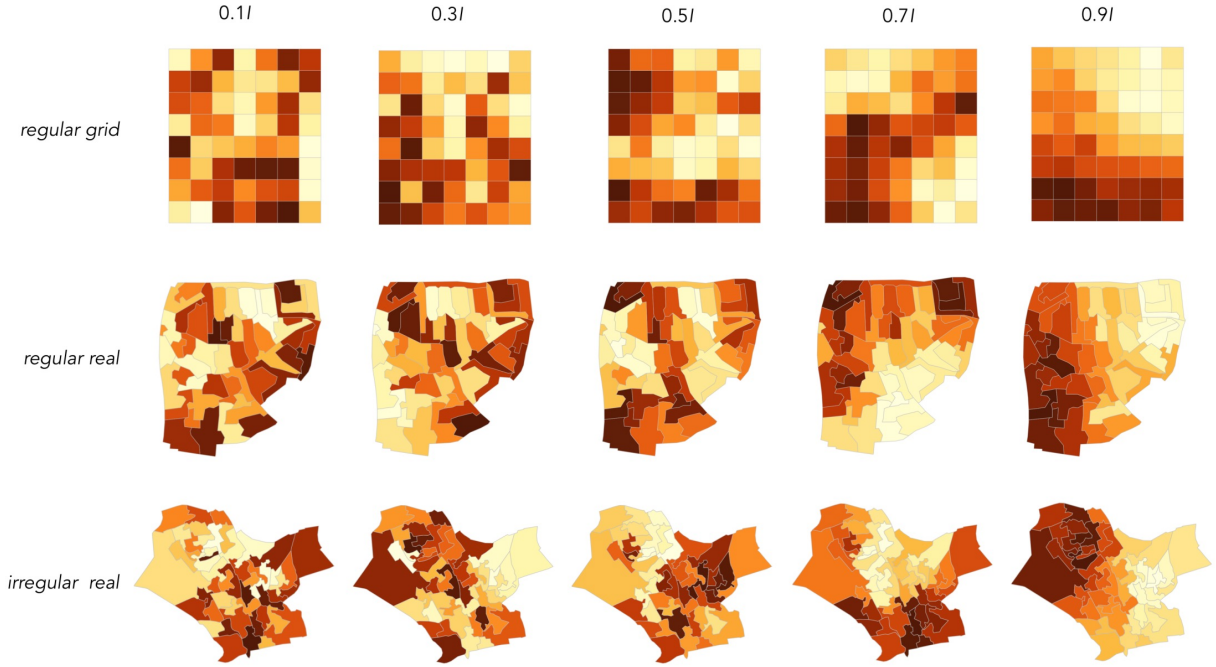


Designing Progressive and Interactive Analytics Processes for High-Dimensional Data Analysis

Turkay, C., Kaya, E., Balcisoy, S., Hauser, H., IEEE TVCG 2017

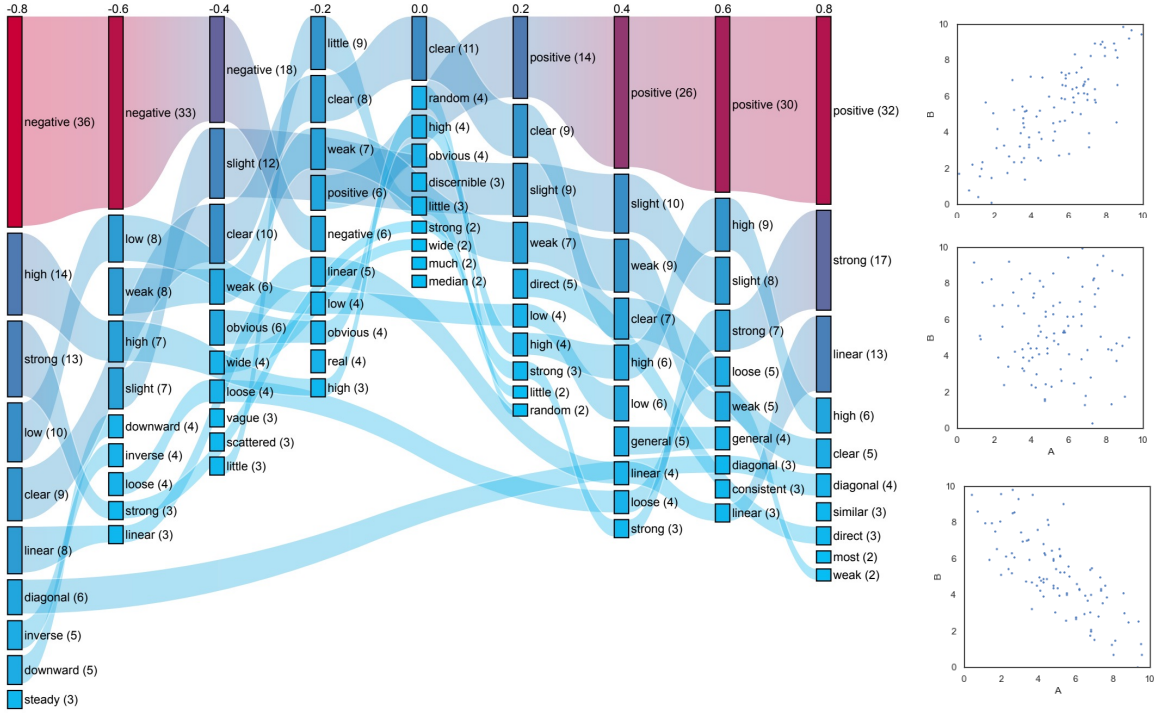
Enhancing a Social Science Model-building Workflow with Interactive Visualisation

Turkay, C., Slingsby, A., Lahtinen, K., Butt, S., & Dykes, J., ESANN 2016 (& Neurocomputing 2017)



Map LineUps: effects of spatial structure on graphical inference

Beecham, R., Dykes, J., Meulemans, W., Slingsby, A., Turkay, C. & Wood, J. (2016).
IEEE Transactions on Visualization and Computer Graphics.



Words of Estimative Correlation: Studying Verbalizations of Scatterplots

Henkin, R., Turkay, C.(2021).
IEEE Transactions on Visualization and Computer Graphics.

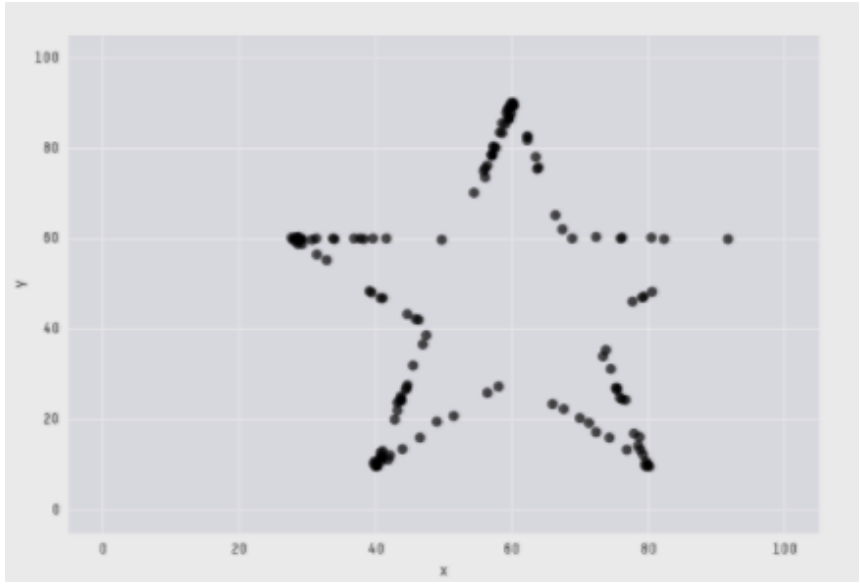
Today?

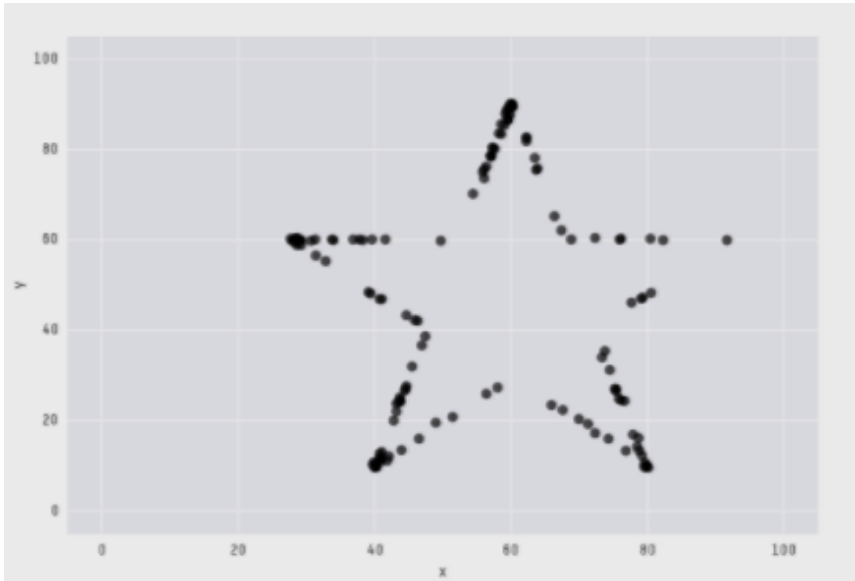
- A brief intro to visualization
  - Elements of a visualization
  - Channels and marks
  - Grammar of visualization (which motivated Altair et al)

- Altair primer
  - Hands on with Altair
  - Basic plotting with examples for further exploration

WHY VISUALISE?

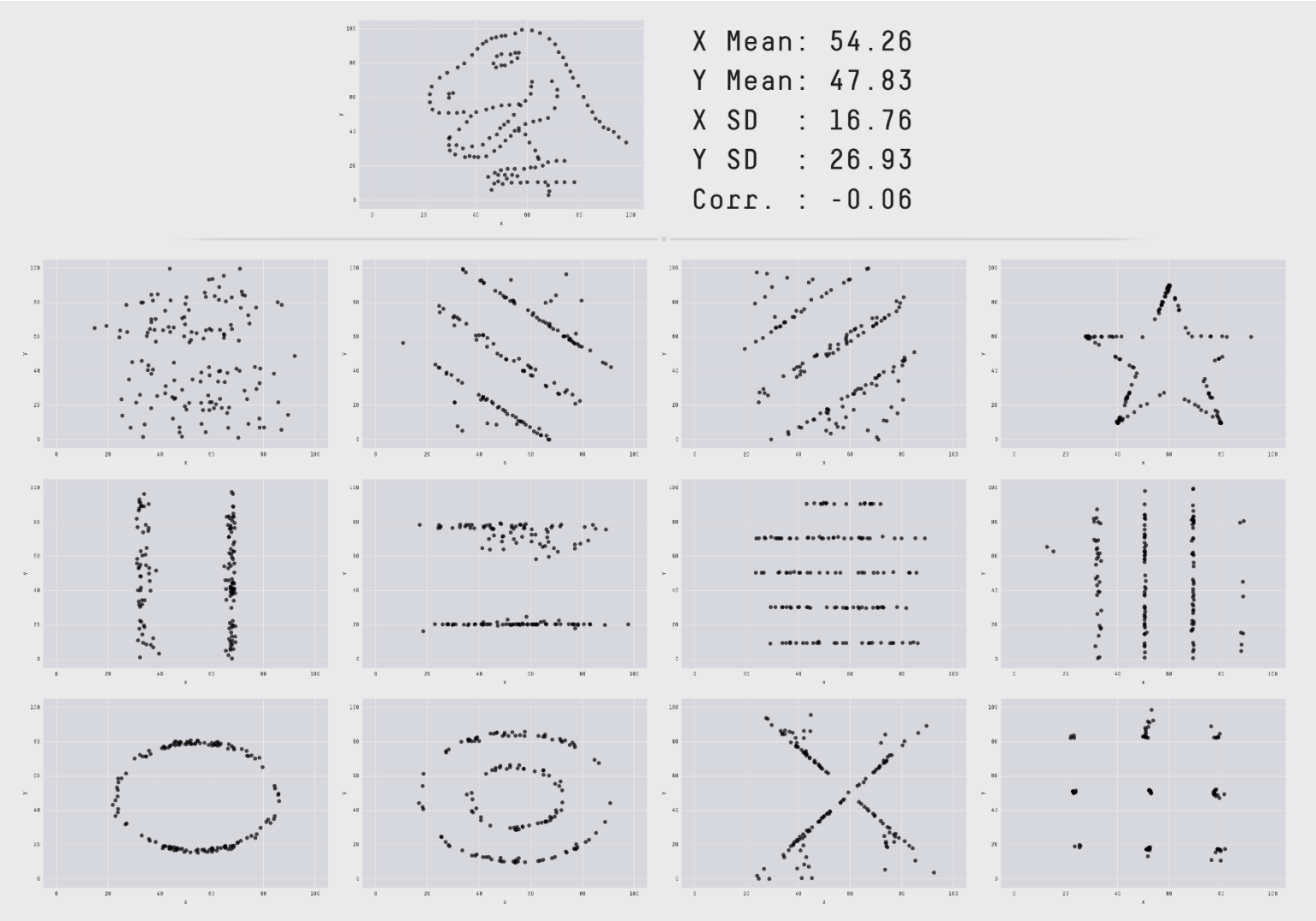
Do you see any similarities?






```
X Mean: 54.26
Y Mean: 47.83
X SD  : 16.76
Y SD  : 26.93
Corr. : -0.06
```

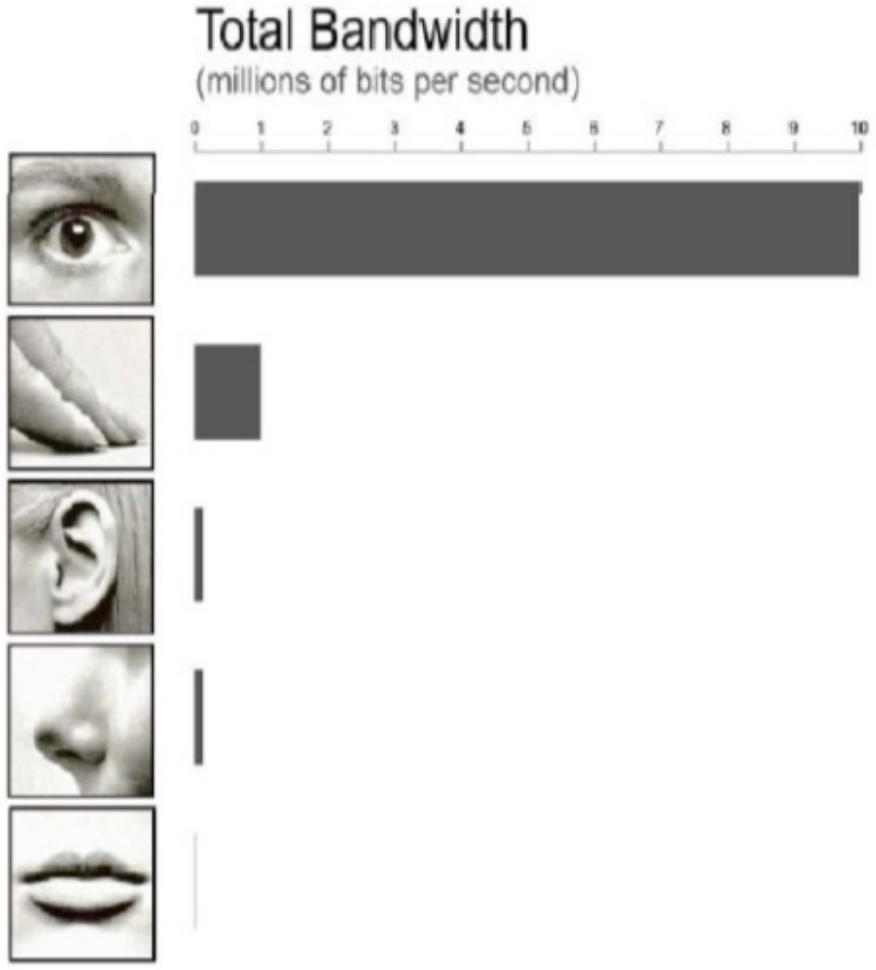
The ***Datasaurus Dozen***



https://www.autodeskresearch.com/publications/samestats

Why Visual?

- Figures are **richer**; provide more information with less clutter and in less space.
- Figures provide the *gestalt* effect: they give an overview; **make structure more visible**.
- Figures are more **accessible**, easier to understand, **faster to grasp**, more **comprehensible**, more **memorable**, more **fun**, and **less formal**.



List & slide adapted from: [Stasko et al. 1998 & Alexander Lex]

ENCODING DECODING ACTION

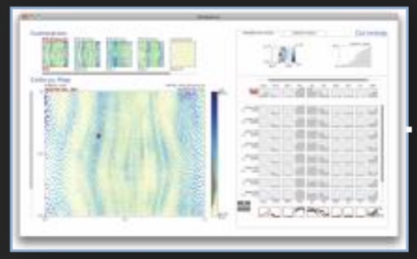
X1, Y1,
x2, y2 …

Engage

"Wow, X & Y looks amazing!"

"I need to find out more!"

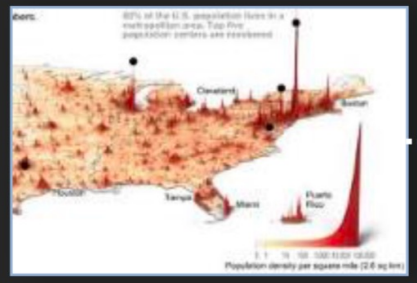
Explore



"I wonder how x relates to y"

"Maybe z is important?"

Explain



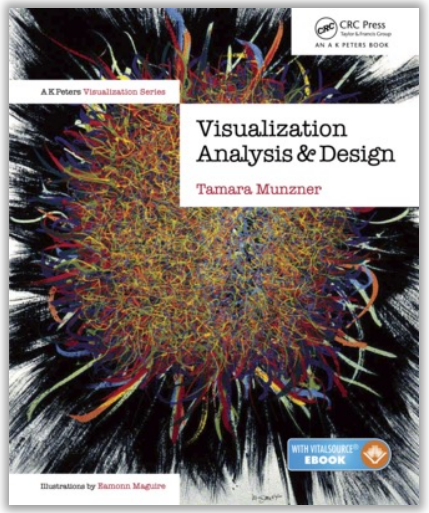
"X does y"

"if I do x then…"

Slide adapted from Greg McInerny

Visualisation?

"Computer-based visualization systems provide visual representations of ***datasets*** *designed to help* ***people*** *carry out* ***tasks*** *more effectively."* *[Tamara Munzner, 2014]*



BUILDING BLOCKS OF VISUALISATIONS

VISUAL ELEMENTS

Marks & Channels

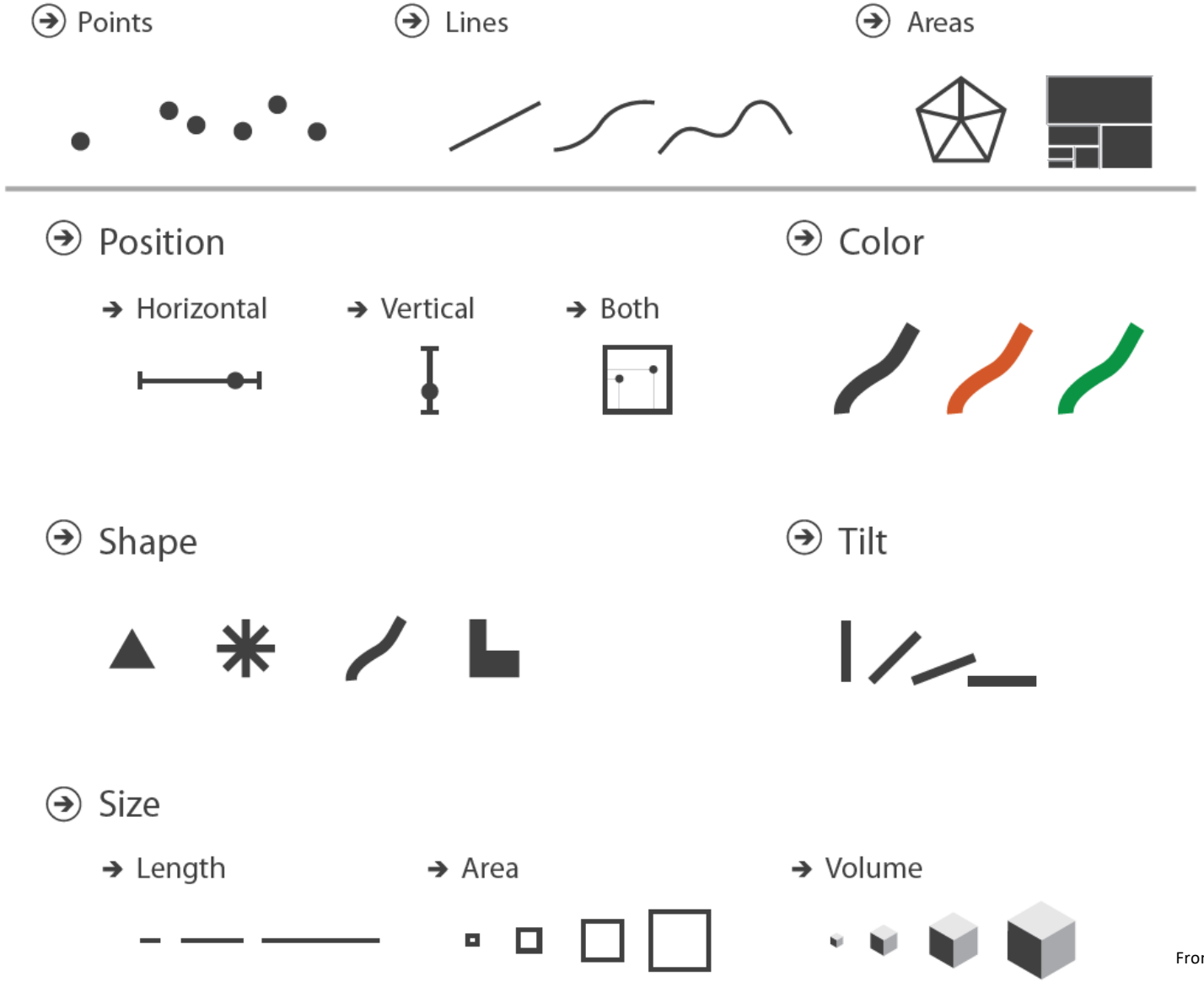
Visual Marks: represent **items** (or **links**)

Channels: change **appearance** based on **attribute**

Channel = Visual Variable

Basic visual elements

VISUAL MARKS
i.e., primitives

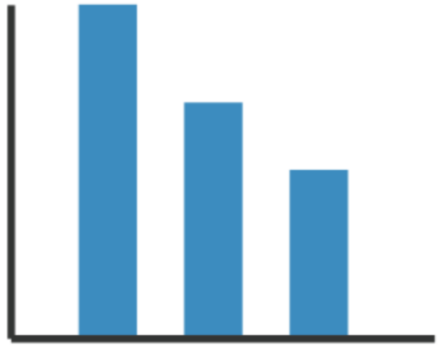


VISUAL CHANNELS
i.e., how they look

From Visualization Analysis & Design by T. Munzner, 2014

Visual encoding

Def. *Representing information visually as combination of marks and channels*



1:
vertical position

mark: line

2:
vertical position
horizontal position

mark: point

3:
vertical position
horizontal position
color hue

mark: point

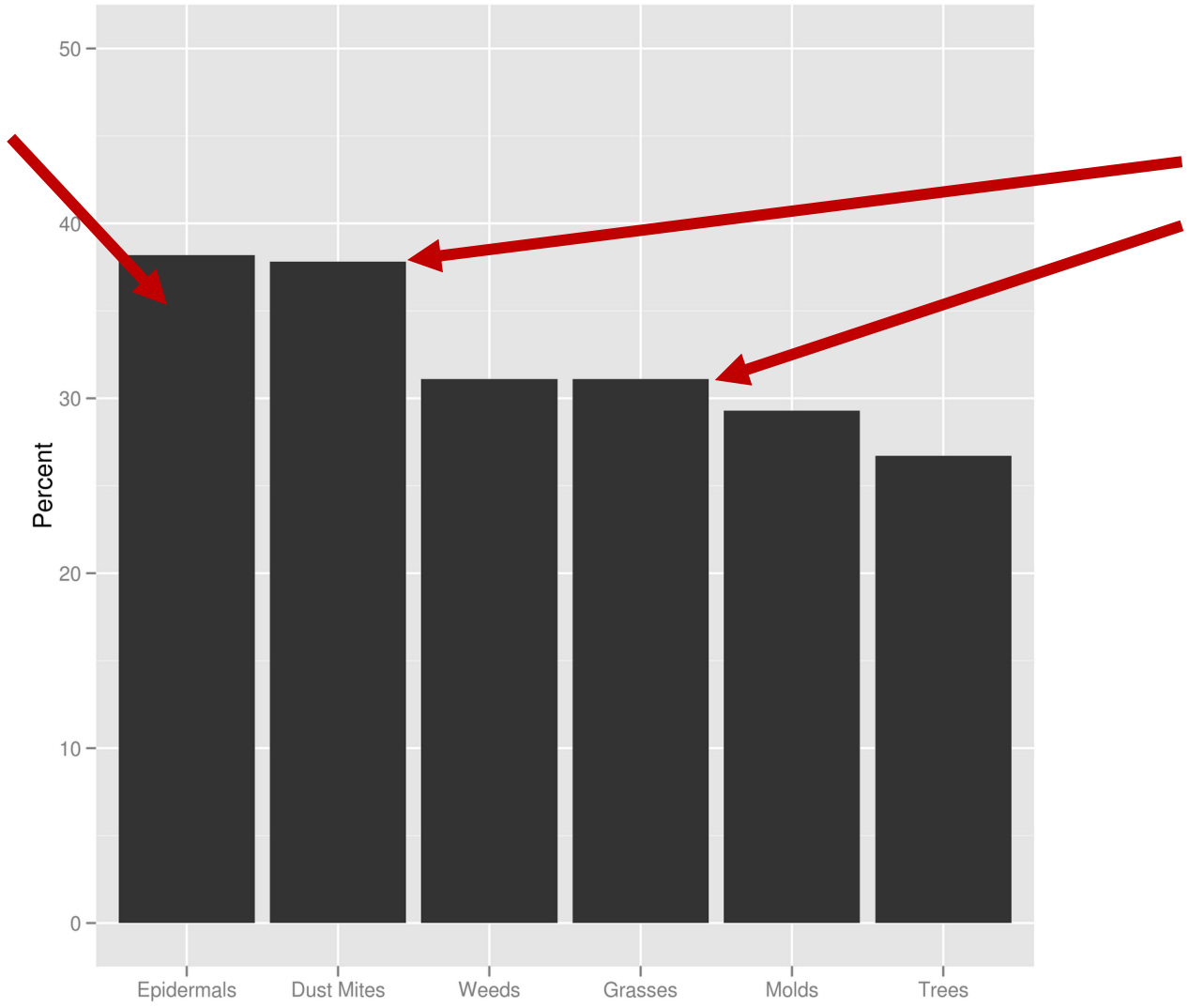
4:
vertical position
horizontal position
color hue
size (area)

mark: point

Basic visual elements in action

VISUAL MARK

lines (**thick** ones)



VISUAL CHANNEL

length (of bars)

http://rrubyperlundich.blogspot.co.uk/2012_09_01_archive.html

Basic visual elements in action

VISUAL CHANNELS

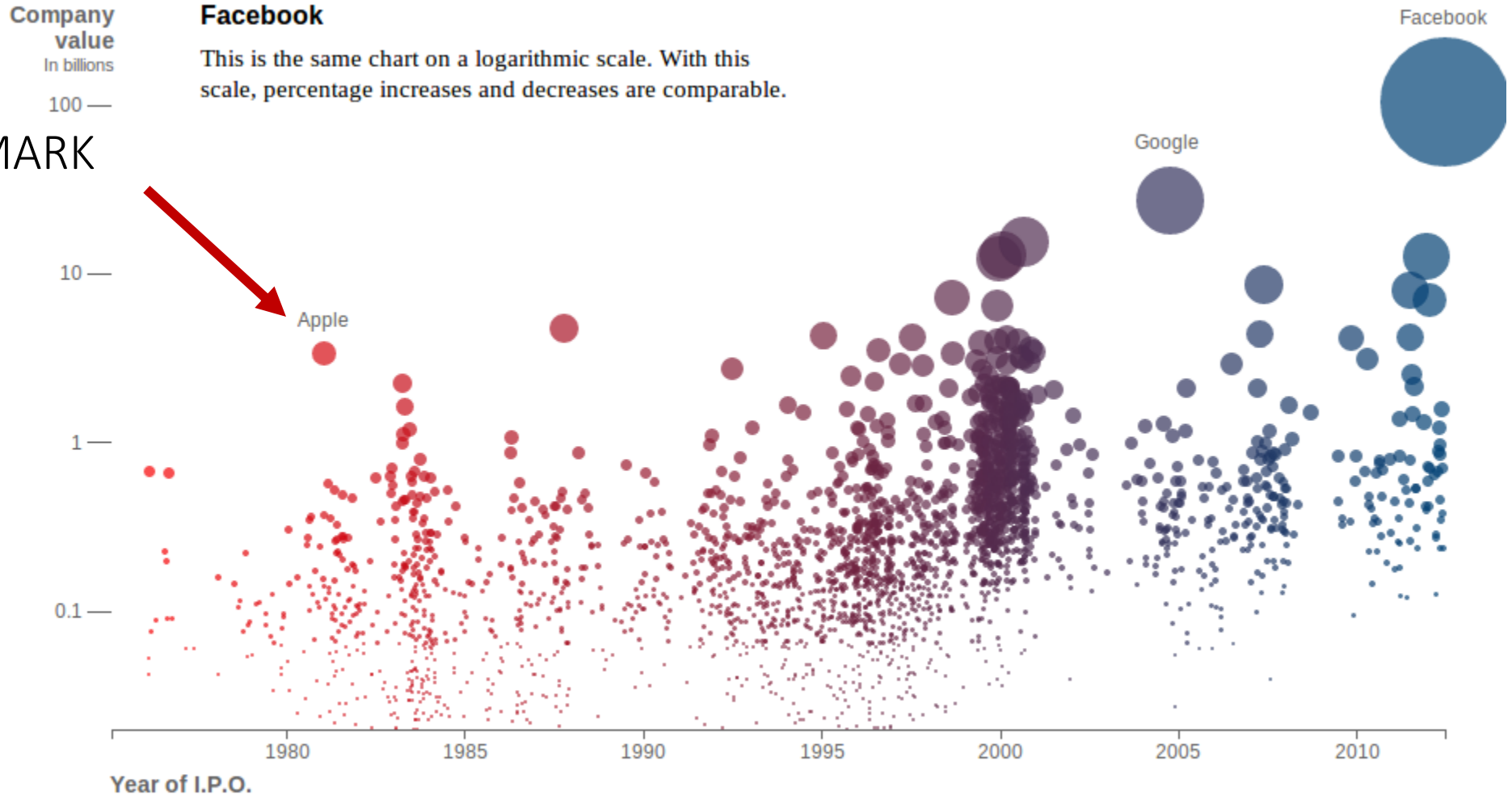
position x and y

colour

Area (redundant)

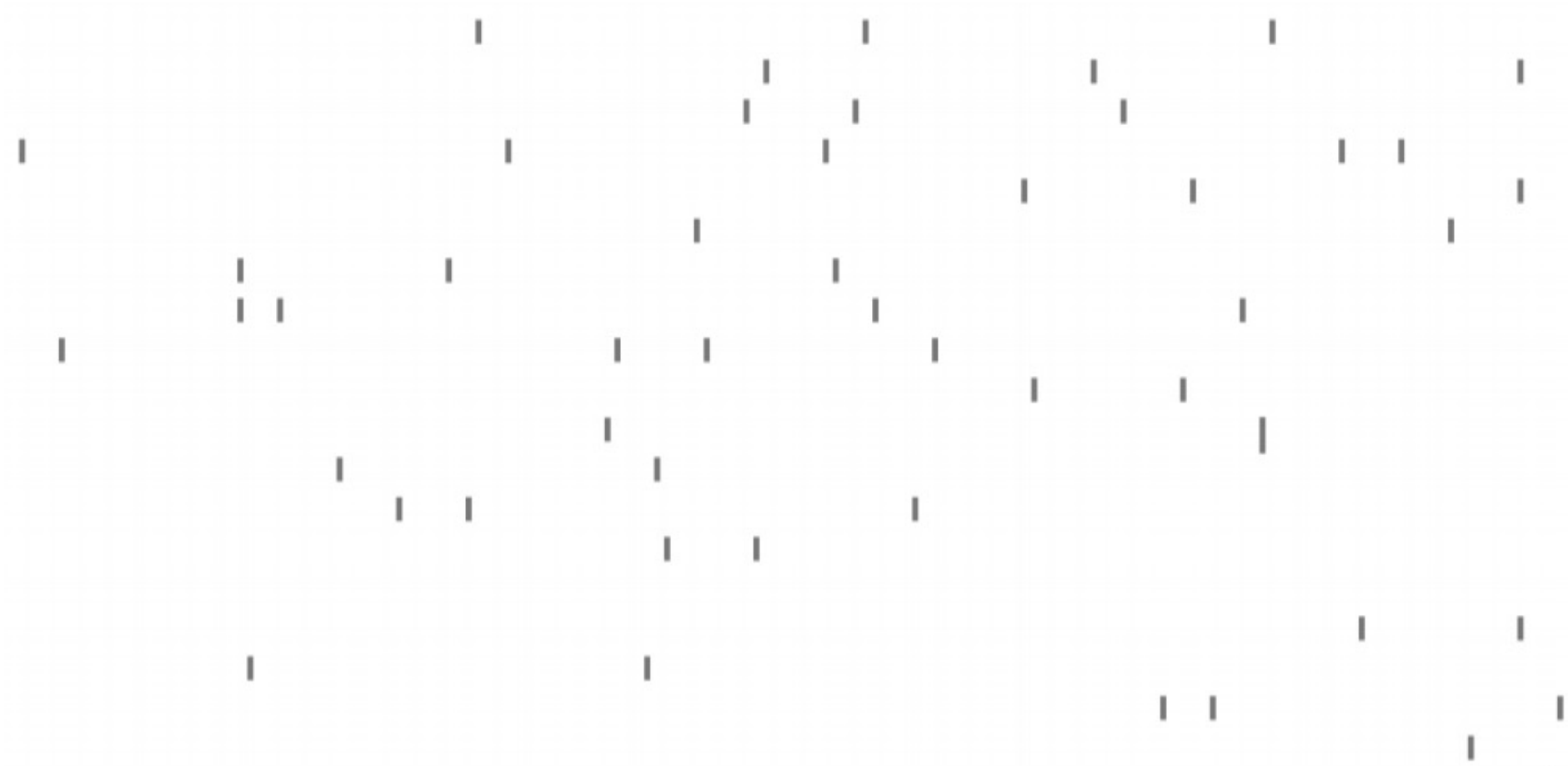
VISUAL MARK

points



http://www.nytimes.com/interactive/2012/05/17/business/dealbook/how-the-facebook-offering-compares.html

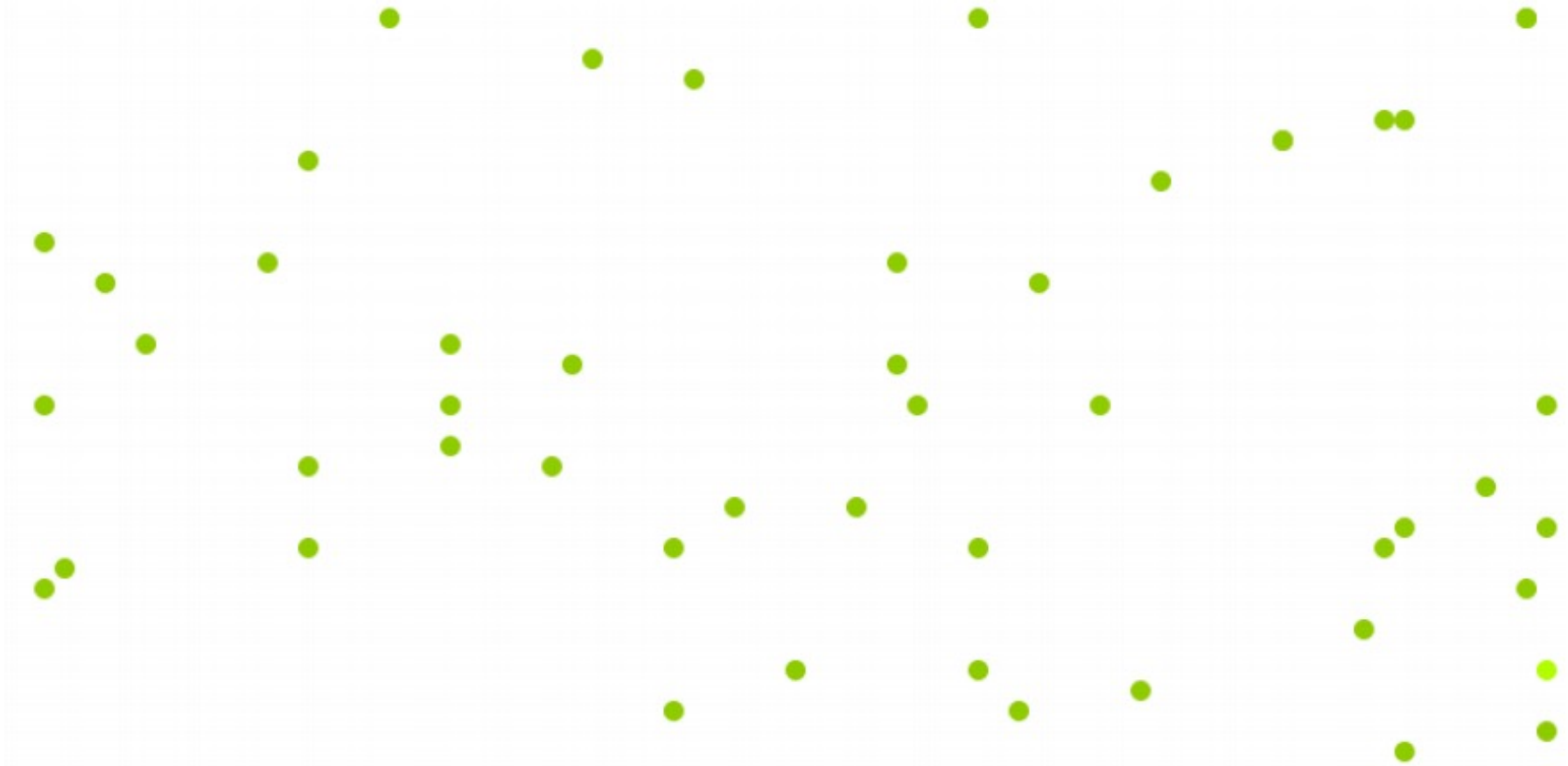
FIND THE ODD ONE OUT!



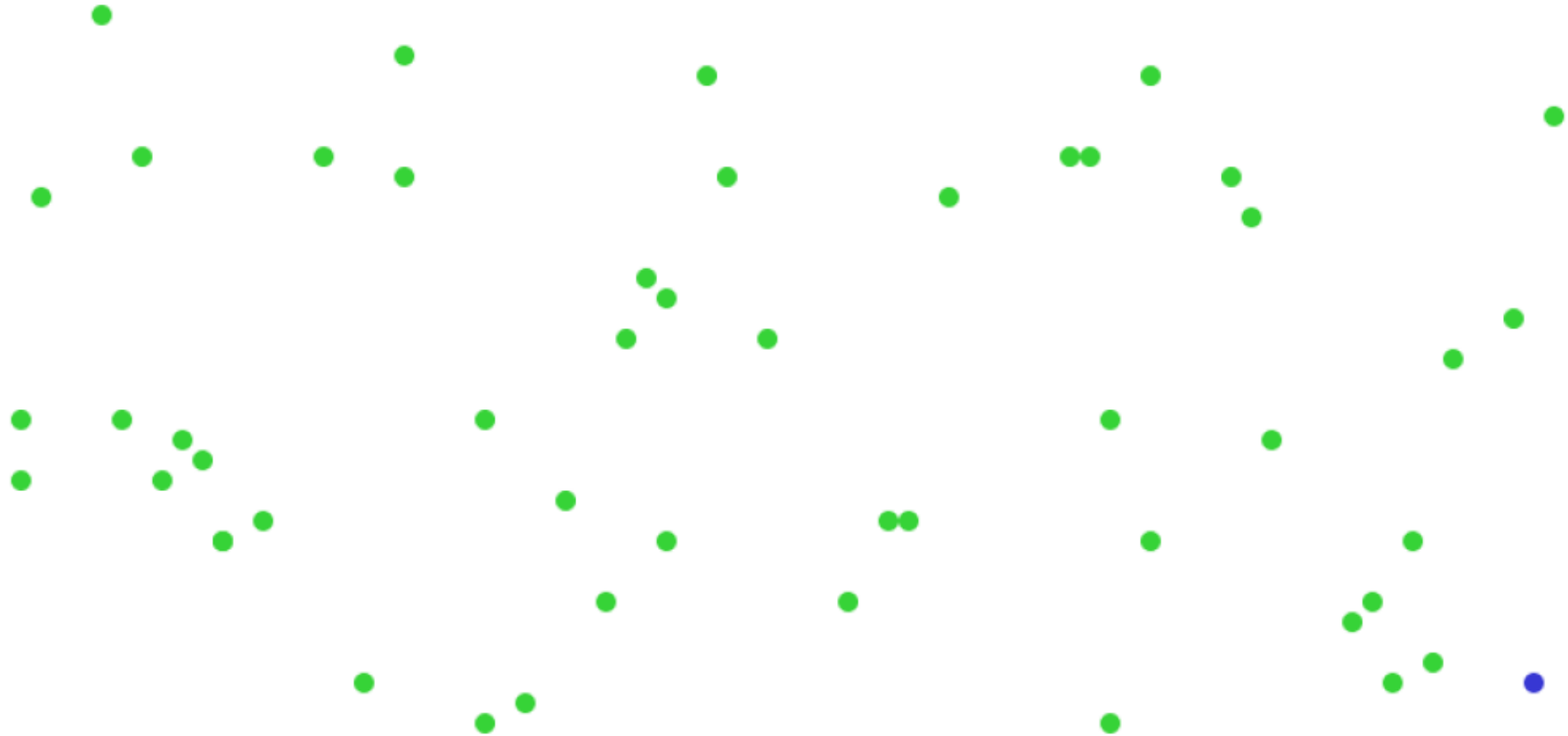
Length

Size

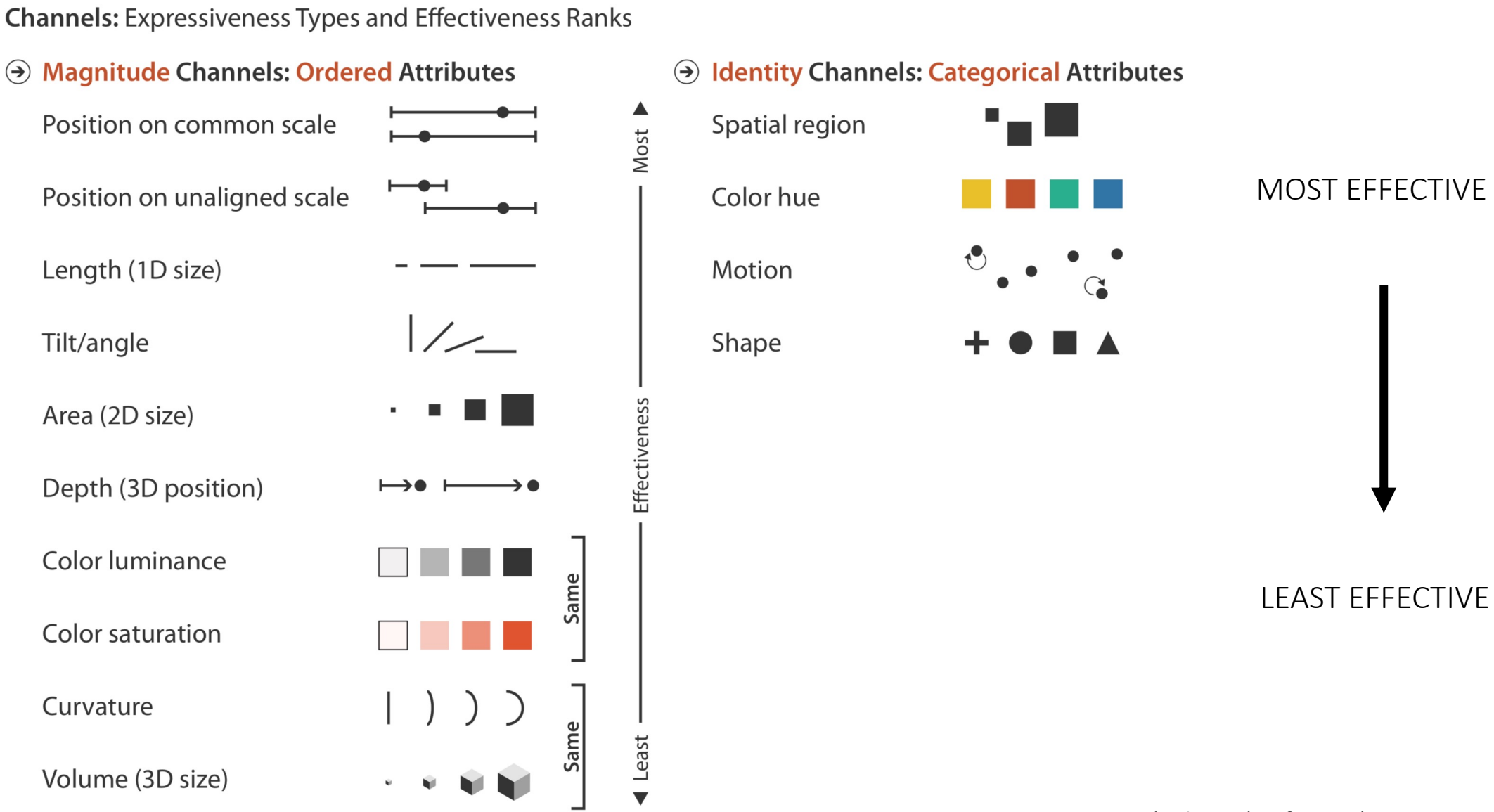
Orientation

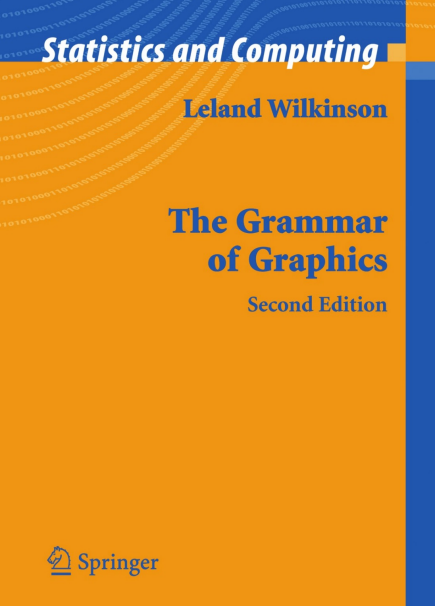


Lightness


Hue

Visual channels don’t work equally well!



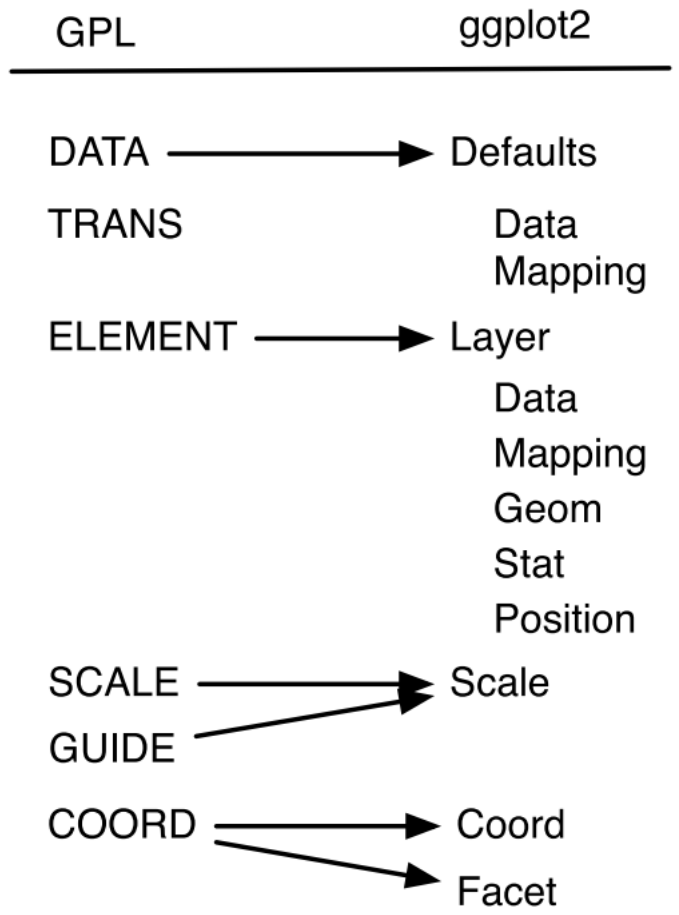


A Layered Grammar of Graphics

Hadley WICKHAM

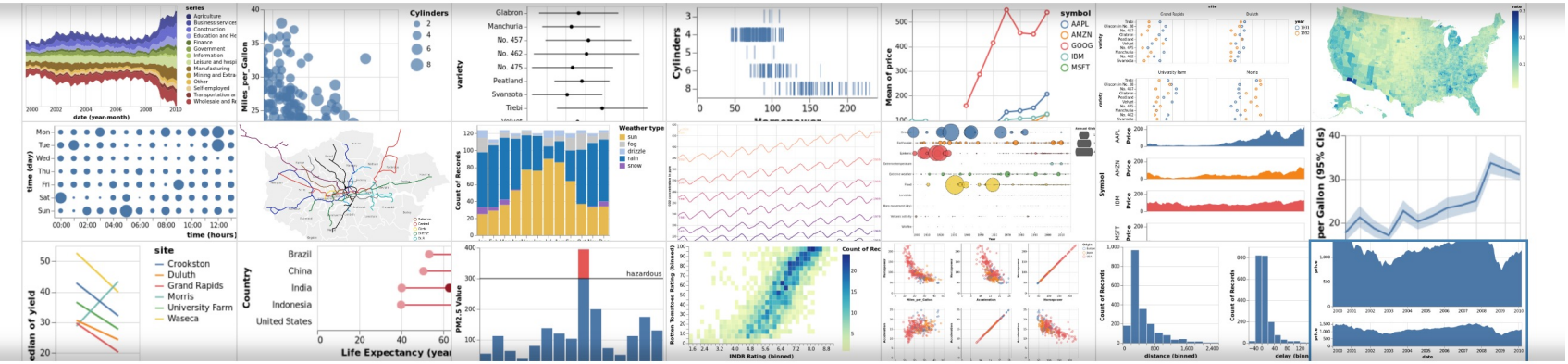
Statistical graphic specifications are expressed in six statements:

1) DATA: a set of data operations that create variables from datasets,
2) TRANS: variable transformations (*e.g., rank*),
3) SCALE: scale transformations (*e.g.*, *log*),
4) COORD: a coordinate system (*e.g.*, *polar*),
5) ELEMENT: graphs (*e.g.*, *points*) and their aesthetic attributes (*e.g.*, *color*),
6) GUIDE: one or more guides (*axes*, *legends*, etc.).



Vega-Lite Examples Tutorials Documentation Usage Ecosystem GitHub Try Online

Vega-Lite – A Grammar of Interactive Graphics



Vega-Lite is a high-level grammar of interactive graphics. It provides a concise, declarative JSON syntax to create an expressive range of visualizations for data analysis and presentation.

Vega-Lite specifications describe visualizations as encoding mappings from data to **properties of graphical marks** (e.g., points or bars). The Vega-Lite compiler **automatically produces visualization components** including axes, legends, and scales. It determines default properties of these components based on a set of **carefully designed rules**. This approach allows Vega-Lite specifications to be concise for quick visualization authoring, while giving user control to override defaults and customize vari
As we also designed Vega-Lite to support data analysis, Vega-Li **transformations** (e.g., aggregation, binning, filtering, sorting) an (e.g., stacking and faceting). Moreover, Vega-Lite specifications layered and multi-view displays, and made **interactive with sele**

Compared to Vega, Vega-Lite provides a more concise and conv specifications to Vega specifications, users may use Vega-Lite a level Vega for advanced use cases.

For more information, read our introduction article to Vega-Lite v Lite v2, see the documentation and take a look at our example g

Get started
Latest Version: 5.2.0

Vega-Lite: A Grammar of Interactive Graphics

Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer

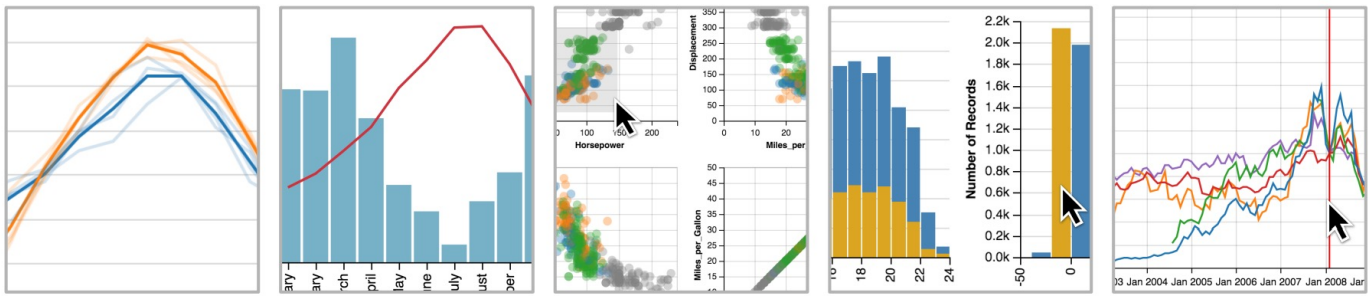
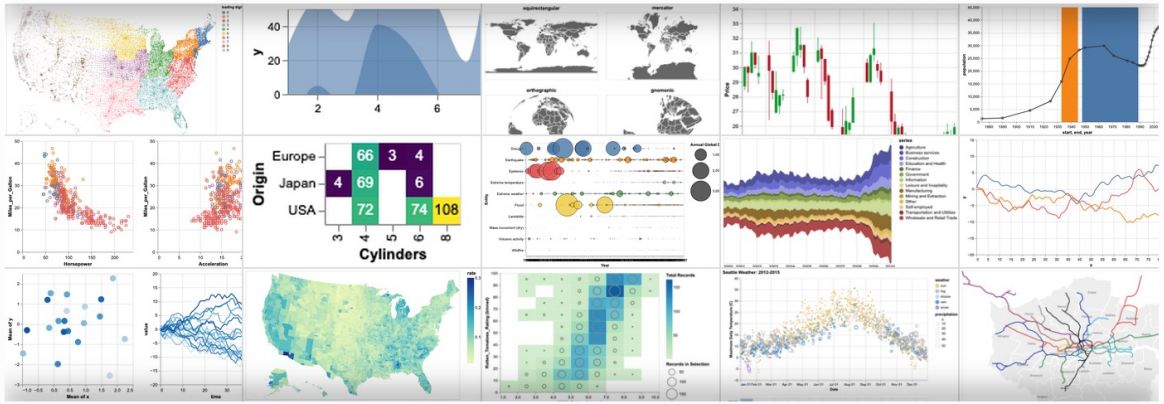


Fig. 1. Example visualizations authored with Vega-Lite. From left-to-right: layered line chart combining raw and average values, dual-axis layered bar and line chart, brushing and linking in a scatterplot matrix, layered cross-filtering, and an interactive index chart.

Altair: Declarative Visualization in Python



Altair is a declarative statistical visualization library for Python, based on Vega and Vega-Lite, and the source is available on GitHub.

With Altair, you can spend more time understanding your data and its meaning. Altair's API is simple, friendly and consistent and built on top of the powerful Vega-Lite visualization grammar. This elegant simplicity produces beautiful and effective visualizations with a minimal amount of code.

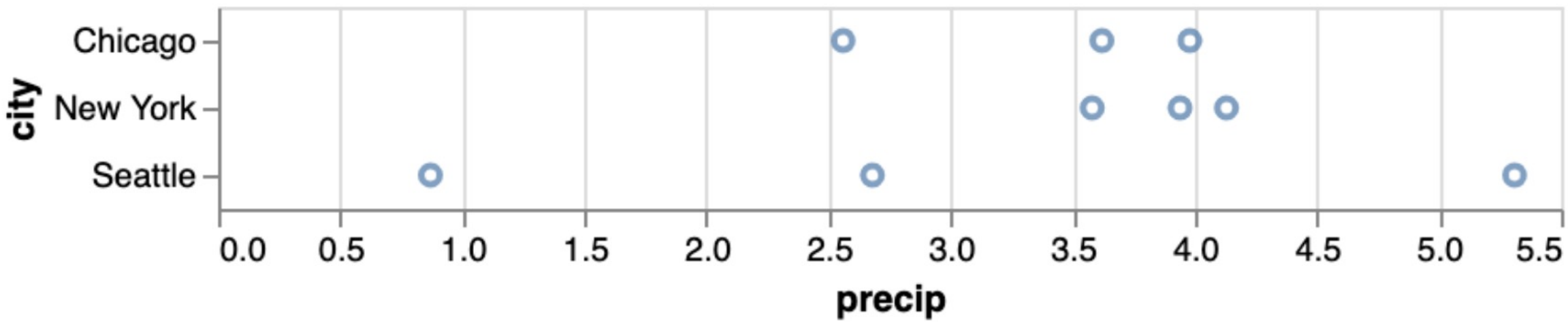
Getting Started

- Overview
- Installation
- Dependencies
- Development Install
- Basic Statistical Visualization

Gallery

- Example Gallery

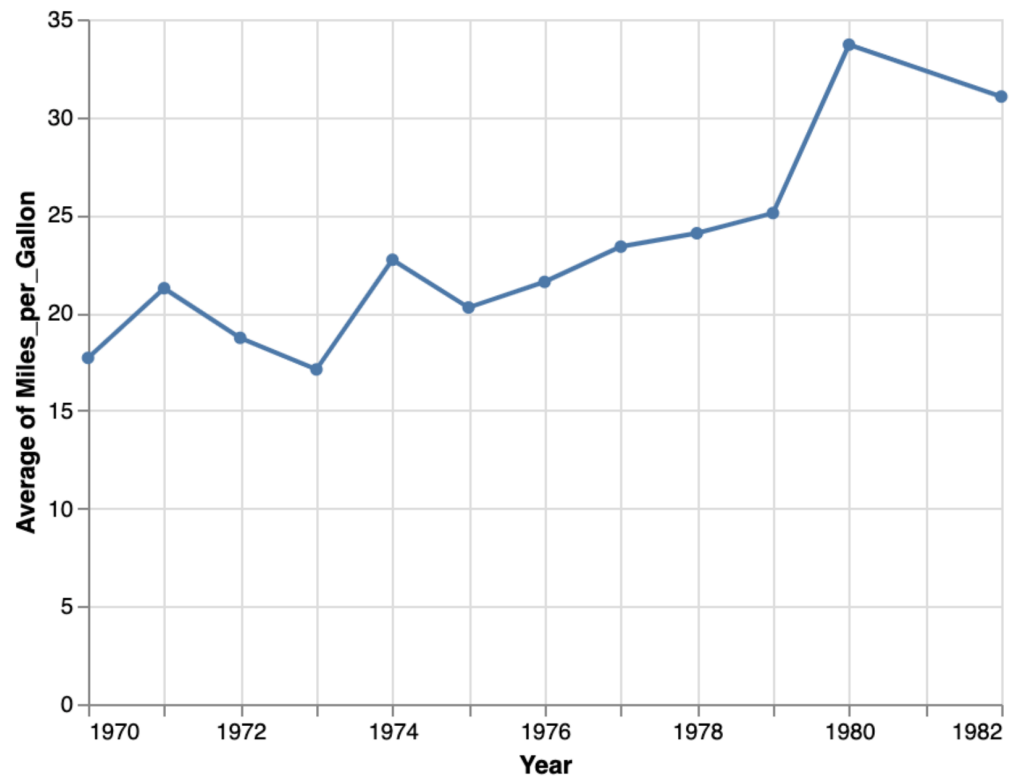
```
alt.Chart(df).mark_point().encode(
    alt.X('precip'),
    alt.Y('city')
)
```



```
line = alt.Chart(cars).mark_line().encode(
    alt.X('Year'),
    alt.Y('average(Miles_per_Gallon)')
)

point = alt.Chart(cars).mark_circle().encode(
    alt.X('Year'),
    alt.Y('average(Miles_per_Gallon)')
)

line + point
```



Let’s get our hands on some code!

Take-aways?

Altair as a declarative, powerful and versatile visualisation toolkit

- Rich specifications, faceting, interaction
- Working seamlessly with Pandas

Active and supportive community (see -- https://github.com/altair-viz)

A few key things to think about when designing a visualisation

- Think first what the objective is? Who is your audience? What message to deliver?
- What the data affords, what works best for the data?
- Which encodings, marks, layout work?
- How to activate the visualisation -- title, colour, style?

github repo

Doing Data Visualisation in Python with Altair, A Primer

Cagatay Turkay

Associate Professor,
Centre for Interdisciplinary Methodologies
University of Warwick

DSSGxUK 2022, University of Warwick, 29 June 2022