Data Visualisation

Narrative & Interactive Data-Vis in Python

18th & 19th Sept 2019
Mumbai, IN

Workshop Notes

Amit & Vikrant

Workshop Introduction

- **Objectives:** Intent & learning outcomes
- **Context:**: Purpose & role of data visualisation, narrative & interactive types, tools & python
- **Approach**: Conceptual & exercise based, group formations
- **Intros:** Participant & facilitator introductions
- **Expectations**: Alignment & discussion
- **Checkin**: Python setup and installation readiness

Objectives

- Understand the value of data visualisation and the role it plays in business analytics and decision making
- Learn the theory of data visualisation including grammar, types, color, annotation, flow, animation, interaction etc.
- Build an understanding of visual perception and cognition to gain an intuitive sense of how data visualisation work
- Get exposure to tools that can be used to create data-visualisation in python.

Context

Learn through practice the two contexts in which data visualisation is used with business stakeholders

– Narrative visualisation (say, telling a compelling data-story in a presentation)
– Interactive visualisation (say, allowing business user to visually explore a complex data-set or a model)

Approach

- The workshop is structured with a mix of conceptual learning (40%) and practice sessions (60%)
- This is a hands-on workshop and we will learn by using python and notebook
- There will be a case-study based approach - Telco Churn Example - to learn from.
- Each day will have four main sessions: see session plan next

Day 1 Sessions

— **Workshop Introduction** (0930 - 1000)
— **Session #1: Value of Data Visualisation** (1000 - 1120)
— Break (1120 - 1140)
— **Session #2: Tools & Abstractions for Data Visualisation** (1140 - 1300)
— Lunch (1300 - 1400)
— **Session #3: Theory of Data Visualisation** (1400 - 1520)
— Break (1520 - 1540)
— **Session #4: Guidelines for Better Data Visualisation** (1540 - 1700)
— **Day One Summary** (1700 - 1730)
— **Data-Story Group Exercise** (1730 - 1900)

Day 2 Sessions

— **Recap & Questions** (0930 - 0950)
— **Data-Story Presentations** (0950 - 1020)
— **Session #5: Crafting Visual Stories with Data** (1020 - 1120)
— Break (1120 - 1140)
— **Data-Story Rework** (1140 - 1210)
— **Session #6: Interactivity** (1210 - 1320)
— Lunch (1320 - 1420)
— **Session #7: Explorable Vis for Business Users** (1420 - 1540)
— Tea Break (1540 - 1600)
— **Session #8: Putting together an Interactive Application** (1600 - 1720)
— **Overall Summary & Way Forward** (1720 - 1800)

Introduction & Expectations

- Tell us about yourself
  - Name & Role
  - Experience with Python
  - Experience with Analytics
  - One thing you want to learn from the workshop

Group Formation

— Groups of 4 or 5
— Mixed experience of Python & Analytics
— For the data-story work & presentation

Installation Check In

Workshop Repo
https://github.com/amitkaps/data-vis-workshop

— Did you follow the instruction in the README.md?
— Any issues you faced?

Narrative Visualisation

#1: Value of Data Visualisation

War Stories & Killer Charts

analysis

numbers

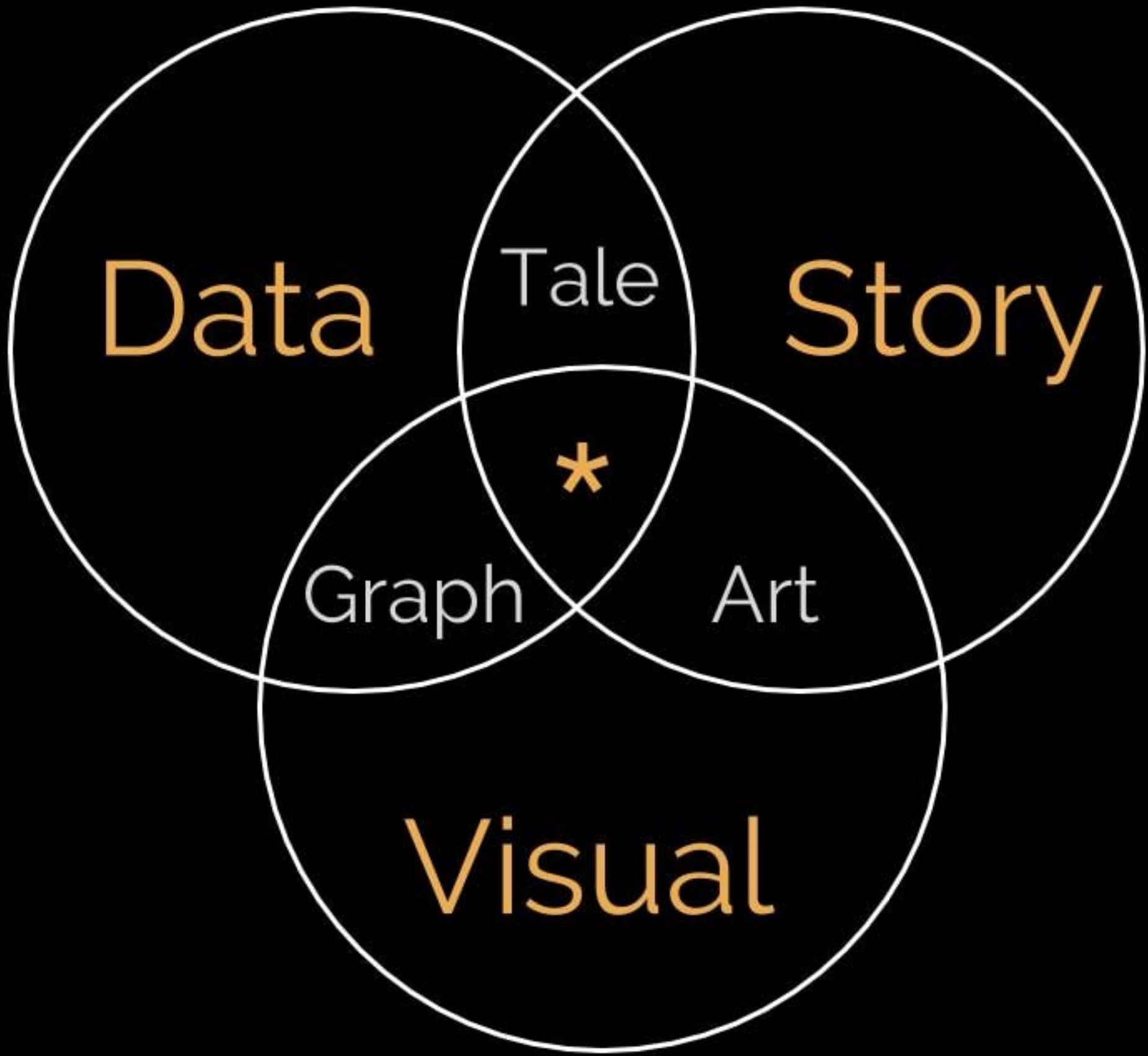
argument

Humans are
pattern-seeking
story-telling
animals.

~~analysis~~ Synthesis

~~numbers~~ Visuals

~~argument~~ Story


Data
Tale
Story
*
Graph
Art
Visual

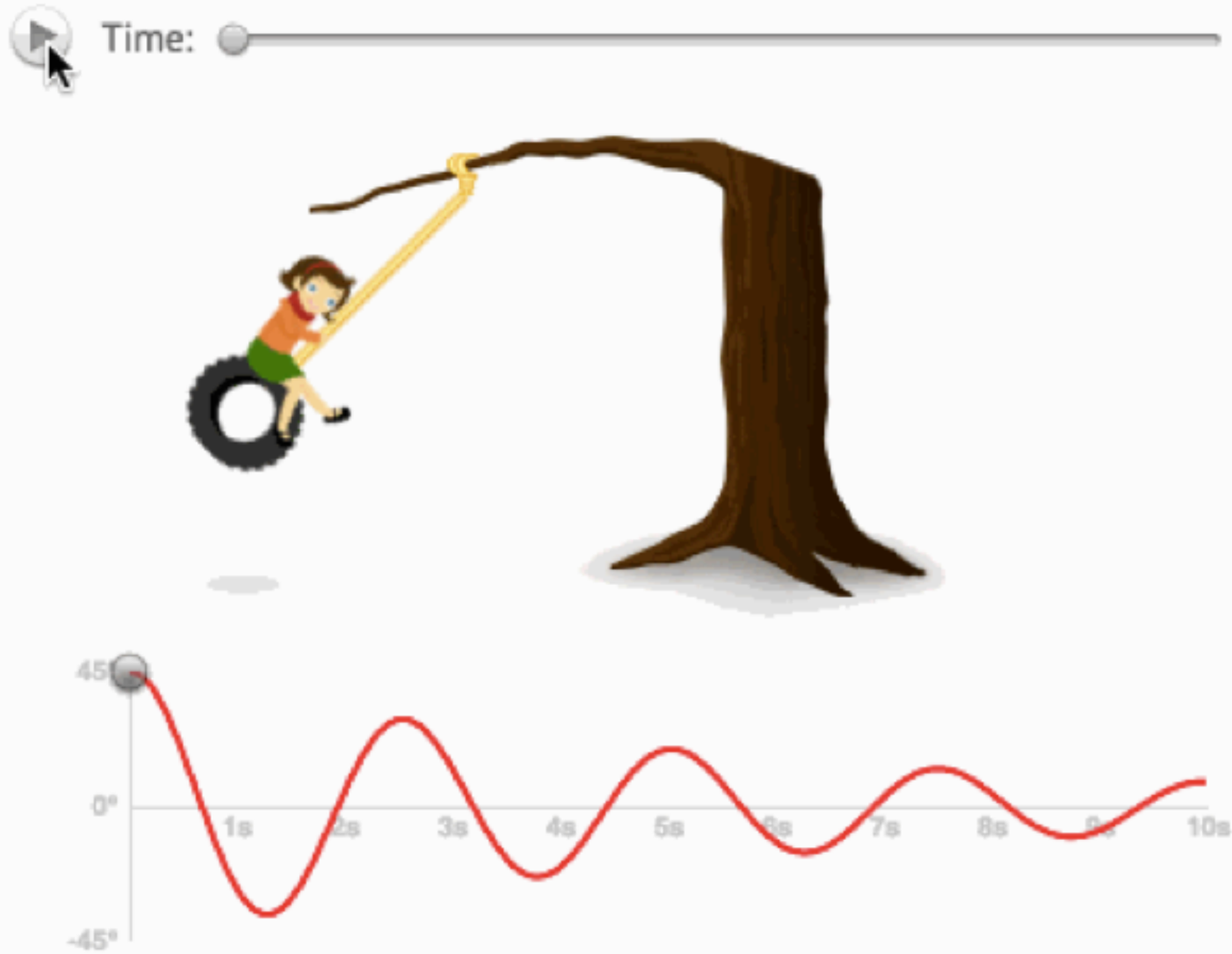
What is visualisation?

Exercise: Visualise this data in at least 10 different ways - using paper & pen

Area	Sales	Profit
North	5	2
East	25	8
West	15	6
South	20	5
Central	10	3

Visualisation is the **transformation** of the **symbolic** into **geometric.**

— McCormick et al.



$$\theta(t) = 0.25\pi\, e^{-0.2t} cos(2.55t)$$

Start with Data

Data

Area	Sales
North	5
East	25
West	15
South	20
Central	10

Identify the Variables

Data		Variables	
Area	Sales	x	y
North	5	1	5
East	25	2	25
West	15	3	15
South	20	4	20
Central	10	5	10

x (Nominal) = Area
y (Quantitative) = Sales

Pick Marks and Scales

Data		Variables		Mark & Scale	
Area	Sales	x	y	x	y
North	5	1	5	20	
East	25	2	25	60	
West	15	3	15	100	
South	20	4	20	140	
Central	10	5	10	180	

x (N) = Area
y (Q) = Sales

x - position
y - bar
scale - 200 x 200

Map to Coordinate

Variables

x	y
1	5
2	25
3	15
4	20
5	10

x (N) = Area
y (Q) = Sales

Mark & Scale

x	y
20	
60	
100	
140	
180	

x - position
y - bar
scale - 200 x 200

Coordinate

cartesian

Example #1

Points

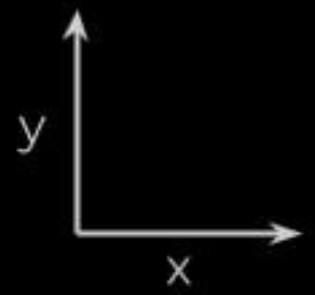
Line

Bar

Bar - Stacked

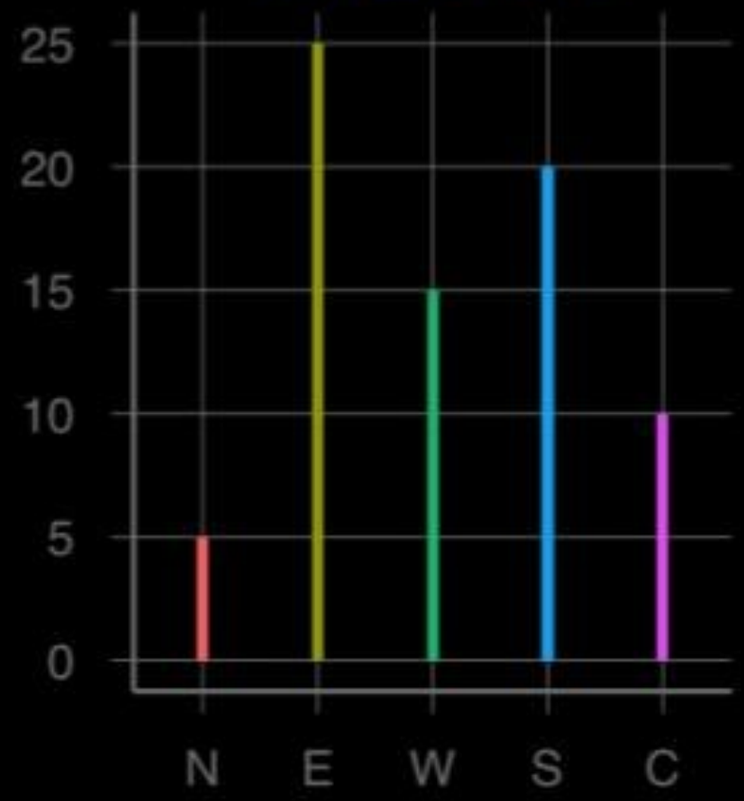
Bar - Stagger

Coordinates
Cartesian

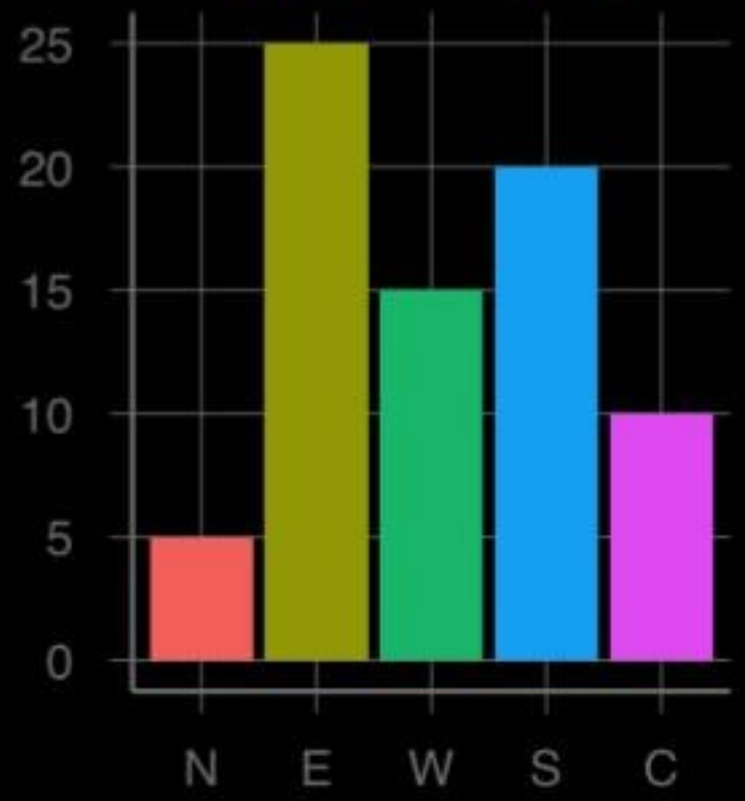


Dot Plot

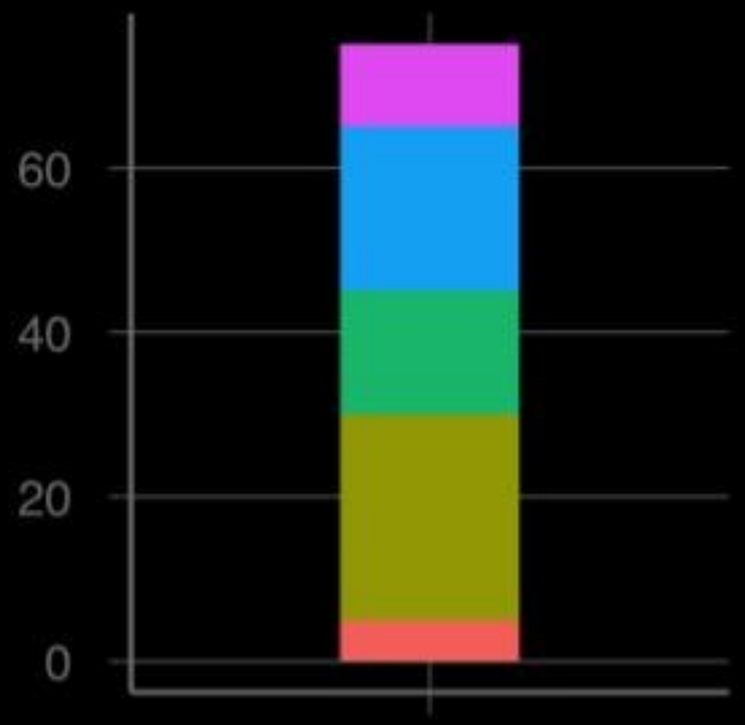
Line Chart



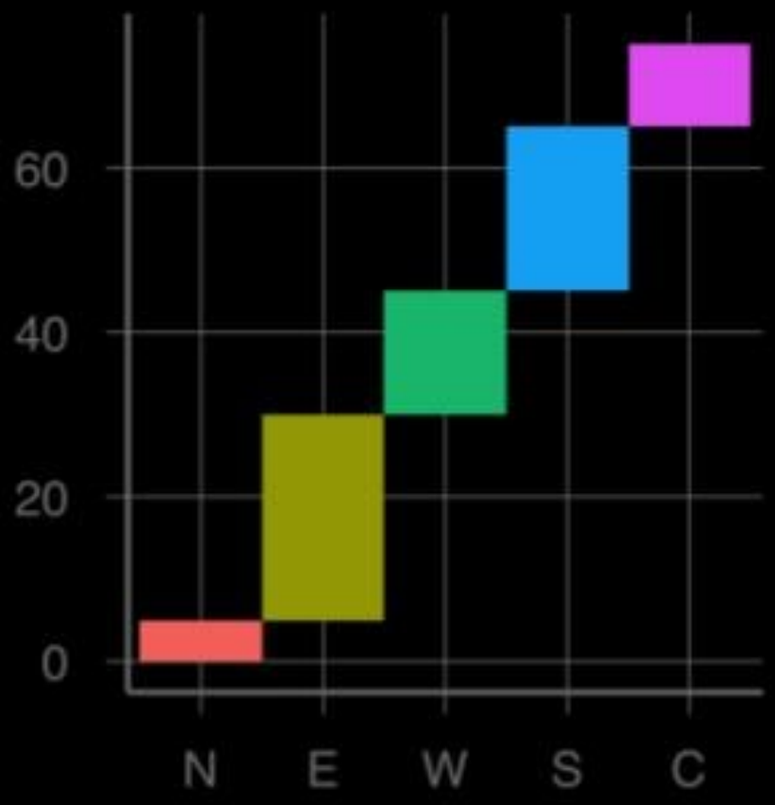
Column Chart

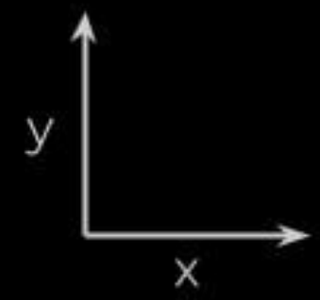


Stacked Column



Waterfall





Vis Examples: 1-5

Points

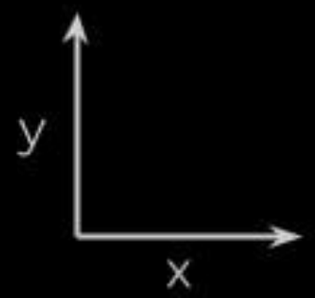
Line

Bar

Bar - Stacked

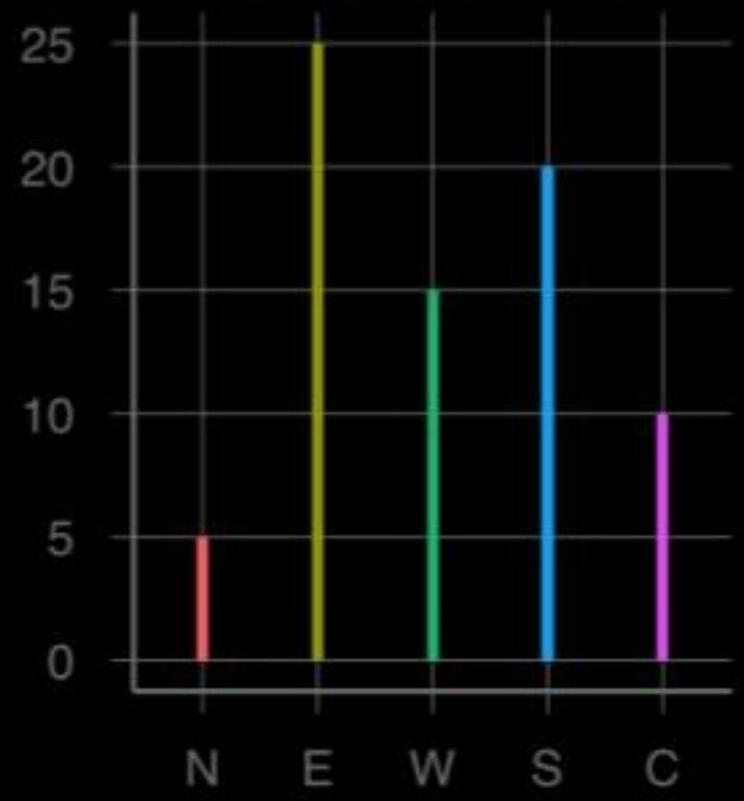
Bar - Stagger

Coordinates
Cartesian

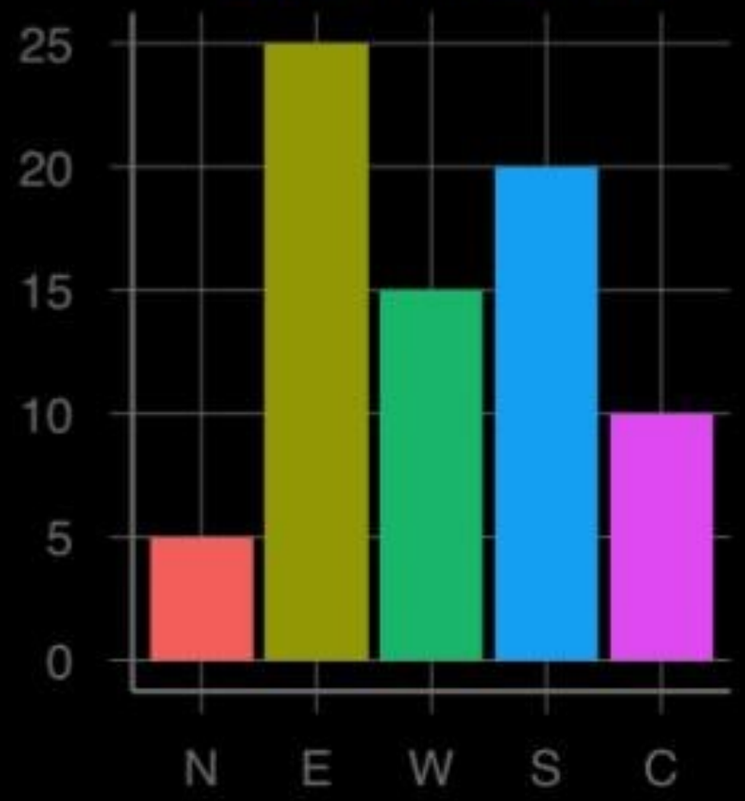


Dot Plot

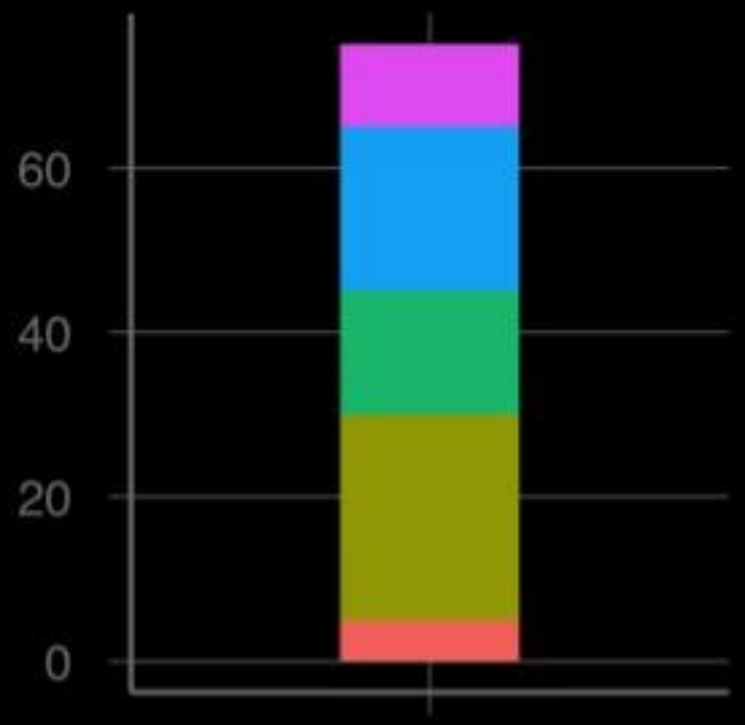
Line Chart



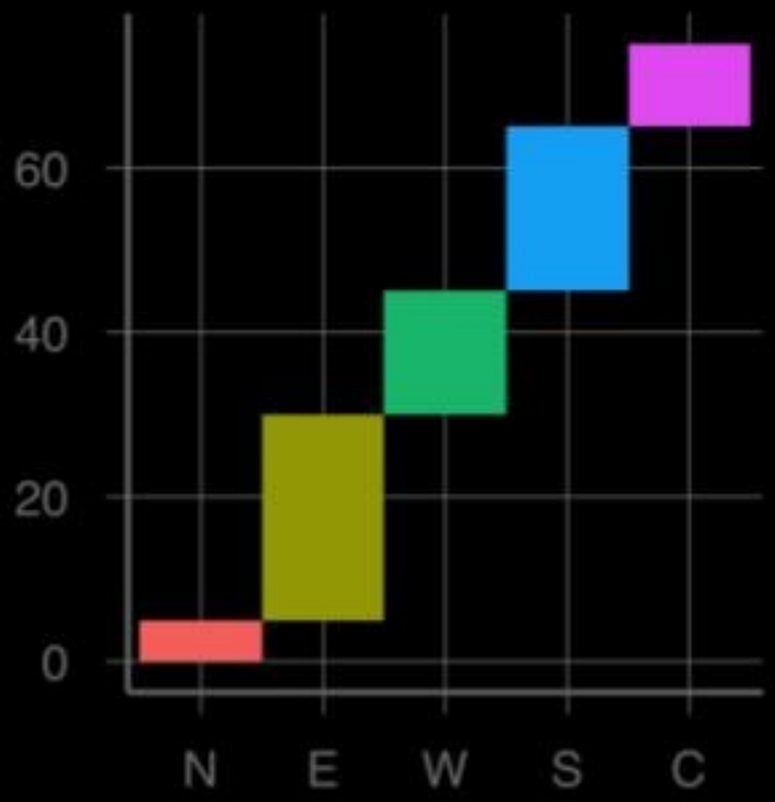
Column Chart

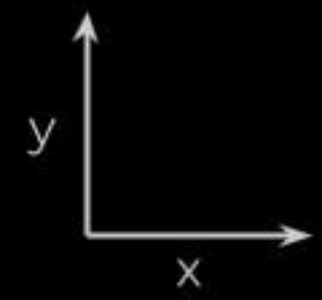


Stacked Column



Waterfall





Vis Examples: 6-10

Points

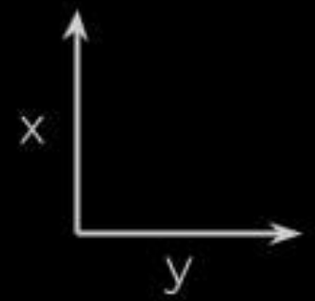
Line

Bar

Bar - Stacked

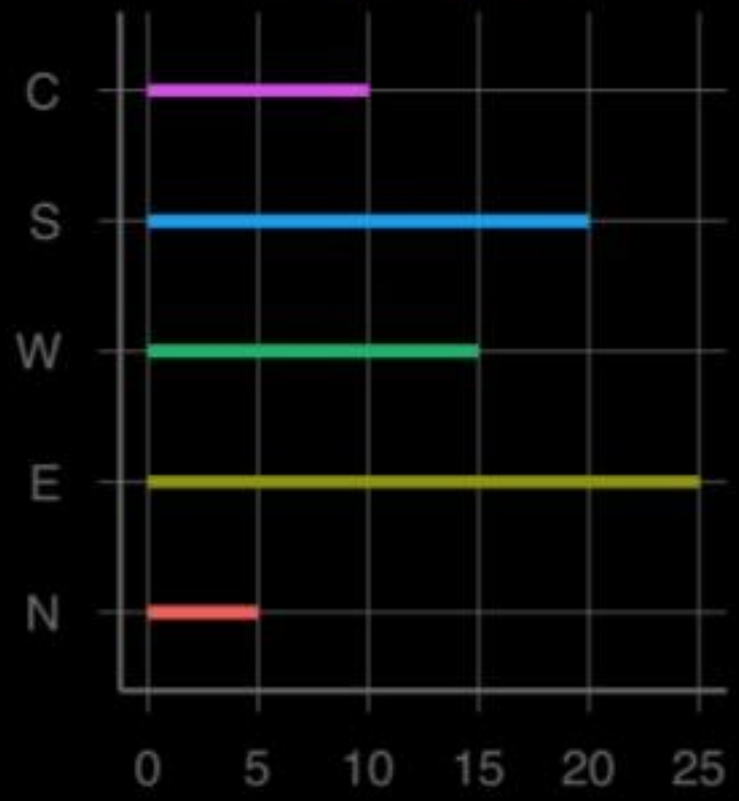
Bar - Stagger

Coordinates
Cartesian - Flip

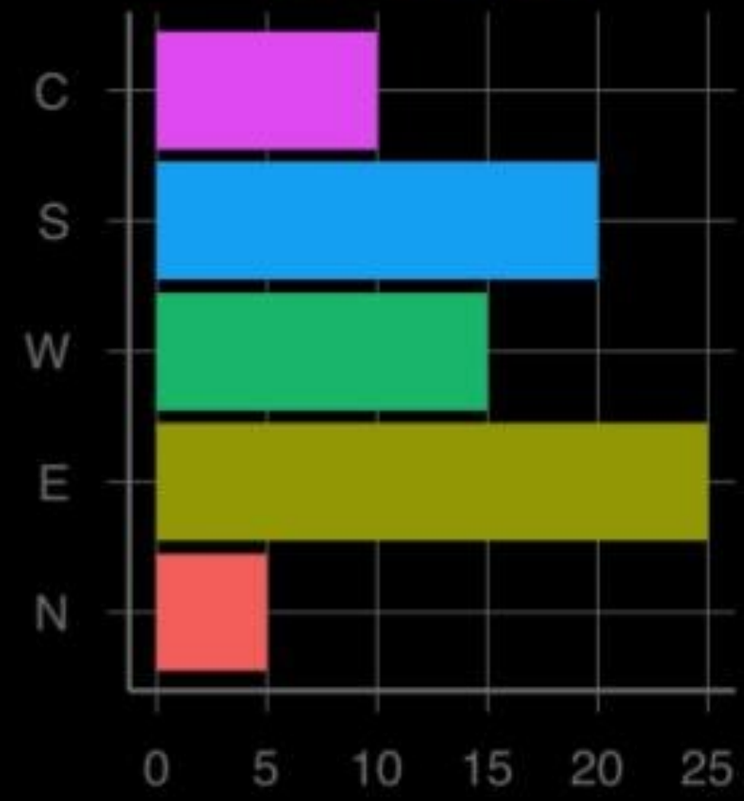


Dot Plot

Line Chart

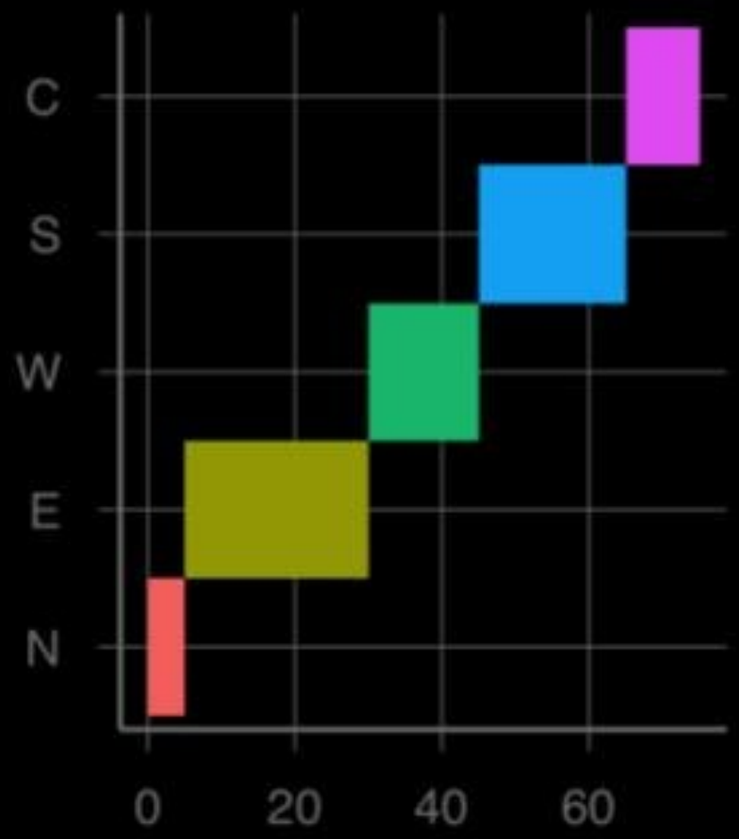


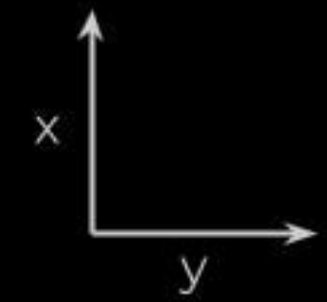
Bar Chart



Stacked Bar

Cascade





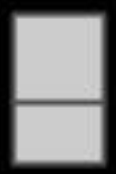
Vis Examples: 11-15

Points

Line

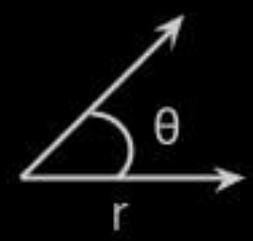
Bar

Bar - Stacked



Bar - Stagger

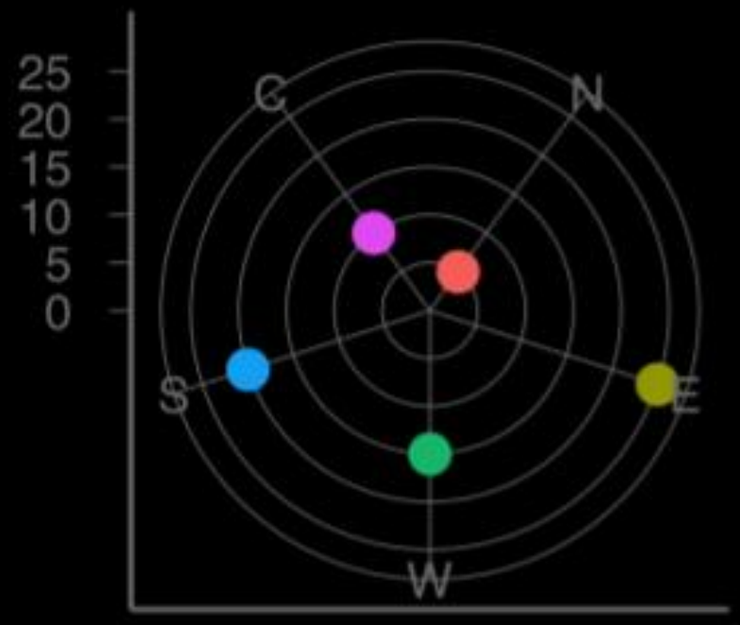
Coordinates
Polar - x



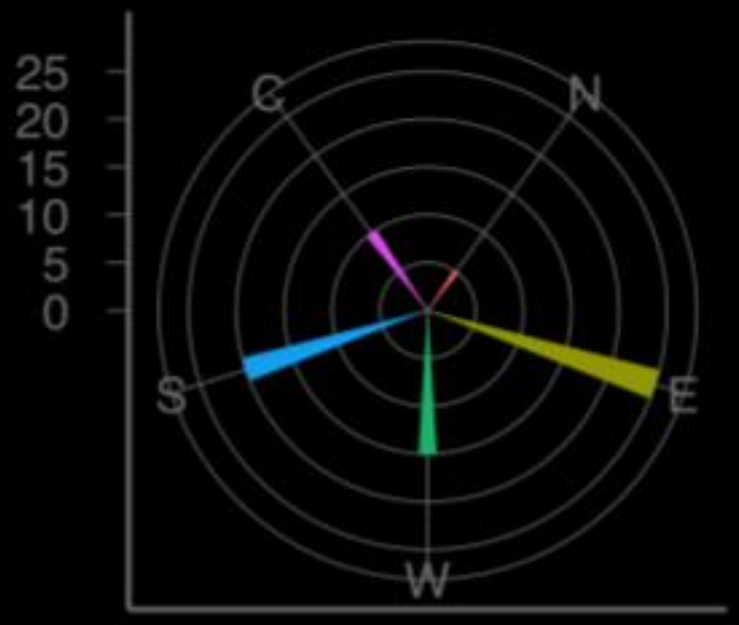
$x = \theta$

$y = r$

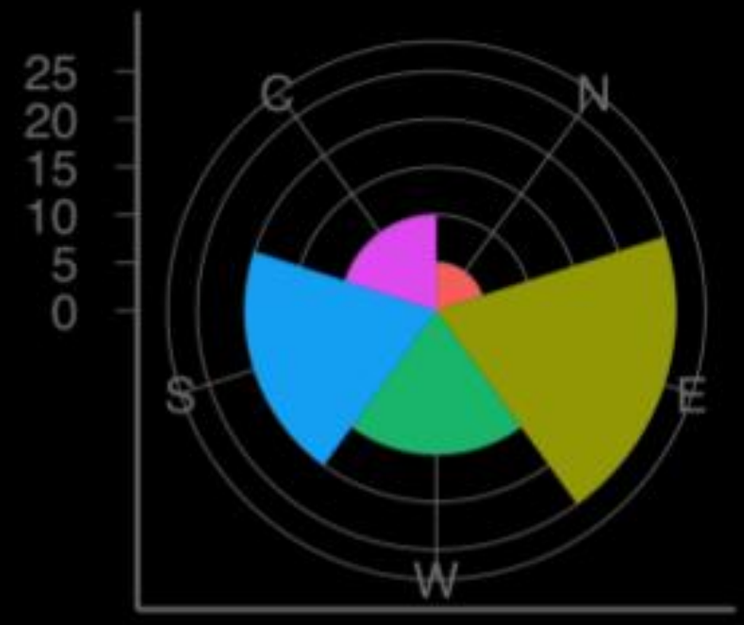
Marked Radar



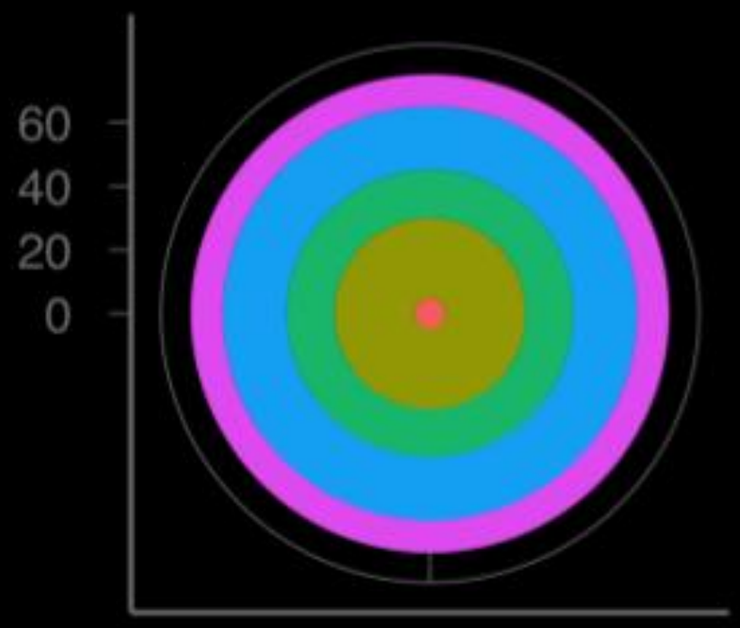
Line Radar?



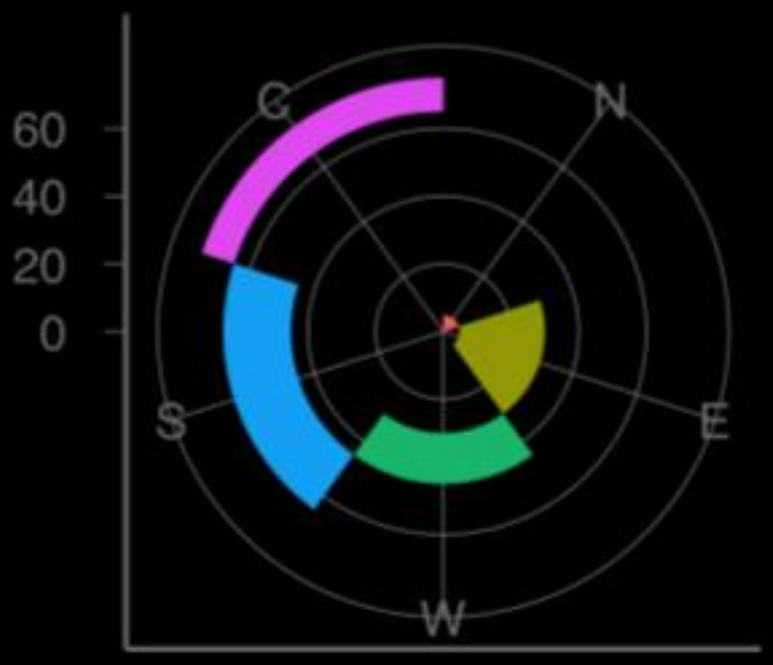
CoxComb

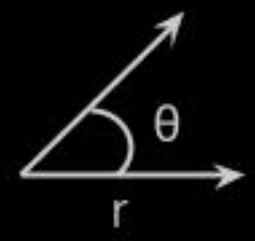


Bullseye



Polar Waterfall?





$x = \theta$

$y = r$

Vis Examples: 16-20

Points

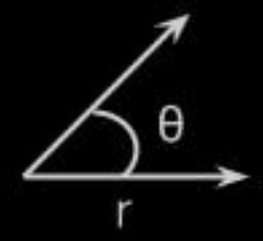
Line

Bar

Bar - Stacked

Bar - Stagger

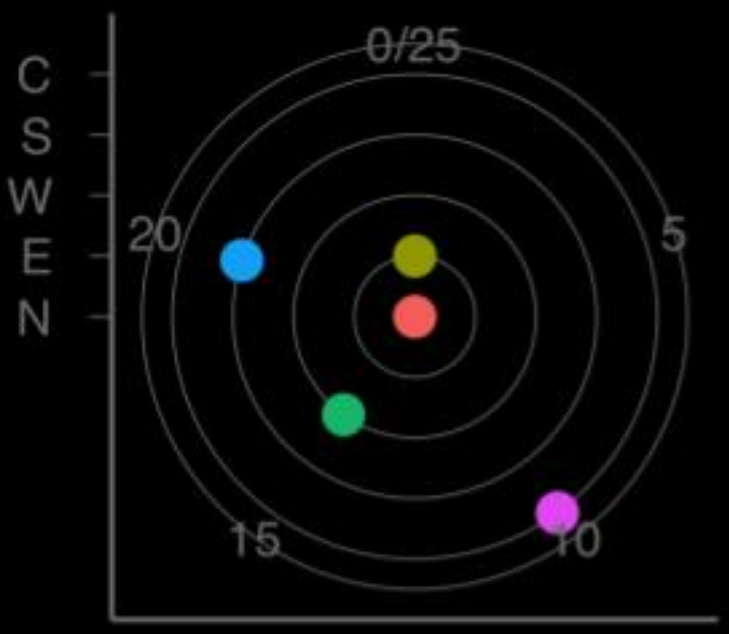
Coordinates
Polar - Y



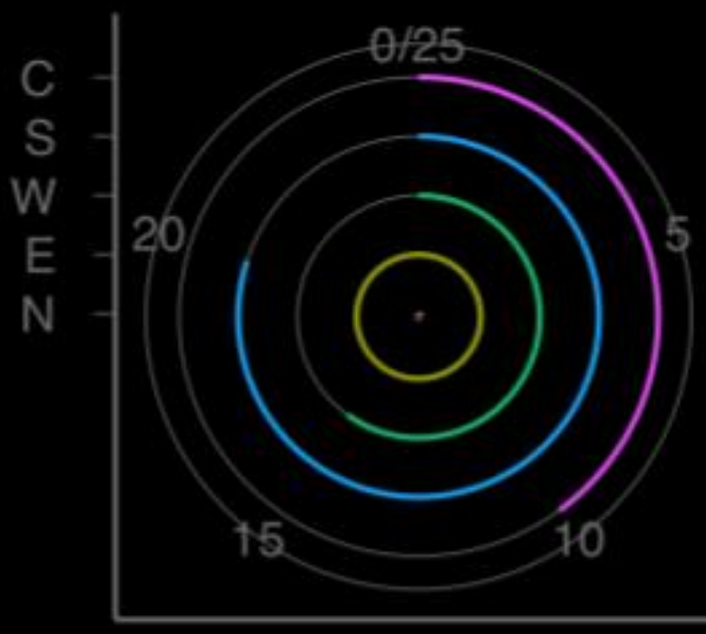
$x = r$

$y = \theta$

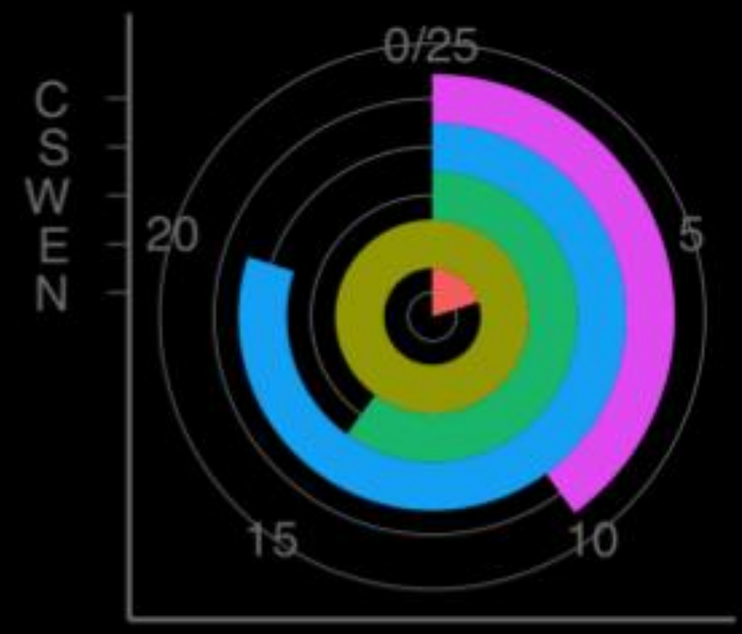
Target?



Wind Rose?

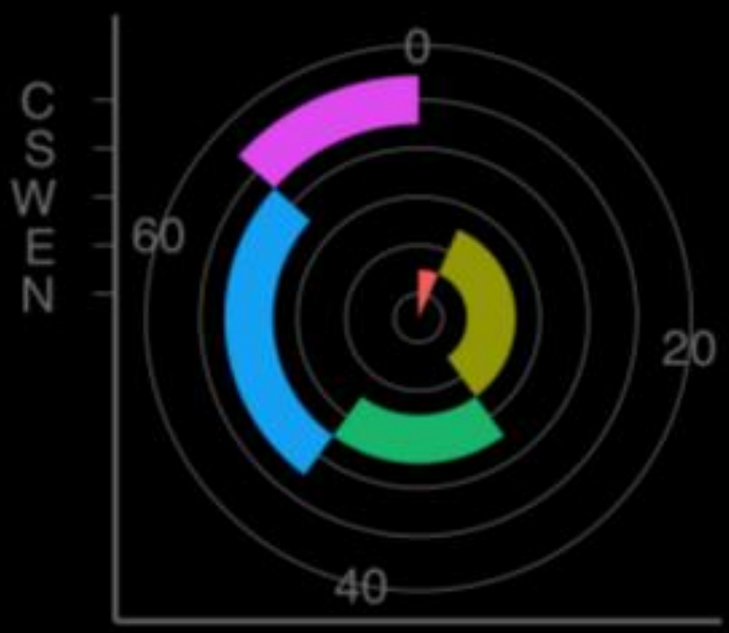


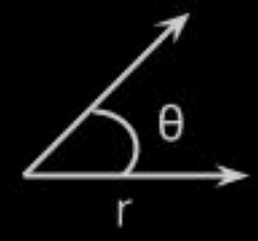
Wind Rose



Pie Chart

Polar Cascade?





$x = r$

$y = \theta$

Learning to See

Visualisation is the use of computer-generated, **interactive, visual representations** of **abstract data** to **amplify cognition.**

— Card, Mackinlay, & Shneiderman

Types of data visualisation with examples

— Explore & Explain
— Static & Interactive
— Analytical & Emotive

Explanatory

Narrative

Exploratory

Interactive

Make decisions

Persuade | Tell a story

Share | Collaborate

Inspire

Expand memory

Answer questions

Find patterns

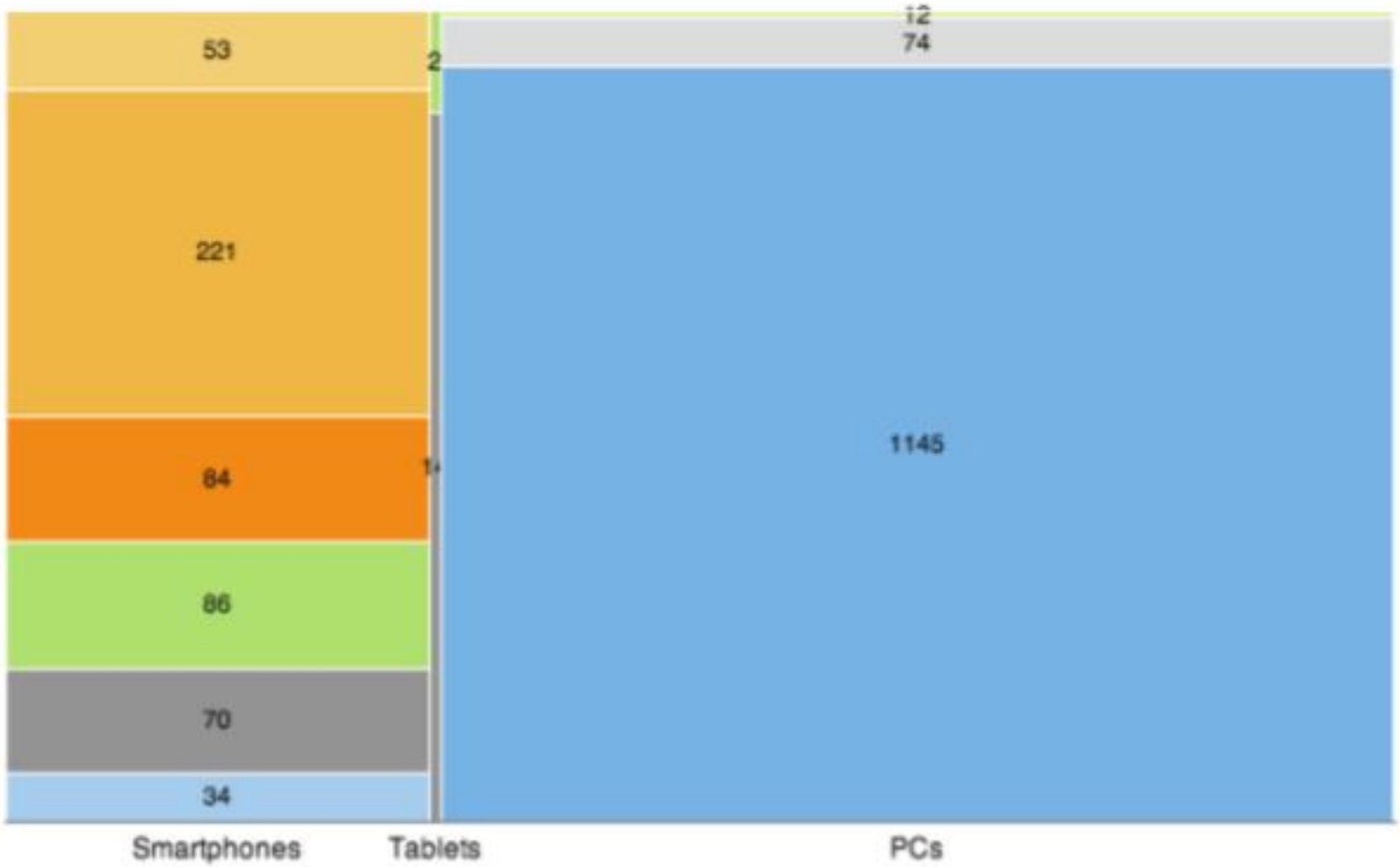
See data in context

Explanatory

narrativeVIZ

Growth of Android and Apple Ecosystem

Installed device estimates by OS in million units by Amit Kapoor



Choose the year to see the transition

2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017

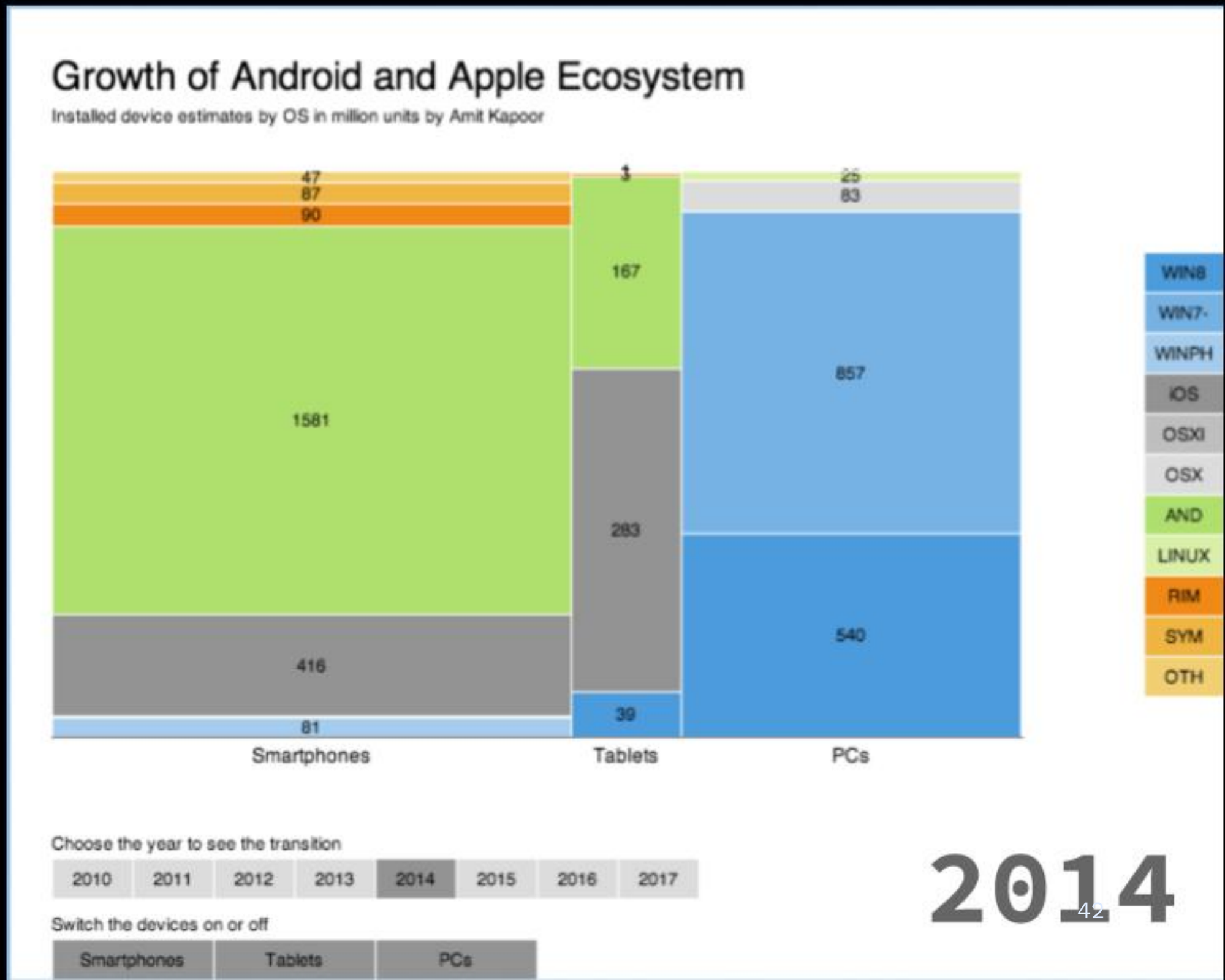
2010

Switch the devices on or off

Smartphones | Tablets | PCs

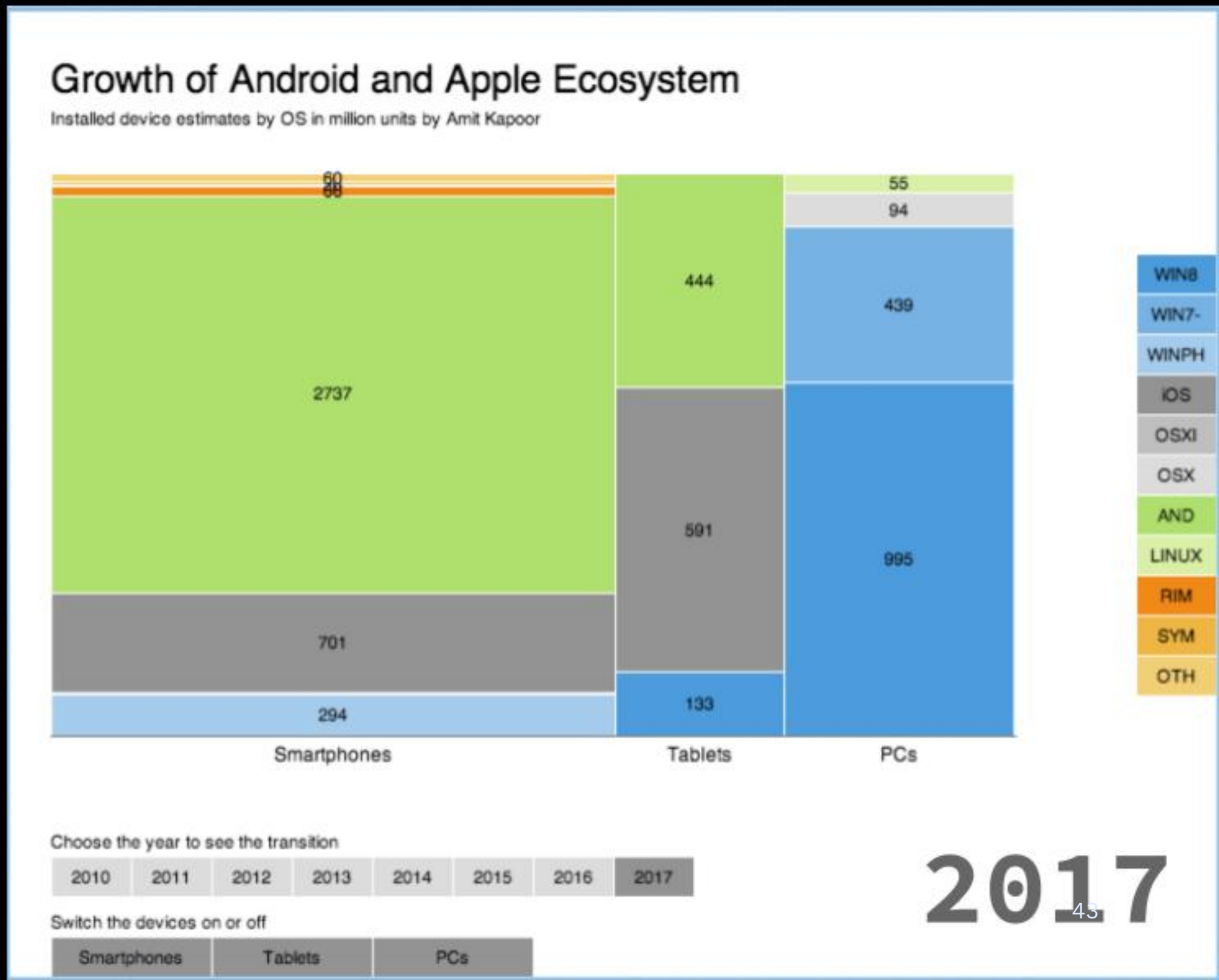
Explanatory

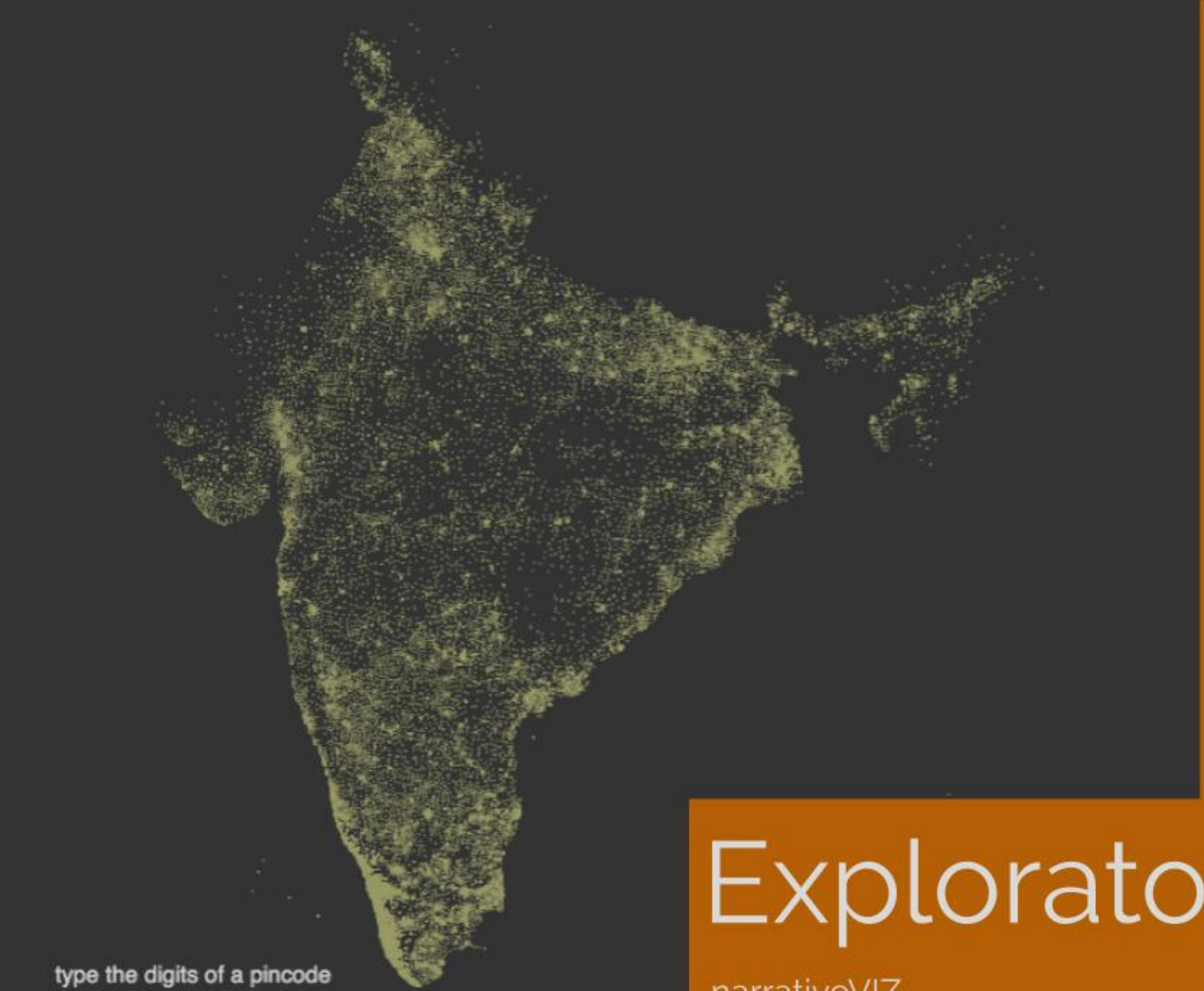
narrativeVIZ



Explanatory

narrativeVIZ





Exploratory

narrativeVIZ

Exploratory

narrativeVIZ

Exploratory

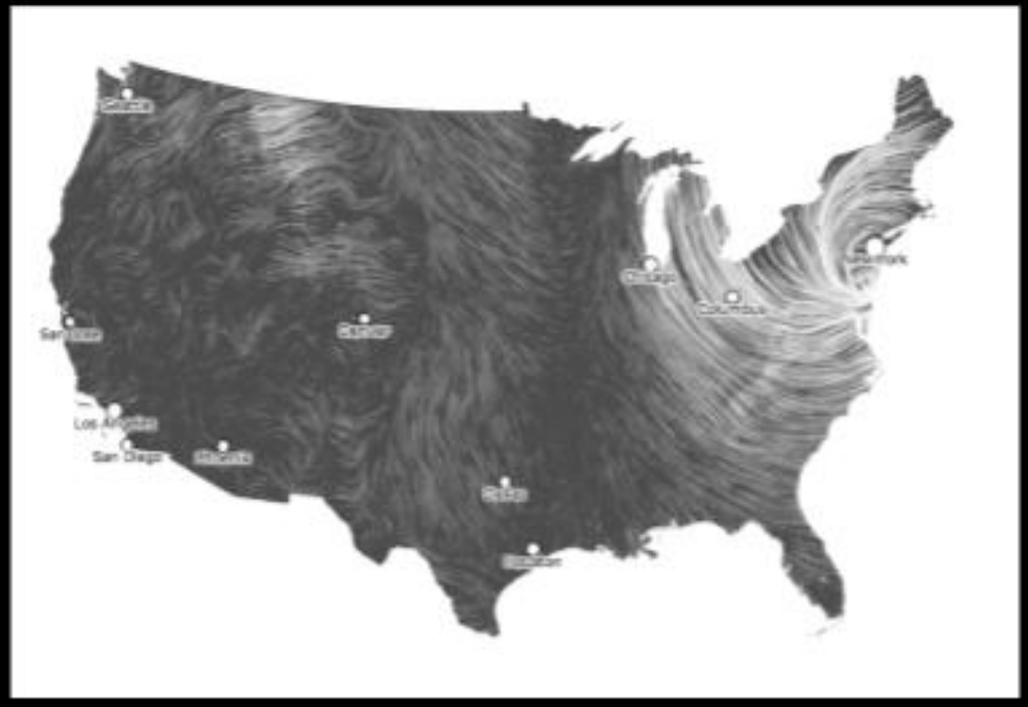
narrativeVIZ

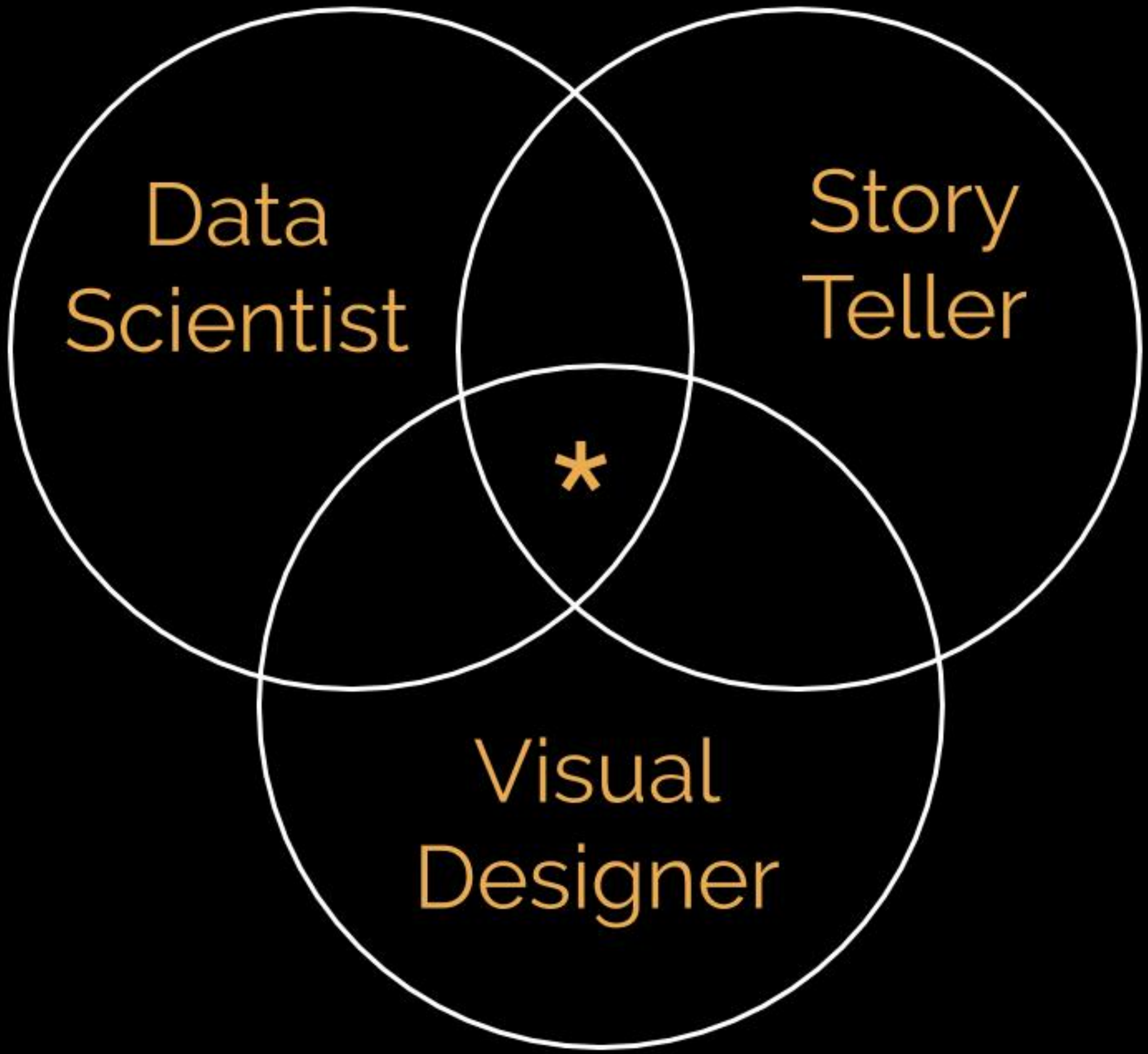
Explanatory	Exploratory
Narrative	Interactive
Strong Order	Weak Order
Heavy Messaging	Light Messaging
Limited Interactivity	Free Interactivity
Author Driven	Reader Driven

Tone of Visualization

Analytical & Pragmatic

Emotive & Abstract




Data
Scientist
Story
Teller
*
Visual
Designer

Analytics Process, Visualisation & Audience

- Linkage with analytical problem solving process: Why - What - How - So What - So Why
- Audience needs and understanding: type, data & visual literacy
- Engagement in different context: personal, presentation & participation

WHY

WHAT

HOW

SO WHY

SO WHAT

Analytics Process

1. **The Why** — Define the Problem
2. **The What** — Frame the Analytical Questions
3. **The How** — Conduct Analytics: Data–Model–Vis
4. **The So What** — Build the Insights
5. **The So Why** — Explain through Narrative

Why

PROBLEM DOMAIN
SITUATION
IMPACT / VALUE

What

QUESTIONS
HYPOTHESES

How

DATA – ACQUIRE / TRANSFORM / EXPLORE
MODEL – INQUISITIVE
PREDICTIVE, CAUSAL

So What

FINDINGS
INSIGHT

So Why

STORY VISUAL
CASE FOR CHANGE

Visualisation & Analytics

— Visual Exploration: Exploratory Data Analysis
— Insight Dashboard: Interactive Visualisation
— Visual Explanation: Narrative Visualisation

Why

PROBLEM DOMAIN
SITUATION, PROBLEM
IMPACT / VALUE

What

QUESTIONS
HYPOTHESES

How

DATA
ACQUIRE
REFINE
TRANSFORM

VISUAL EXPLANATION

INSIGHT DASHBOARD

VISUAL EXPLORATION

MODEL
INFERENTIAL
PREDICTIVE
CAUSAL

So Why

STORY
CASE FOR CHANGE

So What

FINDINGS
INSIGHT

Audience Types

- Analytics: Rich data and modelling experience, high comfort with complex usage
- Operational: Mainly metrics driven with detailed drill-down requirements to build understanding
- Business: CxO or Business roles, looking for explanation and actions

Audiences & Participation Mode

Visual Exploration:
Personal mode with Analytics Audience

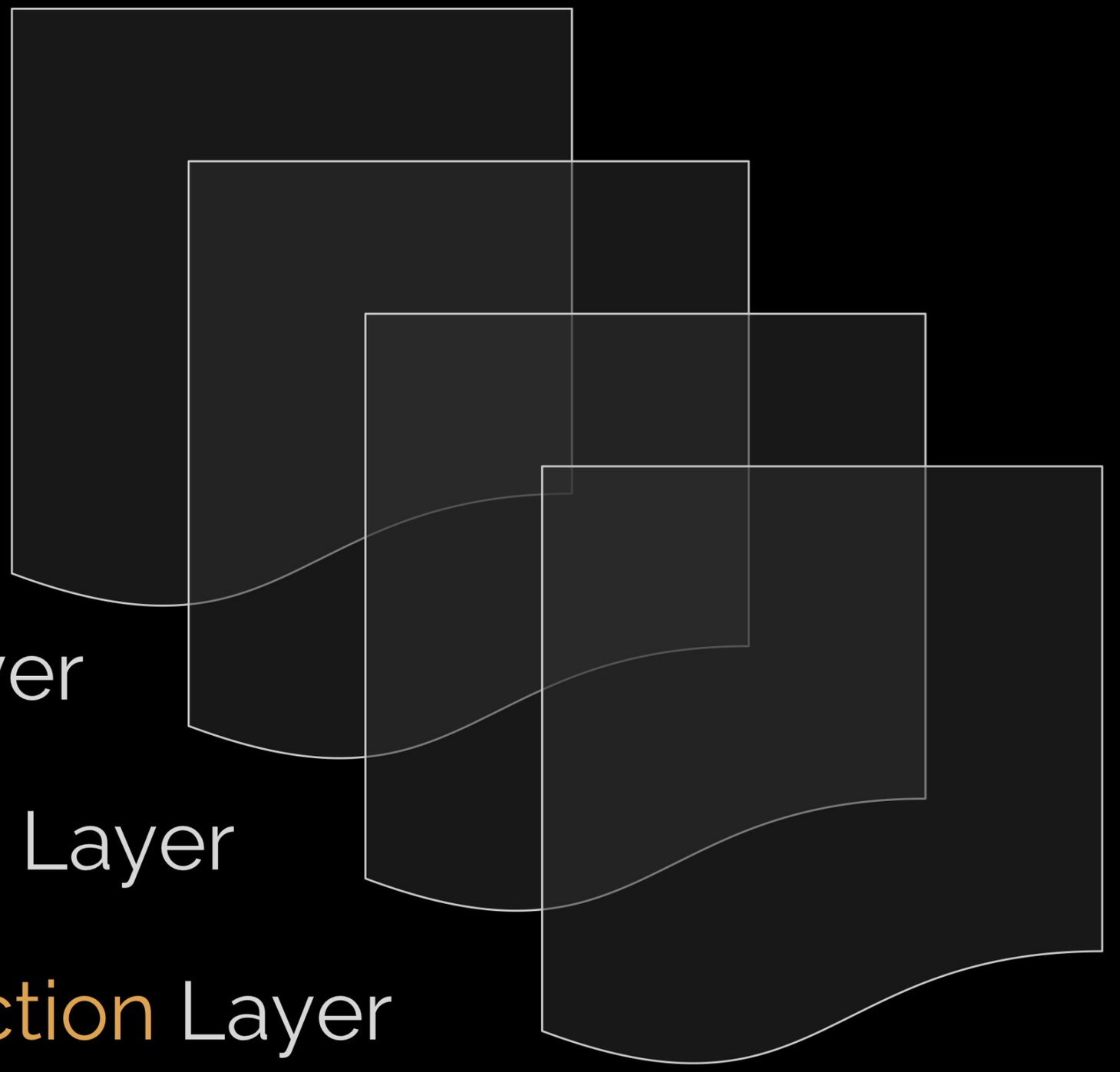
Insight Dashboard:
Participative mode with Operational Audience

Visual Explanation
Presentation mode with Business Audience

Four Layers of Visual Abstraction

— Data Layer
— Visual Layer
— Annotation Layer
— Interaction Layer

Layers of Abstraction



Data Layer

Visual Layer

Annotation Layer

Interaction Layer

Data Layer

Types
- Categorical: Nominal, Ordinal
- Continuous: Temporal, Quantitative

Transforms
- Reshape (e.g. tall <-> wide)
- Aggregation (e.g. bins)
- Basic Stats (e.g. min, max, sum, ...)
- Calculate e.g. New Variables, Window
- Filtering, Sampling

Visual Layer

— Marks: Points, Ticks, Lines, Bar, Area, Glyphs,
Polygon, ...

— Channels: Position-X, Position-Y, Size, Color, Shape, Path, ...

— Scale: Linear, Log, ...

— Coordinate: Cartesian, Polar, Geo, Parallel

— Layout: Single, Facet, Multi-Chart

Annotation Layer

— Title and Labels
— Axis and Tick marks
— Legends
— Grids and Reference Marks
— Text Annotation
— Story Elements

Interaction Layer

— Select e.g. Highlight
— Explore & Navigate e.g. Pan, Zoom, Scale, Rotate
— Connect e.g Brushing & Linking
— Filter & Conditions e.g. Dynamic Queries
— Reconfigure e.g. Sorting
— Transition e.g. Scrolling, Layers
— Staging & Animation

Data Types

— What are the types of data on which we are learning?

— Can you give example of say measuring temperature?

Data Types e.g. Temperature

Categorical

- Nominal: Burned, Not Burned
- Ordinal: Hot, Warm, Cold

Continuous

- Interval: 30 °C, 40 °C, 80 °C
- Ratio: 30 K, 40 K, 50 K

Data Types

Categorical

- Nominal (N) e.g. OSX, Windows, Android
- Ordinal (O) e.g. Good, Better, Best

Continuous

- Interval (zero arbitrary) e.g.
  -> Temporal (T): dates & time
  -> Geographic: latitude & longitude
- Ratio (Q) (zero fixed) e.g. length, mass

Data Types: Operations

Categorical

- Nominal: = , !=
- Ordinal: =, !=, >, <

Continuous

- Interval: =, !=, >, <, -, % of diff
- Ratio: =, !=, >, <, -, +, %

Data Structures

— Tabular (2d arrays) e.g. spreadsheet
— n-Dimension arrays e.g. images, videos
— Hierarchical data e.g. folders of text
— Spatial data e.g. for maps
— ...

Common Data Types & Structure

Matrix	Example	Shape
2D	Tabular	(samples, features)
3D	Sequence (Time & Text)	(samples, steps, features)
4D	Images	(samples, height, width, channels)
5D	Videos	(samples, frames, height, width, channels)

#2: Tools & Abstractions for Data Visualisation

Tools Abstraction:

— Charting vs. Grammar vs. Canvas
— Static vs. Interactive
— SVG vs Canvas vs WebGL Rendering
— Data loading & transformation strategy

Charting

Collection of fixed charts that require data to be shaped in a particular way

Pixel

Paint directly on a canvas. Design & manage every pixel & interaction

Charting

Collection of fixed charts that require data to be shaped in a particular way

Pixel

Paint directly on a canvas. Design & manage every pixel & interaction

Grammar

Collection of graphical primitives for composing data driven graphics

Charting

Collection of fixed charts that require data to be shaped in a particular way

Pixel

Paint directly on a canvas. Design & manage every pixel & interaction

Grammar

Collection of graphical primitives for composing data driven graphics

Charting

Collection of fixed charts that require data to be shaped in a particular way

Code

(imperative)

Describe how each interaction needs to be computed

Visual

GUI tools allowing flexibility to make interactions

Pixel

Paint directly on a canvas. Design & manage every pixel & interaction

Grammar

Collection of graphical primitives for composing data driven graphics

Charting

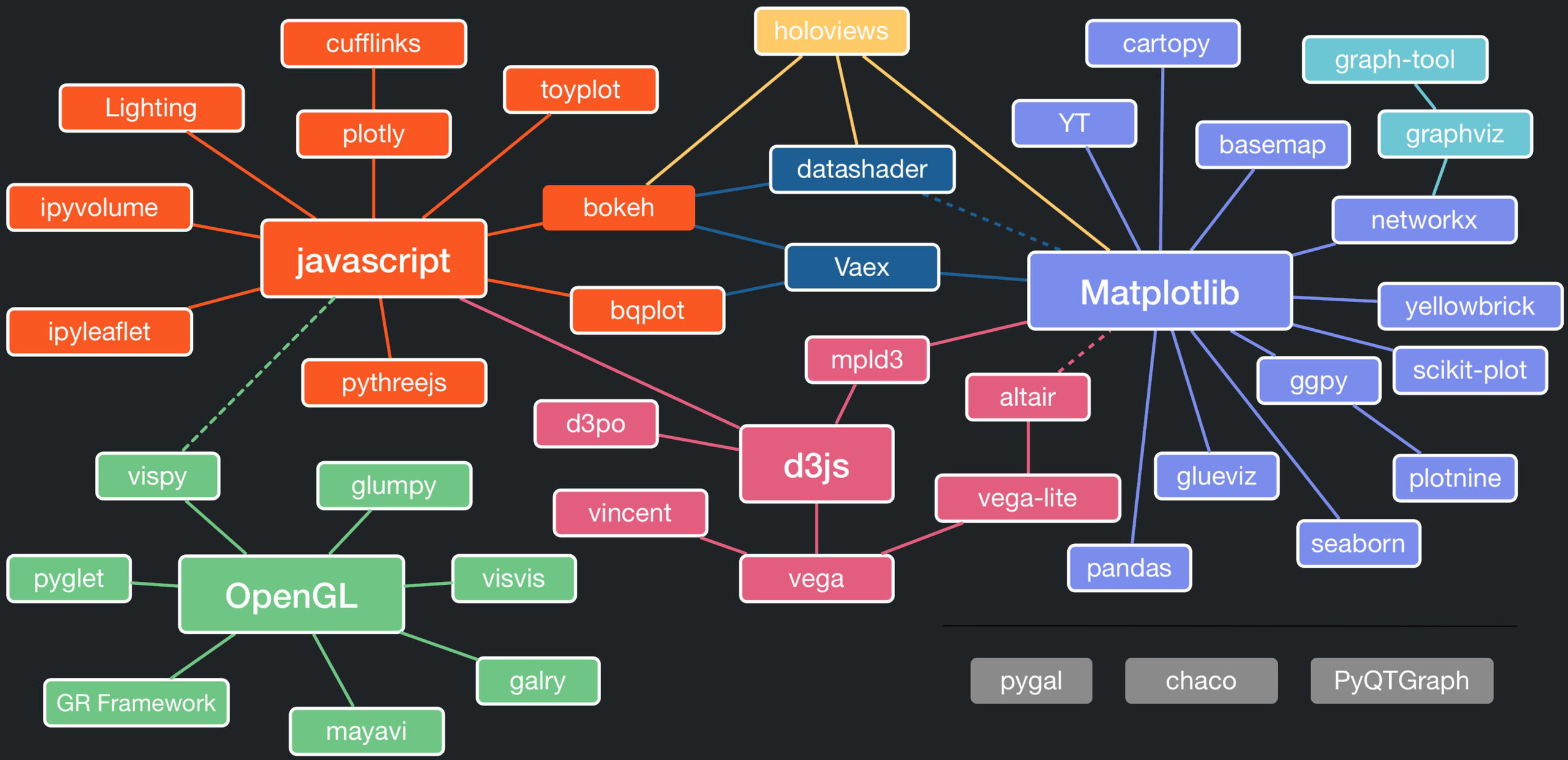
Collection of fixed charts that require data to be shaped in a particular way

Code

(imperative)

Describe how each interaction needs to be computed

Python Data-Vis Libraries



Python Libraries

Static e.g.
`matplotlib`, `seaborn`, `pandas`

Interactive e.g.:
`altair`, `bokeh`, `plotly`, `holoviews`

Rendering Capabilities

Number of Data Points
— SVG: ~10^3
— Canvas: ~10^4
— Web.gl: ~10^6

Exercise: Start with the small dataset

— Getting started with Altair
— Let us use the sample dataset
— Please make at least 2 visualisation from the hand-drawn examples

Case Introduction

— Context: Telco Churn Data
— Start with Why & What
— Dataset introduction

WHY

WHAT

HOW

SO WHY

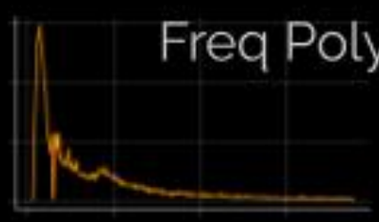
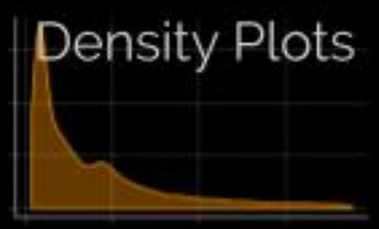
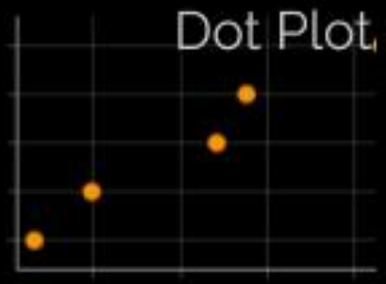
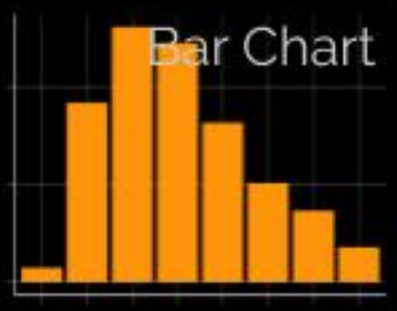
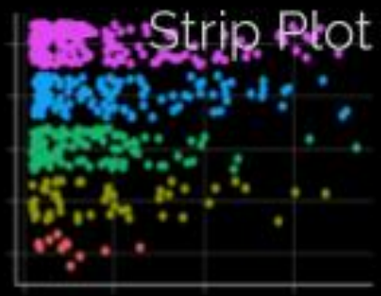
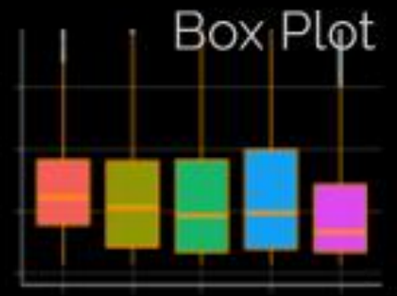
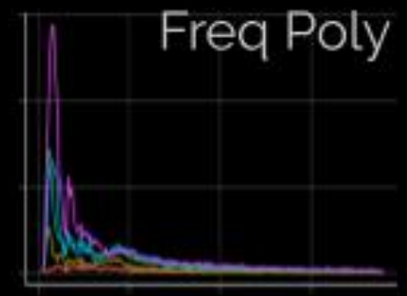
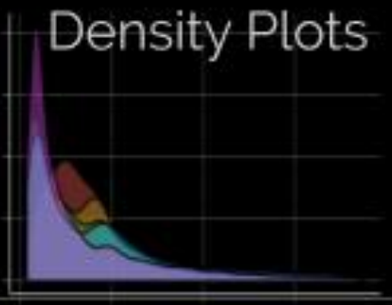
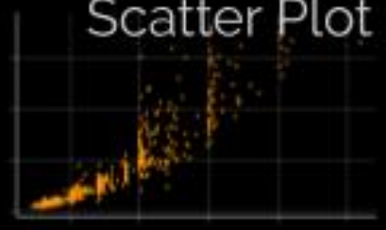
SO WHAT

Exercise – Basic Visualisation

Create simple static visualisations

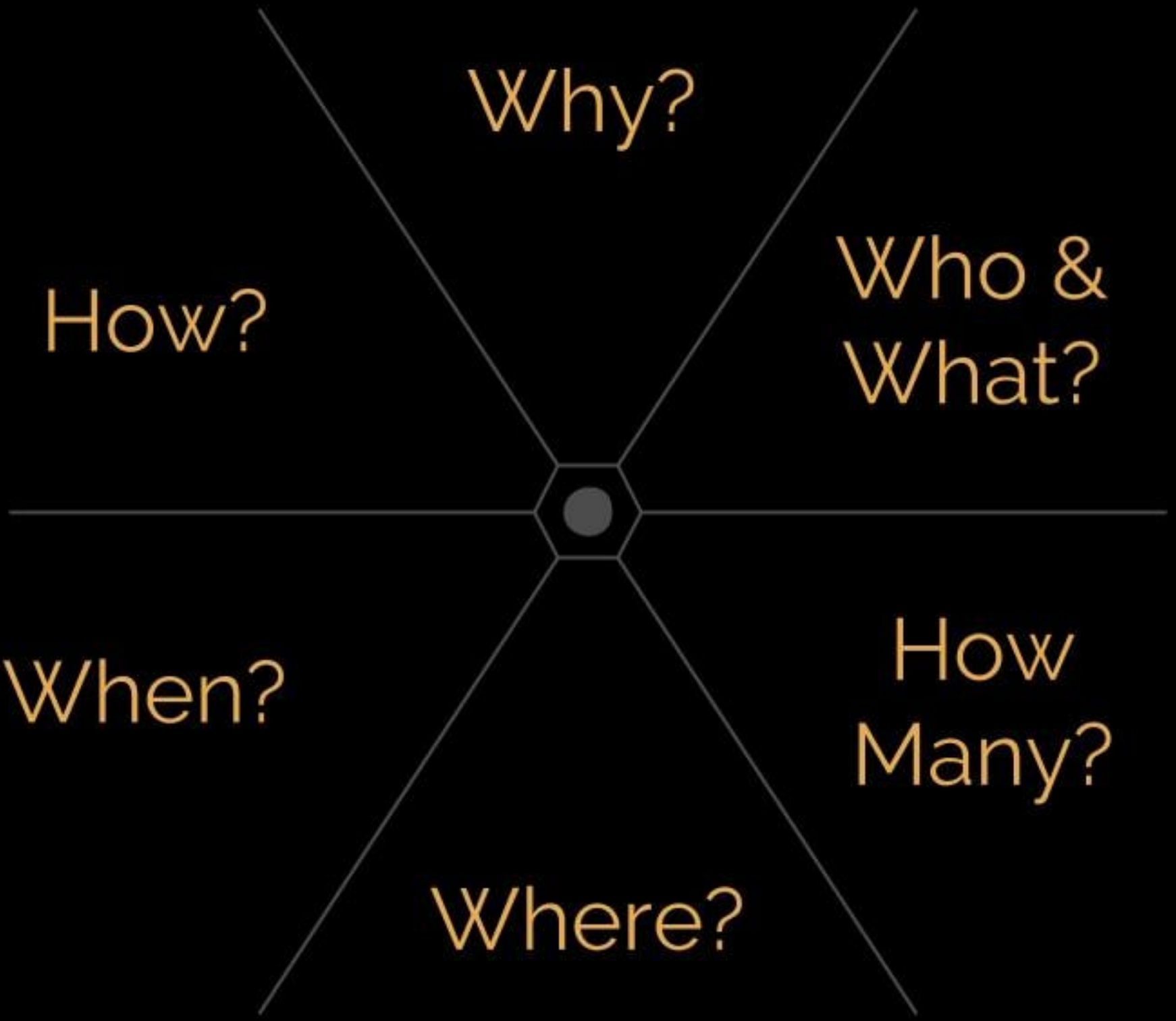
- 1D Continuous
- 1D Categorical
- 2D Continuous and Categorical
- 2D Continuous & Continuous
- 2D Categorical & Continuous

Chart Options	Points	Bars	Lines	Areas
1d Quantitative				
1d Categorical				
2d Quantitative + Categorical				
2d Categorical + Categorical				
2d Quantitative + Quantitative				

Chart Options	Points	Bars	Lines	Areas
1d Quantitative	 Strip Plot	 Histogram	 Freq Poly	 Density Plots
1d Categorical	 Dot Plot	 Bar Chart	Avoid	Avoid
2d Quantitative + Categorical	 Strip Plot	 Box Plot	 Freq Poly	 Density Plots
2d Categorical + Categorical	Avoid	 Bar Chart	Avoid	 Mosaic Plot
2d Quantitative + Quantitative	 Scatter Plot	 Table Lens	 Slopegraph	Avoid

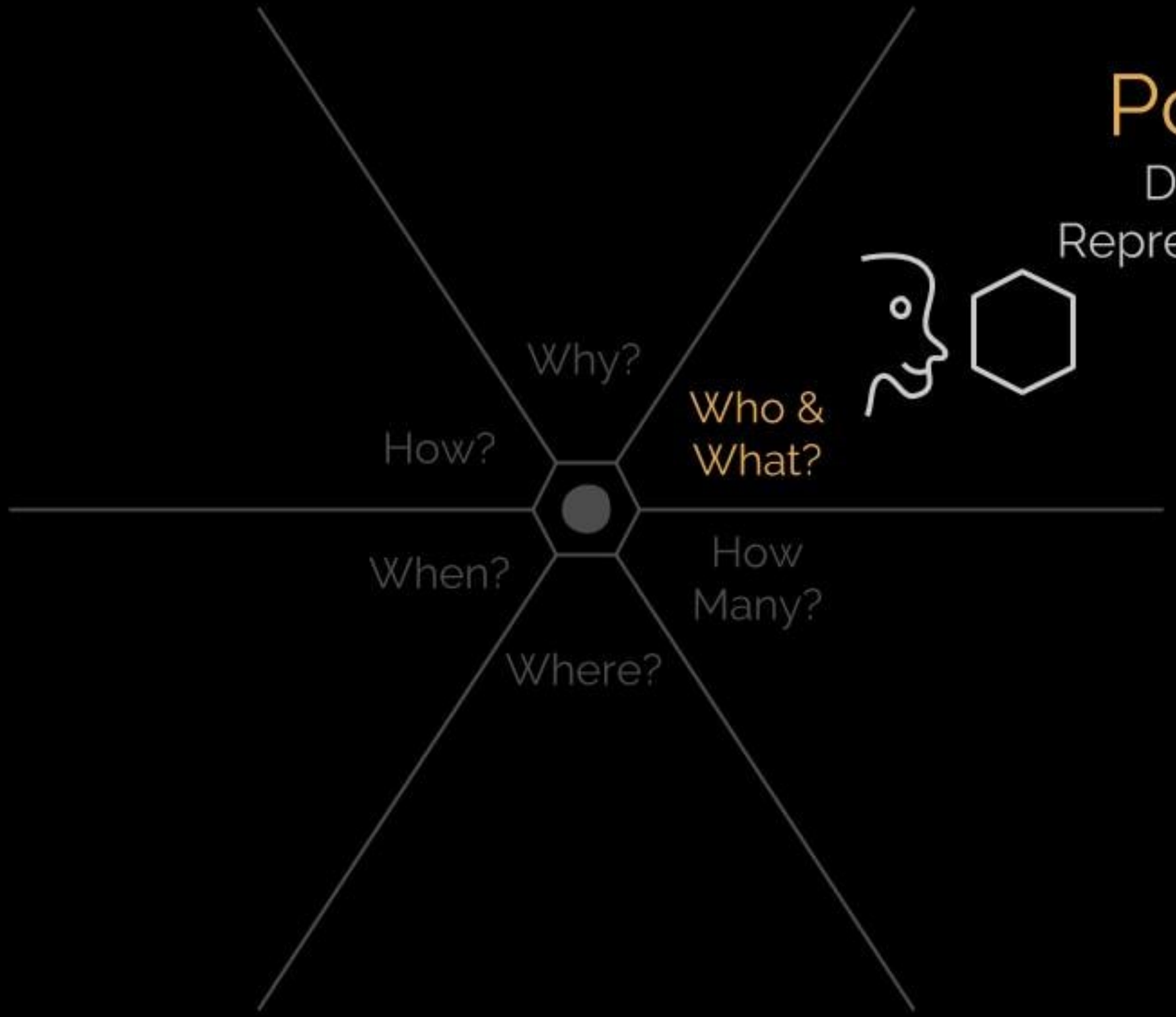
Visualisation Guides

Visualization gives you **answers to questions you didn't know you had.**
— Ben Schneiderman

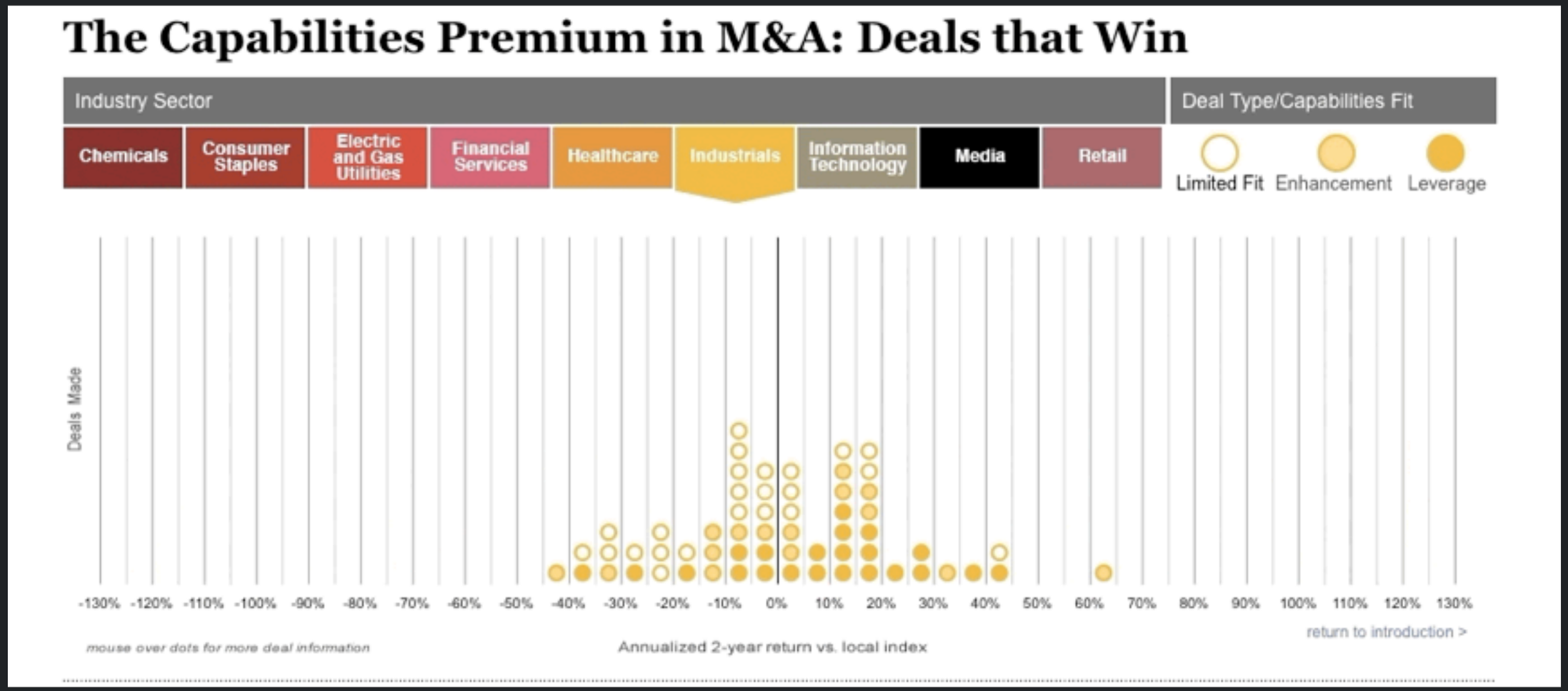

Why?
How?
Who &
What?
When?
How
Many?
Where?

Portrait

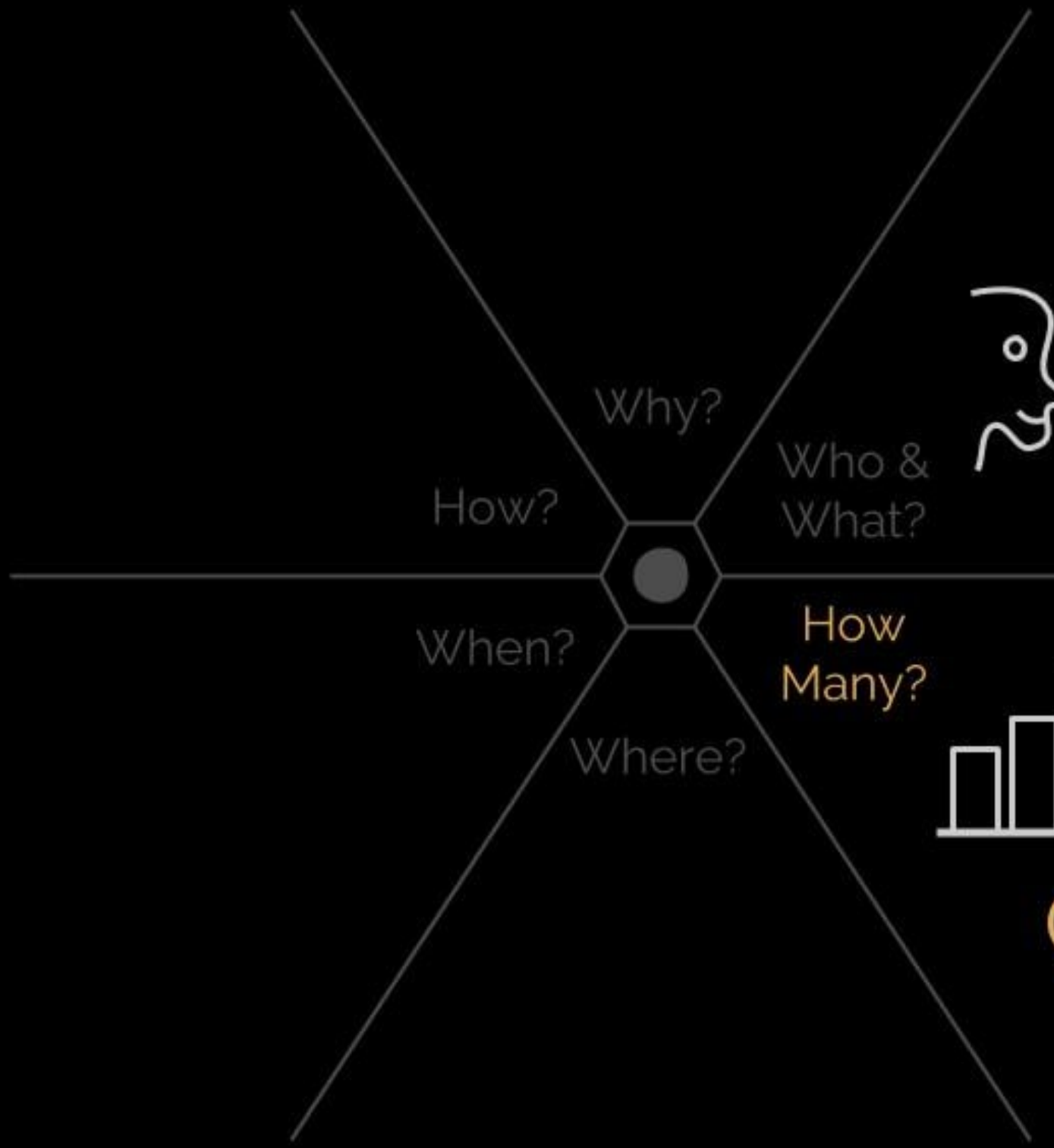
Distribution
Representation



Who & What: Distribution[1]



[1] Capabilities Premium



Portrait

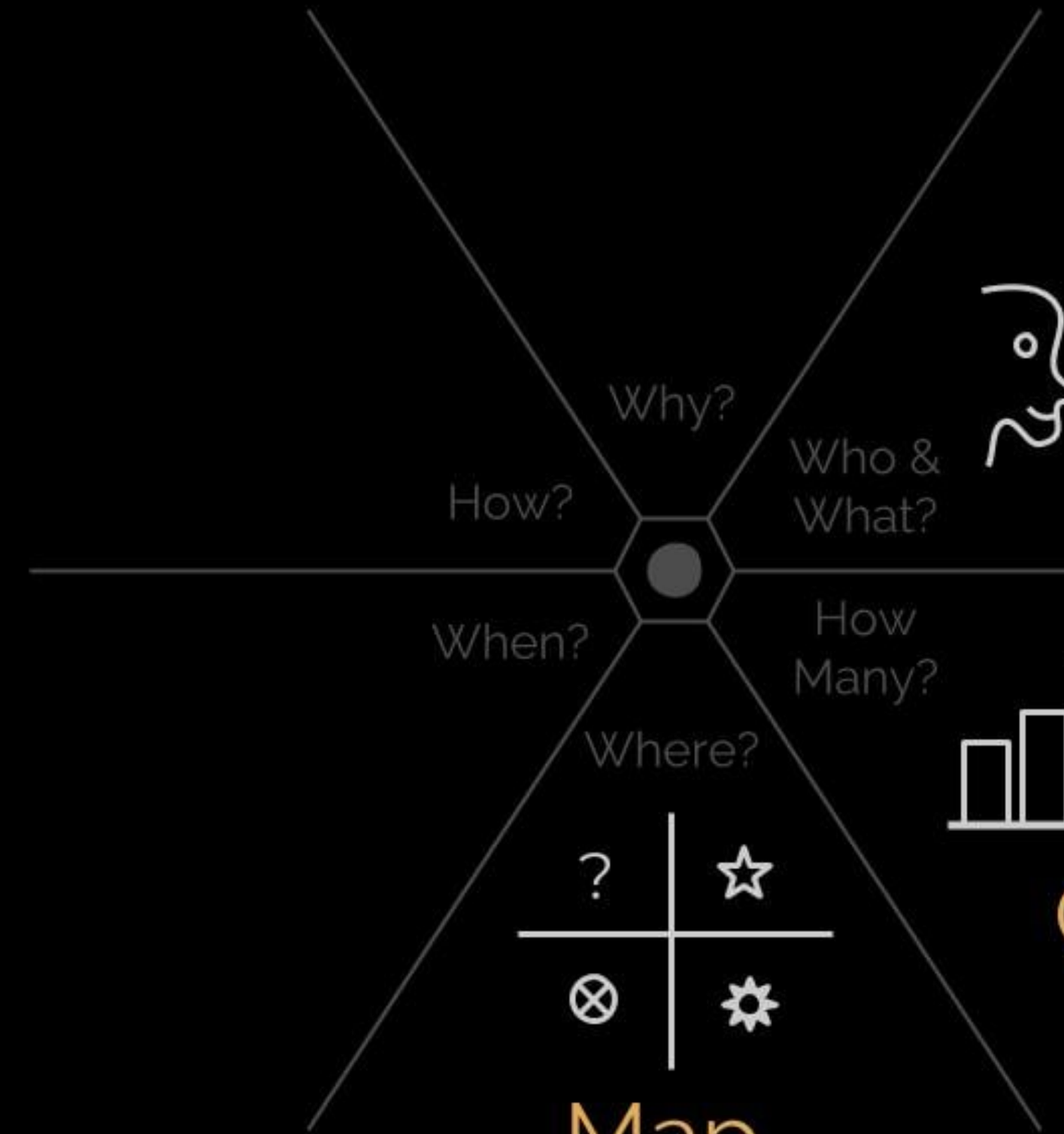
Distribution
Representation

Comparison

Comparative
Representation

How Many: Comparison[2]

[2] Working Capital Profiler



Portrait

Distribution
Representation

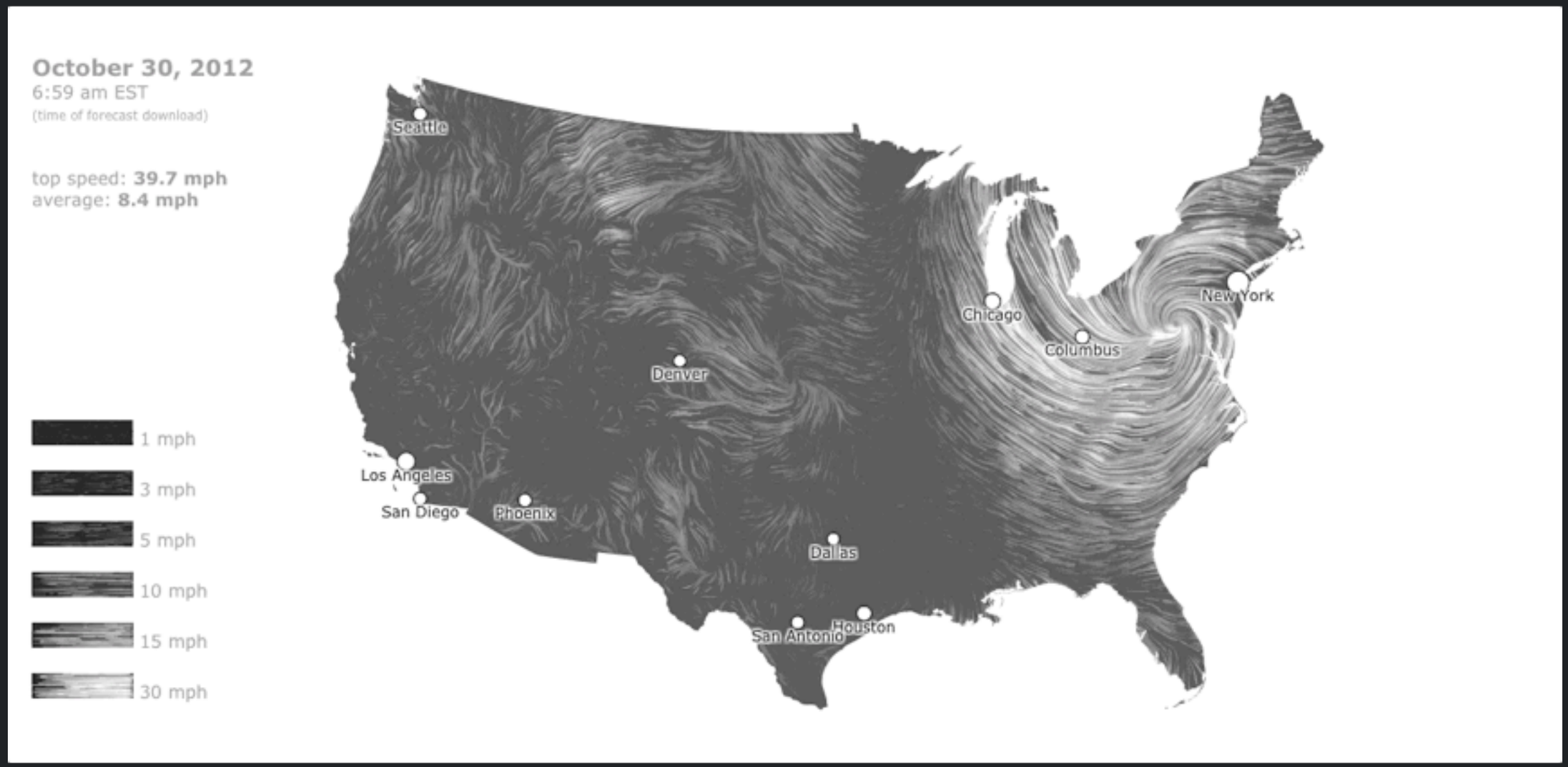
Comparison

Comparative
Representation

Map

Position in Space

Where: Maps[3]



[3] Wind Map

Why?

How?

Who &
What?

Portrait

Distribution
Representation

When?

How
Many?

Where?

Timeline

Position in
Time

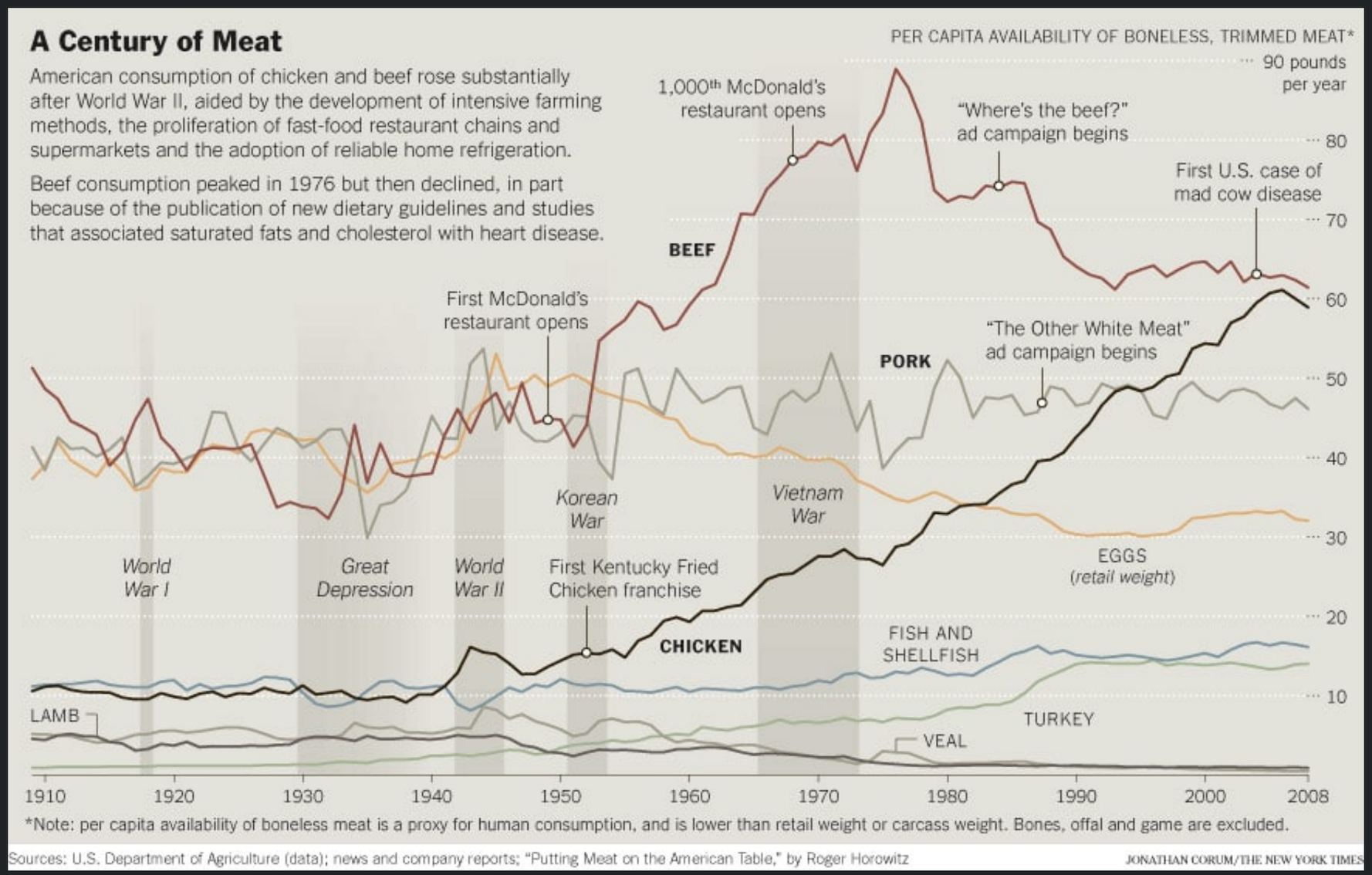
Comparison

Comparative
Representation

Map

Position in Space

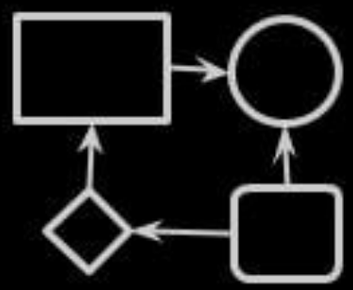
When: Timeline[4]



[4] New York Times

Flowchart

Relationship,
Hierarchy



Portrait

Distribution
Representation

Why?

How?

Who &
What?

When?

How
Many?

Where?

Timeline

Position in
Time

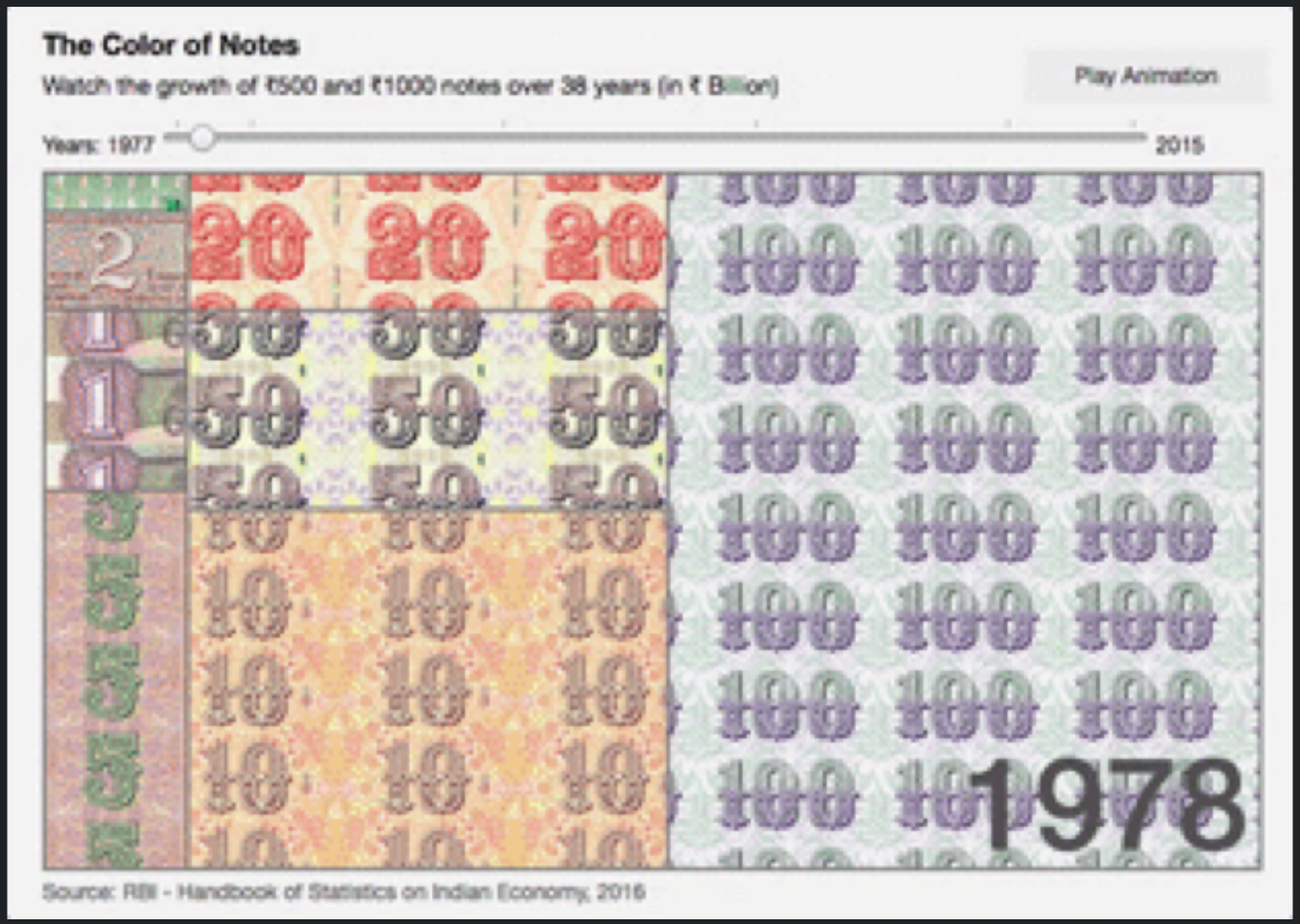
Comparison

Comparative
Representation

Map

Position in Space

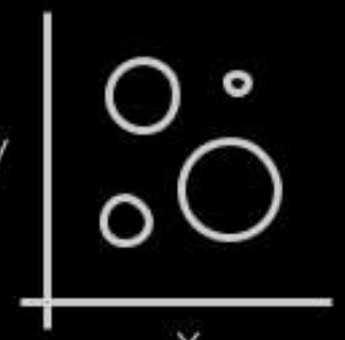
How: Relationship[5]



[5] Amit Kapoor

Multi Variable

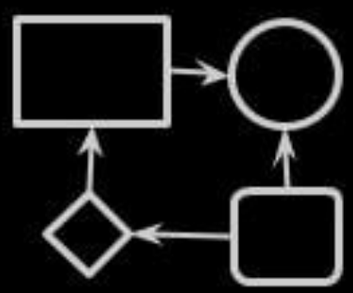
Deduction & Prediction



Why?

Flowchart

Relationship,
Hierarchy



How?

Portrait

Distribution
Representation

Who &
What?

When?

How
Many?

Where?

Timeline

Position in
Time

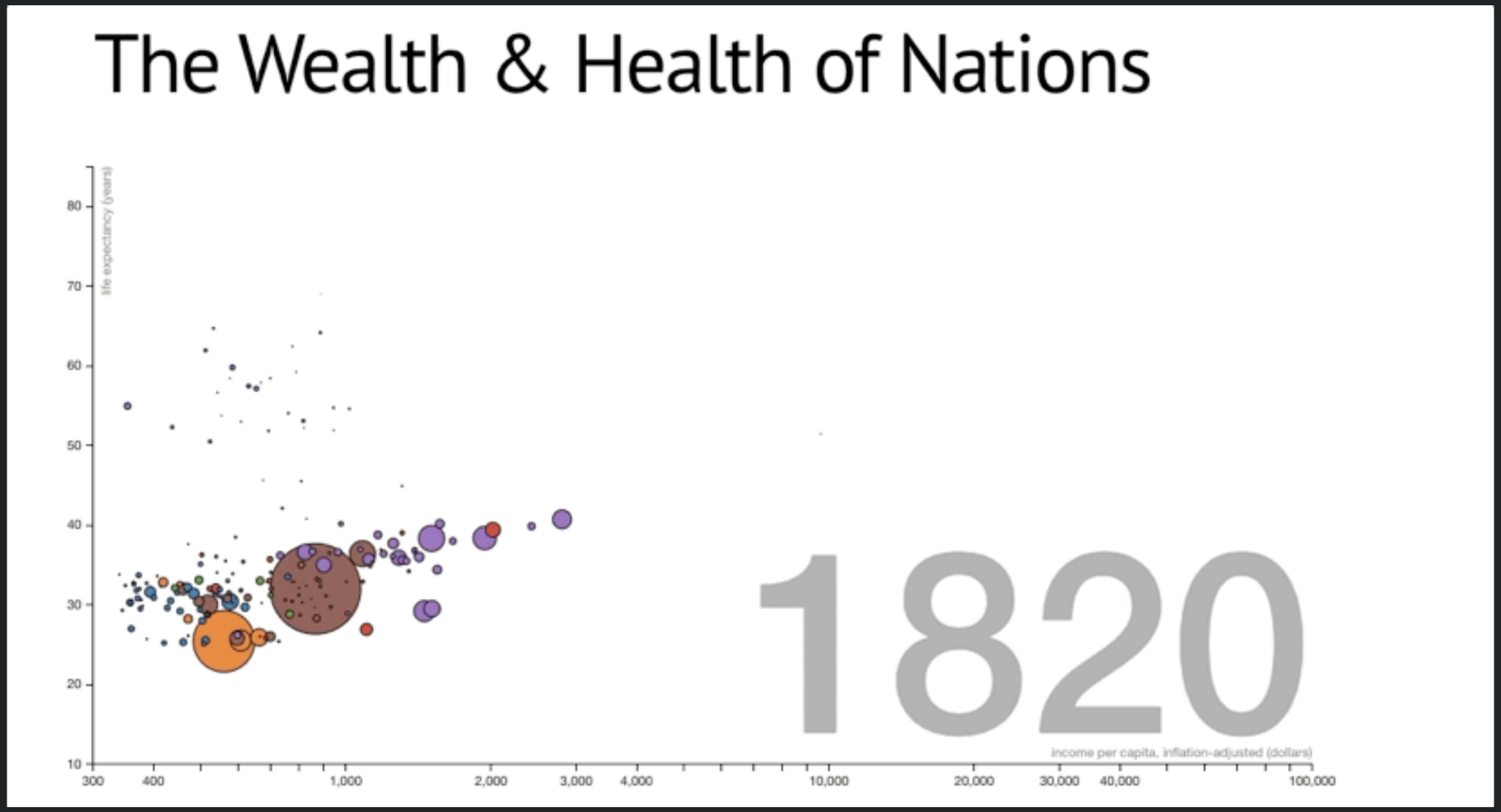
Comparison

Comparative
Representation

Map

Position in Space

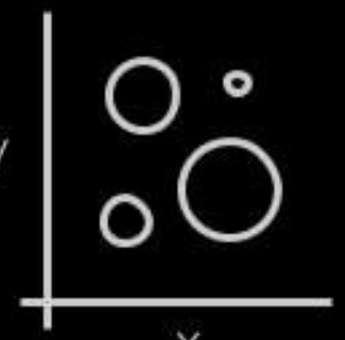
Why: Deduction & Prediction[6]



[6] Mike Bostock

Multi Variable

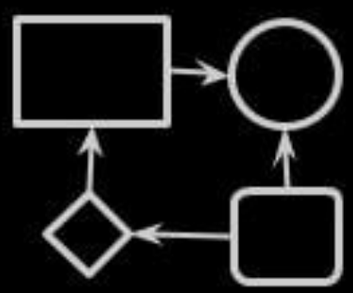
Deduction & Prediction



Why?

Flowchart

Relationship,
Hierarchy



How?

Portrait

Distribution
Representation

Who &
What?

When?

Timeline

Position in
Time

Where?

Map

Position in Space

How
Many?

Comparison

Comparative
Representation

Guide to choosing appropriate visualisation

FT Visual Vocabulary
ft.com/vocabulary

Vega-lite version
gramener.github.io/visual-vocabulary-vega/

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg; Paul McCallum; Martin Stabe
Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

Spine chart

Splits a single value into two contrasting components (eg male/female).

Surplus/deficit filled line

The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation & unemployment, income & life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Line + Column

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data, less good at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Lollipop chart

Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution

Histogram

The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Boxplot

Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Dot plot

A simple way of showing the change or range (min/max) of data across multiple categories.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Cumulative curve

A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

Columns work well for showing change over time - but usually best with only one series of data at a time.

Line + column

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Stock price

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Connected scatterplot

A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (daily, weekly, monthly) – at the expense of showing precision in quantity.

Priestley timeline

Great when date and duration are key elements of the story in the data.

Circle timeline

Good for showing discrete values of varying size across multiple categories (eg earthquakes by continent).

Vertical timeline

Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram

Another alternative to the circle timeline for showing series where there are big variations in the data.

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses
Commodity production, market capitalisation

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Bar

See above. Good when the data are not time series and labels have long category names.

Paired column

As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Paired bar

See above.

Proportional stacked bar

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or seeing fine differences between data is not so important.

Isotype (pictogram)

Excellent solution in some instances – use only with whole numbers (do not slice off an arm to represent a decimal).

Lollipop chart

Lollipop charts draw more attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Radar chart

A space-efficient way of showing value of multiple variables– but make sure they are organised in a way that makes sense to reader.

Parallel coordinates

An alternative to radar charts – again, the arrangement of the variables is important. Usually benefits from highlighting values.

Bullet chart

Good for showing a measurement against the context of a target or performance range.

Grouped symbol

An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Proportional stacked bar

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg total).

Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segments.

Voronoi

A way of turning points into areas – any point within each area is closer to the central point than any other centroid.

Arc

A hemicycle, often used for visualising political results in parliaments.

Gridplot

Good for showing % information, they work best when used on whole numbers and work well in multiple layout form.

Venn

Generally only used for schematic representation.

Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from locator maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rate/ratio)

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base geography.

Proportional symbol (count/magnitude)

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values

Equalised cartogram

Converting each unit on a map to a regular and equally-sized shape – good for representing voting regions with equal value.

Scaled cartogram (value)

Stretching and shrinking a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to annotate any patterns the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not snapped to an admin/political unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey

Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network

Used for showing the strength and inter-connectedness of relationships of varying types.

Exercise - Purpose:

Creating visualisation with purpose & objective

- Distribution
- Ranking
- Magnititude
- Correlation

#3: Theory of Data Visualisation

"We do data visualisations to **learn** something new, not just to **confirm**"

— Edward Tufte

Encoding Grammar

Deep dive into the grammar of graphics

- Data: wide tabular, long tabular
- Transform: bin, sort, filter, calculate, age
- Mark: symbol-type, length-type, area-type
- Channels: X, Y, size, colour, shape, text
- Scales: continuous, discrete, discretising
- Guides: axis, legends, labels
- Coordinates: cartesian, geographic

Encoding

See **Representation Slide Deck** for more concepts

Exercise – Encoding

Encode the telco churn data visualisation using different marks & channels
- Alternate mark representation
- Alternate channel choices

Decoding Principles

Understand how visual perception works
- Gestalt & visual perception
- Ranking of channel effectiveness & efficiency

Apply to select appropriate visualisations
- Error in decoding & empirical evidence

Aesthetics

See **Aesthetics Slide Deck** for more concepts

Exercise – Decoding

Visualisation decoding critique and improving an existing visualisation
- Alternate transformation & representations
- Alternate scale representations

#4: Guidelines for Better Data Visualisation

Resources

See **Aesthetics Slide Deck** for more concepts

Concept – Enhancing Visualisation

Guidelines for enhancing static data visualisation

- Comparing & sorting
- Add encoding variables
- Optimal scales & reference lines
- Layering & facets
- Over-plotting reduction

Exercise – Enhance

Enhancing an existing static visualisation in python

— Adding facets & layers exercise
— Handling over-plotting reduction

Concept – Color & Guides

Effective use of color & guides in data visualisations
- Chart junk & data-ink ratio
- Guidelines on axes and legends
- Color for categorical & continuous encoding
- Aesthetics & theming to have consistent style

Exercise – Redesign

Redesign of an existing visualisation to improve guides, color & aesthetics

Narrative Group Exercise

Data-Story Exercise

Building narrative with purpose & audience context

- To prepare a three-slide data-visual-story to be presented tomorrow morning.
- Presentation timing is 2 mins
- Groups to get started on the assignment

Day One Summary

— Recap of day one concepts and lessons
— Set the context for day two sessions
— Questions & Answers on Day 1 scope

Day 2 Sessions

— **Recap & Questions** (0930 - 0950)
— **Data-Story Presentations** (0950 - 1020)
— **Session #5: Crafting Visual Stories with Data** (1020 - 1120)
— Break (1120 - 1140)
— **Data-Story Rework** (1140 - 1210)
— **Session #6: Interactivity** (1210 - 1320)
— Lunch (1320 - 1420)
— **Session #7: Explorable Vis for Business Users** (1420 - 1540)
— Tea Break (1540 - 1600)
— **Session #8: Putting together an Interactive Application** (1600 - 1720)
— **Overall Summary & Way Forward** (1720 - 1800)