

05

## Task Abstraction

# Notice

- **Author**

- ◆ **João Moura Pires (jmp@fct.unl.pt)**

- **This material can be freely used for personal or academic purposes without any previous authorization from the author, provided that this notice is kept with.**

- **For commercial purposes the use of any part of this material requires the previous authorisation from the author.**

# Bibliography

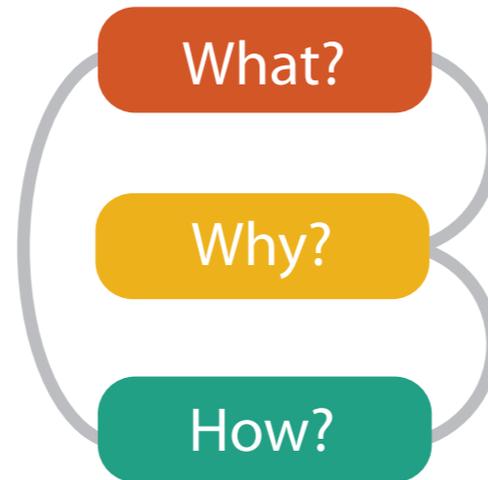
- Many examples are extracted and adapted from
  - ◆ Visualization Analysis & Design,  
Tamara Munzner, 2015

# Table of Contents

---

- **Introduction**
- **Why Analyze Tasks Abstractly?**
- **Who: Designer or User**
- **Actions**
- **Targets**
- **How: A Preview**

## Introduction

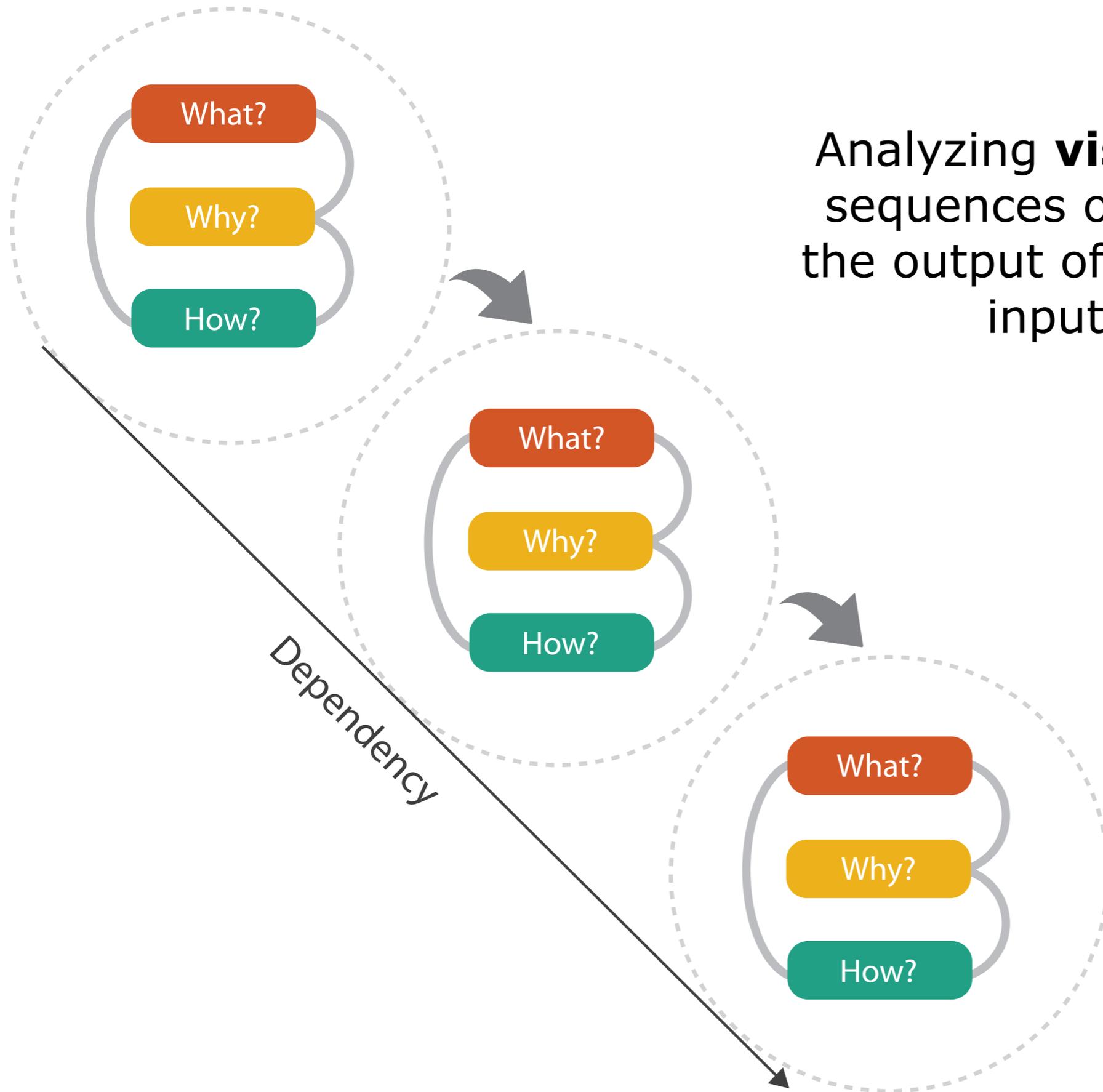


Three-part analysis **framework** for a vis instance:

**why** is the task being performed

**what** data is shown in the views,

and **how** is the vis idiom constructed in terms of design choices.



Analyzing **vis** usage as chained sequences of instances, where the output of one instance is the input to another.

# Why?

## Actions

## Targets

### → Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



### → Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

### → Query

→ Identify



→ Compare



→ Summarize



### → All Data

→ Trends



→ Outliers



→ Features



### → Attributes

→ One

→ Distribution



→ Extremes



→ Many

→ Dependency



→ Correlation



→ Similarity



### → Network Data

→ Topology

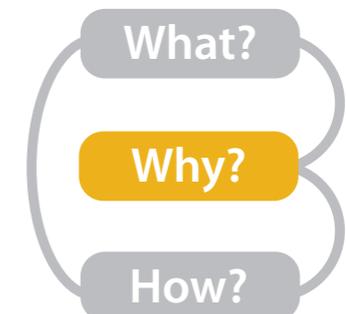


→ Paths



### → Spatial Data

→ Shape



## Why Analyze Tasks Abstractly?

# Why Analyze Tasks Abstractly?

- Consider tasks in **abstract form**, rather than the **domain-specific** way that users typically think about them
  - ◆ Transforming **task descriptions** from domain-specific language **into abstract form** allows you to reason about similarities and differences between them
  - ◆ **Example:**
    - Contrast the prognosis of patients who were intubated in the ICU more than one month after exposure to patients hospitalized within the first week
    - See if the results for the tissue samples treated with LL-37 match up with the ones without the peptide”.
    - **compare values between two groups**

# Why vocabulary

## ■ Why people is using a VIS

### Why?

#### 👉 Actions

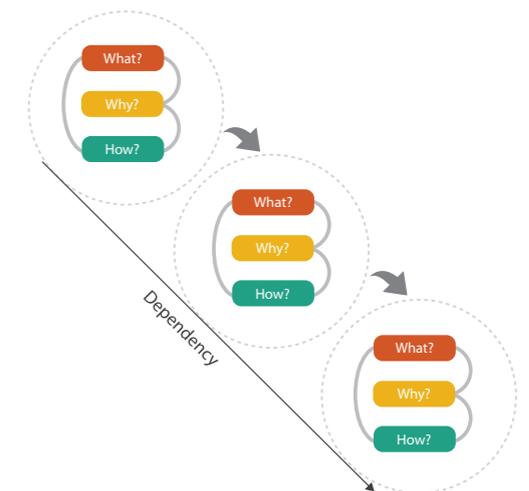
- Analyze
- Search
- Query

#### 🎯 Targets

- All Data
- Attributes
- Network Data
- Spatial Data

## ■ Consider **only one Goal at time**

- ◆ **Complex activities as a chained sequences of tasks**
  - Data abstraction



Who: Designer or User?

# Who: Designer or User?

## ■ Specific tool versus General tool

### ◆ Specific: tools are narrow.

- The designer has built many choices into the design of the tool itself in a way that the user cannot override
- They are limited in the kinds of data and tasks that they can address, but their strength is that users are not faced with an overwhelming array of design choices.

### ◆ General: tools are flexible

- The breadth of choices is both a strength and a limitation: users have a lot of power, but they also may make ineffective choices if they do not have a deep understanding of many vis design issues.

## Actions

➔ Analyze High-level

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search Mid-level

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ Query Low-level

➔ Identify



➔ Compare



➔ Summarize



The choices at each of these three levels are independent from each other, and it's usually useful to **describe actions at all three of them.**

# High-level: Analyze

## Actions

### → Analyze

→ Consume **existing information**

→ Discover



→ Present



→ Enjoy



→ Produce **new information**

→ Annotate



→ Record



→ Derive



Most common case

# High-level: Analyze - Consume

## Actions

### → Analyze

→ Consume **existing information**



→ Present



→ Enjoy



finding completely new things;

- the outcome is to generate a new hypothesis.
- to verify or disconfirm an existing hypothesis

sophisticated interactive idioms, because the vis designer doesn't know in advance what the user will need to see

# High-level: Analyze - Consume

## Actions

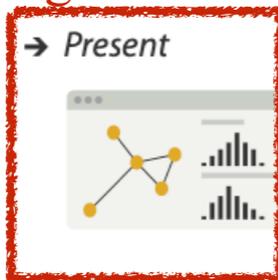
### → Analyze

→ Consume *existing information*

→ Discover



→ Present



→ Enjoy



succinct communication of information, for telling a story with data, or guiding an audience through a series of cognitive operations.

the present goal is not intrinsically limited to a static visual encoding idiom; it's very possible to pursue this goal with dynamic vis idioms that include interaction and animation

# High-level: Analyze - Consume

## Actions

### → Analyze

→ Consume **existing information**

→ Discover



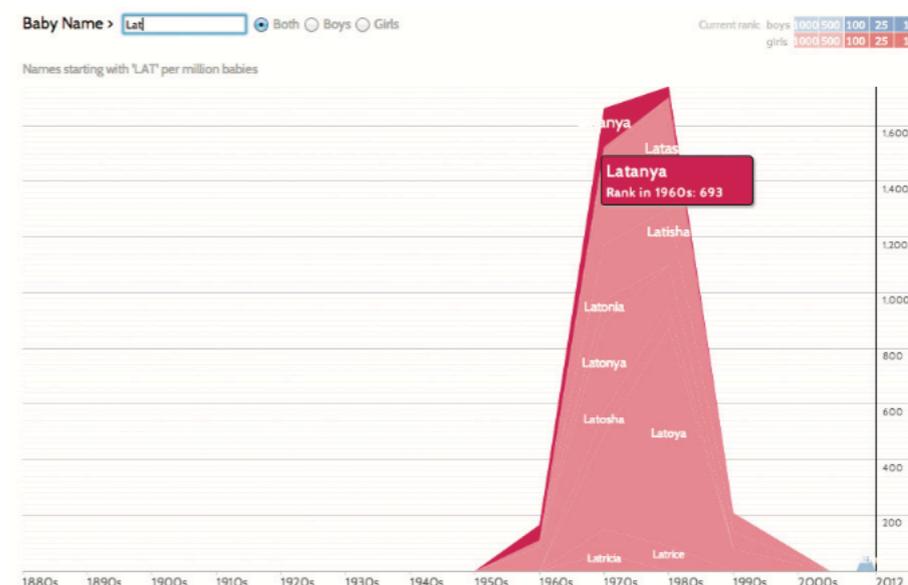
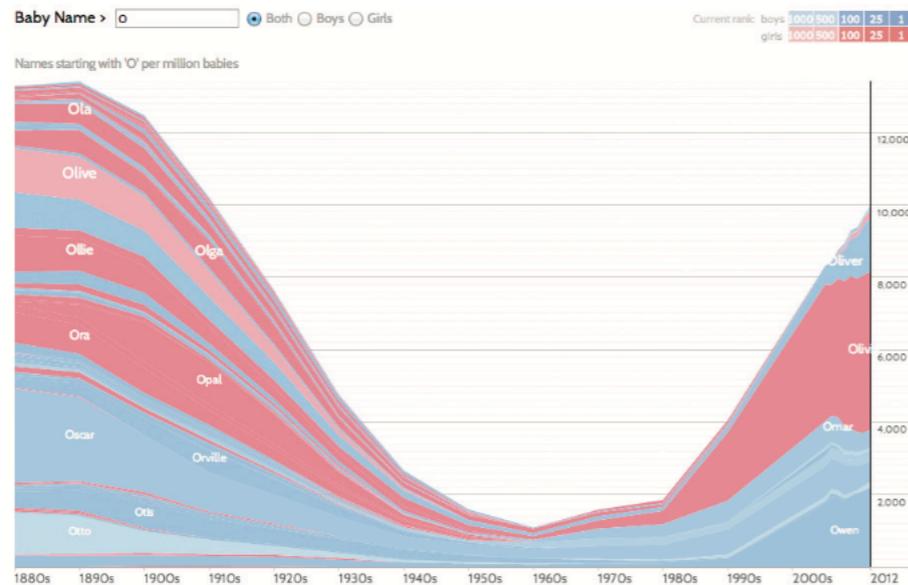
→ Present



→ Enjoy



Who will be the user?



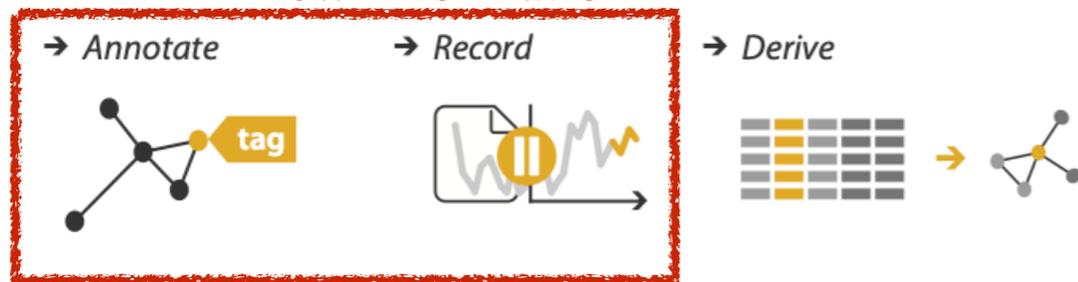
**Figure 3.3.** Name Voyager, a vis tool originally intended for parents focused deciding on what to name their expected baby, ended up being used by many nonparents to analyze historical trends for their own enjoyment. Left: Names starting with 'O' had a notable dip in popularity in the middle of the century. Right: Names starting with 'LAT' show a trend of the 1970s. After [Wattenberg 05, Figures 2 and 3], using <http://www.babynamewizard.com>.

# High-level: Analyze - Produce

## Actions

### → Analyze

→ Produce **new information**



Demo on Tableau (Annotations and Histories)

# Population and Birth Rate (2012)

**Birth Rate Bin**

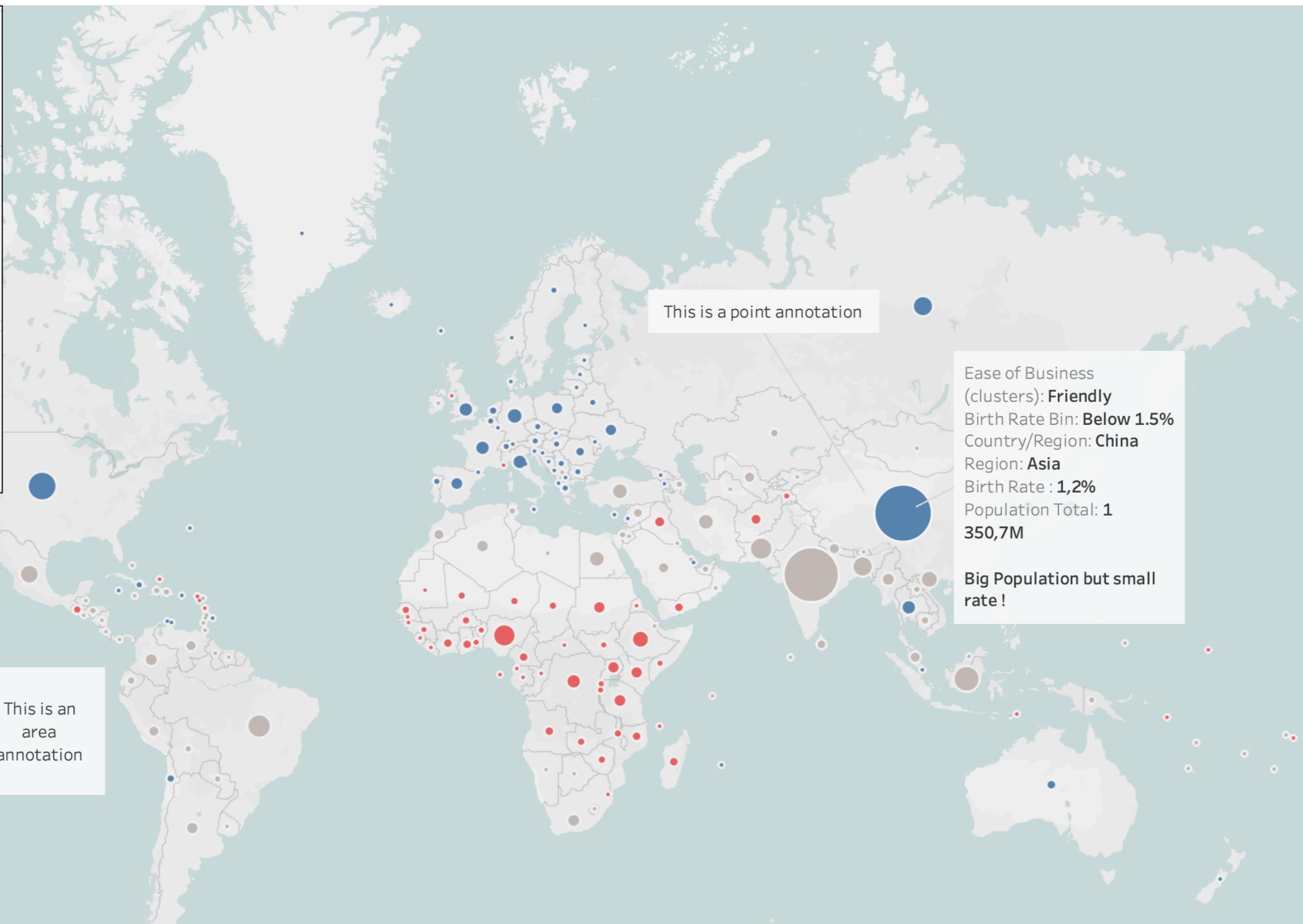
- Above 3%
- 1.5-3%
- Below 1.5%

**Caption**

Data Source: <http://data.worldbank.org/data-catalog/world-development-indicators>

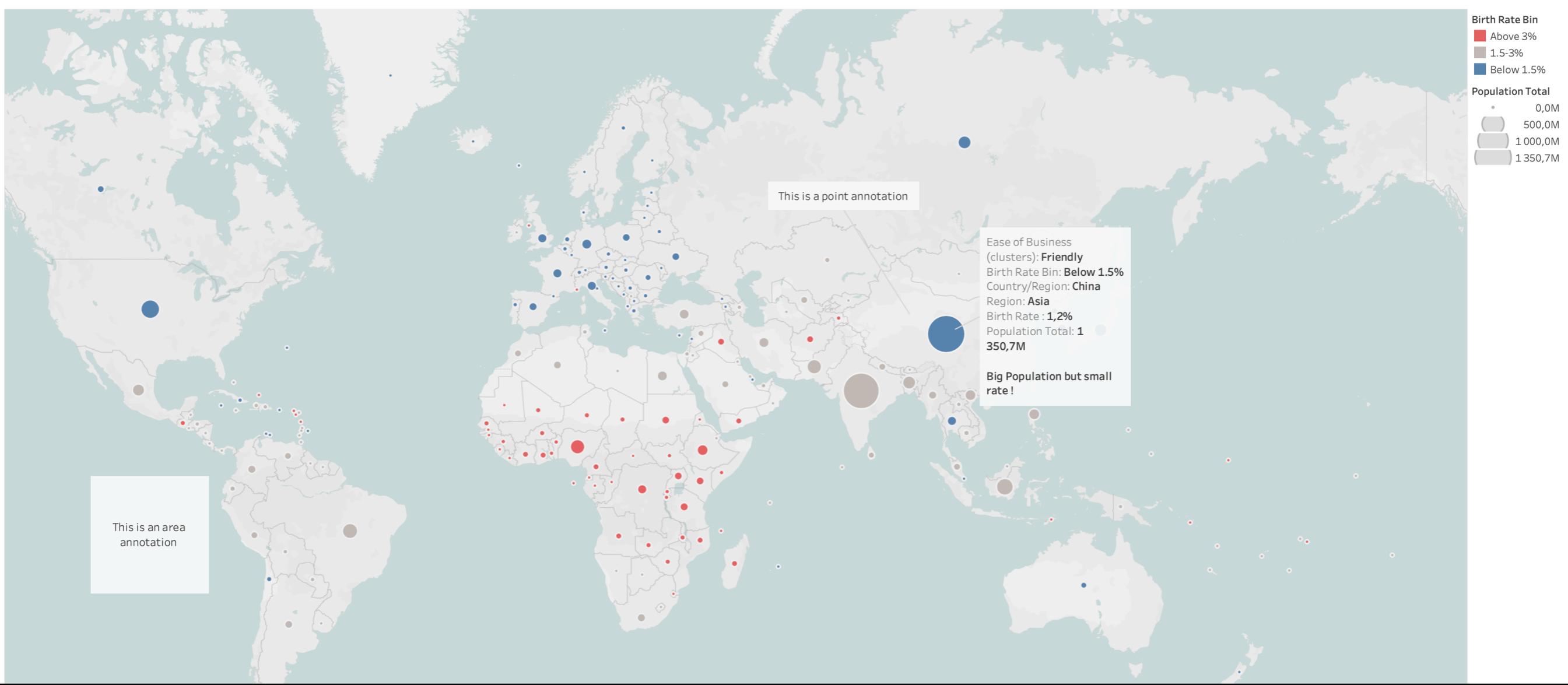
**AVG(Population Total)**

- 0,0M
- 500,0M
- 1 000,0M
- 1 350,7M



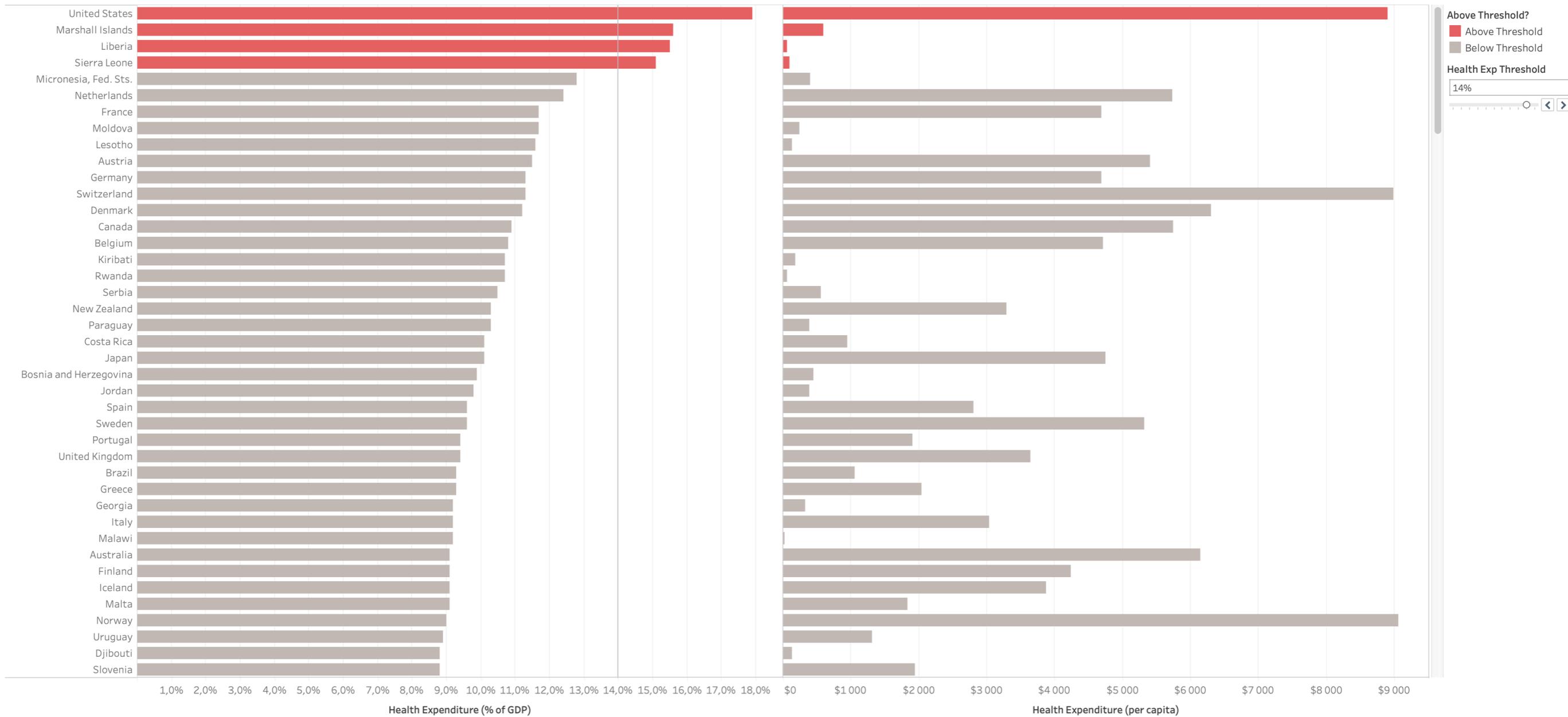
# Indicators of Global Growth and Development

< Global population and birth rate are imbalanced. A few countries spend more than 14% GDP on health care. Mobile phone usage grows much faster than internet usage. Following negative economic growth, we see interest rates decline. Europe continues to be the most popular tourist destination. High GDP correlates with ease of business. >



# Indicators of Global Growth and Development

<
A few countries spend more than 14% GDP on health care.
>



# Indicators of Global Growth and Development

Global population and birth rate are imbalanced.

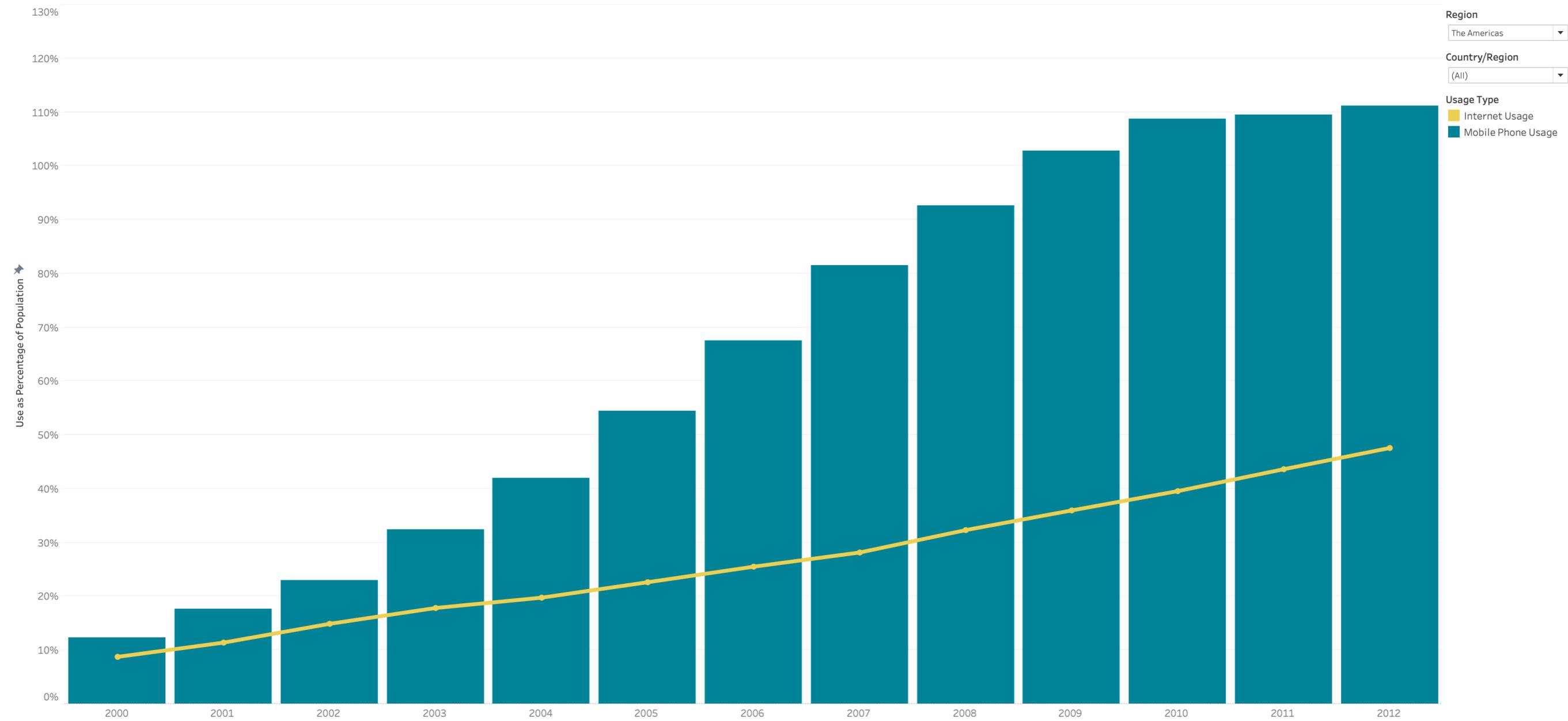
A few countries spend more than 14% GDP on health care.

**Mobile phone usage grows much faster than internet usage.**

Following negative economic growth, we see interest rates decline.

Europe continues to be the most popular tourist destination.

High GDP correlates with ease of business.



Region  
The Americas

Country/Region  
(All)

Usage Type  
Internet Usage  
Mobile Phone Usage

# Indicators of Global Growth and Development

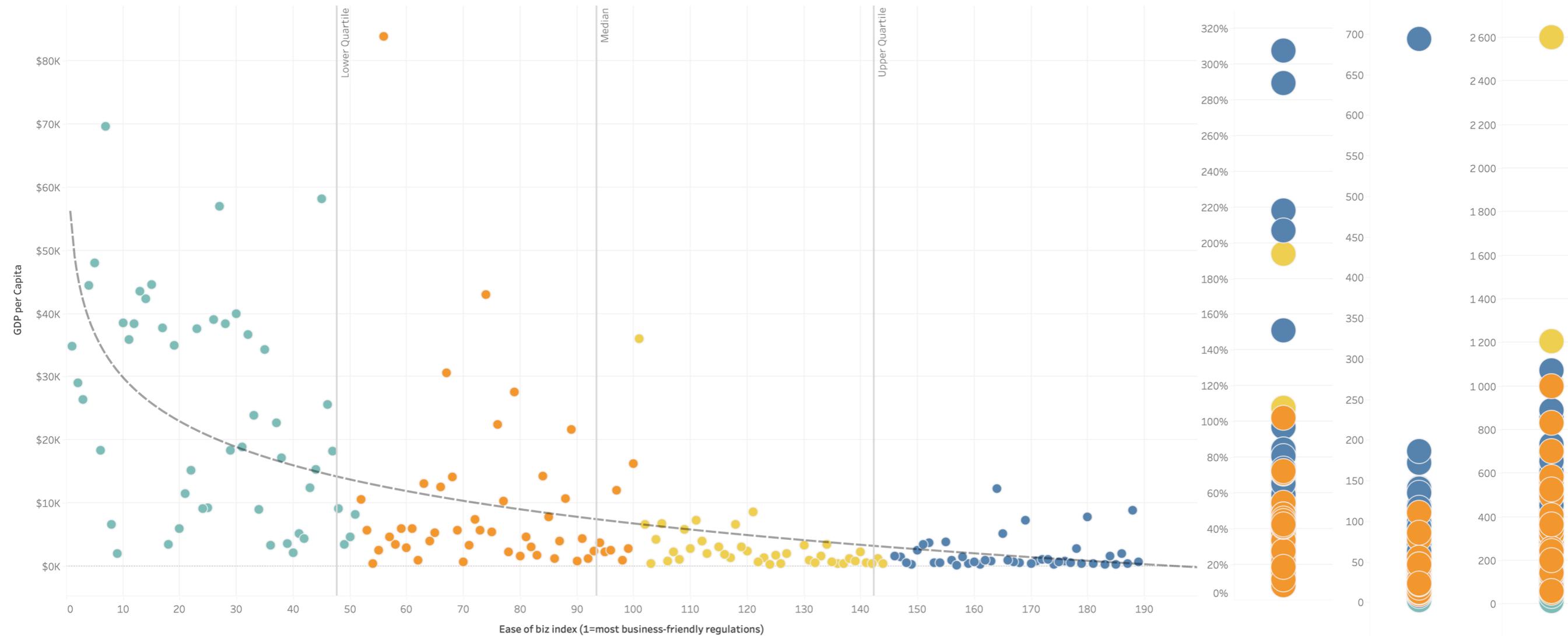
- Global population and birth rate are imbalanced.
- A few countries spend more than 14% GDP on health care.
- Mobile phone usage grows much faster than internet usage.
- Following negative economic growth, we see interest rates decline.
- Europe continues to be the most popular tourist destination.
- High GDP correlates with ease of business.

## Ease of Business

Region: 
 Highlight Country/Region: 
 Ease of Business:
 

- Very Friendly
- Friendly
- Moderate
- Low

## Correlation to GDP Per Capita



# High-level: Analyze - Produce

## 👉 Actions

### ➔ Analyze

➔ Produce **new information**

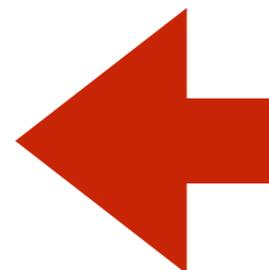
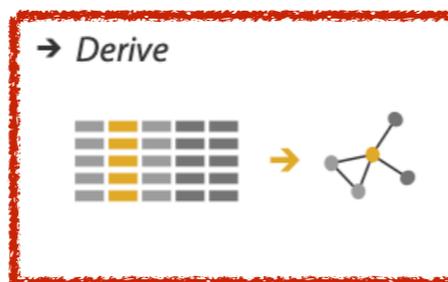
➔ Annotate



➔ Record



➔ Derive



A dataset often needs to be **transformed** beyond its original state in order to create a visual encoding that can solve the desired problem.

To do so, we can create **derived attributes** that extend the dataset beyond the original set of attributes that it contains.

## Derive new dataset types

# High-level: Analyze - Produce

## 👉 Actions

### ➔ Analyze

➔ Produce **new information**

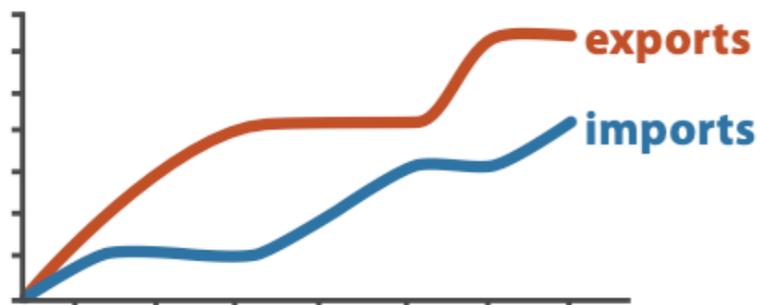
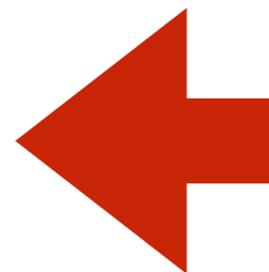
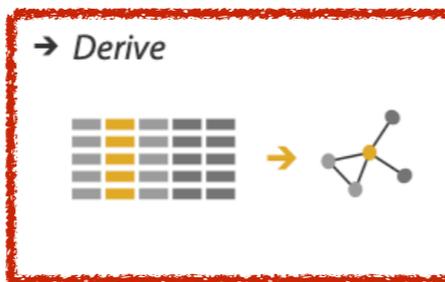
➔ Annotate



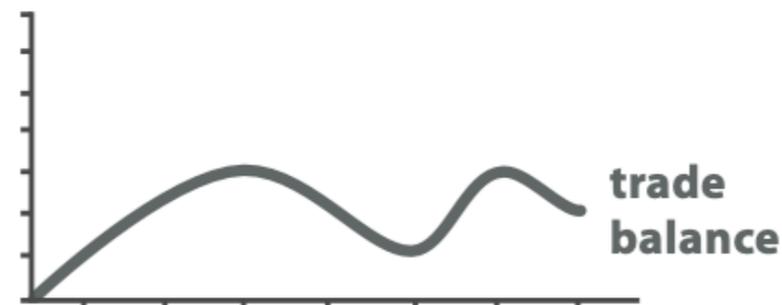
➔ Record



➔ Derive



Original Data



Derived Data

$$\text{trade balance} = \text{exports} - \text{imports}$$

➔ Analyze

High-level

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search

Mid-level

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

The choices at each of these three levels are independent from each other, and it's usually useful to **describe actions at all three of them.**

➔ Query

Low-level

➔ Identify



➔ Compare



➔ Summarize



# Mid-level: Search

## Actions

### Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

All of the high-level analyze cases require the user to **search** for elements of interest within the vis as a mid-level goal

# Mid-level: Search

## Actions

### → Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

- **Lookup** (target and location known)

- **Locate** (target known and **location unknown**)

- ◆ To find a known target at an unknown location, the search type is locate: that is, find out where the specific object is.

- **Browse** (target unknown and **location known**)

- ◆ users are searching for one or more items that fit some kind of specification, such as matching up with a particular range of attribute values.

# Mid-level: Search

## Actions

### → Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

## ■ Explore (target and location unknown)

- ◆ Searching for characteristics without regard to their location, often beginning from an overview of everything.
- ◆ Examples include searching for outliers in a scatterplot, for anomalous spikes or periodic patterns in a line graph of time-series data

➔ Analyze

High-level

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search

Mid-level

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ Query

Low-level

➔ Identify



➔ Compare



➔ Summarize



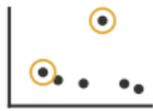
The choices at each of these three levels are independent from each other, and it's usually useful to **describe actions at all three of them.**

# Low-level: Query

## Actions

### → Query

#### → Identify



#### → Compare



#### → Summarize



Once a target or set of **targets** for a search has **been found**, a low-level user goal is to **query these targets** at one of three scopes: **identify**, **compare**, or **summarize**.

**Increase in the amount of search targets** under consideration: one, some, or all.

**identify** refers to a **single** target,

**compare** refers to **multiple** targets,

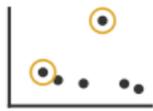
**summarize** refers to the **full set of possible targets**.

# Low-level: Query

## Actions

### → Query

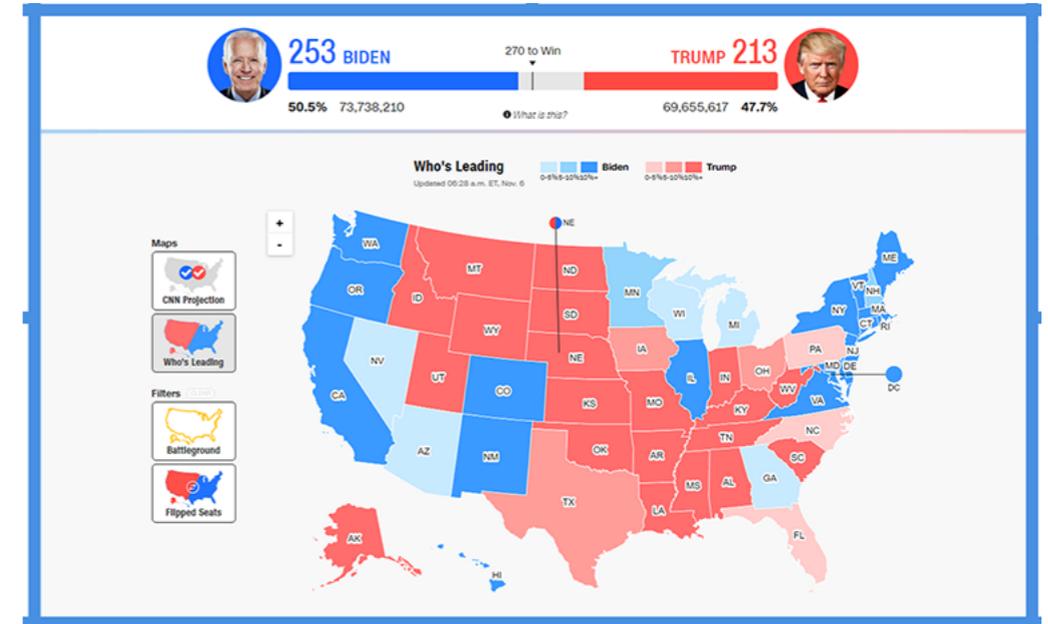
#### → Identify



#### → Compare



#### → Summarize



Different uses of a choropleth map of US election results, where each state is color-coded by the party that won.

- a user can **identify** the election results for one state,
- **compare** the election results of one state to another,
- **summarize** the election results across all states to determine how many favored one candidate or the other.

➔ Analyze

High-level

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search

Mid-level

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ Query

Low-level

➔ Identify



➔ Compare



➔ Summarize



The choices at each of these three levels are independent from each other, and it's usually useful to **describe actions at all three of them.**

➔ **Analyze**

High-level

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ **Search**

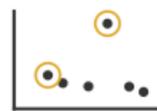
Mid-level

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

➔ **Query**

Low-level

➔ Identify



➔ Compare

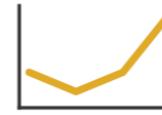


➔ Summarize

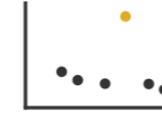


➔ **All Data**

➔ Trends



➔ Outliers



➔ Features



➔ **Attributes**

➔ One

➔ Distribution



➔ Extremes

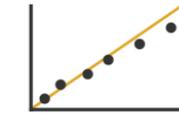


➔ Many

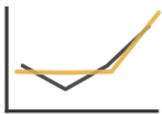
➔ Dependency



➔ Correlation



➔ Similarity



➔ **Network Data**

➔ Topology



➔ Paths



➔ **Spatial Data**

➔ Shape

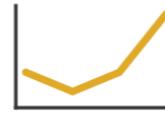


## Targets

➔ All Data

High-level

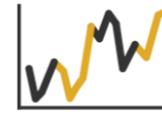
➔ Trends



➔ Outliers



➔ Features



➔ Attributes

➔ One

➔ Distribution



➔ Extremes

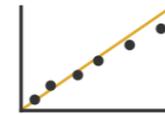


➔ Many

➔ Dependency



➔ Correlation

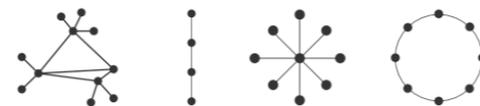


➔ Similarity



➔ Network Data

➔ Topology



➔ Paths



➔ Spatial Data

➔ Shape



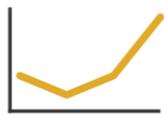
specific types of datasets

# Targets: All Data

## Targets

### → All Data

#### → Trends



#### → Outliers



#### → Features



A **trend** is a high-level characterization of a pattern in the data.

Ex: increases, decreases, peaks, troughs, and plateaus

Some data doesn't fit well with that backdrop; those elements are the **outliers**

The exact definition of **features** is **task dependent**, meaning any particular structures of interest.

# Targets: All Data

## Targets

### → Attributes

→ One

→ *Distribution*



→ *Extremes*

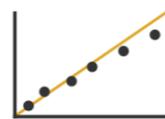


→ Many

→ *Dependency*



→ *Correlation*



→ *Similarity*



Attributes are specific properties that are visually encoded

➔ All Data

➔ Trends



➔ Outliers



➔ Features



➔ Attributes

➔ One

➔ Distribution



➔ Extremes



➔ Many

➔ Dependency



➔ Correlation

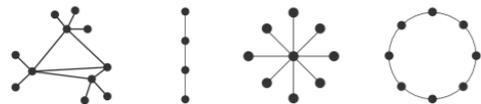


➔ Similarity



➔ Network Data

➔ Topology

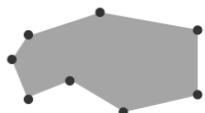


➔ Paths



➔ Spatial Data

➔ Shape



specific types of datasets

The abstract tasks of understanding **trends, outliers, distributions, and correlations** are extremely common reasons to use vis.

Each of them can be expressed in very diverse terms using domain-specific language.

## How to: a preview

# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

*Direction, Rate, Frequency, ...*



## Manipulate

### → Change



### → Select

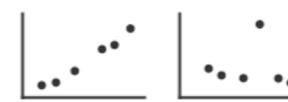


### → Navigate



## Facet

### → Juxtapose



### → Partition



### → Superimpose



## Reduce

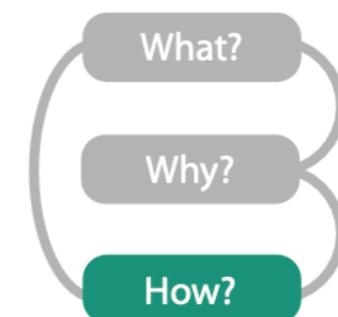
### → Filter



### → Aggregate



### → Embed



# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



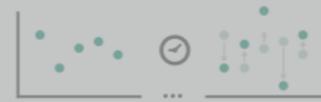
→ Motion

*Direction, Rate, Frequency, ...*



## Manipulate

### → Change



### → Select



### → Navigate



## Facet

### → Juxtapose



### → Partition



### → Superimpose



## Reduce

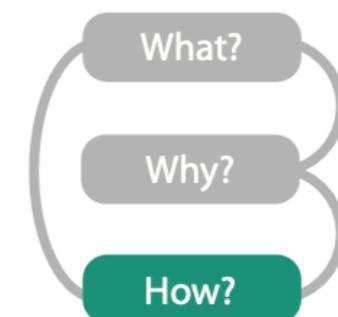
### → Filter



### → Aggregate



### → Embed

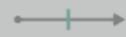


# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



→ Motion

*Direction, Rate, Frequency, ...*



## Manipulate

### → Change



### → Select



### → Navigate



## Facet

### → Juxtapose



### → Partition



### → Superimpose



## Reduce

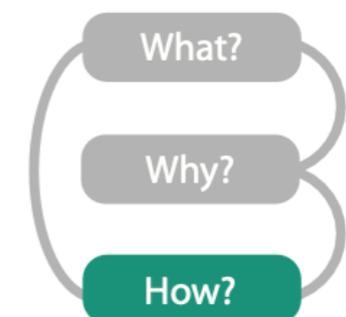
### → Filter



### → Aggregate



### → Embed

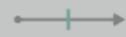


# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



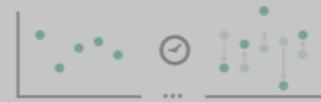
→ Motion

*Direction, Rate, Frequency, ...*



## Manipulate

### → Change



### → Select

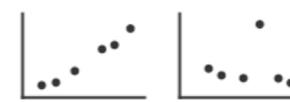


### → Navigate

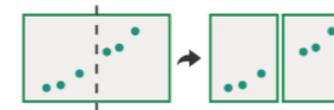


## Facet

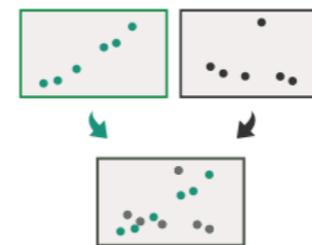
### → Juxtapose



### → Partition



### → Superimpose



## Reduce

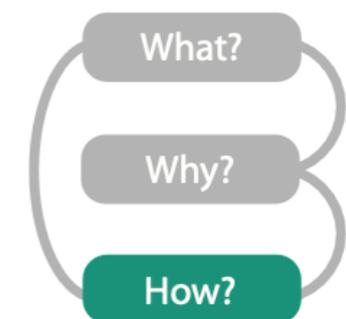
### → Filter

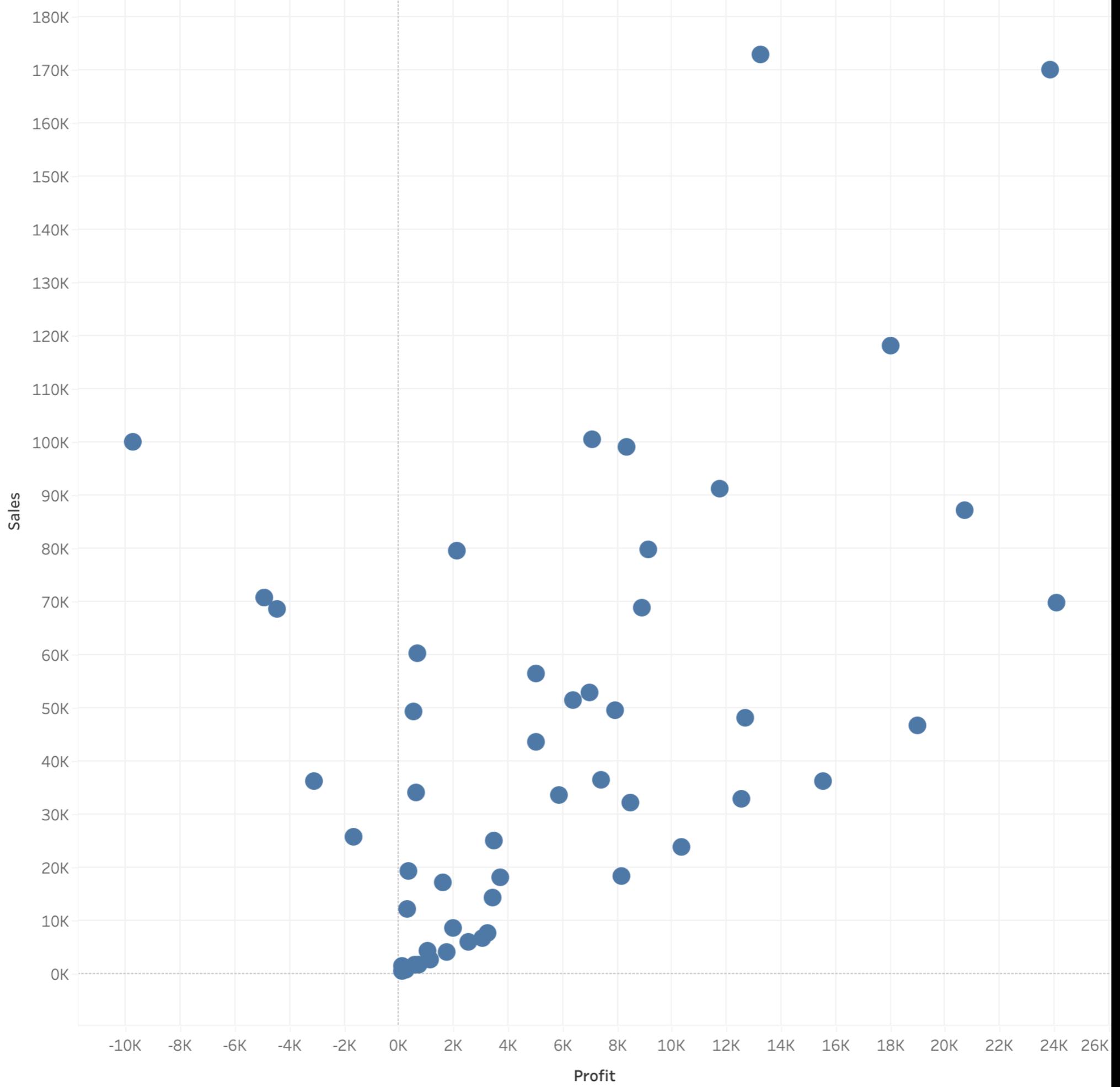


### → Aggregate

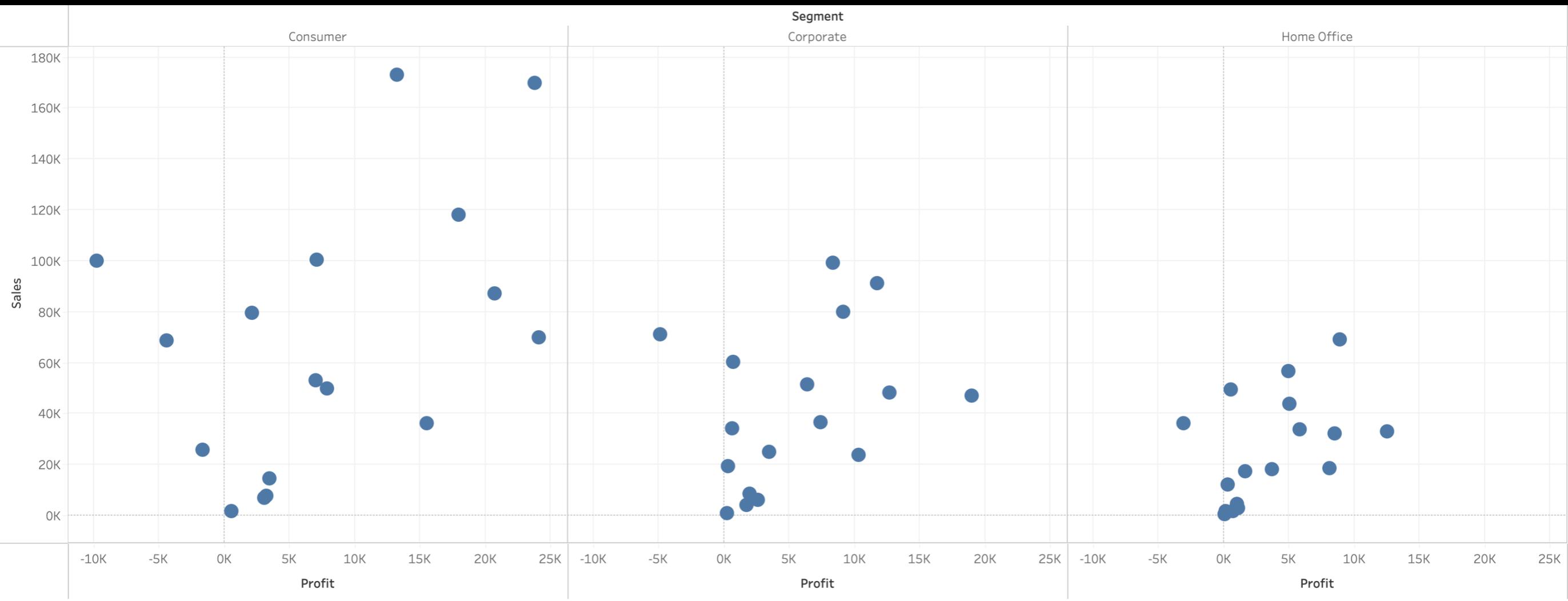


### → Embed

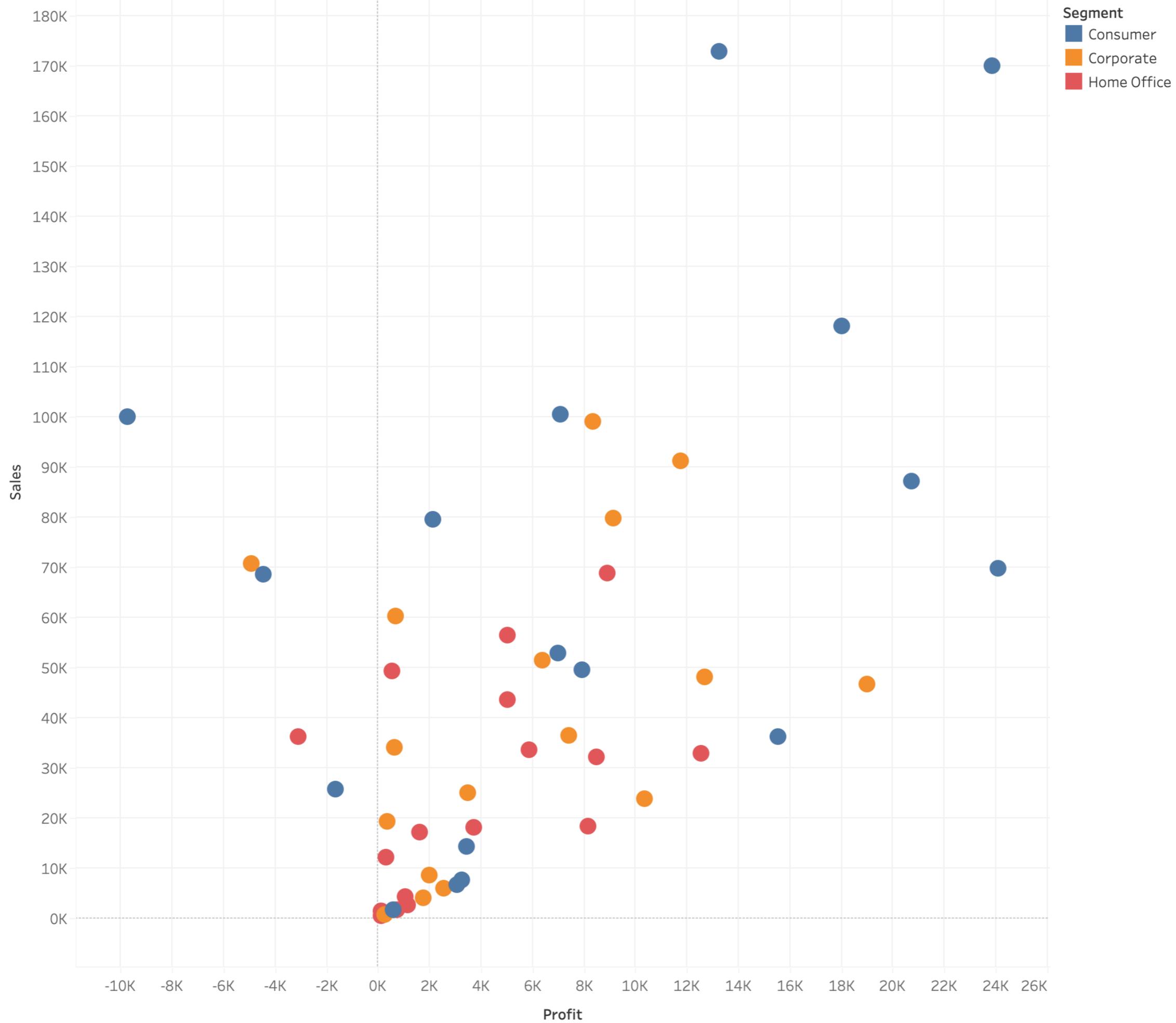




Sum of Profit vs. sum of Sales. Details are shown for Sub-Category and Segment.



Sum of Profit vs. sum of Sales broken down by Segment. Details are shown for Sub-Category.



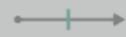
Sum of Profit vs. sum of Sales. Color shows details about Segment. Details are shown for Sub-Category.

# How?

## Encode

### → Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Saturation



→ Luminance



→ Size, Angle, Curvature, ...



→ Shape



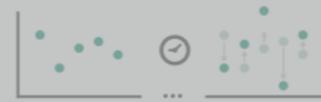
→ Motion

*Direction, Rate, Frequency, ...*



## Manipulate

### → Change



### → Select



### → Navigate



## Facet

### → Juxtapose



### → Partition



### → Superimpose



## Reduce

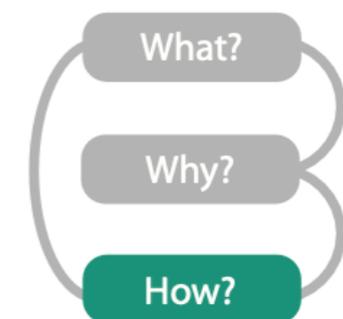
### → Filter



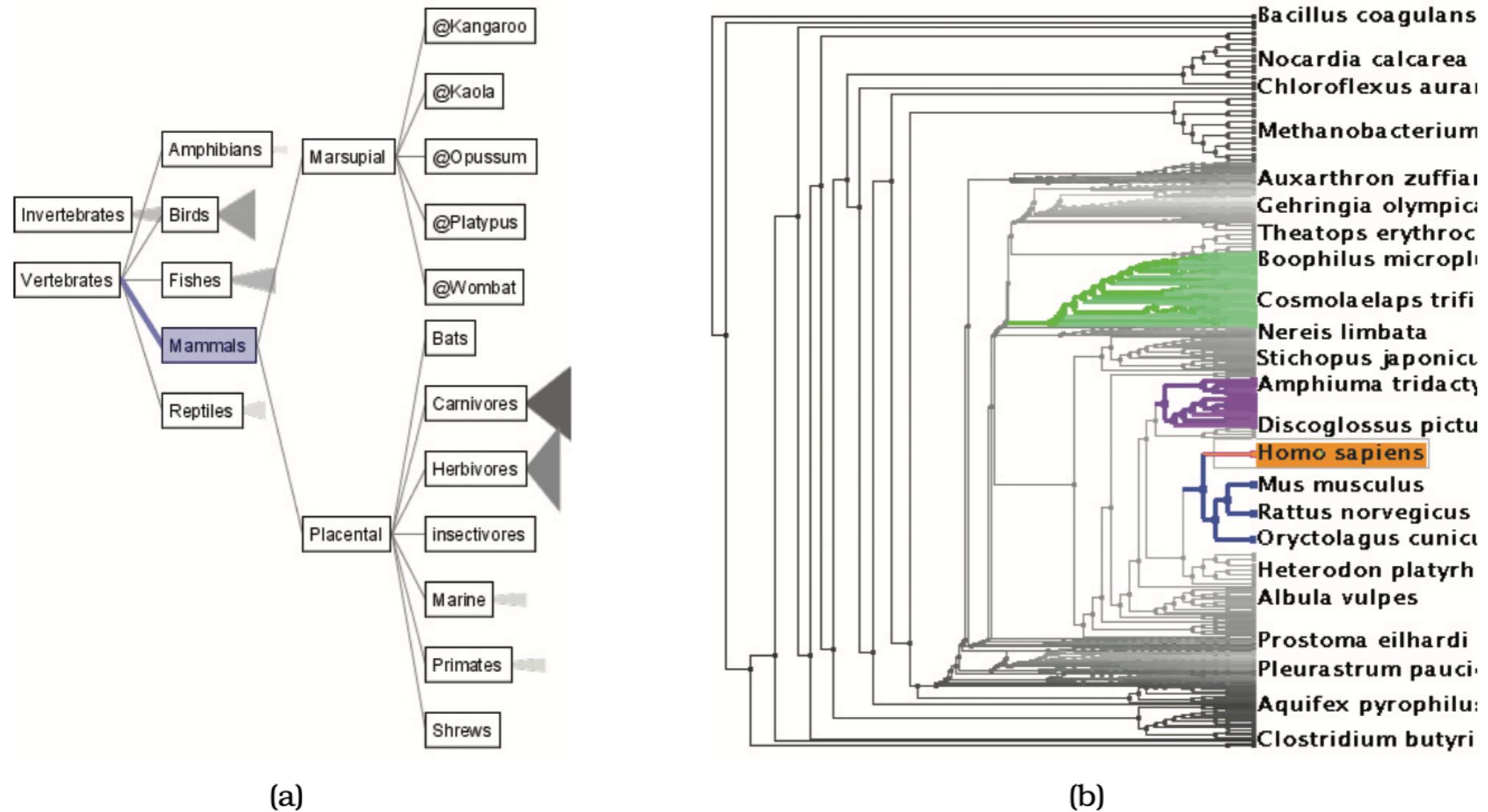
### → Aggregate



### → Embed



# Comparing Two Idioms



**Figure 3.8.** Comparing two idioms. (a) SpaceTree [Plaisant et al. 02]. (b) TreeJuxtaposer. From <http://www.cs.umd.edu/hcil/spacetree> and [Munzner et al. 03, Figure 1].

# Comparing Two Idioms

- **What:** Input data: a large tree composed of nodes and links.
- **Why:** these tools are being used is for the same goal
  - ◆ to **present** a path traced between two nodes of interest to a colleague.
  - ◆ In more detail, to **locate** paths between nodes and **identify** them.

## What?

### → Tree



## Why?

### → Actions

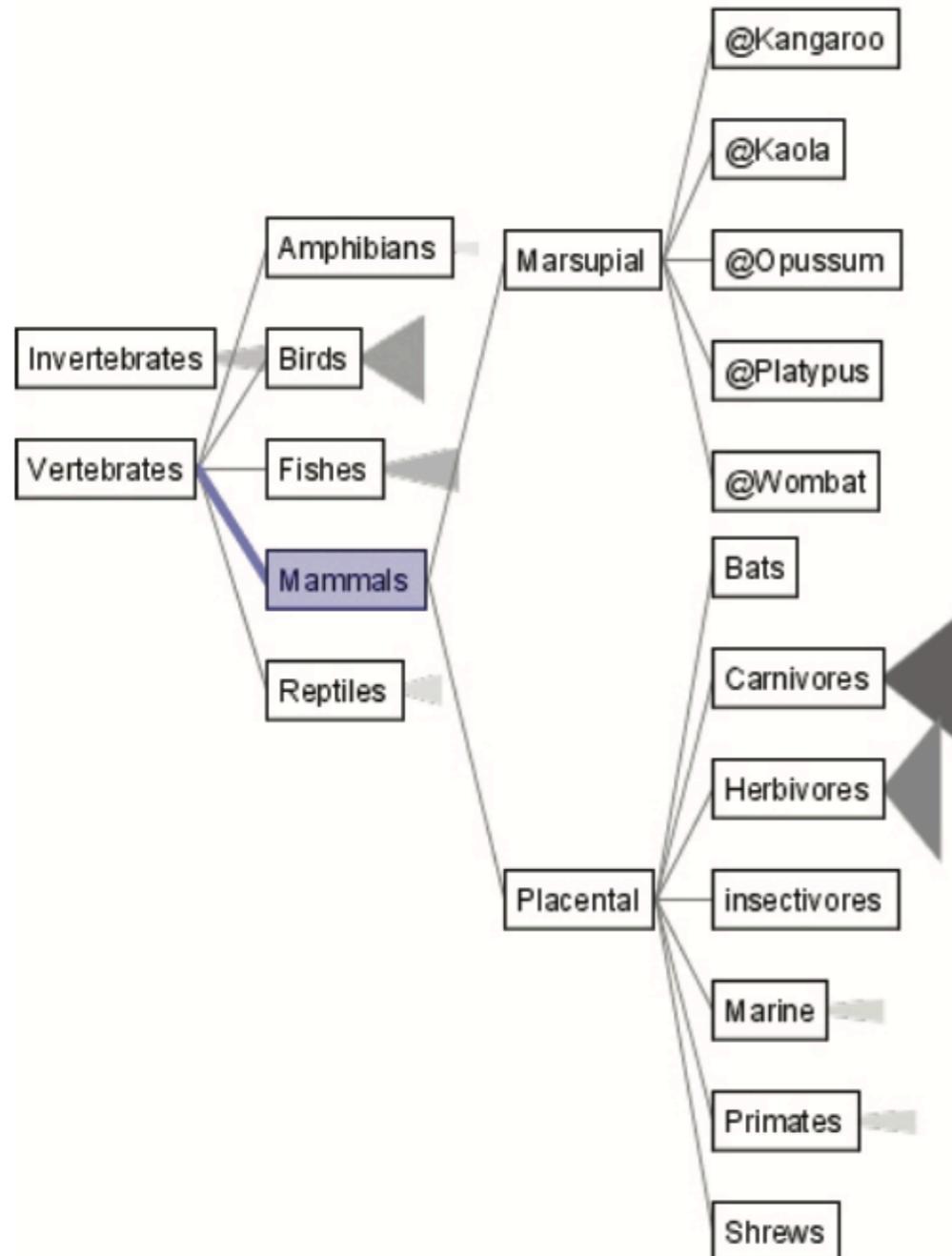
→ Present → Locate → Identify



### → Targets

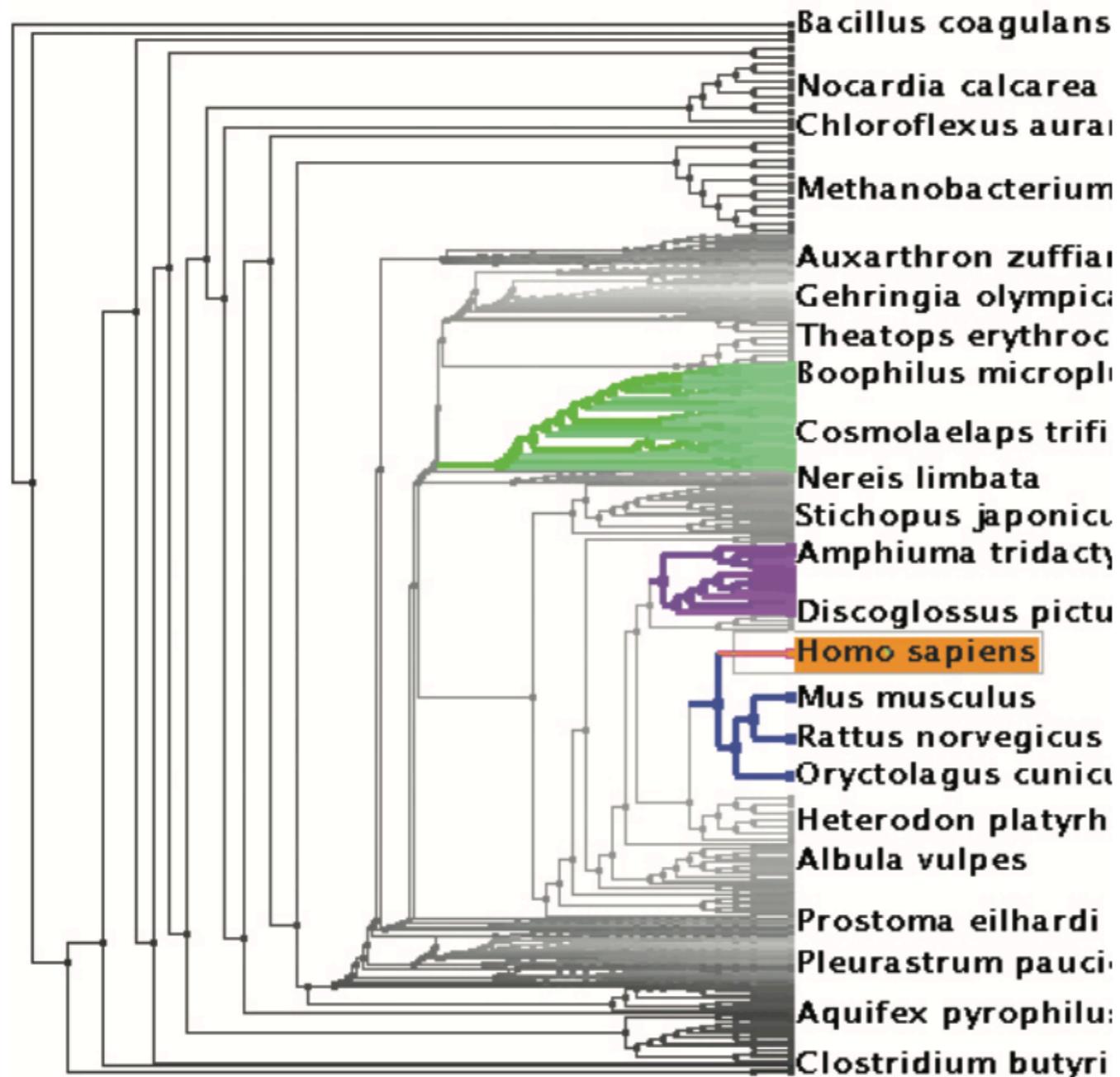
→ Path between two nodes





allow the user to **navigate** and to **select a path**

SpaceTree ties the act of selection to a change of what is shown by automatically aggregating and filtering the unselected items.



allow the user to  
**navigate** and to  
**select a path**

allows the user to arrange areas  
of the tree to ensure visibility  
for areas of interest.

# Comparing Two Idioms

- **What:** Input data: a large tree composed of nodes and links.
- **Why:** these tools are being used is for the same goal
  - ◆ to **present** a path traced between two nodes of interest to a colleague.
  - ◆ In more detail, to **locate** paths between nodes and **identify** them.

## What?

### → Tree



## Why?

### → Actions

→ Present → Locate → Identify



### → Targets

→ Path between two nodes



## How?

### → SpaceTree

→ Encode → Navigate → Select → Filter → Aggregate

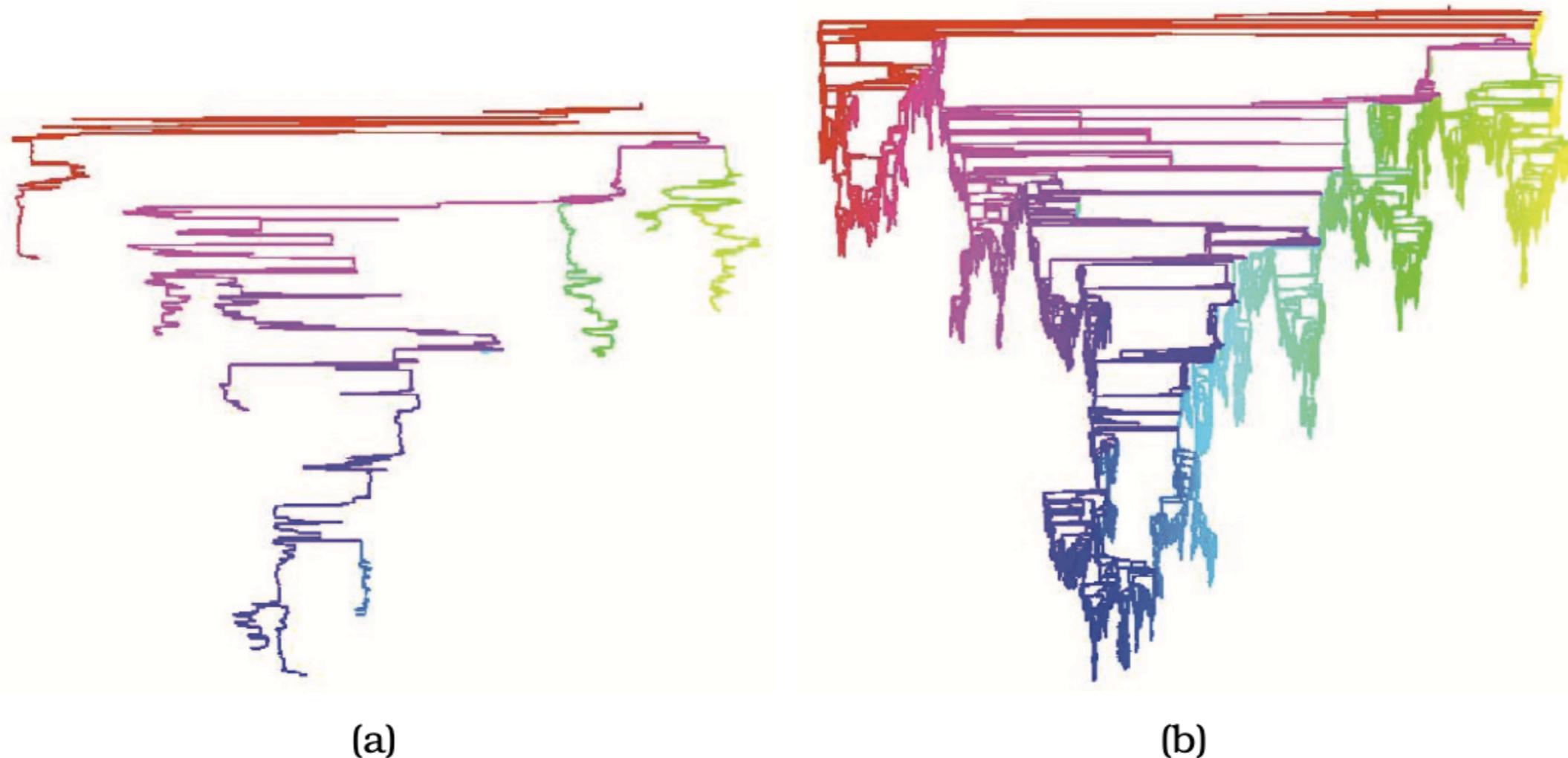


### → TreeJuxtaposer

→ Encode → Navigate → Select → Arrange

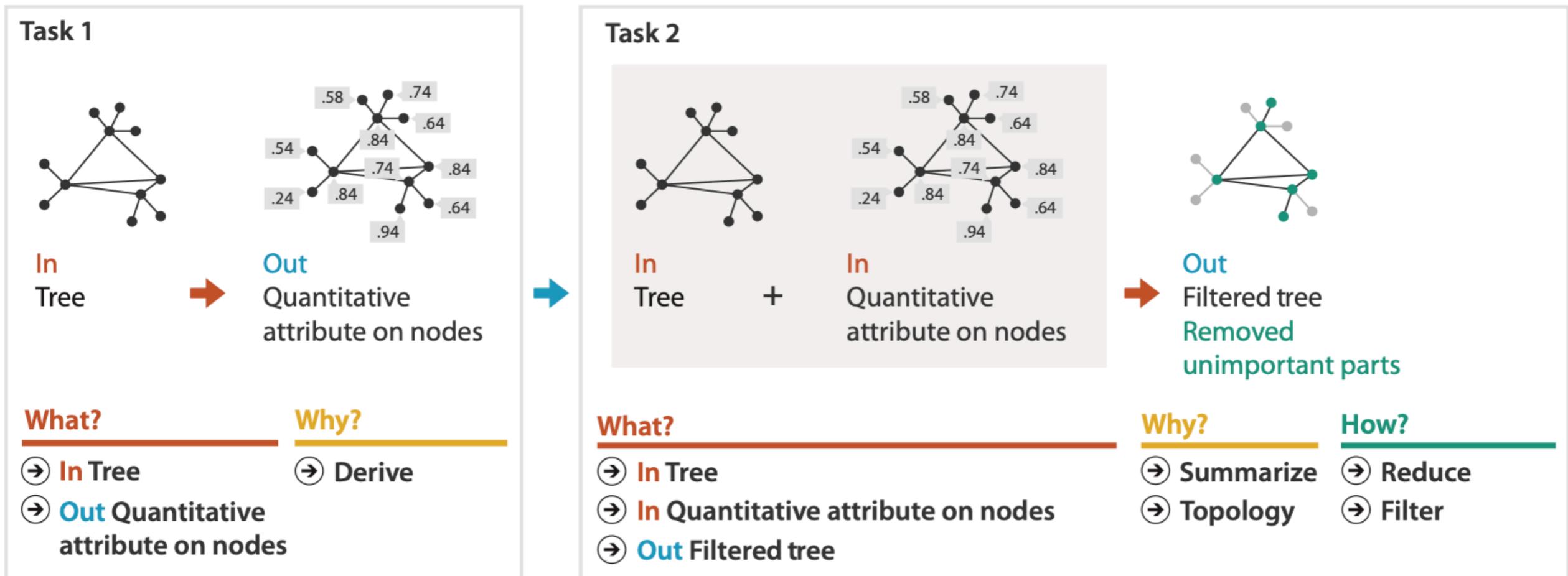


# Deriving a new attribute



**Figure 3.10.** The derived quantitative attribute of Strahler numbers is used to filter the tree in order to create a recognizable summary. (a) The important skeleton of a large tree is visible when only 5000 of the highest-ranked nodes are drawn. (b) The full tree has over a half million nodes. From [Auber 02, Figures 10 and 13].

# Deriving a new attribute: a chained sequence



## Further Reading and Summary



**Q&A**

# Further Reading

---

- **Chapter 3 of Visualization Analysis & Design, Tamara Munzner, 2015**