How to make good visualizations?

Design Criteria

Expressiveness

Tell the truth, the whole truth and nothing but the truth!

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language **express** all **the facts** in the set of data, and only the facts in the data.

Effectiveness

Use visual encodings that people can decode better (faster and more accurate).

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

McKinlay, 1986

Effectiveness
“Use visual encodings that people can decode better”

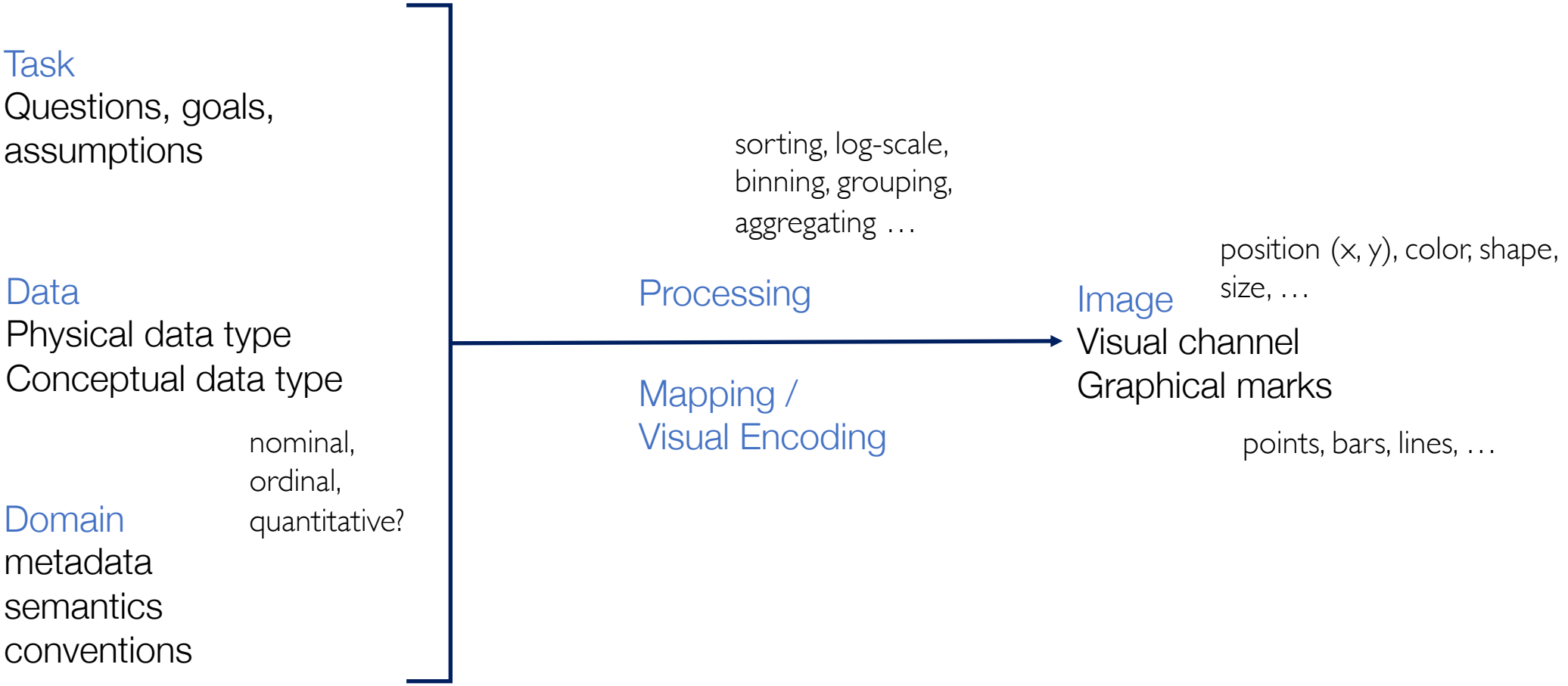
1. Learn about Data Models

2. Learn about Visual Encoding

3. Learn the Rules

4. Examples

5. Summarize with practical tips



Outline

Data Models

Data Model
Formal Description of the data set

Conceptual Model
Mental constructions include semantics and support reasoning

Latitude	Longitude	Surface	Date	Time (UTC)	Temperature (°C)
41.25	-120.9762	Land	Jan 1, 2022	05:00	21.78
89.42	170.7629	Sea	Jan 1, 2022	06:20	-40.80
0.70	130.5643	Land	Jan 1, 2022	08:20	45.60
decimal	decimal	string	date	time	decimal (float)
Spatial location		surface	Time of measurement		Temperature

Data Types

Data Type

A categorization that helps us determine which visual encoding to use

N - Nominal (labels or categories)

Fruits: apples, oranges, ...

O - Ordered

Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

Dates: Jan, 19, 2006;
Location: (LAT 33.98, LONG -118.45)
Only differences (i.e., intervals) may be compared

Q - Ratio (zero fixed)

Physical measurement: length, mass, time duration, counts and amounts
Can measure ratios or proportions

1D (sets and sequences)

Temporal

2D (maps)

3D (shapes)

nD (relational)

Trees (hierarchies)

Networks (graphs)

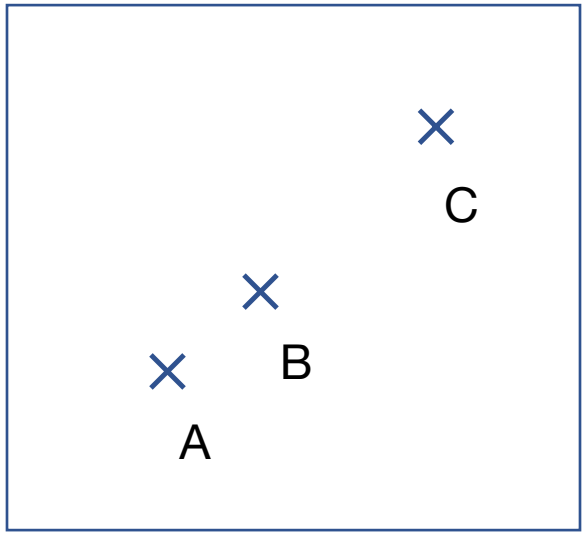
Data Types

Latitude	Longitude	Surface	Date	Time (UTC)	Temperature (°C)	Temperature
41.25	-120.9762	Land	Jan 1, 2022	05:00	21.78	Warm
89.42	170.7629	Sea	Jan 1, 2022	06:20	-40.80	Cold
0.70	130.5643	Land	Jan 1, 2022	08:20	45.60	Hot
Data Model	decimal	decimal	string	date	time	decimal (float)	string
Conceptual Model	Spatial location		surface	Time of measurement		Temperature	Feeling
Data Type	Q-interval		Nominal	Q-interval		Q-ratio	Ordinal
	2-D vector		1-D	1-D		1-D	1-D

Data Type
A categorization that helps us determine which visual encoding to use

- An image is set of signs
- Sender encodes information in signs
- Receiver decodes information from the signs

Jacques Bertin, 1967
Semiology of Graphics



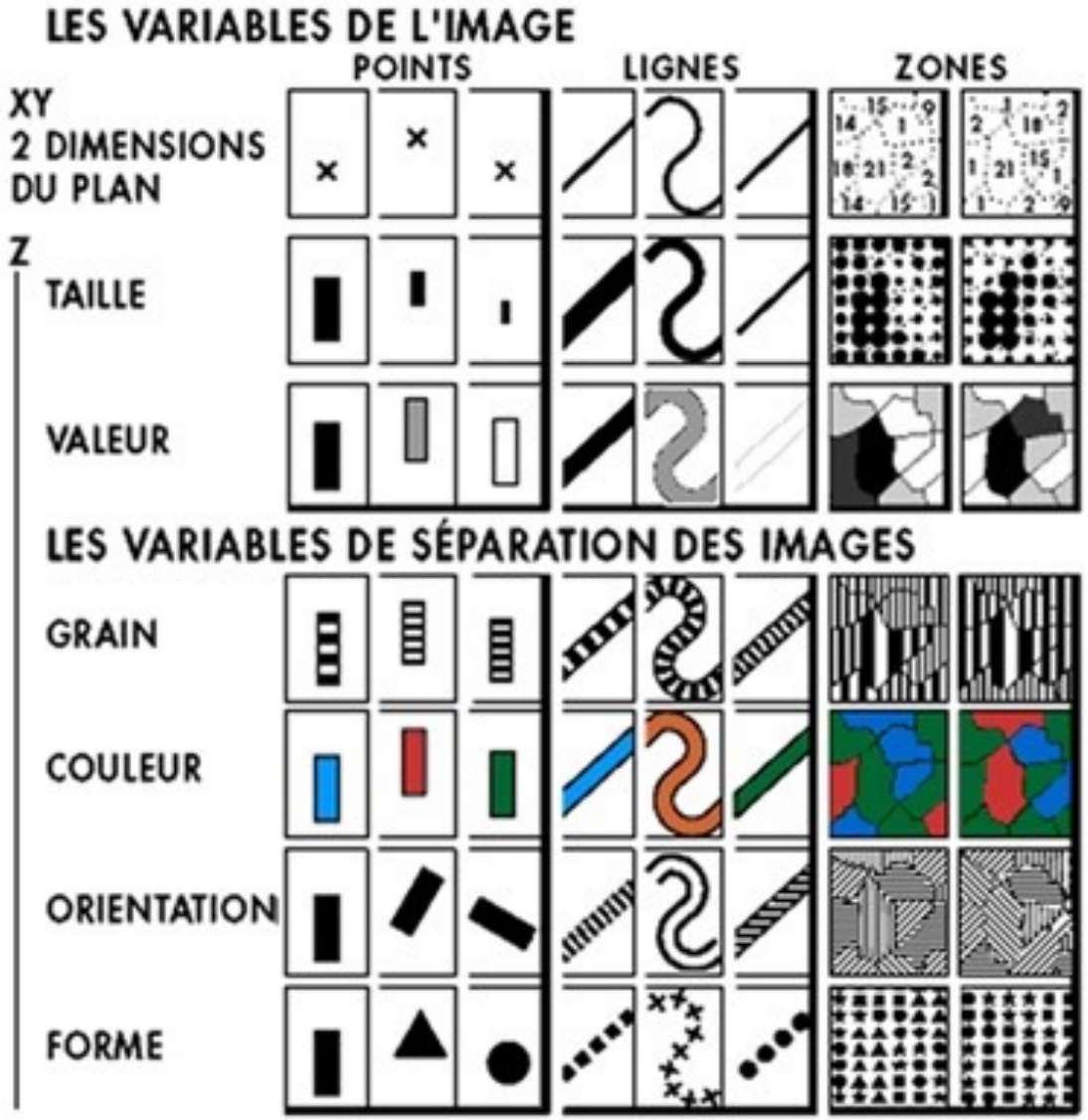
What can I decode from this image?

- A, B, C are distinguishable
- B is between A and C
- The distance between B and C is twice that of AB

What can I encode then using the position (channel) of points (marks)?

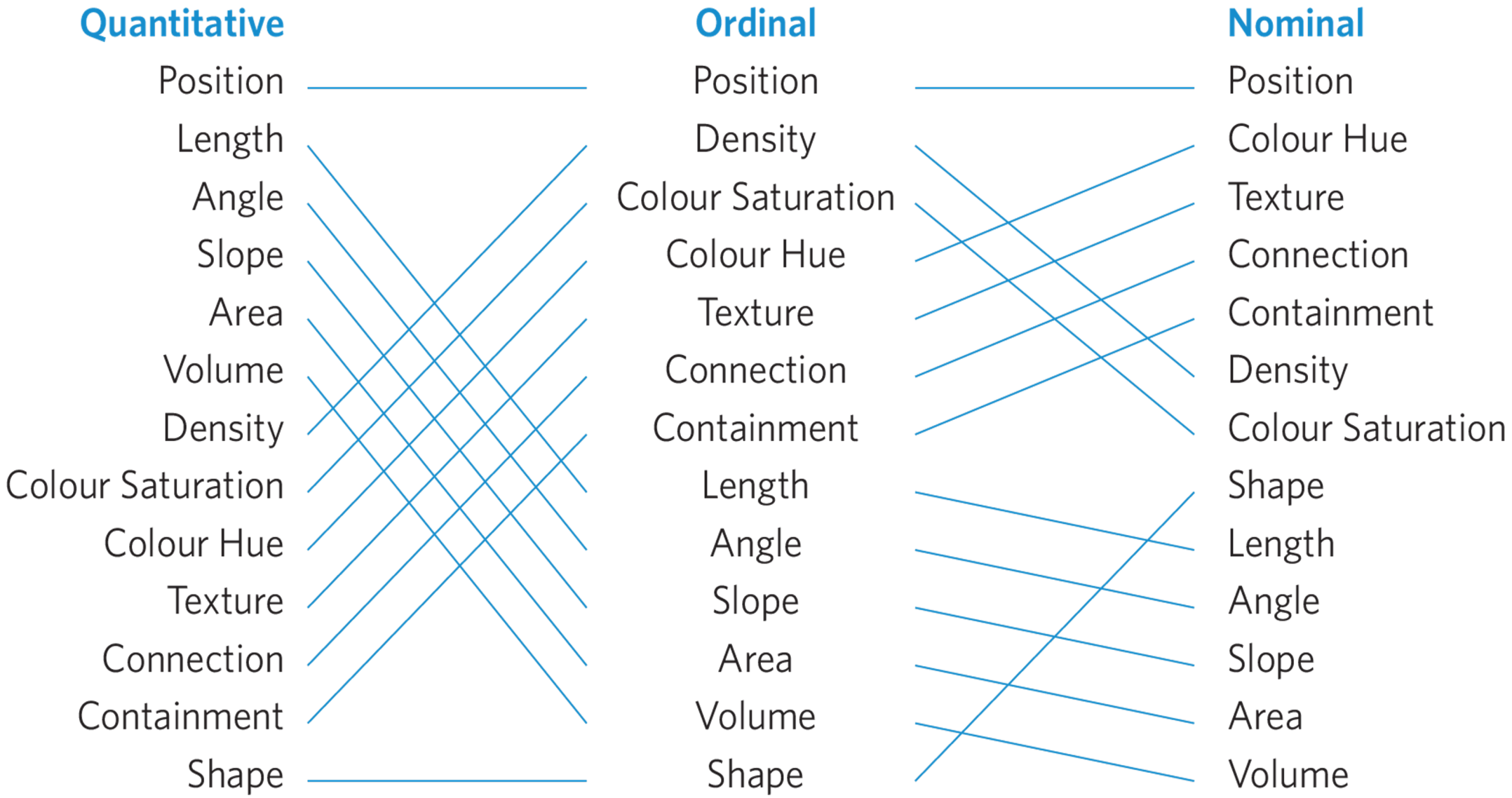
- Quantitative variables

Visual language is a Sign System



Position	N	O	Q
Size	N	O	Q
Color (Value)	N	O	Q
Texture	N	O	
Color (Hue)	N		
Orientation	N		
Shape	N		

Bertin's Visual Encoding Variables & Levels of Organization



Mackinlay's Effectiveness Ranking

Mackinlay's Automated Design Algorithm

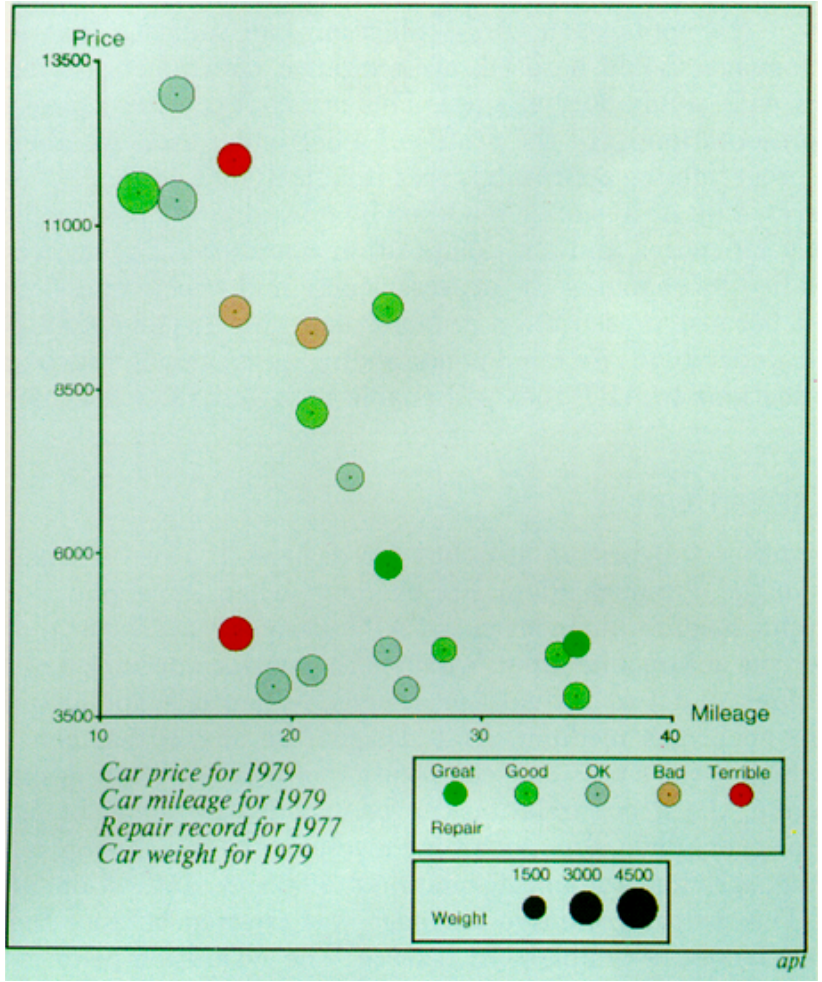
Assume k visual encodings and n data attributes. Pick the "best" encoding among a combinatorial set of possibilities of size $(n+1)^k$

- **Principle of Consistency**
  The properties of the image (visual variables) should match the properties of the data.
- **Principle of Importance Ordering**
  Encode the most important information in the most effective way.

Input: data model and type, ordered list of data variables to show

APT searches over design space
- Test expressiveness of each visual encoding
- Generate encodings that pass test
  Rank by perceptual effectiveness criteria
- **Output** the "most effective" visualization

Input variables:
1. Price 2. Mileage 3. Repair 4. Weight

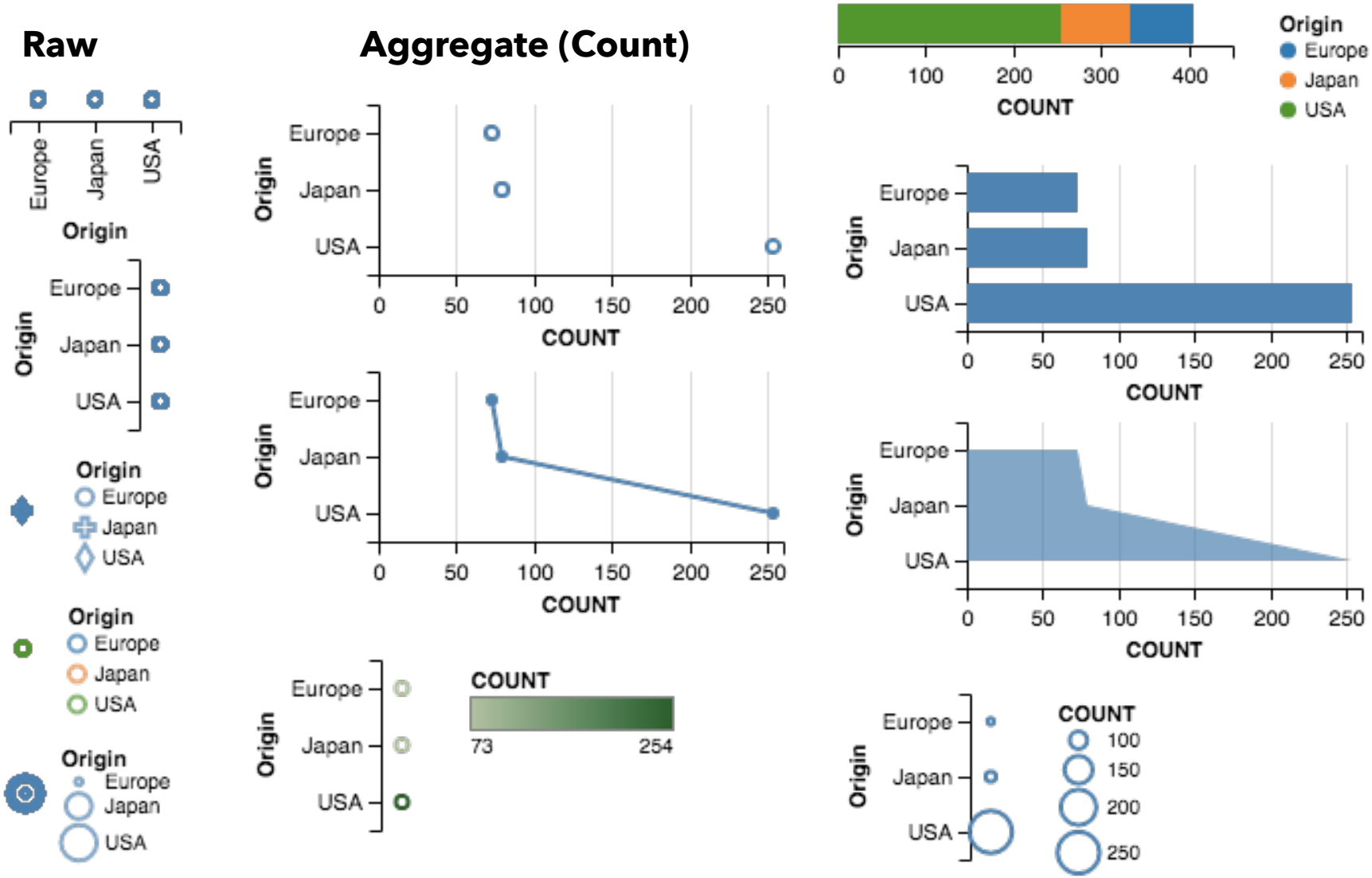


APT - "A Presentation Tool", 1986

Examples

Input
1-D Nominal sequence of car origins

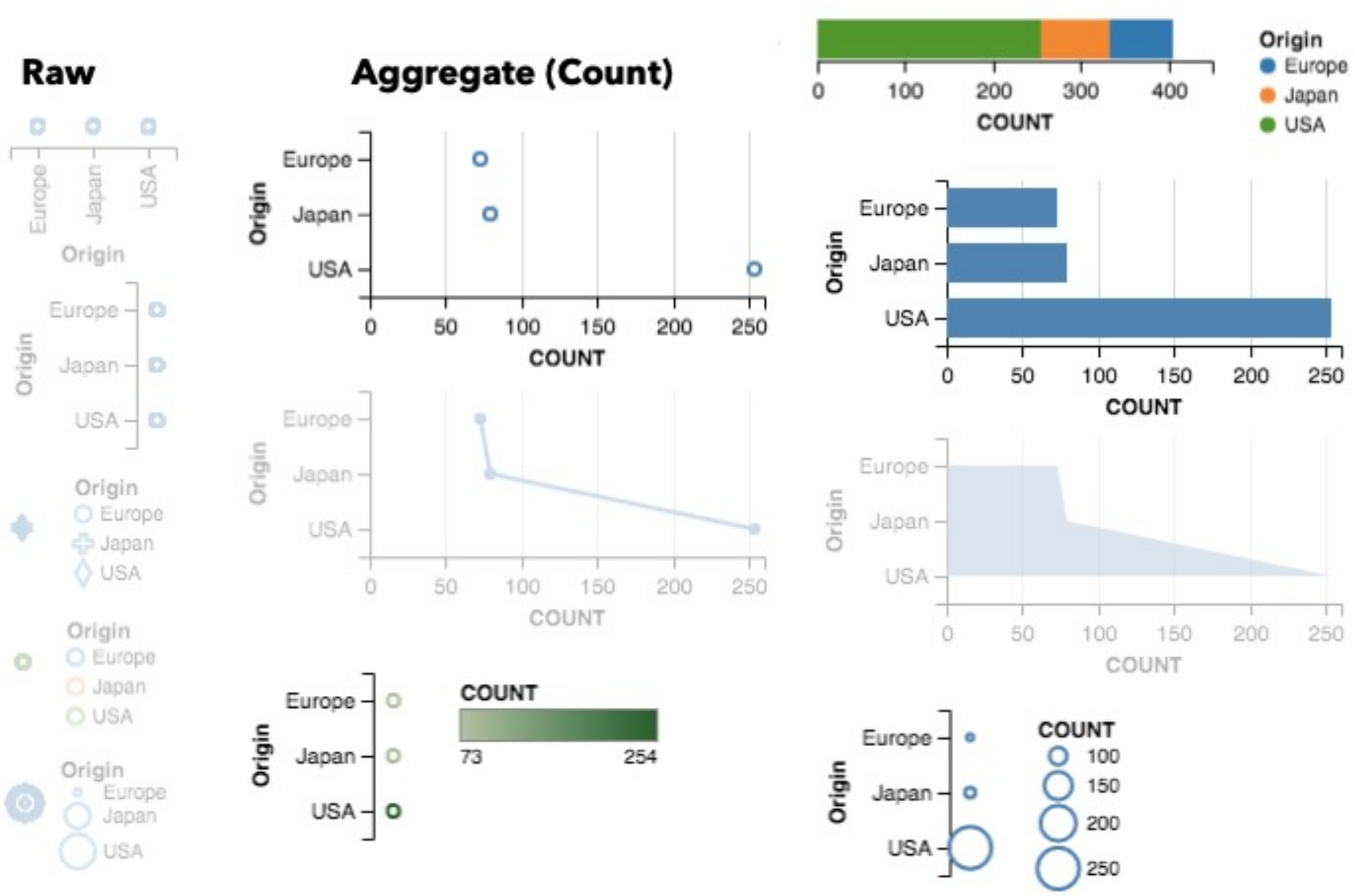
Origin
Europe
Japan
Europe
Japan
USA
USA
USA
Japan
…



Expressive?

Input
1-D Nominal sequence of car origins

Origin
Europe
Japan
Europe
Japan
USA
USA
USA
Japan
…

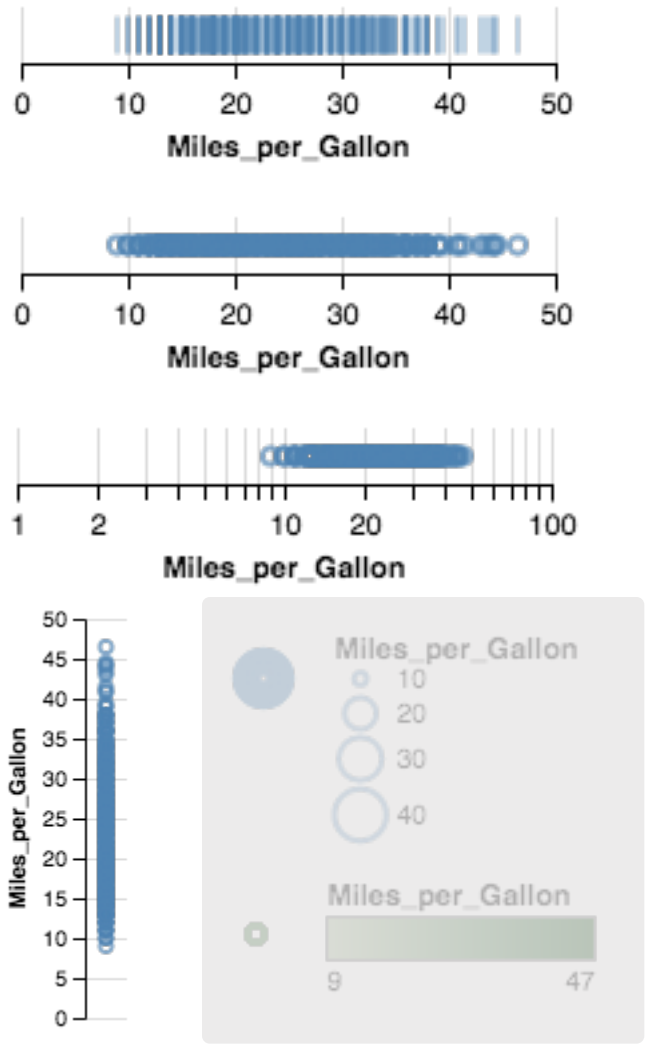


Effective?

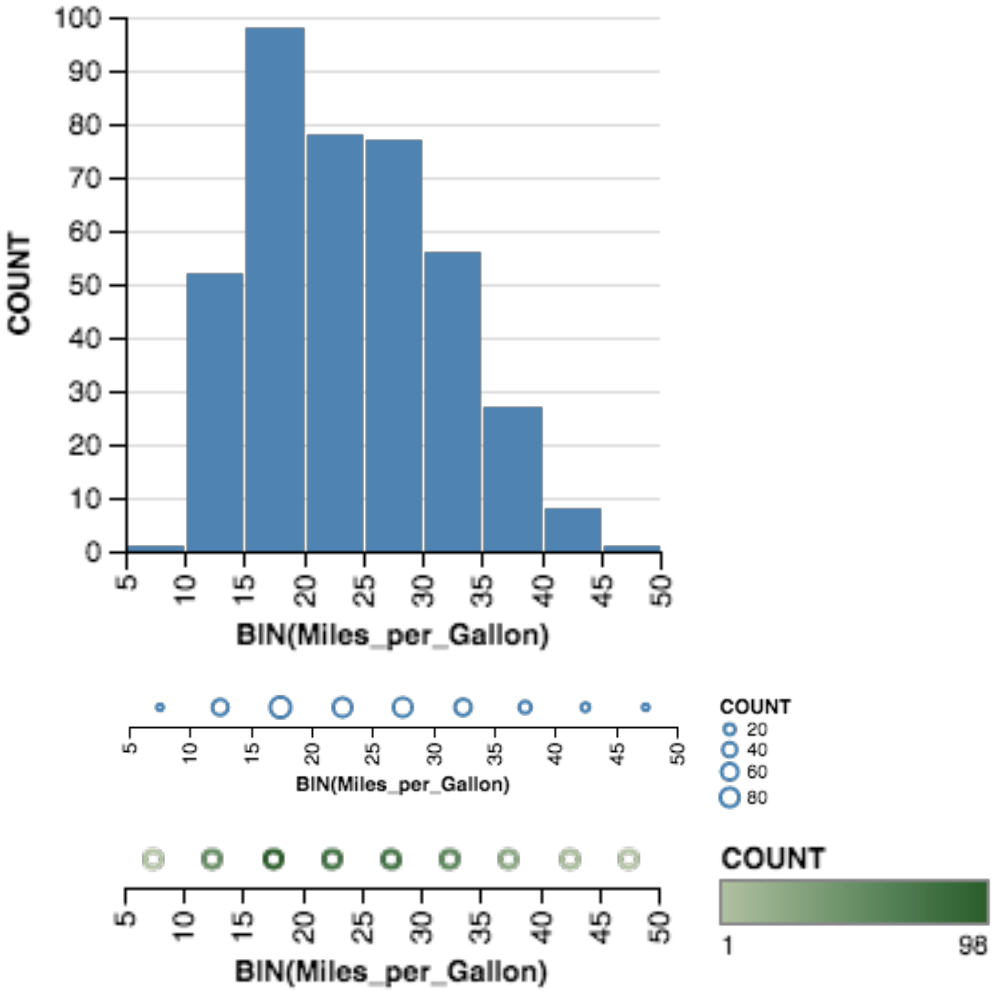
Input
1-D Quantitative
sequence of mileage

Miles per gallon
30.4
48.1
20.5
10.0
35.2
30.8
40.1
21.1
…

Raw



Aggregate (Count)

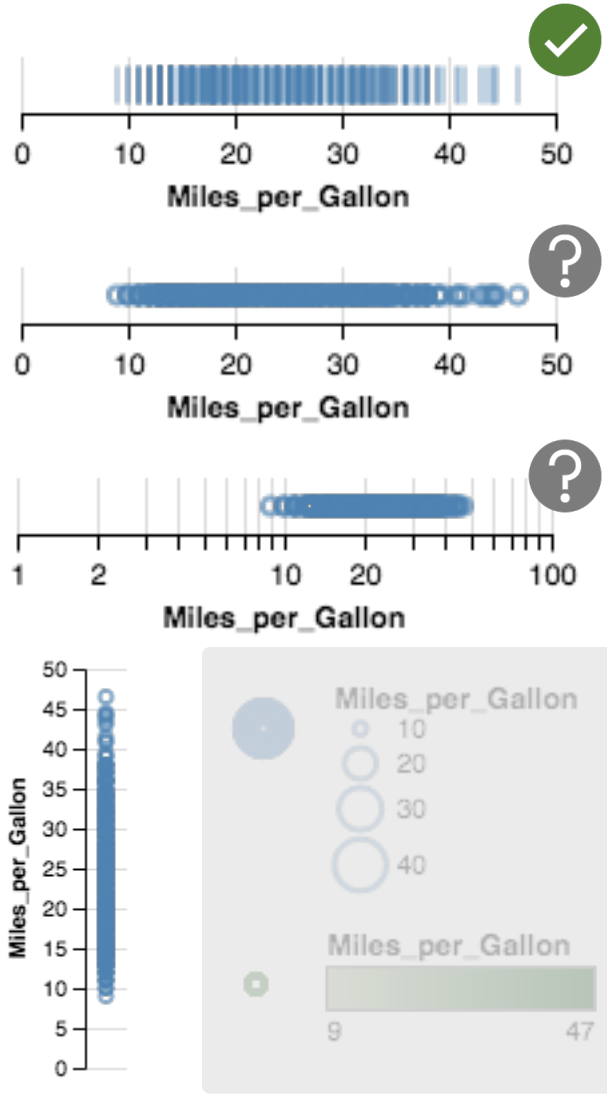


Expressive?

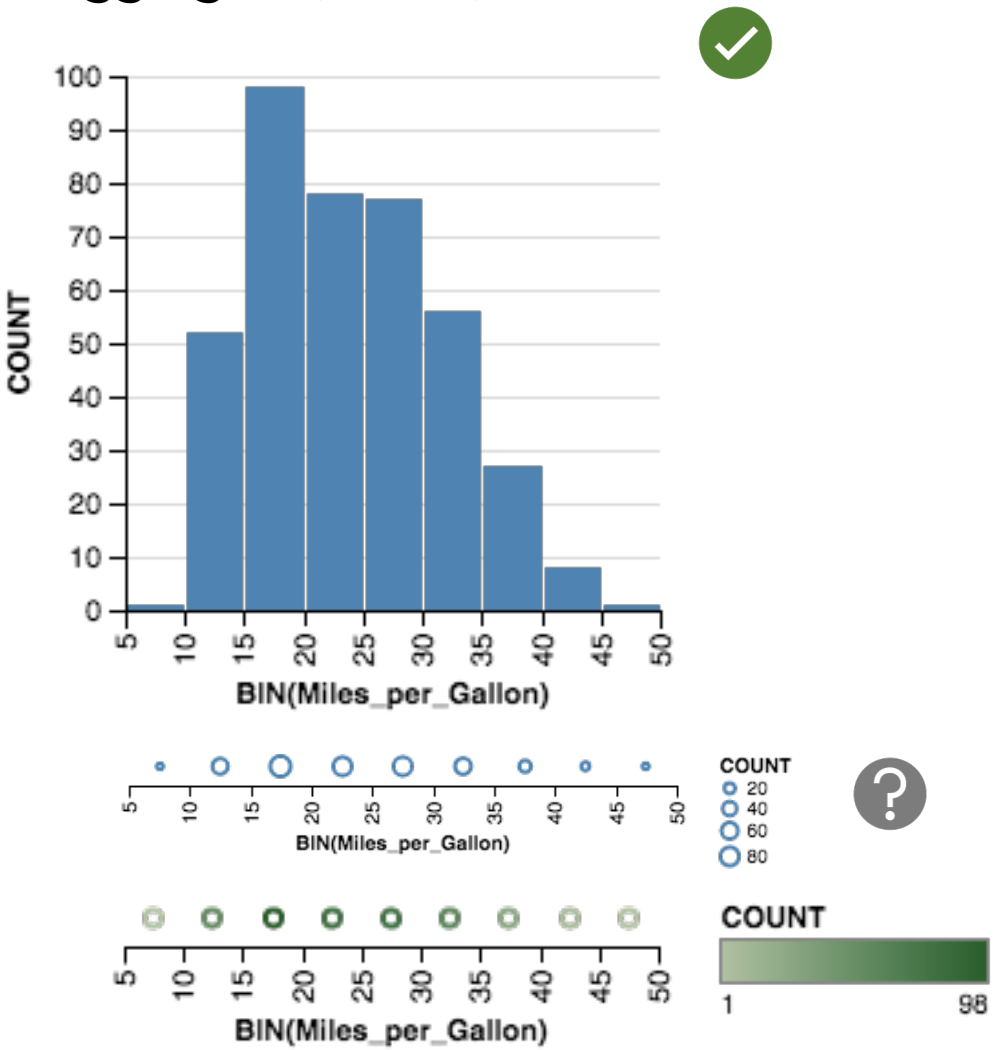
Input
1-D Quantitative
sequence of mileage

Miles per gallon
30.4
48.1
20.5
10.0
35.2
30.8
40.1
21.1
…

Raw



Aggregate (Count)

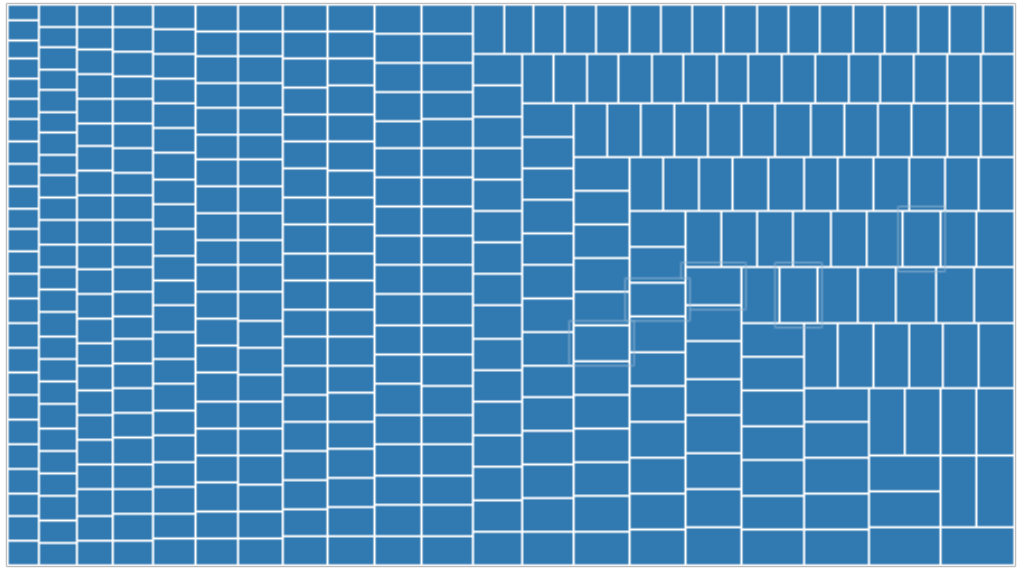


Effective?

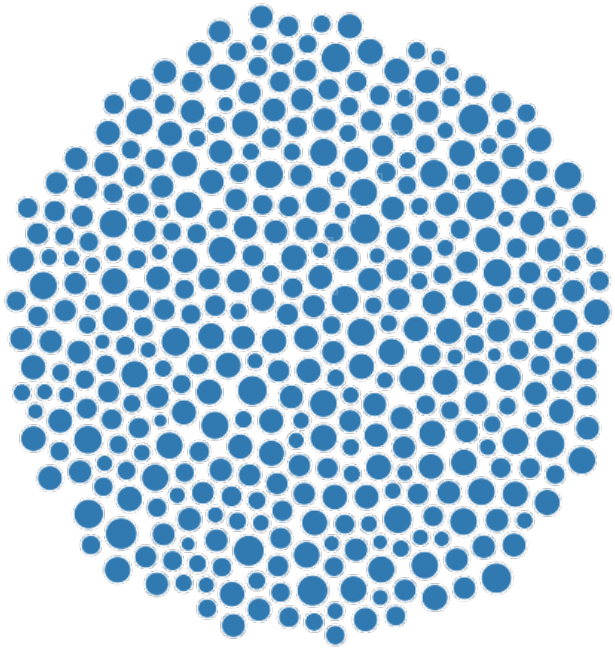
Input
1-D Quantitative
sequence of mileage

Miles per gallon
30.4
48.1
20.5
10.0
35.2
30.8
40.1
21.1
…

Raw (with Layout Algorithm)

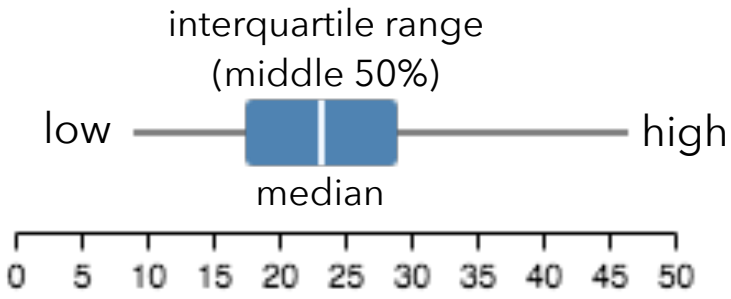


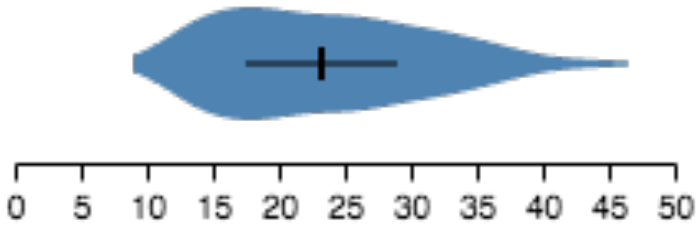
Treemap



Bubble Chart

Aggregate (Distribution)



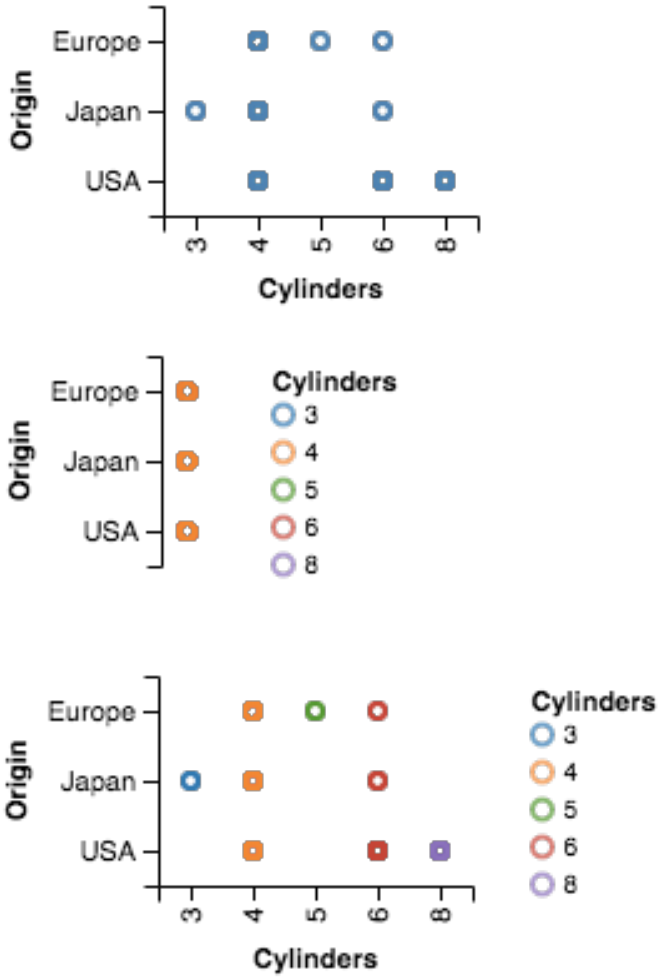


Expressive? Effective?

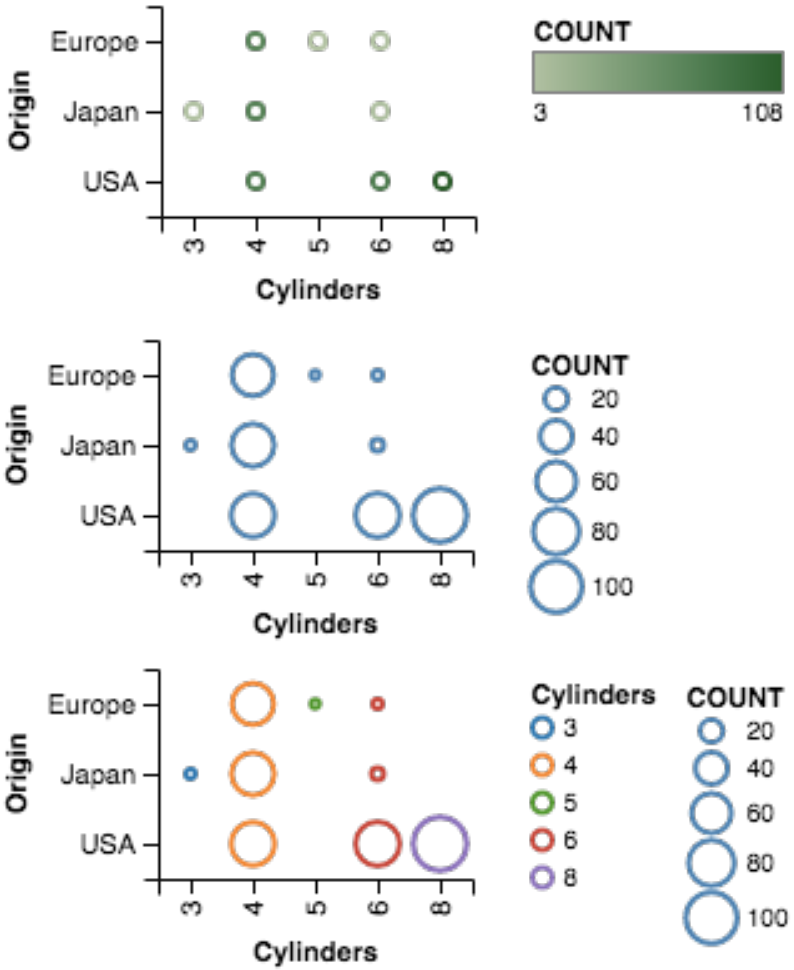
Input
2-D Nominal × Nominal

Origin	Cylinders
Europe	4
Japan	3
Europe	6
Japan	8
USA	5
USA	3
USA	4
Japan	8
…	…

Raw



Aggregate (Count)

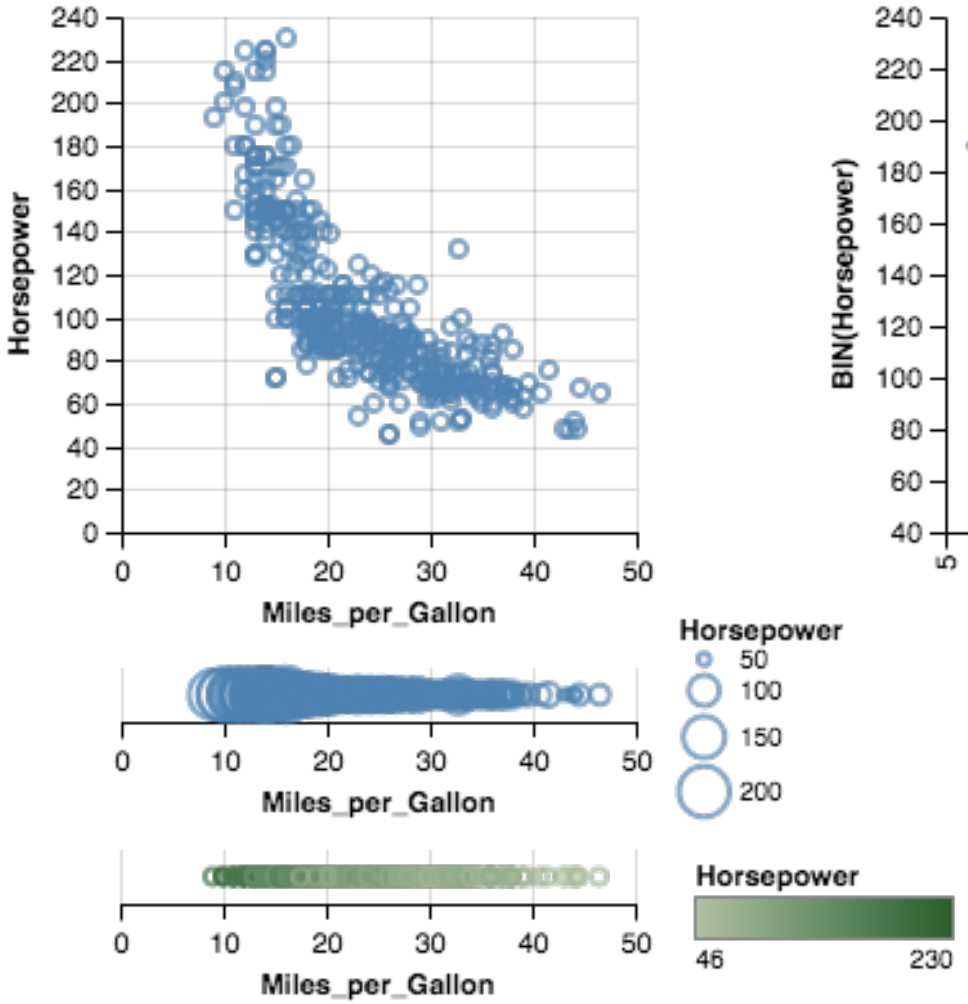


Expressive? Effective?

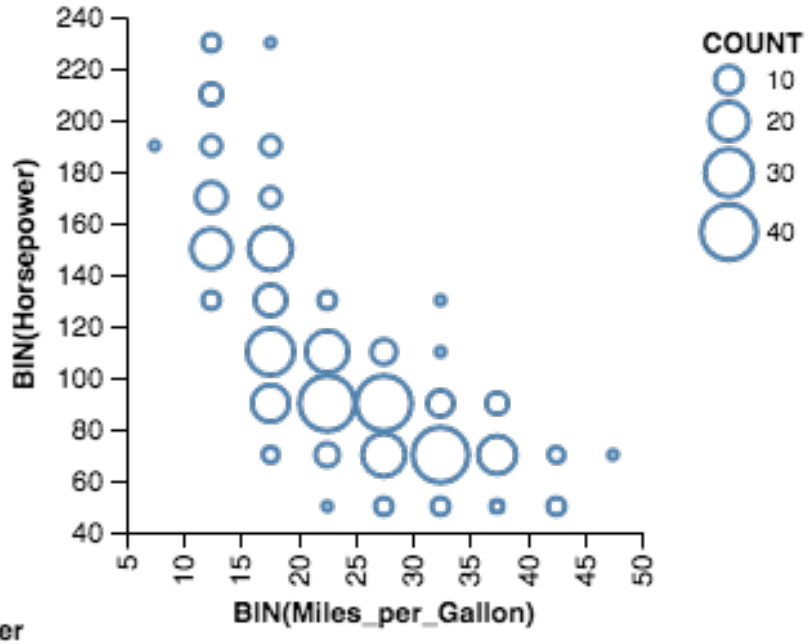
Input
2-D Q × Q

Hourse-power	Mileage
120	28
122	25
144	20
222	14
230	15
80	30
60	40
62	47
…	…

Raw



Aggregate (Count)

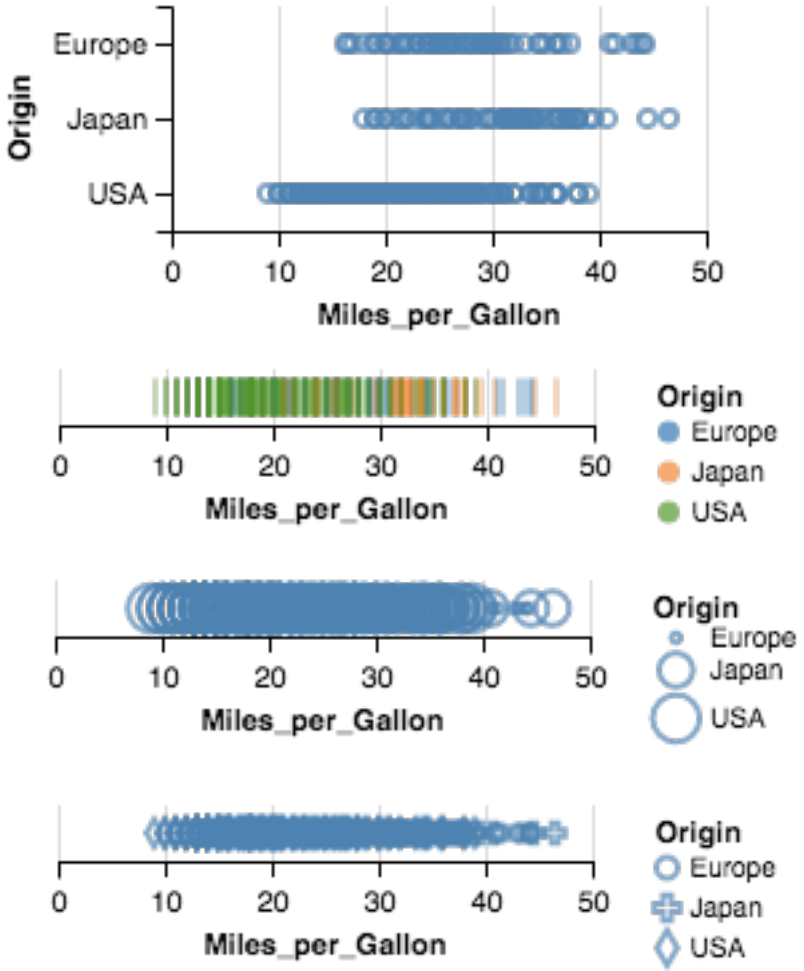


Expressive? Effective?

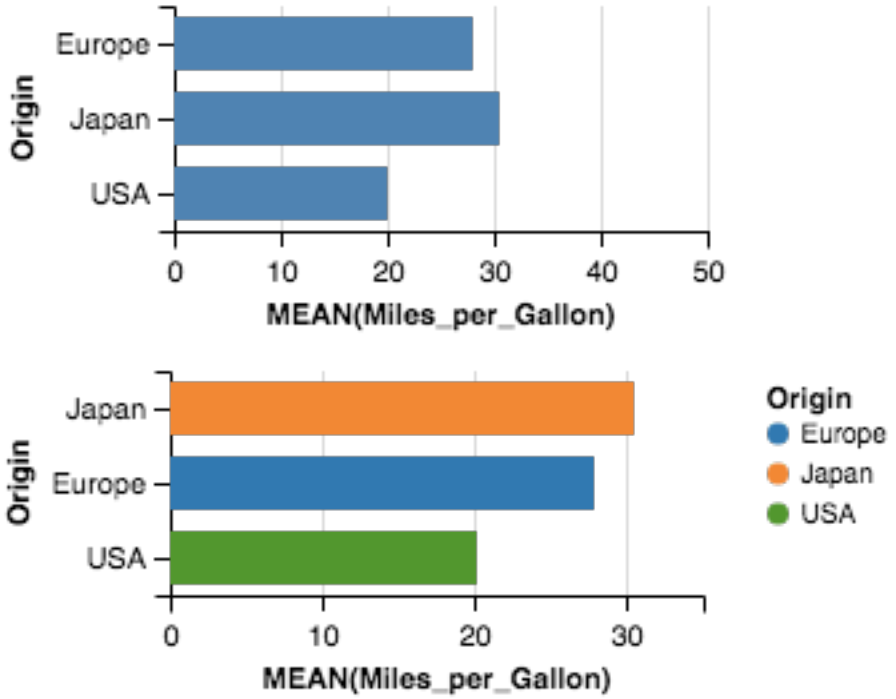
Input
2-D N × Q

Origin	Mileage
Europe	28
Japan	25
Europe	20
Japan	14
USA	15
USA	30
USA	40
Japan	47
…	…

Raw



Aggregate (Mean)

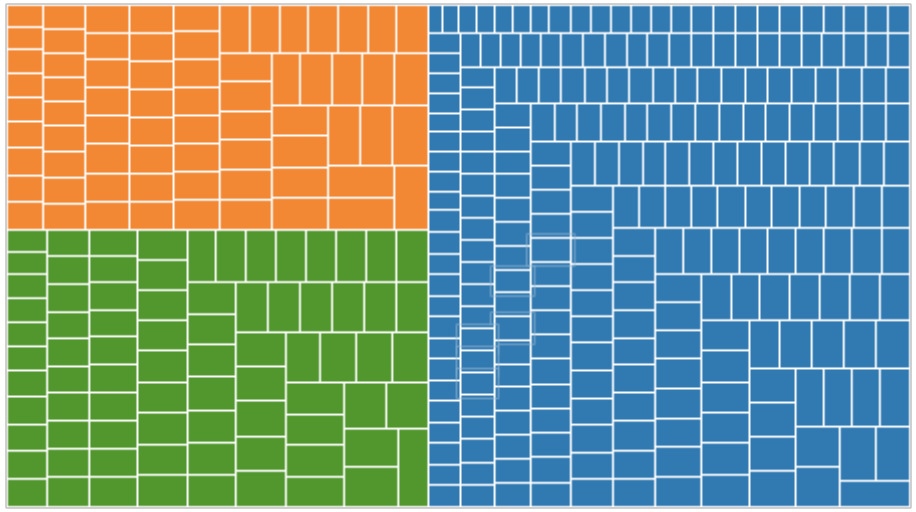


Expressive? Effective?

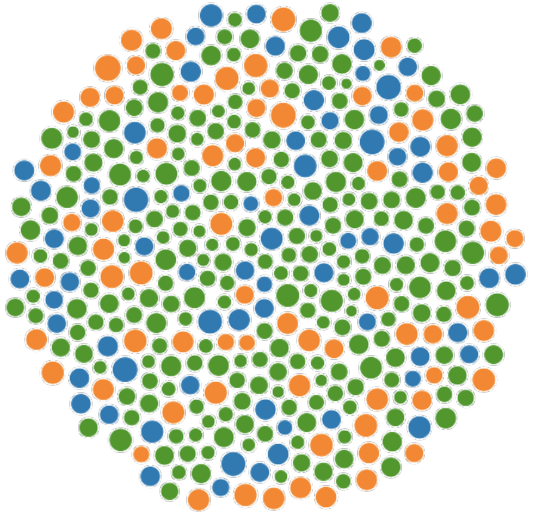
Input
2-D N × Q

Origin	Mileage
Europe	28
Japan	25
Europe	20
Japan	14
USA	15
USA	30
USA	40
Japan	47
…	…

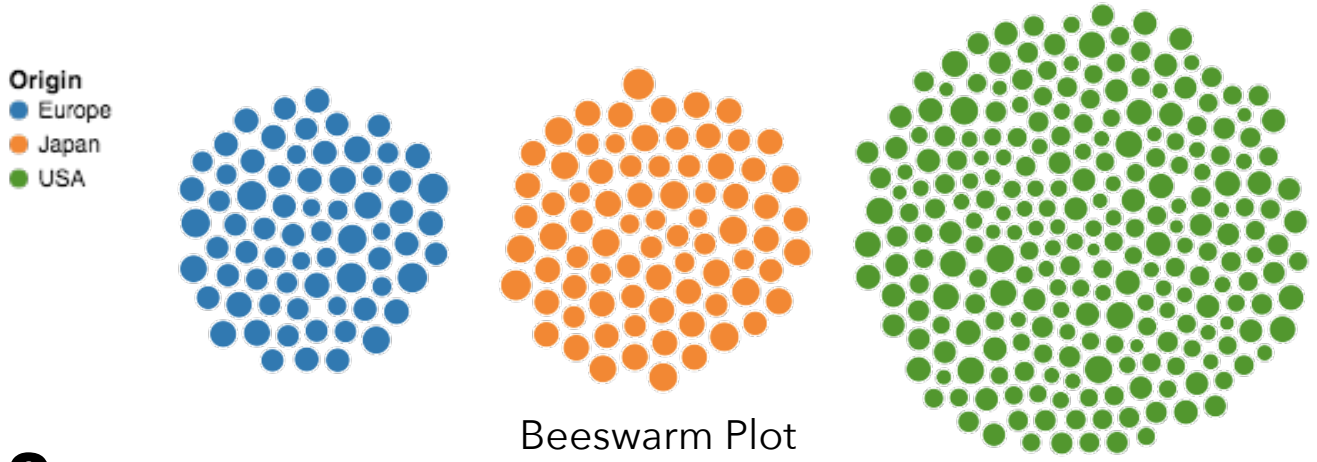
Raw



Treemap



Bubble Chart

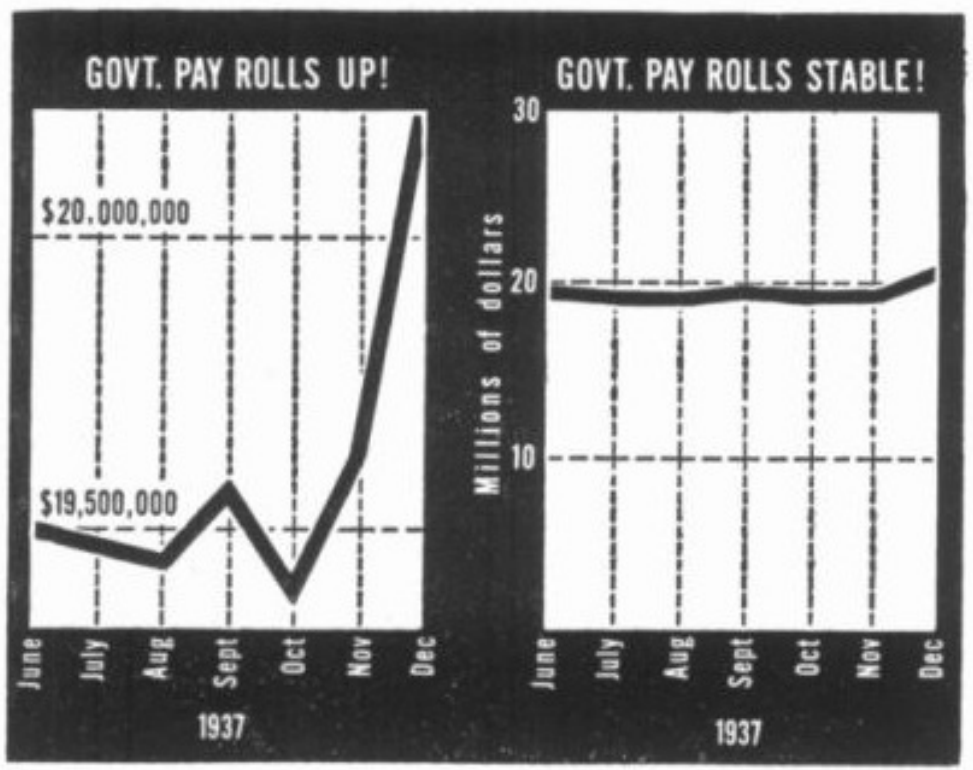


Beeswarm Plot

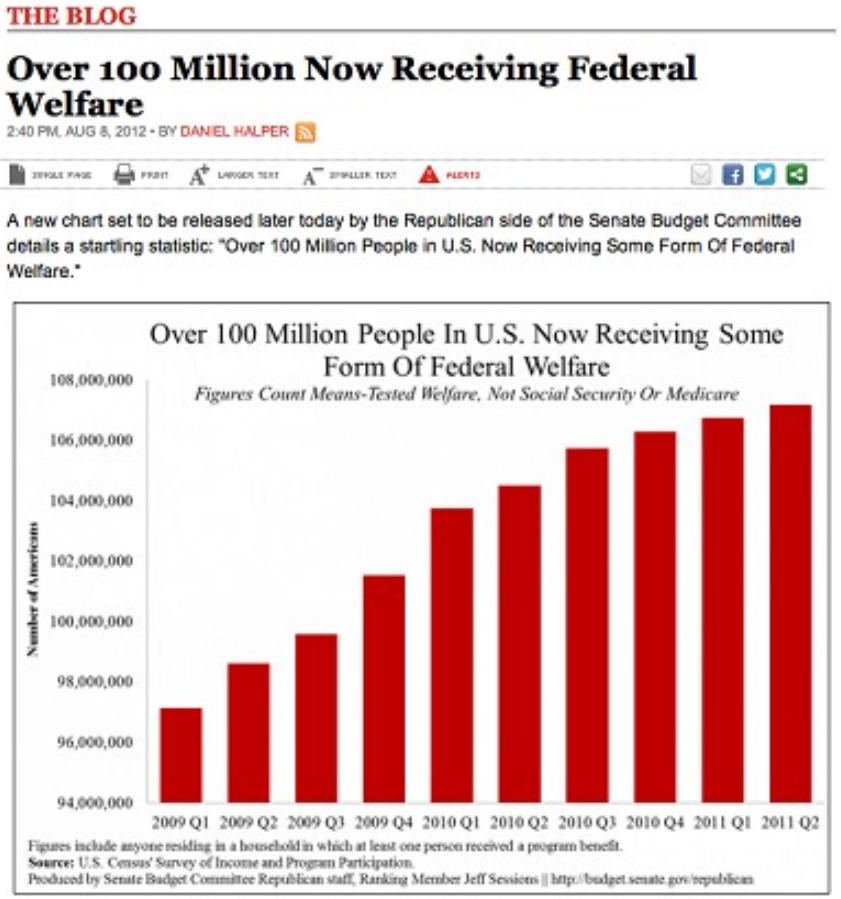
Expressive? Effective?

Design Tips

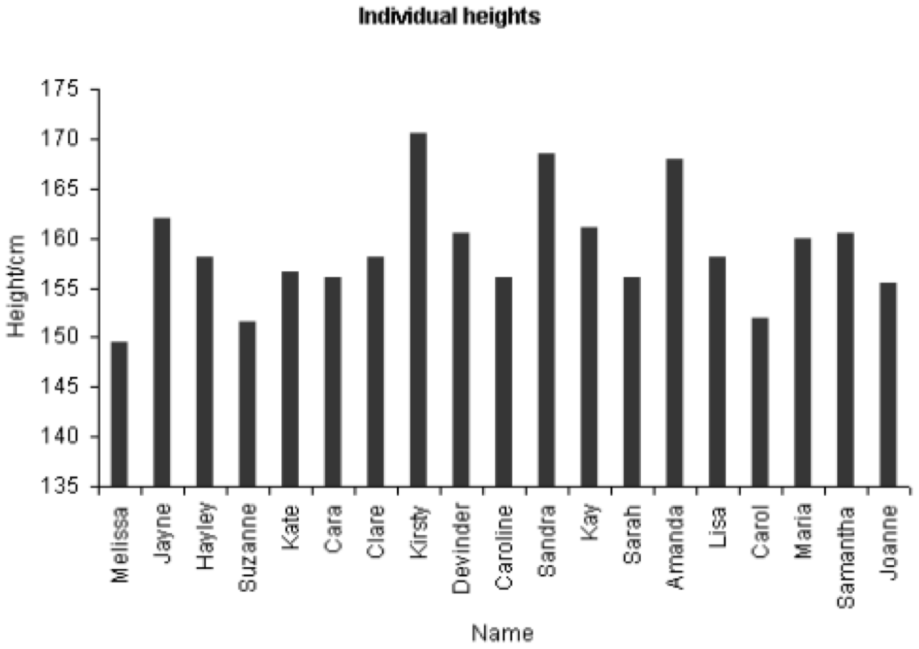
Gee-Whiz Graphs



How to lie with statistics. Huff. Govt payrolls in 1937

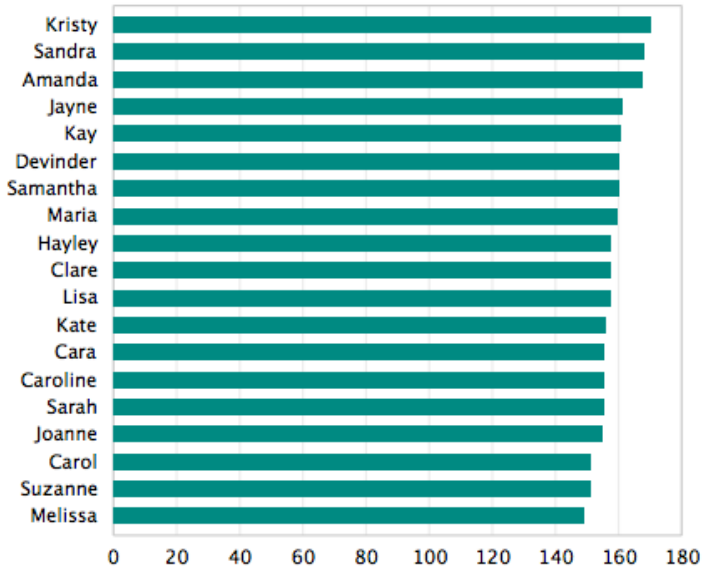


Include Zero?

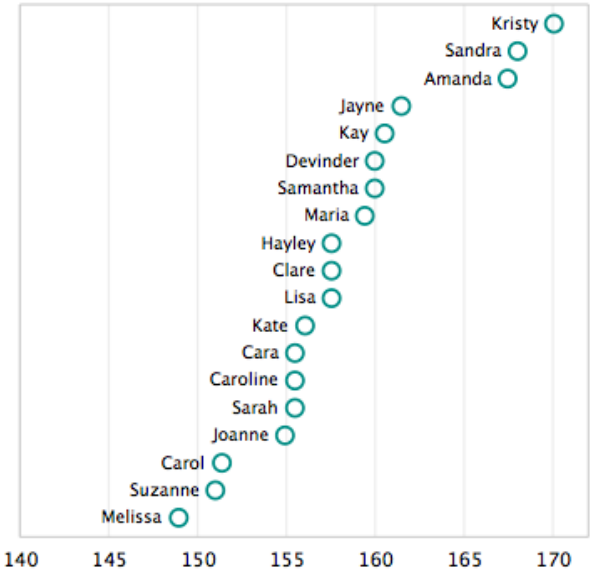


This violates the expressiveness principle!

Bar length encodes the amount

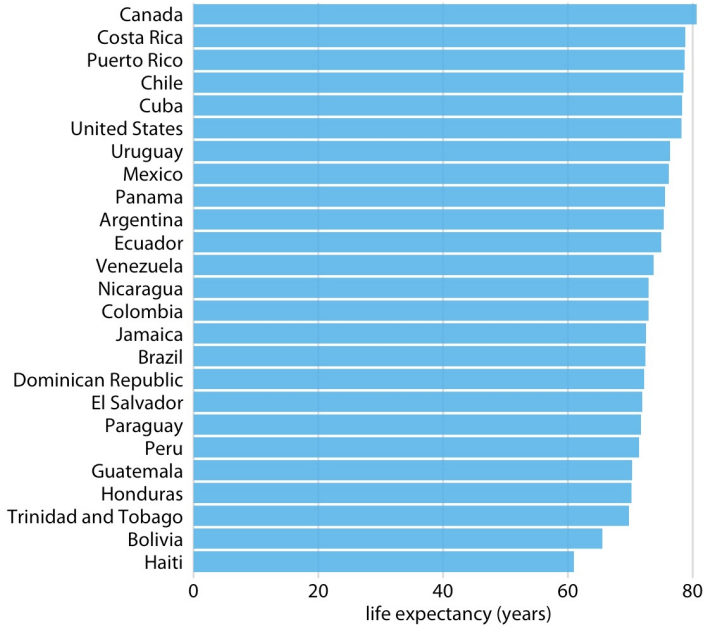


Q-Ratio
“We care about the amount shown”

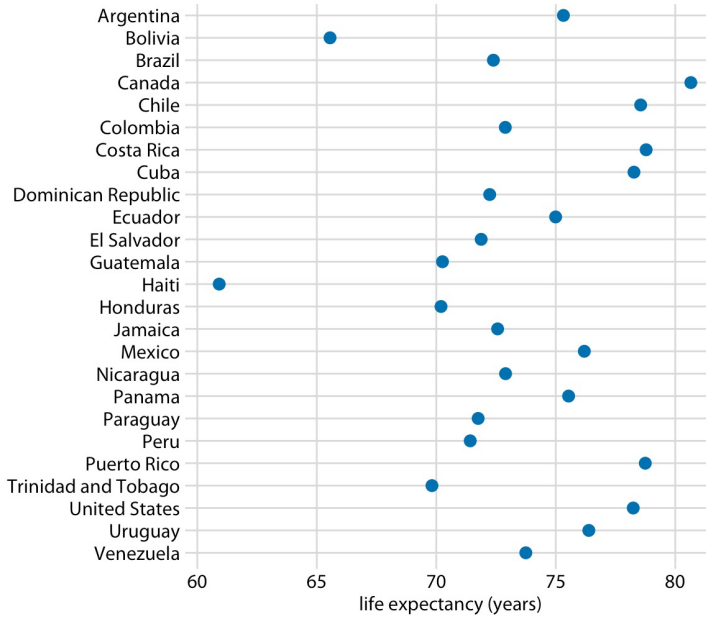


Q-Interval
“We care about the relative position”

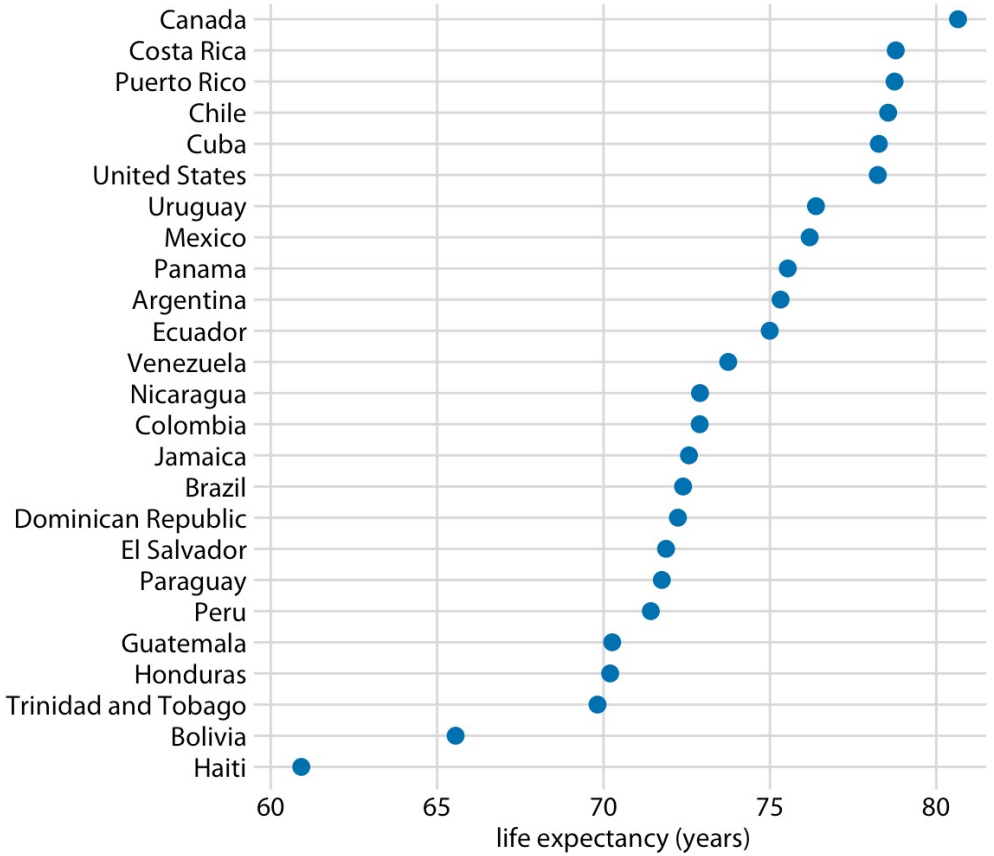
Include Zero?



The bars are draw attention away from the key feature of the data: the differences in life expectancy among the different countries.



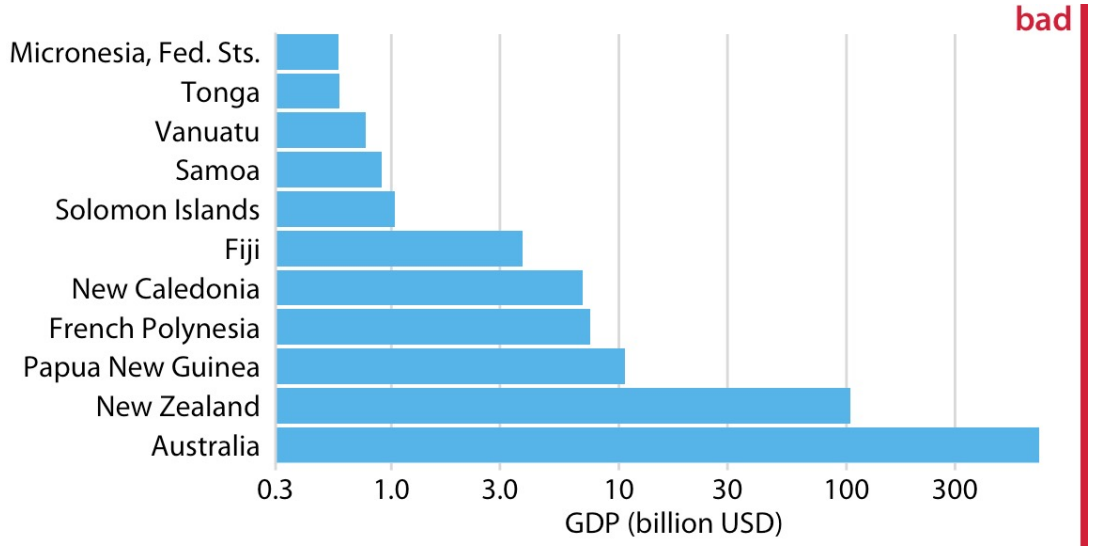
The countries are ordered alphabetically, which causes a dots to form a disordered cloud of points. This makes the figure difficult to read.

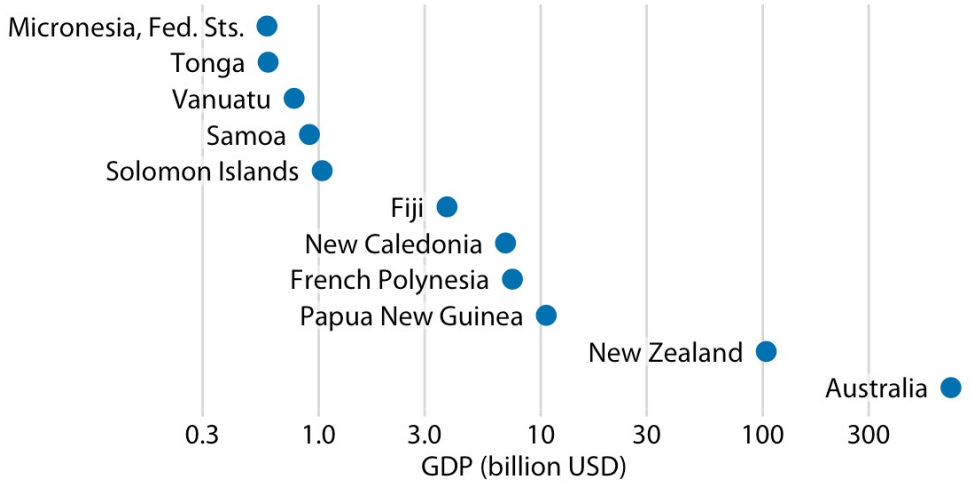


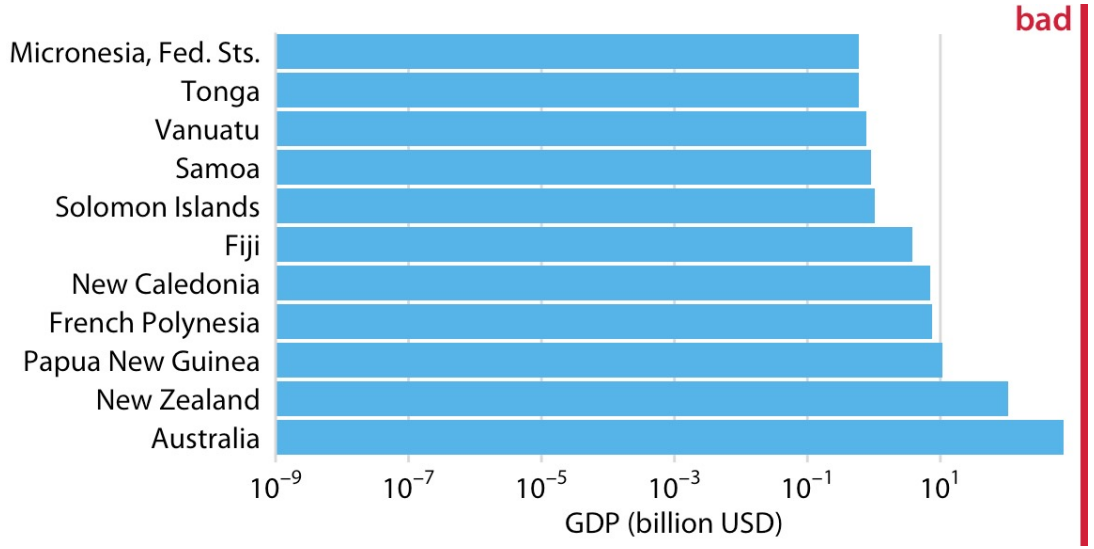
Can be made better by removing the y-axis and labeling each dot: avoids generating the visual perception of a magnitude conveyed by the distance from the name to the dot

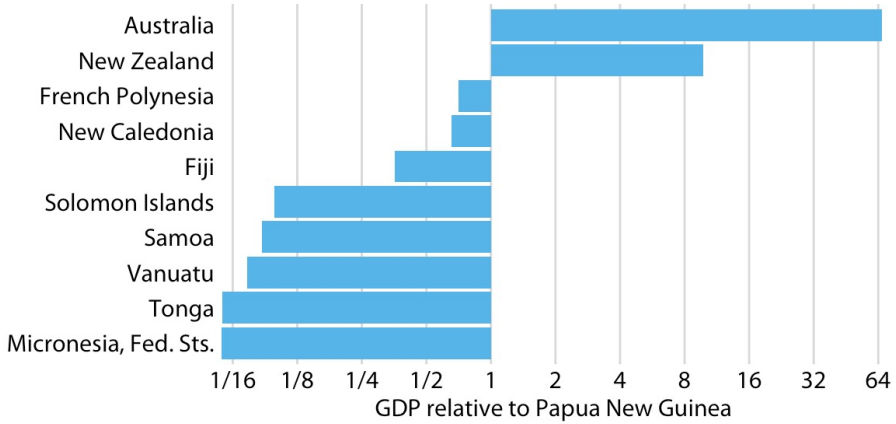
Bar vs. Dot Plots

Fundamentals of Data Visualization, Wilke









Zero, Bars, Dots & Log Scales

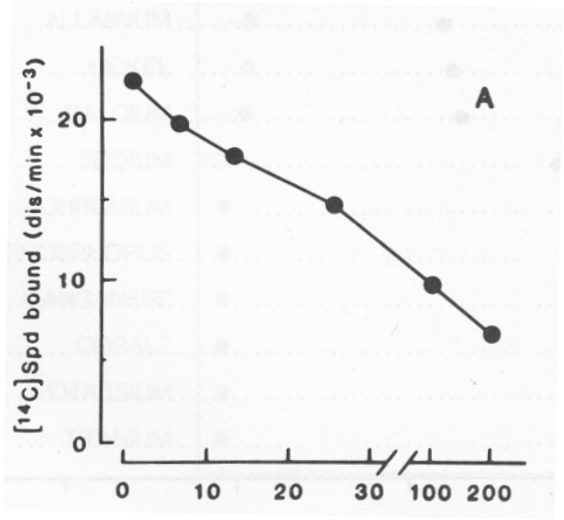
Fundamentals of Data Visualization, Wilke

More about Log Scales

Address data skew

e.g., long tails, outliers

Enables comparison within and across multiple orders of magnitude.



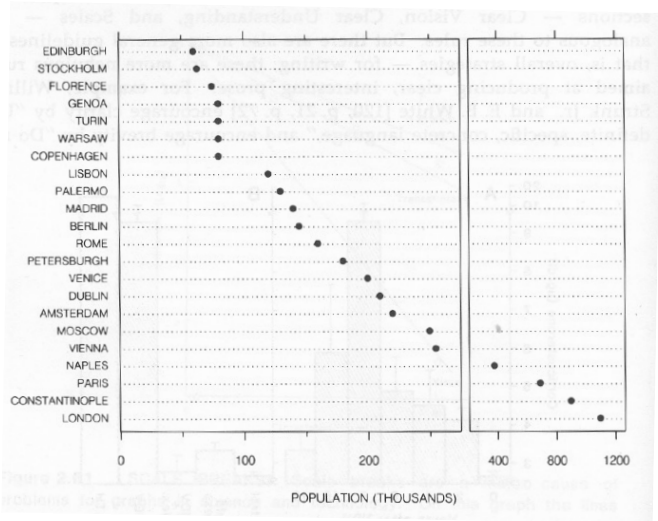
Violates Expressiveness!

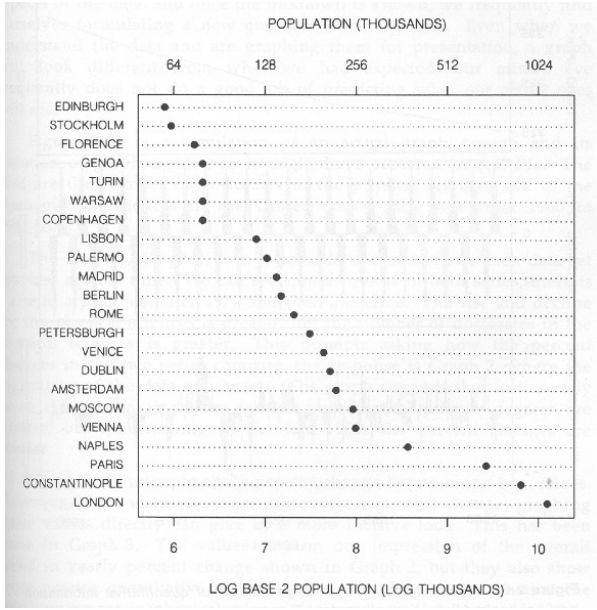
Focus on multiplicative factors

e.g., The GDP of Australia is 64 times that of Papa New Guinea

The logarithm transforms × to +!

Percentage change, not linear difference.



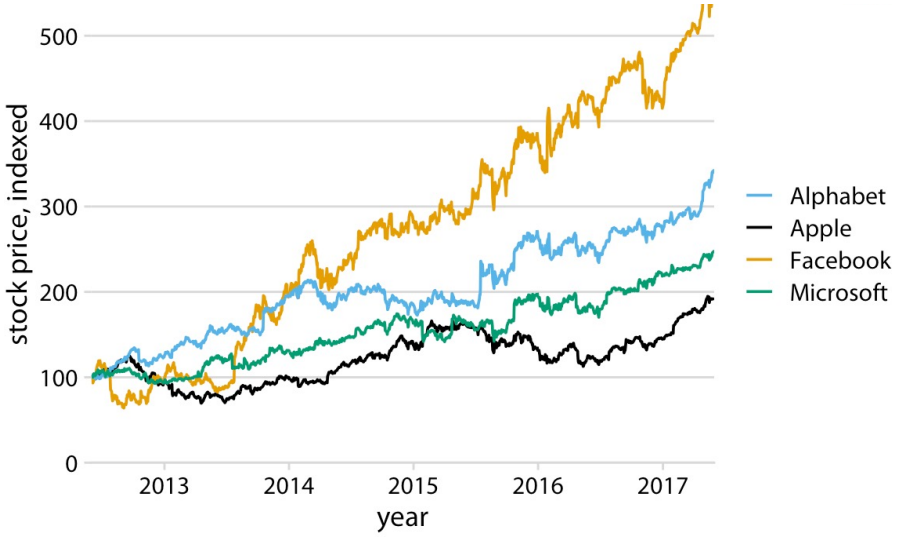


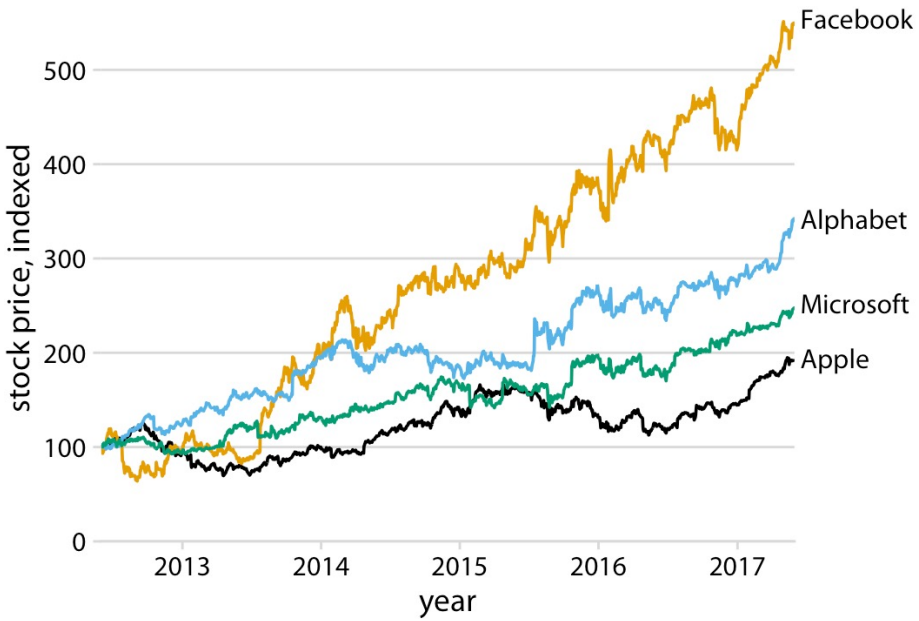
Scale breaks vs. Log scales: Cognitive vs. Perceptual Effort

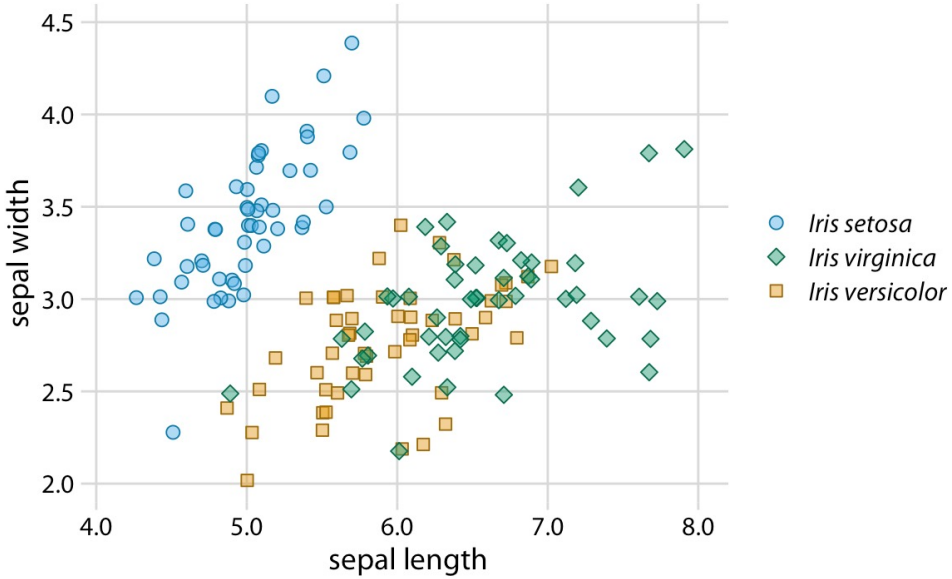
Constraint: **positive, non-zero values**

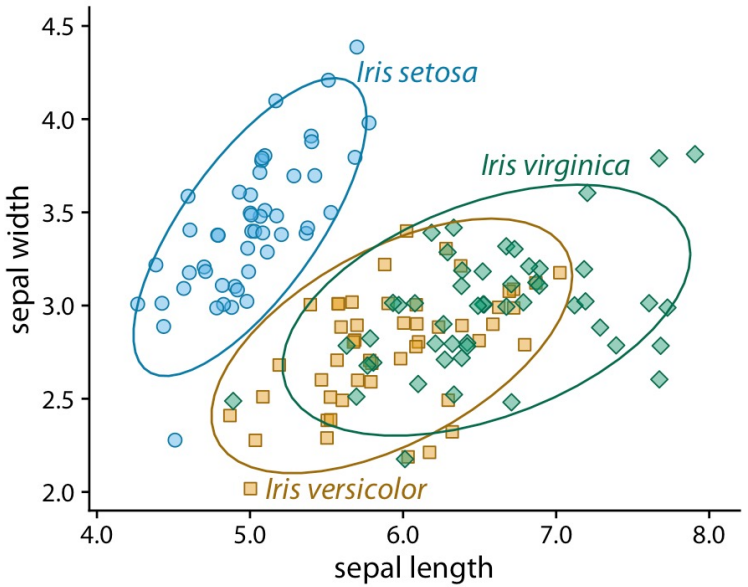
Constraint: **audience familiarity?**

The Elements of Graphing Data [Cleveland, 85]



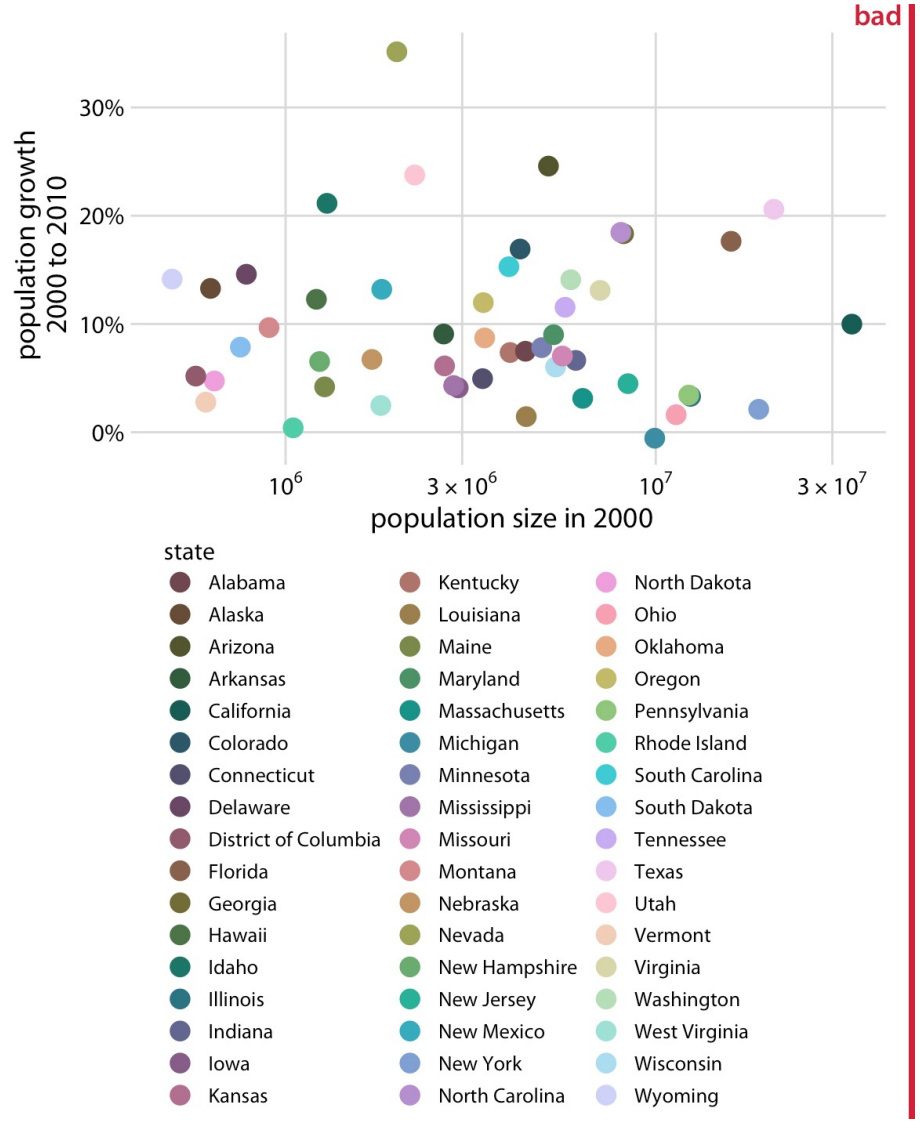


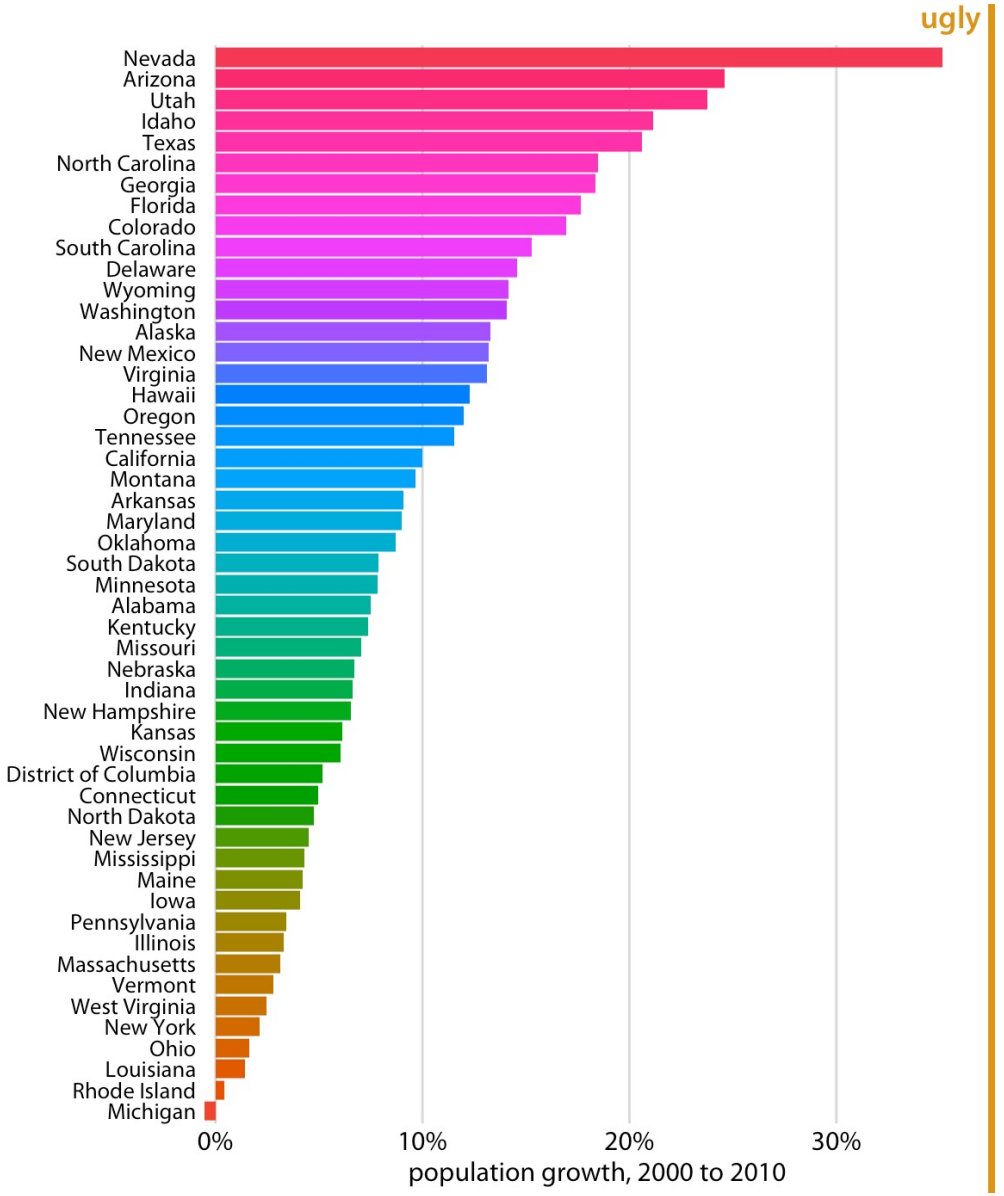




Legends

Fundamentals of Data Visualization, Wilke





Color

Fundamentals of Data Visualization, Wilke

Exercise 2 will be released end of this week

Proposal discussions on Wednesday

1. Form your groups
2. Meet to go over group contracts
3. Come prepared to give us a short discussion of what you want to do
4. Submit a one-page proposal next week

Next in the Course