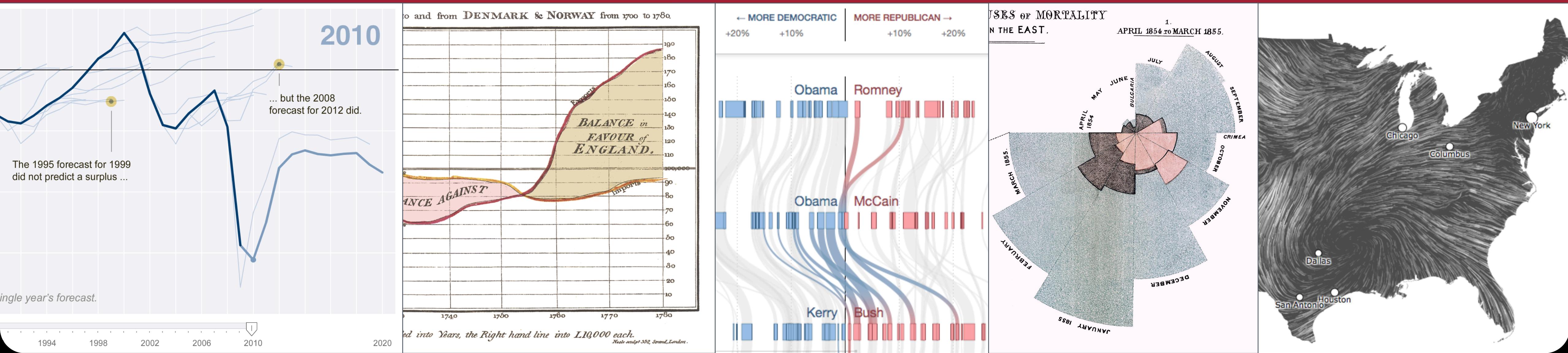


6.894: Interactive Data Visualization Data & Image Models

Arvind Satyanarayan



Activity!

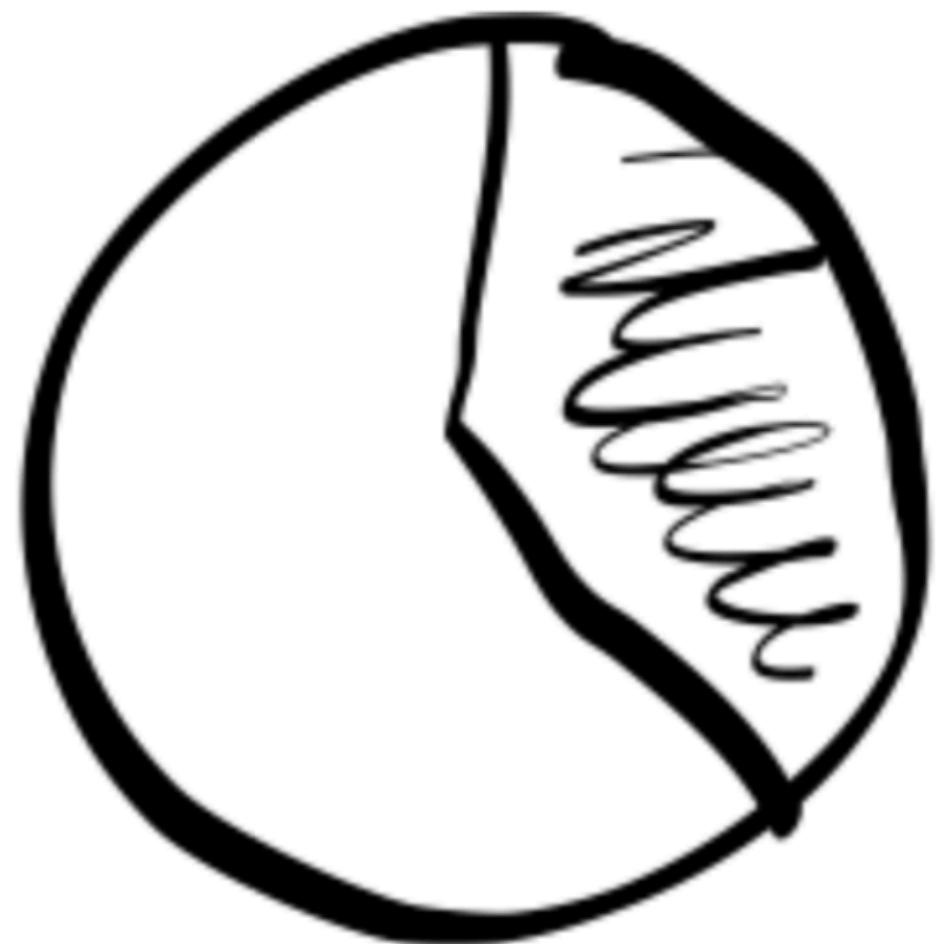
In 1 minute, sketch as many visualizations as possible of these two numbers:

75

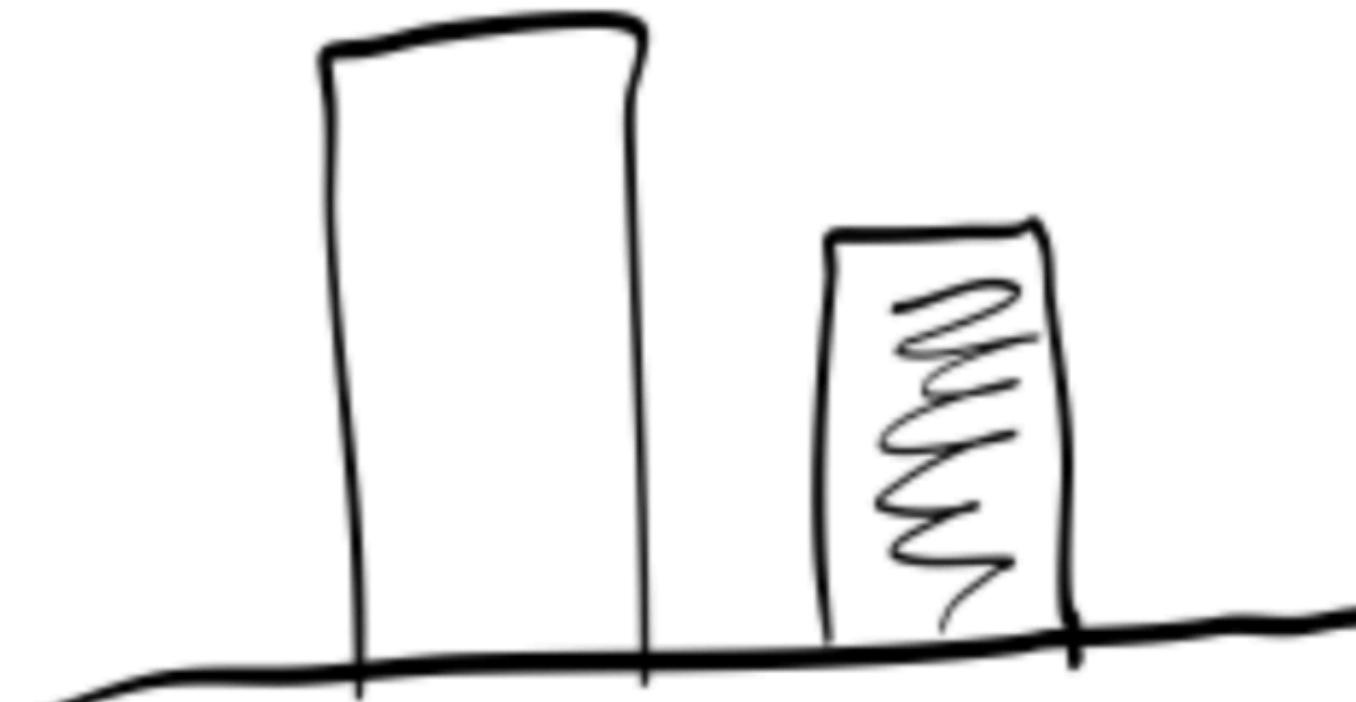
37

Most Likely Results

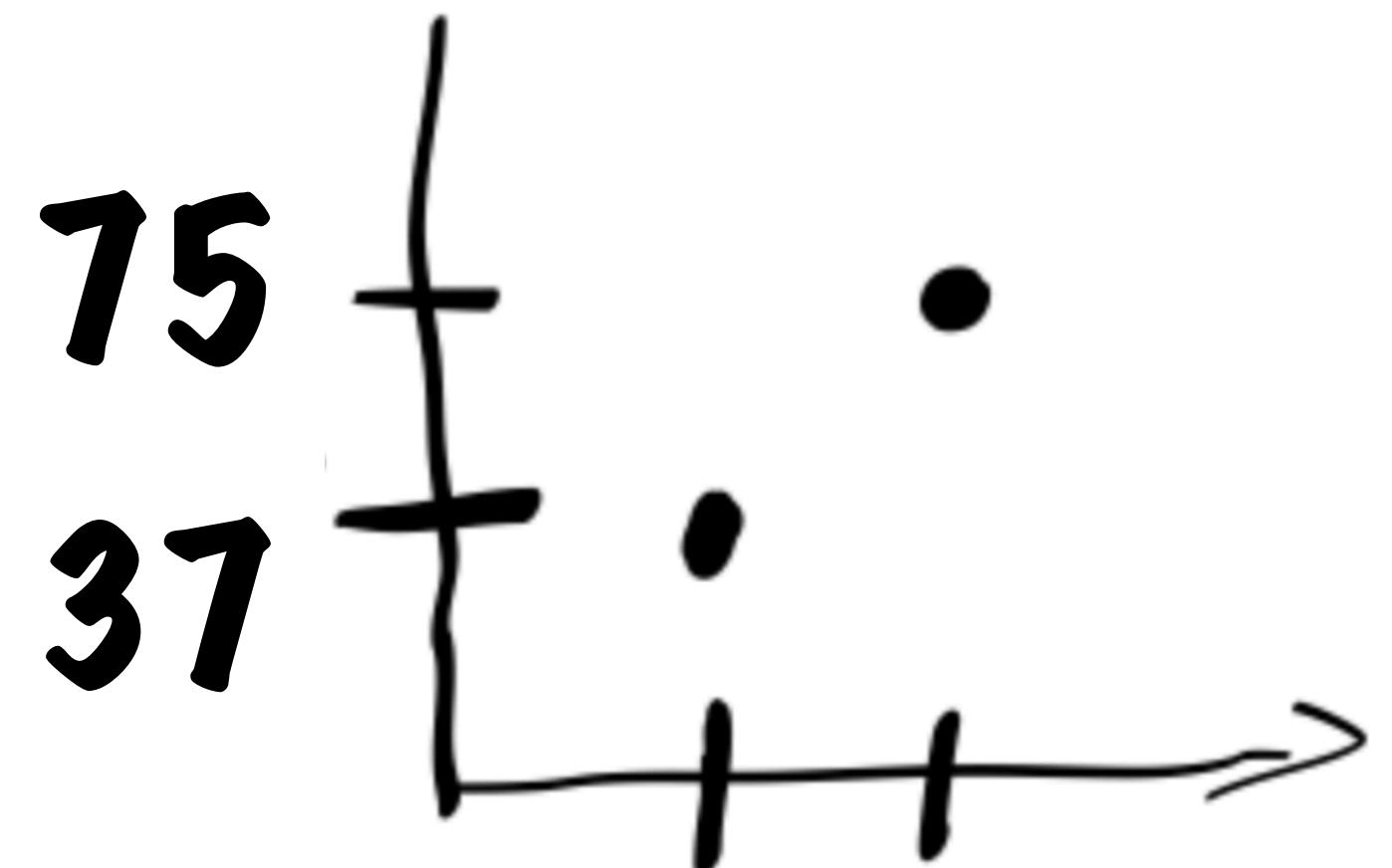
Pie Charts



Bar Charts



Scatterplot



75
37

Arabic Numbers

Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

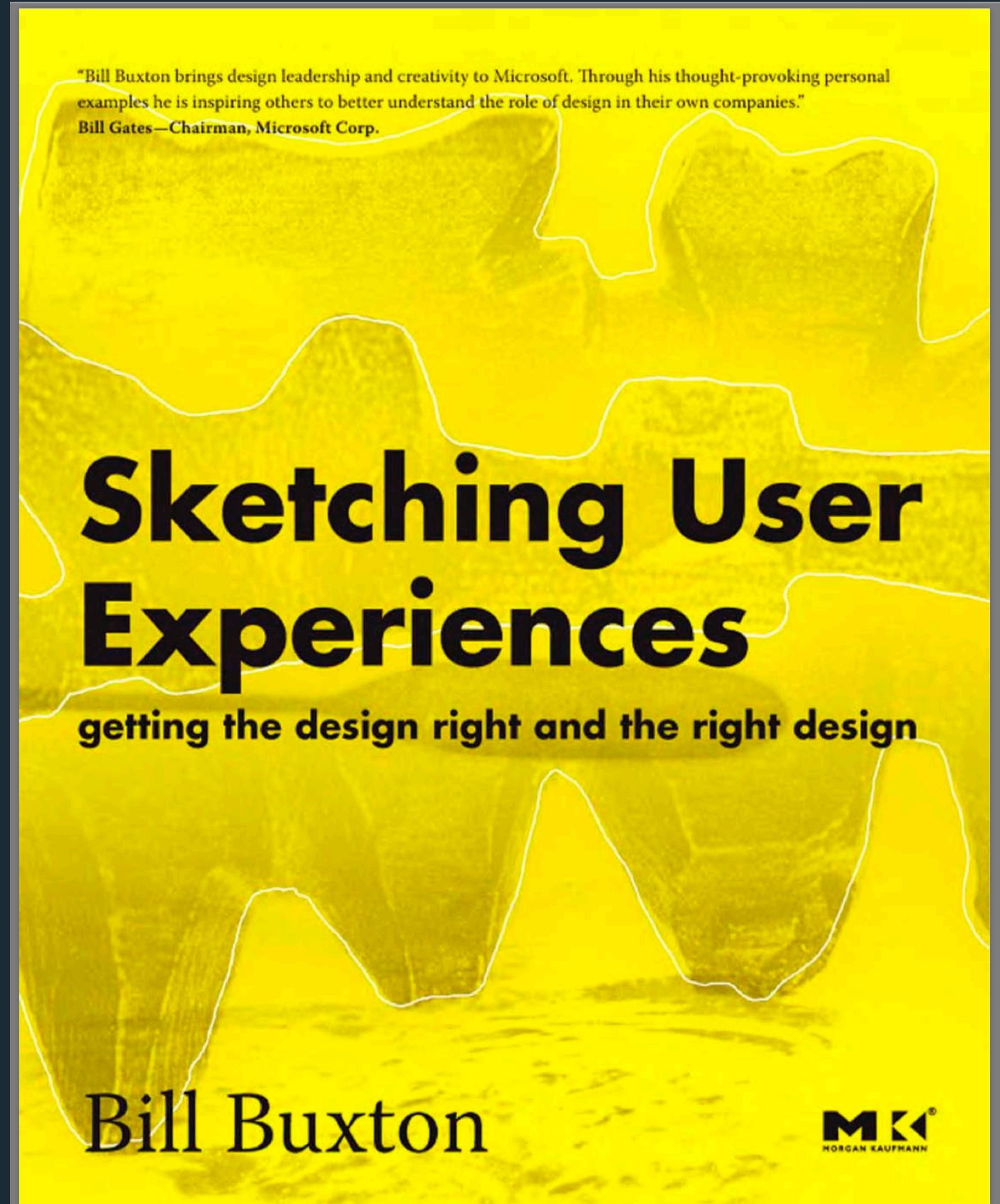
To overcome fixation:

sketch: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas

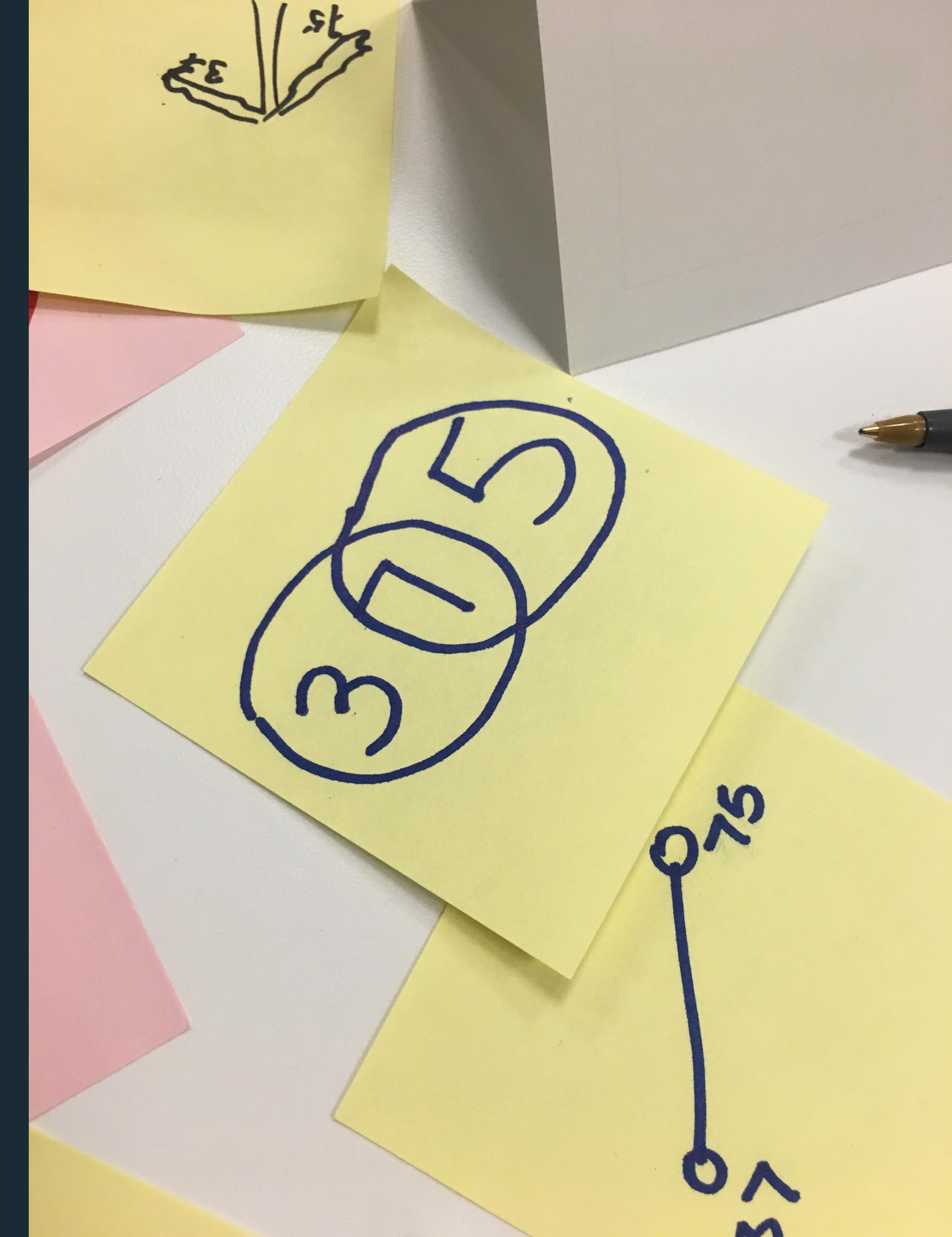
[Buxton 2007]

consult examples: early and repeated exposure to examples improves creativity

[Kulkarni 2012]



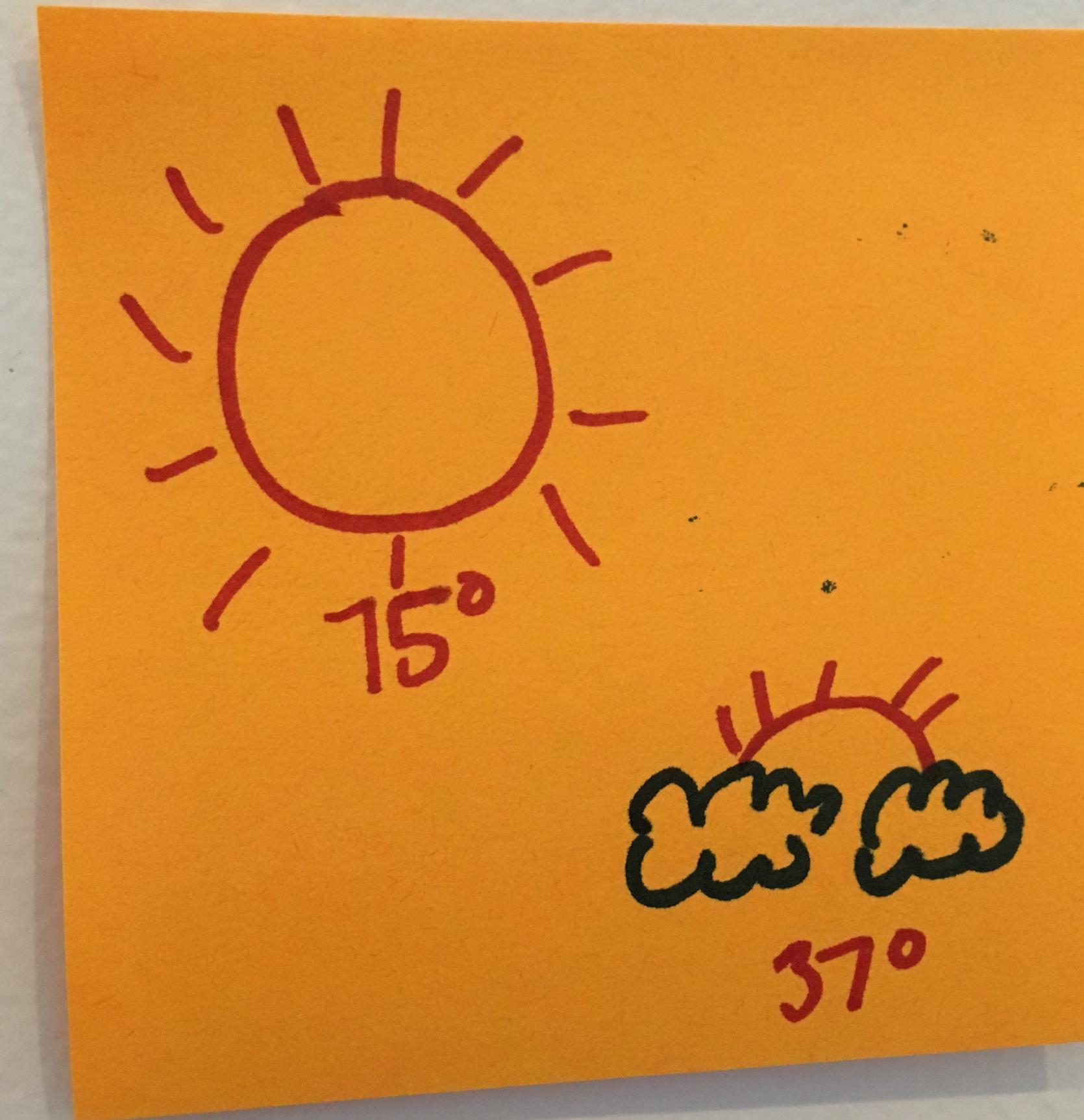
Examples from Jon Schwabish.



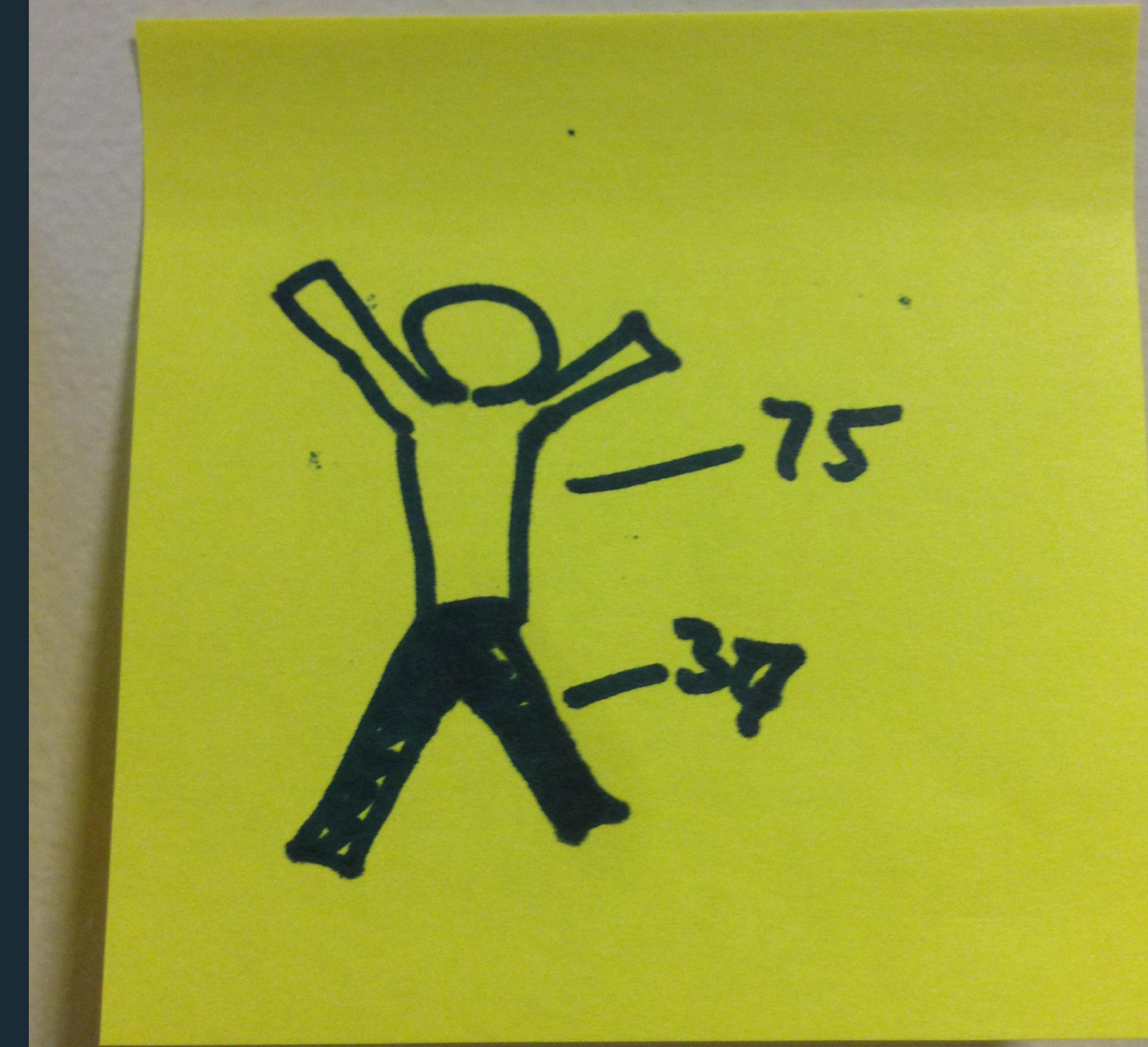
Examples from Jon Schwabish.

3
5
7 } PRIME
#

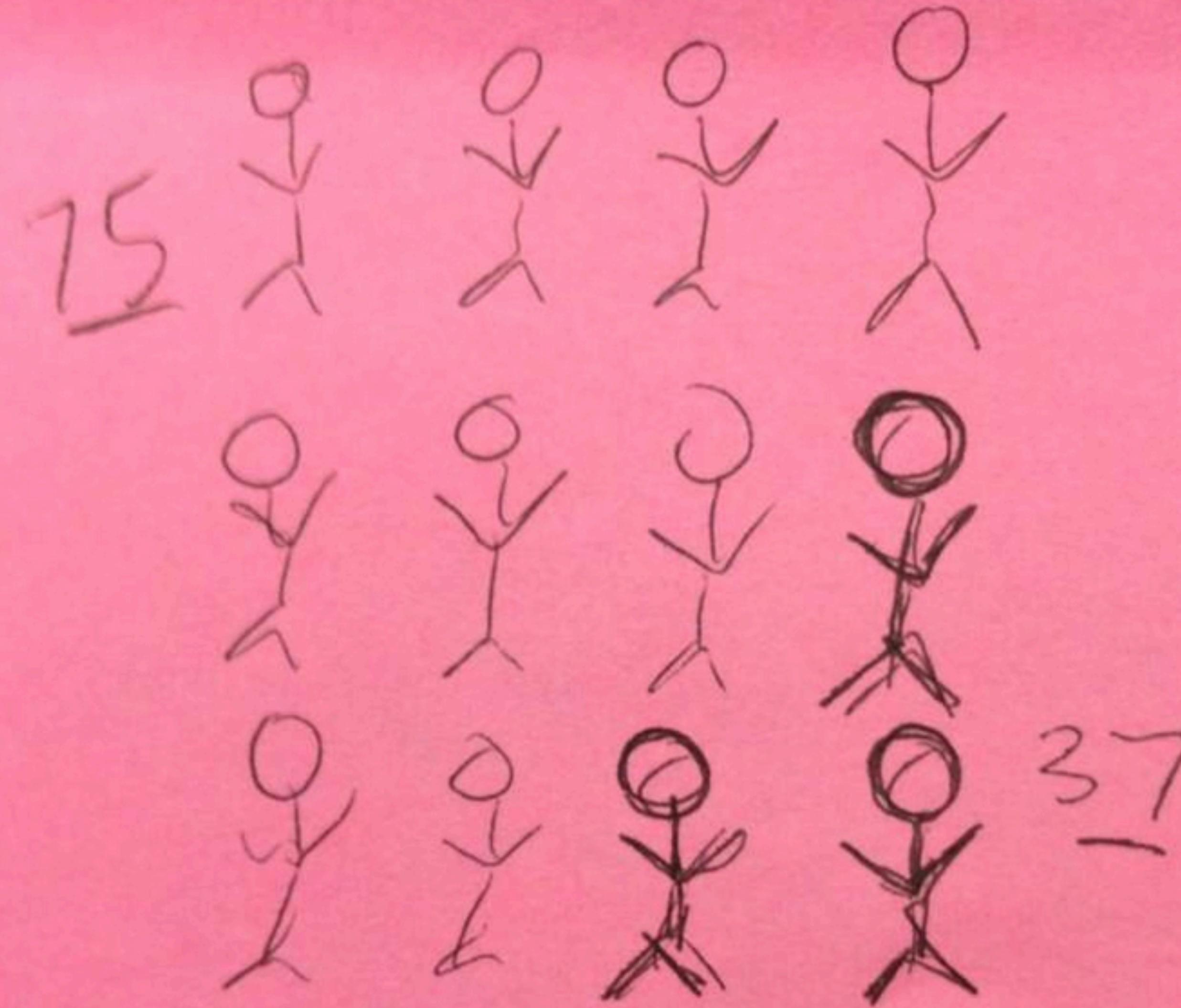
Examples from Jon Schwabish.



Examples from Jon Schwabish.

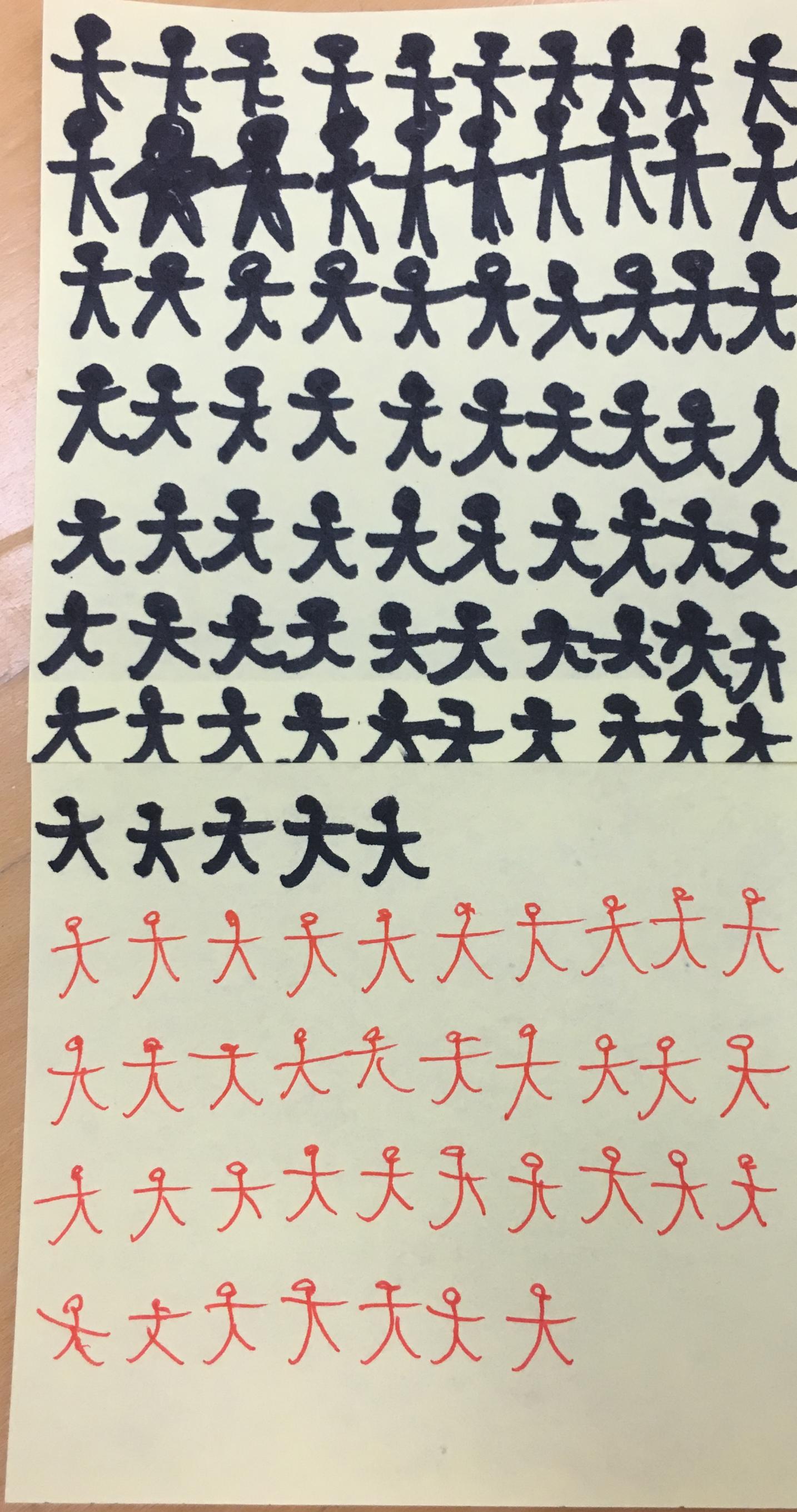


Examples from Jon Schwabish.



NOT TO SCALE .

Examples from Jon Schwabish.



DAIRY
OUTPUT
BY
STATE

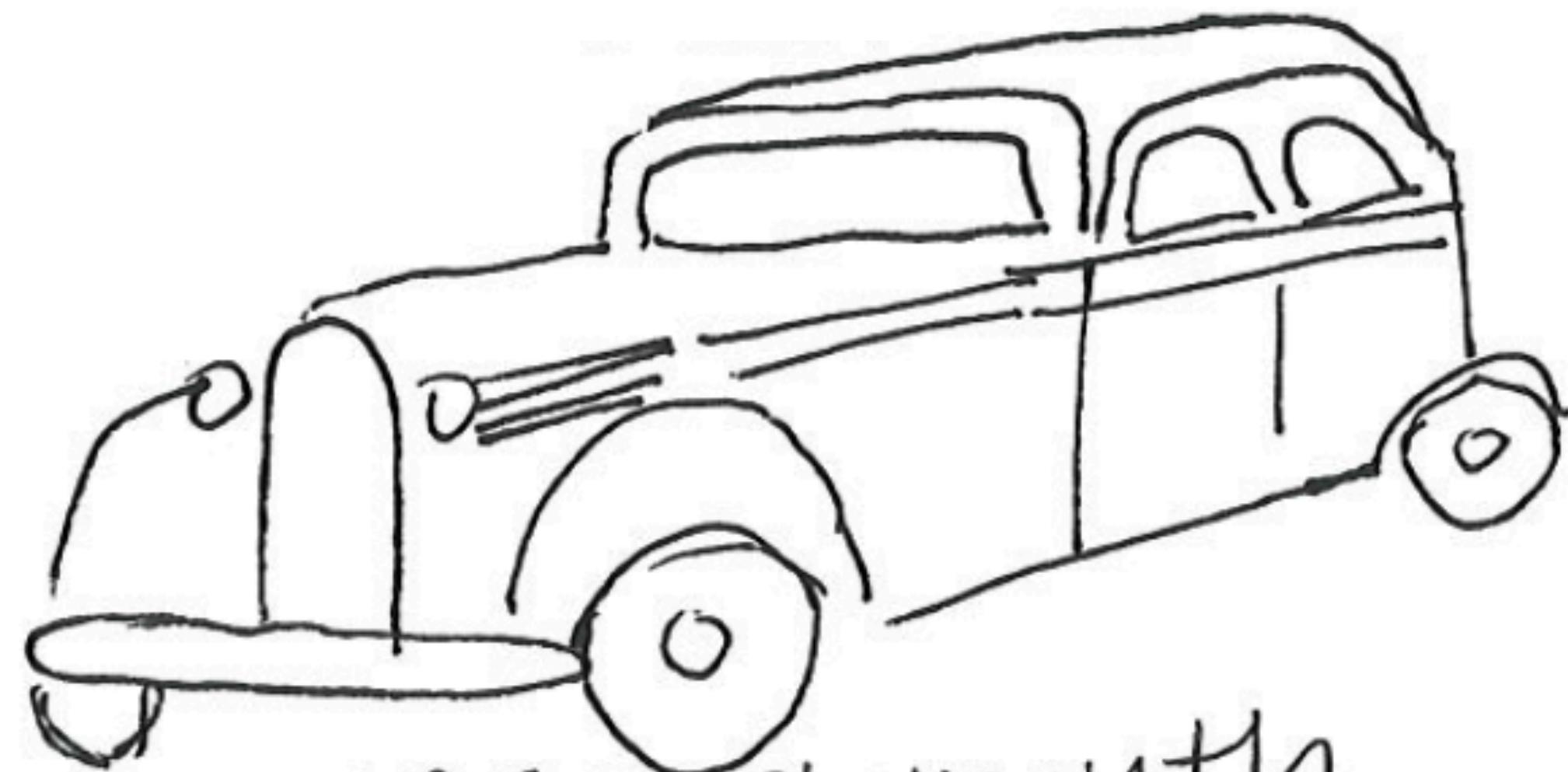
VERMONT



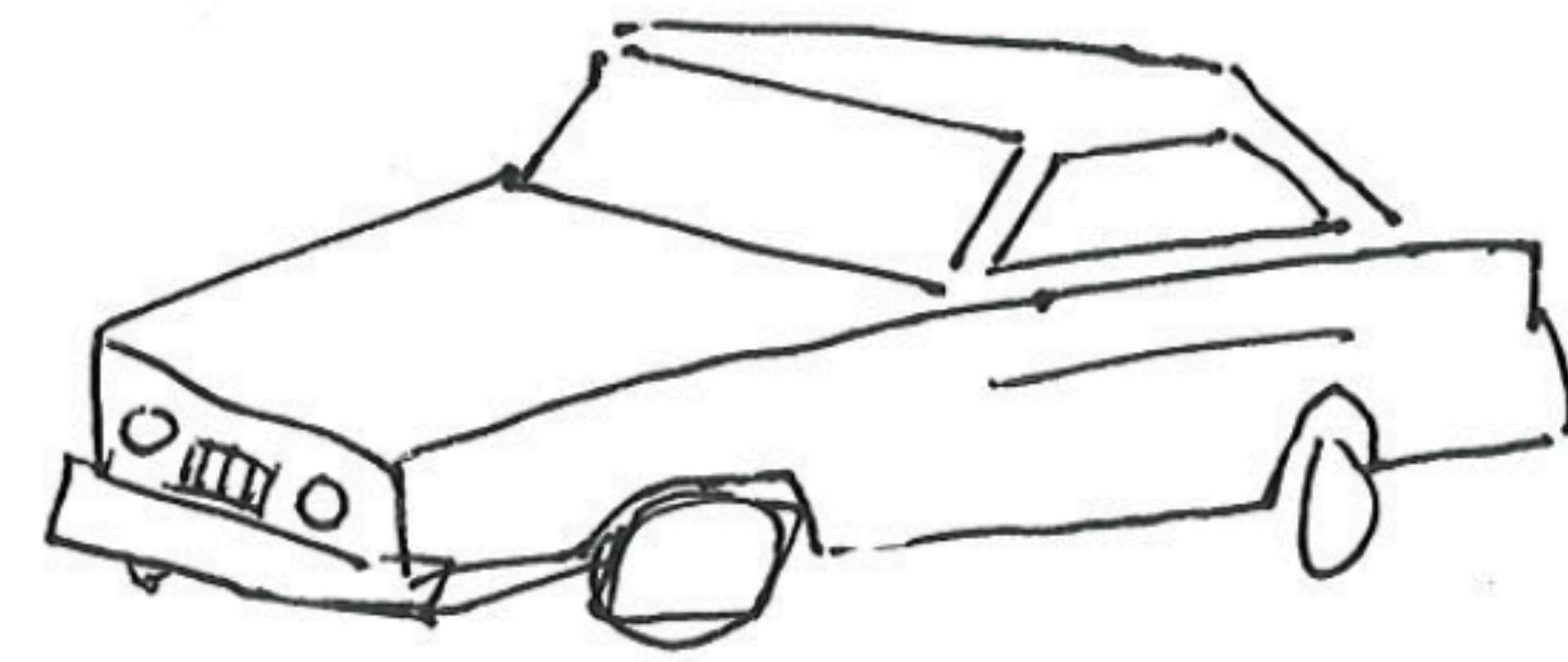
WISCONSIN



Examples from Jon Schwabish.



1937 Plymouth



1975 Plymouth

Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

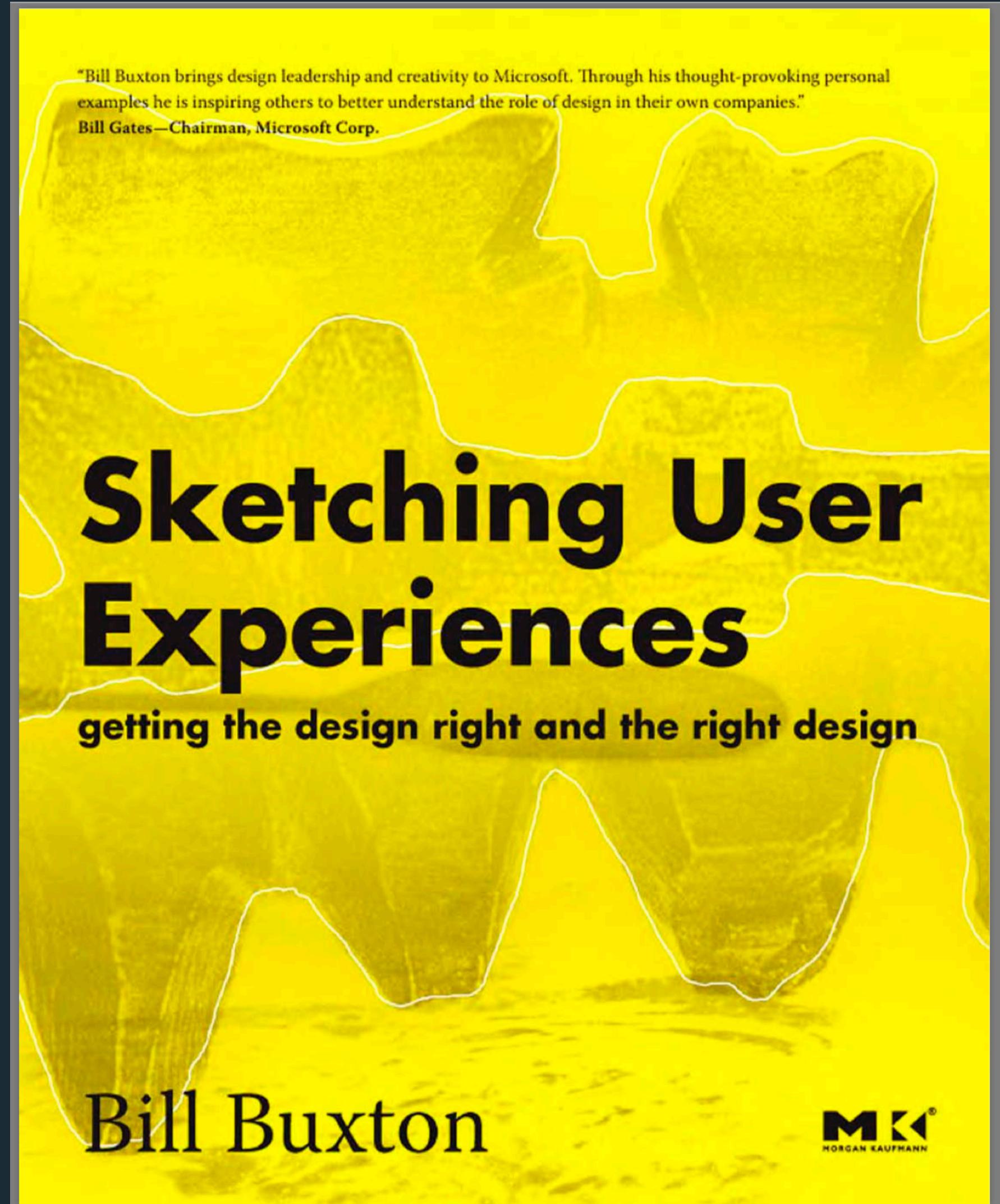
To overcome fixation:

sketch: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas

[Buxton 2007]

consult examples: early and repeated exposure to examples improves creativity

[Kulkarni 2012]



Design Fixation

"A blind adherence to a set of ideas or concepts limiting the output of conceptual design" [Jansson & Smith 1991]

To overcome fixation:

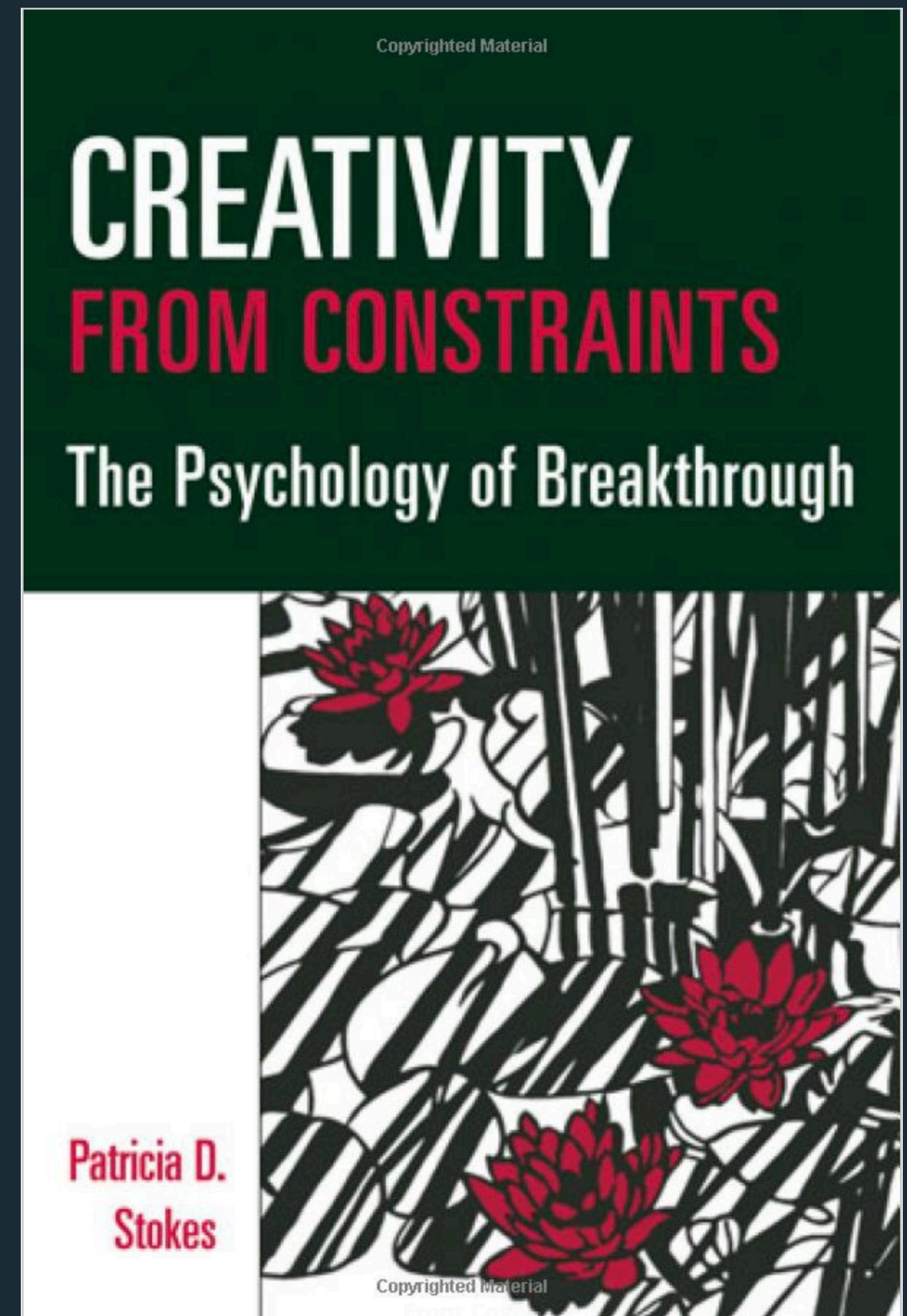
sketch: quick, inexpensive, disposable ways of generating, evaluating, and sharing ideas

[Buxton 2007]

consult examples: early and repeated exposure to examples improves creativity

[Kulkarni 2012]

introduce a constraint: impose new structures to the problem to spur creativity
[Stokes 2006]





Activity!

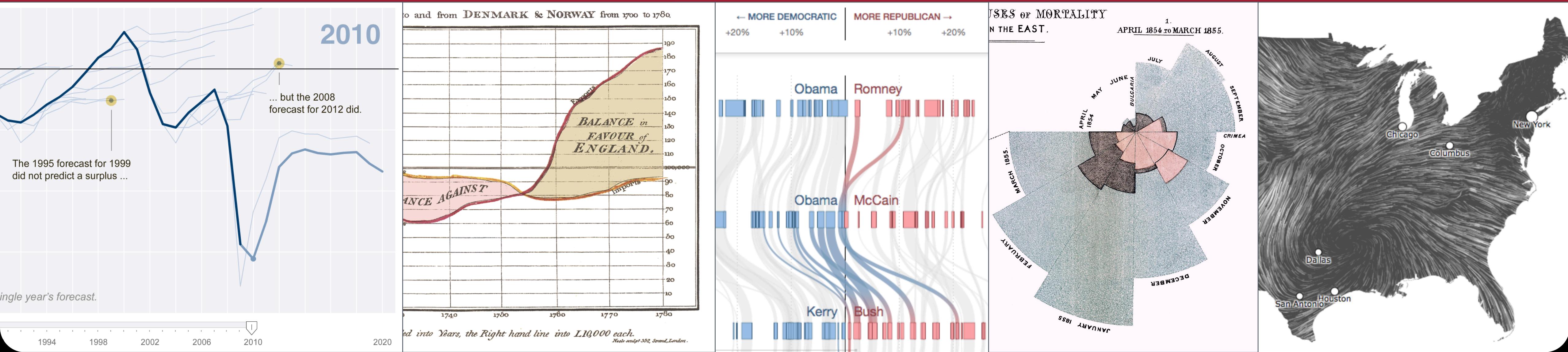
In **5 minutes**, sketch as many **new visualizations** as possible that are different from your previous ideas. If you're stuck, introduce a constraint -- e.g., one line, only black/white, only round objects, etc.

75

37

6.894: Interactive Data Visualization Data & Image Models

Arvind Satyanarayan



Data Visualization

Data → Mapping or Visual Encoding → Visual

Physical Data Types

int, float, string

Conceptual Data Types

temperature, location

Visual **Channels**

x, y, color, opacity

Graphical **Marks**

rect, line, point, area



Expressiveness

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.

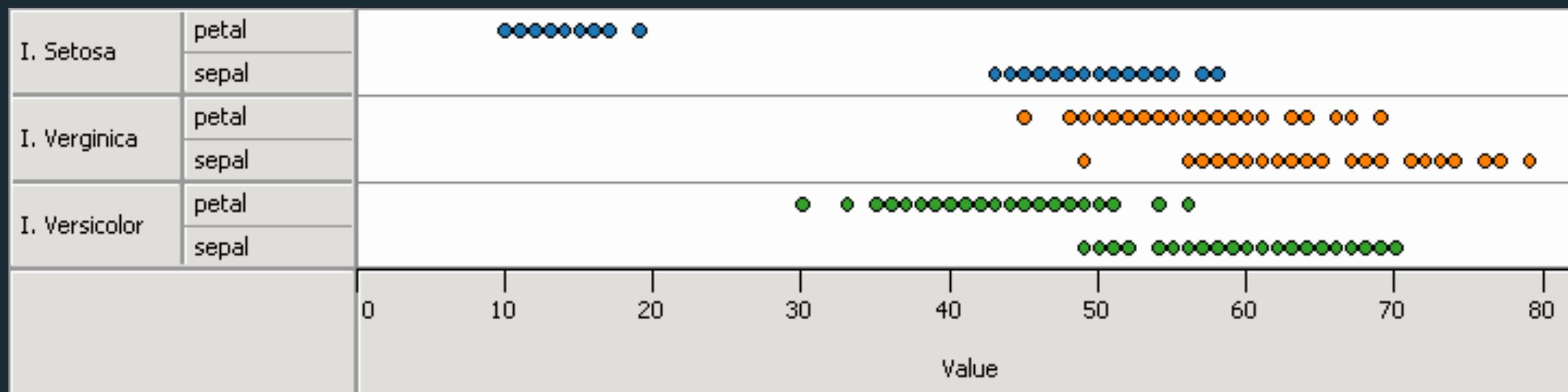
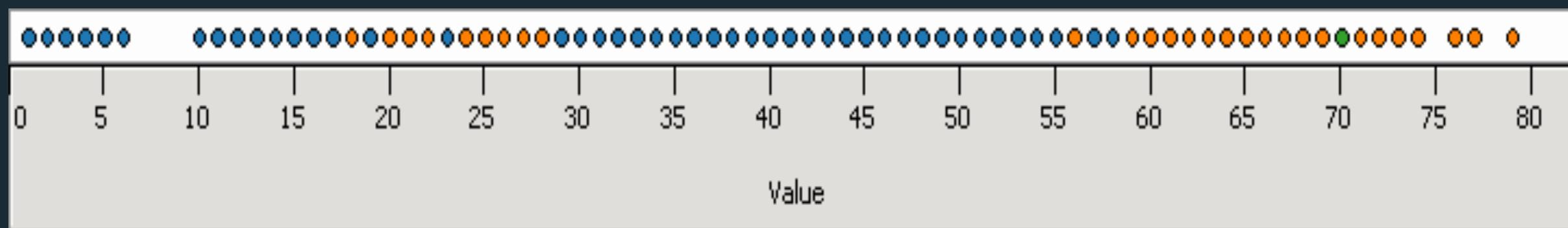
[Mackinlay 1986]



Expressiveness

Cannot express the facts

A multivariate dataset may be *inexpressive* in a single horizontal dot plot because multiple records are mapped to the same position.

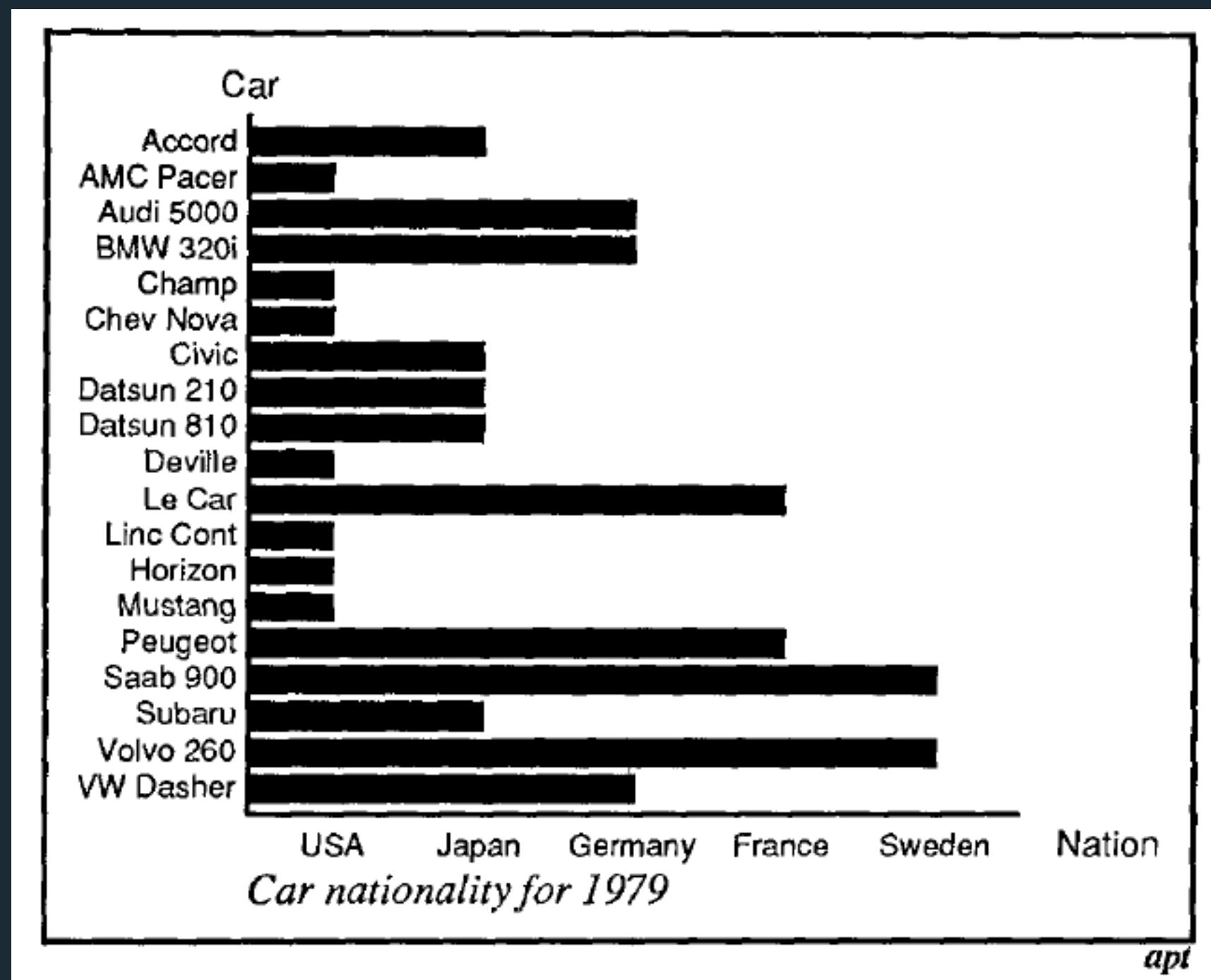


Mapping or Visual Encoding

Data → Visual

Expressiveness

Express facts not in the data



Mapping or Visual Encoding

Data → Visual

Expressiveness

Express facts not in the data

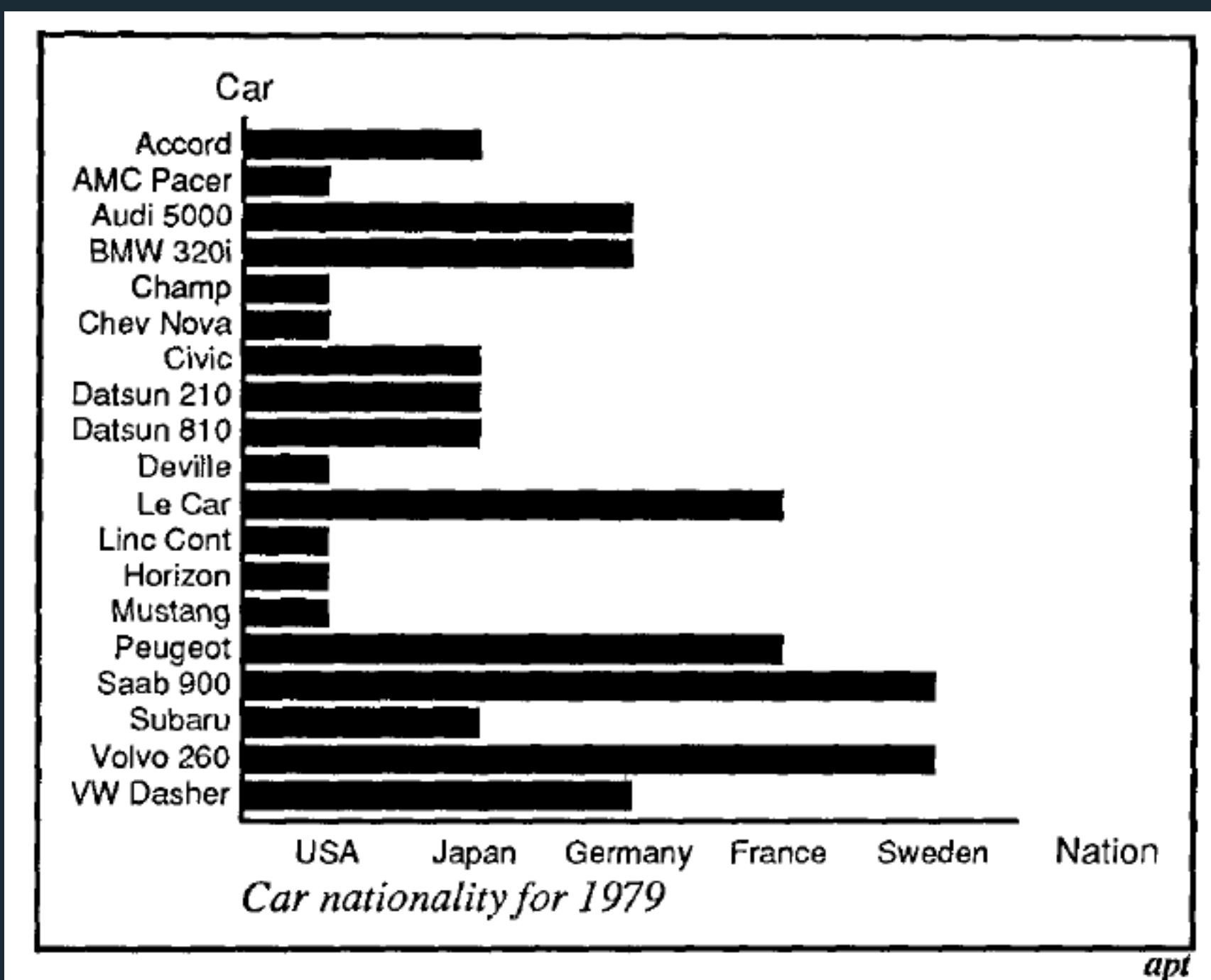


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.



Expressiveness

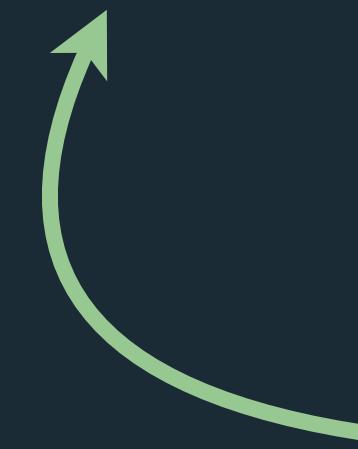
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

[Mackinlay 1986]



Expressiveness

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.*



Data models give us a way of talking about this.

[Mackinlay 1986]

Data Models

Conceptual Models vs Data Models

By "default", data is described in terms of a specific *domain*.

E.g., The average amount of *rain* or *snow* in different *towns, cities, countries*.

E.g., *friends, followers, connections* depending on the social network (or *citations* in academia!).

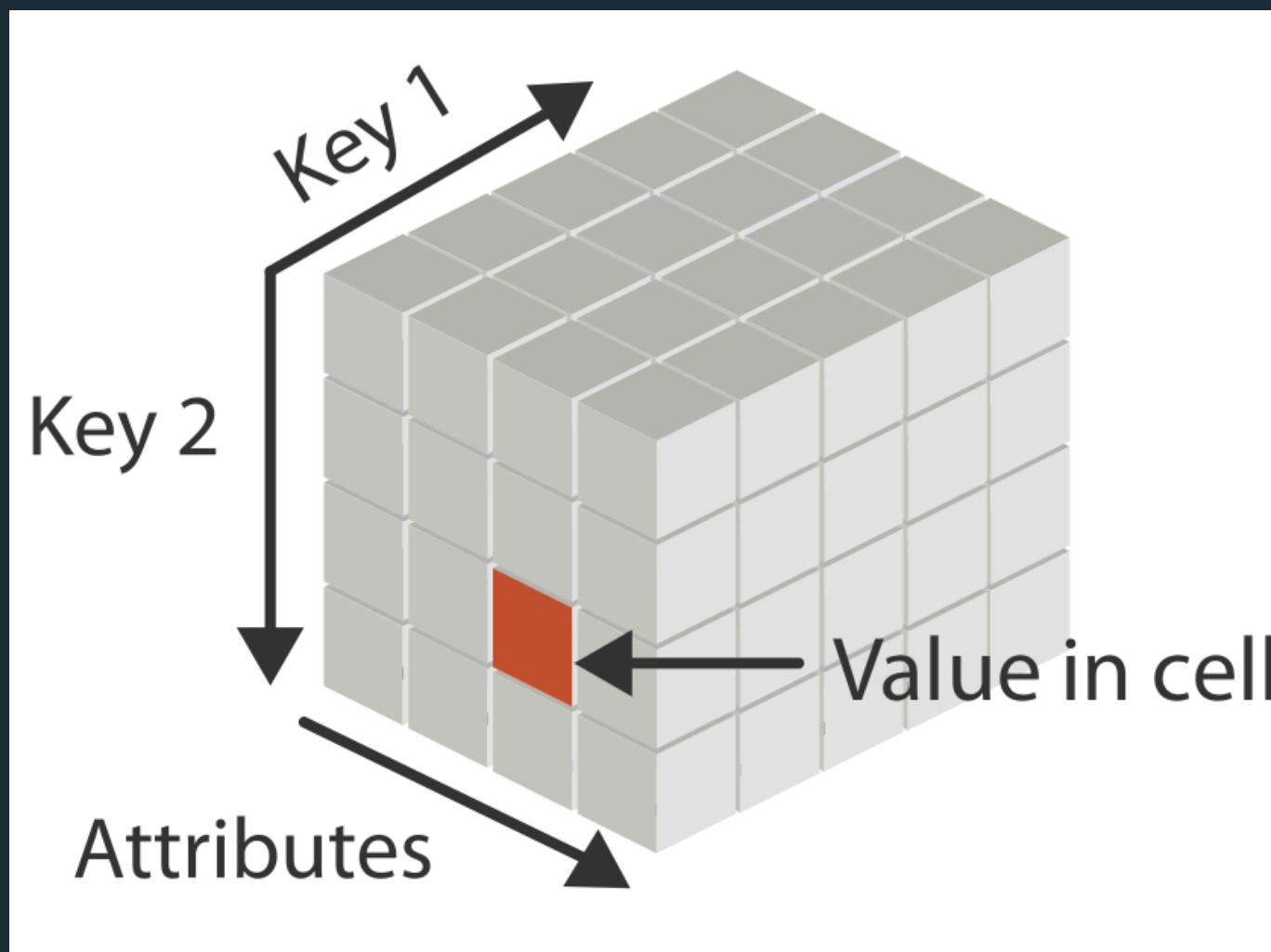
To effectively map data to visuals, we need a level of *abstraction*.

Data abstraction allows us to consistently encode the same "types" of data, even if different domains use different terminology to describe it.

Dataset Types

1. Tabular

rows/records/items



Tamara Munzner, *Visualization Analysis and Design* (2014).

columns/attributes/variables

	A	B	C	D	E	F	G
1	EmployerName	Address	DiffMeanHourlyPercent	DiffMeanBonusPercent	MaleBonusPercent	FemaleBonusPercent	MeanBonusPercent
2	1ST CHOICE STAFF RECRUITMENT LIMITED	8, St. Loyes Street, Bedford, MK40 1EP	-4.5	206.9	2	1	1
3	23.5 DEGREES LIMITED	Charles Watts Way, Hedge End, Southampton,	10	79	4	3	3
4	A. & B. GLASS COMPANY LIMITED	Chilton Industrial Estate, Sudbury, Suffolk,	15	85	61	32	32
5	ABACUS HOTELS LIMITED	20 Station Street, Swaffham, Norfolk,	37.8	-6.6	19.2	16.2	16.2
6	Abbeyfield Wales Society	24 Gold Tops, Newport, NP20 4PG	21.9	0	0	0	0
7	ABERDEEN JOURNALS LIMITED	Mastrick, Aberdeen, United Kingdom,	15.7	44.7	17.1	39.7	39.7
8	ACCESSIBLE TRANSPORT GROUP CONTRACT SERVICES LIMITED	Birmingham, West Midlands, United Kingdom,		0	0	0	0
9	ACEGOLD LIMITED	Norcliffe House, Station Road, Wilmslow, SK9 1BU	-5.1	0	0	0	0
10	Acorns Children's Hospice Trust	Wythall, Birmingham, United Kingdom,	11.2	0	0	0	0
11	AD Astra Academy Trust	Davison Drive, Hartlepool, Cleveland,	9.5	0	0	0	0
12	ADAPT BUSINESS SERVICES LIMITED	Drive, Gorseinon, Swansea, SA4 4QN	3.3	0	0	0	0
13	ADARE INTERNATIONAL LIMITED	Two Colton Square, Leicester, England, Leicester	18.8	71.3	11.6	10.5	10.5

<https://gender-pay-gap.service.gov.uk>

Dataset Types

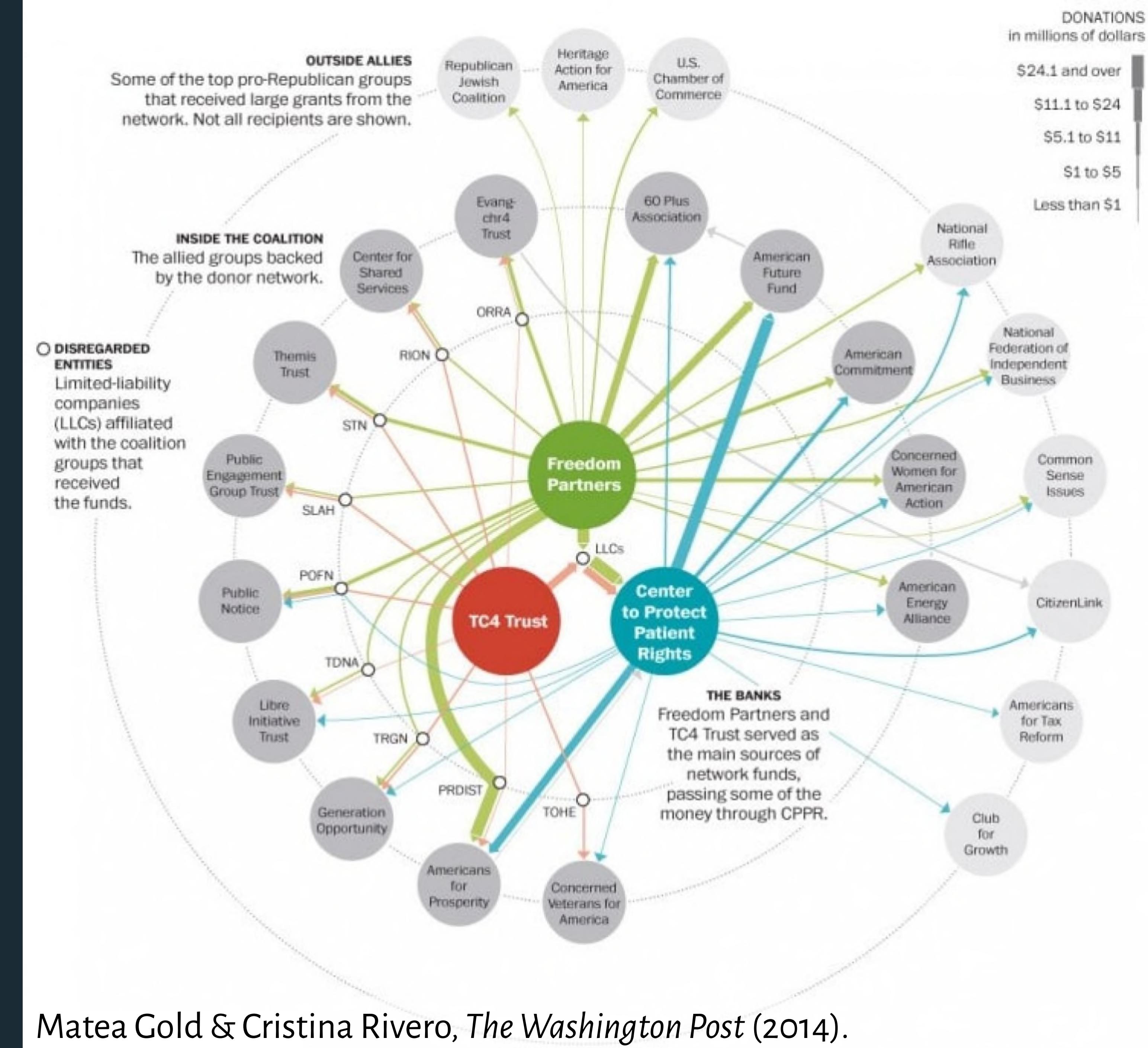
1. Tabular

A collection of records with named attributes.

2. Networks

Nodes and links can also have attributes (e.g., size of nodes, thickness/directionality of links).

Trees are special networks where each node has only one parent.



Dataset Types

1. Tabular

A collection of records with named attributes.

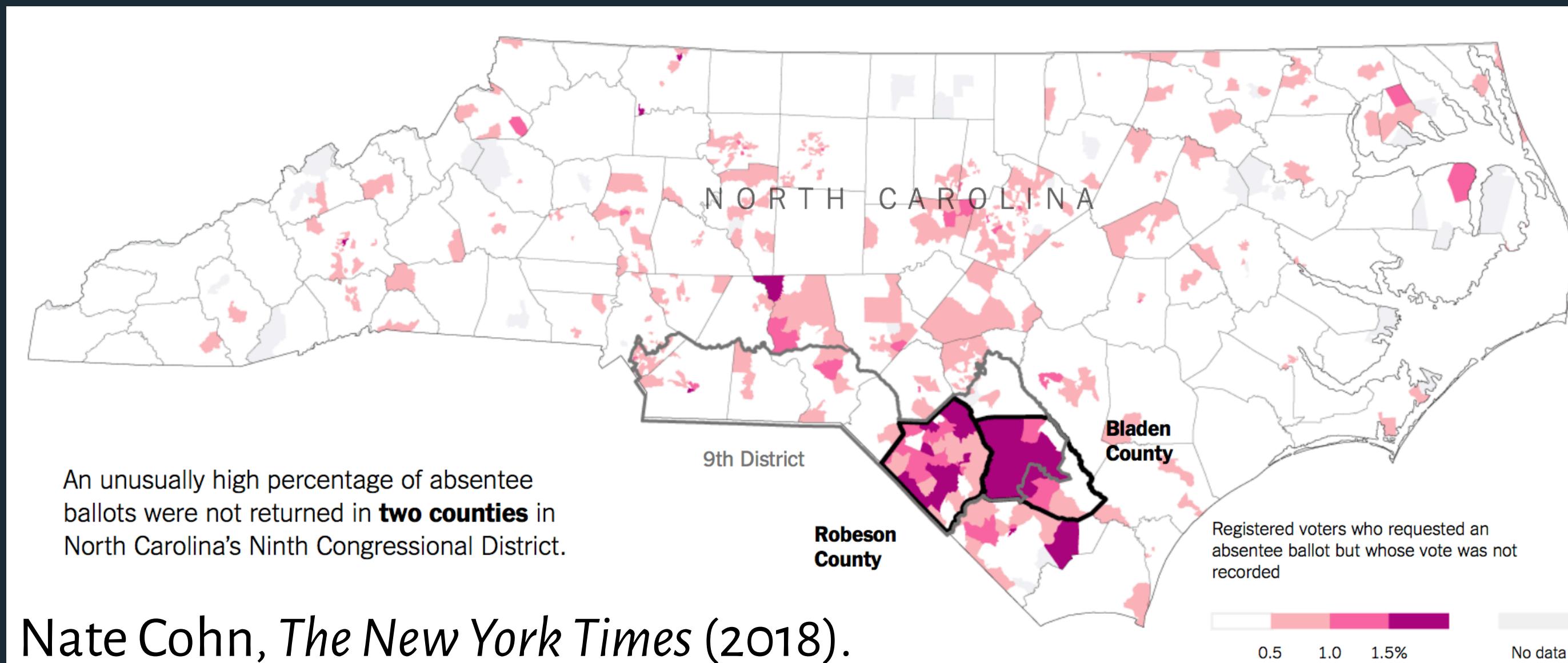
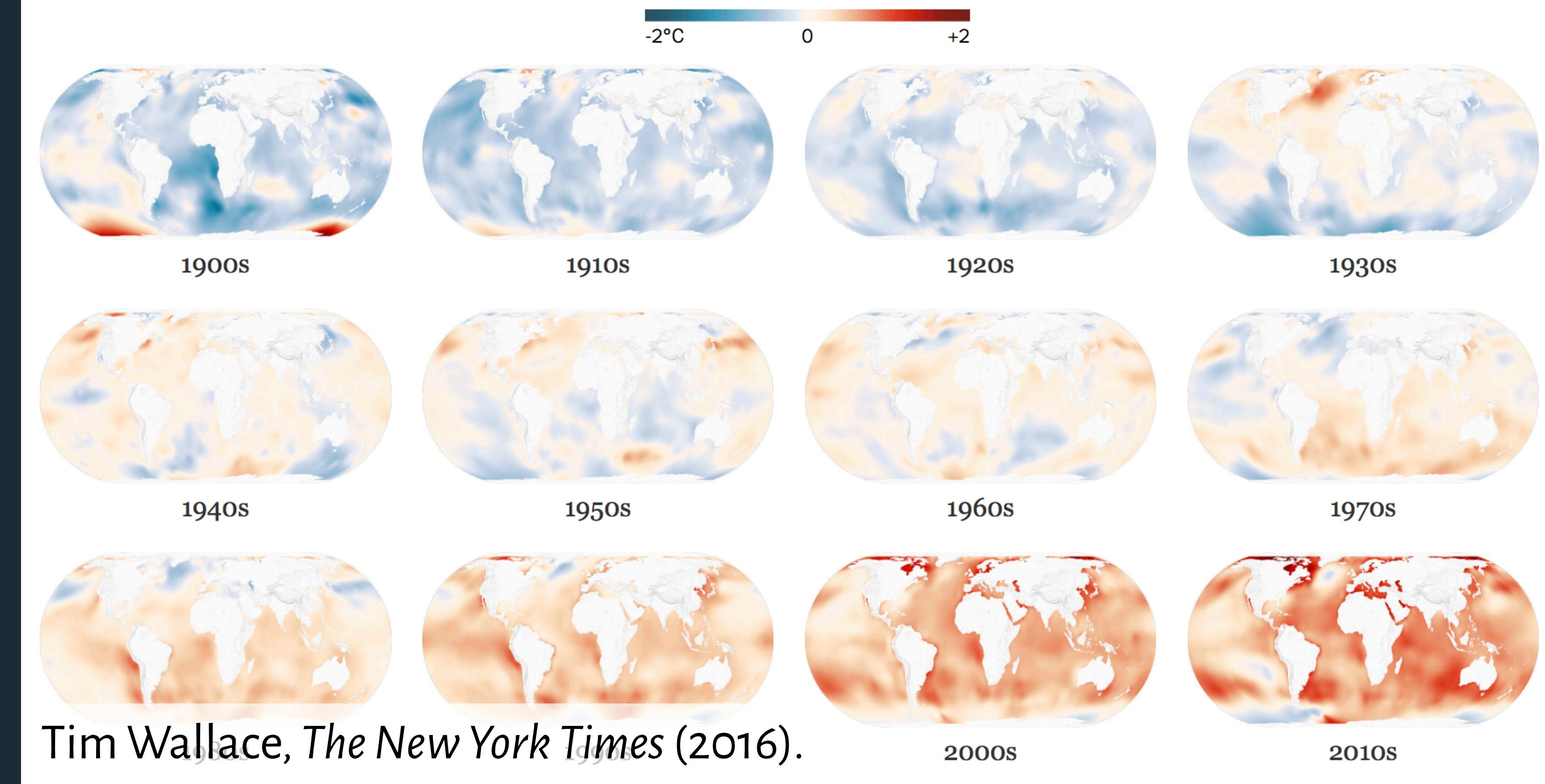
2. Networks

Nodes and links can also have attributes (e.g., size of nodes, thickness/directionality of links).

Trees are special networks where each node has only one parent.

3. Spatial

Continuous "fields" vs discrete "positions"



Attribute Types

Dimensions

~ Independent variables.

Ways of describing the data, often discrete.

E.g., categories, dates, binned quantities.

Can include numerical data, but doesn't make sense to aggregate.

Measures

~ Dependent variables (i.e., their value is a function of one or more dimensions).

Numerical data that can be analyzed and aggregated.

Aggregations including sum, count, avg, std. dev, etc.

Dimensions

- =Abc Above Threshold?
- =Abc Birth Rate Bin
- 🌐 Country
- 🌐 Ease of Business (clusters)
- .dll. GDP per Capita (bin)
- Abc Region

Measures

- Business**
 - # Business Tax Rate
 - # Days to Start Business
 - # Ease of Business
 - # Hours to do Tax
 - # Lending Interest
- Development**
 - # CO2 Emissions
 - # Energy Usage
 - # GDP
 - # Internet Usage
 - # Mobile Phone Usage
 - # Tourism Inbound
 - # Tourism Outbound
- Health**
 - # Health Exp % GDP
 - # Health Exp/Capita
 - # Infant Mortality Rate
 - # Life Expectancy
 - # Life Expectancy Female
 - # Life Expectancy Male
- Population**
 - # Birth Rate

Parameters

- # Health Exp Threshold

Pages

Measure Names

Country

Health Indicators

	Birth Rate	Infant Mortality	Health Exp % GDP	Life Expectancy
Ethiopia	3.8%	0.07	4.4%	58
Malawi	4.2%	0.07	7.4%	50
Eritrea	3.9%	0.05	3.3%	59
Congo, Dem. Rep.	4.6%	0.10	5.8%	48
Niger	5.1%	0.08	7.2%	55
Madagascar	3.7%	0.05	4.7%	62
Mozambique	4.3%	0.09	6.1%	48
Rwanda	3.8%	0.07	8.0%	56
Uganda	4.6%	0.07	8.4%	54
Afghanistan	4.3%	0.08	9.2%	58
Sierra Leone	4.1%	0.13	15.6%	42
Central African Republic	3.7%	0.11	4.1%	46
Guinea	4.0%	0.08	5.8%	53
Tanzania	4.1%	0.06	5.1%	55
Nepal	2.7%	0.05	5.7%	65
Togo	3.8%	0.07	6.5%	55
Guinea-Bissau	4.0%	0.09	5.9%	53
Burkina Faso	4.4%	0.08	6.5%	53
Tajikistan	3.1%	0.05	5.1%	66
Bangladesh	2.3%	0.05	3.3%	68
Mali	4.8%	0.10	6.5%	52
Gambia, The	4.4%	0.06	4.4%	57
Haiti	2.8%	0.07	5.9%	60
Zimbabwe	3.2%	0.06		48
Benin	4.0%	0.07	4.5%	57
Cambodia	2.6%	0.05	6.1%	68
Kenya	3.8%	0.06	4.5%	56
Chad	4.9%	0.10	4.8%	48
Timor-Leste	3.8%	0.06	6.9%	64
Kyrgyz Republic	2.3%	0.03	6.0%	69
Comoros	2.8%	0.07	4.0%	50

Birth Rate

Infant Mortality Rate

Health Exp % GDP

Life Expectancy

Attribute Types

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

1, 2, 3, 4

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).

1, 2, 3, 4, 5

E.g., Date: Jan 10, 2018; Location: Lat 42.35, Lon -71.09

1, 2, 3, 4, 5, 6, 7

Only differences can be utilized (e.g., distances or spans)

Quantitative (Ratio)

Ratio (fixed zero).

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

E.g., Physical measurement: length, mass, temperature

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Counts and amounts. Can measure ratios or proportions.

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative: Interval (zero can be arbitrarily located).

E.g., Dates Jan 10, 2018, location (lat, lon), 71, 00

Quantitative: Ratio (fixed zero). Only differences can be compared (e.g., distance or spans).

Quantitative: Ratio (fixed zero).

E.g., Physical measurement length, mass, temperature

Counts and amounts. Can measure ratios or proportions.

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

Ratio (fixed zero).

E.g., Physical measurement: length, mass, temperature

Counts and amounts. Can measure ratios of proportions.

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

Ratio (fixed zero).

E.g., Physical measurement: length, mass, temperature

Counts and amounts. Can measure ratios or proportions.

Attribute Types

Nominal

=, ≠

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

=, ≠, <, >

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative

(Interval)

=, ≠, <, >, –

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative

(Ratio)

=, ≠, <, >, –, %

Ratio (fixed zero).

E.g., Physical measurement: length, mass, temperature

Counts and amounts. Can measure ratios or proportions.

Data Models

Physical Model

32.5, 54.0, -17.3, ...
Floating point numbers

Conceptual Model

Temperature (°C)

Attribute Type

Burned vs. Not-Burned (N)
Hot, Warm, Cold (O)
Temperature Value (Q)

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: # of people in group

Year: 1850 – 2000 (every decade)

Age: 0 – 90+

Sex: Male, Female

Marital Status: Single, Married,
Divorced, ...

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count:

Year:

Age:

Sex:

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Q-Ratio

Year:

Age:

Sex:

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Q-Ratio

Year:

Age:

Sex: Nominal

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Q-Ratio

Year:

Age:

Sex: Nominal

Marital Status: Nominal

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Q-Ratio

Year: Q-Interval or Ordinal

Age:

Sex: Nominal

Marital Status: Nominal

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Q-Ratio

Year: Q-Interval or Ordinal

Age: Q-Ratio or Ordinal

Sex: Nominal

Marital Status: Nominal

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count:

Year:

Age:

Sex:

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Measure

Year:

Age:

Sex:

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Measure

Year: Dimension

Age:

Sex:

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Measure

Year: Dimension

Age:

Sex: Dimension

Marital Status:

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Measure

Year: Dimension

Age:

Sex: Dimension

Marital Status: Dimension

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Activity: U.S. Census

What are the types of these attributes
(N/O/Q and dimension/measure)?

Think (~1 min), Pair (~2 mins), Share.

People Count: Measure

Year: Dimension

Age: Depends!

Sex: Dimension

Marital Status: Dimension

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343

Data Transformation

Data Transformation

Sorting

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45
10/3	MSFT	74.26
10/4	MSFT	74.69

Data Transformation

Sorting

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45
10/3	MSFT	74.26
10/4	MSFT	74.69

Filtering (price > 100)

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45

Data Transformation

Sorting

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45
10/3	MSFT	74.26
10/4	MSFT	74.69

Filtering (price > 100)

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45

Aggregation (group by, sum, min, max, ...)

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



stock	min(price)
AMZN	965.45
MSFT	74.26

Data Transformation

Sorting

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45
10/3	MSFT	74.26
10/4	MSFT	74.69

Filtering (price > 100)

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price
10/3	AMZN	957.10
10/4	AMZN	965.45

Aggregation (group by, sum, min, max, ...)

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



stock	min(price)
AMZN	965.45
MSFT	74.26

Join

day	stock	price
10/3	AMZN	957.10
10/3	MSFT	74.26
10/4	AMZN	965.45
10/4	MSFT	74.69



day	stock	price	min
10/3	AMZN	957.10	965.45
10/3	MSFT	74.26	74.26
10/4	AMZN	965.45	965.45
10/4	MSFT	74.69	74.26

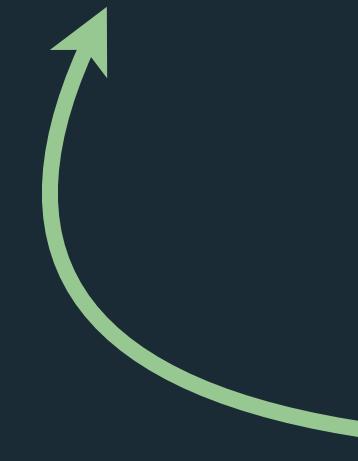


stock	min
AMZN	965.45
MSFT	74.26



Expressiveness

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.*



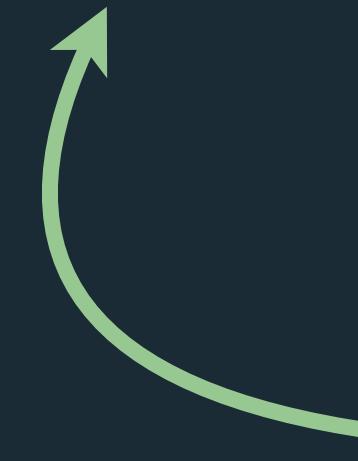
Data models give us a way of talking about this.

[Mackinlay 1986]



Expressiveness

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.*



Data models give us a way of talking about this.

Effectiveness

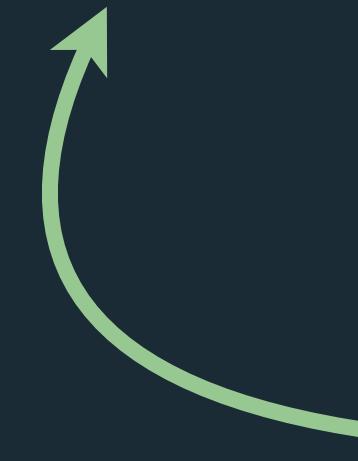
A visualization is more *effective* than another if the information it conveys is more readily perceived than the information in the other visualization.

[Mackinlay 1986]



Expressiveness

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.*



Data models give us a way of talking about this.

Effectiveness

A visualization is more *effective* than another if the information it conveys is *more readily perceived* than the information in the other visualization

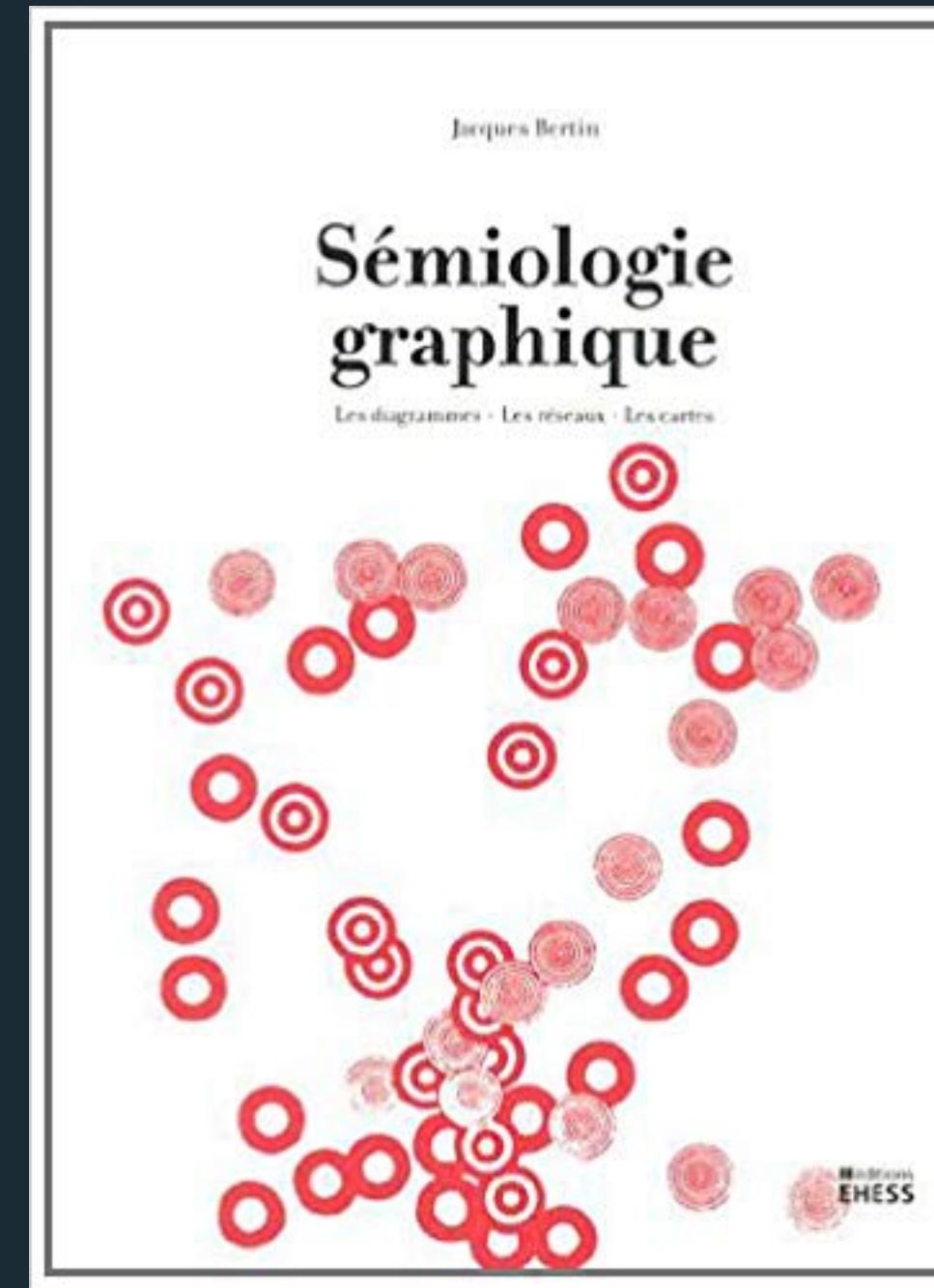


Image models give us a way of talking about this.

[Mackinlay 1986]

Image Models

The Semiology of Graphics (1967)

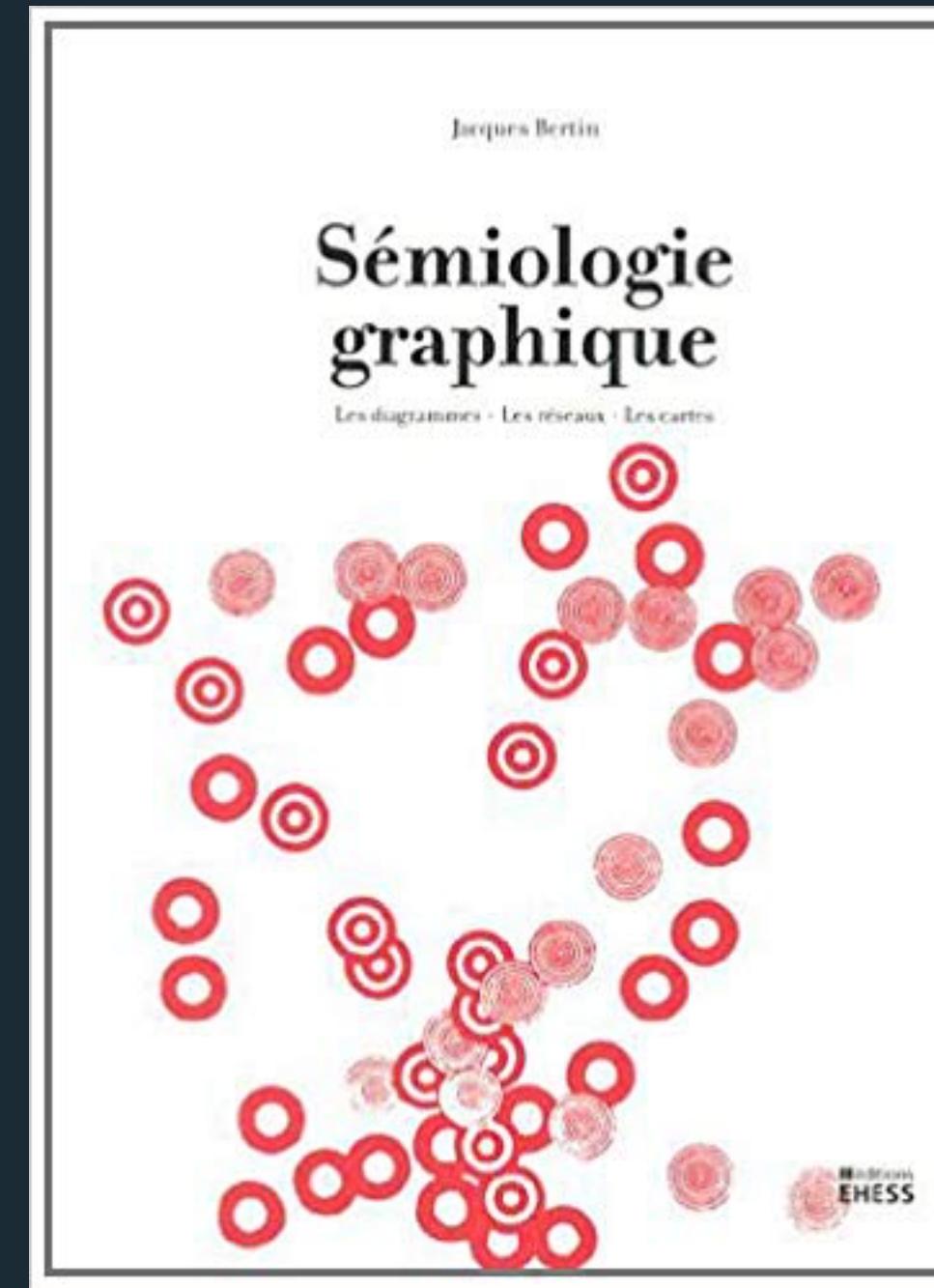


Jacques Bertin (1918 – 2010)
French cartographer

The **Semiology** of Graphics (1967)



Study of signs and how cultures use them.



Jacques Bertin (1918 – 2010)
French cartographer

The **Semiology** of Graphics (1967)



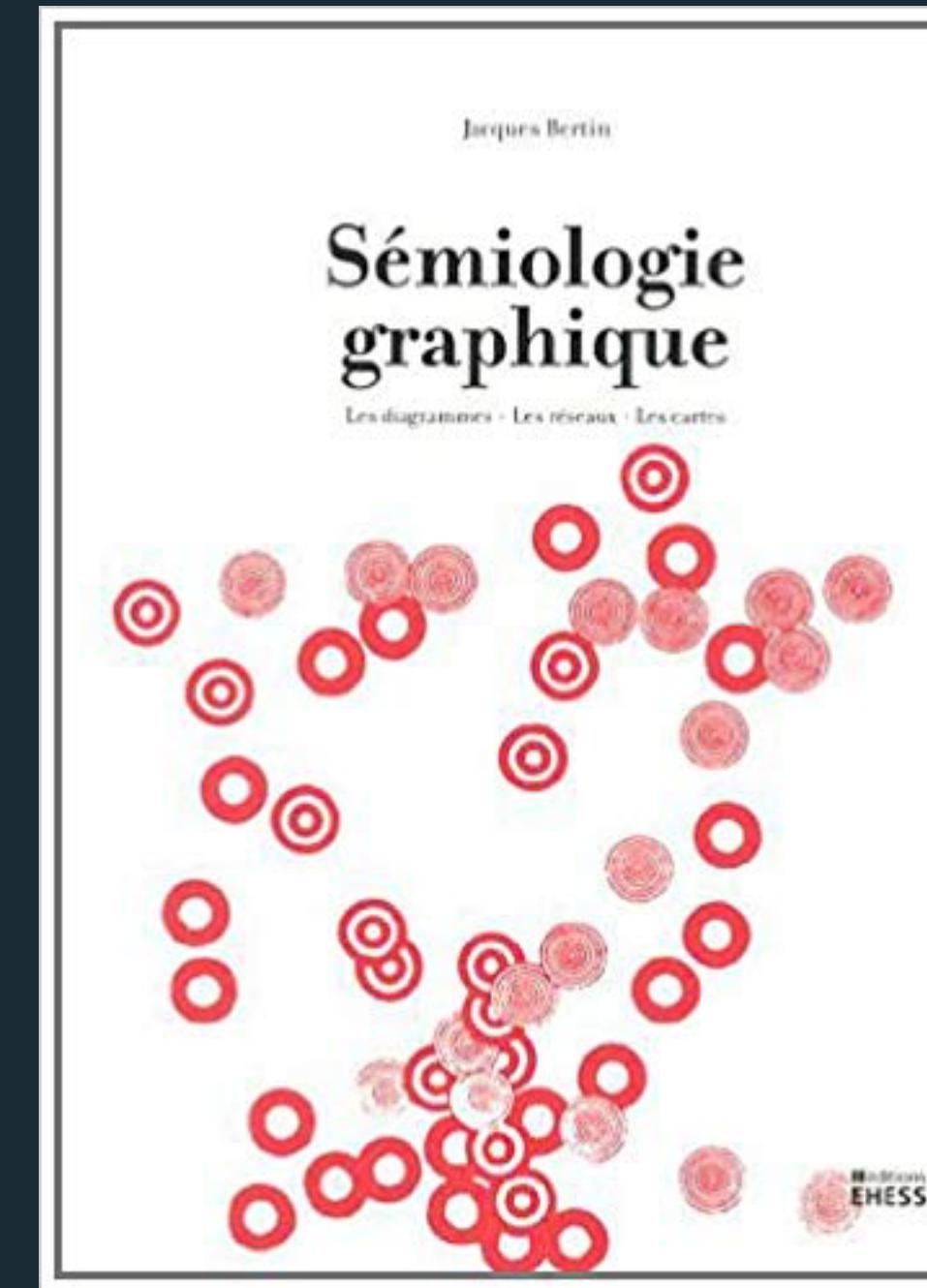
*Study of **signs** and how cultures use them.*



Anything that stands for something other than itself.



Jacques Bertin (1918 – 2010)
French cartographer

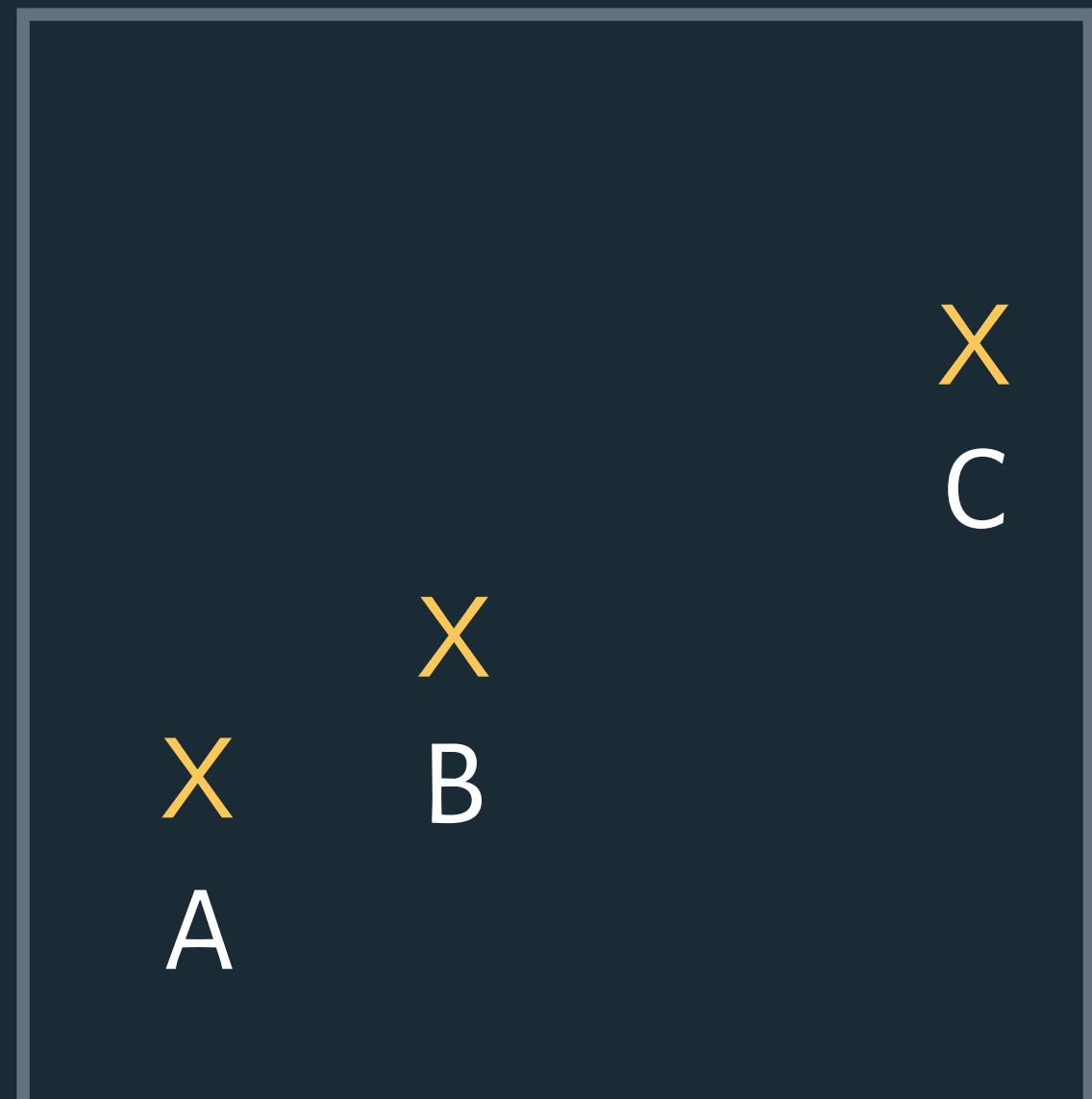


Images are perceived as a set of signs.

Sender encodes information in signs.

Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?



Sender encodes information in signs.

Through visual perception, the receiver decodes the signs for information:

1. What are the elements in question?
2. What are the relationships between them?

What do these signs signify?

1. A, B, C are distinguishable.
2. B is between A and C.
3. BC is twice as long as AB.



"Resemblance, order, and proportionality are the three signfields in graphics."

—Bertin

Visual Variables

Also called visual *channels*.

Used to encode data values as characteristics of marks.

* From 1967, so Bertin only accounted for visualizations that were printable, white paper.

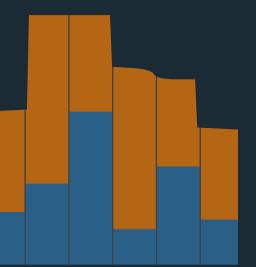
LES VARIABLES DE L'IMAGE						
	POINTS	LIGNES	ZONES			
XY 2 DIMENSIONS DU PLAN	x	x	x			
Z						
TAILLE						
VALEUR						
LES VARIABLES DE SÉPARATION DES IMAGES						
GRAIN						
COULEUR						
ORIENTATION						
FORME						

Marks

Basic graphical elements that represent data items.



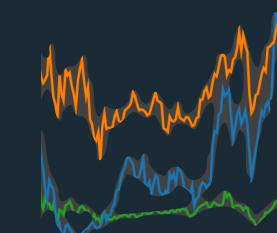
Area



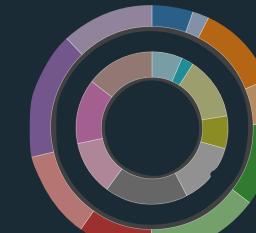
Bar



Point



Line



Arc



Text

Channels: Expressiveness Types and Effectiveness Ranks

→ Magnitude Channels

Position on common scale



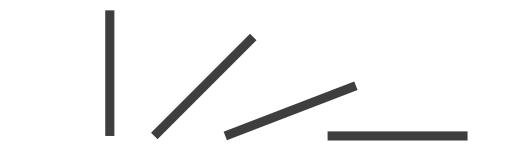
Position on unaligned scale



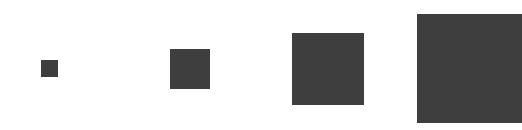
Length (1D size)



Tilt angle



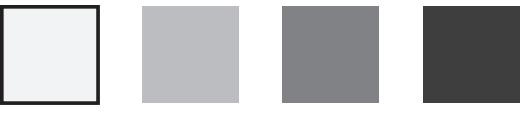
Area (2D size)



Depth (3D position)



Color luminance



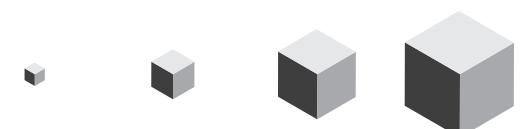
Color saturation



Curvature



Volume (3D size)



→ Identity Channels

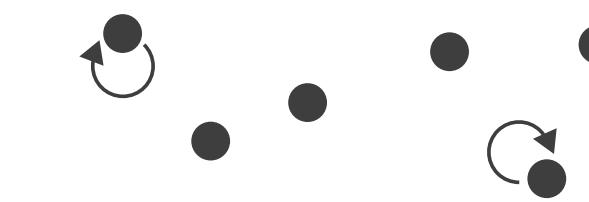
Spatial region



Color hue



Motion



Shape



Channels: Expressiveness Types and Effectiveness Ranks

→ Magnitude Channels: O or Q attributes

Position on common scale



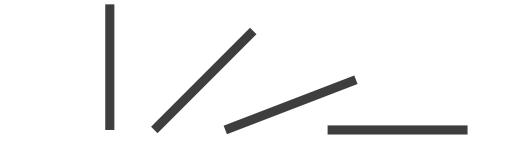
Position on unaligned scale



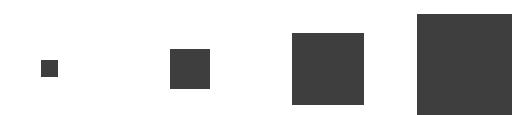
Length (1D size)



Tilt angle



Area (2D size)



Depth (3D position)



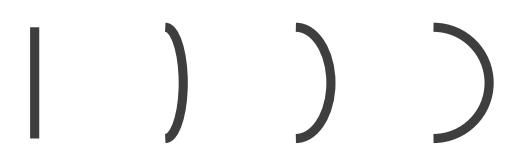
Color luminance



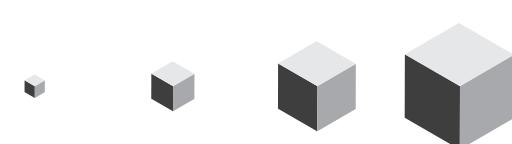
Color saturation



Curvature



Volume (3D size)



→ Identity Channels: N attributes

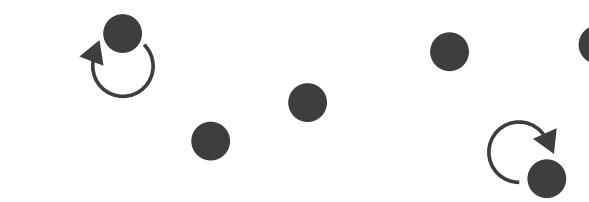
Spatial region



Color hue



Motion



Shape



Channels: Expressiveness Types and Effectiveness Ranks

→ Magnitude Channels: O or Q attributes



→ Identity Channels: N attributes

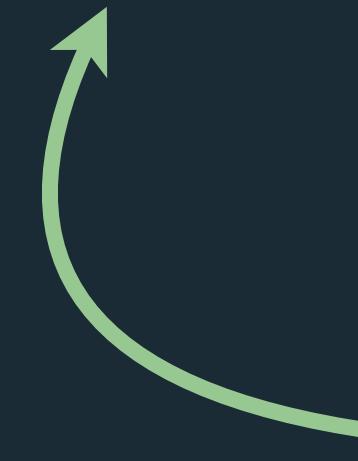


Tamara Munzner, *Visualization Analysis and Design* (2014).



Expressiveness

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.*



Data models give us a way of talking about this.

Effectiveness

A visualization is more *effective* than another if the information it conveys is *more readily perceived* than the information in the other visualization



Image models give us a way of talking about this.

[Mackinlay 1986]

Activity!

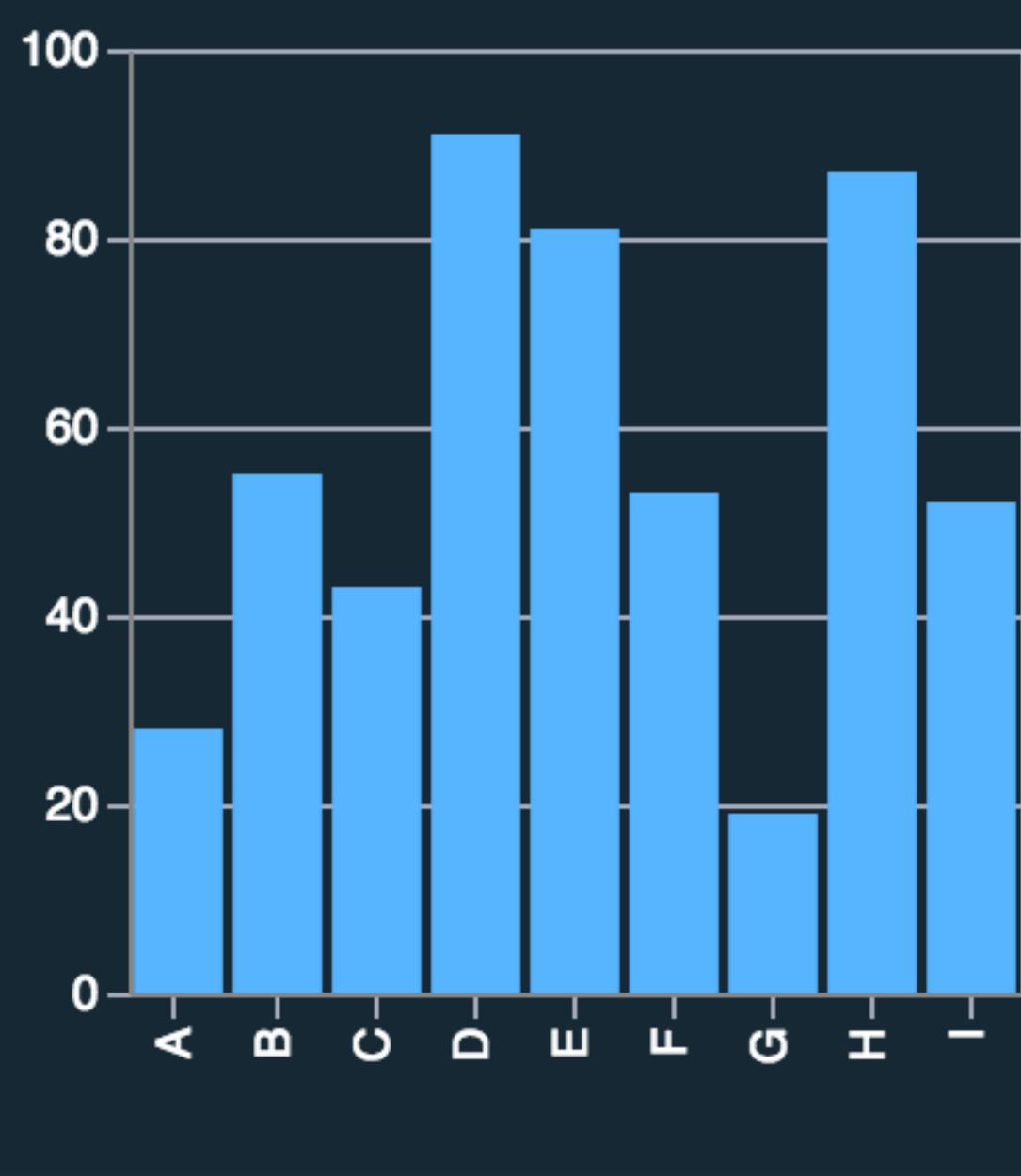
Analyze which channels and marks you've used for the sketches so far. Create at least **3 new sketches** using different channels/marks. **3 minutes.**

75

37

Visual Encoding

Visual Encoding: 1 Nominal, 1 Quantitative

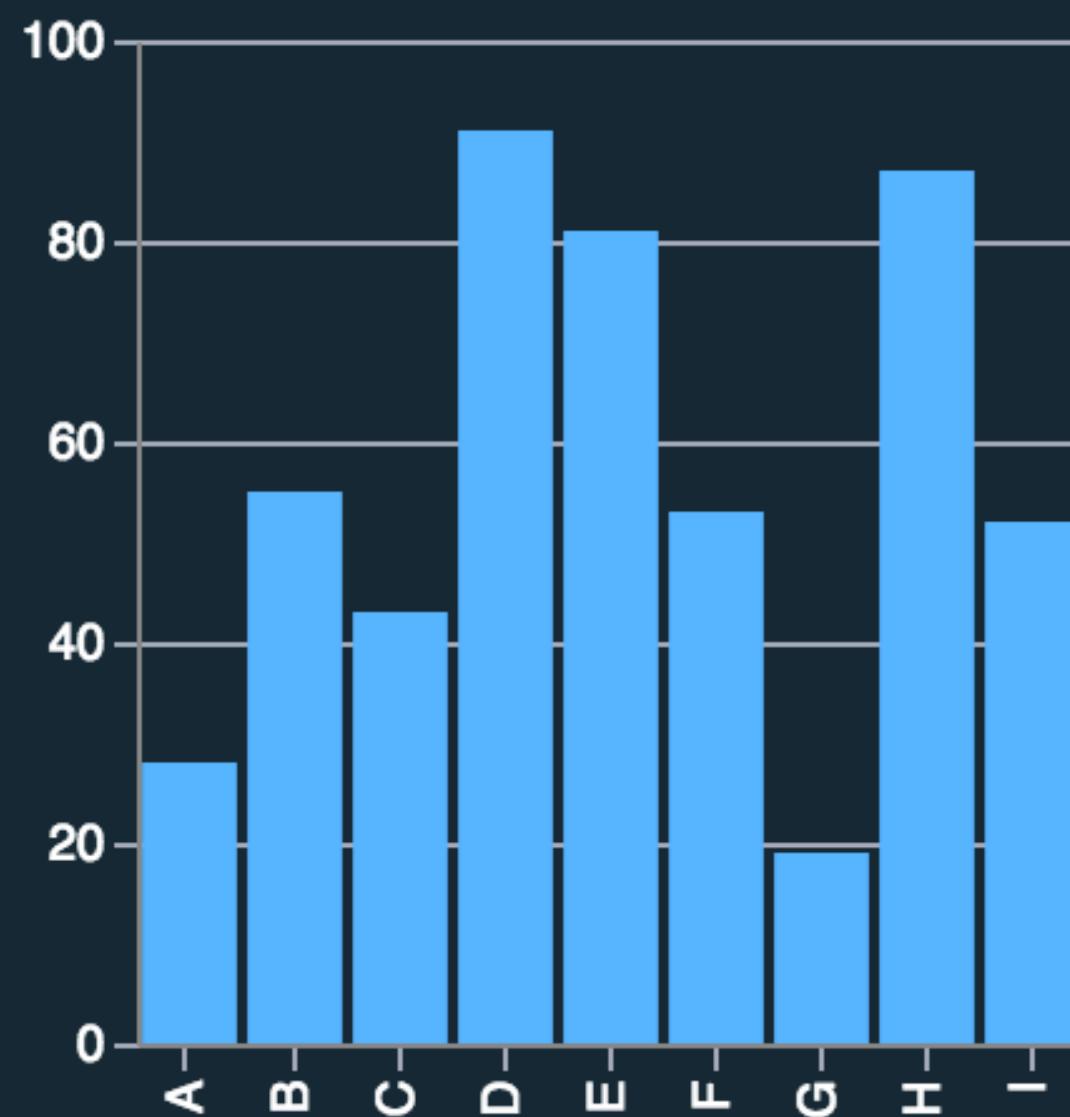


Mark: Bar

$d_{nominal} \rightarrow X$

$d_{quantitative} \rightarrow y$

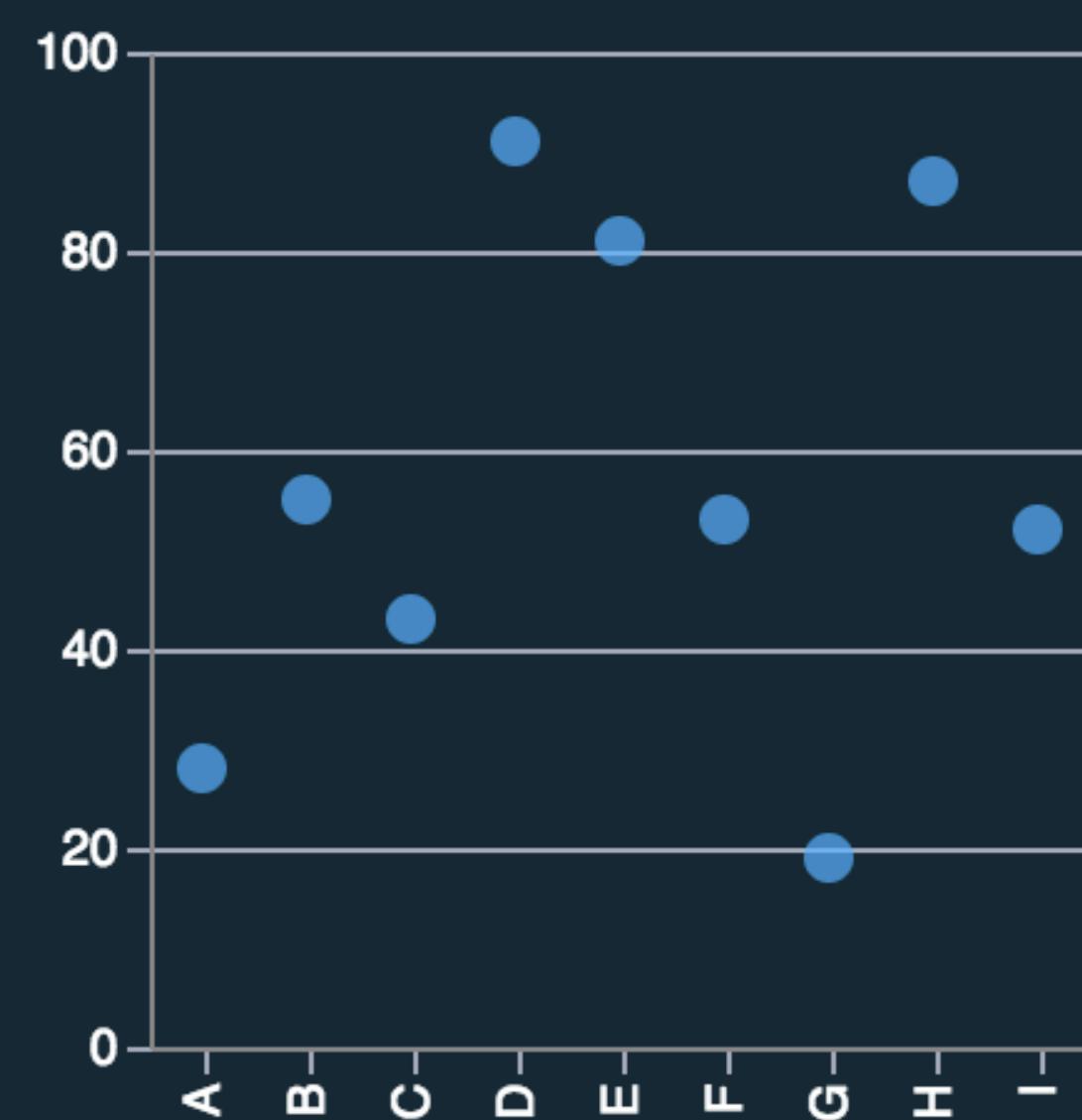
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar

$d_{\text{nominal}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

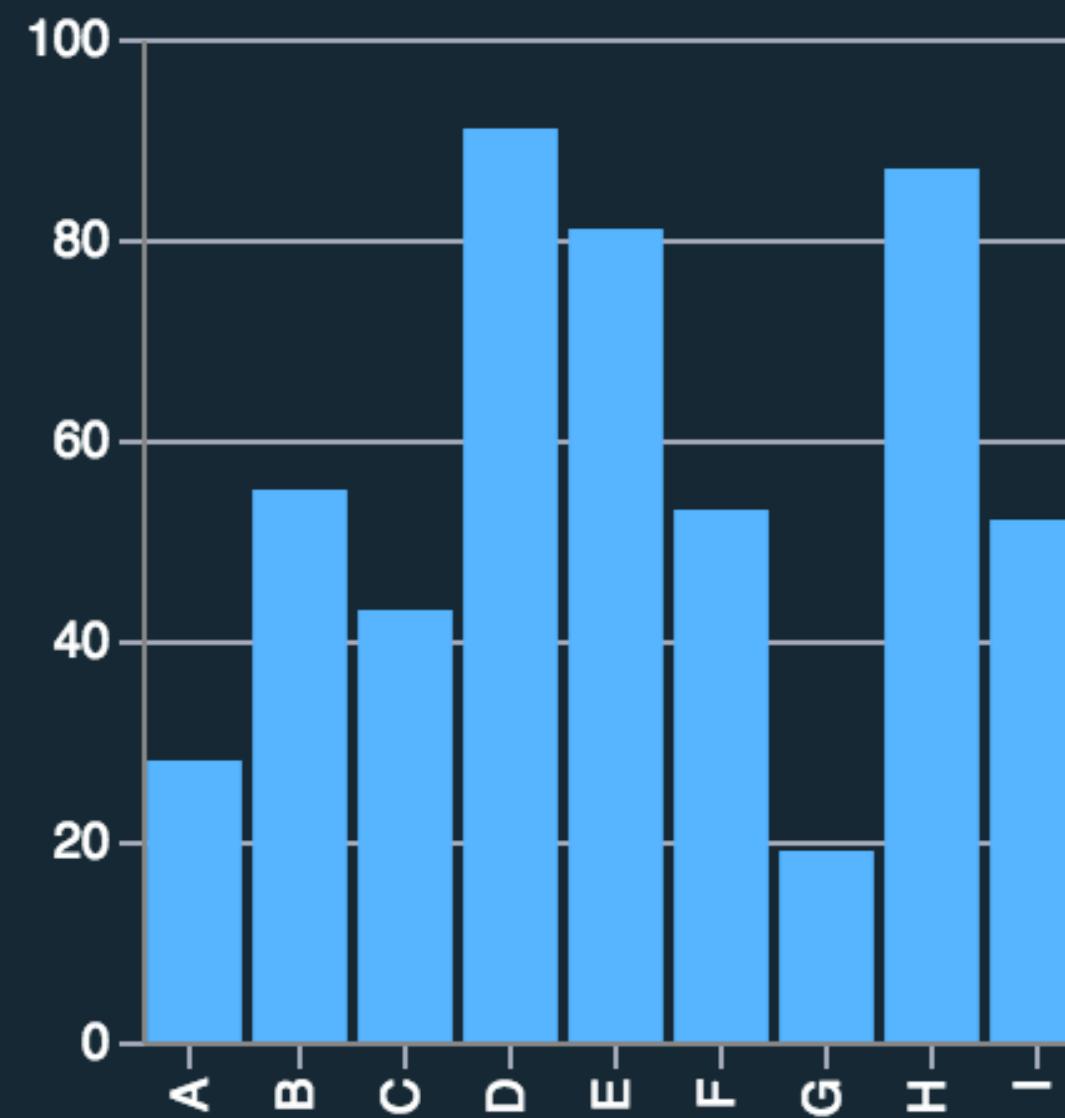


Mark: Point

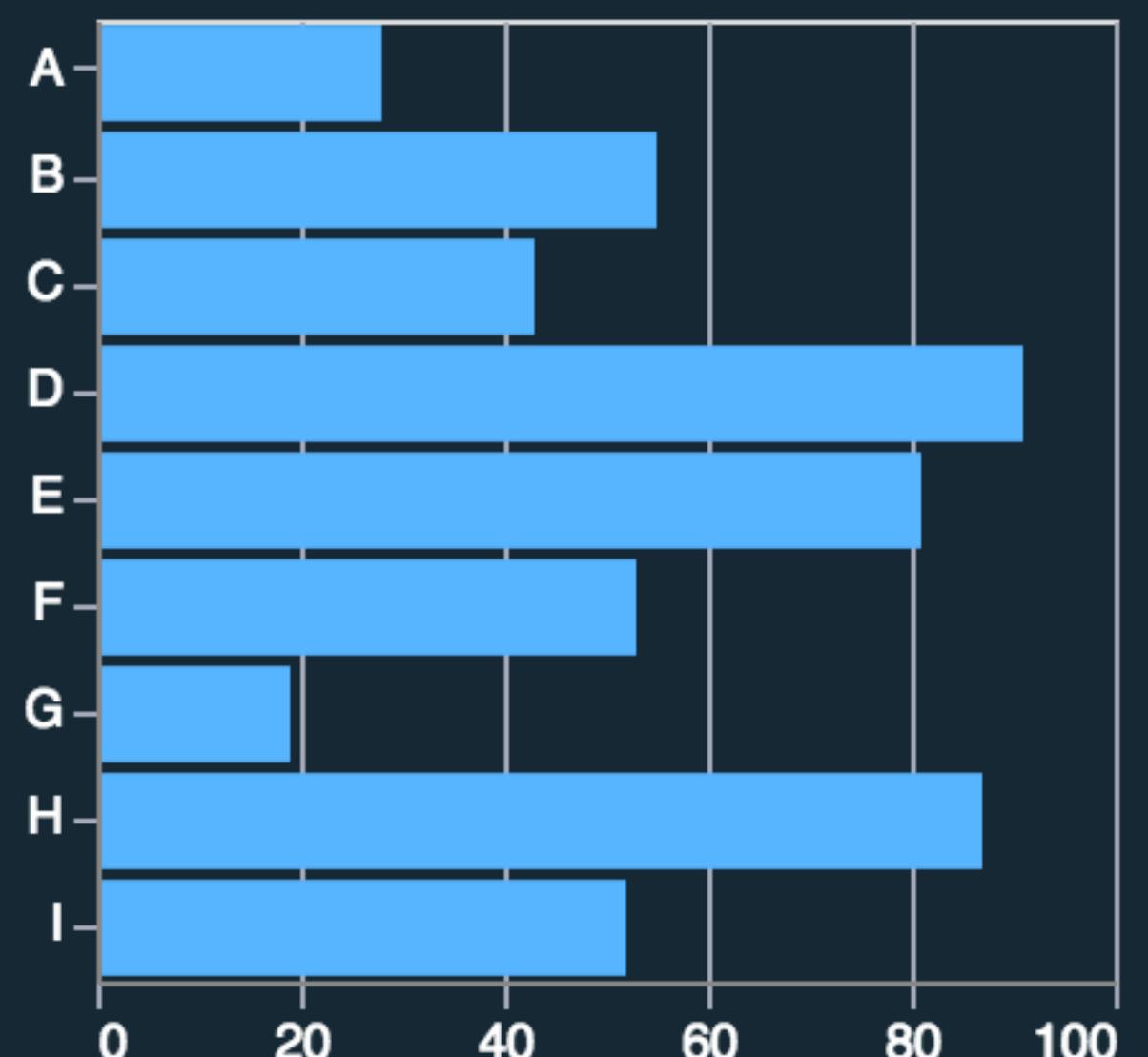
$d_{\text{nominal}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

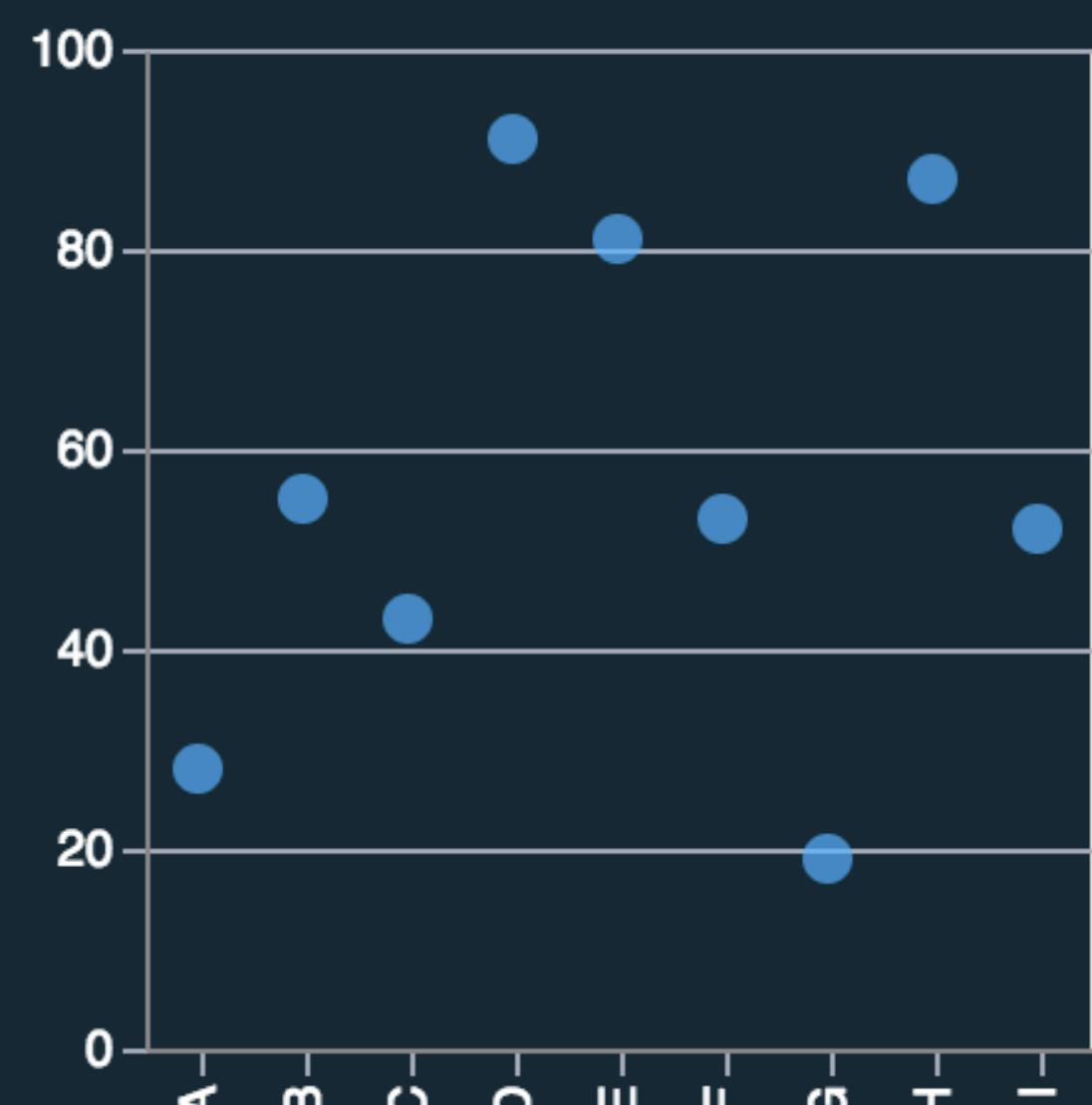
Visual Encoding: 1 Nominal, 1 Quantitative



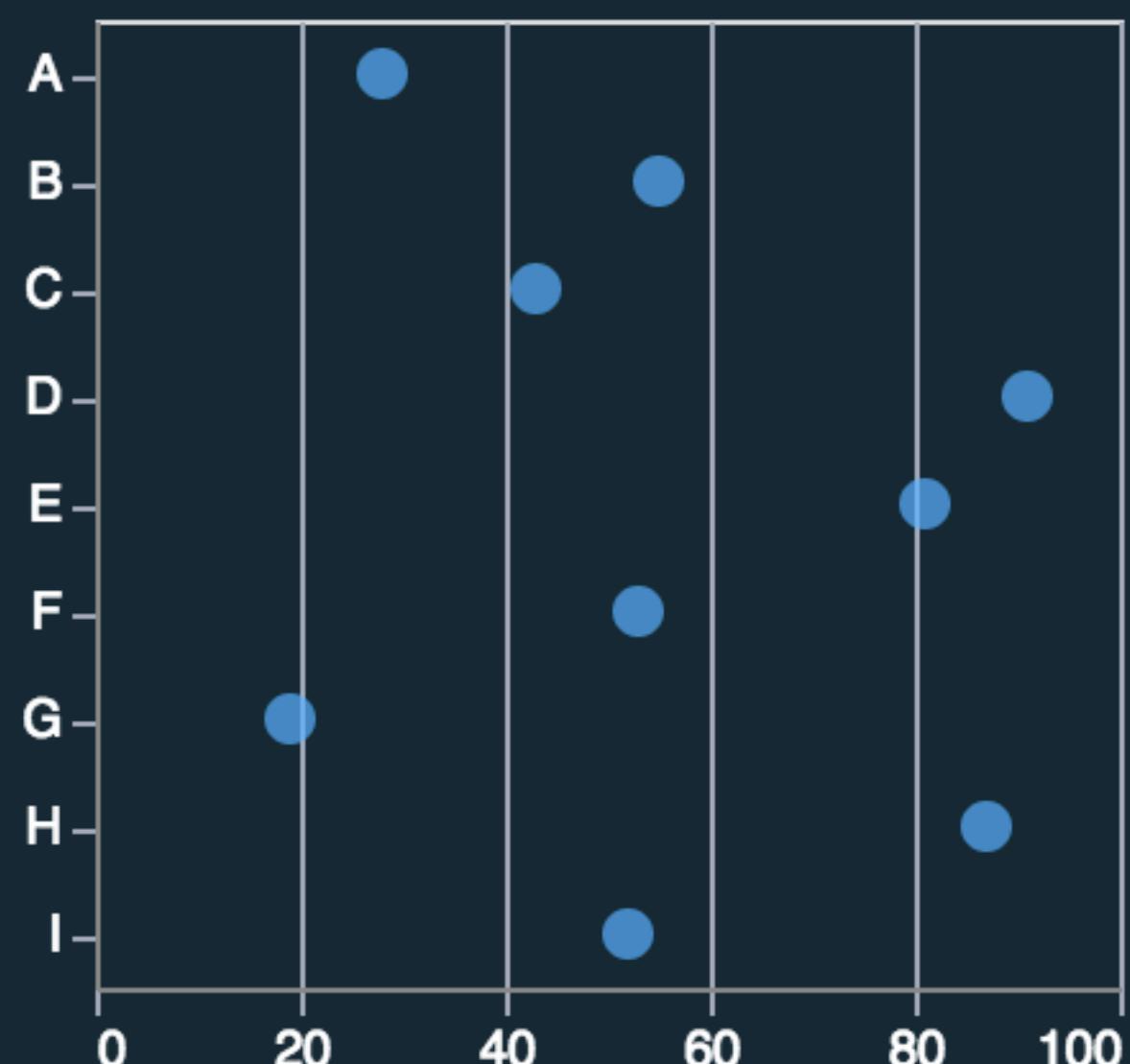
Mark: Bar
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow Y$



Mark: Bar
 $d_{\text{nominal}} \rightarrow Y$
 $d_{\text{quantitative}} \rightarrow X$

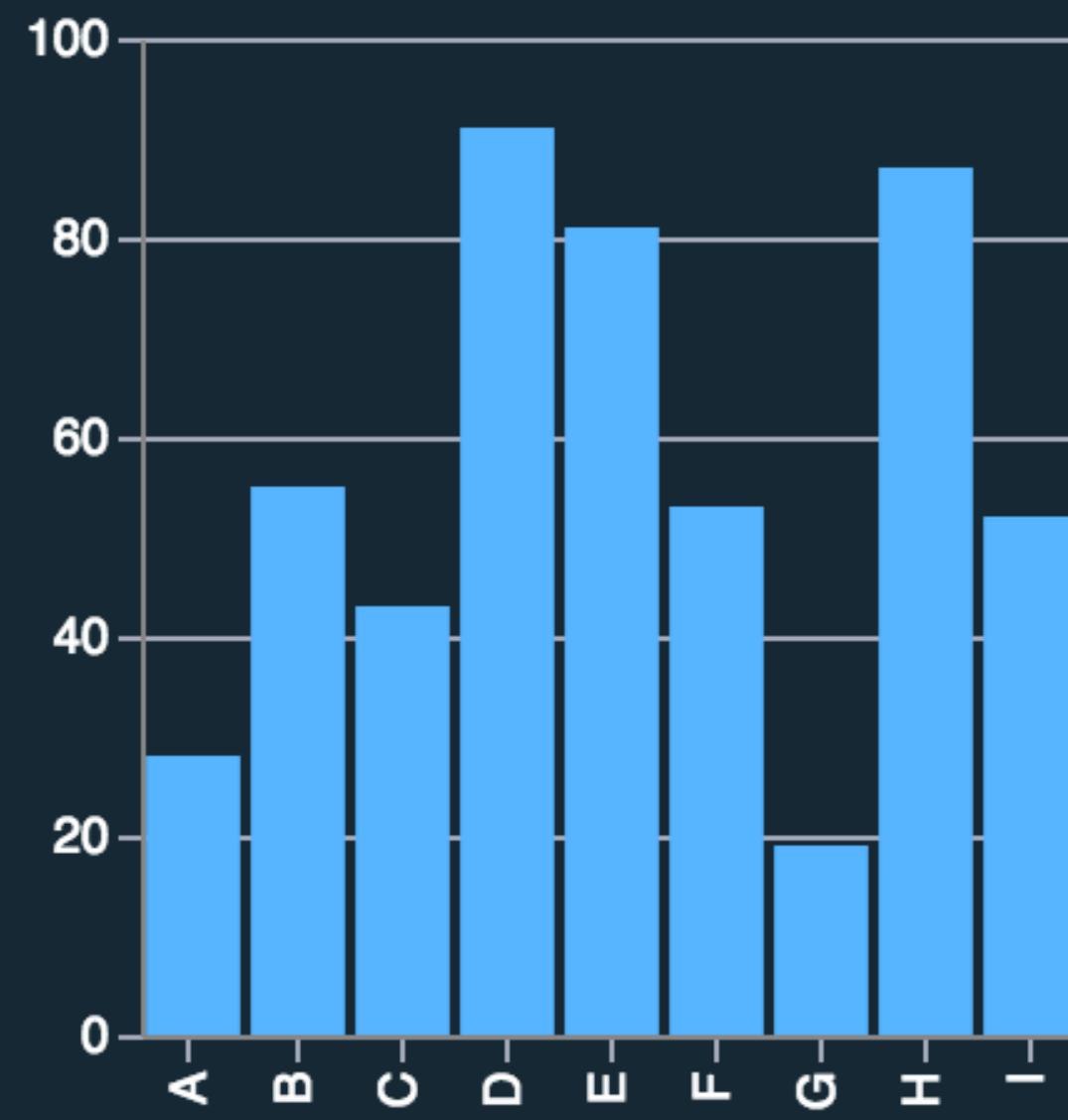


Mark: Point
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow Y$

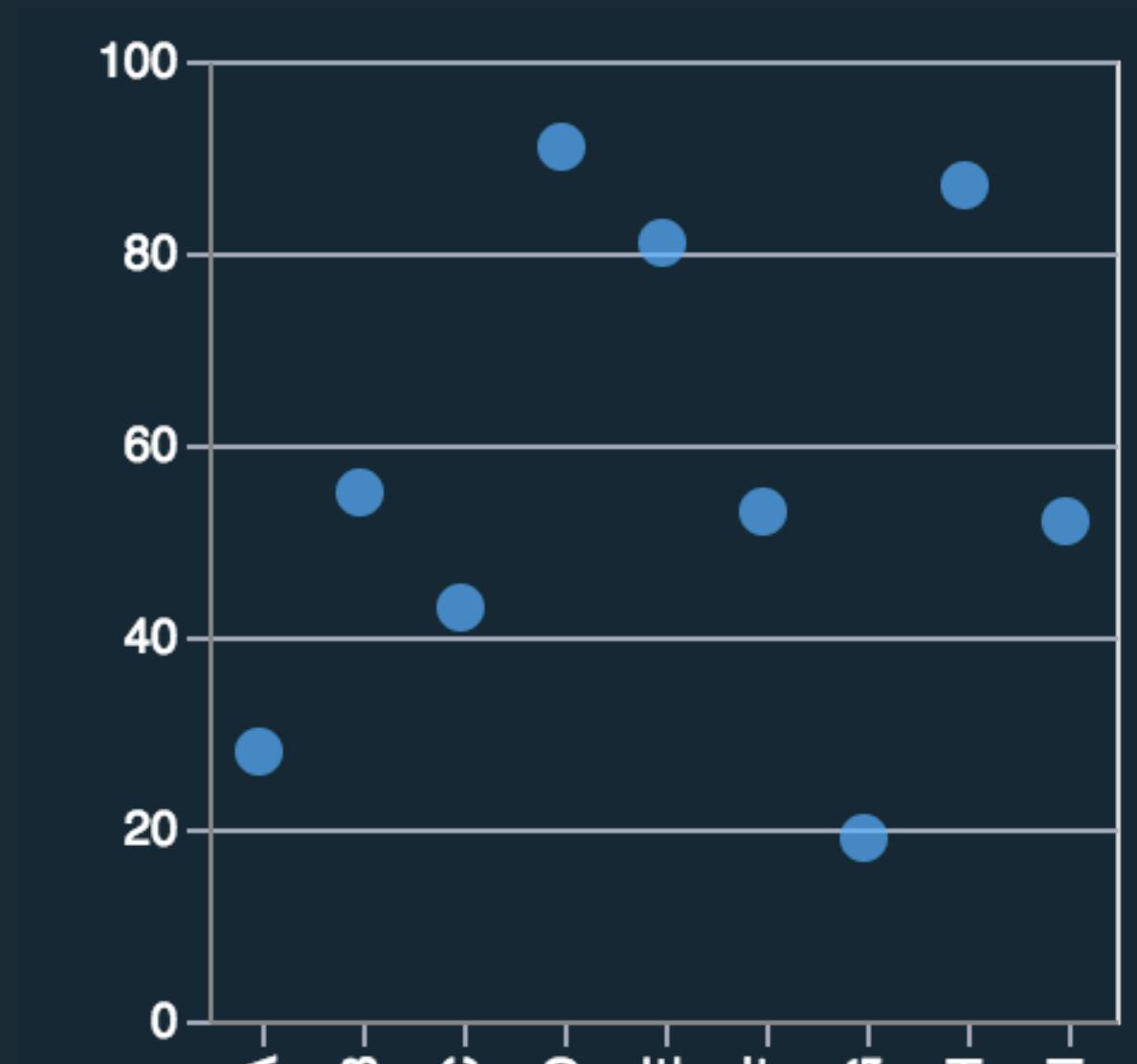


Mark: Point
 $d_{\text{nominal}} \rightarrow Y$
 $d_{\text{quantitative}} \rightarrow X$

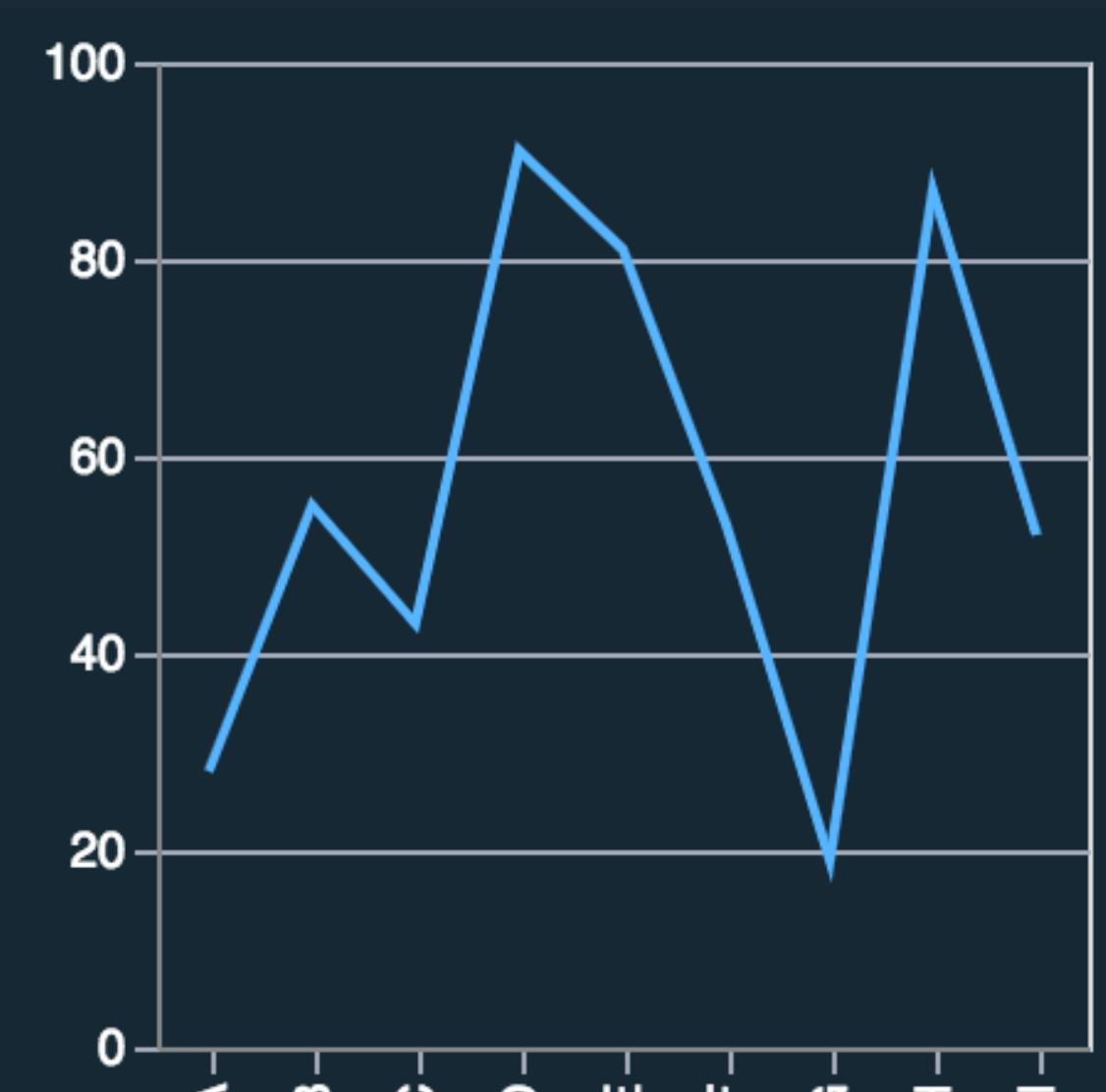
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$

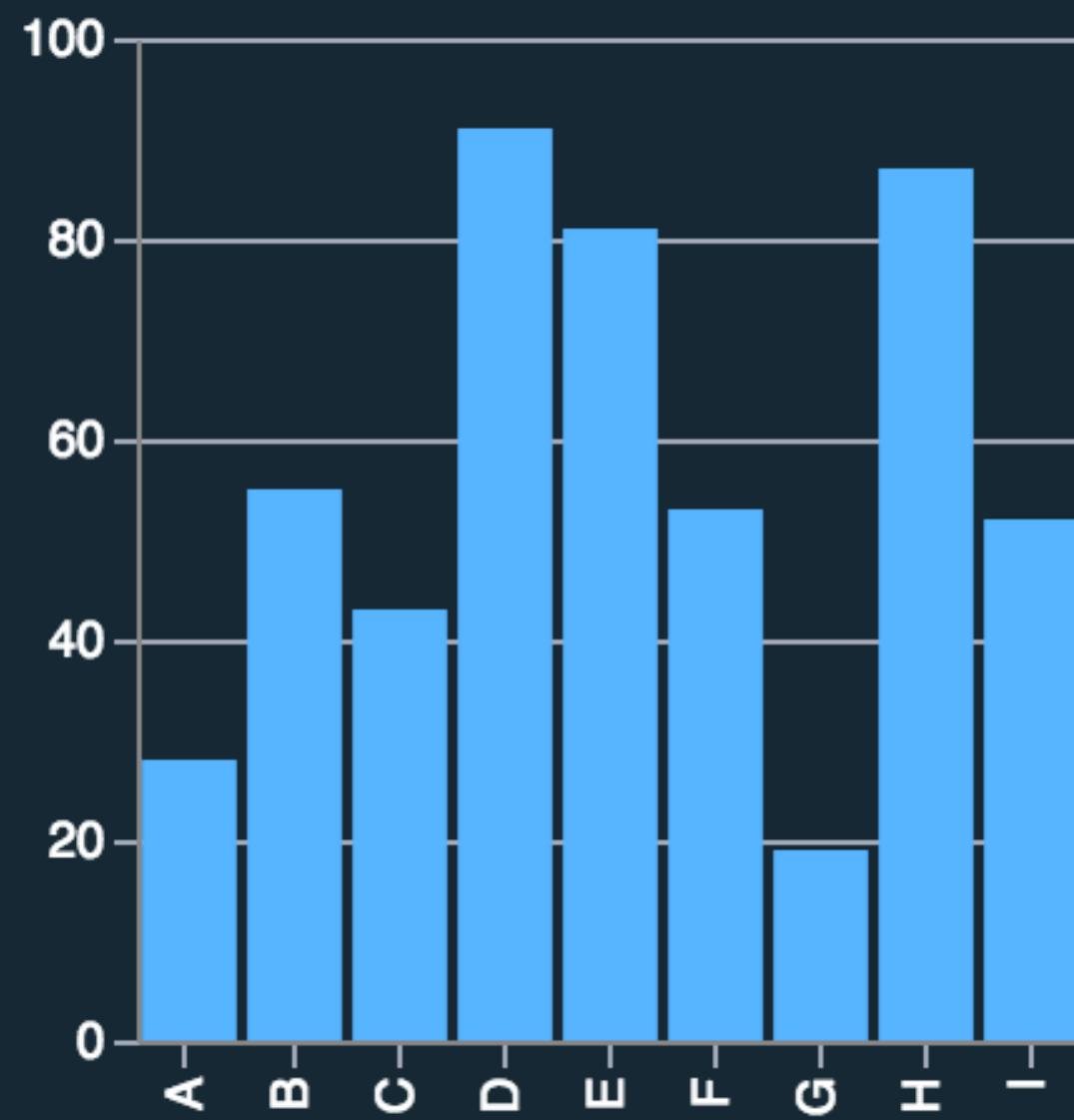


Mark: Point
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$

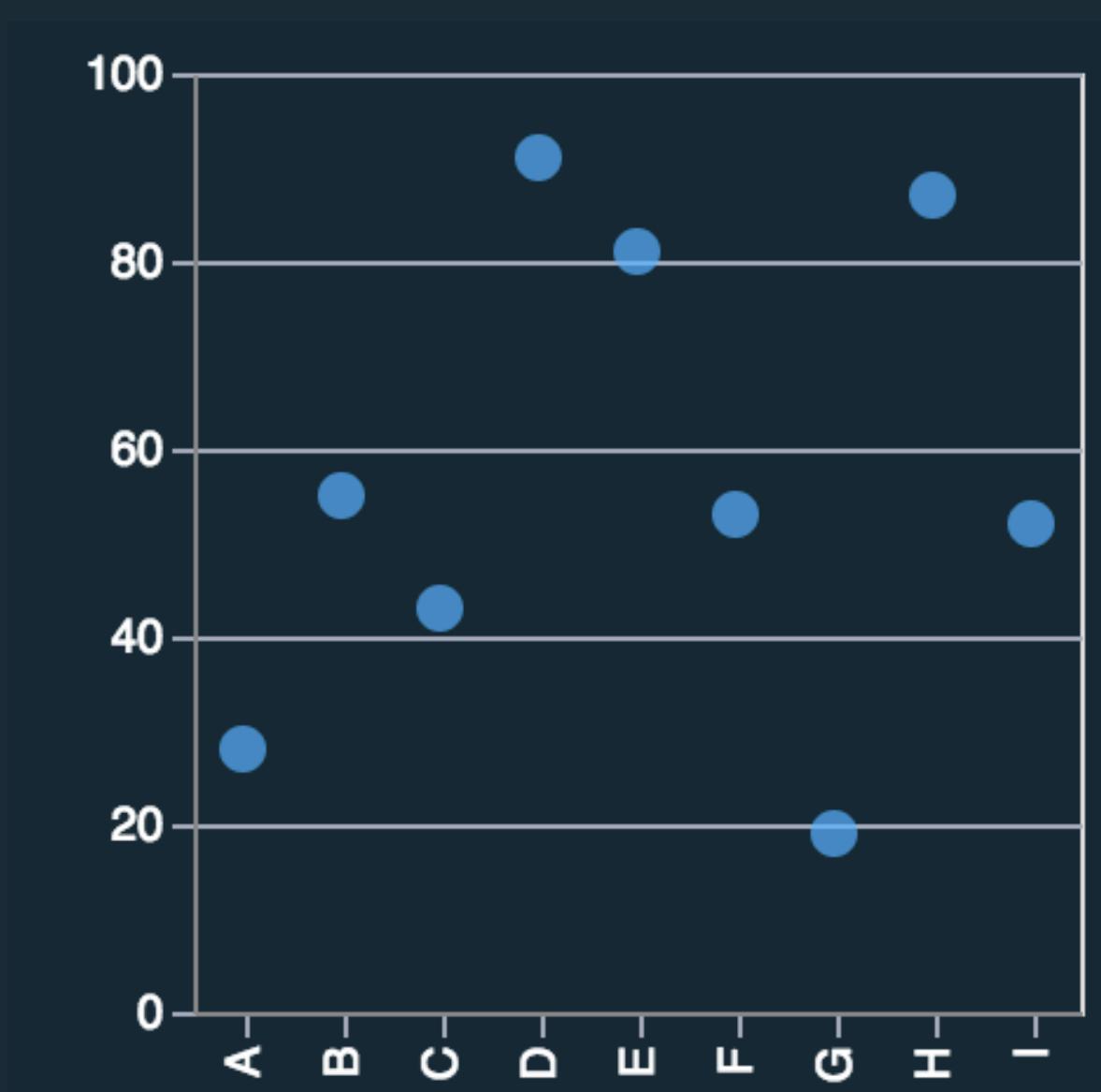


Mark: Line
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$

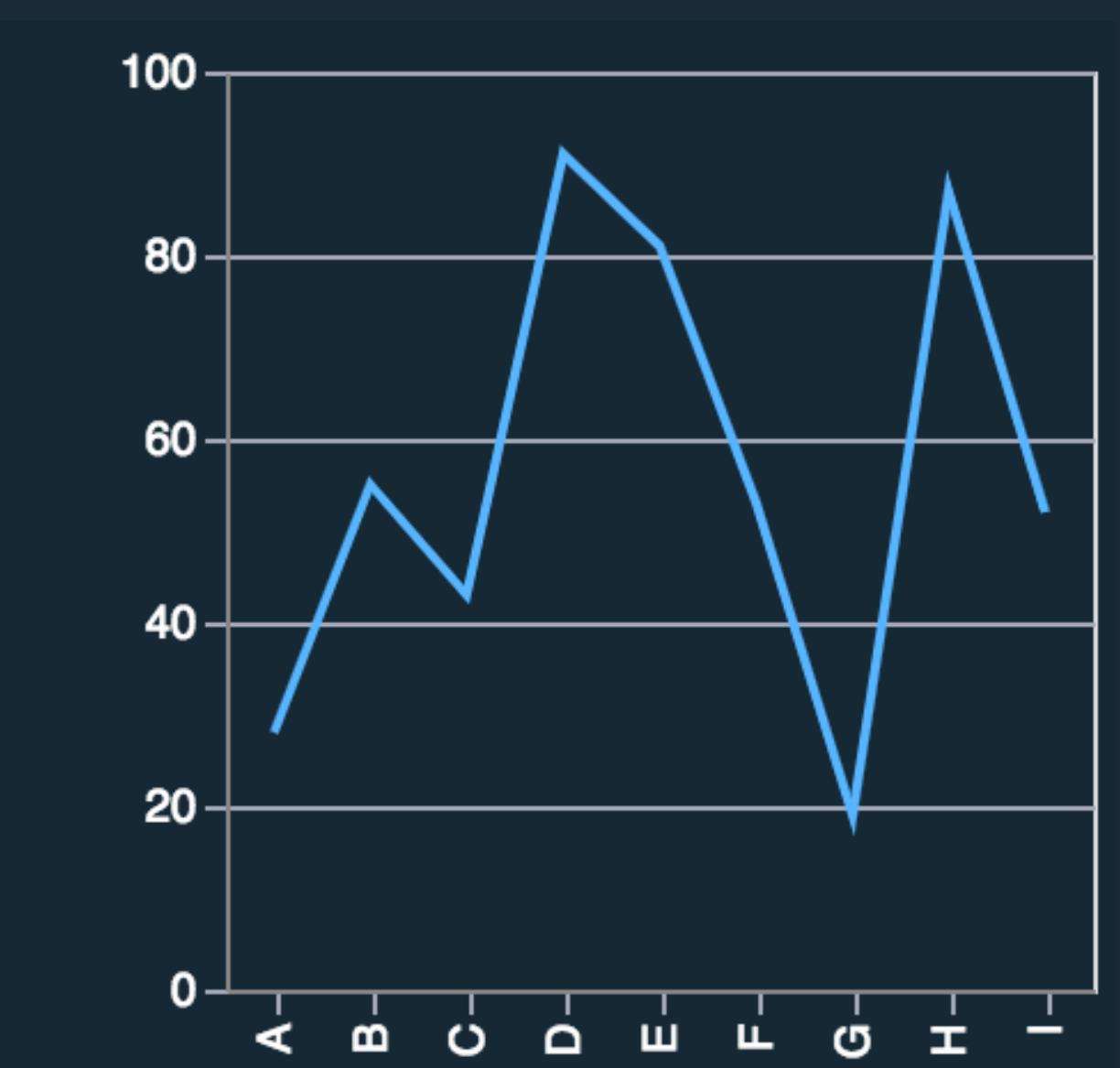
Visual Encoding: 1 Nominal, 1 Quantitative



Mark: Bar
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$



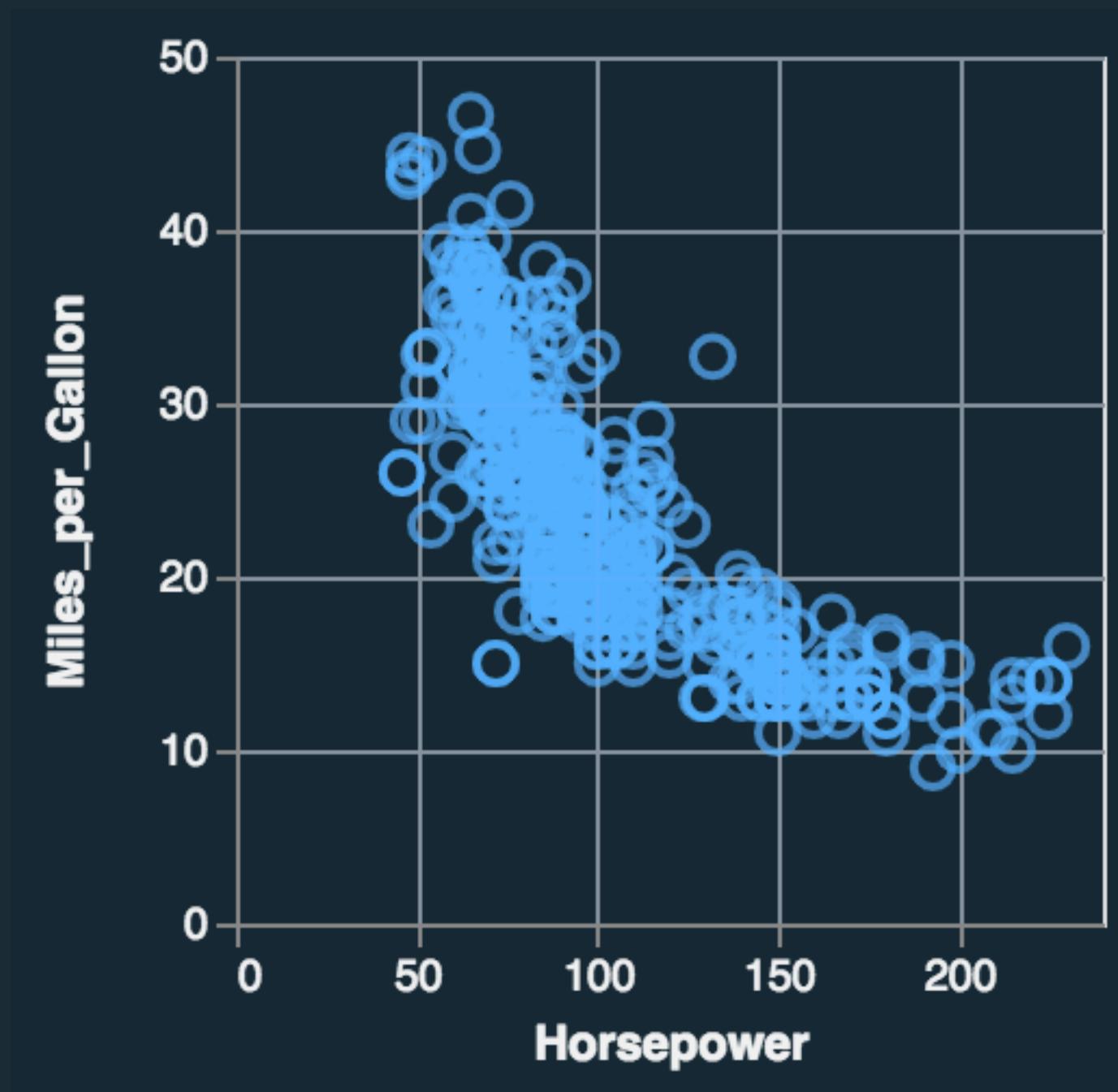
Mark: Point
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$



Mark: Line
 $d_{\text{nominal}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$

Violates expressiveness: the line mark implies a trend across the various categories.

Visual Encoding

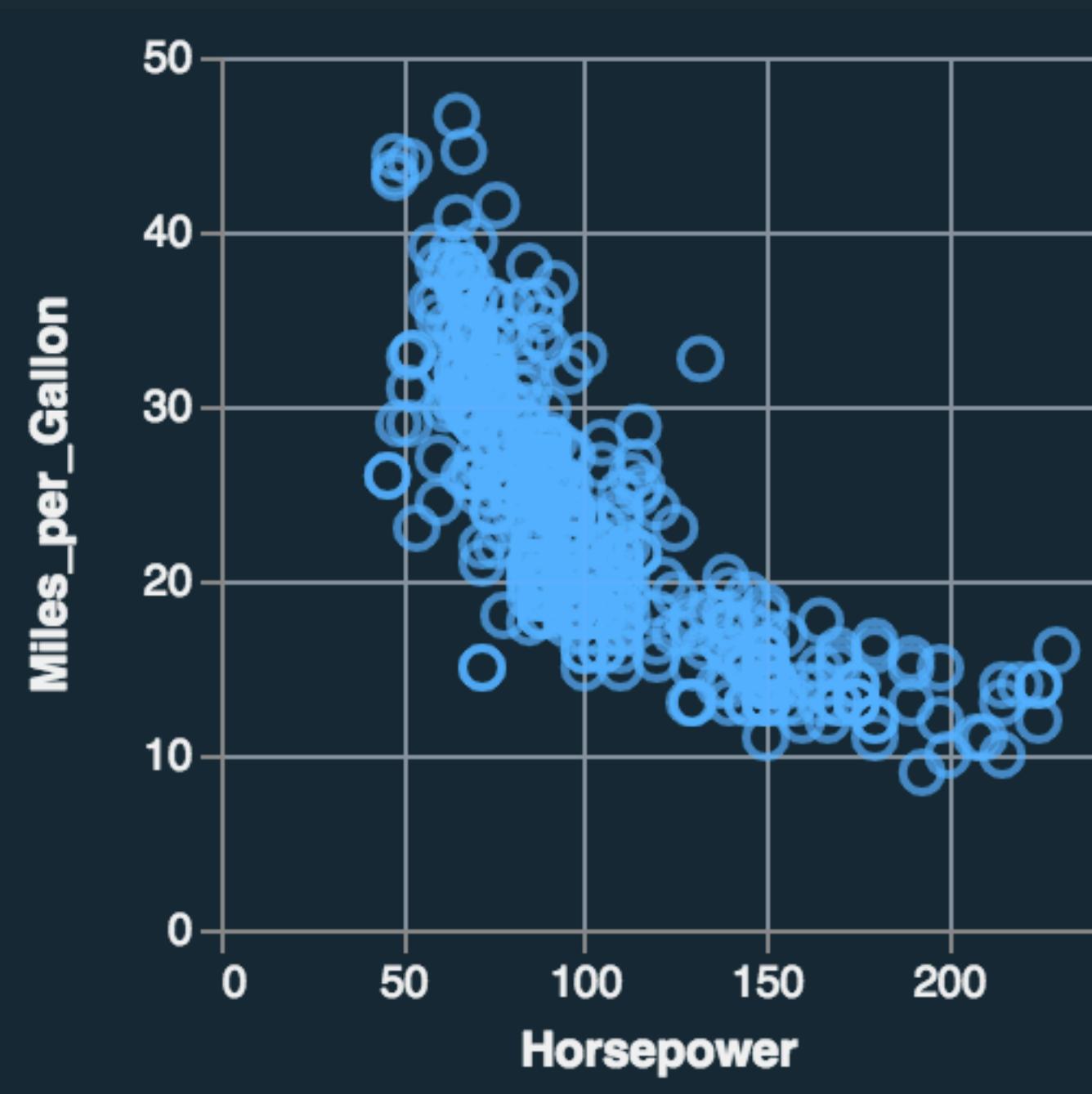


Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

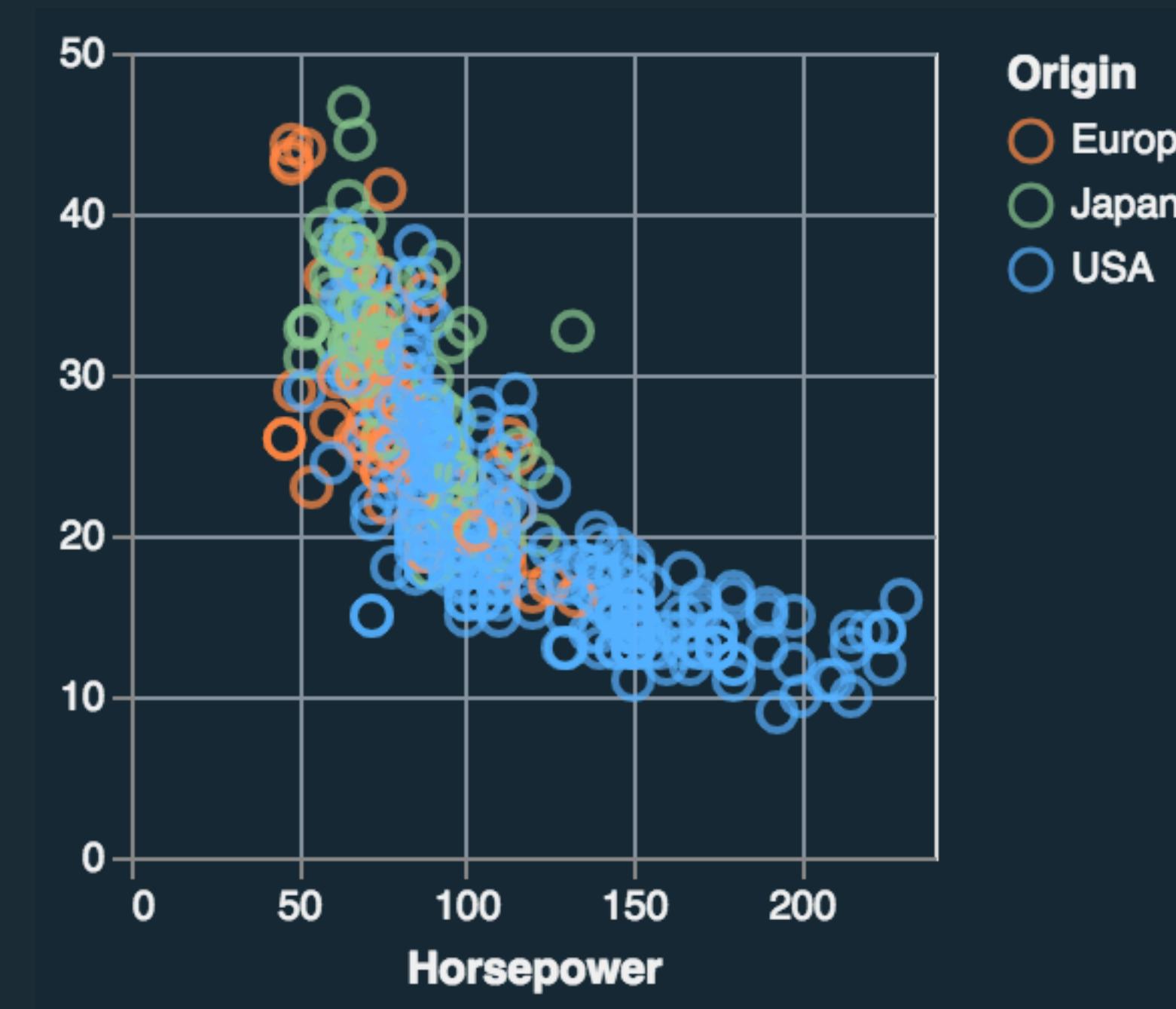
Visual Encoding



Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$



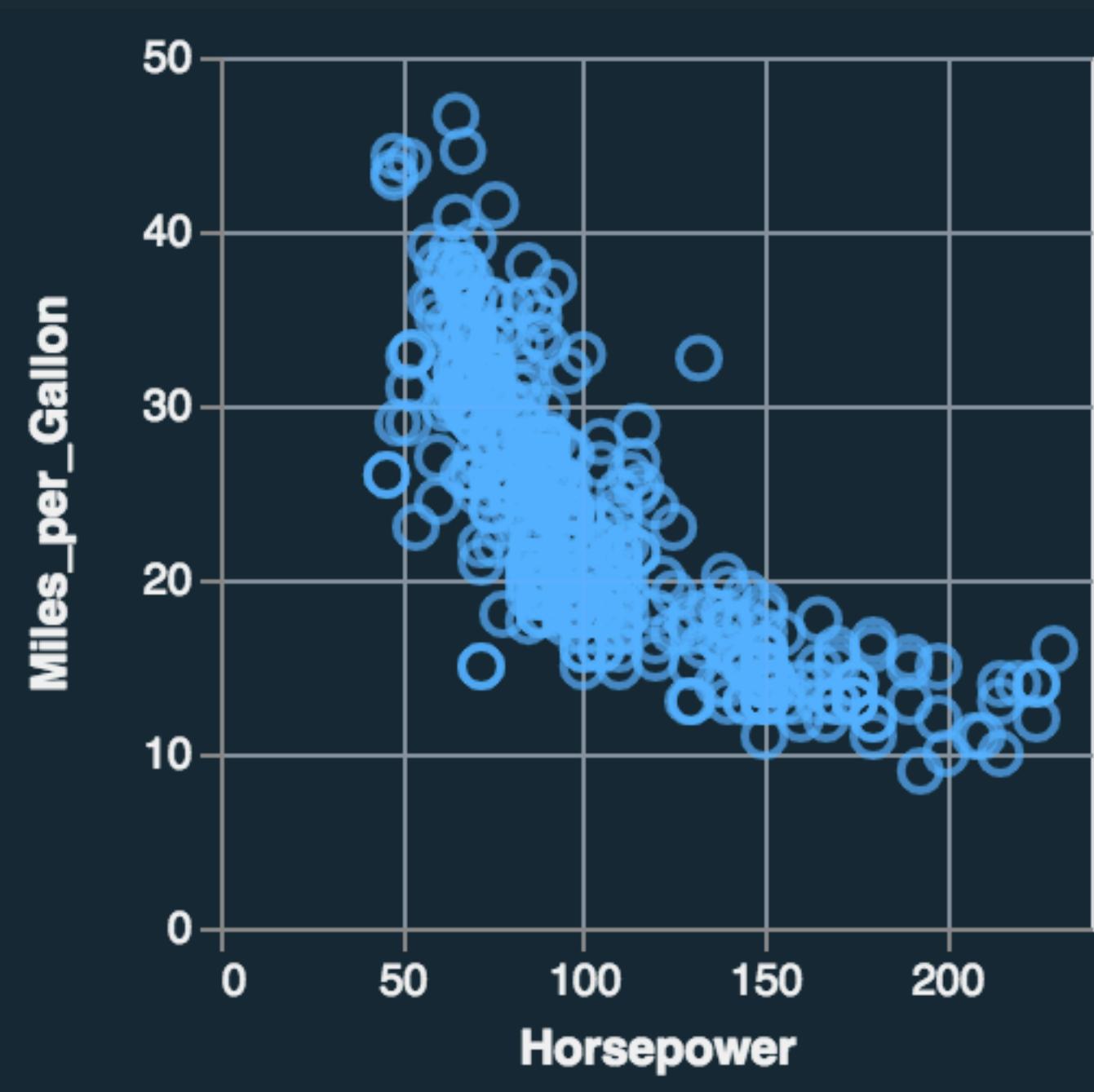
Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

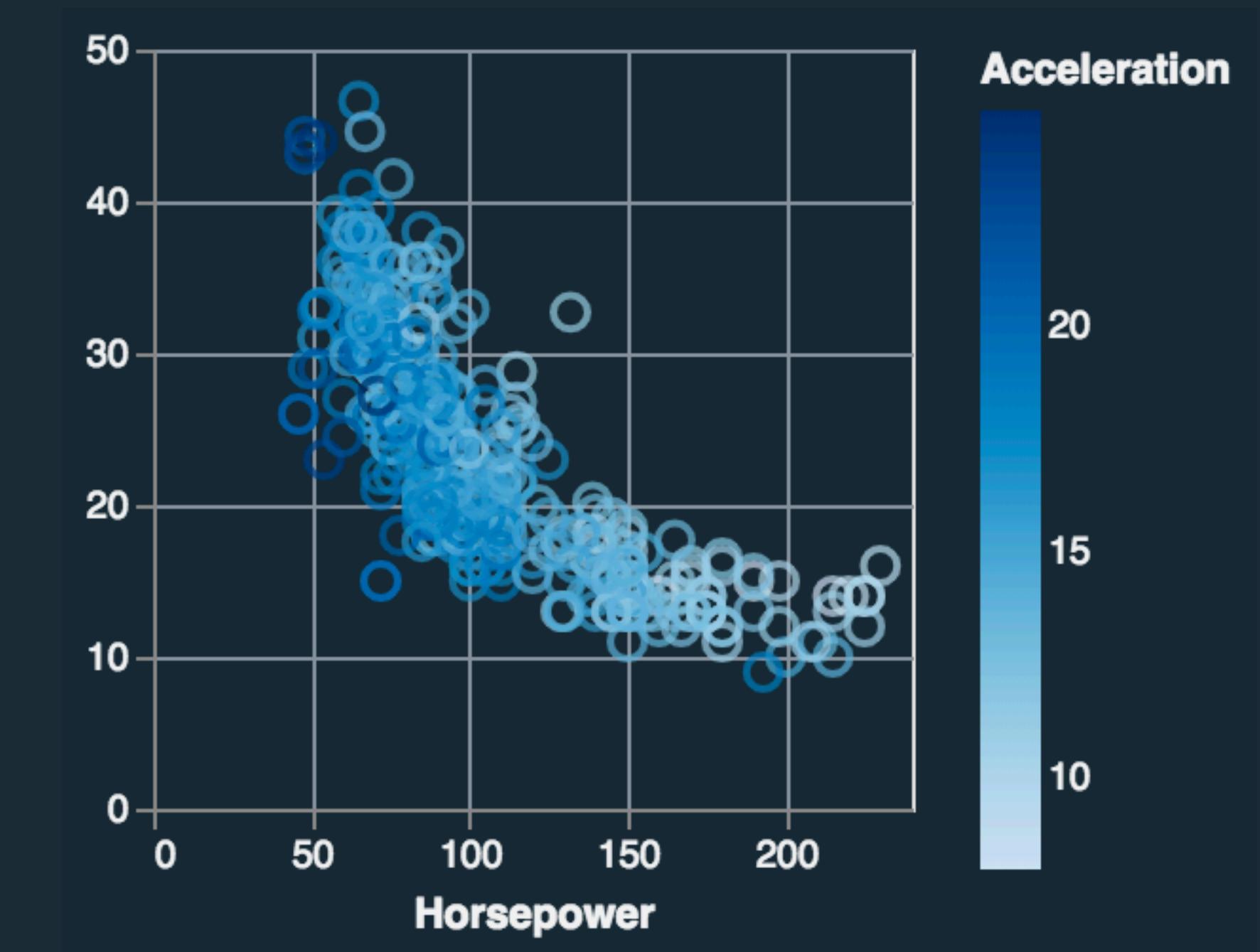
Visual Encoding



Mark: Point
 $d_{\text{quantitative}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$

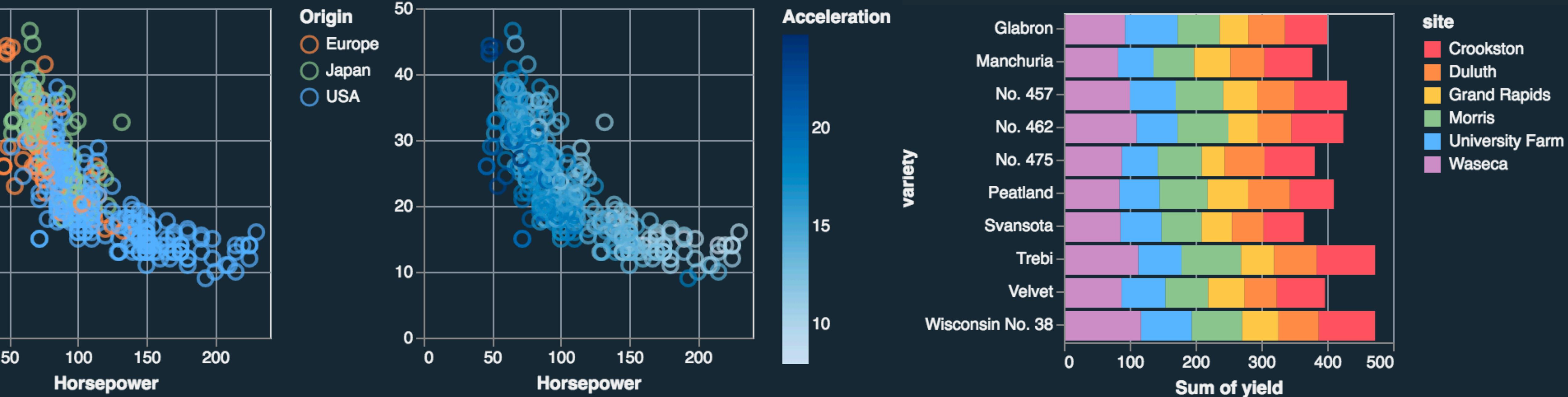


Mark: Point
 $d_{\text{quantitative}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$
 $d_{\text{nominal}} \rightarrow \text{color}$



Mark: Point
 $d_{\text{quantitative}} \rightarrow X$
 $d_{\text{quantitative}} \rightarrow y$
 $d_{\text{quantitative}} \rightarrow \text{color}$

Visual Encoding



Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{quantitative}} \rightarrow \text{color}$

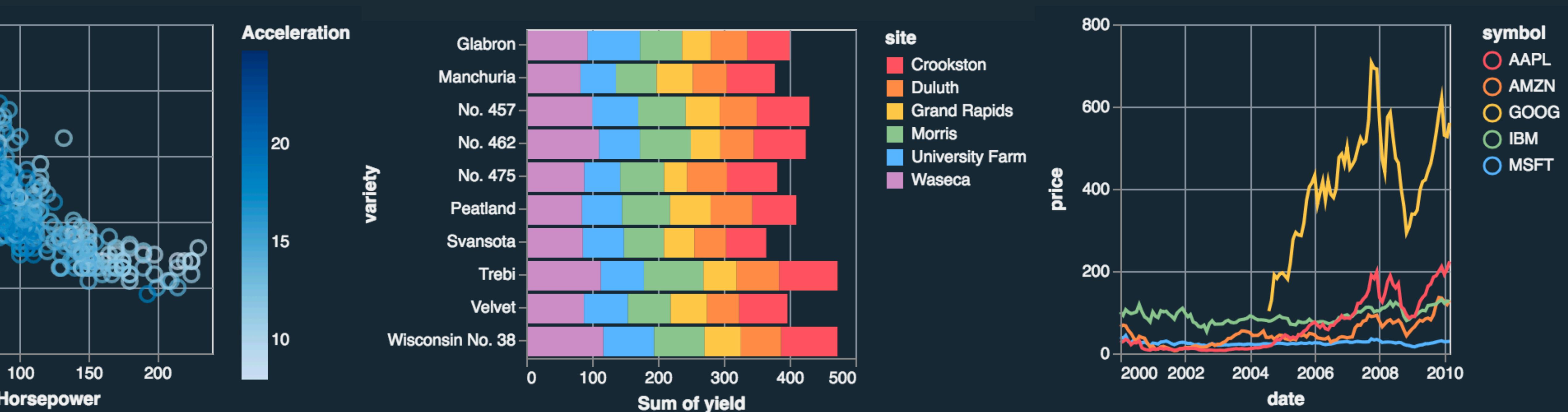
Mark: Bar

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{nominal}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

Visual Encoding



Mark: Point

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{quantitative}} \rightarrow \text{color}$

Mark: Bar

$d_{\text{quantitative}} \rightarrow X$

$d_{\text{nominal}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

Mark: Line

$d_{\text{temporal}} \rightarrow X$

$d_{\text{quantitative}} \rightarrow y$

$d_{\text{nominal}} \rightarrow \text{color}$

Visual Encoding

