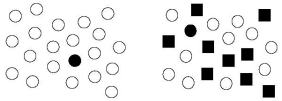
01

Data Models

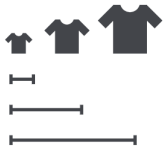
Perceptual Properties



Visual Encoding

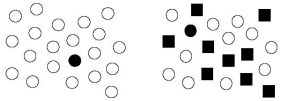
Marks & Channels

Data Types



Coming soon...

Perceptual Properties

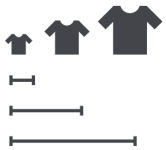


a.k.a Retinal Variables

Marks & Channels

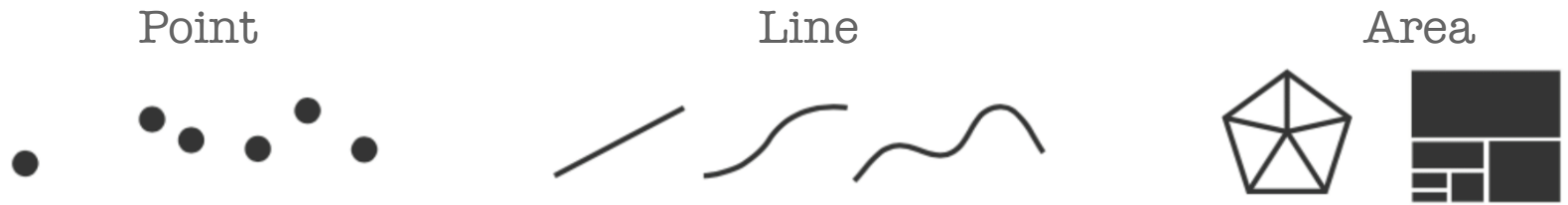
Visual Encoding

Data Types



Marks

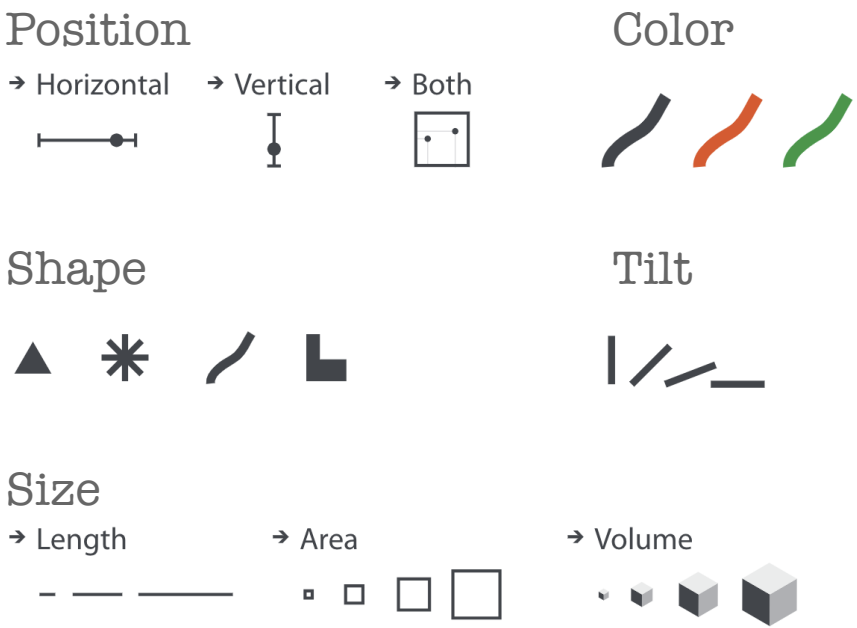
= Basic graphical element in image
(a.k.a. geometry)



Graphics from Munzner. Visualization Analysis and Design (2015)

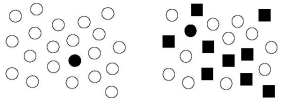
Channels

= Ways to control appearance of marks



Graphics from Munzner. Visualization Analysis and Design (2015)

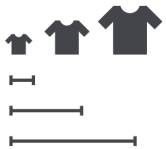
Perceptual Properties



Visual Encoding

Marks & Channels

Data Types



Categorical (a.k.a. Nominal or Qualitative)

Fruit (apple, pear, kiwi…)
Cities (NYC, SF, LA…)

Ordinal

Months (Jan, Feb, Mar…)
Sizes (S, M, L, XL…)

Quantitative

Lengths (1”, 2.5”, 5.14”...)
Population

Graphics from Munzner. Visualization Analysis and Design (2015)

Data Types

What does this mean?

```
14, 2.6, 30, 30, 15, 100001
```

Data Types

: refers to semantic models of data

Many aspects of vis design are driven by the kind of data you have.

```
14, 2.6, 30, 30, 15, 100001
```

Example from Munzner. Visualization Analysis and Design (2015)

Data types are formal descriptions

Math: sets with operations on them

Conceptual models are mental constructions

Include semantics and support reasoning

Example: data vs. conceptual

1D floats vs. temperatures

3D vector of floats vs. spatial location

Example: NYC Daily Temperatures, 2017

Data Type: 21, 28, 47, 55, … (integers)

Conceptual Model: Temperature (°F)

Data Model

Freezing vs. Not-Freezing (C)

Hot, Warm, Cold, Freezing (O)

Temperature Value (Q)

NYC Daily Temperature from https://www.ncdc.noaa.gov/cdo-web/search

Categorical
(a.k.a. Nominal or Qualitative)

Fruit (apple, pear, kiwi…)
Cities (NYC, SF, LA…)

Categorical (a.k.a. Nominal or Qualitative)

Fruit (apple, pear, kiwi…)
Cities (NYC, SF, LA…)

Ordinal

Months (Jan, Feb, Mar…)
Sizes (S, M, L, XL…)

Graphics from Munzner. Visualization Analysis and Design (2015)

Categorical (a.k.a. Nominal or Qualitative)

Fruit (apple, pear, kiwi…)
Cities (NYC, SF, LA…)

Ordinal

Months (Jan, Feb, Mar…)
Sizes (S, M, L, XL…)

Quantitative

Lengths (1”, 2.5”, 5.14”...)
Population

Graphics from Munzner. Visualization Analysis and Design (2015)

Data Models

C: Categorical (labels or or categories, a.k.a. nominal)

Operations: =, ≠

Categories are of equal importance, or “equidistant”

e.g. Eye color: blue, green, dark brown, light brown

Data Models

C: Categorical

Operations: **=**, **≠**

Categories are of equal importance, or “equidistant”

O: Ordered

Operations: =, ≠, **<**, **>**

Items of equal importance, or “equidistant”

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

Data Models

C: Categorical

Operations: **=**, **≠**

Categories are of equal importance, or “equidistant”

O: Ordered

Operations: =, ≠, **<**, **>**

Items of equal importance, or “equidistant”

Q-Interval (location of zero arbitrary)

Operations: =, ≠, <, >, **-**

Can measure distances or spans, only delta (i.e. intervals) may be compared

e.g. Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)

Data Models

C: Categorical

Operations: **=**, **≠**

Categories are of equal importance, or “equidistant”

O: Ordered

Operations: =, ≠, **<**, **>**

Items of equal importance, or “equidistant”

Q-Interval (location of zero arbitrary)

Operations: =, ≠, <, >, **-**

Can measure distances or spans, only delta (i.e. intervals) may be compared

Q-Ratio (zero fixed)

Operations: =, ≠, <, >, -, **%**

Can measure ratios or proportions e.g. Length, Mass, Temp, counts and amounts

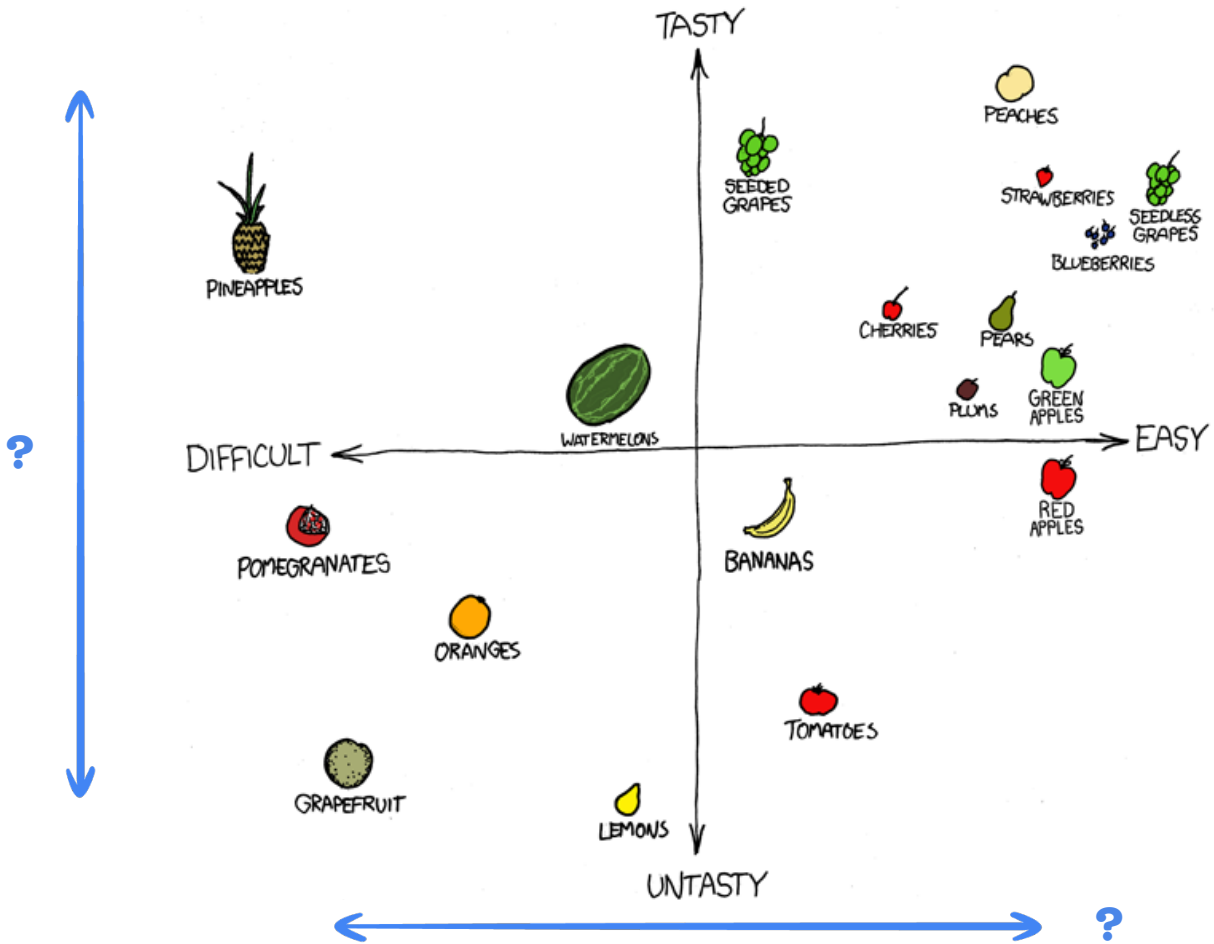
Slide via Jeff Heer

Exercise

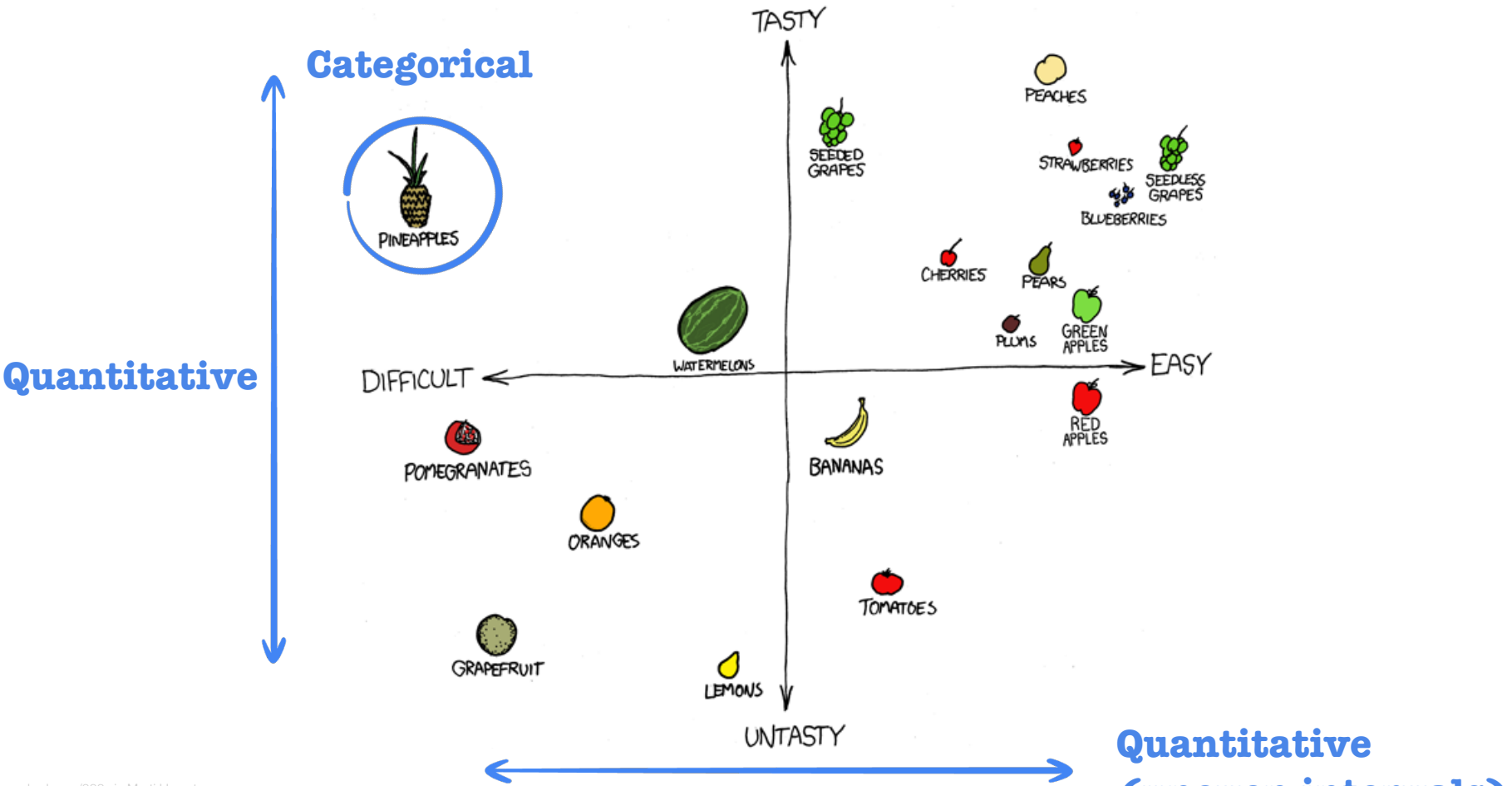
Order ID	Order Priority	Product Container	Product Base Margin
3	5-Low	Large Box	0.8
6	4-Not Specified	Small Pack	0.55
32	2-High	Small Pack	0.79
32	2-High	Jumbo Box	0.72
32	2-High	Medium Box	0.6
32	2-High	Small Box	0.65
35	4-Not Specified	Small Pack	0.52
35	4-Not Specified	Jumbo Box	0.58
36	1-Urgent	Medium Box	0.55
65	1-Urgent	Medium Box	0.49
66	5-Low	Wrap Bag	0.56
69	4-Not Specified	Small Box	0.44
69	4-Not Specified	Small Box	0.6
70	5-Low	Small Pack	0.59
70	5-Low	Wrap Bag	0.82
96	2-High	Small Box	0.55
97	3-Medium	Medium Box	0.38
129	5-Low	Small Box	0.37
130	2-High	Medium Box	0.37

Order ID	Order Priority	Product Container	Product Base Margin
3	5-Low	Large Box	0.8
6	4-Not Specified	Small Pack	0.55
32	2-High	Small Pack	0.79
32	2-High	Jumbo Box	0.72
32	2-High	Medium Box	0.6
32	2-High	Small Box	0.65
35	4-Not Specified	Small Pack	0.52
35	4-Not Specified	Jumbo Box	0.58
36	1-Urgent	Medium Box	0.55
65	1-Urgent	Medium Box	0.49
66	5-Low	Wrap Bag	0.56
69	4-Not Specified	Small Box	0.44
69	4-Not Specified	Small Box	0.6
70	5-Low	Small Pack	0.59
70	5-Low	Wrap Bag	0.82
96	2-High	Small Box	0.55
97	3-Medium	Medium Box	0.38
129	5-Low	Small Box	0.37
130	2-High	Medium Box	0.37

Quantitative
Ordinal
Categorical



xkcd.com/388 via Marti Hearst



xkcd.com/388 via Marti Hearst

02

Example: NYC Daily Temperature

NYC Daily Temperatures, 2017

Data Type: 21, 28, 47, 55, … (integers)

Conceptual Model: Temperature (°F)

Data Model

Freezing vs. Not-Freezing (C)

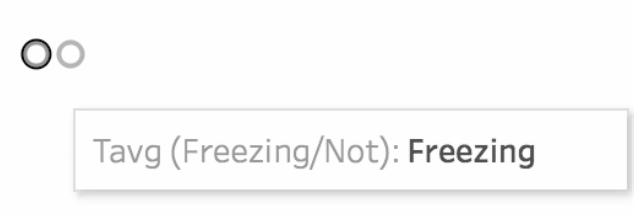
Hot, Warm, Cold, Freezing (O)

Temperature Value (Q)

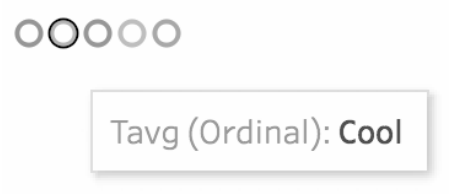
NYC Daily Temperature from https://www.ncdc.noaa.gov/cdo-web/search

Abc nyc-2017dailytemp... Station	Abc nyc-2017dailytemp-1476576... Name	 nyc-2017dailyt... Date	# nyc-2017da... Tavg
USW00014732	LA GUARDIA AIRPOR...	1/1/2017	46
USW00014732	LA GUARDIA AIRPOR...	1/2/2017	40
USW00014732	LA GUARDIA AIRPOR...	1/3/2017	42
USW00014732	LA GUARDIA AIRPOR...	1/4/2017	47
USW00014732	LA GUARDIA AIRPOR...	1/5/2017	34
USW00014732	LA GUARDIA AIRPOR...	1/6/2017	32
USW00014732	LA GUARDIA AIRPOR...	1/7/2017	25
USW00014732	LA GUARDIA AIRPOR...	1/8/2017	23
USW00014732	LA GUARDIA AIRPOR...	1/9/2017	21

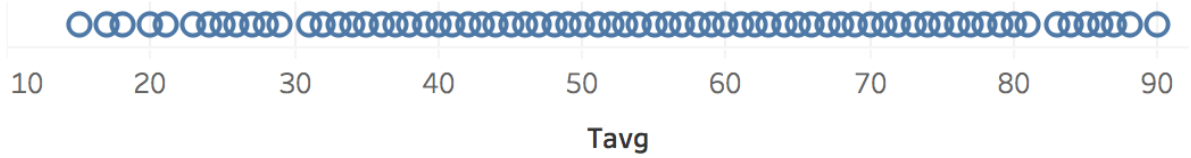
1D Categorical



1D Ordinal

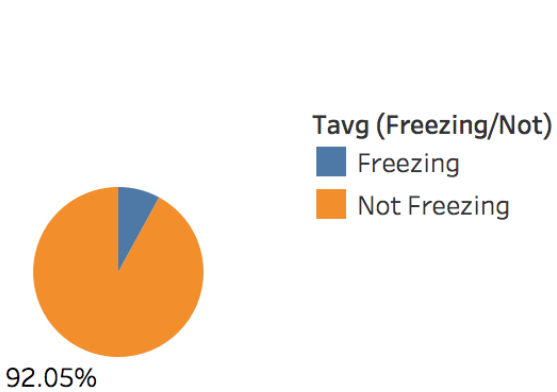


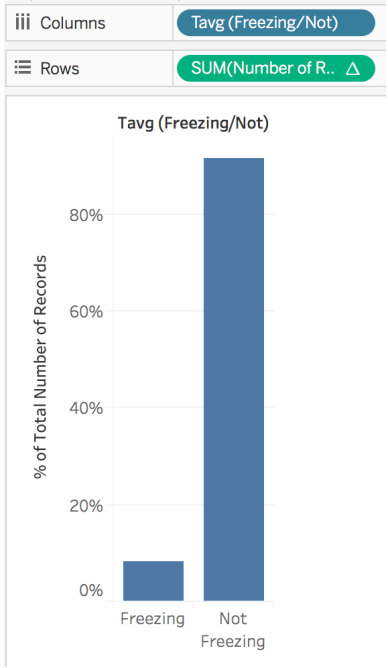
1D Quantitative



2D: Categorical x Quantitative (COUNT)

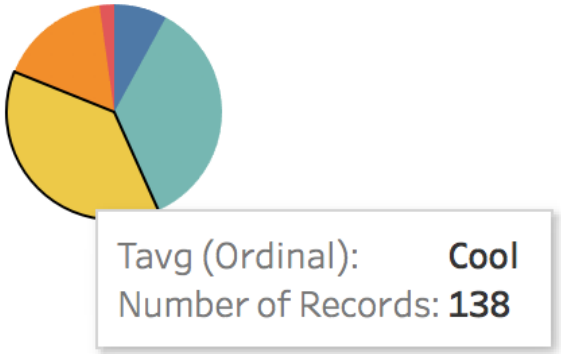
Freezing = under 32 degrees

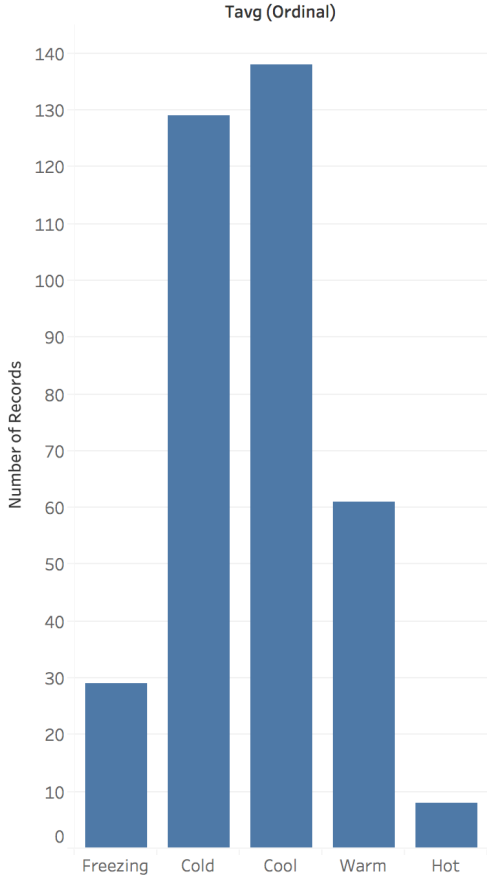




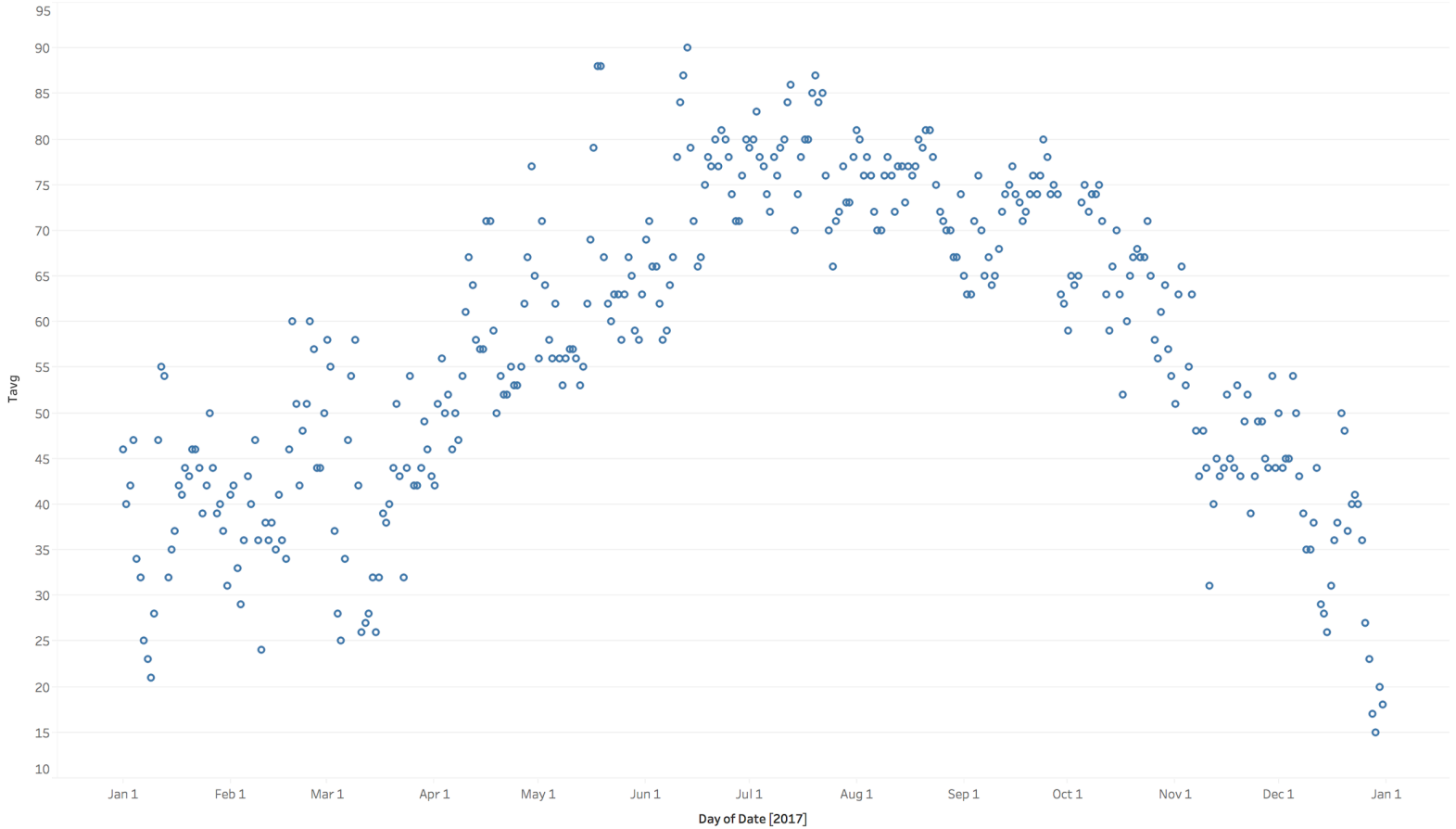
2D: Ordinal x Quantitative (COUNT)

Freezing	<32
Cold	32–54
Cool	55–74
Warm	75–85
Hot	85+



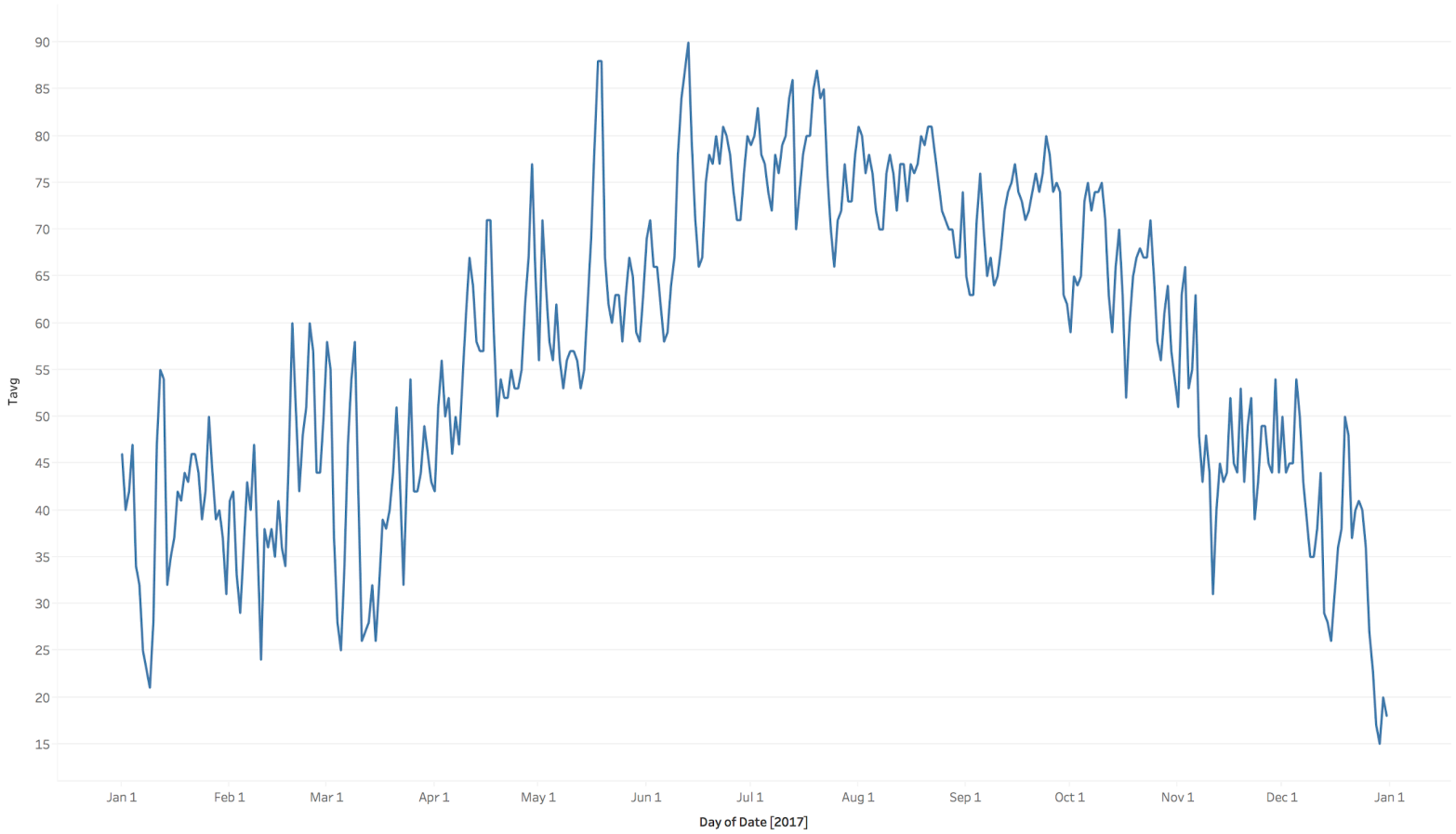


2D: Quantitative x Quantitative (day)

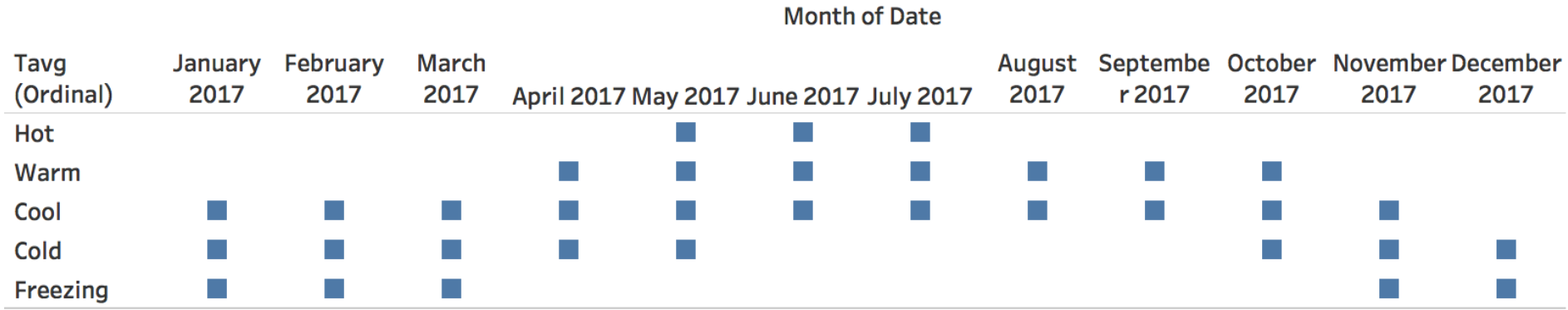


Scatterplot: doesn't make sense for a timeseries. What could make sense?

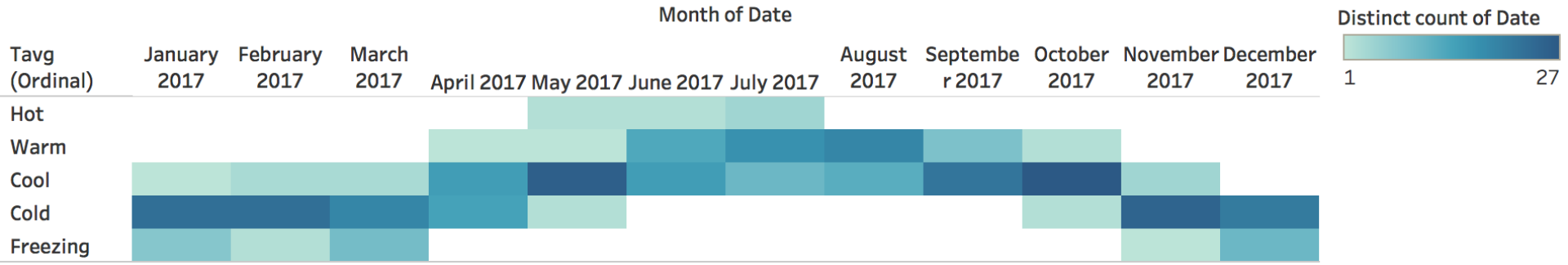
2D: Quantitative x Quantitative (day)



2D: Ordinal x Ordinal (month)



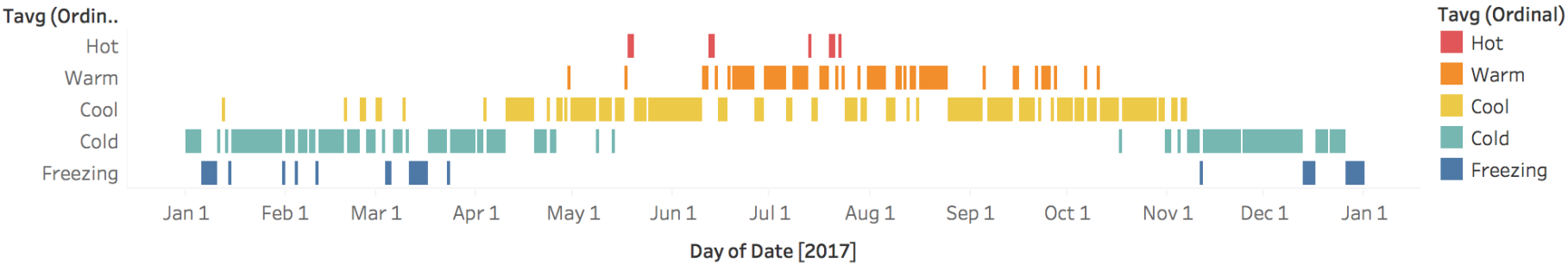
3D: Ordinal x Ordinal (month) x Quant (count)



Color: count of days (Q binned into O)

2D: Ordinal x Quantitative (day)

Gantt chart



Color: temperature (O) *repeat

Revisit: C ⊂ O ⊂ Q

Data Type: 21, 28, 47, 55, … (integers)

Conceptual Model: Daily Avg. Temperature (°F)

Data Model

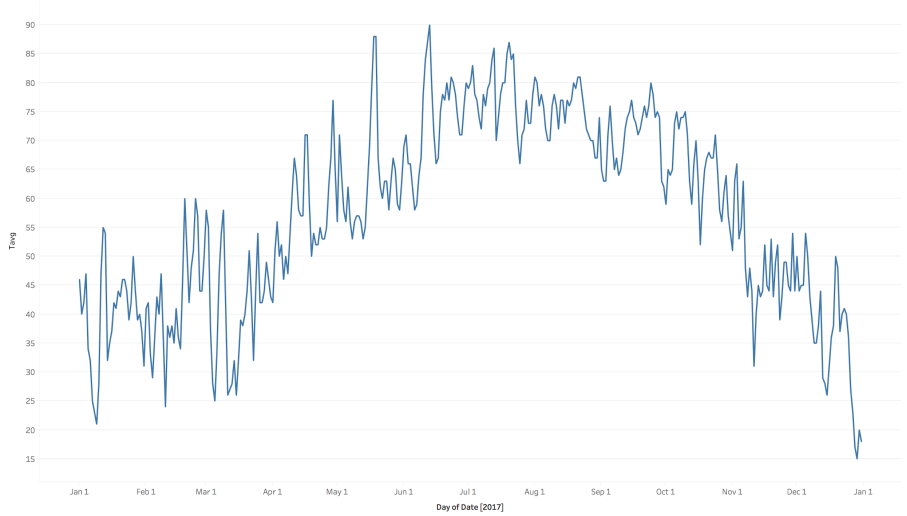
Freezing vs. Not-Freezing (C)

Hot, Warm, Cold, Freezing (O)

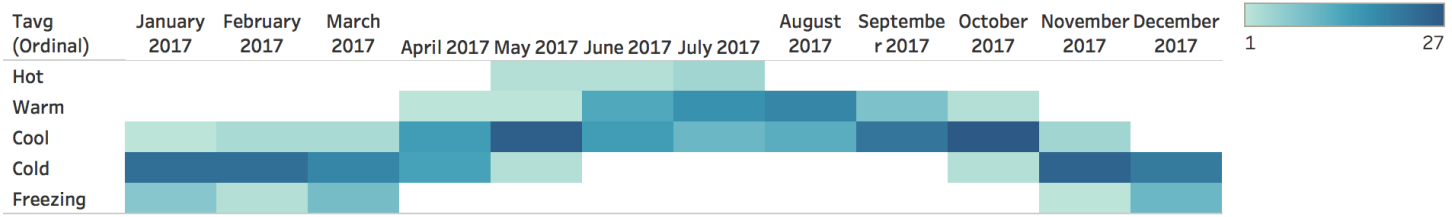
Temperature Value (Q)

NYC Daily Temperature from https://www.ncdc.noaa.gov/cdo-web/search

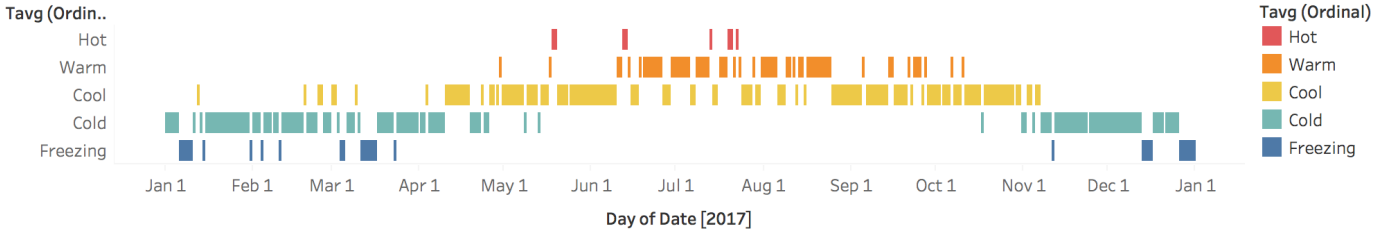
QxQ


Tavg
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
Jan 1
Feb 1
Mar 1
Apr 1
May 1
Jun 1
Jul 1
Aug 1
Sep 1
Oct 1
Nov 1
Dec 1
Jan 1
Day of Date [2017]

OxO


Tavg (Ordinal)
January 2017
February 2017
March 2017
April 2017
May 2017
June 2017
July 2017
August 2017
September 2017
October 2017
November 2017
December 2017
Hot
Warm
Cool
Cold
Freezing
1
27

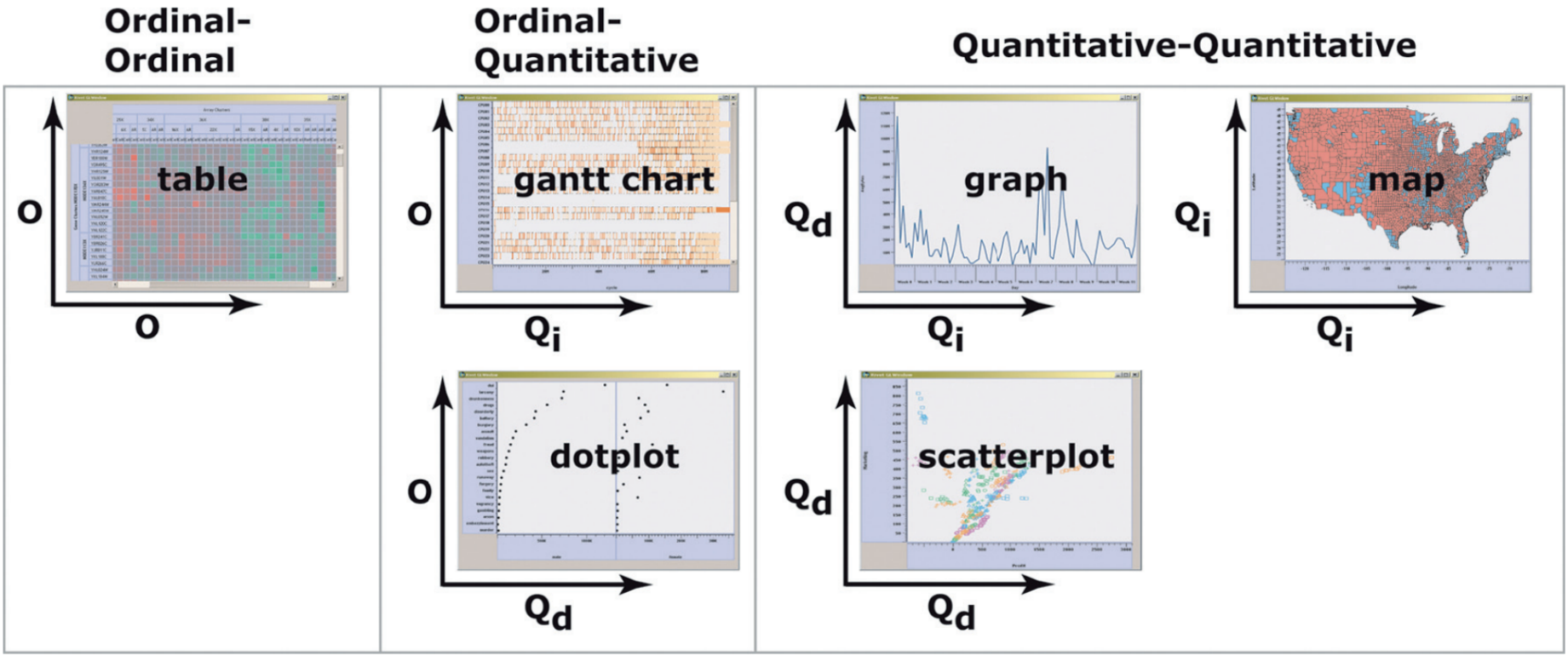
OxQ


Tavg (Ordin..
Hot
Warm
Cool
Cold
Freezing
Jan 1
Feb 1
Mar 1
Apr 1
May 1
Jun 1
Jul 1
Aug 1
Sep 1
Oct 1
Nov 1
Dec 1
Jan 1
Day of Date [2017]
Tavg (Ordinal)
Hot
Warm
Cool
Cold
Freezing

03

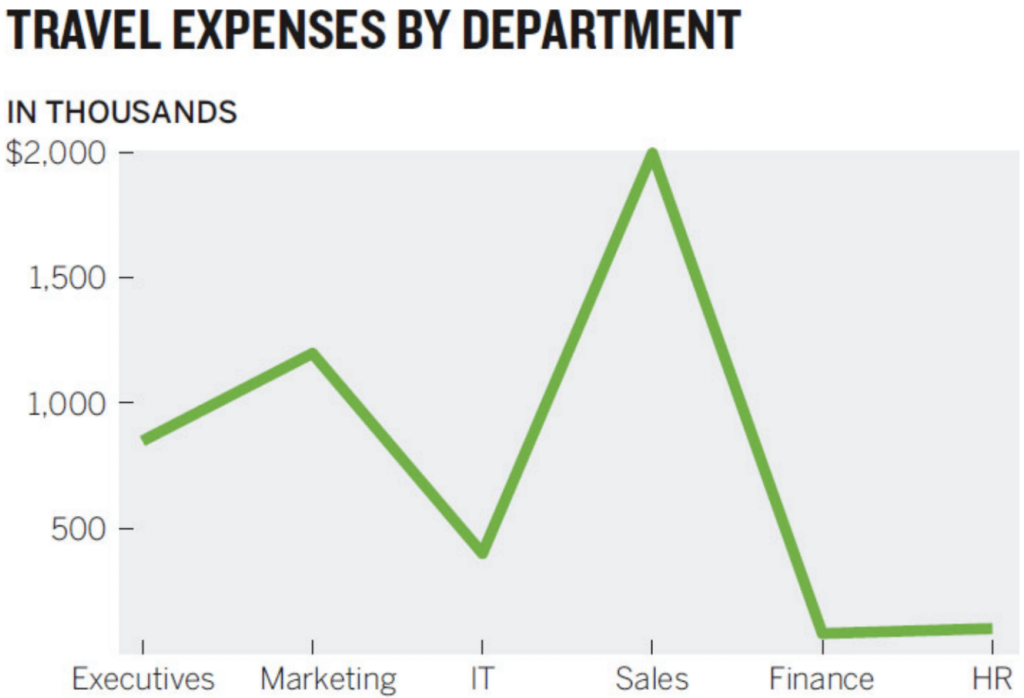
Which Viz Type for Which Data Type

Taxonomy of vis types



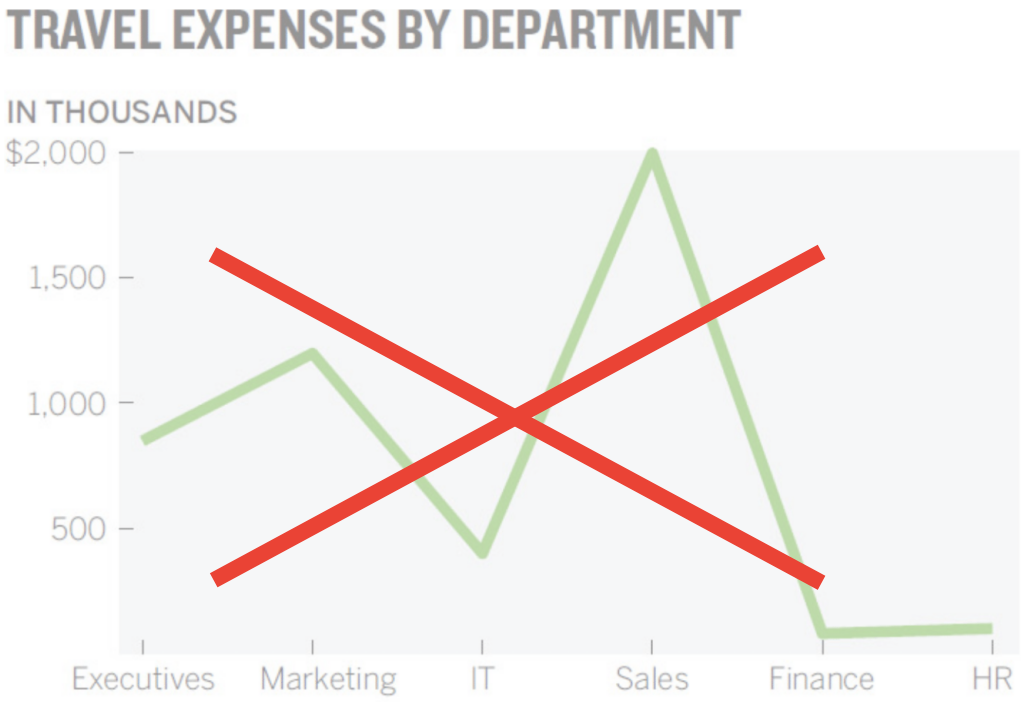
Stolte, et. al. *Polaris*, ACM 2008.

What do you see?



Berinato, Good Charts.

Categorical data should not be connected by a line: it misleadingly suggests an ordering.

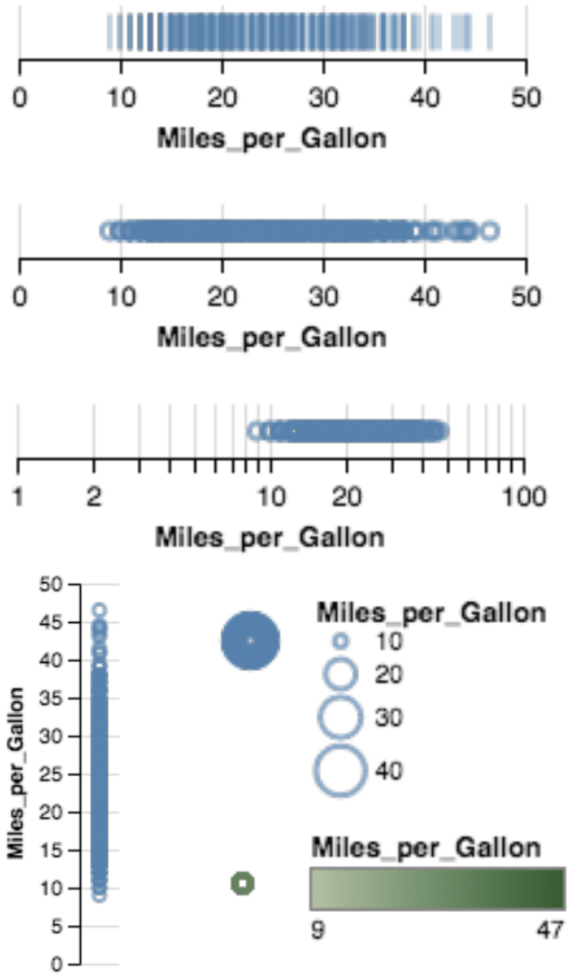


Berinato, Good Charts.

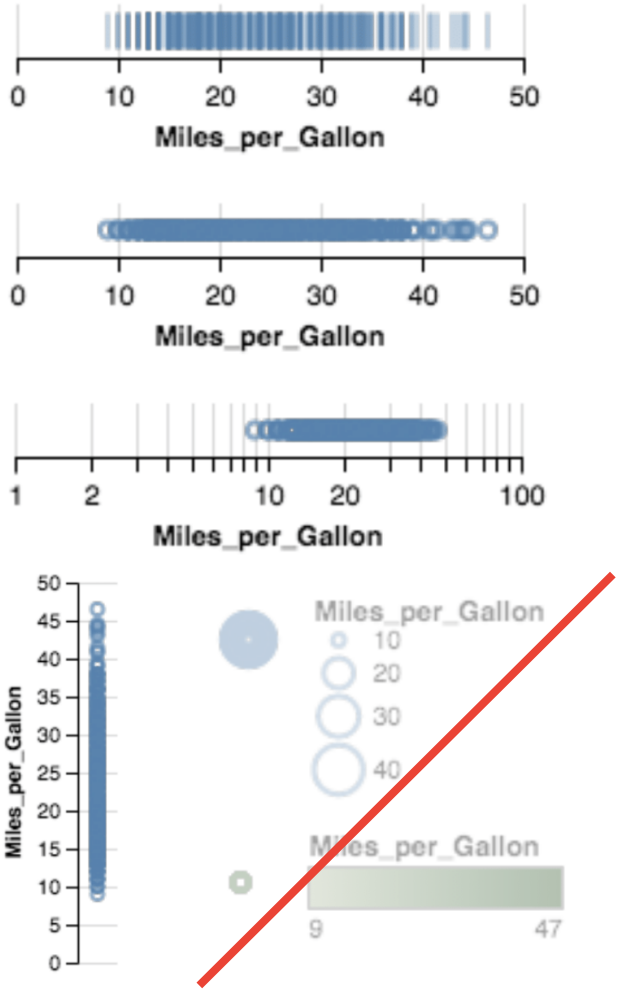
Example data: “mpg”

Model	Origin	Miles per gallon
"ford maverick"	USA	21.0
"datsun pl510"	Japan	27.0
"volkswagen 1131 deluxe sedan"	Germany	26.0
...		

1D Quantitative



1D Quantitative

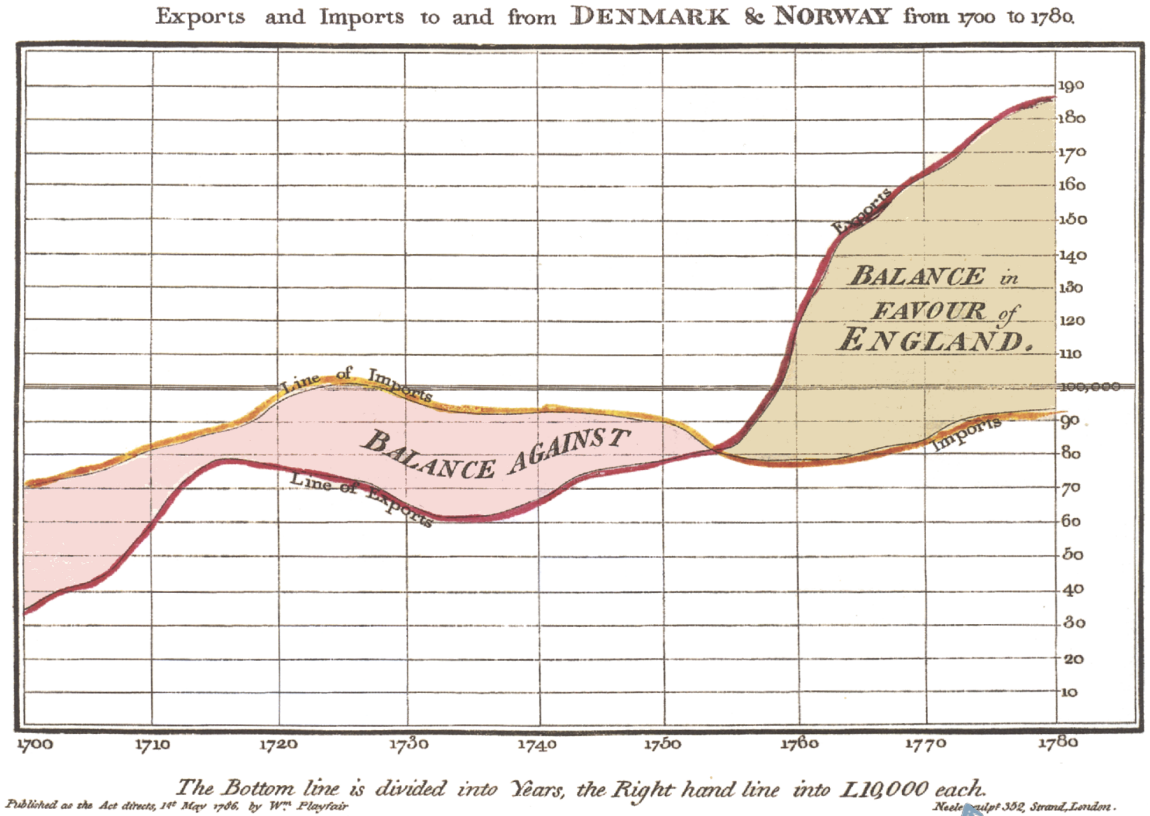


The relationship you do want to highlight is obscured.

(Encoding lecture: quantitative values best mapped to position to be expressive)

03
Deconstruct

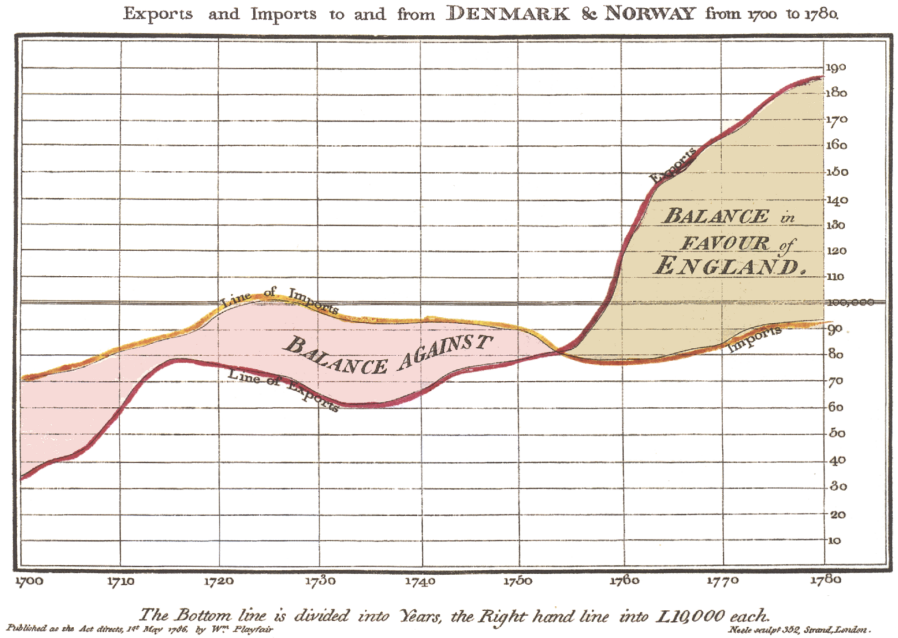
William Playfair (1786)



Inventor of line charts, bar charts, and pie charts.

British pounds

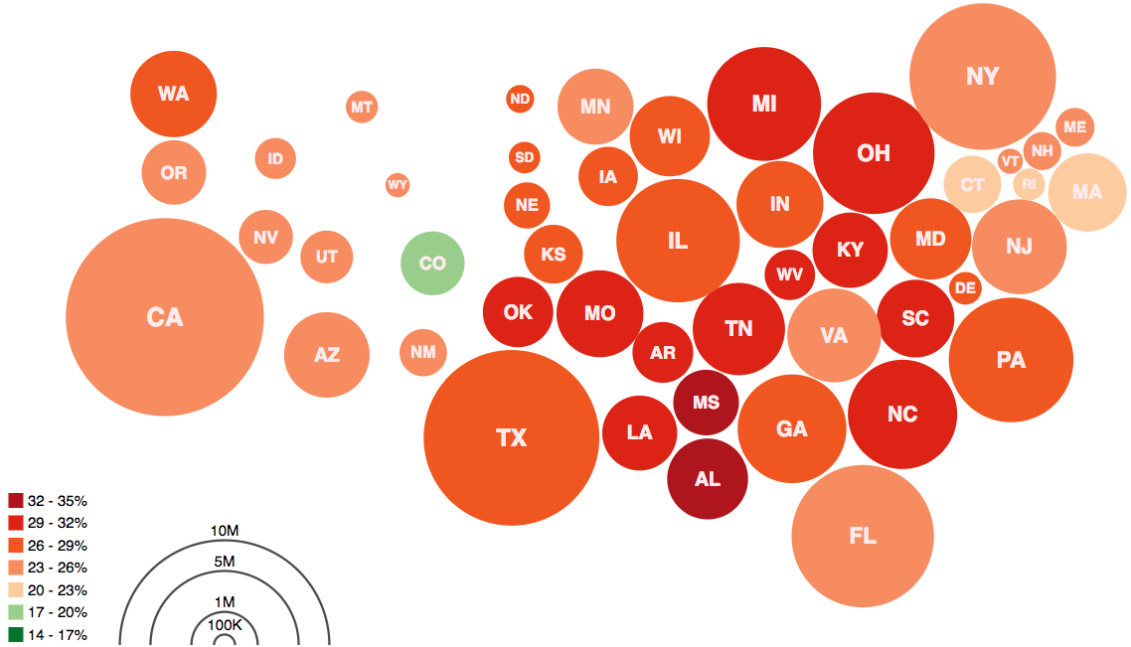
Image via Wikipedia



X-axis: year (Q)
Y-axis: currency (Q)
Color: imports/exports (C)

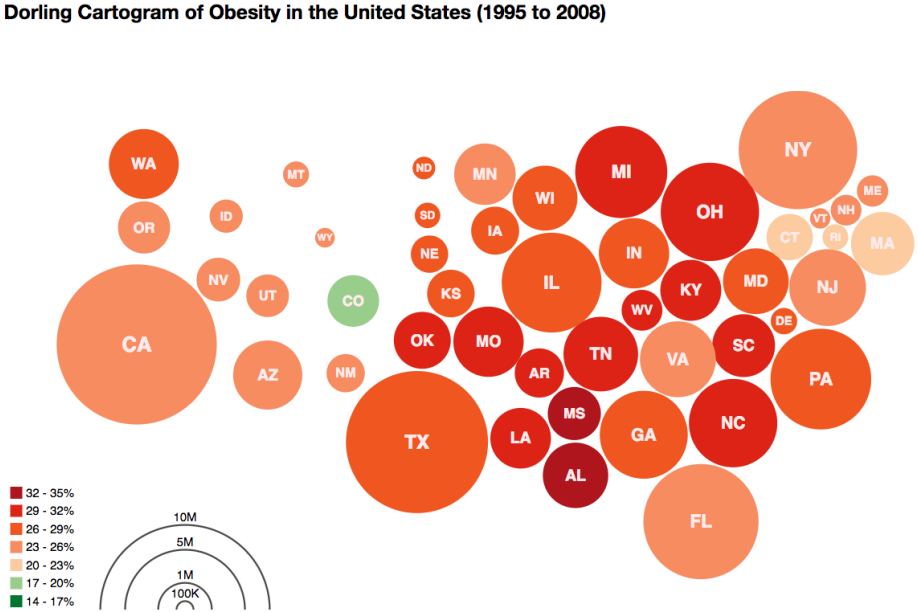
Dorling Cartogram

Dorling Cartogram of Obesity in the United States (1995 to 2008)



Absolute number of obese people per state and percentage of obese people, represented with circle area and color respectively. California dominates the map due to its large population size, while color indicates that Mississippi and Alabama are the states with the highest obesity rate.

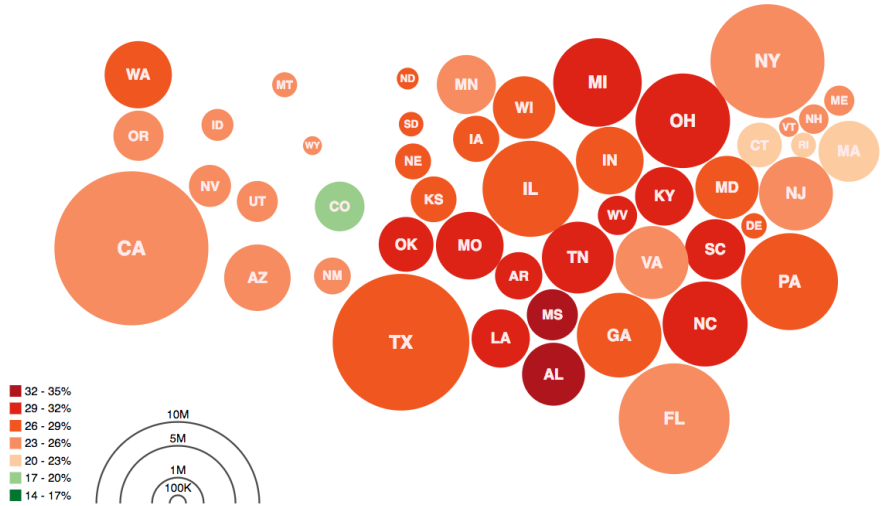
Source: National Center for Chronic Disease Prevention and Health Promotion



Circle Area: state population (Q)
Color: % obese, binned (O)
X-axis: ~longitude of state centroid (Q)
Y-axis: ~latitude of state centroid (Q)

Compare & Contrast

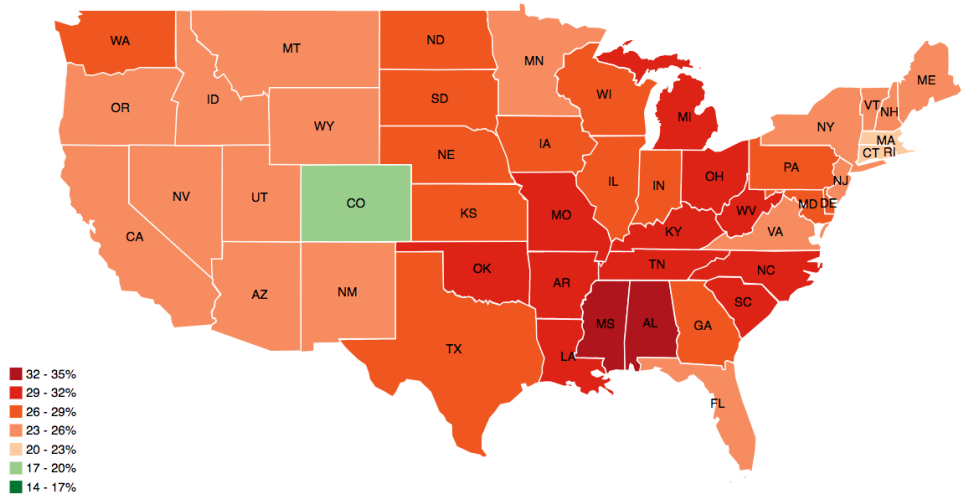
Dorling Cartogram of Obesity in the United States (1995 to 2008)



Absolute number of obese people per state and percentage of obese people, represented with circle area and color respectively. California dominates the map due to its large population size, while color indicates that Mississippi and Alabama are the states with the highest obesity rate.

Source: National Center for Chronic Disease Prevention and Health Promotion

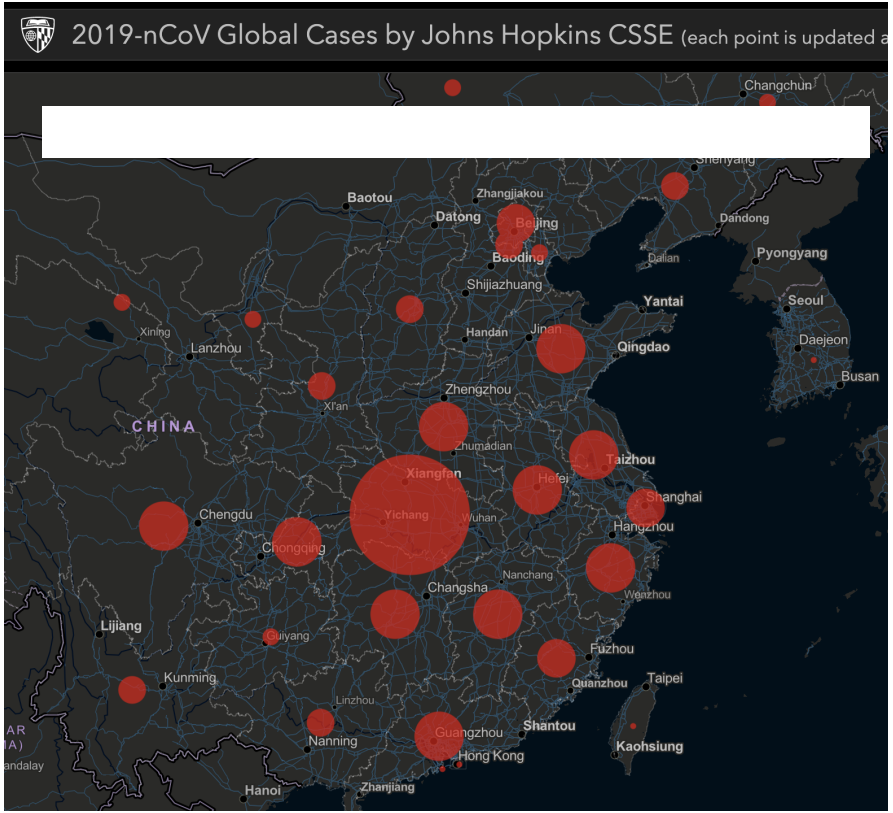
Choropleth Map of Obesity in the United States (1995 to 2008)

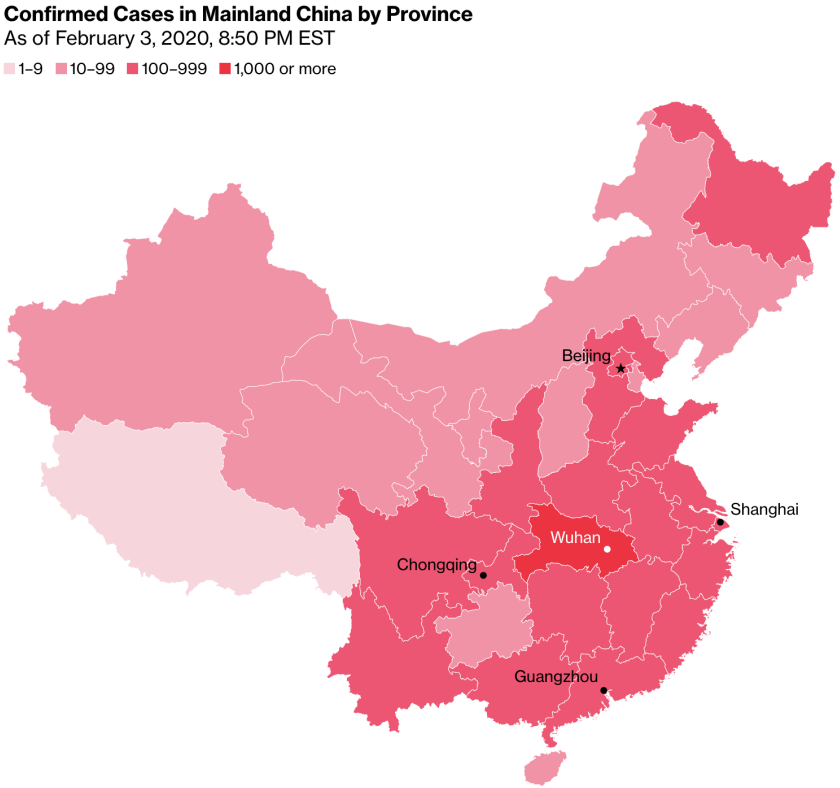


Present of population classified as "Obese" (Body Mass Index in excess of 30), by state. Press the "Play" button to watch the obesity epidemic evolve from 1995 onwards, or drag the slider to a particular year of interest.

Source: National Center for Chronic Disease Prevention and Health Promotion

Compare & Contrast



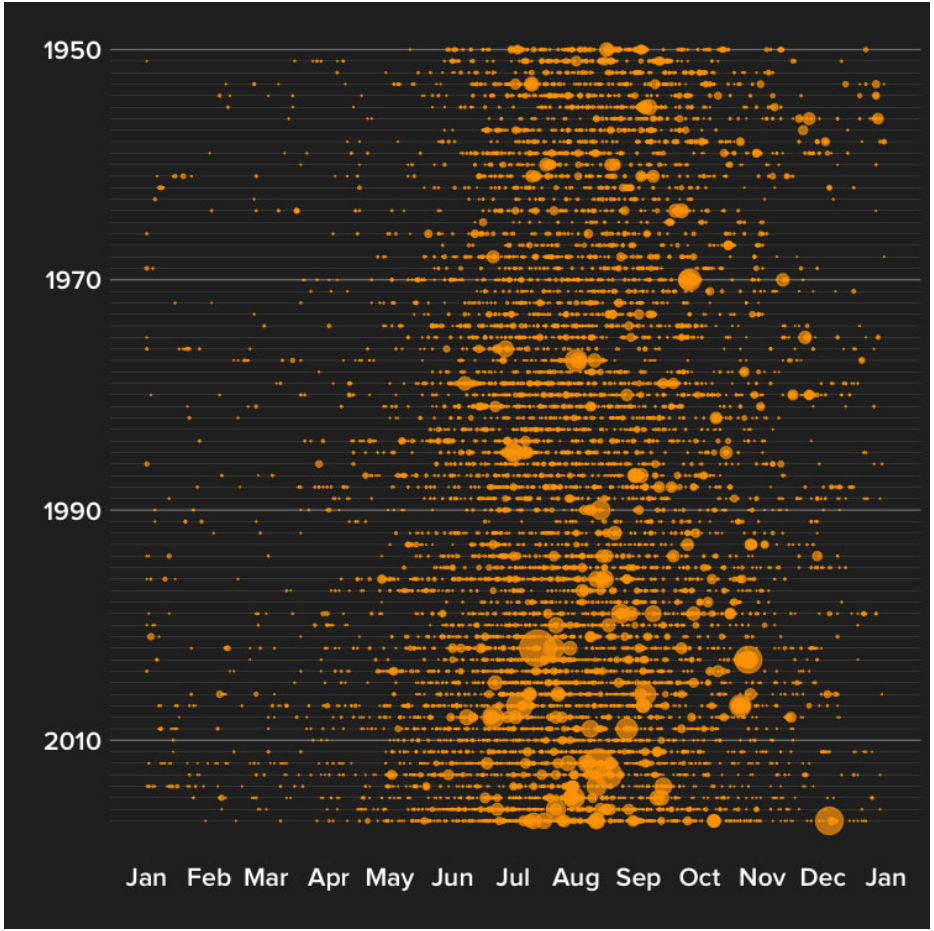


Left: John Hopkins dashboard, right: Bloomberg

Activity: Analyze and Re-design visualization

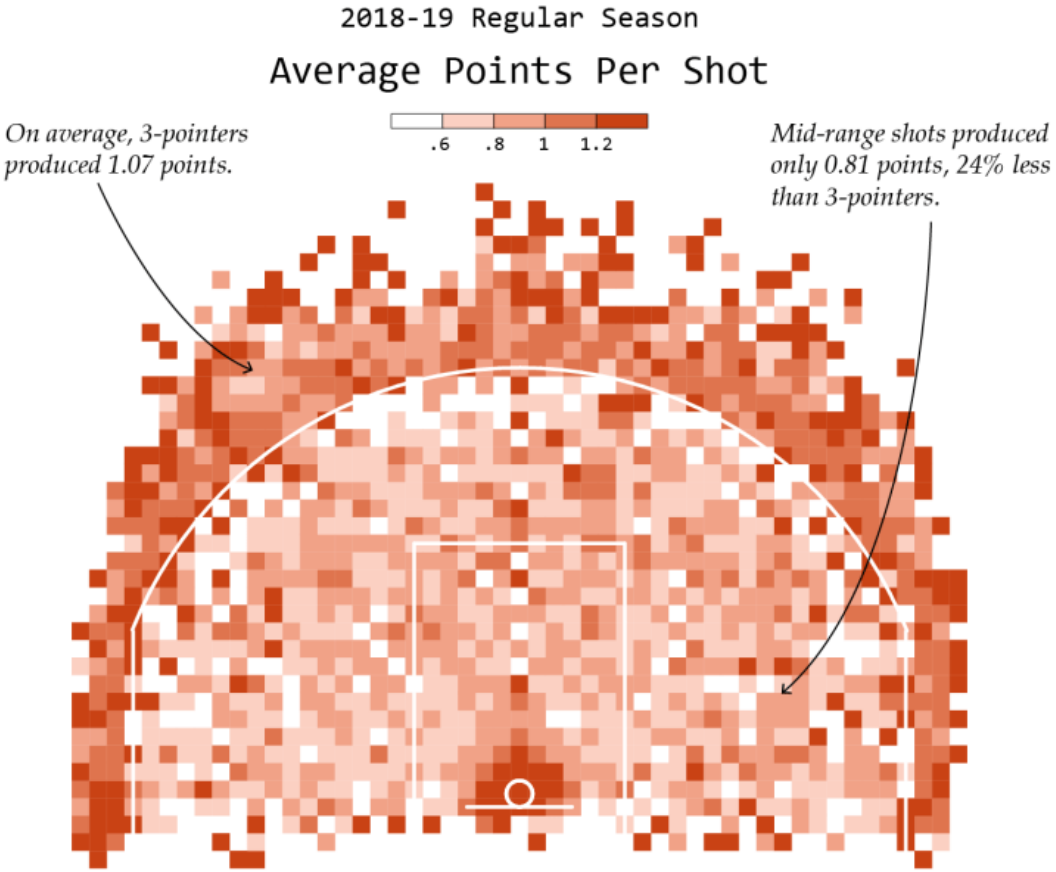
- As a group, choose 1 of the following 5 slides.
- Hop into a vector drawing app then:
  - Identify data variables (C,O,Q) and encodings.
  - Redesign another way to visualize the data. What different message does your redesign prioritize? (Subset of data is OK.)

Analyze and Re-design #1: California Wildfires

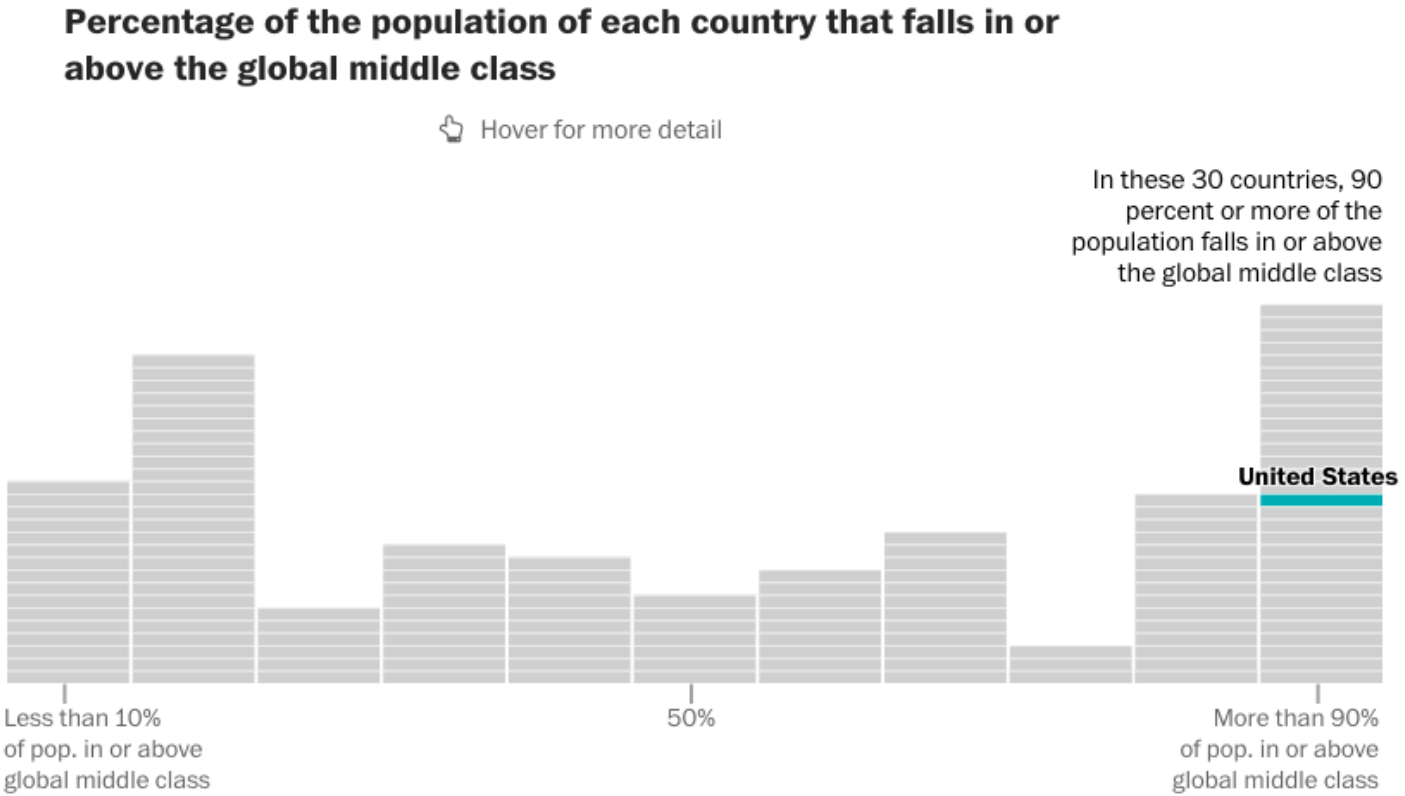


BuzzFeed Peter Aldhous

Analyze and Re-design #2: Basketball

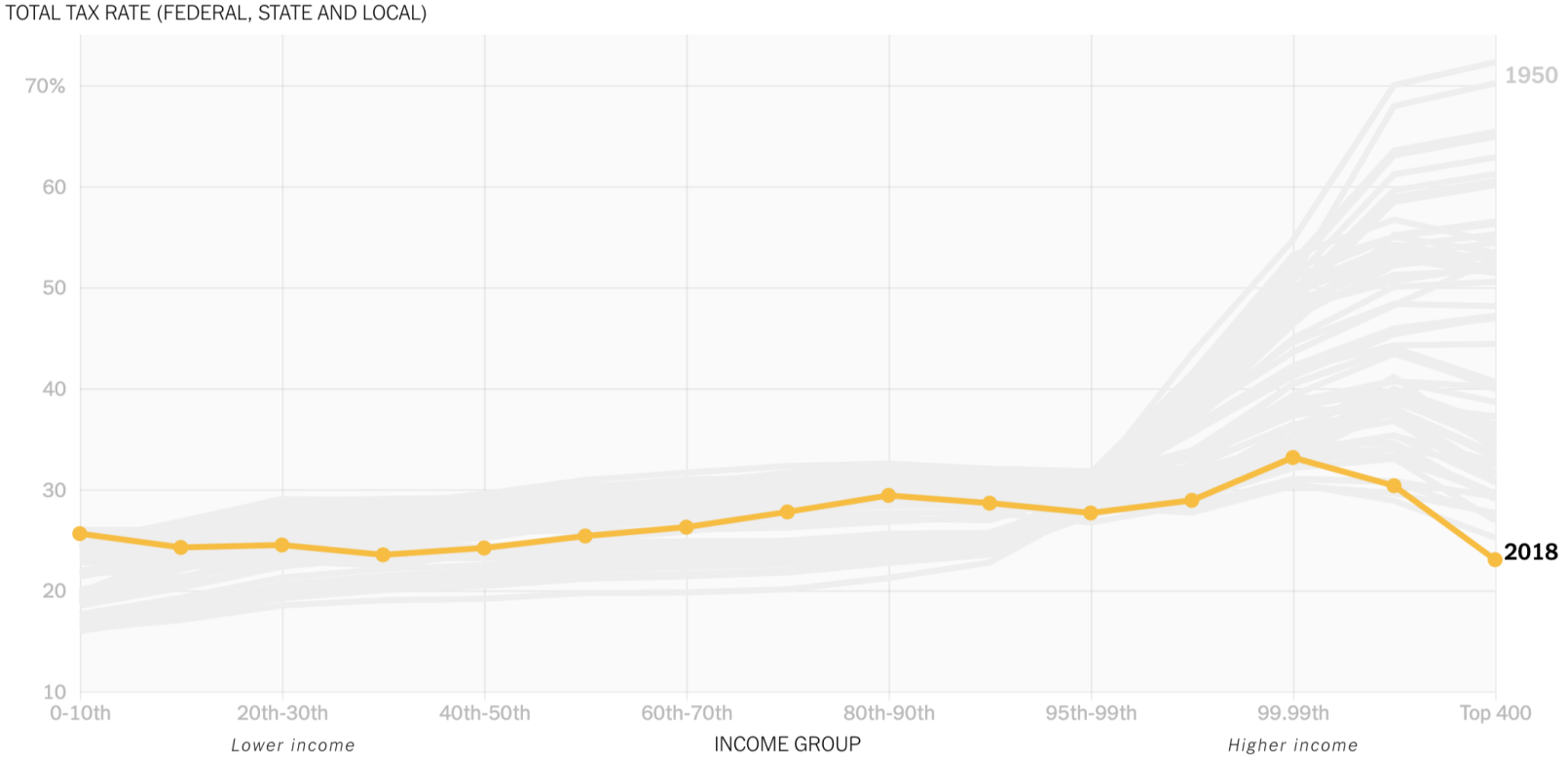


Analyze and Re-design #3: Global Middle Class

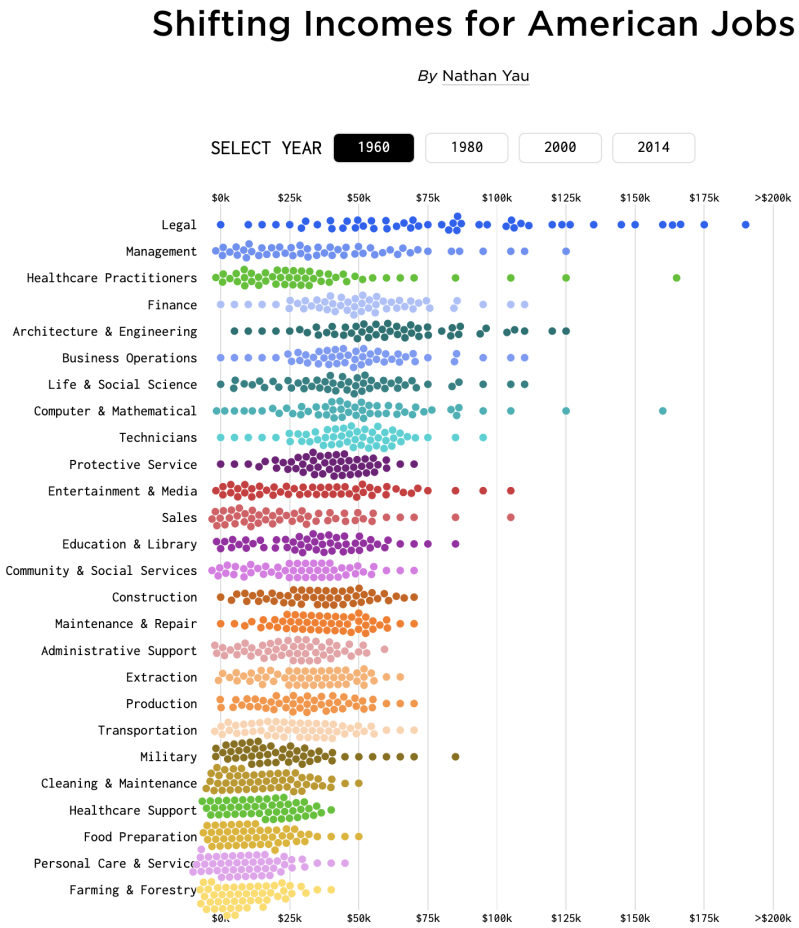


Washington Post

Analyze and Re-design #4: U.S. Total Tax Rate



Analyze and Re-design #5: American Job Incomes

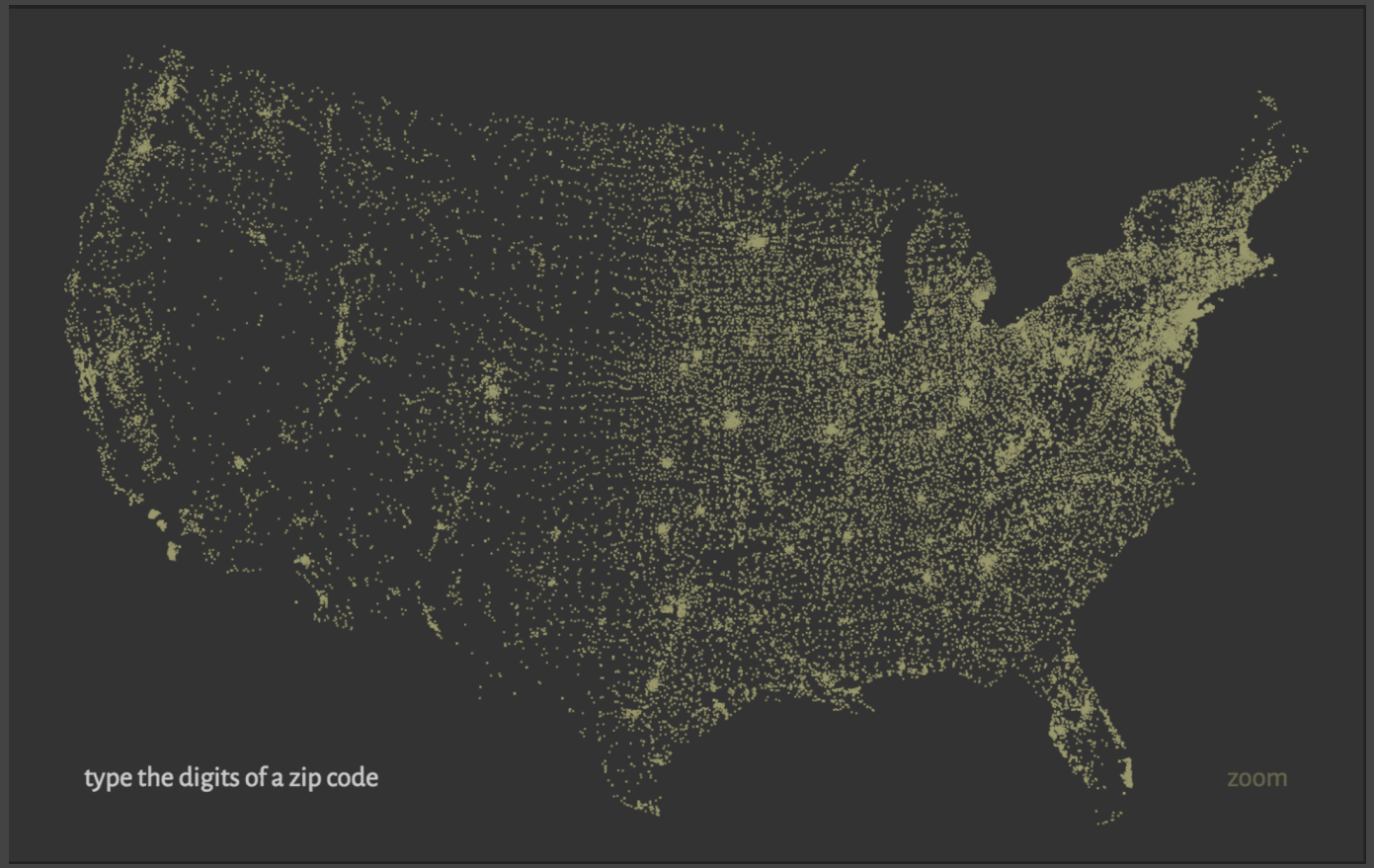


Last exercise: data type of zipcode?

2010_Census_Populations_by_Zip_Code

Zip Code	Total Population	Median Age	Total Males	Total Females	Total Households	Average Household Size
91371	1	73.5	0	1	1	1
90001	57110	26.6	28468	28642	12971	4.4
90002	51223	25.5	24876	26347	11731	4.36
90003	66266	26.3	32631	33635	15642	4.22
90004	62180	34.8	31302	30878	22547	2.73
90005	37681	33.9	19299	18382	15044	2.5
90006	59185	32.4	30254	28931	18617	3.13
90007	40920	24	20915	20005	11944	3
90008	32327	39.7	14477	17850	13841	2.33
90010	3800	37.8	1874	1926	2014	1.87
90011	103892	26.2	52794	51098	22168	4.67
90012	31103	36.3	19493	11610	10327	2.12
90013	11772	44.6	7629	4143	6416	1.26
90014	7005	44.8	4471	2534	4109	1.34

Ben Fry, Zipdecode (1999)



Summary: Data Models

- A dataset measurement can be interpreted as many different semantic data types.
- Different data types determine the appropriate (set of) visual encodings.
  - Visualization software UIs leverage this to help you.
- Different visual encodings highlight different underlying data relationships.