

Data Visualization

Andreas Paepcke

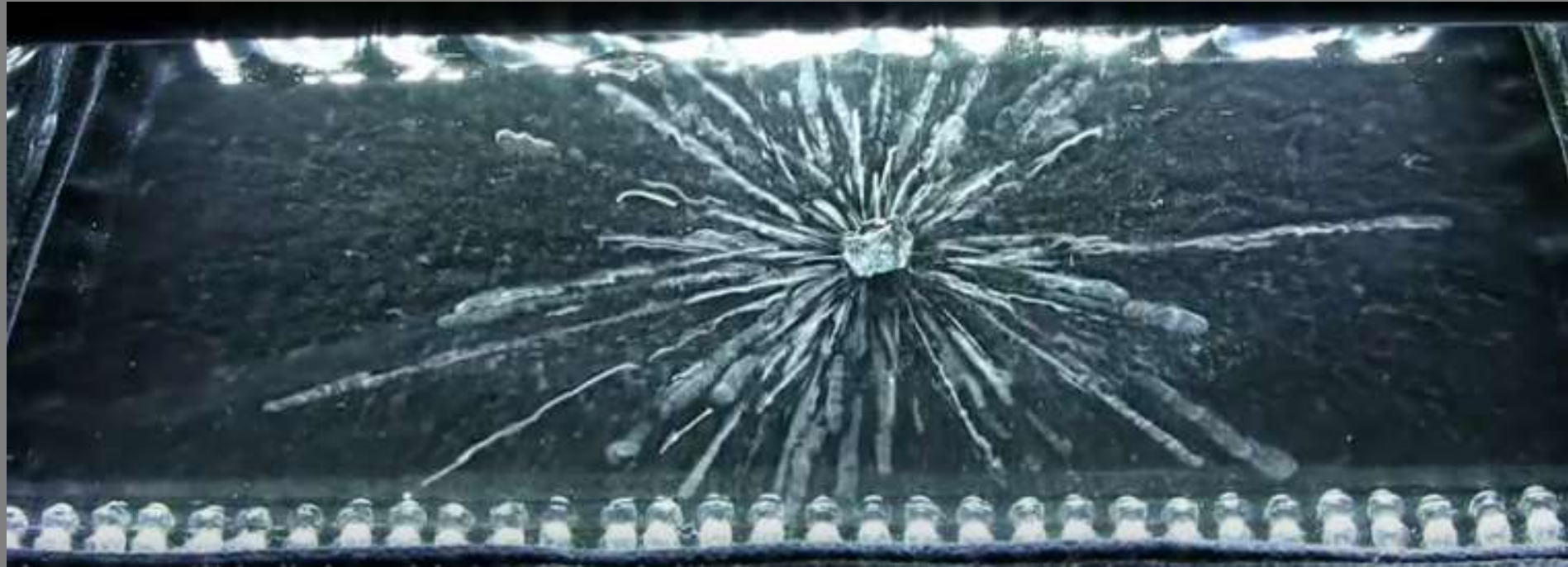
Roadmap

- Purpose of visualization
- Exercise
- History
- Scale types and their viz
- Multidimensional data
- Animation
- What to avoid; deception
- Exercise

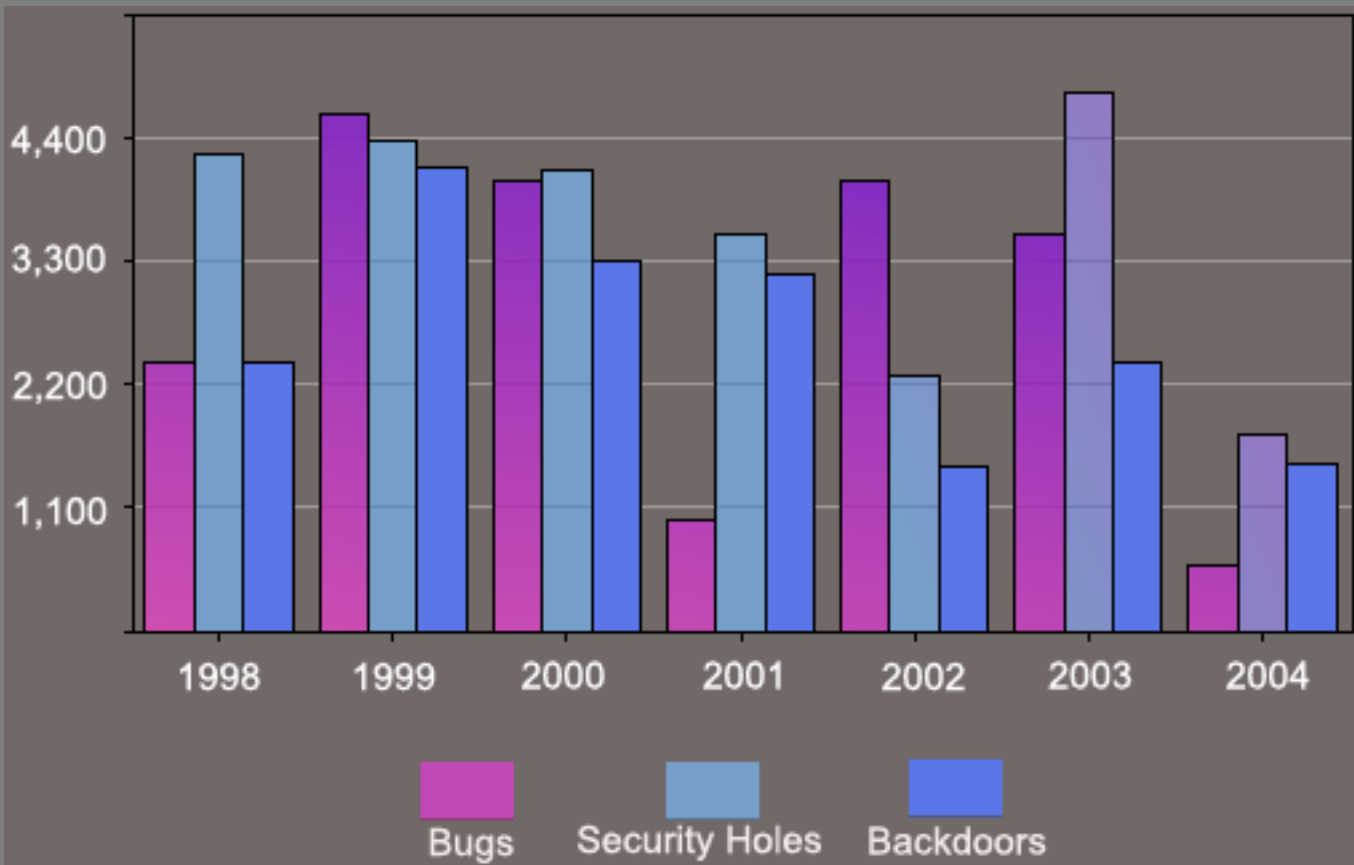
A Piece of Rock



A Piece of Rock in Visualization Device



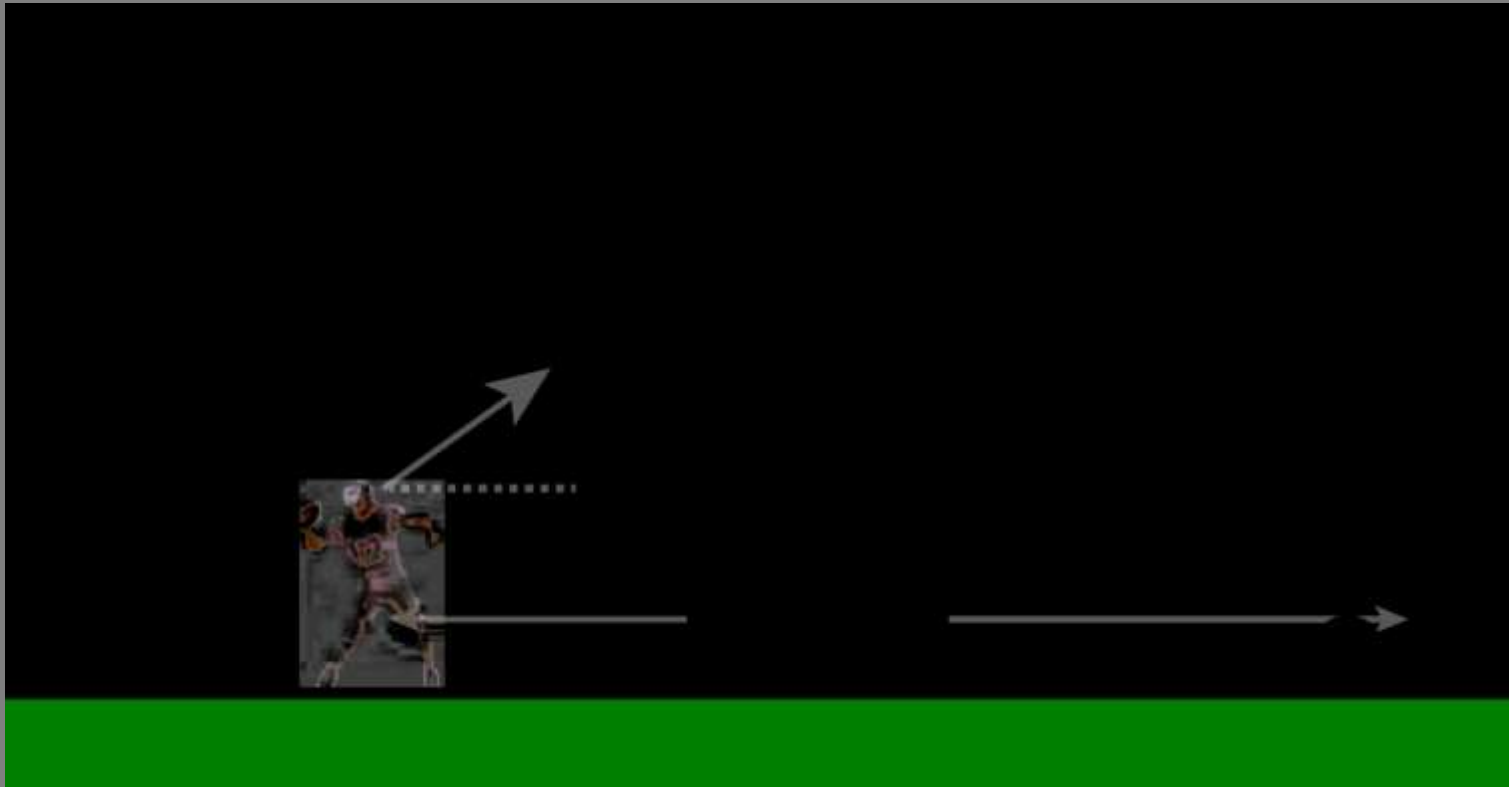




Reasons for visualizing?

Why Visualize?

Explanation



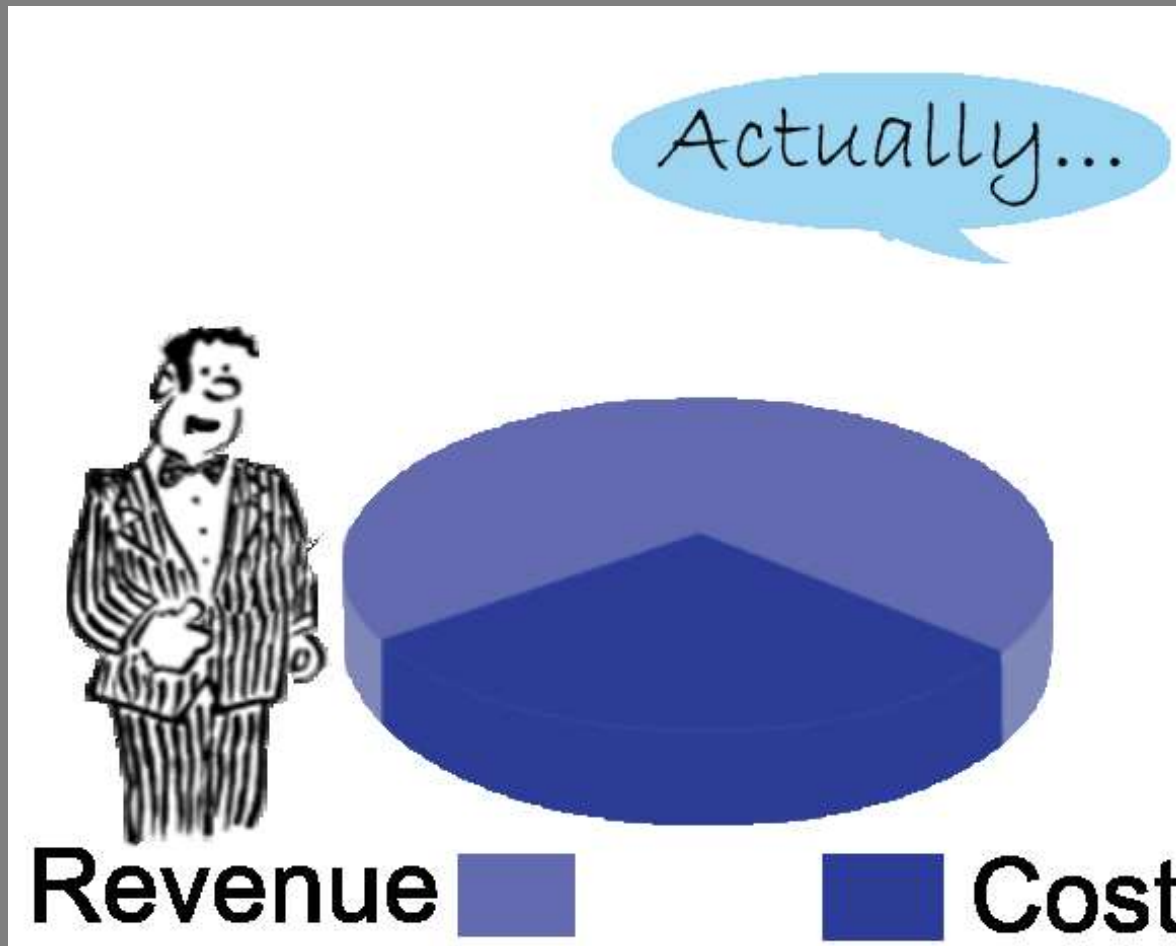
Why Visualize?

Convince



Why Visualize?

Deceive



Why Visualize?

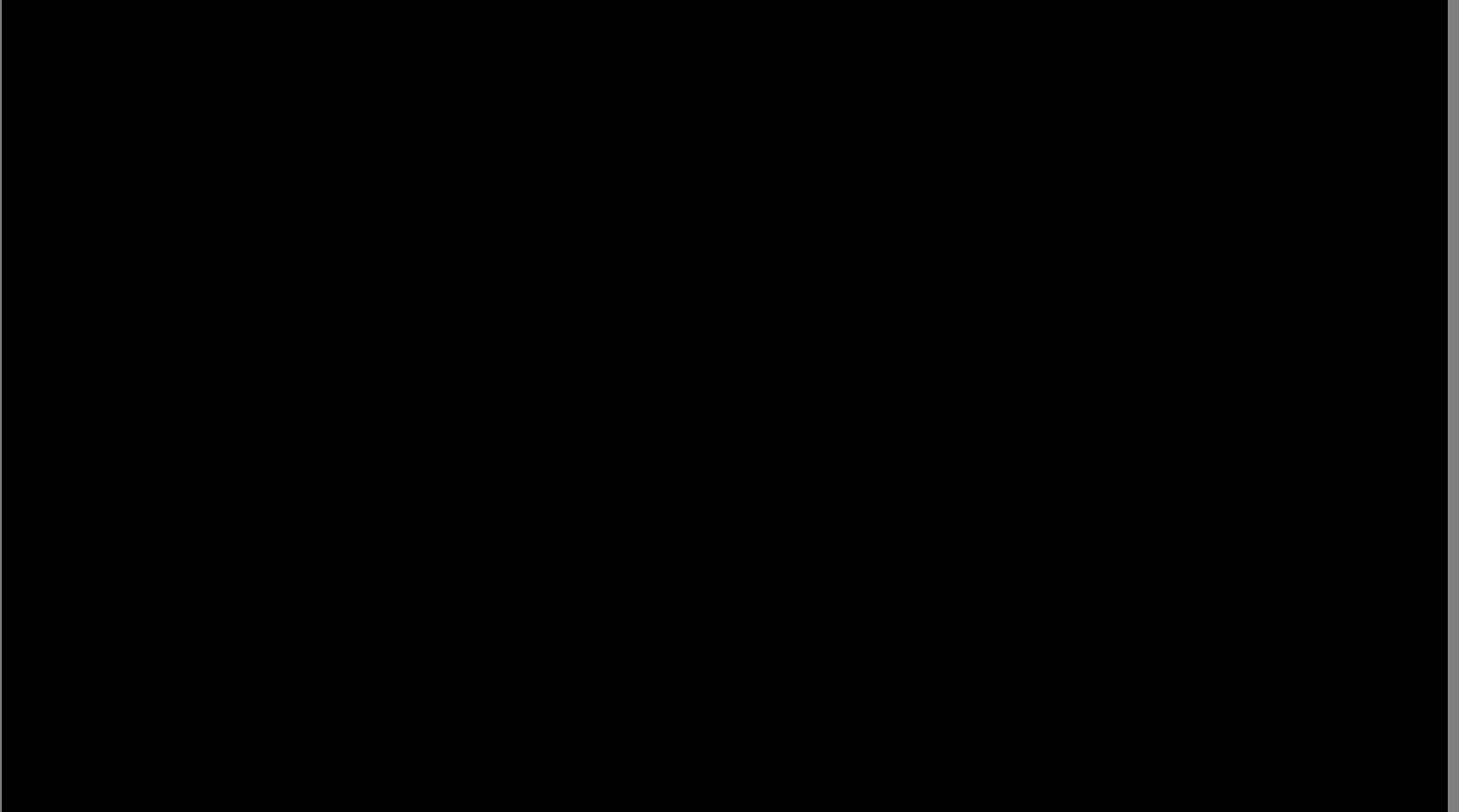
Discover

Wait for it...

a few slides on...

Why Visualize?

Art

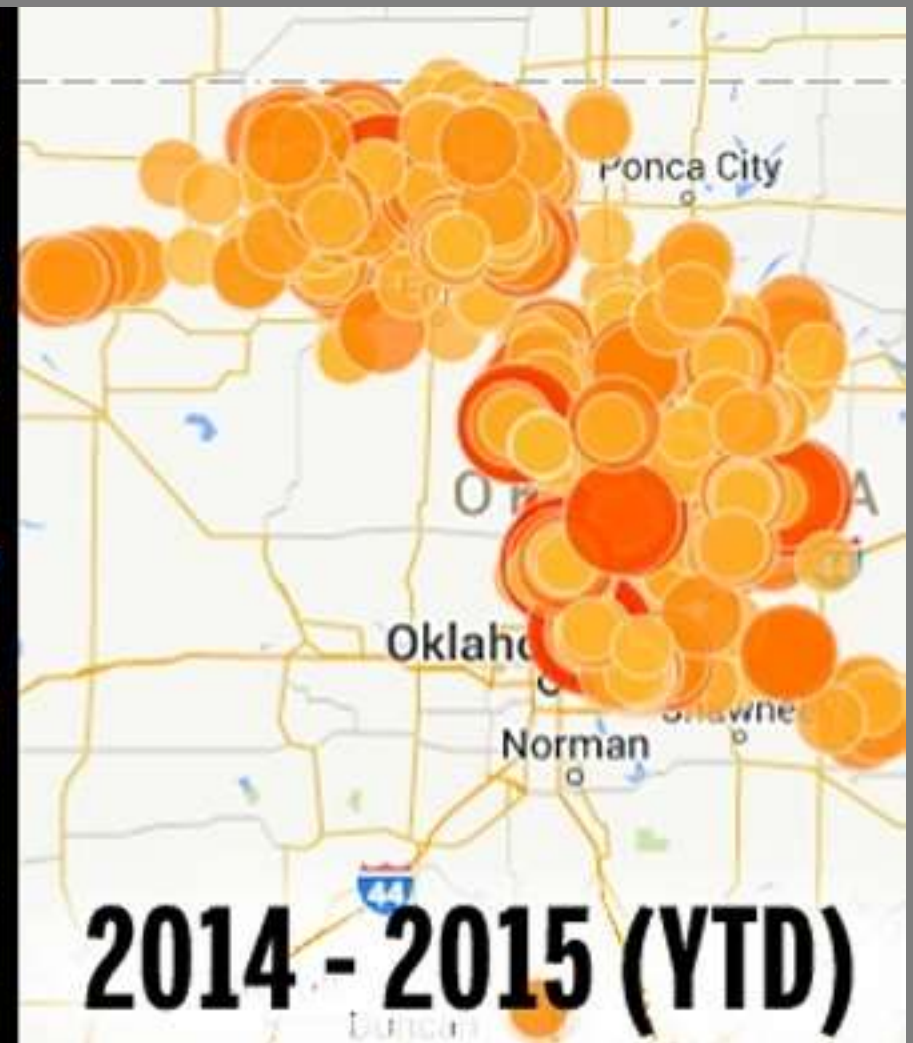


Slide cursor to bottom of black rectangle

Data to Music



2005 – June 2015



Exercise: Evidence of Alzheimer's

| Novel | Age | Distinct Words | Repeated phrases | Indefinite Words (%) |
|--------------|-----|----------------|------------------|----------------------|
| Styles | 28 | 5027 | 7623 | 0.27 |
| Adversary | 32 | 5576 | 7320 | 0.39 |
| Ackroyd | 34 | 4833 | 7905 | 0.48 |
| Orient | 43 | 4692 | 8184 | 0.3 |
| Appointment | 47 | 4985 | 8071 | 0.39 |
| Curtain | 50 | 5131 | 7846 | 0.43 |
| Towards Zero | 51 | 4941 | 7680 | 0.5 |
| Announced | 59 | 5181 | 7596 | 0.45 |
| Destination | 63 | 5442 | 7426 | 0.36 |
| Ordeal | 67 | 4440 | 8221 | 0.58 |
| Clocks | 72 | 4827 | 8014 | 0.61 |
| Endless | 76 | 4159 | 8559 | 0.78 |
| Frankfurt* | 79 | 5583 | 7418 | 0.71 |
| Nemesis | 80 | 4631 | 8103 | 0.66 |
| Elephants | 81 | 3762 | 8821 | 1.02 |
| Postern | 82 | 4275 | 8249 | 1.23 |

Analysis of Agatha Christie novels

Repeated phrases:

“That sort of thing”

“All things considered”

Indefinite words:

“A horrible incident shocked Lord Ackley's staff”

VS.

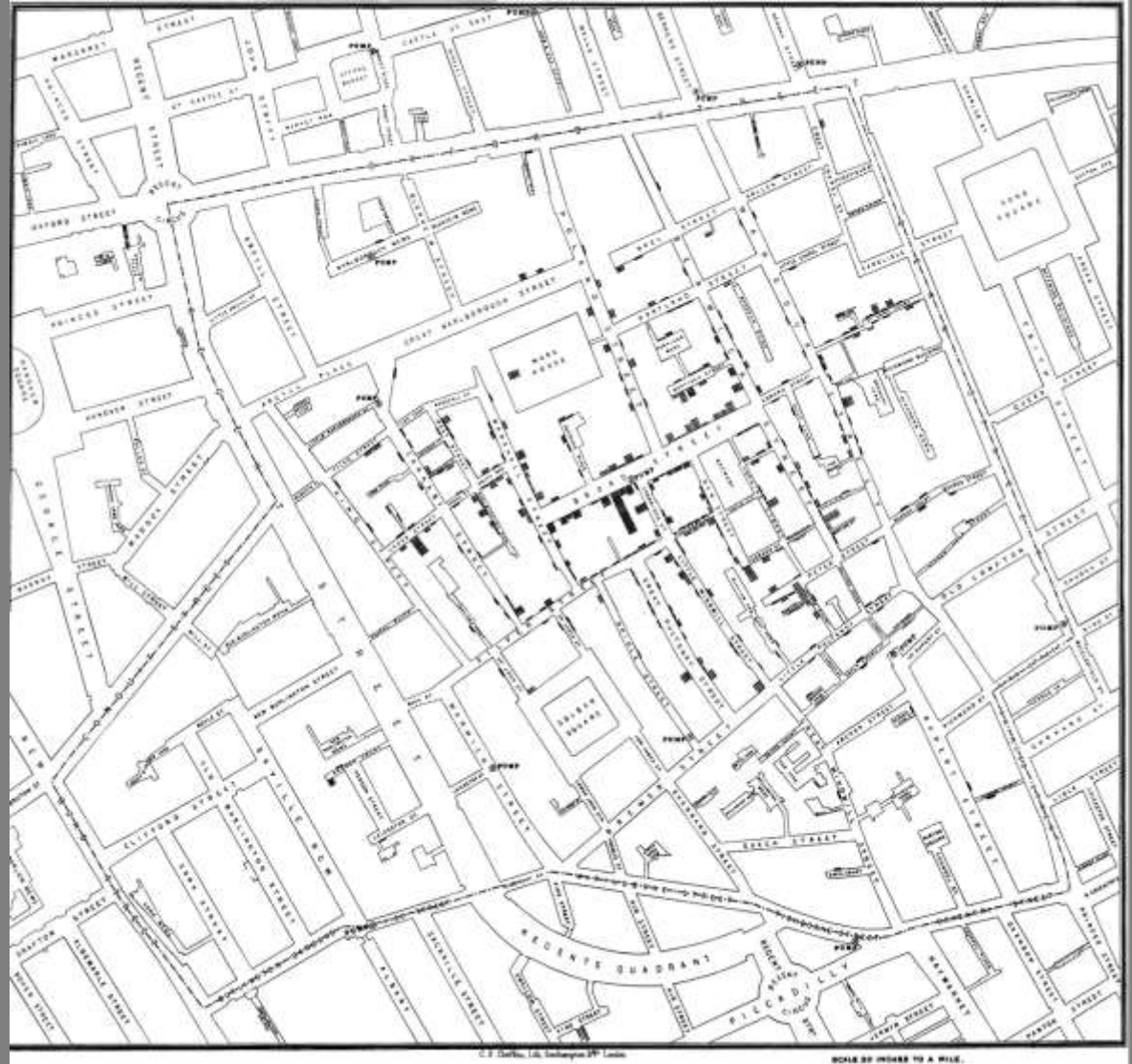
“Something horrible shocked Lord Ackley's staff”

Roadmap

- Purpose of visualization
- Exercise
- **History**
- Scale types and their viz
- Multidimensional data
- Animation
- What to avoid; deception
- Exercise

John Snow: London Cholera Map

1854 Cholera outbreak
Soho, London:



John Snow: London Cholera Map



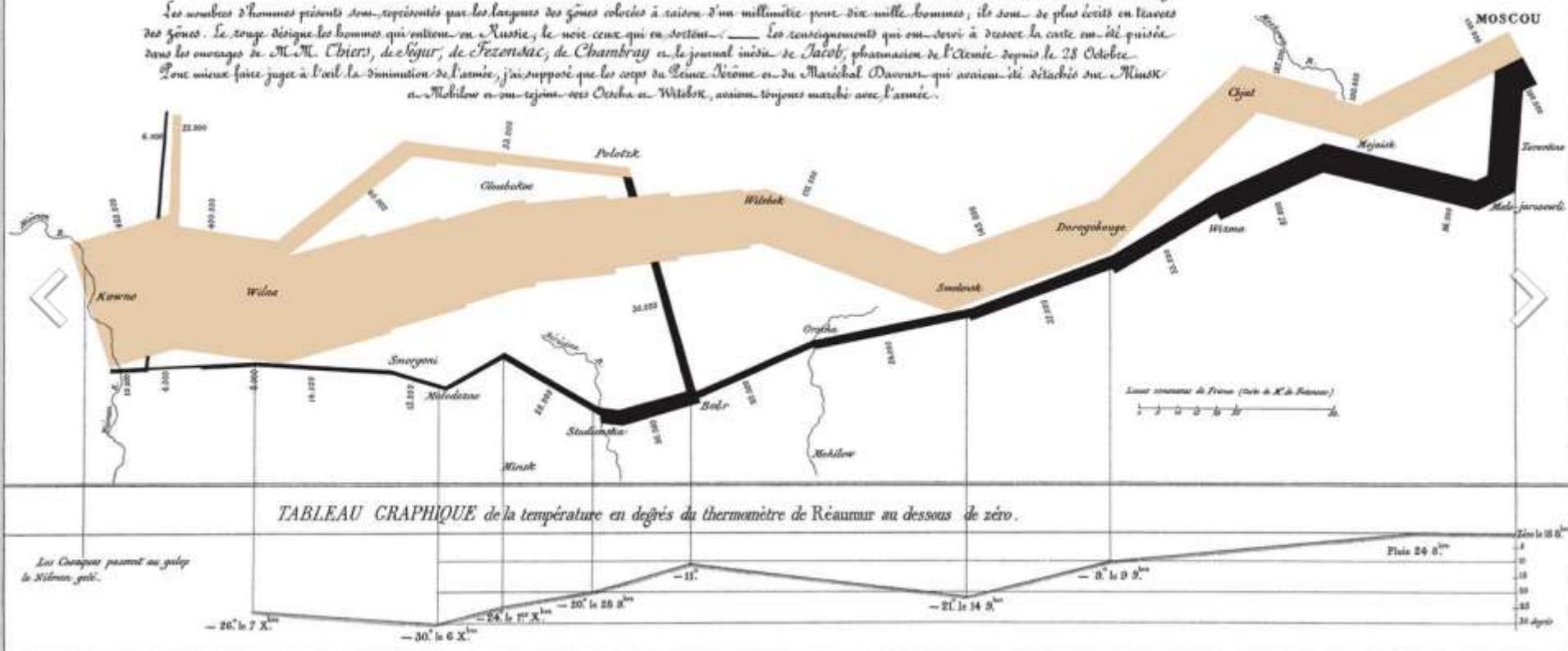
Charles Minard 1869: Napoleon Expedition

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessiné par M. Minard, Ingénieur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

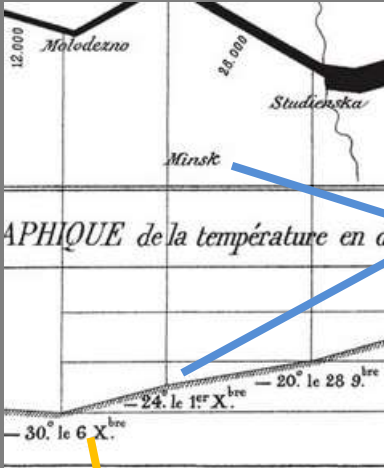
Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes, ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Général Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mohilew n'en rejoignent les Orcha et Witebsk, avaient toujours marché avec l'armée.



Dessiné par Minard, Ingénieur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

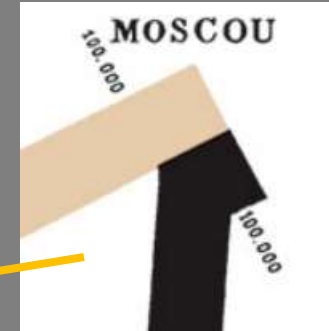
Imp. 101, Bignon et Borel.



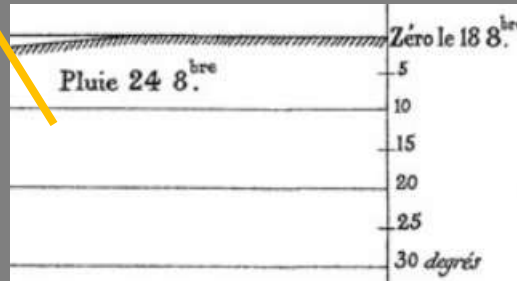
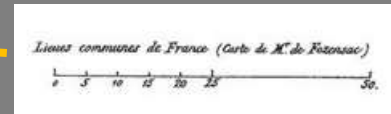
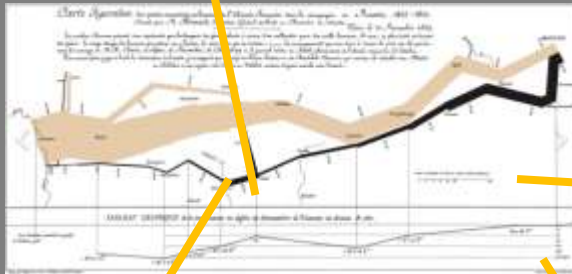
Location (labels)

Direction (color)

Location/Time (relative positioning)



Distance (scale)



Température (grid)



Troup strength (thickness)

Roadmap

- Purpose of visualization
- Exercise
- History
- **Scale types and their viz**
- Multidimensional data
- Animation
- What to avoid; deception
- Exercise

Variables Represent Data of Specific Types

Population and steel production in different world regions (Jan-Aug 2015):

| | Europe | North America | Australia |
|------------------|------------------|-----------------|----------------|
| World Region | Europe | North America | Australia |
| Population | 743M | 529M | 23M |
| Steel Production | 136.792 megatons | 75.516 megatons | 3.265 megatons |

Variable “*worldRegion*” {Europe, North America, Australia}

Variable “*population*” a number ≥ 0

Variable “*steelProduction*” a number ≥ 0

Types of Scales

Nominal \equiv Categorical:

- Mutually exclusive
- No numerical significance

Color:

Green Red Yellow

Products:

Laptops Tables


If only two categories: *Dichotomous*

Nominal: “Name”

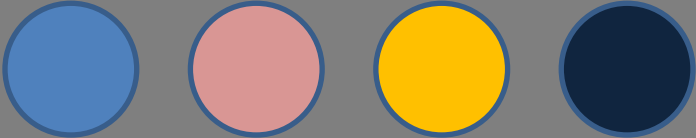
Example Visualizations

Nominal \equiv Categorical:


Shape:



Color:



Labels: Vanilla Chocolate Strawberry



Types of Scales

Ordinal:

- Order is significant...
- Difference is not

Milk shakes:

Small Medium Large

Mood:

Unhappy Neutral Happy Ecstatic

Ordinal: “Order”

Example Visualizations

Ordinal:

Order: Dissatisfied  Satisfied  Highly satisfied 

Color intensity:   

Types of Scales

Interval \equiv Continuous:

- Order is significant...
- Differences known
- No zero value

Time:

| | | |
|------|------|------|
| 10am | 11am | 12pm |
|------|------|------|

Temperature:
(Fahrenheit)

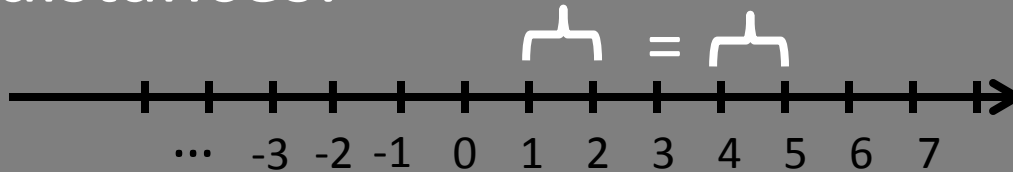
| | | | |
|-----|-----|-----|------|
| 70F | 80F | 90F | 100F |
|-----|-----|-----|------|

Interval: “Inter-value”

Example Visualizations

Interval:

Use distances:



Use height:



Types of Scales

Ratio:

- Order is significant...
- Differences known
- Zero value exists

Width:

| | | |
|-----|------|------|
| 0in | 10in | 20in |
|-----|------|------|

Temperature:
(Kelvin)

| | | | |
|-----|-----|-----|------|
| 70K | 80K | 90K | 100K |
|-----|-----|-----|------|

Ratio: “Comparable”

Putting it Together

click to highlight

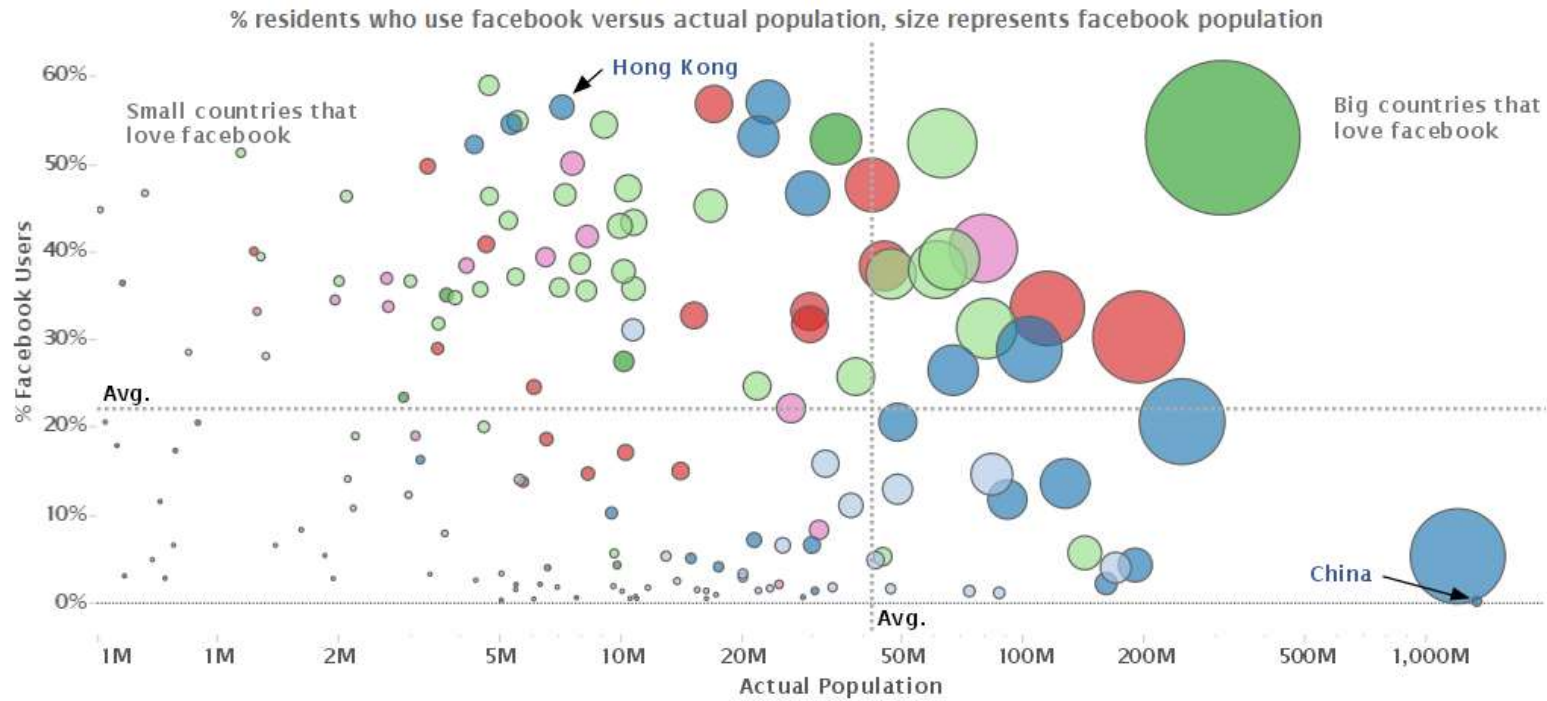
- Africa
- Asia and Australasia
- Europe
- Latin and South America
- Middle East
- North America and Caribbean

Show me these continents:

(All)

Countries of at least 500k population.
Source wikipedia

created by
@pgilks



Putting it Together

click to highlight

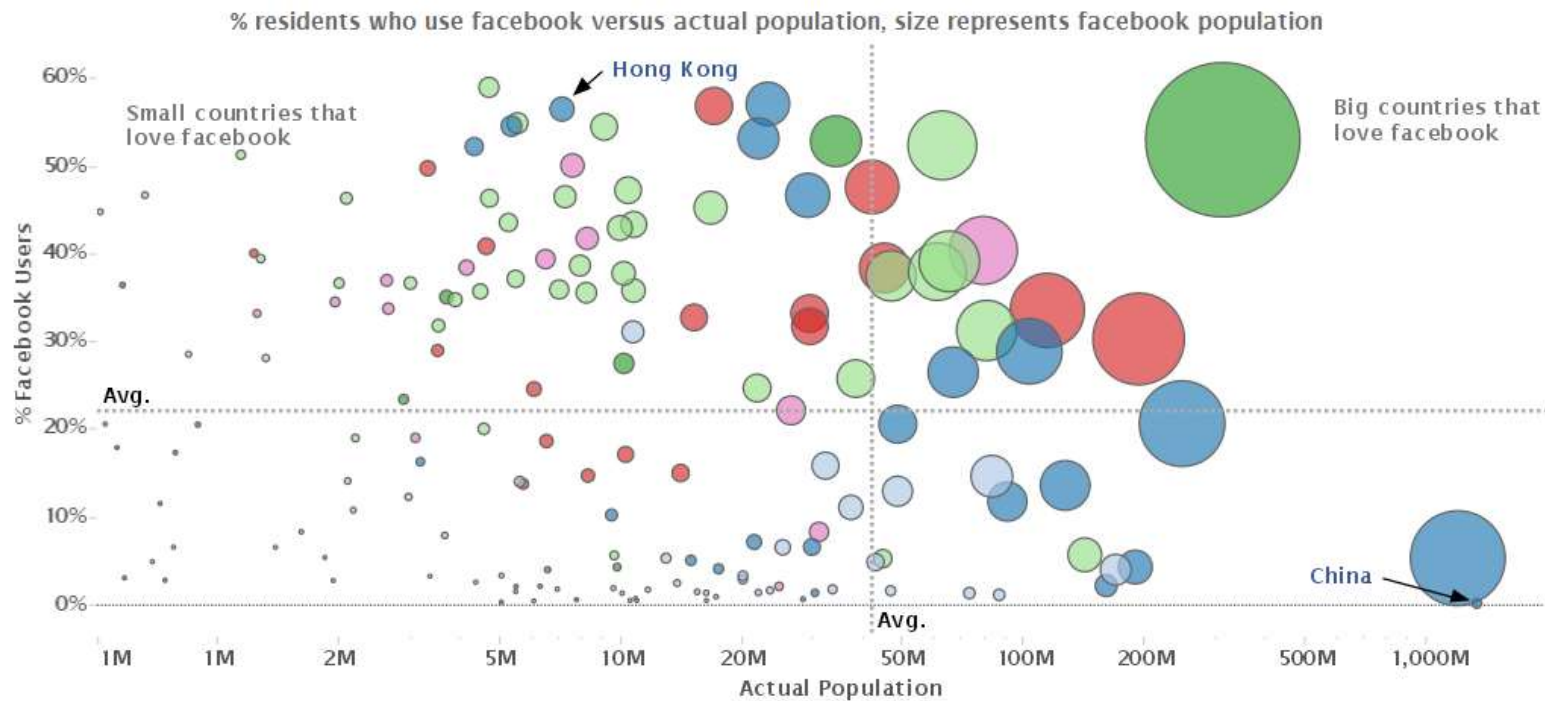
- Africa
- Asia and Australasia
- Europe
- Latin and South America
- Middle East
- North America and Caribbean

Show me these continents:

(All)

Countries of at least 500k population.
Source wikipedia

created by
@pgilks



Position-X:

Actual population; ratio

Position-Y:

Percentage Facebook users; ratio

Color:

Continents; nominal

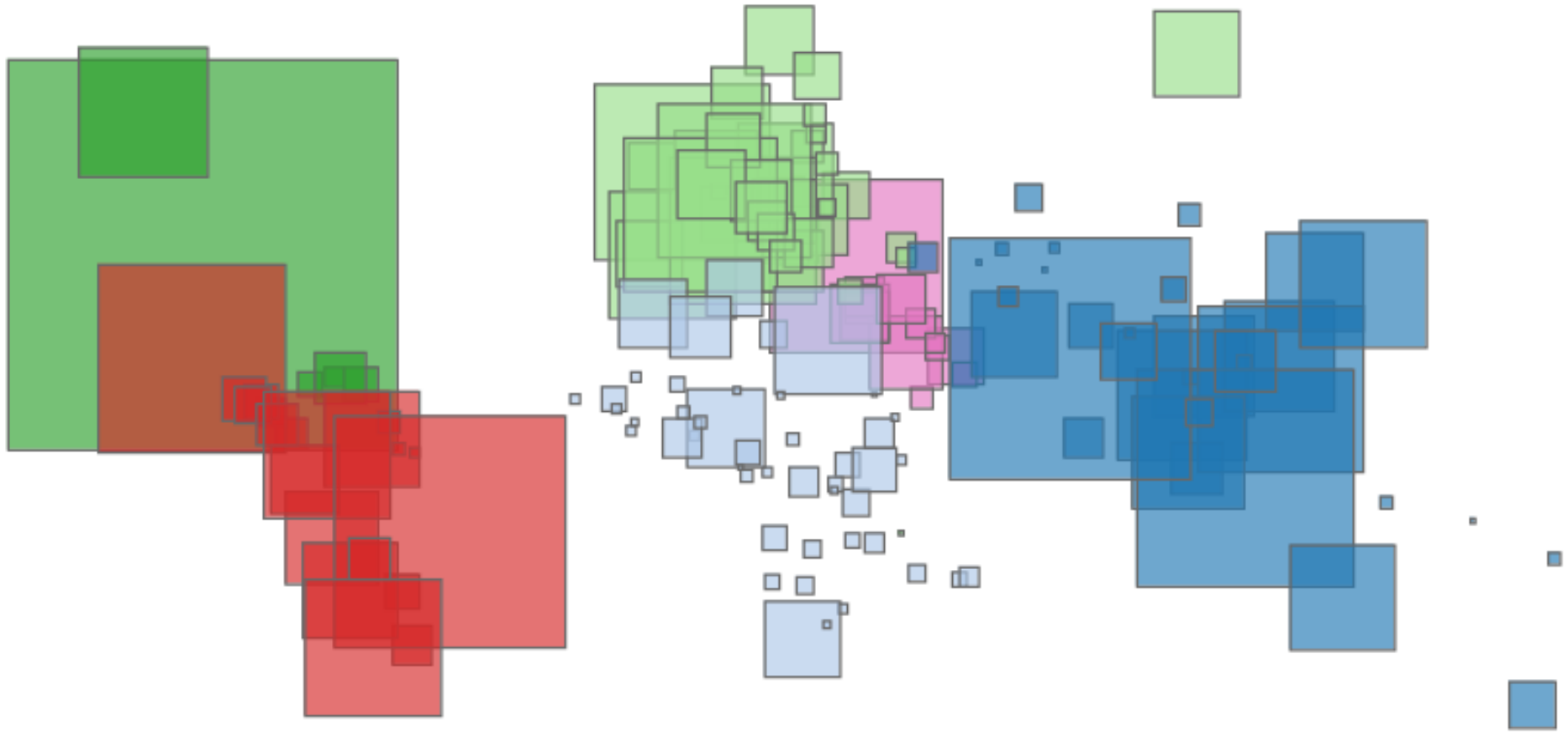
Size:

Facebook population; ratio

Dropline:

Identify point in ratio: Avg

Countries by Facebook Population

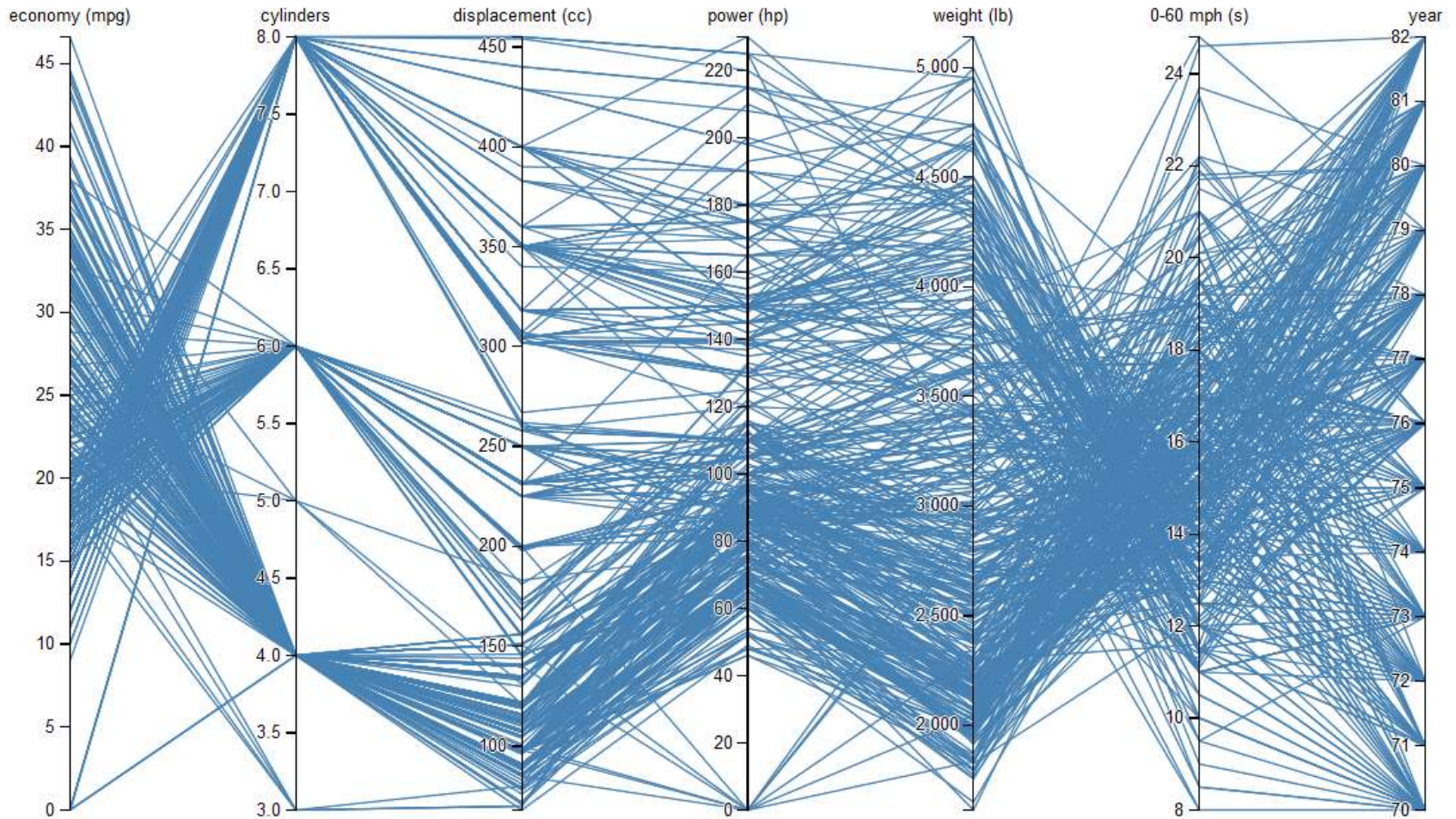


What If I Have More Dimensions?

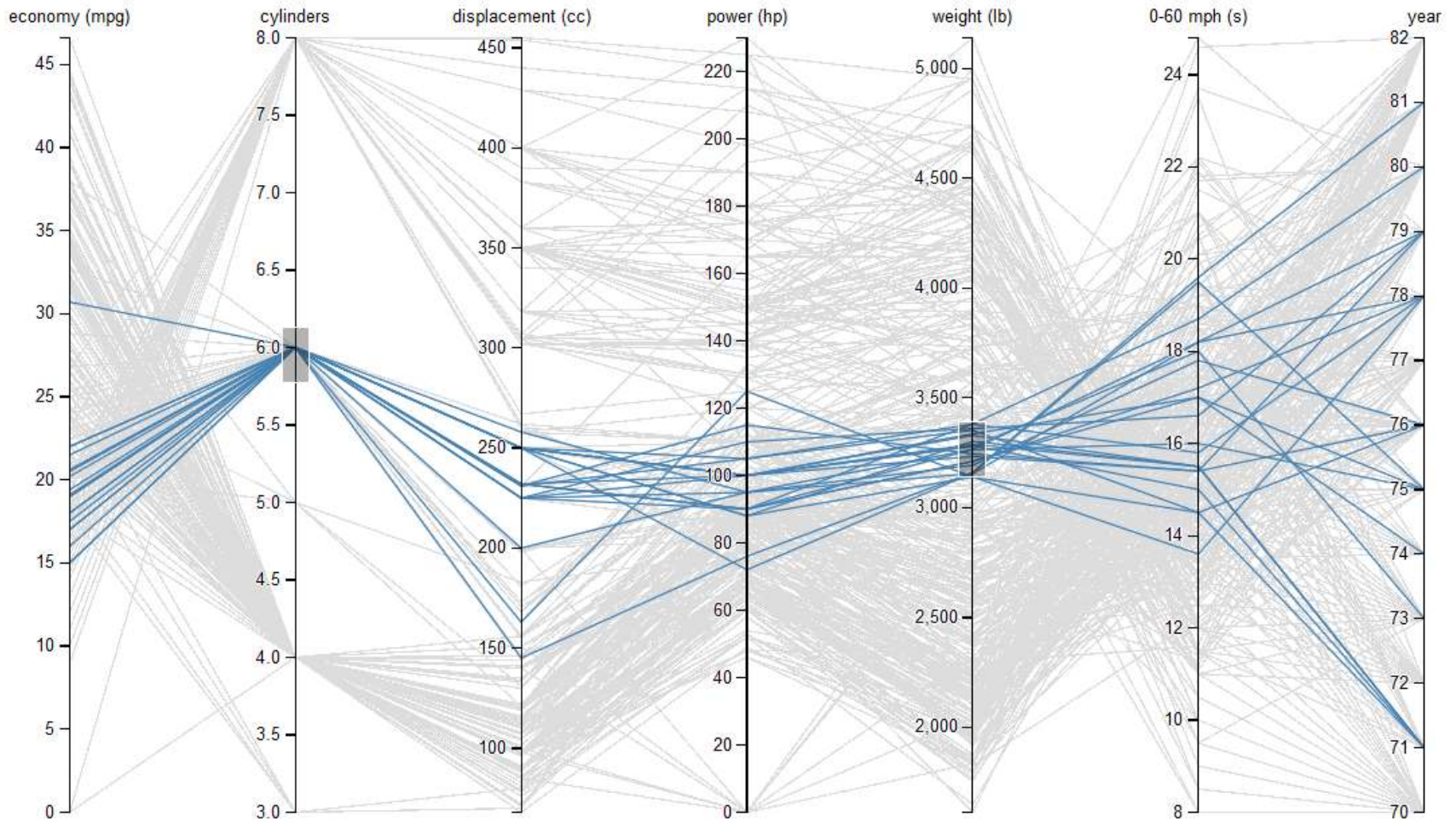


...or: Use Still Images

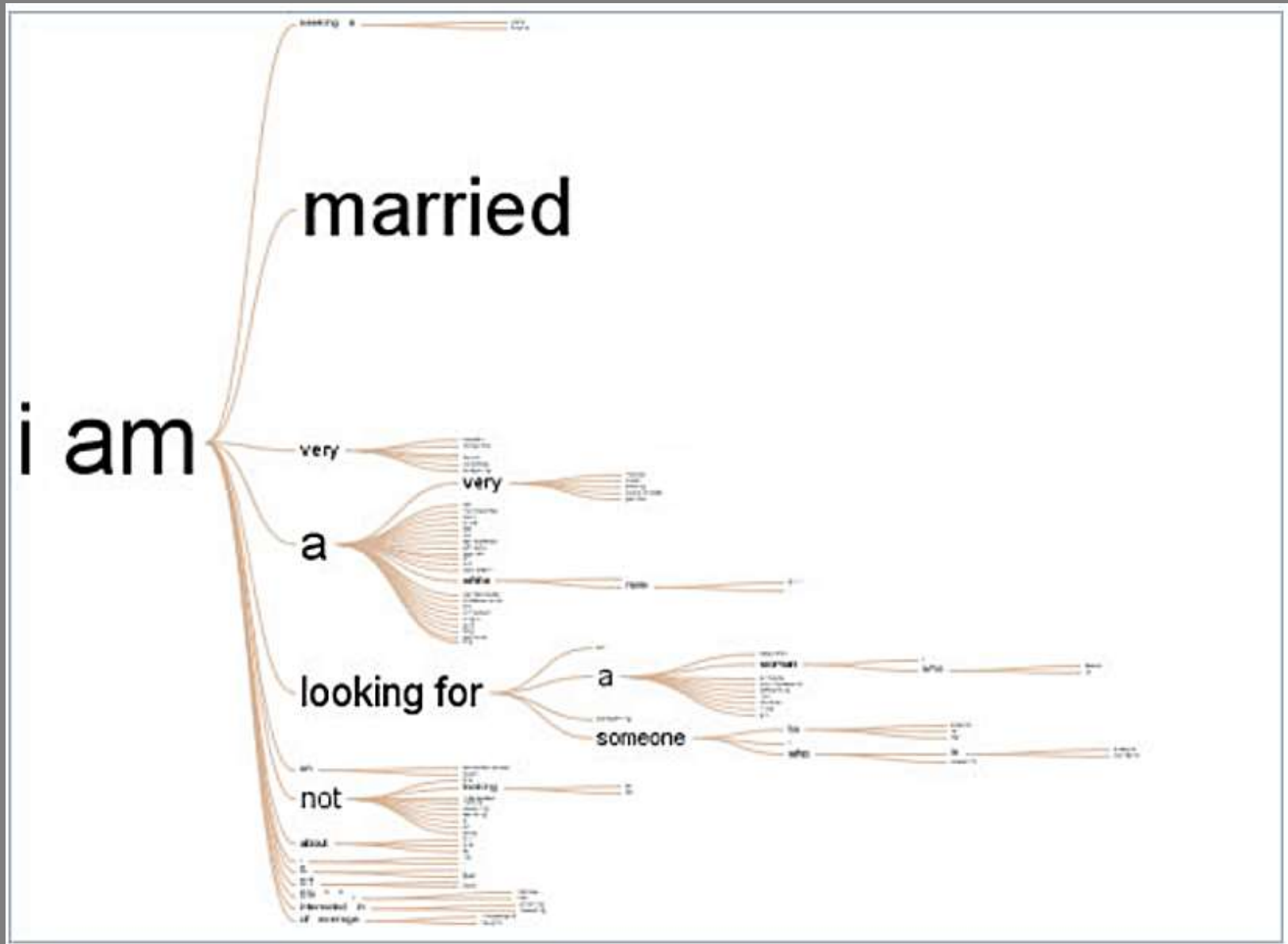
Parallel Coordinates



Parallel Coordinates



Visualizing Text





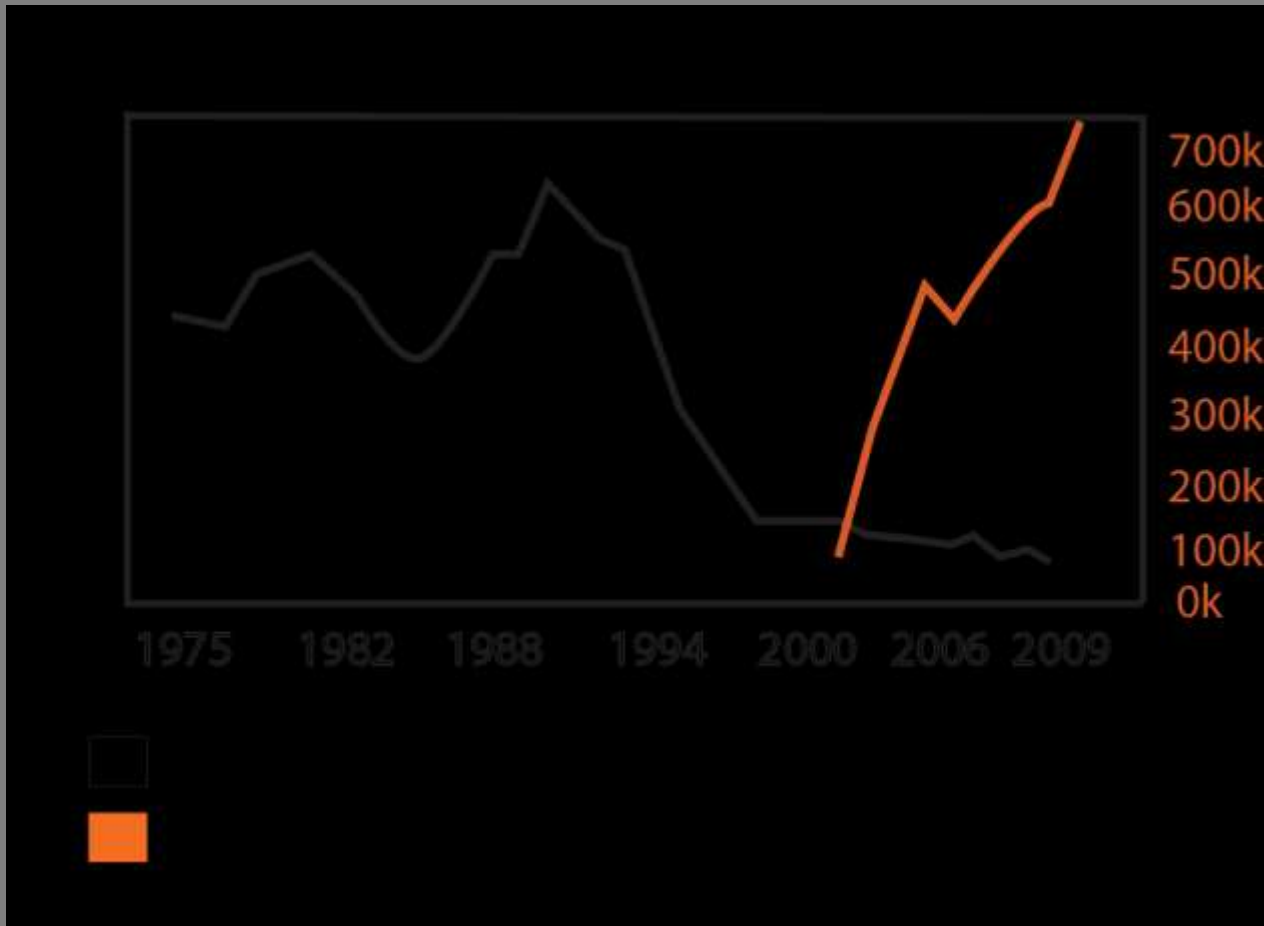
Roadmap

- Purpose of visualization
- Exercise
- History
- Scale types and their viz
- Multidimensional data
- Animation
- **What to avoid; deception**
- Exercise

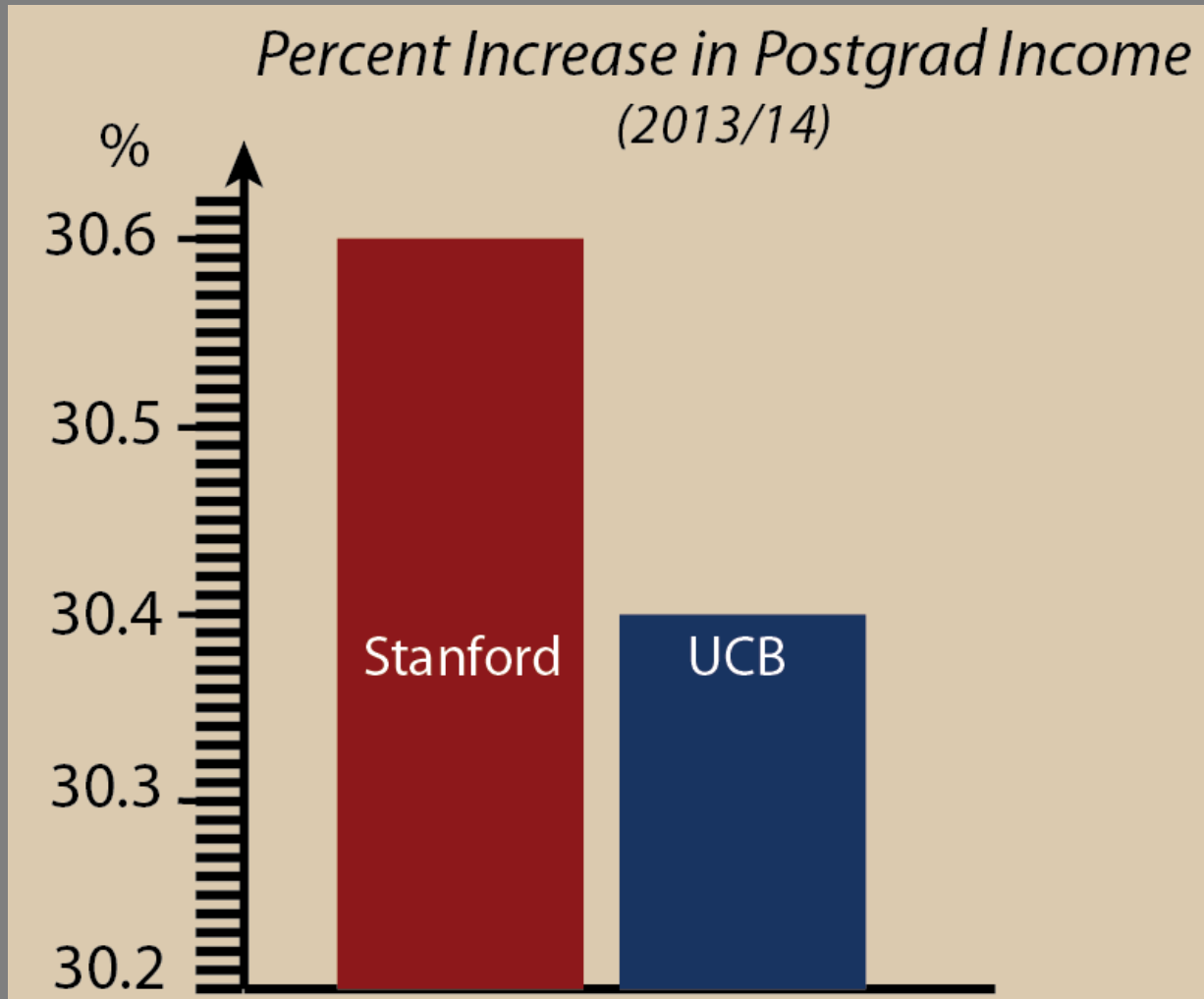
Stop-and-Frisk Prevents Murders



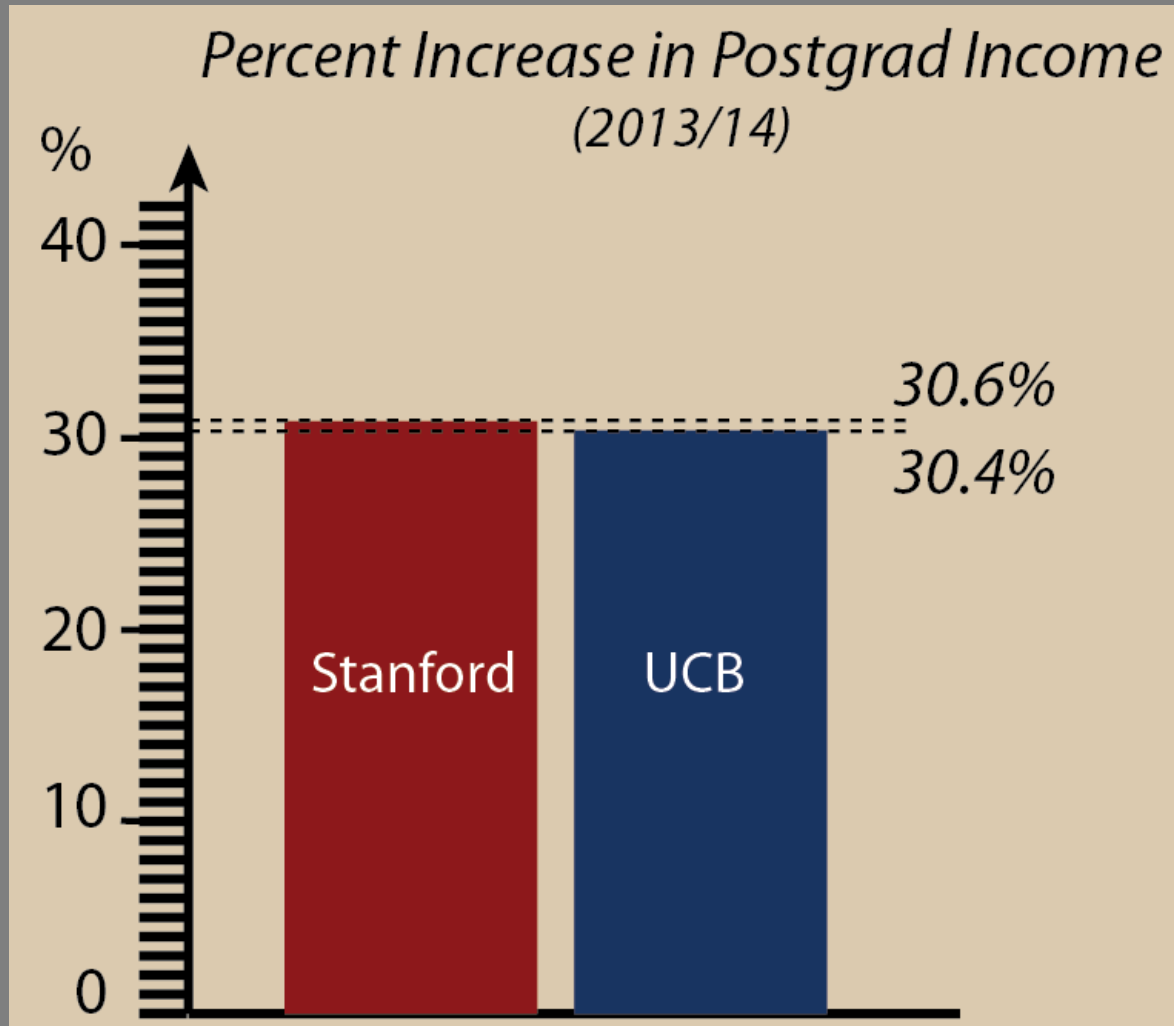
Implied Causation



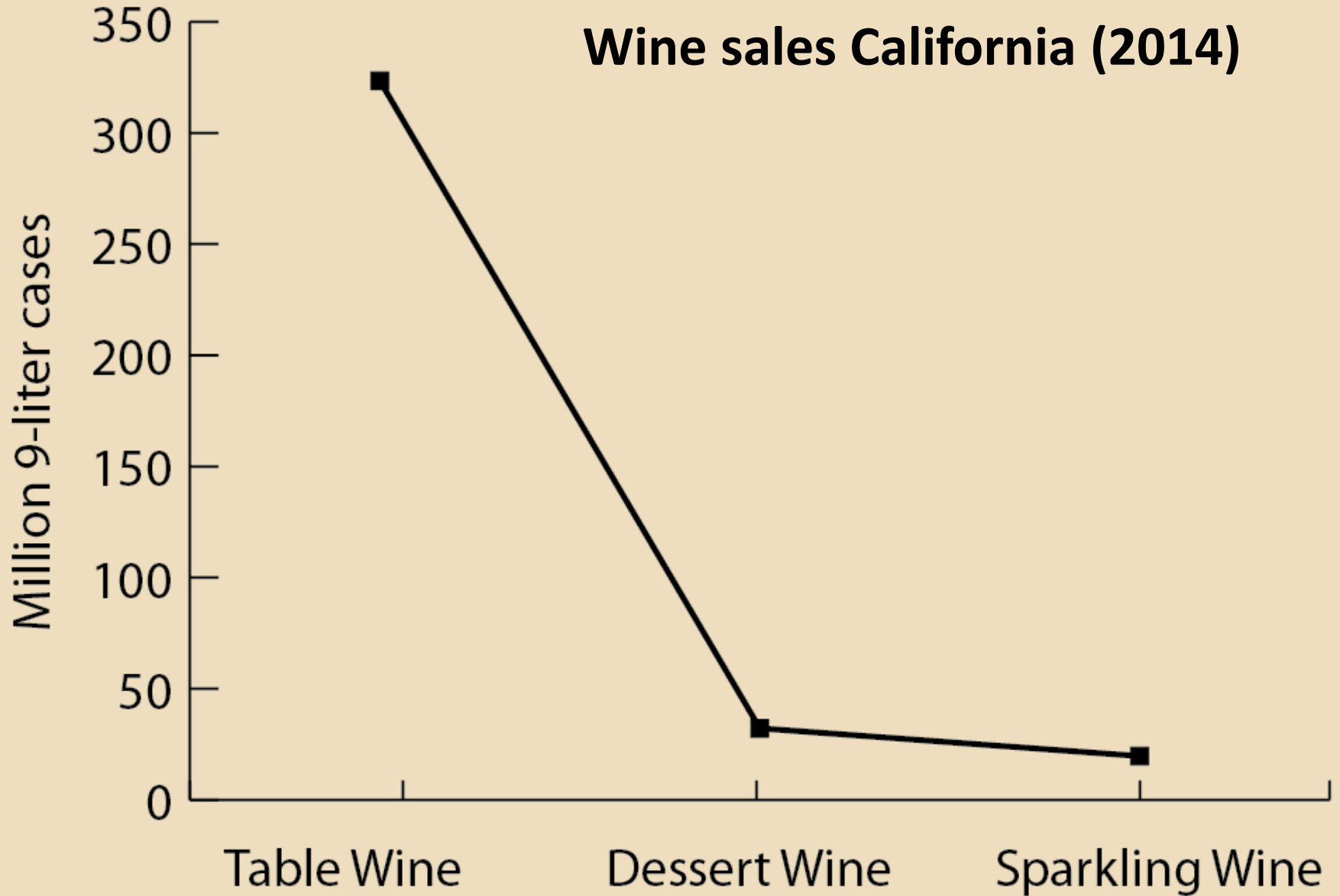
Stanford Grads Richer Quicker



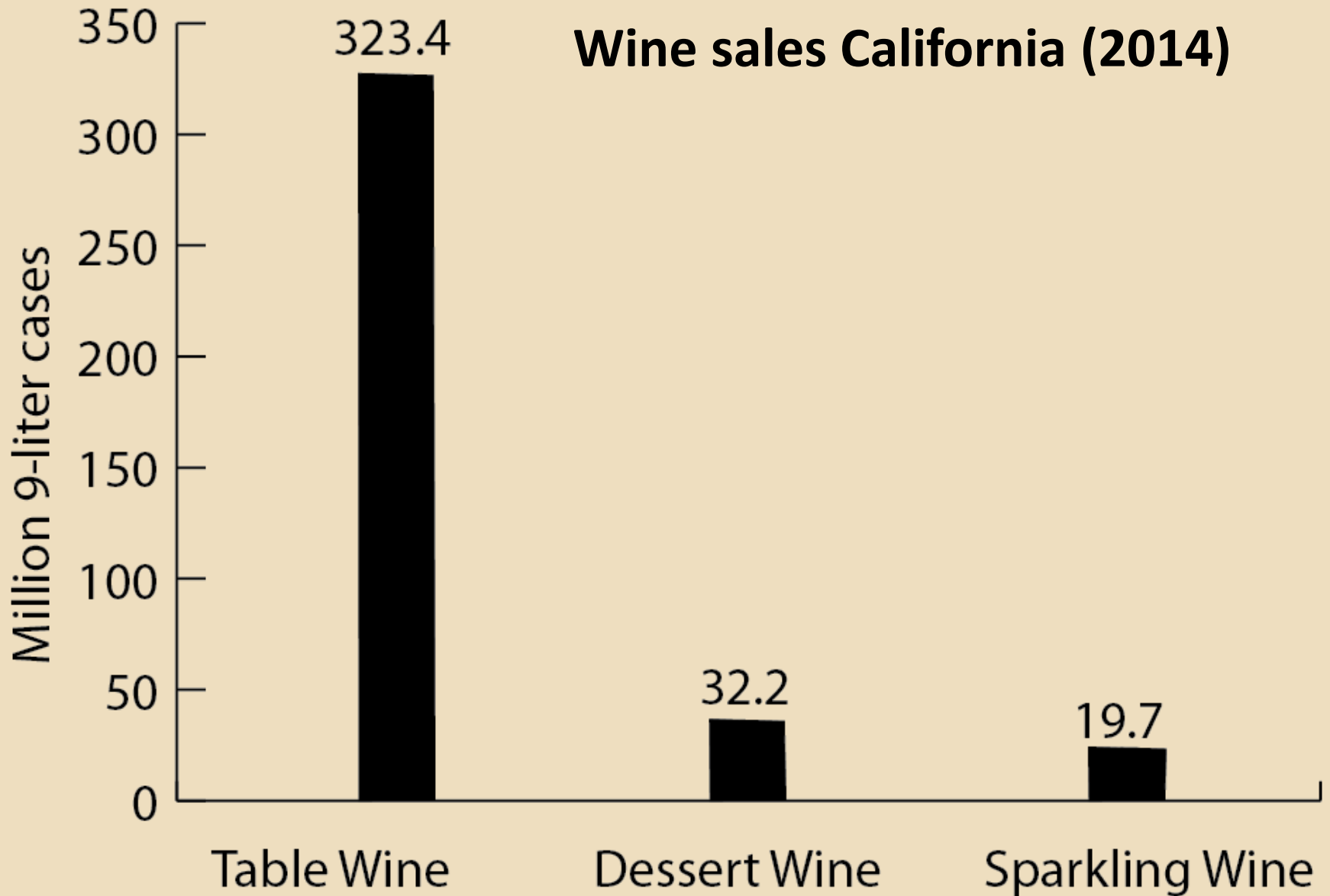
Partial Scale was Used



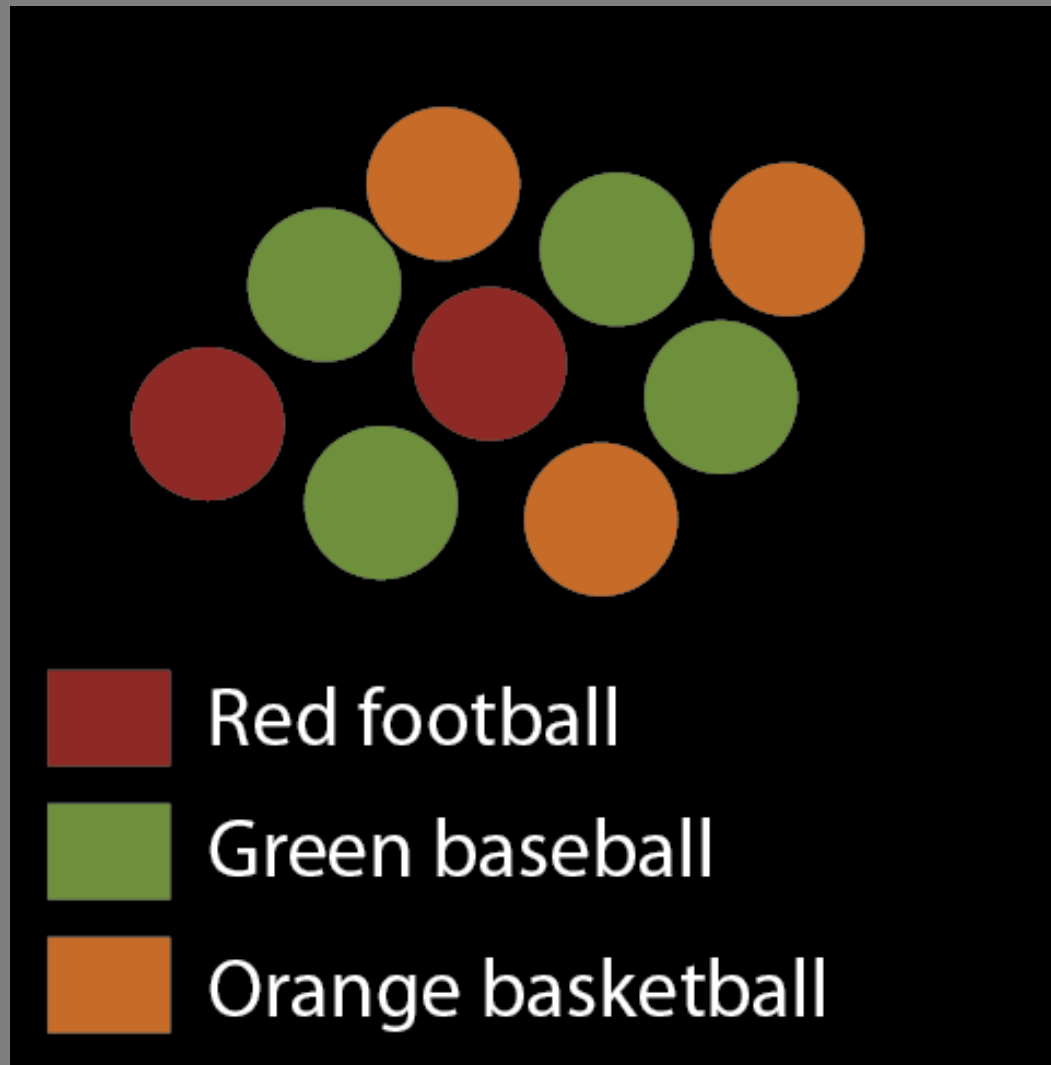
Wine sales California (2014)



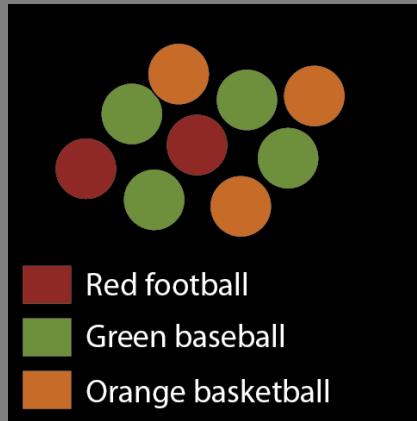
Wine sales California (2014)



Careful With Color



Careful With Color




Original



Test your charts at <http://bit.ly/1qSSJSu>

Exercise: Presidential Greatness/IQ



Click and
hope



Click...and
hope

Class page: Pick up CSV data

Tool: Past CSV data