SOMAIYA VIDYAVIHAR UNIVERSITY
K J Somaiya College of Engineering

Data Visualization
Design principles

Vaibhav P. Vasani

Assistant Professor

Department of Computer Engineering

K. J. Somaiya College of Engineering

Somaiya Vidyavihar University

vaibhav.vasani@gmail.com

Somaiya TRUST

Design principles

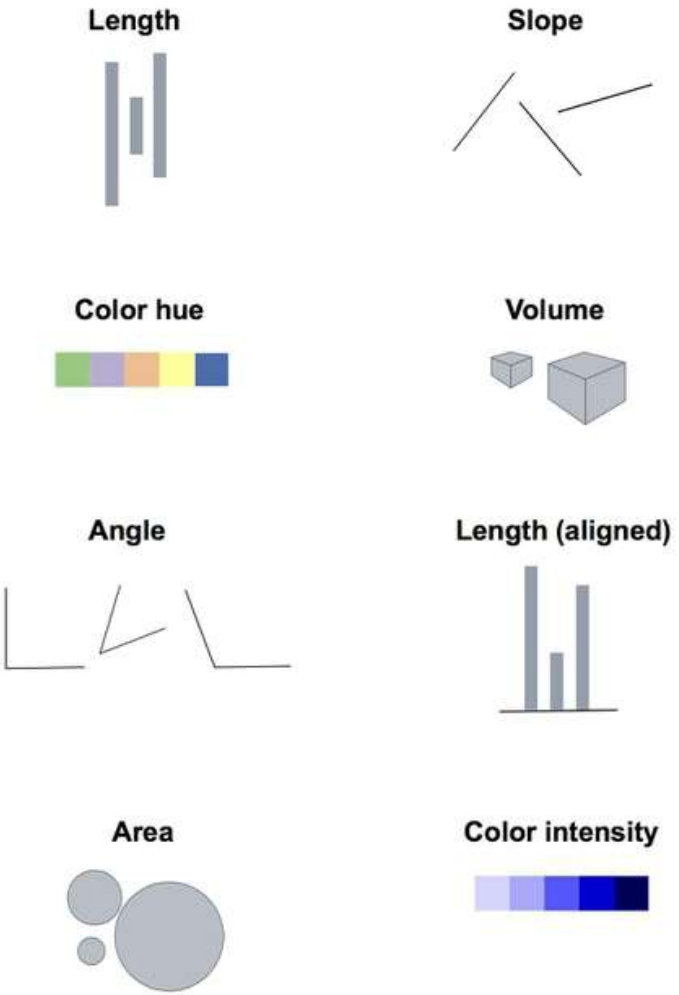
- Design for the human brain!

Data visualization: basic principles

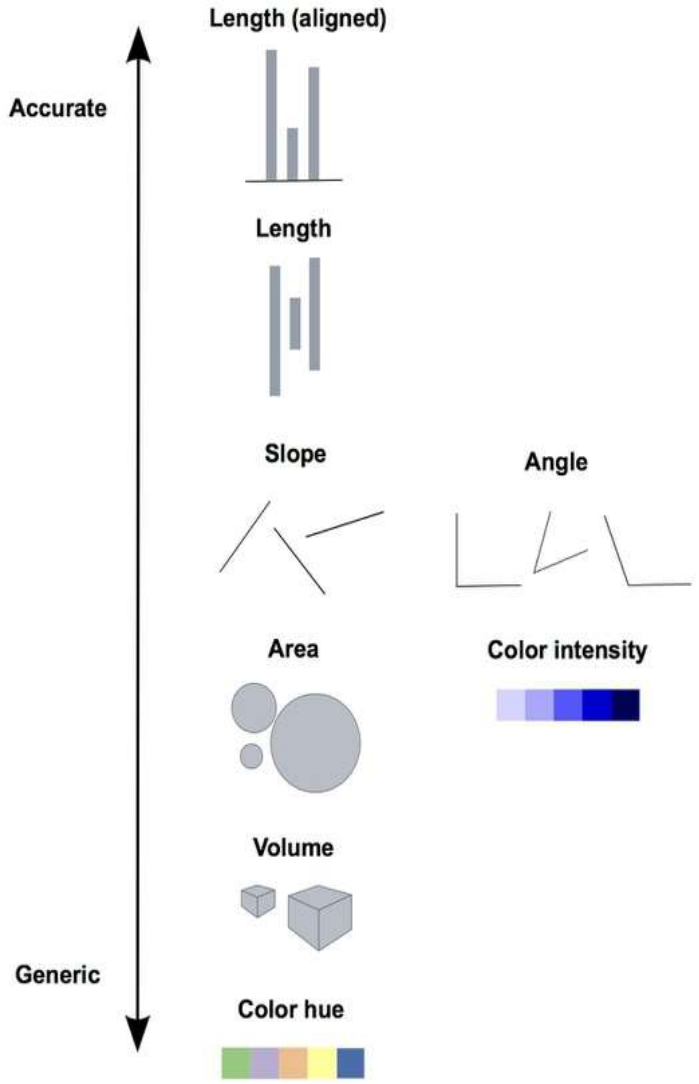
- It is a good way to communicate complex information, because we are highly visual animals, evolved to spot patterns and make visual comparisons.
- To visualize effectively, however, it helps to understand a little about how our brains process visual information.

Visualization: encoding data using visual cues

- Visualization encodes data using visual cues, or “mapping” data onto variation in size, shape or color, and so on.


Length
Slope
Color hue
Volume
Angle
Length (aligned)
Area
Color intensity

- All cues are not equal.
- In the mid-1980s, statisticians William Cleveland and Robert McGill ran some experiments with human volunteers, measuring how accurately they were able to perceive the quantitative information encoded by different cues.
- This perceptual hierarchy of visual cues is important.
- When making comparisons with continuous variables, aim to **use cues near the top** of the scale wherever possible.



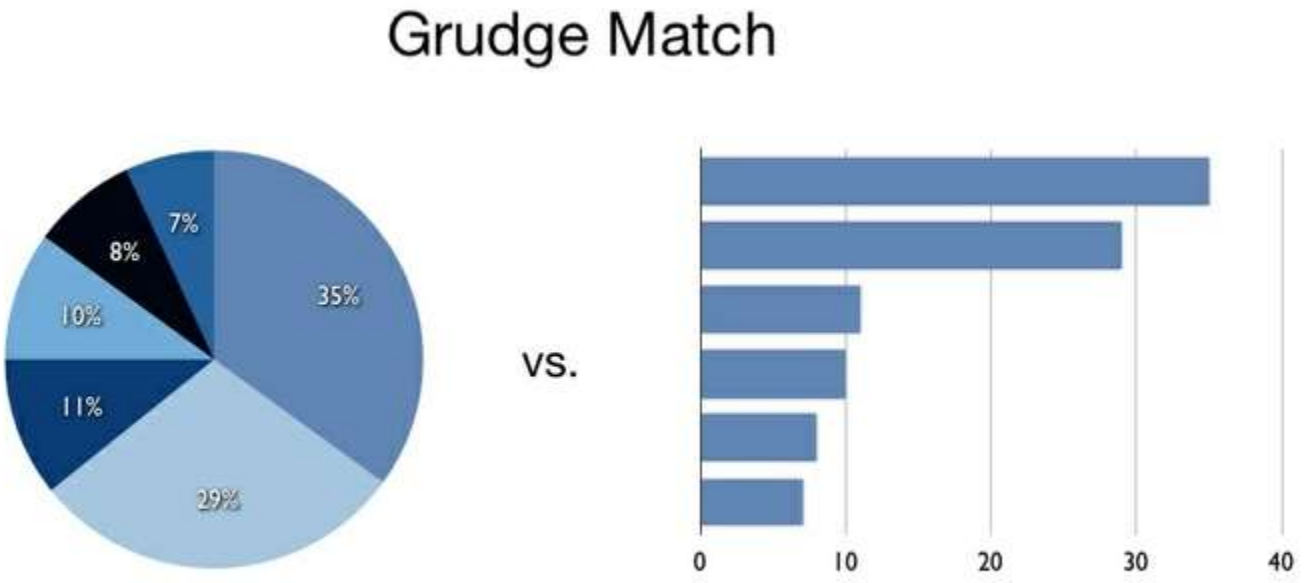
Read Yourself Slide

But this doesn't mean that everything becomes a bar chart

- Length on an aligned scale may be the best option to allow people to compare numbers accurately, but that doesn't mean the other possibilities are always to be avoided in visualization. Indeed, color hue is a good way of encoding categorical data.
- The human brain is particularly good at recognizing patterns and differences. This means that variations in color, shape and orientation, while poor for accurately encoding the precise value of continuous variables, can be good choices for representing categorical data.
- You can also combine different visual cues in the same graphic to encode different variables.
- But always think about the main messages you are trying to impart, and where you can use visual cues near the top of the visual hierarchy to communicate that message most effectively.
- To witness this perceptual hierarchy, look at the following visual encodings of the same simple dataset. In which of the three charts is it easiest to compare the numerical values that are encoded?

- Which of these two charts is easiest to read ?:

  Left or Right ----------→ →

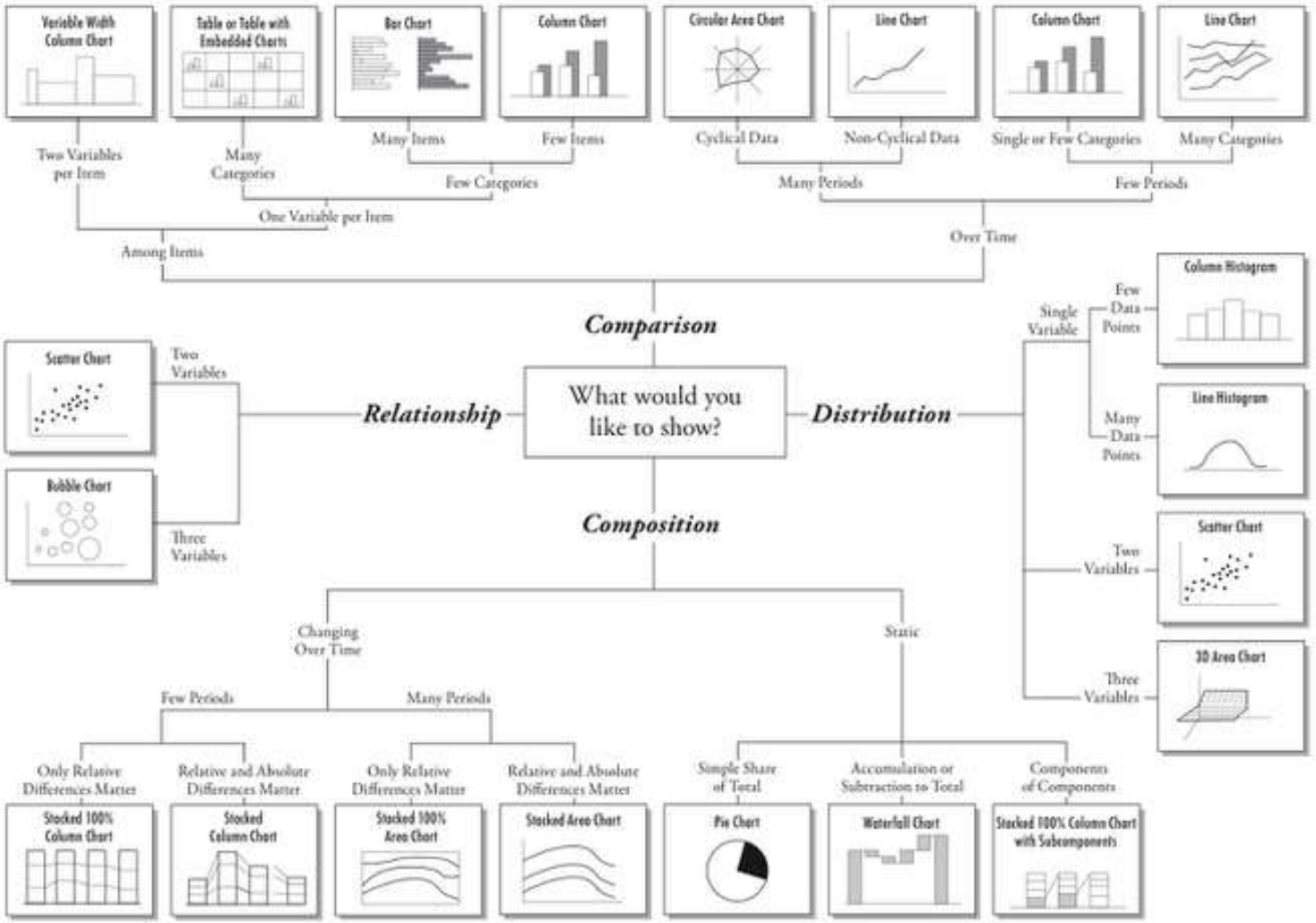


- Pie charts encode continuous variables primarily using the angles made in the center of the circle.
- It is certainly true that angles are harder read accurately than aligned bars.
- However, note that encoding data using the area of circles — which has become a "**fad**" in data visualization in recent years.

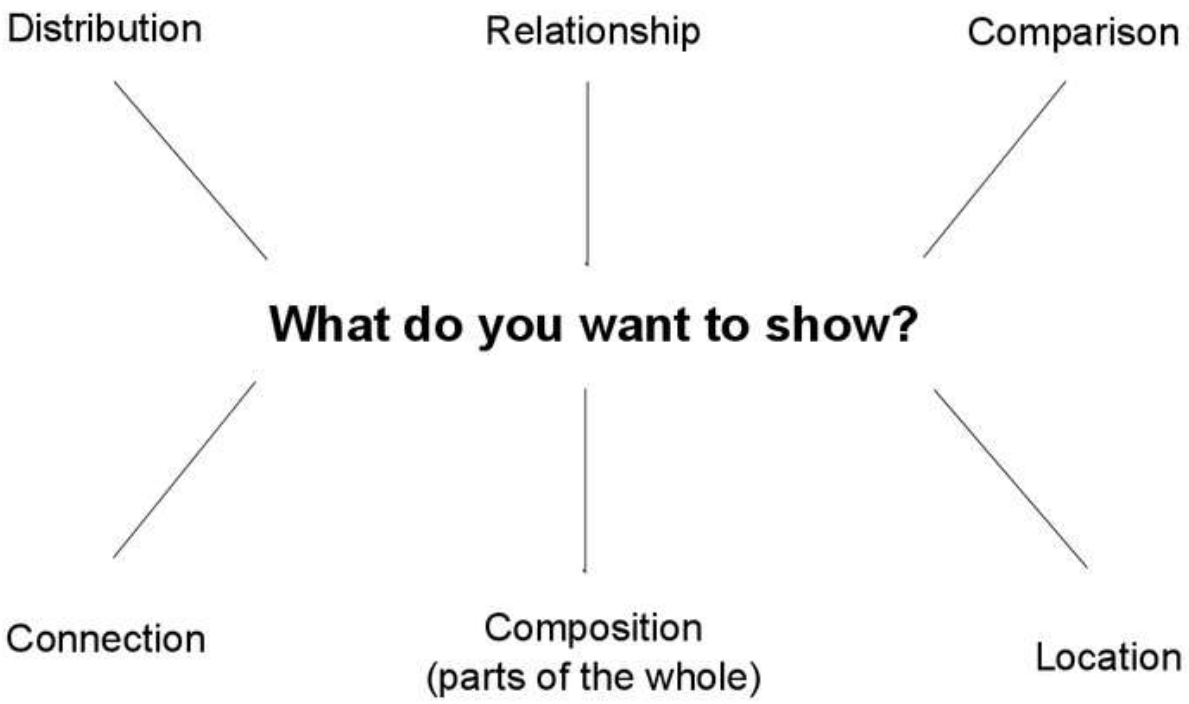
Which chart type should I use?

- Experiment with different charts, to see which works best to liberate the story in your data.
- Tableau Public — will suggest chart types for you to try.

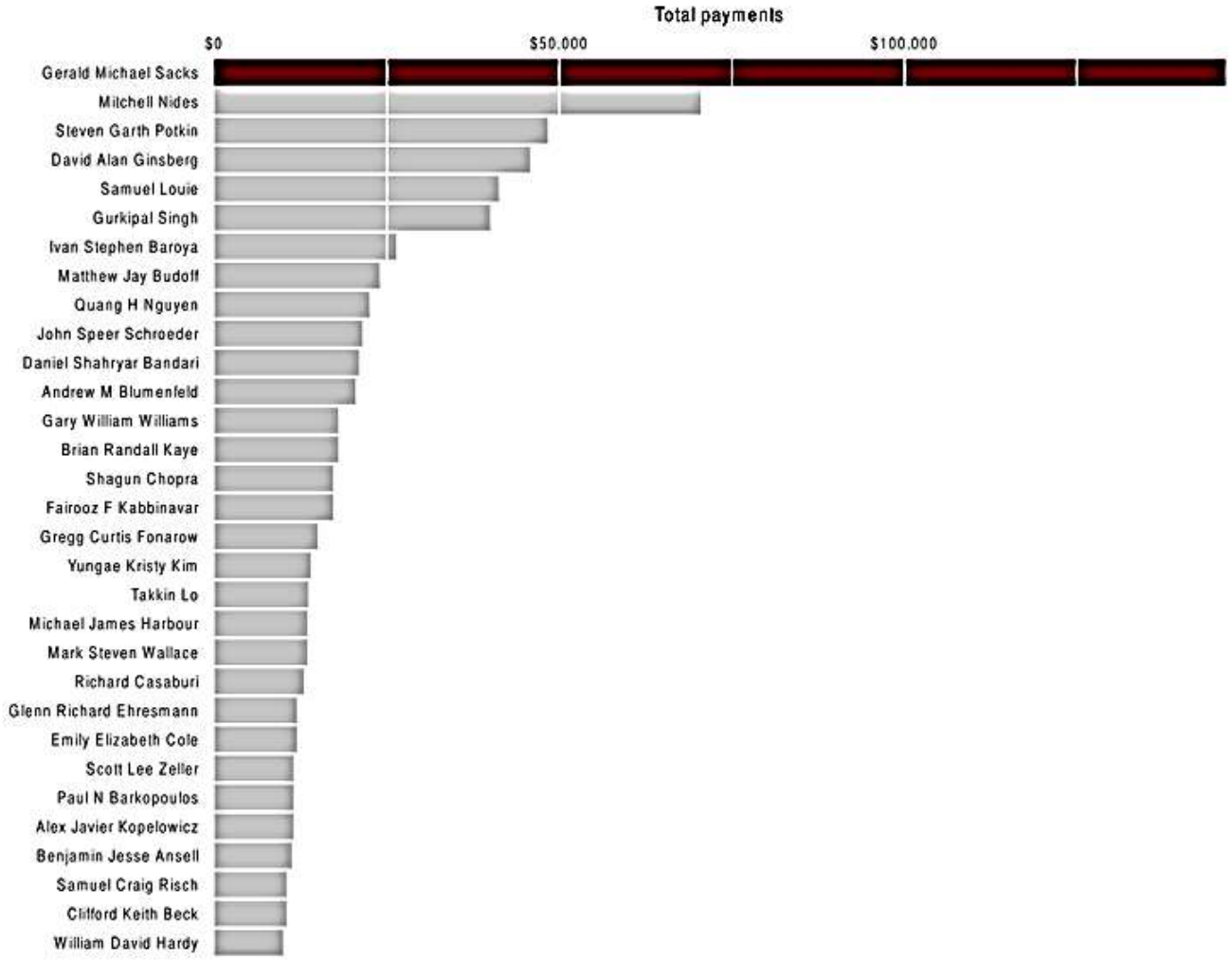
Chart Suggestions—A Thought-Starter



- **connection**, or visualizing how people, things, or organizations relate to one another; and **location**, which covers maps.

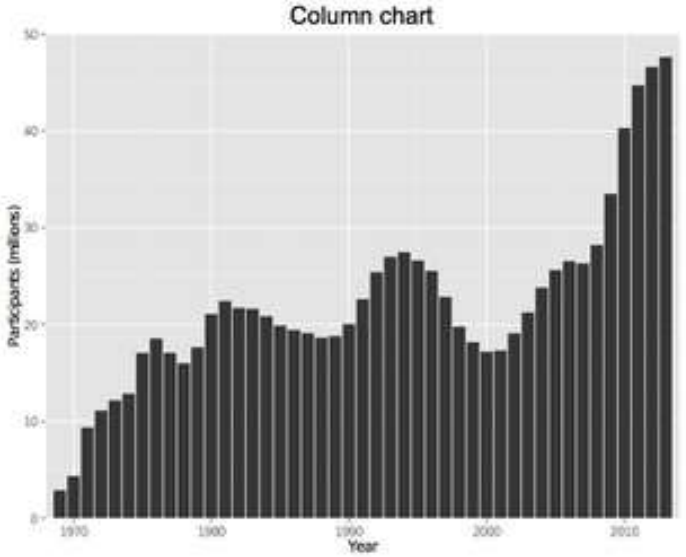


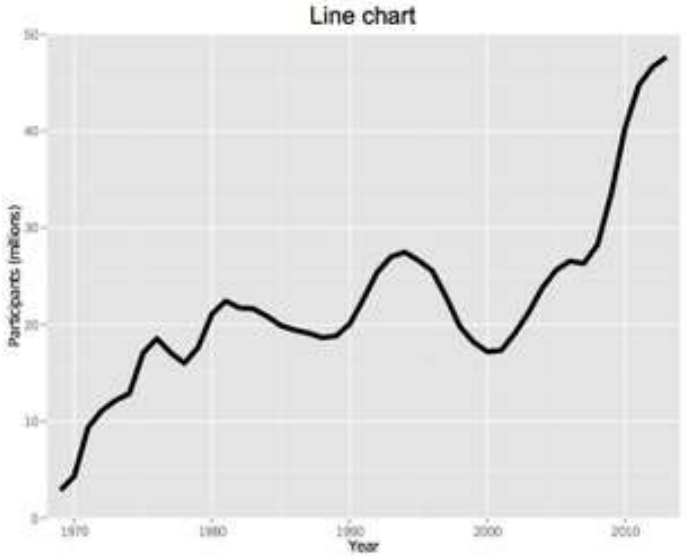
Simple comparisons: bars and columns

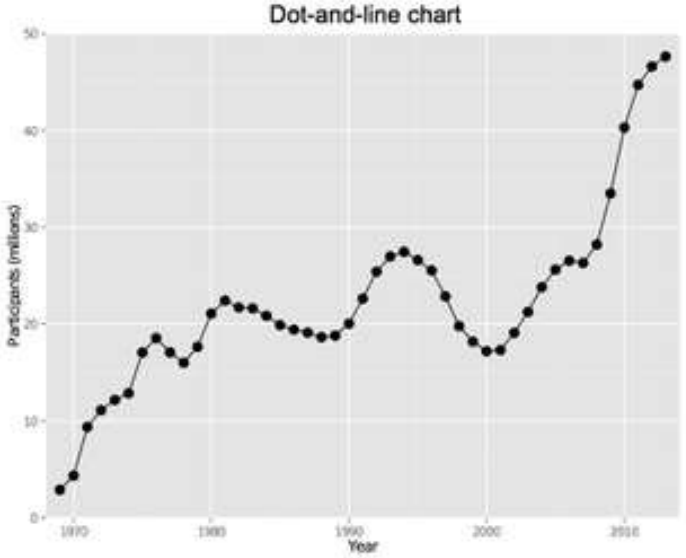


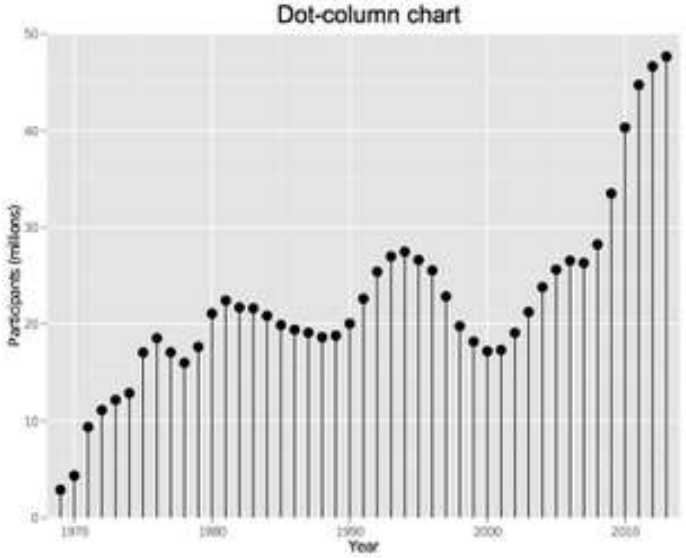
Comparisons: change over time

- Bar or column charts can also be used to illustrate change over time, but there are other possibilities, as shown in these charts showing participation in the federal government's food stamps nutritional assistance program, from 1969 to 2014.


Column chart
Participants (millions)
Year
0
10
20
30
40
50
1970
1980
1990
2000
2010

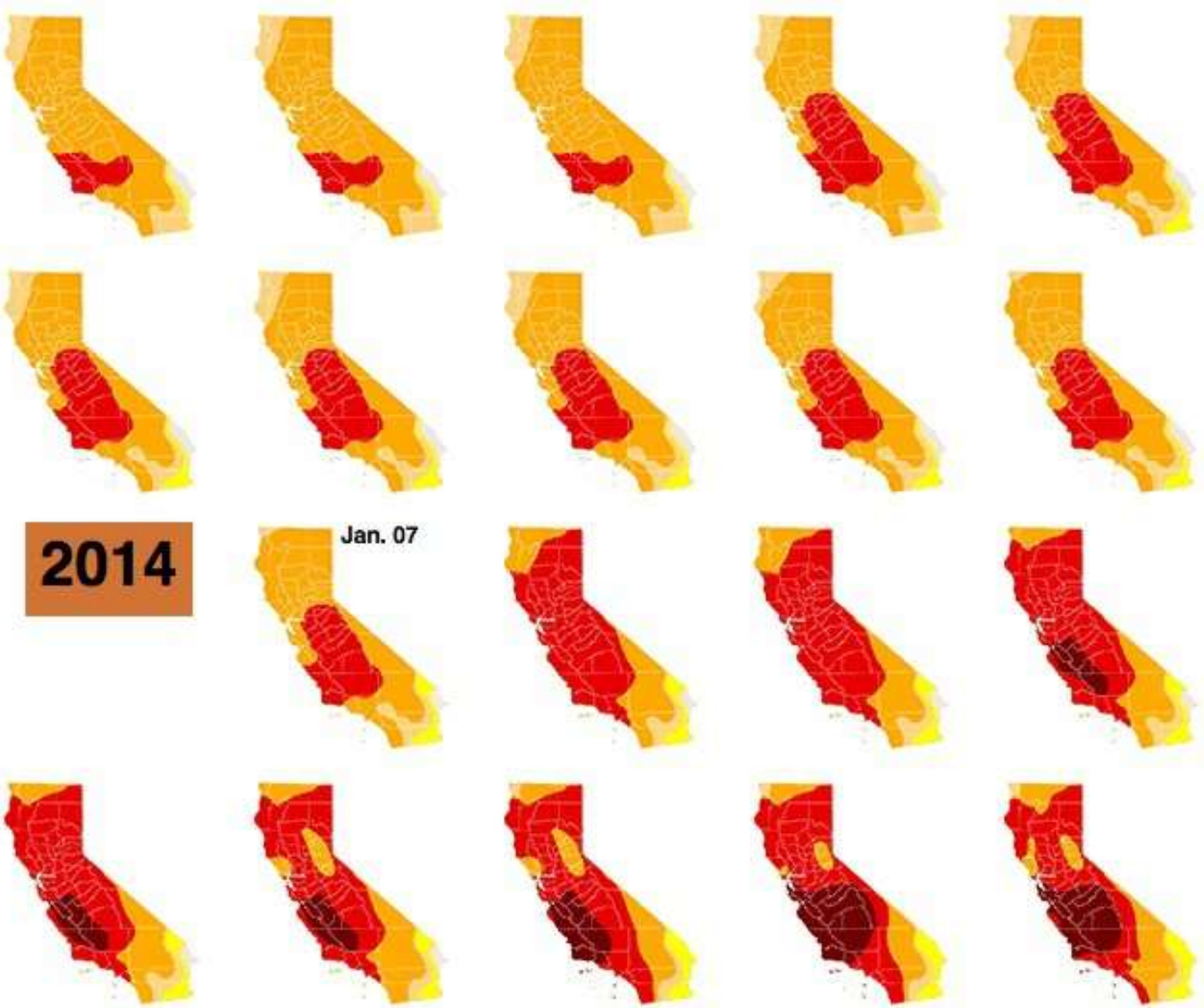

Line chart
Participants (millions)
Year
0
10
20
30
40
50
1970
1980
1990
2000
2010


Dot-and-line chart
Participants (millions)
Year
0
10
20
30
40
50
1970
1980
1990
2000
2010

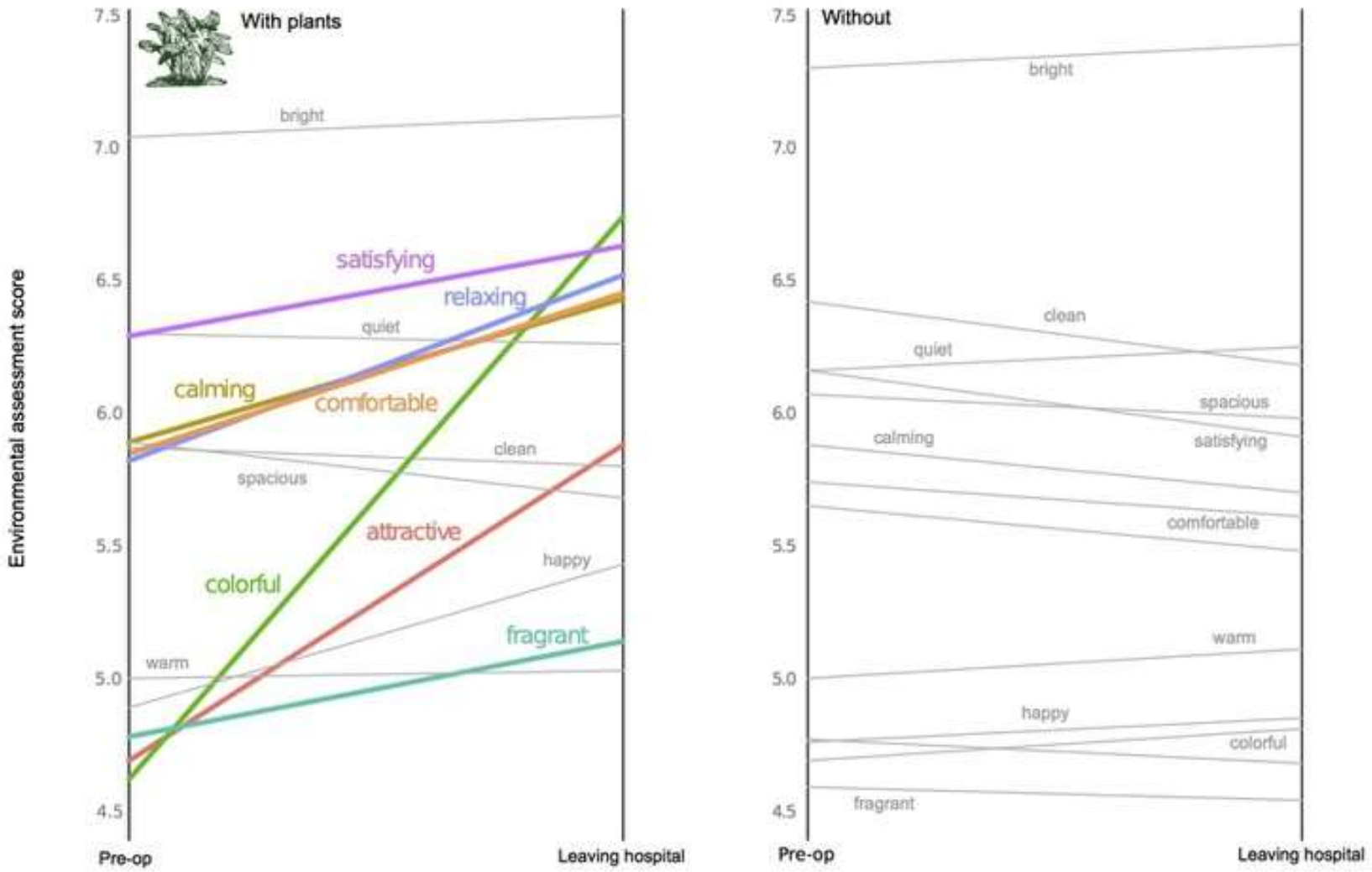

Dot-column chart
Participants (millions)
Year
0
10
20
30
40
50
1970
1980
1990
2000
2010

Multiple comparisons, including over time

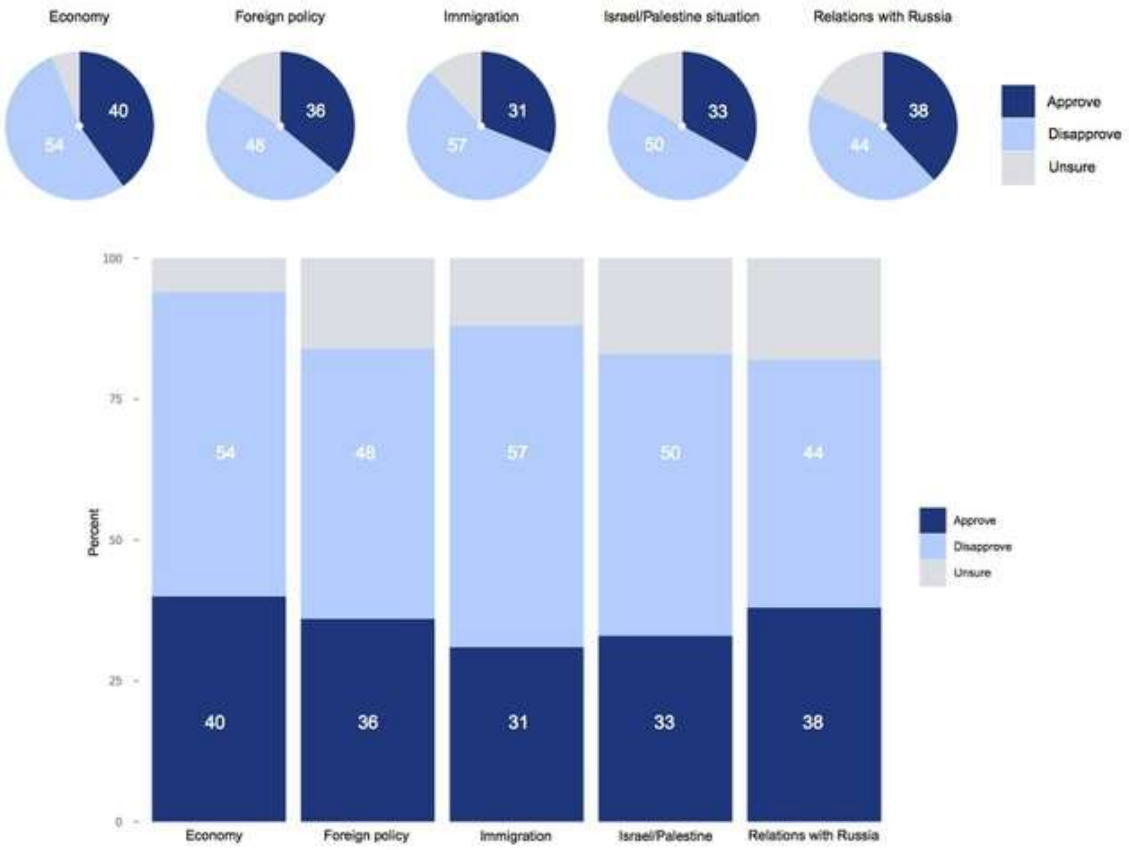
- When comparing very many items, or how one item has changed over time, “small multiples” provide another approach.
- Very successfully in recent years by several news organizations.
- Here is a small section from a larger graphic showing the severity of drought in California in late 2013 and early 2014



(Source: *Los Angeles Times*)

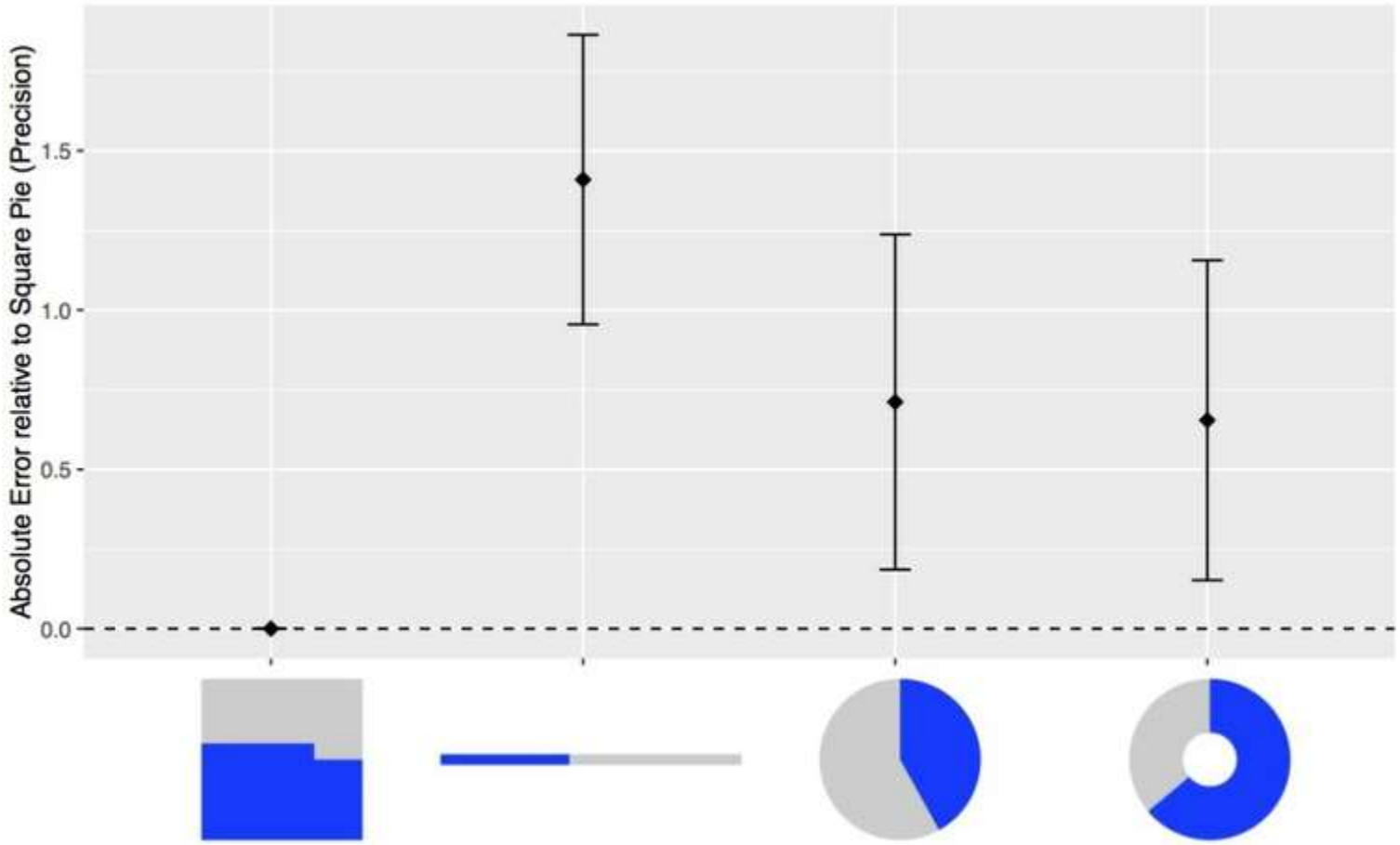

With plants
Without
Environmental assessment score
7.5
7.0
6.5
6.0
5.5
5.0
4.5
bright
satisfying
relaxing
quiet
calming
comfortable
clean
spacious
attractive
happy
colorful
fragrant
warm
Pre-op
Leaving hospital

Composition: parts of the whole

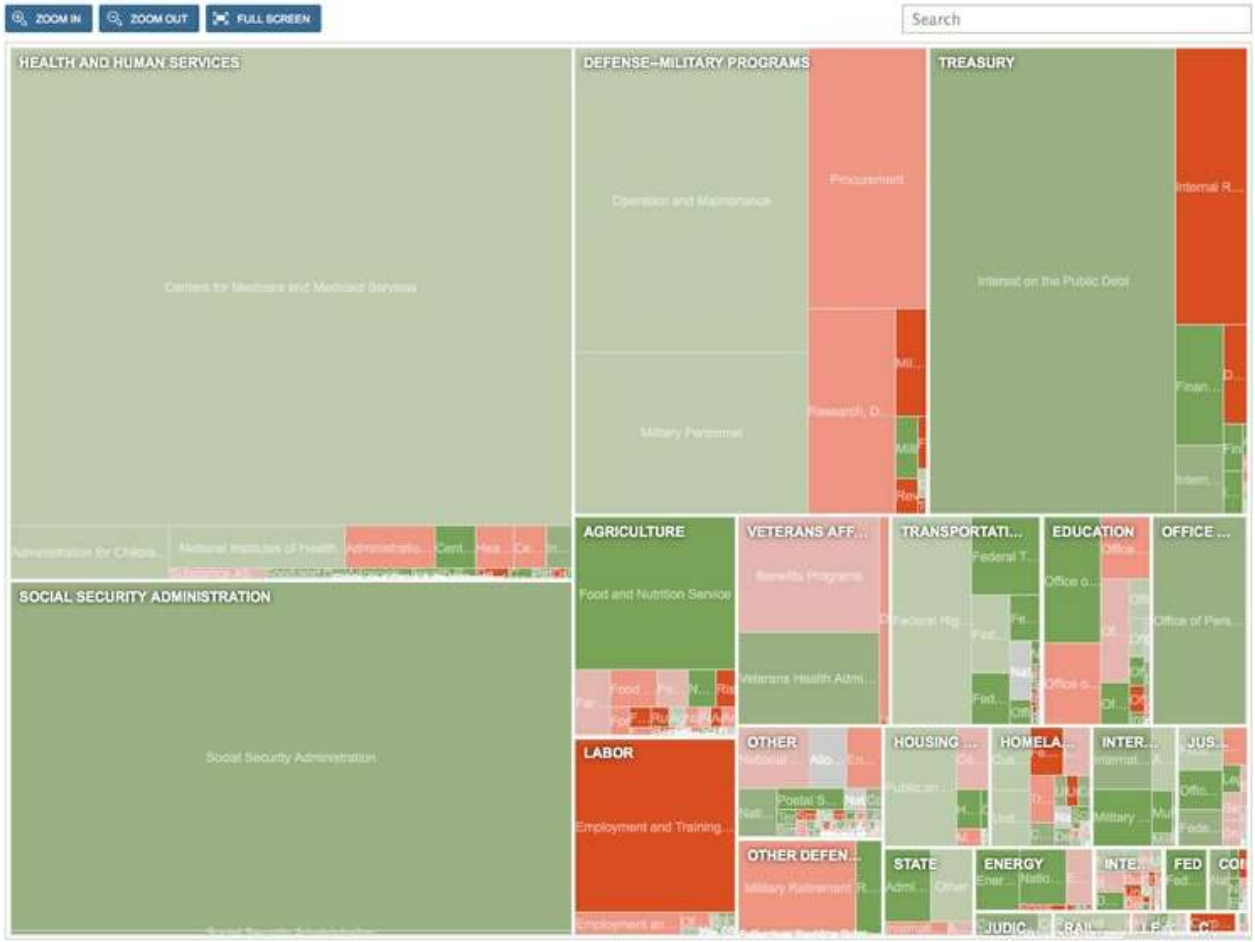


(Source: Peter Aldhous, from CBS poll data, via PollingReport.com)

- Abandon pie charts if there are any more than three parts to the whole, as they become very hard to read when there are many segments.
- ProPublica's graphics style guide goes further, allowing pie charts with two segments only.
- Recent research into how people perceive composition visualizations with just categories suggests that the best approach may actually be a square chart. (encoding of area seems to beat length for accuracy):



(Source: Eagereyes)

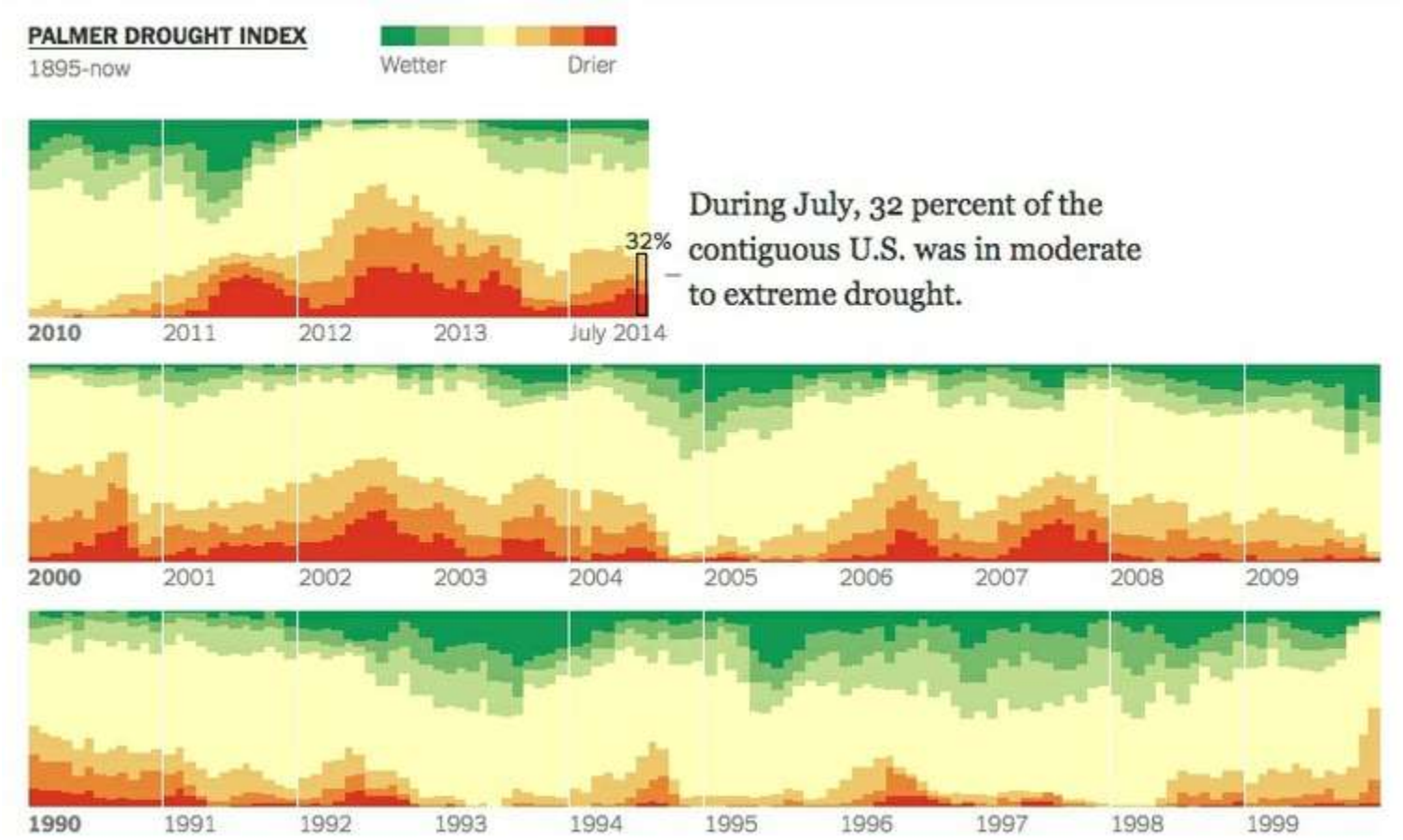


(Source: *The New York Times*)

Composition: change over time

- Data journalists frequently need to show how parts of the whole vary over time.

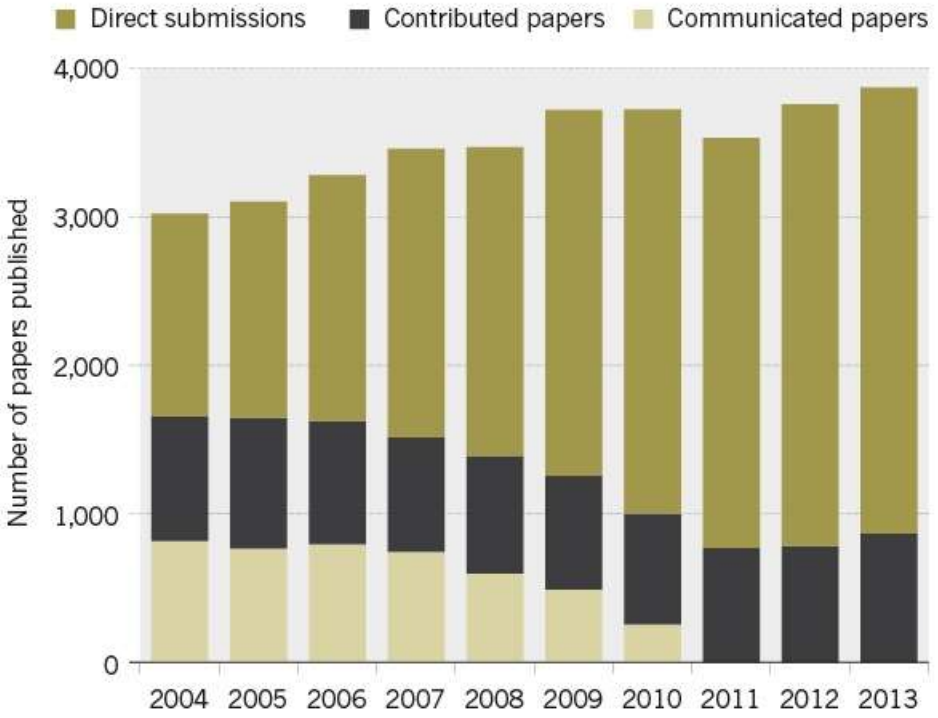
Illustrating the development of drought across the United States, which uses a stacked columns format, in this case with no space between the columns.



(Source: The Upshot, *The New York Times*)

A changing journal

The number of direct submissions to *Proceedings of the National Academy of Sciences* has been increasing steadily over the past decade. Communicated papers were phased out in 2010, but the contributed track has remained constant.



(Source: *Nature*)

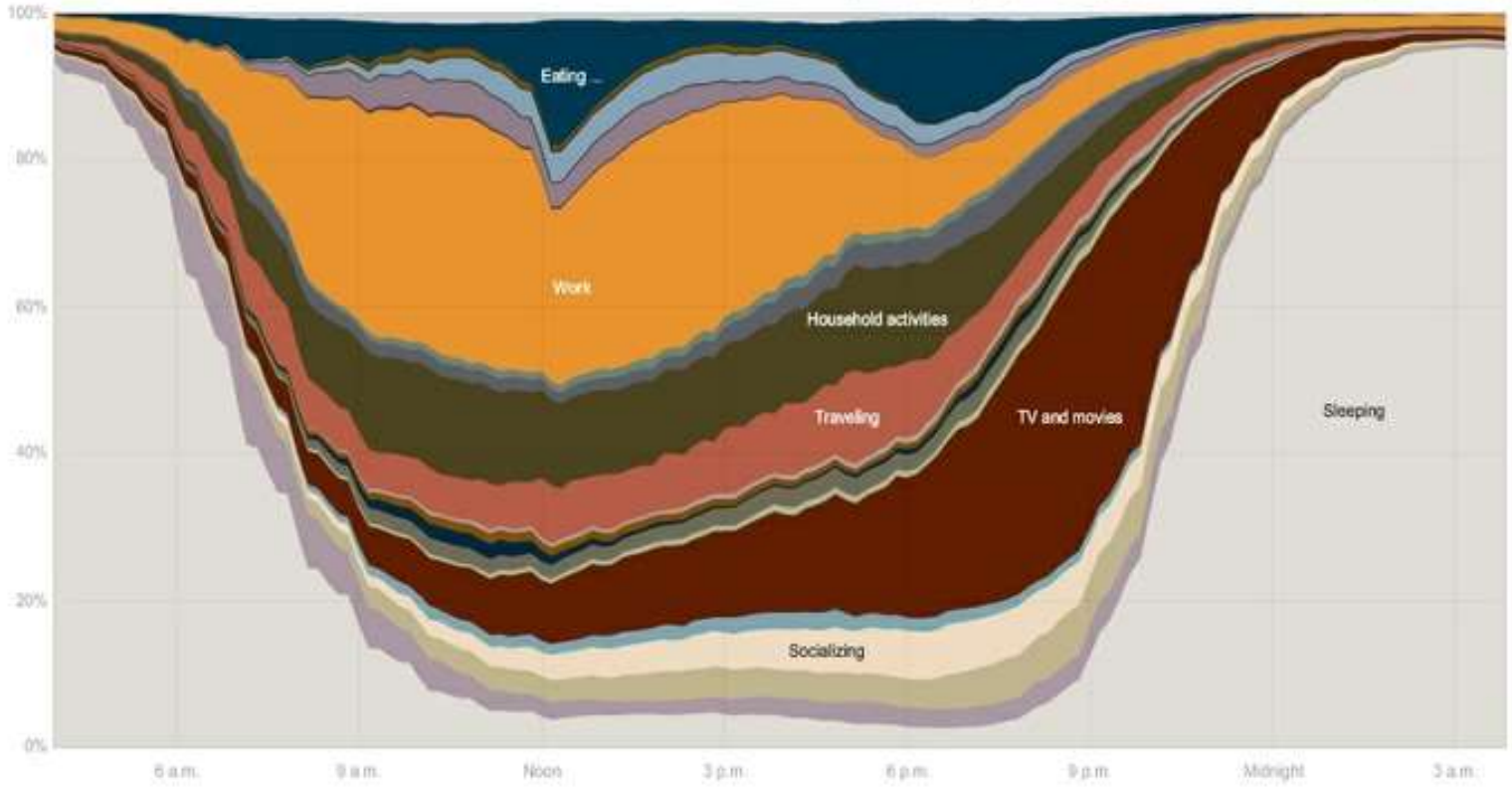
How Different Groups Spend Their Day

The American Time Use Survey asks thousands of American residents to recall every minute of a day. Here is how people over age 15 spent their time in 2008. Related article

Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



(Source: *The New York Times*)

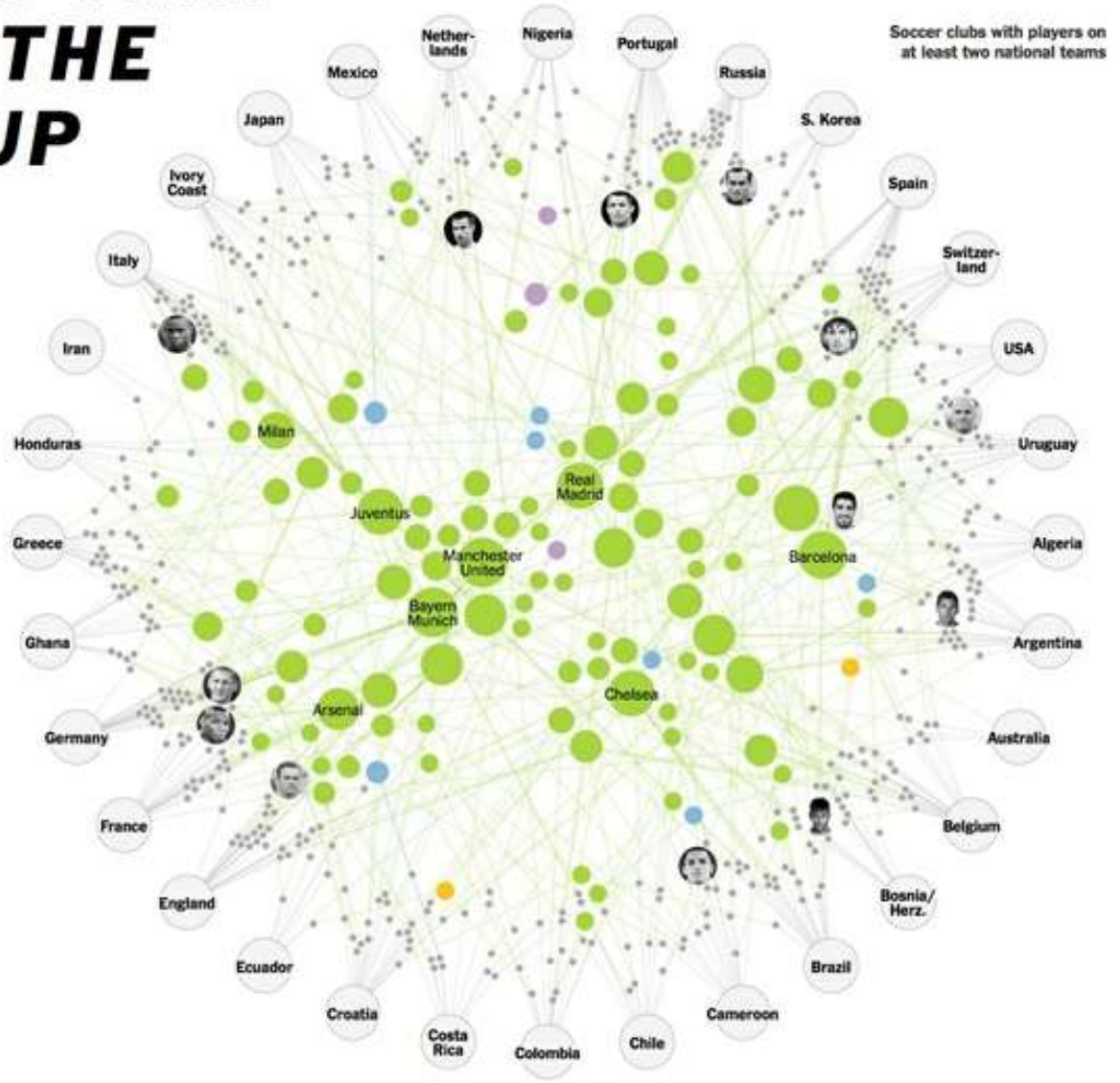
Making connections: network graphs

THE CLUBS THAT CONNECT THE WORLD CUP

By GREGOR AISCH JUNE 20, 2014

The best national teams come together every four years, but the global tournament is mostly a remix of the professional leagues that are in season most of the time. Three out of every four World Cup players play in Europe, and the top clubs like Barcelona, Bayern Munich and Manchester United have players from one end of the globe to the other.

Europe Africa Asia
South America North America



Brazil vs. Argentina

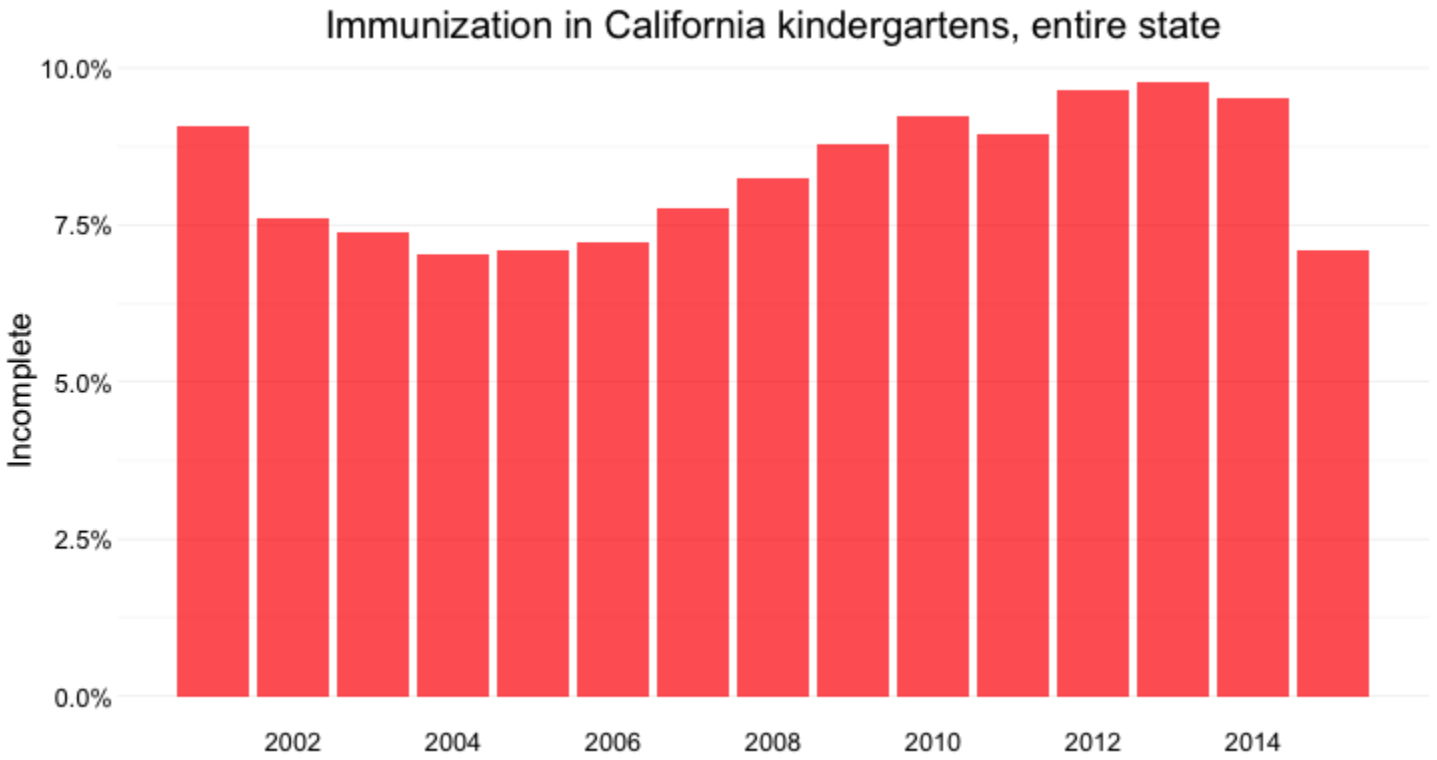
Even archrivals Brazil and Argentina overlap. Neymar, Brazil's star forward, plays alongside Lionel Messi, the Argentine captain, on powerhouse Barcelona. In all, eight Brazilians and 12 Argentines play together on European club teams.

(Source: *The New York Times*)

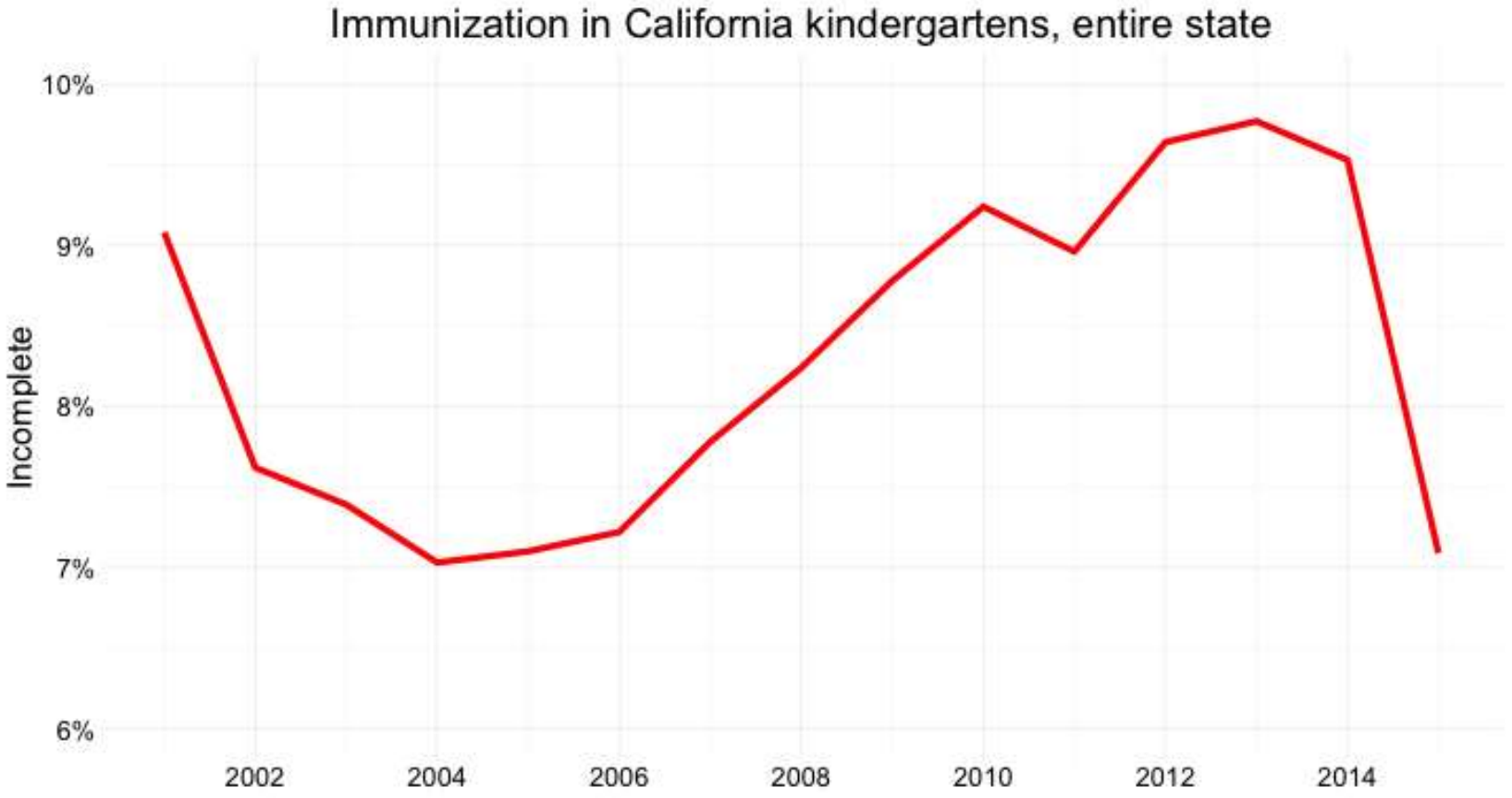
Case study: Immunization in California kindergartens

- A dataset at different levels of analysis, to show how different visual encodings may be needed for different visualization tasks with the same data.

- Data, from the California Department of Public Health, gives numbers on immunization and enrolment at kindergartens across the state.

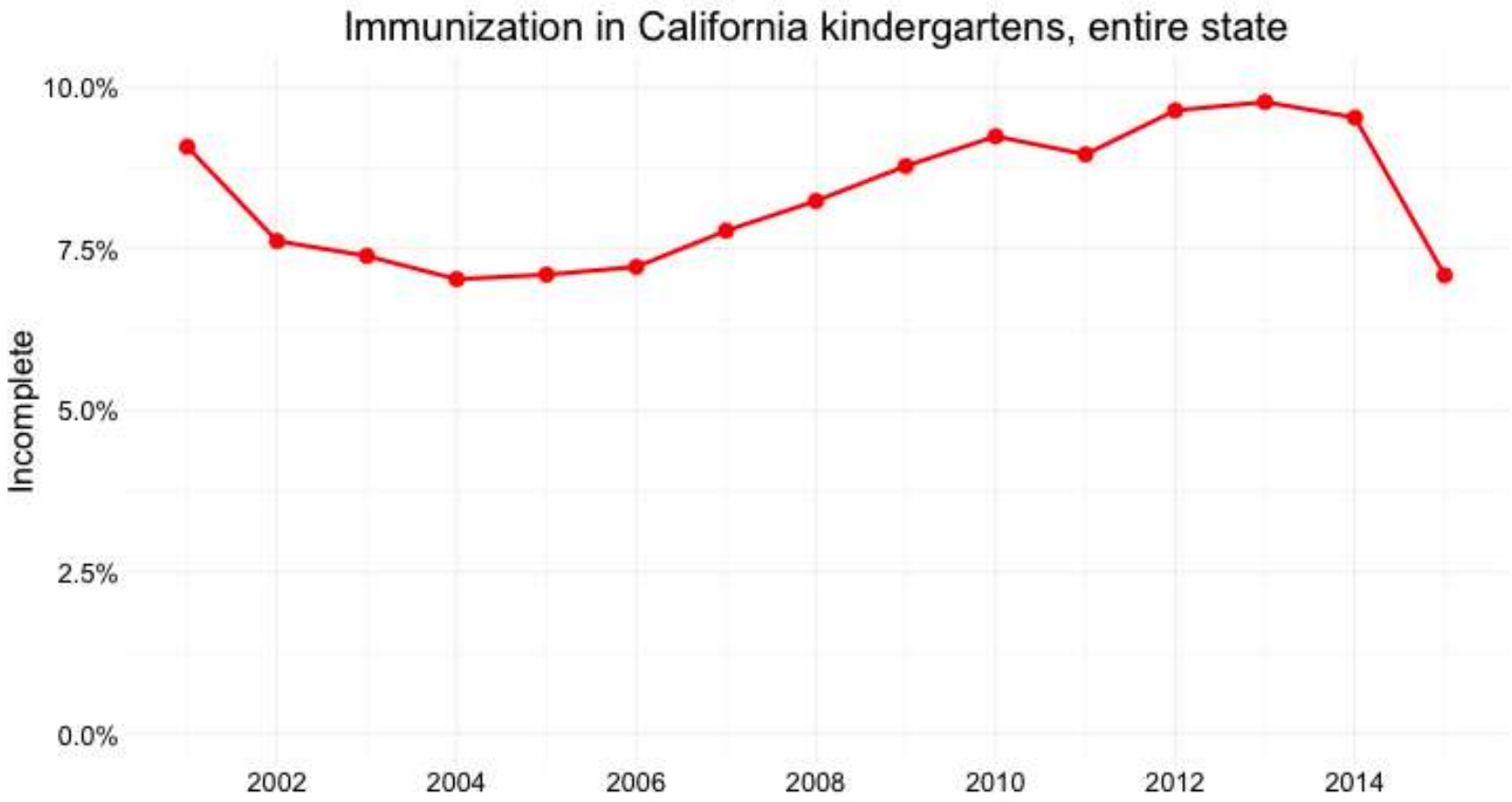


(Source: Peter Aldhous, from California Department of Public Health data)

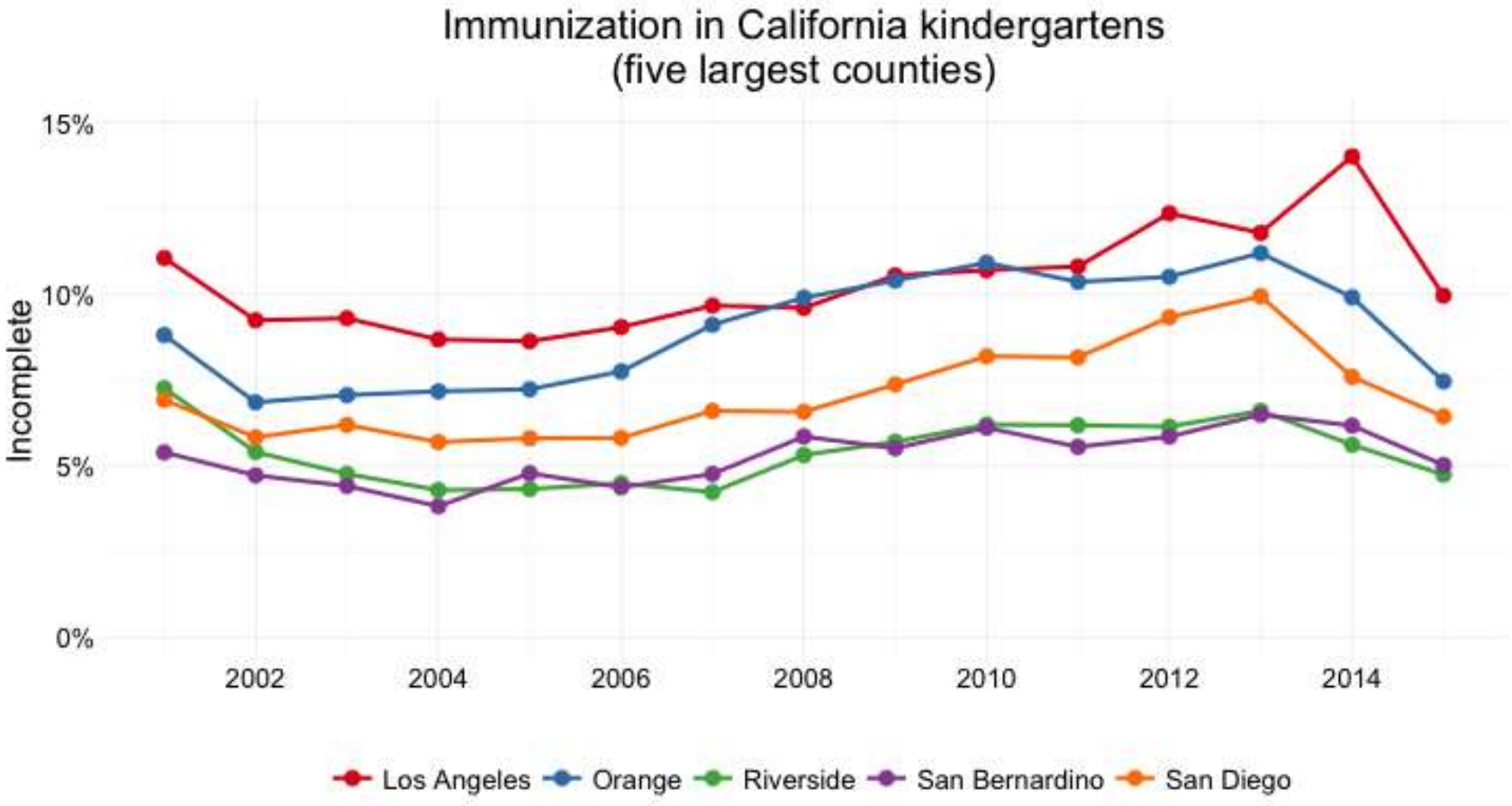


(Source: Peter Aldhous, from California Department of Public Health data)

a dot-and-line chart:

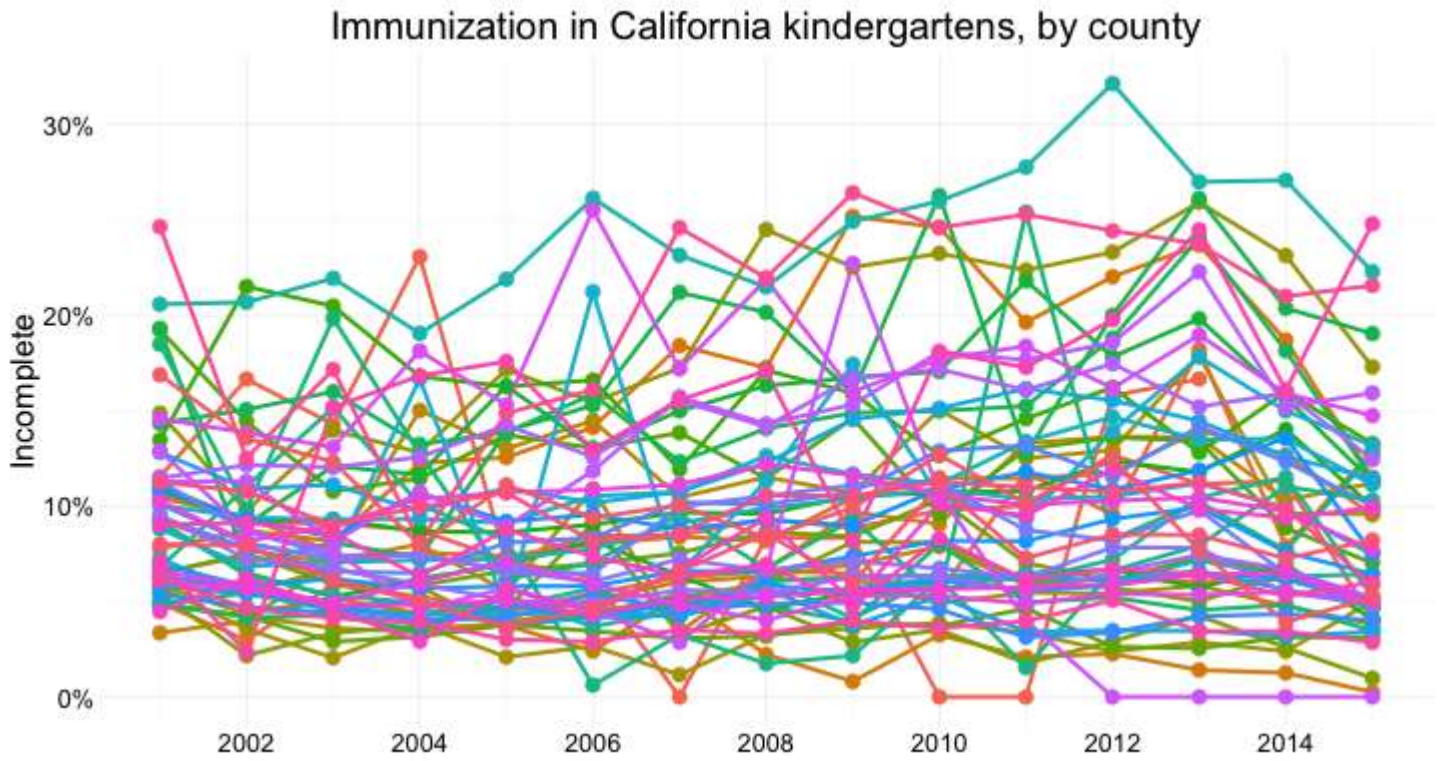


(Source: Peter Aldhous, from California Department of Public Health data)

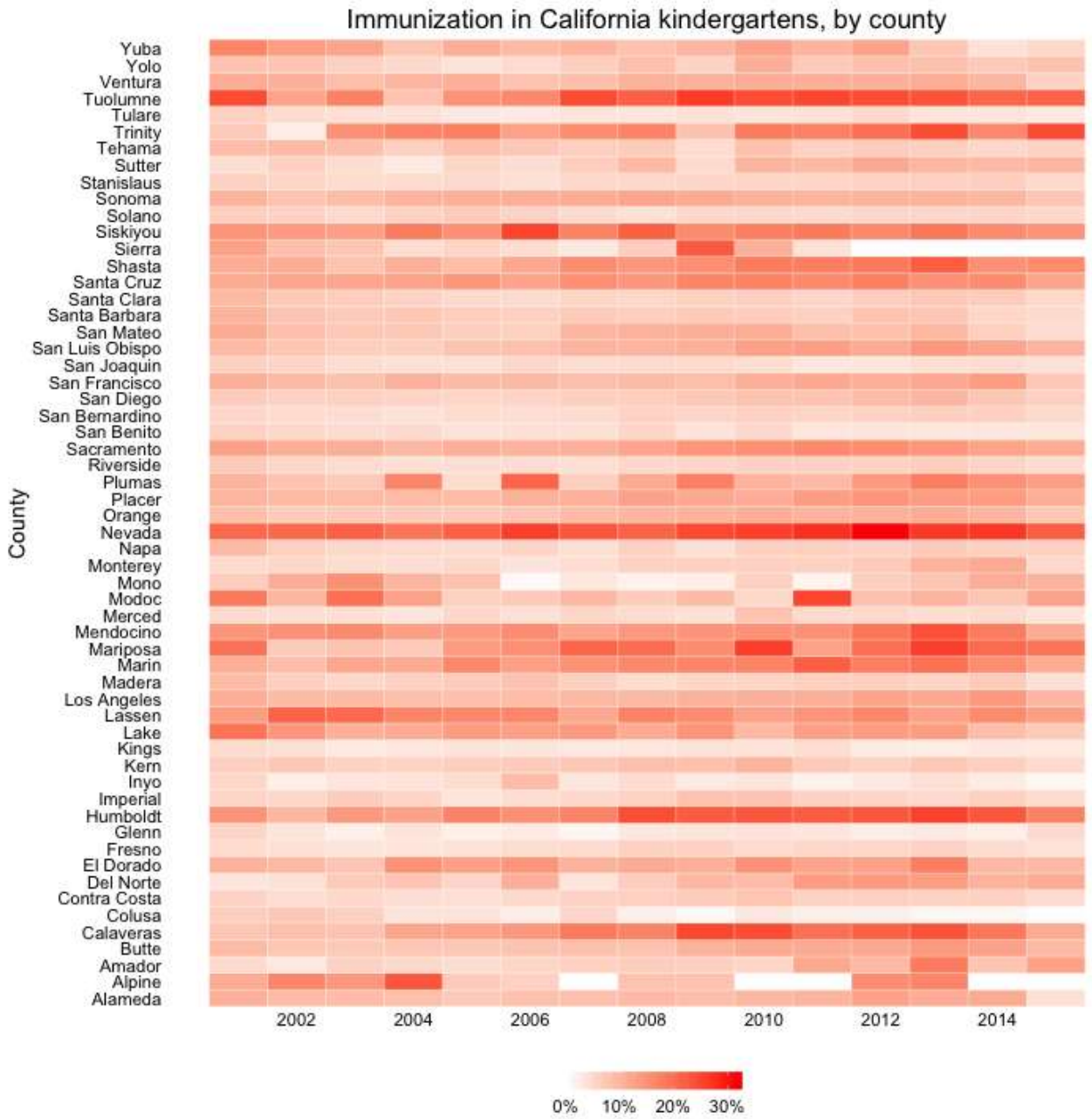


(Source: Peter Aldhous, from California Department of Public Health data)

- But there are 58 counties in California, and trying to compare them all using a dot-and-line chart results in chaos:

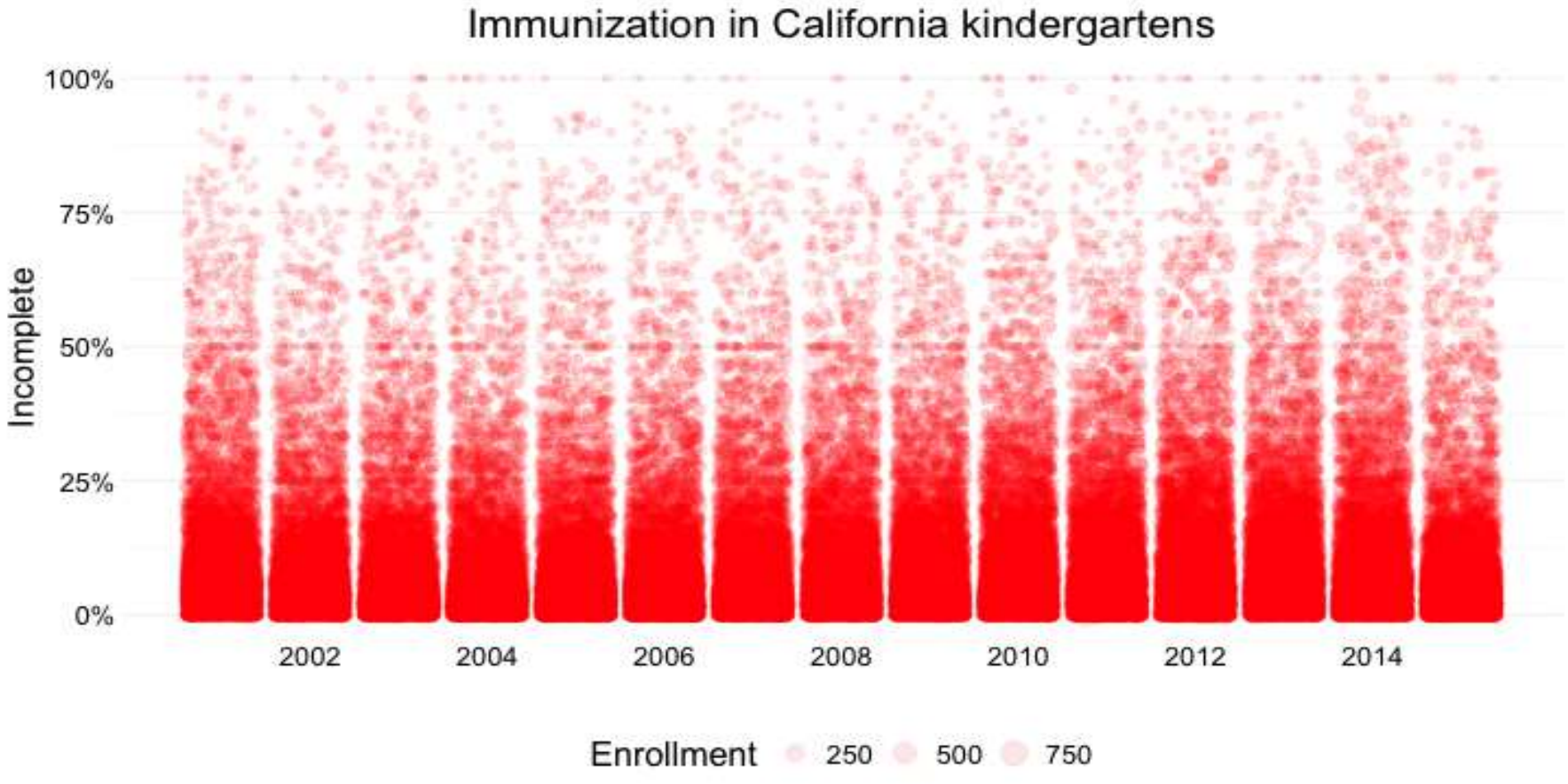


(Source: Peter Aldhous, from California Department of Public Health data)



(Source: Peter Aldhous, from California Department of Public Health data)

- What if we want to visualize the data for every kindergarten on a single chart, to give an overview of how immunization rates vary across schools?



(Source: Peter Aldhous, from California Department of Public Health data)