

DATA VISUALIZATION

Abhijit Dasgupta

WHY VISUALIZE DATA?

To communicate a story

WHY VISUALIZE DATA?

- Data-generating processes can be complex
- The data only mirrors that complexity
- We can use visualization techniques to represent that complexity in a cognitively accessible manner

WHY VISUALIZE DATA?

- Take advantage of our human capability to understand visual patterns quickly and, often, intuitively
- Can be a gateway to interpreting what the data is saying

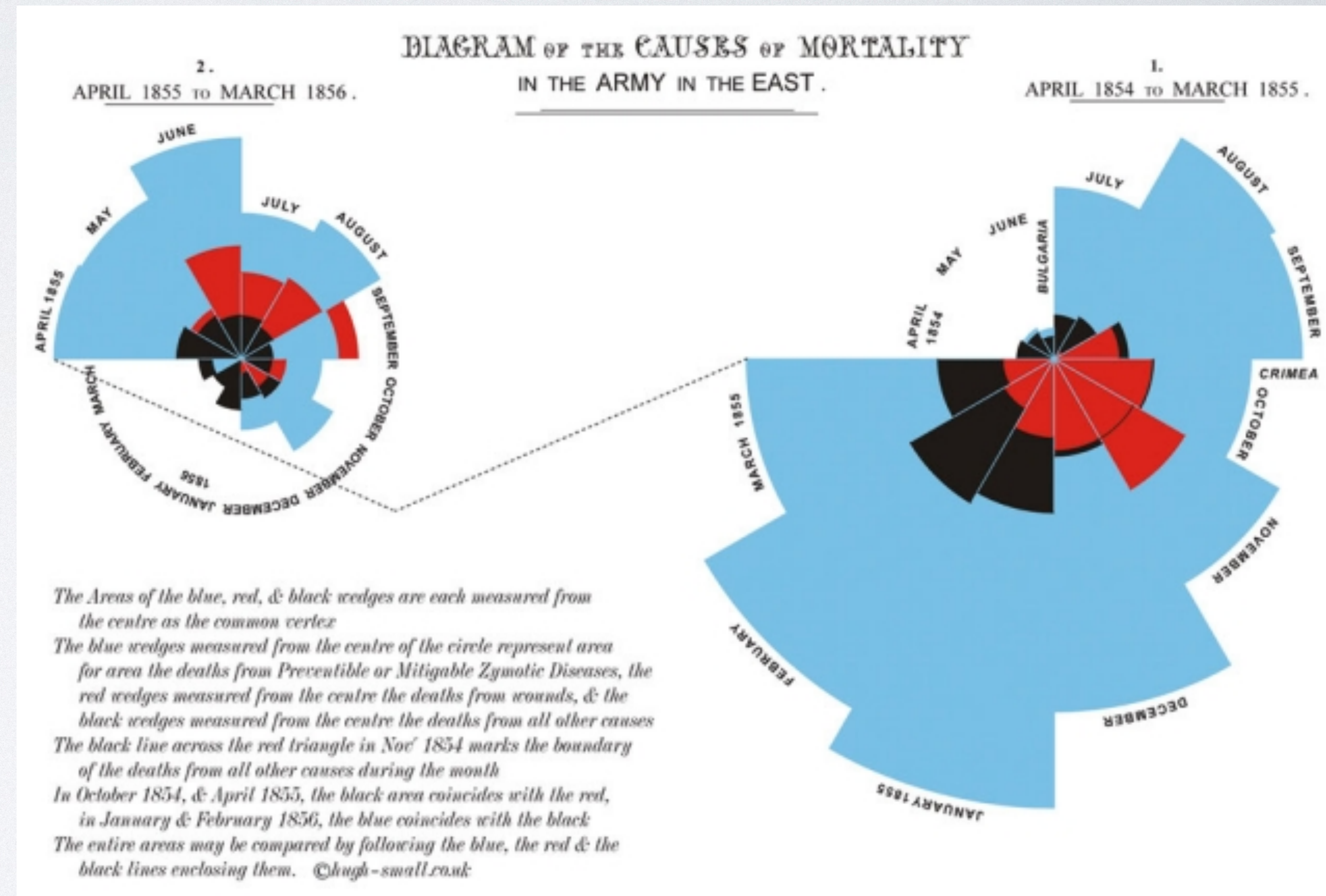
“INSIGHTS”

- Exploration
- Contextual understanding
- Reporting

Data → Information → Evaluation / Action

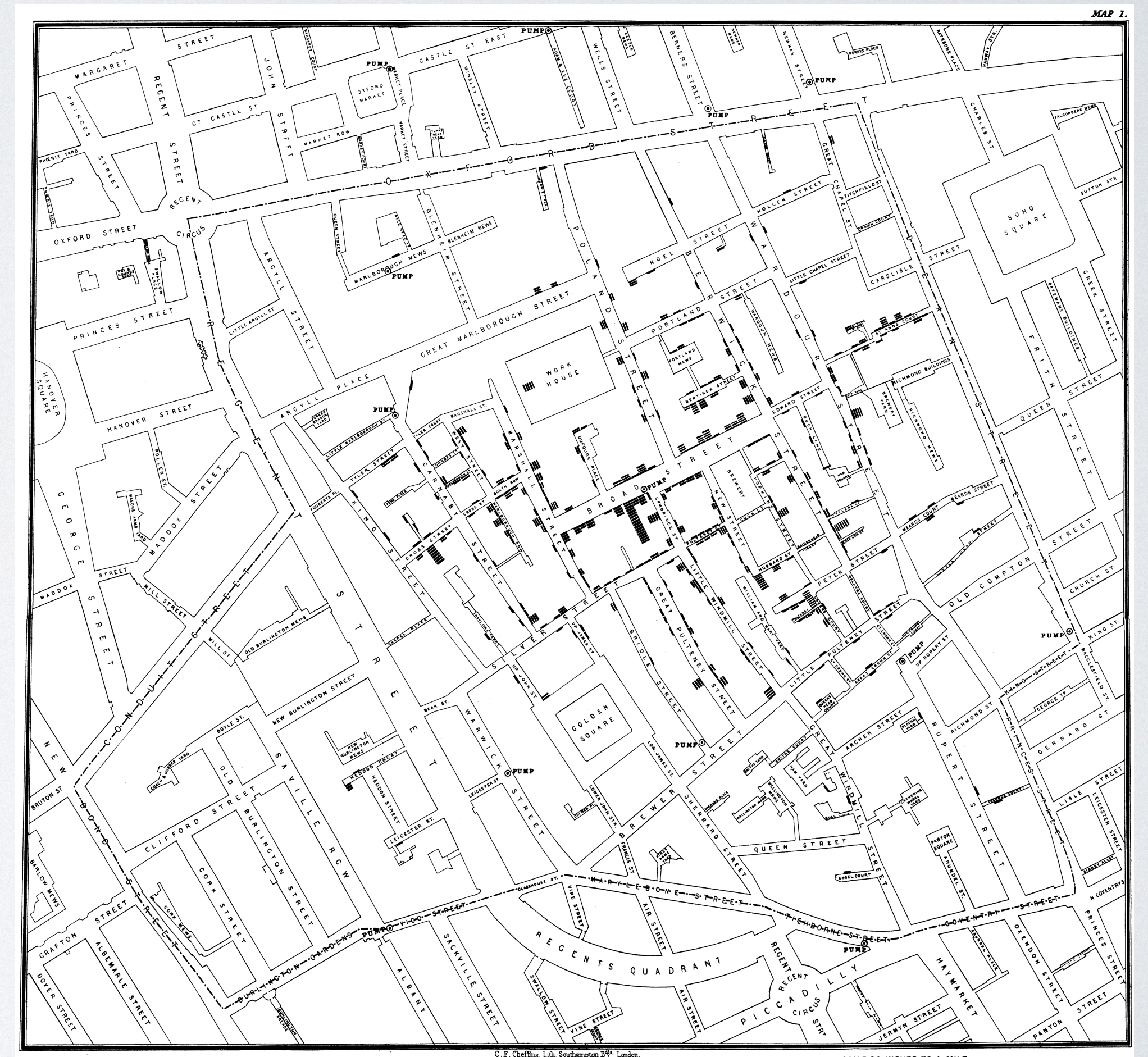
EARLY DAYS

- Florence Nightangle
- Message
- Most military deaths caused by infection

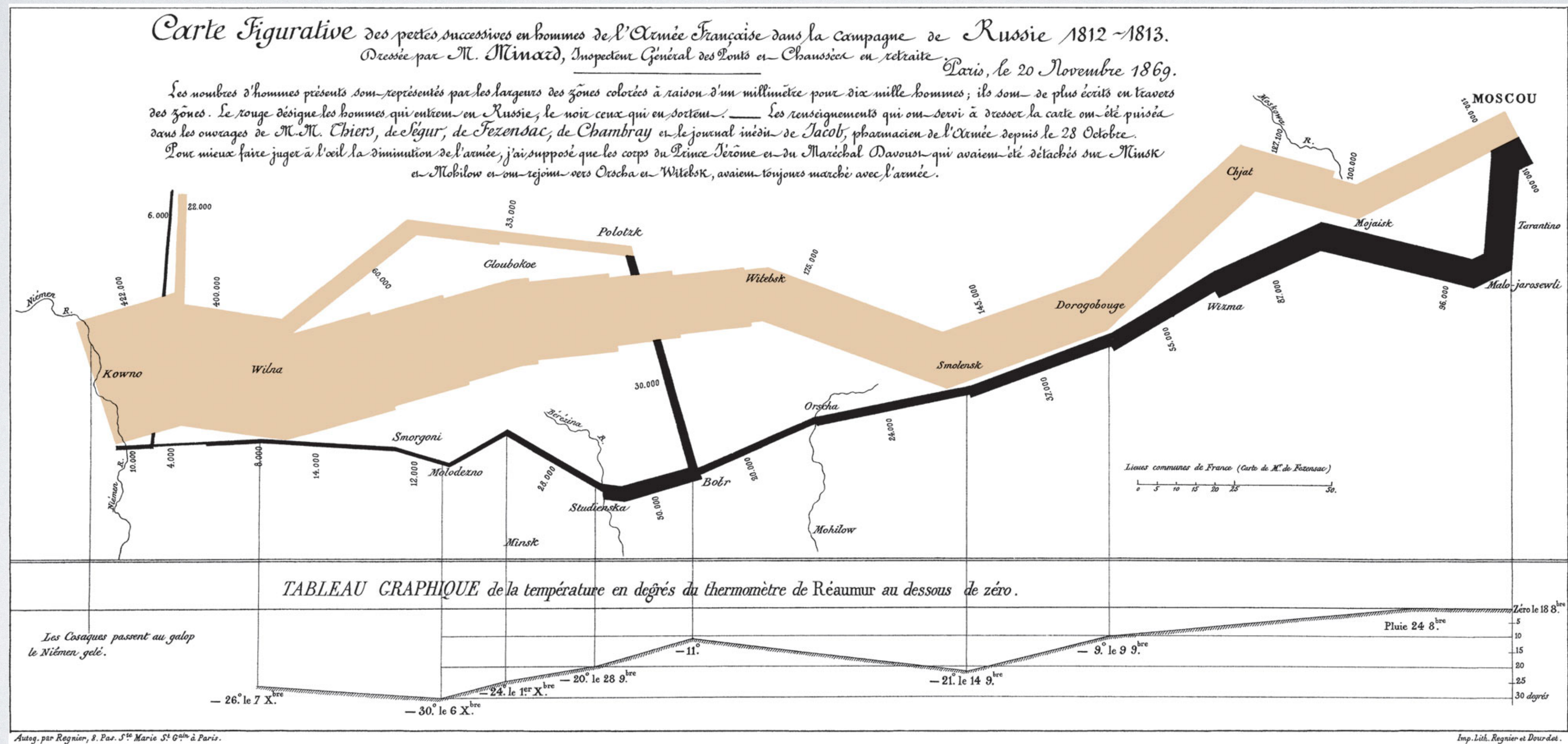


EARLY DAYS

- John Snow, 1854
- Mapping cholera outbreak in London
- Association with certain wells



EARLY DAYS



Minard, 1869

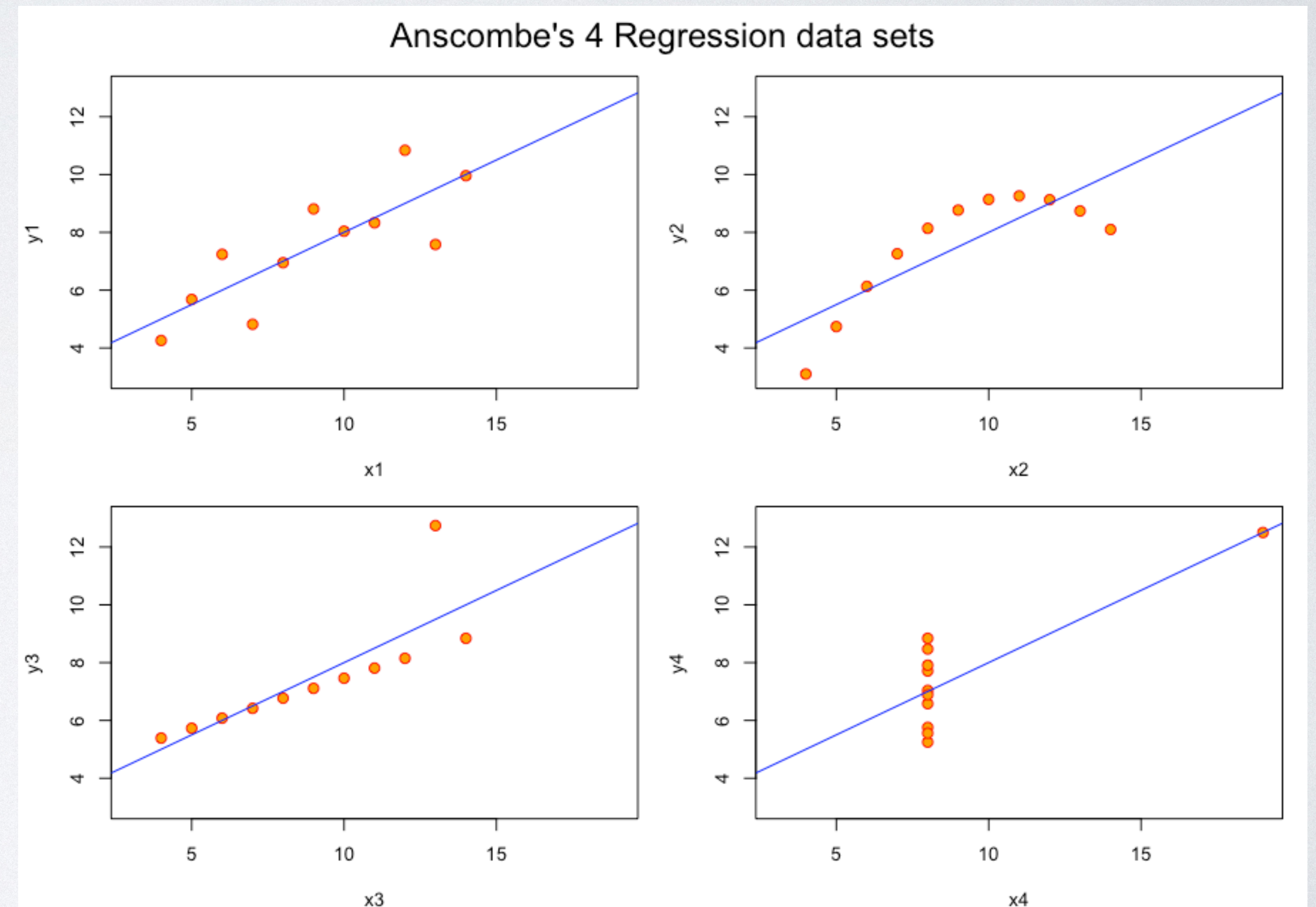
NUMBERS DON'T TELL THE
WHOLE STORY

TYPICAL SUMMARIES

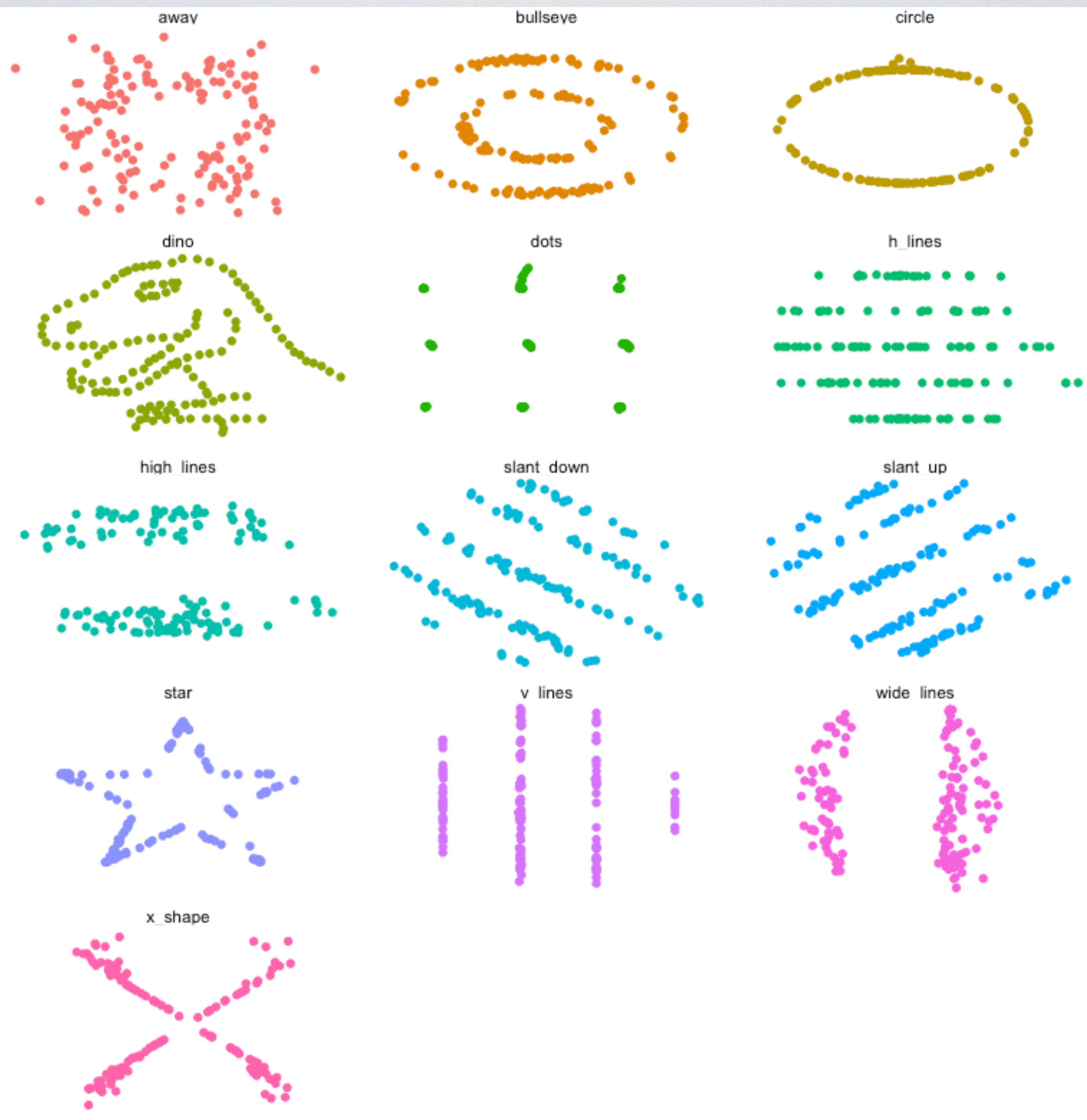
- We compare
 - averages
 - variances
 - correlations

SUMMARIES DON'T DIFFERENTIATE

- Anscombe (1973) created this toy examples
- Averages of x and y are the same
- Correlation between x and y are the same
- Relationships are VERY different

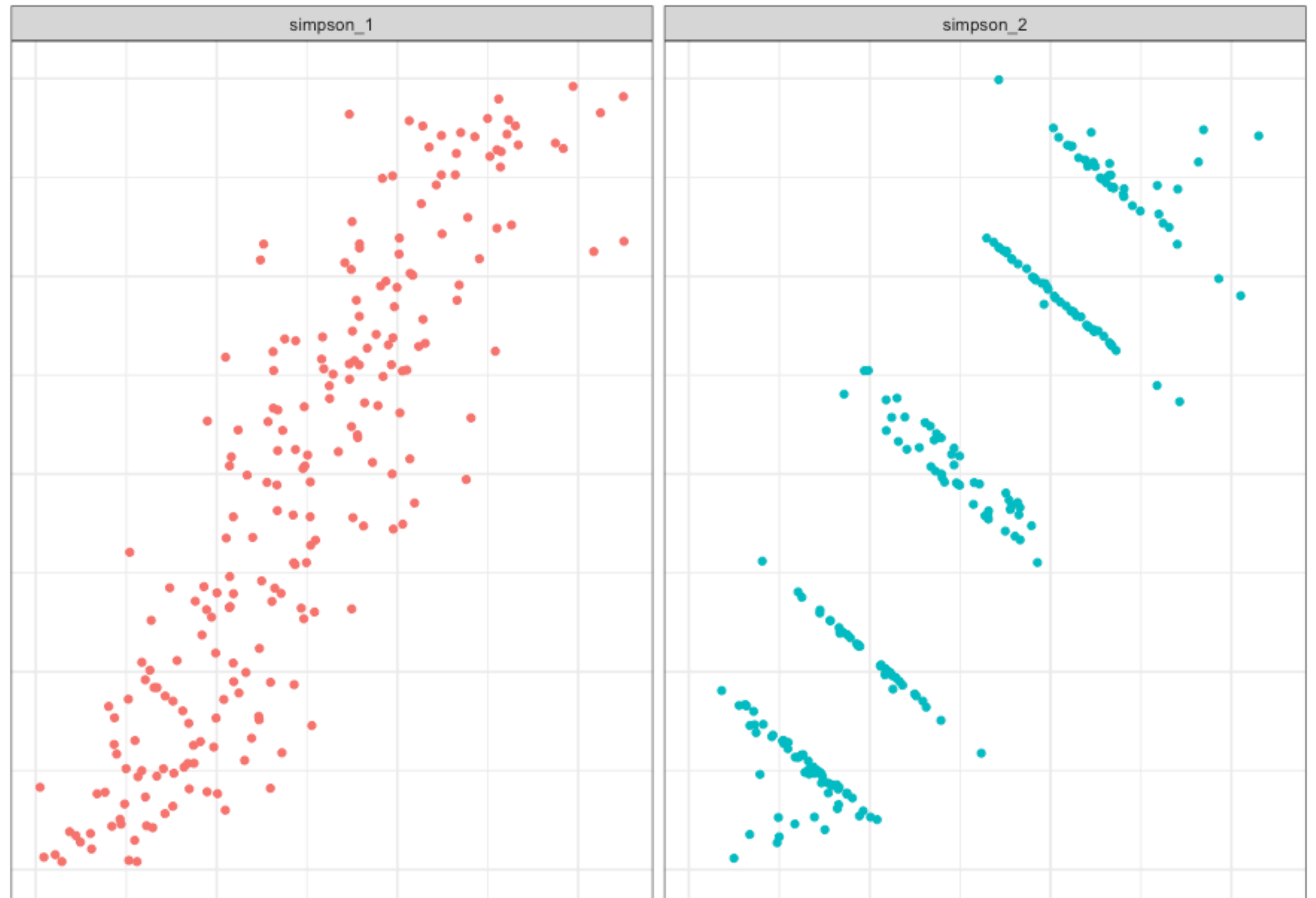


The Datasaurus
Matejka & Fitzmaurice, 2017

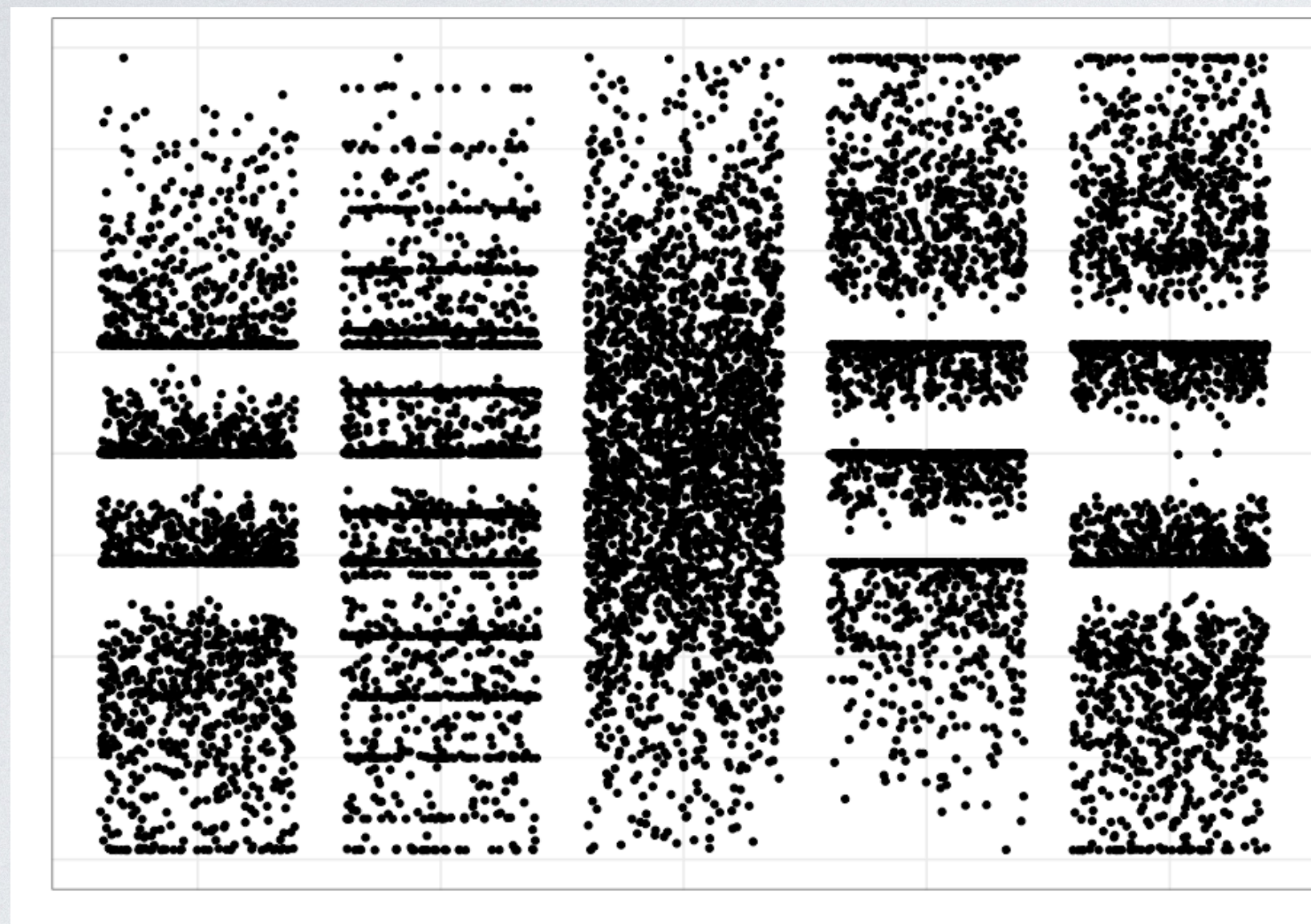


Same positive correlation
overall

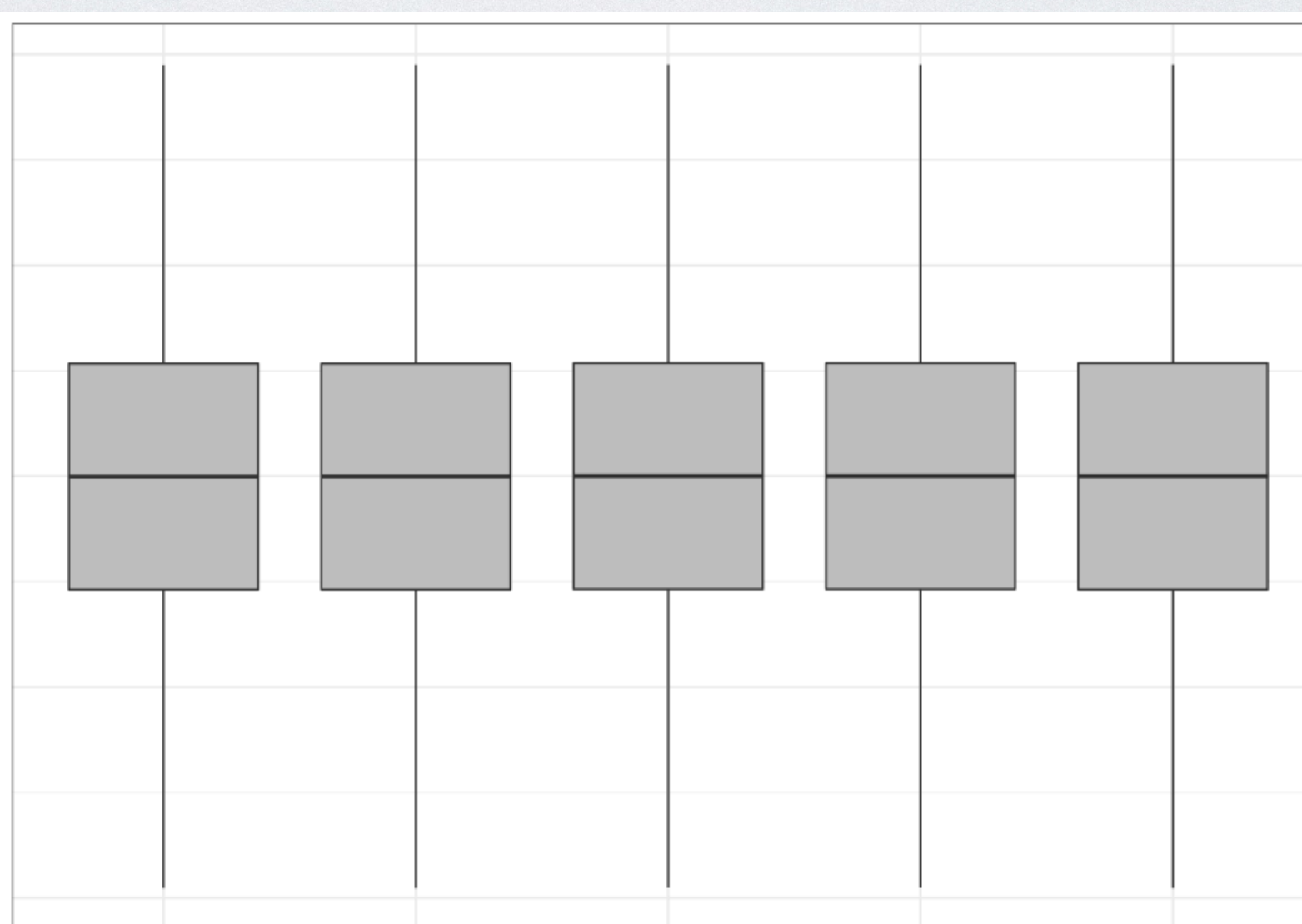
Negative correlation
within groups



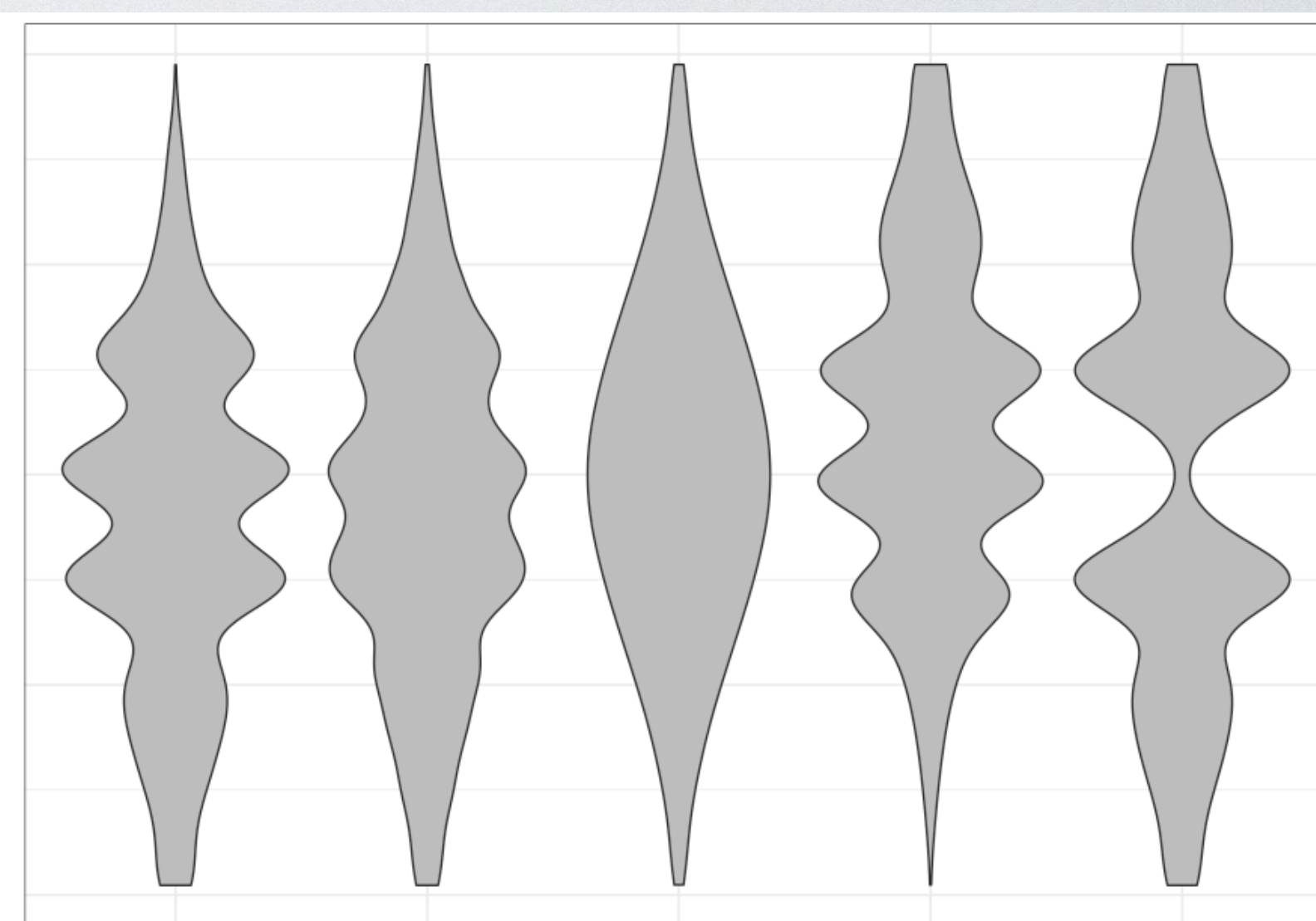
PICTURES DON'T DIFFERENTIATE



Strip plot



Boxplot



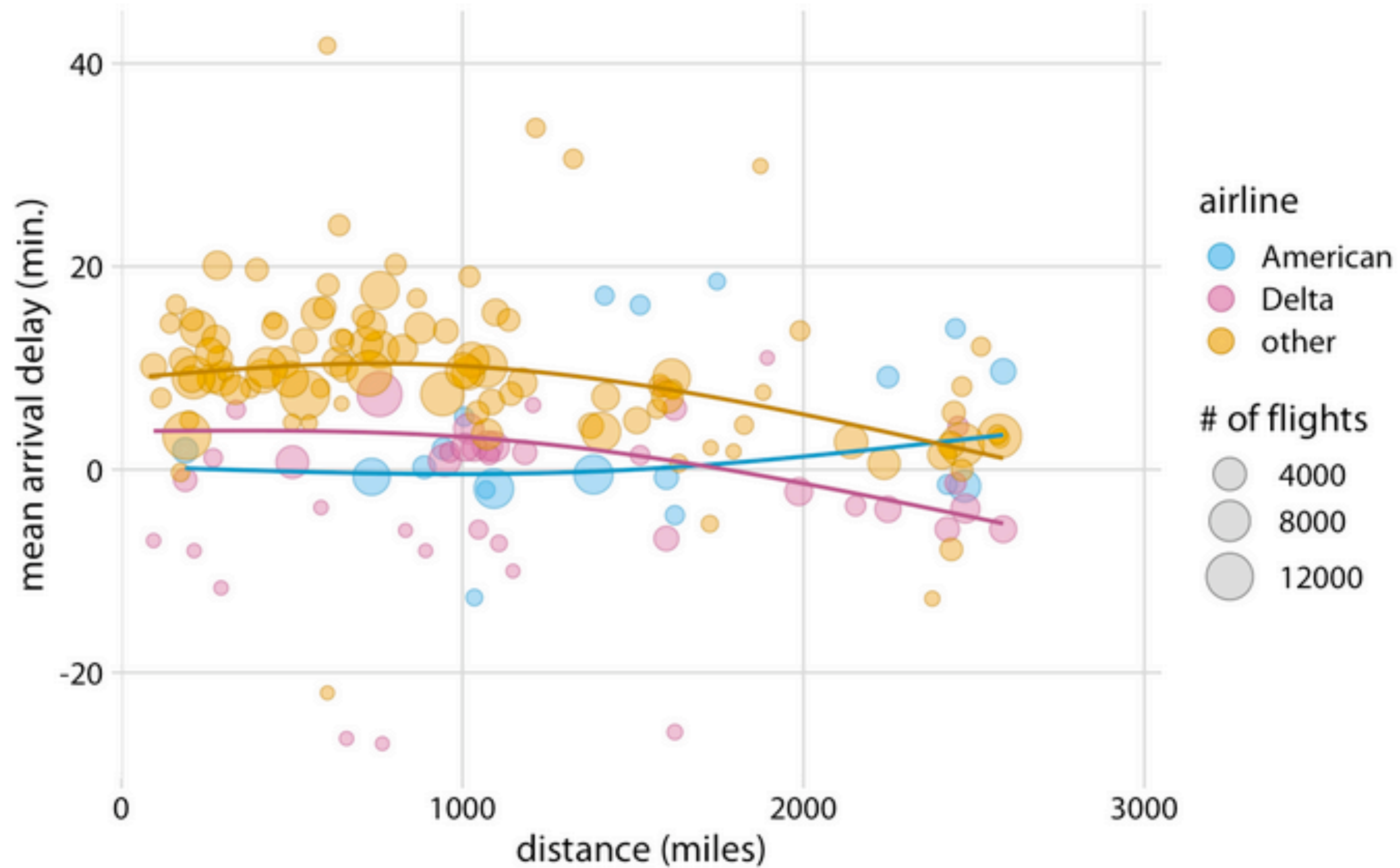
Violin plot

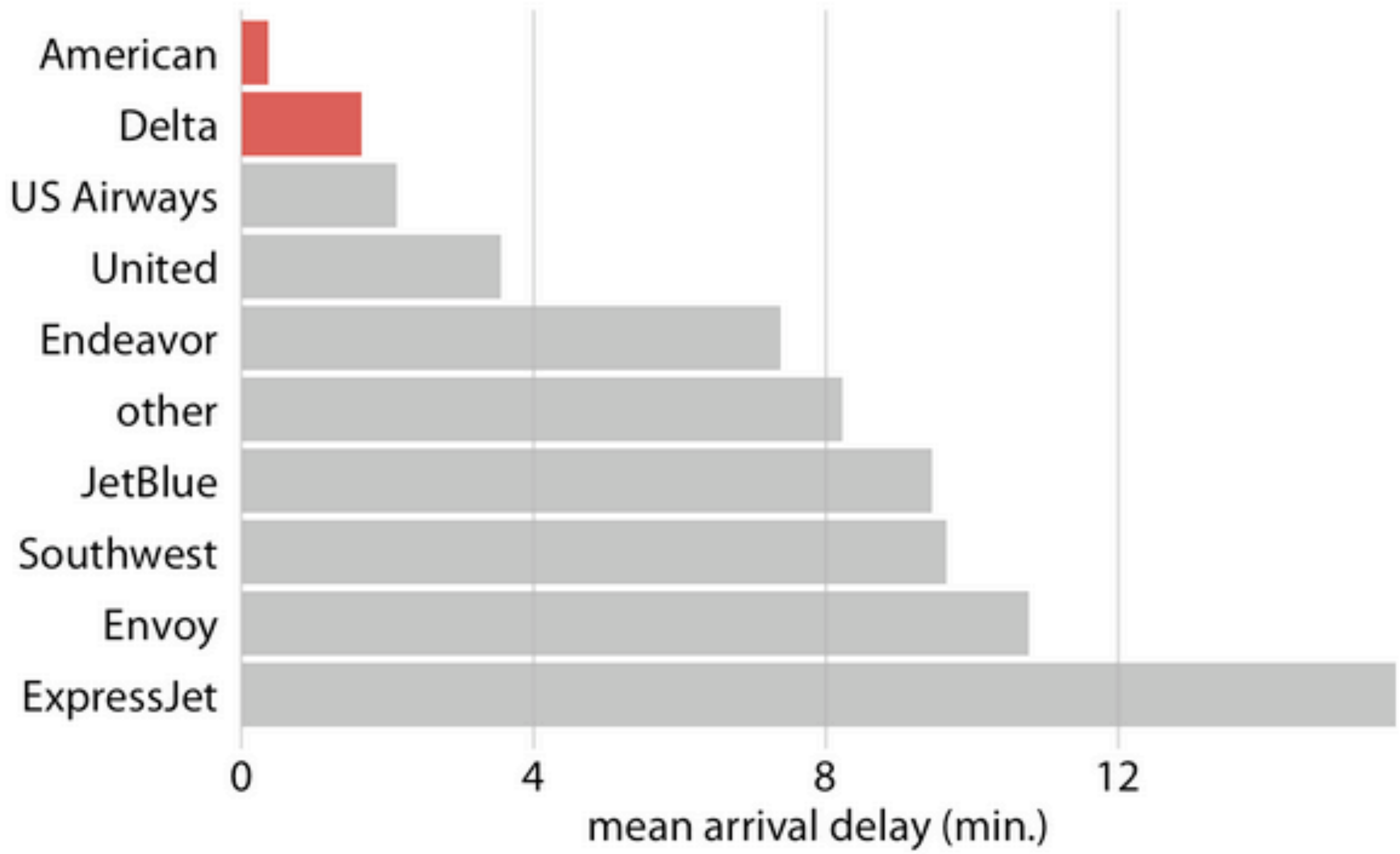
Matejka & Fitzmaurice, 2017

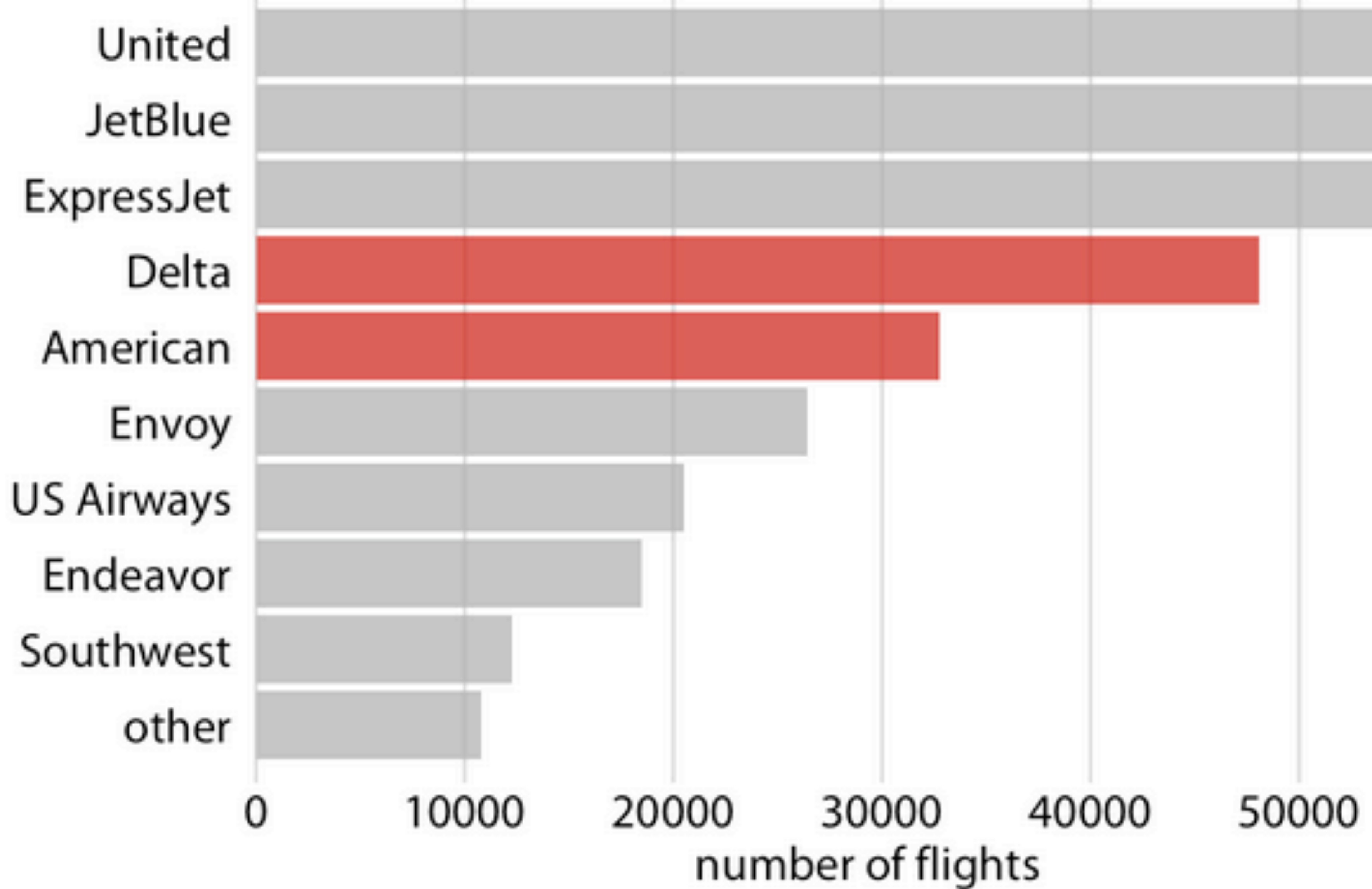
HOW TO THINK ABOUT DATAVIZ

BEGIN WITH THE CONSUMER IN MIND

- You have a deep understanding of the data you're presenting
- The person seeing the visualization DOESN'T
- Develop simpler visualizations first which are easier to explain



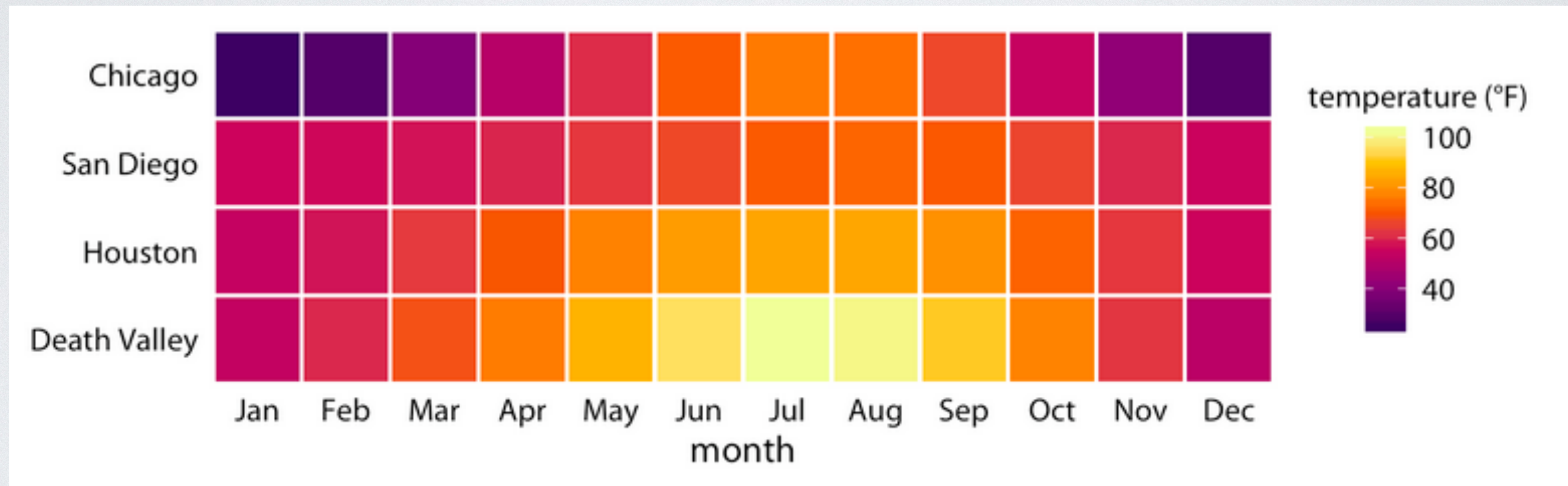




TELL A STORY

- Make sure the graphic is clear
- Make sure the main point you want to make “pops”

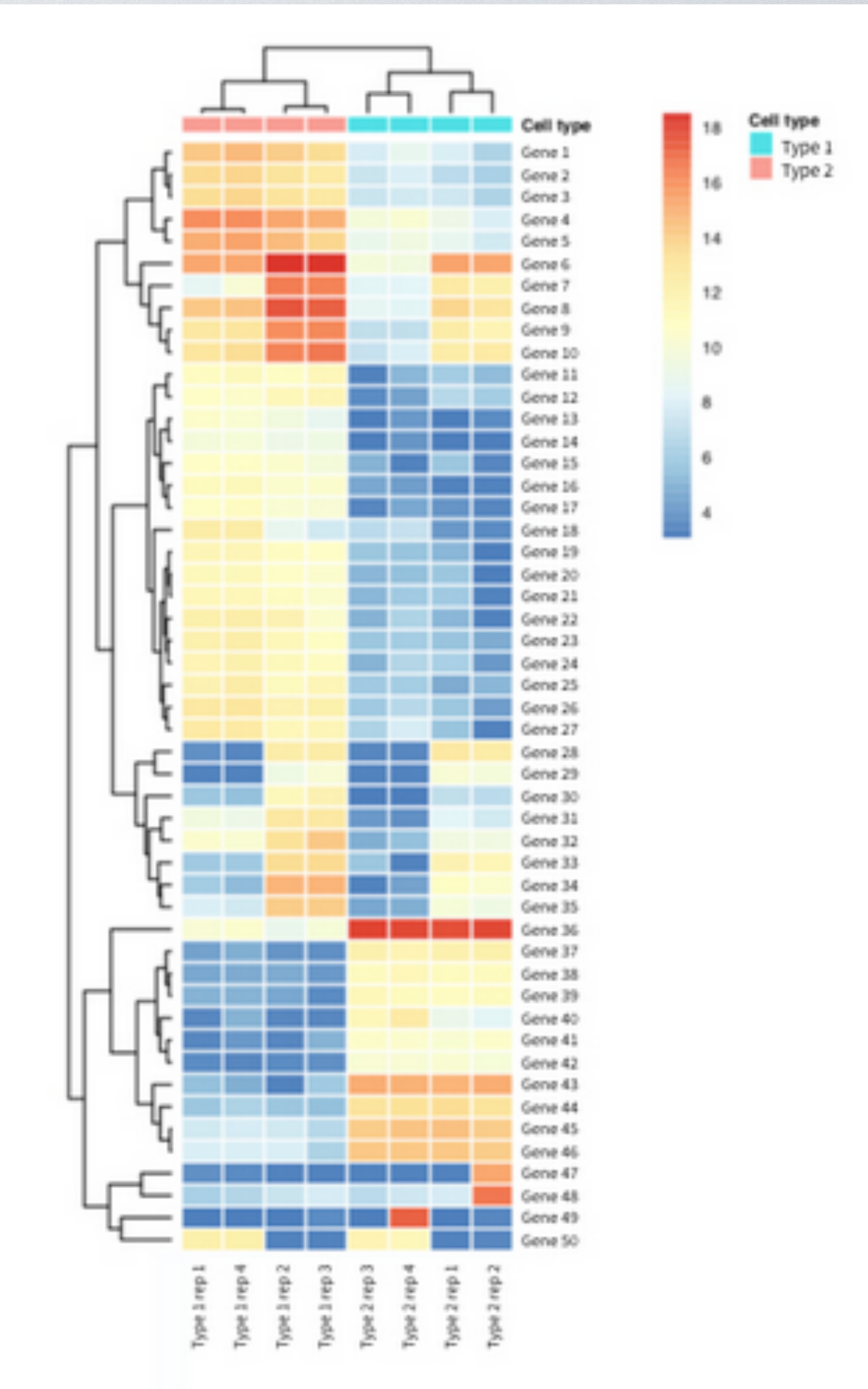
TELL A STORY



TELL A STORY

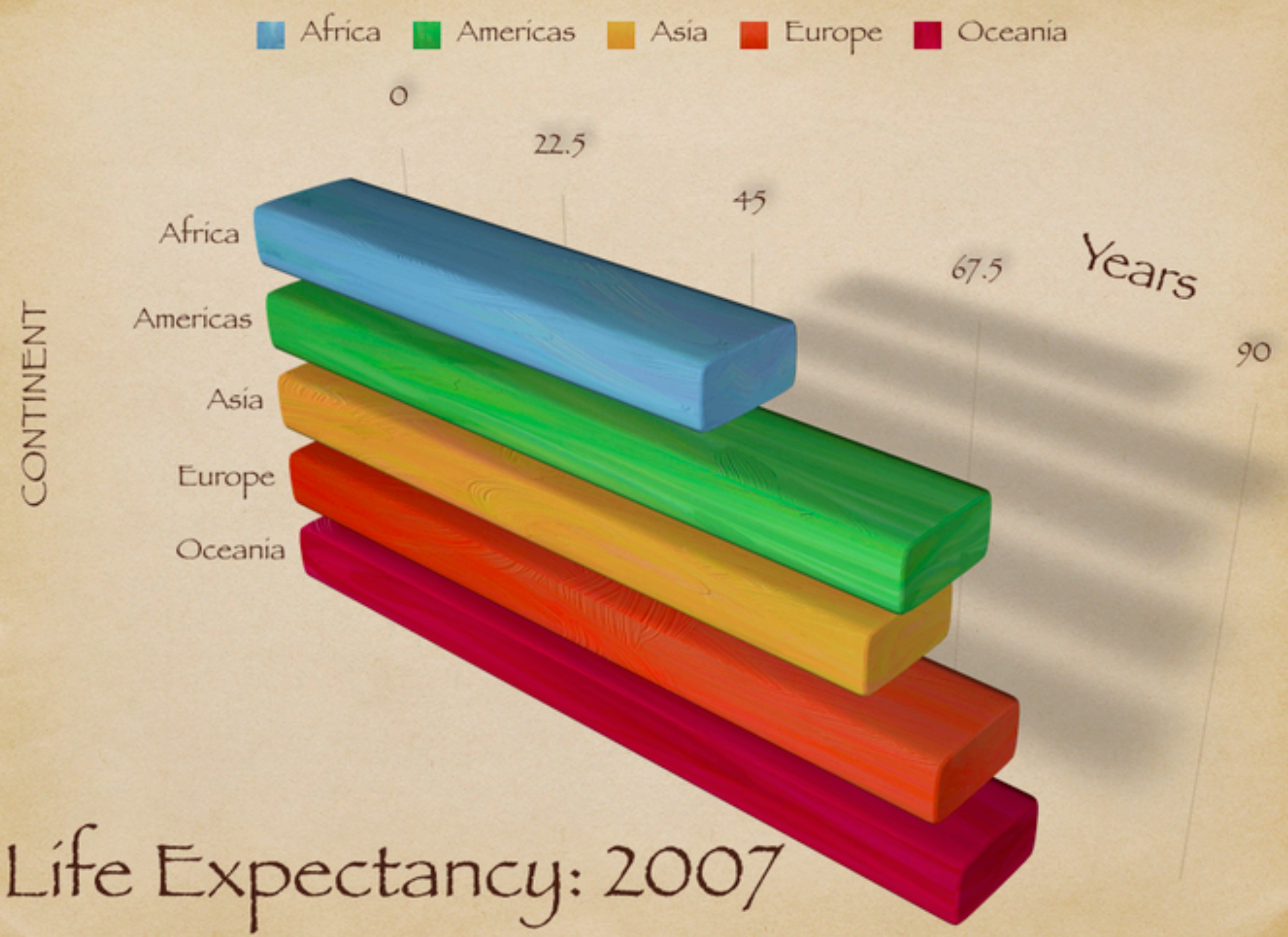


TELL A STORY



WHAT MAKES A GOOD VISUALIZATION?

WHAT MAKES A BAD VISUALIZATION?



Life Expectancy: 2007

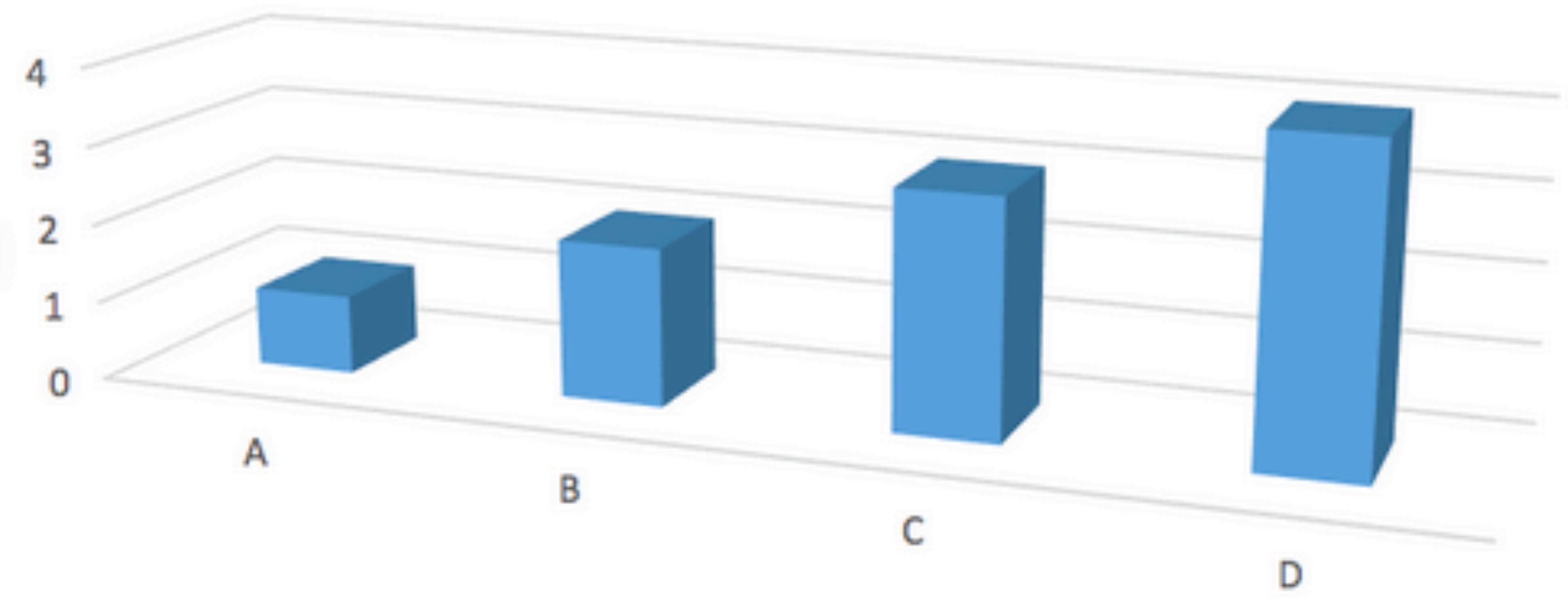
MONSTROUS COSTS

Total House and Senate
campaign expenditures,
in millions



A	1
B	2
C	3
D	4

3D Column Charts in Excel are Awesome





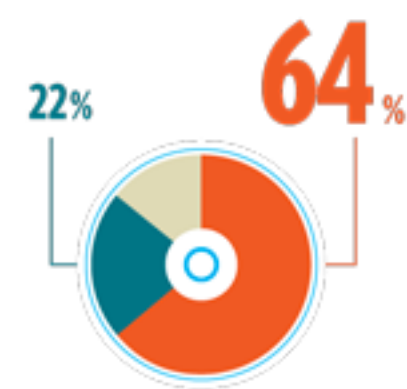


Contrairement au petit écran, les salles de cinéma ne sont pas sujettes à des contraintes de diffusion. Cependant, les obligations faites aux chaînes de financer la production française influence l'offre cinématographique.

SUR LE PETIT ÉCRAN... en 2014

42% des films diffusés à la télévision sont français

37% des films diffusés à la télévision sont américains



Chiffre d'affaires du marché français de films en DVD et Blu-ray en 2015.

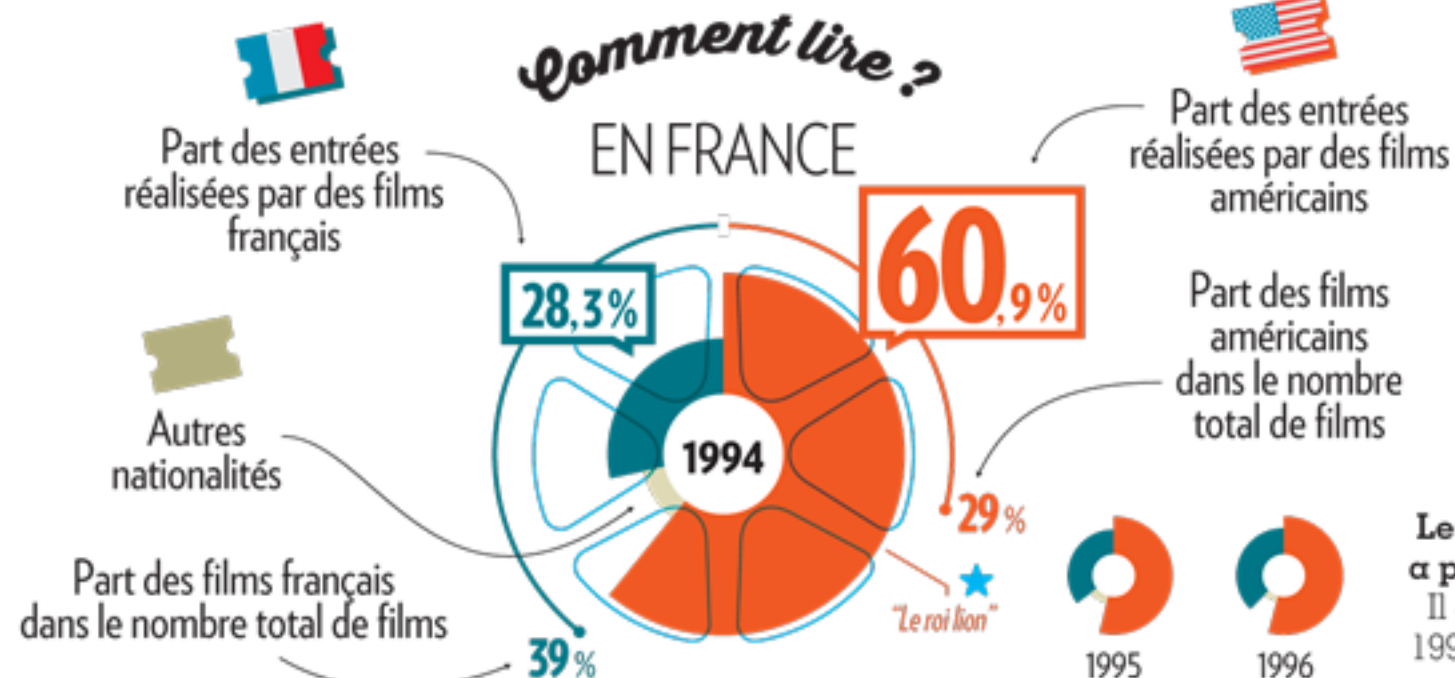


DE MOINS EN MOINS DE FILMS AMÉRICAINS

Depuis vingt ans, **40% des films sortis en salle sont français**. La part des films américains représentait environ un tiers des films dans les années 1990, c'est aujourd'hui plutôt aux alentours d'un quart.

Comment lire ?

EN FRANCE



★ Film succès de l'année

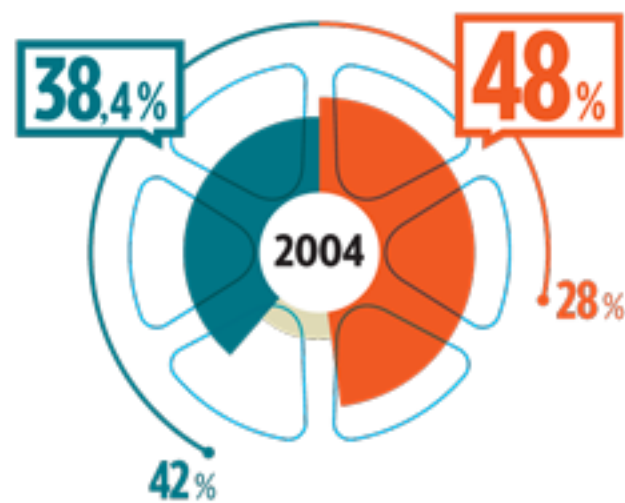
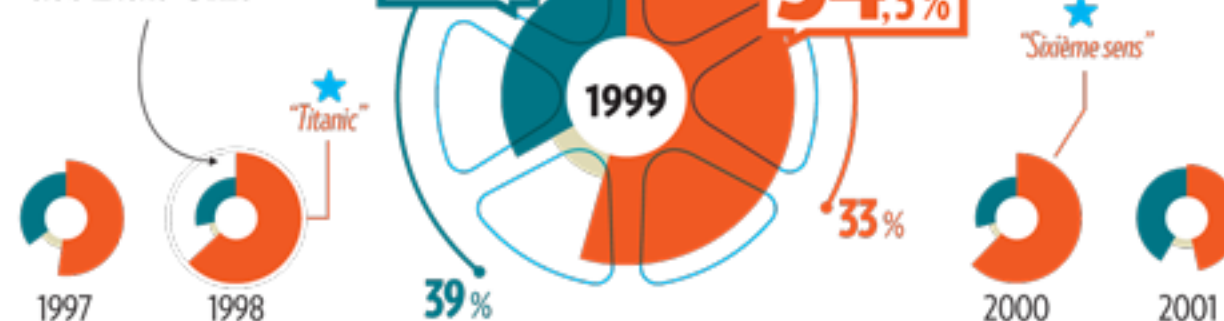


LE CINÉMA FAIT LE PLEIN

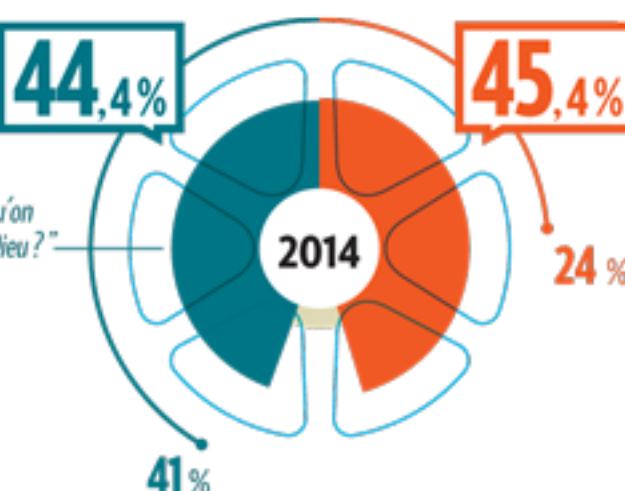
Le nombre d'entrées en salles a presque doublé en vingt ans. Il est passé de 124 millions en 1994, soit 2 entrées en moyenne par personne, à 206 millions, l'équivalent de 3,1 entrées par Français.

Le nombre d'entrées selon la nationalité des films varie beaucoup d'une année à l'autre, car **un grand succès peut bouleverser le classement**.

La meilleure années des Etats-Unis



Le record de la France

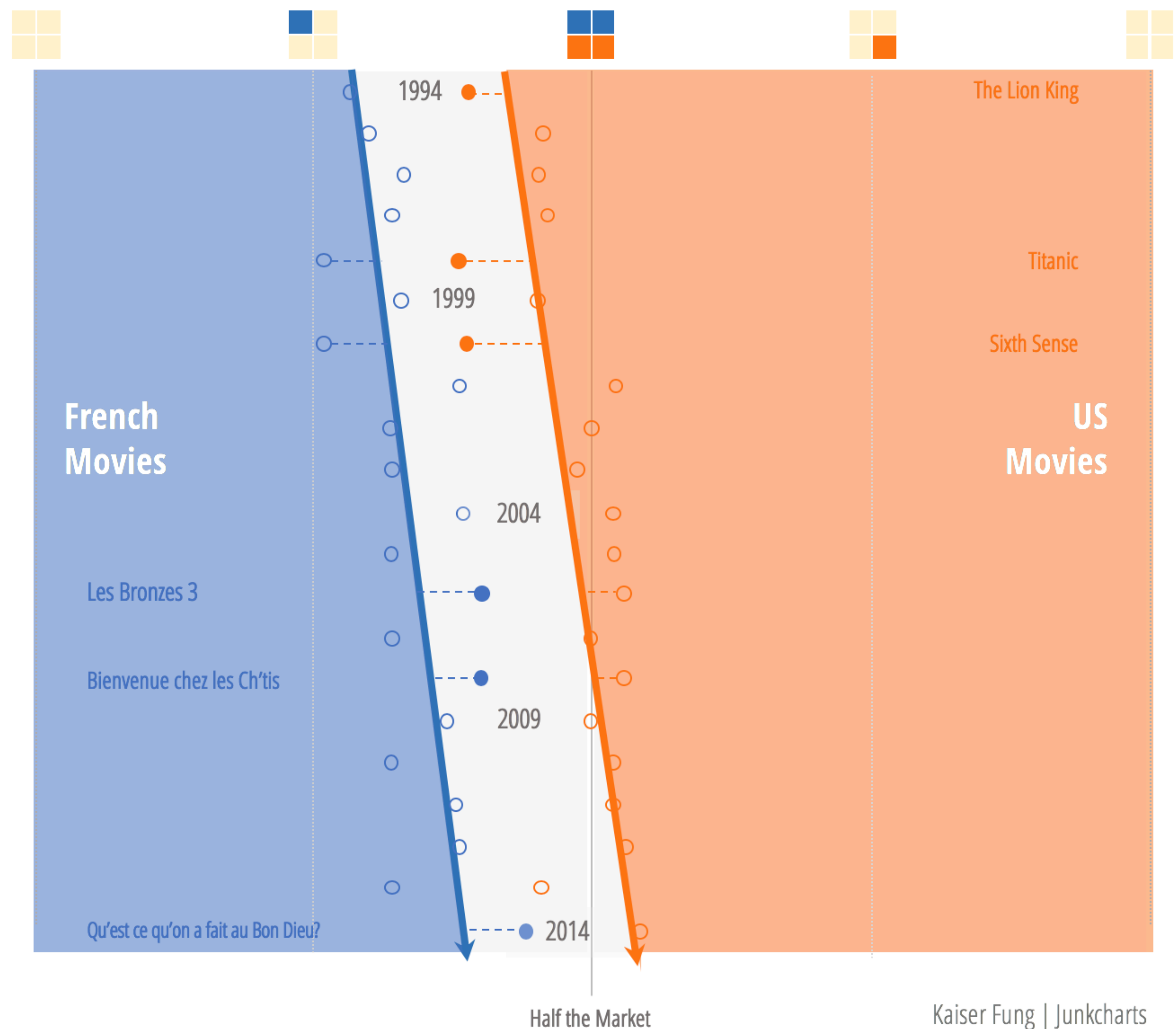


Alors que les films américains ne représentent qu'un quart des films projetés, ils concentrent près de la moitié des entrées en salle.

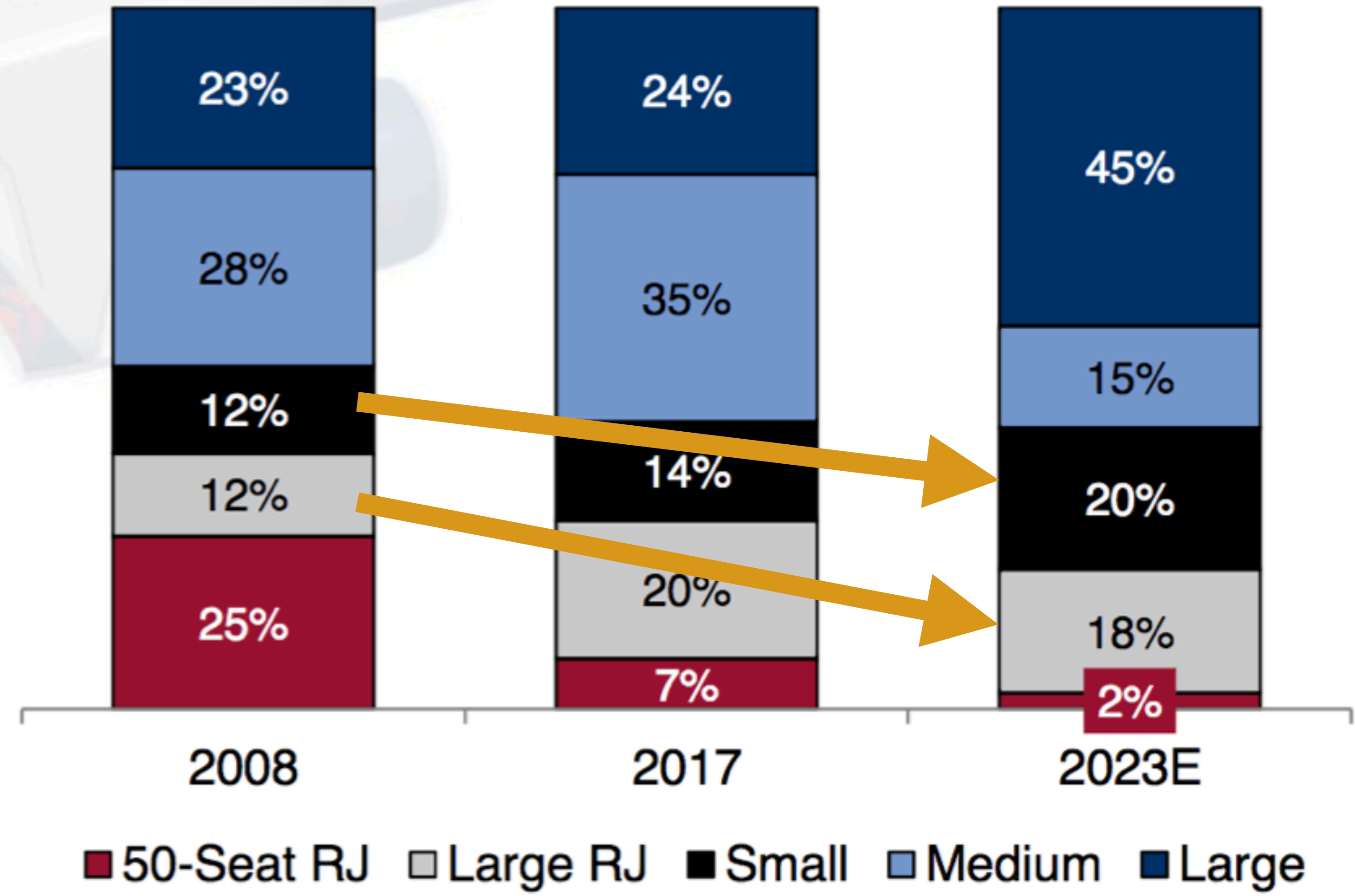
La réponse

Non Si les films américains accueillent toujours près de la moitié du public français, leur part a baissé de presque dix points en vingt ans.

The French are Watching More French Movies

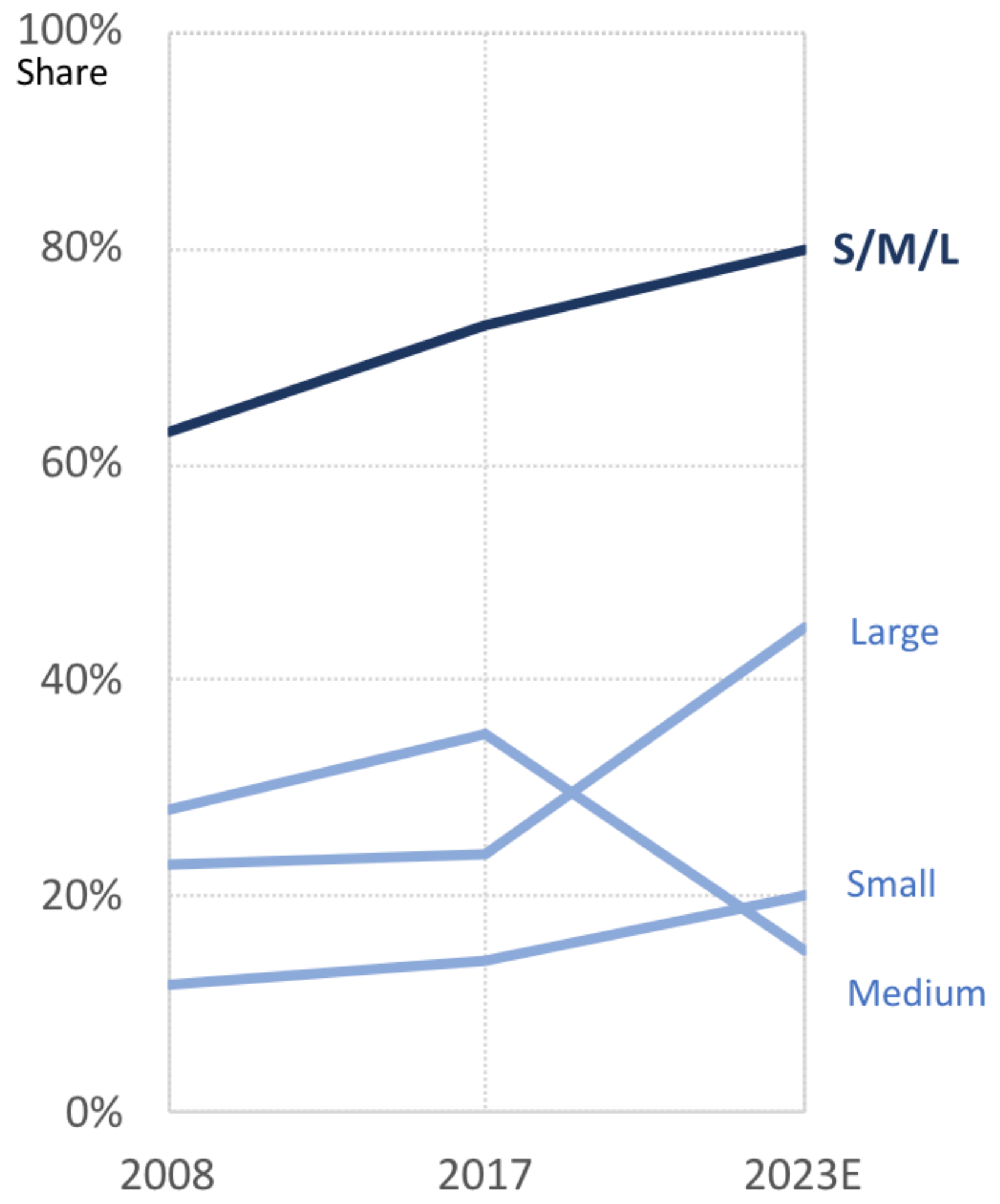


Domestic Seat Departures by Aircraft Type

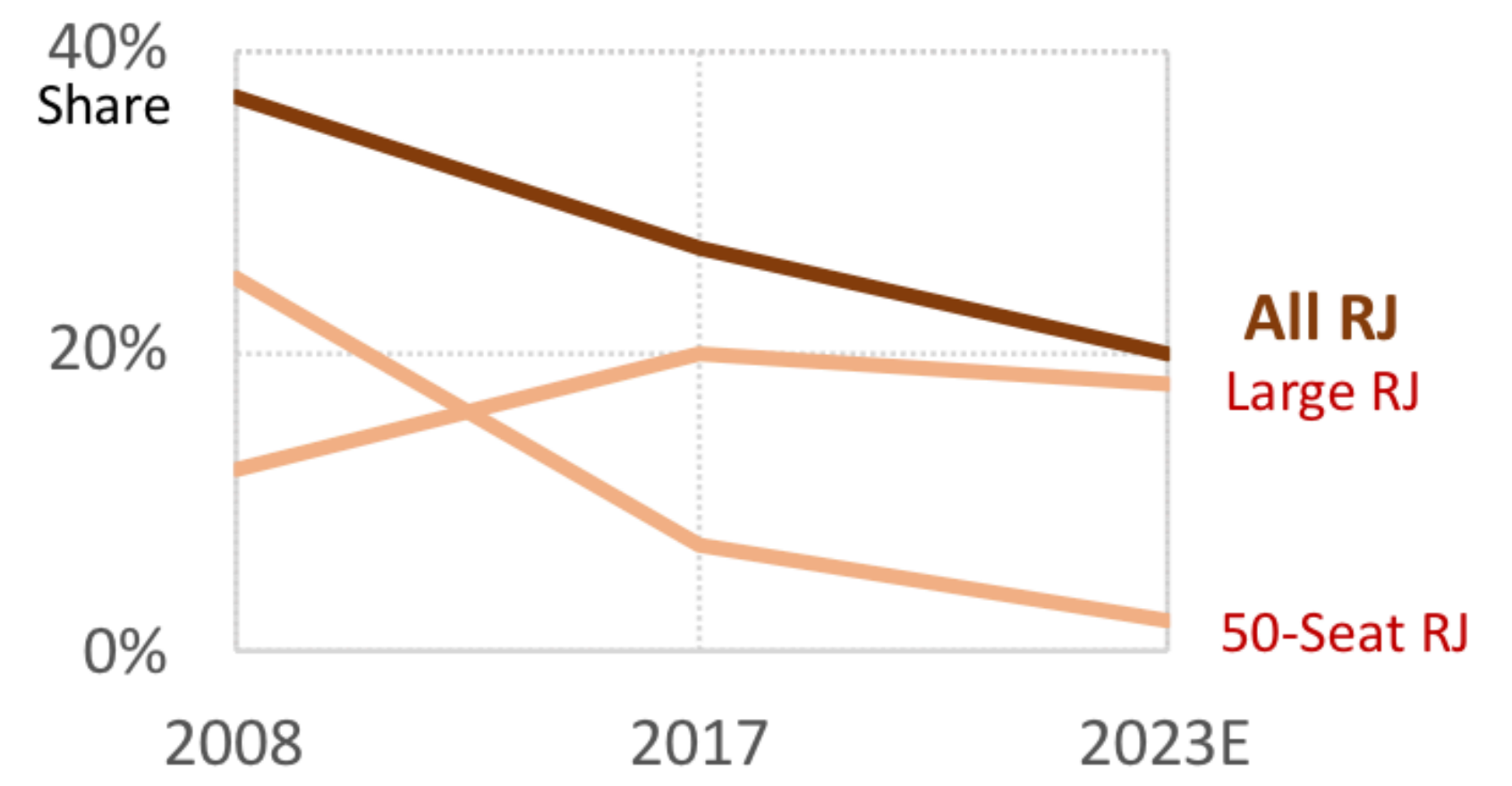


Delta Continues to Upgauge Domestic Aircraft

Growing Small and Large Jets

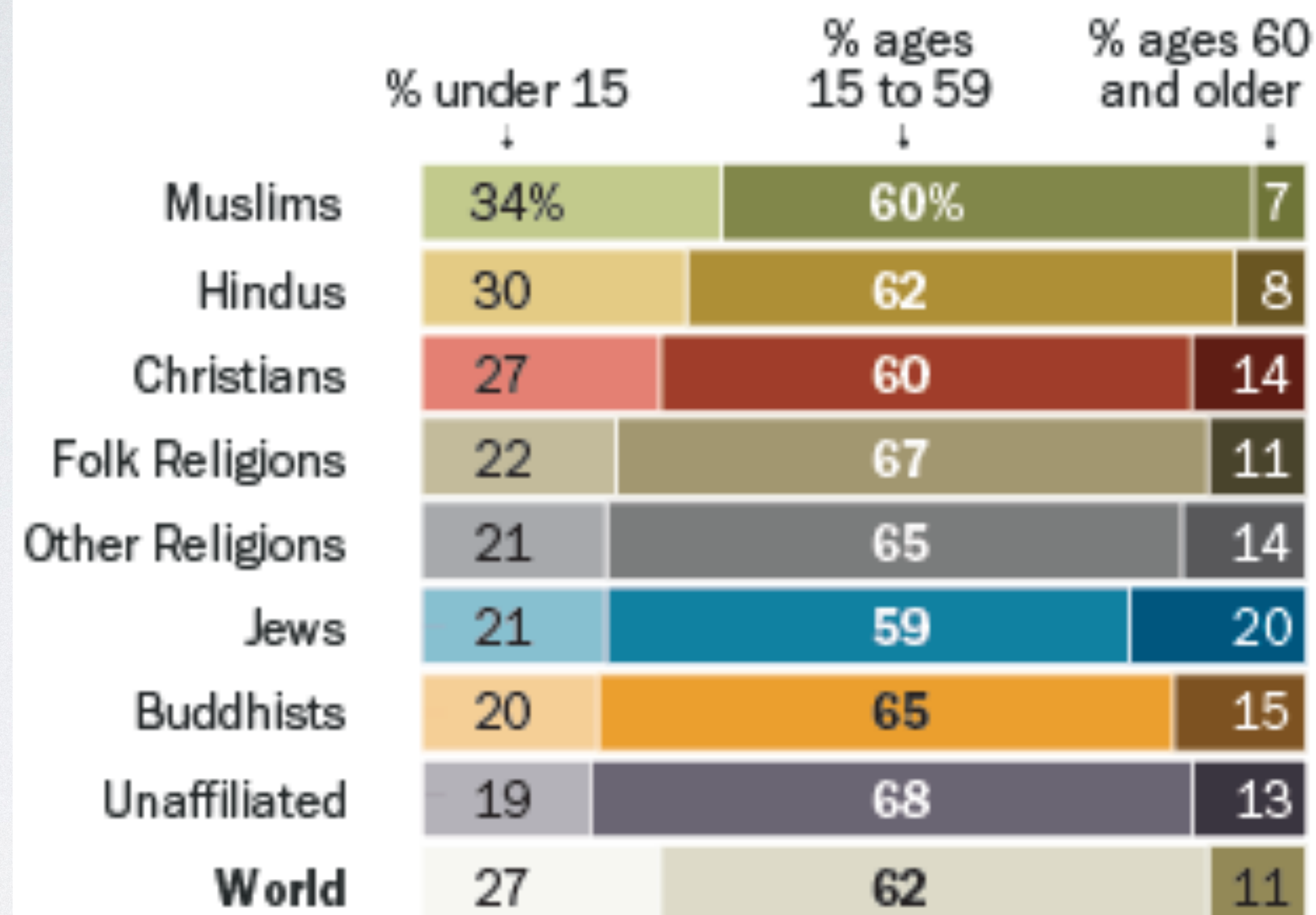


Shrinking RJs



Kaiser Fung / JunkCharts: 2018

Age Distribution of Religious Groups, 2010

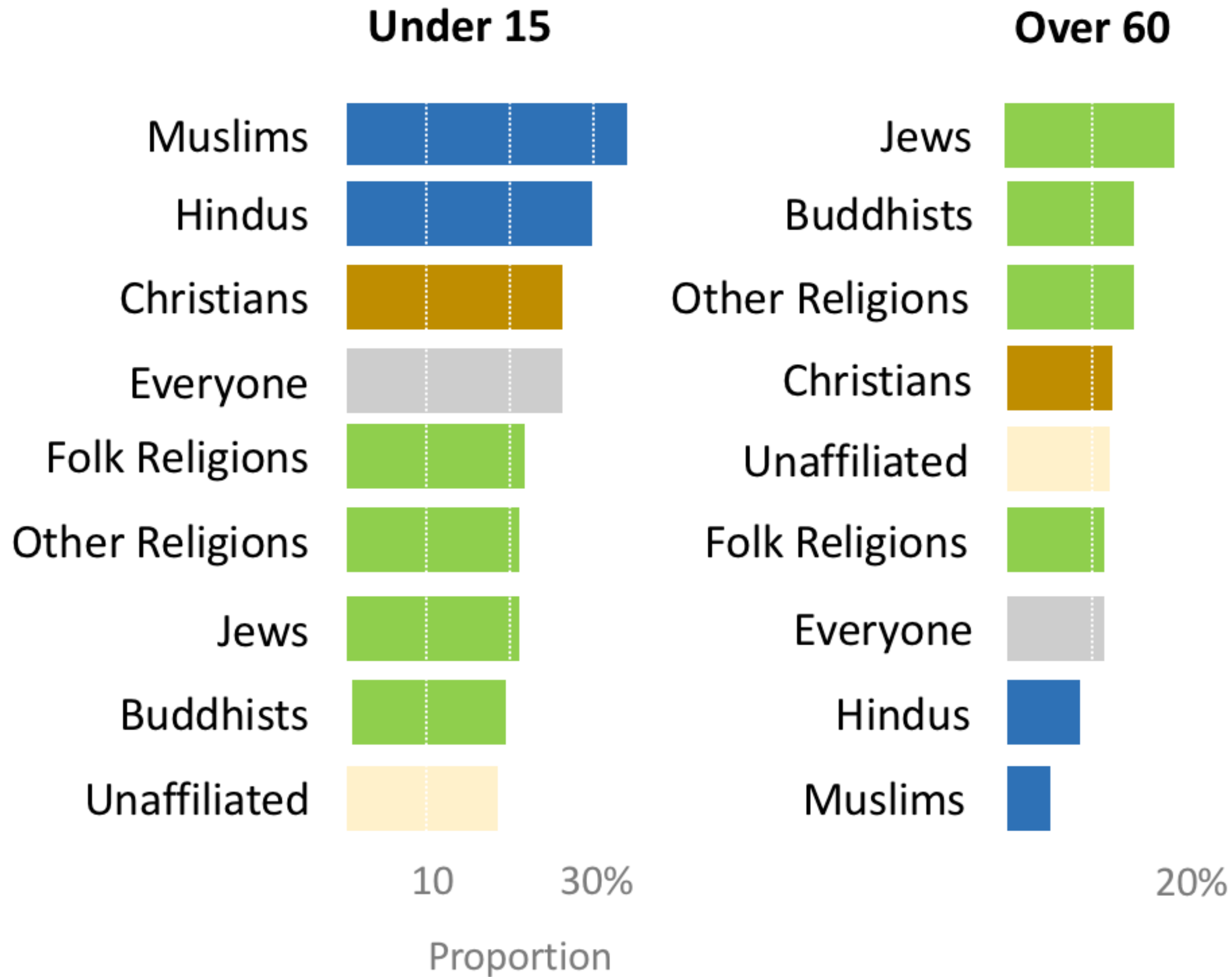


Figures may not add to 100% due to rounding.

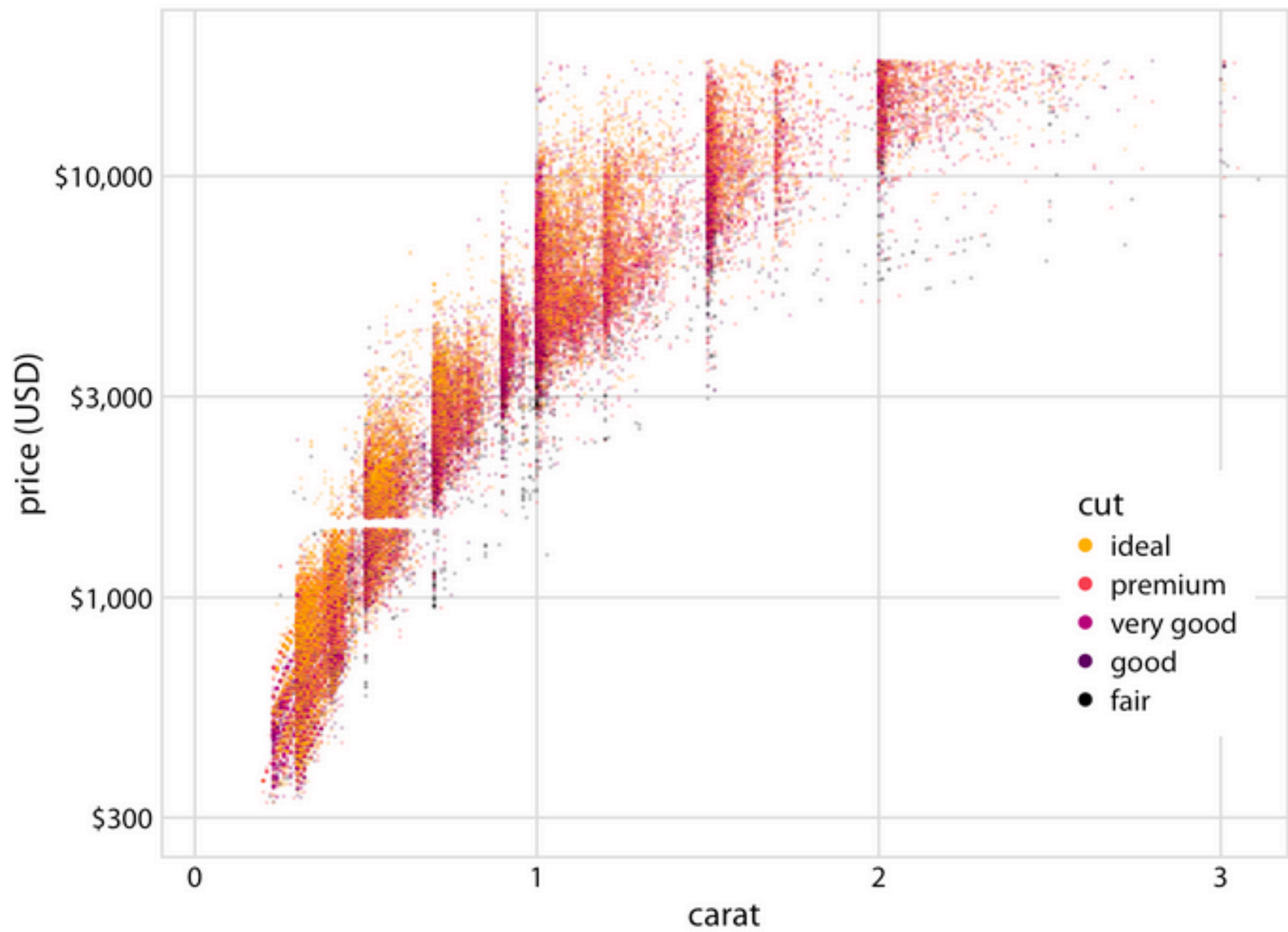
Source: The Future of World Religions: Population Growth Projections, 2010-2050

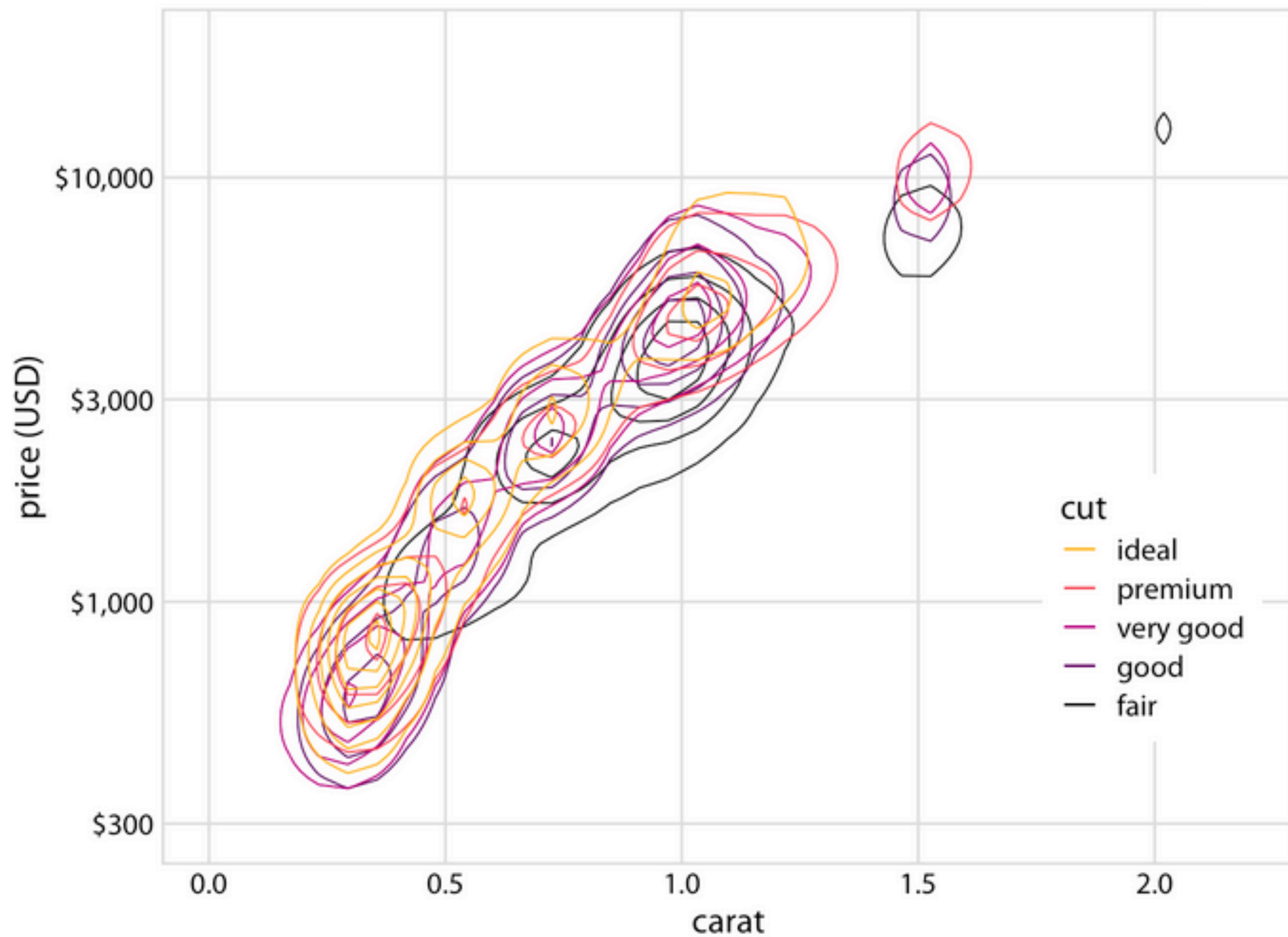
PEW RESEARCH CENTER

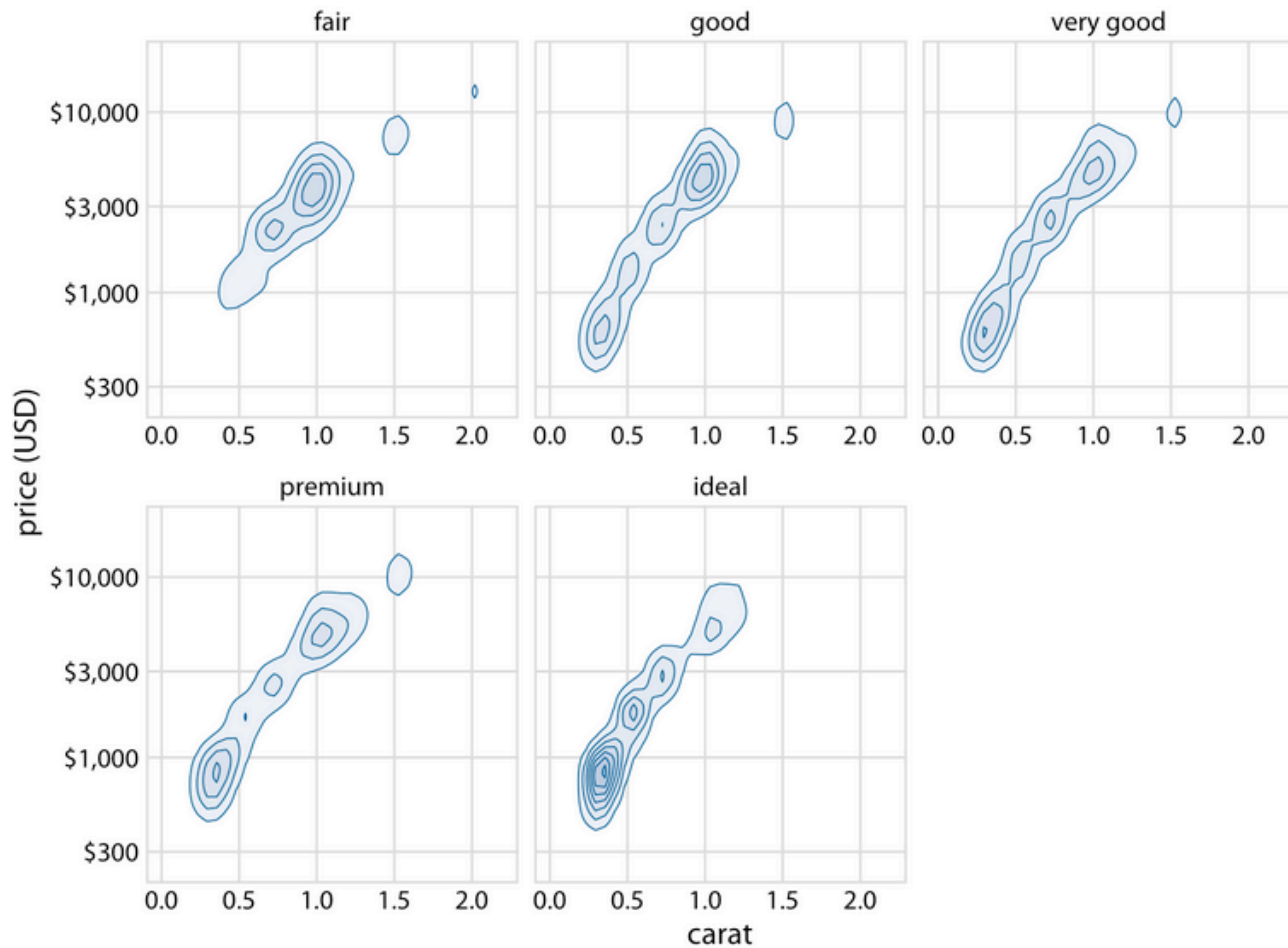
Muslims and Hindus are younger, while Jews and Buddhists are older, relative to average



A MATTER OF PERCEPTION

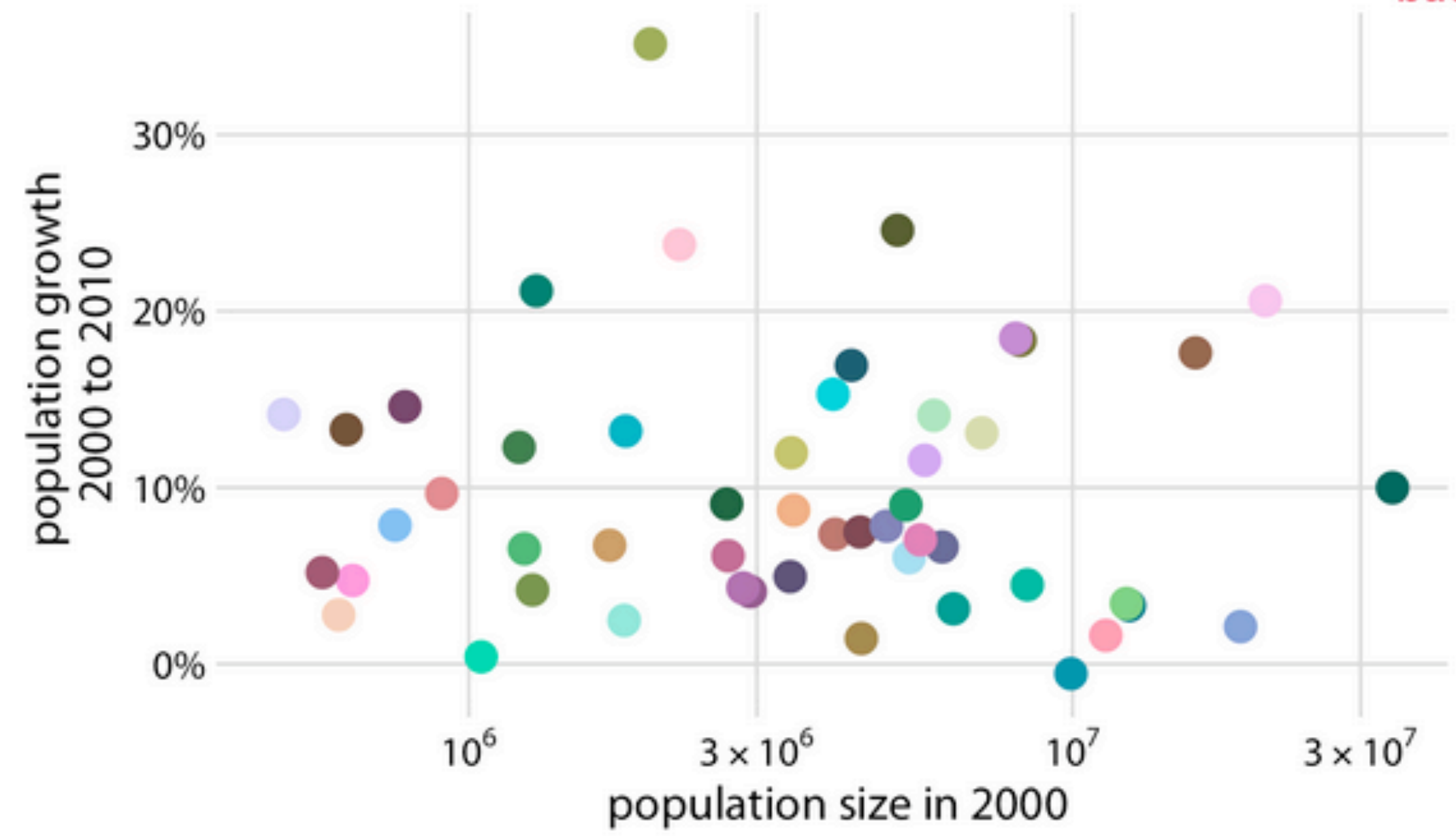




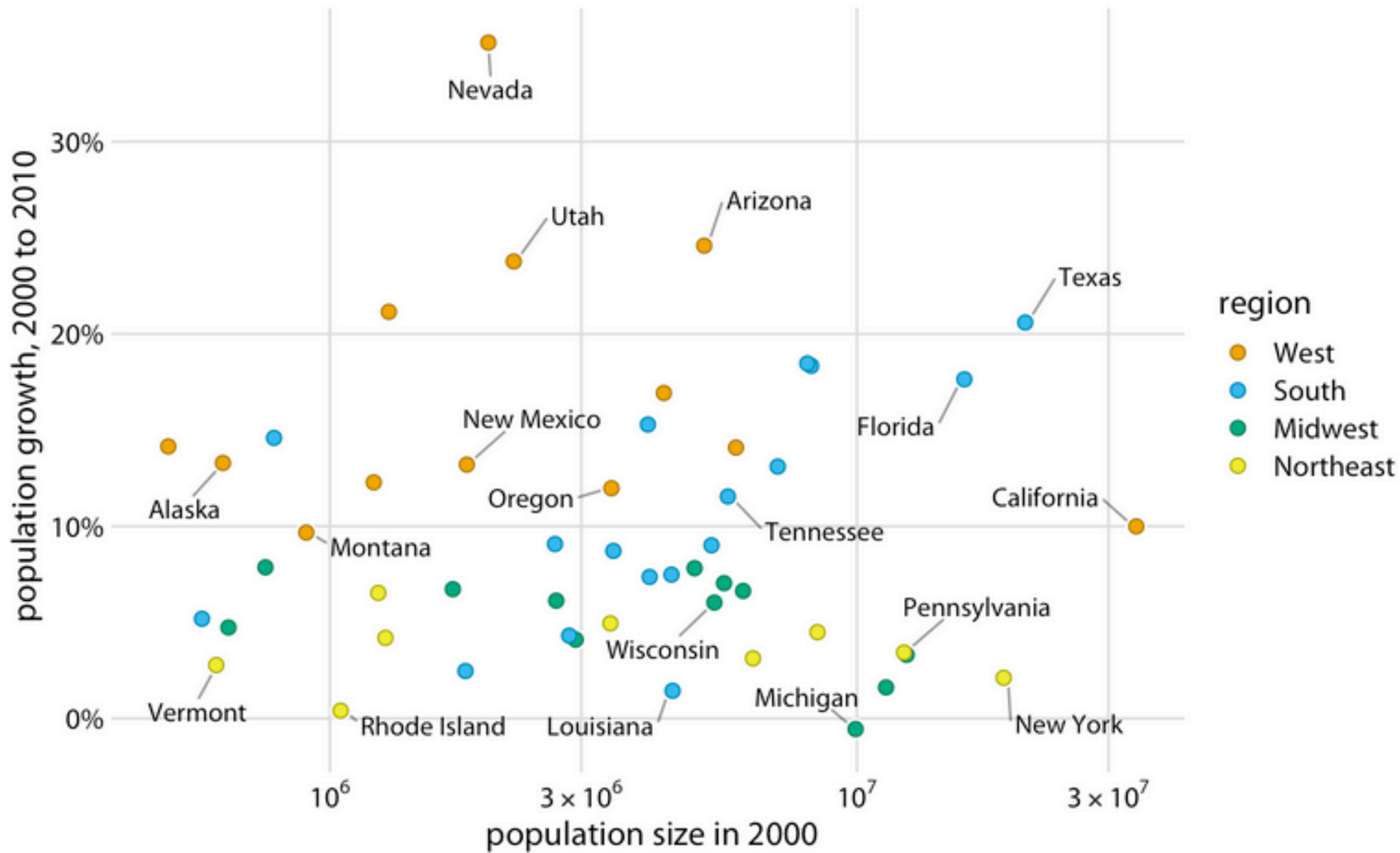


COLORS

bad

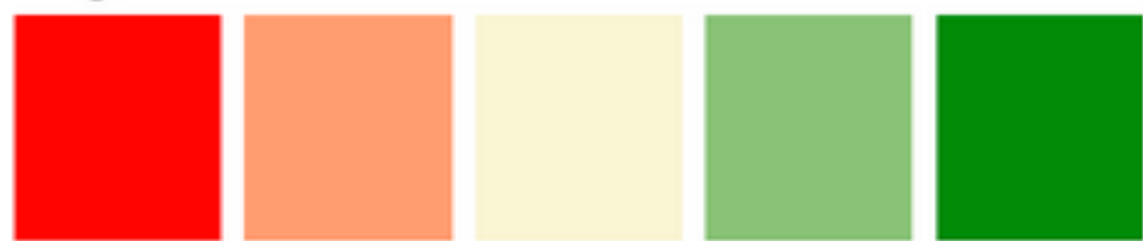


- state
- | | | |
|------------------------|------------------|------------------|
| ● Alabama | ● Kentucky | ● North Dakota |
| ● Alaska | ● Louisiana | ● Ohio |
| ● Arizona | ● Maine | ● Oklahoma |
| ● Arkansas | ● Maryland | ● Oregon |
| ● California | ● Massachusetts | ● Pennsylvania |
| ● Colorado | ● Michigan | ● Rhode Island |
| ● Connecticut | ● Minnesota | ● South Carolina |
| ● Delaware | ● Mississippi | ● South Dakota |
| ● District of Columbia | ● Missouri | ● Tennessee |
| ● Florida | ● Montana | ● Texas |
| ● Georgia | ● Nebraska | ● Utah |
| ● Hawaii | ● Nevada | ● Vermont |
| ● Idaho | ● New Hampshire | ● Virginia |
| ● Illinois | ● New Jersey | ● Washington |
| ● Indiana | ● New Mexico | ● West Virginia |
| ● Iowa | ● New York | ● Wisconsin |
| ● Kansas | ● North Carolina | ● Wyoming |



COLOR-VISION DEFICIENCY

original



deuteranomaly



protanomaly



tritanomaly



original



deuteranomaly



protanomaly



tritanomaly



PRINCIPLES

DATA-INK RATIO

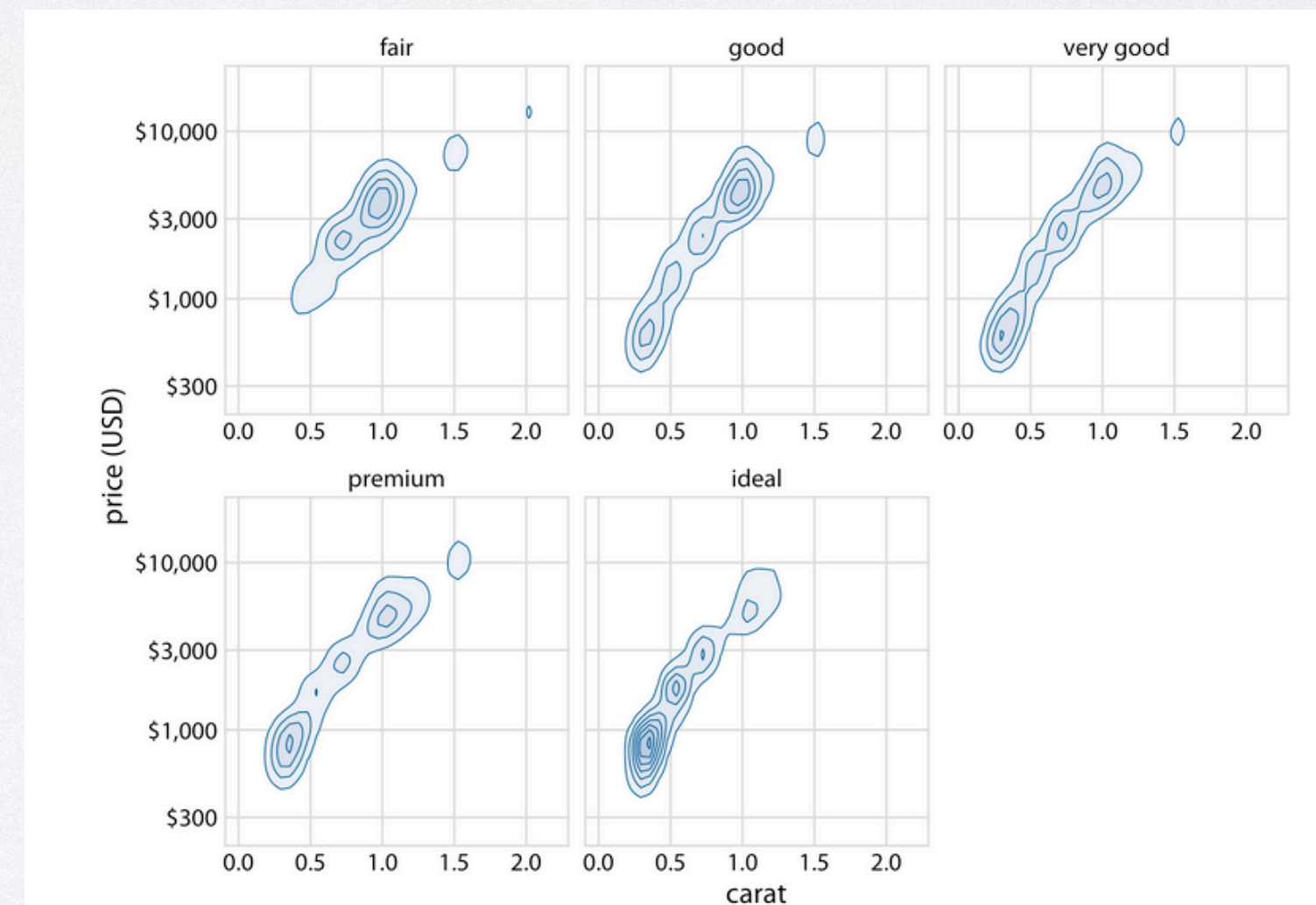
- Make sure most of the “ink” reflects the data
 - **within reason**

NO MENTAL GYMNASTICS

- Self-sufficiency: The graphic should be self-evident without captions
- Context should be clear (labels, axes, etc)

NO MENTAL GYMNASTICS

- Several small graphs can be better than one complicated graph
 - small multiples
 - trellis graphs



NO MENTAL GYMNASTICS

- Is a graph really appropriate
 - can a table be clearer
 - are you doing graphs for graphs' sake

PRESENTING INFORMATION

- Visualizations need to
 - be clear & uncluttered
 - make the point
 - provide easy access to the underlying data “talking points” and metadata
 - allow the consumer to explore a bit and think
 - be geared towards what the customer wants

CONSUMER FOCUS

- What is important to the consumer?
 - Often not what is important to the producer
- We may have a perfectly clear graphic ...
 - which is not what the consumer wants
 - and not what (s)he needs

CONSUMER FOCUS

Harvard
Business
Review

DESIGN

What Happens When Data Scientists and Designers Work Together

by Jon Wettersten and Dean Malmgren

MARCH 05, 2018

“Instead of a version of data science that is narrowly focused on researching new statistical models or building better data visualizations, a design-thinking approach recognizes data scientists as creative problem solvers.”