Computing in Context: Fall 2024

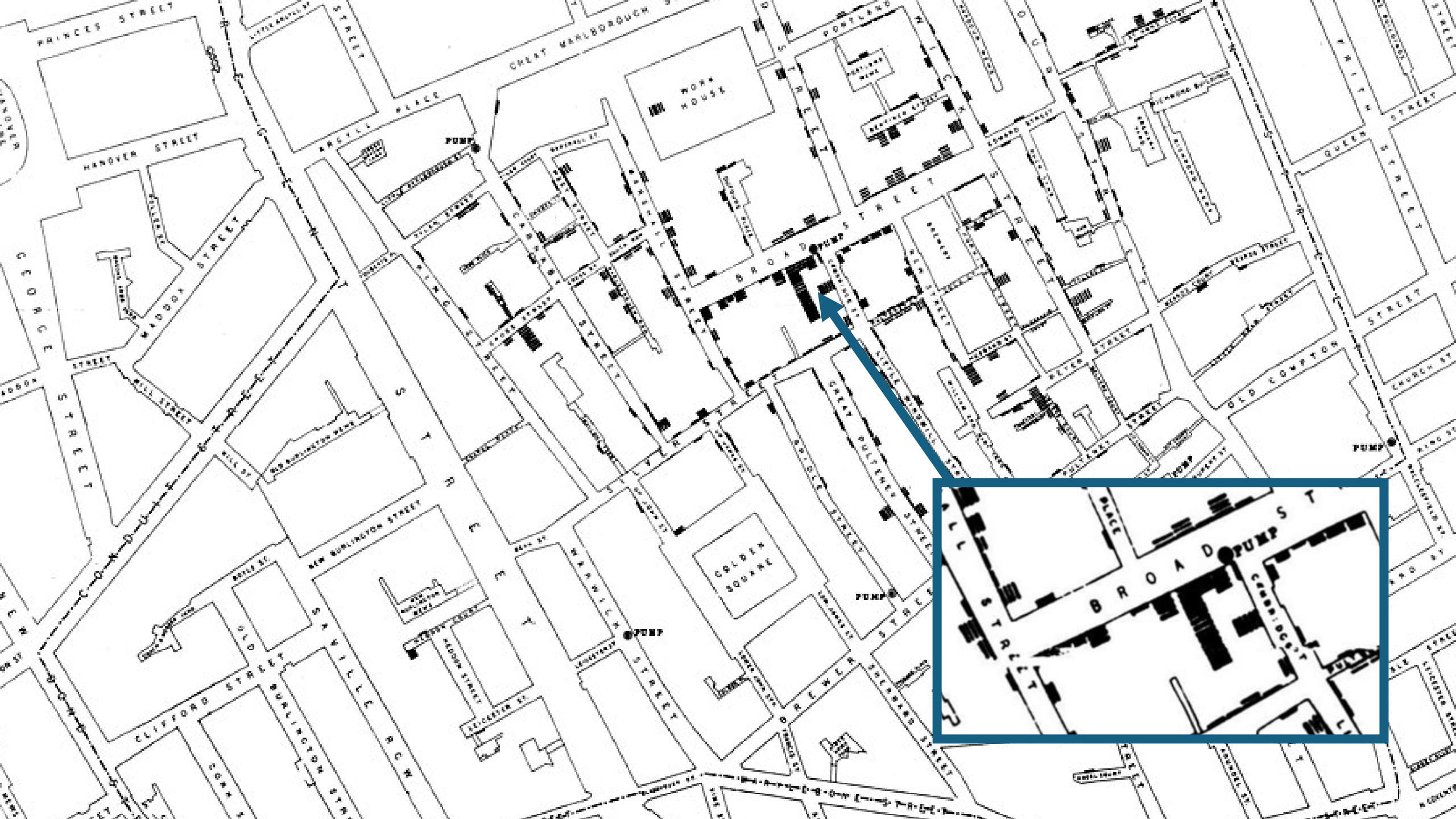
Lecture 5 | Intro to data visualization

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Dr Sarah Ashcroft-Jones | Mailman School of Public Health

Data visualization

Making numbers meaningful


PRINCES STREET
HANOVER STREET
GREAT MARLBOROUGH STREET
ARGYLL PLACE
WORK HOUSE
PUMP
BROAD STREET
PORTLAND PLACE
GEORGE STREET
MADDOX STREET
SAVILLE ROW
NEW BURLINGTON STREET
CLIFFORD STREET
OLD BURLINGTON STREET
GOLDEN SQUARE
WARWICK STREET
GREAT PULTENEY STREET
OLD COMPTON
QUEEN STREET
CHURCH ST
PLACE
CAMBRIDGE

“The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey, 1977

What is data visualization?

The process of representing data graphically

It helps to make data understandable and actionable

Many examples including bar charts, line charts, and scatter plots.

It’s all about insight …

Answers to concrete questions about a dataset (confirmatory)

or

Facts about a given problem we were not aware of (exploratory).

Why is it important?

Simplifies complex data

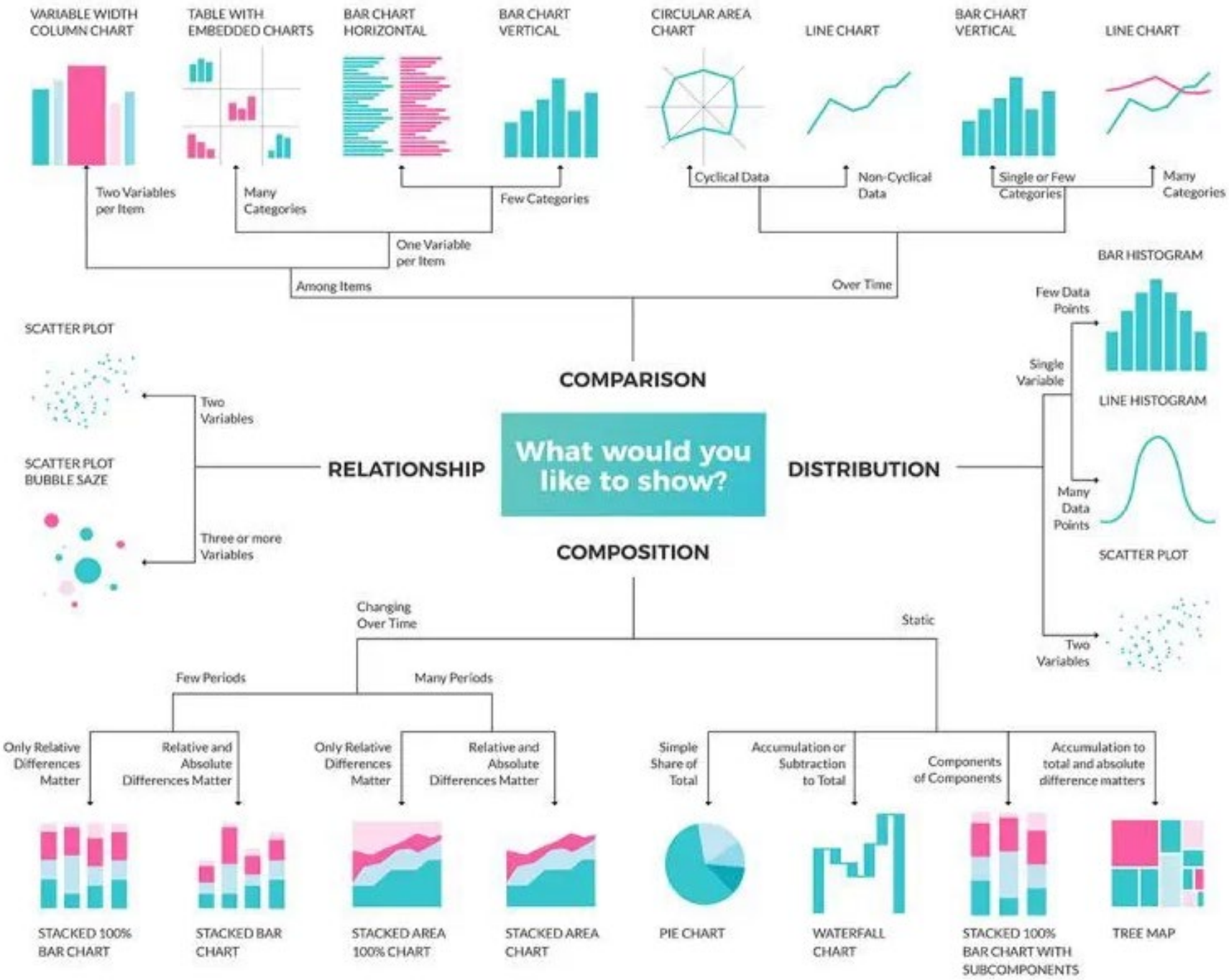
Enhances decision-making by revealing insights

Helps communicate findings effectively to non-technical audiences

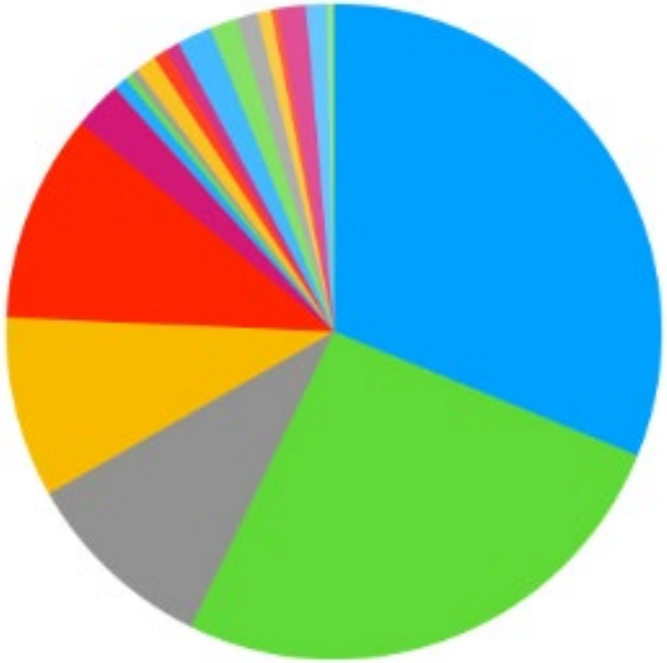
Choosing the right visualization

What’s the purpose of your visualization?
Is it for comparison, trends, relationships, or distribution?

Match data to the chart type:
e.g. Time series → Line chart
e.g. Comparison → Bar chart


What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
Among Items
Over Time
Two Variables per Item
Many Categories
One Variable per Item
Few Categories
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
SCATTER PLOT
Two Variables
SCATTER PLOT BUBBLE SAZE
Three or more Variables
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP

A personal note on pie charts



A note on pie charts - don’t

Best practices

Simplicity: Less is more—avoid information overload

Consistency: Use the same colors and styles across visualizations

Clarity: Include clear titles, axis labels, and legends

Color: Use meaningful colors (e.g., red for losses, green for gains)

Color – a note

Accessibility in data visualization

Check your color palettes – be mindful of colorblindness

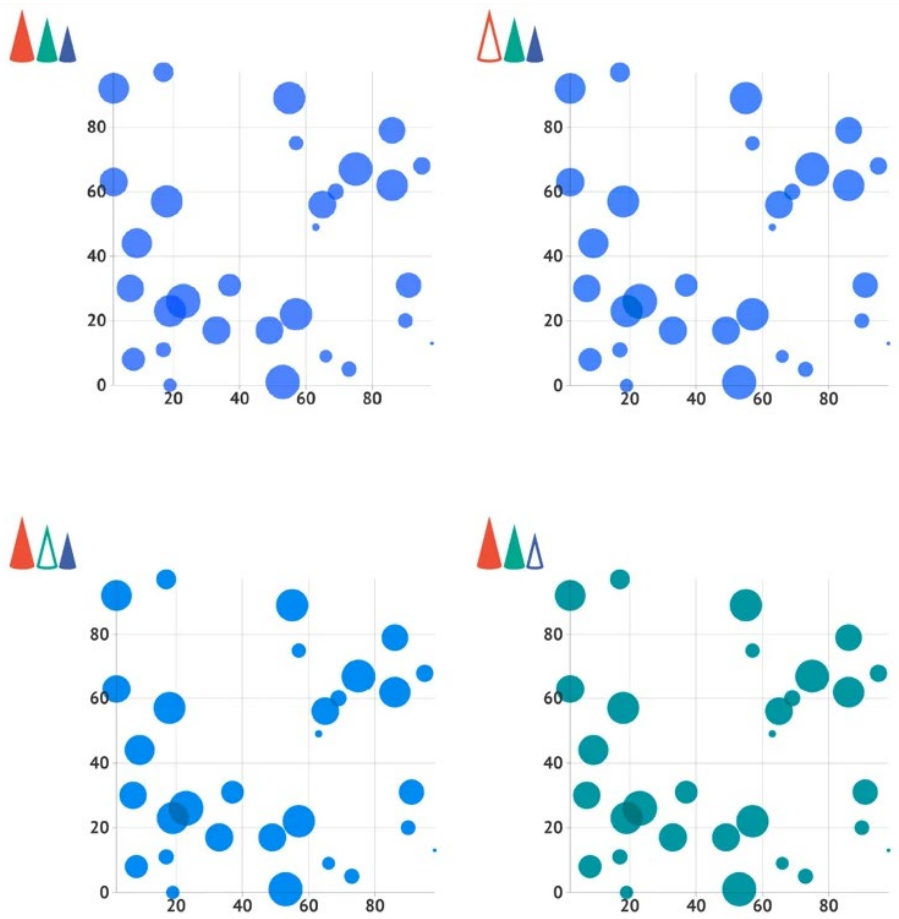
Use icons and symbols in addition to color

Use direct labels

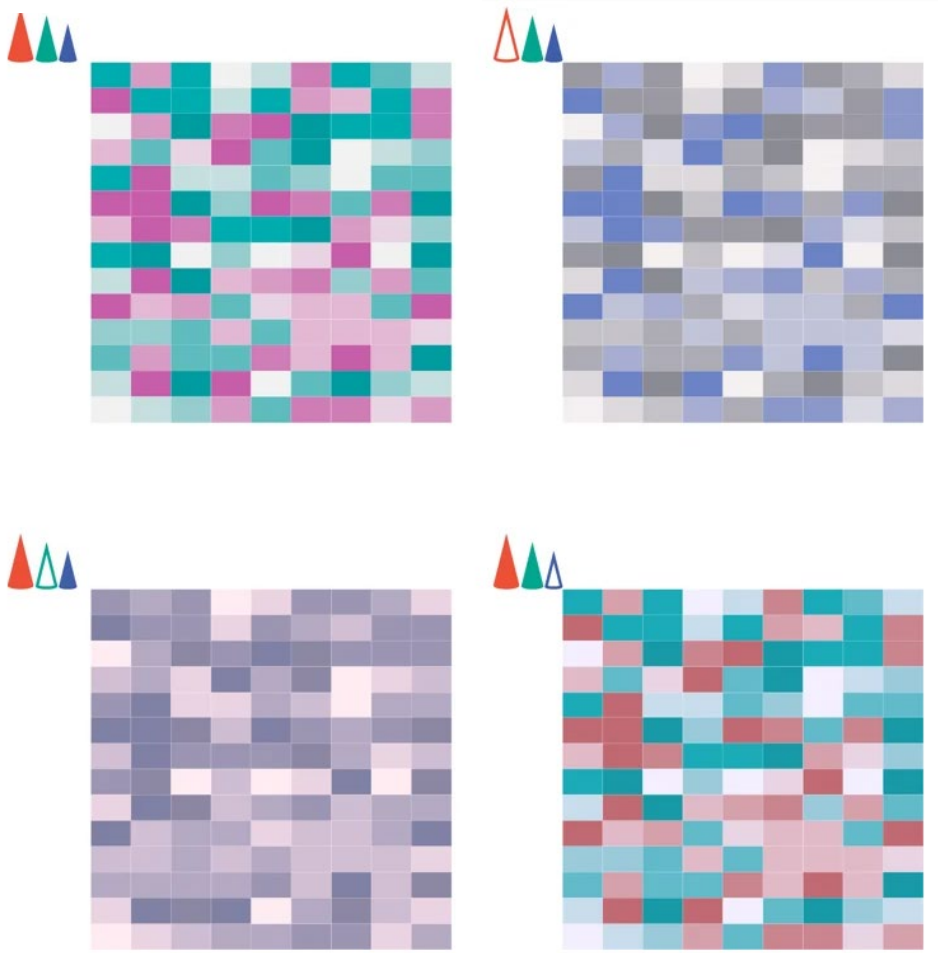
https://www.datylon.com/blog/data-visualization-for-colorblind-readers

Color – a note

✅ Good choice: Bubble chart



❌ Bad choice: Heatmap

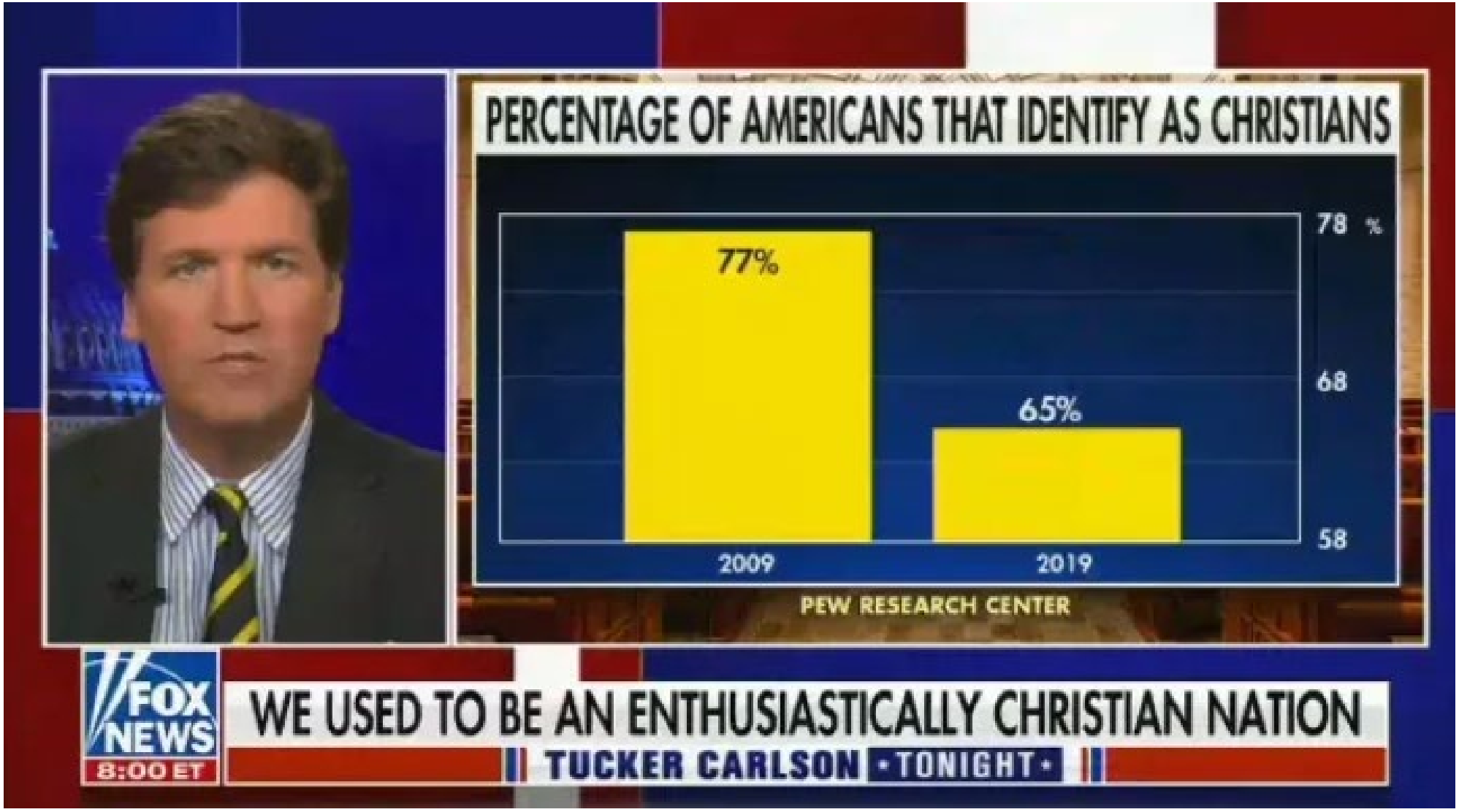


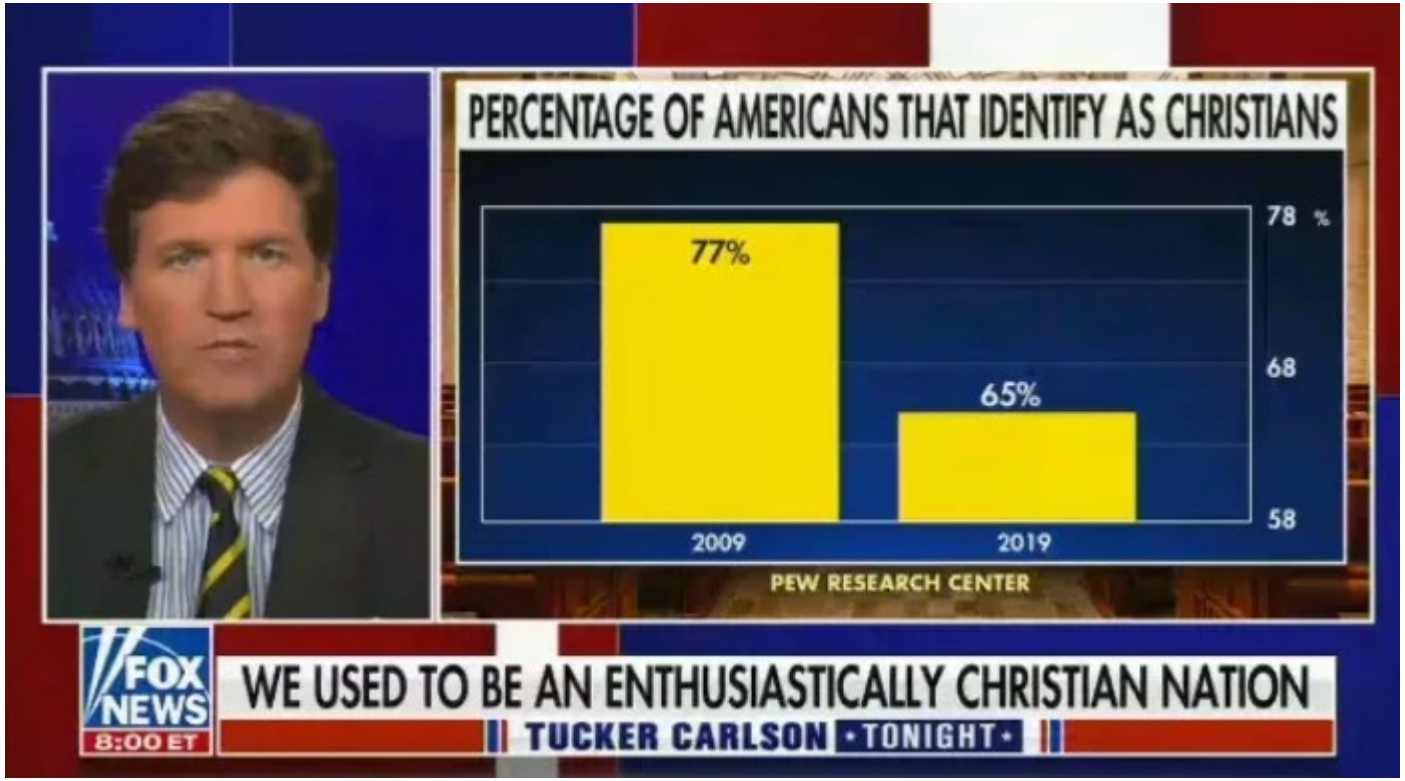
https://www.datylon.com/blog/data-visualization-for-colorblind-readers

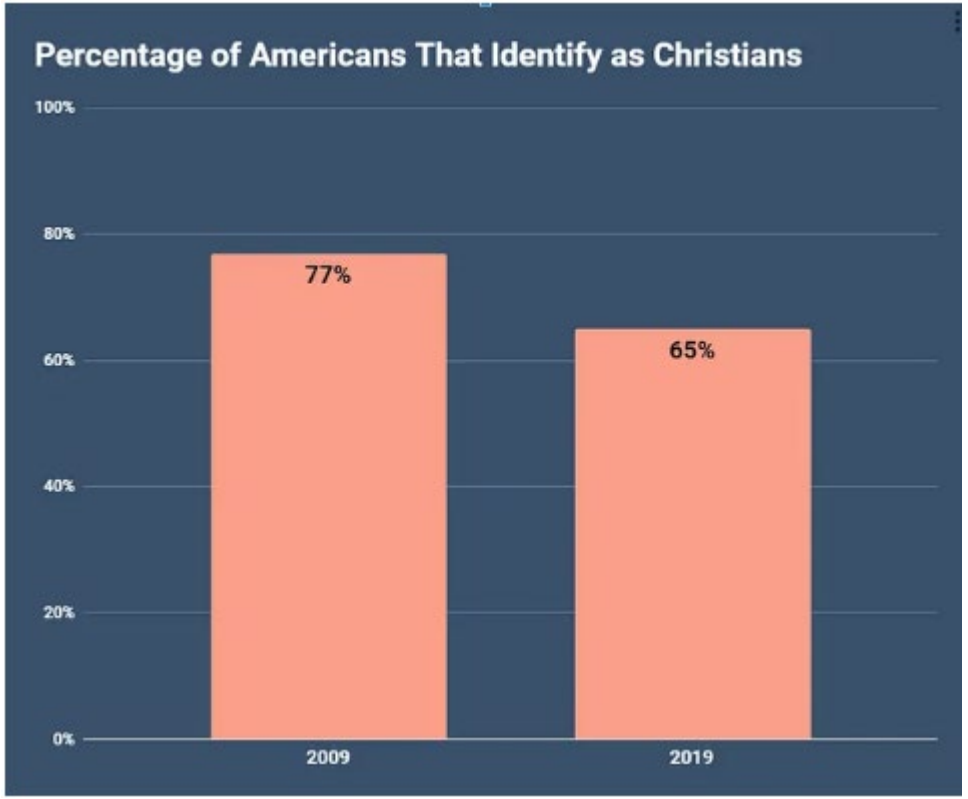
Responsible data visualization

The good, the bad, and the downright dangerous


U.S. Unemployment Rate
5%
10%
15%
2015
2016
2017
2018
2019
2020


PERCENTAGE OF AMERICANS THAT IDENTIFY AS CHRISTIANS
77%
65%
78 %
68
58
2009
2019
PEW RESEARCH CENTER
FOX NEWS
8:00 ET
WE USED TO BE AN ENTHUSIASTICALLY CHRISTIAN NATION
TUCKER CARLSON • TONIGHT •


PERCENTAGE OF AMERICANS THAT IDENTIFY AS CHRISTIANS
77%
65%
78 %
68
58
2009
2019
PEW RESEARCH CENTER
FOX NEWS
8:00 ET
WE USED TO BE AN ENTHUSIASTICALLY CHRISTIAN NATION
TUCKER CARLSON TONIGHT


Percentage of Americans That Identify as Christians
100%
80%
60%
40%
20%
0%
77%
65%
2009
2019

DAVIDsTEA
1d ·

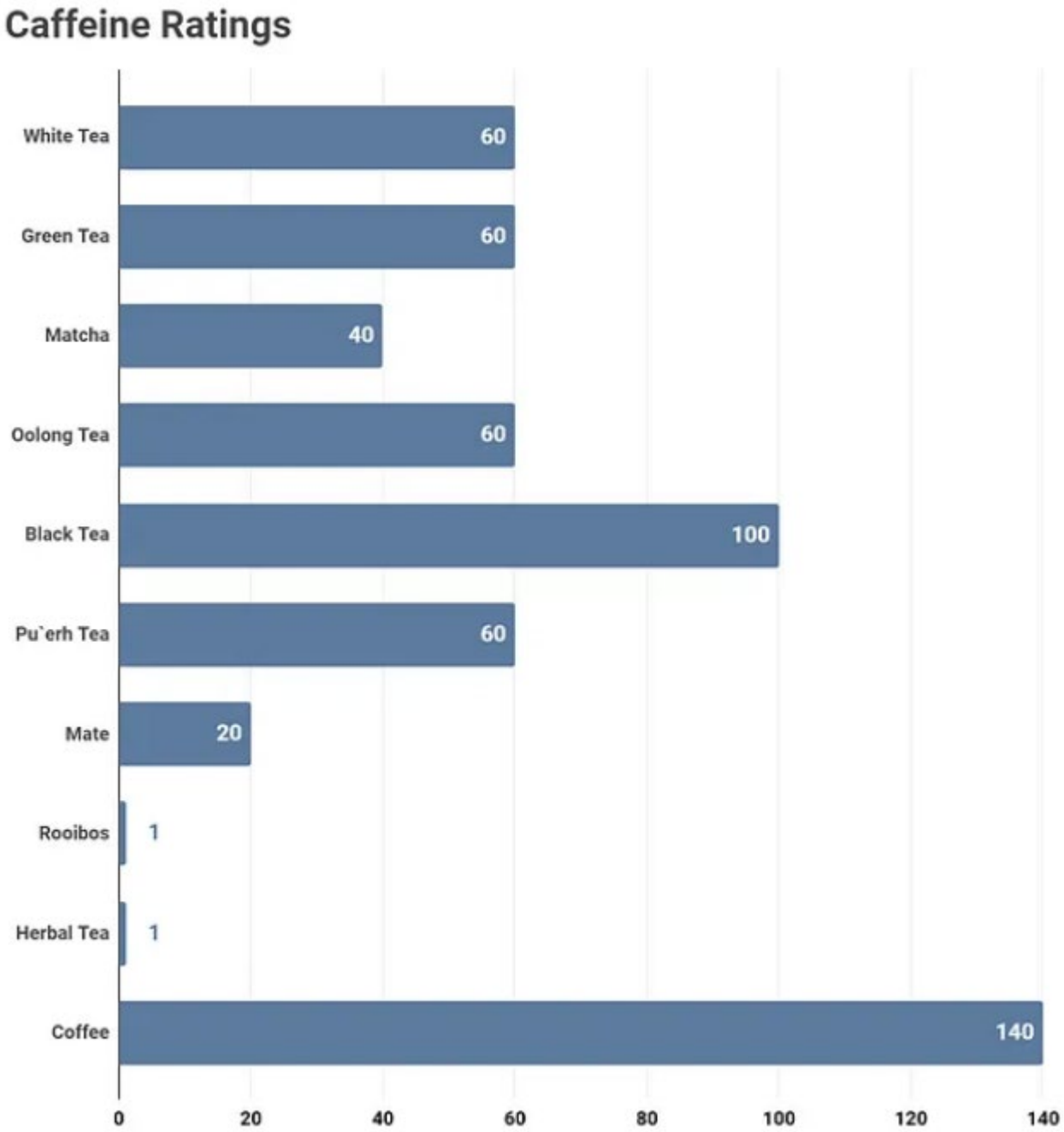
On a scale of 1 to Energizer Bunny, how much caffeine do you need today? Please refer to the chart above and plan accordingly 🤓
https://bit.ly/DTShopAll

DAVIDsTEA

1d ·

On a scale of 1 to Energizer Bunny, how much caffeine do you need today? Please refer to the chart above and plan accordingly 🤓

https://bit.ly/DTShopAll



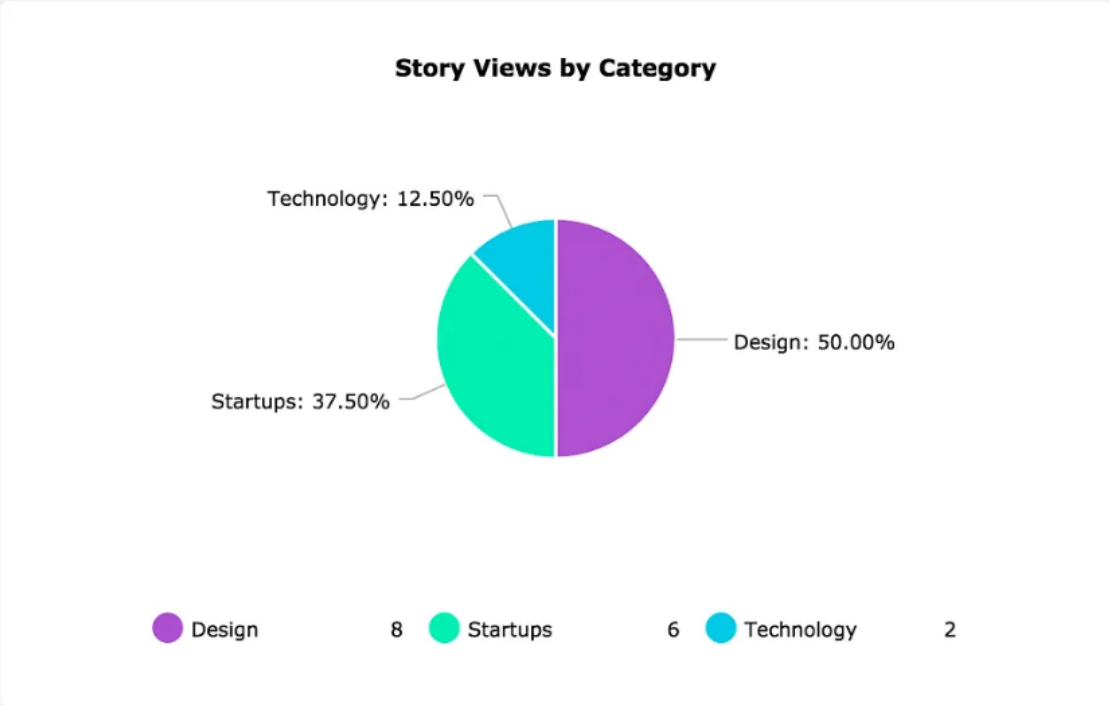
Things to avoid

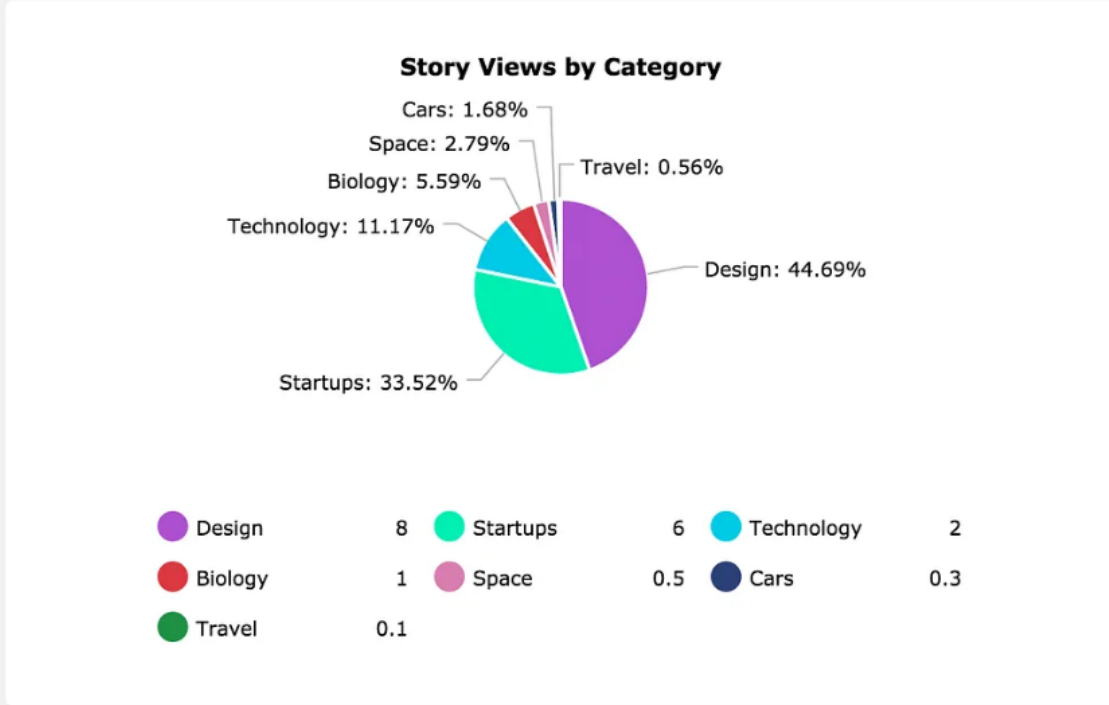
Misleading Axes: Starting axes at non-zero values can distort

Overloading Data: Too much information makes it unreadable

Inconsistent Design: Switching colors or styles confuses viewers

This not that...

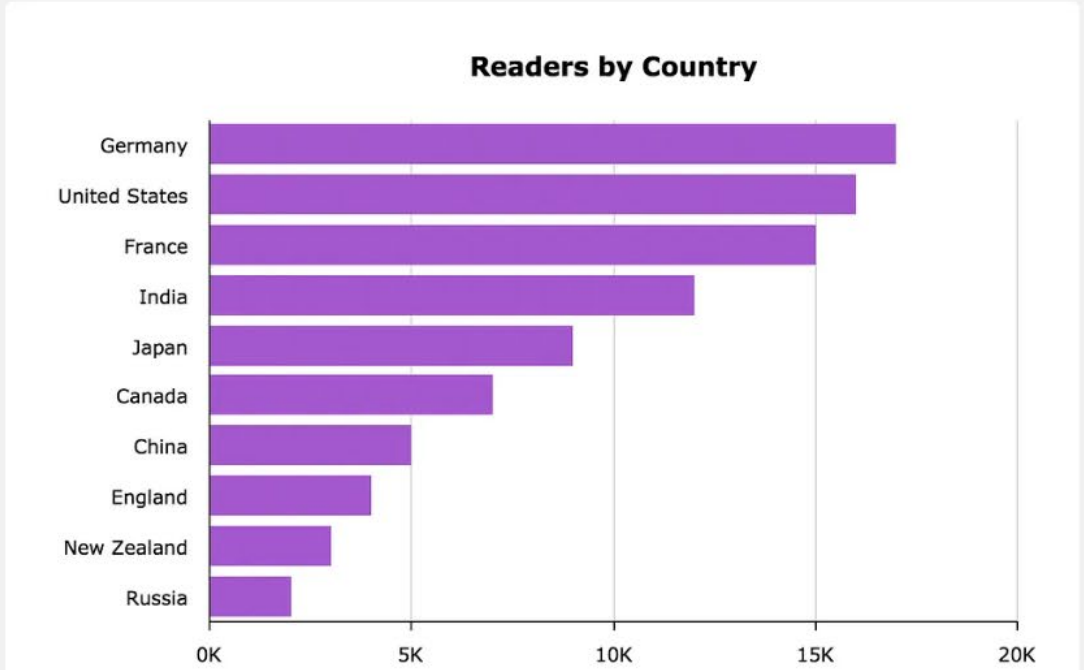


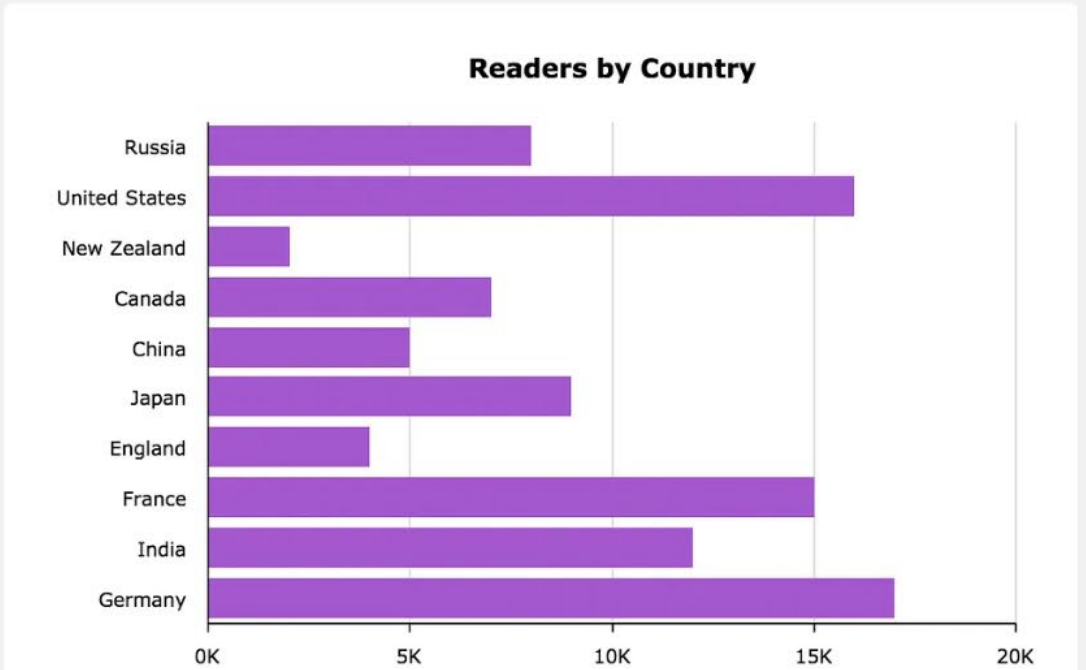


Or just don’t ...

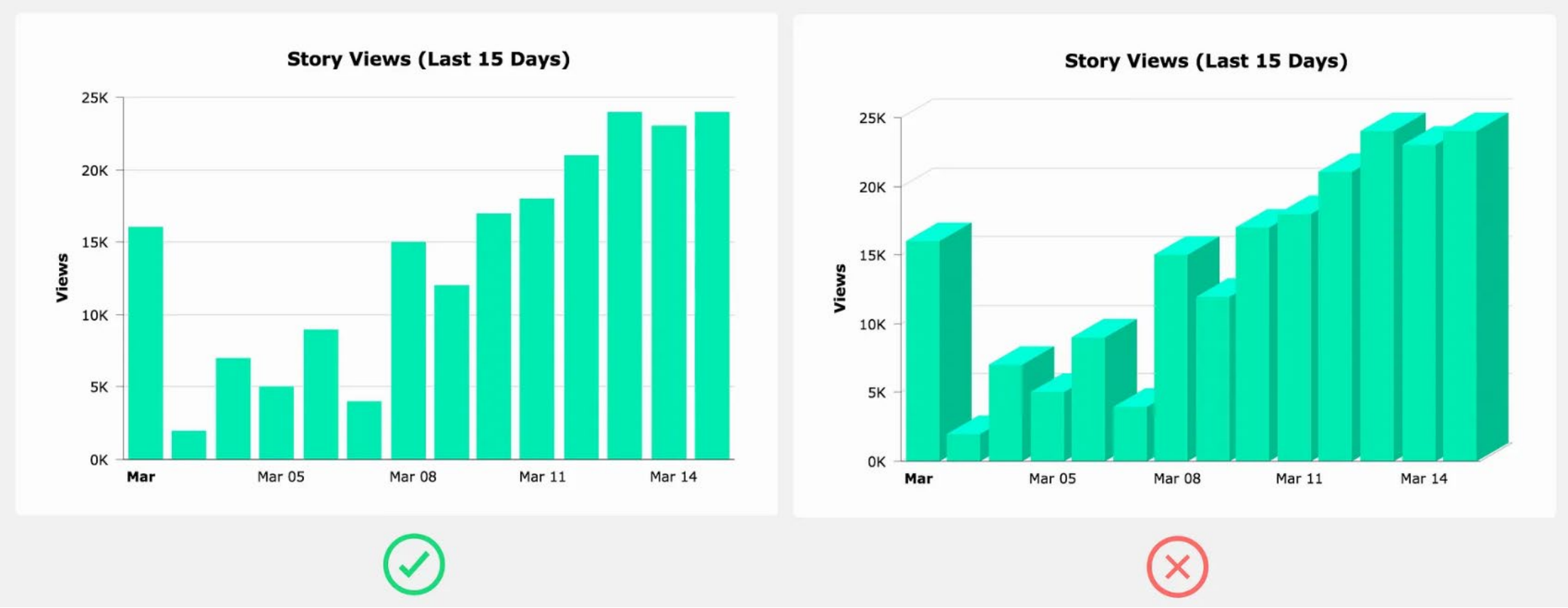

Technology: 12.50%
Startups: 37.50%
Design
8
Travel: 0.56%
Design: 44.69%
6
Technology
2
0.5
Cars
0.3

This not that...

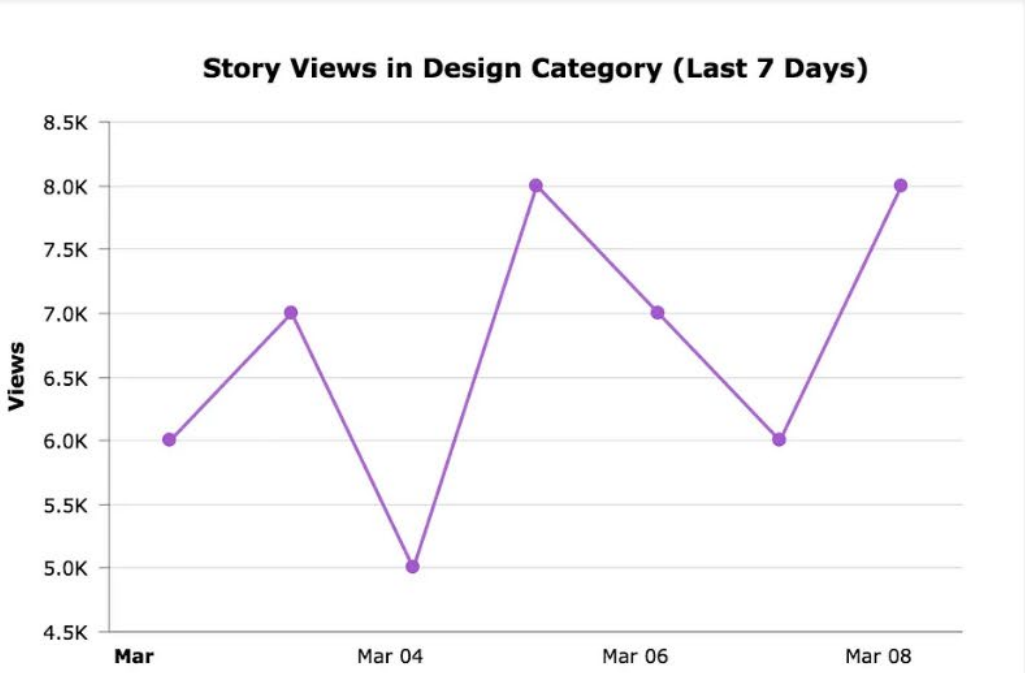


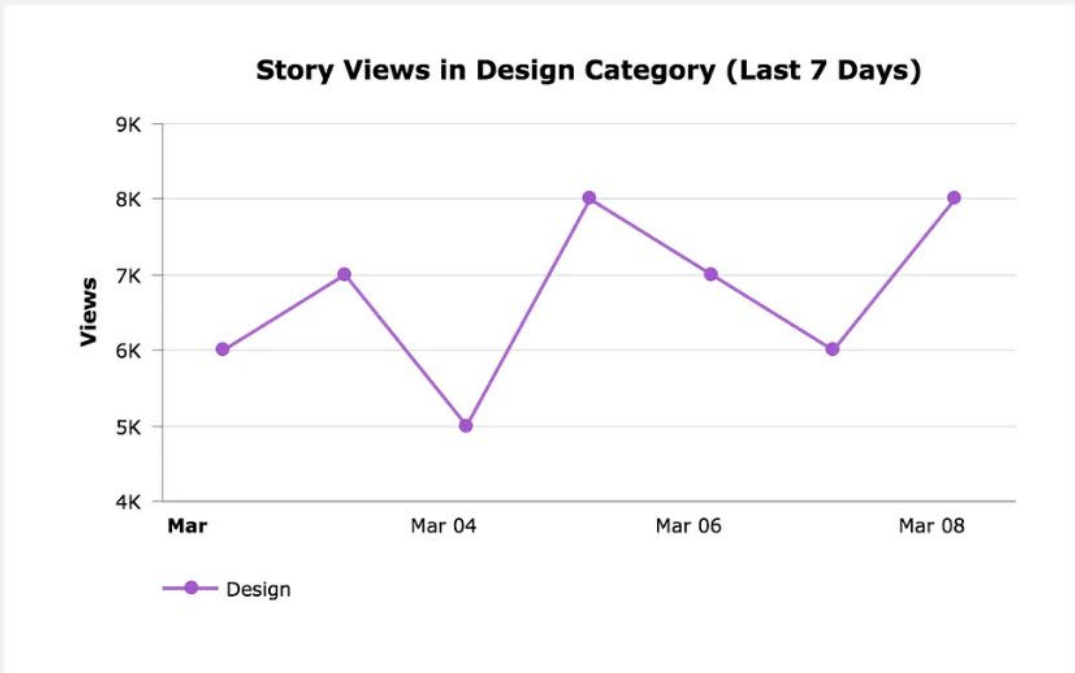


This not that...



This not that...

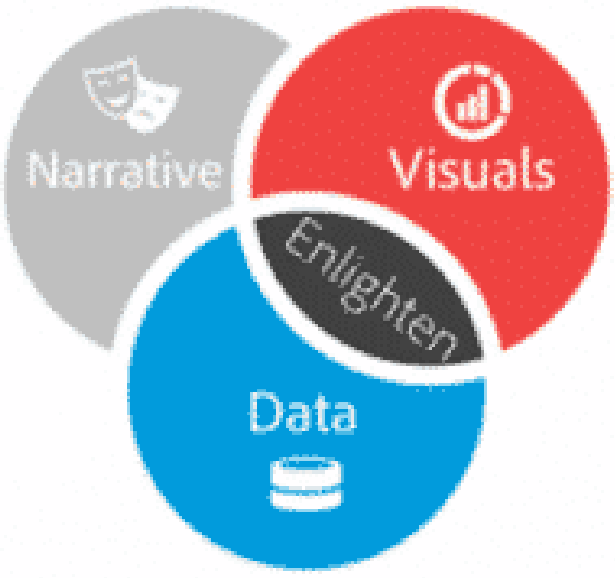




Data storytelling

Data visualization at it’s best

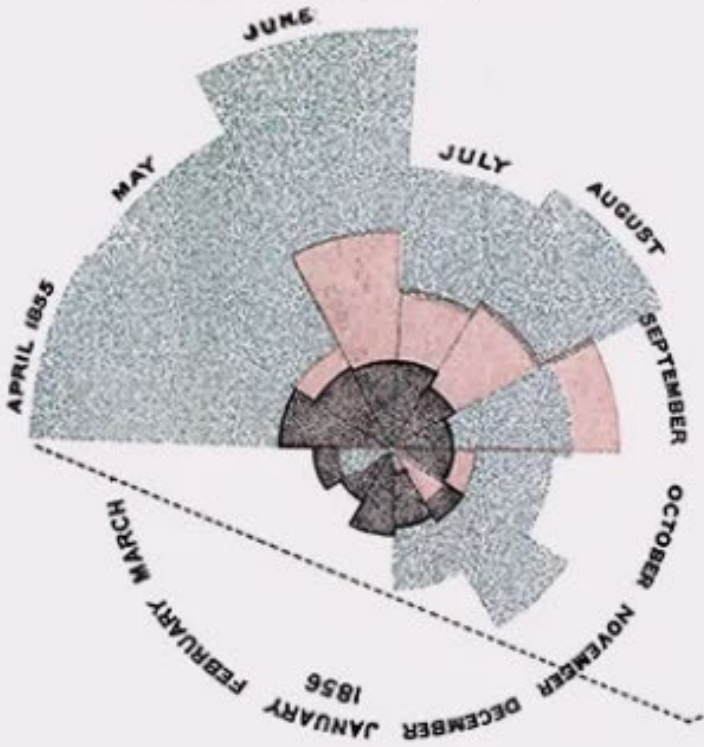

Narrative
Visuals
Explain
Data


Narrative
Visuals
Enlighten
Data

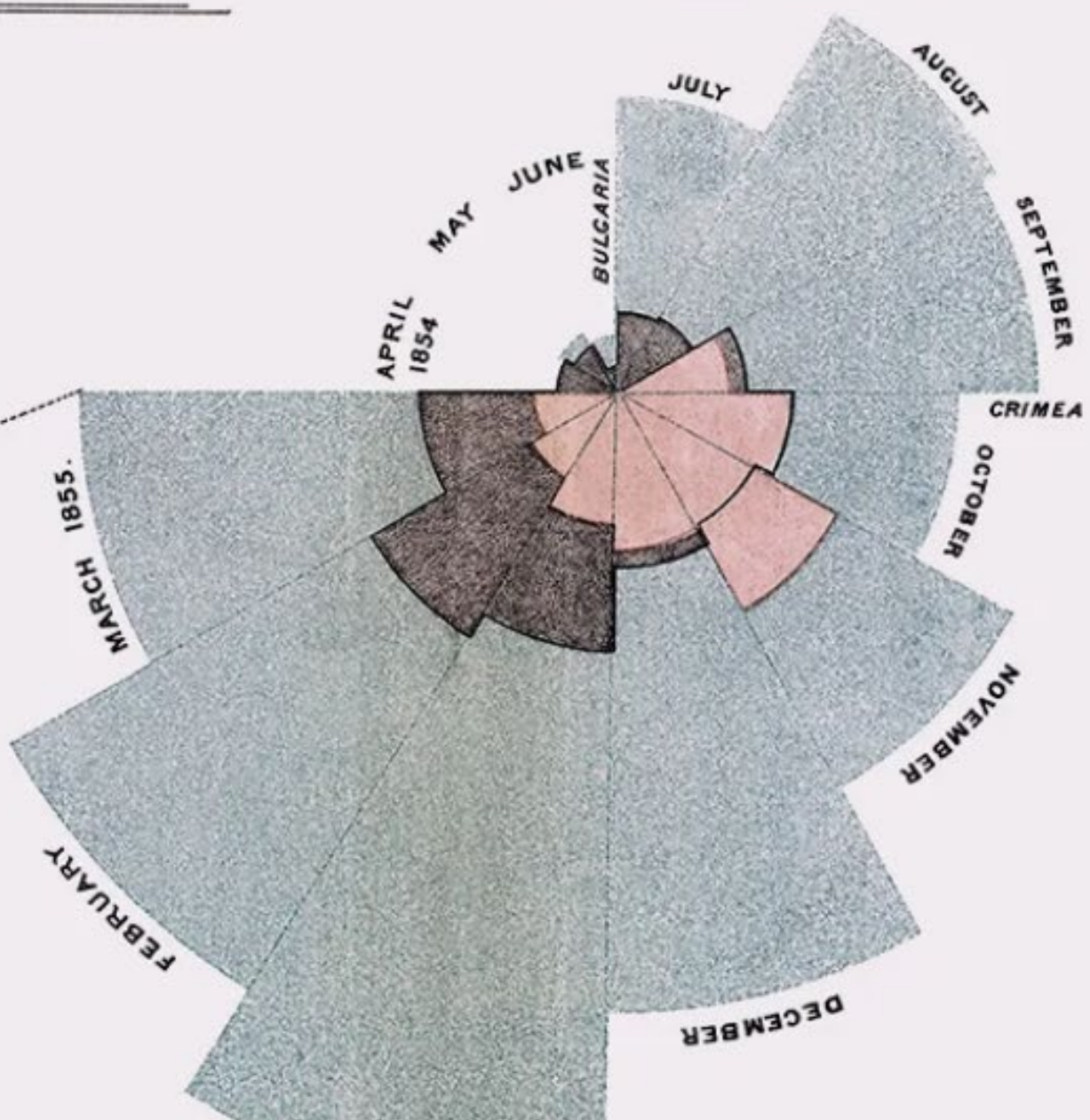

Narrative
Engage
Visuals
Data

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

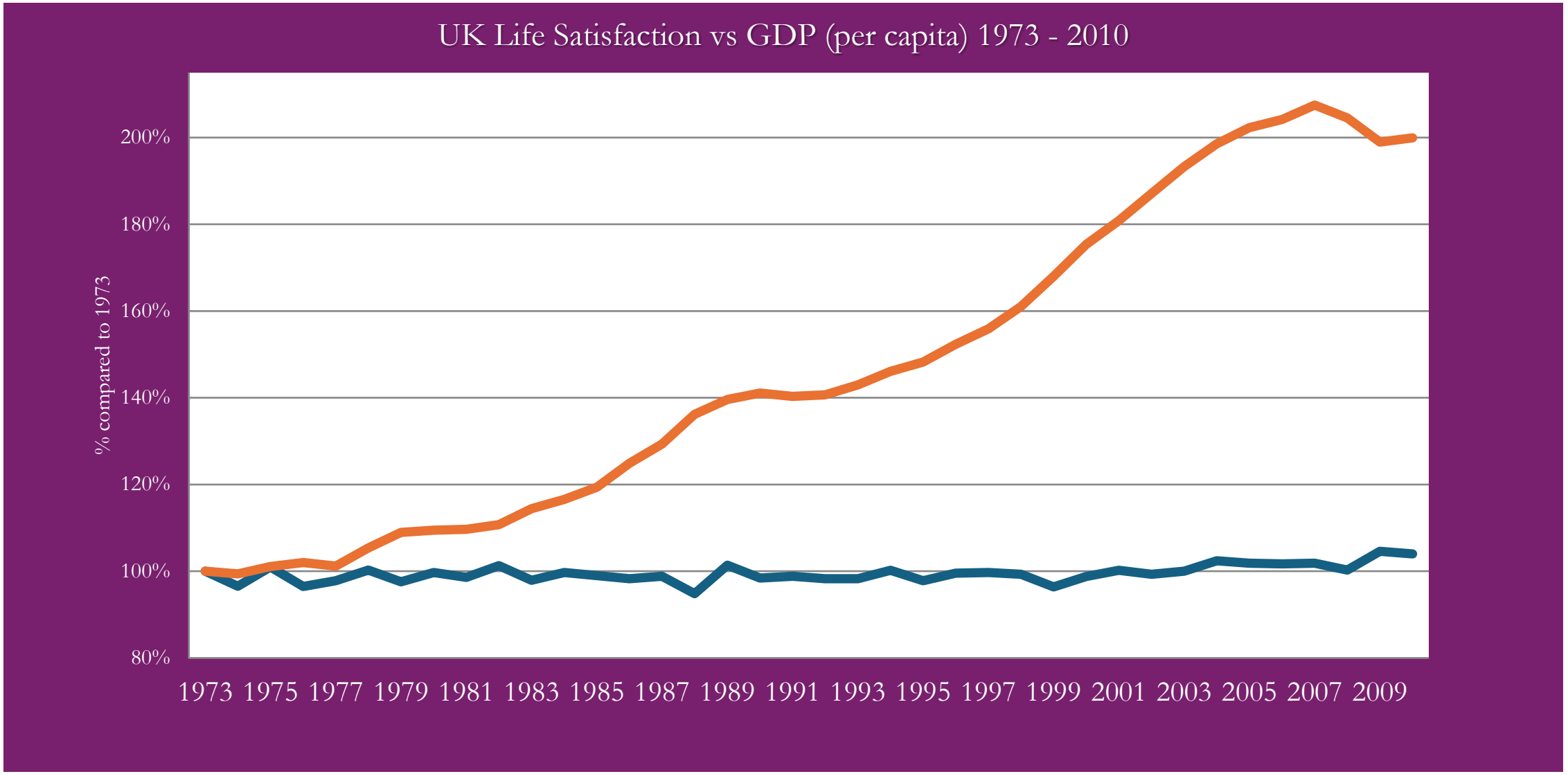
2.
APRIL 1855 TO MARCH 1856.

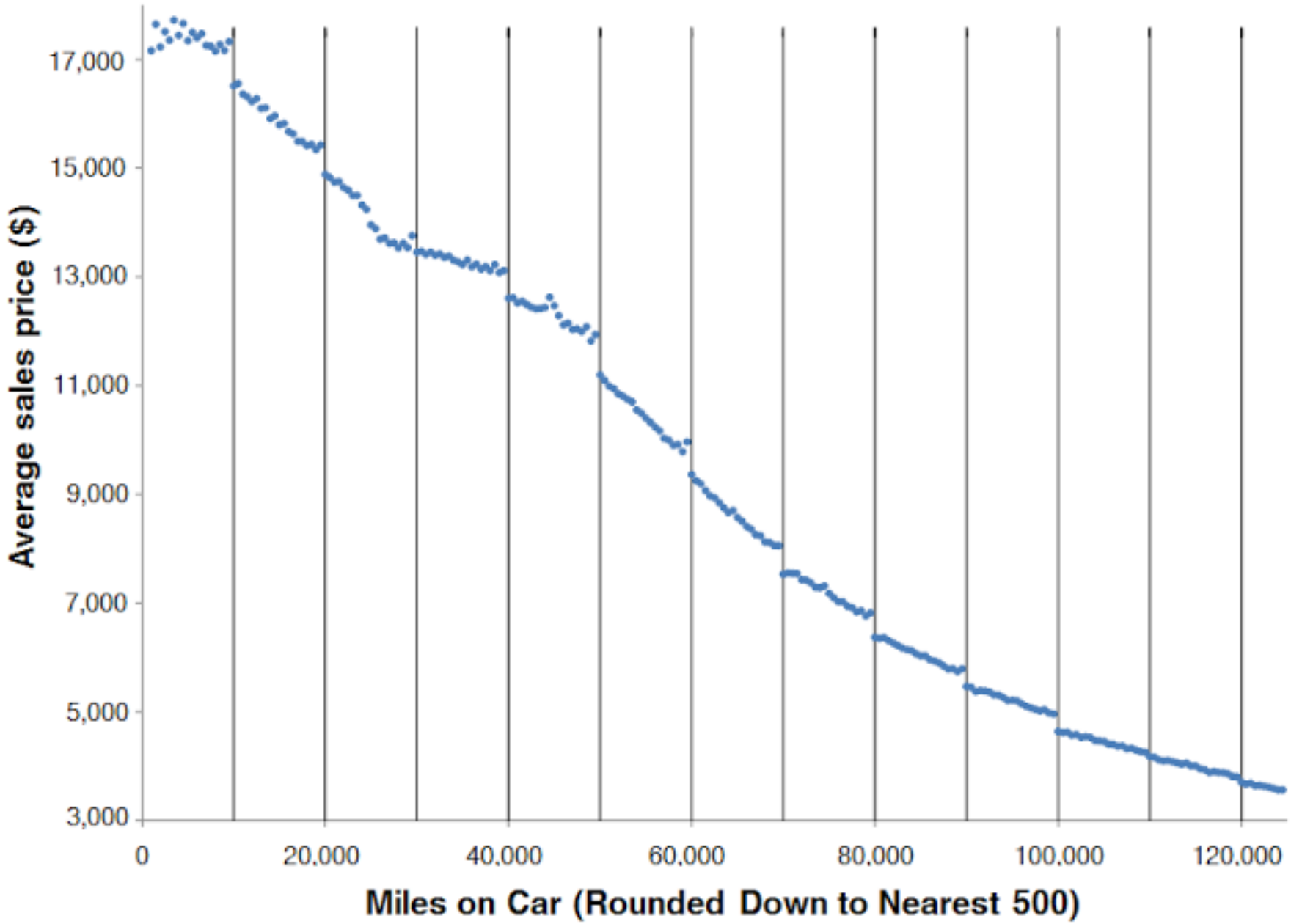


1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.


UK Life Satisfaction vs GDP (per capita) 1973 - 2010
% compared to 1973
200%
180%
160%
140%
120%
100%
80%
1973 1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009


Average sales price ($)
17,000
15,000
13,000
11,000
9,000
7,000
5,000
3,000
0
20,000
40,000
60,000
80,000
100,000
120,000
Miles on Car (Rounded Down to Nearest 500)

“bad data viz” game

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

“Do no harm” discussion

Questions?

Project memo due tonight
Any panic – ask now!