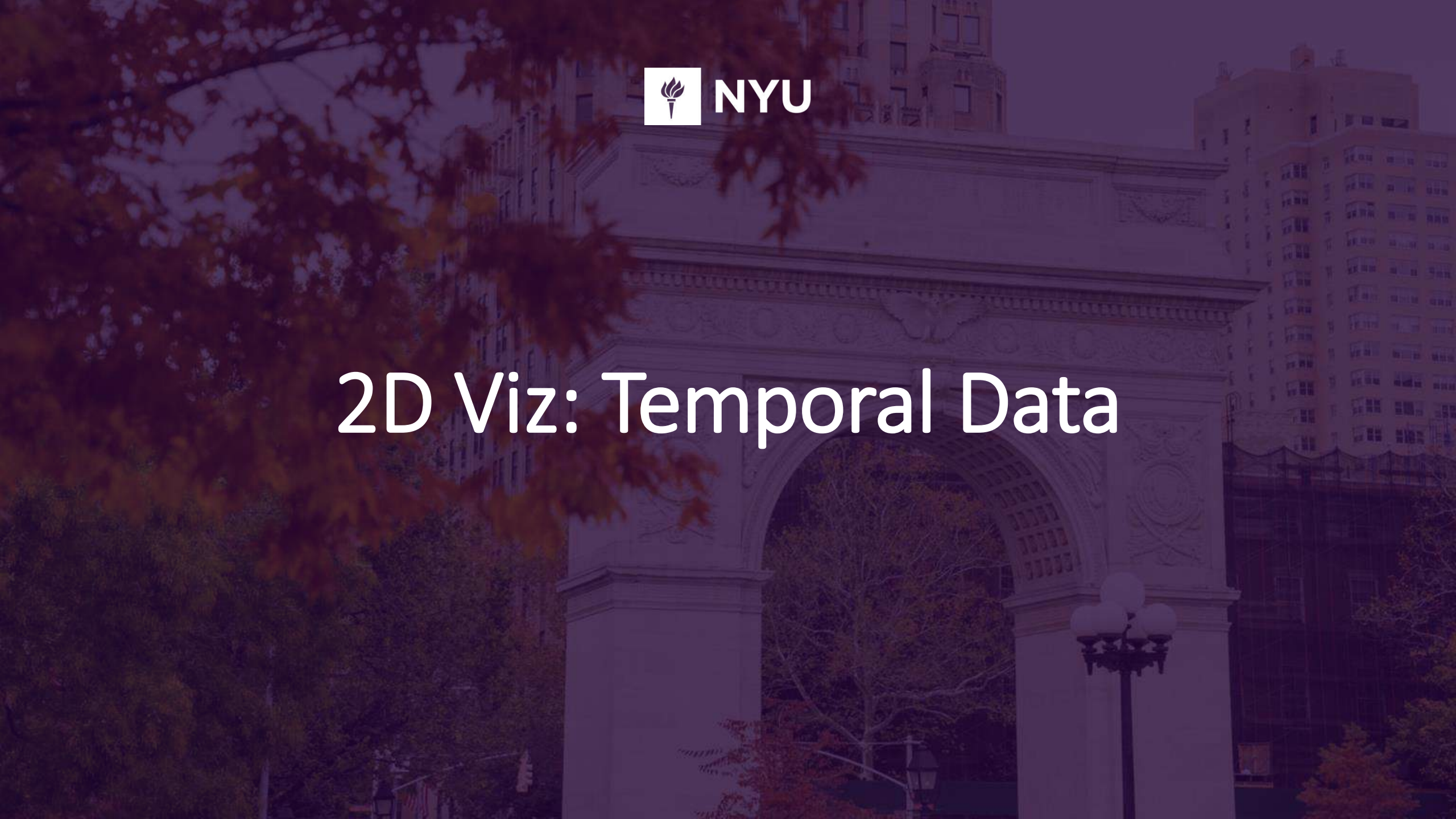
NYU

CS-GY 6313 B: Information Visualization

10/17/2024

Logistics

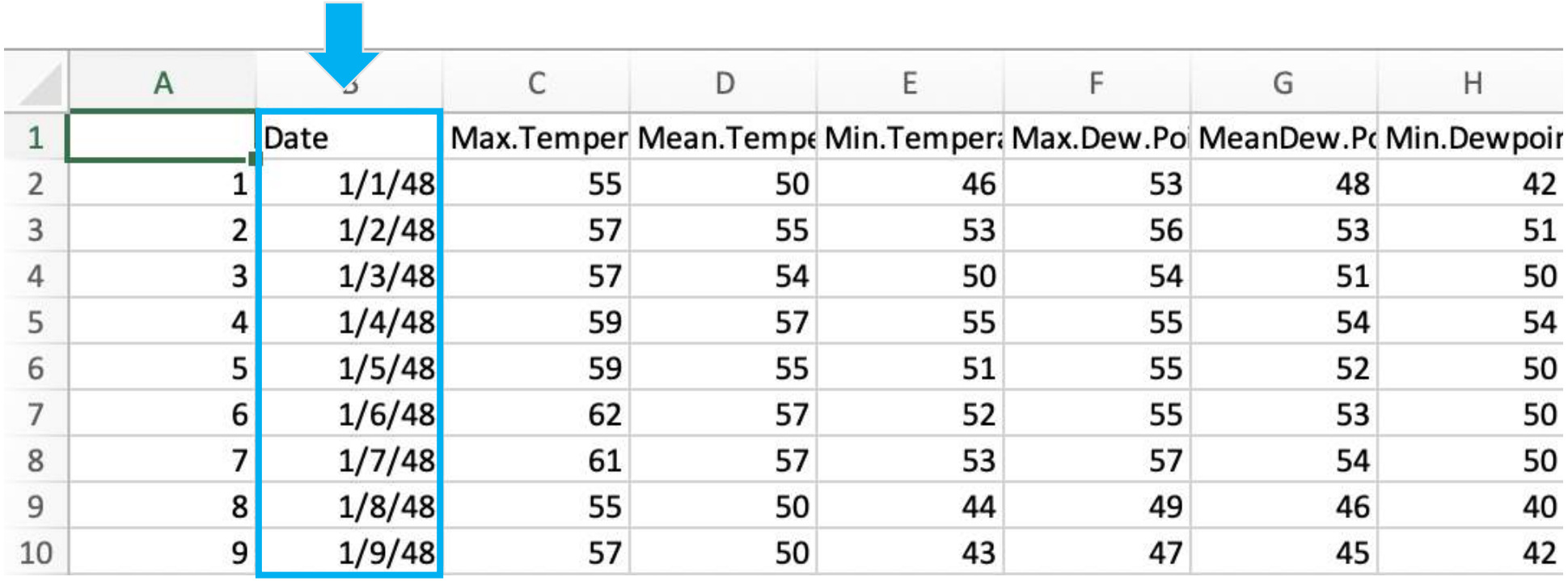
- Assignment 2 due last night
- Assignment 3 released tonight: Interactions
  - Need to find a decently-complex dataset!
- CC Lokesh (TA) for all inquiries about late credits, grading, etc.
- Today:
  - Assignment 2 showcase
  - Temporal data lecture
  - “How to ask interesting questions” tips
  - “Looking for a partner” pitches
  - Project work/question session



2D Viz: Temporal Data

What is temporal data?

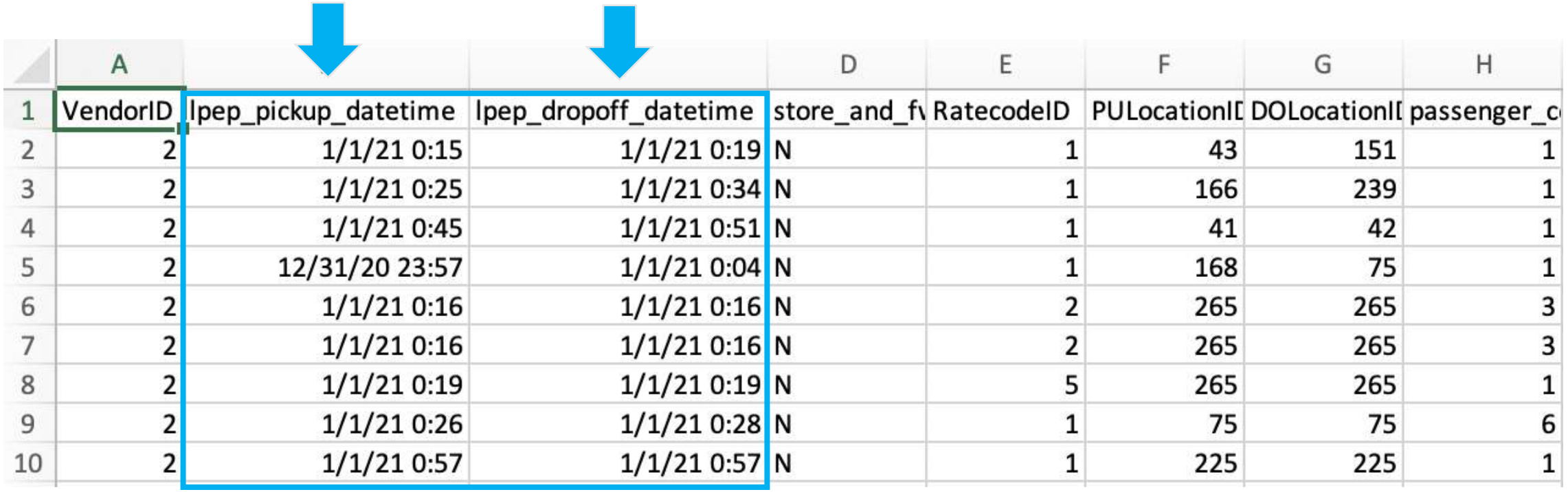
- Data in which the **values depend on time** and **time is explicitly recorded**



	A	B	C	D	E	F	G	H
1		Date	Max.Temper	Mean.Tempe	Min.Tempera	Max.Dew.Po	MeanDew.Pc	Min.Dewpoir
2	1	1/1/48	55	50	46	53	48	42
3	2	1/2/48	57	55	53	56	53	51
4	3	1/3/48	57	54	50	54	51	50
5	4	1/4/48	59	57	55	55	54	54
6	5	1/5/48	59	55	51	55	52	50
7	6	1/6/48	62	57	52	55	53	50
8	7	1/7/48	61	57	53	57	54	50
9	8	1/8/48	55	50	44	49	46	40
10	9	1/9/48	57	50	43	47	45	42

What is temporal data?

- Data in which the **values depend on time** and **time is explicitly recorded**



	A			D	E	F	G	H
1	VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_f\	RatecodeID	PULocationI[	DOLocationI[	passenger_c
2	2	1/1/21 0:15	1/1/21 0:19	N	1	43	151	1
3	2	1/1/21 0:25	1/1/21 0:34	N	1	166	239	1
4	2	1/1/21 0:45	1/1/21 0:51	N	1	41	42	1
5	2	12/31/20 23:57	1/1/21 0:04	N	1	168	75	1
6	2	1/1/21 0:16	1/1/21 0:16	N	2	265	265	3
7	2	1/1/21 0:16	1/1/21 0:16	N	2	265	265	3
8	2	1/1/21 0:19	1/1/21 0:19	N	5	265	265	1
9	2	1/1/21 0:26	1/1/21 0:28	N	1	75	75	6
10	2	1/1/21 0:57	1/1/21 0:57	N	1	225	225	1

Two types of temporal data

- **Measurement data**
  - Value at time T (time + measurement)
  - Eg:
    - Temperature
    - Revenue
    - Stock value
- **Event data**
  - Something happened at time T (time + object)
  - Eg:
    - Tweet
    - Taxi pickup/drop off
    - alarm

Time Structure & Format: ISO 8061

Date

Time

Year Month Day Hour Min. Sec. Time Zone

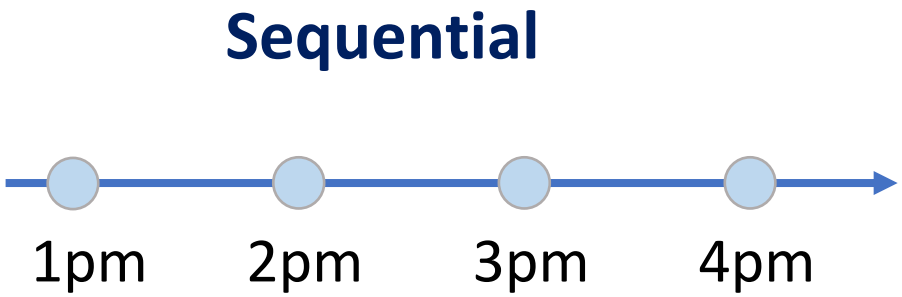
2022-02-28T11:59:59-0500

Other forms:
Feb. 28, 2022
02/28/2022
28/02/2022
2022.16 (decimal date)
Week 9
Monday (week day)
Day 59 (year day)

Other forms:
11:59:59 am
11:59:59.283
11:59:59.283+00:00

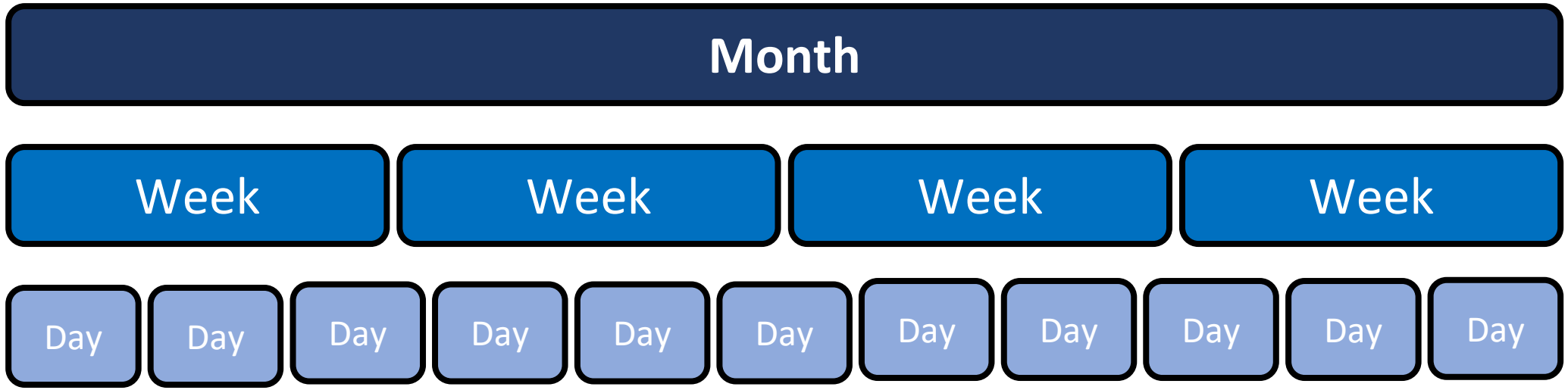
Packages:
R: lubridate
Python: datetime
JavaScript: Date class

Time Structure

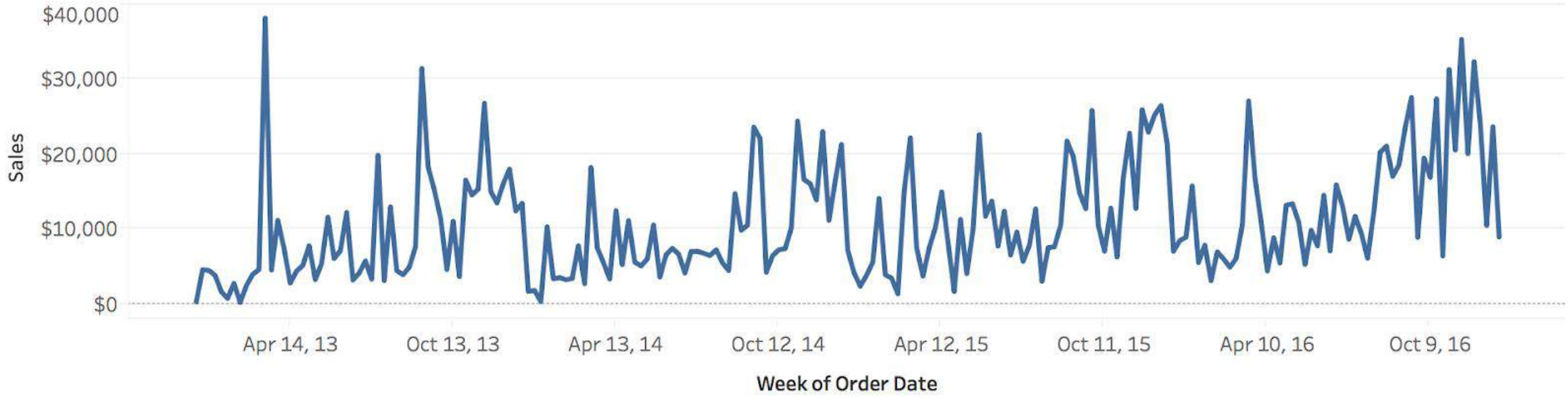


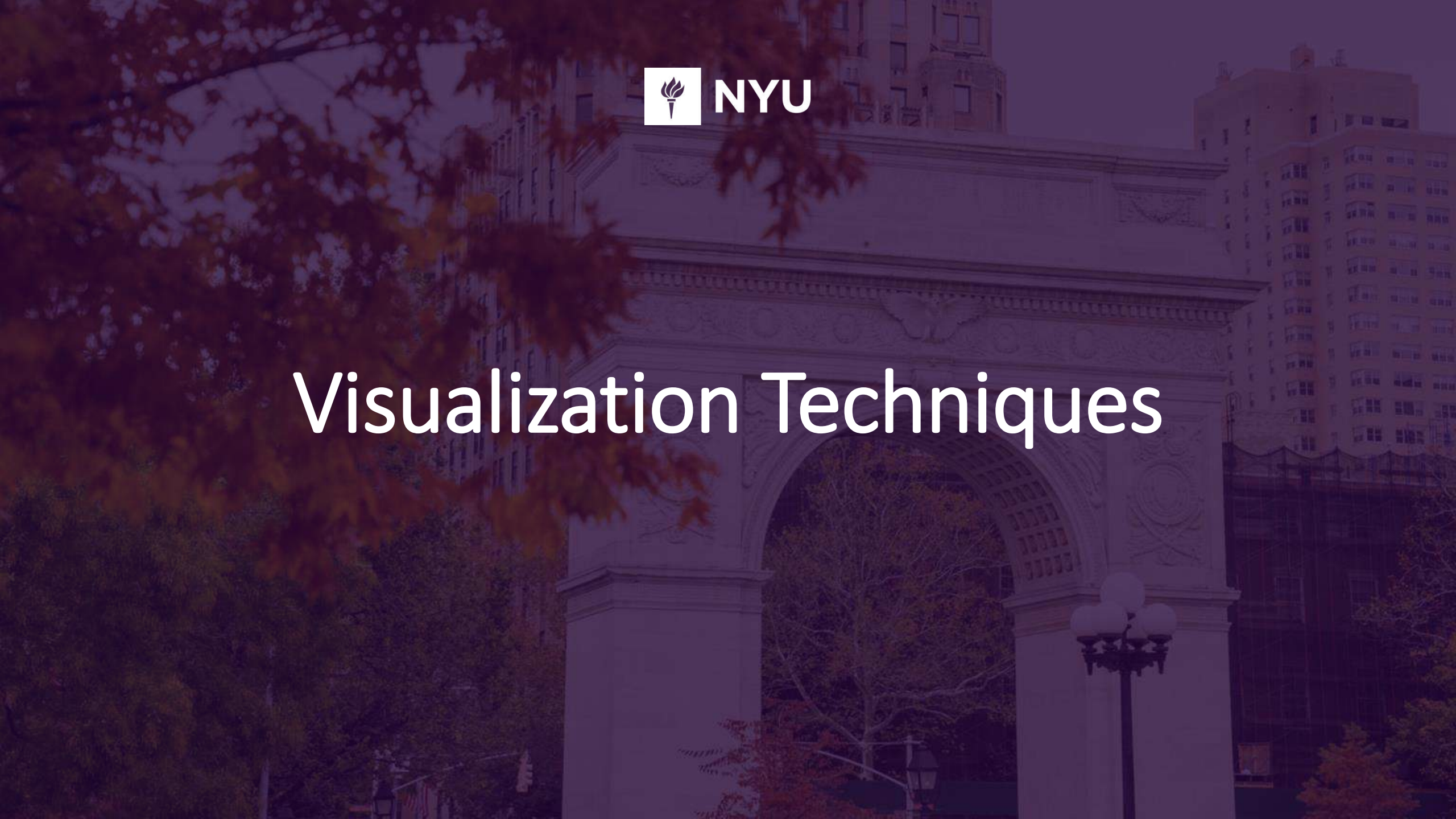
Time Structure

- Time is hierarchical & multi-resolution!



Time Structure: Hierarchical & multi-resolution

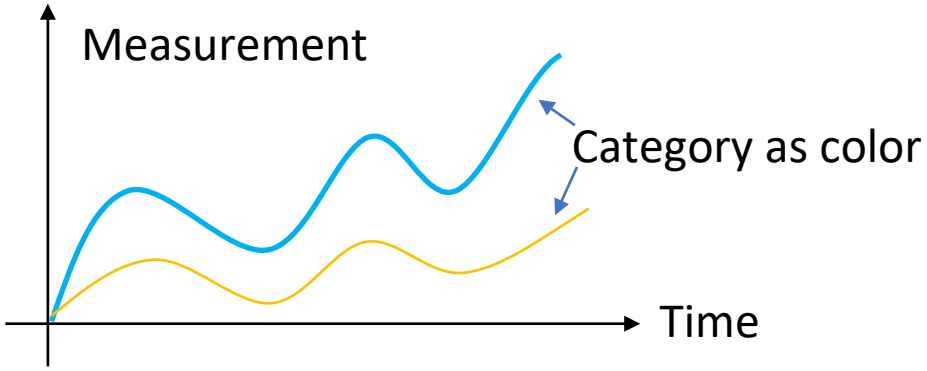


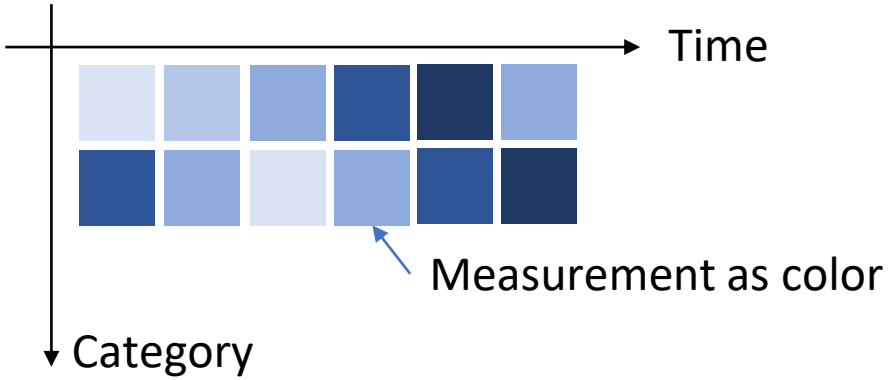


Visualization Techniques

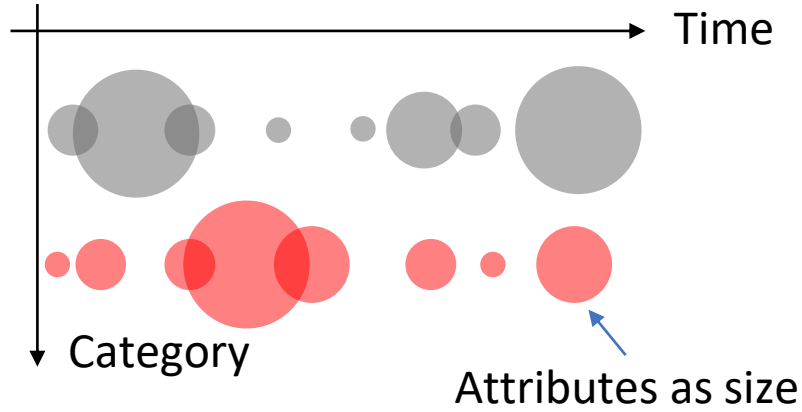
Visualization Techniques

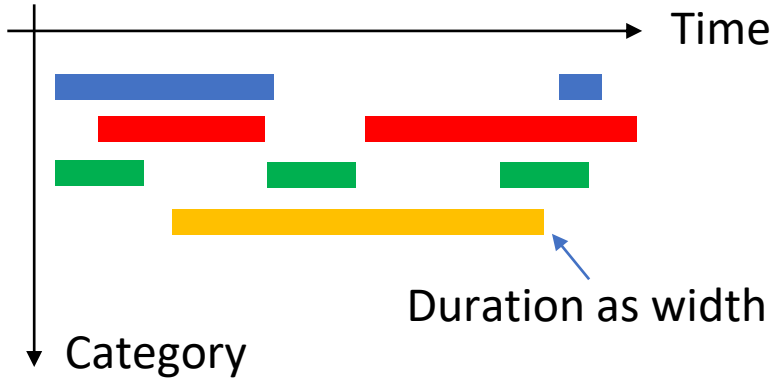
Measurement Data





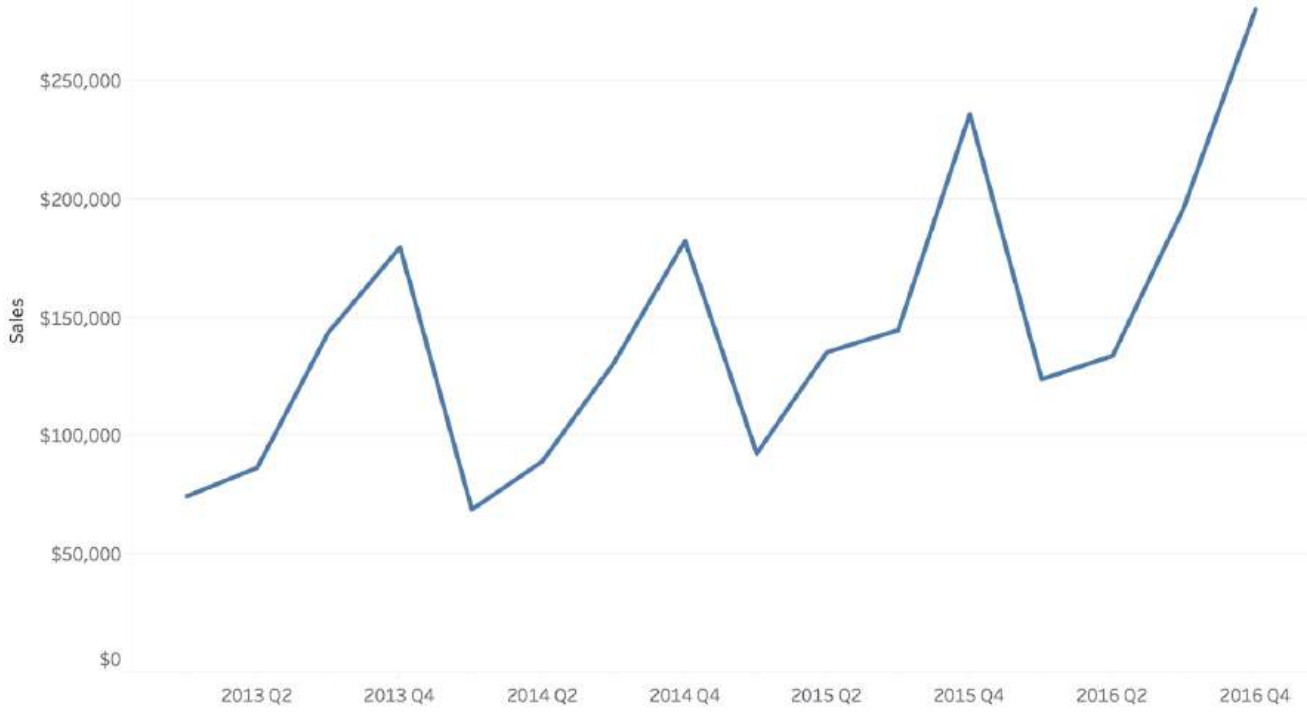
Event Data

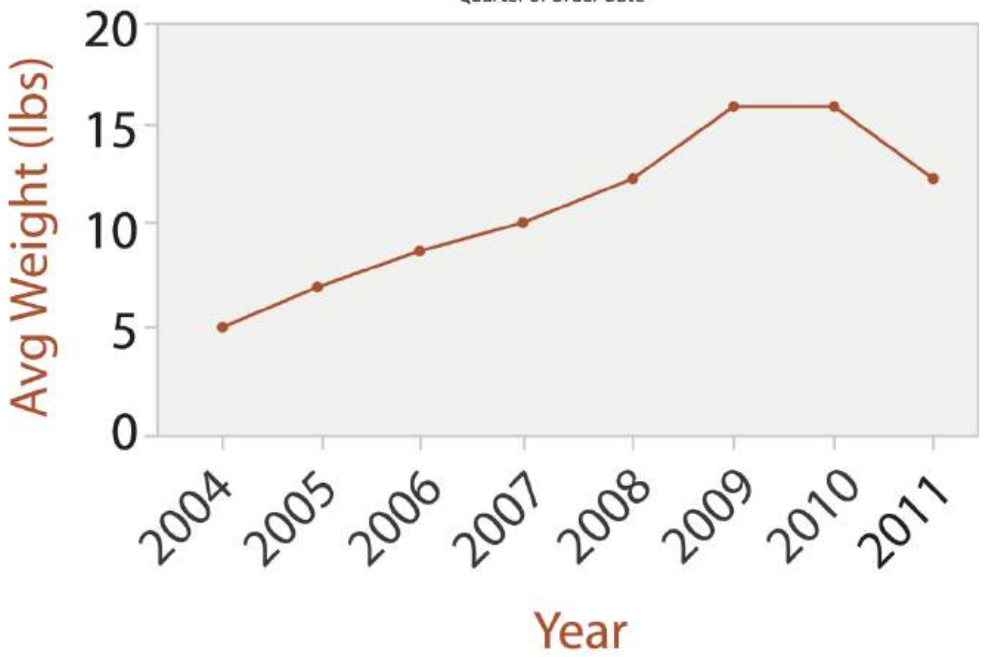




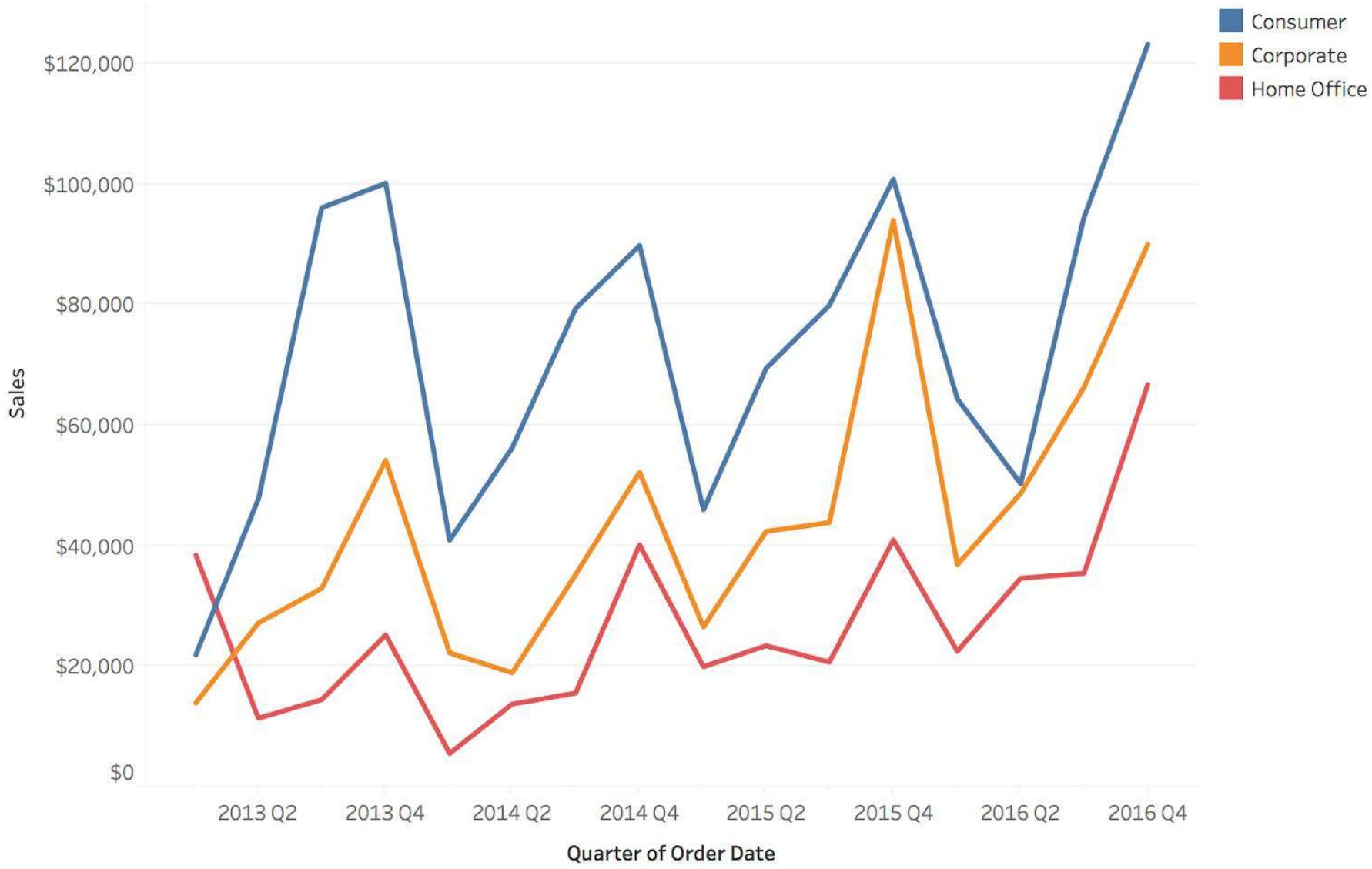
Line Chart

- **Data:** two quantitative attributes
- **Mark:** points and line connection marks between them OR just lines
- **Channels:**
  - Aligned lengths to express quant. value
  - Separated and ordered by key attribute into horizontal regions
- **Task:** find trend
  - Connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
- **Scalability:** hundreds of key levels (x axis), hundreds of value levels (y axis)

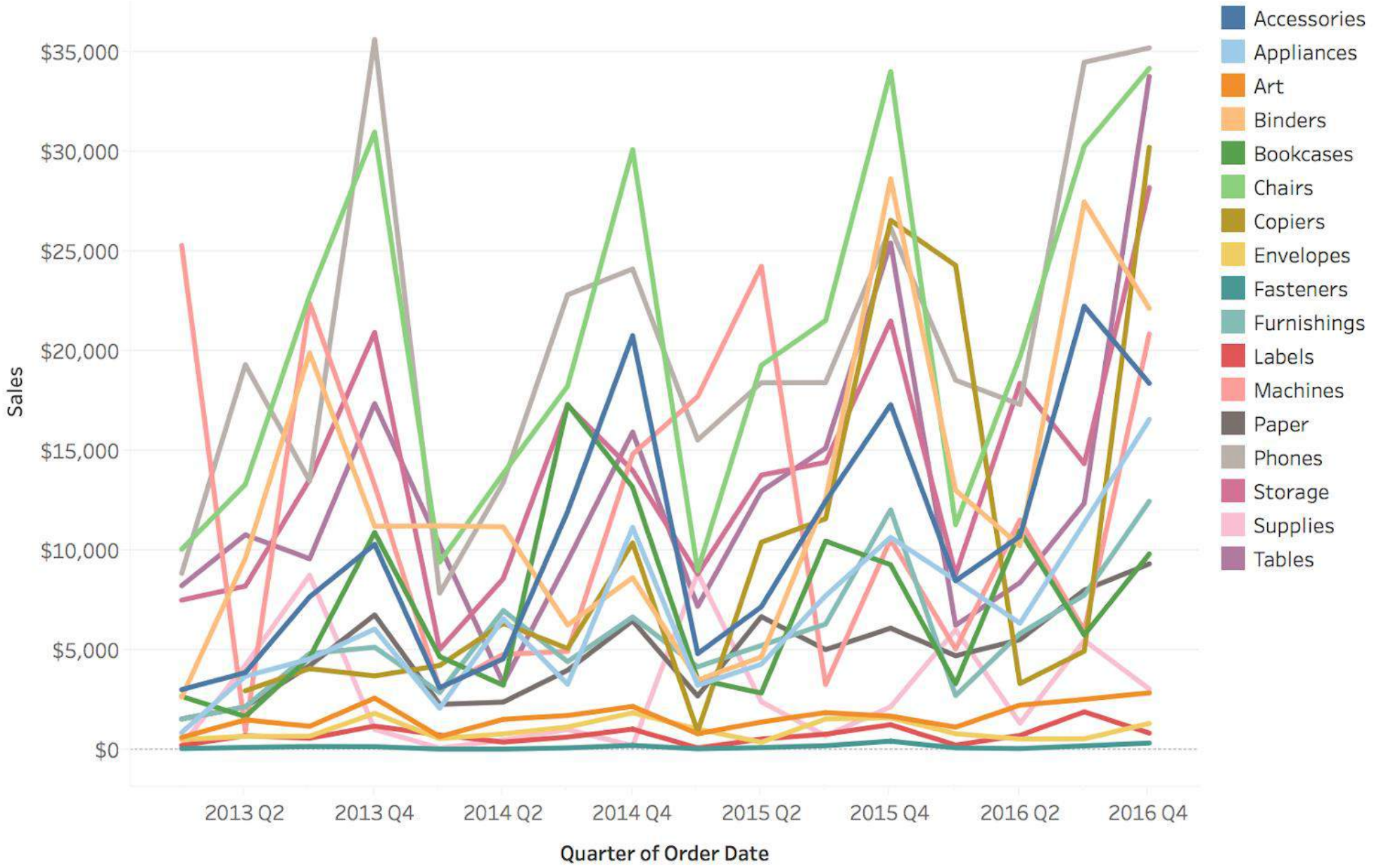




Line Chart


Consumer
Corporate
Home Office
$120,000
$100,000
$80,000
$60,000
$40,000
$20,000
$0
Sales
2013 Q2
2013 Q4
2014 Q2
2014 Q4
2015 Q2
2015 Q4
2016 Q2
2016 Q4
Quarter of Order Date

Line Chart: Beware Overplotting!

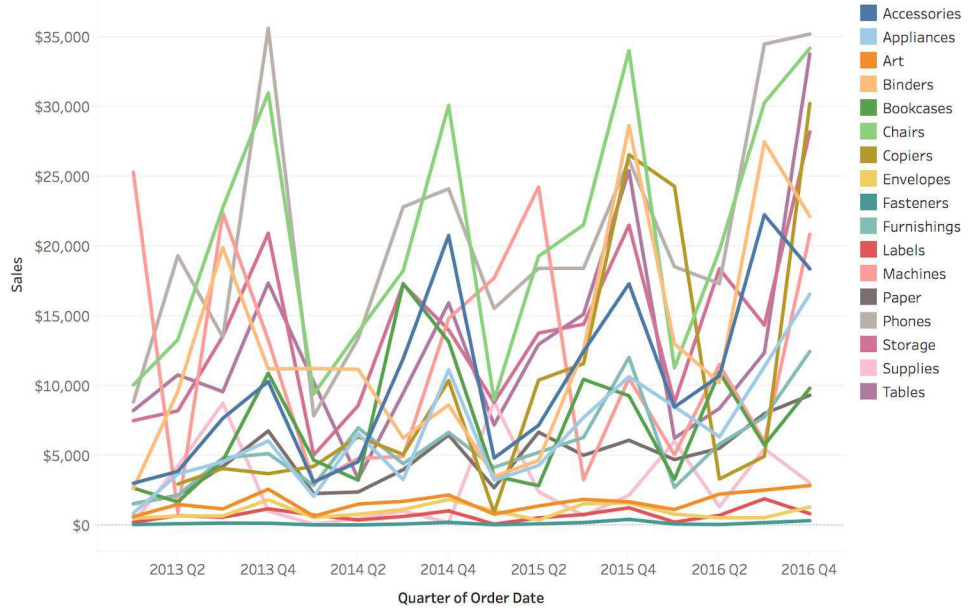


How to avoid overplotting

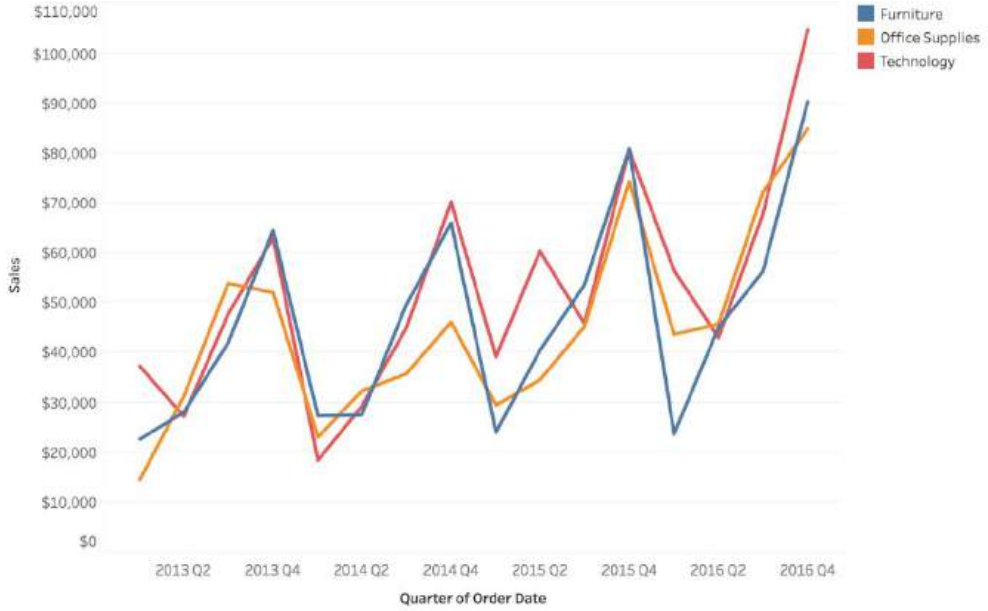
- **Grouping**
  - Combine data so there are fewer categories
- **Filtering/focusing**
  - Only show a few categories of interest
- **Highlighting**
  - Gray out un-interesting categories and highlight the categories of interest with bright color and larger line width.

How to avoid overplotting

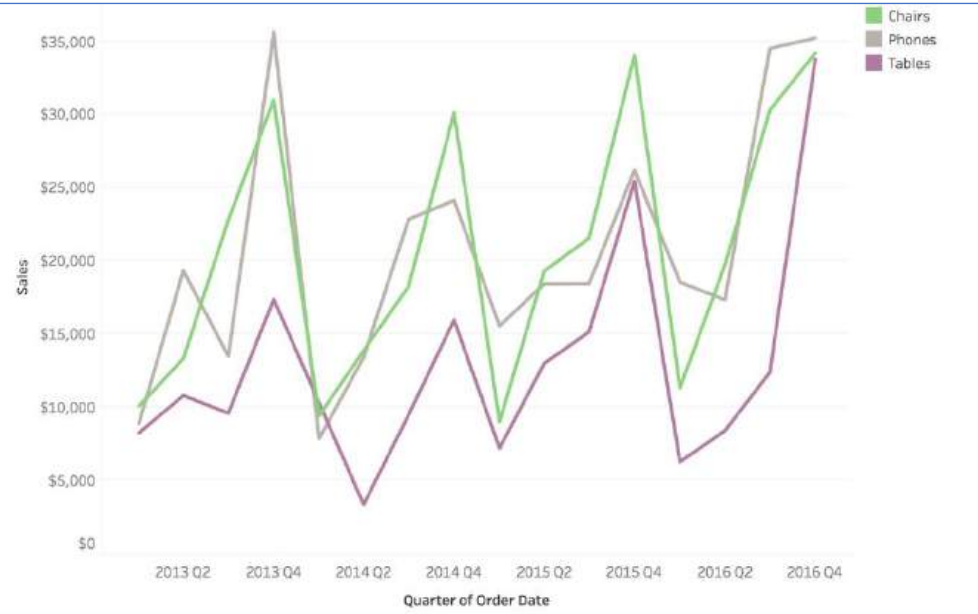
Original



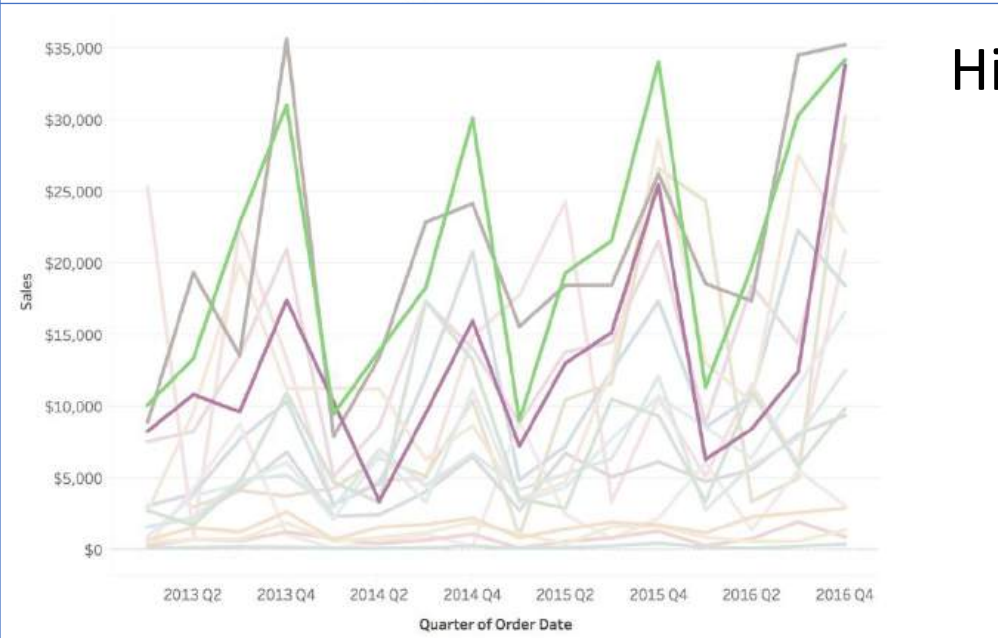
Grouping



Filtering

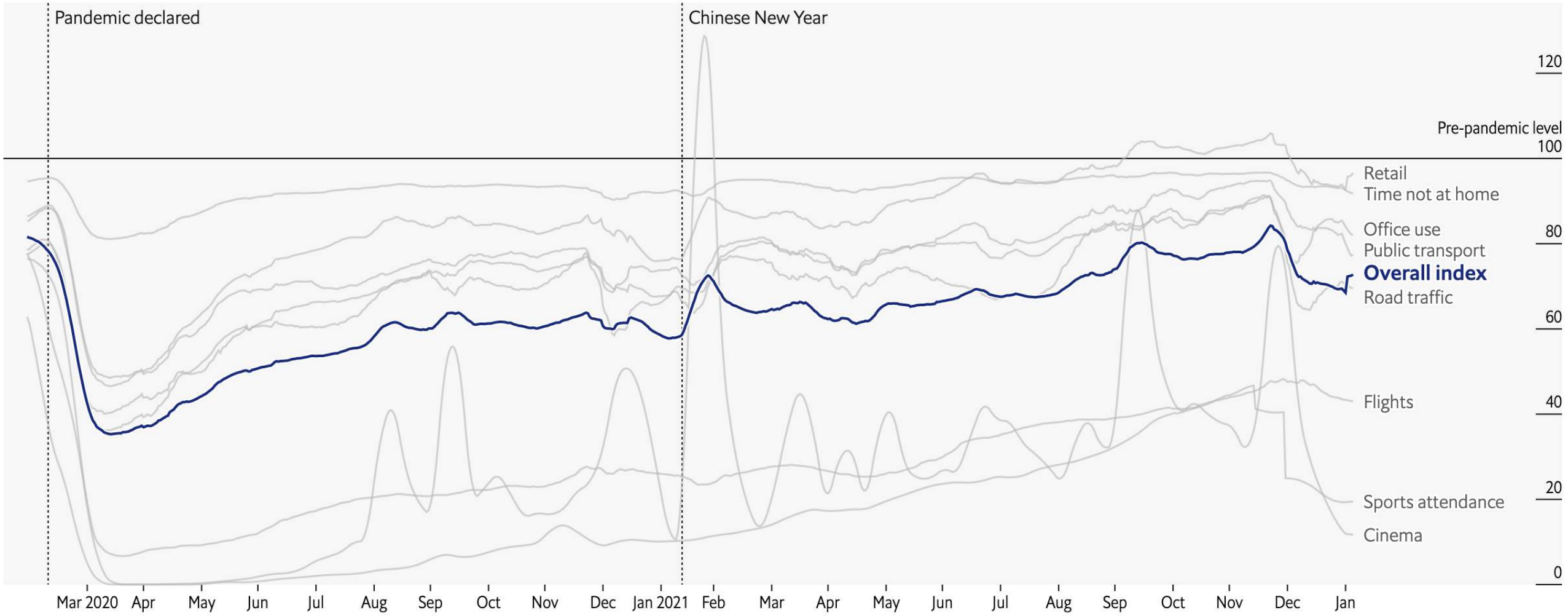


Highlighting



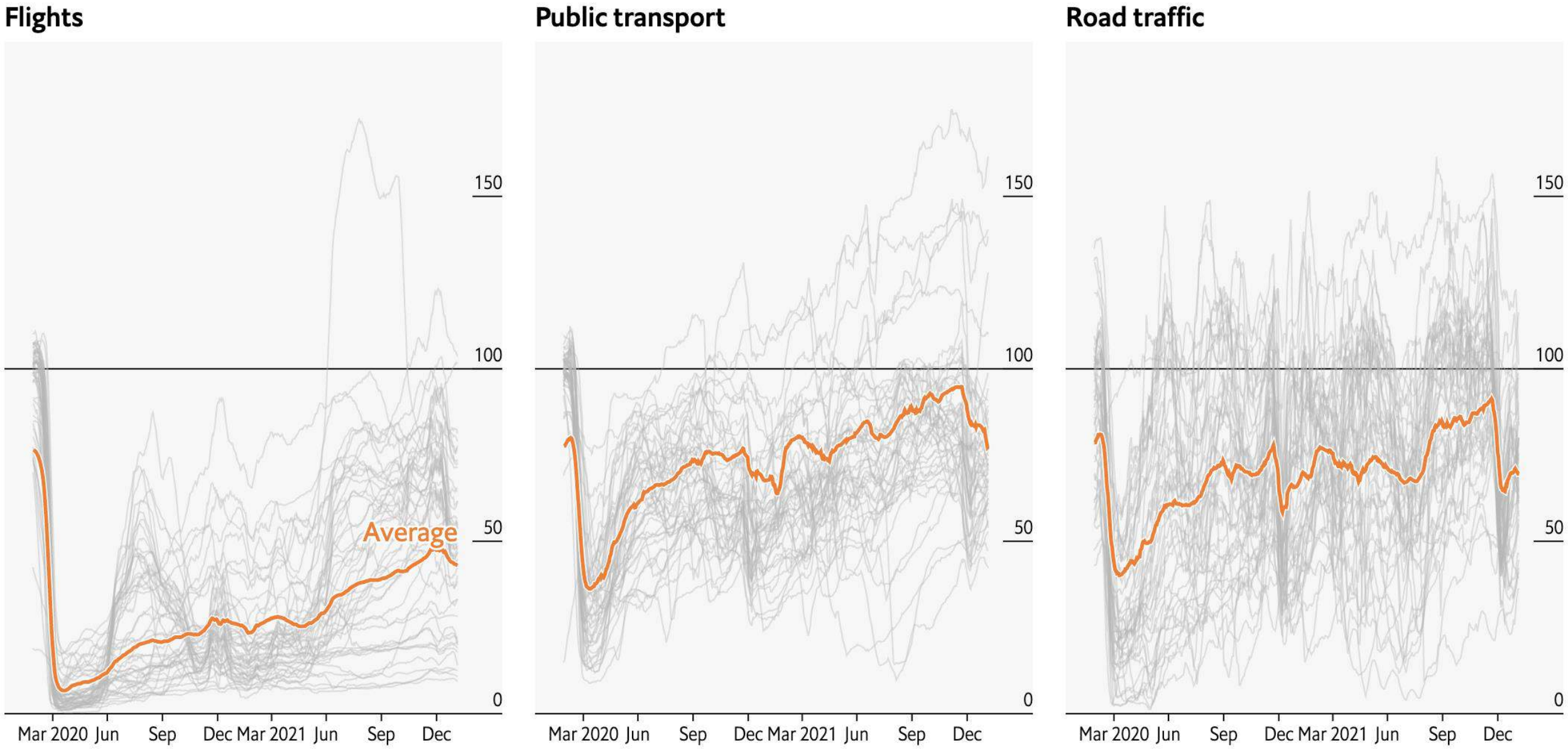
Example: Global normalcy index

- https://www.economist.com/graphic-detail/tracking-the-return-to-normalcy-after-covid-19



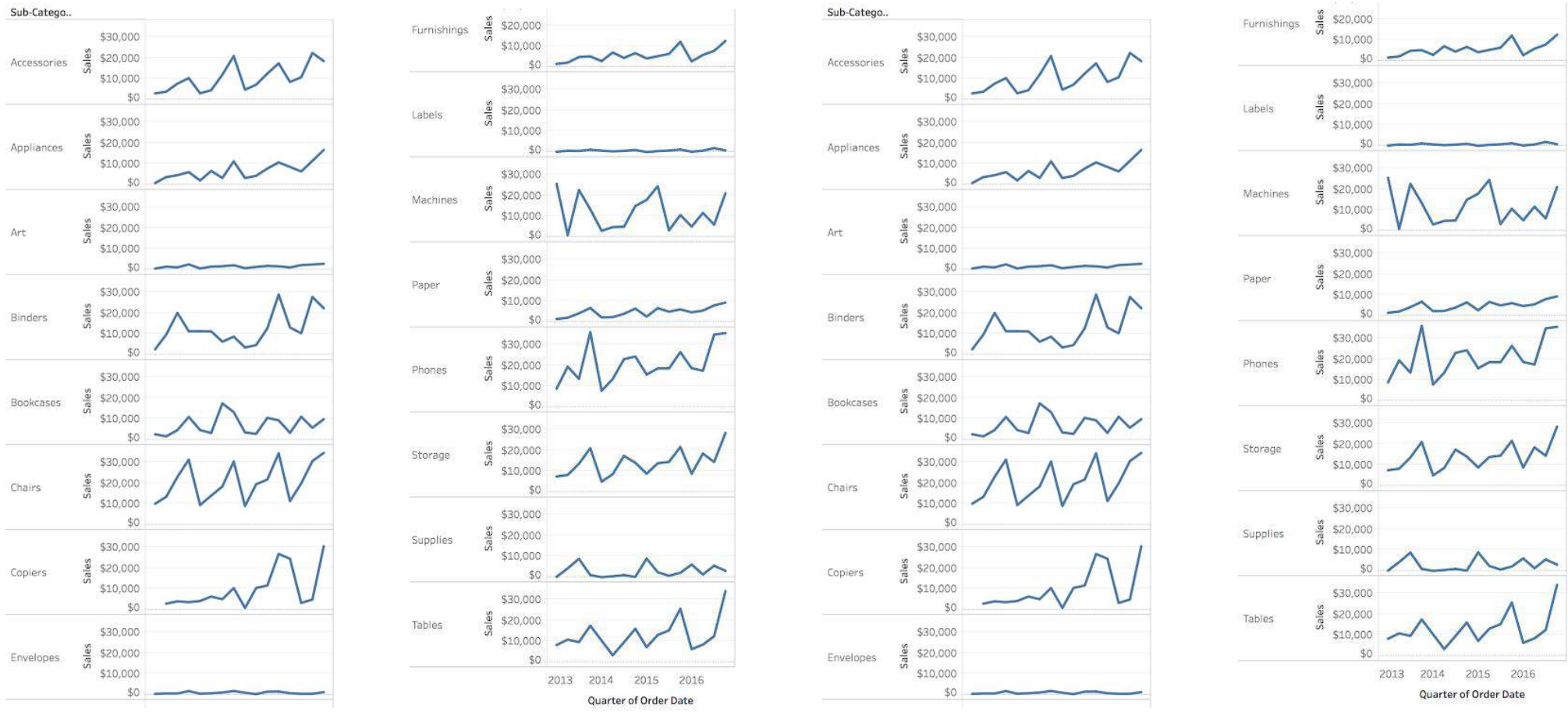
Example: Global normalcy index

- https://www.economist.com/graphic-detail/tracking-the-return-to-normalcy-after-covid-19



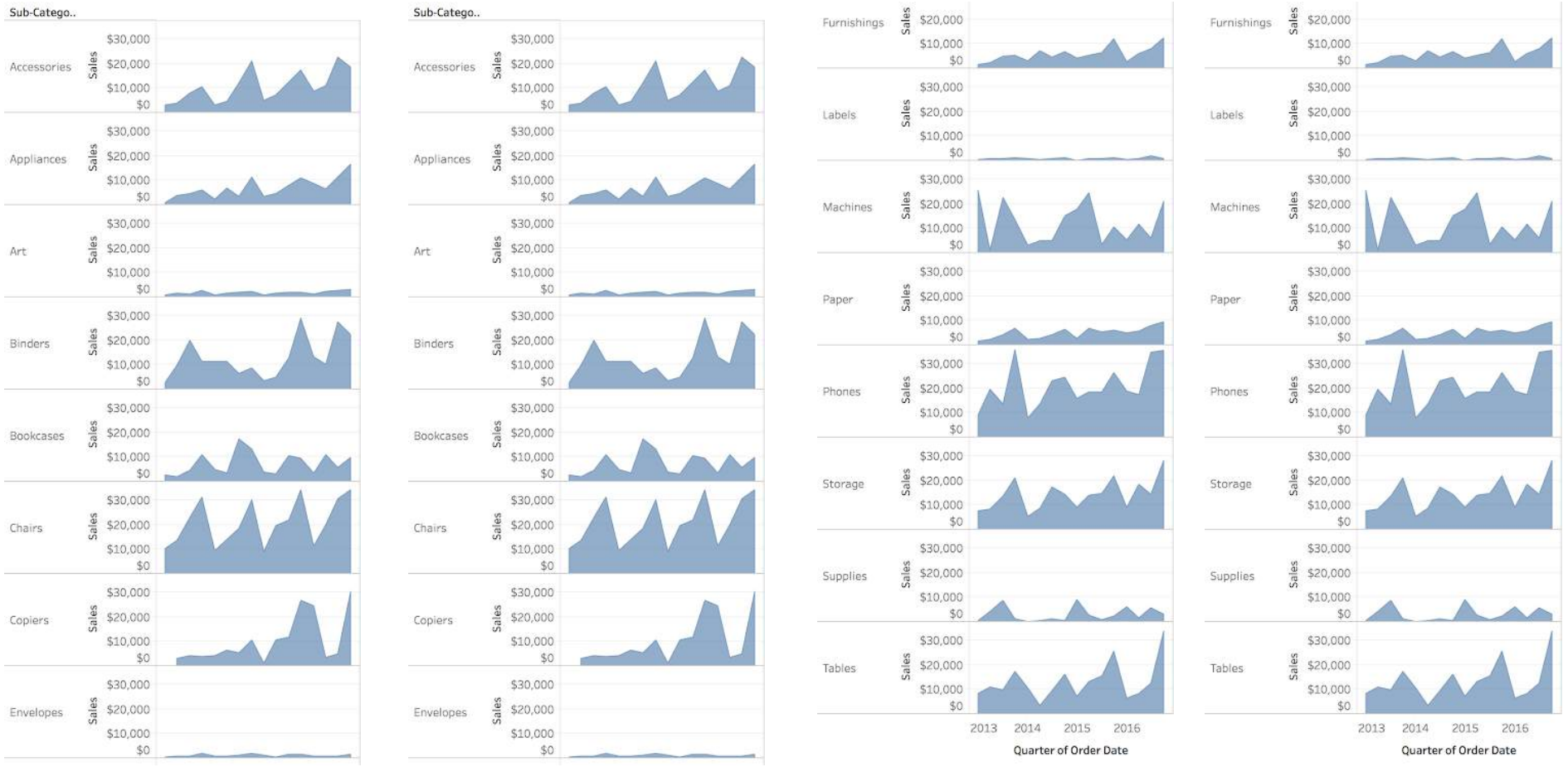
Line Chart: Faceting/multiple small charts

- Good for comparisons of many timelines
  - *"At the heart of quantitative reasoning is a single question: Compared to what? Small multiple designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives. For a wide range of problems in data presentation, small multiples are the best design solution."*



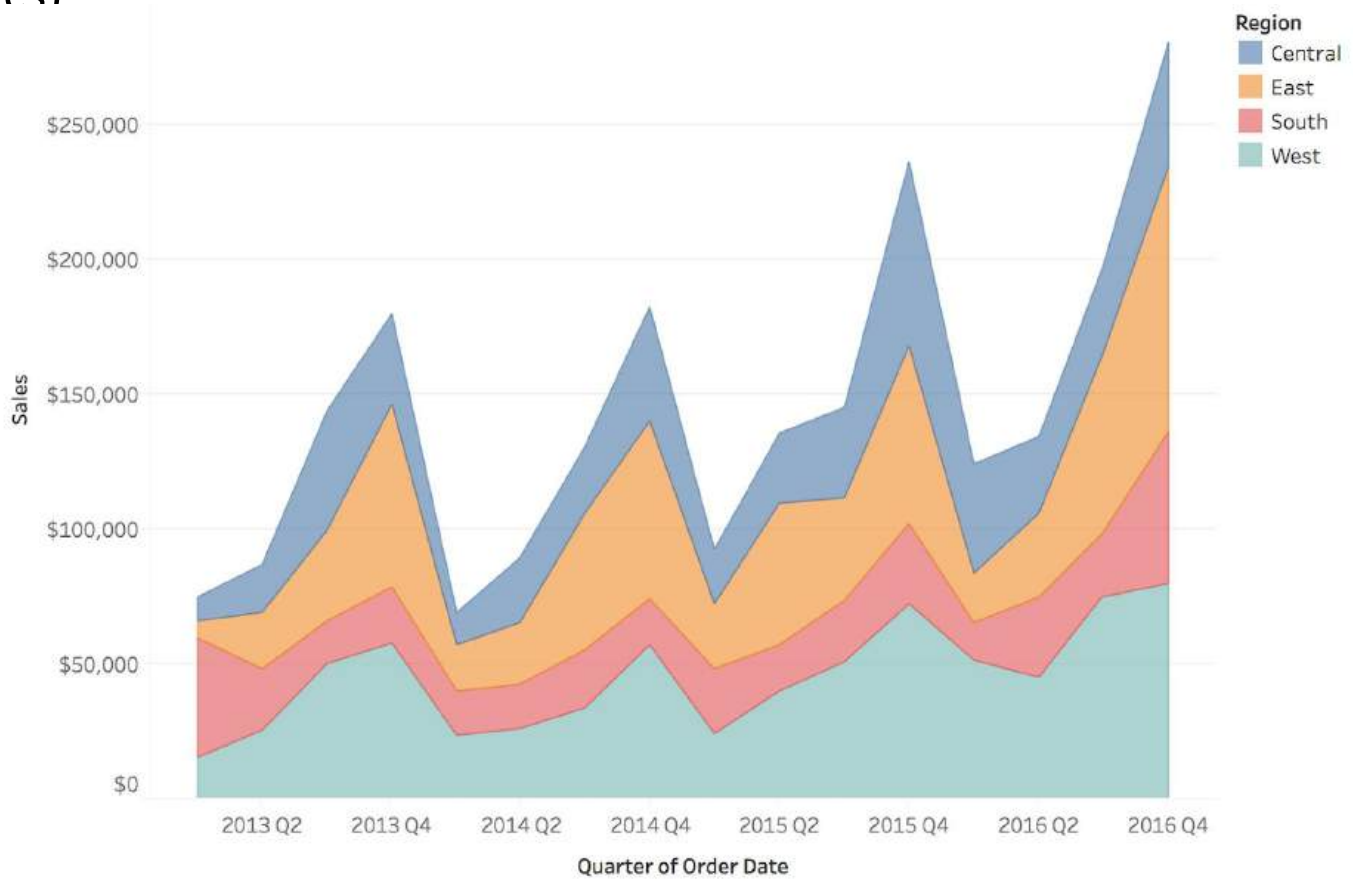
Line Chart: Faceting/multiple small charts

- Good for comparisons of many timelines
  - *"At the heart of quantitative reasoning is a single question: Compared to what? Small multiple designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives. For a wide range of problems in data presentation, small multiples are the best design solution."*



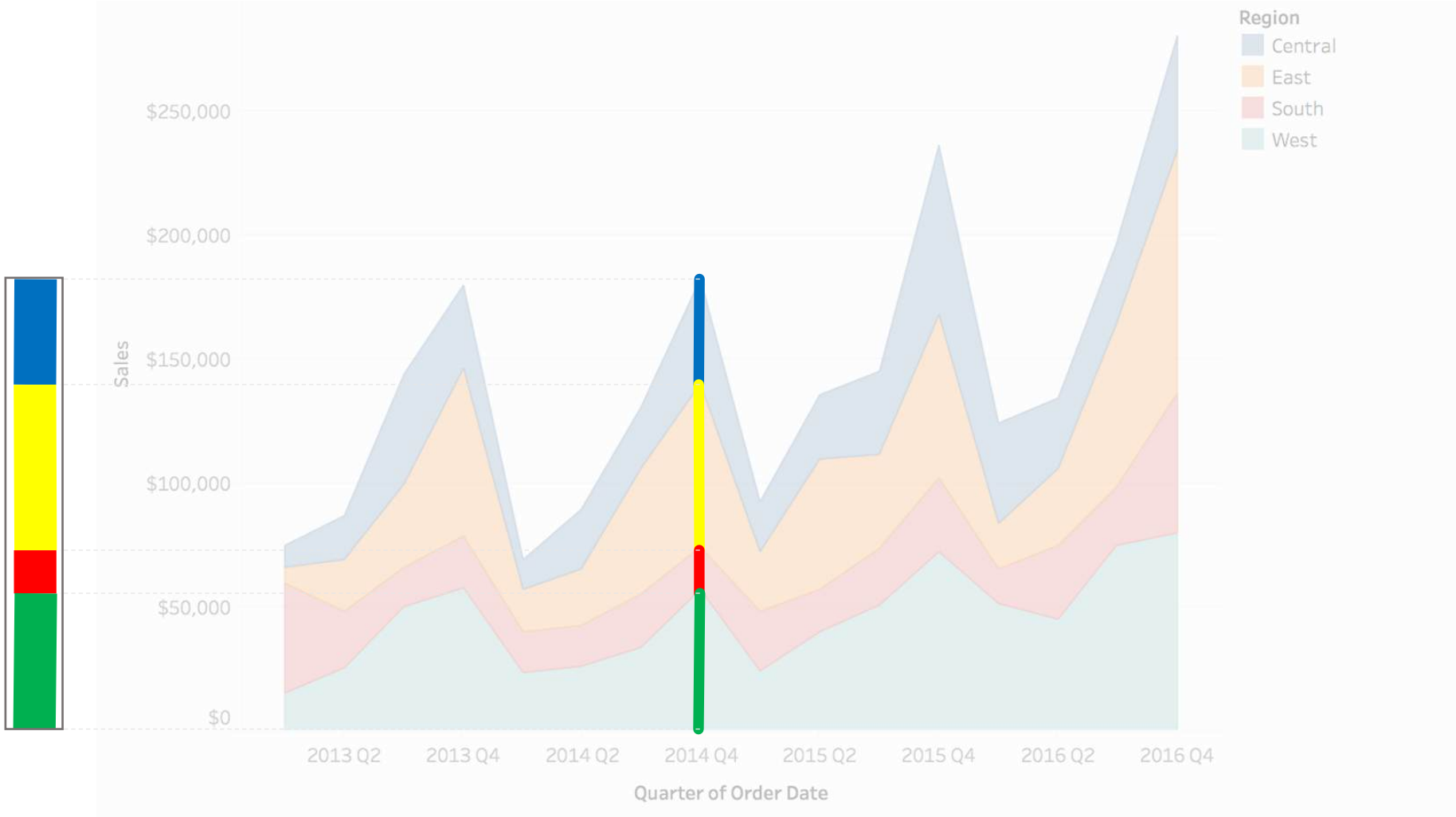
Stacked area chart

- **Data:**
  - One categorical key attribute (area aesthetics)
  - One ordered key attribute (time)
  - One quantitative value attribute
- **Marks:** ?
  - Areas
- **Channels:** ?
  - Color: category
  - Position (height): value
- **Task:** See overall (total) trend, compare proportions

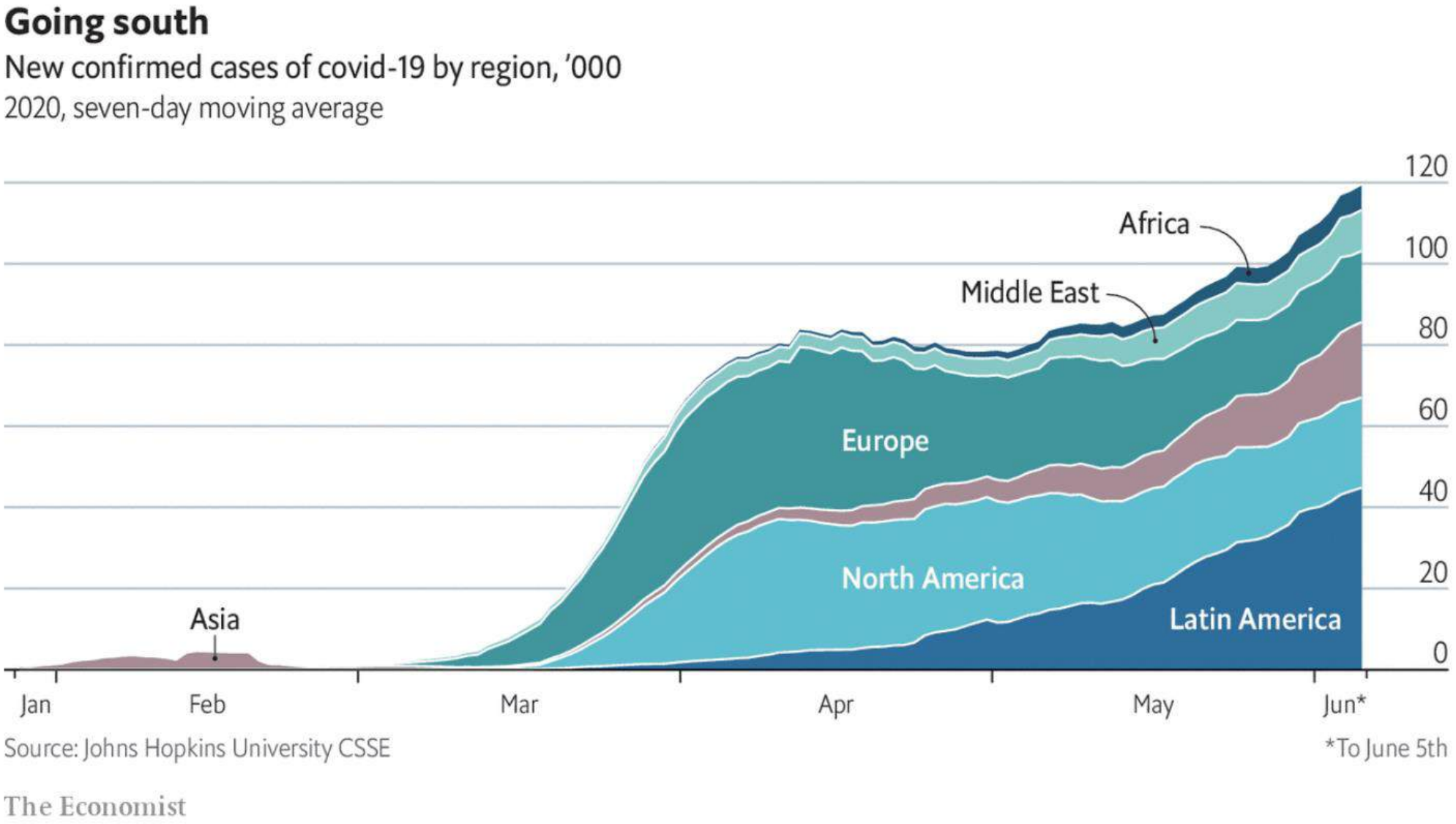


Stacked area chart

- Data corresponds to the same time point the areas are stacked on



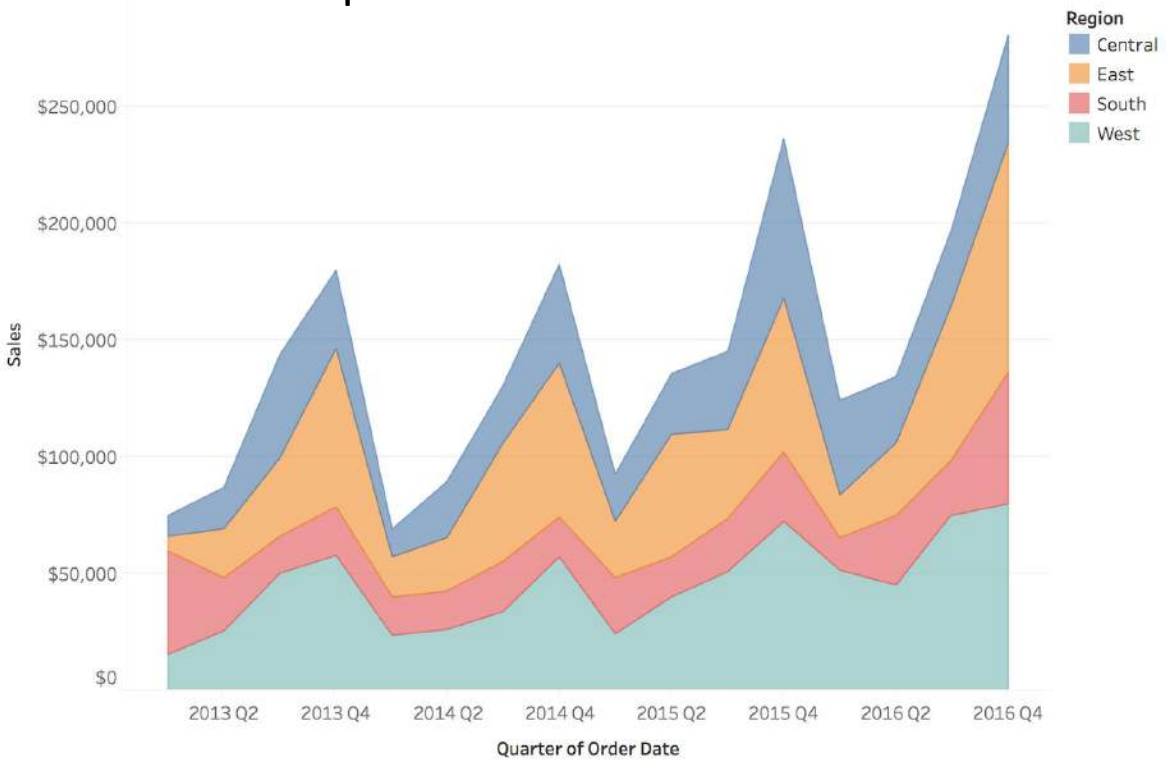
Stacked area chart

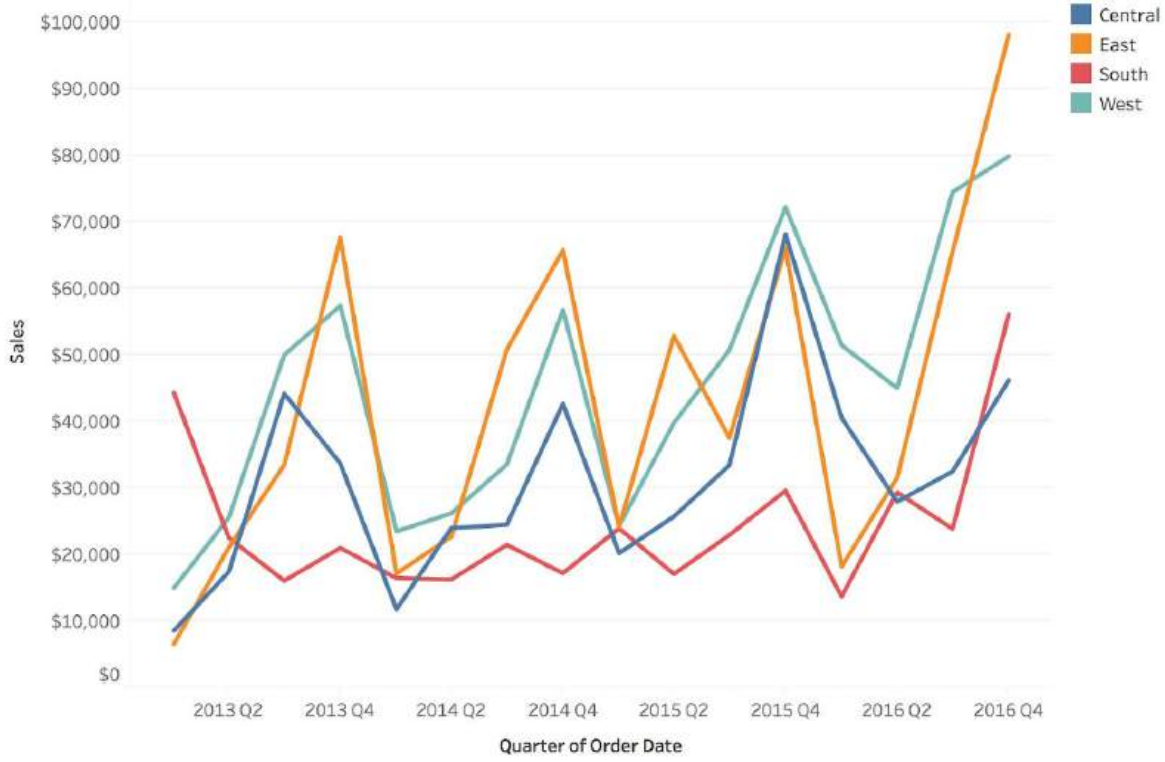


https://www.economist.com/graphic-detail/2020/06/10/coronavirus-cases-are-surging-in-latin-america

Stacked area chart

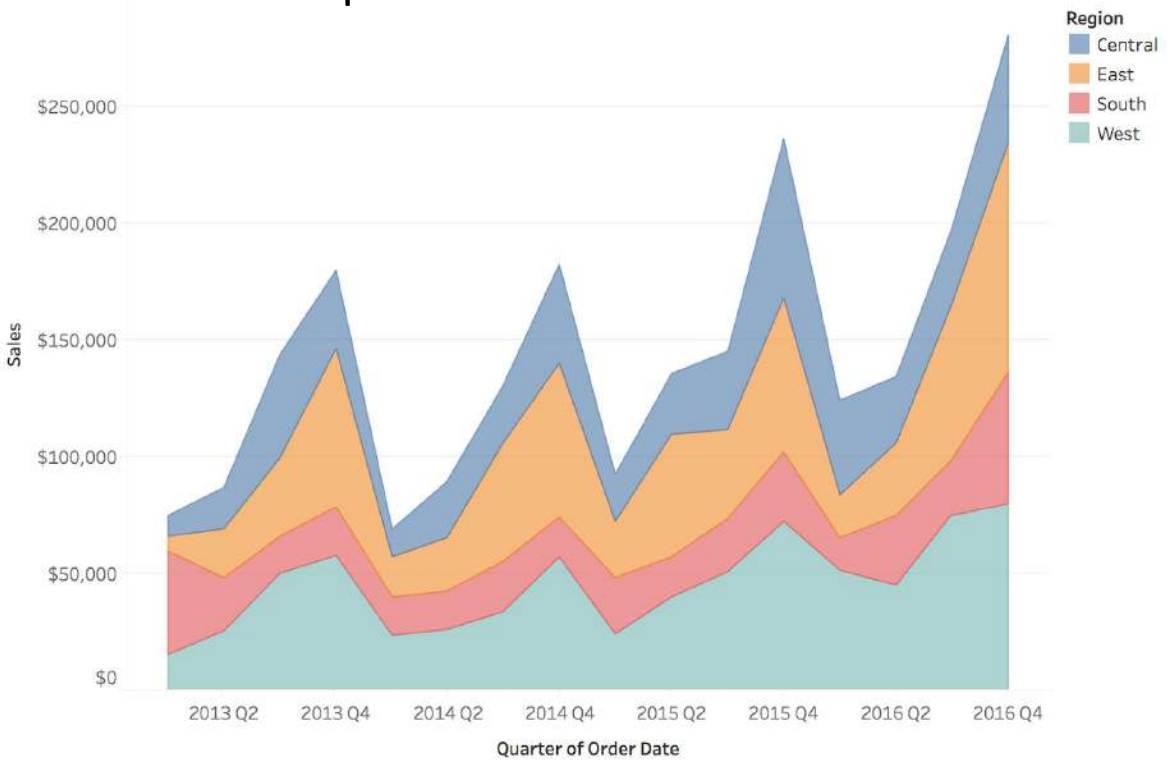
Only vertical width is meaningful.
Temporal trend hard to read.

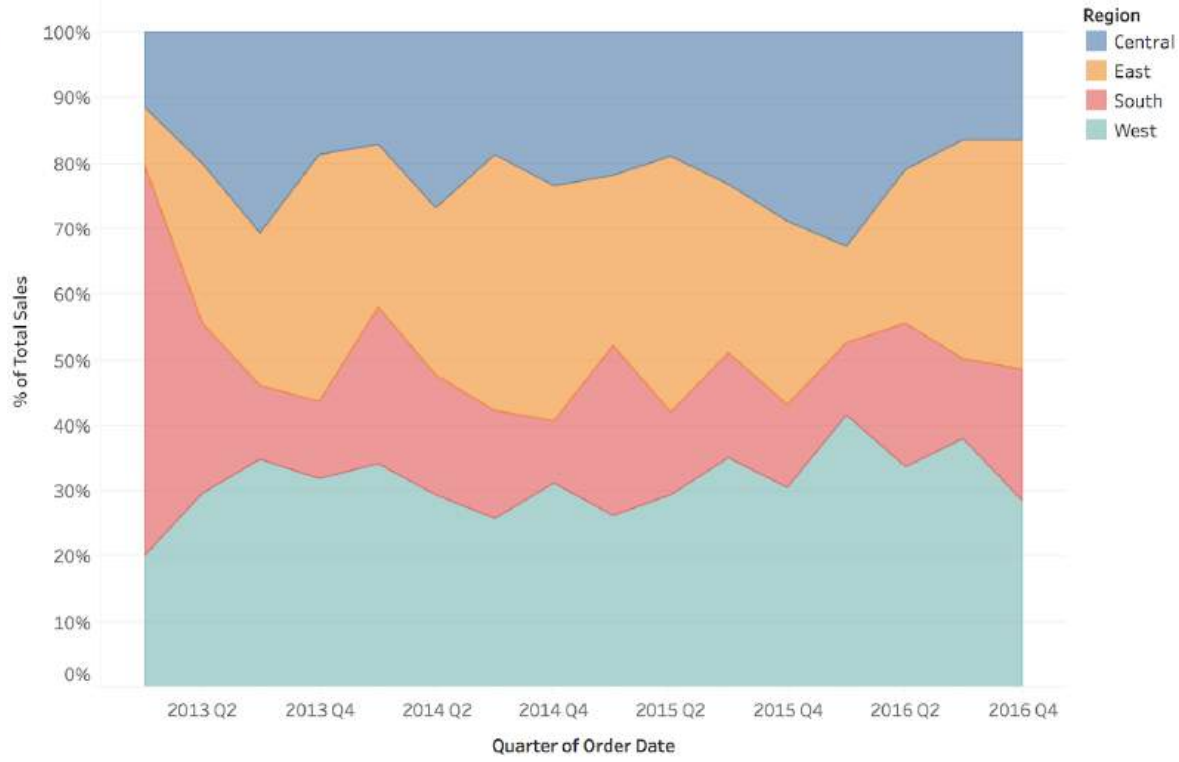




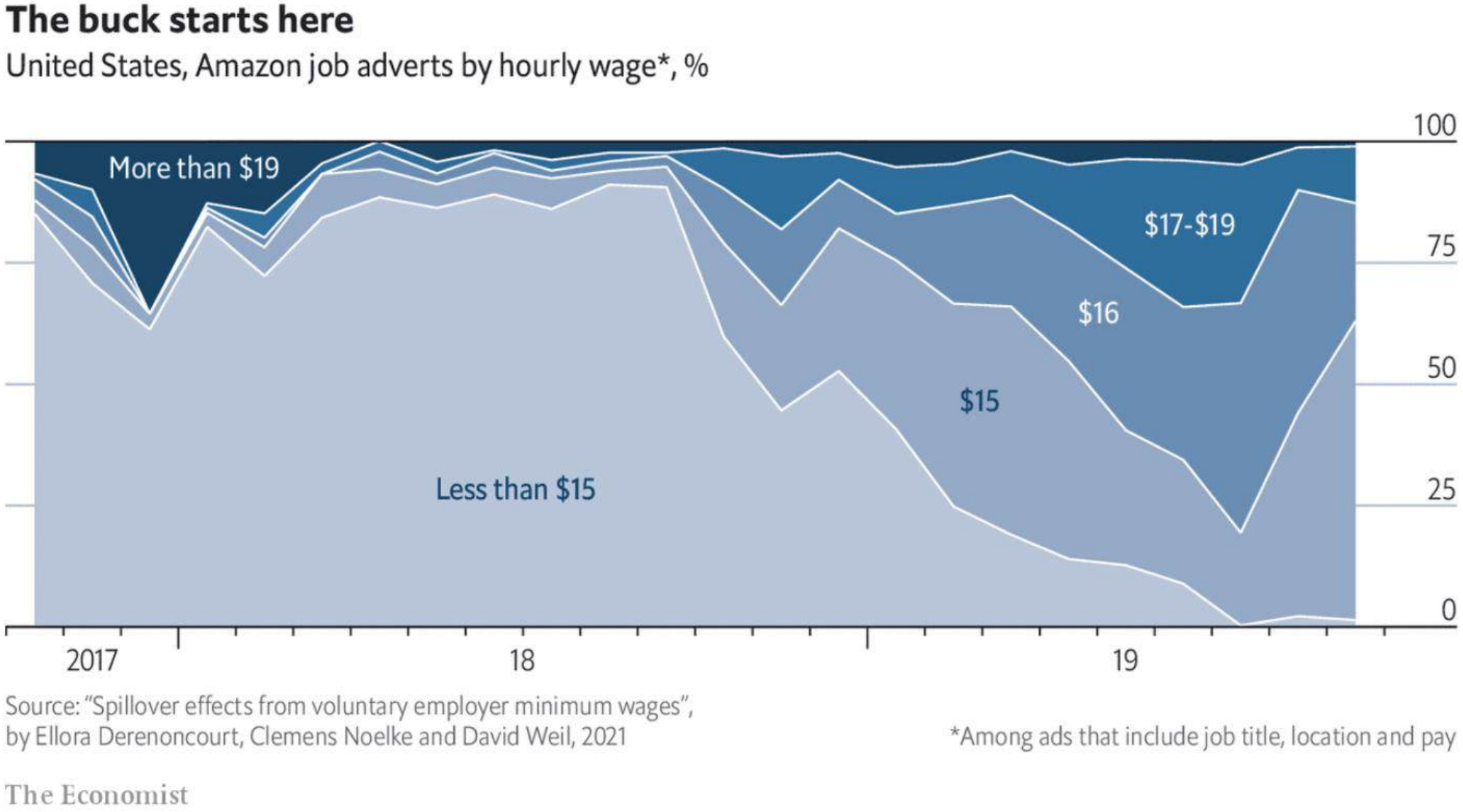
Percent area chart

Only vertical width is meaningful.
Temporal trend hard to read.





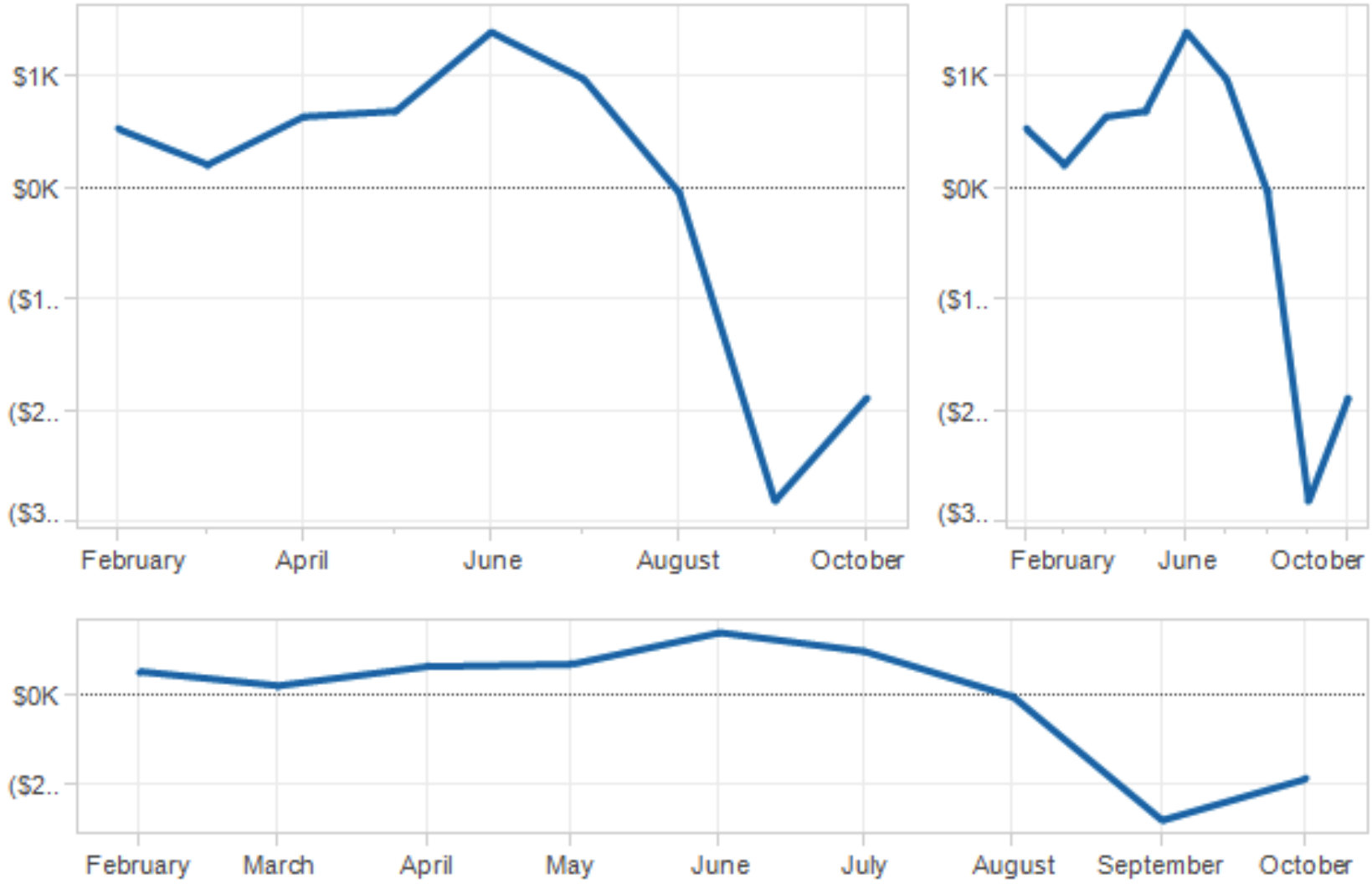
Percent area chart



https://www.economist.com/graphic-detail/2021/04/30/when-amazon-raises-wages-nearby-firms-follow-suit

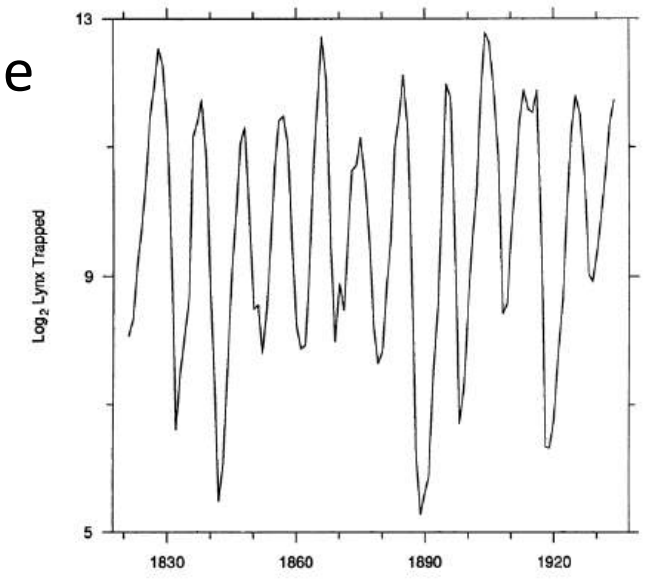
Aspect ratio matters

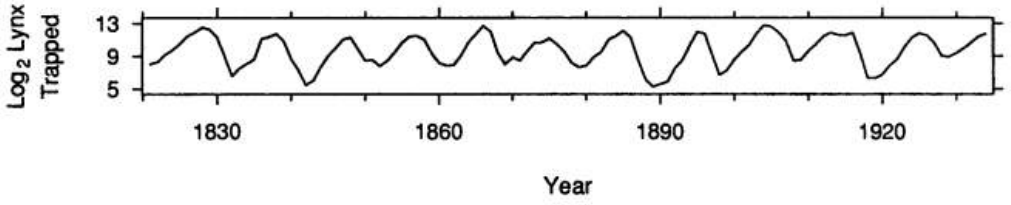
- Affects the perception of rate of change (slope)



Aspect ratio matters

- Always test different aspect ratios and see which one convey your message better.
- Optimization technique for computing the aspect ratio such that the average absolute orientation of line segments in the chart is equal to 45 degrees





The Shape Parameter of a Two-Variable Graph

WILLIAM S. CLEVELAND, MARYLYN E. McGILL, and ROBERT McGILL*

The shape parameter of a two-variable graph is the ratio of the horizontal and vertical distances spanned by the data. For at least 70 years this parameter has received much attention in writings on data display, because it is a critical factor on two-variable graphs that show how one variable depends on the other. But despite the attention, there has been little systematic study. In this article the shape parameter and its effect on the visual decoding of slope information are studied through historical, empirical, theoretical, and experimental investigations. These investigations lead to a method for choosing the shape that maximizes the accuracy of slope judgments.

KEY WORDS: Statistical graphics; Aspect ratio; Graphical perception; Visual perception.

1. INTRODUCTION

Figure 1 gives data on the amount of solar radiation penetrating sea water at different depths: the filled circles are actual measurements and the open circles are estimates (Littler, Littler, Blair, and Norris 1985). The dashed rectangle, which is the *data rectangle*, shows the maximum and minimum values of the data along both the vertical scale and the horizontal scale. Suppose the height of the data rectangle of a graph is h centimeters (cm) and the width is w cm. The *shape parameter*, or *shape*, of the graph is h/w. In Figure 1 the shape is .85.

Shape is a critical factor for two-variable graphs that show the dependence of y on x. Figures 2 and 3 show the Canadian lynx data (Elton and Nicholson 1942), a time series of substantial historical interest because of the many analyses it has inspired (Campbell and Walker 1977). In Figure 2 the shape parameter is 1, and in Figure 3 the shape is .074. In Figure 2 it is impossible to see a critical property of the data that can be seen in Figure 3—the number of lynx trappings rises more slowly than it declines. As we shall explain, this phenomenon of graphical perception—the better perception of the lynx rise and fall in Figure 3—is a result of the effect that the change in shape has on our judgments of slopes.

It is hard to find a statistical-graphics topic more universally discussed than shape. But despite the importance of this parameter and the ubiquity of comments on it, there has been almost no systematic study. In this article we study shape in several ways. Section 2 defines objects critical to the study. In Section 3, comments and recommendations from a sample of twentieth-century writings on graphical data display are reviewed. In Section 4, measurements of the shapes of 481 graphs are analyzed. In Section 5, observations are made about our processing of slope information on a graph; the discussion is critical to the subsequent study of shape and slope, because it determines in a fundamental way how the problem is approached. Section 6 contains theory; we hypothesize that the accuracy of slope judgments depends on what is called *orientation resolution*. The dependence of orientation resolution on other quantities is then investigated. Section 7 describes an experiment that probes the hypothesis of Section 6. In Section 8, the results of the theory and experimentation are used to develop an algorithm for choosing the value of the shape parameter that maximizes both the resolution and accuracy of slope judgments. Section 9 concludes the article with a discussion of the general applicability of our methods, including the significance (for the analysis of the lynx data) of the phenomenon observed in Figure 3.

2. DEFINITIONS

Suppose we have a two-variable graph showing how y depends on x. In such a case, the decoding of quantitative information encoded by the slopes of line segments is a fundamental visual task that we perform. The reason, of course, is that the slopes encode the rate of change of y as a function of x, and decoding the rate of change is important for understanding the dependence of y on x.

2.1 Actual Line Segments and Virtual Line Segments

The line segments on a graph that encode slope information might be *actual* line segments drawn on the graph. For example, the lynx numbers in Figures 2 and 3 are graphed by connecting successive yearly values by line segments. We visually decode the slopes of these segments to infer the local rate of change of the lynx numbers through time. The line segments can also be superimposed on the graph by our visual system. For example, in Figure 1 we can visually superimpose segments connecting successive points on the graph to judge the local rate of change of log radiation as a function of depth. We will follow Marr (1982) and Stevens (1978) and refer to these superimposed segments as *virtual* line segments.

2.2 Coordinate Systems

Consider a point (x, y) inside the data rectangle of a graph. Now, x and y can have the units shown on the two

* William S. Cleveland and Robert McGill are statisticians, AT&T Bell Laboratories, Murray Hill, NJ 07974. Marylyn E. McGill is President, MEM Research, Inc., Murray Hill, NJ 07974. John Chambers, Colin Mallows, and Daryl Pregibon made helpful comments on an earlier draft of this article. We are indebted to the referees, whose comments led to a substantial improvement in the exposition. Joseph Follettie convinced us, from his comments on our earlier experiments in graphical perception, that showing stimuli for short time intervals is an important noise-reduction technique.

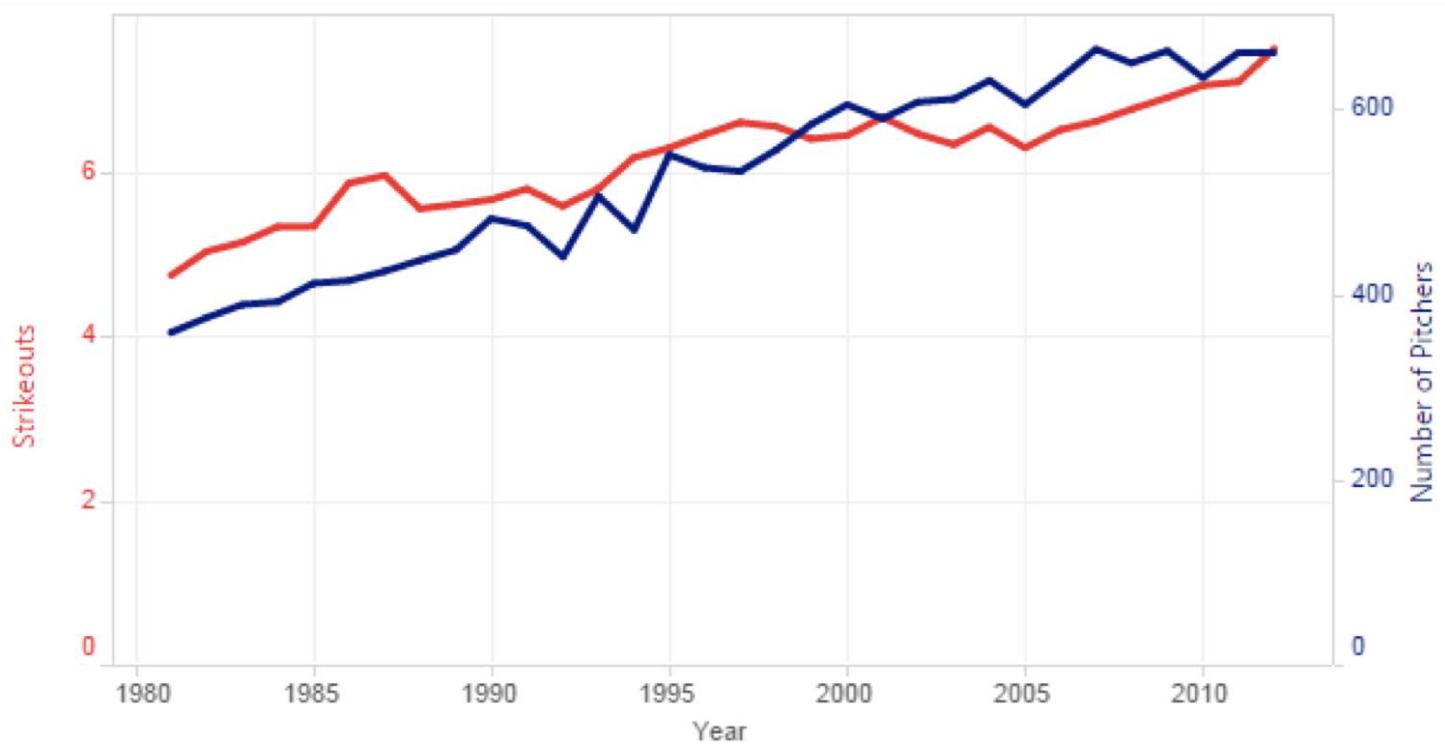
© 1988 American Statistical Association
Journal of the American Statistical Association
June 1988, Vol. 83, No. 402, Applications and Case Studies

This content downloaded from 216.165.95.148 on Thu, 21 Jun 2018 23:07:29 UTC
All use subject to http://about.jstor.org/terms

Cleveland, William S., et al. “The Shape Parameter of a Two-Variable Graph.” *Journal of the American Statistical Association*, vol. 83, no. 402, [American Statistical Association, Taylor & Francis, Ltd.], 1988, pp. 289–300, https://doi.org/10.2307/2288843

Dual-axis line charts

- Controversial
  - Why?
- Acceptable if similar y-axis scales
- Beware, very easy to mislead!



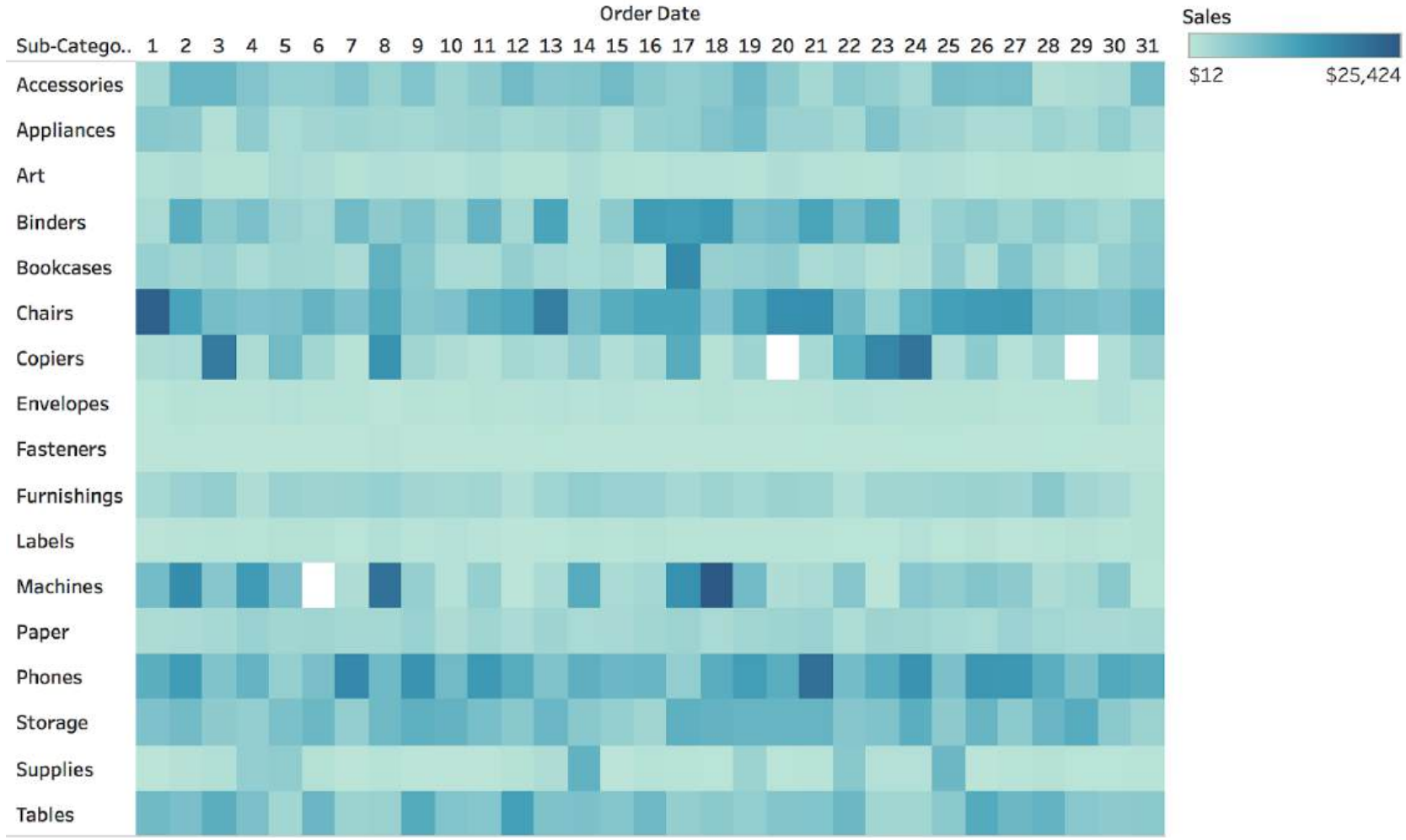
Indexed line chart

- Shows percentage change for a collection of line charts, based on some common starting point in time
- **Data:** two quantitative attributes
- **Derived data:** new quantitative value attribute
  - Percentage change
- **Task:** show **NORMALIZED** change over time
- Good when we care about relative changes, not absolute.
  - Eg: stock data https://mbostock.github.io/protovis/ex/index-chart.html

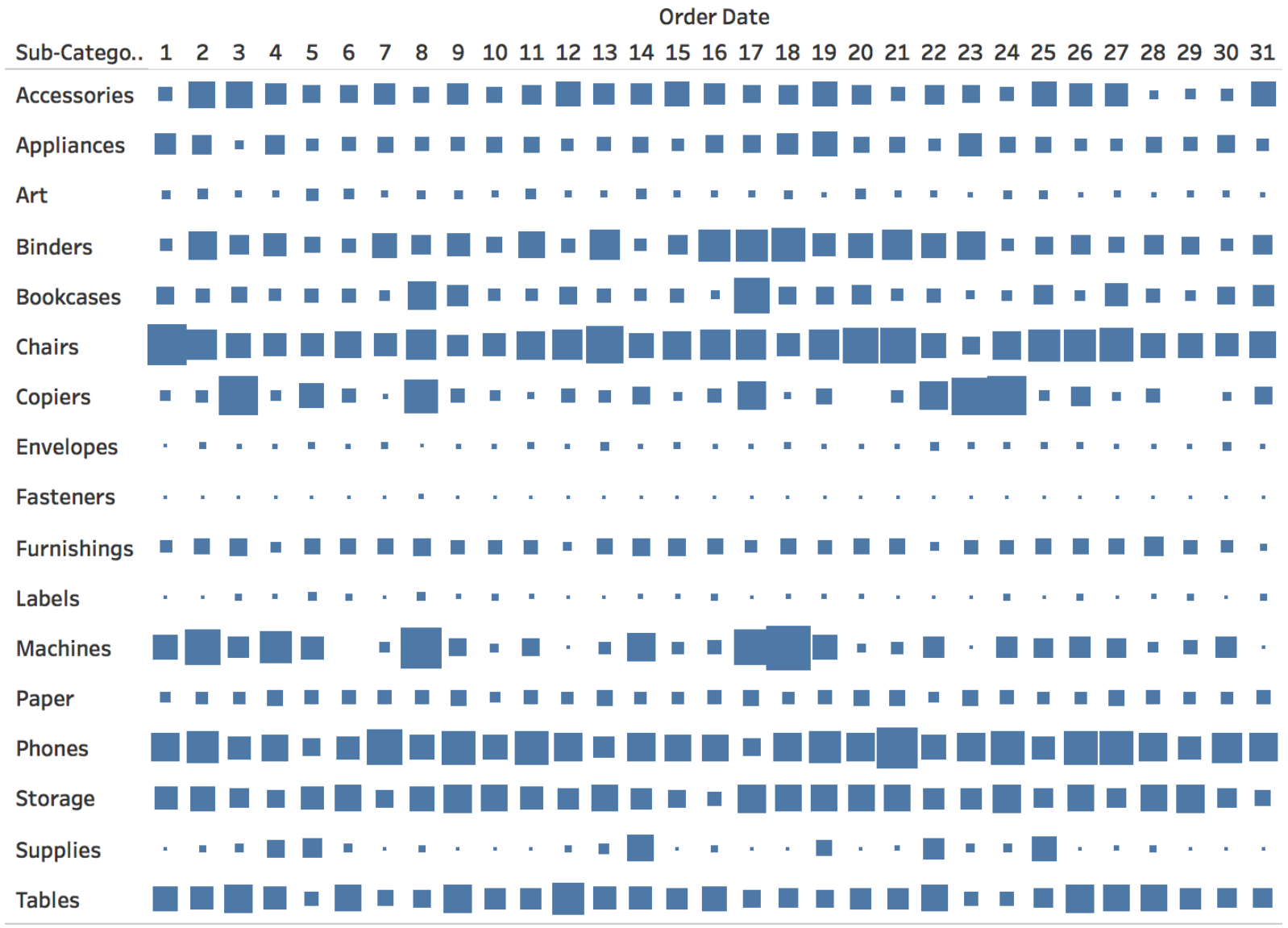
Other Methods

Heat Maps (Matrix)

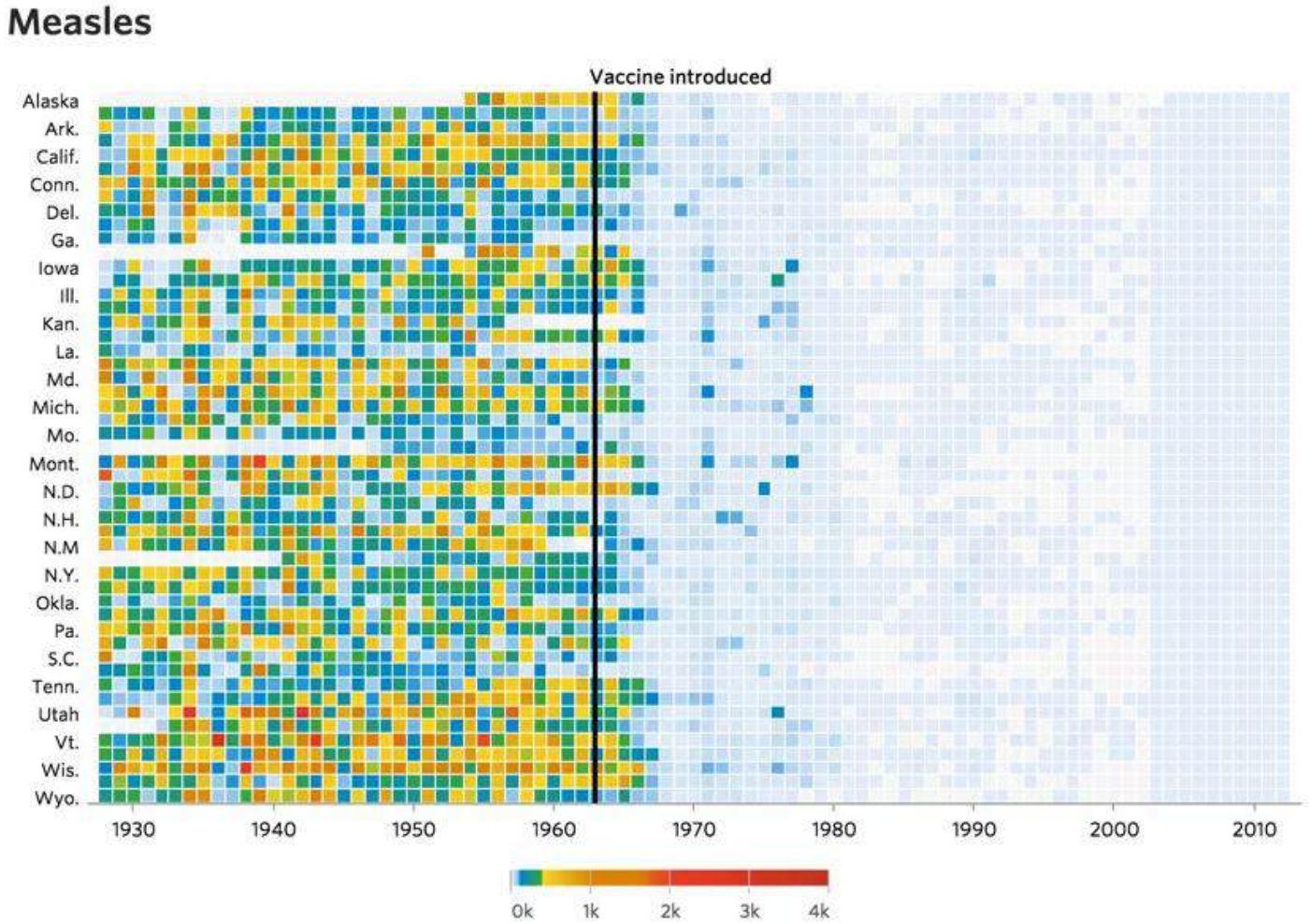
- Time on one axis
- Categories on other axis
- Values in cells (color, size)
- Can be quite space efficient



Heat Maps (Matrix)

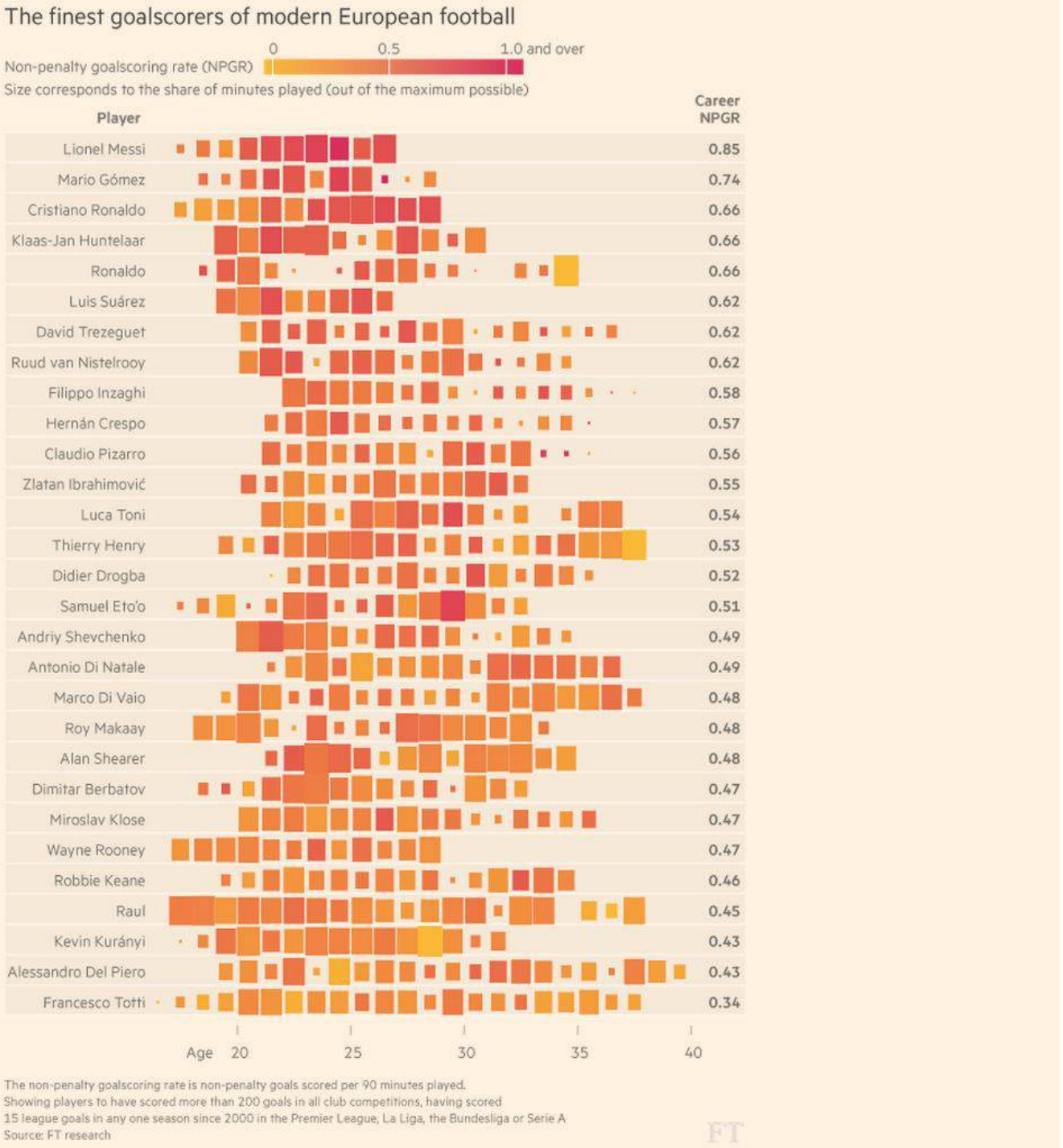


Heat Maps (Matrix)

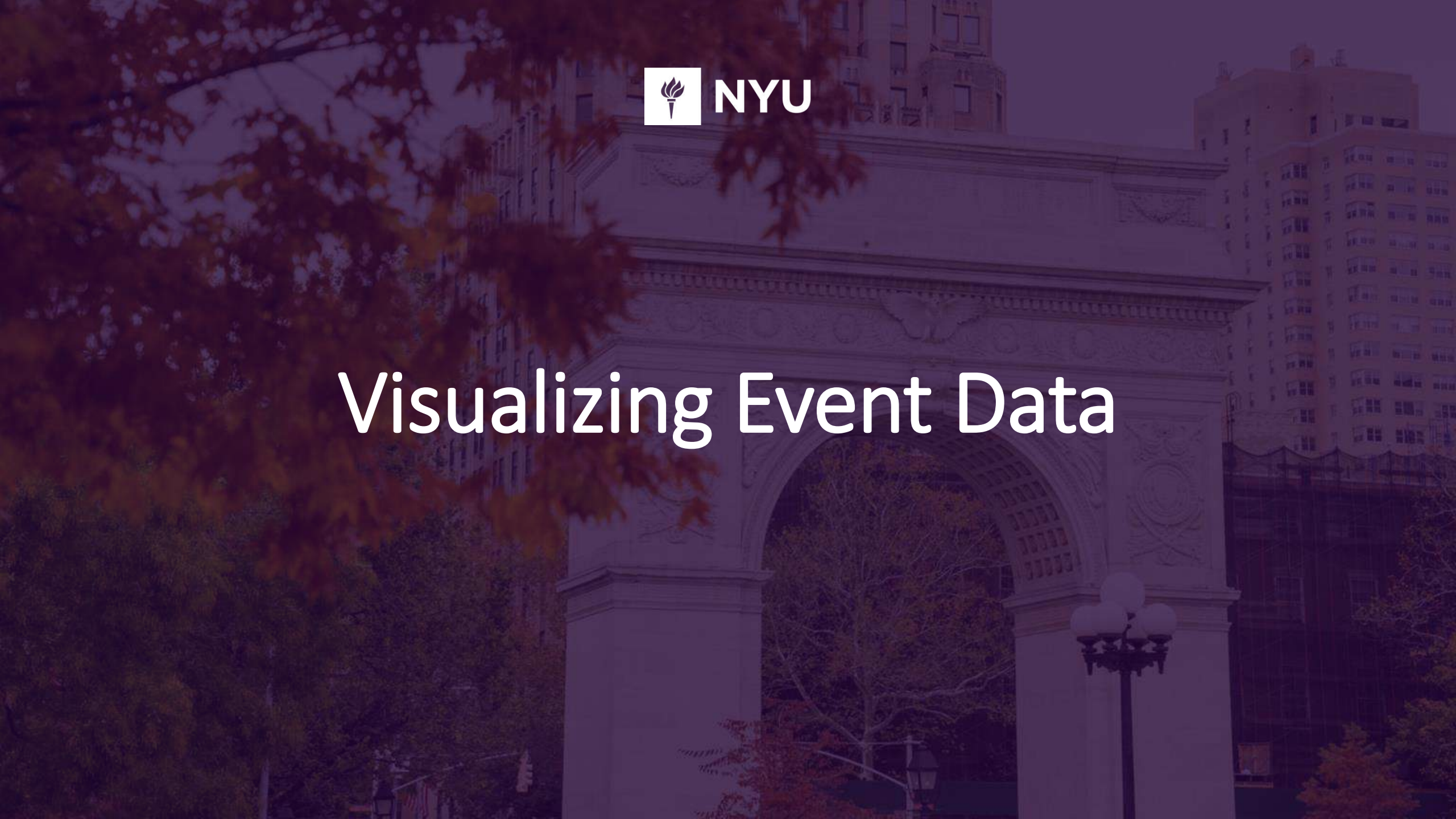


Wall Street Journal: "*Battling Infectious Diseases in the 20th Century: The Impact of Vaccines*"

Heat Maps (Matrix)



Financial Times: The finest goal-scorers of modern European football | Financial Times

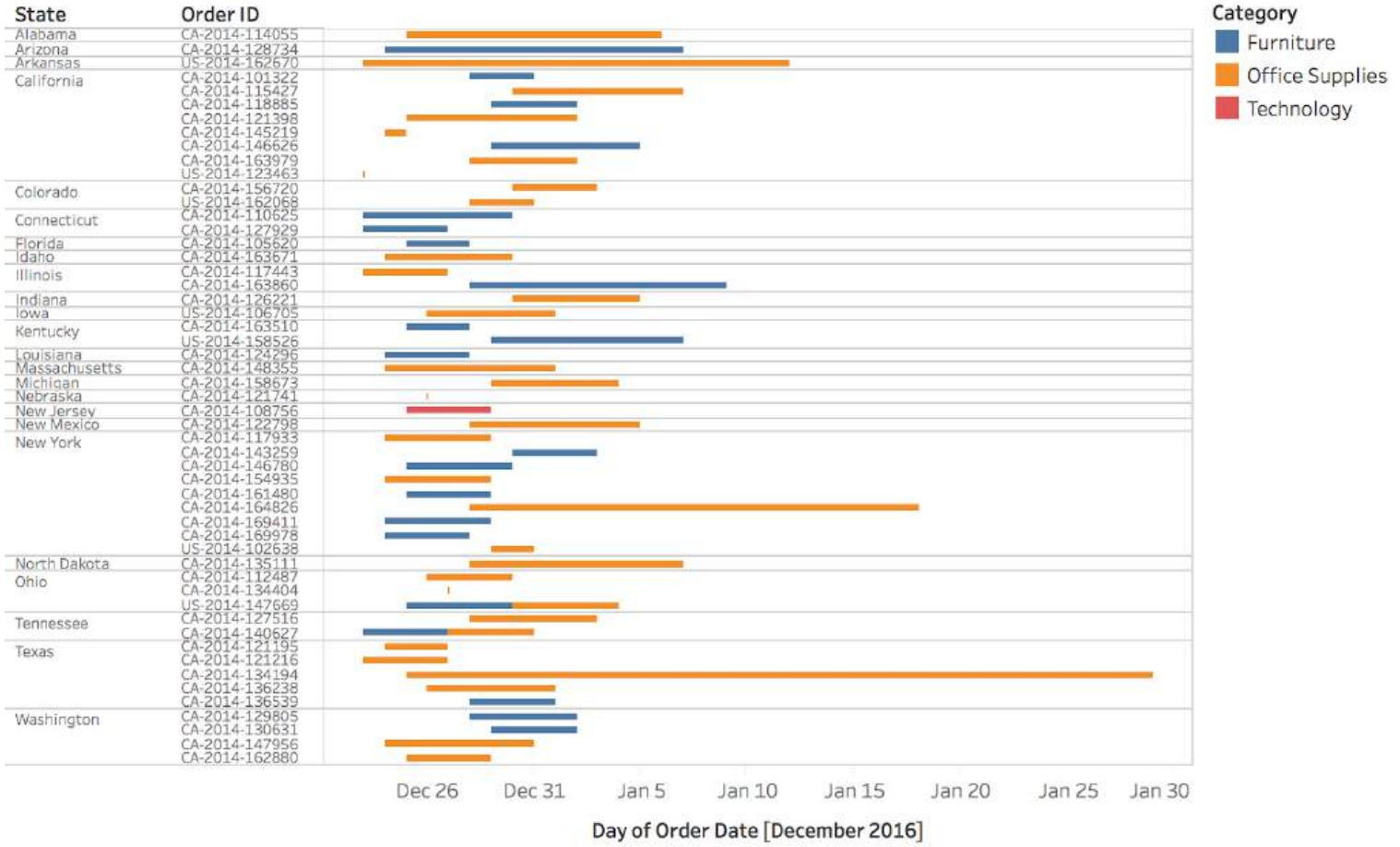


Visualizing Event Data

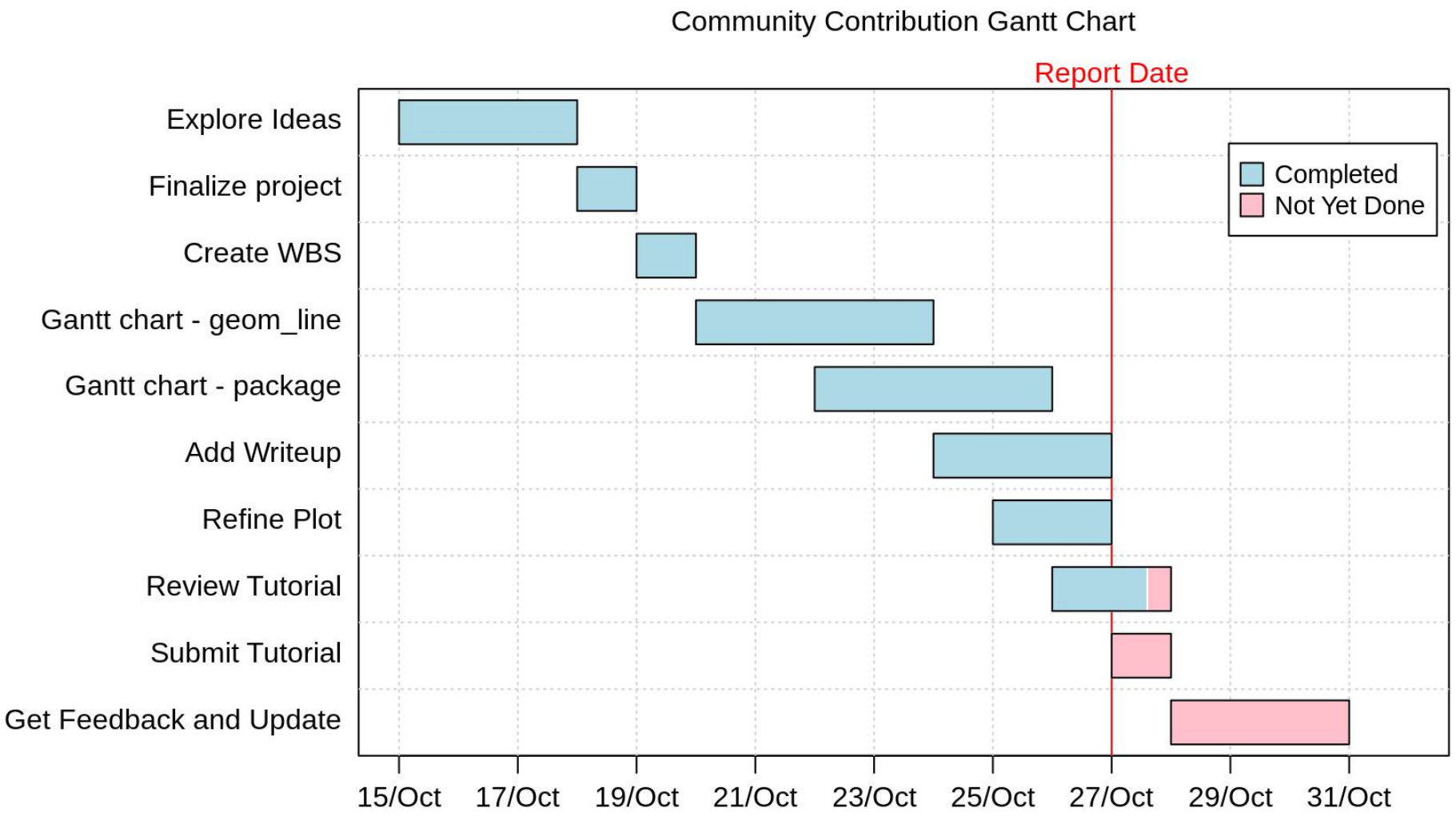
Visualize Event with Duration: Gantt Chart

Henry Gantt

- **Data:** two time attributes (start and end)
- **Marks:**
  - Lines
- **Channels:**
  - Color (category)
  - Position/length (duration)
- **Tasks:**
  - summarize event durations
  - compare events
  - identify intersections/dependencies



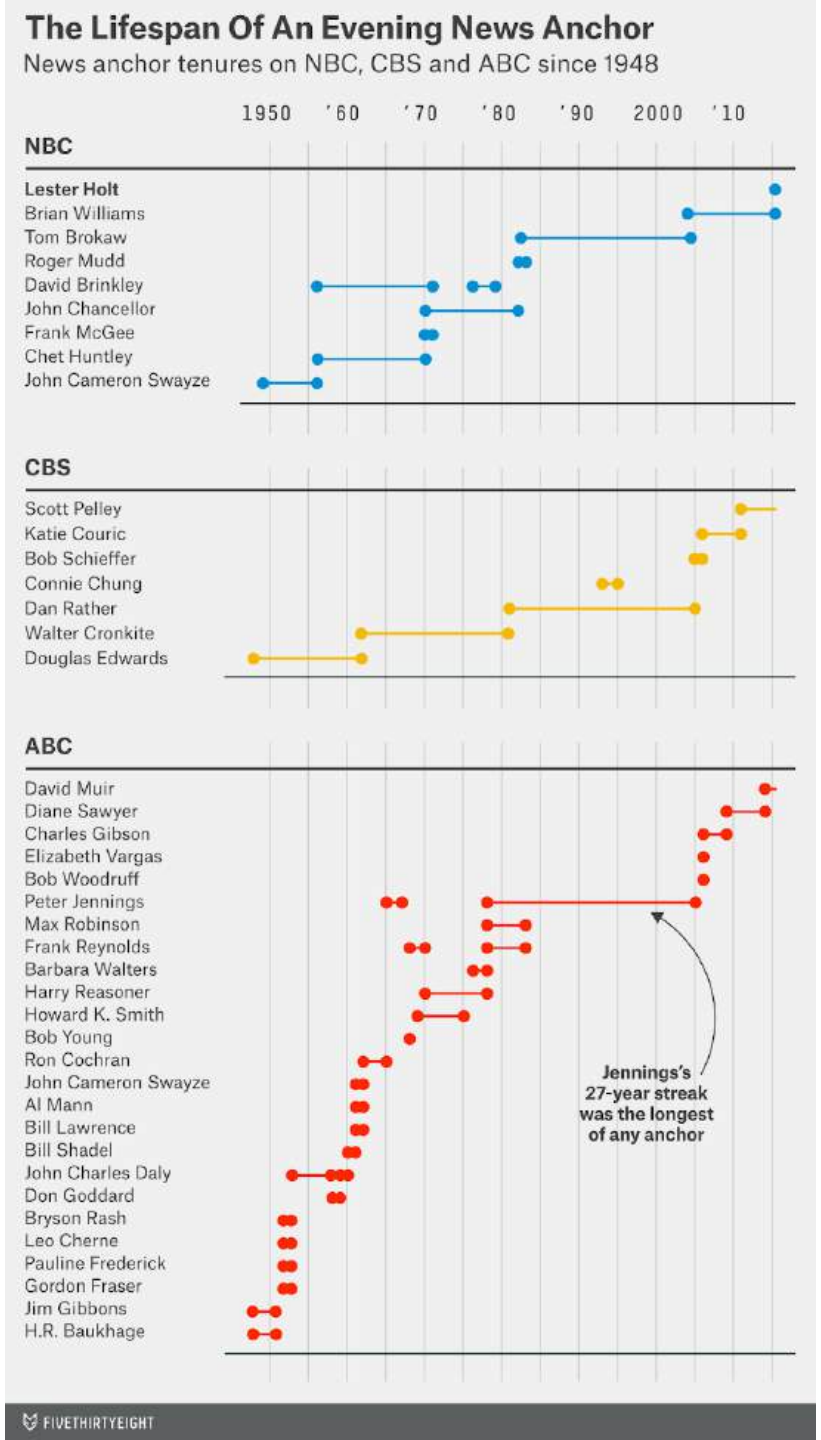
Visualize Event with Duration: Gantt Chart



Source: https://jtr13.github.io/cc19/gantt-charts.html

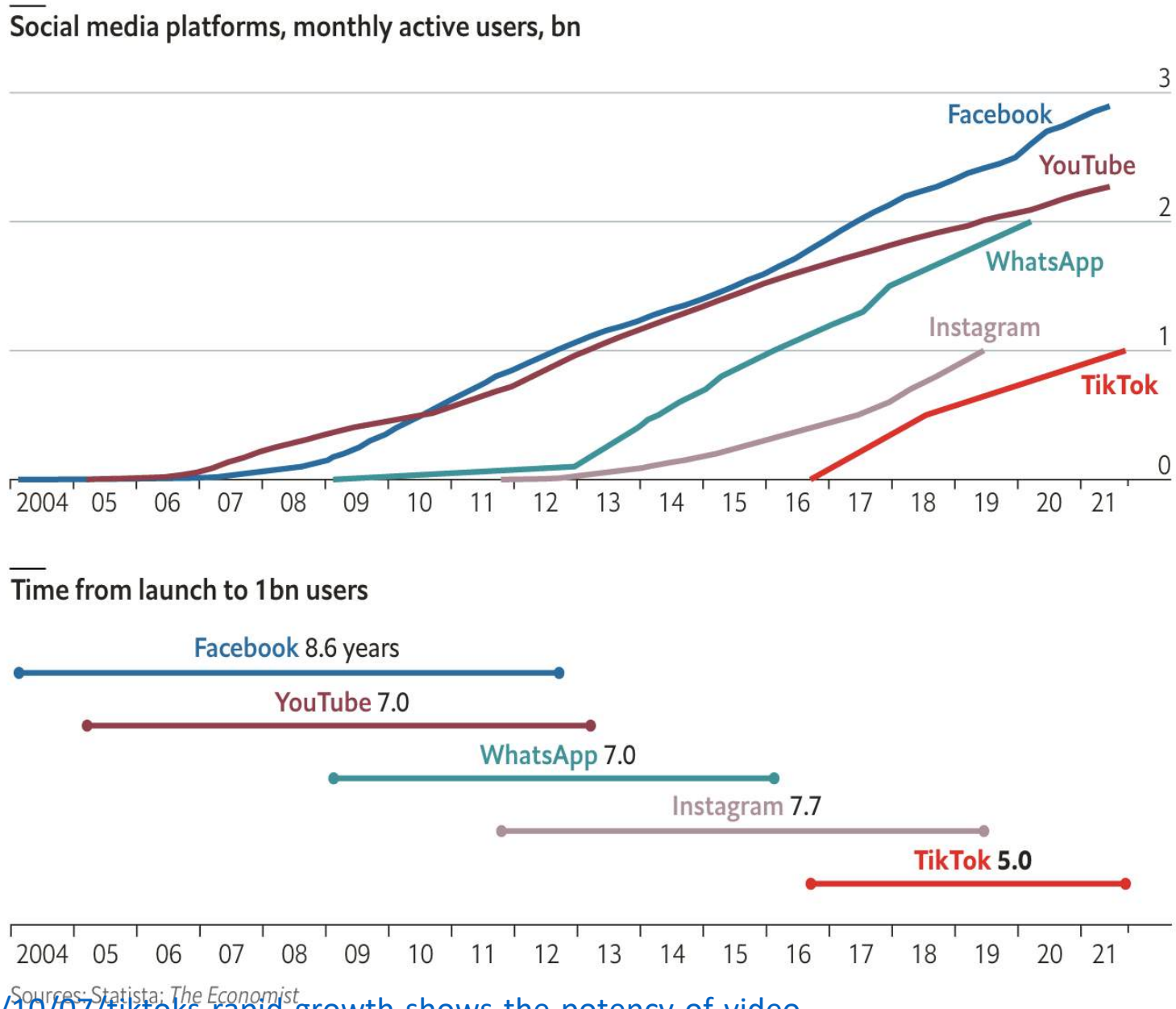
Ranged Dot Plot Lollipop Chart

- Basically a Gantt chart
  - Arguably more visually-appealing
  - Long vs short durations more salient (arguably)
- Source: https://fivethirtyeight.com/features/brian-williams-has-lasted-longer-than-most-in-nbcs-anchor-chair/

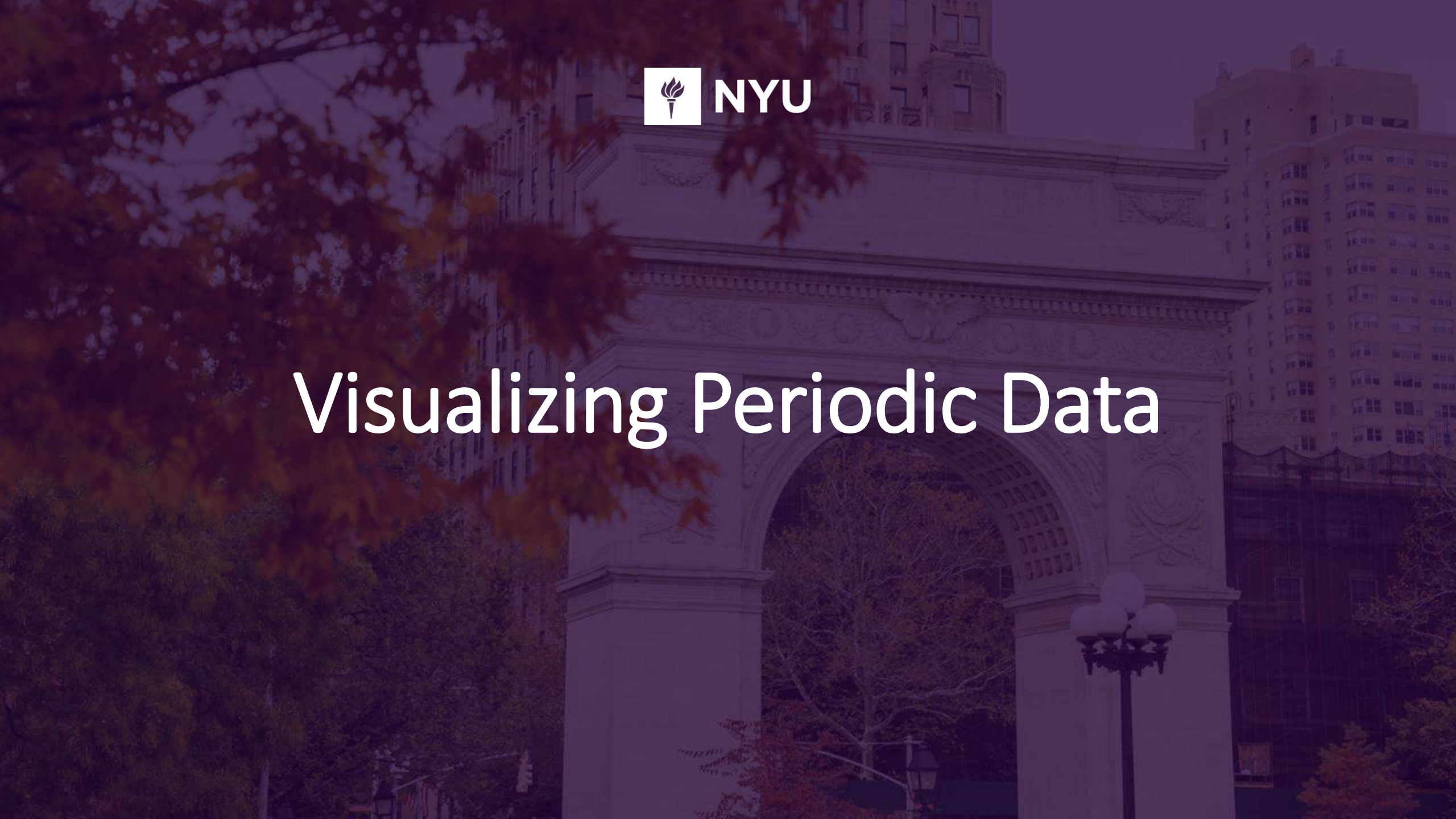


Ranged Dot Plot Lollipop Chart

- **Case Study:** TikTok's rapid growth shows the potency of video
- The Economist Oct 7, 2021



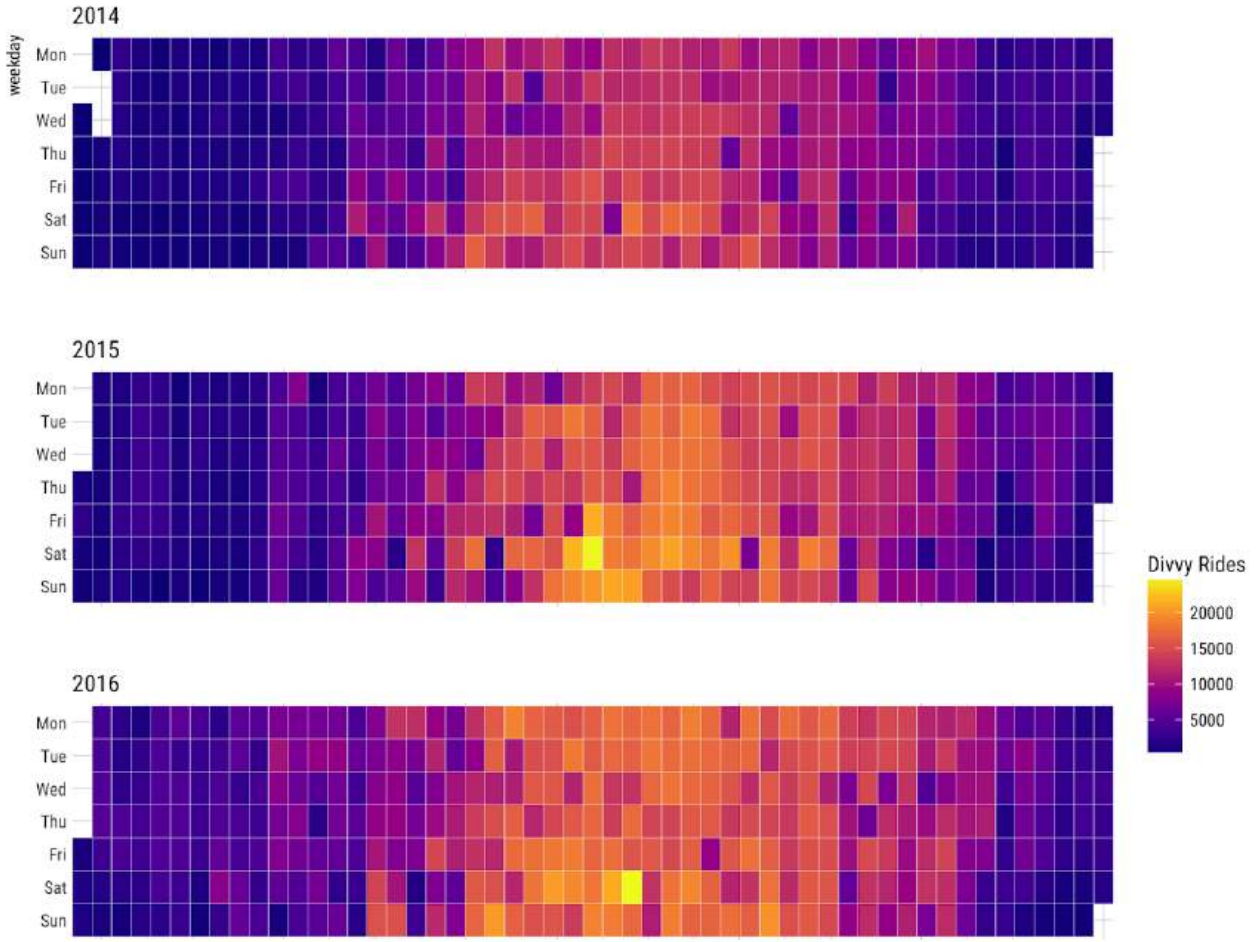
Source: https://www.economist.com/graphic-detail/2021/10/07/tiktoks-rapid-growth-shows-the-potency-of-video



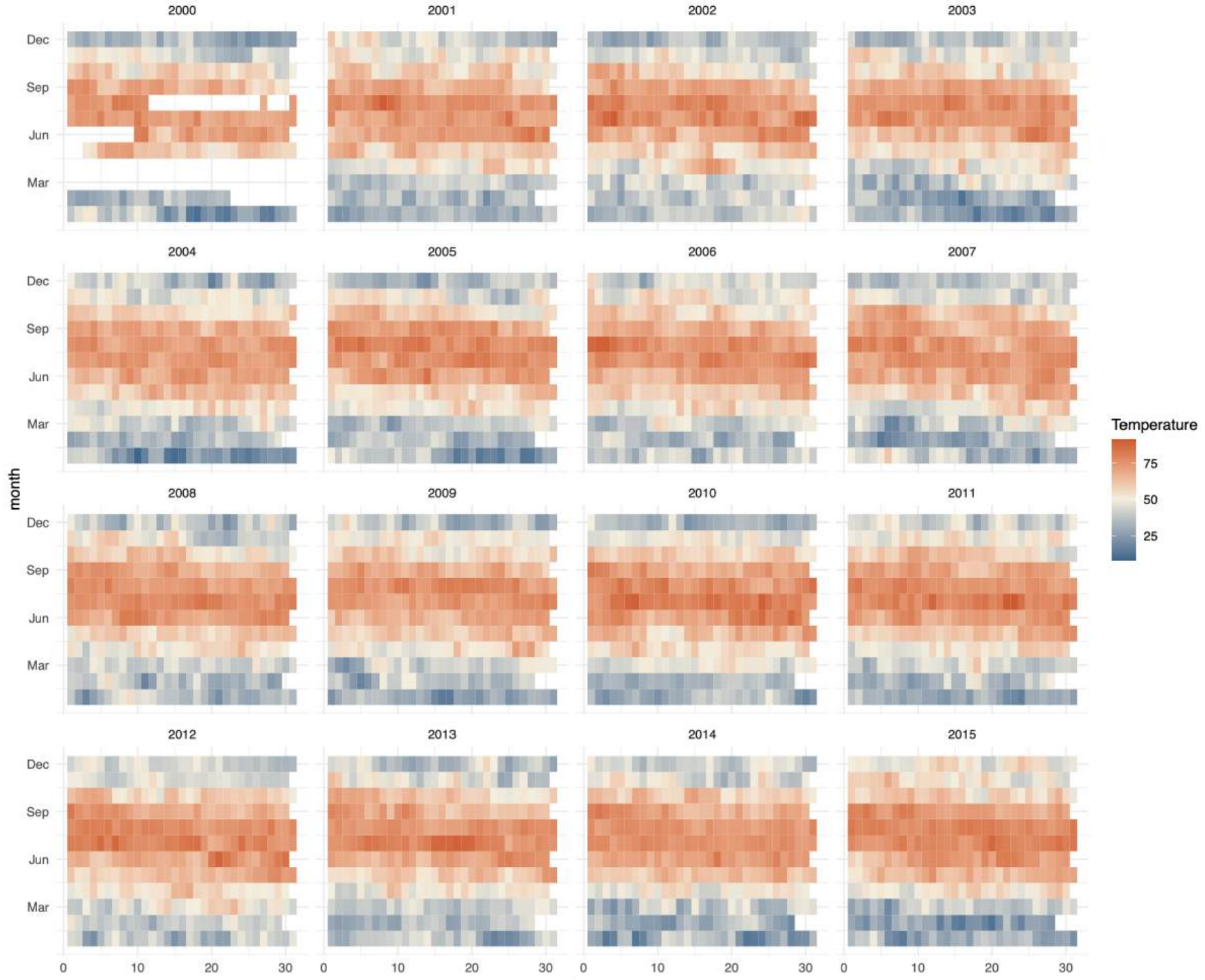
Visualizing Periodic Data

Calendar

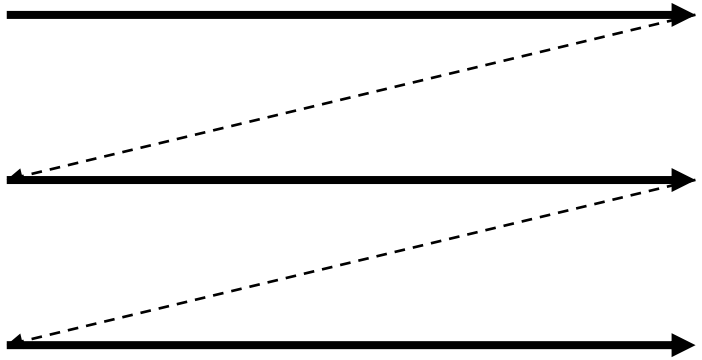
- **Data:** one timeline, dates + quant. value
- **Mark:**
  - Point
- **Channels:**
  - Position (time)
  - Color/size (value)
- **Tasks:**
  - compare trends (by days of the week, month, year)
  - locate outliers
- **Considerations**:
  - natural view for humans
  - focus on time aggregations

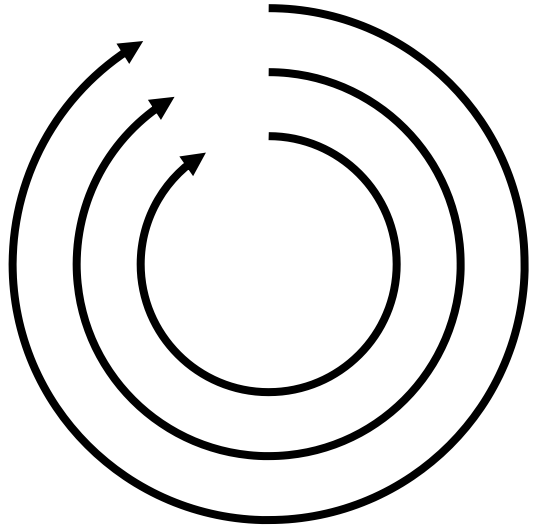


Calendar


2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Dec
Sep
Jun
Mar
month
0
10
20
30
Temperature
75
50
25

Calendar: Radial Layout



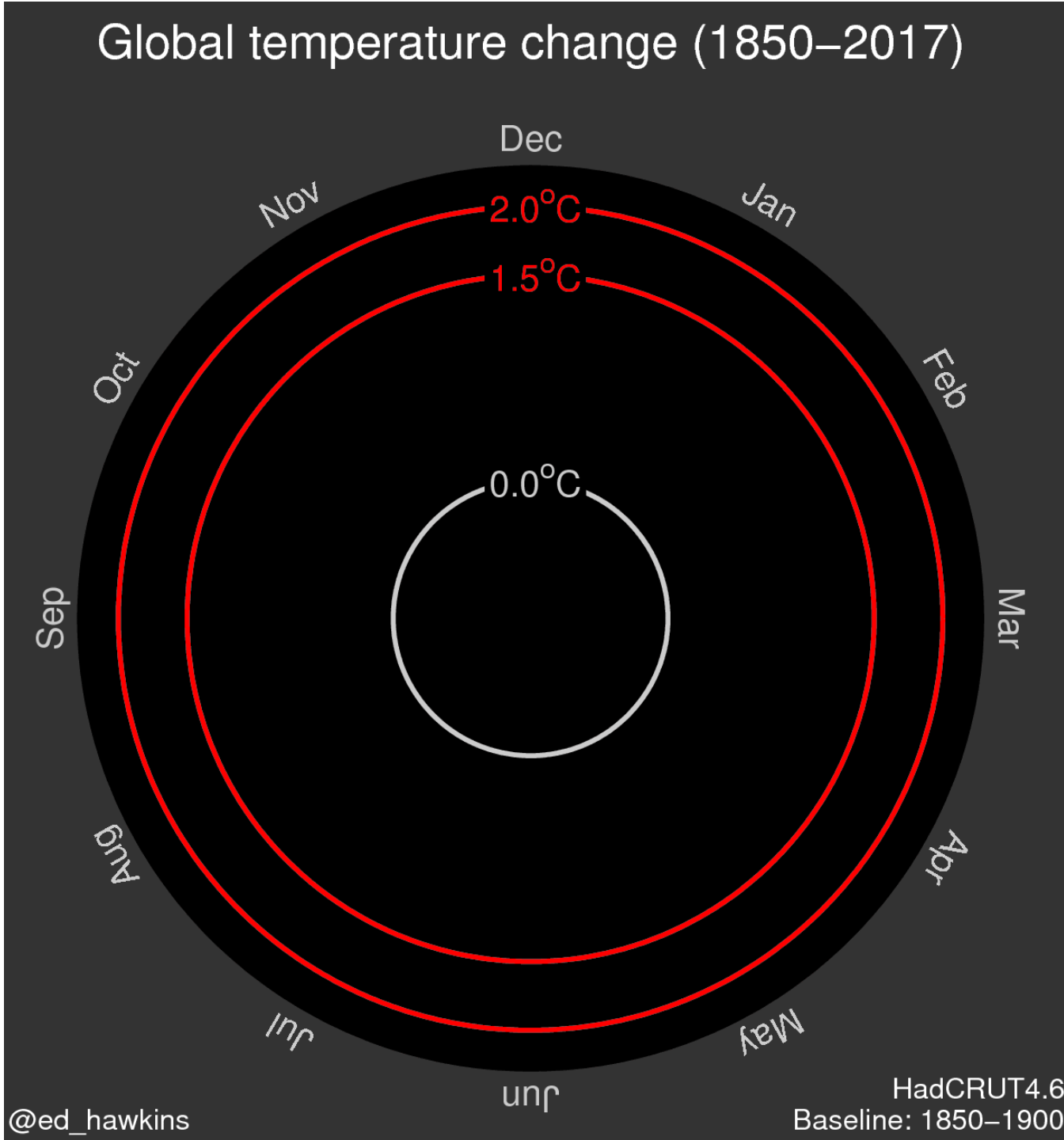


Avoids temporal discontinuity

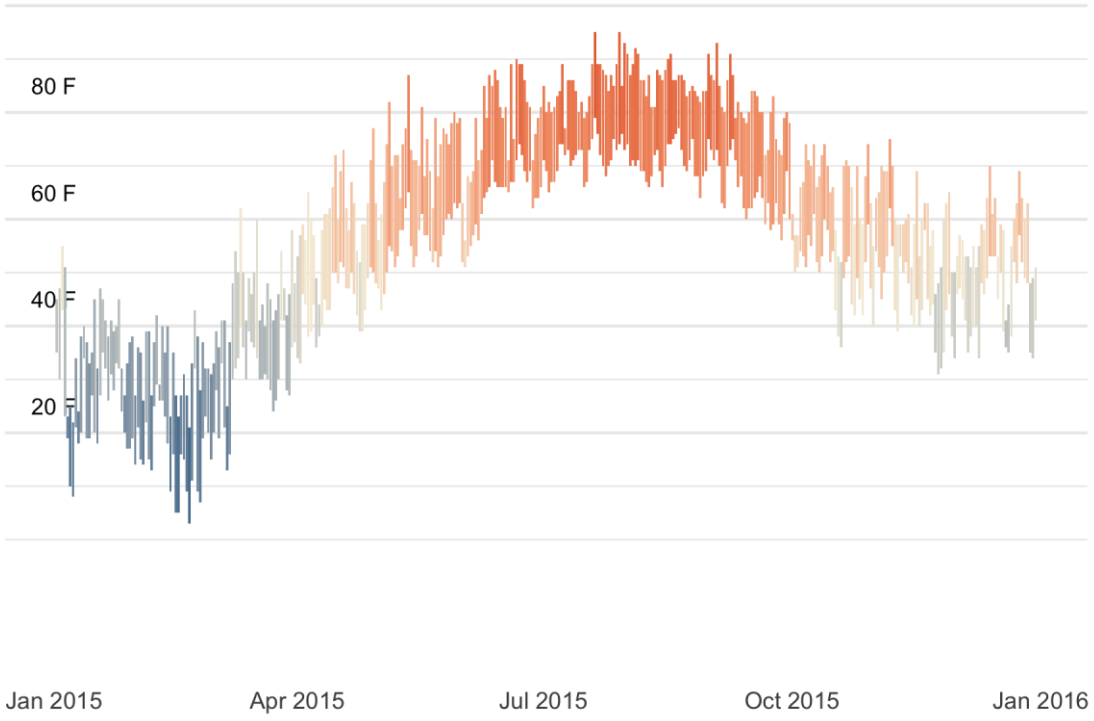
Calendar: Radial Layout

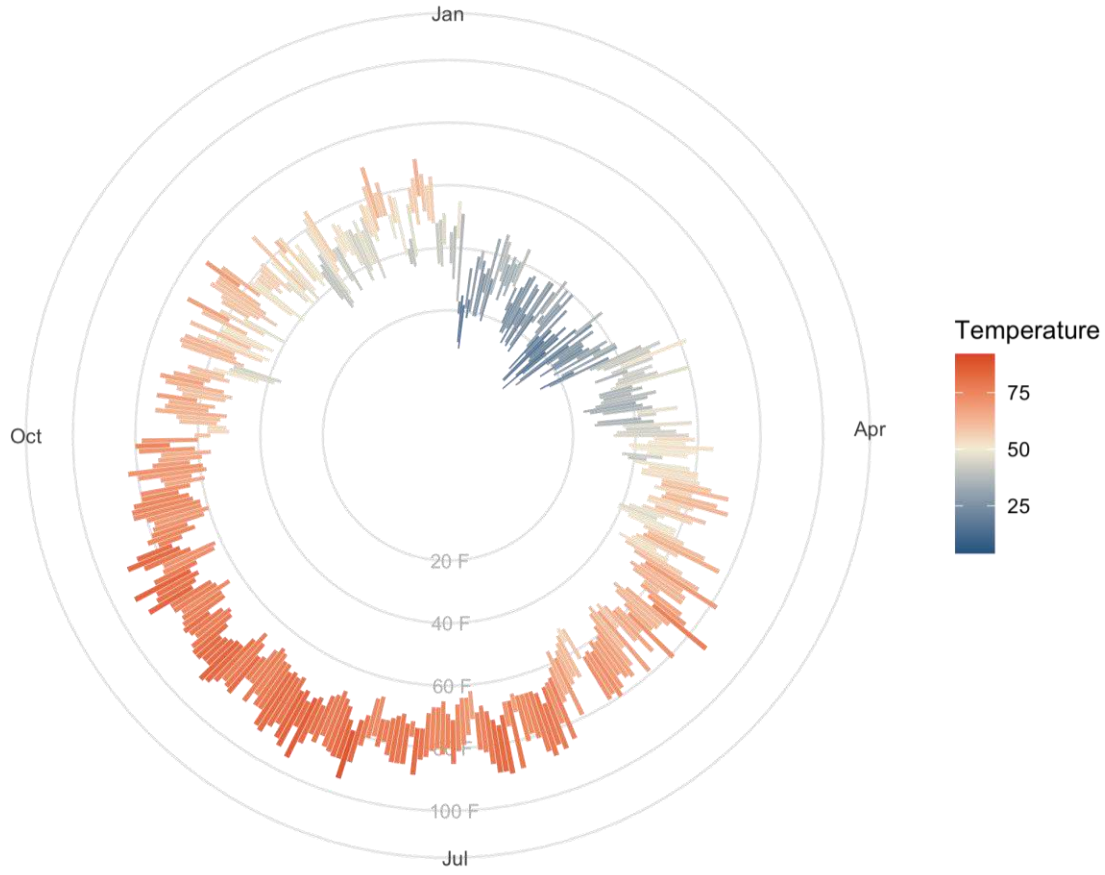
- **Data:** one timeline, dates + quant. value
- **Mark:**
  - line
- **Channels:**
  - Angle/orientation (time)
  - Color/size (value)
  - Position (distance from center) (value)
- **Tasks:**
  - compare trends (by days of the week, month, year)
  - locate outliers
- **Considerations**:
  - Aligns with the cyclic nature of time
  - Space-efficient!

Source: http://www.climate-lab-book.ac.uk/2016/spiralling-global-temperatures/

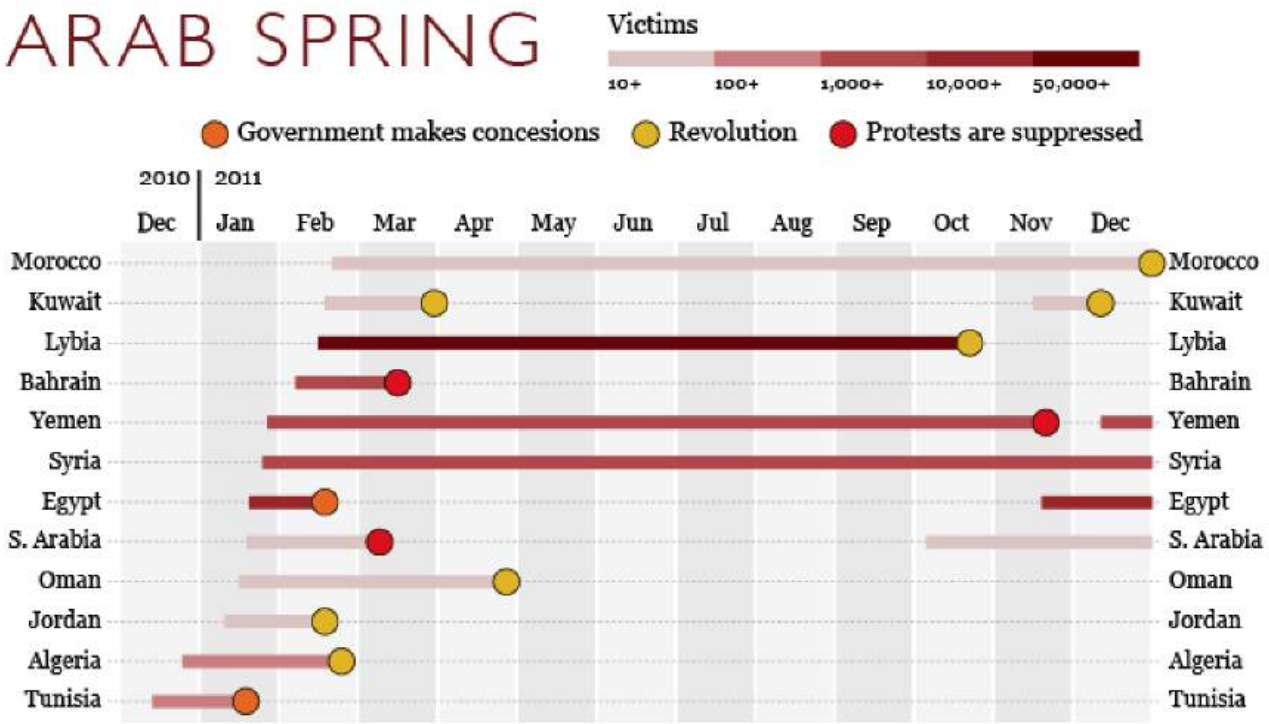


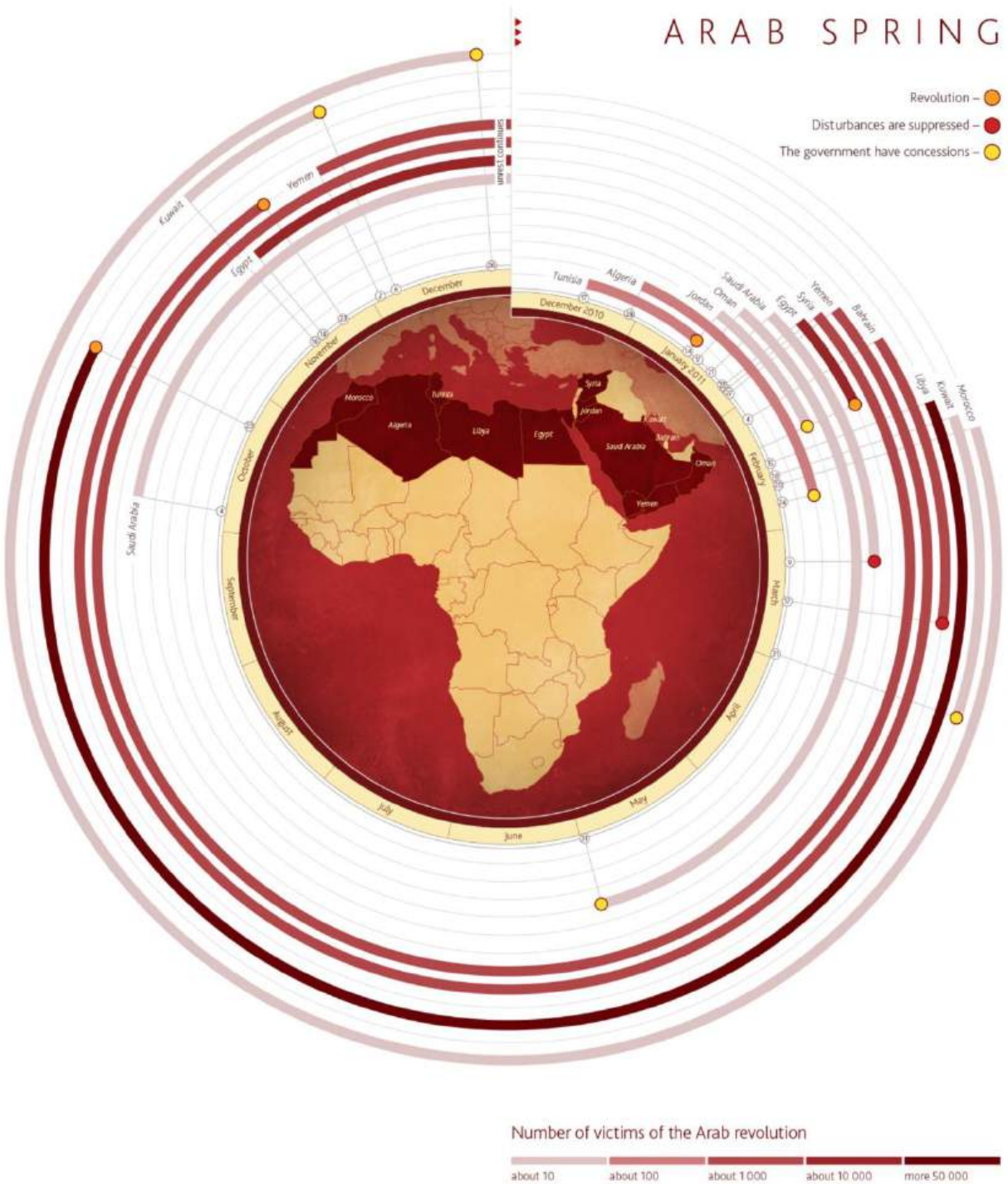
Calendar: Radial Layout



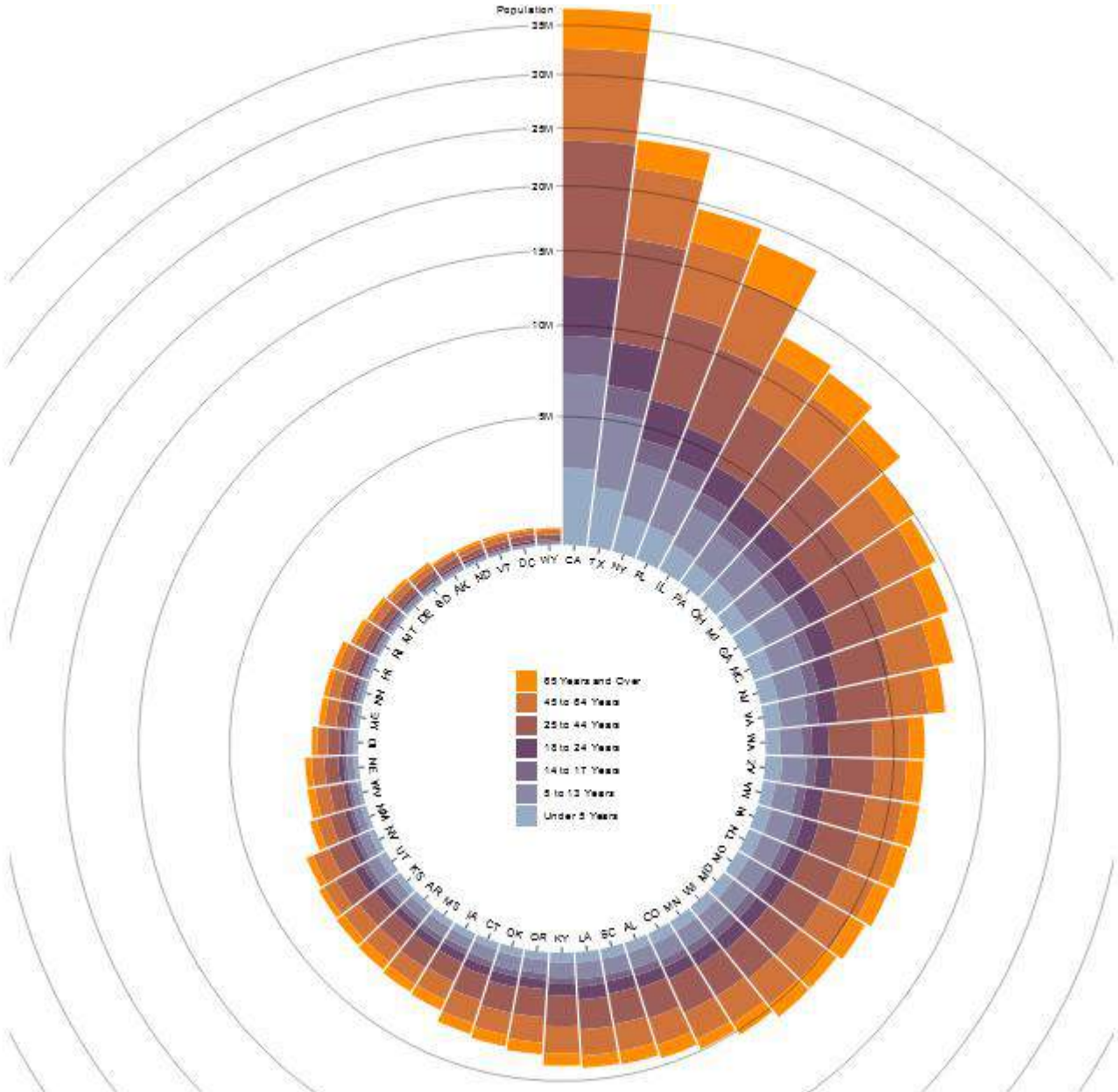


Calendar: Radial Layout





Calendar: Radial Layout (bar chart)



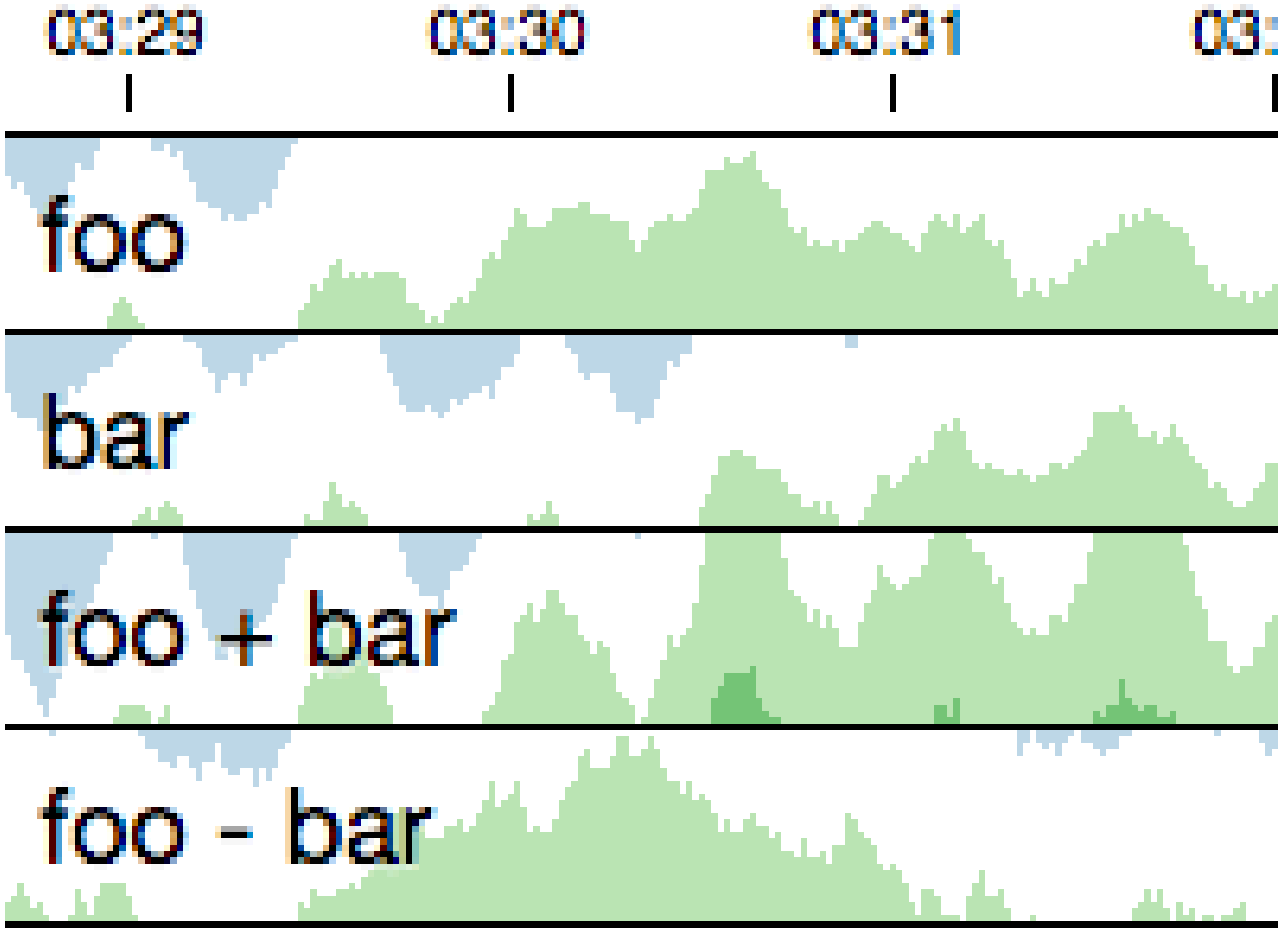
Calendar: Breaking rules

https://public.tableau.com/app/profile/ben.jones/viz/EdgarAllanPoeViz/EdgarAllanPoeViz

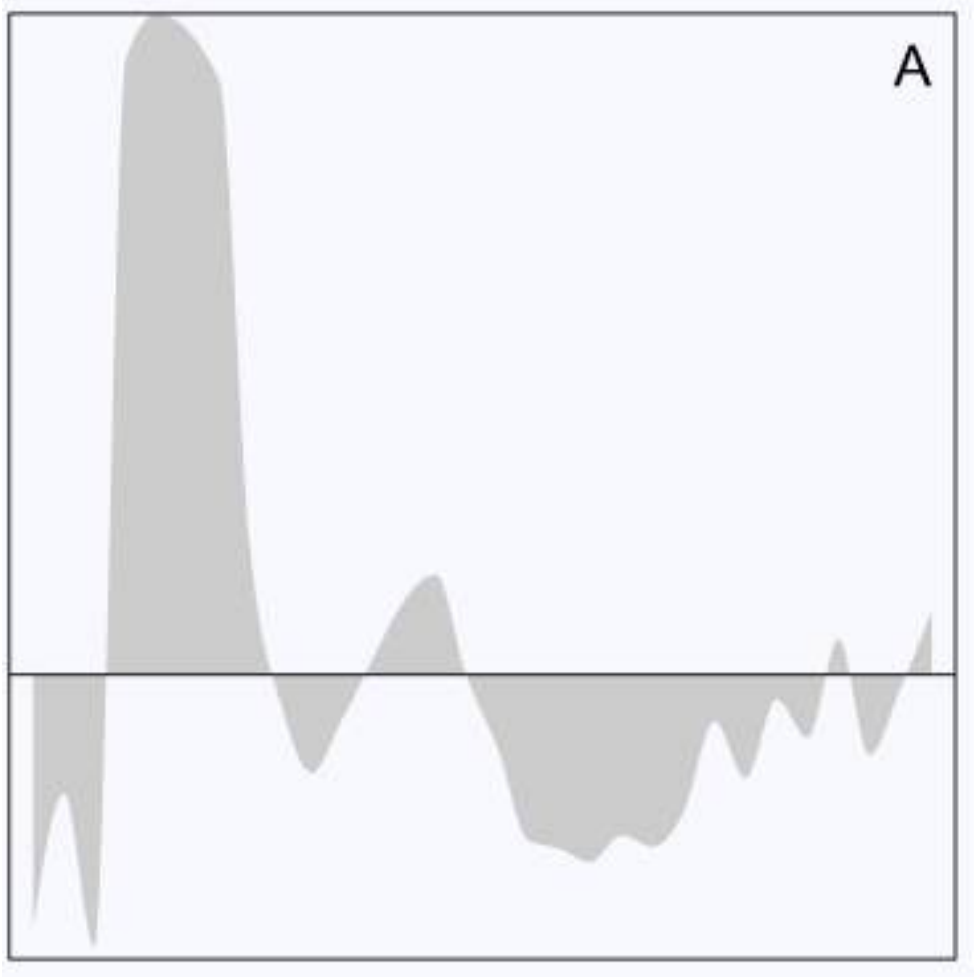
Increase Scalability (space efficiency)

Horizon Chart

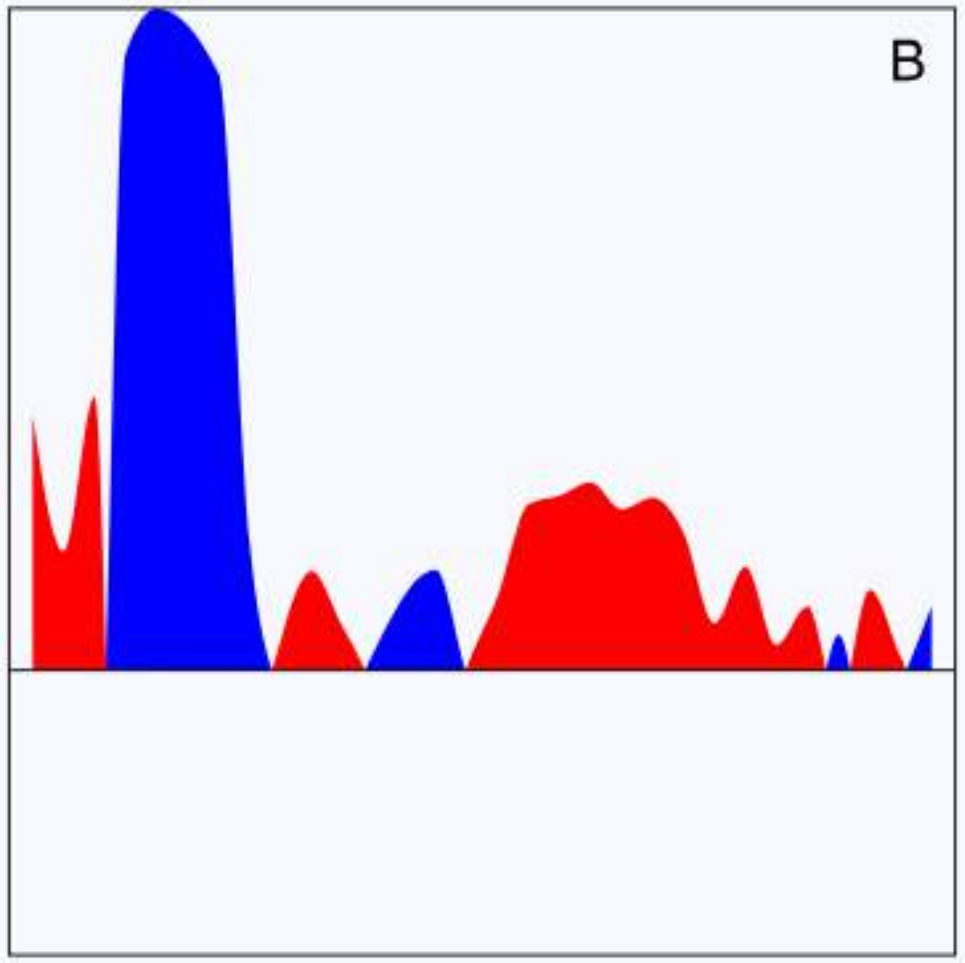
- **Data:** many timeline, dates + quant. values
- **Mark:**
  - Line or areas
- **Channels:**
  - position (time)
  - Position/height (value)
  - Color (value, divergent data)
- **Tasks:**
  - compare trends
  - locate outliers
- **Considerations**:
  - Aligns with the cyclic nature of time
  - Space-efficient!



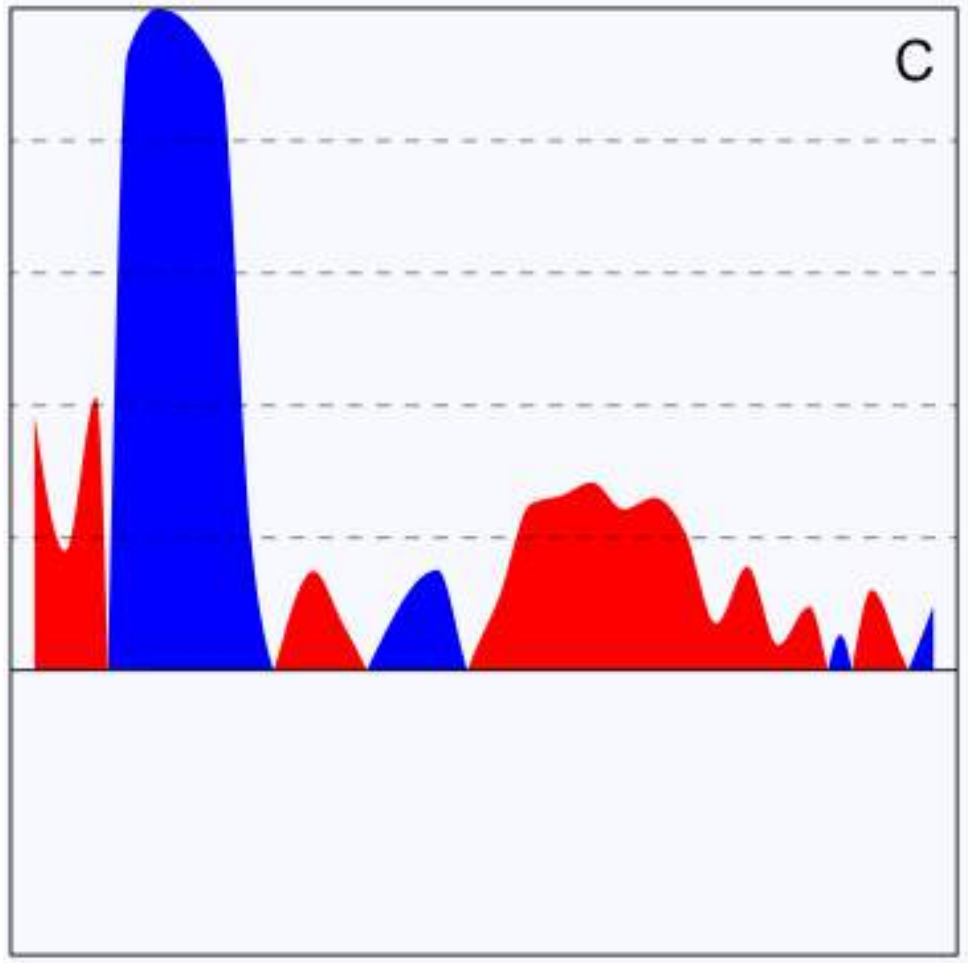
Horizon Chart



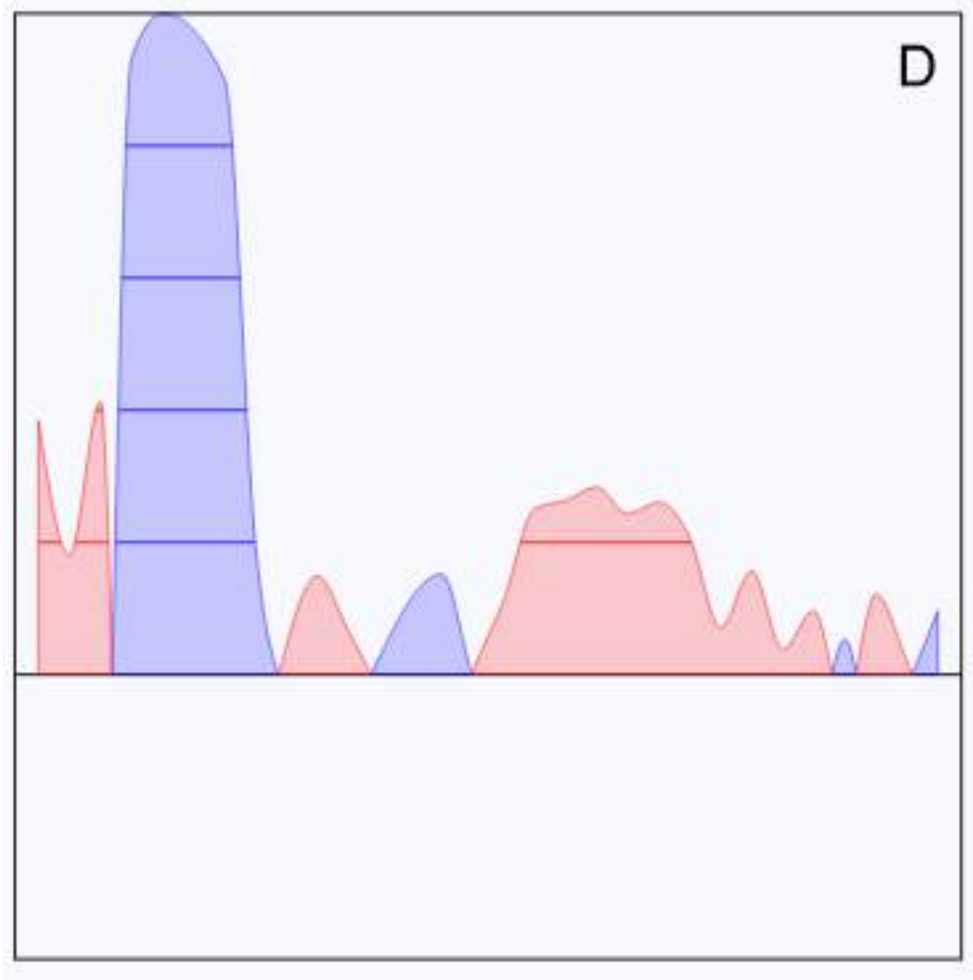
Horizon Chart



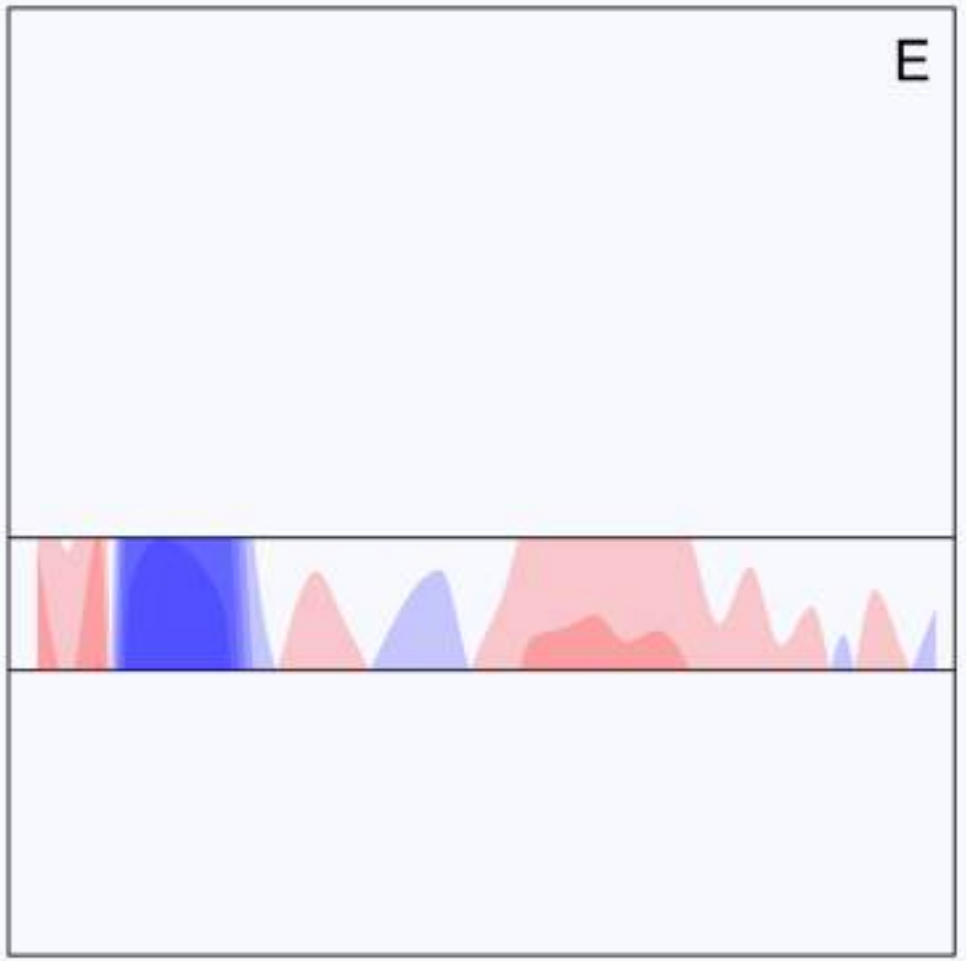
Horizon Chart

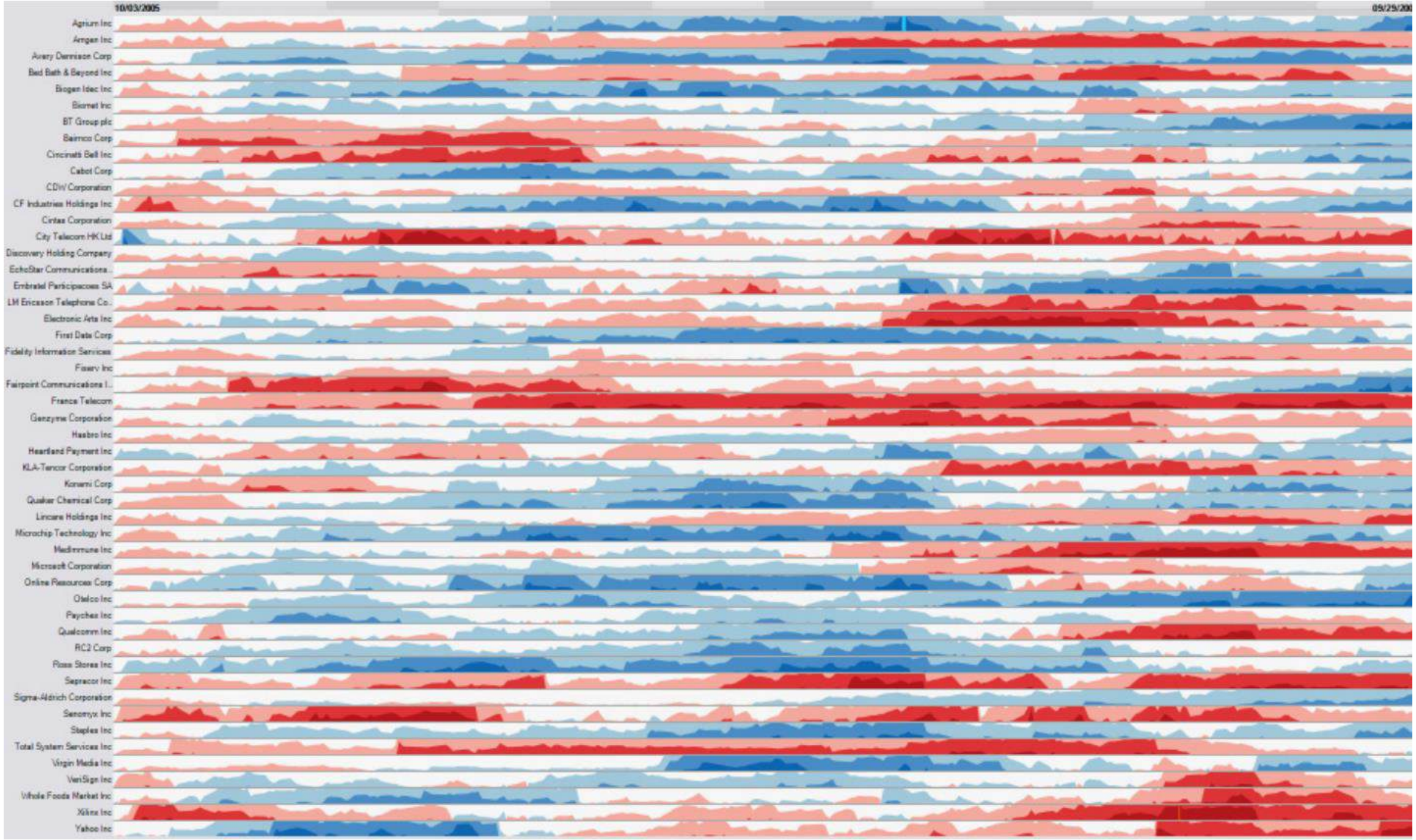


Horizon Chart



Horizon Chart



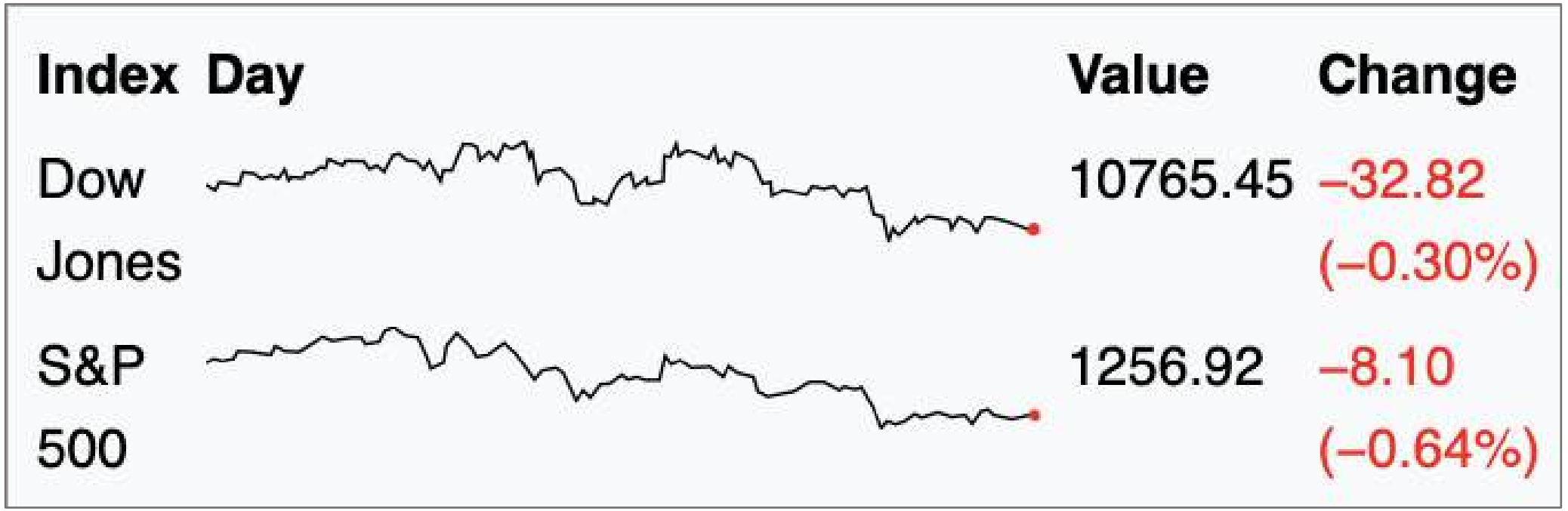


Source: https://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf

Sparklines

- Very small line charts → good for trends, ignores details on purpose → memorable
- *"Sparklines are small, intense, word-sized graphics with typographic resolution. Sparklines are can be placed anywhere that words or numbers or graphics can be placed: in sentences, maps, graphics, tables."*

  --Edward Tufte (History of Sparklines)



Sparklines

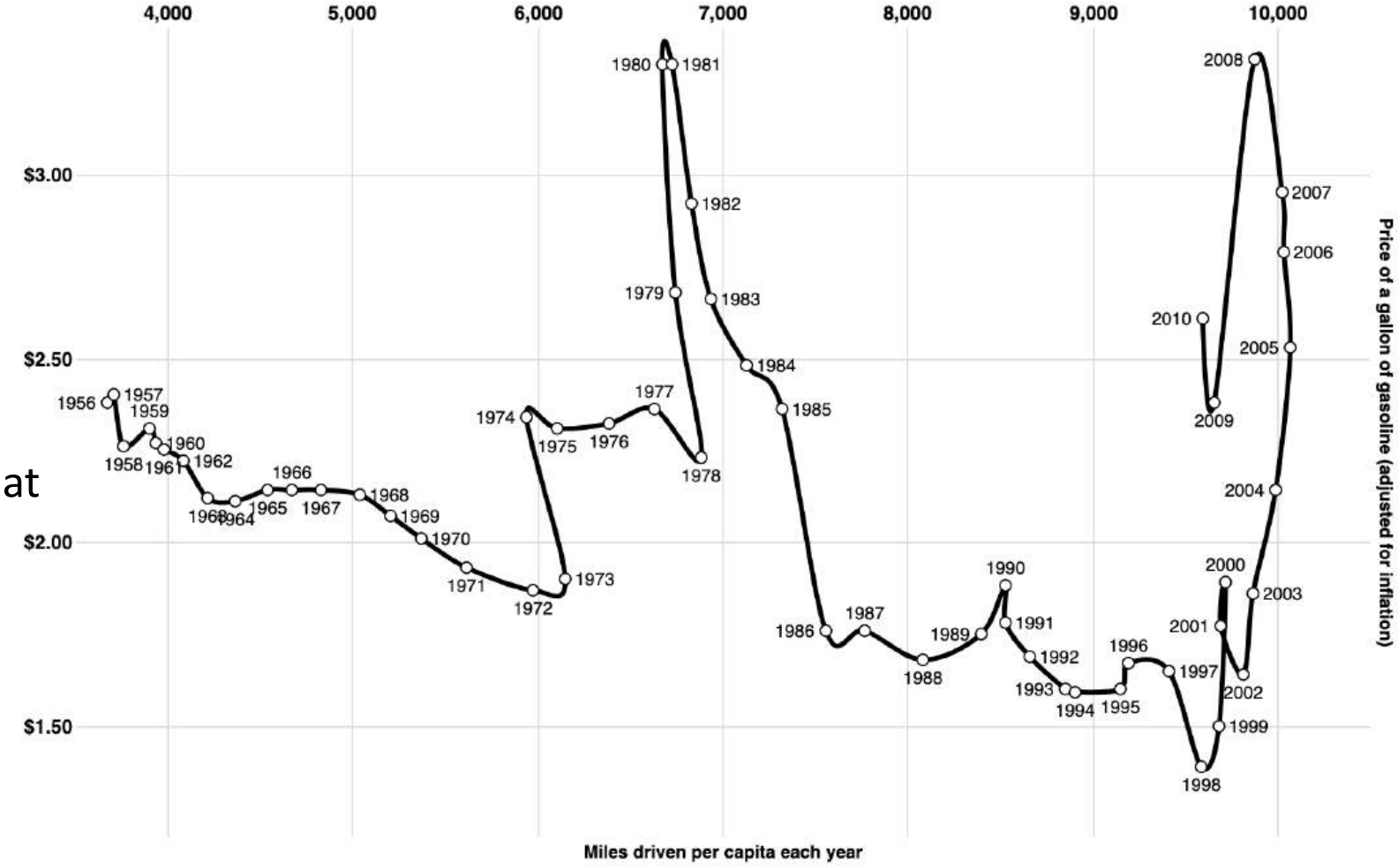
- *“Sparklines are small, intense, word-sized graphics with typographic resolution. Sparklines are can be placed anywhere that words or numbers or graphics can be placed: in sentences, maps, graphics, tables.”*

  --Edward Tufte (History of Sparklines)

Other Techniques

Connected Scatter Plot

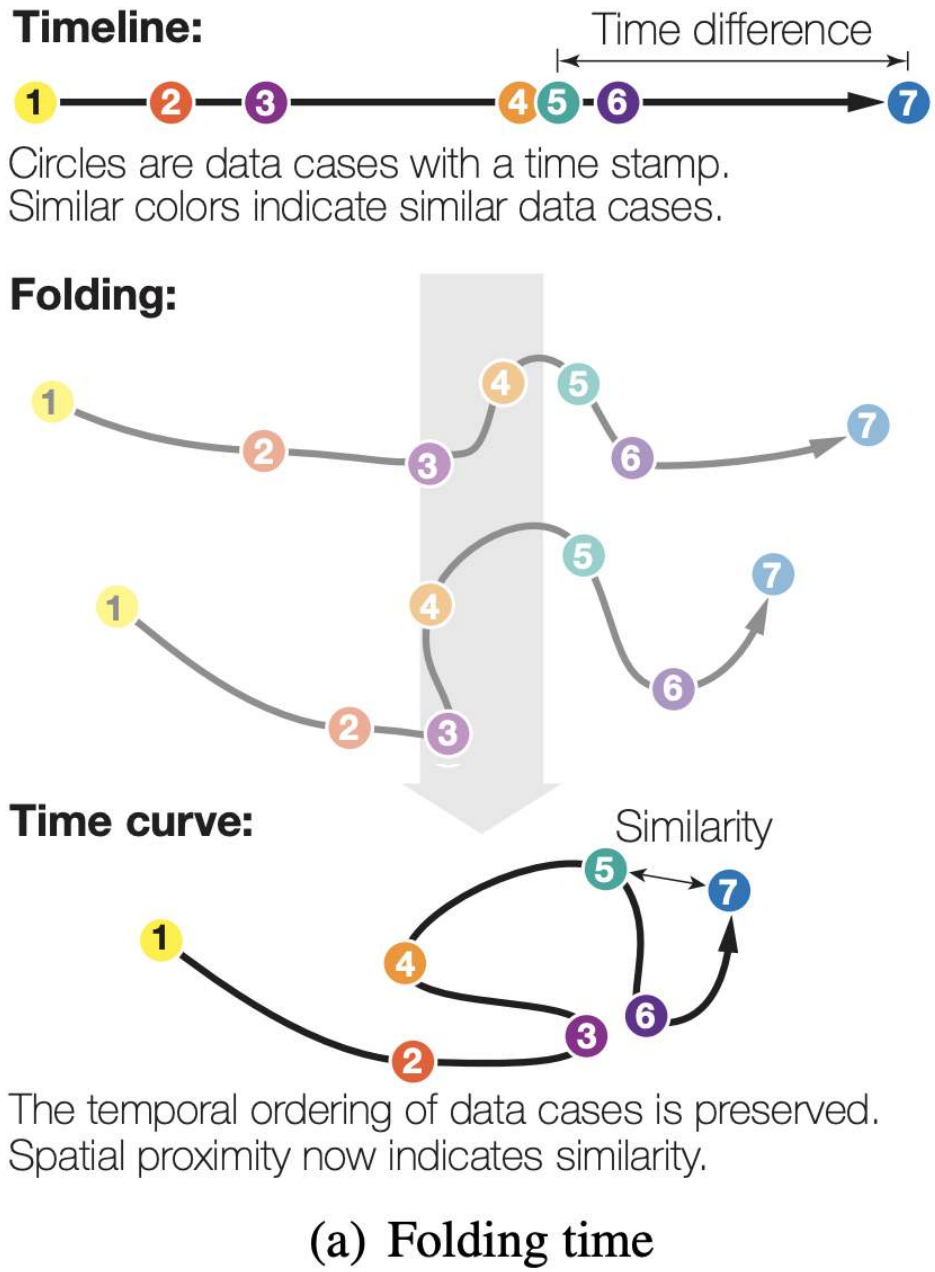
- **Data:** events (time + object)
- **Mark:**
  - Points (objects)
  - Lines (flow of time)
- **Channels:**
  - Position (value)
- **Considerations**:
  - Engaging and interesting to look at
  - Empirical study: correlations difficult to understand



Source: https://vega.github.io/vega/examples/connected-scatter-plot/

Connected Scatter Plot: Time Curves

B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski and P. Dragicevic, "Time Curves: Folding Time to Visualize Patterns of Temporal Evolution in Data," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 559-568, 31 Jan. 2016, doi: 10.1109/TVCG.2015.2467851.



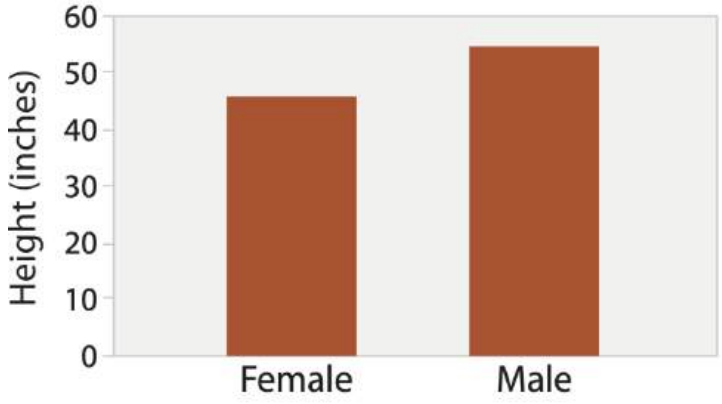
(a) Folding time

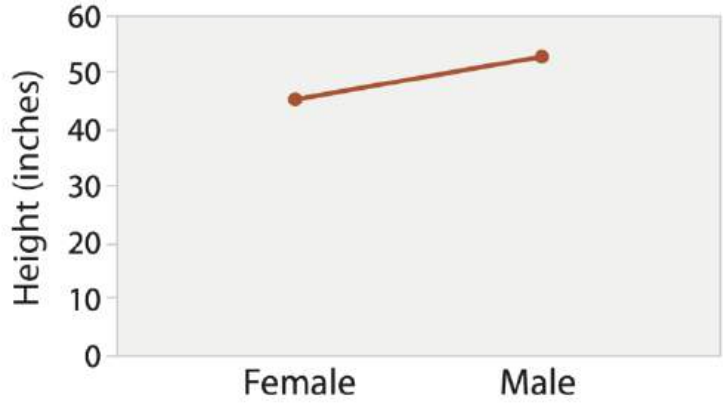
Source: https://vcg.informatik.uni-rostock.de/~ct/timeviz/timeviz.html

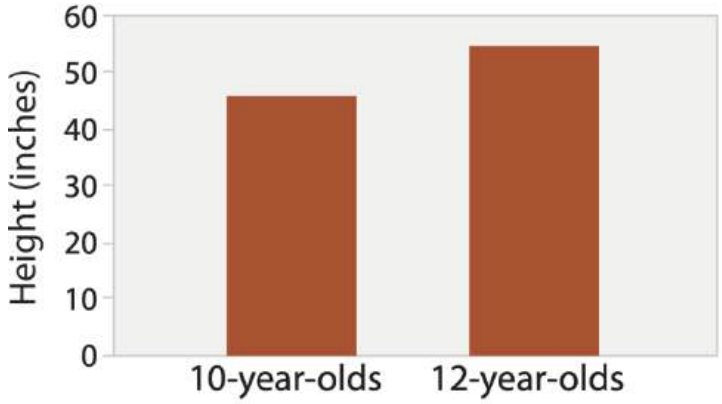
Other Tips

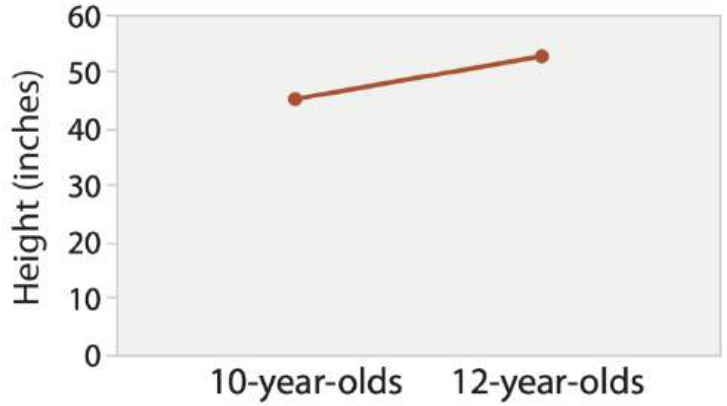
Bar charts vs line charts

- Depends on type of key attribute
  - Bar charts if categorical
  - Line charts if ordered
- Do not use line charts for categorical key attributes
  - Violates expressiveness principle
  - Implication of trend so strong that it overrides semantics!









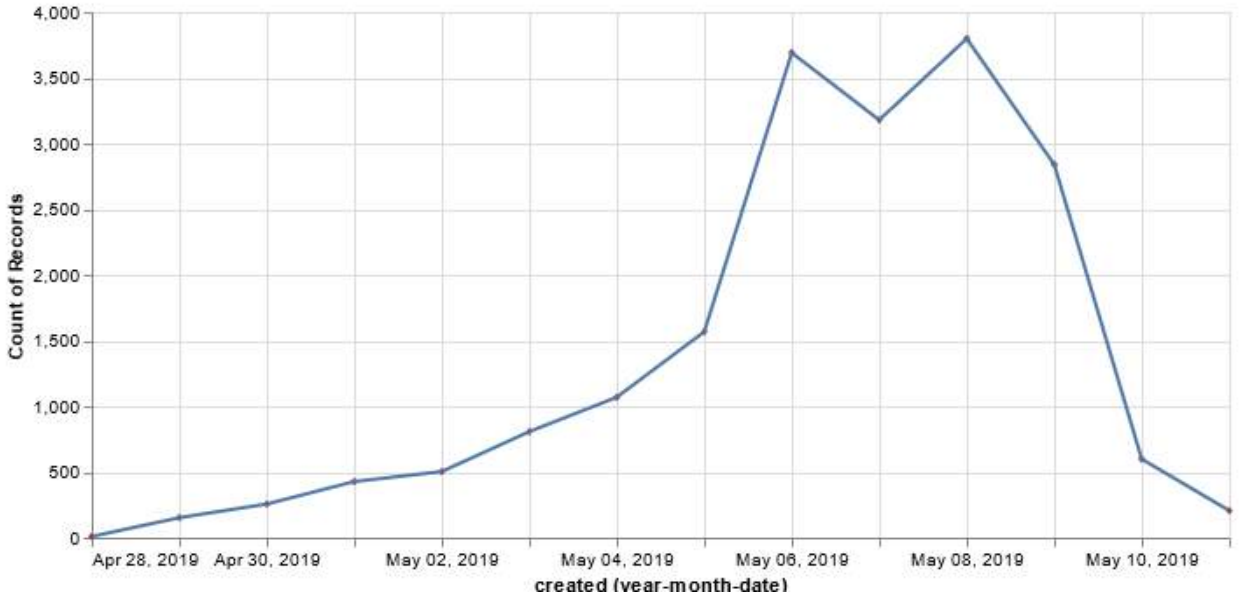
Truncate dates

- Quick way to aggregate data:
  - https://observablehq.com/@berkeleyvis/time-aggregation

Date Truncate

One way of aggregating is taking a part of the date, this is truncating the date to a certain level. Truncate by: *yearmonthdate*

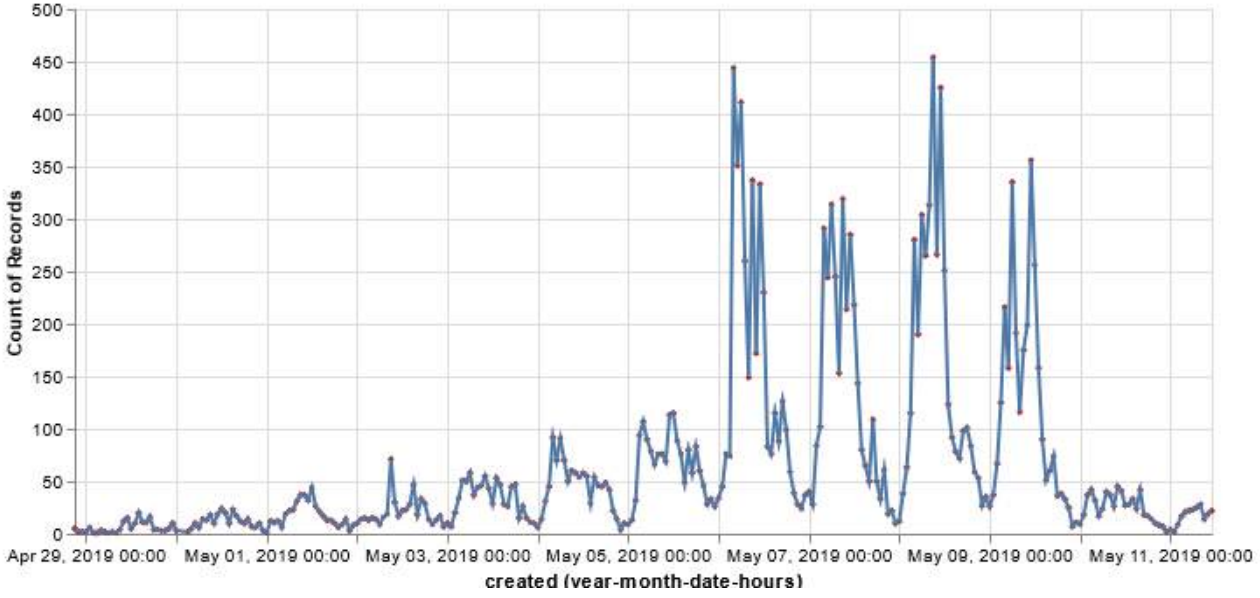
Notice how this reveals sleep patterns of tweeters

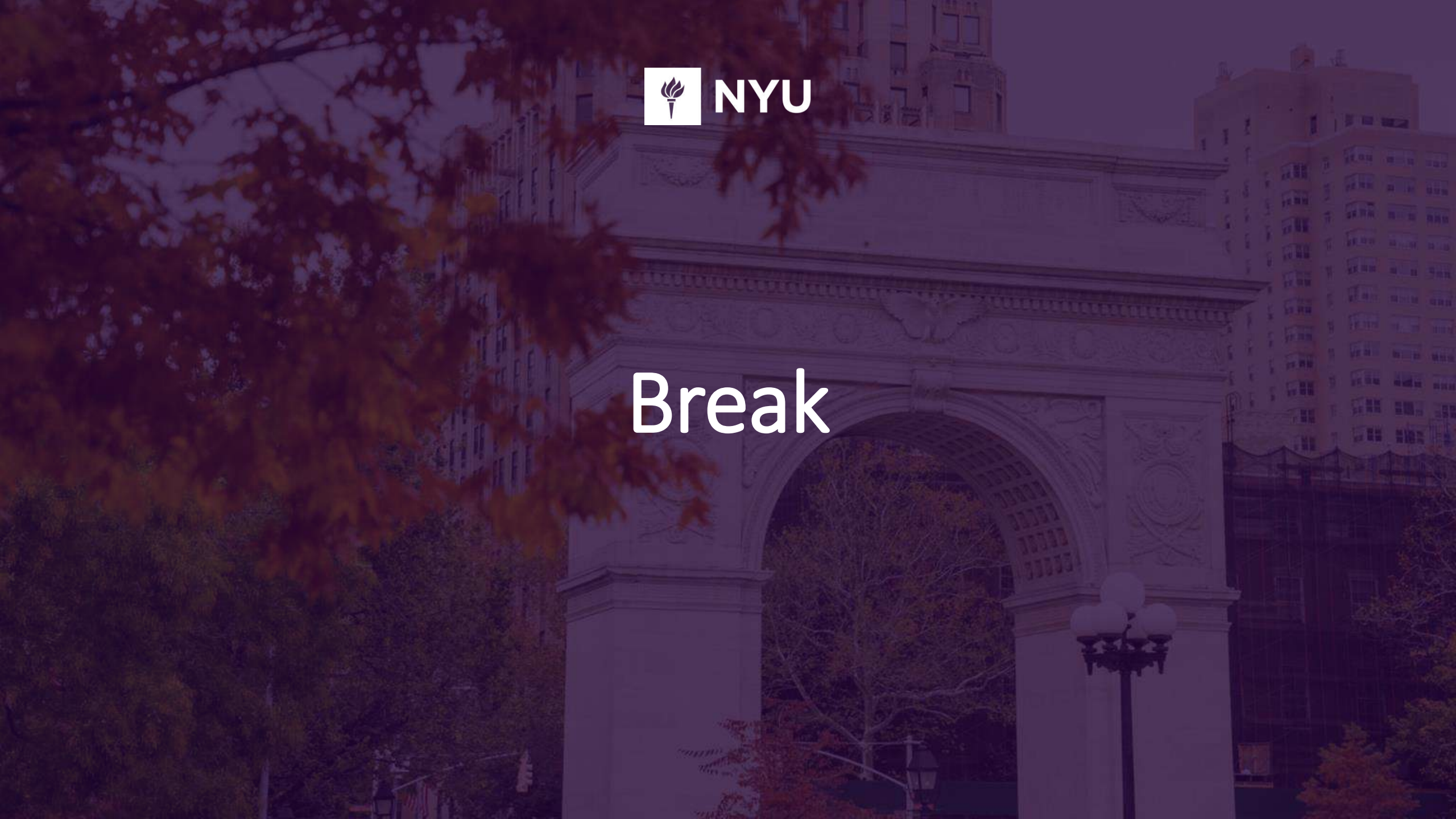


Date Truncate

One way of aggregating is taking a part of the date, this is truncating the date to a certain level. Truncate by: *yearmonthdatehours*

Notice how this reveals sleep patterns of tweeters

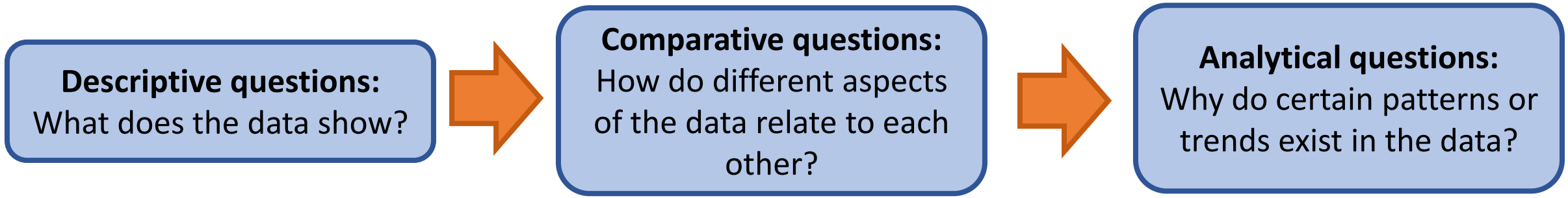




Break

Asking good questions for viz

- Ask questions with increasing complexity



Asking good questions for viz

- Who, What, When, Where, Why, and How
  - Who is involved or affected?
  - What is happening?
  - When did it occur or change?
  - Where is this taking place?
  - Why is this happening?
  - How is it evolving or impacting other factors?

Asking good questions for viz

- Explore multiple dimensions of the data
  - Requires a dataset with enough complexity!

- How do different variables interact?
- Are there any unexpected correlations?
- How do trends change when data is segmented or grouped differently?

Asking good questions for viz

- Consider context:
  - What external factors might influence the data?
  - What are the potential implications of the patterns observed?
  - How does this data relate to real-world issues or decisions?

Asking good questions for viz

- Consider interesting stories:
  - Who would be most interested in or affected by these insights?
  - How could this data challenge common assumptions?

Asking good questions for viz

- Consider practical applications:
  - How could insights from this data be used to solve problems?
  - What decisions could be informed by this analysis?
  - How might different stakeholders interpret or use this information?