

VISUALIZATION TECHNIQUES

Pere-Pau Vázquez

Dept. Computer Science – UPC

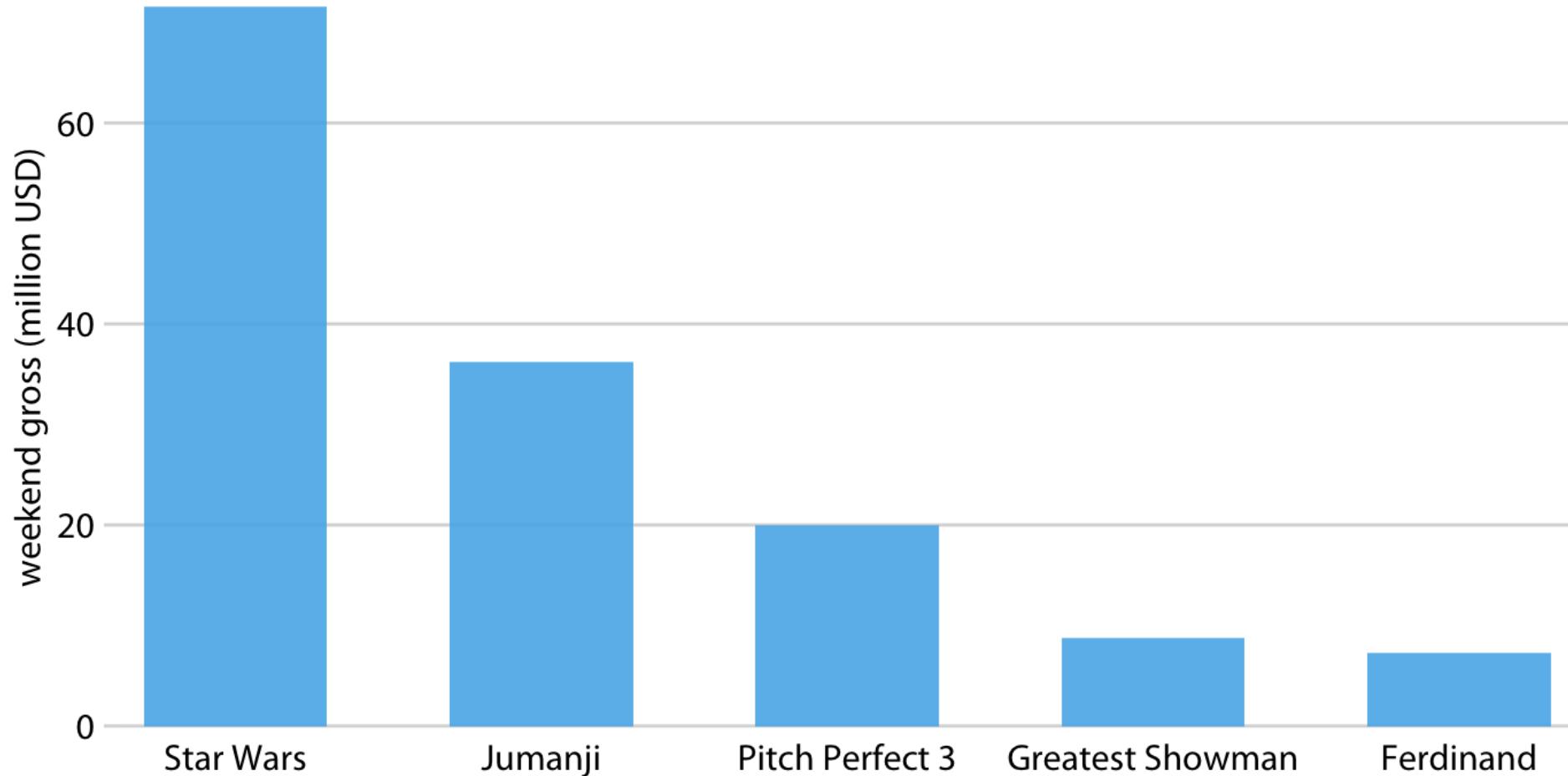
OUTLINE

- Displaying quantities
- Displaying distributions
- Displaying proportions
- Displaying relationships
- Displaying time series
- Displaying geospatial data
- Other charts
- Uncertainty

OUTLINE

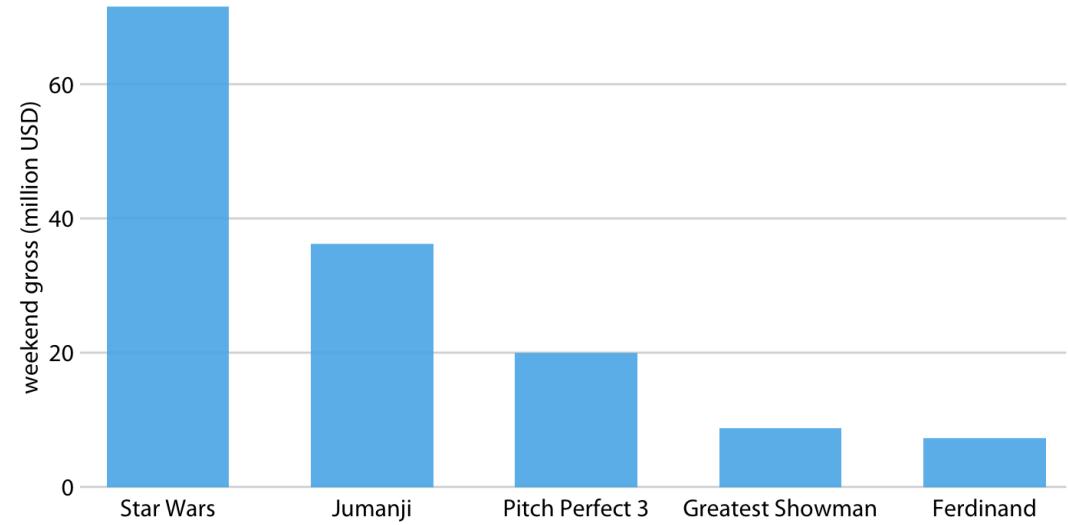
- **Displaying quantities**
- Displaying distributions
- Displaying proportions
- Displaying relationships
- Displaying time series
- Displaying geospatial data
- Other charts
- Uncertainty

DISPLAYING QUANTITIES. BAR CHARTS



DISPLAYING QUANTITIES. BAR CHARTS

- Data: One key, one value
 - Data: 1 category attribute, 1 quantity attribute
 - Mark: lines
- Tasks:
 - Compare/lookup (really easy)
 - Scales to hundreds of elements
- Tips
 - Label orientation
 - Adequate order

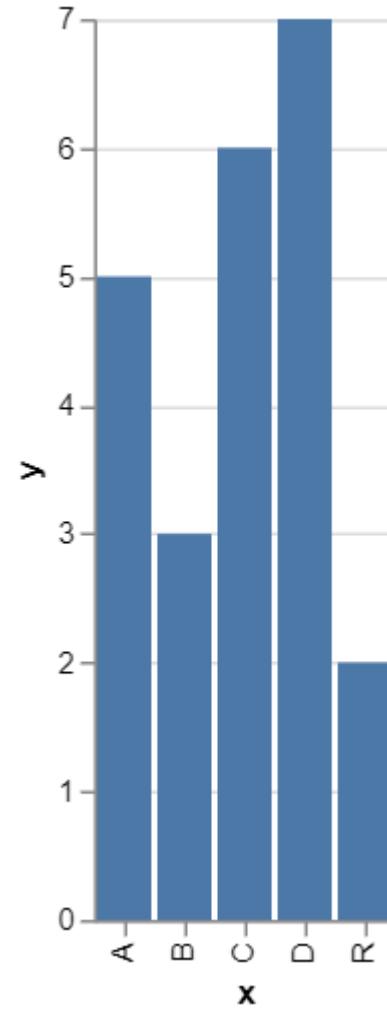


DISPLAYING QUANTITIES. BAR CHARTS

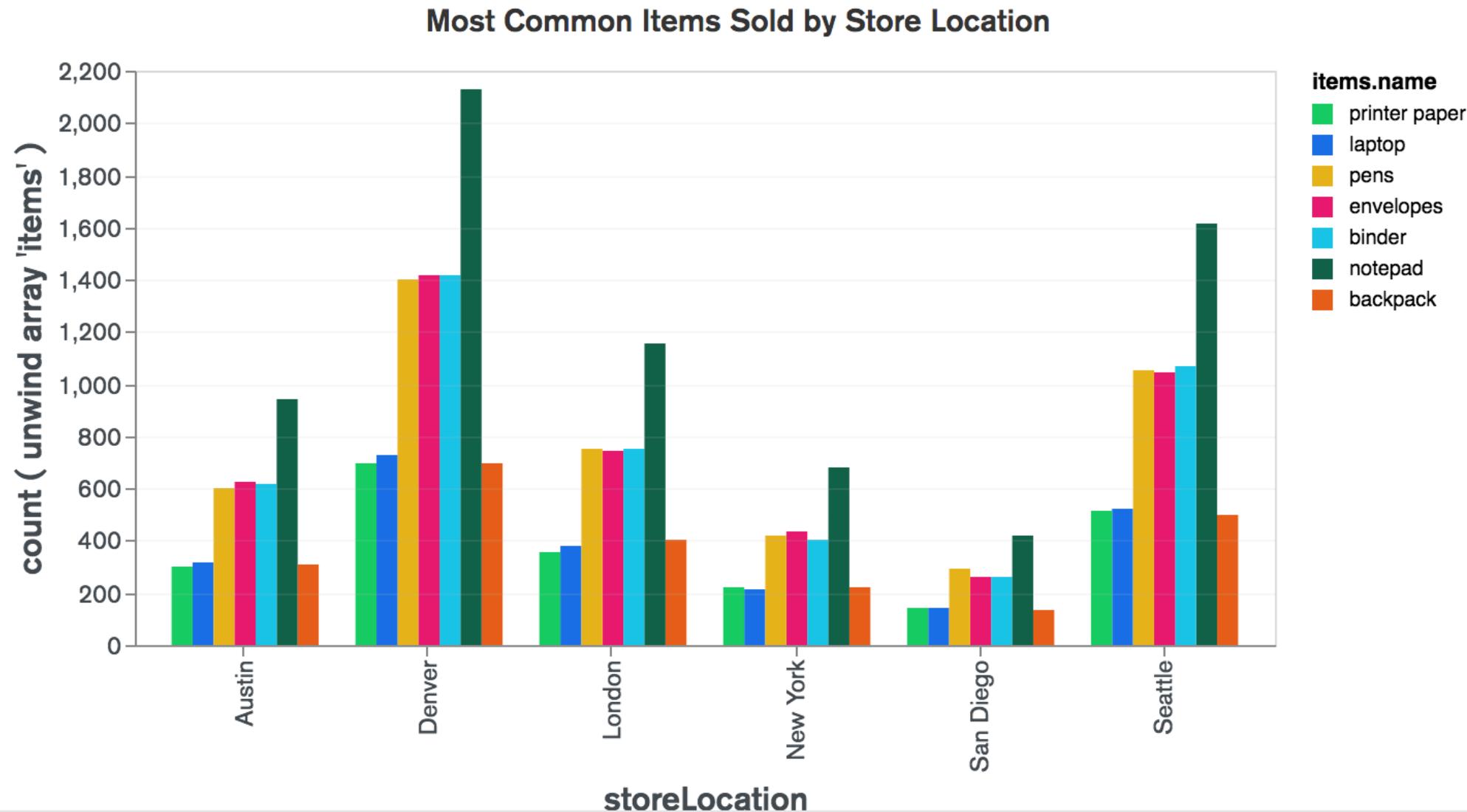
- Bar charts
 - Easy to compare/lookup
 - Scales to hundreds of elements
- Tips
 - Label orientation
 - Adequate order

DISPLAYING QUANTITIES. BAR CHARTS

```
data = alt.Data(values=[{'x': 'A', 'y': 5},  
                     {'x': 'B', 'y': 3},  
                     {'x': 'C', 'y': 6},  
                     {'x': 'D', 'y': 7},  
                     {'x': 'R', 'y': 2}])  
  
alt.Chart(data).mark_bar().encode(  
    x='x:O',  
    y='y:Q'  
)
```



DISPLAYING QUANTITIES. PAIRED BAR CHARTS



DISPLAYING QUANTITIES. PAIRED BAR CHARTS

- Data: One value (quant.), two keys (cat.)
- Marks: lines
- Channels:
 - Length to express quantity
 - Color hue for one key
 - Spatial regions for the other key
 - Separated horizontally, aligned vertically
 - Sometimes ordered by quant attribute, sometimes only on the keys
 - By label (alphabetical)

DISPLAYING QUANTITIES. PAIRED BAR CHARTS

- Tasks:
 - Compare
 - Within same key (easier)
 - Among keys (not so easy)
 - Lookup values
- Scalability:
 - \approx one dozen secondary keys (less than bar charts)

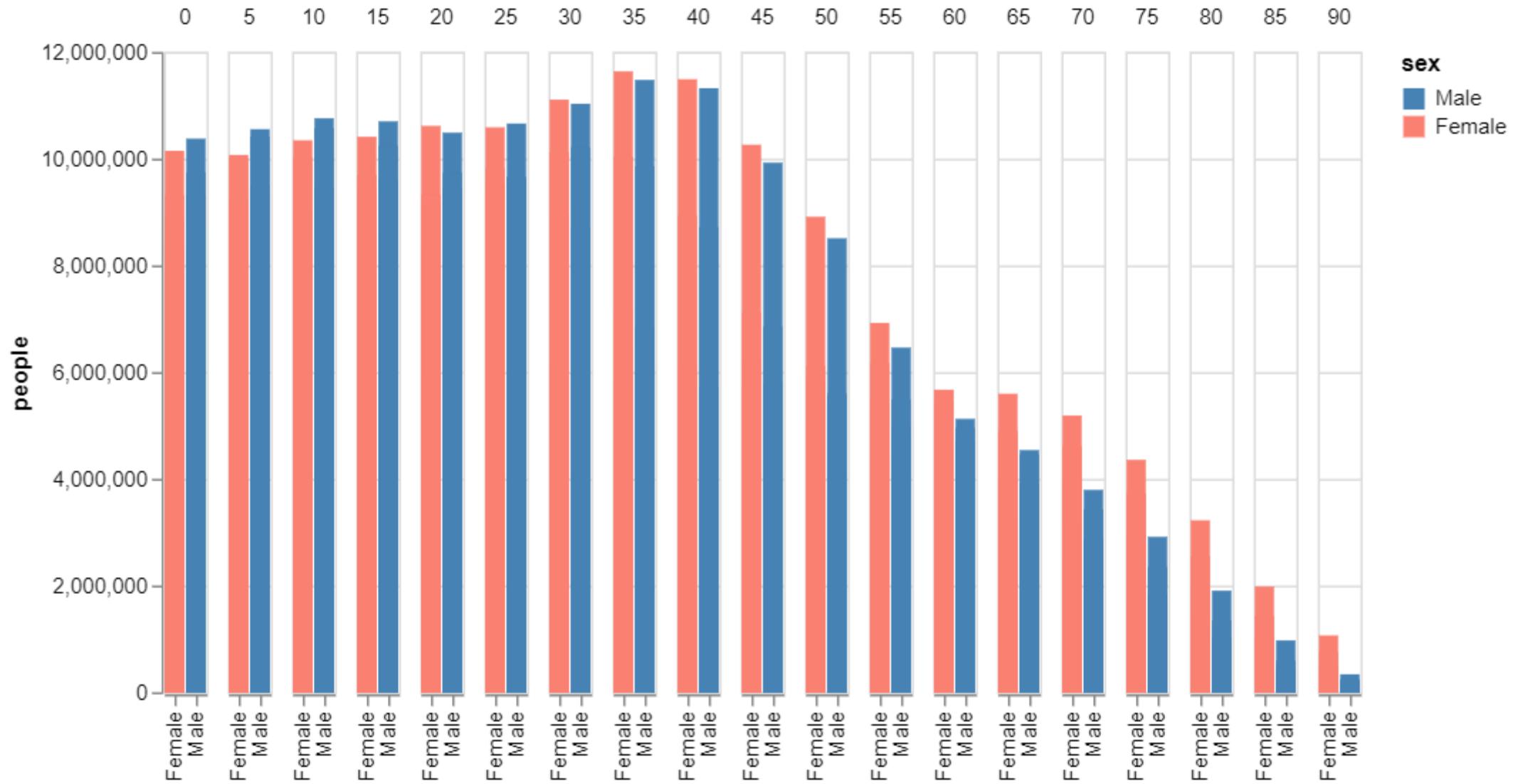
DISPLAYING QUANTITIES. PAIRED BAR CHARTS

```
source = data.population.url

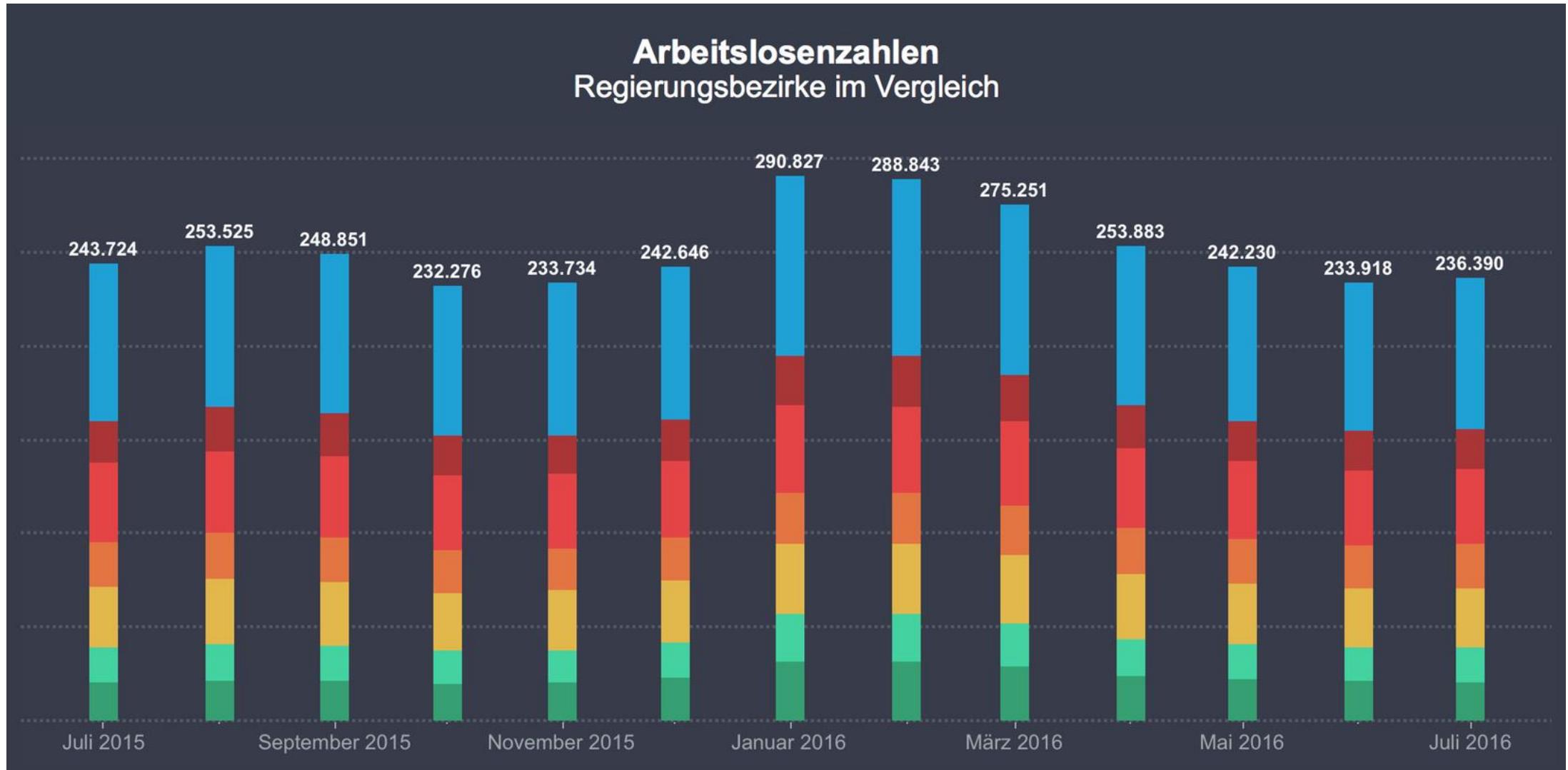
pink_blue = alt.Scale(domain=('Male', 'Female'),
                      range=["steelblue", "salmon"])

alt.Chart(source).mark_bar().encode(
    x=alt.X('sex:N', title=None),
    y=alt.Y('people:Q', scale=alt.Scale(domain=(0, 12000000))),
    color=alt.Color('sex:N', scale=pink_blue),
    column='age:O'
).properties(
    width=20
).transform_calculate(
    "sex", alt.expr.if_(alt.datum.sex == 1, "Male", "Female")
).configure_facet(
    spacing=8
)
```

DISPLAYING QUANTITIES. PAIRED BAR CHARTS



DISPLAYING QUANTITIES. STACKED BAR CHART

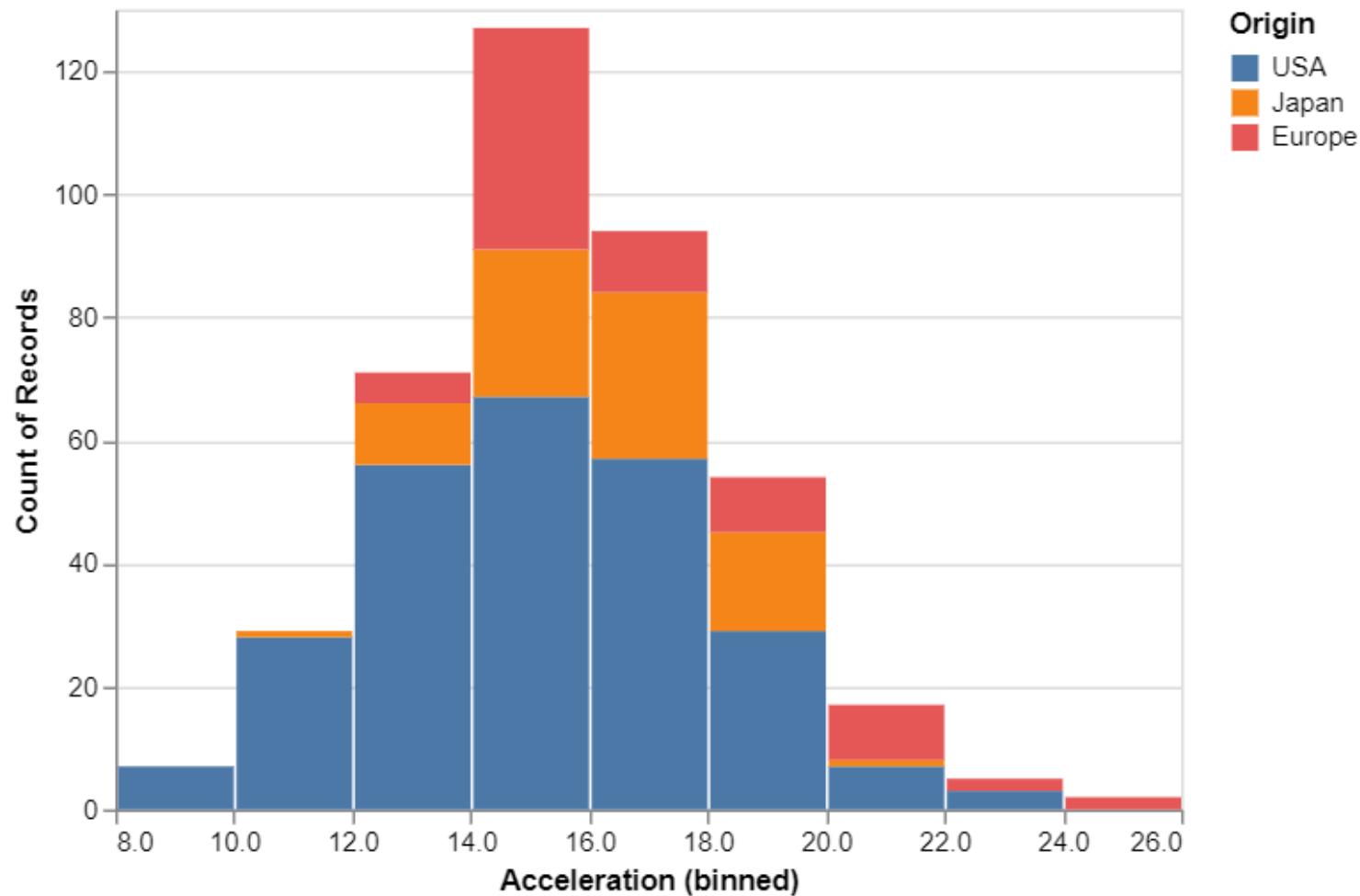


DISPLAYING QUANTITIES. STACKED BAR CHART

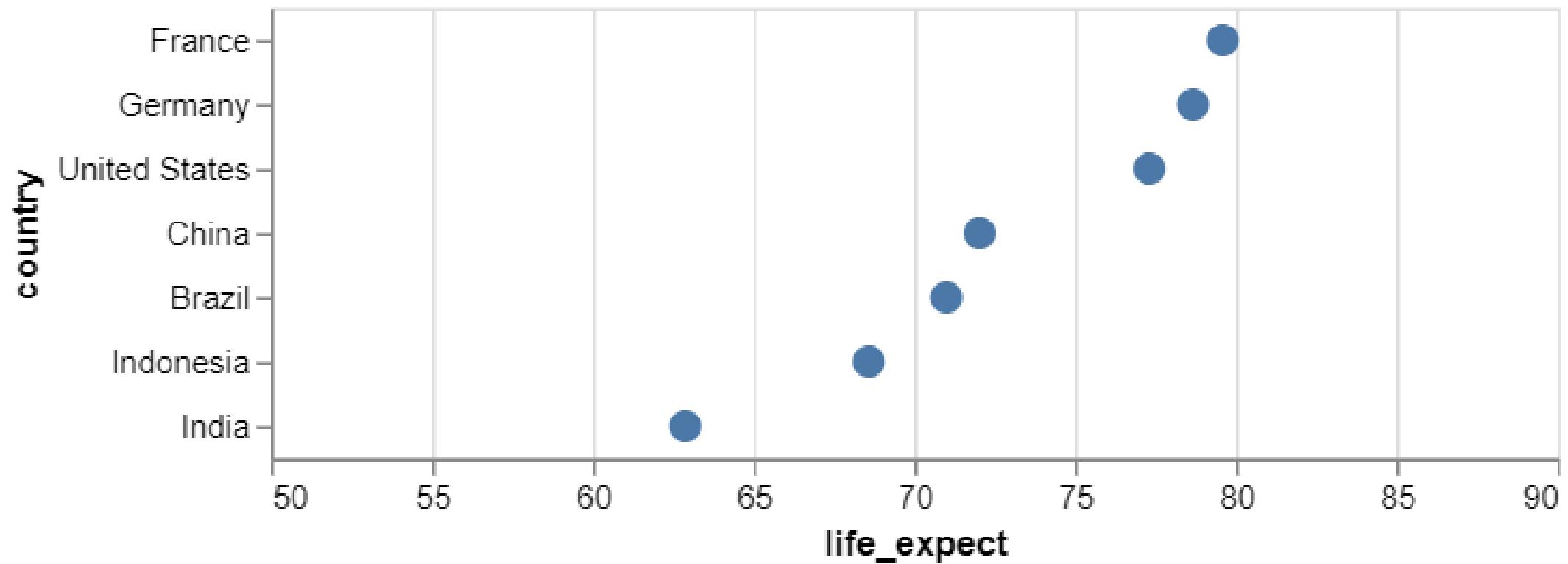
- Data: One value, two keys
- Channels:
 - Same as before, but bars stacked vertically
 - Color hue for one key, spatial regions for the other key
- Tasks:
 - Compare
 - Within same key (difficult), among keys (also difficult), lookup values
- Scalability:
 - \approx one dozen secondary keys (less than bar charts)

DISPLAYING QUANTITIES. STACKED BAR CHART

```
df = data.cars.url  
  
alt.Chart(df).mark_bar()  
.encode(  
    x = alt.X('Acceleration:Q', bin=True),  
    y = alt.Y('count():Q'),  
    color = alt.Color('Origin:N', sort='y'),  
)
```



DISPLAYING QUANTITIES. DOT PLOT



DISPLAYING QUANTITIES. DOT PLOT

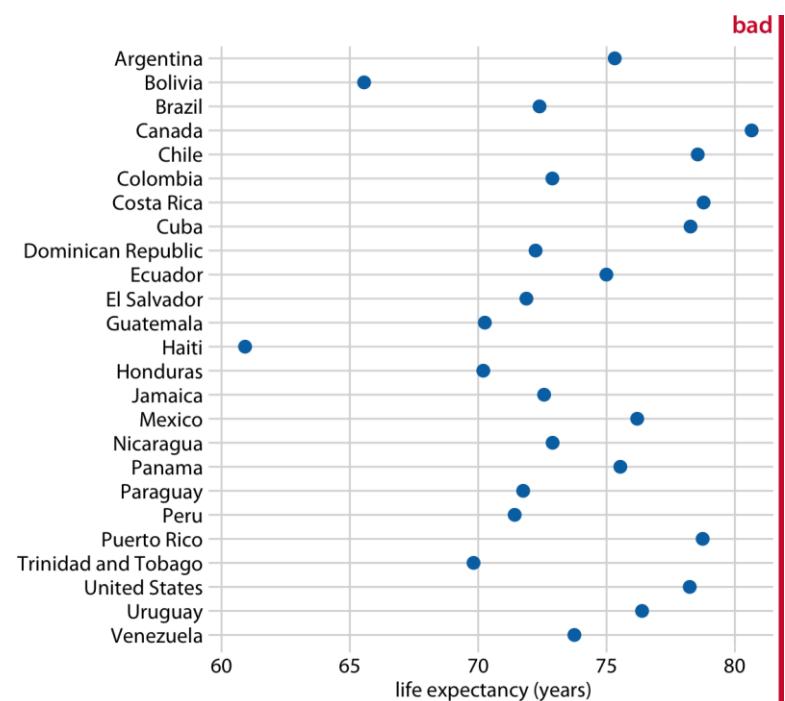
```
import altair as alt
from vega_datasets import data

source = data.countries.url

alt.Chart(source).mark_point(
    size=100,
    opacity=1,
    filled=True
).encode(
    x=alt.X('life_expect:Q', scale = alt.Scale(domain = (50, 90))),
    y=alt.Y('country:N', sort = '-x'),
).transform_filter (alt.datum.year == 2000).transform_filter(
    filter={"field": 'country',
            "oneOf": ["France", "Germany", "China", "India",
                      "United States", "Indonesia", "Brazil"]})
)
```

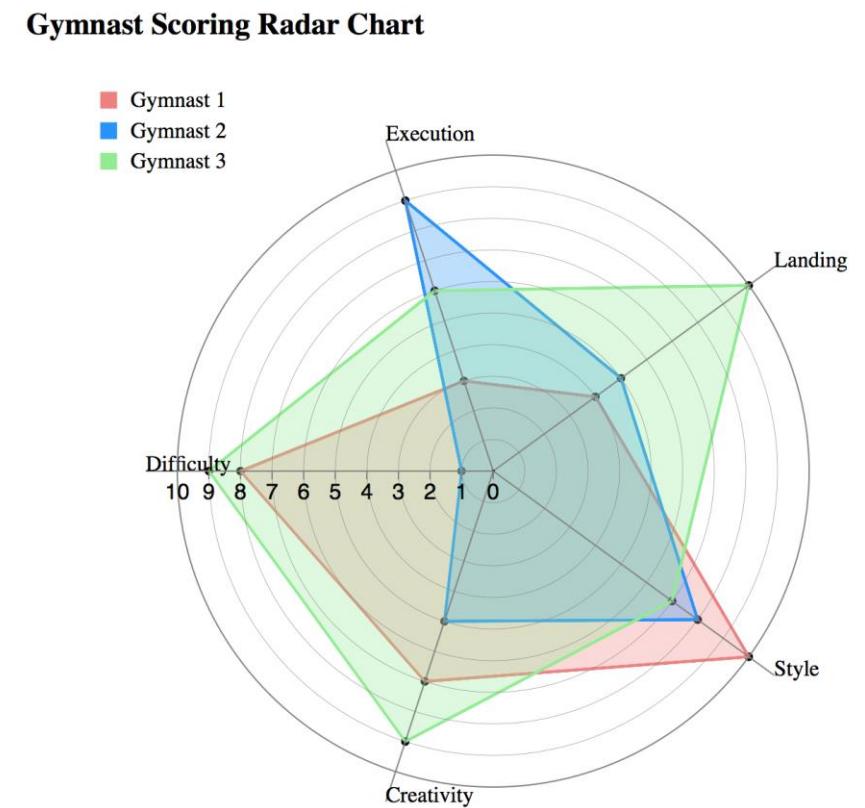
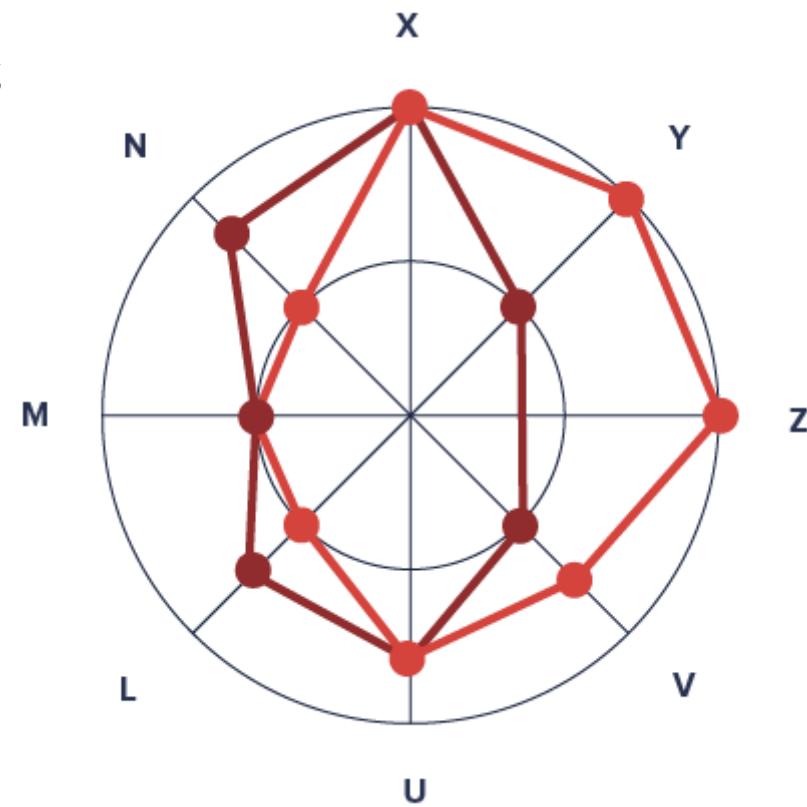
DISPLAYING QUANTITIES. DOT PLOT

- Do not need to start at zero
 - Suitable when small differences must be shown
 - Bar charts might lead the attention away from those differences
- Tips:
 - Must be ordered by quantities
 - The opposite makes the chart difficult to read



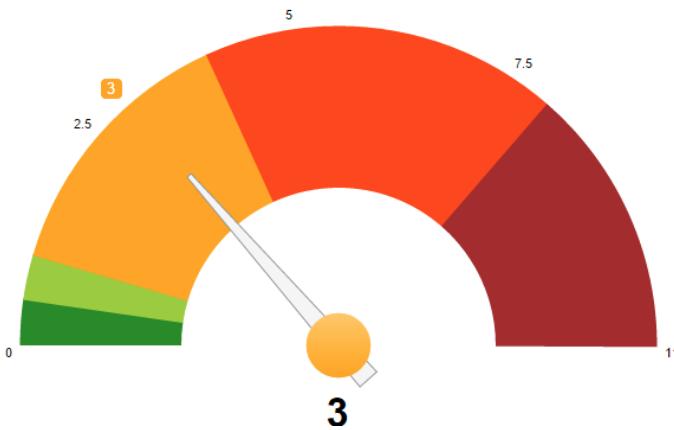
DISPLAYING QUANTITIES. RADAR CHART

- Analogous to paired/grouped column charts
 - Multiple dimensions
 - Space efficient
 - Different designs



DISPLAYING QUANTITIES. GAUGE & BULLET CHARTS

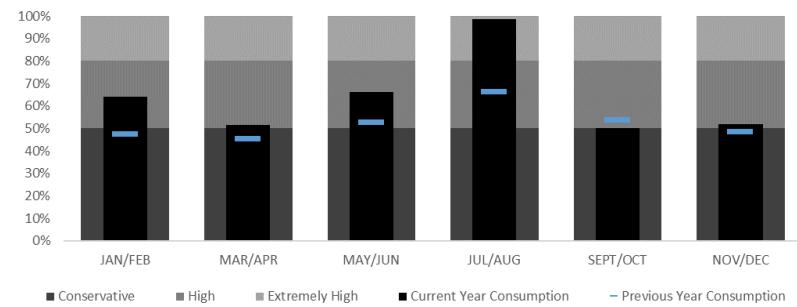
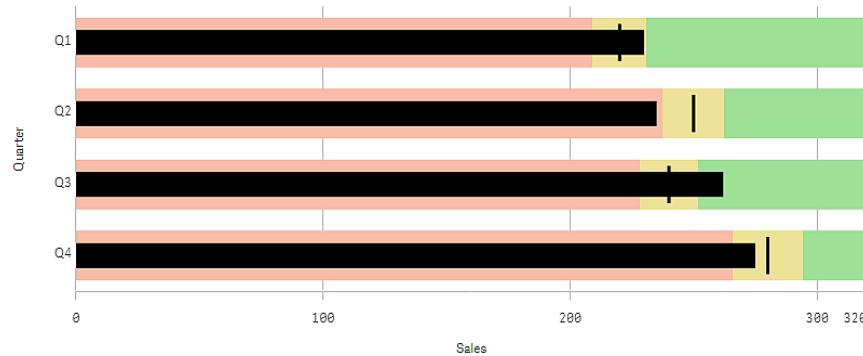
Gauge charts



How satisfied are you with our product?



Bullet charts



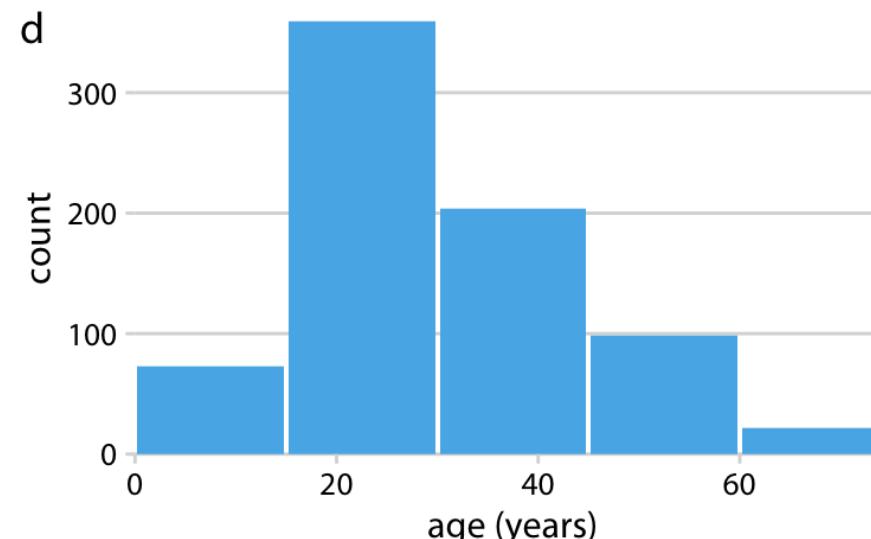
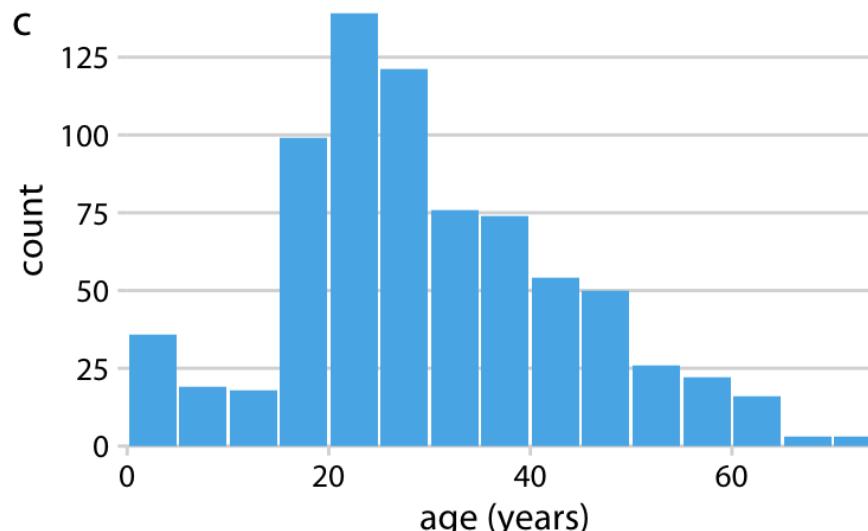
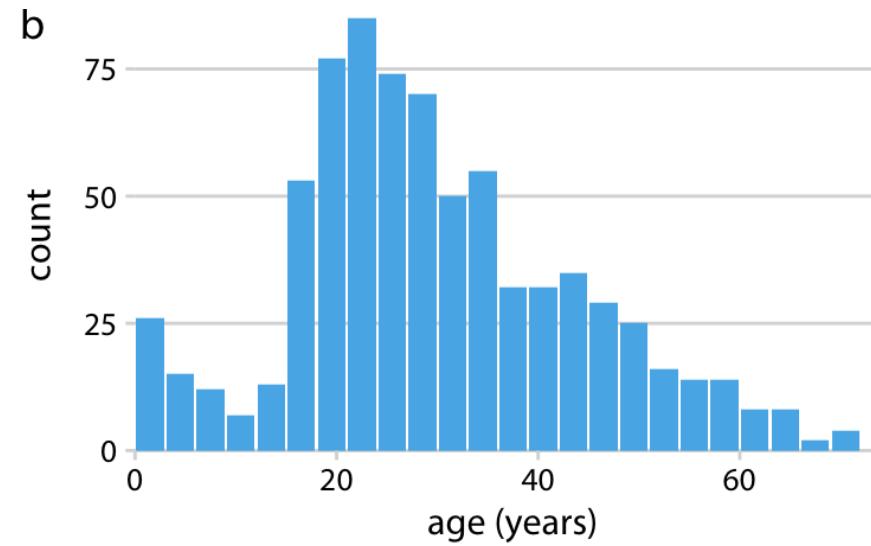
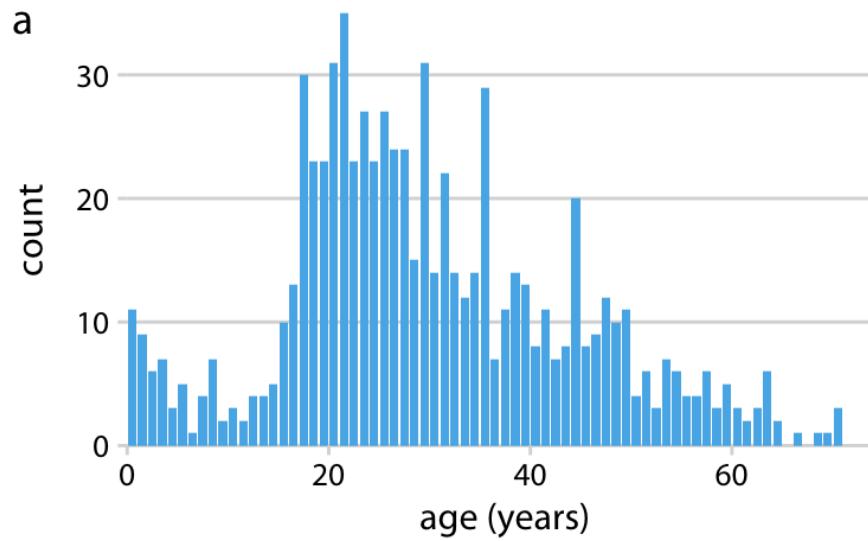
OUTLINE

- *Displaying quantities*
- **Displaying distributions**
- Displaying proportions
- Displaying relationships
- Displaying time series
- Displaying geospatial data
- Other charts
- Uncertainty

DISTRIBUTIONS. HISTOGRAMS

- Individual observations may be not adequate
 - Too many
 - Interested in trends/cumulative values...
- Binning information in value ranges: Histograms
 - Bar charts with binned information
 - Binning sizes may be important: Different distributions may appear
 - Explore different bin sizes

DISTRIBUTIONS. HISTOGRAMS

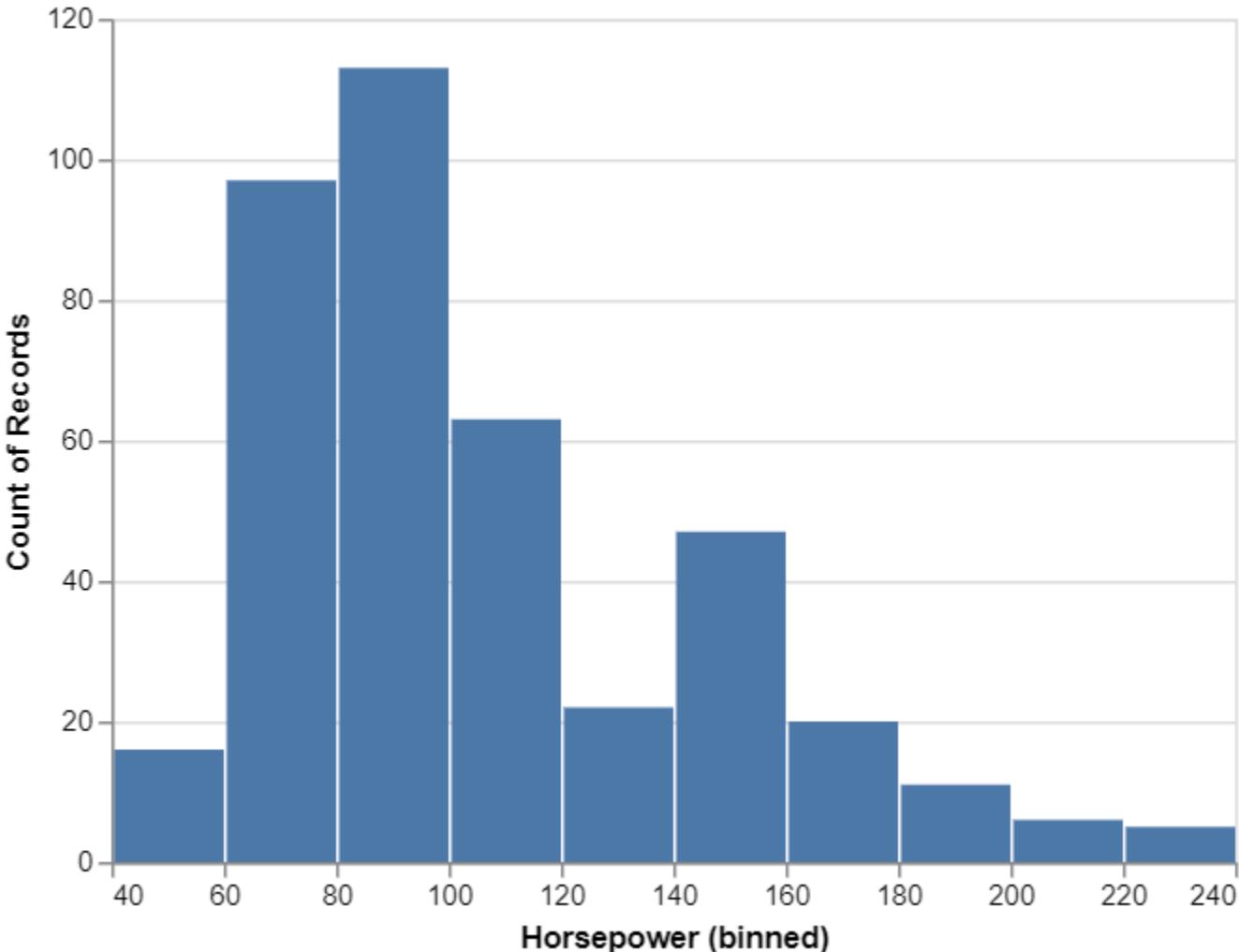


DISTRIBUTIONS. HISTOGRAMS

```
import altair as alt
from vega_datasets import data

source = data.cars()

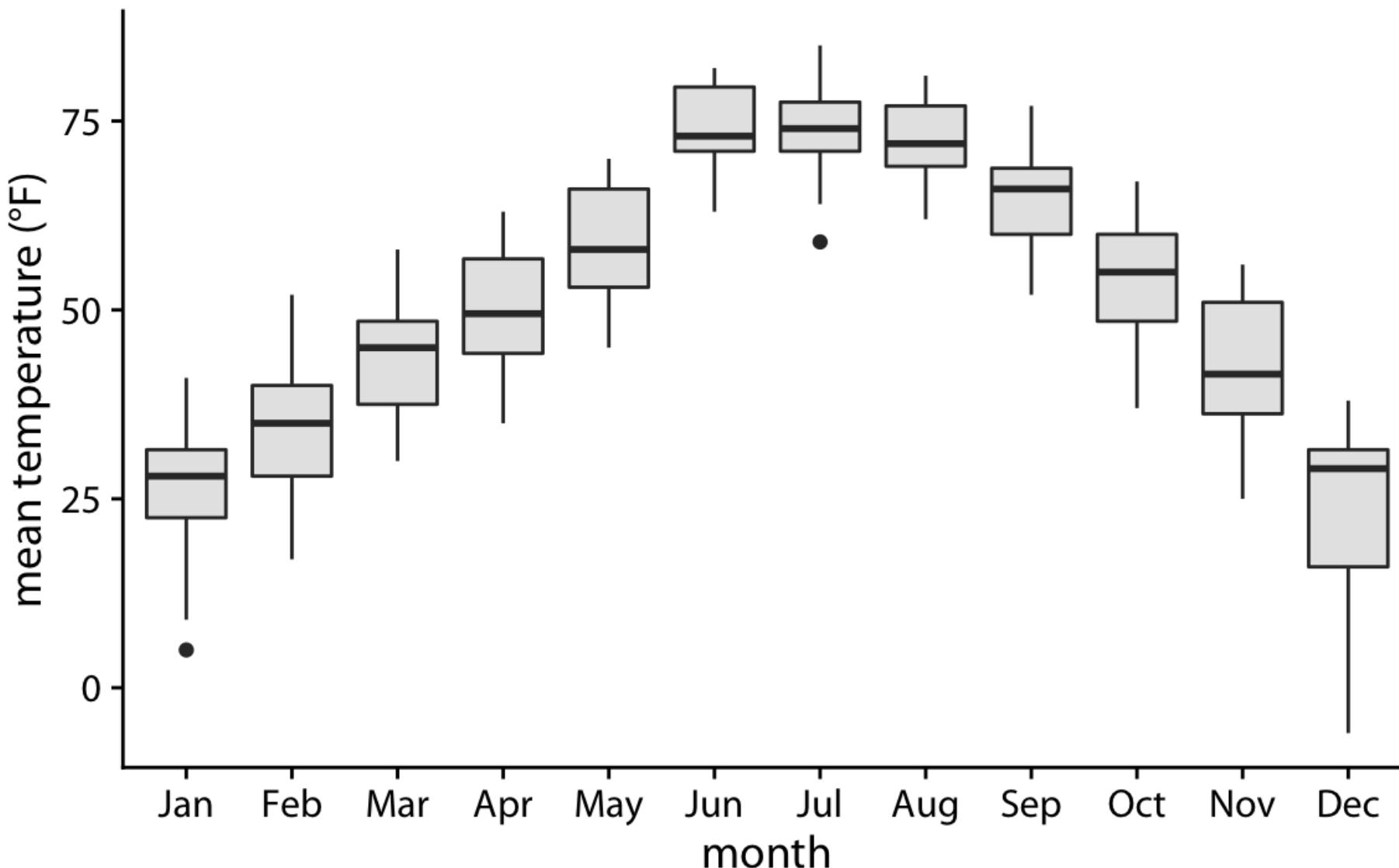
alt.Chart(source).mark_bar().encode(
    alt.X("Horsepower:Q", bin=True),
    y='count()',
)
```



DISTRIBUTIONS. BOX PLOTS

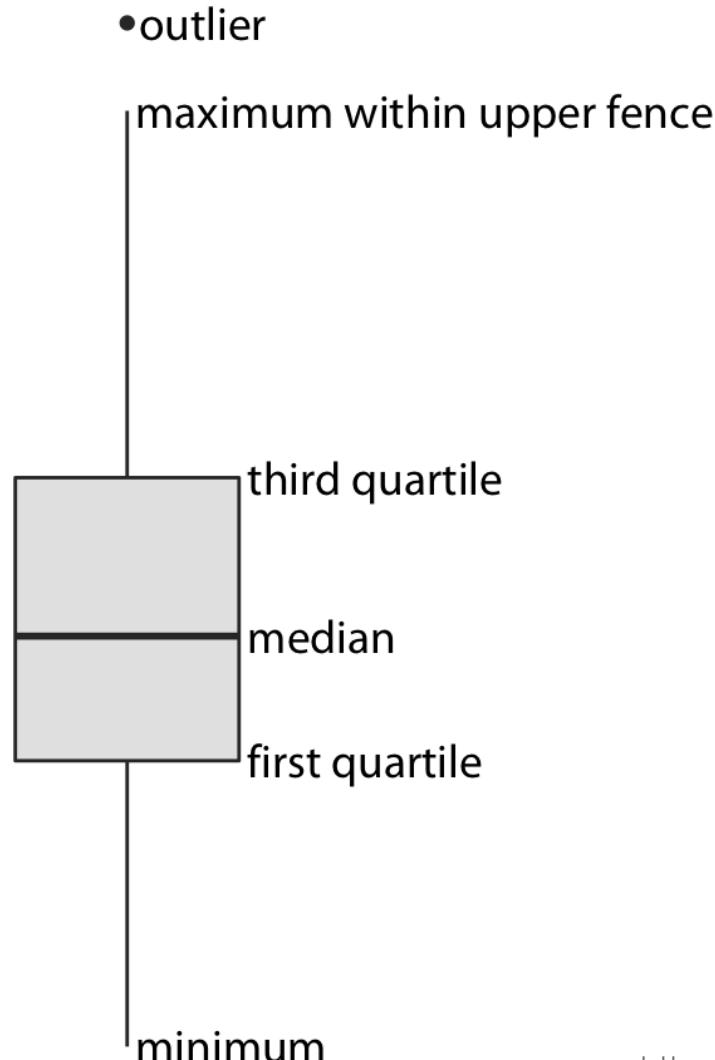
- Useful for several distributions at the same time
- Gives insights on data distribution inside
 - Statistical descriptors
 - Median, minimum, maximum, outliers...

DISTRIBUTIONS. BOX PLOTS



DISTRIBUTIONS. BOX PLOTS

- Construction

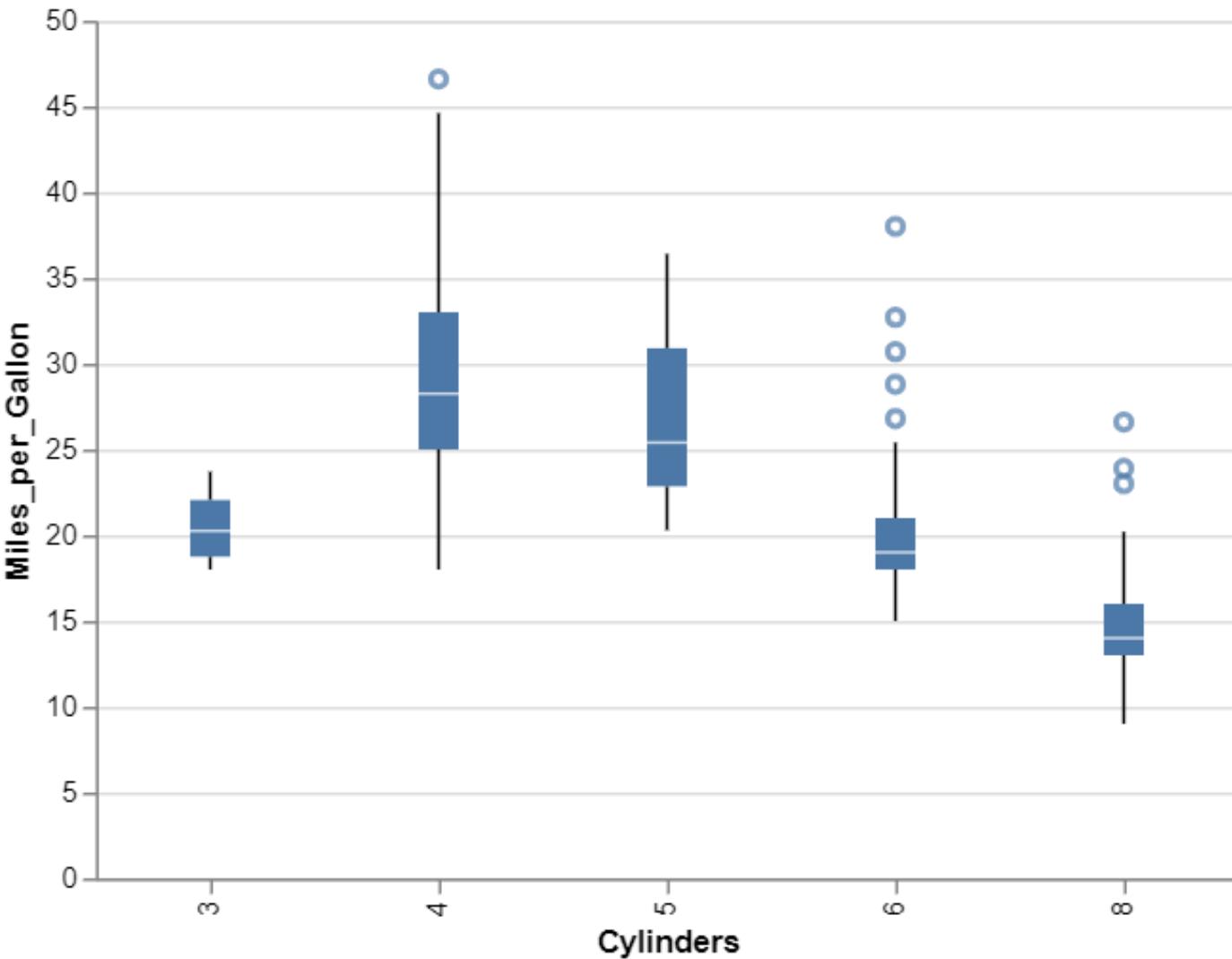


DISTRIBUTIONS. BOX PLOTS

```
import altair as alt
from vega_datasets import data

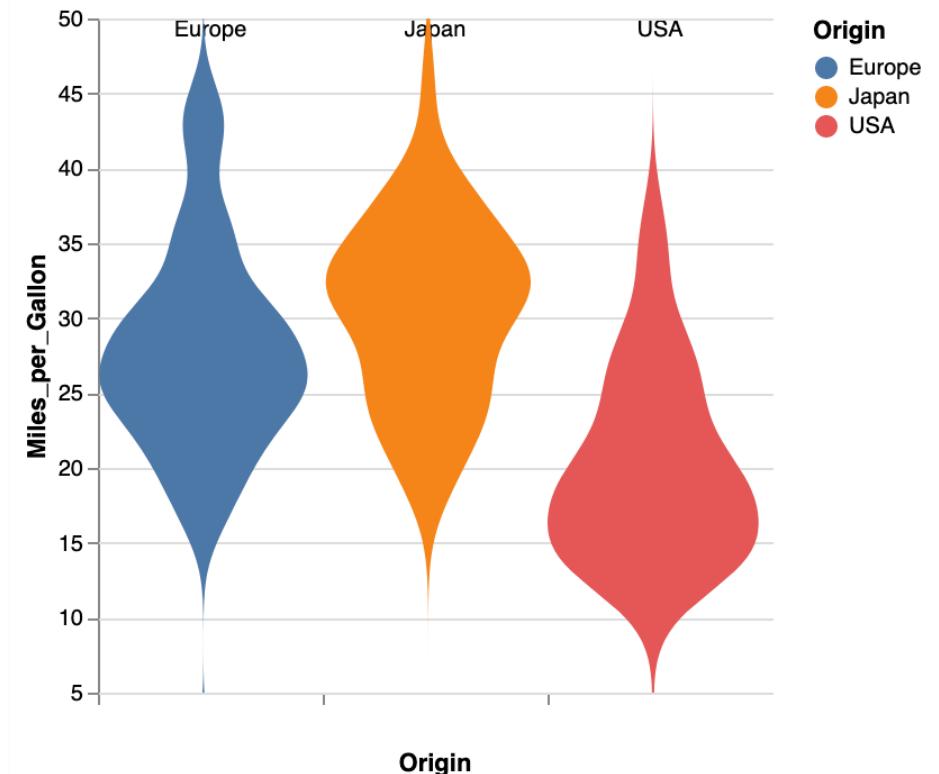
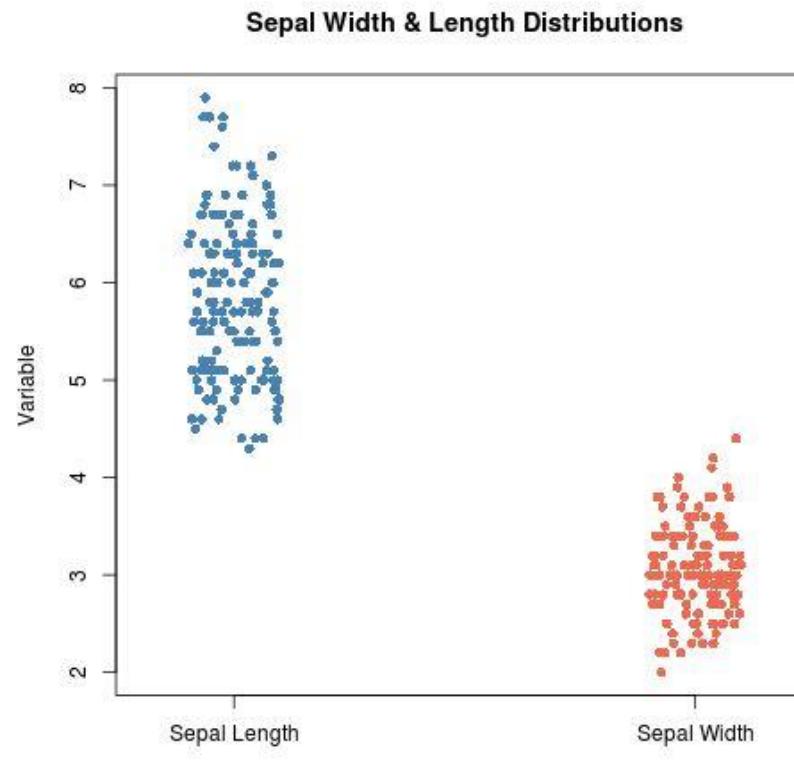
source = data.cars.url

alt.Chart(source).mark_boxplot().encode(
    x='Cylinders:O', y='Miles_per_Gallon:Q'
).properties(
    width=400,
    height=300
)
```



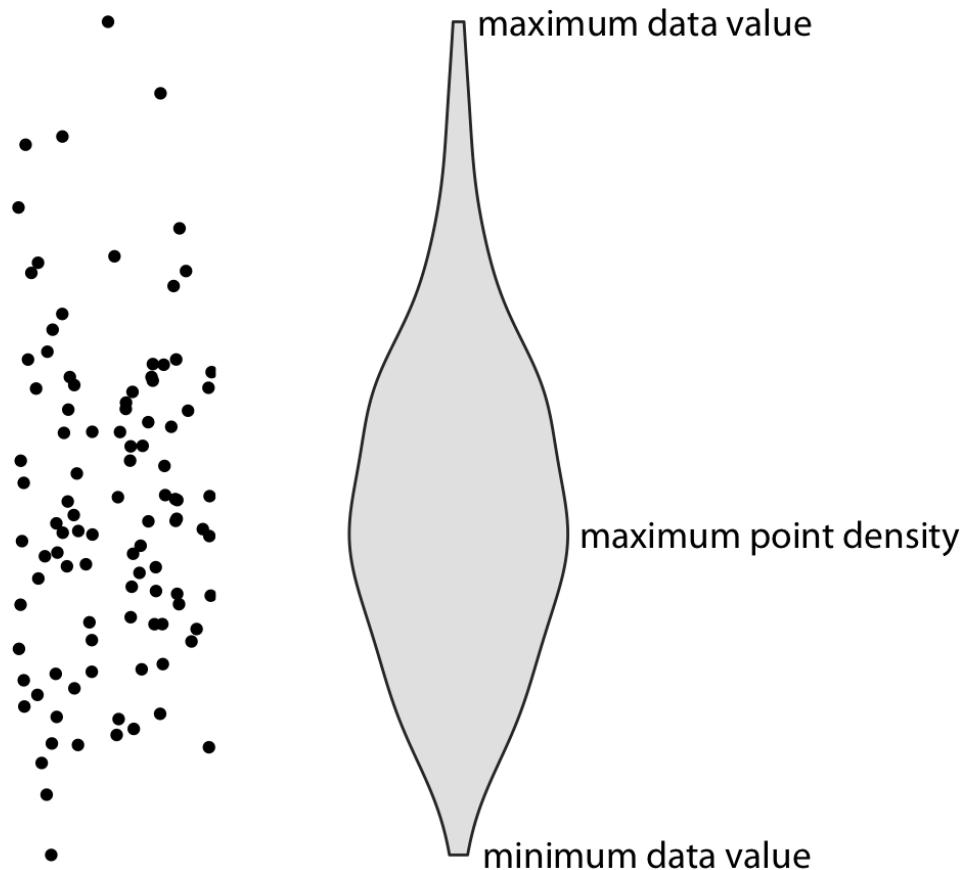
DISTRIBUTIONS. VIOLIN PLOTS & STRIP CHARTS

- Box plots hide/abstract too much data
 - If interested in the internal distribution, alternatives are violin plots, strip charts...



DISTRIBUTIONS. VIOLIN PLOTS & STRIP CHARTS

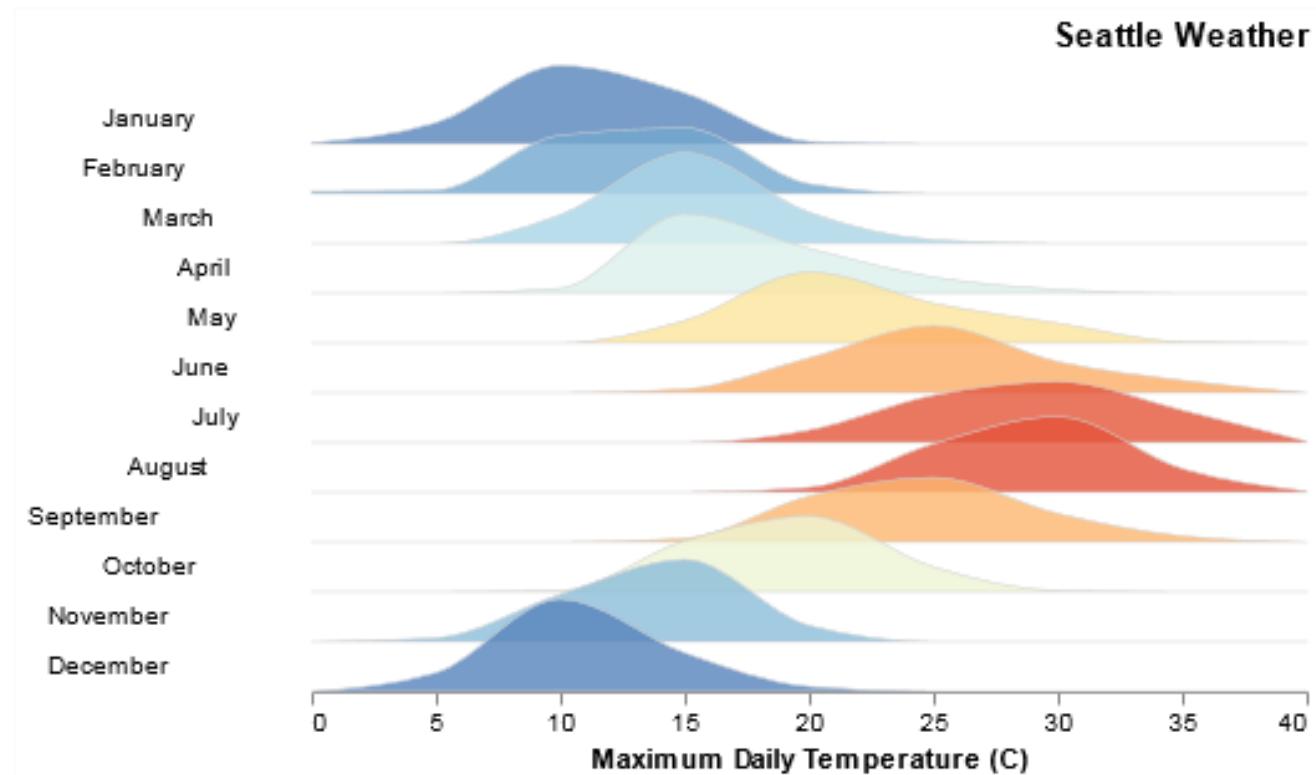
- Violin plots. Construction from a distribution:



```
alt.Chart(data.cars()).transform_density(
    'Miles_per_Gallon',
    as_=['Miles_per_Gallon', 'density'],
    extent=[5, 50],
    groupby=['Origin']
).mark_area(orient='horizontal').encode(
    y='Miles_per_Gallon:Q',
    color='Origin:N',
    x=alt.X(
        'density:Q',
        stack='center',
        impute=None,
        title=None,
        axis=alt.Axis(labels=False, values=[0],grid=False, ticks=True),
    ),
    column=alt.Column(
        'Origin:N',
        header=alt.Header(
            titleOrient='bottom',
            labelOrient='bottom',
            labelPadding=0,
        ),
    )
).properties(
    width=100
).configure_facet(
    spacing=0
).configure_view(
    stroke=None
)
```

DISTRIBUTIONS. RIDGE PLOTS

- Like a half violin plot in horizontal (violin plots are symmetric)
 - Allows more data
 - Even overlapping if done carefully



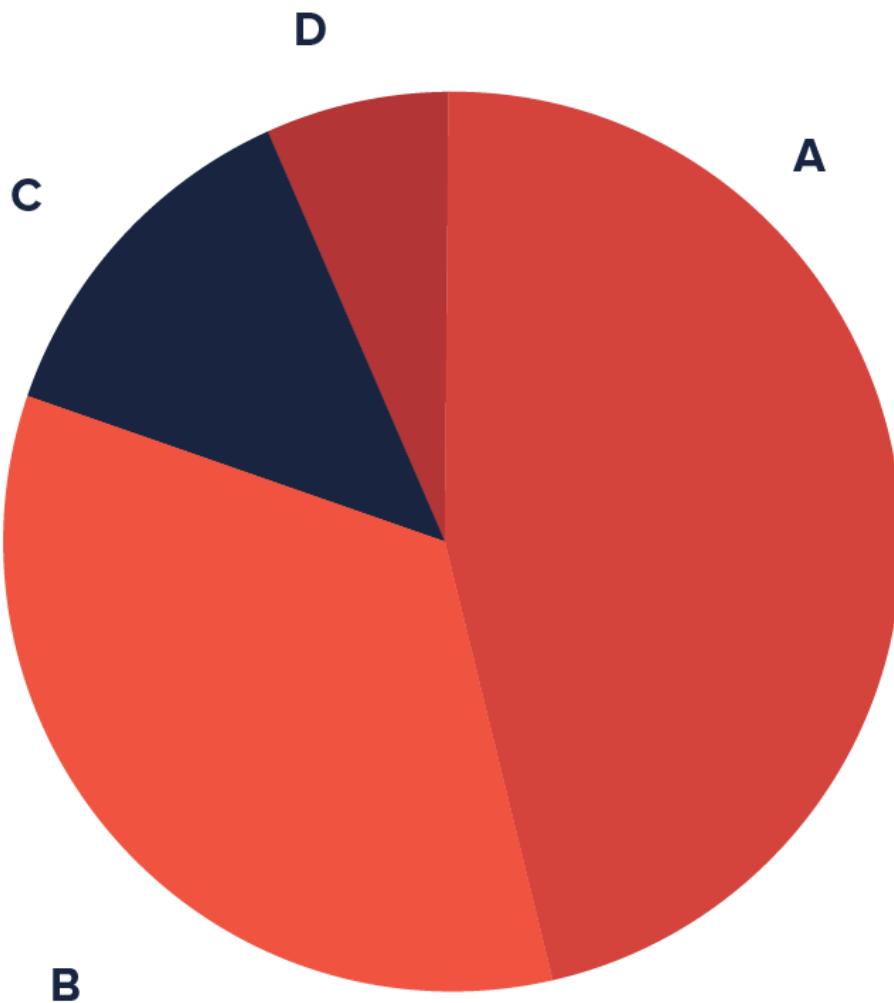
OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- **Displaying proportions**
- Displaying relationships
- Displaying time series
- Displaying geospatial data
- Other charts
- Uncertainty

PROPORTIONS. PIE CHARTS

- Pie chart
 - Area marks with **angle** channel (angle encodes quantity)
 - Accuracy: angle/area much less accurate than line length
- Data
 - 1 key attribute, 1 quantitative value attribute
- Task
 - Part-to-whole judgements
- Perceptual issues

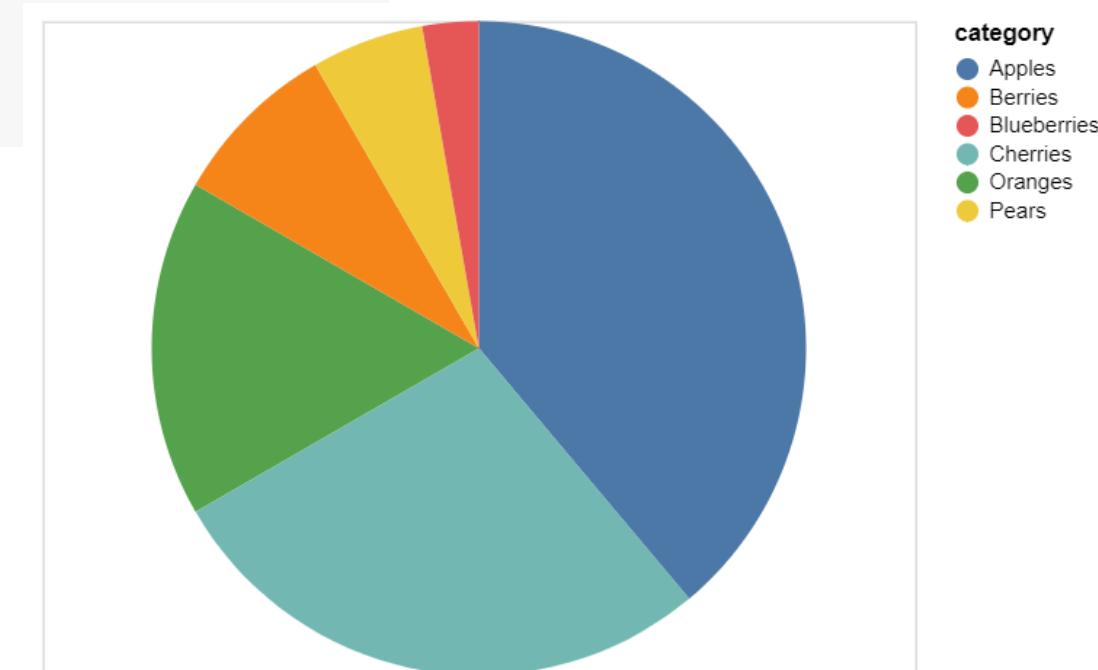
PROPORTIONS. PIE CHARTS

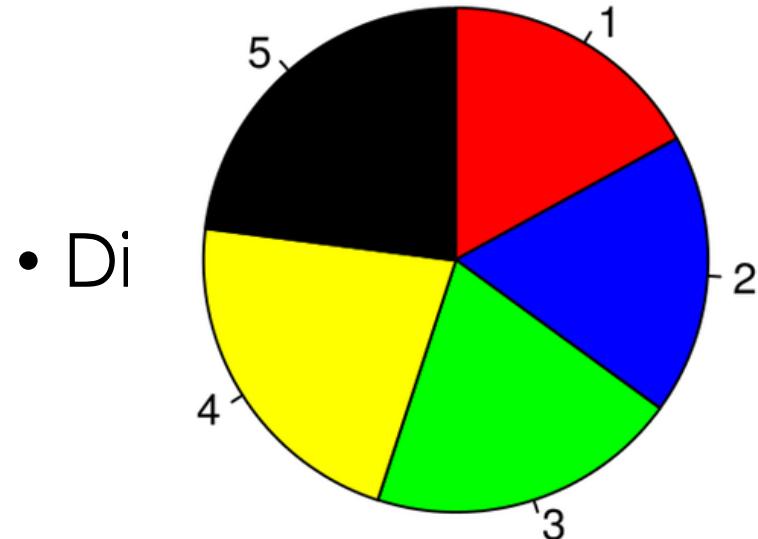
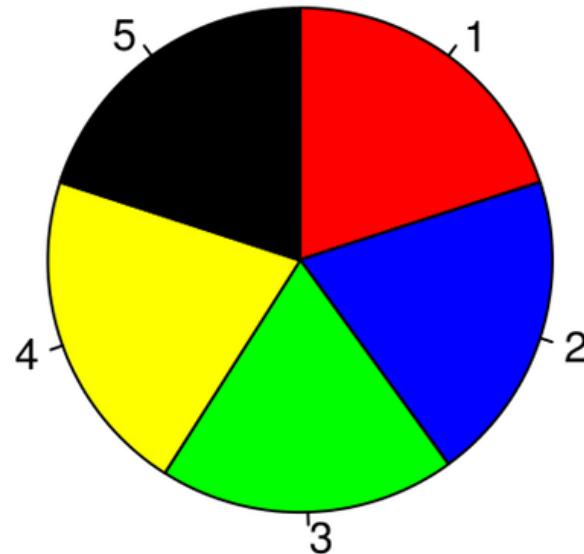
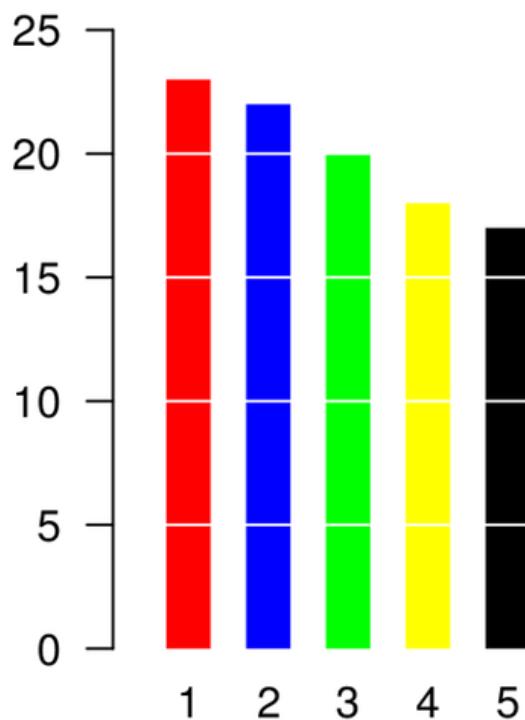
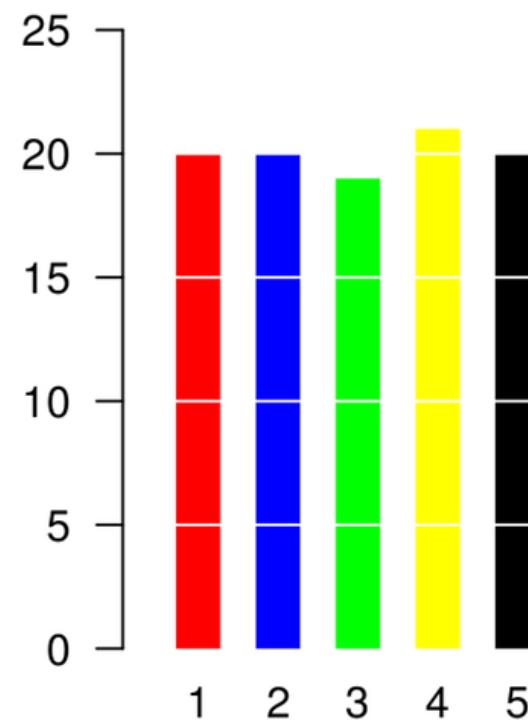
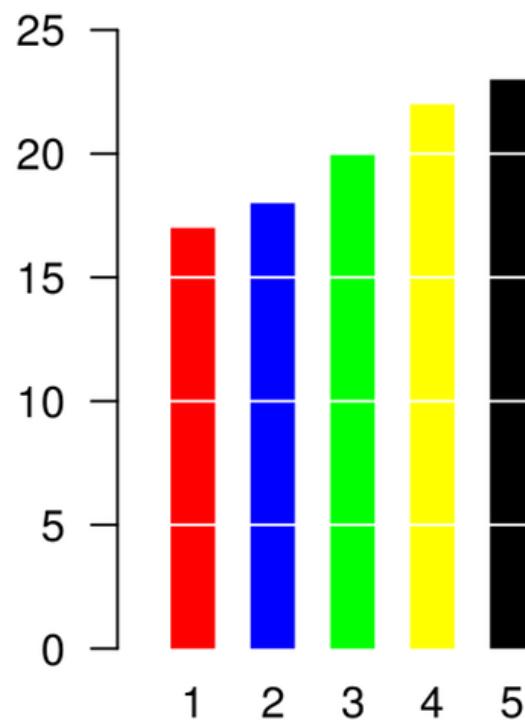
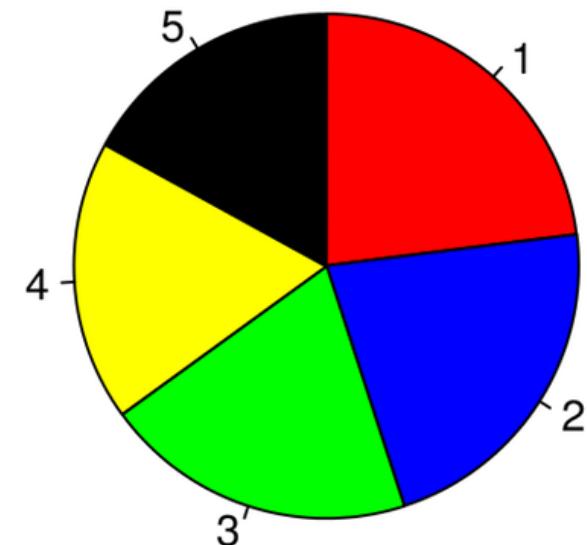


PROPORTIONS. PIE CHARTS

```
source = pd.DataFrame({"category": ['Apples', 'Cherries', 'Oranges',
                                     'Berries', 'Pears', 'Blueberries'],
                       "value": [14, 10, 6, 3, 2, 1]})

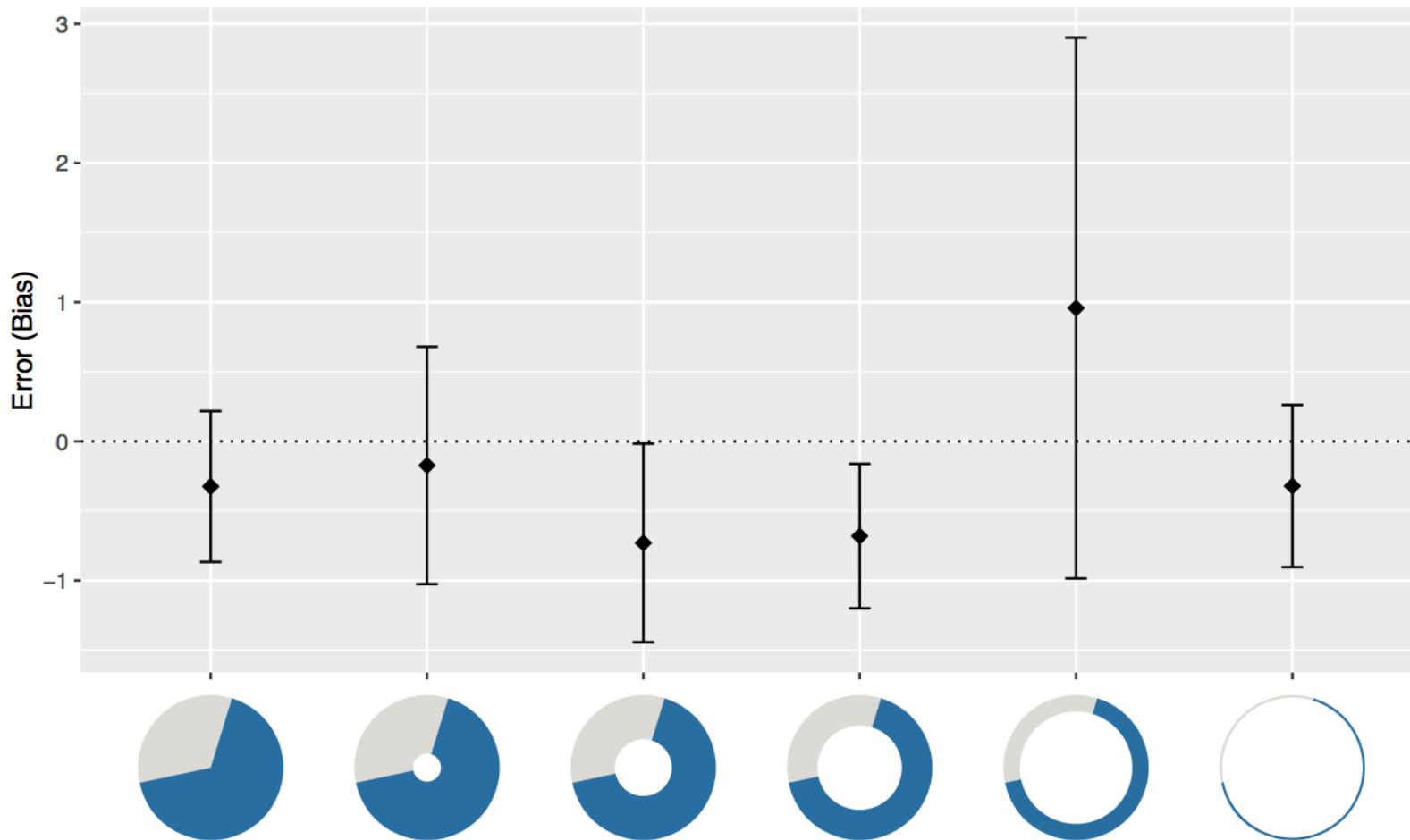
alt.Chart(source).mark_arc().encode(
    theta=alt.Theta(field="value", type="quantitative"),
    order = alt.Order('value', sort = 'descending'),
    color=alt.Color(field="category", type="nominal"),
)
```



**A****B****C**

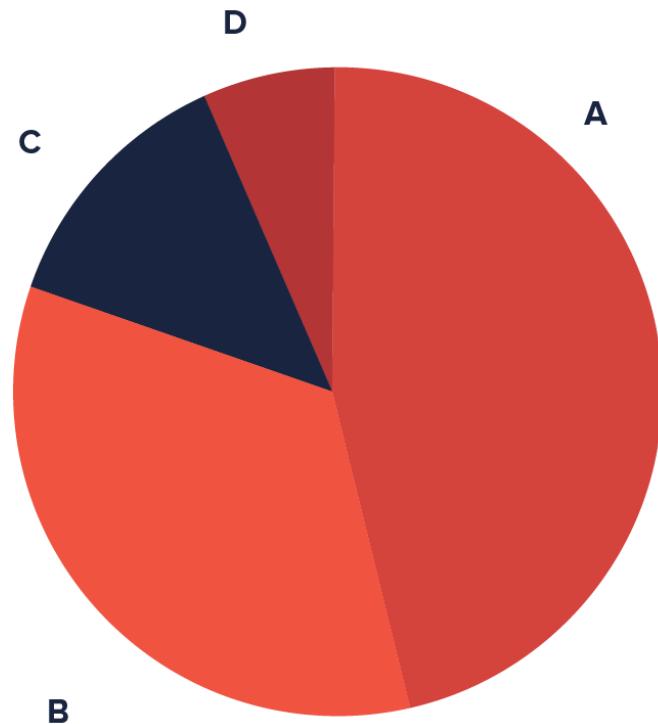
BASIC CHARTS. PIE CHART, POLAR AREA CHARTS

- Donut charts seem to be equally good (cf. [Kosara 2016])

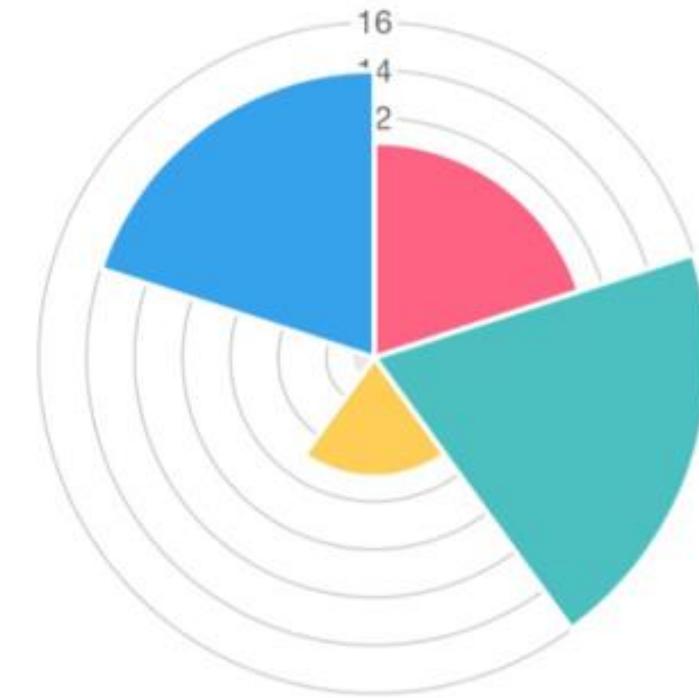


PROPORTIONS. PIE CHARTS

- Different from polar area charts
 - Sectors are equal size, quantities encoded using radius



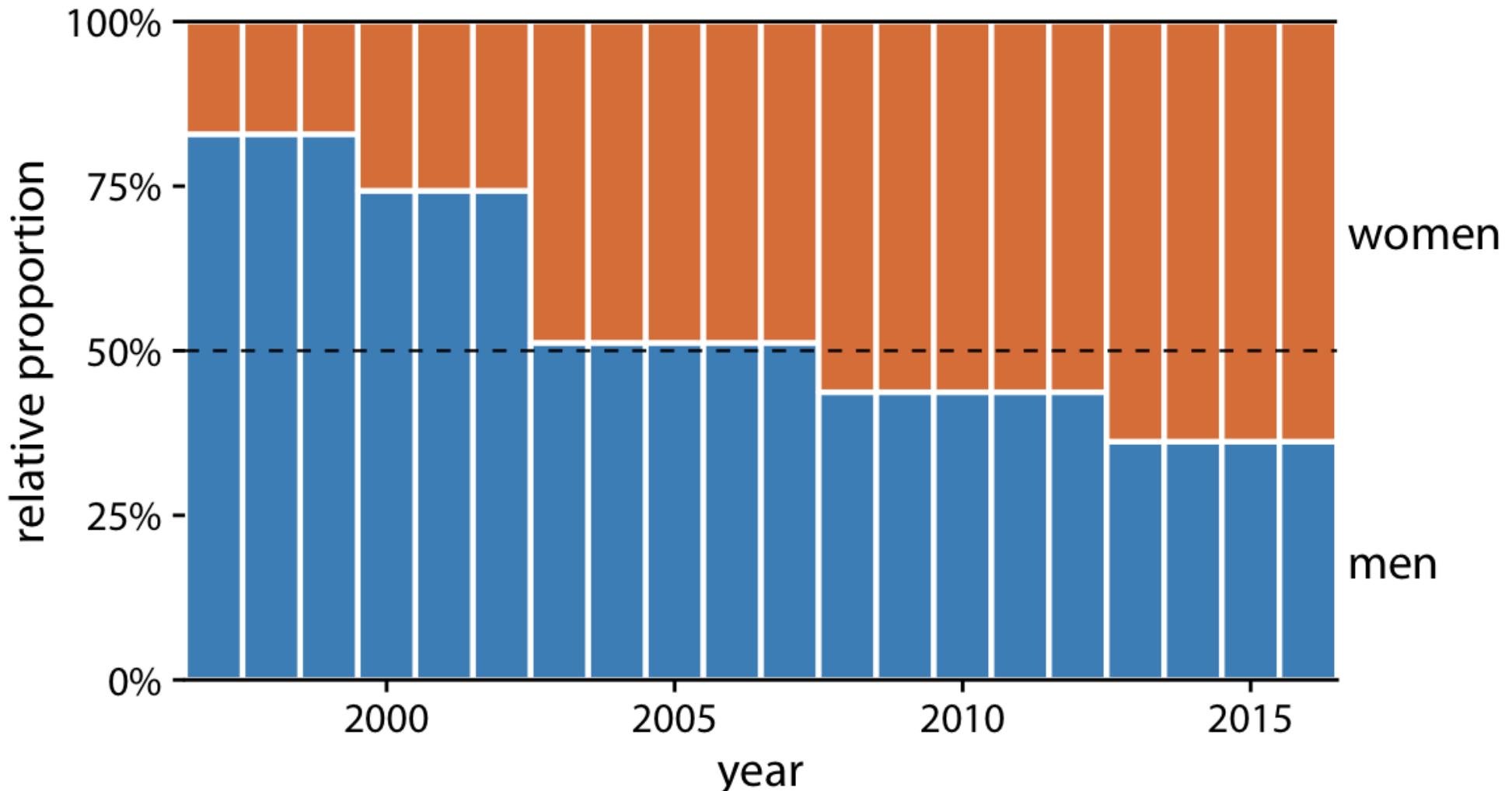
Red Green Yellow Grey Blue



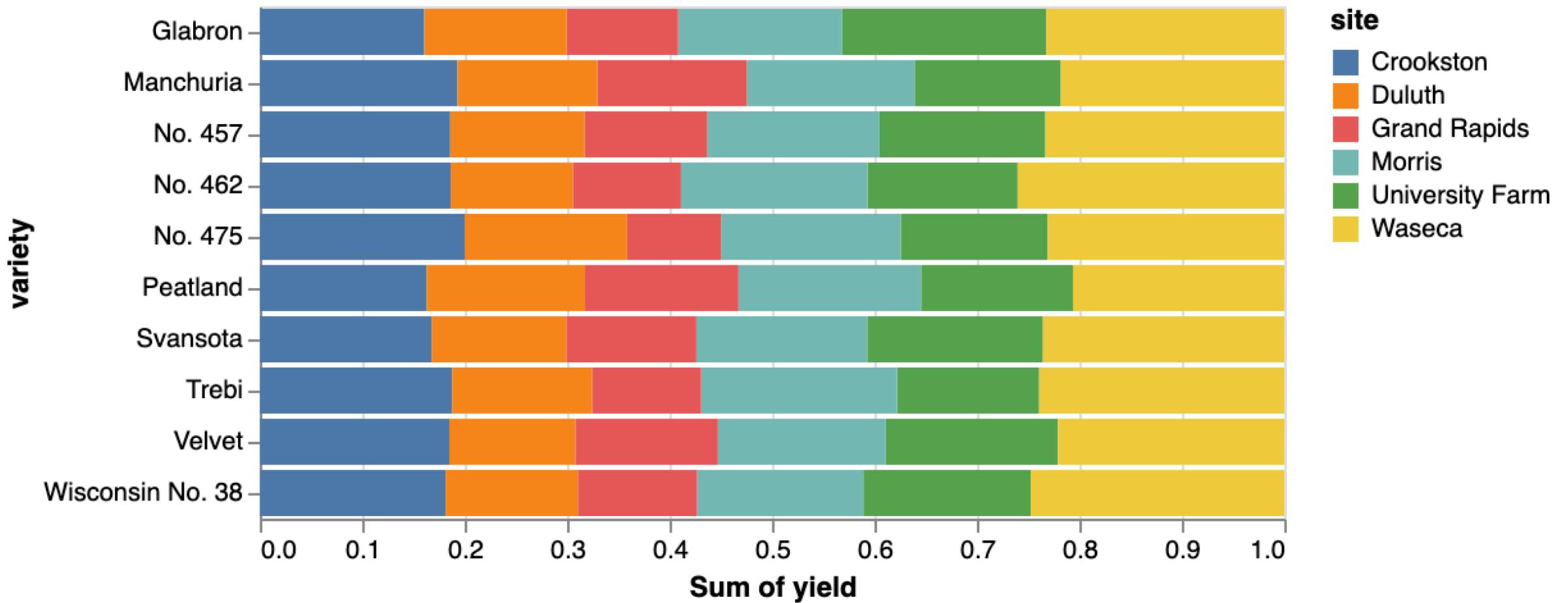
PROPORTIONS. NORMALIZED STACK BARS

- May work for small amount of stacked categories
 - Need to facilitate comparison
 - Many stacked categories are difficult
 - Two categories align on top and bottom
- Stacked area charts also commonly used
 - Same problems than stacked bars
 - Might be useful if converted to small multiples

PROPORTIONS. NORMALIZED STACK BARS



PROPORTIONS. NORMALIZED STACK BARS

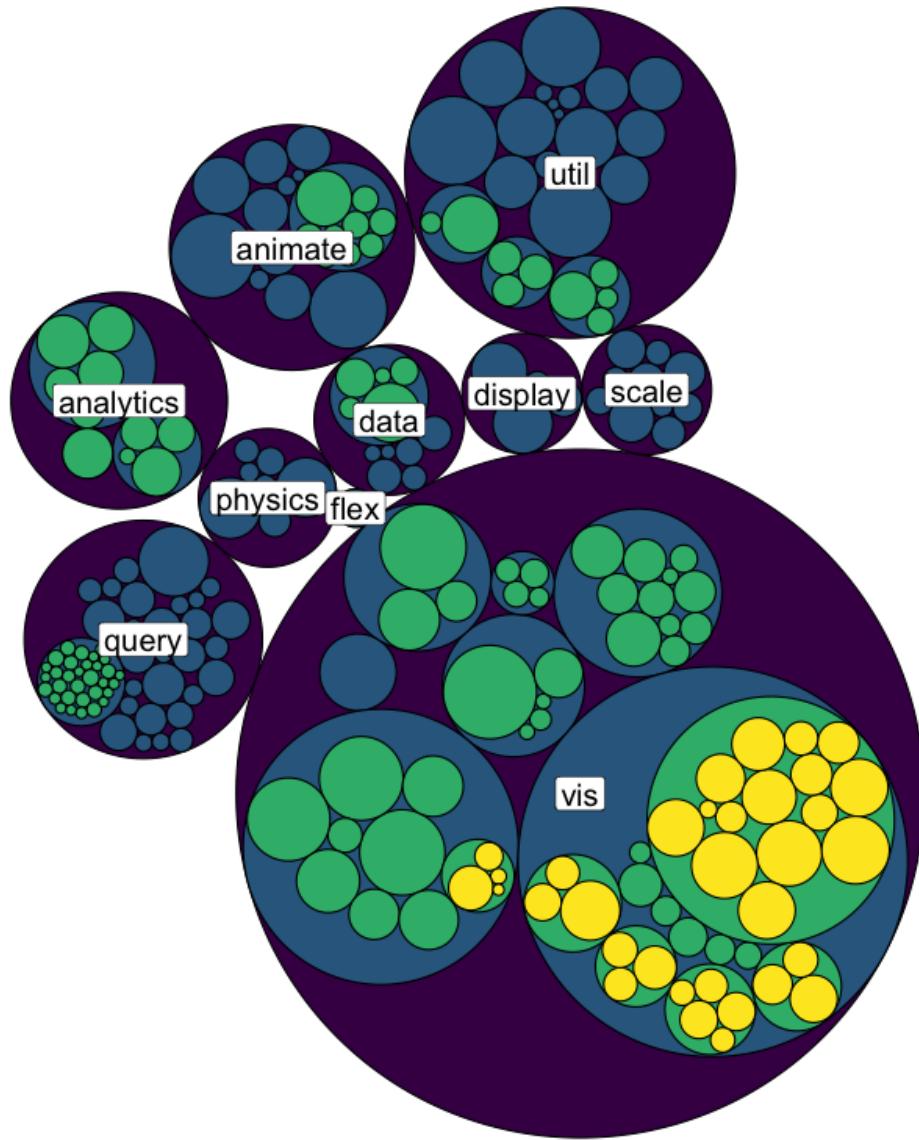


PROPORTIONS. TREEMAPS

- AKA enclosure diagram



PROPORTIONS. TREEMAPS



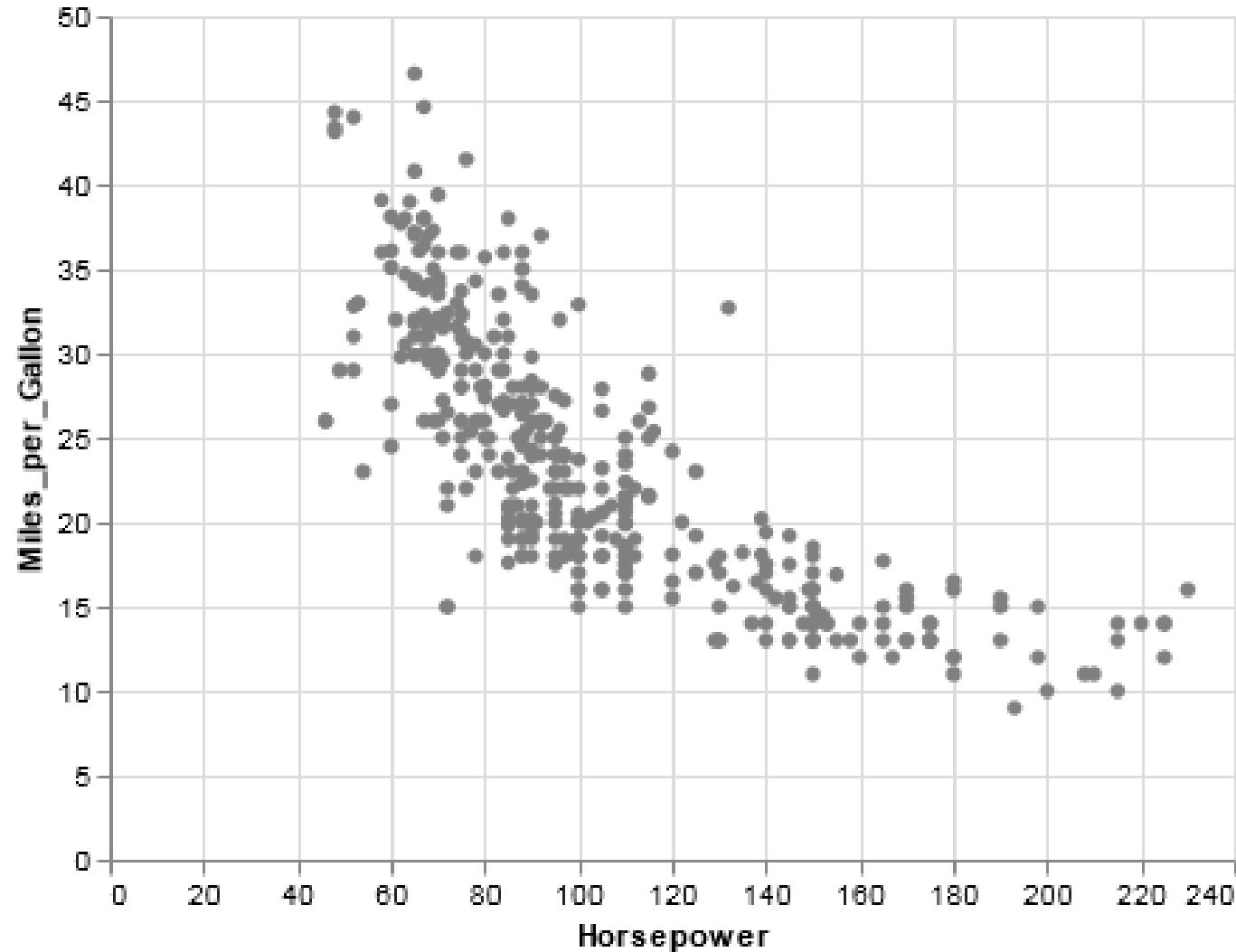
OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- *Displaying proportions*
- **Displaying relationships**
- Displaying time series
- Displaying geospatial data
- Other charts
- Uncertainty

RELATIONSHIPS. SCATTERPLOTS

- Two values, no keys
 - Data
 - 2 quantitative values (e.g., horsepower, miles per gallon)
 - Marks: point
 - Using position
 - Channels
 - Position
 - Task
 - Find correlations

RELATIONSHIPS. SCATTERPLOTS

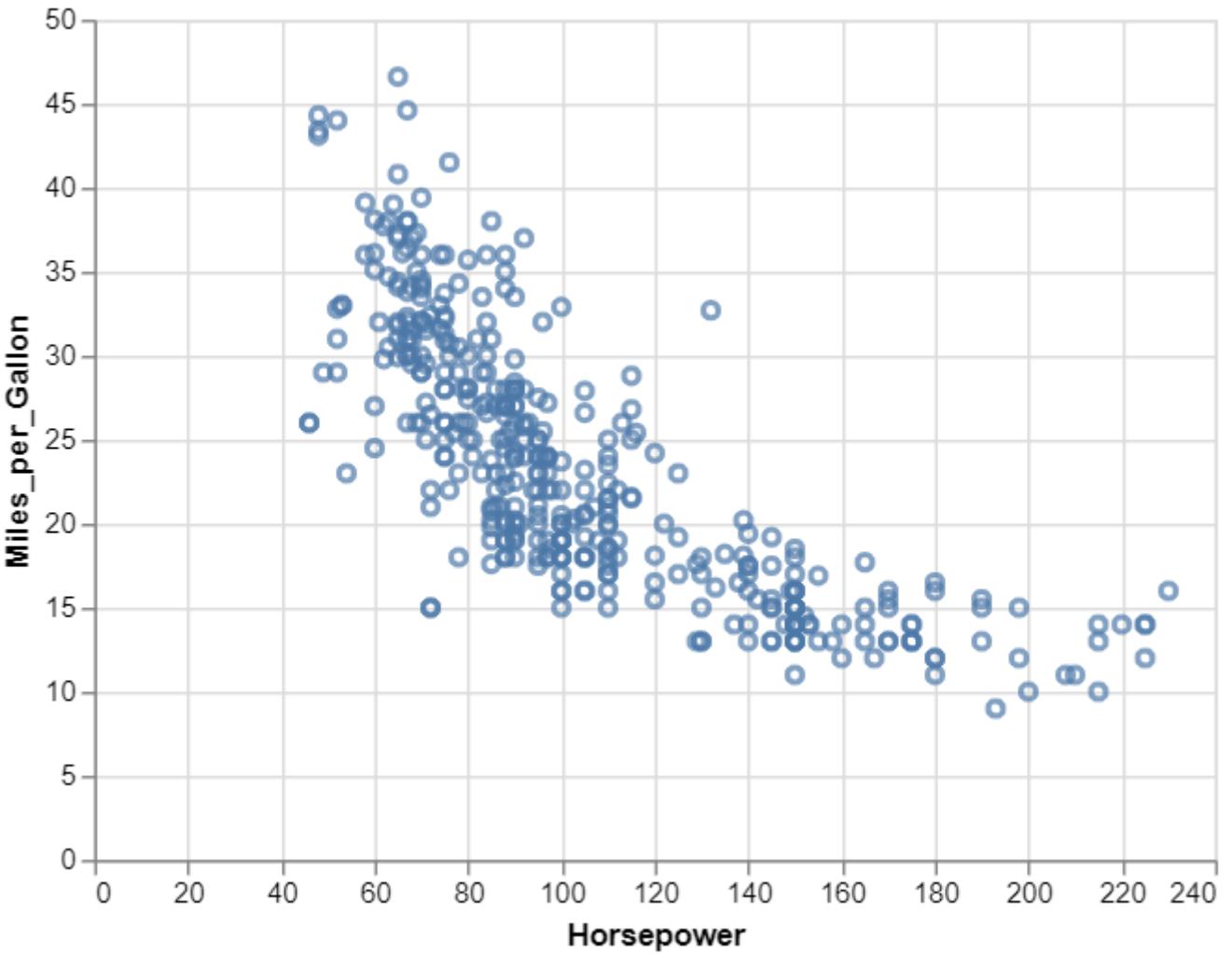


RELATIONSHIPS. SCATTERPLOTS

```
from vega_datasets import data

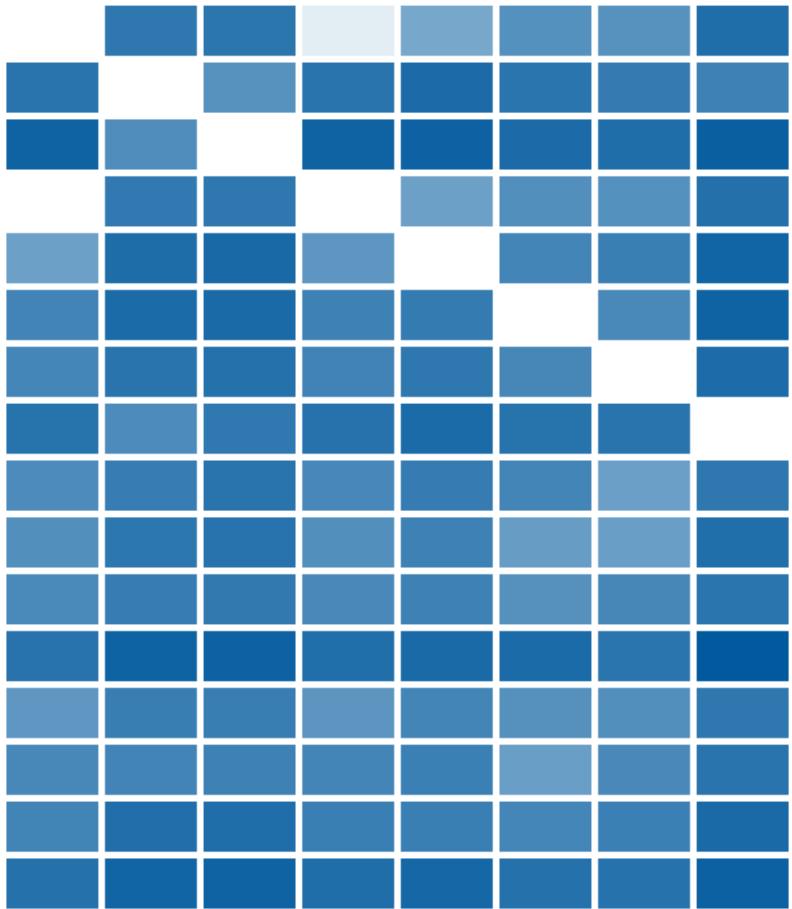
url = data.cars.url

alt.Chart(url).mark_point().encode(
    x='Horsepower:Q',
    y='Miles_per_Gallon:Q'
)
```



RELATIONSHIPS. HEAT MAPS

- Two keys, one value
 - data
 - 2 categorical attributes (article1, article2)
 - 1 quant attribute (difference)
 - Marks: area
 - Separate and align in 2D matrix
 - Indexed by 2 categorical attributes
 - Channels
 - Color by quantitative attribute (ordered colormap)
 - Task
 - Find clusters, outliers
 - Scalability
 - 1M items, 100s of category levels, ~10 quant attribute levels

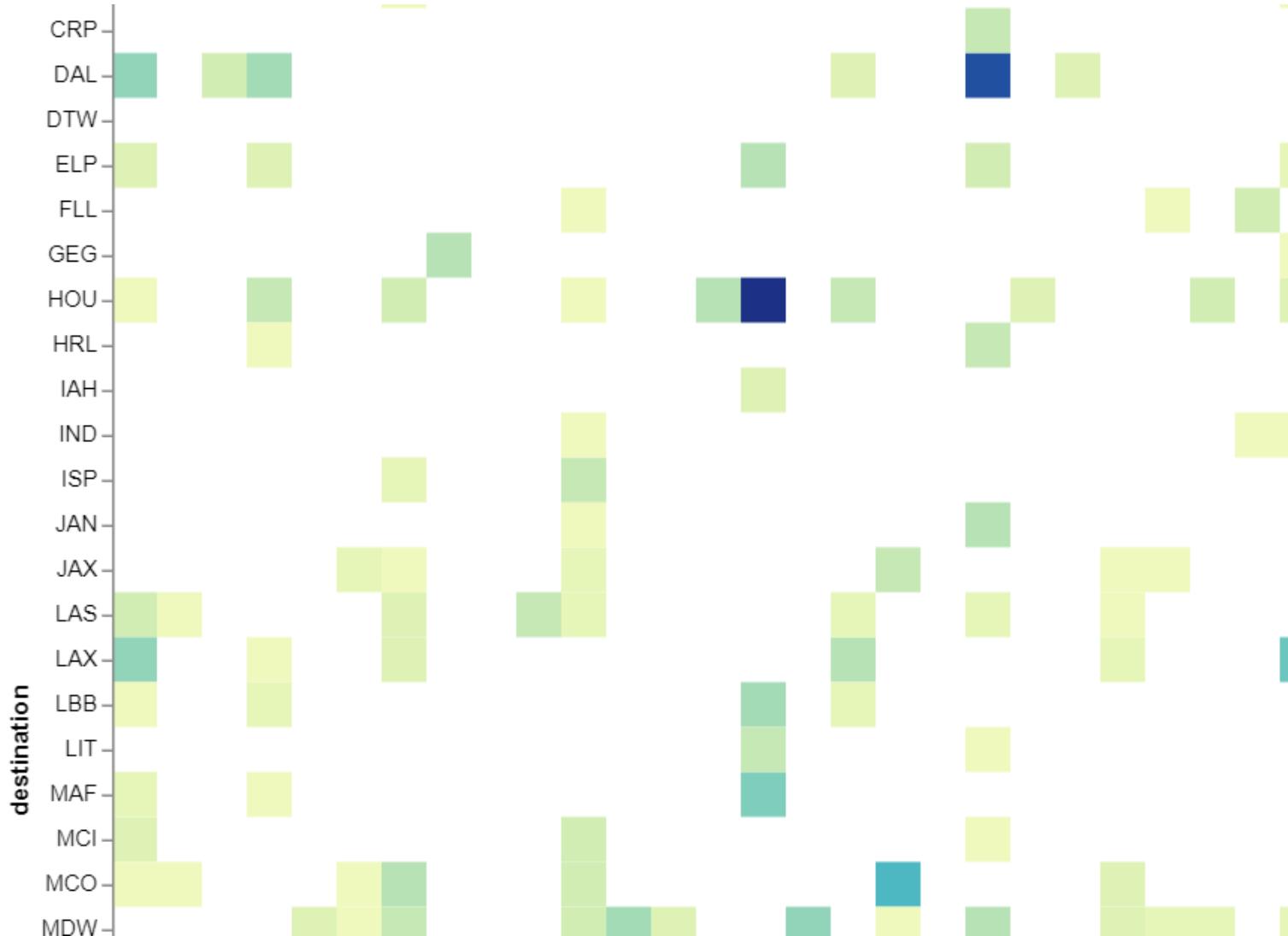


RELATIONSHIPS. HEAT MAPS

```
import altair as alt
from vega_datasets import data

source = data.flights_2k()

alt.Chart(source).mark_rect().encode(
    x='origin:O',
    y='destination:O',
    color = 'count():Q'
)
```



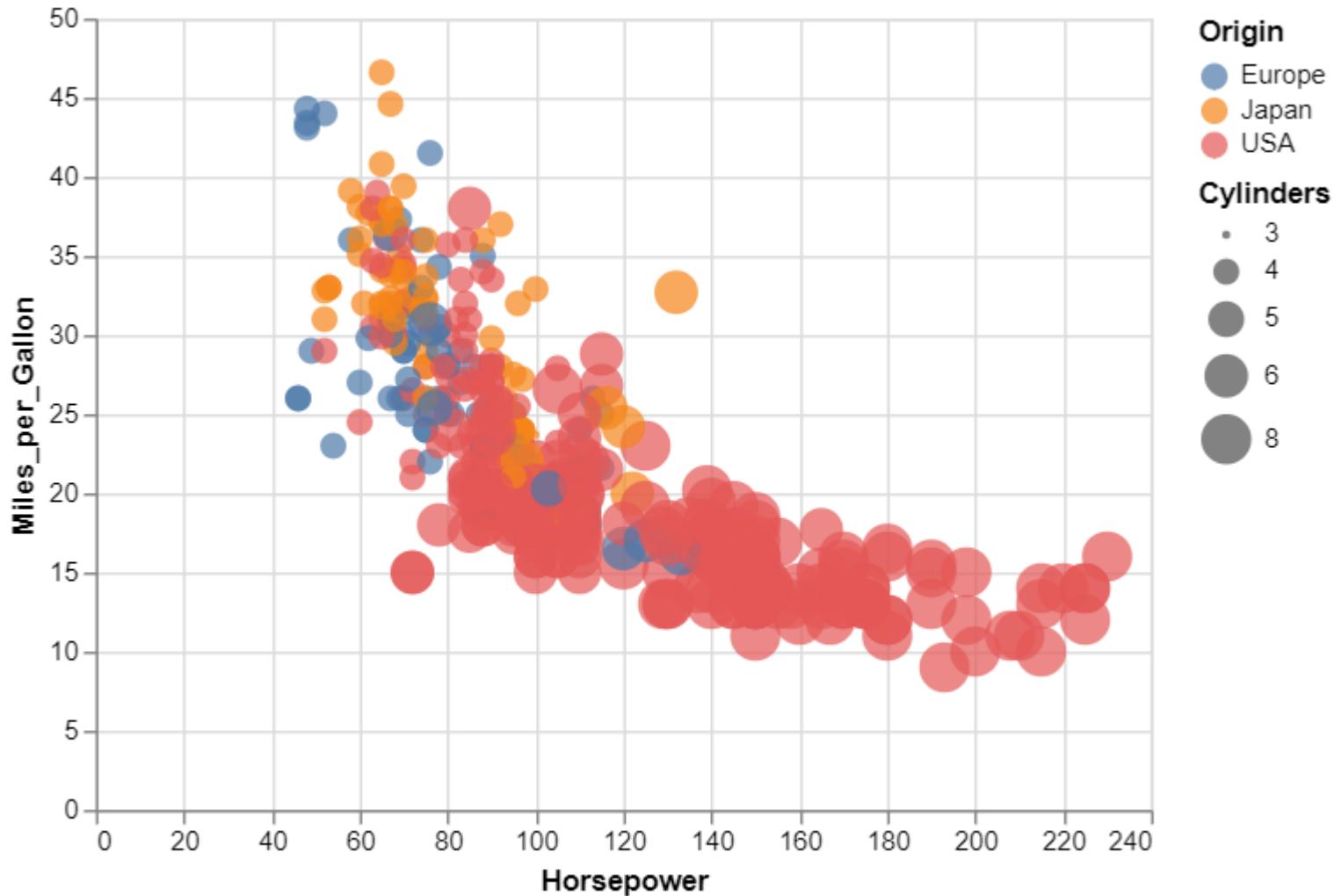
RELATIONSHIPS. BUBBLE CHARTS

- Scatterplots with more attributes as other visual channels
 - Size
 - Color
 - ...

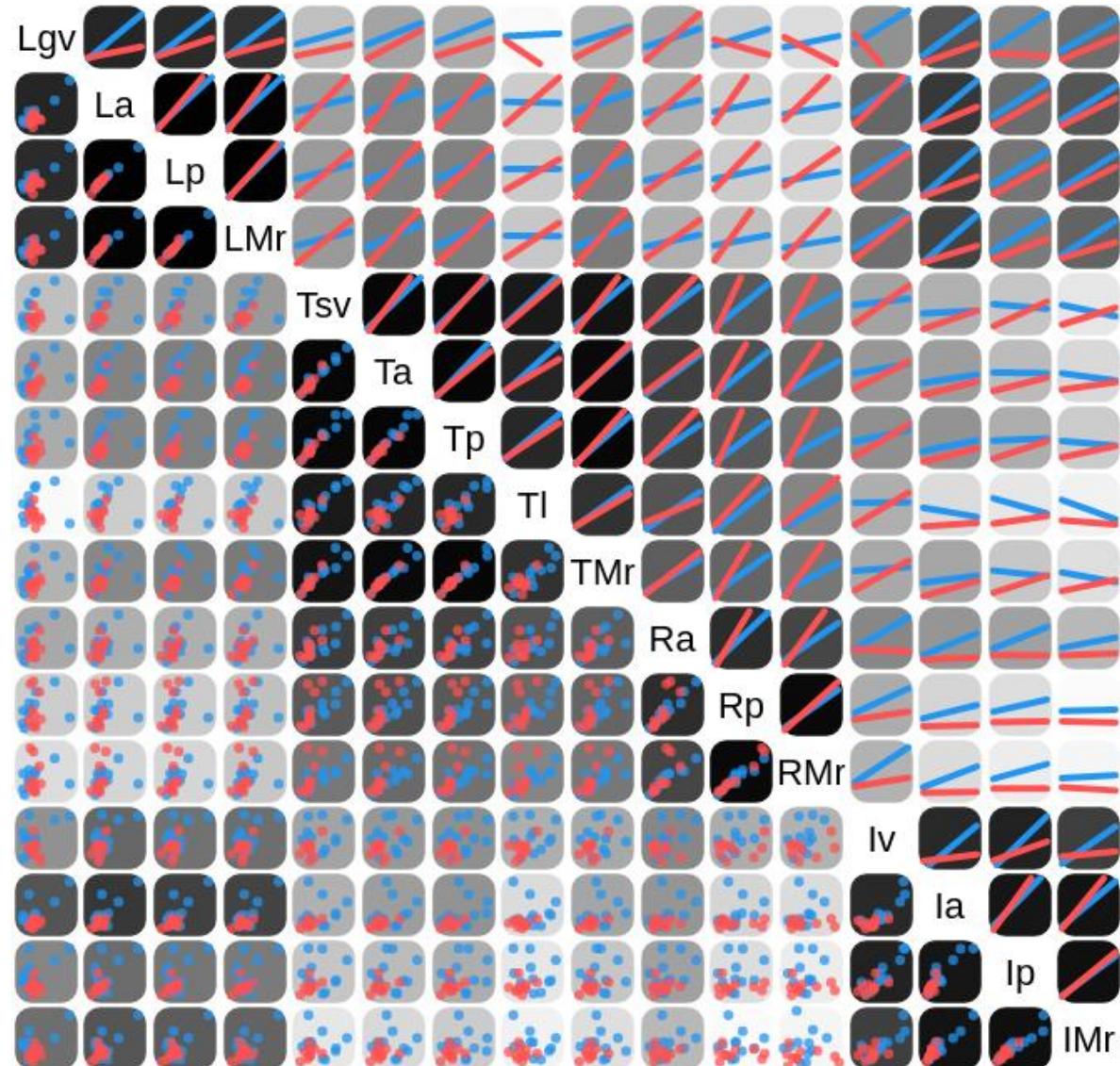
RELATIONSHIPS. BUBBLE CHARTS

```
url = data.cars.url

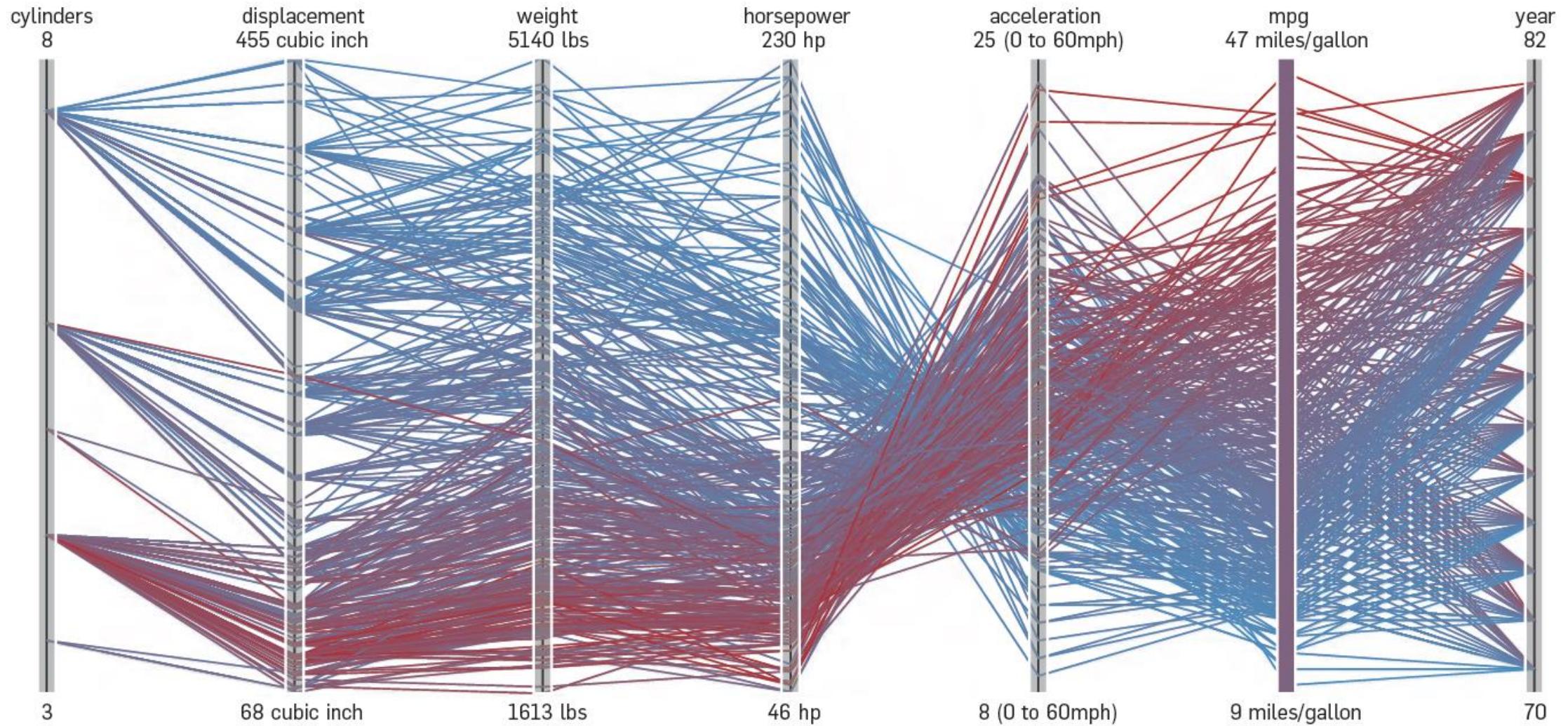
alt.Chart(url).mark_circle().encode(
    x='Horsepower:Q',
    y='Miles_per_Gallon:Q',
    size = 'Cylinders:O',
    color = 'Origin:N'
)
```



RELATIONSHIPS. SCATTERPLOT MATRICES (SPLOM)

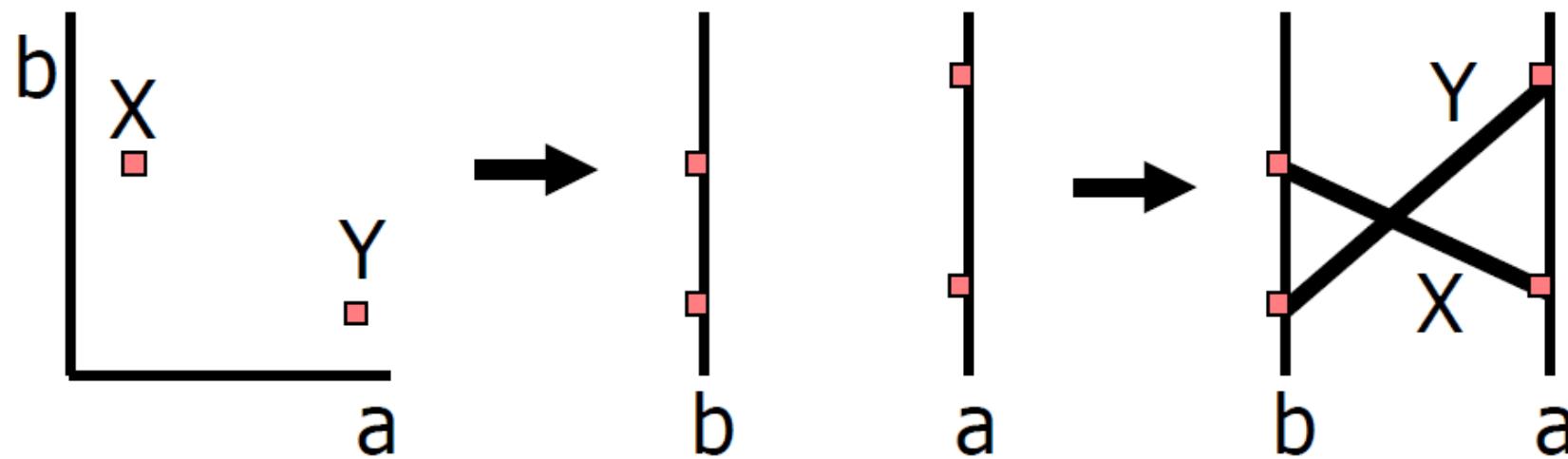


RELATIONSHIPS. PARALLEL COORDINATE PLOTS



RELATIONSHIPS. PARALLEL COORDINATE PLOTS

- All axes parallel
- Each sample is a line in this space
- Axes scaled to min/max range of data
- All dimensions can be shown at the same time
 - Positive / negative correlation == parallel / crossing lines



RELATIONSHIPS. PARALLEL COORDINATE PLOTS

- Data:
 - Several keys, quantitative/categorical values
- Scalability:
 - Up to one dozen of keys
 - Hundreds of items (not thousands)
 - Larger sizes require special techniques, e.g. blending, or hierarchical approaches

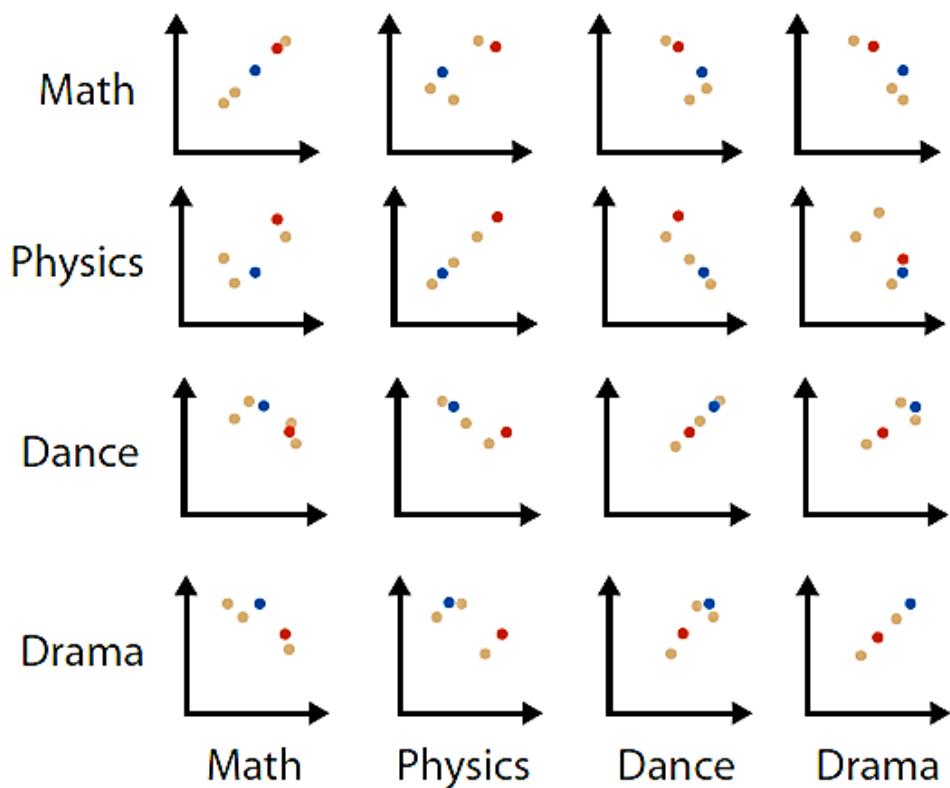
RELATIONSHIPS. PARALLEL COORDINATE PLOTS

- Discussion. Relatively compact

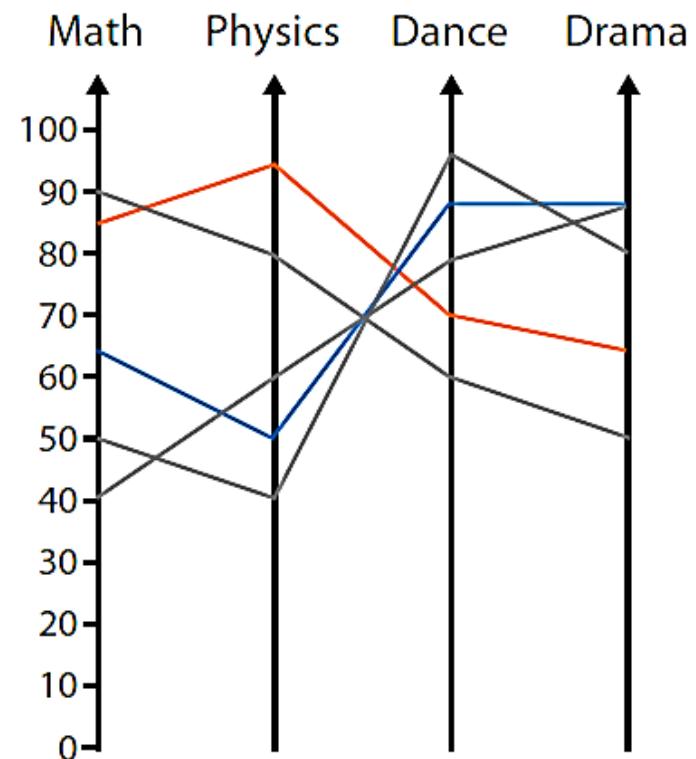
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatterplot Matrix



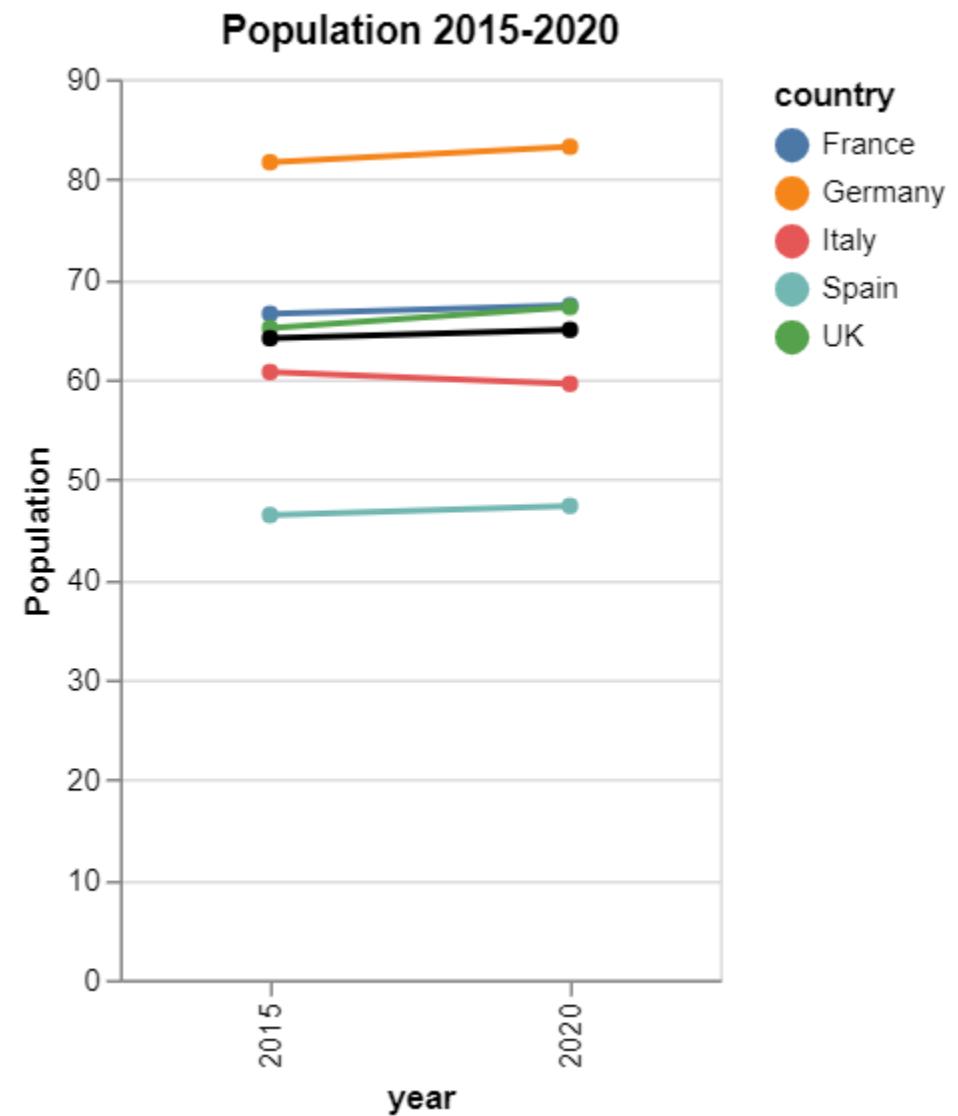
Parallel Coordinates



RELATIONSHIPS. SLOPE CHARTS

- Data:
 - Two values (typically) in two moments of time
- Marks:
 - Points (connected by a line)
- Scalability:
 - Dozens of items?

RELATIONSHIPS. SLOPE CHARTS



RELATIONSHIPS. SLOPE CHARTS

```
ch1 = alt.Chart(data).mark_line(point = True).encode(  
    x='year:N',  
    y=alt.Y('population:Q', title ='Population in millions'),  
    color='country'  
).properties(title = 'Population 2015-2020', width = 200)
```

```
ch1
```

OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- *Displaying proportions*
- *Displaying relationships*
- **Displaying time series**
- *Displaying geospatial data*
- *Other charts*
- *Uncertainty*

REPRESENTATIONS. LINE CHARTS

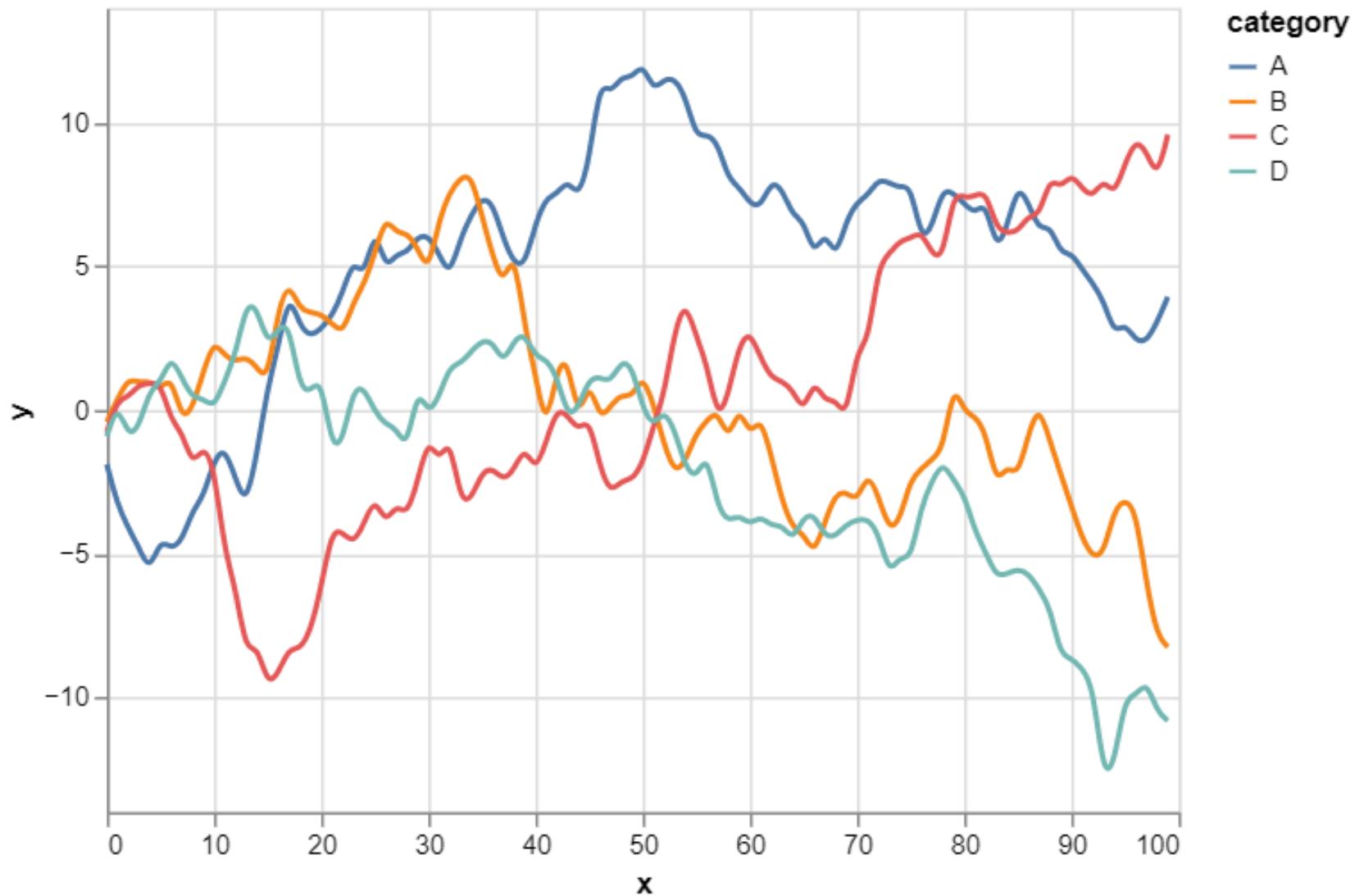
- One key, one value
- Data: 2 quant attributes
- Mark: points
 - Line connection marks between them
- Channels
 - Aligned lengths to express quant value
 - Separated and ordered by key attribute into horizontal regions
- Tasks: Trends
 - Connection marks emphasize ordering of items

TIME SERIES. LINE CHARTS

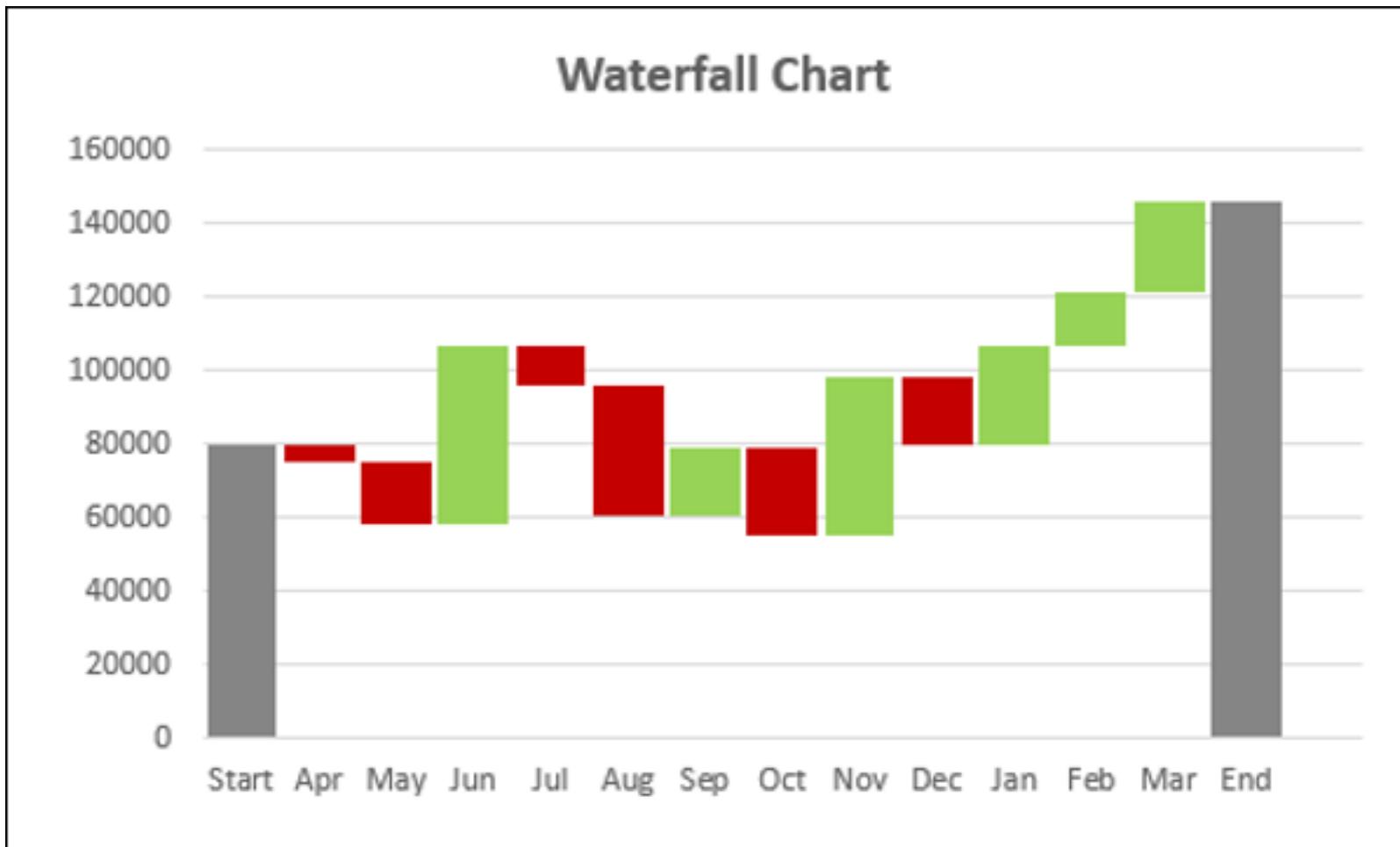
```
np.random.seed(35)
source = pd.DataFrame(np.cumsum(np.random.randn(100, 4), 0).round(2),
                      columns=['A', 'B', 'C', 'D'],
                      index=pd.RangeIndex(100, name='x'))
source = source.reset_index().melt('x', var_name='category', value_name='y')

alt.Chart(source).mark_line(interpolate='basis').encode(
    x='x:Q',
    y='y:Q',
    color='category:N'
)
```

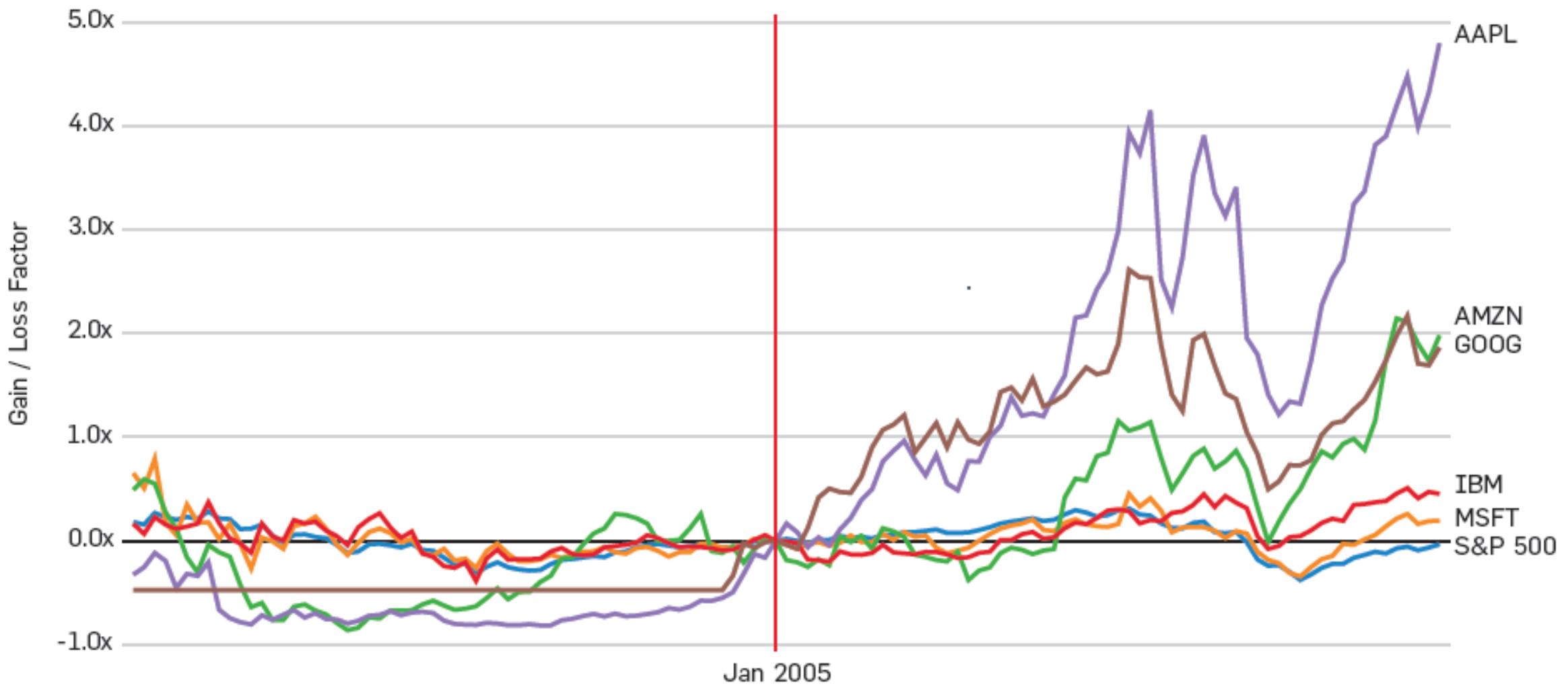
TIME SERIES. LINE CHARTS



TIME SERIES. WATERFALL CHART



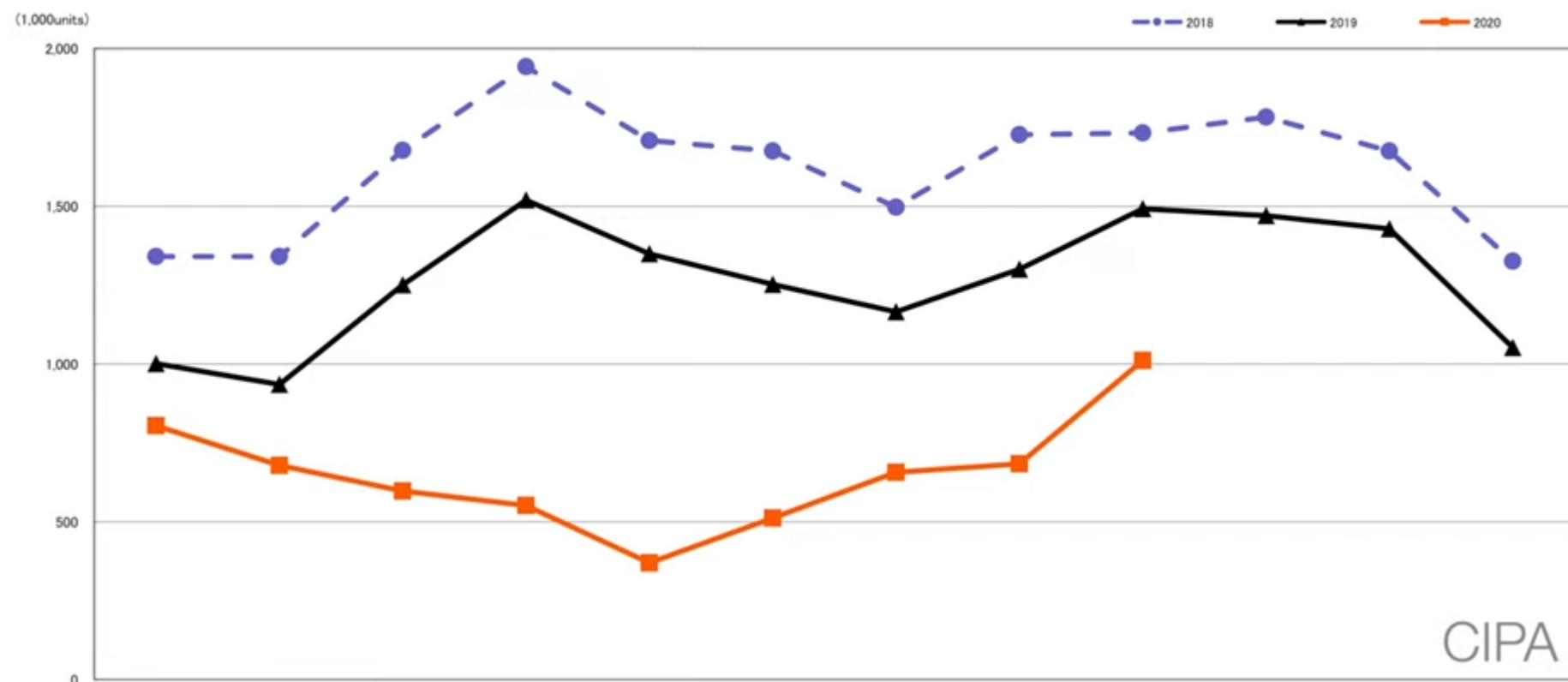
TIME SERIES. INDEX CHARTS



TIME SERIES. INDEX CHARTS

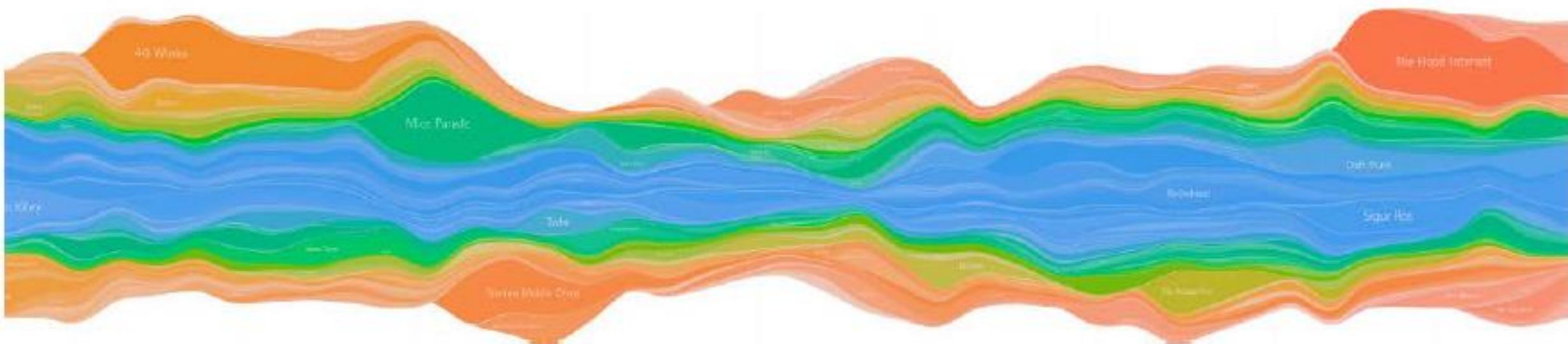
- Indexing over the time axis

Quantity of Total Shipment of DSC [Worldwide]
Comparison of 2018, 2019 and 2020 :Jan.-Aug.



TIME SERIES. STREAMGRAPH

- Generalized stacked graph



TIME SERIES. STREAMGRAPH

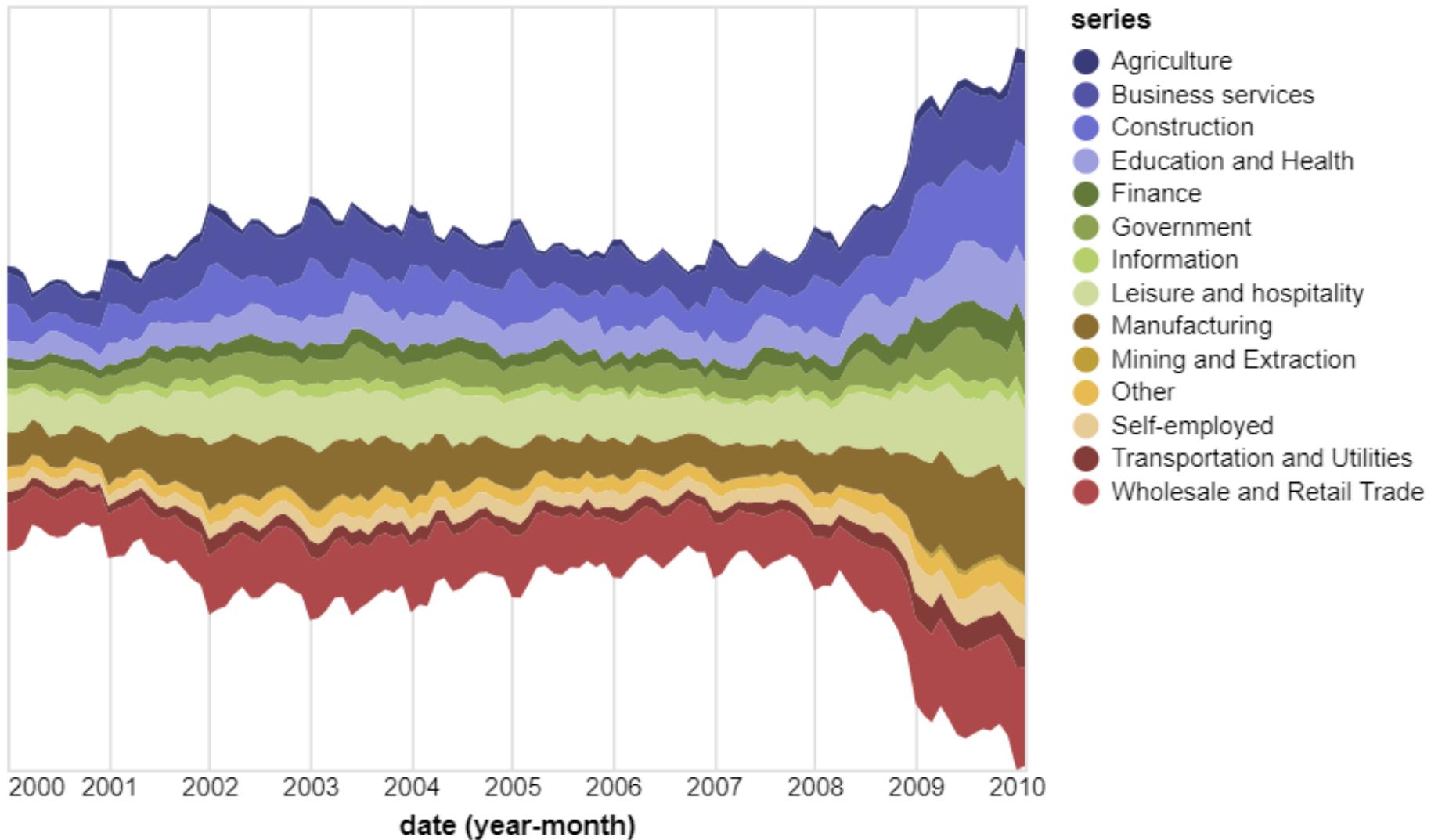
```
import altair as alt
from vega_datasets import data

source = data.unemployment_across_industries.url

selection = alt.selection_multi(fields=['series'], bind='legend')

alt.Chart(source).mark_area().encode(
    alt.X('yearmonth(date):T', axis=alt.Axis(domain=False, format='%Y',
                                                tickSize=0)),
    alt.Y('sum(count):Q', stack='center', axis=None),
    alt.Color('series:N', scale=alt.Scale(scheme='category20b')),
    opacity=alt.condition(selection, alt.value(1), alt.value(0.2))
).add_selection(
    selection
)
```

TIME SERIES. STREAMGRAPH



OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- *Displaying proportions*
- *Displaying relationships*
- *Displaying time series*
- **Displaying geospatial data**
- Other charts
- Uncertainty

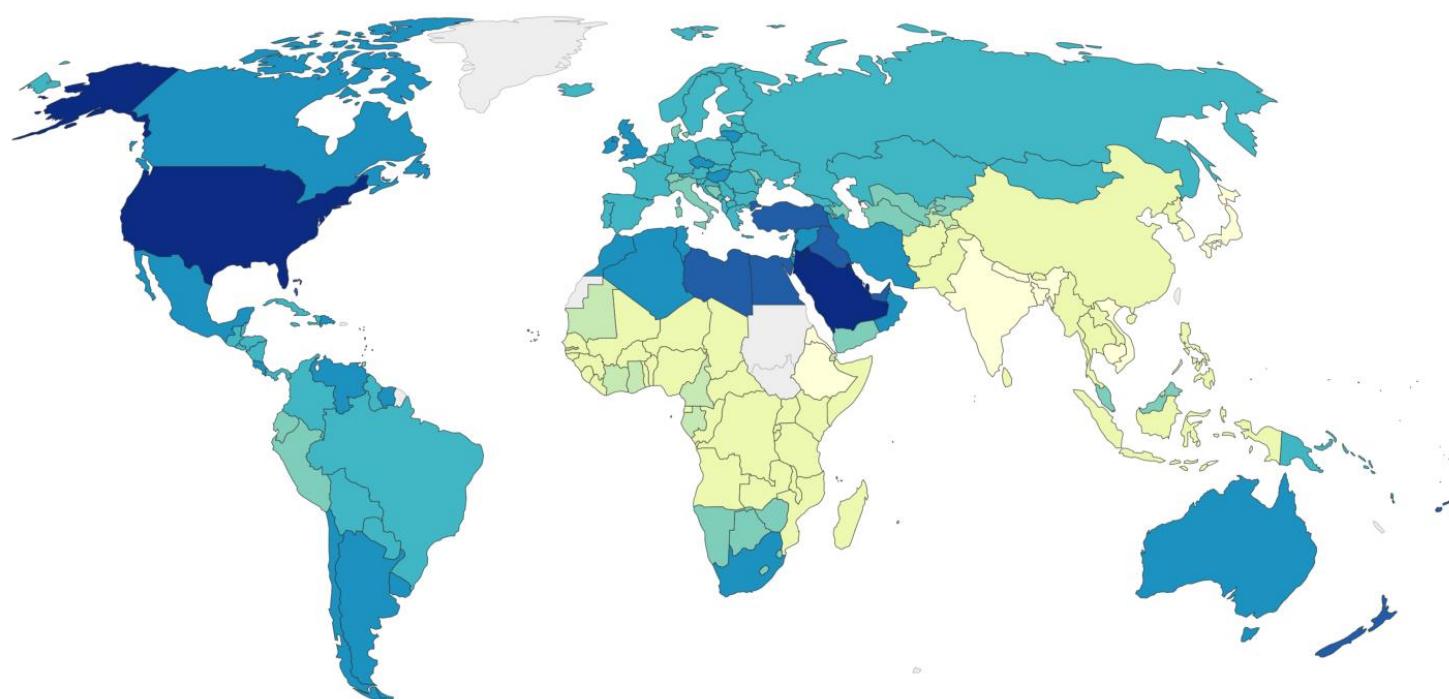
GEOSPATIAL DATA. CHOROPLETH MAPS

- Choropleth maps (Greek: choro = area, pleth = value)
 - Assume that the mapped attribute is uniformly distributed in the regions
 - Used to emphasize the spatial distribution of one or more geographic attributes.
 - Data normalization and color or grayscale mapping are important design decisions
 - Similar types: dasymetric, isarithmic, isometric, isopleth...

GEOSPATIAL DATA. CHOROPLETH MAPS

Share of adults defined as obese, 2016

Percentage of adults aged 18+ years old who are defined as obese based on their body-mass index (BMI). BMI is a person's weight in kilograms (kg) divided by his or her height in metres squared. A BMI greater than or equal to 30 is defined as obese.



GEOSPATIAL DATA. CHOROPLETH MAPS

- Issues

- Larger regions → larger importance
- Must (commonly) take into account population density
- Color issues

PROBLEMS WITH ELECTION MAPS

- <https://www.nytimes.com/interactive/2020/10/30/opinion/election-results-maps.html>

ELECTIONS RESULT

306

Joseph R. Biden Jr. ✓

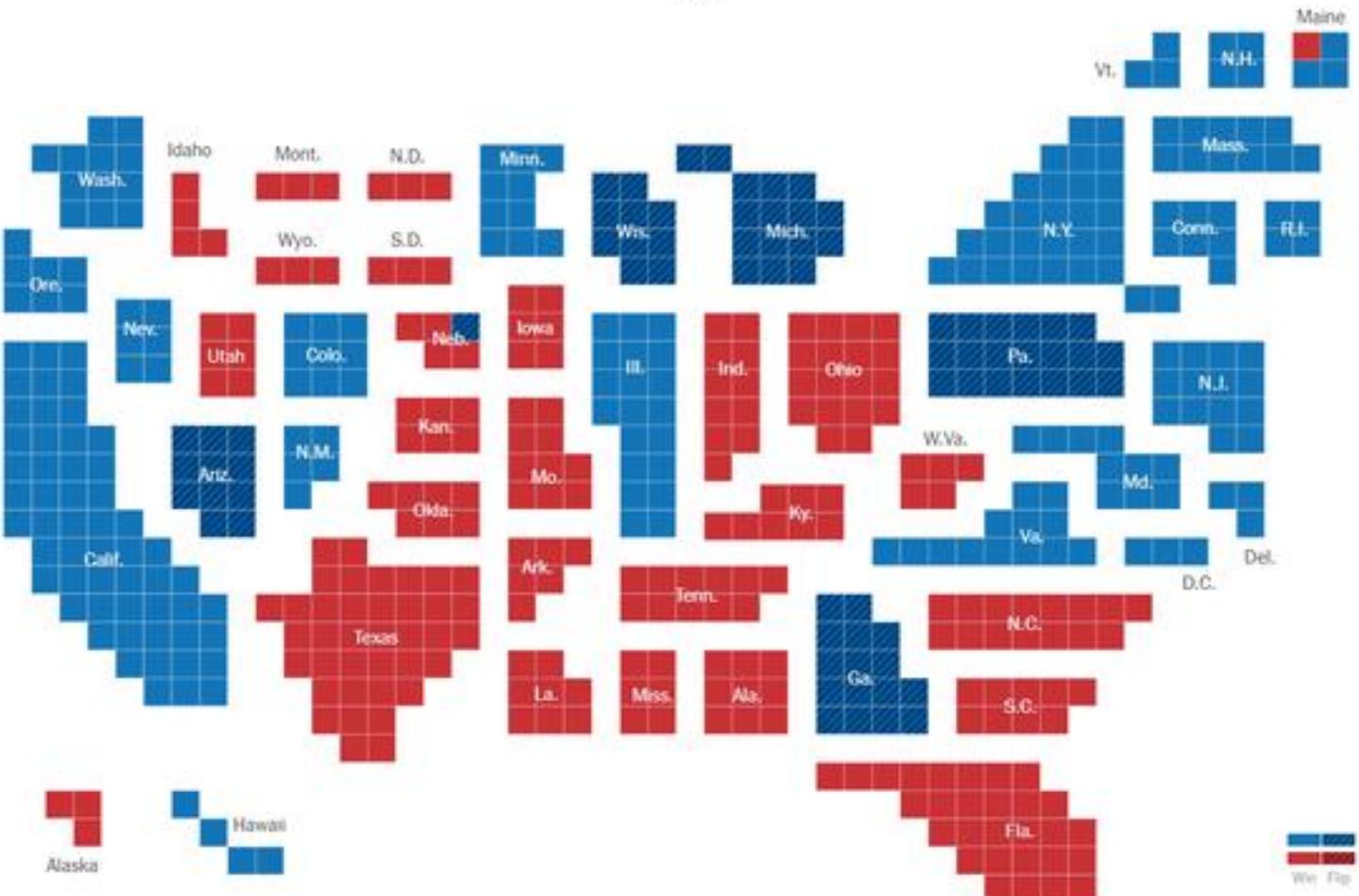
79,554,207 votes (51.0%)

232

Donald J. Trump

270
TO WIN

73,611,180 votes (47.2%)



ELECTIONS RESULT

306

Joseph R. Biden Jr. 

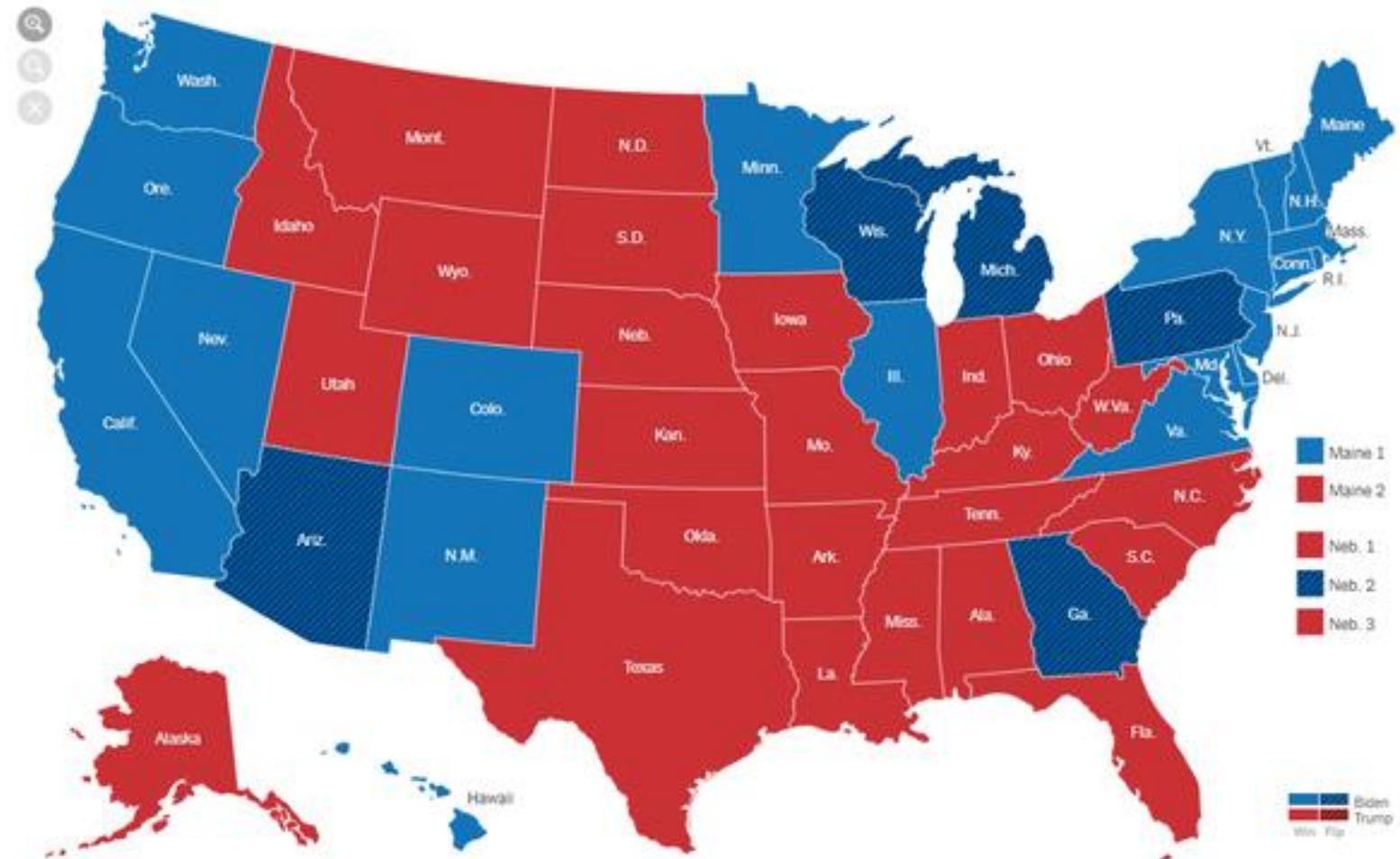
79,554,222 votes (51.0%)

232

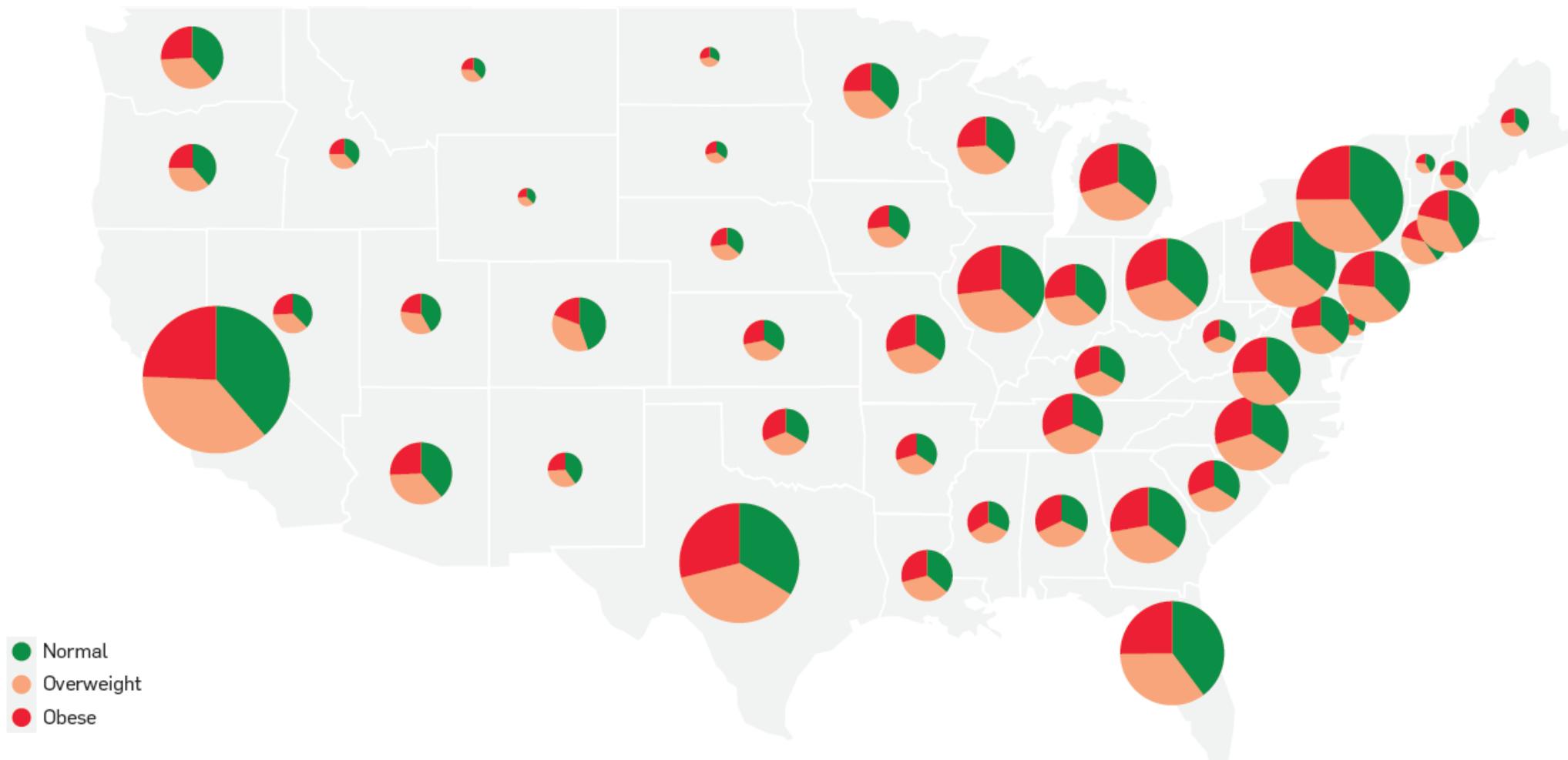
Donald J. Trump

270
TO WIN

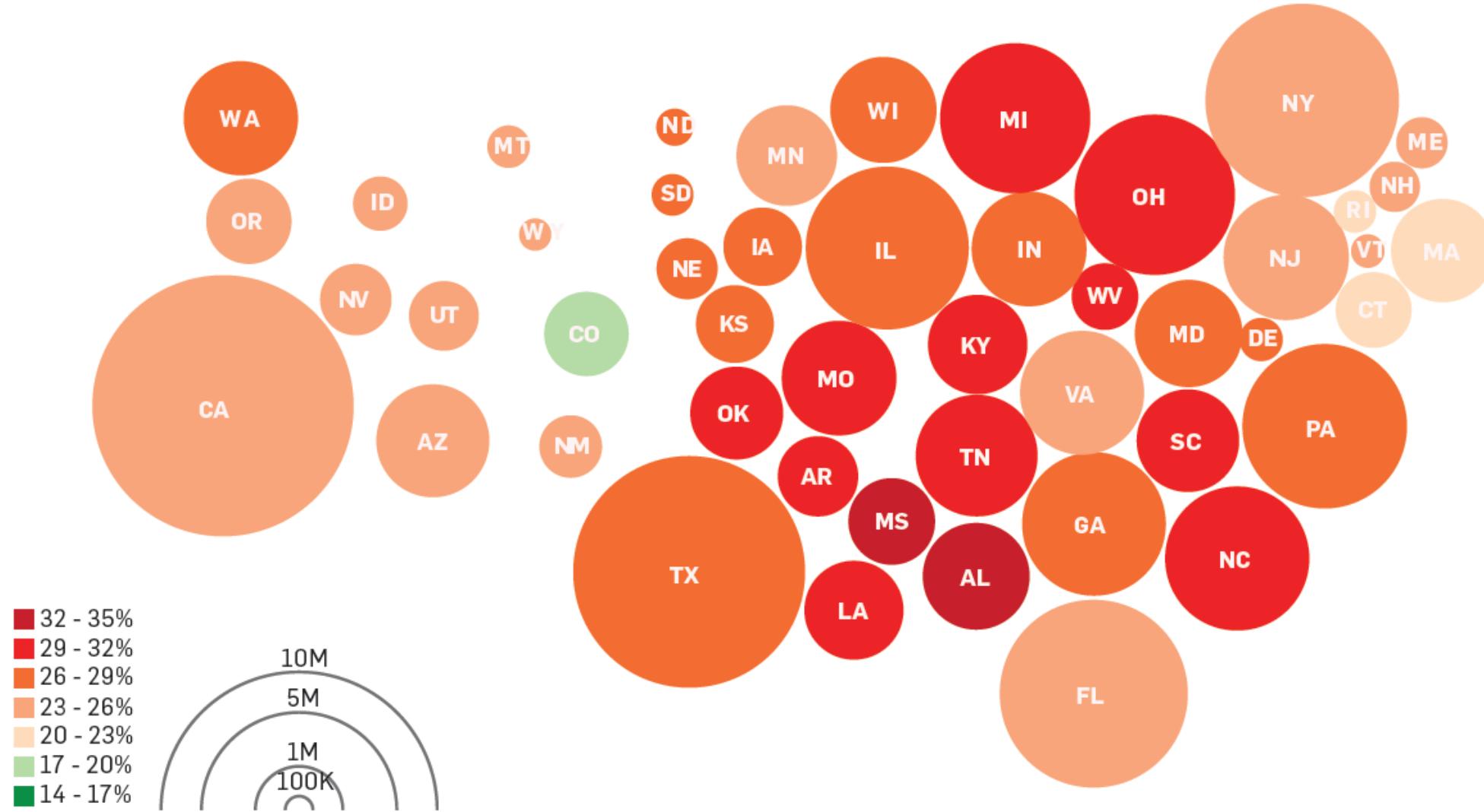
73,611,220 votes (47.2%)



GEOSPATIAL DATA. GRADUATED SYMBOL MAPS



GEOSPATIAL DATA. CARTOGRAMS



States where new cases have increased

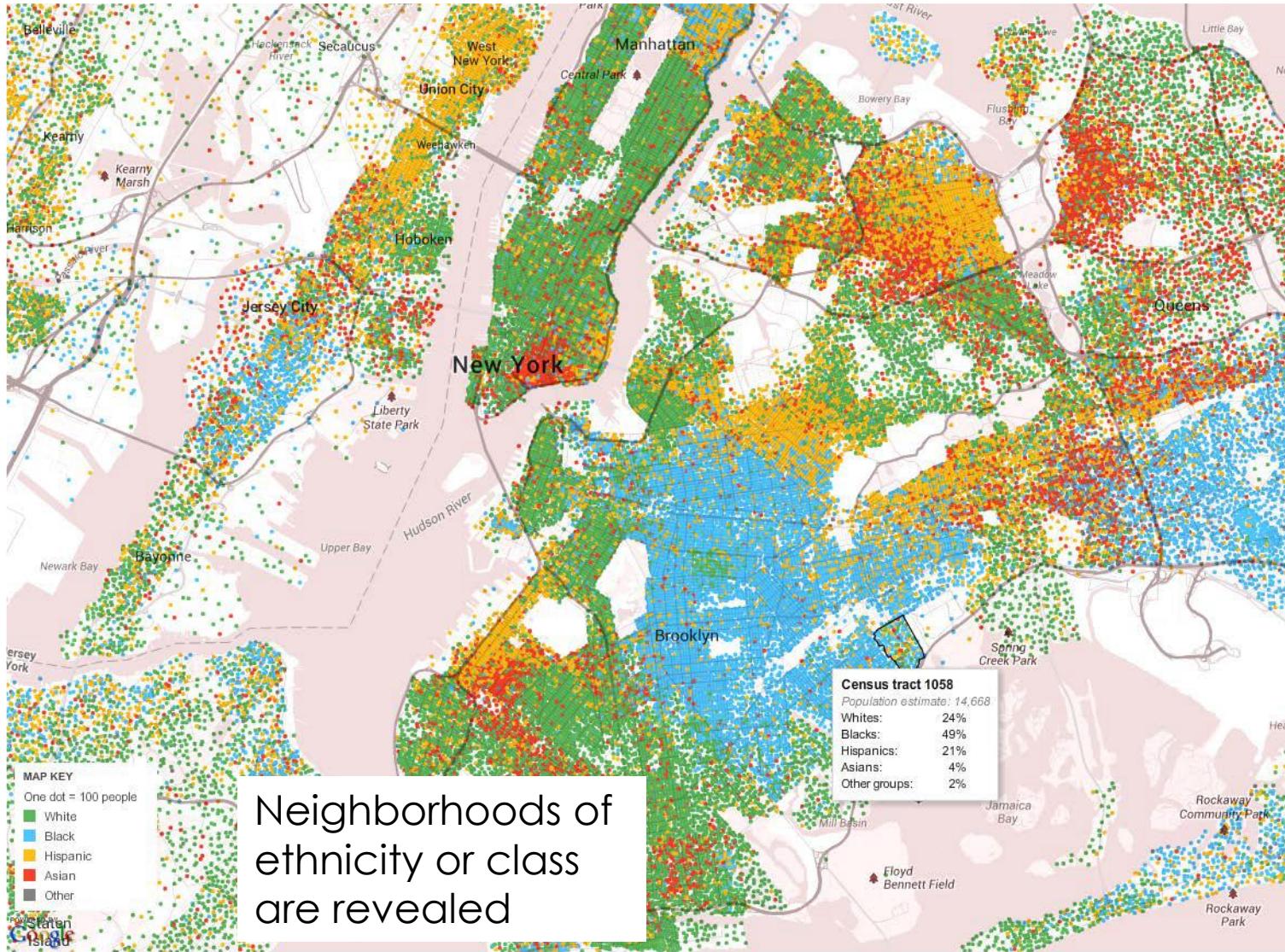
↗ Daily new cases, seven-day average



<https://www.nytimes.com/interactive/2020/05/07/us/coronavirus-states-reopen-criteria.html>

GEOSPATIAL DATA. DOT MAPS

- Census data



GEOSPATIAL DATA. DOT MAPS

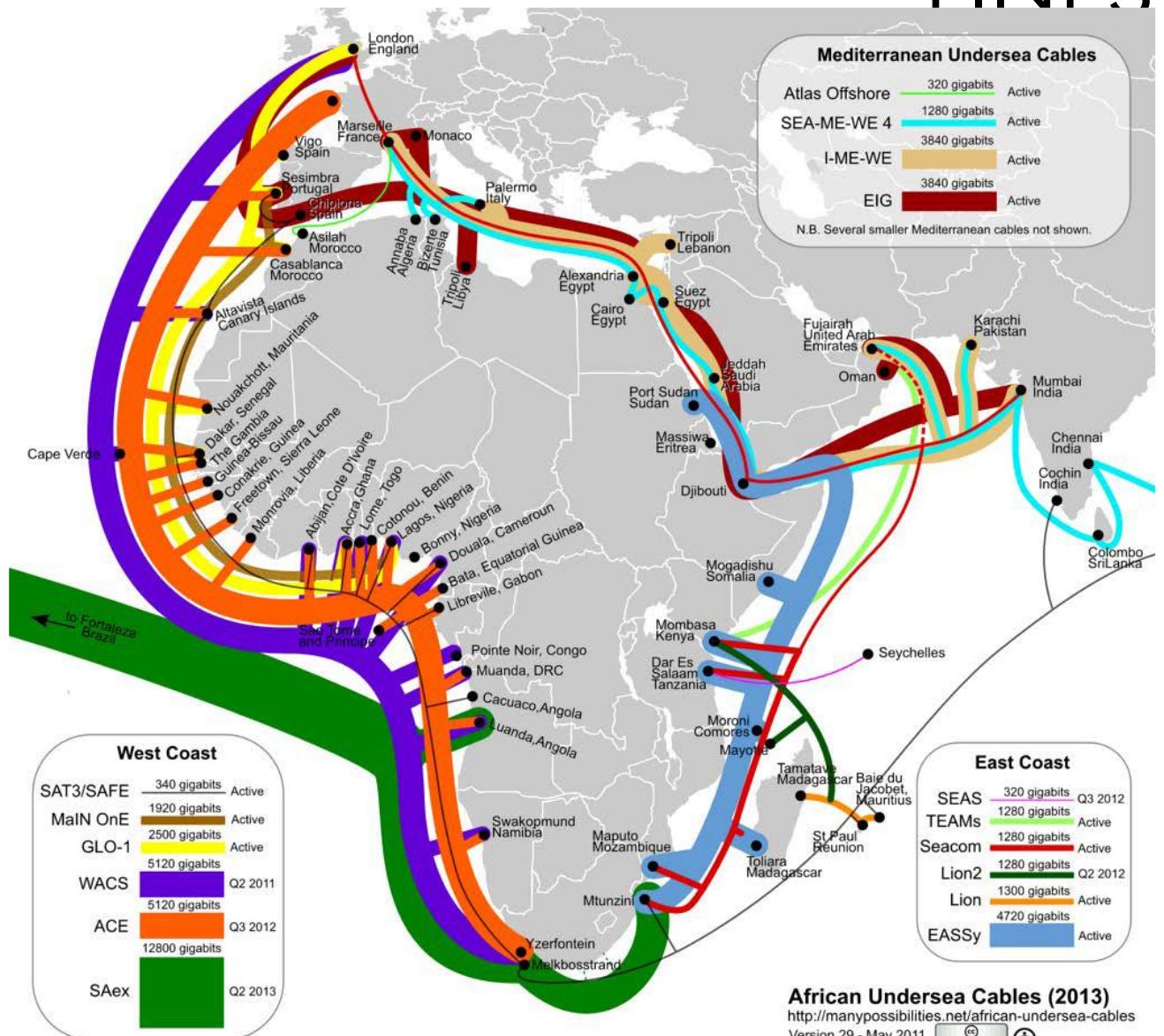
- Dot maps. Issues:
 - If the size of the symbol is used to represent a quantitative parameter, scaling may present perception issues
 - Perceived size also depends on their local neighborhood
 - If color is used to represent a quantitative parameter, problems of color perception also appear
 - Large data cause overlap or overplotting problems
 - Especially in highly populated areas, while low-population areas are virtually empty

GEOSPATIAL DATA. PIXEL MAPS

- Repositions pixels that would otherwise overlap
 - Does not aggregate the data
 - Avoids overlap in the two-dimensional display
 - Quite intuitive result
 - Main idea of the repositioning
 - Recursively partition the data set into four subsets containing the data points in four equal-sized subregions

GEOSPATIAL DATA. LINES

- Mapping attributes to lines



OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- *Displaying proportions*
- *Displaying relationships*
- *Displaying time series*
- *Displaying geospatial data*
- **Other charts**
- Uncertainty

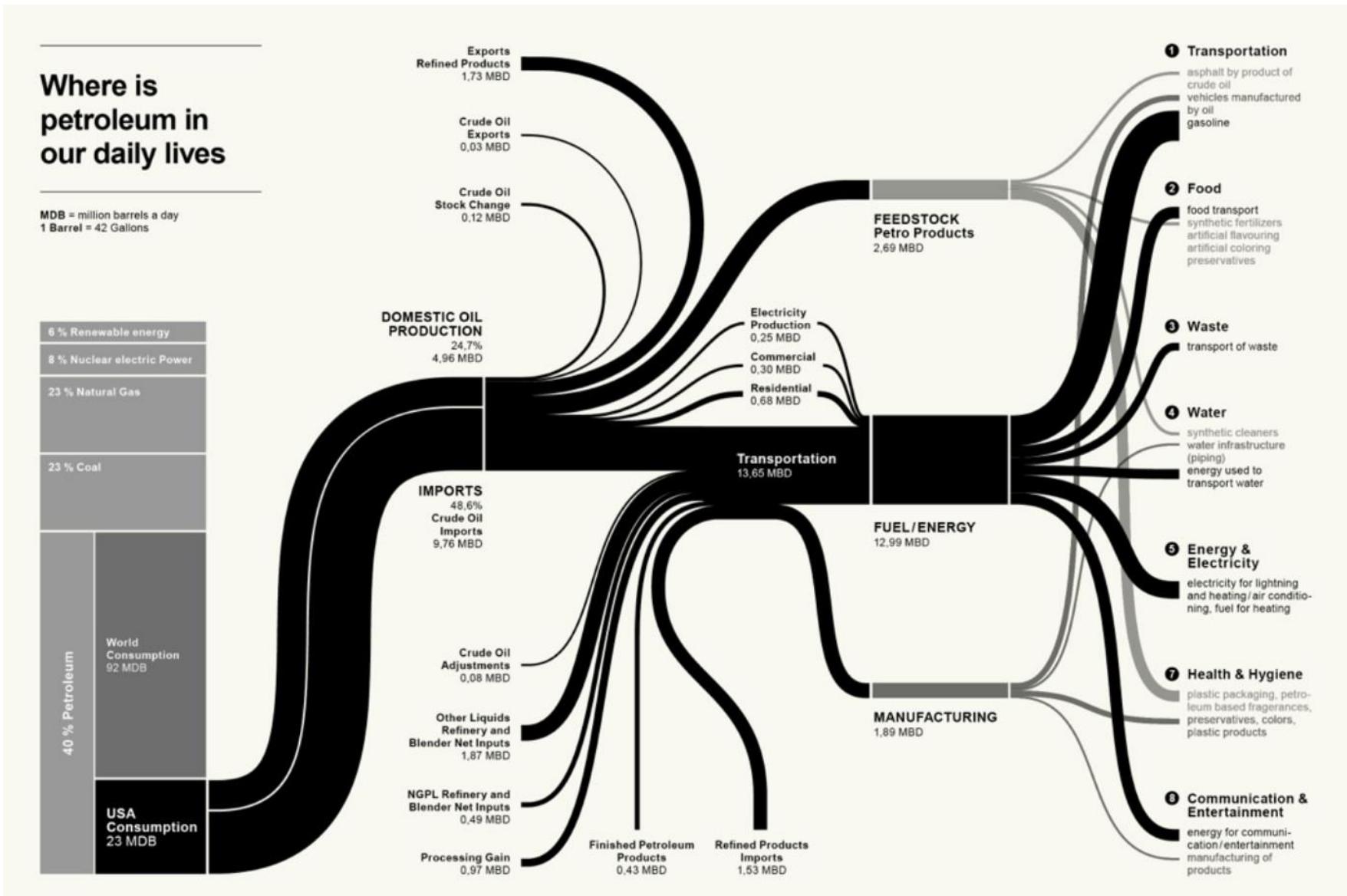
OTHER CHARTS. MULTIPLE VARIABLES. SMALL MULTIPLES

2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support



Orange and green colors correspond to states where support for vouchers was greater or less than the national average. The seven ethnicreligious categories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants. Where a category represents less than 1% of the voters of a state, the state is left blank.

OTHER CHARTS. FLOWS. SANKEY DIAGRAMS

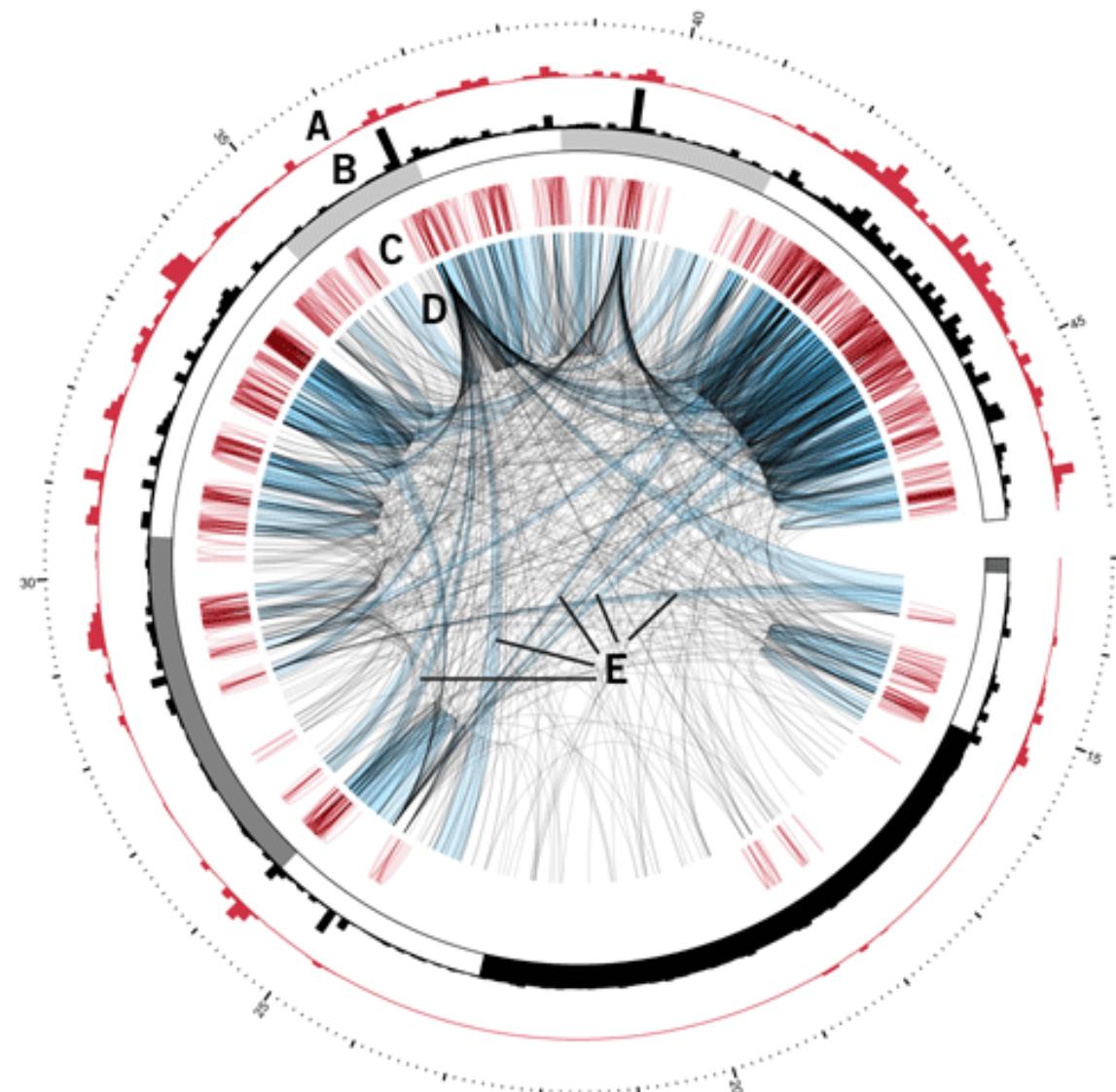


OTHER CHARTS. GRAPHS/NETWORKS

- Node-link diagrams. Circular layout. Discussion
 - Highly regularized and tidy visualization
 - Ordering of nodes possible
 - Edges or nodes never overlay other nodes
 - Easy to visually proceed along edges
 - Collinear edges for a large number of nodes
 - Very long edges and “meaningless” edge length
 - Strong regularity can obscure inherent structures
 - Very dense drawings for complex graphs

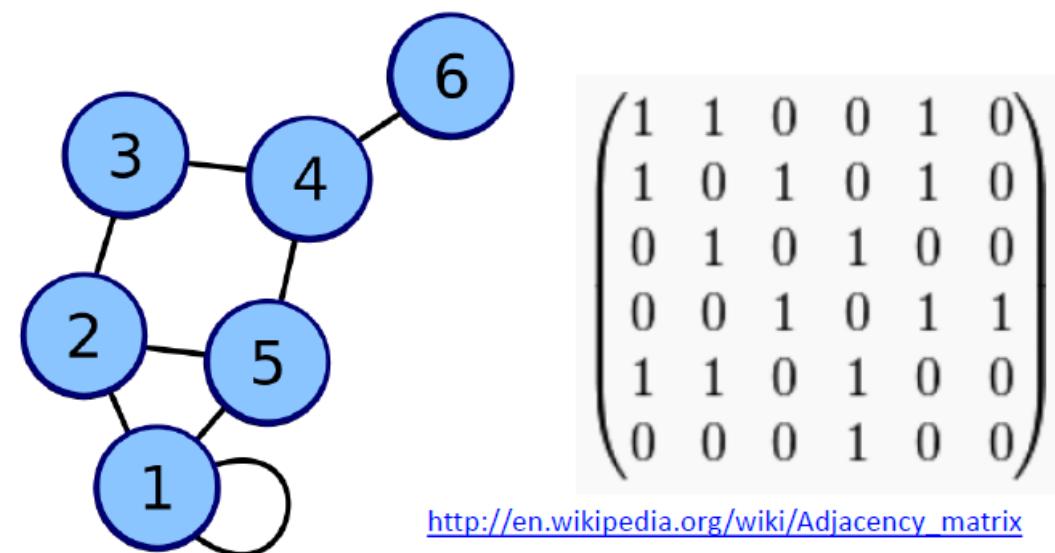
OTHER CHARTS. GRAPHS/NETWORKS

- Node-link diagrams. Circular layout. Can use external space for more data

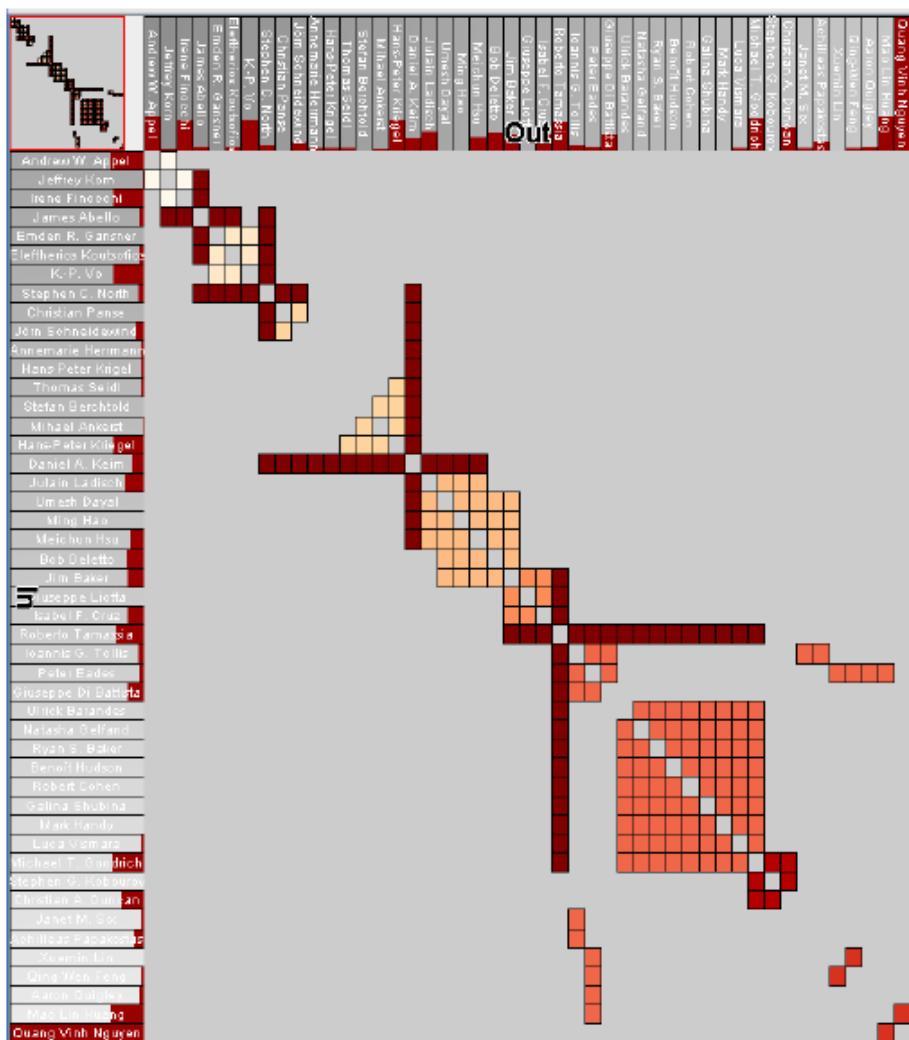
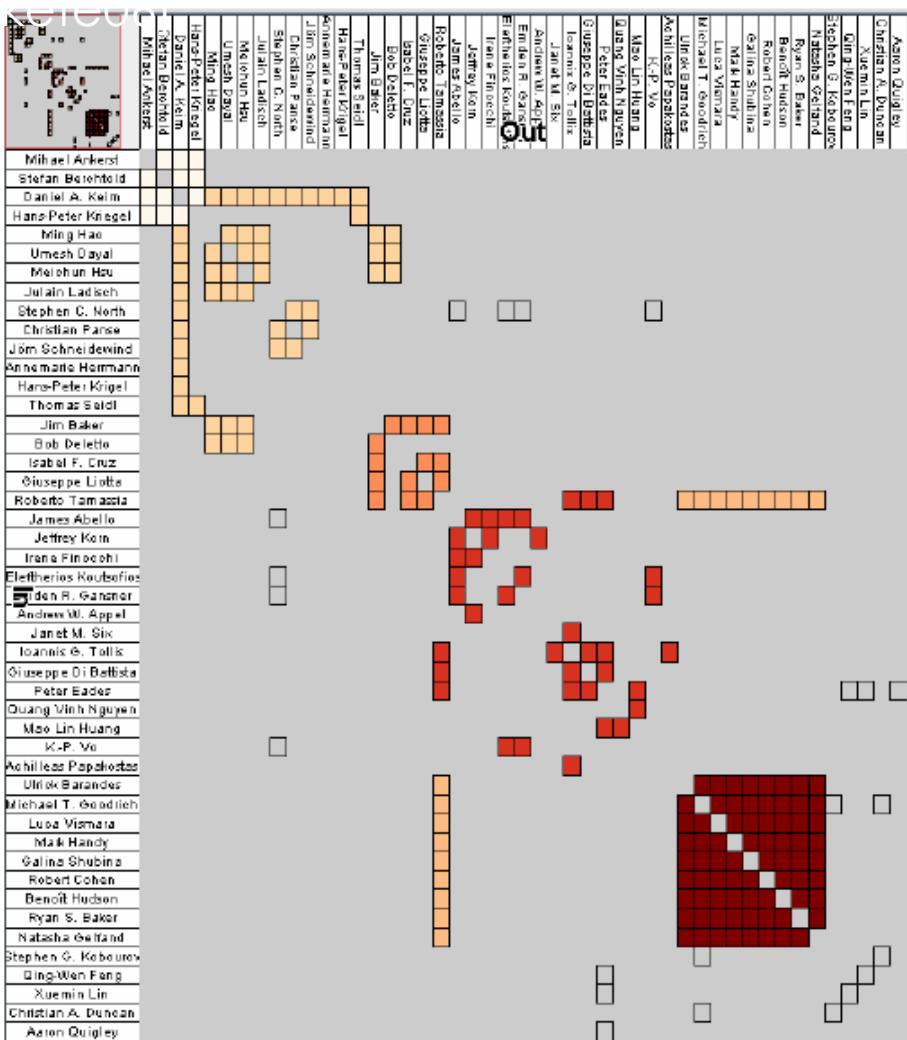


OTHER CHARTS. GRAPHS/NETWORKS

- Matrix charts:
 - Uses adjacency matrix of the graph
 - $n \times n$ adjacency matrix of graph G with n nodes:
 - Entries a_{ij} : number of edges from node i to node j
 - Entries a_{ii} : number of loops from node i to itself



OTHER CHARTS. GRAPHS/NETWORKS

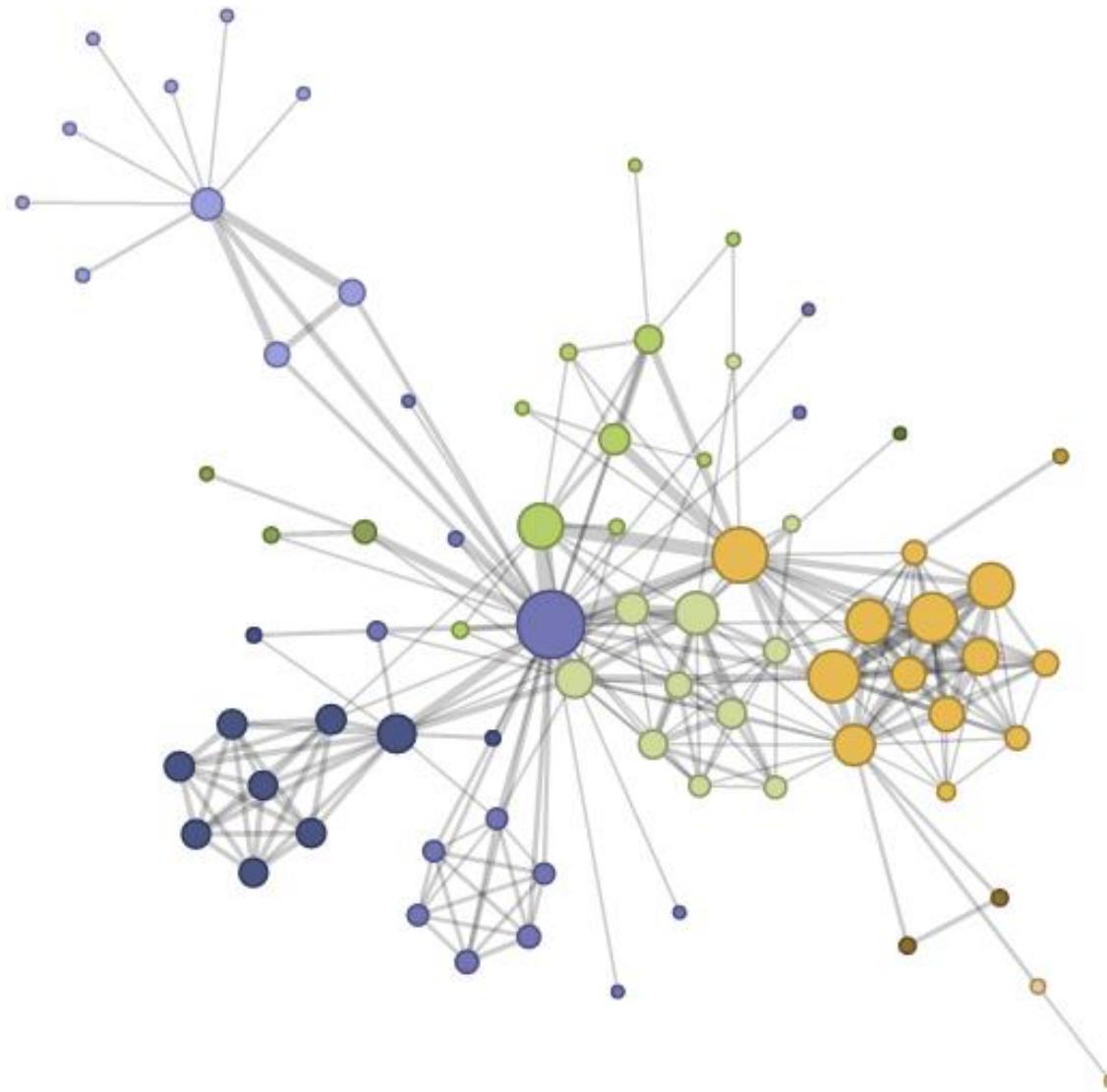


Spotting patterns benefits from node reordering (right).

OTHER CHARTS. GRAPHS/NETWORKS

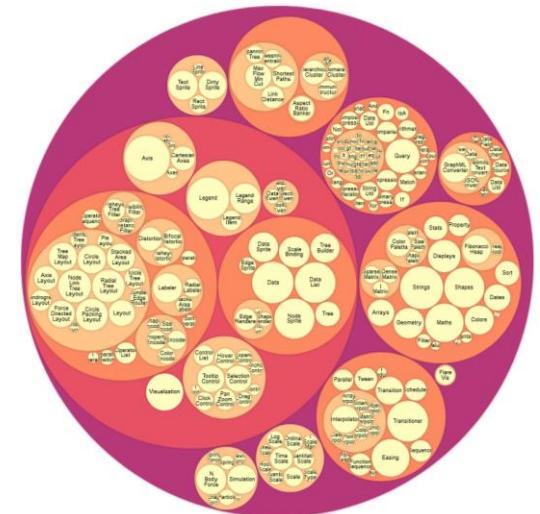
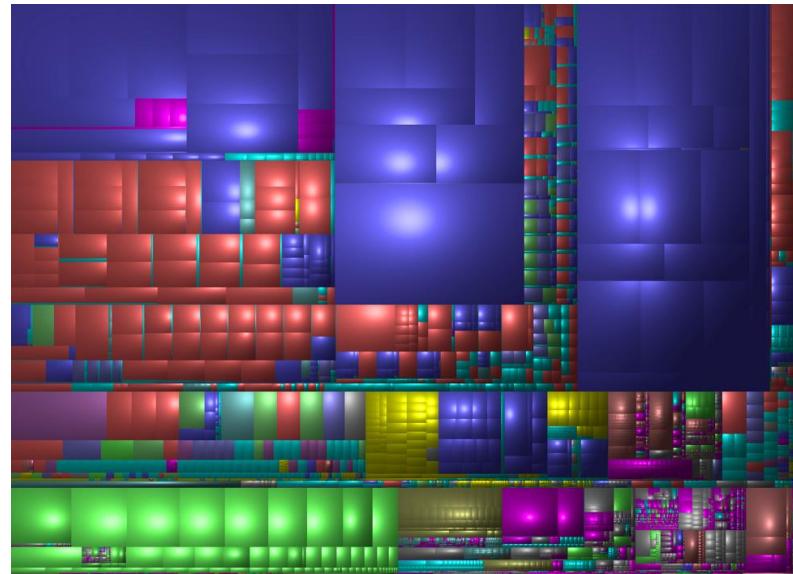
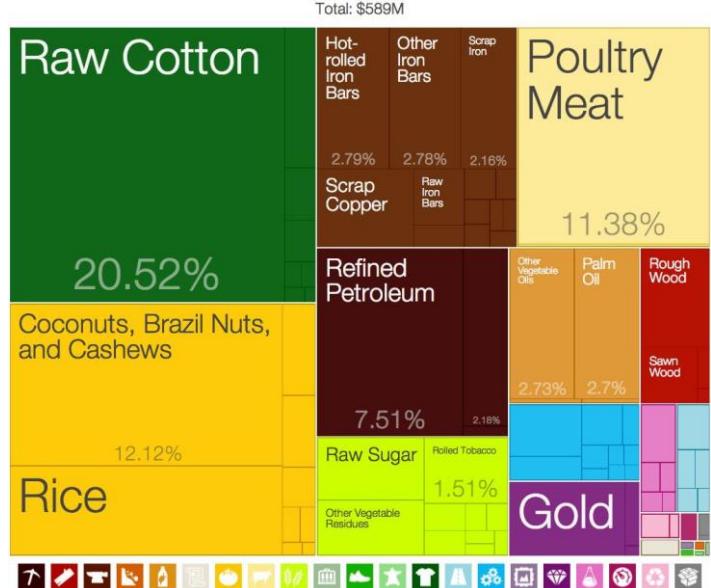
- Matrix charts. Recap. Discussion:
 - Great for dense graphs
 - Visually scalable
 - Can spot patterns and clusters
 - Need to learn them
 - Paths are hard to follow
 - Reordering is costly ($N!$ order of possible orders, an NP problem)

OTHER CHARTS. FORCE-DIRECTED LAYOUT



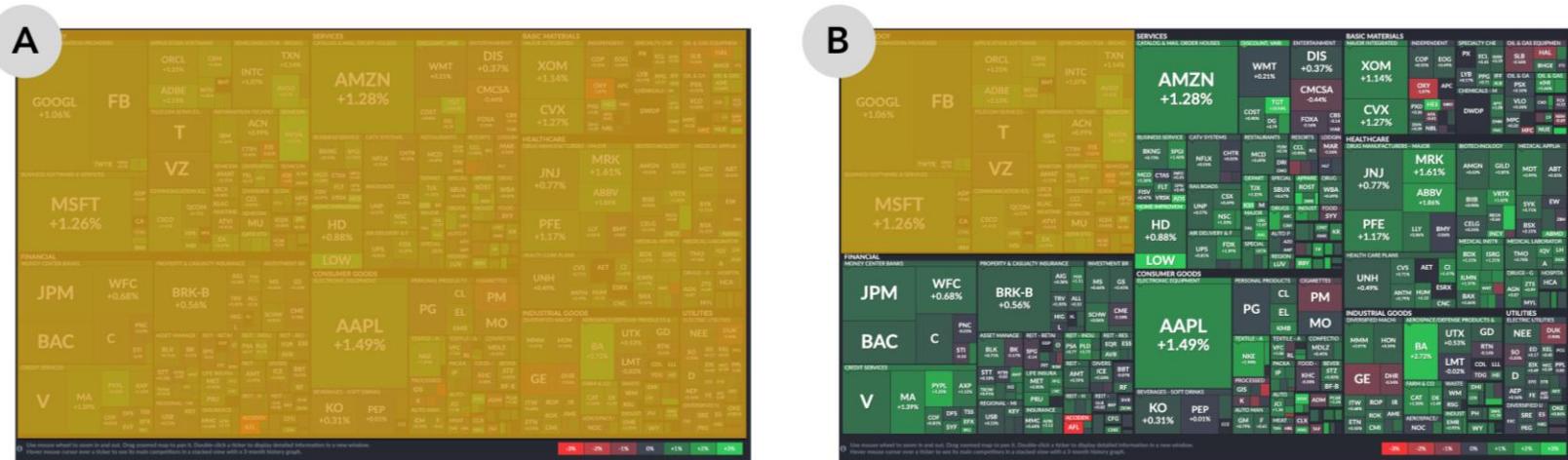
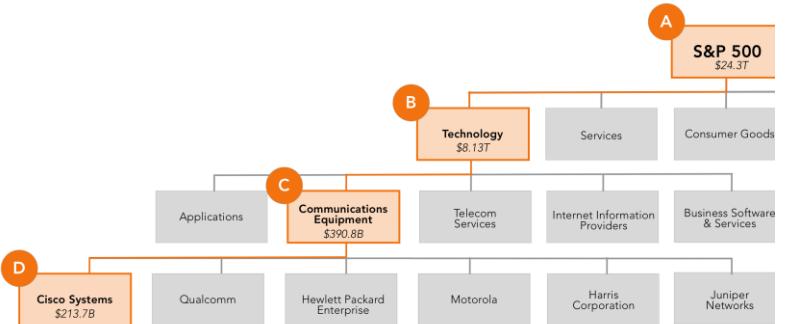
OTHER CHARTS. HIERARCHY. SPACE-FILLING

- Common approaches
 - Rectangular layouts: treemaps, cushion maps...
 - Radial layouts: sunburst, circle packing

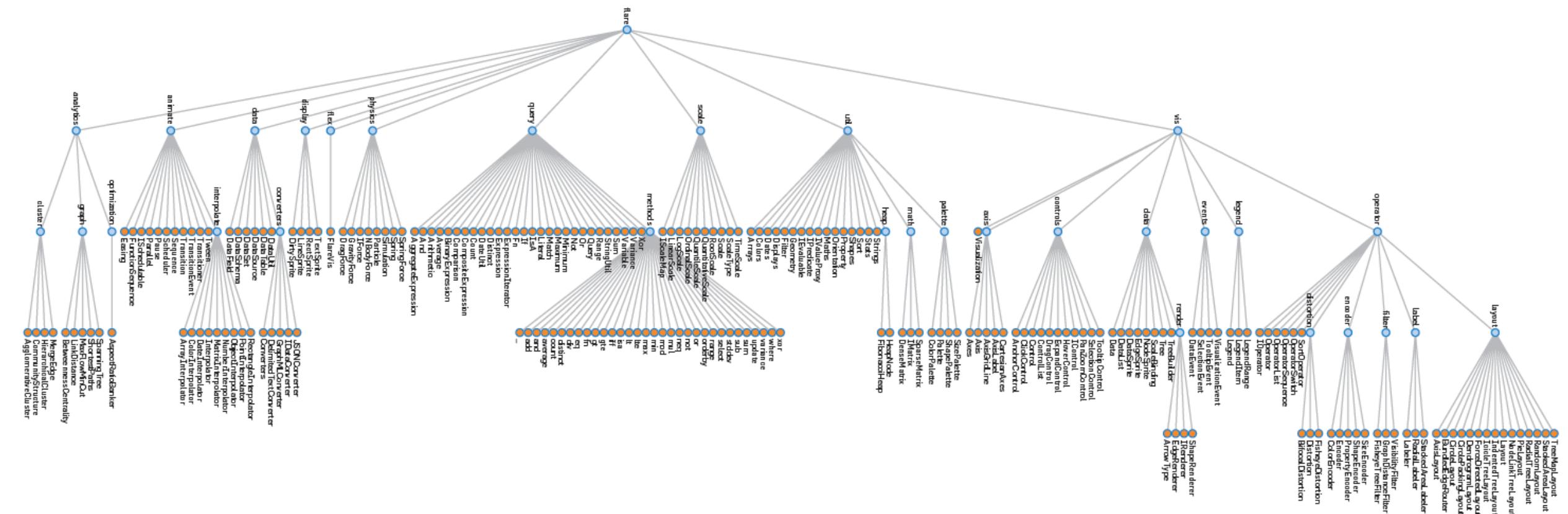


OTHER CHARTS. HIERARCHY. SPACE-FILLING

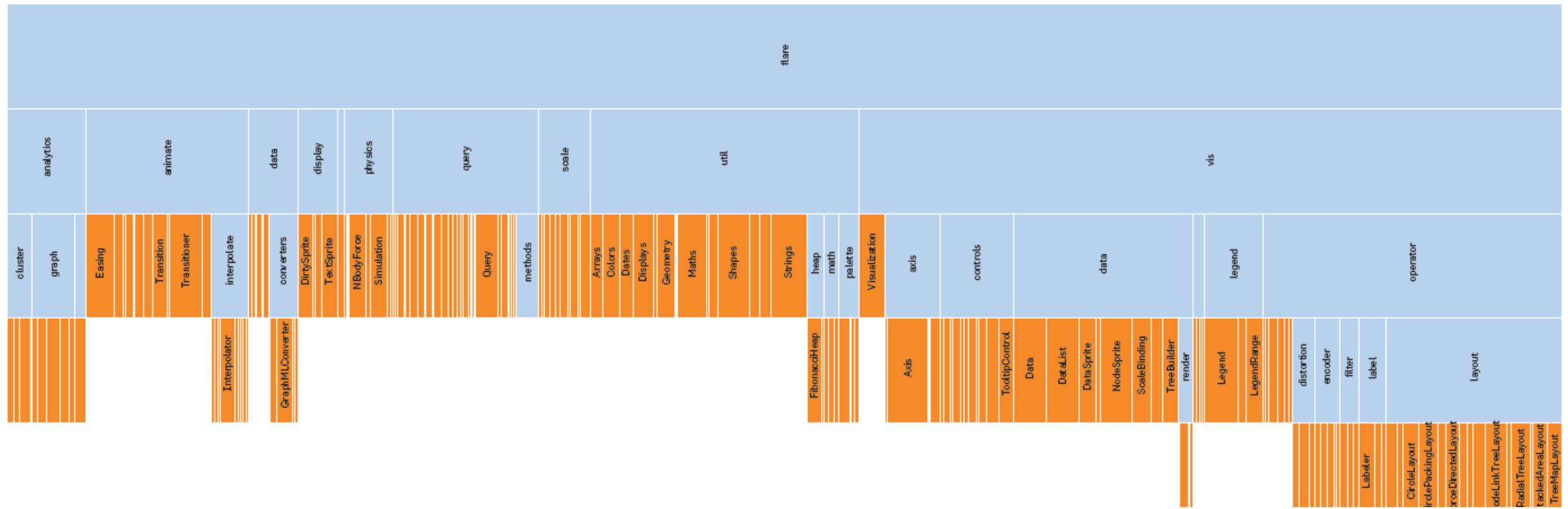
- Treemaps. Construction



OTHER CHARTS. HIERARCHY. NODE-LINK DIAGRAM

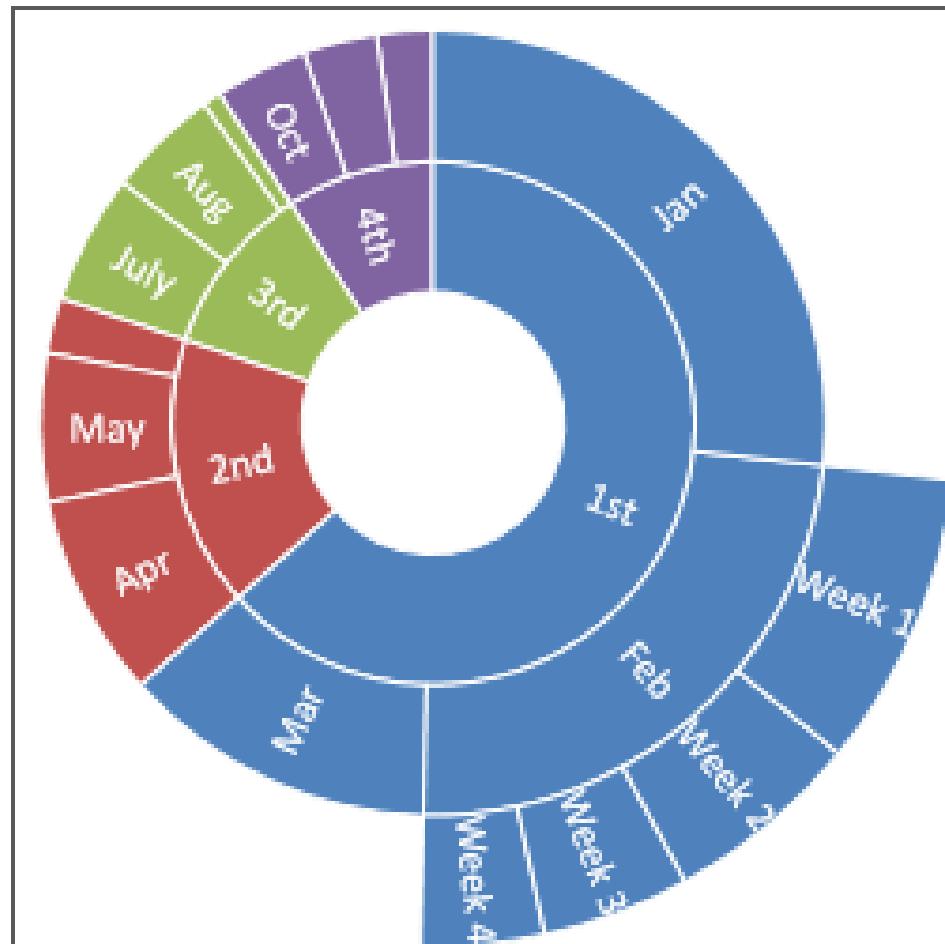


OTHER CHARTS. HIERARCHY. ADJACENCY DIAGRAM



OTHER CHARTS. HIERARCHY. SUNBURST

- Circular version of adjacency diagram



OUTLINE

- *Displaying quantities*
- *Displaying distributions*
- *Displaying proportions*
- *Displaying relationships*
- *Displaying time series*
- *Displaying geospatial data*
- *Other charts*
- **Uncertainty**

UNCERTAINTY

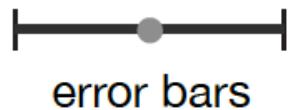
- Data is (very commonly) uncertain to a certain degree
 - We may need to communicate this uncertainty
- It is difficult
 - Regular users not used to uncertainty vis
 - May lead to the readers considering your data is flawed
- But may be highly necessary
 - Depending on the data
 - Depending on the uncertainty levels
 - ...

UNCERTAINTY

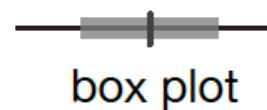
- Most common uncertainty communication methods
- Also violin plots communicate uncertainty to some extent
- Other methods
 - Line ranges

UNCERTAINTY

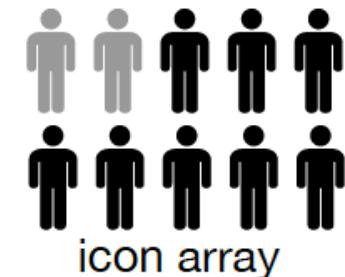
Intervals and Ratios



error bars



box plot



icon array

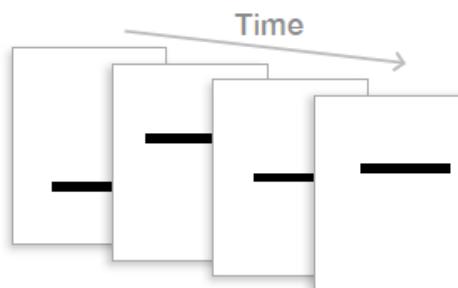
Distributions



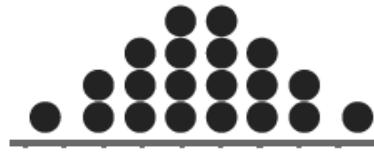
violin plot



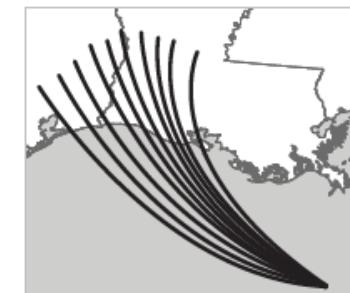
gradient plot



hypothetical outcome plot



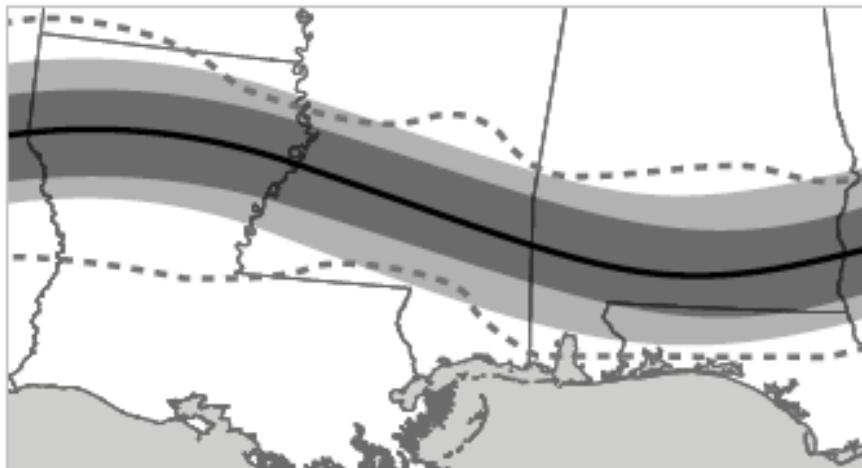
quantile dot plot



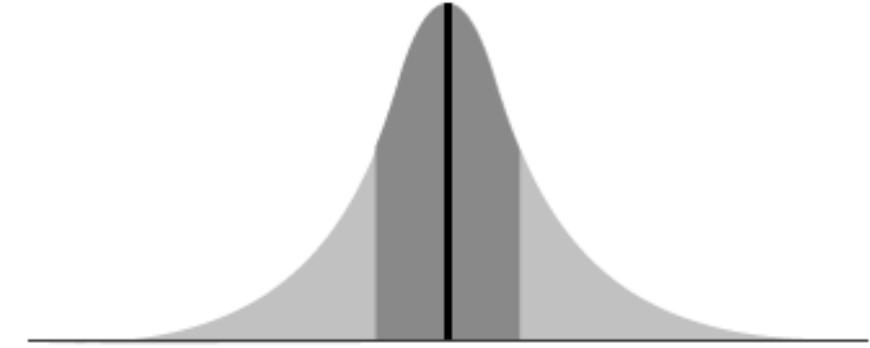
ensemble plot

UNCERTAINTY

- Hybrid methods



contour boxplot



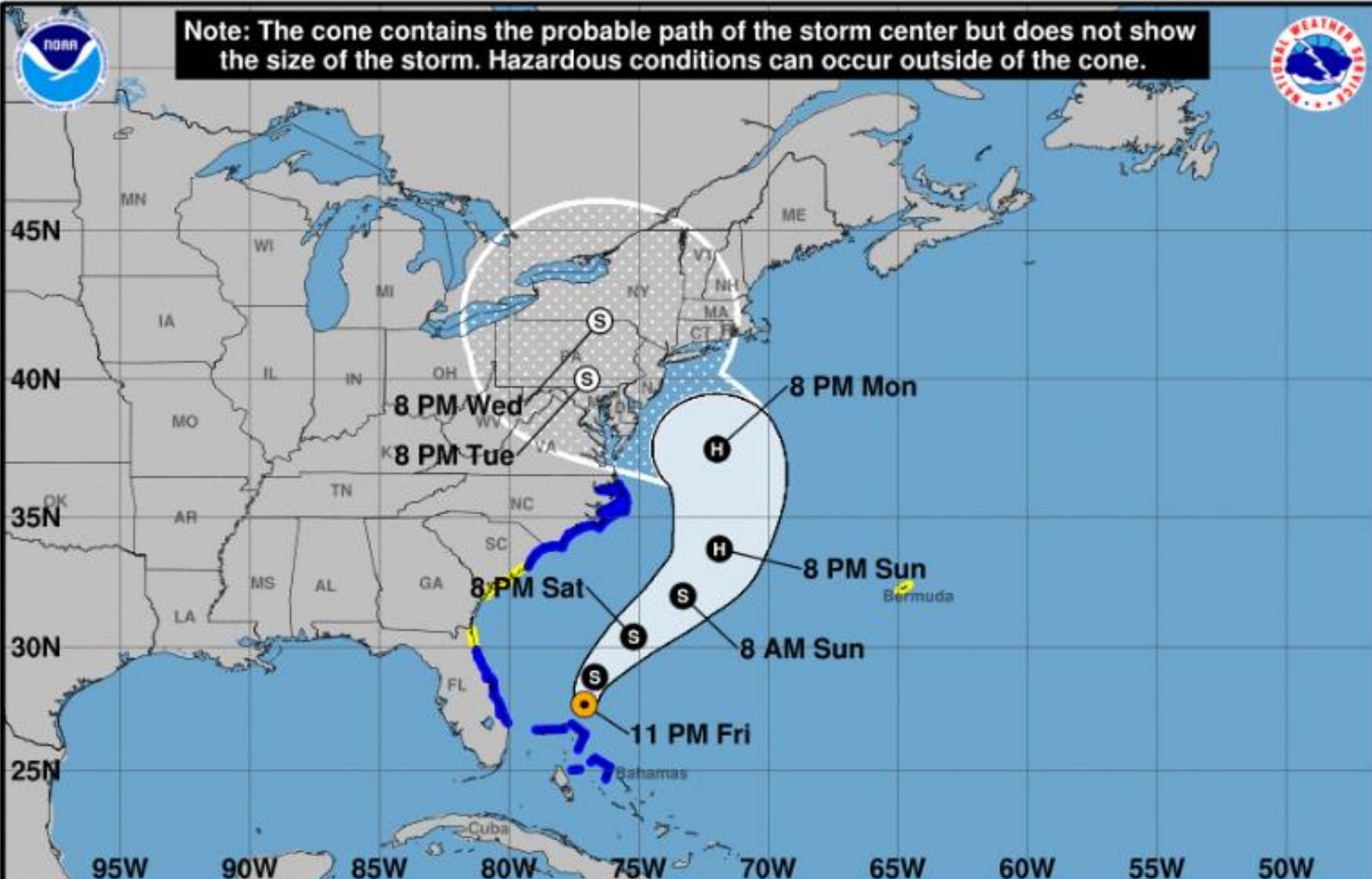
probability density and interval plot

UNCERTAINTY

- Maybe the only uncertainty communication method that has gone to public is hurricane visualization



Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.



Hurricane Sandy

Friday October 26, 2012

11 PM EDT Advisory 19

NWS National Hurricane Center

Current information:

Center location 27.7 N 77.1 W
Maximum sustained wind 75 mph
Movement N at 7 mph

Forecast positions:

● Tropical Cyclone ○ Post-Tropical
Sustained winds: D < 39 mph
S 39-73 mph H 74-110 mph M > 110 mph

Potential track area:



Day 1-3



Day 4-5

Watches:



Hurricane



Trop Storm

Warnings:



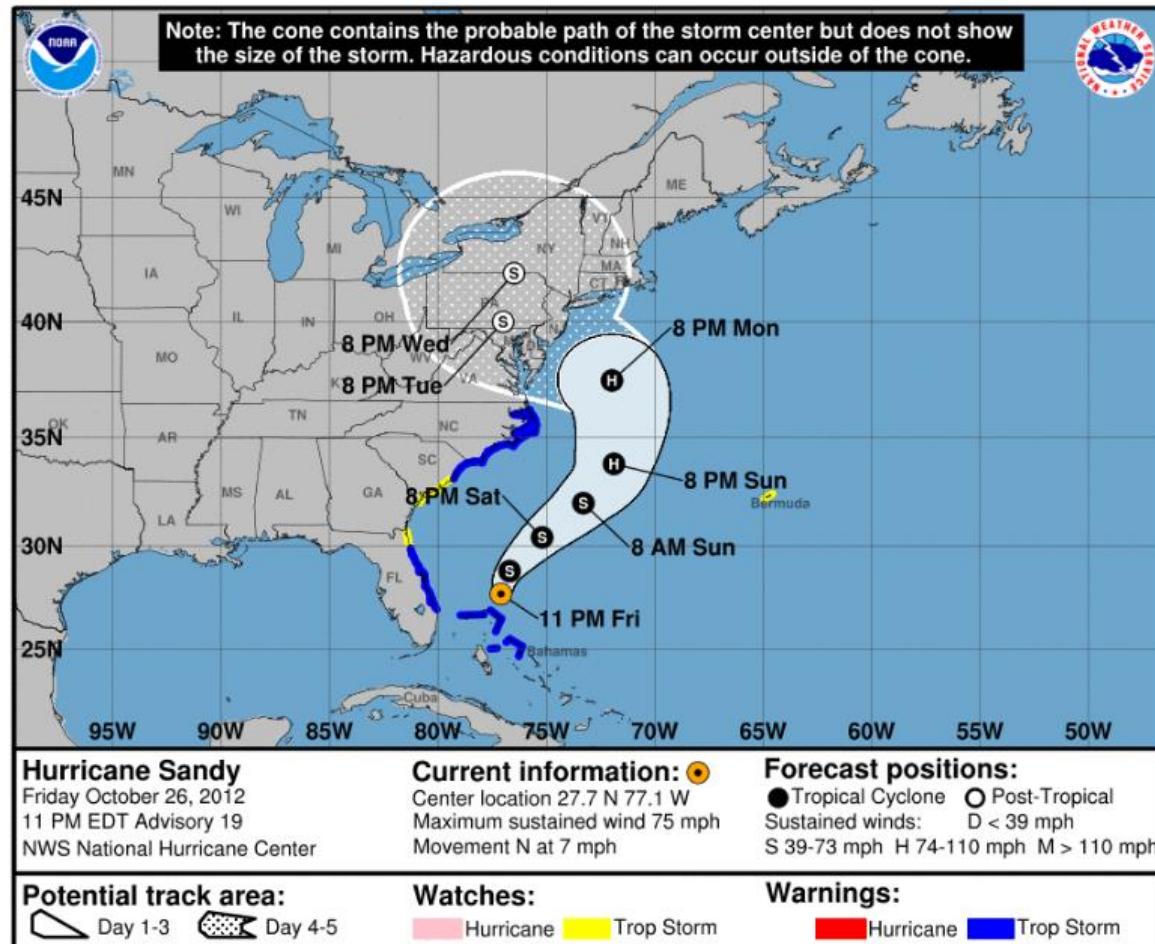
Hurricane

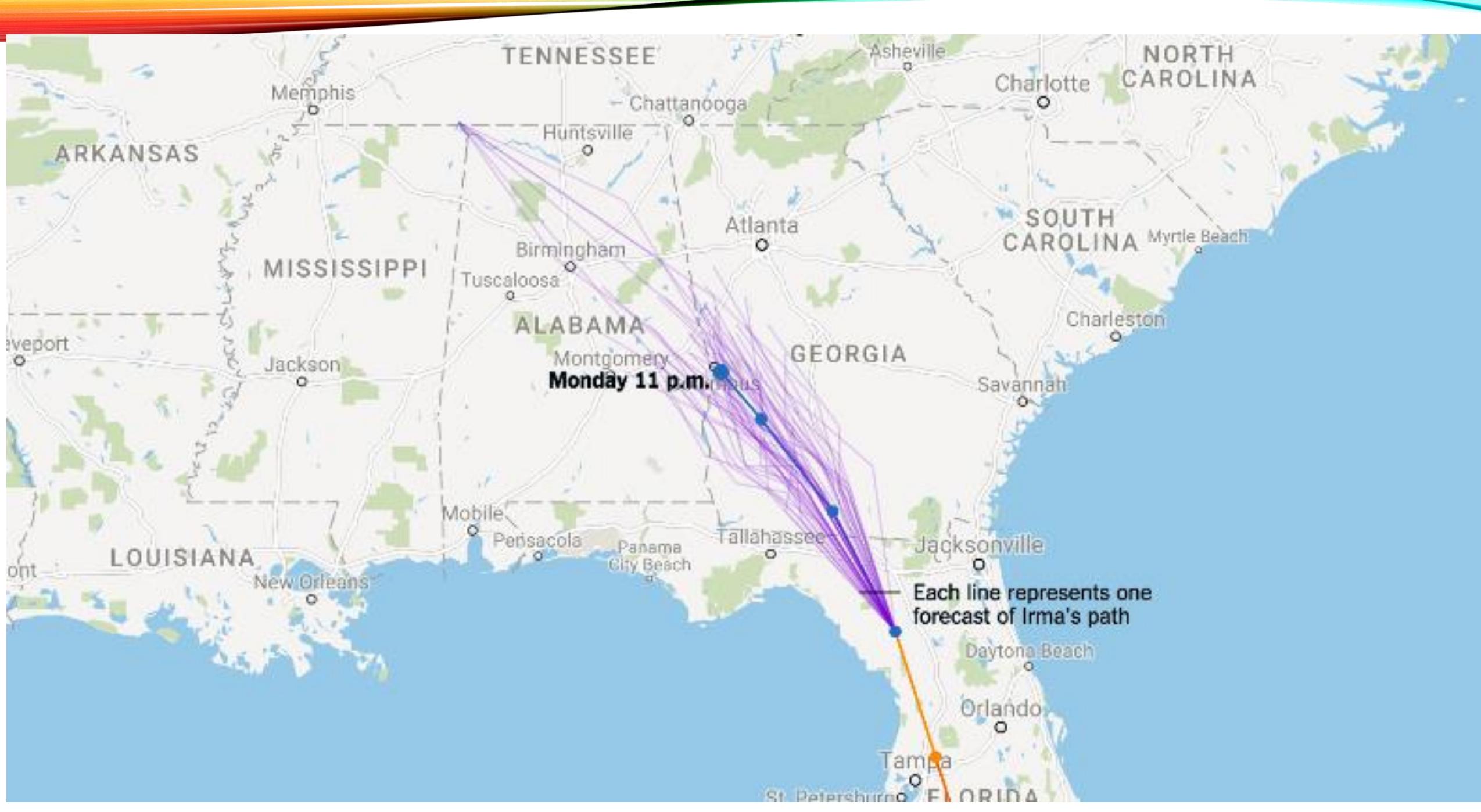


Trop Storm

UNCERTAINTY. HURRICANE TRAJECTORY VIS

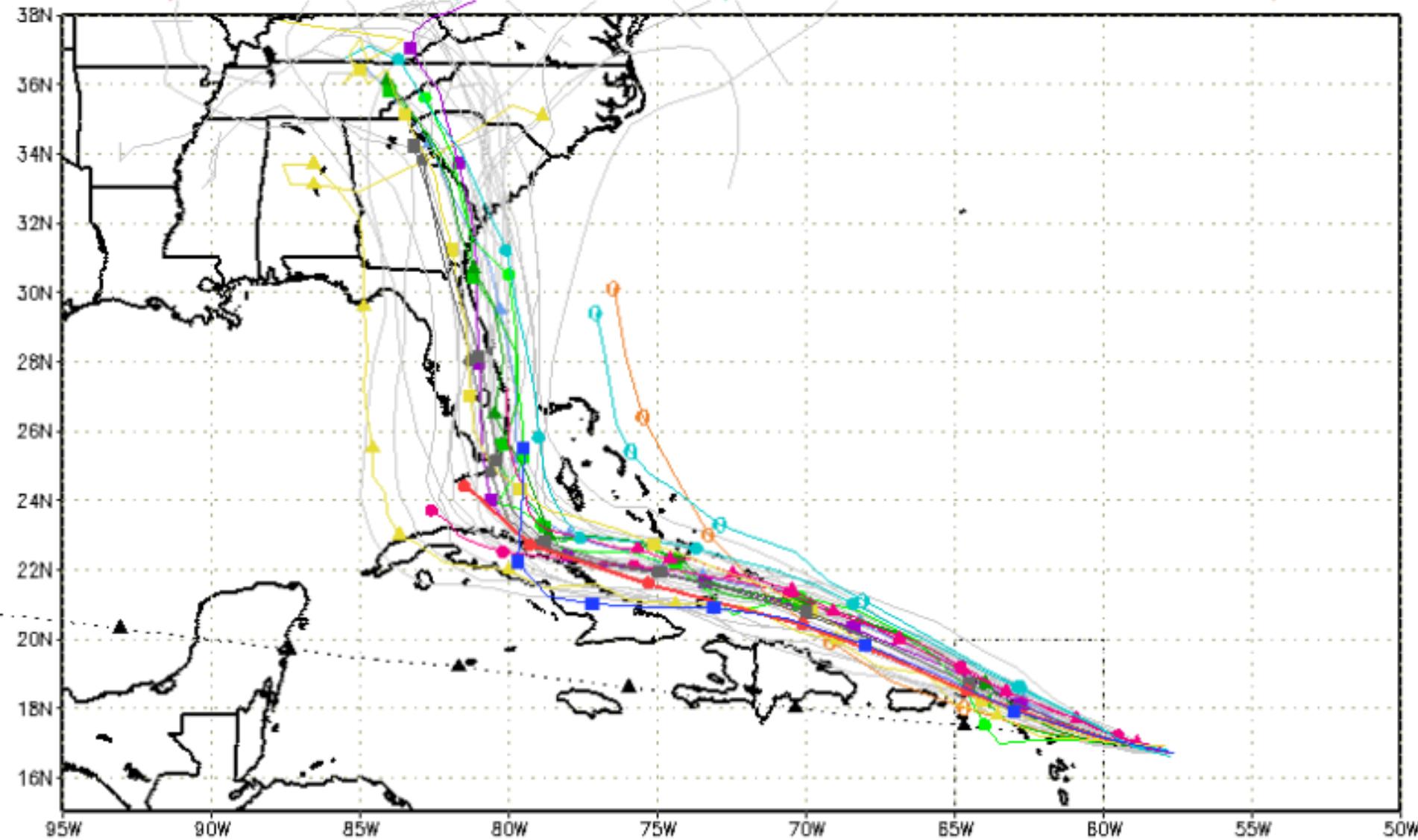
- Cone **contains probable path**
- Uncertainty grows with time
- Forecasting models exhibit different behaviors (paths)







- XTRP 05/1800Z
- TVCN 05/1800Z
- TCVX 05/1800Z
- NHC 05/2100Z
- TABD 05/1800Z
- TABM 05/1800Z
- TABS 05/1800Z
- NVGM 05/1200Z
- HMON 05/1200Z
- GFDN Not Avail
- HWRF 05/1200Z
- OTC 05/1200Z
- AVNO 05/1200Z
- AEMN 05/1200Z
- APxx 05/1200Z
- NAM Not Avail
- CMC 05/1200Z
- CEMN 05/1200Z
- UKM 05/1200Z
- CLP5 05/1800Z



storm_11

sfwmmd.gov

weather@sfwmmd.gov
05-Sep 17:21EDT

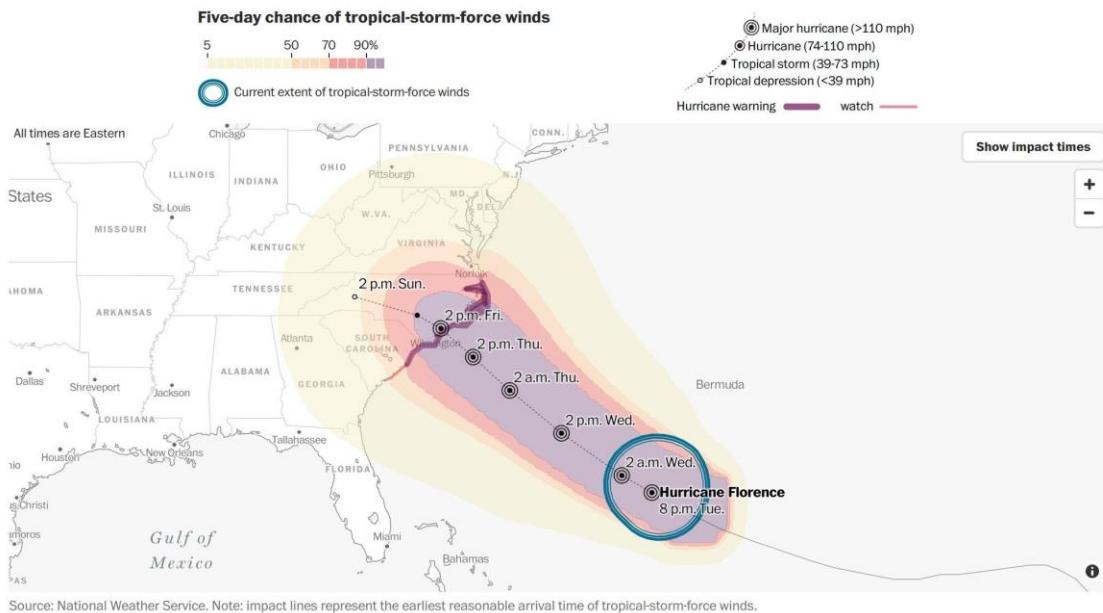
NHC Advisories and County Emergency Management Statements supersede this product.
This graphic should complement, not replace, NHC discussions.
If anything on this graphic causes confusion, ignore the entire product.
For full info, see <http://my.sfwmmd.gov/sfwmmd/common/images/weather/plots.html>



UNCERTAINTY. HURRICANE TRAJECTORY VIS

- Problems with cone-based hurricane forecasting vis:
 - Cone size != hurricane size, hurricane impact outside cone...
 - <https://weather.com/science/weather-explainers/news/tropical-storm-cyclone-forecast-cone-hurricane>
- Problems with spaghetti plots forecasting vis:
 - Some are not models, some are old (e.g., 12 hours old), some are statistical models useless for tracking...
 - <https://arstechnica.com/science/2017/09/please-please-stop-sharing-spaghetti-plots-of-hurricane-models/>

WHAT IF...



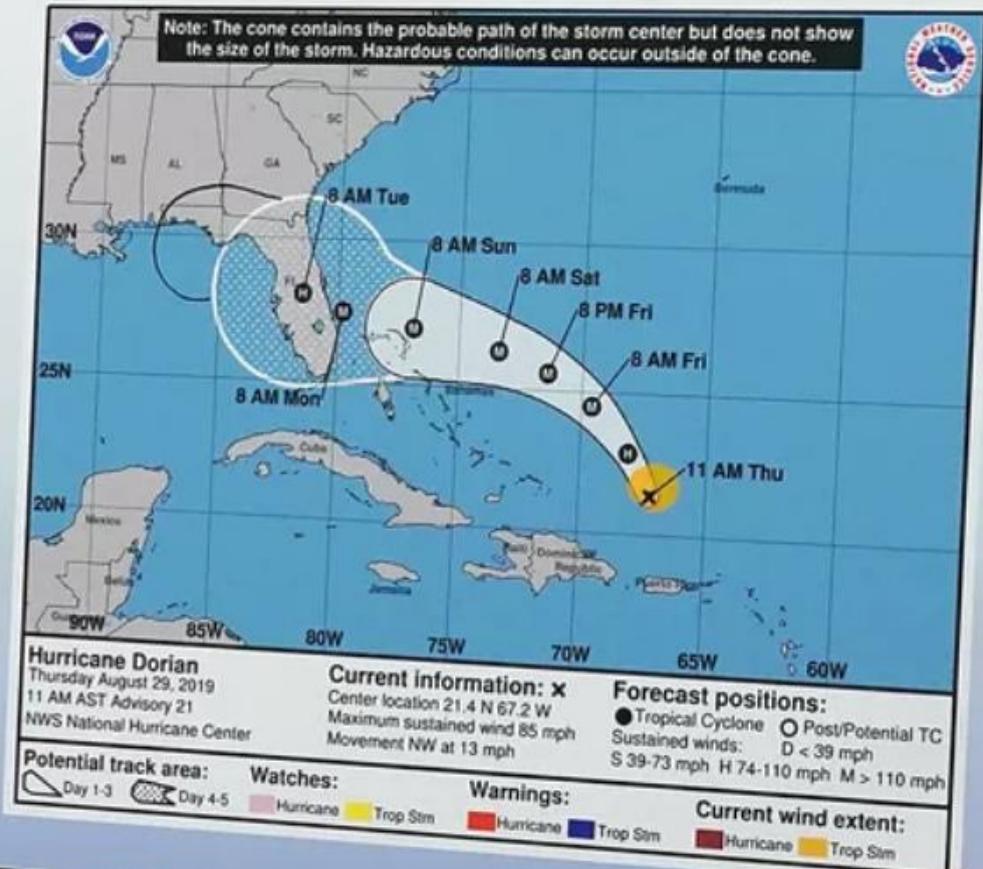


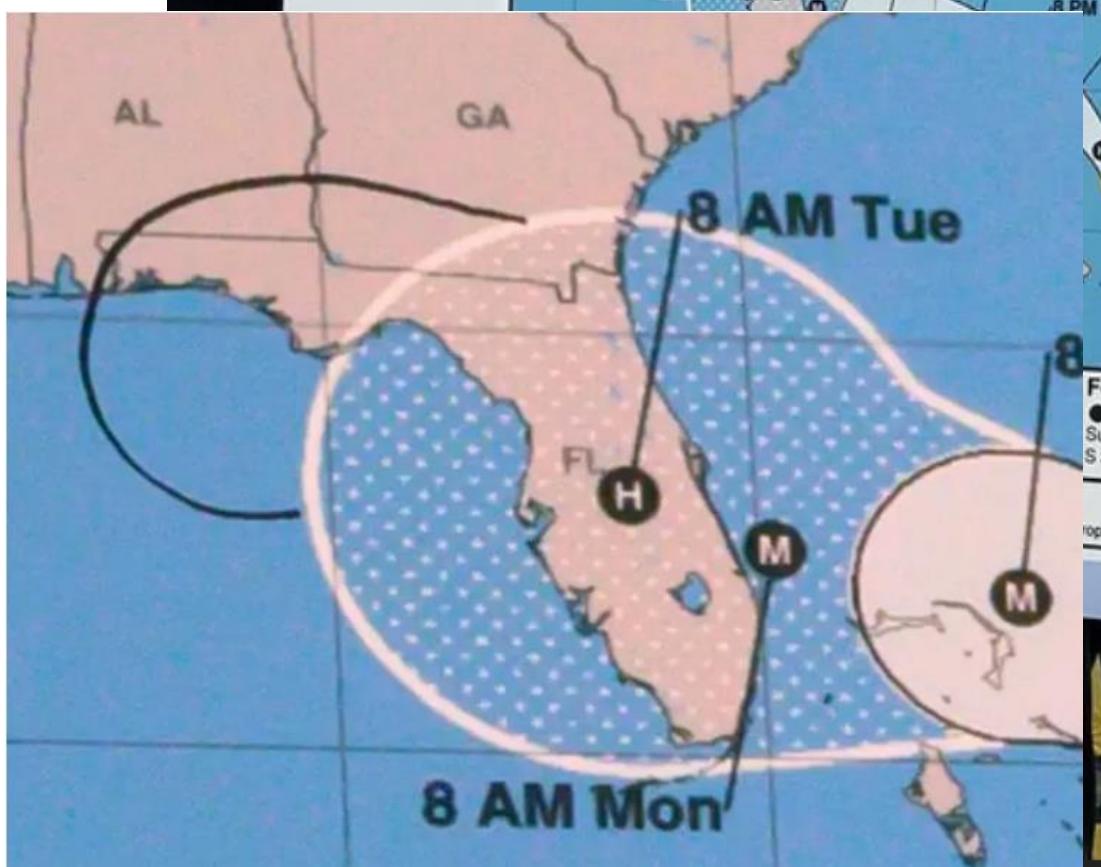
HURRICANES + TRUMP

- Trump (falsely) [tweeted on Sept. 1](#), 2019, that several southern states, **including Alabama**, were “most likely to be hit” by the hurricane after its deadly pass through the Bahamas.
- **Three days later**, Trump [shared a fake map](#) in which a storm track, showed Hurricane Dorian moving toward Alabama.
- When a National Weather Service forecaster tried to set the record straight, its parent agency, NOAA, released an unsigned statement [disavowing the correction](#) — seemingly to appease the White House.



Hurricane Dorian Forecast Track and Intensity







Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.



OTHER RESOURCES

- Jon Schwabish's "One chart at a time" Youtube series:
<https://www.youtube.com/watch?v=gFFj22kjZk>
- <https://datavizproject.com/data-type/>
- <https://public.tableau.com/es-es/s/gallery/visual-vocabulary>
- <https://github.com/ft-interactive/chartdoctor/tree/master/visual-vocabulary>
- <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>
- Timeline of datavis milestones <http://www.datavis.ca/milestones/>

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Heath; Li Faixue; Graham Farish; Billy Ehrenberg-Shannon; Paul McCallum; Martin Stalder
Inspired by the Graphic Continuum by Jim Schatzki and Saverio Ricca



ft.com/vocabulary

Deviation

Emphasise variations (+/-) from a fixed reference point. Can also be a target point if zero but can also be a target or a long-term average. Can also be used to show sentiment. (point-in-time scatterplots)

Example FT uses
Trade surplus/deficit, climate change



Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Ordered bar

Standard bar charts display the ranks of values much more easily than sorted into order.

Histogram

The standard way to show a unimodal distribution - keep the gaps between columns small to highlight the shape of the data.

Dot plot

A good way of showing a discrete range or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many points have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Boxplot

Summarise multiple distributions by showing the median (center) and range of the data.

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple averages).

Lollipop

Lollipop draw more attention to the data value than standard box plots and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates, especially large dates, consider grouping lines using colour.

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution, effectively back-to-back histograms.

Cumulative curve

A good way of showing how unequal a distribution is, x-axis is always cumulative frequency, x-axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a line chart but limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasise individual points in a data series, points can be sized to an additional variable, best with medium-sized dots.

Calendar heatmap

Great when date and duration are key elements of the story in the data.

Priestley timeline

Good for showing the size of varying size across multiple categories (eg earthquakes by comment).

Circle timeline

Good for showing the size of varying size on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram

Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph

A type of area chart; use when seeing changes in proportions over time is more important than individual values.

Correlation

Show the relationship between two or more variables. Note that just because you find a correlation doesn't mean that it's causal (one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy



Ordered column

See above.

Dot plot

Good way of showing a discrete range or range (min/max) of data across multiple categories.

Dot strip plot

Good for showing individual values in a distribution, can be a problem when too many points have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Slope

Perfect for showing how ranks have changed over time or vary between categories.

Boxplot

Summarise multiple distributions by showing the median (center) and range of the data.

Violin plot

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple averages).

Lollipop

Lollipop draw more attention to the data value than standard box plots and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates, especially large dates, consider grouping lines using colour.

Population pyramid

A standard way for showing the age and sex breakdown of a population distribution, effectively back-to-back histograms.

Cumulative curve

A good way of showing how unequal a distribution is, x-axis is always cumulative frequency, x-axis is always a measure.

Frequency polygons

For displaying multiple distributions of data. Like a line chart but limited to a maximum of 3 or 4 datasets.

Beeswarm

Use to emphasise individual points in a data series, points can be sized to an additional variable, best with medium-sized dots.

Calendar heatmap

Great when date and duration are key elements of the story in the data.

Priestley timeline

Good for showing the size of varying size across multiple categories (eg earthquakes by comment).

Circle timeline

Good for showing the size of varying size on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram

Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph

A type of area chart; use when seeing changes in proportions over time is more important than individual values.

Ranking

Show where an item's position in an ordered series is more important than its absolute original value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results



Ordered bar

Standard bar charts display the ranks of values much more easily than sorted into order.

Histogram

The standard way to show a unimodal distribution - keep the gaps between columns small to highlight the shape of the data.

Dot plot

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Dot strip plot

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Barcode plot

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Slope

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Boxplot

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Violin plot

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Lollipop

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Bump

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Population pyramid

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Cumulative curve

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Frequency polygons

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Beeswarm

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Calendar heatmap

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Priestley timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Circle timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Seismogram

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Streamgraph

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Parallel coordinates

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Bullet

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Seismogram

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Streamgraph

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Vertical timeline

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Waterfall

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

Grouped symbol

The standard way to show a discrete distribution - keep the gaps between columns small to highlight the shape of the data.

DATA VISUALIZATION. VISUALIZATION TECHNIQUES

Pere-Pau Vázquez

Dept. Computer Science – UPC