

# **DATA 230 Spring 2023**

## **Week 5**

# **Lecture 5:**

# **One Variable: Count and Category I**

Dr. Seungjoon (Joon) Lee

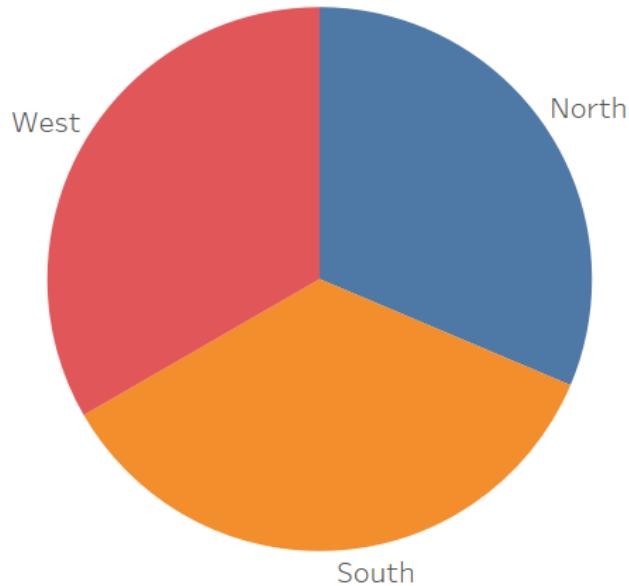
DATA 230: Data Visualization

# Counting Data

We have one attribute (feature/dimension) as categorical data (or finite numeric quantity).

- Yes or No question (binary).
- Highest education level (a few categories).
- Zip code, car manufacturer (finite but many).
- Continuous quantity → discrete quantity due to a certain purpose.

# Binary or a Few Categories



# Bar Chart

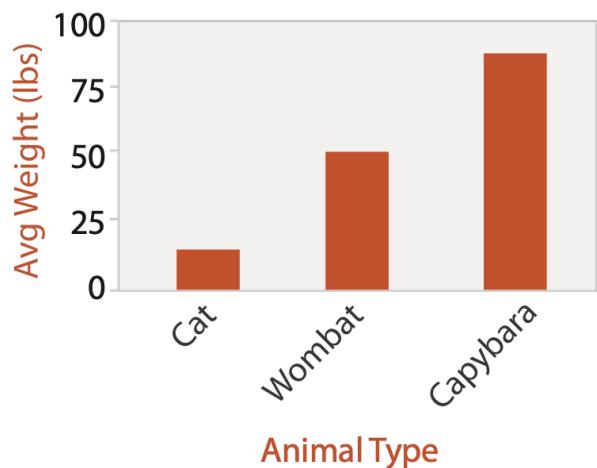
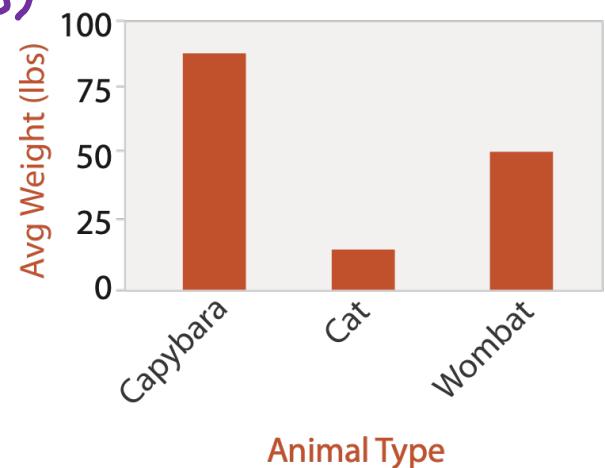
Data: 1 Categorical data, 1 quantitative feature. (Count)

Mark: Line(s)

Channel: size (length) for quant.  
position (Generally horizontal) for categorical. (Vertically aligned)

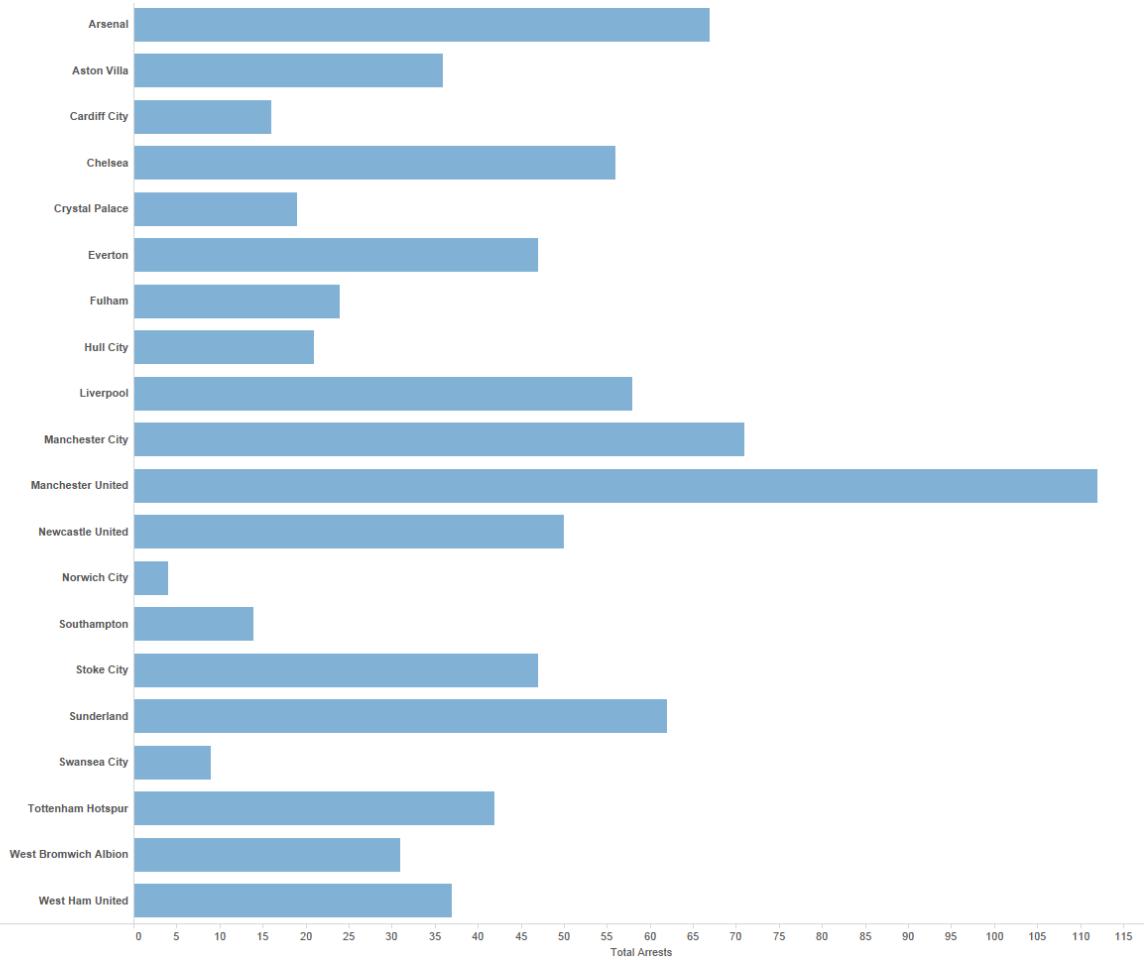
Task: Compare quant. (count) w.r.t. Category

Scalability: dozen to hundreds (depends on display)

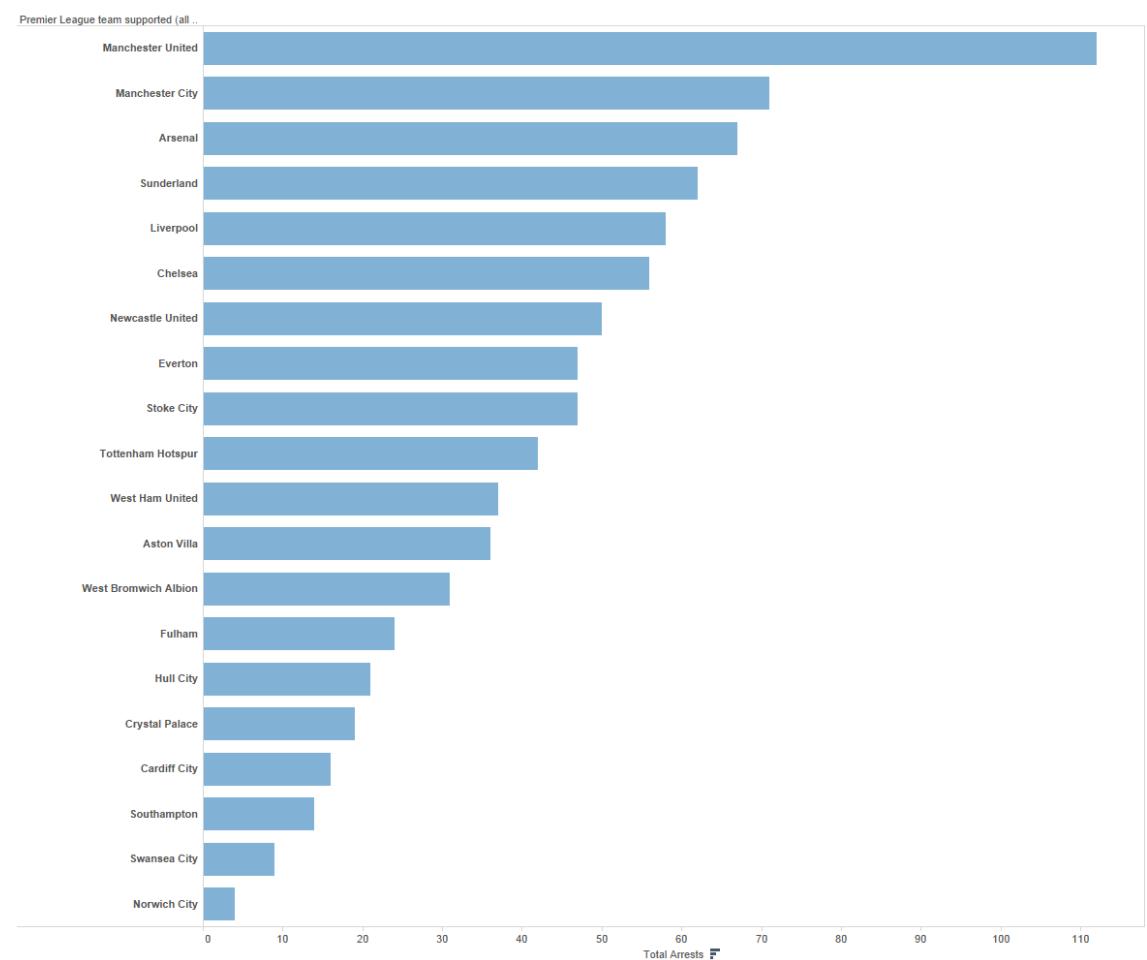


# Bar Chart

ordered by Alphabet (easy to find)



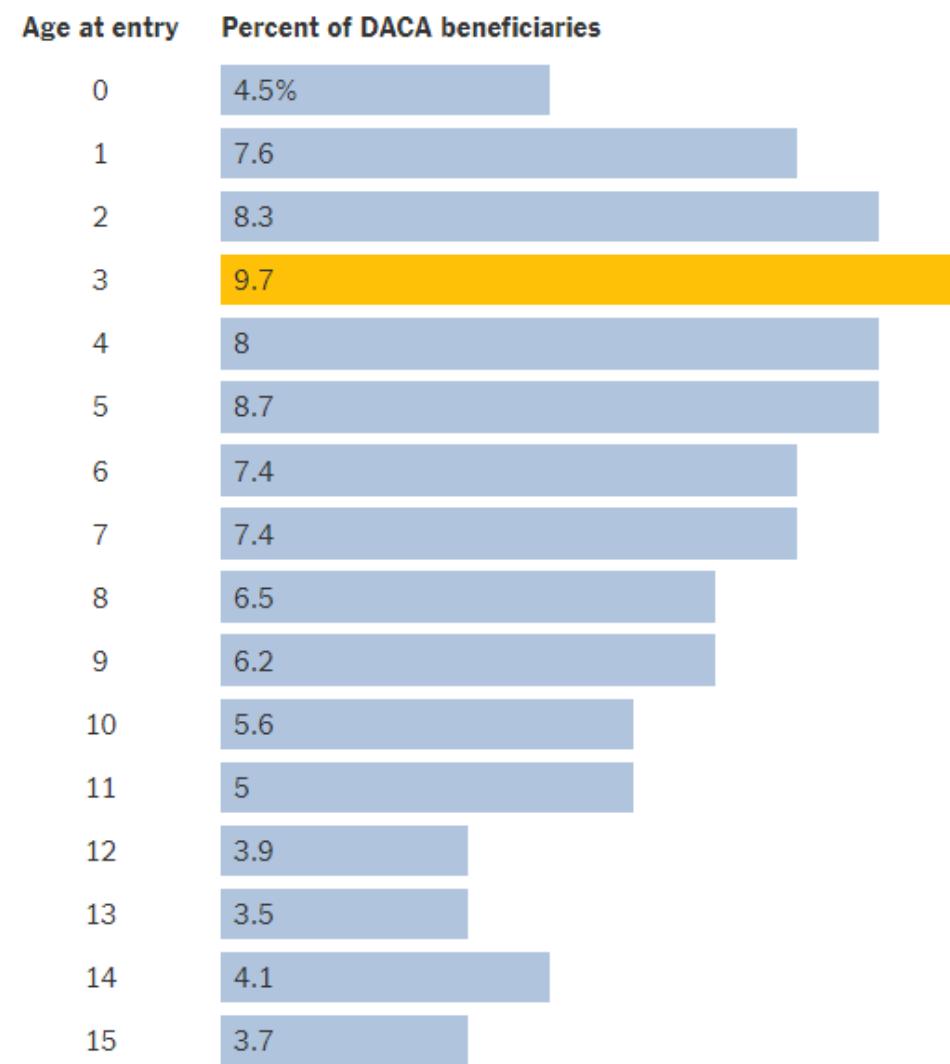
ordered by length (data-driven)



# Bar Chart

Which state does have the most beneficiaries

| State          | DACA beneficiaries |
|----------------|--------------------|
| California     | 222,795            |
| Texas          | 124,300            |
| Illinois       | 42,376             |
| New York       | 41,970             |
| Florida        | 32,795             |
| Arizona        | 27,865             |
| North Carolina | 27,385             |
| Georgia        | 24,135             |
| New Jersey     | 22,024             |
| Washington     | 17,843             |
| Colorado       | 17,258             |
| Nevada         | 13,070             |
| Virginia       | 12,134             |
| Oregon         | 11,281             |



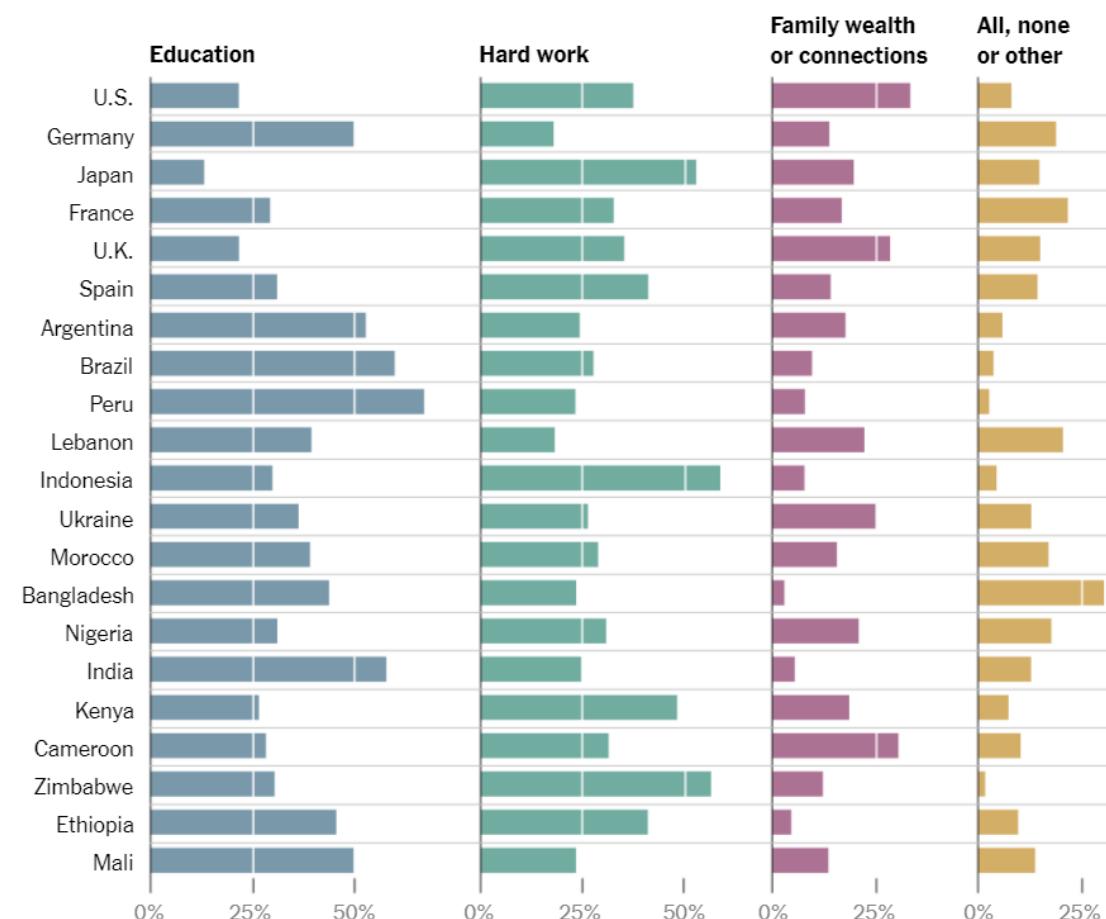
What is an Age distribution?

redundant.

# Bar Chart

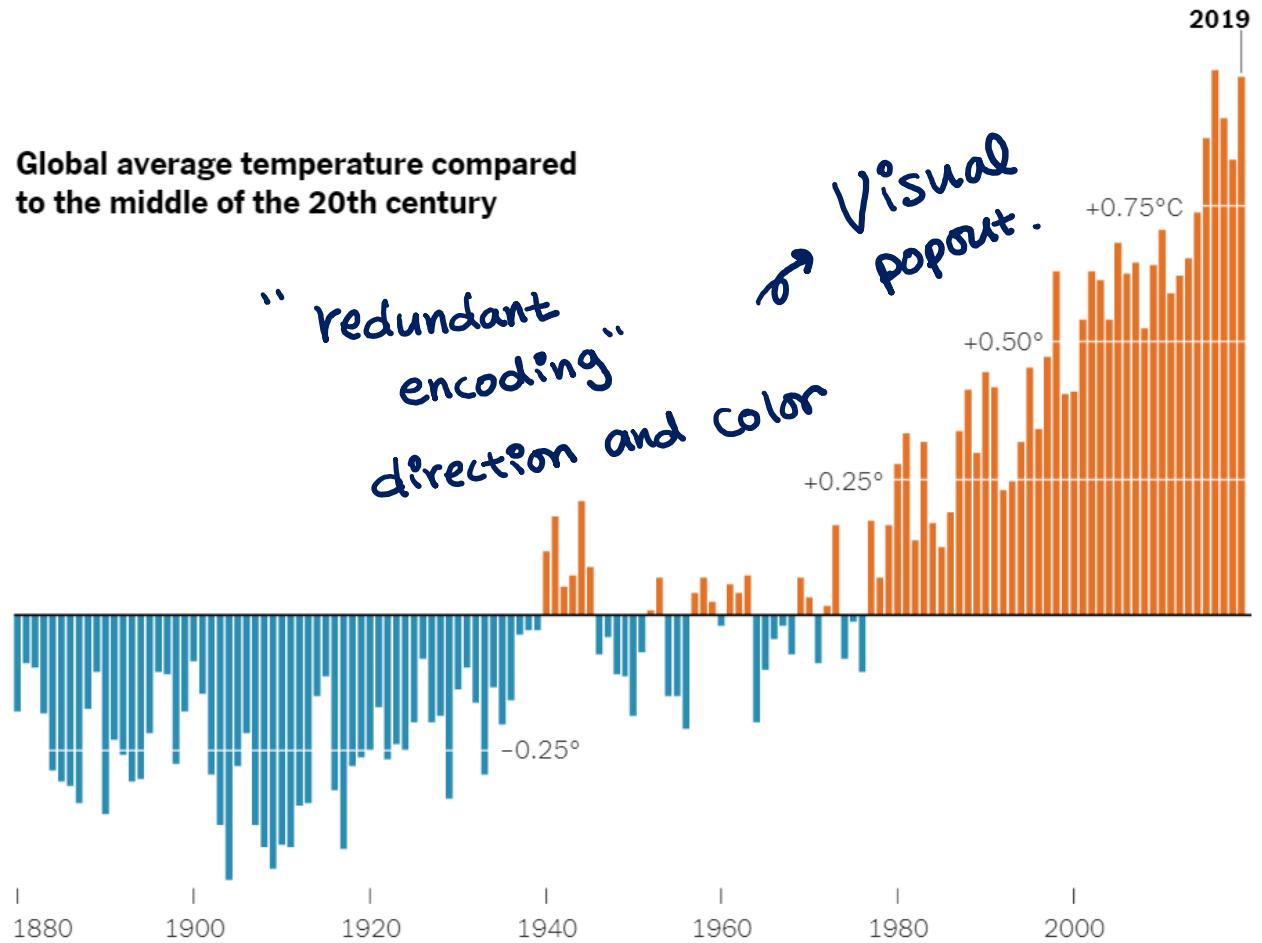
## What Young People Think is the Most Important Factor in Determining Success

How 15- to 24-year-olds in each country responded.



Note: Countries are sorted from richest to poorest according to a measure of gross national income per capita used by the World Bank. • Source: UNICEF-Gallup survey of more than 21,000 people conducted by telephone in 21 countries between February and June 2021.

Global average temperature compared to the middle of the 20th century



# Stacked Bar Chart

Data: 2 Categorical features & 1 quantitative feature (count)

Mark: Vertical stack of lines.

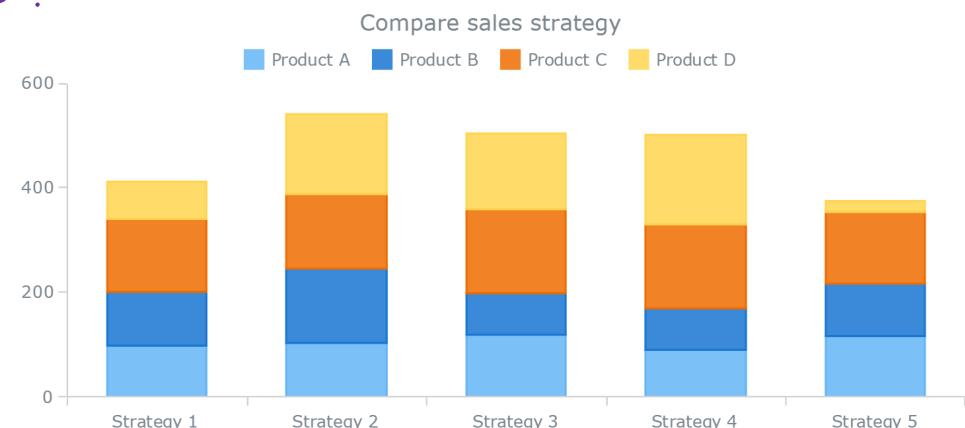
Channel: size(length) for quant.

position(horizontal) for 1 categorical (aligned vertically)

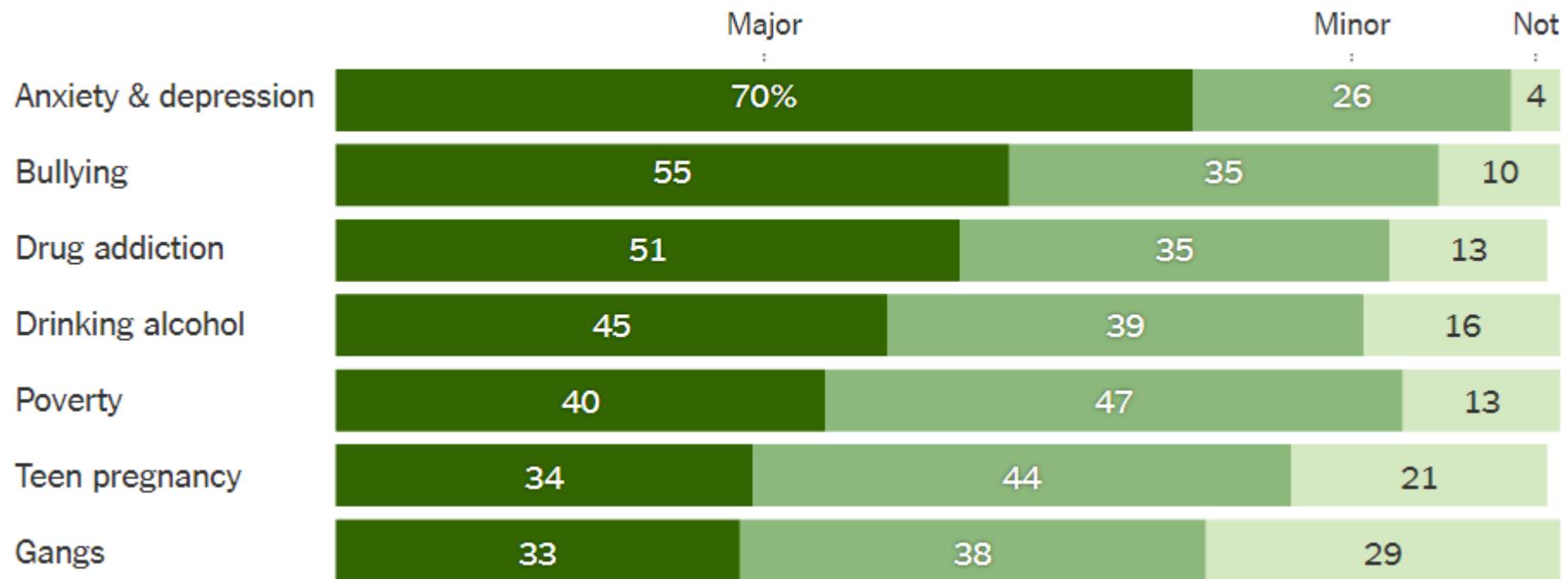
position (vertical) for 1 categorical (unaligned vertically) and color

Task: Summarize , part-to-whole relationship for each category

Scalability: dozen to hundreds (display) , 10 ~ 12 stacks .



# Stacked Bar Chart



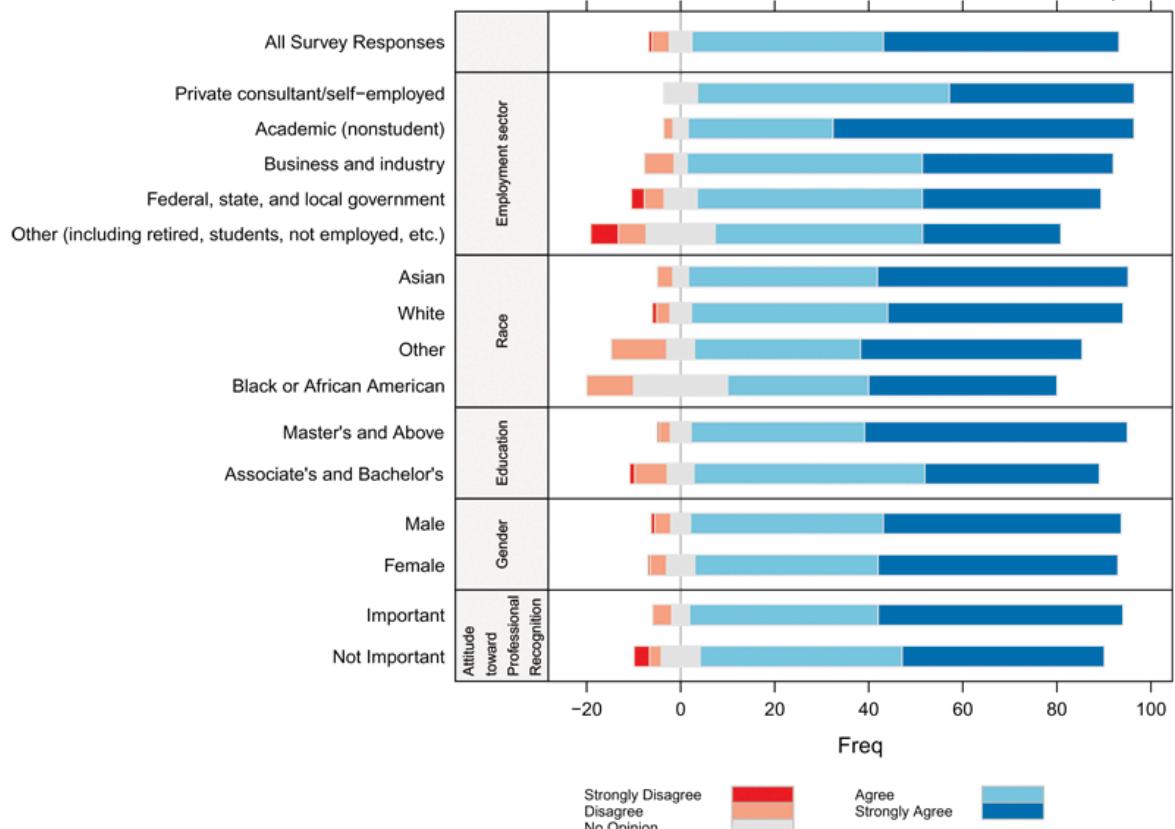
## The Major Issue of Mental Health

This graph shows the percentage of American teenagers, ages 13 – 17, who feel certain problems are major or minor among people their age in their community. The data is from a representative sample of 920 teens surveyed from September to November 2018.

# Stacked Bar Chart

diverging

Is your job professionally challenging?

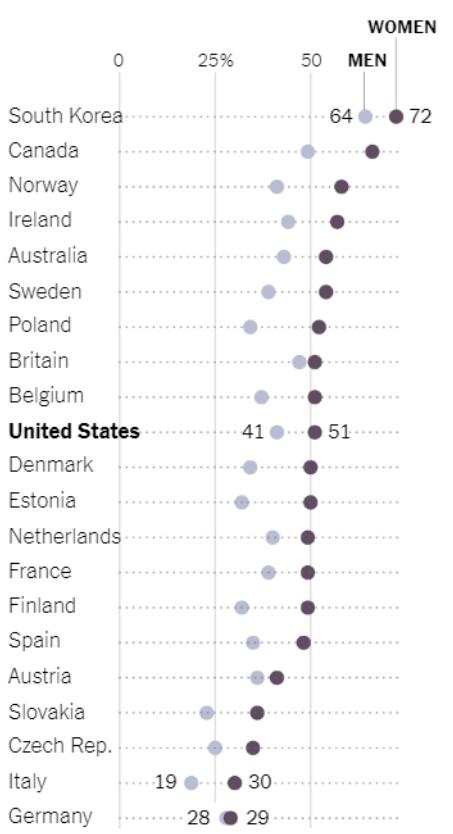


Education attainment in selected countries for 25-34 year old.

Percentage of people who got more, the same or less education than their parents, as of 2012.

|               | MORE<br>(upwardly<br>mobile) | SAME<br>(status<br>quo) | LESS<br>(downwardly<br>mobile) |
|---------------|------------------------------|-------------------------|--------------------------------|
| South Korea   | 61%                          | 35                      | 3                              |
| Italy         | 45                           | 49                      | 5                              |
| Ireland       | 45                           | 44                      | 12                             |
| Russia        | 44                           | 46                      | 11                             |
| Spain         | 41                           | 49                      | 10                             |
| France        | 40                           | 50                      | 10                             |
| Finland       | 39                           | 46                      | 15                             |
| Netherlands   | 38                           | 45                      | 17                             |
| Poland        | 36                           | 57                      | 7                              |
| Australia     | 36                           | 48                      | 16                             |
| Belgium*      | 35                           | 52                      | 13                             |
| Britain*      | 33                           | 51                      | 16                             |
| Denmark       | 28                           | 54                      | 18                             |
| Canada        | 27                           | 54                      | 18                             |
| Japan         | 25                           | 58                      | 18                             |
| Sweden        | 24                           | 47                      | 28                             |
| United States | 24                           | 54                      | 23                             |
| Estonia       | 23                           | 50                      | 27                             |
| Slovakia      | 23                           | 67                      | 10                             |
| Norway        | 22                           | 51                      | 27                             |
| Austria       | 21                           | 57                      | 21                             |
| Germany       | 19                           | 57                      | 24                             |
| Czech Rep.    | 17                           | 71                      | 12                             |

Percentage of people who attained education beyond high school, as of 2014.



Gender data for Russia and Japan not available.

# Normalized Stacked Bar Chart

Data: Same but 1 quant. by normalization.

Mark: SAME

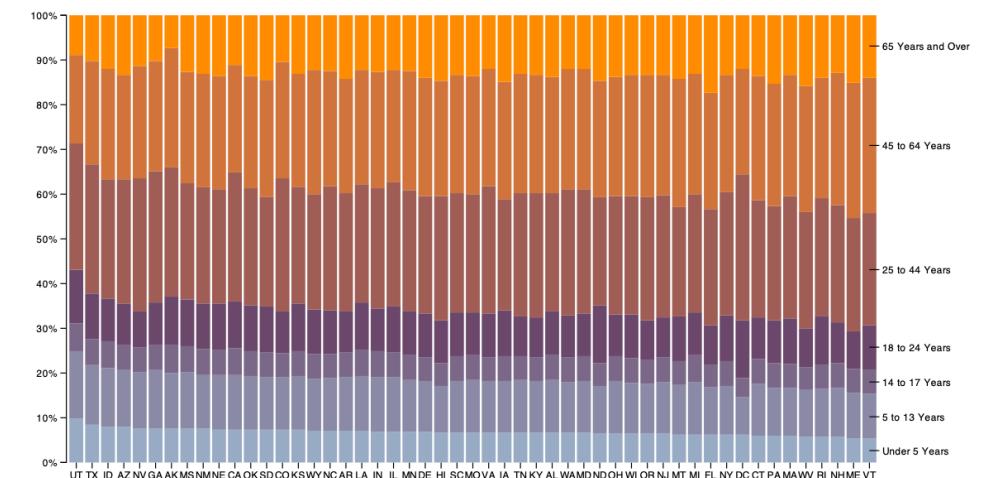
derived data

Channel:

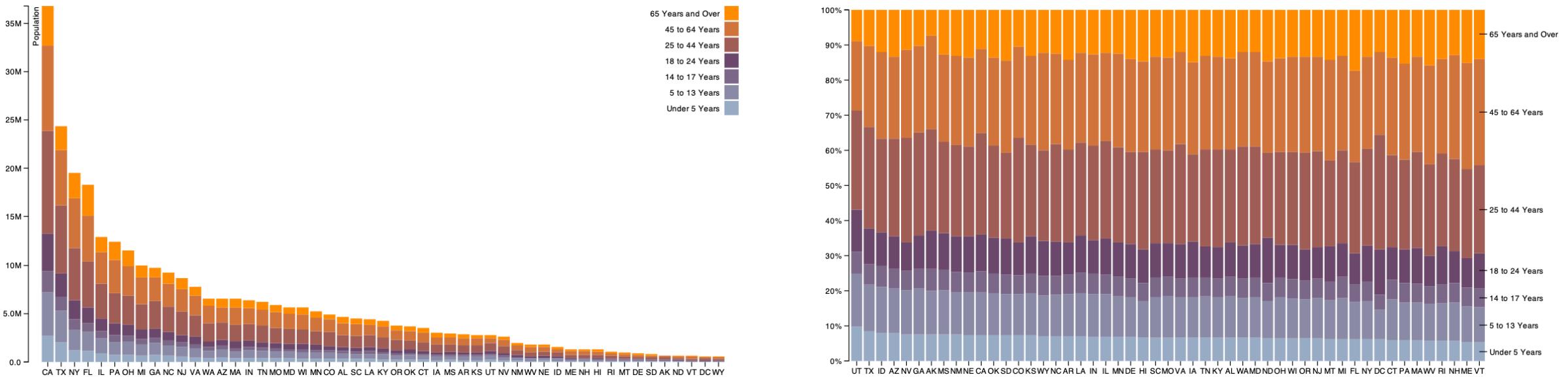
SAME

Task: Part-to-whole judgement. (Summarize)

Scalability: SAME .



# Normalized Stacked Bar Chart



# Marimekko Chart

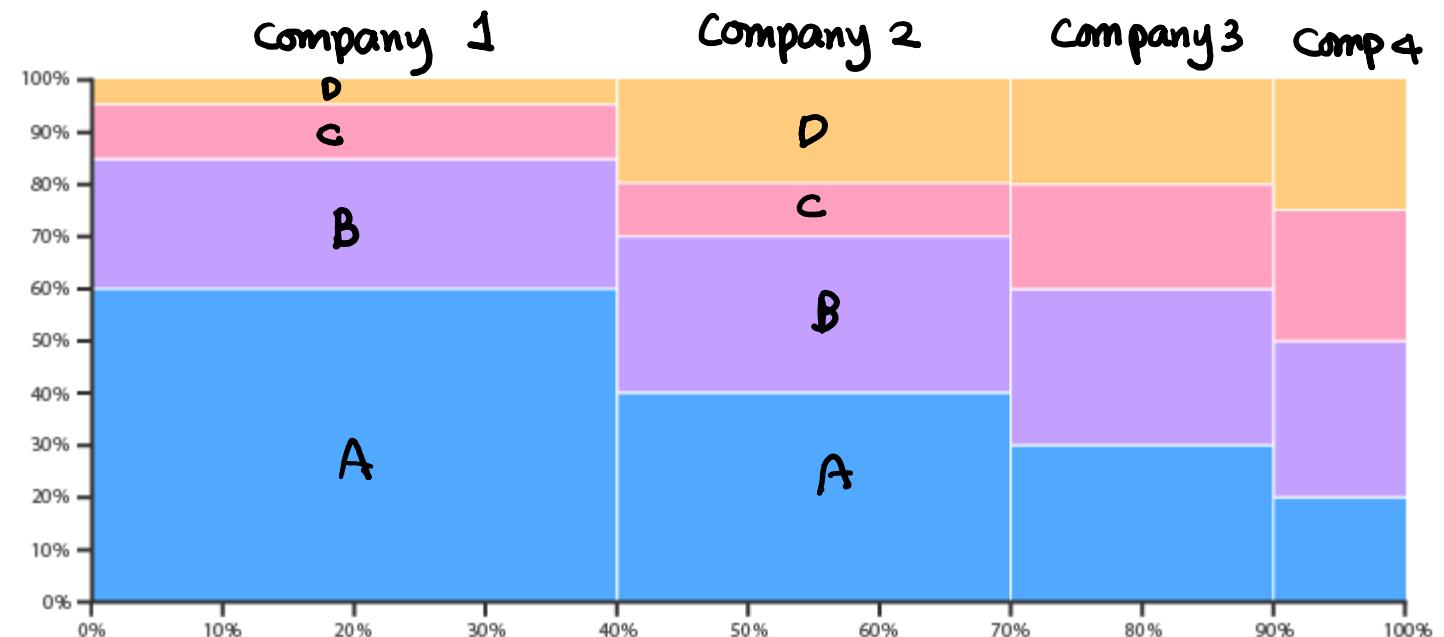
Data: 1 Categorical feature & 2 quantitative features (derived data by normalization)

Mark: Area(s) ↗ 2 lengths

Channel: Size(area) for 2 quantitative (derived) data

Task: Part-to-Whole for 2 different features

Scalability: Less than normalized  
Stacked bar



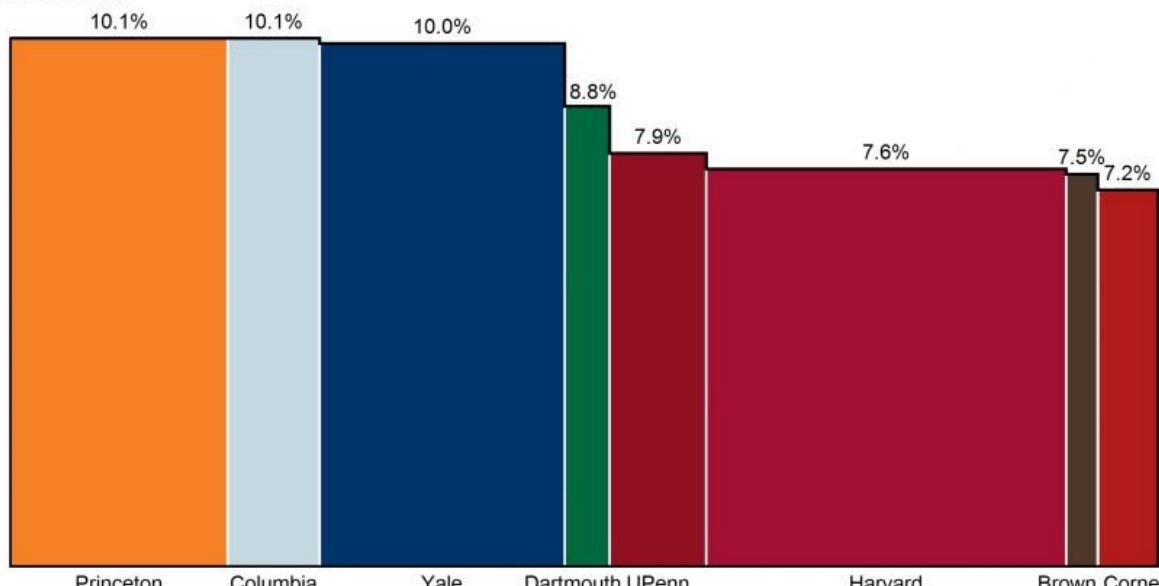
# Marimekko Chart



## Ivy League Endowment Investment Returns

*While Harvard has by far the largest endowment of any ivy, its returns over the last 10 years have been underwhelming.*

10 yr annualized returns



|                               |    |    |    |   |    |    |   |   |
|-------------------------------|----|----|----|---|----|----|---|---|
| Assets Under Management (\$B) | 23 | 10 | 26 | 5 | 10 | 38 | 3 | 6 |
|-------------------------------|----|----|----|---|----|----|---|---|

|                                |    |    |    |   |   |   |   |   |
|--------------------------------|----|----|----|---|---|---|---|---|
| Endowment Chiefs' Tenure (yrs) | 21 | 14 | 31 | 5 | 3 | 0 | 3 | 1 |
|--------------------------------|----|----|----|---|---|---|---|---|

Source: [Wall Street Journal](#)

Subscribe to the [Chart of the Week](#)

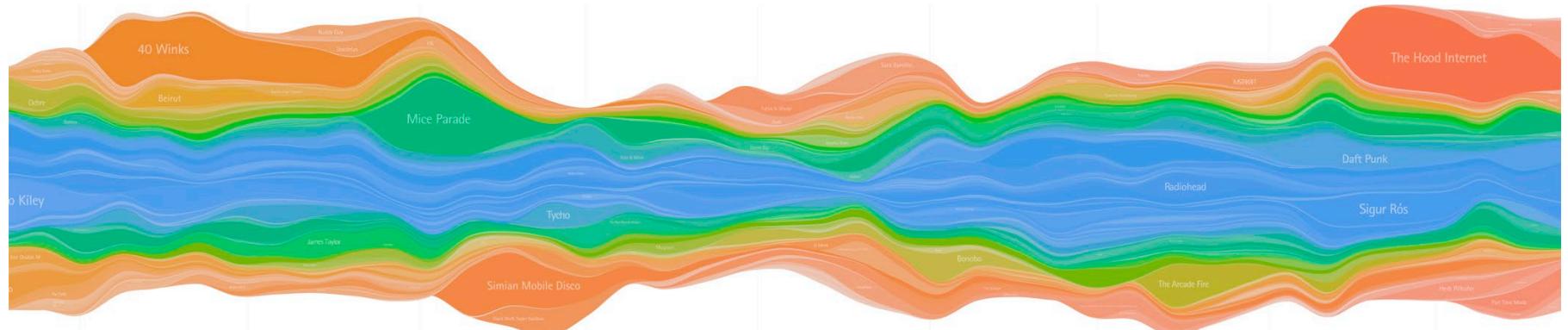
# Streamgraph

Data: 1 categorical feature, 1 ordered feature (time), 1. quant.

Derived data: geometry(layers) and layer ordering

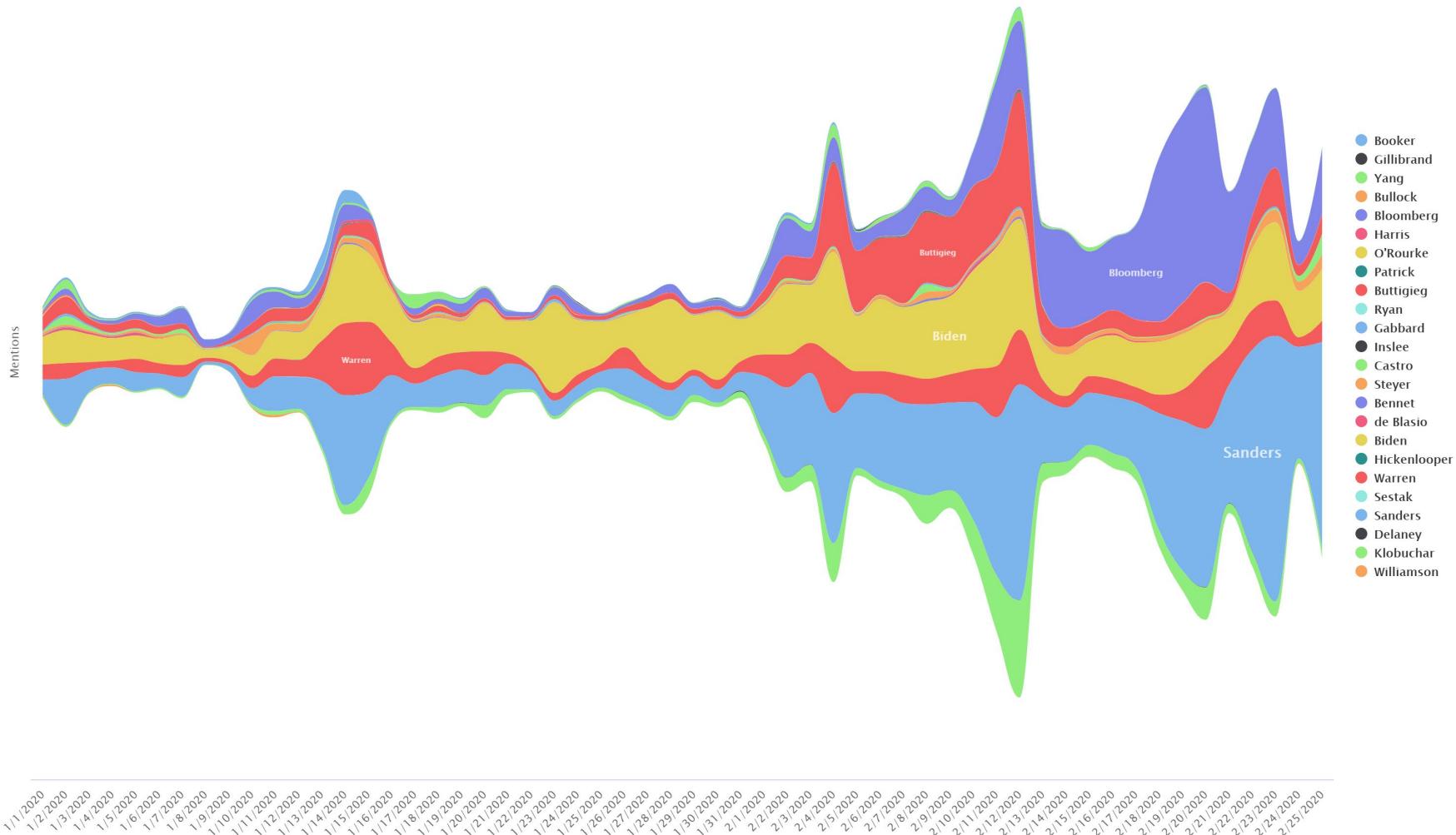
Task: Emphasizing horizontal continuity (in time)

Scalability: hundreds in time, more than stacked bar



# Streamgraph

CNN/MSNBC/FOXNEWS Candidate Mentions  
Total Mentions (1/1/2020 – 2/25/2020)



# Dot/Line Chart

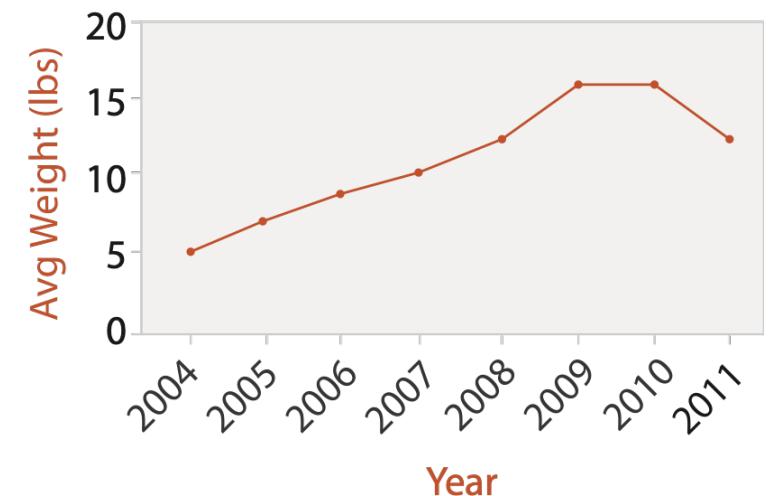
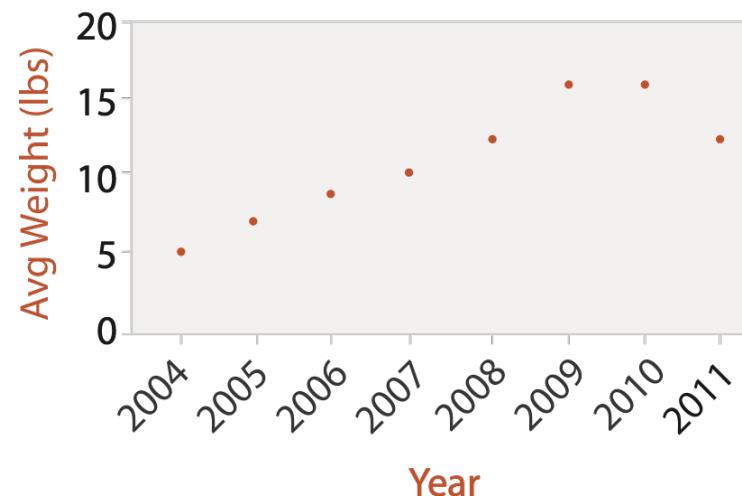
Data: 1 ordered (quant.) feature and 1 quant.

Mark: point(s) and link (line connection)

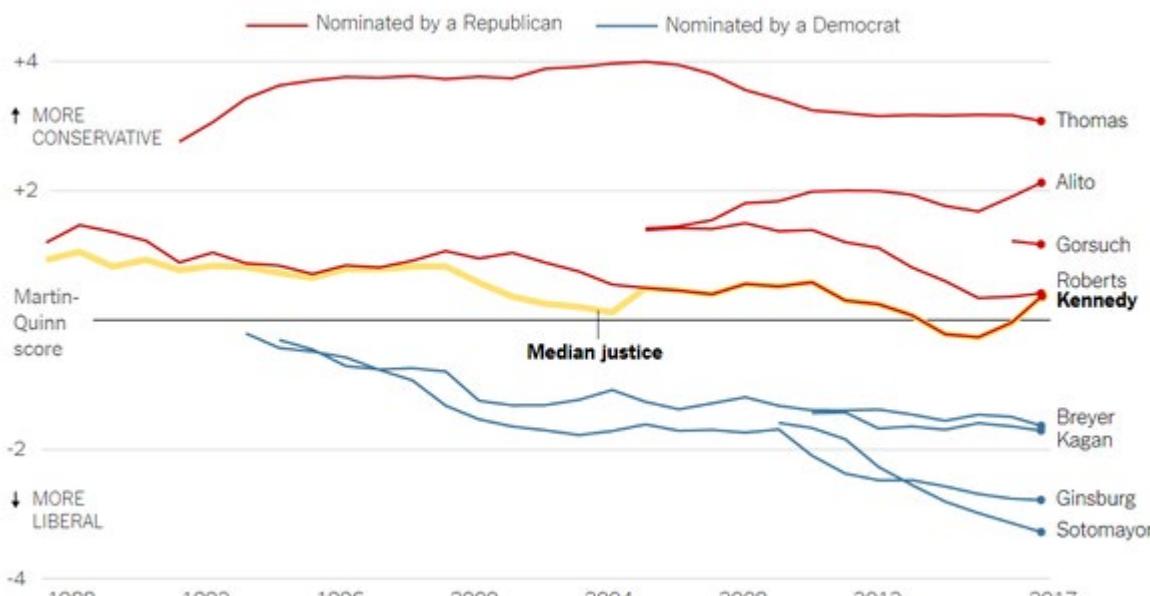
Channel: position (horizontal) for 1 ordered  
position (vertical) for 1 quant.

Task: Find trends, outlier. - link emphasizes ordering of items

Scalability: hundreds.



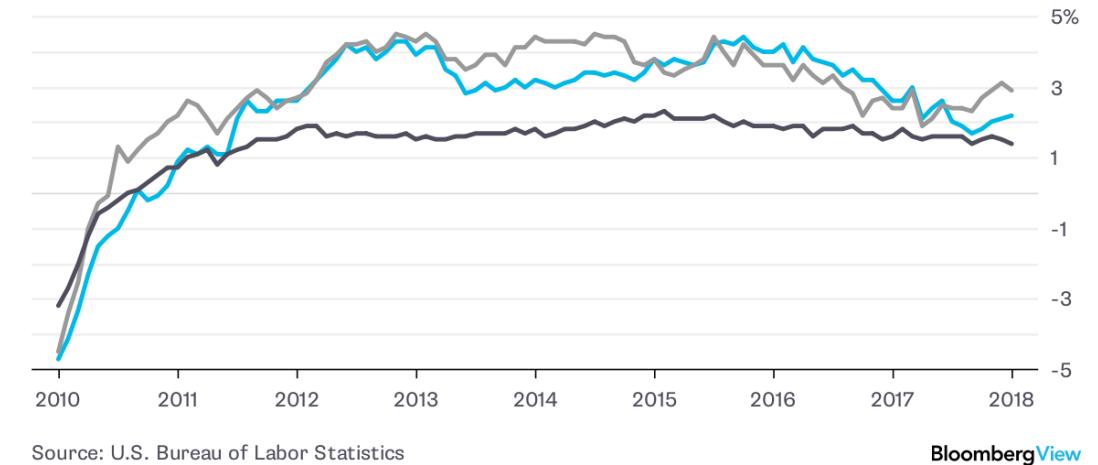
# Dot/Line Chart



## Silicon Valley Is Still Growing Faster Than the U.S.

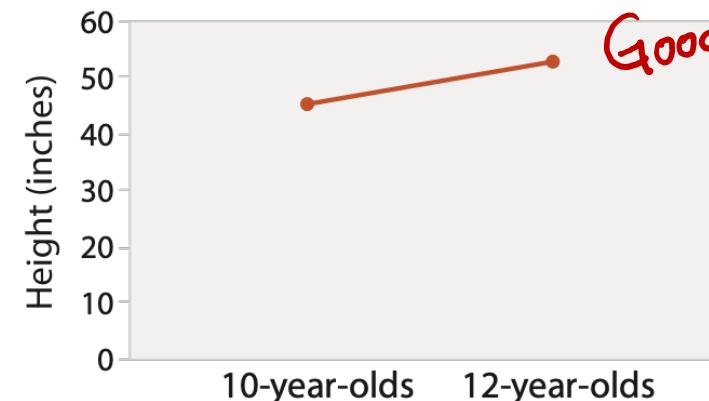
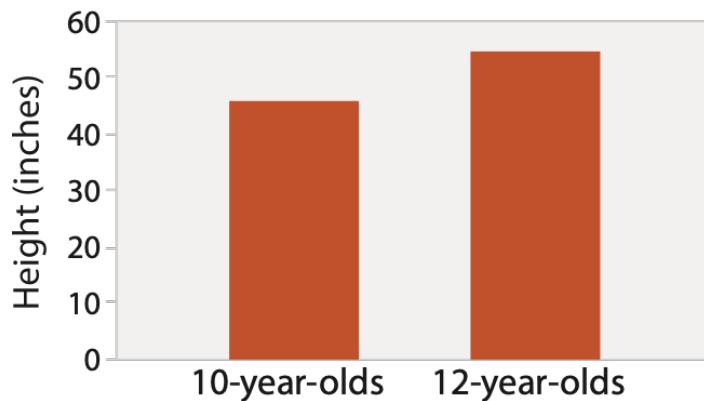
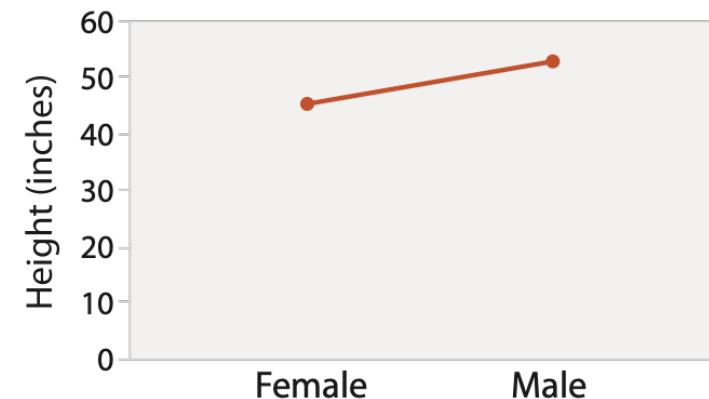
Change in nonfarm payroll employment, trailing 12 months, not seasonally adjusted

San Francisco metropolitan area    San Jose metro area    U.S.



# Bar vs Line

Categorical vs Ordered (quantitative)



# Indexed Chart

Data: 1 ordered feature and 1 derived (quant.) feature

index = change based on baseline

$$= \frac{\text{Value}}{\text{baseline}} \times 100$$

Mark: points) and link(s)

Channel: position (horizontal) for ordered  
position (vertical) for derived quant.

Task: Show "change" in time w.r.t. baseline .

Scalability: SAME

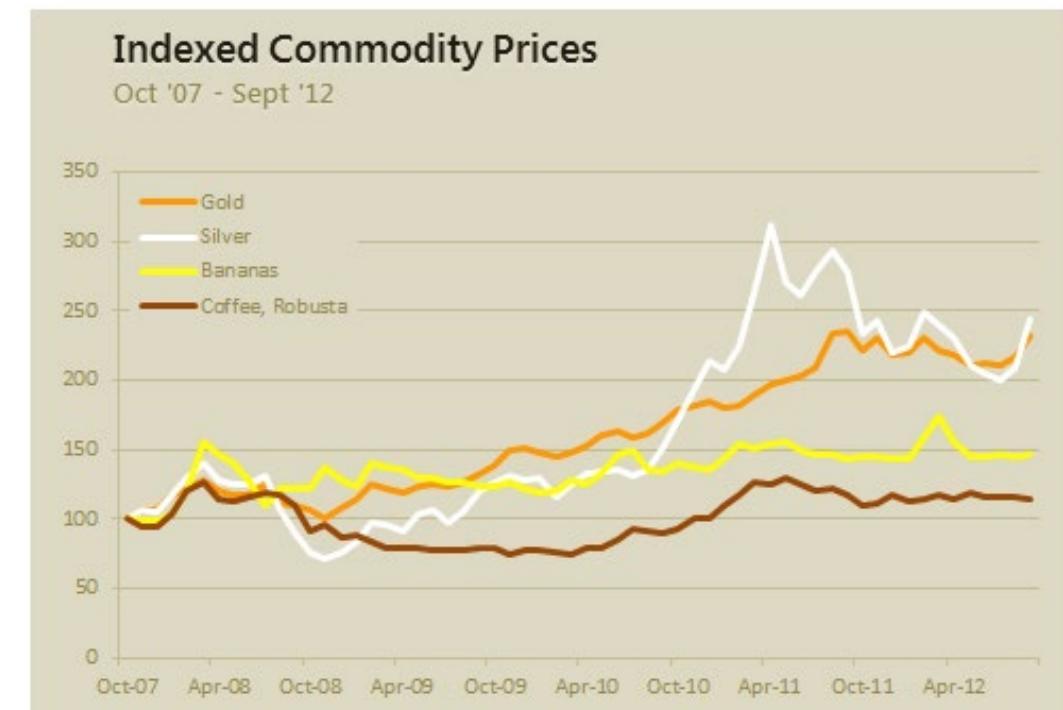
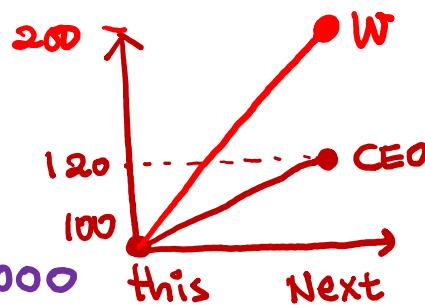
Ethics!!!

CEO \$ 100,000

\$ 120,000

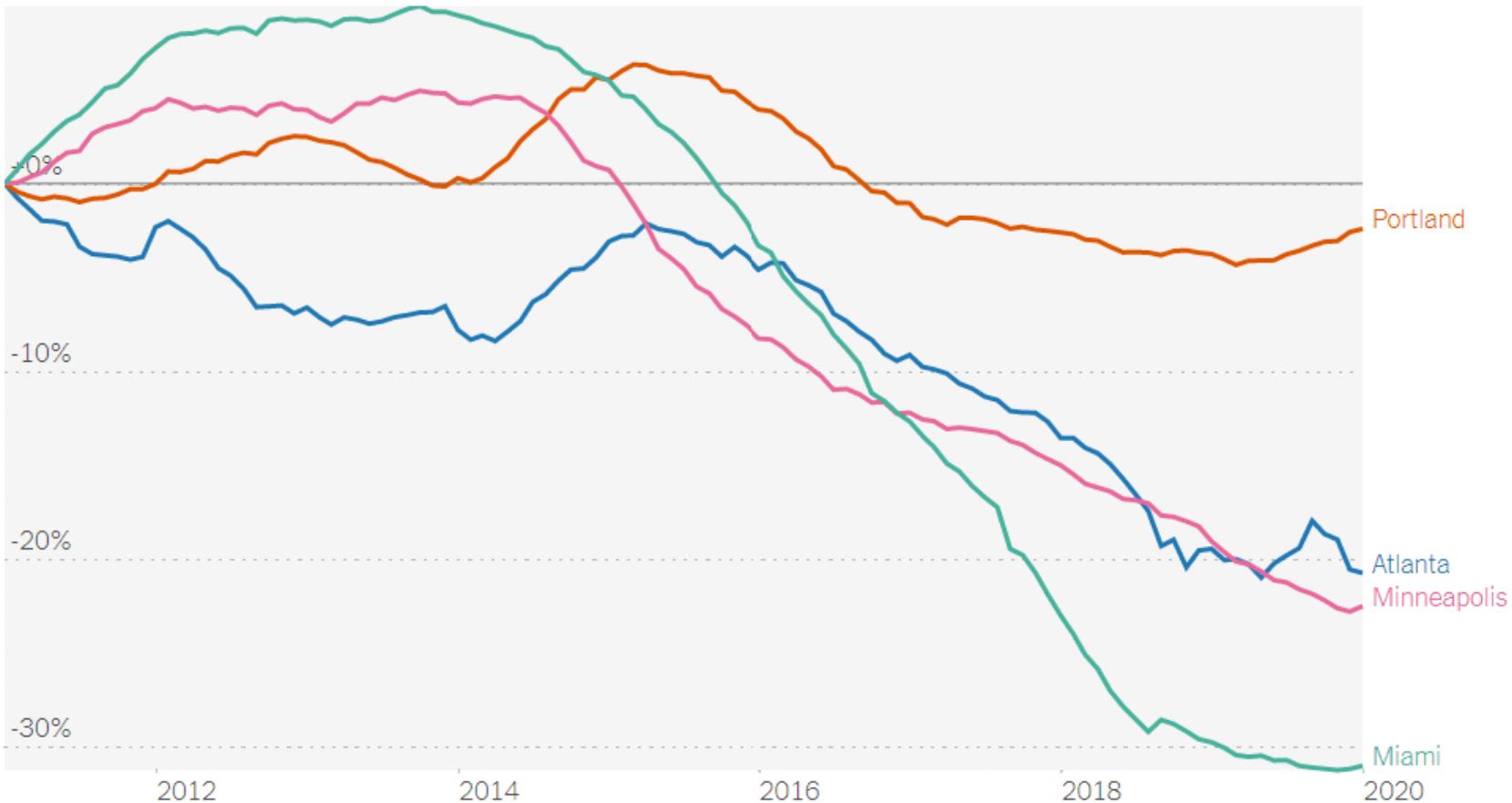
\$ 600

W \$ 300



# Indexed Chart

**Bus ridership in Miami, Minneapolis, Atlanta and Portland, Ore.**



# Pie Chart

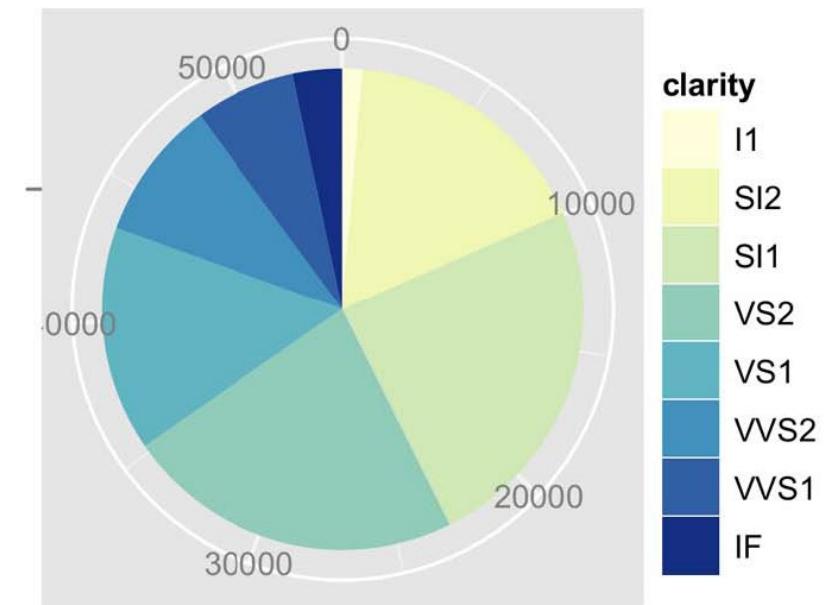
Data: 1 categorical data & 1 quant.

Mark: Area (interlocking) radially separated with uniform height (radius)

Channel: Angle (2D Area)

Task: part - to - whole judgement .

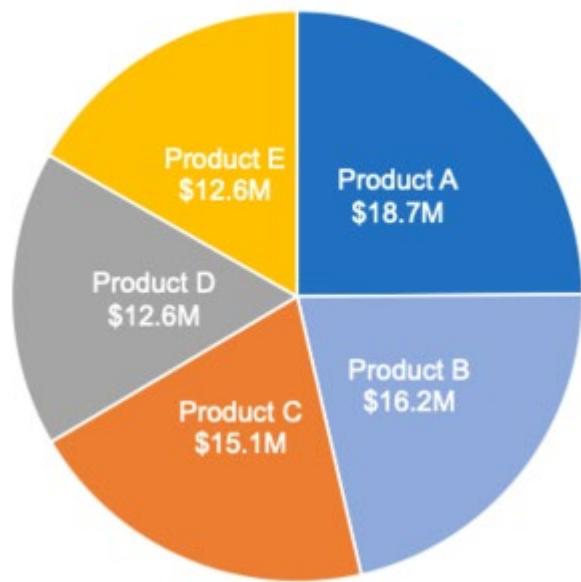
Scalability: Few .



# Pie Chart

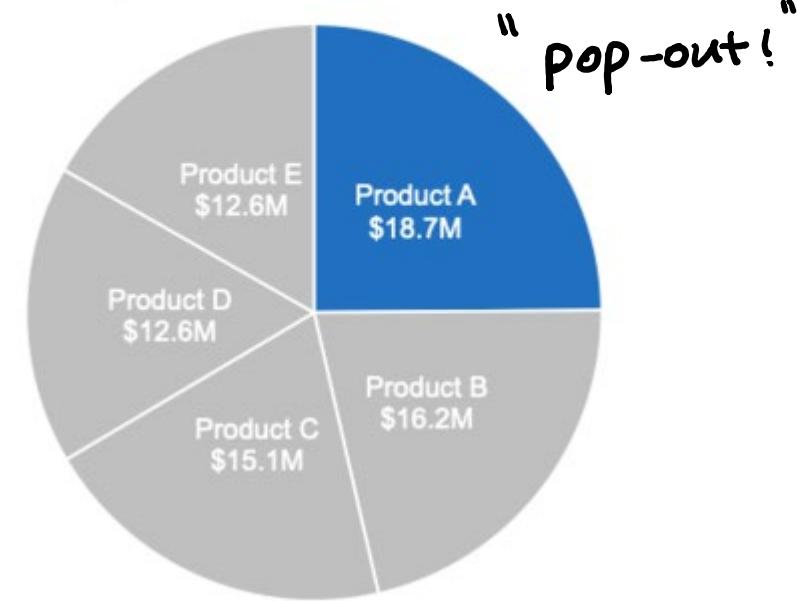
Sales by product

TOTAL = \$75.2M



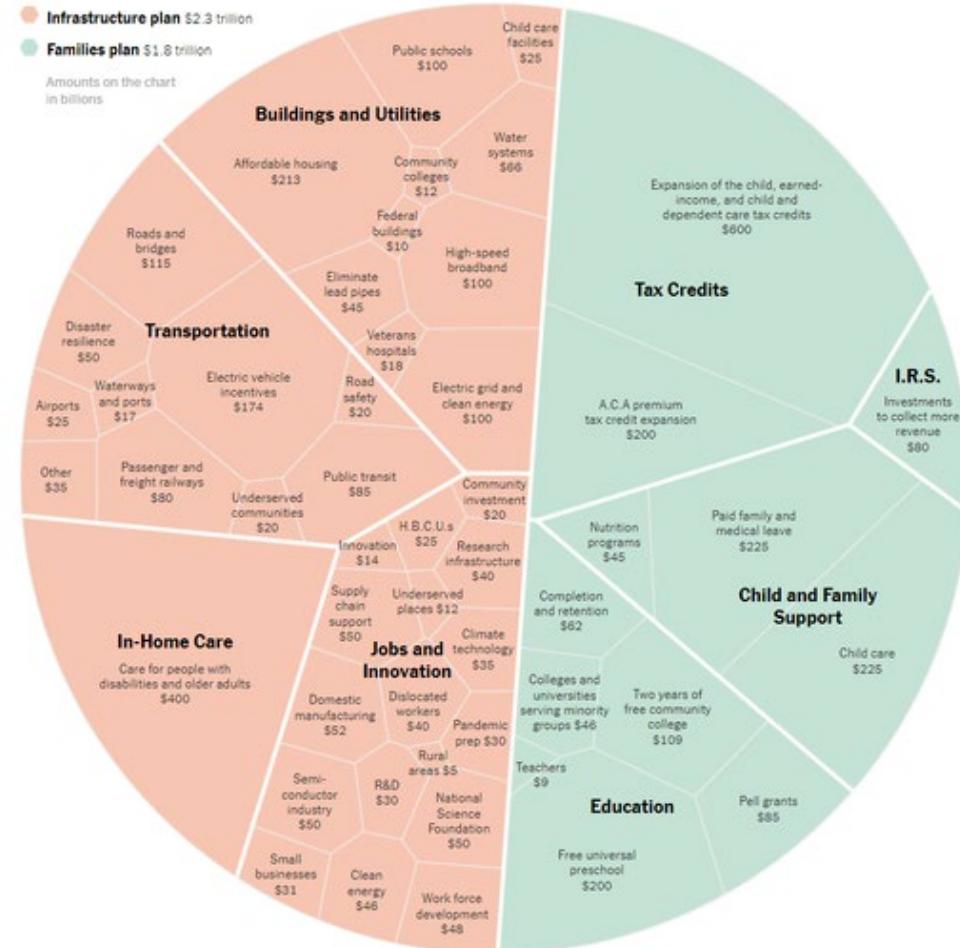
Sales by product

TOTAL = \$75.2M



# Circular Vornoi Treemap

Not Pie chart!



# Coxcomb Chart (Nightingale Rose Chart)

Data: 1 categorical & 1 quant.

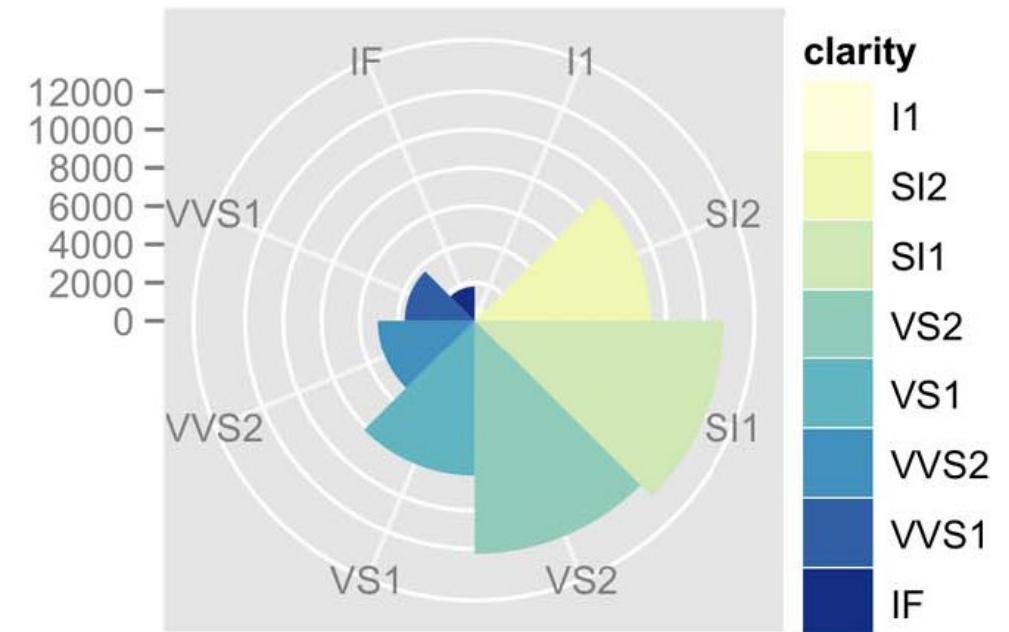
Mark: Line, radially separated with uniform width.

Channel: Size (length)

Task: Part-to-Whole judgement

Scalability: Few

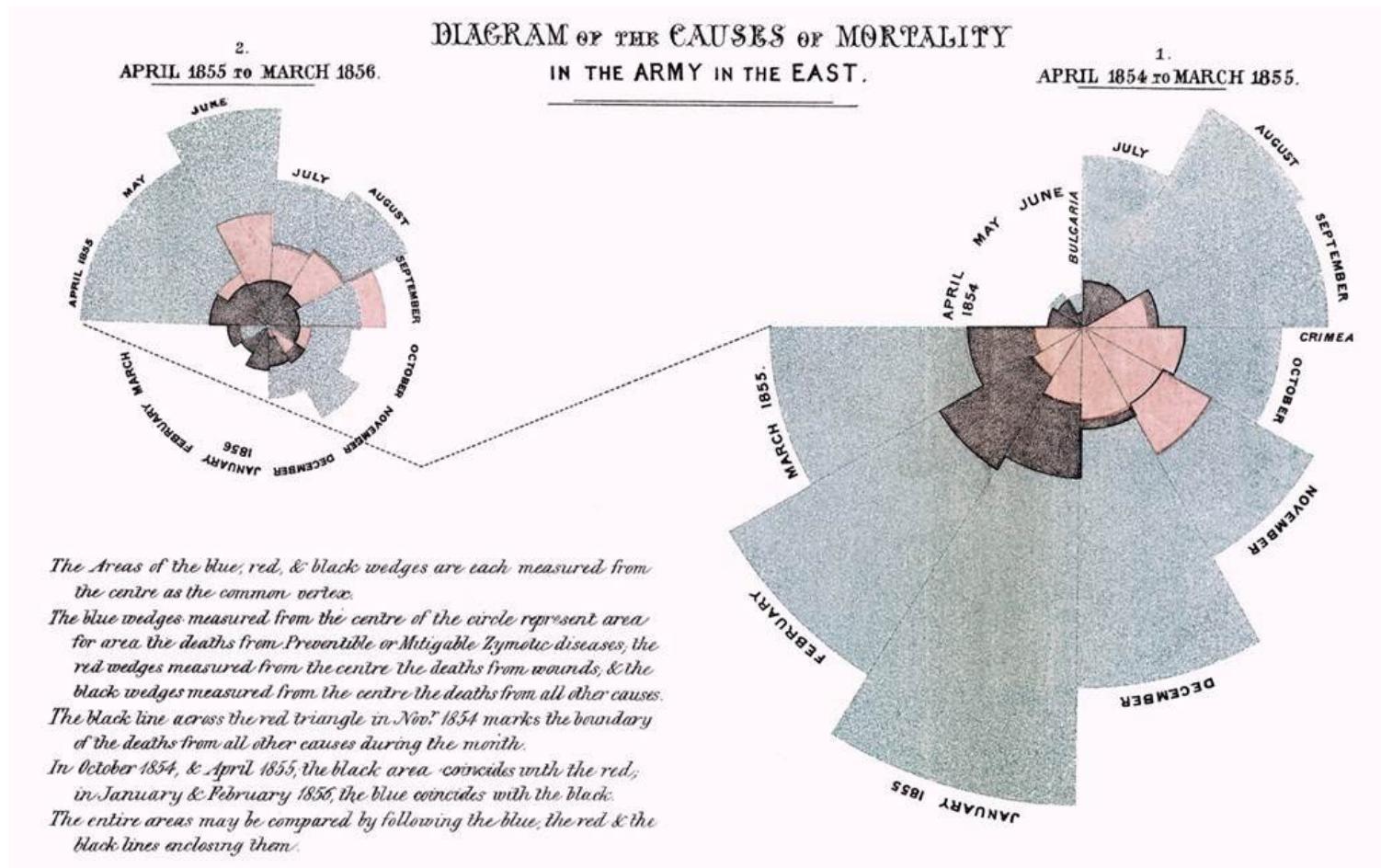
*Direct analog to radial bar chart.*



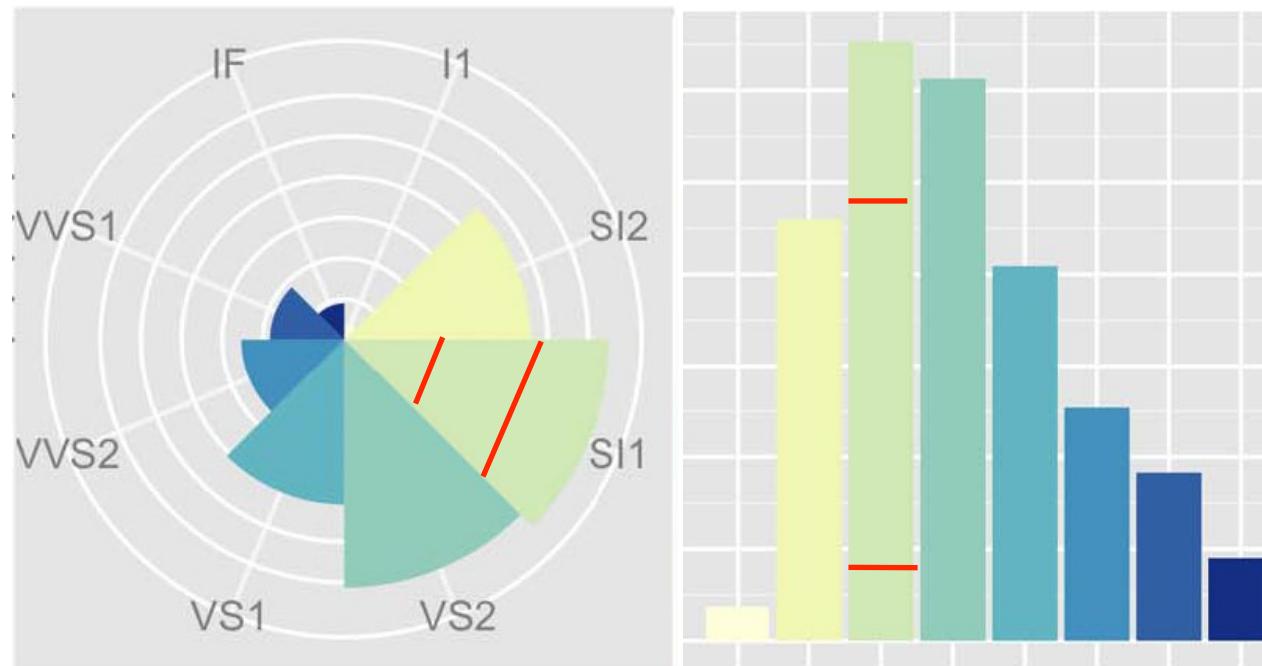
# History of Coxcomb Chart

Invented by Florence Nightingale.

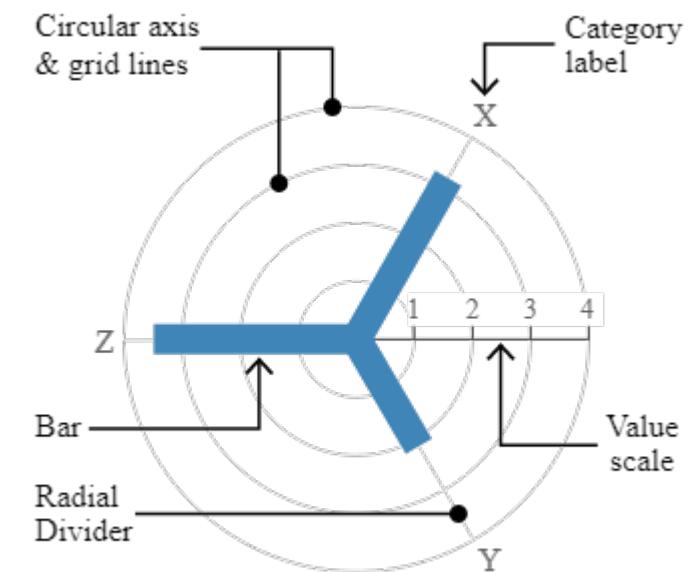
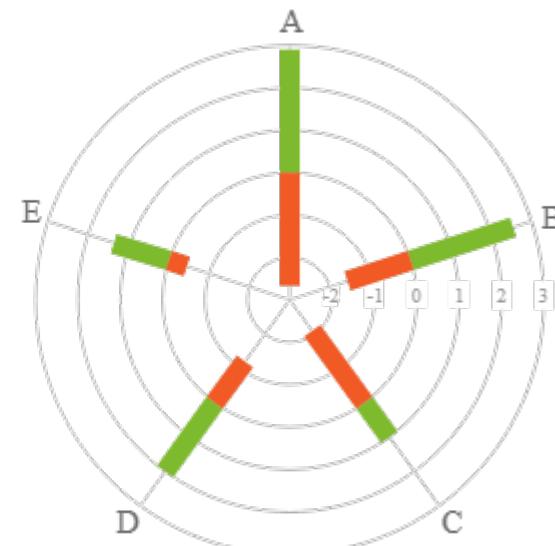
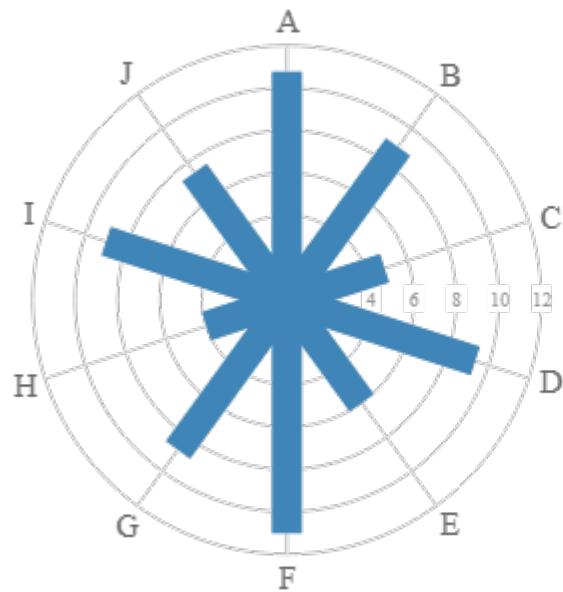
# Diagram of the Causes of Mortality in the Army in the East.



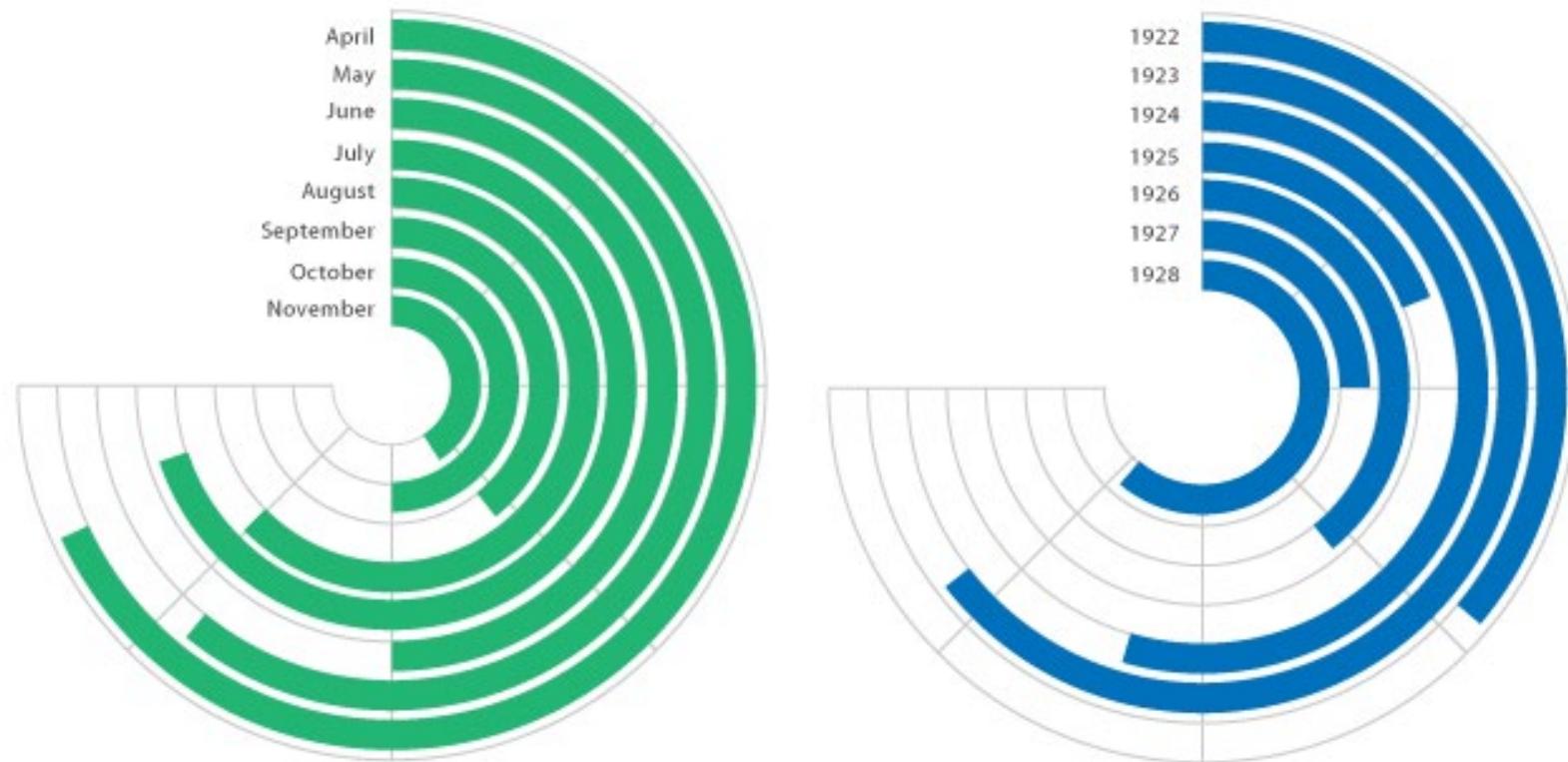
# Bar Chart and Coxcomb Chart



# Radial Column Chart (Star Chart)



# Radar Bar Chart



# Radar Plot

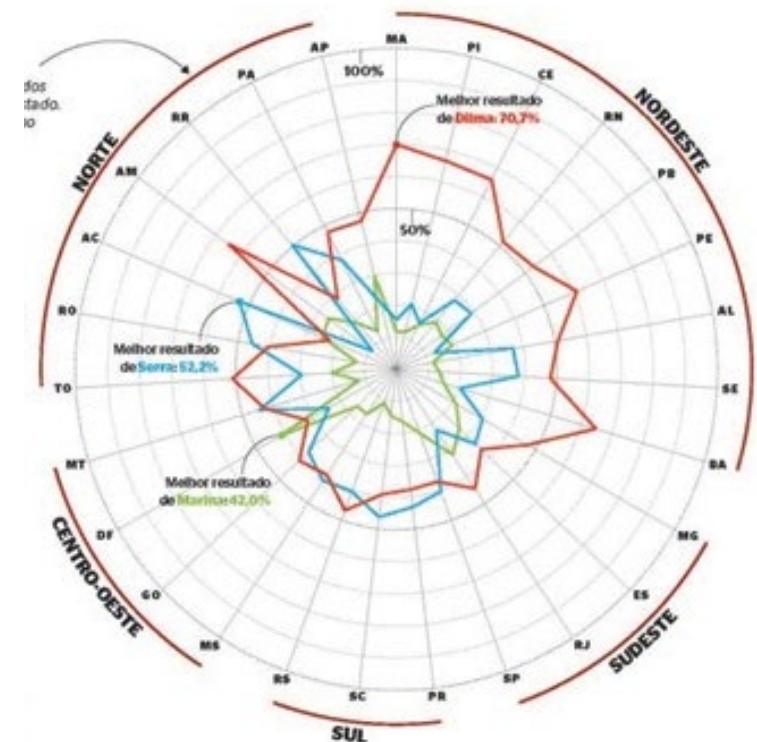
Data:

Mark:

Channel:

Task:

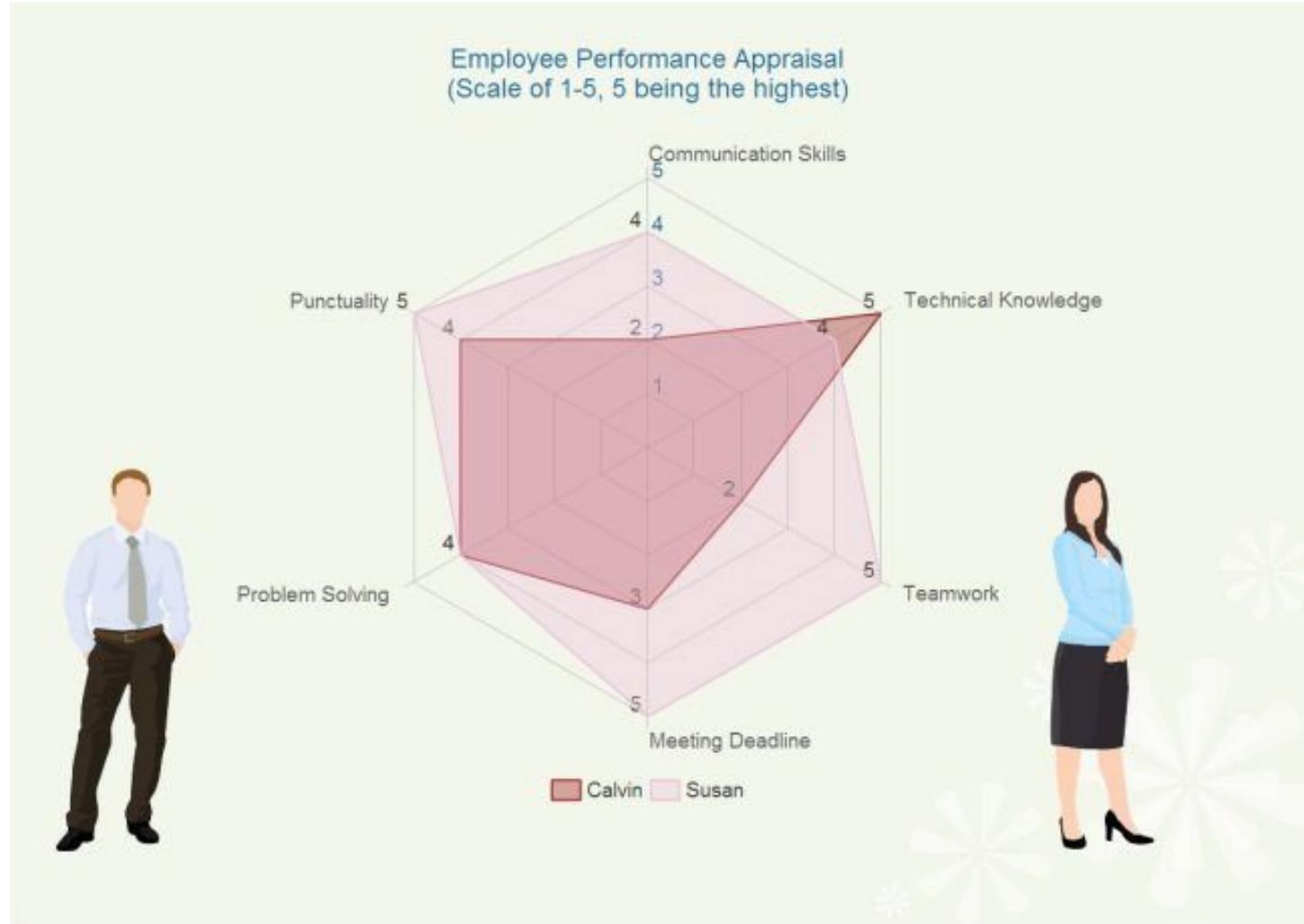
Scalability:



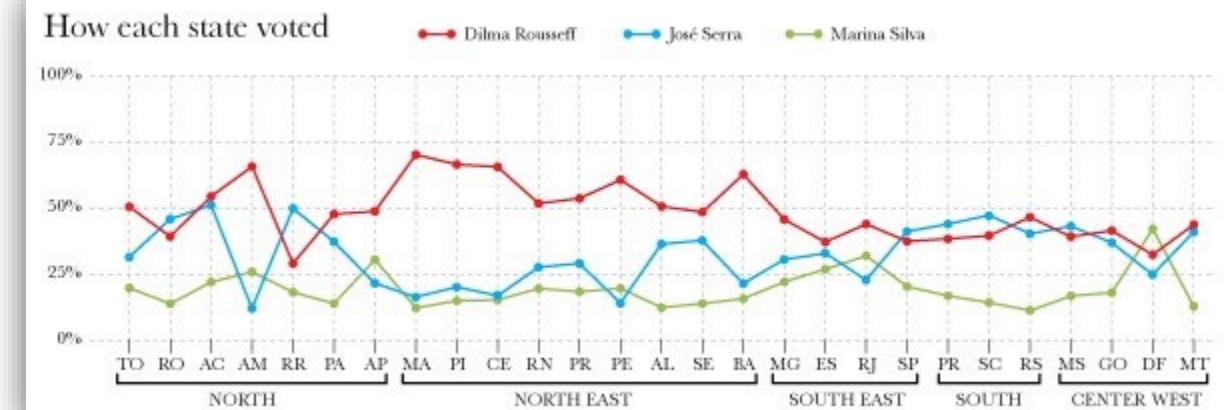
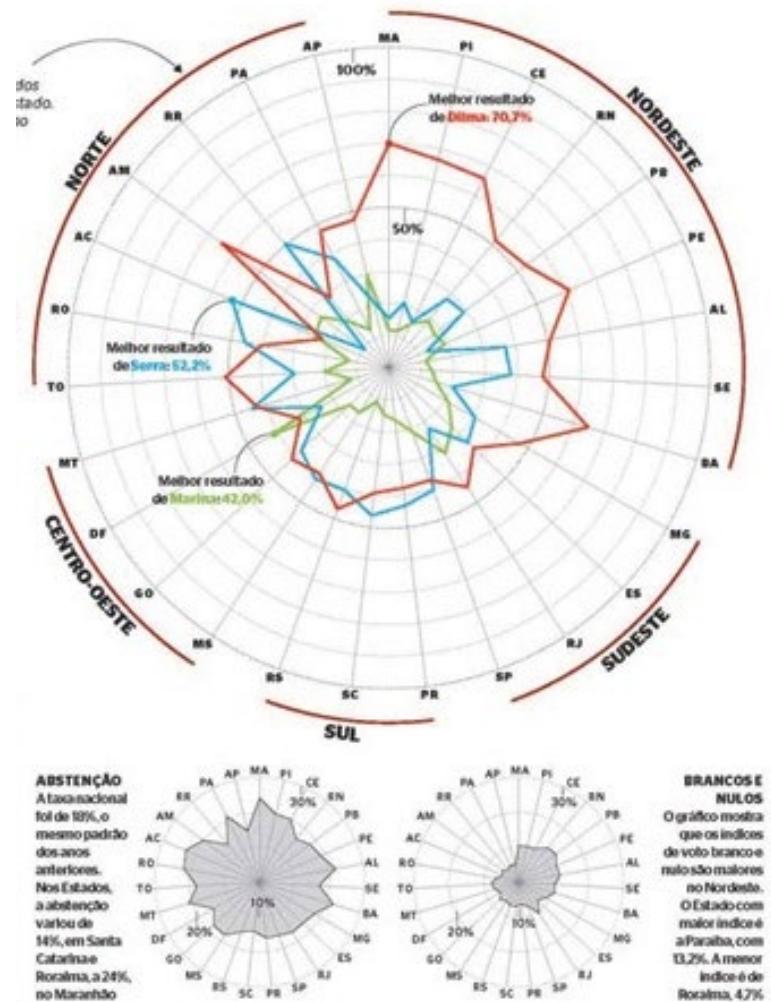
**ABSTENÇÃO**  
A taxa nacional foi de 18%, o mesmo padrão dos anos anteriores. Nos Estados, a abstenção variou de 14%, em Santa Catarina e Roraima, a 24%, no Macapá.

**BRANCOS E NULOS**  
O gráfico mostra que os índices de voto branco e nulo são maiores no Nordeste. O Estado com maior índice é a Paraíba, com 13,2%. A menor índice é de Roraima, 4,7%.

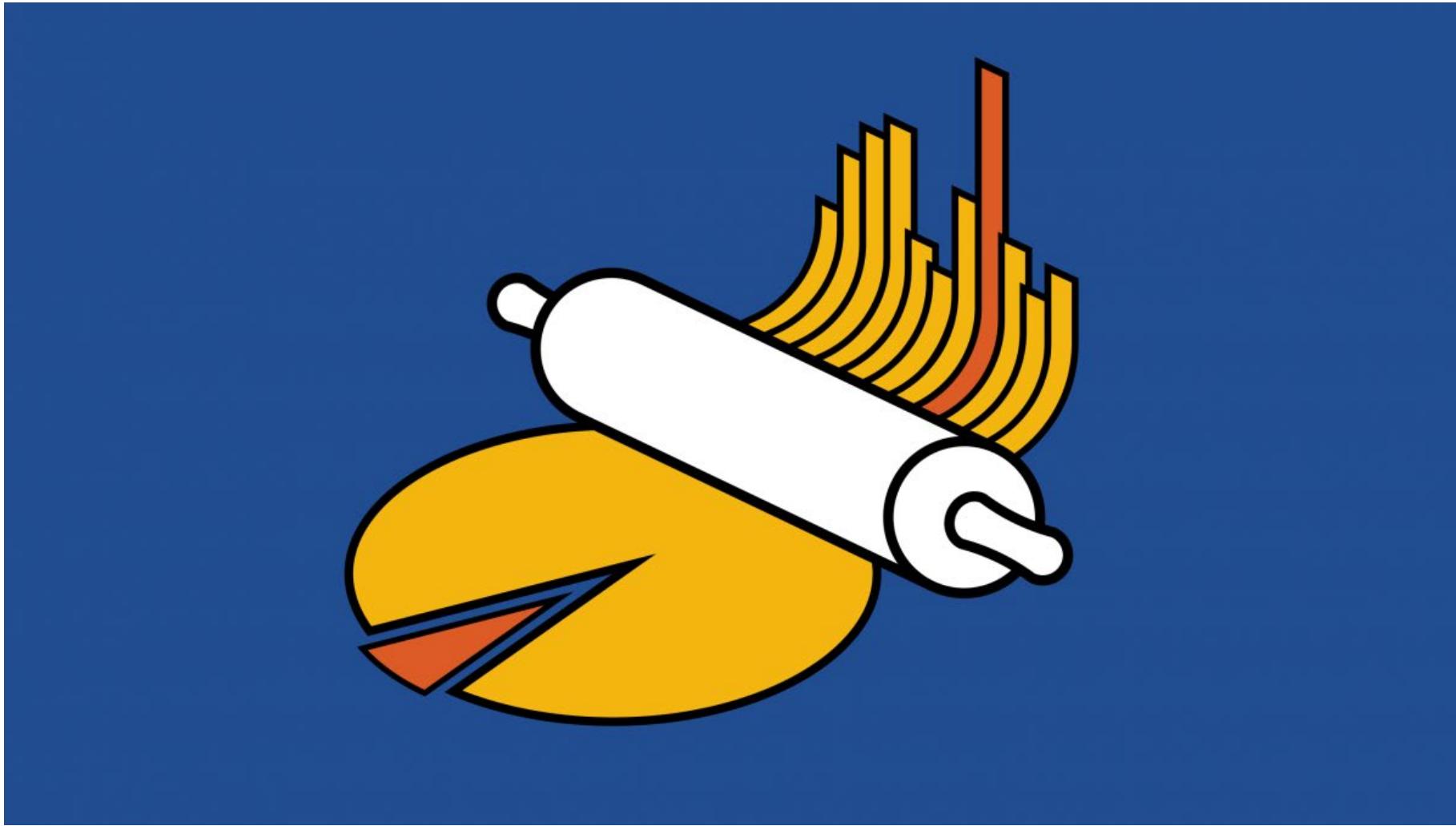
# Radar Plot



# Radar Plot



# Pie vs Bar



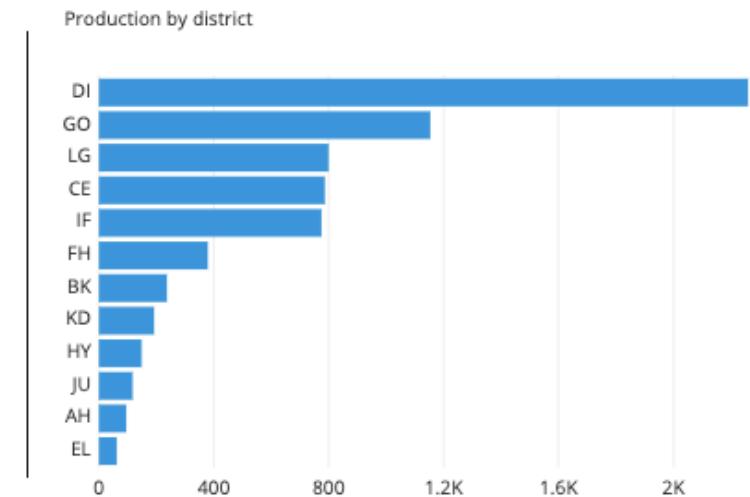
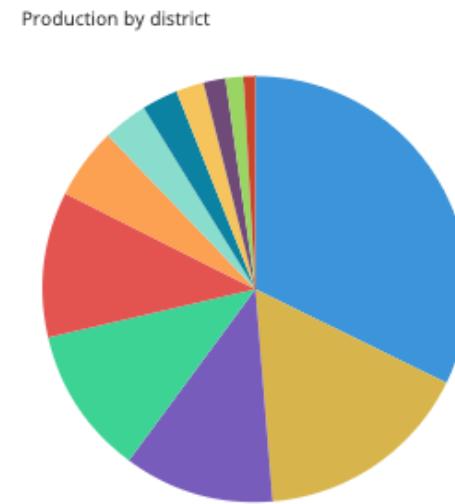
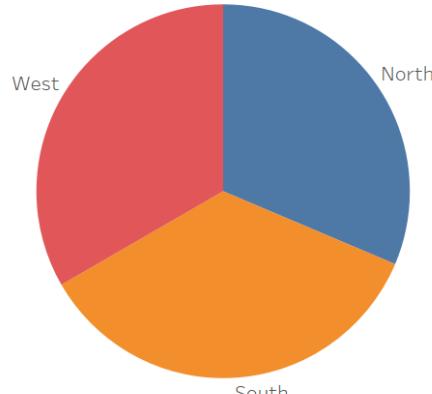
# Pie vs Bar

**Bar chart** is default (unframed and aligned)!

Why not pie chart?

- Multiple slices have similar values: hard to compare (unframed and unaligned).
- Too many slice: hard to understand.

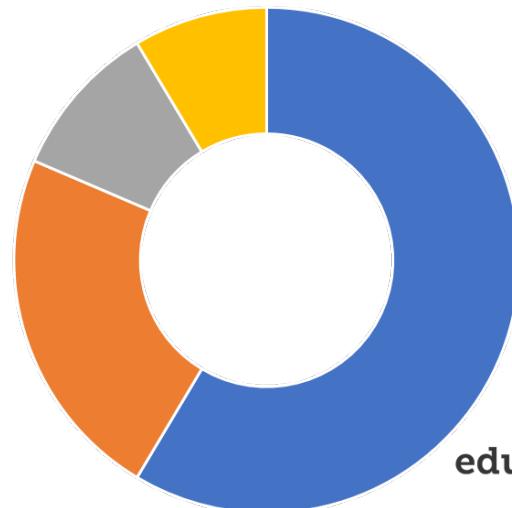
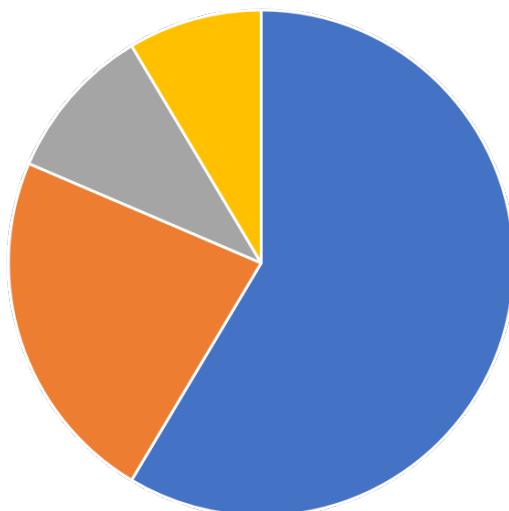
Want to show a part-to-whole comparison?



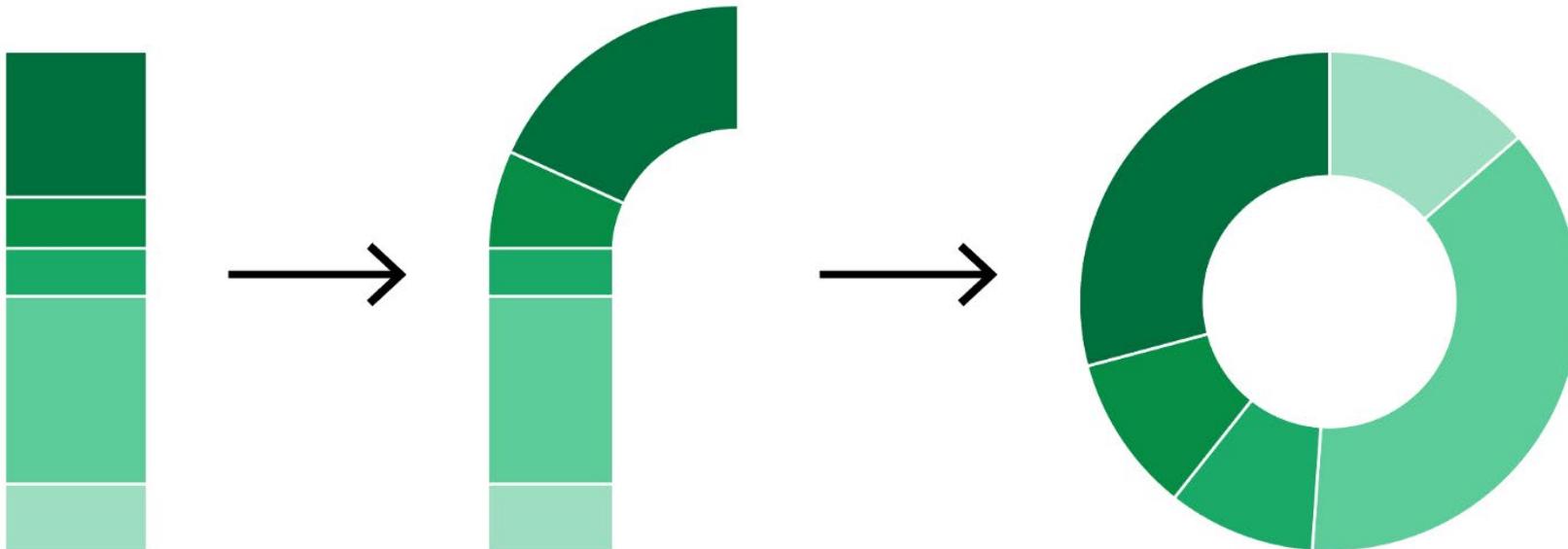
# Pie vs Doughnut

**Do you really want to use a pie chart?**

Use doughnut chart instead (in practice).



# Pie vs Doughnut



# Pie vs Doughnut

