

Data Science II

- Introduction to Data Visualization -

Visualizing Part-to-Whole Relationships and Relationships



Prof. Dr. Eduard Kromer
Summer Semester 2024
University of Applied Sciences Landshut

Visualizing Part-to-Whole Relationships

Part-to-Whole

- this class of charts shows how the shares of some amount relate to the total
- most popular graph in this class is the [pie chart](#)
 - alternatives: [treemap](#) and [sunburst](#)
- graphs in this class can also be used to visualize hierarchical data



Pie Chart



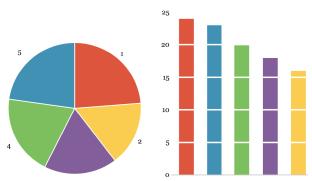
- the pie chart is very familiar to many people and familiarity can be useful
- most important rule for pie charts:** the slices must sum to 100 percent
 - you cannot leave segments out or include segments that sum to more than 100 percent

Pie Chart



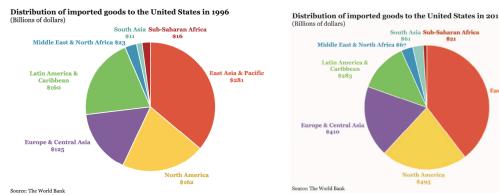
arrangement of slices: order from largest to smallest beginning at the 12-o'clock position
[sometimes it is better to order by category and not by value]

Problems with Pie Charts



- humans cannot easily compare differently sized slices
- if the goal is to help the reader make clear and accurate determinations about the data, the pie chart is not the best choice
- avoid using pie charts with more than 5 slices (more becomes incomprehensible)

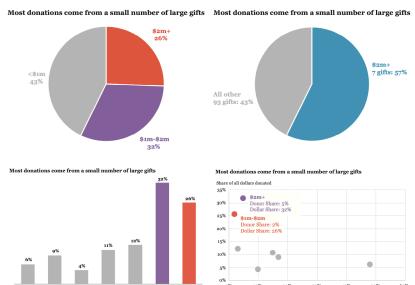
Problems with Pie Charts



Avoid using pairs of pie charts to show changes over time.



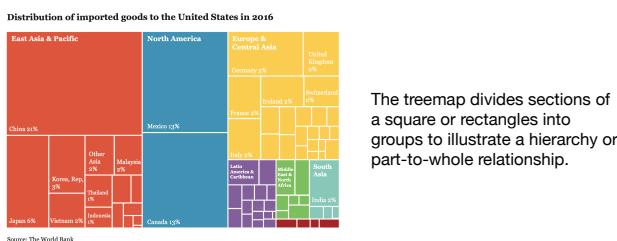
Advantages of Pie Charts



You can read and understand a pie chart almost immediately.



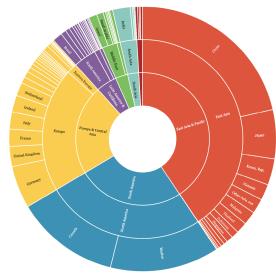
Treemap



The treemap divides sections of a square or rectangles into groups to illustrate a hierarchy or part-to-whole relationship.



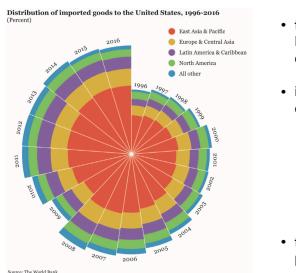
Sunburst Diagram



The sunburst diagram shows the proportions of parts to a whole at several levels in a hierarchy.

Each ring in a sunburst diagram corresponds to a different level in the hierarchy and the slices of each ring refer to the different subgroups.

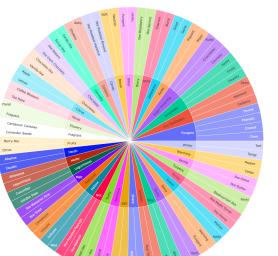
Nightingale Chart



- the **Nightingale chart** (created by Florence Nightingale) is sometimes called the **coxcomb** or **rose diagram**
 - in this chart type the slices have been expanded in different directions
 - ▶ the area of each slice represents its value relative to the whole
 - ▶ the slices are arranged along the time dimension
 - this chart shows **both** changes over time and part-to-whole relationship

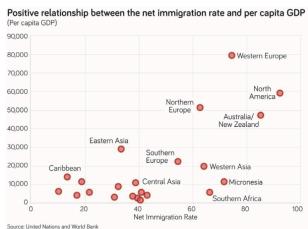
Visualizing Sunburst Charts

- read the [plotly tutorial](#) for sunburst charts
 - use the data in `coffee.csv` to generate a sunburst chart with plotly



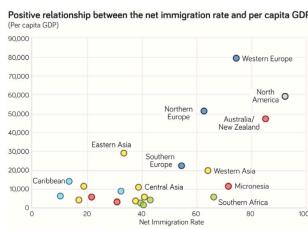
Visualizing Relationships

Scatterplot



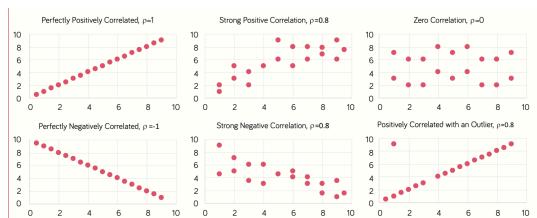
- the most common visualization to illustrate correlations between two variables
 - one variable is plotted along a horizontal axis and the other one along a vertical axis
- the axes do not need to start at zero

Scatterplot

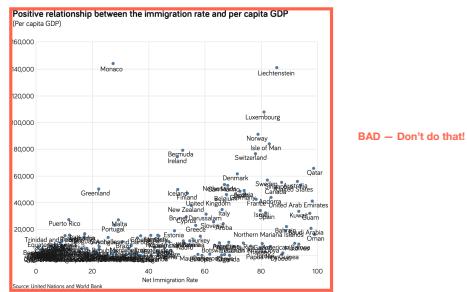


- a scatterplot can help the reader see whether two variables are associated with one another
- two variables can be *positively* or *negatively* or *not correlated*

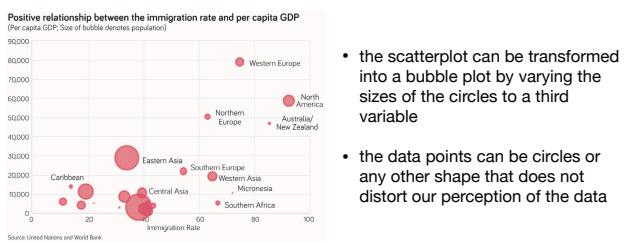
Scatterplot



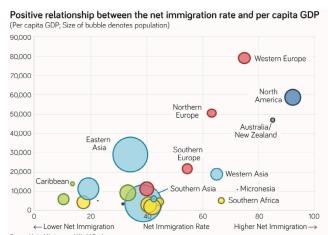
Scatter Plot



Bubble Plot

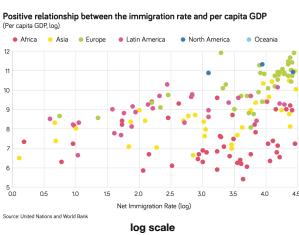
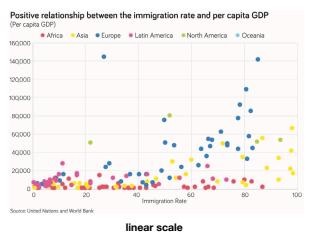


Bubble Plot

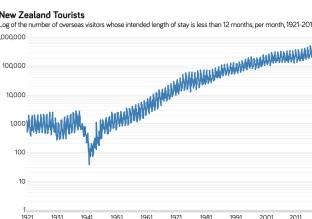
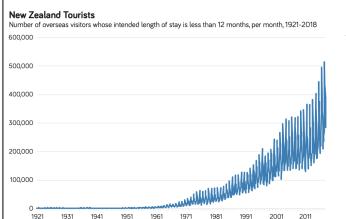


- more colors can be added to denote another variable, such as the region of the world

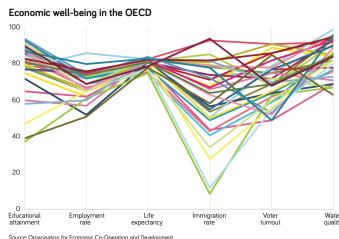
Choosing the Right Scale



Choosing the Right Scale

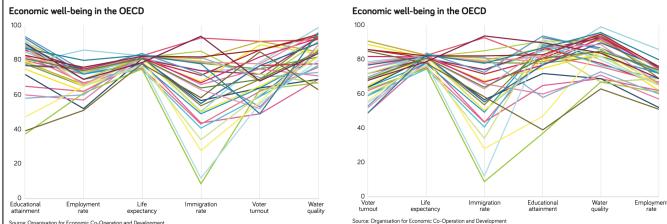


Parallel Coordinates Plot



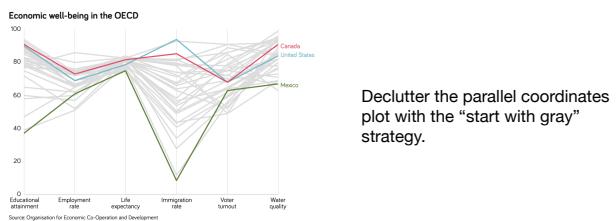
- data values are plotted along multiple vertical axes and connected by lines
- the axes can have different units of measurement (or can be normalized to keep scales uniform)
- the parallel coordinates plot permits multiple correlations within a single view

Parallel Coordinates Plot



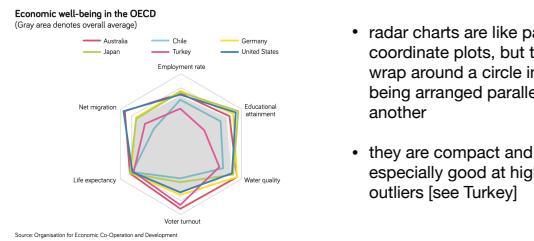
Changing the order of the vertical axes lets us discover different relationships.

Parallel Coordinates Plot



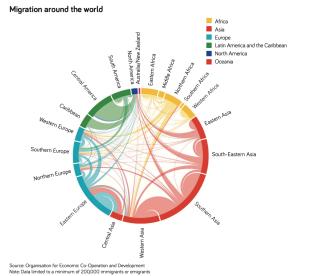
Declutter the parallel coordinates plot with the "start with gray" strategy.

Radar Charts



- radar charts are like parallel coordinate plots, but the lines wrap around a circle instead of being arranged parallel to one another
 - they are compact and especially good at highlighting outliers [see Turkey]

Chord Diagram

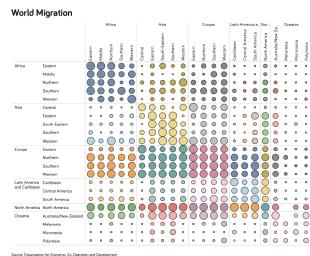


- in chord diagrams, observations [nodes] are located around the circumference of the circle and connected by arcs within the circle to illustrate connections
 - the thickness of the arcs represent the degree of the connection between the different groups

Correlation Matrix

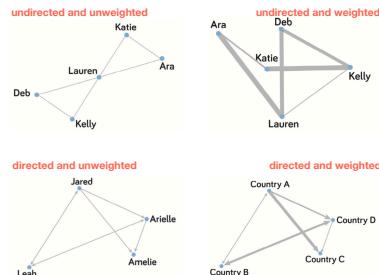
- a correlation matrix is a table with the variables listed along the horizontal and vertical axes
 - numbers in each cell represent the strength of that relationship (Pearson's correlation coefficient)

Correlation Matrix



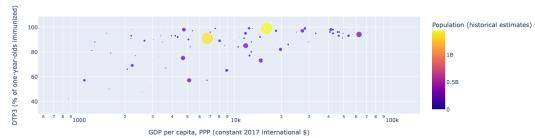
- use circles or other shapes and use color to organize the space

Network Diagrams



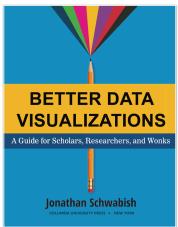
Visualizing Relationships

- use the data set [dtp3-vs-gdp-per-capita.csv](#) to generate a bubble plot similar to this one using [matplotlib](#), [seaborn](#) or [plotly](#)



- What can you improve about this chart?

Literature



References

- Slide 73-80, 83-98; Image Source: J. Schwabisch - Better Data Visualizations, Columbia University Press