

Lecture 21:

Visualization 2

COSC 480 Data Science, Spring 2017
Michael Hay

Logistics

- Project mini-presentations next Wednesday
- See the (updated) `project.md` description for details.

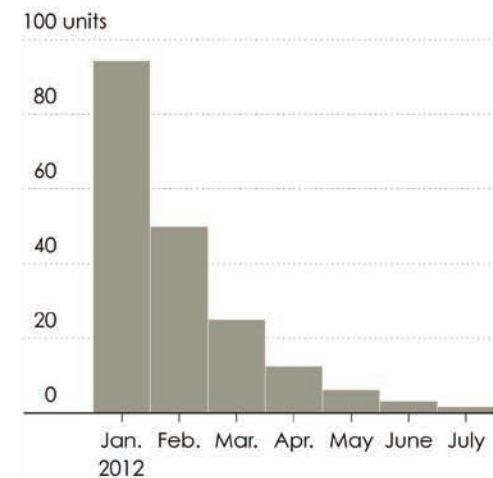
Visualization components

Working parts

Several pieces work together to make a graph. Sometimes these are explicitly shown in the visualization and other times they form a visual in the background. They all depend on the data.

Title of this Graph

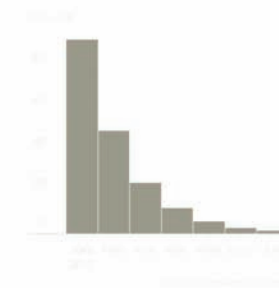
A description of the data or something worth highlighting to set the stage.



Source: Somewhere reputable

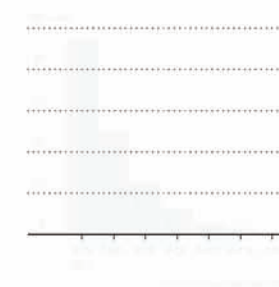
Title of this Graph

A description of the data or something worth highlighting to set the stage.



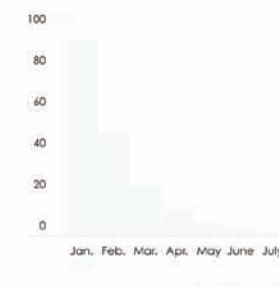
Title of this Graph

A description of the data or something worth highlighting to set the stage.



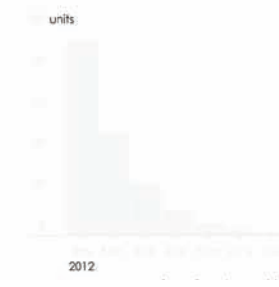
Title of this Graph

A description of the data or something worth highlighting to set the stage.



Title of this Graph

A description of the data or something worth highlighting to set the stage.



Visual Cues

Visualization involves encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals.

Coordinate System

You map data differently with a scatterplot than you do with a pie chart. It's x- and y-coordinates in one and angles with the other; it's cartesian versus polar.

Scale

Increments that make sense can increase readability, as well as shift focus.

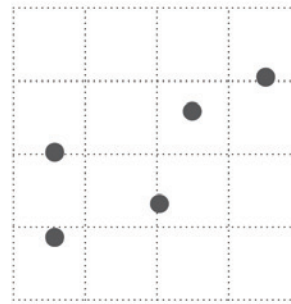
Context

If your audience is unfamiliar with the data, it's your job to clarify what values represent and explain how people should read your visualization.

Visual cues

Position

Where in space the data is



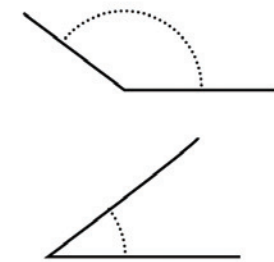
Length

How long the shapes are



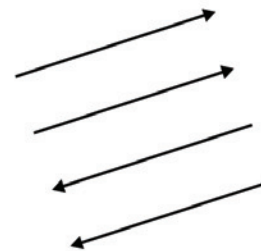
Angle

Rotation between vectors



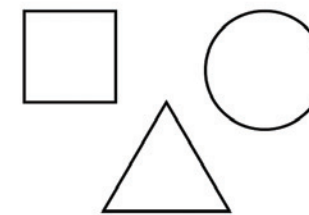
Direction

Slope of a vector in space



Shapes

Symbols as categories

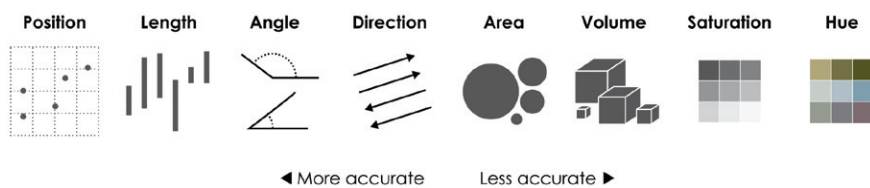


Area

How much 2-D space



Human perception



Volume

How much 3-D space



Color saturation

Intensity of a color hue



Color hue

Usually referred to as color



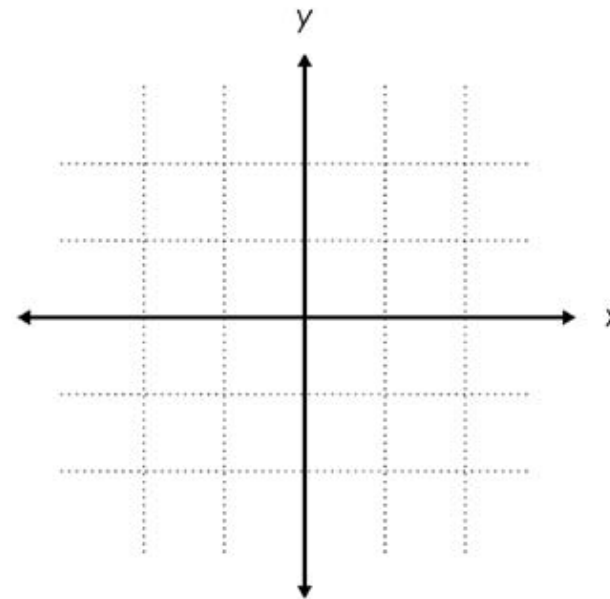
FIGURE 3-3 Visual cues

Recap

Coordinate systems

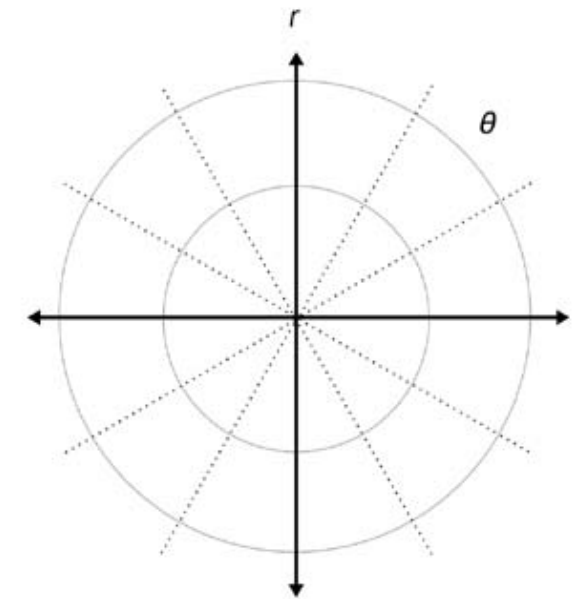
Cartesian

If you've ever made a graph, the x- and y-coordinate system will look familiar to you.



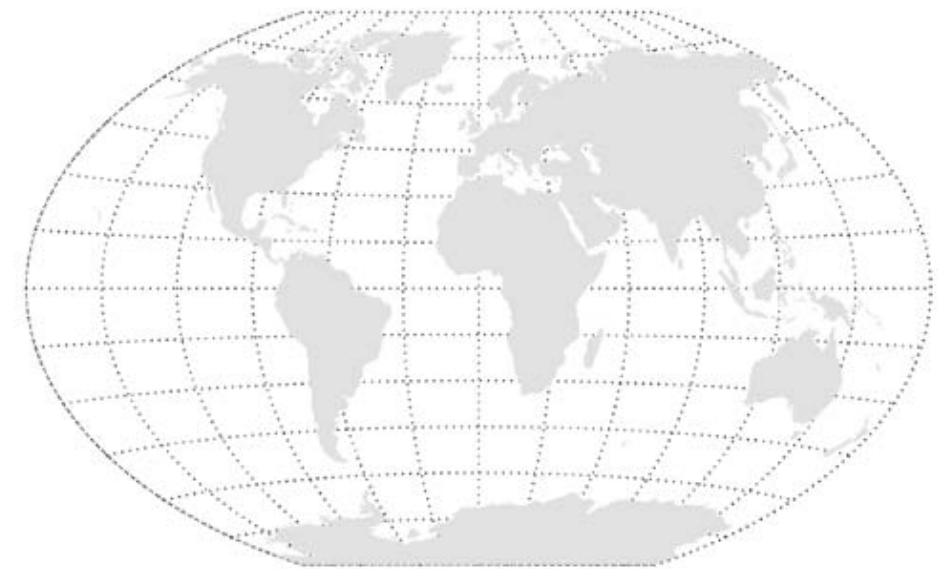
Polar

Pie charts use this system. Coordinates are placed based on radius r and angle θ .



Geographic

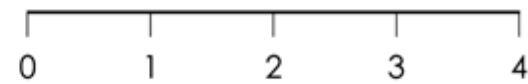
Latitude and longitude are used to identify locations in the world. Because the planet is round, there are multiple projections to display geographic data in two dimensions. This one is the Winkel tripel.



Scales

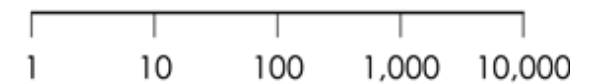
Linear

Values are evenly spaced



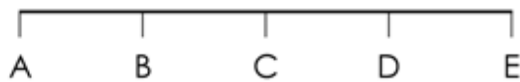
Logarithmic

Focus on percent change



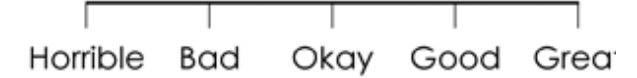
Categorical

Discrete placement in bins



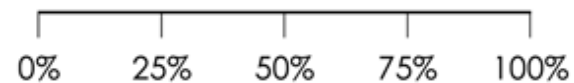
Ordinal

Categories where order matters



Percent

Representing parts of a whole



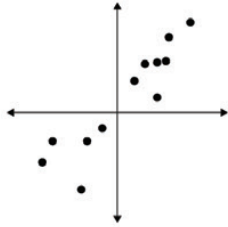

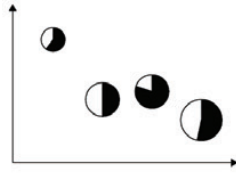

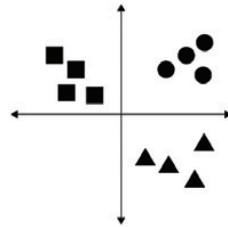
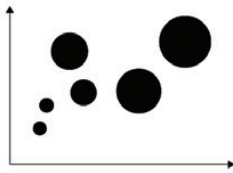
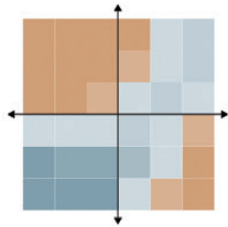
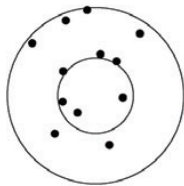


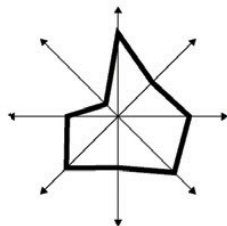
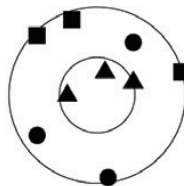
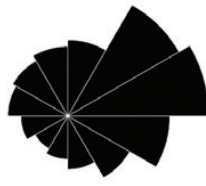
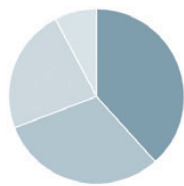







Time

Units of months, days, or hours



Time is special... why?

Recap

	Position	Length	Angle	Direction	Shapes	Area or Volume	Color
Coordinate systems							
Cartesian							
Polar							
Geographic							

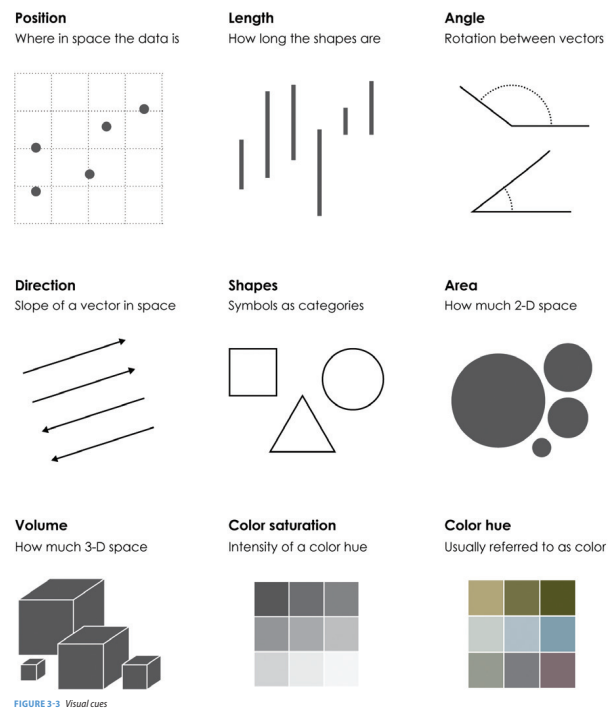
Recap

Question

One of today's readings said that this graphic was *misleading*.

Why is it misleading?

Can you explain how it misleads in terms of visual cues, coordinate systems, scales?



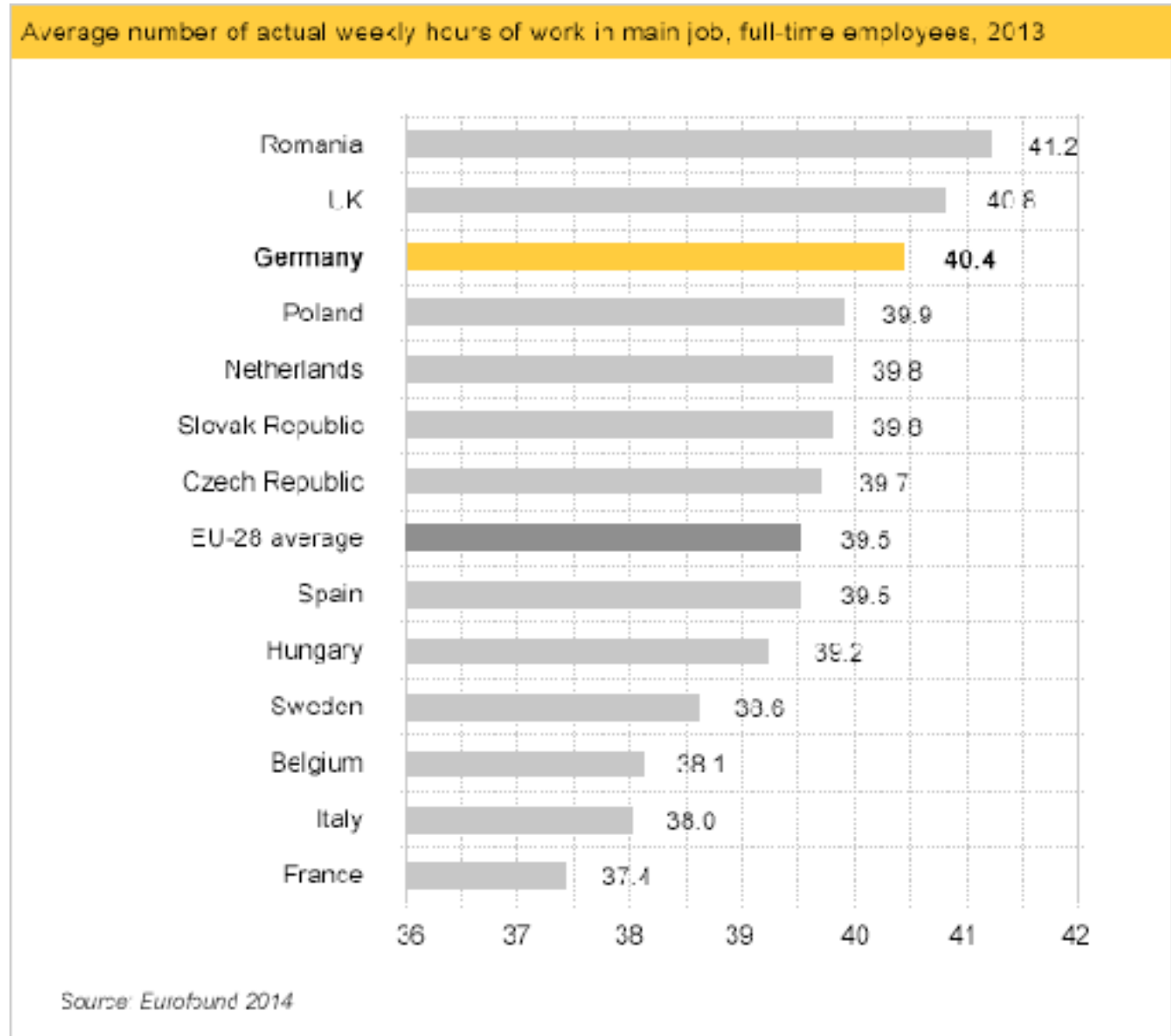
Coordinate systems:

- Cartesian (x,y)
- Polar
- Geographic

Scales:

- linear
- log
- categorical
- ordinal
- percent
- time

Instructions: ~1 minute to think/
answer on your own; then discuss with
neighbors; then I will call on one of you

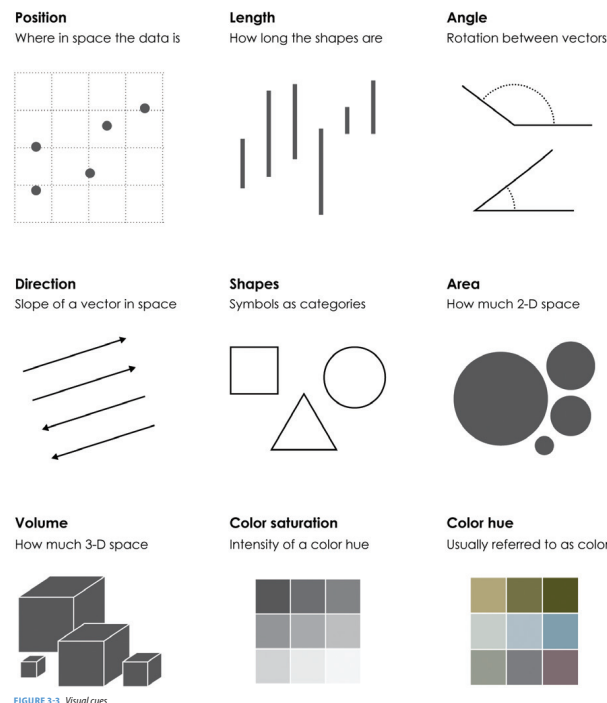
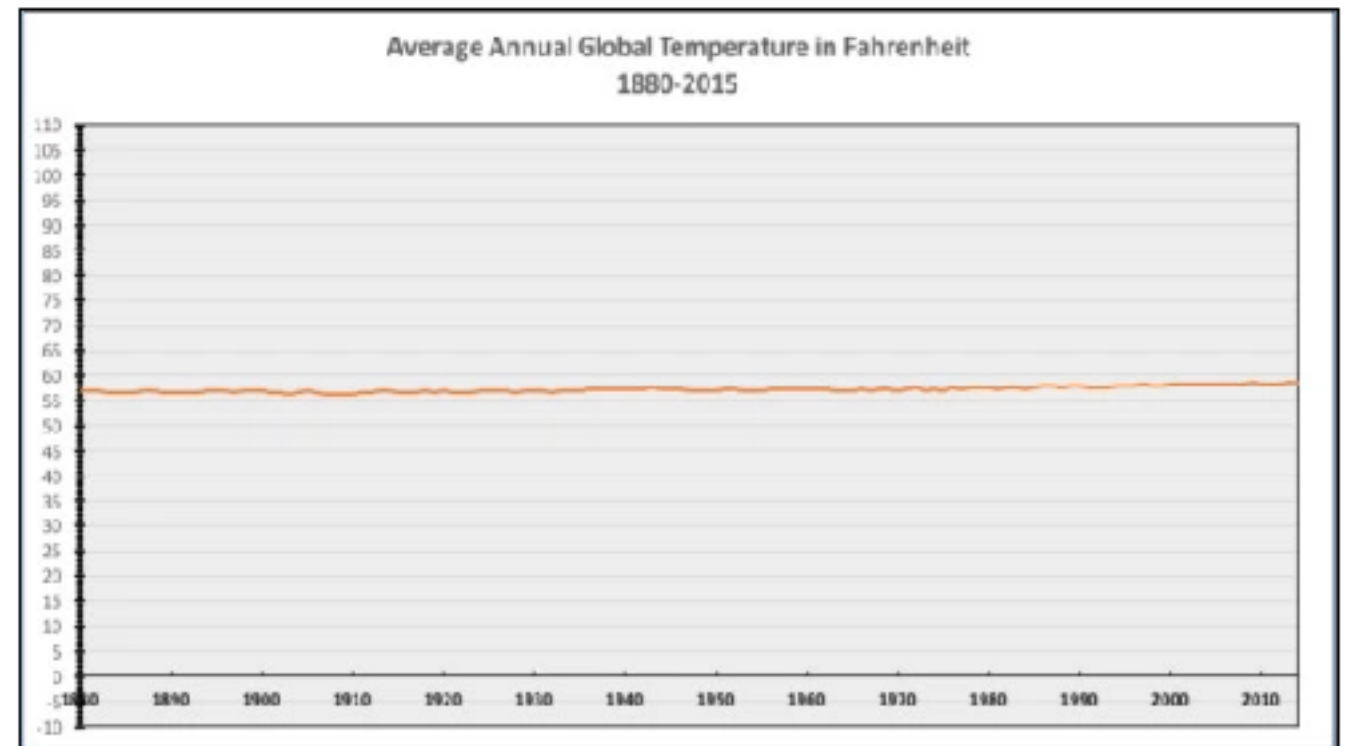


Question

Which of these graphics is misleading?

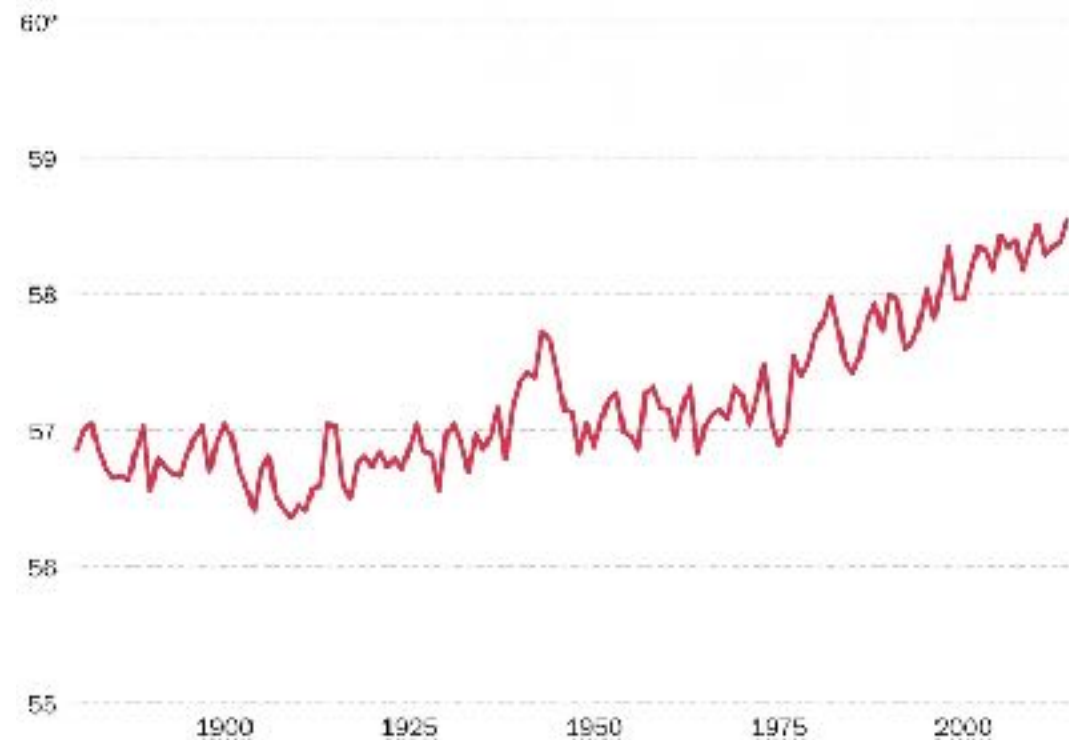
Connect your answer back to visual cues.

Instructions: ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you



Average global temperature by year

Data from NASA/GISS.



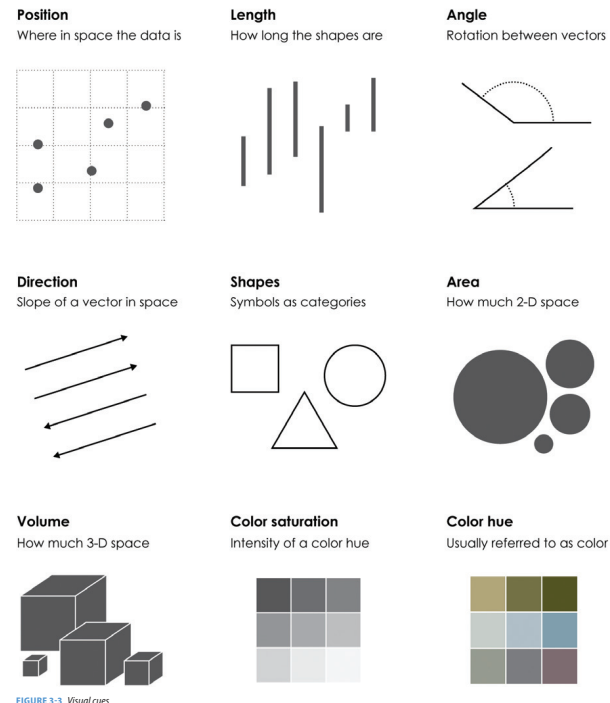
Question

One of today's readings said that this graphic was *misleading*.

Why is it misleading? Hint: compare 45-54 age group to the 65+ age group.

Can you explain how it misleads in terms of visual cues, coordinate systems, scales?

Instructions: ~1 minute to think/answer on your own; then discuss with neighbors; then I will call on one of you

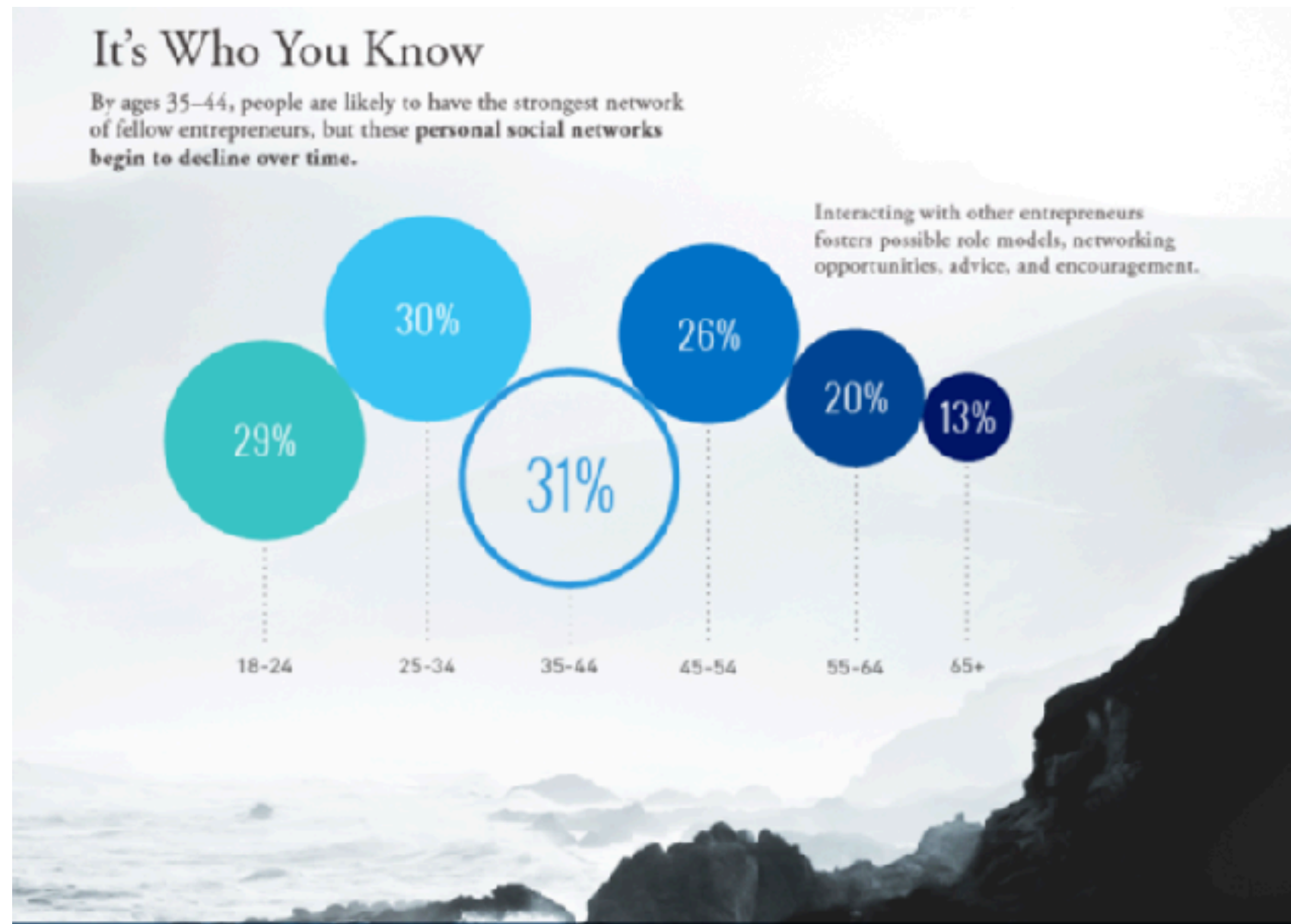


Coordinate systems:

- Cartesian (x,y)
- Polar
- Geographic

Scales:

- linear
- log
- categorical
- ordinal
- percent
- time

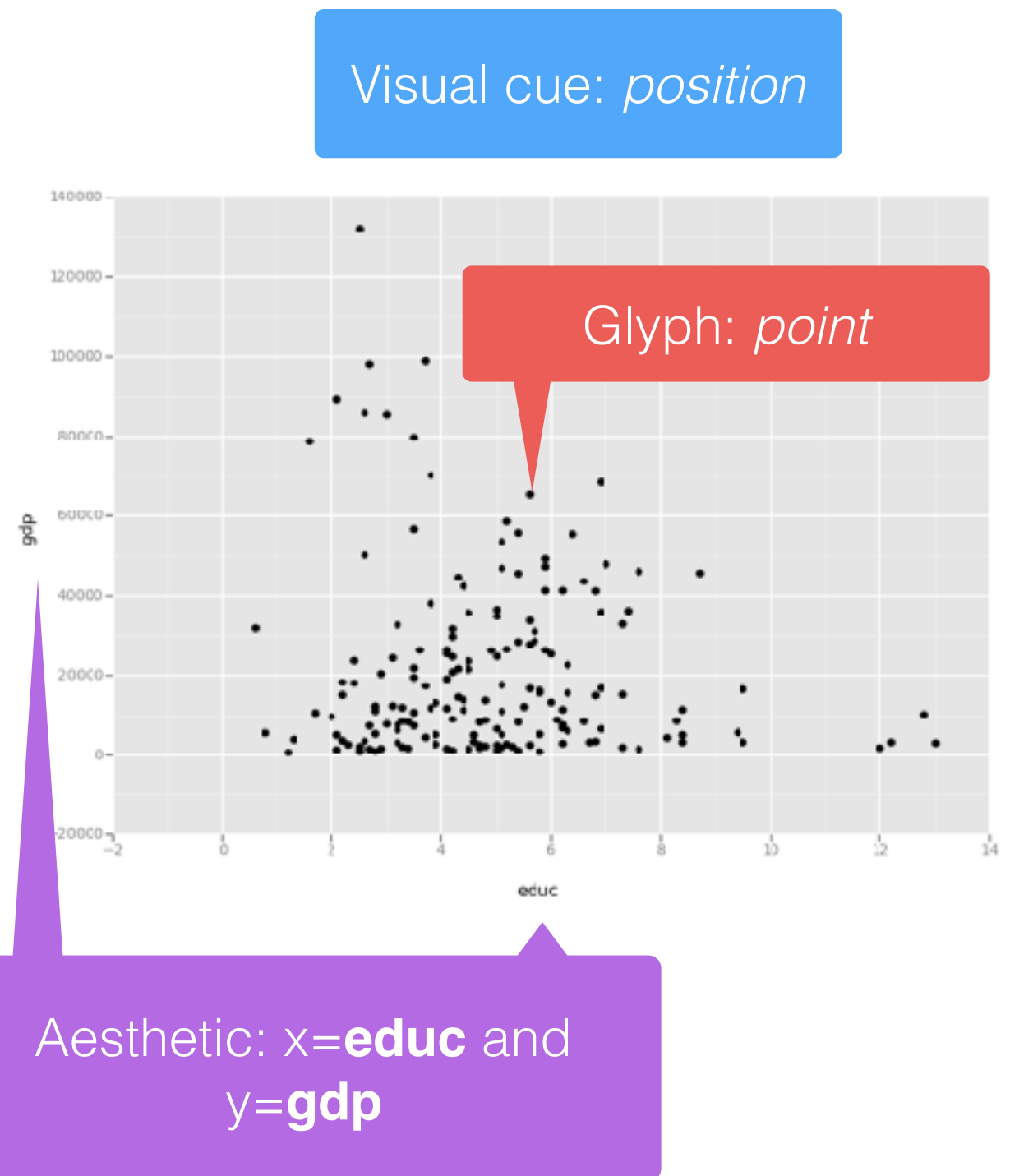


Today

- Conceptual: grammar of graphics
- Practical advice

Terminology

- Glyph (aka mark, symbol, geometric object, geom): basic graphical element
- Visual cues: position, length, angle, etc.
- Aesthetic: explicit mapping between a variable and visual cue



Tidy Data

- Rows: *observations*, each refers to specific, unique, and similar sorts of things
 - Each row will become a glyph in viz
- Columns: *attributes/variables*, all values in a column must be of the same type and represent a descriptive characteristic of an observation.
 - Some (but not necessarily all) of the columns will be mapped to aesthetics
 - Types: numerical, categorical, strings, etc.

country	oil_prod	gdp	educ	roadways	net_users
Afghanistan	0.0	1900.0	NaN	0.064624	>5%
Albania	20510.0	11900.0	3.3	0.626131	>35%
Algeria	1420000.0	14500.0	4.3	0.047719	>15%
American	0.0	13000.0	NaN	1.211055	NaN

Data manipulation

- Data manipulation comes up *a lot* with visualization.
- You often need to munge your data into a form that's amenable to viz.
- This may include doing statistical transformations (e.g., histograms)
- Viz software often provides support for some transformations

Key ideas

- **Glyphs**: geoms
- **Aesthetics**: mapping data to visual cues
- **Scales**: determines mapping from data to aesthetic property.
Example: using a log scale for y axis changes how data is mapped to a y coordinate.
- **Coordinate system**: cartesian, polar, geographic
- **Faceting**: idea of "small multiples"
- **Layers**: composing multiple visualization components into single graphic

ggplot

- Walk thru ggplot example in python notebook.
- <http://ggplot.yhathq.com/>
- Note: ggplot is a library developed in R. The R version is more full-featured than the python port.
- You can use R's version inside a python notebook by installing R and then getting the rpy module to "talk" to R from python.

Tools

- With any tool, there is a learning curve
- You have to decide how much time you want to spend on visualization

Apps

- Google spreadsheets
 - Choose from among fairly limited set of visualizations
 - Learning curve: very low
- Tableau
 - Powerful tool: many chart types, works "automagically", designed for non-technical audience
 - Available on Mac desktops in lab, can also get your own student license
 - Learning curve: need to understand Tableau's terminology (dimensions, measures, attributes), online tutorials are quite good though
- plot.ly
 - Freemium online service
 - Also has python hooks

Code (within notebook)

- matplotlib
 - lots of built-in chart types, easy-to-use for basic stuff, DSFS code provides many examples
 - lots of documentation
- ggplot <http://ggplot.yhathq.com/>
 - a *limited* adaptation of R's ggplot2 library (only cartesian coordinates, missing some statistical transformations)
 - based on *grammar of graphics*
 - once you understand grammar, it's fairly easy to manipulate graphics
- others: plot.ly, bokeh, ...

Code (outside notebook)

- D3 (<https://d3js.org/>)
 - javascript library, very elegant and powerful
 - *low, low level*; very high learning curve
- Vega (<https://vega.github.io/>)
 - *declarative* language ("what" not "how")
 - based on grammar, incredibly elegant
 - low level: for example, charts don't have titles (you place a text "mark")
 - can design your own visualizations by combining marks, scales, data transformations, etc.
 - learning curve: high
- Vega-lite (<https://vega.github.io/>)
 - Higher level language for standard visualizations (bar chart, scatter plot, etc.)

Recommendations

- If you want to...
 - ... keep it simple: matplotlib with standard chart types (or use google spreadsheets)
 - ... create *many and/or complex* visualizations and avoid coding: Tableau
 - ... learn about a visualization grammar: ggplot and vega
 - ... design your own visualizations: vega
 - ... pad your resume: D3