Data stories: Expanding your graphical repertoire
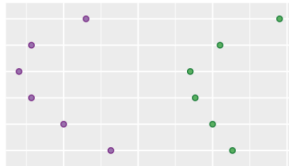
Richard Layton
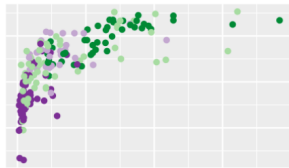
Session 2, 2022–02–21

Expanding your graphical repertoire: Four main topics[1]
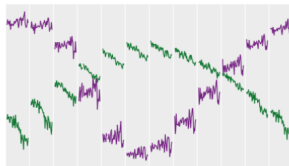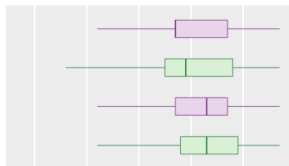
Comparing data

Revealing correlations

Showing evolution
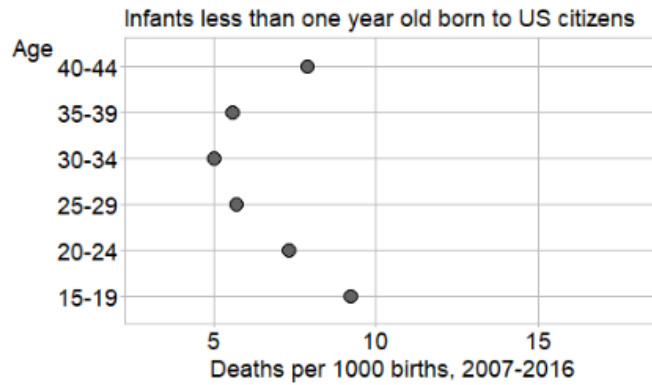
Displaying distributions

I suggest you have a printed copy of these worksheets to write in during the workshop. We have a number of think-write-share activities that for many people work best when thoughts are written down.

Notes

§ Comparing data

Dot plot

Infant mortality in the US by maternal age group.[2]

Infants less than one year old born to US citizens

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: (1) quamtity, (1) category

Argument: Comparison

- What questions does the chart raise?

Multiway dot chart, superposed

Infant mortality in the US by maternal age group, maternal race, and geographical region.[3]

Infants less than one year old born to US citizens

- Describe the main idea(s) this chart conveys to you.

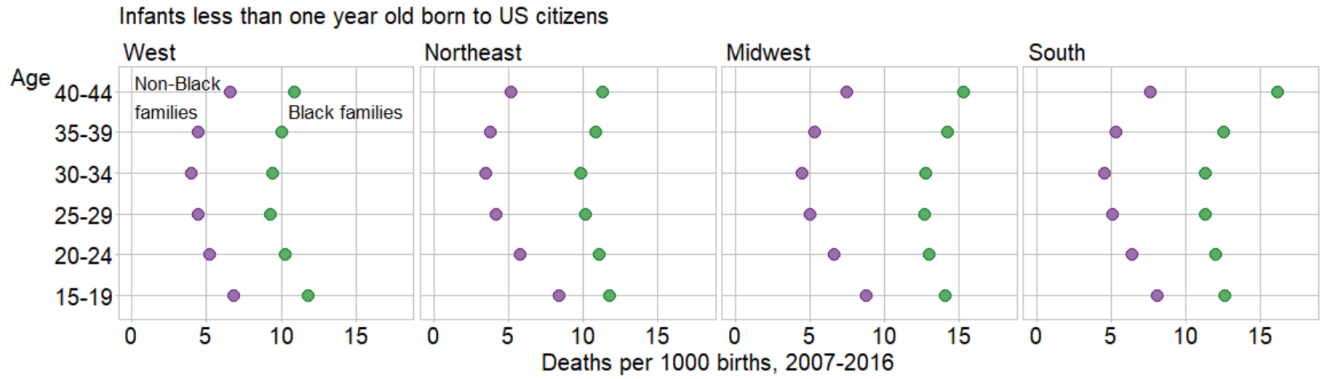*Variables + Argument → Design*

Variables: (1) quamtity, (3) categories

Argument: Comparison

- What questions does the chart raise?

§ Revealing correlations

Scatterplot

GDP[4] and life expectancy at birth[5] by country.
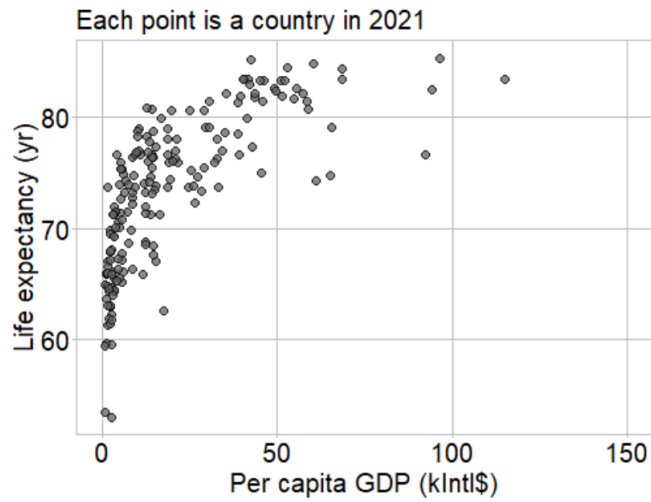
Each point is a country in 2021

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: (2) quamtities

Argument: Correlation

- What questions does the chart raise?

Scatterplot, faceted

GDP[6] and life expectancy at birth[7] by region and country.

[6] Data source: Gapminder [2022a]

[7] Data source: Gapminder [2022b]

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: (2) quamtities, (1) category

Argument: Correlation & comparison

- What questions does the chart raise?

§ *Showing evolution*

Time series

Life expectancy at birth[8] in the US.
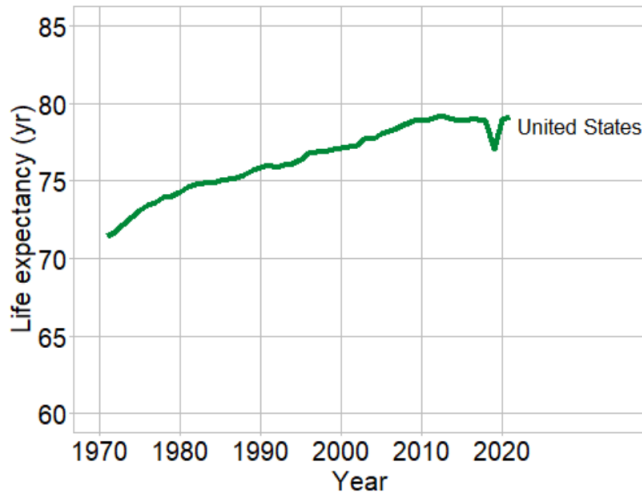
- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: Discrete time, (1) quantity

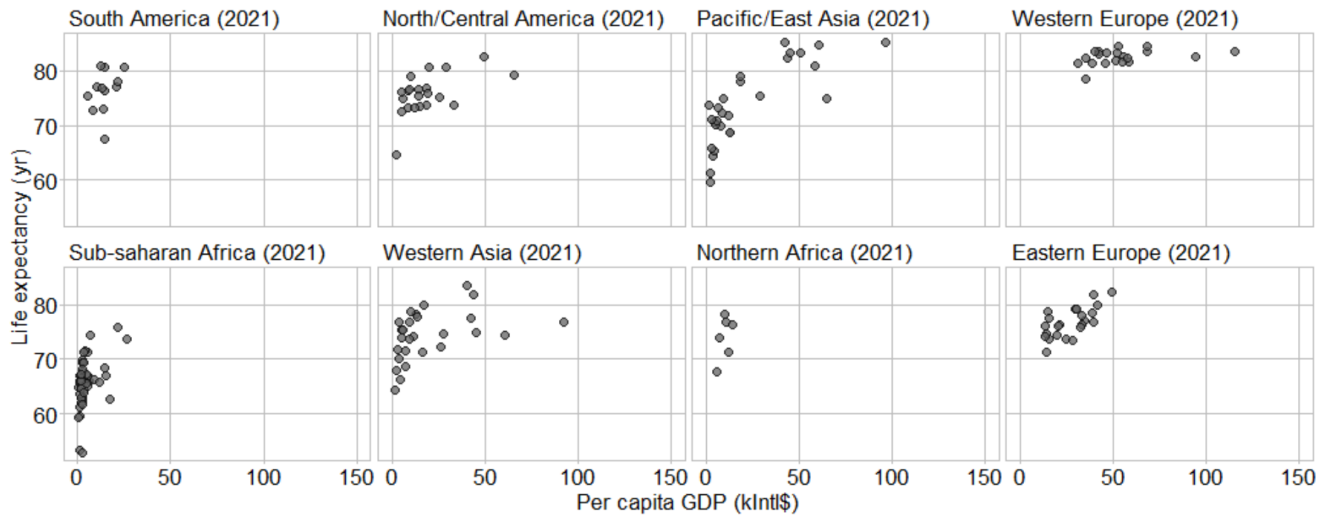*Argument*: Evolution

- What questions does the chart raise?

Time series, faceted

Life expectancy at birth[9] in selected countries

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: Discrete time, (1) quantity, (1) category

Argument: Evolution & comparison

- What questions does the chart raise?

Connected scatterplot, faceted

GDP[10] and life expectancy at birth[11] over time in selected countries.
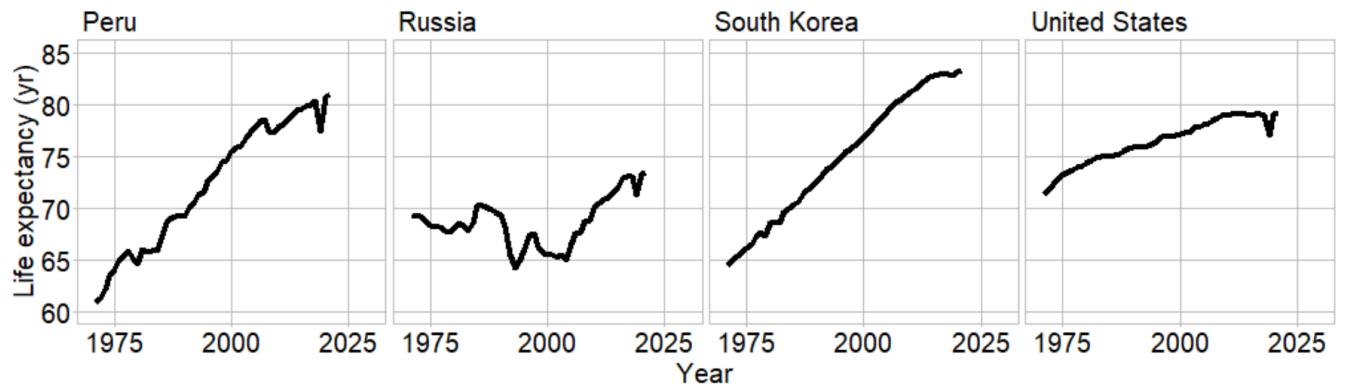
[10] Data source: Gapminder [2022a]

[11] Data source: Gapminder [2022b]

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: Discrete time, (2) quantities, (1) category

Argument: Evolution, correlation, & comparison

- What questions does the chart raise?

Cyclic time series, superposed

Extent of polar ice by month over a span of 43 years.[12]
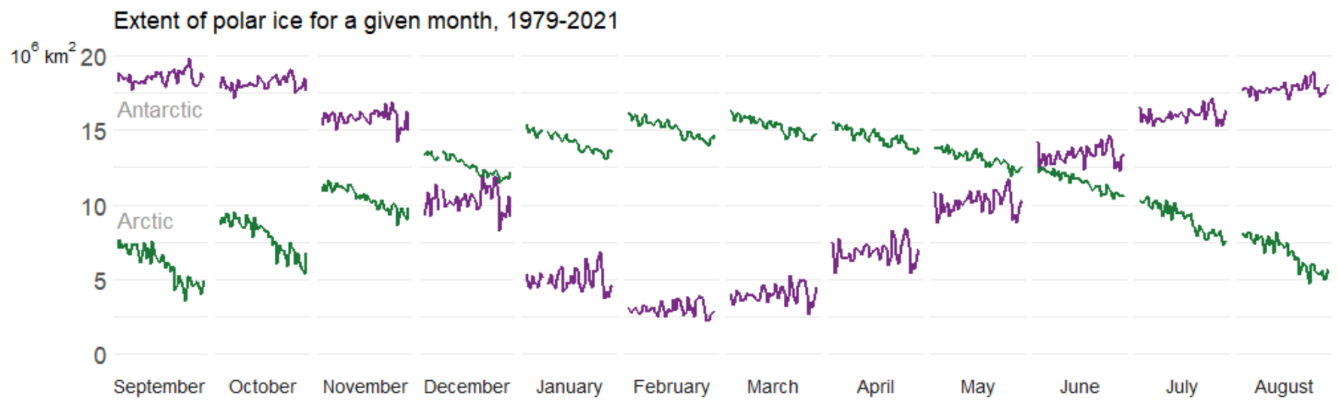
Extent of polar ice for a given month, 1979-2021

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: Discrete time,
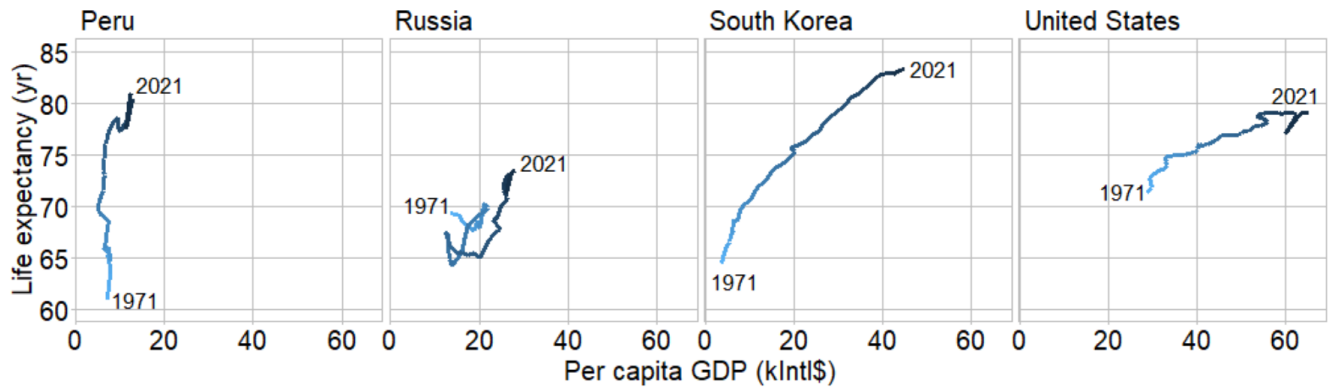(1) quantity, (2) categories

Argument: Evolution & comparison

- What questions does the chart raise?

§ Displaying distributions

Strip chart, superposed

Speeds by event and sex of skiers at the world speed competition in Verbier, Switzerland, April 2011.[13]

[13] Data source: Unwin [2015]

Each marker is one skier

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: (1) quantity, (2) categories

Argument: Distribution & comparison

- What questions does the chart raise?

Box and whisker chart

Years to graduate for US undergraduate students[14,15] by sex and path (traditional or non-traditional experience).

Graduates of 4-year US universities (N = 269,057)

- Describe the main idea(s) this chart conveys to you.

Variables + Argument → Design

Variables: (1) quantity, (2) categories
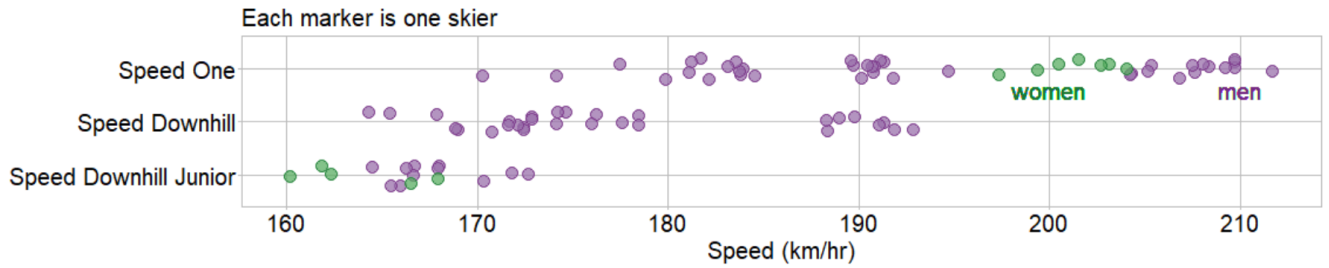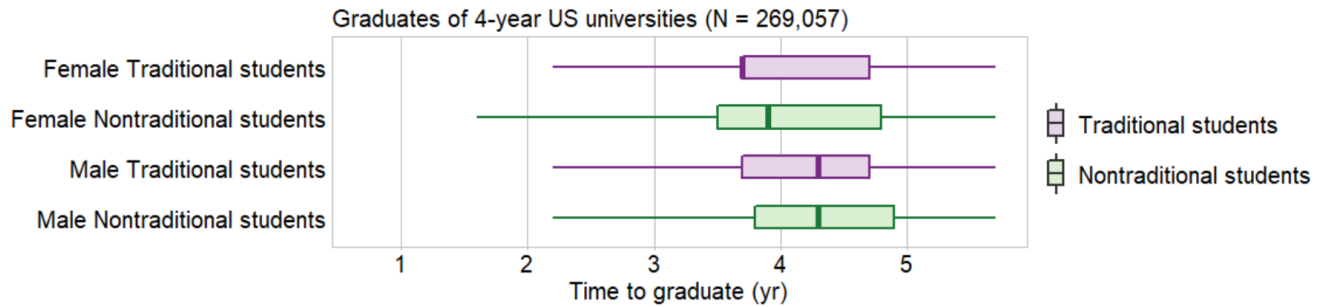
Argument: Distribution & comparison

- What questions does the chart raise?

Ideas to consider

Chart selection

- What are your variables, by name?
- Is a variable quantitative or categorical?
- Is a categorical variable naturally ordered (ordinal) or not (nominal)?
- Starting with a small number of variables, what chart types match the data structure?
- How does the chart type change as you add new variables?

Chart aesthetics

- Superposed designs work best with small numbers of subsets or when the subsets visually cluster.
- Faceted designs permit a greater number of subsets to be compared.
- Deliberately assign the size, shape, and color of every visual element.
- Use color deliberately. Choose colors that are safe for color-vision-deficient viewers.

Audience and message

- What is your story?
- Does the visual evidence directly support your verbal argument?
- Have you placed the story in context visually?
- Who is your audience?
- Will the audience resist your conventions?
- If so, is overcoming audience resistance worth the effort?

Ethics of visual rhetoric

- Is your design equitable and inclusive?
- Are you seeing only what you want to see? What the audience wants to see?
- All there alternative explanations for what the chart shows?
- Are your data dubious? Insufficient?
- Have you concealed information? Concealed a large uncertainty?
- Does your chart suggest misleading patterns?

References

CDC Wonder. Linked Birth/Infant Death Records, 2022-01. URL `https://wonder.cdc.gov/lbd-current.html`.

Jean-luc Doumont. *Trees, Maps, and Theorems*. Principiae, Belgium, 2009.

F. Fetterer, K. Knowles, W.N. Meier, M. Savoie, and A.K. Windnagel. Sea ice index, version 3, Sea ice extent and area organized by year. 2017. DOI: https://doi.org/10.7265/N5K072F8. URL `https://nsidc.org/arcticseaicenews/sea-ice-tools/`.

Gapminder. Gross domestic product per person adjusted for differences in purchasing power, 2022a. URL `https://www.gapminder.org/data/`.

Gapminder. Life expectancy at birth, 2022b. URL `https://www.gapminder.org/data/`.

Richard Layton. *graphclassmate: Companion materials for a course in data visualization*, 2021. URL `https://github.com/graphdr/graphclassmate`. R package version 0.1.0.

Matthew Ohland and Russell Long. Multiple-Institution Database for Investigating Engineering Longitudinal Development, 2021. URL `https://midfield.online/`.

Antony Unwin. *GDAdata: Datasets for the book Graphical Data Analysis with R*, 2015. URL `https://CRAN.R-project.org/package=GDAdata`. R package version 0.93.