[redacted]



# Chapter 2: Many ways to Show: a Case Study.

## Topics

- Many ways to visualise a dataset
- Examples of bad visualisation and Recommendations
- TED talk "The Beauty of Data Visualization"

## Many Ways to Show: a Case Study

- Suppose that you are working high up in a glass tower for some multi-national organisation.
- You are asked to prepare a report describing population trends in the world since 1960.
  - How do you start?

## Dataset

- Download some data from the World Bank.
- For each country:
  - country name (e.g. India)
  - country code (e.g. IND)
  - region (e.g. South Asia)
  - population in 1960 (pop1960)
  - population in 2014 (pop2014)

**How to explore & present this information?**

## Data → Questions

- Are there obvious errors in the data?
- Which countries have largest/smallest populations?
- Populations of which countries have grown the most/least? Why?
- Do different regions show different patterns?
- What are meaningful groups of countries with similar population growth?

## Scatterplot

- Your exploratory scatterplot:
  log pop 1960 Vs log pop 2014
- Shows 422 numbers, the region of each country is displayed as a color. (could use shape as well...)

- The reference comparison might be no change in population,
  - which is a diagonal line of slope 1.

Can we replot so reference comparison is horizontal?

- Most of the space is taken up showing that different countries have different populations:
  can we use more space to indicate growth?

## Population Growth Rate

- growth = pop 2014 / pop 1960

- Yes, there is a country with a population 100 times larger in 2014 than 1960!
- Check:
  - Is this accurate?
  - which country?
- We need to plot growth on a log scale...

## Population Growth Rate - log scale

- Better.
- Reference comparison is horizontal: groups of countries stand out better.
- Horizontal scale is pop1960:
  - but there is no particular reason to compare population size to growth rate.
- Horizontal scale is just spreading the points out.
- we want to see country names

## Selected Countries

- We have selected countries with a population over 10,000,00 in 2014.
- Allows us to see trends in large countries more clearly.
- We could show unusual smaller countries in a separate plot...
- About 80 countries shown - few enough to show names.

## Country Names

- This is a useful exploratory plot - but not so good for presentation because the codes are obscure, and text overlaps. (unclear).

Vertical scale is a log scale: the **gradient** of a line represents the growth in population.

- Height at left is population in 1960;
- height at right is population in 2014.

- Same information as on the scatterplots
  - but now each point has become a line, and the lines cross!

- Too much ink!
  - we are using 10 times the ink for each country, and
  - the graph is more confusing.
- Sometimes you find an interactive d3 application like this, where you click on a line to reveal more information.
- Human ability to read gradient is very bad.
  - Don't use gradient to represent important information unless it is unavoidable.

## Bar chart.

- Dots are usually better than bars – a bar has many times more ink!
- Having 2 bars for each country would be a mess, but 2 dots were fine.
- The countries are ordered alphabetically:
  – this makes the graph unreadable.
- For a graph, order a categorical variable in a visually meaningful way.
- (Alphabetic ordering can be useful for data in tables,
  – if readers actually need to look up numeric values.
- In graphs, you want to pick out information visually, and then find the labels.

- Better.
  – Now we can see patterns of growth by region.
  – one data error stands out.

## Cleveland Dot Plot

– Names are connected to dots, with minimum ink.
– Ordered by population in 2014; population in 1960 shown as red dot.
– Gap between the dots is the log growth rate.
– Hard to see any patterns – colour is used for year & not for region.

## Facetted cleveland Dot Plot

– Countries in different regions shown separately, but on the same horizontal scale.
– Now some patterns emerge...
– We are still trying to plot population size & growth on the same graph.
– Maybe we should plot them on different graphs?

## Cleveland Dot Plot of Growth Rate, Facetted by Region.

- Allow some patterns are obvious, & surprises & questions emerge.
  - Is the region "Europe & Central Asia" reasonable?
  - Use a linear growth scale for emphasis?

## Example of bad visualisation & Recommendations

### Dots or Bars?

- Each bar is one number - look how much ink!
- Include 0 on the scale or not? Depends on your audience.
- A bar looks like a solid mass, so IMO a bar should start at 0, but a dot doesn't have to.

- Excessive ink!!
- Each curve connects a year to a number of earthquakes...
- Use a plot!!!

- Each curve links 2 countries with one number.
- Curves hard to follow.
- Few comparisons can be made.
- Curves could be replaced by dots.
- Selection & Summary needed.

### Recommendations

- Use full range of space available; rescale if necessary.
- Use minimum ink.
  - Prefer dots to bars, or lines that can be made into a dot.
- Select, simplify, & summarise.
- Figure out what comparisons you want to show;
  - choose graphs that can be read accurately
  - for these comparisons & information.

- (Usually) don't show a categorical variable
  - (e.g. country name) in alphabetical or a meaningless order:
  - order it meaningfully!
- More than one plot may be needed:
  - combining population size & population growth on one chart was not clear.
- Use facetting
  - → to show comparisons between subsets of data.

Remarks: graphics & text

When writing a report:

1) First gather & construct your hard, precise information
   - graphics
   - tables
   - technical definitions & equations
2) After that write your text around the hard information
   - (before & after) commenting on it to make your argument.

- Present hard info in a clear & precise way
  - and the make interesting remarks about it.
- The reader can refer to your clear, well designed graphics, technical definitions etc,
- Instead of wading through waffly description...
- You don't expect to write good text first time:
  - much revision & reorganisation is needed.
- Similary, a good graph is hard to produce:
  - expect at least 10 or many more attempts
  - before you find the right way to present the right information.