ibes non-data ink. Design elements!

The plotting space you are using

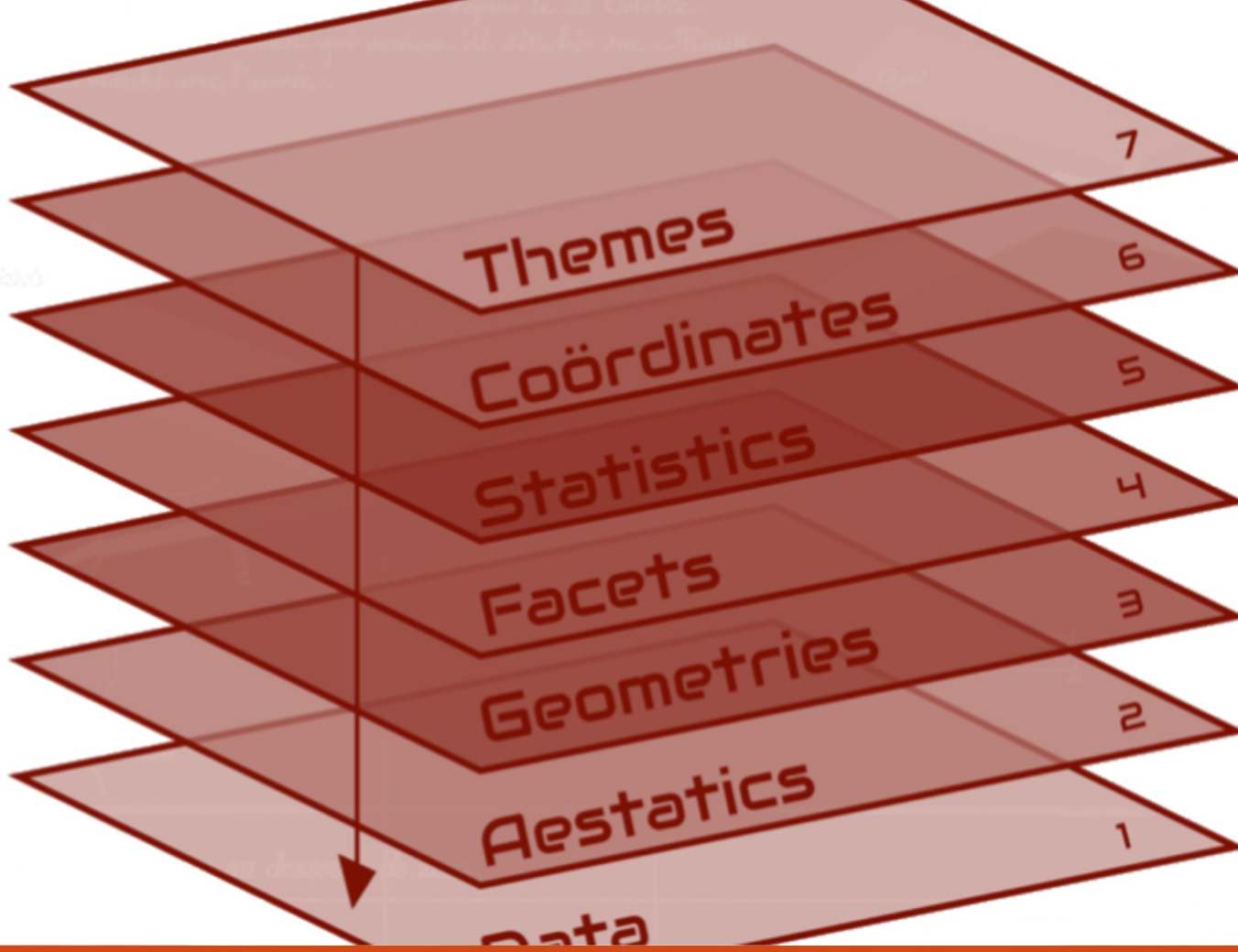
Statistical models & summaries

Rows and columns of sub-plots

hapes used to represent your data

scales on which the data is mapped

The actual variables to be plotted



10. The Grammar of Graphics

Grammar of Graphics

It is one thing to recognize when charts are **effective** and their aesthetics make them **easy to read**, and when they are laid out in a dashboard which tells a **compelling** visual story (and when they are not).

It is another thing altogether to learn how to **build** such charts.

The **grammar of graphics** [Wilkinson, 1999; Wickham, 2009] provides a reliable path to do so.

[D.J. Sarkar, A Comprehensive Guide to the Grammar of Graphics for Effective Visualization of Multi-dimensional Data]

Grammar of Graphics

"A **grammar** is defined as a set of structural rules which helps define and establish the **components of a language**.

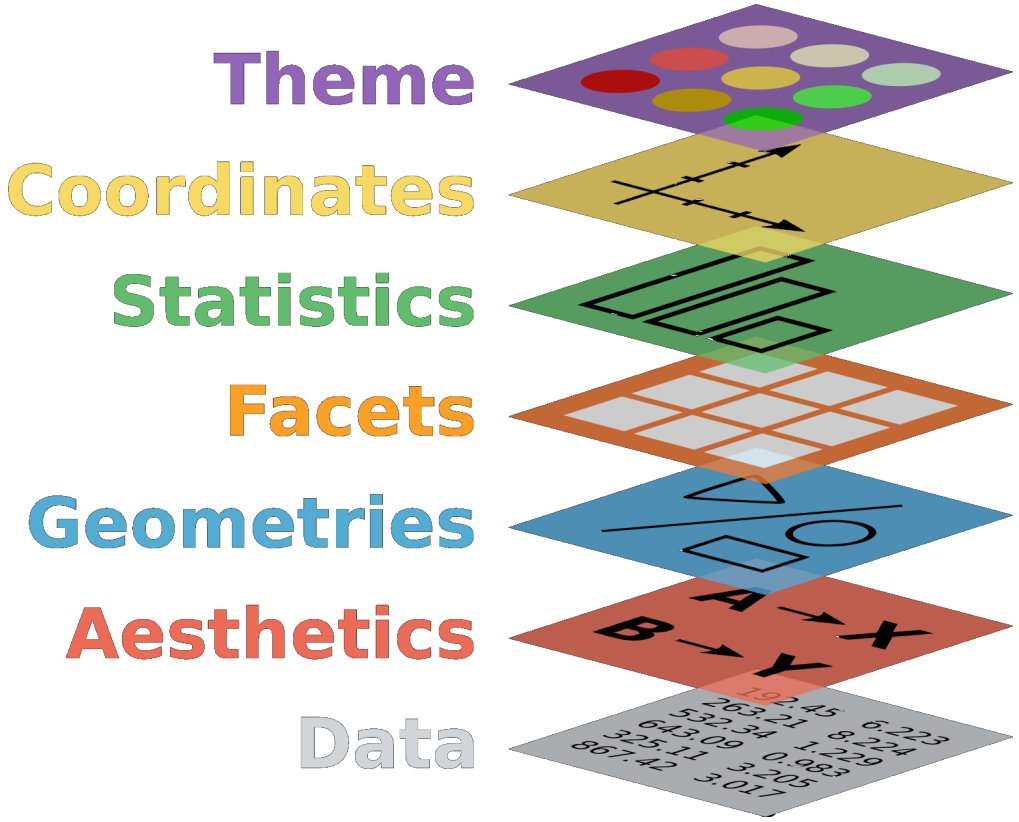
A language's system/structure usually consists of **syntax** and **semantics**.

A grammar of graphics is a framework which follows a **layered** approach to describe and construct visualizations or graphics in a **structured manner**.

The layered grammar of graphics uses pre-defined components to build charts (instead of random trials and errors)."

[Introduction to data visualisation with ggplot2, *QCBS R Workshop Series;* T. De Beus, Think About the Grammar of Graphics When Improving Your Graphs]

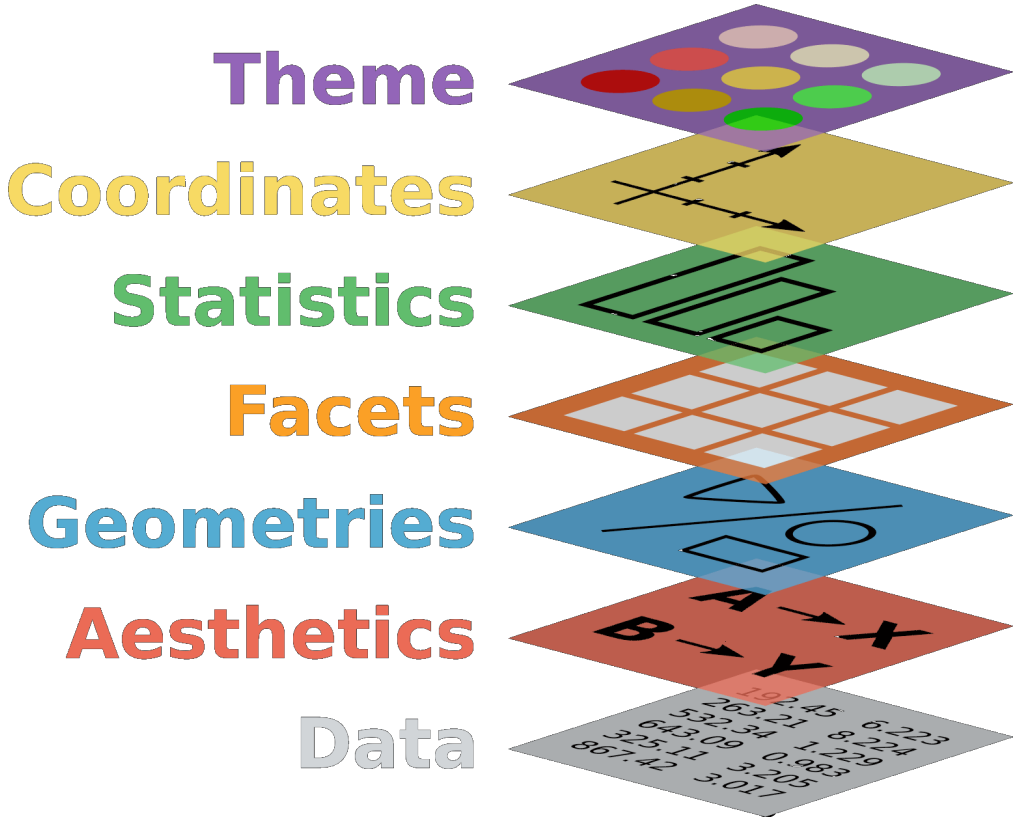
Grammar of Graphics Layers



1. **Data** (required): the plotting observations are found in rows, the variables in columns
2. **Aesthetics** (required): the mapping of the dataset's variables to the chart's scales (position, shape, size, colour, etc.)
3. **Geometry** (required): the type of chart on which the data is represented (bars, lines, points, etc.)
4. **Facets** (optional): the subsets of the data represented on the chart (levels)

[Introduction to data visualisation with ggplot2, *QCBS R Workshop Series;* T. De Beus, Think About the Grammar of Graphics When Improving Your Graphs]

Grammar of Graphics Layers



5. **Statistics** (optional): the measures that could provide context to the chart (centrality, dispersion, trend, etc.)

6. **Coordinates** (required): the chart plotting space (axes, scale, etc.)

7. **Themes** (required): the design choices that are used to create a visual identity (fonts, colours, etc.)

Examples – Gapminder Dataset

The Gapminder dataset (https://gapminder.org) contains socio-demographic information (upwards of 500 variables) for the Earth's nations, for years ranging from 1800 to 2020.

We deconstruct 8 charts built from this dataset using the layered grammar of graphics framework.

Session 4

Data: Gapminder countries, 2009

Geometry: stacked density chart

Aesthetics:

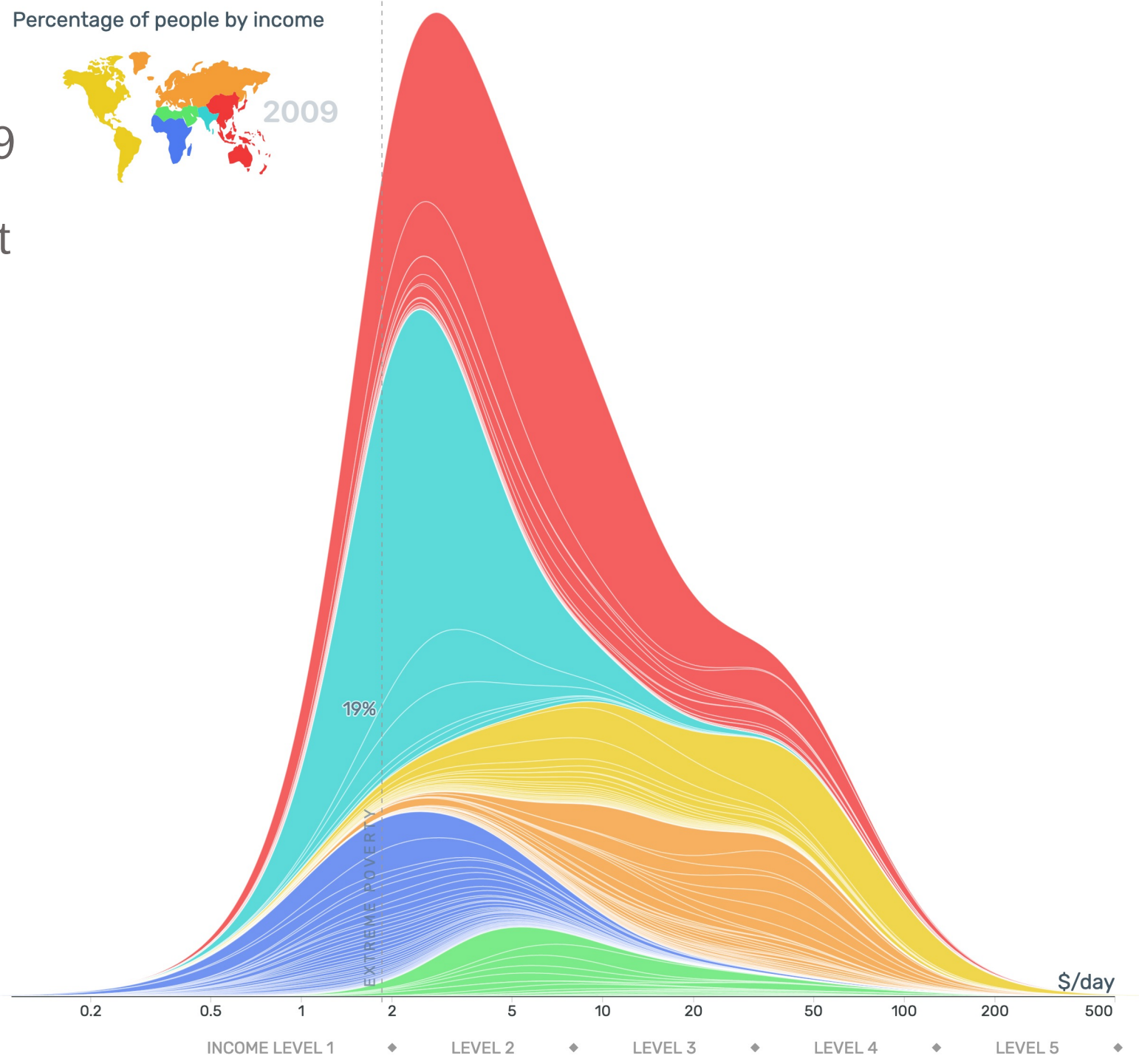
- x: daily income
- (y: percentage per country)
- fill: region

Facets: none

Statistics: extreme poverty prop

Coordinates: logarithmic (x)

Theme: Gapminder Tools; adornment (Extreme Poverty)



Data: Gapminder countries, 2009

Geometry: bubble chart

Aesthetics:

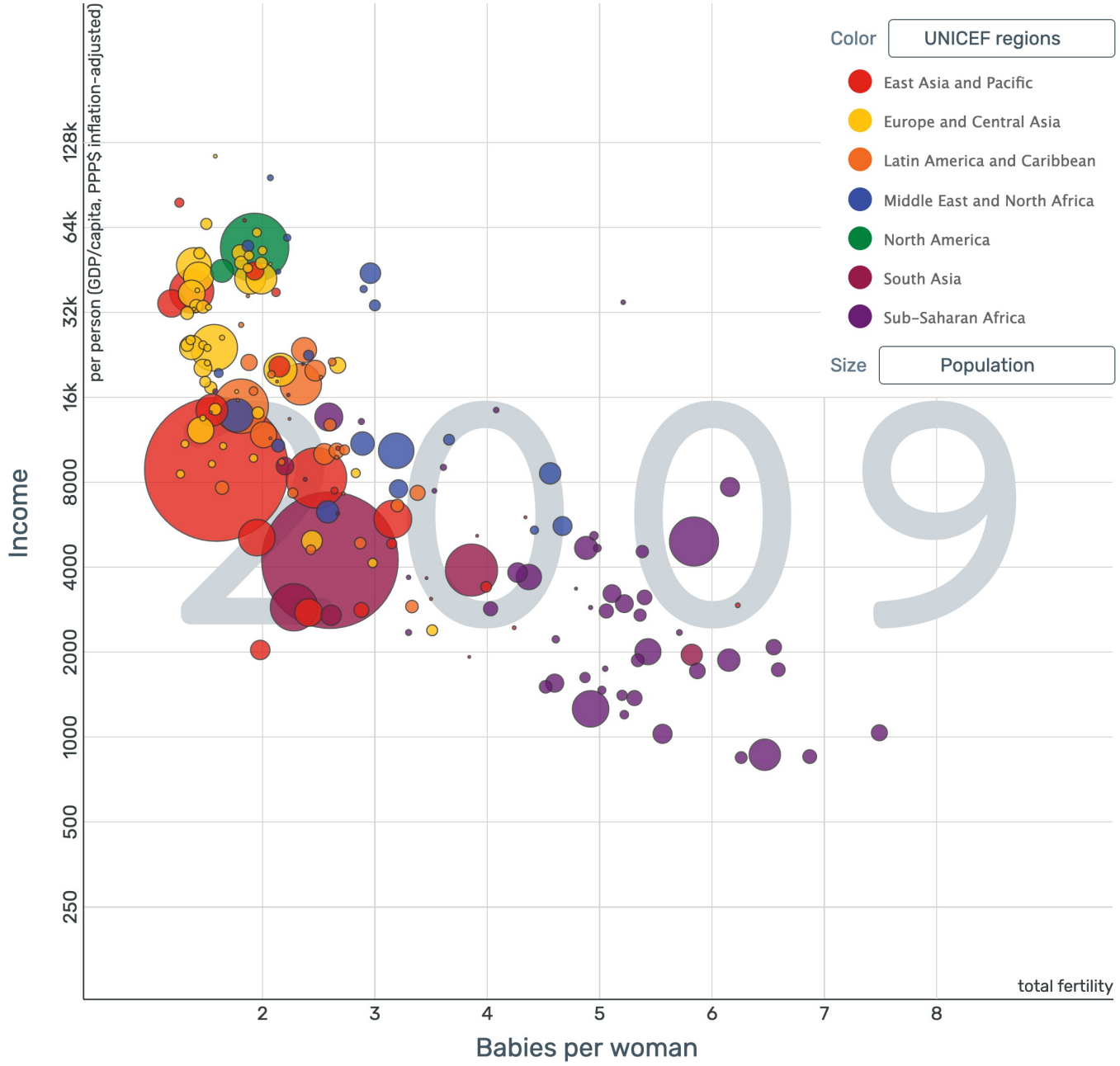
- x: total fertility
- y: income per person
- fill: UNICEF region
- size: population

Facets: none

Statistics: none

Coordinates: logarithmic (x, y, size)

Theme: Gapminder Tools



Data: Gapminder countries, 2009

Geometry: scatterplot chart

Aesthetics:

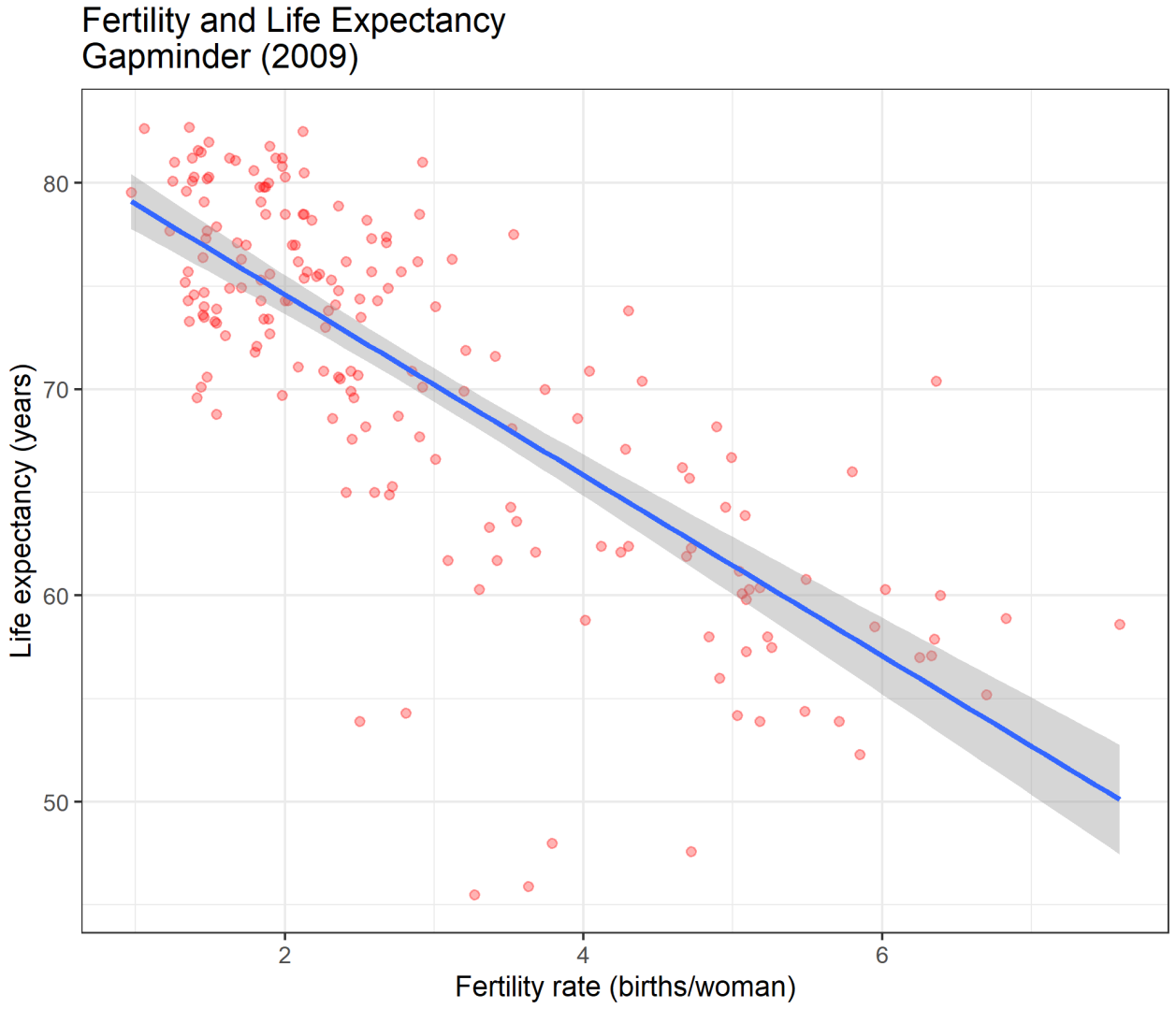
- x: total fertility
- y: life expectancy

Facets: none

Statistics: line of best fit, confidence interval

Coordinates: linear (x, y)

Theme: ggplot2 default



Data: Gapminder countries, 2009

Facets: continent

Geometry: density chart

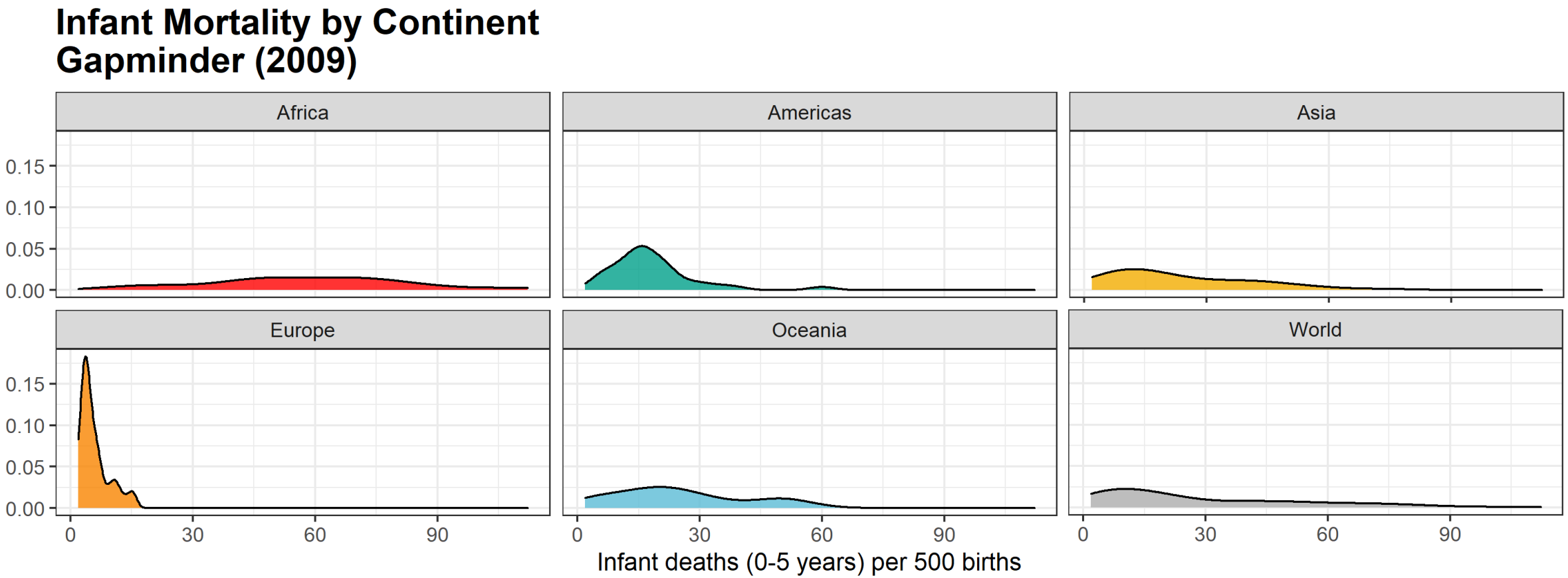
Statistics: none

Aesthetics:

- x: infant mortality
- fill: continent

Coordinates: linear (x)

Theme: Darjeeling1



Data: Gapminder countries, 2009

Geometry: bubble chart

Aesthetics:

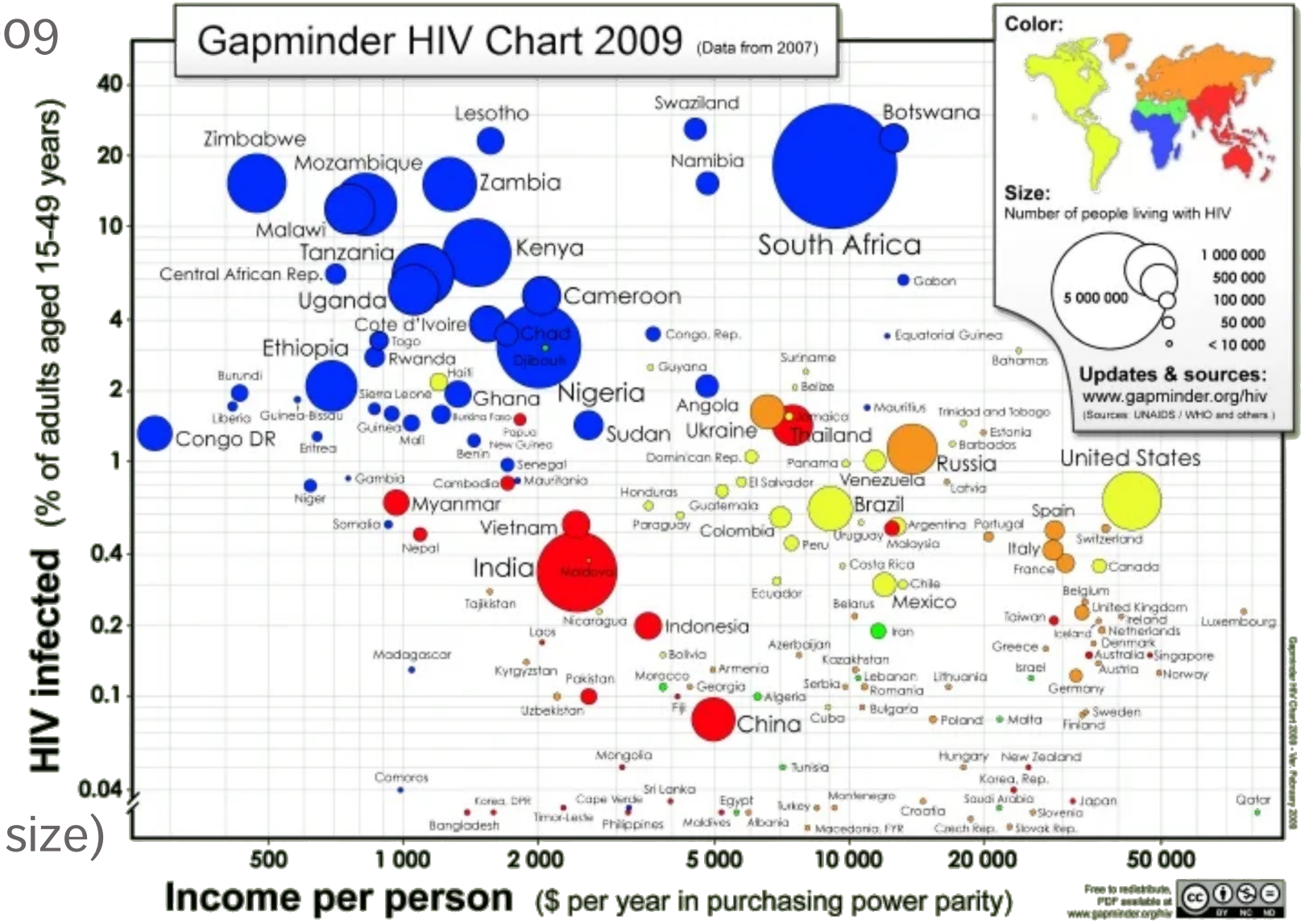
- x: income per person
- y: HIV infection rate
- fill: WHO region
- size: HIV infected population

Facets: none

Statistics: none

Coordinates: logarithmic (x, y, size)

Theme: old Gapminder World



Data: Gapminder countries, 2009

Geometry: boxplot chart

Aesthetics:

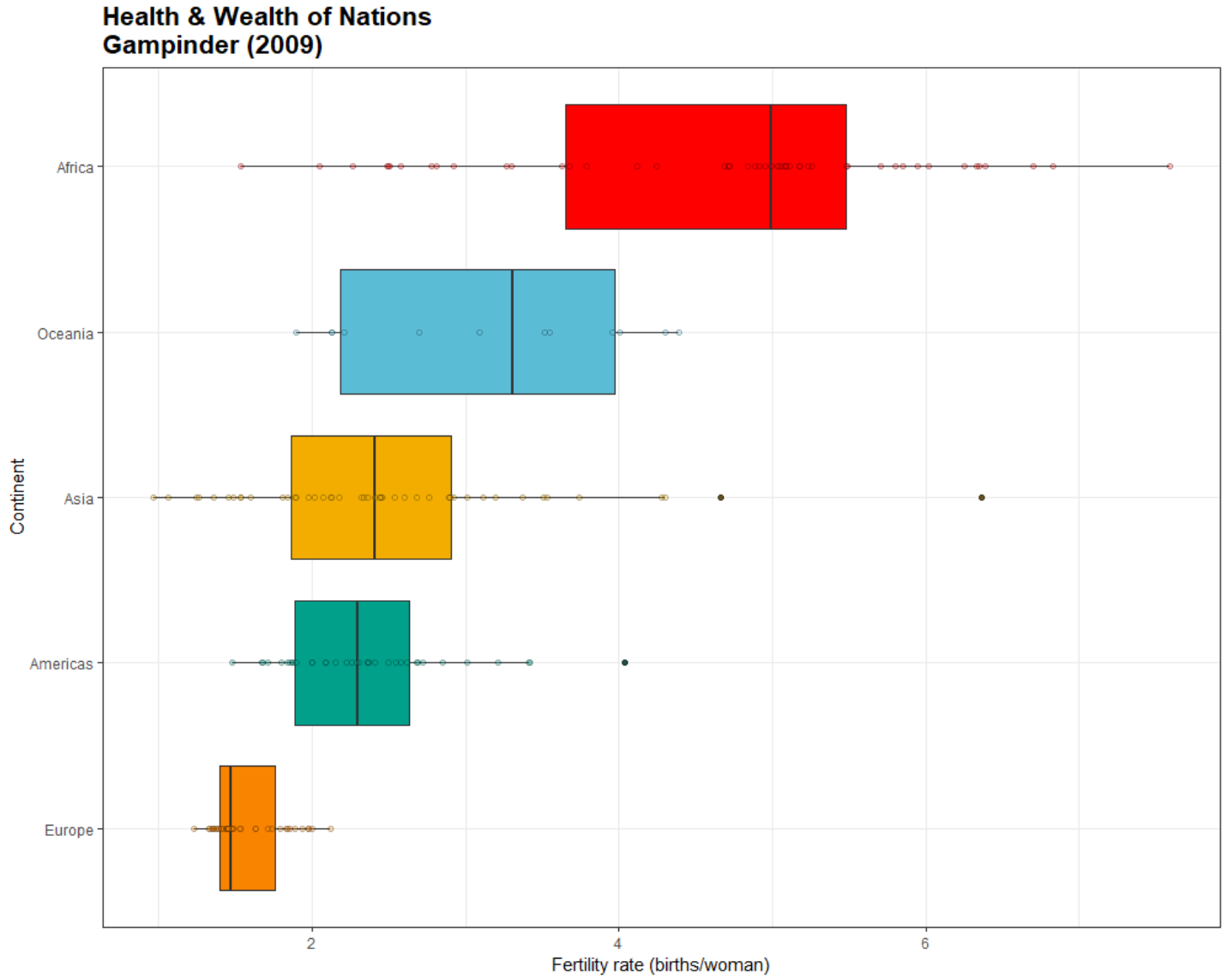
- x: fertility rate
- y: continent
- fill: continent

"Facets": continent

Statistics: 5-pt summary

Coordinates: linear (x)

Theme: Darjeeling1



Data: selected Gapminder countries, 1960-2011

Geometry: line chart

Aesthetics:

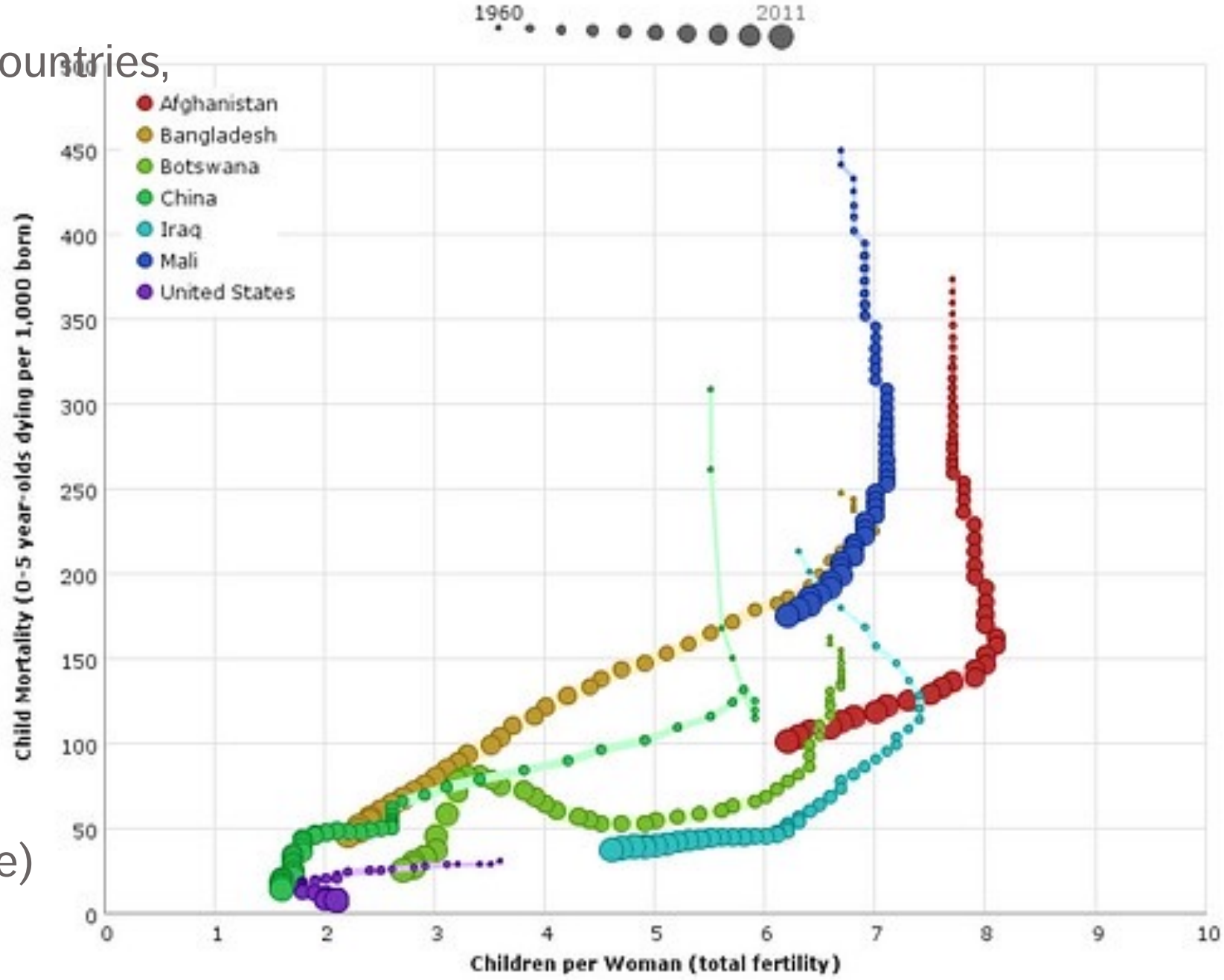
- x: total fertility
- y: percentage per country
- colour: country
- size: year

Facets: none

Statistics: none

Coordinates: linear (x, y, size)

Theme: custom



Data: Gapminder countries, 2012

Geometry: bubble chart

Aesthetics:

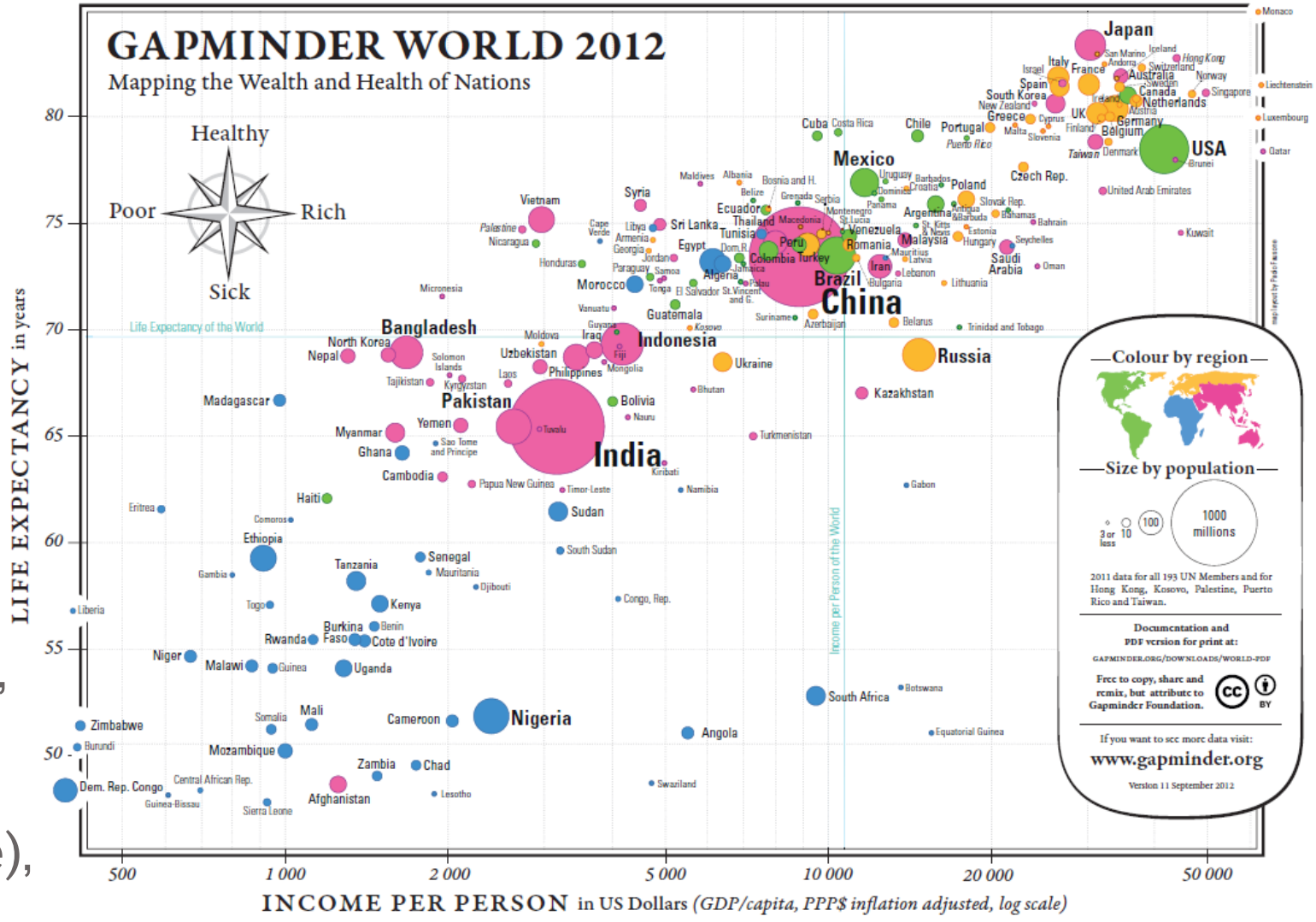
- x: income per person
- y: life expectancy
- fill: region
- size: population

Facets: none

Statistics: world life expectancy, world income per person

Coordinates: logarithmic (x, size), linear (y)

Theme: old Gapminder World



Suggested Reading

The Grammar of Graphics

The Practice of Data Visualization
Essentials of Visual Design

Visual Design

- The Grammar of Graphics

Exercises

The Grammar of Graphics

Deconstruct the charts introduced in the first 9 modules in terms of the grammar of graphics.

What do some of the most effective charts have in common? What about the least effective ones?

Does that suggest a strategy?