

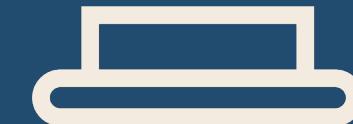
**EMMANUEL OLAMIJUWON**

*Lecturer in Health and Data Literacy*

X: eolamijuwon;

email: emmanuel@olamijuwon.com

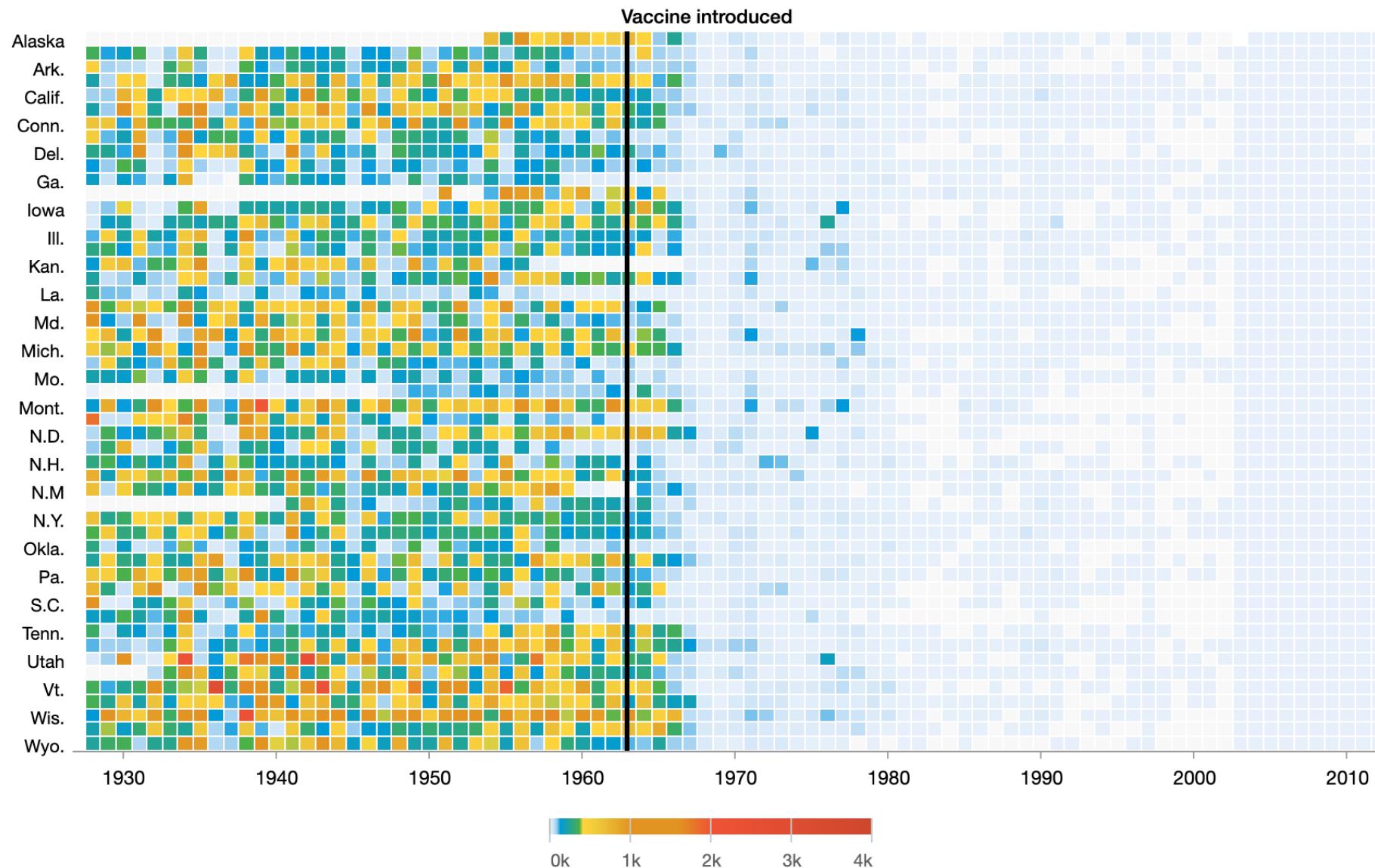
# Basic Types of Plots – ggplot2 basics



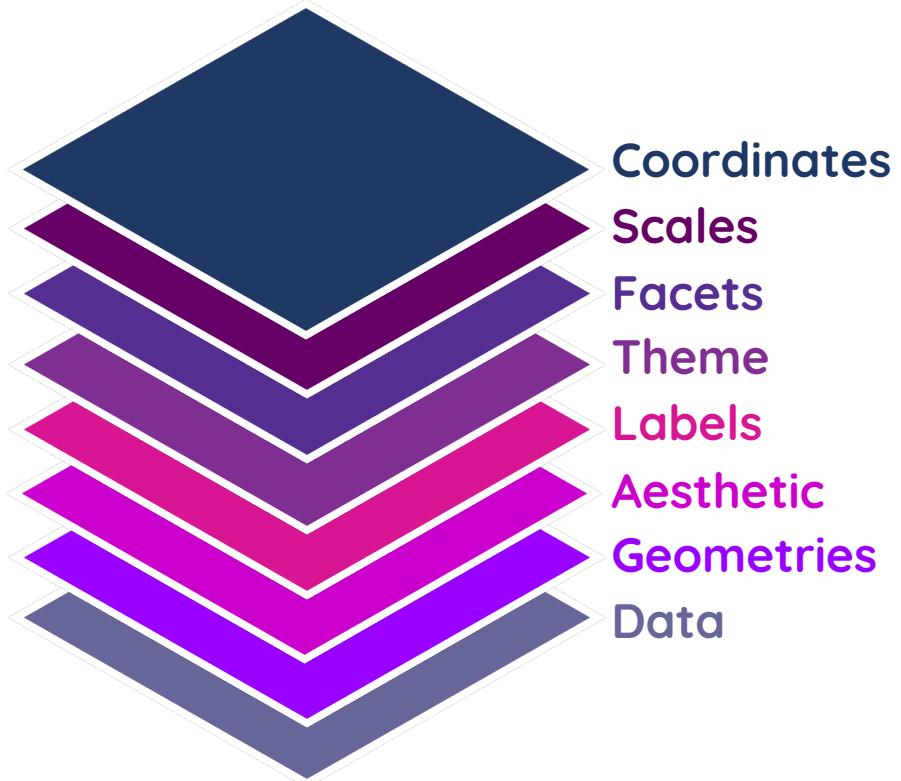
University of  
St Andrews

School of Geography and  
Sustainable Development

# The Impact of Vaccines – Measles

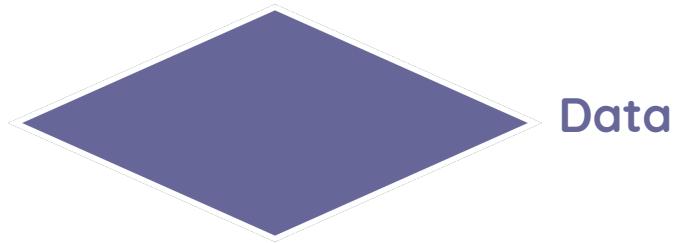


# Plotting in Layers



WorldBank: <https://www.youtube.com/watch?v=LYIvh0oATeU>

# Plotting in Layers - Data



**WorldBank:** <https://data.worldbank.org/>

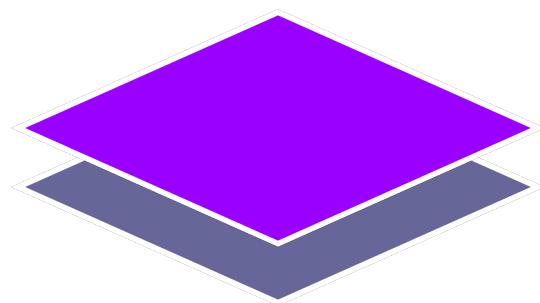
**UK Open Data:** <https://www.data.gov.uk/>

**DataFirst:** <https://www.datafirst.uct.ac.za/>



**giphy:** <https://giphy.com/explore/data-everywhere>

# Plotting in Layers - Geometries (1/2)



Geometries  
Data

## Geometric objects that represent data

Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecca

Deviation Correlation Change v Time Ranking Distribution Part to whole Magnitude Spatial Flow

### Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data

#### Examples of use

Income distribution, population (age/sex) distribution

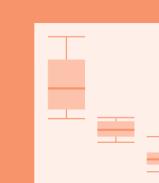
#### Chart types

histogram



The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

boxplot



Summarise multiple distributions by showing the median (centre) and range of the data

violin



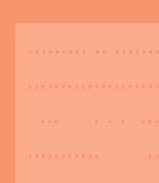
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

population-pyramids



A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms

dot-plot-strip



Good for showing individual values in a distribution, can be a problem when too many dots have the same value

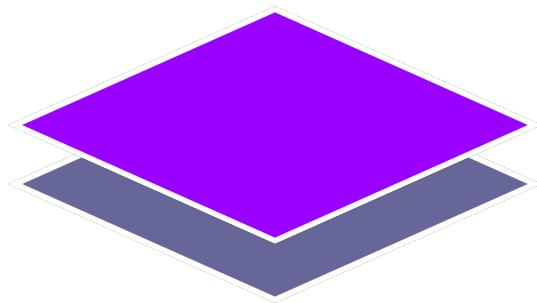
dot-plot



A simple way of showing the range (min/max) of data across multiple categories.

READ MORE: <https://ft-interactive.github.io/visual-vocabulary/>

# Plotting in Layers - Geometries (2/2)



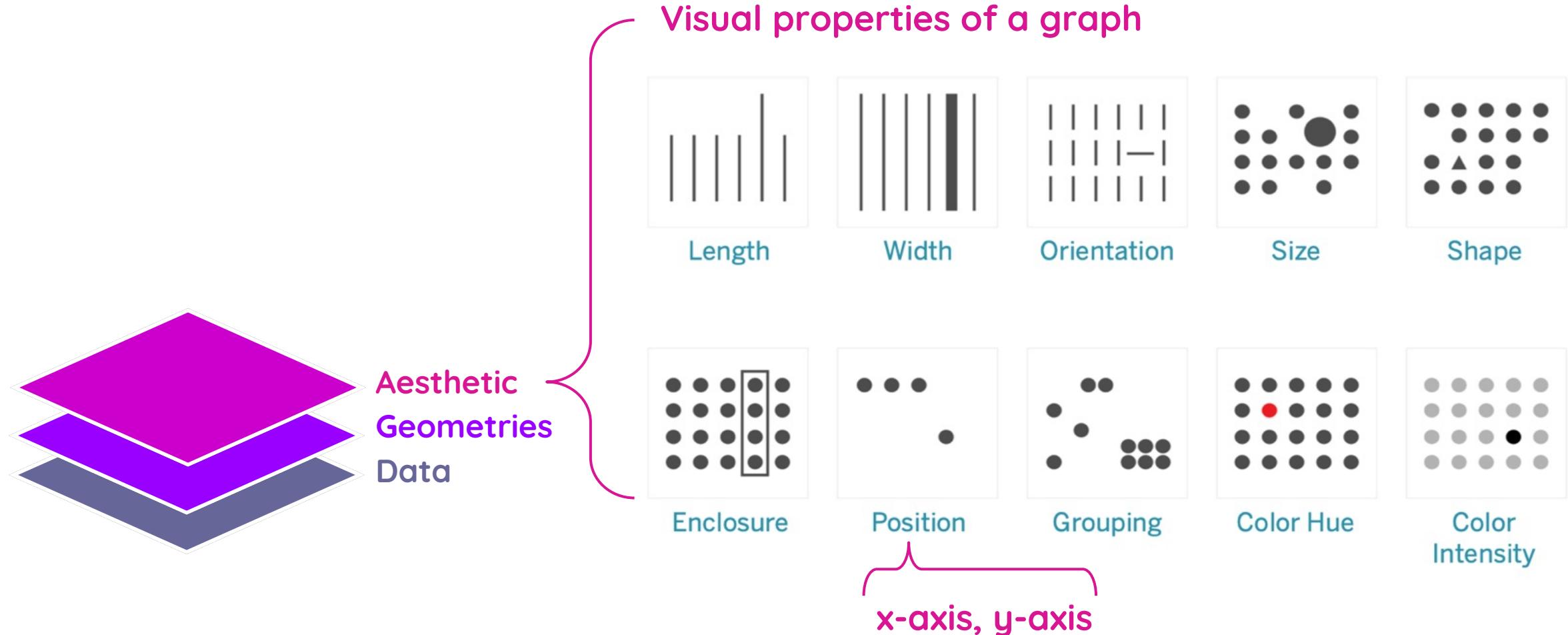
Geometries  
Data

Geometric objects that represent data

- `geom_col`
- `geom_point`
- `geom_bar`
- `geom_line`
- `geom_text`
- `geom_tile`
- `geom_vline`
- `geom_jitter`

READ MORE: <https://ggplot2.tidyverse.org/reference/index.html#geoms>

# Plotting in Layers - Aesthetics (1/2)



READ MORE: <https://ggplot2.tidyverse.org/reference/index.html#aesthetics>

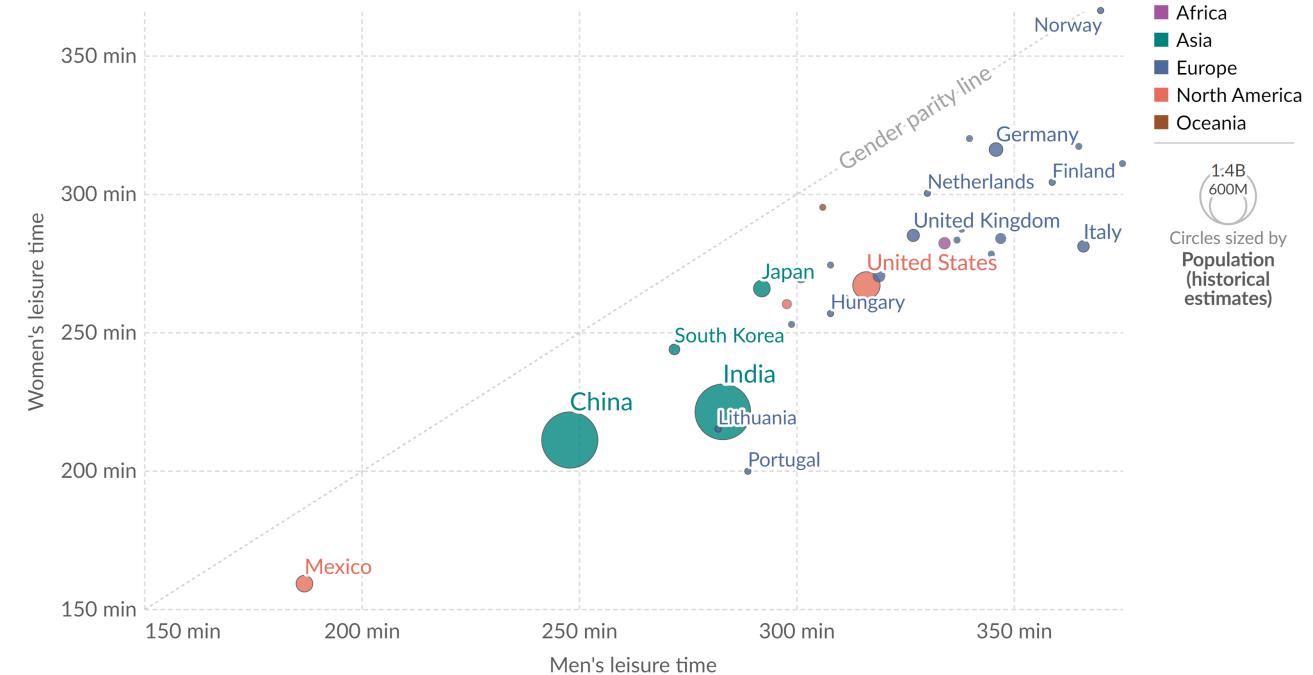
# Plotting in Layers - Aesthetics (2/2)



## Visual properties of a graph

### Gender gap in leisure time, 2020

Average minutes spent on leisure activities, per day, by sex (ages 15-65). For most countries surveys were conducted between 2009 and 2016, but surveys for some countries are older.



Data source: OECD Gender Data

Note: Leisure activities include: sports, attending events, visiting friends, watching TV and other leisure activities.

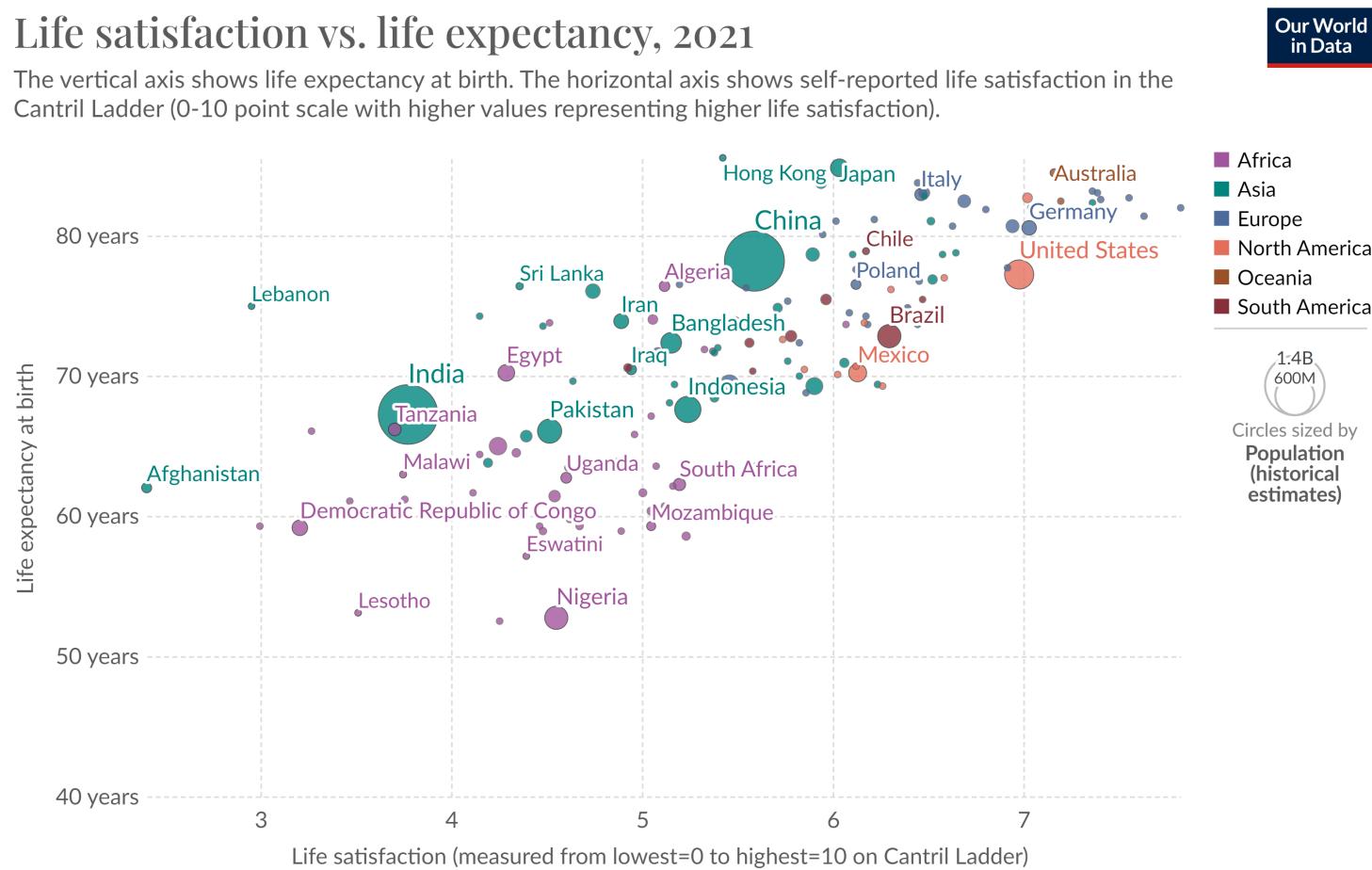
[OurWorldInData.org/happiness-and-life-satisfaction](https://ourworldindata.org/happiness-and-life-satisfaction) | CC BY

READ MORE: <https://ggplot2.tidyverse.org/reference/index.html#aesthetics>

# Mapping: Data → Aesthetics → Graphics (1A/4)

## Life satisfaction vs. life expectancy, 2021

The vertical axis shows life expectancy at birth. The horizontal axis shows self-reported life satisfaction in the Cantril Ladder (0-10 point scale with higher values representing higher life satisfaction).



Data source: UN, World Population Prospects (2022); World Happiness Report (2023)

[OurWorldInData.org/happiness-and-life-satisfaction](https://OurWorldInData.org/happiness-and-life-satisfaction) | CC BY

**READ MORE:** <https://ourworldindata.org/grapher/life-satisfaction-vs-life-expectancy?time=2021>

## Data:

- Life expectancy at birth
- Life satisfaction
- Countries
- Continent
- Population

## Aesthetic: aes()

- y-axis
- x-axis
- text-label
- fill
- size

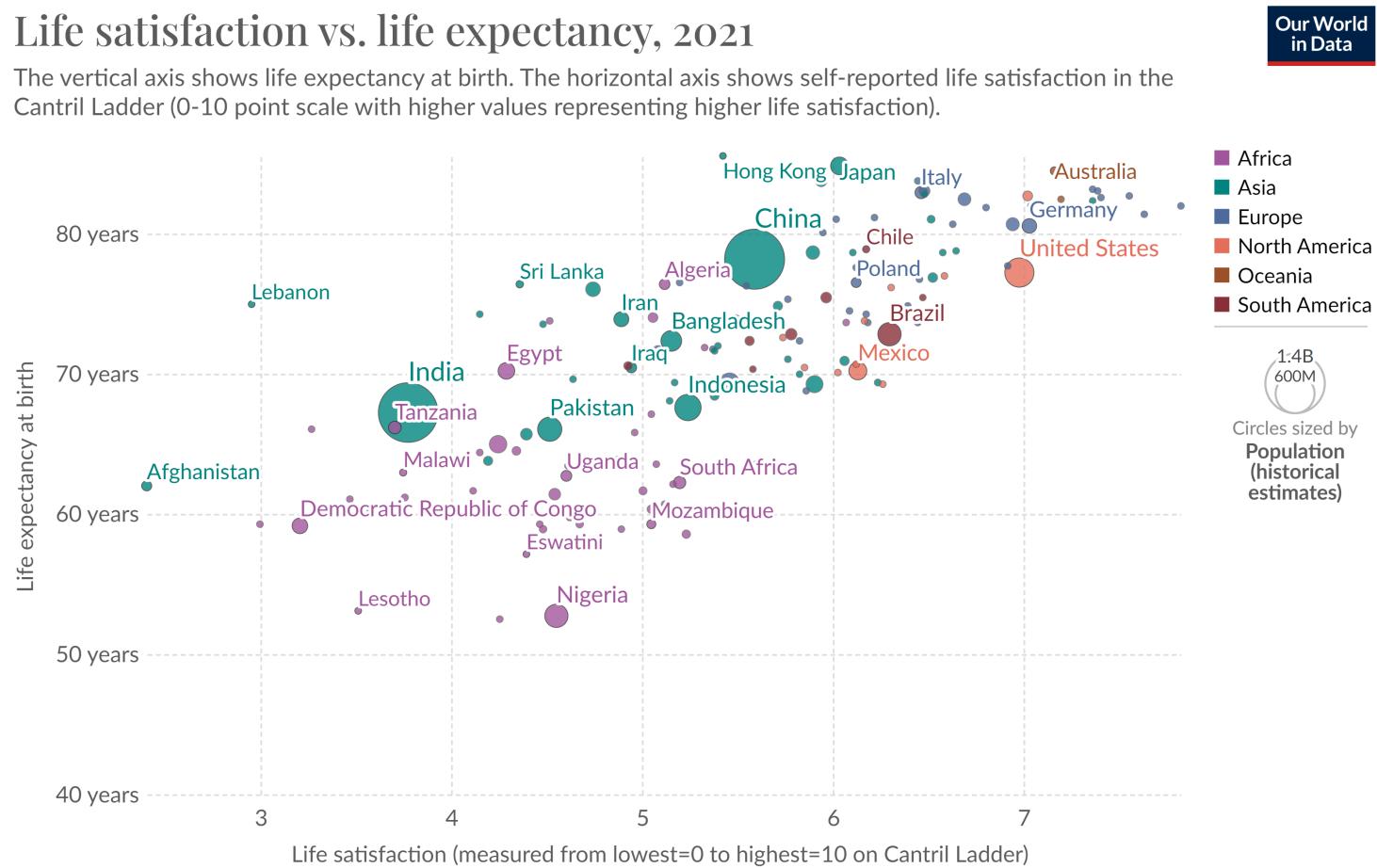
## Grammar of graphics: geom\_

- \_point/\_text

# Mapping: Data → Aesthetics → Graphics (1B/4)

## Life satisfaction vs. life expectancy, 2021

The vertical axis shows life expectancy at birth. The horizontal axis shows self-reported life satisfaction in the Cantril Ladder (0-10 point scale with higher values representing higher life satisfaction).



Data source: UN, World Population Prospects (2022); World Happiness Report (2023)

[OurWorldInData.org/happiness-and-life-satisfaction](https://OurWorldInData.org/happiness-and-life-satisfaction) | CC BY

**READ MORE:** <https://ourworldindata.org/grapher/life-satisfaction-vs-life-expectancy?time=2021>

## Data:

- Life expectancy at birth
- Life satisfaction
- Countries
- Continent
- Population

## Aesthetic: aes()

- y-axis
- x-axis
- text-label
- fill
- size

## Grammar of graphics: geom\_

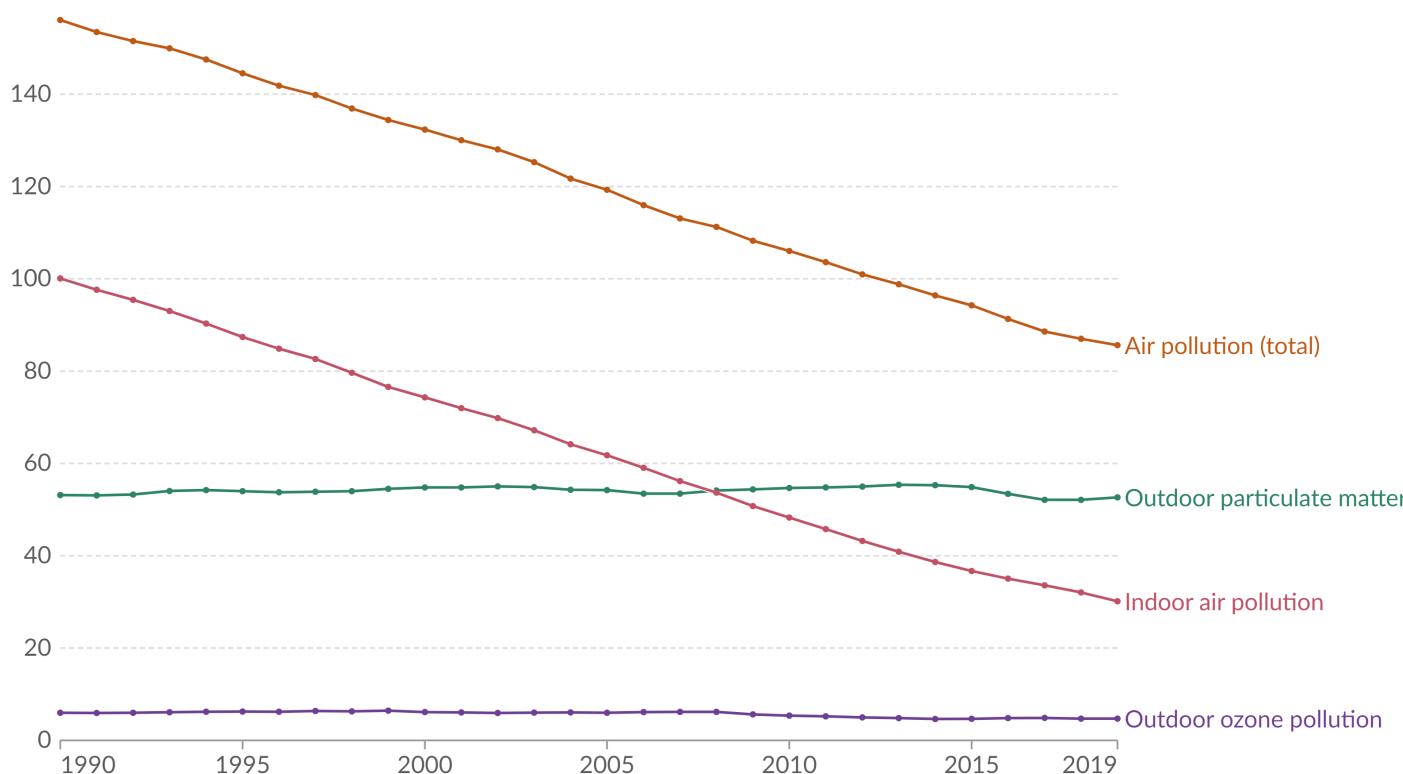
- \_point/\_text

# Mapping: Data → Aesthetics → Graphics (2A/4)

## Death rate from air pollution, World, 1990 to 2019

Death rates are given as the number of attributed deaths from pollution per 100,000 population. These rates are age-standardized, meaning they assume a constant age structure of the population: this allows for comparison between countries and over time.

Our World  
in Data



Data source: IHME, Global Burden of Disease (2019)

[OurWorldInData.org/air-pollution](https://ourworldindata.org/air-pollution) | CC BY

### Data:

- Death rates (per 100,000 population)
- Year
- Air pollution

### Aesthetic: aes()

- y-axis
- x-axis
- color

### Grammar of graphics: geom\_

- \_line

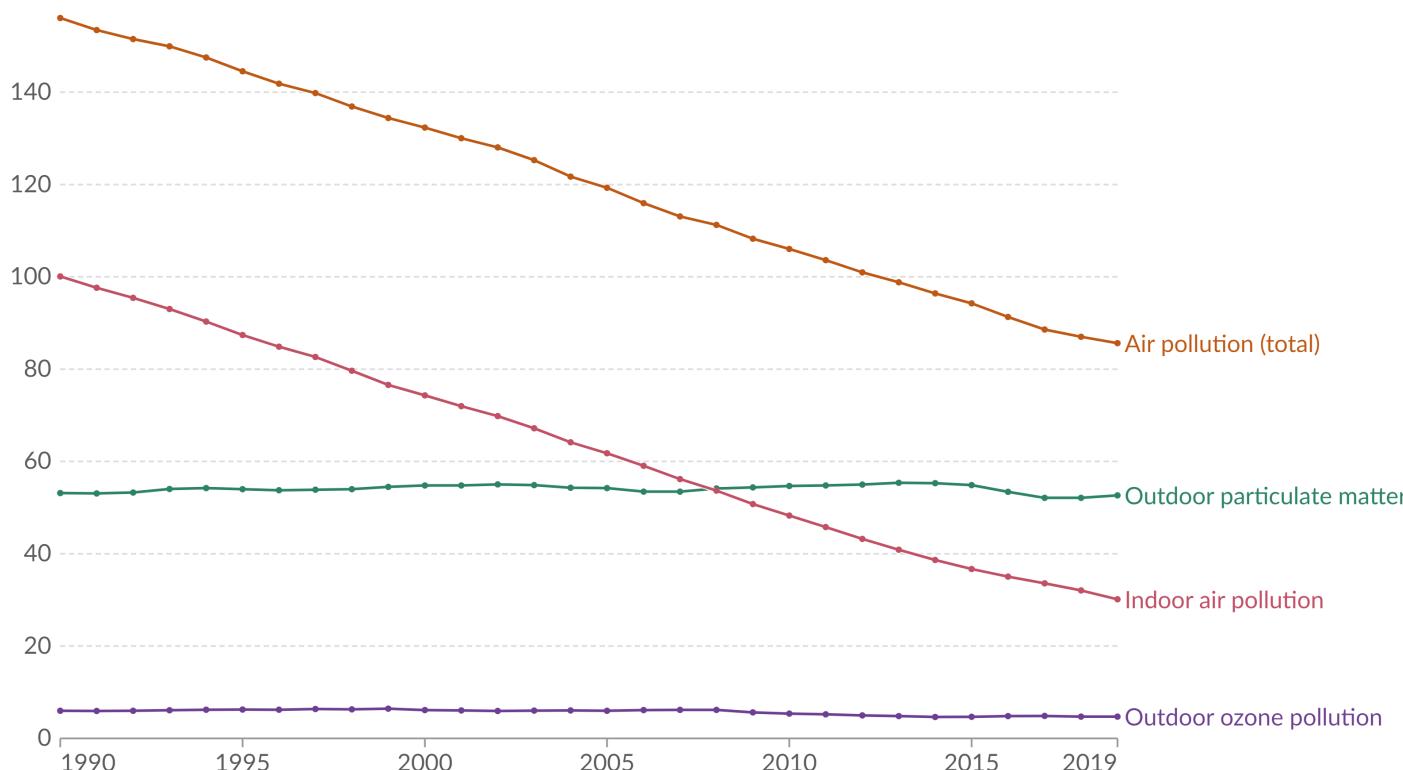
**READ MORE:** <https://ourworldindata.org/grapher/death-rates-from-air-pollution?time=earliest..2019>

# Mapping: Data → Aesthetics → Graphics (2B/4)

## Death rate from air pollution, World, 1990 to 2019

Our World  
in Data

Death rates are given as the number of attributed deaths from pollution per 100,000 population. These rates are age-standardized, meaning they assume a constant age structure of the population: this allows for comparison between countries and over time.



Data source: IHME, Global Burden of Disease (2019)

[OurWorldInData.org/air-pollution](https://OurWorldInData.org/air-pollution) | CC BY

### Data:

- Death rates (per 100,000 population)
- Year
- Air pollution

### Aesthetic: aes()

- y-axis
- x-axis
- color

### Grammar of graphics: geom\_

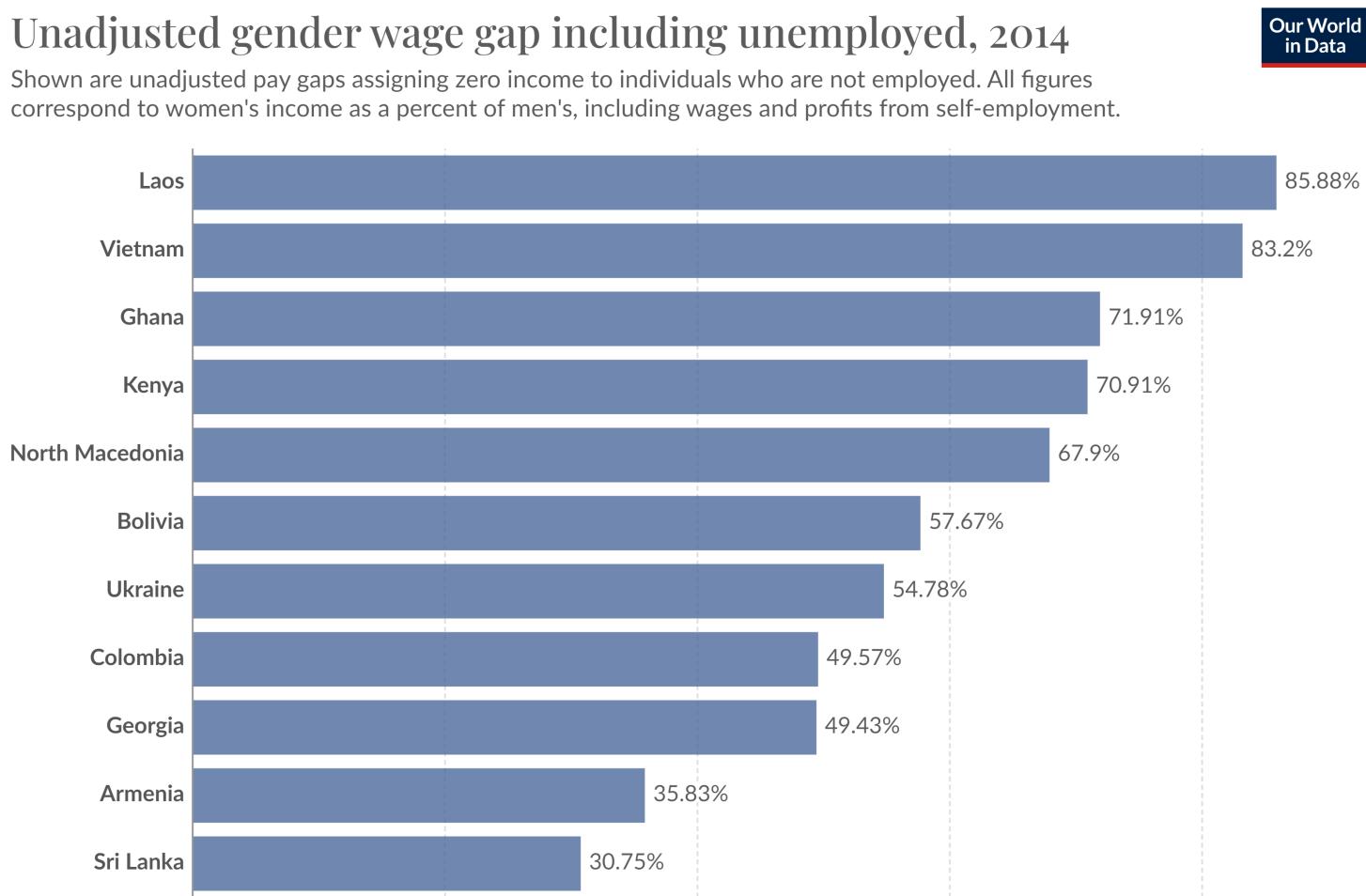
- \_line

**READ MORE:** <https://ourworldindata.org/grapher/death-rates-from-air-pollution?time=earliest..2019>

# Mapping: Data → Aesthetics → Graphics (3A/4)

## Unadjusted gender wage gap including unemployed, 2014

Shown are unadjusted pay gaps assigning zero income to individuals who are not employed. All figures correspond to women's income as a percent of men's, including wages and profits from self-employment.



Data source: Center for Global Development (2018)

[OurWorldInData.org/economic-inequality-by-gender](http://OurWorldInData.org/economic-inequality-by-gender) | CC BY

### Data:

- Country
- Women's income (% of men's income)

### Aesthetic: aes()

- y-axis
- x-axis

### Grammar of graphics: geom\_

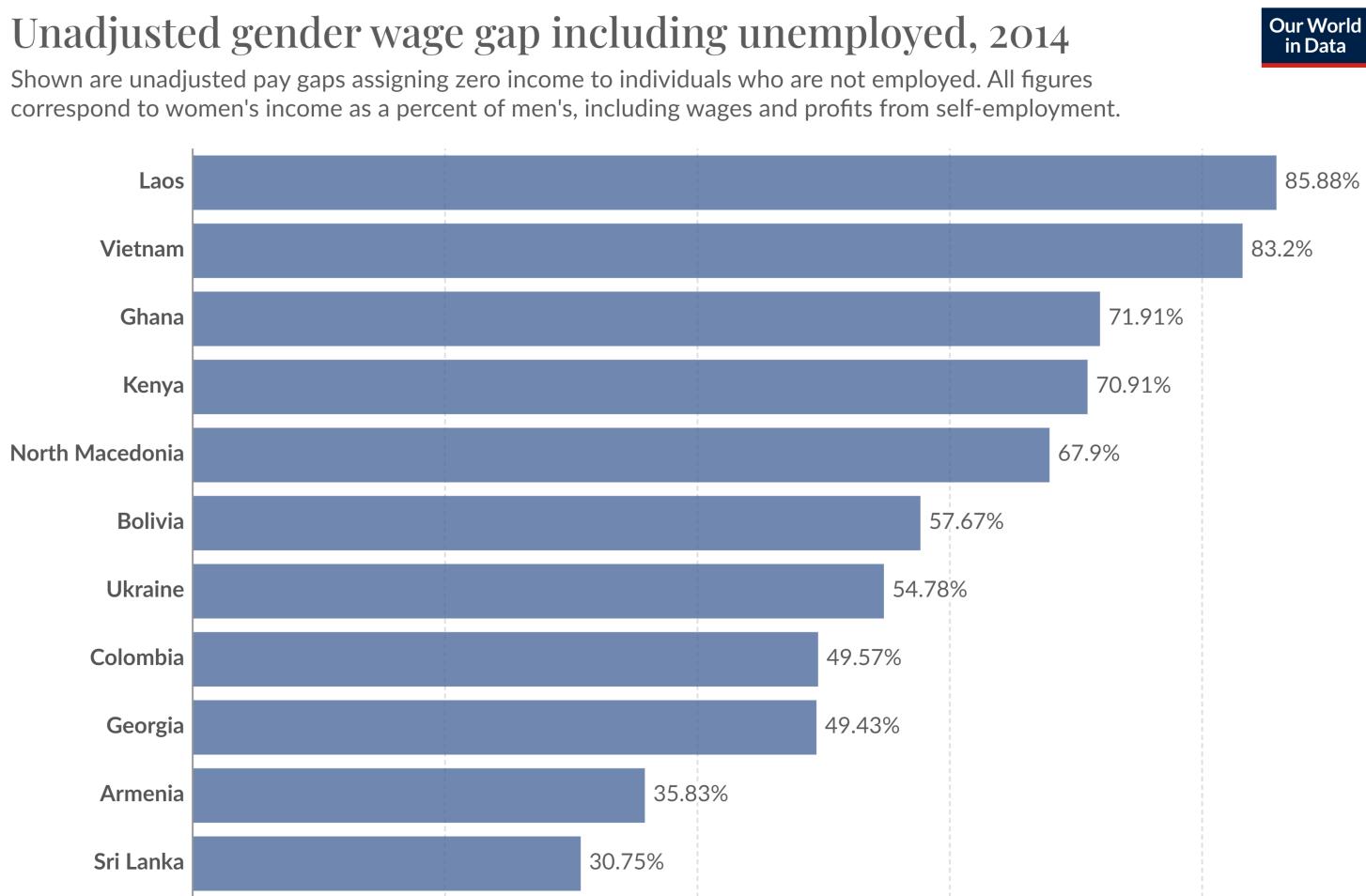
- \_bar/ \_col

**READ MORE:** <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

# Mapping: Data → Aesthetics → Graphics (3B/4)

## Unadjusted gender wage gap including unemployed, 2014

Shown are unadjusted pay gaps assigning zero income to individuals who are not employed. All figures correspond to women's income as a percent of men's, including wages and profits from self-employment.



Data source: Center for Global Development (2018)

[OurWorldInData.org/economic-inequality-by-gender](http://OurWorldInData.org/economic-inequality-by-gender) | CC BY

### Data:

- Country
- Women's income (% of men's income)

### Aesthetic: aes()

- y-axis
- x-axis

### Grammar of graphics: geom\_

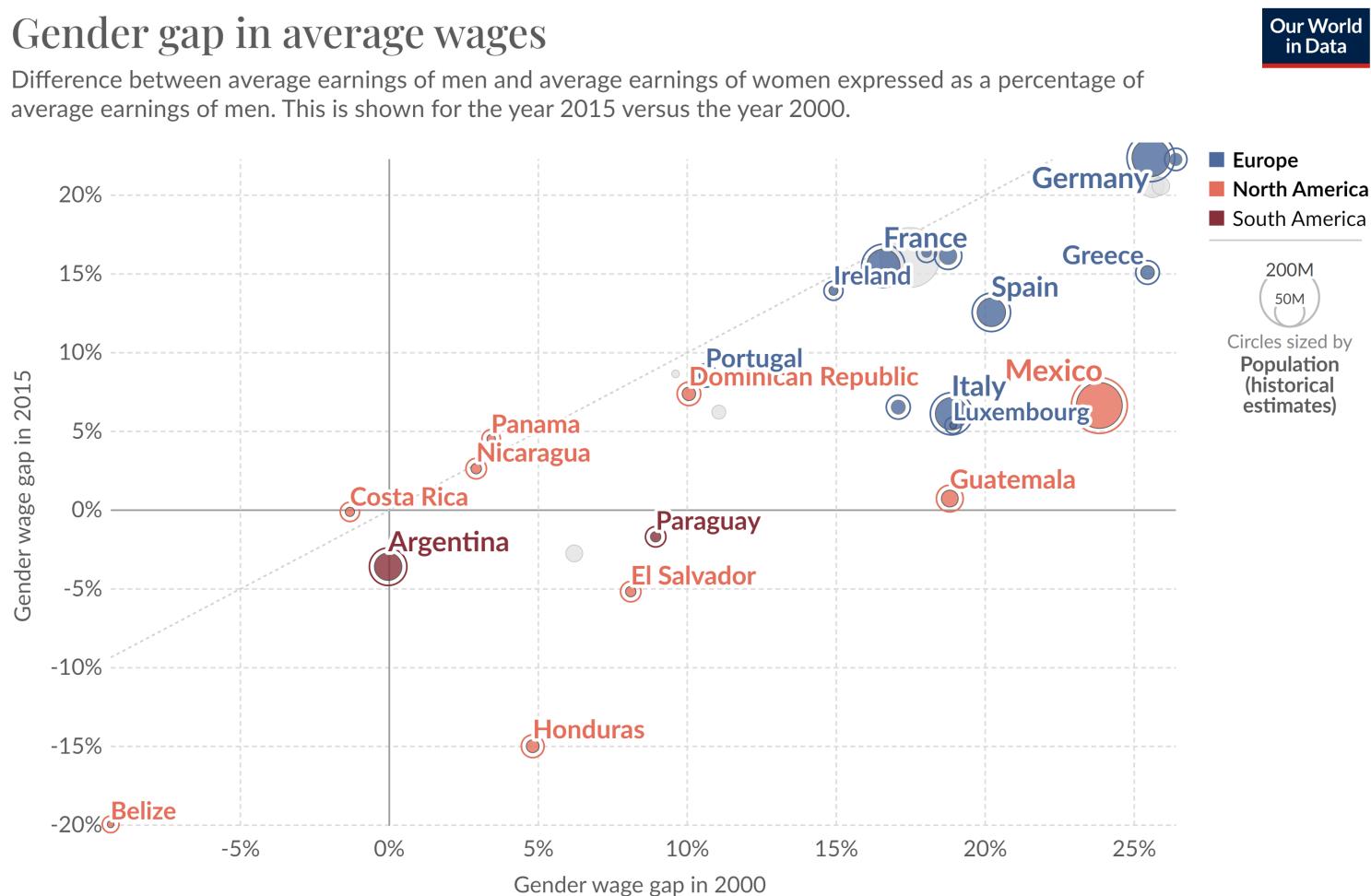
- \_bar/ \_col

**READ MORE:** <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

# Mapping: Data → Aesthetics → Graphics (4A/4)

## Gender gap in average wages

Difference between average earnings of men and average earnings of women expressed as a percentage of average earnings of men. This is shown for the year 2015 versus the year 2000.



Data source: ILOSTAT

[OurWorldInData.org/economic-inequality-by-gender](http://OurWorldInData.org/economic-inequality-by-gender) | CC BY

**READ MORE:** <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

## Data:

- Gender wage gap (in 2015)
- Gender wage gap (in 2000)
- Continent
- Population
- Countries

## Aesthetic: aes()

- y-axis
- x-axis
- color
- size
- label

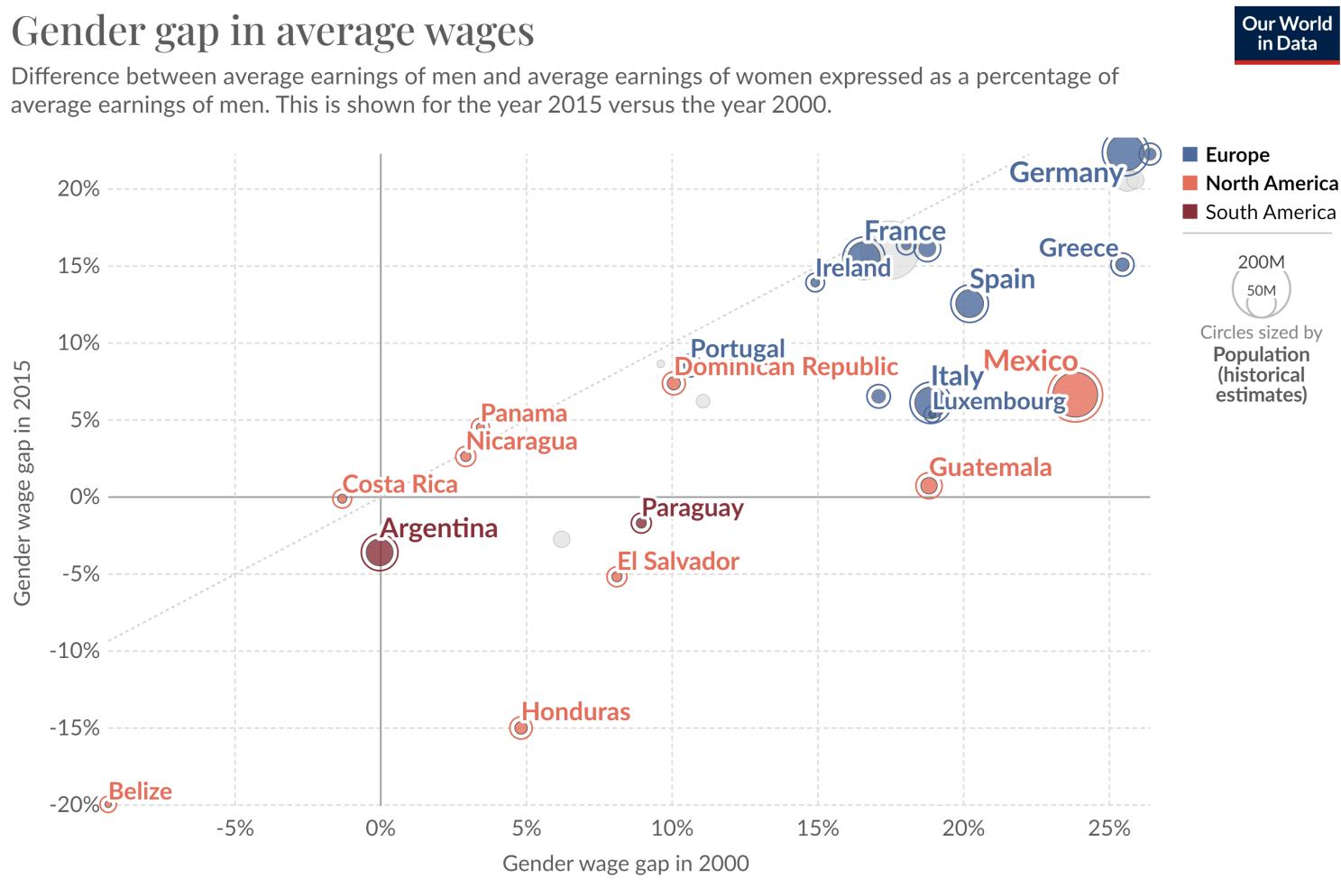
## Grammar of graphics: geom\_

- \_point/ \_text

# Mapping: Data → Aesthetics → Graphics (4B/4)

## Gender gap in average wages

Difference between average earnings of men and average earnings of women expressed as a percentage of average earnings of men. This is shown for the year 2015 versus the year 2000.



Data source: ILOSTAT

[OurWorldInData.org/economic-inequality-by-gender](http://OurWorldInData.org/economic-inequality-by-gender) | CC BY

**READ MORE:** <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

## Data:

- Gender wage gap (in 2015)
- Gender wage gap (in 2000)
- Continent
- Population
- Countries

## Aesthetic: aes()

- y-axis
- x-axis
- color
- size
- label

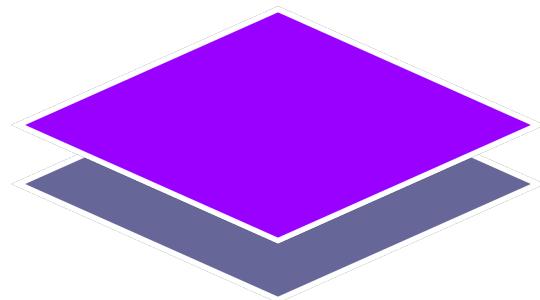
## Grammar of graphics: geom\_ \_point/\_ \_text

# NEXT:

## How to create data visualizations in R



# Mapping: Data → Aesthetics → Graphics (in R)



```
install.packages("ggplot")
library(ggplot)
Data %>% ggplot () +
  geom_* ()
```

{

- geom\_col()
- geom\_point()
- geom\_bar ()
- geom\_line ()
- geom\_text ()
- geom\_tile ()

**READ MORE:** <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

# Mapping: Data → Aesthetics → Graphics (in R)



```
install.packages("ggplot")
library(ggplot)

Data %>% ggplot () +
  geom_*(mapping = aes()) +
  labs () +
  theme_()
```

Geometries (yellow boxes):

- geom\_col()
- geom\_point()
- geom\_bar ()
- geom\_line ()
- geom\_text ()
- geom\_tile ()

Themes (purple boxes):

- theme\_bw()
- theme\_minimal()
- theme\_light()
- theme\_dark()
- theme\_gray()
- theme\_classic()

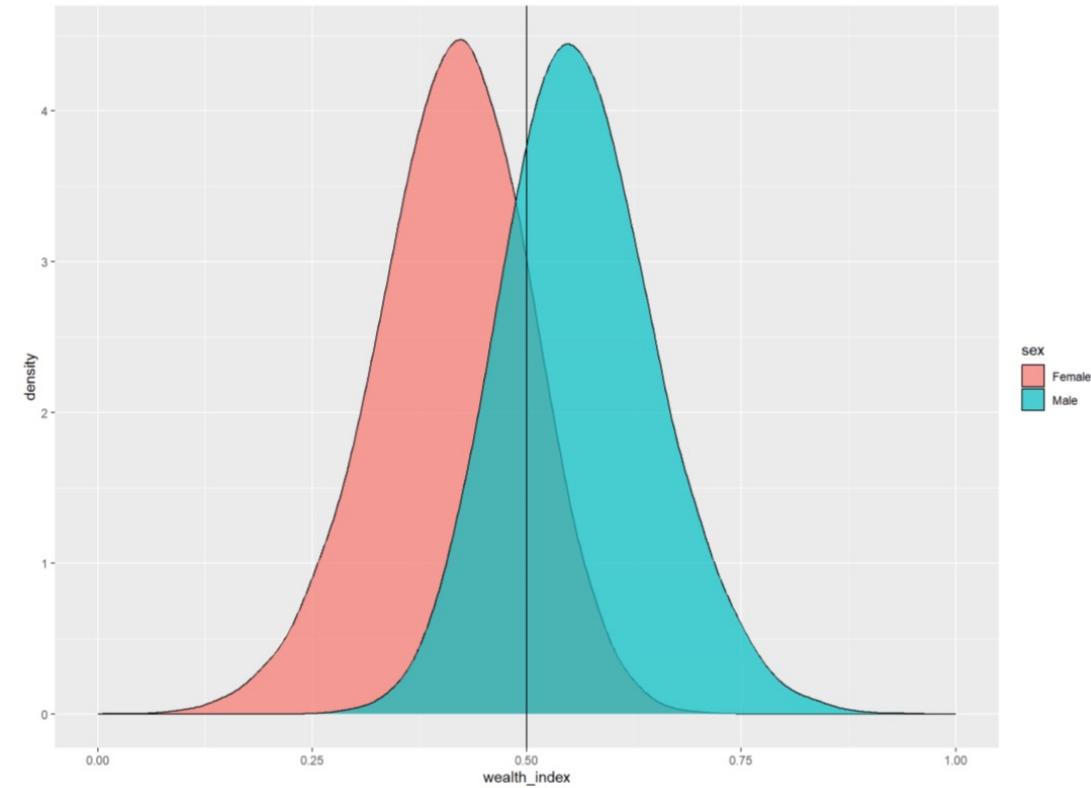
READ MORE: <https://ourworldindata.org/grapher/unadjusted-gender-wage-gap-including-unemployed>

# Visualising distributions with density plots (1/3): geom\_density()

```
Data     study_dta %>%  
        ggplot () +  
        geom_density(aes(group = sex,  
                           fill = sex,  
                           x = wealth_index),  
                       adjust=1.5, alpha = 0.7) +  
        geom_vline(xintercept = 0.5)
```

Geometries

Aesthetic



READ MORE: <https://ggplot2.tidyverse.org/>

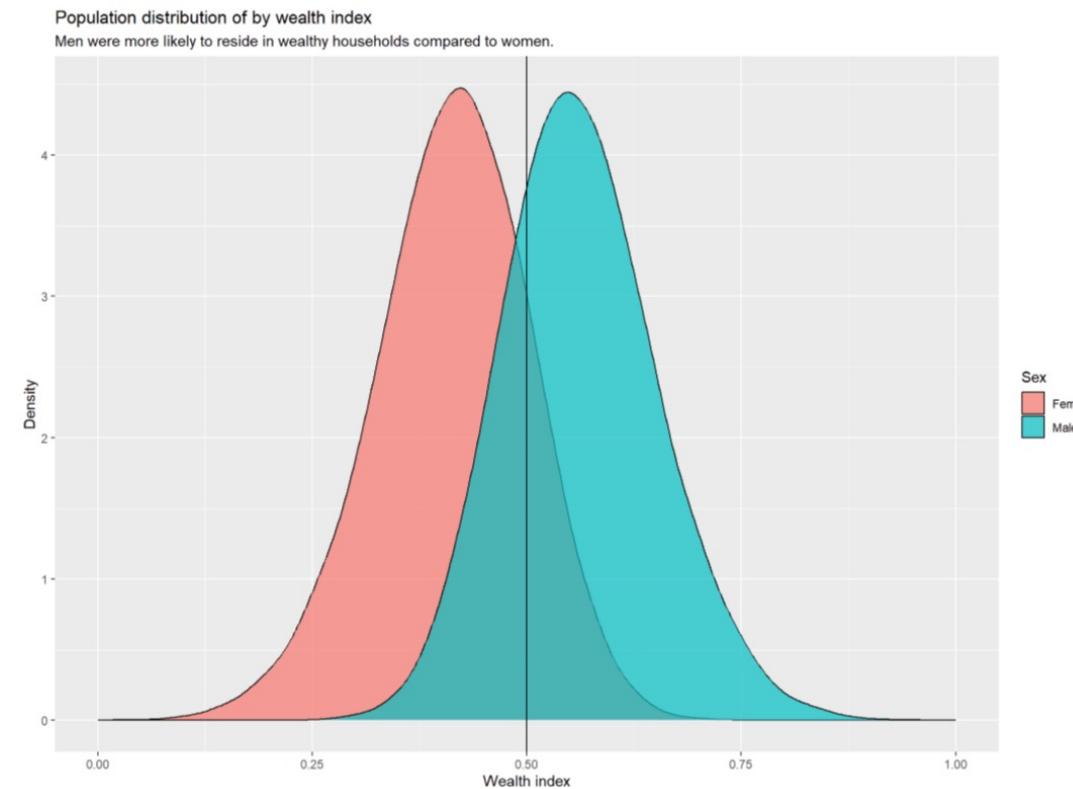
# Visualising distributions with density plots (2/3): geom\_density()

```
Data     study_dta %>%  
        ggplot () +  
        geom_density(aes(group = sex,  
                           fill = sex,  
                           x = wealth_index),  
                       adjust=1.5, alpha = 0.7) +  
        geom_vline(xintercept = 0.5) +  
        labs (x = "Wealth index", y = "Density",  
              fill = "Sex", title = "Population  
distribution of by wealth index",  
              subtitle = "Women were more  
likely to reside in wealthy  
households compared to men.")
```

Geometries {

Aesthetic }

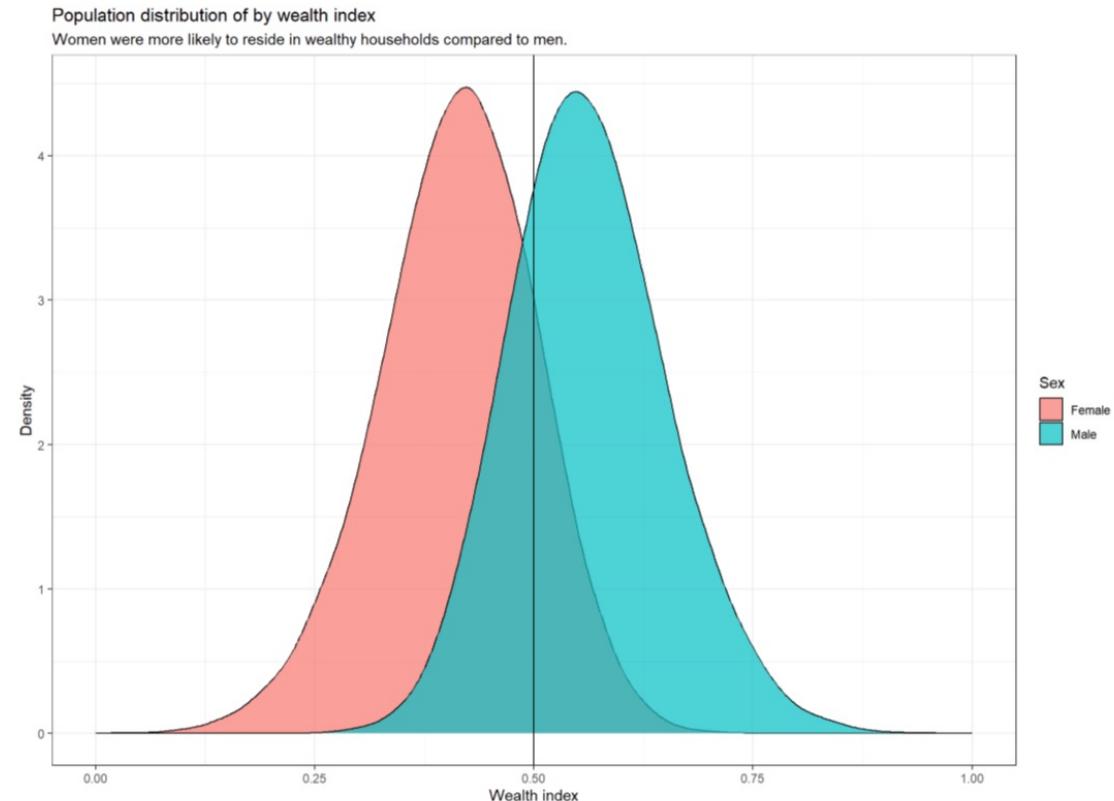
Labels {



READ MORE: <https://ggplot2.tidyverse.org/>

# Visualising distributions with density plots (3/3): geom\_density()

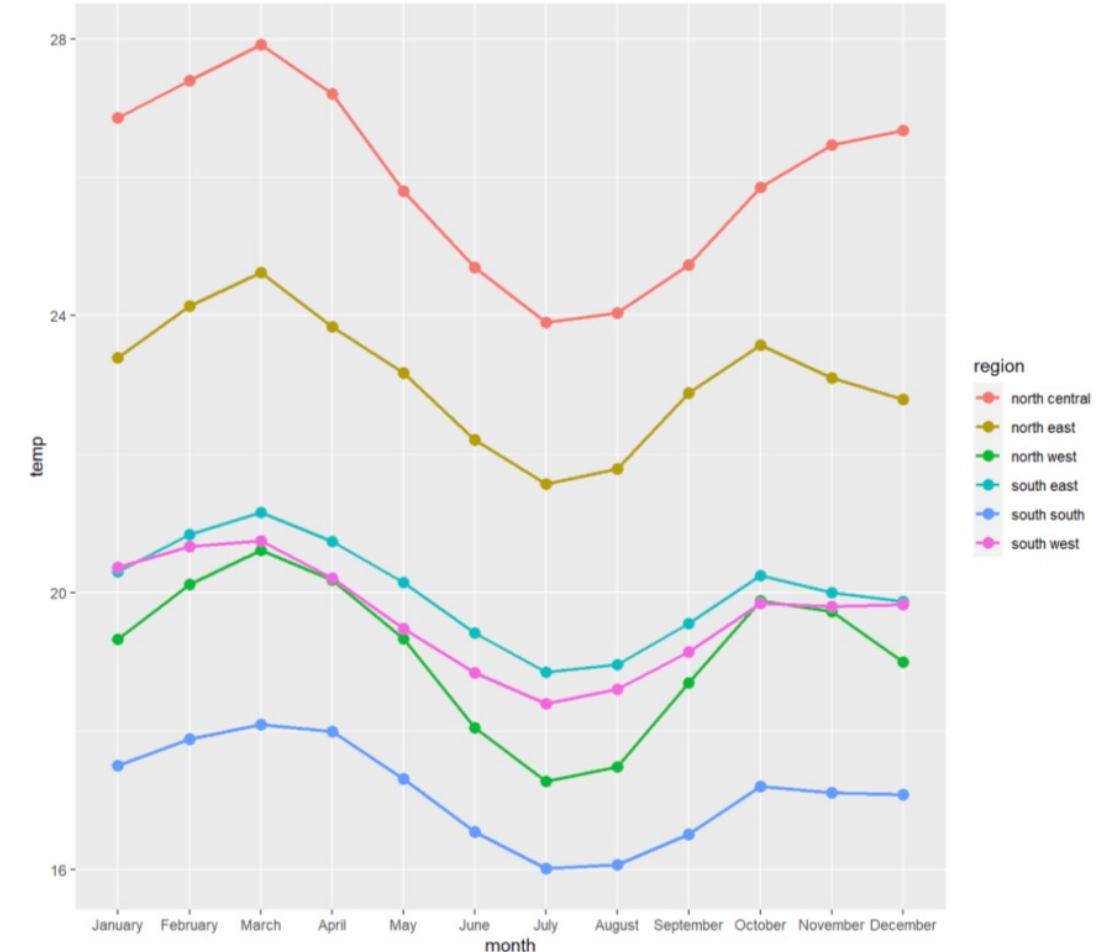
```
Data      study_dta %>%  
ggplot () +  
geom_density(aes(group = sex,  
                 fill = sex,  
                 x = wealth_index),  
             adjust=1.5, alpha = 0.7) +  
geom_vline(xintercept = 0.5) +  
labs (x = "Wealth index", y = "Density",  
      fill = "Sex", title = "Population  
distribution of by wealth index",  
      subtitle = "Women were more  
likely to reside in wealthy  
households compared to men.") +  
theme_bw()
```



READ MORE: <https://ggplot2.tidyverse.org/>

# Visualising trends with line graphs (1/3): geom\_line()

```
Data   study_dta %>%  
      ggplot (aes(x = month, y = temp,  
                     color = region,  
                     group = region)) +  
  
Geometries {geom_line (size = 0.9) +  
            geom_point(size = 3)}
```



READ MORE: <https://ggplot2.tidyverse.org/>

# Visualising trends with line graphs (2/3): geom\_line()

Data

```
study_dta %>%  
  ggplot (aes(x = month, y = temp,  
              color = region,  
              group = region)) +
```

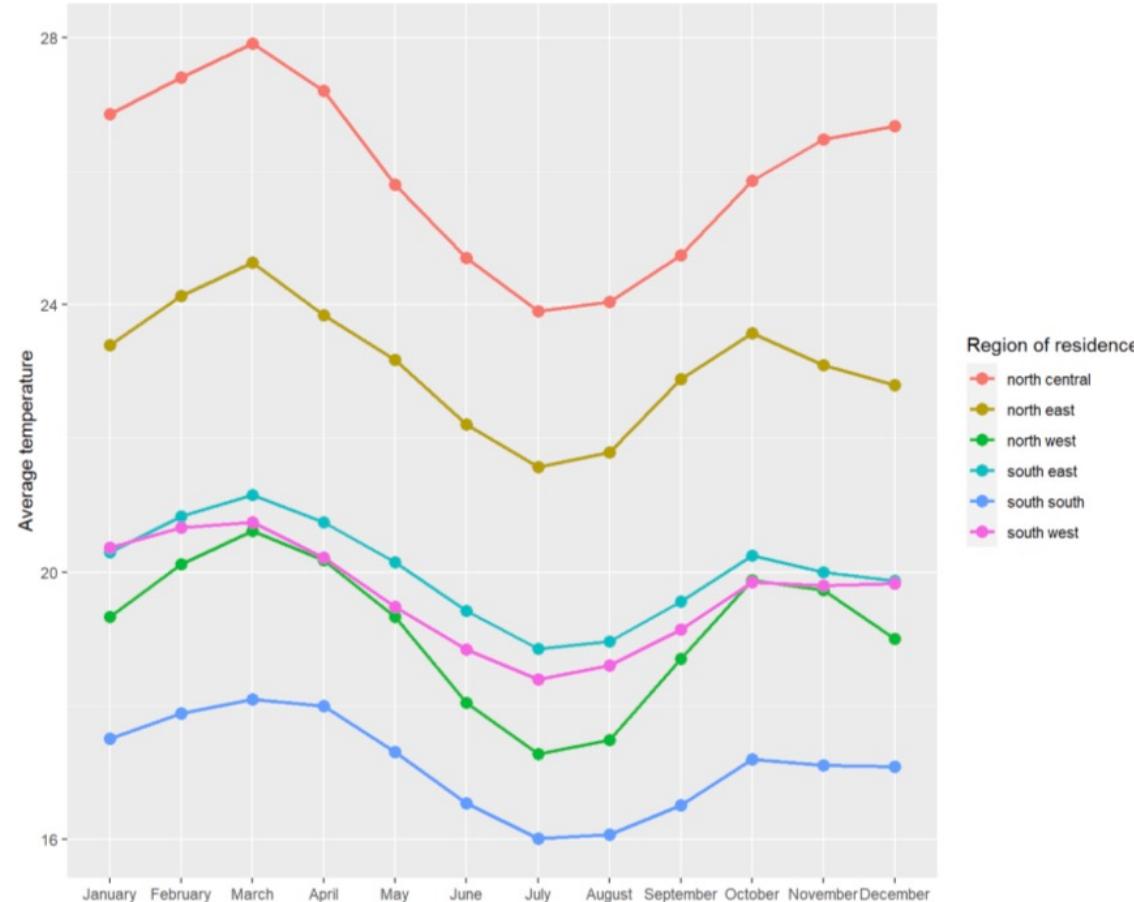
Geometries

```
{geom_line (size = 0.9) +  
 geom_point(size = 3) +
```

Labels

```
labs (y = "Average temperature",  
      x = "", color= "Region of residence",  
      title = "March is the hottest month  
      of the year in Kenya",  
      subtitle = "Average temperature  
      varies significantly across regions  
      in Kenya. \nLess hot in the South-  
      South and hottest in the North-  
      Central regions")
```

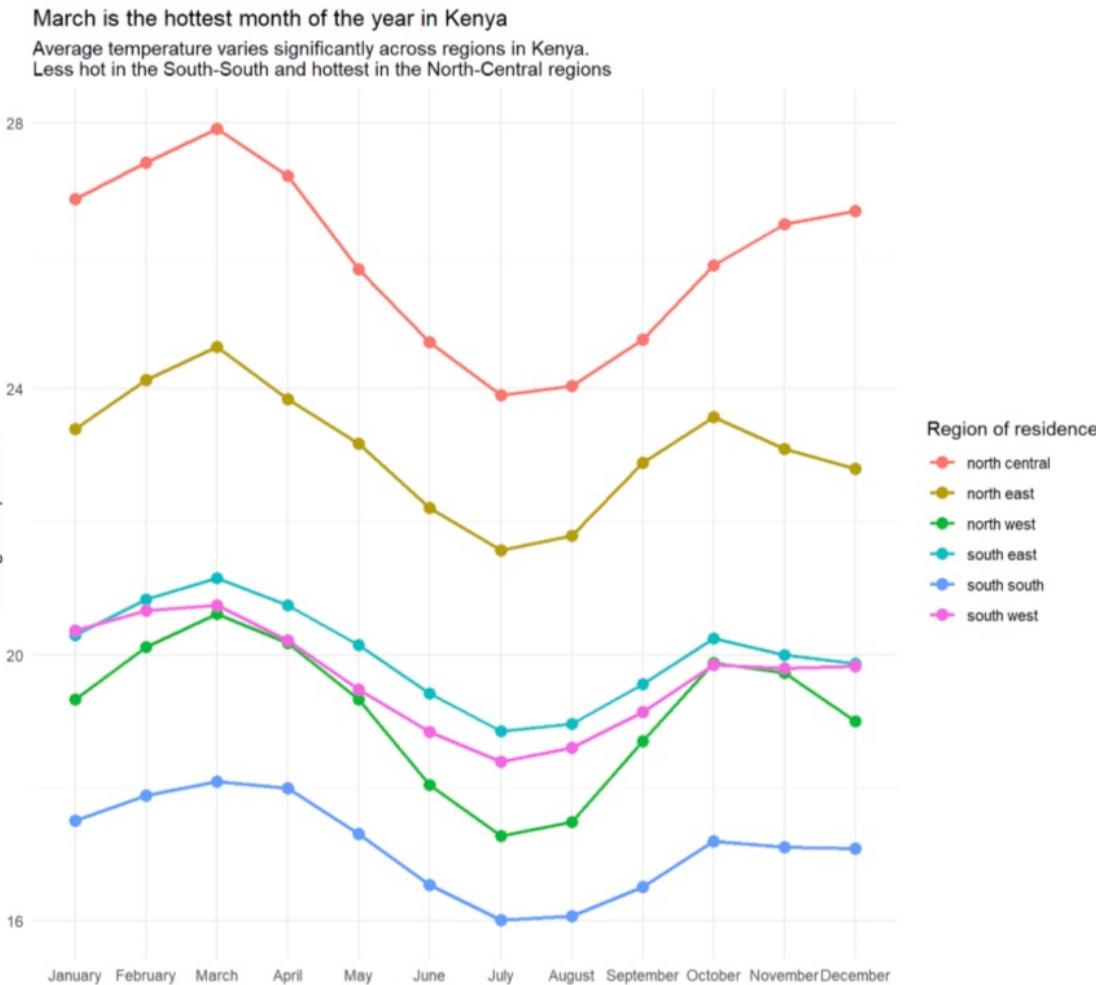
March is the hottest month of the year in Kenya  
Average temperature varies significantly across regions in Kenya.  
Less hot in the South-South and hottest in the North-Central regions



READ MORE: <https://ggplot2.tidyverse.org/>

# Visualising trends with line graphs (3/3): geom\_line()

```
Data   study_dta %>%  
      ggplot (aes(x = month, y = temp,  
                     color = region,  
                     group = region)) +  
  
Geometries { geom_line (size = 0.9) +  
             geom_point(size = 3) +  
             labs (y = "Average temperature",  
                   x = "", color= "Region of residence",  
                   title = "March is the hottest month  
                           of the year in Kenya",  
                   subtitle = "Average temperature  
                           varies significantly across regions  
                           in Kenya. \nLess hot in the South-  
                           South and hottest in the North-  
                           Central regions") +  
  
Labels }  
  
Theme { theme_minimal()
```



READ MORE: <https://ggplot2.tidyverse.org/>

# Your Turn

Open the file  
`intro_ggplot.Rmd`  
in `04-practicals`

Got questions?

Email me: `e.olamijuwon@st-andrews.ac.uk`