

# **RESEARCH METHODS FOR GLOBAL STUDIES**

**CLASS 18: PRESENTING QUANTITATIVE INFORMATION VISUALLY**

John Palmer

# TODAY'S SLIDES

Today's slides will be more detailed than usual so that you can get everything you need from this class regardless of whether you are here in person or on strike and reading this later. You will find notes at the bottom of each slide to walk you through the material. (Note that you may need to scroll down in the notes to see all the text.)

These are the speaker notes. Remember to scroll down to see all of the text here...

*There is no penalty for not coming to class today.*

Keep scrolling...

Keep scrolling...

Good!

# MOTIVATING EXAMPLE

Alice earns 25,000 euros per year; Bob earns 31,000. How can we best communicate this information in a chart?

For small amounts of relatively straightforward quantitative information, a chart or table may not be necessary. That would certainly be the case here. The numbers are clear in the text. In addition, note that we are talking here about two individual wages whereas we will often be more interested in presenting distributions or differences between distributions. We will get to that.

wages

100000

75000

50000

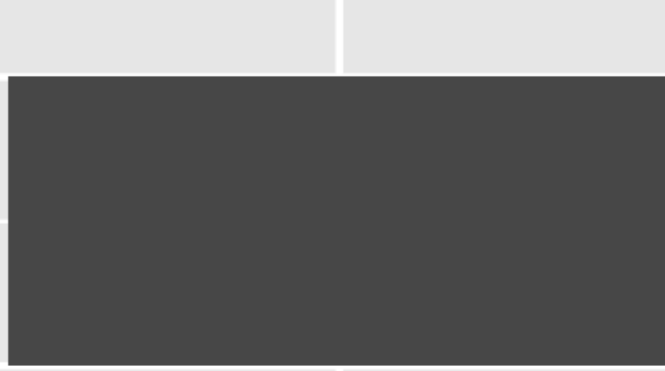
25000

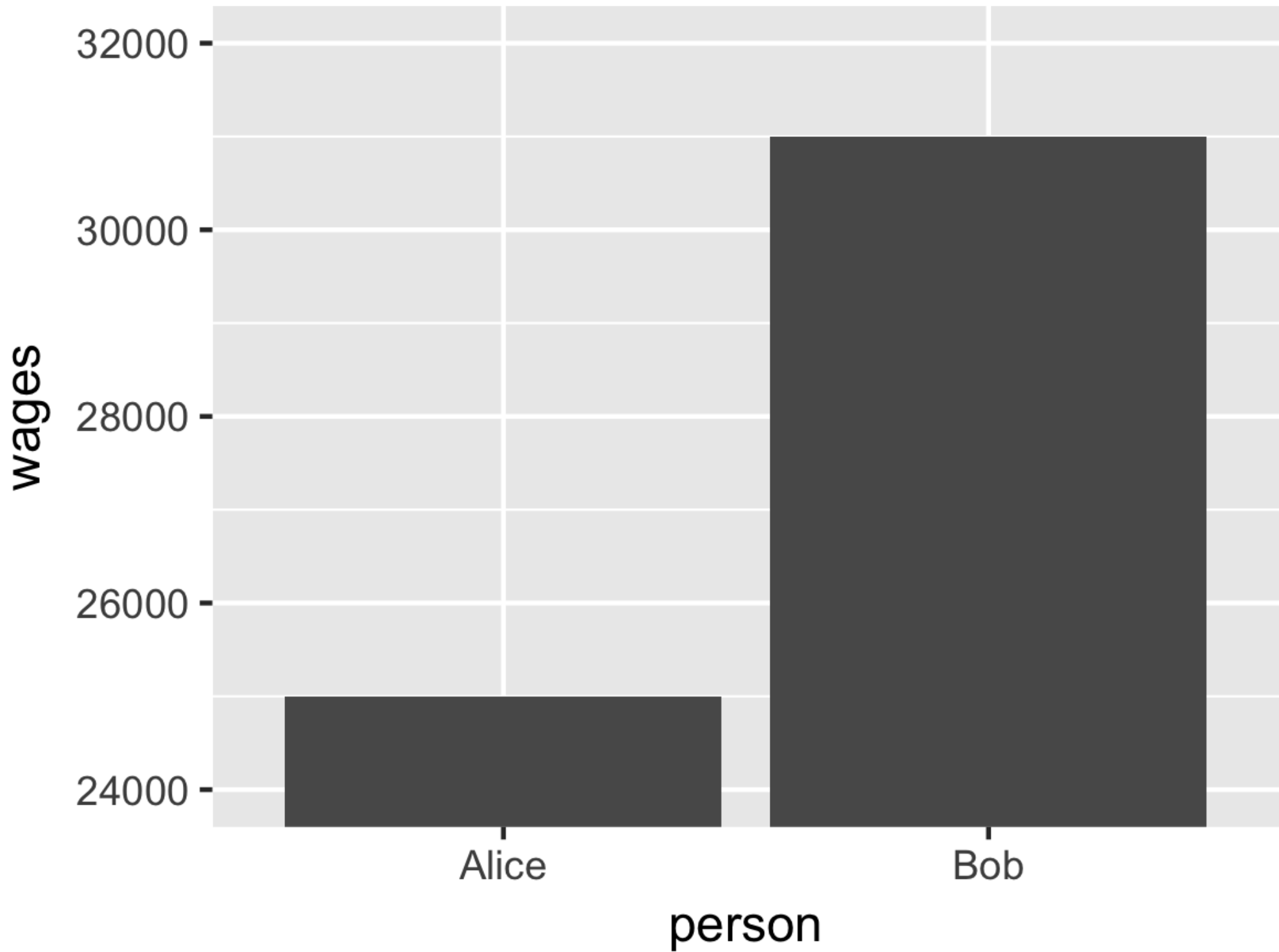
0

Alice

Bob

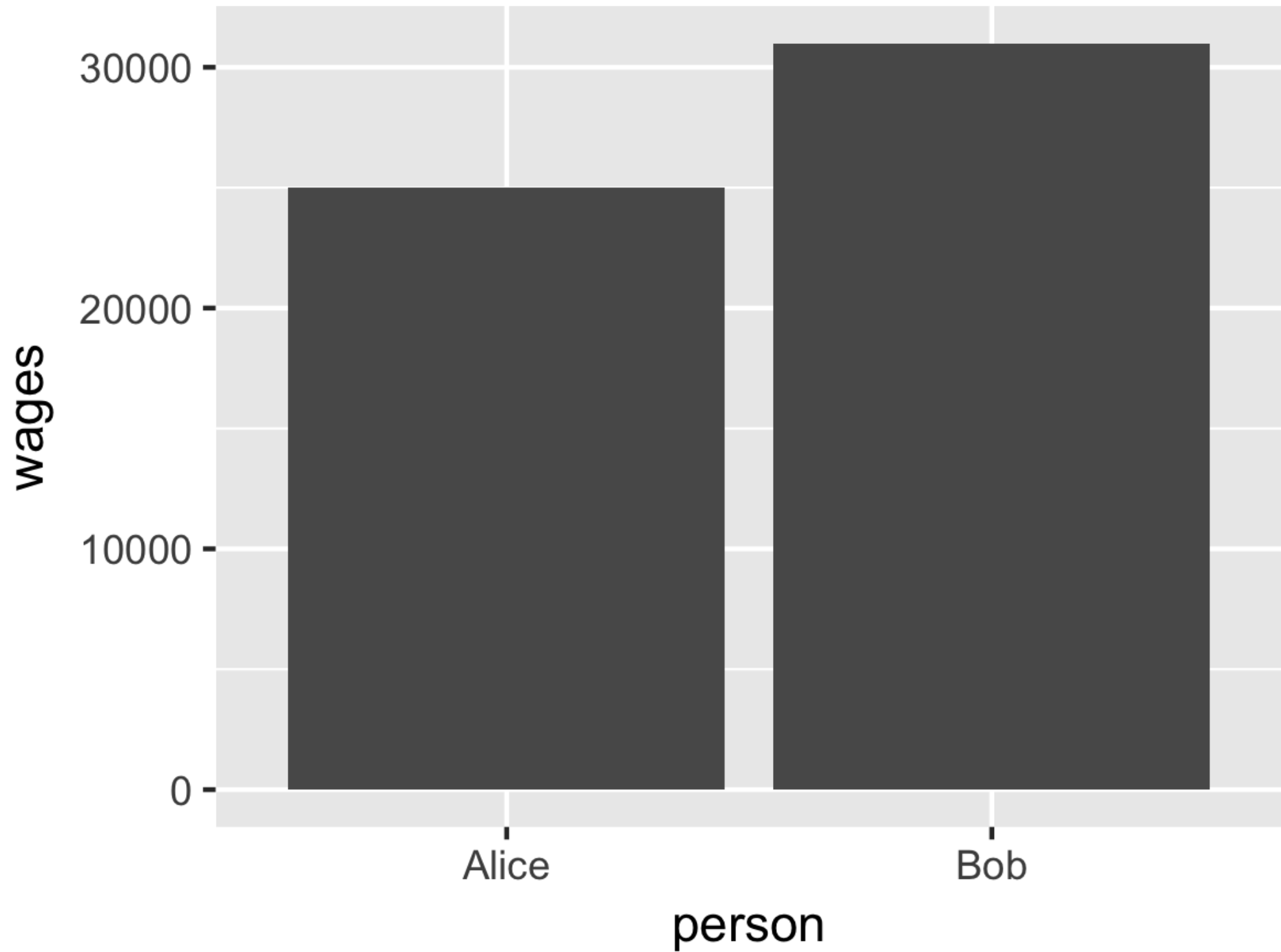
person





## What are the problems with these charts?

Even when displaying simple information, it is easy to make mistakes. The problem with these two charts is that the y-axis is drawn in such a way as to distort the difference. In the first chart, the axis ranges from 0 to 100,000, making the difference look relatively inconsequential. In the second, the axis ranges from 24,000 to 32,000, making the difference look huge. It is possible that there could be a reason to draw the chart in one of these ways, but a better approach is almost certainly to start at 0 and end slightly above the highest value.



## What about this last one?

By letting the y-axis range from 0 to just over 31,000 we can better compare the two wages to each other and compare the difference to each of them. Our interest, of course, may be on the difference or gap in wages here. We could calculate this by subtracting Alice's wages from Bob's, but then we would be left with one value (6,000) and it might be hard to put it into context. If this represents a wage gap in a country we are familiar with, then perhaps we do not need more information. But if we do not know the wage structure or generally how much things cost, then it is hard to know if 6,000 is a large or small amount in practical terms. By showing the two earnings side by side, at least we can compare the gap to the earnings themselves.

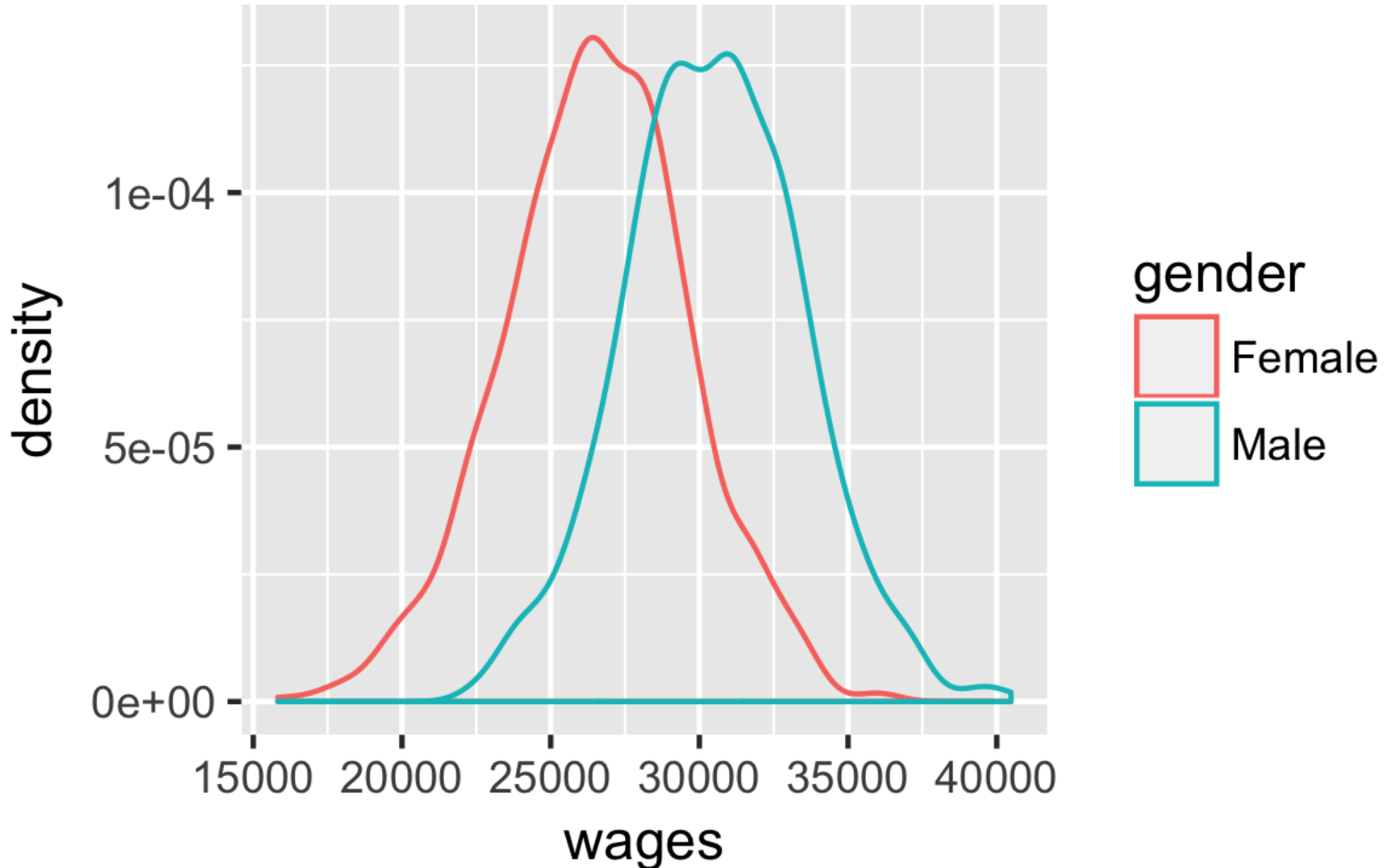


Don't we care more about population distributions than individuals?

Of course that depends on what we are talking about and for what purpose, but in this case, understanding the difference in the distribution of mens' and womens' wages is more interesting (for many purposes) than understanding the difference in the wages of two individuals. Here is one example, using the actual mean wages for industry, construction, and services in Spain in 2014, and generating a synthetic distribution of individual wages by drawing randomly from Guassian distributions with arbitrary standard deviations. (So what follows in the next two plots is just for demonstration purposes.)

# Hypothetical Wage Distributions

(Not real data)



Can we improve this plot?

Since there are only two categories, it is unnecessary to use the side legend. We could easily place the labels directly on the plot, making it easier to read. We can also play with different color combinations to improve the look.

# Hypothetical Wage Distributions





(Not real data)

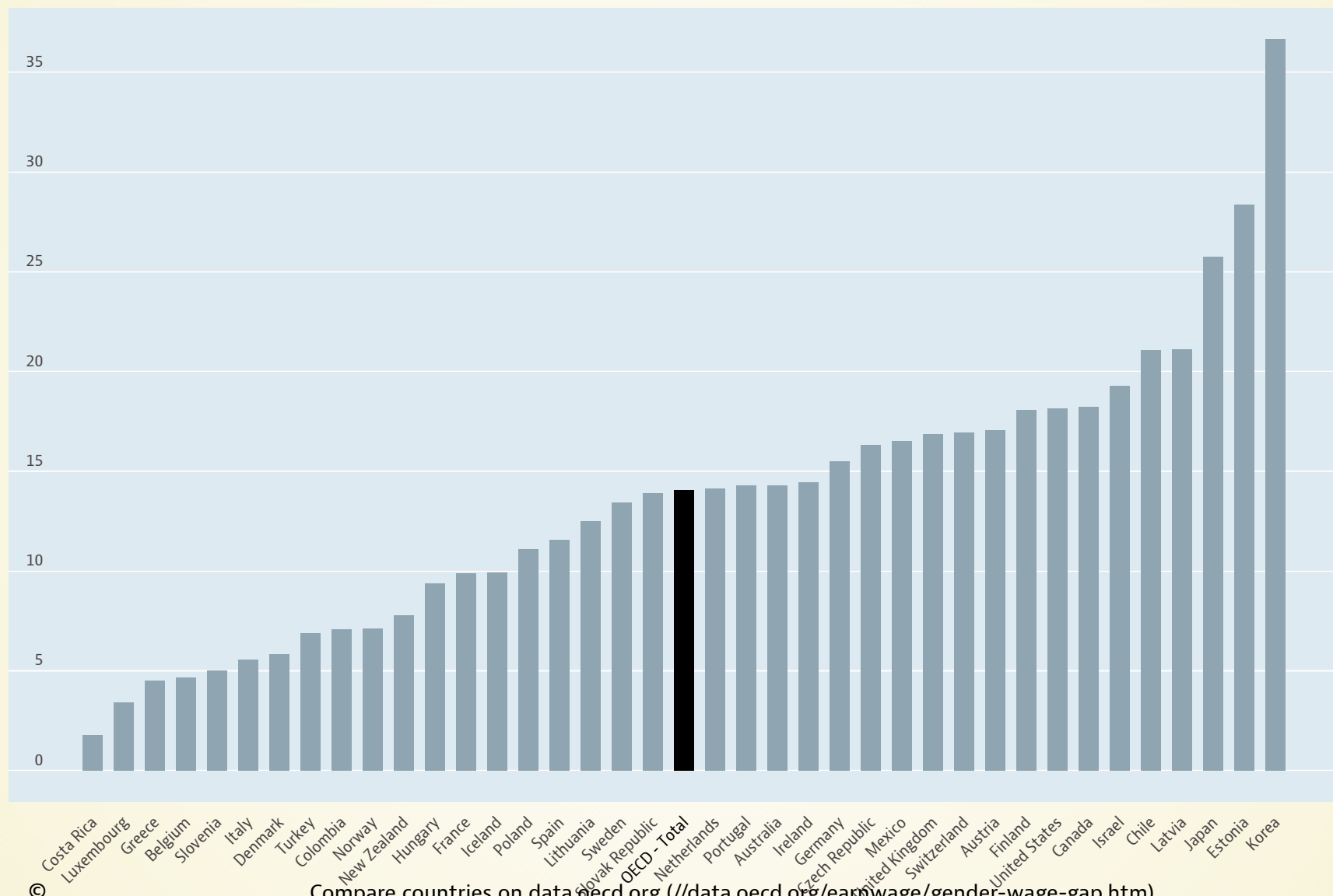


What if we want to compare multiple countries or years?

Density plots are nice for visualizing full distributions, but for making multiple comparisons, we often prefer to simplify things by simply taking means or medians from each distribution. What follows is a chart using real data for the past 5 years.

Gender wage gap ([//data.oecd.org/earnwage/gender-wage-gap.htm](http://data.oecd.org/earnwage/gender-wage-gap.htm)) Total, % of male median wage, 2016 or latest available



©

Compare countries on [data.oecd.org \(//data.oecd.org/earnwage/gender-wage-gap.htm\)](http://data.oecd.org/earnwage/gender-wage-gap.htm)

What is this chart showing us? Can you write a more descriptive caption for it?

# PLANNING A FIGURE

- Qualitative or quantitative information?
- Variables, variable types, subsets
- Format
- Audience
- Message

In planning your visual presentation, you will want to start by considering:

- What type of information is this? Is it qualitative or quantitative?
- If quantitative, how many variables do you want to represent? What types of variables are they? Are there subsets that information should be grouped in?
- What is the format of this presentation? A scientific paper? A policy paper? A poster? Slides?
- Who is your audience? How sophisticated are they in terms of interpreting quantitative data? What types of charts are they used to?
- What message are you trying to convey? What is the purpose of this chart?



# LINKING FIGURES TO TEXT

- Describe every figure in text.
- Refer to figures as, e.g., Figure 1, Table 1...

# MAKE FIGURES SELF-EXPLANATORY

- Add enough information in each figure caption so that figure can be understood without the main text.

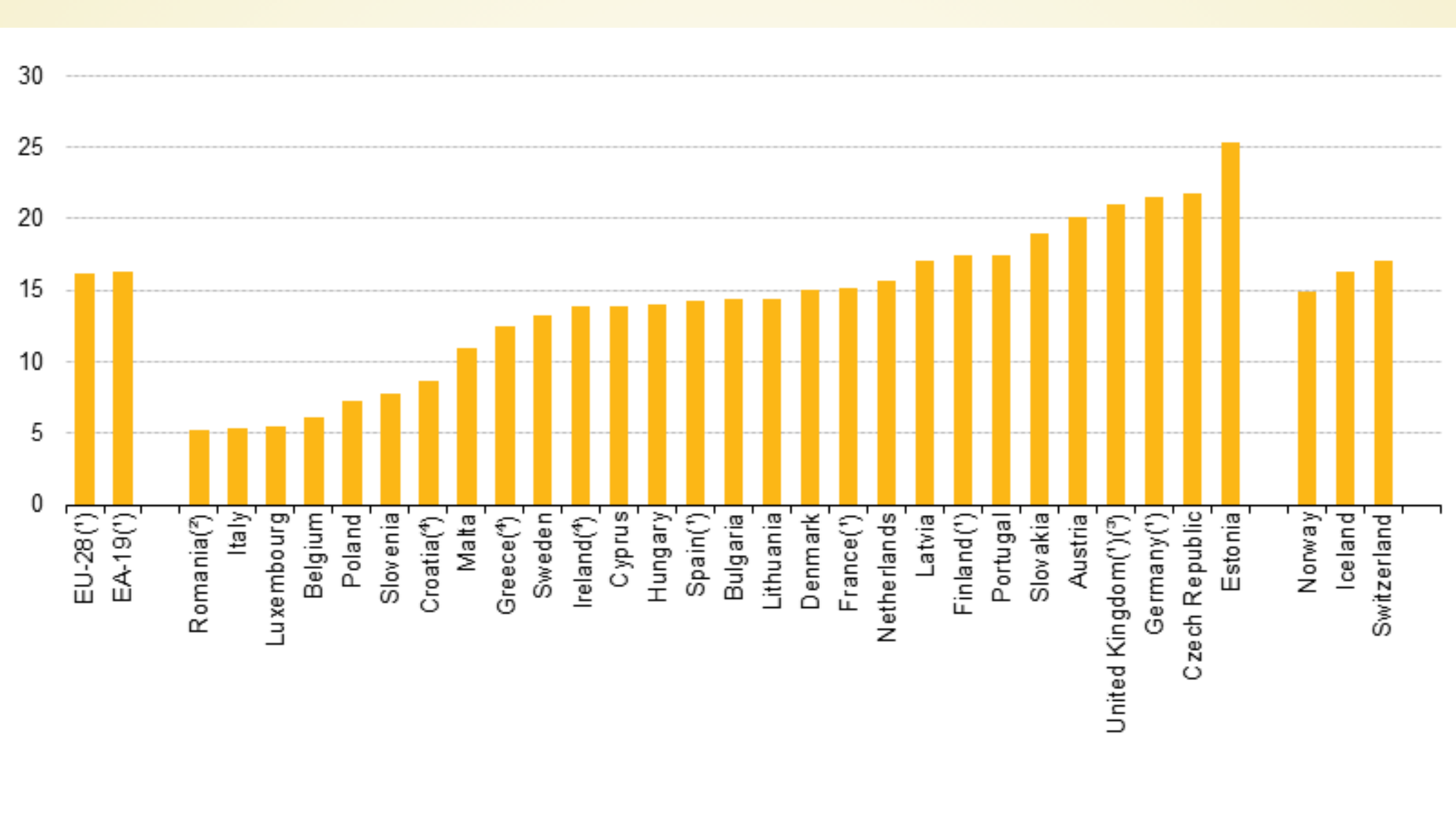
The information in the caption should make it possible for the reader to interpret the figure but should not include the interpretation itself.

# MAKE FIGURES TELL THE STORY

- Ideally, reader should be able to read only the figures and understand the "story" you are telling.

Making this possible is a matter of creating the right charts and adding clear, informative captions

Consider the charts used in the most recent Eurostat report on the gender wage gap. What works well here? What would you improve?



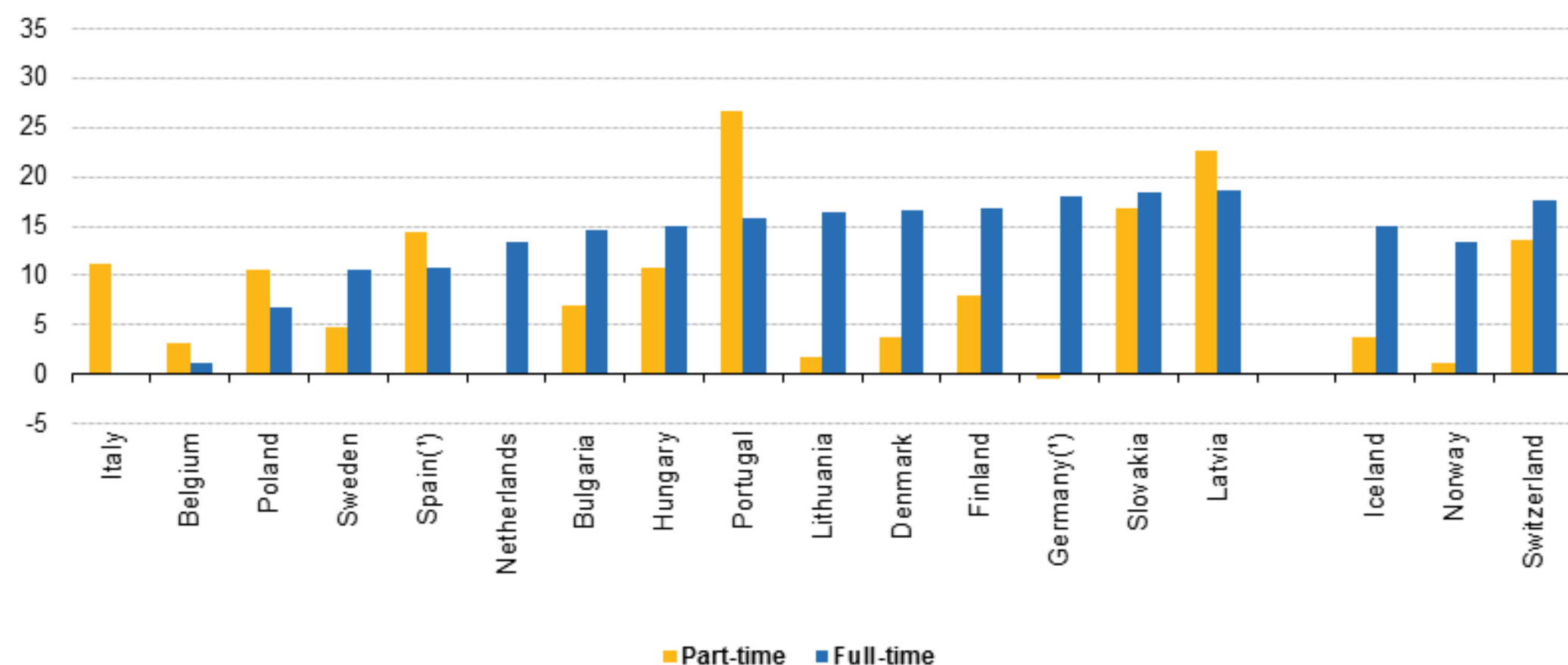
Note: For all countries except the Czech Republic: data for enterprises employing 10 or more employees, NACE Rev. 2 B to S (-O); Czech Republic: data for enterprises employing 1 or more employees, NACE Rev. B to S

<sup>(1)</sup> Provisional data

<sup>(2)</sup> Estimated

<sup>(3)</sup> Estimated by Eurostat

<sup>(4)</sup> 2014 data



Note: Enterprises employing 10 or more employees; NACE Rev. 2 B to S (-O); data not available for the Czech Republic, Estonia, Ireland, Greece, France, Croatia, Cyprus, Luxembourg, Malta, Austria, Romania, Slovenia and the United Kingdom.

(<sup>\*)</sup> Provisional data

	< 25 years	25 - 34	35 - 44	45 - 54	55 - 64	65 years +
Belgium	-2.7	0.6	5.3	7.1	15.4	:
Bulgaria	6.9	13.9	20.0	17.6	6.2	2.0
Czech Republic	11.4	13.8	27.4	23.8	15.1	20.4
Denmark	6.5	10.9	16.1	18.2	16.0	11.0
Spain	10.3	7.7	10.4	16.7	22.7	48.6
France	-3.8	8.4	12.8	18.9	21.1	29.2
Italy	16.0	5.1	5.7	6.4	5.7	:
Cyprus	6.1	-0.6	9.7	23.8	26.3	51.9
Latvia	13.2	16.1	21.0	14.9	13.5	17.1
Lithuania	14.1	17.3	20.0	11.2	10.3	13.8
Hungary	3.7	10.5	19.7	15.8	7.7	22.3
Malta	2.7	9.0	13.1	11.9	7.1	21.1
Netherlands	3.6	1.0	9.5	20.1	21.1	15.9
Poland	7.6	9.0	12.0	6.9	2.3	-13.7
Portugal	8.0	8.3	14.9	23.5	23.2	43.0
Romania(*)	-1.1	1.1	6.7	4.8	2.8	17.5
Slovenia	6.8	7.0	9.7	12.2	6.4	0.0
Slovakia	10.9	12.6	24.2	21.8	17.0	22.1
Finland	4.4	11.4	18.0	19.1	21.3	21.2
Sweden	4.5	8.4	14.8	16.4	16.1	14.3
United Kingdom(*)	4.5	12.7	22.0	27.7	26.9	26.8
Iceland	2.3	8.6	18.4	24.4	21.0	20.1
Norway	1.9	7.8	14.8	18.4	20.8	19.6
Switzerland	3.4	7.1	15.4	22.1	22.9	31.5

Note: For all countries except the Czech Republic: data for enterprises employing 10 or more employees, NACE Rev. 2 B to S (-O); for the Czech Republic: enterprises employing 1 or more employees, NACE Rev. B to S; data not available for Germany, Estonia, Ireland, Greece, Croatia, Luxembourg and Austria.

: data not available

(\*) Provisional data

(\*) Estimated

(\*) Estimated by Eurostat

	Business economy (B to N)	Manufacturing (C)	Electricity, gas, steam and air conditioning supply (D)	Water supply; sewerage, waste management and remediation activities (E)	Construction (F)	Information and communication (J)	Financial and insurance activities (K)	Real estate activities (L)	Professional, scientific and technical activities (M)
Belgium	10.7	6.2	21.8	-2.2	-7.1	11.7	19.9	3.5	15.2
Bulgaria	14.1	24.8	13.8	19.9	-13.0	19.1	21.9	17.0	1.1
Czech Republic	16.2	25.8	14.3	4.7	11.7	32.8	39.8	7.4	25.4
Denmark	16.2	10.7	18.6	1.4	9.5	17.4	19.5	10.6	20.3
Germany <sup>(1)</sup>	25.0	25.0	21.5	5.7	12.9	25.3	28.1	15.9	31.3
Estonia	25.6	30.0	15.0	10.2	15.8	26.4	34.8	14.6	22.6
Spain <sup>(1)</sup>	20.3	19.2	16.1	14.7	3.3	13.2	18.2	22.8	19.6
France <sup>(1)</sup>	12.5	14.4	12.1	3.7	-5.8	11.6	30.6	11.6	22.4
Italy	16.0	16.7	:	:	:	18.9	18.1	:	28.9
Cyprus	21.4	29.0	2.4	-6.2	8.8	13.7	24.8	8.9	33.9
Latvia	15.4	18.6	20.6	18.9	5.3	15.0	25.9	19.2	7.9
Lithuania	17.8	26.0	9.3	14.7	1.6	29.9	38.3	13.9	16.0
Luxembourg	13.9	14.9	-1.3	-4.0	-0.6	17.2	23.0	19.1	25.4
Hungary	14.0	22.0	15.8	1.9	-11.5	20.6	34.2	10.9	17.1
Malta	14.3	21.5	:	7.1	-15.9	20.4	30.4	20.5	25.2
Netherlands	21.6	18.2	15.5	1.1	11.6	17.6	28.8	17.0	23.0
Austria	21.7	21.8	18.4	7.0	6.8	21.9	26.6	25.5	30.8
Poland	15.4	20.7	4.6	-4.3	-14.8	25.9	30.4	7.4	14.9
Portugal	21.1	31.0	7.6	-8.9	-6.2	11.2	23.7	19.2	19.9
Romania <sup>(2)</sup>	7.0	18.6	2.6	-2.0	-20.5	16.8	29.2	10.0	-1.4
Slovenia	8.9	14.8	0.5	-11.1	-22.3	15.8	21.7	-5.5	-1.9
Slovakia	19.8	26.4	8.5	-5.6	2.5	27.6	35.3	26.0	15.9
Finland	17.1	10.2	15.0	1.1	13.4	12.9	33.4	16.0	15.0
Sweden	10.2	4.1	6.9	-2.8	-1.4	11.7	26.9	7.5	14.0
United Kingdom <sup>(1)(3)</sup>	21.8	17.9	20.6	-1.3	16.5	18.5	35.9	22.5	23.2
Iceland	16.4	15.0	13.2	-1.5	-18.8	11.9	39.6	:	:
Norway	17.5	11.7	10.1	-3.9	-4.3	15.2	30.4	20.3	20.8
Switzerland	19.5	17.7	10.4	6.2	6.4	20.9	30.5	20.3	22.7

Note: For all the countries except the Czech Republic: data for enterprises employing 10 or more employees, NACE Rev. 2 B to S (-O); for the Czech Republic: enterprises employing 1 or more employees, NACE Rev. B to S; data not available for Ireland, Greece and Croatia.

: Not available

<sup>(1)</sup> Provisional data

<sup>(2)</sup> Estimated

<sup>(3)</sup> Estimated by Eurostat



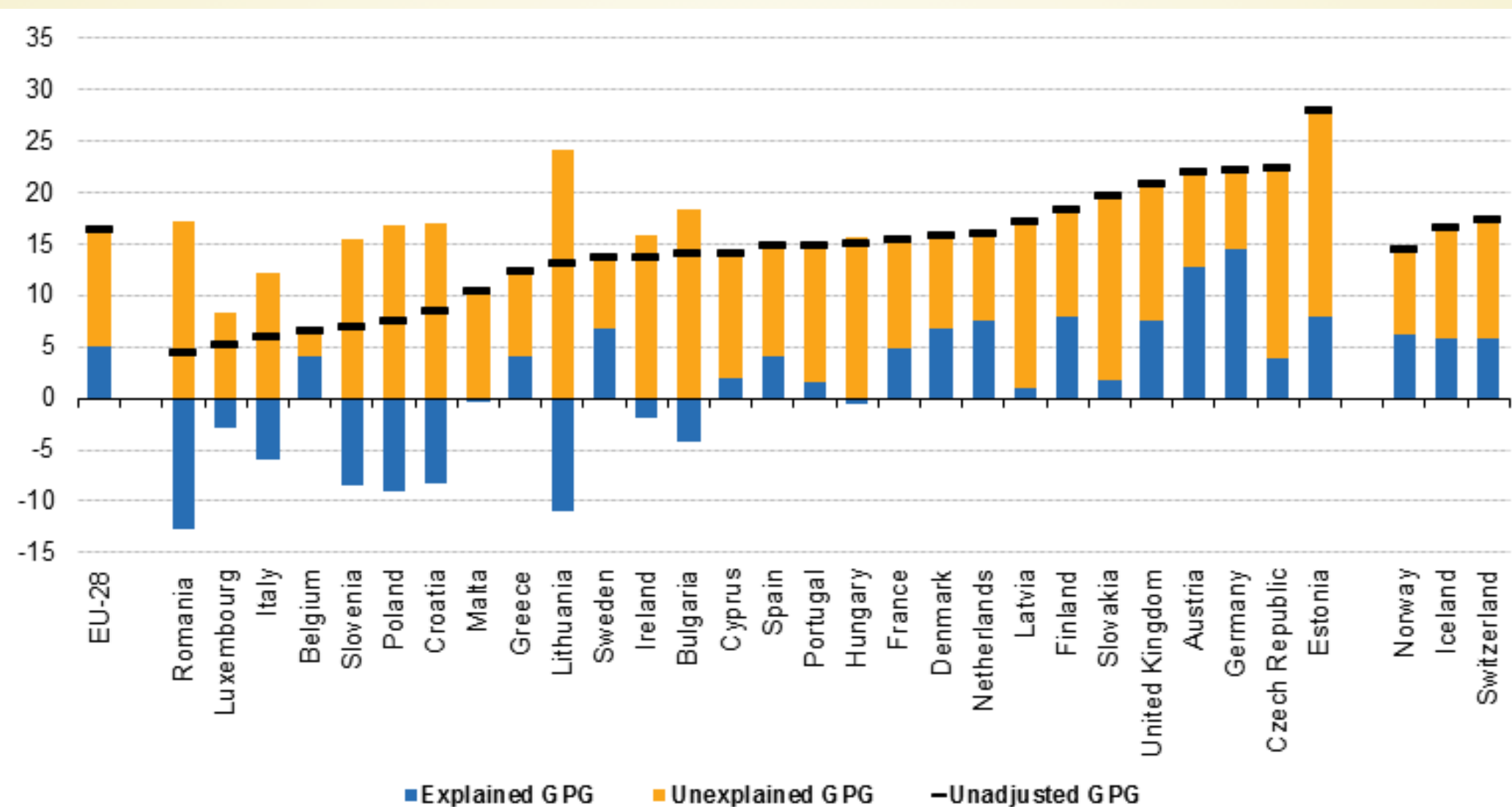
	Public	Private
Belgium	0.1	9.2
Bulgaria	20.6	12.3
Czech Republic	20.5	22.8
Denmark	11.6	15.6
Germany	13.0	24.0
Spain	13.0	19.0
Italy	4.4	17.9
Cyprus	-6.6	23.0
Latvia	16.9	14.1
Lithuania	13.7	17.6
Hungary	11.8	15.0
Netherlands	12.7	21.6
Poland	2.8	16.1
Portugal	13.4	22.6
Romania	9.9	6.8
Slovenia	11.3	7.9
Slovakia	12.7	20.4
Finland	17.7	17.2
Sweden	10.2	11.8
United Kingdom	24.4	22.2
Iceland	12.2	16.4
Norway	8.5	17.8
Switzerland	17.5	17.7

Note: For all countries except the Czech Republic: data for enterprises employing 10 or more employees, NACE Rev. 2 B to S (-O); for the Czech Republic: enterprises employing 1 or more employees, NACE Rev. B to S; data not available for Estonia Ireland, Greece, France Croatia, Luxembourg, Malta and Austria.

(<sup>1</sup>) Provisional data

(<sup>2</sup>) Estimated

(<sup>3</sup>) Estimated by Eurostat



Note: The results are based on Eurostat's statistical working paper that will be published in the first half of 2018; The study uses 2014 data of the Structure of Earnings Survey data are used.

# IMPLEMENTATION IN EXCEL

Download the source data for this report [here](#). Try to replicate and/or improve the charts using Excel.

# IMPLEMENTATION IN R

## R Script

(I will update this script online after the class with all final changes that we make to the script in class.)