ECS7024 Statistics for Artificial Intelligence and Data Science

Topic 2: Bar Chart, Proportions and Distribution

William Marsh

Outline

- Aims: Understand the use of distributions of values (by category)

- Types of data: categorical, continuous
- Bar charts
- Proportions
- Distributions: showing proportion of whole

Types of Data

Categorical Data

- Categorical data
  - Values from (a finite) set
  - Also called ‘factors’, ‘nominal data’ or ‘enumerated
  - Often ‘string’; can be integer ‘code’
- Binary data
  - Values are true and false (many representations)
- Ordinal data
  - Categories are ordered

Continuous Data

- Continuous data
  - Any **numeric** value in an interval
  - *Mathematical abstraction: never completely true*
  - Practical view: too many distinct (ordered) values

- Discrete (continuous)
  - Integer values only (e.g. counts)
  - Warning: ‘discrete’ can be synonym for ‘categorical’

Example: London Population

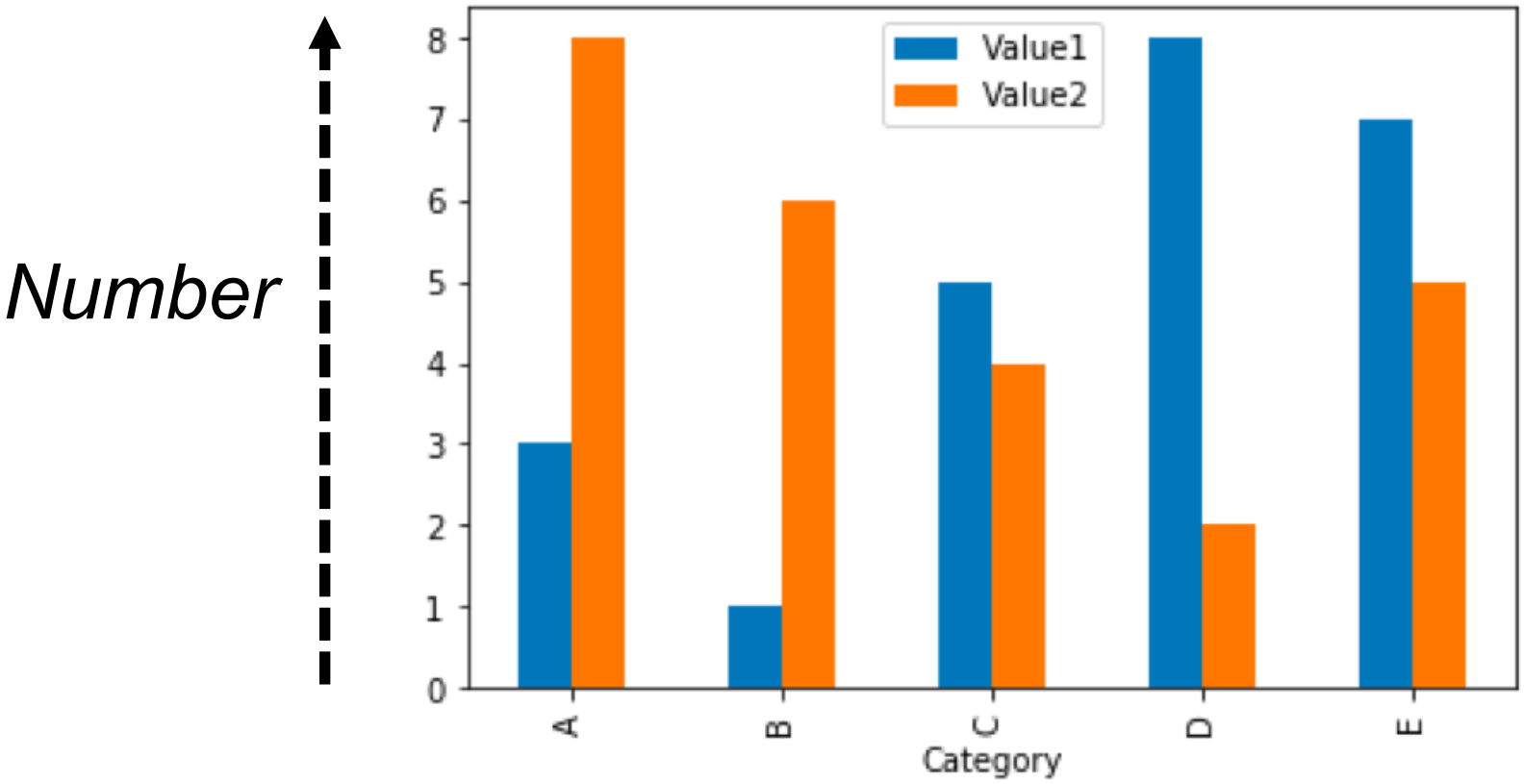
- What type of data

Variable	Description	Type
Area	Includes London Boroughs	Categorical
Age	The ages in a number of bands	Ordinal
Sex	Males and Females	Binary
Usual Residents	An integer	Continuous (discrete)
BirthCountry	A country	Categorical
BirthRegion	A region e.g. Africa or Europe	Categorical

Bar Charts for Discrete Data

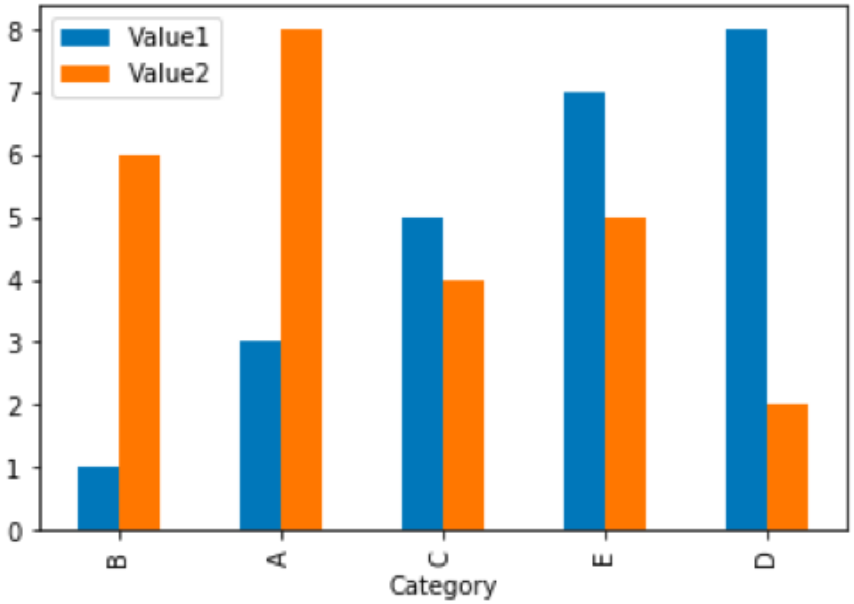
Bar Charts

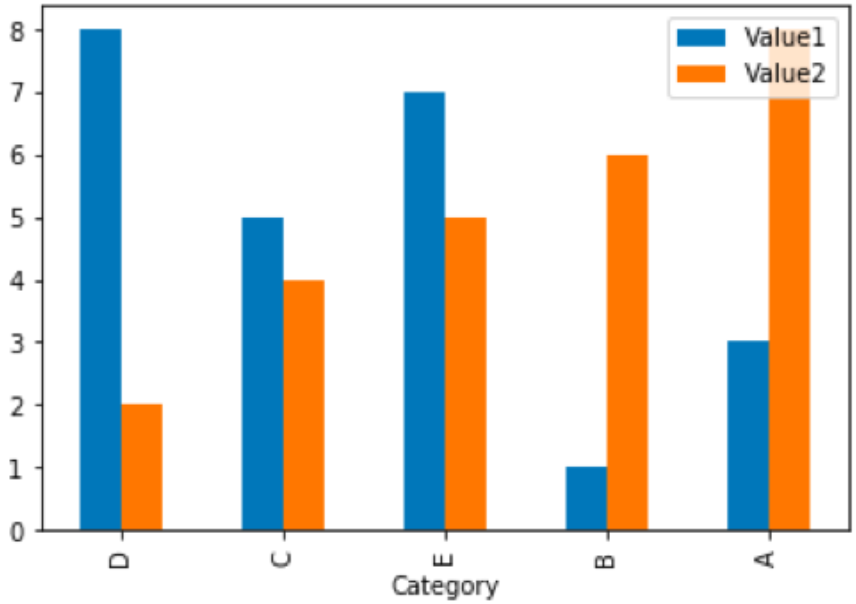
- Bar chart shows values in categories



Same Data

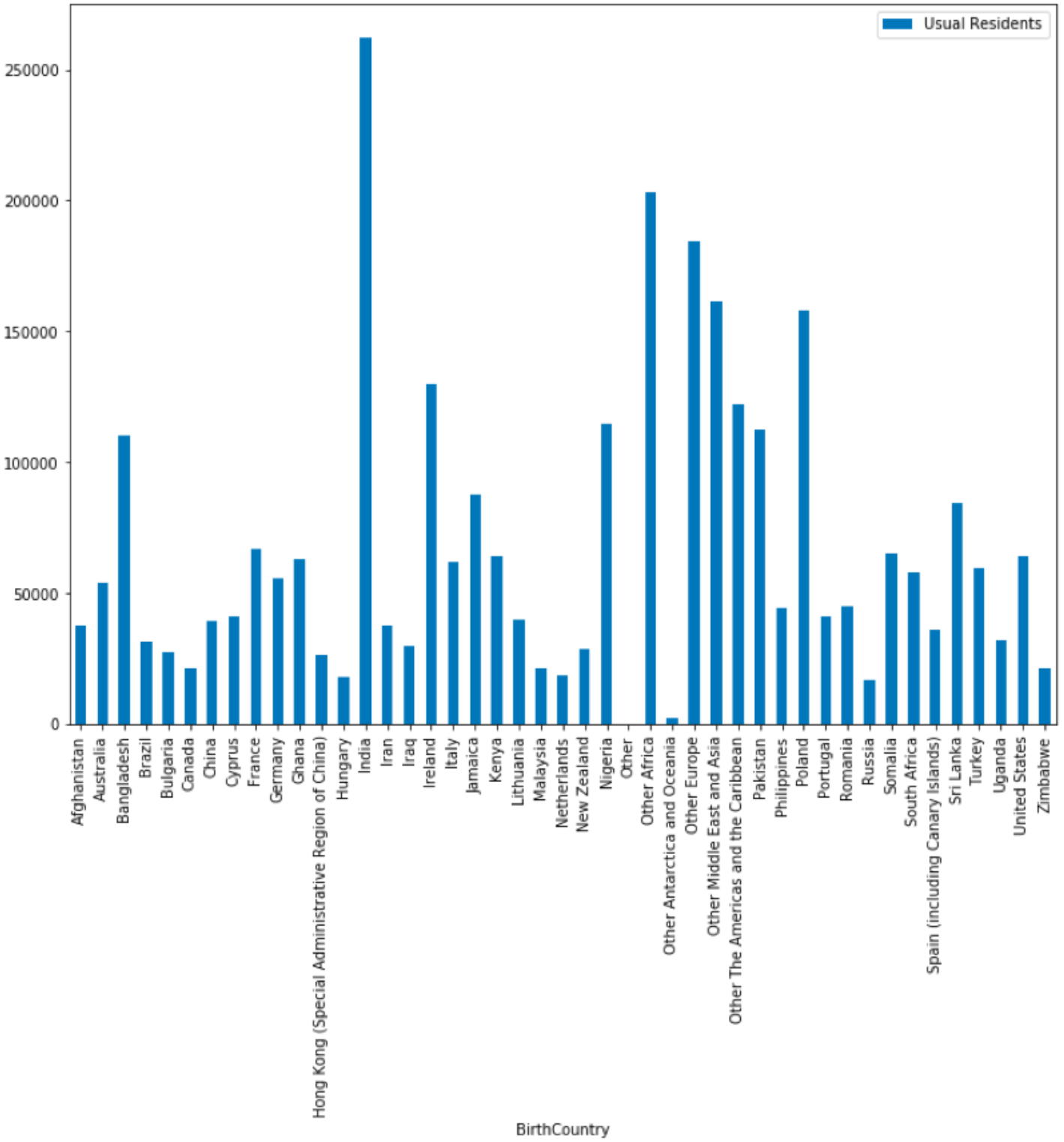
- Changing order of categories makes no difference





Example

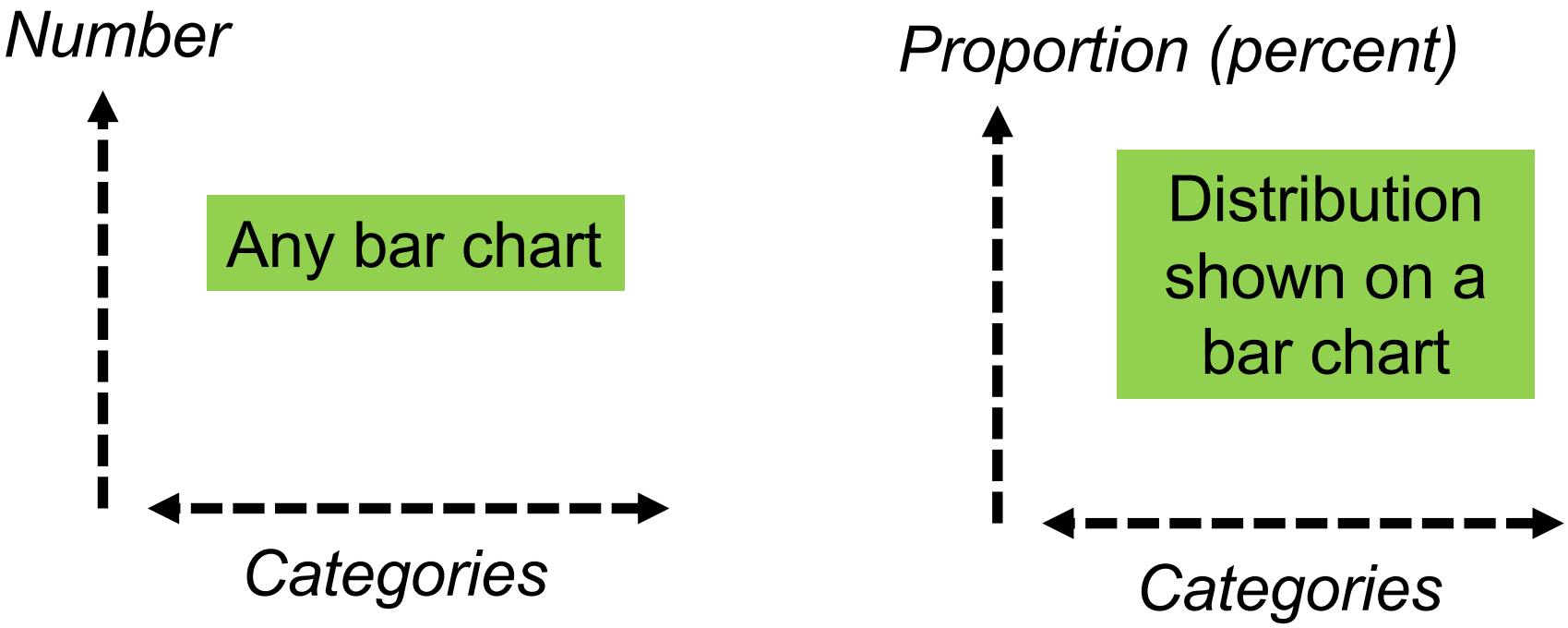
- Number of residents born outside UK
- All London



(Probability) Distribution

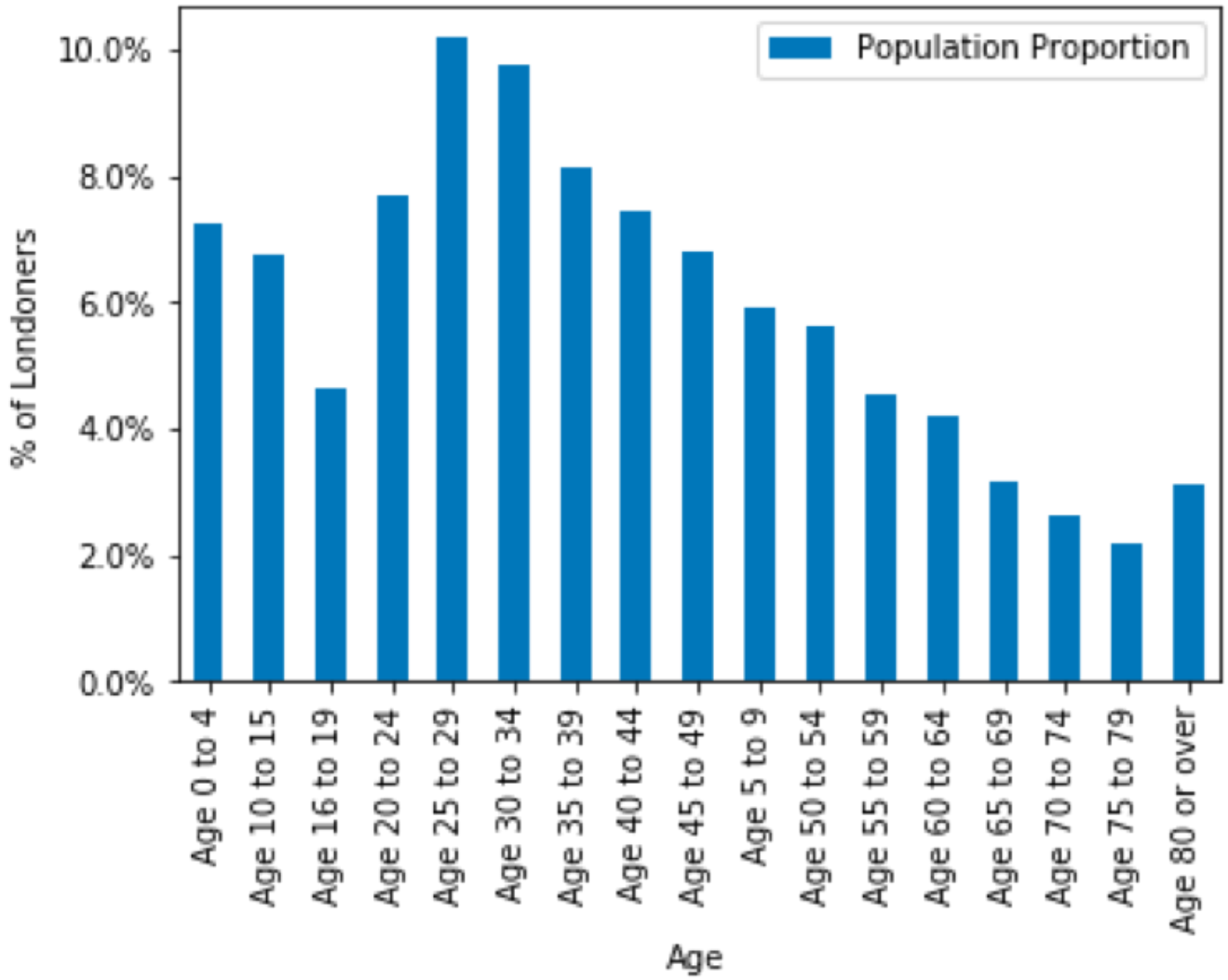
Distribution

- Show composition of whole in parts
- Can be interpreted as a probability
- Proportions combine to 100%



Distribution Example

- How would you describe age structure of London residents?



Quiz

Asking Questions

A study skill

A skill for life

Every lecture will have a ‘learning reflection’ slide

Asking Questions

Why ask questions

- Your aim is to achieve a deep understanding
- Being active
- Understanding arrives in stages
  - Confusion
  - Misconceptions
  - Partial understanding
- Thinking about what you do not understand helps to achieve understanding

How to ask questions

- It's hard. Why?
- Think about where you understanding stops
- Use an example

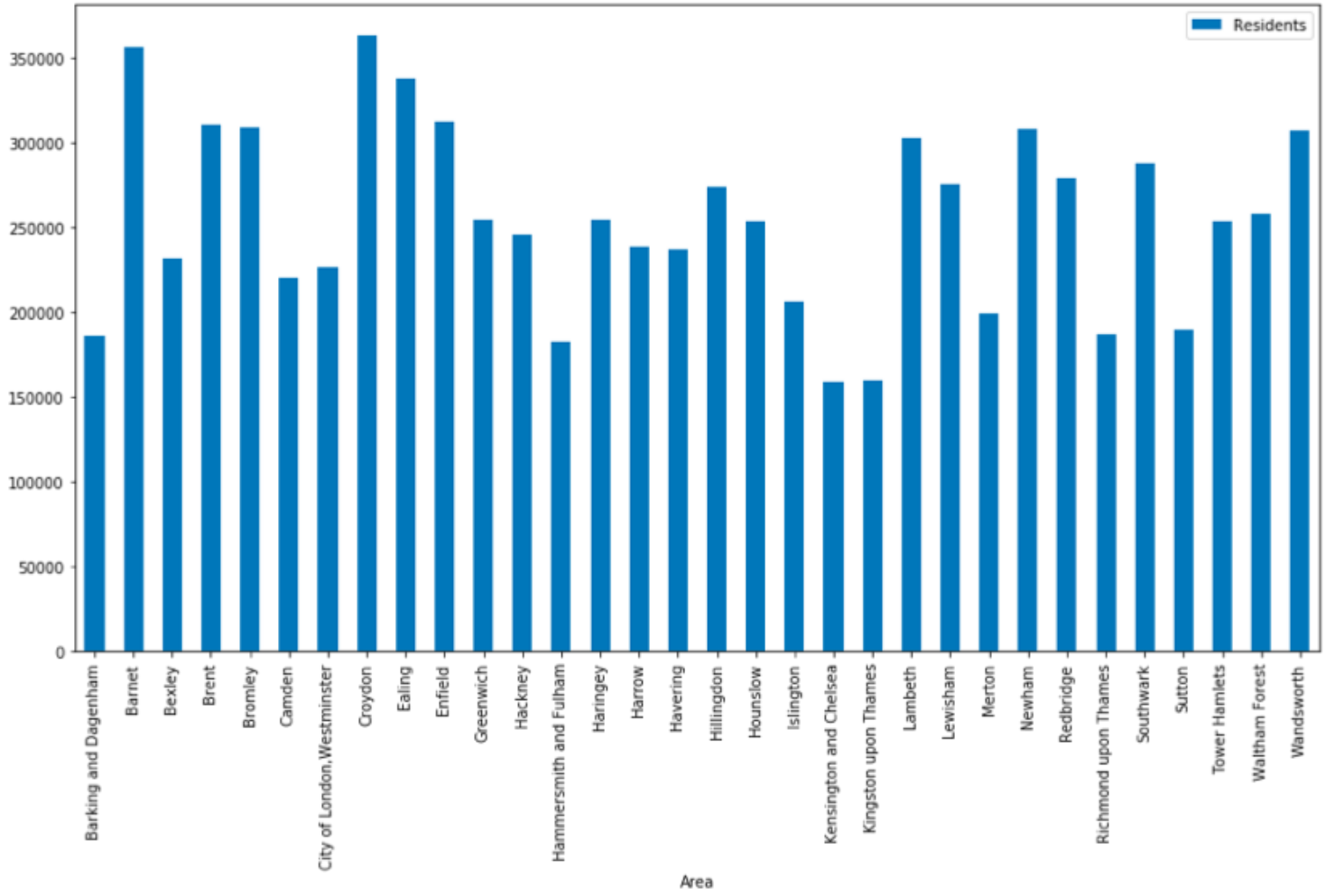
- Practically
  - During practical sessions
  - On the QM+ forum, or by email, at any time

Thinking About Proportions

How do we compare the population of people with a particular birth country in different London Boroughs?

Population of London Boroughs

- Borough vary in size
  - Brent about twice Kensington and Chelsea



Where are Londoners Born in Ireland?

- Two questions we could ask:
  - [**Q1 London proportion**] What proportion of the total Irish population in London lives in each Borough?
  - [**Q2: Borough proportion**] What proportion of the population of each Borough was born in Ireland?

- If those born in Ireland uniformly distributed across London, then we expect more to be in Brent (a large Borough) than Kensington (a smaller one)

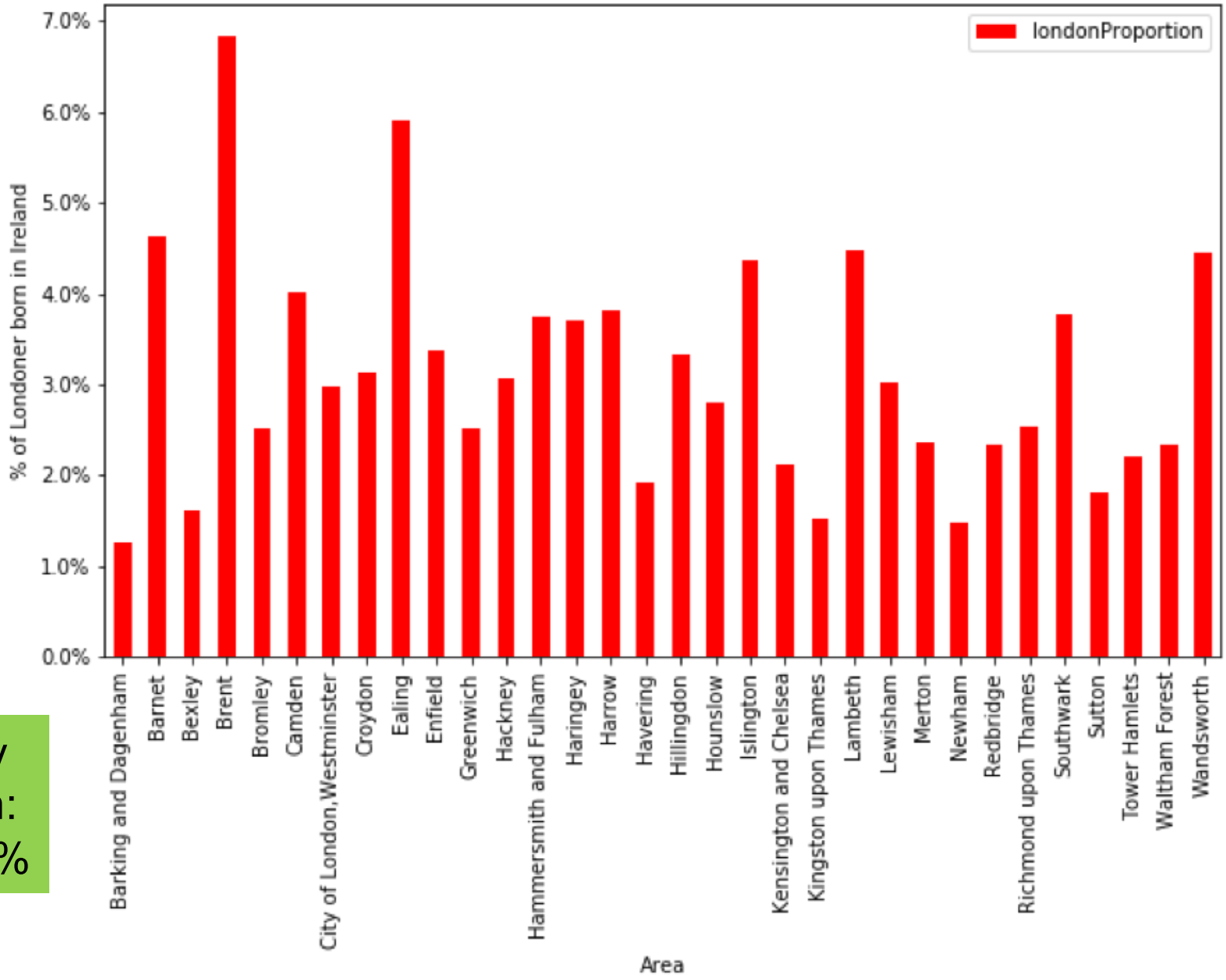
We need to be very precise in the question. MUST write!

Proportions Require Division

Barking, Age 5 to 9, Females has 28 with BirthCountry == Ireland

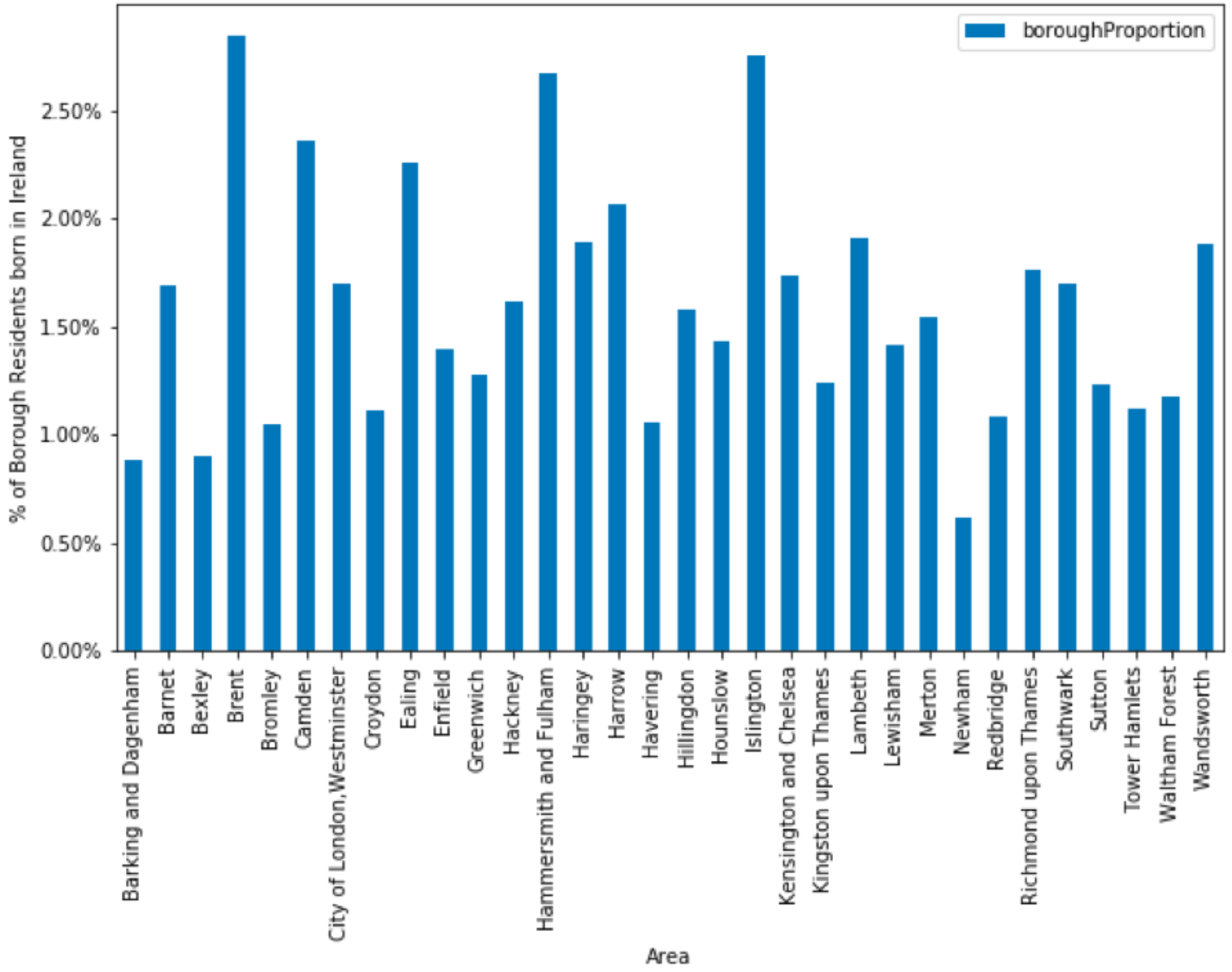
- [Q1 London proportion]
  - 28 / all Londoners born in Ireland
- [Q2: Borough proportion]
  - 28 / all Barking residents

Q1: % Irish-born Londoners in each Borough



Probability distribution: total is 100%

Q2: % Borough Population born in Ireland



Sources of Variability

- [Q1 London proportion]
  - Values depend both on size of Borough and where those born in Ireland like
  - Two sources of variability confused

- [Q2: Borough proportion]
  - Eliminate variability of Borough size
  - Often preferred

Quiz

Summary

- Data can be categorical or continuous
  - BirthCountry is categorical
  - Age is ordinal – categories with an order

- Distribution of values in categories can be shown on a bar chart
  - If the distribution shows part of a whole, it can be interpreted as ‘probability distribution’

- To make comparisons, we often need proportions. Take care!