

# Quality assurance of non-official statistical sources for SDG

**Atanaska Nikolova**

Senior analyst  
UK SDG data

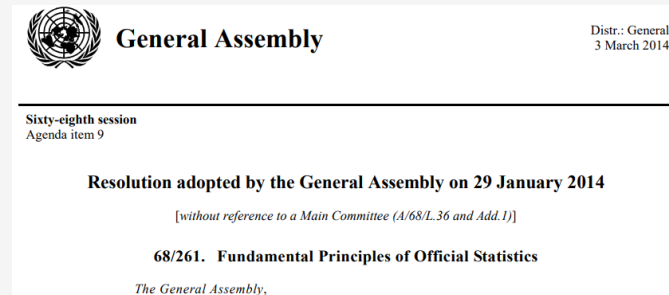
06 April 2022



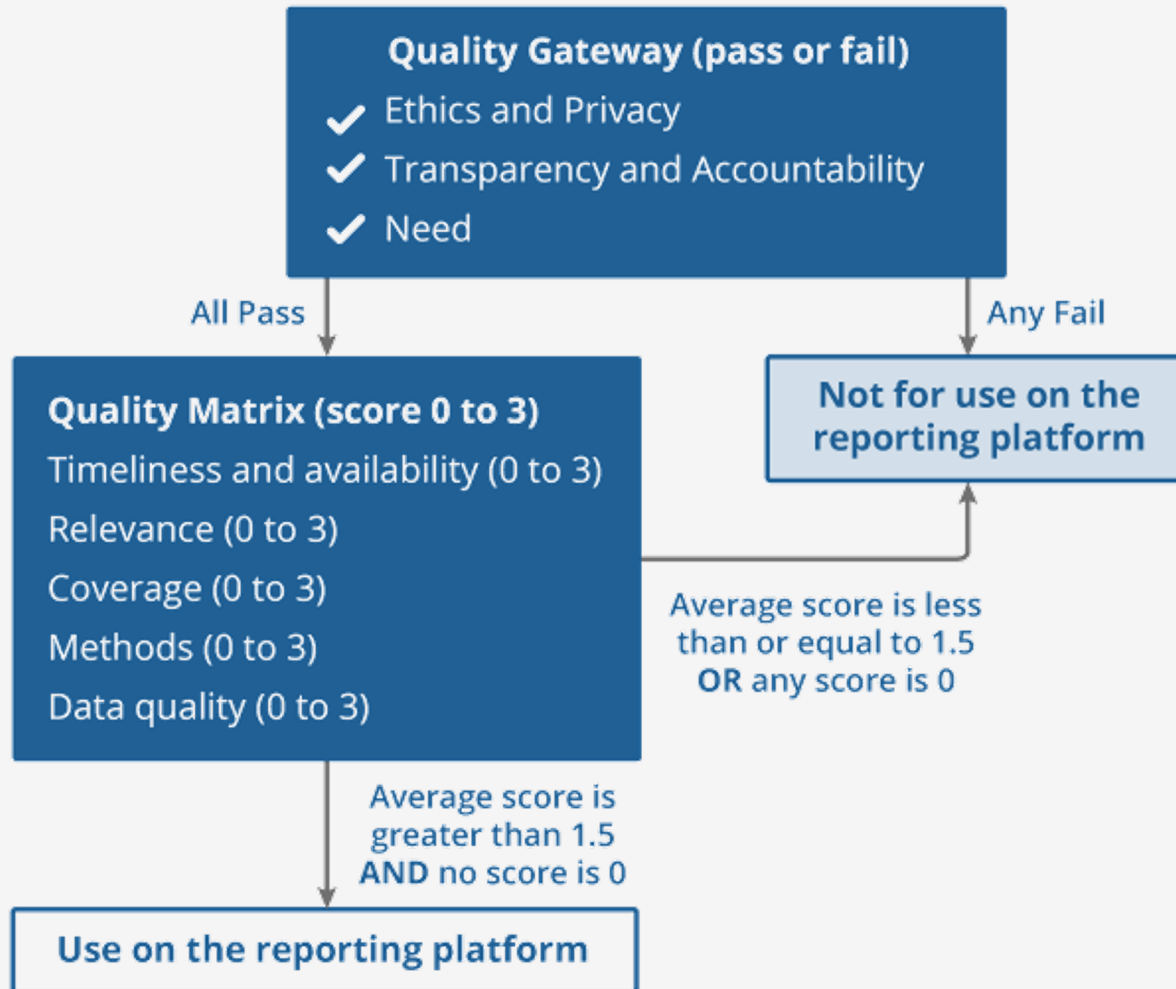
# What are non-official statistical sources?

- An output that does not come from a UK governmental department or government-related body, local or devolved authority, or an official international reporting body (e.g. Eurostat, UN custodian, OECD).
- Examples: charities, businesses, academia, citizen data

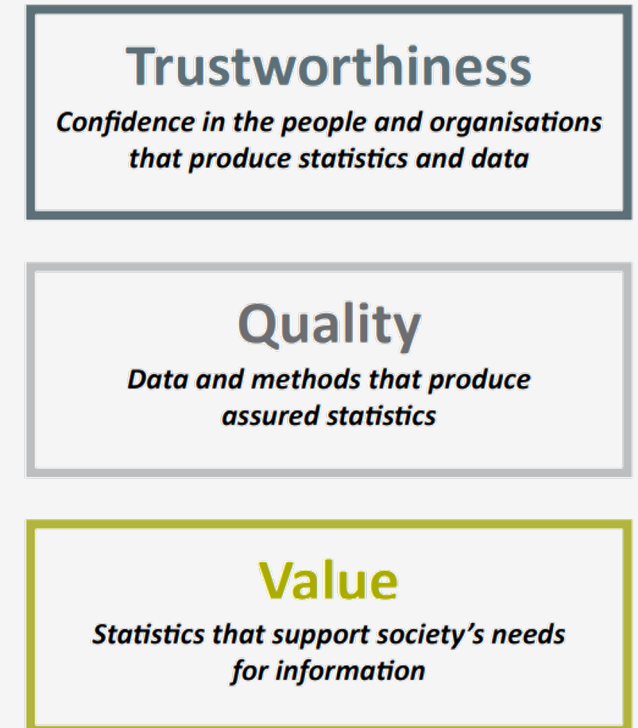
[UK Code of Practice for Statistics](#)



# Protocol for assessment of non-official sources



Official publication (revised March 2022):  
[UK Sustainable Development Goals: use of non-official sources](#)



# Case study – Marine Conservation Society

- Beach litter citizen data for 14.1.1 (b)
- Organised beach cleans by volunteers
- Survey a 100 metre stretch along strandline – record all items found
- Upload to website – national database of beach litter data since 1994

- ✓ Ethics and Privacy
- ✓ Transparency and Accountability
- ✓ Need

# Assessment template – quality gateway

Criterion	Criteria		Assessment	
	Pass	Fail	Source score	Explanation
<b>Ethics and Privacy</b>	No ethical concerns, or any concerns are fully documented and actions are in place to minimise identified risks. Fully compliant with all parts of the Code principle T1 (honesty and integrity) and T6 (data governance). Use is in line with the terms and conditions of the source. Privacy policy is compliant with the General Data Protection Regulation (GDPR) for the UK and the Data Protection Act 2018.	Significant ethical concerns without any mitigations or considerations. Not compliant with all parts of principle T1 and T6 of the Code. The source's terms and conditions prevent use of the data as required. Not compliant with GDPR and/or the Data Protection Act 2018. Any of these conditions would lead to a fail.	Pass	No personal data involved and summary statistics of the data are publicly available
<b>Transparency and Accountability</b>	Source meets principles T4.1 and T4.5 of the Code; ensuring processes for all parts of the data journey are transparent. If metadata information is not already in the public domain, permission must be granted to place this in the public domain.	Source does not meet T4.1 and T4.5 of the Code. Source may not be fully transparent about any data quality issues and there is no metadata available.	Pass	MCS have information about how the GBBC is run. Methodology of analysis done by ONS will be publicly available
<b>Need</b>	There is a clear identified need for the source, either due to a headline data gap or a priority disaggregation gap. Alternatively, the source may be a better fit than a source already being used on the Platform. Sources suitable for non-priority disaggregations may be considered if all priority gaps have been filled and there is sufficient resource for additional assessment.	The proposed source does not fill a headline gap. The source is unlikely to improve on information already on the Platform.	Pass	Fills a data gap

# Assessment template – matrix scoring

**Quality Matrix (score 0 to 3)**  
 Timeliness and availability (0 to 3)  
 Relevance (0 to 3)  
 Coverage (0 to 3)  
 Methods (0 to 3)  
 Data quality (0 to 3)

	3 - high	2 - medium	1 - low	0 - not acceptable	Score	Explanation
<b>Timeliness and ongoing availability</b>	Source is sufficiently current to be informative, with a time series from at least 2015 and no time lag greater than 15 months for annual data, or 6 months for monthly data. A wider gap of 2 years is acceptable when the impact of any statistical change may take longer to be observed, e.g. for some environmental statistics. No gaps (missing data) in the time series. The source is reasonably expected to be regularly updated and available in the future. There must be a record of previous data points (i.e. the source provides a time series)	Source is sufficiently up to date to be informative, with a time lag no greater than 2 years (3 years for statistical changes that may take longer to be observed, such as some environmental statistics). There are no gaps in the time series. There must be a record of previous data points (i.e. the source provides a time series). New timely sources without previous data points that are expected to be updated and available in the future would be included.	Source is older than 2 years (3 years for statistical changes that may take longer to be observed, such as some environmental statistics), but is still meaningful in the social, environmental, or economic context of the indicator. The time series may have gaps, or only one data point has been produced.	Source is too old to be meaningful, with the latest data point(s) before 2015, or has too great a time lag, and/or no reasonable expectation of future updates. The source does not provide access to existing historic data (time series)	3	Time series provided, and agreement on annual supply of data is in place

# Assessment template – matrix scoring

Quality Matrix (score 0 to 3)  
 Timeliness and availability (0 to 3)  
 Relevance (0 to 3)  
 Coverage (0 to 3)  
 Methods (0 to 3)  
 Data quality (0 to 3)




	3 - high	2 - medium	1 - low	0 - not acceptable	Score	Explanation
<b>Relevance</b>	Close match with UN SDG metadata, or gives more detail than the metadata requires. Fully compliant with the Code principles Q1.1 and Q1.5 on suitable data sources. Disaggregations specified in the SDG indicator title are reported as a minimum, potentially supported by additional Inclusive Data Charter (IDC) disaggregations.	Partial match with UN metadata and/or disaggregation in indicator title. Fully compliant with the Code principles Q1.1 and Q1.5 May enable reporting of additional IDC disaggregations.	Does not fully report the indicator but is an appropriate proxy relevant to the UK national context. Fully compliant with the Code principles Q1.1 and Q1.5. May enable reporting of relevant disaggregations.	Does not align well with the metadata for the indicator and/or provides no appropriate proxy for headline or relevant disaggregation gaps.	3	There are minor differences to the UN methodology, which suggests use of a geospatial model to calculate density per km <sup>2</sup> , whilst this data gives median count of litter items
<b>Coverage</b>	Data robustly and reliably measures the entire UK population and/or the entire UK geography (as appropriate to the indicator).	Metadata is clear and transparent about the identifiable population covered or not covered. The population is a suitable representation for reporting the indicator, though may not cover all UK nations.	Metadata is available on the identifiable population covered, but the limitations of partial coverage may not be fully considered by the source producer. The population is adequate for reporting the indicator, even if this could potentially be improved.	Metadata is unclear about the identifiable population and/or coverage is not specified. The population may be of too limited representation or the sample too small to be appropriate for reporting the indicator.	2	Northern Ireland is included but has a small sample size. Some countries are overrepresented in the UK figure (estimate does not control for number of beaches given the sum length of beaches for

# Assessment template – matrix scoring

	3 - high	2 - medium	1 - low	0 - not acceptable	Score	Explanation
<b>Methods</b>	Internationally comparable in line with the UN metadata, the methods are appropriately applied, fit-for-purpose and transparently described. Fully compliant with the Code principle Q2 (sound methods) by using the best available methods openly.	The methods are appropriately applied, and well described. Largely compliant (compliant with half or more of the principles) with the Code principles under Q2, including Q2.3 and Q2.4.	Methodology described in detail and transparent. Methods are justifiable, but might be lacking scientific proof, or alternative methods may be available that might produce improved results. Largely compliant with the Code principles under Q2.	Methodology is neither described nor justified.	2	Data are not comparable in line with UN metadata though <i>are</i> internationally comparable in terms of pre-calculation methods. Compliant with Q2 principles. Methodology will be published in full on e.g. Github once data are on the platform, and key limitations are stated in the NatCap publication
<b>Data quality</b>	Data validation procedures are outlined. It is clear how the data were collected and pre-processed (if relevant). Largely compliant with standards in the Code principle Q3 (assured quality), specifically outlining aspects of accuracy and reliability	Compliant with principle Q3.2 from the Code - transparency about the quality assurance approach taken throughout the preparation of the statistics. Any issues with quality of the data and statistics are transparently outlined	Some basic checks have been conducted, such that accuracy and reliability of the data source can be established, but no formal quality assurance available.	No information on data quality or quality assurance of the statistics	3	MCS clean the data that is sent to them, and further checks and processing occurs in ONS. See comments in Methods above.
<b>Overall comments and average score</b>	Data passes the Non-official protocol before further processing in ONS, but score is improved through the extra work.				2.60	Pass



# Current non-official sources examples

- Indicator 8.10.1 – ATM numbers from LINK 
- Indicator 6.6.1 – Ground water levels from British Geological survey
- Indicator 3.1.1 – Maternal mortality from The National Perinatal Epidemiology Unit and MBRRACE charity 
- Indicator 12.3.1 – Food waste per capita from WRAP charity 

# Thank you!

Contact: [SustainableDevelopment@ons.gov.uk](mailto:SustainableDevelopment@ons.gov.uk)

Website: [sdgdata.gov.uk](https://sdgdata.gov.uk)

Publication: [UK Sustainable Development Goals: use of non-official sources](#)

# UK 14.1.1b MCS Beach litter data

**Emma Wood**

Senior analyst, UK SDG data

April 2022





We're

**#BeachCleanReady**

are you?

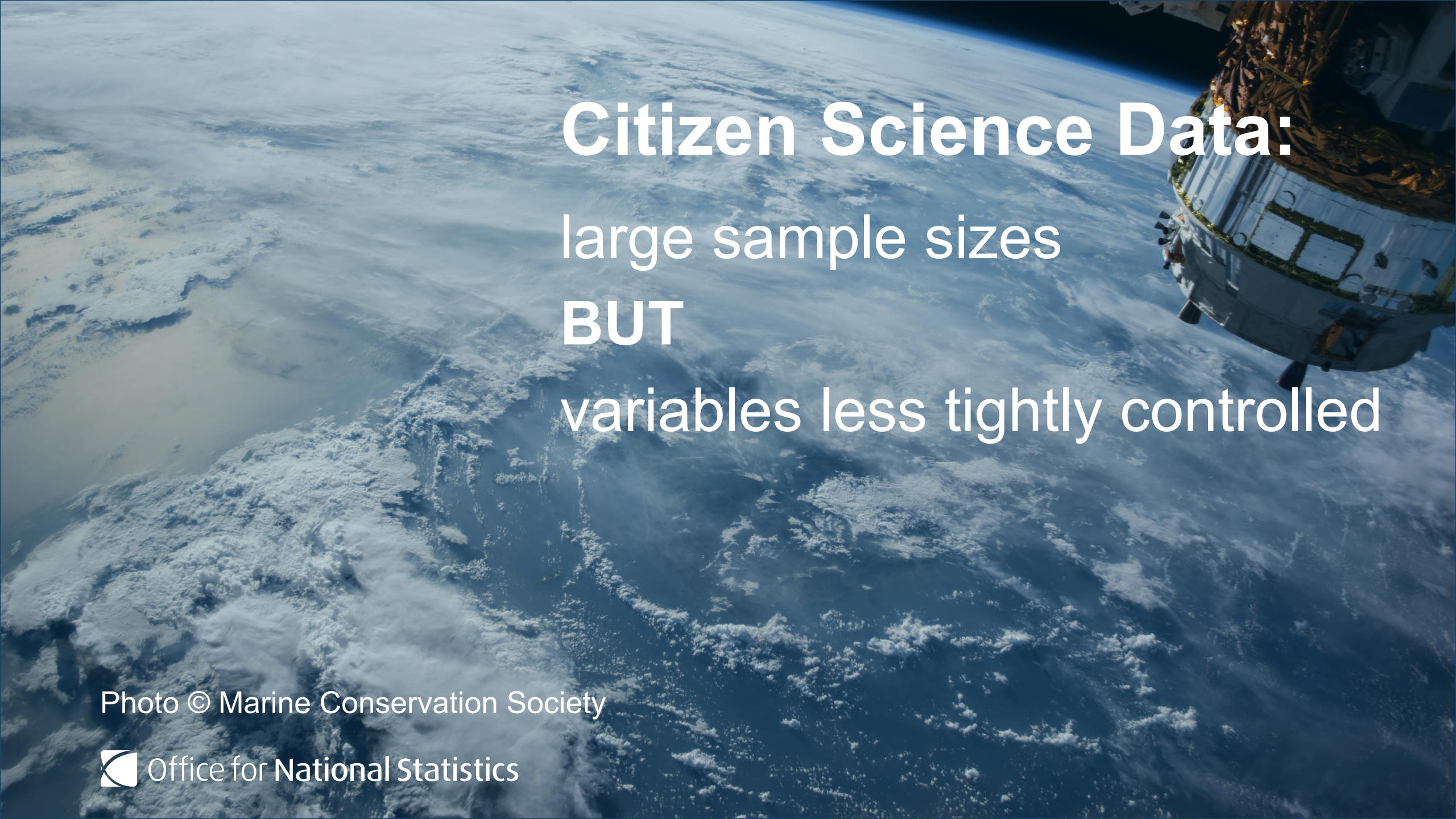
MARINE  
CONSERVATION  
SOCIETY

Great British  
Beach Clean

Friday 17 – Sunday 26 September

# Process

- Met with MCS to discuss methods
- Ingested raw data
- Assessed the data
- Produced estimates and wrote up methods
- To be published in the ONS Natural Capital accounts



**Citizen Science Data:**  
large sample sizes  
**BUT**  
variables less tightly controlled

Photo © Marine Conservation Society

 Office for National Statistics

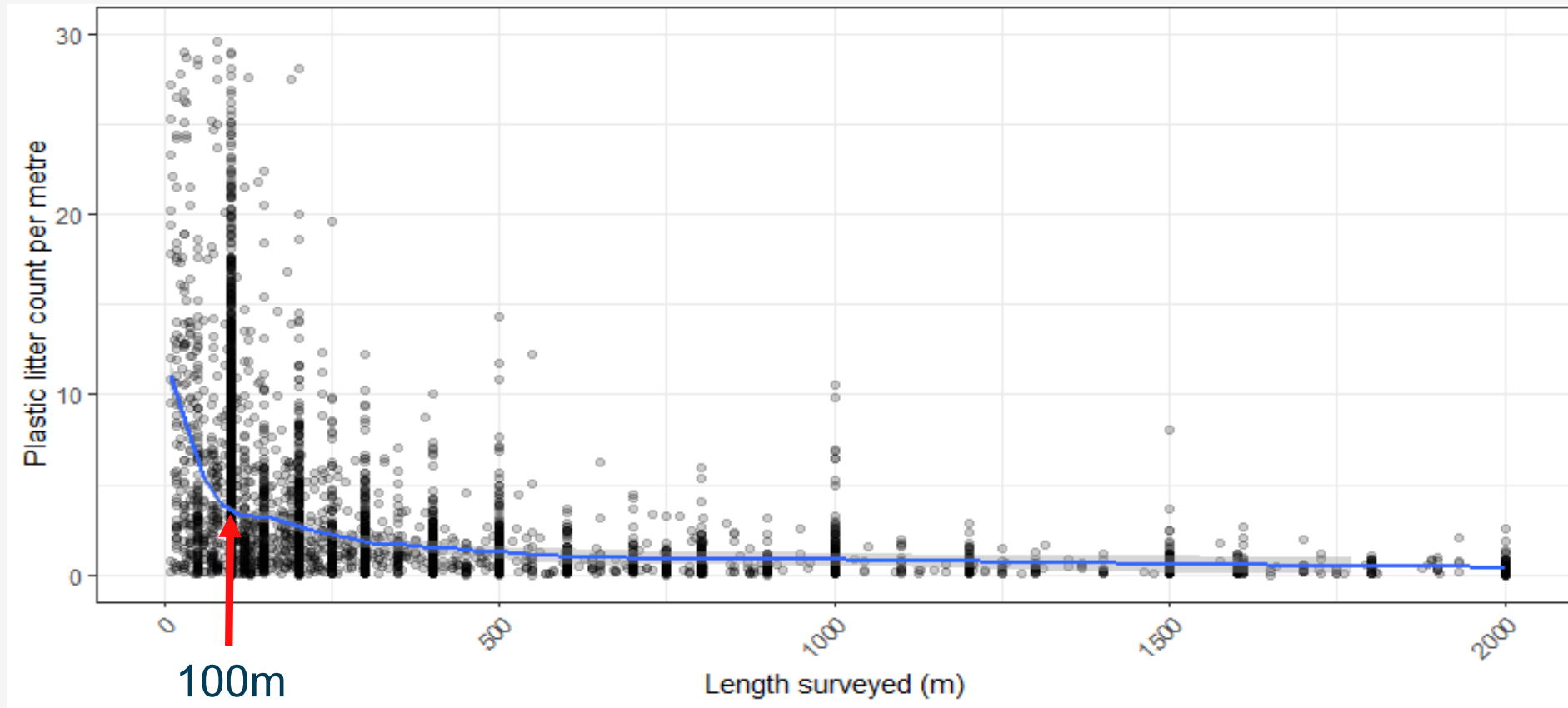
# What is driving trends in litter density estimate?

- Actual change in litter density  
--- OR ---
- Length of beach surveyed
- Number of volunteers
- Composition of beaches
- Weather etc.



# Length surveyed impacts litter density estimate

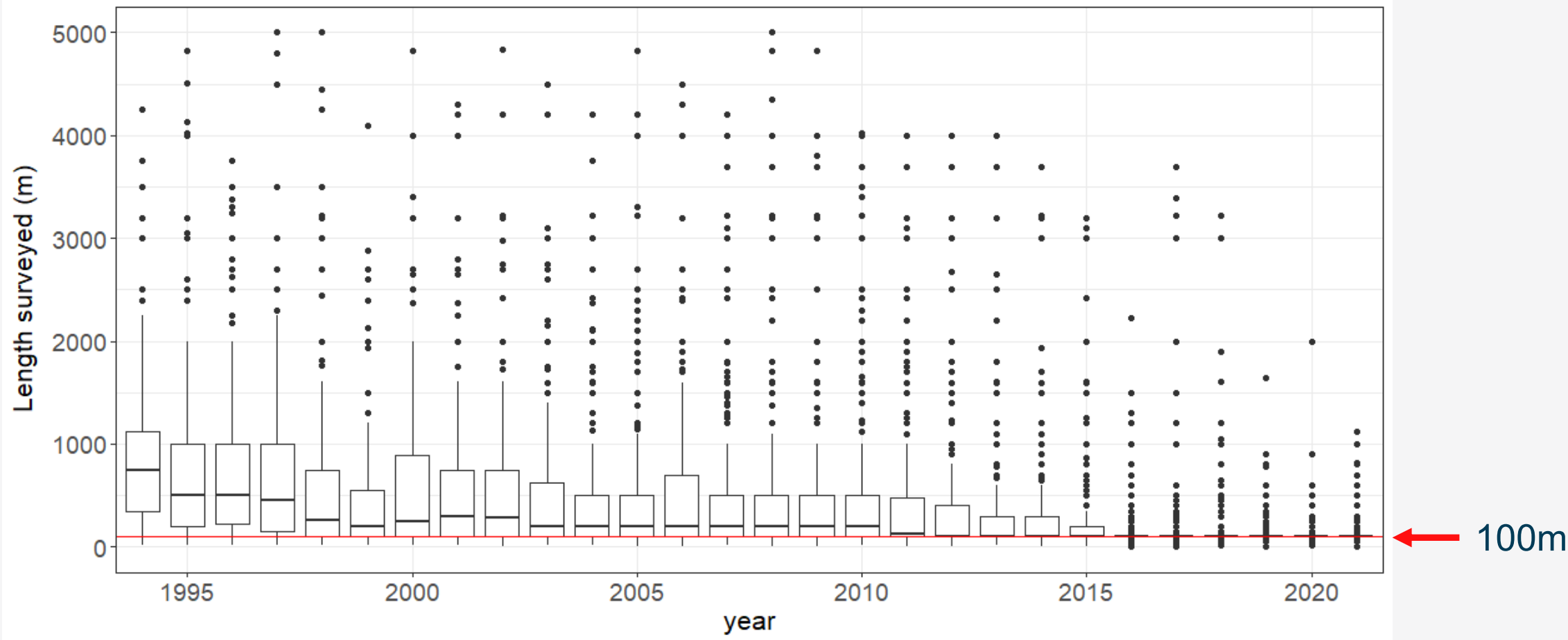
After adjustment for length, shorter stretches have higher litter densities





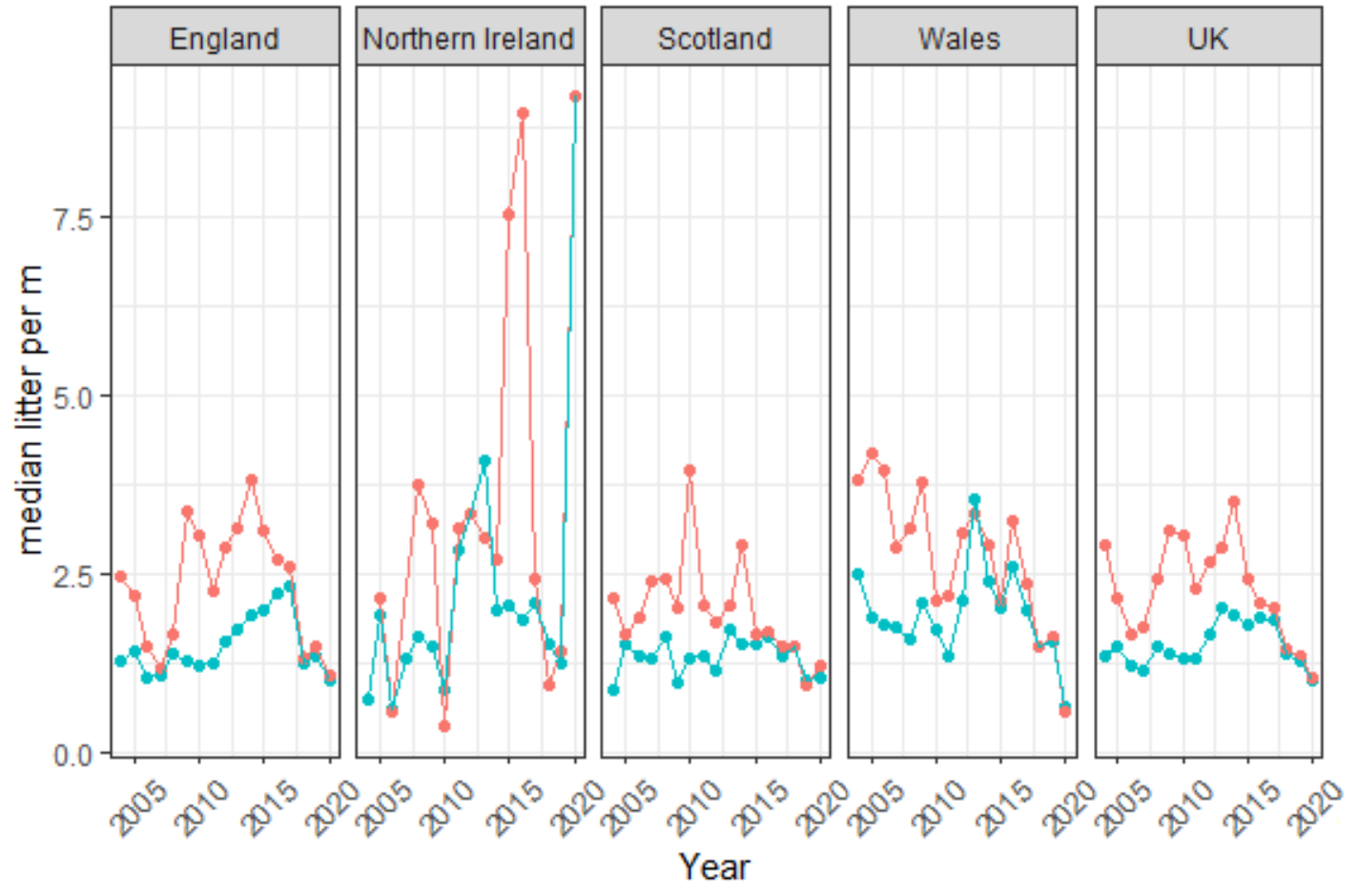
# Trend in length of beach surveyed, UK

Stretch surveyed was longer in earlier years



# Non-standard-length surveys impact the trend

Increase in litter count from 2011 is misleading



dataset

- surveys of 100m
- surveys of any length

The trend in median litter density is affected by temporal changes in the length of beach surveyed.

# Adjustments made

- 100m surveys only
- One survey per beach per year
- Median used, not mean (skewed data with outliers)
- Quality notes (e.g.):
  - 2020 – low volunteer numbers so not comparable with other years
  - Counts not comparable across countries (average volunteer counts differ)
  - Can't control for length of time since beach was last cleaned

# Thank you!

Contact: [SustainableDevelopment@ons.gov.uk](mailto:SustainableDevelopment@ons.gov.uk)

Website: [sdgdata.gov.uk](https://sdgdata.gov.uk)

Publication: [UK Sustainable Development Goals: use of non-official sources](#)