

REPORT OF A BREACH OF THE CODE OF PRACTICE FOR STATISTICS



1. Core Information

Title and link to statistical output	FinBins – granular classification of the UK’s financial sector https://datasciencecampus.ons.gov.uk/finbins-granular-classification-of-the-uks-financial-sector/
Name of producer organisation	Data Science Campus, Office for National Statistics
Name and contact details of person dealing with report	Gareth Pryce, ONS Data Science Campus
Link to published statement about the breach (if relevant)	N/A
Date of breach report	5 December 2018

2. Circumstances of breach

Relevant principle(s) and practice(s)	Principle T6: Data Governance, Practice T6.4 “Appropriate disclosure control methods should be applied before releasing statistics and data”.
Date of occurrence of breach	29 November 2018
<p>A project report, describing findings and methods used, was prepared by the person in charge of delivering the project. After a review, a suggestion was made that the input data should be described in more detail for completeness of the report. A chart showing the pairwise distribution of the most discriminative features (considered good practice in explorative analysis of data) was included in the report. Although the points on the graph do not have direct mapping to individual companies, a single company and some approximate values of the features can be read from the graph, making it disclosive. This wasn’t identified by the author of the report.</p> <p>The report went through several review cycles but the issue in the graph wasn’t identified. Subsequently a blog post and a Twitter message, containing the graph, were derived from the report.</p> <p>After the blog post and report were published on the Data Science Campus website on 29 November 2018, the issue was raised by an ONS employee and a user on Twitter. The graph, and associated tweet, were removed from the website and Twitter for investigations.</p>	

3. Impact of the breach

<p>According to analytics, the blog and report were viewed 72 times on the Data Science Campus website, from the time of publication on Thursday 29th November, 12pm, to the time they were removed from the website. Removal of the content was completed by 8:30am on Friday 30 November. As stated above, one member of staff at ONS raised a concern about the graph directly with the Data Science Campus and one user on Twitter replied to a Data Science Campus tweet that contained a screenshot of the graph. No further activity or comment has been identified, and there are no other known impacts.</p>

4. Corrective actions (taken or planned) to prevent re-occurrence

Steps for replacing the graph with one containing only aggregate information are being undertaken. The graph will be permanently removed from the report altogether if a replacement is not suitable. This replacement will be reviewed by a lead data scientist within the campus, and statistical disclosure experts in ONS.

Further reports to be published will include a formal sign off, with a step included to ensure any statistical disclosure issues are identified prior to the report being made live.

In the longer term, we will undertake a review of processes over the next couple of months to ensure that consistent quality assurance procedures are followed for every project to prevent such issues in the future. We will also instigate a formal sign off for all published articles on the website, that will include a check for any statistical disclosure issues.

We will also arrange training for colleagues on statistical disclosure control and data protection, and the principles related to these areas in the Code of Practice for Statistics to reinforce responsibilities with data when undertaking projects.