

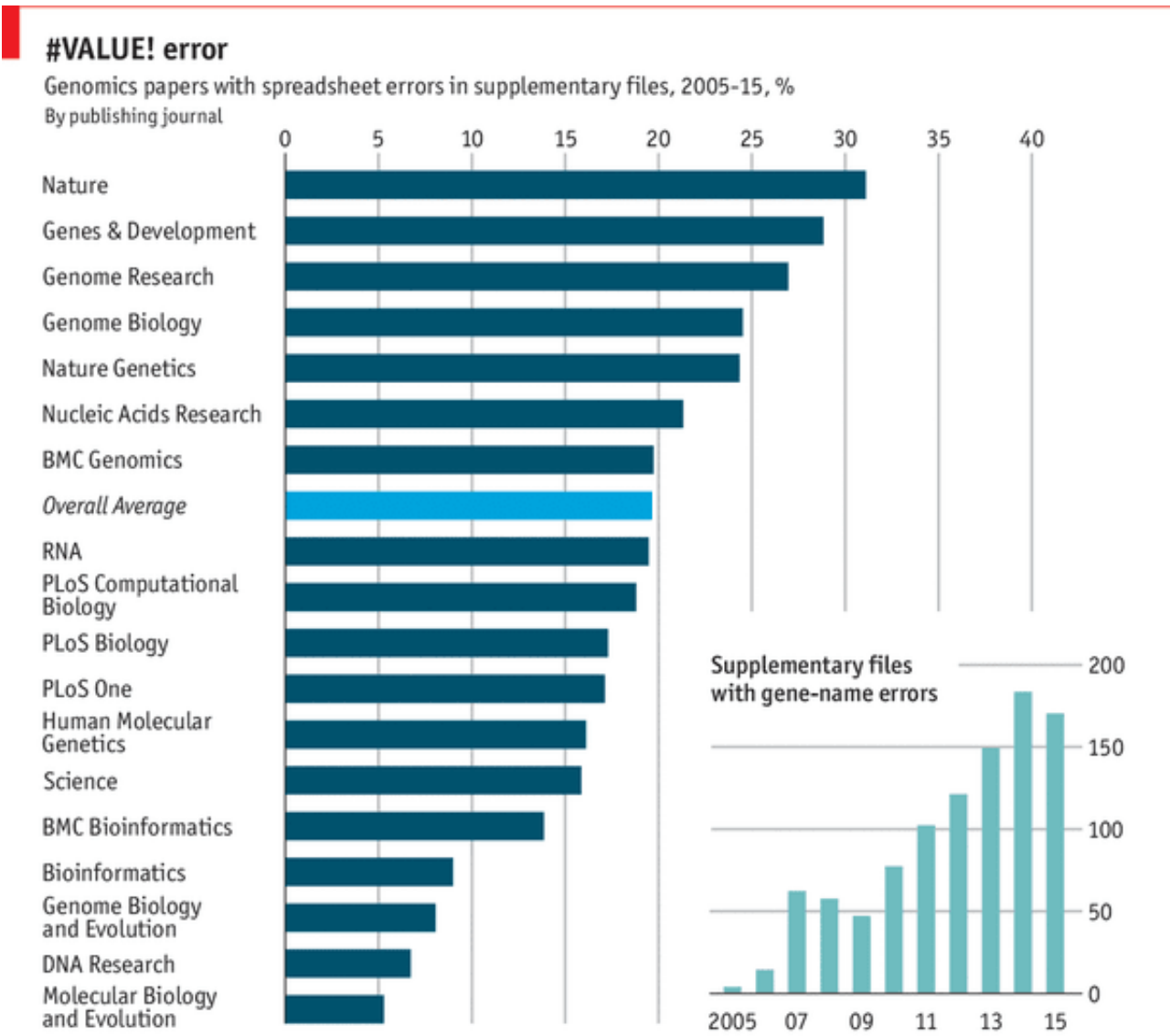
# Graphic detail

Daily chart

## Excel errors and science papers

Spreadsheets are playing havoc with scientists

Sep 7th 2016



Source: "Gene name errors are now widespread in the scientific literature", Ziemann, Eren and El-Osta, 2016

Economist.com

THREE years ago Thomas Herndon, a young graduate student from the University of Massachusetts Amherst, made a splash among economists. Given an assignment to replicate the analysis behind a published academic paper, he pored over the data used for an influential study on government debt written by Carmen Reinhart and Kenneth Rogoff, both professors at Harvard. Much to the authors' embarrassment, Mr Herndon found the most elementary of mistakes: they had accidentally omitted five rows of their spreadsheet when calculating an average. When included, the missing figures weakened the paper's conclusion substantially.

Unsurprisingly, spreadsheet snafus are not unique to economics. A [recent study](#) in the journal *Genome Biology* looked at papers published between 2005 and 2015, and found spreadsheet-related errors in fully one-fifth of articles on genomics that provided supplementary data alongside their text. Although the papers themselves were not necessarily affected, such bugs can create complications for other scientists trying to replicate or build on previous work.

The syntax of genomics makes it particularly difficult for off-the-shelf software to digest. First, spreadsheets often confuse gene symbols for dates. The authors found that Microsoft Excel would often interpret "SEPT2", which corresponds to the gene Septin 2, as "September 2nd". The programme also tended to mistake identification codes like "2310009E13" for numbers in scientific notation—in this particular instance, the code would be read as 2.310009 times  $10^{19}$ . Such hiccups can be quickly resolved by diligent researchers. But they are easy to overlook, and analysts using more sophisticated software may not know to check for them.

The paper's most worrisome finding is that mistakes seem to be proliferating. The authors find that the number of genomics papers packaged with error-ridden spreadsheets is increasing by 15% a year, far above the 4% annual growth rate in the number of genomics papers published. If we extrapolate current trends indefinitely into the future, then by 2025 every spreadsheet attached to a genetics paper will have an error—unless, of course, there is an error in the spreadsheet we used for this calculation.

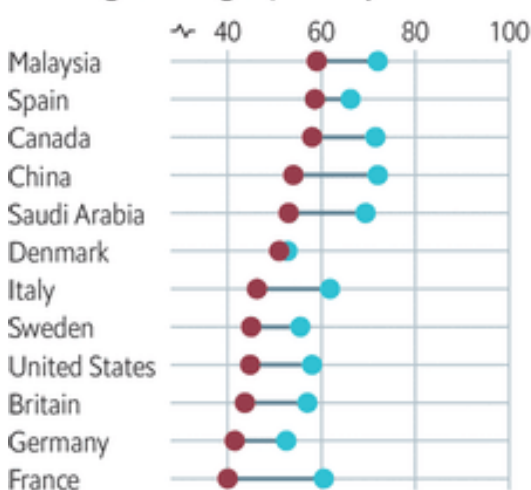
## More from Graphic detail

### A spoonful of sugar

% of respondents who say they are:

Polled in April 2020 (blue dot) September 2020 (red dot)

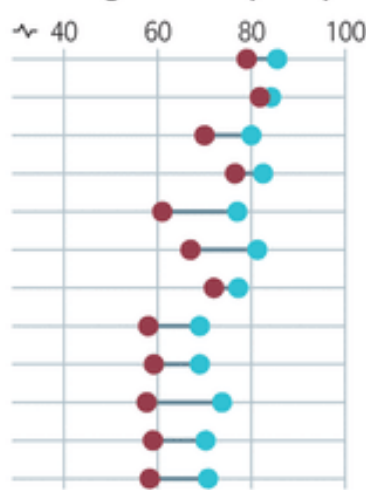
#### Avoiding touching objects in public



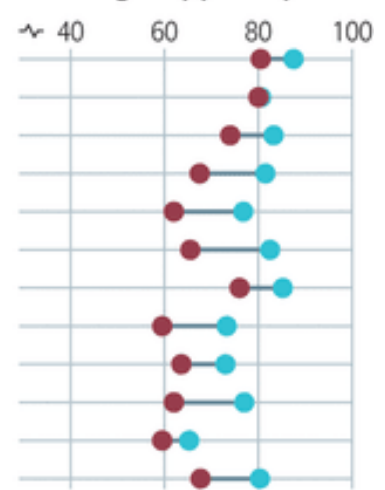
Source: YouGov

The Economist

#### Washing hands frequently



#### Avoiding busy public places



Daily chart

## Pandemic fatigue may be setting in across much of the world

Daily Chart

### Fewer Americans have filed for bankruptcy in 2020 than in 2019

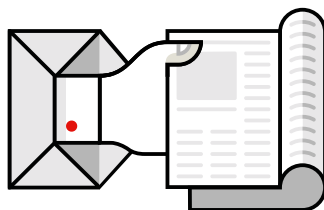
#### Worse than it looks

United States, daily bankruptcy filings in 2020, % change on a year earlier\*



A rising tide

It's better to be a poor pupil in a rich country than the reverse



## The best of our journalism, hand-picked each day

*Sign up to our free daily newsletter, The Economist today*

→ **Sign up now**

**Subscribe**

**Help**

**Group subscriptions**

---

**Keep updated**



Published since September 1843 to take part in “*a severe contest between intelligence, which presses forward, and an unworthy, timid ignorance obstructing our progress.*”

---

## The Economist

About

Advertise

Press centre

## The Economist Group

The Economist Group

The Economist Intelligence Unit

The Economist Store

Careers

Which MBA?

GMAT Tutor

GRE Tutor

Executive Jobs

Executive Education Navigator

---

[Terms of Use](#) [Privacy](#) [Cookie Policy](#) [Manage Cookies](#) [Accessibility](#) [Modern Slavery Statement](#)

[Do Not Sell My Personal Information](#)

Copyright © The Economist Newspaper Limited 2020. All rights reserved.