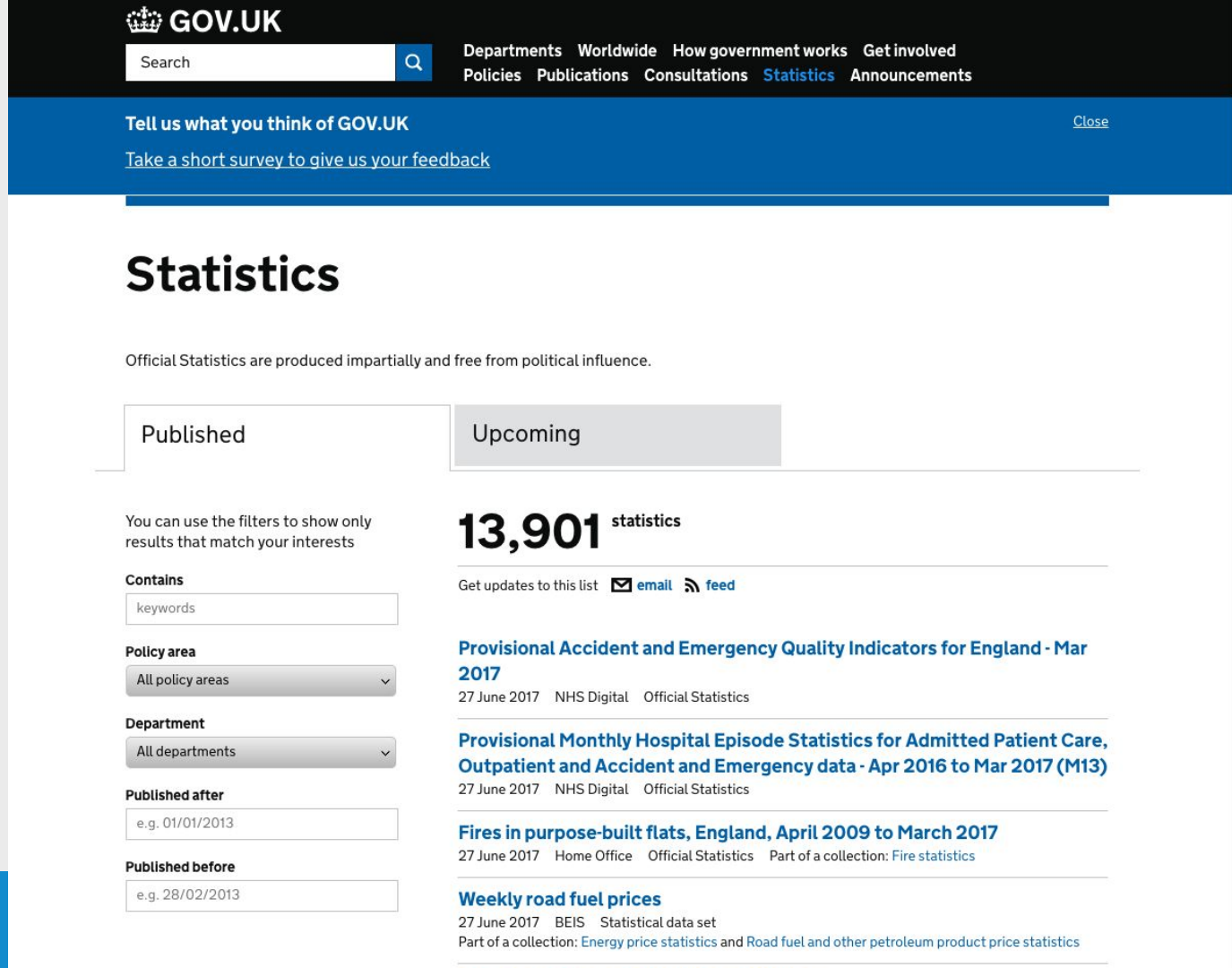
Cabinet Office

Matthew Gregory
Data Scientist
Government Digital Service
@mammykins_

Reproducible Analytical Pipelines

The problem

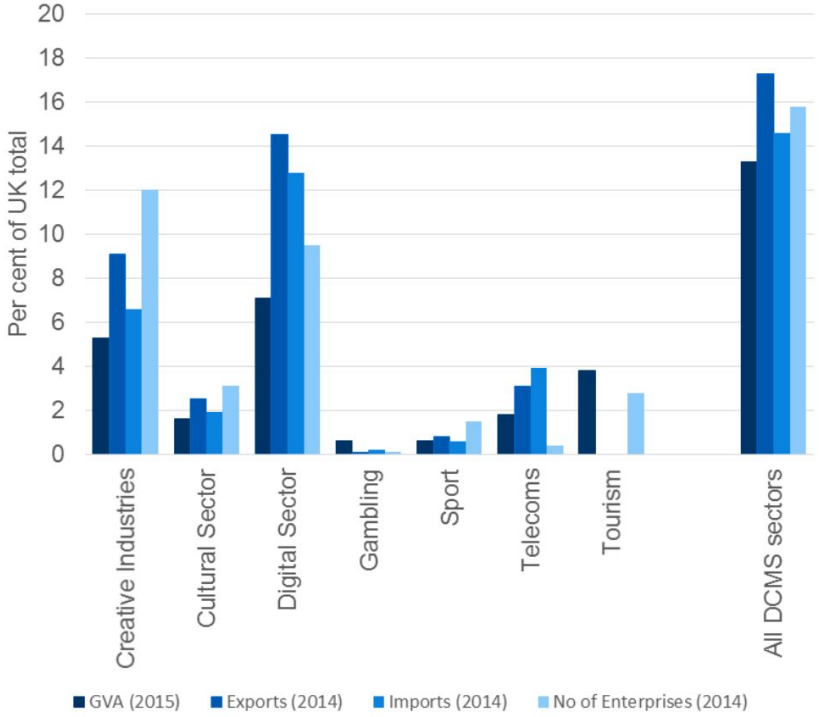

GOV.UK
Search
Departments Worldwide How government works Get involved
Policies Publications Consultations Statistics Announcements
Tell us what you think of GOV.UK
Take a short survey to give us your feedback
Close
Statistics
Official Statistics are produced impartially and free from political influence.
Published
Upcoming
You can use the filters to show only results that match your interests
Contains
keywords
Policy area
All policy areas
Department
All departments
Published after
e.g. 01/01/2013
Published before
e.g. 28/02/2013
13,901 statistics
Get updates to this list email feed
Provisional Accident and Emergency Quality Indicators for England - Mar 2017
27 June 2017 NHS Digital Official Statistics
Provisional Monthly Hospital Episode Statistics for Admitted Patient Care, Outpatient and Accident and Emergency data - Apr 2016 to Mar 2017 (M13)
27 June 2017 NHS Digital Official Statistics
Fires in purpose-built flats, England, April 2009 to March 2017
27 June 2017 Home Office Official Statistics Part of a collection: Fire statistics
Weekly road fuel prices
27 June 2017 BEIS Statistical data set
Part of a collection: Energy price statistics and Road fuel and other petroleum product price statistics

The same kind of input data is used to output the same kind of report, periodically.

Number of enterprises

- In 2014, DCMS sectors accounted for 15.8 per cent of all UK enterprises, approximately 331 thousand enterprises. The number of enterprises in all DCMS sectors has increased by 4.1 per cent between 2013 and 2014, and by 19.5 per cent since 2008.
- The Digital Sector had the largest growth in number of enterprises, increasing by 5.7 per cent between 2013 and 2014. A number of sectors had a reduction in the number of enterprises. The Gambling sector had the largest decrease since 2008 (29.1%). Tourism had the largest decrease since 2013 (11.4%)

Figure 2.1: Key economic measures as per cent of UK totals



RAP came about by recognising two things:

1. Workflows with a large number of manual steps are time consuming and potentially error prone.

“After subtracting the old rate from the new rate, the spreadsheet divided by their sum instead of their average, as the modeler had intended. This error likely had the effect of muting volatility by a factor of two and of lowering the [Value at Risk].”

JPMorgan Embarrassed Over $2 Billion in Losses [Update]

By **Adam Pasick**

Photo: Mario Tama/2012 Getty Images

JPMorgan Chase CEO Jamie Dimon made a shocking disclosure Thursday night that some of the company’s bets on credit markets have gone bad to the "significant" tune of $2 billion. The losses stemmed from a hedging

2. Statisticians increasingly have programming skills, and are using statistical programming tools.


Microsoft
SQL Server


R


APACHE
Spark


sas


SPSS


PostgreSQL


python

Whilst the tools may have moved forward, in general, the way we manage our code has not.

There are many tools and techniques we can learn from software developers.

Transparency and Auditability

with version control

What, Who, Why, When

Commits on Mar 7, 2017

Merge pull request #85 from ukgovdatascience/fix/extract_DCMS_sectors ...

ivyleavedtoadflax committed on GitHub on Mar 7 ✓ 85.45%

5e91f35

fixed extract_DCMS_sectors (closed #60)

DCMSstats committed with ivyleavedtoadflax on Mar 7 ✓ 85.45%

Verified

69fe06e

Commits on Mar 6, 2017

Merge pull request #84 from ukgovdatascience/feature/appendSectors ...

ivyleavedtoadflax committed on GitHub on Mar 6 ✓ 85.40%

a269c65

Re-factored part of appendSectors.R function

ivyleavedtoadflax committed on Mar 6 ✓ 85.40%

Verified

a47120e

Add utils function to elongate SIC

ivyleavedtoadflax committed on Mar 6 ✓ 85.45%

Verified

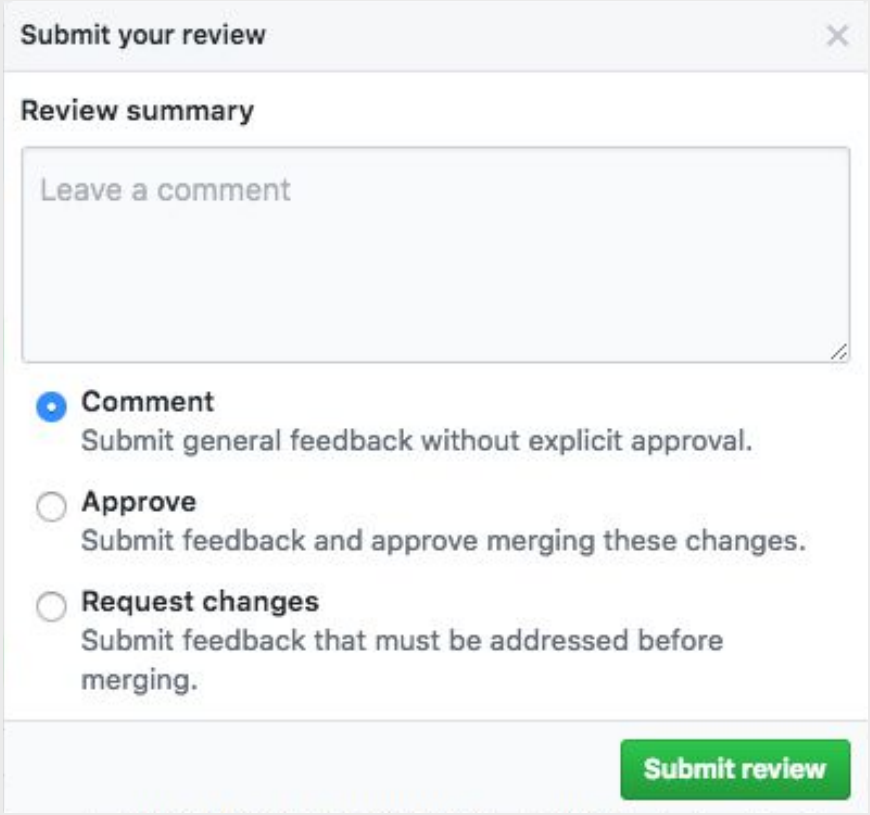
b90c973

added appendSectors function

DCMSstats committed on Mar 6 ✓ 85.37%

79a012f

Quality assurance as you go, not just at the end

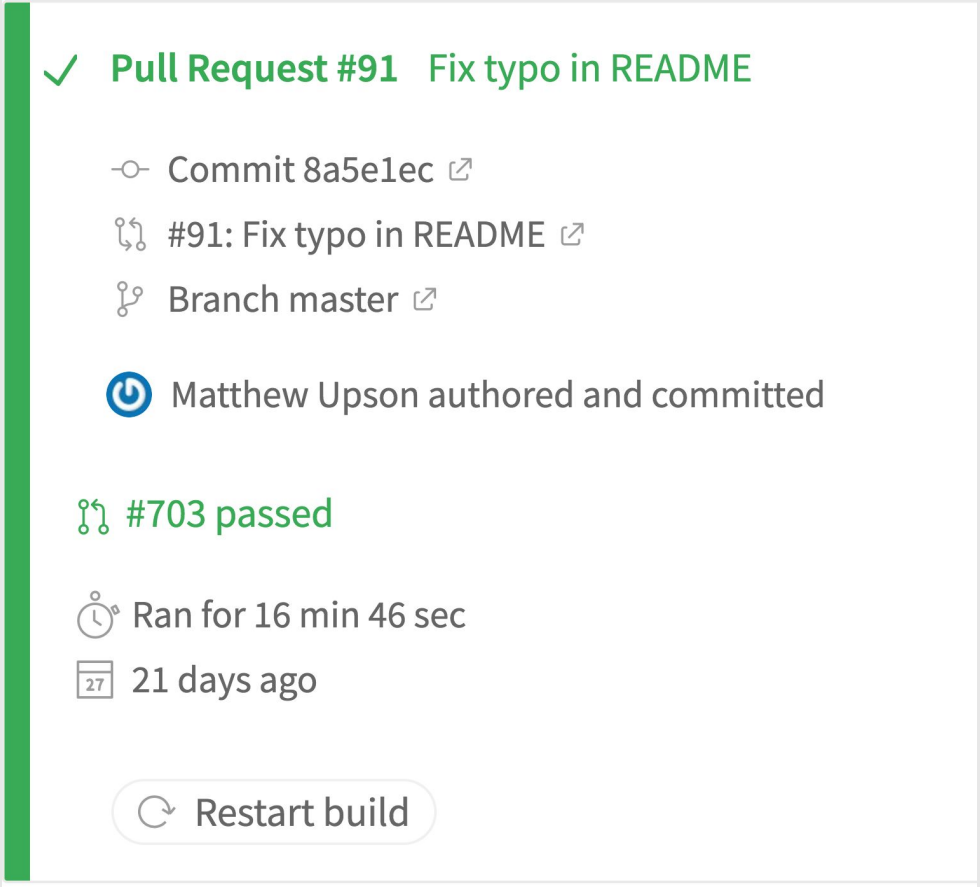


Quality Assurance

with automated testing

The aim is to formalise the informal tests already conducted by those who know the data best.

These tests can be automated

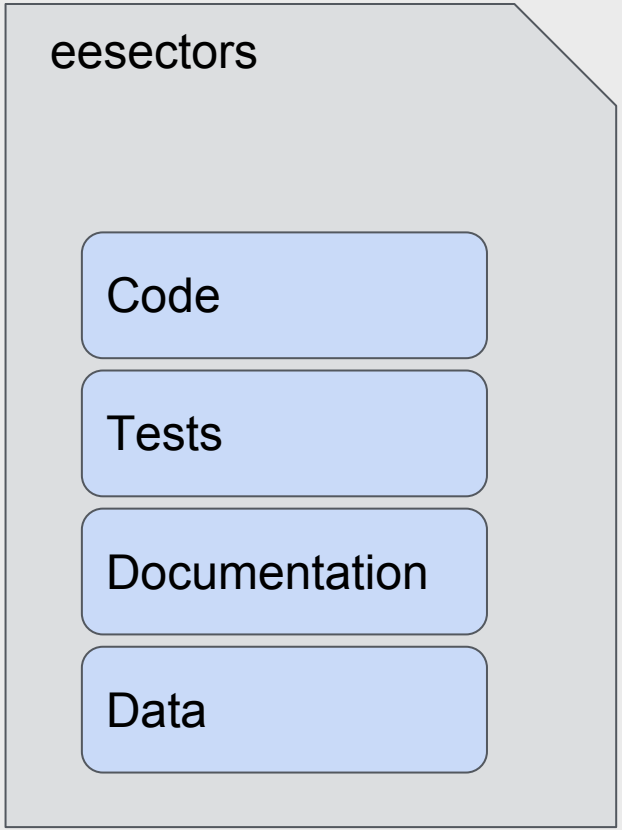


Efficiency

writing software instead of scripts

Everything is in one place.

This helps with institutional knowledge management.



Automating the repetitive processes can reduce production time by 75%

Where are we now?

We’re going to hear from our collaborators across Government.

Department for Education

Department for Culture Media & Sport

Ministry of Justice

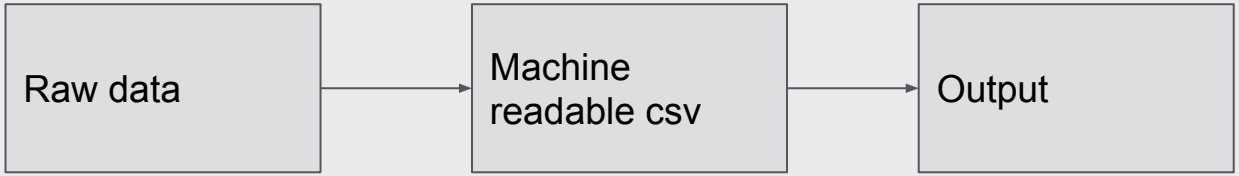
Department
for Education

Laura Selby
Statistician

Our aim when we started:

- To release more data to users via machine readable format
- Drive efficiencies through automation of production
- Use code effectively to store knowledge

Our SFR pipeline:



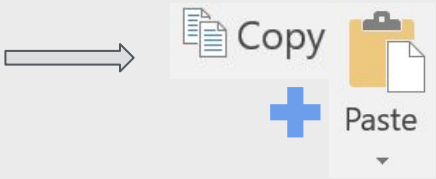
What we wanted to avoid

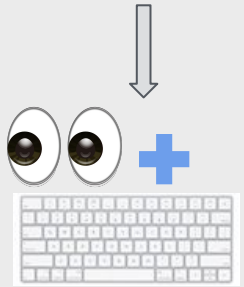
Manually adding numbers

Table 2.2
Pupil enrolments with one or more session of absence by reason
State-funded primary, secondary and special schools
2011/12, five half terms and 2012/13 to 2015/16, six half terms
England

	Five half terms	Six half terms			
	2011/12	2012/13	2013/14	2014/15	2015/16
Percentage of pupils with one or more session of (3):					
Overall absence	**91.0**	**93.8**	**88.4**	**92.0**	**91.7**
Authorised absence	**89.1**	**92.0**	**85.8**	**89.5**	**89.0**
Illness (NOT medical or dental appointments)	80.3	84.6	80.6	82.7	80.8
Medical/dental appointments	36.3	38.8	37.7	37.0	36.3
Religious observance	2.8	3.9	6.8	4.9	5.1
Study leave	2.1	1.5	1.5	1.2	1.2
Traveller absence	0.1	0.1	0.1	0.1	0.1
Agreed family holiday (2)	13.0	15.1	4.9	3.4	3.3
Excluded, no alternative provision	2.4	2.4	2.3	2.5	2.6
Other authorised circumstances	20.8	24.9	20.6	20.1	19.7
Unauthorised absence	**28.6**	**35.1**	**29.7**	**36.6**	**38.6**
Family holiday not agreed	5.8	8.8	10.1	10.4	11.9
Arrived late	6.7	7.5	7.4	7.7	7.8
Other unauthorised circumstances	17.4	21.4	21.5	22.1	23.6
No reason yet	6.8	8.2	7.4	6.7	6.0

Source: School Census



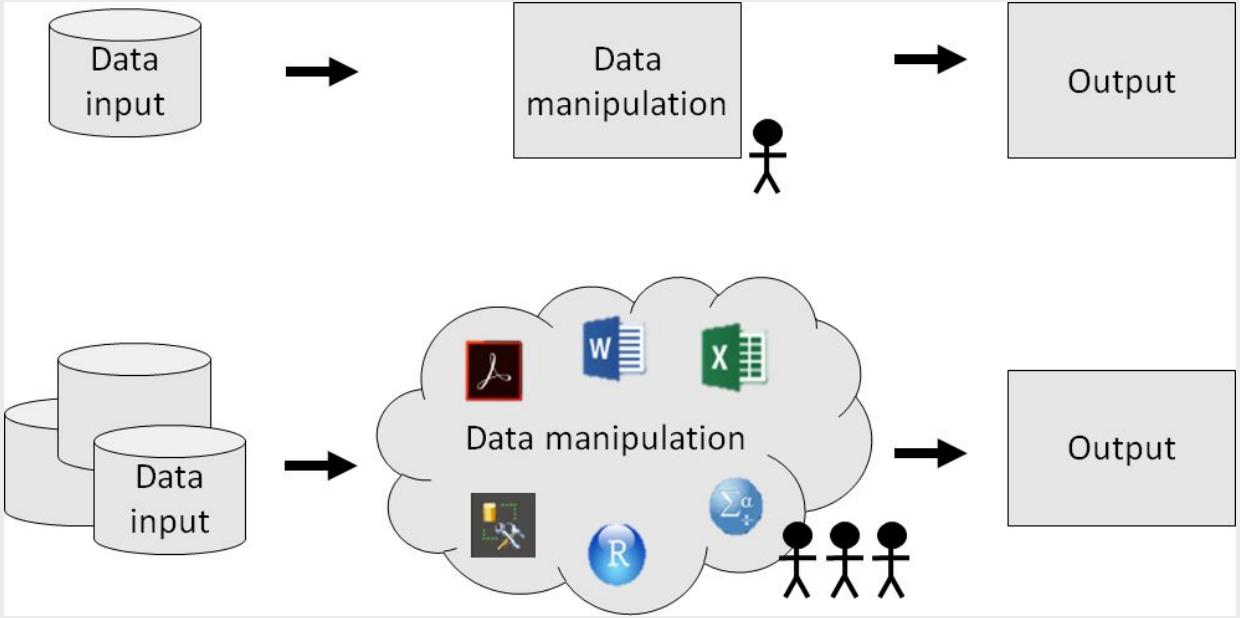


3. Reasons for absence (Tables 2.1, 2.2, 2.3 & underlying data)

Illness remained the most common reason for absence in 2015/16, accounting for 57.3 per cent of all absence. In 2015/16, of all pupils in state-funded primary, secondary and special schools, 80.8 per cent missed at least one session due to illness, and 36.3 per cent missed at least one session due to medical/dental appointments (see chart 4.)

Moving across software

- Reports take a long time to produce
- Reports are hard to reproduce



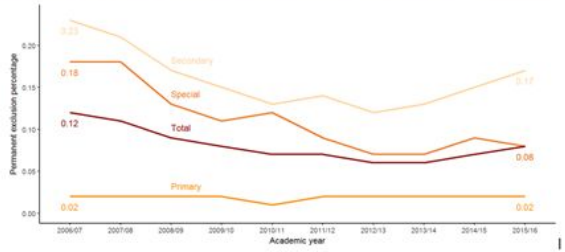
Using rmarkdown

Department for Education

Permanent and Fixed Period Exclusions in England: 2015 to 2016

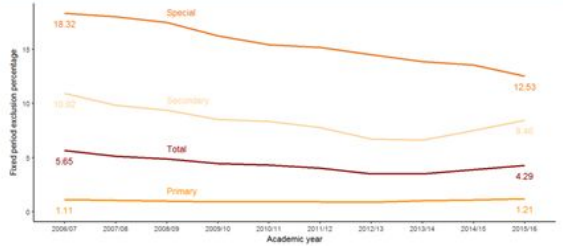
SFR 35/2017, 20 July 2017

The number and rate of permanent exclusions have increased since last year



The overall rate of permanent exclusions has increased from 0.07 per cent of pupil enrolments in 2014/15 to 0.08 per cent in 2015/16.

The number and rate of fixed period exclusions have increased since last year



The overall rate of fixed period exclusions increased, from 3.88 per cent of pupil enrolments in 2014/15 to 4.29 per cent in 2015/16.

Department for Education

Permanent and Fixed Period Exclusions in England: date

Title, date

The number and rate of permanent exclusions changed since last year

plot

The overall rate of permanent exclusions has changed from number per cent of pupil enrolments in year A to number per cent in year B

The number and rate of fixed period exclusions have changed since last year

plot

The overall rate of fixed period exclusions changed from number per cent of pupil enrolments in year A to number per cent in year B

A basic example

example_rmd.Rmd*

Knit | Insert | Run

```
1  ---
2  title: "An example of Rmarkdown"
3  output: word_document
4  ---
5
6  ## My Rmarkdown document
7
8  This is an R Markdown document. You can write text, like a normal word editor.
9
10 ### Add a sub header
11
12 You can also add data tables. The first 5 rows in the 'cars' data set are:
13
14 ```{r cars table, echo=FALSE, warning=FALSE}
15 library(pander)
16
17 pander(head(cars))
18 ```
19
20 You can also add a plot.
21
22 ```{r cars plot, echo=FALSE}
23 plot(cars)
24 ```
25
26 Or add numbers inline:
27
28 The total number of rows in the 'cars' dataset is `r nrow(cars)`. The average
   speed of the cars is `r mean(cars$speed)`.
29
```

28:121 | Add a sub header | R Markdown

An example of Rmarkdown

My Rmarkdown document

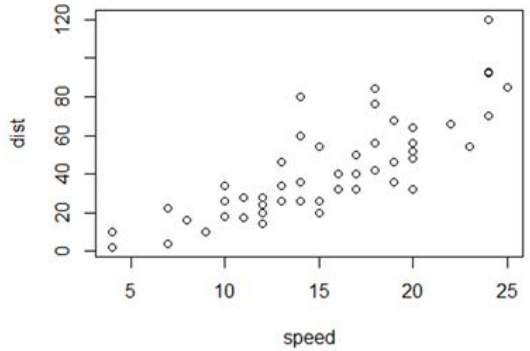
This is an R Markdown document. You can write text, like a normal word editor.

Add a sub header

You can also add data tables. The first 5 rows in the 'cars' data set are:

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10

You can also add a plot.



Or add numbers inline:

The total number of rows in the 'cars' dataset is 50. The average speed of the cars is 15.4.

Applying to National Statistics

Using functions

```
change_ed <- function(numA, numB) {

  if(numA < numB) {return ('increased')}

  if(numA > numB) {return ('decreased')}

  else {return('stayed the same')}

}

# change_ed(10,20)
perm_exc_rate_ref <- function(refyear) {

  data <- filter(main_ud, year == refyear)

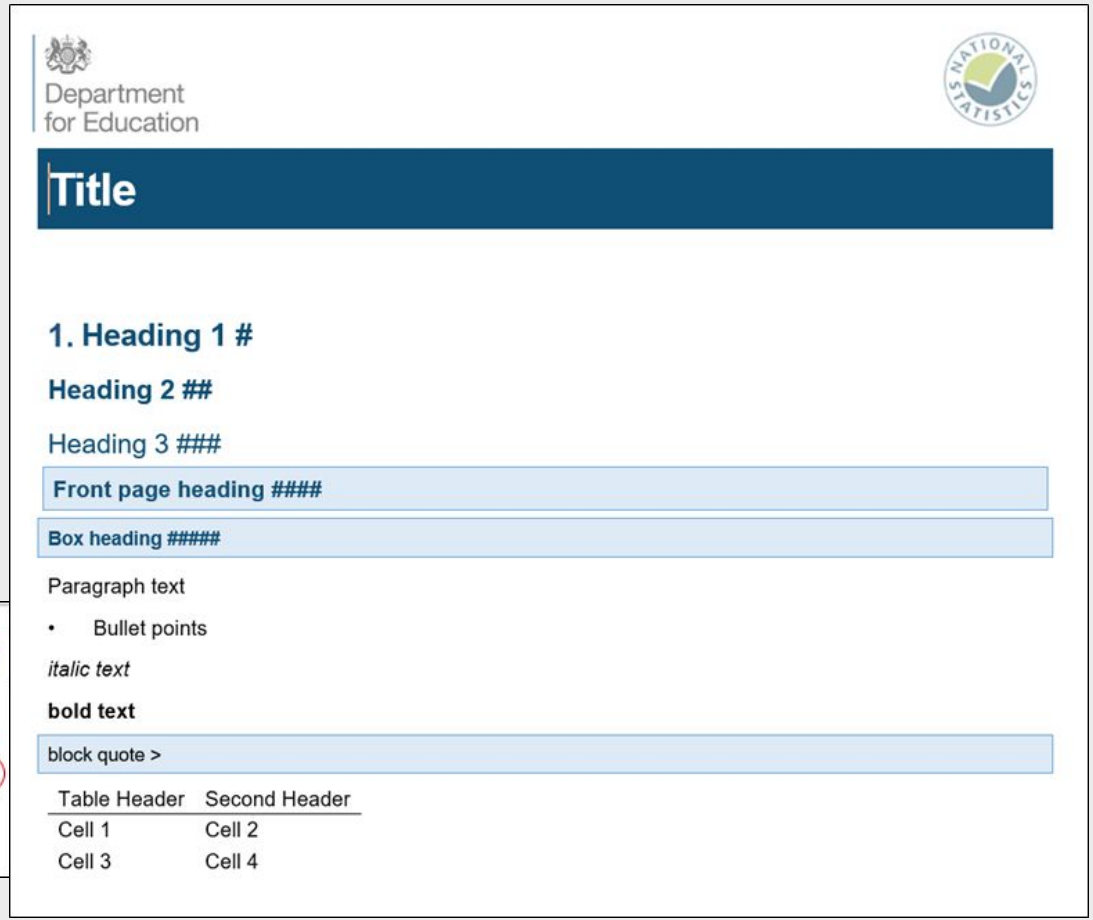
    return(round(filter(data, Level == 'National', School_type == 'Total') %>%
                   dplyr::select(Perm_excl_percent),2))
}

# example
# perm_exc_rate_ref(latest_year)
```

Add some style

You can create a style sheet in Word which forces R to output your document in the desired format.

```
1  ---
2  title: "An example of Rmarkdown"
3  output:
4    word_document:
5      reference_docx: mystyle.docx
6      keep_md: true
7  ---
```



```
## SFR 35/2017, 20 July 2017


#### The number and rate of permanent exclusions have `r change_ed(exc_rate$perm_last,exc_rate$perm_latest)` since last year

```{r echo=FALSE, fig.width=10, fig.height=4.2, dpi=144}
# plot, comparing fixed period exclusion rate to one or more fixed period exclusion rate
# input (dataframe, start year)

permplot(nat_summary,200607)

```

The overall rate of permanent exclusions has `r change_ed(exc_rate$perm_last,exc_rate$perm_latest)` from `r exc_rate$perm_last` per cent of pupil enrolments in `r last_year_f` to `r exc_rate$perm_latest` per cent in `r latest_year_f`.

The number and rate of fixed period exclusions have `r change_ed(exc_rate$fixed_last,exc_rate$fixed_latest)` since last year

```{r echo=FALSE, fig.width=10, fig.height=4.2, dpi=144}
# plot, comparing fixed period exclusion rate to one or more fixed period exclusion rate
# input (dataframe, start year)

fixedplot(nat_summary,200607)

```

The overall rate of fixed period exclusions `r change_ed(exc_rate$fixed_last,exc_rate$fixed_latest)`, from `r exc_rate$fixed_last` per cent of pupil enrolments in `r last_year_f` to `r exc_rate$fixed_latest` per cent in `r latest_year_f`.
```

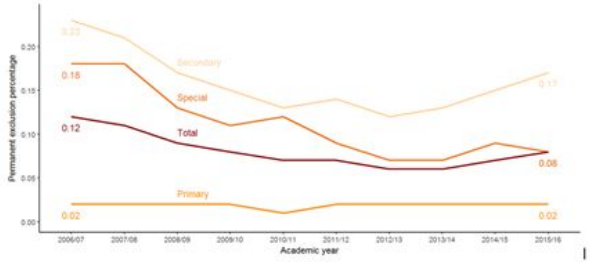
Final output

Department for Education

Permanent and Fixed Period Exclusions in England: 2015 to 2016

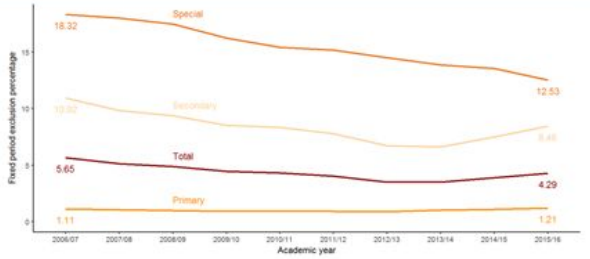
SFR 35/2017, 20 July 2017

The number and rate of permanent exclusions have increased since last year



The overall rate of permanent exclusions has increased from 0.07 per cent of pupil enrolments in 2014/15 to 0.08 per cent in 2015/16.

The number and rate of fixed period exclusions have increased since last year



The overall rate of fixed period exclusions increased, from 3.88 per cent of pupil enrolments in 2014/15 to 4.29 per cent in 2015/16.

1. Permanent exclusions

Permanent exclusion rate definition

A permanent exclusion refers to a pupil who is excluded and who will not come back to that school (unless the exclusion is overturned). The permanent exclusion rate is calculated as follows:

$$\frac{Number\ of\ permanent\ exclusions\ recorded\ across\ the\ academic\ year}{Number\ of\ sole\ and\ dual\ main^{1}\ registered\ pupils\ on\ roll\ as\ at\ January\ census\ day} \times 100$$

The number of permanent exclusions across all state-funded primary, secondary and special schools has increased from 5,795 in 2014/15 to 6,685 in 2015/16. This corresponds to around 35.2 permanent exclusions per day[2] in 2015/16, up from an average of 30.5 per day in 2014/15.

The rate of permanent exclusions across all state-funded primary, secondary and special schools has also increased slightly from 0.07 per cent to 0.08 per cent of pupil enrolments, which is equivalent to 8 pupils per 10,000.

Most (81 per cent) permanent exclusions occurred in secondary schools. The rate of permanent exclusions in secondary schools increased from 0.15 per cent in 2014/15 to 0.17 per cent in 2015/16, which is equivalent to 17 pupils per 10,000

The rate of permanent exclusions stayed the same in primary schools, at 0.02 per cent, and decreased in special schools from 0.09 per cent in 2014/15 to 0.08 per cent in 2015/16

Looking at longer-term trends, the rate of permanent exclusions across all state-funded primary, secondary and special schools has followed a generally downward trend since 2006/07 when the rate was 0.12 per cent.

2. Fixed period exclusions

Fixed period exclusion rate definition

A fixed period exclusion refers to a pupil who is excluded from a school for a set period of time. A fixed period exclusion can involve a part of the school day and it does not have to be for a continuous period. A pupil may be excluded for one or more fixed periods up to a maximum of 45 school days in a single academic year. The fixed period exclusion rate is calculated as follows:

$$\frac{Number\ of\ fixed\ period\ exclusions\ recorded\ across\ the\ academic\ year}{Number\ of\ sole\ and\ dual\ main\ registered\ pupils\ on\ roll\ as\ at\ January\ census\ day} \times 100$$

A pupil may receive more than one fixed period exclusion, so pupils with repeat exclusions can inflate fixed period exclusion rates.

The number of fixed period exclusions across all state-funded primary, secondary and special schools has increased from 302,975 in 2014/15 to 339,360 in 2015/16. This corresponds to around 1,790 fixed period exclusions per day in 2015/16, up from around 1,590 per day in 2014/15.

The rate of fixed period exclusions across all state-funded primary, secondary and special schools has also increased from 3.88 per cent to 4.29 per cent of pupil enrolments, which is equivalent to 429 pupils per 10,000.

There were increases in the number and rate of fixed period exclusions for both state-funded primary and secondary schools but rates decreased in special schools:

[1] For pupils registered at more than one school only their main registration is counted when calculating exclusion rates
[2] Calculated by dividing the total number of exclusions by 190 school days.

Use in DfE

We were the first Government Department to publish Official Statistics in this way.

But this isn’t limited to just Official Statistics -

- Adhoc pieces of analysis
- Quality Assurance reports
- Bespoke reports for individual schools and local authorities

Ministry
of Justice

Christopher Fairbanks
Statistician

Automated tables

Introducing Offender Management Statistics Quarterly (OMSQ)

- National Statistic publication
- Quarterly and Annual editions
- 1 quarterly bulletin (produced using RAP)
- 40 - 60 formatted tables (automated or soon to be automated)

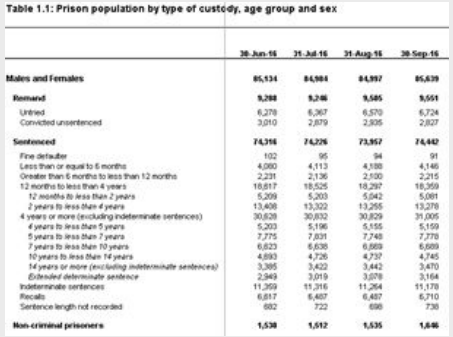


Table 1.1: Prison population by type of custody, age group and sex

	30-Jun-16	31-Jul-16	31-Aug-16	30-Sep-16
Males and Females	**85,134**	**84,984**	**84,997**	**85,639**
Remand	**9,298**	**9,246**	**9,585**	**9,551**
Untried	6,278	6,367	6,570	6,724
Convicted unsentenced	3,010	2,879	2,935	2,827
Sentenced	**74,316**	**74,226**	**73,957**	**74,442**
Fine defaulter	102	95	94	91
Less than or equal to 6 months	4,080	4,113	4,188	4,146
Greater than 6 months to less than 12 months	2,231	2,196	2,100	2,215
12 months to less than 4 years	18,617	18,525	18,297	18,359
12 months to less than 2 years	5,209	5,203	5,042	5,081
2 years to less than 4 years	13,408	13,322	13,255	13,278
4 years or more (excluding indeterminate sentences)	30,828	30,832	30,829	31,005
4 years to less than 5 years	5,203	5,196	5,155	5,159
5 years to less than 7 years	7,775	7,831	7,748	7,778
7 years to less than 10 years	6,823	6,836	6,869	6,889
10 years to less than 14 years	4,693	4,726	4,737	4,745
14 years or more (excluding indeterminate sentences)	3,395	3,422	3,442	3,470
Extended determinate sentence	2,949	3,019	3,078	3,164
Indeterminate sentences	11,359	11,316	11,264	11,178
Recalls	6,817	6,467	6,487	6,710
Sentence length not recorded	682	722	698	738
Non-criminal prisoners	**1,538**	**1,512**	**1,535**	**1,646**

Ministry of Justice

27 July 2017

Offender Management statistics quarterly, England and Wales

Quarter: January to March 2017, Prison population: 30 June 2017

1. Main points

The prison population has been relatively stable for the past five years. This differs to the increasing prison population trend that was observed between the 1990's and early 2000's (see Figure 1). Our most recent extracts indicate that there were:

85,663 prisoners in England and Wales as at 30 June 2017 — There was a slight (1%) increase in the total prison population, compared to the same point in the previous year.

35,517 admissions of which 21,357 were first receptions into prison in the latest quarter — Admissions have increased by 3% on the previous quarter, and decreased by 9% on the same quarter last year. First receptions have increased by 2% on the previous quarter and decreased by 3% on the same quarter last year.

47,767 adjudication outcomes in the last quarter — This is an increase of 16% on the same quarter of the previous year. A total of 5,163 additional days were added to prisoner's sentences.

268,062 offenders on probation as at 31 March 2017 — The number of offenders on probation at the end of March 2017 was 1% higher than the same point in the previous year.

5,347 offenders recalled to prison in the latest quarter — This is a 1% increase on the previous quarter and a 3% increase compared to the same quarter in the previous year.

17,760 releases of which 17,445 were from determinate sentences in the last quarter — The number of releases has remained relatively stable with a slight decrease of 1% compared with the same point in the previous year.

This publication gives offender management statistics for the latest date available and provides comparison with fixed points of time in the previous year. For full and detailed commentary which looks into longer term trends, please refer to the annual publication, published in July, and 'The Story of the Prison Population 1993 - 2016'. For technical detail please refer to the accompanying guide, 'Guide to offender management statistics'.

We are changing how our quarterly bulletins look, and would welcome any feedback to commentary.changing@justice.gsi.gov.uk

For other feedback related to the content of this publication, please let us know at statistics.enquiries@justice.gsi.gov.uk

What do the tables look like?

- Consistent formatting
- Quarters are appended on
- Percentage difference column
- Fonts, indentation

How are they manually produced?

- Outputting tables from SAS
- Copying and pasting values
- Using VLOOKUPS()
- Manual formatting

Table 1.3: Prison population by type of custody, age and sex

Contents

	30-Sep-16	31-Dec-16	31-Mar-17	30-Jun-17	30-Sep-17	Percentage change September 2016 to September 2017
Males and Females	**85,639**	**84,307**	**85,513**	**85,863**	**85,997**	***0%***
15-17	652	600	619	649	598	*-8%*
18-20	4,468	4,357	4,451	4,570	4,549	*2%*
21-24	10,853	10,464	10,481	10,393	10,323	*-5%*
25-29	15,733	15,361	15,587	15,627	15,637	*-1%*
30-39	25,559	25,374	25,866	25,894	25,948	*2%*
40-49	15,428	15,183	15,252	15,354	15,341	*-1%*
50-59	8,386	8,386	8,521	8,564	8,749	*4%*
60 and over	4,560	4,582	4,736	4,812	4,852	*6%*
60-69	*3,075*	*3,066*	*3,175*	*3,213*	*3,251*	*6%*
70 and over	*1,485*	*1,516*	*1,561*	*1,599*	*1,601*	*8%*
Remand	**9,551**	**9,251**	**9,419**	**9,638**	**9,902**	***4%***
15-17	130	127	154	179	143	*10%*
18-20	901	851	960	968	970	*8%*
21-24	1,429	1,371	1,386	1,398	1,423	*0%*
25-29	1,910	1,842	1,814	1,818	1,944	*2%*
30-39	2,882	2,854	2,871	2,903	2,970	*3%*
40-49	1,483	1,458	1,472	1,549	1,554	*5%*
50-59	596	582	600	633	695	*17%*
60 and over	220	166	162	190	203	*-8%*
60-69	*164*	*124*	*125*	*155*	*163*	*-1%*
70 and over	*56*	*42*	*37*	*35*	*40*	**
Sentenced	**74,442**	**73,588**	**74,623**	**74,803**	**74,635**	***0%***
15-17	522	473	465	470	455	*-13%*
18-20	3,492	3,423	3,428	3,541	3,515	*1%*
21-24	9,218	8,916	8,888	8,827	8,739	*-5%*
25-29	13,457	13,184	13,447	13,512	13,392	*0%*
30-39	22,102	21,998	22,498	22,490	22,415	*1%*
40-49	13,641	13,466	13,511	13,525	13,509	*-1%*
50-59	7,687	7,723	7,823	7,837	7,983	*4%*
60 and over	4,323	4,405	4,563	4,601	4,627	*7%*
60-69	*2,897*	*2,931*	*3,039*	*3,037*	*3,067*	*6%*
70 and over	*1,426*	*1,474*	*1,524*	*1,564*	*1,560*	*9%*

What do we want to achieve?

A process that:

- Produce tables quickly
- Is easy to use and maintain
- Reads data from a central, quality assured dataset
  - Easy to change the central dataset
- Output information consistently each quarter:
  - Calculated consistently
  - Consistent user friendly format
- Final output is Indestiguisable from manually creating tables

What did we come up with?

Process flow - producing automated tables

Canonical Data (simplest data form)

Read in a .csv or .sas7bdat

	year	building	trans_method	employee_gender	value
1	2013	Building_1	Bus	M	23
2	2013	Building_1	Cycle	F	19
3	2013	Building_1	Tube	M	32
4	2013	Building_1	Train	F	38
5	2013	Building_1	Walk	M	46
6	2013	Building_1	Bus	F	48
7	2013	Building_1	Cycle	M	33
8	2013	Building_1	Tube	F	5
9	2013	Building_1	Train	M	29
10	2013	Building_1	Walk	F	14
11	2013	Building_1	Bus	M	11
12	2013	Building_1	Cycle	F	1
13	2013	Building_1	Tube	M	14
14	2013	Building_1	Train	F	26
15	2013	Building_1	Walk	M	41
16	2013	Building_1	Bus	F	49
17	2013	Building_1	Cycle	M	34
18	2013	Building_1	Tube	F	38
19	2013	Building_1	Train	M	14
20	2013	Building_1	Walk	F	36
21	2013	Building_1	Bus	M	31
22	2013	Building_1	Cycle	F	8
23	2013	Building_1	Tube	M	18

Cross-tabulation

Using reshape2::dcast

	employee_gender	trans_method	2013	2014	2015	2016
1	Females	(all)	318	226	196	301
3	Females	Bus	77	92	38	87
5	Females	Cycle	53	12	13	38
7	Females	Train	87	33	56	31
9	Females	Tube	69	54	78	79
11	Females	Walk	32	35	11	66
2	Males	(all)	247	285	362	262
4	Males	Bus	39	58	98	77
6	Males	Cycle	49	65	59	54
8	Males	Train	25	73	54	63
10	Males	Tube	36	26	74	30
12	Males	Walk	98	63	77	38

Apply styles

Using xltabr and ‘style sheet’

How people get to work by gender and method of transport, 2013 to 2016

	2013	2014	2015	2016	Percentage change, 2013 to 2016
Females	**267**	**210**	**335**	**368**	**38%**
Bus	67	54	87	97	45%
Cycle	113	33	87	32	-72%
Train	34	9	66	51	50%
Tube	34	46	39	73	115%
Walk	19	68	56	115	505%
Males	**383**	**255**	**358**	**295**	**-23%**
Bus	64	49	54	61	-5%
Cycle	79	28	65	70	-11%
Train	65	83	44	107	65%
Tube	94	54	63	37	-61%
Walk	81	41	132	20	-75%

Introducing ‘xltabr’

- R package, created in-house at MoJ

What can ‘xltabr’ do?

- Turns R dataframes into .xlsx formatted tables using a ‘style sheet’

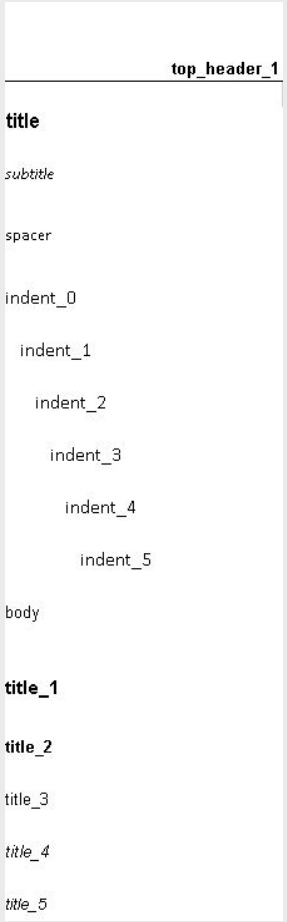
Is this available to everyone?

- Yes, https://github.com/moj-analytical-services/xltabr

Style sheet

Can specify that a cell has any combination of the following styles:

- Fonts
- Bold/Italics/Underlined
- Justification
- Row height
- Column width
- Cell colour
- Borders
- ...

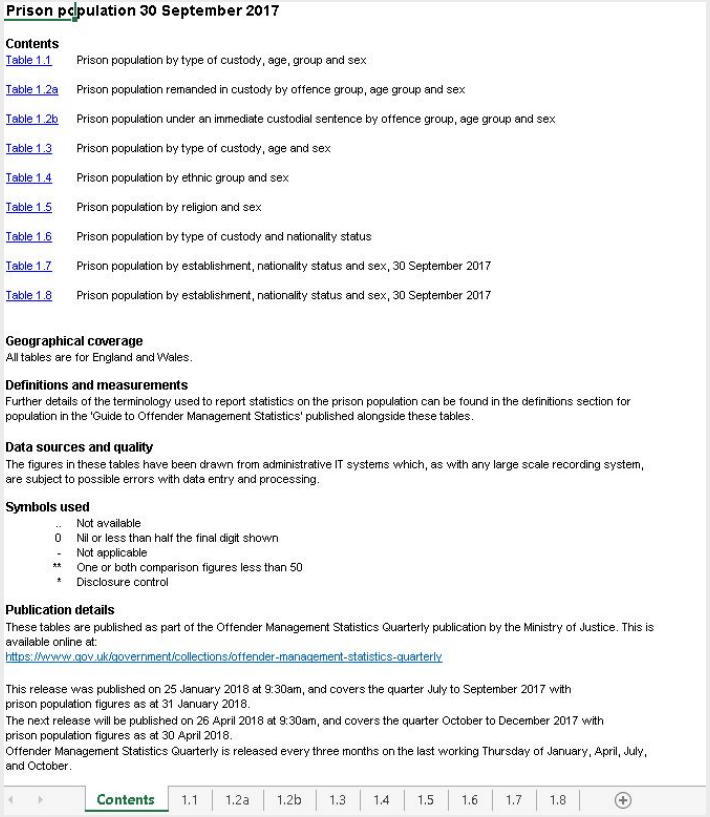


Final product

The Final Product

.xlsx workbook:

- Contents page updates when tables are created
- Publication dates update, including the next publication date
- Tab for each table
- Hyperlinks link to tables



Prison population 30 September 2017

Contents

Table 1.1	Prison population by type of custody, age, group and sex
Table 1.2a	Prison population remanded in custody by offence group, age group and sex
Table 1.2b	Prison population under an immediate custodial sentence by offence group, age group and sex
Table 1.3	Prison population by type of custody, age and sex
Table 1.4	Prison population by ethnic group and sex
Table 1.5	Prison population by religion and sex
Table 1.6	Prison population by type of custody and nationality status
Table 1.7	Prison population by establishment, nationality status and sex, 30 September 2017
Table 1.8	Prison population by establishment, nationality status and sex, 30 September 2017

Geographical coverage
All tables are for England and Wales.

Definitions and measurements
Further details of the terminology used to report statistics on the prison population can be found in the definitions section for population in the 'Guide to Offender Management Statistics' published alongside these tables.

Data sources and quality
The figures in these tables have been drawn from administrative IT systems which, as with any large scale recording system, are subject to possible errors with data entry and processing.

Symbols used

..	Not available
0	Nil or less than half the final digit shown
-	Not applicable
**	One or both comparison figures less than 50
*	Disclosure control

Publication details
These tables are published as part of the Offender Management Statistics Quarterly publication by the Ministry of Justice. This is available online at:
https://www.gov.uk/government/collections/offender-management-statistics-quarterly

This release was published on 25 January 2018 at 9:30am, and covers the quarter July to September 2017 with prison population figures as at 31 January 2018.
The next release will be published on 26 April 2018 at 9:30am, and covers the quarter October to December 2017 with prison population figures as at 30 April 2018.
Offender Management Statistics Quarterly is released every three months on the last working Thursday of January, April, July, and October.

Contents | 1.1 | 1.2a | 1.2b | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8

MoJ

The Final Product

Tables:

- Consistent formatting
- Automated suppression of percentage change based on values less than 50
- Specified row/column height/widths
- Specified borders and indentation

Table 1.2a: Prison population remanded in custody by offence group, age group and sex

	30-Sep-16	31-Dec-16	31-Mar-17	30-Jun-17	30-Sep-17	Percentage change from 30-Sep-16 to 30-Sep-17
Males and Females	**9,551**	**9,251**	**9,419**	**9,638**	**9,902**	**4%**
Violence against the person	2,048	2,000	2,095	2,126	2,249	10%
Sexual offences	892	882	823	871	956	7%
Robbery	698	754	751	797	791	13%
Theft Offences	1,595	1,496	1,471	1,481	1,542	-3%
Criminal damage and arson	333	352	327	363	345	4%
Drug offences	1,473	1,370	1,545	1,449	1,530	4%
Possession of weapons	470	457	450	519	509	8%
Public order offences	186	215	193	255	244	31%
Miscellaneous crimes against society	605	581	565	495	513	-15%
Fraud Offences	119	106	126	125	130	9%
Summary Non-Motoring	864	860	814	893	827	-4%
Summary motoring	34	34	42	30	40	**
Offence not recorded	234	144	217	234	226	-3%
Adults	**8,520**	**8,273**	**8,305**	**8,491**	**8,789**	**3%**
Violence against the person	1,796	1,721	1,785	1,798	1,941	8%
Sexual offences	809	797	740	797	870	8%
Robbery	565	621	592	632	650	15%
Theft Offences	1,428	1,365	1,347	1,353	1,395	-2%
Criminal damage and arson	306	316	300	338	325	6%
Drug offences	1,310	1,223	1,359	1,270	1,332	2%
Possession of weapons	400	408	383	439	430	8%
Public order offences	174	197	178	231	221	27%
Miscellaneous crimes against society	561	545	518	457	480	-14%
Fraud Offences	116	106	122	123	129	11%
Summary Non-Motoring	809	807	748	814	774	-4%
Summary motoring	30	28	36	25	38	**
Offence not recorded	216	139	197	214	204	-6%

What are the benefits of automating tables?

Accuracy:

- Information held in central datasets feed **directly** into the tables
  - Less room for human error (typos, copy + paste)

Efficiency:

- Once automation code has been written, tables can be created at the touch of a button
  - Alleviating time pressure, during busy periods

Department
for Culture
Media & Sport

Olivia Christopherson

Head of Profession

MoJ

Automating DCMS Sector Economic Estimates

- New, high profile publication (various economic measures for 7 DCMS sectors & numerous sub-sectors)
- Very manual process, resource intensive, risk of errors & repetitive work
- Each year produced separately and then QA’ed
- Significant demands for additional breakdowns
- Risk sectors will be redefined

Things to consider

- Software – access to R and Rstudio
- Installation of git on developers' laptops needed for collaborative development
- Data scientist or equivalent with sufficient knowledge
- Dedicated resource (can't be done alongside day job!)
- Importance of good documentation
- Knowing when to stop perfecting code
- Good way to secure data science resource
- Wider benefits - working with GDS; R skills

Ministry
of Justice

Vicky Hughes
Data Scientist

Culture of change

Choose where adds most value

Senior buy-in

Make it less scary

Use the cross Government group

Create an internal group

Cabinet Office

Department for Education

Department for Culture Media & Sport

Ministry of Justice

Thanks! Questions?

https://ukgovdatascience.github.io/rap_companion/