Analysis Journal—RAC Parking Report Automation

mz

Saturday 9.3.2019

- Initialised public github repo.

- Setup folder structure, currenlty looks like this (output of `tree --charset=ascii -d`):

```
|-- code
|-- data
|   |-- 01-raw
|   |-- 02-interim
|   `-- 03-processed
|-- docs
|   |-- admin
|   |-- journal
|   `-- original-reports
|-- figures
`-- outputs
    |-- instructions
    |-- reports
    `-- technical-appendix
```

- available reports used for development, downloaded from RAC press website, pdfs stored in `/docs/original-reports`, see appendix for full table

Next steps: find data

Thursday 14.3.2019

- Look through back reports available, saving all and saving links here, in appendix 1

- Start query file for Ivo.

- OK, got all reports, despite missing years, it looks like most of the data is included in other reports as they have a window of usually 4 years previous that they include data for.

- Now figure out data sources for England, and detail them

- Outline typical England report

England data requirements and availability?

- 2017/18 uses "MHCLG tables on parking income and expenditure"

This looks like the data https://www.gov.uk/government/collections/local-authority-revenue-expenditure-and-financing.

- For each year there is normally budget, provisional outturn and outturn data, the latter needs to be for individual local authorities, which does not seem to exist for 2007/08, but is available for all other years.

- The files are grouped under the title *Local authority revenue expenditure and financing England: 2010 to 2011 individual local authority data - outturn* for each year

- The relevant Excel file is usually called `Revenue outturn RO2 (highways and transport services) + year.xls`.

- There are three columns we are interested in: expenditures, incomes and capital charges, these are usually on separate sheets. There is one of these for both on street and off street parking. So that's 6 columns in total.

- There are also summaries, but I don't need them, maybe only for manually double-checking.

- Appendix 2 lists all the files, their links, the number or rows (local authorities) in the table and the sheets on which the data can be found.

- TODO Add the starting row and the actual columns where the data in question can be found and I can write a funciton to extract the data automatically.

- Download all `.xls` files to `/data/raw`

- "Earlier, councils also submitted budgets for the 2018-19financial year but on a less detailed basis." where are these budgets? Do I need to include them as well? Ah, yes.

- At one point the England data stops including capital charges..2013/14 is last one. This means current reports are less data rich, and if I stick to the brief, it means that the code won't be able to replicate the old reports.

- At one point the Englad data starts including penalty charges for individual Local authorities, although the reports don't do anything special with that, just use the aggregates.

- England totals do not include the national parks and the Nottingham levy. Not sure what to do with that!?

Friday 15.3.2019

- OK, so England outturn data is tabled in Appendix 2.
- Now England Budget data
  - not in *Local authority revenue expenditure and financing tables*
  - under *Budget estimates of local authority revenue expenditure and financing for the financial year* the table *Revenue account budget (RA)* has the parking surplus down as 792, but the report says 782. There have been a few other differnces, not sure if these are typos/errors?
- download data into `\data\raw` and fill out appendix two table with relevant info.
- careful because at one point there are capital charges and then not. So make sure you've got the correct column. Because these are sums, I would even set up a check?
- OK, England budget files are all saved and relevant info extracted manually.
- Now back to outlining the England report as it stands.
- E17/18 says national park surplus is 1.7 million, sum of cells DF405:411 is 1.774.
- Comparing Table 1 for E17/18 to excel file 17/18:
  - on street parking is all the same
  - off street parking: there is a discrepancy. Income is the same, but Leibling's expenditure value is higher than the Excel one (making the surplus lower): 362 vs 356?!

Saturday 16.3.2019

- Comparing Table 1 for E17/18 to excel file 16/17

  - on street parking is all the same
  - off street parking: there is a discrepancy. Leibling's income is 693 vs 689 in the table, and Leibling's expenditures are are 349 vs 343 in the table.

- Comparing Table 1 for E17/18 to excel file 15/16
  - on street parking is all the same
  - off street parking: there is a discrepancy. Leibling's income is 670 vs 674 in the table, and Leibling's expenditures are are 340 vs 337 in the table.

- Trying to figure out where these discrepancies come from: It must be the nottingham workplace levy? Which I don't have data for.

- Second thing, presumably the difference is also due to not including the national parks. Although I have the data for that, I cannot confirm, because I don't know the levy numbers.

- TODO write (find) a funtion to convert excel column names to column numbers.

- Oh, congestion charge data needs to be read as well, that's in a different couple of cells than what i've been looking at until now, so add that to the appendix.

Monday 18.3.2019

- Clean up queries for Ivo for England.
- Look at Scotland data. 2015/16 is the last year with a report but the data on [this link](https://www2.gov.scot/Topics/Statistics/Browse/Local-Government-Finance/PubScottishLGFStats) seems to go to 2016/17
- I'm looking at `Annex A by LA` excel files, which have LA level data.
- OK, so data checking: compare surplusses in individual sheets of `sco-15-16.xlss` with table 9, parking surpluses in report 15/16.
  - Edinburg is same, so is Glasgow, Dundee, Renfrewshire, Fife, South Lank, East Ayr, Highland, Argyle, Aberdeenshire, Angus, South Ary, Moray,Perth, Stirling,
  - Aberdeen city is 4.89 vs 0?
  - If I add Aberdeen city form Leibling to the Excel totals get the Leibling totals for Surplus.OK
- But if I add the Aberdeen city incomes to the Excel ones, it's still not enough. And same for expenditures... So there is another LA that has a discrepancy.
  - Clackmanshire discrepancy: .014 -> .02, but that's minor, could be poor rounding. + South Ayershire discrepancy: .0 -> .80
  - And of course Aberdeen city: .0 -> 9.20
  - This fixes the total income discrepancies.
- OK, now for the expenditure discrepancies:
  - Aberdeen city: 0 -> 4.32
  - South Ayershire expenditure discrepancy: -.327 -> .470
  - ANd these now also add up.
- Start Scotish data query file for Ivo
- Start outline of Scotland report based on *newly discovered* 16/17 report
- Penalty notice charge data is apparenlty out of a scanned pdf in 2015/16.... And 16/18 data in a regular pdf. But perhaps the tabulizer package could help me out? Unfortunatley there are some sort of java issues in installing the package...
- solved issue using this link, installed rJava, but first had to also run `sudo R CMD javareconf`. This is presumably only an ubuntu issue, so should work fine on Windows. * Tried tabulizer on the 17/18 pdf and it works!

PNC data

Decriminalised Parking Enforcement - Local Authorities - Income and Expenditure

- pdf's exist for 2016/17 and 2017/18. Additionally a scanned pdf exists for 2013/14/15/16. All have:
  - a table of which LAs are doing DPE, and which are considering it etc, this is in the report

– a table of PNC incomes. However this last table also has other incomes and expenditures in it, so you'd think it would match with the gov.sco data in the Excel tables, but it doesn't. SOme are quite close, e.g. Glasgow, but others not.

- Added table of PCN data to the appendix, created table to summarise the sources.

*Todo: outline scotland data.

Tuesday 19.3.2019

- Found the 16/17 scotland files, they're not under press, but under publications...
- Add them to app1
- Actually clean up the data source tables and move them to code, these are manual data entry files.
- Saving the tables into rds? Probably, it is serialised, and as opposed to csv it retains the data types. Let's me cleanly load individual objects and assign then names, and there's no danger of overwriting an object of the same name. But is not human readable.
- This also involves using here::here() to load the data in the child .rmd documents for the appendices.

Scotland

- Mainly pure text chunk.
- Available data: fisc.year
- Paragraph on data availability on Penalty Charne Notices, incl. source, which is a scanned pdf...

1. Introduction

England

1. Introduction

- Mainly pure text chunk.
- Available data: fisc.year
- Missing data: national parks, Notthingham workplace parking levy? (17/18)

2. Summary

Table 1 - Parking income and expenditure England

- Unit: whole country
- Available data:
    – Off-street: expenditures (sum from LA tables)
    – Off-street: incomes (sum from LA tables)
    – On-street: expenditures (sum from LA tables)
    – On-street: incomes (sum from LA tables)
    – On-street: penalties - single cell on summary sheet
    – Net expenditure on transport - single cell on summary sheet
    – Budget estimates for next year
- That's all, everything else is calculated from this

Text

- Conditional tree based on data from Table 1.

Figure 1

- Simple line chart based on data from Table 1.

Table 2

- Data from Table 1 for current year, split by London and non-London.

Text

- Conditional tree based on data from Table 2.

3. Income

Text

- Conditional tree based on data from Tables 1 and 2
- Includes footnote on penalty tarrifs.

4. Expenditure

Text

- Conditional tree based on data from Tables 1 and 2

5. Surpluses

Text

- Conditional tree based on data from Tables 1 and 2
- Additionally there is the total for on national parks, that data is also available
- Missing data: the total for Nottingham workplace levy.

Table 3a surpluses for London councils

- Individual LA data for 33 London councils

Table 3b surpluses for top 20 non London councils

- Individual LA data for 33 London councils

Text

- Conditional tree based on data from Tables 3a and 3b

6. Comparison with budgets for2017-18and 2018-19

- Conditional text based on comparison of indovidual LA surpluses with individual LA budgets

7. Congestion charge

Text

- Conditional text based on table below

Table

- Simple table based on data from outturn table.

Appendix 1

- Table of all LA surpluses for 5 years, alphabetical

Appendix 2

- Table of all LA surpluses for 5 years, by surplus size.
- link https://statswales.gov.wales/Catalogue/Local-Government/Finance/Revenue/Transport/RoadsAndTransportRevenu
  by-authority

Appendix 1

links to all reports online

country	year	htmls	pdfs	tab	window
England	2010/11	link		link	
	2011/12	link	link	link	2(4)
	2012/13	link	link	link	3(5)
	2013/14	link	link	<-	4
	2014/15	link*	link	link	4
	2015/16	link	link	link	4
	2016/17	link	link	link	4
	2017/18	link	link	link	4
Scotland	2010/11				
	2011/12	link	link		1
	2012/13				
	2013/14	link	link		3
	2014/15	link	link		4
	2015/16	link	link		4
	2016/17	link	link		
	2017/18				
Wales	2010/11				
	2011/12				
	2012/13	link	link		5
	2013/14				
	2014/15				
	2015/16	link	link		4
	2016/17	link	link		4
	2017/18	link	link		4

∗ link missing from RAC press releases page

Columns headings:

- *country*
- *year*–period covered in report
- *html*–link to press release on RAC website
- *pdf*–link to pdf report if exists
- *tab*–link to additional tables if exist (only England)
- *window*–number of previous years' data included in report, number in brackets is for aggregates, not at LA level

Appendix 2

data overview - England

England–Outturn

year	link	rows	las	first	e.sh	e.on	e.off	i.sh	i.on	i.off	pen.sh	pen.on	pen.1	tot.1
2008/09	xls	405	388	0	3	CU	DB	3	CX	DE	NA	–	G64	J45
2009/10	xls	370	353	0	3	BH	BL	4	BH	BL	NA	–	G64	J45
2010/11	xls	371	353	0	3	BH	BL	4	BH	BL	NA	–	G64	J45
2011/12	xls	371	353	0	3	BH	BL	4	BH	BL	NA	–	G64	J45
2012/13	xls	371	353	12	3	BH	BL	4	BH	BL	NA	–	G65	J46
2013/14	xls	371	353	12	3	BH	BL	4	BH	BL	NA	–	G65	J46
2014/15	xls	371	353	12	3	BE	BI	4	BE	BI	NA	–	I68	L49
2015/16	xlsx	444	353	8	3	CU	DB	3	CX	DE	3	FX	H67	K48
2016/17	xlsx	446	353	8	3	CU	DB	3	CX	DE	3	FX	H67	K48
2017/18	xlsx	445	353	8	3	CU	DB	3	CX	DE	3	FX	H67	K48

- *year*–period covered in report
- *link*–link to data file
- *rows*–number of rows (not all are local authorities though!)
- *las*–number of local authorities (i.e. not ones with class "O"
- *first*–first row of data on sheets 3,4 and 5
- *e.sh*–which sheet are expenditures on
- *e.on*–which column on street expenditures
- *e.off*–which column are off street expenditures
- *e.cc*–which column are congestion charge expenditures on (row = Greater London Authority)
- *i.sh*–which sheet are incomes on
- *i.on*–which column on street incomes
- *i.off*–which column are off street incomes
- *i.cc*–which column are congestion charge incomes on (row = Greater London Authority)
- *pen.sh*–which sheet are penalty charge incomes for LAs on
- *pen.on*–which column are penalty charge incomes for LAs
- *pen.1*–which cell is the totaL Penalty charge notice income for on street parking data on in sheet 2
- *tot.1*–which cell is the Net current expenditures for Total transport and Highways services on in sheet 2

England–Budget

year	link	rows	first	las	budg.tot	budg.la
2008/09		0	0	0	-	-
2009/10		0	0	0	-	-
2010/11	xls	444	10	353	E27	U
2011/12	xls	444	10	353	E28	V
2012/13	xls	444	10	353	E28	V
2013/14	xls	444	10	353	E28	V
2014/15	xls	444	9	353	F34	U
2015/16	xls	444	9	353	F34	U
2016/17	xlsx	443	8	353	E35	U
2017/18	xlsx	446	8	353	E35	V
2018/19	xlsx	443	8	353	E35	V

- *year*–period covered in report
- *link*–link to data file
- *rows*–number of local authorities
- *budg.tot*–wich cell on sheet 2 is the total budgeted surplus in
- *budg.la*–which column on sheet 3 are the LA parking budgeted surpluses on

Appendix 2

data overview - Scotland

year	report	file.name	link	start.sh	end.sh	exp.cell	inc.cell	t.exp.cell	t.inc.cell
2011/12	yes		–	NA	NA	-	-	-	-
2012/13	-	Annex A	xlsx	0	0	F57	G57	F50	G50
2013/14	yes	Annex A	xlsx	2	33	F40	G40	F33	G33
2014/15	yes	Annex A	xlsx	2	33	E40	F40	E33	F33
2015/16	yes	Annex A	xlsx	2	33	B41	C41	B34	C34
2016/17	-	Annex A	xlsx	2	33	B41	C41	B34	C34
2017/18	-	LFR 05	xlsx	3	34	L13	L23	U13	U23

- *year*–period covered in report
- *report*–does the report exist for this year?
- *file.name*–name on link to file on the website
- *link*–link to data file
- *start.sh*–number of first sheet with LA data
- *end.sh*–number of last sheet with LA data
- *exp.cell*–cell with expenditure data
- *inc.cell*–cell with income data
- *t.exp.cell*–total expenditure for all transport and roads cell
- *t.exp.cell*–total income for all transport and roads cell

DPE and PNC data

year	file.type	link	dpe.tab	pcn.tab	e.i.tab
2011/12	–	–	NA	NA	NA
2012/13	–	–	NA	NA	NA
2013/14/15/16	scan	pdf	NA	NA	NA
2016/17	pdf	pdf	2	3	4
2017/18	pdf	pdf	2	3	4

- *year*–period covered in report
- *report*–does the report exist for this year?
- *file.name*–name on link to file on the website
- *link*–link to data file
- *dpe.tab*–which table (including the table of contents) contains the table of LAs that have/haven't DPE
- *pcn.tab*–which table (including the table of contents) contains data on no of PCNs issued
- *i.e.tab*–which table (including the table of contents) contains data on incomes and expenditures of DPE LAs