

Data Science for Actuaries

Paul King, ATRC 28 June 2019

					11° 1						Č.			H	11	1						E .	1		11		1.3		13	~													34																								
																	2 1.	- 4	. ?	• •	• •		2 1	f	·		•	-	T	-																																					
-												_					-													-																																					
	0 0	3	00	10	0	8 9	10	0	0 0	3 14	15	16 1	17 1	3 0	20	21	22 2	3 24	1 25	26	27	28	29 3		1 32	33	0	35	36	37	38	0 (39 4			43	0	45	46 4	7 4	8 49	9 50	0	52	53 5	0 C	5 5	6 51	7 58	0 8 59	60	61 6		3 64	65	Q 66	67 6	8 69	70	0	72	0 C	4 75	0	0	78 7	0 C	
S			11																																																																
-	2 2	2 2	2 2	2 2	2	2 2	2 2	2	2 2	2 2	2			2 2	2	2	2	2	2 2	2	2	2	2 :	2 :	2 2	2	2	2	2	2	2	2 2	2 2	2 2	2	2	2	2 :	2 2	2 2	2	2	2	2	2 2	2 2	2 2	2 2	2 2	2	2 :	2 2	2 2	2	2	2 2	2 2	2	2	2	2 2	2 2	2	2	2	2 2	
01 - N (3 3	3 3	3 3	3 3		3		3	3		3	3	3 3	3 3	3	3		3 3		3	3			3 :	3		3	3	3	3	3	3 3	3 3	3 3	3	3	3	3 :	3 3	3 3	3	3	3	3 :	3 3	3 3	3 3	3 3	3	3	3 :	3 3	3 3	3	3	3 3	3 3	3	3	3	3 3	3 3	3	3	3	3 3	
SIACO	4 4	. 4	4 4	+ 4	4	44	4	4	46	4 4	4	4	4 1	+ 4	4	4	41	4 4	4	4	4	4	4	41	+ 4	4	4	4		4	4	46	• 4	+ 4	4	4	4	4 1	4 4	. 4	4	4	4	4	4 4	4 4	+ 4	+ 4	4	4	4 !	4 4	+ 4	4	4	4 4	+ 4	4	4	4	4 4	. 4	4	4	4	4 4	
	5 5	5 5	5 5	5 5	5	5 5	5	5	5 5	5 5	5	5	5	5	5	5	5 5	5 5	5 5			5	5 !	5 !	5 5	5	5	5	5		5	5 5	5 5	5 5	5	5	5	5 !	5 5	5 5	5	5	5	5	5 5	5 5	5 5	5 5	5	5	5 !	5 5	5 5	5	5	5 5	5 5	5	5	5	5 5	5 5	5	5	5 !	5 5	
	6 6	6 6	6 8	6 6	6	6 6	6	6	6 8	6 6	6	6	6 6	5 6	6	6	6 6	6 6	6	6	6	6	6 1	6 1	6 6	6	6	6	6	6	6	6 8	6 6	5 6	6	6	6	6 1	6 6	6 6	6	6	6	6 1	6 8	6 6	6 6	6 6	6	6	6 (6 6	6 6	6	6	6 6	6 6	6	6	6	6 6	6	6	6	6 1	6 6	
	7 7	77	7 7	7	7	77	7	7	7 7	17	7	7	7 7	17	7	7	7 7	77	7	7	7	7	7		7	7			7	7	7	7 7	77	7	7	7	7	7	7 7	17	7	7	7	7	7 7	7 7	17	1 7	7	7	7	77	17	7	7	7 7	17	7	7	7	77	7	7	7	7	77	
	8 8	3 8	8 8	38	8	8 8	8	8	8 8	8 8	8	8	8	8	8	8	8	3 8		8	8	8	1	8 8	3 8		8	8	8		8	8 8	3 8	3 8	8	8	8	8	3 8	3 8	8	8	8	8	8 8	3 8	3 8	8 8	8	8	8 8	8 8	8 8	8	8	8 8	3 8	8	8	8	8 8	8	8	8	8 8	8 8	
	2 9 9	3	4 5 9 5	6	9	8 9 9 9	10	11 9	12 13	3 14 9 9	15 9	16 1 9	9 9	9 19	20 9	21 2 9	9 9	3 24	25	26 9	27 :	28 î 9	29 3 9 9	9 5	3 32	33	34	35 9	36 9	37	38 : 9	39 4 9 9	0 41 3 5	3 9	43	44	45 J	9 9	7 4	8 49 9 9	50	51 9	52 s	53 5 9 9	9 5	5 5	6 57 9 9	7 58 9 9	3 59 9	60 9	9 9	9 9	3 64 9 9	65 9	66 I	67 6 9 9	8 69	1 70	71 9	72 7 9 1	13 74 9 9	1 75	76 9	9	78 7	9 80	

Outline

- What do I mean by data science
- Data skills
- Case study
- Tools and infrastructure



What do I mean by data science?



When you train as an actuary you'll learn how to analyse data, evaluate financial risks, and communicate this data to non-specialists.

Source: Institute and Faculty of Actuaries



University of Leicester

Curriculum 2019, CM1

- Describe the possible aims of a data analysis (e.g. descriptive, inferential, and predictive).
- Describe the stages of conducting a data analysis to solve real-world problems in a scientific manner and describe tools suitable for each stage.
- Describe sources of data and explain the characteristics of different data sources, including extremely large data sets.
- Explain the meaning and value of reproducible research and describe the elements required to ensure a data analysis is reproducible.



It's always been data science

Age.	Per-	Age.	Per-	Age.	Per-	Age.	Per-	Age.	Per-	Age.	Pet-	Age,	Perfons.
Curt. 1 2 3 4 50	fons. 1000 855 798 760 732	Curt. 9 10 11 12	680 670 661 653 646	Lurt. 15 16 17 18 19	10ns 628 622 616 610 604	Curt. 22 23 24 25 26	fons 585 579 573 567 560	Curt. 29 30 31 32 33	fens 539 531 523 515 507	Curt. 30 37 38	fons 481 472 463 454 445	7 14 28 35 42	5547 4584 4270 3564 3604 3178
7 Age Cort	710 692 Per- fons.	I4 Age. Curt.		21 Age. Curt.	fons	28 Age Curt	fons	34 35 Age. Curt.	fons	42 Age. Curt.	fons	49 56 63 70	2709 2194 1694 1204
43 44 45 45 47 48	417 407 397 387 377 367	50 51 53 53 54 55	34* 335 324 313 302 292	57 58 59 60 61 62	272 262 252 242 232 232	64 65 65 67 68 69	202 192 182 172 162 152	72 73 74 75	131 120 109 98 88 78	111	58 49 41 34 28 23	77 84 100	692 253 107 34000
49	357	56	282	63	212	70	142	77	68	84	20	Sun	n Total.

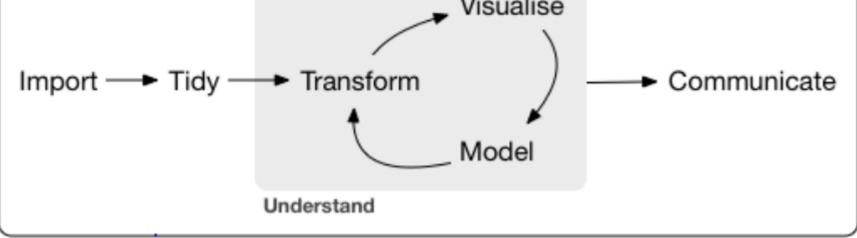


Source: Wikipedia









Program



Data science skills



Case study: from this...

- Clean, structured data
- 5 sheets
- 264 country codes
- 1599 indicators
- 422,136 rows

	5.0.	&			¢	WDIEX	CEL.xls	x - Excel
	le Home	Insert	Page Layou	t Formulas	Data	Review	Vie	w $$
A1	Ŧ	: ×	$\checkmark f_x$	Country N	Name			
			А			В		
1	Country Nam	e	A		C		Code	Indicator Name
_	Arab World	-				RB		2005 PPP conversion factor, GDP (LCU per international \$)
3	Arab World				А	RB		2005 PPP conversion factor, private consumption (LCU per inter
4	Arab World				A	RB		Access to clean fuels and technologies for cooking (% of popula
5	Arab World				A	RB		Access to electricity (% of population)
6	Arab World				A	RB		Access to electricity, rural (% of rural population)
7	Arab World				A	RB		Access to electricity, urban (% of urban population)
8	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
9	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
10	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
11	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
12	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
13	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
14	Arab World				A	RB		Account ownership at a financial institution or with a mobile-m
15 (Arah World Dat	a Coun	try Series	Country-S		RR Series-1	Гime	Account ownershin at a financial institution or with a mobile-m
Read	y							III II - + 100



Case study: to this (reproducibly)...

Income_Group	Region <fctr></fctr>	Av_Fem_Life_Exp <dbl></dbl>	
High income	East Asia & Pacific	83.61193	
Upper middle income	East Asia & Pacific	77.08483	
Low income	East Asia & Pacific	75.07100	
Lower middle income	East Asia & Pacific	71.70246	
High income	Europe & Central Asia	83.34810	
Upper middle income	Europe & Central Asia	77.75850	
Lower middle income	Europe & Central Asia	75.48883	
Low income	Europe & Central Asia	74.18700	•
High income	Latin America & Caribbean	80.21331	
Upper middle income	Latin America & Caribbean	77.25928	
	ی ارت - و0 -		Income_Group High income Low income Lower middle income Upper middle income

Attended



Case study: skills required

- Read in data
- Change data type
- Select and filter
- Search using regular expressions
- Reshape
- Plot and categorise
- Group and summarise



Case study



University of Leicester

Module outline

- Introduction & infrastructure.
 Reproducible workflows and collaborative working.
- Reading tabular files (CSV, Excel). Data structures: data frames and vectors. Simple plots.
- Tidy data: wide vs tall tables: pivoting
- Calculations on tabular data
- Visualizing data

- Putting it together a first complete project
- Checking data; data ethics, governance, and regulation.
- Non-tabular data (XML, JSON, text)
- Working with databases: SQL and relational databases; noSQL types & uses
- Big data tools



Tools and infrastructure



Tools

- R
- Rstudio
- Rmarkdown
- Shiny
- Leaflet
- GitHub
- Bookdown / Blogdown



Bibliography

- R for Data Science
 https://r4ds.had.co.nz/index.html
- Efficient R programming
 https://bookdown.org/csgillespie/efficientR/
- R Markdown: The Definitive Guide https://bookdown.org/yihui/rmarkdown/
- Geocomputation with R
 <u>https://geocompr.robinlovelace.net/</u>
- See the Bookdown site
 <u>https://bookdown.org/</u>





Data Science for Actuaries

Paul King, ATRC 28 June 2019

					1	Ĥ			- N	CI			22	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	ĤI		-1 3	1.1.	freed.	н.		·····		ž.L.	X	100	R*.		X		1									j															1				6						1				
						1					1										-	1						-																																									
	0 0	0 0	0	0 0	0			0 0						0	0	0 1	0	0	0					0		0						0 0	0	0	0	0 0	0 0	0	0	0	0	0 1	0 0	0 0	0 0	0	0	0	0	0 0	0 0	0	0	0	0 0		0 0	0	0	0 (0 0	0 0	0	0	0	0	0 1	0	
	2																																																																				
-	2 2	2 2	2	2 2	2	2 2	2 2	2 2	2	2	2 2	2		2	2	2 :	2 2	2	2	2	2	2 1	2 2	2 2	2	2	2	2	2	2 :	2 2	2 2	2	2	2	2 2	2 2	2	2	2	2	2	2 :	2 2	2	2	2	2	2	2 2	2 2	2	2	2	22	2 2	2	2	2	2 :	2 :	2 2	2	2	2	2	2	2	
01 - N (3 3	3 3	3	3 3		3		3	3		3	3 3	3	3	3	3 :	3	3	3		3	3		3	3			3	3	3 :	3 3	3 3	3	3	3	3 3	3 3	3	3	3	3	3 :	3 :	3 3	3	3	3	3	3	3 3	3 3	3	3	3	3 3	3 3	3	3	3	3 3	3 3	3 3	3	3	3	3	3	3	
SIACO	4 4	4.4	4	4 4	4	41	• 4	4	4	4	44	. 4	. 4	4	4	41	4 4	, 4	4	4	4	4 6	4 4	4	4	4	4	4	4		4 4	. 4	4	4	4	4 1	4 4	4	4	4	4	4	4 4	4 4	4	4	4	4	4	4 4	4	4	4	4	44	+ 4	. 4	4	4	4 1	4 1	. 4	4	4	4	4	4	4	
	5 5	5 5	5	5 5	5	5 5	5 5	5 5	5	5	5 5	5 5	5		5	5 !	5 5	5 5	5	5			5 5	5 5	5	5	5	5	5	5		5.5	5	5	5	5 5	5 5	5	5	5	5	5 !	5 5	5 5	5	5	5	5	5	5 5	5 5	5	5	5	5 5	5 5	5	5	5	5 !	5 !	5 5	5	5	5	5	5 :	5	
	6 8																																																																				
	77																				1																																																
	88	2 3	4	5 6	7	8 9	10	1 11	12	13 1	4 15	5 16	17	18	19 2	0 2	1 22	2 23	24	25	26 2	7 2	8 29	3 30	31	32	33	34	35 3	6 3	1 3	8 39	40	41	42 0	13 4	4 45	46	47	48	49	50 5	1 5	2 53	3 54	55	56	57	58 5	6 6	0 61	62	63 (64 8	5 6	6 67	68	69	70	71 7	27	3 74	75	76	77	78 7	79 8	80	
	9 9	99	9	9 9	9	9 9	3 9	9 9		9	99	3 9	9	9	9	9 9	9 5	9	9	9	9	9 9	3 5	9	9	9	9	9	9 !	9 5	3 5	9 9	9	9	9	9 9	3 9	9	9	9	9	9 :	3 5	3 9	9	9	9	9	9 9	9 5	9 9	9	9	9 :	9 9	3 9	9	9	9	9 9	9 9	19	9	9	9	9 :	9 9	3	