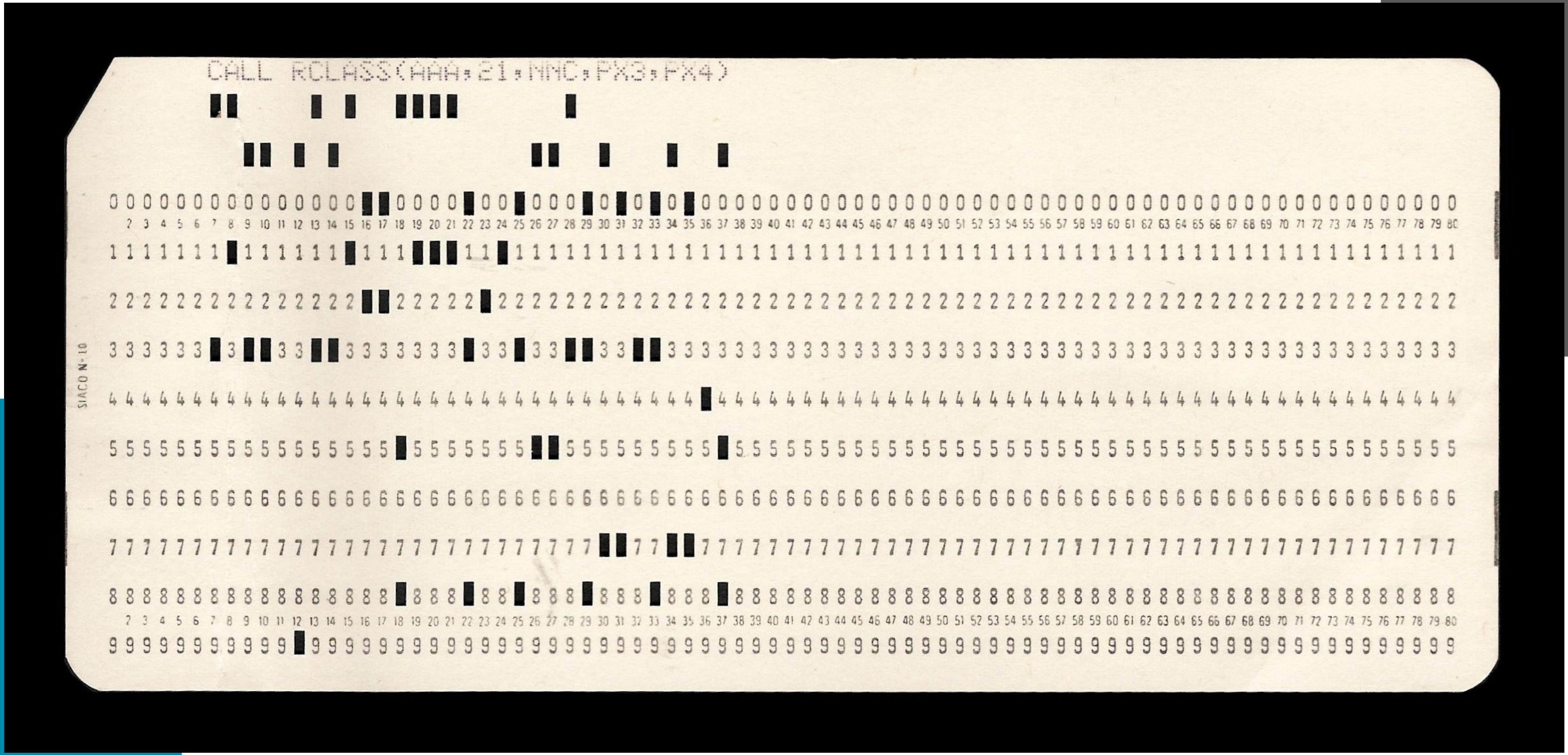
UNIVERSITY OF LEICESTER

Department of Mathematics

Data Science for Actuaries

Paul King, ATRC 28 June 2019



```
CALL RCLASS(AAA,21,NNC,PX3,PX4)
```

Outline

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

What do I mean by data science?

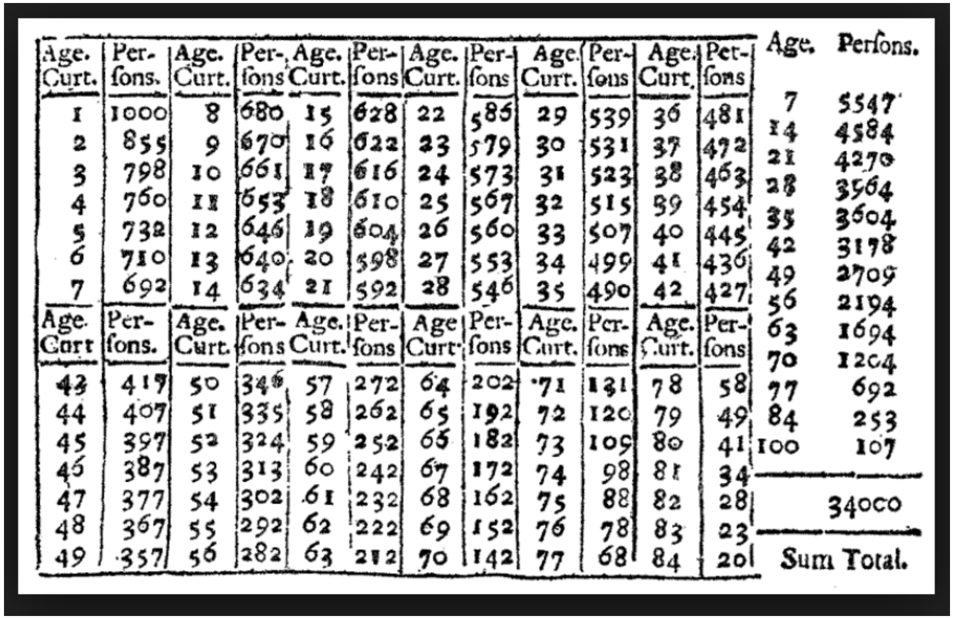
When you train as an actuary you’ll learn how to analyse data, evaluate financial risks, and communicate this data to non-specialists.

Source: Institute and Faculty of Actuaries

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

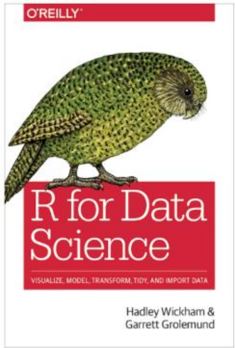


Age. Curt.	Perſons.	Age. Curt.	Perſons	Age. Curt.	Perſons	Age. Curt.	Perſons	Age. Curt.	Perſons	Age. Curt.	Petſons
1	1000	8	680	15	628	22	585	29	539	36	481
2	855	9	670	16	622	23	579	30	531	37	472
3	798	10	661	17	616	24	573	31	523	38	463
4	760	11	653	18	610	25	567	32	515	39	454
5	732	12	646	19	604	26	560	33	507	40	445
6	710	13	640	20	598	27	553	34	499	41	436
7	692	14	634	21	592	28	546	35	490	42	427
Age. Curt	Perſons.	Age. Curt.	Perſons	Age. Curt.	Perſons	Age Curt	Perſons	Age. Curt.	Perſons	Age. Curt.	Perſons
43	417	50	346	57	272	64	202	71	131	78	58
44	407	51	335	58	262	65	192	72	120	79	49
45	397	52	324	59	252	66	182	73	109	80	41
46	387	53	313	60	242	67	172	74	98	81	34
47	377	54	302	61	232	68	162	75	88	82	28
48	367	55	292	62	222	69	152	76	78	83	23
49	357	56	282	63	212	70	142	77	68	84	20

Age.	Perſons.
7	5547
14	4584
21	4270
28	3964
35	3604
42	3178
49	2709
56	2194
63	1694
70	1204
77	692
84	253
100	107
	34000
Sum Total.	

Source: Wikipedia

The data science process

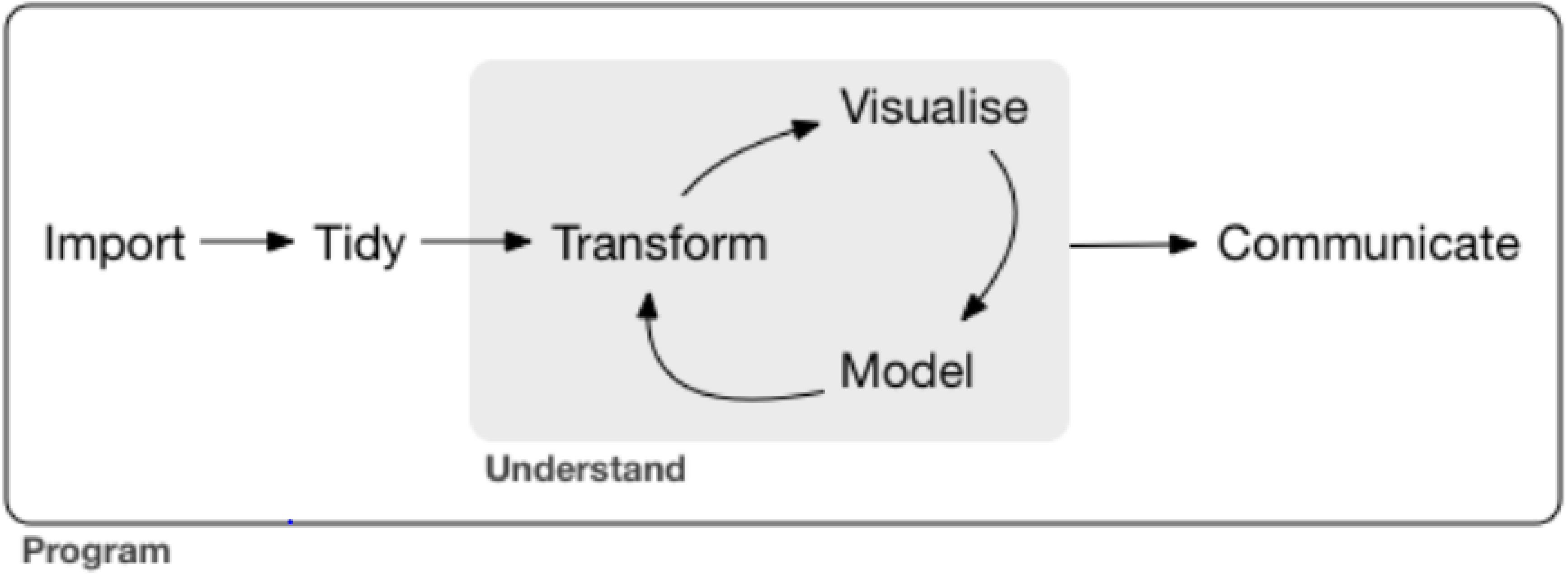


O'REILLY

R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

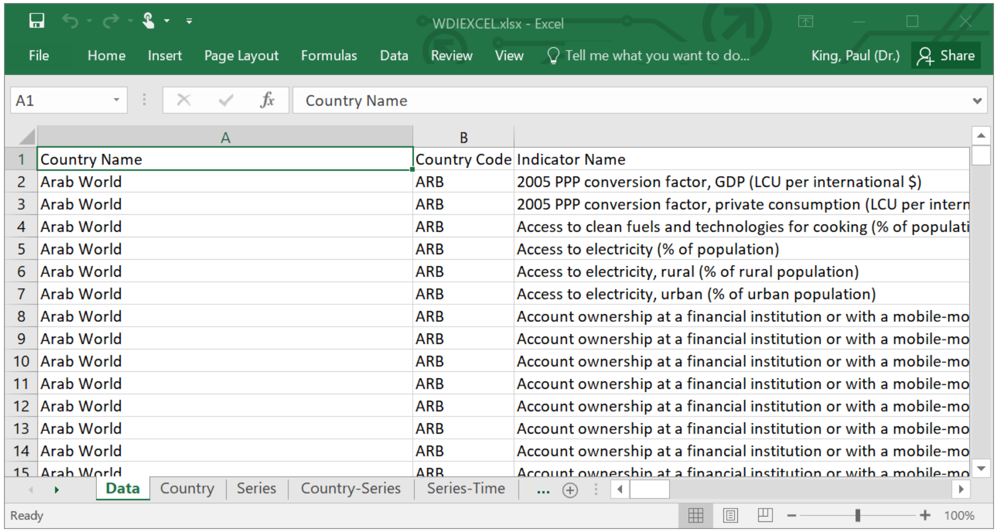
Hadley Wickham & Garrett Grolemund



Data science skills

Case study: from this…

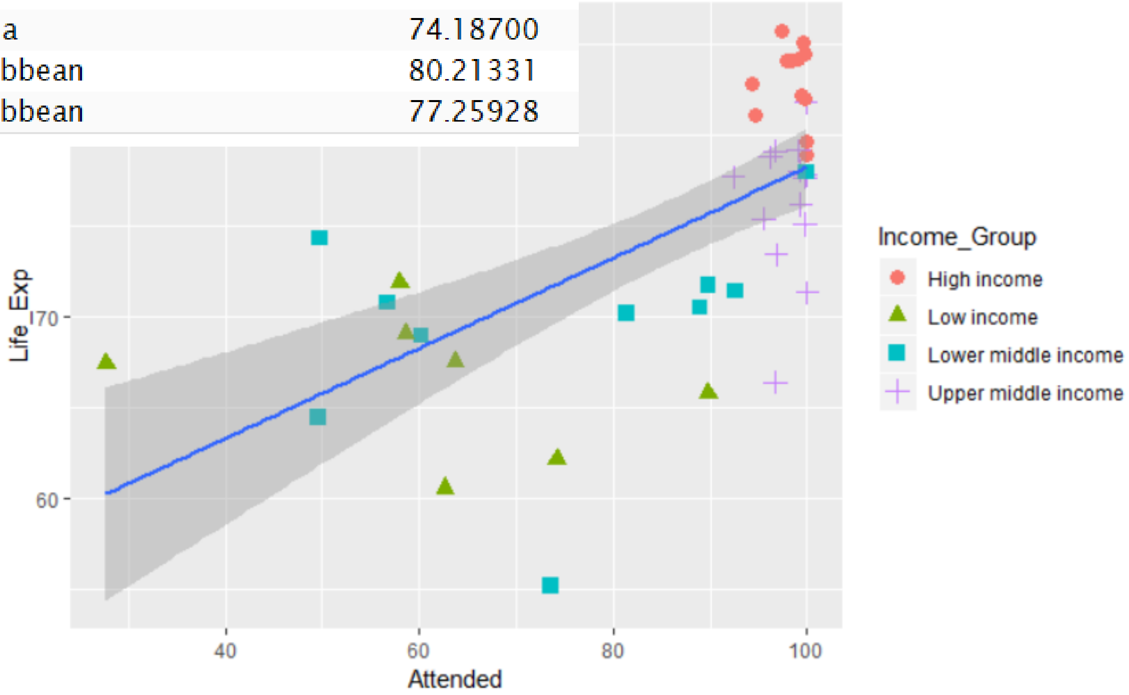
- Clean, structured data
- 5 sheets
- 264 country codes
- 1599 indicators
- 422,136 rows



	A	B	
1	Country Name	Country Code	Indicator Name
2	Arab World	ARB	2005 PPP conversion factor, GDP (LCU per international $)
3	Arab World	ARB	2005 PPP conversion factor, private consumption (LCU per intern
4	Arab World	ARB	Access to clean fuels and technologies for cooking (% of populati
5	Arab World	ARB	Access to electricity (% of population)
6	Arab World	ARB	Access to electricity, rural (% of rural population)
7	Arab World	ARB	Access to electricity, urban (% of urban population)
8	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
9	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
10	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
11	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
12	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
13	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
14	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo
15	Arab World	ARB	Account ownership at a financial institution or with a mobile-mo

Case study: to this (reproducibly)…

Income_Group <fctr>	Region <fctr>	Av_Fem_Life_Exp <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



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

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

UNIVERSITY OF
LEICESTER

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
  https://geocompr.robinlovelace.net/
- See the Bookdown site
  https://bookdown.org/

UNIVERSITY OF
LEICESTER

UNIVERSITY OF LEICESTER

Department of Mathematics

Data Science for Actuaries

Paul King, ATRC 28 June 2019