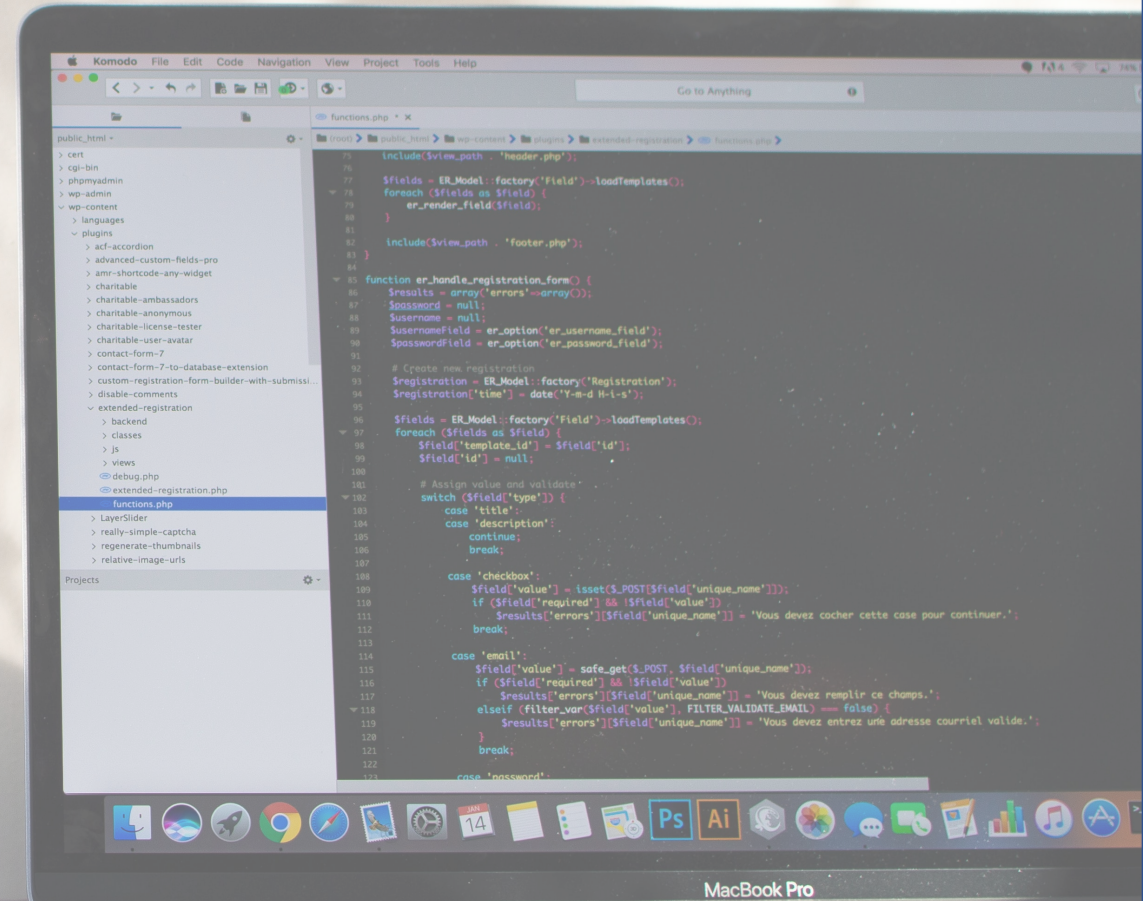


Principles of Spatial Analysis

SHORT LECTURE 01, WEEK 05: EXPLORATORY (SPATIAL) DATA ANALYSIS



Tidy Data

80 percent

of your time goes to data cleaning and preparation ('data wrangling')

3 principles

of what we can call 'tidy data'

country	year	cases	population
Afghanistan	1999	1745	19957071
Afghanistan	2000	2666	20595360
Brazil	1999	31737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	21166	128028583

Each variable must have its own column

country	year	cases	population
Afghanistan	1999	745	19987000
Afghanistan	2000	2000	20000000
Brazil	1999	57707	172000000
Brazil	2000	60400	174000000
China	1999	212200	127201021
China	2000	210700	128042000

Each observation must have its own row

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20594360
Brazil	1999	37737	172004362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

Each value must have its own cell

ukmidyear estimates Population 2019 2020

	A	B	C	D	E	F	G	H	I	J	K	L
1	Contents											
2												
3	MYE2: Population estimates: Persons by single year of age and sex for local authorities in the UK, mid-2019							Please click to e-mail us your opinion:		This met my needs, please produce it next year	I need something different (please)	
4												
5	Code	Name	Geography¹	All ages	0	1	2	3	4	5	6	7
6	K02000001	UNITED KINGDOM	Country	66 796 807	722 881	752 554	777 309	802 334	802 185	809 152	827 149	852 059
7	K03000001	GREAT BRITAIN	Country	64 903 140	700 160	729 146	753 103	777 260	777 225	784 154	801 776	825 785
8	K04000001	ENGLAND AND WALES	Country	59 439 840	649 388	676 412	698 837	720 721	719 821	726 317	742 744	765 225
9	E92000001	ENGLAND	Country	56 286 961	618 858	644 056	665 596	686 135	684 992	691 122	706 742	727 938
10	E12000001	NORTH EAST	Region	2 669 941	26 621	27 612	28 621	29 575	29 315	30 224	30 960	31 956
11	E06000047	County Durham	Unitary Authority	530 094	4 890	5 085	5 292	5 483	5 597	5 826	5 993	5 978
12	E06000005	Darlington	Unitary Authority	106 803	1 100	1 152	1 112	1 218	1 208	1 257	1 317	1 368
13	E06000001	Hartlepool	Unitary Authority	93 663	1 006	1 009	1 031	1 125	1 058	1 126	1 147	1 198
14	E06000002	Middlesbrough	Unitary Authority	140 980	1 802	1 944	1 979	1 886	1 972	1 957	1 990	1 969
15	E06000057	Northumberland	Unitary Authority	322 434	2 665	2 952	2 996	3 146	3 006	3 129	3 346	3 406
16	E06000003	Redcar and Cleveland	Unitary Authority	137 150	1 313	1 372	1 508	1 421	1 492	1 589	1 692	1 694
17	E06000004	Stockton-on-Tees	Unitary Authority	197 348	2 149	2 131	2 337	2 366	2 479	2 418	2 567	2 695
18	E11000007	Tyne and Wear (Met County)	Metropolitan County	1 141 469	11 696	11 967	12 366	12 930	12 503	12 922	12 908	13 648
19	E08000037	Gateshead	Metropolitan District	202 055	2 005	1 981	2 133	2 167	2 222	2 260	2 139	2 295
20	E08000021	Newcastle upon Tyne	Metropolitan District	302 820	3 244	3 220	3 310	3 483	3 269	3 400	3 544	3 588
21	E08000022	North Tyneside	Metropolitan District	207 913	2 233	2 259	2 244	2 386	2 293	2 439	2 361	2 348
22	E08000023	South Tyneside	Metropolitan District	150 976	1 495	1 635	1 619	1 828	1 662	1 721	1 740	1 829
23	E08000024	Sunderland	Metropolitan District	277 705	2 719	2 872	3 060	3 066	3 057	3 102	3 124	3 588
24	E12000002	NORTH WEST	Region	7 341 196	81 258	83 359	86 681	89 238	89 101	90 059	90 982	93 708
25	E06000008	Blackburn with Darwen	Unitary Authority	149 696	2 029	2 041	2 105	2 208	2 192	2 145	2 195	2 342
26	E06000009	Blackpool	Unitary Authority	139 446	1 597	1 601	1 703	1 650	1 711	1 696	1 768	1 783
27	E06000049	Cheshire East	Unitary Authority	384 152	3 646	4 060	4 104	4 302	4 195	4 183	4 431	4 608
28	E06000050	Cheshire West and Chester	Unitary Authority	343 071	3 348	3 547	3 726	3 833				

column headers are values
not variable names

multiple values
are stored in one column

variables

are stored in both rows and columns

multiple observational units
are stored in the same column

a single observation
is stored in multiple tables

country	year	type	count
Afghanistan	2019	cases	745
Afghanistan	2019	population	19 987 071
Afghanistan	2020	cases	2 666
Afghanistan	2020	population	20 595 360
Brazil	2019	cases	3,7737
Brazil	2019	population	172 006 362
Brazil	2020	cases	80 488
Brazil	2020	population	174 504 898
China	2019	cases	212 258
China	2019	population	1 272 915 272
China	2020	cases	213 766
China	2020	population	1 280 428 583

country	year	rate
Afghanistan	2019	745 / 19,987,071
Afghanistan	2020	2,666 / 20,595,360
Brazil	2019	3,7737 / 172,006,362
Brazil	2020	80,488 / 174,504,898
China	2019	212,258 / 1,272,915,272
China	2020	213,766 / 1,280,428,583

Cases

country	2019	2020
Afghanistan	745	2 666
Brazil	3,7737	80 488
China	212 258	213 766

Population

country	2019	2020
Afghanistan	19 987 071	20 595 360
Brazil	172 006 362	174 504 898
China	1 272 915 272	1 280 428 583

let's make it tidy

country	year	type	count
Afghanistan	2019	745	19 987 071
Afghanistan	2020	2 666	20 595 360
Brazil	2019	3,7737	172 006 362
Brazil	2020	80 488	174 504 898
China	2019	212 258	1 272 915 272
China	2020	213 766	1 280 428 583

let's put it into practice