World Cities

Figure A.2 has two layers:

1. Country borders
2. A scatter plot mapping the urban population of each country to the area of a dot.

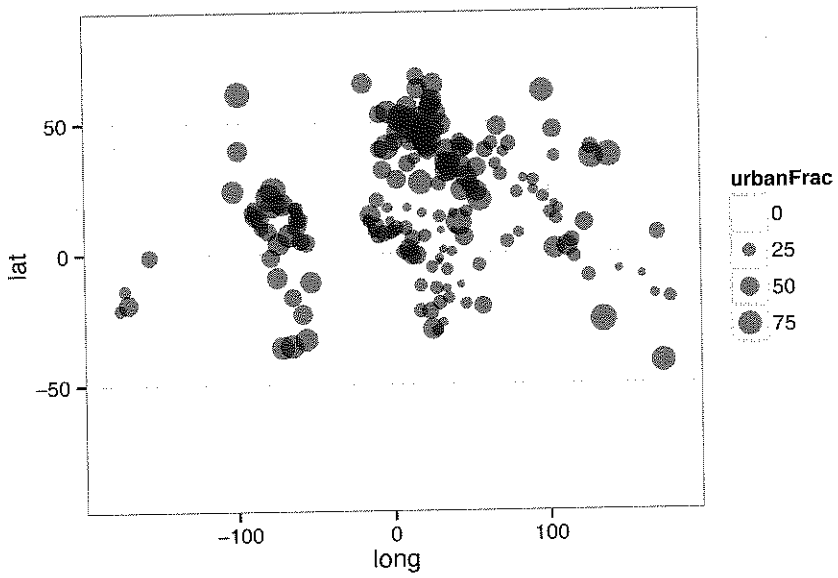

Figure A.2: The end result for this project: each country's urban population fraction shown geographically.

The data behind the graphic in Figure A.2 comes from distinct data tables.

- `WorldCities` gives the population of each city.
- `CountryData` gives, among other things, the population of each country.
- `CountryCentroids` gives the latitude and longitude of the center of each country.

`WorldCities`

code	country	population
3040051	AD	15853
3041563	AD	20430
290594	AE	44411
291074	AE	115949
291696	AE	33575
... and so on for 23,018 rows		

`CountryData`

country	area	pop
Afghanistan	652230	31822848
Akrotiri	123	15700
Albania	28748	3020209
Algeria	2381741	38813722
American Samoa	199	54517
... and so on for 256 rows		

- `Codes` translates between ISO-A2 and ISO-A3 codes.

Your task is to join these sources of information together to create a glyph-ready data table suited for making the scatterplot layer of Figure A.2.

`CountryCentroids`

name	iso_a3	long	lat
Afghanistan	AFG	66.17	33.78
Aland	ALA	19.97	60.20
Albania	ALB	20.26	41.14
Algeria	DZA	2.83	28.14
American Samoa	ASM	-170.72	-14.30
... and so on for 241 rows			

`Codes`

ISO2	ISO3
AF	AFG
AL	ALB
DZ	DZA
AD	AND
AO	AGO
... and so on for 194 rows	

Envision the Output

Analyze Figure A.2 to determine a suitable form for glyph-ready data:

- What variables form the frame?
- What is the glyph? (Ignore the country borders, which are really a guide, not a glyph.)
- What graphical properties does the glyph have? Which variables are mapped to those properties?

Sketch out the form of a glyph-ready data table.

- What are the variables?
- What is the physical meaning of a case?

Write the variable names and a case or two of made-up data in the usual rectangular format.

Origins of the variables

Each of the variables in the glyph-ready table originates in one of the four "raw" tables. Determine which table or tables contains the information needed to generate each variable in the glyph-ready table.

The meaning of a case

The meaning of "case" in the glyph-ready data matches that of three of the four original tables. Which table doesn't have the same meaning for case as the glyph-ready data?

Joining the tables

You're starting with four tables. To end up with the single glyph-ready data, you'll need to perform several joins, each of which involves two tables. For each of the joins,

- Specify the two tables to be involved.
- Name the variables from each table to be used for matching.

Give a name to the output table. (They are labelled A, B, and C below.)

Output Name	Table 1 Name	Table 2 Name	Variables to match
A			
.			
B			
.			
C			