

### 12.3.3 Choropleth maps

Choropleth maps (8.5.1) are helpful for visualizing geographic data. In this example, we use data from the built-in R dataset, `USArrests`, which includes United States arrests in 1973 per 100,000 inhabitants in various categories by state.

To use the data in SAS, we'll save it to an external file in R. We'll use the Stata format for the external file (see 1.2.6).

```
> library(foreign)
> USArrests.st =
  transform(USArrests, region=tolower(rownames(USArrests)))
> write.dta(USArrests.st, "c:\\book\\USArrests.dta",
  convert.factors="string")
```

Note that the state names are provided only as the row names in the R dataframe. It's easier to get them out of R if they are instead stored as a variable. The `transform()` function above adds them to the dataset (2.2). The default is to convert string variables to numbers with Stata labels containing the strings. These are correctly imported to SAS as value labels, but value labels are not useful for our purposes. The `convert.factors="string"` option is used to retain the state names directly instead. Then we can read it into SAS (see 1.1.9).

```
proc import datafile="C:\book\usarrests.dta"
  out=usarrests dbms=dta replace;
run;

proc print data=usarrests (obs=5); run;
```

			Urban		
Obs	Murder	Assault	Pop	Rape	region
1	13.2	236	58	21.2	alabama
2	10.0	263	48	44.5	alaska
3	8.1	294	80	31.0	arizona
4	8.8	190	50	19.5	arkansas
5	9.0	276	91	40.6	california

To make the map, we'll use a built-in US map provided with SAS. This comes with the two-letter US postal codes identifying the states. To match with the lower-case state names in the input dataset, we'll use the `stname1` function to convert the postal codes to long names and the `lowcase` function (2.2.17) to match the values coming from R.

Then the `gmap` procedure makes the choropleth.

```
pattern1 v=s c=grayff;
pattern2 v=s c=grayda;
pattern3 v=s c=grayaa;
pattern4 v=s c=gray68;
pattern5 v=s c=gray22;
data mymap;
set maps.us;
  region = lowcase(stname1(statecode));
run;

proc gmap data=usarrests map=mymap;
  id region;
  choro murder / levels=5;
run; quit;
```

Note that the map dataset and the plot dataset remain separate, but are linked by a commonly named variable specified in the `id` statement. The `pattern` statements change the colors from the default blue shades to print-friendly grays. SAS maps are stored with variables `x` and `y` describing the boundary points. If longitude and latitude values are available, a variety of projections can be applied using the `gproject` procedure.

The results are displayed in Figure 12.4.

In R, we'll use the `ggmap` package. Its functions build on the `ggplot2` package, which implements ideas related to the “grammar of graphics” [196]. The package uses a syntax where specific elements of the plot are added to the final product using special functions connected by the `+` symbol. Some additional work is needed to merge the dataset with the state information (2.3.11) and to sort the resulting dataframe (2.3.10) so that the shape data for the states is plotted in order.

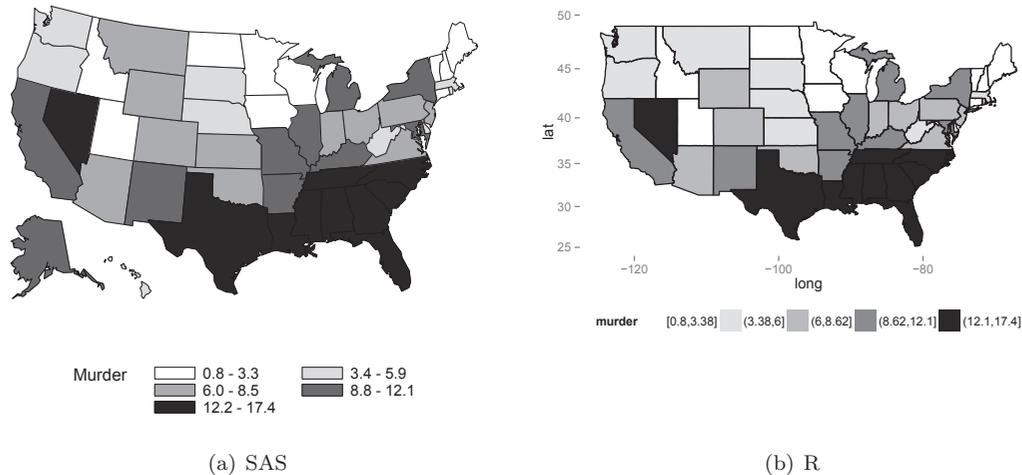


Figure 12.4: Choropleth map

```

> library(ggmap)
> USArrests.st = transform(USArrests.st, murder = cut_number(Murder, 5))
> us_state_map = map_data('state')
> map_data = merge(USArrests.st, us_state_map, by="region")
> map_data = map_data[order(map_data$order),]
> p0 = ggplot(map_data, aes(x=long, y=lat, group=group)) +
  geom_polygon(aes(fill = murder)) +
  geom_path(colour='black') +
  theme(legend.position = "bottom",
        panel.background=element_rect(fill="transparent",
        color=NA)) +
  scale_fill_grey(start=1, end =.1) + coord_map();
> plot(p0)

```

The `scale_fill_grey()` function changes the colors from the default unordered multiple colors to an ordered and print-friendly black and white (see also `scale_file_brewer`). The `ggmap` package uses the Mercator projection (see `coord_map()` in the `ggplot2` package and `mapproject` in the `mapproject` package).

Note that the binning algorithms used by the SAS `gmap` procedure and the `cut_number()` function differ slightly, so that some states are shifted. As always, the choice of groupings can have an impact on the message conveyed by the graphical display.