Paper RV115-2019
Graphics for Univariate Data: Pie is Delicious but Not Nutritious

Peter L. Flom, Peter Flom Consulting, New York, NY

ABSTRACT

When you have univariate data, that is, a single measure on a variety of units, the most common statistical graphic is a pie chart. But pie charts should not be used. Ever. When there are a lot of units, pie charts are unreadable. When there are only a few units, pie charts waste space. And research Cleveland (1993, 1994) shows that, even with a moderate number of units, pie charts can distort the data (for example, using different colors leads to different estimates of the size of the wedges). Fortunately, there are better methods.

Keywords: pie dot univariate graphics.

WHY USE A GRAPH FOR UNIVARIATE DATA?

A statistical graphic is not meant to be a reproduction of a table. Tables are good for looking things up, graphics are good for looking around. In this paper, I use data on the populations of different states of the United States. If I needed to know (for example) the exact population of Montana, I could look it up. For that, a table would be ideal. But if I wanted to know how the population of Montana relates to all the other states, a graphic is better. It is hard to look at 50 lines in a table and get a good sense of the relationships.

WHY PIE CHARTS ARE BAD

Pie charts are used to display univariate categorical data; that is, data on the frequency of various possible levels of a single variable. For example, you may wish to display the population of different regions (states or larger regions) of the United States. Sometimes, the variable can take many levels (e.g. 52 states plus DC and Puerto Rico), sometimes a moderate number (e.g. the nine census divisions) and sometimes only a small number of levels (e.g. the four census divisions). In each case, there are better ways to show the data than with pie charts.

With 52 categories you can use this code:

```
ods pdf file = "c:\personal\presentations\Graphics\pie50.pdf";
ods graphics/height = 11in width = 8.5in;
proc gchart data = statepop;
 pie geographical_area/sumvar = pop08 legend other = 0 noheading;
run;
ods pdf close
```

To get something like Figure 1.

Nothing need be said. That's unreadable. (Can you find Maryland's slice?)

With a more reasonable number of categories, the pie chart is at least readable (see figure 2)

However, research has shown that changing the order of the categories and changing their color can influence perception of how large each portion is. A table may be better, but there are better graphical alternatives, as well.

With even fewer categories (e.g. 4) the pie chart is a waste of space, and it is better to use a table or even text.

A BETTER ALTERNATIVE: THE DOT CHART

A Cleveland dot chart Cleveland (1993, 1994) is clearly a better alternative for moderate or large number of categories. You can produce one in SAS with the following code:

```
ods pdf path="c:\personal\Graphics\dotv1.pdf";
ods graphics/height = 11in width = 8.5in;
proc sgplot data = divisionpop;
```

```
    dot geographical_area/response = pop08 nostatlabel;
    yaxis discreteorder = data;
 run;
 ods pdf close;
```

Which produces Figure 3.

But with a little work, we can produce a much better version; this is still trying to reproduce a table, which isn't what we want to do. First, order the data from largest to smallest and create a new variable of population in millions:

```
proc sort data = divisionpop out = divisionpop2;
 by pop08;
run;
data divisionpop3;
 set divisionpop2;
  pop08millions = pop08/1000000;
run;
```

Producing Figure 4.

With data like this, a log scale may be preferable, because division population spans a broad range and because we may be more interested in the ratio of populations than the arithmetical difference. The difference between the two largest divisions (South Atlantic and Pacific) is about 10 million people, but this seems to mean less than a similar difference between two smaller divisions. (This point will be even clearer when we look at population by state).

```
ods pdf file ="c:\personal\presentations\Graphics\dotv3.pdf";
proc sgplot data = divisionpop3;
  dot geographical_area/response = pop08millions nostatlabel;
  yaxis discreteorder = data label = 'Census division';
  xaxis type = log logbase = 10 logstyle = linear
    label = 'Population of division in millions (log scale)';
 run;
ods pdf close;
```

Which produces Figure 5.

But when we try something similar with the state data, we have a problem: The default type size is too big, and only half the states are labeled! Also, associated grid lines have been removed, making these states even less visible. The simplest version also highlights some other problems, as shown in Figure 6.

First we apply the earlier fixes, leaving the problem of the labels. We can fix this with the following code:

```
/* Reduce the font size of the GraphValueFont */
proc template;
  define Style styles.mystyle;
  parent = styles.default;
    style GraphFonts from GraphFonts
      "Fonts used in graph styles" /
      'GraphTitleFont' = (", ",10pt,bold)
      'GraphFootnoteFont' = (", ",8pt)
      'GraphLabelFont' = (", ",8pt)
      'GraphValueFont' = (", ",4pt)
      'GraphDataFont' = (", ",8pt);
  end;
run;

/* Increase the height of the graph. */
ods graphics / height=600px;
```

Now, use SGPLOT with the new style:

```
ods pdf style= styles.mystyle file = "c:\personal\presentations\Graphics\dotv5.pdf";
proc sgplot data = statepop3;
  dot geographical_area/response = pop08million nostatlabel;
  yaxis discreteorder = data label = 'State';
  xaxis label = 'Population (in millions)';
run;
ods pdf close;
```

Producing Figure 7.

Or, with a log scale:

```
ods pdf file="c:\personal\presentations\Graphics\dotv6.pdf" style=styles.mystyle;
proc sgplot data = statepop3;
  dot geographical_area/response = pop08million nostatlabel;
  yaxis discreteorder = data label = 'State';
  xaxis type = log logbase = 10 logstyle = linear label = "Population (millions, log scale)"
run;
ods pdf close;
```

Producing Figure 8.

CONCLUSIONS

Pie charts are not a good method for displaying univariate data. Better methods are available.

CONTACT INFORMATION

Peter L. Flom
515 West End Ave
Apt 8C
New York, NY 10024 peterflomconsulting@mindspring.com
(917) 488 7176

ACKNOWLEDGEMENTS

SAS® and all other SAS Institute Inc., product or service names are registered trademarks ore trademarks of SAS Institute Inc., in the USA and other countries. ® indicates USA registration. Other brand names and product names are registered trademarks or trademarks of their respective companies.

REFERENCES

Cleveland, W. S. (1993), *Visualizing Data*, Hobart Press, Princeton, NJ.

Cleveland, W. S. (1994), *The Elements of Graphing Data*, Hobart Press, Princeton, NJ.

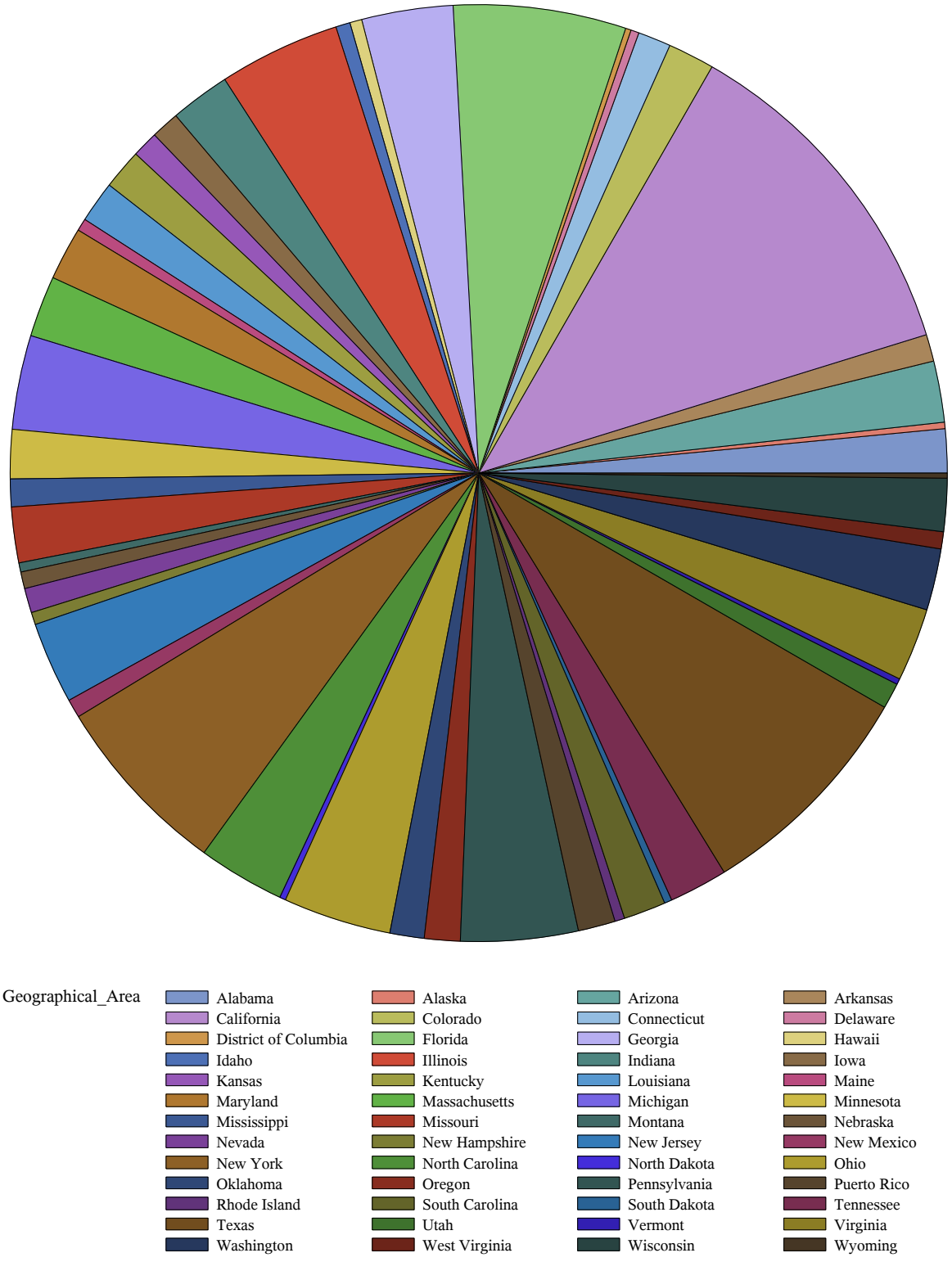

Figure 1: Pie chart with 51 categories

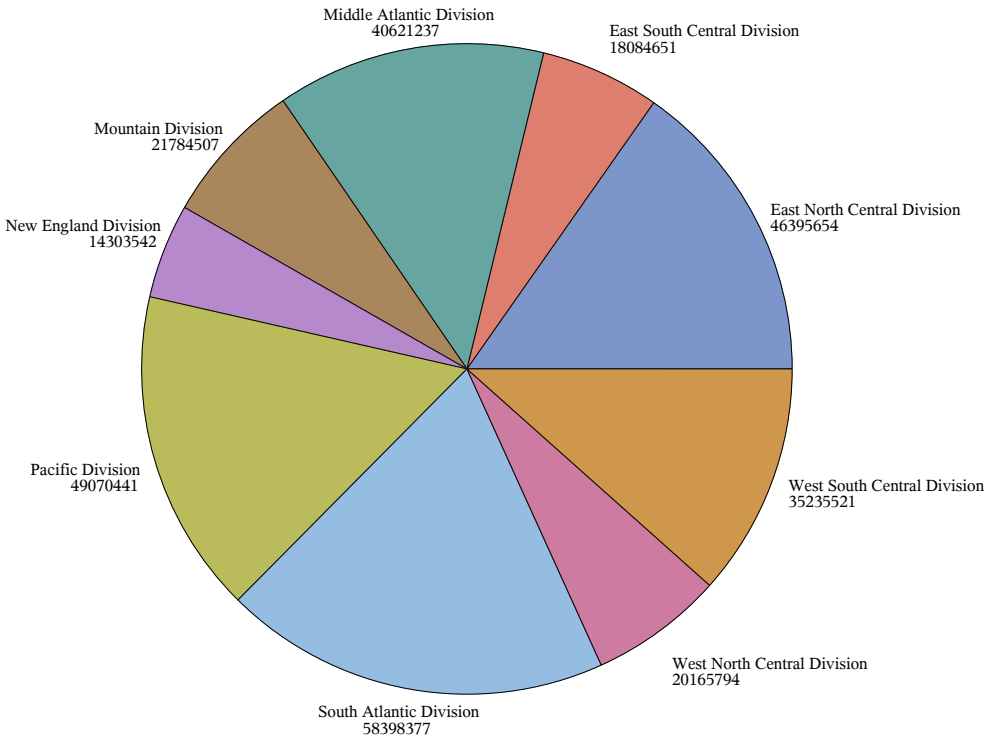

Figure 2: Pie chart with 9 categories

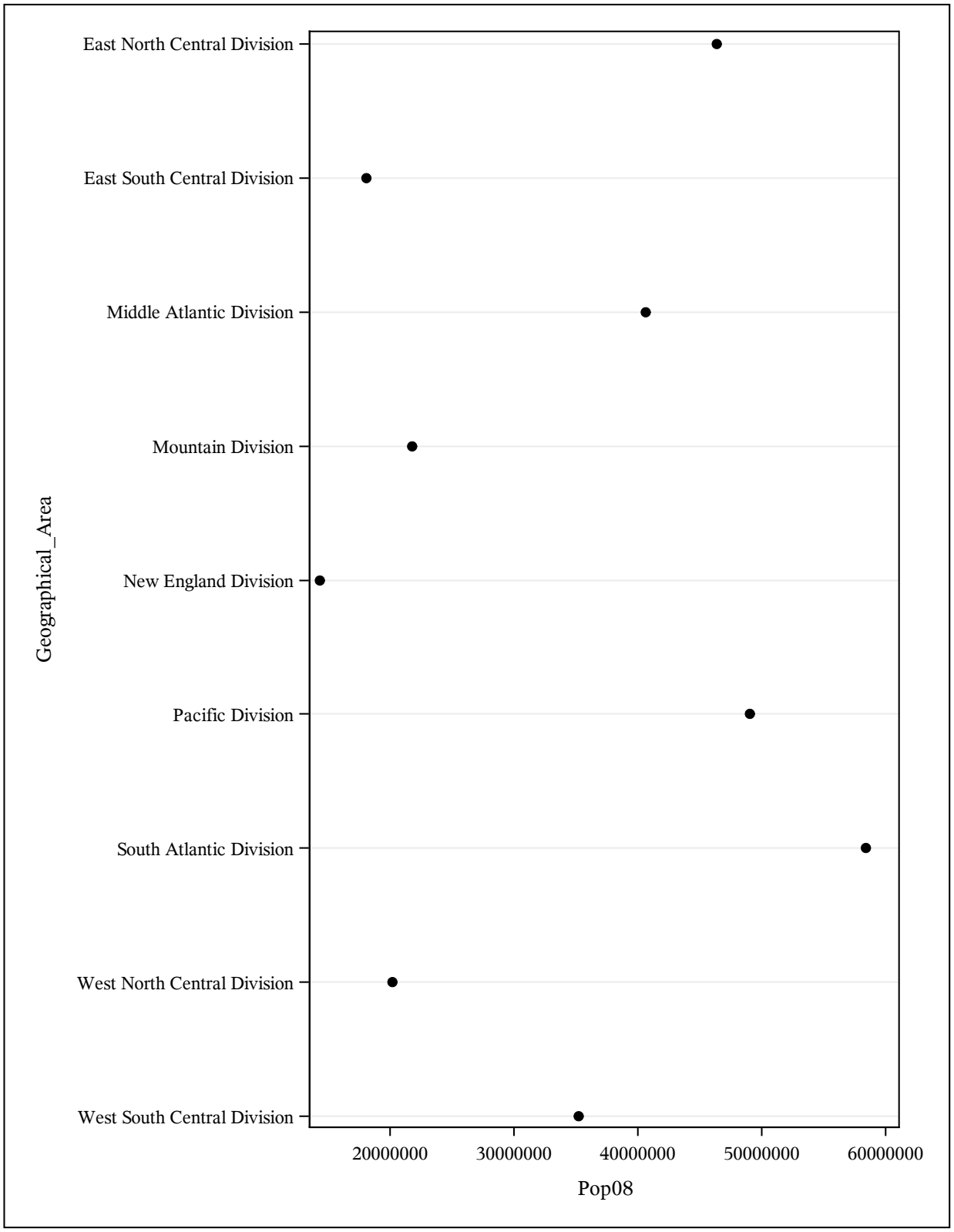

Figure 3: Dot chart chart with 9 categories

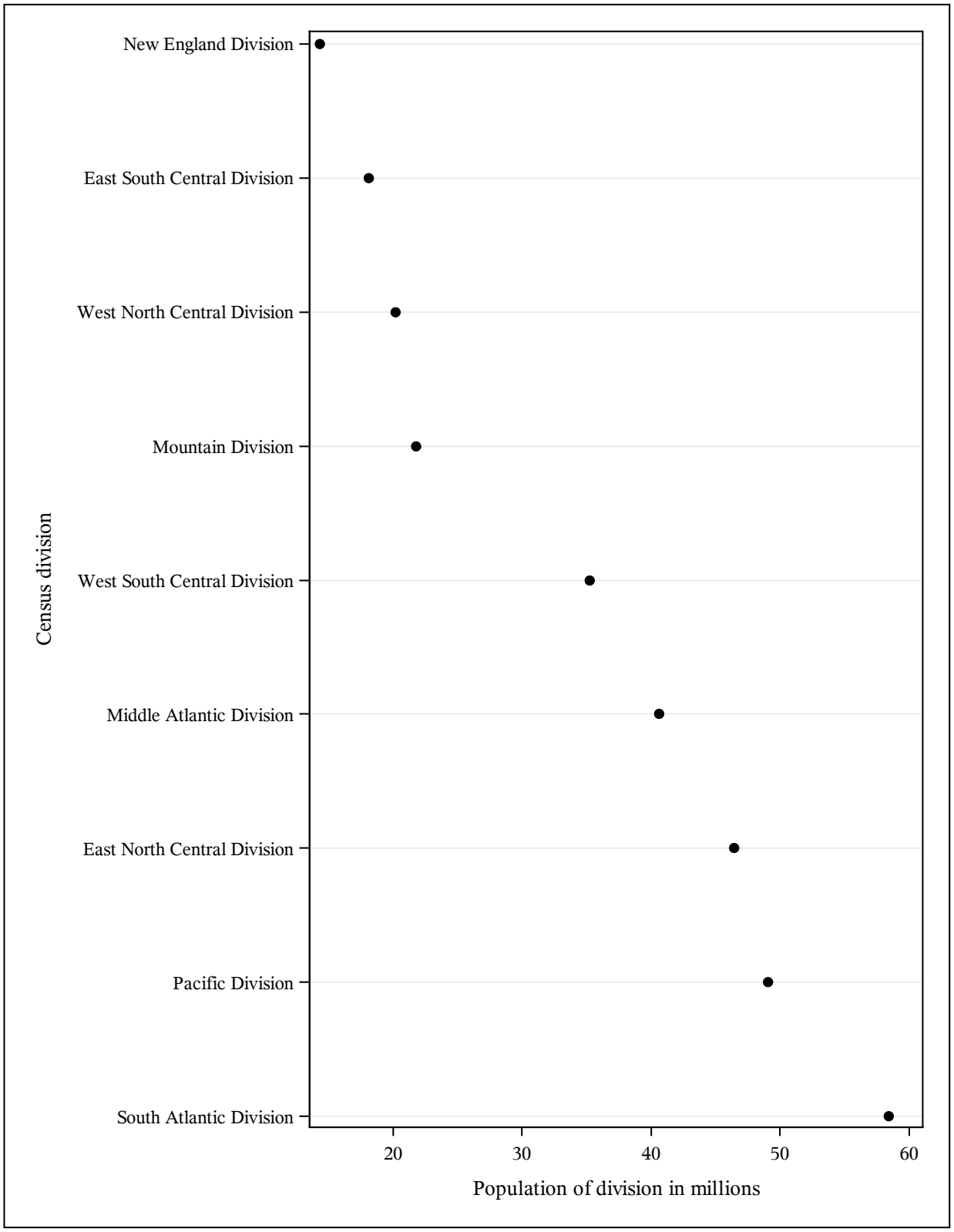

Figure 4: Dot chart chart with 9 categories, version 2

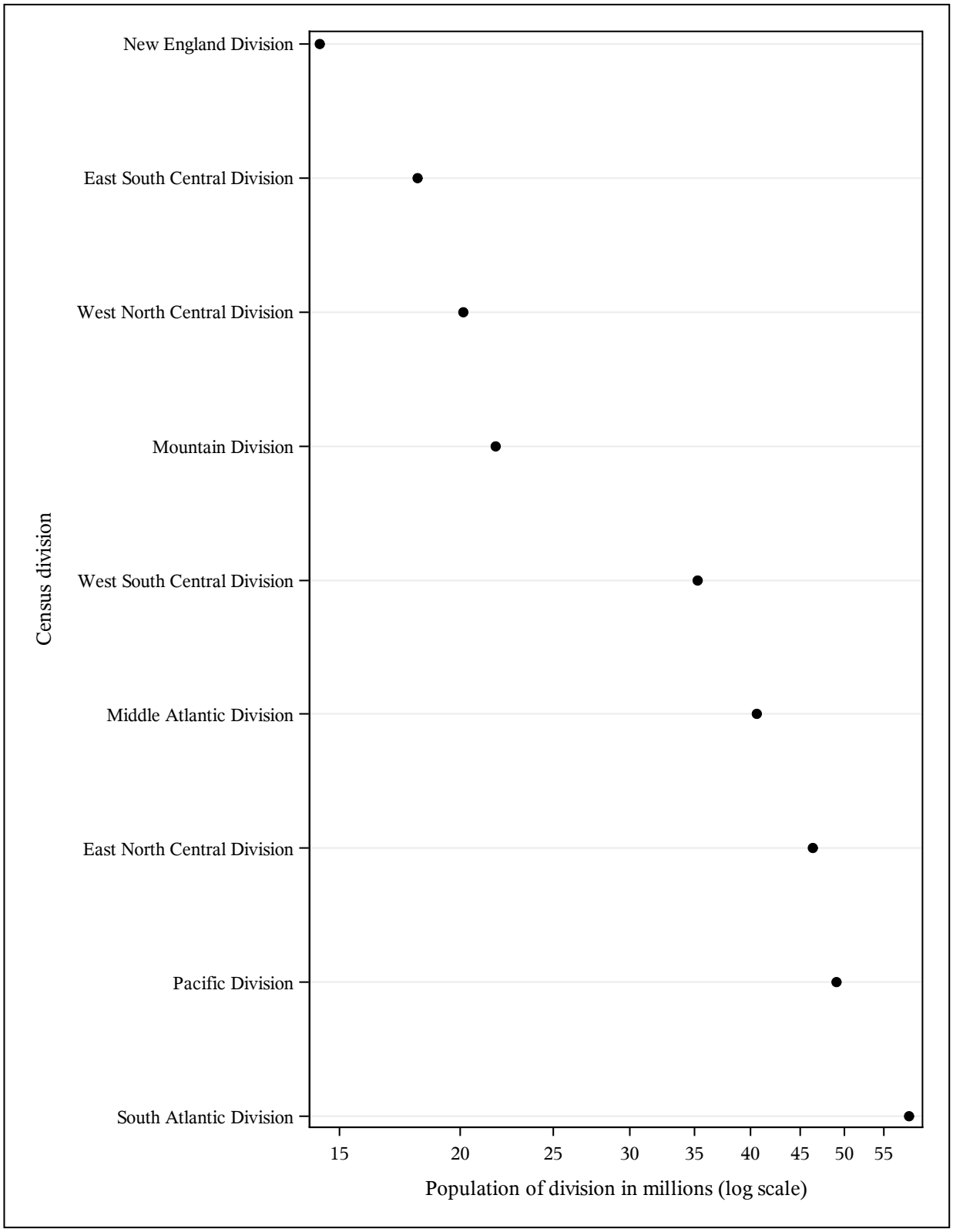

Figure 5: Dot chart chart with 9 categories, version 3

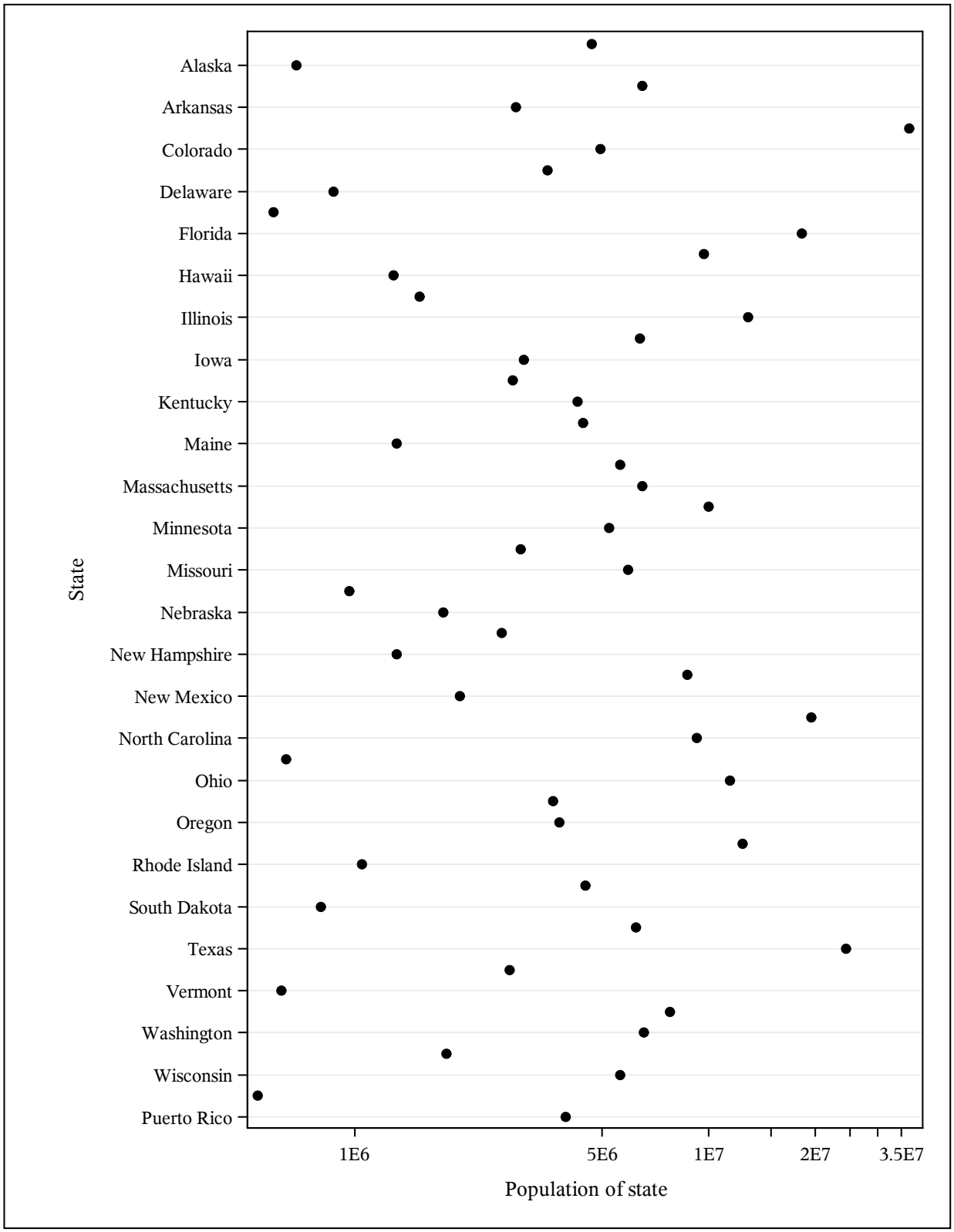

Figure 6: Dot chart chart with 51 categories, version 1

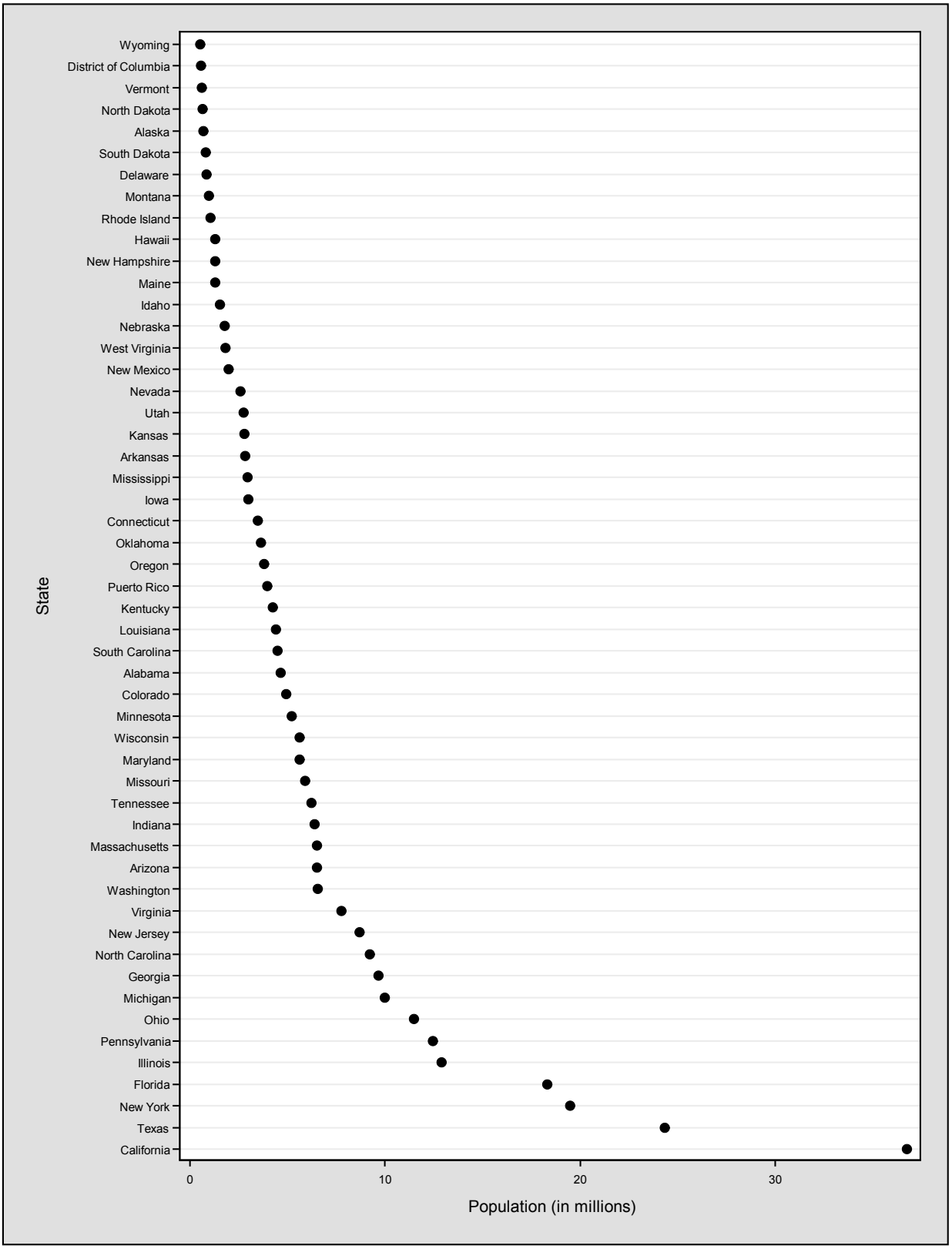

Figure 7: Dot chart chart with 51 categories, version 2

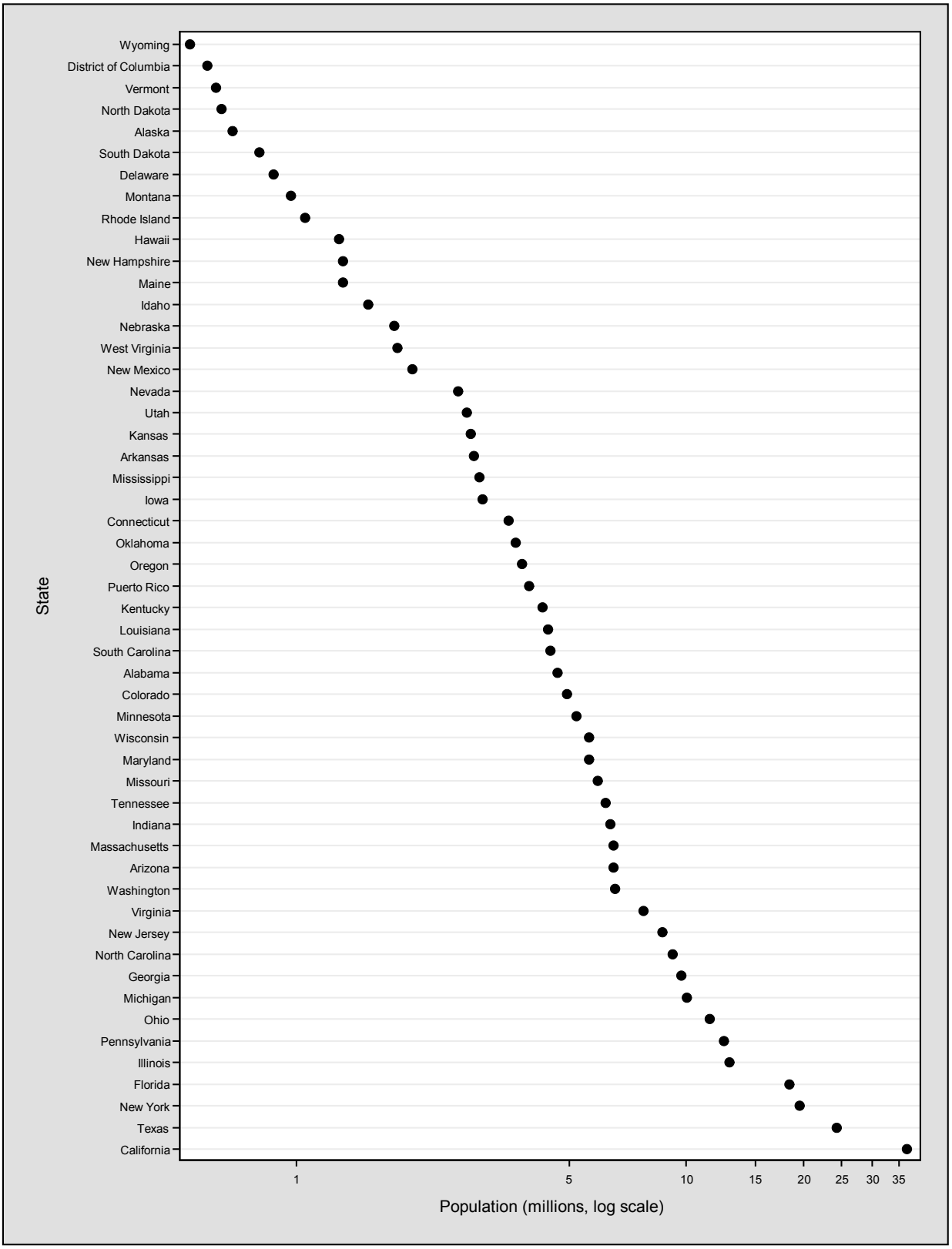

Figure 8: Dot chart chart with 51 categories, version 3