

---

## Analysis of some newsgroups traffic

by Nasser M. Abbasi, Updated Nov 15, 2010

Data was imported from Google newsgroup. I could not obtain data for R so I do not have it here.

Plots are made for the following groups

```
In[2]:= TableForm[groups[[All, 1 ;; 2]], TableHeadings -> {None, {"group", "URL"}}]
```

```
Out[2]//TableForm=
```

group	URL
fortran	<a href="http://groups.google.com/group/comp.lang.fortran/about">http://groups.google.com/group/comp.lang.fortran/about</a>
matlab	<a href="http://groups.google.com/group/comp.soft-sys.matlab/about">http://groups.google.com/group/comp.soft-sys.matlab/about</a>
mathematica	<a href="http://groups.google.com/group/comp.soft-sys.math.mathematica/about">http://groups.google.com/group/comp.soft-sys.math.mathematica/about</a>
maple	<a href="http://groups.google.com/group/comp.soft-sys.math.maple/about">http://groups.google.com/group/comp.soft-sys.math.maple/about</a>
scilab	<a href="http://groups.google.com/group/comp.soft-sys.math.scilab/about">http://groups.google.com/group/comp.soft-sys.math.scilab/about</a>
symbolic	<a href="http://groups.google.com/group/sci.math.symbolic/about">http://groups.google.com/group/sci.math.symbolic/about</a>
ada	<a href="http://groups.google.com/group/comp.lang.ada/about">http://groups.google.com/group/comp.lang.ada/about</a>
python	<a href="http://groups.google.com/group/comp.lang.python/about">http://groups.google.com/group/comp.lang.python/about</a>
lisp	<a href="http://groups.google.com/group/comp.lang.lisp/about">http://groups.google.com/group/comp.lang.lisp/about</a>
java	<a href="http://groups.google.com/group/comp.lang.java.programmer/about?">http://groups.google.com/group/comp.lang.java.programmer/about?</a>
ruby	<a href="http://groups.google.com/group/comp.lang.ruby/about">http://groups.google.com/group/comp.lang.ruby/about</a>
cobol	<a href="http://groups.google.com/group/comp.lang.cobol/about">http://groups.google.com/group/comp.lang.cobol/about</a>
scheme	<a href="http://groups.google.com/group/comp.lang.scheme/about">http://groups.google.com/group/comp.lang.scheme/about</a>
sagesupport	<a href="http://groups.google.com/group/sage-support/about">http://groups.google.com/group/sage-support/about</a>
sagedevel	<a href="http://groups.google.com/group/sage-devel/about">http://groups.google.com/group/sage-devel/about</a>
dsp	<a href="http://groups.google.com/group/comp.dsp/about?">http://groups.google.com/group/comp.dsp/about?</a>
num-analysis	<a href="http://groups.google.com/group/sci.math.num-analysis/about?">http://groups.google.com/group/sci.math.num-analysis/about?</a>
tex	<a href="http://groups.google.com/group/comp.text.tex/about?">http://groups.google.com/group/comp.text.tex/about?</a>
math	<a href="http://groups.google.com/group/sci.math/about?">http://groups.google.com/group/sci.math/about?</a>
physics	<a href="http://groups.google.com/group/sci.physics/about?">http://groups.google.com/group/sci.physics/about?</a>
maxima	<a href="http://www.math.utexas.edu/pipermail/maxima/">http://www.math.utexas.edu/pipermail/maxima/</a>

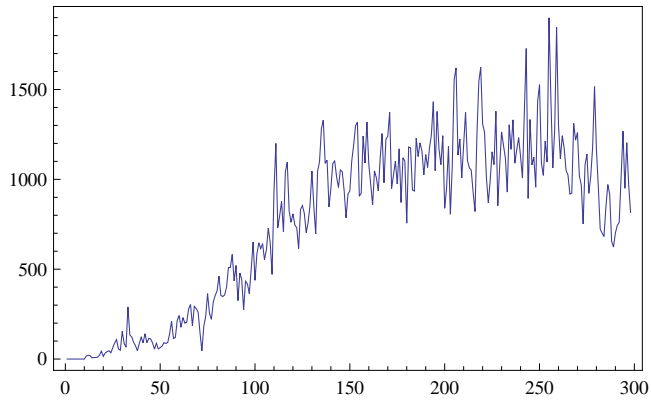
- Show trends of each group on its own, y-axis is average number of posts per months, x-axis is month number starting the count from from 1988. Notice some groups are older than others.

```
In[36]:= Labeled[Grid[Partition[Table[ListPlot[groups[[i, 5]][[1 ;; -3]],  
    Joined -> True, FrameLabel -> {{None, None}, {None, Style[groups[[i, 1], 16]}},  
    FrameTicks -> {{Automatic, None}, {Automatic, None}}, Frame -> True, ImageSize -> 300],  
    {i, 1, Length[groups]}], 3], Frame -> None, Spacings -> {2, 2}],  
    Style["Average posts per month trend", Bold, 18], Top]
```

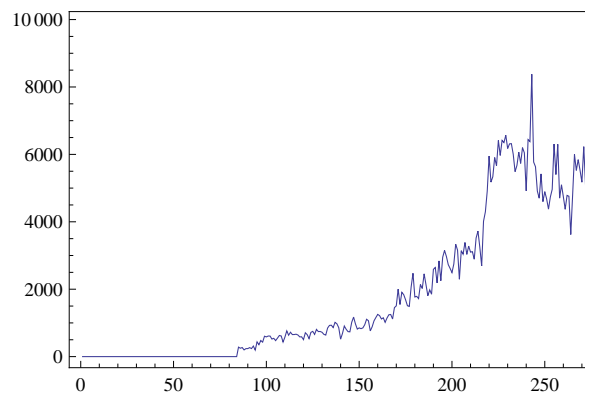
Average posts per month trend

Average posts per month of

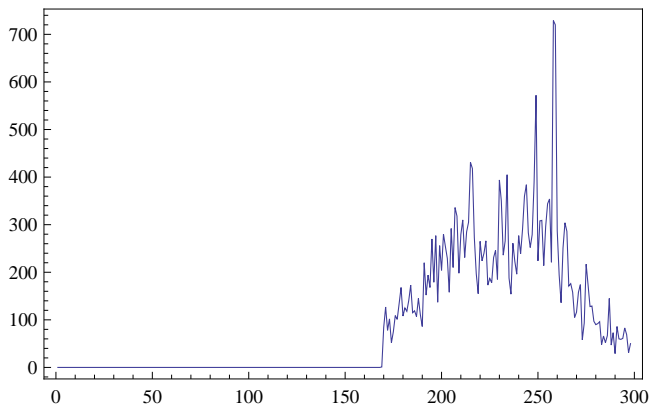
fortran



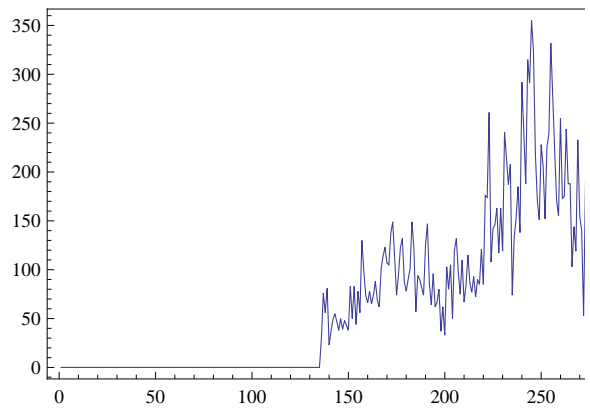
matlab



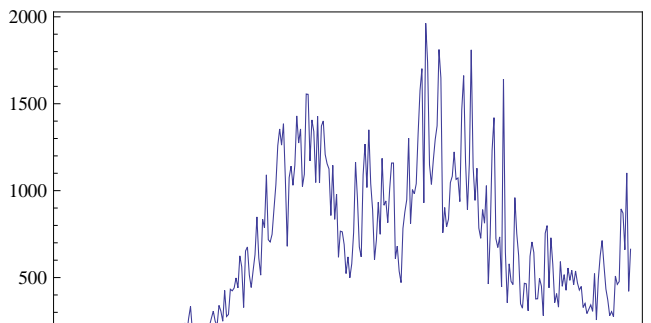
maple



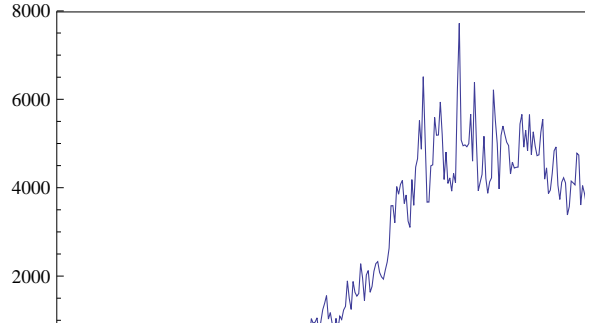
scilab

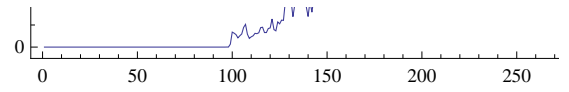
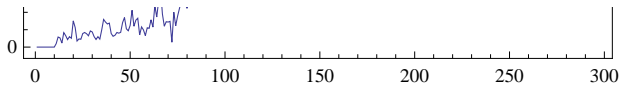


ada



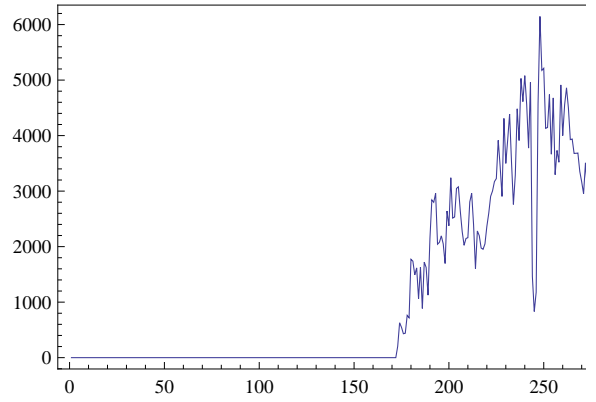
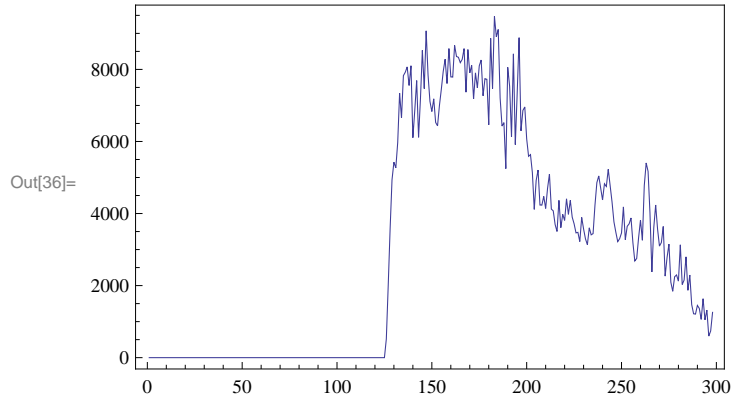
python





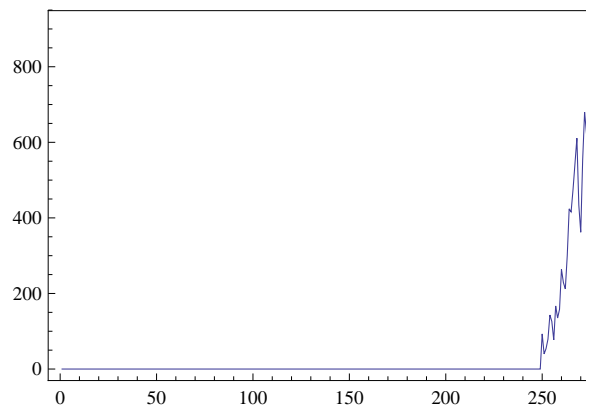
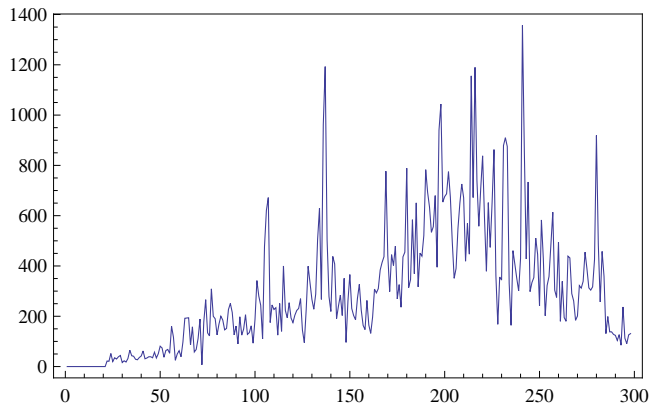
java

ruby



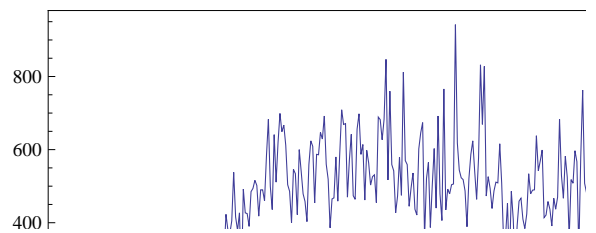
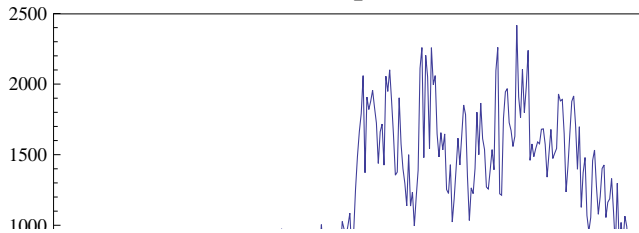
scheme

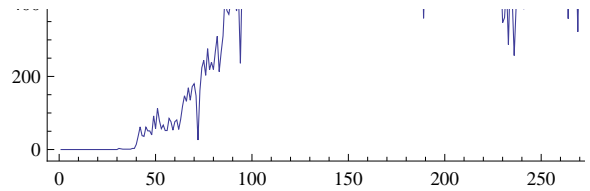
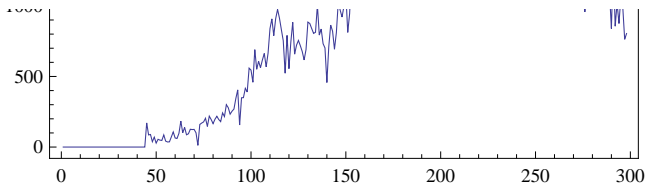
sagesupport



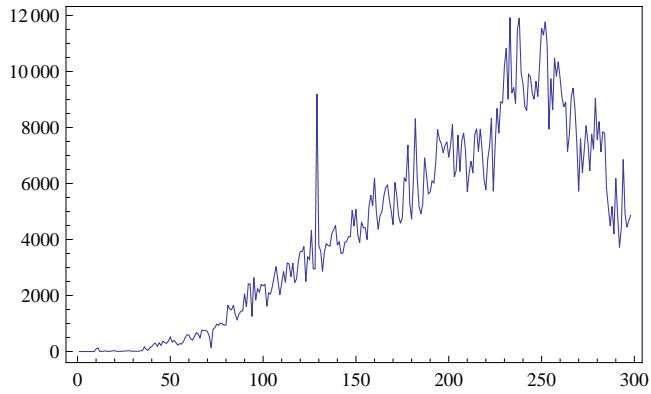
dsp

num-analysis

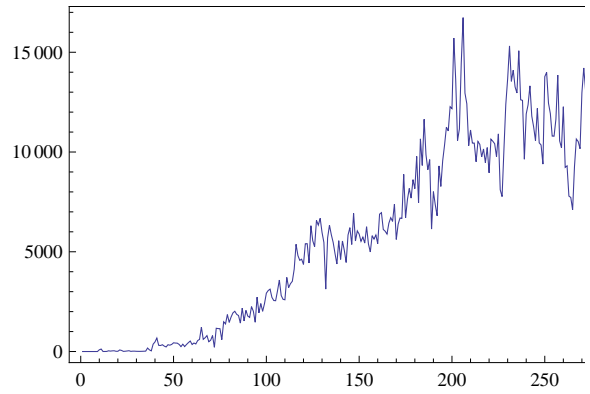




math



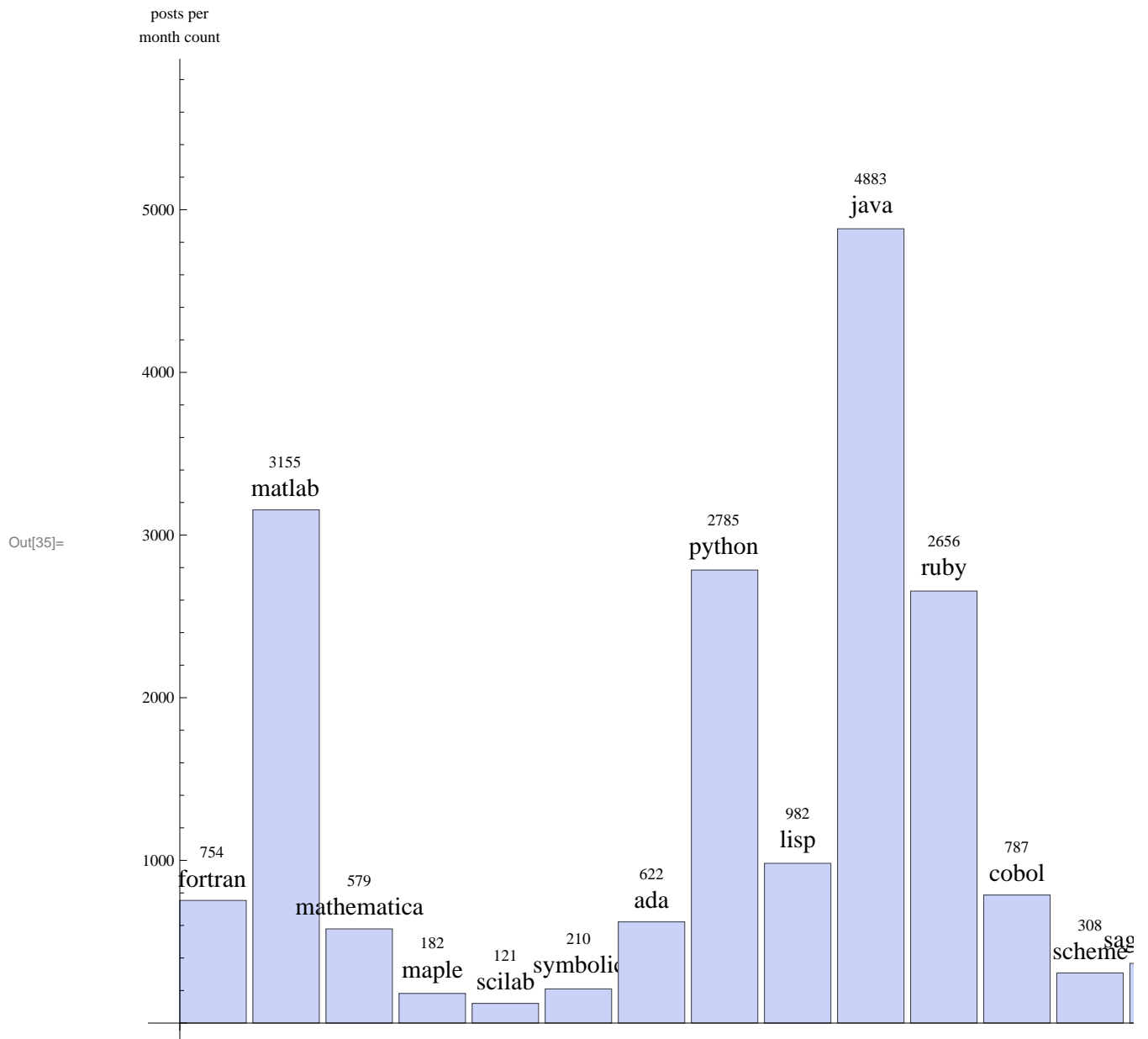
physics



- Make bar chart showing average posts per month for each group

```
In[35]:= BarChart[groups[[All, 4]], ChartLabels -> Placed[groups[[All, 1]], None],
  LabelingFunction -> (Placed[Column[{{#1, Style[#, 14]}}, Alignment -> Center], Above] &),
  PlotLabel -> Style["Average posts per month for number of news groups", 16],
  AxesLabel -> {"group", "posts per\nmonth count"}, ImagePadding -> {{40, 40}, {50, 40}}]
```

## Average posts per month for number of ne

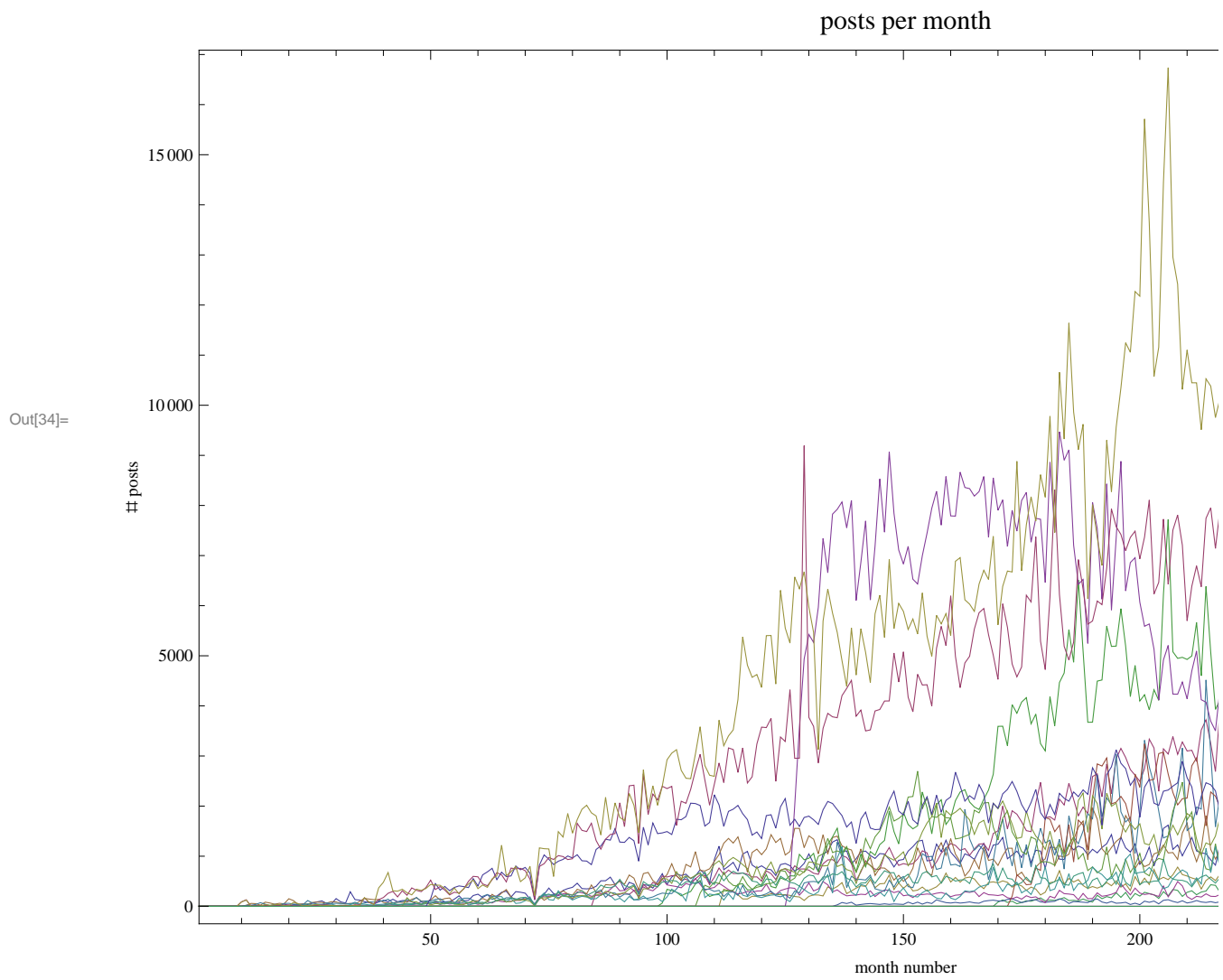


### ■ Show all traffic on one plot

```

In[31]:= Needs["PlotLegends`"];
maxLen =
  Max@Map[Length, Table[Flatten[groups[[i, 3]][[All, 2 ;; -1]], 2], {i, 1, Length[groups]}]];
Do[{
  groups[[i, 5]] = Table[0, {i, maxLen}];
  data = Flatten[groups[[i, 3]][[All, 2 ;; -1]], 2];
  len = Length[data];
  groups[[i, 5]][[maxLen - len + 1 ;; -1]] = data
}, {i, 1, Length[groups]}];
ListPlot[Table[groups[[i, 5]][[1 ;; -3]], {i, 1, Length[groups]}],
  Joined -> True, Frame -> True, PlotRange -> {{1, maxLen}, All},
  PlotLegend -> groups[[All, 1]], LegendPosition -> {.9, -0.2}, LegendSize -> {.5, 1},
  FrameLabel -> {"# posts", None}, {"month number", Style["posts per month", 14]}]

```



---

## How was the plots generated?

The data above was obtained for each system. Each entry in the list is the number of posts per month. Matlab started in 1993. Fortran in 1988, The Other systems did not start until later. A list plot was generated for each data, comparing systems against each others.

Sage has 2 groups. Support and Developer. It was suggested that only Support be used. Hence in the plots below, only Support is used. Average number of posts per month is also compared between all the systems. Data used is cut off at end of October 2010. Does not include November 2010.

Plots are all shown below.

---

## Function to download the data

Thanks goes to mark mcclure for providing these functions on the newsgroup. Made downloading data much easier.

```
In[3]:= getNewsGroupData [newGroupURL_String] := Module[{in, data},
  in = Import[newGroupURL, "Data"];
  data = ToExpression[Rest[Cases[in, {"Archive", _}, Infinity][[1, 2]]]];
  data[[1]] = Prepend[PadLeft[Rest[data[[1]]], 12], First[data[[1]]]];
  data[[-1]] = PadRight[Drop[data[[-1]], -1], 13];
  data
];

getMaximaNewsGroupData [] := Module[{yeardata, year, month, s}, Table[yeardata =
  Import["http://www.math.utexas.edu/pipermail/maxima/" <> ToString[year] <> ".txt.gz"];
  yeardata = StringCases[yeardata, StartOfLine ~~ "From: " ~~ Shortest[x_] ~~
  StartOfLine ~~ "Date: " ~~ Shortest[y_] ~~ EndOfLine];
  yeardata = Table[Last[StringSplit[s, "\n"]], {s, yeardata}];
  Prepend[Table[Total[StringCount[yeardata, month]], {month, {"Jan", "Feb", "Mar", "Apr",
  "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"}}, year], {year, 2006, 2010}]
];
```

---

## Get all the data into arrays and process

```
In[23]:= Do[{
  groups[[i, 3]] = getNewsGroupData [groups[[i, 2]]];
  nyears = Length[groups[[i, 3]]];
  groups[[i, 4]] = Round[Total[groups[[i, 3]][[All, 2 ;; -1], 2] / (nyears * 12)]
}, {i, 1, Length[groups] - 1}
];

groups[[-1, 3]] = getMaximaNewsGroupData [];
nyears = Length[groups[[-1, 3]]];
groups[[-1, 4]] = Round[Total[groups[[-1, 3]][[All, 2 ;; -1], 2] / (nyears * 12)];
```