

Analysis of some newsgroups traffic

by Nasser M. Abbasi, Updated Nov 4, 2012

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

```
(*extra fields: table from url, mean posts per month, running data*)
groups = {
  {"fortran", "http://groups.google.com/group/comp.lang.fortran/about", 0, 0, 0},
  {"matlab", "http://groups.google.com/group/comp.soft-sys.matlab/about", 0, 0, 0},
  {"mathematica",
   "http://groups.google.com/group/comp.soft-sys.math.mathematica/about", 0, 0, 0},
  {"maple", "http://groups.google.com/group/comp.soft-sys.math.maple/about", 0, 0, 0},
  {"scilab",
   "http://groups.google.com/group/comp.soft-sys.math.scilab/about", 0, 0, 0},
  {"symbolic", "http://groups.google.com/group/sci.math.symbolic/about", 0, 0, 0},
  {"ada", "http://groups.google.com/group/comp.lang.ada/about", 0, 0, 0},
  {"python", "http://groups.google.com/group/comp.lang.python/about", 0, 0, 0},
  {"lisp", "http://groups.google.com/group/comp.lang.lisp/about", 0, 0, 0},
  {"java",
   "http://groups.google.com/group/comp.lang.java.programmer/about?", 0, 0, 0},
  {"ruby", "http://groups.google.com/group/comp.lang.ruby/about", 0, 0, 0},
  {"cobol", "http://groups.google.com/group/comp.lang.cobol/about", 0, 0, 0},
  {"scheme", "http://groups.google.com/group/comp.lang.scheme/about", 0, 0, 0},
  {"sagedevel", "http://groups.google.com/group/sage-devel/about", 0, 0, 0},
  {"dsp", "http://groups.google.com/group/comp.dsp/about?", 0, 0, 0},
  {"num-analysis",
   "http://groups.google.com/group/sci.math.num-analysis/about?", 0, 0, 0},
  {"tex", "http://groups.google.com/group/comp.text.tex/about?", 0, 0, 0},
  {"math", "http://groups.google.com/group/sci.math/about?", 0, 0, 0},
  {"physics", "http://groups.google.com/group/sci.physics/about?", 0, 0, 0},
  {"javascript",
   "http://groups.google.com/group/comp.lang.javascript/about?", 0, 0, 0},
  {"c++", "http://groups.google.com/group/comp.lang.c++/about?", 0, 0, 0},
  {"c++ moderated",
   "http://groups.google.com/group/comp.lang.c++.moderated/about?", 0, 0, 0},
  {"comp.lang.functional",
   "http://groups.google.com/group/comp.lang.functional/about?", 0, 0, 0},
  {"maxima", "http://www.math.utexas.edu/pipermail/maxima/", 0, 0, 0}
};
```

```
TableForm[groups[[All, 1 ;; 2]], TableHeadings → {None, {"group", "URL"}}]
```

group	URL
fortran	http://groups.google.com/group/comp.lang.fortran/about
matlab	http://groups.google.com/group/comp.soft-sys.matlab/about
mathematica	http://groups.google.com/group/comp.soft-sys.math.mathematica/about
maple	http://groups.google.com/group/comp.soft-sys.math.maple/about
scilab	http://groups.google.com/group/comp.soft-sys.math.scilab/about
symbolic	http://groups.google.com/group/sci.math.symbolic/about
ada	http://groups.google.com/group/comp.lang.ada/about
python	http://groups.google.com/group/comp.lang.python/about
lisp	http://groups.google.com/group/comp.lang.lisp/about
java	http://groups.google.com/group/comp.lang.java.programmer/about?
ruby	http://groups.google.com/group/comp.lang.ruby/about
cobol	http://groups.google.com/group/comp.lang.cobol/about
scheme	http://groups.google.com/group/comp.lang.scheme/about
sagedevel	http://groups.google.com/group/sage-devel/about
dsp	http://groups.google.com/group/comp.dsp/about?
num-analysis	http://groups.google.com/group/sci.math.num-analysis/about?
tex	http://groups.google.com/group/comp.text.tex/about?
math	http://groups.google.com/group/sci.math/about?
physics	http://groups.google.com/group/sci.physics/about?
javascript	http://groups.google.com/group/comp.lang.javascript/about?
c++	http://groups.google.com/group/comp.lang.c++/about?
c++ moderated	http://groups.google.com/group/comp.lang.c++.moderated/about?
comp.lang.functional	http://groups.google.com/group/comp.lang.functional/about?
maxima	http://www.math.utexas.edu/pipermail/maxima/

- 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.

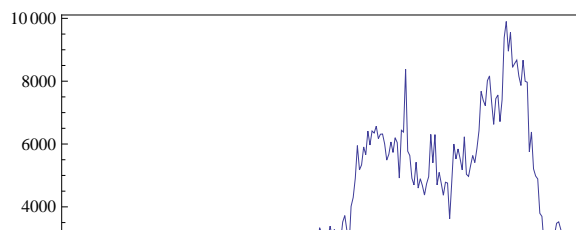
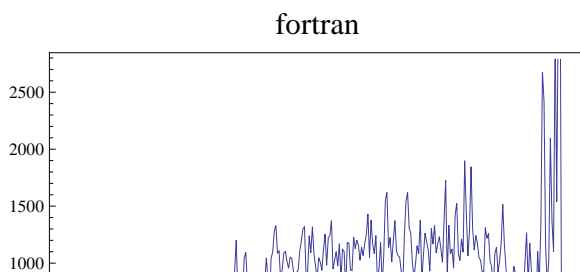
```
(*evaluate all functions at the end of this notebook before running this*)
```

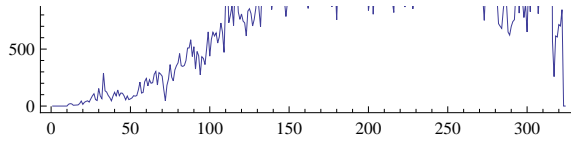
```
tbl = Table[ListPlot[Flatten[Drop[#, 1] & /@ groups[[i, 3]],
  Joined → True,
  FrameLabel → {{None, None}, {None, Style[groups[[i, 1], 16]}]},
  FrameTicks → {{Automatic, None}, {Automatic, None}},
  Frame → True, ImageSize → 300],
  {i, 1, Length[groups]}];
```

```
g = Grid[Partition[tbl, 3],
  Frame → None,
  Spacings → {2, 2}];
```

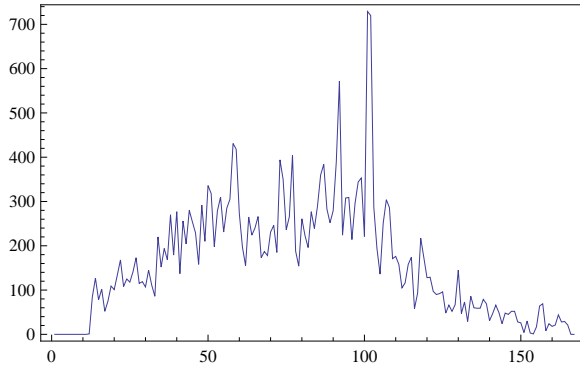
```
Labeled[g, Style["Average posts per month trend", Bold, 18], Top]
```

Average posts per month trend

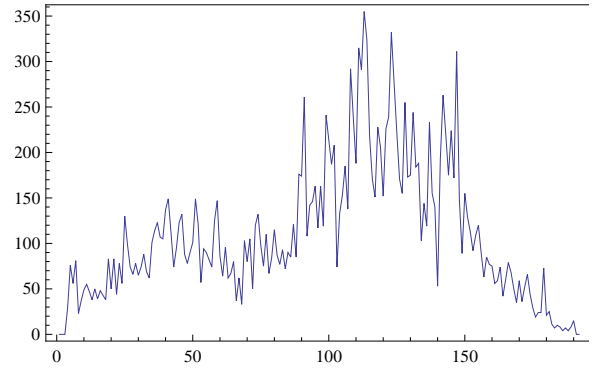




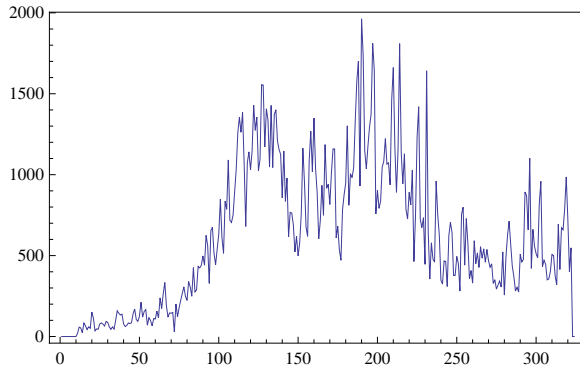
maple



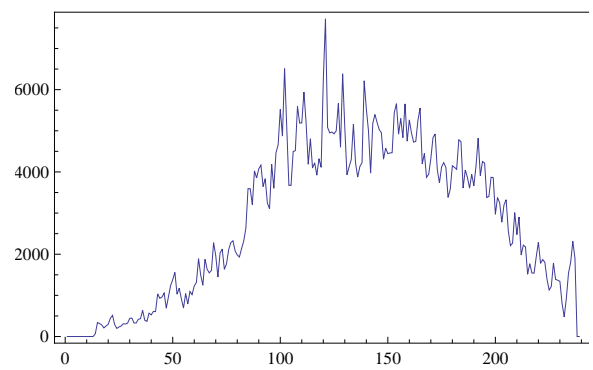
scilab



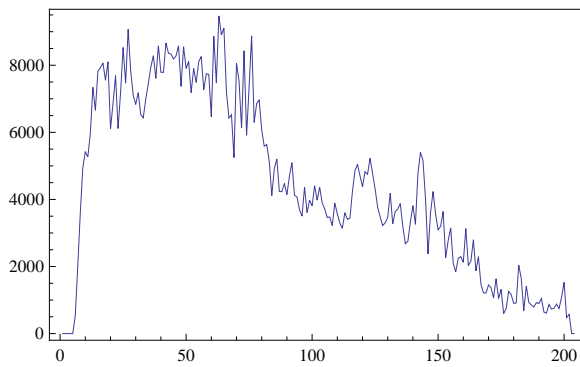
ada



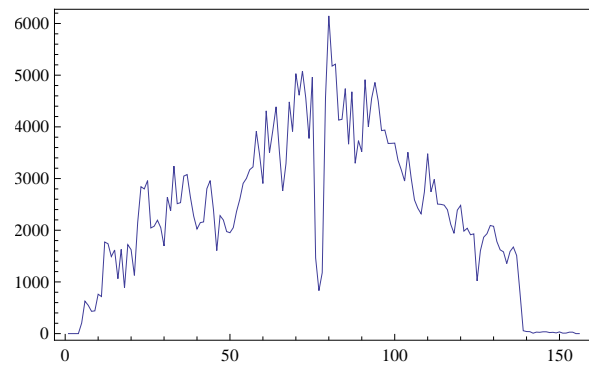
python



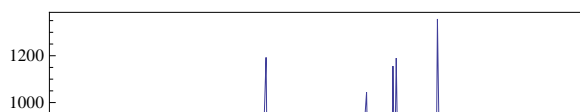
java



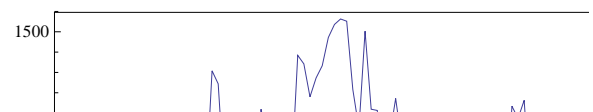
ruby

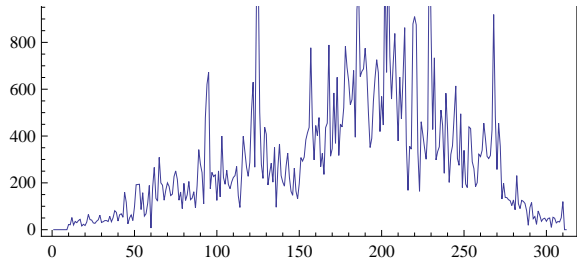


scheme

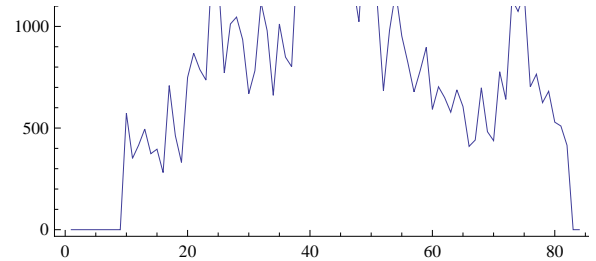


sagedevel

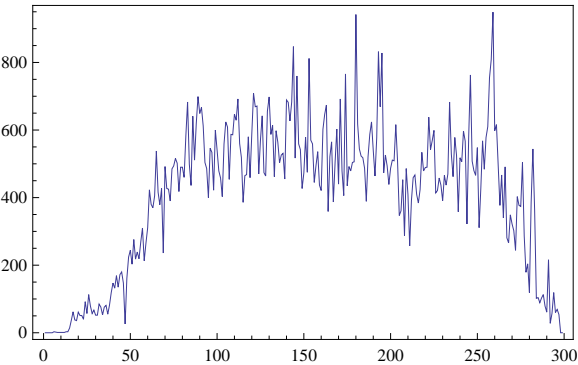




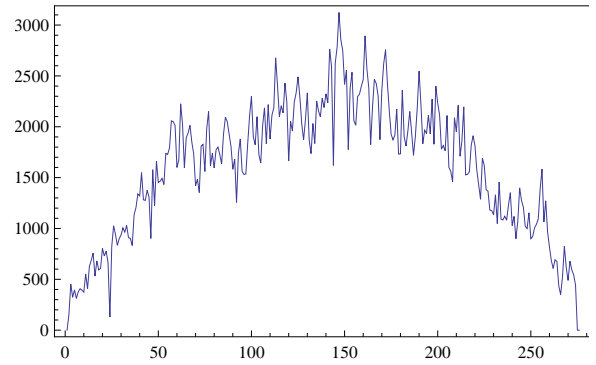
num-analysis



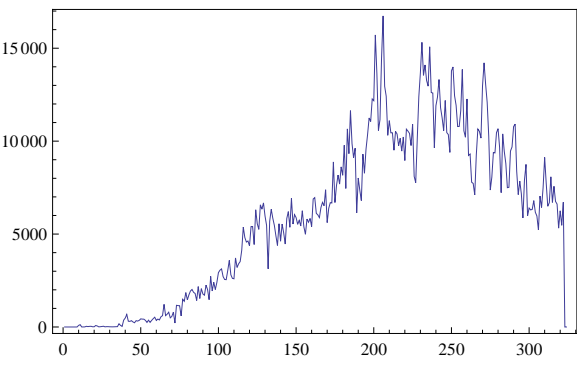
tex



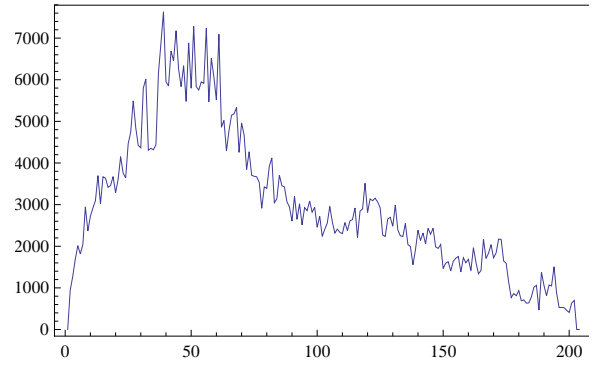
physics



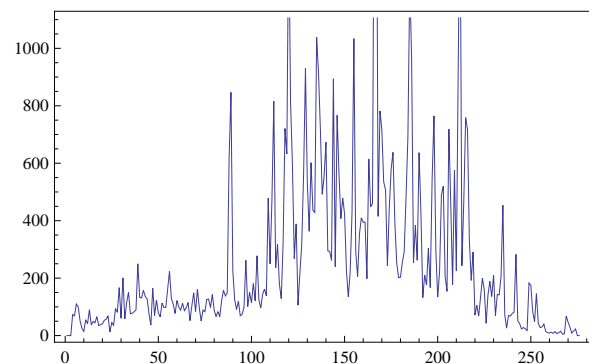
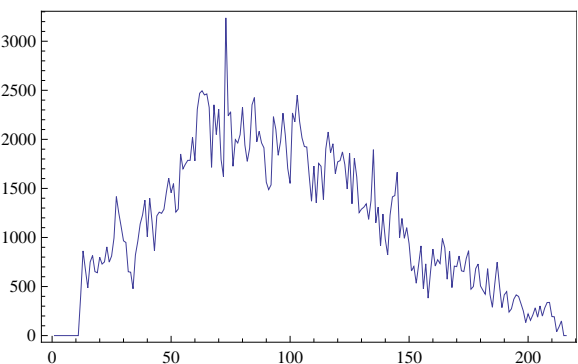
javascript



c++ moderated

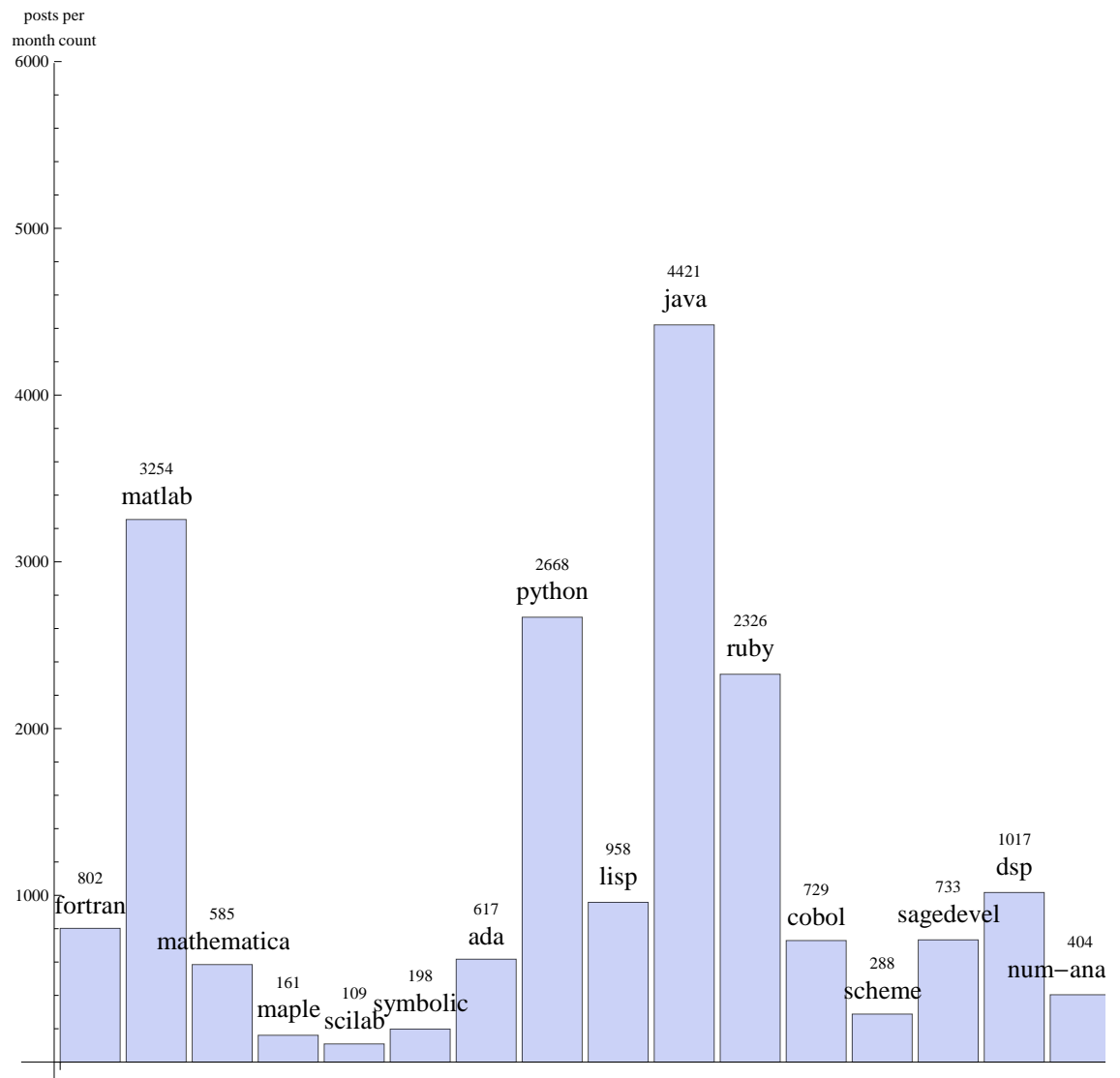


comp.lang.functional

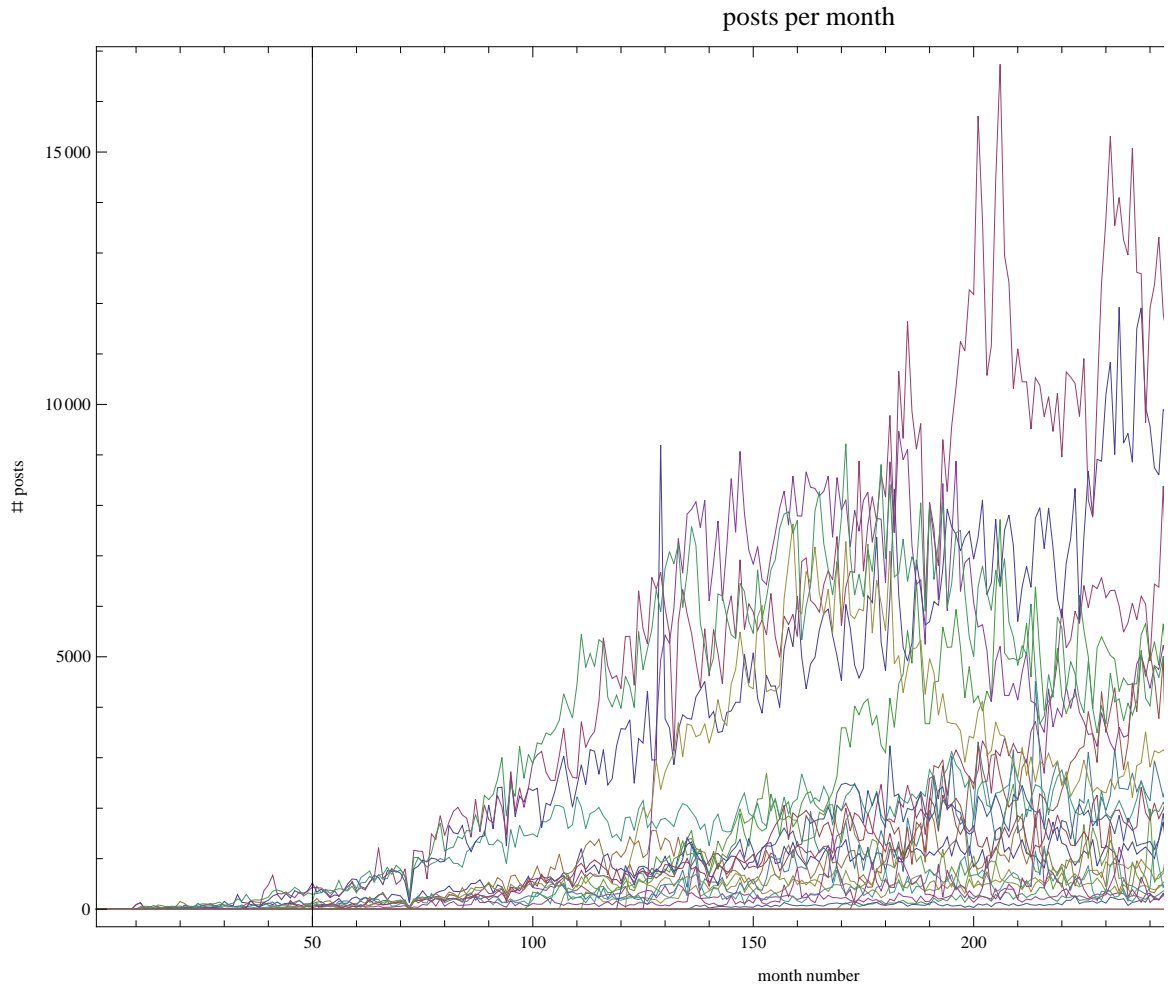


■ Make bar chart showing average posts per month for each group

Average posts per month for number of news gr



- Show all traffic on one plot



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 WReach for providing the *Mathematica* code to help with a problem downloading the data from Google usenet. The post is here.

Thanks also for Mark McClure for providing original function and the parsing for the data.. These made downloading data much easier.

```
(*extra fields: table from url, mean posts per month, running data*)
groups = {
  {"fortran", "http://groups.google.com/group/comp.lang.fortran/about", 0, 0, 0},
  {"matlab", "http://groups.google.com/group/comp.soft-sys.matlab/about", 0, 0, 0},
  {"mathematica",
   "http://groups.google.com/group/comp.soft-sys.math.mathematica/about", 0, 0, 0},
  {"maple", "http://groups.google.com/group/comp.soft-sys.math.maple/about", 0, 0, 0},
  {"scilab",
   "http://groups.google.com/group/comp.soft-sys.math.scilab/about", 0, 0, 0},
  {"symbolic", "http://groups.google.com/group/sci.math.symbolic/about", 0, 0, 0},
  {"ada", "http://groups.google.com/group/comp.lang.ada/about", 0, 0, 0},
  {"python", "http://groups.google.com/group/comp.lang.python/about", 0, 0, 0},
  {"lisp", "http://groups.google.com/group/comp.lang.lisp/about", 0, 0, 0},
  {"java",
   "http://groups.google.com/group/comp.lang.java.programmer/about?", 0, 0, 0},
  {"ruby", "http://groups.google.com/group/comp.lang.ruby/about", 0, 0, 0},
  {"cobol", "http://groups.google.com/group/comp.lang.cobol/about", 0, 0, 0},
  {"scheme", "http://groups.google.com/group/comp.lang.scheme/about", 0, 0, 0},
  {"sagedevel", "http://groups.google.com/group/sage-devel/about", 0, 0, 0},
  {"dsp", "http://groups.google.com/group/comp.dsp/about?", 0, 0, 0},
  {"num-analysis",
   "http://groups.google.com/group/sci.math.num-analysis/about?", 0, 0, 0},
  {"tex", "http://groups.google.com/group/comp.text.tex/about?", 0, 0, 0},
  {"math", "http://groups.google.com/group/sci.math/about?", 0, 0, 0},
  {"physics", "http://groups.google.com/group/sci.physics/about?", 0, 0, 0},
  {"javascript",
   "http://groups.google.com/group/comp.lang.javascript/about?", 0, 0, 0},
  {"c++", "http://groups.google.com/group/comp.lang.c++/about?", 0, 0, 0},
  {"c++ moderated",
   "http://groups.google.com/group/comp.lang.c++.moderated/about?", 0, 0, 0},
  {"comp.lang.functional",
   "http://groups.google.com/group/comp.lang.functional/about?", 0, 0, 0},
  {"maxima", "http://www.math.utexas.edu/pipermail/maxima/", 0, 0, 0}
};
```

```
Needs["JLink`"]

httpGet[url_String] := JavaBlock@
Module[{http, get}, http = JavaNew["org.apache.commons.httpclient.HttpClient"];
get = JavaNew["org.apache.commons.httpclient.methods.GetMethod", url];
http@executeMethod[get]; get@getResponseBodyAsString[]

getNewsGroupData[newGroupURL_String] := Module[{in, data},
in = ImportString[httpGet[newGroupURL], {"HTML", "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

```
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)];
```