Tool for Quantitative Analysis of IETF mailinglists
What

- Data collection (~34 GB!)
- Basic statistics
- Word analysis
- Network analysis
How

- Jupyter notebooks
- Python
  - pandas
  - nltk
  - networkx
  - etc
Issues

- ??, 12 1 2006 18:24:14 +0100
- iso-f675-5 unknown encoding in message
- ?????, 01 ????? 2002 20:15:33 +0800
- x-unknown unknown encoding in message
- Fri, 32 Jan 2008 19:05:18 +0900
Where

https://github.com/dataactive/bigbang/
Who
Statistics

1,944,019 e-mails
Statistics

1,944,019 e-mails
From 192,046 mailaddresses
Statistics

1,944,019 e-mails
From 192,046 mailaddresses
On 955 analyzed mailinglists
Statistics

1,944,019 e-mails
from 19204 mailaddresses
On 955 analyzed mailinglists

1 mailaddress sent 80,533 messages
Statistics

1,944,019 e-mails
from 19204 mailaddresses
On 955 analyzed mailinglists

1 mailaddress sent 80533 messages
Most contributors sent 1 or 2 messages (spam?)
Length of time people participate

Mailinglists are NOT dominated by early participants
Number of messages sent by contributor
Threads on ietf@ietf.org

Longest threads:

• Concerns about Singapore (201 emails)
• Request for community guidance on issue concerning a future meeting of the IETF (156 emails)
• Hotel situation (141 emails)
• Proposed Photography Policy (138 emails)
• DMARC: perspectives from a listadmin of large open-source lists (135 emails)
• Background on Singapore go/no go for IETF 100 (124 emails)
• Diversity of IETF Leadership (103 emails)
• RFP for Remote Participation Services Specifications Development (94 emails)
• IETF Diversity Question on Berlin Registration? (89 emails)
Degree centrality on ietf@ietf.org
Word trends on ietf@ietf.org
Join us!

Mailinglists:
https://lists.ghserv.net/mailman/listinfo/bigbang-dev
https://lists.ghserv.net/mailman/listinfo/bigbang-user

GitHub
https://github.com/dataactive/bigbang/