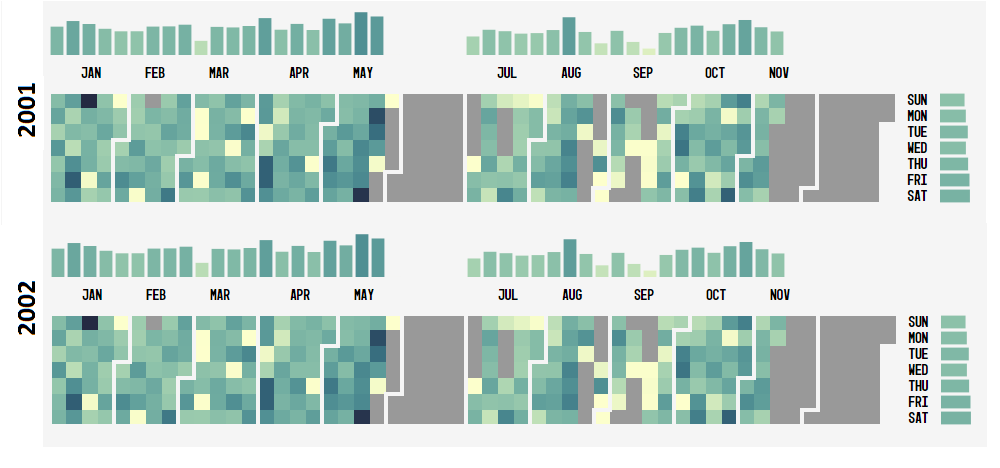
SQL Queries

**LEVEL 1:**

**Calendar Heatmap**



Requirements:

1. Data from January 2001 to December 2001 (total emails by each day)

|  |  |
| --- | --- |
| Date | Total Emails = Sent+Received+FW+CC |
| 2001-1-30 | 808 |
| 2001-1-29 | 894 |
| 2001-1-28 | 866 |
| 2001-1-27 | 872 |
| - |  |
| - |  |

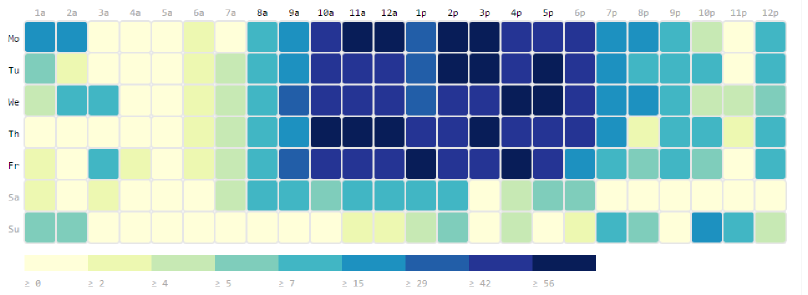
2. Need both CSV and json format

*1.json / 1.csv*

3. SQL query

**select** *DATE*(**date**) **AS** emailDate, *count*(*\**) **AS** totalEmails  
**FROM** message  
**WHERE** *YEAR*(**date**) = 2001  
**GROUP BY** *DATE*(**date**)

**Day/Hour Heatmap**



Requirements:

1. Data from January 2001 to December 2001 (total emails by each day and 24 hours)

|  |  |  |
| --- | --- | --- |
| Day | Hour | Total Emails = Sent+Received+FW+CC |
| 2001-01-30 | 1 | 21 |
| 2001-01-30 | 2 | 44 |
| 2001-01-30 | 3 | 31 |
| 2001-01-30 | - | 87 |
| 2001-01-30 | - |  |
| 2001-01-30 | 24 | 86 |
| 2001-01-29 | 1 | 18 |
| - | - | - |

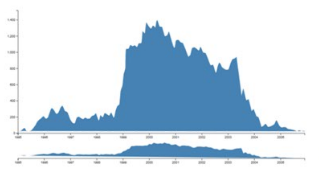
2. Need both CSV and json format

2.csv / 2.json

3. SQL query

**select** *DATE*(**date**) **AS** emailDate, *HOUR*(**date**) **AS** emailHour, *count*(*\**) **AS** totalEmails  
**FROM** message  
**WHERE** *YEAR*(**date**) = 2001  
**GROUP BY** *DATE*(**date**), *HOUR*(**date**)

**Line-area graph**



Requirements:

1. Data from January 2001 to December 2001 (aggregate of emails for each month)

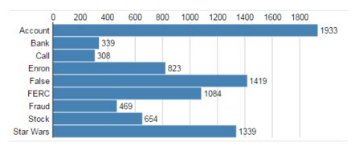
|  |  |
| --- | --- |
| Date | Total Emails = Sent+Received+FW+CC |
| Jan 2001 | 1808 |
| Feb 2001 | 1894 |
| Mar 2001 | 1866 |
| Apr 2001 | 1872 |
| - |  |
| - |  |

2. Need both CSV and json format

3. SQL query

**select** *YEAR*(**date**) **AS year**, *MONTH*(**date**) **as month** , *count*(*\**) **as** totalEmails  
**FROM** message  
**WHERE** *YEAR*(**date**) = 2001  
**GROUP BY** *YEAR*(**date**), *MONTH*(**date**)

**Horizontal Bar Charts**



Requirements:

1. Data from January 2001 to December 2001 (total emails sent/received by each individual using keywords) -**Individuals** could be Kenneth Lay, Jeffrey Skilling, Andrew Fastow, Richard Causay, Micheal Kapper, Lea Fastow, Ben Glisan, Dave Delainey, Mart Koenig, and Lou Lung Pai. **Keywords** could be FERC, Affair, Devastating, Investigation, Disclosure, Bonus, Meeting, Plan, Services, and Report.

|  |  |
| --- | --- |
| Individuals | Total Emails = Sent+Received+FW+CC |
| Andy | 808 |
| Lay | 894 |
| Phillips | 866 |
| Jack | 872 |
| - |  |
| - |  |

2. Need both CSV and json format

**4.csv / 4.json**

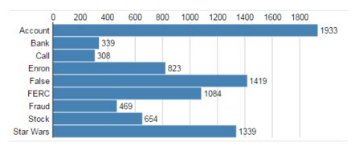
3. SQL query

**The list here has all the individuals who has sent emails with one the terms you’ve identified ordered according to the frequency. The names you’ve stated have not been in much activity so you can choose from here.**

I generated the query with a script so can be reused with other keywords, users, etc.

**SELECT** m.**sender**, e.**firstName**, e.**lastName**, *COUNT*(*\**) **AS** emailCounts  
**FROM** message m, employeelist e  
**WHERE** m.**sender** = e.**Email\_id  
AND** *YEAR*(m.**date**) = 2001  
**AND** m.**mid IN** (  
 **SELECT mid FROM** message  
 **WHERE** *LOWER*(message.**body**) **LIKE** "%ferc%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%affair%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%devastating%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%investigation%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%disclosure%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%bonus%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%meeting%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%plan%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%services%"  
 **OR** *LOWER*(message.**body**) **LIKE** "%report%"  
)  
**GROUP BY** m.**sender ORDER BY** emailCounts **DESC**

Horizontal Bar Charts



Requirements:

1. Data from January 2001 to December 2001 (total emails sent/received using each word by key individuals) - **Keywords** could be FERC, Affair, Devastating, Investigation, Disclosure, Bonus, Meeting, Plan, Services, and Report. **Individuals** could be Kenneth Lay, Jeffrey Skilling, Andrew Fastow, Richard Causay, Micheal Kapper, Lea Fastow, Ben Glisan, Dave Delainey, Mart Koenig, and Lou Lung Pai.

|  |  |
| --- | --- |
| Keywords | Total Emails = Sent+Received+FW+CC |
| Fraud | 808 |
| Enron | 894 |
| FERC | 866 |
| Regulations | 872 |
| - |  |
| - |  |

2. Need both CSV and json format

Here is the output from a script I’ve written:

FERC(8077,)

Affair(4680,)

Devastating(232,)

Investigation(2776,)

Disclosure(8671,)

Bonus(1744,)

Meeting(29319,)

Plan(35361,)

Services(22062,)

3. SQL quer

I generated the query with a script so can be reused with other keywords, users, etc.

**select** *count*(*\**)  
**from** message  
**where** *LOWER*(**body**) **LIKE** "%ferc%"

**select** *YEAR*(**date**) **AS year**, *count*(*\**) **as** totalEmails  
**FROM** message  
**WHERE** *YEAR*(**date**) = 2001

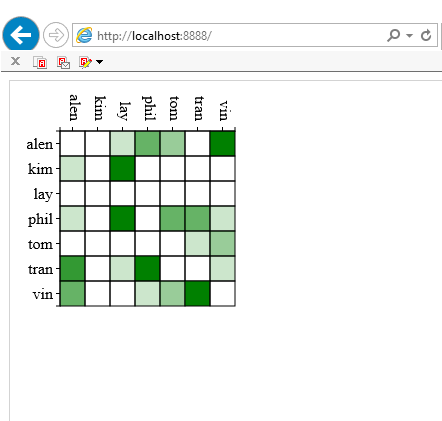
**where** *LOWER*(**body**) **LIKE** "%ferc%"

AND YEAR(m.date) = 2001

  AND MONTH(m.date) = 1

**LEVEL 2:**

**Adjacency Matrix Heatmap - contacts versus contacts**



1. Data from January 2001 to December 2001 (all the days) - **Individuals** could be Kenneth Lay, Jeffrey Skilling, Andrew Fastow, Richard Causay, Micheal Kapper, Lea Fastow, Ben Glisan, Dave Delainey, Mart Koenig, and Lou Lung Pai.

2. Need both CSV and json format

**This is also generated by a script.**

**5.csv**

3. SQL query

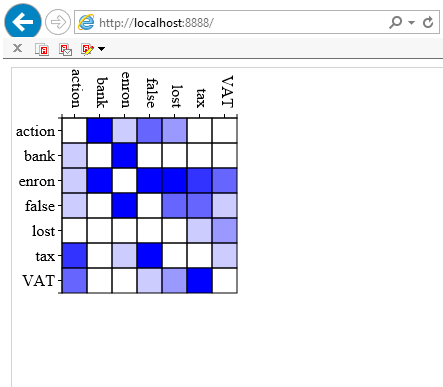
nodelist.csv

|  |  |  |
| --- | --- | --- |
| Individuals | Sent | Received |
| Andy | 80 | 808 |
| Lay | 89 | 894 |
| Phillips | 86 | 866 |
| Jack | 87 | 872 |
| - |  |  |
| - |  |  |

edgelist.csv

|  |  |  |
| --- | --- | --- |
| **source** | **target** | **weight** |
| lay | philip | 24 |
| philip | andy | 32 |
| lay | andy | 12 |
| philip | lay | 18 |
| andy | robert | 15 |

**Adjacency Matrix Heatmap - keywords versus keywords**



1. Data from January 2001 to December 2001 (all the days)- **Keywords** could be FERC, Affair, Devastating, Investigation, Disclosure, Bonus, Meeting, Plan, Services, and Report.

2. Need both CSV and json format

3. SQL query

nodelist.csv

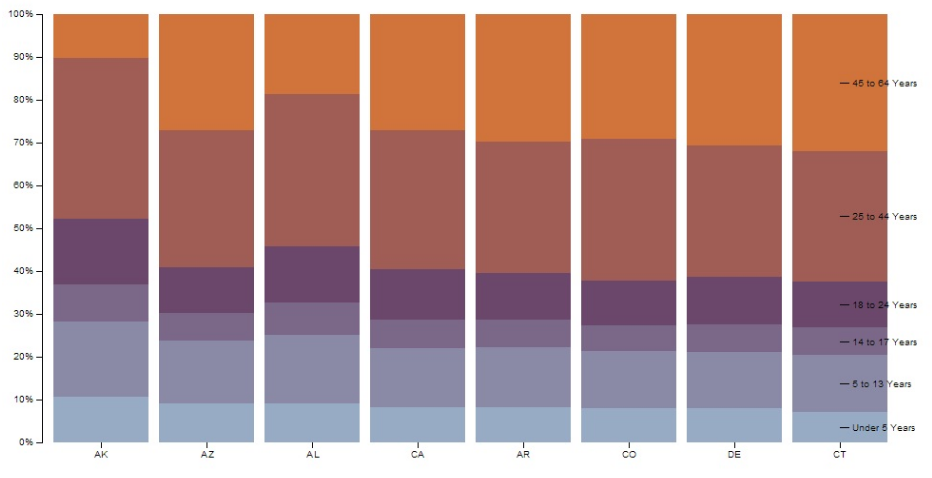
|  |  |  |
| --- | --- | --- |
| Keywords | Sent | Received |
| Fraud | 80 | 808 |
| Enron | 89 | 894 |
| FERC | 86 | 866 |
| Regulations | 87 | 872 |
| - |  |  |
| - |  |  |

edgelist.csv

|  |  |  |
| --- | --- | --- |
| **source** | **target** | **weight** |
| action | FALSE | 3 |
| action | bank | 8 |
| action | lost | 2 |
| action | enron | 1 |
| bank | enron | 5 |

**LEVEL 3:**

**Stacked barcharts**



Requirements:

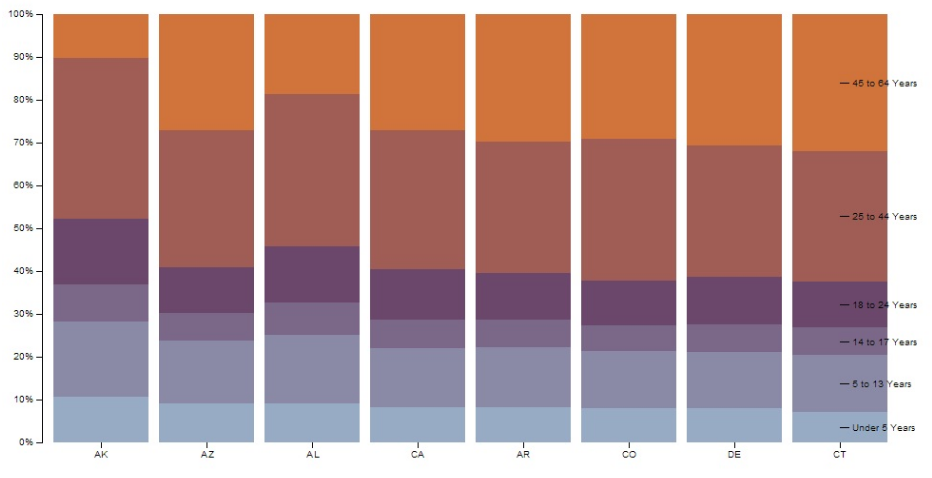
1. Data from January 2001 to December 2001 (all the days) **- Keywords** could be FERC, Affair, Devastating, Investigation, Disclosure, Bonus, Meeting, Plan, Services, and Report. **Individuals** could be Kenneth Lay, Jeffrey Skilling, Andrew Fastow, Richard Causay, Micheal Kapper, Lea Fastow, Ben Glisan, Dave Delainey, Mart Koenig, and Lou Lung Pai.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Individuals | Keyword A | Keyword B | Keyword C | Keyword D |
| Andy | 808 | 808 | 808 | 808 |
| Lay | 894 | 894 | 894 | 894 |
| Phillips | 866 | 866 | 866 | 866 |
| Jack | 872 | 872 | 872 | 872 |
| - |  |  |  |  |
| - |  |  |  |  |

2. Need both CSV and json format

3. SQL query

**Stacked barcharts**



Requirements:

1. Data from January 2001 to December 2001 (all the days) - **Keywords** could be FERC, Affair, Devastating, Investigation, Disclosure, Bonus, Meeting, Plan, Services, and Report. **Individuals** could be Kenneth Lay, Jeffrey Skilling, Andrew Fastow, Richard Causay, Micheal Kapper, Lea Fastow, Ben Glisan, Dave Delainey, Mart Koenig, and Lou Lung Pai.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Keywords | Individual A | Individual B | Individual C | Individual D |
| Fraud | 808 | 808 | 808 | 808 |
| Enron | 894 | 894 | 894 | 894 |
| FERC | 866 | 866 | 866 | 866 |
| Regulations | 872 | 872 | 872 | 872 |
| - |  |  |  |  |
| - |  |  |  |  |

2. Need both CSV and json format

3. SQL query