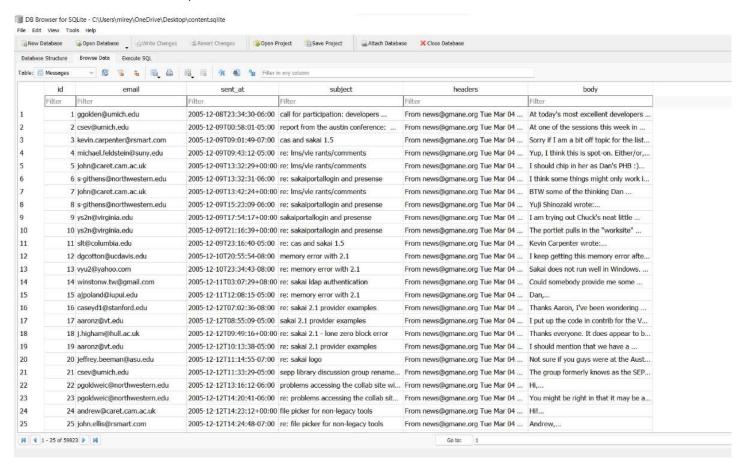
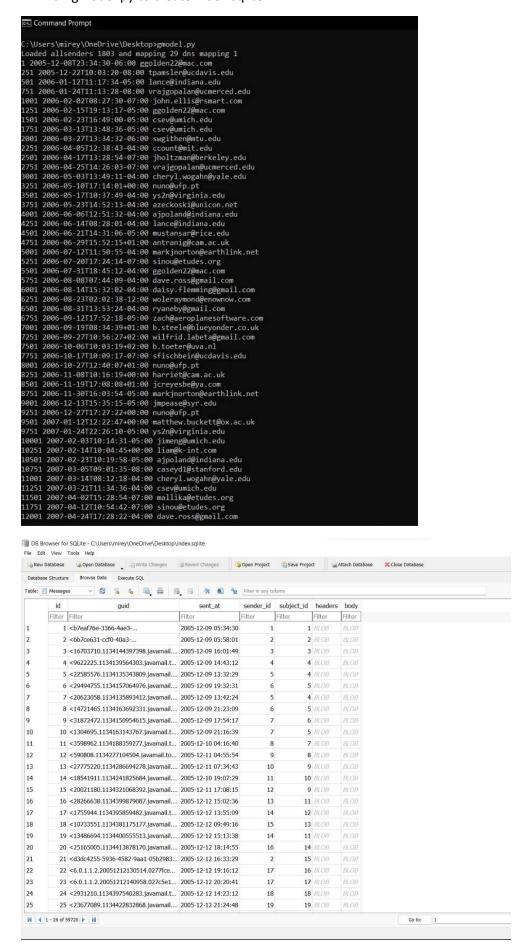
## Mailing List Data - Part I

In this assignment you will download some of the mailing list data from http://mbox.dr-chuck.net/ and run the data cleaning / modeling process and take some screen shots. All screenshots should use the same data set.

## 1. Download Content.sqlite



## 2. Run gmodel.py to create index.sqlite



## 3. gbasic.py program to compute basic histogram data

```
Command Prompt
59251 2015-04-22T10:08:08+02:00 bgarcia@entornosdeformacion.com
59501 2015-05-13T10:50:56+00:00 adam.marshall@ox.ac.uk
:\Users\mirey\OneDrive\Desktop>gbasic.py
How many to dump? 20
Loaded messages= 59720 subjects= 29472 senders= 1800
Top 20 Email list participants
steve.swinsburg@swinsborg.com 3337
azeckoski@unicon.net 1907
ian@cam.ac.uk 1591
csev@umich.edu 1490
david.horwitz@uct.ac.za 1221
matthew@longsight.com 1209
neal.caidin@apereo.org 1035
stephen.marquard@uct.ac.za 931
adam.marshall@ox.ac.uk 929
arwhyte@umich.edu 868
adrian.r.fish@gmail.com 764
slt@columbia.edu 732
jimeng@umich.edu 709
hedrick@rutgers.edu 672
clay.fenlason@gatech.edu 670
markjnorton@earthlink.net 669
ottenhoff@longsight.com 653
chmaurer@indiana.edu 619
bkirschn@umich.edu 599
swgithen@mtu.edu 585
Top 20 Email list organizations
umich.edu 6805
gmail.com 5683
swinsborg.com 3337
cam.ac.uk 2626
uct.ac.za 2582
indiana.edu 2333
longsight.com 2315
unicon.net 2310
ox.ac.uk 1544
berkeley.edu 1388
stanford.edu 1355
ucdavis.edu 1184
rsmart.com 1109
apereo.org 1038
rutgers.edu 908
etudes.org 895
gatech.edu 865
virginia.edu 765
columbia.edu 746
earthlink.net 670
C:\Users\mirey\OneDrive\Desktop>_
```