## further analysis

## May 16, 2016

```
In [1]: %matplotlib inline
        import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np
        # Make the graphs a bit prettier, and bigger
        pd.set_option('display.mpl_style', 'default')
        plt.rcParams['figure.figsize'] = (15, 5)
        plt.rcParams['font.family'] = 'sans-serif'
        # This is necessary to show lots of columns in pandas 0.12.
        # Not necessary in pandas 0.13.
        pd.set_option('display.width', 5000)
        pd.set_option('display.max_columns', 60)
In [2]: cd md
/home/raisa/md
In [54]: all_df=[]
         nfiles=15
         for i in range(nfiles):
             filename = 'msample%d.csv' % i
             print i
             all_df.append(pd.read_csv(filename, header=None))
0
1
2
3
4
5
6
7
8
9
10
11
12
13
In [55]: all_df[0]
```

Date [55]: 0	Out [55]:		0		2	3	4	
1 9 U1018D0M1 U1018D0M1 C1862 C1862 C1862 C2663 33 C202588D0M1 C202588D0M1 C202588D0M1 C2653 C2663 C26	out [00].	0						
2 33 47 C205580DM1 C265380DM1 C265380DM1 C2653 C2653 C565   4 54 C265380DM1 C265380DM1 C265380DM1 C2653 C2653 C2653   5 55 C265380DM1 C265380DM1 C265380DM1 C2653 C2653   6 95 U668DDM1 U668DDM1 C111480DM1 C1115 C1114   8 164 C111480DM1 C111480DM1 C1115 C1114   9 174 C26928DM1 C26928DM1 C26928DM1 C2627 C1315   10 205 U2528DDM1 U5528DDM1 C528 C528   11 213 C5998BDM1 C3998DDM1 C3998DDM1 C528 C528   12 239 C39908DDM1 C3998DDM1 C3998DDM1 C539 C3392   13 243 U228DDM1 U5528DDM1 C3998DDM1 C3392 C3392   13 243 U228DDM1 U228DDM1 C4677 C457   15 308 C206587 C299587 C299587 C25240   16 335 U44DDM1 C11480DM1 C1122   17 355 C171489DDM1 C16778DDM1 C467 C457   17 355 C171489DM1 C16778DDM1 C312   19 454 C2096587 C299587 C299587 C457 C457   20 489 C2905880M1 C3908DDM1 C3633 C2653   21 523 C43348DDM1 C15778DDM1 C1527 C612   21 523 C43348DDM1 C15778DDM1 C2692 C1065   22 554 C26538DDM1 C26538DDM1 C26538DDM1 C2692   23 571 U152567 U152567   24 623 U228DDM1 C26538DDM1 C26538DDM1 C2663   25 641 C36538DDM1 C36538DDM1 C36538DDM1 C3666   26 673 C20438BDDM1 C36538DDM1 C2653   26 673 C20438BDDM1 C36538DDM1 C2653   26 673 C20438BDDM1 C36538DDM1 C3666   27 677 C27598DDM1 C26538DDM1 C26538DDM1 C3666   26 673 C20438BDDM1 C36538DDM1 C3666   26 673 C20438BDDM1 C36538DDM1 C3666   26 673 C20438BDDM1 C36538DDM1 C3666   27 677 C27598DDM1 C26538DDM1 C26538DDM1 C3666   27 677 C27598DDM1 C36538DDM1 C3666   27 677 C27598DDM1 C36538DDM1 C3666   27 677 C2759BDM1 C36538DDM1 C3658DDM1 C3666   27 677 C2759BDM1 C36538DDM1 C3666   27 677 C2759BDM1 C36538DDM1 C3666   27 677 C2759BDM1 C36538DDM1 C3666   27 677 C2759BDM1 C3658BDM1 C3666   27 676 C375 C375 C375 C375 C375 C3								
3         47         C26538BDDM1         C2653BDDM1         C2653BDDM1         C2653BDDM1         C11148BDDM1         C1114BDDM1         C1115         C1114         C1114BDDM1         C1114BDDM1         C1115         C1114BDDM1         C1114BDDM1         C1114BDDM1         C1114BDDM1         C2627         C12540         C2553         C553         C553         C553         C33902BDDM1         C2290EDDM1         C229040         C2540         C2540         C25640         <								
4         54         C265388DDM1         C26538DDM1         C2653         C586           5         55         C26538DDM1         C26533BDM1         C2653         C2653           6         95         U668DDM1         U1668DDM1         C2613         C2613           7         128         C11148BDDM1         C11148BDDM1         C1114         C1114           8         164         C11148BDDM1         C26928BDDM1         C116         C528         C528           10         205         U2526DDM1         U2526DDM1         C152         C1315         C114         C114         C252         C1315         C114         C293         C33908BDDM1         C38998BDDM1         C3592         C1315         C152         C3392         C2391         C33902         C3392         C25240         C267         C267         C457         C4								
5         55         C265380D0M1         C265380D0M1         C2632           7         128         C1114\$0D0M1         C1114\$0D0M1         C1115         C1114           8         164         C1114\$0D0M1         C1114\$0D0M1         C1114         C1114           9         174         C2692\$80D0M1         C2692\$80D0M1         C528         C528           10         205         U25280D0M1         U25280D0M1         C5998*0D0M1         C5593         C553           11         213         C5998*0D0M1         C5998*0D0M1         C553         C553           12         239         C33098*0D0M1         C3309*0E0D0M1         C3309*0E0D0M1         C3309*0E0D0M1         C457         C457           14         286         C1607\$0D0M1         C1607*0E0D0M1         C457         C457         C457           15         308         C20688*7         C2068**         C25240         C25240         C25240           16         335         U49D0M1         C1612         C612         C612         C229           17         355         C1714\$80D0M1         C1714\$80D0M1         C152**0E0DM1         C152**0E0DM1         C152**0E0DM1         C152**0E0DM1         C152**0E0DM1         C152**0E0DM1								
66         95         UGGGDUM1         UGGGDUM1         C312         C322         C322           7         128         C11148DDM1         C11148DDM1         C1114         C1								
7 128 C1114\$BDDM1 C1114\$BDDM1 C1115 C1114  8 164 C1114\$BDDM1 C1115 C1114  9 174 C2692\$BDDM1 C2692\$BDDM1 C1553 C528  10 205 UZ52BDDM1 UZ52BDDM1 C5653 C528  11 213 C599\$BDDM1 C599\$BDDM1 C5653 C553  12 239 C3309\$BDDM1 C3309\$BDDM1 C3309\$BDDM1 C3392 C3392  13 243 U22BDDM1 U22BDDM1 C477 U22  14 226 C1607\$BDDM1 C1607\$BDDM1 C477 U22  14 226 C1607\$BDDM1 C1607\$BDDM1 C477 U22  15 308 C209\$\$97 C2096\$\$97 C25240 C25240  16 335 U4BDDM1 U4BDDM1 C427 C25240  17 355 C1714\$BDDM1 C1774\$BDDM1 C1527\$BDDM1 C512 C612  18 363 C1527\$BDDM1 C1527\$BDDM1 C1527\$BDDM1 C512 C612  19 454 C2095\$BDM1 C3902\$BDM1 C2902\$BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2903\$BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2902\$C255BDM1 C2904\$BDM1 C2904\$C568\$BDM1 C2902\$C255BDM1 C2904\$BDM1 C2904\$C568\$BDM1 C2902\$C255BDM1 C2904\$BDM1 C2904\$C568\$BDM1 C2902\$C255BDM1 C2904\$BDM1 C2904\$C568\$BDM1 C2902\$C252\$C222\$C222\$BDM1 C2900M1 C2900M1 C2904\$C568\$BDM1 C2904\$C568\$BDM1 C2902\$C252\$C222\$C222\$BDM1 C2900M1 C2900M1 C2900M1 C2904\$C568\$C568\$BDM1 C2904\$BDM1 C2902\$C252\$C222\$C222\$C222\$C222\$C222\$C222\$C2								
8         164         C1114\$000M1         C1114\$000M1         C1115         C1114           9         174         C2692\$00M1         C262         C528           10         205         U352\$000M1         C2625         C1315           11         213         C599\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C399\$\$000M1         C397\$\$000M1         C477         U22         C22								
9 174   C2692\$mD0M1   C2692\$mD0M1   C528   C528   10 205								
10								
11								
12   239								
13 243 U22@DOM1 U22@DOM1 C477 U22 14 286 C1607\$@DM1 C1607\$@DM1 C457 C457 15 308 C2096\$@? C2096\$@? C2096\$@? C2594 16 335 U4@DOM1 U4@DOM1 C229 C229 17 355 C1714\$@DM1 C1507\$@DM1 C1527\$@DM1 C								
14         286         C1607\$@DDM1         C1607\$@DDM1         C457         C457           15         308         C2096\$@?         C2540         C25240         C25240           16         335         U4@DDM1         U4@DDM1         C229         C229           17         355         C1714\$@DDM1         C1714\$@DDM1         C612         C612           18         363         C1527\$@DDM1         C1527\$@DDM1         C612         C612           19         454         C2096\$@?         C2096\$@?         C457         C457           20         469         C2902\$@DDM1         C2906\$@?         C612         C6612           24         623         U22@DDM1         U22@DDM1         C2663         C2653         C2653         C2612         C467         C467         C467         C467         C467         C467         C467         C467         C4664         C457         C467         C4612         C4612         C								
15   308								
16         335         U40D0M1         U40D0M1         C229         C229           17         355         C1714\$0D0M1         C1714\$0D0M1         C1527\$0D0M1         C1527         C612           19         454         C2096\$80?         C2092\$0D0M1         C1927\$0D0M1         C1527\$0D0M1         C1527         C457           20         489         C2902\$0D0M1         C2902\$0D0M1         C2902\$0D0M1         C2902\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C2653\$0D0M1         C26612         C612								
17								
18         363         C1527\$@D0M1         C1527\$@D0M1         C1527\$ C612           19         454         C2096\$@?         C2096\$@?         C457         C457           20         489         C2902\$@D0M1         C2902\$@D0M1         C2902\$@D0M1         C2902         C1065           21         523         C4334\$@D0M1         C4334\$@D0M1         C4334         C2106           22         554         C2653\$@D0M1         C2653\$@D0M1         C2653         C612         C612           24         623         U22@D0M1         U22@D0M1         C506         U22         C529           25         641         C860\$@D0M1         C860\$@D0M1         C506         C457         C529           26         673         C2043\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759\$@D0M1         C2759         C2759         MICROSOFT_AUTY <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>								
19								
20								
21         523         C4334\$@DOM1         C4334\$@DOM1         C4334         C2106           22         554         C2653\$@DOM1         C2653\$@DOM1         C2653         C2653           23         571         U18256?         U18256?         C612         C612           24         623         U22@DOM1         U22@DOM1         C506         U22           25         641         C860\$@DOM1         C260\$@DOM1         C529         C529           27         677         C2759\$@DOM1         C27599         C2759         C2759           28         773         LOCAL SERVICE@C3049         LOCAL SERVICE@C3049         C3049         C3049         C3049           29         834         C2480\$@DOM1         C2480\$@DOM1         C5479         C2479         MICROSOFT_AUT           1 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>								
22         554         C2653\$@DOM1         C2653 & C2653           23         571         U1825@?         U1825@?         C612         C612           24         623         U22@DOM1         U22@DOM1         C566         U22           25         641         C860\$@DOM1         C860\$@DOM1         C560         C457           26         673         C2043\$@DOM1         C2759\$@DOM1         C2759         C2759           27         677         C2759\$@DOM1         C2759\$@DOM1         C2759         C2759           28         773         LOCAL SERVICE@C3049         LOCAL SERVICE@C3049         C3049         C3049           29         834         C2480\$@DOM1         C586\$@DOM1         C2479         C2479         MICROSOFT_AUT                    400682         5010840         U9@DOM1         C586\$@DOM1         C566         C586           400683         5010841         U590?         U590?         C1634         C1634           400684         5010873         U590?         U590?         C1634         C1634           400685         5010874         U90?         U22								
23   571								
24         623         U22@DOM1         U22@DOM1         C506         U22           25         641         C860\$@DOM1         C860\$@DOM1         C860         C457           26         673         C2043\$@DOM1         C2759\$@DOM1         C529         C529           27         677         C2759\$@DOM1         C2759\$@DOM1         C2759         C2759           28         773         LOCAL SERVICE@C3049         LOCAL SERVICE@C3049         C3049         C3049         C3049           29         834         C2480\$@DOM1         C2480\$@DOM1         C2479         C2479         MICROSOFT_AUT  .								
25         641         C860\$@DOM1         C860\$@DOM1         C860         C457           26         673         C2043\$@DOM1         C2043\$@DOM1         C529         C529           27         677         C2759\$@DOM1         C2759\$@DOM1         C2759         C2759           28         773         LOCAL SERVICEGC3049         LOCAL SERVICEGC3049         C3049         C3049           29         834         C2480\$@DOM1         C2480\$@DOM1         C2479         C2479         MICROSOFT_AUTF                      400682         5010840         U9@DOM1         C586 \$@DOM1         C586         C586           400683         5010841         U590?         U590?         C1634         C1634           400684         5010873         U590?         U99?         C222         C222           400685         5010879         U22@DOM1         U22@DOM1         C349         U22           400689         5010849         NETWORK SERVICE@C25102         C25102         C25102         C25102           400690         5010980         NETWORK SERVICE@C25102         C25102         C25102								
26         673         C2043\$@DOM1         C2043\$@DOM1         C529         C529           27         677         C2759\$@DOM1         C2759\$@DOM1         C2759         C2759           28         773         LOCAL SERVICE@C3049         LOCAL SERVICE@C3049         C3049         C3049           29         834         C2480\$@DOM1         C2480\$@DOM1         C2479         MICROSOFT_AUTH                    400682         5010840         U9@DOM1         C586\$@DOM1         C586         C586           400683         5010861         U99029?         U590?         C1634         C1634           400684         5010873         U590?         U590?         C1634         C1634           400685         5010873         U590?         U990?         C222         C222           400687         5010879         U22@DOM1         U22@DOM1         C249         U22           400688         5010879         U990DOM1         U90DOM1         C224         C25102         C25102           400690         5010900         C23484\$@DOM1         C1692\$@DOM1         C1692\$@DOM1         C1692\$@DOM1         C1692								
27         677         C2759\$\$DDM1         C2759\$\$DDM1         C2759         C2759           28         773         LOCAL SERVICE@C3049         LOCAL SERVICE@C3049         C3049         C3049         C3049           29         834         C2480\$@DM1         C2480\$@DM1         C2479         C2479         MICROSOFT_AUTH <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>								
28								
29 834 C2480\$@DOM1 C2480\$@DOM1 C2479 C2479 MICROSOFT_AUTH								
								MICROSOFT AUTH
400682         5010840         U9@DOM1         C586\$@DOM1         C586         C586           400683         5010841         U59@?         U59@?         C1634         C1634           400684         5010861         U8929@?         U8929@?         C19037         C19037           400685         5010873         U59@?         U59@?         C1634         C1634           400686         5010879         U22@DOM1         U9@?         C222         C222           400687         5010879         U9@DOM1         U9@DOM1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@DOM1         C1692\$@DOM1         C1692         C1692           400691         5010907         C1692\$@DOM1         C743\$@DOM1         C743\$@DOM1         C586         C586           400692         5010916         C743\$@DOM1         C743\$@DOM1         C467         C222         C222<								
400683         5010841         U590?         U590?         C1634         C1634           400684         5010861         U89290?         U89290?         C19037         C19037           400685         5010873         U590?         U590?         C1634         C1634           400686         5010874         U90?         U90?         C222         C222           400687         5010879         U220D0M1         U220D0M1         C849         U22           400688         5010879         U90D0M1         U90D0M1         C222         C222           400689         5010884         NETWORK SERVICE0C25102         NETWORK SERVICE0C25102         C25102         C25102           400690         5010900         C23484\$0D0M1         C1692\$0D0M1         C1692         C1692           400691         5010907         C1692\$0D0M1         C1692\$0D0M1         C1692         C1692           400692         5010916         C743\$0D0M1         C743\$0D0M1         C586         C586           400693         5010938         U90?         U90?         C222         C222           400695         5010970         U220D0M1         U220D0M1         C246         U22           400697         5011								
400684         5010861         U8929@?         U8929@?         C19037         C19037           400685         5010873         U59@?         U59@?         C1634         C1634           400686         5010874         U9@?         U9@?         C222         C222           400687         5010879         U22@DOM1         U22@DDM1         C849         U22           400688         5010879         U9@DOM1         U9@DDM1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C234844@DOM1         C1692*@DDM1         C1692 *C1692           400691         5010907         C1692*@DDM1         C743*@DDM1         C586         C586           400692         5010916         C743*@DDM1         C743*@DDM1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222           400694         5010963         C2344*@DDM1         C2344*@DDM1         C457         C457           400695         5010970         U22@DDM1         U22@DDM1         C246         U22           400696         5011005								
400685         5010873         U59@?         U59@?         C1634         C1634           400686         5010874         U9@?         U9@?         C222         C222           400687         5010879         U22@D0M1         U22@D0M1         C849         U22           400688         5010879         U9@D0M1         U9@D0M1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@D0M1         C1692\$@D0M1         C1692         C1692           400691         5010907         C1692\$@D0M1         C743\$@D0M1         C586         C586           400692         5010916         C743\$@D0M1         C743\$@D0M1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222           400694         5010963         C2344\$@D0M1         C2344\$@D0M1         C457         C457           400695         5010970         U22@D0M1         U22@D0M1         C246         U22           400696         5011005         U59@?         U59@?         C1634         C1634           400699         501								
400686         5010874         U9@?         U9@?         C222         C222           400687         5010879         U22@DOM1         U22@DOM1         C849         U22           400688         5010879         U9@DOM1         U9@DOM1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@DOM1         C1692\$@DDM1         C1692\$@DDM1         C1692         C1692           400691         5010907         C1692\$@DDM1         C743\$@DDM1         C586         C586           400692         5010916         C743\$@DDM1         C743\$@DDM1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222         C222           400694         5010963         C2344\$@DDM1         C2344\$@DDM1         C457         C457           400695         5010970         U22@DDM1         U22@DDM1         C246         U22           400696         5011005         U59@?         U59@?         C1634         C1634           400698         5011014         U101@?         U59@?         C589         C589 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>								
400687         5010879         U22@DOM1         U22@DOM1         C849         U22           400688         5010879         U9@DOM1         U9@DOM1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@DOM1         C1692\$@DOM1         C1692         C1692           400691         5010907         C1692\$@DOM1         C1692\$@DOM1         C1692         C1692           400692         5010916         C743\$@DOM1         C743\$@DOM1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222           400694         5010963         C2344\$@DOM1         C2344\$@DOM1         C457         C457           400695         5010970         U22@DOM1         U22@DOM1         U22@DOM1         C246         U22           400696         5011005         U59@?         U59@?         C1634         C1634           400697         5011008         C3188\$@DOM1         C3188\$@DOM1         C3188         C3415           400699         5011015         U59@?         U59@?         C589         C589      <								
400688         5010879         U9@DOM1         U9@DOM1         C222         C222           400689         5010884         NETWORK SERVICE@C25102         NETWORK SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@DOM1         C23484\$@DOM1         C234844         C586           400691         5010907         C1692\$@DOM1         C1692\$@DOM1         C1692         C1692           400692         5010916         C743\$@DOM1         C743\$@DOM1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222           400694         5010963         C2344\$@DOM1         C457         C457           400695         5010970         U22@DOM1         U22@DOM1         C246         U22           400696         5011005         U59@?         U59@?         C1634         C1634           400697         5011008         C3188\$@DOM1         C3188\$@DOM1         C3188\$@DOM1         C3188         C3415           400699         5011015         U59@?         U59@?         C589         C589           400700         5011043         U10107@DOM1         U10107@DOM1         C419         C419           4								
400689         5010884         NETWORK         SERVICE@C25102         NETWORK         SERVICE@C25102         C25102         C25102           400690         5010900         C23484\$@DOM1         C23484\$@DOM1         C23484\$         C586           400691         5010907         C1692\$@DOM1         C1692\$@DOM1         C1692         C1692           400692         5010916         C743\$@DOM1         C743\$@DOM1         C586         C586           400693         5010938         U9@?         U9@?         C222         C222           400694         5010963         C2344\$@DOM1         C2344\$@DOM1         C457         C457           400695         5010970         U22@DOM1         U22@DOM1         C246         U22           400696         5011005         U59@?         U59@?         C1634         C1634           400697         5011008         C3188\$@DOM1         C3188\$@DOM1         C3188\$@DOM1         C3188\$         C3188           400699         5011015         U59@?         U59@?         C589         C589           400700         5011043         U10107@DOM1         U10107@DOM1         C419         C419           400701         5011067         C27118\$@DOM1         C21596\$@DOM1								
400690       5010900       C23484\$@DOM1       C23484\$@DOM1       C23484\$       C586         400691       5010907       C1692\$@DOM1       C1692\$@DOM1       C1692\$       C1692         400692       5010916       C743\$@DOM1       C743\$@DOM1       C586       C586         400693       5010938       U9@?       U9@?       C222       C222         400694       5010963       C2344\$@DOM1       C2344\$@DOM1       C457       C457         400695       5010970       U22@DOM1       U22@DOM1       C246       U22         400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1								
400691       5010907       C1692\$@DOM1       C1692\$@DOM1       C1692       C1692         400692       5010916       C743\$@DOM1       C743\$@DOM1       C586       C586         400693       5010938       U9@?       U9@?       C222       C222         400694       5010963       C2344\$@DOM1       C2344\$@DOM1       C457       C457         400695       5010970       U22@DOM1       U22@DOM1       C246       U22         400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       <			5010900					
400692       5010916       C743\$@DOM1       C743\$@DOM1       C586       C586         400693       5010938       U9@?       U9@?       C222       C222         400694       5010963       C2344\$@DOM1       C2344\$@DOM1       C457       C457         400695       5010970       U22@DOM1       U22@DOM1       C246       U22         400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596       C612		400691	5010907	C1692\$@DOM:	C1692\$@DOM1	C1692	C1692	
400693       5010938       U9@?       U9@?       C222       C222         400694       5010963       C2344\$@DOM1       C2344\$@DOM1       C457       C457         400695       5010970       U22@DOM1       U22@DOM1       C246       U22         400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1				·				
400695       5010970       U22@DOM1       U22@DOM1       C246       U22         400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1		400693	5010938	U9@1	U9@?		C222	
400696       5011005       U59@?       U59@?       C1634       C1634         400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1		400694	5010963	C2344\$@DOM:	C2344\$@DOM1	C457	C457	
400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1		400695	5010970	U22@DOM:	U22@DOM1	C246	U22	
400697       5011008       C3188\$@DOM1       C3188\$@DOM1       C3188       C3188         400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1		400696		U59@1	U59@?			
400698       5011014       U101@?       U101@?       C3415       C3415         400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1								
400699       5011015       U59@?       U59@?       C589       C589         400700       5011043       U10107@DOM1       U10107@DOM1       C419       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1								
400700       5011043       U10107@DOM1       U10107@DOM1       C419         400701       5011067       C27118\$@DOM1       C27118\$@DOM1       C1369         400702       5011071       C21596\$@DOM1       C21596\$@DOM1       C21596\$@DOM1								
4007015011067C27118\$@DOM1C27118\$@DOM1C1369C13694007025011071C21596\$@DOM1C21596\$@DOM1C21596\$@DOM1C21596								
400702 5011071 C21596\$@DOM1 C21596\$@DOM1 C21596 C612								
400703 5011083 U9@? U9@? C222 C222		400702	5011071	C21596\$@DOM:	C21596\$@DOM1	C21596	C612	
		400703	5011083	U9@1	Y U90?	C222	C222	

```
400704 5011087
                                          U9@DOM1
         400705 5011110
                                       C398$@DOM1
         400706 5011116
                                      C1791$@DOM1
         400707 5011120
                                      C1617$@DOM1
         400708 5011157
                                      C7780$@DOM1
         400709 5011161
                                         U22@D0M1
         400710 5011167
                                       U6715@DOM1
         400711 5011195
                                        U199@DOM1
         [400712 rows x 9 columns]
In [56]: Y=[]
         for i in range(nfiles):
             Y.append(all_df[i][8]=='Success')
In [13]: Y[1]
Out[13]: 0
                   False
                   False
         1
         2
                   False
         3
                    True
         4
                    True
         5
                   False
         6
                    True
         7
                    True
         8
                    True
         9
                   False
         10
                   False
         11
                   False
         12
                    True
         13
                   False
                    True
         14
                    True
         15
         16
                   False
                    True
         17
         18
                    True
                    True
         19
         20
                   False
         21
                   False
         22
                    True
         23
                    True
         24
                    True
         25
                    True
         26
                   False
         27
                   False
         28
                    True
         29
                   False
                   . . .
         400276
                    True
         400277
                   False
                    True
         400278
         400279
                   False
         400280
                   False
         400281
                    True
         400282
                    True
```

U9@DOM1

C398\$@DOM1

C1791\$@DOM1

C1617\$@DOM1

C7780\$@DOM1

U1825@DOM1

U22@D0M1

C222

C1767

C1065

C1618

C7780

C965

C1929

U6715@DOM1 C10781 C10781

C222

C1767

C1065

C457

C528

C1929

U22

```
400283
                   False
         400284
                    True
         400285
                   False
         400286
                    True
         400287
                    True
         400288
                   False
         400289
                    True
         400290
                   False
         400291
                    True
         400292
                    True
         400293
                   False
                   False
         400294
         400295
                   False
         400296
                   True
         400297
                   False
         400298
                   False
         400299
                   True
         400300
                   False
         400301
                    True
         400302
                   False
         400303
                   False
         400304
                    True
         400305
                   False
         Name: 8, dtype: bool
In [57]: def map_user(x):
             if x.startswith('C'):
                 return 'C'
             elif x.startswith('U'):
                 return 'U'
             else:
                 return x
In [68]: X=[]
         for i in range(nfiles):
             df=all_df[i]
             df["source_user"], df["source_domain"] = zip(*df[1].str.split('0').tolist())
             df["source_user"] = df["source_user"].str.rstrip('$')
             df["destination_user"], df["destination_domain"] = zip(*df[2].str.split('0').tolist())
             df["destination_user"] = df["destination_user"].str.rstrip('$')
             df['source_class']=df['source_user'].map(map_user)
             df['destination_class']=df['destination_user'].map(map_user)
             x=pd.DataFrame.from_items([
             ('time', (df[0]%(24*60*60)).astype(int))])
             x['same_user'] = (df['destination_user'] == df['source_user'])
             x['same_domain']=(df['destination_domain']==df['source_domain'])
             x['source_user_comp_same']=(df[3]==df['source_user'])
             x['destination_user_comp_same']=(df['destination_user']==df[4])
             x['same_comp']=(df[3]==df[4])
             x['source_domain_comp_same']=(df[3]==df['source_domain'])
             x['destination_domain_comp_same']=(df['destination_domain']==df[4])
             for j in [5,6, 7]:
                 for label in sorted(df[j].unique()):
                     if label=='?':
```

```
if j==5:
                                x['?_authentication type']=(df[j]==label)
                                x['?_logon type']=(df[j]==label)
                       else:
                           x[label]=(df[j]==label)
              for cl in ['source_class', 'destination_class']:
                   for label in sorted(df[cl].unique()):
                       if cl=='source_class':
                           x['source_'+label]=(df[cl]==label)
                           x['destination_'+label]=(df[cl]==label)
              X.append(x)
In [62]: X[1]
Out [62]:
                    time same_user same_domain source_user_comp_same destination_user_comp_same same_comp so
          0
                       2
                               True
                                            True
                                                                   False
                                                                                                   True
                                                                                                             False
                       3
                                                                                                              True
          1
                               True
                                            True
                                                                   False
                                                                                                 False
          2
                      11
                               True
                                                                   False
                                                                                                  False
                                                                                                             False
                                            True
          3
                     140
                               True
                                            True
                                                                     True
                                                                                                 False
                                                                                                             False
          4
                                                                                                 False
                     176
                               True
                                            True
                                                                   False
                                                                                                              True
          5
                     185
                               True
                                                                   False
                                                                                                 False
                                                                                                              True
                                            True
          6
                     224
                               True
                                            True
                                                                    True
                                                                                                 False
                                                                                                             False
          7
                     250
                                                                                                              True
                               True
                                            True
                                                                   False
                                                                                                 False
          8
                                                                                                             False
                     252
                               True
                                            True
                                                                   False
                                                                                                 False
          9
                                                                                                              True
                     333
                               True
                                            True
                                                                    True
                                                                                                   True
          10
                     348
                               True
                                            True
                                                                   False
                                                                                                   True
                                                                                                             False
          11
                     416
                               True
                                            True
                                                                   False
                                                                                                   True
                                                                                                             False
          12
                     459
                               True
                                                                                                   True
                                                                                                              True
                                            True
                                                                     True
          13
                     470
                               True
                                            True
                                                                     True
                                                                                                   True
                                                                                                              True
          14
                     485
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                             False
          15
                     490
                               True
                                            True
                                                                     True
                                                                                                 False
                                                                                                             False
          16
                     510
                               True
                                            True
                                                                   False
                                                                                                   True
                                                                                                             False
          17
                     542
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          18
                     551
                               True
                                            True
                                                                     True
                                                                                                 False
                                                                                                             False
          19
                     570
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          20
                                                                   False
                                                                                                             False
                     588
                               True
                                            True
                                                                                                   True
          21
                                                                                                             False
                     623
                               True
                                            True
                                                                   False
                                                                                                   True
          22
                     679
                               True
                                            True
                                                                                                   True
                                                                                                             False
                                                                   False
          23
                     704
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          24
                     726
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          25
                     745
                               True
                                            True
                                                                     True
                                                                                                 False
                                                                                                             False
          26
                     750
                               True
                                                                                                              True
                                            True
                                                                   False
                                                                                                 False
          27
                     859
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          28
                     936
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          29
                     975
                                                                                                              True
                               True
                                            True
                                                                   False
                                                                                                 False
                                . . .
                                             . . .
                                                                      . . .
                                                                                                               . . .
                     . . .
                                                                                                    . . .
          . . .
          400276
                  86055
                               True
                                                                   False
                                                                                                 False
                                                                                                             False
                                            True
          400277
                  86057
                               True
                                                                   False
                                                                                                 False
                                                                                                              True
                                            True
          400278
                                                                                                              True
                  86061
                               True
                                            True
                                                                    True
                                                                                                   True
          400279
                   86073
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          400280
                   86089
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                              True
          400281
                  86093
                               True
                                            True
                                                                   False
                                                                                                 False
                                                                                                             False
          400282
                                                                                                              True
                  86104
                               True
                                            True
                                                                   False
                                                                                                 False
```

```
True
                                                              False
                                                                                          False
         400284
                 86131
                                         True
         400285 86151
                            True
                                         True
                                                              False
                                                                                          False
         400286 86179
                            True
                                         True
                                                               True
                                                                                          False
         400287
                 86251
                            True
                                         True
                                                              False
                                                                                          False
         400288 86259
                                                                                          False
                            True
                                         True
                                                              False
         400289
                 86267
                            True
                                         True
                                                               True
                                                                                           True
                            True
         400290
                 86267
                                         True
                                                              False
                                                                                          False
         400291
                 86275
                            True
                                         True
                                                              False
                                                                                          False
         400292
                 86280
                            True
                                         True
                                                              False
                                                                                          False
         400293
                 86281
                            True
                                         True
                                                              False
                                                                                          False
         400294
                 86281
                            True
                                         True
                                                              False
                                                                                           True
         400295
                 86287
                            True
                                         True
                                                              False
                                                                                          False
                                                                                          False
         400296 86298
                            True
                                         True
                                                              False
         400297
                 86341
                            True
                                                                                          False
                                         True
                                                              False
         400298
                 86347
                            True
                                         True
                                                              False
                                                                                          False
         400299
                 86354
                            True
                                         True
                                                              False
                                                                                          False
         400300
                 86356
                            True
                                         True
                                                              False
                                                                                          False
         400301 86372
                            True
                                         True
                                                              False
                                                                                          False
         400302 86373
                            True
                                         True
                                                              False
                                                                                          False
         400303 86374
                            True
                                         True
                                                              False
                                                                                          False
         400304 86391
                            True
                                                              False
                                                                                          False
                                         True
         400305 86393
                                                                                          False
                            True
                                         True
                                                              False
         [400306 rows x 56 columns]
In [63]: X[0].columns
Out [63]: Index([u'time', u'same_user', u'same_domain', u'source_user_comp_same', u'destination_user_comp_s
In [64]: [len(entry.columns) for entry in X]
Out[64]: [53, 56, 53, 54, 54, 52, 56, 56, 57, 55, 55, 54, 55, 54, 54]
In [65]: all_col = set(sum([list(entry.columns) for entry in X], []))
         [all_col.difference(list(entry.columns)) for entry in X]
Out [65]: [{'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'MICROSOFT_AUTHENTICATION_PACK'},
          {'CygwinLsa', 'MICROSOFT_AUTHENTICA', 'MICROSOFT_AUTHENTICATION_P'},
          { 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'MICROSOFT_AUTHENTICATION_PAC'},
          { 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'CygwinLsa',
```

400283 86127

True

True

False

False

True

False

False

False

'MICROSOFT\_AUTHENTICA', 'MICROSOFT\_AUTHENTICATION\_P', 'MICROSOFT\_AUTHENTICATION\_PAC'},

```
'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PAC'},
          { 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PAC'
           'MICROSOFT_AUTHENTICATION_PACK',
           'Setuid'},
          {'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA'},
          { 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_PAC'},
          {'CygwinLsa', 'MICROSOFT_AUTHENTICATION_P'},
          { 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_PACKAGE_V1'},
          {'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'MICROSOFT_AUTHENTICATION_PAC'},
          {'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'MICROSOFT_AUTHENTICATION_PAC'},
          {'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P'
           'MICROSOFT_AUTHENTICATION_PA'},
          {'CygwinLsa',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'Setuid'},
          {'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE',
           'MICROSOFT_AUTHENTICA',
           'MICROSOFT_AUTHENTICATION_P',
           'MICROSOFT_AUTHENTICATION_PA',
           'MICROSOFT_AUTHENTICATION_PAC'}]
In [69]: col_set = [set(entry.columns) for entry in X]
         common_subset = set.intersection(*col_set)
         drop_cols = [e.difference(common_subset) for e in col_set]
         for entry, to_drop in zip(X, drop_cols):
             print 'dropping', to_drop
             for item in to_drop:
                 del entry[item]
```

{ 'ACRONIS\_RELOGON\_AUTHENTICATION\_PACKAGE',

```
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACKAGE_V1', 'MICROSOFT_AUTHENTICATION_PAC'])
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICROSO
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACK', 'MICROSOFT_AUTHENTICATION_PACKAGE_V1'])
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'MICROSOFT_AUTHENTICATION_PACKAGE_V1', 'MICROSOFT_
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'MICROSOFT_AUTHENTICATION_PACKAGE_V1', 'MICROSOFT_
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'MICROSOFT_AUTHENTICATION_PACKAGE_V1'])
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACK', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'Cygwi
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'MICROSOFT_AUTHENTICATION_P', 'CygwinLsa', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_AUTHENTICATION_P', '
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICROSO
dropping set(['MICROSOFT_AUTHENTICATION_PA', 'Setuid', 'MICROSOFT_AUTHENTICATION_P', 'MICROSOFT_
dropping set(['Setuid', 'CygwinLsa', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICROSOFT_AUTHENTICATION_
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACK', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICRO
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACK', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICRO
dropping set(['MICROSOFT_AUTHENTICATION_PACK', 'ACRONIS_RELOGON_AUTHENTICATION_PACKAGE', 'MICROSOFT_AUTHE
dropping set(['Setuid', 'MICROSOFT_AUTHENTICATION_PACK', 'CygwinLsa', 'MICROSOFT_AUTHENTICATION_PACKAGE_N
In [70]: col0 = list(X[0].columns)
                      for i in range(1,nfiles):
                                col_i = list(X[i].columns)
                                assert col0 == col_i, 'mismatch in %r:\n%s\n%s' % (i, col0, col_i)
In [71]: from sklearn import linear_model
                      clf_l1_LR = linear_model.LogisticRegression(C=1000, penalty='11', tol=0.001).fit(X[0], Y[0])
                      scores=[]
                      scores.append(clf_l1_LR.score(X[0], Y[0]))
                      print 'score for training set', scores[0]
                      for i in range(1,nfiles):
                                scores.append(clf_l1_LR.score(X[i], Y[i]))
                                print 'score for test set', i, scores[i]
score for training set 0.944072051748
score for test set 1 0.94448247091
score for test set 2 0.943976919929
score for test set 3 0.944386639788
score for test set 4 0.944560448937
score for test set 5 0.943735713999
score for test set 6 0.944166904201
score for test set 7 0.943538001825
score for test set 8 0.944438192553
score for test set 9 0.943566597067
score for test set 10 0.944126539894
score for test set 11 0.944858468573
score for test set 12 0.944788959785
score for test set 13 0.944127431039
score for test set 14 0.943777194073
In [72]: print 'mean', np.mean(scores), 'std', np.std(scores)
mean 0.944173502288 std 0.00039856965612
```

Logistic regression with Lasso (L1 penalty) computed over 15 non-overlapping subsets of auth.txt.gz gave me a score with mean 0.9442 and std 0.0004. I believe I am samplying from a normal distribution, which means I have a very narrow gaussian. This in turn means that further sampling will not change my results significantly.

## In []: