# for a given day, get the popularity of a drive model: # cat data_Q3_2018.zip_folder/2018-07-27.csv | sed '1d' | cut -d',' -f3 | sort | uniq -c | sort -g -k1,1 # for every CSV file, get the date # find . | grep csv | while read fullpath; do echo

```
fullpath | sed 's/\//_/g' | sed 's/\.csv//g' | sed 's/zip_folder/_/g' | sed 's/data_//g'; done

# create a file per day containing the popularity of each model
https://stackoverflow.com/questions/17017732/changing-delimiter-of-the-uniq-command
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_fwf.html

date; find . | grep csv | while read fullpath; do
    filename=`echo
```

fullpath | sed 's/zip_folder/_/g' | sed 's/data_//g' | sed 's/\.csv//g' | sed 's/\//_/g' | sed 's/\.//g' | sed 's/^_//g'`; cat `fullpath | sed '1d' | cut -d',' -f3 | sort | uniq -c | sort -g -k1,1 | sed 's/^ *//;s/ /,/' > count_of_models_on_`{filename}.dat; done; date

In [1]:
```
import pandas
print('pandas',pandas.__version__)
import glob
import pickle
import numpy
import seaborn
import time
import datetime
import matplotlib.pyplot as plt
```

```
pandas 0.23.4
```

In [2]:
```
list_of_dat = glob.glob('data_synthesized_from_csvs/count_of_models_per_
day/count_of_models_on_*.dat')
print(len(list_of_dat))
```

```
2092
```

In [3]:

```
list_of_df=[]
start_time = time.time()
for path_to_dat in list_of_dat:
    date_str = path_to_dat[:-len('.dat')].split('_')[-1]
    date_as_dt = datetime.datetime.strptime(date_str, '%Y-%m-%d')
#    print(path_to_dat)
    try:
        df = pandas.read_csv(path_to_dat,header=None)
        df.columns=[date_as_dt,'model']
        df=df.set_index('model')
        list_of_df.append(df)
    except:
        print(path_to_dat)
print('elapsed:',time.time()-start_time,'seconds')
```

```
data_synthesized_from_csvs/count_of_models_per_day/count_of_models_on_2
014__2014_2014-11-02.dat
data_synthesized_from_csvs/count_of_models_per_day/count_of_models_on_Q
1_2017__2017-01-30.dat
data_synthesized_from_csvs/count_of_models_per_day/count_of_models_on_2
015__2015_2015-11-01.dat
```

list_of_models=[] for df in list_of_df: for model_name in df.index: list_of_models.append(model_name) list_of_models = list(set(list_of_models)) print(len(list_of_models))

In [4]:

```
df = pandas.concat(list_of_df,sort=False,axis=1) # join all the datafram
es into a single df
df = df.reindex(sorted(df.columns), axis=1) # order columns by calendar
 date
```

In [5]:

```
df.shape
```

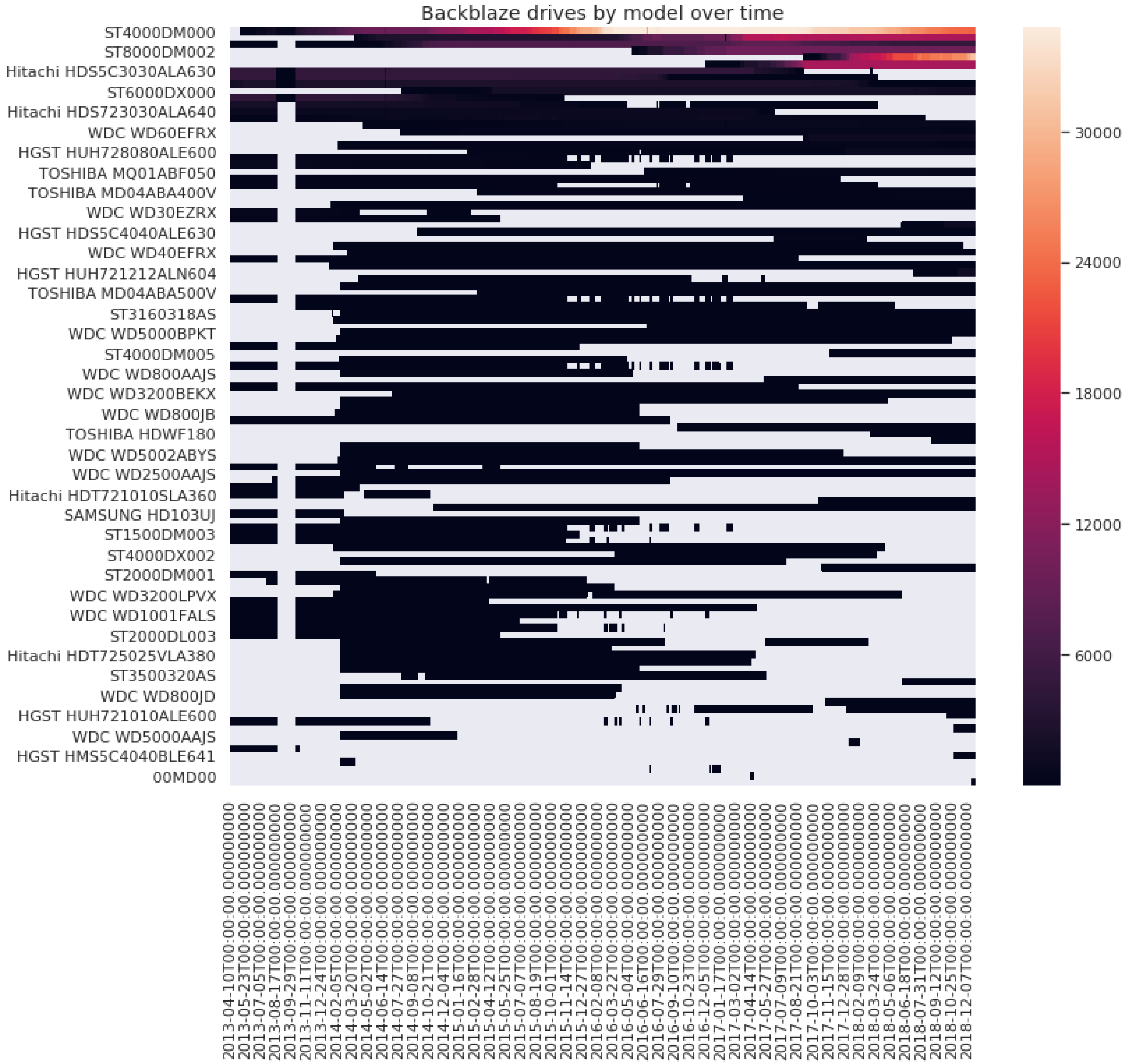
Out[5]:

```
(113, 2089)
```

In [6]:

```
sorted_df = df.loc[df.sum(axis=1).sort_values(ascending=False).index]
```

In [7]:
```
seaborn.set(rc={'figure.figsize':(12,10)})
seaborn.heatmap(sorted_df);
plt.title('Backblaze drives by model over time',fontsize=14);
```



In [11]:
```
len(df.sum(axis=1).sort_values(ascending=False))
```

Out[11]:
```
113
```

In [13]:
```
pandas.options.display.max_rows = 999
```

In [14]:

```
df.sum(axis=1).sort_values(ascending=False)
```

```
Out[14]: ST4000DM000                            45198052.0
         HGST HMS5C4040BLE640                   14872956.0
         HGST HMS5C4040ALE640                   10612497.0
         ST8000DM002                             8198926.0
         ST12000NM0007                           8093190.0
         ST8000NM0055                            7904863.0
         Hitachi HDS5C3030ALA630                 6641559.0
         Hitachi HDS722020ALA330                 5306511.0
         Hitachi HDS5C4040ALE630                 4400563.0
         ST6000DX000                             2517471.0
         ST3000DM001                             2205148.0
         ST31500541AS                            1445217.0
         Hitachi HDS723030ALA640                 1429666.0
         WDC WD30EFRX                            1271769.0
         ST500LM012 HN                            887354.0
         WDC WD60EFRX                             653501.0
         ST10000NM0086                            566937.0
         WDC WD5000LPVX                           451588.0
         HGST HUH728080ALE600                     426811.0
         WDC WD10EADS                             370505.0
         ST31500341AS                             330431.0
         TOSHIBA MQ01ABF050                       303699.0
         ST4000DX000                              293560.0
         ST33000651AS                             222587.0
         TOSHIBA MD04ABA400V                      194619.0
         TOSHIBA MQ01ABF050M                      128920.0
         WDC WD1600AAJS                           126690.0
         WDC WD30EZRX                             123577.0
         ST32000542AS                             119309.0
         TOSHIBA MG07ACA14TA                      108536.0
         HGST HDS5C4040ALE630                      97480.0
         ST4000DM001                               96119.0
         ST9250315AS                               84986.0
         WDC WD40EFRX                              76734.0
         TOSHIBA DT01ACA300                        74177.0
         ST320LT007                                72796.0
         HGST HUH721212ALN604                      71079.0
         WDC WD20EFRX                              67422.0
         ST3160316AS                               64775.0
         TOSHIBA MD04ABA500V                       62640.0
         WDC WD10EACS                              60951.0
         HGST HDS724040ALE640                      58074.0
         ST3160318AS                               49185.0
         ST250LM004 HN                             48456.0
         WDC WD5000LPCX                            47595.0
         WDC WD5000BPKT                            36856.0
         ST9320325AS                               36563.0
         ST1500DL003                               30913.0
         ST4000DM005                               24993.0
         WDC WD800BB                               23656.0
         WDC WD10EADX                              15597.0
         WDC WD800AAJS                             14703.0
         HGST HUS726040ALE610                      14116.0
         Hitachi HDS723030BLE640                   13232.0
         WDC WD3200BEKX                            12656.0
         WDC WD2500BPVT                            11572.0
         WDC WD800AAJB                             11018.0
         WDC WD800JB                               10383.0
         Hitachi HDS723020BLA642                    9620.0
         ST6000DM001                                8990.0
         TOSHIBA HDWF180                            5859.0
         ST500LM030                                 5786.0
         WDC WD1600AAJB                             5757.0
         WDC WD5002ABYS                             5544.0
         WDC WD3200AAJS                             5448.0
         ST320005XXXX                               5032.0
         WDC WD2500AAJS                             4443.0
         ST2000VN000                                4438.0
         WDC WD30EZRS                               4424.0
         Hitachi HDT721010SLA360                    4159.0
         ST8000DM005                                3826.0
         Hitachi HDS724040ALE640                    3724.0
         SAMSUNG HD103UJ                            3710.0
         ST250LT007                                 3593.0
         WDC WD10EARS                               3442.0
         ST1500DM003                                2827.0
         WDC WD10EARX                               2528.0
         WDC WD1600BPVT                             2494.0
         ST4000DX002                                2323.0
         WDC WD5003ABYX                             2254.0
         TOSHIBA HDWE160                            2224.0
         ST2000DM001                                2114.0
         SAMSUNG HD154UI                            1972.0
         WDC WD3200AAJB                             1930.0
         WDC WD3200LPVX                             1808.0
         ST2000DL001                                1447.0
         Hitachi HDS5C3030BLE630                    1415.0
         WDC WD1001FALS                             1366.0
         ST1500DL001                                1314.0
         WDC WD15EARS                               1279.0
         ST2000DL003                                1237.0
         WDC WD3200AAKS                             1188.0
         WDC WD800LB                                1155.0
         Hitachi HDT725025VLA380                    1154.0
         WDC WD2500BEVT                             1140.0
         WDC WD2500AAJB                             1116.0
         ST3500320AS                                 992.0
         ST8000DM004                                 872.0
         WDC WD3200BEKT                              782.0
         WDC WD800JD                                 763.0
         ST6000DM004                                 733.0
         ST1000LM024 HN                              709.0
         HGST HUH721010ALE600                        538.0
         WDC WD10EALS                                532.0
         Seagate BarraCuda SSD ZA500CM10002          490.0
         WDC WD5000AAJS                              327.0
         Samsung SSD 850 EVO 1TB                     234.0
         WDC WD15EADS                                139.0
         HGST HMS5C4040BLE641                         63.0
         WDC WD2500JB                                 39.0
         WDC WD1000FYPS                               20.0
          00MD00                                      14.0
         Seagate BarraCuda SSD ZA2000CM10002          10.0
         dtype: float64
```