EDA_WM-2

November 7, 2017

In [1]:
```
import pymongo
import pandas as pd
import numpy as np

from pymongo import MongoClient
from bson.objectid import ObjectId

import datetime

import matplotlib.pyplot as plt

from collections import defaultdict


%matplotlib inline
import json
plt.style.use('ggplot')

import seaborn as sns
```

In [2]:
```
## Connect to local DB

client = MongoClient('localhost', 27017)
print ("Setup db access")
```

```
Setup db access
```

In [3]:
```
#
# Get collections from mongodb
#
db = client.my_test_db
reponses = db.anon_student_task_responses.find()
```

In [4]:
```
df_responses = pd.DataFrame(list(reponses))
```

In [5]:
```
print (df_responses.head())
```

```
                        _id  bonus  correct      diff          id  incomplete  \
0  59d26fd9d0cd262c1b000001  False     True  0.000000  b6c1c8dLXx       False
1  59d26fd9d0cd262c1b000002  False     True  0.420887  iszM3s-aZG       False
2  59d26fd9d0cd262c1b000003  False     True  0.420887  8UwKe-OymU       False
3  59d26fd9d0cd262c1b000004  False     True  0.782085  FhOINzRAKB       False
4  59d26fd9d0cd262c1b000005   True     True  0.500000  GhmchxnUUV       False

          lesson                                          level_summary  \
0  basic_add_5_1  {'type': 'lesson', 'entered': True, 'time_ente...
1  basic_add_5_1  {'type': 'lesson', 'entered': True, 'time_ente...
2  basic_add_5_1  {'type': 'lesson', 'entered': True, 'time_ente...
3  basic_add_5_1  {'type': 'lesson', 'entered': True, 'time_ente...
4  basic_add_5_1  {'type': 'lesson', 'entered': True, 'time_ente...

                                        problem_set problem_set_id    ...     \
0  lessons/fractions/lesson31_9/part_a/media/prob...     t8suuCs7vN    ...
1  lessons/fractions/lesson31_9/part_a/media/prob...     t8suuCs7vN    ...
2  lessons/fractions/lesson31_9/part_a/media/prob...     t8suuCs7vN    ...
3  lessons/fractions/lesson31_9/part_a/media/prob...     t8suuCs7vN    ...
4  lessons/fractions/lesson31_9/bonus/media/probl...     3jQbmdV9v4    ...

                                    screenshot_url second_try  \
0  http://woot_math_cub.s3.amazonaws.com/ss/12098...      False
1  http://woot_math_cub.s3.amazonaws.com/ss/12098...      False
2  http://woot_math_cub.s3.amazonaws.com/ss/12098...      False
3  http://woot_math_cub.s3.amazonaws.com/ss/12098...      False
4  http://woot_math_cub.s3.amazonaws.com/ss/12098...      False

                             session_id  \
0  720600e1-8969-435b-b16e-bd2c8666f4a7
1  720600e1-8969-435b-b16e-bd2c8666f4a7
2  720600e1-8969-435b-b16e-bd2c8666f4a7
3  720600e1-8969-435b-b16e-bd2c8666f4a7
4  720600e1-8969-435b-b16e-bd2c8666f4a7

                                          student  \
0  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
1  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
2  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
3  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
4  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...

                  sublesson             t time_spent     timestamp  \
0  basic_add_5_1.l31_9_parta  1.506652e+12      24904  1.506652e+12
1  basic_add_5_1.l31_9_parta  1.506652e+12      43222  1.506652e+12
2  basic_add_5_1.l31_9_parta  1.506652e+12      27309  1.506652e+12
3  basic_add_5_1.l31_9_parta  1.506653e+12      30108  1.506653e+12
4       basic_add_5_1.bonus  1.506653e+12      31228  1.506653e+12
```

```
                                                  txt untouched
0  Madelyn ran  1 1/4 miles, stopped, and then ra...     False
1  Juan ran  1 3/6 miles, stopped, and then ran 1...     False
2  Antonio ran  1 1/6 miles, stopped, and then ra...     False
3  Natalie ran  1 3/8 miles, stopped, and then ra...     False
4  The ant traveled 1 1/2  feet, and then  1/6  o...     False

[5 rows x 26 columns]
```

```
In [6]: ## Look act columns
        print (df_responses.columns)
```

```
Index(['_id', 'bonus', 'correct', 'diff', 'id', 'incomplete', 'lesson',
       'level_summary', 'problem_set', 'problem_set_id',
       'problem_set_subspace', 'qual_id', 'randomly_selected', 'response',
       'response_idx', 'retried', 'screenshot_url', 'second_try', 'session_id',
       'student', 'sublesson', 't', 'time_spent', 'timestamp', 'txt',
       'untouched'],
      dtype='object')
```

```
In [7]: ## How many data samples
        print (len(df_responses), "Number of entries")
```

```
52247 Number of entries
```

```
In [8]: ## Example data samle
        print (df_responses.iloc[1])
```

```
_id                                                59d26fd9d0cd262c1b000002
bonus                                                                 False
correct                                                                True
diff                                                               0.420887
id                                                                 iszM3s-aZG
incomplete                                                            False
lesson                                                          basic_add_5_1
level_summary           {'type': 'lesson', 'entered': True, 'time_ente...
problem_set             lessons/fractions/lesson31_9/part_a/media/prob...
problem_set_id                                                    t8suuCs7vN
problem_set_subspace                                              l31_9_parta
qual_id                                 t8suuCs7vN.l31_9_parta.iszM3s-aZG
randomly_selected                                                     False
response                {'numberline_associations': [[{'obj_value': 'A...
response_idx                                                              1
retried                                                                 NaN
screenshot_url          http://woot_math_cub.s3.amazonaws.com/ss/12098...
```

```
second_try                                                        False
session_id                                720600e1-8969-435b-b16e-bd2c8666f4a7
student                  {'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
sublesson                                        basic_add_5_1.l31_9_parta
t                                                             1.50665e+12
time_spent                                                          43222
timestamp                                                     1.50665e+12
txt                      Juan ran  1 3/6 miles, stopped, and then ran 1...
untouched                                                         False
Name: 1, dtype: object
```

```
In [9]: print ("Number of unique lessions", len(df_responses['lesson'].unique()) )
        print ("Unique lessions", df_responses['lesson'].unique())
```

```
Number of unique lessions 221
Unique lessions ['basic_add_5_1' 'add_mixed_1' 'simplify_4' 'simplify_5' 'equiv_1_s
 'review_lesson_1' 'simplify_3' 'simplify_2' 'simplify_0' 'simplify_1'
 'review_lesson_2' 'basic_add_3' 'basic_add_4' 'explore_fract_1_v2'
 'nline_3a' 'nline_3b' 'nline_3c' 'improper_5' 'basic_add_1'
 'model_symbol_3_v3' 'benchmark_2' 'ordering_6' 'benchmark_1'
 'equivalence_5' 'explore_deci_1' 'deci_nline_1' 'benchmark_3' 'ordering_7'
 'ordering_8' 'equivalence_0' 'equivalence_1' 'equiv_1_v2' 'sets_2_v2'
 'sets_2_s1' 'sets_2_s2' 'sets_3_v2' 'sets_3_s1' 'sets_3_s2' 'sets_4'
 'review_lesson_4' 'nline_1b' 'nline_1' 'nline_1c' 'nline_1a' 'nline_2'
 'nline_4' 'mult_whole_frac_review_2' 'mult_whole_frac_review_1'
 'division_2' 'division_3' 'sub_symbol_2' 'mult_whole_mixed_1'
 'add_symbol_3' 'sub_mixed_1' 'sub_model_1' 'basic_ordering_7' 'division_5'
 'division_6' 'nline_0b' 'division_7' 'division_8' 'ordering_2_v3'
 'ordering_2_v3_s1' 'ordering_3' 'adv_ordering_1' 'adv_ordering_2'
 'ordering_5' 'ordering_0' 'equivalence_2' 'advanced_add_1' 'division_1'
 'add_benchmark_1' 'equivalence_6' 'add_uncommon_den_2'
 'add_uncommon_den_3' 'basic_ordering_4' 'ordering_3_1' 'ordering_4'
 'nline_0c' 'improper_4' 'basic_add_2' 'basic_ordering_3'
 'basic_ordering_1' 'ordering_1_v3' 'model_symbol_5' 'name_fract_1_v2'
 'model_symbol_4' 'model_symbol_6' 'num_den_1a' 'model_symbol_1_v2'
 'parts_whole_2' 'model_symbol_3_s1' 'equivalence_3' 'explore_deci_2'
 'advanced_sub_1' 'explore_fract_2' 'name_fract_1_s1' 'sub_mixed_2'
 'basic_sub_5' 'basic_sub_6' 'deci_nline_2' 'deci_nline_3' 'deci_nline_6'
 'deci_nline_4' 'deci_nline_5' 'deci_nline_7' 'equivalence_10'
 'advanced_add_0' 'equivalence_4' 'equivalence_7' 'equivalence_8_v2'
 'equivalence_9' 'sub_symbol_1' 'basic_add_4_1' 'sub_nline_1'
 'mult_whole_frac_5' 'mult_whole_frac_6' 'review_lesson_3' 'model_symbol_2'
 'model_symbol_1_v2_s1' 'explore_deci_5_2' 'explore_deci_6' 'ordering_4_s1'
 'explore_deci_3' 'improper_2' 'label_nline_2' 'improper_1a' 'basic_sub_1'
 'basic_sub_2' 'basic_sub_3' 'name_fract_2_s1' 'mult_whole_frac_7'
 'explore_deci_4' 'model_symbol_4_s1' 'ordering_1_v3_s1' 'explore_deci_10'
 'explore_deci_11' 'explore_deci_12' 'explore_deci_13' 'deci_nline_8'
```

```
 'deci_nline_9' 'deci_nline_10' 'deci_nline_4_s1' 'deci_nline_3_s1'
 'deci_nline_6_s1' 'deci_nline_1_s1' 'deci_nline_2_s1' 'name_fract_2_s2'
 'advanced_sub_0' 'explore_deci_5' 'explore_deci_5_1' 'explore_deci_7'
 'mult_whole_frac_2' 'division_4' 'ordering_6_s1' 'division_10'
 'division_11' 'division_13' 'division_14' 'division_15'
 'add_uncommon_den_1' 'estimating_sums_s1' 'add_uncommon_den_2_s1'
 'sets_1_v2' 'improper_1' 'nline_not_proper_1' 'nline_not_proper_2'
 'model_symbol_2_s1' 'basic_add_5' 'num_den_1a_s1' 'mult_frac_frac_model_1'
 'mult_frac_frac_1' 'mult_frac_frac_2' 'intro_ratio_1' 'intro_ratio_2'
 'intro_ratio_3' 'intro_ratio_4' 'division_17' 'div_mixed_frac_1'
 'unit_rate_1' 'add_estimate_uncommon_den_1' 'sub_model_2' 'sub_symbol_3'
 'sub_mixed_4' 'intro_rat_num_7' 'rat_num_ord_abs_5' 'rat_num_ord_abs_4'
 'rat_num_ord_abs_6' 'rat_num_ord_abs_7' 'rat_num_ord_abs_1'
 'rat_num_ord_abs_2' 'rat_num_ord_abs_3' 'intro_rat_num_4'
 'intro_rat_num_5' 'intro_rat_num_6' 'division_13_0' 'division_13_1'
 'sets_1_s1' 'sets_1_s2' 'explore_deci_8' 'explore_deci_9'
 'add_uncommon_den_5' 'intro_ratio_5' 'add_uncommon_den_4'
 'add_uncommon_den_6' 'mult_whole_frac_2_1' 'mult_whole_frac_4'
 'add_symbol_1' 'add_mixed_2' 'division_9' 'add_symbol_2' 'sub_mixed_3'
 'sub_mixed_5' 'intro_rat_num_1' 'intro_rat_num_2' 'intro_rat_num_3'
 'division_2s1' 'division_16' 'division_13_1s1' 'unit_rate_2' 'unit_rate_3']
```

```
In [10]: print ("Samples of each lesson",df_responses['lesson'].value_counts())

Samples of each lesson review_lesson_1             3399
review_lesson_2              2418
name_fract_1_v2              2247
model_symbol_3_v3            1737
explore_fract_1_v2           1504
nline_1b                     1365
model_symbol_5               1151
nline_2                      1133
ordering_2_v3                1125
ordering_8                   1124
nline_0b                     1011
explore_fract_2               977
basic_ordering_1              867
nline_0c                      759
ordering_1_v3                 750
nline_1                       737
ordering_3                    732
benchmark_2                   704
equivalence_0                 697
nline_1a                      696
parts_whole_2                 671
model_symbol_1_v2             667
ordering_6                    656
```

```
nline_3c                      647
basic_ordering_7              611
ordering_5                    560
nline_1c                      559
benchmark_1                   523
equiv_1_v2                    522
ordering_0                    520
                             ...
review_lesson_4                14
add_uncommon_den_2_s1          14
intro_rat_num_5                14
sub_mixed_5                    14
mult_frac_frac_2               13
mult_frac_frac_model_1         13
mult_whole_frac_2              13
unit_rate_2                    12
mult_frac_frac_1               11
division_10                    11
sets_3_s2                       9
sub_symbol_3                    9
sets_2_s1                       9
division_13_1s1                 8
sub_mixed_4                     8
mult_whole_frac_2_1             7
division_8                      7
intro_rat_num_3                 7
div_mixed_frac_1                7
intro_rat_num_4                 6
division_16                     6
mult_whole_frac_review_2        6
deci_nline_6_s1                 5
sub_mixed_3                     5
division_9                      5
intro_ratio_5                   5
sets_1_s1                       4
sets_1_s2                       4
division_2s1                    4
deci_nline_3_s1                 4
Name: lesson, Length: 221, dtype: int64
```

```
In [11]: print ("Summary sample :", df_responses['level_summary'][0])
```

```
Summary sample : {'type': 'lesson', 'entered': True, 'time_entered': 1506652363265.
```

```
In [12]: ## Promote student info, level summary, level summary problem results
```

```
In [13]: df2 = df_responses.join(pd.DataFrame(df_responses["student"].to_dict()).T)
```

```
In [14]: df2 = df2.join(pd.DataFrame(df2['level_summary'].to_dict()).T)

In [15]: df2 = df2.join(pd.DataFrame(df2['problems'].to_dict()).T)

In [16]: df_student1 = df2.groupby('student_id').agg({ 'lesson':[len,  pd.Series.nu

In [17]: df_student1['percent_correct'] = df_student1['nright']['sum'].astype(float

In [18]: df_student1
```

Out[18]:

	lesson		ntotal	nright	percent_correct
	len	nunique	sum	sum	
student_id					
007D5A2F1	18	2	136	108	0.794118
01Q4G6G2B1	16	1	110	105	0.954545
01Q9N3V2Y1	100	16	523	480	0.917782
01S4P4N2D1	10	2	82	82	1.000000
0205Q2E1	58	4	482	386	0.800830
02B1W2I4S1	35	6	217	217	1.000000
02X8I4D2T1	31	5	151	121	0.801325
032P4C2V1	80	12	383	348	0.908616
03D1R2O4A1	20	3	172	162	0.941860
03S0J2R4J1	16	4	70	57	0.814286
03X1L8B2A1	89	7	745	617	0.828188
03Z0W8Z2D1	157	5	1096	670	0.611314
04R0N2Z4A1	22	3	160	147	0.918750
04Y3F1K4J1	15	2	105	52	0.495238
051X8T2F1	65	7	314	257	0.818471
05W3M0O4F1	88	9	665	476	0.715789
06A6H1Z4N1	83	4	922	637	0.690889
07D4J8U2N1	119	20	643	559	0.869362
07D9Q3V3A1	27	3	199	176	0.884422
07H6J5P2A1	24	4	120	110	0.916667
07J6W1Z4R1	24	3	176	146	0.829545
0823B2G1	101	18	497	492	0.989940
08C0U8Z2T1	15	1	82	51	0.621951
08H1X4R3S1	15	3	75	75	1.000000
099G6L2T1	21	3	149	126	0.845638
0A2N7G5O2T1	116	10	978	734	0.750511
0A5H7R7D2S1	143	15	781	660	0.845070
0B1W2D0C4U1	39	7	195	185	0.948718
0B3Z4Q6F2F1	27	2	293	184	0.627986
0B5E6B5K2O1	88	6	1082	316	0.292052
...	...	...	...	...	...
Y4D0J4R6A2H1	37	1	287	205	0.714286
Y5K4O0F8Z2A1	73	11	407	347	0.852580
Y5Q0B4G4B2D1	19	2	171	152	0.888889
Y6A4Y1C5B2C1	4	1	16	16	1.000000
Y6V4Z6N3J2J1	11	2	61	61	1.000000

```
Y6Z5T9B7O2E1     16       2    112     87        0.776786
Y7B8I8Z3R2B1     36       6    182    173        0.950549
Y7M1U1U7A2B1     27       5    141    129        0.914894
Y8P8X0Y2Q4H1     10       1     90     70        0.777778
Y9T0W4O4T2D1     20       3    136    130        0.955882
Y9X0L8Y4Y2L1    164      11   1280    982        0.767188
Z0G8G7M2Q2Z1      9       1     72     63        0.875000
Z2S9K7S8A3B1     10       2     50     50        1.000000
Z3E3I1N2G4Z1     86       7    904    532        0.588496
Z3H6S2F3M2B1    112      16    642    548        0.853583
Z3M9R3L8D2W1     37       5    207    166        0.801932
Z3Q0O9F3D2G1     50       4    406    364        0.896552
Z3S2E8M4D2X1     17       2    121    104        0.859504
Z4B5P9H7S2V1     18       2    162    162        1.000000
Z4G0K6F4S2U1     38       7    196    184        0.938776
Z4I9P0Y8M2D1     35       3    270    220        0.814815
Z4U5I0O8W2A1    171      24    893    791        0.885778
Z5F3Z5J4S2H1     87       9    725    665        0.917241
Z5N6I4E8R2I1     19       3    131    131        1.000000
Z5R5E6R1K4B1     34       3    274    238        0.868613
Z6J7J6M8O3O1      2       1      4      0        0.000000
Z8H0M1Q2F4C1     24       3    188    187        0.994681
Z8N2Y6S5Z2O1     45       6    348    319        0.916667
Z8T8L4T4A2L1      9       1     81     72        0.888889
Z9N7O6C4D2D1     78       9    659    614        0.931715

[1156 rows x 5 columns]
```

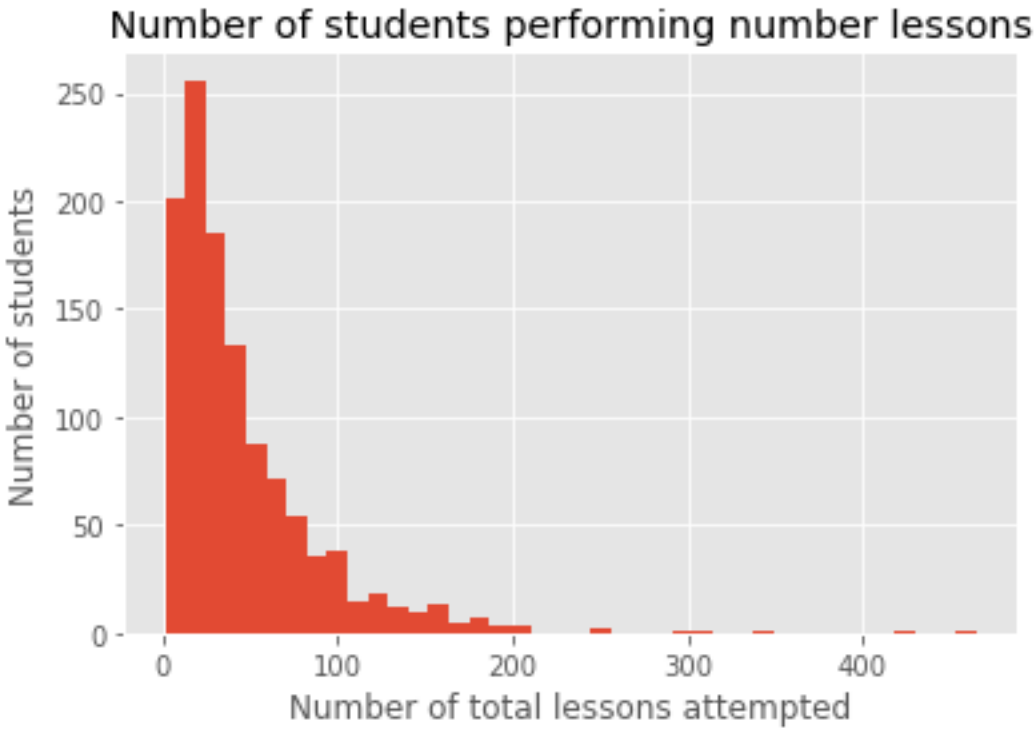
In [19]:
```
y1 = np.array(df_student1['lesson']['len'])
```

In [ ]:

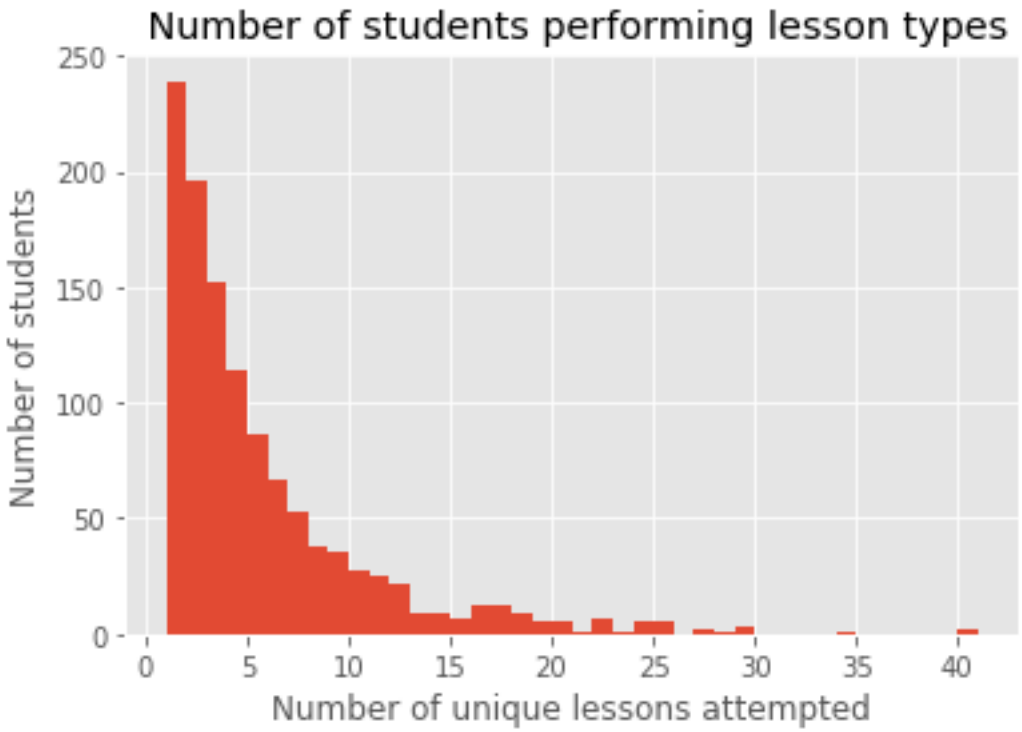
In [20]:
```
# Total Lessons per student

plt.title(' Number of students performing number lessons ')
plt.xlabel('Number of total lessons attempted')
plt.ylabel('Number of students')
plt.hist(y1, bins=40)
```

Out[20]:
```
(array([ 201.,  256.,  185.,  133.,   88.,   71.,   54.,   36.,   38.,
          15.,   18.,   12.,   10.,   13.,    5.,    7.,    3.,    4.,
           0.,    0.,    0.,    2.,    0.,    0.,    0.,    1.,    1.,
           0.,    0.,    1.,    0.,    0.,    0.,    0.,    0.,    0.,
           1.,    0.,    0.,    1.]),
 array([   1. ,   12.6,   24.2,   35.8,   47.4,   59. ,   70.6,   82.2,
          93.8,  105.4,  117. ,  128.6,  140.2,  151.8,  163.4,  175. ,
         186.6,  198.2,  209.8,  221.4,  233. ,  244.6,  256.2,  267.8,
         279.4,  291. ,  302.6,  314.2,  325.8,  337.4,  349. ,  360.6,
         372.2,  383.8,  395.4,  407. ,  418.6,  430.2,  441.8,  453.4,
```

```
        465. ]),
 <a list of 40 Patch objects>)
```



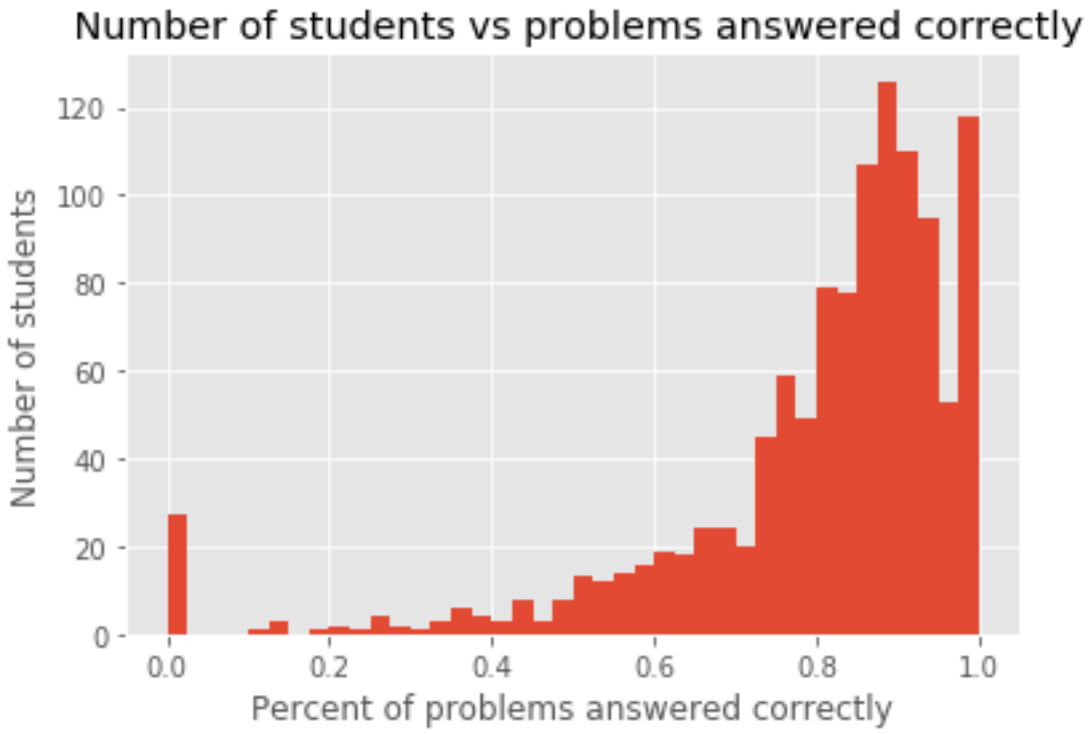

```
In [21]: # Unique students per # of unique lessons
         y2 = np.array(df_student1['lesson']['nunique'])
         plt.title(' Number of students performing lesson types')
         plt.xlabel('Number of unique lessons attempted')
         plt.ylabel('Number of students')
         plt.hist(y2, bins=40)

Out[21]: (array([ 239.,  196.,  152.,  114.,   86.,   67.,   53.,   38.,   36.,
                   28.,   25.,   22.,    9.,    9.,    7.,   12.,   13.,    9.,
                    6.,    5.,    1.,    7.,    1.,    6.,    6.,    0.,    2.,
                    1.,    3.,    0.,    0.,    0.,    0.,    1.,    0.,    0.,
                    0.,    0.,    0.,    2.]),
          array([  1.,   2.,   3.,   4.,   5.,   6.,   7.,   8.,   9.,  10.,  11.,
                  12.,  13.,  14.,  15.,  16.,  17.,  18.,  19.,  20.,  21.,  22.,
                  23.,  24.,  25.,  26.,  27.,  28.,  29.,  30.,  31.,  32.,  33.,
                  34.,  35.,  36.,  37.,  38.,  39.,  40.,  41.]),
          <a list of 40 Patch objects>)
```



In [ ]:

In [22]:
```
# Unique students per # of unique lessons
y3 = np.array(df_student1['percent_correct'])
plt.title(' Number of students vs problems answered correctly')
plt.xlabel('Percent of problems answered correctly')
plt.ylabel('Number of students')
plt.hist(y3, bins = 40)
```

Out[22]:
```
(array([  27.,    0.,    0.,    0.,    1.,    3.,    0.,    1.,    2.,
           1.,    4.,    2.,    1.,    3.,    6.,    4.,    3.,    8.,
           3.,    8.,   13.,   12.,   14.,   16.,   19.,   18.,   24.,
          24.,   20.,   45.,   59.,   49.,   79.,   78.,  107.,  126.,
         110.,   95.,   53.,  118.]),
 array([ 0.   ,  0.025,  0.05 ,  0.075,  0.1  ,  0.125,  0.15 ,  0.175,
         0.2  ,  0.225,  0.25 ,  0.275,  0.3  ,  0.325,  0.35 ,  0.375,
         0.4  ,  0.425,  0.45 ,  0.475,  0.5  ,  0.525,  0.55 ,  0.575,
         0.6  ,  0.625,  0.65 ,  0.675,  0.7  ,  0.725,  0.75 ,  0.775,
         0.8  ,  0.825,  0.85 ,  0.875,  0.9  ,  0.925,  0.95 ,  0.975,  1
 <a list of 40 Patch objects>)
```




```
In [24]: df2.columns

Out[24]: Index(['_id', 'bonus', 'correct', 'diff', 'id', 'incomplete', 'lesson',
                'level_summary', 'problem_set', 'problem_set_id',
                'problem_set_subspace', 'qual_id', 'randomly_selected', 'response',
                'response_idx', 'retried', 'screenshot_url', 'second_try', 'session
                'student', 'sublesson', 't', 'time_spent', 'timestamp', 'txt',
                'untouched', 'grade', 'school_id', 'section_id', 'student_id',
                'blank_slate_mastery', 'description', 'entered', 'lesson_type',
                'lm_stats', 'mastery', 'name', 'path', 'problems', 'stars', 'subjec
                't_elapsed', 'time_entered', 'time_exited', 'title', 'type',
                'unit_name', 'unit_rank', 'nretry_right', 'nretry_wrong', 'nright',
                'ntotal', 'nuntouched', 'nwrong'],
               dtype='object')

In [25]: df2['subject'].unique()

Out[25]: array(['fractions', 'review', 'decimals', 'ratios', 'rational_numbers'], d

In [26]: df2.iloc[0]

Out[26]: _id                                                  59d26fd9d0cd262c1b000001
         bonus                                                               False
         correct                                                              True
```

diff	0
id	b6c1c8dLXx
incomplete	False
lesson	basic_add_5_1
level_summary	{'type': 'lesson', 'entered': True, 'time_ente...
problem_set	lessons/fractions/lesson31_9/part_a/media/prob...
problem_set_id	t8suuCs7vN
problem_set_subspace	l31_9_parta
qual_id	t8suuCs7vN.l31_9_parta.b6c1c8dLXx
randomly_selected	False
response	{'numberline_associations': [[{'obj_value': 'A...
response_idx	0
retried	NaN
screenshot_url	http://woot_math_cub.s3.amazonaws.com/ss/12098...
second_try	False
session_id	720600e1-8969-435b-b16e-bd2c8666f4a7
student	{'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
sublesson	basic_add_5_1.l31_9_parta
t	1.50665e+12
time_spent	24904
timestamp	1.50665e+12
txt	Madelyn ran 1 1/4 miles, stopped, and then ra...
untouched	False
grade	3
school_id	I7N2N0K9
section_id	U3J3J1E8
student_id	F8W8W9U0R2U1
blank_slate_mastery	{'mean': 0.6558873935892596, 'std_dev': 0.3816...
description	In this lesson, students add fractions and mix...
entered	True
lesson_type	l
lm_stats	{'is_new': True, 'created': 1503772100340.0, '...
mastery	{'mean': 0.6558873935892596, 'std_dev': 0.3816...
name	basic_add_5_1
path	basic_add_5_1
problems	{'nwrong': 0, 'ntotal': 5, 'nright': 5, 'nretr...
stars	{'earned': 3, 'old': 0}
subject	fractions
t_elapsed	187518
time_entered	1.50665e+12
time_exited	1.50665e+12
title	Adding Mixed Numbers and Fractions Using Numbe...
type	lesson
unit_name	frac_add_sub_1
unit_rank	4
nretry_right	0
nretry_wrong	0
nright	5

```
ntotal                                                                    5
nuntouched                                                                0
nwrong                                                                    0
Name: 0, dtype: object
```

In [27]:
```
df2.iloc[0]['response']
```

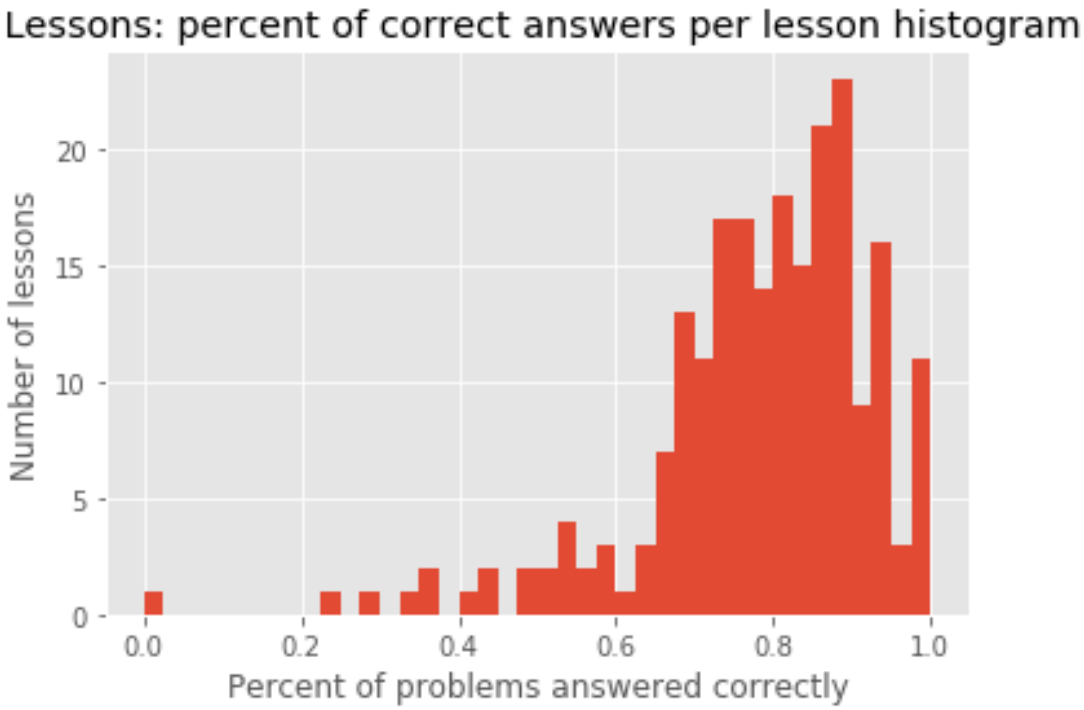
Out[27]:
```
{'fraction_cblock_chains': [{'lcm_sum': {'__as3_type': 'Fraction',
    'denominator': 4,
    'numerator': 3},
   'left': 509,
   'pieces': ['1/4', '1/4', '1/4'],
   'right': 865,
   'sum': {'__as3_type': 'Fraction', 'denominator': 4, 'numerator': 3},
   'top': 355}],
 'fraction_cblock_containment': {},
 'fraction_cblock_counts': {'1/4': 3},
 'fraction_cblock_total_count': 3,
 'input': '8',
 'numberline_associations': [[{'obj_name': 'a_text',
    'obj_value': 'A',
    'pos_value': 1.752535558428964,
    'position': 676.9330855018587}]]}
```

In [28]:
```
df_lesson1 = df2.groupby('lesson').agg({ 'student_id':[len,  pd.Series.nur
```

In [29]:
```
df_lesson1['percent_correct'] = df_lesson1['nright']['sum'].astype(float)
```

In [30]:
```
# Lessons and answers
y3 = np.array(df_lesson1['percent_correct'])
plt.title(' Lessons: percent of correct answers per lesson histogram')
plt.xlabel('Percent of problems answered correctly')
plt.ylabel('Number of lessons')
plt.hist(y3, bins = 40)
```

Out[30]:
```
(array([  1.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   1.,   0.,
          1.,   0.,   1.,   2.,   0.,   1.,   2.,   0.,   2.,   2.,   4.,
          2.,   3.,   1.,   3.,   7.,  13.,  11.,  17.,  17.,  14.,  18.,
         15.,  21.,  23.,   9.,  16.,   3.,  11.]),
 array([ 0.   ,  0.025,  0.05 ,  0.075,  0.1  ,  0.125,  0.15 ,  0.175,
         0.2  ,  0.225,  0.25 ,  0.275,  0.3  ,  0.325,  0.35 ,  0.375,
         0.4  ,  0.425,  0.45 ,  0.475,  0.5  ,  0.525,  0.55 ,  0.575,
         0.6  ,  0.625,  0.65 ,  0.675,  0.7  ,  0.725,  0.75 ,  0.775,
         0.8  ,  0.825,  0.85 ,  0.875,  0.9  ,  0.925,  0.95 ,  0.975,  1
 <a list of 40 Patch objects>)
```



In [34]: df_lesson1

Out[34]:

lesson	student_id len	student_id nunique	ntotal sum	nright sum	percent_corre
add_benchmark_1	272	24	2233	1211	0.5423
add_estimate_uncommon_den_1	120	17	699	635	0.9084
add_mixed_1	54	5	282	194	0.6879
add_mixed_2	60	8	259	224	0.8648
add_symbol_1	71	11	373	329	0.8820
add_symbol_2	63	9	305	258	0.8459
add_symbol_3	65	9	307	224	0.7296
add_uncommon_den_1	146	18	1082	834	0.7707
add_uncommon_den_2	208	24	1419	1037	0.7307
add_uncommon_den_2_s1	14	3	52	48	0.9230
add_uncommon_den_3	153	19	1086	773	0.7117
add_uncommon_den_4	85	14	465	423	0.9096
add_uncommon_den_5	79	10	373	339	0.9088
add_uncommon_den_6	47	9	231	224	0.9696
adv_ordering_1	418	61	2767	2577	0.9313
adv_ordering_2	431	57	3065	2322	0.7575
advanced_add_0	117	18	826	753	0.9116
advanced_add_1	187	20	1204	857	0.7117
advanced_sub_0	215	22	1278	1032	0.8075

advanced_sub_1	130	17	706	512	0.7252
basic_add_1	261	28	1871	1088	0.5815
basic_add_2	95	14	575	517	0.8991
basic_add_3	131	11	702	465	0.6623
basic_add_4	48	7	197	169	0.8578
basic_add_4_1	45	7	229	172	0.7510
basic_add_5	28	4	156	148	0.9487
basic_add_5_1	20	4	100	100	1.0000
basic_ordering_1	867	121	5687	5055	0.8888
basic_ordering_3	327	58	1638	1509	0.9212
basic_ordering_4	330	42	2301	2058	0.8943
...	...	...	...	...	.
sets_1_s1	4	1	16	16	1.0000
sets_1_s2	4	1	16	16	1.0000
sets_1_v2	141	23	876	764	0.8721
sets_2_s1	9	2	36	31	0.8611
sets_2_s2	20	2	160	129	0.8062
sets_2_v2	202	25	1090	898	0.8238
sets_3_s1	16	2	64	42	0.6562
sets_3_s2	9	2	36	31	0.8611
sets_3_v2	167	21	1173	950	0.8098
sets_4	124	19	655	586	0.8946
simplify_0	336	39	2701	1958	0.7249
simplify_1	243	22	1884	1640	0.8704
simplify_2	239	20	1691	1145	0.6771
simplify_3	188	17	1020	604	0.5921
simplify_4	290	27	1413	988	0.6992
simplify_5	236	24	1772	1531	0.8639
sub_mixed_1	58	6	494	343	0.6943
sub_mixed_2	33	4	219	191	0.8721
sub_mixed_3	5	1	25	25	1.0000
sub_mixed_4	8	1	56	0	0.0000
sub_mixed_5	14	1	64	49	0.7656
sub_model_1	77	8	499	333	0.6673
sub_model_2	22	4	104	86	0.8269
sub_nline_1	66	8	556	297	0.5341
sub_symbol_1	35	5	301	99	0.3289
sub_symbol_2	24	3	136	97	0.7132
sub_symbol_3	9	2	36	18	0.5000
unit_rate_1	109	9	1060	386	0.3641
unit_rate_2	12	1	132	108	0.8181
unit_rate_3	15	1	210	60	0.2857

[221 rows x 5 columns]

In [36]:
```
df_lesson1.sort_values('percent_correct')
```

Out[36]:

	student_id		ntotal	nright	percent_correct
	len	nunique	sum	sum	

lesson					
sub_mixed_4	8	1	56	0	0.000000
mult_whole_frac_review_1	17	3	49	12	0.244898
unit_rate_3	15	1	210	60	0.285714
sub_symbol_1	35	5	301	99	0.328904
mult_whole_frac_6	107	6	874	310	0.354691
unit_rate_1	109	9	1060	386	0.364151
deci_nline_4_s1	18	2	99	40	0.404040
division_15	69	6	478	208	0.435146
intro_ratio_1	348	17	4085	1791	0.438433
intro_rat_num_3	7	2	25	12	0.480000
division_6	31	3	149	74	0.496644
sub_symbol_3	9	2	36	18	0.500000
mult_whole_frac_5	28	5	126	65	0.515873
name_fract_2_s1	241	19	2463	1303	0.529030
sub_nline_1	66	8	556	297	0.534173
explore_deci_5_2	33	4	189	102	0.539683
add_benchmark_1	272	24	2233	1211	0.542320
division_14	125	9	1165	656	0.563090
ordering_6	656	60	5113	2881	0.563466
basic_add_1	261	28	1871	1088	0.581507
simplify_3	188	17	1020	604	0.592157
intro_rat_num_2	53	4	607	361	0.594728
rat_num_ord_abs_4	40	5	312	188	0.602564
ordering_2_v3_s1	148	17	743	473	0.636608
explore_deci_12	44	5	257	164	0.638132
equivalence_6	491	53	3085	2001	0.648622
sets_3_s1	16	2	64	42	0.656250
ordering_7	380	39	2647	1750	0.661126
basic_add_3	131	11	702	465	0.662393
sub_model_1	77	8	499	333	0.667335
...	...	...	...	...	...
div_mixed_frac_1	7	1	27	25	0.925926
division_8	7	2	27	25	0.925926
deci_nline_1_s1	17	4	68	63	0.926471
equivalence_4	263	44	1471	1368	0.929980
model_symbol_4	377	52	2556	2378	0.930360
adv_ordering_1	418	61	2767	2577	0.931334
rat_num_ord_abs_2	31	5	191	179	0.937173
model_symbol_2_s1	16	4	64	60	0.937500
equivalence_2	366	57	2297	2157	0.939051
intro_rat_num_7	19	2	171	161	0.941520
explore_deci_13	21	4	105	99	0.942857
explore_deci_9	21	4	105	99	0.942857
improper_4	84	12	456	430	0.942982
ordering_3_1	192	36	946	896	0.947146
model_symbol_3_s1	96	23	381	361	0.947507
basic_add_5	28	4	156	148	0.948718

```
nline_1c                    559     83   3945   3774    0.956654
rat_num_ord_abs_5            30      6    150    145    0.966667
add_uncommon_den_6           47      9    231    224    0.969697
mult_frac_frac_1             11      1     83     81    0.975904
rat_num_ord_abs_6            34      8    130    128    0.984615
model_symbol_4_s1            16      4     64     64    1.000000
basic_add_5_1                20      4    100    100    1.000000
sub_mixed_3                   5      1     25     25    1.000000
division_9                    5      1     25     25    1.000000
division_13_1s1               8      2     32     32    1.000000
sets_1_s2                     4      1     16     16    1.000000
sets_1_s1                     4      1     16     16    1.000000
deci_nline_3_s1               4      1     16     16    1.000000
division_2s1                  4      1     16     16    1.000000

[221 rows x 5 columns]
```

```
In [ ]:
```

```
In [37]: df3  = df2.copy()
```

```
In [38]: df3['percent_correct'] = df3['nright'].astype(float) / df3['ntotal']
```

```
In [ ]: ## Make 'description' a feature wih important words mapped
```

```
In [47]: df3.columns
```

```
Out[47]: Index(['_id', 'bonus', 'correct', 'diff', 'id', 'incomplete', 'lesson',
                'level_summary', 'problem_set', 'problem_set_id',
                'problem_set_subspace', 'qual_id', 'randomly_selected', 'response',
                'response_idx', 'retried', 'screenshot_url', 'second_try', 'session
                'student', 'sublesson', 't', 'time_spent', 'timestamp', 'txt',
                'untouched', 'grade', 'school_id', 'section_id', 'student_id',
                'blank_slate_mastery', 'description', 'entered', 'lesson_type',
                'lm_stats', 'mastery', 'name', 'path', 'problems', 'stars', 'subjec
                't_elapsed', 'time_entered', 'time_exited', 'title', 'type',
                'unit_name', 'unit_rank', 'nretry_right', 'nretry_wrong', 'nright',
                'ntotal', 'nuntouched', 'nwrong', 'percent_correct'],
               dtype='object')
```

```
In [50]: df3.iloc[0]
```

```
Out[50]: _id                                  59d26fd9d0cd262c1b000001
         bonus                                                  False
         correct                                                 True
         diff                                                       0
         id                                                  b6c1c8dLXx
         incomplete                                             False
         lesson                                           basic_add_5_1
```

level_summary	{'type': 'lesson', 'entered': True, 'time_ente...
problem_set	lessons/fractions/lesson31_9/part_a/media/prob...
problem_set_id	t8suuCs7vN
problem_set_subspace	l31_9_parta
qual_id	t8suuCs7vN.l31_9_parta.b6c1c8dLXx
randomly_selected	False
response	{'numberline_associations': [[{'obj_value': 'A...
response_idx	0
retried	NaN
screenshot_url	http://woot_math_cub.s3.amazonaws.com/ss/12098...
second_try	False
session_id	720600e1-8969-435b-b16e-bd2c8666f4a7
student	{'section_id': 'U3J3J1E8', 'school_id': 'I7N2N...
sublesson	basic_add_5_1.l31_9_parta
t	1.50665e+12
time_spent	24904
timestamp	1.50665e+12
txt	Madelyn ran 1 1/4 miles, stopped, and then ra...
untouched	False
grade	3
school_id	I7N2N0K9
section_id	U3J3J1E8
student_id	F8W8W9U0R2U1
blank_slate_mastery	{'mean': 0.6558873935892596, 'std_dev': 0.3816...
description	In this lesson, students add fractions and mix...
entered	True
lesson_type	l
lm_stats	{'is_new': True, 'created': 1503772100340.0, '...
mastery	{'mean': 0.6558873935892596, 'std_dev': 0.3816...
name	basic_add_5_1
path	basic_add_5_1
problems	{'nwrong': 0, 'ntotal': 5, 'nright': 5, 'nretr...
stars	{'earned': 3, 'old': 0}
subject	fractions
t_elapsed	187518
time_entered	1.50665e+12
time_exited	1.50665e+12
title	Adding Mixed Numbers and Fractions Using Numbe...
type	lesson
unit_name	frac_add_sub_1
unit_rank	4
nretry_right	0
nretry_wrong	0
nright	5
ntotal	5
nuntouched	0
nwrong	0
percent_correct	1

```
        Name: 0, dtype: object

In [51]: df3.iloc[0]['txt']

Out[51]: 'Madelyn ran  1 1/4 miles, stopped, and then ran 2/4 of a mile.\nDrag A to

In [52]: df3.iloc[0]['description']

Out[52]: 'In this lesson, students add fractions and mixed numbers using number lin

In [69]: for idx in range(100):
             print (df3.iloc[idx]['lesson'])
             print (df3.iloc[idx]['response'])

basic_add_5_1
{'numberline_associations': [[{'obj_value': 'A', 'obj_name': 'a_text', 'pos_value':
basic_add_5_1
{'numberline_associations': [[{'obj_value': 'A', 'obj_name': 'a_text', 'pos_value':
basic_add_5_1
{'numberline_associations': [[{'obj_value': 'A', 'obj_name': 'a_text', 'pos_value':
basic_add_5_1
{'numberline_associations': [[{'obj_value': 'A', 'obj_name': 'a_text', 'pos_value':
basic_add_5_1
{'bitmap_text_inputs': {'input': ['4']}, 'bitmap_text_interp': {'input': '1 1/2  +
add_mixed_1
{'num': '6', 'den': '8', 'whole': '3', 'fraction_input_value': '3 6/8'}
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_4
```

{'plain_image_groups': [{'url': 'assets/cms/wootmath_fractions/misc_objects/ant_alt
simplify_5
{'bitmap_text_interp': {'input_a': '1/2'}, 'fraction_input_value': '[Fraction] 2/4'
simplify_5
{'bitmap_text_interp': {'input_a': '1/3'}, 'fraction_input_value': '[Fraction] 2/6'
simplify_5
{'bitmap_text_interp': {'input_a': '1/4'}, 'fraction_input_value': '[Fraction] 2/8'
simplify_5
{'bitmap_text_interp': {'input_a': '1/3'}, 'fraction_input_value': '[Fraction] 3/9'
simplify_5
{'bitmap_text_interp': {'input_a': '1/2'}, 'fraction_input_value': '[Fraction] 4/8'
simplify_5
{'bitmap_text_interp': {'input_a': '2/5'}, 'fraction_input_value': '[Fraction] 4/10
simplify_5
{'bitmap_text_interp': {'input_a': '2/3'}, 'fraction_input_value': '[Fraction] 4/6'
simplify_5
{'bitmap_text_interp': {'input_a': '1/2'}, 'fraction_input_value': '[Fraction] 2/4'
simplify_5
{'fraction_circle_groups': [{'scale': 0.9, 'pieces': ['1/2', '1'], 'x': 400, 'y': 3
simplify_5
{'fraction_circle_groups': [{'scale': 0.9000000000000001, 'pieces': ['1/4'], 'x': 8
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '6/6'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '1/2'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/3', 'bitmap_text_interp': {'frac_1': '3/9'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/3', 'bitmap_text_interp': {'frac_1': '2/6'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '4/4'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '4/9'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '3/6'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/4', 'bitmap_text_interp': {'frac_1': '2/9'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/4', 'bitmap_text_interp': {'frac_1': '2/8'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '5/10'}
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '2/4'},
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_1': '6/12'}
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/2', 'bitmap_text_interp': {'frac_2': '5/10',
equiv_1_s1

{'fraction_input_value': '[Fraction] 1/3', 'bitmap_text_interp': {'frac_2': '3/9',
equiv_1_s1
{'fraction_input_value': '[Fraction] 1/4', 'bitmap_text_interp': {'frac_2': '3/12',
equiv_1_s1
{'fraction_input_value': '[Fraction] 3/4', 'bitmap_text_interp': {'frac_2': '9/3',
equiv_1_s1
{'fraction_input_value': '[Fraction] 3/4', 'bitmap_text_interp': {'frac_2': '9/12',
review_lesson_1
{'fraction_circle_containment': {}, 'fraction_circle_groups': [{'scale': 0.9999999
review_lesson_1
{'fraction_circle_containment': {}, 'fraction_circle_groups': [{'scale': 0.9999999
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/2': {'pieces': ['1/4', '1/4'], 'lcm_
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/2': {'pieces': ['1/4', '1/4'], 'lcm_
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/2': {'pieces': ['1/2', '1/4', '1/4']
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/6': {'pieces': ['1/6'], 'lcm_sum': {
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/2': 2, '1/4': 2}, 'fraction_cblock_containme
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/2': 2, '1/4': 3}, 'fraction_cblock_containme
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/6': 9, '1/3': 3}, 'fraction_cblock_containme
review_lesson_1
{'fraction_circle_containment': {'piece_0': {'pieces': ['1/4', '1/4'], 'lcm_sum': {
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/2': {'pieces': ['1/6', '1/6', '1/6']
review_lesson_1
{'fraction_circle_containment': {'[Fraction] 1/2': {'pieces': ['1/6', '1/6', '1/6']
review_lesson_1
{'input_B': '', 'bitmap_text_inputs': {'frac_1': ['', '']}, 'bitmap_text_interp': {
review_lesson_1
{'input_B': '', 'bitmap_text_inputs': {'frac_1': ['', '']}, 'bitmap_text_interp': {
review_lesson_1
{'input_B': '12', 'bitmap_text_inputs': {'frac_1': ['6', '12']}, 'bitmap_text_inter
review_lesson_1
None
review_lesson_1
{'fraction_circle_groups': [{'scale': 0.9999999999999999, 'pieces': ['1/2', '1/4',
review_lesson_1
{'input_B': '6', 'bitmap_text_inputs': {'frac_1': ['2', '6']}, 'bitmap_text_interp'
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/4': 4}, 'bitmap_text_inputs': {'ec
review_lesson_1
{'input_B': '6', 'bitmap_text_inputs': {'frac_2': ['5', '10'], 'frac_1': ['3', '6']
review_lesson_1

{'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/8': 1}, 'fraction_cblock_total_cou
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/8': 1}, 'fraction_cblock_total_cou
review_lesson_1
{'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/4': 1}, 'fraction_cblock_total_cou
review_lesson_1
{'fraction_cblock_chains': [{'pieces': ['1/4'], 'lcm_sum': {'denominator': 4, 'nume
review_lesson_1
{'fraction_cblock_chains': [{'pieces': ['1/6', '1/6'], 'lcm_sum': {'denominator': 6
review_lesson_1
{'fraction_cblock_chains': [{'pieces': ['1/6'], 'lcm_sum': {'denominator': 6, 'nume
simplify_5
{'bitmap_text_inputs': {'input_a': ['4']}, 'bitmap_text_interp': {'input_a': '4/2'}
simplify_5
{'bitmap_text_inputs': {'input_a': ['2']}, 'bitmap_text_interp': {'input_a': '2/2'}
simplify_5
{'bitmap_text_inputs': {'input_a': ['1']}, 'bitmap_text_interp': {'input_a': '1/3'}
simplify_5
{'bitmap_text_inputs': {'input_a': ['']}, 'bitmap_text_interp': {'input_a': '?/2'},
simplify_3
{'whole': '', 'den': '8', 'num': '4', 'fraction_input_value': '4/8'}
simplify_3
{'whole': '', 'den': '2', 'num': '2', 'fraction_input_value': '2/2'}
simplify_3
{'whole': '', 'den': '3', 'num': '3', 'fraction_input_value': '3/3'}
simplify_3
{'whole': '', 'den': '3', 'num': '3', 'fraction_input_value': '3/3'}
simplify_3
{'whole': '', 'den': '32', 'num': '3', 'fraction_input_value': '3/32'}
simplify_3
{'whole': '', 'den': '32', 'num': '3', 'fraction_input_value': '3/32'}
simplify_3
{'whole': '', 'den': '4', 'num': '3', 'fraction_input_value': '3/4'}
simplify_3
{'whole': '', 'den': '4', 'num': '3', 'fraction_input_value': '3/4'}
simplify_3
{'whole': '', 'den': '3', 'num': '2', 'fraction_input_value': '2/3'}
simplify_2
{'fraction_cblock_total_count': 5, 'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/
simplify_2
{'fraction_cblock_total_count': 5, 'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/
simplify_2
{'fraction_cblock_total_count': 6, 'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/
simplify_2
{'fraction_cblock_total_count': 5, 'fraction_cblock_counts': {'1': 2, '1/6': 2, '1/
simplify_2
{'fraction_cblock_total_count': 7, 'fraction_cblock_counts': {'1': 2, '1/2': 1, '1/
simplify_2

```
{'mult_d': '2', 'mult_n': '2'}
simplify_2
{'mult_d': '3', 'mult_n': '3'}
simplify_2
{'mult_d': '4', 'mult_n': '4'}
simplify_2
{'mult_d': '2', 'mult_n': '2'}
simplify_2
None
simplify_2
{'fraction_cblock_total_count': 8, 'fraction_cblock_counts': {'1': 2, '1/6': 4, '1/
simplify_2
{'fraction_cblock_total_count': 8, 'fraction_cblock_counts': {'1': 2, '1/6': 4, '1/
simplify_2
{'fraction_cblock_total_count': 6, 'fraction_cblock_counts': {'1': 2, '1/9': 3, '1/
simplify_2
{'fraction_cblock_total_count': 6, 'fraction_cblock_counts': {'1': 2, '1/12': 3, '1
simplify_2
{'fraction_cblock_total_count': 7, 'fraction_cblock_counts': {'1': 2, '1/12': 4, '1
simplify_2
{'mult_d': '2', 'mult_n': '2'}
```

```
In [100]: my_val = (str(df3.iloc[0]['response']))
          my_val = my_val.replace("': ","_")
          my_val = my_val.replace("_{"," ")
          my_val = my_val.replace("_[",", ")
          for c in [']','[','{','}',"'",""]:
              my_val = my_val.replace(c,'')

In [101]: my_val

Out[101]: 'numberline_associations, obj_value_A, obj_name_a_text, pos_value_1.75253

In [95]: str(df3.iloc[0]['response'])

Out[95]: "{'numberline_associations': [[{'obj_value': 'A', 'obj_name': 'a_text', 'p

In [124]: def stringify_response(resp):
              my_val = str(resp).replace("': ","_")
              my_val = my_val.replace("_{"," ")
              my_val = my_val.replace("_[",", ")
              for c in [']','[','{','}',"'",""",","]:
                  my_val = my_val.replace(c,'')
              return my_val

In [125]: stringify_response(df3.iloc[0]['response'])

Out[125]: 'numberline_associations obj_value_A obj_name_a_text pos_value_1.7525355
```

```
In [126]: df3['response_str'] = df3['response'].apply(stringify_response)

In [127]: for idx in range(20):
              print (idx, df3['response_str'].iloc[idx])

0 numberline_associations obj_value_A obj_name_a_text pos_value_1.752535558428964 p
1 numberline_associations obj_value_A obj_name_a_text pos_value_1.674953531598513 p
2 numberline_associations obj_value_A obj_name_a_text pos_value_1.6671953289154677
3 numberline_associations obj_value_A obj_name_a_text pos_value_1.6206461128171974
4 bitmap_text_inputs input 4 bitmap_text_interp input_1 1/2  +  1/6  = 1 4/6 input_
5 num_6 den_8 whole_3 fraction_input_value_3 6/8
6 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf tot
7 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf tot
8 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf tot
9 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf tot
10 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
11 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
12 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
13 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
14 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
15 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
16 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
17 plain_image_groups url_assets/cms/wootmath_fractions/misc_objects/ant_alt.swf to
18 bitmap_text_interp input_a_1/2 fraction_input_value_Fraction 2/4 input_A_1 bitma
19 bitmap_text_interp input_a_1/3 fraction_input_value_Fraction 2/6 input_A_1 bitma


In [129]: df3.columns

Out[129]: Index(['_id', 'bonus', 'correct', 'diff', 'id', 'incomplete', 'lesson',
                 'level_summary', 'problem_set', 'problem_set_id',
                 'problem_set_subspace', 'qual_id', 'randomly_selected', 'response'
                 'response_idx', 'retried', 'screenshot_url', 'second_try', 'sessio
                 'student', 'sublesson', 't', 'time_spent', 'timestamp', 'txt',
                 'untouched', 'grade', 'school_id', 'section_id', 'student_id',
                 'blank_slate_mastery', 'description', 'entered', 'lesson_type',
                 'lm_stats', 'mastery', 'name', 'path', 'problems', 'stars', 'subje
                 't_elapsed', 'time_entered', 'time_exited', 'title', 'type',
                 'unit_name', 'unit_rank', 'nretry_right', 'nretry_wrong', 'nright'
                 'ntotal', 'nuntouched', 'nwrong', 'percent_correct', 'response_str
                dtype='object')

In [131]: df3.to_csv('data_frame_with_string_response.csv')

In [132]: df_lesson1.to_csv('lesson_summary.csv')

In [ ]:
```