

In [5]:

```
import pandas as pd
import seaborn as sns
%matplotlib inline

pd.options.display.max_columns = 70
pd.options.display.max_rows = 20

top10_data = pd.read_csv('top_10.csv', header = None)
top10_data = top10_data.transpose()

top10_data.columns = top10_data.iloc[0] # set column header
top10_data = top10_data.drop(0) # drop duplicated row
top10_data = top10_data.drop(1) # drop duplicated row
top10_data = top10_data.drop(8) # drop duplicated row
top10_data['cat'] = "top 10" # create top/overall cat
top10_data.rename(columns={'industry': 'thermometer'}, inplace=True)

random_data = pd.read_csv('random_comma_data.csv', header = None)
random_data = random_data.transpose()

random_data.columns = random_data.iloc[0] # set column header
random_data = random_data.drop(0) # drop duplicated row
random_data = random_data.drop(1) # drop duplicated row
random_data = random_data.drop(8) # drop duplicated row
random_data['cat'] = "random" # create top/overall cat
random_data.rename(columns={'industry': 'thermometer'}, inplace=True)

top10_data
```

Out[5]:

	thermometer	automotive- automotive-other	automotive- automotive-other	automotive- automotive-other	automotive- automotive-other	automotive- automotive-other	automotive- automotive-other	automotive- automotive-other
2	a_weighted_rating	0.800508673406529	0.899243370723261	0.617104285874213	1.0	0.996440454544235	0.840841115691487	0.84931235014502
3	a_visibility	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	a_spread	0.45	0.88	0.598076923076923	0.5	0.348245614035088	0.88	0.88
5	a_volume	0.45	0.275	0.675	0.475	0.475	0.6	0.65
6	a_time	1.0	0.285714285714286	1.0	0.642857142857143	1.0	0.821428571428571	1.0
7	a_length	0.708812260536399	0.461538461538462	0.459302325581395	0.736842105263158	0.192982456140351	0.628571428571429	0.75

6 rows x 600 columns

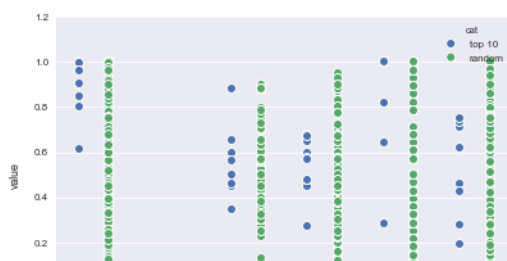
In [ ]:

```
import matplotlib.pyplot as plt

prev = None # print unique columns

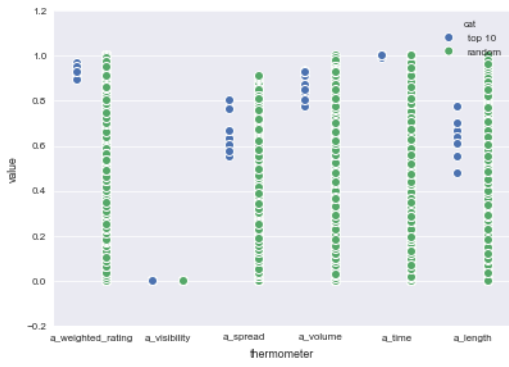
for column in top10_data:
    if column != "cat" and column != "thermometer" and column != prev:
        print column
        prev = column
        temp = pd.melt(top10_data, id_vars=['thermometer', 'cat'], value_vars=[column])
        temp_random = pd.melt(random_data, id_vars=['thermometer', 'cat'], value_vars=[column])
        result = temp.append(temp_random)
        result['value'] = result['value'].astype(float)
        sns.set()
        sns.set_context("paper")
        sns.stripplot(data = result, x="thermometer", y='value', hue='cat')
        plt.show()
```

automotive-automotive-other

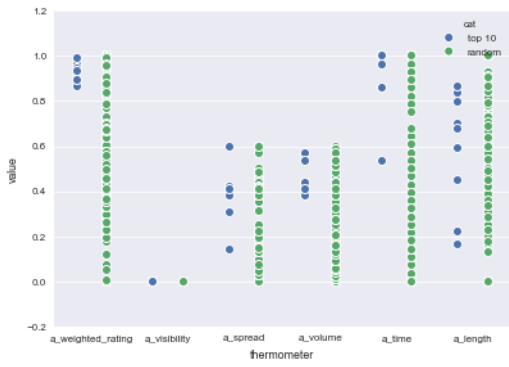




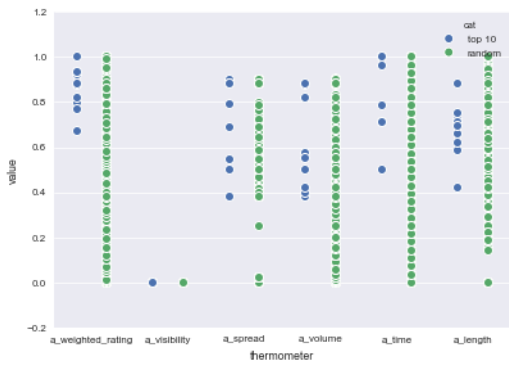
automotive-dealer



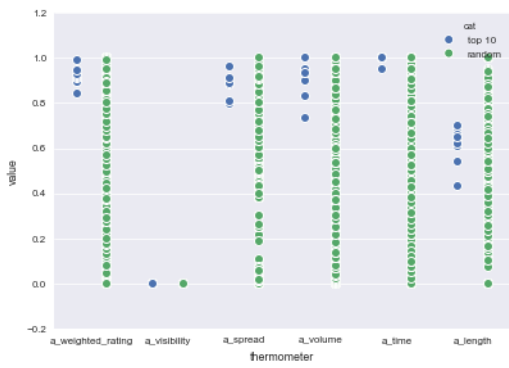
automotive-oem



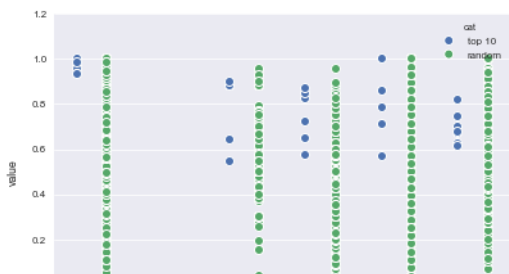
automotive-parts



automotive-rental

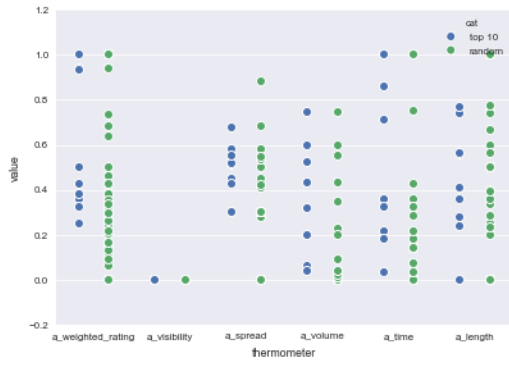


automotive-repair-&-service

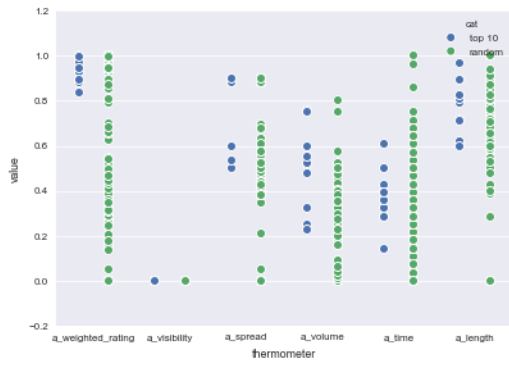




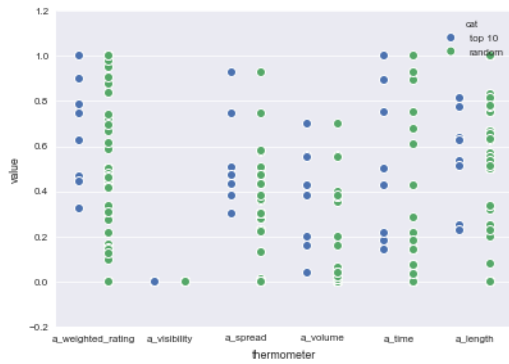
automotive-tires



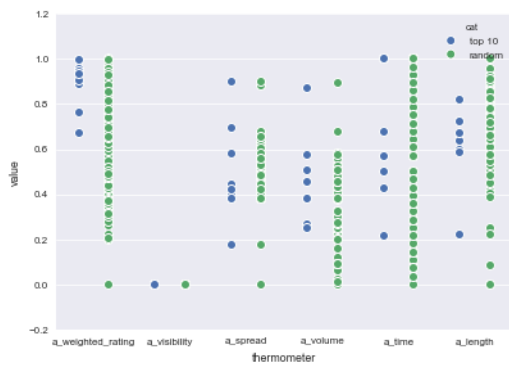
beauty-beauty-other



beauty-hair-grooming



beauty-skin-care

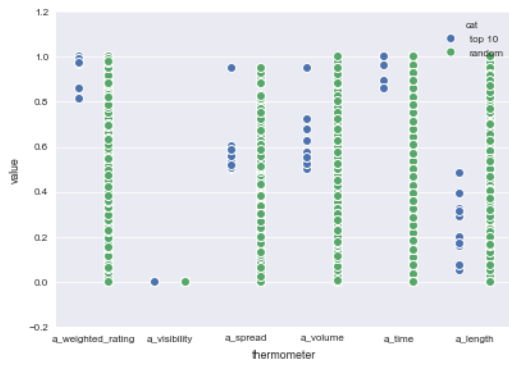


beauty-spa

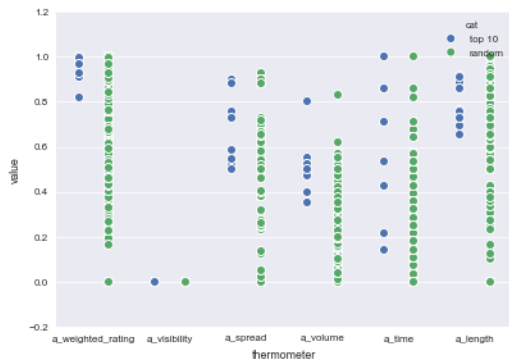




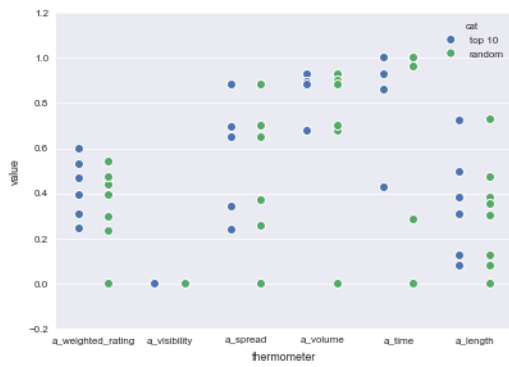
default-default



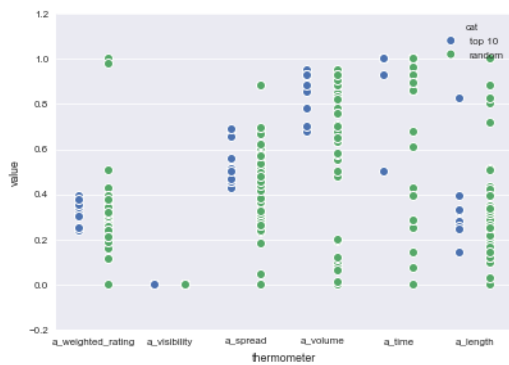
education-education-other



entertainment-amusement-park



entertainment-live-performance-&-sports

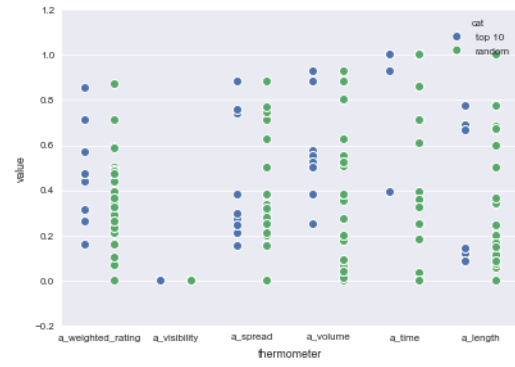


entertainment-museums-and-parks

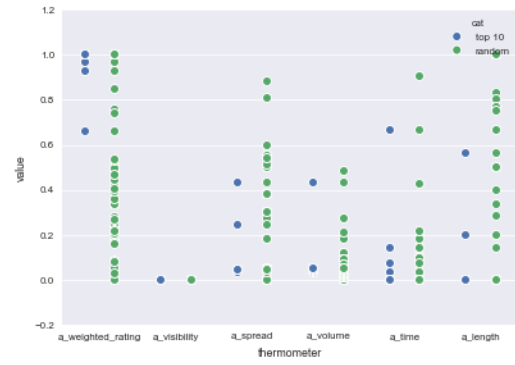




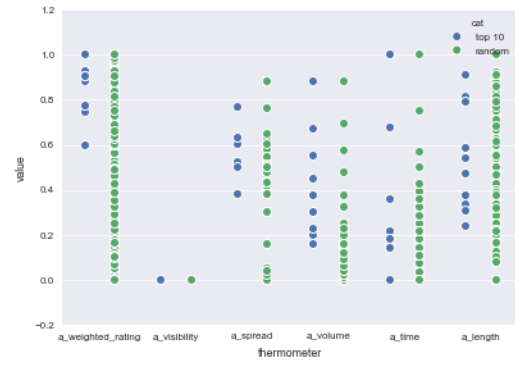
entertainment-other



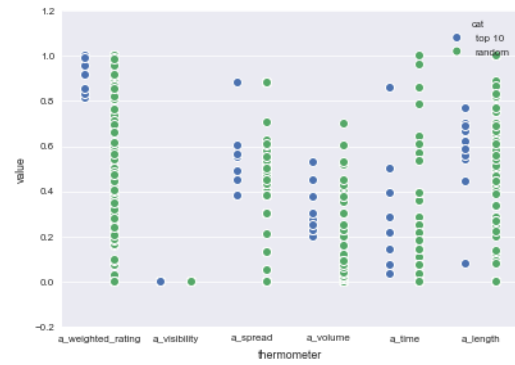
financial-services-accounting



financial-services-banks

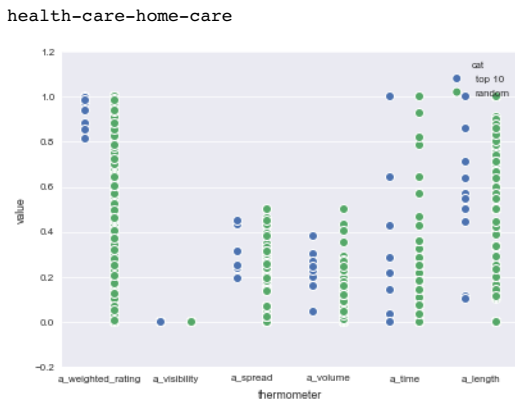
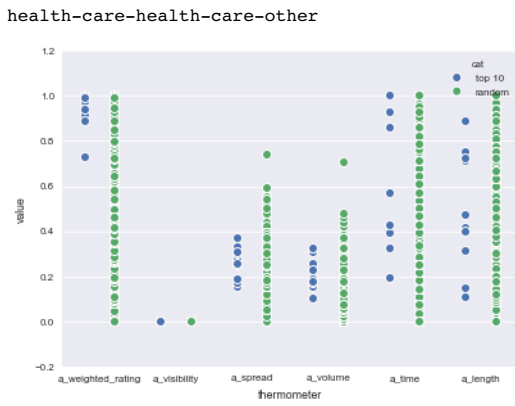
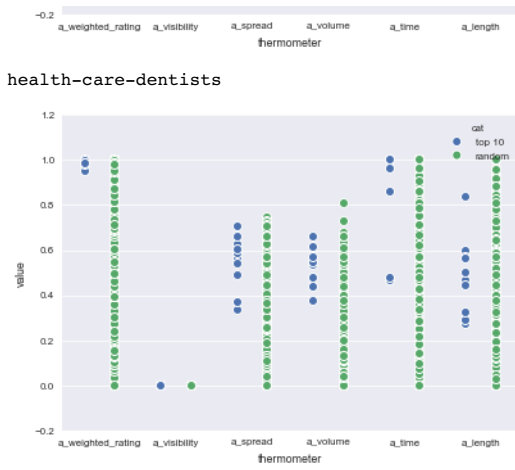


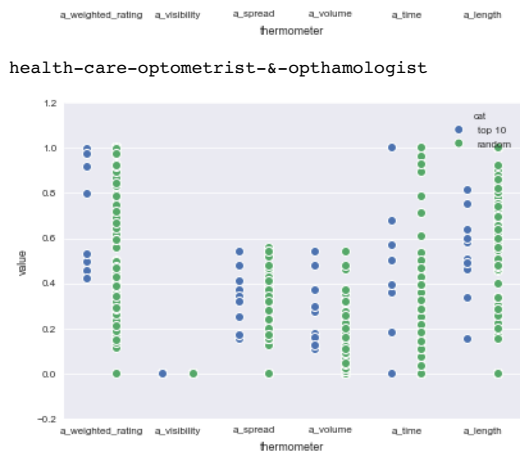
financial-services-financial-services-other



financial-services-insurance







In [ ]: