

databricksCaseStudy_ALT

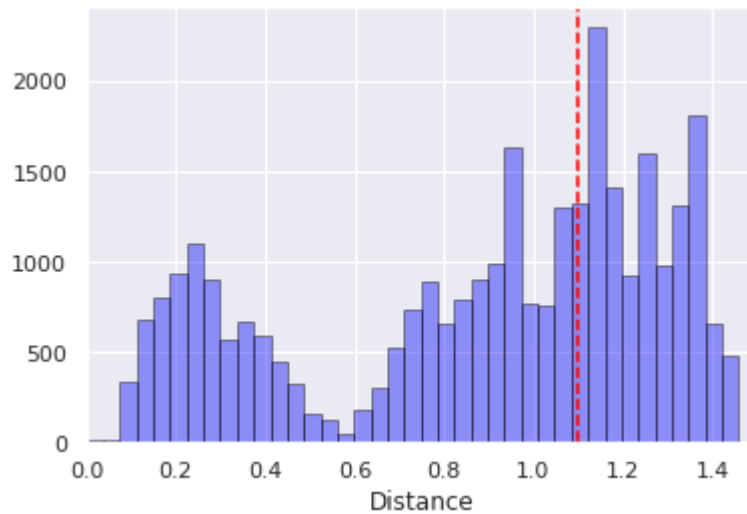
```
import numpy as np
import pandas as pd
import seaborn as sns
import time
import re
import os
import pyspark
import matplotlib.pyplot as plt
from pyspark import SparkContext
sns.set(style="darkgrid")
```

```
#df = spark.read.csv('/FileStore/tables/Adatok.csv', sep = ';', header =
"True")
df = spark.read.csv('/FileStore/tables/Distances_4p.csv', sep = ';', header
= "True")
df = df.toPandas()
df.Distance = df.Distance.astype("double")
df.head(5)
```

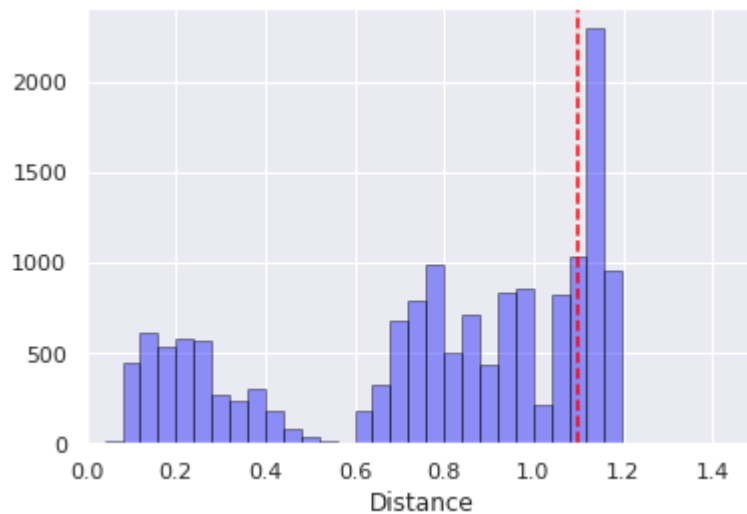
Out[26]:

	Ref_User	Ref_User_Emotion	Ref_User_Emotion_Percentage	Eval	Eval_User_Emotion	Eval
0	Alpha	A	9	Alpha	A	
1	Alpha	A	10	Alpha	A	
2	Alpha	A	8	Alpha	A	
3	Alpha	A	9	Alpha	A	
4	Alpha	A	8	Alpha	A	

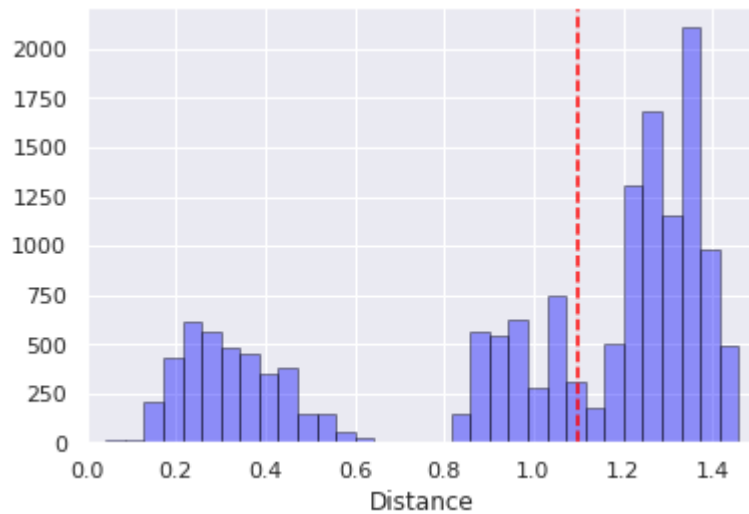
```
distp = sns.distplot(df['Distance'], hist=True, kde=False,color =
'blue',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



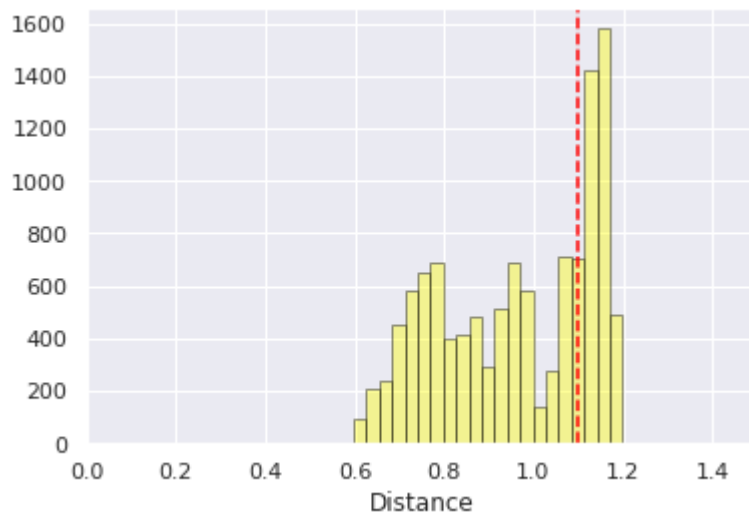
```
df_128 = df[df.Dimension == '128']
distp_128 = sns.distplot(df_128['Distance'], hist=True, kde=False, color =
'blue',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



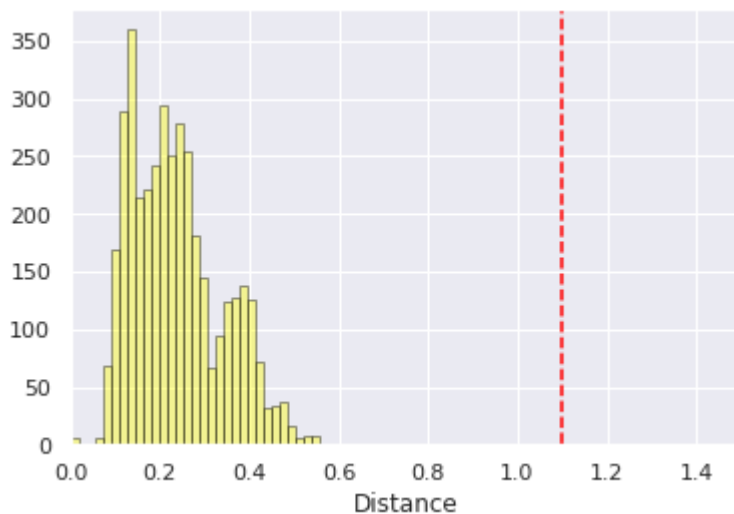
```
df_512 = df[df.Dimension == '512']
distp = sns.distplot(df_512['Distance'], hist=True, kde=False, color =
'blue',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



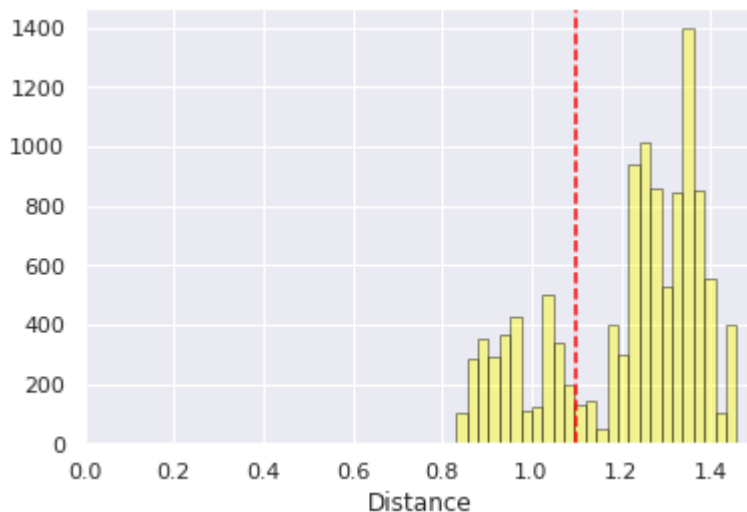
```
df_within_t = df[(df.Within_User == '0') & (df.Dimension == '128')]
distp = sns.distplot(df_within_t['Distance'], hist=True, kde=False, color =
'yellow',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



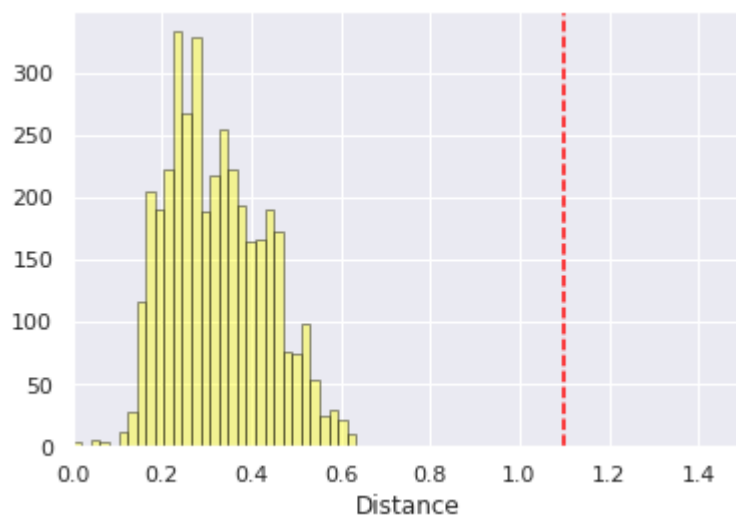
```
df_within_f = df[(df.Within_User == '1') & (df.Dimension == '128')]
distp = sns.distplot(df_within_f['Distance'], hist=True, kde=False, color =
'yellow',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



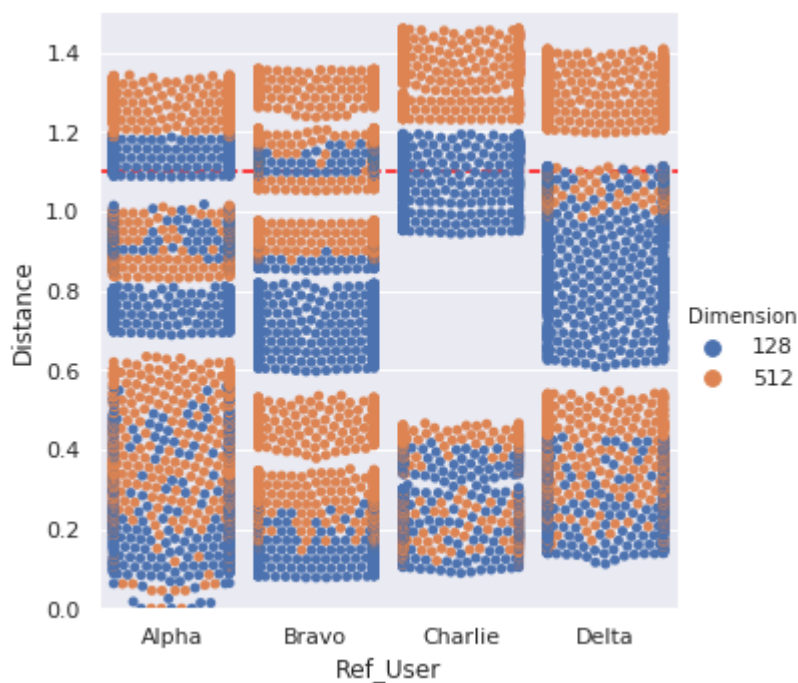
```
df_within_f = df[(df.Within_User == '0') & (df.Dimension == '512')]
distp = sns.distplot(df_within_f['Distance'], hist=True, kde=False, color =
'yellow',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



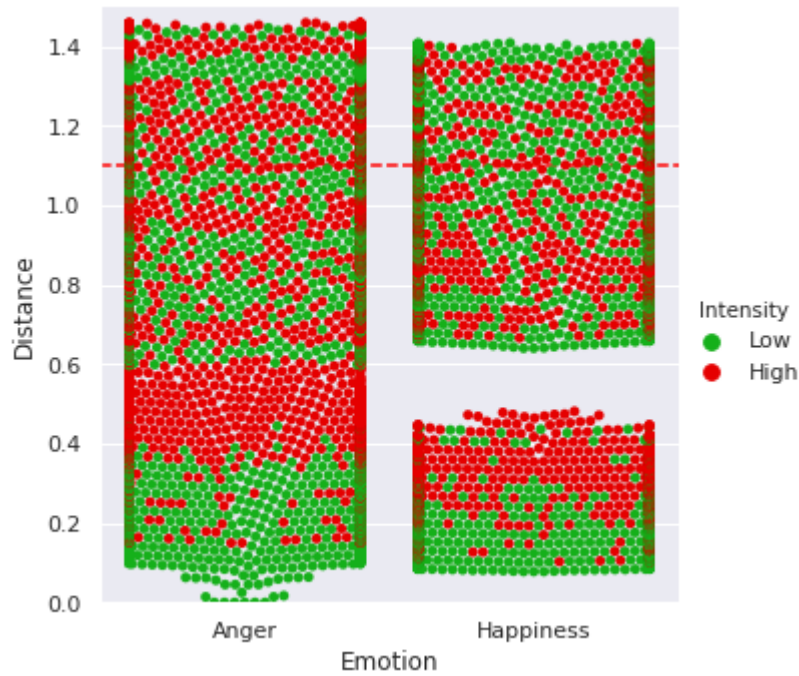
```
df_within_f = df[(df.Within_User == '1') & (df.Dimension == '512')]
distp = sns.distplot(df_within_f['Distance'], hist=True, kde=False, color =
'yellow',hist_kws={'edgecolor':'black'})
plt.xlim(0.0, 1.5)
plt.axvline(1.1, color="red", linestyle="--");
```



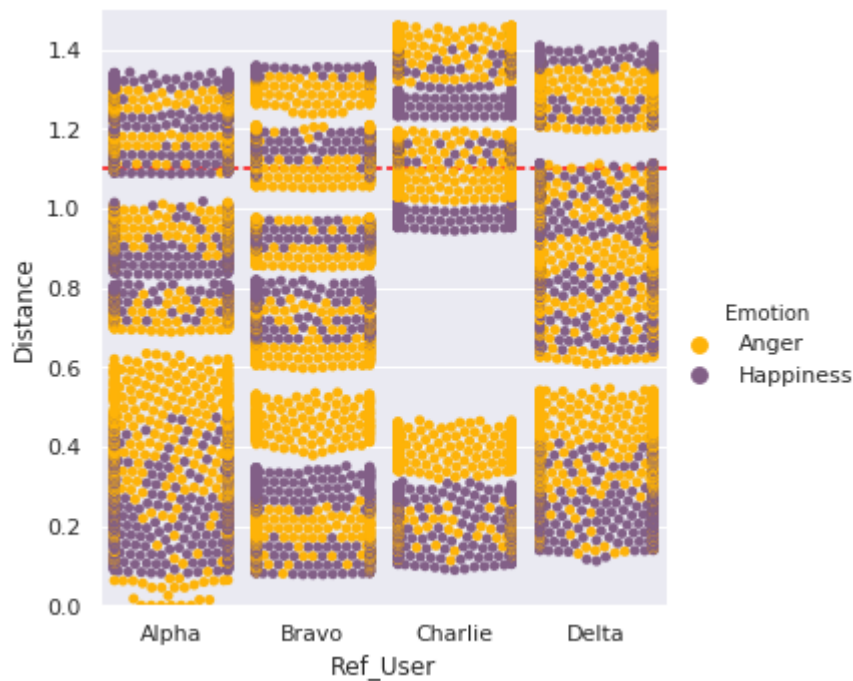
```
cat = sns.catplot(x="Ref_User", y="Distance", hue="Dimension", kind="swarm",
data=df);
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



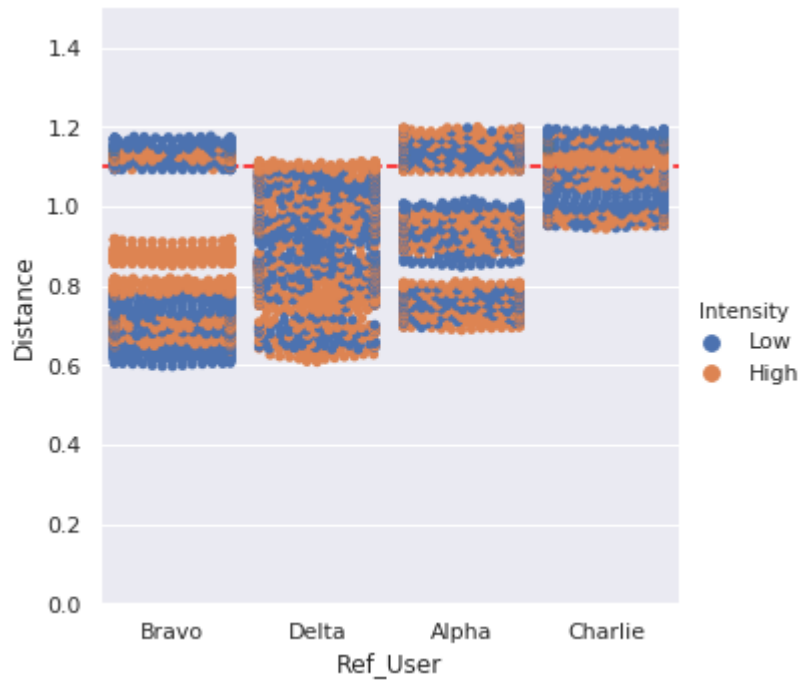
```
cat = sns.catplot(x="Emotion", y="Distance", hue="Intensity", kind="swarm",
data=df, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



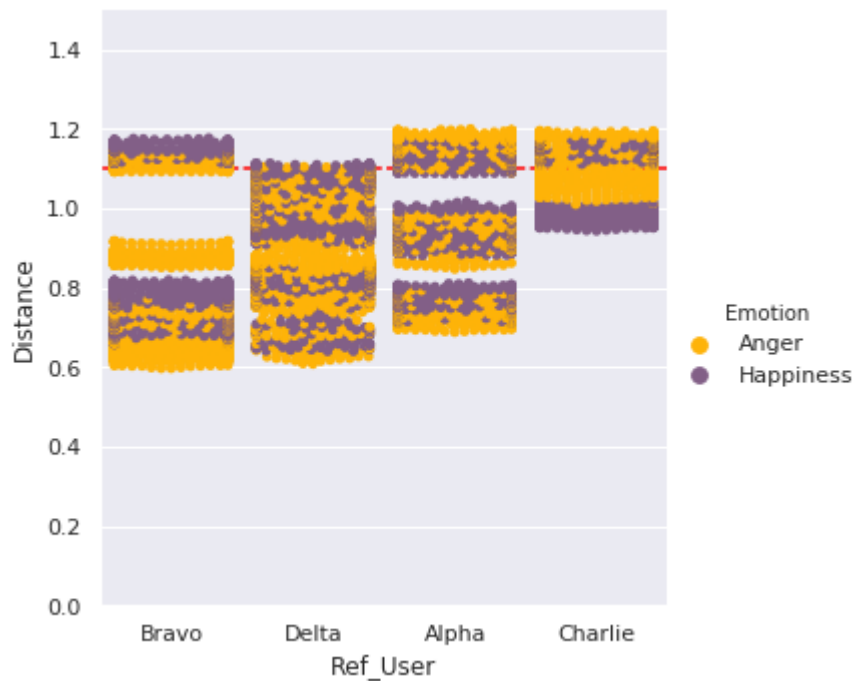
```
cat = sns.catplot(x="Ref_User", y="Distance", hue="Emotion", kind="swarm",
data=df, palette=sns.xkcd_palette(["amber","dusty purple"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



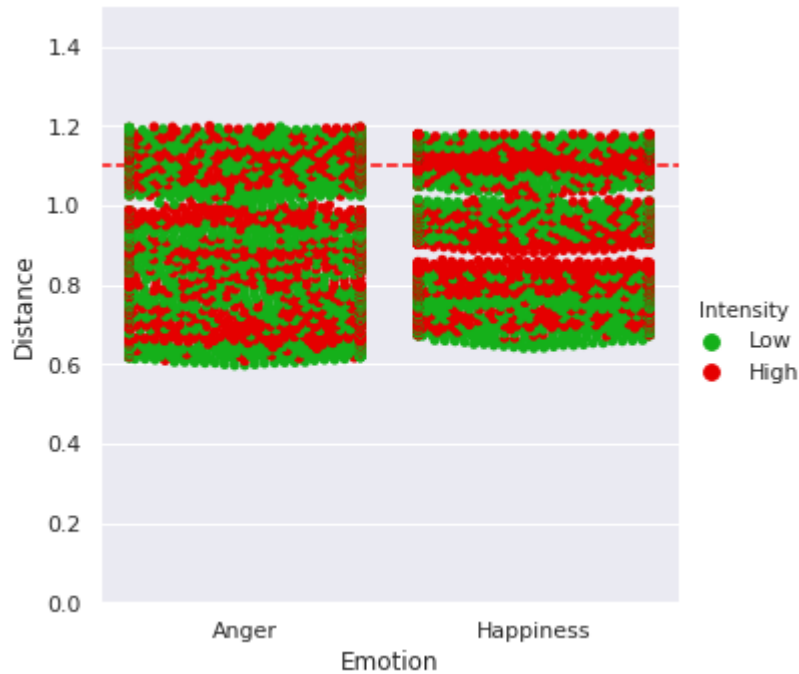
```
df_128 = df[(df.Within_User == '0') & (df.Dimension == '128')]
cat_128 = sns.catplot(x="Ref_User", y="Distance", hue="Intensity",
kind="swarm", data=df_128);
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



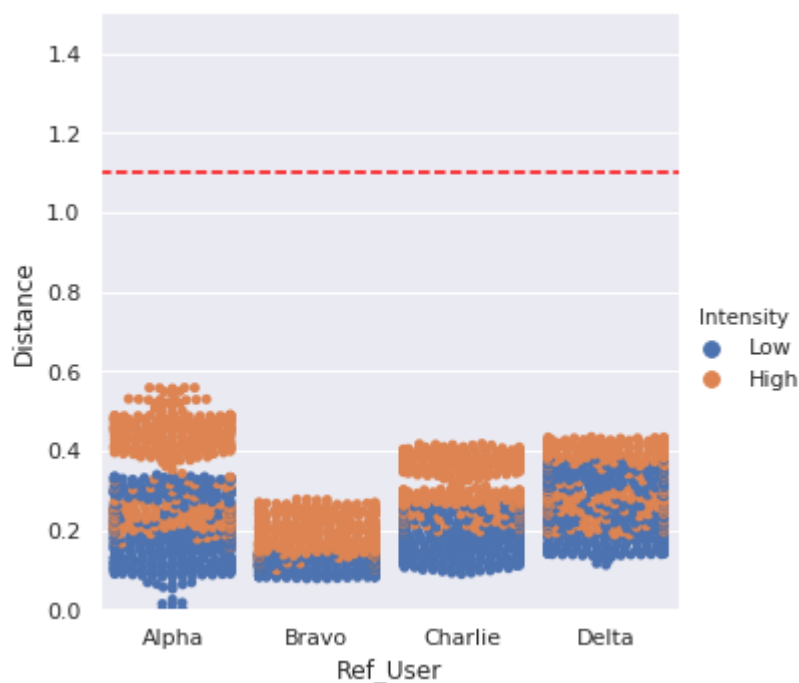
```
df_128 = df[(df.Within_User == '0') & (df.Dimension == '128')]
cat_128 = sns.catplot(x="Ref_User", y="Distance", hue="Emotion",
kind="swarm", data=df_128, palette=sns.xkcd_palette(["amber", "dusty
purple"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



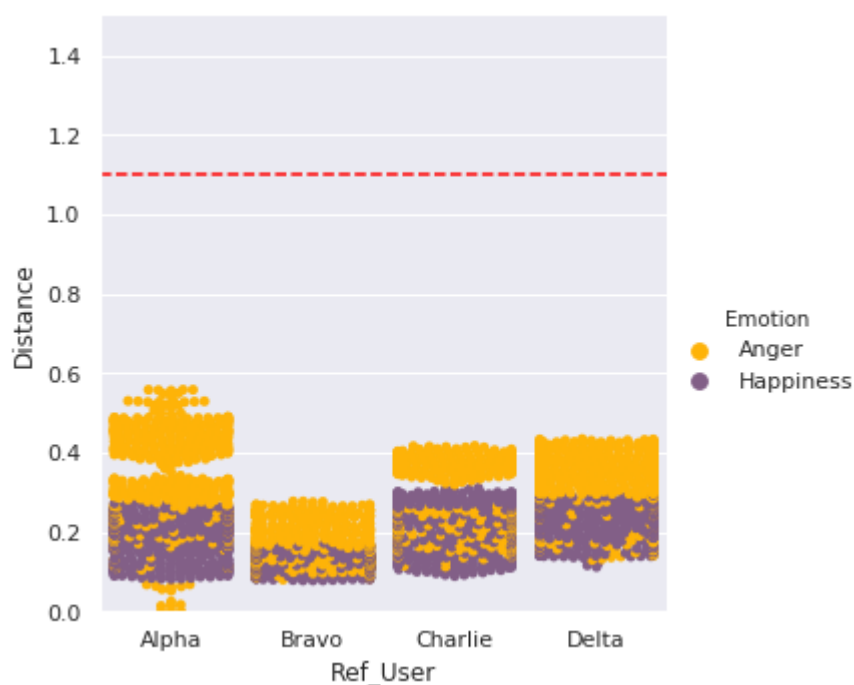
```
df_128 = df[(df.Within_User == '0') & (df.Dimension == '128')]
cat = sns.catplot(x="Emotion", y="Distance", hue="Intensity", kind="swarm",
data=df_128, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



```
df_128 = df[(df.Within_User == '1') & (df.Dimension == '128')]
cat_128 = sns.catplot(x="Ref_User", y="Distance", hue="Intensity",
kind="swarm", data=df_128);
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```

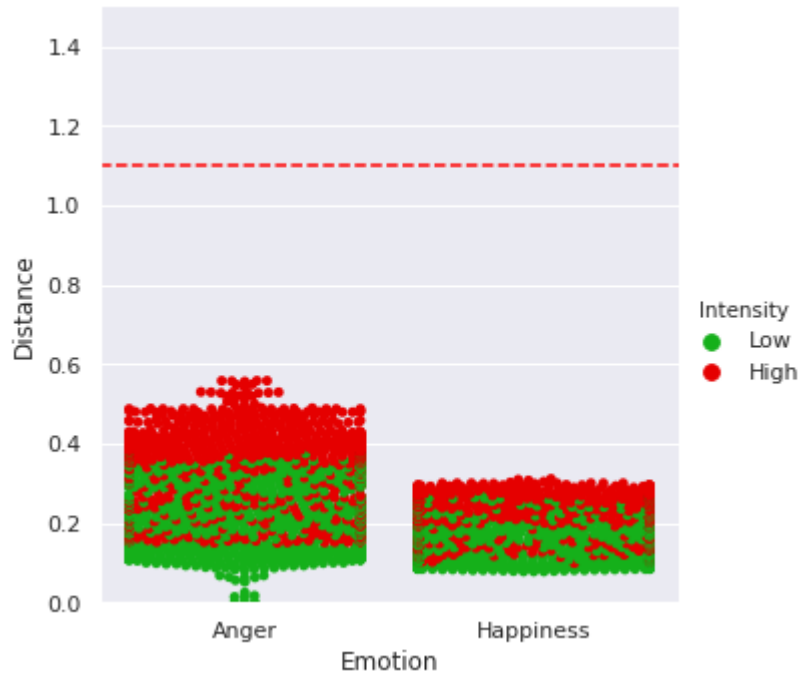
```
df_128 = df[(df.Within_User == '1') & (df.Dimension == '128')]
cat_128 = sns.catplot(x="Ref_User", y="Distance", hue="Emotion",
kind="swarm", data=df_128, palette=sns.xkcd_palette(["amber", "dusty
purple"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



```

df_128 = df[(df.Within_User == '1') & (df.Dimension == '128')]
cat = sns.catplot(x="Emotion", y="Distance", hue="Intensity", kind="swarm",
data=df_128, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");

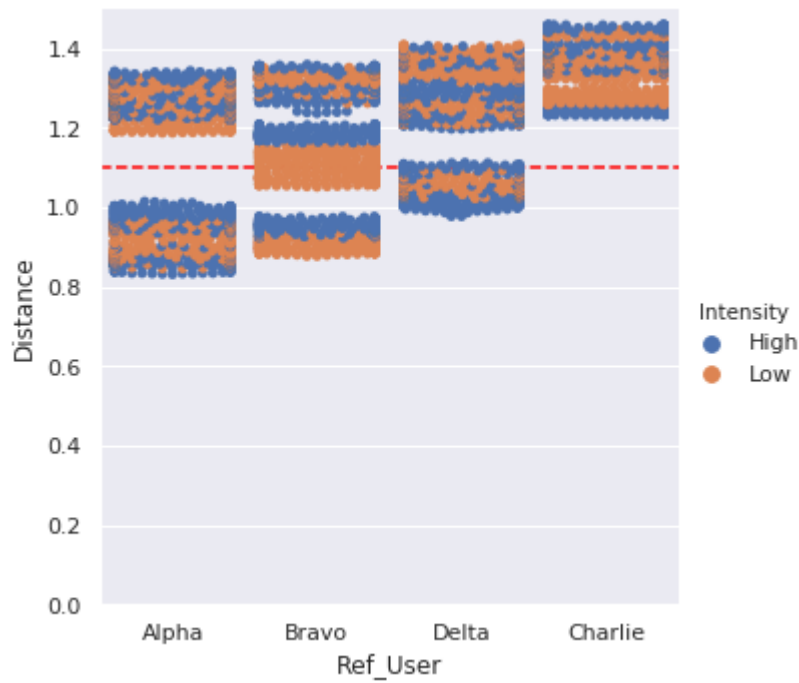
```



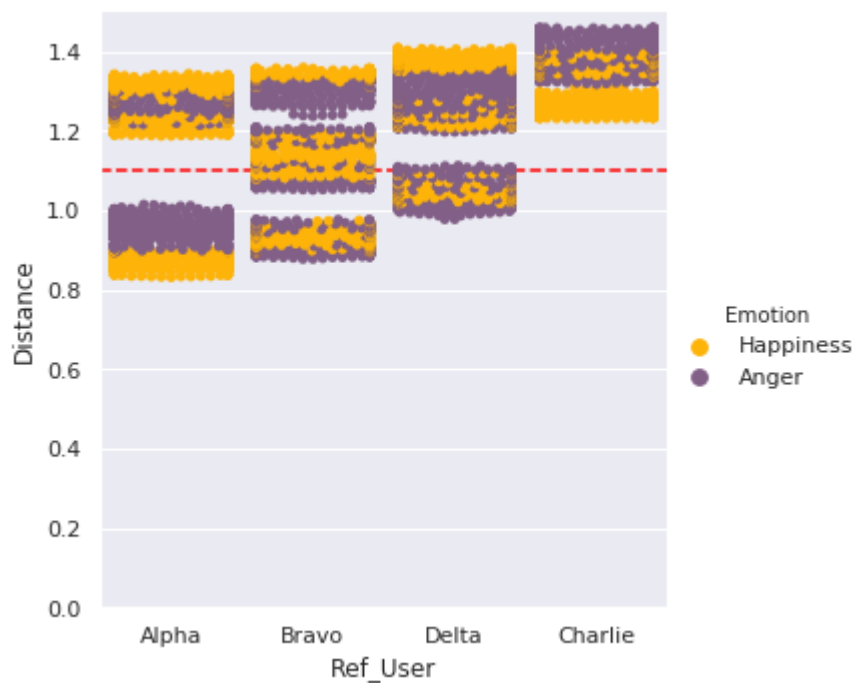
```

df_512 = df[(df.Within_User == '0') & (df.Dimension == '512')]
cat_512 = sns.catplot(x="Ref_User", y="Distance", hue="Intensity",
kind="swarm", data=df_512);
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");

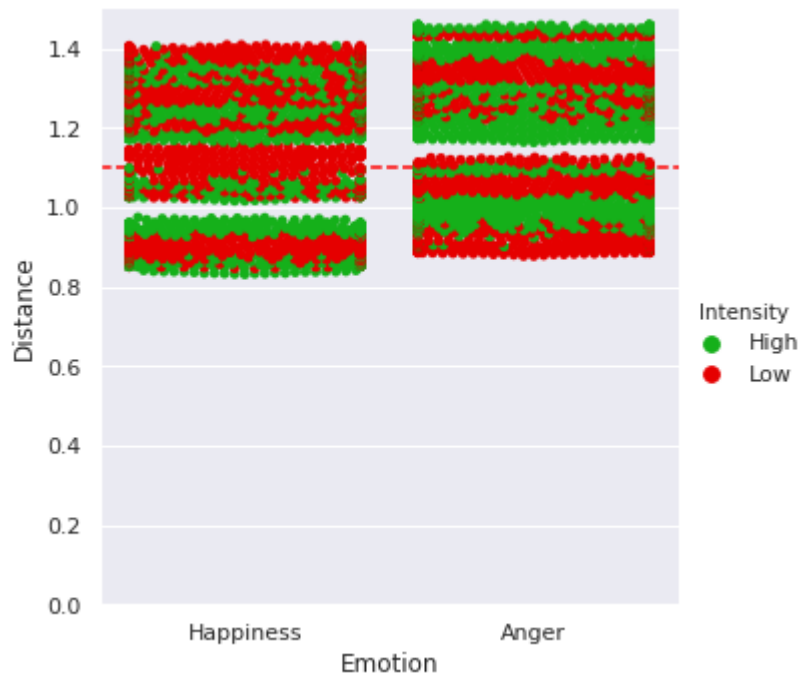
```



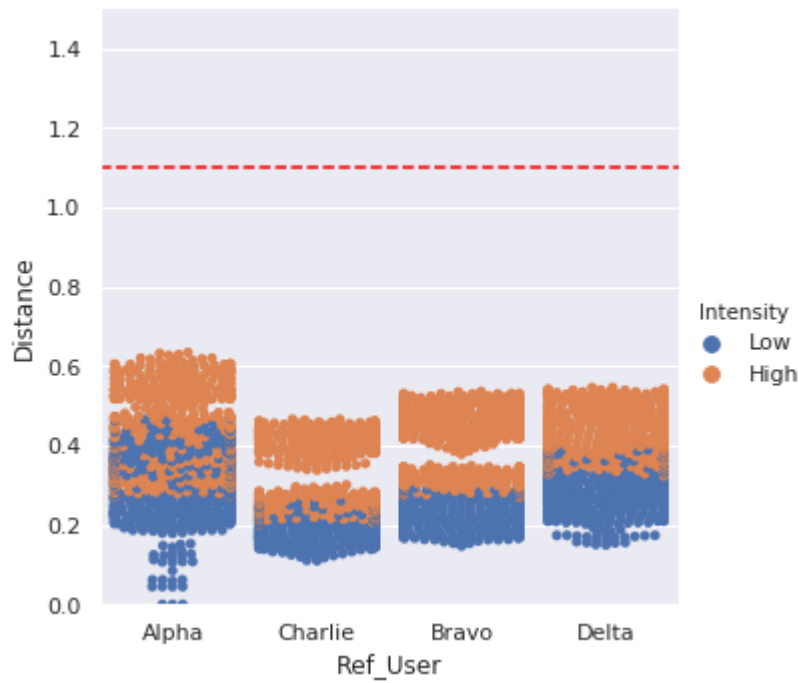
```
df_512 = df[(df.Within_User == '0') & (df.Dimension == '512')]
cat_512 = sns.catplot(x="Ref_User", y="Distance", hue="Emotion",
kind="swarm", data=df_512, palette=sns.xkcd_palette(["amber", "dusty
purple"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



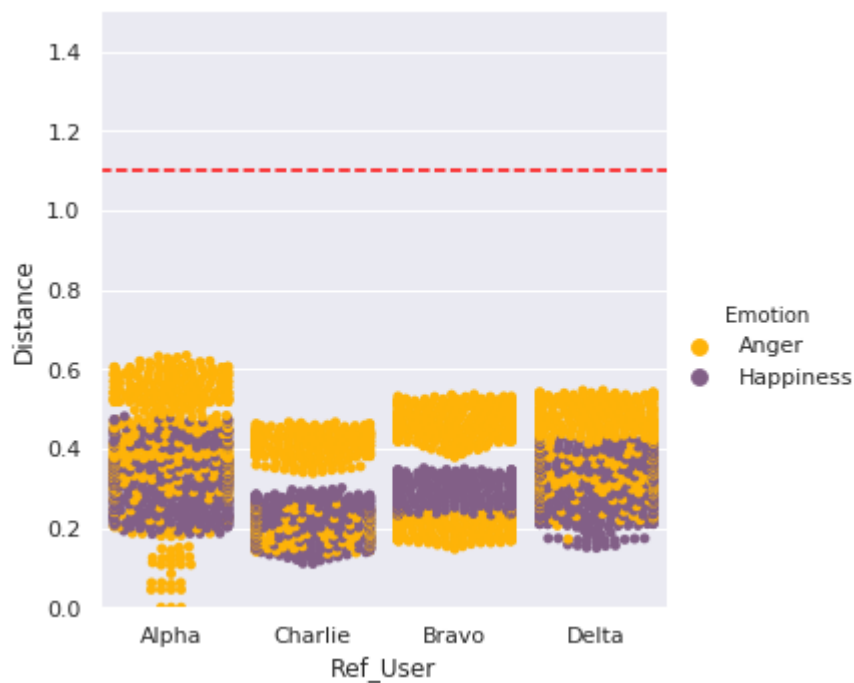
```
df_512 = df[(df.Within_User == '0') & (df.Dimension == '512')]
cat = sns.catplot(x="Emotion", y="Distance", hue="Intensity", kind="swarm",
data=df_512, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



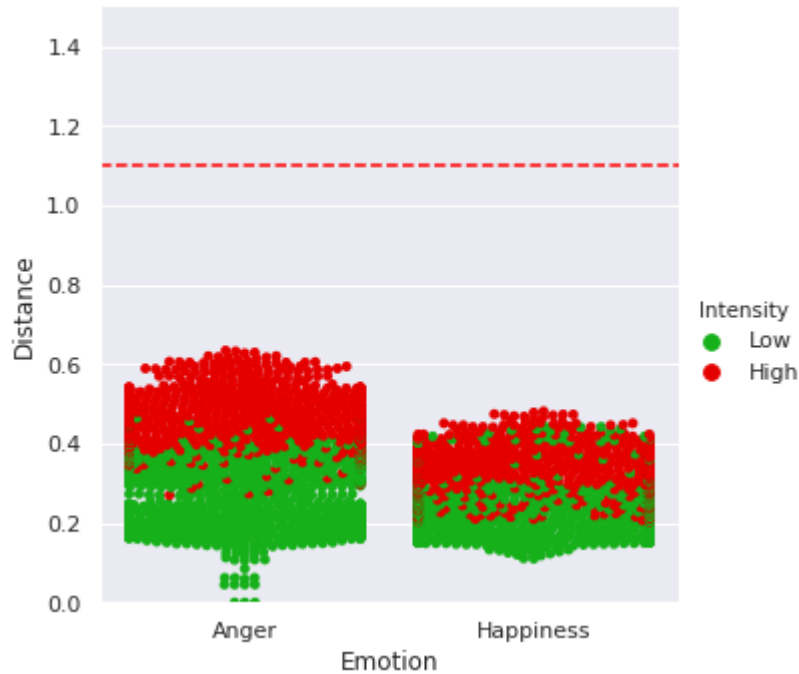
```
df_512 = df[(df.Within_User == '1') & (df.Dimension == '512')]
cat_512 = sns.catplot(x="Ref_User", y="Distance", hue="Intensity",
kind="swarm", data=df_512);
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



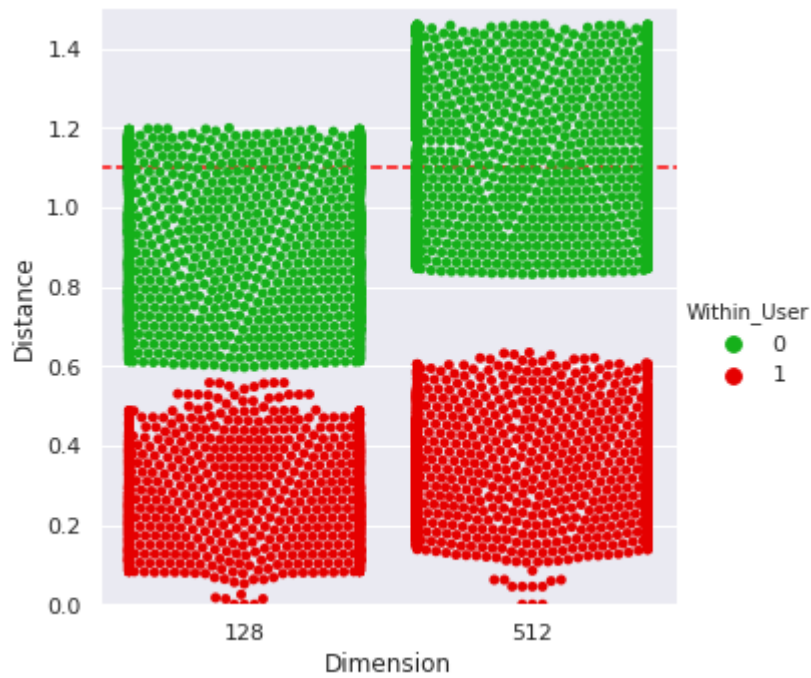
```
df_512 = df[(df.Within_User == '1') & (df.Dimension == '512')]
cat_512 = sns.catplot(x="Ref_User", y="Distance", hue="Emotion",
kind="swarm", data=df_512, palette=sns.xkcd_palette(["amber", "dusty
purple"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



```
df_512 = df[(df.Within_User == '1') & (df.Dimension == '512')]
cat = sns.catplot(x="Emotion", y="Distance", hue="Intensity", kind="swarm",
data=df_512, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```

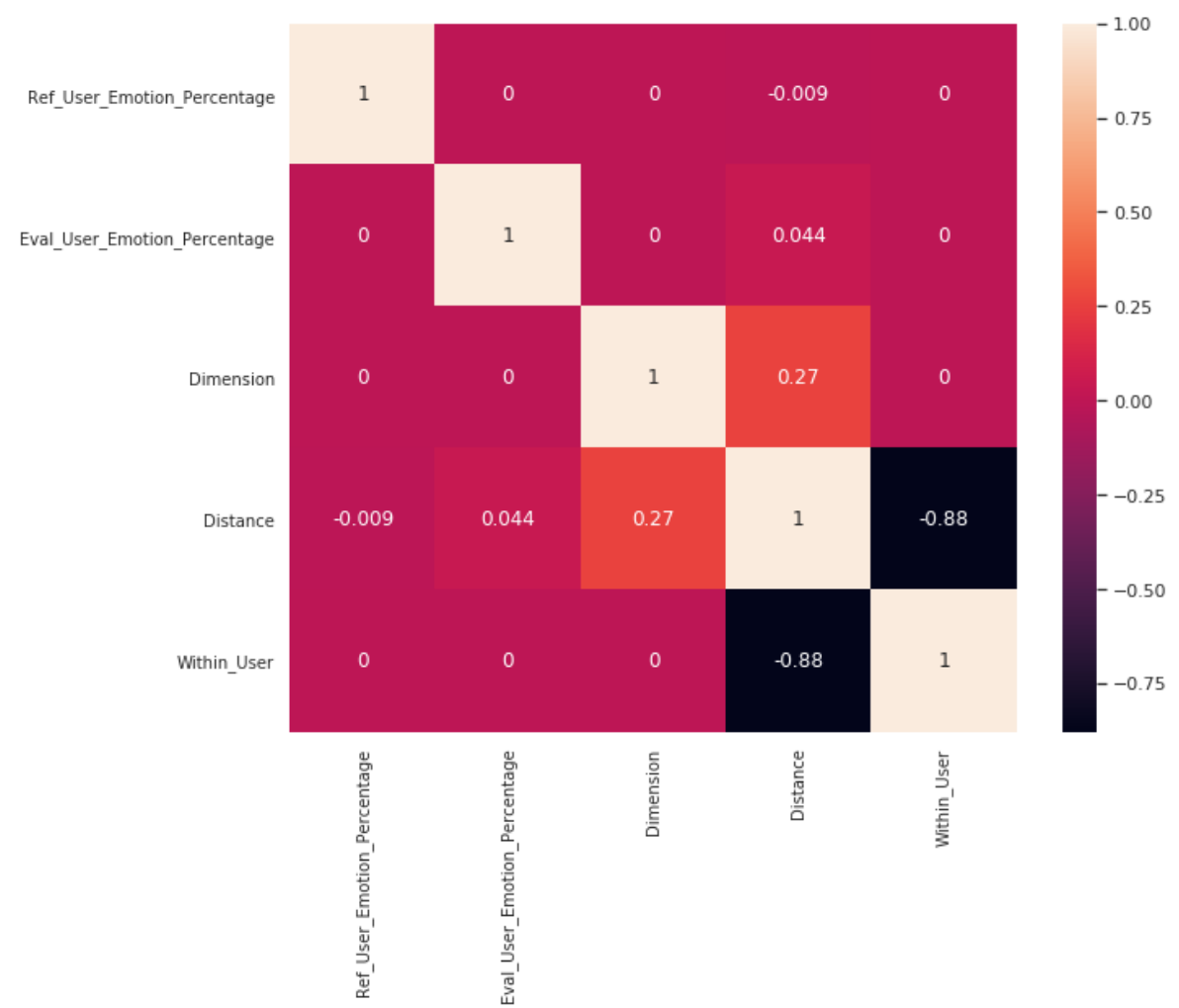


```
df_512_128 = df
cat = sns.catplot(x="Dimension", y="Distance", hue="Within_User",
kind="swarm", data=df_512_128, palette=sns.xkcd_palette(["green","red"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```



```
df.Ref_User_Emotion_Percentage =
df.Ref_User_Emotion_Percentage.astype("int32")
df.Eval_User_Emotion_Percentage =
df.Eval_User_Emotion_Percentage.astype("int32")
df.Dimension = df.Dimension.astype("int32")
df.Within_User = df.Within_User.astype("int32")
```

```
corr_df = df.corr()
sns.heatmap(corr_df, xticklabels=corr_df.columns.values,
yticklabels=corr_df.columns.values, annot = True, annot_kws={'size':12})
heat_map=plt.gcf(); heat_map.set_size_inches(10,8)
plt.xticks(fontsize=10); plt.yticks(fontsize=10);
plt.show()
```



```
df.Emotion = df.Emotion.astype("str")
df.Intensity = df.Intensity.astype("str")
df['Emotion_And_Intensity'] = df[['Emotion', 'Intensity']].apply(lambda x:
'_' .join(x), axis=1)
df.head(5)
```

Out[59]:

	Ref_User	Ref_User_Emotion	Ref_User_Emotion_Percentage	Eval	Eval_User_Emotion	Eva
0	Alpha	A	9	Alpha	A	
1	Alpha	A	10	Alpha	A	
2	Alpha	A	8	Alpha	A	
3	Alpha	A	9	Alpha	A	
4	Alpha	A	8	Alpha	A	


```
df_128_0 = df[(df.Within_User == '0') & (df.Dimension == '128')]
cat_128_0 = sns.catplot(x="Ref_User",
y="Distance",hue="Emotion_And_Intensity", kind="swarm", data=df_128_0,
palette=sns.xkcd_palette(["amber","dusty purple","tea","royal"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```

ValueError: min() arg is an empty sequence

```
df_128_1 = df[(df.Within_User == '1') & (df.Dimension == '128')]
cat_128_1 = sns.catplot(x="Ref_User",
y="Distance",hue="Emotion_And_Intensity", kind="swarm", data=df_128_1,
palette=sns.xkcd_palette(["amber","dusty purple","tea","royal"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```

Command skipped

```
df_512_0 = df[(df.Within_User == '0') & (df.Dimension == '512')]
cat_512_0 = sns.catplot(x="Ref_User",
y="Distance",hue="Emotion_And_Intensity", kind="swarm", data=df_512_0,
palette=sns.xkcd_palette(["amber","dusty purple","tea","royal"]));
plt.ylim(0.0, 1.5)
plt.axhline(1.1, color="red", linestyle="--");
```

Command skipped

ValueError: min() arg is an empty sequence