Separate_EDA

May 27, 2021

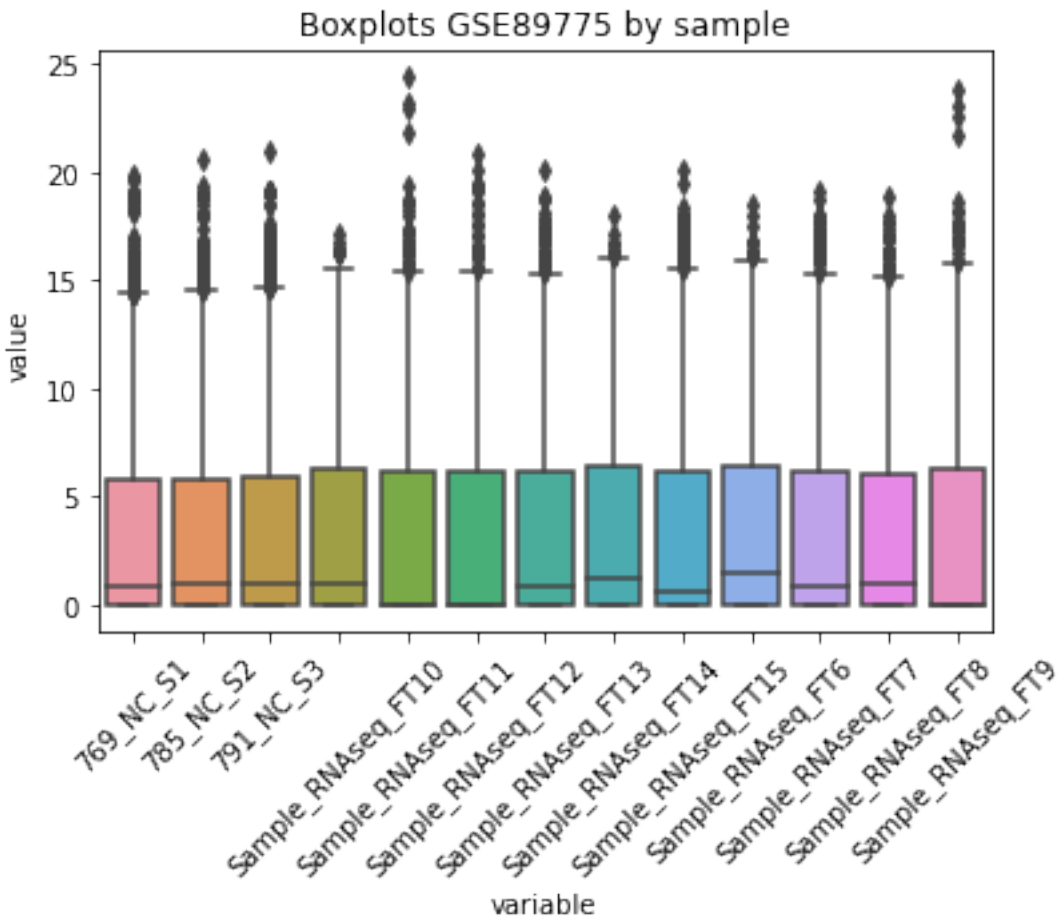
[7]:
```
import numpy as np
import pandas as pd
from scipy import stats
import statsmodels.stats.multitest as multi
from tqdm import tqdm
import matplotlib.pyplot as plt
import seaborn as sns
```

[8]:
```
# Import metadata
metadata_dict = {}
metadata = pd.read_csv('/home/mar/Documents/TFM/May/Task1_v2/Metadata_HB/
↪HB_joint_METADATA.tsv',sep='\t')
lst = []
for i in metadata['type']:
    if str(i).startswith('Hepatoblastoma'):
        lst.append('case')
    elif str(i).startswith('Normal'):
        lst.append('control')
    else:
        lst.append('other')
metadata['class'] = lst
metadata_dict = pd.Series(metadata['class'].values,index=metadata['sample']).
↪to_dict()
```

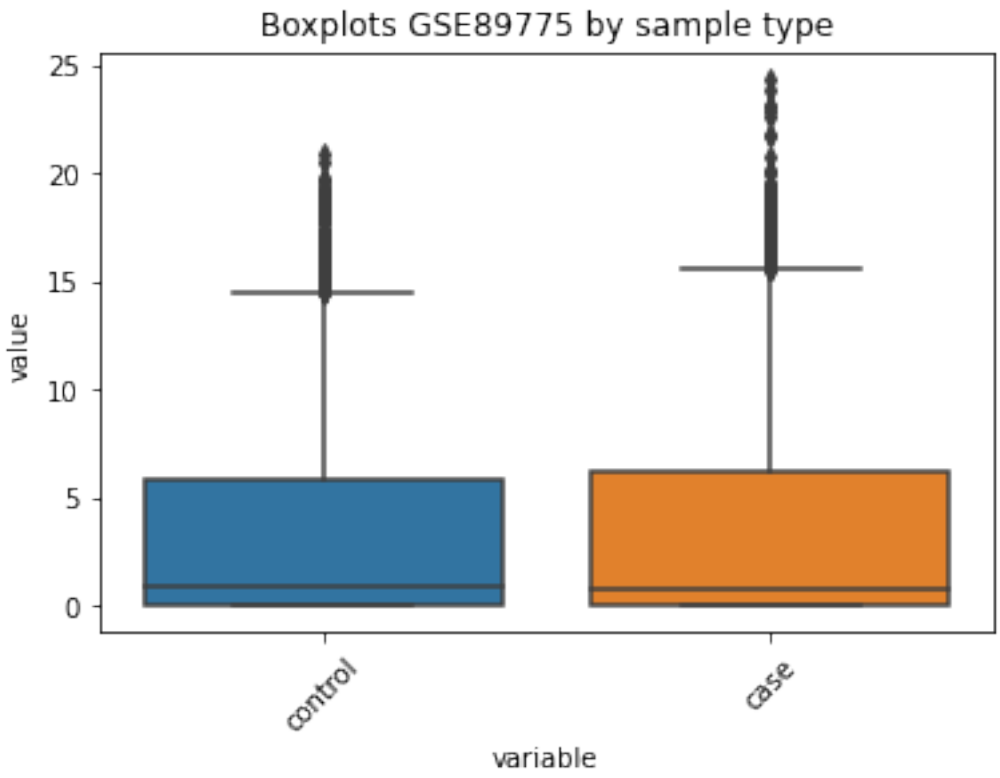
0.1 GSE89775

[9]:
```
# Import data
data = pd.read_csv('/home/mar/Documents/TFM/May/Task1_v2/Matrices_HB/
↪GSE89775_matrix.txt', sep=';', index_col=0)
cols = data.columns.tolist()
```

[10]:
```
# Boxplot by sample
# sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
↪GSE89775 by sample')
# plt.xticks(rotation=45)
# plt.savefig('Figures/GSE89775_samples');
```



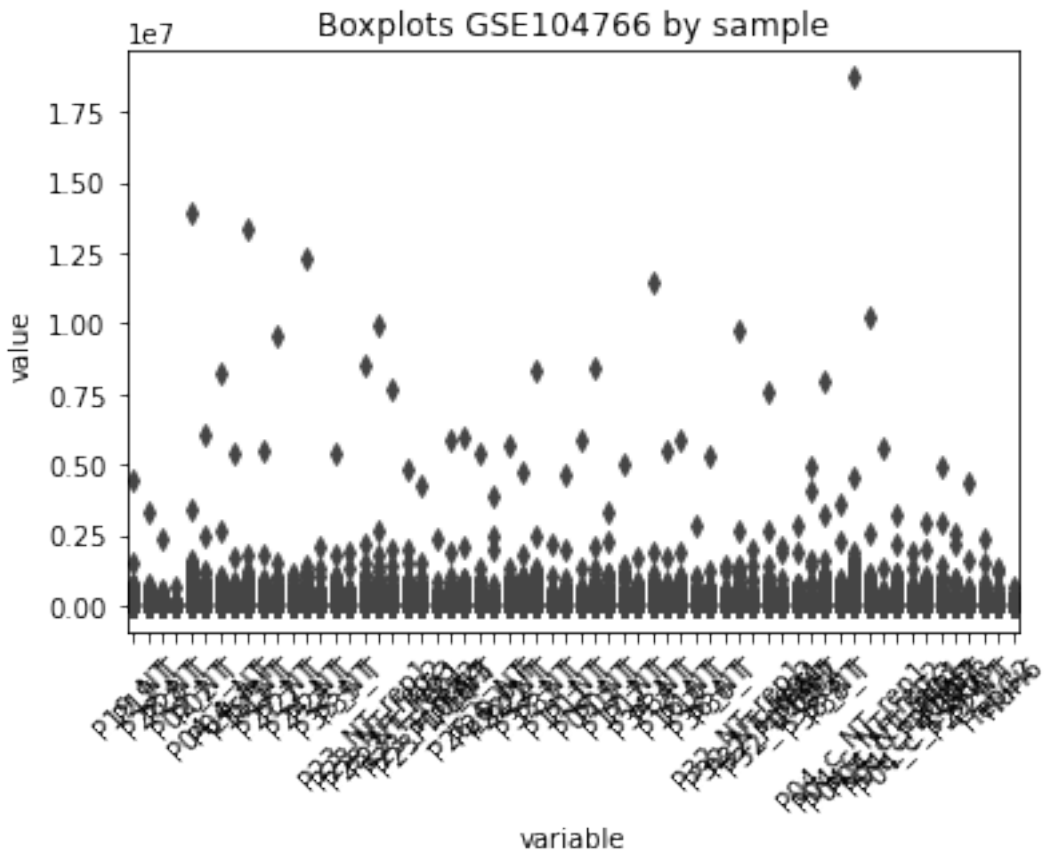
[11]:
```
# Boxplot by sample type
data.columns = data.columns.map(metadata_dict)
cols = data.columns.tolist()
sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
 ↪GSE89775 by sample type')
plt.xticks(rotation=45)
plt.savefig('Figures/GSE89775_types');
```



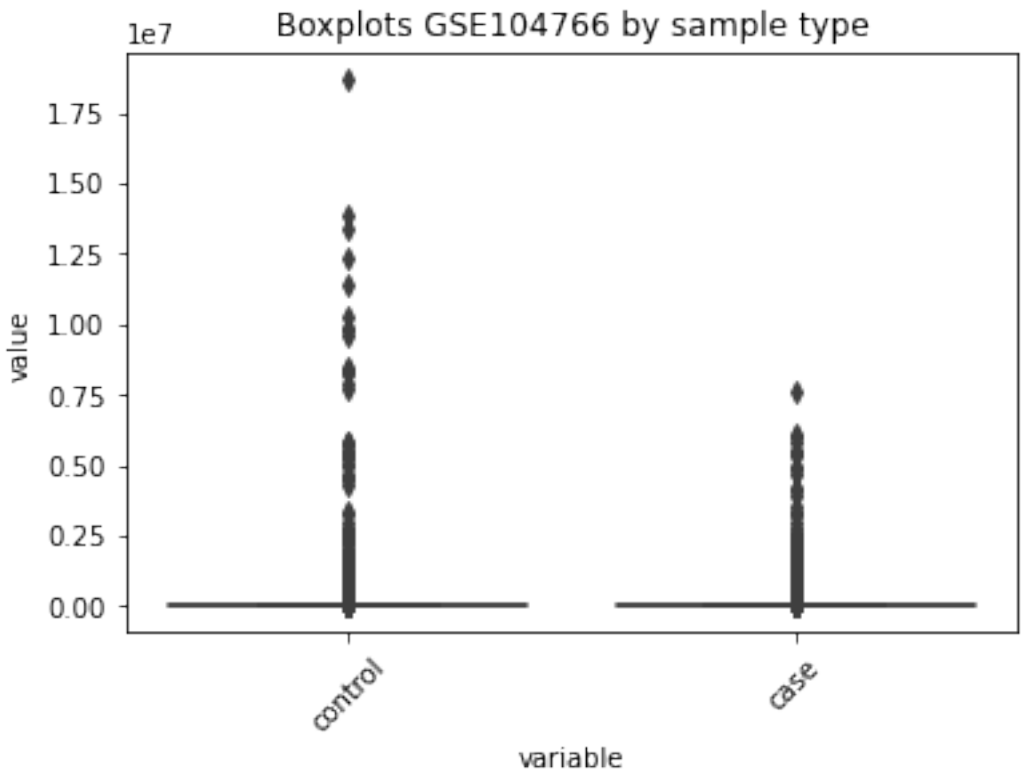
0.2 GSE104766

[12]:
```
# Import data
data = pd.read_csv('/home/mar/Documents/TFM/May/Task1_v2/Matrices_HB/
 ↪GSE104766_matrix.txt', sep=';', index_col=0)
cols = data.columns.tolist()
```

[13]:
```
# Boxplot by sample
# sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
 ↪GSE104766 by sample')
# plt.xticks(rotation=45)
# plt.savefig('Figures/GSE104766_samples');
```



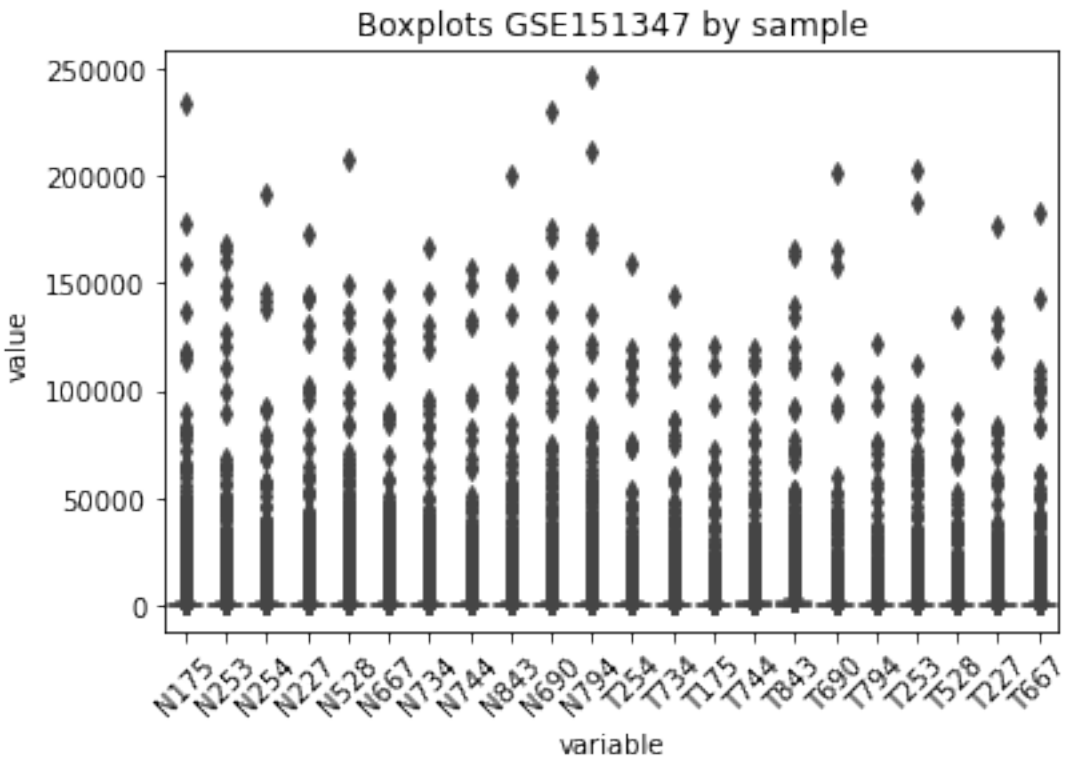
[14]:
```
# Boxplot by sample type
data.columns = data.columns.map(metadata_dict)
cols = data.columns.tolist()
sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
↪GSE104766 by sample type')
plt.xticks(rotation=45)
plt.savefig('Figures/GSE104766_types');
```



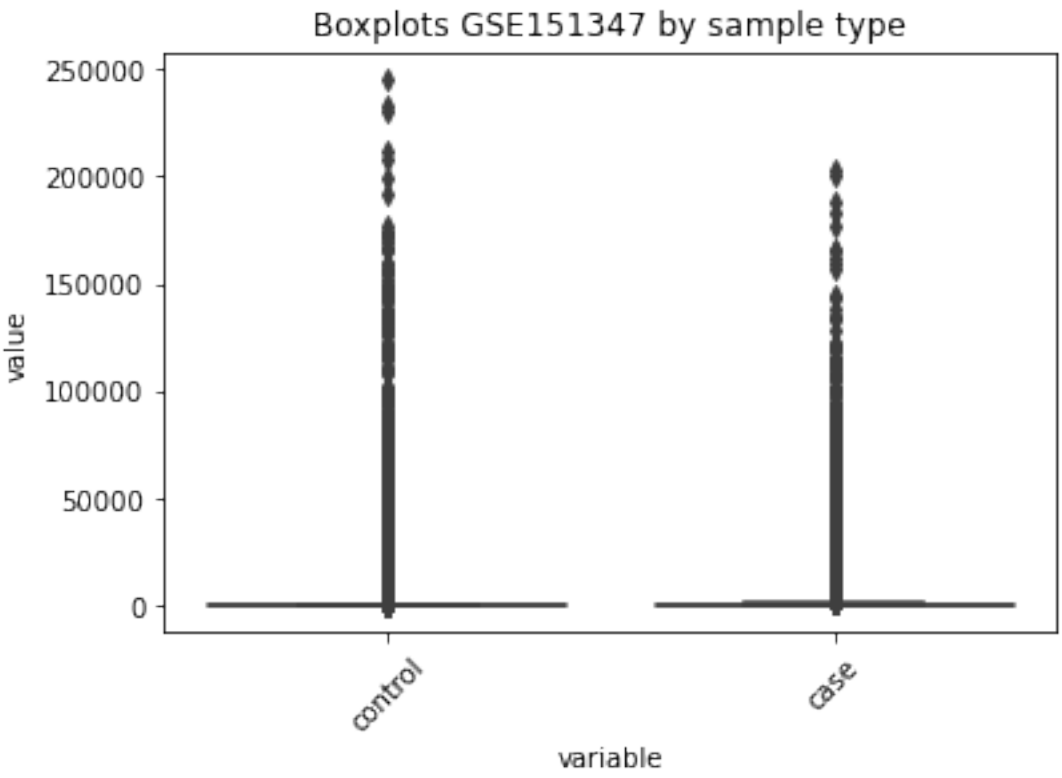
0.3 GSE151347

[15]:
```
# Import data
data = pd.read_csv('/home/mar/Documents/TFM/May/Task1_v2/Matrices_HB/
 ↪GSE151347_matrix.txt', sep=';', index_col=0)
cols = data.columns.tolist()
```

[16]:
```
# Boxplot by sample
# sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
 ↪GSE151347 by sample')
# plt.xticks(rotation=45)
# plt.savefig('Figures/GSE151347_samples');
```



[17]:
```
# Boxplot by sample type
data.columns = data.columns.map(metadata_dict)
cols = data.columns.tolist()
sns.boxplot(x="variable", y="value", data=pd.melt(data)).set_title('Boxplots␣
↪GSE151347 by sample type')
plt.xticks(rotation=45)
plt.savefig('Figures/GSE151347_types');
```


Boxplots GSE151347 by sample type
250000
200000
150000
100000
50000
0
value
control
case
variable