In [1]:
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns # data visualization library
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv('Gender_Age_Dataset.csv')
```

In [3]:
```
gender_mapper = {'M': 0, 'F': 1}
df['Gender'].replace(gender_mapper, inplace=True)
```

In [4]:
```
col = df.columns       # .columns gives columns names in data
print(col)
```

```
Index(['Gender', 'Left-Hippocampus', 'Left-Amygdala', 'CSF',
       'Left-Accumbens-area', 'Left-VentralDC', 'Left-vessel',
       'Left-choroid-plexus', 'Right-Lateral-Ventricle', 'Right-Inf-Lat-Vent',
       'Right-Cerebellum-White-Matter', 'Right-Cerebellum-Cortex',
       'Right-Thalamus-Proper', 'Right-Caudate', 'Right-Putamen',
       'Right-Pallidum', 'Right-Hippocampus', 'Right-Amygdala',
       'Right-Accumbens-area', 'Right-VentralDC', 'Right-vessel',
       'Right-choroid-plexus', '5th-Ventricle', 'WM-hypointensities',
       'Left-WM-hypointensities', 'Right-WM-hypointensities',
       'non-WM-hypointensities', 'Left-non-WM-hypointensities',
       'Right-non-WM-hypointensities', 'Optic-Chiasm', 'CC_Posterior',
       'CC_Mid_Posterior', 'CC_Central', 'CC_Mid_Anterior', 'CC_Anterior',
       'BrainSegVol', 'BrainSegVolNotVent', 'VentricleChoroidVol',
       'lhCortexVol', 'rhCortexVol', 'CortexVol', 'lhCerebralWhiteMatterVol',
       'rhCerebralWhiteMatterVol', 'CerebralWhiteMatter', 'SubCortGrayVol',
       'TotalGrayVol', 'SupraTentorialVol', 'SupraTentorialVol.1', 'MaskVol',
       'BrainSegVol-to-eTIV', 'MaskVol-to-eTIV', 'lhSurfaceHoles',
       'SurfaceHoles', 'EstimatedTotalIntraCranialVol',
       'SupraTentorialVolNotVent', 'rhSurfaceHoles', 'Age'],
      dtype='object')
```

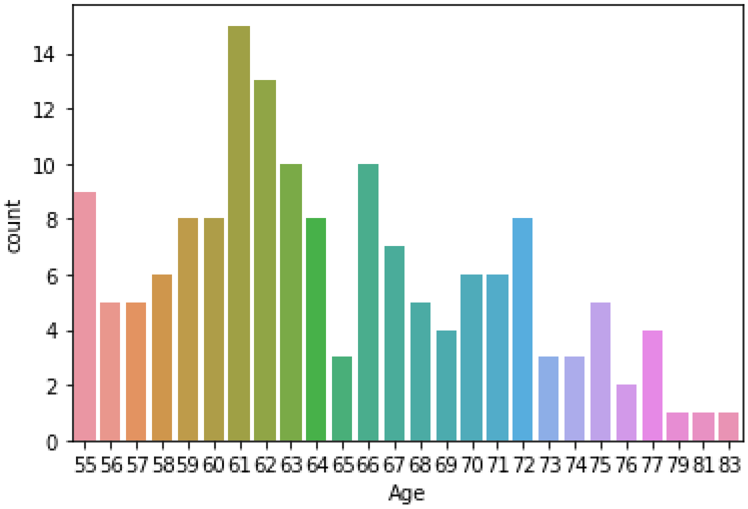
In [5]:
```
y = df.Age
list = [ 'Age', 'Gender']
x = df.drop(list, axis=1)
x.head()
```

Out[5]:

	Left-Hippocampus	Left-Amygdala	CSF	Left-Accumbens-area	Left-VentralDC	Left-vessel	Left-choroid-plexus	Right-Lateral-Ventricle	Righ Inf-La Ve
0	2.926839	1.367249	0.727630	0.425195	2.710304	0.025737	0.204519	8.281764	0.40027
1	3.001051	1.509614	0.773824	0.434145	2.905637	0.027376	0.199077	6.947701	0.52276
2	3.245917	1.161810	1.106959	0.457489	2.938934	0.034745	0.396917	7.000642	0.36040
3	2.612891	1.072513	0.726905	0.330019	2.624964	0.049685	0.193368	11.501124	0.82276
4	3.102019	1.428946	0.860813	0.435614	2.770801	0.038458	0.308704	5.114321	0.49987

5 rows × 55 columns

In [6]:
```
ax = sns.countplot(y,label="Count")
```



In [7]:
```
x.describe()
```

Out[7]:

	Left-Hippocampus	Left-Amygdala	CSF	Left-Accumbens-area	Left-VentralDC	Left-vessel	Left-choroid-plexus	R Lat Ven
count	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.000000	156.00
mean	2.583331	1.063702	0.854409	0.368432	2.457840	0.035732	0.232246	9.18
std	0.399179	0.184307	0.139047	0.082408	0.272450	0.024276	0.064789	3.79
min	0.142967	0.601452	0.531626	0.142119	1.796788	0.001863	0.109082	3.27
25%	2.359207	0.943053	0.754097	0.313678	2.281328	0.023206	0.190805	6.44
50%	2.572450	1.063956	0.837804	0.361898	2.436119	0.031584	0.220110	8.29
75%	2.781072	1.163174	0.921226	0.425371	2.621530	0.044020	0.265365	11.34
max	3.565410	1.658588	1.291225	0.624529	3.266108	0.252521	0.561918	23.17

8 rows × 55 columns

In [8]:
```
x.shape
```

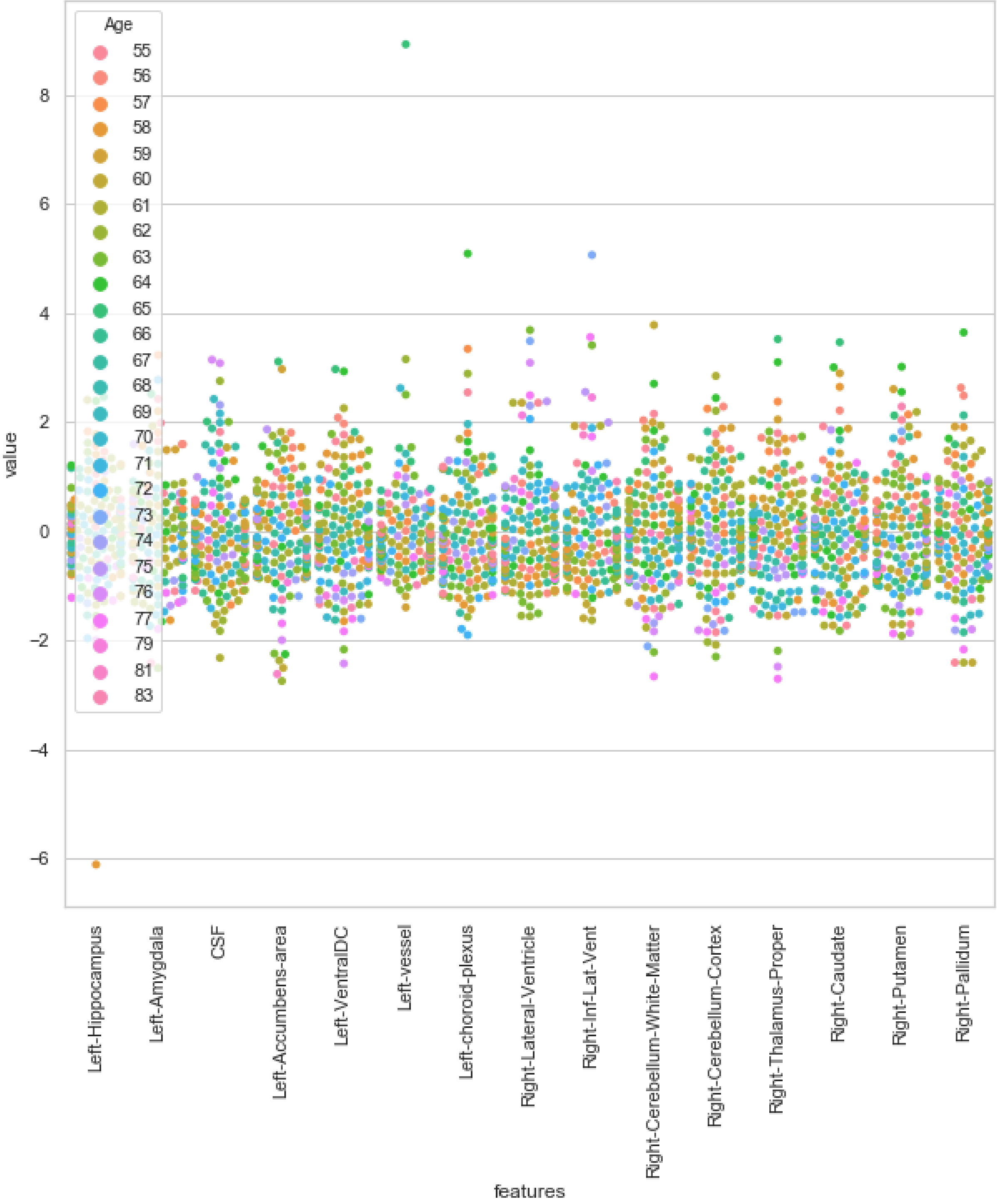
Out[8]:
```
(156, 55)
```

In [9]:
```
import time

sns.set(style="whitegrid", palette="muted")
data_dia = y
data = x
data_n_2 = (data - data.mean()) / (data.std())              # standardization
data = pd.concat([y,data_n_2.iloc[:,0:15]],axis=1)
data = pd.melt(data,id_vars="Age",
                    var_name="features",
                    value_name='value')
```

```
plt.figure(figsize=(10,10))
tic = time.time()
sns.swarmplot(x="features", y="value", hue="Age", data=data)

plt.xticks(rotation=90)
```

Out[9]:

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14]),
 <a list of 15 Text xticklabel objects>)
```



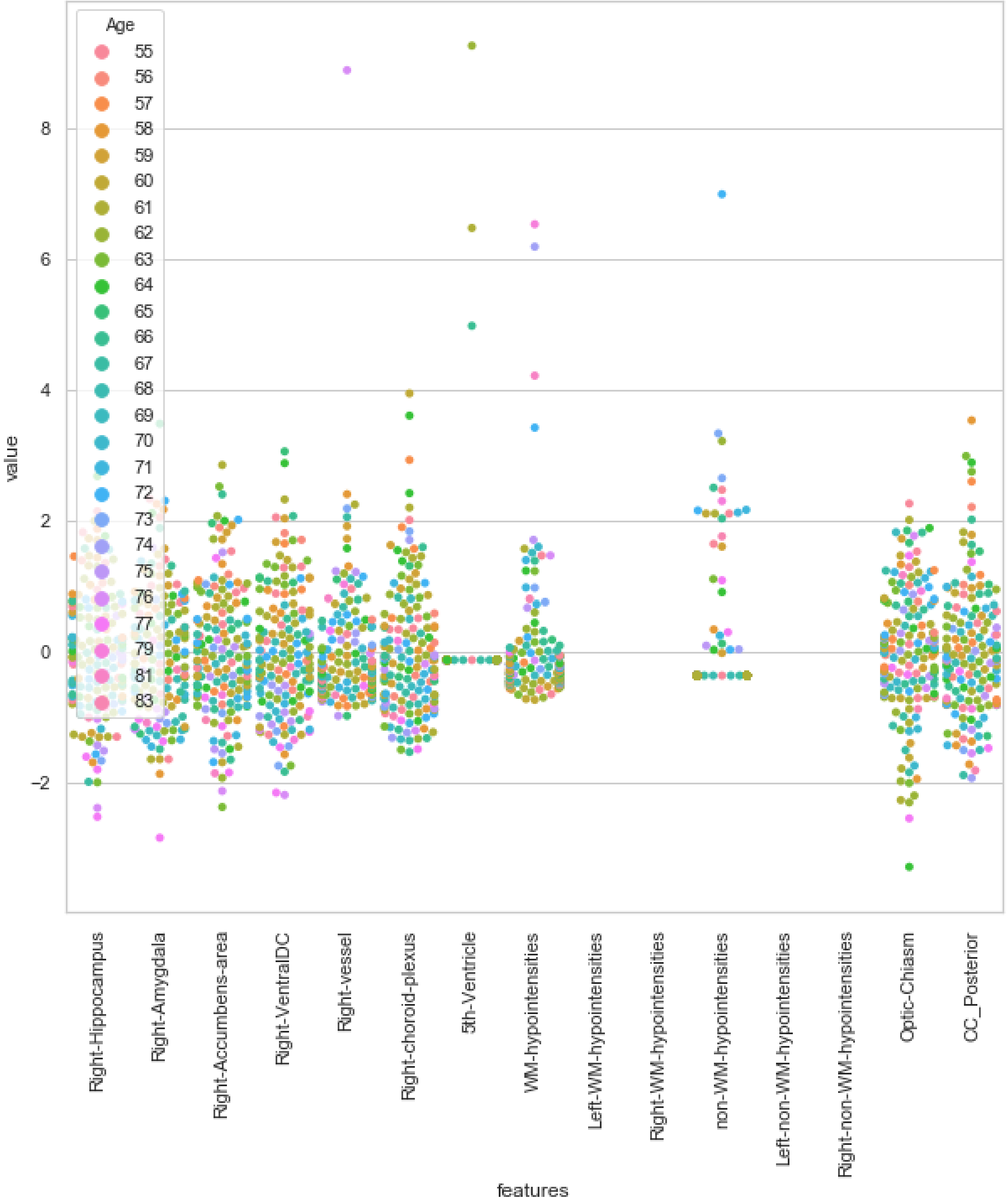
In [10]:

```
data = pd.concat([y,data_n_2.iloc[:,15:30]],axis=1)
data = pd.melt(data,id_vars="Age",
                    var_name="features",
                    value_name='value')
plt.figure(figsize=(10,10))
```

```
sns.swarmplot(x="features", y="value", hue="Age", data=data)
plt.xticks(rotation=90)
```

Out[10]:

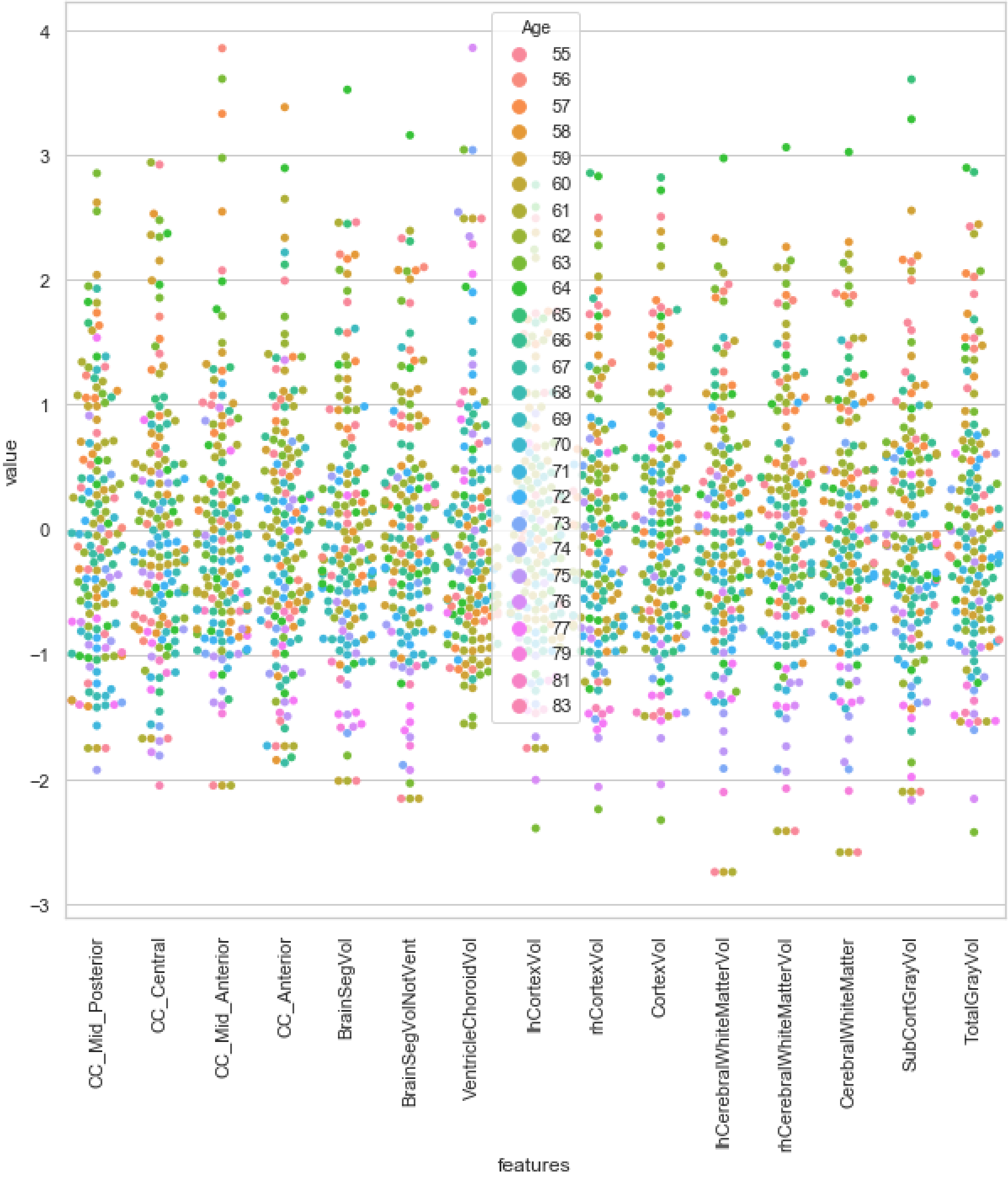
```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14]),
 <a list of 15 Text xticklabel objects>)
```



In [11]:

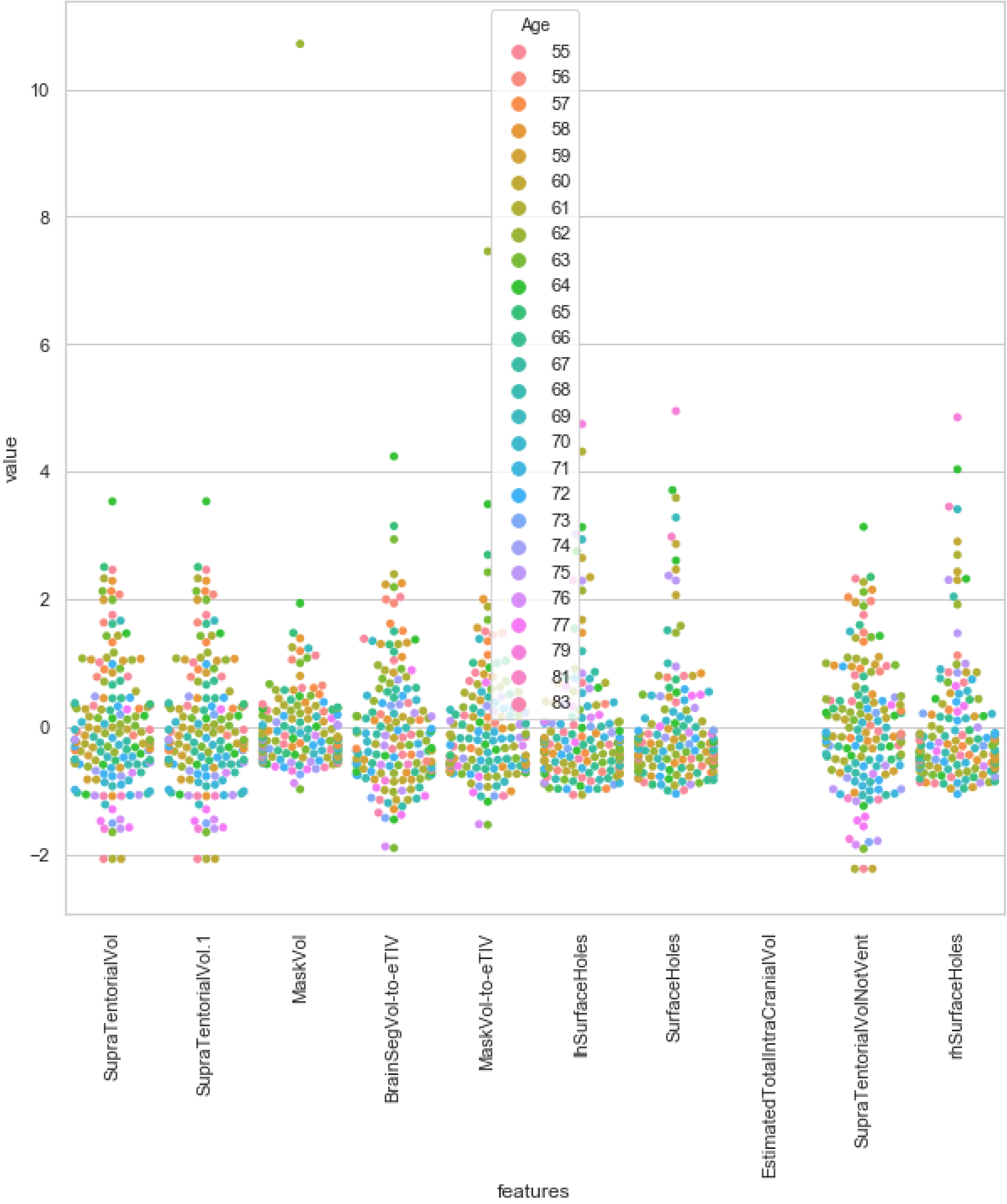
```
data = pd.concat([y,data_n_2.iloc[:,30:45]],axis=1)
data = pd.melt(data,id_vars="Age",
                    var_name="features",
                    value_name='value')
plt.figure(figsize=(10,10))
sns.swarmplot(x="features", y="value", hue="Age", data=data)
plt.xticks(rotation=90)
```

Out[11]:
```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14]),
 <a list of 15 Text xticklabel objects>)
```



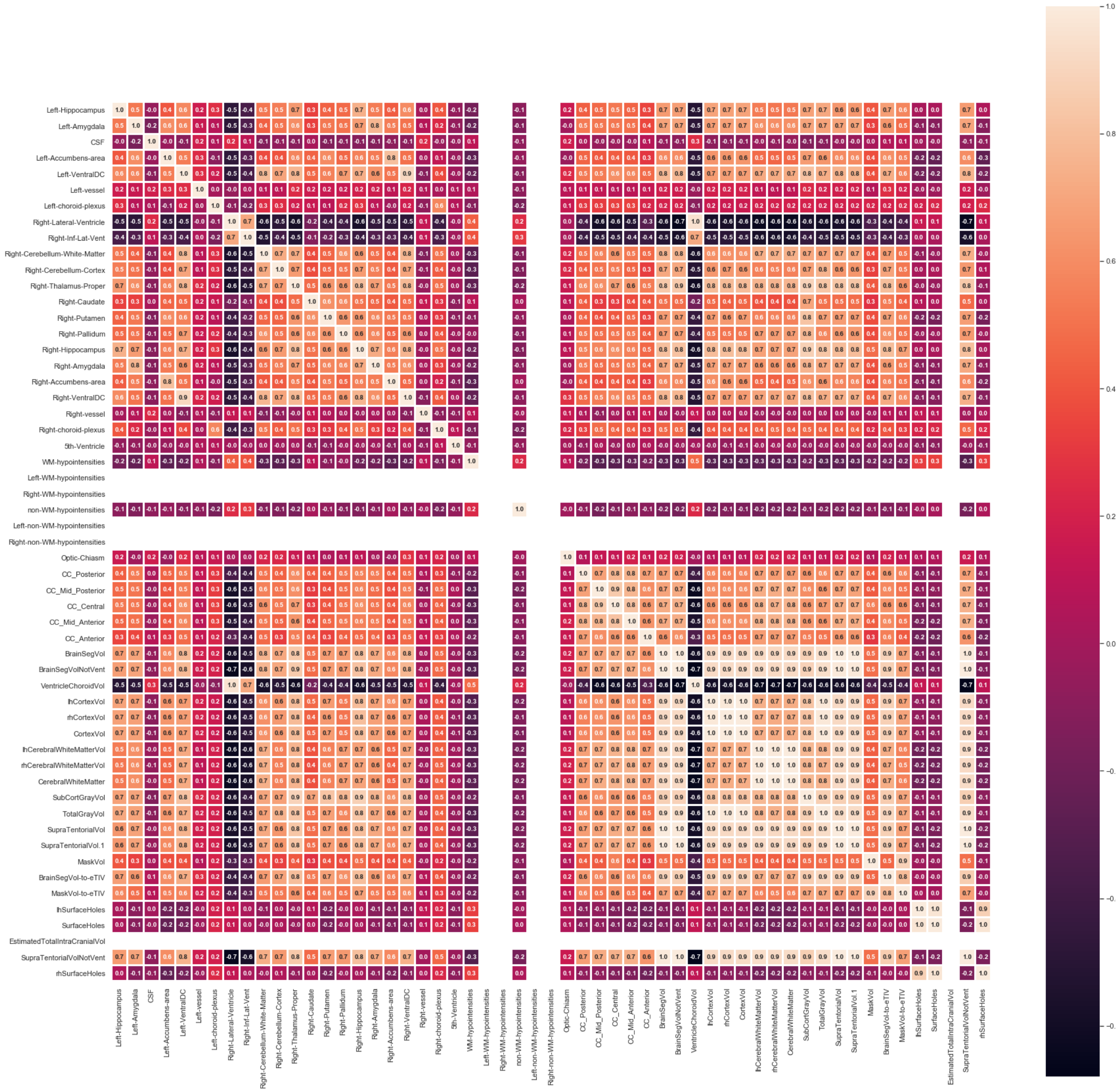
In [12]:
```
data = pd.concat([y,data_n_2.iloc[:,45:65]],axis=1)
data = pd.melt(data,id_vars="Age",
                    var_name="features",
                    value_name='value')
plt.figure(figsize=(10,10))
sns.swarmplot(x="features", y="value", hue="Age", data=data)
toc = time.time()
plt.xticks(rotation=90)
print("swarm plot time: ", toc-tic ," s")
```

```
swarm plot time:  56.35338830947876  s
```



In [39]:

```
#correlation map
f,ax = plt.subplots(figsize=(30, 30))
sns_plot = sns.heatmap(x.corr(), annot=True, linewidths=2, square =True, fmt= '.1f',ax
```


Left-Hippocampus
Left-Amygdala
CSF
Left-Accumbens-area
Left-VentralDC
Left-vessel
Left-choroid-plexus
Right-Lateral-Ventricle
Right-Inf-Lat-Vent
Right-Cerebellum-White-Matter
Right-Cerebellum-Cortex
Right-Thalamus-Proper
Right-Caudate
Right-Putamen
Right-Pallidum
Right-Hippocampus
Right-Amygdala
Right-Accumbens-area
Right-VentralDC
Right-vessel
Right-choroid-plexus
5th-Ventricle
WM-hypointensities
Left-WM-hypointensities
Right-WM-hypointensities
non-WM-hypointensities
Left-non-WM-hypointensities
Right-non-WM-hypointensities
Optic-Chiasm
CC_Posterior
CC_Mid_Posterior
CC_Central
CC_Mid_Anterior
CC_Anterior
BrainSegVol
BrainSegVolNotVent
VentricleChoroidVol
lhCortexVol
rhCortexVol
CortexVol
lhCerebralWhiteMatterVol
rhCerebralWhiteMatterVol
CerebralWhiteMatter
SubCortGrayVol
TotalGrayVol
SupraTentorialVol
SupraTentorialVol.1
MaskVol
BrainSegVol-to-eTIV
MaskVol-to-eTIV
lhSurfaceHoles
SurfaceHoles
EstimatedTotalIntraCranialVol
SupraTentorialVolNotVent
rhSurfaceHoles
Left-Hippocampus
Left-Amygdala
CSF
Left-Accumbens-area
Left-VentralDC
Left-vessel
Left-choroid-plexus
Right-Lateral-Ventricle
Right-Inf-Lat-Vent
Right-Cerebellum-White-Matter
Right-Cerebellum-Cortex
Right-Thalamus-Proper
Right-Caudate
Right-Putamen
Right-Pallidum
Right-Hippocampus
Right-Amygdala
Right-Accumbens-area
Right-VentralDC
Right-vessel
Right-choroid-plexus
5th-Ventricle
WM-hypointensities
Left-WM-hypointensities
Right-WM-hypointensities
non-WM-hypointensities
Left-non-WM-hypointensities
Right-non-WM-hypointensities
Optic-Chiasm
CC_Posterior
CC_Mid_Posterior
CC_Central
CC_Mid_Anterior
CC_Anterior
BrainSegVol
BrainSegVolNotVent
VentricleChoroidVol
lhCortexVol
rhCortexVol
CortexVol
lhCerebralWhiteMatterVol
rhCerebralWhiteMatterVol
CerebralWhiteMatter
SubCortGrayVol
TotalGrayVol
SupraTentorialVol
SupraTentorialVol.1
MaskVol
BrainSegVol-to-eTIV
MaskVol-to-eTIV
lhSurfaceHoles
SurfaceHoles
EstimatedTotalIntraCranialVol
SupraTentorialVolNotVent
rhSurfaceHoles
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6