In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
data = pd.read_csv("heart_desease.csv")
```

In [3]:

```
data.head()
```

Out[3]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalki
0	No	16.60	Yes	No	No	3	30	
1	No	20.34	No	No	Yes	0	0	
2	No	26.58	Yes	No	No	20	30	
3	No	24.21	No	No	No	0	0	
4	No	23.71	No	No	No	28	0	Y

In [4]:

```
data.tail()
```

Out[4]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	Dif
319790	Yes	27.41	Yes	No	No	7	0	
319791	No	29.84	Yes	No	No	0	0	
319792	No	24.24	No	No	No	0	0	
319793	No	32.81	No	No	No	0	0	
319794	No	46.56	No	No	No	0	0	

In [5]:

```
data.shape
```

Out[5]:

```
(319795, 18)
```

In [6]:

```
data.columns
```

Out[6]:

```
Index(['HeartDisease', 'BMI', 'Smoking', 'AlcoholDrinking', 'Stroke',
       'PhysicalHealth', 'MentalHealth', 'DiffWalking', 'Sex', 'AgeCategor
y',
       'Race', 'Diabetic', 'PhysicalActivity', 'GenHealth', 'SleepTime',
       'Asthma', 'KidneyDisease', 'SkinCancer'],
      dtype='object')
```

In [7]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 319795 entries, 0 to 319794
Data columns (total 18 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   HeartDisease      319795 non-null  object
 1   BMI               319795 non-null  float64
 2   Smoking           319795 non-null  object
 3   AlcoholDrinking   319795 non-null  object
 4   Stroke            319795 non-null  object
 5   PhysicalHealth    319795 non-null  int64
 6   MentalHealth      319795 non-null  int64
 7   DiffWalking       319795 non-null  object
 8   Sex               319795 non-null  object
 9   AgeCategory       319795 non-null  object
 10  Race              319795 non-null  object
 11  Diabetic          319795 non-null  object
 12  PhysicalActivity  319795 non-null  object
 13  GenHealth         319795 non-null  object
 14  SleepTime         319795 non-null  int64
 15  Asthma            319795 non-null  object
 16  KidneyDisease     319795 non-null  object
 17  SkinCancer        319795 non-null  object
dtypes: float64(1), int64(3), object(14)
memory usage: 43.9+ MB
```

In [8]:

```
data.describe()
```

Out[8]:

	BMI	PhysicalHealth	MentalHealth	SleepTime
count	319795.000000	319795.00000	319795.000000	319795.000000
mean	28.325399	3.37171	3.898366	7.097075
std	6.356100	7.95085	7.955235	1.436007
min	12.020000	0.00000	0.000000	1.000000
25%	24.030000	0.00000	0.000000	6.000000
50%	27.340000	0.00000	0.000000	7.000000
75%	31.420000	2.00000	3.000000	8.000000
max	94.850000	30.00000	30.000000	24.000000

In [9]:

```
data.isnull().sum()
```

Out[9]:

```
HeartDisease        0
BMI                 0
Smoking             0
AlcoholDrinking     0
Stroke              0
PhysicalHealth      0
MentalHealth        0
DiffWalking         0
Sex                 0
AgeCategory         0
Race                0
Diabetic            0
PhysicalActivity    0
GenHealth           0
SleepTime           0
Asthma              0
KidneyDisease       0
SkinCancer          0
dtype: int64
```

In [11]:

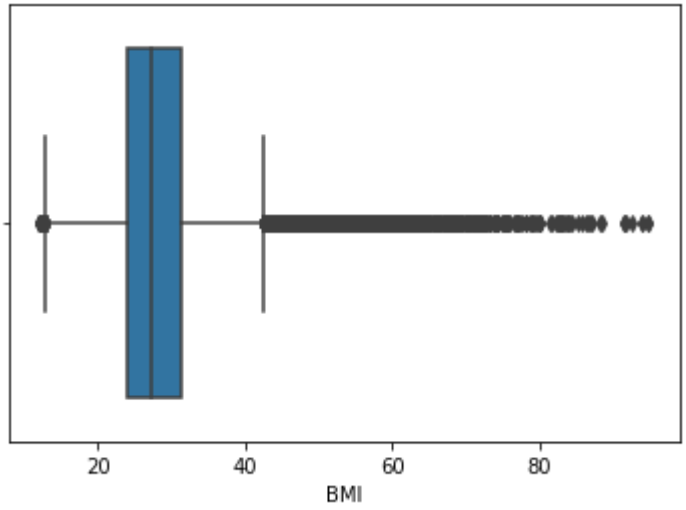
```
data_cat = data[['HeartDisease', 'Smoking', 'AlcoholDrinking', 'Stroke', 'DiffWalking',
                 'Sex', 'AgeCategory', 'Race', 'Diabetic', 'PhysicalActivity', 'GenHealth
                 'Asthma', 'KidneyDisease', 'SkinCancer']]
```

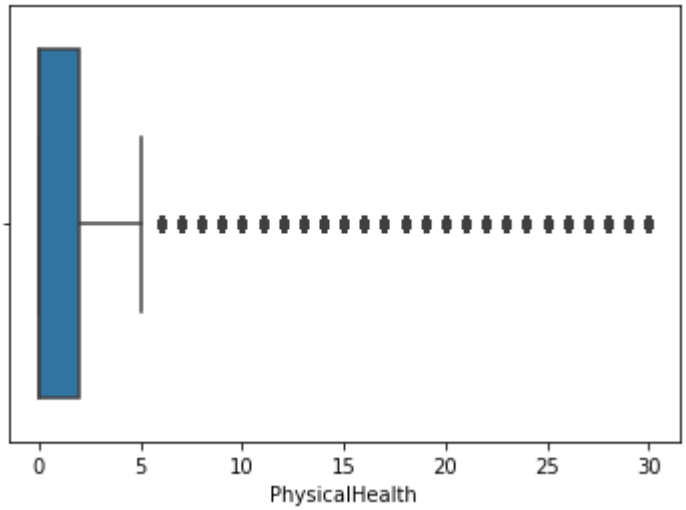
In [12]:

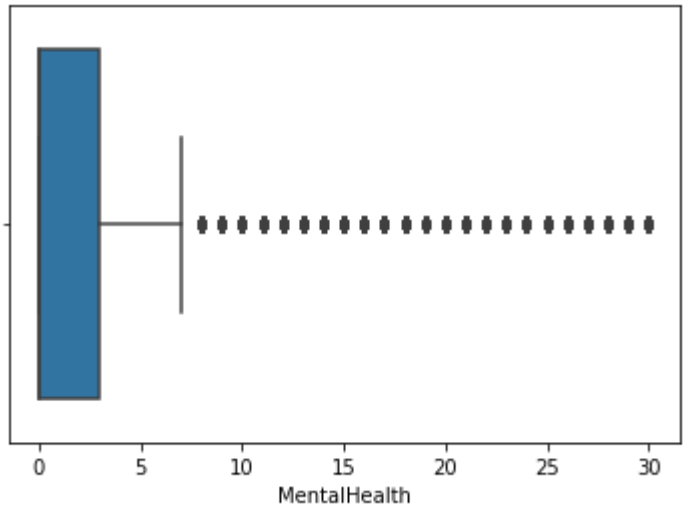
```
data_num = data[['BMI', 'PhysicalHealth', 'MentalHealth', 'SleepTime']]
```

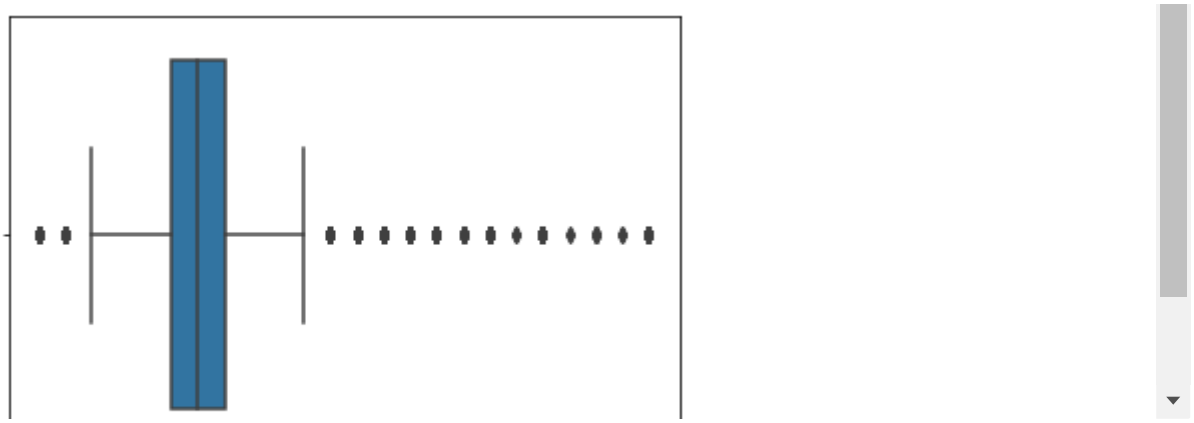
In [13]:

```
for i in data_num.columns:
    sns.boxplot(x=data_num[i])
    plt.show()
plt.show()
```



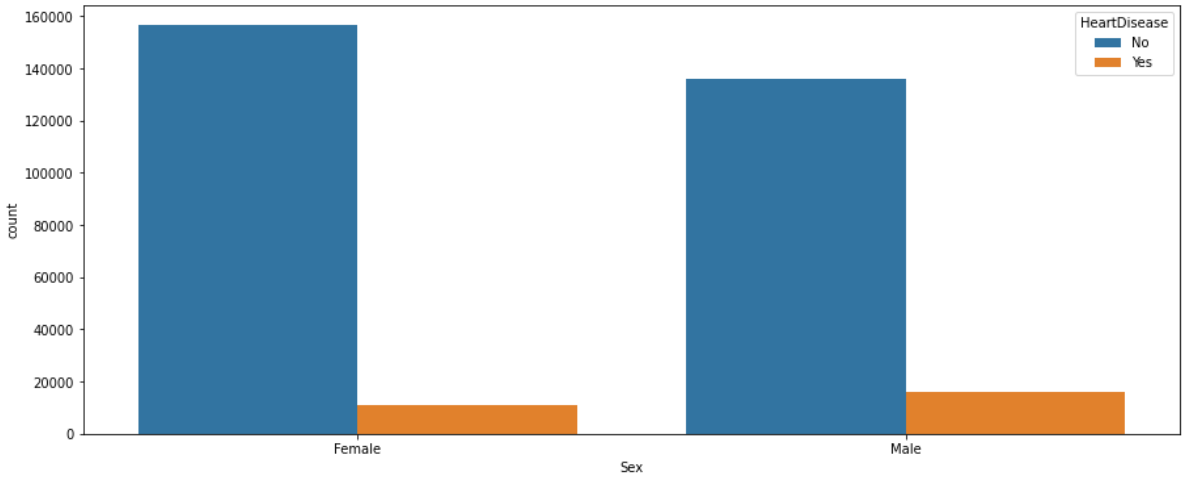






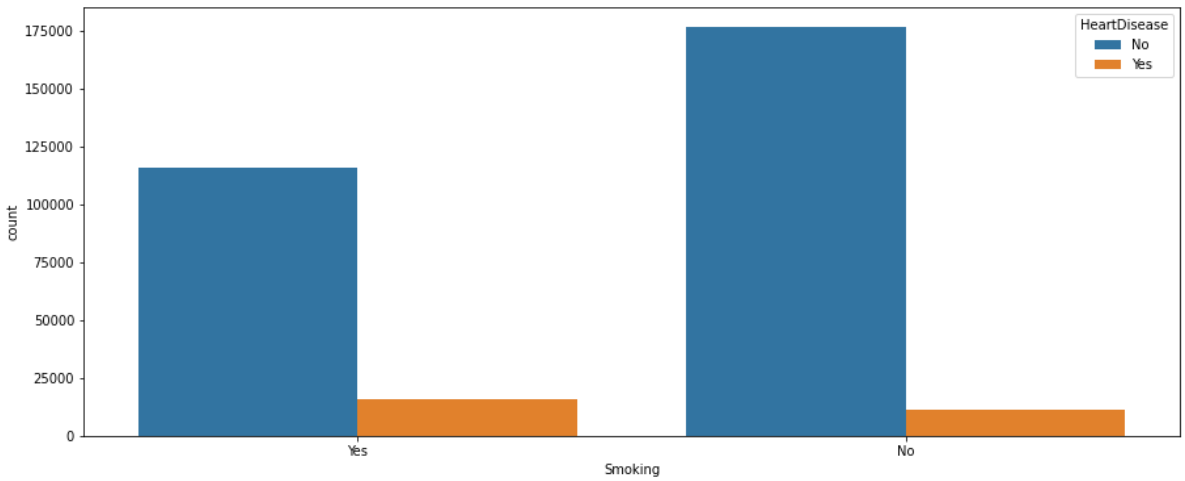
In [27]:

```
plt.figure(figsize=(15,6))
sns.countplot('Sex',hue='HeartDisease', data = data)
plt.xticks(rotation = 0)
plt.show()
```



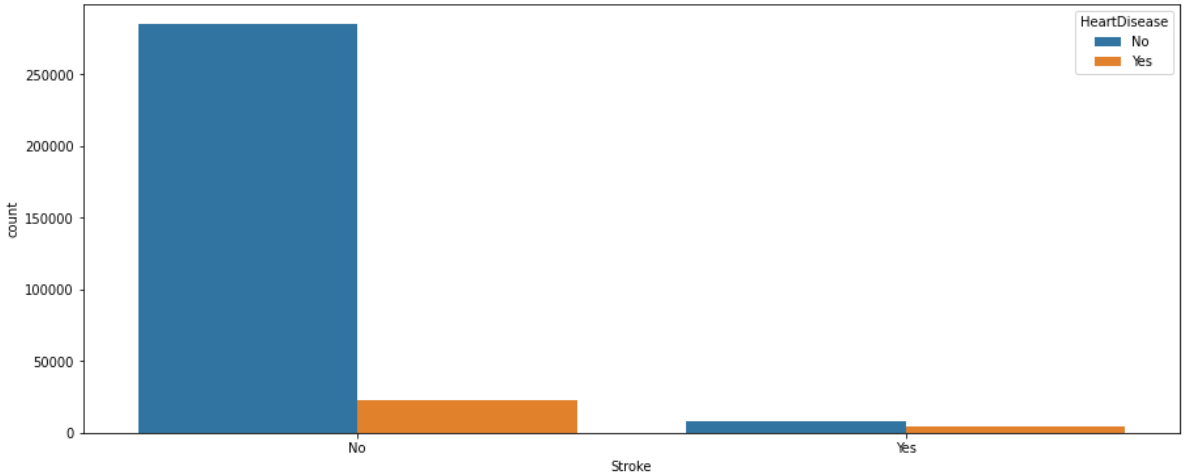
In [26]:

```
plt.figure(figsize=(15,6))
sns.countplot('Smoking',hue='HeartDisease',data=data)
plt.xticks(rotation = 0)
plt.show()
```



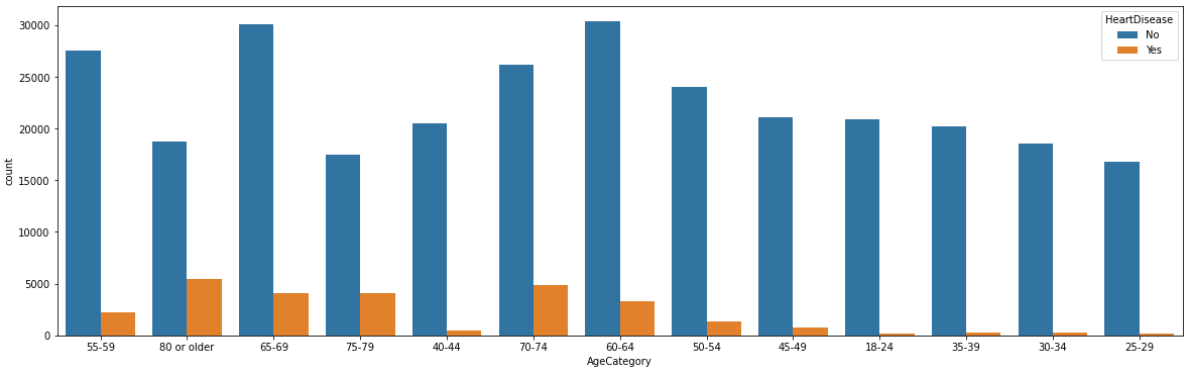
In [25]:

```
plt.figure(figsize=(15,6))
sns.countplot('Stroke',hue='HeartDisease',data=data)
plt.xticks(rotation = 0)
plt.show()
```



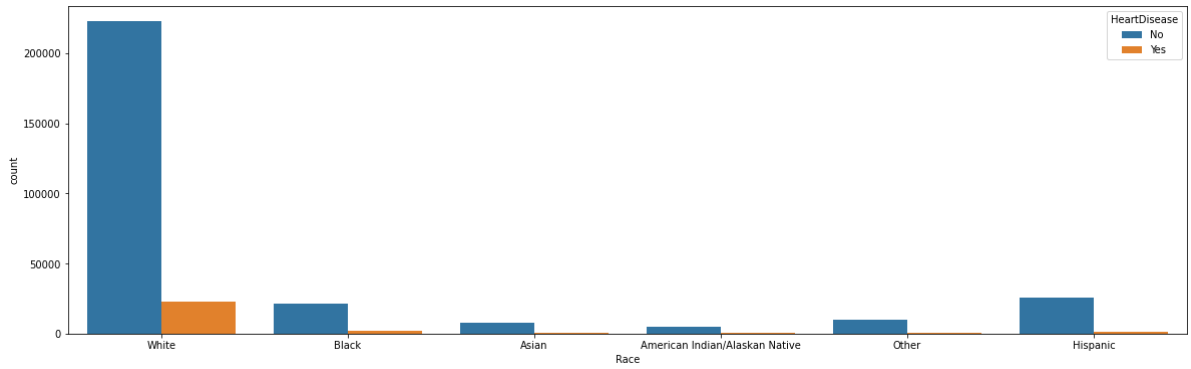
In [23]:

```
plt.figure(figsize=(20,6))
sns.countplot('AgeCategory',hue='HeartDisease',data=data)
plt.xticks(rotation = 0)
plt.show()
```



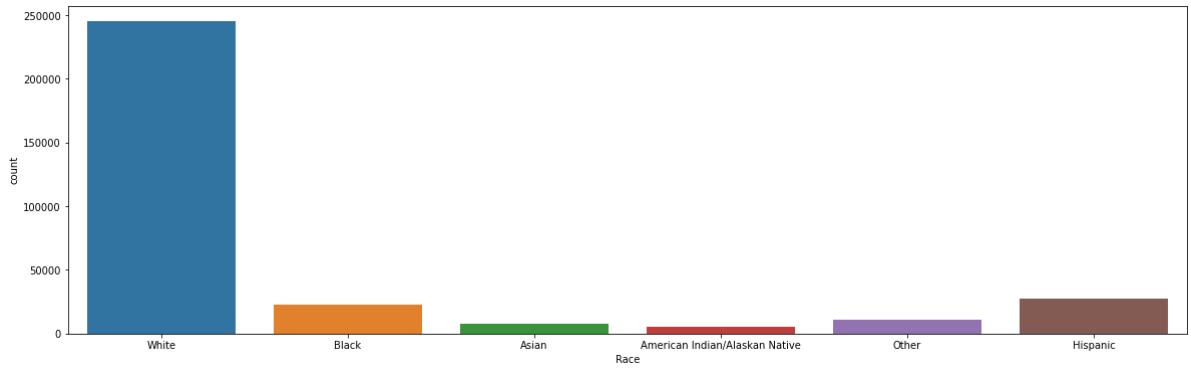
In [22]:

```
plt.figure(figsize=(20,6))
sns.countplot('Race',hue='HeartDisease',data=data)
plt.xticks(rotation = 0)
plt.show()
```



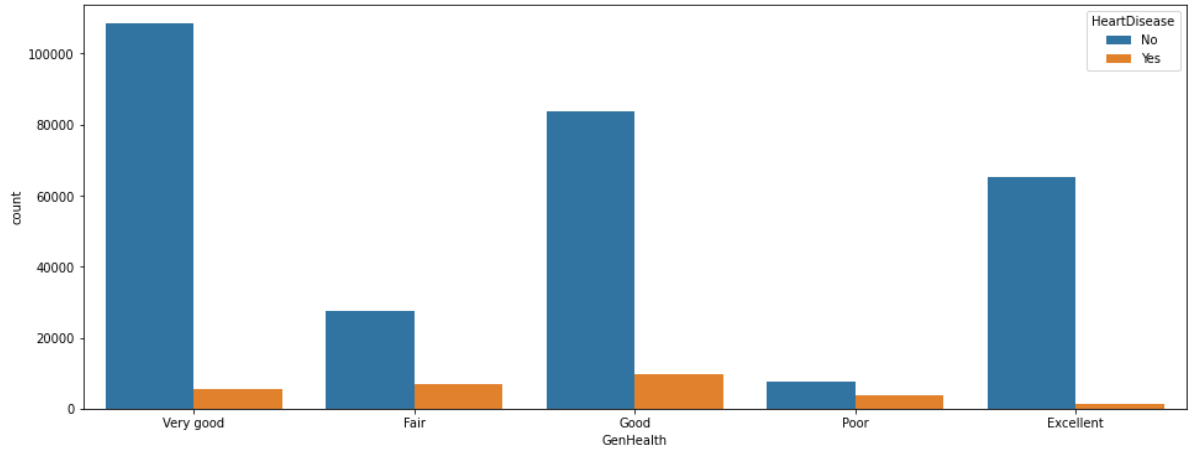
In [28]:

```
plt.figure(figsize=(20,6))
sns.countplot('Race',data=data)
plt.xticks(rotation = 0)
plt.show()
```



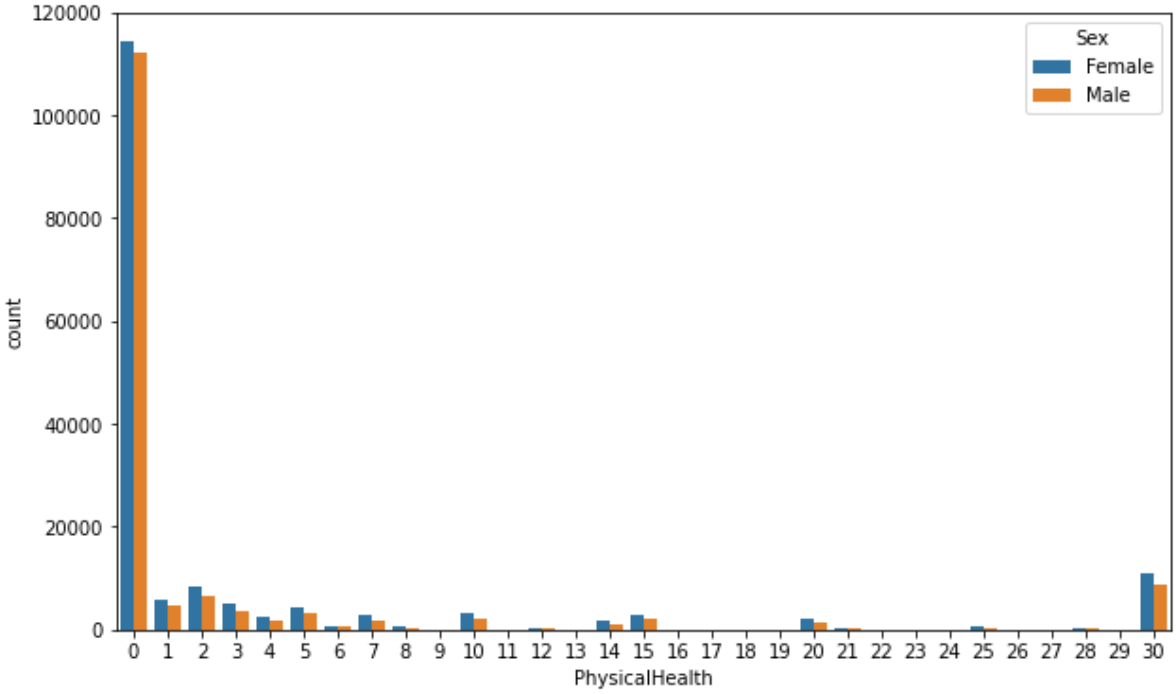
In [29]:

```
plt.figure(figsize=(16,6))
sns.countplot('GenHealth',hue='HeartDisease',data=data)
plt.xticks(rotation = 0)
plt.show()
```



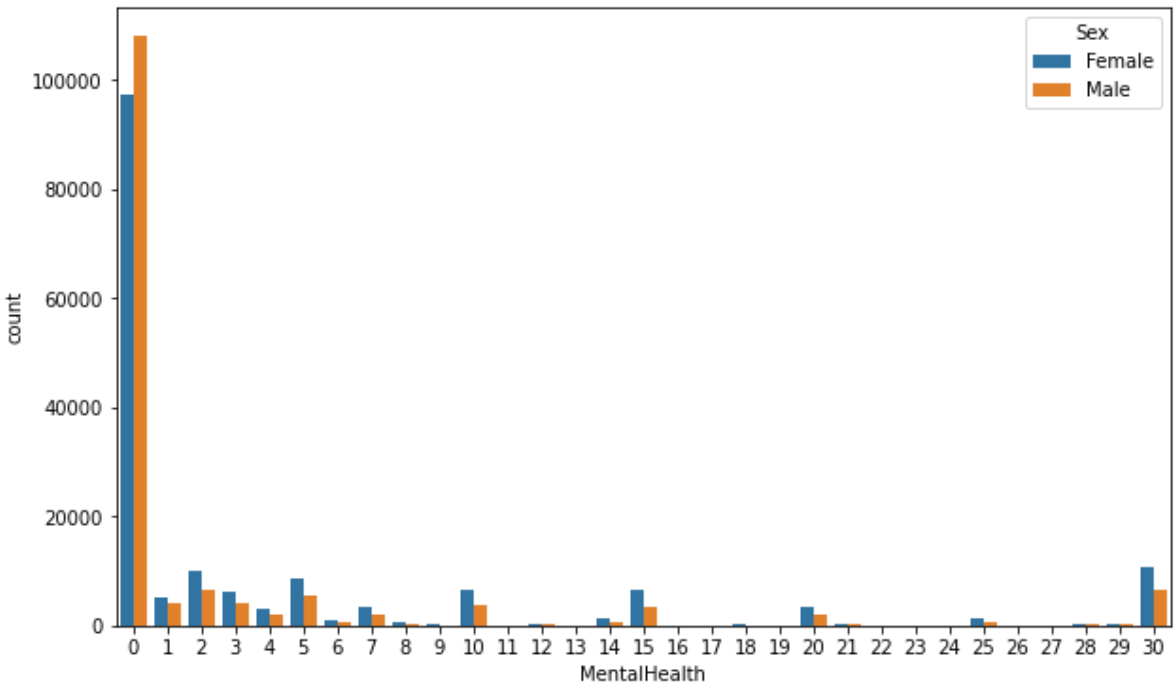
In [34]:

```
plt.figure(figsize=(10,6))
sns.countplot('PhysicalHealth',hue='Sex',data=data)
plt.xticks(rotation = 0)
plt.show()
```



In [35]:

```
plt.figure(figsize=(10,6))
sns.countplot('MentalHealth',hue='Sex',data=data)
plt.xticks(rotation = 0)
plt.show()
```



In [37]:

```
from sklearn.preprocessing import LabelEncoder
```

In [38]:

```
for i in data_cat.columns:
    le=LabelEncoder()
    label=le.fit_transform(data_cat[i])
    data_cat[i]=label
```

In [39]:

```
data_cat.head()
```

Out[39]:

	HeartDisease	Smoking	AlcoholDrinking	Stroke	DiffWalking	Sex	AgeCategory	Race	Diabeti
0	0	1	0	0	0	0	7	5	
1	0	0	0	1	0	0	12	5	
2	0	1	0	0	0	1	9	5	
3	0	0	0	0	0	0	11	5	
4	0	0	0	0	1	0	4	5	

In [40]:

```
data1=pd.concat([data_cat,data_num],axis=1)
```

In [41]:

```
data1.head()
```

Out[41]:

	HeartDisease	Smoking	AlcoholDrinking	Stroke	DiffWalking	Sex	AgeCategory	Race	Diabeti
0	0	1	0	0	0	0	7	5	
1	0	0	0	1	0	0	12	5	
2	0	1	0	0	0	1	9	5	
3	0	0	0	0	0	0	11	5	
4	0	0	0	0	1	0	4	5	

In [42]:

```
data1.corr()
```

Out[42]:

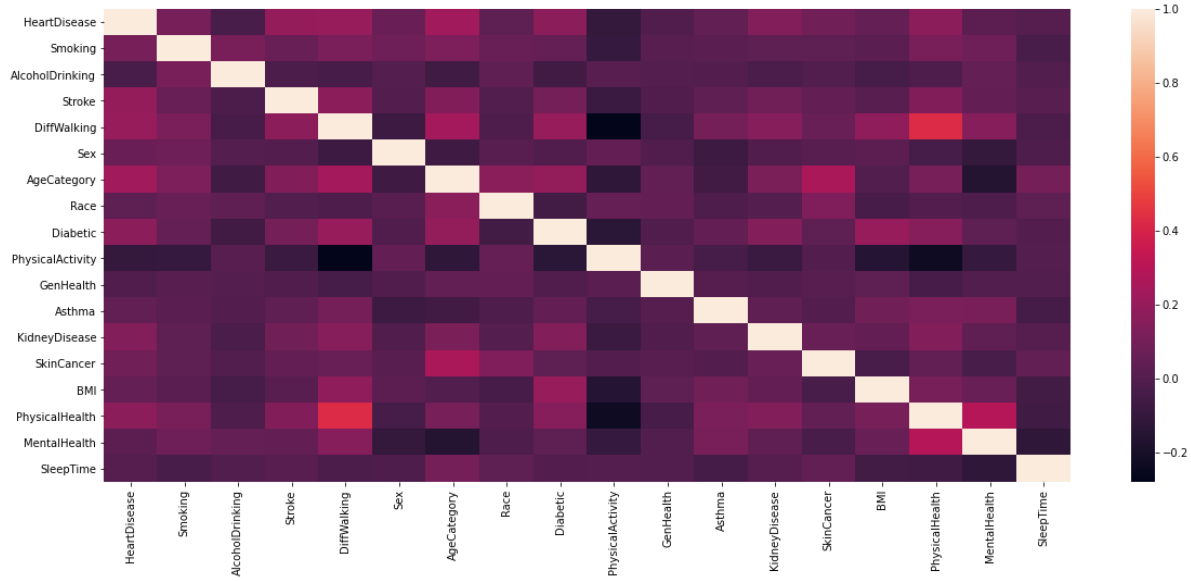
	HeartDisease	Smoking	AlcoholDrinking	Stroke	DiffWalking	Sex	Age(
HeartDisease	1.000000	0.107764	-0.032080	0.196835	0.201258	0.070040	(
Smoking	0.107764	1.000000	0.111768	0.061226	0.120074	0.085052	(
AlcoholDrinking	-0.032080	0.111768	1.000000	-0.019858	-0.035328	0.004200	-(
Stroke	0.196835	0.061226	-0.019858	1.000000	0.174143	-0.003091	(
DiffWalking	0.201258	0.120074	-0.035328	0.174143	1.000000	-0.068860	(
Sex	0.070040	0.085052	0.004200	-0.003091	-0.068860	1.000000	-(
AgeCategory	0.233432	0.128331	-0.059528	0.137822	0.243263	-0.067478	'
Race	0.034854	0.065499	0.036702	-0.003956	-0.015831	0.018855	(
Diabetic	0.168553	0.053847	-0.057372	0.101518	0.205502	-0.013456	(
PhysicalActivity	-0.100030	-0.097174	0.017487	-0.079455	-0.278524	0.048247	-(
GenHealth	-0.011062	0.020625	0.001629	-0.009335	-0.043552	-0.010283	(
Asthma	0.041444	0.024149	-0.002202	0.038866	0.103222	-0.069191	-(
KidneyDisease	0.145197	0.034920	-0.028280	0.091167	0.153064	-0.009084	(
SkinCancer	0.093317	0.033977	-0.005702	0.048116	0.064840	0.013434	(
BMI	0.051803	0.023118	-0.038816	0.019733	0.181678	0.026940	-(
PhysicalHealth	0.170721	0.115352	-0.017254	0.137014	0.428373	-0.040904	(
MentalHealth	0.028591	0.085157	0.051282	0.046467	0.152235	-0.100058	-(
SleepTime	0.008327	-0.030336	-0.005065	0.011900	-0.022216	-0.015704	(

In [45]:

```
plt.figure(figsize=(20,8))
sns.heatmap(data1.corr())
```

Out[45]:

```
<AxesSubplot:>
```



In [46]:

```
data1.drop(['Race','BMI'],axis=1,inplace=True)
```

In [47]:

```
X=data1.iloc[:,1:]
y=data1.iloc[:,0]
```

In [48]:

```
X.shape
```

Out[48]:

```
(319795, 15)
```

In [49]:

```
y.shape
```

Out[49]:

```
(319795,)
```

In [50]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(X,y,test_size=0.20)
```

In [51]:

```
from sklearn.linear_model import LogisticRegression
```

In [52]:

```
model=LogisticRegression()
model.fit(x_train,y_train)
```

Out[52]:

```
LogisticRegression()
```

In [54]:

```
y_pred = model.predict(x_test)
```

In [55]:

```
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_test, y_test))
```

```
Training Accuracy : 0.9146758079394612
Testing Accuracy : 0.916055598117544
```