Set-up

In [1]:

```
! pip install seaborn==0.9.0
```

```
Requirement already satisfied: seaborn==0.9.0 in c:\users\surface\ananew\lib\site-packages (0.9.0)
Requirement already satisfied: scipy>=0.14.0 in c:\users\surface\ananew\lib\site-packages (from
seaborn==0.9.0) (1.0.0)
Requirement already satisfied: matplotlib>=1.4.3 in c:\users\surface\ananew\lib\site-packages
(from seaborn==0.9.0) (2.1.2)
Requirement already satisfied: pandas>=0.15.2 in c:\users\surface\ananew\lib\site-packages (from
seaborn==0.9.0) (0.22.0)
Requirement already satisfied: numpy>=1.9.3 in c:\users\surface\ananew\lib\site-packages (from
seaborn==0.9.0) (1.17.0)
Requirement already satisfied: six>=1.10 in c:\users\surface\ananew\lib\site-packages (from
matplotlib>=1.4.3->seaborn==0.9.0) (1.12.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\surface\ananew\lib\site-packages
(from matplotlib>=1.4.3->seaborn==0.9.0) (2.6.1)
Requirement already satisfied: pytz in c:\users\surface\ananew\lib\site-packages (from
matplotlib>=1.4.3->seaborn==0.9.0) (2017.3)
Requirement already satisfied: cycler>=0.10 in c:\users\surface\ananew\lib\site-packages (from
matplotlib>=1.4.3->seaborn==0.9.0) (0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in
c:\users\surface\ananew\lib\site-packages (from matplotlib>=1.4.3->seaborn==0.9.0) (2.2.0)
```

```
WARNING: You are using pip version 19.2.1, however version 19.3.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

In [2]:

```
import matplotlib.pyplot as plt
import matplotlib.style as style

import numpy as np
import pandas as pd
import os
import seaborn as sns
```

In [3]:

```
# style.available
style.use('seaborn-poster') #sets the size of the charts
style.use('ggplot')
```

In [11]:

```
# sns.set(style="whitegrid")
```

Data Process

In [4]:

```
df = pd.read_csv('examples/HAM10000_metadata.csv')
df.head(10)
```

Out[4]:

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp

	lesion_id	image_id	dx	dx_type	age	sex	localization
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear
5	HAM_0001466	ISIC_0027850	bkl	histo	75.0	male	ear
6	HAM_0002761	ISIC_0029176	bkl	histo	60.0	male	face
7	HAM_0002761	ISIC_0029068	bkl	histo	60.0	male	face
8	HAM_0005132	ISIC_0025837	bkl	histo	70.0	female	back
9	HAM_0005132	ISIC_0025209	bkl	histo	70.0	female	back

In [5]:

```
df.isnull().sum()
```

Out[5]:

```
lesion_id         0
image_id          0
dx                0
dx_type           0
age              57
sex               0
localization      0
dtype: int64
```

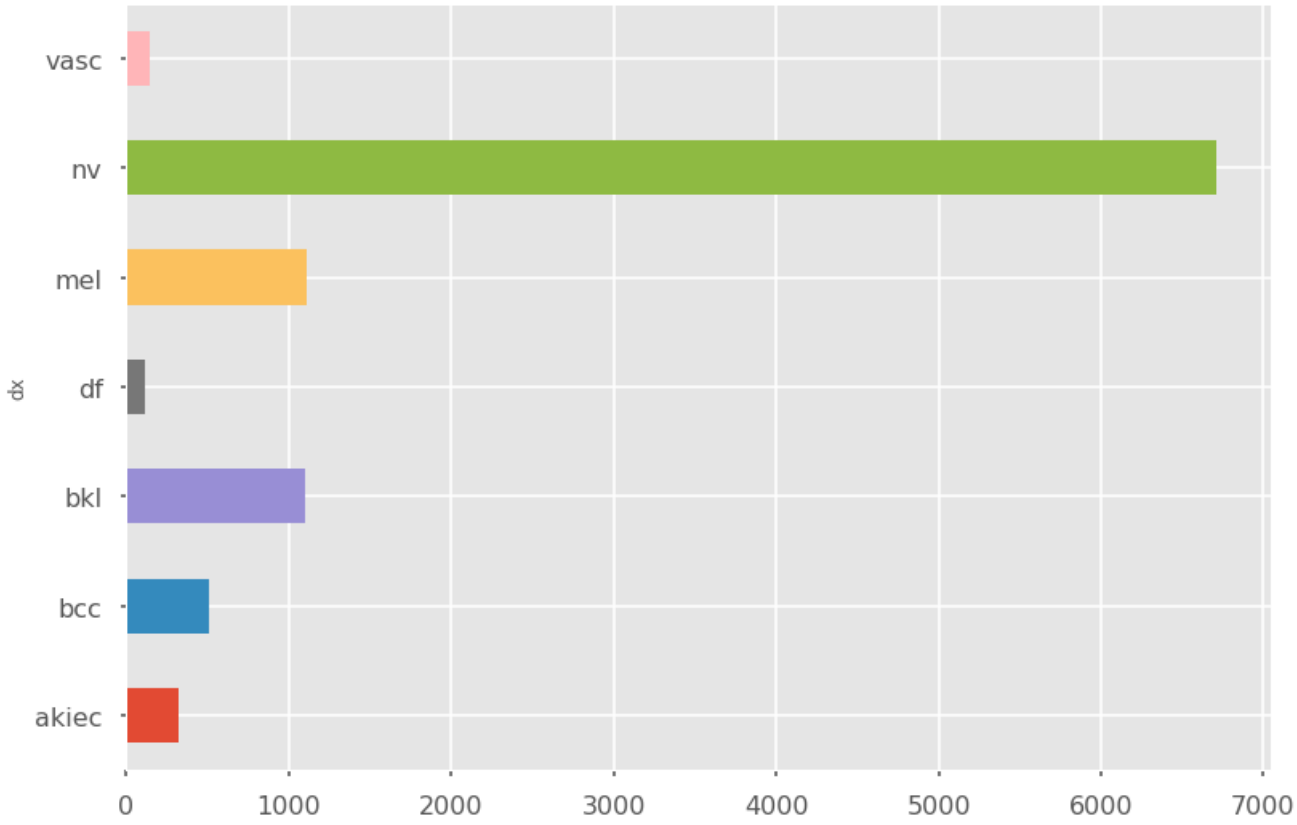
Explanatory Analysis

In [6]:

```
df.groupby('dx').size().plot.barh()
```

Out[6]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x265d1663668>
```

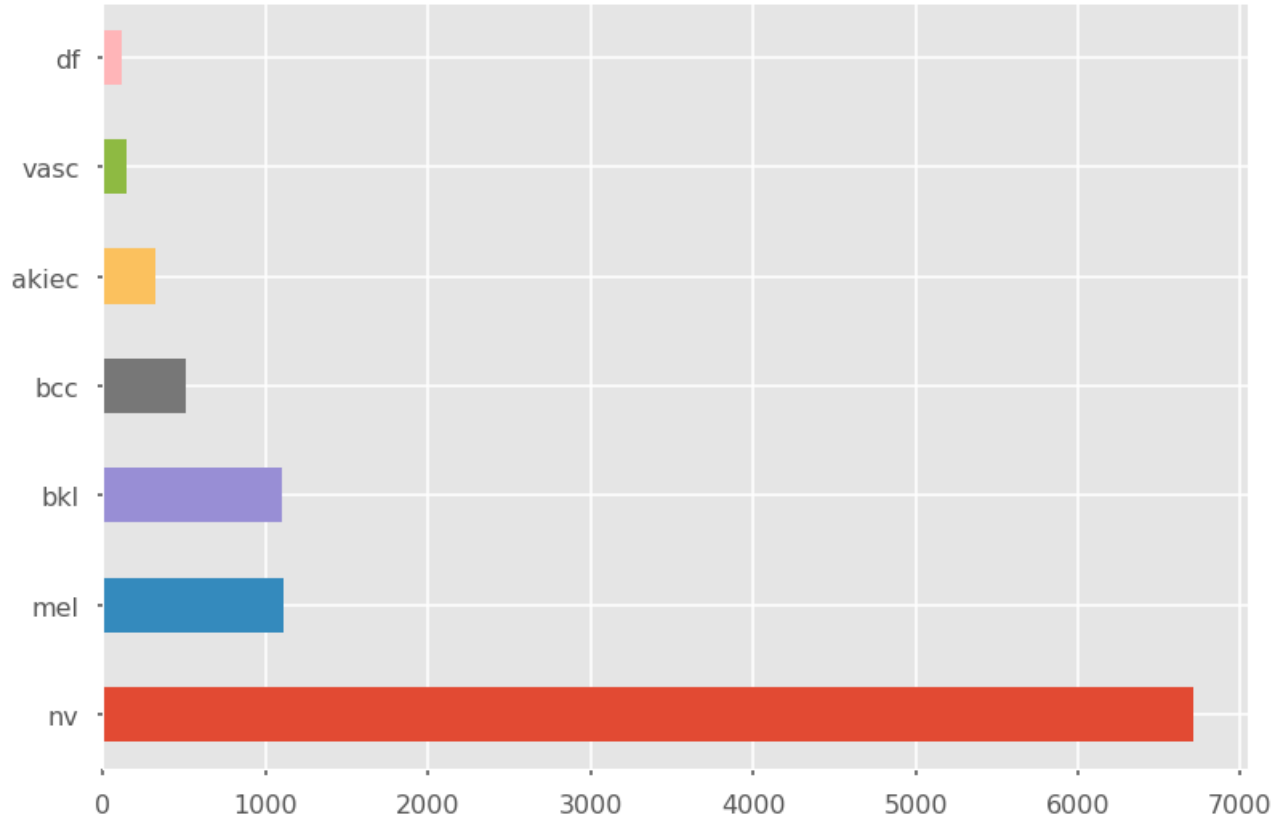


In [19]:

```
df['dx'].value_counts().plot(kind='barh')
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x265d4ae8c50>
```



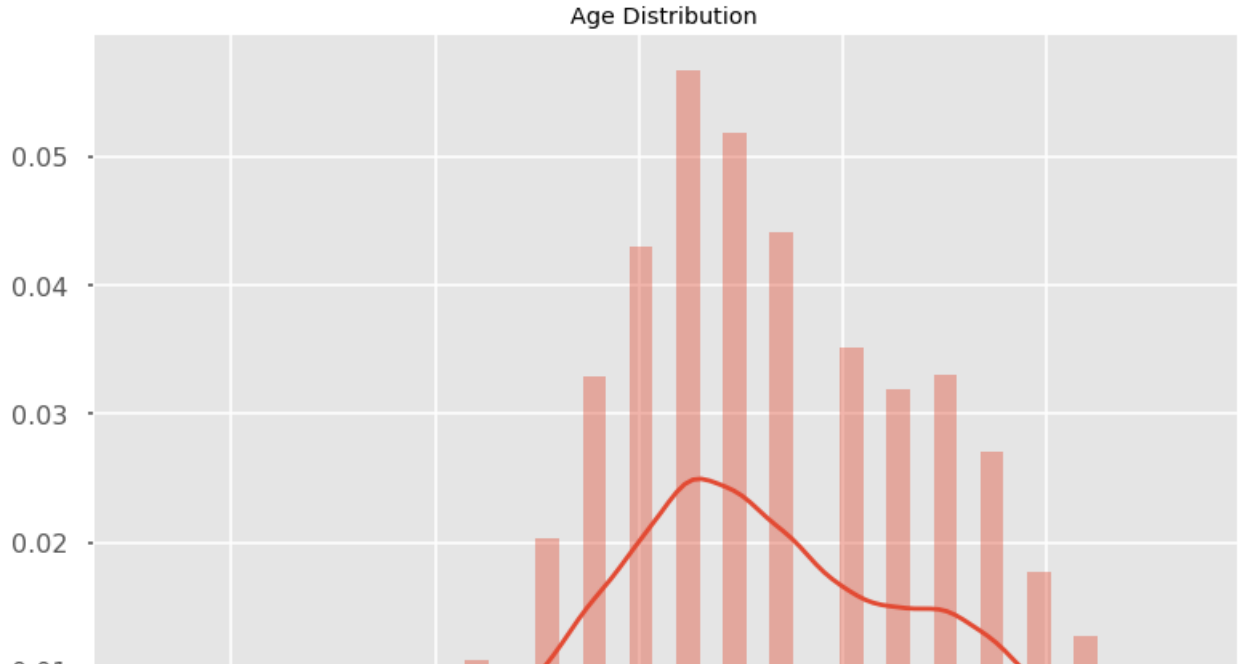
In [18]:

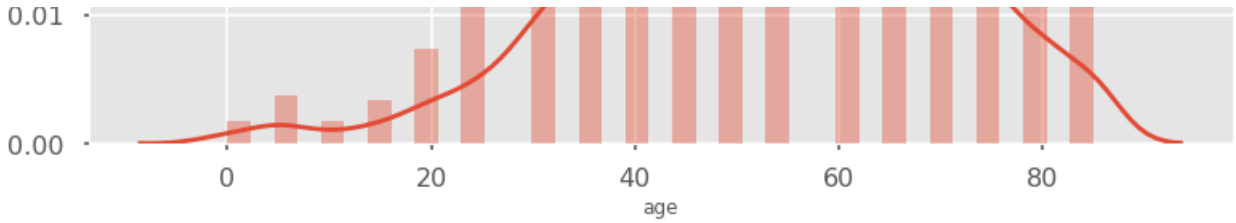
```
sns.distplot(df.age.dropna()).set_title("Age Distribution")
```

```
C:\Users\surface\ananew\lib\site-packages\scipy\stats\stats.py:1706: FutureWarning: Using a non-tu
ple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[se
q]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will re
sult either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

Out[18]:

```
Text(0.5,1,'Age Distribution')
```



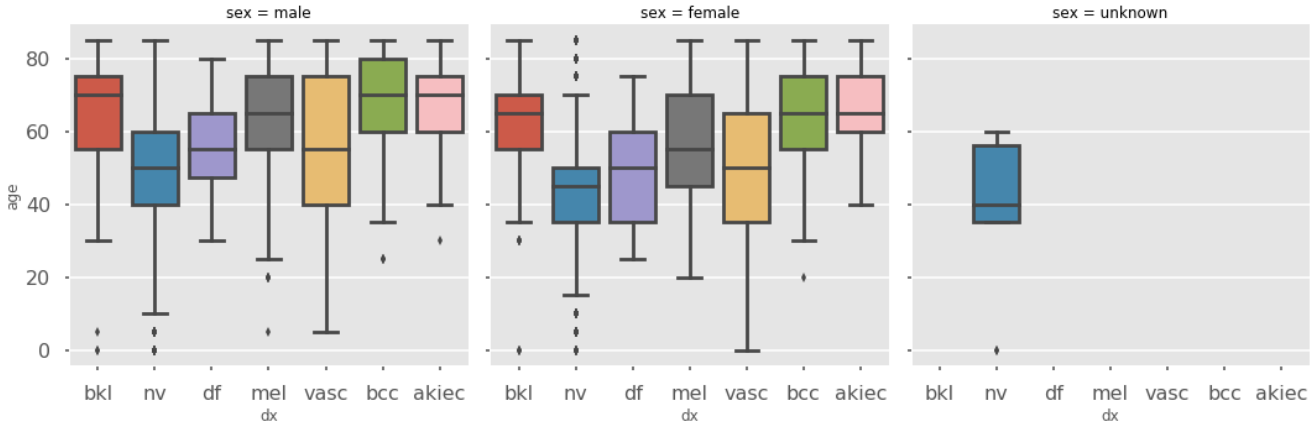


In [28]:

```
sns.catplot(x="dx", y="age", col="sex", kind="box", data=df)
```

Out[28]:

```
<seaborn.axisgrid.FacetGrid at 0x265d5ceb198>
```

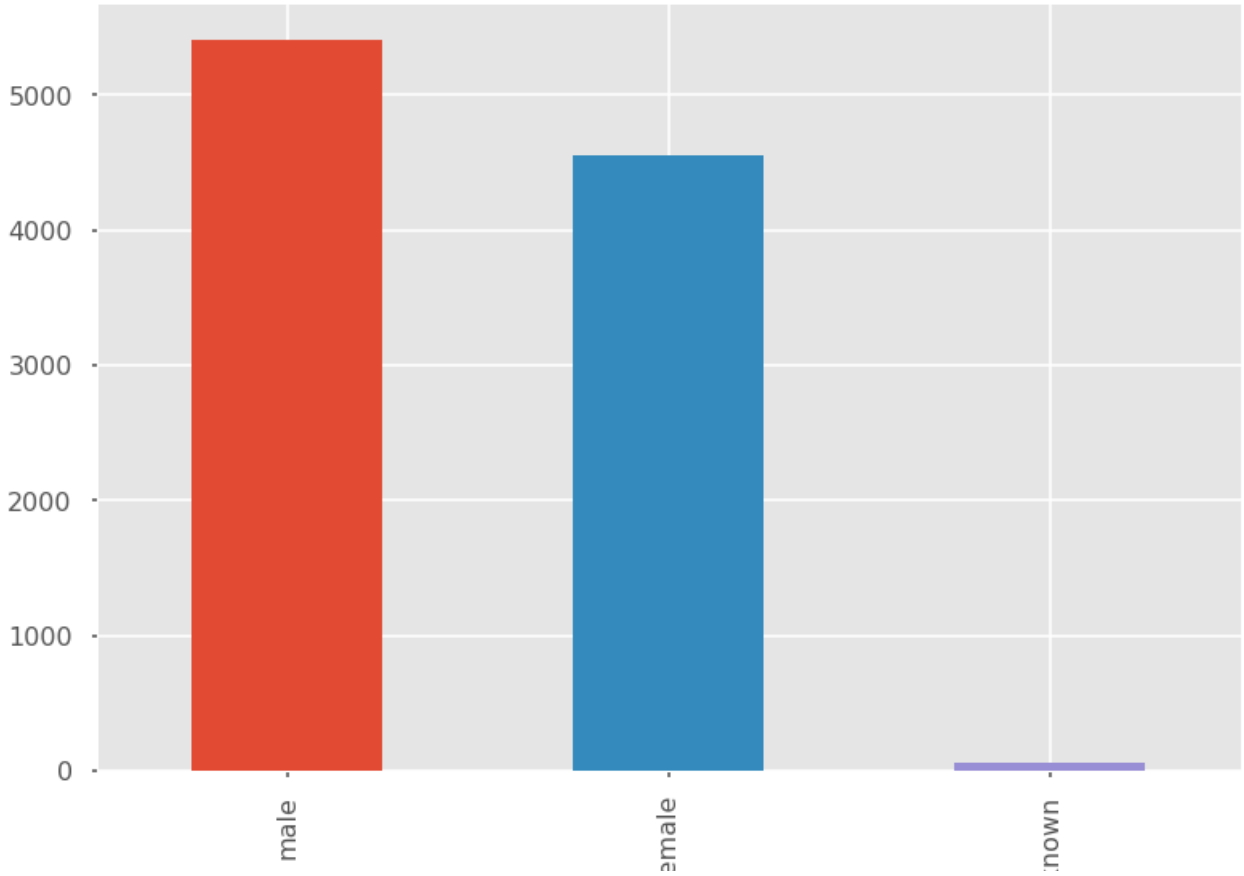


In [8]:

```
# Distribution of males and females
df['sex'].value_counts().plot(kind='bar')
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x265d14711d0>
```

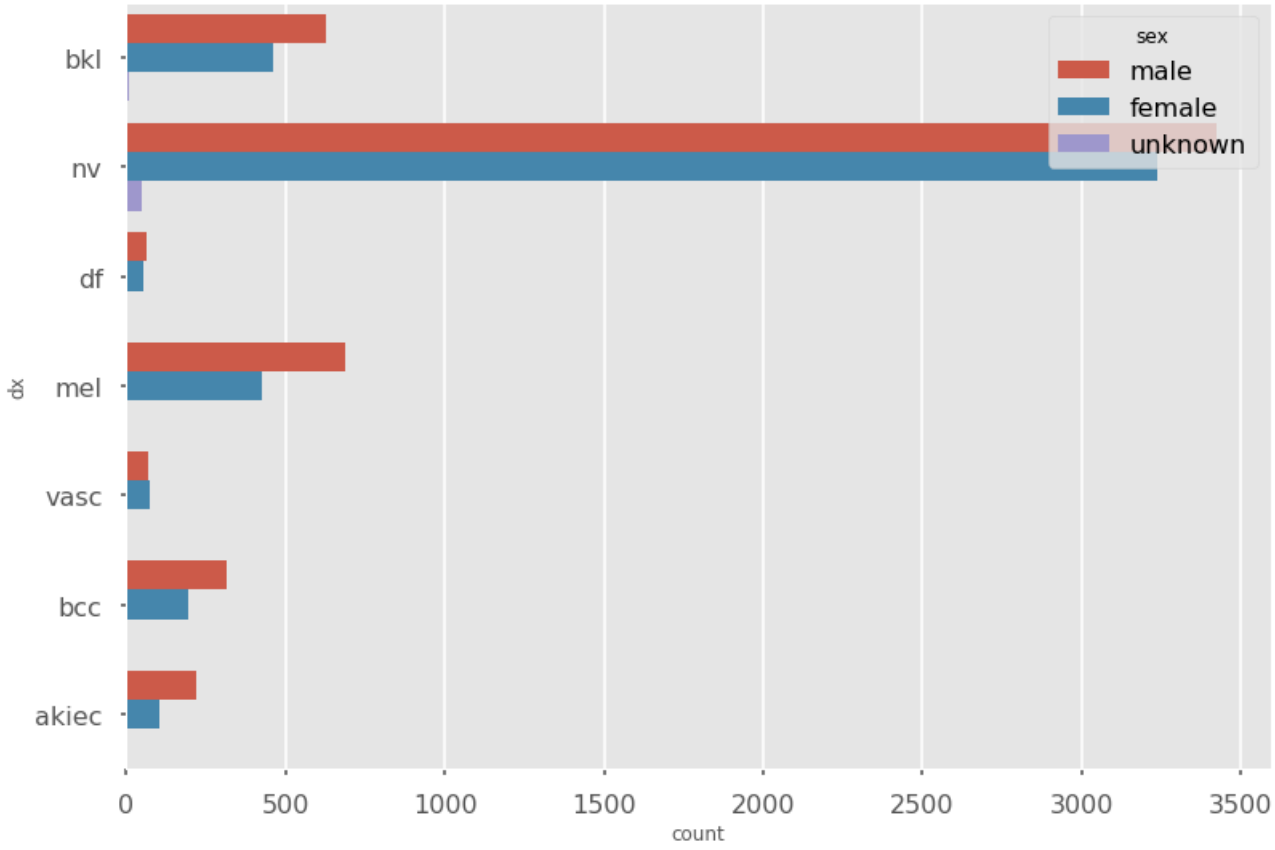


In [20]:

```
sns.countplot(y="dx", hue="sex", data=df)
```

Out[20]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x265d4b452b0>
```



In [27]:

```
sns.catplot(x="dx", hue="sex", col="dx_type", data = df, kind="count", height=4, aspect=1.0)
```

Out[27]:

```
<seaborn.axisgrid.FacetGrid at 0x265d5d446a0>
```