In [30]:

```
# Import libraries
from matplotlib import pyplot as plt
import numpy as np
import pandas as pd
file=pd.read_csv('C:/Users/wit5/Downloads/skin_cancer.csv')
file
```

Out[30]:

	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
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear
...	...	...	...	...	...	...	...
10010	HAM_0002867	ISIC_0033084	akiec	histo	40.0	male	abdomen
10011	HAM_0002867	ISIC_0033550	akiec	histo	40.0	male	abdomen
10012	HAM_0002867	ISIC_0033536	akiec	histo	40.0	male	abdomen
10013	HAM_0000239	ISIC_0032854	akiec	histo	80.0	male	face
10014	HAM_0003521	ISIC_0032258	mel	histo	70.0	female	back

10015 rows × 7 columns

In [13]:

```
file.isna().sum()
```

Out[13]:

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

In [18]:

```
mean_age=int(file['age'].mean())
```

In [19]:

```
file["age"].fillna(mean_age, inplace = True)
```

In [20]:

```
file.isna().sum()
```

Out[20]:

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

In [24]:

```
print("Number of male cancer patients:",file['sex'].value_counts()['male'])
print("Number of female cancer patients:",file['sex'].value_counts()['female'])
```

```
Number of male cancer patients: 5406
Number of female cancer patients: 4552
```

In [25]:

```
file['dx']
```

Out[25]:

```
0          bkl
1          bkl
2          bkl
3          bkl
4          bkl
         ...
10010    akiec
10011    akiec
10012    akiec
10013    akiec
10014      mel
Name: dx, Length: 10015, dtype: object
```

In [ ]:

```
file['dx']
```

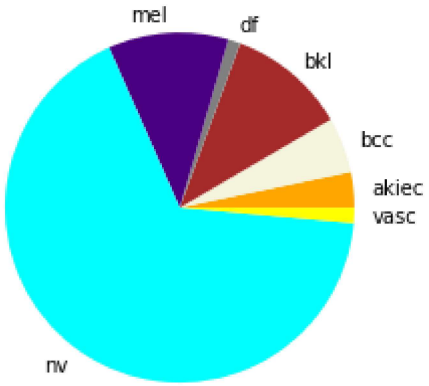
In [34]:

```
dx = file['dx'].value_counts().sort_index()
print(dx)
```

```
akiec     327
bcc       514
bkl      1099
df        115
mel      1113
nv       6705
vasc      142
Name: dx, dtype: int64
```

In [38]:

```
dxtypes=['akiec','bcc','bkl','df','mel','nv','vasc']
colors = ( "orange", "beige", "brown","grey", "indigo", "cyan","yellow")
plt.pie(dx, labels = dxtypes,colors=colors)
plt.show()
```



In [ ]:

```

```