In [39]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

In [14]:
```
df = pd.read_csv("exo.csv",comment='#')
```

In [15]:
```
df.head()
```

Out[15]:

	pl_name	hostname	default_flag	sy_snum	sy_pnum	discoverymethod	disc_year	
0	11 Com b	11 Com	1	2	1	Radial Velocity	2007	)
1	11 Com b	11 Com	0	2	1	Radial Velocity	2007	)
2	11 Com b	11 Com	0	2	1	Radial Velocity	2007	)
3	11 UMi b	11 UMi	0	1	1	Radial Velocity	2009	L
4	11 UMi b	11 UMi	0	1	1	Radial Velocity	2009	L

5 rows × 92 columns

In [16]:
```
cols = ['pl_name', 'hostname', 'pl_rade', 'pl_bmasse', 'pl_orbper',
       'pl_orbsmax', 'discoverymethod', 'disc_year', 'sy_dist']
```

In [17]:
```
df = df[cols]
```

In [18]:
```
df = df.dropna(subset=['pl_rade','pl_orbper','pl_orbsmax','disc_year'])
```

In [21]:
```
nums_cols = ['pl_rade','pl_orbper','pl_orbsmax','disc_year','sy_dist']
df[nums_cols] = df[nums_cols].apply(pd.to_numeric,errors='coerce')
```

In [24]:
```
df = df[df['pl_rade']<50]
df = df[df['pl_orbper']<5000]
df = df[df['sy_dist']<5000]
```

In [25]:
```
df.describe()
```

Out[25]:

	pl_rade	pl_bmasse	pl_orbper	pl_orbsmax	disc_year	sy_d
count	18003.000000	2252.000000	18003.000000	18003.000000	18003.000000	18003.0000
mean	3.528413	380.713521	25.265705	0.134360	2015.833917	719.0534
std	3.970319	801.763434	61.817680	0.155891	3.202177	459.4328
min	0.270000	0.070000	0.176893	0.005000	1999.000000	6.5312
25%	1.521000	11.775000	4.396596	0.051300	2014.000000	375.6470
50%	2.240000	160.500575	9.931436	0.087100	2016.000000	657.2230
75%	3.140000	380.606883	22.939815	0.156000	2016.000000	973.1210
max	48.785000	8899.000000	3650.000000	4.500000	2025.000000	3460.5100

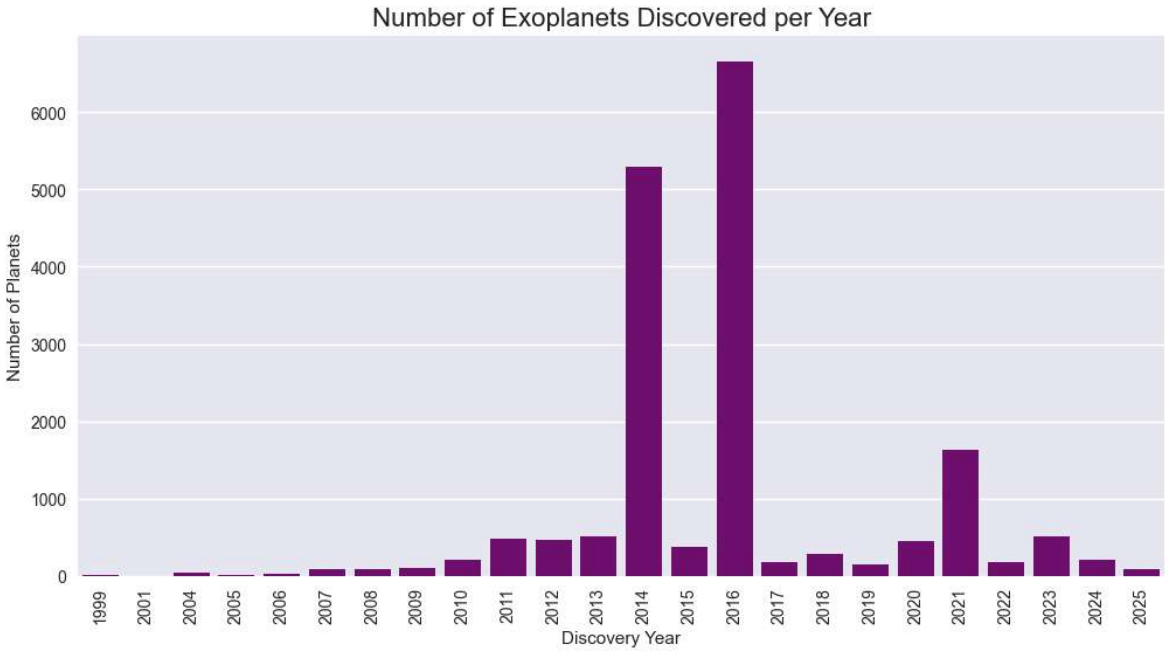
In [28]:
```
plt.style.use('seaborn-v0_8')
```

In [29]:
```
sns.set_palette('viridis')
```

In [44]:
```
df['pl_bmasse'] = df['pl_bmasse'].fillna(0.1)  # Small mass for missing values
df.shape
```

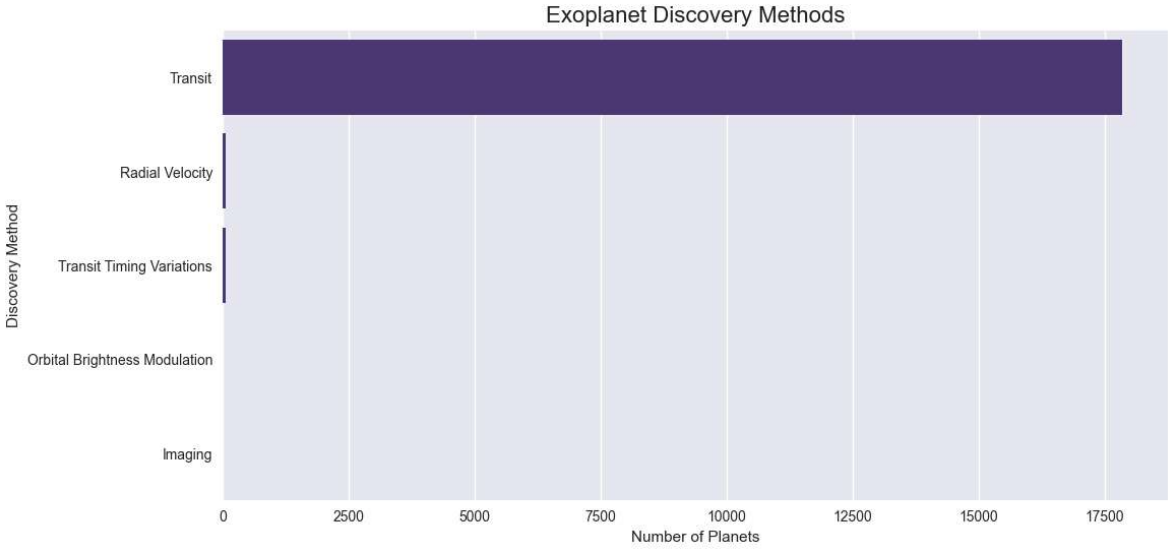
Out[44]: (18003, 9)

In [32]:
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='disc_year', color='purple')
plt.xticks(rotation=90)
plt.title('Number of Exoplanets Discovered per Year', fontsize=16)
plt.xlabel('Discovery Year')
plt.ylabel('Number of Planets')
plt.show()
```

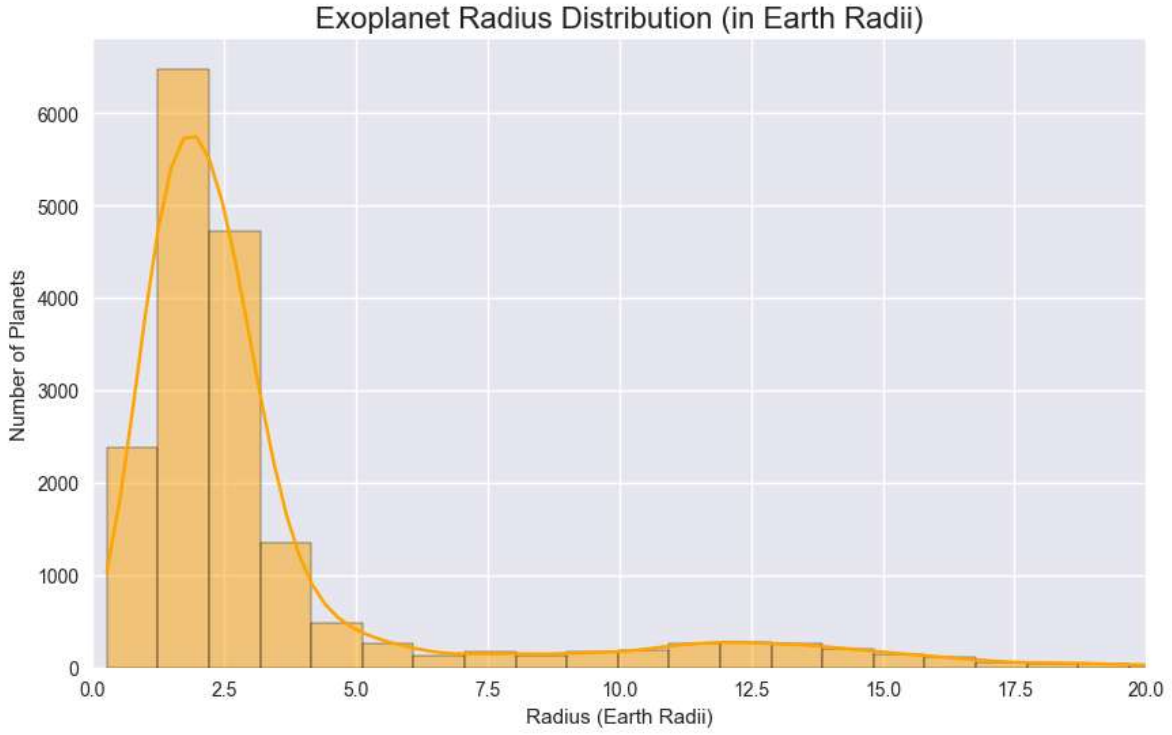


In [33]:
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, y='discoverymethod', order=df['discoverymethod'].value_co
plt.title('Exoplanet Discovery Methods', fontsize=16)
```

```
plt.xlabel('Number of Planets')
plt.ylabel('Discovery Method')
plt.show()
```

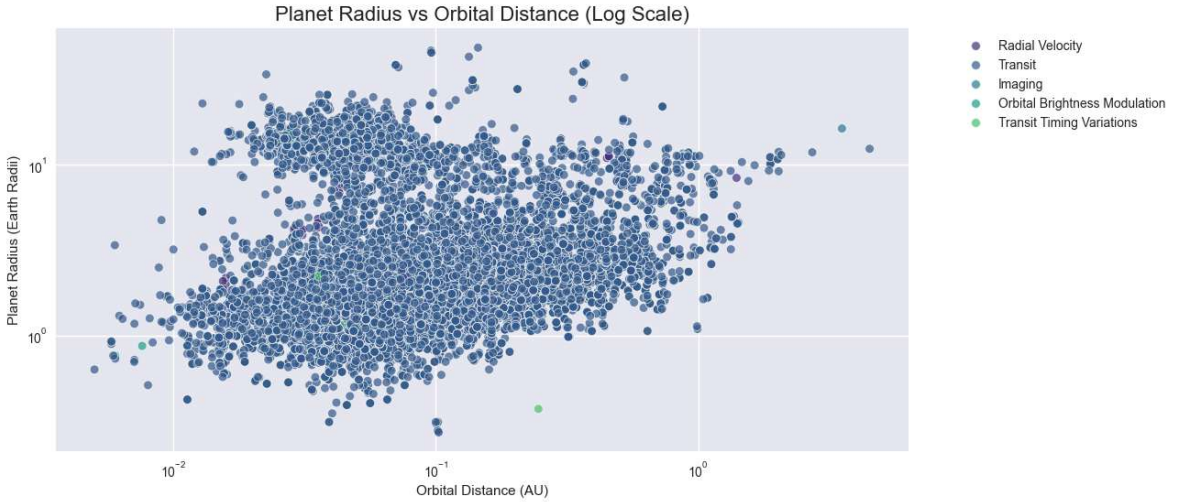


In [34]:
```
plt.figure(figsize=(10,6))
sns.histplot(df['pl_rade'], bins=50, kde=True, color='orange')
plt.xlim(0,20)  # Focus on smaller planets
plt.title('Exoplanet Radius Distribution (in Earth Radii)', fontsize=16)
plt.xlabel('Radius (Earth Radii)')
plt.ylabel('Number of Planets')
plt.show()
```

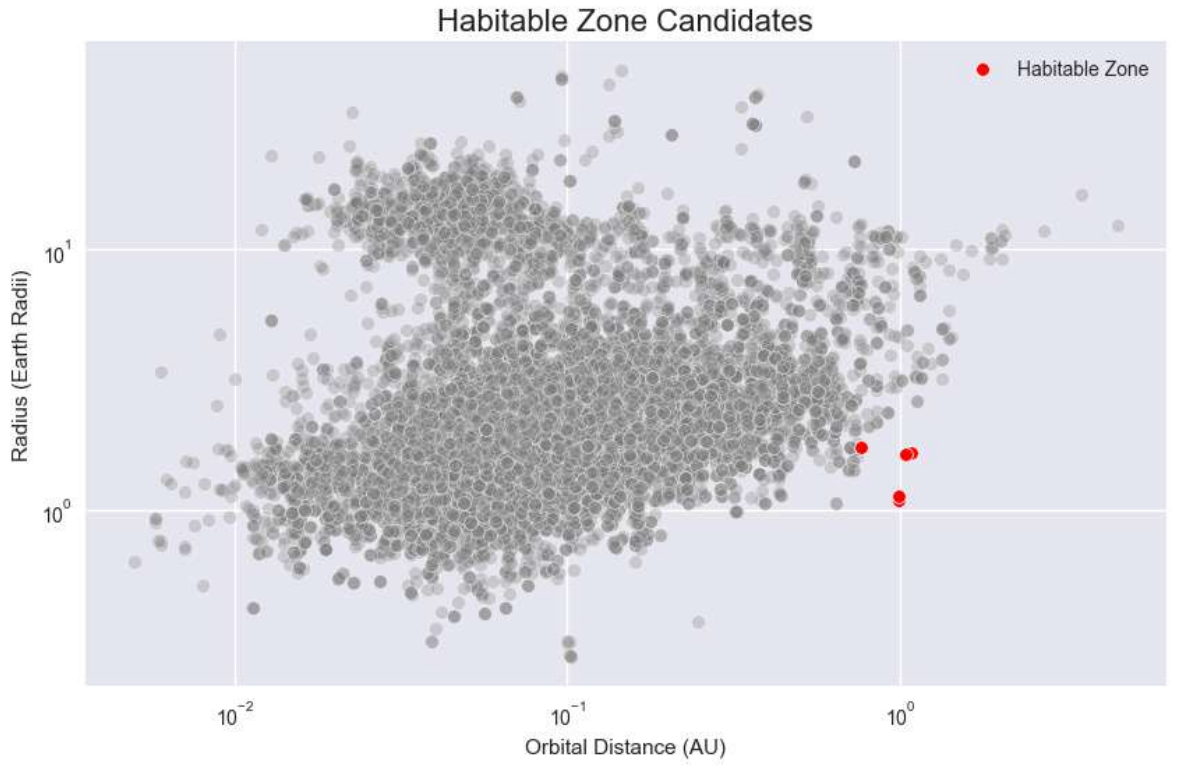


In [35]:
```
plt.figure(figsize=(12,6))
sns.scatterplot(data=df, x='pl_orbsmax', y='pl_rade', hue='discoverymethod', alp
plt.xscale('log')
plt.yscale('log')
plt.title('Planet Radius vs Orbital Distance (Log Scale)', fontsize=16)
plt.xlabel('Orbital Distance (AU)')
plt.ylabel('Planet Radius (Earth Radii)')
```

```
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```

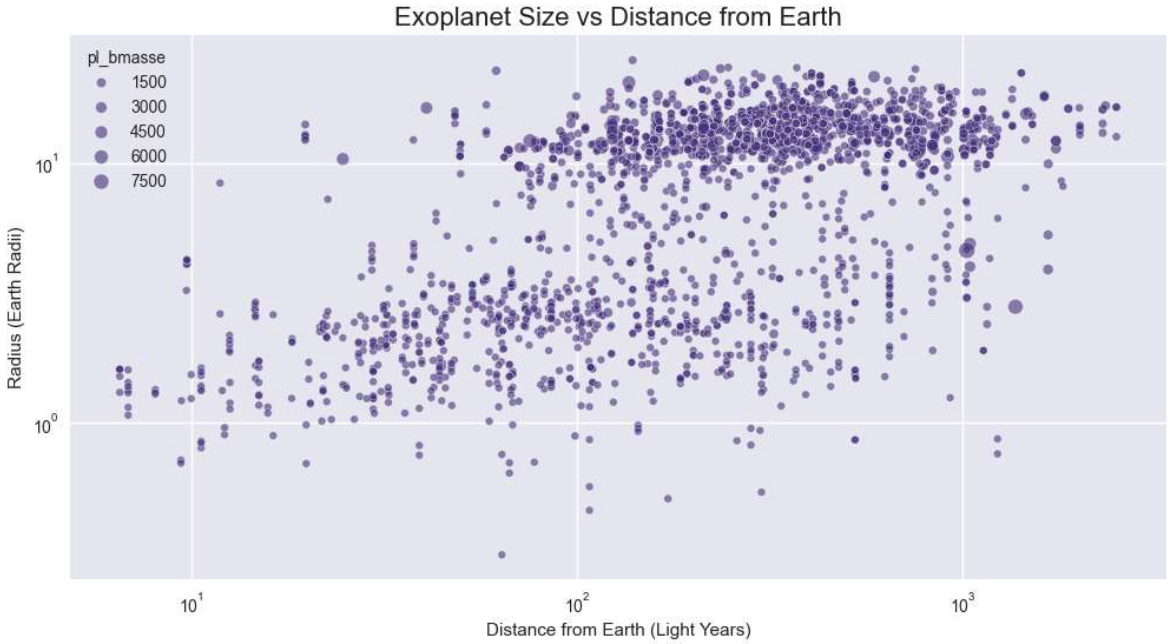


In [36]:
```
habitable = df[(df['pl_orbsmax'] > 0.75) & (df['pl_orbsmax'] < 1.5) &
               (df['pl_rade'] > 0.5) & (df['pl_rade'] < 2)]

plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x='pl_orbsmax', y='pl_rade', color='gray', alpha=0.3)
sns.scatterplot(data=habitable, x='pl_orbsmax', y='pl_rade', color='red', label=
plt.xscale('log')
plt.yscale('log')
plt.title('Habitable Zone Candidates', fontsize=16)
plt.xlabel('Orbital Distance (AU)')
plt.ylabel('Radius (Earth Radii)')
plt.legend()
plt.show()

print("Number of potential habitable planets:", len(habitable))
```




```
Number of potential habitable planets: 6
```

In [37]:
```
plt.figure(figsize=(12,6))
sns.scatterplot(data=df, x='sy_dist', y='pl_rade', size='pl_bmasse', alpha=0.6)
plt.xscale('log')
plt.yscale('log')
plt.title('Exoplanet Size vs Distance from Earth', fontsize=16)
plt.xlabel('Distance from Earth (Light Years)')
plt.ylabel('Radius (Earth Radii)')
plt.show()
```



In [45]:
```
df_plot = df.dropna(subset=['pl_bmasse'])
```

In [46]:
```
import plotly.express as px

fig = px.scatter_3d(
    df_plot,  # Use df if you filled NaNs, else df_plot
    x='pl_orbsmax',    # Orbital distance
    y='pl_rade',       # Planet radius
    z='disc_year',     # Discovery year
    color='discoverymethod',
    size='pl_bmasse',  # Now clean
    hover_name='pl_name',
    log_x=True,
    log_y=True
)

fig.update_layout(title="3D Exoplanet Visualization: Distance, Size & Discovery
fig.show()
```

In [47]:
```
habitable = df[(df['pl_orbsmax'] > 0.75) & (df['pl_orbsmax'] < 1.5) &
               (df['pl_rade'] > 0.5) & (df['pl_rade'] < 2)]

fig = px.scatter_3d(
    habitable,
    x='pl_orbsmax',    # Orbit distance (AU)
    y='pl_rade',       # Radius (Earth radii)
    z='sy_dist',       # Distance from Earth (light years)
    color='disc_year', # Color by discovery year
    hover_name='pl_name',
```

```
    log_x=True,
    log_y=True
)

fig.update_layout(title="3D Habitable Zone Planets")
fig.show()
```