2022 Data Analysis

In [1]:
```
%load_ext autoreload
%autoreload 2

import seaborn as sns
sns.set()

# import utility functions
import sys
sys.path.append('../')
from src.utils import *
```

In [2]:
```
%matplotlib inline
import matplotlib as mpl
mpl.rcParams['figure.dpi']= 300
```

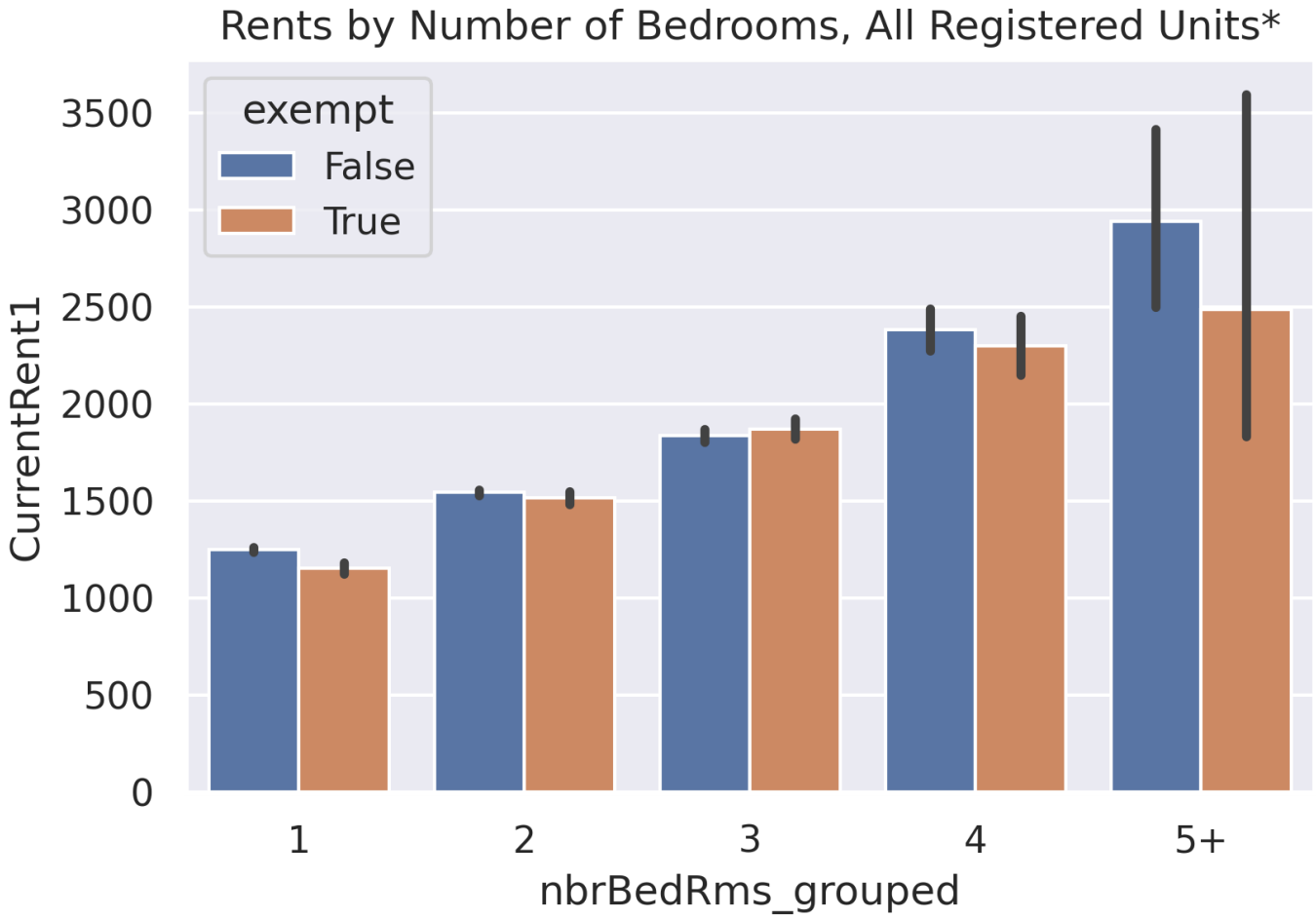
In [3]:
```
# load the 2022 data
df = get_data(2022)
```

```
Breakdown by Outlier Condition:
        Outlier Rents: 6335 (34%)
        Outlier Increase vs Base: 777 ( 4%)
        Outlier Increase vs Previous: 609 ( 3%)
        Overall: 6682 (36%)

Breakdown by Subset:
        6682 outliers (36%)
        11890 non-outliers (64%)


        4142 rent increase (22%)
        14430 no rent increase (78%)


        7036 exempt (38%)
        11536 not exempt (62%)
```

Outliers Removed

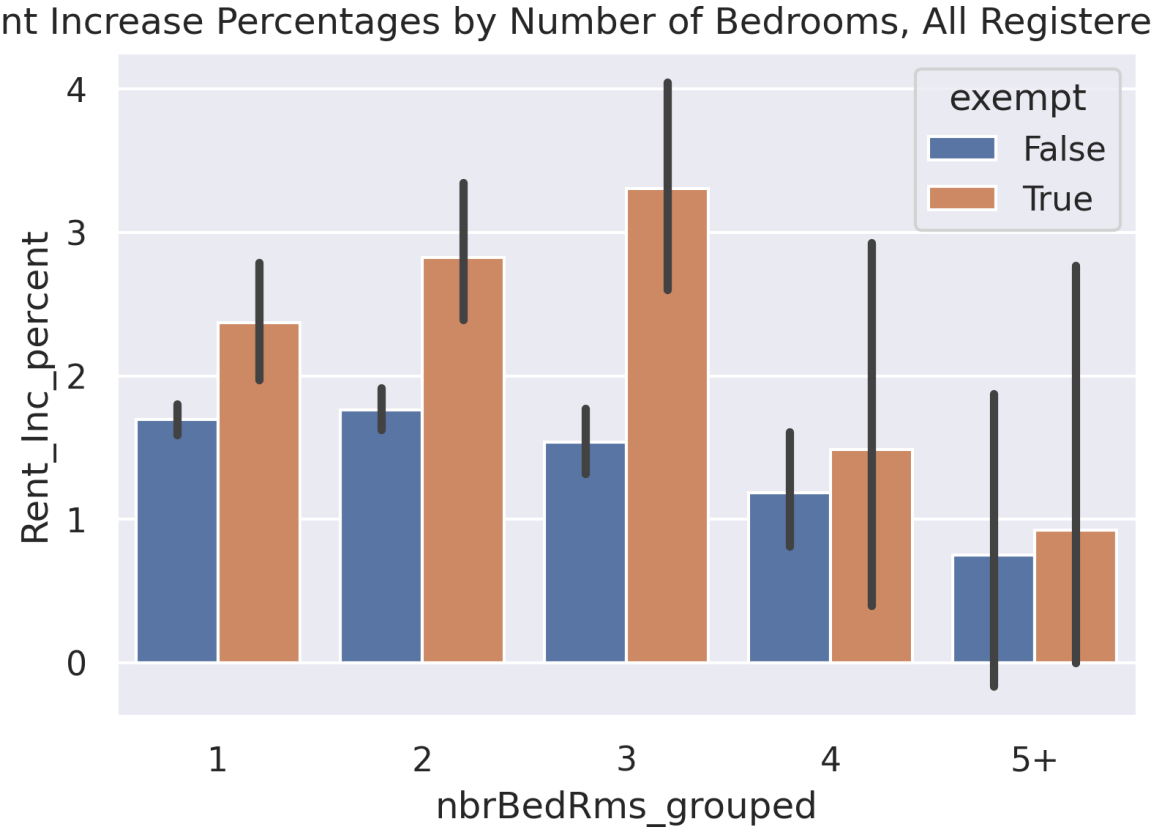
In [4]:
```
ax = sns.barplot(
    data=df[~df["outlier"]].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="CurrentRent1",
    hue="exempt"
    )
ax.set_title("Rents by Number of Bedrooms, All Registered Units*")
```

Out[4]:
```
Text(0.5, 1.0, 'Rents by Number of Bedrooms, All Registered Units*')
```



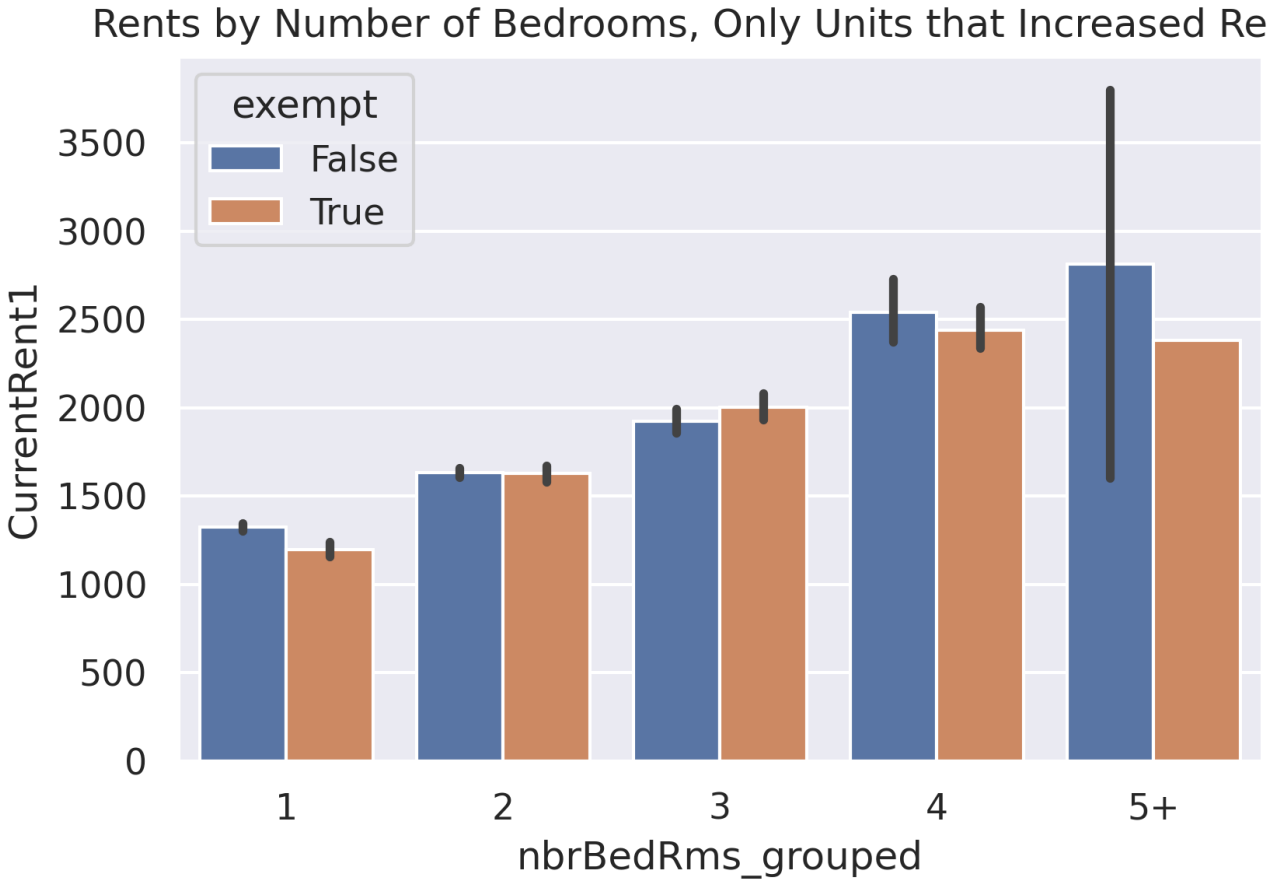
In [5]:
```
ax = sns.barplot(
    data=df[~df["outlier"]].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="Rent_Inc_percent",
    hue="exempt"
    )
ax.set_title("Rent Increase Percentages by Number of Bedrooms, All Registered Units*")
```

Out[5]:
```
Text(0.5, 1.0, 'Rent Increase Percentages by Number of Bedrooms, All Registered Units*')
```



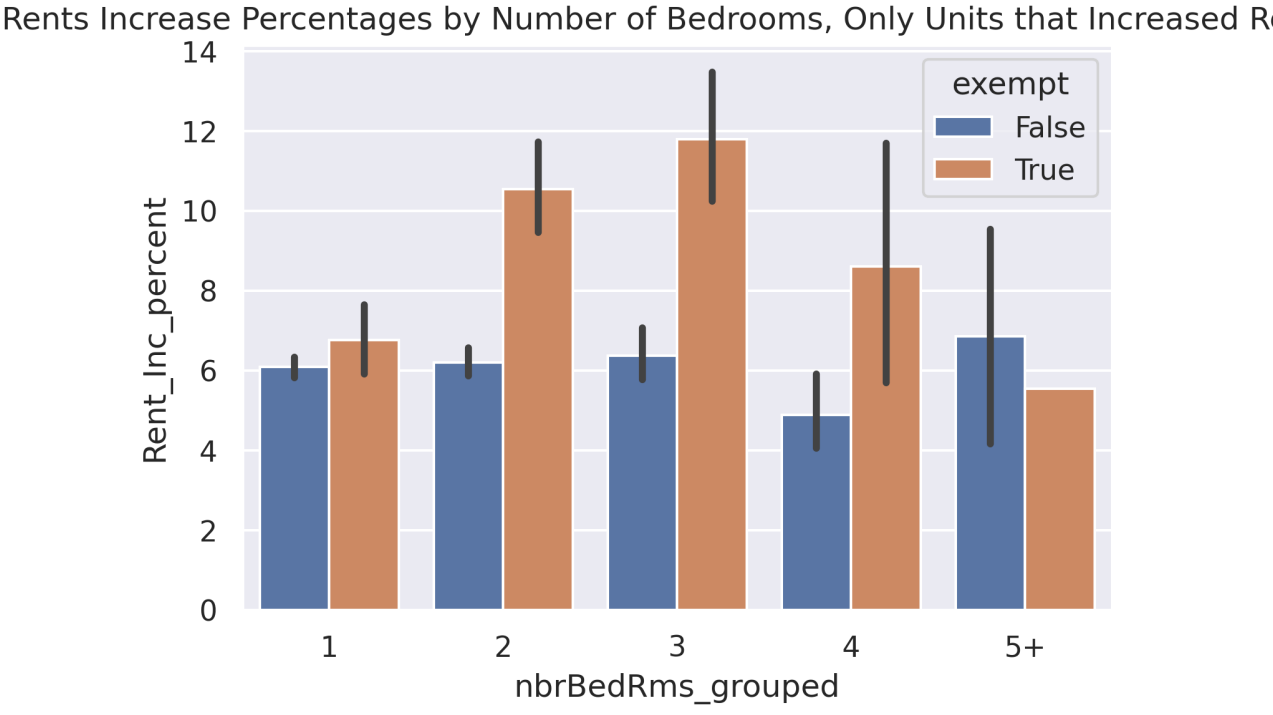
In [6]:
```
ax = sns.barplot(
    data=df[~df["outlier"] & (df["Rent_Inc"] > 0)].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="CurrentRent1",
    hue="exempt"
    )
ax.set_title("Rents by Number of Bedrooms, Only Units that Increased Rents*")
```

Out[6]:
```
Text(0.5, 1.0, 'Rents by Number of Bedrooms, Only Units that Increased Rents*')
```



In [7]:
```
ax = sns.barplot(
    data=df[~df["outlier"] & (df["Rent_Inc"] > 0)].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="Rent_Inc_percent",
    hue="exempt"
    )
ax.set_title("Rents Increase Percentages by Number of Bedrooms, Only Units that Increase
```

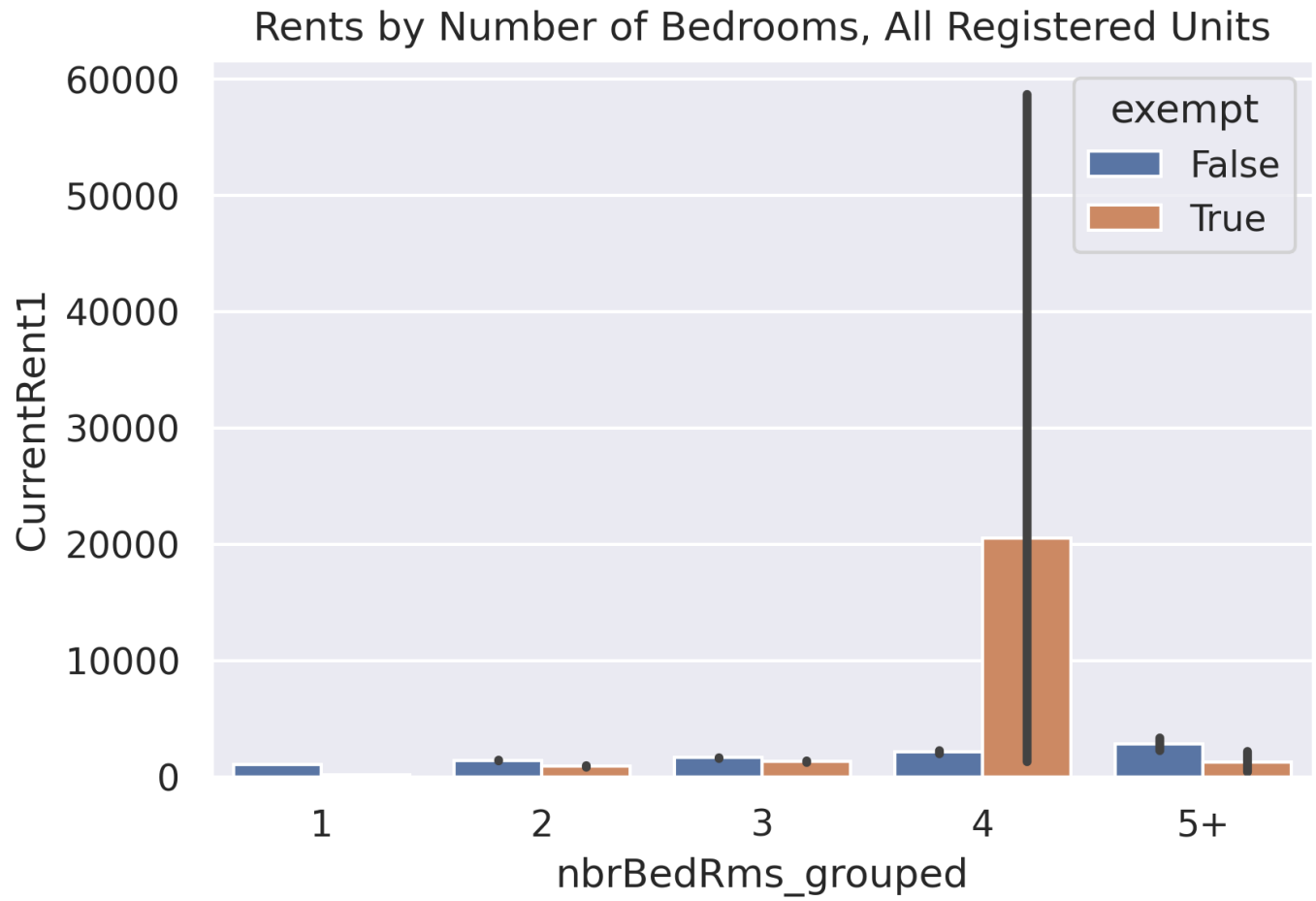
Out[7]:
```
Text(0.5, 1.0, 'Rents Increase Percentages by Number of Bedrooms, Only Units that Increa
sed Rents*')
```



Outliers Included

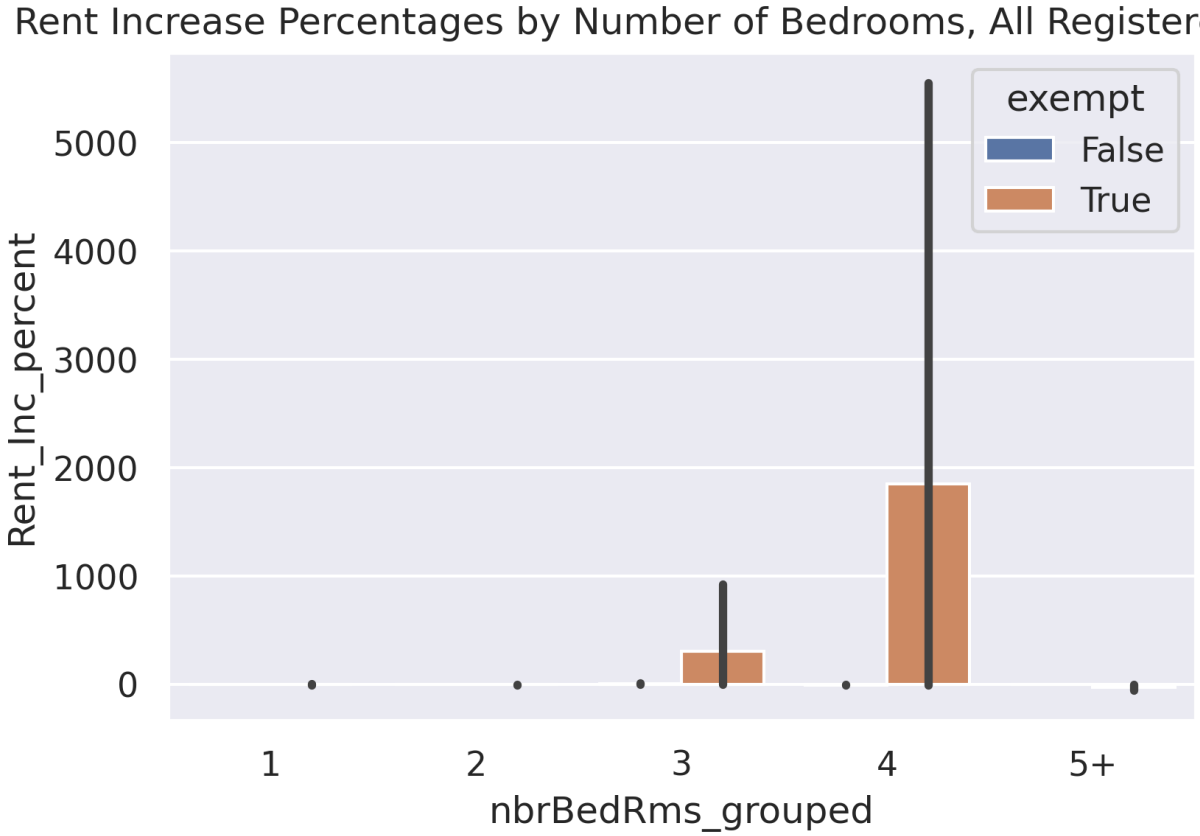
In [8]:
```
ax = sns.barplot(
    data=df.sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="CurrentRent1",
    hue="exempt"
    )
ax.set_title("Rents by Number of Bedrooms, All Registered Units")
```

Out[8]:
```
Text(0.5, 1.0, 'Rents by Number of Bedrooms, All Registered Units')
```



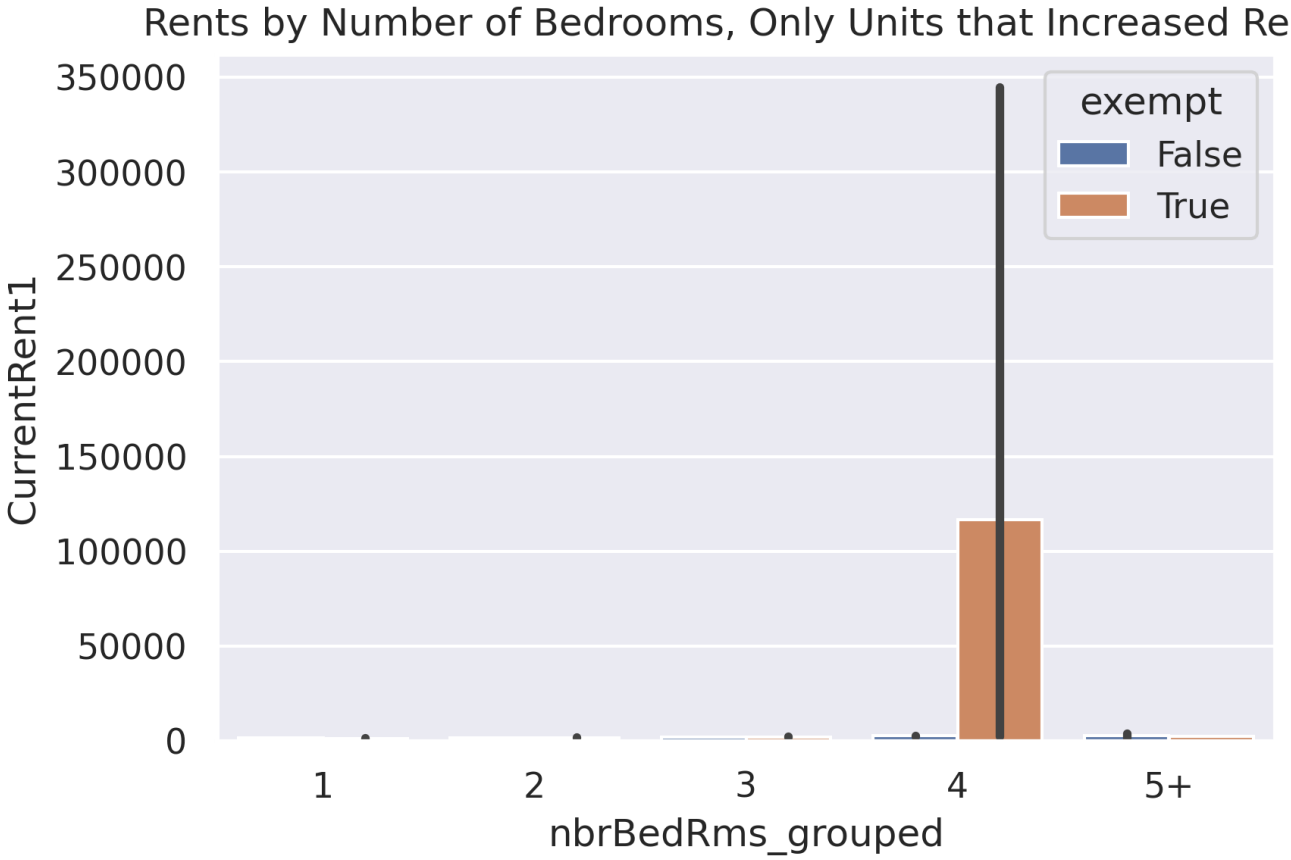
In [9]:
```
ax = sns.barplot(
    data=df.sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="Rent_Inc_percent",
    hue="exempt"
    )
ax.set_title("Rent Increase Percentages by Number of Bedrooms, All Registered Units")
```

Out[9]:
```
Text(0.5, 1.0, 'Rent Increase Percentages by Number of Bedrooms, All Registered Units')
```



In [10]:
```
ax = sns.barplot(
    data=df[df["Rent_Inc"] > 0].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="CurrentRent1",
    hue="exempt"
    )
ax.set_title("Rents by Number of Bedrooms, Only Units that Increased Rents")
```

Out[10]:
```
Text(0.5, 1.0, 'Rents by Number of Bedrooms, Only Units that Increased Rents')
```



In [11]:
```
ax = sns.barplot(
    data=df[(df["Rent_Inc"] > 0)].sort_values("nbrBedRms_grouped"),
    x="nbrBedRms_grouped",
    y="Rent_Inc_percent",
    hue="exempt"
    )
ax.set_title("Rents Increase Percentages by Number of Bedrooms, Only Units that Increase
```

Out[11]:
```
Text(0.5, 1.0, 'Rents Increase Percentages by Number of Bedrooms, Only Units that Increa
sed Rents')
```