

```
In [4]: import pandas as pd
import numpy as np
```

```
In [5]: data=pandas.read_csv("dataset.csv")
```

```
In [6]: data
```

```
Out[6]:
```

	encounter_id	patient_id	hospital_id	hospital_death	age	bmi	elective_surgery	eth
0	66154	25312	118	0	68.0	22.730000	0	Cauc
1	114252	59342	81	0	77.0	27.420000	0	Cauc
2	119783	50777	118	0	25.0	31.950000	0	Cauc
3	79267	46918	118	0	81.0	22.640000	1	Cauc
4	92056	34377	33	0	19.0	NaN	0	Cauc
...
91708	91592	78108	30	0	75.0	23.060250	0	Cauc
91709	66119	13486	121	0	56.0	47.179671	0	Cauc
91710	8981	58179	195	0	48.0	27.236914	0	Cauc
91711	33776	120598	66	0	NaN	23.297481	0	Cauc
91712	1671	53612	104	0	82.0	22.031250	1	Cauc

91713 rows × 186 columns



```
In [7]: Patient_df = df = pd.read_csv("dataset.csv")
```

```
In [8]: Patient1_df = pd.DataFrame(data = [1,2,3,4,5,6], columns = ["Hospital_Death"] ,)
```

```
In [9]: Patient1_df
```

```
Out[9]:
```

	Hospital_Death
0	1
1	2
2	3
3	4
4	5
5	6

```
In [ ]:
```

In []:

In [10]: Patient_df.shape

Out[10]: (91713, 186)

In [11]: Patient_df.head()

Out[11]:

	encounter_id	patient_id	hospital_id	hospital_death	age	bmi	elective_surgery	ethnicity	g
0	66154	25312	118	0	68.0	22.73	0	Caucasian	
1	114252	59342	81	0	77.0	27.42	0	Caucasian	
2	119783	50777	118	0	25.0	31.95	0	Caucasian	
3	79267	46918	118	0	81.0	22.64	1	Caucasian	
4	92056	34377	33	0	19.0	NaN	0	Caucasian	

5 rows × 186 columns

In [12]: Patient_df.tail(8)

Out[12]:

	encounter_id	patient_id	hospital_id	hospital_death	age	bmi	elective_surgery	eth
91705	111411	7848	195	0	67.0	28.876843	0	Am
91706	127138	59223	121	0	54.0	19.770448	0	Am
91707	276	34638	183	0	NaN	33.933518	0	Cauc
91708	91592	78108	30	0	75.0	23.060250	0	Cauc
91709	66119	13486	121	0	56.0	47.179671	0	Cauc
91710	8981	58179	195	0	48.0	27.236914	0	Cauc
91711	33776	120598	66	0	NaN	23.297481	0	Cauc
91712	1671	53612	104	0	82.0	22.031250	1	Cauc

8 rows × 186 columns

In [13]: Patient_df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91713 entries, 0 to 91712
Columns: 186 entries, encounter_id to apache_2_bodysystem
dtypes: float64(170), int64(8), object(8)
memory usage: 130.1+ MB
```

In [14]: data.dtypes

```
Out[14]: encounter_id      int64
patient_id      int64
hospital_id     int64
hospital_death  int64
age            float64
...
leukemia        float64
lymphoma        float64
solid_tumor_with_metastasis float64
apache_3j_bodysystem object
apache_2_bodysystem object
Length: 186, dtype: object
```

In [15]: data.head(18)

```
Out[15]:
```

	encounter_id	patient_id	hospital_id	hospital_death	age	bmi	elective_surgery	ethnicity
0	66154	25312	118	0	68.0	22.730000	0	Caucasian
1	114252	59342	81	0	77.0	27.420000	0	Caucasian
2	119783	50777	118	0	25.0	31.950000	0	Caucasian
3	79267	46918	118	0	81.0	22.640000	1	Caucasian
4	92056	34377	33	0	19.0	NaN	0	Caucasian
5	33181	74489	83	0	67.0	27.560000	0	Caucasian
6	82208	49526	83	0	59.0	57.450000	0	Caucasian
7	120995	50129	33	0	70.0	NaN	0	Caucasian
8	80471	10577	118	1	45.0	NaN	0	Caucasian
9	42871	90749	118	0	50.0	25.710000	0	Native American
10	105427	125898	77	0	72.0	28.257052	1	Hispanic
11	91609	78266	83	0	80.0	27.382812	1	Caucasian
12	76356	41311	118	0	48.0	NaN	0	Caucasian
13	34030	103766	118	0	65.0	NaN	1	Caucasian
14	108388	98174	118	0	81.0	38.189067	1	Caucasian
15	120677	124688	118	0	78.0	NaN	0	Caucasian
16	115771	71252	81	0	30.0	23.383178	0	Caucasian
17	22471	112115	118	0	46.0	25.845717	0	Hispanic

18 rows × 186 columns



In [16]: `data.columns`

Out[16]: Index(['encounter_id', 'patient_id', 'hospital_id', 'hospital_death', 'age',
'bmi', 'elective_surgery', 'ethnicity', 'gender', 'height',
...
'aids', 'cirrhosis', 'diabetes_mellitus', 'hepatic_failure',
'immunosuppression', 'leukemia', 'lymphoma',
'solid_tumor_with_metastasis', 'apache_3j_bodysystem',
'apache_2_bodysystem'],
dtype='object', length=186)

In []:

In [17]: `data.info(verbose=True, null_counts=True)`

In [18]: `data.describe()`

Out[18]:

	encounter_id	patient_id	hospital_id	hospital_death	age	bmi	elective_surgery
count	91713.000000	91713.000000	91713.000000	91713.000000	87485.000000	88284.000000	91713.000000
mean	65606.079280	65537.131464	105.669262	0.086302	62.309516	29.185818	0.000000
std	37795.088538	37811.252183	62.854406	0.280811	16.775119	8.275142	0.000000
min	1.000000	1.000000	2.000000	0.000000	16.000000	14.844926	0.000000
25%	32852.000000	32830.000000	47.000000	0.000000	52.000000	23.641975	0.000000
50%	65665.000000	65413.000000	109.000000	0.000000	65.000000	27.654655	0.000000
75%	98342.000000	98298.000000	161.000000	0.000000	75.000000	32.930206	0.000000
max	131051.000000	131051.000000	204.000000	1.000000	89.000000	67.814990	0.000000

8 rows × 178 columns

In [19]: `data.describe(include="object")`

Out[19]:

	ethnicity	gender	hospital_admit_source	icu_admit_source	icu_stay_type	icu_type	apac
count	90318	91688	70304	91601	91713	91713	
unique	6	2	15	5	3	8	
top	Caucasian	M	Emergency Department	Accident & Emergency	admit	Med-Surg ICU	
freq	70684	49469	36962	54060	86183	50586	

In [20]: `data.isnull().sum(axis=0).sort_values(ascending=False)`

Out[20]:

```

h1_bilirubin_min      84619
h1_bilirubin_max      84619
h1_lactate_max        84369
h1_lactate_min        84369
h1_albumin_max       83824
...
icu_type              0
pre_icu_los_days      0
readmission_status    0
apache_post_operative 0
encounter_id          0
Length: 186, dtype: int64

```

In [21]: `print("Rows with missing values:", data.isnull().any(axis=1).sum())`

Rows with missing values: 91688

In [22]: `missing_values = data.isnull().sum(axis=0).sort_values(ascending=False)[data.isnull().any(axis=1).sum():]`
`print("Features missing", 27000, "missing values:", len(missing_values))`

Features missing 27000 missing values: 74

```
In [23]: data.drop(missing_values.index.tolist() + ["encounter_id", 'hospital_admit_source'],
                  axis=1,
                  inplace = True)

data
```

Out[23]:

ight	...	aids	cirrhosis	diabetes_mellitus	hepatic_failure	immunosuppression	leukemia	lymphoma
73.9	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0
70.2	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0
35.3	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31.7	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NaN	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
72.9	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0
58.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
78.9	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0
55.9	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
56.4	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

```
In [24]: data = data[data[['bmi', 'weight', 'height']].isna().sum(axis=1) == 0]
data
```

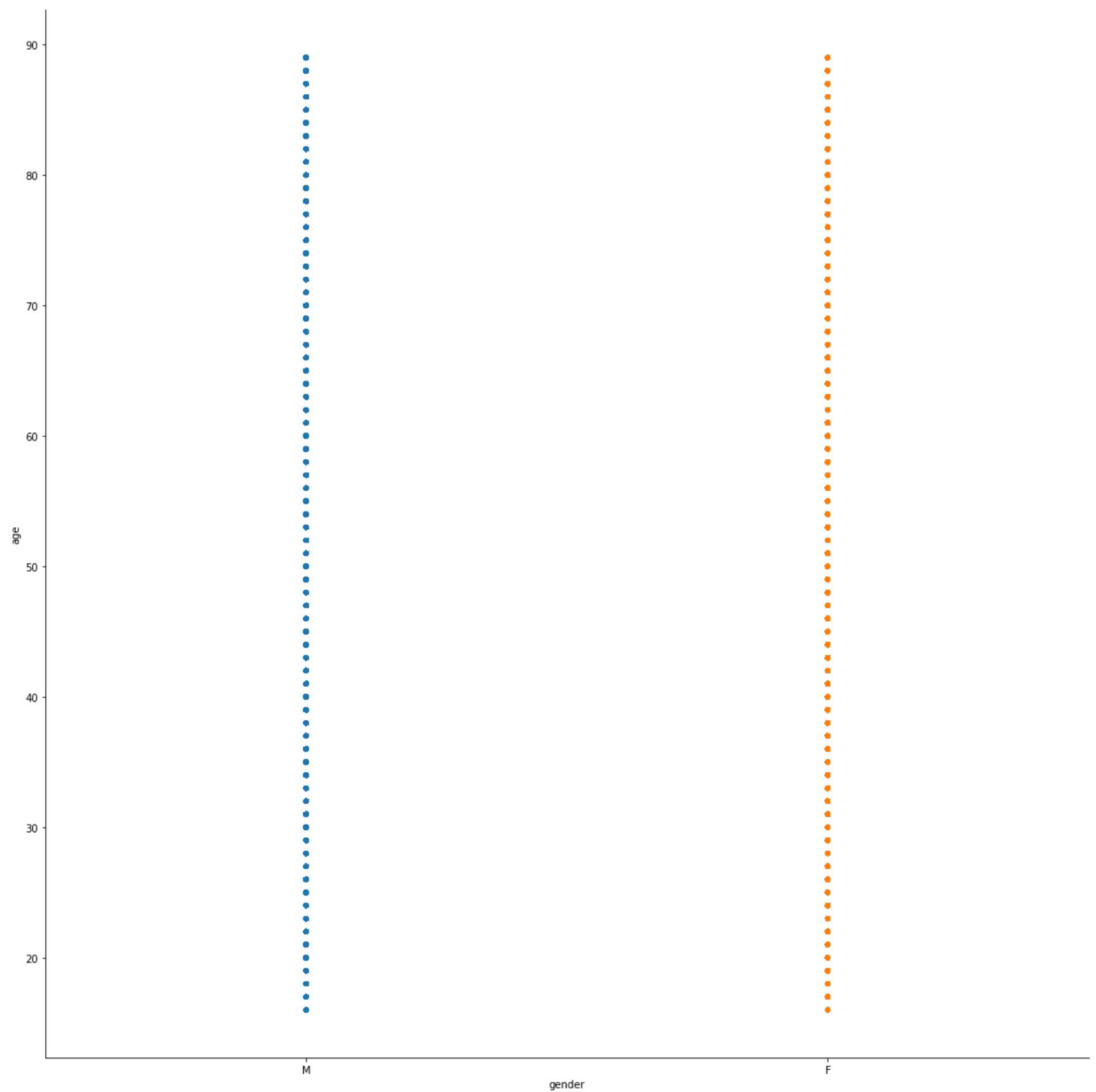
5	0	67.0	27.560000	0	Caucasian	M	190.5	Surg ICU	0.000694
...
91708	0	75.0	23.060250	0	Caucasian	M	177.8	Cardiac ICU	0.298611
91709	0	56.0	47.179671	0	Caucasian	F	183.0	Med-Surg ICU	0.120139
91710	0	48.0	27.236914	0	Caucasian	M	170.2	Med-Surg ICU	0.046528
91711	0	NaN	23.297481	0	Caucasian	F	154.9	Med-Surg ICU	0.081944
91712	0	82.0	22.031250	1	Caucasian	F	160.0	Med-Surg ICU	0.018056

```
In [26]: import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]:
```

```
In [27]: sns.catplot(x="gender", y="age", order=["M", "F"], jitter=False, height=15, data=
```

```
Out[27]: <seaborn.axisgrid.FacetGrid at 0x217ce86a7f0>
```



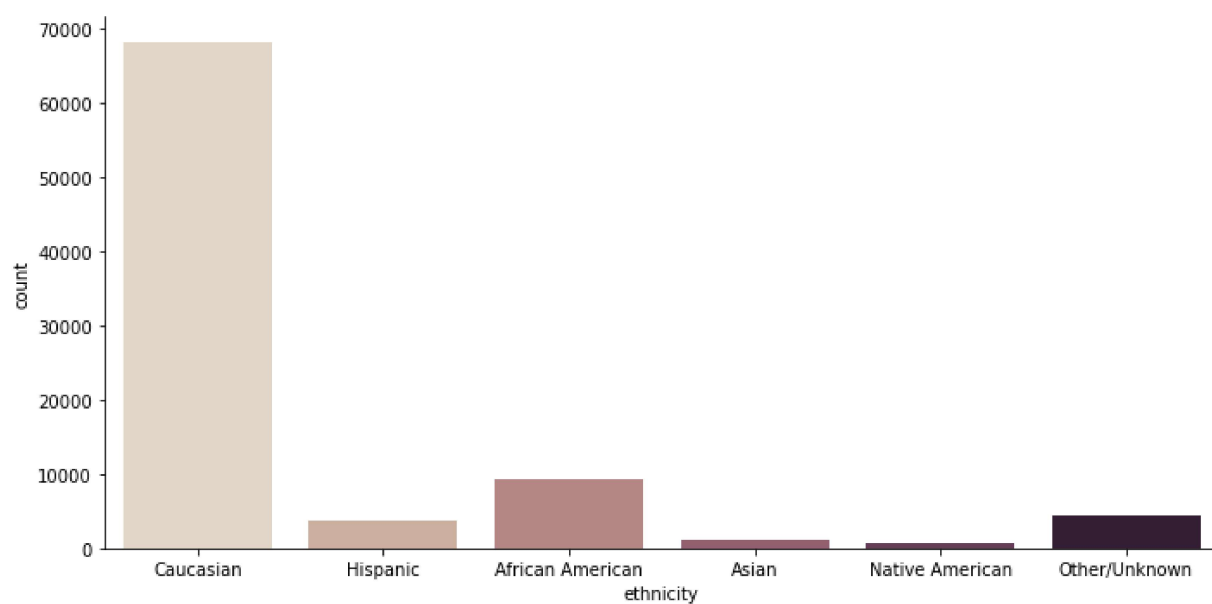
In []:

In []:

In []:

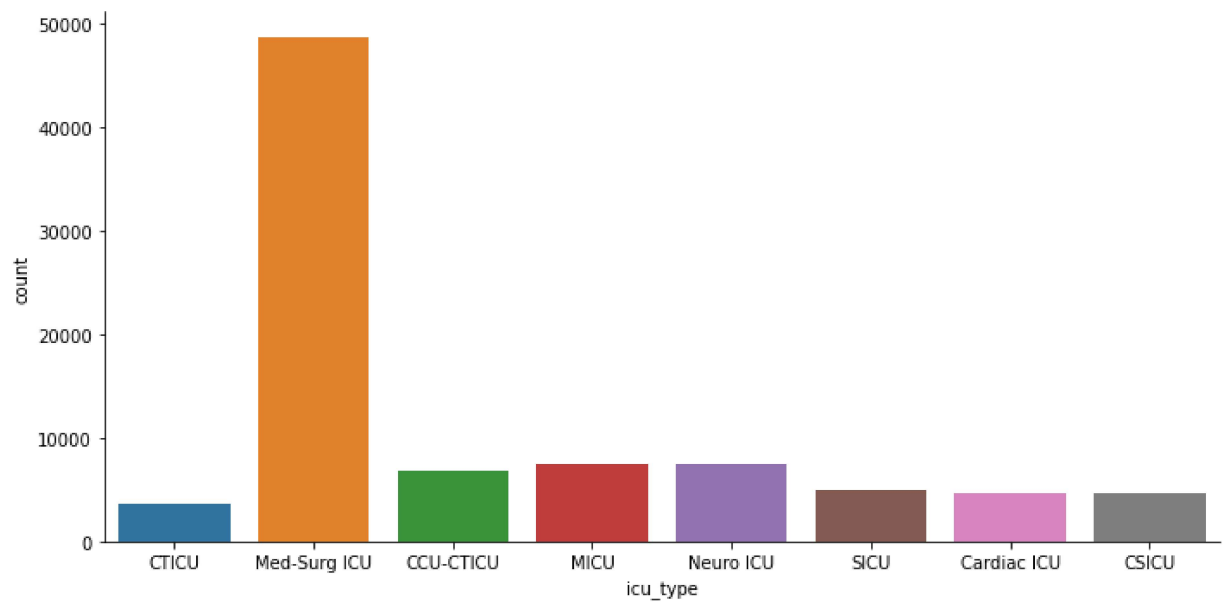
```
In [139]: sns.catplot(x="ethnicity", kind="count", aspect=2, palette="ch:.25", data=data)
```

```
Out[139]: <seaborn.axisgrid.FacetGrid at 0x217df690280>
```



```
In [29]: sns.catplot(x="icu_type", kind="count", aspect=2, data=data)  
sns.color_palette("Set2")
```

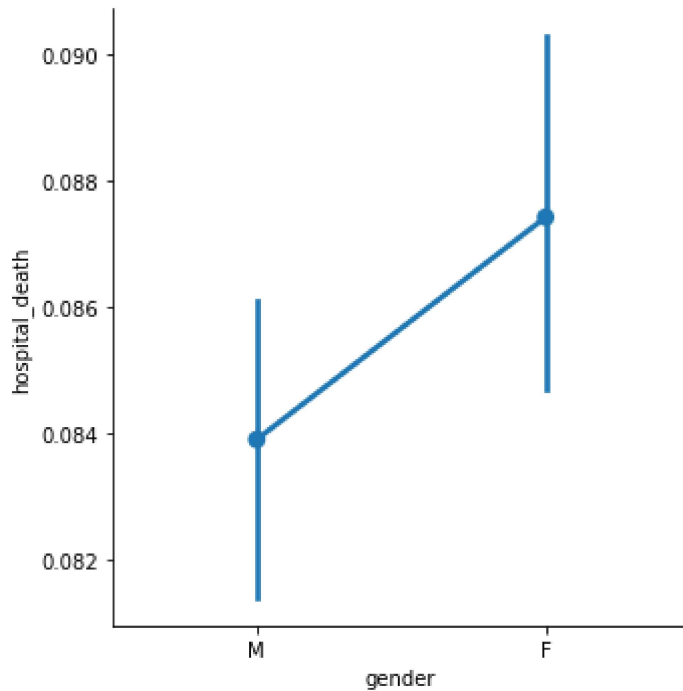
Out[29]:



In []:

```
In [44]: sns.catplot(x="gender", y="hospital_death", kind="point", data=data)
```

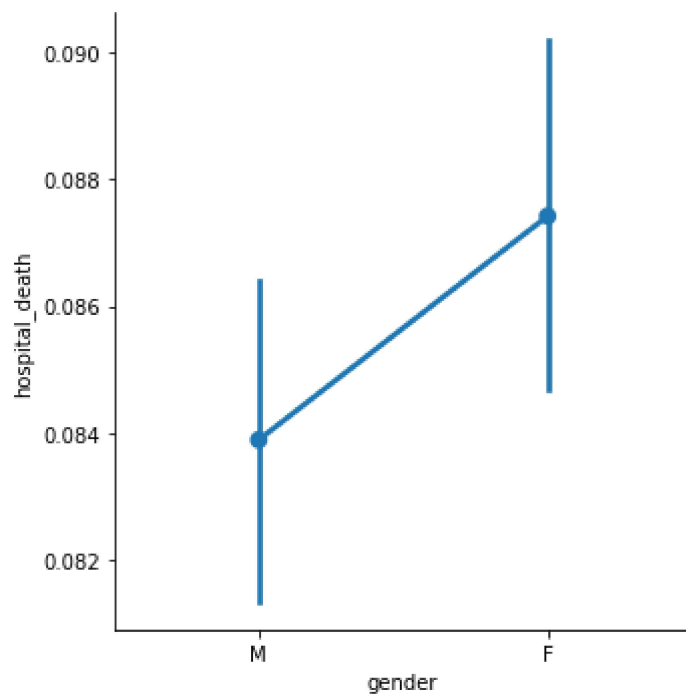
```
Out[44]: <seaborn.axisgrid.FacetGrid at 0x2179bccc310>
```



```
In [137]: def custom_function_death(): sns.catplot(x="gender", y="hospital_death", kind="point", data=data)
```

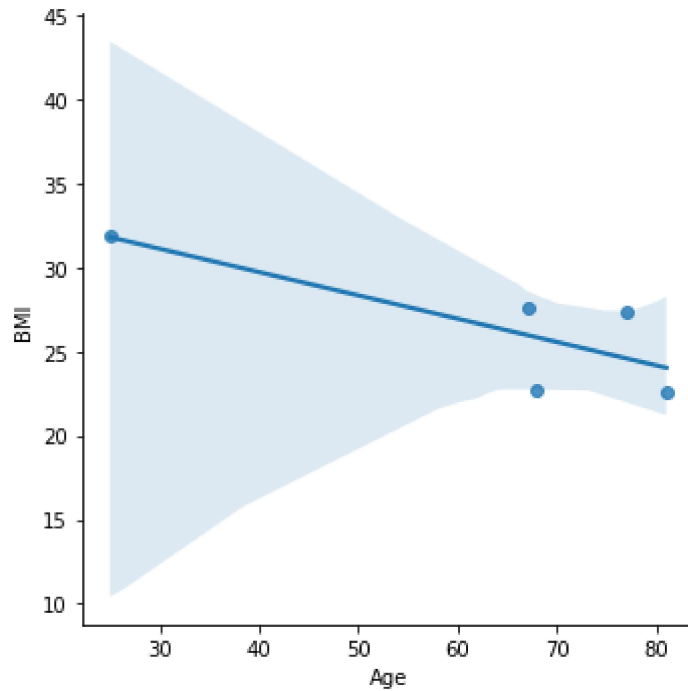
name

```
In [138]: custom_function_death()
```



```
In [78]: patientz = {"age" : [25, 67, 68, 77, 81], "bmi" : [31.95, 27.56, 22.73, 27.41, 22.73]}

df =pd.DataFrame(patientz)
sns.lmplot(x="age",y="bmi",data=df)
plt.xlabel("Age")
plt.ylabel("BMI")
jitter=False
plt.show()
```



In []:

In [98]:

In []:

In []:

In []:

In []: