

Actividad 2 Estadísticas

Equipo:

Ricardo Yadel Murillo Pérez,

José Luis Méndez Becerril,

Martín Ramírez Fuentes y

Plata Luna Iveth Vanessa.

Análisis estadístico

Los campos que se encuentran en el archivo es el año, son los siguientes: el estado, el tipo de crimen, el crimen y el número de crímenes.

```
In [111]: listacamos = crimen[["Year", "Type of Crime", "Crime", "State", "Count"]]
listacamos.head(10)
```

Out[111]:

	Year	Type of Crime	Crime	State	Count
0	1960	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	406
1	1961	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	427
2	1962	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	316
3	1963	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	340
4	1964	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	316
5	1965	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	395
6	1966	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	384
7	1967	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	415
8	1968	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	421
9	1969	Violent Crime	Murder and nonnegligent Manslaughter	Alabama	485

```
In [112]: listacamos.tail(10)
```

Out[112]:

	Year	Type of Crime	Crime	State	Count
16412	1996	Property Crime	Motor vehicle theft	Wyoming	669
16413	1997	Property Crime	Motor vehicle theft	Wyoming	647
16414	1998	Property Crime	Motor vehicle theft	Wyoming	669
16415	1999	Property Crime	Motor vehicle theft	Wyoming	596
16416	2000	Property Crime	Motor vehicle theft	Wyoming	573
16417	2001	Property Crime	Motor vehicle theft	Wyoming	696
16418	2002	Property Crime	Motor vehicle theft	Wyoming	743
16419	2003	Property Crime	Motor vehicle theft	Wyoming	798
16420	2004	Property Crime	Motor vehicle theft	Wyoming	799

A partir de estos datos realizamos las siguientes preguntas.

¿En qué rango de años comprende el análisis?

En este análisis de datos se presentan datos desde 1960 hasta el 2005.

```
In [108]: listaaños = crimen[["Year"]]
          listaaños.head(10)

Out[108]:
```

	Year
0	1960
1	1961
2	1962
3	1963
4	1964
5	1965
6	1966
7	1967
8	1968
9	1969

```
In [109]: listaaños.tail(10)

Out[109]:
```

	Year
16412	1996
16413	1997
16414	1998
16415	1999
16416	2000
16417	2001
16418	2002
16419	2003
16420	2004
16421	2005

Los estados que se encuentran en este análisis son los siguientes:

```
: crimenes.State.value_counts()

: Alabama 322
  Pennsylvania 322
  Nevada 322
  New Hampshire 322
  New Jersey 322
  New Mexico 322
  New York 322
  North Carolina 322
  North Dakota 322
  Ohio 322
  Oklahoma 322
  Oregon 322
  Rhode Island 322
  Montana 322
  South Carolina 322
  South Dakota 322
  Tennessee 322
  Texas 322
  Utah 322
  Vermont 322
  Virginia 322
  Washington 322
  West Virginia 322
  Wisconsin 322
  Nebraska 322
  Missouri 322
  Alaska 322
  Idaho 322
  Arizona 322
  Arkansas 322
  California 322
  Colorado 322
  Connecticut 322
  Delaware 322
  District of Columbia 322
  Florida 322
  Georgia 322
  Hawaii 322
  Illinois 322
  Mississippi 322
  ..
```

```

Indiana      322
Iowa         322
Kansas       322
Kentucky     322
Louisiana    322
Maine        322
Maryland     322
Massachusetts 322
Michigan     322
Minnesota    322
Wyoming      322
Name: State, dtype: int64

```

Los tipos de crímenes que se presentan en la estadística son los siguientes:

```

In [37]: crímenes.Crime.value_counts()
Out[37]: Murder and nonnegligent Manslaughter  2346
Forcible rape 2346
Robbery 2346
Aggravated assault 2346
Burglary 2346
Larceny-theft 2346
Motor vehicle theft 2346
Name: Crime, dtype: int64

```

En el 2005 se presentaron la siguiente cantidad de crímenes en las diversas categorías.

La más alta fue hurto.

```

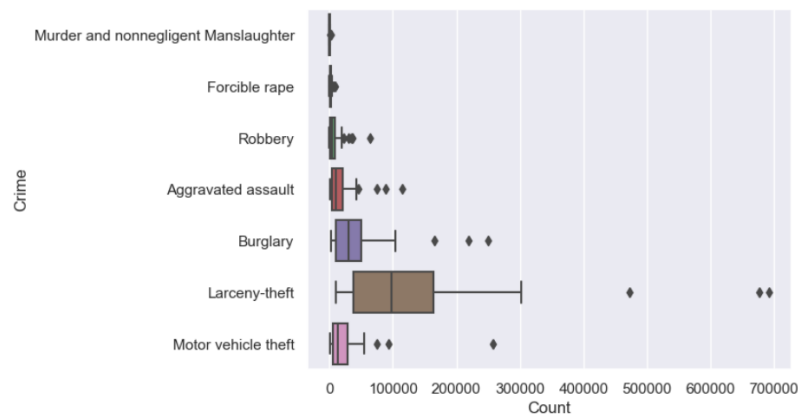
In [103]: # usar el criterio para extraer solo los crímenes de 2005
yr2005 = crímenes[crímenes.Year == 2005]
sns.boxplot(data=yr2005, x='Count', y='Crime')

```

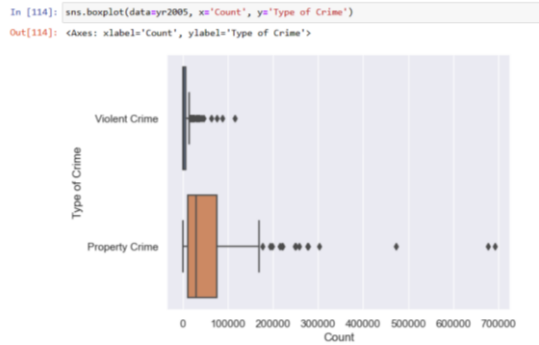
```

Out[103]: <Axes: xlabel='Count', ylabel='Crime'>

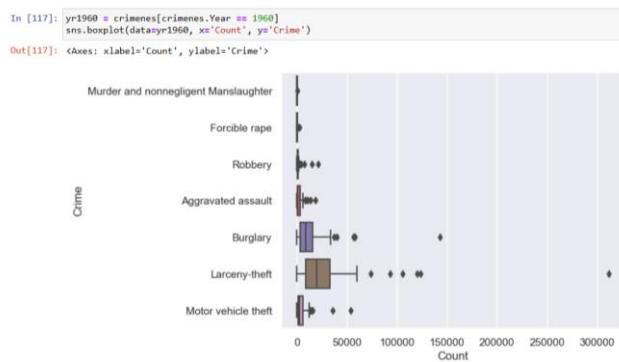
```



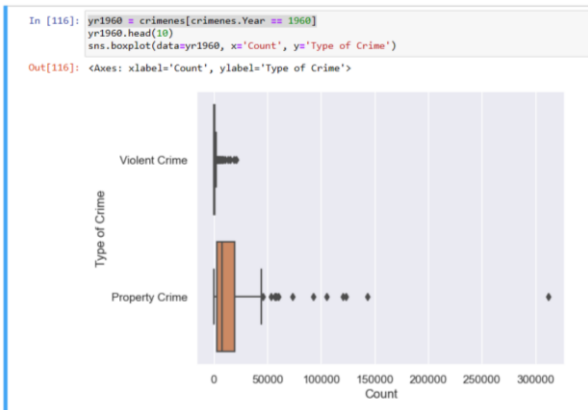
El tipo de crimen más alto en el 2005 fue el crimen a propiedad.



En 1960 el más alto crimen fue el hurto.



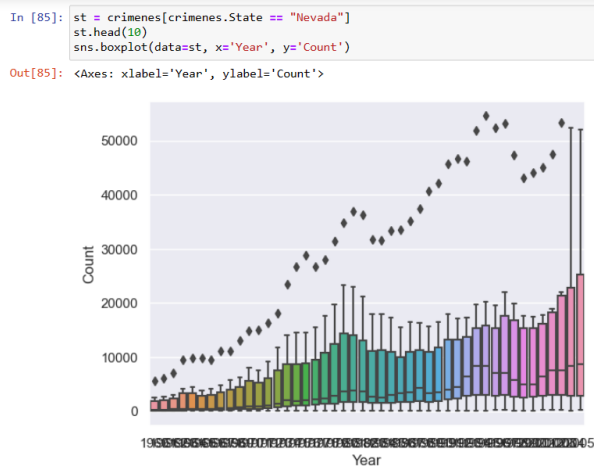
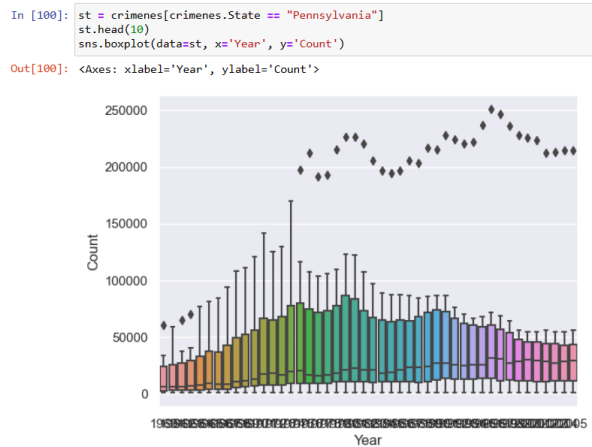
El más alto tipo de crimen en 1960 fue a propiedad.



Específicamente en el estado de Alaska fue el más alto el hurto.

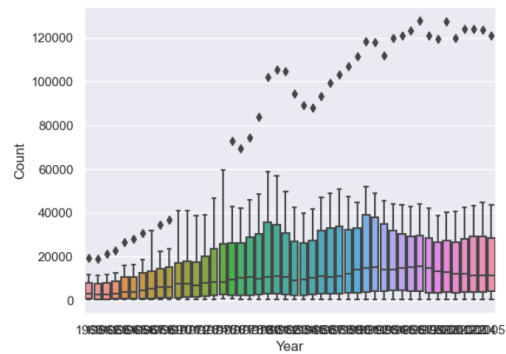


En los estados de Pennsylvania, Nevada y Alabama las estadísticas del número de crimen tienden a incrementarse con forme pasa el tiempo. De estos tres estados Pennsylvania es el más alto.



```
In [118]: st = crimenes[crimenes.State == "Alabama"]
st.head(10)
sns.boxplot(data=st, x='Year', y='Count')

Out[118]: <Axes: xlabel='Year', ylabel='Count'>
```

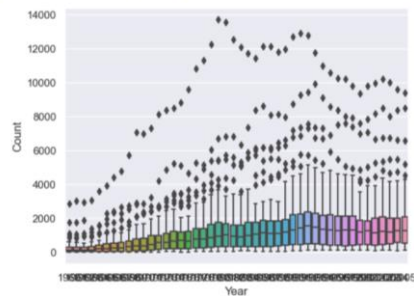


El crimen por asesinato tiende a tener un máximo por los años 80's.

```
In [44]: # usar el criterio para extraer solo crímenes del crimen
cm = crimenes[crimenes.Crime == "Murder and nonnegligent manslaughter"]
```

```
In [75]: sns.boxplot(data=cm, x='Year', y='Count')

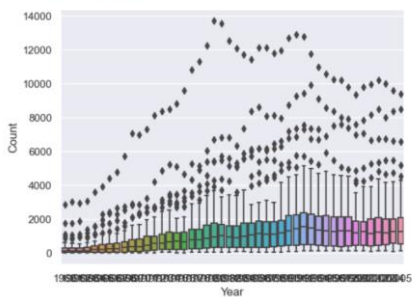
Out[75]: <Axes: xlabel='Year', ylabel='Count'>
```



El crimen de violación tiende a tener su máximo también en los años 80's.

```
In [76]: cm = crimenes[crimenes.Crime == "Forcible rape"]
cm.head(10)
sns.boxplot(data=cm, x='Year', y='Count')

Out[76]: <Axes: xlabel='Year', ylabel='Count'>
```

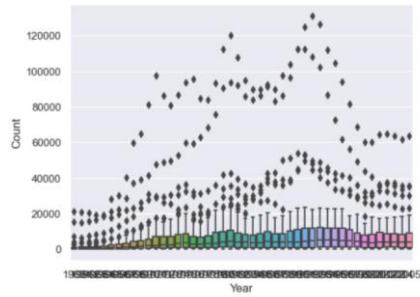


El crimen de robo tiende su altura máxima en los años 90's.

```

In [77]: cm = crimenes[crimenes.Crime == "Robbery"]
cm.head(10)
sns.boxplot(data=cm, xs='Year', ys='Count')
Out[77]: <Axes: xlabel='Year', ylabel='Count'>

```

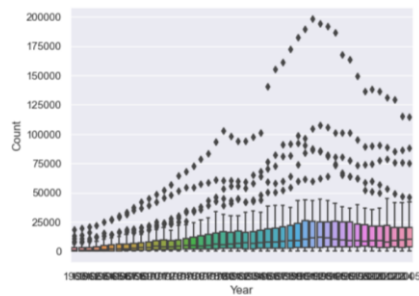


El asalto agravado tiende a incrementarse en los 90's.

```

In [78]: cm = crimenes[crimenes.Crime == "Aggravated assault"]
cm.head(10)
sns.boxplot(data=cm, xs='Year', ys='Count')
Out[78]: <Axes: xlabel='Year', ylabel='Count'>

```

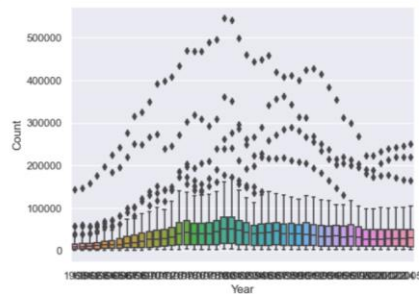


El robo tiende su máximo en los 80's.

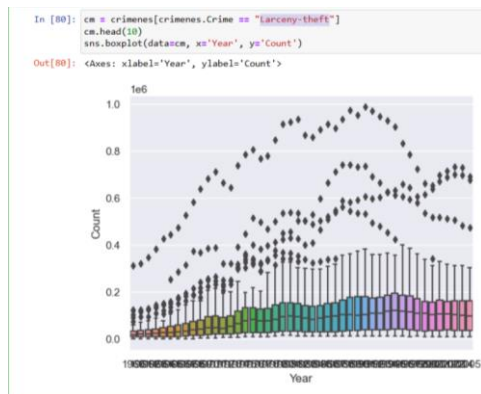
```

In [79]: cm = crimenes[crimenes.Crime == "Burglary"]
cm.head(10)
sns.boxplot(data=cm, xs='Year', ys='Count')
Out[79]: <Axes: xlabel='Year', ylabel='Count'>

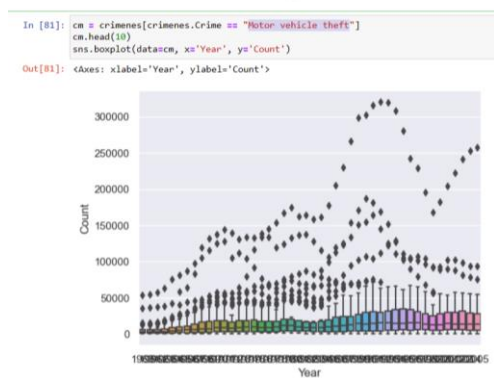
```



El hurto y robo su máximo fue entre los 80's y 90's.



El robo de vehículos se incremento en los 80's y 90's.



El crimen tiende a bajar y posteriormente a subir.

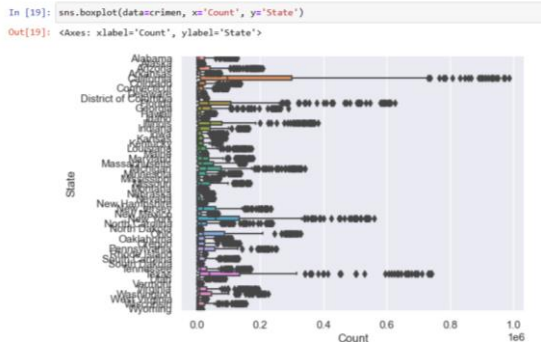
El estado que presenta más crímenes es Alabama.

```
In [26]: # Notación equivalente a la anterior
crimen['Count'].describe()
```

```
Out[26]: count    16422.000000
         mean     29605.496164
         std      74399.826883
         min        0.000000
         25%      669.000000
         50%     4630.000000
         75%     24114.000000
         max     986120.000000
         Name: Count, dtype: float64
```

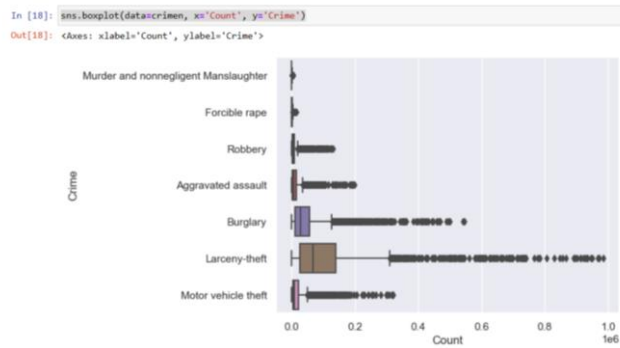
```
In [27]: # Notación equivalente a la anterior
crimen['State'].describe()
```

```
Out[27]: count      16422
         unique        51
         top    Alabama
         freq       322
         Name: State, dtype: object
```

Conclusión

El crimen tipo de crimen que se comete más es el de crimen a hurto a propiedad, en su mayoría se cometieron entre los años 80's y 90's.



Código

```
import pandas as pd

crimen = pd.read_csv('crime.csv')

crimen.head(16421)

crimen.describe()

crimen.State.describe()

crimen.Crime.describe()

crimen.Year.describe()

# Gráfica del histograma

crimen.Year.hist();

# box plot

crimen.boxplot(column='Year');

import seaborn as sns

import numpy as np

crimen = pd.read_csv('crime.csv')

sns.set(color_codes = True)

sns.set_theme(style="darkgrid")

# dropna() devuelve una serie sin

# valores nulos en el campo Age

yearv = crimen.Year.dropna()

sns.histplot(yearv, kde=True);

crimenv = crimen[["State", "Year", "Count"]]

crimenv.head()
```

```
sns.boxplot(data=crimen, x='Count', y='Type of Crime')
```

```
sns.boxplot(data=crimen, x='Count', y='Type of Crime')
```

```
sns.boxplot(data=crimen, x='Count', y='Type of Crime')
```

```
sns.boxplot(data=crimen, x='Count', y='Crime')
```

```
sns.boxplot(data=crimen, x='Count', y='State')
```

```
crimen.Crime.describe()
```

```
listatipocrimen = crimen[["Type of Crime"]]
```

```
listatipocrimen.head(1000)
```

```
listaestado = crimen[["State"]]
```

```
listaestado.head(1000)
```

```
listacrimen = crimen[["Crime"]]
```

```
listacrimen.head(1000)
```

```
listacrimen = crimen[["Crime"]]
```

```
listacrimen.head(1000)
```

```
listaaños = c
```

```
crimen[["Count"]]
```

```
listaaños.head(1000)
```

```
# Notación equivalente a la anterior
```

```
crimen['Count'].describe()
```

```
# Notación equivalente a la anterior
```

```
crimen['State'].describe()
```

```
# Notación equivalente a la anterior
```

```
crimen['Type of Crime'].describe()
```

```

# Calculo de la media
crimen[['Count']].median(axis=0, skipna=True, numeric_only=False)

listaaños = crimen[["Year"]]

listaaños.head(10)

listaaños.tail(10)

listacount = crimen[["Count"]]

listacount.head(10)

listacount.tail(10)

# crea un diccionario
dates_dic = {'Year': listaaños, 'Crime':listacrimen,
             'Type of Crime': listatipocrimen,
             'Count': listacount}

dates_dic

listacampos = crimen[["Year", "Type of Crime", "Crime", "State", "Count"]]

listacampos.head(10)

listacampos.tail(10)

crimenes = pd.D
ataFrame(listacampos)

crimenes

crimenes.State.value_counts()

crimenes.Year.value_counts()

crimenes.Count.value_counts()

# usar el criterio para extraer solo los crímenes de 2005

```

```

yr2005 = crimenes[crimenes.Year == 2005]

sns.boxplot(data=yr2005, x='Count', y='Crime')

yr2005.head(10)

sns.boxplot(data=yr2005, x='Count', y='Type of Crime')

yr1960 = crimenes[crimenes.Year == 1960]

sns.boxplot(data=yr1960, x='Count', y='Crime')

yr1960 = crimenes[crimenes.Year == 1960]

yr1960.head(10)

sns.boxplot(data=yr1960, x='Count', y='Type of Crime')

# usar el criterio para extraer solo crimenes del estado de Alaska

st = crimenes[crimenes.State == "Alaska"]

sns.boxplot(data=st, x='Count', y='Crime')

st.head(10)

st = crimenes[crimenes.State == "Pennsylvania"]

st.head(10)

sns.boxplot(data=st, x='Year', y='Count')

st = crimenes[crimenes.State == "Nevada"]

st.head(10)

sns.boxplot(data=st, x='Year', y='Count')

st = crimenes[crimenes.State == "Alabama"]

st.head(10)

sns.boxplot(data=st, x='Year', y='Count')

# usar el criterio para extraer solo crimenes del crimen

```

```

cm = crimenes[crimenes.Crime == "Murder and nonnegligent Manslaughter"]

# box plot

cm.boxplot(column='Year');

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Forcible rape"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Robbery"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Aggravated assault"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Burglary"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Larceny-theft"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

cm = crimenes[crimenes.Crime == "Motor vehicle theft"]

cm.head(10)

sns.boxplot(data=cm, x='Year', y='Count')

```