In [1]:

```
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as st
import numpy as np
from scipy.stats import linregress

#study data file
DemoT = "DemoClean.csv"
DemoIL = "DemoIL.csv"

#reading the data
CountryDemo = pd.read_csv(DemoT)
Illinois = pd.read_csv(DemoIL)

CountryDemo.head()

#Country data
```

Out[1]:

	Date	State	Cases_Total	Cases_White	Cases_Black	Cases_LatinX	Cases_Asian	Cases_
0	20200412	AL	3582	1657.0	1279.0	NaN	37.0	
1	20200412	AR	1280	814.0	324.0	NaN	20.0	
2	20200412	AZ	3539	661.0	65.0	286.0	NaN	
3	20200412	CA	21794	4323.0	945.0	4729.0	1756.0	
4	20200412	CT	11510	3030.0	993.0	1277.0	124.0	

5 rows × 34 columns

In [2]:

```
# Percentage of deathWhitePeople to total cases - per State

avgPercdeaths = CountryDemo["%DeathsTotalCases"].mean()
print(f"Average percentage of total deaths = {round(avgPercdeaths,2)}%")

avgPercdeaths = CountryDemo["%DeathsWhite"].mean()
print(f"Average percentage of white deaths = {round(avgPercdeaths,2)}%")

CountryWork = CountryDemo.loc[:, ["State", "%DeathsWhite"]]

mean = CountryWork.groupby(["State"]).mean()["%DeathsWhite"]
median = CountryWork.groupby(["State"]).median()["%DeathsWhite"]
variance = CountryWork.groupby(["State"]).var()["%DeathsWhite"]
sdev = CountryWork.groupby(["State"]).std()["%DeathsWhite"]
sem = CountryWork.groupby(["State"]).sem()["%DeathsWhite"]

sumStatDf = pd.DataFrame({"mean %deaths":mean, "median %deaths":median, "variance":varianc
e, "sd":sdev, "sem":sem})

sumStat = sumStatDf.round(2)

sumStat.head()
```

```
Average percentage of total deaths = 3.98%
Average percentage of white deaths = 1.97%
```

Out[2]:

State	mean %deaths	median %deaths	variance	sd	sem
AK	0.64	0.90	0.20	0.45	0.09
AL	1.50	1.60	0.22	0.47	0.09
AR	0.88	0.95	0.11	0.33	0.06
AZ	1.57	1.75	0.31	0.56	0.10
CA	1.10	1.15	0.05	0.22	0.04

In [3]:

```
# Percentage of deathBlackPeople to total cases - per State

avgPercdeaths = CountryDemo["%DeathsBlack"].mean()
print(f"Average percentage of black deaths = {round(avgPercdeaths,2)}%")

CountryWork = CountryDemo.loc[:, ["State", "%DeathsBlack"]]

mean = CountryWork.groupby(["State"]).mean()["%DeathsBlack"]
median = CountryWork.groupby(["State"]).median()["%DeathsBlack"]
variance = CountryWork.groupby(["State"]).var()["%DeathsBlack"]
sdev = CountryWork.groupby(["State"]).std()["%DeathsBlack"]
sem = CountryWork.groupby(["State"]).sem()["%DeathsBlack"]

sumStatDf = pd.DataFrame({"mean %deaths":mean, "median %deaths":median, "variance":varianc
e, "sd":sdev, "sem":sem})

sumStat = sumStatDf.round(2)

sumStat.head()
```

```
Average percentage of black deaths = 0.74%
```

Out[3]:

State	mean %deaths	median %deaths	variance	sd	sem
AK	0.00	0.00	0.00	0.00	0.00
AL	1.38	1.50	0.15	0.39	0.07
AR	0.50	0.50	0.05	0.23	0.04
AZ	0.11	0.10	0.00	0.04	0.01
CA	0.33	0.35	0.01	0.08	0.02

In [4]:

```
# Percentage of deathLatinPeople to total cases - per State

avgPercdeaths = CountryDemo["%DeathsLatin"].mean()
print(f"Average percentage of latin deaths = {round(avgPercdeaths,2)}%")

CountryWork = CountryDemo.loc[:, ["State", "%DeathsLatin"]]

mean = CountryWork.groupby(["State"]).mean()["%DeathsLatin"]
median = CountryWork.groupby(["State"]).median()["%DeathsLatin"]
variance = CountryWork.groupby(["State"]).var()["%DeathsLatin"]
sdev = CountryWork.groupby(["State"]).std()["%DeathsLatin"]
sem = CountryWork.groupby(["State"]).sem()["%DeathsLatin"]

sumStatDf = pd.DataFrame({"mean %deaths":mean, "median %deaths":median, "variance":varianc
e, "sd":sdev, "sem":sem})

sumStat = sumStatDf.round(2)

sumStat.head()
```

```
Average percentage of latin deaths = 0.21%
```

Out[4]:

State	mean %deaths	median %deaths	variance	sd	sem
AK	0.00	0.0	0.00	0.00	0.00
AL	0.00	0.0	0.00	0.00	0.00
AR	0.00	0.0	0.00	0.00	0.00
AZ	0.59	0.6	0.02	0.15	0.03
CA	1.20	1.3	0.05	0.22	0.04

In [5]:

```
# Percentage of deathAsianPeople to total cases - per State

avgPercdeaths = CountryDemo["%DeathsAsian"].mean()
print(f"Average percentage of asian deaths = {round(avgPercdeaths,2)}%")

CountryWork = CountryDemo.loc[:, ["State", "%DeathsAsian"]]

mean = CountryWork.groupby(["State"]).mean()["%DeathsAsian"]
median = CountryWork.groupby(["State"]).median()["%DeathsAsian"]
variance = CountryWork.groupby(["State"]).var()["%DeathsAsian"]
sdev = CountryWork.groupby(["State"]).std()["%DeathsAsian"]
sem = CountryWork.groupby(["State"]).sem()["%DeathsAsian"]

sumStatDf = pd.DataFrame({"mean %deaths":mean, "median %deaths":median, "variance":varianc
e, "sd":sdev, "sem":sem})

sumStat = sumStatDf.round(2)

sumStat.head()
```

```
Average percentage of asian deaths = 0.08%
```

Out[5]:

State	mean %deaths	median %deaths	variance	sd	sem
AK	0.25	0.2	0.05	0.22	0.04
AL	0.01	0.0	0.00	0.03	0.00
AR	0.00	0.0	0.00	0.00	0.00
AZ	0.05	0.0	0.00	0.05	0.01
CA	0.49	0.5	0.01	0.11	0.02

In [6]:

```
#Illinois data - Data Analysis

Illinois.head()
```

Out[6]:

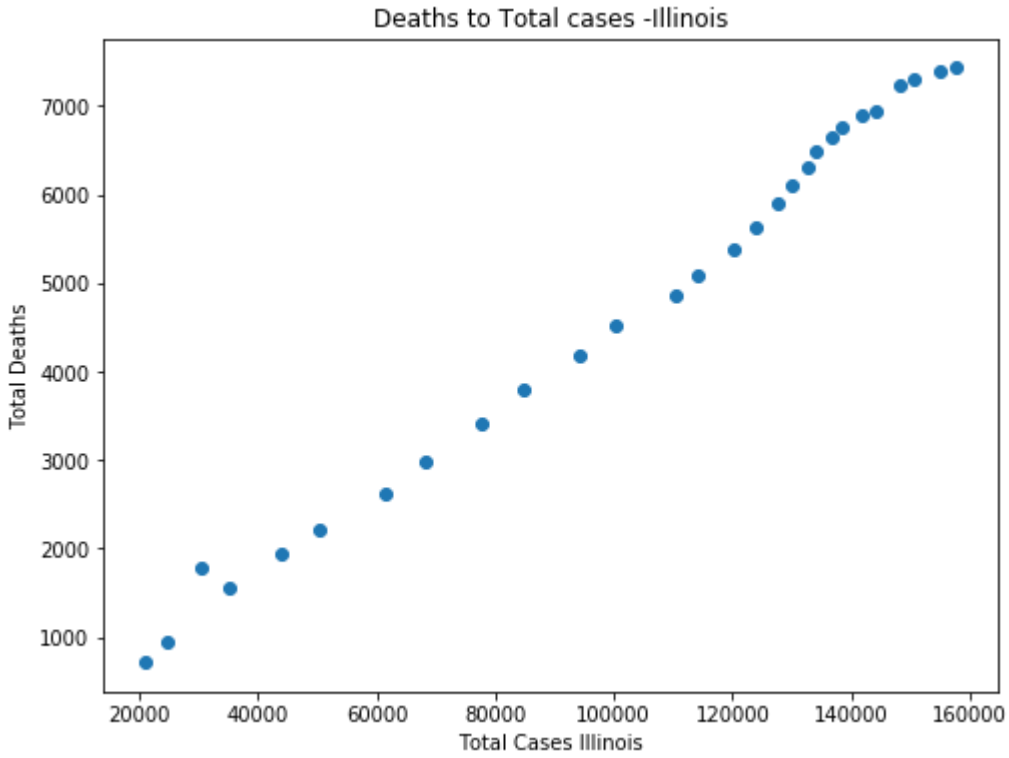
	Date	State	Cases_Total	Cases_White	Cases_Black	Cases_LatinX	Cases_Asian	Cases_
0	20200412	IL	20852	5214	5458	2513	691	
1	20200415	IL	24593	6127	6331	3357	847	
2	20200419	IL	30357	7554	7498	4913	1085	
3	20200422	IL	35108	8573	8504	6195	1244	
4	20200426	IL	43903	10427	9918	8743	1523	

5 rows × 34 columns

In [7]:

```
# Total deaths to total cases -Illinois

avgPercdeaths = Illinois["%DeathsTotalCases"].mean()
print(f"Average percentage of total deaths = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["Cases_Total"], Illinois["Deaths_Total"])
plt.xlabel("Total Cases Illinois")
plt.ylabel("Total Deaths")
plt.title("Deaths to Total cases -Illinois")

plt.show()
```

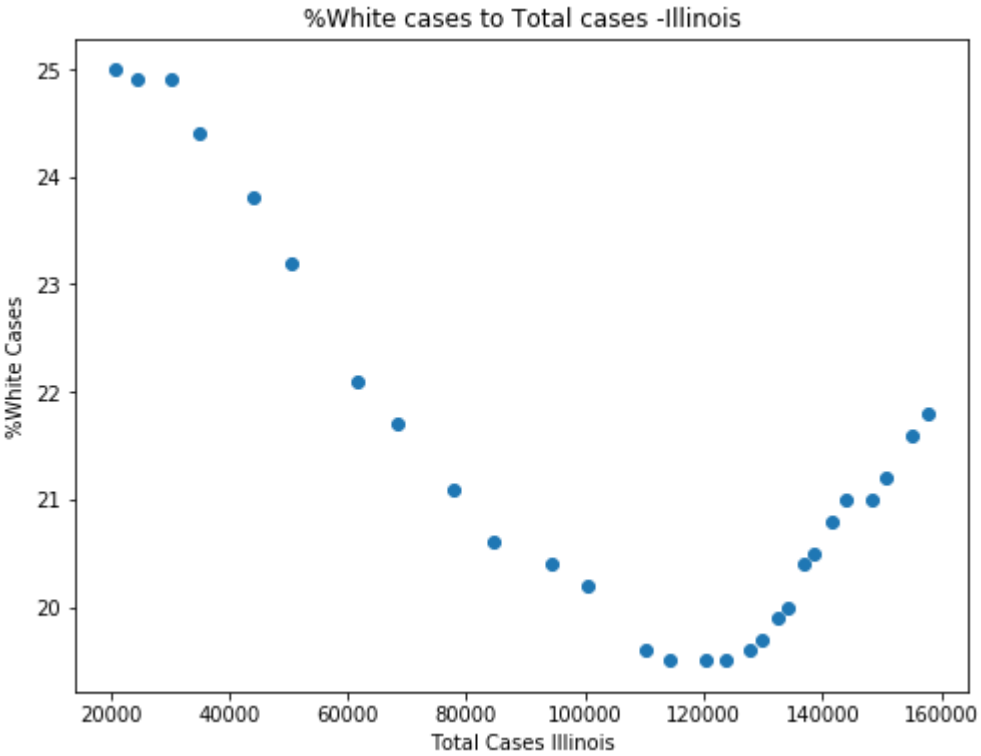
```
Average percentage of total deaths = 4.59%
```



In [28]:

```
# Analysis: % of cases for each race to cases total - Illinois

# % of Whites to Total Cases - Illinois

avgPercdeaths = Illinois["%WhitesTotalCases"].mean()
print(f"Average percentage of white cases = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["Cases_Total"], Illinois["%WhitesTotalCases"])
plt.xlabel("Total Cases Illinois")
plt.ylabel("%White Cases")
plt.title("%White cases to Total cases -Illinois")

plt.show()
```

```
Average percentage of white cases = 21.35%
```



In [18]:

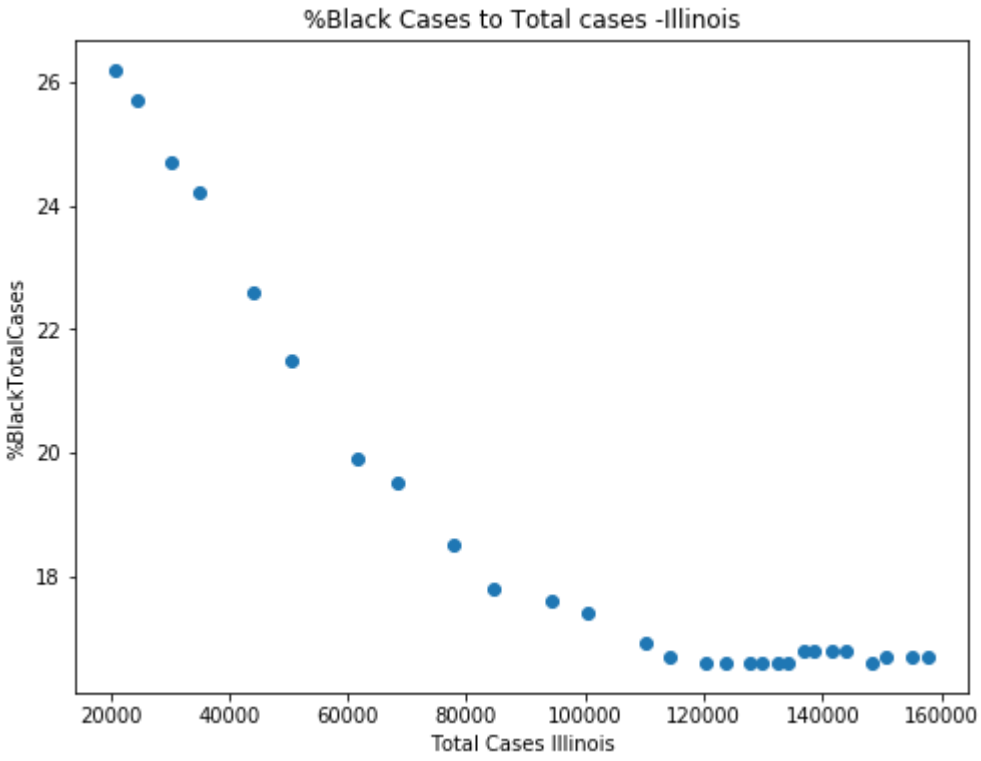
```
#Coefficient of correlation
x = Illinois["Cases_Total"]
y = Illinois["%WhitesTotalCases"]

CoefCor = x.corr(y)
print(f"Coef. of Corr. %White cases to Total cases = {round(CoefCor,2)}")
```

```
Coef. of Corr. %White cases to Total cases = -0.78
```

In [29]:

```
# % of Blacks to Total Cases - Illinois

avgPercdeaths = Illinois["%BlackTotalCases"].mean()
print(f"Average percentage of black cases = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["Cases_Total"], Illinois["%BlackTotalCases"])
plt.xlabel("Total Cases Illinois")
plt.ylabel("%BlackTotalCases")
plt.title("%Black Cases to Total cases -Illinois")

plt.show()
```

```
Average percentage of black cases = 18.67%
```



In [19]:

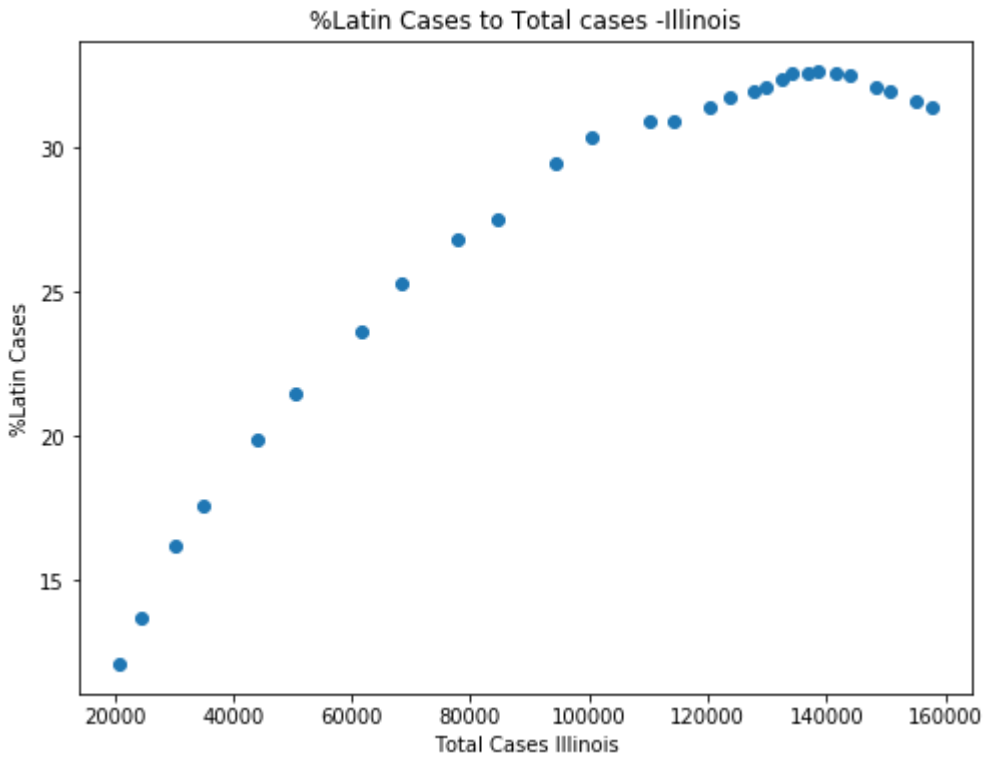
```
#Coefficient of correlation
x = Illinois["Cases_Total"]
y = Illinois["%BlackTotalCases"]

CoefCor = x.corr(y)
print(f"Coef. of Corr. %Black cases to Total cases = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Black cases to Total cases = -0.92
```

In [30]:

```
# % of Latin cases to Total Cases -Illinois

avgPercdeaths = Illinois["%LatinTotalCases"].mean()
print(f"Average percentage of latin cases = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["Cases_Total"], Illinois["%LatinTotalCases"])
plt.xlabel("Total Cases Illinois")
plt.ylabel("%Latin Cases")
plt.title("%Latin Cases to Total cases -Illinois")

plt.show()
```

```
Average percentage of latin cases = 27.7%
```



In [20]:

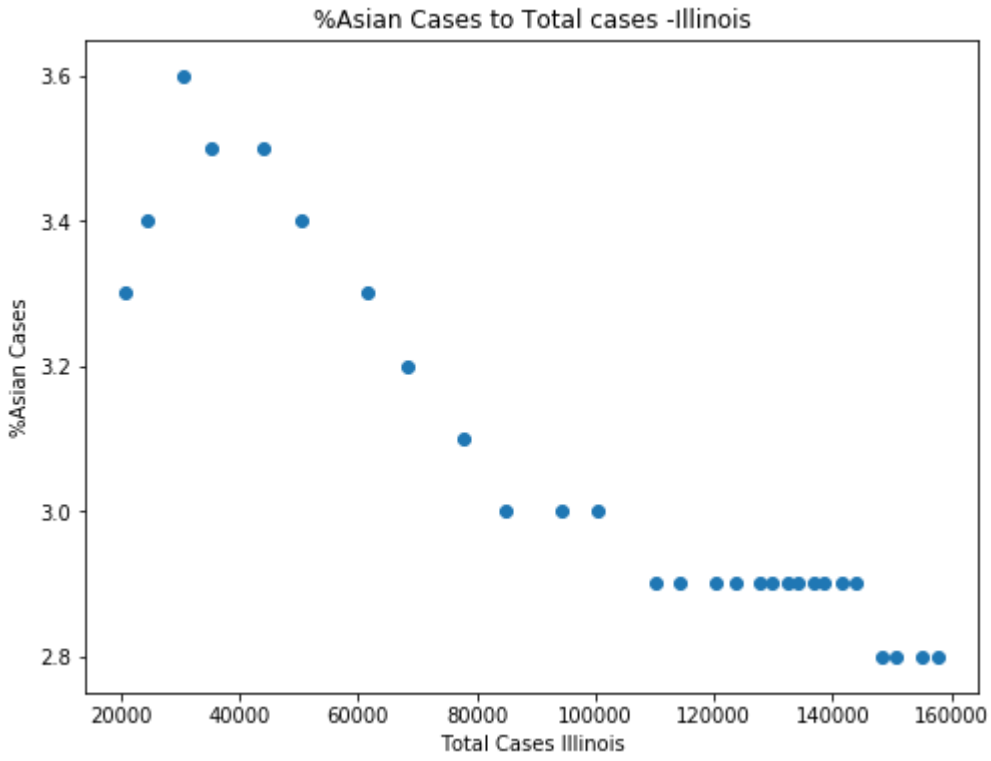
```
#Coefficient of correlation
x = Illinois["Cases_Total"]
y = Illinois["%LatinTotalCases"]

CoefCor = x.corr(y)
print(f"Coef. of Corr. %Latin cases to Total cases = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Latin cases to Total cases = 0.94
```

In [31]:

```
# % of Asian cases to Total Cases -Illinois

avgPercdeaths = Illinois["%AsianTotalCases"].mean()
print(f"Average percentage of asian cases = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["Cases_Total"], Illinois["%AsianTotalCases"])
plt.xlabel("Total Cases Illinois")
plt.ylabel("%Asian Cases")
plt.title("%Asian Cases to Total cases -Illinois")

plt.show()
```

```
Average percentage of asian cases = 3.05%
```



In [21]:

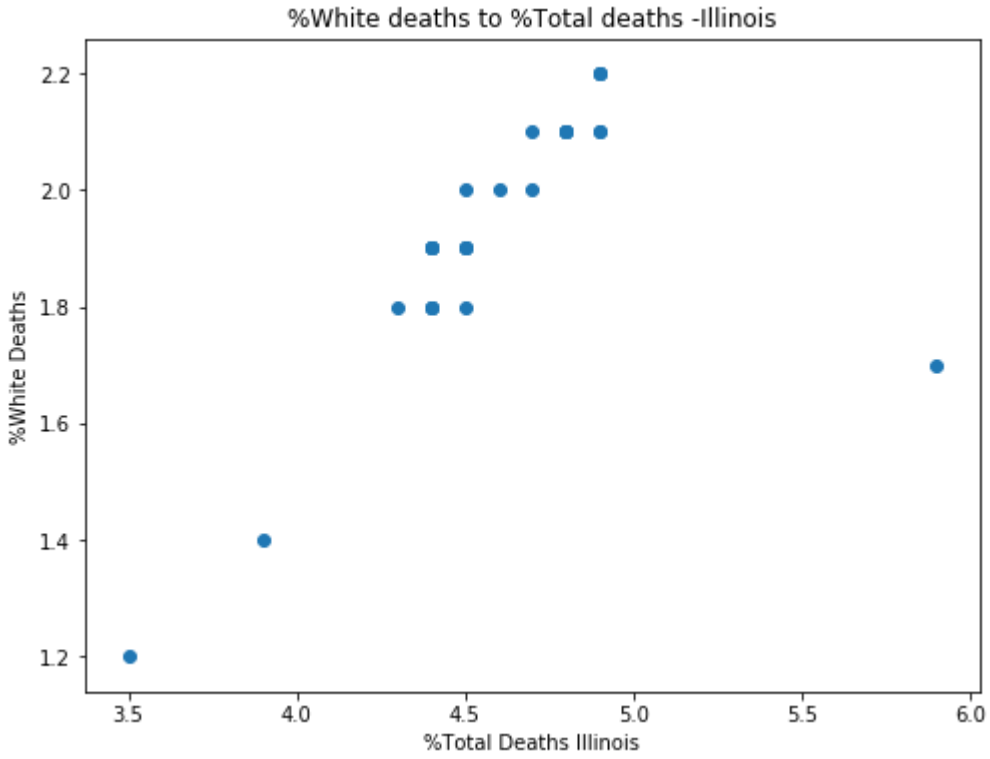
```
#Coefficient of correlation
x = Illinois["Cases_Total"]
y = Illinois["%AsianTotalCases"]

CoefCor = x.corr(y)
print(f"Coef. of Corr. %Asian cases to Total cases = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Asian cases to Total cases = -0.94
```

In [26]:

```
# Analysis: % of deaths for each race to % of deaths total

# % of White deaths to % Total deaths -Illinois

avgPercdeaths = Illinois["%DeathsWhite"].mean()
print(f"Average percentage of White deaths = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["%DeathsTotalCases"], Illinois["%DeathsWhite"])
plt.xlabel("%Total Deaths Illinois")
plt.ylabel("%White Deaths")
plt.title("%White deaths to %Total deaths -Illinois")

plt.show()
```

```
Average percentage of White deaths = 1.93%
```



In [22]:

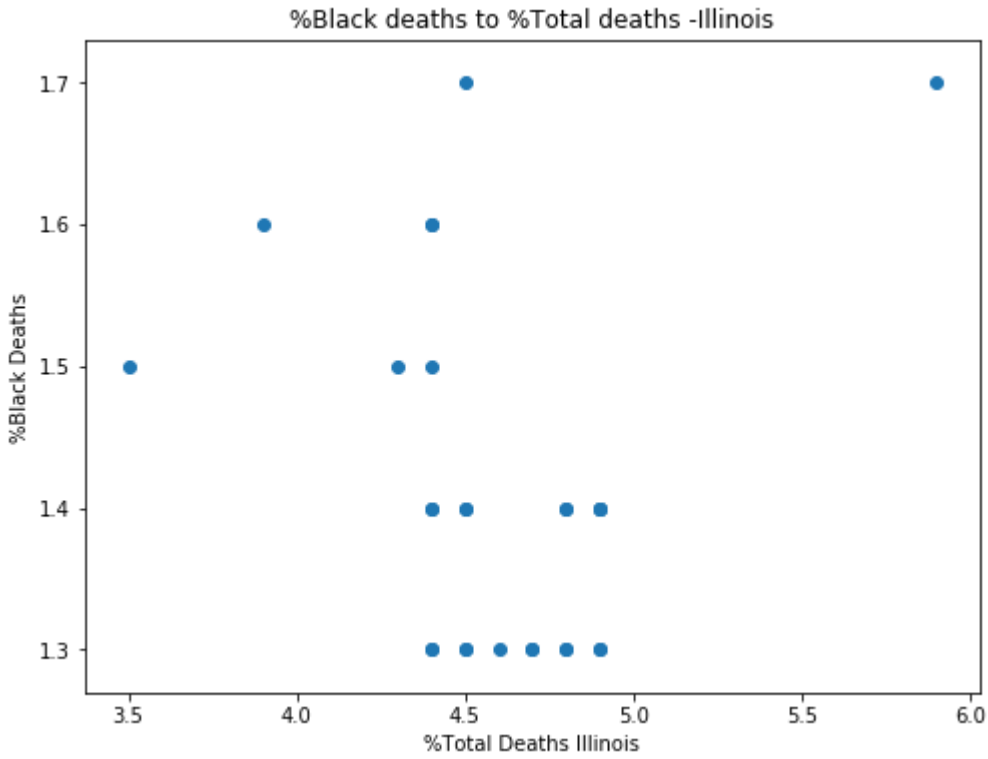
```
#Coefficient of correlation
x = Illinois["%DeathsTotalCases"]
y = Illinois["%DeathsWhite"]
CoefCor = x.corr(y)
print(f"Coef. of Corr. %Deaths Whites to %Deaths total = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Deaths Whites to %Deaths total = 0.62
```

In [27]:

```
# % of Black deaths to % Total deaths -Illinois

avgPercdeaths = Illinois["%DeathsBlack"].mean()
print(f"Average percentage of Black deaths = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["%DeathsTotalCases"], Illinois["%DeathsBlack"])
plt.xlabel("%Total Deaths Illinois")
plt.ylabel("%Black Deaths")
plt.title("%Black deaths to %Total deaths -Illinois")

plt.show()
```

```
Average percentage of Black deaths = 1.41%
```



In [24]:

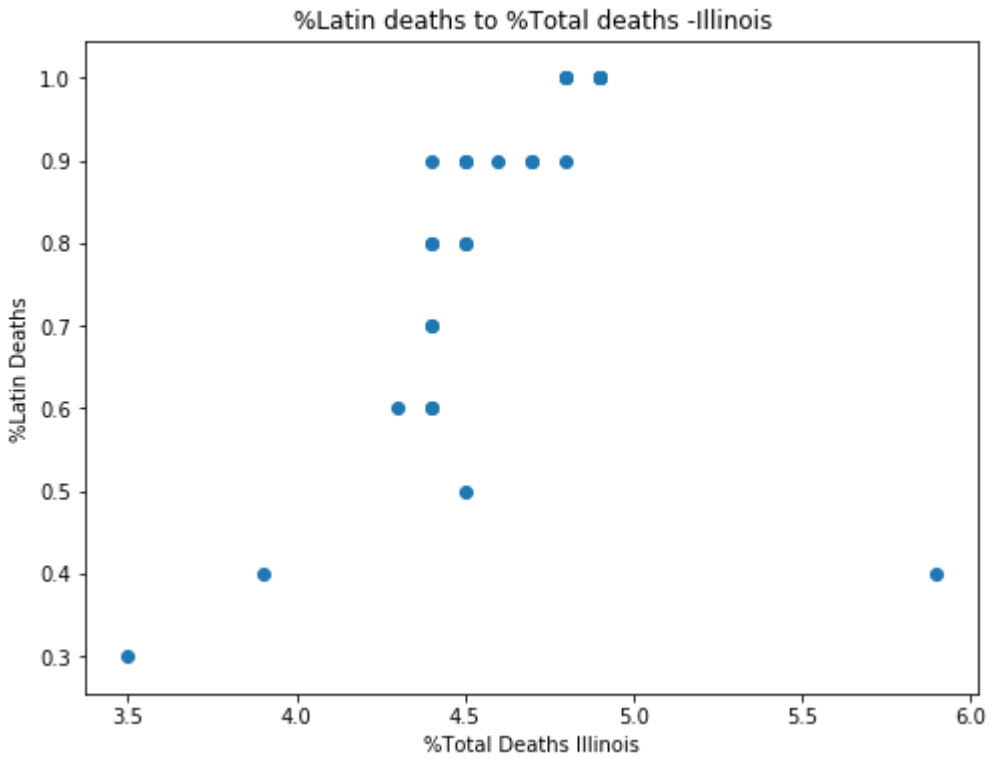
```
#Coefficient of correlation
x = Illinois["%DeathsTotalCases"]
y = Illinois["%DeathsBlack"]
CoefCor = x.corr(y)
print(f"Coef. of Corr. %Deaths Blacks to %Deaths total = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Deaths Blacks to %Deaths total = -0.06
```

In [32]:

```
# % of Latin deaths to % Total deaths -Illinois

avgPercdeaths = Illinois["%DeathsLatin"].mean()
print(f"Average percentage of Latin deaths = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["%DeathsTotalCases"], Illinois["%DeathsLatin"])
plt.xlabel("%Total Deaths Illinois")
plt.ylabel("%Latin Deaths")
plt.title("%Latin deaths to %Total deaths -Illinois")

plt.show()
```

```
Average percentage of Latin deaths = 0.8%
```



In [33]:

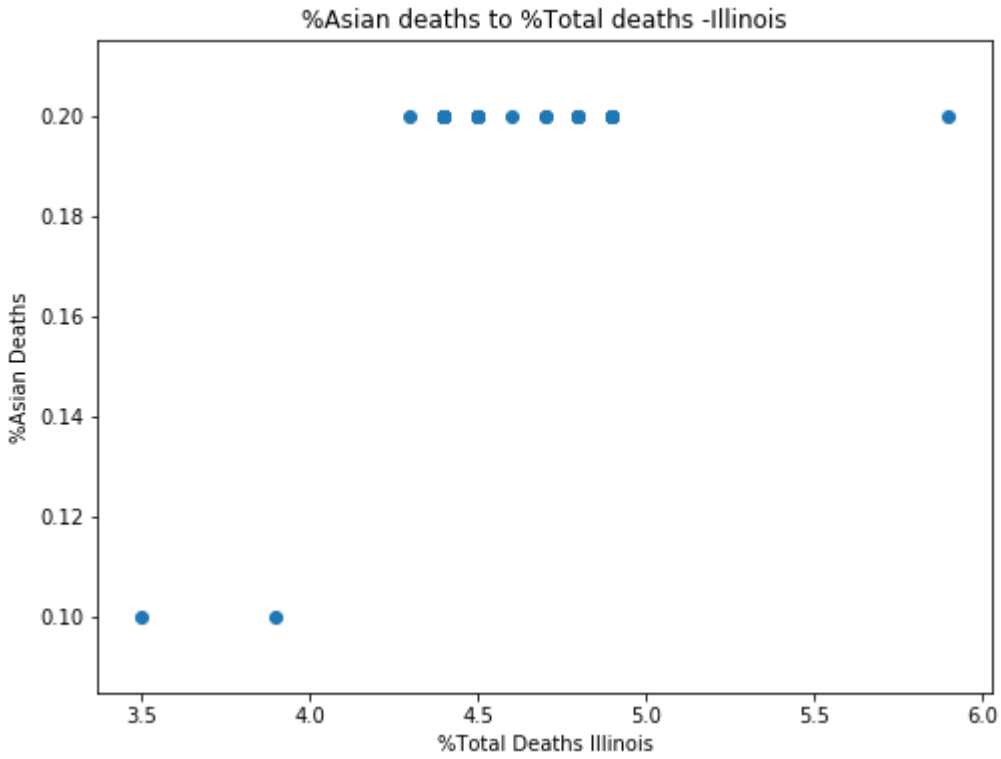
```
#Coefficient of correlation
x = Illinois["%DeathsTotalCases"]
y = Illinois["%DeathsLatin"]
CoefCor = x.corr(y)
print(f"Coef. of Corr. %Deaths Latin to %Deaths total = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Deaths Latin to %Deaths total = 0.41
```

In [34]:

```
# % of Asian deaths to % Total deaths -Illinois

avgPercdeaths = Illinois["%DeathsAsian"].mean()
print(f"Average percentage of Asian deaths = {round(avgPercdeaths,2)}%")

plt.figure(figsize = (8,6))

plt.scatter(Illinois["%DeathsTotalCases"], Illinois["%DeathsAsian"])
plt.xlabel("%Total Deaths Illinois")
plt.ylabel("%Asian Deaths")
plt.title("%Asian deaths to %Total deaths -Illinois")

plt.show()
```

```
Average percentage of Asian deaths = 0.19%
```



In [35]:

```
#Coefficient of correlation
x = Illinois["%DeathsTotalCases"]
y = Illinois["%DeathsAsian"]
CoefCor = x.corr(y)
print(f"Coef. of Corr. %Deaths Asian to %Deaths total = {round(CoefCor,2)}")
```

```
Coef. of Corr. %Deaths Asian to %Deaths total = 0.62
```

In [ ]: