WST PROJECT

Applications in Data Science

Kgotso Lehari
20530545

Questions

1. List the number of drug related crimes per province in descending order

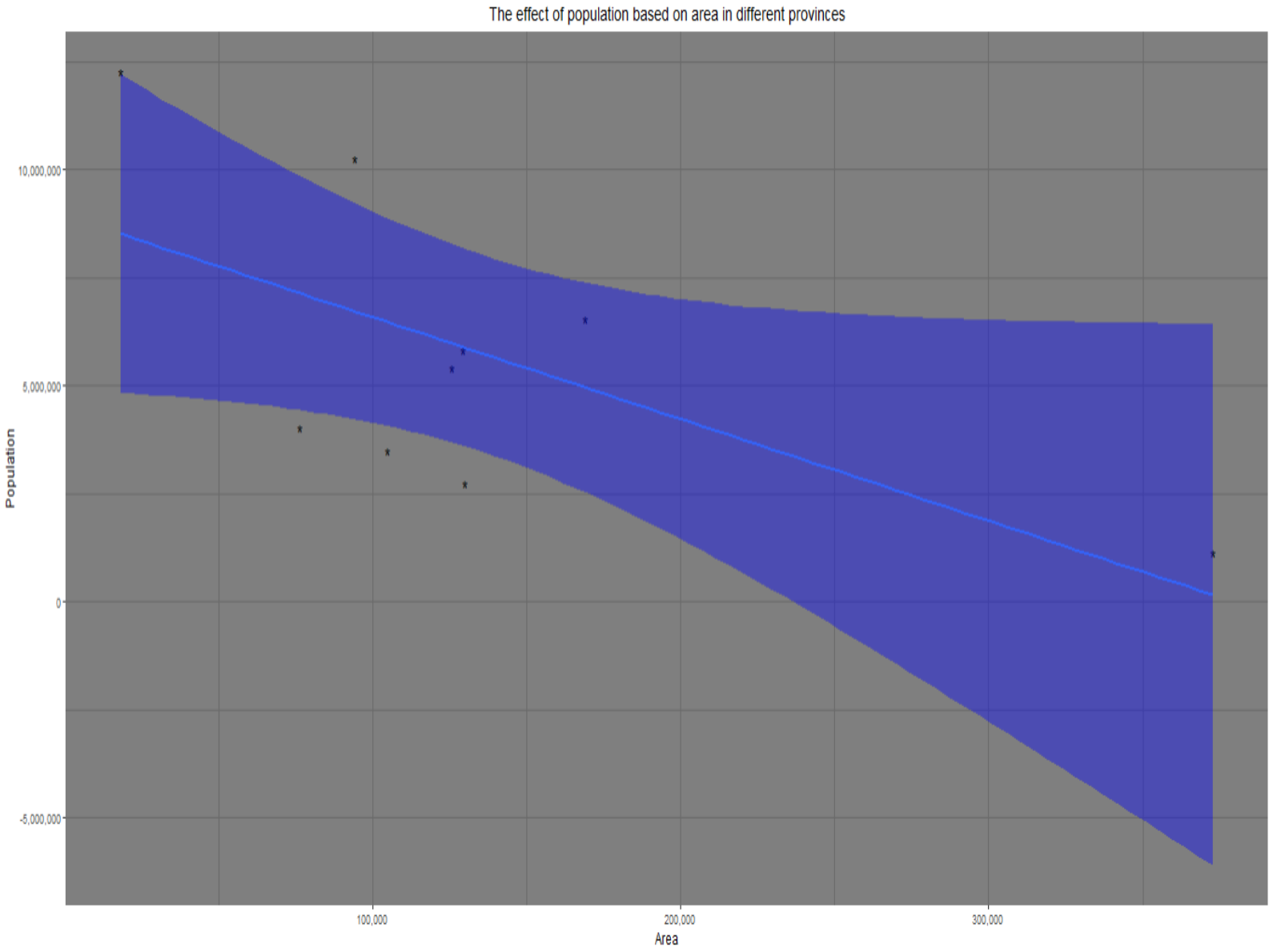
	Province	No_of_crimes
1	Eastern Cape	196
2	Kwazulu/Natal	186
3	Western Cape	150
4	Gauteng	143
5	Free State	110
6	Limpopo	99
7	Northern Cape	91
8	Mpumalanga	86
9	North West	82

2. Count the number of different crimes that occur in Western Cape

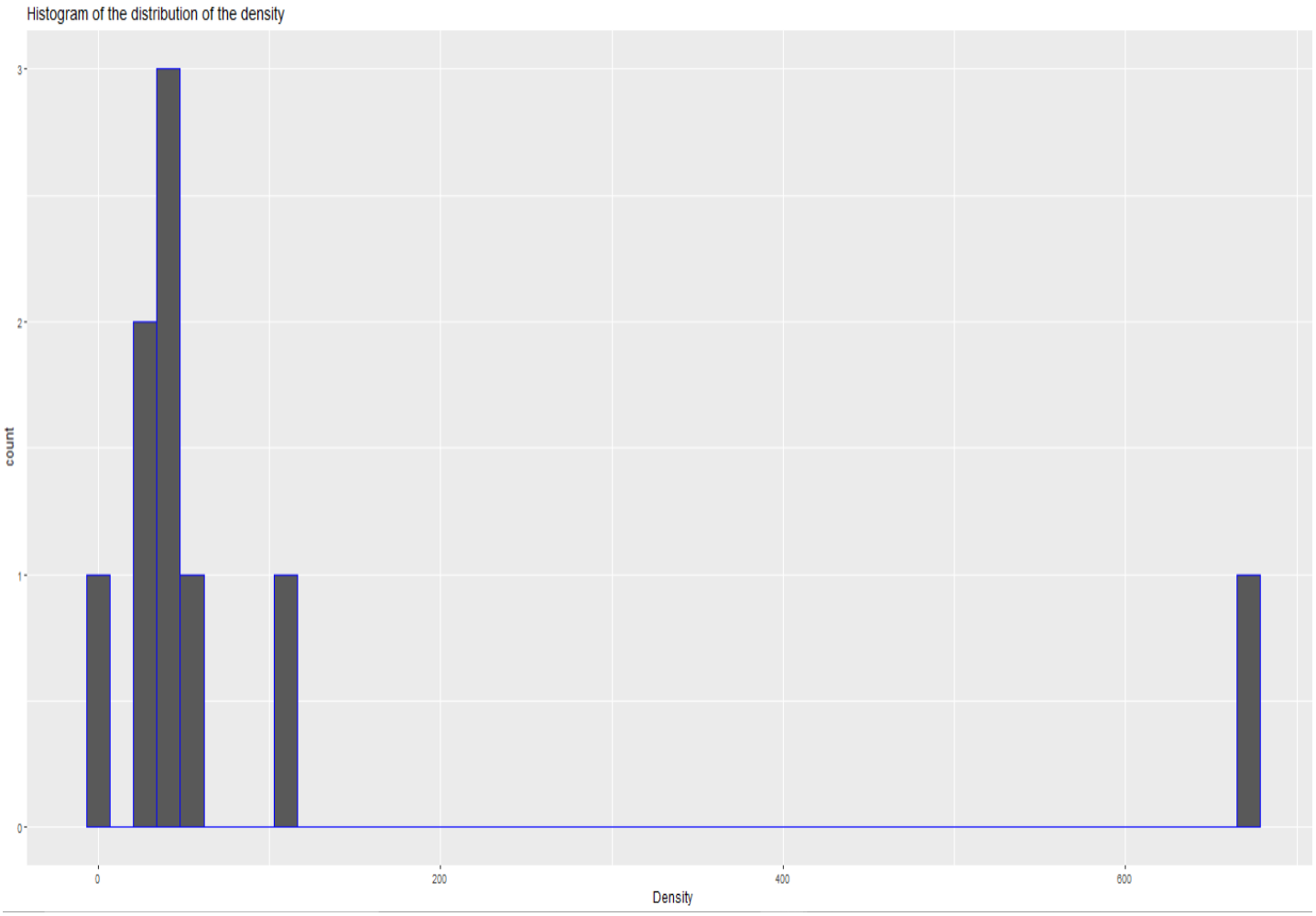
	Category	Num_of_crimes
1	Shoplifting	150
2	Theft of motor vehicle and motorcycle	150
3	Arson	150
4	Truck hijacking	150
5	Sexual offences as result of police action	150
6	Murder	150
7	Common assault	150
8	Stock-theft	150
9	All theft not mentioned elsewhere	150
10	Assault with the intent to inflict grievous bodily harm	150
11	Bank robbery	150
12	Illegal possession of firearms and ammunition	150
13	Robbery of cash in transit	150
14	Burglary at non-residential premises	150
15	Burglary at residential premises	150
16	Drug-related crime	150
17	Theft out of or from motor vehicle	150

18	Malicious damage to property	150
19	Robbery with aggravating circumstances	150
20	Carjacking	150
21	Driving under the influence of alcohol or drugs	150
22	Common robbery	150
23	Attempted murder	150
24	Commercial crime	150
25	Robbery at residential premises	150
26	Sexual Offences	150
27	Robbery at non-residential premises	150

3. What is the effect of population based on different areas?



4. The distribution of the density among provinces



5. Provide the summary of the crimes that occurred during 2015-2016

Min.	0.00
1st Qu.	1.00
Median	11.00
Mean	70.74
3rd Qu.	58.00
Max.	5176.00

Appendix

Q1:

```
q2<- sqldf("select Province , count(Category) AS No_of_crimes
           FROM Crime
          WHERE Category = 'Drug-related crime'
          GROUP BY Province
          ORDER BY No_of_crimes DESC")
```

Q2:

```
q3 <- sqldf("select Category, count(Category) AS Num_of_crimes
           FROM Crime
           WHERE Province = 'Western Cape'
           GROUP BY Category")
```

```
Q3 : area <- Province$Area
pop <- Province$Population

q4 <-  ggplot(data = Province, mapping = aes(x = area, y = pop )) +
  geom_point(shape = '*', size = 5 ) +
  scale_x_continuous( labels = scales::comma) +
  scale_y_continuous(labels = scales::comma)


q4b<- q4 + labs(title = "The effect of population based on area in different provinces", x = "Area", y = "Population") +
geom_smooth(method = lm , fill = 'Blue') +
  theme_dark() +
  theme(plot.title = element_text(hjust = 0.5))
```

Q4 :

```
q5 <- ggplot(aes(x = Density),  data = Province) +
  geom_histogram(bins = 50,color = 'blue') +
  ggtitle('Histogram of the distribution of the density')
```

Q5 : summary(Crime$`2015-2016`)