Data Visualization: Police Arrests In Texas

❖ Introduction

In this project, the goal is to study the police arrests in different areas of Texas and understand the different locations, premises and time that have maximum police arrests from the year 2014 to Spring 2020. This information is important from a safety point of view and to gain understanding of the different locations and time susceptible to crime. With this understanding, it will be easier to implement safety protocols in such locations and also assess whether which area is safe to reside.

The project uses the following database from data.texas.gov (https://www.dallasopendata.com/Public-Safety/Police-Arrests/sdr7-6v3j). This dataset has 84,892 rows and 27 columns.

To better use the information of this project, it is important to be aware of few assumptions:

1) This dataset is not biased in any way.
2) The arrest's current location is not taken into consideration as it has codes like DX and LS whose codes are not known.
3) Age and AgeAtArrestTime have the same values. Therefore, I have taken into consideration AgeAtArrestTime for my visualizations.

❖ Data Preparation / Preprocessing

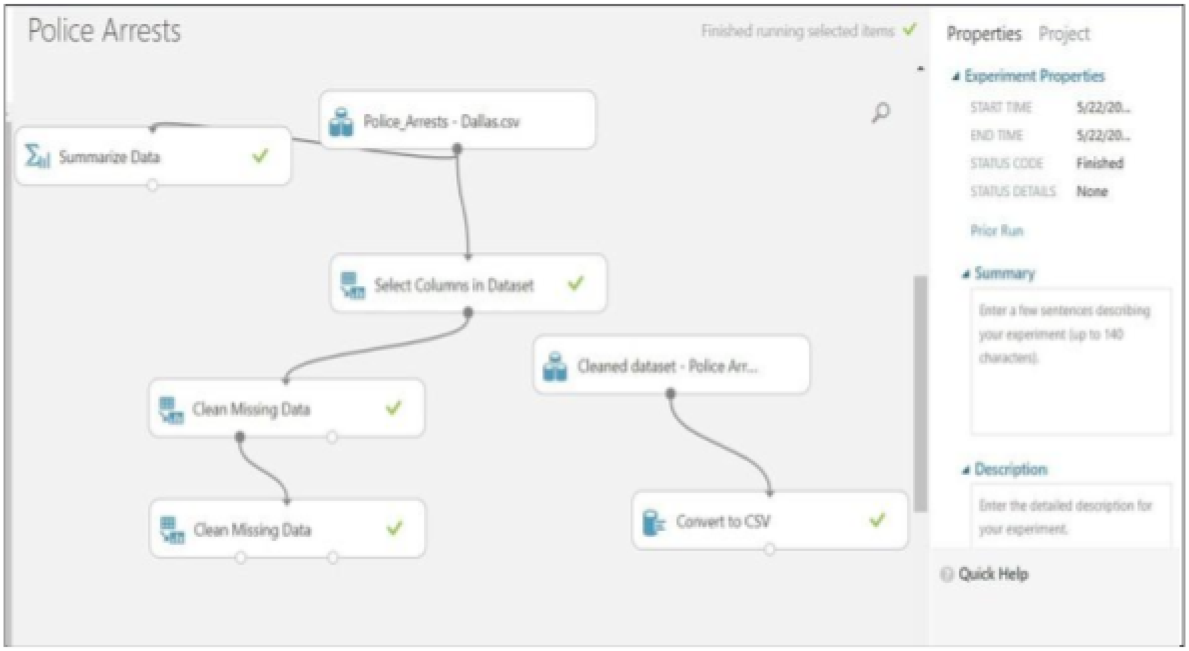
1) Real world data is generally incomplete and lacking attributes values, lacking certain attributes of interest and containing only aggregate data. Noise usually contains errors and outliers, which if not resolved would result in inconsistent data containing discrepancies in codes or names.

2) One of the tasks in data preprocessing is data cleaning. This dataset consisted of some missing values, statistical outliers and noise which required resolution before creating visualization. For this preprocessing task, Microsoft Azure ML was used in the following steps.

Step 1 - The ‘summarize data’ module from Azure ML was used to get an overview of the columns and rows in the dataset. Missing values in features and features of interest were identified.
Step 2 - Using the ‘Select Columns’ module, features that had more than 50% of missing data were removed along with some features like ‘name’ and ‘Phone’ that do not affect the data.
Step 3 - Finally the ‘Clean Missing Data’ module was used to fill missing data by appropriate statistical method.

Replace with median: Calculates the column median value, and uses the median value as the replacement for any missing value in the column. Applies only to columns that have Integer or Float data types. It was used for filling Age, AgeAtArrestTime and Weight features. **Replace with mode**: This method was used to fill the missing data in categorical features.

Step 4 – Finally the dataset was converted to csv file and exported from Azure ML to be used for Tableau. The following figure shows the data preprocessing stages in Azure ML.



START TIME 5/22/20...
END TIME 5/22/20...
STATUS CODE Finished
STATUS DETAILS None

Enter a few sentences describing your experiment (up to 140 characters).

Enter the detailed description for your experiment.

❖ Analysis

Latitude and Longitude were generated by changing the role of the features ArState (Arrest State) and HState (Home State) to geographic.

1) Line graph (Figure 1) – A line graph was used for plotting average age at arrest time and the year wise arrest trend for the period from 2014 to Spring 2020.
A linear trend line was also plotted.

Under Marks tab:

Sum (Number of Records) – was used for labeling purpose Avg
(Age At Arrest Time) was used for detailing purpose.

This graph shows that year wise arrests are decreasing over time. It is also observed that the
The average age at time of arrest is in the range of 34 to 36 years during the period from 2014 to Spring 2020.

2) Maps (Figure 2) – Dual Axis map was used to compare the home states of suspects and the states where they were arrested.

Under Marks tab: For latitude 1, a number of records was used for labeling purposes. HStategroup(Homestategroup) was used for color coding and Hstate was used for detailing and labelingpurpose.Forlatitude 2, number of records was used for labeling purpose. Hstate group (Homestategroup)was used for color coding and ArState was used for detailing and labeling purposes.

3) CombinationCharts:
A barandlinechartwithdualaxis, synchronize axis (Figure 3) – This was used to plot a graph of thetimeofthedaywheremost crime was committed. Under Marks tab: The1stSum(NumberofRecords) – A bar graph was selected and Ar L County was used to colorcodethebargraphcontaining different colors for each county.
The2ndSum(NumberofRecords) - A line graph was selected and a number of records was used for labeling purposes. It wasobservedthatmostcrime was committed at night (especially at 1 am). Moreover, Dallas countyhadthehighestnumber of crime records compared to other counties. LollipopChart(Figure4)–This is again a dual axis chart and was used to plot a graph of crime committed per day.

Under Marks tab: In the2ndSum(NumberofRecords) – Sum (Number of records) was used for labelling purpose.AfilterofArLCity(Arrest Location City) and a single value dropdown option was selectedtofilterthearrestlocation cities. With the help of this filter, the crime committed per day in aparticularcityselected can be visualized.

4) Tree Chart (Figure 5 and Figure 7):
Figure 5 – Tree chart was used to get an overview of arrest premises.

Under Marks tab:

Sum (Number of Records) was used for color coding, detailing and labeling purposes.
Ar Premises (Arrest Premise) was also used for labeling purposes. This graph shows that most arrests (46,386) occurred on highway, street, and alley ETC.

Figure 7 – Tree chart was used to get an overview of whether the suspect was armed or unarmed at the time of arrest.

Under Marks tab:
Sum (Number of Records) was used for color coding, detailing and labeling purposes. Ar Weapon (Arrest Weapon) was also used for labeling purposes. This graph shows that most suspects either had no weapons (48,696) or were unarmed (25, 431) at the time they were arrested.

5) Stacked bar graph (Figure 6) – Crime breakdown by race and gender was plotted using stacked bar graphs.

Under Marks tab:

Variable Sex1 i.e. Gender was used for color coding and Sum (Number of Records) was used for labeling purposes. A filter was used for variable Race1 and a single value (dropdown) option was used. With the help of a single value(dropdown) option, the crime committed by each race and gender can be observed.

6) Dual axis area chart (Figure 8) – This was plotted to study the total of drug related cases. Sum (Number of Records) – under the dropdown arrow, contains quick table calculation and a running total was selected to compute this graph.

Under Marks tab:

For 1st Sum (Number of Records) – Variable drug related was used for color coding. Variable drug related was also used as a filter. This graph gives a running total of Sum (Number of Records). This graph shows that most crime was not related to drugs during the period from 2014 to Spring 2020.

7) Figure 9 – This graph shows the arrest action taken every year over the period of 2014 to Spring 2020.

Under Marks tab:

Variable Sum (Number of Records) – was used for color coding and labeling purposes. Variable Arrest Yr (Arrest Year) – was used for detailing purposes.
Variable Ar Action (Arrest Action) was used for labeling purposes.

A filter of variable Arrest Yr (Arrest Year) was applied to filter the range of values i.e. the time period from 2014 to Spring 2020.

8) Dashboard -

The three dashboards (dashboard 1, dashboard 2, dashboard 3) summarizes the information already given. It serves as a comprehensive representation of the police arrests that occurred during the period of 2014 to Spring 2020 mainly in Texas. This is useful in the sense that the decision makers or anyone using the Tableau file can see everything related to the kind of crime committed and public safety and they don't need to rely on their memory to see the big picture.

9) Story – The story represented in tableau gives the decision makers an overview of the police arrests in Texas. It highlights the crime committed by three categories:

   i. By time – gives details about the overall arrest trend in the period from 2014 to Spring 2020, the susceptible time of the day and day of the week where most crime is committed.
   ii. By category – highlights the categories of crime like the details about possession of an arrest weapon, whether the crime was drug related or not, what kind of arrest action was taken each year and which race and gender committed most crime.

iii. By location – highlights the premises where most arrests took place and also tells us the details of home states of suspects and the states where they were arrested.

❖ Implications

Police officers in Texas make a considerable number of arrests per year. Although arrests are made to enforce laws and protect public safety, it can have wide ranging consequences – including the risk of injury to both officers and suspects. Despite the effects on individuals and the broader community, little is known about the factors underlying arrest trends. Concerns about racial disparities and ongoing debates about policing and community relations - further highlight the need for a better understanding of law enforcement and use of public resources in making arrests. In this report, we examine factors in statewide arrest trends as well as differences in arrest rates and racial disparities across counties in Texas. We analyze the role of crime rates, action taken by officials and county-level factors such as demographics.

With a database of over 84000 entries representing the police arrests majorly in Dallas, Texas in the time frame from 2014 to Spring 2020. There was a need to translate this raw data to gain insights about the crime in the area and implement public safety measures accordingly.

We find that there are three factors influencing this dataset.

The crime committed at different times, different categories and different locations.
Time: It has been observed that the trend of year wise arrest has been decreasing. The number of arrests in 2019 was 11,533 was high as compared to that in Spring 2020 (5,047). More research needs to be conducted whether the decline in Spring 2020 is due to stay at home orders and social distancing measures or due to increased safety vigilance. The average age of the suspects was in the range of 34 – 36 years. It has also been observed that most of the crime is committed post-midnight. Overall, Saturday (14, 891) is the day where most crime is committed.

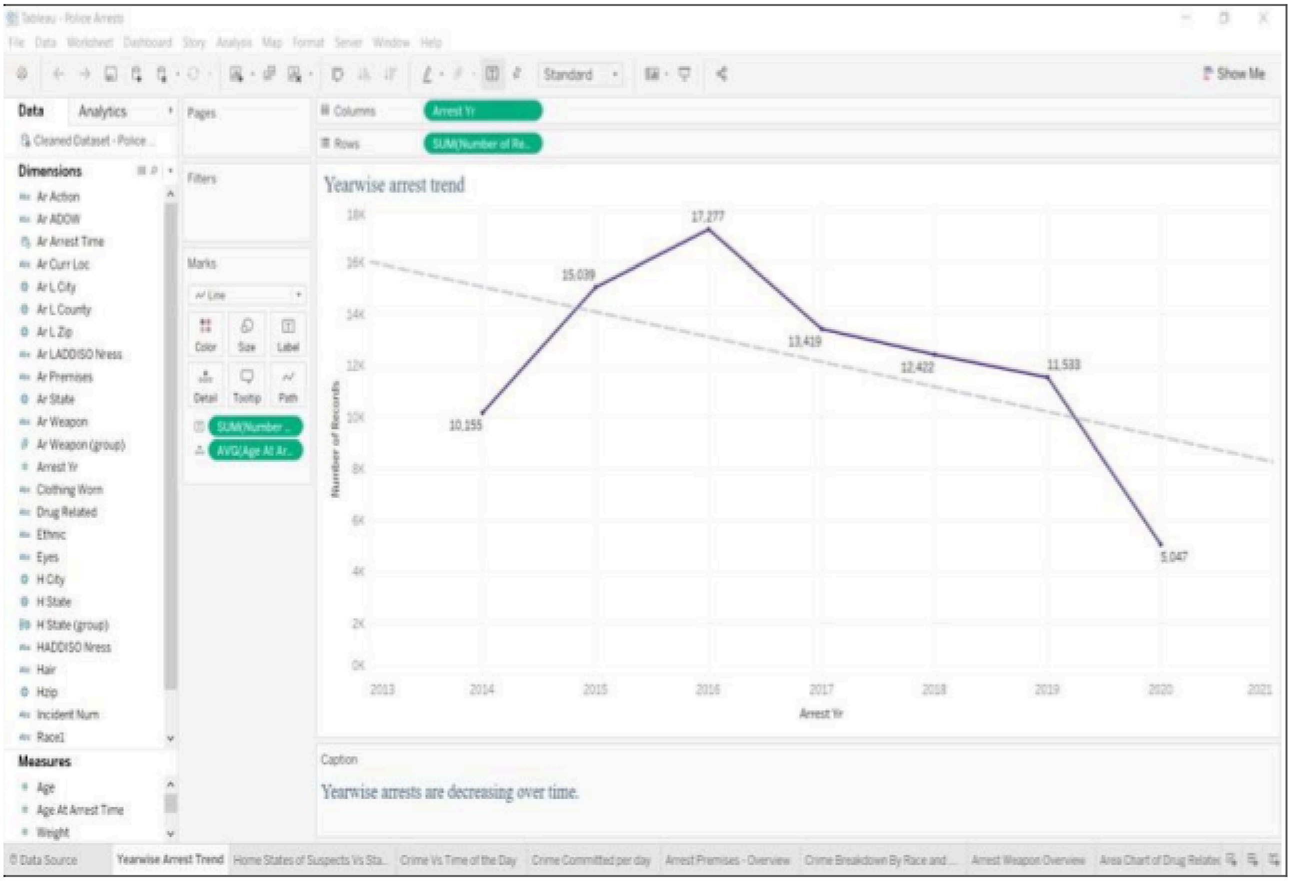
Category: Weapon / Unarmed – Most suspects did not possess any weapon. Drug Related – Most cases were not drug related.
Location: Most suspects were arrested in Texas followed by Tennessee, Utah and Denver. Arrest Premises – Most of the arrests took place on highway, street, alley etc..
Moreover, crime rates vary substantially over counties. More crime was committed in Dallas county as compared to other counties.

- ## Conclusion:

The visualizations created of the police arrests in Texas dataset reveal interesting trends in the data. Based on the number of arrests made since 2014 it can be concluded that crime has gradually reduced in the DFW metropolitan area. The number of arrests dramatically decreased in the spring of 2020. The most likely reason for this dramatic reduction could be associated with the social distancing rules and stay at home orders being enforced amongst the ongoing Covid -19 pandemic or due to increased police vigilance. Interestingly, based on the number of arrests made, the crime peaks during the midnight hours of weekends and is at its lowest during early morning hours on Tuesdays. About half of the arrests made were unarmed suspects and most arrests were non-drug related. Highways and streets were locations identified to be crime hotspots and most arrests were made on these locations. The average age group of the suspects was between 34 to 36 years of age. Finally, Dallas county reported the majority of crimes as compared to other counties in the DFW metropolitan area.

❖ Appendix

Figure 1: Line Graph – Year wise arrest trend and average age at the time of arrest

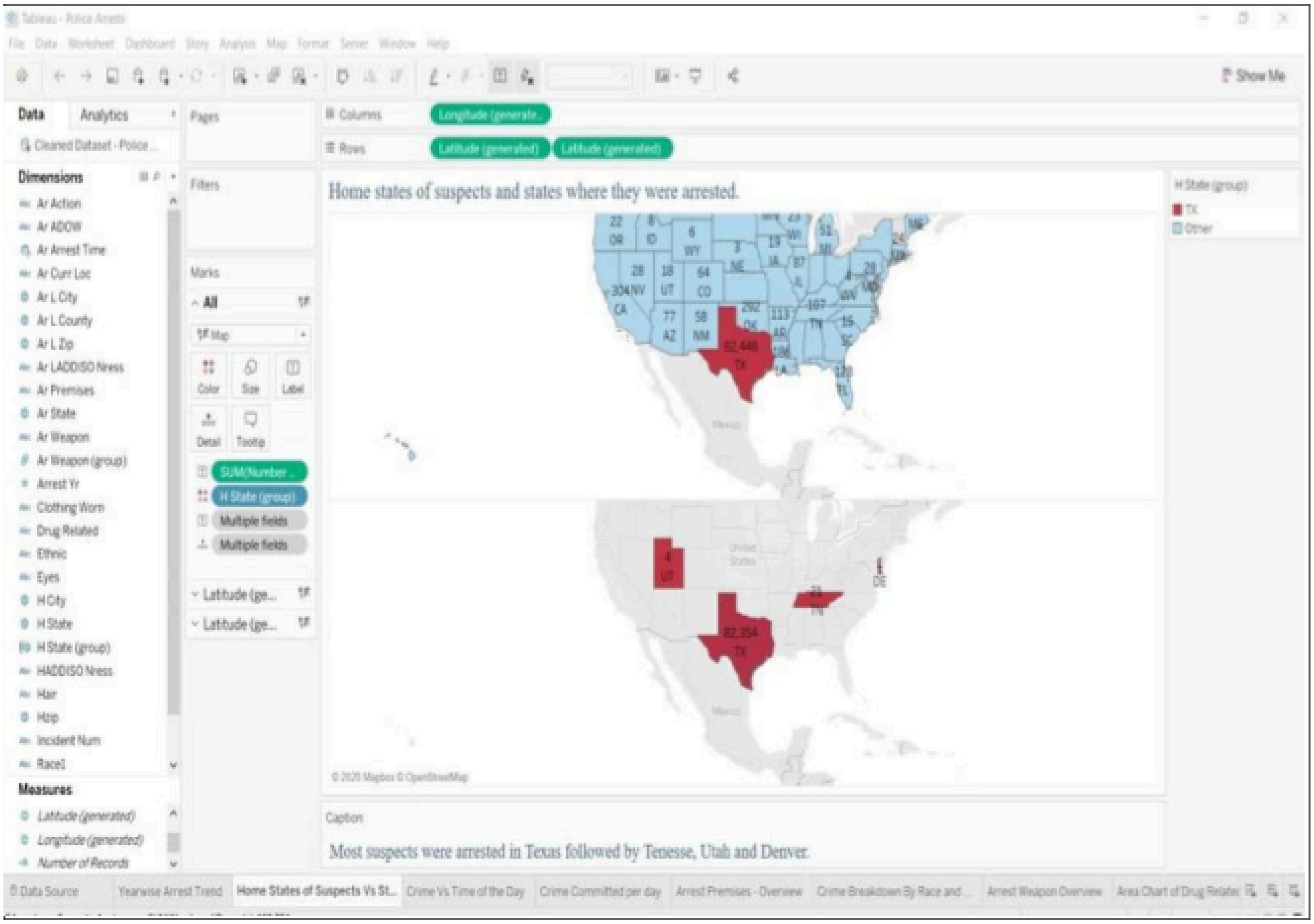


Yearwise arrest trend

Caption

Yearwise arrests are decreasing over time.

Figure 2: Dual Axis Map – Home States of suspects and the states where they were arrested.



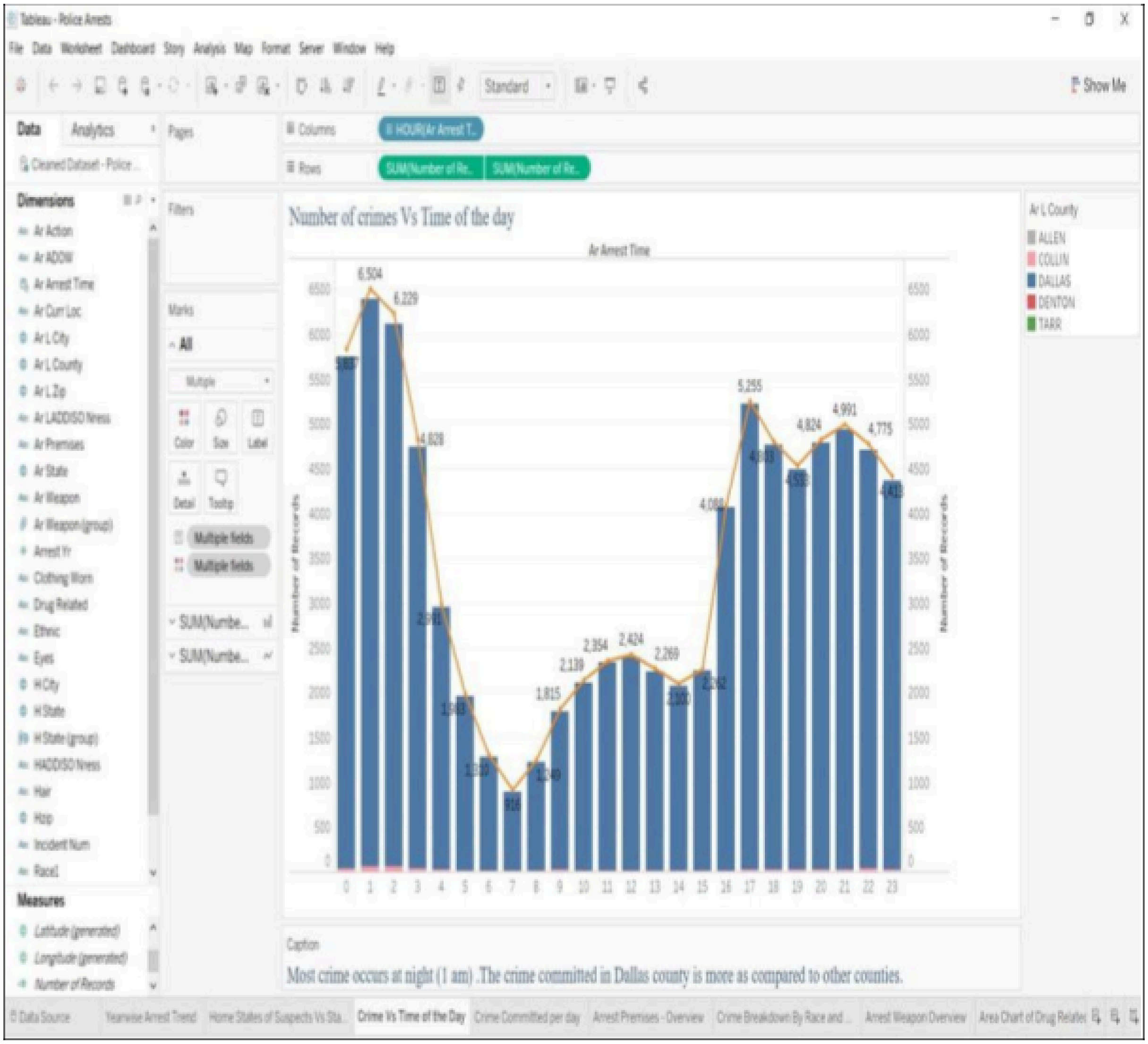
Home states of suspects and states where they were arrested.

Caption

Most suspects were arrested in Texas followed by Tenesse, Utah and Denver.

- Combination charts:

Figure 3: A bar and line chart with dual axis, synchronize axis – Number of crimes Vs Time of the day

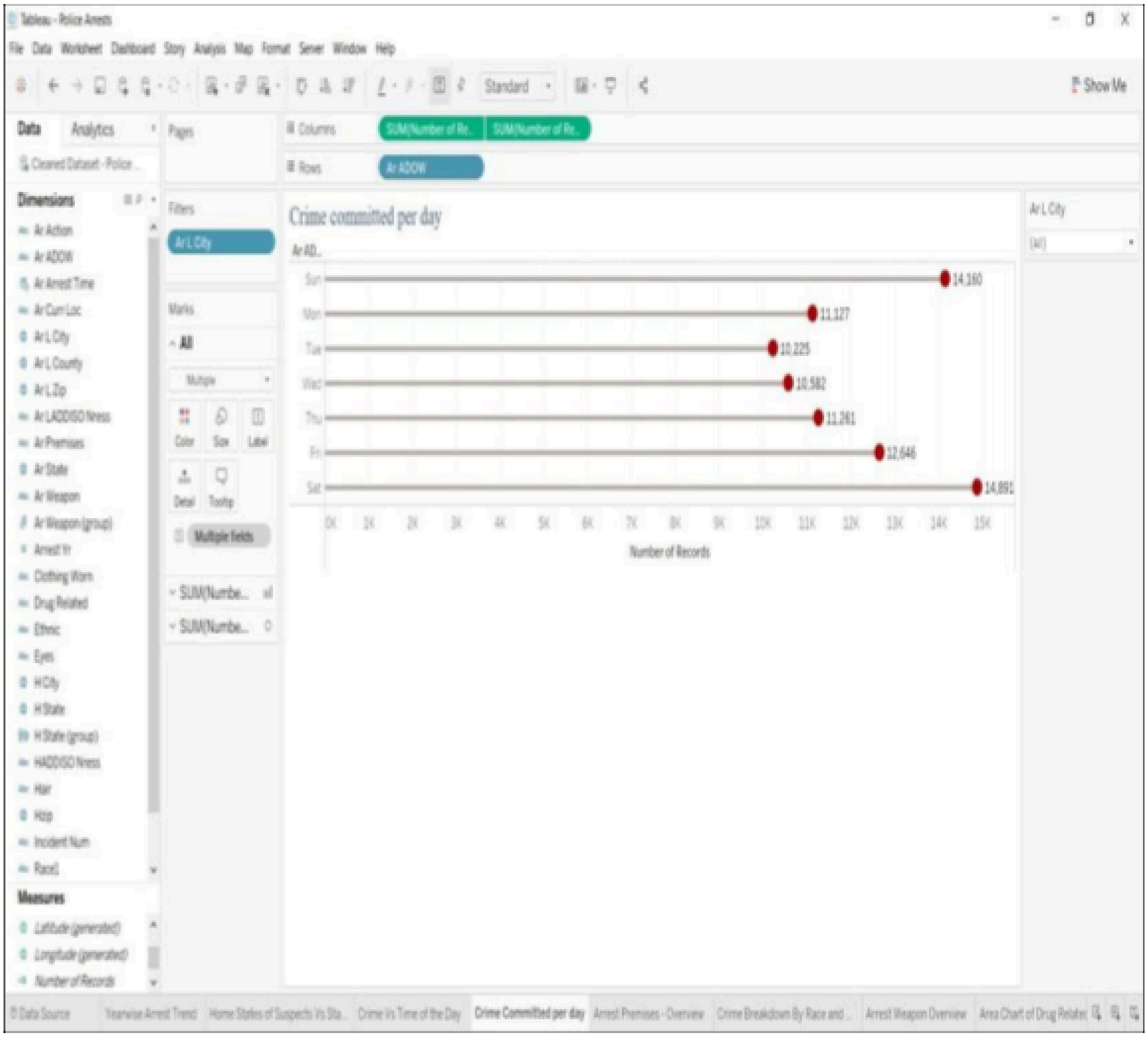


Number of crimes Vs Time of the day

Caption

Most crime occurs at night (1 am) .The crime committed in Dallas county is more as compared to other counties.

Figure 4: Lollipop Chart – Crime committed per day



Columns: SUM(Number of Re.. SUM(Number of Re..

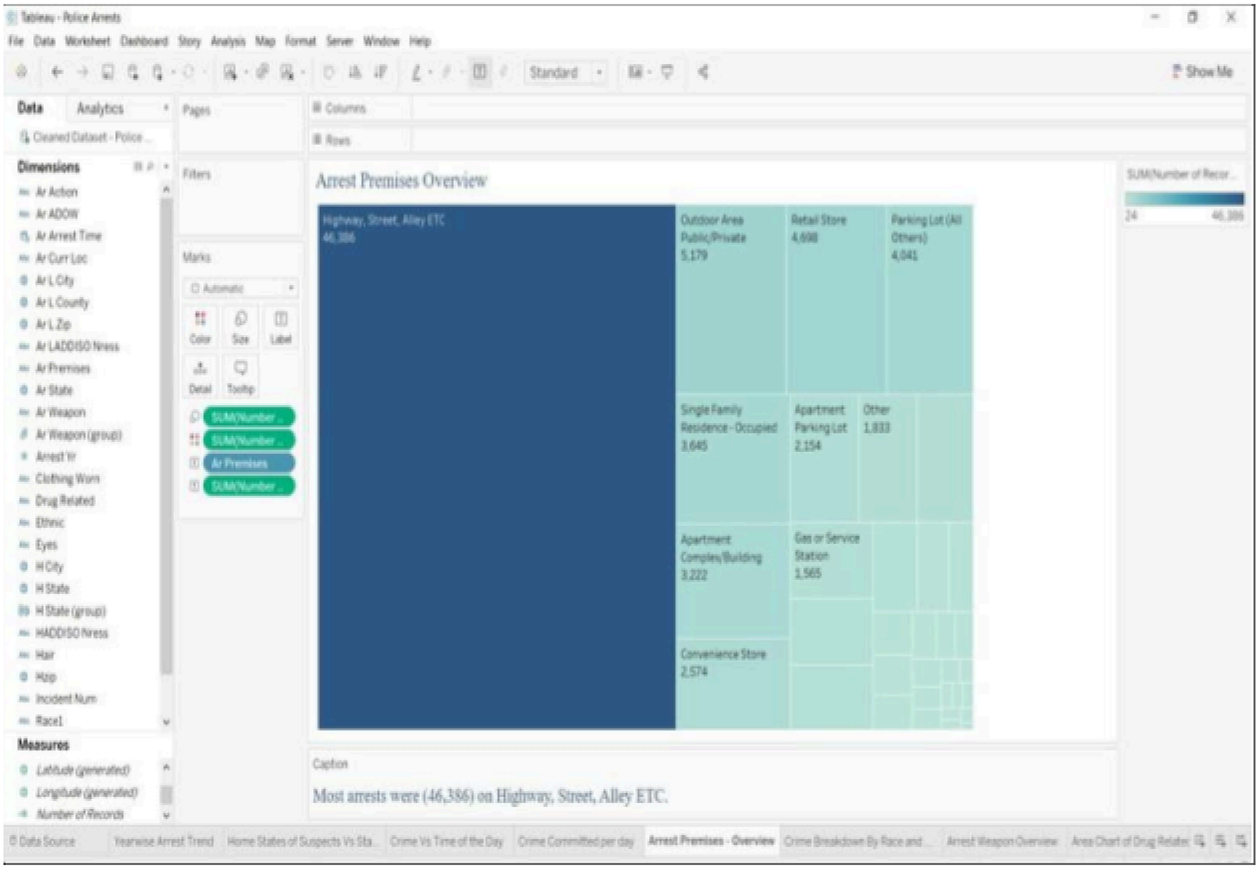
Rows: Ar ADOW

Filters: Ar L City

Ar L City: (All)

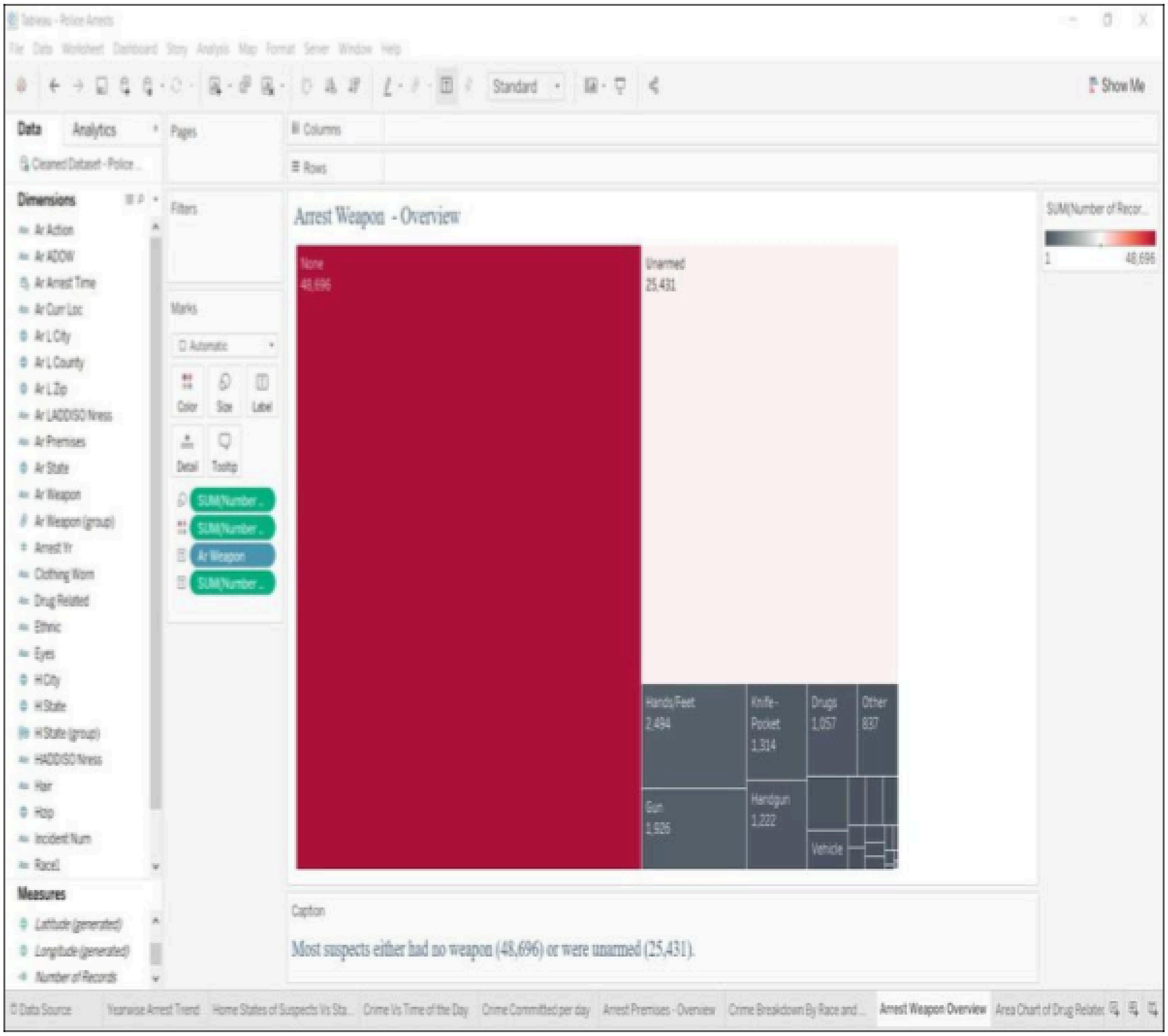
- Tree Chart

Figure 5: Arrest Premises Overview



Most arrests were (46,386) on Highway, Street, Alley ETC.

Figure 7: Arrest Weapon - Overview



Most suspects either had no weapon (48,696) or were unarmed (25,431).

Figure 6: Stacked Bar Graph – Crime Breakdown by Race and Gender

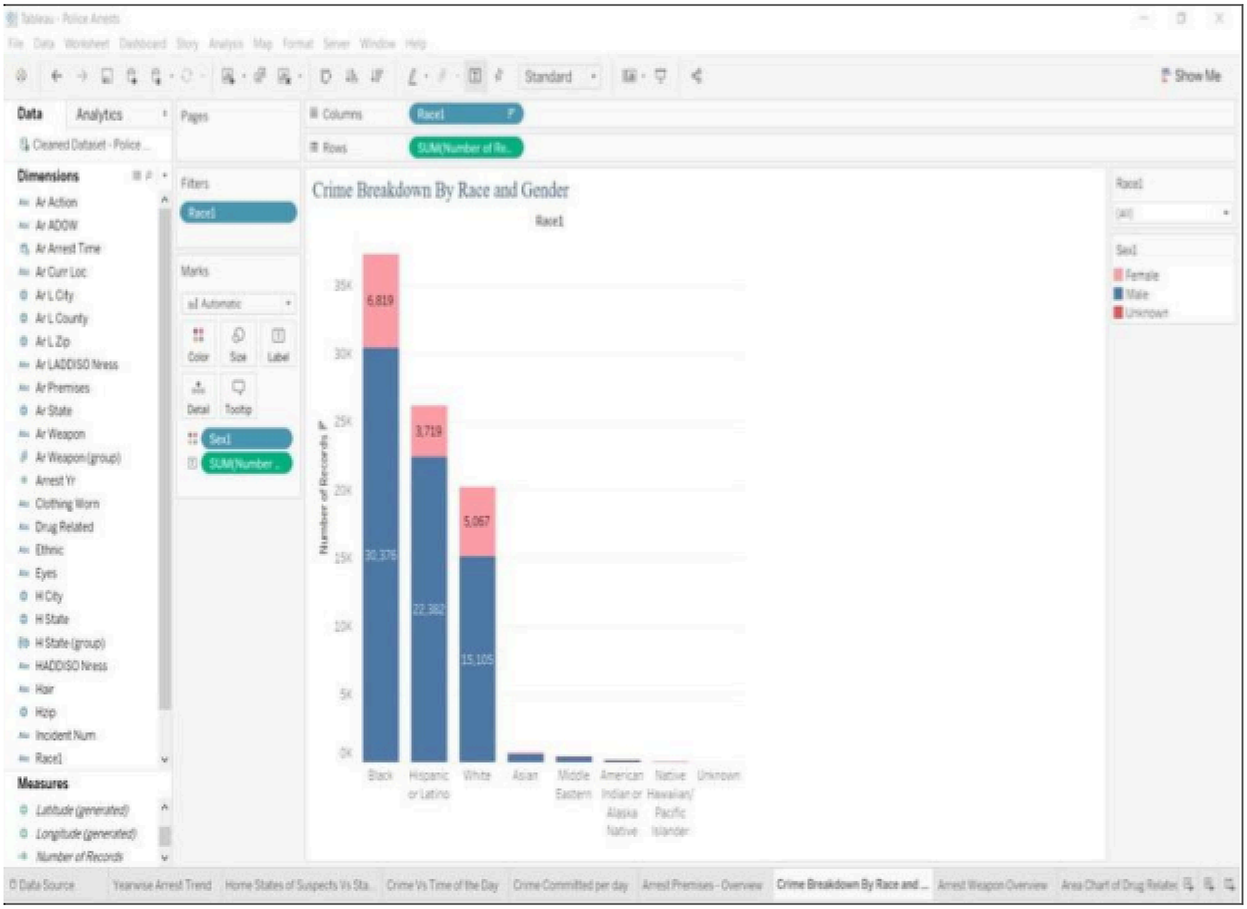


Figure 8: Dual axis area chart with a total of drug related cases

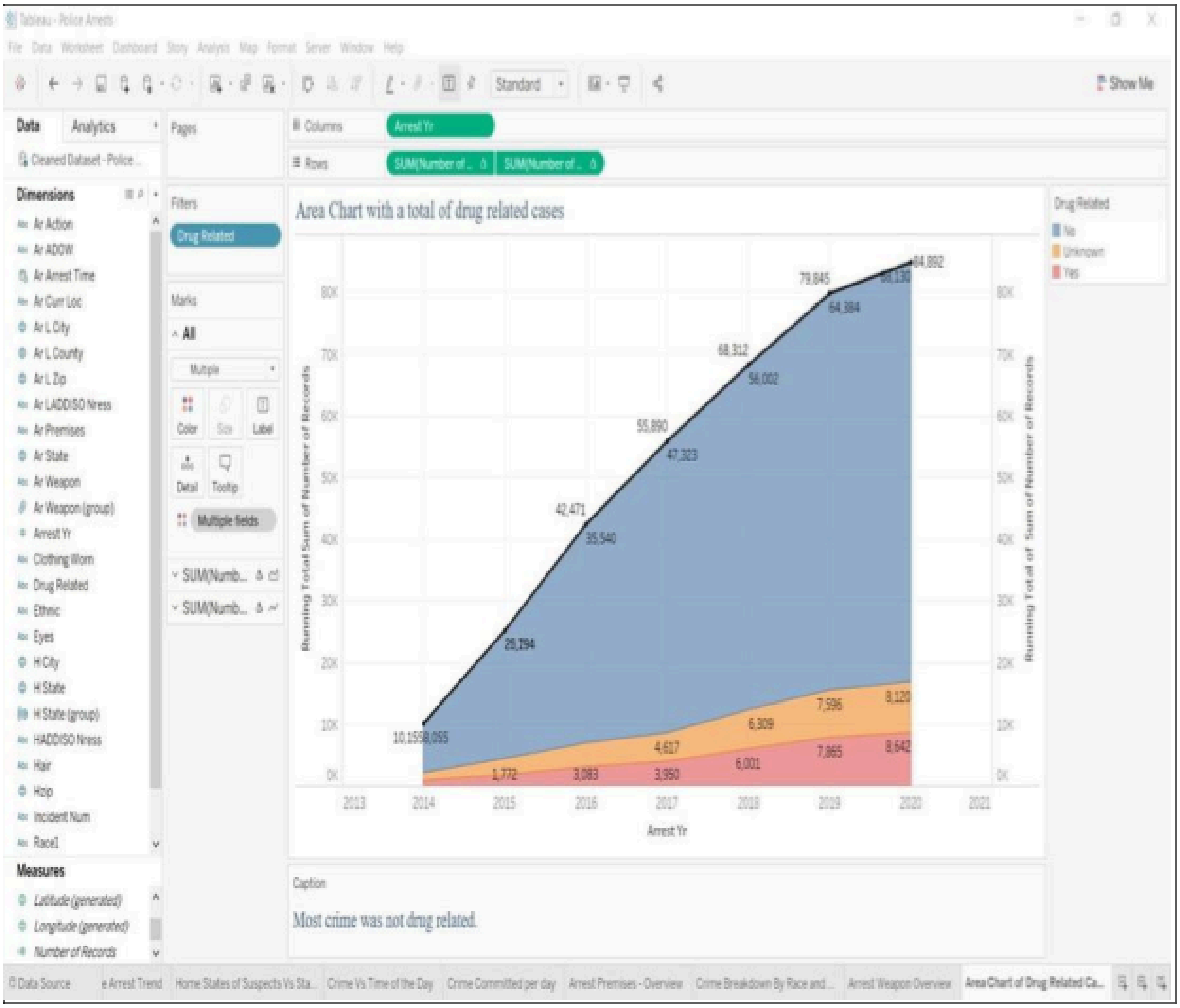


Tableau - Police Arrests

Columns: Arrest Yr

Rows: SUM(Number of ... Δ SUM(Number of ... Δ

Filters: Drug Related

Marks: All — Multiple — Multiple fields

SUM(Numb... Δ

SUM(Numb... Δ

Area Chart with a total of drug related cases

Caption

Most crime was not drug related.

Figure 9: Arrest Action Taken Every Year

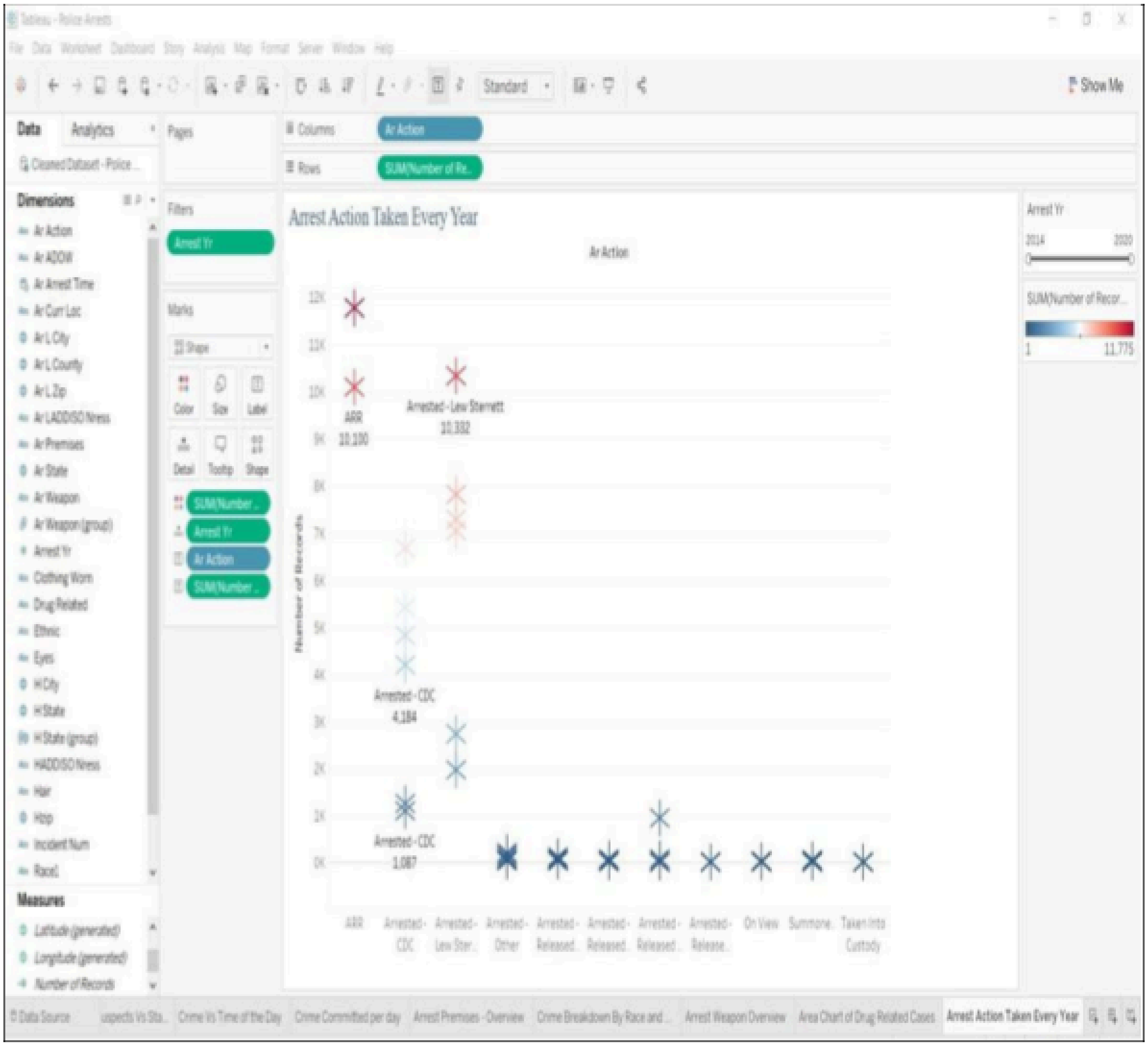
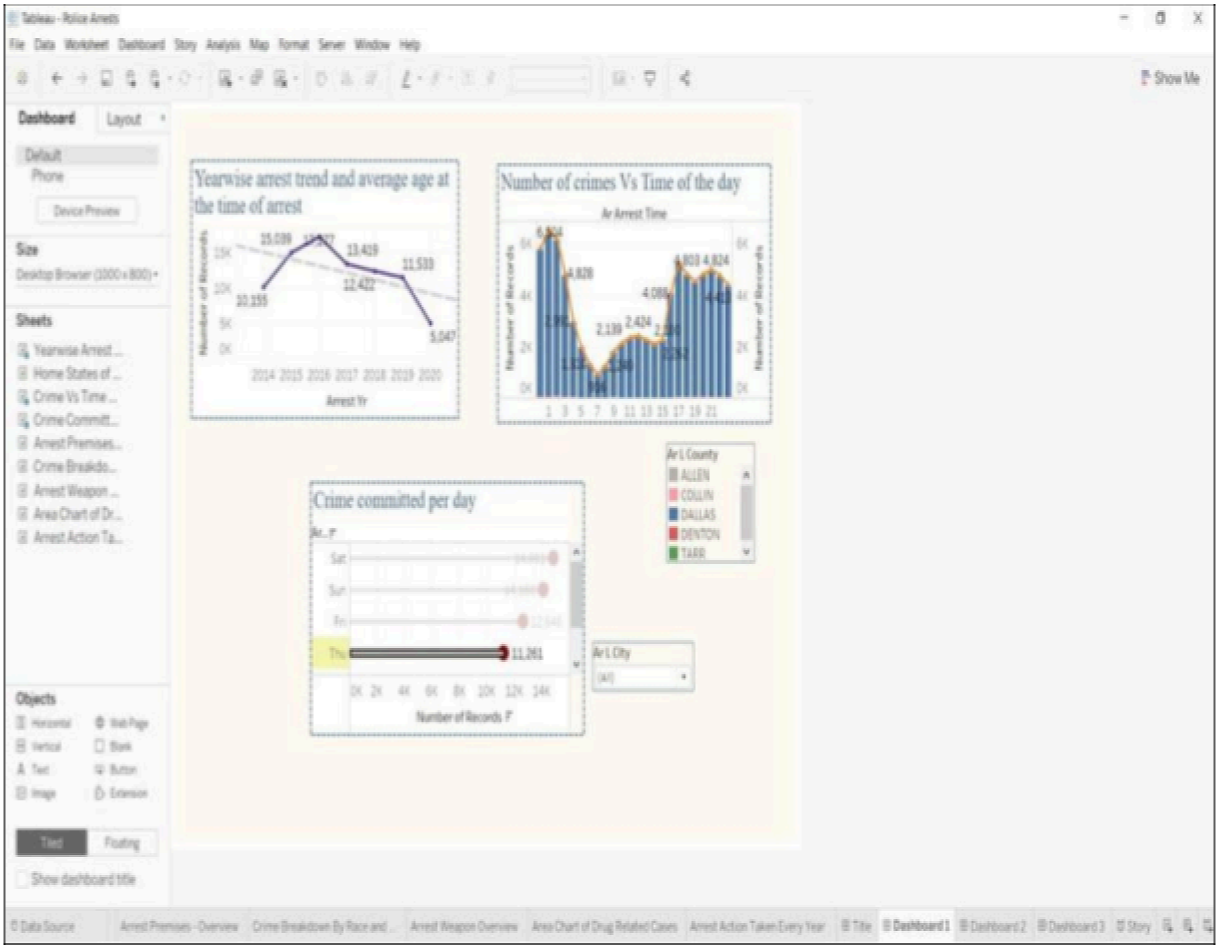


Figure 10 – Dashboard 1, Dashboard 2, Dashboard 3

Dashboard 1



Dashboard 2

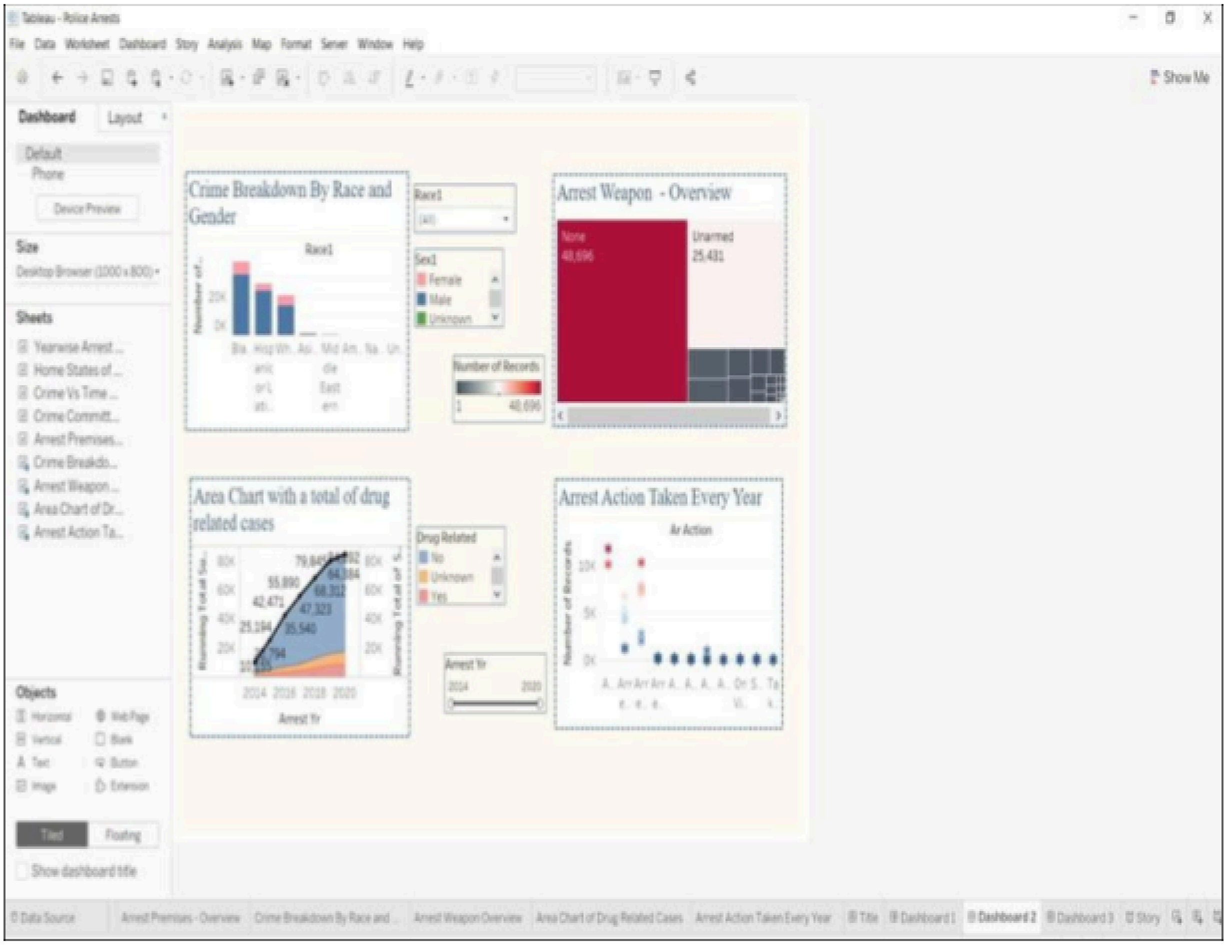


Tableau - Police Arrests
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard
Layout
Default
Phone
Device Preview
Size
Desktop Browser (1000 x 800)
Sheets
Yearwise Arrest...
Home States of...
Crime Vs Time...
Crime Committ...
Arrest Premises...
Crime Breakdo...
Arrest Weapon...
Area Chart of Dr...
Arrest Action Ta...
Objects
Horizontal
Web Page
Vertical
Blank
Text
Button
Image
Extension
Tiled
Floating
Show dashboard title
Crime Breakdown By Race and Gender
Race1
Number of...
20K
0K
Bla.. Hisp anic or L ati.. Wh.. Asi.. Mid dle East ern Am.. Na.. Un..
Race1
(All)
Sex1
Female
Male
Unknown
Number of Records
1
40,696
Arrest Weapon - Overview
None
40,696
Unarmed
25,431
Area Chart with a total of drug related cases
Running Total of Su...
80K
60K
40K
20K
79,845
81,882
64,384
55,890
68,312
42,471
47,323
25,194
35,540
794
2014 2016 2018 2020
Arrest Yr
Drug Related
No
Unknown
Yes
Arrest Yr
2014
2020
Arrest Action Taken Every Year
Ar Action
Number of Records
10K
5K
0K
A.. Arr e.. Arr e.. Arr e.. A.. A.. A.. A.. Ot Vi.. S.. Ta k..
Data Source
Arrest Premises - Overview
Crime Breakdown By Race and...
Arrest Weapon Overview
Area Chart of Drug Related Cases
Arrest Action Taken Every Year
Title
Dashboard 1
Dashboard 2
Dashboard 3
Story

Dashboard 3

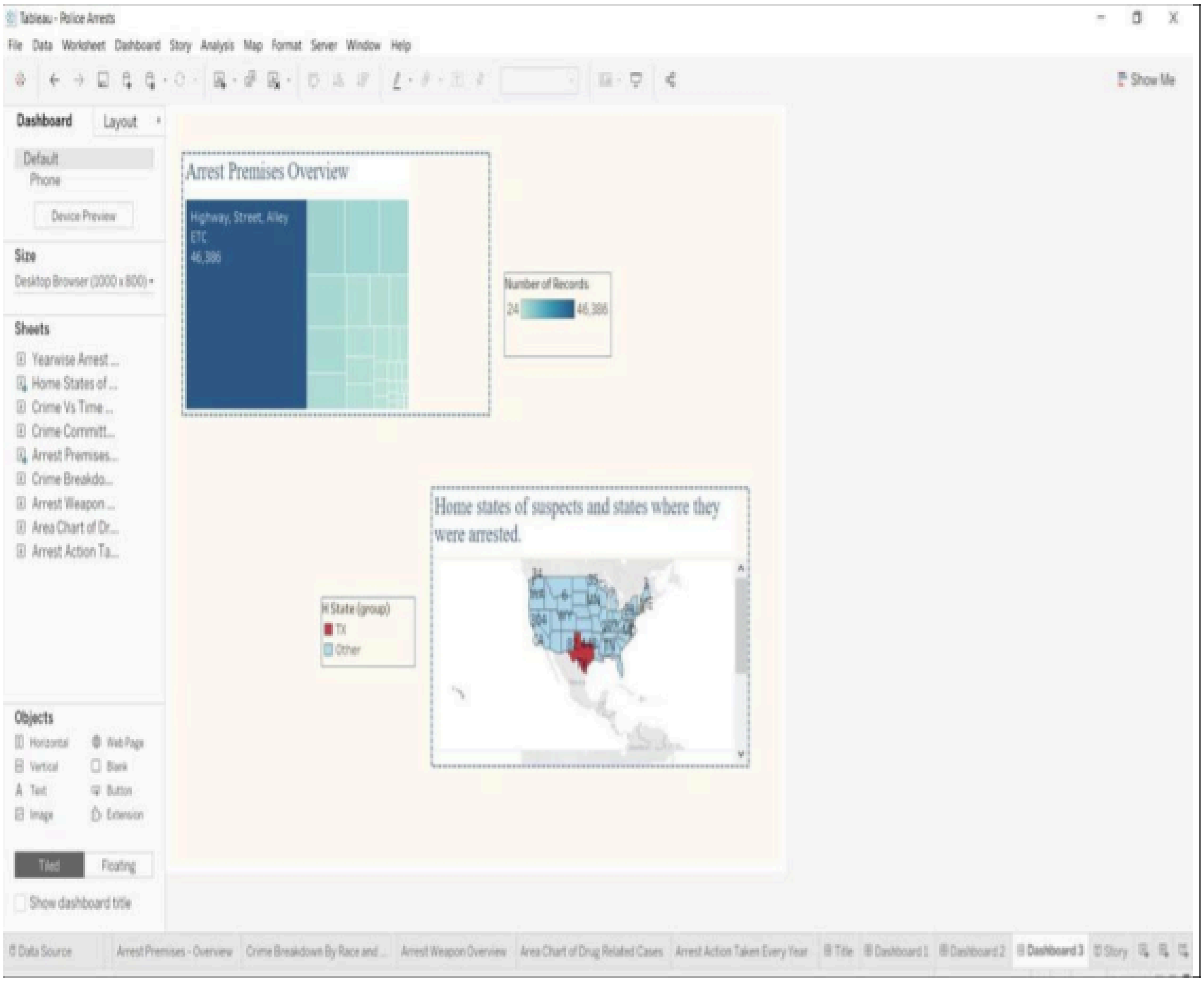


Tableau - Police Arrests
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard
Layout
Default
Phone
Device Preview
Size
Desktop Browser (1000 x 800)
Sheets
Yearwise Arrest ...
Home States of ...
Crime Vs Time ...
Crime Committ...
Arrest Premises...
Crime Breakdo...
Arrest Weapon ...
Area Chart of Dr...
Arrest Action Ta...
Objects
Horizontal
Web Page
Vertical
Blank
Text
Button
Image
Extension
Tiled
Floating
Show dashboard title
Arrest Premises Overview
Highway, Street, Alley
ETC
46,386
Number of Records
24
46,386
Home states of suspects and states where they were arrested.
H State (group)
TX
Other
Data Source
Arrest Premises - Overview
Crime Breakdown By Race and ...
Arrest Weapon Overview
Area Chart of Drug Related Cases
Arrest Action Taken Every Year
Title
Dashboard 1
Dashboard 2
Dashboard 3
Story

Figure 11 – Story

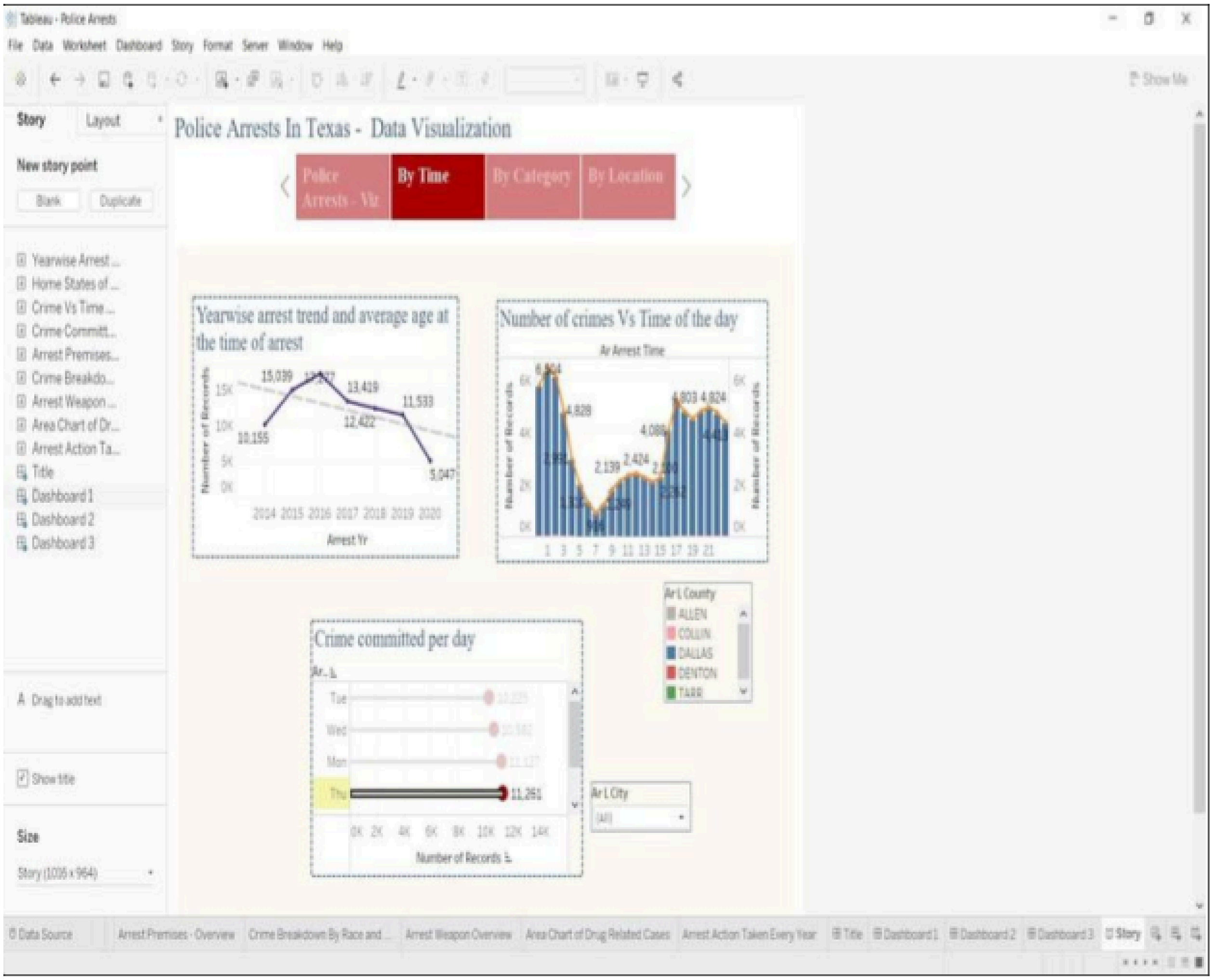


Tableau - Police Arrests
File Data Worksheet Dashboard Story Format Server Window Help
Show Me
Story Layout
New story point
Blank Duplicate
Yearwise Arrest ...
Home States of ...
Crime Vs Time ...
Crime Committ...
Arrest Premises...
Crime Breakdo...
Arrest Weapon ...
Area Chart of Dr...
Arrest Action Ta...
Title
Dashboard 1
Dashboard 2
Dashboard 3
Drag to add text
Show title
Size
Story (1016 x 964)
Police Arrests In Texas - Data Visualization
Police Arrests - Viz
By Time
By Category
By Location
Yearwise arrest trend and average age at the time of arrest
Number of Records
15K
10K
5K
0K
15,039
13,419
12,422
11,533
10,155
5,047
2014 2015 2016 2017 2018 2019 2020
Arrest Yr
Number of crimes Vs Time of the day
Ar Arrest Time
Number of Records
6K
4K
2K
0K
4,828
4,089
2,139
2,424
4,803 4,824
1 3 5 7 9 11 13 15 17 19 21
Ar L County
ALLEN
COLLIN
DALLAS
DENTON
TARR
Crime committed per day
Tue
Wed
Mon
Thu
11,261
0K 2K 4K 6K 8K 10K 12K 14K
Number of Records
Ar L City
(All)
Data Source
Arrest Premises - Overview
Crime Breakdown By Race and ...
Arrest Weapon Overview
Area Chart of Drug Related Cases
Arrest Action Taken Every Year
Title
Dashboard 1
Dashboard 2
Dashboard 3
Story