SAFE TRAVELS IN SOUTH AFRICA

CSC 465 Data Visualization – Final Project Report

SARAH CUMMINGS | MITALI INGOLE | JOY MBOYA

TABLE OF CONTENTS

Abstract 2

Initial Exploratory Analysis 2

Crime type analysis and creation of severity variable 5

Refined Techniques 7

Approach 1: Heat map of type by category 7

Approach 2: Choropleth 8

Approach 3: Glyph Plot 9

Approach 4: Contingency/mosaic plot 10

Results and Discussion 10

Appendix 11

R Code: 11

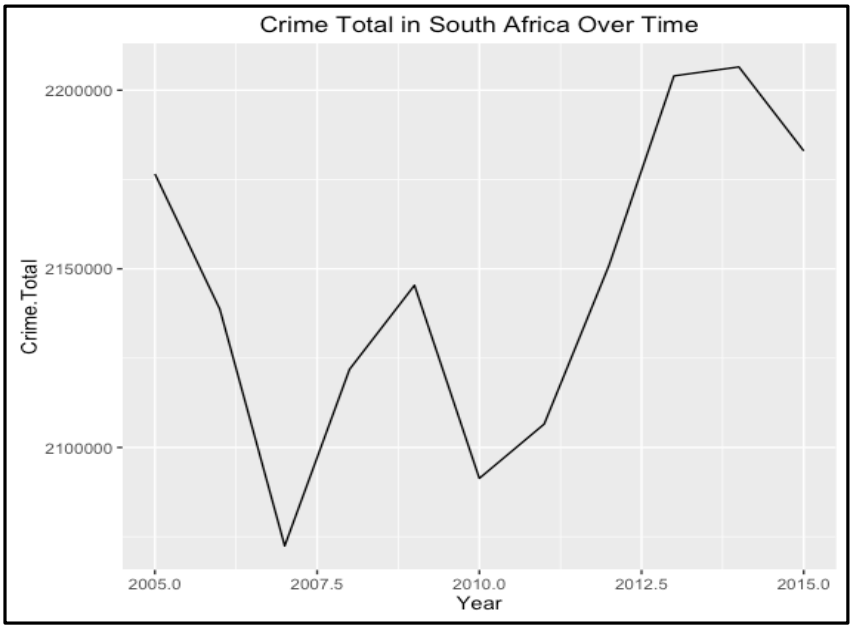
ABSTRACT

In this report, we analyse the crime in South Africa over the span over ten years (mid_2004-mid_2016). We are interested to see if we can find underlying patterns and/or interesting facts about crime in the southern African country that would provide insight into the progression or regression of crime.
The dataset we use for this analysis was sourced from kaggle.com[1]. There are 3 categorical variables and eleven numerical variables. The categorical variables are broken down into province, station (referring to the station where the crime was reported), and category (refers to the kind of crime reported). The numerical variables are broken down by year range, 2004-2005 until 2015-2016.

INITIAL EXPLORATORY ANALYSIS

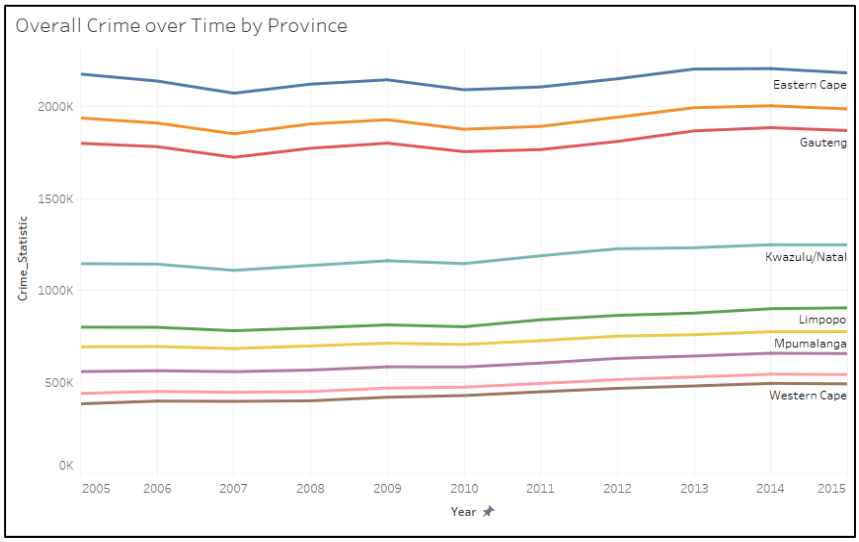
In this exploratory analysis, we experiment with various plots to gain better understanding of the dynamic of crime in South Africa. Below we evaluate what each exploratory plot revealed.

1. Line graphs of total crime over time, and total crime over time by province.



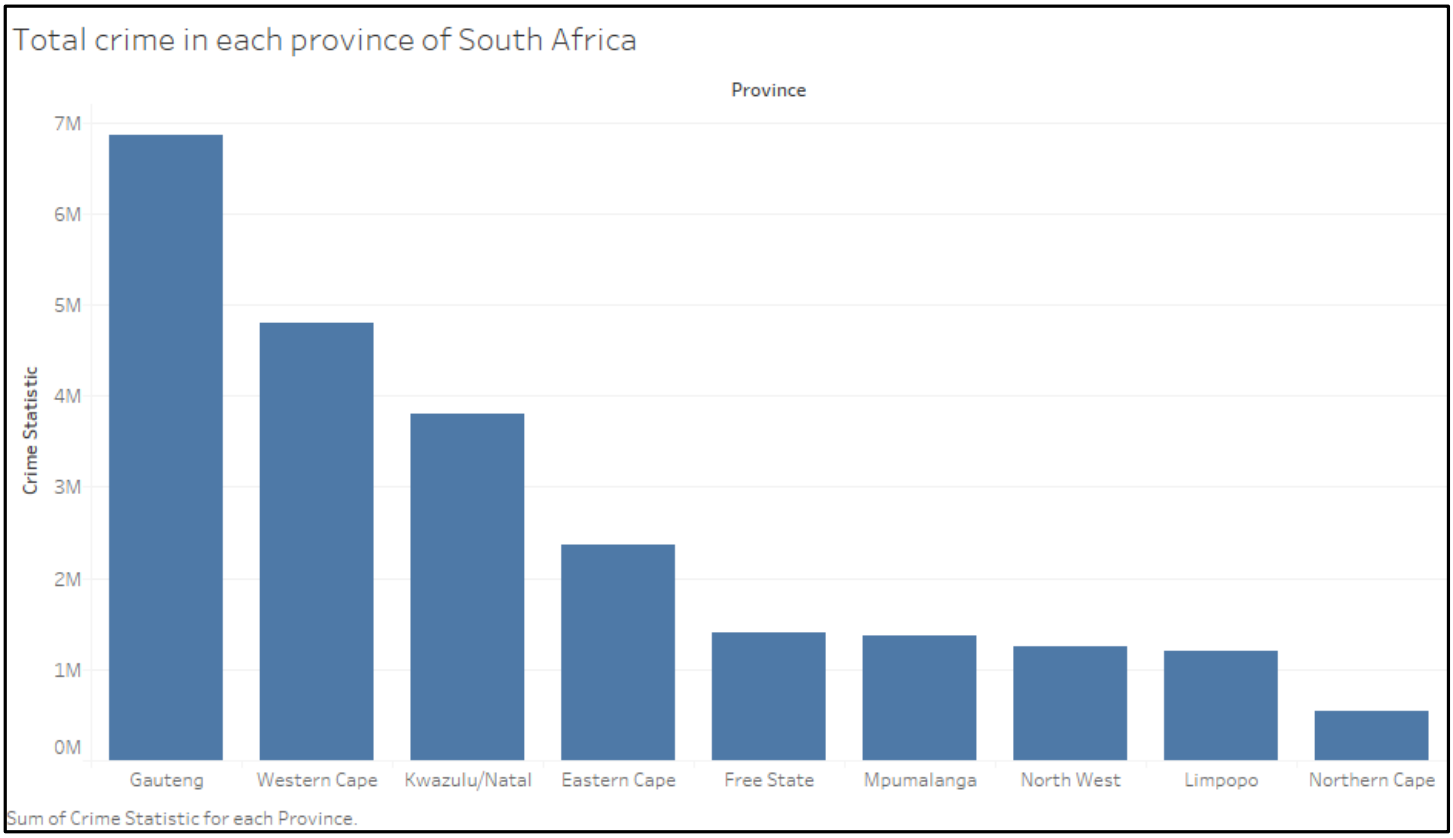
- There seems to an overall drop in crime in 2007 and 2010 based on initial line plot of crime over time
- Plotting each province by itself seems like a natural next step to help determine the provinces that influenced the gradual (though very slow) rise in crime.

[1] South African Crime Data Source - https://www.kaggle.com/slwessels/crime-statistics-for-south-africa



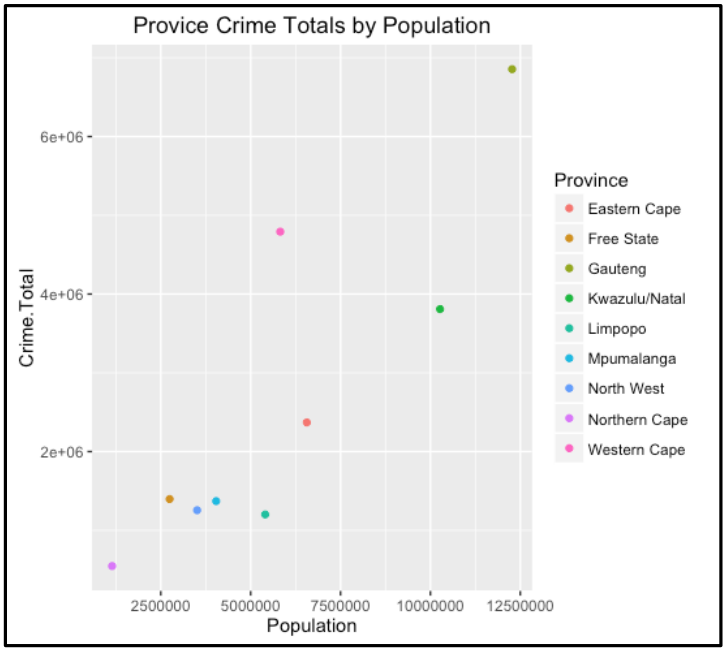
- Eastern Cape Province appears to have the highest rate of crime.
- The highest contributors to crime in South Africa are Eastern Cape, and Gauteng. The other provinces crime caps at a little over 100,000.

2. Bar chart of crime by province



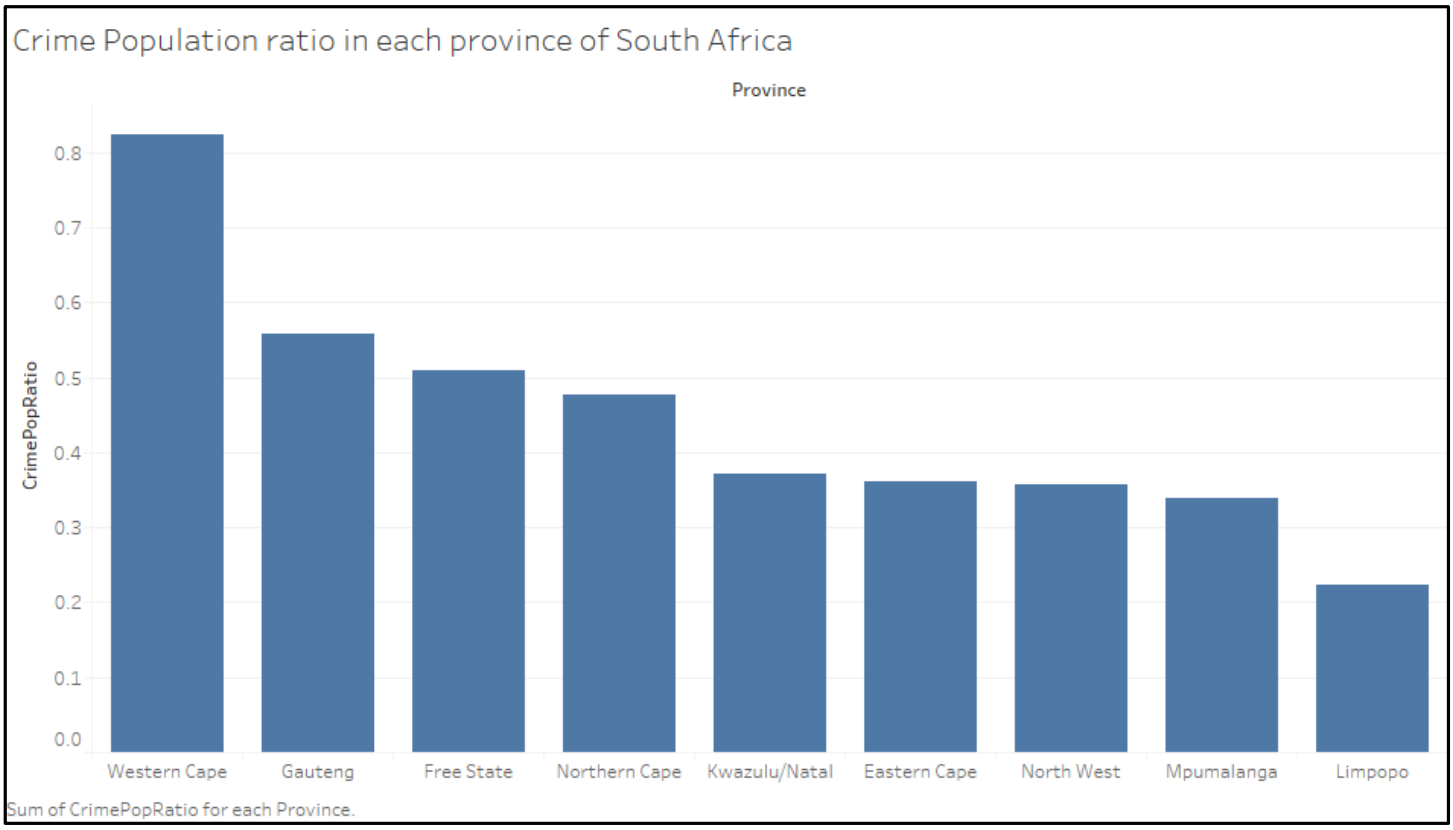
- Gauteng province appears to have the highest rated crime out of all provinces, with West Cape being second. This point is interesting because the line plot of crime over time by province revealed that Western Cape had the highest rate of crime over time with Gauteng second. Northern Cape appears to have the least overall crime rate.

3. Scatterplot of crime totals by population



- There is a positive correlation between crime and population. As population increases, crime rises. Because of this, we created a new variable, CrimePopRatio, that divides crime statistic by population. We focus on Crime Pop Ratio for the rest of our analysis.

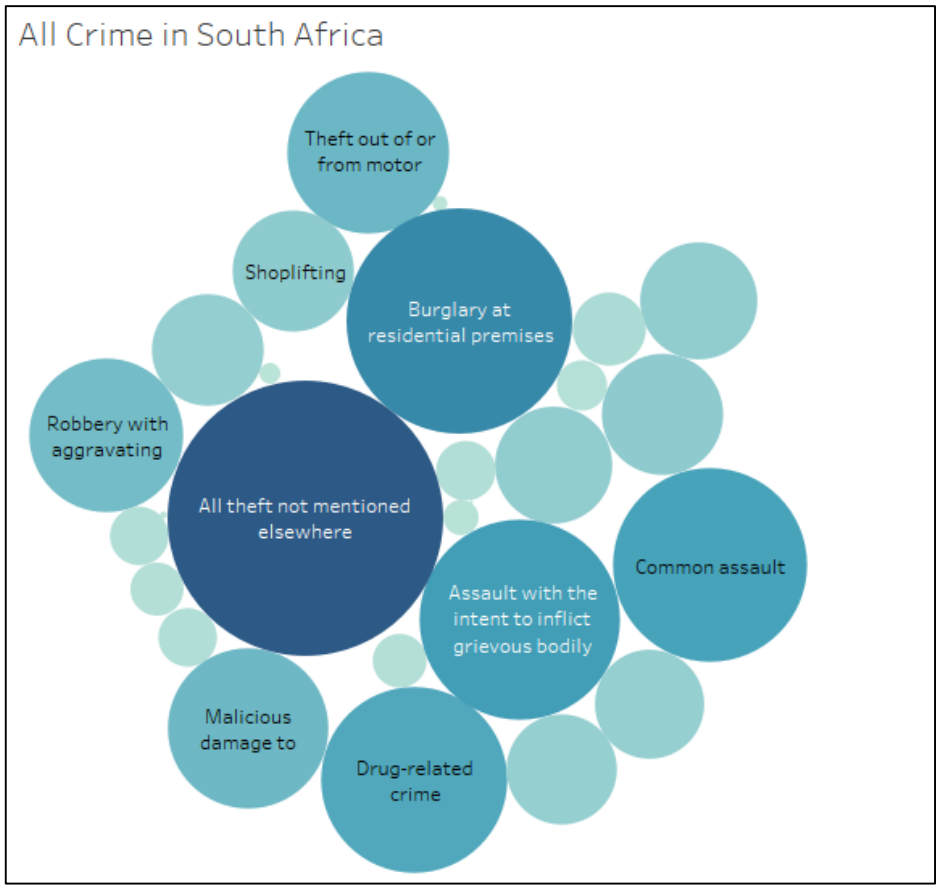
4. Bar chart of crime population ratio by province



- With our new variable, shown in the bar plot above, we can see that Western cape has the most crime per person. Limpopo has the least crime per person.
- From this exploratory analysis, we wonder: what provinces are more dangerous or have more crime per person? Which counties tend to have more crime in categories we consider particularly extreme?

CRIME TYPE ANALYSIS AND CREATION OF SEVERITY VARIABLE

Bubble chart

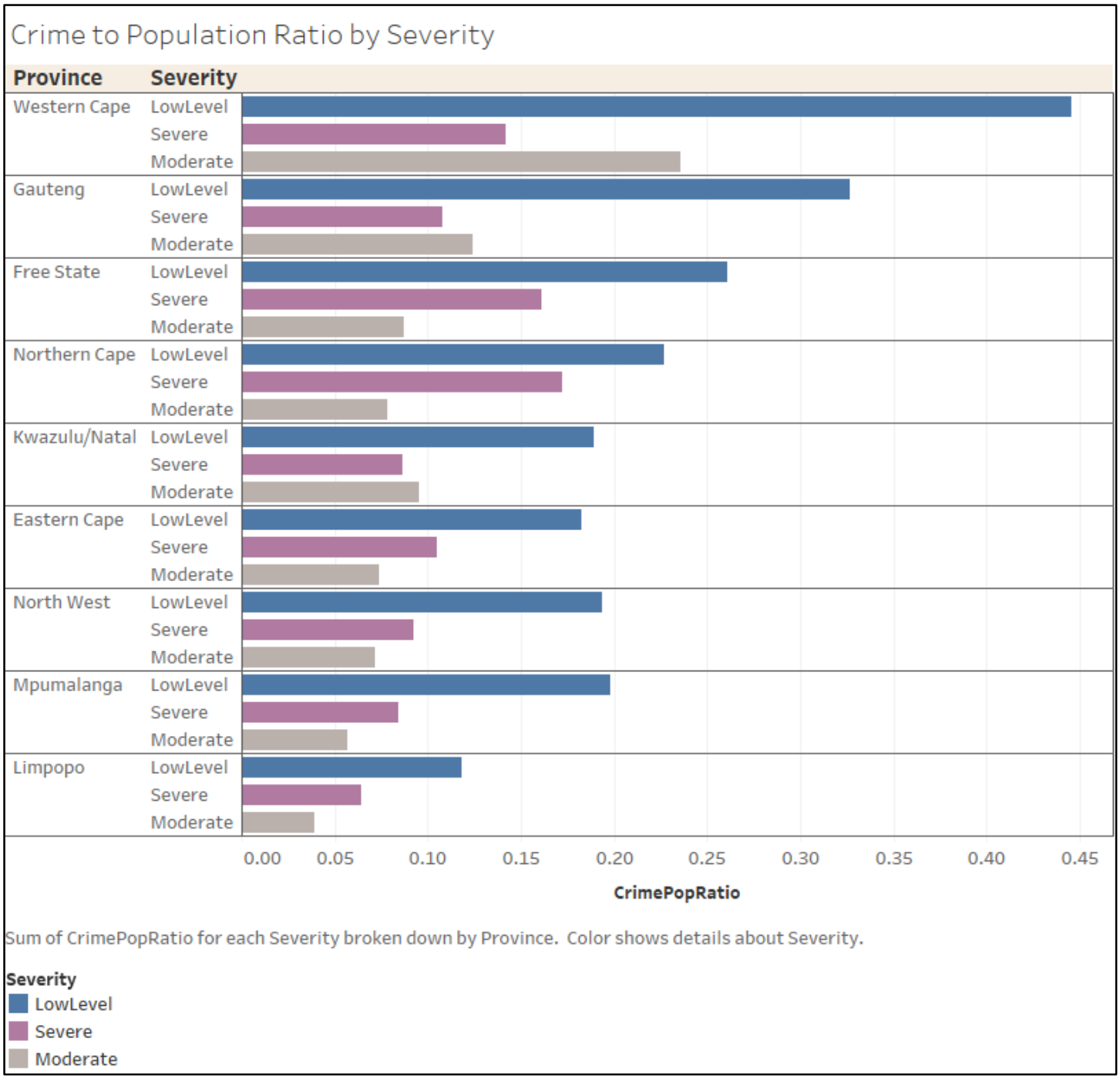


After our initial exploratory analysis, we decided we wanted to further assess the distribution of crime types in the nine provinces. Our dataset included 27 categories of crime, which we visualized in a bubble plot during our initial exploratory phase. One can clearly see that 'All theft not mentioned elsewhere' is the most reported category of crime.

We also see that there are some similar categories of crime that could be grouped together. Because of this, and due the extensive range of crime categories reported, we decided to break these categories down into 3 major severity categories, with 1 being the least heinous crimes and 3 being the most.

- ***Severe Crimes (Category 3***) crimes include murder, attempted murder, sexual related crimes, arson, assault with a deadly weapon and common assault.
- ***Moderate Crimes (Category 2)*** crimes include drug related crimes, bank robbery, driving under the influence, Illegal possession of fire arms and aggravated robbery.
- ***Low Level Crimes (Category 1***) crimes include all theft not mentioned elsewhere, Burglary at non-residential premises, Burglary at residential premises, Carjacking, Commercial crime, Common robbery, Robbery at non-residential premises, Robbery at residential premises, Robbery of cash in transit, Shoplifting, Stock-theft, Theft of motor vehicle and motorcycle, Theft out of or from motor vehicle, Truck hijacking.

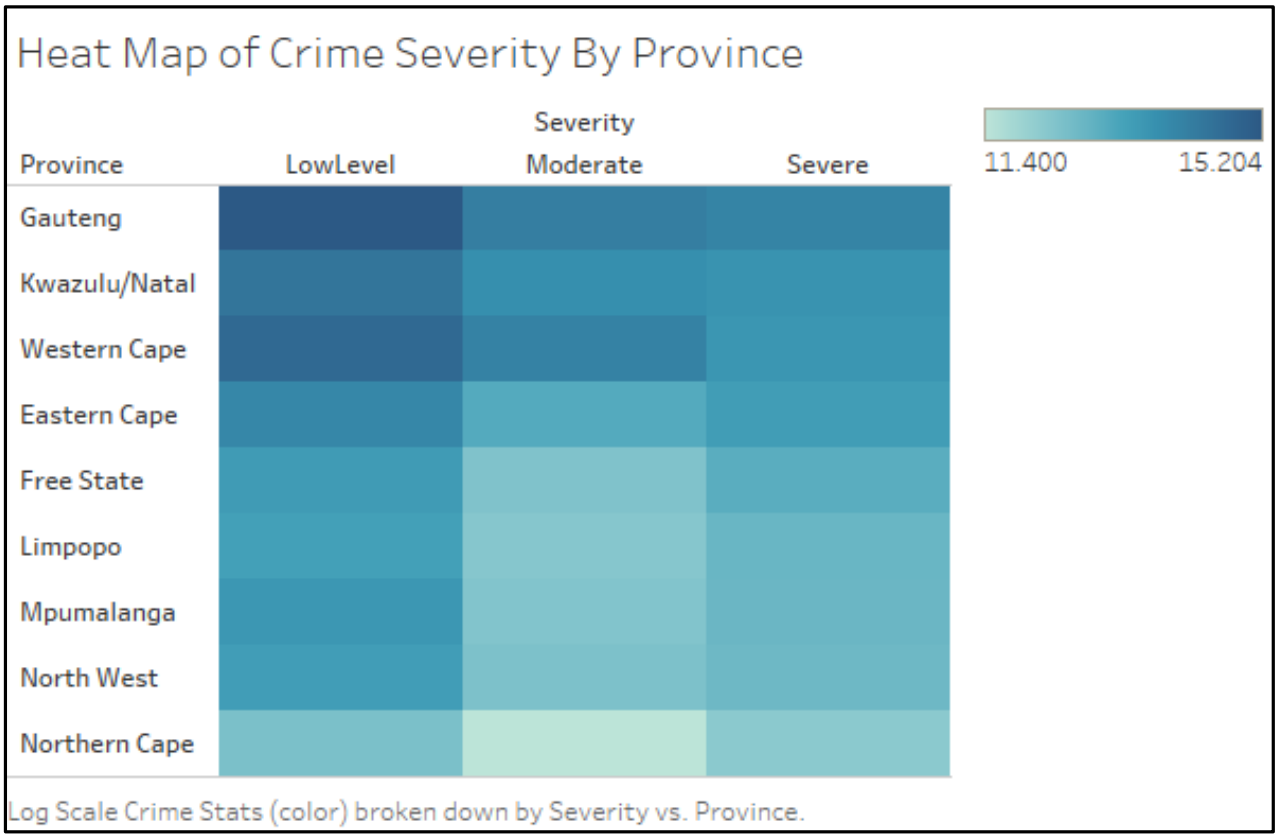
Hierarchical Bar Plot



- The graph above visualizes the distribution of crimes by severity and province.
- Recall western Cape and Gauteng provinces have the highest rate of total crime, as revealed in our exploratory analysis. However, the crimes in those provinces are mostly low level severity.
- It is good to note that even though Gauteng has the highest total crime, Western Cape has the higher rate of heinous crimes in comparison.

REFINED TECHNIQUES

APPROACH 1: HEAT MAP OF TYPE BY CATEGORY

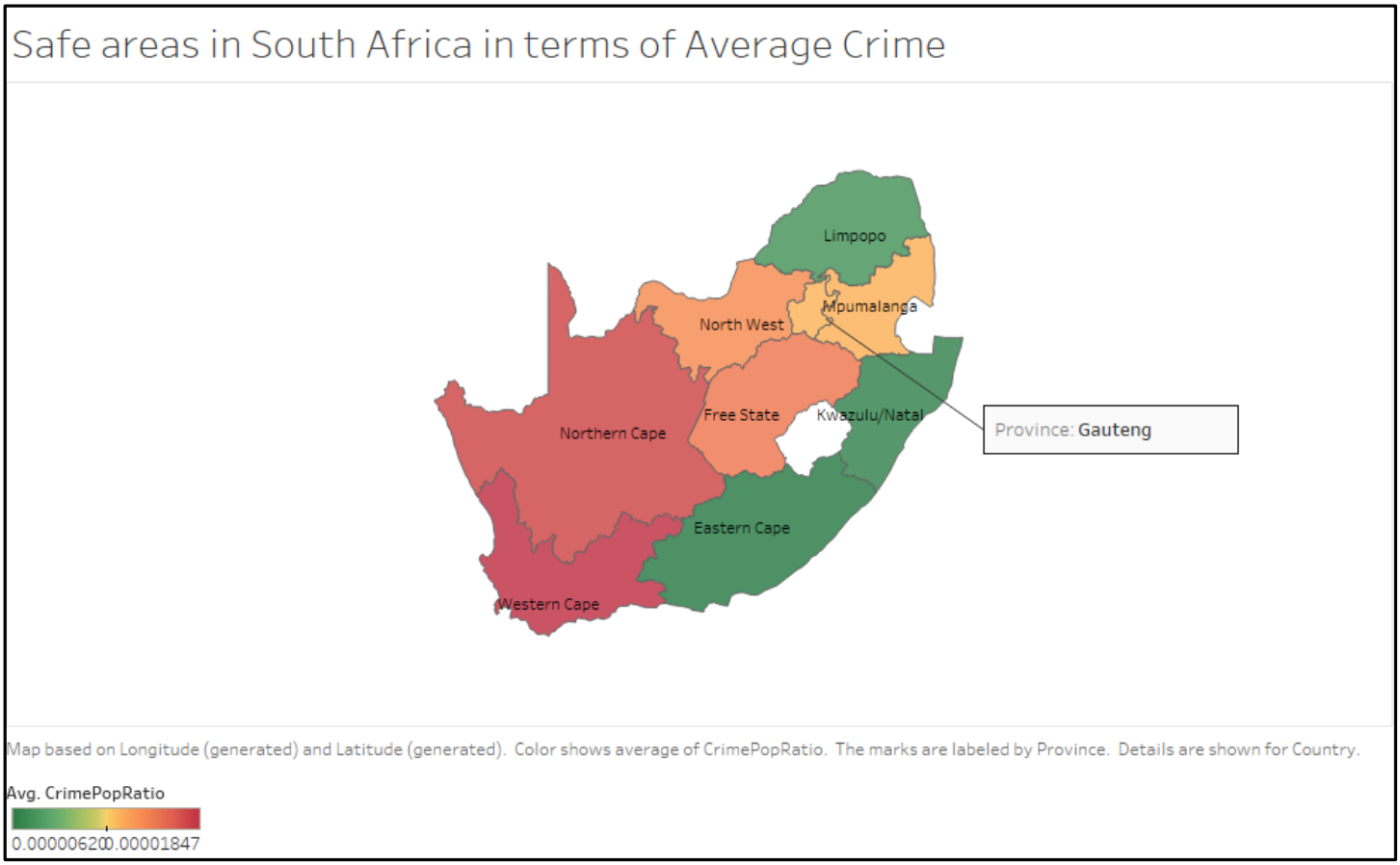


This heat map shows severities in each province with respect to crime statistic by color. We chose to use the log scale of the original crime statistic variable to show the raw distribution of the different severities of crimes across the provinces.

It is clear that low level crimes are the most recorded, with moderate and severe crimes close in numbers. Gauteng still holds the highest numbers in low level, moderate and severe crimes. Northern Cape appears to be the safest with the lowest rate of crime in all 3 levels crime.

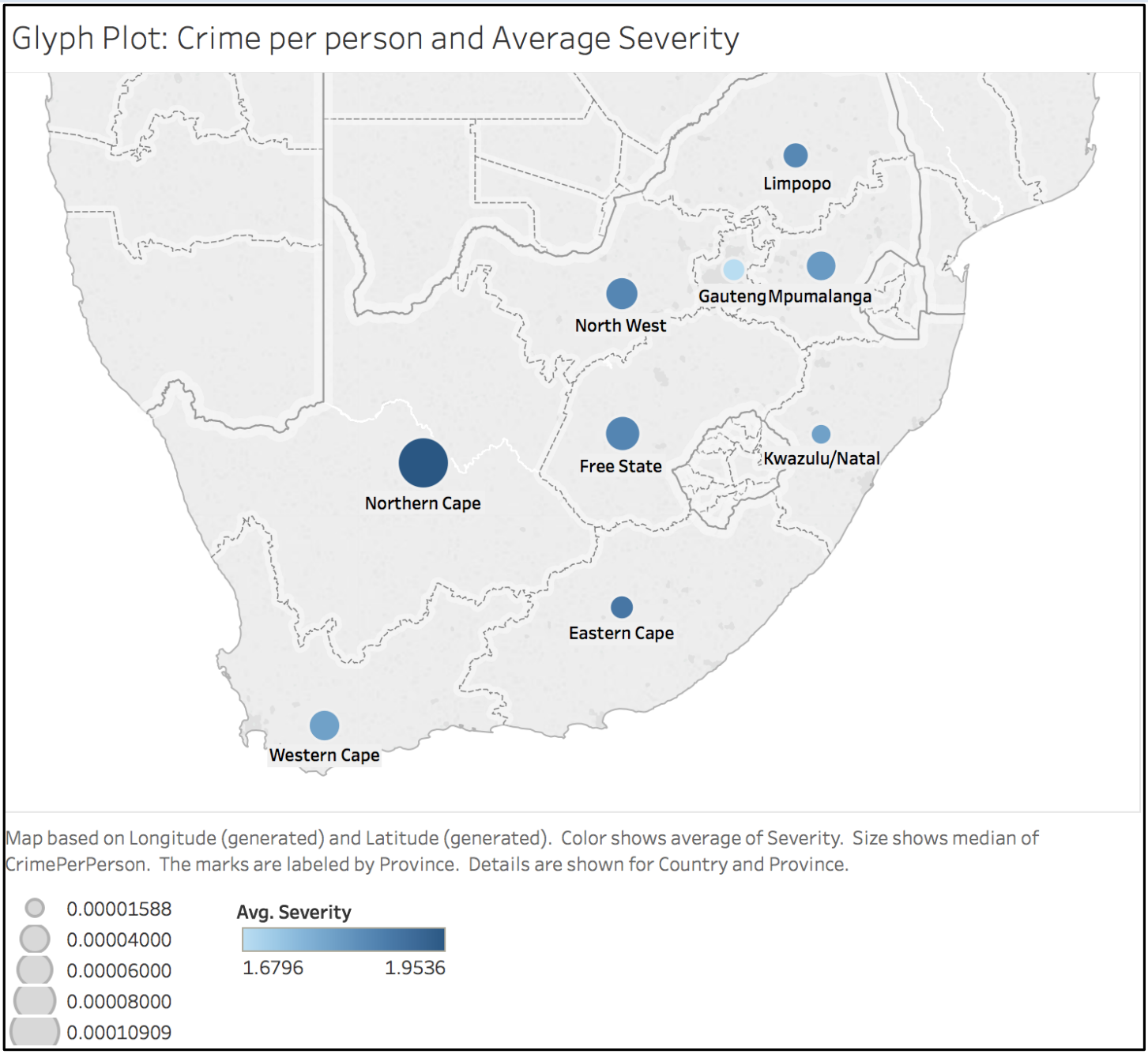
NOTE: This is the log of the initial total crime statistic variable, and not the crime per person variable (CrimePopRatio) we later created.

APPROACH 2: CHOROPLETH



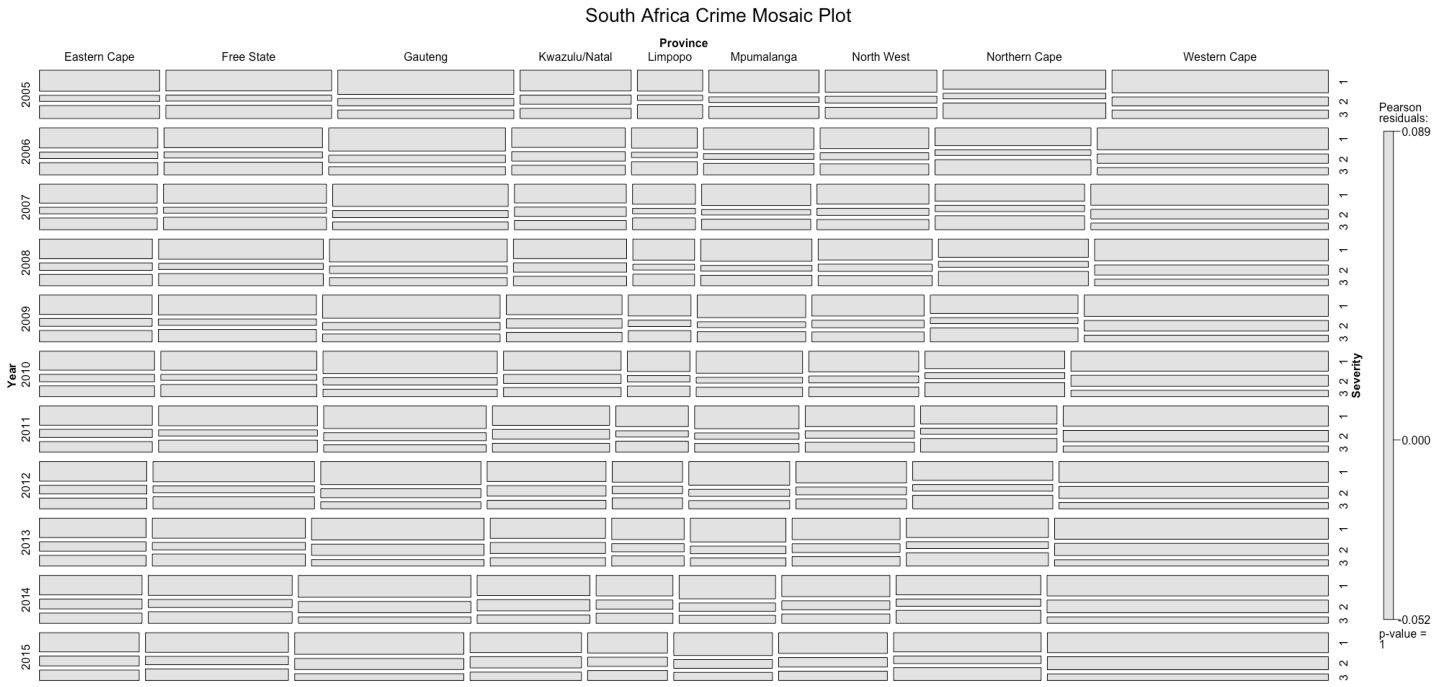
This choropleth depicts safe and unsafe areas in terms of crime and population ratio. It shows that Northern and Western Cape provinces have the most amount of crimes per person as opposed to Easter Cape, Limpopo and Kwazulu/Natal. The color choice for this choropleth is a diverging one that shows green as safe, red as dangerous and yellow as neutral.

APPROACH 3: GLYPH PLOT



This glyph plot was inspired by our choropleths. In addition to showing the crime population ration for each province, we wanted to show the average severity of the crimes committed in each province. Average severity was mapped to color, and crime population ratio (aka crime per person) was mapped to size. This allows us to see both average severity and crime per person at once, so we can have a fuller picture of how safe each province is.

APPROACH 4: CONTINGENCY/MOSAIC PLOT



This mosaic plot, though a very dense visualization of the data, paints a thorough picture of the relationship of year, province and severity for each of the nine provinces.
Note: We attempted to color the plot using Pearson's residuals using shade=TRUE parameter. However, since the range for the residuals was too small, the colors did not show up.

RESULTS AND DISCUSSION

We analysed the crime in South Africa over the span of ten years (mid_2004-mid_2016). Our initial findings were that there was no linear relationship between crime totals and time, and there was a positive correlation between population and crime rate. Also, we binned the 27 categories of crime into 3 levels of "severity" which proved more meaningful. We were most proud of our glyph plot, which allowed us show which provinces are safe or unsafe with respect to crime per person and average severity of crime.

From the glyph plot, one can see that Northern Cape has the most crime per person and also the highest average severity of crime. It is thus the most dangerous province in South Africa, and we would recommend avoiding it if traveling to South Africa. Gauteng has the least average severity of its crimes and a fairly low crime per person ratio, so we consider it the safest. In the future, we could analyse the same variables using demographic data or the police station variable from the original data.

APPENDIX

R CODE:

```
library(ggplot2)
#set directory
setwd("/Users/sarahcummings/Documents/csc465")

# read in our three aggregated files
data1= read.csv('ProjectData_465.csv')
data2=read.csv('Crime by Province with population.csv')
data3=read.csv('Crime by time.csv')

#create total crime over time line graph
t<-ggplot(data3,aes(Year,Crime.Total))+geom_line()
t+ggtitle("Crime Total in South Africa Over Time")

#create scatterplot of crime total by province
r<-ggplot(data2, aes(Poulation,Crime.Total))+geom_point()+xlab('Population')+ggtitle('Provice Crime Totals by Population')
#color the points according to province
r2<-ggplot(data2, aes(Poulation,Crime.Total,colour=Province))+geom_point()+xlab('Population')+ggtitle('Provice Crime Totals by Population')

#mosaic plot
d3<-read.csv('SouthAfricaCrimeStats_v2_Reshaped_NEW_06.csv')
d3$Year<-d3$Year2
tab2=xtabs(CrimPopRatio~Year+Province+Severity,data=d3)
mosaic(data=tab2,~Year+Province+Severity, shade=TRUE,cex=2.5,main="South Africa Crime Mosaic Plot",legend=TRUE
```