HAMLINE UNIVERSITY

Data Mining: Problem Set 1

Your Name*

Insert Date Here

> The purpose of this document is to simulataneously analyze data on US crime rates and become more familiar with the syntax and abilities of R-markdown to combine code and analysis in a progressional document. Blockquotes look better in HTML typically, but you can see their general effect in any document. The text is highlighted differently in RStudio so you know its part of the block quote. Also, the margins of the text in the final document are narrower to separate the block quote from normal text.

The Structure of the Data

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

```
## 'data.frame':    50 obs. of  4 variables:
##  $ Murder  : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
##  $ Assault : int  236 263 294 190 276 204 110 238 335 211 ...
##  $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
##  $ Rape    : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus

---

*Email youremail@hamline.edu. Position Your position here

nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.
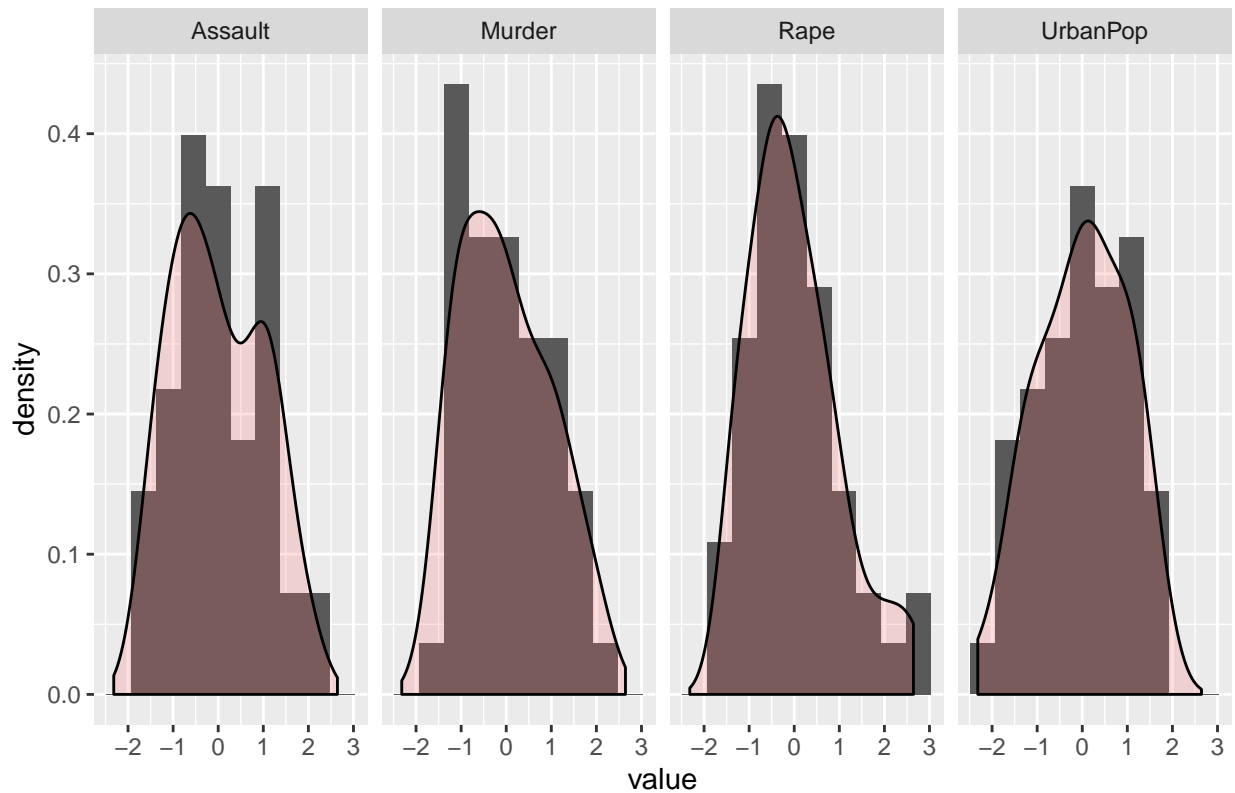
Summary of Features

Murder	Assault	UrbanPop	Rape
Min. : 0.800	Min. : 45.0	Min. :32.00	Min. : 7.30
1st Qu.: 4.075	1st Qu.:109.0	1st Qu.:54.50	1st Qu.:15.07
Median : 7.250	Median :159.0	Median :66.00	Median :20.10
Mean : 7.788	Mean :170.8	Mean :65.54	Mean :21.23
3rd Qu.:11.250	3rd Qu.:249.0	3rd Qu.:77.75	3rd Qu.:26.18
Max. :17.400	Max. :337.0	Max. :91.00	Max. :46.00

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

```r
# Make sure that this code block shows up in the final document
# and that the resulting plot does also.
library(ggplot2)
library(tidyr)
scaled_data = as.data.frame(sapply(USArrests, scale))
ggplot(gather(scaled_data, cols, value), aes(x = value)) +
    geom_histogram(aes(y=..density..), bins = 10) +
    geom_density(alpha=.2, fill="#FF6666") +
    facet_grid(.~cols) +
    ggtitle("Feature Histograms for the Scaled US Arrests Data")
```

Feature Histograms for the Scaled US Arrests Data

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

Relationships Between Features

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus mi arcu, eleifend at tellus nec, imperdiet pellentesque nulla. Duis mollis diam et tincidunt faucibus. Fusce imperdiet consectetur nulla sit amet vehicula. Phasellus vitae porttitor nisl, non tempus ipsum. Aenean

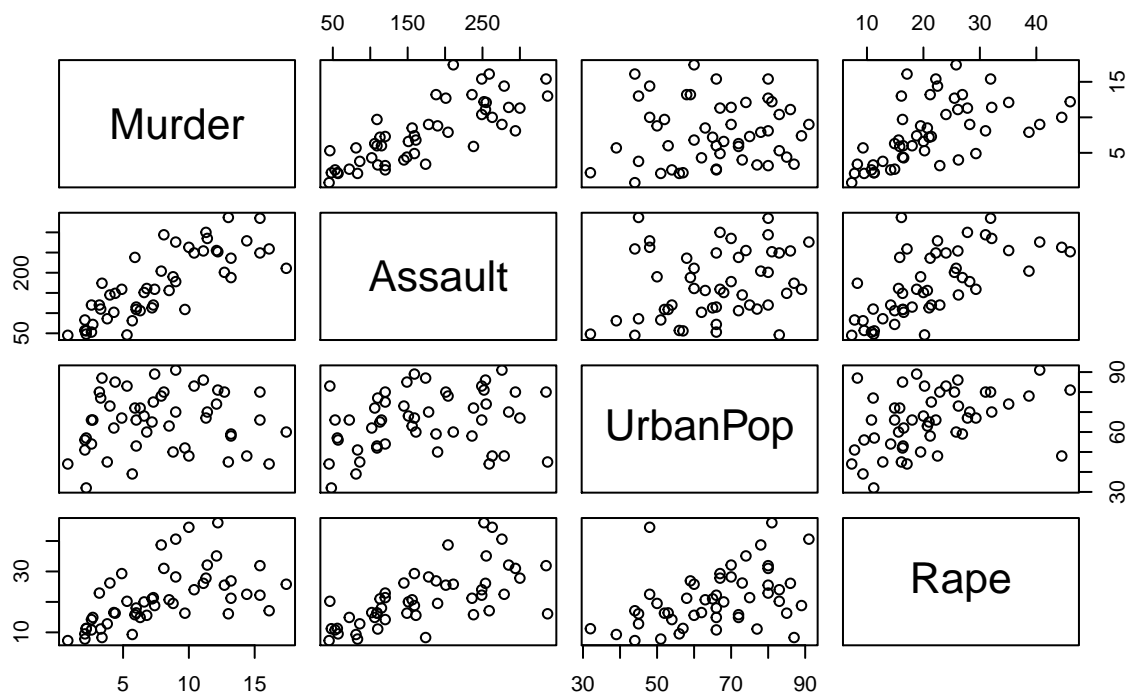## Scatter Plots of Crime Rates and Urban Population

Figure 1: Facet Grid of Scatter Plots

viverra massa et nisl viverra euismod. Nam in lobortis libero. Sed laoreet erat ac diam dapibus malesuada.

Variable	Mean
Murder	7.788
Assault	170.76
UrbanPop	65.54
Rape	21.232