DSC640 WEEK 7 & 4

Cindy Herrera

1/31/2020

4.2 Exercises: Scatterplots, Bubble Charts & Density Plots

You need to submit 3 scatterplots, 3 bubble charts and 3 density plot charts using Tableau or PowerBI, Python and R using the data below (or your own datasets). You can also submit using D3. You can choose which library to use in Python or R, documentation is provided to help you decide and as you start to play around in the libraries, you will decide which you prefer.

Exercise 4.2 Datasets https://content.bellevue.edu/cst/dsc/640/datasets/ex4-2.zip

```
library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(magrittr)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v tibble  2.1.3     v purrr   0.3.3
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ t
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
```

```
library(hrbrthemes)
```

```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
```

```
##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
```

```
##       if Arial Narrow is not on your system, please see http://bit.ly/arialnarrow
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(ggrepel)
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

IMPORTING DATASET

here we are setting up the file paths for crimerates-by-state-2005.csv

```
## 'data.frame':    52 obs. of  9 variables:
##  $ state              : Factor w/ 52 levels "Alabama","Alaska",..: 45 1 2 3 4 5 6 7 8 9 ...
##  $ murder             : num  5.6 8.2 4.8 7.5 6.7 6.9 3.7 2.9 4.4 35.4 ...
##  $ forcible_rape      : num  31.7 34.3 81.1 33.8 42.9 26 43.4 20 44.7 30.2 ...
##  $ robbery            : num  140.7 141.4 80.9 144.4 91.1 ...
##  $ aggravated_assault : num  291 248 465 327 387 ...
##  $ burglary           : num  727 954 622 948 1085 ...
##  $ larceny_theft      : num  2286 2650 2599 2965 2711 ...
##  $ motor_vehicle_theft: num  417 288 391 924 262 ...
##  $ population         : int  295753151 4545049 669488 5974834 2776221 35795255 4660780 3477416 83990
```

here we are setting up the file paths for life-expectancy.csv

```
le <- read.csv("~/Desktop/00 data640/ex4-2/life-expectancy.csv")
str(le)
```

```
## 'data.frame':    187 obs. of  3 variables:
##  $ country   : Factor w/ 187 levels "Afghanistan",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ year      : int  2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 ...
##  $ expectancy: int  42 73 71 46 74 76 70 82 80 68 ...
```

```
head(cbs)
```

```
##           state murder forcible_rape robbery aggravated_assault burglary
## 1 United States    5.6          31.7   140.7             291.1    726.7
## 2       Alabama    8.2          34.3   141.4             247.8    953.8
## 3        Alaska    4.8          81.1    80.9             465.1    622.5
```

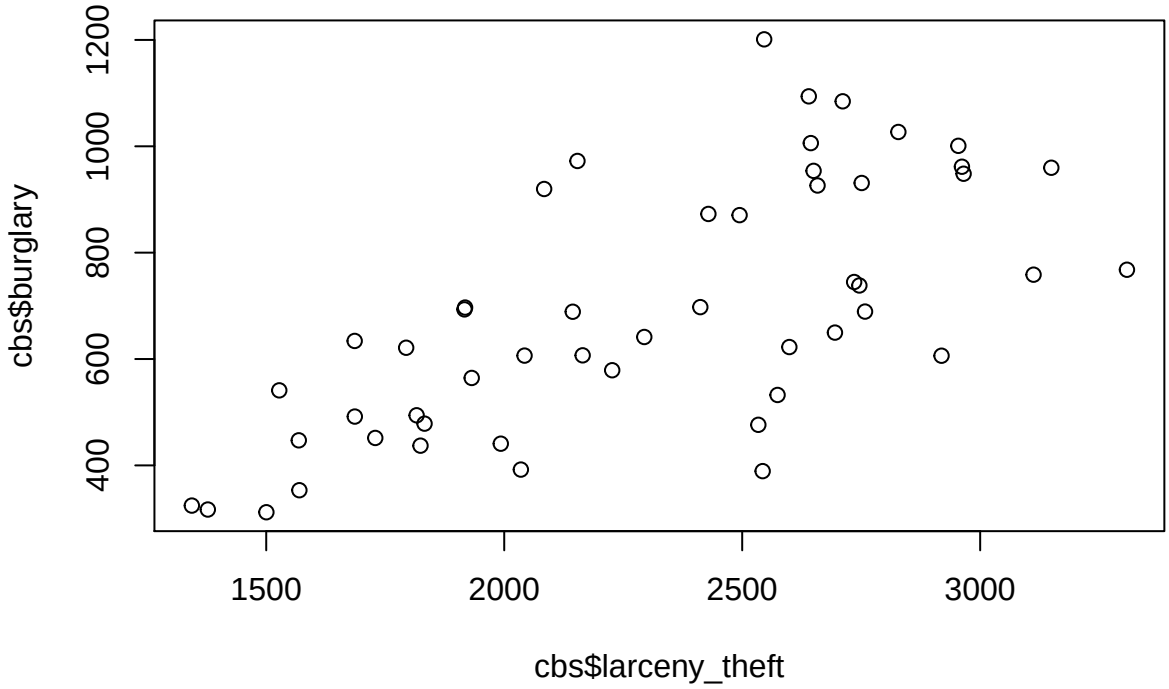
```
## 4       Arizona    7.5          33.8   144.4              327.4    948.4
## 5      Arkansas    6.7          42.9    91.1              386.8   1084.6
## 6    California    6.9          26.0   176.1              317.3    693.3
##   larceny_theft motor_vehicle_theft population
## 1        2286.3               416.7  295753151
## 2        2650.0               288.3    4545049
## 3        2599.1               391.0     669488
## 4        2965.2               924.4    5974834
## 5        2711.2               262.1    2776221
## 6        1916.5               712.8   35795255
```

```
cbs = cbs[-1,]

head(cbs)
```

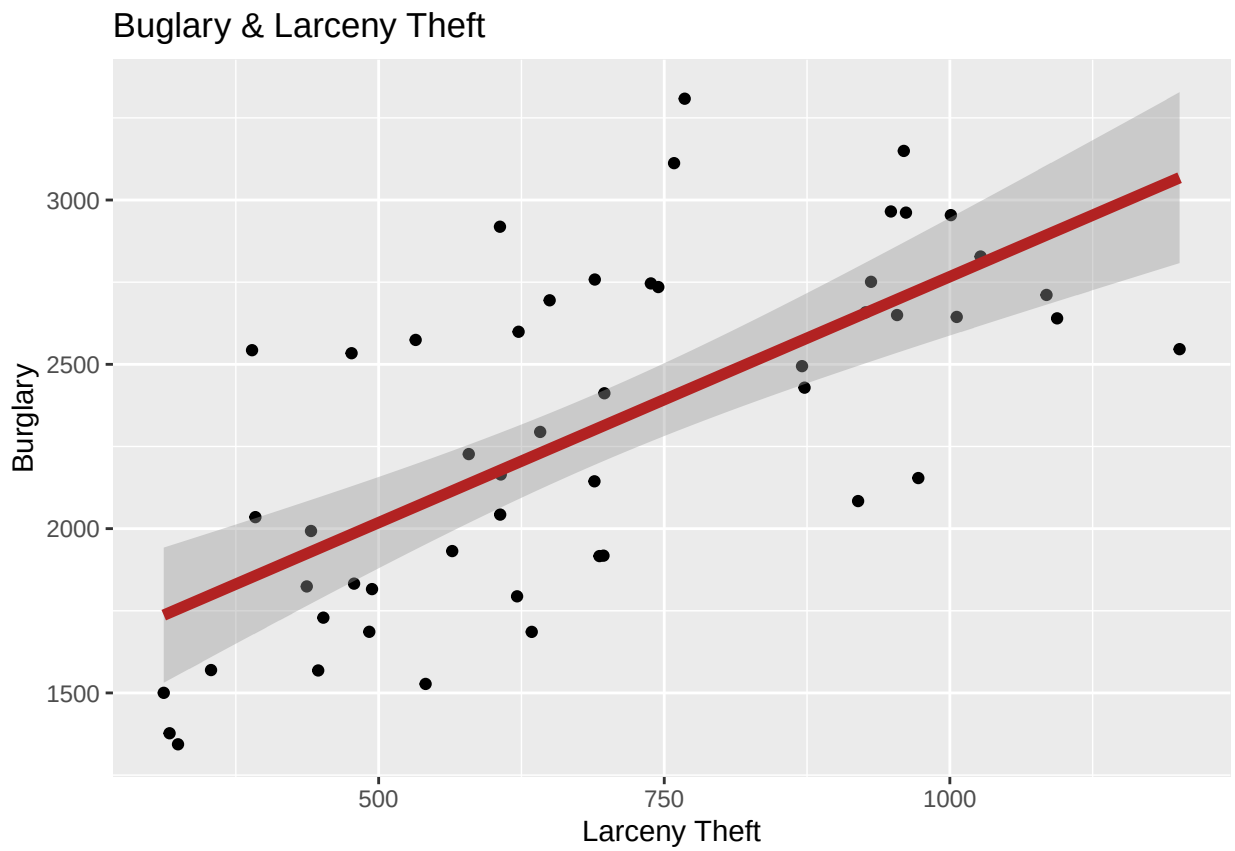
```
##        state murder forcible_rape robbery aggravated_assault burglary
## 2    Alabama    8.2          34.3   141.4              247.8    953.8
## 3     Alaska    4.8          81.1    80.9              465.1    622.5
## 4    Arizona    7.5          33.8   144.4              327.4    948.4
## 5   Arkansas    6.7          42.9    91.1              386.8   1084.6
## 6 California    6.9          26.0   176.1              317.3    693.3
## 7   Colorado    3.7          43.4    84.6              264.7    744.8
##   larceny_theft motor_vehicle_theft population
## 2        2650.0               288.3    4545049
## 3        2599.1               391.0     669488
## 4        2965.2               924.4    5974834
## 5        2711.2               262.1    2776221
## 6        1916.5               712.8   35795255
## 7        2735.2               559.5    4660780
```

```
plot(cbs$larceny_theft, cbs$burglary)
```

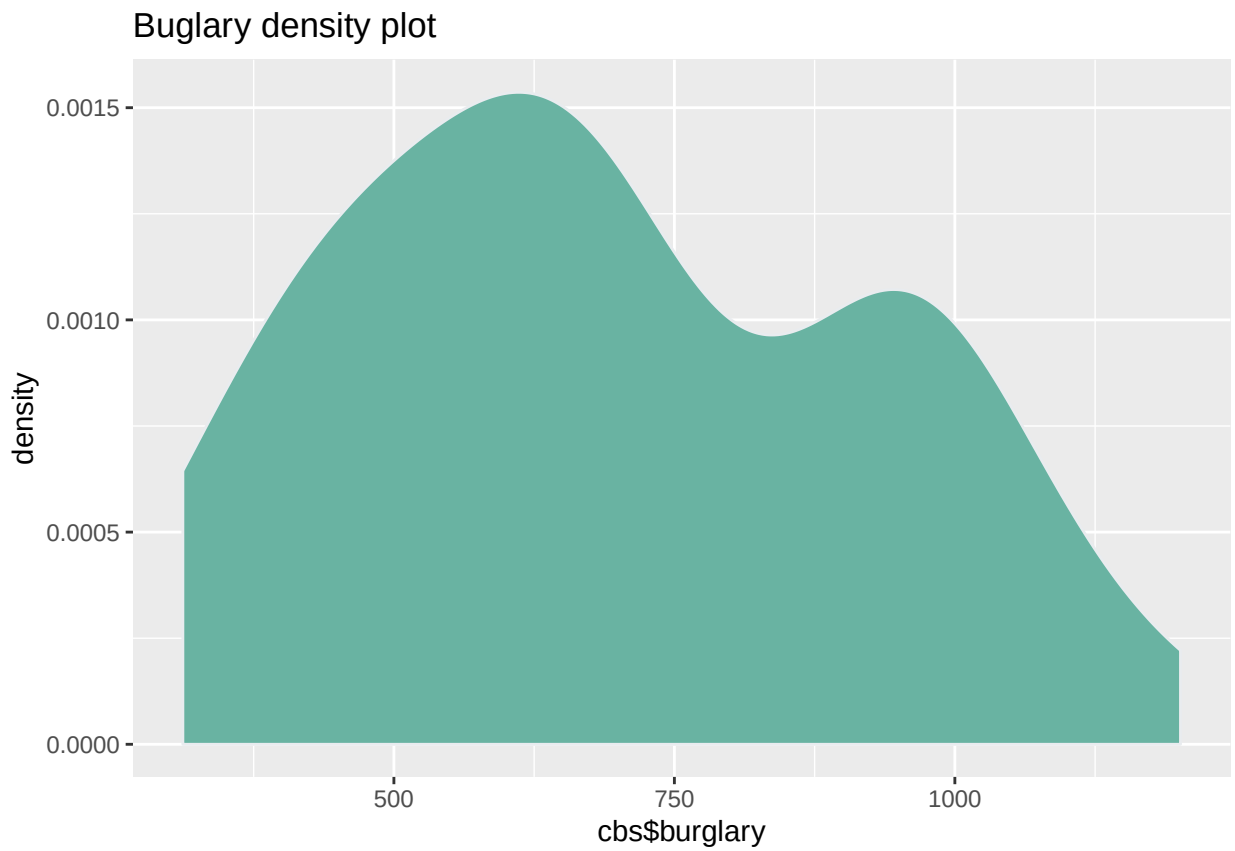


Add some color to the scatter plot

```
ggplot(cbs, aes(x = cbs$burglary, y = cbs$larceny_theft)) +
  geom_point() +
  geom_smooth(method = "lm", col = "firebrick", size =  2) +
  labs(title = "Buglary & Larceny Theft", y = "Burglary", x = "Larceny Theft")
```




```
ggplot(cbs, aes(x = cbs$burglary)) +
  geom_density(fill="#69b3a2", color="#e9ecef") +
  labs(title = "Buglary density plot" )
```


Buglary density plot
0.0015
0.0010
0.0005
0.0000
density
500
750
1000
cbs$burglary