Assignment 6.2_Python:

Import modules

```
In [1]:    1  import pandas as pd
           2  import matplotlib.pyplot as plt
           3  import seaborn as sns
           4  import math
           5  from matplotlib.ticker import FuncFormatter
           6  import plotly
           7  import plotly.figure_factory as ff
           8  from pandas.plotting import parallel_coordinates
           9  import numpy as np
          10
          11  %matplotlib inline
```

Data load and transformation

```
In [2]:    1  education = pd.read_csv('education.csv')
           2  crime = pd.read_csv('crimeratesbystate-formatted.csv')
           3  birthrate = pd.read_csv('birth-rate.csv')
           4
           5  # remove whitespaces from crime dataset (sine we have already encounte
           6  education = education.applymap(lambda x: x.strip() if type(x) is str e
           7  crime = crime.applymap(lambda x: x.strip() if type(x) is str else x)
           8  birthrate = birthrate.applymap(lambda x: x.strip() if type(x) is str e
```

Histogram

```
In [3]:    1  # Distribution of birth rate
           2  birthrate_hist = pd.melt(birthrate, id_vars="Country", var_name="Year"
           3  birthrate_hist["BirthRate_int"] = birthrate_hist["BirthRate"].apply(la
           4  birthrate_hist.head()
```

Out[3]:

	Country	Year	BirthRate	BirthRate_int
0	Aruba	1960	36.400	37
1	Afghanistan	1960	52.201	53
2	Angola	1960	54.432	55
3	Albania	1960	40.886	41
4	Netherlands Antilles	1960	32.321	33

```
In [4]:  ▶  1  sns.distplot( birthrate_hist["BirthRate_int"] )
```

C:\Users\debas\anaconda3\lib\site-packages\seaborn\distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in
a future version. Please adapt your code to use either `displot` (a figur
e-level function with similar flexibility) or `histplot` (an axes-level f
unction for histograms).
  warnings.warn(msg, FutureWarning)

Out[4]: <AxesSubplot:xlabel='BirthRate_int', ylabel='Density'>
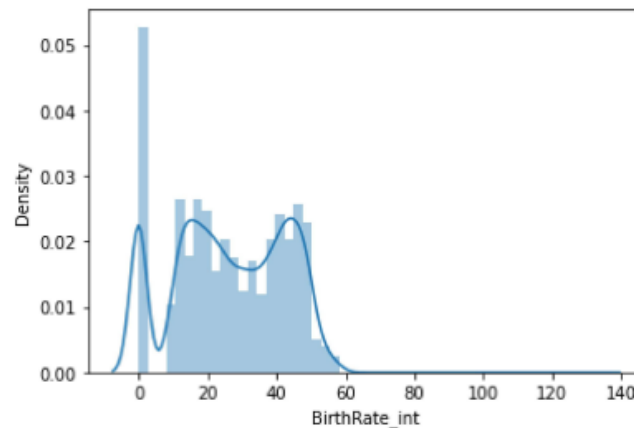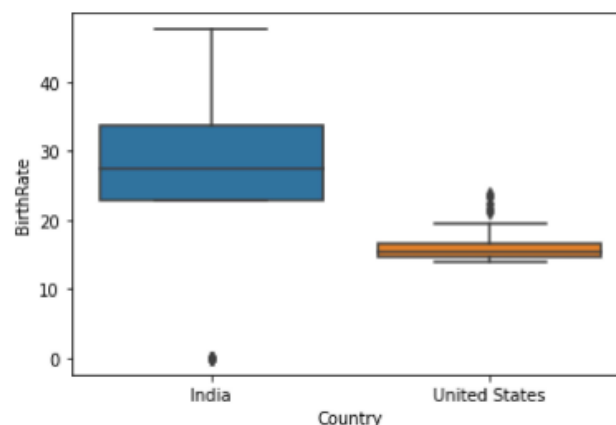
Box plot

Comparison of birthrate betwen India and USA

```
In [5]:  ▶  1  birthrate_box = birthrate_hist[(birthrate_hist["Country"]=="United Sta
            2  sns.boxplot(x = birthrate_box["Country"], y=birthrate_box["BirthRate"]
```

Out[5]: <AxesSubplot:xlabel='Country', ylabel='BirthRate'>

Bullet chart

In [7]: ▶|

```python
 1  ## US burglary statistics against some dummy benchmark
 2  # transform data
 3  crime_bullet = crime[crime["state"]=="United States"][["state","burgla
 4  crime_bullet['target'] = 500
 5  crime_bullet_tuple = [tuple(x) for x in crime_bullet.values][0]
 6
 7
 8  # set parameter for bullet chart
 9  limits = [300, 500, 1000]
10  palette = sns.color_palette("Blues_r", len(limits))
11  fig, ax = plt.subplots()
12  ax.set_aspect('equal')
13  ax.set_yticks([1])
14  ax.set_yticklabels='United States'
15
16  prev_limit = 0
17  for idx, lim in enumerate(limits):
18      ax.barh([1], lim-prev_limit, left=prev_limit, height=75, color=pal
19      prev_limit = lim
20
21  # draw the value we're measuring
22  ax.barh([1], crime_bullet_tuple[1], color='black', height=45)
23
24  ax.axvline(crime_bullet_tuple[2], color="gray", ymin=0.10, ymax=0.9)
```

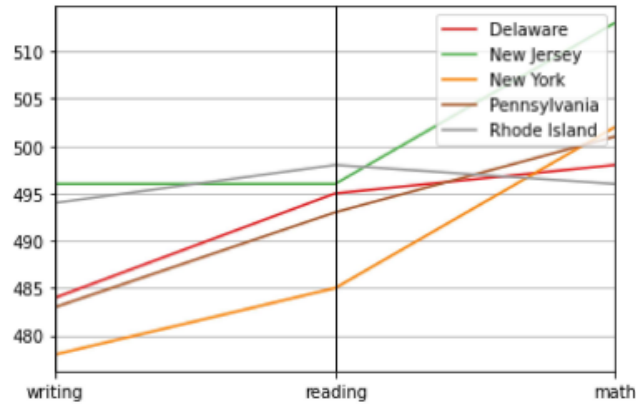Out[7]: <matplotlib.lines.Line2D at 0x1b9e031af70>

Parallel Coordinate plot

```
2  # transform data
3  education_parallel = education[education['state'].isin(['New York','Ne
4
5  # make the plot
6  parallel_coordinates(education_parallel, 'state', colormap=plt.get_cma
7  plt.show()
```
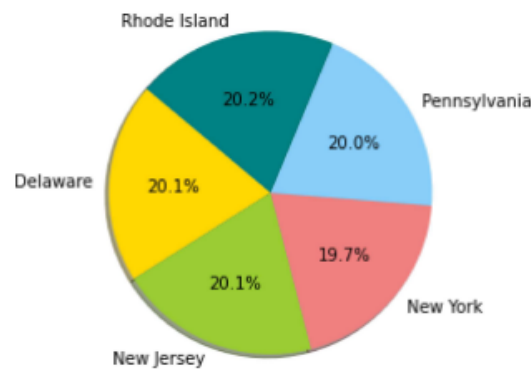

Pie chart

```
1  ##Comparison of reading numbers between 5 states
2  # transform data
3  education_pie = education_parallel[['state','reading']]
4
5  # set colors
6  colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue','teal']
7
8  # plot
9  plt.pie(education_pie['reading'], labels=education_pie['state'], color
10 autopct='%1.1f%%', shadow=True, startangle=140)
11
12 plt.axis('equal')
13 plt.show()
```

Donought chart

In [11]:

```python
#Comparison of reading, writing and math numbers between 5 states
# transform data
education_donut = education_pie

# create a pieplot
plt.pie(education_donut['reading'], labels=education_donut['state'])

# add a circle at the center
my_circle=plt.Circle( (0,0), 0.7, color='white')
p=plt.gcf()
p.gca().add_artist(my_circle)

plt.show()
```


In [ ]:

R:
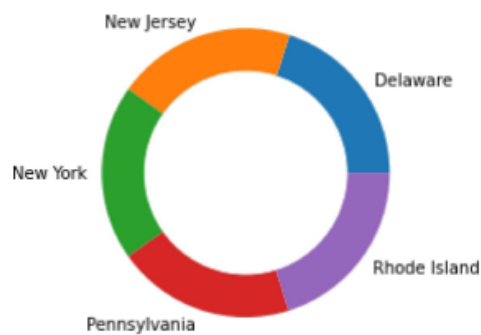
```r
# Course:DSC640
# Name:Debasish Panda
# Exercise 6.2


# Set Directory path
setwd("C:/Users/debas/OneDrive/Desktop/Bellevue/DSC 640/Week_11_12/ex6-2")

# Import required packages
library('magrittr')
library('ggmap')
library('ggplot2')
library('usmap')


# load birth rate data
birthrate <- read.csv('birth-rate.csv')

# load crime data
crime <- read.csv('crimeratesbystate-formatted.csv')

# load education data
education <- read.csv('education.csv')

# check column names
colnames(birthrate)

# format year columns
colnames(birthrate) <- gsub("X", "", colnames(birthrate))

# check column names
colnames(birthrate)

### Plot1: Histogram
options(repr.plot.width = 4, repr.plot.height = 4)

# create pivotted data for plotting
birthrate_hist <- reshape2::melt(birthrate, id=c("Country")) %>%
  dplyr::mutate("Country" = as.character(Country),
                "Year" = as.character(variable),
                "Population" = value,
                "Population_int" = ceiling(value)) %>%
  dplyr::select(c("Country","Year","Population", "Population_int"))

# create histogram of population data
hist(birthrate_hist$Population_int, col=rgb(0.2,0.8,0.5,0.5) , border=F ,
main="")

### Plot2: Box Plot
# create box plot of population data
birthrate_box <- birthrate_hist %>%
  dplyr::filter(Country %in% c("United States", "India"))
boxplot(birthrate_box$Population ~ birthrate_box$Country , col="grey")
```

```
### Plot3: Bullet Chart

# create bullet chart with crime data
crime_bullet <- crime %>%
  dplyr::filter(stringr::str_trim(state, "both") == "United States") %>%
  dplyr::select(c(state, burglary))
bulletgraph(x=crime_bullet$burglary,ref=650,limits=c(0,500,750,1000),
            name= "USA Crime metrics",subname="Burglary",
            col="steelblue4",shades="firebrick")


### Plot4: Donut Chart

# donut chart using USA crime data
education_donut <- education %>%
  dplyr::filter(stringr::str_trim(state, "both") == "United States") %>%
  reshape2::melt(id=c("state")) %>%
  dplyr::rename("category" = variable) %>%
  dplyr::filter(category %in% c("reading","math","writing")) %>%
  dplyr::select(-state)


# add addition columns, needed for drawing with geom_rect
education_donut$fraction = education_donut$value / sum(education_donut$value)
education_donut = education_donut[order(education_donut$fraction), ]
education_donut$ymax = cumsum(education_donut$fraction)
education_donut$ymin = c(0, head(education_donut$ymax, n=-1))

# make the plot
ggplot2::ggplot(education_donut, ggplot2::aes(fill=category, ymax=ymax,
ymin=ymin, xmax=4, xmin=3)) +
  ggplot2::geom_rect() +
  ggplot2::coord_polar(theta="y") +
  ggplot2::xlim(c(0, 4)) +
  ggplot2::theme(panel.grid=ggplot2::element_blank()) +
  ggplot2::theme(axis.text=ggplot2::element_blank()) +
  ggplot2::theme(axis.ticks=ggplot2::element_blank()) +
  ggplot2::annotate("text", x = 0, y = 0, label = "Education Components") +
  ggplot2::labs(title="")


### Plot5: Pie Chart

# pie chart
slices <- education_donut$value
lbls <- education_donut$category
pie(slices, labels = lbls, main="Education Components")
```
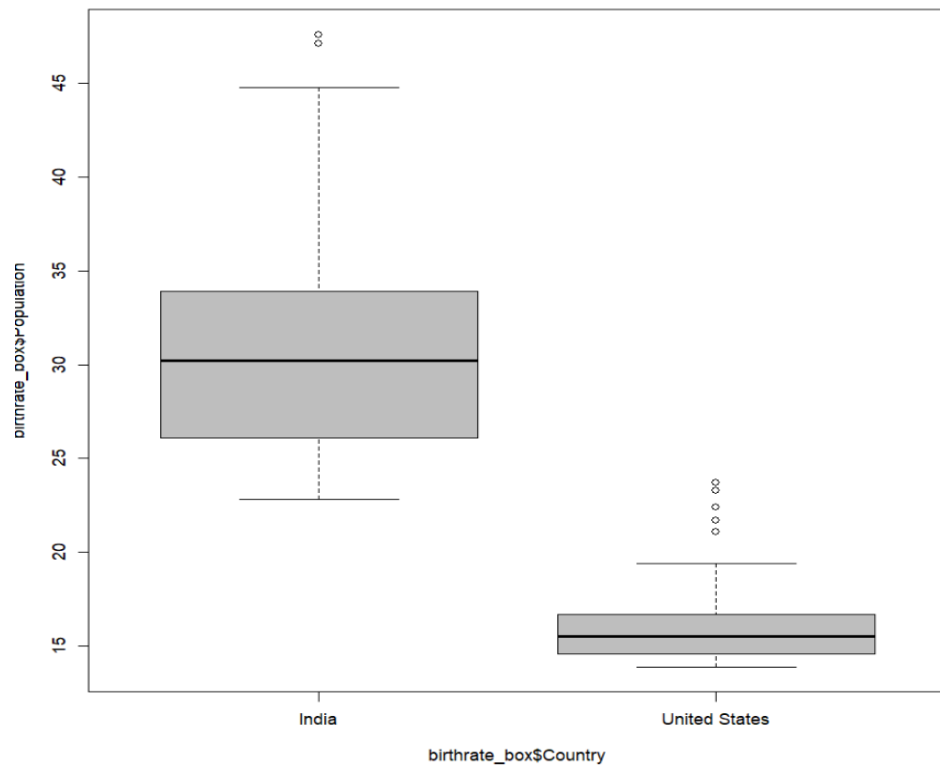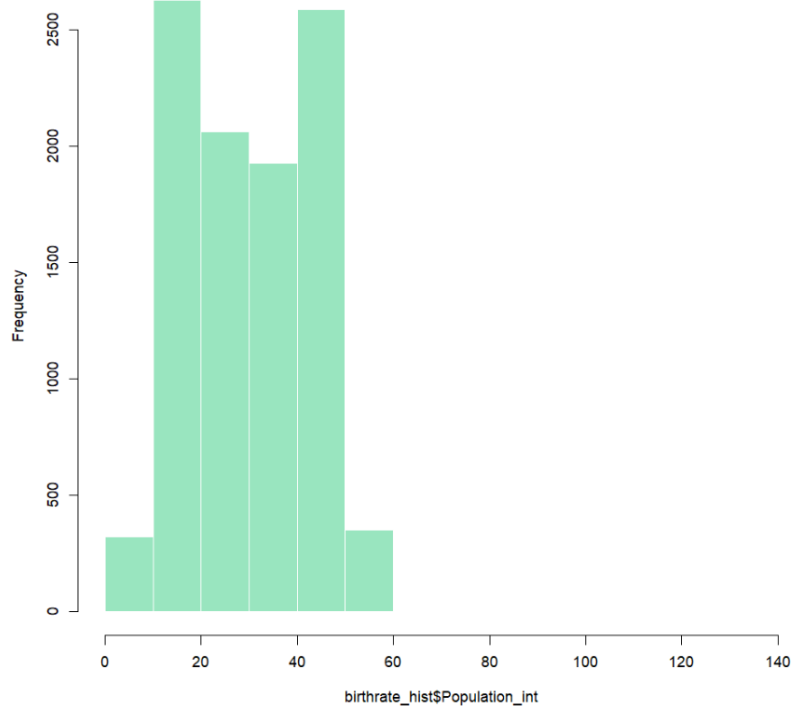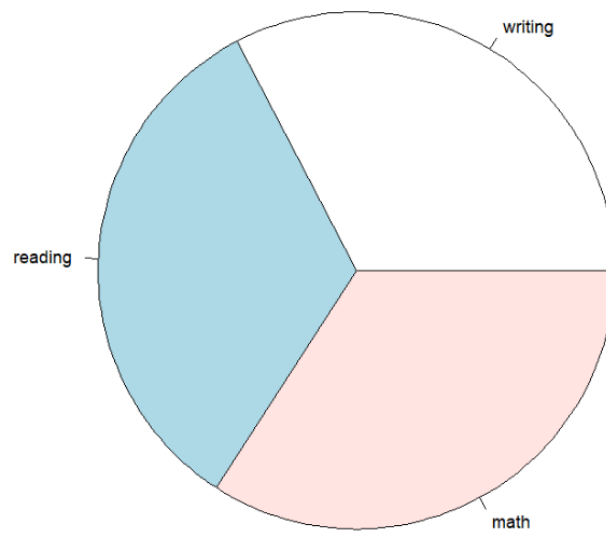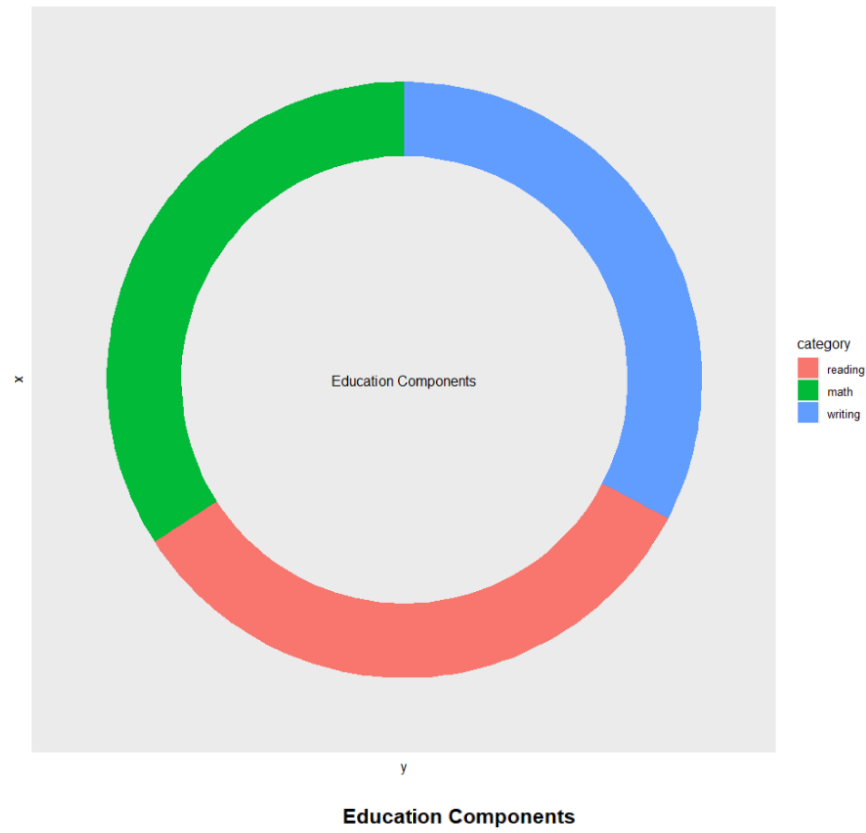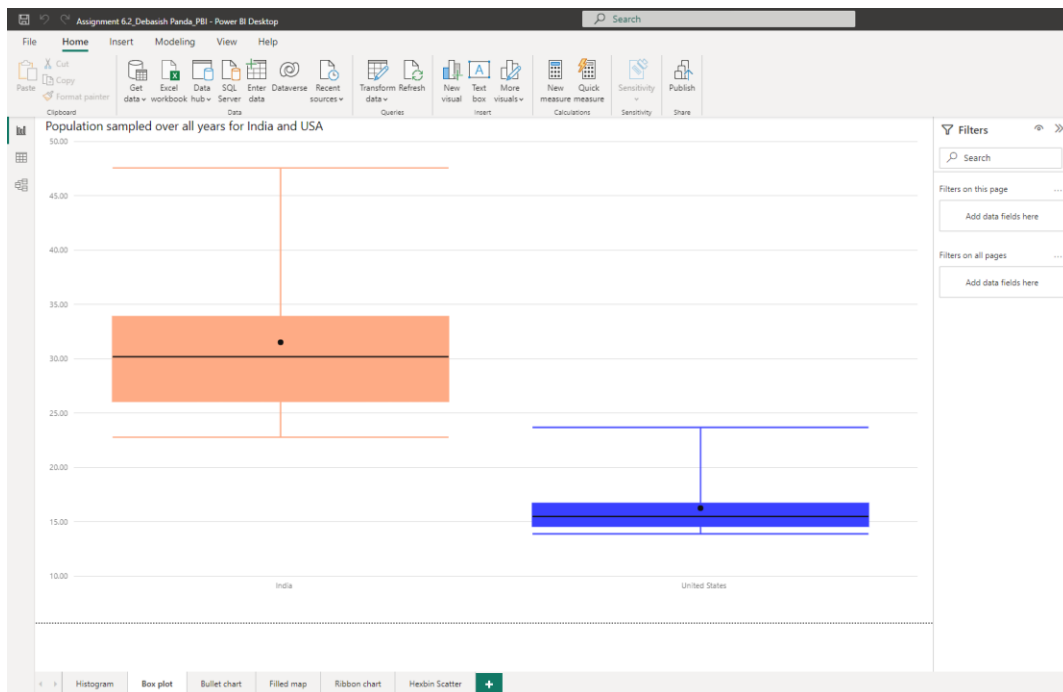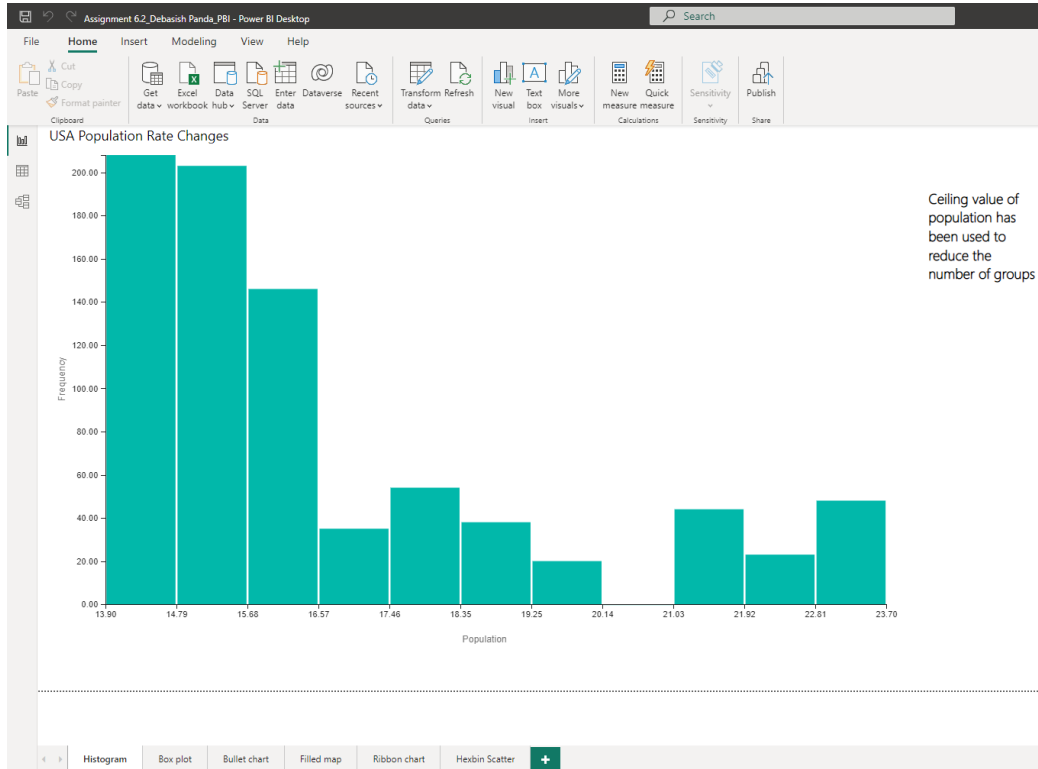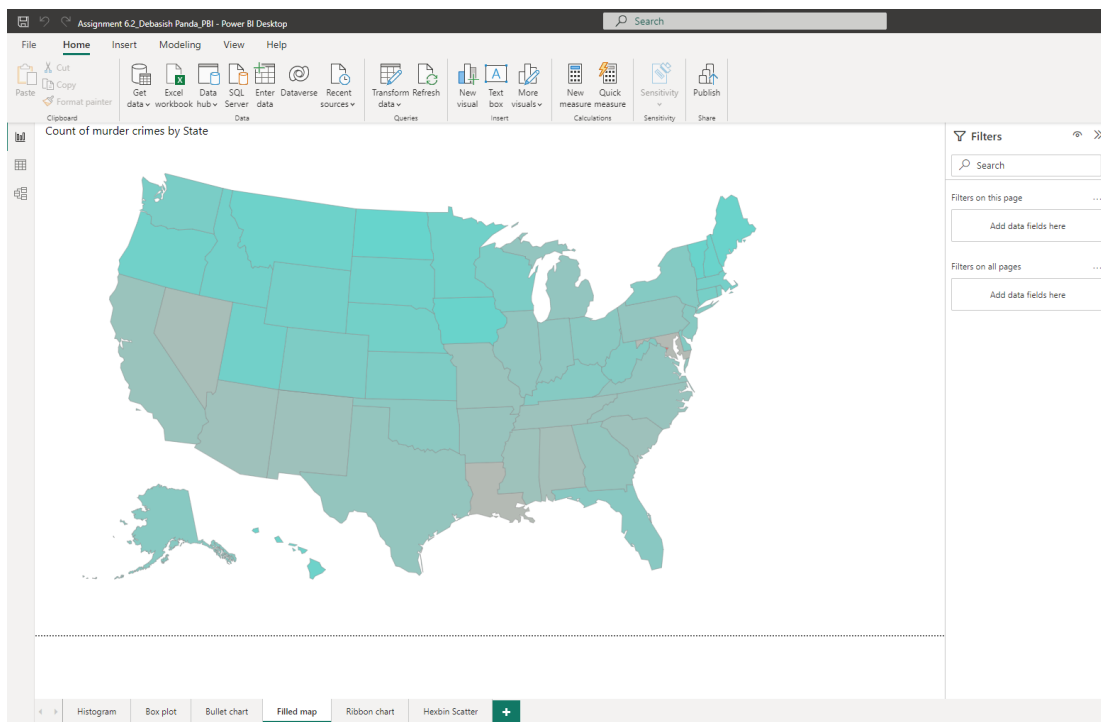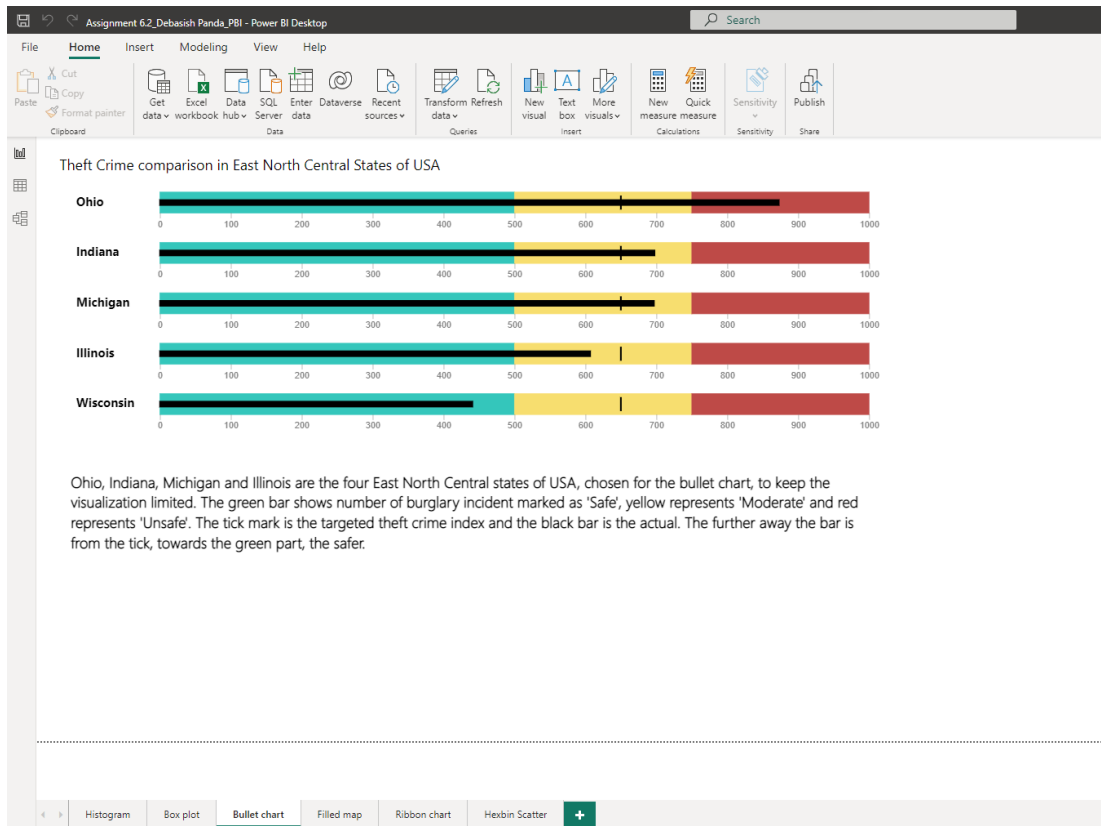
x

y

Education Components

Power BI:

Theft Crime comparison in East North Central States of USA

Ohio, Indiana, Michigan and Illinois are the four East North Central states of USA, chosen for the bullet chart, to keep the visualization limited. The green bar shows number of burglary incident marked as 'Safe', yellow represents 'Moderate' and red represents 'Unsafe'. The tick mark is the targeted theft crime index and the black bar is the actual. The further away the bar is from the tick, towards the green part, the safer.

Count of murder crimes by State

Population and population changes over year - Comparison between Portugal, South Africa and USA for the period of 1990 to 2008

Correlation of robbery and murder plotted in a hex bin scatter chart where each hexagon shows in density by color intensity

Bullet Graph: