

SumbarajuAditya_ex1_2_R_Refresh

SumbarajuAditya

8/30/2021

R Markdown

```
# Load Libraries
```

```
library(blsAPI)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
```

```
## 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(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

Load BLS data

```
#Consumer Price Index-All Urban Consumers (Current Series)
```

```
urban_cpi <- blsAPI('CUUR0000SA0', return_data_frame = TRUE)
```

```
# Chained CPI-All Urban Consumers
```

```
chained_cpi <- blsAPI('SUUR0000SA0', return_data_frame = TRUE)
```

check for right grain of data

```
summary(urban_cpi)
```

```
##      year          period      periodName      value
## Length:33      Length:33      Length:33      Length:33
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##      seriesID
## Length:33
```

```
## Class :character
## Mode :character
```

```
summary(chained_cpi)
```

```
##      year      period      periodName      value
## Length:31      Length:31      Length:31      Length:31
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##      seriesID
## Length:31
## Class :character
## Mode :character
```

```
str(urban_cpi)
```

```
## 'data.frame': 33 obs. of 5 variables:
## $ year : chr "2021" "2021" "2021" "2021" ...
## $ period : chr "M07" "M06" "M05" "M04" ...
## $ periodName: chr "July" "June" "May" "April" ...
## $ value : chr "273.003" "271.696" "269.195" "267.054" ...
## $ seriesID : chr "CUUR0000SA0" "CUUR0000SA0" "CUUR0000SA0" "CUUR0000SA0" ...
```

```
str(chained_cpi)
```

```
## 'data.frame': 31 obs. of 5 variables:
## $ year : chr "2021" "2021" "2021" "2021" ...
## $ period : chr "M07" "M06" "M05" "M04" ...
## $ periodName: chr "July" "June" "May" "April" ...
## $ value : chr "153.424" "152.720" "151.405" "150.221" ...
## $ seriesID : chr "SUUR0000SA0" "SUUR0000SA0" "SUUR0000SA0" "SUUR0000SA0" ...
```

```
dim(urban_cpi)
```

```
## [1] 33 5
```

```
dim(chained_cpi)
```

```
## [1] 31 5
```

Data cleansing and Merging Step

```
#Merging two datasets
```

```
df1 <- merge(urban_cpi, chained_cpi, by=c("year","period"))
```

```
#Renaming the columns
```

```
names(df1)[names(df1) == "value.x"] <- "urban_cpi_Rate"
```

```
names(df1)[names(df1) == "value.y"] <- "chained_cpi_Rate"
```

```
#Convert string representations to numbers
```

```
df1$urban_cpi_Rate <- as.numeric(df1$urban_cpi_Rate)
```

```
df1$chained_cpi_Rate <- as.numeric(df1$chained_cpi_Rate)
```

```
summary(df1)
```

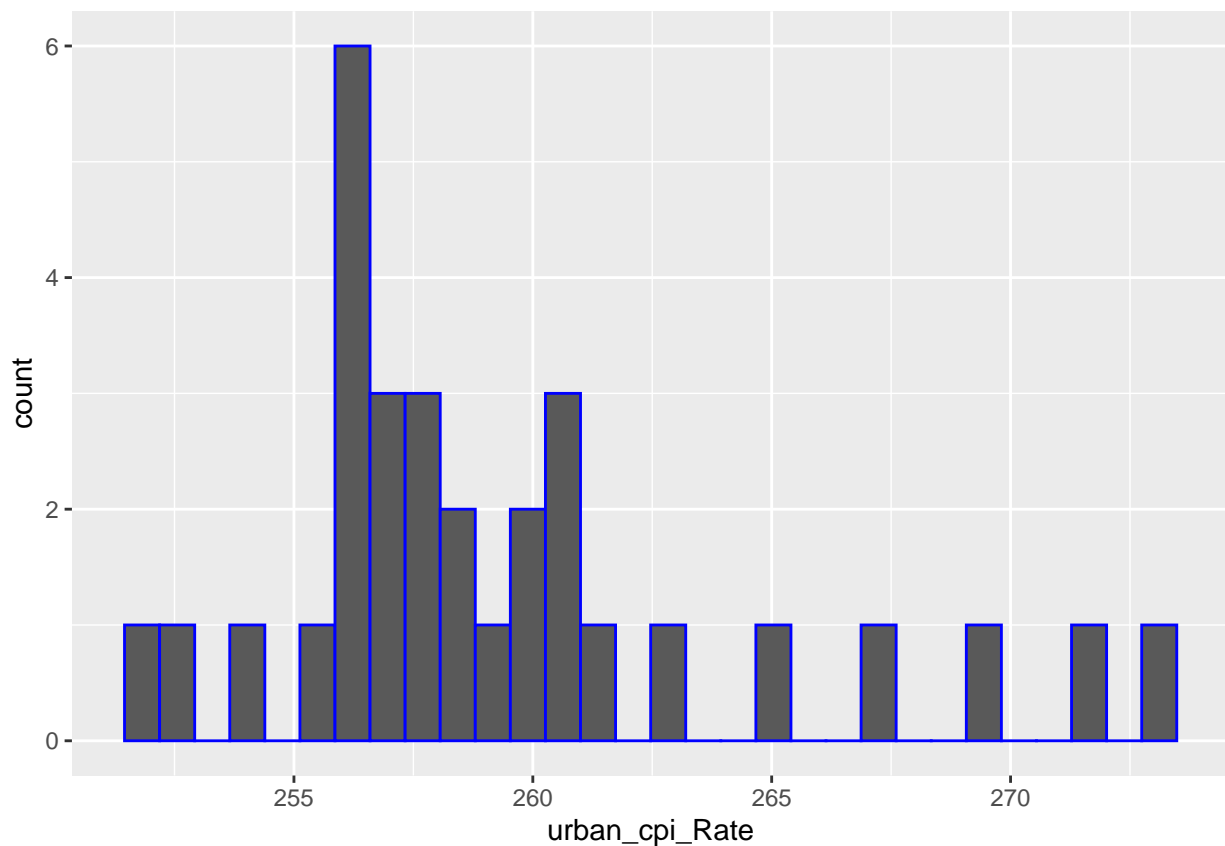
```
##      year      period      periodName.x      urban_cpi_Rate
## Length:31      Length:31      Length:31      Min. :251.7
```

```
## Class :character   Class :character   Class :character   1st Qu.:256.5
## Mode  :character   Mode  :character   Mode  :character   Median :258.0
##                                         Mean  :259.5
##                                         3rd Qu.:260.4
##                                         Max.   :273.0
## seriesID.x         periodName.y        chained_cpi_Rate seriesID.y
## Length:31          Length:31           Min.    :142.0   Length:31
## Class :character    Class :character    1st Qu.:144.2   Class :character
## Mode  :character    Mode  :character    Median :144.8   Mode  :character
##                                         Mean   :145.9
##                                         3rd Qu.:146.5
##                                         Max.   :153.4
```

#Histograms of the 2 variables

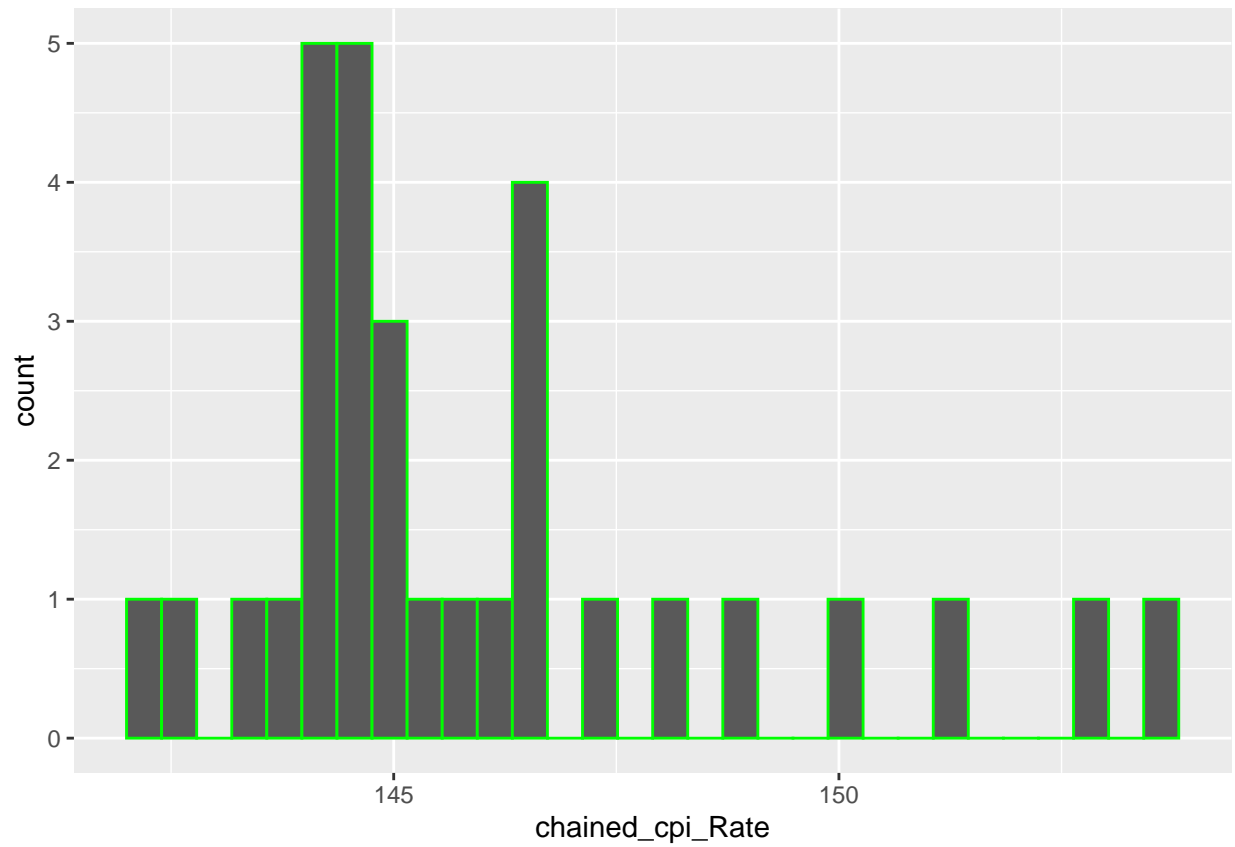
```
ggplot(df1, aes(x = urban_cpi_Rate)) + geom_histogram (col = c("blue"))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



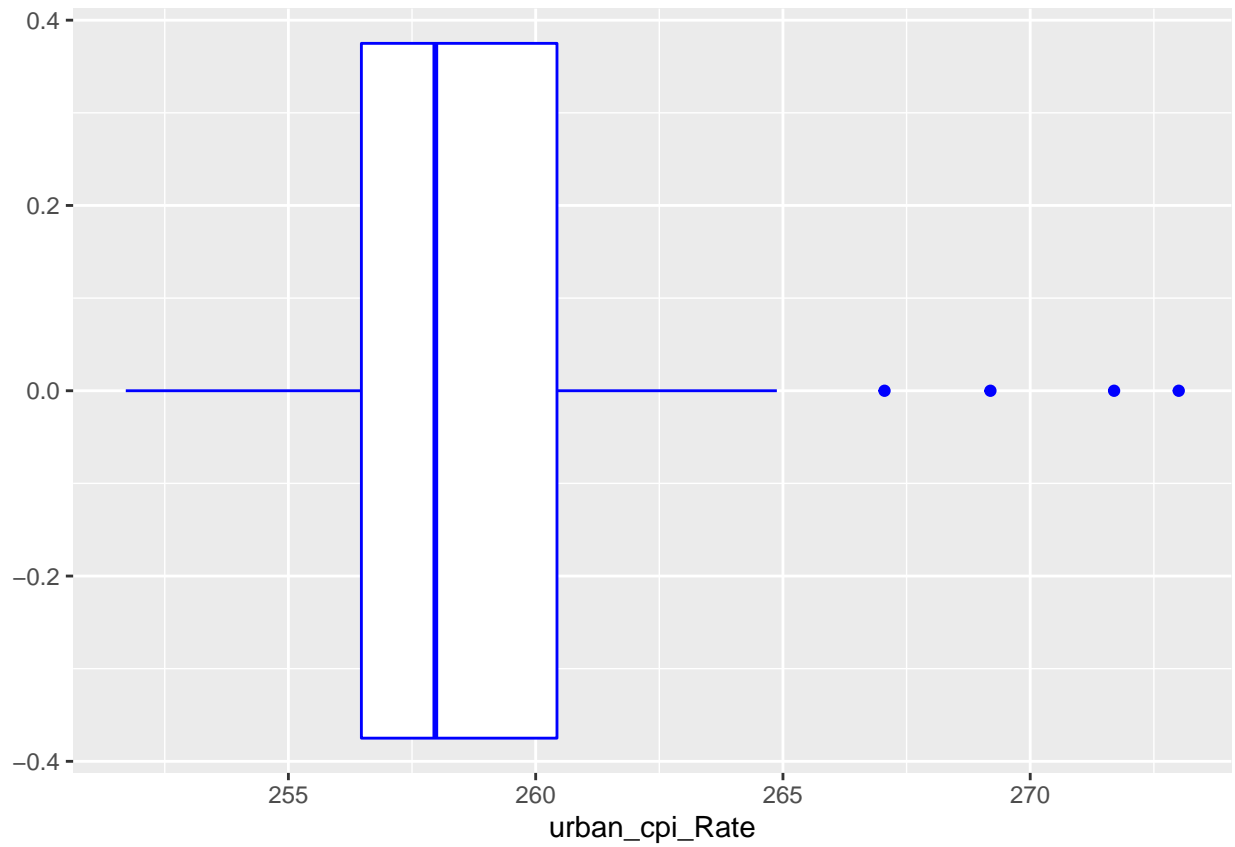
```
ggplot(df1, aes(x = chained_cpi_Rate)) + geom_histogram(col = c("green"))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

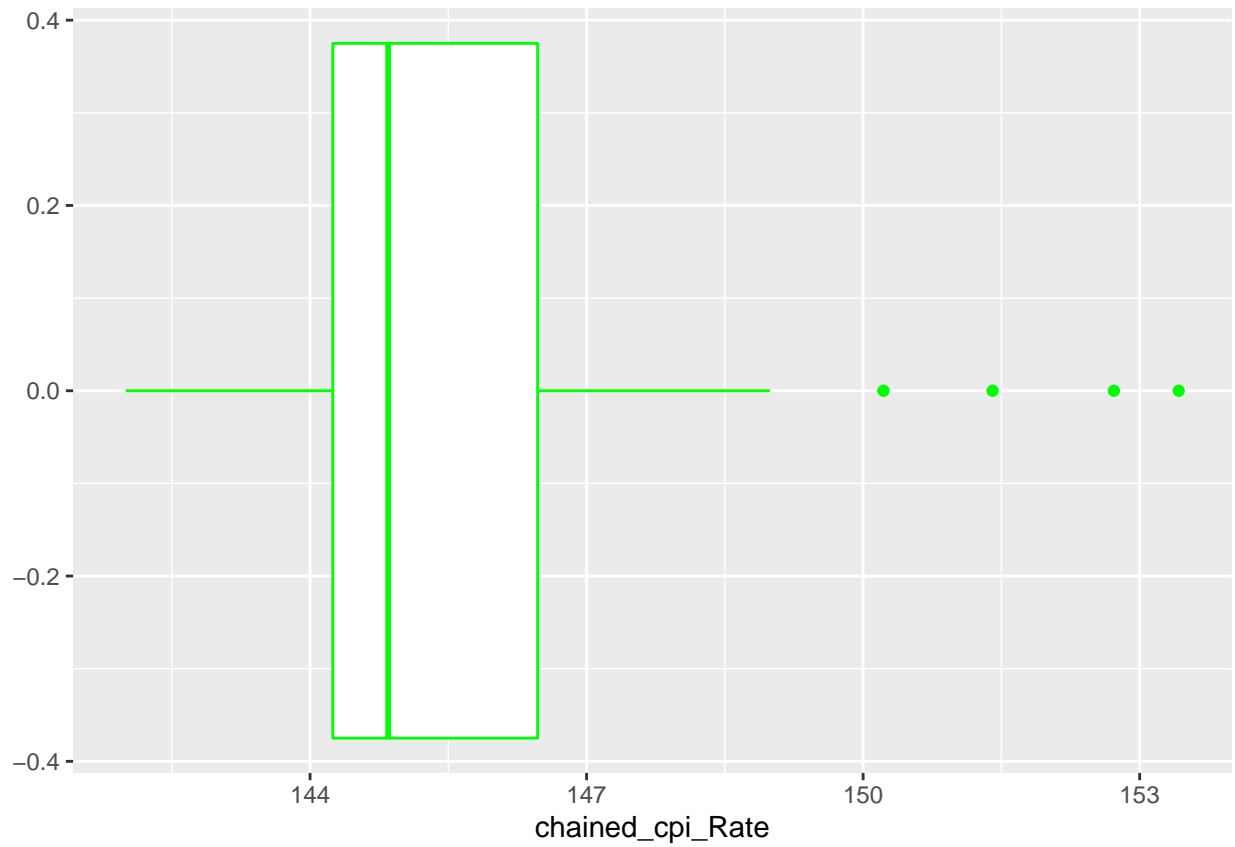


#Boxplots of the 2 variables

```
ggplot(df1, aes(x = urban_cpi_Rate)) + geom_boxplot(col = c("blue"))
```

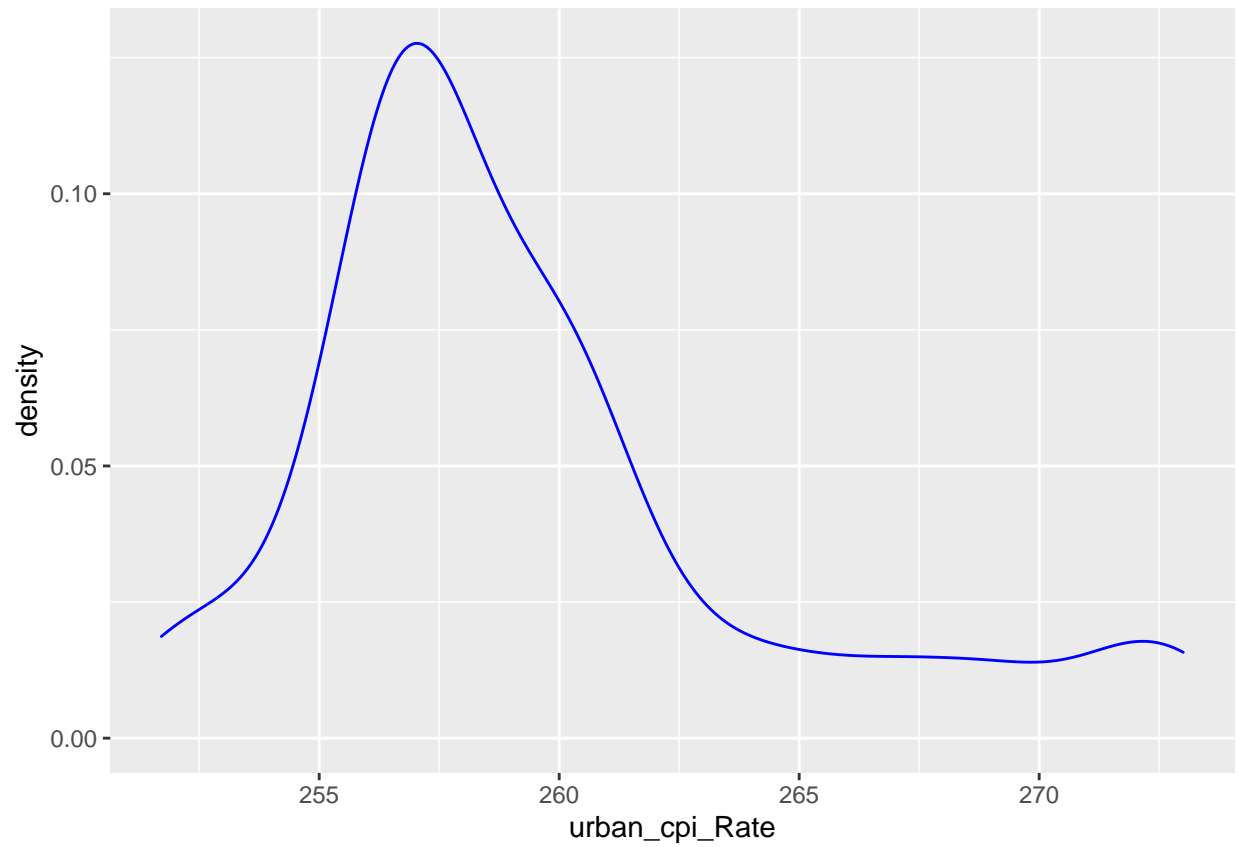


```
ggplot(df1, aes(x = chained_cpi_Rate)) + geom_boxplot(col = c("green"))
```

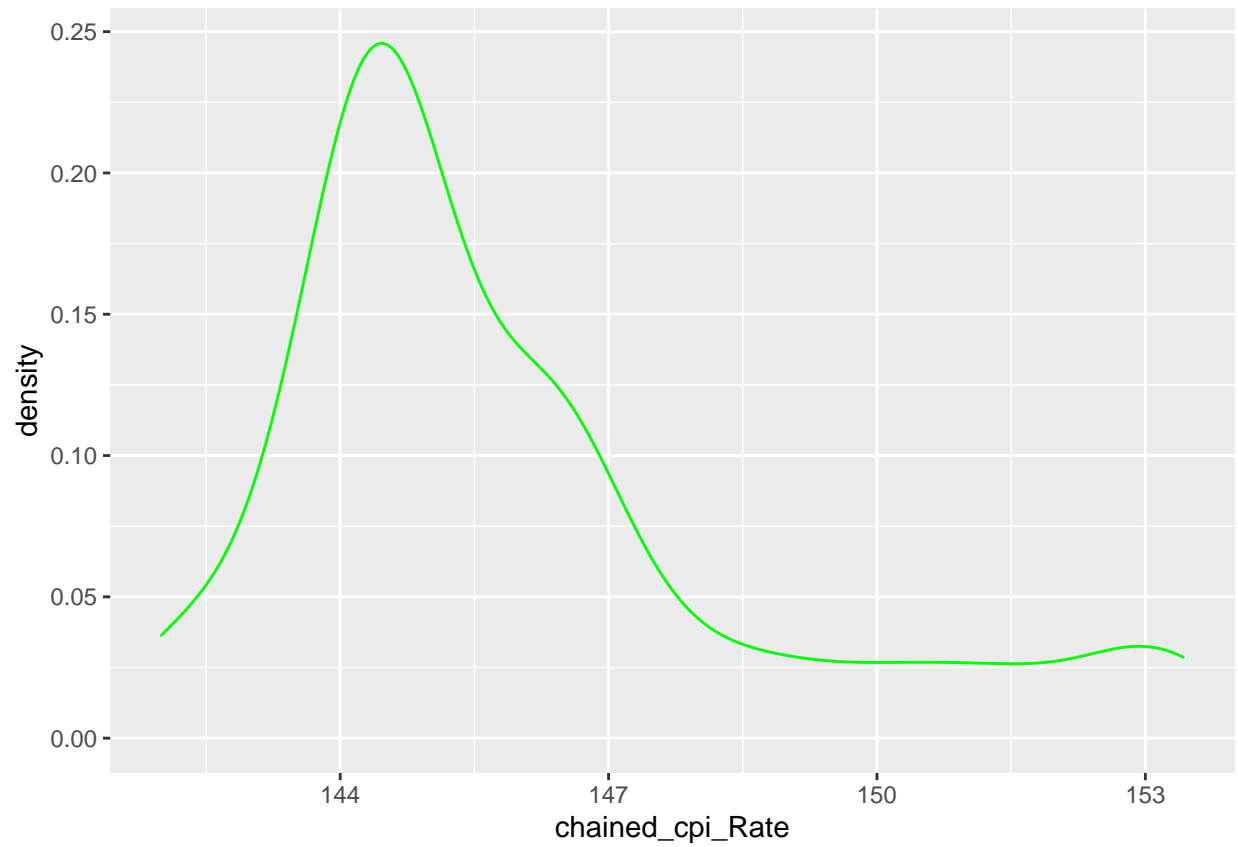


#Density plots of the 2 variables

```
ggplot(df1, aes(x = urban_cpi_Rate)) + geom_density(col = c("blue"))
```



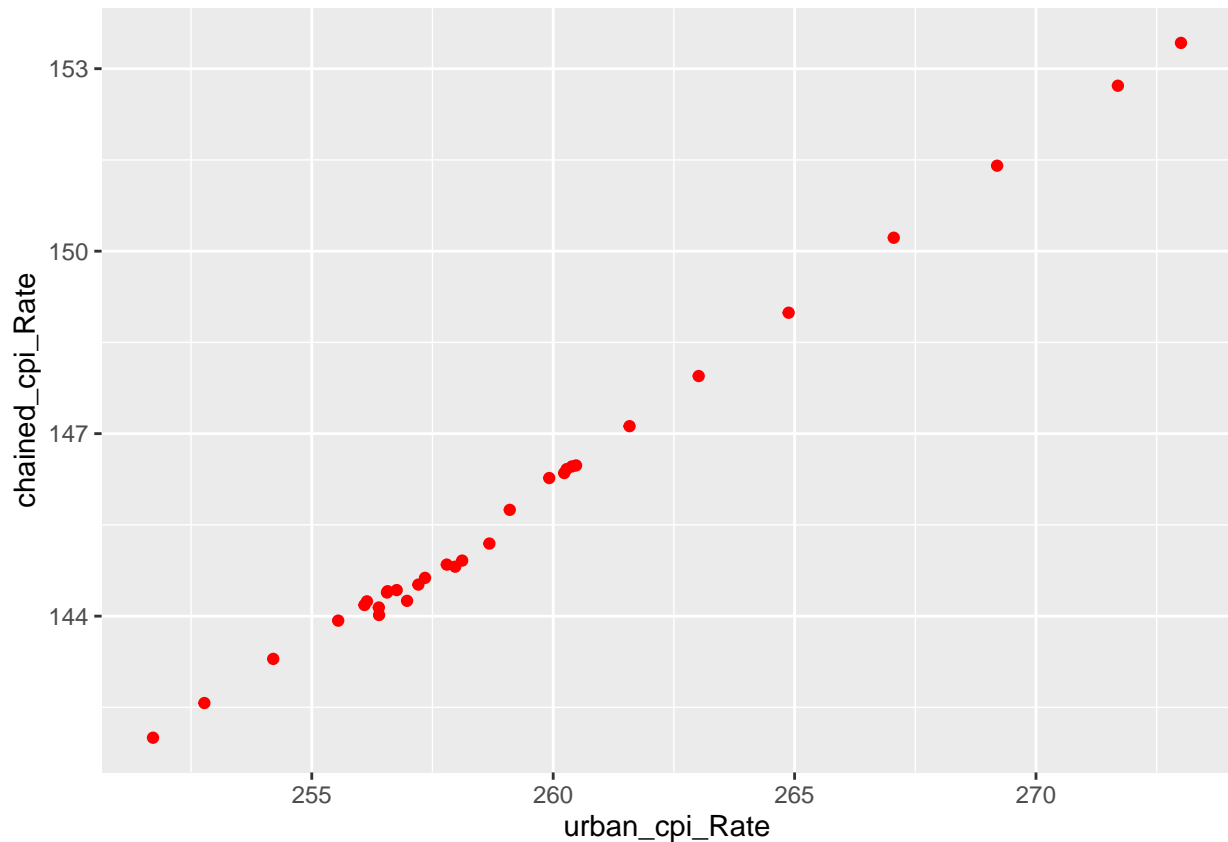
```
ggplot(df1, aes(x = chained_cpi_Rate)) + geom_density(col = c("green"))
```



```
#Save the data-frame as a CSV file
```

```
write.csv(df1, 'C:/BU/DSC630/Consumer_Price_Index.csv')
```


2. Bivariate plots and correlation



```
## Warning: package 'tidyverse' was built under R version 4.0.5
## -- Attaching packages ----- tidyverse 1.3.1 --
## v tibble  3.1.4      v purrr   0.3.4
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
## Warning: package 'tibble' was built under R version 4.0.5
## Warning: package 'tidyr' was built under R version 4.0.5
## Warning: package 'readr' was built under R version 4.0.5
## Warning: package 'forcats' was built under R version 4.0.5
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## [1] 0.9981076
## Warning: package 'gmodels' was built under R version 4.0.5
##
##
##      Cell Contents
## |-----|
## |                               N |
```

```

## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
##
##
## Total Observations in Table:  31
##
##
##           | df1$chained_cpi_Rate
## df1$urban_cpi_Rate | 142.001 | 142.571 | 143.297 | 143.926 | 144.018 | 144.142 | 144.143
## -----|-----|-----|-----|-----|-----|-----|-----
##           251.712 |      1 |      0 |      0 |      0 |      0 |      0 |
##           | 29.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032
##           | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.032 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           252.776 |      0 |      1 |      0 |      0 |      0 |      0 |
##           | 0.032 | 29.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032
##           | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.032 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           254.202 |      0 |      0 |      1 |      0 |      0 |      0 |
##           | 0.032 | 0.032 | 29.032 | 0.032 | 0.032 | 0.032 | 0.032
##           | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.032 | 0.000 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           255.548 |      0 |      0 |      0 |      1 |      0 |      0 |
##           | 0.032 | 0.032 | 0.032 | 29.032 | 0.032 | 0.032 | 0.032
##           | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 0.032 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           256.092 |      0 |      0 |      0 |      0 |      0 |      0 |
##           | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           256.143 |      0 |      0 |      0 |      0 |      0 |      0 |
##           | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000
## -----|-----|-----|-----|-----|-----|-----|-----
##           256.389 |      0 |      0 |      0 |      0 |      0 |      1 |
##           | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 | 29.032
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000
##           | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.032

```


##	-----	-----	-----	-----	-----	-----	-----	-----
##	263.014	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	264.877	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	267.054	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	269.195	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	271.696	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	273.003	0	0	0	0	0	0	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##		0.000	0.000	0.000	0.000	0.000	0.000	0.000
##	-----	-----	-----	-----	-----	-----	-----	-----
##	Column Total	1	1	1	1	1	1	
##		0.032	0.032	0.032	0.032	0.032	0.032	0.032
##	-----	-----	-----	-----	-----	-----	-----	-----
##								
##								

3. Organize a Data Report

```
summary(df1)
```

```
##      year      period      periodName.x      urban_cpi_Rate
## Length:31    Length:31    Length:31      Min.   :251.7
## Class :character Class :character Class :character 1st Qu.:256.5
## Mode  :character Mode  :character Mode  :character Median :258.0
##                                     Mean   :259.5
##                                     3rd Qu.:260.4
##                                     Max.   :273.0
```

```
##   seriesID.x      periodName.y      chained_cpi_Rate  seriesID.y
## Length:31      Length:31      Min.   :142.0      Length:31
## Class :character Class :character 1st Qu.:144.2      Class :character
## Mode  :character Mode  :character Median :144.8      Mode  :character
##                                     Mean  :145.9
##                                     3rd Qu.:146.5
##                                     Max.   :153.4

str(df1)

## 'data.frame':   31 obs. of  8 variables:
## $ year          : chr  "2019" "2019" "2019" "2019" ...
## $ period        : chr  "M01" "M02" "M03" "M04" ...
## $ periodName.x   : chr  "January" "February" "March" "April" ...
## $ urban_cpi_Rate : num  252 253 254 256 256 ...
## $ seriesID.x     : chr  "CUUR0000SA0" "CUUR0000SA0" "CUUR0000SA0" "CUUR0000SA0" ...
## $ periodName.y   : chr  "January" "February" "March" "April" ...
## $ chained_cpi_Rate: num  142 143 143 144 144 ...
## $ seriesID.y     : chr  "SUUR0000SA0" "SUUR0000SA0" "SUUR0000SA0" "SUUR0000SA0" ...

dim(df1)

## [1] 31  8
```

Discussion points

1. The median of urban_cpi_Rate is 258 where as chained_cpi_Rate stands 144.8
2. Bivariate plots resembles the positive correlation.
3. The pearson correlation coefficient is calculated as 0.9981076. it means every positive increase in urban_cpi_Rate there is a positive increase in chained_cpi_Rate
4. The box plots depicts the positive skewness.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.