Assignment 1.2 - R Refresher

Ragunath Gunasekaran

2021-03-21

# 1. Import, Plot, Summarize, and Save Data and summary for every variable, structure and type of data elements

The below is the Structure of datasets which explains the detail of metadata.

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.800 5.100 6.850 6.905 8.550 11.700

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2001 2003 2004 2004 2006 2008

## Length Class Mode   
## 192 character character

## 'data.frame': 180 obs. of 7 variables:  
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ year : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...  
## $ period : chr "M12" "M11" "M10" "M09" ...  
## $ periodName: chr "December" "November" "October" "September" ...  
## $ value : num 6.9 7.1 6.8 7 7.1 7 7 7.1 6.9 7.1 ...  
## $ footnotes : logi NA NA NA NA NA NA ...  
## $ seriesID : chr "LNS13327709" "LNS13327709" "LNS13327709" "LNS13327709" ...

## 'data.frame': 192 obs. of 7 variables:  
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ year : int 2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 ...  
## $ period : chr "M12" "M11" "M10" "M09" ...  
## $ periodName: chr "December" "November" "October" "September" ...  
## $ value : num 13.6 12.7 11.8 11.1 10.8 10.5 10 9.7 9.2 9.1 ...  
## $ footnotes : logi NA NA NA NA NA NA ...  
## $ seriesID : chr "LNS13327709" "LNS13327709" "LNS13327709" "LNS13327709" ...

## 'data.frame': 192 obs. of 7 variables:  
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...  
## $ period : chr "M12" "M11" "M10" "M09" ...  
## $ periodName: chr "December" "November" "October" "September" ...  
## $ value : num 9.2 9.4 9.6 9.7 9.6 9.6 9.5 9.9 9.9 9.8 ...  
## $ footnotes : logi NA NA NA NA NA NA ...  
## $ seriesID : chr "LNS13327709" "LNS13327709" "LNS13327709" "LNS13327709" ...

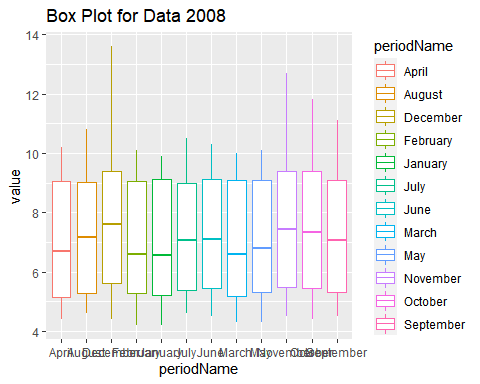
## [1] 180 7

## [1] 192 7

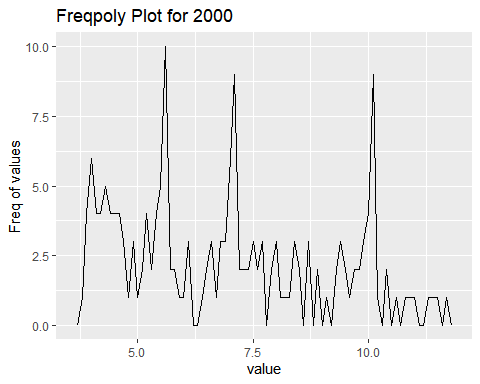
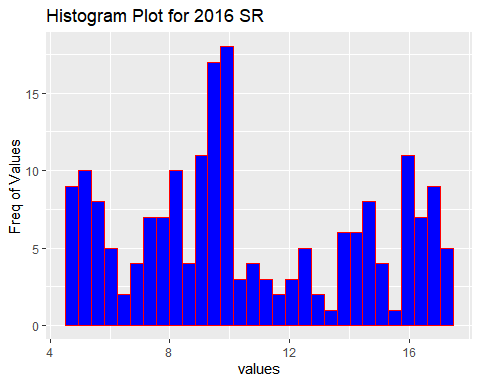
## [1] 192 7

The Dataset contains 180 rows and 6 Variables in 2000 data

## 1. Histograms, Box plots



## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## X year period periodName   
## Min. : 1.00 Min. :1993 Length:180 Length:180   
## 1st Qu.: 45.75 1st Qu.:1995 Class :character Class :character   
## Median : 90.50 Median :1997 Mode :character Mode :character   
## Mean : 90.50 Mean :1997   
## 3rd Qu.:135.25 3rd Qu.:1999   
## Max. :180.00 Max. :2000   
## value footnotes seriesID   
## Min. : 3.800 Mode:logical Length:180   
## 1st Qu.: 5.100 NA's:180 Class :character   
## Median : 6.850 Mode :character   
## Mean : 6.905   
## 3rd Qu.: 8.550   
## Max. :11.700

## X year period periodName   
## Min. : 1.00 Min. :2001 Length:192 Length:192   
## 1st Qu.: 48.75 1st Qu.:2003 Class :character Class :character   
## Median : 96.50 Median :2004 Mode :character Mode :character   
## Mean : 96.50 Mean :2004   
## 3rd Qu.:144.25 3rd Qu.:2006   
## Max. :192.00 Max. :2008   
## value footnotes seriesID   
## Min. : 4.200 Mode:logical Length:192   
## 1st Qu.: 5.300 NA's:192 Class :character   
## Median : 7.300 Mode :character   
## Mean : 7.229   
## 3rd Qu.: 9.200   
## Max. :13.600

## X year period periodName   
## Min. : 1.00 Min. :2009 Length:192 Length:192   
## 1st Qu.: 48.75 1st Qu.:2011 Class :character Class :character   
## Median : 96.50 Median :2012 Mode :character Mode :character   
## Mean : 96.50 Mean :2012   
## 3rd Qu.:144.25 3rd Qu.:2014   
## Max. :192.00 Max. :2016   
## value footnotes seriesID   
## Min. : 4.700 Mode:logical Length:192   
## 1st Qu.: 7.775 NA's:192 Class :character   
## Median : 9.750 Mode :character   
## Mean :10.568   
## 3rd Qu.:14.325   
## Max. :17.200

## X year period periodName   
## Min. : 1.00 Min. :2009 Length:192 Length:192   
## 1st Qu.: 48.75 1st Qu.:2011 Class :character Class :character   
## Median : 96.50 Median :2012 Mode :character Mode :character   
## Mean : 96.50 Mean :2012   
## 3rd Qu.:144.25 3rd Qu.:2014   
## Max. :192.00 Max. :2016   
## value footnotes seriesID   
## Min. : 4.700 Mode:logical Length:192   
## 1st Qu.: 7.775 NA's:192 Class :character   
## Median : 9.750 Mode :character   
## Mean :10.568   
## 3rd Qu.:14.325   
## Max. :17.200

### 2. Explore Some Bivariate Relations

## Warning: package 'tidyverse' was built under R version 4.0.2

## -- Attaching packages ----------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v tibble 3.0.3 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0  
## v purrr 0.3.4

## Warning: package 'tibble' was built under R version 4.0.2

## Warning: package 'tidyr' was built under R version 4.0.2

## Warning: package 'dplyr' was built under R version 4.0.2

## Warning: package 'stringr' was built under R version 4.0.3

## -- Conflicts -------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x tidyr::extract() masks pastecs::extract()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks pastecs::first()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks pastecs::last()

## [1] 0.007213379

## [1] -0.5425109

## [1] -0.5425109

### 3. Organize a Data Report

summary(bls\_2000)

## X year period periodName   
## Min. : 1.00 Min. :1993 Length:180 Length:180   
## 1st Qu.: 45.75 1st Qu.:1995 Class :character Class :character   
## Median : 90.50 Median :1997 Mode :character Mode :character   
## Mean : 90.50 Mean :1997   
## 3rd Qu.:135.25 3rd Qu.:1999   
## Max. :180.00 Max. :2000   
## value footnotes seriesID   
## Min. : 3.800 Mode:logical Length:180   
## 1st Qu.: 5.100 NA's:180 Class :character   
## Median : 6.850 Mode :character   
## Mean : 6.905   
## 3rd Qu.: 8.550   
## Max. :11.700

summary(bls\_2008)

## X year period periodName   
## Min. : 1.00 Min. :2001 Length:192 Length:192   
## 1st Qu.: 48.75 1st Qu.:2003 Class :character Class :character   
## Median : 96.50 Median :2004 Mode :character Mode :character   
## Mean : 96.50 Mean :2004   
## 3rd Qu.:144.25 3rd Qu.:2006   
## Max. :192.00 Max. :2008   
## value footnotes seriesID   
## Min. : 4.200 Mode:logical Length:192   
## 1st Qu.: 5.300 NA's:192 Class :character   
## Median : 7.300 Mode :character   
## Mean : 7.229   
## 3rd Qu.: 9.200   
## Max. :13.600

# Conclusion - Summary of Analysis

As part of the analysis, we have compared US Unemployment Rate from 2000 to 2020 with various period.

# References:

1. <https://cran.r-project.org/web/packages/blscrapeR/index.html>
2. <http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization>