

Assignment_Week_9&10_Venkidusamy_KesavAdithya

Kesav Adithya Venkidusamy

2022/08/05

```
knitr::opts_chunk$set(echo = TRUE)
```

```
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(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
```

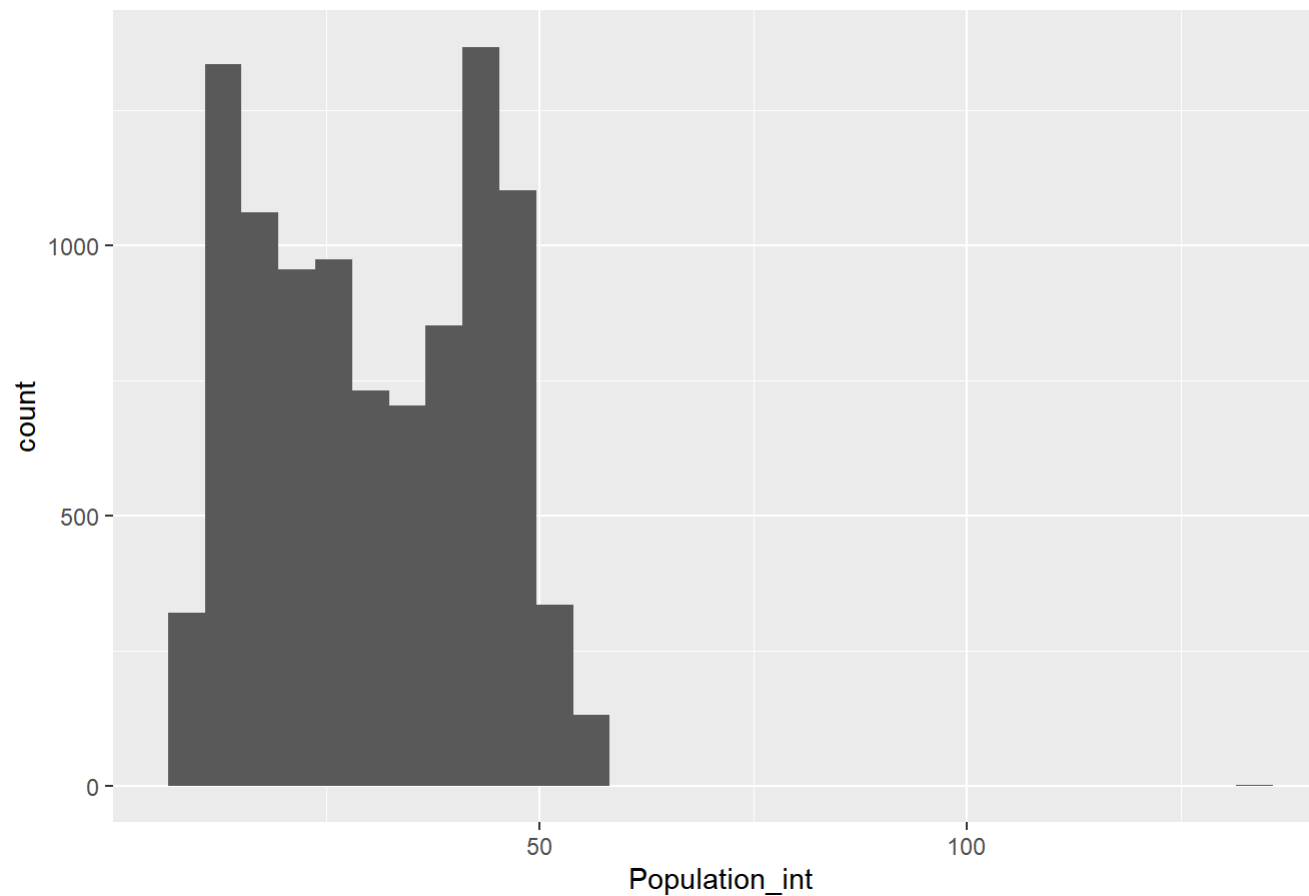
R: Histogram Plot

```
# Creating dataframe  
birth_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 9&10/birth-rate.csv")  
  
# Format year column  
colnames(birth_df) <- gsub("X", "", colnames(birth_df))  
  
## Pivoting the birth dataframe  
birtht_df <- reshape2::melt(birth_df, 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"))  
  
ggplot(birtht_df, aes(x=Population_int)) + geom_histogram() + ggtitle("R - Histogram plot to show the count of Birth Rate")  
+ theme(plot.title = element_text(hjust=0.5))
```

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

```
## Warning: Removed 1596 rows containing non-finite values (stat_bin).
```

R - Histogram plot to show the count of Birth Rate



R: Box Plot

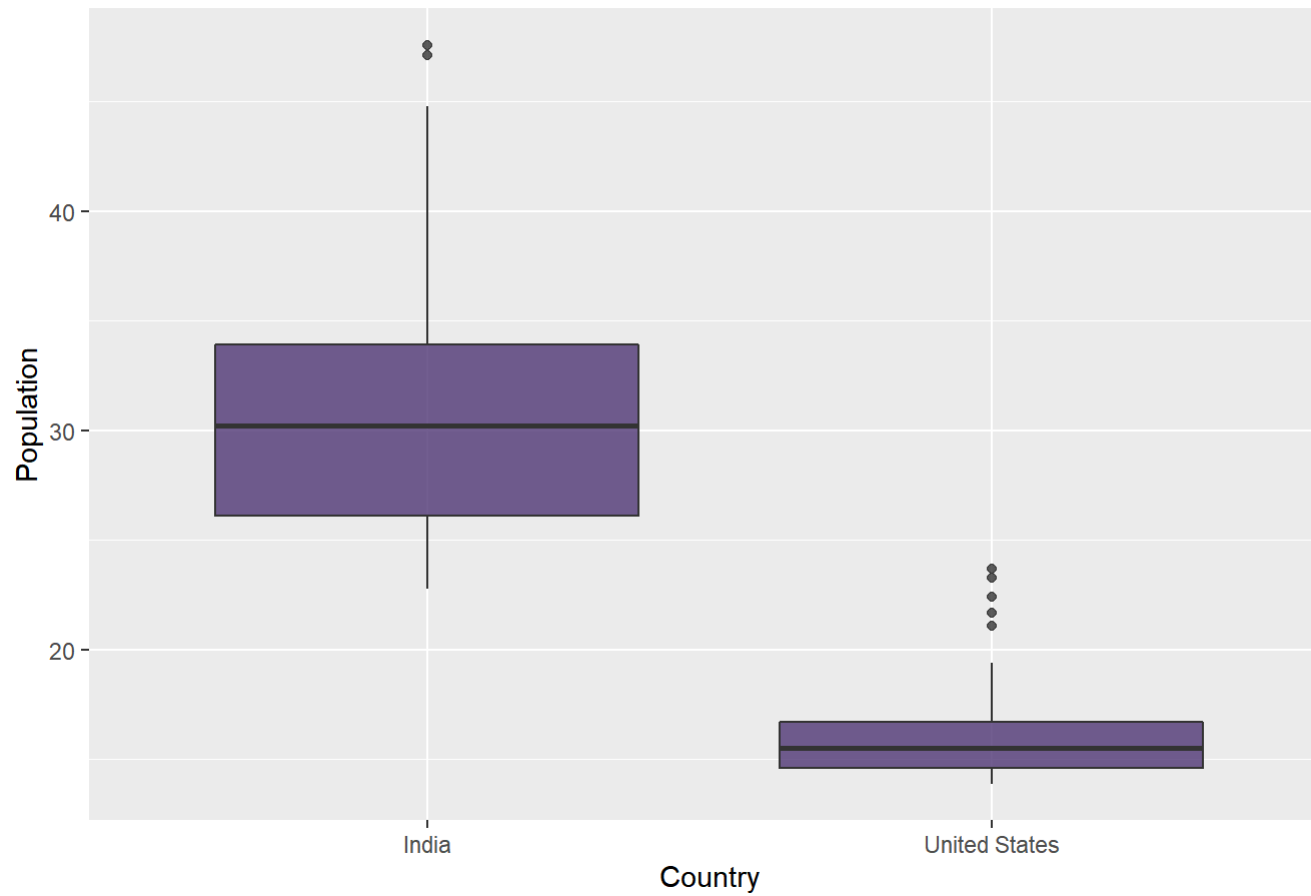
```
## Create box plot
```

```
birth_box_df <- birtht_df %>% dplyr::filter(Country %in% c("United States","India"))
```

```
ggplot(birth_box_df, aes(x=Country, y=Population)) +  
  geom_boxplot(fill="#4f3674", alpha=0.8) + ggtitle("R - Box plot to show outliers in Birth Rate for India and US")
```

```
## Warning: Removed 12 rows containing non-finite values (stat_boxplot).
```

R - Box plot to show outliers in Birth Rate for India and US



R: Bullet Chart

```
# Creating dataframe
crime_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 9&10/crimeratesbystate-formatted.csv")

crime_bullet <- crime_df %>% dplyr::filter(stringr::str_trim(state, 'both') == "Texas") %>% dplyr::select(c(state, burglary))

maxburglary <- max(crime_df$burglary)

fig <- plot_ly(
  type = "indicator",
  mode = "number+gauge+delta",
  value = crime_bullet$burglary,
  textposition = 'middle left',
  domain = list(x = c(0, 1), y= c(0, 1)),
  title = list(text = "Texas \nBurglary", font = list(size = 12)),
  delta = list(reference = 300),
  gauge = list(
    shape = "bullet",
    axis = list(range = list(NULL, 1500)),
    threshold = list(
      line = list(color = "red", width = 2),
      thickness = 0.75,
      value = maxburglary),
    steps = list(
      list(range = c(0, 500), color = "gray"),
      list(range = c(500, 1000), color = "lightgray"),
      list(range = c(1000, 1500), color = "white")),
    bar = list(color = "black")),
  height = 100, width = 800)
fig <- fig %>%
  layout(margin = list(l= 100, r= 10))
fig <- fig %>%
  layout(title="R: Bullet Chart to show Burglary in Texas Compared to US Max Score", font = list(align = 'left'))

fig
```

```
## Warning: 'indicator' objects don't have these attributes: 'textposition'
## Valid attributes include:
## 'align', 'customdata', 'customdatasrc', 'delta', 'domain', 'gauge', 'ids', 'idssrc', 'legendgrouptitle', 'legendrank', 'meta', 'metasrc', 'mode', 'name', 'number', 'stream', 'title', 'transforms', 'type', 'uid', 'uirevision', 'value', 'visible', 'key', 'set', 'frame', 'transforms', '_isNestedKey', '_isSimpleKey', '_isGraticule', '_bbox'
```

R: Bullet Chart to show Burglary in Texas Compared to US Max Score



R: Word Cloud

```
# Load Libraries
library(tm)
```

```
## Warning: package 'tm' was built under R version 4.1.3
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

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

```
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 4.1.3
```

```
## Loading required package: RColorBrewer
```

```
library(SnowballC)
options(warn=-1)

# Read the data from file
airline_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 9&10/airline_safety.txt")

# Create Corpus
corp <- VCorpus(VectorSource(airline_df))

# Clean up text data
corp <- tm_map(corp, removeNumbers)
corp <- tm_map(corp, removePunctuation)
corp <- tm_map(corp, stripWhitespace)
corp <- tm_map(corp, content_transformer(tolower))
corp <- tm_map(corp, removeWords, stopwords("english"))

# Create a document-term-matrix
dtm <- TermDocumentMatrix(corp)
matrix <- as.matrix(dtm)
words <- sort(rowSums(matrix), decreasing = TRUE)
df <- data.frame(words=names(words), freq=words)

# Generate word cloud
wordcloud(words = df$words, freq=df$freq, min.freq = 1, max.words = 100, random.order = FALSE, colors = brewer.pal(8, "Dark 2"))
```

