Pradeep Sahoo Home Work 1

Sahoo Pradeep

6/12/2018

My Github repository for my assignments can be found at this URL: My Github

```
library(mdsr)
library(tidyverse)
```

```
data(WorldCities)
```

QUESTION 1: How many observations and variables are there in this data set? What are some of the variable names? Type up your answer in your document using complete sentences.

```
glimpse(WorldCities)
```

```
## Observations: 23,018
## Variables: 10
## $ code          <int> 3040051, 3041563, 290594, 291074, 291696, 292223...
## $ name          <chr> "les Escaldes", "Andorra la Vella", "Umm al Qayw...
## $ latitude      <dbl> 42.50729, 42.50779, 25.56473, 25.78953, 25.33132...
## $ longitude     <dbl> 1.53414, 1.52109, 55.55517, 55.94320, 56.34199, ...
## $ country       <chr> "AD", "AD", "AE", "AE", "AE", "AE", "AE", "AE", ...
## $ countryRegion <chr> "8", "7", "7", "5", "6", "3", "4", "6", "1", "4"...
## $ population    <dbl> 15853, 20430, 44411, 115949, 33575, 1137347, 263...
## $ regionCode    <int> 1033, 1037, 2, 2, 20, 11, 4, 6, 16, 15, 275, 4, ...
## $ region        <chr> "Europe/Andorra", "Europe/Andorra", "Asia/Dubai"...
## $ date          <chr> "10/15/08", "5/30/10", "11/3/12", "11/30/12", "1...
```

```
WorldCities <- head(WorldCities, 200) # 200 rows
```

```
country_col <- WorldCities$country
unique(country_col)
```

```
## [1] "AD" "AE" "AF" "AG" "AI" "AL" "AM" "AO" "AR"
```

QUESTION 2: There is a column called region in the data set. Can you extract this and show only the unique values?

```
country_col <- select(WorldCities, country)
country_col <- WorldCities %>% select(country)
```

```
WorldCities %>% select(region) %>% head(5)
```

```
##           region
## 1 Europe/Andorra
## 2 Europe/Andorra
## 3     Asia/Dubai
## 4     Asia/Dubai
```

```
## 5     Asia/Dubai
```

QUESTION 3: Can you extract and show the unique entries from the country column in WorldCities using one line of code and two %>% operators?

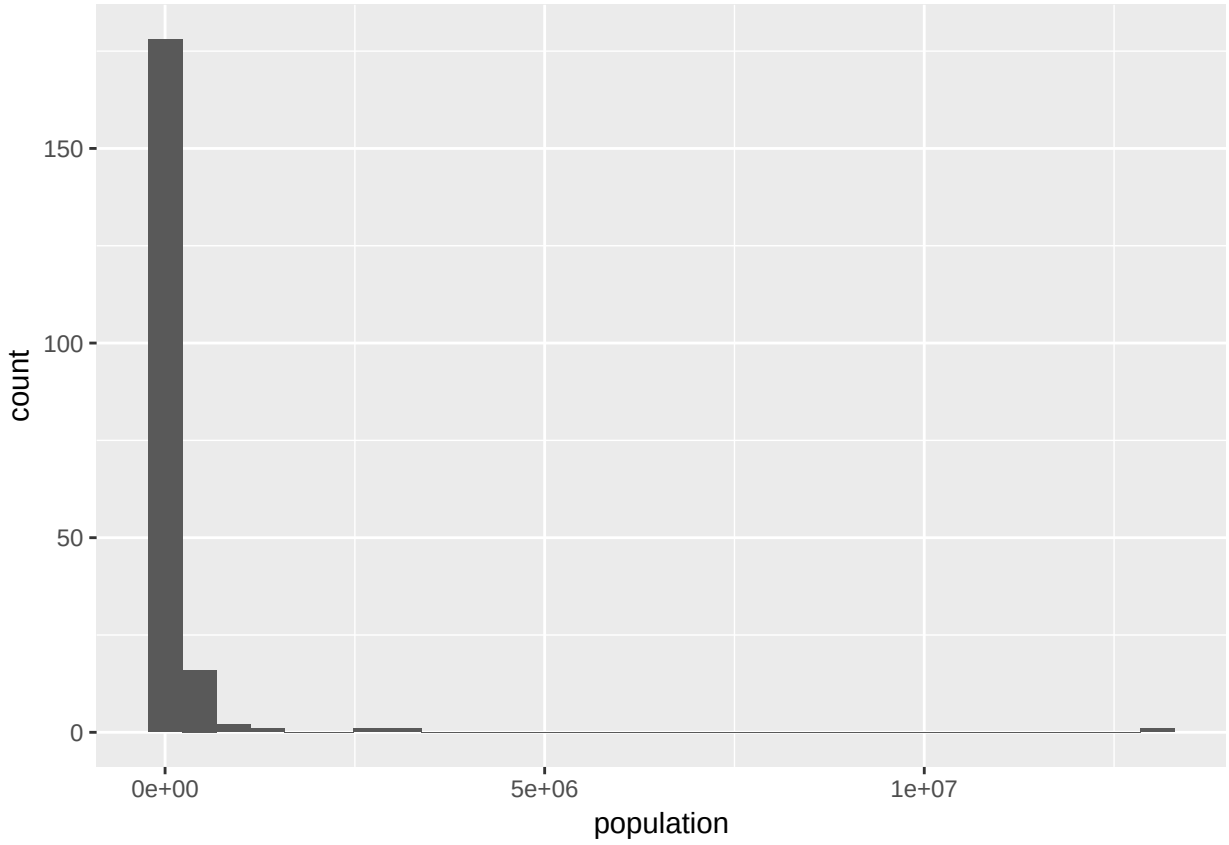
```
WorldCities %>% select(country) %>% unique()
```

```
##     country
## 1        AD
## 3        AE
## 15       AF
## 65       AG
## 66       AI
## 67       AL
## 87       AM
## 104      AO
## 131      AR
```

Visualize it

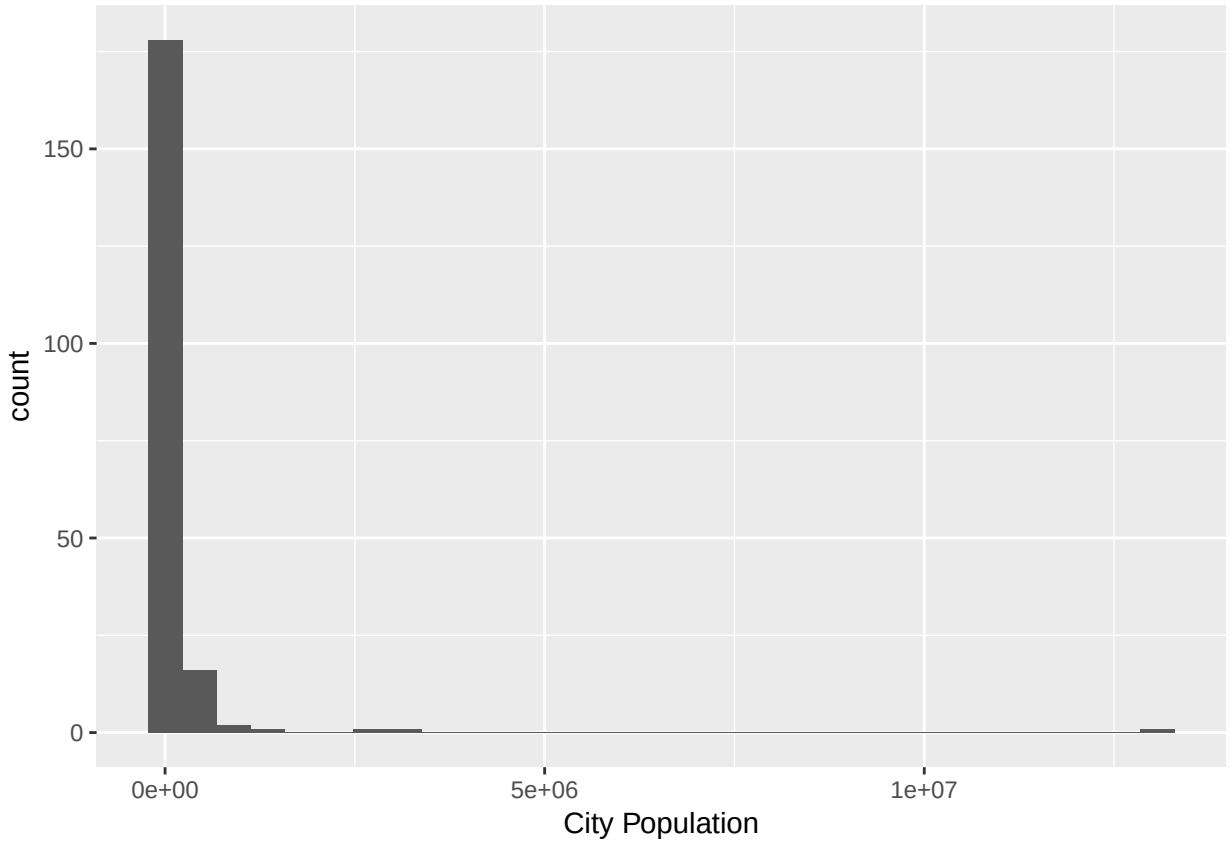
```
WorldCities %>% ggplot(aes(x = population)) +
  geom_histogram()
```

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



```
WorldCities %>% ggplot(aes(x = population)) +
  geom_histogram() +
  xlab('City Population')
```

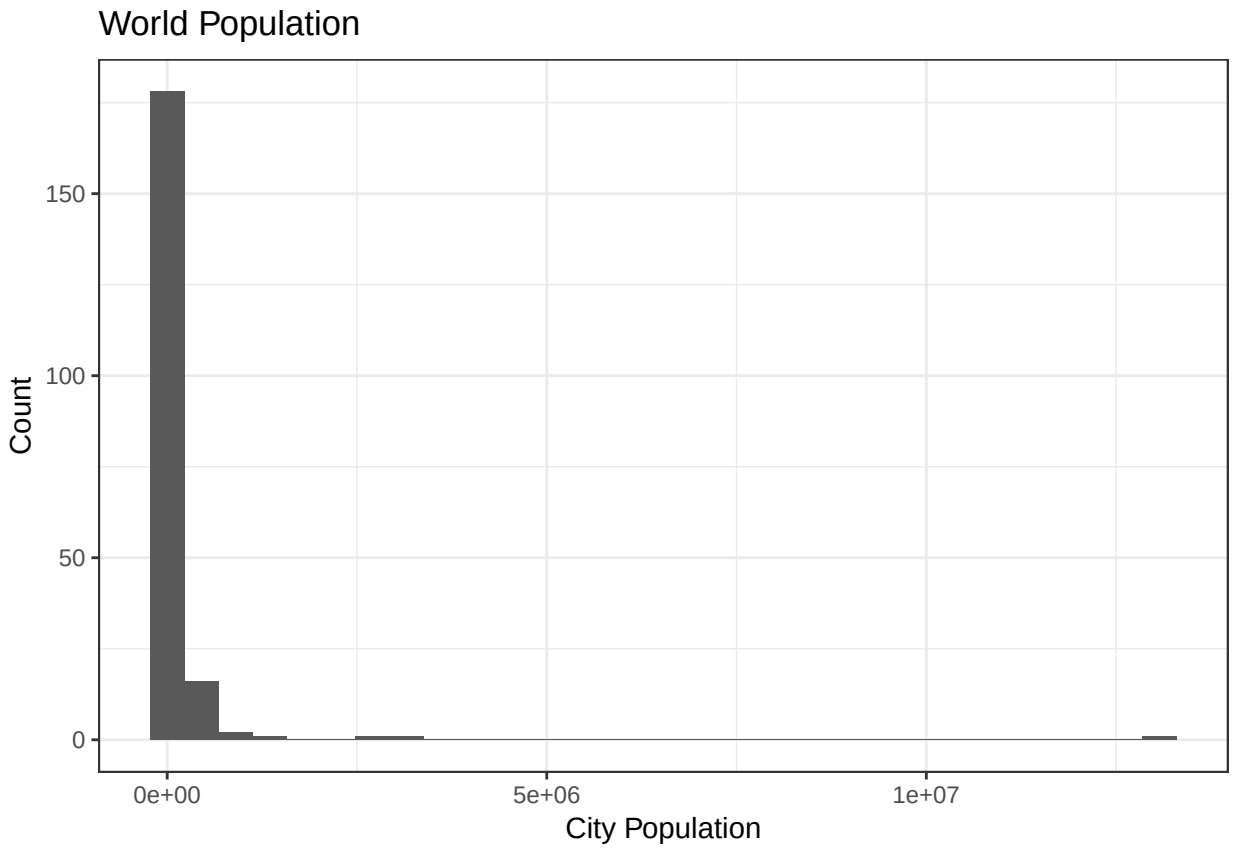
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



QUESTION 4: Make one more histogram of the population and add a new x-label, a new y-label, a new title (use ggtitle('my title')), and change the theme of the plot using theme_bw().

```
WorldCities %>% ggplot(aes(x = population)) +
  geom_histogram() +
  xlab('City Population') +
  ylab('Count') +
  ggtitle('World Population') +
  theme_bw()
```

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


World Population
Count
150
100
50
0
0e+00
5e+06
1e+07
City Population